Computers and Electronics in Agriculture 236 (2025) 110426

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Combining spatio-temporal pest risk prediction and decision theory to improve pest management in smallholder agriculture

Johannes Kopton [a,*], Sytze de Bruin [b], Dario Schulz [c,d], Eike Luedeling [a]

[a] Institute for Crop Science and Resource Conservation (INRES), University of Bonn, Auf dem Hügel 6, 53121 Bonn, Germany
[b] Laboratory of Geo-Information Science and Remote Sensing, Wageningen University, Droevendaalsesteeg 3, 6708 PB Wageningen, The Netherlands
[c] Institute for Food and Resource Economics (ILR), University of Bonn, Nußallee 21, 53115 Bonn, Germany
[d] European Forest Institute (EFI) Bonn, Platz der Vereinten Nationen 7, 53113 Bonn, Germany

ARTICLE INFO

Dataset link: https://github.com/johanneskopton/pest-risk-decision

Keywords:
Pest risk prediction
Decision analysis
Value of Information
Deep Gaussian process
Spatio-temporal modeling
Species distribution modeling
Smallholder agriculture

ABSTRACT

Pests and diseases are a major cause of crop loss, affecting food security and, in particular, the livelihoods of smallholder farmers. While some pest management practices are widely adopted, poorly informed decisions, such as over-application of pesticides, can severely impact human and environmental health as well as farm profits. Frequent crop monitoring is often recommended for making interventions more effective, but highly intensive monitoring is beyond the capacity of many farmers. By combining pest risk prediction and decision analysis, we developed a framework to support the decision of whether to apply pesticides preventively, monitor crops, or omit any crop protection measures for a given day and geographic location. We used a deep Gaussian process classification model for spatiotemporal pest risk prediction, incorporating new observations in near real-time. We then applied decision analysis to determine the best intervention for a given pest risk prediction. Monitoring is recommended when the Value of Information (VoI) exceeds the cost of monitoring. We applied this method to a case study of *Tuta absoluta* infestations in tomato production in Andhra Pradesh, India. Our model-based decision strategy would reduce average pest-related costs by $25.4 \pm 4.3\%$ and pesticide use by $58.8 \pm 2.7\%$ according to Monte Carlo simulations. When monitoring results are used to update the pest risk model and thus shared with other farmers, additional value can be generated for the community. We found that this *community VoI* exceeded the expected information value for the individual farmer (*individual VoI*). Our open-source Python model can easily be adapted to other crops and pathogens, and serve as a basis for pest risk-aware decision support systems.

1. Introduction

Crop pests are a major threat to food security and rural livelihoods. Most farmers around the world regularly use some form of crop protection (Tang et al., 2021), but misapplication can be detrimental to profits, a particularly sensitive issue for the livelihoods of smallholder farmers. In addition, excessive pesticide use can affect human health (Boedeker et al., 2020), environmental sustainability (Tudi et al., 2021) and beneficial arthropods (Desneux et al., 2007), and it can lead to pests becoming resistant to the pesticides' active ingredients (Whalon et al., 2008). To reduce both yield losses and excessive pesticide use, frequent monitoring for pathogens or their symptoms is considered crucial for integrated pest management (Desneux et al., 2022). While daily monitoring of all fields would provide almost perfect information about pest infestations, such intensive monitoring is usually beyond the labor resources of farmers. Pest risk prediction could therefore help farmers not only decide when and where to apply plant protection measures (Rossi et al., 2023), but also whether and when to monitor. Since monitoring results can be fed back into prediction models used to warn nearby farmers about disease outbreaks, the question can be asked in two ways: When and where is monitoring worthwhile for the individual farmer? When and where is monitoring worthwhile for the community?

As a case study, we selected *Tuta absoluta* infestations in tomato production in Andhra Pradesh, India, between 2018 and 2022. *T. absoluta* (known as tomato leaf miner or South American tomato pinworm) is a devastating invasive pest of Indian tomato, eggplant and potato production, which originated in western South America (Biondi et al., 2018). The common practice among farmers in Andhra Pradesh is calendar-based spraying of chemical pesticides (Buragohain et al., 2021) but this one-size-fits-all approach often results in unnecessary spraying at times and in regions with low actual infestation risk. Therefore, improved

* Corresponding author.
E-mail address: johannes@kopton.org (J. Kopton).

https://doi.org/10.1016/j.compag.2025.110426
Received 12 August 2024; Received in revised form 27 November 2024; Accepted 16 April 2025
Available online 6 May 2025

Nomenclature

T. absoluta	*Tuta absoluta*
VoI	Value of Information
EVPI	Expected Value of Perfect Information
GPR	Gaussian Process Regression
ETL	Economic Threshold Level

targeting of pest control interventions can be expected to deliver both financial and environmental benefits.

The goal of this work is to develop a framework for informing crop protection and monitoring decisions based on predicted pest risks. In the case study, we quantify the potential economic and environmental benefits of our pest risk prediction model. In addition, we quantify the value of monitoring to both the individual farmer and their community.

This study builds upon foundational research on decision making under uncertainty. A central concept for both our decision model and the valuation of monitoring benefits is the Value of Information (VoI), which was introduced by Howard (1966). In decision analysis, the VoI is the price that a rational decision-maker would be willing to pay to reduce uncertainty in order to support a decision. A common formulation of this is the Expected Value of Perfect Information (EVPI), which is defined as the expected utility at the current state of uncertainty minus the expected utility given perfect information (Howard, 1966). This concept has been used in a variety of scientific and business applications. Notable in the context of this work is the study of de Bruin and Hunter (2003) on VoI to evaluate spatial datasets for decision-making in agriculture. de Bruin and Bregt (2012) used expected VoI to analyze the utility of taking measurements at a given point in time and space. This is conceptually very similar to the value of disease monitoring and thus closely aligned with the subject of this paper. However, this concept has only been applied to continuous synthetic data without complex relationships with external covariates. Outside the field of spatially explicit modeling, decision analytic approaches have been applied in a number of similar contexts: Ruett et al. (2022) evaluated different monitoring strategies for disease control in ornamental heather production. Luu et al. (2022) used decision analysis to investigate the potential future benefits of providing agro-climatic information to smallholder farmers in Vietnam. While these studies take a very broad and holistic view, both consider only a one-time decision in an aggregated region (asking only *whether* rather than *when* and *where*). In addition to these prospective considerations, empirical evidence suggests that personalized digital extension services (e.g., providing local information on plant pests) significantly benefit smallholder farmers in India (Rajkhowa and Qaim, 2021).

This work also draws upon the literature related to the modeling of plant pests and diseases. In general, pest infestations occur when pests meet susceptible host plants under the right environmental conditions (Agrios, 2008), a relationship commonly known as the “disease triangle”. While susceptible host plants are expected to be present whenever personalized advice to farmers is provided, past pest occurrences (in nearby locations) and/or environmental conditions are usually used as inputs for prediction models. In their literature review, Fenu and Malloci (2021) provide an overview of the methods used in 46 plant pest and disease forecasting studies. Of these, 29 used weather data as predictors, since weather is arguably the most relevant aspect of environmental conditions. None of the above studies used nearby observations directly as input to take advantage of autocorrelation. Several machine learning methods have been commonly used, including artificial neural networks, support vector machines, Bayesian networks, and random forests. In tomato production, such methods have mostly been used to predict infection with powdery mildew (Ghaffari et al., 2010; Bhatia et al., 2020a,b).

In contrast to models based on machine learning, our approach aims at exploiting spatio-temporal neighborhood effects in addition to weather data in order to be able to warn farmers about local outbreaks that cannot be fully explained by environmental suitability. Geostatistical methods allow the analysis of spatial autocorrelation as well as spatial interpolation (Matheron, 1963). Some studies have begun to explore the application of such methods to pest modeling. Wright et al. (2002) analyzed the spatial distribution of the European corn borer (*Ostrinia nubilalis*), but did not use the results for interpolation. de Carvalho Alves et al. (2011) used variogram-based ordinary kriging to interpolate infestation intensity over a small region. Indicator kriging was successfully used by de Carvalho Alves and Pozza (2010) to interpolate the probability of occurrence (which fits well with the goal of our prediction model). However, both studies only interpolated spatially between sampling locations, without extrapolating temporally into the future (forecasting). Fatemi et al. (2023) used geostatistical methods to analyze the spatial distribution of *T. absoluta*, but only under greenhouse conditions and again without forecasting. GSTAR Kriging was used by Pramoedyo et al. (2020) to combine temporal extrapolation with spatial interpolation, but only 9 sites were observed and no external covariates were considered.

Since plant pest distribution models can be considered as a special case of species distribution models (SDM), this area of the literature is also relevant to the present work. In their review of the literature on spatiotemporal SDMs, Martínez-Minaya et al. (2018) found that most studies relied on presence-only observations and weather/climate covariates. Special attention was paid to imperfections in the data, such as biased sampling (Diggle et al., 2010), spatial misalignment of observations and covariates (Foster et al., 2012), or non-stationarity. Inference has typically been done using frequentist or Bayesian statistical models, which make various assumptions about the relationship between spatial and temporal variation. However, because the time scales were decades rather than days, and temporal dynamics were generally considered to be of secondary importance, the applicability of the methodology to the present study is limited.

Gaussian Process Regression (GPR) can be seen as a generalization of kriging (Rasmussen and Williams, 2005). The kernel function (known in geostatistics as *variogram model*) is usually not fitted to an empirical semivariogram, but rather directly to the data using maximum likelihood estimation or Bayesian inference. GPR can directly incorporate regression via nonstationary (e.g., linear) kernels, similar to regression kriging. Using GPR to estimate probabilities with binary input data is known as Gaussian process classification and can be seen as a generalization of indicator kriging. Many challenges regarding the scalability of Gaussian process-based approaches have been addressed in recent years (Liu et al., 2020), enabling their application to large spatio-temporal datasets (Low et al., 2015). Notable among these innovations are sparse Gaussian processes (Titsias, 2009), stochastic variational Gaussian processes (Hensman et al., 2013), and GPU acceleration (Gardner et al., 2021). Furthermore, GPR can be combined with deep learning by using stationary kernels on neural network derived features (Wilson et al., 2015), which greatly increases the explanatory power of this approach. Since all steps are differentiable, all parameters and hyperparameters can be efficiently learned end-to-end. You et al. (2017) used a combination of deep learning and GPR for crop yield prediction. Goldstein et al. (2019) used Gaussian processes to analyze the spread of invasive insects. Overall, this approach does not seem to be widely used in agricultural modeling, especially not for pest prediction.

Here, we contribute to the field of pest risk prediction in three ways: First, we present a probabilistic pest risk prediction method that combines environmental covariates and spatiotemporal neighborhood effects based on deep Gaussian process classification. Second, we assess the potential economic value of disease prediction for smallholder farmers using such a decision model. Third, we investigate the value of monitoring for both the individual farmer and the surrounding farms by introducing the concept of community VoI.

2. Materials and methods

Our approach consists of three coupled models: A pest risk prediction model, a prescriptive decision model that takes the estimated pest risk as an input, and an agent-based model to explore the effects of farmers following the prescribed decision.

2.1. Data

As a case study, we focus on the important tomato pest *T. absoluta* in Andhra Pradesh over a period from 2018 to 2022 (5 years). As our primary data source for the distribution of *T. absoluta* infestations on tomato farms in Andhra Pradesh, we used crowdsourced observation data from the Plantix smartphone app provided by PEAT GmbH (Berlin, Germany). The app allows farmers to classify plant pathogens from a photo of symptoms or visible pathogens through ML-based image classification. To do that, smallholder farmers usually take a picture of the plant using their smartphone directly in the field and upload it using cellular data. The cloud-based Plantix backend then predicts both the depicted crop and pathogen (if any) using an ensemble of image classification models including convolutional neural networks and vision transformers. Each time a farmer uses this service, the classification results including confidence scores are saved into a database, together with various metadata. Plantix data have been successfully used in several academic publications (Wang et al., 2020; Lee et al., 2022). Our dataset consists of 224,136 data points (high-quality images of tomato plants in Andhra Pradesh) with geolocation, timestamp, and classification results. To ensure that only data points of sufficient quality were used, only classification results that were also returned to the user by the Plantix app were included. This internal quality check in Plantix incorporates, among other metadata, the confidence scores of the image classification models for crops and pathogens. Results from third-party licensees were excluded. After further filtering for only the highest confidence image per user per day, and using only locations where weather data was available daily over the entire time period, a dataset of 68,586 records remained.

Weather data were retrieved from the European Centre for Medium-Range Weather Forecasts (ECMWF) "ERA5-Land Daily Aggregated" dataset (Copernicus Climate Change Service, 2024) via Google Earth Engine with a spatial resolution of 11,132 m. A 0.3 degree buffer was used around Andhra Pradesh so that points near the state border could also benefit from neighboring observations.

2.2. Pest risk prediction model

The goal of this model was to predict the pest risk for a single pest. The pest risk is defined as the probability of reaching the Economic Threshold Level (ETL), i.e., infestation at a level where treatment is profitable (Stern, 1973). In this form, the risk could serve directly as an input to the decision model in the next step. However, we had no real ground truth data for this risk. The Plantix crowdsourced data was inherently presence-only. We assumed that infestations detected in a smartphone image corresponded to infestations exceeding the ETL (see Section 2.4 for details on the case study).

Since Plantix data were sampled opportunistically, i.e., sampling was not done completely at random or systematically, it depended on the location of Plantix users and the seasonal distribution of tomato fields. We therefore followed the literature on SDMs and sampled from other Plantix detections in tomato (other pests, diseases, or healthy plants) as pseudo-absence points. This method is known as "target-group sampling" (Phillips et al., 2009; Zbinden et al., 2024) and was described by Martínez-Minaya et al. (2018) as interpreting the data as a marked point pattern. However, even within the tomato fields of Plantix users, sampling was unbalanced. The app is typically consulted when symptoms are already present, leading to a preferential sampling of unhealthy plants. In the raw data, 1.67% of observations were *T. absoluta* (=1146). We chose to sample 5 times as many absence points as there were presence points to mitigate zero-inflation, while at the same time keep enough observations to thoroughly cover the parameter space. To scale these results to the actual overall relative frequencies, we used external literature data for average infestation risk (over the entire study area and time period), assuming additional unobserved *healthy* crops were proportional to total observations. This can be thought of as predicting the spatio-temporal distribution conditional on a known marginal distribution. We represented the presence data as a boolean variable with 1 for presence and 0 for absence.

To predict pest risk at a given point in time and space, we used two types of correlations: correlation with abiotic environmental conditions and spatio-temporal autocorrelation. Significant correlation with weather data was found using logistic regression and random forest classification (detailed results of these preliminary analyses are shown in Figs. A.1 and A.2). For each Plantix observation, 7 weather variables and 2 leaf area index variables were used in time series, each containing the previous 30 days. The spatio-temporal autocorrelation in the Plantix observation data can be demonstrated with covariograms (as shown in Fig. 1). The autocovariance could only be partially explained by autocorrelated environmental variables. The highest covariance (4.8% on the residuals) was found for spatial lags less than 2 km and temporal lags less than 20 days. With increasing spatial and temporal lags, the covariance decreased until it was not significantly different from zero at 100 days and 10 km. This information can be used in spatio-temporal variogram/covariogram kriging, as shown by Snepvangers et al. (2003), but their method exhibits limitations in incorporating complex correlations with external covariates.

We decided to use Gaussian process classification with stochastic variational deep kernel learning (Wilson et al., 2016) as a modeling approach. This method has several advantages over both conventional machine learning approaches and pure geostatistical methods: Unlike parametric machine learning approaches such as neural networks, GPR models can utilize new data without retraining the entire model (i.e., updating the posterior distribution without re-fitting the kernel function). Like their special case *kriging*, they are able to exploit (spatio-temporal) autocorrelation, while being less prone to overfitting than parametric classifiers. Deep kernel learning allows the extraction of meaningful features from the high-dimensional covariate space (Wilson et al., 2015). Stochastic variational inference helps overcome the performance limitations previously associated with GPR and kriging.

For our pest risk prediction model, we designed the architecture as follows (see Fig. 2): A 3-dimensional feature vector x_{weather} was extracted from the weather data using a very small multilayer perceptron (artificial neural network) with one hidden layer of 5 neurons and a ReLU activation function (Fukushima, 1969). Together with the spatial coordinates and timestamps, these features were then used in a Gaussian process layer. For the spatio-temporal autocorrelation, we used Gaussian kernels for space (k_{space}), time (k_{time}) and a metric combination of both (k_{metric}), analogous to the sum-metric variogram model presented by Graeler et al. (2016). In addition, a periodic kernel, multiplied with another Gaussian kernel was used to capture seasonal correlation effects (k_{season}) as shown by HajiGhassemi and Deisenroth (2014). Weather features were jointly included using another Gaussian kernel (k_{weather}) with Automatic Relevance Determination (ARD). As final similarity metric, the sum of all kernels was used.

$$k_{\text{space}}(x,x') = \sigma^2_{\text{space}} \exp\left(-\frac{(x_{\text{lat}}-x'_{\text{lat}})^2+(x_{\text{lon}}-x'_{\text{lon}})^2}{2\ell_{\text{space}}}\right) \tag{1}$$

$$k_{\text{time}}(x,x') = \sigma^2_{\text{time}} \exp\left(-\frac{(x_{\text{time}}-x'_{\text{time}})^2}{2\ell_{\text{time}}}\right) \tag{2}$$

$$k_{\text{metric}}(x,x') = \sigma^2_{\text{metric}} \cdot \exp\left(-\frac{(x_{\text{lat}}-x'_{\text{lat}})^2+(x_{\text{lon}}-x'_{\text{lon}})^2+\alpha(x_{\text{time}}-x'_{\text{time}})^2}{2\ell_{\text{time}}}\right) \tag{3}$$

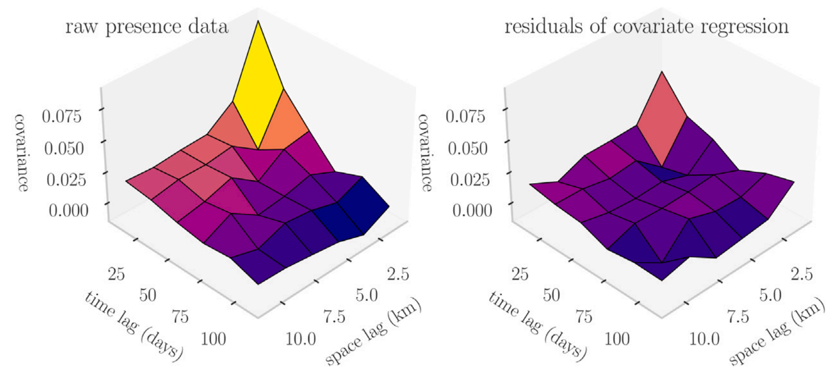

Fig. 1. Spatio-temporal covariogram showing the autocorrelation in the input data. The left subplot is based on raw observation data (0: absence, 1: presence). The right subplot is based on the residuals of a neural network regression, to control for correlation with autocorrelated covariates. Both were computed using 2×10^9 randomly sampled pairs of data points respectively.

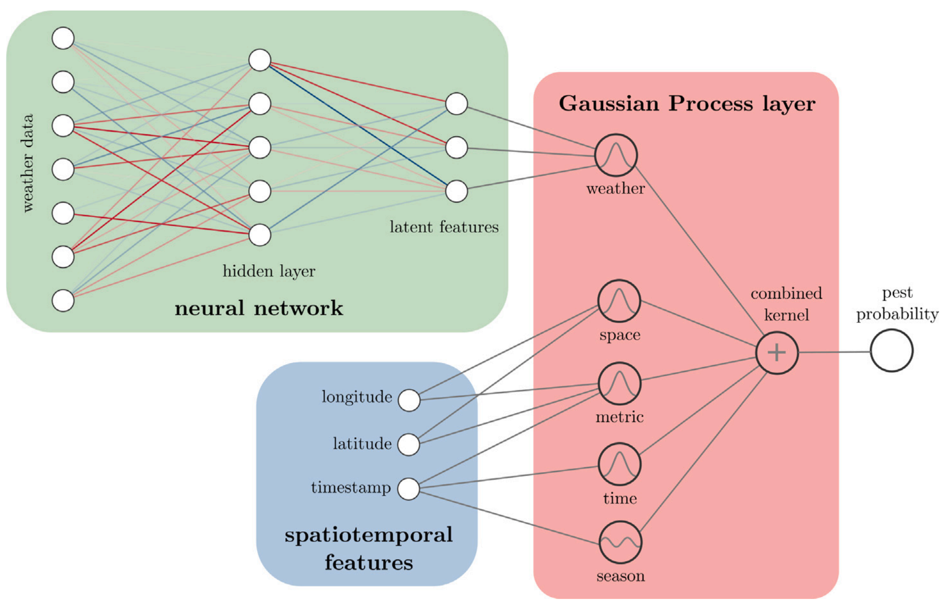

Fig. 2. Schematic representation of the model architecture. The dimensionality of the weather data input (=270) was reduced for this visualization.

$$k_{\text{season}}(x, x') = \sigma^2_{\text{season}} \exp\left(-2\frac{\sin^2\left(\frac{\pi}{p}(x_{\text{time}} - x'_{\text{time}})\right)}{\ell_{\text{season}}}\right) \cdot \exp\left(-\frac{(x_{\text{time}} - x'_{\text{time}})^2}{2\ell_{\text{season\_rbf}}}\right) \quad (4)$$

$$k_{\text{weather}}(x, x') = \sigma^2_{\text{weather}} \exp\left(-\frac{1}{2}\sum_{i=1}^{3}\frac{(x_{\text{weather},i} - x'_{\text{weather},i})^2}{\ell_{\text{weather},i}}\right) \quad (5)$$

$$k(x, x') = k_{\text{space}}(x, x') + k_{\text{time}}(x, x') + k_{\text{metric}}(x, x') + k_{\text{season}}(x, x') + k_{\text{weather}}(x, x') \quad (6)$$

The parameters of the Gaussian process model consist of the output variances σ^2, lengthscale parameters ℓ, period length p, and anisotropy coefficient α. Together with the neural network weights, they were fitted end-to-end using Adam (Kingma and Ba, 2017). The posterior distribution of the Gaussian process was estimated using stochastic variational inference (SVI) with 24 inducing points. For that we used the Natural Gradient Descent optimization algorithm, which has been shown to be effective for variational inference in Gaussian process models (Salimbeni et al., 2018). We selected common learning rates of 0.02 for the kernel parameters, 0.001 for the neural network and 0.1 for SVI. We implemented the model in Python using pyTorch (Paszke et al., 2019) and gPytorch (Gardner et al., 2021). All code was executed on the Google Colab platform (Google, 2024) using a Nvidia Tesla T4 GPU for training. The covariograms were created using the STIF Python package (Kopton, 2024). All our model and visualization source code, as well as synthetic example data, is available at https://github.com/johanneskopton/pest-risk-decision.

We evaluated the accuracy of the pest risk prediction model by 2-fold time series cross-validation, i.e., the data were divided into two sets each at two different time points. Each time, the earlier set was used to train the hyperparameters and the later set was used to test the accuracy of the prediction. To fit the posterior distribution for each day in the test set, data from the previous days in the test set were also included to resemble a real-world prediction scenario.

2.3. Decision model

For the decision model, we followed the expected utility hypothesis, i.e., we assumed that decision-makers are von Neumann–Morgenstern rational (Neumann et al., 2007) agents who try to maximize the expected value (probability-weighted average) of a utility function u. In the following section, we first consider the decision whether or not to apply a pesticide (*treatment*). Then we extend this to the decision of whether to do nothing (*inaction*), to monitor the crop (*monitoring*) for the given pest, or to apply crop protection immediately. We used a simple economic utility function, representing the difference in farm profits compared to a hypothetical situation without the pest:

$$u(\alpha,\beta) = \alpha \cdot (\beta \cdot -C_{\text{loss}|\beta} + (1-\beta)\cdot -C_{\text{loss}|\neg\beta}) + \beta \cdot -C_{\text{treatment}} \quad (7)$$

with u as utility, α as pest presence indicator (0=absence, 1=presence), β as treatment indicator (0=no treatment, 1=treatment), $C_{\text{treatment}}$ as treatment cost, $C_{\text{loss}|\beta}$ as yield loss cost given treatment, and $C_{\text{loss}|\neg\beta}$ as yield loss cost given no treatment. Without loss of generality, this can be shifted to obtain a simpler transformed utility function with $\tilde{u}(\beta = 0) = 0$. Treatment is then the optimal decision option if $\tilde{u}$ is positive. Its expected value is now given by:

$$E[\tilde{u}](p_\alpha, \beta = 1) = p_\alpha \cdot E[\Delta C_{\text{loss}}] - E[C_{\text{treatment}}] \quad (8)$$

with $p_\alpha = E[\alpha]$ as expected value for the Bernoulli-distributed pest presence variable α, and $E[\Delta C_{\text{loss}}]$ as shorthand for expected avoidable yield loss $E[C_{\text{loss}|\neg\beta}] - E[C_{\text{loss}|\beta}]$. Treatment is profitable under the following condition:

$$E[\tilde{u}](p_\alpha, \beta = 1) > 0 \quad \Leftrightarrow \quad p_\alpha > \frac{E[C_{\text{treatment}}]}{E[\Delta C_{\text{loss}}]} \quad (9)$$

From this, the expected value without perfect information (PI) on α can be derived. Depending on the prior probability of pest presence p_α, a prior optimal decision on treatment β (denoted as $\beta_{\neg\text{PI}(\alpha)}$) is assumed:

$$\begin{aligned} E[\tilde{u}](p_\alpha, \beta = \beta_{\neg\text{PI}(\alpha)}) &= \min_\beta E[\tilde{u}](p_\alpha, \beta) \\ &= \begin{cases} E[\tilde{u}](p_\alpha, \beta = 1), & p_\alpha > \frac{E[C_{\text{treatment}}]}{E[\Delta C_{\text{loss}}]} \\ E[\tilde{u}](p_\alpha, \beta = 0), & \text{otherwise} \end{cases} \end{aligned} \quad (10)$$

The expected value given PI on α is then calculated, assuming a posterior optimal decision on treatment β (denoted as $\beta_{\text{PI}(\alpha)}$), i.e. treatment if the pest was found and no treatment otherwise:

$$\begin{aligned} E[\tilde{u}](p_\alpha, \beta = \beta_{\text{PI}(\alpha)}) &= p_\alpha \cdot E[\tilde{u}](p_\alpha = 1, \beta = 1) \\ &+ (1 - p_\alpha) \cdot E[\tilde{u}](p_\alpha = 0, \beta = 0) \end{aligned} \quad (11)$$

From this, we obtained the expected value of perfect information (EVPI) for α, i.e. the value of monitoring. This can be thought of as the expected (opportunity) cost of making a suboptimal decision due to imperfect information and is given by the difference between the expected value given PI and the expected value without PI (Howard, 1966, Eq. 13):

$$\text{EVPI}_\alpha(p_\alpha) = E[\tilde{u}](p_\alpha, \beta = \beta_{\text{PI}(\alpha)}) - E[\tilde{u}](p_\alpha, \beta = \beta_{\neg\text{PI}(\alpha)}) \quad (12)$$

The EVPI reaches its maximum where the prior optimal decision $\beta_{\neg\text{PI}(\alpha)}$ is indifferent (i.e. at the decision boundary):

$$\begin{aligned} \max_{p_\alpha} \text{EVPI}_\alpha(p_\alpha) &= \text{EVPI}_\alpha \left(p_\alpha = \frac{E[C_{\text{treatment}}]}{E[\Delta C_{\text{loss}}]} \right) \\ &= E[C_{\text{treatment}}] - \frac{E[C_{\text{treatment}}]^2}{E[\Delta C_{\text{loss}}]} \end{aligned} \quad (13)$$

The third decision option *monitoring* is thus optimal, if the EVPI exceeds the monitoring costs (see Fig. 6 in Section 3.2 for the application on the case study). Finally, the following overall utility for the three decision options was obtained:

$$E[\tilde{u}](p_\alpha)_{\text{inaction}} = 0 \quad (14)$$

$$E[\tilde{u}](p_\alpha)_{\text{monitoring}} = p_\alpha \cdot (E[\Delta C_{\text{loss}}] - E[C_{\text{treatment}}]) - E[C_{\text{monitoring}}] \quad (15)$$

$$E[\tilde{u}](p_\alpha)_{\text{spraying}} = p_\alpha \cdot E[\Delta C_{\text{loss}}] - E[C_{\text{treatment}}] \quad (16)$$

An optimal crop protection strategy (for the individual farmer) is achieved by maximizing the expected utility given a predicted pest risk from the prediction model ($\max E[\tilde{u}](p_\alpha)$).

2.4. Assumptions for the case study farm

The decision model was applied to the case study of *T. absoluta* control in tomato production. Even within the state of Andhra Pradesh, production conditions and practices vary widely between farms. These variable characteristics include tomato variety (Prashanthi et al., 2022), farm size (Depenbusch et al., 2023), pest pressure (Lakshmi et al., 2019), crop protection practices (Ribka et al., 2020) and marketing (Yesdhanulla and Aparna, 2018). To assess the potential benefits of our pest prediction model and demonstrate the value of monitoring, we modeled an archetypical smallholder tomato farm using estimates from the literature and from experts.

Since we relied on Plantix data, visible symptoms were considered *pest presence* (rather than e.g. observations from a pheromone trap). Visible symptoms also correspond to the economic threshold level (Shiberu and Getu, 2018), so yield is assumed to be impaired, whenever *pest presence* is observable. Most farmers in Andhra Pradesh are able to identify *T. absoluta* without help (Depenbusch et al., 2023), so we do not have to assume a digital pest recognition tool like Plantix for successful monitoring. Total yield loss caused by *T. absoluta* is modeled as a function of days with pest presence. Due to the year-round tomato production in the study area, it is unknown, in which crop growth stage the pest observations took place. We therefore assume a uniform prior. Since the damage associated with each day of pest presence needs to sum to the total observed damage, we assume that (on average) one day of pest presence causes the total observed damage divided by the number of cultivation days with pest presence. Without treatment, *T. absoluta* symptoms are present in about 11% of tomato plants in Andhra Pradesh (Buragohain et al., 2021). Using an average production period of 116 days (Depenbusch et al., 2023), on average *T. absoluta* was present on 10.5 days per production cycle. Without treatment, avoidable yield loss from *T. absoluta* is assumed to be 9100 INR/ha (Buragohain et al., 2021). This results in 868 INR/ha of yield loss per day with pest presence. Treatment is modeled following conventional farmers' practices. While in the academic literature, integrated pest management (IPM) is considered a favorable approach for mitigating damage from *T. absoluta* in India (Sridhar et al., 2019; Buragohain et al., 2021), this framework has not yet been widely adopted in Andhra Pradesh, e.g. less than 1% of the quantity of pesticides applied are biopesticides (Depenbusch et al., 2023). Thus, application of the insecticide Chlorantraniliprole was assumed as the only treatment, as is common practice in AP (Buragohain et al., 2021). According to Buragohain et al. (2021), 7 applications of pesticides in calendar-based spraying (every 10 days, after 4 weeks) cost on average 4167 INR/ha per season, so 595 INR/ha per application. This practice was taken as reference for zero avoidable yield loss caused by *T. absoluta*. Therefore, we modeled pesticide application aimed at full protection of the plants during the following 10 days. We assumed 600 INR/day (INR Labour Bureau, 2024) as value of working time. Applying pesticides using a knapsack sprayer was assumed to take one third of a working day per hectare for a single worker. Assuming an agricultural worker could monitor 20 fields per day, with an average field size of 0.42 ha (calculated from Depenbusch et al. (2023)), this results in 72 INR/ha monitoring costs. Since treatment was effective for 10 days while decisions are made daily, the cost of treatment per day per hectare in the decision model was the cost per treatment divided by 10. To correct for the possibility of omitting monitoring during these 10 days, in the decision model, monitoring and treatment costs are reduced by one third of the monitoring costs, i.e. by 24 INR/ha.

2.5. Population model

To investigate the average benefit of using the pest prediction model to inform crop protection decisions, we modeled a population of farms in Andhra Pradesh over the course of the year 2022 using a simple agent-based model. All farms are assumed to have the archetypical characteristics described in Section 2.4. We used a set of 8925 farm locations with high-quality Plantix tomato observations in the area during that year. We randomly sampled transplanting dates from a uniform distribution, constrained so that the corresponding Plantix observation fell within the growing period. "True" pest occurrences were sampled from random realizations of the multivariate normal distribution defined by the Gaussian process fine-tuned to 2022. Thus, although the true distribution of pests could not be known for all locations, this generated ground truth has the same properties in terms of average infestation, as well as autocorrelation and correlation with weather data. We evaluated 5 scenarios in which all agents follow one of 5 crop protection strategies respectively. In the *inaction* scenario, all agents simply accept any damage caused by *T. absoluta* without intervention or monitoring. In the *calendar spraying* scenario, all agents follow the conventional farmers' practice of applying the insecticide chlorantraniliprole every 10 days. In the *daily monitoring* scenario, all agents follow the common Integrated Pest Management advice to check for infestations every day and apply pesticides only if they find positive results. During the 10-day protected period, no monitoring is done by the agents. In the *model* scenario, all agents followed the model's advice to choose the decision option that maximizes expected utility. Whenever monitoring was performed by an agent, the results were used to update the pest risk model to inform future predictions. In the *model (no updating)* scenario, the model trained only on data through December 2021 was used. The entire simulation was repeated 10 times as a Monte Carlo simulation to average out sampling effects.

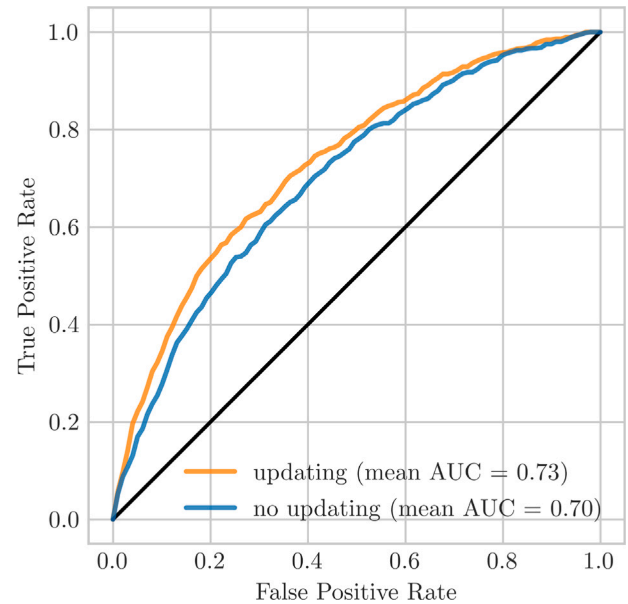

Fig. 3. Receiver operating characteristic (ROC) for the time series cross-validation. In the *updating* case, new observations were incorporated in near-realtime, while in the *no updating* case, only observations from the training set in the past were used. The black line represents a theoretical random classifier for reference. As a summary statistic, the area under the curve (AUC) is given (0.5: random classifier, 1: perfect classifier).

3. Results and discussion

3.1. Pest risk prediction model

For the purpose of informing decisions under uncertainty, the model must not only accurately discriminate between presence and absence (discriminative power), but also quantify its own uncertainty in terms of Bayesian probabilities. Therefore, we evaluated the model in two ways: The discriminatory power was assessed using the receiver operating characteristic (ROC), as shown in Fig. 3. With a mean area under the curve (AUC) of 0.73 in the *updating* case, the model shows significant class separation. The difference to the AUC of 0.70 in the *no updating* case clearly shows that the near-realtime utilization of new observation data crucially improved the model. These findings indicate that monitoring can not only benefit the individual farmer, but also other users of the prediction model. This community VoI is quantified in Section 3.3. As expected for complex systems such as pest propagation, the predictive accuracy is far from 100%, so we have placed particular emphasis on quantifying the model uncertainty in the further analysis.

The accuracy of the uncertainty quantification was evaluated by comparing the percentage of observed presence samples to the predicted presence probabilities (see left part of Fig. 4). For the majority of predictions, the uncertainty quantification was accurate (e.g., $\approx 10\%$ of samples with a predicted risk of about $\approx 10\%$ were *presence* samples). The expected calibration error (ECE) was 5.0%. Compared to other classification models, especially deep learning methods (Guo et al., 2017), the uncertainty quantification was highly accurate. The model produced risk predictions ranging from 0% to about 50%, and the distribution of predictions was significantly skewed so that the most likely predictions were lower than the defined mean of 11% (see histogram in Fig. 4).

The model predictions demonstrate that the model was able to capture the spatial distribution of the pest risk, with local hotspots due to recent outbreaks and weather patterns (as shown for three exemplary dates in Fig. 5). The model also captured the temporal dynamics, with lower risks in the monsoon season (June to October).

Since our model shows not only good class separation but also accurate uncertainty quantification, it is well suited to serve as a basis for probabilistic decision modeling. However, some limitations must be considered: We assumed that Plantix observations of *T. absoluta* were representative of the distribution of all *T. absoluta* occurrences, but the data was preferentially sampled: The Plantix Health Check is likely to be used more often for symptoms that are not widely known, so not only healthy crops but also common pathogens such as *T. absoluta* may be underrepresented. While we scaled pest risk using the average probability of occurrence from an external source to mitigate this, the bias could vary in space and time (e.g., the disease has become better known over time, farmers in certain regions are more sensitized to this pest, etc.). In addition, we did not have data on the autocorrelation of infestations within the same field, so we simply modeled the same field (on another day) as a field with zero spatial distance. While the available training data was not perfectly unbiased or complete, it is, to our knowledge, the most complete and representative database of pest observations that exists.

3.2. Decision model

The results of the decision model show that both the value of monitoring (i.e., the EVPI for pest risk) and the recommended decision option were highly dependent on pest risk. At 52 INR ha^{-1}, the maximum EVPI (at $p_\alpha = 6.4\%$) was not much higher than the monitoring cost of 48.0 INR ha^{-1} (see Fig. 6). However, *monitoring* was the optimal decision option in a range from 5.9% to 13.5% pest risk, which included 15.8% of predictions (see Fig. 4). As can be seen, these results were very sensitive to the assumptions we made for the archetypical tomato farm (as detailed in Section 2.4), so they need to be interpreted with caution.

In terms of expected profits (as given by Eqs. (14)–(16)), the *monitoring* and *treatment* options were very similar (see Fig. 7), although their implications for agricultural practice (in terms of pesticide use, labor demand, etc.) were drastically different. Since the *inaction* option yielded significantly higher expected utility for low (i.e., $< 5.9\%$) pest risks, the decision to do either monitoring or direct treatment, or to do nothing, had a substantial impact on farm profitability. This indicates that prior information on pest risk even before monitoring

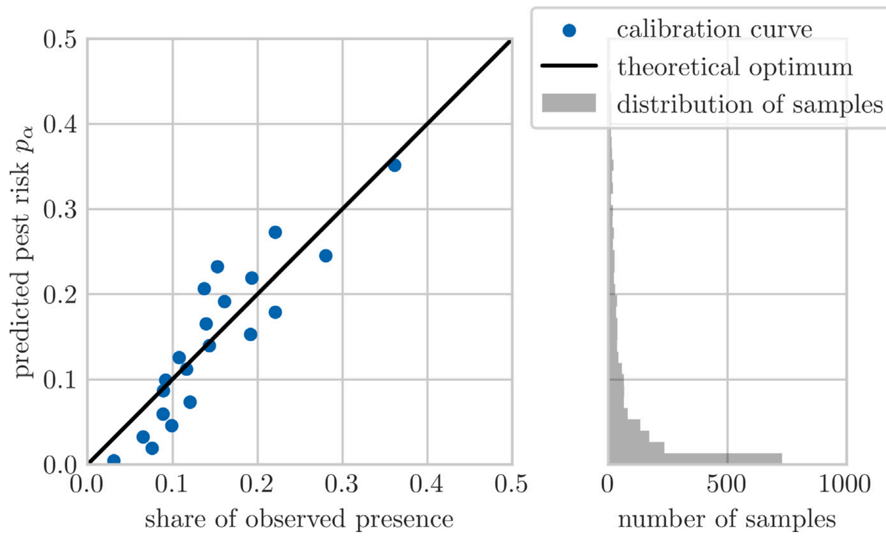

Fig. 4. Calibration curve (left) and distribution of predictions (right). The blue dots show the percentage of observed presence samples as a function of the predicted probability of presence (both scaled by the same factor to fit the defined global mean of 11%). The black line shows the theoretically optimal uncertainty quantification for reference. The histogram shows the number of samples per bin of predicted probability.

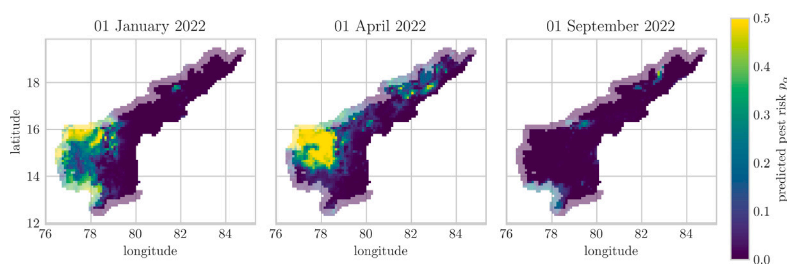

Fig. 5. Predicted pest risk (probability of pest presence) on three sample days in Andhra Pradesh. The hyperparameters were trained once on data for the years 2018–2021, and the probability distribution was fitted using additional data from the respective previous days of 2022. Results in the buffer around the state are shown with lower opacity.

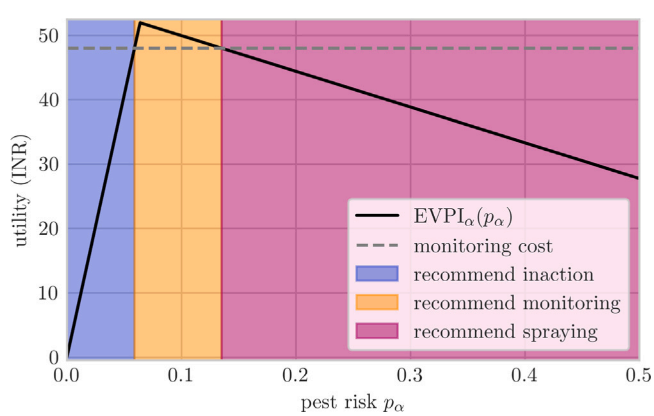

Fig. 6. Expected Value of Perfect Information (EVPI) and monitoring costs as a function of pest risk. The decision recommendation is indicated by the background color.

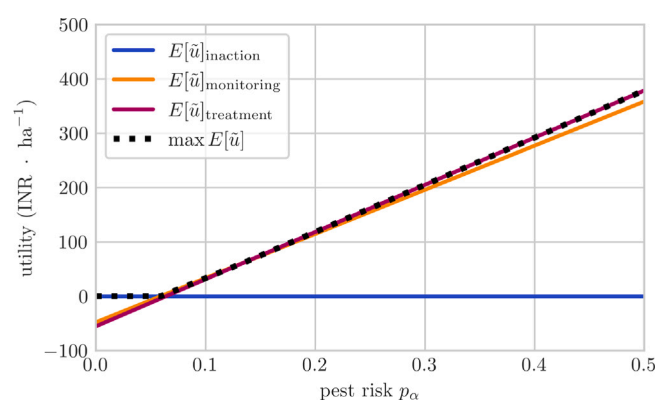

Fig. 7. Expected utility for the 3 decision options over pest risk. The expected utility of the optimal decision option is shown with a black dotted line.

(as can be provided by the predictive model) can be very valuable to decision-makers.

In regions with very low pest risk, a low individual EVPI can be observed (see Fig. 8). As expected, EVPI was highest at the borders of high-risk regions. While the preventive application of pesticides is common practice among tomato growers in Andhra Pradesh (Buragohain et al., 2021), on these exemplary days this option was considered optimal only in parts of the state (see Fig. 9). Where the disease risk was predicted to be low, monitoring or even inaction minimized operation costs by saving pesticides and working time. These results could be valuable recommendations to be used in digital extension services, as suggested by Taylor et al. (2023).

While our findings provide valuable insights into the economics of monitoring, and the decision model could already be useful for farm-level decision support, some further steps need to be taken to realize its full potential in practice. Perhaps the most relevant simplification was our restriction of the scope of the analysis to only a single pest. In reality, a single field inspection could reveal infestations of other pathogens at little or no additional cost. Furthermore, insecticides such as the one considered in our case study (chlorantraniliprole) can be

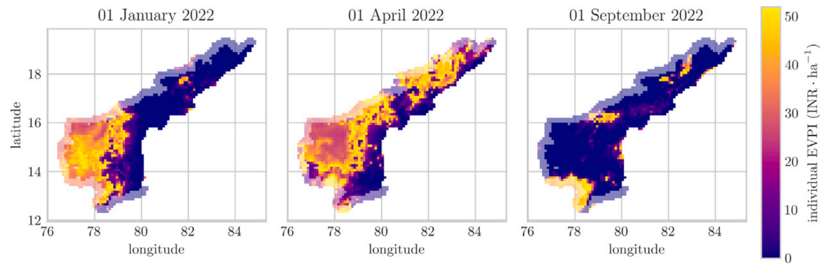

Fig. 8. Expected monetary benefit of pest monitoring for the individual farm, i.e. individual Expected Value of Perfect Information (EVPI) for pest risk. The calculations are based on the model predictions shown in Fig. 5.

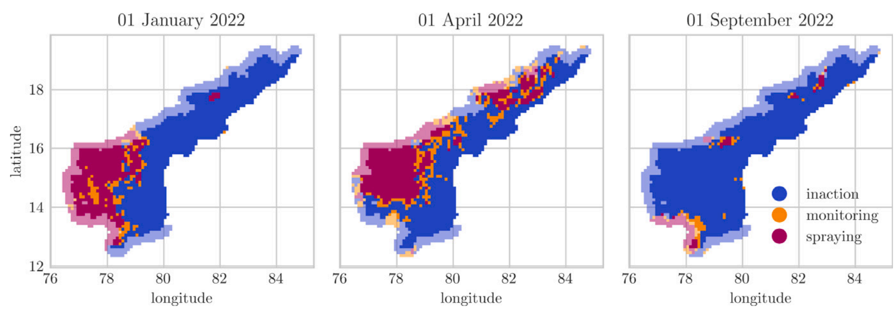

Fig. 9. Decision recommendation based on our decision model using the predicted pest risks (as shown in Fig. 5) as input. Monitoring is the recommended action, whenever the individual EVPI (as shown in Fig. 8) is greater than the cost of monitoring.

effective against multiple pests simultaneously. By extending our risk model to jointly predict the risk of multiple pathogens (as was done, for example, by Patil and Thorat (2016) for grapes), and by modifying our decision model to find the optimal strategy given this joint distribution of risk, farmers can be advised more holistically.

Another simplification is the lack of spatiotemporal dynamics in our pest damage model. Instead of assuming a linear dependence on pest presence, plant growth models can capture complex nonlinear relationships with pathogens (Bregaglio et al., 2021). However, it is unclear whether such complex model extensions would significantly improve the overall results.

In addition, we considered only one exemplary set of farm characteristics, modeling an archetypical farm. To provide more meaningful decision support for individual farmers, we recommend including farm-specific information on tomato variety, management practices, labor availability, production paradigm (e.g., organic, IPM, conventional, etc.), transplanting time, farmer knowledge, product availability, and local input and output prices. Such targeted utilization of our prediction and decision model could be achieved through integration with existing Farm Management Information Systems (FMIS) (Balkrishna et al., 2023) where such data are already available. Since our flexible, open-source model can be easily applied to other crops and pathogens, it can serve as the basis for farm-level decision support systems, especially given user-specific information on farming conditions and practices.

3.3. Potential benefits of pest risk prediction and value of monitoring

To assess the potential economic benefits of following our decision recommendations throughout the year, we used our agent-based population model.

The strategy using our predictive decision model had the lowest total cost (see Table 1). The cost was lower by 1713.8 ± 237.3 INR ha^{-1} a^{-1} (or 25.4 ± 4.3 %) than the farmers' common practice *calendar spraying*. This difference represents the possible expected annual monetary benefit of using our model. In addition to these monetary benefits, we found even more significant environmental and health benefits because the composition of the costs was drastically different (see Fig. 10). While in *calendar spraying* all the costs were caused by the purchase and application of pesticides, in the *model* strategy the costs also included working hours for monitoring as well as minor avoidable yield losses. This way, we obtained a possible reduction in pesticide use of 58.8 ± 2.7 %, demonstrating a significant potential for improving production sustainability. At 8388.9 ± 8.9 INR ha^{-1} a^{-1}, the *daily monitoring* strategy was almost as favorable as the *calendar spraying* strategy, while still saving 53.4 ± 0.4 % in pesticide use. At 9568.7 ± 374.2 INR ha^{-1} a^{-1} of total cost (all due to yield loss), *inaction* performed the worst. With zero standard deviation, the costs of the *calendar spraying* strategy were insensitive to the distribution of pests, which could explain the attractiveness of this strategy to (risk-averse) farmers.

Table 1
Average decision-relevant costs (i.e., negative utility) per hectare for the prediction period (Jan 1, 2022 to Dec 31, 2022). All results are presented as mean ± standard deviation of the Monte Carlo samples. Each row represents a scenario in which all agents follow the respective decision strategy.

Strategy	Average costs per ha and year in INR
inaction	9568.7 ± 374.2
daily monitoring	8388.9 ± 8.9
calendar spraying	8490.2 ± 0.0
model	6776.3 ± 237.3
model (no updating)	7580.3 ± 163.8

The mean individual EVPI for pest risk (i.e. the expected monetary benefits of pest monitoring for the individual farm) was 26.6 ± 4.5 INR ha^{-1} for monitoring in the *model* scenario and 18.8 ± 2.9 INR ha^{-1} in the *model (no updating)* scenario. This difference can be explained by model updating leading to better predictions and thus fewer pesticide applications, therefore monitoring becomes more useful. Since *T. absoluta* infestations were most often observed during the rabi season (November to May), inaction was less often considered optimal during this period (see Fig. 11). However, even during the times of peak pest pressure, only half of the agents were advised to apply pesticides preventively. Thus, the use of pesticides was considerably lower throughout the year than in the *calendar spraying* scenario.

The mean community EVPI for pest risk (i.e. the expected monetary benefits of pest monitoring for all the other farms) was calculated as follows: The *model (no updating)* scenario, in which all farms follow

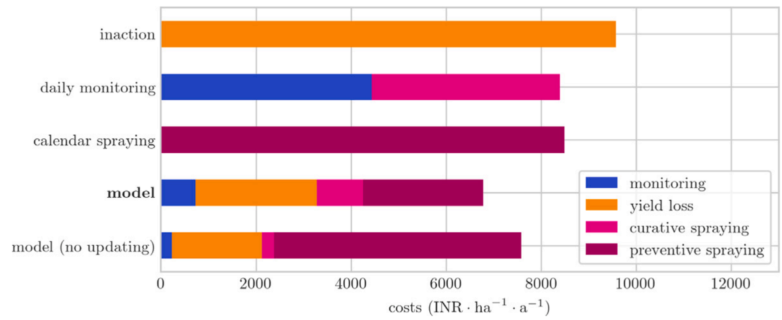

Fig. 10. Average decision-relevant costs per agent (farm) for 2022, when all agents follow the respective decision strategy. Cost contributions are indicated by color. *Curative spraying* refers to pesticide application as a result of positive findings in monitoring.

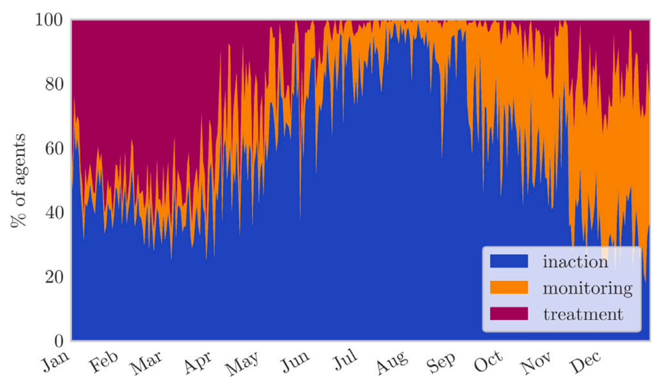

Fig. 11. Percentage of agents for whom the respective decision options were considered optimal in the *model* scenario over time. The simulation was run from January to December of 2022.

the decision recommendation, but the prediction model is not updated with new observations, led to 7580.3 ± 163.8 INR ha^{-1} a^{-1} of total costs. This is 804.0 ± 134.6 INR ha^{-1} a^{-1} more than the *model* strategy (with updating). Divided by the number of monitoring events (10.1 ± 2.6 · a^{-1}), this results in a mean community EVPI for monitoring of 88.2 ± 34.2 INR ha^{-1}. Thus, on average each monitoring provided even more value to the surrounding farms than it did for the monitoring farm itself (the individual EVPI can never exceed 51.95 INR ha^{-1}, see Fig. 6). As shown in Fig. 1, risks were significantly correlated for farms in close spatial proximity, so farms benefited most from nearby observations. However, due to the nature of the Gaussian process model, also more distant locations that shared similar environmental conditions were informed. Therefore, the *community* includes all farms using and contributing to the same pest risk prediction model. Monitoring has diminishing marginal returns, since the first observations are significantly more informative than later ones. Therefore, additional monitoring runs would likely provide less value than this mean community EVPI. In addition, the community EVPI is likely highly variable, depending on how many observations from the same region and time period are already available to the model, and how many other tomato growing farms are in the spatiotemporal neighborhood to benefit from them. Further analyzing the dynamics of community EVPI would be a valuable topic for future research.

In practice, farmers tend to be risk-averse, and this has been shown to be the case in India (Sawosri and Mußhoff, 2020; Senapati, 2020). This implies that the occurrence of avoidable losses, along with visible symptoms and the associated risk of pest spread, may lead farmers to rely on heuristic calendar-based spraying. That is, they may be unwilling to follow the monitoring and intervention scheme recommended by the model, even though the expected total costs could be lower. In addition, in many rural areas of India, temporal fluctuations in labor supply, coupled with imperfect labor markets and labor market transaction costs (Foster and Rosenzweig, 2017), may limit the monitoring capacity of medium-size farms. In such cases, calendar-based spraying offers the added benefit of planning certainty.

In the context of a free mobile application, the quality of pest risk prediction can be seen as a public good (non-excludable, non-rival) whose supply depends on monitoring. Monitoring thus provides a positive externality with diminishing marginal returns, in the sense that each successive observation provides less uncertainty reduction than the previous one. Since individuals can benefit from pest risk prediction without contributing costly monitoring observations, this is a classic free-riding situation. Rational individuals will engage in monitoring only if the private benefits outweigh the costs. Neglecting the public benefits leads to a socially suboptimal undersupply of monitoring activities. We therefore discuss some approaches to remedy this undersupply. First, public extension agents can contribute monitoring activities in regions and at times where it is beneficial. Second, the positive externality can be internalized, so that a farmer receives additional compensation or some kind of subsidy for monitoring based on the value it provides to the surrounding community. Apart from monetary compensation, which may be unrealistic in practice due to transaction costs and unclear payment levels, non-monetary compensation can also be used, such as social prestige rewards. Such rewards have previously been shown to induce the sharing of pest management information in Ethiopia (Balew et al., 2023). In practice, such rewards could be implemented as badges for mobile app users that contributors can earn and display publicly if they wish.

While our model clearly demonstrates the potential benefits of using a probabilistic pest risk prediction model and the importance of community VoI in evaluating monitoring efforts, the specific model results depend on a number of assumptions and simplifications. The limitations of the underlying risk prediction and decision models are discussed in Sections 3.1 and 3.2, respectively. In the population model, we assumed that the Plantix tomato observation sites are a representative sample of tomato farmer sites in Andhra Pradesh, but they are likely to be biased due to socioeconomic drivers of technology adoption (Schulz and Börner, 2023). In addition, the homogeneity of agents (all modeled as archetypical farms) is also a limitation of our population model. For more reliable policy advice, we suggest drawing populations from a distribution of representative farms. This would be similar to the approach of Luu et al. (2022), who used a fully probabilistic model derived from extensive expert consultation. In our case, further improvements could be achieved by regionalizing agronomic and economic variables such as farm sizes, wages, and commodity prices. In addition to the aforementioned model improvements, another promising avenue for advancing research in this area could be the development of long-term pest forecasts based on weather forecasts. This approach could enable the implementation of environmentally friendly preventive measures, thereby further reducing the need for pesticides.

4. Conclusion

Crop pests continue to be a major problem for farmers around the world. At the same time, overuse of pesticides poses a threat to human and environmental health. This study presents a framework to support decisions on whether to apply pesticides, monitor crops, or omit crop protection measures on a given day and at a given geographic location. To achieve this, a spatio-temporal pest risk prediction model was developed using deep Gaussian process classification. In a case study of tomato production in Andhra Pradesh, the prediction model showed considerable discriminative power and accurate uncertainty quantification. For an archetypical tomato farm, our model-based decision strategy proved highly valuable in a Monte Carlo simulation, reducing average pest-related costs by 25.4$\pm$4.3% and pesticide use by 58.8$\pm$2.7%. The information value of monitoring one hectare of tomato crop was found to be 26.6 $\pm$ 4.5 INR on average for the individual farmer. We found 88.2$\pm$34.2 INR ha^{-1} of additional information value in monitoring for the community, as other farmers are warned of disease outbreaks. For this purpose, we introduced the novel decision analytic concept of community VoI. To apply our framework in a decision support context, we suggest utilizing current, farm-specific data instead of the literature-based data used in this case study. Furthermore, integrating multiple crop pests and diseases is anticipated to enhance our approach for practical decision-making. Overall, our adaptable, open-source risk prediction and decision models provide a robust foundation for pest risk-based decision support systems.

CRediT authorship contribution statement

Johannes Kopton: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Sytze de Bruin:** Writing – review & editing, Methodology, Conceptualization. **Dario Schulz:** Writing – review & editing, Writing – original draft, Conceptualization. **Eike Luedeling:** Writing – review & editing, Supervision, Funding acquisition.

Funding

This work was partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC 2070–390732324-PhenoRob. Open Access funding enabled and organized by Projekt DEAL.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank all the kind people at Plantix (especially Rob Strey, Thomas Friedel, Stefanos Papaiordanidis, Andjela Komnenic and Dimo Dimov) for providing the pest observation data and for their generous support and advice.

Appendix

See Figs. A.1 and A.2.

Data availability

The pest observation data that has been used is confidential. Processed weather data and the model code are available at https://github.com/johanneskopton/pest-risk-decision.

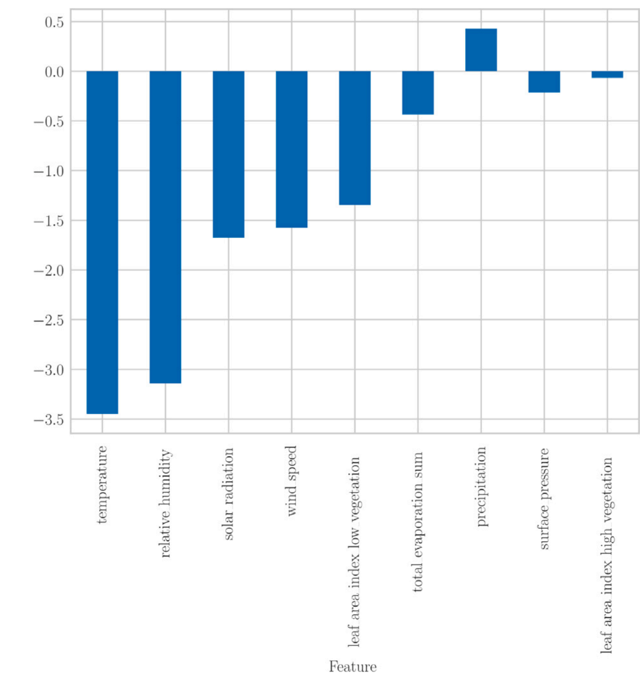

Fig. A.1. Coefficients of a logistic regression model. Positive coefficients correspond to positive correlation with pest risk. Negative coefficients correspond to negative correlation with pest risk. Values close to zero indicate little linear correlation with pest risk but these variables could still be valuable predictors in more complex nonlinear models (such as a random forest classifier, see Fig. A.2) or the deep GPR model, we presented in the paper. Daily weather variables of the previous 30 days were aggregated to averages.

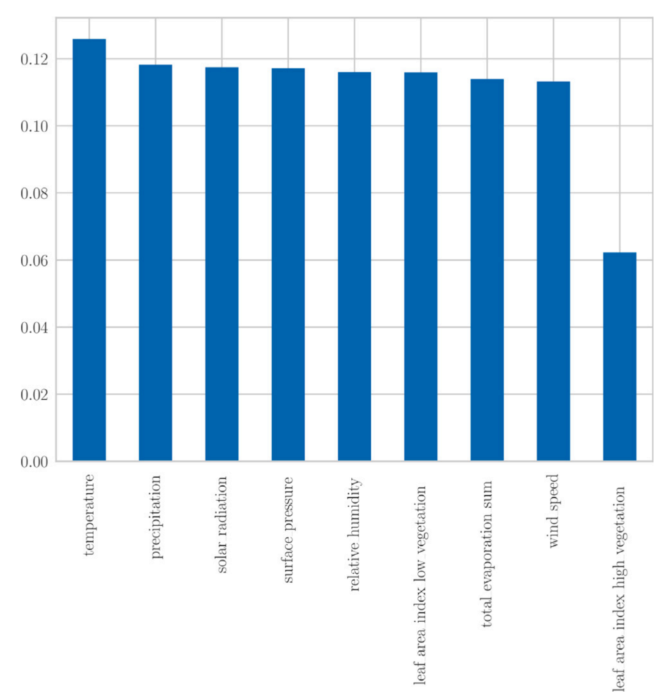

Fig. A.2. Feature importance in a random forest classification model with 100 trees. Daily weather variables of the previous 30 days were aggregated to averages. Apart from *leaf area index high vegetation*, all variables showed similar importance.

References

Agrios, George N., 2008. Plant pathology, 5th ed., [3rd print] Elsevier Academic Press, Amsterdam, ISBN: 978-0-12-044565-3.

Balew, S., Bulte, E., Abro, Z., Kassie, M., 2023. Incentivizing and nudging farmers to spread information: Experimental evidence from Ethiopia. Am. J. Agric. Econ. 105, 994–1010. http://dx.doi.org/10.1111/ajae.12346, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/ajae.12346.

Balkrishna, A., Pathak, R., Kumar, S., Arya, V., Singh, S.K., 2023. A comprehensive analysis of the advances in Indian Digital Agricultural architecture. Smart Agric. Technol. 5, 100318. http://dx.doi.org/10.1016/j.atech.2023.100318, URL: https://www.sciencedirect.com/science/article/pii/S2772375523001478.

Bhatia, A., Chug, A., Singh, A.P., 2020a. Application of extreme learning machine in plant disease prediction for highly imbalanced dataset. J. Stat. Manag. Syst. URL: https://www.tandfonline.com/doi/abs/10.1080/09720510.2020.1799504.

Bhatia, A., Chug, A., Singh, A.P., 2020b. Hybrid SVM-LR classifier for powdery mildew disease prediction in tomato plant. In: 2020 7th International Conference on Signal Processing and Integrated Networks (SPIN). IEEE, Noida, India, pp. 218–223. http://dx.doi.org/10.1109/SPIN48934.2020.9071202, URL: https://ieeexplore.ieee.org/document/9071202/.

Biondi, A., Guedes, R.N.C., Wan, F.H., Desneux, N., 2018. Ecology, worldwide spread, and management of the invasive South American tomato pinworm, Tuta absoluta : Past, present, and future. Annu. Rev. Entomol. 63, 239–258. http://dx.doi.org/10.1146/annurev-ento-031616-034933, URL: https://www.annualreviews.org/doi/10.1146/annurev-ento-031616-034933.

Boedeker, W., Watts, M., Clausing, P., Marquez, E., 2020. The global distribution of acute unintentional pesticide poisoning: Estimations based on a systematic review. BMC Public Health 20, 1875. http://dx.doi.org/10.1186/s12889-020-09939-0.

Bregaglio, S., Willocquet, L., Kersebaum, K.C., Ferrise, R., Stella, T., Ferreira, T.B., Pavan, W., Asseng, S., Savary, S., 2021. Comparing process-based wheat growth models in their simulation of yield losses caused by plant diseases. Field Crop. Res. 265, 108108. http://dx.doi.org/10.1016/j.fcr.2021.108108, URL: https://www.sciencedirect.com/science/article/pii/S037842902100054X.

Buragohain, P., Saikia, D.K., Sotelo-Cardona, R., 2021. Development and validation of an integrated pest management strategy against the invasive South American tomato leaf miner, Tuta absoluta in South India. Crop. Prot. 139, 105348. http://dx.doi.org/10.1016/j.cropro.2020.105348, URL: https://www.sciencedirect.com/science/article/pii/S0261219420302817.

Copernicus Climate Change Service, 2024. ERA5-Land daily data from 1950 to present. URL: https://developers.google.com/earth-engine/datasets/catalog/ECMWF_ERA5_LAND_DAILY_AGGR.

de Bruin, S., Hunter, G.J. and, 2003. Making the trade-off between decision quality and information cost. Photogramm. Eng. Remote Sens. 69, 91–98. http://dx.doi.org/10.14358/PERS.69.1.91.

de Bruin, D., Bregt, A.K., 2012. Where and when should sensors move? Sampling using the expected value of information. Sensors 12, 16274–16290. http://dx.doi.org/10.3390/s121216274, URL: https://www.mdpi.com/1424-8220/12/12/16274.

de Carvalho Alves, M., Pozza, E.A., 2010. Indicator kriging modeling epidemiology of common bean anthracnose. Appl. Geomat. 2, 65–72. http://dx.doi.org/10.1007/s12518-010-0021-1.

de Carvalho Alves, M., da Silva, F.M., Moraes, J.C., Pozza, E.A., de Oliveira, M.S., Souza, J.C.S., Alves, L.S, 2011. Geostatistical analysis of the spatial variation of the berry borer and leaf miner in a coffee agroecosystem. Precis. Agric. 12, 18–31. http://dx.doi.org/10.1007/s11119-009-9151-z.

Depenbusch, L., Sequeros, T., Schreinemachers, P., Sharif, M., Mannamparambath, K., Uddin, N., Hanson, P., 2023. Tomato pests and diseases in Bangladesh and India: Farmers' management and potential economic gains from insect resistant varieties and integrated pest management. Int. J. Pest Manag. 1–15. http://dx.doi.org/10.1080/09670874.2023.2252760.

Desneux, N., Decourtye, A., Delpuech, J.M., 2007. The sublethal effects of pesticides on beneficial arthropods. Annu. Rev. Entomol. 52, 81–106. http://dx.doi.org/10.1146/annurev.ento.52.110405.091440, URL: https://www.annualreviews.org/content/journals/10.1146/annurev.ento.52.110405.091440.

Desneux, N., Han, P., Mansour, R., Arnó, J., Brévault, T., Campos, M.R., Chailleux, A., Guedes, R.N.C., Karimi, J., Konan, K.A.J., Lavoir, A.v., Luna, M.G., Perez-Hedo, M., Urbaneja, A., Verheggen, F.J., Zappalà, L., Abbes, K., Ali, A., Bayram, Y., Cantor, F., Cuthbertson, A.G.S., De Vis, R., Erler, F., Firake, D.M., Haddi, K., Hajjar, M.J., Ismoilov, K., Jaworski, C.C., Kenis, M., Liu, H.t., Madadi, H., Martin, T., Mazih, A., J. Messelink, G., Mohamed, S.A., Nofemela, R.S., Oke, A., Ramos, C., Ricupero, M., Roditakis, E., Shashank, P.R., Wan, F.H., Wang, M.h., Wang, S., Zhang, Y.B., Biondi, A., 2022. Integrated pest management of Tuta absoluta: Practical implementations across different world regions. J. Pest Sci. 95, 17–39. http://dx.doi.org/10.1007/s10340-021-01442-8.

Diggle, P.J., Menezes, R., Su, T.l., 2010. Geostatistical inference under preferential sampling. J. R. Stat. Soc. Ser. C. Appl. Stat. 59, 191–232. http://dx.doi.org/10.1111/j.1467-9876.2009.00701.x, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9876.2009.00701.x.

Fatemi, S.A., Mohammadi, D., Eivazian Kary, N., 2023. Using Nearest Neighbor, Ripley's K Function and Kriging methods to identify distribution pattern of tomato leafminer moth, Tuta absoluta in greenhouse conditions. J. Appl. Res. Plant Prot. 12, 143–152. http://dx.doi.org/10.22034/arpp.2022.15651, URL: https://arpp.tabrizu.ac.ir/article_15651_en.html.

Fenu, G., Malloci, F.M., 2021. Forecasting plant and crop disease: An explorative study on current algorithms. Big Data Cogn. Comput. 5, 2. http://dx.doi.org/10.3390/bdcc5010002, URL: https://www.mdpi.com/2504-2289/5/1/2.

Foster, A.D., Rosenzweig, M.R., 2017. Are there too many farms in the world? Labor-market transaction costs, machine capacities and optimal farm size. https://papers.ssrn.com/abstract=3049733.

Foster, S.D., Shimadzu, H., Darnell, R., 2012. Uncertainty in spatially predicted covariates: Is it ignorable? J. R. Stat. Soc. Ser. C. Appl. Stat. 61, 637–652. http://dx.doi.org/10.1111/j.1467-9876.2011.01030.x, URL: https://academic.oup.com/jrsssc/article/61/4/637/7079603.

Fukushima, K., 1969. Visual feature extraction by a multilayered network of analog threshold elements. IEEE Trans. Syst. Sci. Cybern. 5, 322–333. http://dx.doi.org/10.1109/TSSC.1969.300225, URL: https://ieeexplore.ieee.org/document/4082265.

Gardner, J.R., Pleiss, G., Bindel, D., Weinberger, K.Q., Wilson, A.G., 2021. GPyTorch: Blackbox matrix-matrix Gaussian process inference with GPU acceleration. http://dx.doi.org/10.48550/arXiv.1809.11165, URL: arXiv:1809.11165.

Ghaffari, R., Zhang, F., Iliescu, D., Hines, E., Leeson, M., Napier, R., Clarkson, J., 2010. Early detection of diseases in tomato crops: An Electronic Nose and intelligent systems approach. In: The 2010 International Joint Conference on Neural Networks. IJCNN, pp. 1–6. http://dx.doi.org/10.1109/IJCNN.2010.5596535, URL: https://ieeexplore.ieee.org/document/5596535.

Goldstein, J., Park, J., Haran, M., Liebhold, A., Bjørnstad, O.N., 2019. Quantifying spatio-temporal variation of invasion spread. Proc. R. Soc. B: Biol. Sci. 286.

Google, 2024. Google colaboratory. URL: https://colab.research.google.com/.

Graeler, B., Pebesma, E., Heuvelink, G., 2016. Spatio-temporal interpolation using gstat. R Journal 8, 204–218. http://dx.doi.org/10.32614/RJ-2016-014.

Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q., 2017. On calibration of modern neural networks. http://dx.doi.org/10.48550/arXiv.1706.04599, URL: arXiv:1706.04599.

HajiGhassemi, N., Deisenroth, M., 2014. Analytic long-term forecasting with periodic Gaussian processes. In: Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics. PMLR, pp. 303–311, URL: https://proceedings.mlr.press/v33/hajighassemi14.html.

Hensman, J., Fusi, N., Lawrence, N.D., 2013. Gaussian processes for big data. http://dx.doi.org/10.48550/arXiv.1309.6835, URL: arXiv:1309.6835.

Howard, R.A., 1966. Information value theory. IEEE Trans. Syst. Sci. Cybern. 2, 22–26. http://dx.doi.org/10.1109/TSSC.1966.300074.

INR Labour Bureau, C.D., 2024. Average daily wage rate: rural: agricultural: ploughing and tilling: men: Andhra Pradesh. URL: https://www.ceicdata.com/en/india/average-daily-wage-rate-rural-agricultural-by-state-ploughing-and-tilling/average-daily-wage-rate-rural-agricultural-ploughing-and-tilling-men-andhra-pradesh.

Kingma, D.P., Ba, J., 2017. Adam: A method for stochastic optimization. http://dx.doi.org/10.48550/arXiv.1412.6980, URL: arXiv:1412.6980.

Kopton, J., 2024. STIF: Space-Time Interpolation and Forecasting. Zenodo, http://dx.doi.org/10.5281/zenodo.13271982, URL: https://zenodo.org/records/13271982.

Lakshmi, R., Kumar, P.G., Kumar, M.R., 2019. Biotic constraints in tomato production in Chittoor district of Andhra Pradesh: A review. J. Èntomol. Zoöl. Stud. URL: https://www.semanticscholar.org/paper/Biotic-constraints-in-tomato-production-in-Chittoor-Lakshmi-Kumar/bbae9fa413475ef5b5cbafe158d290228afea56e.

Lee, J.Y., Wang, S., Figueroa, A.J., Strey, R., Lobell, D.B., Naylor, R.L., Gorelick, S.M., 2022. Mapping sugarcane in Central India with smartphone crowdsourcing. Remote. Sens. 14, 703. http://dx.doi.org/10.3390/rs14030703, URL: https://www.mdpi.com/2072-4292/14/3/703.

Liu, H., Ong, Y.S., Shen, X., Cai, J., 2020. When Gaussian process meets big data: A review of scalable GPs. IEEE Trans. Neural Netw. Learn. Syst. 31, 4405–4423. http://dx.doi.org/10.1109/TNNLS.2019.2957109, URL: https://ieeexplore.ieee.org/abstract/document/8951257.

Low, K.H., Chen, J., Hoang, T.N., Xu, N., Jaillet, P., 2015. Recent advances in scaling up Gaussian process predictive models for large spatiotemporal data. In: Ravela, S., Sandu, A. (Eds.), Dynamic Data-Driven Environmental Systems Science. Springer International Publishing, Cham, pp. 167–181. http://dx.doi.org/10.1007/978-3-319-25138-7_16.

Luu, T.T.G., Whitney, C., Biber-Freudenberger, L., Luedeling, E., 2022. Decision analysis of agro-climate service scaling — A case study in Dien Bien District, Vietnam. Clim. Serv. 27, 100313. http://dx.doi.org/10.1016/j.cliser.2022.100313, URL: https://linkinghub.elsevier.com/retrieve/pii/S2405880722000310.

Martínez-Minaya, M., Conesa, D., Pennino, M.G., 2018. Species distribution modeling: A statistical review with focus in spatio-temporal issues. Stoch. Environ. Res. Risk Assess. 32, 3227–3244. http://dx.doi.org/10.1007/s00477-018-1548-7.

Matheron, G., 1963. Principles of geostatistics. Econ. Geol. 58, 1246–1266. http://dx.doi.org/10.2113/gsecongeo.58.8.1246, URL: http://pubs.geoscienceworld.org/economicgeology/article/58/8/1246/17275/Principles-of-geostatistics.

Neumann, J.V., Rubinstein, A., Morgenstern, O., Kuhn, H.W., 2007. Theory of Games and Economic Behavior: 60th Anniversary Commemorative Edition. Afterword: Ariel Rubinstein. Princeton University Press, Princeton, NJ, 16th anniverary edition ed.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: An imperative style, high-performance deep learning library. http://dx.doi.org/10.48550/arXiv.1912.01703, URL: arXiv:1912.01703.

Patil, S.S., Thorat, S.A., 2016. Early detection of grapes diseases using machine learning and IoT. In: 2016 Second International Conference on Cognitive Computing and Information Processing (CCIP). pp. 1–5. http://dx.doi.org/10.1109/CCIP.2016.7802887, URL: https://ieeexplore.ieee.org/document/7802887.

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S., 2009. Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. Ecol. Appl. 19, 181–197. http://dx.doi.org/10.1890/07-2153.1, URL: https://onlinelibrary.wiley.com/doi/abs/10.1890/07-2153.1.

Pramoedyo, H., Ashari, A., Fadliana, A., 2020. Application of GSTAR kriging model in forecasting and mapping coffee berry borer attack in Probolinggo district. J. Phys.: Conf. Ser. 1563, 012005. http://dx.doi.org/10.1088/1742-6596/1563/1/012005.

Prashanthi, S., Srikar, K., Theodore, R.K., 2022. Adoption pattern of 448 varieties of tomato among the farmers of Ananthapuramu District. J. Res. ANGRAU 50, 17–24. http://dx.doi.org/10.5555/20230107861, URL: https://www.cabidigitallibrary.org/doi/full/10.5555/20230107861.

Rajkhowa, P., Qaim, M., 2021. Personalized digital extension services and agricultural performance: Evidence from smallholder farmers in India. PLOS ONE 16, e0259319. http://dx.doi.org/10.1371/journal.pone.0259319.

Rasmussen, C.E., Williams, C.K.I., 2005. Gaussian Processes for Machine Learning. The MIT Press, http://dx.doi.org/10.7551/mitpress/3206.001.0001, URL: https://direct.mit.edu/books/book/2320/Gaussian-Processes-for-Machine-Learning.

Ribka, D., Mahendran, K., Lavanya, S.M., Senthilnathan, S., 2020. An explorative study on usage pattern of pesticides among tomato growing farmers in Anantapur district of Andhra Pradesh. Int. J. Chem. Stud. 8, 50–53. http://dx.doi.org/10.22271/chemi.2020.v8.i6a.10917.

Rossi, V., Caffi, T., Salotti, I., Fedele, G., 2023. Sharing decision-making tools for pest management may foster implementation of Integrated Pest Management. Food Secur. 15, 1459–1474. http://dx.doi.org/10.1007/s12571-023-01402-3.

Ruett, M., Dalhaus, T., Whitney, C., Luedeling, E., 2022. Assessing expected utility and profitability to support decision-making for disease control strategies in ornamental heather production. Precis. Agric. http://dx.doi.org/10.1007/s11119-022-09909-z, URL: https://link.springer.com/10.1007/s11119-022-09909-z.

Salimbeni, H., Eleftheriadis, S., Hensman, J., 2018. Natural gradients in practice: non-conjugate variational inference in Gaussian process models. http://dx.doi.org/10.48550/arXiv.1803.09151, URL: arXiv:1803.09151.

Sawosri, A.W., Mußhoff, O., 2020. Risk and time preferences of farmers in India and Indonesia. Working Paper 32. EFForTS Discussion Paper Series. https://www.econstor.eu/handle/10419/217206.

Schulz, D., Börner, J., 2023. Innovation context and technology traits explain heterogeneity across studies of agricultural technology adoption: A meta-analysis. J. Agric. Econ. 74, 570–590. http://dx.doi.org/10.1111/1477-9552.12521, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/1477-9552.12521.

Senapati, A.K., 2020. Evaluation of risk preferences and coping strategies to manage with various agricultural risks: Evidence from India. Heliyon 6, http://dx.doi.org/10.1016/j.heliyon.2020.e03503, URL: https://www.cell.com/heliyon/abstract/S2405-8440(20)30348-0.

Shiberu, T., Getu, E., 2018. Determination of the economic threshold level of tomato leaf miner, Tuta absoluta Meyrick (Lepidoptera: Gelechiidae) on tomato plant under glasshouse conditions. J. Hortic. For. 10, 9–16. http://dx.doi.org/10.5897/JHF2018.0522, URL: http://academicjournals.org/journal/JHF/article-abstract/0C6499556123.

Snepvangers, J.J.J.C., Heuvelink, G.B.M., Huisman, J.A., 2003. Soil water content interpolation using spatio-temporal kriging with external drift. Geoderma 112, 253–271. http://dx.doi.org/10.1016/S0016-7061(02)00310-5, URL: https://www.sciencedirect.com/science/article/pii/S0016706102003105.

Sridhar, V., Naik, S.O., Nitin, K.S., Asokan, R., Swathi, P., Gadad, H., 2019. Efficacy of integrated pest management tools evaluated against Tuta absoluta (Meyrick) on tomato in India. J. Biol. Control. 264–270. http://dx.doi.org/10.18311/jbc/2019/23254, URL: https://ischolar.sscldl.in/index.php/jbc/article/view/189077.

Stern, V.M., 1973. Economic thresholds. Annu. Rev. Entomol. 18, 259–280. http://dx.doi.org/10.1146/annurev.en.18.010173.001355, URL: https://www.annualreviews.org/content/journals/10.1146/annurev.en.18.010173.001355.

Tang, F.H.M., Lenzen, M., McBratney, A., Maggi, F., 2021. Risk of pesticide pollution at the global scale. Nat. Geosci. 14, 206–210. http://dx.doi.org/10.1038/s41561-021-00712-5, URL: https://www.nature.com/articles/s41561-021-00712-5.

Taylor, B., Tonnang, H.E.Z., Beale, T., Holland, W., Oronje, M., Abdel-Rahman, D., Finegold, C., Zhu, J., Pozzi, S., Murphy, S.T., 2023. Leveraging data, models & farming innovation to prevent, prepare for & manage pest incursions: Delivering a pest risk service for low-income countries. In: von Braun, J., Afsana, K., Fresco, L.O., Hassan, M.H.A. (Eds.), Science and Innovations for Food Systems Transformation. Springer International Publishing, Cham, pp. 439–453. http://dx.doi.org/10.1007/978-3-031-15703-5_23.

Titsias, M., 2009. Variational learning of inducing variables in sparse gaussian processes. In: van Dyk, D., Welling, M. (Eds.), Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics. PMLR, Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA, pp. 567–574, URL: https://proceedings.mlr.press/v5/titsias09a.html.

Tudi, M., Daniel Ruan, H., Wang, L., Lyu, J., Sadler, R., Connell, D., Chu, C., Phung, D.T., 2021. Agriculture development, pesticide application and its impact on the environment. Int. J. Environ. Res. Public Heal. 18, 1112, URL: https://www.mdpi.com/1660-4601/18/3/1112.

Wang, S., Di Tommaso, S., Faulkner, J., Friedel, T., Kennepohl, A., Strey, R., Lobell, D.B., 2020. Mapping crop types in Southeast India with smartphone crowdsourcing and deep learning. Remote. Sens. 12, 2957. http://dx.doi.org/10.3390/rs12182957, URL: https://www.mdpi.com/2072-4292/12/18/2957.

Whalon, M.E., Mota-Sanchez, D., Hollingworth, R.M., 2008. Global Pesticide Resistance in Arthropods. CABI.

Wilson, A.G., Hu, Z., Salakhutdinov, R., Xing, E.P., 2015. Deep kernel learning. http://dx.doi.org/10.48550/arXiv.1511.02222, URL: arXiv:1511.02222.

Wilson, A.G., Hu, Z., Salakhutdinov, R., Xing, E.P., 2016. Stochastic variational deep kernel learning. http://dx.doi.org/10.48550/arXiv.1611.00336, URL: arXiv:1611.00336.

Wright, R.J., Devries, T.A., Young, L.J., Jarvi, K.J., Seymour, R., 2002. Geostatistical analysis of the small-scale distribution of European corn borer (Lepidoptera: Crambidae) larvae and damage in whorl stage corn. Environ. Èntomol. 31, 160–167. http://dx.doi.org/10.1603/0046-225X-31.1.160.

Yesdhanulla, S., Aparna, B., 2018. Marketing channels and price spread of tomato in Chittoor district of Andhra Pradesh. J. Pharmacogn. Phytochem. 7, 873–876, URL: https://www.phytojournal.com/archives/2018.v7.i2.3433/marketing-channels-and-price-spread-of-tomato-in-chittoor-district-of-andhra-pradesh.

You, J., Li, X., Low, M., Lobell, D., Ermon, S., 2017. Deep Gaussian process for crop yield prediction based on remote sensing data. In: Proceedings of the AAAI Conference on Artificial Intelligence. p. 31. http://dx.doi.org/10.1609/aaai.v31i1.11172, URL: https://ojs.aaai.org/index.php/AAAI/article/view/11172.

Zbinden, R., van Tiel, B., Hughes, L., Tuia, D., 2024. On the selection and effectiveness of pseudo-absences for species distribution modeling with deep learning. http://dx.doi.org/10.48550/arXiv.2401.02989, URL: arXiv:2401.02989.