

Kevin Schneider



# Perspectives on Mutual Dependence in Plant Health

Perspectives on Mutual Dependence in Plant Health

Kevin Schneider

Wageningen University

# Propositions

1. Without data on true pest absences all species distribution modelling analyses are bound to remain modelling exercises.  
(this thesis)
2. Capturing economic heterogeneity by stratification does not only refine the overall estimate of pest impact but also signals whether unequal consequences from pest spread arise to the different strata.  
(this thesis)
3. Good predictive performance does not imply model correctness.
4. Spatially explicit data opens new avenues for interdisciplinary research that address key challenges for society.
5. Lunch time is the most creative hour in the day.
6. Strengthening collective efforts requires sacrificing personal freedom.

Propositions belonging to the thesis, entitled

“Perspectives on Mutual Dependence in Plant Health”

Kevin Schneider

Wageningen, 27<sup>th</sup> of October 2021

# Perspectives on Mutual Dependence in Plant Health

Kevin Schneider



## **Thesis committee**

### **Promotor**

Prof. Dr A.G.J.M. Oude Lansink  
Professor of Business Economics  
Wageningen University & Research

### **Co-promotors**

Dr M.C.M. Mourits  
Associate professor, Business Economics Group  
Wageningen University & Research

Dr W. van der Werf  
Associate professor, Centre for Crop Systems Analysis  
Wageningen University & Research

### **Other members**

Prof. Dr E. Bulte, Wageningen University & Research  
Dr S. Parnell, University of Salford, UK  
Dr J. Schans, Netherlands Food and Consumer Product Safety  
Authority, Utrecht  
Dr S. Janssen, Wageningen University & Research

This research was conducted under the auspices of the Wageningen  
School of Social Sciences (WASS).

# Perspectives on Mutual Dependence in Plant Health

Kevin Schneider

## **Thesis**

submitted in fulfilment of the requirements for the degree of doctor  
at Wageningen University  
by the authority of the Rector Magnificus,  
Prof. Dr. A.P.J. Mol,  
in the presence of the  
Thesis Committee appointed by the Academic Board  
to be defended in public  
on Wednesday 27 October 2021  
at 11 a.m. in the Aula.

Kevin Schneider  
Perspectives on Mutual Dependence in Plant Health  
**334** pages

PhD thesis, Wageningen University, Wageningen, Netherlands (2021)  
With references, with summary in English.

ISBN: 978-94-6395-976-6

DOI: <https://doi.org/10.18174/553468>



*Dedicated to Birgit Schneider and  
Michael Beuttel*



# Acknowledgements

I remember vividly how compliments of my Bachelor thesis supervisor, slowly but steadily, convinced me that I really understood a lot of what I was doing. Over the course of my Masters, I started reading all recommended textbooks for each course I took. It slowly started daunting on me that I really knew close to nothing and that I should better start to convey my uncertainty whenever I was asked to express my opinion/knowledge on anything at all. I spent the last four years working towards this PhD. I read scientific literature and textbooks, completed coursework, and applied newly acquired skills to address scientific questions. Thanks to all this hard work, I can confidently say that I am now certain that I really do not know very much at all. Fortunately, I realized that working with the right people still allows you to combine knowledge, overcome knowledge-gaps, and collectively achieve great results. Here, I want to thank everyone that helped me in the last years.

First, I would like to thank my promotor Prof. Dr Alfons Oude Lansink. On my first day in office, I addressed Alfons following German standards with Professor Lansink. He quickly brushed it off and informed me that I can call him Alfons. I did not know it at the time, but this gesture very much predicted my future working relationship with Alfons. Throughout the years, Alfons has always been a positive force behind the work I was doing. He not only supported me tremendously by giving scientific input, but very often Alfons found words that reignited my motivation and prevented me from throwing

everything into the (digital) trash bin. Despite Alfons being a very busy man, I could always rely on open doors and prompt responses whenever I had questions or concerns. Through his optimism, eye for detail, and his can-do-attitude, Alfons was not only a mentor to me when it comes to scientific rigour, but more broadly on how I want to approach teamwork and the management of people. I could not have asked for a better promotor.

I would also like to thank my daily supervisor Dr Monique Mourits. From the start, Monique has found encouraging and motivating words that helped me a lot through difficult stages of my research. Initially to my surprise, Monique was always protective of my weekends and my holidays. Her attention to my schedule and her cautionary words prevented me from overworking several times. I truly believe that Monique's caring was critical to my endurance throughout the PhD. In dozens of meetings, Monique had deeply reflected on my drafts and jumped in to have my back whenever I struggled to explain my ideas to others. Monique has shown me the patience and time-commitment that is demanded of teachers at a university. The organizational ability of successfully managing courses with several hundred students is a skill that I did not even have on my radar when I started as a PhD student. Monique perfectly complemented my supervising team and gave me a humbling understanding of the diverse skillset that is required to succeed in academia.

I would also like to thank my co-promotor Dr Wopke van der Werf. During my hiring-interview, Wopke threw questions at me that I was very much unprepared for and during the start of my project these types of critical questions would frequently come up in meetings with him. Wopke struck me as a character, and someone I had problems fully understanding in the beginning. However, little did I know that these questions were, in a sense, intellectual sparring exercises that greatly equipped me to better reflect on my decisions and communicate my ideas in clearer ways. If science is music and a collaboration is a band, Wopke is playing jazz and once you understand that working with him is incredibly rewarding. While brainstorming with Wopke, I felt completely at ease pitching (not seldom stupid) ideas on how to approach problems. Wopke would do the same and while doing so utilize his broad exposure to different scientific disciplines and modelling approaches. As Monique and Alfons, Wopke has always

been a very encouraging and motivating force behind my work. Wopke had a very strong influence in shaping most of my research chapters, he introduced me to many people I hope to work with in the future, and through his lobbying I ended up presenting in the right meeting at the right time which resulted in an exciting next step for my career. For all these things, I will always be gratefully indebted to him.

My time at Wageningen University was positively shaped by many others. Over the four years, I have learned to appreciate the importance of having a strong administrative backbone to enable a great scientific group. Anne Houwers and Jeannette Lubbers-Poortvliet always had open doors for any questions I had. With their help, I ventured across Europe and even to Florida effortlessly. I also want to thank a few colleagues of the Business Economics Group. Conversations with Frederic Ang have always been incredible. If all scholars would share Frederic's passion for research and his commitment to quality, academia would be a better place. I thank Melina Lamkowsky, Francis Edwardes, Thomas Slijper, Xinxin Wang, Zhengcong Wang, and Xiaomei Yue for being great colleagues. I thank Robbert van den Dool and Lia Hemerik for great discussions and their reflection on my work. I thank my Spanish colleagues from the POnTE project Martina Cendoya, Antonio Vicent, and Juan A. Navas-Cortés for their invaluable input and their hospitality while welcoming me into the project.

I thank Ruozhu Han for all her support and caring. I know for a fact that my compulsive character and my, sometimes excessive, fixation on work causes strain on those close to me. Without Ruozhu's understanding and her support, I could not have achieved many of the sprints that I felt were necessary to get my research to where I wanted it to be. Lastly, I want to thank my mother Birgit Schneider and her partner Michael Beuttel for supporting me throughout my life and for enabling this journey of mine. Without you, I would not be where I am today. I dedicate this book to both of you.



# Chapter Publications

Schneider, K., Skevas, I. and Oude Lansink, A., 2021. Spatial Spillovers on Input-specific Inefficiency of Dutch Arable Farms. *Journal of Agricultural Economics*, 72(1), pp.224-243. <https://doi.org/10.1111/1477-9552.12400>

Schneider, K., Van der Werf, W., Cendoya, M., Mourits, M., Navas-Cortés, J.A., Vicent, A. and Oude Lansink, A., 2020. Impact of *Xylella fastidiosa* subspecies *pauca* in European olives. *Proceedings of the National Academy of Sciences*, 117(17), pp.9250-9259. <https://doi.org/10.1073/pnas.1912206117>

Schneider, K., Mourits, M., van der Werf, W. and Oude Lansink, A., 2021. On consumer impact from *Xylella fastidiosa* subspecies *pauca*. *Ecological Economics*, 185, p.107024. <https://doi.org/10.1016/j.ecolecon.2021.107024>

Schneider, K., Makowski, D. and van der Werf, W., 2021. Predicting hotspots for invasive species introduction in Europe. *Environmental Research Letters* (revised and submitted)



# Contents

<b>1</b>	<b>General Introduction</b> .....	1
1.1	Plant Health Policies .....	3
1.2	Economic Dependence .....	4
1.3	Spatial Dependence .....	6
1.4	Models and Uncertainties .....	7
1.5	Problem Statement .....	9
1.6	Objective and Research Questions .....	9
<b>2</b>	<b>Spatial Spill Overs on Input Specific Inefficiency</b> .....	13
2.1	Introduction .....	14
2.2	Materials and Methods .....	17
2.2.1	Directional Distance Function .....	17
2.2.2	Determinants of Inefficiency .....	19
2.2.3	Estimation .....	21
2.2.4	Data .....	23
2.3	Results .....	26
2.3.1	Directional Distance Function .....	26
2.3.2	Determinants of Inefficiency .....	27
2.4	Discussion .....	32
2.5	Conclusions .....	36
<b>3</b>	<b>Impact of <i>Xylella fastidiosa</i> subspecies <i>pauca</i> in European Olives</b> .....	39
3.1	Introduction .....	40

3.2	Results	43
3.3	Discussion	54
3.4	Materials and Methods	58
3.4.1	Climatic Suitability Map	58
3.4.2	Disease Spread Simulation	60
3.4.3	Economic Model	62
3.4.4	Global Sensitivity Analysis	66
3.4.5	Economic Data	67
3.4.6	Data Archival	69
3.5	Supplementary Material	71
<b>4</b>	<b>On Consumer Impact from <i>Xylella fastidiosa</i> subspecies <i>pauca</i></b>	<b>89</b>
4.1	Introduction	90
4.2	Materials and Methods	92
4.2.1	Climatic Suitability Map	92
4.2.2	Disease Spread Simulation	94
4.2.3	Economic Model	95
4.2.4	Data	98
4.2.5	Sensitivity Analyses	100
4.3	Results	101
4.4	Discussion	109
4.5	Conclusions	112
4.6	Supplementary Material	115
<b>5</b>	<b>Predicting hotspots for invasive species introduction in Europe</b>	<b>121</b>
5.1	Introduction	121
5.2	Results and Discussion	124
5.2.1	European Hotspots for Pest Introductions	125
5.2.2	Feature Contributions	125
5.2.3	Model Performance	128
5.2.4	Sensitivity	129
5.2.5	Implications for Pest Surveys	130
5.3	Materials and Methods	136
5.3.1	Data	136
5.3.2	Data processing	139
5.3.3	Cross-validation techniques	139
5.3.4	Algorithm and hyperparameter tuning	140

5.3.5	Prediction and Mapping .....	142
5.4	Supplementary Material .....	144
<b>6</b>	<b>General Discussion</b> .....	<b>211</b>
6.1	Implications of mutual dependence and heterogeneity	215
6.2	Agri-environmental Data Integration .....	223
6.2.1	Opportunities .....	225
6.2.2	Challenges .....	228
6.3	Policy Implications .....	231
6.4	Limitations and Future Work .....	235
6.5	Main Conclusions .....	240
6.5.1	Main Policy and Research Implications .....	241
<b>References</b>	.....	<b>243</b>
<b>Summary</b>	.....	<b>285</b>
<b>Glossary</b>	.....	<b>291</b>
<b>Index</b>	.....	<b>297</b>
<b>Curriculum Vitae</b>	.....	<b>299</b>
<b>Book Cover</b>	.....	<b>301</b>



# List of Figures

3.1	Binary suitability maps and olive production sites for Italy. ....	44
3.2	Binary suitability maps and olive production sites for Greece. ....	45
3.3	Binary suitability maps and olive production sites for Spain. ....	46
3.4	Simulated geographical distribution of $X_{fp}$ in year 5 (a) and 50 (b) with the median of the elicited rate of radial range expansion (5.18 km per year) and a climate suitability threshold of 0.132 in the ensemble SDM. ....	48
3.5	All randomized points of introduction for Greece and Spain. ....	71
3.6	Points of introduction which resulted in an infected area less than (blue) and more or equal to (red) 30 percent of the Greek area of production for the 5.18 km per year spread rate and a climatic suitability threshold of 0.132 (suitable area in yellow). ....	72
3.7	Uncertainty of the Greek area of production infected over time for the 1.1 km (a), 5.18 km (b) and 12.35 km (c) per year spread rate and a climatic suitability threshold (T2) of 0.132. ....	73

3.8	Points of introduction which resulted in an infected area less than (blue) and more or equal to (red) 50 percent of the Spanish area of production for the 5.18 km per year spread rate and a climatic suitability threshold of 0.132 (suitable area in yellow). . . . .	74
3.9	Uncertainty of the Spanish area of production infected over time for the 1.1 km (a), 5.18 km (b) and 12.35 km (c) per year spread rate and a climatic suitability threshold (T2) of 0.132. . . . .	75
3.10	Points of introduction which resulted in an infected area less than (blue) and more or equal to (red) 50 percent of the Spanish area of production for the 12.35 km per year spread rate and a climatic suitability threshold of 0.093 (suitable area in yellow). . . . .	76
3.11	Uncertainty of the Spanish area of production infected over time for the 1.1 km (a), 5.18 km (b) and 12.35 km (c) per year spread rate and a climatic suitability threshold (T3) of 0.093. . . . .	77
3.12	Distribution of economic results across all random points of introduction in Greece for all spread scenarios.	78
3.13	Distribution of economic results across all random points of introduction in Spain for all spread scenarios.	79
4.1	Change in the consumption of olive oil for spread of Xfp in Italy beyond the current extent. . . . .	102
4.2	Distribution of changes in total welfare and consumer surplus following introductions of Xfp into Greece (a) or Spain (b) and spread over 50 years with a suitability threshold of 0.135. . . . .	106
4.3	Stochastic evaluation of welfare changes under uncertainty of supply elasticity (a), yield decline (b) and cost change (c). The vertical line depicts the deterministic value used. . . . .	108
4.4	Change in the European exports of olive oil for spread of Xfp in Italy beyond the current extent. . . . .	120
5.1	Probability of new introduction of at least one invasive species across Europe. . . . .	126

5.2	Spearman correlation of the European prediction across all combinations for cross-validation and background data generation.....	131
5.3	Difference between maximum and minimum risk level values for Europe across different background generation approaches for models tuned using temporal (a) and continental spatial-blocks (b) for cross-validation. ....	132
5.4	Fifty most important features across different background generation approaches for models tuned using temporal split cross-validations. ....	133
5.5	Fifty most important features across different background generation approaches for models tuned using spatial-block split cross-validations.....	134
5.6	Overview of presence points and randomly generated background points. ....	145
5.7	Overview of presence points and kdbias background points. ....	146
5.8	Overview of presence points and background points generated by the combined approach of biasing and geographically excluding.....	147
5.9	Number of presence records over time and chosen cut-offs for the temporal cross-validation. ....	148
5.10	Presence and random background points used for training (cumulative over time) and validating in the first fold of the temporal cross-validation. ....	149
5.11	Presence and random background points used for training (cumulative over time) and validating in the second fold of the temporal cross-validation. ....	150
5.12	Presence and random background points used for training (cumulative over time) and validating in the third fold of the temporal cross-validation. ....	151
5.13	Presence and random background points used for training (cumulative over time) and validating in the fourth fold of the temporal cross-validation. ....	152
5.14	Presence and random background points used for training (cumulative over time) and validating in the fifth fold of the temporal cross-validation. ....	153

- 5.15 Presence and random background points used for training (cumulative over time) and validating in the sixth fold of the temporal cross-validation. . . . . 154
- 5.16 Overview of presences and random background points falling into the different continental spatial blocks. . . . . 155
- 5.17 Cross validation performance, in terms of area under the ROC, for all approaches for cross validation and background point generation. . . . . 156
- 5.18 Cross validation performance, in terms of Sensitivity, for all approaches for cross validation and background point generation. . . . . 157
- 5.19 Cross validation performance, in terms of Specificity, for all approaches for cross validation and background point generation. . . . . 158
- 5.20 Hyperparameter tuning for random split cross validation and kd05dfar background points. . . . . 159
- 5.21 Hyperparameter tuning for random split cross validation and kdbias background points. . . . . 160
- 5.22 Hyperparameter tuning for randomly split cross validation and random background points. . . . . 161
- 5.23 Hyperparameter tuning for spatial-block cross validation and kd05dfar background points. . . . . 162
- 5.24 Hyperparameter tuning for spatial-block cross validation and kdbias background points. . . . . 163
- 5.25 Hyperparameter tuning for spatial-block cross validation and random background points. . . . . 164
- 5.26 Hyperparameter tuning for temporal cross validation and kd05dfar background points. . . . . 165
- 5.27 Hyperparameter tuning for temporal cross validation and kdbias background points. . . . . 166
- 5.28 Hyperparameter tuning for temporal cross validation and random background points. . . . . 167
- 5.29 Prediction for model tuned using random splits for cross-validation and kd05dfar background points. . . . . 168
- 5.30 Prediction for model tuned using random splits for cross-validation and kdbias background points. . . . . 169
- 5.31 Prediction for model tuned using random splits for cross-validation and random background points. . . . . 170

5.32	Average prediction for model tuned using random splits for cross-validation across all approaches to generate background points. ....	171
5.33	Range between maximum and minimum of the predictions for model tuned using random splits for cross-validation across all approaches to generate background points. ....	172
5.34	Prediction for model tuned using spatial-block splits for cross-validation and kd05dfar background points. ...	173
5.35	Prediction for model tuned using spatial-block splits for cross-validation and kdbias background points. ....	174
5.36	Prediction for model tuned using spatial-block splits for cross-validation and random background points. ...	175
5.37	Average prediction for model tuned using spatial-block splits for cross-validation across all approaches to generate background points. ....	176
5.38	Range between maximum and minimum of the predictions for model tuned using spatial-block splits for cross-validation across all approaches to generate background points. ....	177
5.39	Prediction for model tuned using temporal splits for cross-validation and kd05dfar background points. ....	178
5.40	Prediction for model tuned using temporal splits for cross-validation and kdbias background points. ....	179
5.41	Prediction for model tuned using temporal splits for cross-validation and random background points. ....	180
5.42	Average prediction for model tuned using temporal splits for cross-validation across all approaches to generate background points. ....	181
5.43	Range between maximum and minimum of the predictions for model tuned using temporal splits for cross-validation across all approaches to generate background points. ....	182
5.44	Average prediction across all nine approaches and GBIF presences (red). ....	183
5.45	Fifty most important features across all approaches. ...	184

- 5.46 Fifty most important features across different approaches to generating background points for models tuned on random splits for cross validations. . . . 185
- 5.47 Fifty most important features across different approaches to generating background points for models tuned on spatial-block splits for cross validations. 186
- 5.48 Fifty most important features across different approaches to generating background points for models tuned on temporal splits for cross validations. . . 187
- 5.49 Fifty most important features across different approaches to cross-validation for models trained on kd05dfar background points. . . . . 188
- 5.50 Fifty most important features across different approaches to cross-validation for models trained on kdbias background points. . . . . 189
- 5.51 Fifty most important features across different approaches to cross-validation for models trained on random background points. . . . . 190
  
- 6.1 Tree shape as a raster after removing the background. . 302
- 6.2 Generated geographic points across Europe. . . . . 303
- 6.3 Final cover logo and the corresponding legend. . . . . 306

# List of Tables

2.1	Descriptive statistics.....	24
2.2	Average annual inefficiency scores in percent for the output ( $\beta_y$ ), productive inputs ( $\beta_x$ ) and damage abatement inputs ( $\beta_a$ ). .....	27
2.3	Bootstrapped regression results for productive inputs ( $\beta_x$ ). .....	28
2.4	Bootstrapped regression results for damage abatement inputs ( $\beta_a$ ).....	29
2.5	Marginal effects on productive inputs and damage abatement inputs inefficiency.....	31
3.1	Economic impact over 50 years in billion Euro. ....	51
3.2	First and total order sensitivity indices of significant economic parameters without replanting (S1), with replanting (S2) and resistance benefits (RB) for impact in Italy. ....	54
3.3	Economic parameters for the deterministic analysis as well as the global sensitivity analysis.....	80
3.4	Economic impact over 50 years in billion Euro for all spread scenarios. ....	81
3.5	First order sensitivity indices for all parameters for the economic impact without replanting. ....	82
3.6	Total order sensitivity indices for all parameters for the economic impact without replanting.....	83

3.7	First order sensitivity indices for all parameters for the economic impact with replanting. ....	84
3.8	Total order sensitivity indices for all parameters for the economic impact with replanting. ....	85
3.9	First order sensitivity indices for all parameters for the benefit of resistance breeding. ....	86
3.10	Total order sensitivity indices for all parameters for the benefit of resistance breeding. ....	87
4.1	Overview of the economic parameter values. ....	100
4.2	Change in prices and consumption over 50 years. ....	103
4.3	Changes in welfare over 50 years in billion Euro for all spread scenarios. ....	105
4.4	First and total order sensitivity indices of statistically significant economic parameters ( $P < 0.05$ ). ....	107
4.5	Regression output for the elasticity estimation. ....	115
4.6	Country specific changes in welfare over 50 years in billion Euro for spread scenarios in Italy. ....	116
4.7	Country specific changes in welfare over 50 years in billion Euro for spread scenarios in Greece averaged over all random points of introduction. ....	117
4.8	Country specific changes in welfare over 50 years in billion Euro for spread scenarios in Spain averaged over all random points of introduction. ....	118
4.9	First and total order sensitivity indices for all economic parameter. ....	119
5.1	Overview of model performances for all cross-validation and background generation technique. ....	129
5.2	Overview of features. ....	191
5.3	Overview of raw data with spatial information. ....	203
5.4	Overview of tuned hyperparameters for all approaches to cross validation and the generation of background points. ....	209

# Abbreviations & Acronyms

AUC	Area under the receiver operating characteristic curve
CORINE	Coordination of Information on the Environment
CRS	Constant returns to scale
DEA	Data Envelopment Analysis
DMU	Decision-making unit
EFSA	European Food Safety Authority
EKE	Expert knowledge elicitation
EL	Greece
EPPO	European and Mediterranean Plant Protection Organization
ES	Spain
EU	European Union
FADN	Farm Accountancy Data Network
GBIF	Global Biodiversity Information Facility
GDP	Gross Domestic Product
HHI	Herfindahl-Hirschmann index
IHD	Irrigated-high-density
II	Irrigated-intensive
IPPC	International Plant Protection Convention
ISPMs	International Standards for Phytosanitary Measures
IT	Italy
IT	Irrigated-traditional
IVD	Inverse distance spatial weight matrix

KNN	K-nearest neighbors spatial weight matrix
NDVI	Normalized difference vegetation index
NPPO	National Plant Protection Organizations
PO <sub>n</sub> TE	Pest Organisms Threatening Europe
RHD	Rainfed-high-density
RI	Rainfed-intensive
RoEU	Rest of Europe
ROW	Rest of the World
RT	Rainfed-traditional
SDM	Species distribution model
SLX	Spatial lag of X model
VRS	Variable returns to scale
Xf	<i>Xylella fastidiosa</i>
Xfp	<i>Xylella fastidiosa</i> subspecies <i>pauca</i>

# Chapter 1

## General Introduction

Given the continuous increase in population across the globe, mankind is faced with the need to optimize food production [1]. Depending on the crop, between 10 to 41 percent of the world's harvests are lost to weeds, insects, and pathogens [2, 3]. Hence, minimization of harvest losses has great potential. Globalization of production and trade continuously leads to the introduction of new pests and poses a challenge to safeguarding production against yield losses [4–6]. Species which are non-native to an area and which have adverse effects on the ecosystem or economies are commonly referred to as *invasive species*. Invasive species introductions are driven by global trade and travel [6, 7]. While the majority of introductions do not result in significant impacts [8], introductions of hazardous organisms can have severe consequences. The agricultural sector is particularly vulnerable to introductions of plant pests [9]. Here, invasive species can lead to a reduction of food supply which can adversely affect consumers through higher prices [10, 11], as well as reduced food quality, food safety [12], and food security [13].

One such invasive species is *Xylella fastidiosa* which is a bacterium from the family *Xanthomonadaceae* first described by Wells et al. [14]. The list of host plants for *Xylella fastidiosa* currently comprises 563 plant species from the Americas, Europe, the Middle East and Asia [15]. In the European Union, at least 84 host plants have been identified [16]. This species is considered one of the most dangerous plant pathogenic bacteria worldwide [15, 17]. The bacterium is naturally transmitted

by insect vectors which feed on the xylem of host plants [18, 19]. If expressed in susceptible plant hosts, symptoms of *Xylella fastidiosa* include, among others, leaf marginal necrosis, leaf abscission, dieback, delayed growth and death of plants through the obstruction of the xylem and a lack of sufficient water flow through the host [20, 21]. The multiplication of the bacteria with the associated clogging of the xylem will first result in declining yields and reduced fruit quality due to a decrease in water and nutrient flow [22]. Eventually, this shortage will result in the host's death [23]. In 2013, *Xylella fastidiosa* subspecies *pauca* was detected in *Olea europaea* (olive), *Nerium oleander* (oleander) and *Prunus dulcis* (almond) in Italy [24]. The detection led to the enactment of control measures, including vector control and tree felling. The latter resulted in great societal unrest in the affected region [25, 26].

In general, the management of plant pests is, among other strategies, achieved through the application of approximately 3.5 million tons of pesticides worldwide [27]. The environmental and health consequences of these agricultural inputs are of societal concern [28]. Various analyses have been conducted to assess whether farmer utilize such inputs efficiently [29–31], however, usually under the implicit assumption that farmers operate in isolation from their peers. This modelling practice contradicts evidence that suggests that peers interact during the decision making [32–35]

Epidemics of native and invasive pests resemble quite closely problems of pollution [36, 37]. In many cases, it is not possible to determine the exact point of origin. As control within one area influences density and dispersal of the pest, management actions directly influence the costs of control for neighbors [37–39]. This results in mutual dependence in control costs and economic impact among stakeholders. By restricting analyses to an individual decision-making unit, the best strategy for pest control might not align with the social optima [37, 40]. Failing to account for the economic and spatial dependence among farmers, regions, countries, and markets could very well result in wrong conclusions.

## 1.1 Plant Health Policies

Internationally, the design of the General Agreement on Tariffs and Trade and the associated Sanitary and Phytosanitary Agreement and Agreement on Technical Barriers to Trade strongly shaped the current practice of managing invasive species [41, 42]. Public policy aimed at reducing risks associated with pests often tries to achieve this by regulating international trade as a primary way of preventing domestic control costs [43]. Countries have the right to “*take sanitary and phytosanitary measures necessary for protection of human, animal or plant life or health*” (Article 2, WTO Agreement on the Application of Sanitary and Phytosanitary measures). In addition, the International Plant Protection Convention (IPPC) provides International Standards for Phytosanitary Measures (ISPMs) that assist countries in reaching agreements as a basis for trade relationships [42].

In Europe, the decision-making process on the management of plant pests is distributed over multiple actors. The European Commission, the European parliament, and the European Council act as the legislative body with the Standing Committee on Plants, Animals, Food and Feed supporting the European Commission through opinions on draft measures. Scientific advice is provided by the European Food Safety Authority (EFSA). The European and Mediterranean Plant Protection Organization (EPPO) assists regulators by developing international standards and risk assessments. National Plant Protection Organizations (NPPOs) plan and implement management in practice. While NPPOs follow European requirements in the context of organisms with *quarantine* status, the management of non-quarantine species is decided upon at the national level. Most plant pests fall in the latter category. This includes species that are either already widely present in Europe, or that are expected to have low *economic* impact. Appending emerging threats to the list of quarantine species is not easy due to the required depth in analysis [41]. The fact that many hazardous species are simply unknown, or not on the radar, prior to them causing impact aggravates the situation [42].

Protecting plant health is important for Europe for two reasons. First, Europe is one of the centers for international trade in general [44], and for plant products in particular [45]. As invasive species are externalities of trade [6, 46], Europe is greatly at risk of suffering

unintended costs from trade through spread of hazardous species. Second, Europe's governmental structure results in a distribution of mandates regarding the regulation of trade and the control of invasive pests. On the one hand, Europe is characterized by a very high integration of countries in terms of free movement of products, services, and people. On the other hand, member states remain partially in control in their enforcement of legislation [47], which can create weak links that undermine countries' control efforts. For example, while the Netherlands designed advanced frameworks and procedures to prevent the introduction and spread of invasive species other member states do relatively little [48].

The control of pest invasions represents a *weakest-link public good* [46, 49]. Public goods generally describe goods for which benefits are neither *rival* nor *exclusive*. Due to the non-excludability, free riding on pest control is incentivized [50]. Weakest-link public good are public goods for which the outcome is largely determined by the least effective peer (i.e., the weakest link). In weakest-link public goods, decision-making units' best interest is to cooperate with the weaker links to incentivize their private control efforts [37].

Over 11,000 species have already spread into Europe with the average annual rate of establishment progressively increasing over the last century [41]. Estimates suggest that these species already cost taxpayers 12.5 billion Euro annually [51]. Considering that for almost 90 percent of invasive species information on impact in Europe is missing, this arguably represents a rather conservative estimate [52]. Insights into the mutual dependence in the context of pest invasions might motivate efforts toward increased cooperation on an international level. Information on the nature of the mutual dependence, on beneficiaries and losers from pest invasions, and the general risk of member states to pest introductions could contribute to a more informed discussion on management strategies.

## 1.2 Economic Dependence

A market is a place in which suppliers provide products and services to consumers in exchange for money [53]. Prices for goods are deter-

mined by the intersection of aggregate supply and demand. Economic well-being is generated by producers through profits following prices exceeding their operational costs. Similarly, consumers derive economic benefits through prices below their willingness to pay for the product [54]. Arguably, the need of having an opposite side on each trade results in mutual dependence among suppliers and consumers. Moreover, as the *aggregate* behavior determines equilibria, which in turn influences economic well-being, dependence arises also among individuals of the same actor group. In the context of a large number of peers, individuals may be influenced by structural changes in behavior of a large number of peers. In markets with a smaller number of peers, strategic considerations emerge precisely because of this mutual dependence [55].

Globalization has resulted in interconnected economies [56]. The European Union is *the* prime example of successful integration and harmonization of markets across countries [57]. As value chains increasingly become international [58], the aforementioned mutual dependence of market actors has gained in complexity by adding a geographic, multi-national, component. Globalization increases the groups of potential suppliers and consumers, thereby, creating ripple effects that not seldom are difficult to foresee [59]. While traditional, physical, marketplaces fell within the jurisdiction of one regulating entity, in an international context regulatory conditions may vary for suppliers and consumers located in different countries [60]. Arguably, many agreements have been reached to harmonize rules [61]. Nevertheless, fundamentally the mutual dependence in economic well-being across geographic spaces with different legislators can result in adverse incentives [62, 63]. For example, changes in trade following changes in aggregate supply, or demand, may benefit one country at the expense of another. While this insight was already embedded in modelling efforts during negotiations of several trade agreements [64], models in a plant health context rarely emphasize this dependence and thereby fail to investigate whether there might be diverging incentives toward management among stakeholders.

### 1.3 Spatial Dependence

As noted by Tobler [65, p.236], *"Everything is related to everything else, but near things are more related than distant things"*. In general, due to the interaction with the surrounding environment, analyses of the agricultural sector can be expected to strongly benefit from the inclusion of spatial effects [66]. Spatial effects comprise spatial heterogeneity and spatial dependence [67]. While spatial heterogeneity concerns differences in the operational environment which consequently lead to different input requirement and output possibility sets, spatial dependence arises from mutual dependence of measurements in space.

To illustrate, the availability and price of a plot determines whether a farmer decides to purchase or lease in a particular location. Consequently, fields are usually scattered around a farm and directly intertwined with plots of other farmers. Focusing on the case of Dutch farming, around 90 percent of farmers purchase their land within a 6.7 km radius [68]. In turn, fields are not only exposed to environmental conditions [69], but also to management practices on neighboring fields.

This mutual dependence is easily comprehensible on a field-level. However, it similarly holds on a landscape or even on country level. The increasing degree of interconnectedness in the age of globalization strengthens this mutual dependence [48]. Pest populations are spatial phenomena by nature [70]. Consequently, Knipling [71] introduced the idea of area-wide pest management via collective actions. The lack of collective actions is a major obstacle to successful pest management [33, 50]. As nearby control of pest populations affects pest pressure in the landscape, inefficiency of a farmer in employing pest control agents is likely influenced by neighboring peers. Similarly, risk mitigation on a regional, or country, level is very much depending on actions of surrounding areas [37]. As plant protection products have adverse effects on the ecosystem, their total use over the landscape, as opposed to farm level, should be optimized [38, 72]. Next to the mutual dependence arising from the spatial nature of pest propagation, farmers might try to improve their decisions by orientating themselves on peers' actions [35, 73]. Such knowledge spillovers can result in meaningful impacts on the operational efficiency of farms [32, 34], and influence risk perception [74].

Spatial dependence arises from spatial structures within the environment [75]. Spatial dependence in abundance of organisms can result from large-scale processes such as erosion, meso-scale processes such as wildfires and fine-scale processes such as organism interactions, for example predator-prey and host-parasitoid dynamics, and distance-limited dispersal [76, 77]. Similarly, patches of comparable nutrient availability or other environmental factors will influence the spatial pattern of pest occurrence. Clearly, contagious biotic processes will result in spatially autocorrelated occurrences through dispersal [76, 78, 79].

Spatially explicit insights may benefit the design of policies by acknowledging local differences, fostering synergies, and allocating funds in the most cost effective manner [80]. Analyses on the control of invasive pests must acknowledge population dynamics and dispersal of the invader [40, 81]. In other words, uniform risk mitigation measures over a heterogeneous space are likely to be less effective and more costly than they need to be [80, 82]. Ultimately, the spatial structure of invaders, hosts, and the environment will have crucial implications for the design of effective control strategies [39, 83], and must therefore be accounted for in economic analyses.

## 1.4 Models and Uncertainties

A scientific model is an attempt to capture a complex system in a comprehensible analogue [84]. By studying the system, gathering data or qualitative information and imposing boundaries, modelers attempt to formulate mathematical relationships which capture the essential elements on a level of granularity suitable to address the research question. Models may take theories as input and through the addition of empirical data allow scholars to generate predictions that theories alone would now allow for [84]. However, the objective of models varies widely from developing hypotheses [85], over testing hypotheses using data [86], to simulating possible futures to inform decisions in the now [87], among others. In many cases, the process of developing the model is at least as, if not more, insightful than the numeric results computed.

Analyses of potential impacts of, and consequently the design of management strategies against, pests are subject to inherent uncertainties [36, 88, 89]. These uncertainties may be due to absence of knowledge on model inputs or the proper model structure [90]. Pest pressure in the landscape is critically affected by complex interactions with biotic and abiotic conditions [76]. Variability in environmental conditions across space will consequently influence the efficacy of management strategies and lead to uncertainty on best practices [83]. As temporal and spatial variation in pest severity is to be expected [91, 92], models for decision support may incorporate such variation [72, 93]. In addition, the production cycle of arable crops results in farmers, and regulators, having to make decision on the appropriate management prior to, by humans, perceivable signals on risks and rewards. Here, simulations of possible scenarios may provide decision makers with probabilistic advice [94].

Evaluations of past performance are critical to optimize production systems going forward. Such analyses are important for the continuous improvements of the environmental and economic costs associated with agricultural practices [95]. While post-hoc assessments allow deriving insights on past performance, they often do not allow to support decision making in the context of novel, invasive, pests that may not have arrived yet [96]. Simulations and scenario-based approaches of potential pest risk are commonly employed to generate versions of possible futures and thereby aid decision makers in the now [97]. Again, such analyses are plagued by underlying uncertainties. Therefore, the sensitivity of results must find central attention [91, 98]. For the most part, there is no such thing as one correct model [84]. Instead, problems should be analyzed through multiple lenses [99]. Here, it is argued that models on plant health threats must range from post-hoc analyses of past managerial performance to simulations of possible futures. Insights of models into either temporal direction are critical to inform decisions in the present.

## 1.5 Problem Statement

Analyses of the economic impact of, and possible risk mitigation strategies against, pests often fail to account for spatial and economic dependencies among the evaluated decision-making units and heterogeneity in the environment they operate in. Individuals' actions in the presence of mutual dependence may not result in the socially optimal outcome. The spread of pests may result in diverging consequences to different actors. Consequently, accounting for the mutual dependence of actors, regions, and countries is critical for analyzing management of pests. In this thesis, I develop methodological approaches to include the spatial nature of pest spread and the mutual dependence of farmers, countries, and markets in bio-economic analyses of plant pests and their management, to contribute to an informed discussion on plant health policies in Europe.

## 1.6 Objective and Research Questions

The objective of this thesis is to develop and implement new methodological approaches to account for spatial and economic dependencies in analyses of economic impact of, and mitigation strategies against, plant pests. To achieve this objective, four questions are addressed.

1. Do neighbors' characteristics associate with farmers' managerial performance?

Question one aims at providing a methodological approach for accounting for possible mutual dependence among farms when evaluating their managerial performance. To achieve this, the dissertation starts by assessing past performance of Dutch arable farms, with a focus on pest control agents, by making use of data from the Farm Accountancy Data Network. The frequently applied two-stage Data Envelopment Analysis framework is extended to incorporate farm characteristics of neighboring observations while optimizing the spatial neighborhood structure. The results highlight the spatial spillovers among farmers and the need to employ multiple neighborhood structures when analyzing spatial dependence.

2. What are the potential economic impacts from *Xylella fastidiosa* subspecies *pauca* to European olive grower?

Question two aims at providing a forward-looking approach for economic impact assessments in interdependent markets. Here, a suite of spatially explicit models is developed with the aim of predicting climatic suitability, simulating potential future spread, and computing economic impact for the invasive pest *Xylella fastidiosa* subspecies *pauca*. A dynamic net-present value model on the olive market is developed to highlight the mutual dependence among growers in different countries. The results show the economic dependence among growers and stress the diverging incentives toward mitigation strategies.

3. Who benefits and who loses from the control of *Xylella fastidiosa* subspecies *pauca*?

To answer question three, the spatially explicit climatic suitability and spread simulations mentioned above are used and the dynamic net-present value model developed under question two is substituted with a partial equilibrium model on the olive oil market. The results highlight the economic dependence among olive oil processors and consumers in different countries. As affected processors and consumers are found to jointly bear the economic consequences from further spread of *Xylella fastidiosa* subspecies *pauca*, the study emphasizes that pest epidemics should be contextualized as a societal challenge, as opposed to one that affects only producers, in public debates on management strategies.

4. Can joint analyses of various pests help identify weak-links and thereby support collective control?

Lastly, to address question four, a prediction of hotspots for pest introduction in Europe is made to identify areas that could be classified as weak links due to their elevated risk scores. A machine learning model was trained on a dataset covering 248 invasive species to map risk of new pest introduction in Europe as a function of climate, soils, water, and anthropogenic factors. Results show that the BeNeLux states, Northern Italy, the Northern Balkans, and the United Kingdom, and areas around container ports such as Antwerp, London, Rijeka, and Saint Petersburg were at higher risk for introductions. The analysis

shows that machine learning can produce hotspot maps for plant pest introduction with a high predictive accuracy, but that systematically collected data on species' presences *and* absences are required to further validate and improve these maps.

The thesis consists of six chapters. Following this first chapter, which provided a general introduction, chapter two to five correspond to research articles that address the research questions one to four in that order. Chapter six provides a general discussion. There, all findings are placed into the context of several related streams of research, opportunities and challenges of georeferenced economic data are discussed, policy implications suggested, and avenues for future work proposed.



## Chapter 2

# Spatial Spill Overs on Input Specific Inefficiency of Dutch Arable Farms

### Abstract

Traditional benchmarking implicitly assumes that decision making units operate in isolation from their peers. For arable production systems in particular, this assumption is unlikely to hold in reality. This paper quantifies spatial s on input-specific inefficiency using Data Envelopment Analysis and a second-stage bootstrap truncated regression model. The bootstrap algorithm is extended to allow for the estimation of the parameter of the spatial weight matrix, which captures the proximity between producers. The empirical application concerns Dutch arable farms for which latitudes and longitudes are available. Farmers were found to be fully efficiency in output. The average inefficiency across years was 3.87% for productive inputs and 2.98% for damage abatement inputs under variable returns to scale. For productive inputs technical inefficiency, statistically significant spillover effects from neighbors' age and their degree of specialization depended on the type of the spatial weight matrix used (inverse distance or k-nearest neighbors), statistically significant spillover effects of subsidy payments were adverse and statistically significant spillover effects from insurance payments were beneficial. For damage abatement inputs technical inefficiency, statistically significant adverse effects were found for neighbors' age and subsidy payments and beneficial effects from neighbors' insurance payments and their degree of specialization.

## 2.1 Introduction

As noted by Tobler [65, p.236], "Everything is related to everything else, but near things are more related than distant things". In general, due to the interaction with the surrounding environment, analyses of the agricultural sector can be expected to strongly benefit from the inclusion of spatial effects [66]. Spatial effects comprise spatial heterogeneity and spatial dependence [67]. While spatial heterogeneity concerns differences in the operational environment which consequently lead to different input requirement and output possibility sets, spatial dependence arises from interdependencies of measurements in space.

The availability and price of a plot determines whether a farmer decides to purchase or lease in a particular location. As a consequence, fields are usually scattered around a farm and directly intertwined with plots of other farmers. Focusing on the case of Dutch farming, around 90 percent of farmers purchase their land within a 6.7 km radius [68]. In turn, fields are not only exposed to environmental conditions [69], but also to management practices on neighboring fields. For example, the control of pathogens on nearby fields can be expected to suppress the populations' ability to disperse into other territories. In turn, spillover effects can be generated by neighbors' management practices. In addition, spillover effects might be generated through the social-network via the transfer of knowledge among farmers [32, 34].

The environmental consequences of agricultural inputs such as fertilizer and plant protection agents are of societal concern [28]. In the agricultural economics literature, pesticides are commonly referred to as damage abatement inputs [100]. Damage abatement inputs reduce potential shortfall rather than further increase output [29]. Parametric [29] as well as non-parametric [30, 31] approaches have been applied to assess whether farmers utilize such inputs efficiently. The need to account for such environmental differences was acknowledged by T. Skevas, Oude Lansink, & Stefanou [101] and T. Skevas & Serra [95]. However, under the implicit assumption that farmers operate in isolation from their peers. Pest populations are spatial phenomena by nature [70]. Knipling [71] introduced the idea of area-wide pest management via collective actions. Similarly, we stress that nearby control of pest populations affects pest pressure in the landscape, which

in turn influences the inefficiency of a farmer in employing damage abatement inputs. Hence, different farm characteristics can be expected to generate externalities for the surrounding farmers. Through social networks, knowledge and experiences might be transferred among farmers [34]. This can foster improvements to input or output efficiency through observation and conversations with peers [32].

The necessity to control for spatial heterogeneity has already been emphasized in the seminal work of Farrell [102]. While the rise in georeference data has greatly benefited scientific efforts to improve the measurement of productivity and efficiency by accounting for unobserved spatial heterogeneity in recent years [103, 104], the issue of spatial interdependence has found little attention with the exception of a handful of studies. The spatial econometric literature is rich in applications on spatial interdependencies [67]. The developed techniques have started to attract the attention of research working on productivity and efficiency. The first contribution in this regard was developed by Druska & Horrace [105] by modelling spatially correlated error terms within the stochastic frontier setting. Various studies measured spatial dependence in efficiency or productivity in non-agricultural applications [106–108]. For the agricultural sector, Areal, Balcombe, & Tiffin [109] identified spatial dependence in technical efficiency of dairy farms in the UK, Martínez-Victoria, Maté-Sánchez-Val, & Oude Lansink [110] found spatial spillovers in productivity growth for Spanish agri-food companies and T. Skevas & Grashuis [111] identified spatial spillover effects on efficiency scores among farm cooperatives in the US. I. Skevas & Oude Lansink [112] and I. Skevas [113] found spatial spillover effects on dynamic inefficiency of Dutch dairy farms. The study makes three distinct contributions beyond I. Skevas & Oude Lansink [112]. Firstly, and most importantly, it advocates for the measurement of spatial spillovers on output and input-specific scores rather than one composite measure of farm performance. While this has clear benefits for policy design and farm management, previous studies on both non-parametric and parametric approaches have solely relied on measuring spillovers on one composite farm-level efficiency score. Thereby, possibly diverse spatial influences have been convoluted into one composite effect that might well provide erroneous insights. Secondly, we contribute by exploiting the advantages of the SLX model by estimating the spatial weight matrix as opposed to the rule of thumb approach

employed in I. Skevas & Oude Lansink [112]. Thirdly, in contrast to I. Skevas & Oude Lansink [112] we employ both a k-nearest-neighbor and an inverse distance approach. Subsequently, we communicate and discuss the different results for the two approaches and thereby stress the need for practitioners to make us of different spatial weight matrix or more clearly motivate their choice for either one. While the importance of accounting for spatial spillovers is stressed in all the aforementioned studies, none of the studies simultaneously measured spillovers on input and output specific inefficiency. Furthermore, the above studies used arbitrary rule of thumbs to define neighboring farmers and construct the so-called spatial weights matrix, in their attempt to estimate spatial spillovers.

The objective of this study is to quantify the effects of spatial spillovers on input and output specific technical inefficiency in Dutch arable crop farms. We address two gaps in the literature. First, we measure spatial spillovers on input and output-specific inefficiency. This allows for more refined insights regarding which outputs or inputs are influenced by neighbors' characteristics in contrast to the aforementioned studies which did not include input and output-specific inefficiency scores into their analyses. Second, rather than making an ad hoc selection of the spatial weight matrix, we estimate the parameter of the spatial weight matrix and report the results for the two most commonly used types (i.e. inverse distance and k-nearest neighbours). Previous studies have found results to be sensitive to the chosen spatial weight matrix, which defines the structure of the spatial relationship between decision-making units (DMUs) [108, 109]. The ad-hoc selection of the spatial weight matrix is frequently criticized in the econometric literature [114–116]. This problem is addressed by estimating the parameter of the spatial weight matrix empirically using farm-level information on coordinates. For this purpose, a two-stage Data Envelopment Analysis (DEA) approach is used. First, a non-parametric directional distance function is computed to estimate inefficiency scores for output, productive inputs and damage abatement inputs. Second, a spatial econometric model is defined, which incorporates regressors for spatial lags of farm characteristics alongside other non-lagged explanatory variables and time-period fixed effects. In contrast to non-spatial efficiency analyses, this framework

extends the farm-level assessment by relaxing the assumption that DMUs operate in isolation from their peers.

## 2.2 Materials and Methods

### 2.2.1 Directional Distance Function

Suppose  $N$  farmers produce  $Q$  outputs from  $I$  productive inputs,  $B$  damage abatement inputs and  $F$  quasi-fixed factors. The damage abatement inputs are exclusively related to plant health within this study. Non-negative vectors of outputs, productive inputs, damage abatement inputs and quasi-fixed factors are denoted by  $y \in \mathfrak{R}_+^Q$ ,  $x \in \mathfrak{R}_+^I$ ,  $a \in \mathfrak{R}_+^B$  and  $k \in \mathfrak{R}_+^F$ , respectively. The production technology for a DMU is fully represented by the input requirement set as  $T(y : k) = \{(x, a) \in \mathfrak{R}_+^I \times \mathfrak{R}_+^B \mid (x, a) \text{ can produce } y, \text{ given } k\}$ . A non-parametric representation of the technology can be depicted as  $T(y : k) = \{(x, a) : Y'\lambda \geq y_i, X'\lambda \leq x_i, A'\lambda \leq a_i, K'\lambda \leq k_i, L'\lambda = 1, \lambda \geq 0\}$

Where  $Y$  denotes a  $N \times Q$  matrix of observed outputs and  $y_i$  is a vector of observed outputs for farm  $i$ .  $X$  is the  $N \times I$  matrix of observed productive inputs and  $x_i$  is the vector of productive inputs used by farm  $i$ .  $A$  is the  $N \times B$  matrix of observed damage abatement inputs and  $a_i$  is the vector of damage abatement inputs used by farm  $i$ .  $K$  is the  $N \times F$  matrix of observed quasi-fixed factors and  $k_i$  is the vector of quasi-fixed factors used by farm  $i$ .  $\lambda$  denotes a  $N \times 1$  vector of intensity variables (farm weights) and  $L$  denotes the  $N \times 1$  unity vector. Constraining the sum of  $\lambda$  to unity enforces variable returns to scale [117].

To estimate the input and output specific inefficiency scores, a directional distance function is computed. Following Chambers, Chung, & Färe [118],  $(g = -g_x, -g_a, g_y)$  denotes the directional vector. The distance function aims to expand output and contract productive as well as damage abatement inputs, simultaneously. The distance of DMUs to the frontier (i.e. the inefficiency score) will generally depend on the chosen directional vector. Choosing the observed quantities ( $g = g_x = x, g_a = a, g_y = y$ ) allows for a direct interpretation in percentages. Furthermore, the measure is more in line with the Farrell

[102] measure of efficiency as noted in Färe & Grosskopf [119]. The distance function can formally be depicted as follows:

$$\vec{D}(x, a, y; g) = \sup \left\{ \beta_x, \beta_a, \beta_y : (x - \beta_x g_x, a - \beta_a g_a, y + \beta_y g_y) \in T(y : k) \right\} \quad (2.1)$$

Within all years  $T$ , the mathematical program aims to identify the maximum attainable expansion of outputs in direction  $g_y$  as well as the maximum feasible contraction of productive inputs and damage abatement inputs in direction  $g_x$  and  $g_a$ , respectively. To achieve this, the following linear programming problem is solved for all  $N$  observations separately for all years. By solving model 2.2 separately for every year, the reference technology is allowed to vary from year to year. This is necessary to account for year specific weather conditions and changes in the technology over time.

$$\vec{D}(x, a, y, k, e; g) = \max_{\beta_x, \beta_a, \beta_y, \lambda^i} \{ \beta_x + \beta_a + \beta_y \} \quad (2.2a)$$

s.t.

$$\sum_{i=1}^N \lambda_i y_i \geq y + \beta_y g_y \quad (2.2b)$$

$$\sum_{i=1}^N \lambda_i x_{ip} \leq x_p - \beta_x g_x \quad (2.2c)$$

$$\sum_{i=1}^N \lambda_i a_{ib} \leq a_b - \beta_a g_a \quad (2.2d)$$

$$\sum_{i=1}^N \lambda_i k_{if} \leq k_f \quad (2.2e)$$

$$\sum_{i=1}^N \lambda_i = 1 \quad (2.2f)$$

$$\lambda_i \geq 0 \quad (2.2g)$$

where  $\beta$  is the percentage value of the expansion (contraction) of outputs (inputs). Constraint (2.2b), (2.2c), (2.2d) and (2.2e) impose free disposability of outputs, productive inputs, damage abatement inputs

and quasi-fixed factors, respectively. Constraint (2.2f) imposes variable returns to scale.

## 2.2.2 Determinants of Inefficiency

The association of farm characteristics with the computed inefficiency scores is measured with the widely used bootstrap truncated regression model [120]. As it is customary, farm characteristics are included in the second stage regression to test for their associations with inefficiency scores. This has been done in the context of, both, radial distance functions [121, 122] and directional distance functions [101, 123, 124]. However, this study also tests for spatial spillovers by further including spatially weighted regressors of the farm characteristics. This specification is commonly referred to as the *spatial lag of X model* (SLX) [116]. The specification is a reduced form approach for measuring spatial interdependency. In contrast to the spatial lag and the spatial error model, the SLX model allows for the estimation of the parameter of the spatial weight matrix and thereby enables practitioners to circumvent rule of thumb approaches. In addition, in contrast to the spatial lag model the signs of direct and indirect effects are not restricted to be similar when employing the SLX model [116]. Lastly, if error terms are spatially structured yet this structure is not accounted for the estimates remain unbiased [125]. To avoid overestimation of the spatial spillovers and to account for the fact that the reference technology is different across years, temporal fixed effects are included as dummy variables. The truncated regression model can formally be depicted as follows:

$$\beta = \alpha I + \eta T + \delta Z + \theta WZ + \epsilon \quad (2.3)$$

where  $\beta$  is a vector of the dependent variable (i.e. pooled inefficiency scores for  $N$  farmers). Following T. Skevas et al.'s [101] study on Dutch arable crop farms, we use the inefficiency scores under variable returns to scale for the second stage. This is motivated by the fact that the variable returns to scale technology represents a less restrictive formulation of the technology. While the constant returns to scale formulation requires global adherence to this property, variable returns to scale allows for increasing, constant or decreasing returns to scale

locally [30, p.64]. While the observed distribution of  $\beta$  is censored at zero, true inefficiency remains unobserved. Therefore, the dependent variable in equation 2.3 must be treated as having a truncated distribution with a point of truncation at zero.  $I$  is the vector of ones associated with the constant term parameter  $\alpha$ .  $T$  depicts the temporal fixed effects with the vector of response parameters  $\eta$ .  $Z$  denotes the matrix of  $J$  explanatory variables and  $\delta$  denotes the vector of unknown parameters to be estimated.  $W$  is the spatial weight matrix which captures the spatial proximity between farmers.  $WZ$  depicts the linear combinations of neighbors' characteristics obtained by inner products of the spatial weight matrix with a variable of interest.  $\theta$  denotes the vector of parameters of the spatially lagged farm characteristics and  $\epsilon$  denotes a vector of independent and identically distributed error terms with zero mean and variance  $\sigma^2$ . Despite panel data at hand, fixed or random effects could not be included into equation 2.3 due to the use of the Simar & Wilson [120] bootstrapping algorithm [31, 101, 112, 121, 122].

The spatial weight matrix ( $W$ ) is constructed based on geographic proximity.  $W$  is always symmetric.  $w_{ij}$  denotes the elements of  $W$ . We employ two common types of spatial weight matrices. In the inverse distance spatial weight matrix ( $IVD$ ), the value of  $w_{ij}$  is the inverse distance between farmers  $i$  and  $j$ . In the  $k$ -nearest neighbor spatial weight matrix ( $KNN$ ), distances between farmers  $i$  and  $j$  are computed. Subsequently, a binary matrix is constructed in which, for every farm, the  $k$  smallest distances receive a value of 1 while all others a value of 0 (see e.g. [110]). While  $IVD$  results in larger weights on characteristics of DMUs in closer proximity, this weight matrix implicitly assumes that spatial influences extent far beyond the nearby vicinity. In contrast,  $KNN$  restricts the spillovers to  $k$  neighbors, but the spatial influences from these  $k$  neighbors are assumed to be of equal importance. Diagonal elements ( $w_{ij}$  where  $i = j$ ) are always set to zero.  $IVD$  is standardized by dividing every element by the maximum eigenvalue of  $W$ , whereas  $KNN$  is standardized by dividing  $W$  by its row-sums [116, 126, 127]. As mentioned above, using  $IVD$  ensures that nearby DMUs exert larger influence compared to distant DMUs. Nonetheless, a distance cut-off ( $\gamma$ ) from which onward no spatial influence is assumed to exist is usually arbitrarily determined by the scholar (e.g. [111]). In contrast to previous work, we estimate the

optimal distance cut-off empirically instead of choosing an arbitrary value. For *KNN*, we estimate the optimal number of neighbors which we also denote with  $\gamma$  for simplicity. The absence of information on the true spatial weight matrix is one of the major hurdles of applied spatial econometrics [114, 115]. A data-driven approach for the selection of  $W$  is therefore one of the major advantages of the SLX model [116]. The ad-hoc selection of either *IVD* or *KNN* is approached by estimating both to assess the robustness of our results. In traditional spatial econometric applications, ordinary least squares residuals are minimized to estimate the optimal spatial structure for the SLX model [128]. In line with this, we maximize the log-likelihood of observing the data within the maximum likelihood estimation of the Simar & Wilson [120, pp.41-42] bootstrap algorithm to search for the optimal distance cut-off or the optimal number of neighbors. This selection goes beyond testing a number of pre-defined spatial structures, as it allows practitioners to optimize the parameter of the spatial weight matrix empirically.

### 2.2.3 Estimation

To estimate the parameter of the spatial weight matrix, a non-standardized (inverse) distance weight matrix is generated first. The optimization algorithm either searches for the optimum distance cut-off between 2.5 and 100 kilometers or for the optimum number of neighbors between 2 and 70. Setting lower and upper bounds ensures feasible values (e.g.  $\gamma > 0$  for *IVD*;  $\gamma < N$  for *KNN*). If the lower or upper limit is found to be binding (i.e. the evaluated quantiles for the bootstrapped distribution of  $\gamma$  fall on one of the limits), the search-range for  $\gamma$  should be widened. The following steps are taken within the optimization of the spatial weight matrix.

1. For *IVD*,  $w_{ij}$  smaller than  $1/\gamma$  are set to zero. In other words, spatial influences from neighbors which are further away than the drawn cut-off value are removed. For *KNN*, for every farm the  $\gamma$  smallest distances are set to 1 and others to 0. In other words, only the spatial influences from the  $\gamma$  neighbors are retained.

2. The spatial weight matrix is standardized by the maximum eigenvalue for *IVD* and row-sums for *KNN*.
3. Spatially lagged variables are generated by computing inner products of the rows of the particular spatial weight matrix and the farm characteristics at hand.
4. Equation 2.3 is computed and the AIC returned.

As true inefficiency scores are unobserved and the estimates serially correlated, we implement the second stage using Algorithm 1 developed by Simar & Wilson [120, pp.41-42]. First, inefficiency estimates are computed using model (2.2a). Second, maximum likelihood in a truncated regression setting is used to obtain estimates of the environmental response parameters as well as the variance of the error term for the inefficient DMUs. At this stage, the aforementioned optimization routine is performed once. Subsequently, the inefficiency scores are replaced by linear predictions using the environmental response parameters for the optimal value of  $\gamma$ . Third, for 2,000 iterations errors are sampled out of a truncated normal distribution, the optimization of  $\gamma$  performed and the environmental variables regressed onto the predicted inefficiencies. Lastly, confidence intervals are constructed for the empirical distributions of the coefficients as well as  $\gamma$  obtained from the bootstrap.

Following Singbo et al. [124], the bootstrapped coefficients are used to compute marginal effects at the mean of the variables in  $Z$  as follows:

$$\frac{\partial E(\beta|Z, \beta > 0)}{\partial Z} = \left\{ 1 - \frac{Z' \hat{\delta}^*}{\hat{\sigma}^*} \times \frac{\phi(Z' \hat{\delta}^* / \hat{\sigma}^*)}{\Phi(Z' \hat{\delta}^* / \hat{\sigma}^*)} - \left[ \frac{\phi(Z' \hat{\delta}^* / \hat{\sigma}^*)}{\Phi(Z' \hat{\delta}^* / \hat{\sigma}^*)} \right]^2 \right\} \hat{\delta}^* \quad (2.4)$$

where  $\beta$  is the estimated inefficiency score,  $Z$  is the mean of a particular environmental variable,  $\hat{\delta}^*$  is the bootstrapped coefficient for the environmental variable,  $\hat{\sigma}^*$  is the estimated variance of the error term,  $\phi(\cdot)$  is the standard normal distribution and  $\Phi(\cdot)$  is the standard normal cumulative distribution function.

### 2.2.4 Data

The balanced panel data<sup>1</sup> on Dutch arable farms are provided by Wageningen Economic Research and cover the period from 2011 to 2016. The dataset comprises of farm-level information on revenues, expenses and balance sheet items as well as geographical information in the form of longitude and latitude coordinates. Furthermore, characteristics of the primary operator are at hand. As coordinates were rounded at one minute by the data provider, we added random noise by sampling out of a uniform distribution of minus one minute to plus one minute to prevent DMUs with the exact same coordinate.<sup>2</sup> Since this study focuses on farms engaged primarily in the arable crop production, we have selected farms whose revenue from sales of arable crops comprises at least 66 percent of their total revenues within every year the farm is observed. The final dataset constitutes a balanced panel of 75 farms with 450 observations. Table 2.1 presents the descriptive statistics. While a larger number of DMUs would have been desirable, the parsimony of model 2.2 justifies the use of annual reference technologies. DEA is frequently used in the context of a small number of DMUs [129]. However, the resulting spatial coverage requires care when extrapolating the results.

In our data on Dutch arable crop farms, the vast majority of total revenue is generated by potatoes, barley, sugar-beet, wheat, onions

---

<sup>1</sup> The data was balanced to ensure that the spatial weight matrix does not change over time. Using unbalanced data would allow for the estimation of the spatial weight matrix as described above if only one overall  $\gamma$  for all year-specific weight matrices is used. Alternatively, one could estimate year-specific  $\gamma_t$ . However, this would significantly increase the complexity of the optimization problem and might result in numerical instability.

<sup>2</sup> The storing of location information in the rounded format resulted in DMUs with the same location. This would have resulted in (infeasible) implausible values when computing the (inverse) distances. Omitting duplicate coordinates is highly undesirable as this would remove DMUs within close proximity which are expected to be critical in generating spillovers. While adding random noise between minus and plus one-degree minute means that the spatial weight matrix inherited a random aspect, in practice the consequences were found to be minimal. We computed 10000 draws and constructed distance matrices to test the spearman correlation of DMU distances between them. We found a correlation of 99.96 percent suggesting that the ordering of importance among DMUs is virtually unaffected by the random noise.

**Table 2.1** Descriptive statistics.

<b>Variable</b>	<b>Dimension</b>	<b>Mean</b>	<b>S.D.</b>
Output	1000 Euro	672.07	608.21
Productive Inputs	1000 Euro	110.99	81.37
Damage Abatement Inputs	1000 Euro	64.01	51.02
Buildings & Machinery	1000 Euro	776.78	817.39
Labor	100 hours	48.53	30.01
Area	100 hectare	1.27	0.92
Age of farmer	10 years	5.23	1.00
Subsidies per ha	100 Euro	3.80	1.73
Insurance per ha	100 Euro	1.14	0.54
HHI	[0,1]	0.34	0.13

and vegetables.<sup>3</sup> Using 2010 as the base year, a Törnqvist index is constructed. The deflated total revenue, excluding subsidies, is used as output ( $Y$ ). Five categories of inputs are used. First, productive inputs ( $X$ ) comprise expenses of seeds and plants, fertilizers, energy, other crop-specific costs and contract work, which were deflated with a Törnqvist index. Second, chemical and biological crop protection agents ( $A$ ) are measured by deflating the aggregated expenditures for both using the price index for *crop protection agents*. Third, buildings and machinery are measured in deflated book values using a Törnqvist index. Fourth, total labor is measured in man-hours and consists of family and hired labor. Fifth, total utilized agricultural area is measured in hectares and includes owned, as well as rented land. Capital, labor and area are included in the matrix of quasi-fixed factors ( $K$ ).

For the second stage, information on the farmers' age, the received subsidies and insurance payments are obtained from the dataset. Subsidies and insurance payments are included as payments per hectare to avoid measuring farm-size effects [130]. The Herfindahl-Hirschman Index (HHI) is computed as proxy for the farm specialization (see e.g. [129, 131, 132]). The HHI is computed by summing the squared revenue shares of ware potatoes, energy crops, barley, grass-seed, oats, other arable crops, other cereals, pulse, seed potatoes, rye, sugar-beet, wheat, fodder crops, onions, starch potatoes, flower bulbs, turnips, vegetables,

<sup>3</sup> Differences in the revenue shares from these crops were found not to be associated with differences in technical inefficiency.

other horticulture, cattle, cut flowers, pigs, poultry and other sources of revenue. The effects of farm specialization, or diversification, have previously been analyzed not only in single-output models [129, 133], but also in the context of one overall farm-level score [124, 134–136]. Despite our one-output approach, economies of scope would become apparent through positive estimates for the coefficient with respect to HHI. This would reflect that lower input-specific technical inefficiencies are associated with lower scores for the HHI (i.e. more diversified farms). Finally, the available latitude and longitude coordinates are used to calculate the distance between farmers. Within the previously described algorithm, the spatially lagged variables for age, subsidies per hectare, insurance payments per hectare and the HHI are computed as inner products of the spatial weights matrix with the farms' characteristics.<sup>4</sup>

Age can be associated with lower inefficiency through the accumulated knowledge from learning-by-doing. On the other hand, it can be associated with higher inefficiency due to decreased motivation or health [137]. The literature is split regarding the potential effects of subsidies on farm-level efficiency [130]. Subsidies may improve the ability to invest in new technology which would have beneficial effects on efficiency. Alternatively, subsidies can deteriorate the eagerness to make economically rational decisions and thereby decrease efficiency [135]. A higher degree of insurance coverage might result in farmers undertaking more risky investments into new technology [138, 139], which could reduce inefficiency. However, larger payments might be associated with higher inefficiency of pesticides as farmers' perception of yield risk might influence their degree of insurance coverage [140], as well as their tendency to overuse damage abatement inputs. The degree of specialization is expected to be associated with lower inefficiency due to more experience in producing the particular product as well as the ability to better optimize production processes [141]. In terms of neighbors' characteristics, the neighbors' age could be associated with lower inefficiency due to knowledge spillovers [32, 142]. Theory regarding the expected effect of the neighbors' subsidies per hectare are absent from the literature. One exception is Storm et al.

---

<sup>4</sup> Ideally, additional farm characteristics such as education and agricultural training would be included in the second stage regression. However, such information is not available in the utilized dataset.

[125] who found adverse effects on farm survival from increased subsidy payments to neighboring farms. Storm et al. [125] hypothesized that neighbors' subsidy payments increase land prices in the vicinity which negatively affects the ability to optimize the scale of production. Risk averse farmers are more likely to have higher insurance coverage [143]. Spatial effects from higher insurance payments of neighboring farms might measure spillovers of risk attitudes. On the one hand, this could result in adverse effects on pesticide inefficiency through social pressure to safeguard against pathogen multiplication. On the other hand, risk averse neighbors might increase vigilance toward pests and thereby improve technical inefficiency of damage abatement inputs through collective efforts as well as through a reduced pest pressure in the landscape as a result of improved phytosanitary conditions on their own fields. Finally, the neighbors' degree of specialization is expected to be associated with reduced inefficiency due to experience spillovers [32].

## 2.3 Results

### 2.3.1 Directional Distance Function

Table 2.2 presents the annual average inefficiency for output, productive inputs and damage abatement inputs under Constant Returns to Scale (*CRS*) and Variable Returns to Scale (*VRS*). The mean inefficiency scores under *CRS* of 0, 4.52 and 4.54 suggest that the potential for producing output is fully exploited whereas farmers can decrease the use of productive inputs and damage abatement inputs by 4.52 and 4.54 percent, respectively. Under *VRS*, the average inefficiency across years are slightly smaller indicating that the farmers operate at an almost optimal size with average scale inefficiencies of 0 percent for output, 0.65 percent for productive inputs and 1.56 percent for damage abatement inputs.

It is important to note that the technical and scale inefficiency scores are not comparable between years, due to the fact that the within-year computations result in a different reference technology. This is accounted for in the second stage bootstrap truncated regres-

**Table 2.2** Average annual inefficiency scores in percent for the output ( $\beta_y$ ), productive inputs ( $\beta_x$ ) and damage abatement inputs ( $\beta_a$ ).

	2011	2012	2013	2014	2015	2016	Mean
<i>Constant returns to scale</i>							
$\beta_y$	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$\beta_x$	5.85	3.72	2.56	6.24	3.60	5.14	4.52
$\beta_a$	4.17	3.53	4.44	4.40	5.10	5.61	4.54
<i>Variable returns to scale</i>							
$\beta_y$	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$\beta_x$	5.91	3.60	3.06	5.47	2.55	2.61	3.87
$\beta_a$	2.51	2.35	2.98	4.00	3.73	2.34	2.98
<i>Scale inefficiency</i>							
$\beta_y$	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$\beta_x$	-0.06	0.13	-0.51	0.78	1.05	2.53	0.65
$\beta_a$	1.66	1.19	1.46	0.40	1.37	3.27	1.56

sion by including temporal fixed effects. The technical inefficiencies of 2014 are comparatively high with means of 5.47 percent for productive inputs and 4 percent for damage abatement inputs under *VRS*. In contrast, year 2016 was comparatively low with means of 2.61 percent for productive inputs and 2.34 percent for damage abatement inputs, respectively. Across years, the average productive inputs inefficiencies fluctuated between 2.55 and 5.91 percent under *VRS*. Damage abatement input inefficiencies varied between 2.34 and 4 percent.

### 2.3.2 Determinants of Inefficiency

Table 2.3 and 2.4 present the results of the bootstrap truncated regression of productive inputs and damage abatement inputs, respectively. The bootstrap truncated regression was not feasible for outputs due to lack of variation in output specific inefficiency.

For the regression of *VRS* technical inefficiency of productive inputs, 80 percent and 66 percent of the parameters are significant (at the 10 percent level or lower) for the *IVD* and *KNN* model, respectively. For the scale inefficiency on the other hand, only 33 and 20 percent of the parameters are significant for the *IVD* and *KNN* models. Table 2.3 also shows that the results of the productive input-specific

**Table 2.3** Bootstrapped regression results for productive inputs ( $\beta_x$ ).

	Technical		Scale	
	IVD	KNN	IVD	KNN
Intercept	0.2033***	0.9103***	-0.0522	0.4230
2012	-0.0502***	-0.0314***	-0.0317	-0.0138
2013	-0.0547***	-0.0349**	0.0142	0.0463
2014	-0.0037	0.0087	-0.0268	0.0002
2015	-0.0651***	-0.0351**	0.0033	0.0477
2016	-0.0614***	-0.0185	0.0541***	0.1186*
age	-0.0103***	-0.0173***	0.0079*	0.0076
subsidies	-0.1558***	-0.1288***	0.0450	0.0543
insurance	-0.0878	-0.0354	0.1548*	0.0937
HHI	-0.0893***	0.0022	-0.1167***	-0.0921**
W_age	0.0247**	-0.1286***	-0.0070	-0.0178
W_subsidies	-0.1267	0.1774**	0.0375	-0.3206
W_insurance	-1.0656***	-0.4869	0.2875	-1.8914
W_HHI	0.1381*	-0.2188***	-0.0860	-0.2389
sigma	0.0380***	0.0393***	0.0367***	0.0373***
AIC	-1397	-1382	-1411	-1384
$\gamma$	42.3	16	54.3	38
$\gamma$ 90% CI	[15.4, 87.1]	[11, 19]	[14.7, 91.2]	[17, 56]

\*\*\*=p<0.01; \*\*=p<0.05; \*p<0.10

technical and scale inefficiency are sensitive to whether an *IVD* or a *KNN* spatial weight matrix was used. The results in Table 2.3 show that the signs of the statistically significant parameters generally (with the exception of *W\_age* and *W\_HHI*) do not change when using either *IVD* or *KNN*. However, some variables are only significant in one of the models (e.g. *HHI*, *W\_subsidies* and *W\_insurance* for the *VRS* inefficiency and *age* and *insurance* in the scale inefficiency).

For damage abatement input inefficiency, a similar pattern arises with 60 percent and 66 percent of the parameters being significant (at the 10 percent level or lower) for the *VRS* technical inefficiency regression of the *IVD* and *KNN* models, respectively. For scale inefficiency, only 33 percent and 13 percent of the parameter were significant for the *IVD* and *KNN* models, respectively. The signs of the statistically significant parameters generally do not change when using the *IVD* or *KNN* model, but the statistical significance of some variables does depend on choosing *IVD* or *KNN* (e.g. *W\_age*, *W\_subsidies* and

**Table 2.4** Bootstrapped regression results for damage abatement inputs ( $\beta_a$ ).

	Technical		Scale	
	IVD	KNN	IVD	KNN
Intercept	0.0894**	0.5338	-0.0410	0.3392
2012	-0.0096	0.0018	-0.0164	-0.0067
2013	0.0272*	0.0427*	-0.0073	0.0105
2014	0.0521***	0.0664***	-0.1025***	-0.0777**
2015	0.0473***	0.0968***	-0.0141	0.0056
2016	0.0215	0.1177***	0.0231**	0.0427
age	-0.0156***	-0.0135**	0.0085**	0.0054
subsidies	-0.2316***	-0.2182***	0.0291	0.0280
insurance	0.1213	-0.0350	0.1113*	0.1123*
HHI	0.0135	-0.0069	-0.0361	-0.0239
W_age	0.0276**	0.0271	-0.0006	-0.0676
W_subsidies	0.1687	0.2552**	-0.0818	-0.1742
W_insurance	-1.0186***	-3.5406***	0.1504	-0.1728
W_HHI	-0.1404	-0.8167***	0.0077	0.1478
sigma	0.0394***	0.0470***	0.0381***	0.0370
AIC	-1530	-1437	-1786	-1798
$\gamma$	57.5	29	54.8	30
$\gamma$ 90% CI	[25.3, 86.7]	[27, 31]	[15.2, 94.2]	[17, 49]

\*\*\*=p<0.01; \*\*=p<0.05; \*=p<0.10

$W\_HHI$  for the *VRS* inefficiency, and farmers' *age* for scale inefficiency).

For productive inputs, the optimal distance cut-off was estimated to be 42.3 km for technical inefficiency and 54.3 km for scale inefficiency. However, the 90 percent confidence interval obtained from the bootstrap suggests rather large intervals ranging 15.4 to 87.1 and 14.7 to 91.2 km for technical and scale inefficiency, respectively. This could be caused by the sizable error terms sampled within the bootstrap. In addition, it is critical to note that only a sub sample of the population is included in the data. Consequently, the estimation of distance decay effects is certainly aggravated. Arguably, the large confidence intervals suggest a minor influence of the distance cut-off on model performance. This seems plausible given the strong weight of close-by DMUs when constructing spatially lagged regressors using the *IVD* weight matrix. The optimal number of neighbors was estimated to be 16 for technical inefficiency and 38 for scale inefficiency. The 90 percent confidence

interval ranged 11 to 19 for technical inefficiency and 17 to 56 for scale inefficiency. For damage abatement inputs, the optimal distance cut-off was estimated to be 57.5 km for technical inefficiency and 54.8 km for scale inefficiency. The 90 percent confidence interval ranged 25.3 to 86.7 km and 15.2 to 94.2 km for technical and scale inefficiency, respectively. The optimal number of neighbors was estimated to be 29 for technical inefficiency and 30 for scale inefficiency. The confidence interval ranged 27 to 31 neighbors for technical inefficiency and 17 to 49 neighbors for scale inefficiency.

In terms of model performance, the *IVD* based regressions obtained a lower AIC score compared to the *KNN* equivalents for all inefficiencies except the scale inefficiency of damage abatement inputs. This suggests that including influences from near-by DMUs, and weighting them more, was able to explain the inefficiency scores better than including influences of the near-by community of  $k$  farmers. However, practitioners should be careful when extrapolating this result to other data. The *IVD* and *KNN* approach could very well perform differently if a more complete spatial coverage would be available. The balanced panel data used for this analysis comprises only 75 DMUs which are distributed across the Netherlands. Consequently, measuring influences from  $k$  nearest neighbors does not necessarily reflect  $k$  tightly connected farms in space. The *IVD* matrix takes distance into account more directly and places majority of the weight on the farmer(s) in close proximity. As near things tend to be more related than distant things [65], it might well be that *IVD* was able to explain the data better than *KNN* given the spatial coverage at hand.

To allow for interpretation of the association, marginal effects are calculated at the mean of the data after equation 2.4. The marginal effects are depicted in Table 2.5. An increase of the farmers age by ten years is associated with a decrease in productive input technical inefficiency by around 0.17 percent and an increase of 0.53 percent in scale inefficiency. For damage abatement inputs, a ten-year increase in farmers' age is associated with a decrease in technical inefficiency of 0.17 to 0.2 percent and an increase in scale inefficiency by 0.58 percent. A rise in subsidy payments of 1000 Euro per hectare is associated with a decrease of productive input technical inefficiency by 2.22 to 2.26 percent. For damage abatement input, the results suggest an association with a decrease in technical inefficiency of 2.35 to 2.85 percent. An

**Table 2.5** Marginal effects on productive inputs and damage abatement inputs inefficiency.

	Technical		Scale	
	IVD	KNN	IVD	KNN
<i>Productive inputs</i>				
age	-0.0016***	-0.0017***	0.0053*	0.0049
subsidies	-0.0226***	-0.0222***	0.0215	0.0272
insurance	-0.0272	-0.0121	0.0745*	0.0403
HHI	-0.0200***	0.0008	-0.0221***	-0.0201**
W_age	0.0223**	-0.0004***	-0.0017	-0.0016
W_subsidies	-0.0290	0.1460**	0.0158	-0.0199
W_insurance	-0.1240***	-0.0765	0.1457	-0.0469
W_HHI	0.0775*	-0.0270***	-0.0228	-0.0241
<i>Damage abatement inputs</i>				
age	-0.0017***	-0.0020**	0.0058**	0.0031
subsidies	-0.0235***	-0.0283***	0.0125	0.0121
insurance	0.0542	-0.0121	0.0492*	0.0500*
HHI	0.0052	-0.0024	-0.0108	-0.0076
W_age	0.0261**	0.0268	-0.0002	-0.0007
W_subsidies	0.1093	0.2314**	-0.0217	-0.0218
W_insurance	-0.1201***	-0.0464***	0.0649	-0.0457
W_HHI	-0.0313	-0.0197***	0.0029	0.1095

\*\*\*=p<0.01; \*\*=p<0.05; \*=p<0.10

increase in insurance payments of 1000 Euro per hectare is associated with an increase in scale inefficiency for productive inputs by 7.45 percent. For damage abatement inputs, higher insurance payments are associated with increased scale inefficiency by 4.92 to 5 percent. For *IVD*, a one unit increase in the degree of specialization is associated with a decrease in productive input technical inefficiency by 2 percent. For both spatial weight matrices, a one unit increase in the degree of specialization was associated with a 2.01 to 2.21 percent decrease in scale inefficiency for productive inputs.

For *IVD*, a cumulative increase in the neighbors' age is associated with an increase in productive input technical inefficiency of 2.23 percent. For *KNN*, a statistical association with a decline in productive input technical inefficiency of 0.04 percent is obtained. For damage abatement inputs, a cumulative increase in the neighbors' age is asso-

ciated with increased technical inefficiency by 2.61 percent if an *IVD* matrix was used.

For *KNN*, a cumulative increase of 1000 Euro in the subsidy payments per hectare is associated with an increase in productive input technical inefficiency by 14.60 percent. For damage abatement inputs, an increase in the neighbors' subsidy payments is associated with an increase in technical inefficiency by 23.14 percent when using *KNN*.

For *IVD*, a cumulative increase in the neighbors insurance payments per hectare by 1000 Euro is associated with a decrease in productive input technical inefficiency by 12.40 percent. For damage abatement inputs, a cumulative increase in neighbors' insurance payments is associated with a decrease in technical inefficiency by 4.64 percent for *KNN* and 12.01 percent for *IVD*.

Lastly, for *IVD* a cumulative increase in the neighbors' degree of specialization by one unit is associated with an increase in the technical inefficiency for productive inputs by 7.75 percent. However, for *KNN* the regression suggests a decrease in productive input technical inefficiency by 2.7 percent. For damage abatement inputs, a cumulative increase in the neighbors' degree of specialization is associated with a decrease in technical inefficiency by 1.97 percent if *KNN* was used.

## 2.4 Discussion

The estimated technical inefficiency scores for productive inputs and damage abatement inputs are in line with earlier findings by T. Skevas et al. [101] and T. Skevas et al. [31] who identified lower technical inefficiency scores (0.03 to 0.10) for productive inputs in the Netherlands during 2003 to 2007. The slight difference between results on the productive input technical inefficiency can be explained by our access to more recent data as well as T. Skevas et al.'s [101] inclusion of undesirable inputs and outputs. For the years 2003 to 2007, T. Skevas et al. [31] estimated annual averages of output technical inefficiency to range between 7 and 13 percent for Dutch arable farmers. Our results suggest that Dutch arable crop farms were efficient in terms of their output during the years 2011 to 2016 with the annual average

inefficiency of 0 percent. A low output-specific technical inefficiency in Dutch farms was also found in other studies [144]. Furthermore, to ease the estimation of the spatial weight matrix we decided to focus on a balanced panel. Consequently, we restricted our sample to commercial arable farms which stayed within the FADN reporting system for a significant number of years. On the one hand, this justifies comparing the DMUs under one reference technology. On the other hand, these criteria might result in a selection of DMUs which is more homogeneous. This could also explain why we find no inefficiency in output and very low inefficiencies in inputs.

The results from the second stage regression suggest that older farmers are associated with lower technical inefficiency of productive inputs and damage abatement inputs. This could stem from their accumulated knowledge and past experiences [137]. However, a cumulative increase in the neighbors' age is associated with higher input technical inefficiency scores. Tveteras & Battese [32] suggest that firms which operate next to knowledge-intensive producers less likely make managerial decisions that increase technical inefficiency. Younger farmers might be more up to date with recent developments and in turn could provide signals to neighboring peers that improve their decision making. Age is also associated with a higher scale inefficiency suggesting that older farmers operate farms at a sub-optimal scale. This finding could be explained from the shorter time horizon for older farmers, resulting in a lower incentive to invest in scale changes [145].

Our results suggest that higher subsidies are associated with lower technical inefficiencies of productive and damage abatement inputs. The literature is divided regarding the effects of subsidies on farm-level efficiency [130, 136]. For the Netherlands, previous studies have identified small impacts of subsidies or a significant positive associations between subsidies and technical inefficiency [146, 147]. The conflicting results in our study might be related to differences between our approach of measuring output and input-specific inefficiency scores and the approach used in previous studies. Our findings could suggest that subsidies allow for investments in improved technologies that might operate more efficiently [136]. Similarly, Reidsma, Ewert, Oude Lansink, & Leemans [148] found direct effects of subsidies per hectare on input intensity per hectare and further argue that increased intensity can lead to a more profitable use of the area. In terms of spatial spillovers,

our results for subsidies were statistically insignificant for *IVD*. However, for *KNN* the results suggest a statistically significant positive association with technical inefficiency of productive inputs and damage abatement inputs. The sensitivity of results to the choice of the spatial weight matrix was also stressed in previous studies [108, 109]. Our results highlight the importance of employing multiple approaches and to report on the robustness of results. The spatial spillover effect of subsidies does not occur for scale inefficiency. Storm et al. [125] found adverse effects of neighbors' subsidy payments on farm survival. Higher subsidy payments can improve farmers' ability to purchase land and thereby increase land prices in the vicinity [125]. Our results for scale inefficiency reject Storm et al.'s [125] hypothesis.

Higher insurance payments per hectare are associated with larger scale inefficiency. This might stem from the need of farmers with sub-optimal scales of production to more rigorously safeguard their income. Alternatively, this could suggest that a base-level of insurance is seen as essential by Dutch arable farmers. Consequently, farms with sub-optimal scales might have to stem larger payments per hectare due to the inability to increase their area of production. Whereas the farm's own insurance payments are negatively associated with scale inefficiency (and not with the farm's technical inefficiency), our results on the spatial spillover of neighbors' insurance payments suggest statistically significant negative relations with the technical inefficiency of productive input and damage abatement input. Farmers with a high perception of yield risk might opt for higher insurance coverage [140]. At the same time, these farmers might be more likely to control diseases in their fields more rigorously to avoid a shortfall in yield. This extra effort could improve the bio-security in the vicinity and could thereby benefit their neighbors.

Consistent with the literature, the degree of specialization is associated with lower technical inefficiency in productive inputs. Specialization comes along with expertise in producing the particular good [134, 135]. Arguably, expertise allows for better judgment regarding input use. Furthermore, larger concentration on a small number of products allows to optimize operational processes [141]. As with the spillover effect of subsidies, our results on productive inputs were sensitive to the choice of the spatial weight matrix. For productive inputs, having more specialized neighbors is associated with an increase in

inefficiency if an *IVD* is used or a decrease in inefficiency if *KNN* is used. Albeit that our results for *IVD* were only marginally significant at the 90 percent confidence interval. For damage abatement inputs, results suggest beneficial effects from more specialized neighbors if an *KNN* matrix was used. This could be related to exchange of knowledge on best practices or improved phytosanitary control on specialized neighboring farms and in turn reduced pressure of pathogens in own fields.

The mixed results for *IVD* and *KNN* are likely related to the different nature of the spatial weight matrices. While *IVD* strongly emphasizes the degree of proximity, *KNN* treats the selected number of neighbors as equally important. In turn, our results shed light on the different effects that could arise from an individual versus a community of neighbors. The mixed results for the spatial spillovers could suggest that distance itself is of greater importance when measuring effects of certain farm characteristics. We argued that many of the spillover effects on damage abatement input technical inefficiency from different farm characteristics are rooted in the interdependence among fields which arises through pathogen multiplication and dispersal. In contrast, spatial spillovers on technical inefficiency of productive inputs are hypothesized to arise through the social network of farmers. Consequently, it might well be that different effects arise depending on whether proximity is taken into account directly, as in *IVD*, or whether a near-by community of  $k$  neighbors is investigated jointly. Certainly, in line with Areal et al. [109] and Pede et al. [108] our analysis stresses the need to communicate the robustness of results in spatial econometric applications in dependence of different formulations of the spatial weight matrix.

While some research has evaluated community-effects on individuals' behavior [149–152], more work is needed on such effects within the context of production economics. Signals for improving operational processes could very well differ depending on whether individual peers or the general neighborhood characteristics are referenced by the decision making unit. Certainly, a clearer distinction between individual-centric versus community-based spillovers is necessary to improve the design of policy. As evident from our results on the spillover effect from subsidies, adverse effects might go unnoticed if analyses do not aim at capturing the different channels of influence.

## 2.5 Conclusions

The objective of this study was to empirically quantify the effects of spatial spillovers on output and input-specific technical inefficiency in Dutch arable crop farms. For this purpose, a two-stage Data Envelopment Analysis (DEA) approach was used. First, a non-parametric directional distance function was computed to estimate technical and scale inefficiency scores for output, productive inputs and damage abatement inputs. Second, a spatial econometric model was estimated which incorporates regressors for spatial lags of farm characteristics alongside other non-lagged explanatory variables and time-period fixed effects. The paper avoids an ad hoc selection of the type of weight matrix, by using both the inverse distance weight matrix and the binary  $k$ -nearest-neighbors weight matrix. In addition, the distance cutoff and the optimal number of neighbors were estimated empirically rather than imposing rules of thumb.

The average technical inefficiency across years was found to be 0 percent for output, 3.87 percent for productive inputs and 2.98 percent for damage abatement inputs. Results of spatial spillovers were sensitive to the choice of the spatial weight matrix. This stresses the need for scholars to apply multiple lenses towards estimating the spatial spillovers in spatial econometric applications. The differences in the results of the two approaches are attributed to the different types of spillovers generated, where the inverse distance approach emphasizes close-by neighbors and the  $k$ -nearest neighbors assigns equal importance to every farmer in the community of  $k$  neighbors. For productive inputs technical inefficiency, statistically significant spillover effects from neighbors' age and their degree of specialization depended on the type of the spatial weight matrix used, statistically significant spillover effects of subsidy payments were adverse and statistically significant spillover effects from insurance payments were beneficial. For damage abatement inputs technical inefficiency, statistically significant adverse effects were found for neighbors' age and subsidy payments and beneficial effects from neighbors' insurance payments and their degree of specialization. For scale inefficiency, no spatial spillover effects were found.

Accounting for spillover effects in estimating the determinants of technical and scale inefficiency relaxes the assumption that farmers

operate in isolation from their peers. Fostering the influx of young farmers is often emphasized by EU policy makers (e.g. [153]). Our results suggest that young farmers could not only lead to more optimal scales of production but benefit the close-by network of peers. The need for farm subsidy payments is often strongly debated in the literature (e.g. [130]). We found sizable adverse spillover effects from subsidy payments on the technical inefficiency of both productive inputs and damage abatement inputs. Hence, the discussion on the need for subsidies should be broadened to also include spillovers to the near-by community of peers. We found that insurance payments are not statistically associated with the technical inefficiency of the insured. However, sizable beneficial spillover effects were found for both productive inputs inefficiency and damage abatement inputs technical inefficiency. The spatial insurance coverage could inform insurance design by signaling the risk awareness of a community of farmers. The beneficial spillover effects might suggest that risk premia could be lowered if a community of farmers is insured. The spatial nature of pathogens certainly results in a mutual dependence between farmers, which is best approached through collective actions [71]. The optimal degree of specialization is subject to discussion in the agricultural economics literature [131, 154]. While results for productive inputs technical inefficiency differed for the two spatial weight matrices, having a community of specialized neighbors seems to benefit own inefficiency for productive inputs and damage abatement inputs. Accounting for these beneficial spillover effects is necessary when designing policy that could affect farm specialization.

Certainly, acknowledging externalities generated by different farm characteristics is crucial for policy design. Traditional farm-level assessments do not allow to account for positive or negative externalities which are generated by different farm characteristics. In turn, insights are obtained that could very well communicate an incomplete picture to the regulators. As stressed in this analysis, farm characteristics will not only influence the own operation but also the community of peers neighboring it. Future research could investigate whether other farm characteristics exert influence on inefficiency. The developed framework could also very well be applied to other industries in which input and output-specific assessments of spatial spillovers are expected.



## Chapter 3

# Impact of *Xylella fastidiosa* subspecies *pauca* in European Olives

### Abstract

*Xylella fastidiosa* is the causal agent of plant diseases that cause massive economic damage. In 2013, a strain of the bacterium was for the first time detected in the European territory (Italy) causing the Olive Quick Decline Syndrome. We simulate future spread of the disease based on climatic suitability modeling and radial expansion of the invaded territory. An economic model is developed to compute impact based on discounted foregone profits and losses in investment. The model projects impact for Italy, Greece and Spain as these countries account for around 95 percent of the European olive oil production. Climatic suitability modeling indicates that, depending on the suitability threshold, 95.5 to 98.9, 99.2 to 99.8 and 84.6 to 99.1 percent of the national areas of production fall into suitable territory in Italy, Greece and Spain, respectively. For Italy, across the considered rates of radial range expansion the potential economic impact over 50 years ranges from 1.9 to 5.2 billion Euro for the economic worst-case scenario in which production ceases after orchards die off. If replanting with resistant varieties is feasible, the impact ranges from 0.6 to 1.6 billion Euro. Depending on whether replanting is feasible, between 0.5 and 1.3 billion Euro can be saved over the course of 50 years if disease spread is reduced from 5.18 km to 1.1 km per year. The analysis stresses the necessity to strengthen the ongoing research on cultivar

resistance traits and application of phytosanitary measures including vector control and inoculum suppression by removing host plants.

## Significance Statement

*Xylella fastidiosa* is one of the most dangerous plant pathogenic bacteria worldwide. Regulatory measures were enacted in response to the detection of the subspecies *pauca* (*Xfp*) in Italian olives in 2013, but the current impact is nevertheless major. We developed a spatially explicit bio-economic model to compute potential future economic impact of the *Xfp* strain. Uncertainty on spread is accounted for by simulating different scenarios. The majority of orchards were found to be within climatically suitable territory. Even under slow disease spread and the ability to replant with resistant cultivars, projections of future economic impact in affected countries run in the billions of Euro. Our findings highlight the importance of minimizing disease spread and implementing adaptation measures in affected areas.

## 3.1 Introduction

*Xylella fastidiosa* (*Xf*) is a bacterium from the family *Xanthomonadaceae* and was first described by Wells et al. [14]. The list of host plants for *Xf* currently comprises 563 plant species from the Americas, Europe, the Middle East and Asia [15]. In the European Union (EU), at least 84 host plants for *Xf* have been identified [16]. This species is considered one of the most dangerous plant pathogenic bacteria worldwide [15, 17]. The bacterium is naturally transmitted by insect vectors which feed on the xylem of host plants [18, 19]. If expressed in susceptible plant hosts, symptoms of *Xf* include, among others, leaf marginal necrosis, leaf abscission, dieback, delayed growth and death of plants through the obstruction of the xylem and a lack of sufficient water flow through the host [20, 21]. The multiplication of the bacteria with the associated clogging of the xylem will first result in declining yields and reduced fruit quality due to a decrease in water

and nutrient flow [22]. Eventually, this shortage will result in the host's death [23].

In 2013, *Xf* subspecies (subsp.) *pauca* (*Xfp*) was detected in *Olea europaea* (olive), *Nerium oleander* (oleander) and *Prunus dulcis* (almond) in Italy [24]. The detection led to the enactment of control measures, including vector control and tree felling. The latter resulted in great societal unrest in the affected region [25, 26]. Unfortunately, the size of the area currently affected and the hidden reservoir of symptomless but infectious host plants is likely to hinder any attempts of disease eradication [155]. Furthermore, recent studies suggest that the tight network of olive orchards in Apulia (Italy) can be expected to serve as a European reservoir of *Xfp* [156]. Nevertheless, the removal of infected trees and vector control along the border of the infected area may act as a cordon sanitaire reducing disease spread.

Currently in the EU, *Xf* is present in Italy, France, Spain and Portugal including the subsp. *pauca*, *multiplex* and *fastidiosa* [157]. Since there is no practical cure for *Xf* under field conditions [158, 159], control strategies applied in the EU focus on eradication or containment of the disease by host removal, vector control and restrictions on the production and movement of plant materials for planting. Research efforts are currently targeting the identification of resistance traits and biological control [160–168]. The use of non-host species or resistant cultivars of host species seems the most feasible and promising long-term strategy to adapt to *Xf* in affected regions [17, 22]. Important advances have been made with regard to the identification of resistant cultivars. In particular, symptom expression in the olive varieties FS-17 and Leccino is drastically reduced compared to other cultivars. The enacted regulatory measures prohibit replanting of hosts within the infected zone. Exceptions were made for FS-17 and Leccino which are currently the only olive cultivars that may be replanted in the infected zone [169].

Here, we develop a spatially explicit bio-economic model that accounts for disease spread and economic characteristics of olive cultivation systems in different European countries. The model projects impact for Italy, Greece and Spain as these countries account for around 95 percent of the European production [170]. Impact is computed over a 50-year time horizon employing a suite of models. The climatically suitable territory is assessed using an ensemble predic-

tion based on ten species distribution models (SDMs). The spatial distribution of olive orchards is obtained from land cover data. The disease spread is simulated using a cellular automaton model with mixed neighborhood processes (rook's and queen's case contiguity) to approximate a radial spread process at spread rates obtained from expert knowledge elicitation. We account for the uncertainty in the annual rate of dispersal by using three quantiles of the expert-elicited distribution of spread rates [157]. An economic model is developed to compute impact to growers as discounted foregone profits and losses in investment due to the premature death of infected trees. Additional profits to non-affected growers, as a result of price responses to changes in the European supply, are accounted for. For all disease spread scenarios, two economic scenarios are explored. In the first, production is assumed to cease once production in an orchard becomes unprofitable due to the disease. In the second, infected orchards are replanted with a resistant cultivar. These two extremes bracket the plausible range of impact. The bio-economic scenarios are compared to a baseline in which *Xfp* is absent. The difference between both economic scenarios approximates the benefit from resistant cultivars.

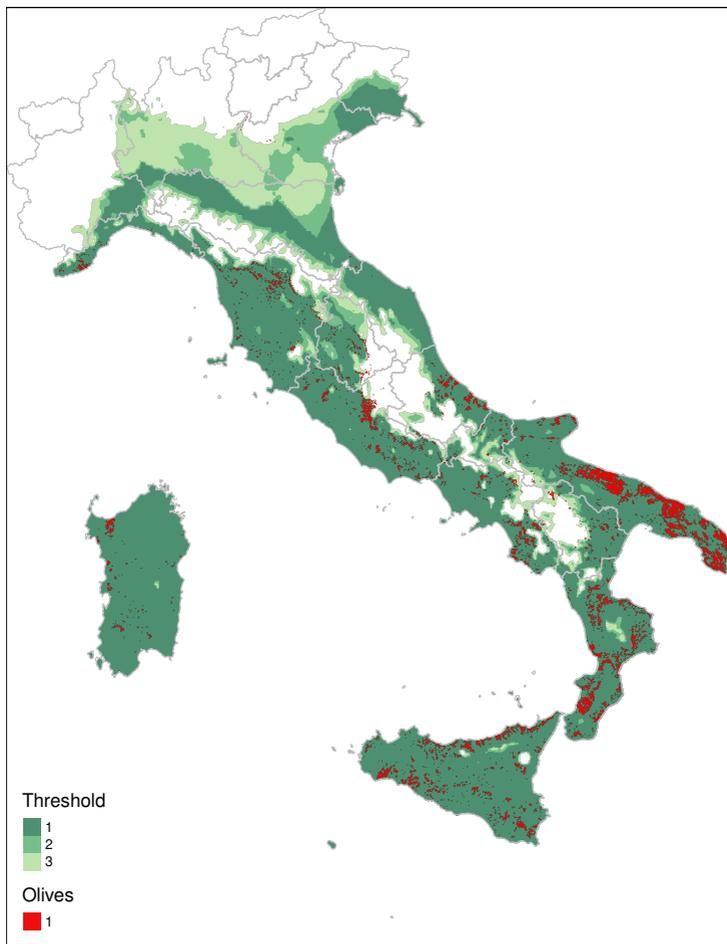
Consequently, this study derives various insights. First, we report on the climatic suitability of the European olive production sites for establishment and spread of *Xf*. Second, we explore the bandwidth of economic impact that results from uncertainty regarding the annual rate of dispersal. Here, we also compare results for different spread rates to provide insights on the economic benefit that might be secured through means of reducing the rate of spread such as vector control and host removal. Third, we analyze economic impact from possible introductions into Greece and Spain. This allows us to identify high-risk areas for disease introduction and establishment, discuss differences between countries in the sensitivity of results with regard to the uncertainty on the annual rate of dispersal as well as compare the magnitude of the potential future economic impact between Italy, Greece and Spain.

The uncertainty on various aspects was taken into consideration. First, while previous work explored the importance of long-distance jumps for spread of *Xfp* [171], data to accurately parameterize such jumps is currently not available. Therefore, our spread model simplifies the dispersal process into a composite spread comprising local dis-

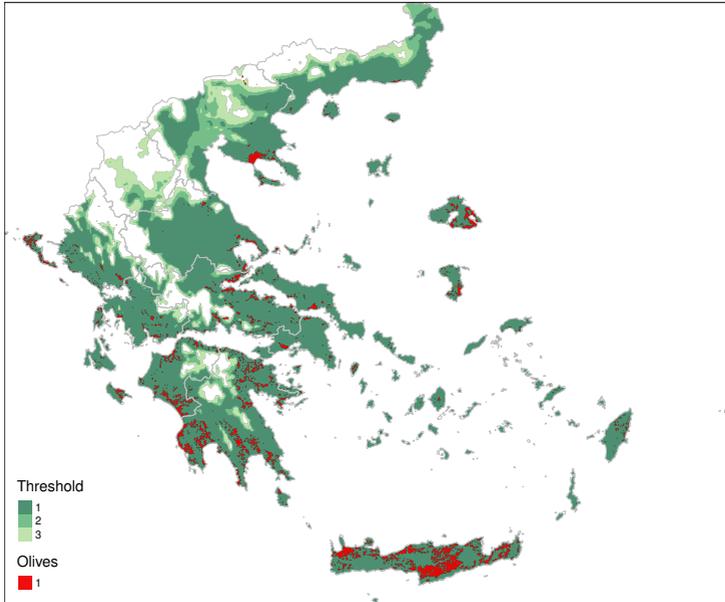
persal and probabilistic jumps. Second, we make use of country-wide averages of prices and operational costs per ton of olives due to the unavailability of data at finer spatial resolution. Third, the economic model intends to derive insights into the potential impacts to olive growers. Processors and consumers are not included into the analysis. We discuss consequences of this simplification and address expected market effects in more detail below. Fourth, changes in fruit quality due to  $Xfp$  are not considered. Reductions in oil yield per ton of olives may reduce the willingness to pay on the side of processors. Consequently, the periods for which continued production on infected plots is profitable would be shortened, and impacts slightly higher as a result. Fifth, for the replanting scenario, we assume that the replanted trees are fully resistant or tolerant to the pathogen and produce the same profits as their susceptible predecessors. While full resistance is achievable, partial resistance or reductions in full-bearing yields might be the outcome of breeding for resistance or tolerance. Lastly, we present an economic best-case scenario in which infected orchards are replanted with resistant equivalents. Many of the olive tree cultivars have been inherited from generation to generation over the last centuries. Arguably, these trees represent a sizable cultural heritage value for many growers and other citizens across Europe [25, 26]. Furthermore, olive orchards provide a landscape value which benefits other sectors such as tourism [172]. Due to the difficulties of quantifying cultural heritage and landscape values in monetary terms, we omit these aspects from our analysis. Nevertheless, this study intends to contribute to a more informed discussion among stakeholders by exploring the direct economic impact that can be expected from  $Xfp$  for European olive growers.

## 3.2 Results

The results on climatic suitability were obtained in the form of a continuous variable which can range from zero to one for a given location in Europe. The continuous scores were converted to a binary prediction (suitable or unsuitable) based on three different thresholds. The thresholds are numerical cut-offs such that locations with higher

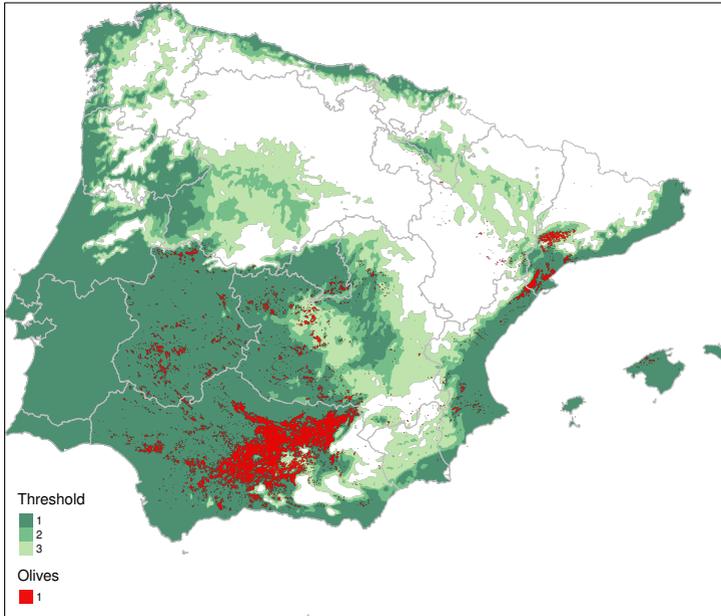


**Fig. 3.1** Binary suitability maps and olive production sites for Italy.



**Fig. 3.2** Binary suitability maps and olive production sites for Greece.

values are classified as suitable. A threshold of 0.165 (T1) was used to ensure that the model correctly predicted at least 90 percent of the locations in which  $Xf$  was confirmed to be present as suitable (90 percent sensitivity). A threshold of 0.132 (T2) maximizes the sum of the accuracy of predicting occupied sites to be suitable and unoccupied sites to be unsuitable (i.e. maximizing the sum of sensitivity and specificity) and a value of 0.093 (T3) minimizes the difference between the accuracy of predicting occupied sites to be suitable and unoccupied sites to be unsuitable (i.e. minimum difference between sensitivity and specificity) [157]. Spatially explicit information on the distribution of European olive orchards was incorporated via the Coordination of Information on the Environment (CORINE) land cover map which was aggregated to a 1 km resolution. The ensemble prediction from the SDMs suggest that for the three thresholds between 95.5 to 98.9, 99.2 to 99.8 and 84.6 to 99.1 percent of the national area of production



**Fig. 3.3** Binary suitability maps and olive production sites for Spain.

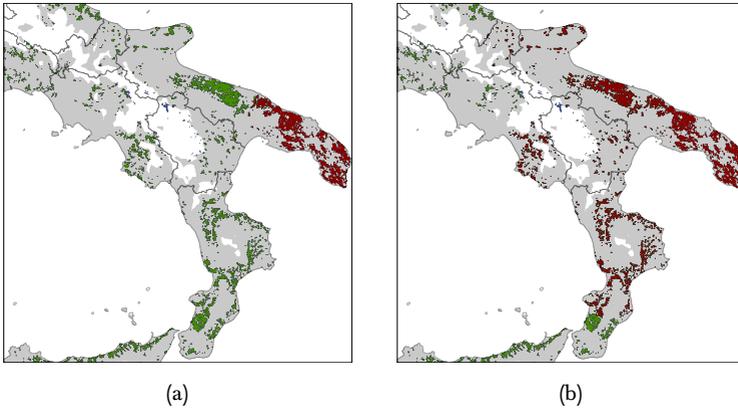
falls into climatically suitable territory in Italy, Greece and Spain, respectively (see Figure 3.1, 3.2, 3.3).

The radial range expansion model requires a point of origin and simulates radial dispersal around this point assuming that all climatically suitable cells within this range may be affected. In our model, spread of the disease is realized through cell-to-cell transmissions. While these transmissions require that the invaded cells are climatically suitable, they do not depend on olives to be present. Consequently, our spread model simulates temporal as well as spatial spread by acknowledging the spatially explicit distribution of olive production sites under the assumption that alternative hosts assist dispersal through climatically suitable habitat. There is still very incomplete knowledge on the rate of spread of *Xfp* [171, 173]. While monitoring data on the outbreak in Apulia is available, the sampling design in the region drastically changed over time since the first detection. This presumably

is related to a change in priorities of authorities from detection to containment of the disease. Consequently, estimating epidemiological parameters is severely aggravated if not impossible. In addition, the observed epidemic in Apulia is likely to be a worst-case scenario due to the tight network of orchards and the abundance of suitable vectors [156, 171]. Therefore, extrapolating estimated parameters for Apulia to continental Europe would be questionable.

To take the uncertainty on the spread rate into account, we utilize the 5, 50 and 95 percentiles of a distribution of spread rates obtained by formal expert knowledge elicitation. The quantiles correspond to a rate of radial range expansion of 1.10 (RR05), 5.18 (RR50) and 12.35 (RR95) km per year [157]. The elicited rates account for the heterogeneous landscape in Europe, differences in vector abundance and application of control measures as is currently done. Long-distance jumps due to plant trade are not considered in the spread simulations within countries. However, they are accounted for when studying introductions into Spain or Greece (see below). For Italy, we analyzed 9 different spread scenarios resulting from the combination of the three spread rates and three thresholds for the binary climatic suitability map. As the elicited rates intend to gauge the pace of spread beyond the current extent, we classified the extent known to be affected in 2019 as infected at the starting point of our modeling time horizon. Subsequently, spread was simulated beyond the infected zone for the three different rates (see Figure 3.4a and 3.4b). For Greece and Spain, the uncertainty on the potential point of introduction of the disease was accounted for by randomly infecting one olive cell within climatically suitable territory. For every suitability threshold, 1000 points of introduction were generated per country. Subsequently, for each point of introduction spread was simulated for RR05, RR50 and RR95. In turn, 9000 spread scenarios were analyzed for Greece and Spain, respectively. A visualization of all generated points of introduction is provided in the Supplementary Material (Fig. 3.5).

For RR05, around 22 percent of the Italian area of production was affected at the end of the 50-year time horizon. The suitability thresholds did not alter these results, because cells within the extent reached fall into suitable territory regardless of the chosen threshold. For RR50, depending on the climatic suitability threshold between 50.3 and 52.9 percent of the Italian area of production was infected



**Fig. 3.4** Simulated geographical distribution of *Xfp* in year 5 (a) and 50 (b) with the median of the elicited rate of radial range expansion (5.18 km per year) and a climate suitability threshold of 0.132 in the ensemble SDM. Grey depicts suitable territory. Blue depicts non-suitable olives. Green depicts suitable olives. Red depicts infected olives.

at the end of the 50-year time horizon. For RR95, depending on the climatic suitability threshold between 68.6 and 75.3 percent of the Italian area of production was infected at the end of the 50-year time horizon. For Greece, on average across the random points of introduction, depending on the climatic suitability threshold, between 7.7 to 8.0, 28.2 to 28.9 and 34.5 to 38.5 percent of the national area of production was infected for RR05, RR50 and RR95, respectively. For Spain, on average across the random points of introduction, depending on the climatic suitability threshold, between 11.4 to 12.3, 62.9 to 69.6 and 74.5 to 94.8 percent of the national area of production was infected for RR05, RR50 and RR95, respectively.

The uncertainty on the point of introductions of the disease in Greece and Spain sensitively influenced the area reached by the pathogen and in turn the economic impact. Irrespective of the annual rate of dispersal, the share of the national area infected was found to differ substantially across the randomized points of introduction (Supplementary Material, Fig. 3.7, 3.9, 3.11). For Greece, introductions into Crete, Attica and western or central Greece resulted in sizable

shares of the national area infected whereas introduction onto islands in the northern or southern Aegean as well as central Macedonia remained isolated from the country's main areas of olive production. For Greece, the consequences of differences in the annual rate of spread were not as pronounced as for the other two countries. The Greek main areas of production are divided between mainland regions and Crete. The sea, as a natural barrier for spread, prevented the epidemic from reaching more than 38.5 percent of the national area of production, on average across the random points of introduction, even for larger annual rates of spread. For Spain, the different climatic thresholds not only influenced the share of the national area of production within climatically suitable territory more strongly compared to the other two countries, but also more sensitively determined the area reached by the epidemic compared to Italy and Greece. The spatial continuity of the climatically suitable area in central Spain influenced whether introductions into Catalonia and the Valencian Community were contained within those regions or whether they spread over Castilla La Mancha into the country's main area of olive production in Andalusia (Figure 3.3). Visualizations of changes in high-risk points of introduction depending on the suitability threshold are provided in the Supplementary Material (Fig. 3.8, 3.10).

The spread scenarios resulted in different extents of the European olive production lost and in turn different price responses following reductions in European supply if replanting is not feasible. For spread in Italy, depending on the climatic suitability threshold around 5.5, 12.6 to 13.2 and 17.7 to 18.9 percent of the European supply was lost in year 50 for RR05, RR50 and RR95, respectively. Consequently, prices were estimated to increase by approximately 2.9, 6.5 to 6.9 and 9.2 to 9.8 percent across Europe. For spread in Greece, depending on the climatic suitability threshold around 1.4, 5.3 to 5.4 and 6.7 to 7.4 percent of the European supply was lost in year 50 for RR05, RR50 and RR95, respectively. As a result, prices were estimated to increase by 0.7, 2.8 and 3.8 percent across Europe. For spread in Spain, depending on the climatic suitability threshold 5.4 to 5.8, 32.4 to 35.4 and 40.8 to 50.6 percent of the European supply was lost in year 50. Prices were estimated to increase by 2.8 to 3.0, 16.8 to 18.4 and 21.2 to 26.3 percent across Europe.

Table 3.1 depicts the economic impact over the course of 50 years in terms of present value. For Italy, the differences in the economic results for the different climatic suitability thresholds were negligible. To improve readability, only the results for T2 are presented. Results for all thresholds can be found in the Supplementary Material (3.4). For Italy, total impact ranged between 1.86 billion for RR05 and 5.17 billion Euro for RR95 if replanting is not feasible. Notably, the increase in producer prices following reductions in Italian supply positively affected Greek and Spanish growers. Depending on the spread rate and in turn the magnitude of the Italian supply reduction, the additional profits ranged between 0.68 billion to 1.59 billion Euro for Greece and 1.71 billion to 3.99 billion Euro for Spain. Summed over the countries additional profits to growers ranged between 0.74 billion to 1.02 billion Euro. Under the replanting scenario, impact in Italy ranged between 0.59 billion and 1.57 billion Euro. The recovery of the Italian supply diminished the price increase which resulted in reduced additional profit flows for Greek and Spanish growers when compared to the scenario without replanting. Nevertheless, the additional profits ranged between 0.10 billion to 0.24 billion Euro for Greece and 0.26 billion to 0.62 billion Euro for Spain. Summed over the three countries losses in profits ranged between 0.02 billion and 0.1 billion Euro. Regardless of the economic scenario, between 0.21 to 0.61 billion Euro worth of investments were lost due to the premature death of trees in Italy. The rainfed-intensive and rainfed-traditional cropping systems were responsible for most of these losses with shares on the losses in investment of 28 and 27 percent, respectively. The benefit of resistant cultivars to Italy ranged between 1.27 billion to 3.60 billion Euro. Evidently, non-affected producers would benefit if Italian growers were not able to recover their supply using resistant trees.

For Greece and Spain, all impacts presented below represent averages across the 1000 random points of the future introduction scenarios. The distributions of all economic results for each spread scenario can be found in the Supplementary Material (Fig. 3.12 & 3.13). For Greece, total impact ranged between 0.21 billion for RR05 to 1.94 billion Euro for RR95 if *Xfp* is introduced and replanting is not feasible. Again, the increase in producer prices following reductions in Greek supply positively affected producer profits in the other two included countries. Depending on the spread rate and in turn the magnitude

Table 3.1 Economic impact over 50 years in billion Euro.

Scenario	Without Replanting			With Replanting			Lost Investment			Benefits of Resistance						
	Total Impact		Change in Profits	Total Impact		Change in Profits	Lost Investment		Benefits of Resistance							
	EL	ES	IT	EL	ES	IT	EL	ES	IT	EL	ES	IT				
<i>Xfp spread in Italy</i>																
RR05	0.68	1.71	-1.86	0.68	1.71	-1.65	0.10	0.26	-0.59	0.10	0.26	-0.38	-0.21	-0.58	-1.45	1.27
RR50	1.04	2.61	-3.15	1.04	2.61	-2.75	0.17	0.42	-1.05	0.17	0.42	-0.65	-0.40	-0.87	-2.18	2.10
RR95	1.59	3.99	-5.17	1.59	3.99	-4.56	0.24	0.62	-1.57	0.24	0.62	-0.96	-0.61	-1.34	-3.37	3.60
<i>Xfp spread in Greece*</i>																
RR05	-0.21	0.18	0.09	-0.17	0.18	0.09	-0.09	0.03	0.01	-0.04	0.03	0.01	-0.04	0.13	-0.15	-0.07
RR50	-1.08	0.89	0.42	-0.89	0.89	0.42	-0.37	0.12	0.06	-0.18	0.12	0.06	-0.19	0.71	-0.77	-0.37
RR95	-1.94	1.60	0.76	-1.65	1.60	0.76	-0.58	0.19	0.09	-0.29	0.19	0.09	-0.29	1.36	-1.41	-0.67
<i>Xfp spread in Spain*</i>																
RR05	0.21	-0.71	0.25	0.21	-0.53	0.25	0.05	-0.39	0.06	0.05	-0.21	0.06	-0.17	-0.16	0.32	-0.19
RR50	1.90	-7.83	2.27	1.90	-6.55	2.27	0.37	-2.93	0.44	0.37	-1.65	0.44	-1.27	-1.53	4.90	-1.83
RR95	3.76	-16.86	4.50	3.76	-14.74	4.50	0.61	-4.98	0.73	0.61	-2.85	0.73	-2.12	-3.15	11.88	-3.77

\* averaged over all random points of introduction. EL=Greece, ES=Spain, IT=Italy.  
All results are for a climatic suitability threshold of 0.132.

of the Greek supply reduction, the additional profits ranged between 0.18 billion to 1.60 billion Euro for Spain and 0.09 billion to 0.76 billion Euro for Italy. Summed over the countries additional profits to growers ranged between 0.10 billion to 0.71 billion Euro. Under the replanting scenario, Greek impacts ranged between 0.09 billion and 0.58 billion Euro. The recovery of the Greek supply diminished the price increase which reduced the additional profits to the other countries compared to the scenario were replanting is not feasible. The additional profits ranged between 0.03 billion to 0.19 billion Euro for Spain and 0.01 billion to 0.09 billion Euro for Italy. Regardless of the economic scenario, between 0.04 billion to 0.29 billion Euro worth of investments were lost due to the premature death of trees in Greece. The irrigated-intensive and irrigated-traditional cropping systems were responsible for most of these losses with shares on the total losses in investment of 35 and 27 percent, respectively. The benefit of resistant cultivars to Greece ranged between 0.13 billion to 1.36 billion Euro. For Spain, total impact ranged between 0.71 billion for RR05 to 16.86 billion Euro for RR95 if *Xfp* is introduced and replanting is not feasible. Again, the increase in producer prices following reductions in Spanish supply positively affected the other two included countries. Depending on the spread rate and in turn the magnitude of the Spanish supply reduction, the additional profits ranged 0.21 billion to 3.76 billion Euro for Greece and 0.25 billion to 4.50 billion Euro for Italy. Summed over the countries foregone profits to growers ranged between 0.07 billion to 6.48 billion Euro. Under the replanting scenario, Spanish impact ranged between 0.39 billion and 4.98 billion Euro. The recovery of the Spanish supply diminished the price increase which reduced the additional profits to the other countries compared to the scenario were replanting is not feasible. The additional profit flows ranged 0.05 to 0.61 for Greece and 0.06 billion to 0.73 billion Euro for Italy. Summed over the countries foregone profits to growers ranged between 0.10 billion to 1.51 billion Euro. Regardless of the economic scenario, between 0.17 billion to 2.12 billion Euro worth of investments were lost due to the premature death of trees in Spain. The irrigated-traditional and rainfed-traditional cropping systems were responsible for most of these losses with shares on the total losses in investment of 32 and 22 percent, respectively. The benefit of resistant cultivars to Spain ranged between 0.32 billion to 11.88 billion Euro.

Evidently, the magnitude of the economic impact differed between the three countries. This can be attributed to the differences in the total area of production, distribution of this area into the different cropping systems and country-specific profitability of olive production per hectare. However, as described above, spatial characteristics of the three countries crucially determined the area of production that was reached by  $Xfp$  within the time horizon for the different spread rates. On average across the random points of introduction, the natural barriers around areas of production in Greece prevented impacts above 1.94 billion Euro. Due to the spatial continuity of the climatically suitable territory in Spain and the spatial concentration of olive production in Andalusia, impact drastically exceeded the impact computed for Italy for RR50 and RR95. The calculated impacts for Spain and Greece are contingent on introduction of the pathogen within these countries at the start of the time horizon.

The economic benefit that might be secured by reducing the annual rate of spread was found to differ among the evaluated countries. For Italy, a reduction in the rate of spread from 5.18 km to 1.1 km per year was found to reduce the overall impact by around 41 and 44 percent in the scenarios without replanting and with replanting, respectively. In other words, around 1.29 or 0.46 billion Euro would be saved. For Greece and Spain, the economic benefit from reducing the annual rate of dispersal is larger with reductions in the overall impact by around 81 and 91 percent, respectively. This corresponds to economic savings of 0.87 billion Euro for Greece and 7.12 billion Euro for Spain if the rate of spread would be reduced from 5.18 km to 1.1 km per year and replanting is not feasible. The difference in the sensitivity of the results with regard to the annual spread rate can be explained by the different types of starting conditions. For Italy, we initiated spread beyond the currently known infected zone which already comprises 17 percent of the national area of production. For Greece and Spain, the simulations were initiated on a randomly generated suitable olive cell. In turn, the area infected within the time horizon is more sensitively influenced by the annual rate of spread which resulted in the differences in the economic benefit from delaying the further dispersal of the disease.

A global sensitivity analysis of the economic model based on spread in Italy using variance decomposition showed that only a few out of 31 parameters had statistically significant first order indices at

**Table 3.2** First and total order sensitivity indices of significant economic parameters without replanting (S1), with replanting (S2) and resistance benefits (RB) for impact in Italy.

Parameter	First Order			Total Order		
	S1	S2	RB	S1	S2	RB
Price Italy	0.664	0.619	0.480	0.717	0.756	0.629
Costs Italy	0.165	0.142	0.117	0.169	0.223	0.149
Discount Rate	0.033		0.074	0.070		0.160
Yield decline			0.039			0.191
Cost change			0.039			0.163

5 percent level. Uncertainty regarding the prices and costs per ton of olives in Italy, the discount rate as well as the changes in yield and operational costs due to *Xfp* were found to sensitively influence the results. Table 3.2 depicts the first and total order sensitivity indices for the statistically significant parameters. Detailed results for all parameters can be found in the Supplementary Material (Tab. 3.5 to 3.10). Price per ton of olives was most influential and caused 72 and 76 percent of the variance in impact with and without replanting, respectively. The profitability per ton of olives will crucially determine the profits foregone and in turn the total impact. This indicates that the effects of the observed empirical variation in prices and costs outweigh the uncertainty in replanting costs as well as other orchard-specific parameters such as the longevity of the cropping systems. Research on the expected annual decline of yield following infection with *Xfp* as well as data on changes in operational costs following infection would benefit further modeling efforts.

### 3.3 Discussion

While impacts were not sensitively influenced by the climatic limits for Italy and Greece, the different climatic suitability thresholds did more strongly influence the maximum extent and the dispersal path of *Xfp* in Spain. Locations and timing of future introductions of the pathogen, if any, are highly uncertain. More research on the climatic suitability for *Xf* in Castilla La Mancha could provide important information on

the spatial continuity of the suitable area. This continuity will crucially determine whether non-detected introductions into coastal areas can be expected to be contained by unsuitable climatic barriers or whether the disease is able to relatively quickly spread into the main olive production sites in Andalusia.

Our analysis revealed that sizable impact can be expected from new introductions of *Xfp* into olive-dense production areas, irrespective of the annual rate of spread. This stresses the need for growers to be vigilant and promptly report possible infections to the national plant protection organizations. Unfortunately, the ability to promptly report introductions and initiate actions to prevent further dispersal crucially depends on the length of the asymptomatic period following infection.

The results show that the economic benefit that might be secured through reducing the rate of dispersal depends on the existence of natural barriers for disease spread and the distribution of olive production sites in a country. Once *Xfp* is well established in an area and has reached a large geographical extent, eradication is considered not practical [156]. Therefore, phytosanitary regulations focus mainly on reducing the rate of disease spread by felling trees and suppressing vectors at the border of the infected area [169, 174]. In particular for Italy and Spain, our results suggest sizable benefit from reducing the annual spread rate. This indicates that current phytosanitary measures to reduce disease spread via inoculum suppression and vector control are of great importance. Further efforts to identify additional effective measures, as done in Serio et al. [159], are called for. However, our results also indicate that introductions into islands of Greece might be managed by early detection, containment and eradication. While reductions in the spread rate still resulted in a sizable benefit in Greece, the natural barriers contained even spread at a larger rate comparatively well. This could render eradication and containment efforts more feasible. Hence, countries with a more continuous climatically suitable territory and a concentration of olive production sites seem to benefit more strongly from means of reducing the annual spread rate. Certainly, more work is needed to provide a sound analysis of territory-based control strategies against *Xfp*. The development of vector-based spread models will greatly benefit future work on this [175].

The planting of resistant cultivars or the substitution of olive production by other land uses seems the most feasible and promising strategy to control *Xf* in those regions where the pathogen is no longer eradicable [17, 22]. Current regulations allow planting of two resistant olive cultivars within the infected zones in the containment areas in Italy [169]. Our analysis revealed a clear benefit, for affected countries, of replanting with resistant varieties. The olive cultivars FS-17 and Leccino present promising points of departure [160, 161, 164, 166]. However, more research is needed on their performance under field-conditions in different cropping systems and different parts of Europe. To prevent landscapes with genetically uniform trees, further breeding efforts are crucial.

Earlier studies on perennials did not account for the possibility of a continued production on partially infected plots [176]. We found that continuing cultivation for a limited period was economically rational. However, early clearing of infected trees might limit the spread of the disease [177]. Therefore, the social benefits derived from the removal of entire infected orchards can be viewed as a public good, suggesting that the costs of eradication warrant compensation from authorities [49]. The analysis revealed that our conservative assumption on the annual decline (increase) of yield (operational cost) by 10 percent resulted in negative profit margins within two to four years after first infection, depending on the country and cropping system. In turn, commercial farmers can be expected to cease production relatively quickly following symptom expression of *Xfp*. This might increase the European loss in production beyond what is expected when solely focusing on the biological yield decline [157]. If olives harvested from infected trees result in lower oil yield per ton of olives, willingness to pay on the side of the processor might be reduced which would slightly increase impacts by shortening the periods for which continued production under infection is profitable.

Our analysis focused on the impacts to growers to narrow down on the intersection of biology and economics. By taking the cropping system specific conditions into account, we were able to simulate the replanting scenario and derive impacts which arise solely through the premature death of trees. In general, invasive species tend to result in reductions in yield which, if the extent of the epidemic is sufficiently large, result in changes in the country's total production. In cases in

which the affected country is a significant contributor to the European (or world) supply, the reduction in the production is likely to result in an increase in price [178]. As highlighted within this analysis, higher producer prices will benefit non-affected growers [10, 179, 180]. Most olives are used as an input for processing into olive oil. In turn, the simulated price increase would result in higher costs of production for oil processors. This could affect the consumer price for olive oil. However, the degree to which the change in production costs could be transferred to consumers and the degree to which higher consumer prices for olive oil would be transmitted back to olive growers depends on magnitude, speed and asymmetry of price transmissions in the supply chain [181]. Among other aspects, these factors are influenced by the existence of market power of processors [182], as well as the consumers' willingness to switch to alternative products. Future work could build on the framework developed within this study and narrow down on modeling these supply chain related aspects. This could add insights into the potential impacts to processors, consumers and competing markets.

Some of our assumptions may be too optimistic. First, the profit flows understate the value generated through olive processing. If the level of analysis is extended to olive oil, the economic impact would be greater due to the larger profit margins in the oil production [183]. In 2017, the production value of olives was around 2.4 billion Euro whereas the production value of olive oil was around 6.7 billion Euro [184]. Second, farms were assumed to be able to replant. However, the olive sector in Europe is characterized by relatively small-scale farming and some farms may not have the financial means for replanting [183, 185]. In our study, resistant cultivars were defined as those not suffering reductions in yield or quality when planted in an infected area. This applies both to completely resistant cultivars, where the host and the pathogen are incompatible, or completely tolerant cultivars, where the host is infected but without yield loss [186]. While full resistance is achievable, partial resistance or reductions in full-bearing yields might be the outcome of breeding efforts. In addition, tolerant cultivars remain hosts for the pathogen and are inoculum reservoirs [187], which might support disease spread. Lastly, vigilant growers might aim to stay ahead of the disease by additional monitoring efforts and preventive measures prior to their orchards being infected.

Consequently, associated increases in operational costs due to these preventive measures outside of the radial range of infection would increase total impact beyond what was computed here.

Replanting sometimes centuries old or even millennial trees by young trees has severe consequences in terms of cultural heritage and provision of a landscape that is attractive for tourism and recreation. Quantifying these losses in monetary terms was not within the scope of this analysis. Furthermore, the slow development of olive orchards can be expected to result in considerable nurturing costs. Additional income support schemes might be necessary to ensure that farmers remain financially capable to nurture the orchards back into a productive state that contributes to cultural heritage and an attractive landscape in the affected areas.

*Xfp* is known to affect various economically important hosts, including besides olives, also cherries and almonds. Additional assessments on other host species would inform the discussion on risks associated with new introductions of *Xf* or the further dispersal of the strain detected in Italy. Certainly, the importance of the European wine sector calls for an assessment of subsp. *fastidiosa* in grapevine. The modelling framework developed within this study could very well be used for this. The overall potential impact of *Xf* in Europe may thus far exceed the impact evaluated here for the subspecies *pauca* in one host, olive.

## 3.4 Materials and Methods

### 3.4.1 Climatic Suitability Map

Species distribution models (SDMs) explore the relationship between geographical occurrences of species and environmental variables [188, 189]. SDMs draw statistical inference on drivers of species ranges from a snapshot of occurrence data by finding statistical correlations between species' distributions and environmental factors. We make use of occurrence data of *Xf* from the Update of the *Xf* host database [15], local datasets of outbreaks in Italy, France and Spain which were obtained from the national plant protection organizations (i.e. Osservatorio Fitosanitario Regione Puglia, Italy; Servicio de Sanidad

Vegetal, Generalitat Valenciana, Spain; Conselleria de Medi Ambient, Agricultura i Pesca del Govern de les Illes Balears, Spain; Bureau de la Santé des Végétaux, Ministère de l'Agriculture et de l'Alimentation, France) as well as recent records of  $Xf$  in Porto, Tuscany and Hula Valley [190–192].

The presence records were filtered in three ways: first, by selecting only records from infection observed under natural inoculum pressure either during surveys or research activities on natural habitat. Thereby, omitting records from greenhouse, screenhouse or interceptions; second, by selecting records with precise geographic coordinates; third, by only using records with confirmed positives. To reduce spatial autocorrelation, the presence records were further submitted to a spatial filtering approach. In this procedure, the presence records are randomly selected according to a minimum nearest neighbor distance of at least 5 km between each locality. This distance is equal to the spatial resolution used for the climatic data. The procedure was repeated four times obtaining four different spatially filtered data sets. We generated weighted pseudo-absence data to simulate a prevalence of 0.1. To explore and reduce the uncertainty of the random sampling, we repeated this process four times to generate four pseudo-absence datasets per model replication.

Climate data was obtained from Chelsa Climatology [193]. The data ranged from 1979 to 2013 and is a downscaled version of the European Centre for Medium-Range Weather Forecasts Reanalysis Interim (ERA-Interim) global circulation model. We use data at a 5 km resolution. Nineteen bioclimatic variables were analyzed out of which nine were included into the prediction after controlling for multicollinearity (variance inflation factor <10). The ensemble prediction followed the methodology described within Bragard et al. [157]. However, for this study we refined the spatial prediction from a resolution of 10 km to 5 km. We made use of ten modeling techniques, namely, bioclim, boosted and regression trees, classification and regression trees, domain, generalized additive models, multivariate adaptive regression splines, maximum entropy, random forest, recursive partitioning and regression trees and support vector machines. Model performance was evaluated using the true skill statistic [194]. In total, we computed 800 models comprising four spatially filtered data sets, four pseudo-absence sampling replicates, ten modeling techniques and

five cross-validation runs. The final prediction combined the individual predictions with a true skill statistic larger than or equal to 0.7 [195].

The output of this prediction is a continuous variable that can range from zero to one for a given location in Europe. To better integrate the suitability map with the needs of the disease spread simulation, the continuous scores were downscaled from the 5 km resolution to a 1 km resolution using bilinear interpolation. This ensures that unsuitable barriers such as waterbodies are accurately accounted for when simulating spread. Furthermore, this improves the coverage of the predicted area in coastal areas with irregular shapes such as in Greece which is crucial as many of the olive cells are located near the coast. Lastly, the downscaled map was converted into a binary prediction (suitable or unsuitable) for each 1 by 1 km cell using three different thresholds. Threshold 1 (0.165) is particularly informative for models based on presence-only data and ensures that a correct prediction on species presence of at least 90 percent is made. Threshold 2 (0.132) was used to maximize the sum of the accuracy of predicting occupied sites to be suitable and unoccupied sites to be unsuitable (i.e. sum of sensitivity and specificity) and 0.093 (T3) was used to minimize the difference between the accuracy of predicting occupied sites to be suitable and unoccupied sites to be unsuitable (i.e. minimum difference between sensitivity and specificity) [196].

### 3.4.2 Disease Spread Simulation

Data on the olive production sites in Europe was obtained from the CORINE land cover map<sup>1</sup> and aggregated to a 1 km resolution to reduce the computational time. To simulate spread, we use a basic radial range expansion model proposed for risk analyses in Robinet et al. [197]. The model is the mathematical solution of a two-dimensional population growth model (exponential or logistic) with random dispersal, also known as the Skellam model [198–200]. Despite the model’s simplicity, past population expansions have been found to compare reasonably well to such a radial range expansion approach [201]. The model has a single parameter called the rate of radial range expansion

---

<sup>1</sup> <https://land.copernicus.eu/pan-european/corine-land-cover>

( $rr$ ) which depends on population growth and dispersal characteristics which are collapsed into a single parameter. For the value of  $rr$ , we used the 5, 50 and 95 percent quantiles of a distribution elicited from experts using formal methods for expert knowledge elicitation (EKE) [157]. The quantiles correspond to radial rates of range expansion of 1.10, 5.18 and 12.35 km per year.

The structured EKE followed the methodology described in the European Food Safety Authority Guidance on Uncertainty [202]. Our concise overview of the approach follows the description in Baker et al. [203]. The ad hoc group included experts who defined the methodology as well as internationally recognized experts on the disease and on relevant agricultural practices. First, the parameter was reviewed and clarifications provided to the experts if needed. Second, evidence was provided and discussed to derive a list of evidence and uncertainties. The corresponding evidence table is published in Bragard et al. [157]. Third, overall uncertainties were summarized. Fourth, the parameter was elicited by a structured expert judgement following the Quartile Method of the Sheffield protocol [204]. Here, each of the seven invited experts was asked to individually estimate the following quantiles in this order 1) 1st and 99th percentile, 2) median value, 3) inter quantile ranges. Afterwards, estimates were discussed and a consensus distribution agreed upon by the group. Lastly, the fitted distribution was reviewed and agreed upon. In this EKE, the rate of radial range expansion was defined as the *mean distance (km) which will comprise 90 percent of the area containing the newly infected plants around an infected area within 1 year* [157]. Assumptions focused on disease spread by infected vectors, due to their natural dissemination or human-assisted movements but not plant movements for trade. The estimates consider the heterogeneity of the European territory, the differences in vector abundance as well as current control measures.

Radial range expansion is modelled using a cellular automaton model with mixed neighborhood sequences (rook and queen) to generate cell to cell spread on a grid of 1 by 1 km cells. For this purpose, the discrete annual time steps are further broken down into within-year time steps. The number of within-year time steps depends on  $rr$ . For example, a rate of 5.18 km per year is approximated by 41 years of 5 and 9 years of 6 within-year time steps. This generates 259 steps in total and 5.18 steps per year on average. The ordering of the number

of within-year steps across years is randomized. By ensuring that a proportion of  $2 - \sqrt{2}$  of steps are taken in rook's fashion and  $\sqrt{2} - 1$  in queen's fashion, a spread pattern is generated which conforms to a regular octagon which closely encloses a circle [205]. The step-type (rook or queen) is randomly assigned to every step while ensuring that, over the course of the time horizon, the aforementioned proportions of rook and queen steps are obtained exactly. As corners of the octagon marginally over-estimate spread, the area infected at every time step is constrained by a radial range model with the radius expanding at the elicited rate. Invasion into a cell is only accepted if it is climatically suitable which ensures that unsuitable territories are not travelled through.

### 3.4.3 Economic Model

To account for economic differences between the countries and cropping systems, it was decided to stratify the population into  $I$  cropping systems indexed with  $i$  in  $C$  countries indexed with  $c$ . The total olive growing area in Europe ( $A_0$ ) was assumed to be constant at 4.6 million hectares for the entire planning horizon of 50 years [206]. A planning horizon of 50 years was chosen due to the slow development of olive orchards as well as their natural longevity. The percentage distribution of the area of production into density (<140, 140-399, >400 trees per hectare) and age (<5, 5-11, 12-49, >50 years) classes for all countries was obtained from Eurostat [206]. Additionally, Eurostat information on the national percentage of irrigated olive hectares was obtained. The density classes were further sub-divided into rainfed and irrigated on the assumption that the irrigated share is the same across density classes. Within age-classes, hectares were uniformly distributed. A dynamic model was built and run over a planning horizon of  $T$  years at an annual time step indexed by  $t$ . Orchard age is indexed by  $a$ . The area of production for country  $c$ , cropping-system  $i$ , of age  $a$  is denoted  $A_{cia}$ . Once the maximum longevity was reached, it was assumed that replanting is undertaken in the following year.

The prices are denoted ( $p_{ct}$ ). Future monetary flows were discounted by the discount rate  $r$ . The replanting costs ( $RC_i$ ) were ac-

counted for by equivalent annual costs ( $ARC_i$ ). Equation 3.1 converts the establishment costs of an orchard to annual costs in dependence of the different longevities ( $m_i$ ) of the evaluated cropping systems.

$$ARC_i = \frac{RC_i \cdot r}{1 - (1 + r)^{-m_i}} \quad (3.1)$$

Yield in tons per hectare is specific for the country, cropping system and tree age ( $Y_{cia}$ ). We use cropping-system specific information on full-bearing ages and full-bearing yield potential to linearly interpolate yield for all tree ages. Subsequently, we re-scaled yields such that the simulated total production equals the 5-year averages prior to the detection of the pathogen (2007 to 2013) after FAOstat data. This equates to around 3.19 million, 2.37 million, 6.69 million tons in Italy, Greece and Spain, respectively. The yield was multiplied with the price ( $p_{ct}$ ) to get the yearly revenue per hectare. The country, cropping system and orchard age specific operating costs ( $C_{cia}$ ) as well as the equivalent annual costs for replanting ( $ARC_i$ ) were subtracted from the revenue to obtain the profit ( $\pi_{ciat}$ ) in Euro per hectare. Inner products of  $A_{cia}$  and  $\pi_{ciat}$  were computed to obtain the annual profit flow for the particular cropping system in country  $c$  and year  $t$  ( $\Pi_{cit}$ ). In other words, we multiplied the country, cropping-system and age specific area with the corresponding profit and aggregated across all orchard ages. The total annual profit flows from olive production within Europe ( $\Pi_t$ ) in year  $t$ , was computed by summing the annual profit flows for all cropping systems across countries and across cropping systems. The net present values for the annual profit flows were obtained by discounting with the rate  $r$  by multiplying ( $\Pi_t$ ) with  $(1 + r)^{-t}$ .

The total production area affected by  $Xfp$  in country  $c$  at time  $t$  is denoted as  $A_{ct}^x$  and was obtained by the spread simulations in percent of the total national olive cells. The susceptible but disease-free area ( $Sx_{cit}$ ) for cropping system  $i$ , country  $c$ , orchard age  $a$  at time  $t$  was calculated by subtracting the total area affected in year  $t$  from the total area ( $A_{cia}$ ). The inner products of  $Sx_{cit}$  and the  $\pi_{ciat}$  were computed to obtain the annual profit flow of disease-free hectares for the particular cropping system ( $\Pi_{cit}^h$ ). In other words, the area across orchards ages within a given country and cropping system was multiplied with the corresponding profits and the resulting profits aggregated over all orchard ages.

To store the disease progression within infected trees, tensors ( $M_{cit_x}^x$ ) were generated in which rows depict the different orchard-ages, columns depict different points in time and each element depicts the vector of different ages-of-infection. We denote by  $\Delta A_{ct}^x$  the area of olives that was newly found to be infected in country  $c$  in year  $t$ , calculated by the difference of the cumulative percentage infected in succeeding discrete time steps. The newly infected area for cropping system  $i$ , country  $c$ , orchard age  $a$  in year  $t$ , was obtained by multiplying the percent of newly infected cells with the total number of hectares for the particular combination. The yield of infected orchards ( $Y_{cia,a_x}^x$ ) declines with every discrete time step under infection ( $a_x$ ). The country, cropping system and age-of-infection specific operating costs ( $C_{ci,a_x}^x$ ) as well as the equivalent annual costs for replanting ( $ARC_i$ ) were subtracted from the revenue to obtain the profit in Euro per hectare ( $\pi_{cia_x}^x$ ). Orchards remain in production as long as their profit margins are non-negative. Thenceforth, it was assumed that the production ceases and no profit is generated from the infected orchards. Inner products of the  $\Delta A_{ct}^x$  and the  $\pi_{cia_x}^x$  were computed to obtain the annual profit flow of infected hectares for the particular cropping system ( $\Pi_{cit}^x$ ). In other words, the infected area across orchards ages within a given country and cropping system was multiplied with the corresponding profits under infection and the resulting profits aggregated over all orchard ages.

As most of the orchards die off prematurely compared to the natural production cycles, additional costs arise due to the loss of investment. For an orchard of age  $a$  at the point of death, the farmer will have utilized  $a$  periods of the equivalent annual replanting costs. If orchards die before their maximum longevity ( $m_i$ ) is reached, the system-specific equivalent annual replanting costs for the remainder of periods are accounted for as losses of the investment ( $L_{ia}$ ) (equation 3.2). To compute all possible losses of investment, vectors for all combinations of cropping system ( $i$ ) and orchard age ( $a$ ) were generated ( $LI_i$ ). The inner products of the died-off population ( $v_{cit}^d$ ) and  $LI_i$  were computed to obtain the total amount of lost investments in year  $t$  ( $L_{cit}$ ). In other words, the area of unprofitable orchards across orchards ages within a given country and cropping system was multiplied with

the corresponding losses in investment and the resulting losses were aggregated over all orchard ages.

$$L_{ia} = [ARC_i \cdot (m_i - a)] \quad (3.2)$$

To model a price response following the decrease in European supply, we computed the reduction in supply in percent ( $\Delta Q$ ) at the end of every time step. Subsequently, prices were updated by adding ( $\Delta Q \times pr \times p_{ct}$ ) and profits ( $\pi_{ciat}$  and  $\pi_{cidx}^x$ ) recomputed for use in the following year. To estimate the price response parameter ( $pr$ ), we collected panel data from FAOstat on produced quantities of olives, the area of production and price indices for Italy, Spain and Greece from 1991 to 2017. As prices may influence the produced quantities and vice versa, parameter estimates obtained from ordinary least squares estimates would suffer from endogeneity bias. We addressed this problem by instrumenting the produced quantities with the area of production. We estimated equation 3.3. Where  $\log(P)$  and  $\log(Q)$  are log-transformed prices and produced quantities, respectively. *Year* represents a time-trend, *Country* dummy variables for country-effects,  $\log(P_{t-1})$  a lag to control for autocorrelation.  $\tau$ ,  $\theta$ ,  $\rho$  and  $\beta$  are parameters to be estimated and  $\epsilon$  the independent and identically distributed error term. The coefficient ( $\beta$ ) of  $\log(Q)$  ( $-0.52$ ,  $P < 0.001$ ) was used as an estimate for the price response parameter in our model.

$$\log(P) = \tau Year + \theta Country + \rho \log(P_{t-1}) + \beta \log(Q) + \epsilon \quad (3.3)$$

To obtain the total annual profit flows from olive production within Europe in year  $t$  under the *Xfp* epidemic, all profits from disease-free ( $\Pi_{cit}^h$ ) and infected ( $\Pi_{cit}^x$ ) hectares are to be considered as well as the losses of investment ( $L_{cit}$ ). For the total annual profit flow ( $\Pi_t^x$ ), it was aggregated across countries, cropping systems and discounted with rate  $r$  (equation 3.4).

$$\Pi_t^x = \sum_{c=1}^C \sum_{i=1}^I [\Pi_{cit}^h + \Pi_{cit}^x - L_{cit}] \cdot (1+r)^{-t} \quad (3.4)$$

We denote by  $\Delta A_{cit}^r$  the area of olives that is newly replanted in year  $t$ . The areas of replanted orchards are denoted  $A_{cit,ar}^r$ . The yield and profit were assumed to be similar to the susceptible equivalents. For the best-case economic scenario, the total annual profit flows in year  $t$  under the *Xfp* epidemic is obtained by considering profits from disease-free ( $\Pi_{cit}^h$ ), infected ( $\Pi_{cit}^x$ ) and resistant ( $\Pi_{cit}^r$ ) hectares as well as the losses of investment ( $L_{cit}$ ). For the total annual profit flow ( $\Pi_t^r$ ), it was aggregated across countries, cropping systems and discounted with rate  $r$  (equation 3.5).

$$\Pi_t^r = \sum_{c=1}^C \sum_{i=1}^I [\Pi_{cit}^h + \Pi_{cit}^x + \Pi_{cit}^r - L_{cit}] \cdot (1+r)^{-t} \quad (3.5)$$

Economic impact for both scenarios ( $EI^x$  and  $EI^r$ ) was computed by aggregating the differences between the profit flows without *Xfp* and profit flows  $\Pi_t^x$  and  $\Pi_t^r$  over  $T$ . The difference between  $\Pi_t^x$  and  $\Pi_t^r$  is expected to provide an exploration of the potential economic benefit associated with ongoing research on resistance traits ( $RB$ ). While  $EI^x$ ,  $EI^r$  and  $RB$  will depend on the choice of  $T$ , the discounting effect in the later years of the time horizon will result in only small differences if the number of years is slightly reduced or increased.

### 3.4.4 Global Sensitivity Analysis

To assess the parameter sensitivity, we conducted a global sensitivity analysis using a variance decomposition method [207]. Sensitivity indices report the variance in the output  $Y$  attributable to variation in input  $X_i$  (*first-order indices*) as well as through higher-order interactions between this variable and other inputs (e.g. *second-order indices*). The total effect on the output caused by the input  $X_i$  is called the *total-order sensitivity index* [208]. The conditional variance of input  $X_i$  and model output  $Y$  can be written as  $V_{X_i}(E_{X_{\sim i}}(Y|X_i))$  with  $X_{\sim i}$  denoting the matrix of all inputs except  $X_i$ . The inner expectation operator is the mean of output  $Y$  taken over all possible values of the input matrix except  $X_i$ . The outer variance is taken over all possible values

of  $X_i$ . This is generally referred to as the *top marginal variance* of input  $X_i$  [209]. The top marginal variance determines the reduction in output variation if the input is fixed with the true value. In contrast,  $E_{X_{-i}}(V_{X_i}(Y|X_{-i}))$  is the expected variance that would remain if all inputs except  $X_i$  would be fixed. This is generally referred to as the *bottom marginal variance*. The first-order sensitivity indices are bound between 0 to 1 and provide a measure of relative importance with higher values implying larger effects on the outcome. The first-order sensitivity indices ( $S_i$ ) and the total-order sensitivity indices ( $S_{Ti}$ ) can be obtained by the following equations [209].

$$S_i = \frac{V_{X_i}(E_{X_{-i}}(Y|X_i))}{V(Y)} \quad (3.6)$$

$$S_{Ti} = \frac{E_{X_{-i}}(V_{X_i}(Y|X_{-i}))}{V(Y)} = 1 - \frac{V_{X_{-i}}(E_{X_i}(Y|X_{-i}))}{V(Y)} \quad (3.7)$$

First-order and total-order indices were computed after Sobol [210]. To sample the input parameter space, 10,000 draws were generated from each input distribution. The computational time was improved by applying the Saltelli [208] sampler which generated an input matrix of length  $N(k + 2)$  where  $N$  is the number of draws and  $k$  is the number of model inputs. The implementation used the improved formulas of Jansen [211] and Saltelli et al. [209]. In total, 31 parameters were included which resulted in 330,000 rows of input values for which economic impact was computed using one spread model (T2 - RR50 in Italy).

### 3.4.5 Economic Data

A detailed overview is provided in the Supplementary Material (Tab. 3.3). Prices and costs for olives in Euro per ton were obtained from the European Commission [185] with average prices in 2000 to 2009 of 481 and 497 Euro in Spain and Italy, respectively. The average costs of 247 and 316 Euro per ton for Spain and Italy comprise specific costs (fertilizers, crop protection, fuel, water, other specific costs), farming overheads (building and machinery upkeep, energy, contract work,

other direct costs), depreciation and external factors (wages, rent, interest). Prices of Spain were above Italian prices from 2000 to 2005. Since then, Italian prices have been higher than Spanish prices. This might be related to the recent droughts in Italy. In addition, we might expect the higher prices to be related to (perceived) differences in product quality as well as the culinary focus of the Italian culture which might result in consumers that are willing to pay more for their food products. Due to absence of published information on prices and costs from Greece we consulted an expert which resulted in an estimated price of 560 Euro and estimated costs of 387 Euro per ton. For the global sensitivity analysis, Italian prices (and costs) were sampled out of a normal distribution based on the computed means and standard deviations of 115 (52). The discount rate was set to 3 percent for the deterministic computation. For the global sensitivity analysis, the discount rate was sampled out of a uniform distribution between 3 and 7 percent which comprises values frequently used in similar studies [212, 213]. The prices were estimated to increase by 0.52 percent following a one percent decrease in supply. For the global sensitivity analysis, the estimated standard deviation of 0.1 was used to sample out of a normal distribution.

There is uncertainty in the agronomical literature regarding the longevity of the different cropping systems. The effects of higher tree densities on the longevity of orchards is not yet fully understood [214]. Rallo et al. [215] reported minimum values for the longevity of >100, >100, >40, >40, >20, >15 years, for the rain-fed-traditional, irrigated-traditional, rain-fed-intensive, irrigated-intensive, rain-fed-high-density and irrigated-high-density system, respectively. Data from Eurostat suggests that around 23.47 and 0.68 percent of orchards in the density classes intensive and high density, respectively, are older than 50 years. To allow for the empirically observed ages within the different density classes, the longevities were set to 75 and 55 years for the intensive and high-density systems, respectively. To acknowledge the differences between cropping systems after Rallo et al. [215], the longevity of the traditional systems was set to 135 years for the deterministic computation. For the global sensitivity analysis, the longevities were sampled from uniform distributions ranging from 135-270, 75-150, 55-110 for the traditional, intensive and high-density systems, respectively.

The full bearing tree ages were obtained from Rallo et al. [215]. For the global sensitivity analysis, the ranges reported by Rallo et al. [215] were sampled out of a uniform distribution. The full bearing yield potential was obtained from Rallo et al. [215]. For the global sensitivity analysis, the ranges reported by Rallo et al. [215] were sampled from a uniform distribution. The full bearing ages and full bearing yields were used to linearly interpolate the yields across ages. Subsequently, yields were rescaled to result in the empirically observed total production of olives according to FAOstat. For this, we made use of the 5-year averages (2007 to 2013) for Italy (3.19 million tons), Greece (2.37 million tons) and Spain (6.69 million tons). For the global sensitivity analysis, the estimated standard deviation for Italian supply (2007 to 2013) was used to sample out of a normal distribution. The replanting costs in Euro per hectare were obtained from Rallo et al. [215] for the deterministic assessment. For the global sensitivity analysis, the uncertainty regarding possible costs for uprooting as well as geographic differences in replanting costs was approached by sampling uniformly between the reported costs and double the amount.

The annual yield decline due to  $Xfp$  was set to 10 percent. The annual increase in cost due to  $Xfp$  was set to 10 percent. For the global sensitivity analysis, the absence of knowledge on these parameters was approached by sampling out of a uniform distribution minus 5 to minus 50 percent and minus 25 to plus 25 percent, respectively.

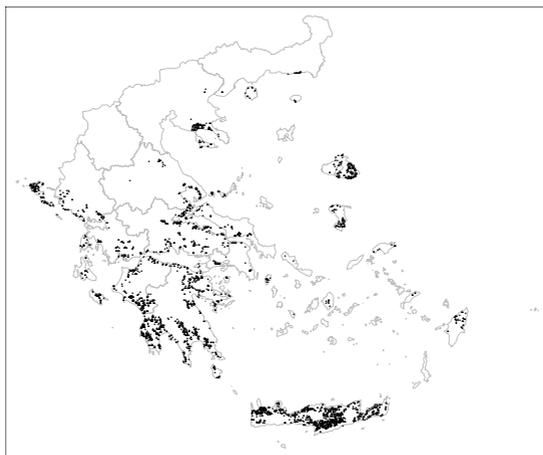
### 3.4.6 Data Archival

All computations were performed on a high-performance computing cluster. Data and R scripts are available at <https://doi.org/10.5281/zenodo.3672794>.

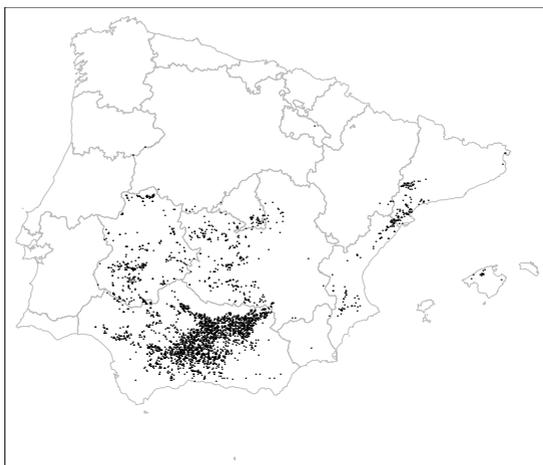
## Acknowledgements

This work was supported by funding from the European Union's Horizon 2020 Pest Organisms Threatening Europe (POnTE) research and innovation program under grant agreement number 635646. M.C. held an IVIA grant partially funded by the European Social Fund Comunitat Valenciana 2014-2020. We thank the European Food Safety Authority for the expert knowledge elicitation on the spread rate, Donato Boscia for sharing his expertise on the pathosystem, Lia Hemerik for her input on the spread model and the high-performance computing cluster of Wageningen University and Research (Anunna) for allowing us to perform all computations free of charge.

### 3.5 Supplementary Material

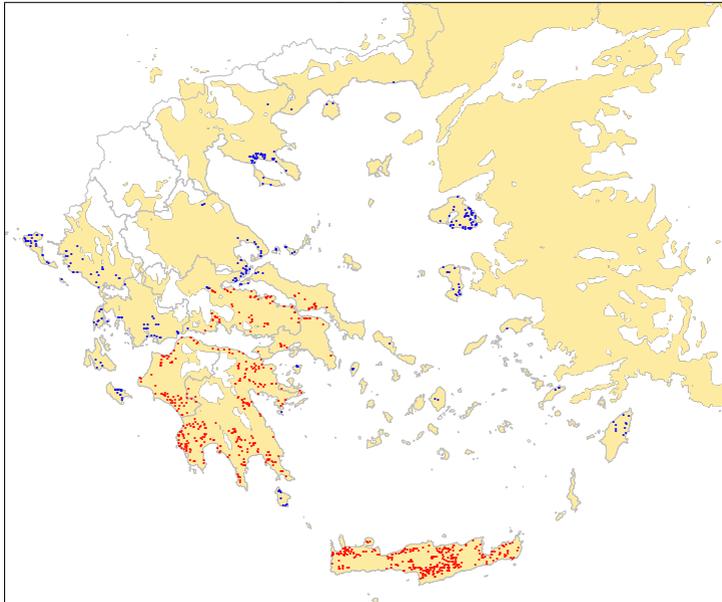


(a)

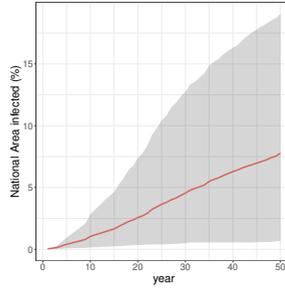


(b)

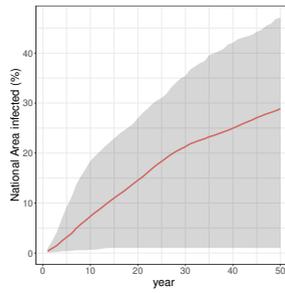
**Fig. 3.5** All randomized points of introduction for Greece and Spain.



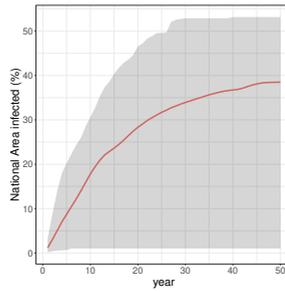
**Fig. 3.6** Points of introduction which resulted in an infected area less than (blue) and more or equal to (red) 30 percent of the Greek area of production for the 5.18 km per year spread rate and a climatic suitability threshold of 0.132 (suitable area in yellow).



(a)

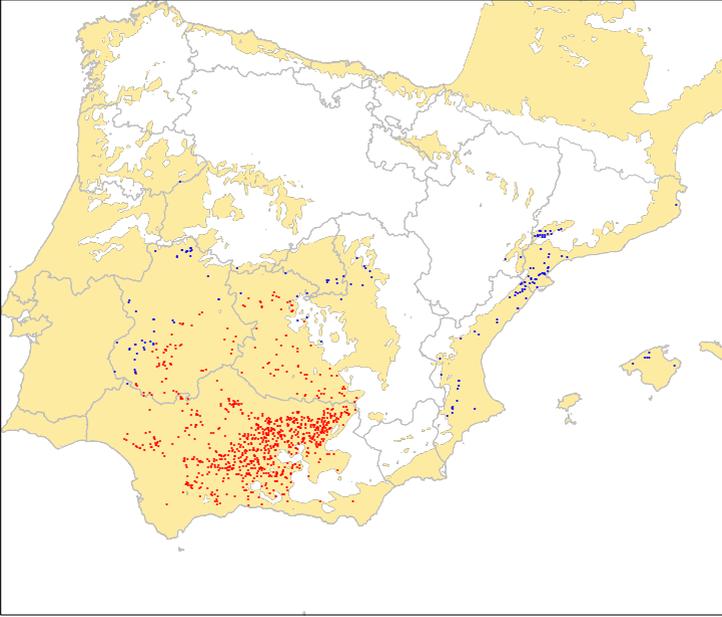


(b)

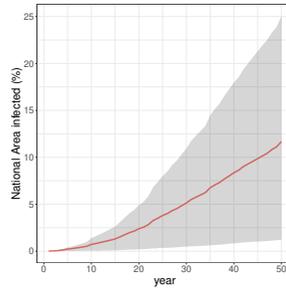


(c)

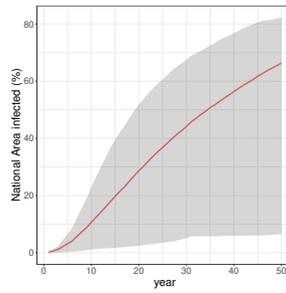
**Fig. 3.7** Uncertainty of the Greek area of production infected over time for the 1.1 km (a), 5.18 km (b) and 12.35 km (c) per year spread rate and a climatic suitability threshold ( $T_2$ ) of 0.132. The mean of 1000 epidemics from 1000 random points of introduction is indicated in red. The grey area represents the 90 percent range (from the 5 to 95 percentile) of the 1000 simulated epidemics.



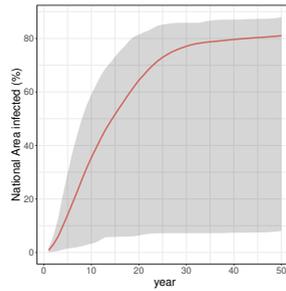
**Fig. 3.8** Points of introduction which resulted in an infected area less than (blue) and more or equal to (red) 50 percent of the Spanish area of production for the 5.18 km per year spread rate and a climatic suitability threshold of 0.132 (suitable area in yellow).



(a)

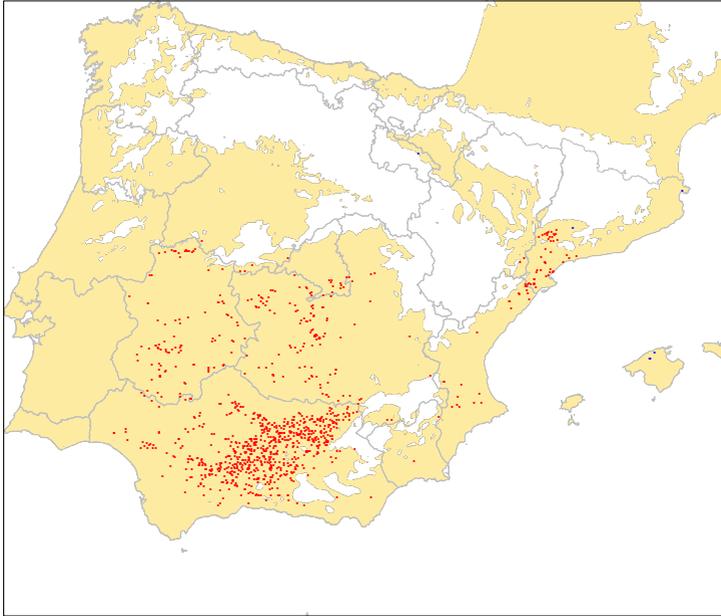


(b)

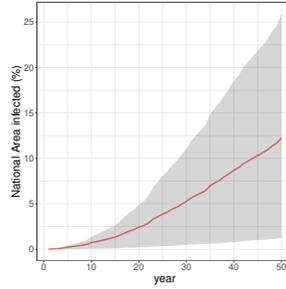


(c)

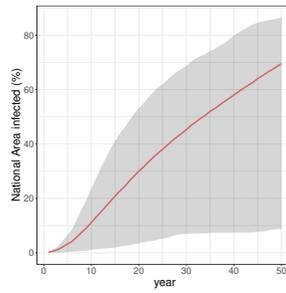
**Fig. 3.9** Uncertainty of the Spanish area of production infected over time for the 1.1 km (a), 5.18 km (b) and 12.35 km (c) per year spread rate and a climatic suitability threshold ( $T_2$ ) of 0.132. The mean of 1000 epidemics from 1000 random points of introduction is indicated in red. The grey area represents the 90 percent range (from the 5 to 95 percentile) of the 1000 simulated epidemics.



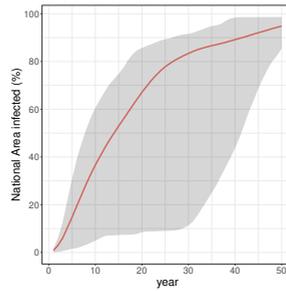
**Fig. 3.10** Points of introduction which resulted in an infected area less than (blue) and more or equal to (red) 50 percent of the Spanish area of production for the 12.35 km per year spread rate and a climatic suitability threshold of 0.093 (suitable area in yellow).



(a)

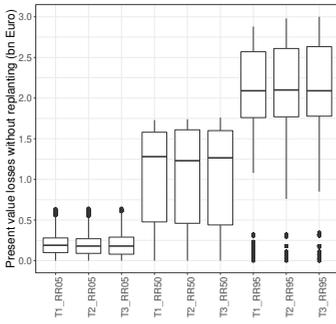


(b)

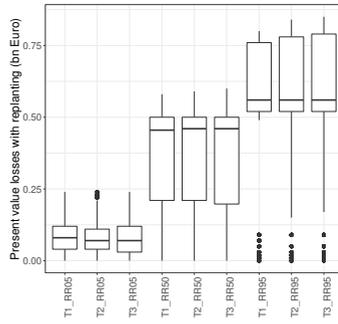


(c)

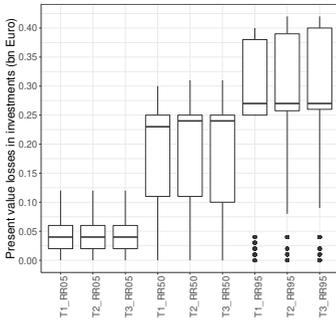
**Fig. 3.11** Uncertainty of the Spanish area of production infected over time for the 1.1 km (a), 5.18 km (b) and 12.35 km (c) per year spread rate and a climatic suitability threshold ( $T_3$ ) of 0.093. The mean of 1000 epidemics from 1000 random points of introduction is indicated in red. The grey area represents the 90 percent range (from the 5 to 95 percentile) of the 1000 simulated epidemics.



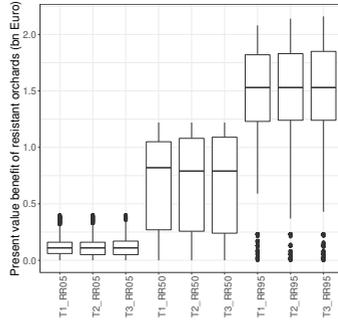
(a) Total impact without replanting



(b) Total impact with replanting

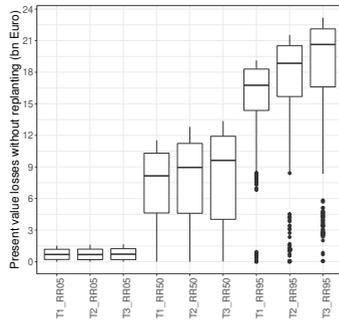


(c) Losses in investment

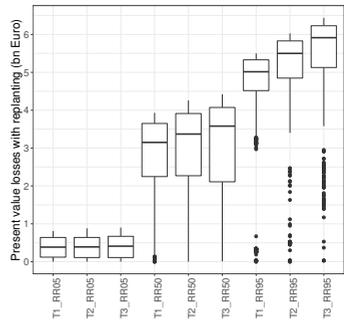


(d) Benefit of resistant orchards

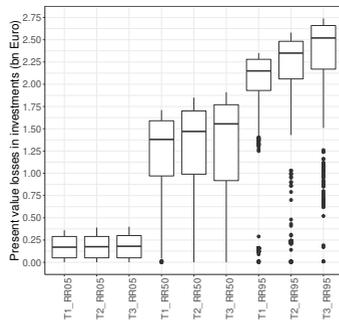
**Fig. 3.12** Distribution of economic results across all random points of introduction in Greece for all spread scenarios.



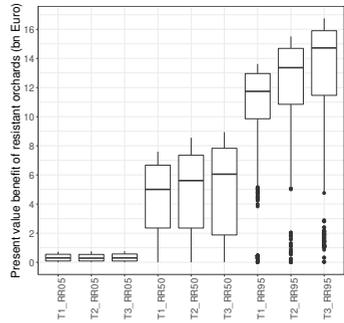
(a) Total impact without replanting



(b) Total impact with replanting



(c) Losses in investment



(d) Benefit of resistant orchards

**Fig. 3.13** Distribution of economic results across all random points of introduction in Spain for all spread scenarios.

**Table 3.3** Economic parameters for the deterministic analysis as well as the global sensitivity analysis.

Parameter	Deterministic	Global Sensitivity Analysis		
		Distribution	Mean/Min.	SD/Max.
Planning Horizon ( <i>years</i> )	50	-	-	-
Total area of production in Europe ( <i>ha</i> )	4,600,000	-	-	-
Total production in Italy ( <i>tons</i> )	3,188,712	normal	3,188,712	175,940
Total production in Greece ( <i>tons</i> )	2,367,466	-	-	-
Total production in Spain ( <i>tons</i> )	6,689,447	-	-	-
Price in Italy ( <i>Euro ton<sup>-1</sup></i> )	496	normal	496	116
Price in Greece ( <i>Euro ton<sup>-1</sup></i> )	560	-	-	-
Price in Spain ( <i>Euro ton<sup>-1</sup></i> )	481	-	-	-
Operating Costs in Italy ( <i>Euro ton<sup>-1</sup></i> )	316	normal	316	52
Operating Costs in Greece ( <i>Euro ton<sup>-1</sup></i> )	387	-	-	-
Operating Costs in Spain ( <i>Euro ton<sup>-1</sup></i> )	247	-	-	-
Price response (%)	0.52	normal	0.52	0.1
Discount rate (%)	3	uniform	3	7
Xfp yield-decline (%)	10	uniform	-50	-5
Xfp cost-change (%)	10	uniform	-25	25
Full-bearing age RT ( <i>years</i> )	10	uniform	10	15
Full-bearing age IT ( <i>years</i> )	8	uniform	8	10
Full-bearing age RI ( <i>years</i> )	6	uniform	6	8
Full-bearing age II ( <i>years</i> )	5	uniform	5	7
Full-bearing age RHD ( <i>years</i> )	4	uniform	4	5
Full-bearing age IHD ( <i>years</i> )	3	uniform	3	6
Yield potential RT ( <i>ton ha<sup>-1</sup></i> )	2.5	uniform	1	5
Yield potential IT ( <i>ton ha<sup>-1</sup></i> )	5	uniform	5	10
Yield potential RI ( <i>ton ha<sup>-1</sup></i> )	5	uniform	4	6
Yield potential II ( <i>ton ha<sup>-1</sup></i> )	8	uniform	8	10
Yield potential RHD ( <i>ton ha<sup>-1</sup></i> )	5	uniform	4	6
Yield potential IHD ( <i>ton ha<sup>-1</sup></i> )	10	uniform	10	16
Replanting cost RT ( <i>Euro ha<sup>-1</sup></i> )	500	uniform	500	1000
Replanting cost IT ( <i>Euro ha<sup>-1</sup></i> )	2500	uniform	2500	5000
Replanting cost RI ( <i>Euro ha<sup>-1</sup></i> )	750	uniform	750	1500
Replanting cost II ( <i>Euro ha<sup>-1</sup></i> )	3000	uniform	3000	6000
Replanting cost RHD ( <i>Euro ha<sup>-1</sup></i> )	3000	uniform	3000	6000
Replanting cost IHD ( <i>Euro ha<sup>-1</sup></i> )	6000	uniform	6000	12000
Longevity RT ( <i>years</i> )	135	uniform	135	270
Longevity IT ( <i>years</i> )	135	uniform	135	270
Longevity RI ( <i>years</i> )	75	uniform	75	150
Longevity II ( <i>years</i> )	75	uniform	75	150
Longevity RHD ( <i>years</i> )	55	uniform	55	110
Longevity IHD ( <i>years</i> )	55	uniform	55	110

Cropping systems are rainfed-traditional (RT), irrigated-traditional (IT), rainfed-intensive (RI), irrigated-intensive (II), rainfed-high-density (RHD) and irrigated-high-density (IHD)

Table 3.4 Economic impact over 50 years in billion Euro for all spread scenarios.

Spread Scenario	Without Replanting			With Replanting			Lost Investment		Benefits of Resistance							
	Total Impact ES	Change in Profits ES	IT	Total Impact ES	Change in Profits ES	IT	Investment	EL	ES	IT						
<i>Xfp spread in Italy</i>																
T1RR05	0.68	1.71	-1.86	0.68	1.71	-1.65	0.10	0.26	-0.59	0.10	0.26	-0.38	-0.21	-0.58	-1.45	1.27
T2RR05	0.68	1.71	-1.86	0.68	1.71	-1.65	0.10	0.26	-0.59	0.10	0.26	-0.38	-0.21	-0.58	-1.45	1.27
T3RR05	0.68	1.71	-1.86	0.68	1.71	-1.65	0.10	0.26	-0.59	0.10	0.26	-0.38	-0.21	-0.58	-1.45	1.27
T1RR50	1.03	2.59	-3.13	1.03	2.59	-2.73	0.16	0.42	-1.04	0.16	0.42	-0.64	-0.40	-0.86	-2.17	2.09
T2RR50	1.04	2.61	-3.15	1.04	2.61	-2.75	0.17	0.42	-1.05	0.17	0.42	-0.65	-0.40	-0.87	-2.18	2.10
T3RR50	1.05	2.64	-3.21	1.05	2.64	-2.80	0.17	0.43	-1.08	0.17	0.43	-0.66	-0.41	-0.88	-2.21	2.14
T1RR95	1.57	3.94	-5.10	1.57	3.94	-4.50	0.24	0.61	-1.55	0.24	0.61	-0.95	-0.60	-1.33	-3.34	3.55
T2RR95	1.59	3.99	-5.17	1.59	3.99	-4.56	0.24	0.62	-1.57	0.24	0.62	-0.96	-0.61	-1.34	-3.37	3.60
T3RR95	1.64	4.12	-5.41	1.64	4.12	-4.77	0.25	0.64	-1.64	0.25	0.64	-1.00	-0.63	-1.39	-3.48	3.77
<i>Spread in Greece<sup>a</sup></i>																
T1RR05	-0.22	0.18	0.09	-0.17	0.18	0.09	-0.09	0.03	0.01	-0.04	0.03	0.01	-0.05	0.13	-0.16	-0.07
T2RR05	-0.21	0.18	0.09	-0.17	0.18	0.09	-0.09	0.03	0.01	-0.04	0.03	0.01	-0.04	0.13	-0.15	-0.07
T3RR05	-0.22	0.18	0.09	-0.17	0.18	0.09	-0.09	0.03	0.01	-0.04	0.03	0.01	-0.05	0.13	-0.15	-0.07
T1RR50	-1.09	0.90	0.43	-0.90	0.90	0.43	-0.37	0.12	0.06	-0.18	0.12	0.06	-0.19	0.72	-0.78	-0.37
T2RR50	-1.08	0.89	0.42	-0.89	0.89	0.42	-0.37	0.12	0.06	-0.18	0.12	0.06	-0.19	0.71	-0.77	-0.37
T3RR50	-1.07	0.88	0.42	-0.88	0.88	0.42	-0.36	0.12	0.06	-0.18	0.12	0.06	-0.18	0.70	-0.76	-0.37
T1RR95	-1.87	1.56	0.75	-1.60	1.56	0.75	-0.54	0.18	0.09	-0.28	0.18	0.09	-0.27	1.33	-1.38	-0.66
T2RR95	-1.94	1.60	0.76	-1.65	1.60	0.76	-0.58	0.19	0.09	-0.29	0.19	0.09	-0.29	1.36	-1.41	-0.67
T3RR95	-1.91	1.58	0.75	-1.63	1.58	0.75	-0.57	0.19	0.09	-0.29	0.19	0.09	-0.28	1.34	-1.39	-0.66
<i>Spread in Spain<sup>a</sup></i>																
T1RR05	0.21	-0.70	0.25	0.21	-0.53	0.25	0.05	-0.38	0.06	0.05	-0.21	0.06	-0.17	-0.16	0.32	-0.20
T2RR05	0.21	-0.71	0.25	0.21	-0.53	0.25	0.05	-0.39	0.06	0.05	-0.21	0.06	-0.17	-0.16	0.32	-0.19
T3RR05	0.22	-0.73	0.26	0.22	-0.55	0.26	0.05	-0.40	0.06	0.05	-0.22	0.06	-0.18	-0.17	0.33	-0.20
T1RR50	1.79	-7.21	2.14	1.79	-6.01	2.14	0.35	-2.76	0.42	0.35	-1.56	0.42	-1.21	-1.44	4.45	-1.72
T2RR50	1.90	-7.83	2.27	1.90	-6.55	2.27	0.37	-2.93	0.44	0.37	-1.65	0.44	-1.27	-1.53	4.90	-1.83
T3RR50	1.96	-8.20	2.35	1.96	-6.88	2.35	0.38	-3.04	0.46	0.38	-1.72	0.46	-1.32	-1.58	5.16	-1.89
T1RR95	3.54	-15.33	4.24	3.54	-13.34	4.24	0.57	-4.63	0.69	0.57	-2.65	0.69	-1.99	-2.97	10.69	-3.56
T2RR95	3.76	-16.86	4.50	3.76	-14.74	4.50	0.61	-4.98	0.73	0.61	-2.85	0.73	-2.12	-3.15	11.88	-3.77
T3RR95	4.01	-18.59	4.80	4.01	-16.26	4.80	0.66	-5.48	0.80	0.66	-3.15	0.80	-2.33	-3.35	13.11	-4.00

<sup>a</sup>averaged over all random points of introduction. EL=Greece, ES=Spain, IT=Italy. All results are for a climatic suitability threshold of 0.132.

**Table 3.5** First order sensitivity indices for all parameters for the economic impact without replanting.

Parameter	Index	Bias	Std.Error	Min. CI	Max. CI
Price in Italy ( <i>Euro ton</i> <sup>-1</sup> )	0.6643	-0.0001	0.0069	0.6509	0.6785
Price response (%)	0.0034	-0.0003	0.0145	-0.0234	0.0348
Costs in Italy ( <i>Euro ton</i> <sup>-1</sup> )	0.1649	-0.0008	0.0134	0.1388	0.1925
Total production IT ( <i>ton</i> )	-0.0004	-0.0003	0.0147	-0.0274	0.0315
Discount rate (%)	0.0331	0.0000	0.0165	0.0024	0.0687
Xfp yield-decline (%)	0.0226	-0.0004	0.0146	-0.0044	0.0526
Xfp cost-change (%)	0.0196	-0.0005	0.0146	-0.0069	0.0504
Longevity RT ( <i>years</i> )	0.0008	-0.0003	0.0146	-0.0264	0.0328
Longevity IT ( <i>years</i> )	0.0007	-0.0003	0.0146	-0.0259	0.0318
Longevity RI ( <i>years</i> )	0.0038	-0.0003	0.0145	-0.0228	0.0351
Longevity II ( <i>years</i> )	-0.0002	-0.0003	0.0146	-0.0272	0.0318
Longevity RHD ( <i>years</i> )	0.0000	-0.0003	0.0146	-0.0269	0.0315
Longevity IHD ( <i>years</i> )	0.0002	-0.0003	0.0146	-0.0265	0.0319
Replanting cost RT ( <i>Euro ha</i> <sup>-1</sup> )	0.0008	-0.0003	0.0145	-0.0260	0.0319
Replanting cost IT ( <i>Euro ha</i> <sup>-1</sup> )	0.0000	-0.0003	0.0145	-0.0269	0.0316
Replanting cost RI ( <i>Euro ha</i> <sup>-1</sup> )	0.0002	-0.0003	0.0146	-0.0266	0.0311
Replanting cost II ( <i>Euro ha</i> <sup>-1</sup> )	0.0004	-0.0003	0.0145	-0.0266	0.0317
Replanting cost RHD ( <i>Euro ha</i> <sup>-1</sup> )	0.0001	-0.0003	0.0146	-0.0268	0.0316
Replanting cost IHD ( <i>Euro ha</i> <sup>-1</sup> )	0.0002	-0.0003	0.0146	-0.0266	0.0319
Full-bearing age RT ( <i>years</i> )	0.0002	-0.0003	0.0145	-0.0266	0.0318
Full-bearing age IT ( <i>years</i> )	0.0002	-0.0003	0.0146	-0.0266	0.0318
Full-bearing age RI ( <i>years</i> )	0.0001	-0.0003	0.0145	-0.0268	0.0317
Full-bearing age II ( <i>years</i> )	0.0002	-0.0003	0.0146	-0.0266	0.0318
Full-bearing age RHD ( <i>years</i> )	0.0002	-0.0003	0.0146	-0.0266	0.0318
Full-bearing age IHD ( <i>years</i> )	0.0002	-0.0003	0.0146	-0.0266	0.0318
Yield Potential RT ( <i>ton ha</i> <sup>-1</sup> )	0.0001	-0.0003	0.0145	-0.0268	0.0319
Yield Potential IT ( <i>ton ha</i> <sup>-1</sup> )	0.0000	-0.0003	0.0146	-0.0271	0.0316
Yield Potential RI ( <i>ton ha</i> <sup>-1</sup> )	0.0003	-0.0003	0.0145	-0.0267	0.0319
Yield Potential II ( <i>ton ha</i> <sup>-1</sup> )	0.0001	-0.0003	0.0145	-0.0269	0.0318
Yield Potential RHD ( <i>ton ha</i> <sup>-1</sup> )	-0.0001	-0.0003	0.0146	-0.0270	0.0315
Yield Potential IHD ( <i>ton ha</i> <sup>-1</sup> )	0.0001	-0.0003	0.0146	-0.0268	0.0317

Cropping systems are denoted:

rainfed-traditional (RT), irrigated-traditional (IT)

rainfed-intensive (RI), irrigated-intensive (II)

rainfed-high-density (RHD) and irrigated-high-density (IHD)

**Table 3.6** Total order sensitivity indices for all parameters for the economic impact without replanting.

Parameter	Index	Bias	Std.Error	Min. CI	Max. CI
Price in Italy ( <i>Euro ton<sup>-1</sup></i> )	0.7171	0.0007	0.0123	0.6922	0.7410
Price response (%)	0.0039	0.0000	0.0001	0.0037	0.0040
Costs in Italy ( <i>Euro ton<sup>-1</sup></i> )	0.1689	0.0002	0.0034	0.1620	0.1760
Total production IT ( <i>ton</i> )	0.0044	0.0000	0.0001	0.0041	0.0046
Discount rate (%)	0.0703	0.0000	0.0018	0.0667	0.0737
Xfp yield-decline (%)	0.0727	0.0001	0.0023	0.0679	0.0770
Xfp cost-change (%)	0.0577	0.0001	0.0018	0.0542	0.0612
Longevity RT ( <i>years</i> )	0.0020	0.0000	0.0000	0.0019	0.0021
Longevity IT ( <i>years</i> )	0.0008	0.0000	0.0000	0.0008	0.0008
Longevity RI ( <i>years</i> )	0.0030	0.0000	0.0001	0.0028	0.0031
Longevity II ( <i>years</i> )	0.0007	0.0000	0.0000	0.0007	0.0007
Longevity RHD ( <i>years</i> )	0.0002	0.0000	0.0000	0.0002	0.0002
Longevity IHD ( <i>years</i> )	0.0000	0.0000	0.0000	0.0000	0.0000
Replanting cost RT ( <i>Euro ha<sup>-1</sup></i> )	0.0010	0.0000	0.0000	0.0009	0.0010
Replanting cost IT ( <i>Euro ha<sup>-1</sup></i> )	0.0005	0.0000	0.0000	0.0005	0.0005
Replanting cost RI ( <i>Euro ha<sup>-1</sup></i> )	0.0008	0.0000	0.0000	0.0007	0.0008
Replanting cost II ( <i>Euro ha<sup>-1</sup></i> )	0.0002	0.0000	0.0000	0.0002	0.0003
Replanting cost RHD ( <i>Euro ha<sup>-1</sup></i> )	0.0001	0.0000	0.0000	0.0001	0.0001
Replanting cost IHD ( <i>Euro ha<sup>-1</sup></i> )	0.0000	0.0000	0.0000	0.0000	0.0000
Full-bearing age RT ( <i>years</i> )	0.0000	0.0000	0.0000	0.0000	0.0000
Full-bearing age IT ( <i>years</i> )	0.0000	0.0000	0.0000	0.0000	0.0000
Full-bearing age RI ( <i>years</i> )	0.0000	0.0000	0.0000	0.0000	0.0000
Full-bearing age II ( <i>years</i> )	0.0000	0.0000	0.0000	0.0000	0.0000
Full-bearing age RHD ( <i>years</i> )	0.0000	0.0000	0.0000	0.0000	0.0000
Full-bearing age IHD ( <i>years</i> )	0.0000	0.0000	0.0000	0.0000	0.0000
Yield Potential RT ( <i>ton ha<sup>-1</sup></i> )	0.0006	0.0000	0.0000	0.0006	0.0007
Yield Potential IT ( <i>ton ha<sup>-1</sup></i> )	0.0003	0.0000	0.0000	0.0002	0.0003
Yield Potential RI ( <i>ton ha<sup>-1</sup></i> )	0.0004	0.0000	0.0000	0.0003	0.0004
Yield Potential II ( <i>ton ha<sup>-1</sup></i> )	0.0001	0.0000	0.0000	0.0001	0.0002
Yield Potential RHD ( <i>ton ha<sup>-1</sup></i> )	0.0001	0.0000	0.0000	0.0001	0.0001
Yield Potential IHD ( <i>ton ha<sup>-1</sup></i> )	0.0000	0.0000	0.0000	0.0000	0.0000

Cropping systems are denoted:

rainfed-traditional (RT), irrigated-traditional (IT)

rainfed-intensive (RI), irrigated-intensive (II)

rainfed-high-density (RHD) and irrigated-high-density (IHD)

**Table 3.7** First order sensitivity indices for all parameters for the economic impact with replanting.

Parameter	Index	Bias	Std.Error	Min. CI	Max. CI
Price in Italy ( <i>Euro ton</i> <sup>-1</sup> )	0.6186	0.0001	0.0112	0.5976	0.6405
Price response (%)	-0.0117	0.0001	0.0192	-0.0481	0.0275
Costs in Italy ( <i>Euro ton</i> <sup>-1</sup> )	0.1421	0.0000	0.0194	0.1050	0.1837
Total production IT ( <i>ton</i> )	-0.0129	0.0001	0.0194	-0.0504	0.0271
Discount rate (%)	-0.0096	0.0001	0.0215	-0.0510	0.0348
Xfp yield-decline (%)	-0.0065	0.0001	0.0212	-0.0482	0.0373
Xfp cost-change (%)	-0.0260	0.0003	0.0219	-0.0687	0.0191
Longevity RT ( <i>years</i> )	-0.0108	0.0001	0.0190	-0.0475	0.0286
Longevity IT ( <i>years</i> )	-0.0129	0.0001	0.0190	-0.0497	0.0269
Longevity RI ( <i>years</i> )	-0.0043	0.0001	0.0189	-0.0408	0.0347
Longevity II ( <i>years</i> )	-0.0153	0.0001	0.0190	-0.0527	0.0241
Longevity RHD ( <i>years</i> )	-0.0148	0.0001	0.0190	-0.0517	0.0248
Longevity IHD ( <i>years</i> )	-0.0151	0.0001	0.0190	-0.0524	0.0244
Replanting cost RT ( <i>Euro ha</i> <sup>-1</sup> )	-0.0126	0.0001	0.0191	-0.0501	0.0268
Replanting cost IT ( <i>Euro ha</i> <sup>-1</sup> )	-0.0148	0.0001	0.0190	-0.0519	0.0249
Replanting cost RI ( <i>Euro ha</i> <sup>-1</sup> )	-0.0102	0.0002	0.0190	-0.0472	0.0296
Replanting cost II ( <i>Euro ha</i> <sup>-1</sup> )	-0.0140	0.0001	0.0190	-0.0509	0.0249
Replanting cost RHD ( <i>Euro ha</i> <sup>-1</sup> )	-0.0146	0.0001	0.0190	-0.0520	0.0255
Replanting cost IHD ( <i>Euro ha</i> <sup>-1</sup> )	-0.0152	0.0001	0.0190	-0.0525	0.0245
Full-bearing age RT ( <i>years</i> )	-0.0156	0.0001	0.0190	-0.0529	0.0240
Full-bearing age IT ( <i>years</i> )	-0.0152	0.0001	0.0190	-0.0526	0.0246
Full-bearing age RI ( <i>years</i> )	-0.0150	0.0001	0.0190	-0.0521	0.0249
Full-bearing age II ( <i>years</i> )	-0.0152	0.0001	0.0190	-0.0524	0.0246
Full-bearing age RHD ( <i>years</i> )	-0.0153	0.0001	0.0190	-0.0525	0.0245
Full-bearing age IHD ( <i>years</i> )	-0.0152	0.0001	0.0190	-0.0526	0.0246
Yield Potential RT ( <i>ton ha</i> <sup>-1</sup> )	-0.0133	0.0001	0.0190	-0.0504	0.0265
Yield Potential IT ( <i>ton ha</i> <sup>-1</sup> )	-0.0161	0.0001	0.0191	-0.0534	0.0239
Yield Potential RI ( <i>ton ha</i> <sup>-1</sup> )	-0.0144	0.0001	0.0190	-0.0515	0.0261
Yield Potential II ( <i>ton ha</i> <sup>-1</sup> )	-0.0151	0.0001	0.0190	-0.0524	0.0246
Yield Potential RHD ( <i>ton ha</i> <sup>-1</sup> )	-0.0152	0.0001	0.0191	-0.0528	0.0248
Yield Potential IHD ( <i>ton ha</i> <sup>-1</sup> )	-0.0153	0.0001	0.0190	-0.0526	0.0245

Cropping systems are denoted:

rainfed-traditional (RT), irrigated-traditional (IT)

rainfed-intensive (RI), irrigated-intensive (II)

rainfed-high-density (RHD) and irrigated-high-density (IHD)

**Table 3.8** Total order sensitivity indices for all parameters for the economic impact with replanting.

Parameter	Index	Bias	Std.Error	Min. CI	Max. CI
Price in Italy ( <i>Euro ton<sup>-1</sup></i> )	0.7565	0.0001	0.0151	0.7266	0.7853
Price response (%)	0.0029	0.0000	0.0001	0.0027	0.0031
Costs in Italy ( <i>Euro ton<sup>-1</sup></i> )	0.2231	0.0002	0.0064	0.2092	0.2352
Total production IT ( <i>ton</i> )	0.0038	0.0000	0.0002	0.0033	0.0042
Discount rate (%)	0.0595	0.0000	0.0030	0.0534	0.0654
Xfp yield-decline (%)	0.0965	0.0002	0.0039	0.0889	0.1039
Xfp cost-change (%)	0.0907	0.0000	0.0040	0.0824	0.0984
Longevity RT ( <i>years</i> )	0.0082	0.0000	0.0002	0.0077	0.0086
Longevity IT ( <i>years</i> )	0.0033	0.0000	0.0001	0.0031	0.0035
Longevity RI ( <i>years</i> )	0.0104	0.0000	0.0003	0.0098	0.0110
Longevity II ( <i>years</i> )	0.0026	0.0000	0.0001	0.0025	0.0028
Longevity RHD ( <i>years</i> )	0.0008	0.0000	0.0000	0.0008	0.0009
Longevity IHD ( <i>years</i> )	0.0001	0.0000	0.0000	0.0001	0.0001
Replanting cost RT ( <i>Euro ha<sup>-1</sup></i> )	0.0049	0.0000	0.0001	0.0046	0.0052
Replanting cost IT ( <i>Euro ha<sup>-1</sup></i> )	0.0023	0.0000	0.0001	0.0022	0.0024
Replanting cost RI ( <i>Euro ha<sup>-1</sup></i> )	0.0056	0.0000	0.0002	0.0051	0.0060
Replanting cost II ( <i>Euro ha<sup>-1</sup></i> )	0.0015	0.0000	0.0000	0.0014	0.0016
Replanting cost RHD ( <i>Euro ha<sup>-1</sup></i> )	0.0005	0.0000	0.0000	0.0004	0.0005
Replanting cost IHD ( <i>Euro ha<sup>-1</sup></i> )	0.0000	0.0000	0.0000	0.0000	0.0000
Full-bearing age RT ( <i>years</i> )	0.0002	0.0000	0.0000	0.0002	0.0002
Full-bearing age IT ( <i>years</i> )	0.0000	0.0000	0.0000	0.0000	0.0000
Full-bearing age RI ( <i>years</i> )	0.0002	0.0000	0.0000	0.0002	0.0002
Full-bearing age II ( <i>years</i> )	0.0000	0.0000	0.0000	0.0000	0.0000
Full-bearing age RHD ( <i>years</i> )	0.0000	0.0000	0.0000	0.0000	0.0000
Full-bearing age IHD ( <i>years</i> )	0.0000	0.0000	0.0000	0.0000	0.0000
Yield Potential RT ( <i>ton ha<sup>-1</sup></i> )	0.0019	0.0000	0.0002	0.0016	0.0022
Yield Potential IT ( <i>ton ha<sup>-1</sup></i> )	0.0005	0.0000	0.0000	0.0004	0.0006
Yield Potential RI ( <i>ton ha<sup>-1</sup></i> )	0.0008	0.0000	0.0001	0.0006	0.0009
Yield Potential II ( <i>ton ha<sup>-1</sup></i> )	0.0002	0.0000	0.0001	0.0001	0.0003
Yield Potential RHD ( <i>ton ha<sup>-1</sup></i> )	0.0001	0.0000	0.0000	0.0001	0.0002
Yield Potential IHD ( <i>ton ha<sup>-1</sup></i> )	0.0000	0.0000	0.0000	0.0000	0.0000

Cropping systems are denoted:

rained-traditional (RT), irrigated-traditional (IT)

rained-intensive (RI), irrigated-intensive (II)

rained-high-density (RHD) and irrigated-high-density (IHD)

**Table 3.9** First order sensitivity indices for all parameters for the benefit of resistance breeding.

Parameter	Index	Bias	Std.Error	Min. CI	Max. CI
Price in Italy ( <i>Euro ton</i> <sup>-1</sup> )	0.4805	-0.0001	0.0128	0.4560	0.5060
Price response (%)	0.0018	-0.0006	0.0164	-0.0303	0.0338
Costs in Italy ( <i>Euro ton</i> <sup>-1</sup> )	0.1166	0.0000	0.0156	0.0856	0.1478
Total production IT ( <i>ton</i> )	-0.0034	-0.0007	0.0166	-0.0357	0.0291
Discount rate (%)	0.0743	-0.0007	0.0189	0.0386	0.1138
Xfp yield-decline (%)	0.0398	-0.0005	0.0163	0.0091	0.0739
Xfp cost-change (%)	0.0388	-0.0009	0.0164	0.0083	0.0729
Longevity RT ( <i>years</i> )	-0.0015	-0.0006	0.0164	-0.0336	0.0309
Longevity IT ( <i>years</i> )	-0.0016	-0.0006	0.0164	-0.0338	0.0309
Longevity RI ( <i>years</i> )	-0.0013	-0.0006	0.0164	-0.0336	0.0309
Longevity II ( <i>years</i> )	-0.0016	-0.0006	0.0164	-0.0337	0.0308
Longevity RHD ( <i>years</i> )	-0.0017	-0.0006	0.0164	-0.0340	0.0308
Longevity IHD ( <i>years</i> )	-0.0015	-0.0006	0.0164	-0.0337	0.0310
Replanting cost RT ( <i>Euro ha</i> <sup>-1</sup> )	-0.0015	-0.0006	0.0164	-0.0332	0.0311
Replanting cost IT ( <i>Euro ha</i> <sup>-1</sup> )	-0.0013	-0.0006	0.0164	-0.0338	0.0308
Replanting cost RI ( <i>Euro ha</i> <sup>-1</sup> )	-0.0019	-0.0005	0.0164	-0.0337	0.0309
Replanting cost II ( <i>Euro ha</i> <sup>-1</sup> )	-0.0017	-0.0006	0.0164	-0.0336	0.0306
Replanting cost RHD ( <i>Euro ha</i> <sup>-1</sup> )	-0.0016	-0.0006	0.0164	-0.0339	0.0309
Replanting cost IHD ( <i>Euro ha</i> <sup>-1</sup> )	-0.0016	-0.0006	0.0164	-0.0337	0.0310
Full-bearing age RT ( <i>years</i> )	-0.0010	-0.0006	0.0164	-0.0332	0.0316
Full-bearing age IT ( <i>years</i> )	-0.0016	-0.0006	0.0164	-0.0337	0.0309
Full-bearing age RI ( <i>years</i> )	-0.0019	-0.0006	0.0164	-0.0339	0.0304
Full-bearing age II ( <i>years</i> )	-0.0016	-0.0006	0.0164	-0.0339	0.0310
Full-bearing age RHD ( <i>years</i> )	-0.0016	-0.0006	0.0164	-0.0338	0.0309
Full-bearing age IHD ( <i>years</i> )	-0.0016	-0.0006	0.0164	-0.0338	0.0309
Yield Potential RT ( <i>ton ha</i> <sup>-1</sup> )	-0.0030	-0.0006	0.0164	-0.0351	0.0296
Yield Potential IT ( <i>ton ha</i> <sup>-1</sup> )	-0.0015	-0.0006	0.0164	-0.0338	0.0308
Yield Potential RI ( <i>ton ha</i> <sup>-1</sup> )	-0.0024	-0.0006	0.0164	-0.0344	0.0302
Yield Potential II ( <i>ton ha</i> <sup>-1</sup> )	-0.0022	-0.0006	0.0164	-0.0341	0.0301
Yield Potential RHD ( <i>ton ha</i> <sup>-1</sup> )	-0.0019	-0.0006	0.0164	-0.0342	0.0307
Yield Potential IHD ( <i>ton ha</i> <sup>-1</sup> )	-0.0017	-0.0006	0.0164	-0.0339	0.0308

Cropping systems are denoted:

rainfed-traditional (RT), irrigated-traditional (IT)

rainfed-intensive (RI), irrigated-intensive (II)

rainfed-high-density (RHD) and irrigated-high-density (IHD)

**Table 3.10** Total order sensitivity indices for all parameters for the benefit of resistance breeding.

Parameter	Index	Bias	Std.Error	Min. CI	Max. CI
Price in Italy ( <i>Euro ton<sup>-1</sup></i> )	0.6297	-0.0002	0.0120	0.6057	0.6541
Price response (%)	0.0045	0.0000	0.0001	0.0043	0.0047
Costs in Italy ( <i>Euro ton<sup>-1</sup></i> )	0.1486	0.0000	0.0036	0.1416	0.1557
Total production IT ( <i>ton</i> )	0.0057	0.0000	0.0002	0.0053	0.0060
Discount rate (%)	0.1603	0.0003	0.0035	0.1530	0.1673
Xfp yield-decline (%)	0.1905	-0.0004	0.0079	0.1758	0.2060
Xfp cost-change (%)	0.1630	-0.0001	0.0066	0.1495	0.1761
Longevity RT ( <i>years</i> )	0.0000	0.0000	0.0000	0.0000	0.0000
Longevity IT ( <i>years</i> )	0.0000	0.0000	0.0000	0.0000	0.0000
Longevity RI ( <i>years</i> )	0.0002	0.0000	0.0000	0.0002	0.0002
Longevity II ( <i>years</i> )	0.0000	0.0000	0.0000	0.0000	0.0000
Longevity RHD ( <i>years</i> )	0.0000	0.0000	0.0000	0.0000	0.0000
Longevity IHD ( <i>years</i> )	0.0000	0.0000	0.0000	0.0000	0.0000
Replanting cost RT ( <i>Euro ha<sup>-1</sup></i> )	0.0004	0.0000	0.0000	0.0004	0.0005
Replanting cost IT ( <i>Euro ha<sup>-1</sup></i> )	0.0002	0.0000	0.0000	0.0002	0.0002
Replanting cost RI ( <i>Euro ha<sup>-1</sup></i> )	0.0011	0.0000	0.0001	0.0008	0.0013
Replanting cost II ( <i>Euro ha<sup>-1</sup></i> )	0.0002	0.0000	0.0000	0.0002	0.0002
Replanting cost RHD ( <i>Euro ha<sup>-1</sup></i> )	0.0001	0.0000	0.0000	0.0001	0.0001
Replanting cost IHD ( <i>Euro ha<sup>-1</sup></i> )	0.0000	0.0000	0.0000	0.0000	0.0000
Full-bearing age RT ( <i>years</i> )	0.0001	0.0000	0.0000	0.0001	0.0002
Full-bearing age IT ( <i>years</i> )	0.0000	0.0000	0.0000	0.0000	0.0000
Full-bearing age RI ( <i>years</i> )	0.0002	0.0000	0.0000	0.0002	0.0002
Full-bearing age II ( <i>years</i> )	0.0000	0.0000	0.0000	0.0000	0.0000
Full-bearing age RHD ( <i>years</i> )	0.0000	0.0000	0.0000	0.0000	0.0000
Full-bearing age IHD ( <i>years</i> )	0.0000	0.0000	0.0000	0.0000	0.0000
Yield Potential RT ( <i>ton ha<sup>-1</sup></i> )	0.0014	0.0000	0.0001	0.0013	0.0016
Yield Potential IT ( <i>ton ha<sup>-1</sup></i> )	0.0005	0.0000	0.0000	0.0004	0.0005
Yield Potential RI ( <i>ton ha<sup>-1</sup></i> )	0.0008	0.0000	0.0000	0.0007	0.0009
Yield Potential II ( <i>ton ha<sup>-1</sup></i> )	0.0003	0.0000	0.0000	0.0002	0.0003
Yield Potential RHD ( <i>ton ha<sup>-1</sup></i> )	0.0002	0.0000	0.0000	0.0001	0.0002
Yield Potential IHD ( <i>ton ha<sup>-1</sup></i> )	0.0001	0.0000	0.0000	0.0000	0.0001

Cropping systems are denoted:

rainfed-traditional (RT), irrigated-traditional (IT)

rainfed-intensive (RI), irrigated-intensive (II)

rainfed-high-density (RHD) and irrigated-high-density (IHD)



## Chapter 4

# On Consumer Impact from *Xylella fastidiosa* subspecies *pauca*

### Abstract

The introduction of *Xylella fastidiosa* in Apulia has resulted in the desiccation of millions of olive trees. Here, we employ a multi-country partial equilibrium model to analyze the possible distribution of economic impacts among olive oil processors and consumers. The results suggest that the majority of the impacts would fall on consumers as a consequence of higher prices. If the disease disperses beyond the current extent in Italy the decline in consumer welfare ranges from 4.1 billion to 10.3 billion Euro over the course of 50 years depending on the rate of disease spread. In other words, each of the 195 million households in Europe would incur additional costs ranging 63 cents to 1.6 Euro every year over the course of 50 years. Introductions of the pathogen into Greece or Spain could cost European consumers between 0.4 billion to 3.3 billion Euro and 1.8 billion to 5.3 billion Euro, respectively. This would correspond to additional annual household costs ranging 6 to 51 cents and 27 cents to 8.2 Euro, respectively. As significant economic consequences from further dispersal of the disease are borne by consumers, the economic threat is not limited to producers but should be contextualized as a societal problem.

## 4.1 Introduction

The introduction of invasive species can have major impacts on economies, ecosystems and societies, and can cost taxpayers billions of dollars annually [216, 217]. New invasive species introductions are driven by global trade and travel [5, 7]. While the majority of introductions do not result in significant impacts [8], introductions of hazardous organisms can have severe consequences. The agricultural sector is particularly vulnerable to introductions of livestock diseases or plant pathogenic organisms [9]. Here, invasive species can lead to a reduction of food supply which can adversely affect consumers through higher prices [10, 11], as well as reduced food quality, food safety [12], and food security [13].

Olives have been at the agronomical, cultural and culinary heart of the Mediterranean basin for centuries [218, 219]. Olive trees can get very old, in the order of hundreds of years [215], and old trees contribute to the agro-ecological landscape and cultural heritage in the Mediterranean basin. The vast majority of the harvested olives are processed into oil [170, 219]. Around 3 billion tonnes of olive oil are produced globally out of which nearly three quarters originate from Italy, Greece and Spain [170].

The European olive production is under threat due to the invasion of the bacterium *Xylella fastidiosa* subspecies *pauca* (Xfp). *Xylella fastidiosa* (Xf) can infect several hundred plant species and is considered one of the most dangerous plant pathogenic bacteria worldwide [15, 17]. The subspecies *pauca* was first detected in the European territory in 2013 and is causing the Olive Quick Decline Syndrome which has spread across the southern part of the province Apulia in Italy [24]. In susceptible hosts, Xf obstructs the xylem leading to desiccation [20, 22]. The detection of the bacterium and the devastating impacts of the disease led to the enactment of control measures including vector control and felling of healthy trees to prevent spread by establishing a cordon sanitaire around the infected area [169, 174]. The implementation of tree felling as a measure raised dramatic societal unrest in the affected region [25, 26]. Insights into potential consequences to citizens from further spread of the pathogen could inform the public debate on appropriate measures against Xfp.

Market responses to further spread of the pathogen, and consequences to consumers, will crucially depend on the ability of suppliers to adapt. The elasticity of supply provides a measure of the responsiveness of a production system to price changes. Various factors determine the ability of producers to adapt the production volume to price changes, such as the length of a production cycle, the mobility of the operation and the ability to store the product. Price responses following reductions in supply due to an invasive species can be particularly severe if the supply of the host plant is inelastic. Inelastic supply is inherently connected with the production of olives, which requires considerable up-front investments [220], putting a ‘natural’ barrier on market entry; in addition, olive trees take several years to reach a full bearing state [215], thereby causing a considerable delay between a price signal and adaptations in production.<sup>1</sup>

This paper aims to inform the ongoing discussion on measures against invasive species by developing insights into potential economic consequences to European consumers in the case of Xfp. It is argued that the problem of invasive species should be contextualized as a societal challenge in public debates on appropriate management strategies since negative consequences arise to both affected producers and consumers. From a methodological point of view, this paper stresses the need to employ multi-country models in the context of biological invasions into interdependent markets [221]. Furthermore, we show how a joint use of scenarios, global sensitivity analyses and stochastic evaluations can prioritize future work for policy makers and scholars under uncertainty. This study contributes to the literature in several ways. To the best of our knowledge, multi-country partial equilibrium models are rarely used in the analysis of the economic consequences of invasive species, whereas global sensitivity analyses have not been used yet in the context of partial equilibrium modeling.

---

<sup>1</sup> Additional production as a result of higher prices might stem from more input usage on existing plots or from the establishment of new orchards. While high tree density systems reach their full-bearing state in three to four years, olives are often cultivated on marginal land due to their robustness. Traditional orchards may take up to a decade to reach their full-bearing state [215]. Environmental and social factors will constrain the establishment of new plots. For example, the availability of ground water for irrigation and learning costs will influence the decision of producers to establish olives as a modern high density system or as a traditional orchard.

Furthermore, this paper adds to the literature by being the first to generate insights on consumer impacts for the case of Xfp in Europe.

## 4.2 Materials and Methods

This study employs a suite of models. First, a prediction of climatic suitability is generated to delineate the national supply which could potentially be affected by the pathogen and the areas through which dispersal might occur. Second, simulations of future spread provide scenario-based estimates of the olive oil supply affected at each time step depending on the spread rate. The disease spread simulations are based on an assumption of radial invasion of the climatically suitable habitat while taking the spatial distribution of olive orchards into account. We account for the uncertainty in the annual rate of dispersal by using three quantiles of an expert-elicited distribution of spread rates [157]. The corresponding spread rates are 1.10 (RR05), 5.18 (RR50) and 12.35 (RR95) km per year. The economic model assumes that the percentage of orchards infected in a country equals the percentage of the country's olive oil supply affected. The simulation results project the expected supply reductions from further dispersal of the pathogen over 50 years in Italy and new introductions into Greece or Spain. Lastly, a multi-country partial equilibrium model on the olive oil market is computed to shed light on the potential economic impact of Xfp to European consumers. The model explicitly projects welfare changes for Italy, Greece, Spain and the Rest of Europe (RoEU). The economic model computes changes in welfare to affected processors, non-affected processors and consumers compared to the baseline in which Xfp is absent. Impact is computed over a 50-year time horizon and presented as present values.

### 4.2.1 Climatic Suitability Map

The main purpose of the climatic suitability map is to predict the area of production within each country that could potentially be affected

by the disease. Furthermore, dispersal paths within the disease spread simulations were limited to climatically suitable habitat. Consequently, the realized speed of the invasion depended on the continuity of the predicted climatically suitable area within a region. In principle, disease introductions into climatically suitable regions could remain contained by climatic barriers if these were to surround the area of introduction. Hence, a prediction of the climatically suitable area is crucial to delineate the potential maximum extent of the invasion as well as to obtain a more realistic spread progression by taking geographic information on climatic suitability into account within the disease spread simulation.

The climatic suitability map is described in more detail within Schneider et al. [222]. Disease occurrences were filtered by keeping only confirmed positives with precise coordinates under natural inoculum pressure (i.e. records from greenhouses, screenhouses and interceptions were omitted). To reduce spatial autocorrelation, the remaining presence points were thinned by enforcing a minimum distance of 5 km between points which is equal to the spatial resolution of the climate data. Due to the randomness associated with the thinning, the procedure was repeated four times. Ten species distribution models (bioclim, boosted and regression trees, classification and regression trees, domain, generalized additive models, multivariate adaptive regression splines, maximum entropy, random forest, recursive partitioning and regression trees and support vector machines) were used to explore the relationship between occurrences of *Xf* and environmental variables. Subsequently, an ensemble prediction was generated taking the relative model performances into account. The ensemble prediction provided a continuous score ranging from zero to one for locations in Europe. These scores were bilinearly interpolate from a 5 km to a 1 km resolution to meet the needs of the disease spread model. Lastly, the downscaled map was converted to a binary map, i.e. indicating whether a given location is suitable or not, using three different thresholds. Threshold 1 (0.165) is particularly informative for models based on presence-only data and ensures that a correct prediction on species presence of at least 90 percent is made. Threshold 2 (0.132) was used to maximize the sum of the accuracy of predicting occupied sites to be suitable and unoccupied sites to be unsuitable (i.e. sum of sensitivity and specificity) and Threshold 3 (0.093) was used to minimize the

difference between the accuracy of predicting occupied sites to be suitable and unoccupied sites to be unsuitable (i.e. minimum difference between sensitivity and specificity) [196].

### 4.2.2 Disease Spread Simulation

The purpose of the disease spread simulation is to provide the dynamic change of the area of production affected by the pathogen over the time horizon of 50 years. This is particularly important in dynamic economic assessments, because future impacts must be discounted [178]. As there are still significant uncertainties on crucial aspects of the dispersal process of Xfp such as long-distance jumps [171], we simulated various spread scenarios.

The disease spread simulation is described in more detail within Schneider et al. [222]. Spread was modelled as a radial range expansion process [197]. Dispersal characteristics are collapsed into a single parameter described as the rate of radial range expansion ( $rr$ ). To circumvent the issue of simulating an expanding circle using grid cells, the radial dispersal is simulated using a cellular automaton model with alternating step-types, namely rook and queen's case contiguity, at a ratio of  $2 - \sqrt{2}$  and  $2 - \sqrt{1}$ , respectively [205]. The annual time-steps are broken down into within-year-time-steps such that the elicited rate is obtained. Step-types are randomly assigned to every within-year-step such that the aforementioned ratios are obtained. As a result, a spread pattern is generated which resembles a regular octagon. As spread is slightly over-predicted into corner-directions of the octagon, the radial range model is used to tightly contain the newly infected cells within the elicited radius by removing over-predicted cells.

Invasion into a cell is only accepted if the cell is climatically suitable, however, does not depend on whether or not olives are present. While this ensures that climatically unsuitable territories are not travelled through, the model implicitly assumes that alternative hosts assist dispersal of Xfp. We justify this assumption by the facts that the host range of the pathogen spans several hundred plant species and any xylem feeding insect is currently considered to be a potential vector [15, 157].

To acknowledge the uncertainty on future disease spread, nine scenarios were analyzed for Italy resulting from three climatic suitability thresholds and three spread rates. While for Italy spread was simulated beyond the currently infected zone, the uncertainty on the point of introduction in Greece and Spain was evaluated by stochastic simulations with points of introduction randomly selected from the olive cells within the climatically suitable territory. For all combinations of climatic suitability threshold and spread rate, 1000 points of introduction were generated for both Greece and Spain. Consequently, 9000 instances of future epidemics were generated for each of the countries.

### 4.2.3 Economic Model

Let  $t$  denote discrete annual points in time and the time horizon within which the impacts are assessed as  $T$ . Italy, Greece, Spain and the Rest of Europe (RoEU) were included as interdependent markets ( $C$ ) into the analysis. The markets within Europe are denoted by the subscript  $c$ . The following set of equations are solved on a European level. Thereby, we implicitly model intra-European trade and assume that supply can freely flow within Europe to meet demand. The European market interacts with the Rest of the World (ROW) through trade of excess supply or demand.<sup>2</sup>

To model the change in supply due to reduced yields and changes in operational costs, the spread simulations were used to compute the newly infected share of the national areas in the four included countries ( $z_{ct,ax}$ ). Different ages of infection are denoted with  $ax$ . We assume a one to one translation of the percentage of orchards infected to the percent of national supply of olive oil affected. We discuss this limitation in more detail below. We denote  $D_t$  and  $S_t$  as the total demand and total supply in the European Union at time  $t$ , respectively.  $SA_t$  and  $SN_t$  denote the supply by the affected and non-affected producers, respectively.  $P_{tc}$  is the equilibrium price in market  $c$  at time  $t$ . The market-specific equilibrium prices are coupled

---

<sup>2</sup> Europe is a net-exporter of olive oil and remained so in the disease spread scenarios.

to the world market price  $WP_t$ , and only differ by a constant country-specific wedge,  $\mu_c$ . The wedges capture, for example, differences in transportation costs among countries and are assumed to be constant over time.  $\eta_c$  and  $\theta_c$  denote the elasticities of demand and supply in the different European markets.  $\chi_c$  and  $\rho_c$  are parameters. The parameter  $h$  denotes the horizontal percentage shift in the supply curve due to a reduction in yield (i.e.  $h$  equals -0.10 for an annual yield decrease of 10 percent).  $g$  denotes a simultaneous vertical percentage shift in the supply curve due to the increased production costs (i.e.  $g$  equals -0.10 for an annual cost increase of 10 percent). Parameters  $h$ ,  $g$ , and  $z_{ct,ax}$  represent the supply shifter  $N_{ct}$  within the equation for the producer surplus ( $iPS_t$ ) below. The European market can be depicted as follows:

$$D_t = \sum_{c \in C} \chi_c \cdot P_{tc}^{-\eta_c} \quad (4.1)$$

$$S_t = SA_t + SN_t \quad (4.2)$$

$$SA_t = \sum_{c \in C} \sum_{ax=1}^T (1+h)^{ax} \cdot \rho_c \cdot ((1+g)^{ax} \cdot P_{tc})^{\theta_c} \cdot z_{ct,ax} \quad (4.3)$$

$$SN_t = \sum_{c \in C} \rho_c \cdot (P_{tc})^{\theta_c} \cdot \left(1 - \sum_{ax=1}^T z_{ct,ax}\right) \quad (4.4)$$

$$WP_t = P_{tc} - \mu_c \quad (4.5)$$

We let  $E_t$  and  $X_t$  denote the excess supply and demand of the European Union, respectively.  $\omega$  denotes the elasticity of excess demand (negative) or supply (positive) in the rest of the world.  $\nu$  is a parameter calibrated on market data. The European Union interacts with the rest of the world via excess domestic supply or demand as follows:

$$E_t = S_t - D_t \quad (4.6)$$

$$X_t = \nu \cdot (WP_t)^\omega \quad (4.7)$$

$$X_t = E_t \quad (4.8)$$

The economic impact is computed as the sum of total discounted future welfare losses ( $iTW$ ), due to the spread of the disease and associated changes in the equilibrium price and quantity [223]. Impact on total welfare comprises the sum of impacts on total producer surplus ( $iPS_t$ ) and total consumer surplus ( $iCS_t$ ) which are expressed in monetary units. The total producer and consumer surpluses on the European level comprise the producer and consumer surpluses in the included countries. The producer surplus, consumer surplus and as a result the overall impact are discounted using the discount rate  $r$ . We let  $S_t^0(\cdot)$  denote the supply function in absence of the infection, while  $S_t^1(\cdot)$  is the supply function after introduction of the pathogen.  $a$  and  $b$  denote the intersection of the supply curves with the y-axis.  $N_{ct}$  denotes the supply shifter for the different countries. The supply and demand function adhere to the following properties [223]:  $\frac{\delta S_t}{\delta P_t} > 0$ ,  $\frac{\delta S_t}{\delta N_t} < 0$  and  $\frac{\delta D_t}{\delta P_t} < 0$

$$iPS_t = \sum_{c \in C} \left( \left( \int_a^{P_{tc}} S_t^0(P_{tc}) \delta P_{tc} - \int_b^{P_{t+1,c}} S_t^1(P_{tc}, N_{ct}) \delta P_{tc} \right) \cdot (1+r)^{-t} \right) \quad (4.9)$$

$$iCS_t = \sum_{c \in C} \left( \left( \int_{P_{tc}}^{P_{t+1,c}} D_t(P_{tc}) \delta P_{tc} \right) \cdot (1+r)^{-t} \right) \quad (4.10)$$

$$iTW = \sum_{t=1}^T (iPS_t + iCS_t) \quad (4.11)$$

#### 4.2.3.1 Estimation of the supply elasticity

The absence of published information on supply elasticities was approached by estimating a European supply elasticity econometrically. We used panel data for the years 1990 to 2016 (see next section), to estimate a partial adjustment model (equation 4.12) which is also known as the Nerlove Supply Response Model [224]. Using the Cobb-Douglas functional form, we included a time trend ( $Y$ ), country-fixed effects

( $C$ ), first lags of prices ( $P_{t-1}$ ) and first lags of production ( $Q_{t-1}$ ) and regressed on current production ( $Q_t$ ).  $\alpha$ ,  $\tau$ ,  $\beta$  and  $\mu$  are parameters and  $\kappa$  a vector of parameters.  $\lambda$  captures the inertia and is generally expected to be  $0 < \lambda < 1$ . The model was estimated as equation 4.13 in which  $\phi$  denote the set of parameters and  $\nu_t$  the error term ( $R^2$  of 0.921). The formulation allows to derive short-run and long-run supply elasticities. The short run elasticity is obtained through the coefficient estimate for first lags of prices ( $\phi_3$ ). The long run elasticity ( $\beta$ ) is obtained by computing the expectation of  $\lambda$  and dividing the coefficient for the first lag prices by this expectation (equation 4.14 and 4.15). In the deterministic calculations, we used the estimate for the long run supply elasticity (0.3241). In the global sensitivity analysis, this estimate was varied by plus and minus 50 percent. The resulting interval comprises the estimated short run supply elasticity (0.2513). Table 4.5 in the Supplementary Material depicts the fitted model.

$$Q_t = \lambda(\alpha + \tau Y + \kappa C + \beta P_{t-1} + \epsilon_t) + (1 - \lambda)\mu Q_{t-1} \quad (4.12)$$

$$Q_t = \phi_0 + \phi_1 Y + \phi_2 C + \phi_3 P_{t-1} + \phi_4 Q_{t-1} + \nu_t \quad (4.13)$$

$$\hat{\lambda} = 1 - \phi_4 \quad (4.14)$$

$$\hat{\beta} = \phi_3 / \hat{\lambda} \quad (4.15)$$

Notably, our approach is not able to capture structural changes of actors' behavior over the 50-year time horizon. The sizable market shock within some of the spread scenarios and our long time horizon of 50 years could fundamentally alter actors' ability to respond to price signals. Our reliance on historical data to estimate the supply elasticity and our assumption that this estimate is constant over the 50-year time horizon are critical limitations of our analysis. We address these limitations through sensitivity analyses (see section 4.2.5), and through further discussion below.

#### 4.2.4 Data

For the climatic suitability modeling, presence data was derived from the Update of the Xf host database [15], data from the national plant

protection organizations and records of outbreaks in Porto, Tuscany and Hula Valley [190–192]. Weighted pseudo-absence data was simulated at a prevalence of 0.1. Climate data ranging from 1979 to 2013 was obtained from Chelsa Climatology at a 5 km spatial resolution [193].

For the disease spread simulations, data on olive production sites was obtained from the Coordination of Information on the Environment (CORINE) land cover map<sup>3</sup> and aggregated to a 1 km resolution to reduce the computational time. For the radial range expansion ( $rr$ ), we used the 5, 50 and 95 percent quantiles of a distribution elicited from experts [157]. The elicitation followed a structured methodology as described in [202] and [203]. The seven invited experts are internationally recognized experts on the disease and on relevant agricultural practices. The parameter was defined as the mean distance, in kilometers, which will comprise 90% of the area containing the newly infected plants around an infected area within one year [157]. The elicited quantiles correspond to a radial range expansion of 1.10, 5.18 and 12.35 km per year.<sup>4</sup>

To calibrate the economic model, various data were gathered (Table 4.1). The total production and consumption for the European Union, Italy, Greece and Spain was obtained from the International Olive Oil Council [225, 226]. Five-year averages, for 2012 to 2016, were constructed and used as estimates of the equilibrium quantities. Corresponding standard deviations were computed for use in the global sensitivity analysis. Country-specific prices for olive oil were obtained from Eurostat. Five-year averages for the same time-period were used as estimates of the equilibrium prices. For the rest of the European Union (RoEU), the five-year average price was constructed as a, by production-share, weighted price of Croatia, Portugal and Slovenia as these countries represent the other European producers besides Italy, Greece and Spain. Again, standard deviations were computed for use in the global sensitivity analysis. The world price was computed as the, by production-share, weighted average European price. This is justified by the fact that the European market is the

---

<sup>3</sup> <https://land.copernicus.eu/pan-european/corine-land-cover>

<sup>4</sup> The climatic suitability map and spread simulations can be accessed via Zenodo (<https://doi.org/10.5281/zenodo.3672794>). Visualizations of both maps are provided in the Supplementary Material of the previous chapter.

**Table 4.1** Overview of the economic parameter values.

Parameter	Symbol	Dimension	Deterministic	Global Sensitivity Analysis		
				Distribution	Mean/Min.	SD/Max.
Demand Italy	$D_t$	1,000 tons	559.96	truncated normal	559.96	67.72
Demand Greece	$D_t$	1,000 tons	139.00	truncated normal	139.00	24.17
Demand Spain	$D_t$	1,000 tons	488.26	truncated normal	488.26	26.26
Demand RoEU	$D_t$	1,000 tons	416.70	truncated normal	416.70	8.69
Supply Italy	$S_t$	1,000 tons	351.62	truncated normal	351.62	124.29
Supply Greece	$S_t$	1,000 tons	260.98	truncated normal	260.98	84.13
Supply Spain	$S_t$	1,000 tons	1187.16	truncated normal	1187.16	413.13
Supply RoEU	$S_t$	1,000 tons	91.28	truncated normal	91.28	21.82
Price Italy	$P_{r,c}$	Euro per ton	4067.22	truncated normal	4067.22	685.98
Price Greece	$P_{r,c}$	Euro per ton	2855.42	truncated normal	2855.42	364.92
Price Spain	$P_{r,c}$	Euro per ton	2611.02	truncated normal	2611.02	498.28
Price RoEU	$P_{r,c}$	Euro per ton	4242.16	truncated normal	4242.16	522.46
Demand elasticity Italy	$\eta_c$	Dimensionless	-0.842	uniform	-1.263	-0.421
Demand elasticity Greece	$\eta_c$	Dimensionless	-0.760	uniform	-1.140	-0.380
Demand elasticity Spain	$\eta_c$	Dimensionless	-0.485	uniform	-0.728	-0.243
Demand elasticity RoEU	$\eta_c$	Dimensionless	-0.350	uniform	-0.525	-0.175
Demand elasticity RoW	$\omega$	Dimensionless	-0.350	uniform	-0.525	-0.175
Supply elasticity Italy	$\theta_c$	Dimensionless	0.3241	uniform	0.1621	0.4862
Supply elasticity Greece	$\theta_c$	Dimensionless	0.3241	uniform	0.1621	0.4862
Supply elasticity Spain	$\theta_c$	Dimensionless	0.3241	uniform	0.1621	0.4862
Supply elasticity RoEU	$\theta_c$	Dimensionless	0.3241	uniform	0.1621	0.4862
Yield decline	$h$	percent	-10	uniform	-100	-5
Cost change	$g$	percent	-10	uniform	-100	-5
Discount rate	$r$	percent	3	uniform	3	7

$D_t$  and  $S_t$  denote European demand and supply at time  $t$  and are obtained by aggregating over the four markets.

world leader in olive oil production and consumption [170]. Hence, the world market price can be expected to be determined by the European market. Country specific price wedges were calibrated by subtracting the country prices from the world price. Information on the demand elasticities was gathered from the scientific literature [227–229]. For the global sensitivity analysis, these estimates were varied by plus and minus 50 percent. To estimate the supply elasticity, the same data from the International Olive Oil Council [225] and Eurostat were used but for the years 1990 to 2016 (see section 4.2.3.1). Notably, highly incomplete timeseries for Croatia and Slovenia forced us to rely on data for Portugal as proxy for the RoEU.

#### 4.2.5 Sensitivity Analyses

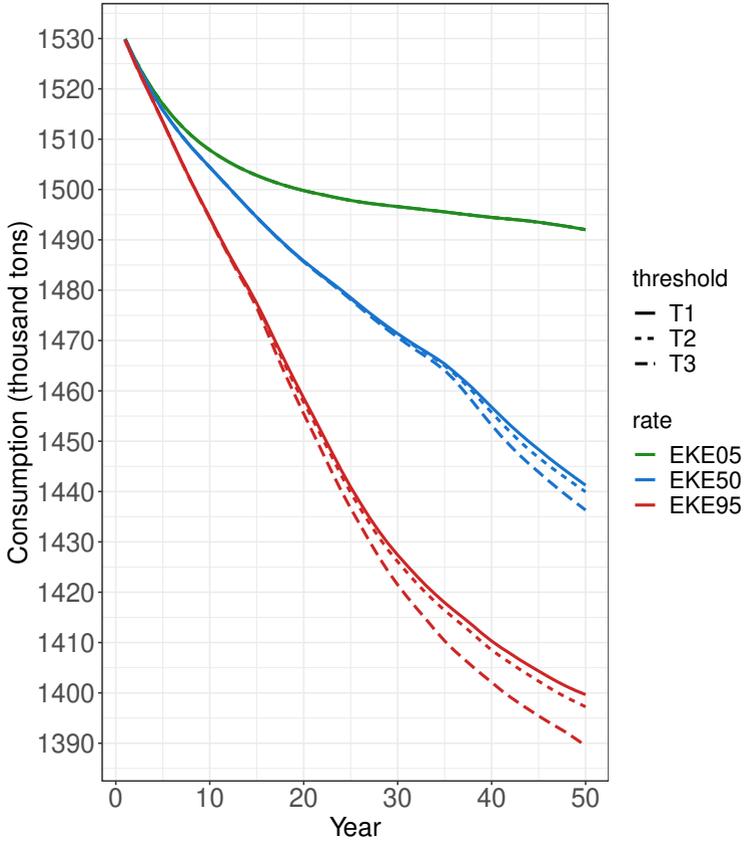
The sensitivity of the results with regard to the parameter values used was explored using a global sensitivity analysis through variance decomposition [209]. To calculate the Sobol sensitivity indices [230], a computational experiment was set up in which input parameters

were sampled out of defined distributions (Table 4.1). To sample the input parameter space, 10000 draws were generated from each input distribution. The computational time was improved by applying the Saltelli [208] sampler which generated an input matrix of length  $N(k + 2)$  where  $N$  is the number of draws (10,000) and  $k$  is the number of model inputs. The implementation was achieved via the *sensitivity* package for R [231], using the improved formulas of Jansen [211] and Saltelli et al. [209]. In total, 24 parameters were included which resulted in 260000 rows of input values for which impact was computed for one spread simulation (Italy following RR50 and threshold T2).

In addition to the global sensitivity analysis, the influence of parameter uncertainty on welfare changes was analyzed in more detail for the supply elasticities ( $\theta_c$ ), the yield decline ( $h$ ) and the cost change ( $g$ ). The reason for this selection is our reliance on historical data for the estimation of the supply elasticity, and the absence of scientific information on the other two parameter. To analyze the influence of uncertainty in these parameters, 10000 draws from their distributions (Table 4.1) were generated. Subsequently, while fixing all other parameter at their deterministic values welfare changes were computed using all draws for  $\theta_c$ ,  $h$  or  $g$ .

## 4.3 Results

The climatic suitability modelling shows that the vast majority of olive orchards are within the climatically suitable territory for establishment and spread of Xfp [222]. Thereby, Xfp is threatening nearly three quarters of the world's production of olive oil. The different spread scenarios resulted in varying price responses across Europe (Table 4.1). Following spread beyond the current extent in Italy, for RR05 the prices for olive oil increased by 5.2 percent in Greece, 5.7 percent in Spain, 3.6 percent in Italy and 3.4 percent in the RoEU. For RR50 prices increased up to 13.6, 14.9, 9.3 and 8.9 percent in Greece, Spain, Italy and the RoEU, respectively. For RR95 prices increased up to 21.3, 23.5, 14.6 and 14.0 percent in Greece, Spain, Italy and the RoEU, respectively. As a consequence of higher prices, the consumption in



**Fig. 4.1** Change in the consumption of olive oil for spread of Xfp in Italy beyond the current extent.

Europe declined between 2.5 percent for RR05 and 9.2 percent for RR95 (Figure 4.1).

Following spread in Greece, on average over the 1000 points of introduction for RR05 prices increased up to 1.2 percent in Greece, 1.3 percent in Spain, 0.8 percent in Italy and 0.8 percent in the RoEU. For RR50, prices increased up to 4.8, 5.3, 3.3 and 3.1 percent in Greece, Spain, Italy and the RoEU, respectively. For RR95, prices increased up to 6.8, 7.5, 4.6 and 4.4 percent in Greece, Spain, Italy and the RoEU,

**Table 4.2** Change in prices and consumption over 50 years.

Spread Scenario Suitability Spread Threshold Rate	Spread in Italy			Spread in Greece*			Spread in Spain*								
	Price Increase (%) ES	IT	RoEU	Price Increase (%) ES	IT	RoEU	Price Increase (%) ES	IT	RoEU						
	EL	EU (%)	Consumption EU (%)	EL	EU (%)	Consumption EU (%)	EL	EU (%)	Consumption EU (%)						
T1 RR05	5.2	5.7	3.6	3.4	-2.5	1.2	1.3	0.8	0.8	-0.6	7.8	8.6	5.3	5.1	-3.6
T2 RR05	5.2	5.7	3.6	3.4	-2.5	1.2	1.3	0.8	0.8	-0.6	8.0	8.8	5.5	5.2	-3.6
T3 RR05	5.2	5.7	3.6	3.4	-2.5	1.2	1.3	0.8	0.8	-0.6	8.4	9.3	5.7	5.5	-3.8
T1 RR50	12.8	14.1	8.7	8.4	-5.8	4.8	5.3	3.3	3.1	-2.3	77.0	85.0	52.6	50.3	-23.7
T2 RR50	13.0	14.3	8.9	8.5	-5.9	4.8	5.3	3.3	3.1	-2.3	86.8	95.8	59.3	56.7	-25.3
T3 RR50	13.6	14.9	9.3	8.9	-6.1	4.7	5.2	3.2	3.1	-2.2	93.1	102.7	63.6	60.8	-26.5
T1 RR95	19.6	21.6	13.4	12.8	-8.5	6.3	7.0	4.3	4.1	-3.0	120.9	133.3	82.6	78.9	-32.1
T2 RR95	20.0	22.0	13.7	13.1	-8.7	6.9	7.6	4.7	4.5	-3.2	145.5	160.5	99.4	95.0	-35.2
T3 RR95	21.3	23.5	14.6	14.0	-9.2	6.8	7.5	4.6	4.4	-3.2	187.6	207.0	128.1	122.5	-40.8

\*averaged over all random points of introduction. EL=Greece, ES=Spain, IT=Italy, RoEU=Rest of Europe.

respectively. The decline in the total European consumption ranged between 0.6 percent for RR05 to 3.2 percent for RR95.

Following spread in Spain, on average over the 1000 points of introduction for RR05 prices increased up to 8.4 percent in Greece, 9.3 percent in Spain, 5.7 percent in Italy and 5.5 percent in the RoEU. For RR50, prices increased up to 93.1, 102.7, 63.6 and 60.8 percent in Greece, Spain, Italy and the RoEU, respectively. For RR95, prices increased up to 187.6, 207.0, 128.1 and 122.5 percent in Greece, Spain, Italy and the RoEU, respectively. The decline in the total European consumption ranged between 3.6 percent for RR05 to 40.8 percent for RR95.

Table 4.3 depicts the welfare changes over 50 years expressed as present value based on a discounting rate of three percent per year. We advise to read the table as follows: The three horizontal blocks correspond to Xfp spread simulations in Italy, Greece and Spain. The rows within each block depict different scenarios for the climatic suitability threshold and the spread rate. Readers might start by looking at the column on the very right hand side which depicts the total impact to welfare across Europe obtained by summing the total welfare changes for each country. As changes in total welfare comprises changes in producer surplus and consumer surplus, readers might want to continue by looking at these columns. Total change to consumer and producer surplus is discussed below and was computed by aggregating over countries. Surplus generated by affected producers pertains to affected producers in the country in which spread was simulated (i.e. Italy, or Greece, or Spain). Results for country specific changes to welfare are depicted in Tables 4.6 to 4.8 in the Supplementary Material.

If the pathogen spreads beyond the current extent in Italy, the decline in consumer welfare due to higher prices for olive oil ranged between 4.1 billion for RR05 to 10.3 billion Euro for RR95 across Europe. While total producer surplus increased between 0.2 to 0.4 billion Euro across Europe, producer surplus in Italy declined between 3.8 and 9.9 billion Euro.

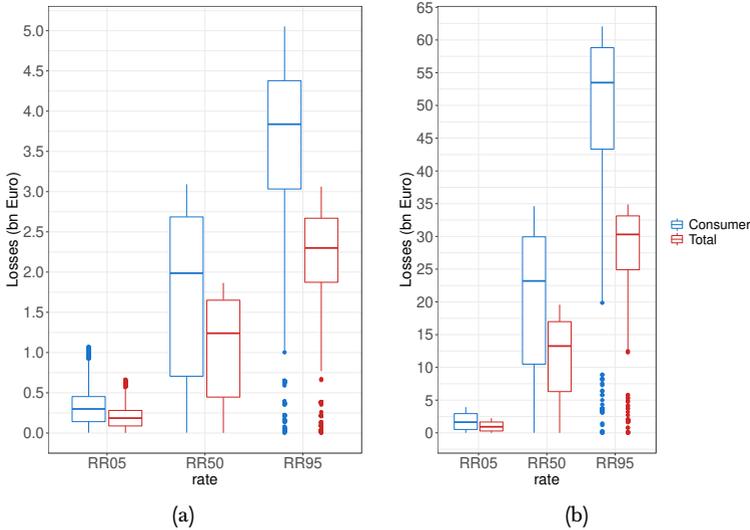
Results from introductions of Xfp into Greece and Spain are presented as averages across the 1000 randomized points of introduction. The distributions of welfare changes over all points of introduction for one climatic suitability threshold (T2) are depicted in Figure 4.2a

**Table 4.3** Changes in welfare over 50 years in billion Euro for all spread scenarios.

<b>Spread Scenario</b>		<b>Consumer</b>	<b>Producer</b>	<b>Non-Affected Producer</b>	<b>Affected Producer</b>	<b>Total</b>
<b>Suitability</b>	<b>Spread</b>					
<b>Threshold Rate</b>						
<i>Spread in Italy</i>						
T1	RR05	-4.09	0.18	-0.50	0.68	-3.91
T2	RR05	-4.09	0.18	-0.50	0.68	-3.91
T3	RR05	-4.09	0.18	-0.50	0.68	-3.91
T1	RR50	-6.27	0.19	-1.02	1.21	-6.08
T2	RR50	-6.32	0.19	-1.03	1.22	-6.12
T3	RR50	-6.40	0.18	-1.07	1.25	-6.20
T1	RR95	-9.80	0.36	-1.46	1.82	-9.45
T2	RR95	-9.91	0.36	-1.49	1.85	-9.56
T3	RR95	-10.29	0.38	-1.56	1.94	-9.91
<i>Spread in Greece*</i>						
T1	RR05	-0.37	0.13	0.07	0.06	-0.24
T2	RR05	-0.36	0.13	0.07	0.06	-0.22
T3	RR05	-0.37	0.13	0.07	0.06	-0.22
T1	RR50	-1.82	0.70	0.43	0.27	-1.12
T2	RR50	-1.79	0.69	0.42	0.26	-1.10
T3	RR50	-1.79	0.69	0.43	0.26	-1.09
T1	RR95	-3.23	1.28	0.87	0.41	-1.96
T2	RR95	-3.31	1.30	0.86	0.43	-2.02
T3	RR95	-3.27	1.29	0.86	0.43	-1.98
<i>Spread in Spain*</i>						
T1	RR05	-1.76	0.77	0.46	0.31	-0.99
T2	RR05	-1.76	0.77	0.45	0.32	-0.99
T3	RR05	-1.81	0.78	0.45	0.33	-1.03
T1	RR50	-18.64	8.02	5.16	2.86	-10.63
T2	RR50	-20.31	8.68	5.55	3.13	-11.63
T3	RR50	-21.33	9.08	5.77	3.31	-12.25
T1	RR95	-42.64	18.62	13.55	5.07	-24.02
T2	RR95	-47.58	20.56	14.83	5.73	-27.03
T3	RR95	-53.00	22.57	15.84	6.72	-30.43

EL=Greece. ES=Spain. IT=Italy. RoEU=Rest of Europe.

\*averaged over all random points of introduction.



**Fig. 4.2** Distribution of changes in total welfare and consumer surplus following introductions of *Xfp* into Greece (a) or Spain (b) and spread over 50 years with a suitability threshold of 0.135.

and 4.2b. Following spread in Greece, consumer welfare across Europe declined between 0.4 billion to 3.3 billion Euro. Total producer surplus increased between 0.1 to 1.3 billion Euro across Europe and declined in Greece between 0.2 and 2.2 billion Euro. Following spread in Spain, consumer welfare across Europe declined between 1.8 billion and 53 billion Euro. Total producer surplus in Europe increased between 0.8 to 22.6 billion Euro, and producer surplus in Spain declined between 0.1 to 13.3 billion Euro.

The results of the sensitivity analysis show that out of 24 parameters only four parameters had statistically significant first order indices at the critical 5 percent level (Table 4.4). Uncertainty regarding supply from Italy and Spain, price of olive oil in Italy and the discount rate had statistically significant first order indices. For example, the variation in the Italian supply caused 33.5 percent in the variance of impact to total welfare through direct effects (21.3 percent) and higher order

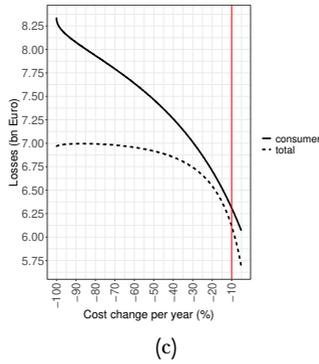
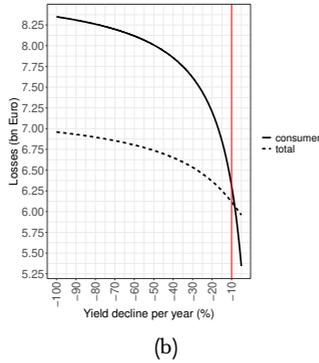
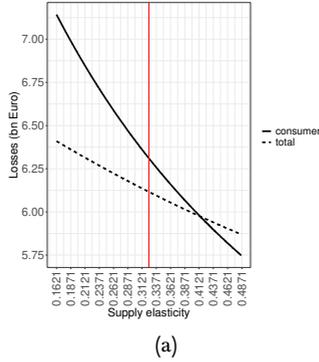
**Table 4.4** First and total order sensitivity indices of statistically significant economic parameters ( $P < 0.05$ ).

Parameter	First Order			Total Order		
	Index	Min.	Max.	Index	Min.	Max.
Supply Italy	0.213	0.175	0.258	0.335	0.316	0.354
Supply Spain	0.219	0.180	0.260	0.345	0.327	0.363
Price Italy	0.108	0.069	0.152	0.186	0.176	0.195
Discount Rate	0.113	0.072	0.160	0.208	0.199	0.217

interactions with other parameter. Sensitivity indices for all economic parameter are shown in Table 4.9 in the Supplementary Material.

For the deterministic economic analyses (Table 4.3), we used an econometrically estimated long run supply elasticity and very conservative values for the yield decline and cost change under Xfp pressure. The absence of scientific information for the yield decline and cost change under infection as well as our reliance on historical data to estimate suppliers' adaptability to price changes was approached through stochastic evaluations for one spread scenario (Figure 4.3a, 4.3b & 4.3c). The deterministic results for spread in Italy, following RR50 with a suitability threshold T2, were a decline of 6.3 billion Euro in consumer welfare and 6.1 billion Euro in total welfare. The evaluated range of the supply elasticity resulted in impacts to consumers ranging 5.8 to 7.1 billion Euro and total impacts ranging 5.9 to 6.4 billion Euro. The evaluated range of the yield decline under Xfp resulted in impacts to consumers ranging 5.3 to 8.4 billion Euro and total impacts ranging 6.0 to 7.0 billion Euro. The evaluated range of the cost change under Xfp resulted in impacts to consumers ranging 6.1 to 8.3 billion Euro and total impacts ranging 5.7 to 7.0 billion Euro. In comparison to the uncertainty on the spread rate or likely points of introductions in Greece and Spain, these intervals for impact are narrow (cf. Table 4.3 and Figure 4.2a & 4.2b).

As evident from Figure 4.3a, more inelastic supply elasticities widened the gap between consumer impact and total impact. In other words, markets characterized by less adaptable producers resulted in higher prices following supply shocks due to Xfp which left consumers worse off as a result. Comparable insights were obtained for the yield decline and the cost change under Xfp (Figure 4.3b and 4.3c). Both parameters determine the annual reduction in supply of affected



**Fig. 4.3** Stochastic evaluation of welfare changes under uncertainty of supply elasticity (a), yield decline (b) and cost change (c). The vertical line depicts the deterministic value used.

producers. Higher values resulted in more drastic reductions and in turn higher prices for olive oil which left consumers significantly worse off compared to the deterministic results. Consequently, while we believe that our derived conclusions are robust to our assumptions on these parameters the deterministic results are likely conservative estimates.

## 4.4 Discussion

Our analysis shows that consumers can be expected to bear the majority of the economic impacts from further dispersal or new introductions of Xfp in Europe. While producer surplus in affected countries might decline, increasing prices for olive oil across Europe would generate additional profits to processors at the expense of consumers. Consequently, control measures to prevent future spread of the disease can be expected to more strongly benefit consumers rather than processors. In turn, regulatory measures such as the preventive felling of trees should find public acceptance with regard to economic considerations. Unfortunately, consumers might not be aware that they are beneficiaries of the control of invasive plant diseases such as Xfp. This stresses the need for improving the communication of pest risk assessments to the public.

The results highlight that introductions of Xfp into Spain can be expected to significantly disrupt the European olive oil market. Spain is the world's largest producer of olive oil. Consequently, reductions in supply from Spain sensitively affect world prices. Public actions to foster vigilance among growers and citizens to prevent an introduction and further spread of Xfp in other olive growing regions of Europe are crucial. Further dissemination of information on Xfp to citizens in olive growing hot-spots could increase awareness and thereby improve the likelihood of spotting possible introductions promptly. Unfortunately, in-field detection is aggravated due to the absence of precise information on the asymptomatic period following infection [22]. Therefore, further research on the epidemiology and early detection methods is important [232, 233].

Preferences for invasive species control vary among citizens [234, 235], and economic concerns could very well not be the first priority for some. Unfortunately, Xfp is currently not curable under field conditions and sensitive cultivars can be expected to die off quickly following infection [157]. Consequently, the establishment of a cordon sanitaire to prevent spread of the disease through felling of healthy trees is likely to spare many more trees than are felled. An intelligible communication of this information to stakeholders maybe could have prevented the societal unrest in the affected region.

The results of our analysis demonstrate that affected producers and consumers jointly carry the economic risk associated with Xfp. To mitigate the risk, research currently targets the identification of resistant traits in olive varieties [160–166], biological control of the vectors [168], early detection of the disease [232], and the establishment of a cordon sanitaire by removing host plants near the infected zone. Evidently, all these measures aim at growers taking action to prevent further impacts. Unfortunately, most strategies have in common that they pose a significant financial burden on growers in the affected area, and with profit margins for olives historically being small already [185], this could result in a significant portion of farmers having to cease production entirely. The prevention of further spread can be seen as a public good and, therefore, compensation payments are justified [49]. Policy makers must ensure that compensation schemes are intelligible and appropriately budgeted to minimize adverse behavior of not reporting possible infections [236–238]. Further research could investigate to which extent market instruments might assist growers' risk mitigation. For example, farmers nearby the infected zone might not yet qualify for official compensation schemes, however, they could already benefit from participation in insurance programs which aim at alleviating possible financial burden associated with replanting orchards with a resistant variety in case Xfp continues to spread.

The case on the Olive Quick Decline Syndrome exemplifies the drastic impact to consumers which can arise from biological invasions solely through price responses following reductions in supply. Fortunately, other aspects such as food safety and food security are not affected here. However, the losses in cultural heritage and landscape value from the desiccation of century old olive trees are additional factors which need to be considered in the design of policy. Given

the existing level of globalization, the risk of biological invasions is likely to prevail in the future [5]. For inelastic production systems in particular, consumers can be expected to bear a sizeable share of the economic consequences. Therefore, regulators should ensure that phytosanitary authorities can react promptly to emerging threats. To improve response, information exchange between public bodies, science and farmers must be fostered. In addition to continued support of policy makers for research on adaptation strategies to Xfp in Europe, it is essential that both public bodies and scientists play a role in involving citizens through intelligible communication of new insights.

While our analysis aims at contributing to the ongoing public debate on Xfp through an exploration of potential consequences to consumers, uncertainty on various aspects on the side of the pathogen as well as absence of more granular economic data resulted in some limitation. In what follows, we discuss these limitations and possible consequences of our assumptions.

We assumed that the percentage of orchards infected in a country equals the percentage of the country's olive oil supply affected. This simplification is required because the used land-cover data only provides a binary indication of whether olives are present in a given grid cell. To obtain a grid cell specific contribution to the national production, information on the olive density within a grid, or ideally data on the grid cell specific production, would be needed. While the inclusion of the binary land-cover data still added significantly to the disease spread simulation, it might well be that our simulation of the olive oil supply affected fails to account for granular differences in the areas' contribution to the national supply.

The disease spread model employed, while being relatively straightforward, predicted past population expansions well [201], and the model has been proposed as a suitable tool for pest risk assessments [197]. Nevertheless, strengthening the integration among all employed models would be desirable. Information on the relation between climatic suitability and the spread rate of Xfp might improve the disease spread simulations by allowing for varying spread rates depending on local conditions. Furthermore, insights into the efficacy of management strategies, such as vector control, on the spread rate would allow to better link economic models with the disease spread simulations for Xfp and thereby enable analyses of different strategies.

The olive oil market is known to be a segmented market with different quality grades. However, due to the absence of data on production volumes of different olive oil quality grades and cross price elasticities between qualities, the model developed for this analysis assumes one composite olive oil market and thereby is not able to demonstrate impacts for producers of high versus low quality olive oils. As a response to price increases, consumers may substitute high quality olive oil with a lower quality alternative which could worsen the economic impacts for high quality olive oil producers in Italy, Greece and Spain.

The absence of information on the countries' supply elasticities was approached with an econometric estimation using panel data ranging the years 1990 to 2016. Through the use of a global sensitivity analysis and stochastic evaluations of the economic impact under different values for this parameter, we provided additional insights and points for discussion. However, market shocks as severe as simulated in some of the disease spread scenarios could very well fundamentally alter actors' behavior and thereby contradict estimates obtained from historical data alone. Nevertheless, our analysis shows that improving producers' adaptability to price changes is likely to reduce total impacts and in particular impacts to consumers. Consequently, regulatory support to foster adaptability of producers is called for.

## 4.5 Conclusions

While previous research has analyzed possible impacts of *Xylella fastidiosa* subspecies *pauca* to olive growers [222], the distribution of economic consequences among olive oil producers and consumers has not yet found attention. By using a multi-country partial equilibrium model, we show that price responses following reductions in supply due to the invader are likely to redistribute the negative economic impacts to consumers. This market response is expected to be particularly strong due to the inelastic nature of the production system. If the pathogen spreads beyond the current extent in Italy, the decline in European consumer welfare, due to higher prices for olive oil, ranged between 4.1 billion and 10.3 billion Euro over the course of 50 years

depending on the spread rate. In other words, each of the 195 million households in Europe [239], would incur additional costs ranging approximately 63 cents to 1.6 Euro every year over the course of 50 years.<sup>5</sup> Introductions into Greece and Spain could cost consumers between 0.4 to 3.3 and 1.8 to 53 billion Euro depending on the spread rate, respectively. This would correspond to additional annual costs ranging approximately 6 to 51 cents and 27 cents to 8.2 Euro for every household in Europe, respectively.

Our analysis stresses the importance of public actions to foster vigilance among growers and citizens to prevent additional introductions and further spread of the disease. Improving the communication of pest risk assessments to the public is crucial to create a better understand of the economic consequences if control measures were to be unsuccessful. Given the mutual ownership, between consumers and producers, of the risks associated with invasive species, the problem should be contextualized as a societal challenge in public debates on appropriate management strategies.

Further research is crucial on various fronts. Insights into the impact the pathogen has on yields and costs for different cultivars under different cultivation practices are important. Our understanding of the epidemiology needs to be deepened through further analyses of drivers for spread and speed with which dispersal might occur. It seems unlikely that *Xylella fastidiosa* subspecies *pauca* will be eradicated from Europe. Consequently, science and stakeholders must collaborate in finding feasible strategies for long term adaptation. Here, the development of resistant varieties and improved means of environmentally responsible vector control are invaluable avenues for further work. The need for enduring surveillance and control calls for additional research on cost effective detection methods.

---

<sup>5</sup> Equivalent annual costs of the losses to consumers were computed by dividing the total present value losses to consumers (see Table 4.3) by the annuity factor [240], using a discount rate of 3 percent and a time horizon of 50 years. The equivalent annual costs were then divided by the number of households in the EU-27 in 2019 [239]

## **Acknowledgements**

This work was supported by funding from the European Union's Horizon 2020 Pest Organisms Threatening Europe (POriTE) research and innovation program under grant agreement number 635646. The species distribution model was developed by Juan A. Navas-Cortes. The spread model was developed in collaboration with Martina Cendoya and Antonio Vicent. We thank the two anonymous reviewers for their constructive comments.

## 4.6 Supplementary Material

**Table 4.5** Regression output for the elasticity estimation.

Regressed on log-transformed production		
	<b>Coefficient</b>	<b>Std. Error</b>
Intercept	-12.721	11.408
year	0.008	0.005
Spain	0.816***	0.144
Italy	0.265**	0.109
Other	-1.503***	0.229
Log-transformed first lag of price	0.251	0.171
Log-transformed first lag of production	0.224**	0.102
R <sup>2</sup>	0.921	
Adj. R <sup>2</sup>	0.916	
Number of observations	101	

\*\*\*  $p < 0.01$ ; \*\*  $p < 0.05$ ; \*  $p < 0.1$ .

Table 4.6 Country specific changes in welfare over 50 years in billion Euro for spread scenarios in Italy.

Spread Scenario Suitability Spread Threshold Rate	Consumer				Producer				Non-Affected Producer				Affected Producer		Total				
	IT	EL	ES	RoEU	IT	EL	ES	RoEU	IT	EL	ES	RoEU	IT	EL	ES	RoEU	Total		
T1	RR05	-1.21	-0.41	-1.46	-1.01	-3.76	0.68	2.99	0.27	-4.44	0.68	2.99	0.27	0.68	-4.97	0.27	1.53	-0.74	-3.91
T2	RR05	-1.21	-0.41	-1.46	-1.01	-3.76	0.68	2.99	0.27	-4.44	0.68	2.99	0.27	0.68	-4.97	0.27	1.53	-0.74	-3.91
T3	RR05	-1.21	-0.41	-1.46	-1.01	-3.76	0.68	2.99	0.27	-4.44	0.68	2.99	0.27	0.68	-4.97	0.27	1.53	-0.74	-3.91
T1	RR50	-1.86	-0.62	-2.23	-1.56	-5.93	1.05	4.65	0.42	-7.14	1.05	4.65	0.42	1.21	-7.79	0.43	2.41	-1.13	-6.08
T2	RR50	-1.87	-0.63	-2.25	-1.57	-5.97	1.06	4.68	0.42	-7.19	1.06	4.68	0.42	1.22	-7.84	0.43	2.43	-1.14	-6.12
T3	RR50	-1.89	-0.64	-2.28	-1.59	-6.06	1.07	4.74	0.43	-7.31	1.07	4.74	0.43	1.25	-7.95	0.44	2.47	-1.16	-6.20
T1	RR95	-2.89	-0.97	-3.49	-2.45	-9.41	1.68	7.42	0.67	-11.23	1.68	7.42	0.67	1.82	-12.31	0.71	3.93	-1.78	-9.44
T2	RR95	-2.93	-0.98	-3.53	-2.47	-9.53	1.70	7.51	0.68	-11.38	1.70	7.51	0.68	1.85	-12.46	0.72	3.98	-1.80	-9.55
T3	RR95	-3.04	-1.02	-3.66	-2.57	-9.92	1.77	7.82	0.71	-11.86	1.77	7.82	0.71	1.94	-12.96	0.75	4.16	-1.86	-9.91

EL=Greece; ES=Spain; IT=Italy; RoEU=Rest of Europe.

**Table 4.7** Country specific changes in welfare over 50 years in billion Euro for spread scenarios in Greece averaged over all random points of introduction.

Spread Scenario Stability Spread Threshold Rate	Consumer			Producer			Non-Affected Producer			Affected Producer			Total						
	IT	EL	ES	RoEU	IT	EL	ES	RoEU	IT	EL	ES	RoEU	IT	EL	ES	RoEU	Total		
T1	RR05	-0.11	-0.04	-0.13	-0.09	0.09	-0.24	0.26	0.02	0.09	-0.30	0.26	0.02	0.06	-0.02	-0.28	0.13	-0.07	-0.24
T2	RR05	-0.10	-0.04	-0.13	-0.09	0.09	-0.23	0.25	0.02	0.09	-0.29	0.25	0.02	0.06	-0.02	-0.27	0.13	-0.06	-0.22
T3	RR05	-0.11	-0.04	-0.13	-0.09	0.09	-0.23	0.25	0.02	0.09	-0.29	0.25	0.02	0.06	-0.02	-0.27	0.13	-0.06	-0.22
T1	RR50	-0.54	-0.18	-0.65	-0.45	0.46	-1.21	1.33	0.12	0.46	-1.48	1.33	0.12	0.27	-0.08	-1.39	0.68	-0.33	-1.12
T2	RR50	-0.53	-0.18	-0.64	-0.44	0.45	-1.19	1.31	0.12	0.45	-1.46	1.31	0.12	0.26	-0.08	-1.37	0.67	-0.32	-1.10
T3	RR50	-0.53	-0.18	-0.64	-0.44	0.45	-1.18	1.30	0.12	0.45	-1.44	1.30	0.12	0.26	-0.08	-1.36	0.67	-0.32	-1.09
T1	RR95	-0.96	-0.32	-1.15	-0.80	0.82	-2.14	2.38	0.22	0.82	-2.55	2.38	0.22	0.41	-0.14	-2.46	1.23	-0.59	-1.96
T2	RR95	-0.98	-0.33	-1.18	-0.82	0.84	-2.19	2.43	0.22	0.84	-2.63	2.43	0.22	0.43	-0.14	-2.53	1.25	-0.60	-2.02
T3	RR95	-0.97	-0.33	-1.16	-0.81	0.83	-2.16	2.40	0.22	0.83	-2.59	2.40	0.22	0.43	-0.14	-2.49	1.24	-0.59	-1.98

EL=Greece, ES=Spain, IT=Italy, RoEU=Rest of Europe.

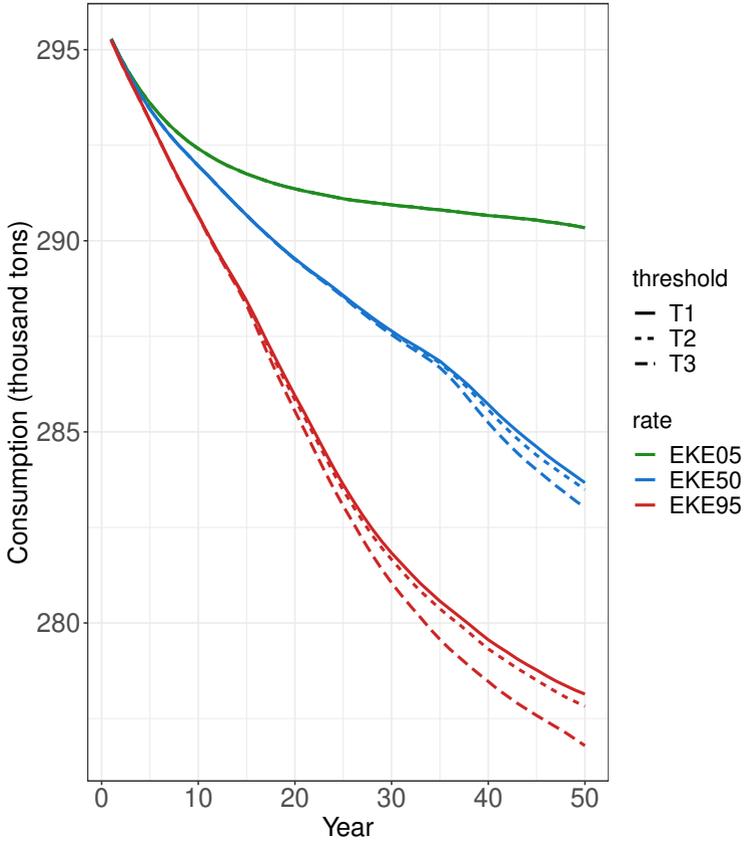
**Table 4.8** Country specific changes in welfare over 50 years in billion Euro for spread scenarios in Spain averaged over all random points of introduction.

Spread Scenario Stability Spread Threshold Rate	Consumer			Producer			Non-Affected Producer			Affected Producer			Total						
	IT	EL	ES	RoEU	IT	EL	ES	RoEU	IT	EL	ES	RoEU	IT	EL	ES	RoEU	Total		
T1	RR05	-0.52	-0.18	-0.63	-0.43	0.45	0.29	-0.09	0.12	0.45	0.29	-0.40	0.12	0.31	-0.07	0.12	-0.72	-0.32	-0.99
T2	RR05	-0.52	-0.18	-0.63	-0.43	0.45	0.29	-0.09	0.12	0.45	0.29	-0.41	0.12	0.32	-0.07	0.12	-0.72	-0.32	-0.99
T3	RR05	-0.54	-0.18	-0.64	-0.45	0.46	0.30	-0.10	0.12	0.46	0.30	-0.43	0.12	0.33	-0.08	0.12	-0.74	-0.33	-1.03
T1	RR50	-5.42	-1.77	-6.61	-4.84	5.46	3.66	-2.53	1.43	5.46	3.66	-5.39	1.43	2.86	0.04	1.88	-9.14	-3.41	-10.63
T2	RR50	-5.88	-1.92	-7.20	-5.31	6.07	4.08	-3.07	1.60	6.07	4.08	-6.20	1.60	3.13	0.19	2.16	-10.27	-3.71	-11.63
T3	RR50	-6.17	-2.01	-7.56	-5.59	6.44	4.33	-3.38	1.69	6.44	4.33	-6.69	1.69	3.31	0.27	2.32	-10.94	-3.90	-12.25
T1	RR95	-12.19	-3.94	-15.09	-11.42	13.74	9.31	-8.04	3.61	13.74	9.31	-13.11	3.61	5.07	1.55	5.37	-23.13	-7.81	-24.02
T2	RR95	-13.52	-4.34	-16.83	-12.89	15.96	10.86	-10.45	4.19	15.96	10.86	-16.18	4.19	5.73	2.44	6.51	-27.28	-8.70	-27.03
T3	RR95	-14.96	-4.79	-18.75	-14.50	18.43	12.58	-13.27	4.83	18.43	12.58	-20.00	4.83	6.72	3.47	7.79	-32.02	-9.67	-30.43

EL=Greece; ES=Spain; IT=Italy; RoEU=Rest of Europe.

**Table 4.9** First and total order sensitivity indices for all economic parameter.

Parameter	First Order			Total Order		
	Index	Min.	Max.	Index	Min.	Max.
Demand Italy	-0.050	-0.090	-0.004	0.008	0.007	0.008
Demand Greece	-0.054	-0.093	-0.008	0.002	0.002	0.002
Demand Spain	-0.055	-0.094	-0.010	0.002	0.002	0.002
Demand RoEU	-0.055	-0.095	-0.009	0.000	0.000	0.000
Supply Italy	0.213	0.175	0.258	0.335	0.316	0.354
Supply Greece	-0.040	-0.080	0.008	0.017	0.016	0.019
Supply Spain	0.219	0.180	0.260	0.345	0.327	0.363
Supply RoEU	-0.054	-0.094	-0.008	0.001	0.001	0.001
Price Italy	0.108	0.069	0.152	0.186	0.176	0.195
Price Greece	-0.055	-0.095	-0.010	0.001	0.001	0.001
Price Spain	-0.050	-0.091	-0.005	0.011	0.010	0.012
Price RoEU	-0.055	-0.095	-0.009	0.001	0.000	0.001
Demand Elasticity Italy	-0.058	-0.099	-0.013	0.004	0.004	0.005
Demand Elasticity Greece	-0.055	-0.095	-0.010	0.001	0.000	0.001
Demand Elasticity Spain	-0.054	-0.095	-0.008	0.003	0.003	0.004
Demand Elasticity RoEU	-0.056	-0.096	-0.011	0.001	0.000	0.001
Demand Elasticity RoW	-0.053	-0.094	-0.007	0.002	0.001	0.002
Supply Elasticity Italy	-0.045	-0.085	0.000	0.015	0.014	0.015
Supply Elasticity Greece	-0.056	-0.096	-0.010	0.000	0.000	0.000
Supply Elasticity Spain	-0.052	-0.092	-0.007	0.005	0.005	0.006
Supply Elasticity RoEU	-0.055	-0.095	-0.010	0.000	0.000	0.000
Yield Decline	-0.054	-0.095	-0.009	0.002	0.002	0.003
Cost Change	-0.054	-0.095	-0.009	0.003	0.002	0.003
Discount Rate	0.113	0.072	0.160	0.208	0.199	0.217



**Fig. 4.4** Change in the European exports of olive oil for spread of Xfp in Italy beyond the current extent.

## Chapter 5

# Predicting hotspots for invasive species introduction in Europe

### Abstract

Plant pest invasions cost billions of Euros each year in Europe. Understanding where hotspots of pest introduction are located could greatly help prevention and management. Here, we assess whether data-driven risk maps produced by machine learning methods could supplement the costly species-specific risk analyses currently conducted by governmental agencies. An elastic-net algorithm was trained on a dataset covering 248 invasive species to map risk of new introductions in Europe as a function of climate, soils, water, and anthropogenic factors. Results revealed that the BeNeLux states, Northern Italy, the Northern Balkans, and the United Kingdom, and areas around container ports such as Antwerp, London, Rijeka, and Saint Petersburg were at higher risk for introductions. Our analysis shows that machine learning can produce hotspot maps for pest introductions with a high predictive accuracy, but that systematically collected data on species' presences *and* absences are required to further validate and improve these maps.

### 5.1 Introduction

Biological invasions describe inadvertent introductions of organisms into new territories. While many entries may not lead to long-term

establishment [8], successful establishments of hazardous species can have major consequences for ecosystems and economies [216, 217]. A reliable prioritization of areas for potential introduction would be invaluable to inform surveillance effort [241, 242].

By definition, *introduction* of a species comprises *entry* and *establishment* [243]. Entry of a pest describes its movement into an area and establishment the perpetuation of the species within an area after successful entry [243]. Species distribution models (SDMs)<sup>1</sup> are popular data-driven tools that aim at predicting species' niches on the basis of environmental characteristics of known locations of occurrences [245]. Subsequently, a prediction of the potential area of establishment is derived by assessing the similarity in environmental conditions in other, possibly unsampled, locations. SDMs are commonly developed for specific species. While results from such analyses help to identify risky areas, estimate potential impact and develop management strategies [222, 246], they require species-specific data acquisition, calibration and validation. As a consequence of the time, effort and expertise required for this task, such species-specific analyses are only available for a few hazardous invaders [41]. A generic approach that could help to identify areas that are generally more at risk for pest introduction, without having to first develop a range of species-specific models, would greatly improve evidence-based prevention and management.

The vast majority of SDMs rely exclusively on climatic data to predict where a particular species may establish and maintain a population without the need for further immigration [247]. For invasive species, a growing body of literature stresses the role of anthropogenic factors in the introduction of species [5, 48, 99, 201, 248, 249]. Consequently, such data could very well improve predictions of hotspots for species introduction [250]. Nevertheless, there have been limited efforts to include anthropogenic features (i.e., predictor variables) into such models [99]. The type of approach determines how to interpret results. SDMs based exclusively on climate data map areas suitability for establishment. While some anthropogenic features are expected to ease establishment, others are related to entry, such as distances to container ports and road density. Consequently, maps derived from

---

<sup>1</sup> Also known as bioclimatic models, climate envelopes, ecological niche models, habitat models, resource selection functions, range maps, among others [244]

such features would not assess the risk of establishment but depict the suitability for introduction (i.e., entry and establishment).

We aim to develop a generic modelling approach to identify hotspots for plant pest introductions. We assess the risk of presence in Europe for the whole group of 248 invasive species on the priority lists (A1 and A2) of the European and Mediterranean Plant Protection Organization (EPPO). The A1 list contains species that are absent from Europe while the A2 list contains species with a geographically limited presence. We obtained worldwide data on the presence of these species from the Global Biodiversity Information Facility (GBIF). Background data<sup>2</sup> were generated using three standard methods recommended by the literature [251, 252]. Global georeferenced data on a wide range of potential predictors related to climate, soils, water, and anthropogenic factors were collected, and an elastic-net machine learning algorithm was trained on around 400,000 observations across the globe to predict new introduction of invasive species as a function of the predictors. The hyperparameters<sup>3</sup> were tuned using three cross-validation techniques. Although the resulting risk maps all have high predictive performance, they show striking differences depending on the background data generating techniques and cross-validation methods considered. Our analysis shows that machine learning can produce hotspot maps for plant pest introduction with a high predictive accuracy, but that systematically collected data on species' presences *and* absences are required to further validate and improve these maps.

---

<sup>2</sup> Background data characterize the feature space and act as pseudo-absences to which presence data are compared within the classification model. They do not necessarily aim to be true absence points, but rather provide a characterization of possible values features could take throughout the studied geographic area.

<sup>3</sup> The term hyperparameter denotes a parameter that controls the learning process of the algorithm but that is not directly inferred from the training (i.e., fitting) of the model as is the case for coefficients. In other words, hyperparameters hold settings that influence the structure of the model. A standard approach is to tune these hyperparameters (i.e., optimize) by running the learning algorithm for different values and choosing the hyperparameter value that results in the best performance according to a cross-validation procedure. The elastic-net has two hyperparameters (section 5.3.4).

## 5.2 Results and Discussion

Irrespective of the cross-validation technique, 20 percent of the data were randomly selected, withheld from training steps, and used for testing performance of the models on unseen data. On the 80 percent remaining data, we implemented and compared three cross-validation techniques to optimize the model hyperparameters (i.e., the parameters of the penalty term of the elastic-net). First, we followed the most widely used approach of randomly splitting the data into folds<sup>4</sup> [253]. Second, we separated data into continental spatial blocks (Supplementary Material: Figure 5.16). This technique intends to assess the transferability of the model across geographic space [253–255]. Lastly, we used temporal splits for cross-validation in which presences were separated by their year of record. Through forward chaining of the temporal folds<sup>5</sup>, we intended to test the model’s ability to predict future introductions.

For each cross-validation technique, we used three methods to generate background data representing pseudo-absences as opposed to true absences corresponding to absences verified by field surveys. First, we followed the most widely used approach of randomly generating background data across the studied area (denoted random) [251], being the entire globe (Supplementary Material: Figure 5.6). Second, we generated data from a biased background (denoted *kdbias*). This approach intends to mimic the geographic bias in the presence database [252, 257]. Lastly, we combined the *kdbias* approach with

---

<sup>4</sup> A fold is a term used in machine learning to describe subsets of the data. For example, five randomly split folds correspond to five data partitions each holding 20 percent of the training data.

<sup>5</sup> We refer to *forward chaining* to describe an out-of-sample approach in which the temporal order of the cross-validation folds is considered. In the first iteration, the cross-validation starts by training the algorithm on data from the first time-period and validating performance on data from the second time-period. In the second iteration, data from the first and the second time-period is used for training and performance validated on data from the third time-period, and so on. Hence, data available for training is growing over time. The first time-period is not used for validation, whereas the last time-period is not used for training within the cross-validation. This cross-validation, and comparable variations, is commonly used for time-series analyses. The technique is reviewed under the name *prequential growing window* within Cerqueira et al. [256]

Barbet-Massin et al.'s [251] recommendation for geographic exclusion (denoted `kd05dfar`). Here, we again generated data from a biased background but subsequently reduced the bias by removing data that were less than 5 decimal degrees away from any presence data (Supplementary Material: Figure 5.8).

### 5.2.1 European Hotspots for Pest Introductions

As our objective is the analysis of hotspots to improve the management of future introductions, we believe that temporal cross-validation most closely represents our objective. However, the spatial-block design best mimics spatial transferability<sup>6</sup> [253, 258, 259]. We will discuss average results, across different background generation approaches, for models tuned on temporal and continental splits.

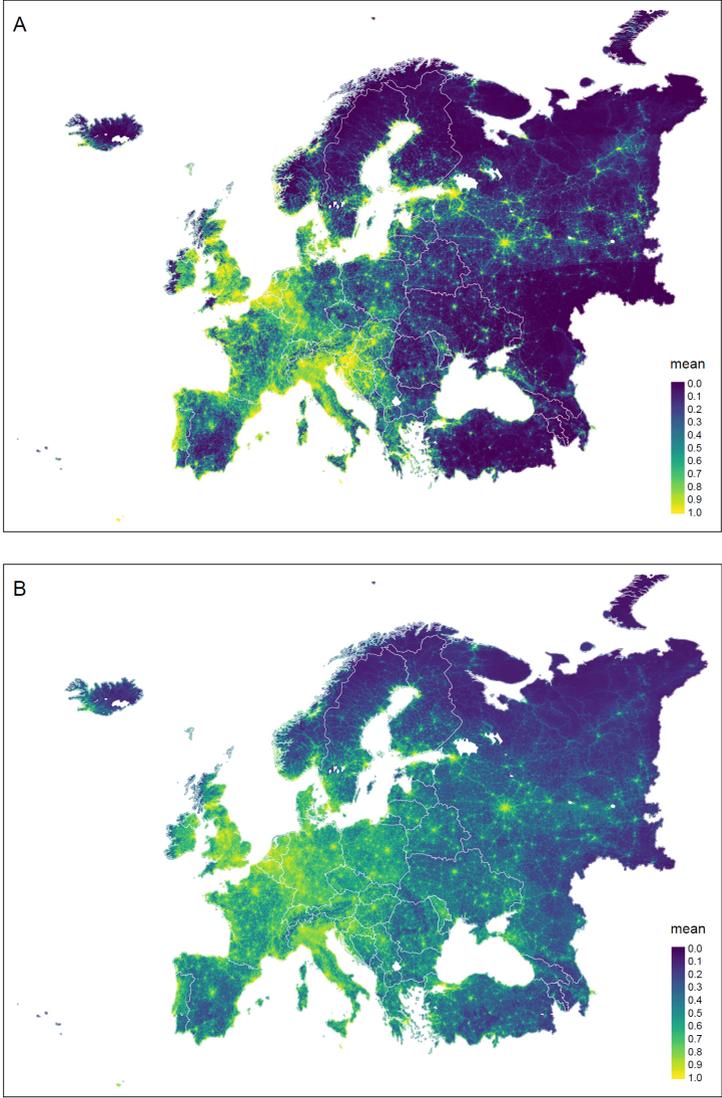
Figure 5.1 depicts the average predicted risk of new introduction, across the three approaches for background data generation, for models tuned using temporal and continental cross-validation. Irregular, polygon-like, surfaces in the maps are results from the input data on water indicators which came in the form of spatial-polygons. More importantly, hotspots, i.e., areas with high probability of presence of at least one invasive species, were consistently predicted to fall into highly anthropogenically-impacted areas. The BeNeLux states, Northern Italy, the Northern Balkans, and the United Kingdom were generally predicted to be at higher risk of future introductions. The contrast between regions at low and high risk was higher in models tuned on temporal folds compared to models tuned on spatial folds.

### 5.2.2 Feature Contributions

The importance of each feature (i.e., variable) was computed using the Feature Importance Ranking Measure [260, 261]. Here, we discuss

---

<sup>6</sup> Transferability describes the ability of the model to generalize and correctly predict new areas or time periods.



**Fig. 5.1** Probability of new introduction of at least one invasive species across Europe. The mean risk levels across different models tuned using temporal (a) and spatial (b) cross-validation.

feature contributions based on their average score across the different background generation approaches.

For models tuned on temporal splits, the highest ranked features were soil sand content and temperature-related features. While locations characterized by sandier soil at a 60 cm depth were associated with higher risk scores, higher sand content at a 30 cm depth was associated with decreased risk. This could be related to different land uses for soils characterized by less sandy topsoil. Many of the analyzed species are forestry pests. As forests are often characterized by sandier soils, it could explain why higher values for the soils' sand content were found to increase risk. Minimum and maximum temperatures in the different months had varying effects. For example, higher minima in February consistently increased risk while higher minima in January consistently decreased it. Next to soil and temperature features, port connectivity, water availability, water withdrawal, access to cities, the minimum distance to a port, a spatial proxy of the Gross Domestic Product (GDP), and the road density of highways were important. Higher values for anthropogenic features were generally associated with higher risk scores.<sup>7</sup>

For models tuned on continental splits, the feature ranking differed considerably compared to the models tuned to predict into future time periods. Here, anthropogenic features dominated the ranking. Across all approaches to background data, the degree of nightlight radiance, being our spatial proxy of GDP, consistently ranked very high as a risk increasing factor. Accordingly, access to cities, minimum distance to a port, the human impact on the environment, road densities for various road types, water withdrawal, population density, and the degree of human modification of terrestrial systems were important features. In general, effect directions again suggested that areas with a higher anthropogenic impact are at a higher risk. Next to anthropogenic features, higher values for drought severity, elevation, seasonal variation in water, precipitation seasonality, and the land cover classification for cultivated and managed cropland decreased risk, while the biome classification for temperate sclerophyll woodland and shrubland, and higher values for flood occurrences, biodiversity intactness, sand con-

---

<sup>7</sup> Access to cities and minimum distance to a port are inversely related to anthropogenic pressure as higher values correspond to longer driving times to a city and larger distances to a port, respectively.

tent in the soil, and precipitation in the driest quarter were associated with increased risk.

### 5.2.3 Model Performance

Model performance depended on both the cross-validation technique and the background generation approach (see Table 5.1). In terms of cross-validation, the highest performance scores were obtained by randomly splitting data into folds, followed by temporal splits and lastly continental splits. Randomly splitting not only resulted in higher average performance scores across validation folds, but also in a severely reduced variation of performance across folds compared to the temporal and continental techniques (Supplementary Material: Figures 5.17 to 5.19). The high performance with random splitting is likely related to spatial clustering of species presence. This violates the independence assumption and leads to models that overfit to residual dependencies, resulting in overoptimistic model performance [253].

Concerning the background generation approach, the highest performance scores were obtained with the random approach, followed by the geographic exclusion approach, and lastly the biased data generation technique. While the very good performance of the random technique is likely inflated by the large geographic scale considered here [262, 263], the lower performance for the biased approach is arguably over-pessimistic as the approach results in a large number of data points in the exact same locations yet opposing classifications for the dependent variable (Supplementary Material: Figure 5.7).

While the temporal and spatial block approaches did result in lower performance scores compared to randomly split folds, they test and optimize traits of the model that are desirable for our purpose which led us to present the models above. The performance of models to predict introduction into new geographic spaces or to provide a prioritization of areas for future introductions, is most appropriately estimated by cross-validation techniques that also simulate those behaviors. In addition, cross-validation techniques that simulate the modelling objective result in hyperparameter values that are optimized

**Table 5.1** Overview of model performances for all cross-validation and background generation techniques. Performance was measured by the area under the ROC curve computed by cross-validation or using an independent test dataset. An AUC of 1 indicates perfect classification while an AUC of 0.5 indicates random classification.

Cross validation	Background data	Cross validation			Test data*		
		AUC	Sens.	Spec.	AUC	Sens.	Spec.
random	random	0.99	0.95	0.94	0.99	0.95	0.94
random	kdbias	0.91	0.82	0.84	0.91	0.82	0.84
random	kd05dfar	0.99	0.95	0.96	0.99	0.95	0.96
spatial	random	0.95	0.78	0.92	0.97	0.92	0.92
spatial	kdbias	0.85	0.64	0.83	0.85	0.73	0.80
spatial	kd05dfar	0.95	0.78	0.88	0.97	0.91	0.93
temporal	random	0.97	0.89	0.94	0.99	0.95	0.94
temporal	kdbias	0.87	0.71	0.86	0.91	0.82	0.84
temporal	kd05dfar	0.98	0.90	0.96	0.99	0.95	0.96

Sensitivity (Sens.) and specificity (Spec.) were computed for a threshold of 0.5.

\*test data refers to 20% of randomly split withheld data.

for the task. As a result of the hyperparameter values, feature selection and model fitting are optimized for the research objective as well.

Performance was reasonably good, even for rigorous validation approaches such as temporal and continental splits, indicating that top-down analyses, through the bundling of species, do not necessarily sacrifice performance per se. Interestingly, global presence patterns were quite stable over time (Supplementary Material: Figures 5.10 to 5.15). As hotspots for pest introductions did not change considerably over the time horizon 1970 to 2021, the cross-validation scores obtained from temporal splits suggest that the models are very much able to predict future introductions based on past ones.

### 5.2.4 Sensitivity

Despite their somewhat comparable accuracies, the generated risk maps as well as the importance of the features of the corresponding models were drastically different (Figure 5.2). Similar to Austin [264], our analysis shows that equivalent performance metrics can result in very different models and outputs.

Figure 5.3 depicts the difference between the maximum and minimum probability values across the three approaches of generating background data for models tuned using temporal and continental cross-validation. The different background generation approaches resulted in sizable changes in predicted risk for large parts of France, Germany, Northern Spain, and Moldova. Individual maps for all approaches are provided in the Supplementary Material (Figures 5.29 to 5.43).

Figures 5.4 and 5.5 depict the 50 most important features, on average across different approaches to generating background data, for models tuned using temporal and continental cross-validation, respectively. The importance of features, and occasionally coefficient directions, varied considerably suggesting that very different models were created. See Figures 5.45 to 5.51 in the Supplementary Material for further examples of feature importance in different models. The results stress the diversity in models that can be built using the same presence data. Arguably, this underscores the need to explore sensitivity of results beyond computing several learning algorithms using the same data generating process, especially if clear data on true species absence are unavailable.

### 5.2.5 Implications for Pest Surveys

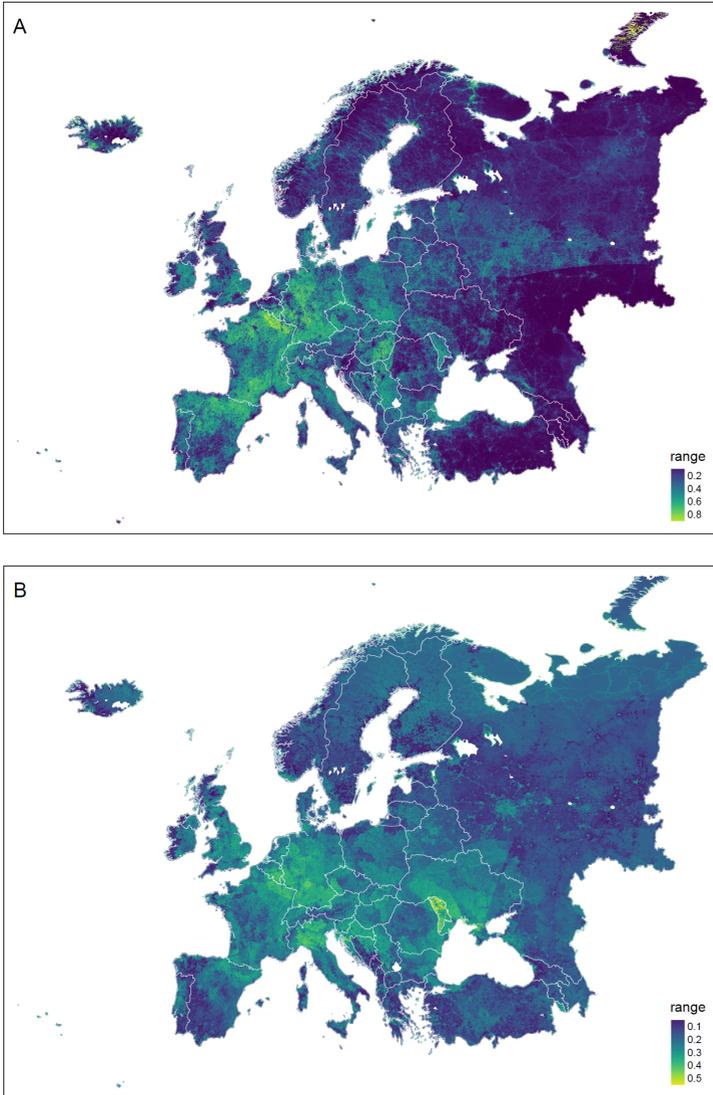
The lack of systematically surveyed species presence and true absence restricted us from disentangling whether predictions were a result of monitoring or reporting bias, or of area characteristics that indeed promote the introduction of invasive species. As an example, areas around container ports such as Antwerp, London, Rijeka, and Saint Petersburg, were generally predicted to be at high risk. The literature frequently discusses the involvement of international trade [5, 46, 201], in particular via boats and roads [48, 265], in the introduction of invasive species. Ecosystems characterized by a high level of anthropogenic disturbance are expected to facilitate species entry and establishment [46, 48]. Consequently, our results would be in line with the expectations from the literature. However, because regulators, scholars, and citizens expect that these areas likely contain new introductions, these

spatialCV_random									1.00									
spatialCV_kd05dfar									1.00	0.97								
spatialCV_kdbias									1.00	0.92	0.92							
temporalCV_random									1.00	0.82	0.88	0.87						
randomCV_random									1.00	1.00	0.82	0.88	0.87					
temporalCV_kd05dfar									1.00	0.93	0.93	0.83	0.89	0.84				
randomCV_kd05dfar									1.00	1.00	0.93	0.93	0.83	0.89	0.84			
temporalCV_kdbias									1.00	0.78	0.78	0.72	0.72	0.64	0.60	0.53		
randomCV_kdbias									1.00	1.00	0.78	0.78	0.71	0.71	0.63	0.60	0.52	
	randomCV_kdbias																	
	temporalCV_kdbias																	
	randomCV_kd05dfar																	
	temporalCV_kd05dfar																	
	randomCV_random																	
	temporalCV_random																	
	spatialCV_kdbias																	
	spatialCV_kd05dfar																	
	spatialCV_random																	

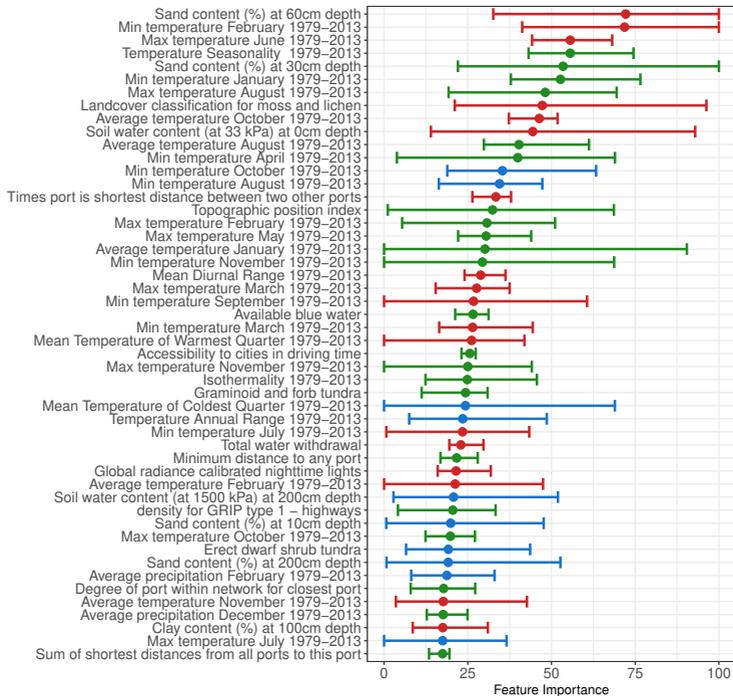
**Fig. 5.2** Spearman correlation of the European prediction across all combinations for cross-validation and background data generation.

locations are also often highly monitored, which could lead to reporting bias (Supplementary Material: Figure 5.44).

Systematic survey data on species presences and, equally important [99, 266], true absences would suspend these concerns entirely. Such data would allow to measure to what extent pest presences are driven by anthropogenic features, without having to ponder whether these characteristics were exclusively, or partially, related to biases [258, 267–269]. The inclusion and analysis of anthropogenic features is critical to further our understanding of externalities from human-driven land-use change, infrastructure, and trade. Efforts to include anthropogenic features into models, except for attempts to correct for data biases,



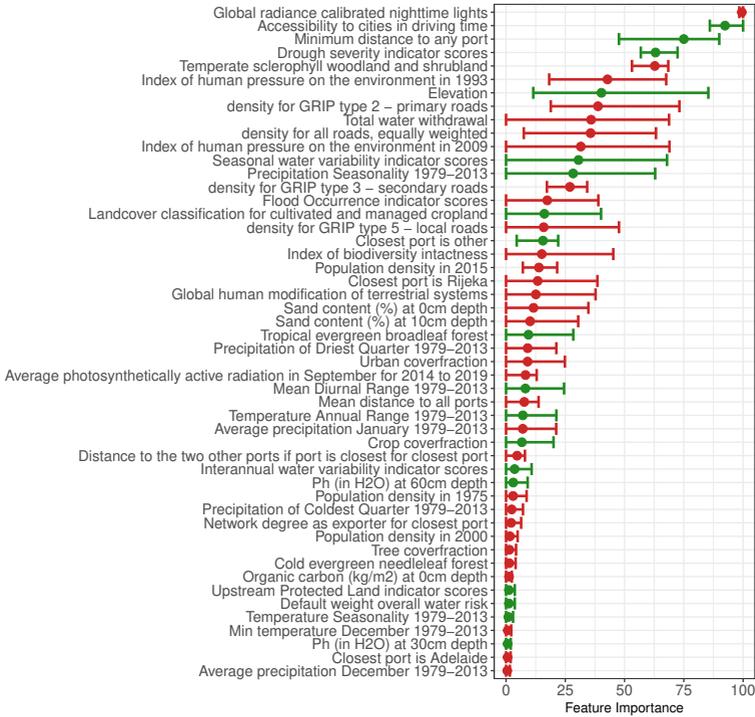
**Fig. 5.3** Difference between maximum and minimum risk level values for Europe across different background generation approaches for models tuned using temporal (a) and continental spatial-blocks (b) for cross-validation.



**Fig. 5.4** Fifty most important features across different background generation approaches for models tuned using temporal split cross-validations. The average is depicted as a dot. The line range shows the minimum and maximum values. If coefficients were consistently negative (i.e., risk decreasing) across all approaches for background data generation, the graph is colored green. If coefficients were consistently positive (i.e., risk increasing) across all approaches, the graph is colored red. If the coefficient direction changed between the different approaches to generating background data, it is colored blue.

are lacking [99]. The unavailability of systematic data for the left-hand side of the equation is likely a major reason for that.

Absence of a species may be due to one of the following causes [266]. First, *environmental absence* describes locations with unsuitable environmental conditions. Second, *contingent absence* describes locations which are suitable per se, but due to dispersal limitations, local extinctions, or an inadequate size of the suitable patch, among other



**Fig. 5.5** Fifty most important features across different background generation approaches for models tuned using spatial-block split cross-validations. The average is depicted as a dot. The line range shows the minimum and maximum values. If coefficients were consistently negative (i.e., risk decreasing) across all approaches for background data generation, the graph is colored green. If coefficients were consistently positive (i.e., risk increasing) across all approaches, the graph is colored red.

factors, they remained free of the species at the time of observation. Lastly, *methodological absence* describes locations which are falsely classified due to underlying biases, or incomplete coverage, in the available calibration data. SDMs predicting the fundamental niche aim to correctly classify environmental absences from presences, whereas contingent absences become particularly important when predicting the realized niche [99, 266]. Methodological absences taint predictions regardless of modelling purpose yet are likely to prevail in most data

used for SDM research [85]. While appropriate surveying for true absences requires considerable labor input [270], without such data, all analyses are bound to remain modelling exercises.

Global estimates suggest that the impact of invasive species runs in the trillions of Dollars [271]. For Europe, conservative estimates of annual impacts range from 12.5 to 20 billion Euro [51]. Several thousand species have already invaded Europe and the annual rates of new establishments are progressively increasing [41, 48]. The continuous rise in flow of products and people will likely only aggravate the risk of biological invasions in the future [4–6, 180]. Nevertheless, compared to estimates on current and future impacts, expenses for management and surveillance remain low [272]. While the process of hazardous invasions will remain random, predictive models in combination with the ever-increasing amount of georeferenced data can improve support of decision making in the future.

Harmonizing species surveys and making the resulting data available for research can further improve the prediction of hotspots. For invasive species on EPPO's priority lists, annual surveys are already conducted by the European member states. These data remain unharmonized across states, inaccessible to researchers, and without records of true absences. The inclusion of true absences in such efforts is as important for predictive models as the systematic collection of presences [99, 266].

While our analysis is a critical call for the need of systematic survey data, we believe the obtained results are a reason for optimism. In the last decades, previously unimaginable advances have been made in the breadth and quality of georeferenced environmental and anthropogenic data and computing technologies. Consequently, the quality of our predictions is more than ever bottlenecked by the lack of open data on results of systematic surveys and records on absence. Considering the current and potential future impact of invasive species to our ecosystems and economies, additional funding for species surveys would likely result in significant paybacks by informing the design of management strategies using predictive models.

## 5.3 Materials and Methods

### 5.3.1 Data

#### 5.3.1.1 Species presence

The list of species was obtained from the A1 and A2 list of EPPO (version 2020-09).<sup>8</sup> Both lists contain species that are recommended for regulation as quarantine pests in Europe. The A1 list contains species that are absent from Europe while the A2 list contains species with limited presence. Subsequently, on the 30th of March 2021, 490,323 occurrences were obtained from the Global Biodiversity Facility (GBIF) (10.15468/dl.fc5kva). The raw data was cleaned by removing all points with any of the following characteristics: reporting year prior to 1970, fossil specimen, literature-based observations, preserved specimen, location falling exactly on the centroids of capitals, or centroids of countries, or into sea, or on biodiversity institutions assuming that those are part of a collection [273]. Furthermore, presences with duplicated values across all features (i.e., input variables of the models) were removed, thereby, we effectively thinned presences at the scale of the finest environmental predictors. Similar criteria are often employed to determine the thinning radius [222]. The final set of 213,616 presence data, for which complete and unique combinations of feature data were available, spans 248 species, 92 families, 52 orders, 21 classes and 13 phyla, or more specifically 138 *Arthropoda*, 37 *Tracheophyta*, 2 *Mollusca*, 18 *Ascomycota*, 1 *Negarnaviricota*, 15 *Basidiomycota*, 19 *Proteobacteria*, 4 *Oomycota*, 2 *Cressdnaviricota*, 5 *Nematoda*, 2 *Actinobacteriota*, 4 *Kitri-noviricota*, and 1 *Chytridiomycota*. For model training, we classified presence of any species as a 1 and pseudo-absence as 0 (see next section).

---

<sup>8</sup> [https://www.eppo.int/ACTIVITIES/plant\\_quarantine/A1\\_list](https://www.eppo.int/ACTIVITIES/plant_quarantine/A1_list);  
[https://www.eppo.int/ACTIVITIES/plant\\_quarantine/A2\\_list](https://www.eppo.int/ACTIVITIES/plant_quarantine/A2_list)

### 5.3.1.2 Background data

The GBIF data come as presence-only and it was thus necessary to generate background data representing pseudo-absences to train and test our models, as commonly done in SDMs [251]. While this is common practice in the SDM literature, there is no consensus regarding the best approach [259]. The issue of generating background data is aggravated when presence data is biased [252]. While GBIF is extensively used in ecological research [274], geographic bias is very likely [268, 275]. We tested three ways to generate pseudo-absence data all of which find support in the literature [251, 252, 257].

First, random data were generated on a global scale covering all parts of the world except the poles. Randomly sampling background data is the default strategy in SDMs and frequently recommended [251]. The approach implicitly assumes that the entire geographic extent is equally relevant for the analysis and that the entire possible feature space should be used as a comparison to the presence data. Second, conceptually close to the bias-file approach of the popular SDM algorithm MaxEnt, we generated data from a biased background which aims at mimicking the geographic bias in the GBIF database [252, 257]. Here, presences were counted within 5 decimal degree grids. Next, a two-dimensional Gaussian kernel density was estimated on the count-grids and rescaled such that all values sum to unity. Subsequently, the resulting value was used as the probability of a background data point being generated in a location. With this second approach, the background data tend to remain close to the presence data, as would be the case if the sampling areas were kept close to each other. Thereby, an implicit assumption is made that only areas nearby known presences are relevant for the analysis and that the feature space used as a comparison to the presences should be restricted to nearby conditions. Lastly, we combined the biasing approach with Barbet-Massin et al.'s [251] recommendation for geographic exclusion. Here, we generated a larger number of data from a biased background and subsequently removed data that were less than 5 decimal degrees away from any presence data. From the remaining background data, a random subset was sampled such that the resulting data had a balanced number of presences and background data. Notably, while Barbet-Massin et al. [251] recommended a distance of two degrees, in latitude or longitude,

this criterion would have resulted in a questionable comparison in our case due to the large number and geographic spread of species presences in our database. Nevertheless, our approach intends to provide more background data within proximity of the presences [252], without heavily overlapping background data with presence data as done in the pure biasing approach as employed in MaxEnt. Hence, as with the second approach, the implicit assumption is made that only areas nearby known presences are relevant for the analysis, however, here the feature space used as a comparison to the presences does not comprise conditions of areas where pest presence is reported. We denote the three approaches as “random”, “kdbias”, and “kd05dfar”, respectively. The data generated by the three approaches are shown in the Supplementary Material (Figure 5.6 to 5.8). The final datasets had a balanced distribution of presences and pseudo-absences, i.e., a sampling prevalence of 50 percent [251, 276].

### 5.3.1.3 Features

Various georeferenced data were gathered. Table 5.2 in the Supplementary Material Provides an overview for the features and Table 5.3 for the raw data. Data on climate were obtained from Karger et al. [193]. Soil characteristics were obtained through OpenLandMap [277–286]. An indicator of erosion risk was obtained from the World Resource Institute [287]. Information on landcover was obtained from Buchhorn et al. [288]. A dataset on water related indicators was obtained from the World Resource Institute [289]. An indicator of biodiversity intactness was obtained from Newbold et al. [290]. Data on population density were obtained from the Joint Research Centre [291]. Data on road densities for different road types were obtained from Meijer, Huijbregts, Schotten, & Schipper [292]. An indicator of anthropogenic pressure on the environment was obtained from Venter et al. [293]. Data on human-driven modification of terrestrial systems was obtained from Kennedy et al. [294]. A spatial layer on accessibility to cities, measured in driving time, was obtained from Weiss et al. [295]. Studies advocated for the use of the Gross Domestic Product (GDP) in analyses of invasive species [41, 265]. However, GDP is generally only available at course, country-level, resolution. Therefore, we decided to proxy

GDP using spatial data on radiance of nightlights which were obtained from Hengl [296] [297–299]. Lastly, georeferenced data for container ports was obtained from Bartholdi, Jarumaneeroj, & Ramudhin [300]. These data comprise longitude and latitude as well as connectivity indices for 200 container ports around the world. For each presence and background point, the minimum distances to a port and the mean distance to all ports were computed. Subsequently, connectivity indices of the closest port, as well as connectivity indices for all ports, weighted by their inverse distances to a particular point, were used as features.

### 5.3.2 Data processing

First, observations with incomplete data were omitted. For categorical features, the 19 most frequent categories were kept, and other categories were grouped into one. Next, categorical features were dummy encoded. In addition to the spatially weighted port connectivity indices, the following continuous features were engineered: the average annual photosynthetically active radiation, the standard deviation of the photosynthetically active radiation across months, the change in population density between 1975 and 2015, and the change in human impact on the environment between 1993 and 2009. All continuous features were transformed to normality, centered, and scaled. The best transformation to normality was estimated from a set of candidate functions using only the training data [189]. The final datasets, for the three approaches to background data, have 427,232 points, half of those being presences and the other half background data, with complete data for 246 features. Out of those, 181 features were continuous, and 65 features were dummy encoded categories. Twenty percent of the data was kept from training for testing model performance.

### 5.3.3 Cross-validation techniques

After exclusion of 20 percent of the data for testing, three different cross-validation techniques were implemented on the remaining data.

First, we followed the most frequently used approach of randomly splitting the data into folds. To manage computational time, we used five folds. Second, we separated data into continental spatial blocks. Here, six folds were generated corresponding to the continents Africa, Asia, Australia, Europe, North America, and South America (Supplementary Material: Figure 5.16). As such, we intended to assess the transferability of the model across geographic space [253–255]. Lastly, we used temporal splits for cross-validation in which presences were separated by their year of record and background data randomly assigned, without replacement, such that a balanced fold was obtained. With the temporal cross-validation, we intended to test whether the model can predict future presences. Due to the exponential increase in presence records over time (Supplementary Material: Figure 5.9), we divided them into unequal time periods corresponding to the years 1970–2005, 2006–2011, 2012–2014, 2015–2016, 2017–2018, 2019, 2020–2021, resulting in seven folds with an approximately equal number of presences in each. Subsequently, models were trained and validated by forward chaining the folds (Supplementary Material: Figure 5.10 to 5.15).

### 5.3.4 Algorithm and hyperparameter tuning

The model is a generalized linear model based on a logit link, equivalent to a logistic model. The model includes regression coefficients that are estimated using a learning algorithm called elastic-net [301, 302]. The algorithm is a regularization technique that combines the L1 (sum of absolute coefficient magnitudes) and L2 (sum of squared coefficient magnitudes) coefficient penalties into the loss function. In doing so, the model is a generalization of the lasso and ridge regression approaches and allows for the estimation of pure versions of the two as well as mixed variants.

We decided to use this training algorithm because of its ability to find an optimal balance between bias and variance. The elastic net reduces variance at the cost of introducing bias to minimize the prediction error. This approach is called *regularization* and is designed to optimize the predictive performance of the model. The algorithm

is computationally relatively fast, memory efficient, and robust to correlated features. It is thus well-adapted to large scale practical applications.

The parameters are estimated using a penalized log likelihood objective function [302]. The likelihood is based on a binomial distribution and the penalization is based on the elastic net penalty. The elastic net includes a penalty term defined by two hyperparameters named  $\alpha$  and  $\lambda$ . The hyperparameter  $\alpha$  describes the mixing of the L1 and L2 penalties. If  $\alpha$  equals 1, the elastic net would essentially be a lasso regression whereas  $\alpha$  equal to 0 would result in a pure ridge regression. The hyperparameter  $\lambda$  denotes the degree of regularization employed. In the elastic-net algorithm, the regularization determines the extent to which coefficient magnitudes affect the loss function. Consequently, the regularization determines the extent to which coefficients are shrunk toward zero. By shrinking coefficient values, a model fit is obtained that might generalize the underlying relationships better.

Both hyperparameters were tuned using a grid search to maximize the AUC value computed successively with the three above-mentioned cross-validation techniques. The AUC metric measures the correctness in rankings between locations which is directly related to our modelling objective of identifying areas at risk [257, 303]. However, whenever true absences are not available, performance metrics represent heuristic measures only and should therefore be cautiously interpreted [254, 304]. Sensitivity and specificity receive equal attention in our results. The presented values correspond to values obtained at a cut-off of 0.5. For  $\alpha$ , values between 0 and 1 were searched at 0.1 increments and for  $\lambda$  values between 0 and 1 at 0.025 increments, resulting in a total of 451 combinations. In the Supplementary Material, Figures 5.20 to 5.28 depict the tuning results, and Table 5.4 depicts the optima, for each cross-validation technique and background data generation approach. Surprisingly, only the spatial block cross-validation resulted in regularized models, while the random and temporal splits suggested that no regularization yielded the best performance. No regularization ( $\lambda = 0$ ) essentially collapses the elastic net into a standard generalized linear model with binomial distribution. The tuning results for random and temporal cross-validation could be related to the spatial clustering of data which resulted in small regularization values in other studies

[303]. Following the hyperparameter tuning, the model coefficients were estimated using the entire training data.

To summarize, 20 percent of the data were withheld for testing performance on unseen observations, whereas 80 percent of the data were used for training. Within the cross-validation techniques, these training data were split into folds to tune the hyperparameters and obtain cross-validation-based performance metrics. Subsequently, using the tuned hyperparameter values, the feature coefficients were estimated on the entire training dataset (i.e., 80 percent of all data). Lastly, performance (AUC, sensitivity, and specificity) of the fitted model on unseen data was computed on the 20 percent of withheld testing data.

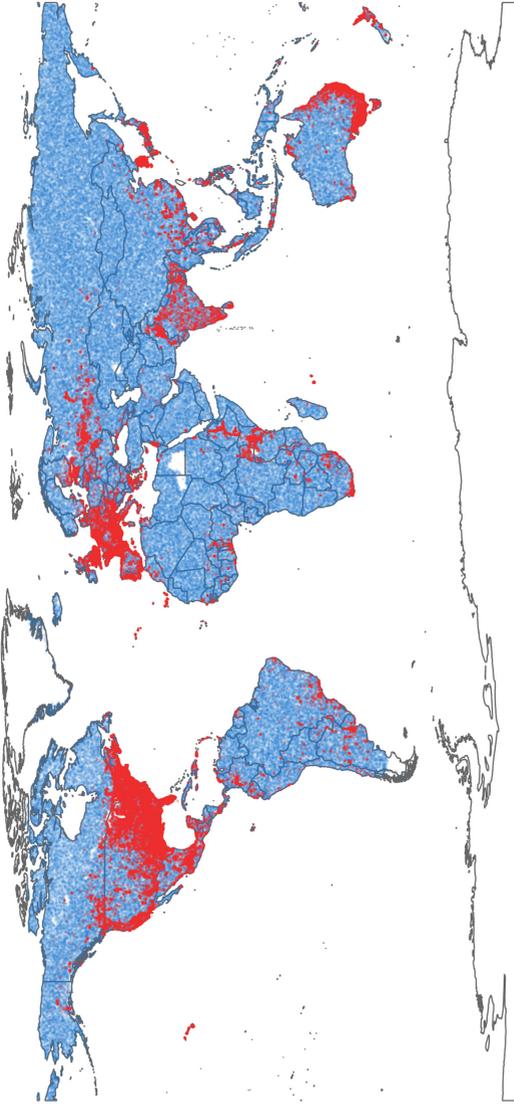
### 5.3.5 Prediction and Mapping

To circumvent the problem of differences in the resolution of input layers, longitude, and latitude coordinates for around 870,000 points across Europe were generated. The number of points was chosen such that the modelling steps are feasible in terms of computational time and memory requirements. Subsequently, for all points feature data were extracted and processed as described above. To minimize empty spaces in the risk map, due to systematically missing input data in certain locations, individual features were imputed for 73,380 points, with values of the geographically closest point within a maximum distance of 1 decimal degree. Points with partially missing data mostly fall on coastlines and on in-land waterbodies. Consequently, missing data is likely due to resolution-related artifacts of pixels which fall on non-linear country borders and unavailable information for some features.

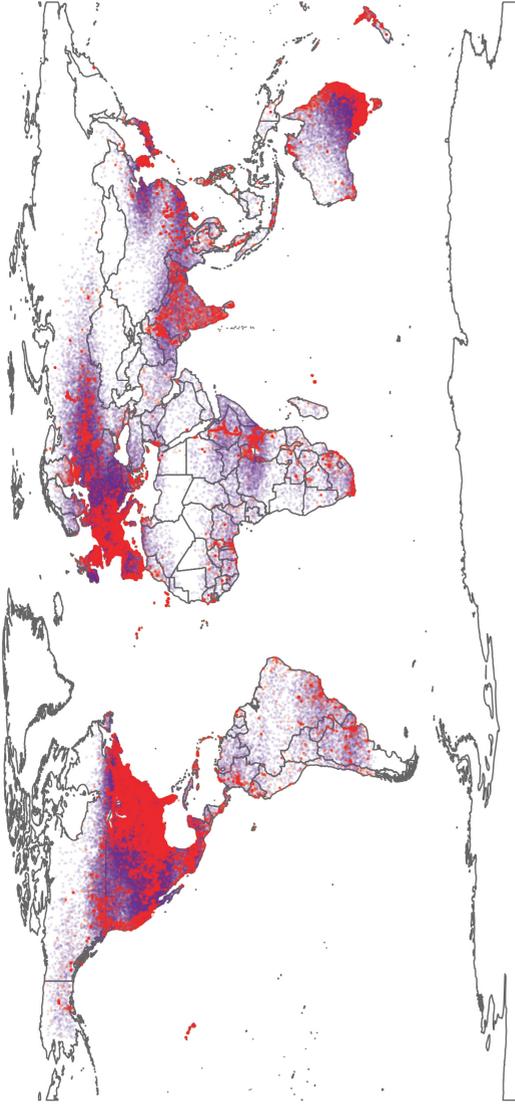
The trained and tuned models were used to generate a continuous probability score for introduction at all points in Europe. All maps are point-based. Each point was colored using the probability score. The figures shown within the manuscript depict the average probability score across the three background data approaches, for models tuned on temporal and continental cross-validation techniques. The sensitivity of this probability to the background data approach is shown through visualizations of the range of the probability score which

was computed by taking the difference between the maximum and minimum values for each point. Individual maps for all approaches are provided in the Supplementary Material (Figures 5.29 to 5.43).

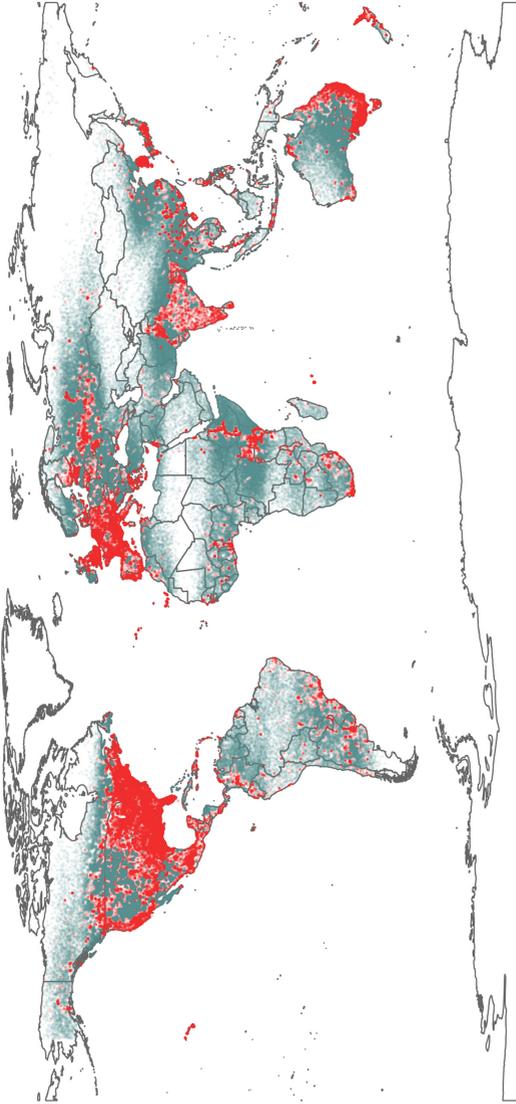
## **5.4 Supplementary Material**



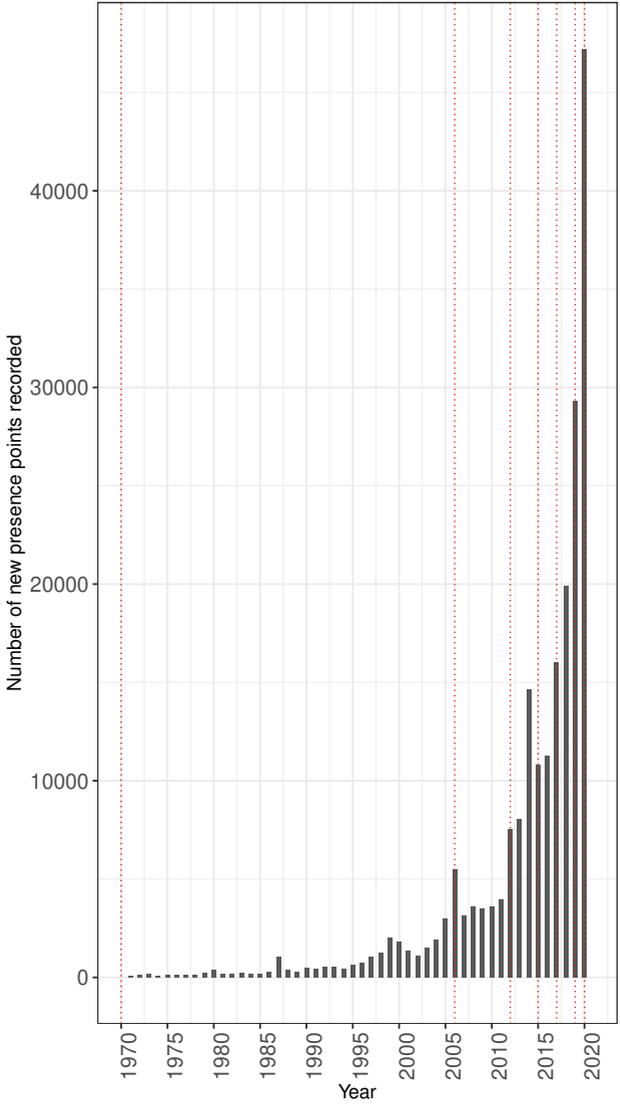
**Fig. 5.6** Overview of presence points (red) and randomly generated background points (blue). The total number of points is 427,232 with half being presences and the other half background. The randomly generated background points should, in theory, reach into every corner of the world. However, as points with missing feature data were removed some areas, such as the South tip of Argentina and Chile and parts of Egypt and Libya, have no points.



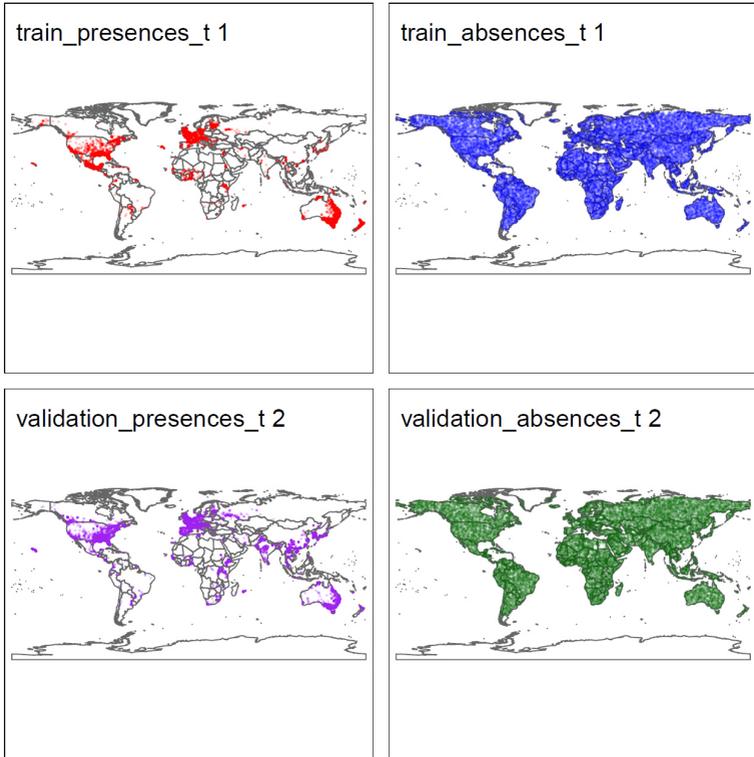
**Fig. 5.7** Overview of presence points (red) and kbias background points (purple). The total number of points is 427,232 with half being presences and the other half background. As a consequence of the biasing approach, many background points are plotted underneath the presence points.



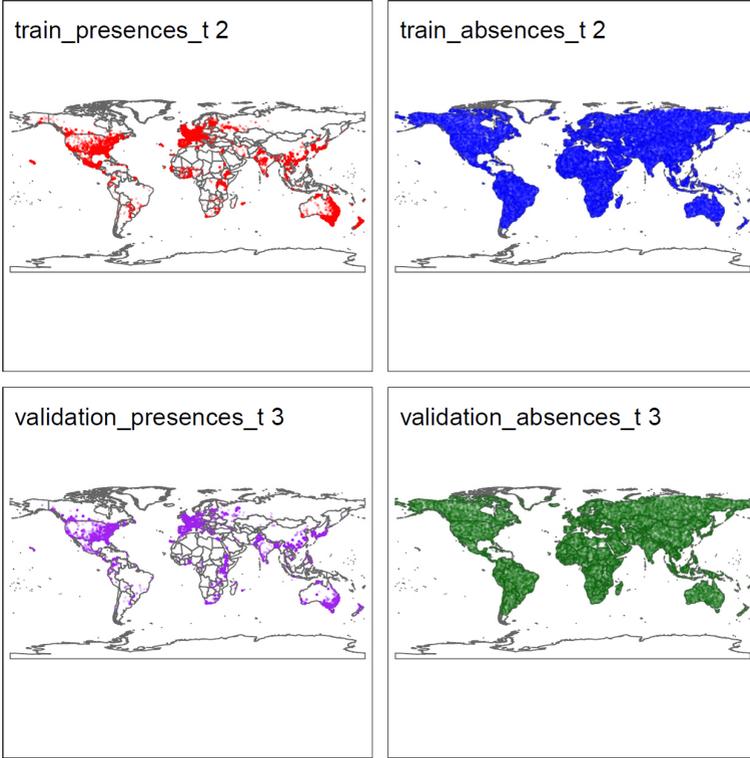
**Fig. 5.8** Overview of presence points (red) and background points generated by the combined approach of biasing and geographically excluding (grey). The total number of points is 427,232 with half being presences and the other half background. While this map might look similar to Figure 5.7 on first glance, the crucial difference is that areas with presences (red) do not have any background (grey) point. As the number of background points is unchanged, yet areas with known presences are prohibited from having background points, areas in near known presences have a larger amount of background points which results in a denser coloration compared to Figure 5.7.



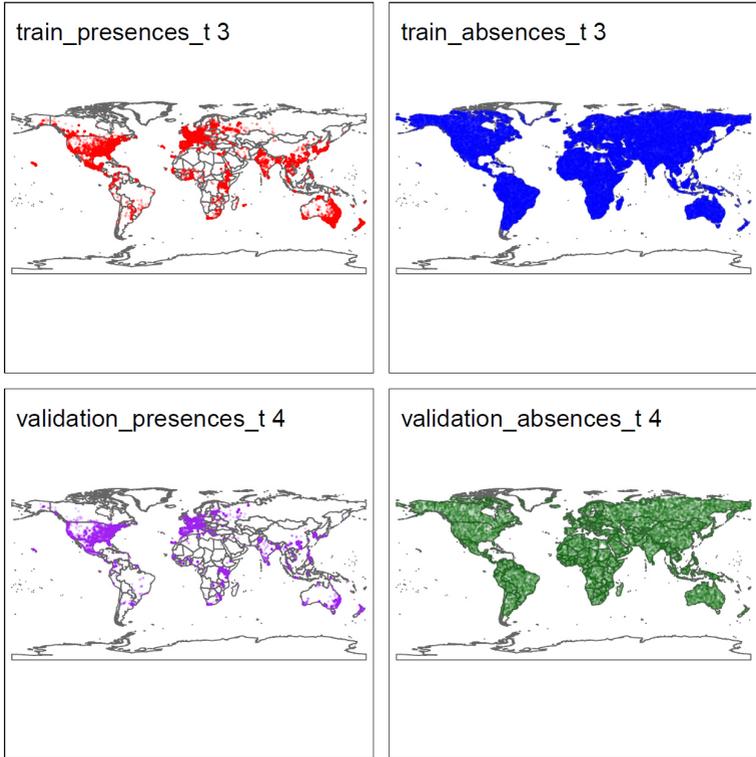
**Fig. 5.9** Number of presence records over time and chosen cut-offs for the temporal cross-validation (red). The temporal folds correspond to 1970-2005, 2006-2011, 2012-2014, 2015-2016, 2017-2018, 2019, and 2020-2021



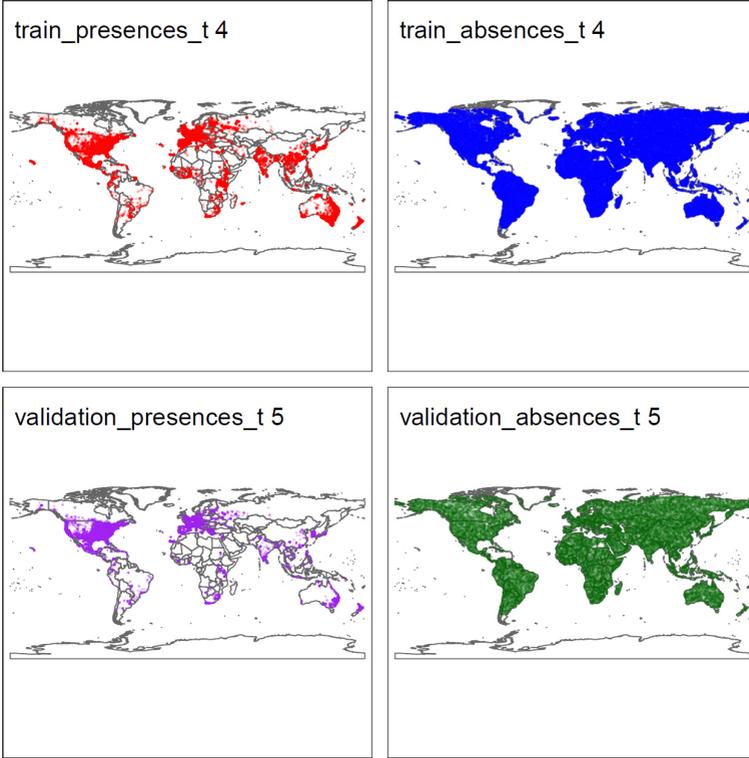
**Fig. 5.10** Presence and random background points used for training (cumulative over time) and validating in the first fold of the temporal cross-validation.



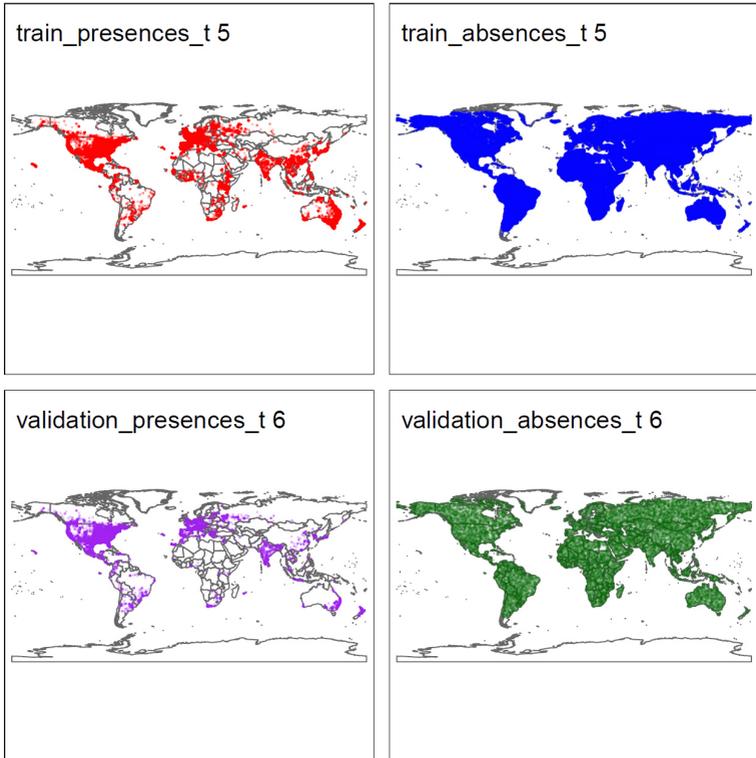
**Fig. 5.11** Presence and random background points used for training (cumulative over time) and validating in the second fold of the temporal cross-validation. Note, the training data is growing over time which means that the top row in the plot shows the data of, both, the first and second time periods.



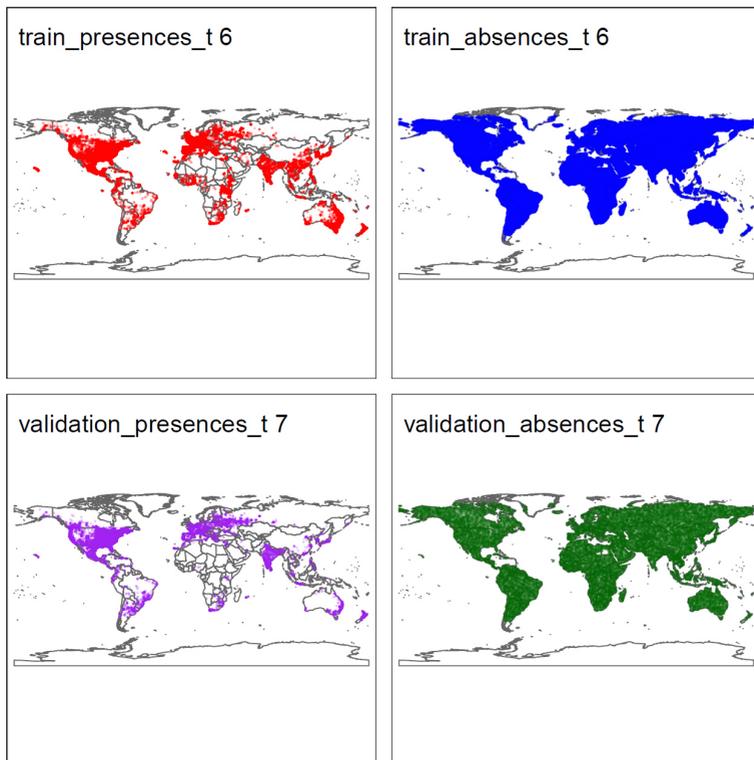
**Fig. 5.12** Presence and random background points used for training (cumulative over time) and validating in the third fold of the temporal cross-validation. Note, the training data is growing over time which means that the top row in the plot shows the data of the first, second and third time periods.



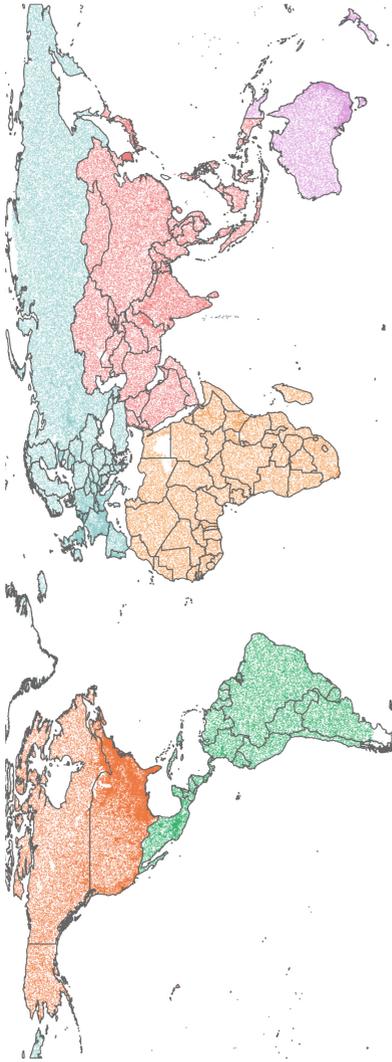
**Fig. 5.13** Presence and random background points used for training (cumulative over time) and validating in the fourth fold of the temporal cross-validation. Note, the training data is growing over time which means that the top row in the plot shows the data from the first to the fourth time periods.



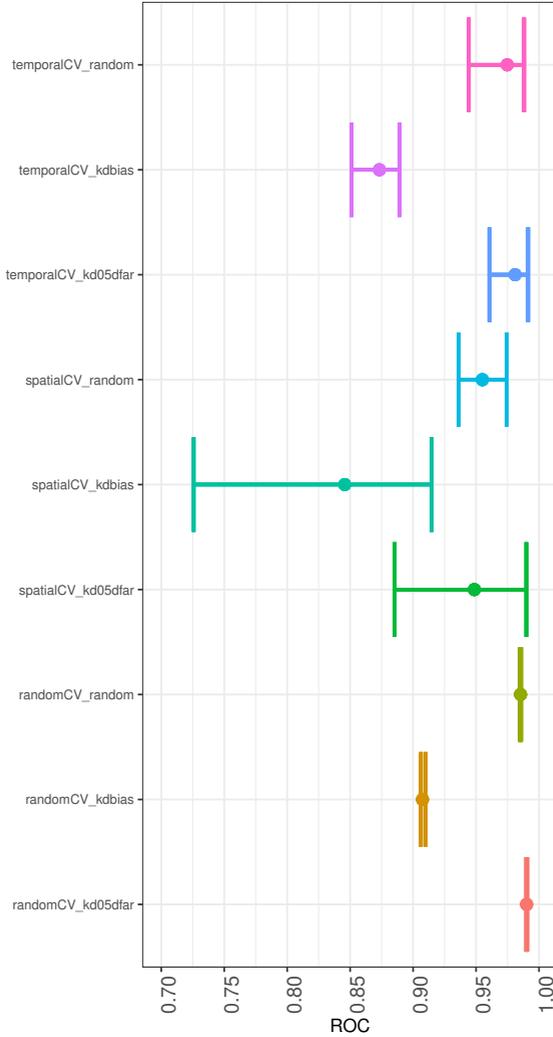
**Fig. 5.14** Presence and random background points used for training (cumulative over time) and validating in the fifth fold of the temporal cross-validation. Note, the training data is growing over time which means that the top row in the plot shows the data from the first to the fifth time periods.



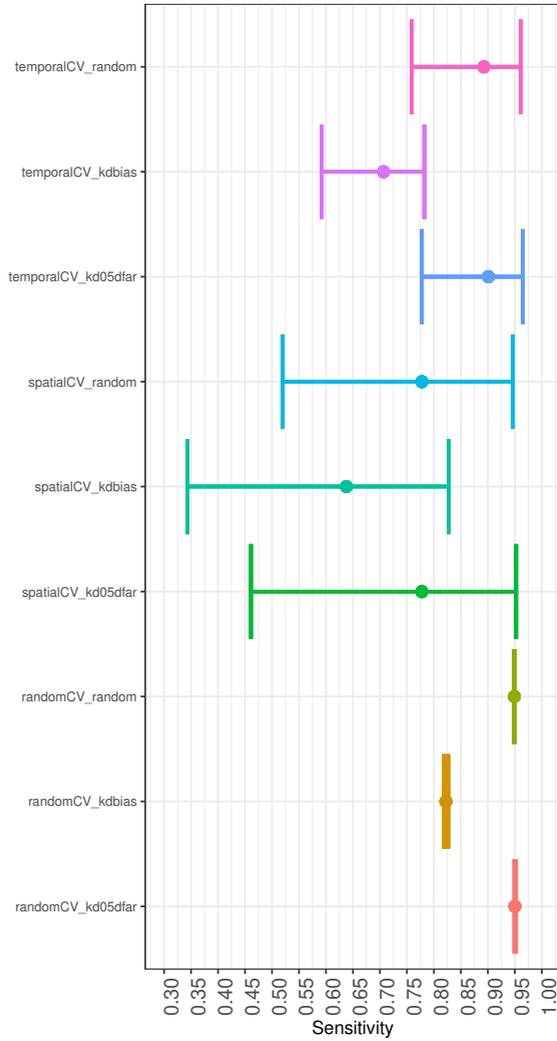
**Fig. 5.15** Presence and random background points used for training (cumulative over time) and validating in the sixth fold of the temporal cross-validation. Note, the training data is growing over time which means that the top row in the plot shows the data from the first to the sixth time periods.



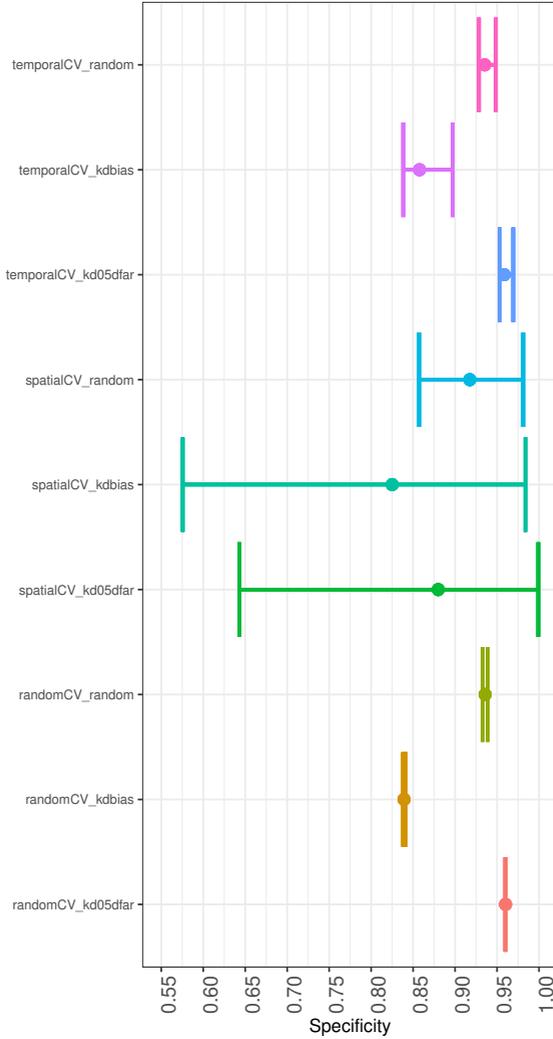
**Fig. 5.16** Overview of presences and random background points falling into the different continental spatial blocks. Here, presences and background points have the same color. Hence, the colors only show the spatial blocks used. The total number of points is 427,232 with half being presences and the other half background. The randomly generated background points should, in theory, reach into every corner of the world. However, as points with missing feature data were removed some areas, such as the South tip of Argentina and Chile and parts of Egypt and Libya, have no points.



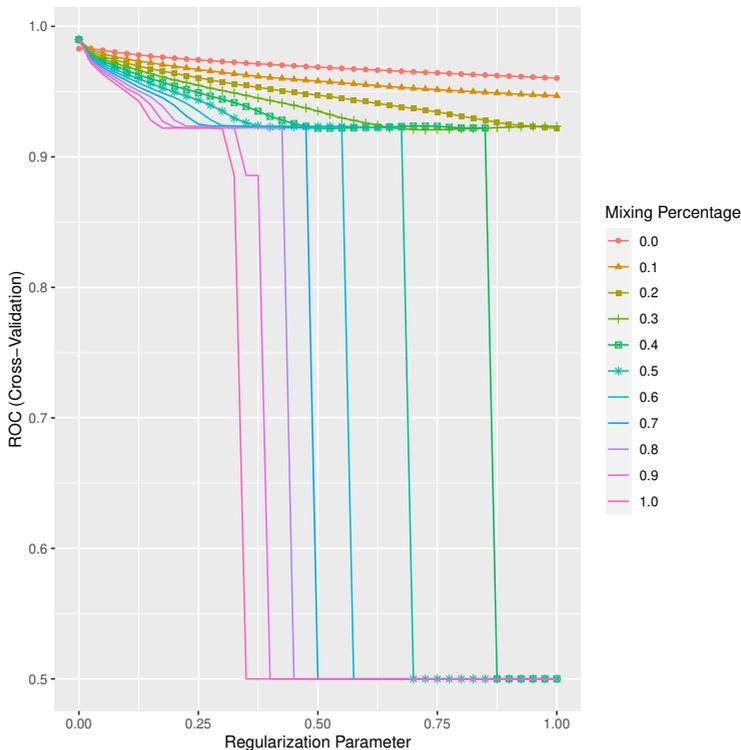
**Fig. 5.17** Cross validation performance, in terms of area under the ROC, for all approaches for cross validation and background point generation. The dots show the average scores, and the line ranges show the minimum and maximum values obtained across the different folds.



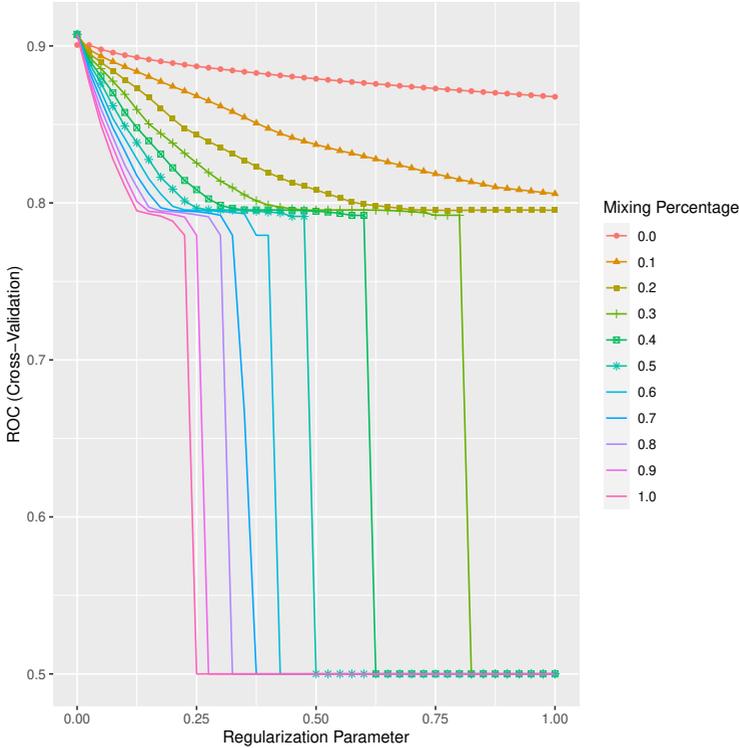
**Fig. 5.18** Cross validation performance, in terms of Sensitivity, for all approaches for cross validation and background point generation. The dots show the average scores, and the line ranges show the minimum and maximum values obtained across the different folds.



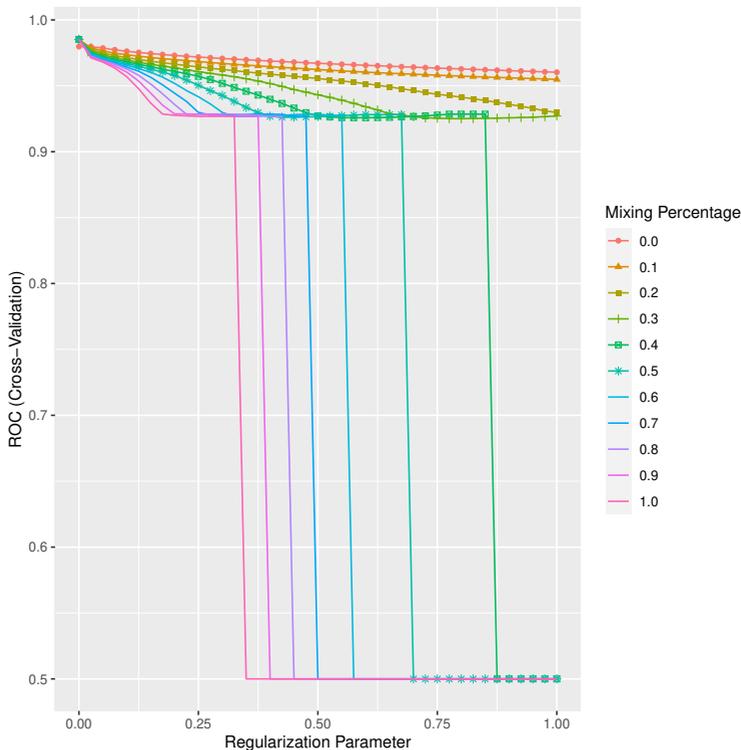
**Fig. 5.19** Cross validation performance, in terms of Specificity, for all approaches for cross validation and background point generation. The dots show the average scores, and the line ranges show the minimum and maximum values obtained across the different folds.



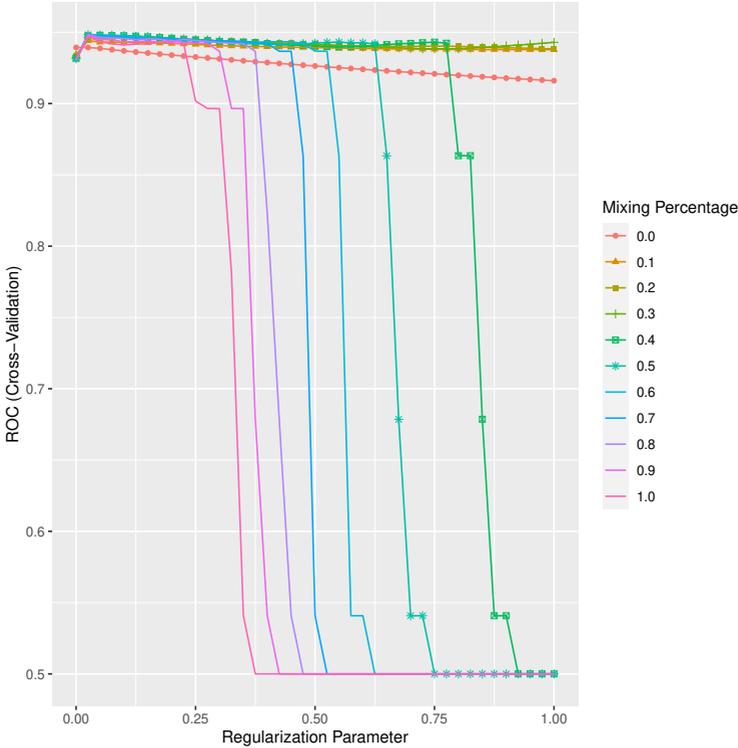
**Fig. 5.20** Hyperparameter tuning for random split cross validation and kd05dfar background points. The objective function of the elastic net includes a penalty term defined by two hyperparameters named  $\alpha$  (“Mixing Percentage”) and  $\lambda$  (“Regularization Parameter”). The hyperparameter  $\alpha$  describes the mixing of the L1 and L2 penalties. If  $\alpha$  equals 1, the elastic net would essentially be a lasso regression whereas  $\alpha$  equal to 0 would result in a pure ridge regression. The hyperparameter  $\lambda$  denotes the degree of regularization employed. Both hyperparameters were tuned using a grid search to maximize the area under the ROC. For  $\alpha$ , values between 0 and 1 were searched at 0.1 increments and for  $\lambda$  values between 0 and 1 at 0.025 increments resulting in a total of 451 combinations. No regularization ( $\lambda = 0$ ) essentially collapses the elastic net into a standard generalized linear model with binomial distribution. Table 3 below depicts the optima, for each cross-validation technique and approach to generating background data.



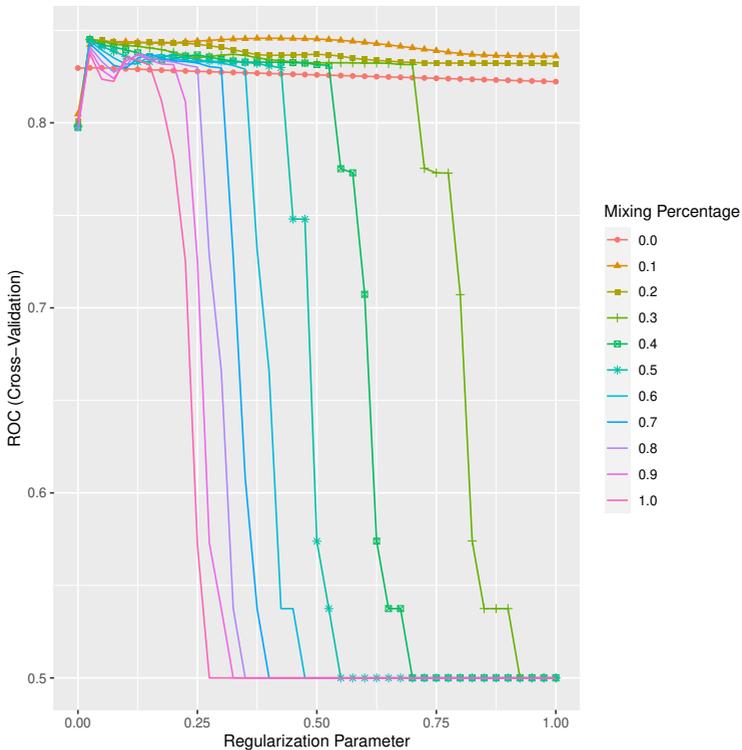
**Fig. 5.21** Hyperparameter tuning for random split cross validation and kdbias background points. The objective function of the elastic net includes a penalty term defined by two hyperparameters named  $\alpha$  (“Mixing Percentage”) and  $\lambda$  (“Regularization Parameter”). The hyperparameter  $\alpha$  describes the mixing of the L1 and L2 penalties. If  $\alpha$  equals 1, the elastic net would essentially be a lasso regression whereas  $\alpha$  equal to 0 would result in a pure ridge regression. The hyperparameter  $\lambda$  denotes the degree of regularization employed. Both hyperparameters were tuned using a grid search to maximize the area under the ROC. For  $\alpha$ , values between 0 and 1 were searched at 0.1 increments and for  $\lambda$  values between 0 and 1 at 0.025 increments resulting in a total of 451 combinations. No regularization ( $\lambda = 0$ ) essentially collapses the elastic net into a standard generalized linear model with binomial distribution. Table 3 below depicts the optima, for each cross-validation technique and approach to generating background data.



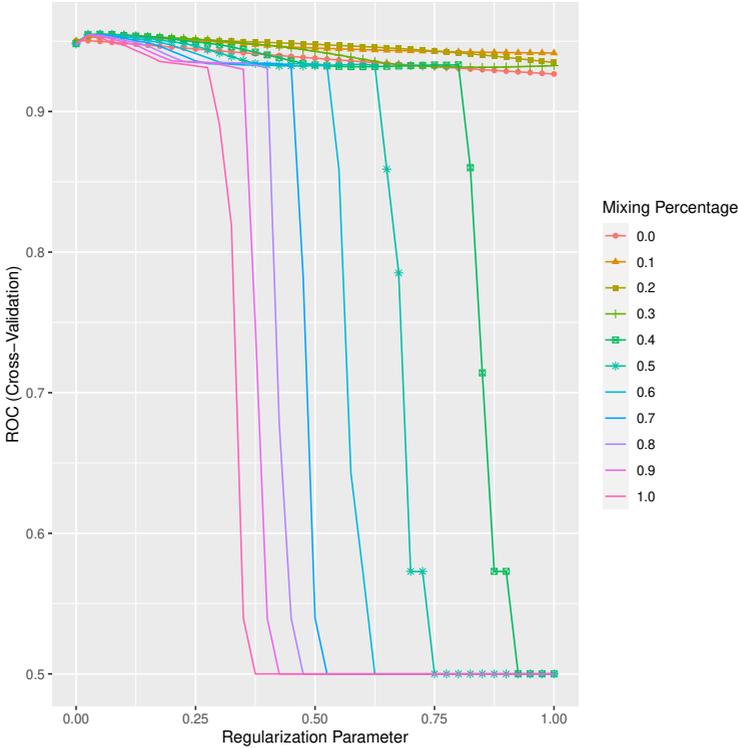
**Fig. 5.22** Hyperparameter tuning for randomly split cross validation and random background points. The objective function of the elastic net includes a penalty term defined by two hyperparameters named  $\alpha$  (“Mixing Percentage”) and  $\lambda$  (“Regularization Parameter”). The hyperparameter  $\alpha$  describes the mixing of the L1 and L2 penalties. If  $\alpha$  equals 1, the elastic net would essentially be a lasso regression whereas  $\alpha$  equal to 0 would result in a pure ridge regression. The hyperparameter  $\lambda$  denotes the degree of regularization employed. Both hyperparameters were tuned using a grid search to maximize the area under the ROC. For  $\alpha$ , values between 0 and 1 were searched at 0.1 increments and for  $\lambda$  values between 0 and 1 at 0.025 increments resulting in a total of 451 combinations. No regularization ( $\lambda = 0$ ) essentially collapses the elastic net into a standard generalized linear model with binomial distribution. Table 3 below depicts the optima, for each cross-validation technique and approach to generating background data.



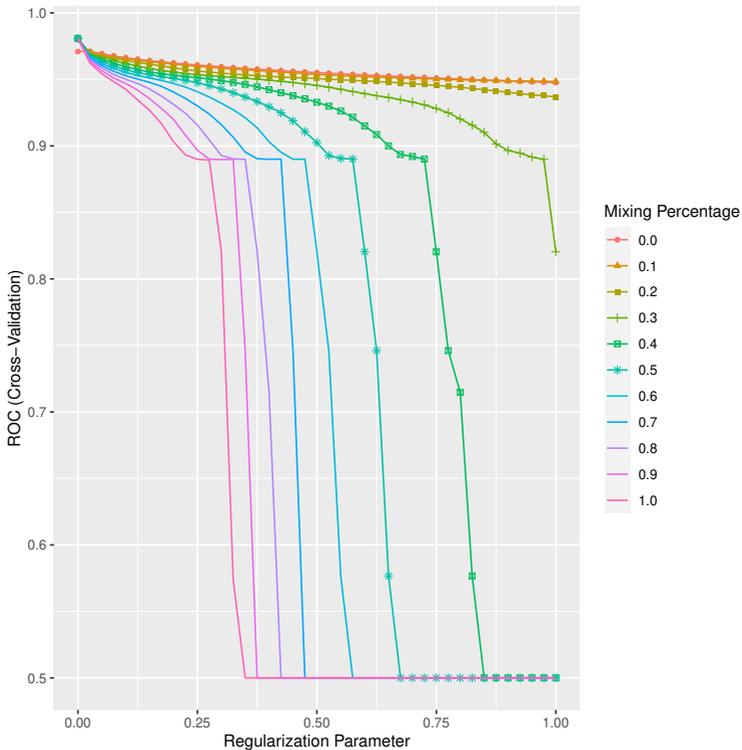
**Fig. 5.23** Hyperparameter tuning for spatial-block cross validation and kd05dfar background points. The objective function of the elastic net includes a penalty term defined by two hyperparameters named  $\alpha$  (“Mixing Percentage”) and  $\lambda$  (“Regularization Parameter”). The hyperparameter  $\alpha$  describes the mixing of the L1 and L2 penalties. If  $\alpha$  equals 1, the elastic net would essentially be a lasso regression whereas  $\alpha$  equal to 0 would result in a pure ridge regression. The hyperparameter  $\lambda$  denotes the degree of regularization employed. Both hyperparameters were tuned using a grid search to maximize the area under the ROC. For  $\alpha$ , values between 0 and 1 were searched at 0.1 increments and for  $\lambda$  values between 0 and 1 at 0.025 increments resulting in a total of 451 combinations. No regularization ( $\lambda = 0$ ) essentially collapses the elastic net into a standard generalized linear model with binomial distribution. Table 3 below depicts the optima, for each cross-validation technique and approach to generating background data.



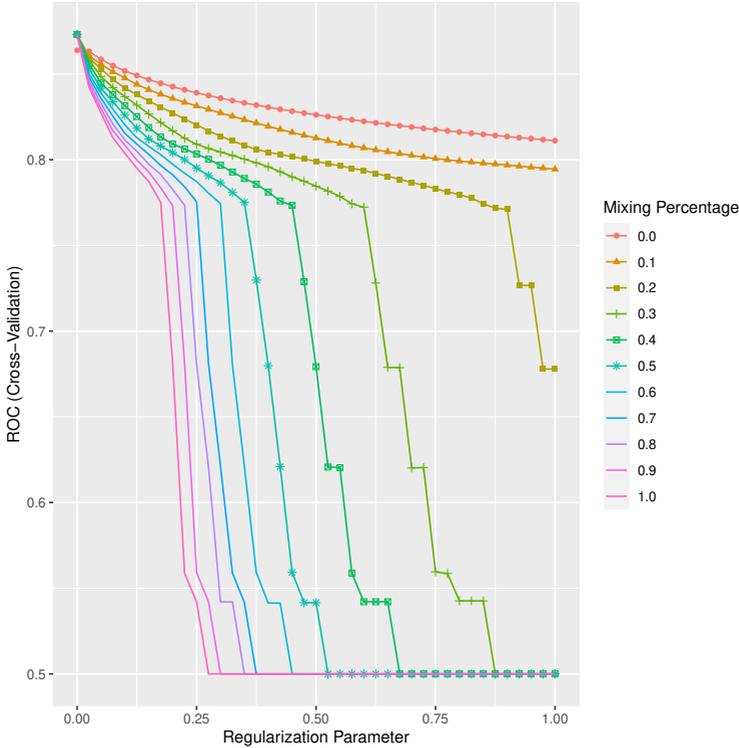
**Fig. 5.24** Hyperparameter tuning for spatial-block cross validation and kdbias background points. The objective function of the elastic net includes a penalty term defined by two hyperparameters named  $\alpha$  (“Mixing Percentage”) and  $\lambda$  (“Regularization Parameter”). The hyperparameter  $\alpha$  describes the mixing of the L1 and L2 penalties. If  $\alpha$  equals 1, the elastic net would essentially be a lasso regression whereas  $\alpha$  equal to 0 would result in a pure ridge regression. The hyperparameter  $\lambda$  denotes the degree of regularization employed. Both hyperparameters were tuned using a grid search to maximize the area under the ROC. For  $\alpha$ , values between 0 and 1 were searched at 0.1 increments and for  $\lambda$  values between 0 and 1 at 0.025 increments resulting in a total of 451 combinations. No regularization ( $\lambda = 0$ ) essentially collapses the elastic net into a standard generalized linear model with binomial distribution. Table 3 below depicts the optima, for each cross-validation technique and approach to generating background data.



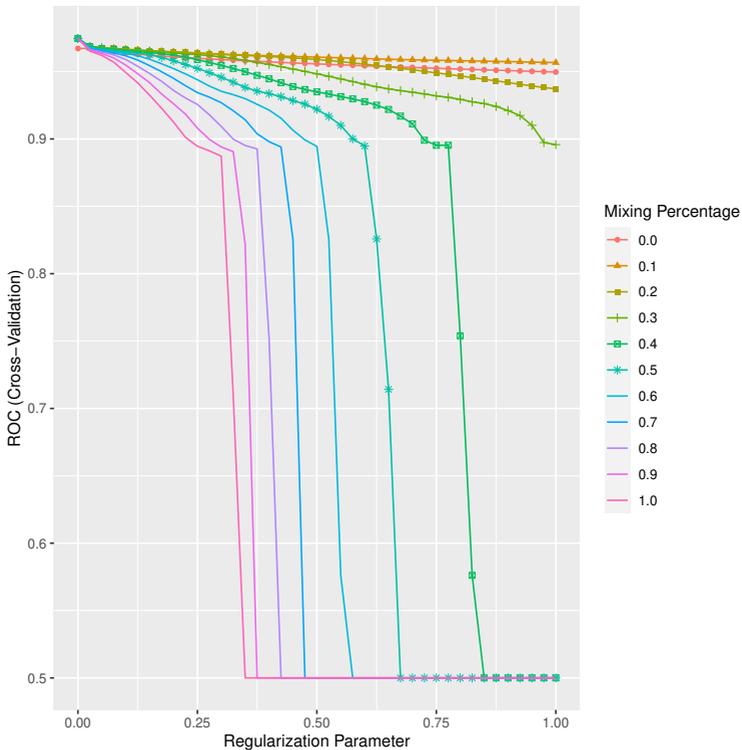
**Fig. 5.25** Hyperparameter tuning for spatial-block cross validation and random background points. The objective function of the elastic net includes a penalty term defined by two hyperparameters named  $\alpha$  (“Mixing Percentage”) and  $\lambda$  (“Regularization Parameter”). The hyperparameter  $\alpha$  describes the mixing of the L1 and L2 penalties. If  $\alpha$  equals 1, the elastic net would essentially be a lasso regression whereas  $\alpha$  equal to 0 would result in a pure ridge regression. The hyperparameter  $\lambda$  denotes the degree of regularization employed. Both hyperparameters were tuned using a grid search to maximize the area under the ROC. For  $\alpha$ , values between 0 and 1 were searched at 0.1 increments and for  $\lambda$  values between 0 and 1 at 0.025 increments resulting in a total of 451 combinations. No regularization ( $\lambda = 0$ ) essentially collapses the elastic net into a standard generalized linear model with binomial distribution. Table 3 below depicts the optima, for each cross-validation technique and approach to generating background data.



**Fig. 5.26** Hyperparameter tuning for temporal cross validation and kd05dfar background points. The objective function of the elastic net includes a penalty term defined by two hyperparameters named  $\alpha$  (“Mixing Percentage”) and  $\lambda$  (“Regularization Parameter”). The hyperparameter  $\alpha$  describes the mixing of the L1 and L2 penalties. If  $\alpha$  equals 1, the elastic net would essentially be a lasso regression whereas  $\alpha$  equal to 0 would result in a pure ridge regression. The hyperparameter  $\lambda$  denotes the degree of regularization employed. Both hyperparameters were tuned using a grid search to maximize the area under the ROC. For  $\alpha$ , values between 0 and 1 were searched at 0.1 increments and for  $\lambda$  values between 0 and 1 at 0.025 increments resulting in a total of 451 combinations. No regularization ( $\lambda = 0$ ) essentially collapses the elastic net into a standard generalized linear model with binomial distribution. Table 3 below depicts the optima, for each cross-validation technique and approach to generating background data.



**Fig. 5.27** Hyperparameter tuning for temporal cross validation and kdbias background points. The objective function of the elastic net includes a penalty term defined by two hyperparameters named  $\alpha$  (“Mixing Percentage”) and  $\lambda$  (“Regularization Parameter”). The hyperparameter  $\alpha$  describes the mixing of the L1 and L2 penalties. If  $\alpha$  equals 1, the elastic net would essentially be a lasso regression whereas  $\alpha$  equal to 0 would result in a pure ridge regression. The hyperparameter  $\lambda$  denotes the degree of regularization employed. Both hyperparameters were tuned using a grid search to maximize the area under the ROC. For  $\alpha$ , values between 0 and 1 were searched at 0.1 increments and for  $\lambda$  values between 0 and 1 at 0.025 increments resulting in a total of 451 combinations. No regularization ( $\lambda = 0$ ) essentially collapses the elastic net into a standard generalized linear model with binomial distribution. Table 3 below depicts the optima, for each cross-validation technique and approach to generating background data.



**Fig. 5.28** Hyperparameter tuning for temporal cross validation and random background points. The objective function of the elastic net includes a penalty term defined by two hyperparameters named  $\alpha$  (“Mixing Percentage”) and  $\lambda$  (“Regularization Parameter”). The hyperparameter  $\alpha$  describes the mixing of the L1 and L2 penalties. If  $\alpha$  equals 1, the elastic net would essentially be a lasso regression whereas  $\alpha$  equal to 0 would result in a pure ridge regression. The hyperparameter  $\lambda$  denotes the degree of regularization employed. Both hyperparameters were tuned using a grid search to maximize the area under the ROC. For  $\alpha$ , values between 0 and 1 were searched at 0.1 increments and for  $\lambda$  values between 0 and 1 at 0.025 increments resulting in a total of 451 combinations. No regularization ( $\lambda = 0$ ) essentially collapses the elastic net into a standard generalized linear model with binomial distribution. Table 3 below depicts the optima, for each cross-validation technique and approach to generating background data.

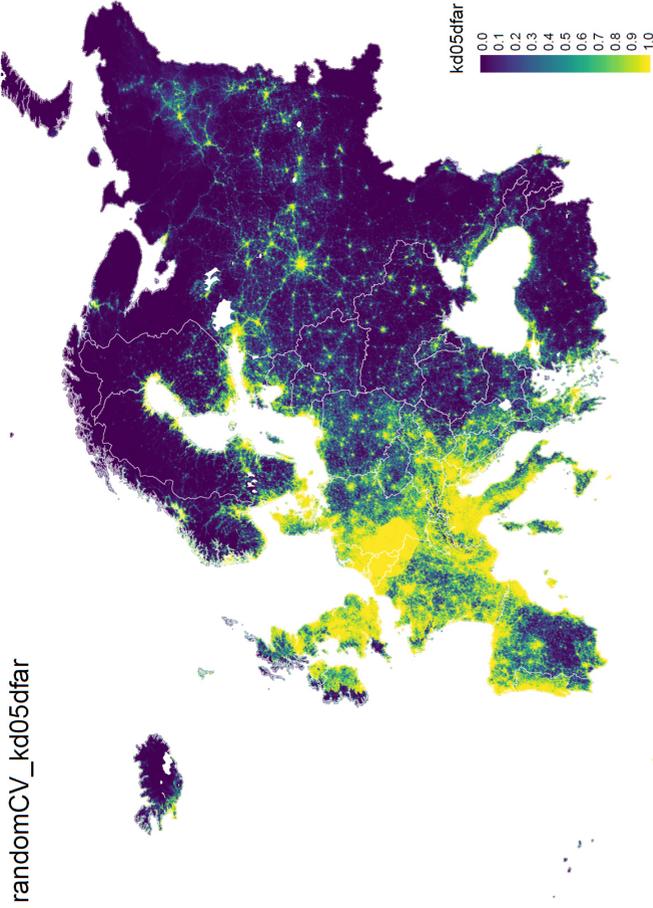
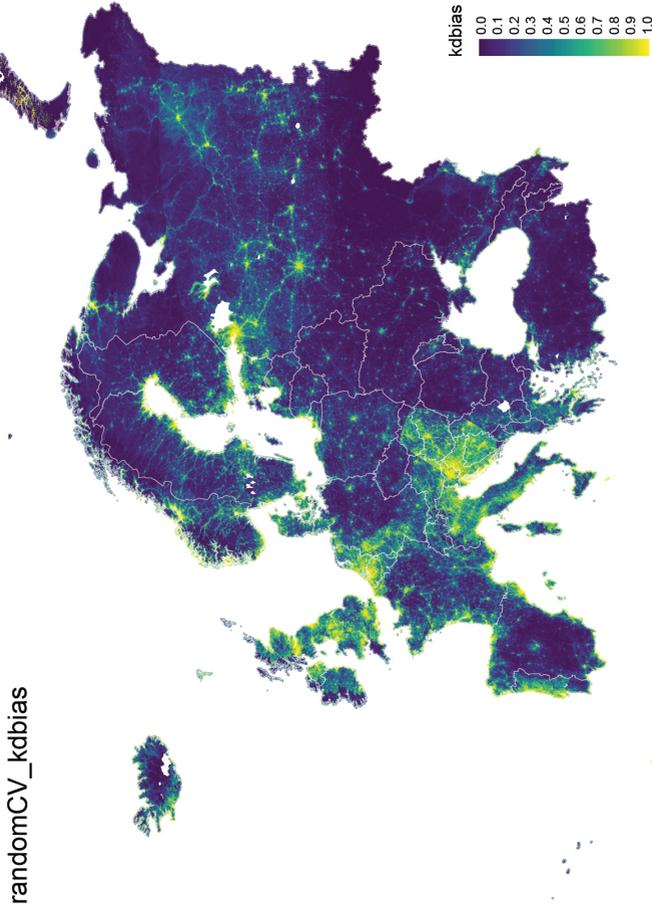


Fig. 5.29 Prediction for model tuned using random splits for cross-validation and kd05dfar background points.



**Fig. 5.30** Prediction for model tuned using random splits for cross-validation and kdbias background points.

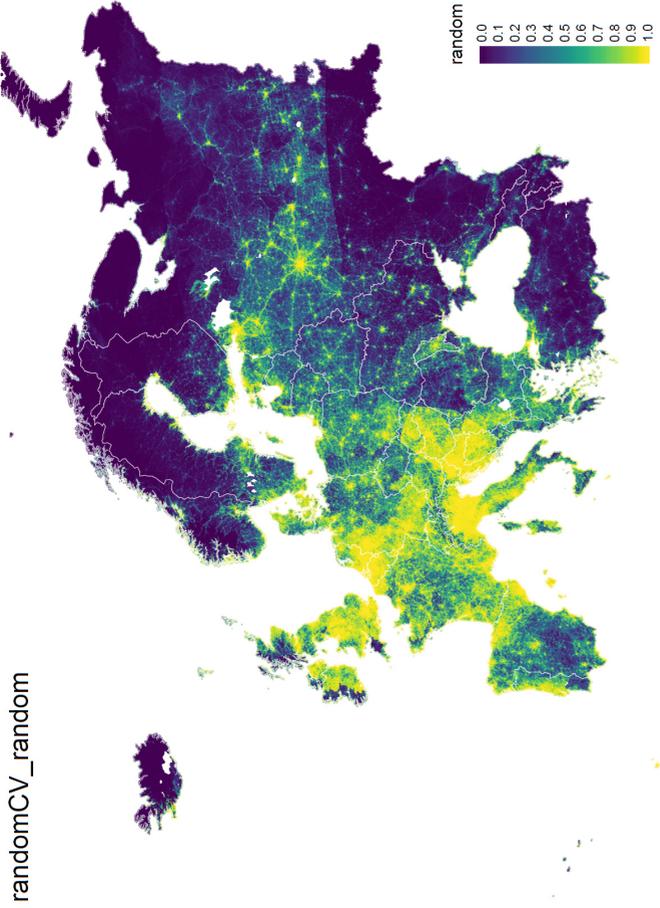
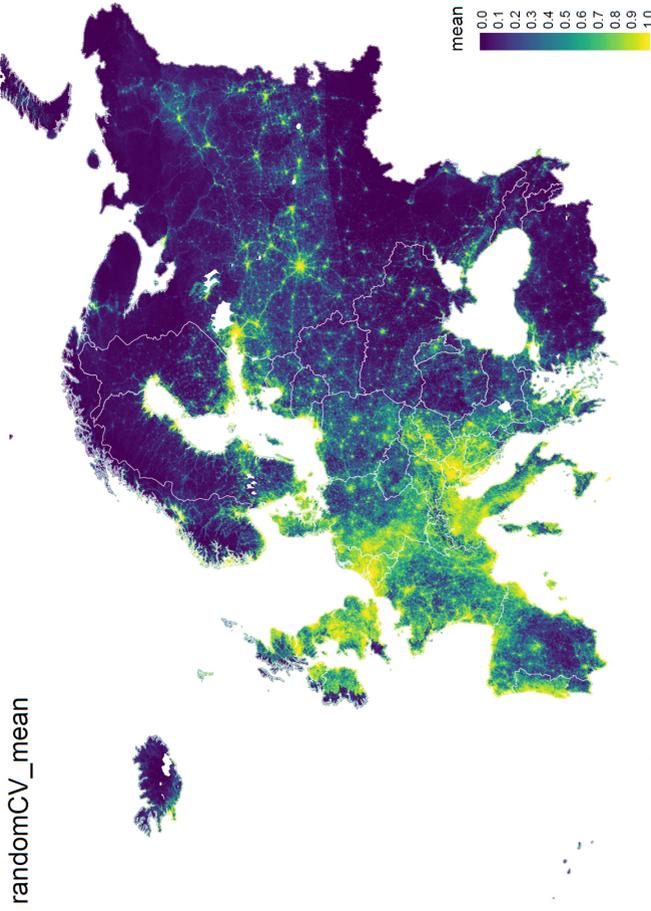
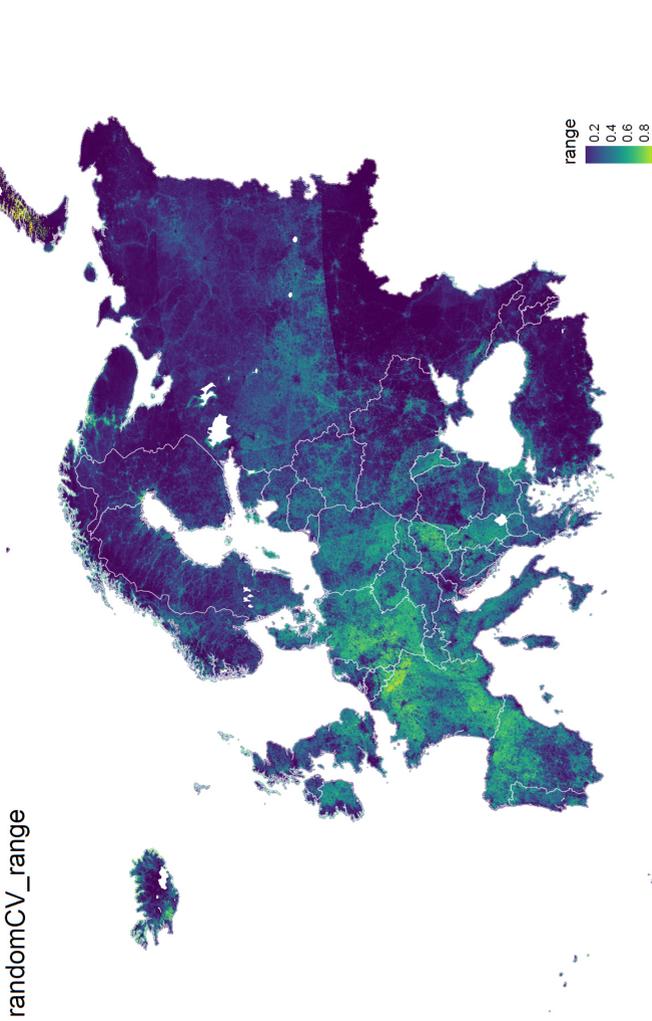


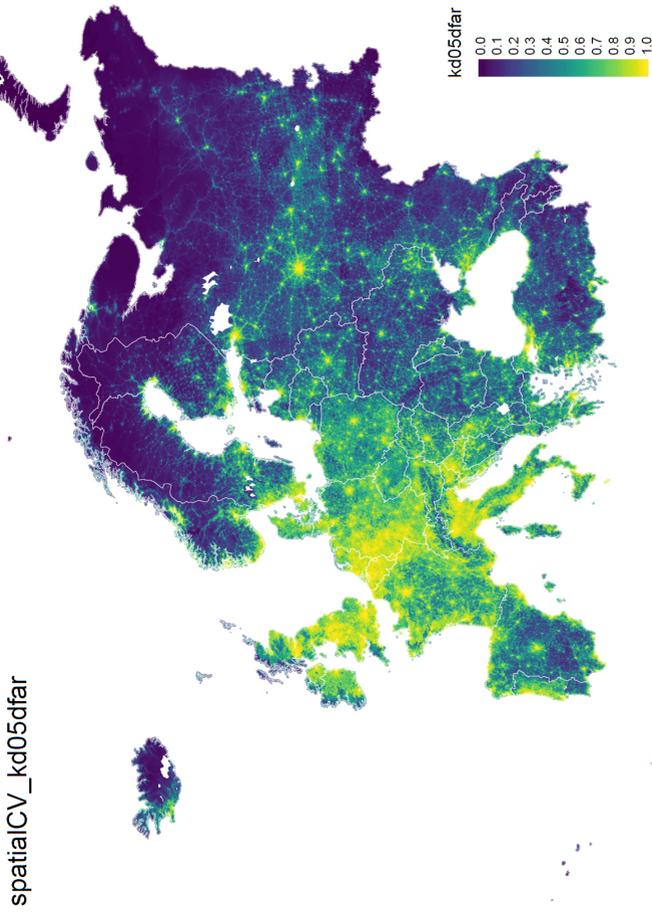
Fig. 5.31 Prediction for model tuned using random splits for cross-validation and random background points.



**Fig. 5.32** Average prediction for model tuned using random splits for cross-validation across all approaches to generate background points.



**Fig. 5.33** Range between maximum and minimum of the predictions for model tuned using random splits for cross-validation across all approaches to generate background points.



**Fig. 5.34** Prediction for model tuned using spatial-block splits for cross-validation and kd05dfar background points.

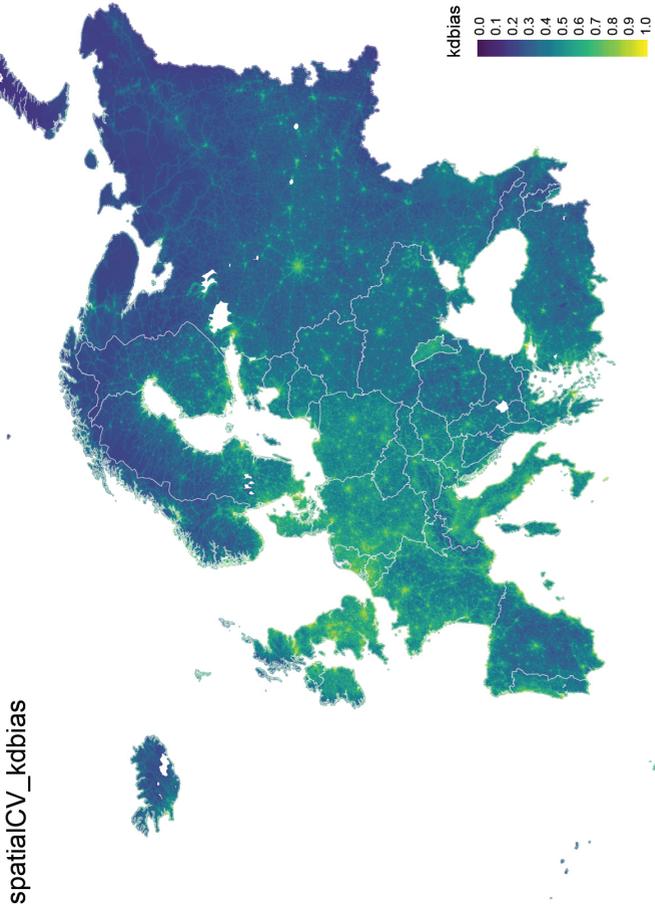
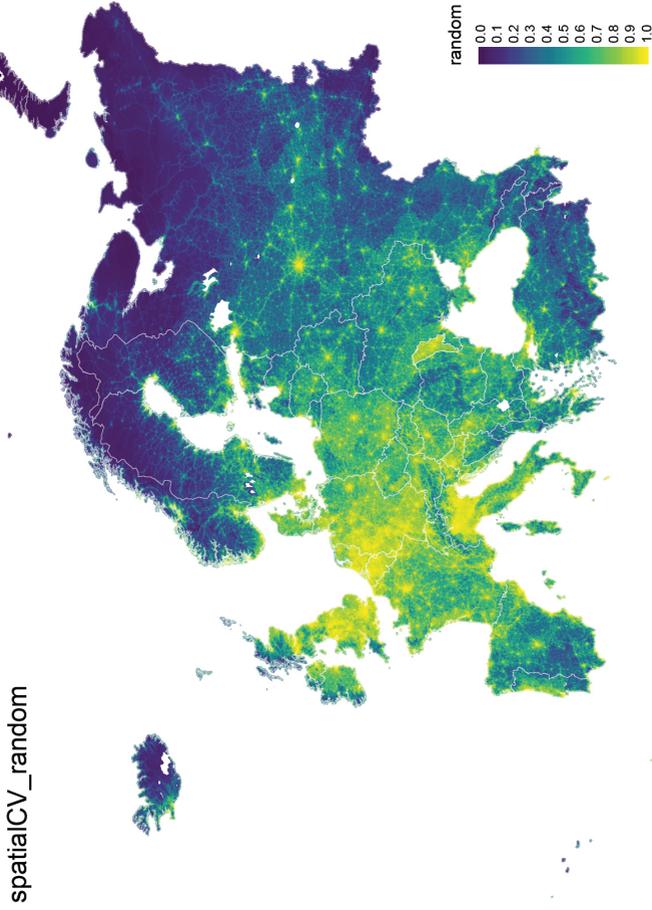
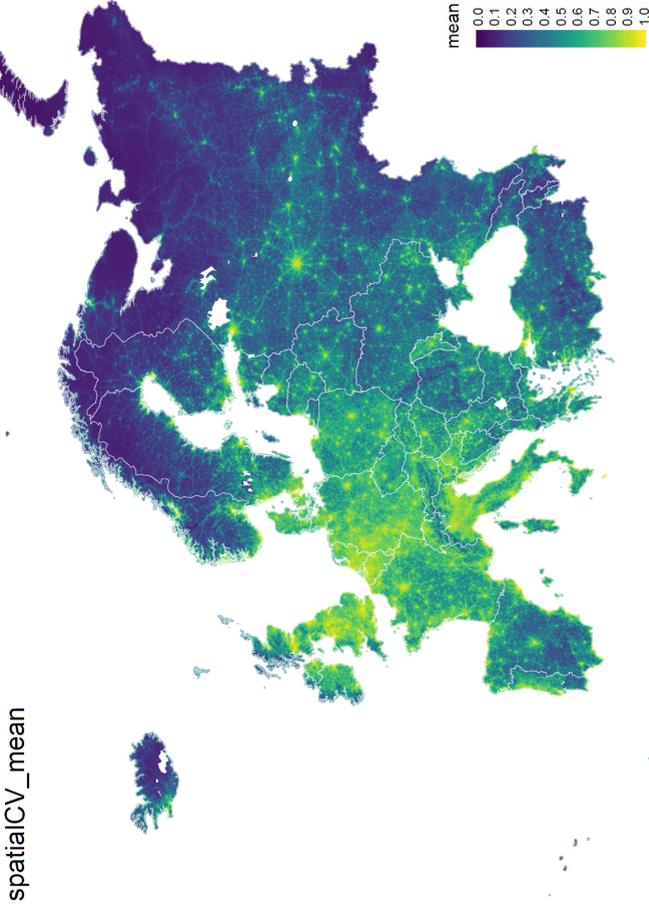


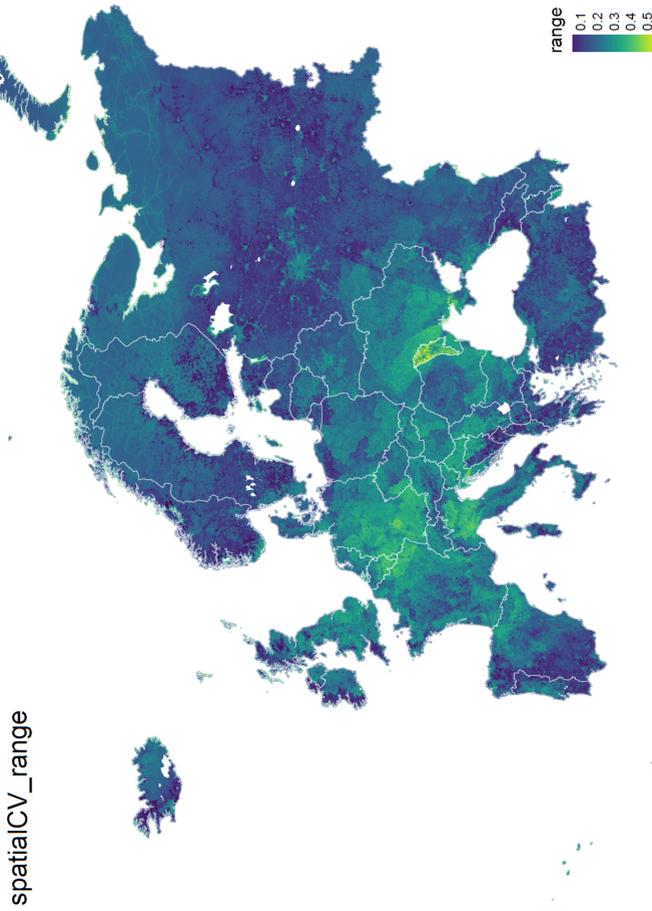
Fig. 5.35 Prediction for model tuned using spatial-block splits for cross-validation and kdbias background points.



**Fig. 5.36** Prediction for model tuned using spatial-block splits for cross-validation and random background points.



**Fig. 5.37** Average prediction for model tuned using spatial-block splits for cross-validation across all approaches to generate background points.



**Fig. 5.38** Range between maximum and minimum of the predictions for model tuned using spatial-block splits for cross-validation across all approaches to generate background points.

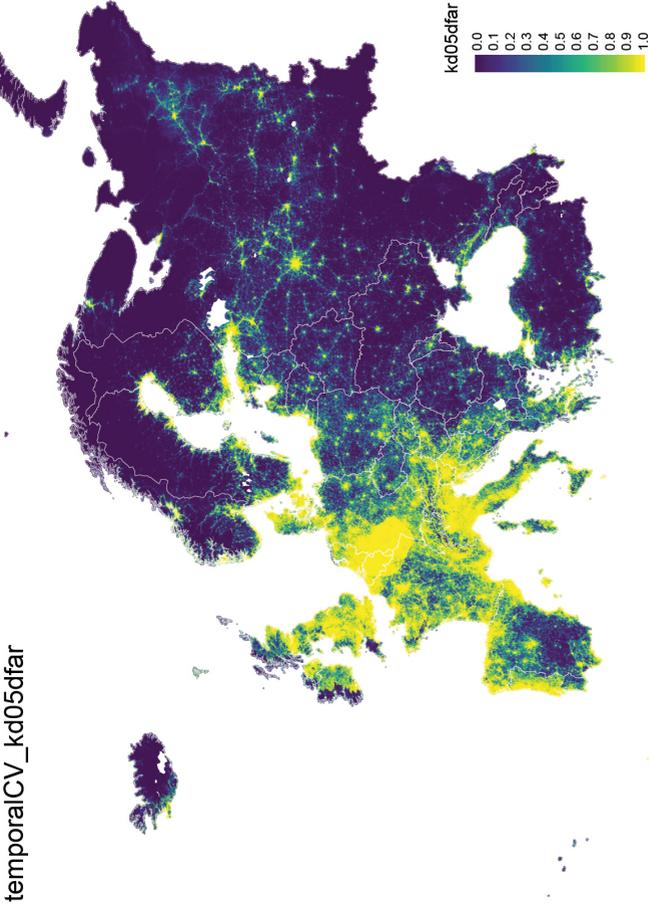
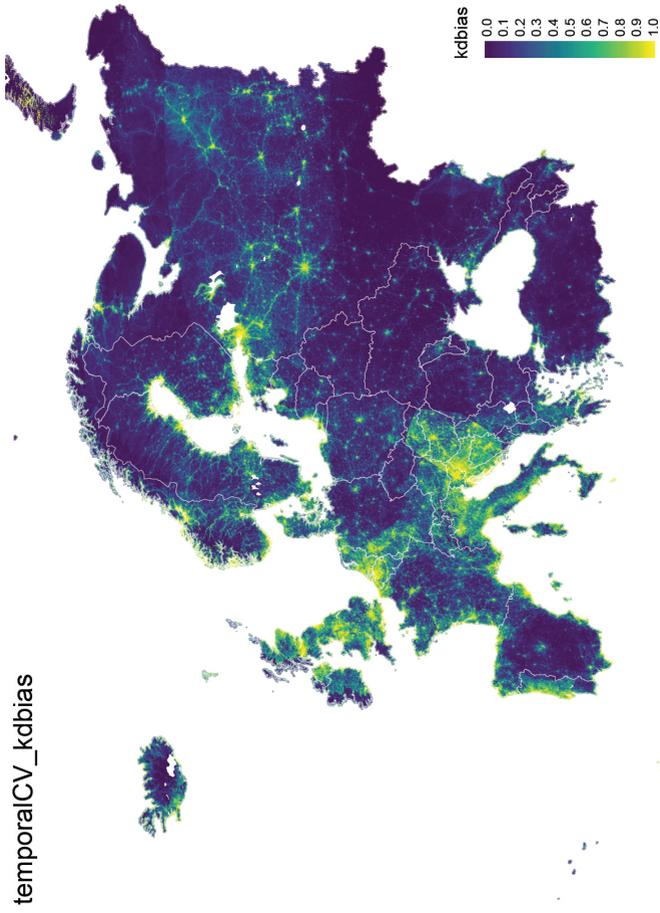


Fig. 5.39 Prediction for model tuned using temporal splits for cross-validation and kd05dfar background points.



**Fig. 5.40** Prediction for model tuned using temporal splits for cross-validation and kdbias background points.

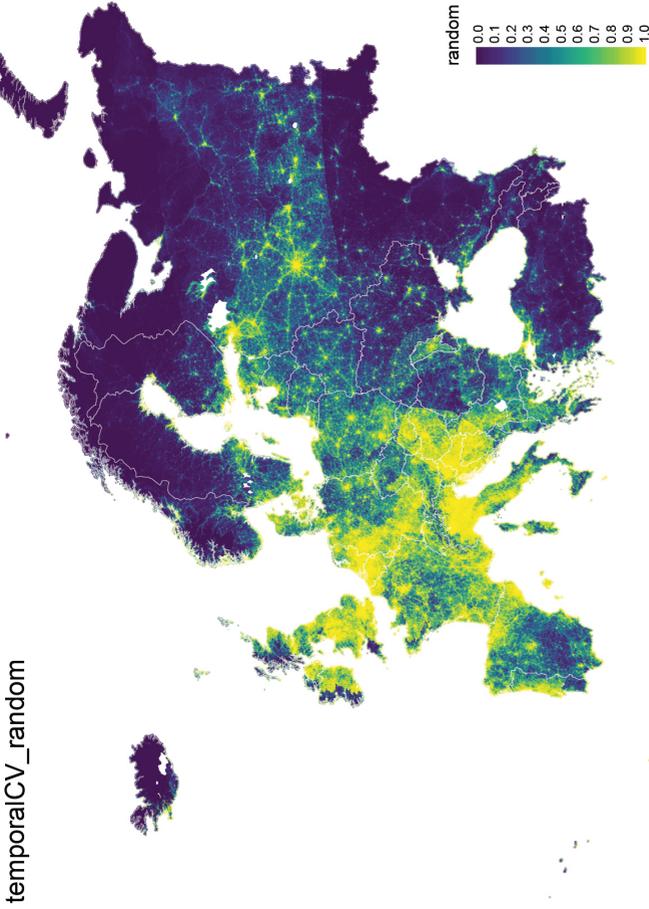
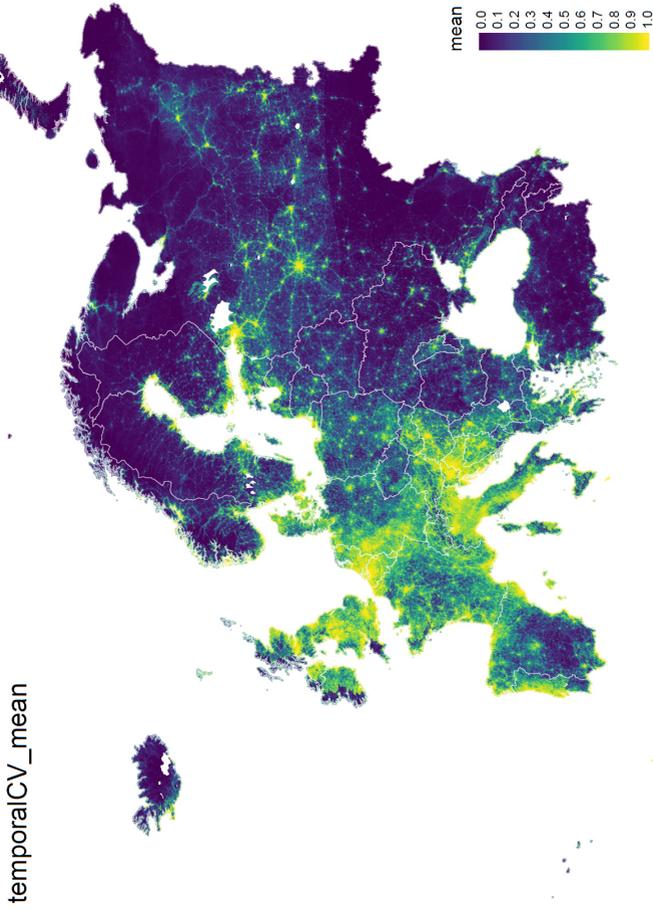
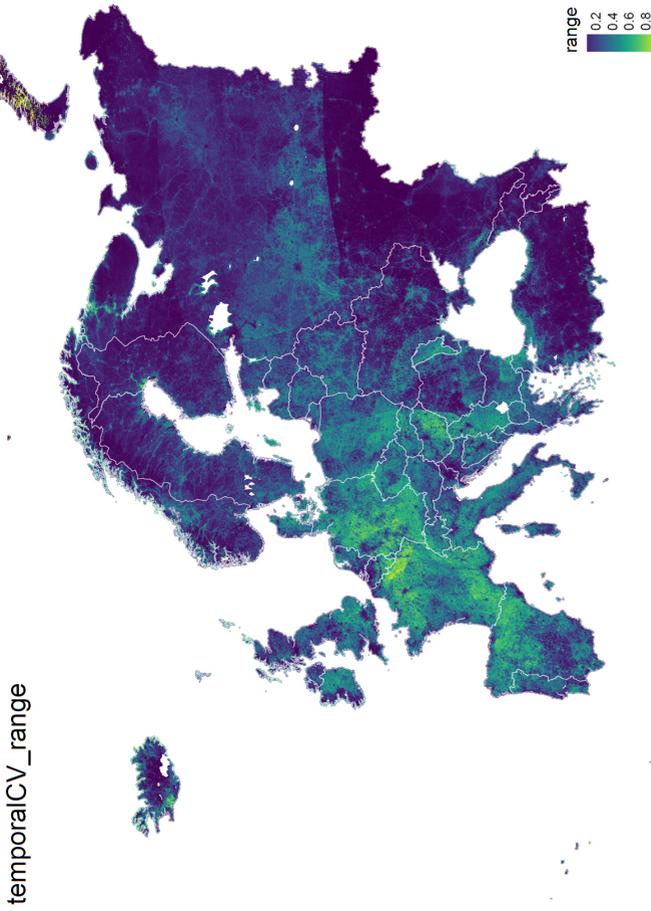


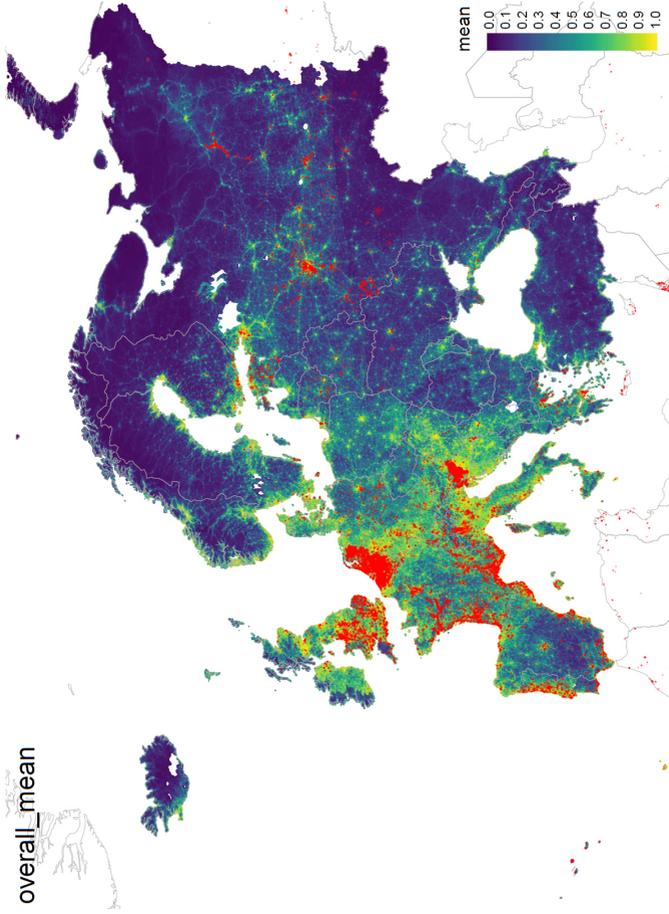
Fig. 5.41 Prediction for model tuned using temporalCV\_random and random background points.



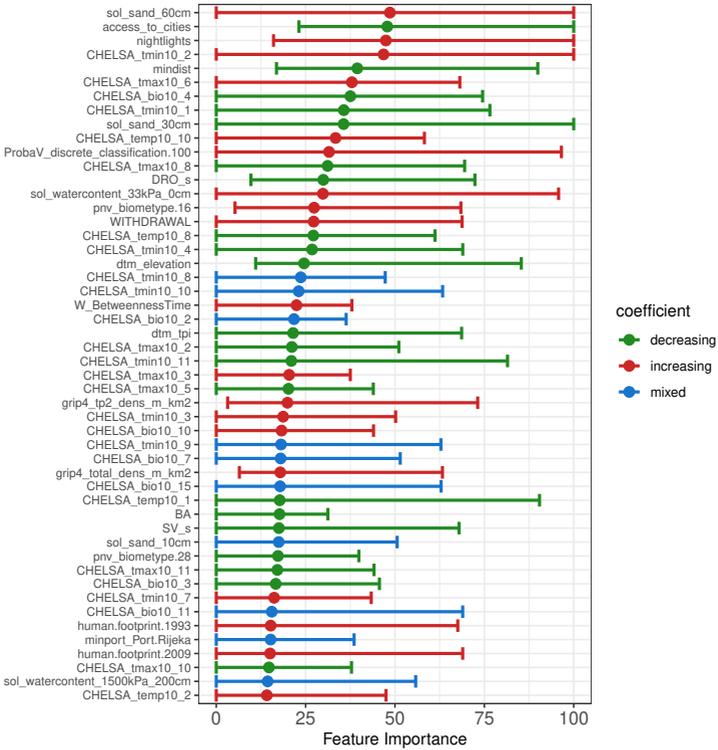
**Fig. 5.42** Average prediction for model tuned using temporal splits for cross-validation across all approaches to generate background points.



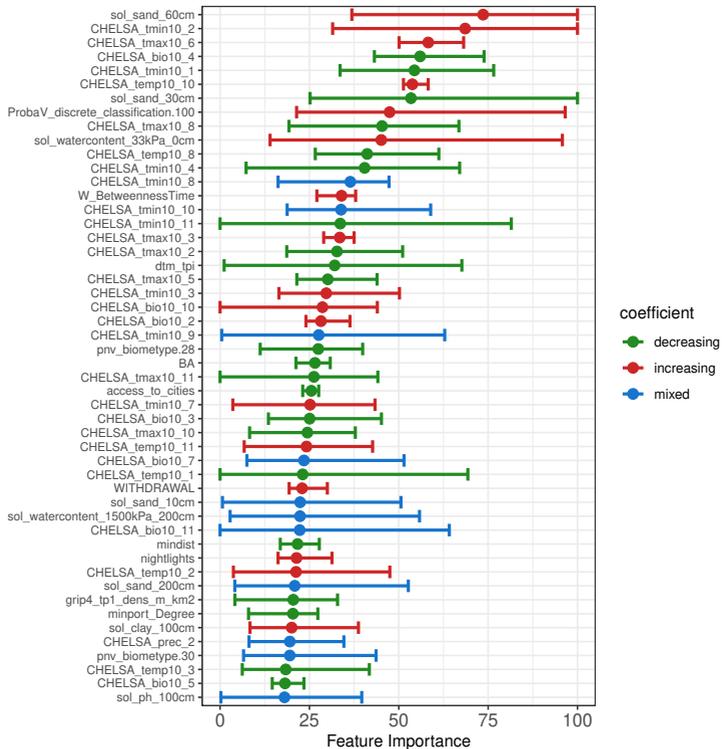
**Fig. 5.43** Range between maximum and minimum of the predictions for model tuned using temporal splits for cross-validation across all approaches to generate background points.



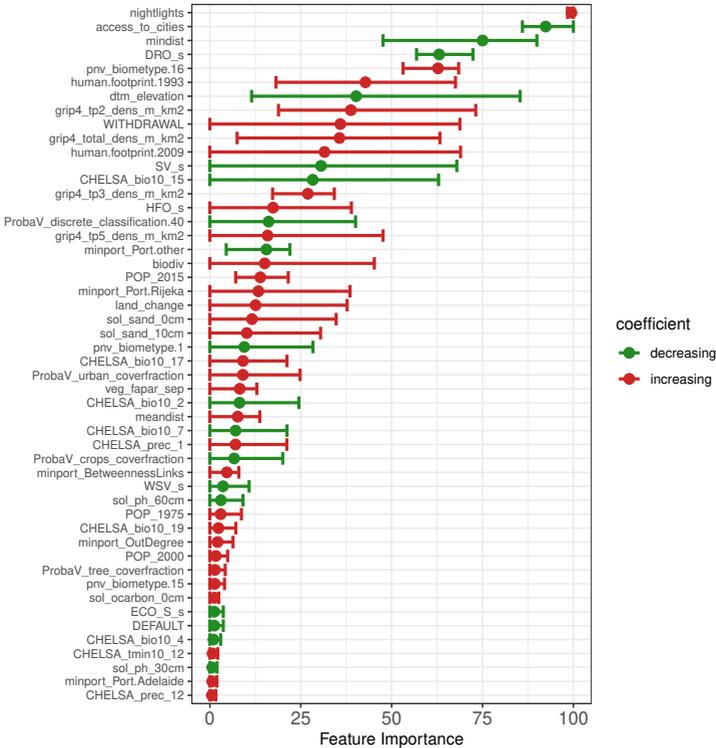
**Fig. 5.44** Average prediction across all nine approaches and GBIF presences (red).



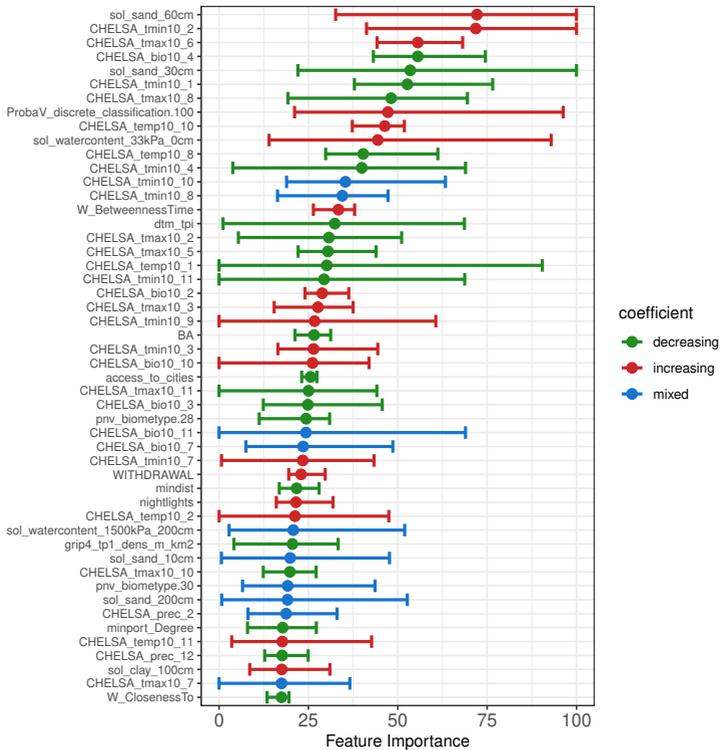
**Fig. 5.45** Fifty most important features across all approaches. The average is depicted as a dot. The line range shows the minimum and maximum values. If coefficients were consistently negative (i.e., risk decreasing) across all approaches, the graph is colored green. If coefficients were consistently positive (i.e., risk increasing) across all approaches, the graph is colored red. If the coefficient direction changed between the different approaches to generating background data, it is colored blue. Descriptions for all features can be found in Table 1 below.



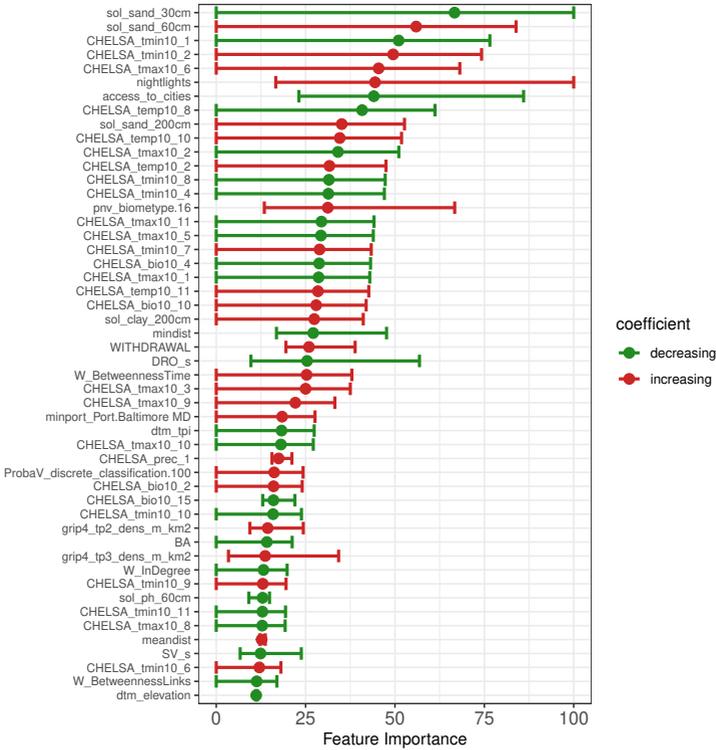
**Fig. 5.46** Fifty most important features across different approaches to generating background points for models tuned on random splits for cross validations. The average is depicted as a dot. The line range shows the minimum and maximum values. If coefficients were consistently negative (i.e., risk decreasing) across all approaches for background data generation, the graph is colored green. If coefficients were consistently positive (i.e., risk increasing) across all approaches, the graph is colored red. If the coefficient direction changed between the different approaches to generating background data, it is colored blue. Descriptions for all features can be found in Table 1 below.



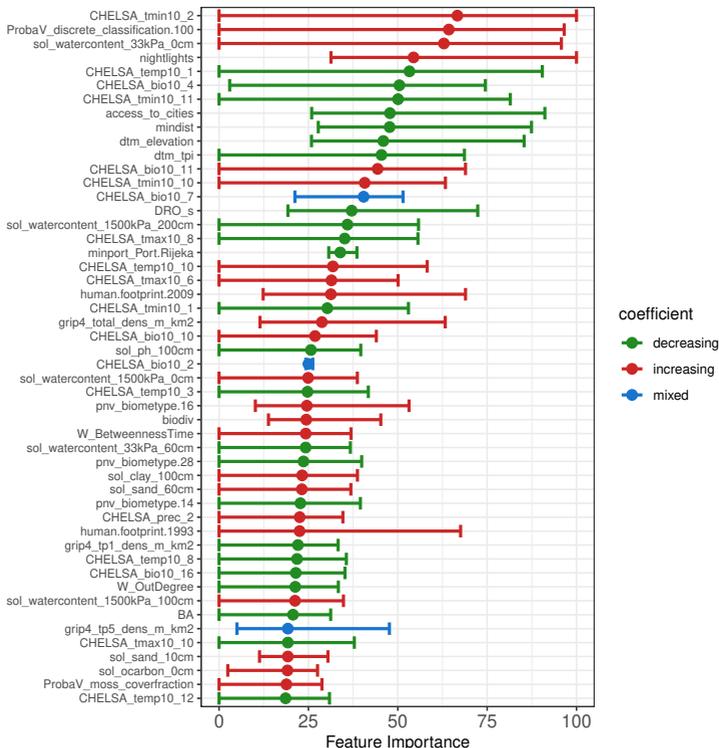
**Fig. 5.47** Fifty most important features across different approaches to generating background points for models tuned on spatial-block splits for cross validations. The average is depicted as a dot. The line range shows the minimum and maximum values. If coefficients were consistently negative (i.e., risk decreasing) across all approaches for background data generation, the graph is colored green. If coefficients were consistently positive (i.e., risk increasing) across all approaches, the graph is colored red. If the coefficient direction changed between the different approaches to generating background data, it is colored blue. Descriptions for all features can be found in Table 1 below.



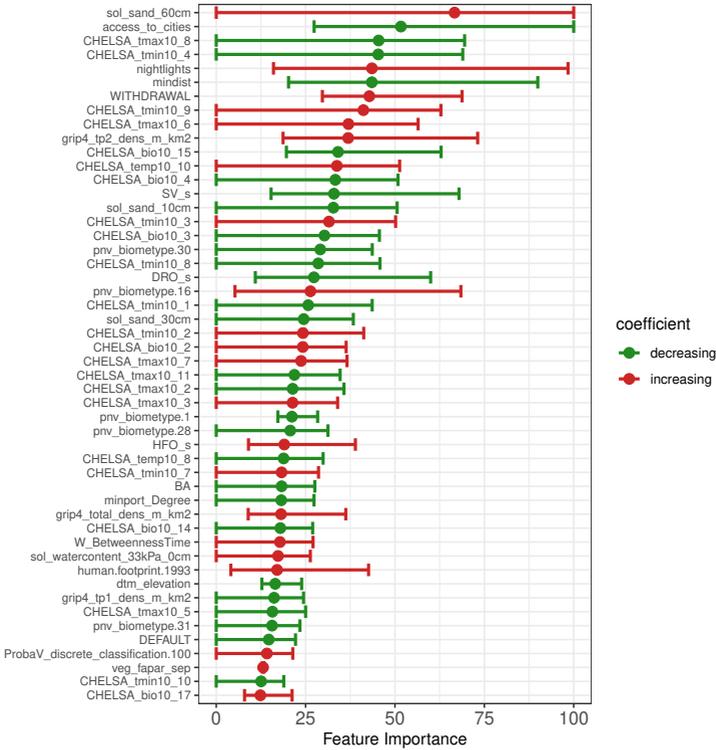
**Fig. 5.48** Fifty most important features across different approaches to generating background points for models tuned on temporal splits for cross validations. The average is depicted as a dot. The line range shows the minimum and maximum values. If coefficients were consistently negative (i.e., risk decreasing) across all approaches for background data generation, the graph is colored green. If coefficients were consistently positive (i.e., risk increasing) across all approaches, the graph is colored red. If the coefficient direction changed between the different approaches to generating background data, it is colored blue. Descriptions for all features can be found in Table 1 below.



**Fig. 5.49** Fifty most important features across different approaches to cross-validation for models trained on kd05dfar background points. The average is depicted as a dot. The line range shows the minimum and maximum values. If coefficients were consistently negative (i.e., risk decreasing) across all approaches, the graph is colored green. If coefficients were consistently positive (i.e., risk increasing) across all approaches, the graph is colored red. If the coefficient direction changed between the different approaches to generating background data, it is colored blue. Descriptions for all features can be found in Table 1 below.



**Fig. 5.50** Fifty most important features across different approaches to cross-validation for models trained on kdbias background points. The average is depicted as a dot. The line range shows the minimum and maximum values. If coefficients were consistently negative (i.e., risk decreasing) across all approaches, the graph is colored green. If coefficients were consistently positive (i.e., risk increasing) across all approaches, the graph is colored red. If the coefficient direction changed between the different approaches to generating background data, it is colored blue. Descriptions for all features can be found in Table 1 below.



**Fig. 5.51** Fifty most important features across different approaches to cross-validation for models trained on random background points. The average is depicted as a dot. The line range shows the minimum and maximum values. If coefficients were consistently negative (i.e., risk decreasing) across all approaches, the graph is colored green. If coefficients were consistently positive (i.e., risk increasing) across all approaches, the graph is colored red. Descriptions for all features can be found in Table 1 below.

Table 5.2: Overview of features.

<b>Variable name</b>	<b>Description</b>
access_to_cities	Accessibility to cities in driving time
biodiv	Index of biodiversity intactness
CHELSA_bio10_1	Annual Mean Temperature 1979-2013
CHELSA_bio10_10	Mean Temperature of Warmest Quarter 1979-2013
CHELSA_bio10_11	Mean Temperature of Coldest Quarter 1979-2013
CHELSA_bio10_12	Annual Precipitation 1979-2013
CHELSA_bio10_13	Precipitation of Wettest Month 1979-2013
CHELSA_bio10_14	Precipitation of Driest Month 1979-2013
CHELSA_bio10_15	Precipitation Seasonality 1979-2013
CHELSA_bio10_16	Precipitation of Wettest Quarter 1979-2013
CHELSA_bio10_17	Precipitation of Driest Quarter 1979-2013
CHELSA_bio10_18	Precipitation of Warmest Quarter 1979-2013
CHELSA_bio10_19	Precipitation of Coldest Quarter 1979-2013
CHELSA_bio10_2	Mean Diurnal Range 1979-2013
CHELSA_bio10_3	Isothermality 1979-2013
CHELSA_bio10_4	Temperature Seasonality 1979-2013
CHELSA_bio10_5	Max Temperature of Warmest Month 1979-2013
CHELSA_bio10_6	Min Temperature of Coldest Month 1979-2013
CHELSA_bio10_7	Temperature Annual Range 1979-2013
CHELSA_bio10_8	Mean Temperature of Wettest Quarter 1979-2013

*Continued on next page*

Table 5.2 - *Continued from previous page*

<b>Variable name</b>	<b>Description</b>
CHELSA_biol0_9	Mean Temperature of Driest Quarter 1979-2013
CHELSA_prec_1	Average precipitation January 1979-2013
CHELSA_prec_10	Average precipitation October 1979-2013
CHELSA_prec_11	Average precipitation November 1979-2013
CHELSA_prec_12	Average precipitation December 1979-2013
CHELSA_prec_2	Average precipitation February 1979-2013
CHELSA_prec_3	Average precipitation March 1979-2013
CHELSA_prec_4	Average precipitation April 1979-2013
CHELSA_prec_5	Average precipitation May 1979-2013
CHELSA_prec_6	Average precipitation June 1979-2013
CHELSA_prec_7	Average precipitation July 1979-2013
CHELSA_prec_8	Average precipitation August 1979-2013
CHELSA_prec_9	Average precipitation September 1979-2013
CHELSA_temp10_1	Average temperature January 1979-2013
CHELSA_temp10_10	Average temperature October 1979-2013
CHELSA_temp10_11	Average temperature November 1979-2013
CHELSA_temp10_12	Average temperature December 1979-2013
CHELSA_temp10_2	Average temperature February 1979-2013

*Continued on next page*

Table 5.2 – *Continued from previous page*

<b>Variable name</b>	<b>Description</b>
CHELSA_temp10_3	Average temperature March 1979-2013
CHELSA_temp10_4	Average temperature April 1979-2013
CHELSA_temp10_5	Average temperature May 1979-2013
CHELSA_temp10_6	Average temperature June 1979-2013
CHELSA_temp10_7	Average temperature July 1979-2013
CHELSA_temp10_8	Average temperature August 1979-2013
CHELSA_temp10_9	Average temperature September 1979-2013
CHELSA_tmax10_1	Max temperature January 1979-2013
CHELSA_tmax10_10	Max temperature October 1979-2013
CHELSA_tmax10_11	Max temperature November 1979-2013
CHELSA_tmax10_12	Max temperature December 1979-2013
CHELSA_tmax10_2	Max temperature February 1979-2013
CHELSA_tmax10_3	Max temperature March 1979-2013
CHELSA_tmax10_4	Max temperature April 1979-2013
CHELSA_tmax10_5	Max temperature May 1979-2013
CHELSA_tmax10_6	Max temperature June 1979-2013
CHELSA_tmax10_7	Max temperature July 1979-2013
CHELSA_tmax10_8	Max temperature August 1979-2013
CHELSA_tmax10_9	Max temperature September 1979-2013
CHELSA_tmin10_1	Min temperature January 1979-2013
CHELSA_tmin10_10	Min temperature October 1979-2013

*Continued on next page*

Table 5.2 - *Continued from previous page*

<b>Variable name</b>	<b>Description</b>
CHELSA_tmin10_11	Min temperature November 1979-2013
CHELSA_tmin10_12	Min temperature December 1979-2013
CHELSA_tmin10_2	Min temperature February 1979-2013
CHELSA_tmin10_3	Min temperature March 1979-2013
CHELSA_tmin10_4	Min temperature April 1979-2013
CHELSA_tmin10_5	Min temperature May 1979-2013
CHELSA_tmin10_6	Min temperature June 1979-2013
CHELSA_tmin10_7	Min temperature July 1979-2013
CHELSA_tmin10_8	Min temperature August 1979-2013
CHELSA_tmin10_9	Min temperature September 1979-2013
human.footprint.1993	Index of human pressure on the environment in 1993
human.footprint.2009	Index of human pressure on the environment in 2009
land_change	Global human modification of terrestrial systems
ProbaV_bare_coverfraction	Bare soil coverfraction
ProbaV_crops_coverfraction	Crop coverfraction
ProbaV_grass_coverfraction	Grass coverfraction
ProbaV_moss_coverfraction	Moss coverfraction
ProbaV_tree_coverfraction	Tree coverfraction
ProbaV_urban_coverfraction	Urban coverfraction
dtm_curvature	Profile curvature
dtm_downslope_curvature	Downslope curvature
dtm_dvm	Digital terrain models
dtm_dvm2	Digital terrain models
dtm_elevation	Elevation
dtm_negopeness	Negative openness
dtm_slope	Slope
dtm_tpi	Topographic position index
dtm_vbf	Valley Bottom Flatness
sol_clay_0cm	Clay content (%) at 0cm depth

*Continued on next page*

Table 5.2 – *Continued from previous page*

<b>Variable name</b>	<b>Description</b>
sol_clay_100cm	Clay content (%) at 100cm depth
sol_clay_10cm	Clay content (%) at 10cm depth
sol_clay_200cm	Clay content (%) at 200cm depth
sol_clay_30cm	Clay content (%) at 30cm depth
sol_clay_60cm	Clay content (%) at 60cm depth
sol_ocarbon_0cm	Organic carbon (kg/m <sup>2</sup> ) at 0cm depth
sol_ocarbon_100cm	Organic carbon (kg/m <sup>2</sup> ) at 100cm depth
sol_ocarbon_10cm	Organic carbon (kg/m <sup>2</sup> ) at 10cm depth
sol_ocarbon_200cm	Organic carbon (kg/m <sup>2</sup> ) at 200cm depth
sol_ocarbon_30cm	Organic carbon (kg/m <sup>2</sup> ) at 30cm depth
sol_ocarbon_60cm	Organic carbon (kg/m <sup>2</sup> ) at 60cm depth
sol_ph_0cm	Ph (in H <sub>2</sub> O) at 0cm depth
sol_ph_100cm	Ph (in H <sub>2</sub> O) at 100cm depth
sol_ph_10cm	Ph (in H <sub>2</sub> O) at 10cm depth
sol_ph_200cm	Ph (in H <sub>2</sub> O) at 200cm depth
sol_ph_30cm	Ph (in H <sub>2</sub> O) at 30cm depth
sol_ph_60cm	Ph (in H <sub>2</sub> O) at 60cm depth
sol_sand_0cm	Sand content (%) at 0cm depth
sol_sand_100cm	Sand content (%) at 100cm depth
sol_sand_10cm	Sand content (%) at 10cm depth
sol_sand_200cm	Sand content (%) at 200cm depth
sol_sand_30cm	Sand content (%) at 30cm depth
sol_sand_60cm	Sand content (%) at 60cm depth
sol_watercontent_1500kPa_0cm	Soil water content (at 1500 kPa) at 0cm depth
sol_watercontent_1500kPa_100cm	Soil water content (at 1500 kPa) at 100cm depth
sol_watercontent_1500kPa_10cm	Soil water content (at 1500 kPa) at 10cm depth

*Continued on next page*

Table 5.2 - *Continued from previous page*

<b>Variable name</b>	<b>Description</b>
sol_watercontent_1500kPa_200cm	Soil water content (at 1500 kPa) at 200cm depth
sol_watercontent_1500kPa_30cm	Soil water content (at 1500 kPa) at 30cm depth
sol_watercontent_1500kPa_60cm	Soil water content (at 1500 kPa) at 60cm depth
sol_watercontent_33kPa_0cm	Soil water content (at 33 kPa) at 0cm depth
sol_watercontent_33kPa_100cm	Soil water content (at 33 kPa) at 100cm depth
sol_watercontent_33kPa_10cm	Soil water content (at 33 kPa) at 10cm depth
sol_watercontent_33kPa_200cm	Soil water content (at 33 kPa) at 200cm depth
sol_watercontent_33kPa_30cm	Soil water content (at 33 kPa) at 30cm depth
sol_watercontent_33kPa_60cm	Soil water content (at 33 kPa) at 60cm depth
veg_fapar_apr	Average photosynthetically active radiation in April for 2014 to 2019
veg_fapar_aug	Average photosynthetically active radiation in August for 2014 to 2019
veg_fapar_dec	Average photosynthetically active radiation in December for 2014 to 2019
veg_fapar_feb	Average photosynthetically active radiation in February for 2014 to 2019
veg_fapar_jan	Average photosynthetically active radiation in January for 2014 to 2019
veg_fapar_jul	Average photosynthetically active radiation in July for 2014 to 2019
veg_fapar_jun	Average photosynthetically active radiation in June for 2014 to 2019

*Continued on next page*

Table 5.2 – *Continued from previous page*

<b>Variable name</b>	<b>Description</b>
veg_fapar_mar	Average photosynthetically active radiation in March for 2014 to 2019
veg_fapar_may	Average photosynthetically active radiation in May for 2014 to 2019
veg_fapar_nov	Average photosynthetically active radiation in November for 2014 to 2019
veg_fapar_oct	Average photosynthetically active radiation in October for 2014 to 2019
veg_fapar_sep	Average photosynthetically active radiation in September for 2014 to 2019
nightlights	Global radiance calibrated night-time lights (proxy for GDP)
POP_1975	Population density in 1975
POP_1990	Population density in 1990
POP_2000	Population density in 2000
POP_2015	Population density in 2015
W_PCInbound	Inverse-distance weighted port connectivity index for incoming shipments
W_PCOutbound	Inverse-distance weighted port connectivity index for outgoing shipments
W_Degree	Inverse-distance weighted degree of port within network
W_InDegree	Inverse-distance weighted network degree as importer
W_OutDegree	Inverse-distance weighted network degree as exporter
W_ClosenessFrom	Inverse-distance weighted reciprocal of sum of shortest distances from a port to all others

*Continued on next page*

Table 5.2 - *Continued from previous page*

<b>Variable name</b>	<b>Description</b>
W_ClosenessTo	Inverse-distance weighted reciprocal of sum of shortest distances from all other ports to this port
W_BetweennessLinks	Inverse-distance weighted distance to the two other ports if port is closest
W_BetweennessTime	Inverse-distance weighted number of times this port appears as shortest distance between two other ports
minport_PCInbound	Port connectivity index for incoming shipments for closest port
minport_PCOutbound	Port connectivity index for outgoing shipments for closest port
minport_Degree	Degree of port within network for closest port
minport_InDegree	Network degree as importer for closest port
minport_OutDegree	Network degree as exporter for closest port
minport_ClosenessFrom	Reciprocal of sum of shortest distances from a port to all others for closest port
minport_ClosenessTo	Reciprocal of sum of shortest distances from all other ports to this port for closest port
minport_BetweennessLinks	Distance to the two other ports if port is closest for closest port
minport_BetweennessTime	Number of times this port appears as shortest distance between two other ports for closest port
mindist	Minimum distance to any port
meandist	Mean distance to all ports
grip4_area_land_km2	m <sup>2</sup> land area per 5 arcminute cell
grip4_total_dens_m_km2	density for all roads, equally weighted

*Continued on next page*

Table 5.2 – *Continued from previous page*

<b>Variable name</b>	<b>Description</b>
grip4_tp1_dens_m_km2	density for GRIP type 1 - highways
grip4_tp2_dens_m_km2	density for GRIP type 2 - primary roads
grip4_tp3_dens_m_km2	density for GRIP type 3 - secondary roads
grip4_tp4_dens_m_km2	density for GRIP type 4 - tertiary roads
grip4_tp5_dens_m_km2	density for GRIP type 5 - local roads
WITHDRAWAL	Total withdrawal
BA	Available blue water
BWS_s	Baseline water stress indicator scores
WSV_s	Interannual water variability indicator scores
SV_s	Seasonal water variability indicator scores
HFO_s	Flood Occurrence indicator scores
DRO_s	Drought severity indicator scores
ECO_S_s	Upstream Protected Land indicator scores
MC_s	Media Coverage indicator scores
DEFAULT	Default weight overall water risk
veg_fapar_mean	Average photosynthetically active radiation for 2014 to 2019 across months
veg_fapar_sd	Standard deviation of photosynthetically active radiation for 2014 to 2019 across months
POP_change	Change in population density between 1995 to 2015
HF_change	Change in human footprint between 1993 to 2009
erosion.0	Erosion risk category 0
erosion.1	Erosion risk category 1
erosion.2	Erosion risk category 2

*Continued on next page*

Table 5.2 - *Continued from previous page*

<b>Variable name</b>	<b>Description</b>
erosion.3	Erosion risk category 3
erosion.4	Erosion risk category 4
erosion.5	Erosion risk category 5
ProbaV_discrete_classification.20	Landcover classification for shrubs
ProbaV_discrete_classification.30	Landcover classification for herbaceous vegetation
ProbaV_discrete_classification.40	Landcover classification for cultivated and managed cropland
ProbaV_discrete_classification.50	Landcover classification for urban areas
ProbaV_discrete_classification.60	Landcover classification for bare or sparse vegetation
ProbaV_discrete_classification.80	Landcover classification for permanent water bodies
ProbaV_discrete_classification.90	Landcover classification for herbaceous wetland
ProbaV_discrete_classification.100	Landcover classification for moss and lichen
ProbaV_discrete_classification.111	Landcover classification for closed evergreen needleleaf forests
ProbaV_discrete_classification.112	Landcover classification for closed evergreen broadleaf forests
ProbaV_discrete_classification.113	Landcover classification for closed deciduous needleleaf forests
ProbaV_discrete_classification.114	Landcover classification for closed deciduous broadleaf forests
ProbaV_discrete_classification.115	Landcover classification for closed mixed forests
ProbaV_discrete_classification.116	Landcover classification for closed unknown forests
ProbaV_discrete_classification.121	Landcover classification for open evergreen needle leaf forests
ProbaV_discrete_classification.122	Landcover classification for open evergreen broad leaf forests

*Continued on next page*

Table 5.2 – *Continued from previous page*

<b>Variable name</b>	<b>Description</b>
ProbaV_discrete_classification.124	Landcover classification for open deciduous broad leafforests
ProbaV_discrete_classification.125	Landcover classification for open mixed forests
ProbaV_discrete_classification.126	Landcover classification for open unknown forests
ProbaV_discrete_classification.other	Landcover classification for any other category
pnv_biometype.1	Tropical evergreen broadleaf forest
pnv_biometype.2	Tropical semi-evergreen broadleaf forest
pnv_biometype.3	Tropical deciduous broadleaf forest
pnv_biometype.4	Warm-temperate evergreen and mixed forest
pnv_biometype.7	Cool-temperate rainforest
pnv_biometype.8	Cool evergreen needleleaf forest
pnv_biometype.9	Cool mixed forest
pnv_biometype.13	Temperate deciduous broadleaf forest
pnv_biometype.14	Cold deciduous forest
pnv_biometype.15	Cold evergreen needleleaf forest
pnv_biometype.16	Temperate sclerophyll woodland and shrubland
pnv_biometype.17	Temperate evergreen needleleaf open woodland
pnv_biometype.18	Tropical savanna
pnv_biometype.20	Xerophytic woods and scrubs
pnv_biometype.22	Steppe
pnv_biometype.27	Desert
pnv_biometype.28	Graminoid and forb tundra
pnv_biometype.30	Erect dwarf shrub tundra
pnv_biometype.31	Low and high shrub tundra
minport_Port.Adelaide	Closest port is Adelaide
minport_Port.Antwerp	Closest port is Antwerp
minport_Port.	Closest port is Baltimore
Baltimore MD	

*Continued on next page*

Table 5.2 - *Continued from previous page*

<b>Variable name</b>	<b>Description</b>
minport_Port. Dutch Harbor AK	Closest port is Dutch Harbor
minport_Port. Houston TX	Closest port is Houston
minport_Port.Karachi	Closest port is Karachi
minport_Port.Kemi	Closest port is Kemi
minport_Port.Melbourne	Closest port is Melbourne
minport_Port.Mobile AL	Closest port is Mobile
minport_Port.Montreal	Closest port is Montreal
minport_Port. New York NY/NJ	Closest port is New York
minport_Port. Novorossiysk	Closest port is Novorossiysk
minport_Port. Prince Rupert BC	Closest port is Prince Rupert
minport_Port.Rijeka	Closest port is Rijeka
minport_Port. St Petersburg	Closest port is is Saint Petersburg
minport_Port.Sydney	Closest port is Sydney
minport_Port. Vostochniy	Closest port is Vostochniy
minport_Port. Xingang/Tianjin	Closest port is Xingang
minport_Port.Yingkou	Closest port is Yinkou
minport_Port.other	Closest port is other port not listed above

Table 5.3: Overview of raw data with spatial information.

<b>Description</b>	<b>Data type/resolution</b>
Name of species	Character string
Latitude	Decimal degrees
Longitude	Decimal degrees
Annual Mean Temperature 1979-2013	30 arc sec
Mean Temperature of Warmest Quarter 1979-2013	30 arc sec
Mean Temperature of Coldest Quarter 1979-2013	30 arc sec
Annual Precipitation 1979-2013	30 arc sec
Precipitation of Wettest Month 1979-2013	30 arc sec
Precipitation of Driest Month 1979-2013	30 arc sec
Precipitation Seasonality 1979-2013	30 arc sec
Precipitation of Wettest Quarter 1979-2013	30 arc sec
Precipitation of Driest Quarter 1979-2013	30 arc sec
Precipitation of Warmest Quarter 1979-2013	30 arc sec
Precipitation of Coldest Quarter 1979-2013	30 arc sec
Mean Diurnal Range 1979-2013	30 arc sec
Isothermality 1979-2013	30 arc sec
Temperature Seasonality 1979-2013	30 arc sec
Max Temperature of Warmest Month 1979-2013	30 arc sec
Min Temperature of Coldest Month 1979-2013	30 arc sec
Temperature Annual Range 1979-2013	30 arc sec
Mean Temperature of Wettest Quarter 1979-2013	30 arc sec
Mean Temperature of Driest Quarter 1979-2013	30 arc sec
Average precipitation January 1979-2013	30 arc sec
Average precipitation October 1979-2013	30 arc sec
Average precipitation November 1979-2013	30 arc sec
Average precipitation December 1979-2013	30 arc sec
Average precipitation February 1979-2013	30 arc sec
Average precipitation March 1979-2013	30 arc sec
Average precipitation April 1979-2013	30 arc sec
Average precipitation May 1979-2013	30 arc sec
Average precipitation June 1979-2013	30 arc sec
Average precipitation July 1979-2013	30 arc sec
Average precipitation August 1979-2013	30 arc sec

*Continued on next page*

Table 5.3 - *Continued from previous page*

Description	Data type/resolution
Average precipitation September 1979-2013	30 arc sec
Average temperature January 1979-2013	30 arc sec
Average temperature October 1979-2013	30 arc sec
Average temperature November 1979-2013	30 arc sec
Average temperature December 1979-2013	30 arc sec
Average temperature February 1979-2013	30 arc sec
Average temperature March 1979-2013	30 arc sec
Average temperature April 1979-2013	30 arc sec
Average temperature May 1979-2013	30 arc sec
Average temperature June 1979-2013	30 arc sec
Average temperature July 1979-2013	30 arc sec
Average temperature August 1979-2013	30 arc sec
Average temperature September 1979-2013	30 arc sec
Max temperature January 1979-2013	30 arc sec
Max temperature October 1979-2013	30 arc sec
Max temperature November 1979-2013	30 arc sec
Max temperature December 1979-2013	30 arc sec
Max temperature February 1979-2013	30 arc sec
Max temperature March 1979-2013	30 arc sec
Max temperature April 1979-2013	30 arc sec
Max temperature May 1979-2013	30 arc sec
Max temperature June 1979-2013	30 arc sec
Max temperature July 1979-2013	30 arc sec
Max temperature August 1979-2013	30 arc sec
Max temperature September 1979-2013	30 arc sec
Min temperature January 1979-2013	30 arc sec
Min temperature October 1979-2013	30 arc sec
Min temperature November 1979-2013	30 arc sec
Min temperature December 1979-2013	30 arc sec
Min temperature February 1979-2013	30 arc sec
Min temperature March 1979-2013	30 arc sec
Min temperature April 1979-2013	30 arc sec
Min temperature May 1979-2013	30 arc sec
Min temperature June 1979-2013	30 arc sec
Min temperature July 1979-2013	30 arc sec

*Continued on next page*

Table 5.3 – *Continued from previous page*

Description	Data type/resolution
Min temperature August 1979-2013	30 arc sec
Min temperature September 1979-2013	30 arc sec
Profile curvature	250 meter
Downslope curvature	250 meter
Digital terrain models DVM	250 meter
Digital terrain models DVM2	250 meter
Elevation	250 meter
Negative openness	250 meter
Slope	250 meter
Topographic position index	250 meter
Valley Bottom Flatness	250 meter
Predicted biome type	250 meter
Clay content (%) at 0cm depth	250 meter
Clay content (%) at 10cm depth	250 meter
Clay content (%) at 100cm depth	250 meter
Clay content (%) at 200cm depth	250 meter
Clay content (%) at 30cm depth	250 meter
Clay content (%) at 60cm depth	250 meter
Organic carbon (kg/m <sup>2</sup> ) at 0cm depth	250 meter
Organic carbon (kg/m <sup>2</sup> ) at 10cm depth	250 meter
Organic carbon (kg/m <sup>2</sup> ) at 100cm depth	250 meter
Organic carbon (kg/m <sup>2</sup> ) at 200cm depth	250 meter
Organic carbon (kg/m <sup>2</sup> ) at 30cm depth	250 meter
Organic carbon (kg/m <sup>2</sup> ) at 60cm depth	250 meter
Ph (in H <sub>2</sub> O) at 0cm depth	250 meter
Ph (in H <sub>2</sub> O) at 10cm depth	250 meter
Ph (in H <sub>2</sub> O) at 100cm depth	250 meter
Ph (in H <sub>2</sub> O) at 200cm depth	250 meter
Ph (in H <sub>2</sub> O) at 30cm depth	250 meter
Ph (in H <sub>2</sub> O) at 60cm depth	250 meter
Sand content (%) at 0cm depth	250 meter
Sand content (%) at 10cm depth	250 meter
Sand content (%) at 100cm depth	250 meter
Sand content (%) at 200cm depth	250 meter
Sand content (%) at 30cm depth	250 meter

*Continued on next page*

Table 5.3 – *Continued from previous page*

Description	Data type/resolution
Sand content (%) at 60cm depth	250 meter
Soil water content (at 1500 kPa) at 0cm depth	250 meter
Soil water content (at 1500 kPa) at 10cm depth	250 meter
Soil water content (at 1500 kPa) at 100cm depth	250 meter
Soil water content (at 1500 kPa) at 200cm depth	250 meter
Soil water content (at 1500 kPa) at 30cm depth	250 meter
Soil water content (at 1500 kPa) at 60cm depth	250 meter
Soil water content (at 33 kPa) at 0cm depth	250 meter
Soil water content (at 33 kPa) at 10cm depth	250 meter
Soil water content (at 33 kPa) at 100cm depth	250 meter
Soil water content (at 33 kPa) at 200cm depth	250 meter
Soil water content (at 33 kPa) at 30cm depth	250 meter
Soil water content (at 33 kPa) at 60cm depth	250 meter
Average photosynthetically active radiation in April for 2014 to 2019	250 meter
Average photosynthetically active radiation in August for 2014 to 2019	250 meter
Average photosynthetically active radiation in December for 2014 to 2019	250 meter
Average photosynthetically active radiation in February for 2014 to 2019	250 meter
Average photosynthetically active radiation in January for 2014 to 2019	250 meter
Average photosynthetically active radiation in July for 2014 to 2019	250 meter
Average photosynthetically active radiation in June for 2014 to 2019	250 meter
Average photosynthetically active radiation in March for 2014 to 2019	250 meter

*Continued on next page*

Table 5.3 – *Continued from previous page*

<b>Description</b>	<b>Data type/resolution</b>
Average photosynthetically active radiation in May for 2014 to 2019	250 meter
Average photosynthetically active radiation in November for 2014 to 2019	250 meter
Average photosynthetically active radiation in October for 2014 to 2019	250 meter
Average photosynthetically active radiation in September for 2014 to 2019	250 meter
Bare soil coverfraction	100 meter
Crop coverfraction	100 meter
Classification of land uses	100 meter
Grass coverfraction	100 meter
Moss coverfraction	100 meter
Tree coverfraction	100 meter
Urban coverfraction	100 meter
Erosion risk	15 arc sec
Total withdrawal (m3)	Spatial polygons
Available blue water (m3)	Spatial polygons
Baseline water stress indicator scores (0-5)	Spatial polygons
Interannual Variability indicator scores (0-5)	Spatial polygons
Seasonal variability indicator scores (0-5)	Spatial polygons
Flood Occurrence indicator scores (0-5)	Spatial polygons
Drought severity indicator scores (0-5)	Spatial polygons
Upstream Protected Land indicator scores (0-5)	Spatial polygons
Media Coverage indicator scores (0-5)	Spatial polygons
Default weight overall water risk	Spatial polygons
Biodiversity intactness	1 kilometer
Population density in 1975	250m
Population density in 1990	250m
Population density in 2000	250m
Population density in 2015	250m
m2 land area per 5 arcminute cell	5 arc minutes
density for all roads, equally weighted	5 arc minutes
density for GRIP type 1 - highways	5 arc minutes
density for GRIP type 2 - primary roads	5 arc minutes

*Continued on next page*

Table 5.3 - *Continued from previous page*

Description	Data type/resolution
density for GRIP type 3 - secondary roads	5 arc minutes
density for GRIP type 4 - tertiary roads	5 arc minutes
density for GRIP type 5 - local roads	5 arc minutes
Human pressure on the environment in 1993 (index)	1 kilometer
Human pressure on the environment in 2009 (index)	1 kilometer
Ports' Latitude	Decimal degrees
Ports' Longitude	Decimal degrees
Port name	Character string
Port connectivity index for incoming shipments	Numeric
Port connectivity index for outgoing shipments	Numeric
Degree of port within network	Numeric
Network degree as importer	Numeric
Network degree as exporter	Numeric
Reciprocal of sum of shortest distances from a port to all others	Numeric
Reciprocal of sum of shortest distances from all other ports to this port	Numeric
Distance to the two other ports if port is closest	Numeric
Number of times this port appears as shortest distance between two other ports	Numeric
Global human modification of terrestrial systems	1 kilometer
Accessibility to cities in driving time	1 kilometer
Global radiance calibrated nighttime lights	30 arc sec

**Table 5.4** Overview of tuned hyperparameters for all approaches to cross validation and the generation of background points.

<b>Cross-Validation</b>	<b>Background points</b>	$\alpha$	$\lambda$
random	random	0.90	0.00
random	kdbias	0.40	0.00
random	kd05dfar	0.90	0.00
spatial blocks	random	0.40	0.05
spatial blocks	kdbias	0.10	0.40
spatial blocks	kd05dfar	0.70	0.03
temporal	random	1.00	0.00
temporal	kdbias	0.70	0.00
temporal	kd05dfar	0.90	0.00



## Chapter 6

# General Discussion

The thesis provided four research chapters that addressed distinct yet connected objectives. The methodologies used are diverse and, consequently, the research articles that resulted out of the work presented here contributed to several research streams. The contributions made were in part methodological and in part through critical insights obtained from applications to damage abatement input use, potential impacts from the invader *Xylella fastidiosa* subspecies *pauca*, and risk toward pest introduction across Europe.

Chapter 2 provided an empirical approach to measure spatial spillover effects of decision-making units' characteristics with managerial performance of neighbors. This allowed to relax the assumption that decision-making units operate in isolation from their peers. The study extended the widely used bootstrap truncated regression model to estimate the parameter of the spatial weight matrix. The steps proposed for the estimation of this parameter are transferable to other modelling work, possibly unrelated to efficiency measurement, in which a maximum likelihood estimator is used and a *spatial lag in X* (SLX) model employed. In addition, the application contributed to the ongoing discussion on spatial dependence in managerial performance. In contrast to previous work, the chapter stressed the benefits of measuring spatial spillovers for input- and output-specific inefficiencies separately, as opposed to using an overall efficiency score.

**The first research question was: Do neighbors' characteristics associate with farmers' managerial performance? I addressed**

**this question in chapter 2 and found that: Yes, neighbors' characteristics were found to associate with farmers' managerial performance. However, *how* the characteristics associate with farmers' performance was found to depend on the definition of the neighborhood.**

Chapter 3 provides an integrated framework that derives insights from climatic suitability, spread modelling, and economic modelling. Building on previous work, we computed an ensemble climatic suitability model based on 10 machine learning algorithms. Furthermore, we extended the radial range expansion model to (i) take the climatic suitability into account within the dispersal process, (ii) use a mixed neighborhood structure for the pixel-to-pixel transmission to result in a radial spread pattern despite the underlying square grids, (iii) include host presence, and (iv) capture uncertainty in the point of introduction via stochastic simulations using a High-Performance Computing Cluster. The economic model showed that (a) losses in investments of perennial hosts can be captured through foregone annuities, (b) heterogeneity among different cropping systems can be integrated using regional statistics on host-density and irrigation, (c) net-present value models can be extended to include price responses, and (d) worst- and best-case economic scenarios allow to bracket a range of impacts that can provide information on the value of mitigation strategies. In terms of application, the chapter has produced crucial information for risk managers, scientists, and the public on the potential economic impact further spread of *Xylella fastidiosa* subspecies *pauca* may cause to European olive growers.

**The second research question was: What are the potential economic impacts from *Xylella fastidiosa* subspecies *pauca* to European olive grower? I addressed this question in chapter 3 and found that: For Italy, across the considered spread rates the potential economic impact over 50 years ranged from 1.86 to 5.17 billion Euro for the economic worst-case scenario in which production ceases after orchards die off. If replanting with resistant varieties is feasible, the impact ranged from 0.59 to 1.57 billion Euro. For Greece, across the considered spread rates the potential economic impact over 50 years ranged from 0.21 to 1.94 billion, if replanting is not feasible, and 0.09 to 0.58 billion, if replanting is feasible. For Spain, across the considered spread**

**rates the potential economic impact over 50 years ranged from 0.71 to 16.86 billion, if replanting is not feasible, and 0.39 to 4.98 billion, if replanting is feasible. Even under slow spread rates and the ability to replant with resistant cultivars, economic impact from further spread of *Xylella fastidiosa* subspecies *pauca* was sizable and warrants regulatory response.**

Chapter 4 has contributed to the literature by illustrating how spatially explicit pest spread models can be translated to suit the needs of partial equilibrium models. Furthermore, the chapter has shown how global sensitivity analyses can be used in the context of partial equilibrium models. The application highlighted that the largest share of the potential future impact from *Xylella fastidiosa* subspecies *pauca* in the olive oil market would fall on consumers because of higher prices following reductions in supply due to the invasion. This redistribution of impact to consumers is expected to be particularly severe due to the inelastic nature of the market. The analysis highlights that the problem of invasive pests should be contextualized as a societal challenge as opposed to one that affects only producers. The chapter stressed the fact that consumers are beneficiaries of pest control. I argued that this must be communicated to the public to raise awareness that invasive pest control efforts benefit citizens.

**The third research question was: Who benefits and who loses from the control of *Xylella fastidiosa* subspecies *pauca*? I addressed this question in chapter 4 and found that: Due to the inelastic supply and demand, the main beneficiaries of control measures against *Xylella fastidiosa* subspecies *pauca* are consumers. Under disease spread, competitors in areas unlikely to be affected are worse off from control efforts.**

Chapter 5 has shown that a joint analysis of several hundred pests is able to produce hotspot maps with a high accuracy. This approach to mapping might be one way to address the overwhelming absence of information for many hazardous organisms which results out of the considerable time and labor requirements for species-specific analyses. By combining spatial data from various sources ranging climate, soil, water, and anthropogenic factors, the chapter has produced a solid base of georeferenced data that can easily be used for future work. The chapter has also contributed to discussions within the species distribution modelling literature on appropriate techniques for pseudo-absence

generation and cross-validation. In terms of application, the results presented in chapter 5 clearly stress the need to generate harmonized systematically collected presences. Furthermore, I highlighted the need to include true absences into pest surveys and make these data publicly available.

**The fourth research question was: Can joint analyses of various pests help identify weak-links and thereby support collective control? I addressed this question in chapter 5 and found that: Yes, joint analyses of various pests could help to identify weak-links and inform collective control. However, harmonized, systematic, species survey data comprising also true absences are required to further validate and improve these maps.**

The thesis, as a collection of these articles, has contributed to the literature by providing methodological approaches which (i) capture spatial dependencies, (ii) account for environmental and economic heterogeneity at the level of granularity feasible under the available data, (iii) acknowledge the temporal nature of pest spread and economic impact in perennial hosts, (iv) highlight the actionable insights sensitivity analyses can generate, and (v) propose cost-effective modelling strategies to address the absence of risk maps for many invasive species.

In the remainder of this chapter, I will first discuss implications of economic and spatial dependencies and heterogeneity for pest risk assessments and analyses of optimal pest control by placing the results of this thesis into the context of related research. In section 6.2, I will discuss opportunities and challenges for agri-environmental data integration which can be achieved through a proper spatial indexation of economic data. In section 6.3, I derive policy implications. In section 6.4, I discuss limitations of the work presented in this thesis and propose avenues for future work. Lastly, section 6.5 lists the main conclusions.

## 6.1 Implications of mutual dependence and heterogeneity

The foundational economic principle that demand and supply jointly determine price which in turn influences economic wellbeing of market participants is widely understood. The pricing mechanism is the link that creates mutual dependence in outcome among producers and between producers and consumers. The economic dependence via prices results in competition and strategic decisions of market participants [305]. Thereby, economic dependence among individuals lays the foundation for the very vast stream of research on game theory [55, 62]. Yet, this dependence is often ignored in analyses on pest control and impact. Invasive species control requires a coordination among stakeholders and across countries. Hence, as shown in chapter 3 and 4, modelling economic dependencies in pest risk assessments can generate critical information on possibly diverging incentives among the different actors. This can support discussions and allow to find common ground on, for example, budget allocation for management, the design of policy such that vulnerable stakeholders are supported especially, and by engaging stakeholders that lack private (economic) incentives to ensure collective efforts.

Traditionally, farm-level decision making of pest control was analyzed under the damage abatement framework [100]. In a nutshell, models derived optimal control levels such that marginal costs of actions equate to their marginal benefits [223, 306]. The underlying theoretical foundation of such modelling efforts was the idea that costs and benefits are readily observable by the decision-making units. The assumption that one decision maker can decide upon the level of control effectively constrains the geographic space analyzed to the unit's own property. Such analyses fail to recognize that marginal costs of pest control are a function of pest pressure which itself is a function of neighbors' actions [97, 307]. Our identification of spatial spillovers in chapter 2 provides empirical support of this. Consequently, the marginal cost and benefit functions of pest control extend beyond a single decision-making unit and farm-level decision analyses can be expected to fail in capturing the entirety of costs and benefits which may result in suboptimal strategies.

Compared to spatial dependence, temporal dependence has found significantly more attention in the literature. Economic modelling work has recognized that on-farm profit maximization might not be appropriately modelled as a one-period problem, but rather a dynamic optimization of long-term profits is needed to better capture a farmer's objective function [308]. Similarly, control of pests in the current year might influence the populations' reproductive cycle and thereby result in spillover effects which influence the pest pressure in the following year [31].

There certainly is a case to be made on the need to refine temporal analyses. Impact assessments in perennial hosts require temporal considerations. Chapter 3 contributed to temporal aspects by showing how losses in undertaken investments, due to the premature death of hosts, could be captured through the foregone annuities. However, often, as shown in chapter 3 and 4, temporal models use discrete annual time steps and thereby fail to acknowledge within-year dynamics, such as the different development stages of a pest, that might be highly relevant for optimizing control [175].

While there is a need to improve the temporal resolution of analyses, I believe the spatial dimension remains significantly more underutilized in economics. Just as temporal problems involve choosing a strategy that comprises a path of decisions which are interdependent across time [309], optimal control in a spatial system involve simultaneously choosing actions across an interconnected landscape [39]. It is crucial to understand that the, in economics, traditionally asked questions on *when* and *how much* control should be employed are inseparable from the complex spatial question on *where* these efforts should be targeted [39].

Through chapter 2, I showed that managerial performance of a farm can be associated to neighbors' characteristics. While I did find that spatial spillovers were associated with input-specific farm performance, the coefficient directions for some characteristics depended on the definition of the neighborhood. It seems plausible that there might be differences in community effects versus individuals influencing their immediate neighbors. However, while our study advocated for practitioners to test different model specifications, I must acknowledge the lack of sound theoretical work on possible peer effects.

The spatial econometric techniques are traditionally computed using geographic distances as a basis for defining the neighborhood structure among observations [126]. While some work has been done using alternative proxies for closeness such as the size of trade flows between regions [310], there is a critical lack of work on how economists could capture knowledge spillover effects in a world in which *social networks* are digital and easily span across the globe. Neighborhood structures which are based on geographic distances will remain appropriate for capturing biological interdependencies such as effects from pest spread, but they may be meaningless for modelling exchanges of knowledge in a world in which people feel more closely connected to a digital representation of a person sitting on the other end of the world compared to a neighbor living right next door. Nevertheless, chapter 2 contributes to the discussion on the need of extending decision analyses beyond a single farm, certainly in analyses of optimal pest control.

Spatial heterogeneity results from differences in the operational environment that lead to different input requirement and output possibility sets. The need to correct for heterogeneity is widely acknowledged in econometric work and most commonly approached through panel estimators that aim to control heterogeneity implicitly through random or fixed effects [311, 312]. Similarly, the need to carefully consider environmental differences between decision-making units when comparing their managerial performance has already been stressed in the seminal work of Farrell [102]. While some work has addressed spatial heterogeneity in efficiency and productivity analyses (e.g., [103, 104]), majority of benchmarking studies do not include any environmental data. While the inclusion of quasi-fixed factors is often done to acknowledge economic differences among decision-making units, incorporating data to account for differences in the physical environment remains to be the exception. Fortunately, this issue has started to attract the attention of scholars in recent years (e.g., [104, 313–317]). Ignorance of potential differences in the operational environment among decision-making units risks that (spatial) heterogeneity is falsely attributed to managerial inefficiency [316]. Unfortunately, despite initially planning to do so, in chapter 2 I was also not able to appropriately correct for environmental heterogeneity as no data on relevant environmental factors were available. After receiving

access to the data, I realized that the geographic information provided was shifted to retain anonymity of farmers. Consequently, external environmental data could not be added to the analysis.

Acknowledging heterogeneity is also important for pest risk assessments. Differences in the environmental conditions of locations, i.e., spatial heterogeneity, are fundamental to species distribution models which predict differences in areas suitability for establishment precisely because of variation in environmental conditions [244, 318]. Similarly, spread modelling work has long recognized the need to capture environmental heterogeneity in the landscape [75, 93]. Despite this, economic pest risk assessments often compute potential impacts using aggregate markets (e.g., [4, 179, 223, 319, 320]). Arguably, heterogeneity will result in varying impacts and consequences from management practices. In chapter 3, I have stratified the olive production into different cropping systems. This turned out to be particularly relevant when simulating replanting of olive orchards as high-density system reach a profitable state significantly faster than traditional, low-density, orchards.

Hence, while stratification is useful to refine the aggregated country-level estimates, in my view, the more important benefit of acknowledging heterogeneity, e.g., through stratification, is that it allows to assess whether there are unequal consequences from pest spread to different strata. For example, due to the differences in the orchard's growth speed [215], highly commercialized high-density systems could potentially recover relatively quickly from further spread of the pest if resistant cultivars were available. Small-scale growers that cultivate traditional orchards might not only lack the financial means for replanting, due to the system's lower profitability, but they would also have to bear a longer nurturing period following replanting in which trees are not producing significant yields. Such insights are not possible if homogeneity of the population is assumed. Yet, these results are highly relevant for the design of policies that may aim at supporting more vulnerable strata especially. Having said that, the stratification as done in chapter 3 is far from perfect. I will revisit this point in more detail in the next section and in the section on limitations below.

The lack of spatial modelling work in economic analyses of pest impact and control is unlikely to be related to the unavailability of spatial biological models. Economic models severely lag behind in

their ability to integrate spatial components. Yet, spatial economic models are very much needed to derive optimal strategies [321]. Many economic analyses use non-spatial numerical simulations that provide estimates of the supply affected in non-spatial, aggregated, markets [4, 96, 320]. However, as shown in chapter 3 and 4 even if spatially explicit models for pest spread are at hand, results of them must often either be translated into a non-spatial representation or one that aggregates from precise geographic locations to country-level estimates. I will discuss my view on why this might be the case at the end of the section and more extensively in the next section of this chapter.

Non-spatial numerical simulations are unable to acknowledge potential spatial clustering of hosts and they are unable to derive insights on locations in which a pest introduction would result in drastic impact. Not seldom, the production of host plants is somewhat concentrated in certain areas. This aspect is critical to the realization of pest spread and impact [321]. Spatial clustering of host cultivation results out of differences in habitat suitability, as well as potential economic benefits from agglomeration [32], or an area's reputation for a particular product which can generate price premia and thereby incentivize cultivation [322, 323]. While the effect of monocultures on pest risk traditionally received attention [324], here I do not necessarily refer to monocultures but merely to differences in the regional focus on certain crops. As evident from the used landcover data in chapter 3, olive cultivation is heavily concentrated in certain parts of Europe. In Spain, this concentration is in Andalusia, whereas in Greece the concentration is on Crete.

The spatial clustering of hosts has crucial implication for pest epidemics for two reasons. First, depending on the pathosystem, and the dispersal process of the pest, the likelihood of pest introduction might be higher in areas with more hosts present. The larger the number of hosts within an area, the higher might be the likelihood of introduction there, which aggravates risk. Second, if hazardous pests are introduced into a host hotspot even slow, very localized, spread rates can result in devastating impacts. This was evident from the stochastic results of spread over time for *Xylella fastidiosa* subspecies *pauca* following simulated introductions into Greece or Spain. Similarly, Strona et al. [325] show that a clustered network of hosts may result in vast geographic areas essentially become a *small world*. If the network

of hosts is sufficiently tight epidemiological uncertainties may become close to irrelevant [156]. As host locations matter, the geographic point of pest introduction becomes a critical component of the modelling work [39]. Consequently, deriving the sensitivity of results with respect to different points of introduction allows for actionable insights on areas that should be prioritized in pest surveys. The simplification of pest spread using non-spatial models essentially results in a smoothed progression of the invasion over time which implicitly assumes that hosts are homogeneously spread over the landscape. As many host crops are cultivated in a clustered way, such simplified spread modelling ignoring the spatial patterns of host cover can result in wrong insights on the potential epidemiological progression and, as a result, economic impact.

The spatial dimension comprises more than intangible interactions of individuals' actions across geographic areas. Including spatial aspects into modelling work means acknowledging the heterogeneity of areas and understanding that the landscape configuration itself is critical for the epidemiological progression, pest control, and impact assessments [81, 83, 93, 326]. Chapter 3 discussed economic benefits from resistant olive cultivars to control the spread of *Xylella fastidiosa* subspecies *pauca*. It was assumed that replanting followed the pest affecting an orchard. Thereby, spread continued throughout the landscape and replanting was merely a reactive response. Resistant varieties may still contribute to transmission [187], however, if the inoculum provided by resistant cultivars is sufficiently small the spatial configuration of *where* hosts are replanted, possibly before the pest arrives, can become a critical factor in managing the invasion.

Differences in the landscape configuration might also result in differences in the environmental burden from control measures such as pesticide applications [72, 327]. Therefore, not only environmental impact assessments for pesticides must take the landscape configuration, and possible future changes of it, into account [328], but also analyses of optimal pest control strategies should take the heterogeneity of the analyzed landscape and its configuration into account to minimize environmental burden. As pesticide applications are in the millions of tons worldwide [27], this is a critical issue that deserves more attention in future modelling work.

As a silver lining: including information on the landscape configuration does not solely add complexity without providing benefits. Epanchin-Niell & Wilen [39] show that control strategies can become more effective when using information on host locations, natural barriers, and dispersal geometries. This was similarly visible in our spread modelling efforts for *Xylella fastidiosa* subspecies *pauca*. In Greece, the country topography, with many waterbodies separating areas of production, might be exploitable and thereby could result in eradication efforts becoming more feasible compared to other regions in Europe if the pest were to be introduced in the future. In addition, by constraining spread to climatically suitable habitat I simulated epidemics that had to travel around non-suitable hurdles such as mountain ranges. This not only delayed realized spread, through the additional path that was required to spread around the hurdle, which is relevant for measuring discounted economic impact, but non-penetrable hurdles also opened opportunities for control strategies that exploit these boundaries. By strategically incorporating landscape features into the control strategy, the invasion front and consequently impact can be reduced, and control strategies can become more cost effective [39]. While certainly far from perfect, through inclusion of information on climatic suitability and host locations I did succeed in capturing some of these spatial aspects in chapters 3 and 4. An integration of additional landscape features into the spread model developed in chapter 3 could easily be achieved using gridded data which provides a binary indication of whether a given pixel can be spread through or not.

A strategic placement of hosts can reduce the pest risk [321]. The fact that the distances between hosts matters has been introduced to everyone around the globe in the recent COVID crises through *social distancing* measures. The same applies to plant pests. By acknowledging the landscape configuration of hosts, the continuity of spread can be disrupted and risk for future pest invasions can be reduced. This also allows to minimize the application of ecotoxic control agents such as pesticides and fungicides as these could be placed more strategically as opposed to large-scale spraying on every field in a region [321]. While the idea of spatial barriers between hosts is already applied in intercropping and strip-cropping techniques [329–331], achieving this across a landscape would require a coordination of collective efforts across many individual decision-making units. Minimizing risks

associated with spatially clustered hosts could result in discontinuous, patchy, surfaces of host availability such that the maximum extent of a pest epidemic would be somewhat limited in absence of long-distance jumps [331]. The decrease in economic risk and the benefits of a slowing down of potential neighbor to neighbor transmission should be considered in analyses on sound landscape configurations for agricultural production [326, 332–334].

The dissertation has provided four research chapters which emphasize the crucial importance of space in economic analyses. The discussion within this section has further substantiated why “*space matters*” [321, p. 395]. Hence, the question emerges: why do many economic analyses continue to ignore the spatial dimension? I believe this is likely related to two factors of which none has to do with the added complexity of spatial modelling work, and both need to be remedied. First, in contrast to temporal models, spatial modelling techniques remain largely absent from undergraduate and graduate curricula in economics. Consequently, many young economists have limited to no exposure to spatial techniques before starting their research career. This likely constrains the conceptual breath in which many researchers think when approaching a problem that may have underlying spatial dynamic structure such as pest spread. Second, most economic data lack a proper spatial indexation. Including the time, in years, months, days, or even hours and minutes, a record was observed alongside the measured value has become common practice. Yet, the majority of economic data has either no spatial index at all, or very coarse information indicating the country or, if lucky, the NUTS-2 region in which a measurement was taken. In the rare cases in which precise geographic information is available in microeconomic data, confidentiality restrictions often prevent that these data are provided to researchers. Time-series analysts would likely not be able to learn very much from data with a temporal resolution in decades or centuries. The same applies to coarse spatial information. I believe in economic research in general, but for agricultural economics in particular, we need to drastically change this norm and acknowledge that not only *when* but also *where* a measurement was taken is crucial for analyses. In what follows, I will take a closer look at opportunities and challenges for agri-environmental data integration which can be achieved through a proper spatial indexation.

## 6.2 Agri-environmental Data Integration

Above, I have argued that spatial heterogeneity and dependence matter in analyses on impact and control of pests. I have also noted that many economic analyses remain non-spatial, and that this fact is likely not related to the unavailability of spatial biological models. Carrasco et al. [97] noted that improving the integration of biological and economic models would require information on the marginal and average cost function over the landscape. This directly relates to the point made above on the need to extend analyses on optimal control beyond a single farm. In chapter 2, I was unable to incorporate environmental data to account for spatial heterogeneity. In chapters 3 and 4, the precise geographic spread simulations had to be aggregated to country-level estimates. The sole reason for these limitations was the unavailability of a finer spatial indexation for economic data.

Despite early recognition of the importance of time in empirical models [311, 312], prior to 1970 even simple regression models took up to a day to compute on the available electromechanical desk computers which, among other factors, resulted in mostly cross-sectional analyses. In subsequent years, the statistical developments proliferated and over time economic research more and more tried to replace cross-sectional analyses by panel techniques which account for heterogeneity among observations [335]. Arguably, these developments were enabled by better computing technologies and driven by an increasing understanding that temporal effects cannot only be exploited to control for individual heterogeneity, but that the temporal dimension itself might hold valuable information that can be analyzed.

Geographic information systems are direct successors of the centuries-old practice of map making [336]. Therefore, the idea of having georeferenced information is by no means new. While early use of spatial data was driven by environmental disciplines in which space itself is fundamental such as geography, soil science, and climatology, the value of spatial data is more and more recognized in interdisciplinary research [337]. In the social sciences, the regional sciences are probably the most prominent advocates and users of spatial economic data. Here, undergraduate and graduate textbooks provide students with knowledge on diverse spatial modelling techniques [338, pp. 1105-1674]. Most general economists and agricultural economists

lag behind in their expertise on including spatial data into modelling work. Similarly, methodological contributions by scholars focusing on spatial econometric models were for a long time marginalized and not taken seriously by traditional econometricians [67]. Whether spatial econometric models are useful tools remains contested [114].

In what follows, I will discuss opportunities of including longitude and latitude information alongside economic data, as well as challenges that must be overcome to ensure proper use and allow for the full benefits. While I believe this point relates to many, if not all, economic datasets, to keep the discussion focused, I will highlight opportunities and challenges using the case of the Farm Accountancy Data Network (FADN).

The FADN data underlie the empirical application in chapter 2. The microeconomic data is collected every year throughout Europe and managed by the European Commission. The data comprises representative samples for different farm types and holds information on the generated revenues, expenses, balance sheet positions, and general farm characteristics. To access the microeconomic data, researchers must file a request and motivate every variable they would like to get access to. The FADN data is extensively used in productivity and efficiency analyses (e.g., [101, 112, 136, 339]), but also in other areas within the agricultural economics domain (e.g., [340, 341]). For the project under which chapter 2 was enabled I obtained shifted longitude and latitude information that retained exact distances among farmers without providing us the real locations. Interestingly, a closer look into the FADN suggests that the exact longitude and latitude data are commonly collected when surveying the farms.<sup>1</sup> Hence, these data are already available but simply not utilized besides the provision of regional indicators. Considering data privacy, it appears defensible to provide only regional dummies when data is handed to external researcher. However, there are serious disadvantages from the lack of environmental data integration into the FADN. Considering the

---

<sup>1</sup> After requesting access to the microeconomic data at DG AGRI, an Excel sheet with thousands of variables is provided out of which a selection is to be made. Three variables labeled *A\_LO\_40\_N0*, *A\_LO\_40\_N1*, *A\_LO\_40\_N2* may provide NUTS-0, NUTS-1, or NUTS-2 level indicators which are “*estimate by DG AGRI based on geographic information of farm*”. Through my work on chapter 2, I am aware that similarly labeled variables denote the precise longitudes and latitudes in the raw database.

sensitivity of the geographic information, this data integration could at least be achieved within the European Commission. The new *Farm to Fork strategy* discusses the intended transformation from the Farm *Accountancy* Data Network to the Farm *Sustainability* Data Network which will add data on biodiversity targets and sustainability indicators [342]. However, it appears as if the strategy fails to recognize the need for utilizing the already available geographic information for this pursuit because it is suggested that these additional data will be “collected”.

### 6.2.1 Opportunities

Most data are broadly considered to be static, meaning that researchers often view a given compilation of measurements as fixed. Scholars may work on the same dataset over several projects, or different teams of researchers try to obtain various insights out of the same dataset. However, in both scenarios the data remain unchanged, and the insights scholars can derive are somewhat predetermined by what had been collected when the dataset was compiled. Having a spatial indexation turns a static dataset into one that can dynamically grow, forever. The reason for this is that essentially *all* other data that is collected, predicted, or simulated on anything that might relate to the original data can be retroactively merged into the dataset based on a shared geographic indexation such as longitudes and latitudes. In my view, it is difficult to overemphasize how this simple fact fundamentally alters opportunities for societal knowledge generation. Our ability to identify and explain patterns in data would no longer be constrained to the set of variables that were judged to be relevant, and therefore collected, at the time. Instead, historic datasets would be allowed to grow alongside new technologies, methodological developments, and our understanding of which aspects might matter.

As an example, in recent years the work on satellite imagery proliferated [343]. While satellite images have already been collected for decades, the relatively recent development of powerful cloud computing technologies in combination with methodological advancements in machine learning now allows for large scale processing of the collec-

tions [344]. These advancements opened up opportunities to process decades of data and generate predictions of environmental properties in space and time [345]. Newly generated predictions of environmental properties such as temperatures [193], precipitation [193], extreme weather events [346], soil characteristics [279, 281, 283], pollution [347, 348], or other factors that are relevant to, or directly caused by, agricultural production could all be integrated into historic FADN records if a spatial indexation would be available. It may not have been possible to collect some of these data, for example, in the 1970s. Of course, this principle holds for any future developments that would provide new measurements of variables in space and time where both space and time coincide with historically collected, otherwise static, economic observations.

Consequently, georeferencing FADN data would result in an explosion of explanatory variables that could be investigated and allow for a genuine integration of knowledge, in the form of concrete data, from different research disciplines into agricultural economics. In addition, by sharing a spatial index, the burden of data collection does not exclusively fall on the people conducting the surveys. For example, in recent years there rightfully is a growing interest in including environmental aspects into productivity and efficiency analyses (e.g., [349]). This requires information on emissions or other pollutants which are difficult to measure. Farmers might not necessarily be aware of these data and even if asked during the data collection process, they would not be able to provide sound estimates. As the list of variables demanded by scientists increases, the time needed to complete the survey would drastically exceed what farmers could cope with. At the same time, sophisticated models are being developed to predict emissions across space [350, 351]. Hence, there is no need to include such questions into surveys if data can be retroactively merged through other sources. A shared spatial index, such as longitude and latitude, allows for data collection to become a multi-disciplinary collective effort.

In chapters 3 and 4, I conducted pest impact assessments based on country-level economic data that was used to calibrate the economic models. While I tried to capture heterogeneity by modelling different cropping systems and countries, georeferenced economic data would clearly allow to better capture the heterogeneity in economic conditions in different areas and therefore allow to obtain more precise estimates

of potential pest impact. Deriving differences in the regional impacts could deliver actionable insights that would allow to prioritize areas for management. Having georeferenced FADN data would indeed provide empirical information on the marginal and average cost function across the landscape [97]. In addition to improving pest risk assessments by allowing to better capture economic heterogeneity, spatially indexed economic data might also allow deriving insights on the biological process of pest spread in data scarce situations.

Not seldom, analyses of novel pest epidemics are hampered by the unavailability of information on the biology of the pathosystem. As seen in chapters 3 and 4, information on the spread rate of *Xylella fastidiosa* subspecies *pauca* was highly uncertain and experts' opinions were elicited to get a sense of possible speeds with which future spread might occur. The first detection of the bacteria was in 2013 [24]. In the following years, more and more olive orchards were reported to be infected in Apulia. Olive growers are routinely included in the FADN surveys and, presumably, the infected region in Apulia was included in the Italian surveys. Under availability of longitude and latitude information, it likely would have been possible to exploit the panel nature of the FADN data and econometrically measure unusual changes in profits for olive growers in the infected zone over time. This would not only have allowed for an estimation of realized impacts, based on empirical data, which would greatly inform local authorities and risk managers, but this estimation could also have provided an indication of the speed with which spread might have occurred over the landscape by revealing spatio-temporal patterns of unusual changes in profits. In absence of better biological information, this estimate could be used to parameterize spread models.

By estimating relationships based on historic data, real-time predictions can be made to inform decision makers and farmers on the ongoing production cycle. While many data, such as FADN, are collected and recorded at discrete annual time steps, many modern data sources such as satellite imagery provide new data every few days. If historic data were used to train models to relate, for example, the *normalized difference vegetation index* (NDVI) to realized crop yields, the higher frequency of satellite imagery could be exploited to provide close to real-time insights on the current status of production [352, 353]. This information can help facility decision making, allow

for preparedness management in case of disastrous weather events or pest outbreaks, and thereby help to safeguard food security in vulnerable regions [354–356]. While such tools are already being heavily developed, the lack of data is often approached by using synthetic data which is generated based on process-based models that relate vegetation indices to simulated crop yields [343]. Georeferenced economic data would add valuable calibration data by providing on-the-ground information on realized revenues which can be deflated to implicit quantities,<sup>2</sup> as well as local differences in technologies, measured for example by the capital stock, and farmers' input usage.

## 6.2.2 Challenges

In my view, there are three domains that require careful attention in the pursuit of agri-environmental data integration. The first domain is on aspects related to data privacy and governance. Precise geographic information clearly undermines anonymity of farmers in the FADN. For spatial econometric models such as the one presented in chapter 2, distances among farmers suffice. However, for the aforementioned opportunities of georeferenced economic data precise information on locations is required. Arguably, these data should not be openly shared. However, data scientists at the European Commission could assist external researchers in merging desired spatial databases to records within the FADN based on longitudes and latitudes. This would allow to utilize many of the advantages of a spatial indexation without having to provide modelers with the sensitive data on locations. In addition, protocols with increased security checks could be established such that scholars may be granted access to the geographic information under strict security standards.

Regardless of these two suggestions, at least scientists directly employed under the European Commission should make use of the ability to integrate environmental data for their research on the FADN. The new Farm to Fork strategy mentions that additional data on

---

<sup>2</sup> Deflation describes the process of computing quantities by dividing monetary variables, such as expenses and revenues, by corresponding price indices. The obtained estimates are usually referred to as implicit quantities.

biodiversity targets and sustainability indicators will be “collected” and added to the FADN [342]. The huge variety, and high quality, of available spatial data on these aspects should be utilized by using the already available geographic information to match external data to FADN records. Solely relying on additional survey questions and FADN-based calculations of indicators would be a huge loss of opportunity. As the additional value of having spatially indexed data is immense, in my opinion the question is not whether European authorities should foster such a spatial indexation scheme but rather in which way this must be enabled to comply with privacy laws and address security concerns. Clearly, this requires further work from experts on governance, law, and information technology.

The second domain that, in my view, requires further attention is the spatial granularity of the indexation. The longitude and latitude information that is already collected in the FADN surveys describe the centroid of a farm. As discussed in the General Introduction and chapter 2, fields are usually scattered around a farm. Hence, centroids are suboptimal indices to match with environmental data as they only allow for point estimates at locations that are not necessarily relevant for the production process. Furthermore, centroid points do not allow to capture the spatial heterogeneity in the environmental conditions across the farmer’s fields. Ideally, FADN would include spatial geometry (i.e., polygon) objects on all individual fields for each farm. This would allow to clearly relate individual fields to economic observations, account for fields’ heterogeneity by enabling field-level data integration on environmental properties and, very importantly, provide spatially explicit data on planted crops. As highlighted in the previous section, spatial information on hosts is critical for spread simulations and economic analyses. The olive landcover data used in chapter 3 and 4 – in fact, to the best of my knowledge all landcover products – are *predictions* of land use which are based on classification algorithms that assign a land use class based on the pixel composition of a satellite image [357]. Of course, these approaches come with classification errors that are unobservable to the user. Furthermore, most landcover products provide classifications of categories that are differentiable by the algorithms such as forests versus cropland versus urban areas. Correctly classifying different arable crops from space is significantly more difficult and consequently spatially explicit

information on, for example, potatoes versus wheat versus corn is unavailable for Europe as far as I am aware. Notably, the United States Department of Agriculture is already publishing arable crop-specific spatial data which is updated annually.<sup>3</sup> These predictions were enabled by field-level labels on planted crops.

Field-level polygons with ground-truth on which crop was planted in the field each year would provide three invaluable benefits. First, for areas in which this information is available modelers could use these data directly and thereby avoid having to build spread and economic models on top of predictions that most likely contain errors. Pest risk assessments typically develop a map for the potential establishment of the evaluated pest. Overlaying this map with data on where hosts are cultivated would provide valuable insights on the actual risk. Second, these data would provide valuable calibration data for algorithms that are being trained to provide large-scale land cover classifications. With these additional data, arable crop-specific predictions across Europe might be achievable in the future. Lastly, field polygons allow for landscape-wide field-level analyses which could, and arguably should, be used to derive insights that directly benefit farmers. The new Farm to Fork strategy explicitly mentions the need to provide “*tailored advisory services*” to farmer [342]. Field-level polygons that can be related to economic observations would enable scalable solutions to allow for tailored advisory services in a cost-effective way.

Lastly, the third domain that requires attention is the establishment of *proper* pilot projects that clearly communicate the immense value a spatial indexation scheme can generate for scientists and farmers, as well as society at large. Currently, there are a few pilot projects that aim to include satellite imagery into spatial farm-level analyses. For example, the Sen4CAP<sup>4</sup> project aims to improve the spatial monitoring of farmers using satellite imagery to better enforce environmental regulations. In my view, I struggle to find a worse objective if the goal is to get stakeholders excited about what these technologies have to offer. Farmers are already heavily scrutinized on their environmental performance while facing increasing price pressure. The success of establishing, for example, a field-level spatially indexed FADN database depends on the willingness of farmers to participate. By introducing

---

<sup>3</sup> <https://nassgeodata.gmu.edu/CropScape/>

<sup>4</sup> <http://esa-sen4cap.org/>

these technologies as a sophisticated tool to spy on farmers and enforce regulations, a narrative is created that there is nothing to gain but much to lose if farmers were to accept increased levels of transparency through georeferenced information. There is a critical need to initiate pilot projects that place farmers' benefits first. This could, for example, be achieved by providing field-level analyses that generate actionable information to participating farmers free of charge.

### 6.3 Policy Implications

Plant health is a public good and is generally considered a positive externality [49]. The “*consumption*” of plant health does not reduce its availability to others (non-rivalness) and excluding individuals is not feasible (non-excludability) [49]. However, its protection not only has positive spillover effects but also negative ones. While insufficient control can result in increased pest pressure in the landscape, too much control, for example through pesticide applications, may unnecessarily pollute the environment and might inadvertently select for resistance in the pest species [306]. As often the case with public goods, market allocation is likely inefficient and policy interventions are warranted [43, 46, 49]. As shown in chapter 4, in markets characterized by inelastic supply and demand consumers are the main beneficiaries of pest control efforts due to the economic dependencies among producers and consumers. Hence, using public funds for policy interventions is justified as it benefits taxpayers.

What complicates the public good characteristic of plant health is the fact that it is a so called *weakest-link* public good [46, 49]. As the lack of performance of a single decision-making unit can jeopardize the biosecurity level for a whole area, the success of control is determined by the worst performing unit (i.e., the weakest link). Through areas' differences in climate, ecosystems, and the frequency of introductions, among other factors, heterogeneity in control costs arise that may result in decision-making units with higher costs not acting sufficiently. Differences in the hosts value at risk similarly result in heterogeneity in direct benefits from control efforts which can deteriorate individuals' motivation to control pests [37, 39]. As the number of non-acting units

becomes larger, the economic incentives to all other decision-making units for establishing pest control approach zero [38, 46, 307]. Game theoretical modelling suggests that public signals of own control efforts are required to overcome this deadlock [307]. Ideally, cooperative efforts should be spear-headed by decision-making units that have the highest economic incentive to prevent pest spread [307]. In fact, it is in the best interest of the decision-making units to cooperate with the weaker link and increase their incentives to control [37]. As a result, earlier work has often stressed the need for a global coordination and cooperation in the management of invasive species [46]. By mapping hotspots for pest introductions, as done in chapter 5, discussions among decision-making units can be informed and decision making on collective management supported.

In chapters 3 and 4, the diverging interest between stakeholders in controlling the pest became very apparent. The economic dependence between competitors results in economic benefits from shortages in supply. Consequently, control strategies, such as the breeding of resistant cultivars, were disadvantageous to unaffected producers that would have profited from price increases due to supply shortages under pest spread. By including price responses into a pest risk modelling work, economic dependencies are captured and diverging interests can be spotted by computing changes in monetary flows to different stakeholders. Arguably, the computed gains in profit to some growers as a result of affected growers not being able to replant do not consider any empathy or feelings of solidarity such growers might have, nor do they account for the possibility that even non-affected growers might derive value from merely knowing that a *Xylella fastidiosa* subspecies *pauca* resistant cultivar exists in case they ever needed one. Furthermore, non-affected growers could profit in the long term from resistant varieties because such varieties could slow down pest spread such that it would never reach the non-affected growers. The pure competitive nature of the economic computation might not apply in its entirety to the real world. Nevertheless, such information can support public discourse by providing context to why stakeholders might respond differently to proposed management strategies.

For the invader *Xylella fastidiosa* subspecies *pauca*, chapter 3 and 4 provide economic evidence that regulatory measures are warranted. Both chapters indicated that the further spread of the disease must be

avoided and introductions into new territories, especially into Spain, should be prevented. As consumers were the main beneficiaries of the control of *Xylella fastidiosa* subspecies *pauca*, using public funds to support containment and surveillance efforts is warranted. Chapter 3 has highlighted the sizable economic benefits adaptation strategies, such as resistant cultivars, would generate to affected countries. Hence, decision makers must continue their support of research that aims to deliver long-term adaptation strategies to *Xylella fastidiosa* subspecies *pauca* in Europe. Unfortunately, the continued cultivation under pest pressure was found to likely be unprofitable in as little as three to four years due to the olive orchards' relatively low profitability. Depending on the cropping system, financial support for up to a decade might be needed to ensure that farmers who are keen to transition to a resistant cultivar are able to do so as olives require a sizable period in which inputs must be employed but no, or very limited amounts, of output is generated. While the complete ceasing of production following the desiccation of the orchards is a rather strong worst-case scenario, many olive cropping systems operate on marginal land due to the trees' robustness. Hence, it is likely that not many, if any, alternative crops could be cultivated in affected areas. Therefore, management strategies must prevent spread and, if available, promptly move to resistant cultivars to avoid large losses in investment and an accumulation of foregone profits in the future.

The general susceptibility of areas to pest invasion is, among other factors, determined by the anthropogenic disturbance of the ecosystem [46]. The volume and direction of trade are good empirical predictors where invasive pests are introduced [6, 46]. Hence, the extent to which countries should hold an interest in fostering multi-national collaboration should be related to their involvement in, and derived benefits from, international trade. As pest invasions are externalities from international trade, solutions should aim to internalize, as best as possible, these costs [46]. Strategies that heavily rely on border inspections fail to acknowledge that costs for the strategy will largely be determined by the source countries' ability to control pest pressure. As evident in the case of *Xylella fastidiosa* subspecies *pauca*, hazardous species often emerge from areas where the capacity and incentive to

control may be limited.<sup>5</sup> Consequently, the question arises whether a sink country's best strategy is indeed to put most (monetary) efforts into border inspection, or if extension services in source countries could yield better returns on investment. Analyses like the hotspot prediction of chapter 5 allow identifying which areas require increased vigilance. This could be the basis for additional work that derives optimal budget allocation among member states on grounds of areas' risk of pest introduction.

The complexity of spatial spillovers and interdependency in actions of individual decision-making units might lead to the insight that "*pest management decisions may be better made at a regional level rather than at the individual farm level*" [306, p.228]. Joint efforts may not only increase the likelihood of success, but also decreases the required monitoring and information costs for individual farmers [306]. As deriving insights through data-driven information systems is easily scalable across vast areas, the cost per user is a function of the number of participants. Arguably, the more decision-making units are involved the more difficult it becomes to coordinate control [38]. A shared information exchange system would greatly assist such coordination in particular if accompanied by tools that foster interactive communication among peers [50, 359].

Spatially uniform policies over a spatially heterogeneous landscape likely result in inefficient outcomes [80]. While this holds true for many domains, the agricultural sector, and pests in particular, might be the most obvious example of this. As pest growth and dispersal is spatially heterogeneous, management response must reflect this. The feasibility of spatially heterogeneous policies is often debated on the grounds of the required information that would result in such efforts being too costly [306]. In addition to georeferenced economic data, insights obtained through hotspot analyses such as chapter 5 could be used to inform spatially heterogeneous policies through the design of legislation that takes areas' risk scores into account. Unfortunately, nature is stochastic as, for example, the weather and pest pressure changes from year to year. Hence, ideally this would require policies to be dynamic as well which is likely difficult in practice. Due to spillover effects, traditional policy impact analyses conducted on individual

---

<sup>5</sup> *Xylella fastidiosa* subspecies *pauca* might have been introduced through coffee plants sourced from Costa Rica [358]

farm-level may be inappropriate and should be extended to include potential community effects [35].

While involving citizens in data collection has found significant attention and gained in popularity [360, 361], I believe the full potential of citizen science remains largely underutilized. Chapter 5 highlighted the sizable amounts of data that is already produced by citizen science using one example, the Global Biodiversity Information Facility. The chapter stressed the statistical problems that arise in citizen science data due to the opportunistic sampling scheme [275]. Citizen science data has a huge potential because it can scale with the number of people that participate [360]. Species surveying is a time-consuming task that, consequently, costs significant amounts of money. In chapter 5, I emphasized the need to include true absences into surveys which would make those even more time-consuming [270]. While some life forms, such as bacteria, viruses, and nematodes are difficult to spot and therefore certainly require experts for surveying, other species such as trees, plants, birds, or even insects could very likely be correctly identified by laymen if given proper instructions. Regulators may strengthen citizen science by supporting an app-based coordination of laymen's search efforts [362–364]. Through a coordination of collective search efforts, it might be possible to generate properly sampled datasets based on citizens' reporting.

## 6.4 Limitations and Future Work

Without repeating too much of the previous section, I briefly discuss a couple of limitations that arose due to issues of data availability. In chapter 2, the spatial coverage of the arable crop farms was not ideal. To improve the optimization of the parameter of the spatial weight matrix, I decided to work with a balanced panel that spans six years and comprises 75 farms. Unfortunately, the small number of farms resulted in a sparse spatial coverage that very likely not only affected the estimation of the parameter of the spatial weight matrix, which might be evident from the large confidence intervals for the distance cutoff, but it also resulted in the need to be very careful when extrapolating the obtained insights on the spillover effects. In

addition, the small number of farms may have resulted in issues with the dimensionality of the linear programming problem that could have led to the small estimates for technical inefficiency. In retrospect, it might have been a better decision to extract a balanced dataset for a smaller number of years. Being less restrictive with the number of years each farm needs to be observed would have allowed to extract larger cross-sections that would provide a denser spatial coverage and a better discrimination in the linear programming model. As temporal dynamics were not central to the analysis, the cost of having six years of data (i.e., the smaller number of decision-making units in each year) might have outweighed the benefits. Hence, omitting some years to increase the number of farms in each year might have been a better decision.

As briefly discussed before, the FADN data used for chapter 2 provided centroids of farms. The underlying motivation for the chapter was the idea that management decisions on neighboring *fields* affect, for example, pest pressure on other fields. As data on field locations were not available, the centroids of the farms were used and geographic distances between centroids formed the basis of the proximity structure. However, it would be more in line with the chapter's motivation to compute distances between the actual fields. Furthermore, as briefly discussed above, modelling social knowledge spillovers might be better approached by proximity structures that are not only based on geographic distances. Further work is needed to find a proper measure of *social* distance which could form the basis of proximity structures.

In chapter 3 and 4, country-level data was used to calibrate the economic models. More granular economic data would have significantly improved both analyses as it would have allowed to better capture heterogeneity. The country-level estimates for prices and costs used in chapter 3 were mostly obtained from the Olive Oil Farm Report [185]. The research chapters were not developed in the exact order as presented in this dissertation. Consequently, while the Olive Oil Farm Report provided aggregated statistics based on FADN data, at the time I was only vaguely aware of the FADN data and did not realize that I could request access to the farm-level data. In retrospect, it might have been a good idea to request Italian, Spanish, and Greek microeconomic FADN data for olive growers and, if the spatial coverage would have allowed, calibrate the economic model to producers' economic condi-

tions in the different NUTS-2, or even NUTS-3, regions. While price data is not included in the FADN, at least heterogeneity in operational costs could have been better captured in this way. Furthermore, better data on biological components of the model such as the annual yield decline due to *Xylella fastidiosa* subspecies *pauca* would have improved the analysis.

The main limitation for chapter 5 was the absence of systematic data on pest presences and absences. As this is extensively discussed within the chapter itself, I will not repeat too much here. However, it is important to acknowledge that the underlying motivation for the chapter was the question whether it would be possible to bundle several species together and jointly analyze them to save resources. This approach could be one way to address the overwhelming gap of information on areas' suitability for establishment or introduction that exists for most invasive species. While I found that this modelling approach resulted in very well-performing models, the ability to jointly analyze species of course depends on the availability of harmonized data for every species that one is interested in. As I had to rely on the GBIF database for this information, I was not able to disentangle whether the predictions are a consequence of reporting bias or whether the included anthropogenic factors indeed promote pest introduction. Hence, while joint analyses could save resources in the modelling stage, the feasibility of this strategy would likely require additional resources on the stage of data collection and harmonization.

Next to the limitation on available data, computational limitations also prevented me from estimating several algorithms and, possibly, creating an ensemble prediction in chapter 5. Ensemble predictions are often used to improve model performance. As performance was very good already, I judge this limitation to be minor. Nevertheless, comparing the sensitivity in results related to different algorithms to the sensitivity in results related to the background data approach would have been an interesting contribution to the species distribution modelling literature. Training multiple algorithms on the datasets would be straightforward but would require a High-Performance Computing Cluster. The cluster would require additional funding that, due to the memory requirements that result out of the data's high dimensionality, would likely be expensive.

While work presented in this thesis did partially succeed in capturing spatial and temporal aspects, a critical limitation remains our ability to model feedback mechanisms between the biological and economic systems. One of the underlying motivations for chapter 2 was the idea that farmers' management influences the pest pressure in the vicinity of the farm which consequently might influence neighbors' operational performance. However, in chapters 3 and 4 the modelling pipeline was clearly compartmentalized and unidirectional. First, climatic suitability mapping provided a binary indication whether spread can occur in locations. Next, spread simulations dispersed through suitable habitat under the assumption that control efforts failed. Lastly, the impact to farmers was computed. Arguably, the unidirectional integration of the three models is a simplification. In reality, the continuous score for the climatic suitability might result in different local spread rates, knowledge on the climatic suitability might influence farmers' preparedness, the realized speed with which spread occurs might influence farmers' responses and management strategies which in turn would affect spread rates as well. More sophisticated spread models as employed here tend to acknowledge the role of the population density in spread [171, 365]. Arguably, farmers' actions can influence population density and consequently spread [221].

The inherent interconnectedness between the biological and economic models clearly requires further work [72, 366, 367]. Human responses to epidemics can significantly alter their progression [40, 368]. Ignorance of such feedback mechanisms can result in institutional failures [369]. Certainly, this fact should raise optimism as it essentially means that analyses of appropriately integrated socio-ecological systems may provide improved strategies for pest control. However, this of course also challenges modelers by significantly increasing the complexity of our work. The need for a better integration of epidemiological and economic models has already been stressed for over a decade [40, 321]. Previous work that aimed to account for such feedbacks often used agent based approaches [97, 370, 371]. However, in my view, models including these feedbacks quickly become unintelligible and data limitations often turn such efforts into pure modelling exercise with limited to no ability to validate whether the simulated patterns conform to the real world or merely to our subjective understanding of it. A calibration of more complex models which include feedbacks

requires more in-depth data that allows to tease out the biological and economic relationships. I believe georeferenced economic data would be a great start for this.

Future work could investigate legislative solutions for pest control that might incorporate spatial and temporal variability in better ways as done to date. It might be possible to reach parametric solutions, like weather index insurances [372–374], in which environmental conditions are directly considered through a computation of an pest-risk-related index that is spatially and temporally variable and used to tailor the applicable legislation to the individuals' condition.

The DAISIE<sup>6</sup> project resulted in a database comprising over 11,000 invasive species already present in Europe. While the database intended to hold information on the expected ecological and economic impacts, the researchers were able to compile information on these impacts for 11 and 13 percent of species, respectively [52]. Clearly, the number of new introductions in combination with the depth in analysis required for each individual species results in an overwhelming absence of information that hampers decision making and leaves society in the dark on the realized externalities. I believe there are two approaches to this, which are not necessarily mutually exclusive, regulators could take. First, more scientists could be employed to conduct pest risk assessments. Second, an arsenal of generalized models could be developed to speed up species-specific analyses. Here, an approach comparable to chapter 5 could provide a generic map for potential establishment for a range of species, possibly for different taxonomic groups, and spread models similar to Hudgins et al. [201] could provide a generalized dispersal model. Having both ready to go when faced with an emerging threat could speed up the required risk assessments, albeit at the expense of precision in results. Having an arsenal of generalized maps and models would reflect a top-down approach that enables quick decision making in data scarce situations. This could be an important avenue for future work.

The control of pests is a notoriously difficult task. The inherent randomness of hazardous pest epidemics will continue to challenge risk managers in the future. However, further improvements in the quality and breadth of georeferenced data will increase our chances of establishing sound models that support our decision making. Future

---

<sup>6</sup> Delivering Alien Invasive Species Inventories for Europe

work must therefore continue to push the envelope of what is currently possible in terms of data collection. This need ranges from improving existing methodologies that measure or predict environmental properties, over developing new technologies that obtain additional variables of interest, to increasing the number of feet on the ground for surveying. While challenging, I believe all of this is possible through collective efforts between decision makers, scientists, and citizens.

## 6.5 Main Conclusions

- Neighbors' characteristics associate with farmers' managerial performance. However, *how* the characteristics associate with farmers' performance depends on the definition of the neighborhood (chapter 2).
- Farmers do not operate in isolation from their peers because the outcome of their efforts is partially determined by neighbors' actions (chapter 2, 3).
- The assessment of pest related losses in investments for perennial hosts can be achieved by computing losses in foregone annuities (chapter 3).
- The vast majority of the European olive cultivation is within climatically suitable territory for the establishment of *Xylella fastidiosa* subspecies *pauca* (chapter 3).
- For Italy, across the considered spread rates the potential economic impact from *Xylella fastidiosa* subspecies *pauca* to olive growers over 50 years ranges from 1.86 to 5.17 billion Euro if the current control measures were to fail and replanting with a resistant cultivar would not be feasible (chapter 3).
- Irrespective of the disease spread rate and the ability to replant with resistant cultivars, projections of potential future economic impacts from *Xylella fastidiosa* subspecies *pauca* to olive growers in affected countries are sizable and warrant regulatory response (chapter 3).
- In Europe, introductions of *Xylella fastidiosa* subspecies *pauca* into Spain would likely result in the most drastic economic impact (chapter 3, 4).

- Capturing economic heterogeneity by stratification, does not only refine the overall estimate of pest impact but also signals whether unequal consequences from pest spread arise to the different strata (chapter 3, 4).
- Due to the inelastic supply and demand, consumers are the main beneficiaries of the control of *Xylella fastidiosa* subspecies *pauca* (chapter 4).
- Sensitivity analyses generate actionable insights on areas that should be prioritized in management, market characteristics that must be promoted, and data gaps that need to be addressed (chapter 3, 4, 5).
- Joint analyses of various species may address the overwhelming absence of information on risk maps for invasive species (chapter 5).
- Hotspot maps can identify areas that are at higher risk of invasive pest introductions (chapter 5).
- In Europe, the BeNeLux states, Northern Italy, the Northern Balkans, and the United Kingdom, and areas around container ports such as Antwerp, London, Rijeka, and Saint Petersburg are at higher risk for invasive pest introductions (chapter 5).
- In species distribution models, scholars should not exclusively rely on machine learning performance as a measure of model correctness (chapter 5).
- Systematic species survey data comprising also true absence are required to be able to disentangle reporting bias from true effects and thereby allow for the analysis of the anthropogenic involvement in invasive pest introduction (chapter 5).

### 6.5.1 Main Policy and Research Implications

- Policy impact analyses, for example on the effects of farm subsidies, should also investigate possible community effects and not only direct impacts (chapter 2).
- Analyses of optimal pest control should be expanded beyond individual farm-level (chapter 2, 3).

- Incorporating information on host locations into pest risk assessments is crucial, especially if cultivation may be spatially clustered (chapter 3, 4).
- The inclusion of price responses to supply changes into pest risk assessments is important for deriving insights on economic dependencies and diverging incentives among stakeholders for controlling pest spread (chapter 3, 4).
- Acknowledging environmental heterogeneity in pest risk assessments is important not only for the prediction of the suitability of establishment, but also for simulations of pest spread and, consequently, economic impact (chapter 3, 4, 5).
- An arsenal of generalized models may speed up pest risk assessments, albeit at the expense of precision in results (chapter 3, 4, 5).
- Pest management strategies should take spatial heterogeneity into account by acknowledging geographic information on areas' pest risk, as well as topographic, climatic, and economic conditions (chapter 3, 4, 5).
- Spatial field-level data on host cultivation is needed to improve pest risk assessments (chapter 3, 4, 5)
- A precise spatial indexation of economic data is critical to improve modelling work (chapter 2, 3, 4, 5).

# References

- [1] Cohen, J. E. Human population: the next half century. *Science* **302**, 1172–1175 (2003). URL <https://www.ncbi.nlm.nih.gov/pubmed/14615528>.
- [2] Savary, S. *et al.* The global burden of pathogens and pests on major food crops. *Nature Ecology & Evolution* **3**, 430–439 (2019). URL <http://www.nature.com/articles/s41559-018-0793-y>.
- [3] Oerke, E. C. Crop losses to pests. *The Journal of Agricultural Science* **144** (2005). URL <https://doi.org/10.1017/S0021859605005708>.
- [4] Cook, D. C. Benefit cost analysis of an import access request. *Food Policy* **33**, 277–285 (2008). URL <https://doi.org/10.1016/j.foodpol.2007.09.002>.
- [5] Hulme, P. E. Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of Applied Ecology* **46**, 10–18 (2009). URL <http://doi.wiley.com/10.1111/j.1365-2664.2008.01600.x>.
- [6] Hulme, P. E. Unwelcome exchange: International trade as a direct and indirect driver of biological invasions worldwide. *One Earth* **4**, 666–679 (2021). URL <https://doi.org/10.1016/j.oneear.2021.04.015>.
- [7] Levine, J. M. & D'Antonio, C. M. Forecasting biological invasions with increasing international trade. *Conservation Biology* **17**, 322–

- 326 (2003). URL <https://doi.org/10.1046/j.1523-1739.2003.02038.x>.
- [8] Levine, J. M. *et al.* Mechanisms underlying the impacts of exotic plant invasions. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270**, 775–781 (2003). URL <https://doi.org/10.1098/rspb.2003.2327>.
- [9] Paini, D. R. *et al.* Global threat to agriculture from invasive species. *Proceedings of the National Academy of Sciences* **113**, 7575–7579 (2016). URL <http://doi.org/10.1073/pnas.1602205113>.
- [10] Mangen, M.-J. J. & Burrell, A. M. Who gains, who loses? Welfare effects of classical swine fever epidemics in the Netherlands. *European Review of Agriculture Economics* **30**, 125–154 (2003). URL <https://doi.org/10.1093/erae/30.2.125>.
- [11] Knight-Jones, T. J. & Rushton, J. The economic impacts of foot and mouth disease - what are they, how big are they and where do they occur? *Prev Vet Med* **112**, 161–173 (2013). URL <https://www.ncbi.nlm.nih.gov/pubmed/23958457>.
- [12] Lehner, A. & Stephan, R. Microbiological, epidemiological, and food safety aspects of enterobacter sakazakii. *Journal of food protection* **67**, 2850–2857 (2004). URL <https://doi.org/10.4315/0362-028X-67.12.2850>.
- [13] Cook, D. C. *et al.* Biosecurity and yield improvement technologies are strategic complements in the fight against food insecurity. *PLoS One* **6** (2011). URL <https://doi.org/10.1371/journal.pone.0026084>.
- [14] Wells, J. M. *et al.* *Xylella fastidiosa* gen. nov., sp. nov: Gram-Negative, Xylem-Limited, Fastidious Plant Bacteria Related to *Xanthomonas* spp. *International Journal of Systematic Bacteriology* **37**, 136–143 (1987). URL <https://doi.org/10.1099/00207713-37-2-136>.
- [15] EFSA. Update of the *Xylella* spp. host plant database. *EFSA Journal* **16** (2018). URL <http://doi.wiley.com/10.2903/j.efsa.2018.5408>.
- [16] European Commission. Commission database of host plants found to be susceptible to *Xylella fastidiosa* in the Union territory - update 9. *Ref. Ares(2017)3824160* (2017).

- [17] Hopkins, D. L. & Purcell, A. H. Xylella fastidiosa : Cause of Pierce's Disease of Grapevine and Other Emergent Diseases. *Plant Disease* **86**, 1056–1066 (2002). URL <http://apsjournals.apsnet.org/doi/10.1094/PDIS.2002.86.10.1056>.
- [18] Chatterjee, S., Almeida, R. P. P. & Lindow, S. Living in two worlds: the plant and insect lifestyles of Xylella fastidiosa. *Annu Rev Phytopathol* **46**, 243–271 (2008). URL <https://www.ncbi.nlm.nih.gov/pubmed/18422428>.
- [19] Cornara, D., Bosco, D. & Fereres, A. Philaenus spumarius: when an old acquaintance becomes a new threat to European agriculture. *Journal of Pest Science* **91**, 957–972 (2018). URL <http://link.springer.com/10.1007/s10340-018-0966-0>.
- [20] Hopkins, D. L. Xylella Fastidiosa: Xylem-Limited Bacterial Pathogen of Plants. *Annual Review of Phytopathology* **27**, 271–290 (1989). URL <https://doi.org/10.1146/annurev.py.27.090189.001415>.
- [21] European Food Safety Authority. Scientific Opinion on the risk to plant health posed by Xylella fastidiosa in the EU territory, with the identification and evaluation of risk reduction options. *EFSA Journal* **13** (2015). URL <https://www.efsa.europa.eu/en/efsajournal/pub/3989>.
- [22] Saponari, M. *et al.* Isolation and pathogenicity of Xylella fastidiosa associated to the olive quick decline syndrome in southern Italy. *Scientific Reports* **7**, 17723 (2017). URL <http://www.nature.com/articles/s41598-017-17957-z>.
- [23] Purcell, A. Paradigms: examples from the bacterium Xylella fastidiosa. *Annu Rev Phytopathol* **51**, 339–356 (2013). URL <https://www.ncbi.nlm.nih.gov/pubmed/23682911>.
- [24] Saponari, M. *et al.* Pilot project on Xylella fastidiosa to reduce risk assessment uncertainties. *EFSA Supporting Publications* **13** (2016). URL <https://doi.org/10.2903/sp.efsa.2016.EN-1013>.
- [25] Almeida, R. P. P. Can Apulia's olive trees be saved? *Science* **353**, 346–348 (2016). URL <http://www.sciencemag.org/cgi/doi/10.1126/science.aaf9710>.
- [26] Stokstad, E. Food security: Italy's olives under siege. *Science* **348**, 620 (2015). URL <https://www.ncbi.nlm.nih.gov/pubmed/>

- 25953988.
- [27] Sharma, A. *et al.* Worldwide pesticide usage and its impacts on ecosystem. *SN Applied Sciences* **1**, 1446 (2019). URL <http://link.springer.com/10.1007/s42452-019-1485-1>.
- [28] Kohler, H. R. & Triebskorn, R. Wildlife Ecotoxicology of Pesticides: Can We Track Effects to the Population Level and Beyond? *Science* **341**, 759–765 (2013). URL <http://www.sciencemag.org/cgi/doi/10.1126/science.1237591>.
- [29] Oude Lansink, A. & Carpentier, A. Damage Control Productivity: An Input Damage Abatement Approach. *Journal of Agricultural Economics* **52**, 11–22 (2008). URL <http://doi.wiley.com/10.1111/j.1477-9552.2001.tb00935.x>.
- [30] Oude Lansink, A. & Silva, E. Non-Parametric Production Analysis of Pesticides Use in the Netherlands. *Journal of Productivity Analysis* **21**, 49–65 (2004). URL <http://link.springer.com/10.1023/B:PROD.0000012452.97645.30>.
- [31] Skevas, T., Stefanou, S. E. & Oude Lansink, A. Pesticide use, environmental spillovers and efficiency: A DEA risk-adjusted efficiency approach applied to Dutch arable farming. *European Journal of Operational Research* **237**, 658–664 (2014). URL <https://linkinghub.elsevier.com/retrieve/pii/S0377221714000678>.
- [32] Tveteras, R. & Battese, G. E. Agglomeration Externalities, Productivity, and Technical Inefficiency. *Journal of Regional Science* **46**, 605–625 (2006). URL <http://doi.wiley.com/10.1111/j.1467-9787.2006.00470.x>.
- [33] Parsa, S. *et al.* Obstacles to integrated pest management adoption in developing countries. *Proceedings of the National Academy of Sciences* **111**, 3889–3894 (2014). URL <https://doi.org/10.1073/pnas.1312693111>.
- [34] Lapple, D. & Kelley, H. Spatial dependence in the adoption of organic drystock farming in Ireland. *European Review of Agricultural Economics* **42**, 315–337 (2015). URL <https://doi.org/10.1093/erae/jbu024>.
- [35] Bell, A., Zhang, W. & Nou, K. Pesticide use and cooperative management of natural enemy habitat in a framed field experiment. *Agricultural Systems* **143**, 1–13 (2016).

- URL <https://linkinghub.elsevier.com/retrieve/pii/S0308521X15300524>.
- [36] Horan, R. D., Perrings, C., Lupi, F. & Bulte, E. H. Biological Pollution Prevention Strategies under Ignorance: The Case of Invasive Species. *American Journal of Agricultural Economics* **84**, 1303–1310 (2002). URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-8276.00394>.
- [37] Fenichel, E. P., Richards, T. J. & Shanafelt, D. W. The Control of Invasive Species on Private Property with Neighbor-to-Neighbor Spillovers. *Environmental and Resource Economics* **59**, 231–255 (2014). URL <http://link.springer.com/10.1007/s10640-013-9726-z>.
- [38] Epanchin-Niell, R. S. *et al.* Controlling invasive species in complex social landscapes. *Frontiers in Ecology and the Environment* **8**, 210–216 (2010). URL <http://doi.wiley.com/10.1890/090029>.
- [39] Epanchin-Niell, R. S. & Wilen, J. E. Optimal spatial control of biological invasions. *Journal of Environmental Economics and Management* **63**, 260–270 (2012). URL <https://linkinghub.elsevier.com/retrieve/pii/S0095069611001392>.
- [40] Perrings, C. *et al.* Merging Economics and Epidemiology to Improve the Prediction and Management of Infectious Disease. *EcoHealth* **11**, 464–475 (2014). URL <http://link.springer.com/10.1007/s10393-014-0963-6>.
- [41] Hulme, P. E., Pyšek, P., Nentwig, W. & Vilà, M. Will Threat of Biological Invasions Unite the European Union? *Science* **324**, 40–41 (2009). URL <https://doi.org/10.1126/science.1171111>.
- [42] Vettraino, A., Potting, R. & Raposo, R. EU Legislation on Forest Plant Health: An Overview with a Focus on *Fusarium circinatum*. *Forests* **9**, 568 (2018). URL <http://www.mdpi.com/1999-4907/9/9/568>.
- [43] Olson, L. J. The Economics of Terrestrial Invasive Species: A Review of the Literature. *Agricultural and Resource Economics Review* **35**, 178–194 (2006). URL [https://www.cambridge.org/core/product/identifier/S1068280500010145/type/journal\\_article](https://www.cambridge.org/core/product/identifier/S1068280500010145/type/journal_article).
- [44] De Benedictis, L. & Tajoli, L. The World Trade Network. *The World Economy* **34**, 1417–1454 (2011). URL <http://doi.wiley.com>.

- com/10.1111/j.1467-9701.2011.01360.x.
- [45] Maxwell, A., Vettraino, A. M., Eschen, R. & Andjic, V. International Plant Trade and Biosecurity. In *Horticulture: Plants for People and Places, Volume 3*, 1171–1195 (Springer Netherlands, Dordrecht, 2014). URL [http://link.springer.com/10.1007/978-94-017-8560-0\\_9](http://link.springer.com/10.1007/978-94-017-8560-0_9).
- [46] Perrings, C., Burgiel, S., Lonsdale, M., Mooney, H. & Williamson, M. International cooperation in the solution to trade-related invasive species risks. *Annals of the New York Academy of Sciences* **1195**, 198–212 (2010). URL <http://doi.wiley.com/10.1111/j.1749-6632.2010.05453.x>.
- [47] Van Den Brink, T. The Impact of EU Legislation on National Legal Systems: Towards a New Approach to EU – Member State Relations. *Cambridge Yearbook of European Legal Studies* **19**, 211–235 (2017). URL [https://www.cambridge.org/core/product/identifier/S1528887017000027/type/journal\\_article](https://www.cambridge.org/core/product/identifier/S1528887017000027/type/journal_article).
- [48] Keller, R. P., Geist, J., Jeschke, J. M. & Kühn, I. Invasive species in Europe: ecology, status, and policy. *Environmental Sciences Europe* **23**, 23 (2011). URL <https://enveurope.springeropen.com/articles/10.1186/2190-4715-23-23>.
- [49] Oude Lansink, A. Public and private roles in plant health management. *Food Policy* **36**, 166–170 (2011). URL <http://linkinghub.elsevier.com/retrieve/pii/S0306919210001090>.
- [50] Cieslik, K. *et al.* The role of ICT in collective management of public bads: The case of potato late blight in Ethiopia. *World Development* **140**, 105366 (2021). URL <https://linkinghub.elsevier.com/retrieve/pii/S0305750X20304940>.
- [51] European Commission. Towards an EU strategy on invasive species. *Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions* **789** (2008). URL <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52008DC0789&from=EN>.
- [52] Vilà, M. *et al.* How well do we understand the impacts of alien species on ecosystem services? A pan-European, cross-taxa assessment. *Frontiers in Ecology and the Environment* **8**, 135–144 (2010). URL <http://doi.wiley.com/10.1890/080083>.

- [53] Simon, H. A. Organizations and Markets. *Journal of Economic Perspectives* **5**, 25–44 (1991). URL <https://pubs.aeaweb.org/doi/10.1257/jep.5.2.25>.
- [54] Massell, B. F. Price Stabilization and Welfare. *The Quarterly Journal of Economics* **83**, 284 (1969). URL <https://doi.org/10.2307/1883084>.
- [55] Myerson, R. B. *Game Theory* (Harvard University Press, 2013).
- [56] Levitt, T. The globalization of markets. In *Readings in international business: a decision approach*, 249–252 (MIT Press, 1993). URL [https://edisciplinas.usp.br/pluginfile.php/4180851/mod\\_resource/content/1/Levitt%20Globalization%20Markets.pdf](https://edisciplinas.usp.br/pluginfile.php/4180851/mod_resource/content/1/Levitt%20Globalization%20Markets.pdf).
- [57] Sapir, A. Regional Integration in Europe. *The Economic Journal* **102**, 1491 (1992). URL <https://academic.oup.com/ej/article/102/415/1491-1506/5157140>.
- [58] Gereffi, G., Humphrey, J., Kaplinsky, R. & Sturgeon\*, T. J. Introduction: Globalisation, Value Chains and Development. *IDS Bulletin* **32**, 1–8 (2001). URL <http://doi.wiley.com/10.1111/j.1759-5436.2001.mp32003001.x>.
- [59] Ivanov, D. Simulation-based ripple effect modelling in the supply chain. *International Journal of Production Research* **55**, 2083–2101 (2017). URL <https://www.tandfonline.com/doi/full/10.1080/00207543.2016.1275873>.
- [60] Roy, J. The effect of employment protection legislation on international trade. *Economic Modelling* **94**, 221–234 (2021). URL <https://linkinghub.elsevier.com/retrieve/pii/S0264999320312220>.
- [61] World Trade Organization. World Trade Report. Tech. Rep., World Trade Organization (2011). URL [https://www.wto.org/english/res\\_e/booksp\\_e/anrep\\_e/world\\_trade\\_report11\\_e.pdf](https://www.wto.org/english/res_e/booksp_e/anrep_e/world_trade_report11_e.pdf).
- [62] Abbott, P. C. & Kallio, P. K. S. Implications of Game Theory for International Agricultural Trade. *American Journal of Agricultural Economics* **78**, 738–744 (1996). URL <https://onlinelibrary.wiley.com/doi/abs/10.2307/1243297>.
- [63] McMillan, J. A Game-Theoretic View of International Trade Negotiations: Implications for the Developing Countries. In *Developing Countries and the Global Trading System*, 26–44 (Palgrave

- Macmillan UK, London, 1989). URL [http://link.springer.com/10.1007/978-1-349-20417-5\\_2](http://link.springer.com/10.1007/978-1-349-20417-5_2).
- [64] Hertel, T., Hummels, D., Ivanic, M. & Keeney, R. How confident can we be of CGE-based assessments of Free Trade Agreements? *Economic Modelling* **24**, 611–635 (2007). URL <https://linkinghub.elsevier.com/retrieve/pii/S026499930700003X>.
- [65] Tobler, W. R. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography* **46**, 234 (1970). URL <https://www.jstor.org/stable/143141?origin=crossref>.
- [66] Weiss, M. D. Precision Farming and Spatial Economic Analysis: Research Challenges and Opportunities. *American Journal of Agricultural Economics* **78**, 1275 (1996). URL <https://doi.org/10.2307/1243506>.
- [67] Anselin, L. Thirty years of spatial econometrics. *Papers in Regional Science* **89**, 3–25 (2010). URL <http://doi.wiley.com/10.1111/j.1435-5957.2010.00279.x>.
- [68] Cotteleer, G., Gardebroek, C. & Luijt, J. Market Power in a GIS-Based Hedonic Price Model of Local Farmland Markets. *Land Economics* **84**, 573–592 (2008). URL <http://le.uwpress.org/cgi/doi/10.3368/le.84.4.573>.
- [69] Chambers, R. G., Hailu, A. & Quiggin, J. Event-specific data envelopment models and efficiency analysis. *Australian Journal of Agricultural and Resource Economics* **55**, 90–106 (2011). URL <http://doi.wiley.com/10.1111/j.1467-8489.2010.00517.x>.
- [70] Turchin, P. *Complex Population Dynamics: A Theoretical/Empirical Synthesis (MPB-35)* (Princeton University Press, 2003). URL [www.jstor.org/stable/j.ctt24hqkz](http://www.jstor.org/stable/j.ctt24hqkz).
- [71] Knipling, E. F. Regional Management of the Fall Armyworm: A Realistic Approach? *The Florida Entomologist* **63**, 468 (1980). URL <https://www.jstor.org/stable/3494531?origin=crossref>.
- [72] Topping, C. J. *et al.* Towards a landscape scale management of pesticides: ERA using changes in modelled occupancy and abundance to assess long-term population impacts of pesticides. *Sci Total Environ* **537**, 159–169 (2015). URL <https://www.ncbi.nlm.nih.gov/pubmed/26318547>.

- [73] Bakker, L., Sok, J., Van Der Werf, W. & Bianchi, F. Kicking the habit: What makes and breaks farmers' intentions to reduce pesticide use? *Ecological Economics* **180**, 106868 (2021). URL <https://doi.org/10.1016/j.ecolecon.2020.106868>.
- [74] Milne, A. E., Gottwald, T., Parnell, S. R., Alonso Chavez, V. & van den Bosch, F. What makes or breaks a campaign to stop an invading plant pathogen? *PLOS Computational Biology* **16**, e1007570 (2020). URL <https://dx.plos.org/10.1371/journal.pcbi.1007570>.
- [75] Andersen, R. *et al.* Environmental control and spatial structures in peatland vegetation. *Journal of Vegetation Science* **22**, 878–890 (2011). URL <http://doi.wiley.com/10.1111/j.1654-1103.2011.01295.x>.
- [76] Keitt, T. H., Bjørnstad, O. N., Dixon, P. M. & Citron-Pousty, S. Accounting for spatial pattern when modeling organism-environment interactions. *Ecography* **25**, 616–625 (2002). URL <http://doi.wiley.com/10.1034/j.1600-0587.2002.250509.x>.
- [77] Miller, J., Franklin, J. & Aspinall, R. Incorporating spatial dependence in predictive vegetation models. *Ecological Modelling* **202**, 225–242 (2007). URL <https://linkinghub.elsevier.com/retrieve/pii/S0304380006006247>.
- [78] Tilman, D. *Resource competition and community structure* (Princeton University Press, 1982). URL <https://doi.org/10.1515/9780691209654>.
- [79] Dark, S.J. The biogeography of invasive alien plants in California: an application of GIS and spatial regression analysis. *Diversity and Distributions* **10**, 1–9 (2004). URL <http://doi.wiley.com/10.1111/j.1472-4642.2004.00054.x>.
- [80] Albers, H. J., Fischer, C. & Sanchirico, J. N. Invasive species management in a spatially heterogeneous world: Effects of uniform policies. *Resource and Energy Economics* **32**, 483–499 (2010). URL <https://linkinghub.elsevier.com/retrieve/pii/S0928765510000217>.
- [81] Blackwood, J., Hastings, A. & Costello, C. Cost-effective management of invasive species using linear-quadratic control. *Ecological Economics* **69**, 519–527 (2010). URL <https://linkinghub.elsevier.com/retrieve/pii/S0921800909003590>.

- [82] Burnett, K. M., Kaiser, B. A. & Roumasset, J. A. Invasive Species Control over Space and Time: *Miconia calvescens* on Oahu, Hawaii. *Journal of Agricultural and Applied Economics* **39**, 125–132 (2007). URL [https://www.cambridge.org/core/product/identifi er/S1074070800028996/type/journal\\_article](https://www.cambridge.org/core/product/identifi er/S1074070800028996/type/journal_article).
- [83] Travis, J. M. J. & Park, K. J. Spatial structure and the control of invasive alien species. *Animal Conservation* **7**, 321–330 (2004). URL <http://doi.wiley.com/10.1017/S1367943004001507>.
- [84] Sismondo, S. Models, Simulations, and Their Objects. *Science in Context* **12**, 247–260 (1999). URL [https://www.cambridge.org/core/product/identifi er/S0269889700003409/type/journal\\_article](https://www.cambridge.org/core/product/identifi er/S0269889700003409/type/journal_article).
- [85] Jarnevi ch, C. S., Stohlgren, T. J., Kumar, S., Morisette, J. T. & Holcombe, T. R. Caveats for correlative species distribution modeling. *Ecological Informatics* **29**, 6–15 (2015). URL <https://linkinghub.elsevier.com/retrieve/pii/S1574954115000977>.
- [86] Koch, K. *Parameter estimation and hypothesis testing in linear models* (Springer Science & Business Media, 2013).
- [87] Feddema, J. J. The Importance of Land-Cover Change in Simulating Future Climates. *Science* **310**, 1674–1678 (2005). URL <https://doi.org/10.1126/science.1118160>.
- [88] Perrings, C. Mitigation and adaptation strategies for the control of biological invasions. *Ecological Economics* **52**, 315–325 (2005). URL <https://linkinghub.elsevier.com/retrieve/pii/S0921800904003052>.
- [89] Eiswerth, M. E. & van Kooten, G. C. The Economics of Invasive Species Management: Uncertainty, Economics, and the Spread of an Invasive Plant Species. *American Journal of Agricultural Economics* **84**, 1317–1322 (2002). URL <https://doi.org/10.1111/1467-8276.00396>.
- [90] Douma, J. *et al.* Pathway models for analysing and managing the introduction of alien plant pests - an overview and categorization. *Ecological Modelling* **339**, 58–67 (2016). URL <https://linkinghub.elsevier.com/retrieve/pii/S030438001630312X>.
- [91] Loomis, R. S. & Adams, S. S. Integrative Analyses of Host-Pathogen Relations. *Annual Review of Phytopathology* **21**, 341–

- 362 (1983). URL <https://doi.org/10.1146/annurev.py.21.090183.002013>.
- [92] Olson, L. J. & Roy, S. On Prevention and Control of an Uncertain Biological Invasion. *Review of Agricultural Economics* **27**, 491–497 (2005). URL <https://doi.org/10.1111/j.1467-9353.2005.00249.x>.
- [93] Lustig, A. *et al.* A modeling framework for the establishment and spread of invasive species in heterogeneous environments. *Ecol Evol* **7**, 8338–8348 (2017). URL <https://www.ncbi.nlm.nih.gov/pubmed/29075453>.
- [94] Khanna, M., Isik, M. & Winter-Nelson, A. Investment in site-specific crop management under uncertainty: implications for nitrogen pollution control and environmental policy. *Agricultural Economics* **24**, 9–12 (2000). URL <http://doi.wiley.com/10.1111/j.1574-0862.2000.tb00089.x>.
- [95] Skevas, T. & Serra, T. Derivation of netput shadow prices under different levels of pest pressure. *Journal of Productivity Analysis* **48**, 25–34 (2017). URL <http://link.springer.com/10.1007/s11123-017-0507-5>.
- [96] Cook, D. C., Thomas, M. B., Cunningham, S. A., Anderson, D. L. & De Barro, P. J. Predicting the economic impact of an invasive species on an ecosystem service. *Ecological Applications* **17**, 1832–1840 (2007). URL <http://doi.wiley.com/10.1890/06-1632.1>.
- [97] Carrasco, L. R. *et al.* Towards the integration of spread and economic impacts of biological invasions in a landscape of learning and imitating agents. *Ecological Economics* **76**, 95–103 (2012). URL <https://doi.org/10.1016/j.ecolecon.2012.02.009>.
- [98] Rimbaud, L. *et al.* Improving Management Strategies of Plant Diseases Using Sequential Sensitivity Analyses. *Phytopathology*® **109**, 1184–1197 (2019). URL <https://apsjournals.apsnet.org/doi/10.1094/PHYTO-06-18-0196-R>.
- [99] Venette, R. C. *et al.* Pest Risk Maps for Invasive Alien Species: A Roadmap for Improvement. *BioScience* **60**, 349–362 (2010). URL <https://doi.org/10.1525/bio.2010.60.5.5>.
- [100] Lichtenberg, E. & Zilberman, D. The Econometrics of Damage Control: Why Specification Matters. *American Journal of Agri-*

- cultural Economics* **68**, 261 (1986). URL <https://doi.org/10.2307/1241427>.
- [101] Skevas, T., Oude Lansink, A. & Stefanou, S. E. Measuring technical efficiency in the presence of pesticide spillovers and production uncertainty: The case of Dutch arable farms. *European Journal of Operational Research* **223**, 550–559 (2012). URL <https://linkinghub.elsevier.com/retrieve/pii/S0377221712004894>.
- [102] Farrell, M. J. The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society. Series A (General)* **120**, 253 (1957). URL <http://www.jstor.org/stable/2343100?origin=crossref>.
- [103] Fusco, E. & Vidoli, F. Spatial stochastic frontier models: controlling spatial global and local heterogeneity. *International Review of Applied Economics* **27**, 679–694 (2013). URL <http://www.tandfonline.com/doi/abs/10.1080/02692171.2013.804493>.
- [104] Vidoli, F. & Canello, J. Controlling for spatial heterogeneity in nonparametric efficiency models: An empirical proposal. *European Journal of Operational Research* **249**, 771–783 (2016). URL <https://linkinghub.elsevier.com/retrieve/pii/S0377221715009765>.
- [105] Druska, V. & Horrace, W. C. Generalized Moments Estimation for Spatial Panel Data: Indonesian Rice Farming. *American Journal of Agricultural Economics* **86**, 185–198 (2004). URL <https://www.jstor.org/stable/3697883>.
- [106] Glass, A., Kenjegalieva, K. & Sickles, R. C. A spatial autoregressive stochastic frontier model for panel data with asymmetric efficiency spillovers. *Journal of Econometrics* **190**, 289–300 (2016). URL <https://linkinghub.elsevier.com/retrieve/pii/S0304407615001803>.
- [107] Tsionas, E. G. & Michaelides, P. G. A Spatial Stochastic Frontier Model with Spillovers: Evidence for Italian Regions. *Scottish Journal of Political Economy* **63**, 243–257 (2016). URL <http://doi.wiley.com/10.1111/sjpe.12081>.
- [108] Pede, V. O., Areal, F. J., Singbo, A., McKinley, J. & Kajisa, K. Spatial dependency and technical efficiency: an application of a Bayesian stochastic frontier model to irrigated and rainfed rice

- farmers in Bohol, Philippines. *Agricultural Economics* **49**, 301–312 (2018). URL <http://doi.wiley.com/10.1111/agec.12417>.
- [109] Areal, F. J., Balcombe, K. & Tiffin, R. Integrating spatial dependence into Stochastic Frontier Analysis. *Australian Journal of Agricultural and Resource Economics* **56**, 521–541 (2012). URL <http://doi.wiley.com/10.1111/j.1467-8489.2012.00597.x>.
- [110] Martínez-Victoria, M., Maté-Sánchez-Val, M. & Oude Lansink, A. Spatial dynamic analysis of productivity growth of agri-food companies. *Agricultural Economics* *agec.12486* (2019). URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/agec.12486>.
- [111] Skevas, T. & Grashuis, J. Technical efficiency and spatial spillovers: Evidence from grain marketing cooperatives in the US Midwest. *Agribusiness* *agr.21617* (2019). URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/agr.21617>.
- [112] Skevas, I. & Oude Lansink, A. Dynamic inefficiency and spatial spillovers in dutch dairy farming. *Journal of Agricultural Economics* (2020). URL <https://doi.org/10.1111/1477-9552.12369>.
- [113] Skevas, I. Inference in the spatial autoregressive efficiency model with an application to dutch dairy farms. *European Journal of Operational Research* **283**, 356–364 (2020). URL <https://doi.org/10.1016/j.ejor.2019.10.033>.
- [114] Gibbons, S. & Overman, H. G. Mostly Pointless Spatial Econometrics? *Journal of Regional Science* **52**, 172–191 (2012). URL <http://doi.wiley.com/10.1111/j.1467-9787.2012.00760.x>.
- [115] McMillen, D. P. Perspectives on Spatial Econometrics: Linear Smoothing with Structured Models. *Journal of Regional Science* **52**, 192–209 (2012). URL <http://doi.wiley.com/10.1111/j.1467-9787.2011.00746.x>.
- [116] Halleck Vega, S. & Elhorst, J. P. The SLX Model. *Journal of Regional Science* **55**, 339–363 (2015). URL <http://doi.wiley.com/10.1111/jors.12188>.
- [117] Banker, R. D., Charnes, A. & Cooper, W. W. Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science* **30**, 1078–1092 (1984). URL <http://pubsonline.informs.org/doi/abs/10.1287/mnsc.30.9.1078>.

- [118] Chambers, R. G., Chung, Y. & Färe, R. Profit, Directional Distance Functions, and Nerlovian Efficiency. *Journal of Optimization Theory and Applications* **98**, 351–364 (1998). URL <http://link.springer.com/10.1023/A:1022637501082>.
- [119] Färe, R. & Grosskopf, S. Theory and Application of Directional Distance Functions. *Journal of Productivity Analysis* **13**, 99–103 (2000). URL <https://doi.org/10.1023/A:1007844628920>.
- [120] Simar, L. & Wilson, P. W. Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics* **136**, 31–64 (2007). URL <http://linkinghub.elsevier.com/retrieve/pii/S0304407605001594>.
- [121] Kapelko, M., Lansink, A. O. & Stefanou, S. E. Analyzing the impact of investment spikes on dynamic productivity growth. *Omega* **54**, 116–124 (2015). URL <https://doi.org/10.1016/j.omega.2015.01.010>.
- [122] Reztis, A. N. & Kalantzi, M. A. Investigating technical efficiency and its determinants by data envelopment analysis: An application in the greek food and beverages manufacturing industry. *Agribusiness* **32**, 254–271 (2016). URL <https://doi.org/10.1002/agr.21432>.
- [123] Singbo, A. G. & Lansink, A. O. Lowland farming system inefficiency in benin (west africa): directional distance function and truncated bootstrap approach. *Food Security* **2**, 367–382 (2010). URL <https://doi.org/10.1007/s12571-010-0086-z>.
- [124] Singbo, A. G., Oude Lansink, A. & Emvalomatis, G. Estimating farmers’ productive and marketing inefficiency: an application to vegetable producers in Benin. *Journal of Productivity Analysis* **42**, 157–169 (2014). URL <http://link.springer.com/10.1007/s11123-014-0391-1>.
- [125] Storm, H., Mittenzwei, K. & Heckelei, T. Direct Payments, Spatial Competition, and Farm Survival in Norway. *American Journal of Agricultural Economics* **97**, 1192–1205 (2015). URL <https://doi.org/10.1093/ajae/aau085>.
- [126] LeSage, J. & Pace, R. *Introduction to Spatial Econometrics* (Chapman and Hall/CRC, New York, 2009).
- [127] Elhorst, J. P. *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. SpringerBriefs in Regional Science (Springer

- Berlin Heidelberg, Berlin, Heidelberg, 2014). URL <http://link.springer.com/10.1007/978-3-642-40340-8>.
- [128] Elhorst, J. P. Matlab Software for Spatial Panels. *International Regional Science Review* **37**, 389–405 (2014). URL <http://journals.sagepub.com/doi/10.1177/0160017612452429>.
- [129] Dimara, E., Pantzios, C. J., Skuras, D. & Tsekouras, K. The impacts of regulated notions of quality on farm efficiency: A DEA application. *European Journal of Operational Research* **161**, 416–431 (2005). URL <https://linkinghub.elsevier.com/retrieve/pii/S0377221703006349>.
- [130] Minviel, J.J. & Latruffe, L. Effect of public subsidies on farm technical efficiency: a meta-analysis of empirical results. *Applied Economics* **49**, 213–226 (2017). URL <https://www.tandfonline.com/doi/full/10.1080/00036846.2016.1194963>.
- [131] Kim, K., Chavas, J.-P., Barham, B. & Foltz, J. Specialization, diversification, and productivity: a panel data analysis of rice farms in Korea. *Agricultural Economics* **43**, 687–700 (2012). URL <http://doi.wiley.com/10.1111/j.1574-0862.2012.00612.x>.
- [132] Pope, R. D. & Prescott, R. Diversification in Relation to Farm Size and Other Socioeconomic Characteristics. *American Journal of Agricultural Economics* **62**, 554 (1980). URL <https://doi.org/10.2307/1240214>.
- [133] Baležentis, T. & De Witte, K. One-and multi-directional conditional efficiency measurement–efficiency in lithuanian family farms. *European Journal of Operational Research* **245**, 612–622 (2015). URL <https://doi.org/10.1016/j.ejor.2015.01.050>.
- [134] Skevas, I., Emvalomatis, G. & Brümmer, B. The effect of farm characteristics on the persistence of technical inefficiency: a case study in German dairy farming. *European Review of Agricultural Economics* **45**, 3–25 (2018). URL <http://academic.oup.com/erae/article/45/1/3/4080270>.
- [135] Zhu, X., Demeter, R. & Oude Lansink, A. Technical efficiency and productivity differentials of dairy farms in three EU countries: the role of CAP subsidies. *Agricultural Economics Review* **13**, 66–92 (2012). URL <https://doi.org/10.22004/ag.econ.253496>.

- [136] Zhu, X. & Oude Lansink, A. Impact of CAP Subsidies on Technical Efficiency of Crop Farms in Germany, the Netherlands and Sweden. *Journal of Agricultural Economics* **61**, 545–564 (2010). URL <http://doi.wiley.com/10.1111/j.1477-9552.2010.00254.x>.
- [137] Tauer, L. Age and Farmer Productivity. *Review of Agricultural Economics* **17**, 63 (1995). URL <https://doi.org/10.2307/1349655>.
- [138] Farrin, K. & Miranda, M. J. A heterogeneous agent model of credit-linked index insurance and farm technology adoption. *Journal of Development Economics* **116**, 199–211 (2015). URL <https://doi.org/10.1016/j.jdeveco.2015.05.001>.
- [139] Dercon, S. & Christiaensen, L. Consumption risk, technology adoption and poverty traps: Evidence from ethiopia. *Journal of development economics* **96**, 159–173 (2011). URL <https://doi.org/10.1016/j.jdeveco.2010.08.003>.
- [140] Sherrick, B. J., Barry, P. J., Ellinger, P. N. & Schnitkey, G. D. Factors influencing farmers' crop insurance decisions. *American Journal of Agricultural Economics* **86**, 103–114 (2004). URL <https://www.jstor.org/stable/3697877>.
- [141] Sauer, J. & Latacz-Lohmann, U. Investment, technical change and efficiency: empirical evidence from German dairy production. *European Review of Agricultural Economics* **42**, 151–175 (2015). URL <https://doi.org/10.1093/erae/jbu015>.
- [142] Porter, M. The Economic Performance of Regions. *Regional Studies* **37**, 549–578 (2003). URL <http://www.tandfonline.com/doi/full/10.1080/0034340032000108688>.
- [143] Sulewski, P. & Kloczko-Gajewska, A. Farmers' risk perception, risk aversion and strategies to cope with production risk: an empirical study from poland. *Studies in Agricultural Economics* **116**, 140–147 (2014). URL <https://doi.org/10.7896/j.1414>.
- [144] Dakpo, K. H. & Lansink, A. O. Dynamic pollution-adjusted inefficiency under the by-production of bad outputs. *European Journal of Operational Research* **276**, 202–211 (2019). URL <https://doi.org/10.1016/j.ejor.2018.12.040>.
- [145] Davis, J., Caskie, P. & Wallace, M. Promoting structural adjustment in agriculture: The economics of new entrant schemes

- for farmers. *Food Policy* **40**, 90–96 (2013). URL <https://doi.org/10.1016/j.foodpol.2013.02.006>.
- [146] Reidsma, P., Oude Lansink, A. & Ewert, F. Economic impacts of climatic variability and subsidies on European agriculture and observed adaptation strategies. *Mitigation and Adaptation Strategies for Global Change* **14**, 35 (2009). URL <http://link.springer.com/10.1007/s11027-008-9149-2>.
- [147] Skevas, T. & Serra, T. The role of pest pressure in technical and environmental inefficiency analysis of Dutch arable farms: an event-specific data envelopment approach. *Journal of Productivity Analysis* **46**, 139–153 (2016). URL <http://link.springer.com/10.1007/s11123-016-0476-0>.
- [148] Reidsma, P., Ewert, F., Oude Lansink, A. & Leemans, R. Vulnerability and adaptation of European farmers: a multi-level analysis of yield and income responses to climate variability. *Regional Environmental Change* **9**, 25 (2009). URL <http://link.springer.com/10.1007/s10113-008-0059-3>.
- [149] Wandersman, A. & Giamartino, G. A. Community and individual difference characteristics as influences on initial participation. *American Journal of Community Psychology* **8**, 217–228 (1980). URL <https://doi.org/10.1007/BF00912661>.
- [150] Curry, S. J. *et al.* Assessment of community-level influences on individuals' attitudes about cigarette smoking, alcohol use, and consumption of dietary fat. *American journal of preventive medicine* **9**, 78–84 (1993). URL [https://doi.org/10.1016/S0749-3797\(18\)30744-X](https://doi.org/10.1016/S0749-3797(18)30744-X).
- [151] Stephenson, R. Community influences on young people's sexual behavior in 3 african countries. *American journal of public health* **99**, 102–109 (2009). URL <https://doi.org/10.2105/AJPH.2007.126904>.
- [152] Foster, H. & Brooks-Gunn, J. Neighborhood, family and individual influences on school physical victimization. *Journal of youth and adolescence* **42**, 1596–1610 (2013). URL <https://doi.org/10.1007/s10964-012-9890-4>.
- [153] Rovnêa, P. The analysis of farm population with respect to young farmers in the european union. *Procedia-Social and Behavioral Sciences* **220**, 391–398 (2016). URL <https://doi.org/10.1016/j.sbspro.2016.05.513>.

- [154] Kurosaki, T. Specialization and diversification in agricultural transformation: the case of west punjab, 1903–92. *American Journal of Agricultural Economics* **85**, 372–386 (2003). URL <https://www.jstor.org/stable/1245133>.
- [155] Su, C.-C. *et al.* Pierce's Disease of Grapevines in Taiwan: Isolation, Cultivation and Pathogenicity of *Xylella fastidiosa*. *Journal of Phytopathology* **161**, 389–396 (2013). URL <https://doi.org/10.1111/jph.12075>.
- [156] Strona, G., Carstens, C. J. & Beck, P. S. A. Network analysis reveals why *Xylella fastidiosa* will persist in Europe. *Scientific Reports* **7**, 71 (2017). URL <http://www.nature.com/articles/s41598-017-00077-z>.
- [157] Bragard, C. *et al.* Update of the Scientific Opinion on the risks to plant health posed by *Xylella fastidiosa* in the EU territory. *EFSA Journal* **17** (2019). URL <http://doi.wiley.com/10.2903/j.efsa.2019.5665>.
- [158] Bucci, E. M. *Xylella fastidiosa* , a new plant pathogen that threatens global farming: Ecology, molecular biology, search for remedies. *Biochemical and Biophysical Research Communications* **502**, 173–182 (2018). URL <https://linkinghub.elsevier.com/retrieve/pii/S0006291X18311288>.
- [159] Serio, F. D. *et al.* Collection of data and information on biology and control of vectors of *Xylella fastidiosa*. *EFSA Supporting Publications* **16** (2019). URL <http://doi.wiley.com/10.2903/sp.efsa.2019.EN-1628>.
- [160] De Pascali, M. *et al.* Molecular Effects of *Xylella fastidiosa* and Drought Combined Stress in Olive Trees. *Plants* **8**, 437 (2019). URL <https://www.mdpi.com/2223-7747/8/11/437>.
- [161] Sabella, E. *et al.* *Xylella fastidiosa* induces differential expression of lignification related-genes and lignin accumulation in tolerant olive trees cv. Leccino. *Journal of Plant Physiology* **220**, 60–68 (2018). URL <https://linkinghub.elsevier.com/retrieve/pii/S0176161717302687>.
- [162] Baù, A., Delbianco, A., Stancanelli, G. & Tramontini, S. Susceptibility of *Olea europaea* L. varieties to *Xylella fastidiosa* subsp. *pauca* ST53: systematic literature search up to 24 March 2017. *EFSA Journal* **15** (2017). URL <https://doi.org/10.2903/j.efsa.2017.4772>.

- [163] Luvisi, A., Nicoli, F. & De Bellis, L. Sustainable Management of Plant Quarantine Pests: The Case of Olive Quick Decline Syndrome. *Sustainability* **9** (2017). URL <https://doi.org/10.3390/su9040659>.
- [164] Luvisi, A. *et al.* Xylella fastidiosa subsp. pauca (CoDiRO strain) infection in four olive (*Olea europaea* L.) cultivars: profile of phenolic compounds in leaves and progression of leaf scorch symptoms. *Phytopathologia Mediterranea* **56**, 259–273 (2017). URL <https://www.jstor.org/stable/44809344>.
- [165] Caserta, R., Souza-Neto, R. R., Takita, M. A., Lindow, S. E. & De Souza, A. A. Ectopic Expression of Xylella fastidiosa rpfF Conferring Production of Diffusible Signal Factor in Transgenic Tobacco and Citrus Alters Pathogen Behavior and Reduces Disease Severity. *Molecular Plant-Microbe Interactions* **30**, 866–875 (2017). URL <https://apsjournals.apsnet.org/doi/10.1094/MPMI-07-17-0167-R>.
- [166] Giampetruzzi, A. *et al.* Transcriptome profiling of two olive cultivars in response to infection by the CoDiRO strain of Xylella fastidiosa subsp. pauca. *BMC Genomics* **17**, 475 (2016). URL <http://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-016-2833-9>.
- [167] Kyrkou, I., Pusa, T., Ellegaard-Jensen, L., Sagot, M.-F. & Hansen, L. H. Pierce's Disease of Grapevines: A Review of Control Strategies and an Outline of an Epidemiological Model. *Frontiers in Microbiology* **9** (2018). URL <https://www.frontiersin.org/article/10.3389/fmicb.2018.02141/full>.
- [168] Baccari, C., Antonova, E. & Lindow, S. Biological Control of Pierce's Disease of Grape by an Endophytic Bacterium. *Phytopathology* **109**, 248–256 (2019). URL <https://apsjournals.apsnet.org/doi/10.1094/PHYTO-07-18-0245-FI>.
- [169] European Commission. Commission implementing decision 2017/2352 amending Implementing Decision 2015/789 as regards measures to prevent the introduction into and the spread within the Union of Xylella fastidiosa (Wells *et al.*). *Official Journal* **L336**, 31 (2017). URL <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017D2352&from=EN>.

- [170] Eurostat. Agriculture, forestry and fishery statistics. Tech. Rep., Eurostat (2016). URL <https://doi.org/10.2785/147560>.
- [171] White, S. M., Bullock, J. M., Hooftman, D. A. & Chapman, D. S. Modelling the spread and control of xylella fastidiosa in the early stages of invasion in apulia, italy. *Biological Invasions* **19**, 1825–1837 (2017). URL <https://doi.org/10.1007/s10530-017-1393-5>.
- [172] Holmes, T. P., Aukema, J. E., Von Holle, B., Liebhold, A. & Sills, E. Economic Impacts of Invasive Species in Forests. *Annals of the New York Academy of Sciences* **1162**, 18–38 (2009). URL <http://doi.wiley.com/10.1111/j.1749-6632.2009.04446.x>.
- [173] Saponari, M., Giampetruzzi, A., Loconsole, G., Boscia, D. & Saldarelli, P. Xylella fastidiosa in Olive in Apulia: Where We Stand. *Phytopathology* PHYTO-08-18-031 (2018). URL <https://apsjournals.apsnet.org/doi/10.1094/PHYTO-08-18-0319-FI>.
- [174] European Commission. Commission implementing decision 2015/789 as regards measures to prevent the introduction into and the spread within the Union of Xylella fastidiosa (Wells et al.). *Official Journal* **L125**, 36 (2015). URL <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32015D0789&from=EN>.
- [175] Fierro, A., Liccardo, A. & Porcelli, F. A lattice model to manage the vector and the infection of the Xylella fastidiosa on olive trees. *Scientific Reports* **9**, 8723 (2019). URL <http://www.nature.com/articles/s41598-019-44997-4>.
- [176] Hafi, A. *et al.* Economic impacts of Xylella fastidiosa on the Australian wine grape and wine-making industries (2017). URL [https://www.xfactorsproject.eu/wp-content/uploads/2017/11/EcoImpactsXylellaFastidiosa\\_20171123\\_ABARES.pdf](https://www.xfactorsproject.eu/wp-content/uploads/2017/11/EcoImpactsXylellaFastidiosa_20171123_ABARES.pdf).
- [177] Sisterson, M. S. & Stenger, D. C. Roguing with replacement in perennial crops: conditions for successful disease management. *Phytopathology* **103**, 117–128 (2013). URL <https://www.ncbi.nlm.nih.gov/pubmed/23075167>.
- [178] Rich, K. M., Miller, G. Y. & Winter-Nelson, A. A review of economic tools for the assessment of animal disease outbreaks.

- Revue Scientifique Et Technique-Office International Des Epizooties* **24**, 833–845 (2005).
- [179] Soliman, T., Mourits, M. C. M., Oude Lansink, A. & van der Werf, W. Economic justification for quarantine status - the case study of ‘*Candidatus Liberibacter solanacearum*’ in the European Union. *Plant Pathology* **62**, 1106–1113 (2013). URL <https://doi.org/10.1111/ppa.12026>.
- [180] Cook, D. C. & Fraser, R. W. Trade and invasive species risk mitigation: Reconciling WTO compliance with maximising the gains from trade. *Food Policy* **33**, 176–184 (2008). URL <https://doi.org/10.1016/j.foodpol.2007.07.001>.
- [181] Assefa, T. T., Meuwissen, M. P. & Oude Lansink, A. G. Price Volatility Transmission in Food Supply Chains: A Literature Review. *Agribusiness* **31**, 3–13 (2015). URL <http://doi.wiley.com/10.1002/agr.21380>.
- [182] Cutts, M. & Kirsten, J. Asymmetric price transmission and market concentration: an investigation into four South African agro-food industries. *The South African Journal of Economics* **74**, 323–333 (2006). URL <http://doi.wiley.com/10.1111/j.1813-6982.2006.00064.x>.
- [183] European Commission. Economic analysis of the olive sector. *Directorate-General for Agriculture and Rural Development* (2012).
- [184] Eurostat. Economic accounts for agriculture - values at current prices (2019). URL [https://ec.europa.eu/eurostat/web/products-datasets/-/aact\\_eaa01](https://ec.europa.eu/eurostat/web/products-datasets/-/aact_eaa01).
- [185] European Commission. EU olive oil farms report . *Agriculture and Rural Development* (2012). URL <https://static.oliveoiltimes.com/library/olive-farms-report.pdf>.
- [186] Agrios, G. N. *Plant Pathology* (Academic Press, 2005), fifth edit edn. URL <https://linkinghub.elsevier.com/retrieve/pii/C20090020376>.
- [187] Sisterson, M. S. & Stenger, D. C. Modelling effects of vector acquisition threshold on disease progression in a perennial crop following deployment of a partially resistant variety. *Plant Pathology* (2018). URL <http://doi.wiley.com/10.1111/ppa.12833>.
- [188] Guisan, A. & Zimmermann, N. E. Predictive habitat distribution models in ecology. *Ecological Modelling* **135**, 147–186 (2000).

- URL <https://linkinghub.elsevier.com/retrieve/pii/S0304380000003549>.
- [189] Peterson, A. T. *et al.* *Ecological niches and geographic distributions* (Princeton University Press, 2011).
- [190] EPPO Reporting Service. *X. fastidiosa* subsp. *multiplex* found in Toscana region in November 2018 (2019). URL <https://gd.eppo.int/reporting/article-6446>.
- [191] EPPO Reporting Service. *Xylella fastidiosa* subsp. *multiplex* was first found in December 2018 on lavender plants (in a zoo garden) in the municipality of Vila Nova de Gaia (near Porto) (2019). URL <https://gd.eppo.int/reporting/article-6447>.
- [192] EPPO Reporting Service. Detected for the first time in 2019 in symptomatic almond trees in 3 adjacent commercial orchards in the Hula Valley (Northeastern part). Laboratory analysis confirmed the presence of *X. fastidiosa* subsp. *fastidiosa* (2019). URL <https://gd.eppo.int/reporting/article-6551>.
- [193] Karger, D. N. *et al.* Climatologies at high resolution for the earth's land surface areas. *Scientific Data* **4**, 170122 (2017). URL <http://www.nature.com/articles/sdata2017122>.
- [194] Allouche, O., Tsoar, A. & Kadmon, R. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* **43**, 1223–1232 (2006). URL <https://doi.org/10.1111/j.1365-2664.2006.01214.x>.
- [195] Guisan, A., Thuiller, W. & Zimmermann, N. E. *Habitat Suitability and Distribution Models* (Cambridge University Press, Cambridge, 2017). URL <http://ebooks.cambridge.org/ref/id/CB09781139028271>.
- [196] Jiménez-Valverde, A. & Lobo, J. M. Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecologica* **31**, 361–369 (2007). URL <http://linkinghub.elsevier.com/retrieve/pii/S1146609X07000288>.
- [197] Robinet, C. *et al.* A suite of models to support the quantitative assessment of spread in pest risk analysis. *PLoS One* **7**, e43366 (2012). URL <https://www.ncbi.nlm.nih.gov/pubmed/23056174>.

- [198] Skellam, J. G. Random dispersal in theoretical populations. *Biometrika* **38**, 196–218 (1951). URL <https://doi.org/10.1093/biomet/38.1-2.196>.
- [199] Okubo, A. & Levin, S. A. *Diffusion and Ecological Problems: Modern Perspectives*, vol. 14 of *Interdisciplinary Applied Mathematics* (Springer New York, New York, NY, 2001). URL <http://link.springer.com/10.1007/978-1-4757-4978-6>.
- [200] Shigesada, N. & Kawasaki, K. *Biological invasions : theory and practice* (Oxford University Press, New York, 1997), 1st edn.
- [201] Hudgins, E. J., Liebhold, A. M. & Leung, B. Predicting the spread of all invasive forest pests in the United States. *Ecology Letters* **20**, 426–435 (2017). URL <http://doi.wiley.com/10.1111/ele.12741>.
- [202] Benford, D. *et al.* Guidance on Uncertainty Analysis in Scientific Assessments. *EFSA Journal* **16** (2018). URL <http://doi.wiley.com/10.2903/j.efsa.2018.5123>.
- [203] Baker, R. *et al.* Report on the methodology applied by EFSA to provide a quantitative assessment of pest-related criteria required to rank candidate priority pests as defined by Regulation (EU) 2016/2031. *EFSA Journal* **17** (2019). URL <http://doi.wiley.com/10.2903/j.efsa.2019.5731>.
- [204] European Food Safety Authority. Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment. *EFSA Journal* **12**, 3734 (2014). URL <http://doi.wiley.com/10.2903/j.efsa.2014.3734>.
- [205] Farkas, J., Baják, S. & Nagy, B. Approximating the Euclidean circle in the square grid using neighbourhood sequences. *Pure Math. Appl.* **17**, 309–322 (2006). URL [http://www.bke.hu/puma/17\\_3/FarkasBajakNagy.pdf](http://www.bke.hu/puma/17_3/FarkasBajakNagy.pdf).
- [206] Eurostat. Olive trees - Area by age and density classes (2019). URL <https://ec.europa.eu/eurostat/data/database>.
- [207] Saltelli, A. *et al.* *Global Sensitivity Analysis. The Primer* (John Wiley & Sons, Ltd, Chichester, UK, 2007). URL <http://doi.wiley.com/10.1002/9780470725184>.
- [208] Saltelli, A. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications* **145**, 280–297 (2002). URL <http://linkinghub.elsevier.com/retrieve/pii/S0010465502002801>.

- [209] Saltelli, A. *et al.* Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications* **181**, 259–270 (2010). URL <http://linkinghub.elsevier.com/retrieve/pii/S0010465509003087>.
- [210] Sobol, I. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation* **55**, 271–280 (2001). URL <http://linkinghub.elsevier.com/retrieve/pii/S0378475400002706>.
- [211] Jansen, M. J. Analysis of variance designs for model output. *Computer Physics Communications* **117**, 35–43 (1999). URL <https://linkinghub.elsevier.com/retrieve/pii/S0010465598001544>.
- [212] Alston, J. M., Fuller, K. B., Kaplan, J. D. & Tumber, K. P. Economic consequences of Pierce’s Disease and related policy in the California winegrape industry. *Journal of Agricultural and Resource Economics* **38**, 269–297 (2013). URL <https://www.jstor.org/stable/23496755>.
- [213] Tumber, K. P., Alston, J. M. & Fuller, K. B. The Costs of Pierce’s Disease in the California Winegrape Industry. *California Agriculture* (2014). URL <https://escholarship.org/uc/item/6m53d4ff>.
- [214] Diez, C. M. *et al.* Cultivar and Tree Density As Key Factors in the Long-Term Performance of Super High-Density Olive Orchards. *Frontiers in Plant Science* **7** (2016). URL <http://journal.frontiersin.org/Article/10.3389/fpls.2016.01226/abstract>.
- [215] Rallo, L. *et al.* High-Density Olive Plantations. In *Horticultural Reviews Volume 41*, 303–384 (John Wiley & Sons, Inc., Hoboken, New Jersey, 2014). URL <http://doi.wiley.com/10.1002/9781118707418.ch07>.
- [216] Pimentel, D., Zuniga, R. & Morrison, D. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics* **52**, 273–288 (2005). URL <http://linkinghub.elsevier.com/retrieve/pii/S0921800904003027>.

- [217] Charles, H. & Dukes, J. S. Impacts of invasive species on ecosystem services. In *Biological invasions*, 217–237 (Springer, 2008). URL [https://doi.org/10.1007/978-3-540-36920-2\\_13](https://doi.org/10.1007/978-3-540-36920-2_13).
- [218] Kiritsakis, A., Turkan, K. M. & Kiritsakis, K. Olive Oil. In *Bailey's Industrial Oil and Fat Products*, 1–38 (Wiley, 2020). URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/047167849X.bio029.pub2>.
- [219] Vossen, P. Olive oil: history, production, and characteristics of the world's classic oils. *HortScience* **42**, 1093–1100 (2007). URL <https://doi.org/10.21273/HORTSCI.42.5.1093>.
- [220] Stillitano, T. *et al.* Economic profitability assessment of mediterranean olive growing systems. *Bulgarian Journal of Agricultural Science* **22**, 517–526 (2016). URL <http://www.agrojournals.org/22/04-01.pdf>.
- [221] Petucco, C., Lobianco, A. & Cauria, S. Economic evaluation of an invasive forest pathogen at a large scale: the case of ash dieback in france. *Environmental Modeling & Assessment* **25**, 1–21 (2020). URL <https://doi.org/10.1007/s10666-019-09661-1>.
- [222] Schneider, K. *et al.* Impact of xylella fastidiosa subspecies pauca in european olives. *Proceedings of the National Academy of Sciences* **117**, 9250–9259 (2020). URL <https://doi.org/10.1073/pnas.1912206117>.
- [223] Surkov, I. V., Oude Lansink, A. & van der Werf, W. The optimal amount and allocation of sampling effort for plant health inspection. *European Review of Agricultural Economics* **36**, 295–320 (2009). URL <https://doi.org/10.1093/erae/jbp030>.
- [224] Nerlove, M. Estimates of the elasticities of supply of selected agricultural commodities. *Journal of farm Economics* **38**, 496–509 (1956). URL <https://doi.org/10.2307/1234389>.
- [225] International Olive Oil Council. Olive oil production data (2018). URL <https://www.internationaloliveoil.org/what-we-do/economic-affairs-promotion-unit/#figures>.
- [226] International Olive Oil Council. Olive oil consumption data (2018). URL <https://www.internationaloliveoil.org/what-we-do/economic-affairs-promotion-unit/#figures>.

- [227] Gil, J. M., Dhehibi, B., Kaabia, M. B. & Angulo, A. M. Non-stationarity and the import demand for virgin olive oil in the European Union. *Applied Economics* **36**, 1859–1869 (2004). URL <http://www.tandfonline.com/doi/abs/10.1080/0003684042000227877>.
- [228] Pierani, P. & Rizzi, P. L. An econometric analysis of the olive oil market in Italy. *European Review of Agricultural Economics* **18**, 37–60 (1991). URL <https://doi.org/10.1093/erae/18.1.37>.
- [229] Rosa, M. P. El comportamiento del consumidor y la demanda de aceites vegetales en España: oliva virgen versus oliva/girasol. *Estudios Agrosociales y Pesqueros* **192**, 161–193 (2002). URL <https://doi.org/10.22004/ag.econ.165075>.
- [230] Saltelli, A. *et al.* *Global sensitivity analysis: the primer* (John Wiley & Sons, 2008).
- [231] Iooss, B. *et al.* *sensitivity: Global Sensitivity Analysis of Model Outputs* (2020). URL <https://CRAN.R-project.org/package=sensitivity>. R package version 1.22.1.
- [232] Zarco-Tejada, P. J. *et al.* Previsual symptoms of *Xylella fastidiosa* infection revealed in spectral plant-trait alterations. *Nature Plants* (2018). URL <http://www.nature.com/articles/s41477-018-0189-7>.
- [233] White, S. M., Navas-Cortés, J. A., Bullock, J. M., Boscia, D. & Chapman, D. S. Estimating the epidemiology of emerging xylella fastidiosa outbreaks in olives. *Plant Pathology* **69**, 1403–1413 (2020). URL <https://doi.org/10.1111/ppa.13238>.
- [234] Sharp, R. L., Larson, L. R. & Green, G. T. Factors influencing public preferences for invasive alien species management. *Biological Conservation* **144**, 2097–2104 (2011). URL <https://doi.org/10.1016/j.biocon.2011.04.032>.
- [235] Estévez, R. A., Anderson, C. B., Pizarro, J. C. & Burgman, M. A. Clarifying values, risk perceptions, and attitudes to resolve or avoid social conflicts in invasive species management. *Conservation Biology* **29**, 19–30 (2015). URL <https://doi.org/10.1111/cobi.12359>.
- [236] Porteous, A. H., Rammohan, S. V. & Lee, H. L. Carrots or sticks? improving social and environmental compliance at suppliers through incentives and penalties. *Production and Operations*

- Management* **24**, 1402–1413 (2015). URL <https://doi.org/10.1111/poms.12376>.
- [237] Herzog, I. *et al.* Time to look for evidence: Results-based approach to biodiversity conservation on farmland in Europe. *Land Use Policy* **71**, 347–354 (2018). URL <https://doi.org/10.1016/j.landusepol.2017.12.011>.
- [238] Vergamini, D., Viaggi, D. & Raggi, M. Evaluating the potential contribution of multi-attribute auctions to achieve agri-environmental targets and efficient payment design. *Ecological Economics* **176**, 106756 (2020). URL <https://doi.org/10.1016/j.ecolecon.2020.106756>.
- [239] Eurostat. Household composition statistics (2021). URL [https://ec.europa.eu/eurostat/statistics-explained/index.php/Household\\_composition\\_statistics#More\\_and\\_more\\_households\\_consisting\\_of\\_adults\\_living\\_alone](https://ec.europa.eu/eurostat/statistics-explained/index.php/Household_composition_statistics#More_and_more_households_consisting_of_adults_living_alone). (Date accessed: 22.01.2021).
- [240] Copeland, T. E., Weston, J. F., Shastri, K. *et al.* *Financial theory and corporate policy*, vol. 4 (Pearson Addison Wesley Boston, 2005).
- [241] Kulhanek, S. A., Leung, B. & Ricciardi, A. Using ecological niche models to predict the abundance and impact of invasive species: application to the common carp. *Ecological Applications* **21**, 203–213 (2011). URL <https://doi.org/10.1890/09-1639.1>.
- [242] Maron, J. L. & Vilà, M. When do herbivores affect plant invasion? Evidence for the natural enemies and biotic resistance hypotheses. *Oikos* **95**, 361–373 (2001). URL <https://doi.org/10.1034/j.1600-0706.2001.950301.x>.
- [243] FAO. ISPM 5 Glossary of phytosanitary terms. Tech. Rep., Food and Agricultural Organization of the United Nations (2017). URL [http://www.fao.org/fileadmin/user\\_upload/faoterm/PDF/ISPM\\_05\\_2016\\_En\\_2017-05-25\\_PostCPM12\\_InkAm.pdf](http://www.fao.org/fileadmin/user_upload/faoterm/PDF/ISPM_05_2016_En_2017-05-25_PostCPM12_InkAm.pdf).
- [244] Elith, J. & Leathwick, J. R. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics* **40**, 677–697 (2009). URL <http://www.annualreviews.org/doi/10.1146/annurev.ecolsys.110308.120159>.

- [245] Kearney, M. & Porter, W. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology Letters* **12**, 334–350 (2009). URL <http://doi.wiley.com/10.1111/j.1461-0248.2008.01277.x>.
- [246] Bazzichetto, M. *et al.* Plant invasion risk: A quest for invasive species distribution modelling in managing protected areas. *Ecological Indicators* **95**, 311–319 (2018). URL <https://linkinghub.elsevier.com/retrieve/pii/S1470160X18305788>.
- [247] Grinnell, J. Field Tests of Theories Concerning Distributional Control. *The American Naturalist* **51**, 115–128 (1917). URL <https://www.journals.uchicago.edu/doi/10.1086/279591>.
- [248] Robinet, C. *et al.* Role of Human-Mediated Dispersal in the Spread of the Pinewood Nematode in China. *PLoS ONE* **4**, e4646 (2009). URL <https://dx.plos.org/10.1371/journal.pone.0004646>.
- [249] Tabak, M. A., Piaggio, A. J., Miller, R. S., Sweitzer, R. A. & Ernest, H. B. Anthropogenic factors predict movement of an invasive species. *Ecosphere* **8**, e01844 (2017). URL <http://doi.wiley.com/10.1002/ecs2.1844>.
- [250] Gallardo, B. & Aldridge, D. C. The 'dirty dozen': socio-economic factors amplify the invasion potential of 12 high-risk aquatic invasive species in Great Britain and Ireland. *Journal of Applied Ecology* **50**, 757–766 (2013). URL <http://doi.wiley.com/10.1111/1365-2664.12079>.
- [251] Barbet-Massin, M., Jiguet, F., Albert, C. H. & Thuiller, W. Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution* **3**, 327–338 (2012). URL <http://doi.wiley.com/10.1111/j.2041-210X.2011.00172.x>.
- [252] Vollerling, J., Halvorsen, R., Auestad, I. & Rydgren, K. Bunching up the background better bias in species distribution models. *Ecography* **42**, 1717–1727 (2019). URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/ecog.04503>.
- [253] Roberts, D. R. *et al.* Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **40**, 913–929 (2017). URL <http://doi.wiley.com/10.1111/ecog.02881>.

- [254] Merow, C. *et al.* What do we gain from simplicity versus complexity in species distribution models? *Ecography* **37**, 1267–1281 (2014). URL <http://doi.wiley.com/10.1111/ecog.00845>.
- [255] Phillips, S. J. Transferability, sample selection bias and background data in presence-only modelling: a response to Peterson *et al.* (2007). *Ecography* 080227084236895–0 (2008). URL <http://doi.wiley.com/10.1111/j.2007.0906-7590.05378.x>.
- [256] Cerqueira, V., Torgo, L. & Mozetič, I. Evaluating time series forecasting models: an empirical study on performance estimation methods. *Machine Learning* **109**, 1997–2028 (2020). URL <http://link.springer.com/10.1007/s10994-020-05910-7>.
- [257] Phillips, S. J. *et al.* Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications* **19**, 181–197 (2009). URL <http://doi.wiley.com/10.1890/07-2153.1>.
- [258] El-Gabbas, A. & Dormann, C. F. Improved species-occurrence predictions in data-poor regions: using large-scale data and bias correction with down-weighted Poisson regression and Maxent. *Ecography* **41**, 1161–1172 (2018). URL <http://doi.wiley.com/10.1111/ecog.03149>.
- [259] Renner, I. W. *et al.* Point process models for presence-only analysis. *Methods in Ecology and Evolution* **6**, 366–379 (2015). URL <https://doi.org/10.1111/2041-210X.12352>.
- [260] Greenwell, B. M., Boehmke, B. C. & McCarthy, A. J. A Simple and Effective Model-Based Variable Importance Measure. *arXiv stat.ML* (2018). URL <http://arxiv.org/abs/1805.04755>.
- [261] Zien, A., Krämer, N., Sonnenburg, S. & Rätsch, G. The Feature Importance Ranking Measure. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 694–709 (Springer, 2009). URL [http://link.springer.com/10.1007/978-3-642-04174-7\\_45](http://link.springer.com/10.1007/978-3-642-04174-7_45).
- [262] Barve, N. *et al.* The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling* **222**, 1810–1819 (2011). URL <https://linkinghub.elsevier.com/retrieve/pii/S0304380011000780>.
- [263] VanDerWal, J., Shoo, L. P., Graham, C. & Williams, S. E. Selecting pseudo-absence data for presence-only distribution modeling:

- How far should you stray from what you know? *Ecological Modelling* **220**, 589–594 (2009). URL <https://linkinghub.elsevier.com/retrieve/pii/S0304380008005486>.
- [264] Austin, M. Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling* **200**, 1–19 (2007). URL <https://linkinghub.elsevier.com/retrieve/pii/S0304380006003140>.
- [265] Gallardo, B. Europe's top 10 invasive species: relative importance of climatic, habitat and socio-economic factors. *Ethology Ecology & Evolution* **26**, 130–151 (2014). URL <http://www.tandfonline.com/doi/abs/10.1080/03949370.2014.896417>.
- [266] Lobo, J. M., Jiménez-Valverde, A. & Hortal, J. The uncertain nature of absences and their importance in species distribution modelling. *Ecography* **33**, 103–114 (2010). URL <http://doi.wiley.com/10.1111/j.1600-0587.2009.06039.x>.
- [267] El-Gabbas, A. & Dormann, C. F. Wrong, but useful: regional species distribution models may not be improved by range-wide data under biased sampling. *Ecology and Evolution* (2018). URL <http://doi.wiley.com/10.1002/ece3.3834>.
- [268] Fernández, D. & Nakamura, M. Estimation of spatial sampling effort based on presence-only data and accessibility. *Ecological Modelling* **299**, 147–155 (2015). URL <https://linkinghub.elsevier.com/retrieve/pii/S0304380014006309>.
- [269] Warton, D. I., Renner, I. W. & Ramp, D. Model-Based Control of Observer Bias for the Analysis of Presence-Only Data in Ecology. *PLoS ONE* **8**, e79168 (2013). URL <https://dx.plos.org/10.1371/journal.pone.0079168>.
- [270] Mackenzie, D. I. & Royle, J. A. Designing occupancy studies: general advice and allocating survey effort. *Journal of Applied Ecology* **42**, 1105–1114 (2005). URL <http://doi.wiley.com/10.1111/j.1365-2664.2005.01098.x>.
- [271] Diagne, C. *et al.* High and rising economic costs of biological invasions worldwide. *Nature* (2021). URL <http://www.nature.com/articles/s41586-021-03405-6>.
- [272] Kettunen, M. *et al.* Technical support to EU strategy on invasive species (IAS)-Assessment of the impacts of IAS in Europe and the EU (final module report for the European Commis-

- sion). Tech. Rep., Institute for European Environmental Policy (2008). URL [https://ec.europa.eu/environment/nature/invasivealien/docs/Kettunen2009\\_IAS\\_Task%201.pdf](https://ec.europa.eu/environment/nature/invasivealien/docs/Kettunen2009_IAS_Task%201.pdf).
- [273] Zizka, A. *et al.* CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution* **10**, 744–751 (2019). URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13152>.
- [274] Heberling, J. M., Miller, J. T., Noesgaard, D., Weingart, S. B. & Schigel, D. Data integration enables global biodiversity synthesis. *Proceedings of the National Academy of Sciences* **118**, e2018093118 (2021). URL <http://doi.org/10.1073/pnas.2018093118>.
- [275] Beck, J., Böller, M., Erhardt, A. & Schwanghart, W. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics* **19**, 10–15 (2014). URL <https://linkinghub.elsevier.com/retrieve/pii/S1574954113001155>.
- [276] McPherson, J. M., Jetz, W. & Rogers, D. J. The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology* **41**, 811–823 (2004). URL <http://doi.wiley.com/10.1111/j.0021-8901.2004.00943.x>.
- [277] Amatulli, G., McInerney, D., Sethi, T., Strobl, P. & Domish, S. Geomorpho90m - Global high-resolution geomorphometry layers: empirical evaluation and accuracy assessment. *PeerJ Preprints* (2019). URL <https://peerj.com/preprints/27595v1/>.
- [278] Hengl, T. & Wheeler, I. Soil organic carbon content in x 5 g / kg at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (Version v0.2) [Data set]. *Zenodo* (2018). URL <https://zenodo.org/record/2525553#.YHmBV0gzb-g>.
- [279] Hengl, T. Clay content in % (kg / kg) at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (Version v0.2). *Zenodo* (2018). URL <https://zenodo.org/record/2525663#.YHmBe0gzb-g>.
- [280] Hengl, T. *et al.* Global mapping of potential natural vegetation: an assessment of machine learning algorithms for estimating land potential. *PeerJ* **6**, e5457 (2018). URL <https://peerj.com/articles/5457>.

- [281] Hengl, T. Soil pH in H<sub>2</sub>O at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (2018). URL <https://zenodo.org/record/2525664#.YHmAwegzb-g>.
- [282] Hengl, T. Global Maps of Potential Natural Vegetation at 1 km resolution (2018). URL <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/QQHCIK>.
- [283] Hengl, T. & Wheeler, I. Soil organic carbon content in x 5 g / kg at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (2018). URL [https://zenodo.org/record/2525553#.YH1\\_1egzb-g](https://zenodo.org/record/2525553#.YH1_1egzb-g).
- [284] Hengl, T. Sand content in % (kg / kg) at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (2018). URL <https://zenodo.org/record/2525662#.YHmAT-gzb-g>.
- [285] Hengl, T. & Gupta, S. Soil water content (volumetric %) for 33kPa and 1500kPa suctions predicted at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (2019). URL [https://zenodo.org/record/2784001#.YH1\\_U0gzb-g](https://zenodo.org/record/2784001#.YH1_U0gzb-g).
- [286] Yamazaki, D. *et al.* A high-accuracy map of global terrain elevations. *Geophysical Research Letters* **44**, 5844–5853 (2017). URL <http://doi.wiley.com/10.1002/2017GL072874>.
- [287] World Resources Institute. Erosion (2016). URL <http://water.globalforestwatch.org/map/>.
- [288] Buchhorn, M. *et al.* Copernicus Global Land Service: Land Cover 100m: collection 3: epoch 2019: Globe (2020). URL <https://zenodo.org/record/3939050#.YHmC3-gzb-g>.
- [289] World Resource Institute. Aqueduct Global Maps 2.1 Data (2015). URL <https://www.wri.org/resources/data-sets/aqueduct-global-maps-21-data>.
- [290] Newbold, T. *et al.* Has land use pushed terrestrial biodiversity beyond the planetary boundary? A global assessment. *Science* **353**, 288–291 (2016). URL <https://doi.org/10.1126/science.aaf2201>.
- [291] European Commission, Joint Research Centre (JRC); and Columbia University, Center for International Earth Science Information Network (CIESIN). GHS Population Grid, Derived from GPW4, Multitemporal (1975, 1990, 2000, 2015). (2015). URL [https://data.europa.eu/data/datasets/jrc-ghsl-ghs\\_pop\\_gpw4\\_globe\\_r2015a?locale=en](https://data.europa.eu/data/datasets/jrc-ghsl-ghs_pop_gpw4_globe_r2015a?locale=en).

- [292] Meijer, J. R., Huijbregts, M. A. J., Schotten, K. C. G. J. & Schipper, A. M. Global patterns of current and future road infrastructure. *Environmental Research Letters* **13**, 064006 (2018). URL <https://iopscience.iop.org/article/10.1088/1748-9326/aabd42>.
- [293] Venter, O. *et al.* Global terrestrial Human Footprint maps for 1993 and 2009. *Scientific Data* **3**, 160067 (2016). URL <http://www.nature.com/articles/sdata201667>.
- [294] Kennedy, C. M., Oakleaf, J. R., Theobald, D. M., Baruch-Mordo, S. & Kiesecker, J. Managing the middle: A shift in conservation priorities based on the global human modification gradient. *Global Change Biology* **25**, 811–826 (2019). URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.14549>.
- [295] Weiss, D. J. *et al.* A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature* **553**, 333–336 (2018). URL <http://www.nature.com/articles/nature25181>.
- [296] Hengl, T. Nighttime Lights PCI-4 based on the Version 4 DMSP-OLS Nighttime Lights Time Series 1997–2014 (2018). URL <https://zenodo.org/record/1458947#.YHmGzegzb-g>.
- [297] Bhandari, L. & Roychowdhury, K. Night Lights and Economic Activity in India: A study using DMSP-OLS night time images. *Proceedings of the Asia-Pacific Advanced Network* **32**, 218 (2011). URL <http://journals.sfu.ca/apan/index.php/apan/article/view/59>.
- [298] Forbes, D. J. Multi-scale analysis of the relationship between economic statistics and DMSP-OLS night light images. *GIScience & Remote Sensing* **50**, 483–499 (2013). URL <https://www.tandfonline.com/doi/full/10.1080/15481603.2013.823732>.
- [299] Benedek, J. & Ivan, K. Remote Sensing Based Assessment of Variation of Spatial Disparities. *Geographia Technica* **13** (2018). URL [http://technicalgeography.org/index.php/latest-issue-1-2018/211-01\\_benedek](http://technicalgeography.org/index.php/latest-issue-1-2018/211-01_benedek).
- [300] Bartholdi, J. J., Jarumaneeroj, P. & Ramudhin, A. A new connectivity index for container ports. *Maritime Economics & Logistics* (2016). URL <http://www.palgrave-journals.com/doi/10.1057/mel.2016.5>.

- [301] Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* **33** (2010). URL <http://www.jstatsoft.org/v33/i01/>.
- [302] Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**, 301–320 (2005). URL <http://doi.wiley.com/10.1111/j.1467-9868.2005.00503.x>.
- [303] Anderson, R. P. & Gonzalez, I. Species-specific tuning increases robustness to sampling bias in models of species distributions: An implementation with Maxent. *Ecological Modelling* **222**, 2796–2811 (2011). URL <https://linkinghub.elsevier.com/retrieve/pii/S0304380011002195>.
- [304] Leroy, B. *et al.* Without quality presence-absence data, discrimination metrics such as TSS can be misleading measures of model performance. *Journal of Biogeography* **45**, 1994–2002 (2018). URL <http://doi.wiley.com/10.1111/jbi.13402>.
- [305] McAfee, R. P. & Mcmillan, J. Competition and Game Theory. *Journal of Marketing Research* **33**, 263–267 (1996). URL <http://journals.sagepub.com/doi/10.1177/002224379603300301>.
- [306] Waterfield, G. & Zilberman, D. Pest Management in Food Systems: An Economic Perspective. *Annual Review of Environment and Resources* **37**, 223–245 (2012). URL <http://www.annualreviews.org/doi/10.1146/annurev-environ-040911-105628>.
- [307] Hennessy, D. A. Biosecurity incentives, network effects, and entry of a rapidly spreading pest. *Ecological Economics* **68**, 230–239 (2008). URL <https://linkinghub.elsevier.com/retrieve/pii/S0921800908001134>.
- [308] Wong, R. Profit maximization and alternative theories: a dynamic reconciliation. *The American Economic Review* 689–694 (1975). URL <https://www.jstor.org/stable/1806544>.
- [309] Fousekis, P. & Stefanou, S. E. Capacity utilization under dynamic profit maximization. *Empirical Economics* **21**, 335–359 (1996). URL <http://link.springer.com/10.1007/BF01179862>.
- [310] Glass, A., Kenjegalieva, K. & Paez-Farrell, J. Productivity growth decomposition using a spatial autoregressive frontier model. *Eco-*

- nomics Letters* **119**, 291–295 (2013). URL <https://linkinghub.elsevier.com/retrieve/pii/S0165176513001110>.
- [311] Eisenhart, C. The Assumptions Underlying the Analysis of Variance. *Biometrics* **3**, 1 (1947). URL <https://www.jstor.org/stable/3001534?origin=crossref>.
- [312] Scheffe, H. Alternative Models for the Analysis of Variance. *The Annals of Mathematical Statistics* **27**, 251–271 (1956). URL <http://projecteuclid.org/euclid.aoms/1177728258>.
- [313] Fusco, E., Vidoli, F. & Sahoo, B. K. Spatial heterogeneity in composite indicator: A methodological proposal. *Omega* **77**, 1–14 (2018). URL <https://linkinghub.elsevier.com/retrieve/pii/S0305048317302025>.
- [314] Auteri, M., Guccio, C., Pammolli, F., Pignataro, G. & Vidoli, F. Spatial heterogeneity in non-parametric efficiency: An application to Italian hospitals. *Social Science & Medicine* **239**, 112544 (2019). URL <https://linkinghub.elsevier.com/retrieve/pii/S0277953619305386>.
- [315] Rahman, M. T., Nielsen, R., Khan, M. A. & Asmild, M. Efficiency and production environmental heterogeneity in aquaculture: A meta-frontier DEA approach. *Aquaculture* **509**, 140–148 (2019). URL <https://linkinghub.elsevier.com/retrieve/pii/S0044848619300894>.
- [316] Skevas, I. Measurement of production inefficiency in a technology and inefficiency heterogeneity setting. *Applied Economics* 1–11 (2020). URL <https://www.tandfonline.com/doi/full/10.1080/00036846.2020.1738327>.
- [317] Gadanakis, Y. & Areal, F.J. Accounting for rainfall and the length of growing season in technical efficiency analysis. *Operational Research* **20**, 2583–2608 (2020). URL <http://link.springer.com/10.1007/s12351-018-0429-7>.
- [318] Jiménez-Valverde, A., Rodríguez-Rey, M. & Peña-Aguilera, P. Climate data source matters in species distribution modelling: the case of the Iberian Peninsula. *Biodiversity and Conservation* **30**, 67–84 (2021). URL <http://link.springer.com/10.1007/s10531-020-02075-6>.
- [319] Acquaye, A. K. A., Alston, J. M., Lee, H. & Sumner, D. A. Economic Consequences of Invasive Species Policies in the Presence of Commodity Programs: Theory and Application

- to Citrus Canker\*. *Review of Agricultural Economics* **27**, 498–504 (2005). URL <https://doi.org/10.1111/j.1467-9353.2005.00250.x>.
- [320] Hodda, M. & Cook, D. C. Economic impact from unrestricted spread of potato cyst nematodes in Australia. *Phytopathology* **99**, 1387–1393 (2009). URL <https://www.ncbi.nlm.nih.gov/pubmed/19900005>.
- [321] Gilligan, C. A. & van den Bosch, F. Epidemiological Models for Invasion and Persistence of Pathogens. *Annual Review of Phytopathology* **46**, 385–418 (2008). URL <http://www.annualreviews.org/doi/10.1146/annurev.phyto.45.062806.094357>.
- [322] Deselnicu, O. C., Costanigro, M., Souza-Monteiro, D. M. & McFadden, D. T. A meta-analysis of geographical indication food valuation studies: What drives the premium for origin-based labels? *Journal of Agricultural and Resource Economics* 204–219 (2013). URL <https://doi.org/10.22004/ag.econ.158285>.
- [323] Aprile, M. C., Caputo, V. & Nayga Jr, R. M. Consumers' valuation of food quality labels: the case of the European geographic indication and organic farming labels. *International Journal of Consumer Studies* **36**, 158–165 (2012). URL <http://doi.wiley.com/10.1111/j.1470-6431.2011.01092.x>.
- [324] Andow, D. The extent of monoculture and its effects on insect pest populations with particular reference to wheat and cotton. *Agriculture, Ecosystems & Environment* **9**, 25–35 (1983). URL <https://linkinghub.elsevier.com/retrieve/pii/0167880983900038>.
- [325] Strona, G. *et al.* Small world in the real world: Long distance dispersal governs epidemic dynamics in agricultural landscapes. *Epidemics* **30**, 100384 (2020). URL <https://linkinghub.elsevier.com/retrieve/pii/S1755436519300805>.
- [326] Plantegenest, M., Le May, C. & Fabre, F. Landscape epidemiology of plant diseases. *Journal of The Royal Society Interface* **4**, 963–972 (2007). URL <https://royalsocietypublishing.org/doi/10.1098/rsif.2007.1114>.
- [327] Topping, C. J., Dalby, L. & Skov, F. Landscape structure and management alter the outcome of a pesticide ERA: Evaluating impacts of endocrine disruption using the ALMaSS Euro-

- pean Brown Hare model. *Science of The Total Environment* **541**, 1477–1488 (2016). URL <https://linkinghub.elsevier.com/retrieve/pii/S0048969715308597>.
- [328] Topping, C. J., Aldrich, A. & Berny, P. Overhaul environmental risk assessment for pesticides. *Science* **367**, 360–363 (2020). URL <https://doi.org/10.1126/science.aay1144>.
- [329] Boudreau, M. A. Diseases in Intercropping Systems. *Annual Review of Phytopathology* **51**, 499–519 (2013). URL <http://www.annualreviews.org/doi/10.1146/annurev-phyto-082712-102246>.
- [330] Ditzler, L., Apeldoorn, D. F., Schulte, R. P., Tittonell, P. & Rossing, W. A. Redefining the field to mobilize three-dimensional diversity and ecosystem services on the arable farm. *European Journal of Agronomy* **122**, 126197 (2021). URL <https://linkinghub.elsevier.com/retrieve/pii/S1161030120302045>.
- [331] Juventia, S. D., Rossing, W. A., Ditzler, L. & van Apeldoorn, D. F. Spatial and genetic crop diversity support ecosystem service delivery: A case of yield and biocontrol in Dutch organic cabbage production. *Field Crops Research* **261**, 108015 (2021). URL <https://linkinghub.elsevier.com/retrieve/pii/S0378429020312995>.
- [332] Davis, K. F., Rulli, M. C., Seveso, A. & D’Odorico, P. Increased food production and reduced water use through optimized crop distribution. *Nature Geoscience* **10**, 919–924 (2017). URL <http://www.nature.com/articles/s41561-017-0004-5>.
- [333] Nguyen, T. H., Nong, D. & Paustian, K. Surrogate-based multi-objective optimization of management options for agricultural landscapes using artificial neural networks. *Ecological Modelling* **400**, 1–13 (2019). URL <https://linkinghub.elsevier.com/retrieve/pii/S0304380019300870>.
- [334] Butsic, V. *et al.* Aligning biodiversity conservation and agricultural production in heterogeneous landscapes. *Ecological Applications* **30** (2020). URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/eap.2057>.
- [335] Nerlove, M. *Essays in panel data econometrics* (Cambridge University Press, 2005).
- [336] Bernhardsen, T. *Geographic information systems: an introduction* (John Wiley & Sons, Ltd, 2002).

- [337] Ricker, B. A., Rickles, P. R., Fagg, G. A. & Haklay, M. E. Tool, toolmaker, and scientist: case study experiences using GIS in interdisciplinary research. *Cartography and Geographic Information Science* **47**, 350–366 (2020). URL <https://www.tandfonline.com/doi/full/10.1080/15230406.2020.1748113>.
- [338] Fischer, M. M. & Nijkamp, P. *Handbook of regional science* (Heidelberg: Springer, 2014), 3 edn.
- [339] Emvalomatis, G., Stefanou, S. E. & Oude Lansink, A. A Reduced-Form Model for Dynamic Efficiency Measurement: Application to Dairy Farms in Germany and The Netherlands. *American Journal of Agricultural Economics* **93**, 161–174 (2011). URL <https://doi.org/10.1093/ajae/aaq125>.
- [340] Kelly, E. *et al.* Sustainability indicators for improved assessment of the effects of agricultural policy across the EU: Is FADN the answer? *Ecological Indicators* **89**, 903–911 (2018). URL <https://linkinghub.elsevier.com/retrieve/pii/S1470160X17308439>.
- [341] Bazzani, G. M., Vitali, G., Cardillo, C. & Canavari, M. Using FADN Data to Estimate CO<sub>2</sub> Abatement Costs from Italian Arable Crops. *Sustainability* **13**, 5148 (2021). URL <https://www.mdpi.com/2071-1050/13/9/5148>.
- [342] European Commission. A Farm to Fork Strategy for a fair, healthy and environmentally-friendly food system. *Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions* **381** (2020). URL [https://ec.europa.eu/info/sites/default/files/communication-annex-farm-fork-green-deal\\_en.pdf](https://ec.europa.eu/info/sites/default/files/communication-annex-farm-fork-green-deal_en.pdf).
- [343] Burke, M., Driscoll, A., Lobell, D. B. & Ermon, S. Using satellite imagery to understand and promote sustainable development. *Science* **371**, eabe8628 (2021). URL <https://doi.org/10.1126/science.abe8628>.
- [344] Gorelick, N. *et al.* Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* **202**, 18–27 (2017). URL <https://linkinghub.elsevier.com/retrieve/pii/S0034425717302900>.
- [345] Hansen, M. C. *et al.* High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **342**, 850–853 (2013). URL

- <https://doi.org/10.1126/science.1244693>.
- [346] Thirumurugan, P. & Krishnaveni, M. Flood hazard mapping using geospatial techniques and satellite images—a case study of coastal district of Tamil Nadu. *Environmental Monitoring and Assessment* **191**, 193 (2019). URL <http://link.springer.com/10.1007/s10661-019-7327-1>.
- [347] Chang, C.-H., Liu, C.-C. & Tseng, P.-Y. Emissions Inventory for Rice Straw Open Burning in Taiwan Based on Burned Area Classification and Mapping Using Formosat-2 Satellite Imagery. *Aerosol and Air Quality Research* **13**, 474–487 (2013). URL <https://aaqr.org/articles/aaqr-12-06-0a-0150>.
- [348] Li, X., Chen, W., Cheng, X. & Wang, L. A Comparison of Machine Learning Algorithms for Mapping of Complex Surface-Mined and Agricultural Landscapes Using ZiYuan-3 Stereo Satellite Imagery. *Remote Sensing* **8**, 514 (2016). URL <http://www.mdpi.com/2072-4292/8/6/514>.
- [349] Førsund, F. R. Performance measurement and joint production of intended and unintended outputs. *Journal of Productivity Analysis* (2021). URL <https://link.springer.com/10.1007/s11123-021-00599-9>.
- [350] Li, Y. *et al.* Spatial variability and distribution of N<sub>2</sub>O emissions from a tea field during the dry season in subtropical central China. *Geoderma* **193-194**, 1–12 (2013). URL <https://linkinghub.elsevier.com/retrieve/pii/S0016706112003503>.
- [351] Hamrani, A., Akbarzadeh, A. & Madramootoo, C. A. Machine learning for predicting greenhouse gas emissions from agricultural soils. *Science of The Total Environment* **741**, 140338 (2020). URL <https://linkinghub.elsevier.com/retrieve/pii/S0048969720338602>.
- [352] Shang, J. *et al.* Estimating plant area index for monitoring crop growth dynamics using Landsat-8 and Rapid-Eye images. *Journal of Applied Remote Sensing* **8**, 085196 (2014). URL <http://remotesensing.spiedigitallibrary.org/article.aspx?doi=10.1117/1.JRS.8.085196>.
- [353] Sakuma, A. & Yamano, H. Satellite Constellation Reveals Crop Growth Patterns and Improves Mapping Accuracy of Cropping Practices for Subtropical Small-Scale Fields in Japan. *Re-*

- mote Sensing* **12**, 2419 (2020). URL <https://www.mdpi.com/2072-4292/12/15/2419>.
- [354] Goerner, A., Reichstein, M. & Rambal, S. Tracking seasonal drought effects on ecosystem light use efficiency with satellite-based PRI in a Mediterranean forest. *Remote Sensing of Environment* **113**, 1101–1111 (2009). URL <https://linkinghub.elsevier.com/retrieve/pii/S0034425709000285>.
- [355] Bhattacharya, B. & Chattopadhyay, C. A multi-stage tracking for mustard rot disease combining surface meteorology and satellite remote sensing. *Computers and Electronics in Agriculture* **90**, 35–44 (2013). URL <https://linkinghub.elsevier.com/retrieve/pii/S0168169912002529>.
- [356] Lobell, D. B. *et al.* Eyes in the Sky, Boots on the Ground: Assessing Satellite- and Ground-Based Approaches to Crop Yield Measurement and Analysis. *American Journal of Agricultural Economics* **102**, 202–219 (2020). URL <https://onlinelibrary.wiley.com/doi/abs/10.1093/ajae/aaz051>.
- [357] Phiri, D. & Morgenroth, J. Developments in Landsat Land Cover Classification Methods: A Review. *Remote Sensing* **9**, 967 (2017). URL <http://www.mdpi.com/2072-4292/9/9/967>.
- [358] Jeger, M. *et al.* Updated pest categorisation of *Xylella fastidiosa*. *EFSA Journal* **16** (2018). URL <http://doi.wiley.com/10.2903/j.efsa.2018.5357>.
- [359] Damtew, E. *et al.* Communicative interventions for collective action in the management of potato late blight: evidence from a framed field game experiment in Ethiopia. *Food Security* **13**, 255–271 (2021). URL <https://link.springer.com/10.1007/s12571-020-01120-0>.
- [360] Bonney, R. *et al.* Next Steps for Citizen Science. *Science* **343**, 1436–1437 (2014). URL <https://doi.org/10.1126/science.1251554>.
- [361] Poisson, A. C. *et al.* Quantifying the contribution of citizen science to broad-scale ecological databases. *Frontiers in Ecology and the Environment* **18**, 19–26 (2020). URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/fee.2128>.
- [362] O’Grady, M. J. *et al.* Intelligent Sensing for Citizen Science. *Mobile Networks and Applications* **21**, 375–385 (2016). URL <http://link.springer.com/10.1007/s11036-016-0682-z>.

- [363] Luna, S. *et al.* Developing Mobile Applications for Environmental and Biodiversity Citizen Science: Considerations and Recommendations. In *Multimedia Tools and Applications for Environmental & Biodiversity Informatics*, 9–30 (Springer International Publishing, Cham, 2018). URL [http://link.springer.com/10.1007/978-3-319-76445-0\\_2](http://link.springer.com/10.1007/978-3-319-76445-0_2).
- [364] Hampf, A. C., Nendel, C., Strey, S. & Strey, R. Biotic Yield Losses in the Southern Amazon, Brazil: Making Use of Smartphone-Assisted Plant Disease Diagnosis Data. *Frontiers in Plant Science* **12** (2021). URL <https://www.frontiersin.org/articles/10.3389/fpls.2021.621168/full>.
- [365] Yemshanov, D. *et al.* A bioeconomic approach to assess the impact of an alien invasive insect on timber supply and harvesting: a case study with *Sirex noctilio* in eastern Canada. *Canadian Journal of Forest Research* **39**, 154–168 (2009). URL <https://doi.org/10.1139/X08-164>.
- [366] Finnoff, D., McIntosh, C., Shogren, J. F., Sims, C. & Warziniack, T. Invasive Species and Endogenous Risk. *Annual Review of Resource Economics* **2**, 77–100 (2010). URL <http://www.annualreviews.org/doi/10.1146/annurev.resource.050708.144212>.
- [367] Pacilly, F. C. A., Groot, J. C. J., Hofstede, G. J., Schaap, B. F. & van Bueren, E. T. L. Analysing potato late blight control as a social-ecological system using fuzzy cognitive mapping. *Agronomy for Sustainable Development* **36**, 35 (2016). URL <http://link.springer.com/10.1007/s13593-016-0370-1>.
- [368] Fenichel, E. P. *et al.* Adaptive human behavior in epidemiological models. *Proceedings of the National Academy of Sciences* **108**, 6306–6311 (2011). URL <http://www.pnas.org/cgi/doi/10.1073/pnas.1011250108>.
- [369] Horan, R. D., Fenichel, E. P., Drury, K. L. S. & Lodge, D. M. Managing ecological thresholds in coupled environmental-human systems. *Proceedings of the National Academy of Sciences* **108**, 7333–7338 (2011). URL <http://www.pnas.org/cgi/doi/10.1073/pnas.1005431108>.
- [370] Topping, C. J. *et al.* ALMaSS, an agent-based model for animals in temperate European landscapes. *Ecological Modelling* **167**, 65–82 (2003). URL <https://linkinghub.elsevier.com/retrieve/pii/S030438000300173X>.

- [371] Finnoff, D., Shogren, J. F., Leung, B. & Lodge, D. The importance of bioeconomic feedback in invasive species management. *Ecological Economics* **52**, 367–381 (2005). URL <https://doi.org/10.1016/j.ecolecon.2004.06.020>.
- [372] Chantarat, S., Barrett, C. B., Mude, A. G. & Turvey, C. G. Using Weather Index Insurance to Improve Drought Response for Famine Prevention. *American Journal of Agricultural Economics* **89**, 1262–1268 (2007). URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1467-8276.2007.01094.x>.
- [373] Horton, J. B. Parametric Insurance as an Alternative to Liability for Compensating Climate Harms. *Carbon & Climate Law Review* **12**, 285–296 (2018). URL <http://cclr.lexxion.eu/article/CCLR/2018/4/4>.
- [374] Bucheli, J., Dalhaus, T. & Finger, R. The optimal drought index for designing weather index insurance. *European Review of Agricultural Economics* **48**, 573–597 (2021). URL <https://academic.oup.com/erae/article/48/3/573/5904966>.

# Summary

Analyses of the economic impact of, and possible risk mitigation strategies against, pests often fail to account for spatial and economic dependencies among the evaluated decision-making units, and heterogeneity in the environment they operate in. Awareness of the mutual dependence of actors, regions and countries is critical for proper management of pests. In this thesis, I develop methodological approaches to account for the spatial nature of pest populations and the mutual dependence of farmers, countries, and markets to contribute to a more informed discussion on plant health policies in Europe.

Chapter 2 provides an empirical approach to measure spatial spillover effects of decision-making units' characteristics with managerial performance of neighbors. This allows relaxing the assumption that decision-making units operate in isolation from their peers. The model is applied to data from the Farm Accountancy Data Network. The data comprises expenses, revenues, balance sheet positions, and farm characteristics for 75 Dutch arable crop farms which are observed over six years. The results show that managerial performance of decision-making units is related to neighbors' characteristics such as the degree of farm specialization, received subsidies, insurance payments, and age. However, *how* they associate is found to depend on the definition of the neighborhood. The results imply that analyses of optimal pest control should be expanded beyond individual farm-level.

Chapter 3 provides an integrated framework that derives insights from climatic suitability, spread modelling, and economic modelling.

A pest spread model is developed such that environmental heterogeneity, point of pest introduction, and host locations are fundamental components for the realized dispersal over time. The economic model captures heterogeneity of different cropping systems and countries, includes temporal effects through losses in investments, and highlights economic dependence among growers due to price responses following changes in aggregate supply. The model is applied to the invasive species *Xylella fastidiosa* subspecies *pauca* to compute impacts to olive growers in Europe, with a focus on Italy, Greece, and Spain. For Italy, across the considered spread rates the potential economic impact over 50 years ranges from 1.9 billion to 5.2 billion Euros for the economic worst-case scenario, in which production ceases after orchards die off. If replanting with resistant varieties is feasible, the impact ranges from 0.6 billion to 1.6 billion Euros. Even under slow spread rates and the ability to replant with resistant cultivars, economic impact to olive growers from further spread of *Xylella fastidiosa* subspecies *pauca* is expected to be sizable (0.6 billion Euro) and warrants strong regulatory response.

Chapter 4 translates results of the spatially explicit pest spread model to suit the needs of partial equilibrium models. The chapter shows how global sensitivity analyses can be informative in the context of partial equilibrium models. The model is applied to the invasive species *Xylella fastidiosa* subspecies *pauca* to compute impacts on producers and consumers of olive oil in Europe, with a particular focus on Italy, Greece, and Spain. I find that most of the potential future impact of *Xylella fastidiosa* subspecies *pauca* in the olive oil market would fall on consumers because of higher prices following reductions in supply. The analysis highlights that the problem of invasive pests should be contextualized as a societal challenge as opposed to one that affects only producers. The chapter stresses the fact that consumers are beneficiaries of pest control.

Chapter 5 shows that a joint analysis of several hundred pests can produce hotspot maps with a high accuracy. A machine learning model is trained on a dataset covering 248 invasive species to map risk of new pest introduction in Europe as a function of climate, soils, water, and anthropogenic factors. Due to the considerable time and labor requirements for species-specific analyses, there is no information on area-specific suitability for establishment or introduction for

many hazardous species. The joint analysis of several species could be one approach for addressing this knowledge gap. Furthermore, joint analyses of various pests could help to identify weak-links and thereby inform collective control. Results show that the BeNeLux states, Northern Italy, the Northern Balkans, and the United Kingdom, and areas around container ports such as Antwerp, London, Rijeka, and Saint Petersburg are at higher risk for introductions. However, harmonized, systematic, species survey data comprising also true absences are required to further validate and improve these maps.

The thesis, as a collection of these articles, contributes to the literature by providing methodological approaches which (i) capture spatial dependencies, (ii) account for environmental and economic heterogeneity at the level of granularity feasible under the available data, (iii) acknowledge the temporal nature of pest spread and economic impact in perennial hosts, (iv) highlight the actionable insights sensitivity analyses can generate, and (v) propose cost-effective modelling strategies to address the absence of risk maps for many invasive species.

## Main Conclusions

- Neighbors' characteristics associate with farmers' managerial performance. However, *how* the characteristics associate with farmers' performance depends on the definition of the neighborhood (chapter 2).
- Farmers do not operate in isolation from their peers because the outcome of their efforts is partially determined by neighbors' actions (chapter 2, 3).
- The assessment of pest related losses in investments for perennial hosts can be achieved by computing losses in foregone annuities (chapter 3).
- The vast majority of the European olive cultivation is within climatically suitable territory for the establishment of *Xylella fastidiosa* subspecies *pauca* (chapter 3).
- For Italy, across the considered spread rates the potential economic impact from *Xylella fastidiosa* subspecies *pauca* to olive growers

over 50 years ranges from 1.86 to 5.17 billion Euro if the current control measures were to fail and replanting with a resistant cultivar would not be feasible (chapter 3).

- Irrespective of the disease spread rate and the ability to replant with resistant cultivars, projections of potential future economic impacts from *Xylella fastidiosa* subspecies *pauca* to olive growers in affected countries are sizable and warrant regulatory response (chapter 3).
- In Europe, introductions of *Xylella fastidiosa* subspecies *pauca* into Spain would likely result in the most drastic economic impact (chapter 3, 4).
- Capturing economic heterogeneity by stratification, does not only refine the overall estimate of pest impact but also signals whether unequal consequences from pest spread arise to the different strata (chapter 3, 4).
- Due to the inelastic supply and demand, consumers are the main beneficiaries of the control of *Xylella fastidiosa* subspecies *pauca* (chapter 4).
- Sensitivity analyses generate actionable insights on areas that should be prioritized in management, market characteristics that must be promoted, and data gaps that need to be addressed (chapter 3, 4, 5).
- Joint analyses of various species may address the overwhelming absence of information on risk maps for invasive species (chapter 5).
- Hotspot maps can identify areas that are at higher risk of invasive pest introductions (chapter 5).
- In Europe, the BeNeLux states, Northern Italy, the Northern Balkans, and the United Kingdom, and areas around container ports such as Antwerp, London, Rijeka, and Saint Petersburg are at higher risk for invasive pest introductions (chapter 5).
- In species distribution models, scholars should not exclusively rely on machine learning performance as a measure of model correctness (chapter 5).
- Systematic species survey data comprising also true absence are required to be able to disentangle reporting bias from true effects and thereby allow for the analysis of the anthropogenic involvement in invasive pest introduction (chapter 5).

## Main Policy and Research Implications

- Policy impact analyses, for example on the effects of farm subsidies, should also investigate possible community effects and not only direct impacts (chapter 2).
- Analyses of optimal pest control should be expanded beyond individual farm-level (chapter 2, 3).
- Incorporating information on host locations into pest risk assessments is crucial, especially if cultivation may be spatially clustered (chapter 3, 4).
- The inclusion of price responses to supply changes into pest risk assessments is important for deriving insights on economic dependencies and diverging incentives among stakeholders for controlling pest spread (chapter 3, 4).
- Acknowledging environmental heterogeneity in pest risk assessments is important not only for the prediction of the suitability of establishment, but also for simulations of pest spread and, consequently, economic impact (chapter 3, 4, 5).
- An arsenal of generalized models may speed up pest risk assessments, albeit at the expense of precision in results (chapter 3, 4, 5).
- Pest management strategies should take spatial heterogeneity into account by acknowledging geographic information on areas' pest risk, as well as topographic, climatic, and economic conditions (chapter 3, 4, 5).
- Spatial field-level data on host cultivation is needed to improve pest risk assessments (chapter 3, 4, 5)
- A precise spatial indexation of economic data is critical to improve modelling work (chapter 2, 3, 4, 5).



# Glossary

**Background data** describes environmental data to which pest presence data are compared to. Background data are obtained by generating geographic points and extracting environmental data for these locations.

**Balanced panel** describes data in which all cross-sectional units are observed throughout all time periods.

**Bootstrapping** is a statistical technique that falls under the class of resampling methods and uses random sampling with replacement to generate new datasets.

**Calibration** refers to the practice of adjusting values of model parameters. Data may be used to support values and result in models that more closely mimic reality.

**Cellular automaton** describes a discrete modelling technique in which pixel to pixel interactions are simulated. A cellular automaton consists of a grid of pixels each having one of a finite number of states (e.g., pest-free or infected).

**Centroid** denotes a point at the center of something. In this dissertation, centroid generally refers to a geographic point at the center of an area.

**Citizen science** refers to scientific research conducted by laymen. It is also known as community science, crowd science, crowd-sourced

science, civic science, or volunteer monitoring. In this dissertation, citizen science generally refers to the collection of data by citizens.

**Consumer surplus** describes the economic well-being of consumers and is obtained by integrating the difference between the price paid and the willingness to pay along the aggregate demand curve.

**Cordon sanitaire** denotes a barrier intending to stop the spread of a pest or an infectious disease.

**Cross-validation** describes techniques which aim to assess how well a model generalizes to unseen data. While cross-validation approaches are very diverse, in general they involve splitting the training data into several partitions. Subsequently, the model is trained on parts of the data and performance tested on withheld validation data. Cross-validation is also used to tune (i.e., optimize) the hyperparameters.

**Damage abatement inputs** refer to agricultural inputs that do not increase output, but rather protect against potential shortfall in production. In this dissertation, damage abatement inputs generally refer to pesticides, fungicides, herbicides, and other inputs that aim at reducing pest impact.

**Data Envelopment Analysis** is a non-parametric technique that is based on linear programming. The methodology is rooted in production economics and used in efficiency and productivity measurement. The name comes from the fact that the data are enveloped to derive the production frontier.

**Decision-making unit** denotes an analyzed entity which is the owner of the decision problem. Depending on the context, a decision-making unit could be a person, a company, or even a country.

**Deflation** as used in this dissertation describes the process of computing implicit quantities by dividing monetary variables, such as expenses and revenues, by corresponding price indices. Deflation is often used in methodologies that rely on having input and output quantities while only monetary variables are available to the researcher.

**Discounting** describes the process of obtaining present values of future monetary flows. Due to interest rates, inflation, and opportunity costs, economic theory suggests that money in the future is worth less

than money in the present. Discounting translates future expenses and revenues into present values.

**Elastic-net** is a regularization technique that combines the L1 (sum of absolute coefficient magnitudes) and L2 (sum of squared coefficient magnitudes) coefficient penalties into the loss function. The model is a generalized linear model based on a logit link, equivalent to a logistic model.

**Equivalent annual costs** are the annual cost of owning, operating, and maintaining an asset over its entire life.

**Feature** is machine learning terminology for a variable.

**Fold** is a term used in machine learning to describe subsets of the data. For example, five randomly split folds correspond to five data partitions each holding 20 percent of the data.

**Fundamental niche** describes the full range of environmental conditions that allow for a viable population of a species.

**Georeferenced** describes data which comprises indices that relate the measurement values to a geographic location using a coordinate reference system such as, for example, longitudes and latitudes.

**Herfindahl-Hirschman index** is a measure of concentration. The index is computed by summing the squared shares of individual components. The index is frequently used for assessing the degree of a market's concentration, but also farm specialization.

**Heterogeneity** describes that the analyzed population is different in attributes or outcomes. Heterogeneous is the opposite of homogeneous. Measuring heterogeneity or homogeneity assesses the validity of the commonly employed assumption that statistical properties are equal across observations in the dataset.

**Hyperparameter** denotes parameters that control the learning process of a machine learning algorithm but that are not directly inferred from the training (i.e., fitting) of the model as is the case for coefficients. In other words, the hyperparameters hold settings that influence the structure of the model. A standard approach is to tune these hyperparameters (i.e., optimize) by running the learning algorithm for different

values and choosing the hyperparameter value that results in the best performance according to a cross-validation procedure.

**Hypothesis** is a proposition made as a basis for reasoning without any assumption of its truth. Hypotheses can be tested using data.

**Invasive species** describe species that are non-native to an area. The term is generally associated with species that have adverse effects either economically or on the invaded ecosystem.

**Marginal effect** describe how a dependent variable changes following a marginal change in an independent variable while holding the other regressors constant.

**Model** in a scientific context describes an attempt to turn a complex system into a comprehensible analogue. By studying the system, gathering data, or qualitative information, and imposing boundaries, modelers attempt to formulate mathematical relationships which capture the essential elements on a level of granularity suitable to address a research question.

**Model ensemble** describes techniques that combine several individual machine learning algorithms into one overall prediction. The ensemble may be achieved through simple averaging, or by using approaches that take models' performance into account such as weighted averaging or meta-models.

**Multicollinearity** describes variables which can be closely predicted based on linear combinations of other variables. If multicollinearity is present, the predictive performance of machine learning algorithms is unaffected, but the estimation of standard errors is influenced.

**Non-parametric** describes techniques that do not pose any apriori assumptions regarding the underlying distribution of variables.

**Parameter** denotes various things depending on the context and scientific discipline. In this dissertation, a parameter generally describes a numerical element of a modelling system that is estimated using data.

**Pest entry** describes a pest's movement into an area.

**Pest establishment** refers to the perpetuation of the species within an area after successful entry.

**Pest introduction** describes the joint event of pest entry and establishment.

**Prevalence** describes the proportion of a population that is affected by a disease. Sampling prevalence denotes the proportion of presences to (pseudo-) absences in a dataset.

**Producer surplus** describes the economic well-being of producers and is obtained by integrating the difference between the price obtained and the marginal costs along the aggregate supply curve.

**Productive input** denotes inputs that increase the desired output.

**Pseudo-absence** describes a generated location which is treated as pest absence data. In this dissertation, pseudo-absences generally refer to geographic points which are used to generate background data to which pest presences are compared to.

**Public good** describes a good for which restricting others' consumption is not feasible (i.e., non-excludability) and for which its consumption does not affect the availability to others (i.e., non-rival).

**Realized niche** is based on the fundamental niche of a species. However, additional biotic factors such as intra-species competition and dispersal further constrain the fundamental niche into the realized niche.

**Regularization** in machine learning describes the approach of improving the generalization of models by limiting their flexibility and complexity. The implementation of this concept depends on the learning algorithm. In the elastic-net model, the regularization parameter determines the extent to which coefficient magnitudes are penalized and, consequently, shrunk toward zero.

**Sensitivity analysis** assesses the robustness of model results with respect to assumptions made by the researcher. Sensitivity analyses may evaluate results for different model structures, or for different parameter values.

**Simulation** describes very different things depending on the context and the scientific discipline. In this dissertation, the word simulation denotes the generation of synthetic data following a defined mathematical process. Simulation is often synonymously used with *numerical simulation* which describes the computation of a mathematical model using random draws out of a distribution of parameter values.

**Spatial dependence** describes a statistical dependence between measurement values which were collected across a geographic area. Spatial dependence may be measured by correlating measurements in space (i.e., spatial autocorrelation).

**Spatial lag of X model** denotes a model specification in which spatial lags for the independent variables are included. The spatial autoregressive and spatial error models refer to specifications where spatial lags for the dependent variable and error term are included as explanatory variables, respectively.

**Spatial spillovers** describe spatial effects where a decision-making unit's actions do not only influence his own output but also the output of neighboring units.

**Spatial weight matrix** is a matrix which is used to weight observations during the construction of spatially lagged variables. Such weighting is often based on geographic distances. However, there is a variety of matrix structures, such as inverse distance or k-nearest neighbors, which all come with implicit assumptions on the underlying neighborhood structure.

**Tensor** denotes a multi-dimensional mathematical object. In this dissertation, the tensor used is a multi-dimensional matrix in which elements of a two-dimensional matrix were vectors.

**Weakest-link public good** combine the properties of public goods with the additional characteristic that the realized outcome is largely influenced by the least performing decision-making unit.

# Index

## B

background data 123, 137  
biological invasion 121  
bootstrap 19  
bottom marginal variance 67

## C

calibration 99, 230  
cellular automaton 61, 94  
centroid 229  
citizen science 235  
climatic suitability 43, 48, 92  
constant returns to scale 19  
consumer surplus 97, 104  
contingent absence 133  
cordon sanitaire 41, 90, 110  
cross-validation 124, 139

## D

damage abatement inputs 14, 215  
directional distance function 17  
discounting 62, 63, 97

## E

economic dependence 4, 215, 232

elastic-net 140  
elasticity of supply 91  
ensemble prediction 41, 45, 93  
environmental absence 133  
equivalent annual costs 63  
expert knowledge elicitation 42, 61

## F

features 138  
feedback mechanism 238  
forward chaining 124  
fundamental niche 134

## G

global sensitivity analysis 53, 66, 100

## H

Herfindahl-Hirschman Index 24  
hyperparameter 123, 141  
tuning 141

## I

invasive species 1

## L

losses in investment 65

**M**

marginal effects 22, 30  
 market 4  
 methodological absence 134  
 model 7  
 multicollinearity 59

**O**

Olive Quick Decline Syndrome 39,  
 90, 110

**P**

partial equilibrium model 96  
 pest entry 122  
 pest establishment 122  
 pest introduction 47, 48, 122, 130, 220  
 pest presence 136  
 phytosanitary measures 3, 40, 41, 109  
 prevalence 59  
 producer surplus 97  
 production technology 17  
 productive inputs 17  
 pseudo-absences 124, 137  
 public good 4, 110, 231

**Q**

quasi-fixed factors 17

**R**

radial range expansion 46, 60, 94  
 realized niche 134  
 regularization 140  
 resistant cultivars 41, 56, 110, 220

**S**

satellite imagery 225

sensitivity 8, 66, 100, 106, 129  
 simulation 8, 94, 219  
 social distancing 221  
 spatial clustering 219  
 spatial effects 6, 14  
   spatial dependence 6, 14, 216  
   spatial heterogeneity 6, 14, 217, 229  
 spatial indexation 222, 225, 228  
 spatial lag of X 19  
 spatial weight matrix 20  
 species distribution model 58, 93, 122  
 spillovers 6, 14, 15, 217  
 stratification 62, 218  
 supply elasticity 97, 107

**T**

Törnqvist 24  
 tensor 64  
 top marginal variance 67  
 true absence 131  
 truncated distribution 20

**U**

uncertainty 8, 42, 47, 95, 101, 107

**V**

variable returns to scale 19  
 vectors 40

**W**

weakest-link public good 4, 231

**X**

Xylella fastidiosa 1, 39, 90

# Curriculum Vitae

I was born on April 23, 1991, in Baracaldo, Spain. After two years in diapers on the beach, I relocated and grew up in Lindental, near Stuttgart, in Germany. In 2015, I obtained my undergraduate degree in *Horticultural Sciences* from Humboldt Universität zu Berlin. In 2017, I graduated with *High Distinction* from the M.Sc. program *Life Science Economics & Policy* of the Technische Universität München.

In October 2017, I started as a Ph.D. candidate at the Business Economics Group. Until 2019, the position was embedded in the Pest Organisms Threatening Europe (POnTE) project. The POnTE project dealt with invasive species that pose risks to the European territory. The Business Economics Group was tasked to conduct economic impact assessments. The deliverable for the invasive species *Xylella fastidiosa* subspecies *pauca* laid the foundation for two of the here presented chapters.

Next to invasive species, I am interested in general machine learning applications, efficiency and productivity measurement, and satellite imagery. I am passionate about integrating environmental and economic data and generating insights by bridging scientific disciplines. In future work, I would like to look at questions that deal with safeguarding biodiversity, optimizing water use, preparing food systems for extreme weather events, and assisting disaster recovery. I like scuba diving, coffee, and cats.

**Kevin Schneider**  
**Wageningen School of Social Sciences (WASS)**  
**Completed Training and Supervision Plan**



Wageningen School  
of Social Sciences

Name of the learning activity	Department/institute	Year	ECTS*
<b>A) Project related competences</b>			
Modelling and Simulation of Complex Socio-Technical Systems, INF 51806	WUR	2018	6
Statistical Uncertainty Analysis of Dynamic Models	PE&RC, WUR	2017	1.5
Theory and Practice of Efficiency & Productivity Measurement	BEC, WUR & U Florida	2018	4.5
Economic Modelling of the Bioeconomy	WASS, WUR	2019	3
Dynamic Models in R	PE&RC, WUR	2019	1.8
Geostatistics	PE&RC, WUR	2019	1.5
Machine Learning	Stanford Coursera	2020	2
<b>B) General research related competences</b>			
WASS Introduction	WASS, WUR	2017	1
Writing Project Proposal	BEC, WUR	2018	6
Data Scientist with Python	Datacamp	2018	3
<i>'Measuring the effects of spatial spillovers on farm-level inefficiency'</i>	European Workshop on Efficiency and Productivity Analysis (EWEPA), London	2019	1
<i>'Potential impact of Xylella fastidiosa subsp. pauca in European olives: a bioeconomic analysis'</i>	Second European Conference on Xylella fastidiosa, European Food Safety Authority, Corsica	2019	1
<b>C) Career related competences/personal development</b>			
Assisted teaching of Decision Science 2	BEC/WUR	2019	2
<b>Total</b>			<b>34.3</b>

\*One credit according to ECTS is on average equivalent to 28 hours of study load

# Book Cover

The basis of the book cover was a black and white stock image<sup>7</sup>. I liked the image because it vaguely resembles an olive tree while abstractly visualizing dependencies through molecule-style connections. In QGIS, I scaled and projected the image onto gridded data for land-use classification across Europe from the Copernicus database<sup>8</sup>. I saved the manually reprojected image and performed the following steps in R. Feel free to reach out to me in case you want the corresponding files or the full script.

---

## Code Segment 6.1 Load, read, and prepare.

---

```
1 #- We load some packages and set the working directory
2 library(raster); library(tmap); library(sf); setwd("YOUR_PATH")
3
4 #- We load the manually prepared files for the tree shape and data
5 shape = raster("data/tree.tif")
6 classes = raster("data/landclass.tif")
7
8 #- Removing the white background is straightforward
9 shape[shape>=200] = NA
```

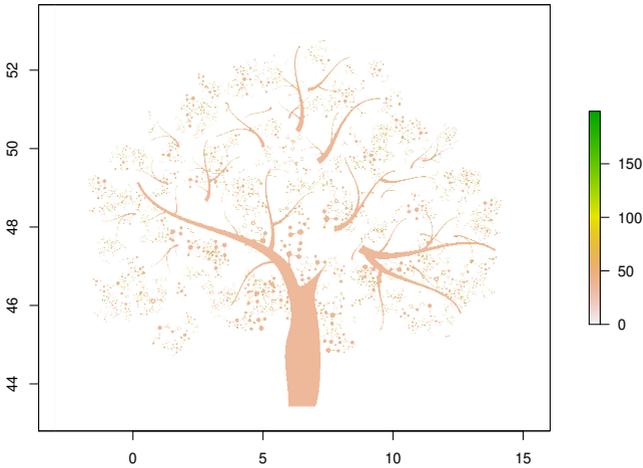
---

We end up with a raster image of our desired tree-shape in which only the black pixels have values. We can visualize this by using *plot(shape)*. The raster looks as follows (see Figure 6.1).

---

<sup>7</sup> <https://www.shutterstock.com/de/image-vector/molecule-tree-87089435>

<sup>8</sup> <https://land.copernicus.eu/global/products/lc>



**Fig. 6.1** Tree shape as a raster after removing the background.

Next, we need to change the tree's pixels to geographic points and extract the land-use data for each point. Because we manually overlaid the shape in QGIS, we know that the coordinate reference system and the general area align with the data. However, note that the manual reprojection is not very scientific! We simply scaled and moved the black and white tree-shape to align nicely over Europe. Do not take any of this too seriously!

**Code Segment 6.2** Turn pixels into geographic points.

---

```

1 #- Convert the drawing to geographic points
2 pts = rasterToPoints(shape, spatial=T)
3 crs(pts) = crs(classes)
4
5 #- If we want we can plot the result
6 eu = shapefile("data/Europe.shp")
7 tm_shape(eu)+tm_borders()+tm_shape(pts)+tm_dots(size=0.001)

```

---

Figure 6.2 shows the generated geographic points. The tree shape is made up of around 2.9 million individual points. Given that we have geographic points now, we can easily extract the land-use data.

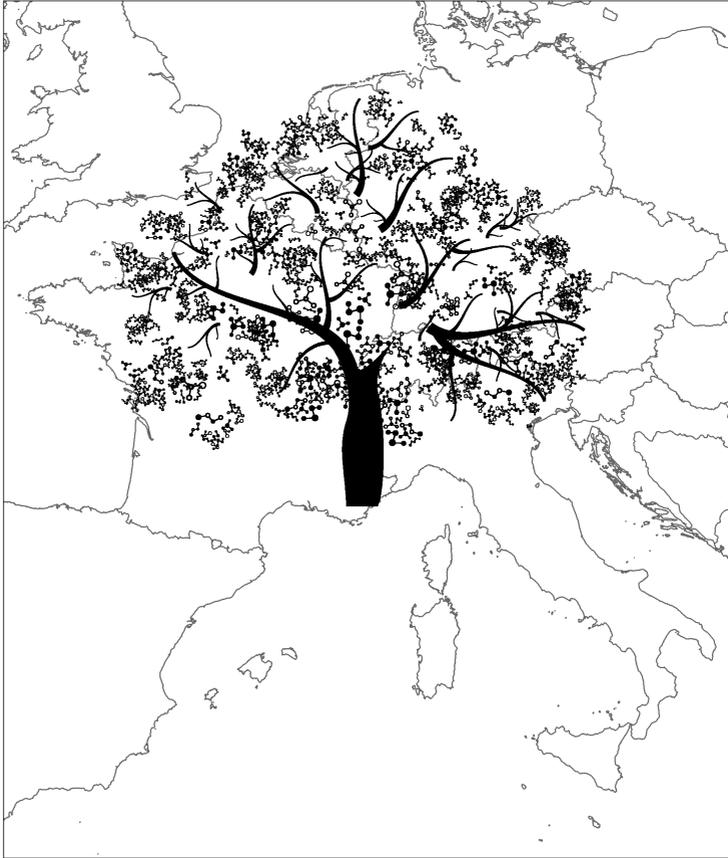


Fig. 6.2 Generated geographic points across Europe.

Because we had moved the tree's top branches into the North sea when manually reprojecting the image, all points get the same land-use classification and therefore color. This results in a witty reference to *Xylella fastidiosa* subspecies *pauca* symptoms. The disease causes branches to desiccate. Of course, you can take any shape you like and follow the same steps. You can also move the shape to wherever you would like it to land! Endless opportunities.

**Code Segment 6.3** Extract land-use data.

---

```

1  #- Extract the land classifications for each "location"
2  pts = data.frame(cbind(pts@coords, landcover=extract(classes, pts)))
3
4  #- Convert data.frame back to a spatial object
5  pts$landcover = as.factor(pts$landcover)
6  coordinates(pts) = ~x+y
7  crs(pts) = crs(classes)
8
9  #- Change the variable's categories from numbers to nicer names
10 levels(pts$landcover) = read.csv("data/levels.csv", sep=";",
11                                header=T)$name

```

---

We now have categorical data for the land-use in each "location". This means we can color the categories to our liking and in doing so create a colorful tree in which the color-pattern is determined by the land-use across European! To do so, we pick a few nice color-palettes and randomly shuffle them. I did this many times and picked a nice result for us in the form of hex-color-codes. Note that the next code segment takes quite some time. Because it took so long, below we simply work with the final color-hex-code!

**Code Segment 6.4** Generate random color-scheme.

---

```

1  #- Get the number of categories
2  cats = length(levels(pts$landcover))
3
4  #- Generate a list of four color-palettes we test
5  colours = list(alphabet = pals::alphabet(cats),
6                glasbey  = pals::glasbey(cats),
7                kelly    = pals::kelly(cats),
8                polychrome = pals::polychrome(cats))
9
10 #- Visualize 50 shuffles for each palette and save hex-codes
11 for(n in c("alphabet", "glasbey", "kelly", "polychrome")){
12   for(i in 1:50){
13     folder = paste("results/colorruns/", n, "/", sep="")
14     cols = sample(colours[[n]])
15     names(cols) = levels(pts$landcover)
16
17     p = tm_shape(pts) + tm_dots(col="landcover", palette=cols, size
18                               =0.001) + tm_layout(legend.show=F, frame=F, bg.color="
19                               transparent")
20     tmap_save(p, paste(folder, i, "_cover.png", sep=""), dpi=3000)
21     write.table(cols, paste(folder, i, "_hex.txt", sep=""))
22   }
23 }

```

---

Lastly, we simply need to generate the final visualization and save it. Of course, we want to know what land-use category each color refers to in the final image. Therefore, we save the legend separately

and print it here for endless exploration of our generated color-pattern (see Figure 6.3b)! To give you a better orientation for starting your journey, the half-moon shape at the top of the tree's trunk is *Lake Geneva* in Switzerland.

---

**Code Segment 6.5** Visualize the colored book cover.

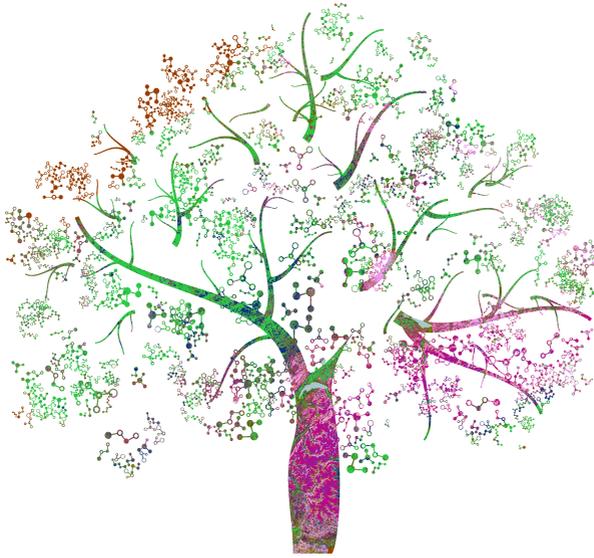
---

```

1  #- Select one of the random color schemes
2  colours = read.table("results/colorruns/alphabet/2_hex.txt",
3                      header=T)$x
4
5  #- Change the names and overwrite one color
6  names(colours) = levels(pts$landcover)
7  colours[[7]]   = "#A6C8CC"
8
9  #- Now we just need to generate the plot
10 p = tm_shape(pts) + tm_dots(col="landcover", palette=colours, size
    =0.001) + tm_layout(legend.show=F, frame=F, bg.color="
    transparent")
11
12 #- .EPS is great for working in Photoshop or Illustrator
13 tmap_save(p, "results/cover.eps", dpi=3000)
14
15 #- Save the color-legend
16 pdf("legend.pdf")
17 plot(NULL, xaxt='n', yaxt='n', bty='n', ylab='', xlab='', xlim=0:1,
    ylim=0:1)
18 legend("topleft", pch=16, pt.cex=3, cex=1.3, bty='n',
    legend=names(colours), col = colours)
19 mtext("Land Cover Class", at=0.2, cex=2)
20 dev.off()
21
22
23 #- Never forget to celebrate your victory
24 print("YAY! YIPPEY! YIIIIHA!")

```

---



(a)

### Land Cover Class

- Shrubs
- Herbaceous vegetation
- Cultivated and managed cropland
- Urban
- Bare
- Snow
- Permanent water bodies
- Herbaceous wetland
- Moss and lichen
- Closed forest, evergreen needle leaf
- Closed forest, deciduous broad leaf
- Closed forest, mixed
- Closed forest, unknown
- Open forest, evergreen needle leaf
- Open forest, deciduous broad leaf
- Open forest, mixed
- Open forest. Unknown
- Open sea

(b)

**Fig. 6.3** Final cover logo and the corresponding legend.

The research described in two chapters of this dissertation was supported by funding from the European Commission's Horizon 2020 Pest Organisms Threatening Europe (POnTE) research and innovation program under grant agreement number 635646.