

Research Article

Multimodel Inference for the Prediction of Disease Symptoms and Yield Loss of Potato in a Two-Year Crop Rotation Experiment

Wim Van den Berg,¹ Jan Vos,² and Johan Grasman³

¹ Applied Plant Research for Arable Farming, Multifunctional Agriculture and Field Production of Vegetables, Wageningen UR, P.O. Box 430, 8200 AK Lelystad, The Netherlands

² Centre for Crop Systems Analysis, Wageningen University, P.O. Box 430, 6700 AK Wageningen, The Netherlands

³ Biometris, Department of Mathematical and Statistical Methods, Wageningen University, P.O. Box 100, 6700 AC, Wageningen, The Netherlands

Correspondence should be addressed to Wim Van den Berg, wim2.vandenberg@wur.nl

Received 23 August 2011; Accepted 16 December 2011

Academic Editor: Robert J. Kremer

Copyright © 2012 Wim Van den Berg et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The second order Akaike information criterion was used for the assessment of 139 regression models for three responses of potato test crops: (a) incidence of *Spongospora subterranea* on the harvested tubers, (b) percentage of haulms infected with *Verticillium dahliae*, and (c) tuber yield. Six variables that are likely related to the response variables were taken into consideration: soil infestations of the fungus *Verticillium dahliae* and of three nematode species (*Globodera pallida*, *Trichodoridae*, and *Meloidogyne* spp.) and, furthermore, soil pH and water soluble phosphor (P). Interactions between *V. dahliae* and the three nematode species were included as well. Based on multimodelling, predictors are given a weight from which one may decide about the need to include them in a prediction of crop yield. The most important predictors were soil infestation levels of *V. dahliae* and *G. pallida* and soil pH. The outcome also showed that tubers suffered more from *S. subterranea* for higher soil pH values. Finally, yield reduction from the presence of *V. dahliae* was enhanced by the presence of higher densities of *G. pallida*.

1. Introduction

To answer research questions in agronomy, statistical hypotheses are formulated and an analysis of variance on experimental results is performed in order to detect the effect of different treatments [1]. In data obtained from common types of experiments, effects of treatments are causal because of randomization of the treatment levels over the experimental units. In crop rotation experiments, a wider approach is needed in the sense that not only controlled treatments are considered, but also other predictors such as weather and soil characteristics. Ultimately, such procedures should lead to a statistical model giving the best prediction of crop yield. In crop rotation experiments on soil borne pathogens, different pathogen species are present and various regression models can be formulated about the harm caused by these species in combination with the effect of abiotic predictors for crop growth. The aim of this study is to select

from a set of plausible predictors for tuber yield the ones that definitely should be included in a prediction model. We will use the multimodel approach in which a model is constructed as a weighted combination of all models. The advantage of this approach is that one uses information from all models instead of choosing the best model. This can be attractive when a number of models are doing about equally well. To calculate the weights, the Akaike Information Criterion is used. In the generalized linear models used in this study, the multimodel method also leads to improved estimates of parameters based on the weights of the different models. The study used data of an experiment by Scholte [2] on the application of trap crops to reduce the damage from potato cyst nematodes.

Farmers grow potatoes as a component in a crop rotation system. Short intervals between successive potato crops may result in a buildup of harmful soil-borne pathogens, including several fungi and nematode species [3]. Above a

	1994	1995	1996	1997	1998
Set 1	Disco	Cycle 1 Year 1	Cycle 1 Year 2	Cycle 2 Year 1	Cycle 2 Year 2
Set 2	Red clover	Disco	Cycle 1 Year 1	Cycle 1 Year 2	Cycle 2 Year 1

Year 1: Trap crop and fallow in spring
Tagetes or oats or fallow in summer and autumn

Year 2: Main potato cultivars Seresta or Karida or Elkana
potato haulms removed or incorporated into the soil

FIGURE 1: The two-year crop rotation scheme, representing the sequence of crops and treatments for one of three blocks. Each block was split into two “sets”: the experimental cycles in Set 1 started with the potato cultivar Disco in 1994. The experimental cycles of Set 2 started with the same potato cultivar in 1995, and red clover was grown in 1994. A cycle consisted of a sequence of crops and treatments lasting two years. Year one of a cycle consisted of two parts: spring treatments and treatments applied in summer plus autumn. Spring treatments were either fallow or the cultivation of a trap crop against potato cyst nematode. Summer/autumn treatments were one of three options: cultivation of *Tagetes*, oats, or fallow. In year 2 of the cycle, potatoes were grown on all plots; treatments consisting of three potato cultivars showed differing degree of resistance to potato cyst nematodes (cvs Seresta, Karida, or Elkana); potato cultivar treatments were split up in two sub treatments consisting of incorporation or removal from the field of potato haulms. See Scholte [2] for more details.

threshold level of infestation, tuber yield will suffer from the presence of potato cyst nematodes [4, 5].

If more than one soil-borne pathogen is present, interaction effects are expected resulting in a further increase in crop damage [6]. Apart from Scholte [2], there is a wealth of other studies reporting on the relation between inoculum density of interacting diseases and crop response. Evans [7], for instance, showed in a pot experiment with several potato cultivars higher infection rates of the fungus *Verticillium dahliae* in the presence of cyst nematode species (*Globodera rostochiensis* or *G. pallida*). Potato early dying [8] is caused by *V. dahliae* and *V. albo-atrum*, but severity aggravates when soil is also infested by *Pratylenchus penetrans*.

Many results have been published on the harming effects of soil-borne pathogens in combination with environmental factors such as temperature, soil pH, and soil texture. The index of powdery scab (*Spongospora subterranea*) tended to take lower levels when the pH of the soil increased by adding lime [9]. Yield loss of potato by *G. pallida* is more severe at high than at low pH values of the soil. This is attributed to a lower availability of phosphate at higher pH values [10]. On sandy soils, damage from *Meloidogyne* spp., *Pratylenchus* spp., and *Trichodoridae* is greater than on clay soils. Temperature affects growth and development of nematodes as well as the agricultural crops themselves. Moreover, the number of generations of *Meloidogyne* may be one or two in cool summers, while at higher temperature three generations are possible [11].

In the past, mathematical models have been introduced to quantify the damage from pathogens and their interactions using multiple linear regression [12, 13]. Wheeler et al. [14] modeled the simultaneous effect of *P. penetrans* and *V. dahliae* on tuber yield of potato with nonlinear regression, while Jacobsen et al. [15] modelled tuber yield as a function of the density of both *M. hapla* (x_1) and *V. albo-atrum* (x_2)

as predictors on the natural log scale, resulting in an additive model plus an interaction term:

$$y = \alpha + \beta \log(x_1) + \gamma \log(x_2) + \delta \log(x_1) \log(x_2), \quad (1)$$

where y is the yield of potato tubers and α , β , γ , and δ are regression coefficients. Here and elsewhere “log” means the natural logarithm.

This paper presents a new statistical analysis of a case study dealing with four factorial treatments applied in a crop rotation experiment (1995–1998) with a rotation cycle of two years [2]. Results of such an experiment can be used to formulate models for the prediction of pathogen dynamics and yield loss based on measured values of predictor variables. Parameters in expressions of this type are estimated by means of Generalized Linear Models (GLMs) using a logarithmic link function. In this study we took an information-theoretical approach for model selection and multimodel inference [16]. First, we select a biologically plausible subset of predictors from the available set including their interactions. The second order Akaike information criterion [17] is used to select the best models. The approach is applied to the prediction of three responses: (a) the incidence of *S. subterranea* at tuber skin, (b) percentage of haulm infection by *V. dahliae*, and (c) tuber yield.

2. Materials and Methods

2.1. Experimental Design. The data used in this study are from a crop rotation study with cycle of two years (Figure 1) [2]. The experiment consisted of two so-called “sets”: preparations for Set 1 started in 1994 and for Set 2 in 1995. In both sets the same treatments were applied in the same sequence. The advantage of having two sets is that observations on main potato crops could be done each year.

Set 1 consisted of two two-year cycles of the crop rotation and Set 2 of 1.5 cycles of the crop rotation. The first year of the crop rotation cycle is divided into the spring period and the period of summer taken together with autumn. In the spring period, two trap cropping treatments were applied: (i) cropping with potato cultivar Kartel acting as a trap crop for potato cyst nematode (PCN), it resulted in a decline of the infestation (see the Section 2.2) and (ii) a fallow control. In summer and autumn (after destroying the potato trap crop), three green manure cropping treatments were applied: (i) growing of *Tagetes*, (ii) growing of oats (*Avena sativa*), and (iii) fallow control. In year two of the crop rotation cycle, potatoes were grown; the choice of cultivar is also a treatment (Seresta, Karida, and Elkana). In addition to the cultivar factor, another treatment factor was applied, immediately after the harvest potato haulms were either incorporated into the soil or removed from the plot.

The two trap cropping treatments, the three green manure cropping treatments and the two haulm removal treatments, were combined in a factorial design. The 12 treatment combinations of trap crop \times green manure crop \times haulm removal were randomized over 12 main plots. The potato cultivars split the main plots into three plots meaning that one replicate of one set consisted of 36 plots. Thus, the design consisted of three blocks each being split into the two sets as mentioned before. Set 1 of the experiment was terminated in November 1998 after growing potato in that year. Set 2 was terminated in March 1999 after sampling the plots for determination of soil infestation with pathogens. The experiment was located in Achterberg (51°59'N, 5°35'E), Netherlands, on a sandy soil.

2.2. Justification of Treatments. The treatments were designed to differentially affect the infestation level of the soil-borne pests and diseases. In the first year, potato itself was used as a trap crop against PCN; the other half of the plots remained fallow. Potato can be used as a trap crop for PCN, provided the crop is killed by glyphosate by the end of June, that is, before PCNs complete their life cycle and multiply [2]. The green manure *Tagetes* was used to reduce the density of *P. penetrans*, oats was expected to enhance antagonists of *Rhizoctonia solani* [18, 19], while fallow was used to reduce the density of all soil-borne pathogens. In the second year, the three potato cultivars were grown: Seresta, Karida, and Elkana. They are, respectively, highly resistant, moderately resistant, and susceptible to *G. pallida*, and all three are highly resistant to the pathotypes of *G. rostochiensis*. Potato haulms are the most important source of *V. dahliae* microsclerotia. Hence, the soil inoculum level of *V. dahliae* can be influenced by incorporating or removing haulms [20].

2.3. Inoculation. In 1994 and 1995, inoculations were made to uniformly infest the plots with the organisms under study. *R. solani*, anastomosis group 3 (AG3), and *Pratylenchus crenatus* were already present in the soil at high densities as a result of previously grown crops. Inoculations were made with *G. pallida*, *M. hapla*, *M. chitwoodi* and *P. penetrans*, and *V. dahliae*; details are in [2].

2.4. Soil Sampling. Before planting main crop potatoes, 60 soil cores were taken at random from each plot with a 2.0 cm diameter auger to a depth of 20 cm early in March. Each sample was gently but thoroughly mixed and then separated into four parts: 300 mL soil was used for a biotest for the presence of root-knot nematodes, 2000 mL to analyze the populations of other nematodes, 250 mL to determine soil infestation with *V. dahliae*, and the remaining soil was used to measure pH and analyze N, P, and K levels. In spring soil pH was measured as well as the soil concentrations of nitrate, ammonium, and water soluble P and K on the plots to be planted with the potato main crop.

2.5. Data Acquisition of Nematode Densities. Cysts of nematodes were extracted from a subsample of moist soil using the method described by Oostenbrink [21]. After crushing the cysts, numbers of viable juveniles were counted using a stereo dissecting microscope. In 1999, Set 2 was treated differently: air-dried soil samples were used from which cysts were separated using the method described by Van Bezooijen et al. [22]. Populations of *Pratylenchus* spp. and *Meloidogyne* spp. were counted after extracting them using the method described by Oostenbrink [21].

2.6. Haulm Infections with *V. Dahliae* and *R. Solani*. Each year, in the second week of August, 20 randomly chosen plants of the potato test crop were harvested, five from each of the inner four rows of each plot. A transverse section of 1–2 mm thickness was cut from the middle of each haulm piece, and eight sections per dish were placed on a selective medium in Petri dishes and incubated for two weeks [20]. Colonies of *V. dahliae* were counted using a stereo dissecting microscope. *Rhizoctonia* haulm canker was recorded at the underground part of each haulm using five classes of disease severity.

2.7. Inoculum Density of *V. Dahliae* in Soil. Microsclerotia are the surviving structures of *V. dahliae* in soil quantified by a colony forming unit (cfu). To that end, soil samples were dried and sieved and were dispersed on a selective agar plate. The plates were incubated and after three weeks the number of cfu of *V. dahliae* was determined using a stereo dissecting microscope; the number of cfu was expressed per g soil, details are in [2].

2.8. Incidence of *S. Subterranea* on Tubers. For each plot the amount of *S. subterranea* was assessed using five severity classes: 0, 0–25%, 25–50%, 50–75%, or 75–100% of tuber skin covered with powdery scab.

2.9. Specification of Plausible Predictors. In the case study, a large number of variables were measured for monitoring soil fertility, density of pathogens in soil, and growth of the potato crop. To save time and effort in the selection process and obtain parsimony using multimodel inference, we only retained predictors in the model that are meaningful as possible predictors of tuber yield: some disease symptoms were measured in the field during crop growth or afterwards on harvested tubers. These quantities have no predictive

value and were discarded from the analysis. Potatoes are host to the fungus *V. dahliae* and to the nematodes *G. pallida*, *Trichodoridae*, and *Meloidogyne* spp. Densities of these pathogens was measured in the soil before planting. *Trichodoridae* stands for all nematodes belonging to the genera *Trichodorus* and *Paratrichodorus*. Also soil fertility measurements were made before planting the crop. Only pH and water soluble P were assumed to be relevant [10] and retained, while total N and water soluble K were discarded from the analysis as these nutrients are too mobile and variable in the soil to have predictive value.

Following the literature [7, 11], the interaction between *V. dahliae*, and the three nematode species were added to the predictor set. We did not include the interaction between pH and nematode density in spring as Mulder [8] did, because there the range of pH values was much larger than in our case study. Three of the nine predictors are interactions. From all 2^9 models, we only considered those in which both predictors were included when their interaction was in the model [23]. Therefore, for each of the three responses (i.e., incidence of *S. subterranea* on harvested tubers, percentage of haulms infected with *V. dahliae* and dry tuber yield) only 139 predictive models were selected to which multimodel inference was applied. The model with all nine predictors is called the global model.

2.10. Regression Models. A simple nonlinear model for yield loss caused by nematodes of one species is formulated as

$$\mu = \mu_0 \rho^x, \quad (2)$$

where μ is the expected tuber yield and x the initial density of the pathogen at the moment potatoes are planted. Parameter μ_0 is the maximum yield when $x = 0$, and the parameter ρ represents the intact fraction of the roots after attack due to one unit of pathogen [24]. (2) has also been used by Mol et al. [25] for the assessment of yield loss by *V. dahliae*, where x was the initial density of cfu g^{-1} soil. Here, we assume that parameter ρ is identical for all three cultivars we consider. This means that the relative damage per unit of pathogen is assumed to be the same for all cultivars. However, we use the parameter μ_{0c} to express the differences in yield of the three cultivars ($c = 1, 2, 3$). In this way we capture the effect of more than one soil-borne pathogen species in a multiplicative yield loss model with a restricted number of parameters

$$\mu = \mu_{0c} \rho_1^{x_1} \rho_2^{x_2} \cdot \rho_n^{x_n}, \quad (3)$$

where x_1, x_2, \dots, x_n are the initial densities of the n soil-borne pathogens [26]. It is noted that the parameter ρ varies with the pathogen type. Yield decrease due to two pathogens, for example, *V. dahliae*, x_1 , and *G. pallida*, x_2 , is modeled as

$$\mu = \mu_{0c} \rho_1^{x_1} \rho_2^{x_2} \rho_{12}^{x_1 x_2}. \quad (4)$$

The parameters of (4) were estimated with a Generalized Linear Model (GLM) with a logarithmic link function and with the response having a normal distribution [27]. The product of x_1 and x_2 can be written as x_3 , so that (4)

transforms into

$$g(\mu) = \log(\mu) = \beta_{0c} + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3, \quad (5)$$

where $\beta_{0c} = \log(\mu_{0c})$, $\beta_i = \log(\rho_i)$, $i = 1, 2$, and $\beta_3 = \log(\rho_{12})$. For an analysis based on a reciprocal link function, we refer to [28, 29].

The environmental predictors, pH and water soluble P content of the soil, can be added in the same way to (5). We have p predictors for which the regression coefficients are identical for the three cultivars. Furthermore, the intercept is assumed to depend on the choice of the cultivar, so that there are $p + 3$ regression parameters. Including the error term, the GLM with a logarithmic link function takes the form

$$g(\mu) = \log(\mu) = \beta_{0c} + \sum_{j=1}^p \beta_j x_j. \quad (6)$$

Besides the tuber yield, we have the two other response variables. Firstly, the percentage of haulms infested with *V. dahliae* for which a logit link function is used with the infested number as response and the total number of haulms as binomial total. Secondly, the distribution of tubers over the five severity classes of coverage with *S. subterranea* is analyzed with ordinal regression [27].

2.11. Second Order Akaike Information Criterion (AIC^c). In our case study, nine predictors are available for prediction of the tuber yield μ . We assume the tuber yield to be normally distributed, so we have to maximize the normal likelihood, L , as function of the regression parameter vector, β , and the standard deviation, σ , of the error term given the data, y , [23]:

$$L(\beta, \sigma | y) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp\left(-\frac{\sum_{i=1}^n (y_i - \mu_i(\beta))^2}{2\sigma^2} \right)$$

$$\text{with } \log(\mu_i(\beta)) = \beta_{0c} + \sum_{j=1}^p \beta_j x_{ij}, \quad (7)$$

where i is the observation index. Replacing σ^2 by its estimate (the residual sum of squares divided by n), we are left with a likelihood function of β only that has to be maximized. For model selection and multimodel inference, we use the second order Akaike Information Criterion (AIC^c):

$$\text{AIC}^c = -2 \log L(\beta | y) + 2K + \frac{2K(K+1)}{(n-K-1)}, \quad (8)$$

where n is the number of observations and $K = p + 4$ the number of parameters (p regression coefficients, 3 intercepts for the cultivars and σ^2 for the error term). Burnham and Anderson [16] demonstrated that AIC^c should be preferred

above the standard Akaike Information Criterion if $n/K < 40$. In our case study, n/K is indeed below this value for most of the candidate models and so we use (8). In the following, we take for the parameters in the model those values that minimize the AIC^c .

2.12. Kullback-Leibler Distance. Let R candidate models be selected in advance. For each model $r = 1, \dots, R$ the AIC_r^c with the best choice of parameters is calculated. Following Burnham and Anderson [16], the model with the lowest AIC^c is denoted by $AIC_{r^*}^c$. The Kullback-Leibler distance Δ_r between AIC_r^c and $AIC_{r^*}^c$ is used to compute an appropriate weight w_r for each of the $r = 1, \dots, R$ models:

$$w_r = \frac{e^{-\Delta_r/2}}{\sum_{r=1}^R e^{-\Delta_r/2}} \quad \text{with } \Delta_r = AIC_r^c - AIC_{r^*}^c. \quad (9)$$

Using w_r , we can order the models: $r_1 = r^*$ and so on. The summed weight of the first k models

$$P(k) = \sum_{j=1}^k w_{r_j} \quad (10)$$

denotes the probability that from all R models the best approximation to truth is among these k models.

2.13. Multimodel Inference. Now, based on the weights w_r , we compose the (multi)model having estimated regression coefficients

$$\hat{\beta}_j = \sum_{r=1}^R w_r \hat{\beta}_{rj}, \quad j = 1, \dots, p, \quad (11)$$

meaning that in a model without predictor j the coefficient $\hat{\beta}_{rj}$ is given the value zero. Furthermore, each predictor can be given a weight

$$v_j = \sum_{r=1}^R w_r I_j(r), \quad j = 1, \dots, p, \quad (12)$$

where $I_j(r) = 1$ if predictor j is in model r , otherwise $I_j(r) = 0$. The weight v_j takes values in the interval $[0, 1]$ and is an indication for the importance of this predictor in the process of constructing the best regression model. Following Lukacs et al. [30], the variance of the regression coefficient j takes the form

$$\begin{aligned} & \text{var}(\hat{\beta}_j) \\ &= \sum_{r=1}^R w_r \left\{ \text{var}(\hat{\beta}_{rj}) + (\hat{\beta}_{rj} - \hat{\beta}_j)^2 \right\}, \quad j = 1, \dots, p. \end{aligned} \quad (13)$$

2.14. Statistical Package. All statistical analyses are carried out using GenStat [31].

3. Results

3.1. Statistics. The index (0–100) for *S. subterranea* varied from 0 to 78.4 with a median equal to 12.7. Percentage of haulms infected with *V. dahliae* varied from 0 to 86.4 with median value 12.8. Tuber dry yield ranged from 910 to 1729 g m⁻². Densities of *V. dahliae*, *G. pallida*, *Trichodoridae*, and *Meloidogyne* spp. showed low medians because only on a fraction of the plots there were higher levels of infestation. Water soluble P varied from 5.5 to 34.0 kg/ha in the tillage layer. A similar range of values was found for these predictors in 1996 and 1997.

In multiple regression analysis, high correlations between two or more predictors, also called multicollinearity [32], may complicate the analysis and the interpretation of the outcomes. In the first place multicollinearity may increase the standard errors of the parameters in a model (“inflation”). Secondly, multicollinearity increases model uncertainty by inducing different estimates across different models, which enlarges the second term in (13). Table 1 gives the correlation matrix for the nine predictors (in 1998). As expected a positive correlation between interaction terms and their corresponding two predictors was found. All correlation coefficients were below 0.7, so multicollinearity appeared not to be a problem.

The goodness of fit of the global model, with all candidate predictors present in the model, (6), was quantified by the coefficient of multiple determination R^2 (Table 2). In the analyses over the three years R^2 ranged from 0.38 for *S. subterranea* in 1997 to 0.66 for the percentage of haulms infected with *V. dahliae* in 1998. The R^2 for the model with only the cultivar term varied from 0.05 for *S. subterranea* in 1996 to 0.37 for tuber yield in 1997. The part of R^2 representing the variance explained by the nine predictors ranged from 0.15 (= 0.52–0.37) for dry tuber yield in 1997 to 0.43 (= 0.55–0.12) for the percentage of haulms infected with *V. dahliae* in 1996. The coefficients of determination in Table 2 are rather low. This may be caused by predictors not taken into consideration and/or measurement errors. Wider ranges in the predictors would have resulted in a wider range of the response, increasing R^2 . The question whether all nine predictors substantially contributed to the size of R^2 was answered in the multimodel approach.

3.2. Akaike Information Criterion. For each year and for each of the three response variables, the 139 candidate models were fitted to the data. The regression coefficients, with corresponding standard errors, were saved for each model and the AIC^c was calculated using (8). For each response variable, the models were sorted according to ascending AIC^c .

3.3. Multimodel Inference. Multimodel inference results are presented from analyses for the three response variables. The summed weights, v_j , for predictor j , $j = 1, \dots, 9$, are given by (12) and lie between 0 and 1; Tables 3, 4, and 5 show their values in the analysis of the three responses. A value close to 1 for a predictor indicates that it has a high predictive value.

TABLE 1: Correlation matrix for the nine predictors in the year 1998. There are 108 pairs in each correlation. Units: *V. dahliae* in cfu g⁻¹ soil; *G. pallida*, *Trichodoridae*, and *Meloidogyne* spp. in mL⁻¹ and Water soluble P in kg ha⁻¹.

Nr	Predictor	1	2	3	4	5	6	7	8	9
1	<i>V. dahliae</i> g ⁻¹	1.00								
2	<i>G. pallida</i> mL ⁻¹	0.00	1.00							
3	<i>Trichodoridae</i> mL ⁻¹	-0.02	-0.09	1.00						
4	<i>Meloidogyne</i> spp. mL ⁻¹	-0.20	-0.17	-0.11	1.00					
5	<i>V. dahliae</i> × <i>G. pallida</i>	0.40	0.60	-0.09	-0.13	1.00				
6	<i>V. dahliae</i> × <i>Trichodoridae</i>	0.51	-0.09	0.67	-0.12	0.08	1.00			
7	<i>V. dahliae</i> × <i>Meloidogyne</i> spp.	-0.01	-0.17	-0.05	0.69	-0.13	0.00	1.00		
8	pH	0.39	-0.08	0.21	-0.12	0.15	0.43	-0.06	1.00	
9	Water soluble P	-0.19	0.01	-0.06	0.06	-0.07	-0.13	0.15	-0.33	1.00

TABLE 2: The coefficient of multiple determination, R^2 , of the global model for the three responses over the three years. In the global model, the nine predictors are present as well as the cultivar term. Between brackets, R^2 is given for the model with only the cultivar term.

Year	Index <i>S. subterranea</i> (0–100)	Haulms infected with <i>V. dahliae</i> (%)	Dry tuber yield (g m ⁻²)
1996	0.44 (0.05)		0.44 (0.16)
1997	0.38 (0.16)		0.52 (0.37)
1998	0.50 (0.09)		0.46 (0.22)

TABLE 3: Summed weight, $v_j = \sum wr I_j(r)$, weighted regression coefficients, $\hat{\beta}_j$, and standard errors of $\hat{\beta}_j$, calculated with (11)–(13) for *S. subterranea* in the years 1996, 1997, and 1998. Units: *V. dahliae* in cfu g⁻¹ soil; *G. pallida*, *Trichodoridae*, and *Meloidogyne* spp. in mL⁻¹ and Water soluble P in kg ha⁻¹.

j	Year Predictor	1996			1997			1998		
		v_j	$\hat{\beta}_j$	se $\hat{\beta}_j$	v_j	$\hat{\beta}_j$	se $\hat{\beta}_j$	v_j	$\hat{\beta}_j$	se $\hat{\beta}_j$
1	<i>V. dahliae</i>	1.00	0.46	0.35	1.00	1.13	0.65	1.00	0.90	0.45
2	<i>G. pallida</i>	0.31	0.02	0.21	0.29	0.00	0.12	0.99	-0.21	0.26
3	<i>Trichodoridae</i>	0.54	0.36	0.62	0.94	0.23	0.50	1.00	1.65	0.62
4	<i>Meloidogyne</i> spp.	0.84	3.42	1.57	0.50	-0.18	0.40	0.99	-0.68	0.48
5	<i>V. d</i> × <i>G.pal</i>	0.13	-0.01	0.11	0.08	-0.02	0.15	0.93	0.16	0.24
6	<i>V. d</i> × <i>Trichod.</i>	0.41	-0.35	0.51	0.44	0.48	0.63	1.00	-5.33	0.90
7	<i>V. d</i> × <i>Meloid.</i> spp.	0.80	-13.51	2.33	0.17	0.23	0.53	0.20	-0.01	0.42
8	pH	1.00	2.52	0.56	1.00	1.10	0.48	1.00	3.80	0.57
9	Water soluble P	1.00	-0.04	0.10	1.00	0.08	0.10	0.99	-0.03	0.09

TABLE 4: Summed weights, $v_j = \sum wr I_j(r)$, weighted regression coefficients, $\hat{\beta}_j$, and standard errors of $\hat{\beta}_j$, calculated with (11)–(13) for percentage haulms infected with *V. dahliae* in the years 1996, 1997, and 1998.

j	Year Predictor	1996			1997			1998		
		v_j	$\hat{\beta}_j$	se $\hat{\beta}_j$	v_j	$\hat{\beta}_j$	se $\hat{\beta}_j$	v_j	$\hat{\beta}_j$	se $\hat{\beta}_j$
1	<i>V. dahliae</i>	1.00	0.35	0.30	0.99	1.19	0.73	1.00	1.27	0.42
2	<i>G. pallida</i>	0.41	-0.02	0.24	0.83	0.00	0.28	0.41	0.02	0.14
3	<i>Trichodoridae</i>	0.30	-0.07	0.40	0.45	0.09	0.38	0.36	0.03	0.34
4	<i>Meloidogyne</i> spp.	0.84	-5.54	1.84	0.40	-0.08	0.38	0.66	-0.38	0.50
5	<i>V. d</i> × <i>G.pal</i>	0.11	-0.01	0.10	0.53	0.31	0.47	0.14	-0.02	0.13
6	<i>V. d</i> × <i>Trichod.</i>	0.08	-0.01	0.20	0.21	-0.26	0.53	0.09	-0.02	0.26
7	<i>V. d</i> × <i>Meloid.</i> spp.	0.23	1.12	1.37	0.21	0.52	0.72	0.23	0.26	0.57
8	pH	0.98	1.11	0.59	1.00	1.03	0.54	0.27	0.06	0.33
9	Water soluble P	1.00	-0.03	0.11	0.24	0.00	0.07	0.47	-0.02	0.09

TABLE 5: Summed weights, $v_j = \sum w_r I_j(r)$, weighted regression coefficients, $\hat{\beta}_j$, and standard errors of $\hat{\beta}_j$, calculated with (11)–(13) for tuber dry yield in the years 1996, 1997, and 1998.

j	Year Predictor	1996			1997			1998		
		v_j	$\hat{\beta}_j$	se $\hat{\beta}_j$	v_j	$\hat{\beta}_j$	se $\hat{\beta}_j$	v_j	$\hat{\beta}_j$	se $\hat{\beta}_j$
1	<i>V. dahliae</i>	1.00	-0.044	0.146	0.91	-0.105	0.233	1.00	-0.071	0.150
2	<i>G. pallida</i>	0.29	0.001	0.069	0.47	0.004	0.060	0.97	0.001	0.070
3	<i>Trichodoridae</i>	1.00	0.075	0.306	0.99	0.005	0.172	0.36	0.008	0.120
4	<i>Meloidogyne</i> spp.	0.30	-0.041	0.329	0.73	-0.038	0.171	0.30	0.003	0.079
5	<i>V. d</i> × <i>G. pal</i>	0.07	0.000	0.029	0.10	-0.001	0.061	0.94	-0.023	0.091
6	<i>V. d</i> × <i>Trichod.</i>	0.65	0.077	0.221	0.21	-0.000	0.160	0.08	0.003	0.101
7	<i>V. d</i> × <i>Meloid.</i> spp.	0.07	0.023	0.263	0.21	-0.037	0.224	0.07	0.000	0.076
8	pH	0.23	0.000	0.108	0.91	-0.093	0.191	0.76	0.057	0.173
9	Water soluble P	1.00	0.005	0.035	0.37	0.001	0.028	0.28	0.000	0.017

The multimodel inference for each of the response variables showed the following.

(a) For *S. subterranean*, the summed weights v_1 , v_8 , and v_9 of the predictors *V. dahliae*, pH, and water soluble P have the value 1 in each of the three years. The severity of *S. subterranea* on the tubers increased with the density of *V. dahliae* and with the pH. It only increased with the water soluble P concentration in 1997 (Set 2) and decreased for higher concentration in the other years. The summed weights, v_2 and v_4 for *G. pallida* and *Meloidogyne* spp. were 0.99 for both of them in 1998. Also in 1998, the summed weight v_5 and v_6 of the interactions between *V. dahliae* and *G. pallida* and between *V. dahliae* and *Trichodoridae* were high: they were 0.93 and 1.00, respectively.

(b) As expected the number of cfu of *V. dahliae* g⁻¹ soil measured in soil in spring was the most important predictor for the percentage of haulm infections. In 1998 *V. dahliae* was the only predictor with a high summed weight. Water soluble P and pH in 1996, and pH in 1997, also showed high summed weights.

(c) For tuber yield loss, *V. dahliae* (cfu g⁻¹ soil) in spring was the most important predictor. The summed weight v_1 took the values 1.00, 0.91, and 1.00 in the three consecutive years. The summed weights v_3 and v_9 , for *Trichodoridae*, and water soluble P were high in 1996; for *Trichodoridae* and pH, the summed weight v_3 and v_8 were high in 1997. In 1998 the summed weights v_2 and v_5 for *G. pallida* and for the interaction between *V. dahliae* and *G. pallida* were high (0.94).

4. Discussion and Conclusions

We dealt with the estimation of crop damage from soil-borne pathogens in potato using predictors that are at hand at the start of the growing season. Predictors included the initial densities of nematode species for which potato is a host and soil properties, such as pH value. In this way we ended up with only predictors that are biologically relevant and were only weakly correlated (Table 1). This resulted in six candidate predictors and three interactions. Instead of considering all possible combinations of these nine

predictors, we discarded the models with interactions for which at least one of the predictors forming the interaction was absent [23]. This reduced the number of models from 512 to only 139. For each model r , a weight, w_r , is calculated using the Akaike information criterion corrected for small datasets. The weight w_r denotes the probability that model r provides the closest approximation to the true model out of all the models considered [33]. Averaging of parameter estimates based on multimodelling is carried out with the use of these model weights. This approach utilizes the information from all models and requires less computational effort than, for example, a Bayesian alternative [34].

Differences in the levels of pathogens were created in four orthogonal treatments. For example, removal of potato haulms from the field in autumn on half of the plots considerably reduced the cfu g⁻¹ soil of *V. dahliae* compared to plots where haulms were incorporated into the soil [2]. Density of *G. pallida* was higher on fields where no trap crop was grown in the second year of each cycle. So the design of the experiment has been successfully given that the experimenter aimed at analyzing separately the effects of several soil-borne pathogens in rotations by varying (via factorial treatments) the soil infestation levels of the target organisms. If continued after 1998 for further years, the experiment would probably have given more insight into possible interactions between the pathogens that were present. The summed weights, as given in Table 3, 4, and 5 for the years 1996, 1997, and 1998, already showed this trend in crop damage from these predictors. In 1998, for the first time the summed weight for the interaction of *V. dahliae* and *G. pallida* was high. If the time series covered more years, the modelling results would probably show a rise of the summed weights for *Trichodoridae*, *Meloidogyne* spp. and their interaction with *V. dahliae* in the years after 1998.

By using data from only one experimental field, results are strongly influenced by the local soil characteristics. Differences in the pH values as well as in the levels of infestation with *S. subterranea* were already present at the start of the experiment. As the incidence of *S. subterranea* is positively correlated with the value of soil pH, the statistical results should be interpreted with care because other factors

in the soil may be correlated to the pH, such as antagonists of *S. subterranea*. Furthermore, inoculation of *P. penetrans* was not successful, may be because *P. crenatus* was already present in the soil. The presence of antagonists of soil-borne pathogens varies between locations and has a strong influence on observed treatment effects [35].

Our statistical analysis on factors that affected the response variables may provide a reliable estimation under conditions comparable with those of the experiment and leads to the following conclusions.

(a) For the incidence of *S. subterranea*, the predictors *V. dahliae*, pH, and water soluble P had the largest effect on the distribution of tubers over the five severity classes (Table 3). No hierarchy within this set of three predictors could be discerned.

(b) The percentage of haulms infected with *V. dahliae* is mainly determined by the number of cfu/g soil measured in spring (Table 4).

(c) *V. dahliae* is the most important predictor of potato yield loss. According to the regression coefficients for *V. dahliae* (Table 5), yield loss was most severe in 1997 when the measured density of cfu was low [2].

The approach we presented is worth considering in other applications, such as the design of new crop experiments and the planning of potato crops based on expert systems. With multimodelling, predictors are given a weight which can be employed in a farmers' decision support system. Identification of influential predictors plays a role in various other disciplines, such as in conservation biology where the persistence of a species depends on the interaction with other species and on the habitat with its different abiotic components. Since no specific model choice has to be made, multimodelling guarantees that no information gets lost in an early stage of the analysis.

Acknowledgment

The authors thank Dr. Ing. K. Scholte for providing the data for our case study.

References

- [1] R. A. Fisher, *The Design of Experiments*, Oliver and Boyd, Edinburgh, UK, 1949.
- [2] K. Scholte, "Effect of potato used as a trap crop on potato cyst nematodes and other soil pathogens and on the growth of a subsequent main potato crop," *Annals of Applied Biology*, vol. 136, no. 3, pp. 229–238, 2000.
- [3] J. Vos and C. D. Van Loon, "Effects of cropping frequency on of potato production," in *Effects of Crop Rotation on Potato Production in the Temperate Zones*, J. Vos, C. D. van Loon, and G. J. Bollen, Eds., vol. 40 of *Developments in Plant and Soil Science*, pp. 1–23, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1989.
- [4] F. G. W. Jones and R. A. Kempton, "Population dynamics, population models and integrated control in plant nematology," in *MAFF/ADAS GDI*, J. F. Southey, Ed., Ministry of Agriculture, Fisheries and Food, 1978.
- [5] J. Schans, *Population dynamics of potato cyst nematodes and associated damage to potato*, Ph.D. thesis, Wageningen Agricultural University, Wageningen, The Netherlands, 1993.
- [6] M. A. Back, P. P. J. Haydock, and P. Jenkinson, "Disease complexes involving plant parasitic nematodes and soilborne pathogens," *Plant Pathology*, vol. 51, no. 6, pp. 683–697, 2002.
- [7] K. Evans, "The interactions of potato cyst nematodes and *Verticillium dahliae* on early and maincrop potato cultivars," *Annals of Applied Biology*, vol. 110, no. 2, pp. 329–339, 1987.
- [8] A. Mulder, *Tolerance of the potato to stress associated with potato cyst nematodes, drought and pH an ecophysiological approach*, Ph.D. thesis, Wageningen University, 1994.
- [9] J. G. Harrison, R. J. Searle, and N. A. Williams, "Powdery scab disease of potato—a review," *Plant Pathology*, vol. 46, no. 1, pp. 1–25, 1997.
- [10] F. J. De Ruijter and A. J. Haverkort, "Effects of potato-cyst nematodes (*Globodera pallida*) and soil pH on root growth, nutrient uptake and crop growth of potato," *European Journal of Plant Pathology*, vol. 105, no. 1, pp. 61–76, 1999.
- [11] V. H. Dropkin, *Introduction to Plant Nematology*, John Wiley & Sons, New York, NY, USA, 1988.
- [12] A. J. Dobson, *An Introduction to Statistical Modeling*, Chapman and Hall, London, UK, 1988.
- [13] K. Scholte and J. J. s'Jacob, "Effect of crop rotation, cultivar and nematicide on growth and yield of potato (*Solanum tuberosum* L.) in short rotations on a marine clay soil," *Potato Research*, vol. 33, no. 2, pp. 191–200, 1990.
- [14] T. A. Wheeler, L. V. Madden, R. C. Rowe, and R. M. Riedel, "Modeling of yield loss in potato early dying caused by *Pratylenchus penetrans* and *Verticillium dahliae*," *Journal of Nematology*, vol. 24, no. 1, pp. 99–102, 1992.
- [15] B. J. Jacobsen, D. H. MacDonalds, and H. L. Bisonnette, "Interaction between *Meloidogyne hapla* and *Verticillium albo-atrum* in the *Verticillium Wilt* Disease of Potato," *Phytopathology*, vol. 69, no. 3, pp. 288–292, 1979.
- [16] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference. A Practical Information-Theoretic Approach*, Springer, New York, NY, USA, 2nd edition, 2002.
- [17] D. R. Anderson, *Model Based Inference. A Primer on Evidence*, Springer, 2008.
- [18] K. Scholte, L. Mol, and M. Lootsma, "Control of *Verticillium dahliae* and *Rhizoctonia solani* by cultural practice," in *Proceedings of the Posters and Demonstrations, 13th Triennial Conference of the European Association for Potato Research*, pp. 134–135, Veldhoven, The Netherlands, July 1996.
- [19] P. C. Struik and F. Bonciarelli, "Resource use at the cropping system level," *European Journal of Agronomy*, vol. 7, no. 1–3, pp. 133–143, 1997.
- [20] L. Mol, K. Scholte, and J. Vos, "Effects of crop rotation and removal of crop debris on the soil population of two isolates of *Verticillium dahliae*," *Plant Pathology*, vol. 44, no. 6, pp. 1070–1074, 1995.
- [21] M. Oostenbrink, "Estimating nematode populations by some selected methods," in *Nematology*, J. N. Sasser and W. R. Jenkins, Eds., pp. 85–102, The University of North Carolina Press, Chapel Hill, NC, USA, 1960.
- [22] J. Van Bezooijen, H. Overmars, and R. Siep, "Bulk extraction of potato-cyst nematodes," *Mededelingen Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen der Rijksuniversiteit Gent*, vol. 64, no. 3, pp. 377–381, 1999.
- [23] Y. Lee, J. A. Nelder, and Y. Pawitan, *Generalized Linear Models with Random Effects: Unified Analysis via H-likelihood*, CRC Press, 2006.

- [24] C. H. Schomaker and T. H. Been, "Plant growth and population dynamics," in *Plant Nematology*, R. N. Perry and M. Moens, Eds., CABI, Oxfordshire, UK, 2006.
- [25] L. Mol, O. C. Huisman, K. Scholte, and P. C. Struik, "Theoretical approach to the dynamics of the inoculum density of *Verticillium dahliae* in the soil: first test of a simple model," *Plant Pathology*, vol. 45, no. 2, pp. 192–204, 1996.
- [26] J. C. Zadoks and R. D. Schein, *Epidemiology and Plant Disease Management*, Oxford University Press, New York, NY, USA, 1979.
- [27] P. McCullagh and J. A. Nelder, *Generalized Linear Models*, Chapman and Hall, London, UK, 2nd edition, 1989.
- [28] J. France and J. H. M. Thornley, *Mathematical Models in Agriculture*, Butterworths, London, UK, 1984.
- [29] D. A. Elston, M. S. Phillips, and D. L. Trudgill, "The relationship between initial population density of potato cyst nematode *Globodera pallida* and the yield of partially resistant potatoes," *Revue Nématologique*, vol. 14, no. 2, pp. 221–229, 1991.
- [30] P. M. Lukacs, K. P. Burnham, and D. R. Anderson, "Model selection bias and Freedman's paradox," *Annals of the Institute of Statistical Mathematics*, vol. 62, no. 1, pp. 117–125, 2010.
- [31] VSN International, "GenStat for Windows 13th Edition," VSN International, Hemel Hempstead, UK, 2010, GenStat.co.uk.
- [32] D. C. Montgomery and E. A. Peck, *Introduction to Linear Regression Analysis*, John Wiley & Sons, New York, NY, USA, 1982.
- [33] R. M. Royall, *Statistical Evidence: a Likelihood Paradigm*, Chapman and Hall, London, UK, 1997.
- [34] J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky, "Bayesian model averaging: a tutorial," *Statistical Science*, vol. 14, no. 4, pp. 382–417, 1999.
- [35] J. Postma, M. T. Schilder, J. Bloem, and W. K. Van Leeuwen-Haagsma, "Soil suppressiveness and functional diversity of the soil microflora in organic farming systems," *Soil Biology and Biochemistry*, vol. 40, no. 9, pp. 2394–2406, 2008.