Eur J Plant Pathol (2009) 124:413–425
DOI 10.1007/s10658-009-9428-0

A predictive model for early-warning of Septoria leaf blotch on winter wheat

D. E. te Beest · M. W. Shaw · S. Pietravalle · F. van den Bosch

Received: 10 January 2008 / Accepted: 19 January 2009 / Published online: 4 March 2009

Abstract Disease–weather relationships influencing Septoria leaf blotch (SLB) preceding growth stage (GS) 31 were identified using data from 12 sites in the UK covering 8 years. Based on these relationships, an early-warning predictive model for SLB on winter wheat was formulated to predict the occurrence of a damaging epidemic (defined as disease severity of 5% or > 5% on the top three leaf layers). The final model was based on accumulated rain > 3 mm in the 80-day period preceding GS 31 (roughly from early-February to the end of April) and accumulated minimum temperature with a 0°C base in the 50-day period starting from 120 days preceding GS 31 (approximately January and February). The model was validated on an independent data set on which the prediction accuracy was influenced by cultivar resistance. Over all observations, the model had a true positive proportion of 0.61, a true negative proportion of 0.73, a sensitivity of 0.83, and a specificity of 0.18. True negative proportion increased to 0.85 for resistant cultivars and decreased to 0.50 for susceptible cultivars. Potential fungicide savings are most likely to be made with resistant cultivars, but such benefits would need to be identified with an in-depth evaluation.

Keywords *Septoria tritici* · *Mycosphaerella graminicola* · Validation · Window Pane · Disease forecasting

D. E. te Beest (✉) · F. van den Bosch
Biomathematics and Bioinformatics division,
Rothamsted Research,
Harpenden, Hertfordshire AL5 2 JQ, UK
e-mail: detebeest@hotmail.com

F. van den Bosch
e-mail: frank.vandenbosch@bbsrc.ac.uk

M. W. Shaw
University of Reading, School of Biological Sciences,
Reading RG6 6AS, UK

S. Pietravalle
Central Science Laboratory,
Sand Hutton, York YO41 1LZ, UK

Introduction

Wheat is both economically and in acreage the most important crop in the UK. The key foliar disease on wheat in England since the 1980s has been Septoria leaf blotch (SLB) caused by *Mycosphaerella graminicola* (anamorph *Septoria tritici*) (Hardwick et al. 2001). Because of the potential yield loss, growers tend to spray fungicides several times each year to protect their crops from this disease. Actual disease levels do not always justify a fungicide spray. In years with a low disease risk, a lower fungicide dose could be used (Hardwick et al. 2001).

If a SLB severity that is not economically damaging could be predicted, fungicide usage could be adjusted accordingly. This may have economic and

environmental benefits. There are several systems available to help a grower decide when to use a fungicide (Frahm and Volk 1993; Jorgensen et al. 1996; Parsons and te Beest 2004). These systems often require an input from a disease assessment to be able to make a prediction. However, the first recommended spray timing (growth stage (GS) 31) is before any meaningful disease assessment can be done. An indication of risk prior to the first fungicide spray would be helpful in guiding spray decisions. In this paper an early-warning predictive model that can give an indication of the risk of a SLB infection was developed.

Weather is known to be of great importance for fungal plant pathogens. Various weather factors influence epidemic progress and disease severity of SLB (Thomas et al. 1989; Shaw 1990; Shaw and Royle 1993; Chungu et al. 2001). By empirically identifying and quantifying the most influential relationships between measurable weather variables and SLB, these relationships can potentially be used for prediction. One method to identify these key weather factors is the iterative search strategy called 'Window Pane' (Coakley and Line 1982). Window Pane is an algorithm that iteratively searches time windows for a selection of weather factors to identify the main relationships between disease severity and weather. This method has been used for plant diseases in various crops (Coakley et al. 1985, 1988; Chuang and Jeger 1987; Pietravalle et al. 2003; te Beest et al. 2008) and in this paper is applied to SLB.

For SLB, quantitative relationships with weather have been identified before by Tyldesley and Thompson (1980), Coakley et al. (1985), Hansen et al. (1994), and Parker et al. (1999). These papers, however, did not look specifically at the occurrence or absence of a damaging epidemic. We define a 'damaging epidemic' as a mean disease severity of 5% or 75% over the top three leaves at GS 75 which is commonly seen as an economically-damaging disease infection (Gladders et al. 2001, Pietravalle et al. 2003). The discriminant analysis was introduced to Window Pane by Pietravalle et al. (2003) for this purpose. However, in Pietravalle et al. (2003) in-field weather data were used. This makes results less useful for practical application as it requires each grower to measure infield weather; we therefore used weather from meteorological stations. An early-warning model for SLB was demonstrated by Pietravalle et al. (2003), but Window Pane was not restricted accordingly. In this paper we restricted Window Pane and analysed relationships prior to the first fungicide spray at GS 31.

In this work, the objectives were: to gain insights into the weather factors in the period preceding GS 31 that contribute to a damaging epidemic of SLB later in the season and to use these insights to formulate an early-warning model for infection by SLB on winter wheat.

Materials and methods

Data

The sources of data for both model development and validation are summarised in Table 1. The model development data were used to find the strongest relationships between weather factors and disease and to build the early-warning model. The validation data were used to validate the early-warning model. All sites were located in England, Scotland or Northern Ireland and were named after the research station where the experiment was located or after a city near the experiment. For most sites observations over a number of years were available. Each combination of year and site is described as a 'year-site'. All disease data were from plots unsprayed with fungicides. The data consisted of observations of the percentage of each leaf covered by the disease. In total, 41 year-sites were available for model development, covering 12 sites and the period from 1994 until 2002, with the exception of 1998 (Table 1). The validation data consisted of 41 year-sites, of which 34 were for the period 2003 to 2005 and 7 for the period 1994 to 2000. These data covered 16 sites in total, 11 of which were not represented in the model development data set.

Disease severity at GS 75, when grain is at the medium-milk stage, is a good predictor of the final yield loss (King et al. 1983) and was thus used as a benchmark for final yield loss. The top three leaves contribute most to yield formation (Seck et al. 1991) and these were used to calculate the overall disease severity according to the equation:

$$Sev_{Total} = \frac{\sum_{i=1}^{3} Sev_i \cdot Area_i}{\sum_{i=1}^{3} Area_i}. \quad (1)$$

in which Sev_i represents the percentage disease on leaf i, $Area_i$ represents the leaf area of leaf i and Sev_{Total} represents the average disease percentage over all three leaves. Sev_i and Sev_{Total} are both in percentage and leaf area ($Area_i$) is in mm. On average 40 plants per observation were assessed, (further details can be found in the references footnoted in Table 1). If it was not possible to classify whether a damaging epidemic had taken place, due to other foliar diseases, that observation was not used. This only occurred for a few observations. In the validation data, measurements of leaf area size were missing.

Table 1 Overview of model development and validation data for prediction of a damaging epidemic of Septoria leaf blotch on wheat

Model development data				Validation data			
Site	Year	Obs[a]	E[b]	Site	Year	Obs[a]	E[b]
Northern Ireland[c]	2000[h]	2	2	Askham Bryan[e]	2003[l]	1	1
Arthur Rickwood[d]	1994[i]	4	0		2004[l]	5	2
	1995[i]	3	2	Blenheim[e]	1996[i]	4	0
	1997[i]	4	0	Boxworth[d]	1994[i]	4	0
Bridgets[d]	1995[i]	4	4		1995[i]	4	0
	1996[i]	4	4		1996[i]	4	0
	1997[i]	4	1		1997[i]	3	1
Gleadthorpe[d]	1994[i]	4	0		2003[l]	1	0
	1995[i]	4	0		2004[l]	4	3
	1996[i]	4	0		2005[l]	4	3
Harpenden[e]	1996[i]	4	0	Newcastle[e]	2004[l]	5	2
High Mowthorpe[d]	1994[i]	4	3		2005[l]	4	1
	1995[i]	4	0	Drayton[e]	2003[l]	1	0
	1996[i]	4	0		2004[l]	5	4
	1997[i]	4	4		2005[l]	4	0
	1999[h]	3	3	Exeter[e]	2003[l]	1	1
	2000[h]	7	6		2004[l]	5	2
	2001[j]	5	1		2005[l]	4	1
	2002[j]	1	1	Gleadthorpe[d]	2003[l]	1	1
Bristol[f]	2000[h]	5	5		2004[l]	5	4
	2001[j]	1	0		2005[l]	4	0
	2002[j]	2	2	Tadcaster[e]	2003[l]	1	1
Morley[d]	1999[h]	2	2	High Mowthorpe[d]	2003[l]	1	0
	2000[h]	2	2		2004[l]	5	5
Rosemaund[d]	1994[k,i]	25	6		2005[l]	4	2
	1995[k,i]	25	22	Morley[d]	2003[l]	1	1
	1996[k,i]	25	24		2004[l]	5	4
	1997[k,i]	23	23		2005[l]	4	0
	1999[h]	3	3	Rosemaund[d]	2003[l]	1	1
	2000[h]	3	3		2004[l]	5	5
	2001[j]	5	0		2005[l]	4	4
	2002[j]	5	5	Starcross[d]	1999[h]	2	2
Aberdeen[g]	1999[h]	2	0		2000[h]	2	2
	2000[h]	2	2	Lavenham[e]	2003[l]	1	0
Tadcaster[e]	1997[i]	4	4		2004[l]	5	4
Terrington[d]	1994[i]	4	0		2005[l]	4	1
	1995[i]	4	0	Terrington[d]	2003[l]	1	0
	1996[i]	4	1		2004[l]	5	2
	1997[i]	4	0		2005[l]	4	2
	1999[h]	3	0	Wye[d]	2003[l]	1	1
	2000[h]	2	2	York[e]	2005[l]	4	4
Total		228		Total		133	

Table 1 (continued)

Model development data				Validation data			
Site	Year	Obs[a]	E[b]	Site	Year	Obs[a]	E[b]
Resistant group		50%		Resistant group		50%	
Intermediate group		34%		Intermediate group		38%	
Susceptible group		16%		Susceptible group		11%	

[a] Obs is the number of observations available for each year-site, each observation is an experiment with a different cultivar

[b] E is the number of damaging epidemics (> 5% severity) in each year-site

[c] Agricultural Research Institute of Northern Ireland

[d] ADAS

[e] Central Science Laboratory

[f] Long Ashton Research Station

[g] Scottish Agricultural Colleges

[h] Derived from the DESSAC validation experiment as described in Milne et al. (2003)

[i] A subset of these data is described in Pietravalle et al. (2003) where the susceptible subset was used for Window Pane analysis

[j] A subset of these data is described in Lovell et al. (2002)

[k] Described in Parker et al. (2004)

[l] originates from the 'crop monitor' programme (www.cropmonitor.co.uk)

The level of disease severity was calculated as the average disease severity on the top three leaf layers in these cases.

Many different cultivars were included in the data. The cultivars were divided into three groups, according to the resistance rating published by the National Institute of Agricultural Botany (NIAB) (Anonymous 1994–2005). The NIAB ranking system uses ratings of 1 to 9, with 1 being very susceptible and 9 very resistant. None of the cultivars in the data set had a resistant rating > 7 or < 3. A 'resistant' group contained all cultivars with a NIAB rating of 6 and 7, an 'intermediate' group contained all cultivars with ratings 4 and 5, and a 'susceptible' group contained all cultivars with a rating of 3. The dynamics of the pathogen on the crop depend on the disease resistance of the cultivar. In wheat, the position of the leaves (e.g. distance between nodes) may prevent pathogen dispersal (Lovell et al. 2004b), or resistance genes may be present (Chartrain et al. 2005). The effect of cultivar resistance was therefore examined in the analysis.

To maximise practical applicability of the early-warning predictive model, weather data from meteorological stations were used. For each year-site, weather data were retrieved from the Biotechnology and Biological Sciences Research Council (BBSRC) meteorological website (http://www.bits.bbsrc.ac.uk/metweb). Weather data at these stations was daily and the data were widely available. All weather data used to develop the model were from weather stations based locally, which means the distance between the weather station and the experimental field varied from a few hundred meters to a maximum of 5 km. For the validation data, local weather data were often not available and weather data were retrieved from the closest meteorological weather station, which means that distances were up to 15 km. For validation purposes this is a realistic distance as often weather data would not be available locally in practical applications. This adds to the 'rigour and realism' of the validation test.

The following daily weather data were retrieved: maximum, minimum, and mean temperature (all in °C; mean temperature was calculated as the average of maximum and minimum temperature), mean % relative humidity (RH), radiation (radiant energy from the sun in MJ m^{-2}), rain (in mm), sunshine (h), vapour pressure (mean in millibars), wind run (daily distance travelled by the wind in km), and wind speed (mean in m s^{-1}). Due to sensor errors and because some sites did not measure all weather factors, some weather data were missing. In the model development data set for vapour pressure, rain, minimum, maximum, and mean temperature, 3% of the weather data was missing; for wind run and speed, 20% of the

weather data was missing; for RH, radiation and sunshine, 32% of the weather data was missing. In the validation data set for minimum, maximum, and mean temperature, 2% of the weather data was missing; for rain and vapour pressure, 10% of the weather data was missing; for RH, radiation and sunshine, 28% of the weather data was missing; for wind run and speed, 40% of the weather data was missing. The weather data that were missing were omitted from the analysis as described in the Materials and methods section.

Model development

The strongest disease–weather relationships were identified with Window Pane, described by Coakley and Line (1982) and Pietravalle et al. (2003). The influence of weather on SLB is dependent on the crop GS. To be in correspondence with the crop phenology of winter wheat, all observations were aligned on GS 31. Window Pane iteratively searched the period preceding GS 31. The time lag marked the total searched period, counting backwards from GS 31, which was set at day 0. A window is formed by first counting the time lag backwards, which is then the starting point of one window, and by then counting the window length forward. The first window was formed from the maximum time lag (230 days) and the maximum window length (120 days). By reducing the window with the window increment (5 days), consecutive windows were formed. When the minimum window length was reached (20 days), the time lag was reduced by its increment (10 days) and the process started again until the minimum time lag (20 days) was reached. In each window a weather variable was calculated according to a specified weather function. If a window contained days with missing weather data, then that window was not used to ensure that each weather variable was calculated accurately in each window. The weather functions used were: (1) number of days the numerical value of the weather data under consideration was above or below a threshold, (2) number of consecutive days above or below a threshold, (3) the average value of the weather data in a window, and (4) accumulation of the numerical value of the weather data above or below a threshold. A range of values were tested for each weather factor. This range was based on range of value each weather factor had in the data (based on summary statistics such as median and quartiles). This range was tested in intervals. For example, minimum temperature was tested in the range −3°C to 15°C in steps of 3°C with all weather functions.

We applied the discriminant analysis as introduced into Window Pane by Pietravalle et al. (2003). The discriminant analysis was used to calculate the proportion misclassification (m) according to the equations:

$$m = \frac{\min\left(\sum_{i=1}^{n}\left|\left(Y_i - X_i^{(-)}\right)\cdot w_i\right|, \sum_{i=1}^{n}\left|\left(Y_i - X_i^{(+)}\right)\cdot w_i\right|\right)}{\sum_{i=1}^{n} w_i} \tag{2}$$

with

$$X_i^{(-)} = \begin{cases} 0 & \text{if } X_i > x_0 \\ 1 & \text{otherwise} \end{cases}$$

$$X_i^{(+)} = \begin{cases} 0 & \text{if } X_i \le x_0 \\ 1 & \text{otherwise} \end{cases}$$

$$w_i = \frac{1}{\#Obs_j} \tag{3}$$

Disease observations were classified into a binary vector (Y_i). Observations with a disease severity > 5% (damaging epidemic) were scored as 1. Observations with a disease severity < 5% (no damaging epidemic) were scored as 0. The contribution of each observation i to the misclassification was scaled with weight w_i according to the number of observations present in the j^{th} year-site (Obs_j) (te Beest et al. 2008). The weather variable (X_i) was classified to a binary vector (X_i^- and X_i^+) according to a variable threshold x_0. X_i^- is 0 if X_i is $\le x_0$ and 1 otherwise, X_i^+ is the opposite of X_i^-. The section (Y_{ii}) quantified for one particular observation i whether the binary value of disease observations (Y_i) and the weather observation (X_i) were different. By summarising these values over all observations (over i), the total number of observations classified incorrectly was calculated.

The division of the weight of the number of observations classified incorrectly by the total weight of all observations resulted in the misclassification (m). The binary value of X_i changes with the threshold x_0 with which it was classified. This threshold x_0 can be set such that it minimises the misclassification (m) in each window. The scaling was introduced because

year-sites were unevenly represented in the data set which gave well-represented year-sites more leverage on the result. With this scaling every year-site was given an equal weight in the analysis.

The effect of cultivar resistance was ignored in the first analysis and the misclassification was calculated over all observations in the model development data set. The three cultivars groups were analysed independently in a second analysis with the same procedure. For each cultivar group the year-site weighting (w_i) was adjusted to that cultivar group so that within the cultivar groups year-sites were also represented evenly.

The strongest disease–weather relationships were selected from the Window Pane analysis. The iterative nature of Window Pane means that a large number of potential relationships are generated, which makes it likely that by pure chance a spurious relationship is found (Pietravalle et al. 2003). This problem can be reduced with strict selection criteria. The first criterion was the misclassification as calculated with the discriminant function (equations 2 and 3). The second criterion was the run length of a relationship. The run length is the consecutive number of windows with a low misclassification (Pietravalle et al. 2003). If a low misclassification does not occur in multiple consecutive windows the relationship is unlikely to be genuine. This criterion is based on the existence of autocorrelation between these consecutive windows. The third criterion is the number of similar relationships occurring with related weather variables. The misclassification selection threshold was set at 0.23. The run length selection criterion for the cultivar groups was set to three or more consecutive windows with a misclassification < 0.25. Over all cultivars this criterion was set to two or more consecutive windows. The criteria were set to select the strongest 1% of the relationships. Finally, all identified relationships should relate to the life-cycle of the *M. graminicola* pathogen. Identified relationships are less likely to be spurious if a biological explanation can be found. As in Pietravalle et al. (2003), the selected weather variables were combined into models using Fisher's linear discriminant function (Mardia et al. 1979).

Plant disease predictive models often lack a thorough validation on an independent data set, the reason being that data are often limited and necessary for model building. Especially for models that are empirically derived, a validation is crucial to check whether the identified relationships hold on an independent data set. In this paper, the derived predictive models were validated on an independent data set equal in size to the model development data set and covering different years and different sites. This means that the validation to some degree took into account the effects of cultivar replacement and pathogen evolution, it reduces the effect of spatial correlations, and there was a greater distance between the meteorological site and the experimental field. This validation therefore forms an excellent test to see if the models hold on an independent data set and in a practical application. In the validation the number of errors made when assigning observations to specific classes was quantified as a misclassification. For unambiguity this misclassification has been called 'validation-misclassification'.

Results

Based on the misclassification and run length (Table 2) the best disease–weather relationships over all cultivars were identified with accumulated rain in the 80- and 110-day period preceding GS 31, and minimum temperature and vapour pressure both in the 50-day period starting from 120 days preceding GS 31. Counting from the average GS 31 date (end of April, Fig. 1) this corresponds with the period from mid-January to the end of April (for rain) and the period from January to February (for vapour pressure and minimum temperature). More rain, a higher temperature, and higher vapour pressure increased the risk of a damaging SLB epidemic. These disease–weather relationships had a run length longer than 3 windows and misclassification < 0.20 over all cultivars. A low misclassification means this variable has good predictive value, and higher run length means this variable is less likely to be spurious. Other disease–weather relationships were found with wind, radiation, and sunshine. An increase in wind, radiation or sunshine reduced the risk of a damaging SLB epidemic (Table 2, Fig. 1). There were differences among the three cultivar groups. In the resistant cultivar group, rain had the lowest misclassification, whereas in the intermediate cultivar group minimum temperature had the lowest misclassification.

Table 2 Disease–weather relationships with Septoria leaf blotch as identified with the Window Pane analysis

Weather variable	Time lag[a]	Window[a]	m[b]	RL[c]
All cultivars				
Rain (accumulation > 6 mm)[d]	110[f]	110[f]	0.18	3
Rain (accumulation > 3 mm)	80[g]	80[g]	0.19	4
Vapour pressure (number of days ≥ 9 mbar)[e]	120[h]	50[h]	0.20	7
Min temperature (accumulation > 0°C)	120[h]	50[h]	0.20	4
Max temperature (number of days ≤ 4°C)	110[h]	60[h]	0.21	13
Sunshine (accumulation > 4 h)	230[i]	40[i]	0.20	2
Radiation (accumulation > 8 MJ)	130[h]	60[h]	0.23	2
Resistant cultivars				
Rain (accumulation > 6 mm)	110[f]	110[f]	0.12	7
Vapour pressure (number of days ≥ 9 mbar)	120[h]	50[h]	0.17	7
Min temperature (accumulation > 0°C)	120[h]	50[h]	0.20	3
Windrun (accumulation < 100 km)	100[j]	80[j]	0.23	4
Windrun (number of days ≤ 100 km)	130[h]	110[h]	0.23	4
Intermediate cultivars				
Min temperature (accumulation > 0°C)	120[h]	50[h]	0.18	4
Max temperature (number of days ≤ 4°C)	120[k]	80[k]	0.20	13
Rain (accumulation > 2 mm)	90[g]	90[g]	0.20	3
Sunshine (accumulation > 4 h)	230[i]	45[i]	0.21	5
Vapour pressure (number of days ≥ 9 mbar)	120[h]	50[h]	0.22	3
Susceptible cultivars				
Min temperature (accumulation > 0°C)	130[h]	60[h]	0.20	3
Rain (consecutive number of days ≥ 6 mm)	90[g]	85[g]	0.22	9
Sunshine (accumulation > 4 h)	230[i]	45[i]	0.22	5

[a] 'Time lag' is the number of days before GS 31 where the window starts, 'Window' is the length of the window, counting forwards from Time lag. Together they mark the period in which the weather variable was calculated

[b] m is the misclassification of the weather variable

[c] RL means 'Run Length' which is the number of consecutive windows during which misclassification remained below the threshold

[d–e] Within parentheses is shown the weather function with threshold

[d] 'accumulation > 6 mm' means daily rain > 6 mm accumulated

[e] 'number of days ≥ 9 mbar' means total number of days with vapour pressure > 9 mb

[e–g] Indicated period corresponds roughly with

[f] mid-January to the end of April

[g] early-February to the end of April

[h] January and February

[i] September and October

[j] mid-January to mid-April

[k] early-January to early-March

Separate models were built over all cultivars, over the resistant, and over the intermediate cultivars. In the susceptible cultivar group the number of observations was too small to build separate models. Over all cultivars, three different models were identified, all with a similar misclassification on the model development data. These models consisted of rain with minimum temperature, rain with wind run, and vapour pressure with wind run (Table 3). Initially we identified only a weak disease–weather relationship with wind run. However, when we combined wind in a model with water vapour pressure or rain, wind run had a strong additive effect on the misclassification of the model. The model with rain and minimum temperature had the lowest validation-misclassification over all cultivars and therefore should be the most useful model. The best model for the intermediate cultivars consisted of the same

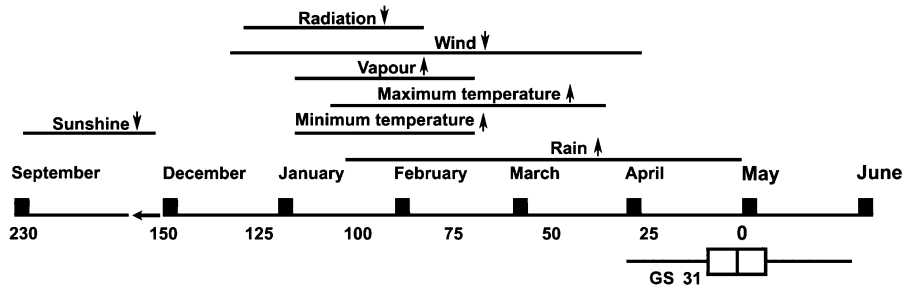

Fig. 1 Overview of the identified disease–weather relationships for Septoria leaf blotch. Time bar indicates time of the season counted back from GS 31, set at time 0. The box plot indicates the variation in the date at which GS 31 occurred. The arrows indicate whether an increase or decrease in the weather variable increases the risk of a damaging epidemic, e.g. whether more rain increased the risk of a damaging epidemic

variables as the overall model. For resistant cultivars the best model on the model development data consisted of accumulated rain > 6 mm in the 110-day period preceding GS 31. In the validation this model had a higher validation-misclassification on the resistant cultivars than the minimum temperature–rain model formulated over all cultivars. For resistant cultivars the overall minimum temperature and rain model should be the most useful. Reformulating the model for each cultivar group (Table 4) did not improve the misclassification of the model and one model over all cultivars was therefore preferred. Rain

Table 3 Models for predicting a damaging epidemic of Septoria leaf blotch on winter wheat.

	TL[a]	W[a]	m[b]	validation-m[c]
Model 1a—All cultivars				
Rain (accumulation > 3 mm)[d]	80	80	0.17	0.36
Min temperature (accumulation > 0°C)[d]	120	50		
Model[e]: 0.046 · Rain + 0.042 · Tmin − 6.69 > 0				
Model 1a—Resistant cultivars			0.17	0.38
Model 1a—Intermediate cultivars			0.16	0.27
Model 1b—All cultivars with logistic regression			0.17	0.36
Model[f]: $p = \frac{1}{1+e^{4.38-0.22 \cdot Rain-0.032 \cdot T\min}}$				
Model 2—All cultivars				
Vapour pressure (Number of days ≥ 9 mbar)[d]	120	50	0.17	0.41
Wind run (Number of days ≤ 100 km) [d]	130	110		
Model[e]: 0.47 · Vapour − Wind · 0.015 − 5.14 > 0				
Model 3—All cultivars				
Rain (accumulation > 6 mm) [d]	110	110	0.17	0.5
Wind run (accumulation < 100 km) [d]	100	80		
Model[e]: 0.11 · Rain − 0.0042 · Wind − 5.17 > 0				
Model 4—Resistant cultivars				
Rain (accumulation > 6 mm)[d]	110	110	0.12	0.41
Model[e]: 0.07 · Rain − 2.94 > 0				

[a] 'Time lag' is the number of days before GS 31 where the window starts, 'Window' is the length of the window, counting forwards from 'Time lag'. Together they mark the period in which the weather variable was calculated

[b] m is the misclassification of the model on the model data, and

[c] 'validation -m' is the validation-misclassification of the model on the validation data

[d] Within parentheses is shown the weather function with threshold, 'accumulation > 3 mm' for example means daily rain > 3 mm accumulated

[e] 'Model' indicates the equation representing the model, the models are conditional and predict a damaging epidemic if the statement is true

[f] Represents the logistic regression model; a damaging epidemic is predicted if $P>0.5$

Table 4 Minimum temperature and rain model reformulated for each cultivar group (Model 1, Table 3)

	m cultivar model[a]	m overall model[b]
Resistant		
Model[c]: $0.053 \cdot \text{Rain}^{d} + 0.038 \cdot \text{Tmin}^{e} - 6.80 > 0$	0.17	0.17
Intermediate		
Model[c]: $0.027 \cdot \text{Rain}^{d} + 0.023 \cdot \text{Tmin} - 3.88 > 0$	0.16	0.16
Susceptible		
Model[c]: $0.046 \cdot \text{Tmin}^{e} - 4.20 > 0$	0.29	0.30

[a] misclassification within cultivar group (m) with the reformulated cultivar models

[b] misclassification within cultivar group (m) with the model formulated over all cultivars as described in Table 3.3

[c] The equation representing the model, the discriminant analysis models are conditional and predict a damaging epidemic if the statement is true

[d] Daily rain accumulated > 3 mm in an 80-day window preceding GS 31

[e] Daily minimum temperature accumulated > 0°C in a 50-day period starting from 120 days preceding GS 31

increased in importance in the resistant cultivar group and decreased in importance as a variable in the susceptible group. Logistic regression gave the same result as Fisher's linear discriminant analysis (Table 3). This confirms that the best classification threshold for the underlying data was found.

The accuracy of the model based on minimum temperature and rain (Model 1, Table 3) was further analysed (Table 5). Described are the estimated conditional probabilities (true positive, false positive, true negative and false negative proportion), sensitivity, and specificity in each cultivar group and over all cultivars on the validation data set. Sensitivity is the fraction of true positives among the positive predictions. Specificity is the fraction of true negatives among the negative predictions. There is a clear shift in prediction accuracy with cultivar resistance as can be seen by a shift in true negative proportion from 0.85 in the resistant group to 0.50 in the susceptible group. Overall the final model has a high sensitivity and a low specificity, especially in the resistant cultivar group (Table 5).

Discussion

A model based on accumulated rain in an 80-day window preceding GS 31 and accumulated minimum temperature > 0°C in a 50-day period starting from 120 days preceding GS 31 (Model 1, Table 3) proved to be best based on both model development data and validation data.

Disease–weather relationships with temperature were identified with accumulated minimum temperature, with a 0°C base roughly in the period January to February and the number of days with a maximum temperature < 4°C in roughly the period from February to March. Both periods are early enough in the season to be used for early-warning predictions. Temperature in this period of the season is most likely to influence the latent period of *M. graminicola*. The latent period of *M. graminicola* is long at low temperatures (Shaw 1990; Lovell et al. 2004a). Low temperatures in winter and spring in the U.K. slow down the growth of the pathogen. If temperature in these periods increases, temperature becomes less limiting and *M. graminicola* growth is increased, resulting in a higher risk of SLB. In previous work, a relationship between SLB and the consecutive number of days with a minimum temperature > 7°C in April was found (Coakley et al. 1985). Another relationship between SLB and the number of days > 7°C was found in the period from early-January until early-March with infield weather data (Pietravalle et al. 2003). Both correspond with the disease–weather relationship found in this paper. Temperatures < −2°C in November were found to negatively influence SLB disease severity (Parker et al. 1999; Gladders et al. 2001). In our analysis no disease–weather relationships with temperature were identified in November. This may be due to the fact that data in these analyses originated from plots treated with fungicides.

We found that high vapour pressure, low wind runs, and high radiation all increased the risk of a damaging

Table 5 Performance of the minimum temperature and rain model (Model 1, Table 3) formulated for the prediction of a damaging epidemic of Septoria leaf blotch on winter wheat

		Obs[a]	YS[b]	Accuracy	Sens[c]/Spec[d]
All cultivars					
Epidemic predicted	Epidemic present	43	16	0.61[e]	0.83[c]
	Epidemic absent	27	12	0.39[f]	
Absence of epidemic predicted	Epidemic present	9	6	0.27[g]	0.18[d]
	Epidemic absent	24	11	0.73[h]	
Resistant cultivars					
Epidemic predicted	Epidemic present	16	8	0.42	0.89
	Epidemic absent	22	11	0.58	
Absence of epidemic predicted	Epidemic present	2	1	0.15	0.06
	Epidemic absent	11	7	0.85	
Intermediate cultivars					
Epidemic predicted	Epidemic present	20	16	0.87	0.80
	Epidemic absent	3	3	0.13	
Absence of epidemic predicted	Epidemic present	5	5	0.31	0.36
	Epidemic absent	11	10	0.69	
Susceptible cultivars					
Epidemic predicted	Epidemic present	7	7	0.78	0.78
	Epidemic absent	2	2	0.22	
Absence of epidemic predicted	Epidemic present	2	2	0.50	0.50
	Epidemic absent	2	2	0.50	

[a] Obs is the number of observations

[b] YS is the number of year-sites

[c] sensitivity and

[d] specificity of the model as described in Hughes et al. (1999)

[e] True positive proportion

[f] False positive proportion

[g] False negative proportion

[h] True negative proportion.

SLB epidemic. In a leaf wetness model for oilseed rape (*Brassica napus*), wind, radiation, and vapour pressure were the main factors influencing leaf wetness (Papastamati et al. 2004). Long periods of leaf wetness, which are essential for disease development, influence latent period and spore production (Chungu et al. 2001; Crowe et al. 1978; Djurle et al. 1996). Because these disease–weather relationships were found in overlapping periods it is likely that they influence leaf wetness. The disease–weather relationships we found indicates that this effect continues over a long period, from January to April. This is approximately the same period as identified for temperature. When temperature is high enough for development, leaf wetness additionally becomes important in the development of *M. graminicola*. Previously, wind in a period close to the period found in this paper was found to influence SLB (Pietravalle et al. 2003). The suggested mechanism was also that much wind shortens leaf wetness periods.

Disease–weather relationships were identified with accumulated rain in an 80- and 110-day window preceding GS 31 (roughly from mid-January to the end of April). This disease–weather relationship is in a period of the season early enough for early-warning predictions. The importance of rain can be explained by the notion that within-season *M. graminicola* disperses mainly through splash-dispersed conidia (Shaw and Royle 1993). More rain leads to more build-up of conidia and a greater risk of a damaging epidemic later in the season. The period we identified with rain starts at the end of January which is after the start of the period we identified with latent period (minimum temperature) and leaf wetness (vapour, wind, and radiation). It is therefore likely that from

this period *M. graminicola* starts growing faster, more spores are available, and that the spread of spores across the plant becomes important. Later in the season, additional to the build-up of conidia, rain also influences the spread to new (emerging) leaf layers. Our analysis stopped at GS 31, but it is reasonable to assume that rain continues to be of importance beyond GS 31. In previous analyses similar quantitative relationships between SLB and weather have been identified. A relationship with rain and disease severity after GS 31 (mid-May to mid-June) was identified by Hansen et al. (1994) and Tyldesley and Thompson (1980). The absence of rain in April was found to be negatively correlated with disease severity (Coakley et al. 1985). Disease severity was also positively correlated with the consecutive number of days with > 10 mm of rain (Thomas et al. 1989). Spring rain (number of days with > 1 mm of rain in May and June) explained part of the long-term variation in *M. graminicola* abundance (Shaw et al. 2008). Finally, a relationship between SLB and the number of days with rain > 5 mm in a period early-March until the end of April with infield weather data was found by Pietravalle et al. (2003).

During September and October more sunshine reduced the risk of a SLB epidemic. Previously, a negative correlation between the disease severity and average sunshine found in August was linked to the abundance of ascospores and the timing of sowing and harvest (Daamen and Stol 1992). The disease-weather relationship with sunshine found in this paper is likely to be related to a similar process. In Shaw et al. (2008) high summer temperature, related to sunshine, was also related to less abundance of *M. graminicola*, explaining part of the long-term variation in *M. graminicola* abundance.

There was a strong effect of cultivar resistance in the validation (Table 5). On susceptible cultivars a damaging epidemic often occurred even though one was not predicted. On resistant cultivars a damaging epidemic often did not occur even though one was predicted. This suggests there is a relation between cultivar resistance and how rain and minimum temperature influence *M. graminicola* that is currently not captured in the model. For this reason cultivar resistance possibly should be taken into account in the model. On the model development data the effect of cultivar resistance was not strong. Separate cultivar models did not perform better than one model formulated over all cultivars (Tables 3, 4). As both the model development and validation data sets were from different periods, a factor in the interaction between cultivar resistance and *M. graminicola* may have changed in time. Cultivar resistance may, for example, have improved, or classification into resistance groups may have improved. It is also possible that classification in resistance ratings were more consistent in the validation data set because validation data were from one data set whereas the model development data were from multiple data sets. In the Window Pane analysis, although all cultivar groups were influenced by similar weather factors, rain appeared to be a more important weather factor for the resistant cultivars (Tables 2, 4) than for intermediate cultivars. It is possible that a damaging epidemic requires more rain for cultivars with disease escape characteristics resulting from canopy and crop structure, compared to susceptible cultivars (Lovell et al. 2004b). Resistance to the pathogen per se may reduce the reproduction rate of *M. graminicola* (Chartrain et al. 2005). This may mean more spores (through dispersal) are needed to cause a similar infection compared to a susceptible cultivar. How dispersal and reproduction rate are influenced by cultivar resistance and how this affects the epidemic growth rate can be analysed according to Segarra et al. (2001). Although cultivar resistance partly explains false positives for resistant cultivars and false negatives for susceptible cultivars, it does not explain the false negatives for resistant cultivars and false positives for susceptible cultivars. Factors outside those included in the model and cultivar resistance play a role. These factors may be local weather events that are not measurable from a meteorological station, weather events happening after GS 31, site-specific events such as the availability of inoculum, or population dynamic factors as yet unidentified. If, without the predictive model, a damaging epidemic was predicted for all susceptible crops and an absence of an epidemic was predicted for all resistant cultivars, then this prediction would be correct in 0.69 and 0.64 of cases. The model increases accuracy to 0.78 and 0.85, respectively.

In comparison with previous analyses (Tyldesley and Thompson 1980; Coakley et al. 1985; Hansen et al. 1994; Parker et al. 1999; Pietravalle et al. 2003) our model gives an early prediction of a damaging epidemic because of the restricted search before GS 31 and because weather data from meteorological stations were

used. Also, results have been validated on an independent data set which increases confidence in the result. Confidence is also increased because results can be compared with results of these previous analyses.

The current threshold is set to maximise classification, as are the selected weather variables. A different threshold may be preferred, depending on the aim of the grower, and possibly other beneficiaries. Lowering the threshold reduces the number of false negatives and increases false positives, and increasing the threshold does the opposite; a full analysis on this trade-off can be made with the ROC curve (Hughes et al. 1999). Potential benefits of the model in terms of financial profit can not be identified with such an analysis, yet this is the most deciding factor in whether a predictive model is likely to be adopted (Royle and Shaw 1988). A more in-depth validation procedure to evaluate these benefits will be introduced in later work.

Acknowledgements This project has been funded by The Department for Environment, Food and Rural Affairs (DEFRA), UK. Rothamsted Research receives support from the UK Biotechnology and Biological Sciences Research Council (BBSRC). Weather data used are Crown Copyright supplied by the UK Meteorological Office.

References

Anonymous. (1994–2005). *Cereal variety handbook : NIAB recommended list of cereals*.

Chartrain, L., Joaquim, P., Berry, S. T., Arraiano, L. S., Azanza, F., & Brown, J. K. M. (2005). Genetics of resistance to Septoria tritici blotch in the Portuguese wheat breeding line TE 9111. *Theoretical and Applied Genetics*, *110*, 1138–1144.

Chuang, T. Y., & Jeger, M. J. (1987). Predicting the rate of development of black sigatoka (*Mycosphaerella fijiensis* var. *difformis*) disease in southern Taiwan. *Phytopathology*, *77*, 1542–1547.

Chungu, C., Gilbert, J., & Townley-Smith, F. (2001). Septoria tritici blotch development as affected by temperature, duration of leaf wetness, inoculum concentration, and host. *Plant Disease*, *85*, 430–435.

Coakley, S. M., & Line, R. F. (1982). Prediction of stripe rust epidemics on winter wheat using statistical models. *Phytopathology*, *72*, 1006.

Coakley, S. M., Line, R. F., & McDaniel, L. R. (1988). Predicting stripe rust severity on winter wheat using an improved method for analyzing meteorological and rust data. *Phytopathology*, *78*, 543–550.

Coakley, S. M., McDaniel, L. R., & Shaner, G. (1985). Model for predicting severity of Septoria tritici blotch on winter wheat. *Phytopathology*, *75*, 1245–1251.

Crowe, M. J., Coakley, S. M., & Emge, R. G. (1978). Forecasting dew duration at Pendleton, Oregon, using simple weather observations. *Journal of Applied Meteorology*, *17*, 1482–1487.

Daamen, R. A., & Stol, W. (1992). Surveys of cereal diseases and pests in the Netherlands .5. Occurrence of *Septoria* spp. in winter wheat. *Netherlands Journal of Plant Pathology*, *98*, 369–376.

Djurle, A., Ekbom, B., & Yuen, J. E. (1996). The relationship of leaf wetness duration and disease progress of glume blotch, caused by *Stagonospora nodorum*, in winter wheat to standard weather data. *European Journal of Plant Pathology*, *102*, 9–20.

Frahm, J., & Volk, T. (1993). Pro_Plant—a computer-based decision-support systems for cereal disease control. *EPPO Bulletin*, *23*, 685–693.

Gladders, P., Paveley, N. D., Barrie, I. A., Hardwick, N. V., Hims, M. J., Langton, S., & Taylor, M. C. (2001). Agronomic and meteorological factors affecting the severity of leaf blotch caused by *Mycosphaerella graminicola* in commercial wheat crops in England. *Annals of Applied Biology*, *138*, 301–311.

Hansen, J. G., Secher, B. J. M., Jorgensen, L. N., & Welling, B. (1994). Thresholds for control of *Septoria* spp. in winter wheat based on precipitation and growth stage. *Plant Pathology*, *43*, 183–189.

Hardwick, N. V., Jones, D. R., & Slough, J. E. (2001). Factors affecting diseases of winter wheat in England and Wales, 1989–98. *Plant Pathology*, *50*, 453–462.

Hughes, G., McRoberts, N., & Burnett, F. J. (1999). Decision making and diagnosis in disease management. *Plant Pathology*, *48*, 147–153.

Jorgensen, L. N., Secher, B. J. M., & Nielsen, G. C. (1996). Monitoring diseases of winter wheat on both a field and a national level in Denmark. *Crop Protection*, *15*, 383–390.

King, J. E., Cook, R. J., & Melville, S. C. (1983). A review of Septoria diseases of wheat and barley. *Annals of Applied Biology*, *103*, 345–373.

Lovell, D. J., Hunter, T., Powers, S. J., Parker, S. R., & Van den Bosch, F. (2004a). Effect of temperature on latent period of Septoria leaf blotch on winter wheat under outdoor conditions. *Plant Pathology*, *53*, 170–181.

Lovell, D. J., Parker, S. R., Hunter, T., Welham, S. J., & Nichols, A. R. (2004b). Position of inoculum in the canopy affects the risk of S*eptoria tritici* blotch epidemics in winter wheat. *Plant Pathology*, *53*, 11–21.

Lovell, D. J., Parker, S. R., Paveley, N. D., & Worland, A. J. (2002). Understanding field resistance mechanisms for improved control of *Septoria tritici*. *Plant Protection Science*, *38*, 165–169.

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. London: Academic.

Milne, A., Paveley, N., Audsley, E., & Livermore, P. (2003). A wheat canopy model for use in disease management decision support systems. *Annals of Applied Biology*, *143*, 265–274.

Papastamati, K., McCartney, H. A., & van den Bosch, F. (2004). Modelling leaf wetness duration during the rosette stage of oilseed rape. *Agricultural and Forest Meteorology*, *123*, 69–78.

Parker, S. R., Lovell, D. J., Royle, D. J., & Paveley, N. D. (1999). Analysing epidemics of *Septoria tritici* for improved estimates of disease risk. In J. A. Lucas, P.

Bowyer, & H. M. Anderson (Eds.), *Septoria on Cereals: A Study of Pathosystems* (pp. 96–107). Oxford: CAB International.

Parker, S. R., Welham, S., Paveley, N. D., Foulkes, J., & Scott, R. K. (2004). Tolerance of Septoria leaf blotch in winter wheat. *Plant Pathology*, *53*, 1–10.

Parsons, D. J., & te Beest, D. E. (2004). Optimising fungicide applications on winter wheat using genetic algorithms. *Biosystems Engineering*, *88*, 401–410.

Pietravalle, S., Shaw, M. W., Parker, S. R., & van den Bosch, F. (2003). Modeling of relationships between weather and *Septoria tritici* epidemics on winter wheat: A critical approach. *Phytopathology*, *93*, 1329–1339.

Royle, D. J., & Shaw, M. W. (1988). The costs and benefits of disease forecasting in relation to its adoption in farming practice. In B. C. Clifford, & E. Lester (Eds.), *Control of plant disease: costs and benefits* (pp. 231–246). Oxford: Blackwell.

Seck, M., Roelfs, A. P., & Teng, P. S. (1991). Influence of leaf position on yield loss caused by wheat leaf rust in single tillers. *Crop Protection*, *10*, 222–228.

Segarra, J., Jeger, M. J., & van den Bosch, F. (2001). Epidemic dynamics and patterns of plant diseases. *Phytopathology*, *91*, 1001–1010.

Shaw, M. W. (1990). Effects of temperature, leaf wetness and cultivar on the latent period of *Mycosphaerella graminicola* on winter wheat. *Plant Pathology*, *39*, 255–268.

Shaw, M. W., Bearchell, S. J., Fitt, B. D. L., & Fraaije, B. A. (2008). Long-term relationships between enviroment and abundance in wheat of *Phaeosphaeria nodorum* and *Mycosphaerella graminicola*. *New Phytologist*, *177*, 229–238.

Shaw, M. W., & Royle, D. J. (1993). Factors determining the severity of epidemics of *Mycosphaerella graminicola* (*Septoria tritici*) on winter wheat in the UK. *Plant Pathology*, *42*, 882–899.

te Beest, D. E., Paveley, N. D., Shaw, M. W., & van den Bosch, F. (2008). Disease–weather relationships for powdery mildew and yellow rust on winter wheat. *Phytopathology*, *98*, 609–617.

Thomas, M. R., Cook, R. J., & King, J. E. (1989). Factors affecting development of *Septoria tritici* in winter wheat and its effect on yield. *Plant Pathology*, *38*, 246–257.

Tyldesley, J. B., & Thompson, N. (1980). Forecasting *Septoria nodorum* on winter wheat in England and Wales. *Plant Pathology*, *29*, 9–20.