

Assessing plant health in a network of experiments on hardy winter wheat varieties in France: multivariate and risk factor analyses

Serge Savary · Céline Jouanin · Irène Félix · Emmanuelle Gourdain · François Piraux · Laetitia Willocquet · François Brun

Accepted: 2 May 2016 / Published online: 11 May 2016
© Koninklijke Nederlandse Planteziektenkundige Vereniging 2016

Abstract A large network of field experiments has been conducted over several years across France to identify combinations of winter wheat cultivars and management practices in which partial resistances under limited chemical protection would achieve adequate disease management, while leading to satisfactory yield performance, and so achieve the double objective of ecological sustainability and economic viability. Little information is available to document the variation in multiple disease levels, a necessary step towards a chemical extensification process, in wheat networked experiments. This article provides a description of disease intensities in a set of 101 experiments totalling 3525 individual wheat plots over eight successive years (2003–2010). The diseases considered are brown rust

(BR, *Puccinia triticina*), yellow rust (YR, *Puccinia striiformis*), fusarium head blight (FHB, *Fusarium graminearum*, *F. culmorum*, and *F. avenaceum*), powdery mildew (PM, *Blumeria graminis*), and septoria tritici blotch (STB, *Zymoseptoria tritici*). Hierarchical cluster analysis led to the identification of three variety groups associated with (1) moderate-low disease levels in general, except for YR (moderate levels) – 16 varieties; (2) moderate-low BR, YR, and FHB levels, and moderate PM and STB levels – 12 varieties; (3) comparatively higher BR, YR, FHB, and STB levels, and moderate PM levels – 17 varieties. The association of disease levels represented as binary categories (i.e., epidemics vs. non-epidemics) with climatic years corresponded to chi-square values ($\chi^2 = 87.0\text{--}1402$) that were one to two orders of magnitude larger than the values corresponding to the associations of diseases with variety groups ($\chi^2 = 6.41\text{--}321$) or with levels of crop management ($\chi^2 = 21.2\text{--}82.1$). Multivariate non parametric analyses indicated the existence of three disease syndromes, two of which being dominated by BR or STB, and a third associated with diverse diseases and frequent FHB. This suggests that STB and BR might each be considered as key-stone species dominating specific wheat disease syndromes. Multiple correspondence analysis highlighted the linkages between multiple epidemic occurrence and the three characterized variety groups. Risk factors analyses conducted through logistic regressions provided quantitative estimates of the contribution of climatic years, variety groups, and crop management, to the likelihood of epidemic occurrence for each of the five diseases

S. Savary (✉) · C. Jouanin · L. Willocquet
INRA, UMR1248 AGIR, 24 Chemin de Borde Rouge, Auzeville,
CS 52627, F-31326 Castanet-Tolosan cedex, France
e-mail: Serge.Savary@toulouse.inra.fr

S. Savary
INPT, UMR AGIR, Université Toulouse, F-31029 Toulouse,
France

I. Félix
ARVALIS, Domaine du Chaumoy, F-18570 Le Subdray, France

E. Gourdain · F. Piraux
ARVALIS, Station expérimentale, F-91720 Boigneville, France

F. Brun
ACTA, 24 Chemin de Borde Rouge, Auzeville, CS52627,
F-31326 Castanet-Tolosan cedex, France

considered. The results indicate that climatic years, wheat varieties, and crop management, in this decreasing order, define disease epidemic risk in the multiple wheat-diseases pathosystem.

Keywords Categorical data · Risk factor · Multiple pathosystem · Crop management · Agricultural extensification · Sustainable agriculture

Introduction

A large network of field experiments has been conducted over several years across France to identify combinations of wheat cultivars and management practices that would achieve the double objective of ecological sustainability and economic viability (Bouchard et al. 2008; Loyce et al. 2008, 2012). Several reports on this important project spearheaded by breeders who selected hardy winter wheat cultivars (Doussinault 1998; Brancourt-Hulmel et al. 2003) have dealt with crop yield (Loyce et al. 2008, 2012) and economic performances (Bouchard et al. 2008; Loyce et al. 2012).

Hardy winter wheat cultivars may be defined as (Doussinault 1998; Bouchard et al. 2008) wheat varieties that (1) have yield performances approaching those of the best high yielding wheat varieties at moderate levels of fertilizer inputs, (2) produce good quality grain (i.e., with a high protein content), and (3) carry multiple partial resistances to the main diseases of wheat. An important objective of this networked experimental project was therefore to assess the extent to which disease might be checked by some varieties under some field conditions: this is a key to chemical extensification, i.e., to a reduction in the use of pesticides. If, in a first stage, some wheat varieties grown at reduced levels of mineral fertilizer (especially nitrogen) and under minimal chemical protection were to lead to reduced disease levels, then such (variety x management) combinations were to be seen as entry points for chemical extensification. If, in a second stage, such combinations of wheat varieties and reduced chemical inputs were to lead to satisfactory yields of good grain quality, the double objective of ecological and economical sustainability would be within reach.

Several reports have indicated that this is the case (Félix et al. 2002; Bouchard et al. 2008; Loyce et al. 2008, 2012). An analysis based on a 2-year period (2001–2002) indicated that disease (yellow rust, brown

rust and septoria tritici blotch) intensity decreased with intensifying chemical inputs and with variety resistance scores (Loyce et al. 2008). Little information however is available to document multiple disease intensities, their variation and distribution, their possible combinations as disease syndromes, and the main factors determining multiple epidemic risks. The goal of this article is to provide information on the first stage of a chemical extensification path, that is, on the level of disease reduction achieved by hardy winter wheat varieties at specified intensification levels.

In this article, we provide quantitative information on the distribution frequencies of five different wheat diseases, and their variability over years. These diseases are: brown rust (*Puccinia triticina*), yellow rust (*Puccinia striiformis*), fusarium head blight (involving a complex of pathogens, including: *Fusarium graminearum*, *F. culmorum*, and *F. avenaceum*; Jennings et al. 2004; Nielsen et al. 2011), powdery mildew (*Blumeria graminis*), and septoria tritici blotch (*Zymoseptoria tritici*). A second objective is to assess the hypothesis of existence of disease syndromes, i.e., of disease associations (Willocquet et al. 2008), from the data set. A third objective is to document the effect of wheat varieties on diseases, and especially to determine if patterns of reactions exist in the several varieties tested in the experimental network. A fourth objective is to assess the strength of association between disease variables (including, if existing, disease syndromes) and risk factors, such as climatic years, wheat varieties, and crop management. A last overall objective of the work reported here is to assess the use of a generic approach to analyse large data sets that involve qualitative and quantitative information, and variable level of data precision, which has been used elsewhere (Savary et al. 1994; Avelino et al. 2004; Savary et al. 2006; Zhang et al. 2006).

Materials and methods

Network of experiments

A large set of winter wheat varieties were tested in France in a network of field experiments established at different locations (Fig. 1), where each variety was tested with four levels of crop management. These experiments were conducted for several years. Here, we report on the years 2003–2010, when a total of 45 wheat varieties were tested.

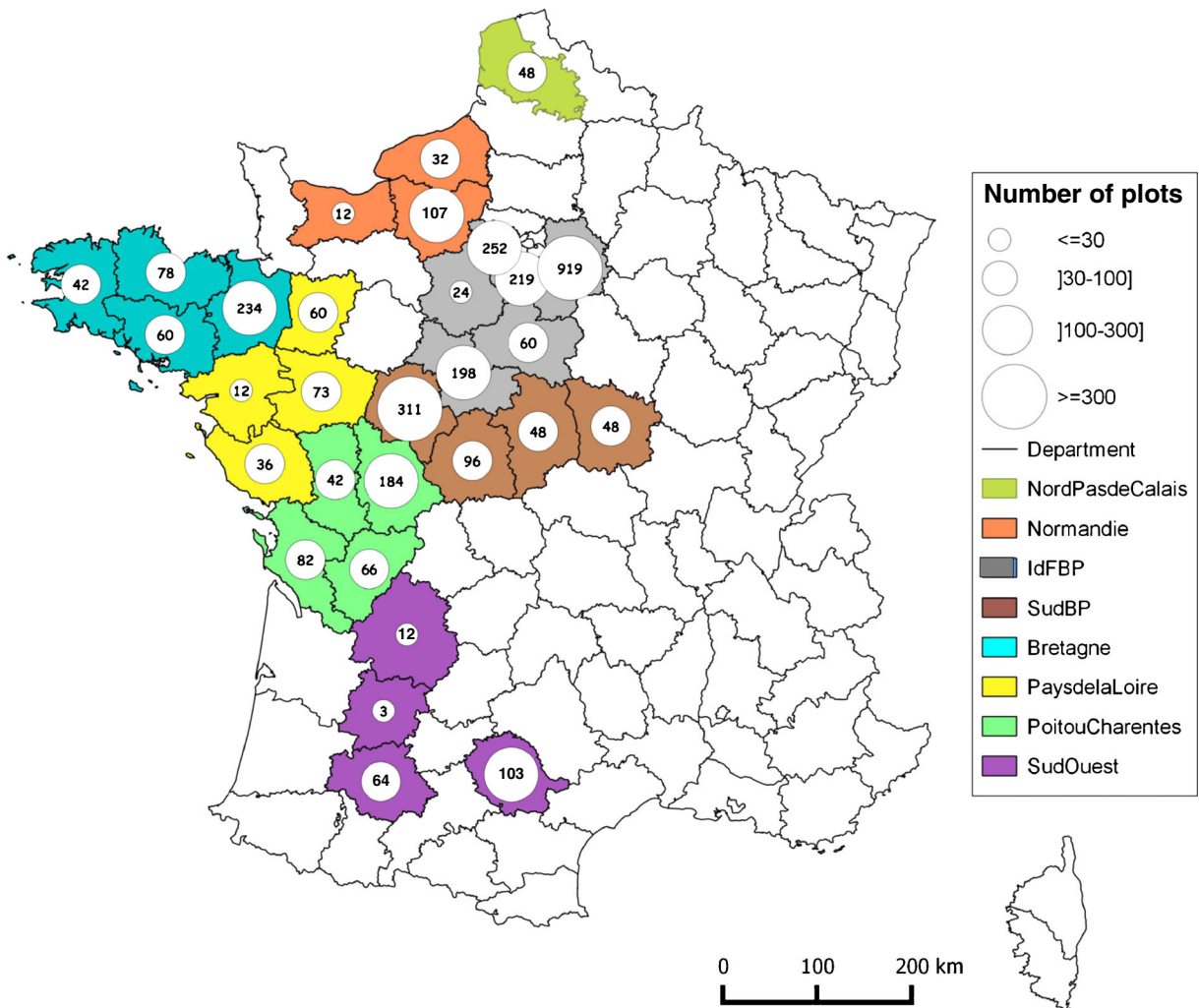


Fig. 1 Overview of the network of field experiments on hardy wheat in 2003–2010. The map shows the distribution of 3525 individual plots corresponding to 101 experiments conducted

during 2003–2010 across French Départements. Colours represent broad French regions where experiments took place

The levels of crop management in the network of experiments, CMGT1 to CMGT4, have been described in detail elsewhere (Rolland et al. 2003; Loyce et al. 2008, 2012; Bouchard et al. 2008). They consist in four levels of agricultural, primarily chemical, extensification, from a highly intensive CMGT1 to a much more extensive CMGT4. The reasoning of extensification and intensification levels is based on CMGT2, which represents current recommendations for high yield performances. The objectives of the different crop management levels can be summarized as follows.

- CMGT2 aims at local high attainable (i.e., un-injured) crop yields under near-optimal mineral

nutrition: CMGT2 thus entails high seeding rates (150–400 seeds.m⁻²; radiation interception at early crop growth maximized), high nitrogen fertilizer inputs (usually 160 kg.ha⁻¹ in three splits; attainable yield maximized), frequent fungicide applications (one to three applications; attainable yield achieved), and a growth regulator when required (one application at most; lodging prevented).

- CMGT1 aims at maximized yield: the nitrogen fertilizer inputs of CMGT2 are increased by 30 kg.ha⁻¹; fungicide application is systematic (two to three applications); and a growth regulator is applied once to twice.

- CMGT3 aims at a grain yield lower than CMGT2: the seeding rate is reduced by 40 %; nitrogen fertilizer is reduced by 30 kg.ha⁻¹; growth regulators are seldom applied; fungicide use is limited (one application at heading at half the recommended dose); and no growth regulator is used.
- CMGT4 aims at a reduction in usage of chemicals: the seeding rate is similar to CMGT3; nitrogen fertilizer is reduced by 60 kg.ha⁻¹ compared with CMGT2; no growth regulator applied; and no fungicides are used.

Since CMGT2 is based on site-specific recommendations, and because the definition of successive levels of crop management also is site-specific, the entire range of crop management levels may therefore vary, within ranges, from one experiment to another.

Experimental designs and field data collection

Experimental designs varied across locations and years. A typical design was a split-plot with crop management levels as main units and varieties as sub-units (Loyce et al. 2008), with four, sometimes three, replications. However, experiments were also conducted using randomized block or strip-plot designs. A large number of experiments also only involved CMGT2 and CMGT3 as management levels. The number of wheat varieties, too, varied depending on the considered experiment. We selected and included in the analysis reported here experiments which entailed (1) at least three different wheat varieties, (2) two management levels, and (3) at least three replications. The number of individual plots (i.e., combinations of variety x crop management level) therefore varied from experiment to experiment and year to year (Fig. 1). The individual plot size (wheat variety x crop management in one replication) was also variable among experiments, from 15 to 68 m².

A variable amount of information was collected in each experiment. All experiments reported yield performances estimated at a grain moisture content of 15 %. Disease measurements were also made, with a range of methods, including severity (proportion of leaf area infected) and incidence (proportion of infected leaves) of diseases of the foliage, and disease incidence of head disease. Disease assessments pertained to five diseases (Table 1): brown rust (BR), yellow rust (YR), fusarium head blight (FHB), powdery mildew (PM) and septoria tritici blotch (STB).

Data set

Experiments were conducted in the network at many locations nation-wide. Figure 1 shows the distribution of individual experimental plots accumulated over the considered period, with large numbers in the center-north, and center-west, but smaller in the west, and south-west of France. These areas account for the bulk of the wheat production in the country.

The experimental network generated a large data set. The experimental data was first sorted and selected using two criteria. First, only experiments where disease assessments had been performed were considered. And second, we retained experiments in which assessments had been made during a consistent range of crop development stages (early booting to early grain filling, i.e., development stage 45 to 85, Zadoks et al. 1974). The latter choice reflects the need to consider (1) comparable disease levels at similar crop development stage, and (2) disease levels that have been reliably assessed, since disease assessment on senescing leaves at the ripening stage is likely to be more inaccurate. Disease information was standardized throughout the entire data set, so that LR, YR, PM, and STB are expressed in disease severity (% leaf area infected), and FHB in disease incidence (% heads infected).

The resulting data set involves 101 experiments conducted during 8 successive years (2003–2010), corresponding to 3525 individual wheat plots.

Analytical strategy

The resulting data set involved a very large number of wheat varieties, some of which had been extensively tested in many experiments over a long fraction of the considered period of time, while others only appeared in a limited number of trials over few years. On the other hand, the data set involved five different wheat diseases, some of which were very frequently encountered (STB, BR), while others were documented in few trials and few individual plots in these trials.

Departing from former analyses on similar data (Bouchard et al. 2008; Loyce et al. 2008), we considered the individual experimental plot as a statistical unit, enabling multivariate approaches. A second decision was to recognize that the data had been collected over a long period of time, by different investigation groups with various emphases; thus categorizing the information (Savary et al. 1995) was considered a main avenue

Table 1 List of variables involved in the analysis of multiple disease data in a multi-site and multi-year evaluation of wheat varieties

Variable type	Variable	Symbol	Notes ^a	
Original disease intensities				
	Brown rust severity	BR	Fraction leaf area infected (%)	
	Yellow rust severity	YR	Fraction leaf area infected (%)	
	Fusarium head blight incidence	FHB	Fraction heads infected (%)	
	Powdery mildew severity	PM	Fraction leaf area infected (%)	
	Septoria tritici blotch severity	STB	Fraction leaf area infected (%)	
Crop management levels				
	Levels of crop management from CMGT4 to CMGT1	CMGT4, CMGT3, CMGT2, CMGT1	Increasing levels of fertilizer, fungicides, crop growth regulators use, and seeding rates, from the lowest, CMGT4, to the highest, CMGT1	
Disease intensities as binary categories				
	Binary categories of BR severity	BRBIN	Non-epidemics BR < 0.102 BRBIN = BR0 n = 2855	Epidemics BR ≥ 0.102 BRBIN = BR1 n = 670
	Binary categories of YR severity	YRBIN	YR < 0.076 YRBIN = YR0 n = 3470	YR ≥ 0.076 YRBIN = YR1 n = 55
	Binary categories of FHB incidence	FHBBIN	FHB < 0.663 FHBBIN = FHB0 n = 3058	FHB ≥ 0.663 FHBBIN = FHB1 n = 467
	Binary categories of PM severity	PMBIN	PM < 0.233 PMBIN = PM0 n = 3322	PM ≥ 0.233 PMBIN = PM1 n = 203
	Binary categories of STB severity	STBBIN	STB < 12.91 STBIN = STB0 n = 2527	STB ≥ 12.91 STBBIN = STB1 n = 998
Grouping metavariables				
	Clusters of individual experimental plots according to disease intensities as binary categories	CLUSDISEASE	Three clusters, X, Y, Z; Generated from hierarchical cluster analysis (Chi-square distance, Ward criterion; see text)	
	Clusters of wheat varieties according to untransformed disease intensities	VARGROUP	Three clusters, VARGROUP A, B, C; Generated from hierarchical cluster analysis of mean variety disease performances (Euclidean distance, Ward criterion; see text)	

^a n number of plots in the dataset

to reduce data noise, and stabilize very large variances. Third, while detail on the behaviour of each individual wheat variety, or characteristics of the five different diseases, or again, the specific disease responses to the different levels of intensification all are important questions, we focused on multiple disease responses (i.e., plant health) of wheat varieties. Therefore, approaches to assemble varieties in broad groups, and to identify disease responses as tractable overall plant health responses, were considered. Grouping methods for wheat varieties and multiple disease responses (i.e., injury profiles, or disease syndromes; Savary et al. 1994, 2006) were therefore sought. Fourth, each year of experiments and data collection corresponded to different climatic conditions. The year of data collection was therefore considered a variable of its own, corresponding to a climatic year. Fifth, the individual disease

responses, or the combined disease syndrome response, were addressed as the realization of given disease risks associated with diverse risk predictors, i.e., wheat varieties, crop management, and climatic years. A risk factor approach (Savary et al. 2011) to data analysis was therefore adopted.

Analytical steps

A first step in the analysis was to address the large variances in measured disease intensities, and devise suitable data transformation. This required consideration of the distribution frequencies of disease intensities, in view of their categorization (Savary et al. 1995).

A second step concerned the identification of plant health syndromes. This was achieved using (categorized) disease data pertaining to each of the

3525 plots, and considering each of them as one realization of one of several possible disease syndromes affecting the entire population. This step involved a hierarchical cluster analysis using the Ward criterion and the chi-square distance.

A third step considered the wheat varieties that had been tested, and their (multiple) disease(s) responses. The disease response of a given variety may be seen as a function of its genotype and the variable climatic year. This response can be further affected by crop management, especially because crop management entailed variable, sometimes high, use of fungicides that can suppress disease response. Ideally, this analysis should have been conducted in absence of any fungicide use, and might therefore have been conducted using the subset of individual plots of CMGT4. However, CMGT4 was represented by only 321 plots. Closer inspection of these 321 plots indicated that some varieties were very poorly represented or absent from this subset. This analysis was therefore conducted on plots belonging to the CMGT3 crop management level (1572 plots), where fungicide use is relatively low, where all five diseases were recorded, with comparatively lower variability thus enabling disease means and variances to represent each variety. Disease response of varieties was characterised by grouping varieties according to disease intensities observed in plots under CMGT3. To achieve this, a hierarchical cluster analysis was performed, based on the mean and standard deviations of the five diseases across all plots grown under CMGT3. The cluster analysis used the Euclidian distance and the Ward criterion (Lames and McCulloch 1990; Wilkinson et al. 2007).

A fourth step considered the level and significance of associations between the generated meta-variables (disease syndromes and variety groups) along with individual variables (individual disease levels and climatic years). These associations were assessed by conducting a series of chi-square tests on the corresponding contingency tables (Savary et al. 1995).

The nature and variability of the nominal crop management levels (CMGT1–4) was addressed in a fifth step, in order to better qualify the nature of this factor in the analyses. For this purpose, a principal component analysis (Hau and Kranz 1990; Lames and McCulloch 1990) was performed, involving the levels of chemicals applied on the wheat plots across years and sites.

A sixth step consisted in generating an overall picture of the multiple links between climatic years and disease

levels, varieties and variety groups, and crop management through a multiple correspondence analysis (Benzécri 1973; Greenacre 1984; Savary et al. 1995). This step was designed to bring together the intermediate results of the former steps.

A risk factor analysis involving logistic regressions (Harrell 2001; Esker et al. 2006; Savary et al. 2011) was followed in the last step. In this phase, the likelihood of occurrence of disease epidemics was considered the outcome of a series of predictors: climatic years, crop management, and variety groups.

Results

Distribution frequencies of disease intensities

The distribution frequencies (Fig. 2) of the five diseases were strongly skewed, with very large proportions of plots showing very little or no disease, and very small proportions with large disease intensities. There were also very strong variations in disease intensities over the successive years (Fig. 2). Brown rust (BR, Table 1) occurred in most years (except 2004), with a maximum level in 2007. Yellow rust (YR) occurred only in four consecutive years (2007–2010) in a total of 55 individual wheat plots, with severities in most cases equal to, or lower than, 2 % (medians in the range of 1 to 10 %). Fusarium head blight (FHB) also occurred unevenly, often with very large variances, as in 2008, when it reached maximum levels. Similarly, powdery mildew (PM) occurred in some years only, and at low levels (medians in the range of a few percent). By contrast, septoria tritici blotch (STB) occurred in all years, with medians often in the range of 10 % severity and with large variances as well.

Overall disease variation across varieties

The very large effect of wheat varieties on disease intensities is shown in Fig. 3. Overall differences in intensities among the five disease is indicated again, with moderate overall levels of brown rust and fusarium head blight, low levels of yellow rust and powdery mildew, and much higher levels of septoria tritici blotch (Fig. 3). Variability among wheat varieties in specific disease intensities suggests the existence of strong differences with respect to susceptibility to brown rust and possibly (given the low levels observed), to yellow rust

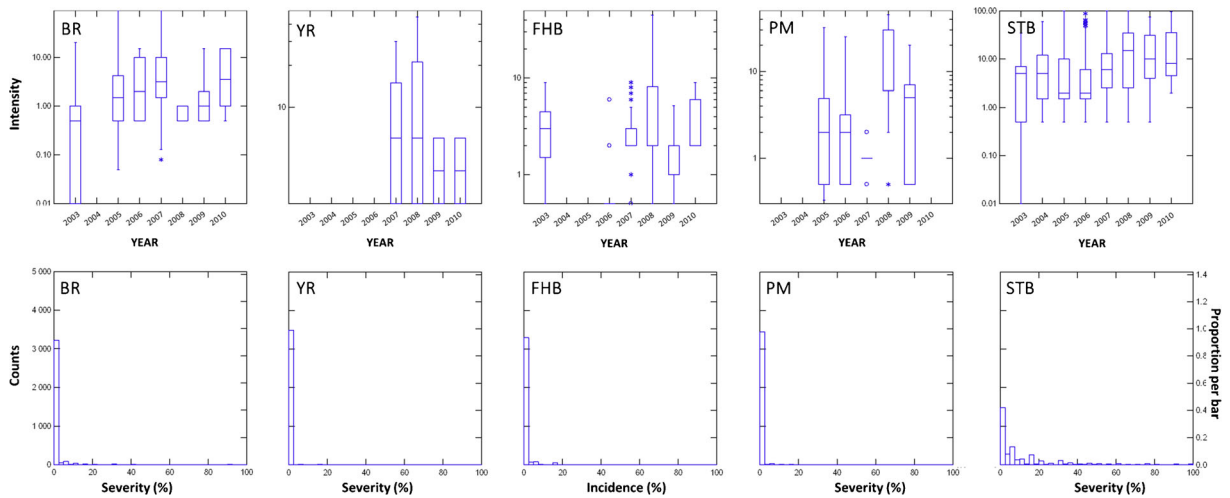


Fig. 2 Distribution frequencies and box-plots of individual disease distributions over years. Top: log-transformed variation in disease intensities (severities or incidences, see Table 1) over years. Abscissa: years, ordinates: log-transformed (base 10) disease intensities. Note the differences in scales of ordinates depending on diseases. Bottom:

distribution frequencies of untransformed disease intensities (disease severities: BR, YR, PM, and STB; or incidence: FHB) in 3525 wheat plots of a network of wheat experiments in France. Top and bottom: *BR*: brown rust, *YR*: yellow rust, *FHB*: fusarium head blight, *PM*: powdery mildew, and *STB*: septoria tritici blotch

and powdery mildew. The response of wheat varieties to FHB seems more homogeneous. Variation of STB intensity among varieties also suggests quantitative varietal differences.

Transformation of disease data in a binary form

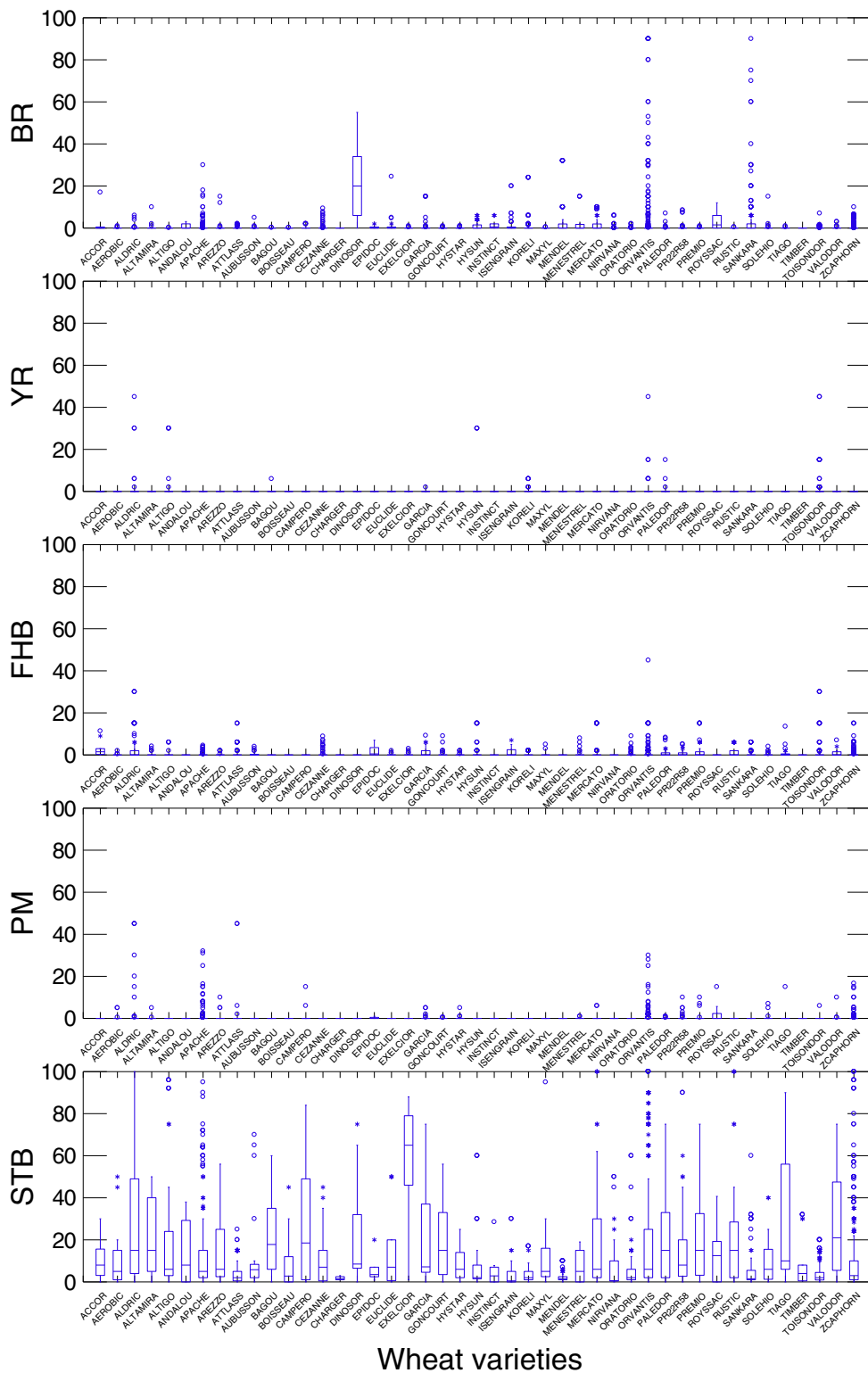
The observed distribution frequencies of disease intensities (Fig. 2) led to considering the disease-related information in a way similar to earlier studies conducted on large samples of monitored crop stands (Savary et al. 1994; Avelino et al. 2004, 2006; Zhang et al. 2006), with the use of categories to account for variability of quantitative measurements (Savary et al. 1995). In such large data sets, disease level data generally may be associated to two broad processes: disease establishment, and disease intensification in the host population (Lenné and Jeger 1994; Savary et al. 1994). Quantitative information pertaining to the observed level of disease can be very useful – reflecting conditions that may have favoured or hampered disease intensification –, but is dependent on the prior of disease establishment.

Examination of raw data indicated that large differences in disease assessment could occur from location to location and from year to year. This may be attributed to differences of emphasis from diverse observation teams, and differences in disease measurement itself (Reddy

et al. 2011). Further, the distribution frequencies of Fig. 2 indicate that the occurrence of any disease (except STB) at any level higher than 5 % severity (incidence in the case of FHB) is infrequent in the analysed data set. The decision was therefore reached to emphasize disease establishment over epidemic intensification, and thus to represent disease data in a binary form (De Wolf et al. 2003; Esker et al. 2006), 0 indicating absence or very low disease level, and 1 indicating any disease level above a given, low, threshold. The thresholds for conversion in a binary form were calculated as: $t = \text{mean} - 0.05 (s/\text{mean})$, where ‘mean’ is the arithmetic mean of disease intensity, and s is its standard deviation. The corresponding values (Table 1) of the thresholds were 0.102, 0.076, 0.663, 0.233, and 12.91 % for BR, YR, FHB, PM, and STB, respectively. Disease levels above or below these threshold were further referred to as ‘epidemics’ or ‘non-epidemics’ (Esker et al. 2006).

Disease syndromes: hierarchical cluster analysis on binarized disease data

Hierarchical cluster analysis on individual plots represented by their levels of disease intensities in a binary form led to the identification of three broad groups of injury profiles (Fig. 4). A large group X (2273 plots) corresponds to the most diverse occurrence of diseases. Only in this first group do YR, FHB, and PM epidemics



Wheat varieties

Fig 3 Box-plots of patterns of disease responses across wheat varieties. Each box-plot indicates a distribution of untransformed disease intensity across wheat varieties. Abscissa: wheat varieties,

ordinates: disease intensities (severities or incidences, see Table 1). BR: brown rust, YR: yellow rust, FHB: fusarium head blight, PM: powdery mildew, STB: septoria tritici blotch

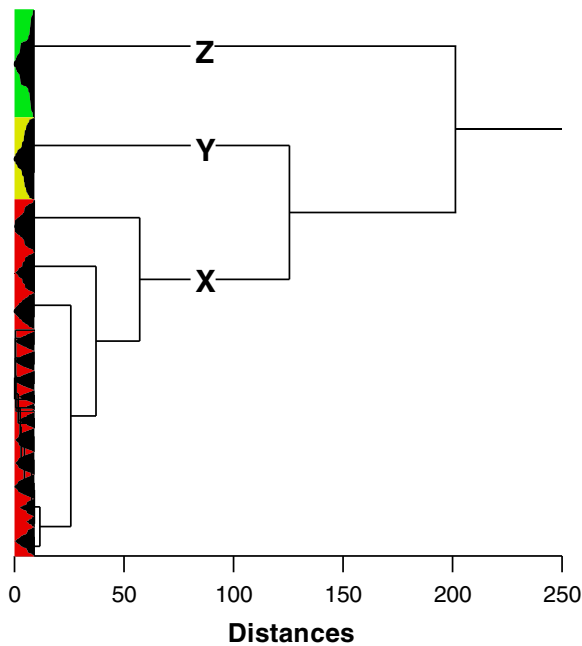


Fig. 4 Hierarchical cluster analysis of wheat plots using binary levels of disease intensities. The cluster tree results from a hierarchical cluster analysis of individual wheat field plots represented by disease levels in a binary form (BRBIN, YRBIN, FHBIN, PMBIN, STBIN, Table 1) using a Chi-square distance and a Ward criterion. Three groups, X, Y, Z, of individual wheat plots are indicated: X: possible BR epidemics; possible YR epidemics; frequent FHB epidemics; possible PM epidemics; possible STB epidemics ($n = 2273$ plots); Y: frequent BR epidemics; no YR epidemics; no FHB epidemics; no PM epidemics; no STB epidemics ($n = 543$ plots); Z: no BR epidemics; no YR epidemics; no FHB epidemics; no PM epidemics; frequent STB epidemics ($n = 709$ plots)

occur. The observed frequency of epidemic occurrence (i.e., binary disease level = 1) makes it possible for multiple epidemics to occur simultaneously in the same individual wheat plot. Group X is associated with possible brown rust epidemics (BRBIN = 1 in 279 cases), possible yellow rust epidemics (YRBIN = 1 in 55 cases), frequent FHB epidemics (FHBIN = 1 in 515 cases), possible PM epidemics (PMBIN = 1 in 215 cases), and possible STB epidemics (STBIN = 1 in 418 cases). Group Y (543 plots) is predominantly associated with the occurrence of brown rust (BR = 1 in 440 cases). Group Z (709 plots), by contrast, is predominantly associated with the occurrence of septoria tritici blotch epidemics (STB = 1 in 634 cases). The three groups therefore correspond to three distinct disease syndromes, the first with a likely combination of multiple wheat diseases, the second with brown rust, and the third with septoria tritici blotch.

Variety groups: hierarchical cluster analysis on untransformed mean disease data in CMGT3

Three wheat variety groups (VARGROUP-A -C) were identified from a hierarchical cluster analysis involving the arithmetic means and standard deviations of each of the five diseases (Fig. 5). From this classification based on variety performances under CMGT3 (i.e., limited pesticide use; Table 1), the three variety groups were characterized as follows:

- VARGROUP-A, consisting of 16 varieties, corresponds to moderate-low disease levels in general, except for YR (moderate levels);
- VARGROUP-B, with 12 varieties, corresponds to moderate-low BR, YR, and FHB levels, and moderate PM and STB levels;
- VARGROUP-C, with 17 varieties, corresponds to comparatively higher BR, YR, FHB, and STB levels, and moderate PM levels.

These differences among variety groups are further shown in Table 2, where the distribution frequencies of epidemics and non-epidemics in each variety group are tested in the entire data set (i.e., considering all four crop management levels). The resulting χ^2 values lead to rejection of the null hypothesis of independence of disease intensity (severity or incidence) across the three groups of varieties ($P < 0.05$) for the five considered diseases. Table 2 also indicates that variety groups are significantly ($P < 0.05$) associated with the occurrence of epidemics or non-epidemics for each of the five diseases. The independence of distribution of individual wheat plots among the three identified disease syndromes (X, Y, or Z) and the three variety groups (VARGROUP-A - B) is further rejected ($\chi^2 = 134.67$, $P < 0.001$).

Strength of variable associations: chi-square tests

Table 2 further provides an overview of the association of disease levels as binary categories (epidemics vs. non-epidemics) or of the occurrences of the three disease syndromes and successive climatic years. Chi-square values obtained indicate rejection ($P < 0.001$) of the null hypotheses of independence between any of the five wheat diseases (or disease syndrome,

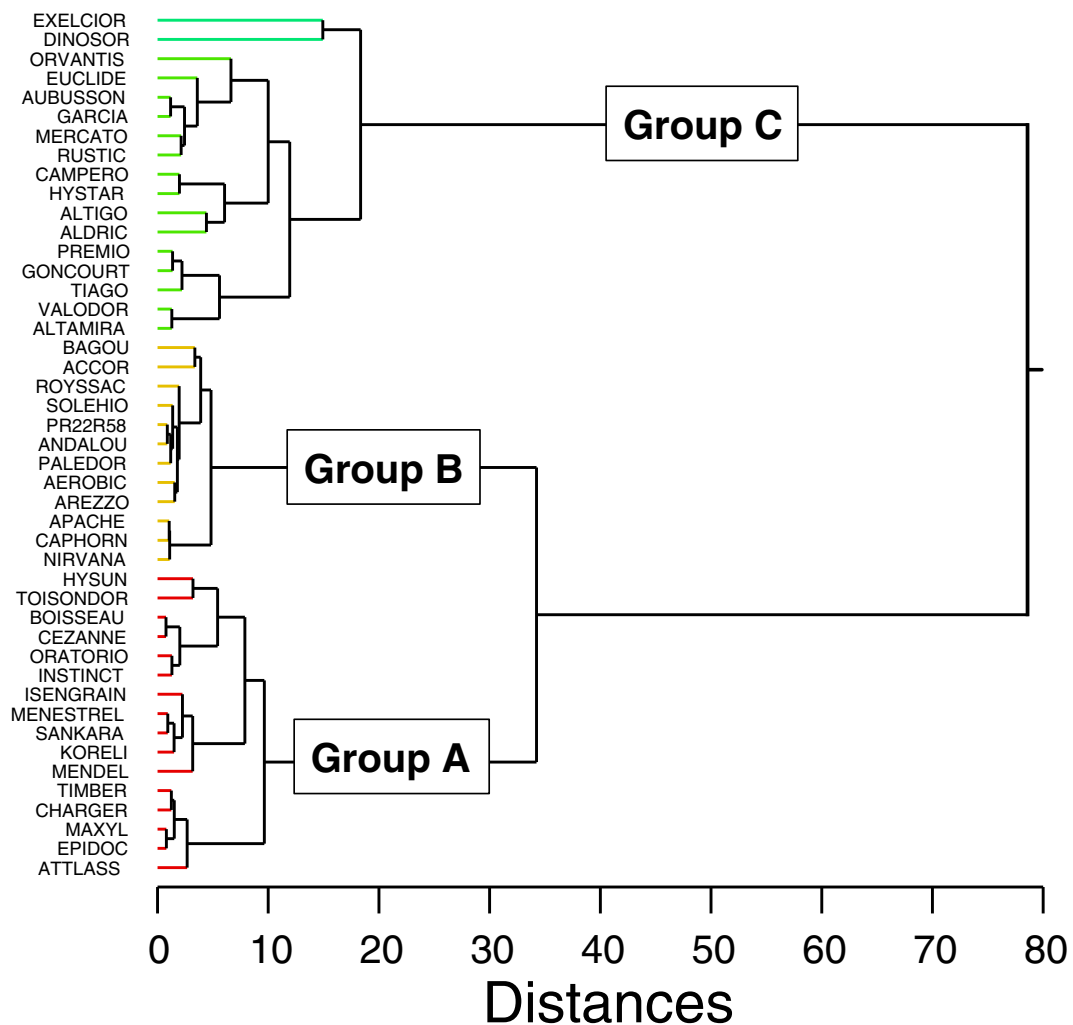


Fig. 5 Hierarchical cluster analysis of wheat varieties based on observed disease intensities. Cluster tree resulting from an analysis using a Euclidean distance and a Ward minimum variance criterion. The analysis was performed on a data set where each wheat

variety is represented by its mean (untransformed) disease severities and the associated standard deviations across all plots under CMGT3 (See Table 1). All five diseases (BR, YR, STB, PM, and FHB, Table 1) were involved in the analysis

CLUSDISEASE) and year (see also Fig. 2). Similarly, the use of binary disease data enables the testing of the independence of disease occurrences in the different crop management treatments, leading to the rejection of this hypothesis ($P < 0.001$) for all five diseases, as well as for disease syndromes. The series of chi-square values listed in Table 2 provide a numerical overview of the strength of relationships among qualitative (e.g., variety groups, crop management treatments), ordinal and qualitative (e.g., climatic years) variables and the observed responses expressed as binary (i.e., 0–1, or: non-epidemic vs. epidemic) disease responses. Overall, the association of climatic years with disease responses

is very large ($\chi^2 = 87.0\text{--}1402$), followed by the association of variety groups ($\chi^2 = 6.41\text{--}321$), while the associations of crop management with disease levels has smaller, but the least variable, chi-squares ($\chi^2 = 21.2\text{--}82.1$). These associations provide the basis of multiple correspondence analysis.

Patterns of epidemics and non-epidemics over years

The variability of multiple disease occurrences over years is further documented in Fig. 6, which displays the distribution frequencies of epidemics and non-epidemics over climatic years. The diagram indicates

Table 2 Chi square tests between levels of wheat diseases as binary categories, crop management, variety groups, and years

X-variable	Y-variable	Pearson Chi-square	df	P
YEAR	BRBIN	1401.81	7	< 0.001
	YRBIN	87.01	7	< 0.001
	FHBBIN	757.95	7	< 0.001
	PMBIN	203.21	7	< 0.001
	STBBIN	384.55	7	< 0.001
	CLUSDISEASE	1005.13	14	< 0.001
CMGT	BRBIN	31.87	3	< 0.001
	YRBIN	50.21	3	< 0.001
	FHBBIN	40.38	3	< 0.001
	PMBIN	21.22	3	< 0.001
	STBBIN	82.06	3	< 0.001
	CLUSDISEASE	32.48	6	< 0.001
CLUSDISEASE	BRBIN	1621.88	2	< 0.001
	YRBIN	30.06	2	< 0.001
	FHBBIN	289.59	2	< 0.001
	PMBIN	115.88	2	< 0.001
	STBBIN	1624.50	2	< 0.001
VARGROUP	BRBIN	6.41	2	0.041
	YRBIN	24.06	2	<0.001
	FHBBIN	24.45	2	< 0.001
	PMBIN	86.90	2	< 0.001
	STBBIN	321.14	2	< 0.001
	CLUSDISEASE	134.67	4	< 0.001

df degrees of freedom of the Chi-square test; *P* *P*-value associated with the Pearson's Chi-square test

the absolute as well as the relative frequencies of epidemic and non-epidemic events. Frequent brown rust epidemics (Fig. 6a) occurred during the studied period, especially in 2007, with additional cases in 2003, 2005, and 2009. Yellow rust epidemics (Fig. 6b) were quite rare, with some cases only in 2008. There were fusarium head blight epidemics (Fig. 6c) in 2003, followed by a long lapse, and a succession of cases in 2007, 2008, 2009, and 2010. Powdery mildew epidemics (Fig. 6d) were rather rare, except in 2005 and 2009, when some cases were observed. Septoria tritici blotch epidemics (Fig. 6e) occurred every year, especially in 2005 and 2007–2009 (nearly half the wheat plots grown in 2008 were considered as epidemic cases). The data structure shown in Fig. 6 provides the framework for further logistic regressions.

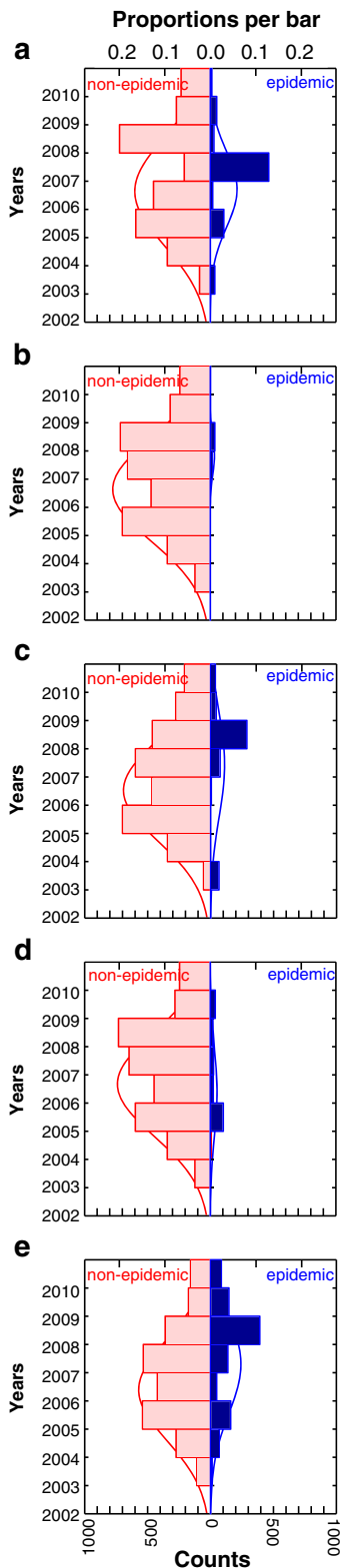
Variability within, and overlap between, crop management levels

The nature of the four crop management (CMGT, Table 1) treatments was investigated in a principal component analysis (PCA) of the levels of chemicals involved. In this case, each plot is seen as one realization of one of the four management levels. PCA led to the identification of two main axes accounting for 34.2 and 21.7 % of total variance, respectively (Fig. 7). The analysis shows the large extent of overlap among crop management treatments, indicated by the successive ellipses of confidence associated with CMGT1 to CMGT4 (Fig. 7, right). This analysis indicates that the nominal differences in crop management level do not always correspond to large differences in terms of frequency of chemical inputs to the wheat crop. Rather, one may interpret the successive crop management levels, CMGT4 to CMGT1, as a gradient of increasing chemical and seed inputs.

Mapping association between disease intensity, climatic year, and varieties: multiple correspondence analysis

Using the contingency tables leading to the chi-square tests listed in Table 2, a Burt table involving the successive years and the binarized levels of disease was built. A Burt table is a series of juxtaposed contingency tables, where, similarly to a correlation matrix, each pair of distributional associations is taken one after the other. In this case, the Burt table involves, in turn, the five diseases represented by their binarized levels, the eight considered years, and the three variety groups (VARGROUP-A, -B, and -C). This multiple contingency table generated a multiple correspondence analysis. The two first axes of the correspondence analysis (Table 3) accounted for 12.73 and 11.42 %, respectively, of the total inertia represented by the variables: BRBIN, YRBIN, FHBBIN, PMBIN, and STBBIN; years 2003 to 2010; and VARGROUP-A, VARGROUP-B, and VARGROUP-C.

This system of axes (Fig. 8) was then used to project additional variables, enabling to map the degree of linkage of these additional variables with the active variables that determine the system of axes. A first group of additional variables are the individual wheat varieties that had been involved in the network of experiments. Figure 8 provides further information pertaining to the variety groups (VARGROUP), which



◀ **Fig. 6** Distribution frequencies of epidemics and non-epidemics of wheat diseases over years. Epidemics and non-epidemics correspond to thresholds used for converting disease data in a binary form (Table 1). Abscissa: absolute counts of epidemic events (*bars*) and proportion of epidemic events as a normalized fraction of total events (*continuous lines*). Ordinates: successive years. Left: non-epidemics; right: epidemics. **a** Brown rust; **b** Yellow rust; **c** Fusarium head blight; **d** Powdery mildew; **e** septoria tritici blotch

had been determined separately, using the mean and standard deviation of disease levels in hierarchical cluster analysis (Fig. 5) on a subset of data consisting of plots from CMGT3. A second group of additional variables are the levels of crop management CMGT1–4.

The two first axes accounted for the inertia of the (binary) disease variables to varying degrees, as indicated by their squared cosines (Table 3). The two first axes therefore accounted well for the variability in BR (axis 1: 19.3 and axis 2: 60.2 %), FHB (24.1 and 10.9 %), and STB (37.0 and 4.5 %); they did not account well for the variability in YR (0.2 and 0.5 %) and PM (1.2 and 1.3 %). The two axes also accounted well for some year-variables: 2004 (1 % and 14 %), 2006 (0.4 % and 14 %), 2007 (23 % and 52 %), 2008 (42 % and 1.4 %). These years correspond to frequent epidemic cases of the different diseases shown in Fig. 6. The system of axes therefore provides a good overall representation of the variables, with a particular emphasis on brown rust, fusarium head blight, and septoria tritici blotch epidemics, as well as some of the years covered by the present analysis.

Variety groups were associated with variable squared cosines (Table 3). The first and second axes accounted for 12.3 and 0.01 %, 2.7 and 5.6 %, and 27.5 and 6.9 % of the inertia of VARGROUP-A, -B, and -C, respectively. Accumulated squared cosines therefore indicate that variety VARGROUP-C (34.4 %) is strongly associated with differing (binarized) disease levels, while VARGROUP-A (12.4 %) and VARGROUP-B (8.3 %) are associated with more uniform distributions of epidemics and non-epidemics. By contrast, crop management levels are associated with uniformly small squared cosines (0.02 to 0.69 % on the first axis, and 0.13 to 1.03 % on the second axis, for CMGT1, 2, 3, and 4), indicating that crop management levels were not in general associated with very large differences in epidemic and non-epidemic occurrence.

Figure 8 displays the positions of the centres of inertia of binarized disease levels, successive years,

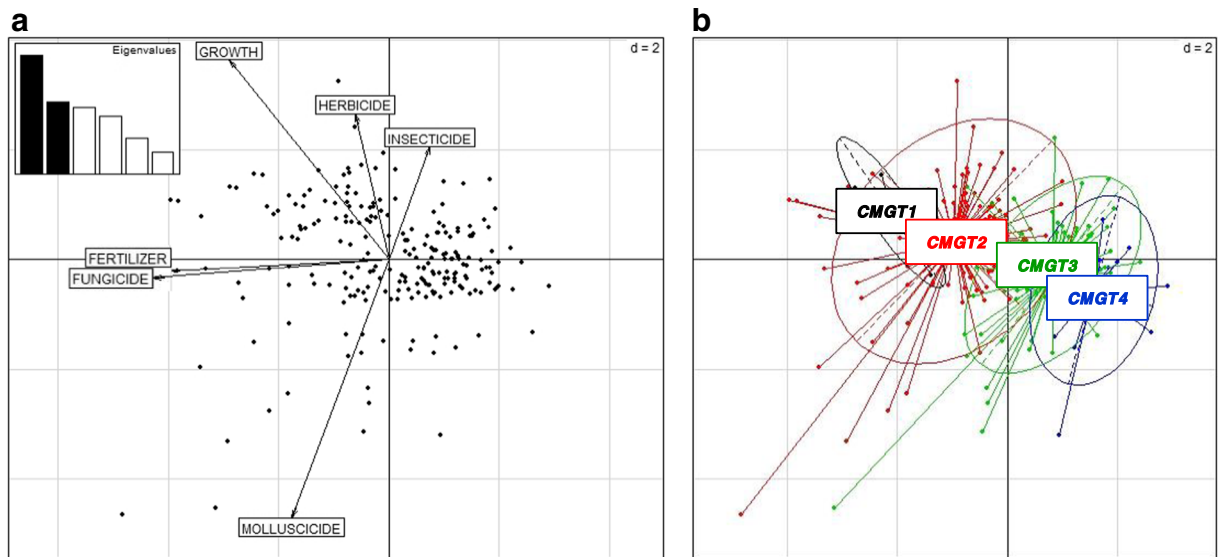


Fig. 7 Principal component analysis of the levels of chemical inputs involved in four levels of wheat crop management. **a** distribution of individual experimental plots involved in the analysis (*dots*) and projection of vectors involved in the analysis along its two first axes. The bar chart indicates successive eigenvalues.

Axes 1 and 2 account for 34.2 and 21.7 % of variance, respectively. Insert: proportion of variances accounted by the successive eigenvalues: 34.2, 21.7, 19.4, 13.6, 9.1, and 5 %. **b** projection of the four crop management levels on the system of axes and ellipses of confidence (95 %) of accurate positioning of individual plots

and variety groups. While VARGROUP-C appears strongly associated with occurrence of FHB and septoria tritici blotch, as well as (more weakly) yellow rust and powdery mildew, VARGROUP-A and -B are only weakly associated with powdery mildew and yellow rust occurrence. The occurrence of brown rust (BR1) plays an important role in the associations displayed in Fig. 8, being strongly associated with 1 year variable (2007). Occurrence of brown rust appears weakly associated with both variety groups A and C, but disconnected from variety group B. Crop management levels (CMGT1–4), being very close to the origin of axes, are not shown in Fig. 8.

Individual wheat varieties are also displayed in Fig. 8. The location of the different varieties belonging to a given variety group can be widely scattered about their respective centres of inertia. VARGROUP-C appears distinct, with its linkages with occurrences of four diseases (FHB1, STB1, BR1, and YR1). VARGROUP-A and B broadly overlap, being however distinguishable on the horizontal axis where varieties of group A appear to be quite apart from multiple disease occurrence. While brown rust occurrence appears very strongly associated with one variety of group C, the vector BR0-BR1 provides additional discrimination between varieties of group A

(which in general are closer to BR1, and away from BR0) and varieties of group B (which in general are away from BR1, and may be close to BR0).

Risk factor analysis: logistic regressions

Logistic regressions (Table 4) provide quantitative estimates of the contribution of climatic years, variety groups, and crop management to the likelihood of epidemic occurrence for each of the five diseases considered. In each of the five logistic regressions tested, climatic years (YEAR), variety groups (VARGROUP) and crop management (CMGT) are therefore considered as possible predictors of disease epidemics detected in the data set.

Depending on the considered disease, the logistic models addressed different numbers of cases, that is, the number of epidemics detected in the data set varied, and was in some cases small. In a total population of 3525 plots (Fig. 1), 670, 55, 467, 203, and 998 epidemics of brown rust, yellow rust, fusarium head blight, powdery mildew, and septoria tritici blotch, respectively, were considered. In spite of this, the logistic models converged for all five diseases, with significant likelihood ratios associated with the complete model tested. Each of the five logistic regressions tested correspond to

Table 3 Multiple correspondence analysis of levels of diseases as binary categories, climatic years, and wheat variety groups

Variables (classes) ^a	Inertia	Axis 1 (inertia: 12.73 %)				Axis 2 (inertia: 11.42 %)		
		Coordinate	Contribution ^b		Coordinate	Contribution ^b		
			To axis	Reciprocal		To axis	Reciprocal	
BR0	0.0136	0.21	0.021	0.193	-0.38	0.072	0.602	
BR1	0.0579	-0.91	0.088	0.193	1.60	0.305	0.602	
YR0	0.0011	-0.01	0.000	0.002	0.01	0.000	0.005	
YR1	0.0703	0.33	0.001	0.002	0.54	0.003	0.005	
FHB0	0.0095	-0.19	0.018	0.241	-0.13	0.009	0.109	
FHB1	0.0620	1.26	0.118	0.241	0.84	0.059	0.109	
PM0	0.0041	-0.03	0.000	0.013	-0.03	0.000	0.012	
PM1	0.0673	0.46	0.007	0.013	0.44	0.007	0.012	
STB0	0.0202	-0.38	0.059	0.370	-0.13	0.008	0.045	
STB1	0.0512	0.97	0.149	0.370	0.33	0.020	0.045	
Y2003	0.0690	-0.11	0.000	0.000	0.59	0.008	0.012	
Y2004	0.0645	-0.36	0.007	0.014	-1.14	0.080	0.141	
Y2005	0.0575	-0.39	0.017	0.037	-0.47	0.027	0.054	
Y2006	0.0619	-0.51	0.020	0.041	-0.96	0.077	0.141	
Y2007	0.0585	-1.02	0.105	0.229	1.53	0.265	0.517	
Y2008	0.0569	1.29	0.190	0.424	0.24	0.007	0.014	
Y2009	0.0652	0.63	0.020	0.038	0.01	0.000	0.000	
Y2010	0.0665	0.73	0.021	0.040	-0.22	0.002	0.004	
VARGROUP-A	0.0473	-0.49	0.046	0.123	-0.03	0.000	0.000	
VARGROUP-B	0.0467	-0.23	0.010	0.027	-0.33	0.023	0.056	
VARGROUP-C	0.0489	0.77	0.106	0.275	0.38	0.030	0.069	

^a Variables are categories (classes) of the original quantitative variables. Disease variables are disease levels as binary categories, i.e., non-epidemics (0) vs. epidemic (1); Successive years (Y2003-Y2010) are processed as categories; and three variety groups (VARGROUP-A -C) are considered

^b Contributions to axes are fractions of axis inertia (sum over contributions of variables to each axis = 1) that are accounted for by each individual variable (class); Reciprocal contributions are fractions of variable (class) inertia that are accounted for by successive axes (sum over axes of successive relative contributions for each variable =1)

areas under the receiver operating characteristic curves (ROC) that are larger than 0.5 and close to 1. This further indicates that the predictors considered account for a fair fraction of the variability of epidemics for the five diseases (Harrell 2001). Each parameter of the logistic models of Table 4 is documented by its estimate, its standard error and associated probability, and its odds ratio. The sign of the estimate indicates the direction of the parameter effect. Large odds ratios indicate a strong association of the predictor with epidemics, while odds ratios close to 0 indicate a strong association of the predictor with non-epidemics (Harrell 2001).

For example, in the case of brown rust, years 2003, 2005, 2007, and 2009 (year 2010 being the reference)

are positively, significantly, associated with the likelihood of epidemics. In particular, very large odds ratios were obtained for years 2003 and 2007, indicating their strong association with epidemics. VARGROUP-A and B (VARGROUP-C being the reference) are significantly associated with non-epidemics, and CMGT1–3 (CMGT4 being the reference) are also significantly associated with non-epidemics.

The other logistic regressions of Table 4 can be interpreted in a similar manner. In the case of yellow rust, climatic years 2007 and 2008, and variety group A, were significantly associated with epidemics, while crop management levels 2 and 3 were associated to non-epidemics.

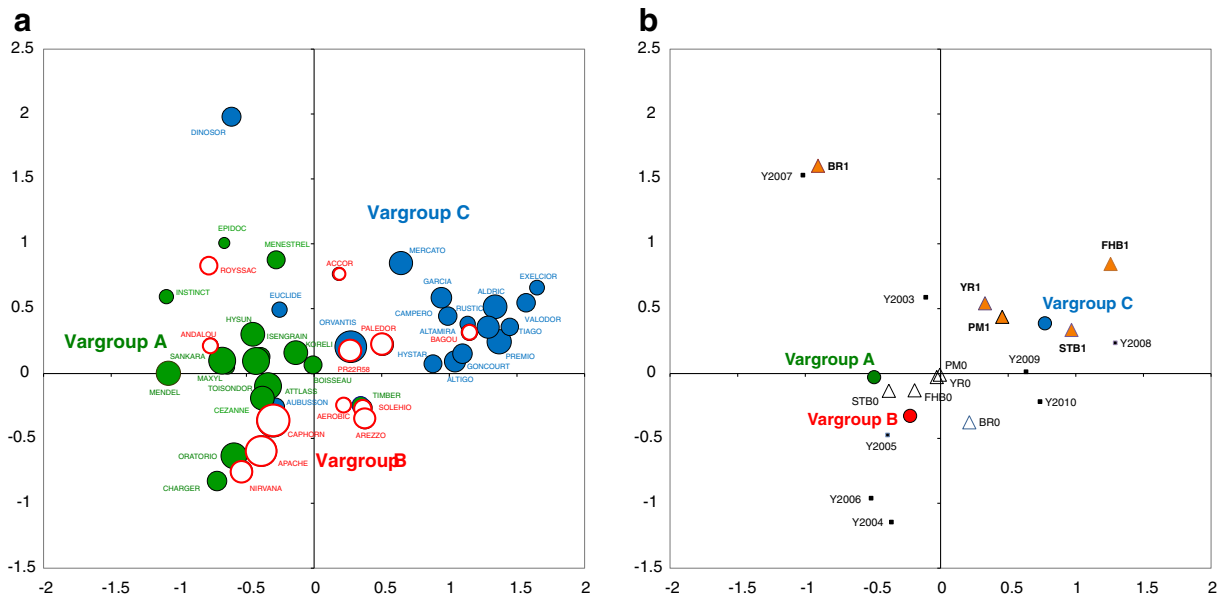


Fig. 8 Projection of wheat varieties characterized by their disease performances onto a multiple correspondence graph of the association of years and binary disease levels. Multiple correspondence analysis involves disease levels in a binary form (Table 1) and successive years (2003–2010) as active variables. The two first axes (accounting for 12.73 and 11.42 % of total inertia, respectively) are used. Wheat varieties are independent variables projected on the system of axes based on their levels of BR, YR, FHB, PM, and FHB in a binary form. **a** location of wheat varieties

on the graph. Colour of labels corresponds to variety groups (A, B, C; Fig. 5). The area of each *circle* is proportional to the number of wheat plots under a given variety. **b** locations of the centres of inertia of years, disease levels (1: epidemics, 0: non-epidemics), and centres of inertia of the projected variety groups (A, B, and C). Note that variety groups (VARGROUP A-C) were not derived from analysis of binary data on the entire data set, but from untransformed means and standard deviations of disease levels assessed in CMGT3 (see text and Fig. 5)

In the case of fusarium head blight, years 2003 and 2008 are significant predictors of epidemics, and 2006, 2007, and 2009 are significant predictors of non-epidemics; VARGROUP-A is a (weak, $P = 0.072$) predictor of non-epidemics, while crop management CMGT1–3 all are significant predictors of epidemics, i.e., CMGT1, CMGT2, and CMGT3 all are significantly associated with FHB occurrence.

Year 2010 could not be used as a reference for the powdery mildew model, since no epidemic occurred that year. Using year 2005 as a reference instead, years 2006–2009 are significant predictors of non-epidemics, VARGROUP-A is a predictor of non-epidemics, while CMGT2 is a predictors of non-epidemics.

Lastly, in the case of septoria tritici blotch, years 2003, 2004, and 2006 are significant predictors of non-epidemics and years 2008 and 2009 are significant predictors of epidemics; both VARGROUP-A and VARGROUP-B are significant predictors of non-epidemics, and CMGT2 is a predictor of non-epidemics.

Discussion

The analyses presented here contribute to understand the rich information generated by this network of experiments on hardy wheat varieties in France. We discuss first the successive results reported here, and then introduce further analyses to be reported.

Distribution frequencies and conversion of disease data in a binary form

Figure 2 illustrates a common finding of many large scale experiments, or of farmers’ field surveys (Savary et al. 1994, 1995): in four of the five wheat diseases considered in this study (BR, YR, FHB, PM), disease intensity (severity or incidence) distributions are very strongly skewed, with a very large majority of plots having no or very low disease intensity, while only a minority have variable disease intensity. The distribution frequency associated to STB severity shows a more complex structure, with three groups of individual

Table 4 Logistic regressions of occurrence of individual wheat disease epidemics with years, variety groups, and crop management levels

Disease	Likelihood ratio ^a		Area under ROC curve ^b	Parameter statistics by component (year, variety group, crop management)				
	LR	<i>P</i>		Parameter ^c	Estimate ^d	SE	<i>P</i>	Odds ratio
BR brown rust <i>n</i> = 670 epidemics	1383.779	<0.001	0.902	Constant	−1.925	0.317	<0.001	-
				VARGROUP-A	−0.272	0.147	0.064	0.762
				VARGROUP-B	−0.391	0.147	0.008	0.676
				CMGT1	−1.568	0.460	0.001	0.208
				CMGT2	−1.972	0.207	<0.001	0.139
				CMGT3	−1.143	0.194	<0.001	0.319
				Year 2003	2.984	0.415	<0.001	19.760
				Year 2004	NA ⁽¹⁾	NA ⁽¹⁾	NA ⁽¹⁾	NA ⁽¹⁾
				Year 2005	1.694	0.331	<0.001	5.442
				Year 2006	0.366	0.395	0.353	1.442
				Year 2007	4.556	0.336	<0.001	95.215
				Year 2008	−0.541	0.397	0.173	0.582
				Year 2009	1.442	0.355	<0.001	4.227
YR yellow rust <i>n</i> = 55 epidemics	152.943	<0.001	0.907	Constant	−3.719	0.732	<0.001	-
				VARGROUP-A	0.760	0.305	0.012	2.140
				VARGROUP-B	−1.217	0.631	0.054	0.296
				CMGT1	NA ⁽²⁾	NA ⁽²⁾	NA ⁽²⁾	NA ⁽²⁾
				CMGT2	−2.928	0.481	<0.001	0.054
				CMGT3	−1.475	0.351	<0.001	0.229
				Year 2003	NA ⁽²⁾	NA ⁽²⁾	NA ⁽²⁾	NA ⁽²⁾
				Year 2004	NA ⁽²⁾	NA ⁽²⁾	NA ⁽²⁾	NA ⁽²⁾
				Year 2005	NA ⁽²⁾	NA ⁽²⁾	NA ⁽²⁾	NA ⁽²⁾
				Year 2006	NA ⁽²⁾	NA ⁽²⁾	NA ⁽²⁾	NA ⁽²⁾
				Year 2007	1.694	0.803	0.035	5.442
				Year 2008	2.051	0.749	0.006	7.777
				Year 2009	0.228	1.016	0.823	1.256
FHB Fusarium head blight <i>n</i> = 467 epidemics	815.548	<0.001	0.864	Constant	−2.682	0.295	<0.001	-
				VARGROUP-A	−0.252	0.140	0.072	0.777
				VARGROUP-B	0.127	0.146	0.384	1.135
				CMGT1	1.278	0.465	0.006	3.589
				CMGT2	0.928	0.266	<0.001	2.530
				CMGT3	1.348	0.264	<0.001	3.850
				Year 2003	1.768	0.300	<0.001	5.861
				Year 2004	NA ⁽³⁾	NA ⁽³⁾	NA ⁽³⁾	NA ⁽³⁾
				Year 2005	NA ⁽³⁾	NA ⁽³⁾	NA ⁽³⁾	NA ⁽³⁾
				Year 2006	−2.944	0.488	<0.001	0.053
				Year 2007	−0.694	0.231	0.003	0.499
				Year 2008	1.179	0.200	<0.001	3.250
				Year 2009	−0.507	0.259	0.050	0.602
PM ^e powdery mildew <i>n</i> = 203 epidemics	253.141	<0.001	0.800	Constant	−1.087	0.281	<0.001	-
				VARGROUP-A	−2.491	0.361	<0.001	0.083
				VARGROUP-B	0.007	0.167	0.967	1.007
				CMGT1	NA ⁽⁴⁾	NA ⁽⁴⁾	NA ⁽⁴⁾	NA ⁽⁴⁾

Table 4 (continued)

Disease	Likelihood ratio ^a		Area under ROC curve ^b	Parameter statistics by component (year, variety group, crop management)					
	LR	<i>P</i>		Parameter ^c	Estimate ^d	SE	<i>P</i>	Odds ratio	
STB septoria tritici blotch <i>n</i> = 998 epidemics				CMGT2	−0.603	0.283	0.033	0.547	
				CMGT3	0.072	0.271	0.791	1.075	
				Year 2003	NA ⁽⁴⁾	NA ⁽⁴⁾	NA ⁽⁴⁾	NA ⁽⁴⁾	
				Year 2004	NA ⁽⁴⁾	NA ⁽⁴⁾	NA ⁽⁴⁾	NA ⁽⁴⁾	
				Year 2006	−1.129	0.241	<0.001	0.323	
				Year 2007	−1.907	0.282	<0.001	0.148	
				Year 2008	−1.817	0.259	<0.001	0.162	
				Year 2009	−0.526	0.212	0.013	0.591	
				Year 2010	NA ⁽⁴⁾	NA ⁽⁴⁾	NA ⁽⁴⁾	NA ⁽⁴⁾	
		715.794	<0.001	0.772	Constant	0.097	0.175	0.581	-
					VARGROUP-A	−1.729	0.117	<0.001	0.178
					VARGROUP-B	−0.365	0.098	<0.001	0.694
					CMGT1	−0.176	0.357	0.622	0.838
					CMGT2	−0.949	0.147	<0.001	0.387
					CMGT3	−0.271	0.144	0.060	0.762
					Year 2003	−0.931	0.366	0.011	0.394
					Year 2004	−0.379	0.207	0.067	0.685
					Year 2005	−0.283	0.173	0.103	0.753
					Year 2006	−1.054	0.212	<0.001	0.349
					Year 2007	−0.294	0.179	0.099	0.745
				Year 2008	1.212	0.166	<0.001	3.360	
				Year 2009	0.506	0.185	0.006	1.659	

The reference disease levels to all logistic regressions are non-epidemics (BR0, YR0, FHB0, PM0, STB0; See Table 1 for threshold values)

^a Sample size: *n* = 3525 plots;

^b Area under the receiver operating characteristic curve

^c Parameters: Years 2003 to 2010 (Year 2010 as reference, except for PM: 2005); variety groups VARGROUP-A and VARGROUP-B (VARGROUP-C as reference); crop management CMGT1, CMGT2, and CMGT3 (CMGT4 as reference)

^d NA⁽¹⁾: no brown rust epidemic under the considered combination of predictors; NA⁽²⁾: no yellow rust epidemic under the considered combination of predictors; NA⁽³⁾: no fusarium head blight epidemic under the considered combination of predictors; NA⁽⁴⁾: no powdery mildew epidemic under the considered combination of predictors

^e The reference year for PM is 2005, as no PM epidemic was recorded in 2010

wheat plots: a first large group with no or very low STB severity, a third one, quite small, with fairly high disease severity, and a small intermediate group where disease severity is average. This distributional pattern has been documented in different pathosystems (Savary et al. 1994; Avelino et al. 2004; Zhang et al. 2006), and has been associated with successive stages of disease dynamics, from establishment to intensification (Savary et al. 1995), at the plot level. Such variable distribution frequencies represent challenges in data analysis, which can be addressed through data categorization (Savary

et al. 1995), or, as here, through data conversion in a binary form (Esker et al. 2006). The transformation of quantitative, continuous data into categories has been discussed, especially in plant pathology. Data categorization entails the loss of some information, but enables useful analyses. Direct benefits from the use of categories in large, multi-year, data sets involving a number of field observers, such as the material of this analysis, are to standardize information across the data set (years, locations, and observers), and to account for observational errors (Savary et al. 1995).

Categorization of data also concerns the experimental years, which are in effect converted into, and processed as, categorical variables, so that ‘climatic years’ become predictors in the course of the analysis. The nature of these ‘climatic years’, in terms of actual climate, e.g., the temperatures and rainfalls that had occurred during the growing season, will be discussed in a companion article.

The very large effect of wheat varieties on disease levels is partly illustrated by Fig. 3. For some diseases such as BR, YR, and PM, these effects seem apparent (with most varieties showing very little or any disease at all), whereas they remain unclear for other diseases, such as FHB and STB (with all varieties showing some disease, albeit at varying levels).

Cluster analyses

Hierarchical cluster analysis of crop health data at the plot level (Fig. 4) yields a clear structure in disease syndromes, with three groups: dominated by STB (group Z), dominated by BR (group Y), and diverse, with fairly high levels of FHB, combined with possible occurrence of other diseases (group X). This analysis therefore suggests that STB and BR might, respectively, be considered key-stone species dominating specific wheat disease syndromes. We are not aware of such finding in the literature on wheat pathology, and this deserves further investigation. The overall result of Fig. 4, with the characterization of crop health syndromes in the case of wheat, concurs with similar results obtained in a range of very diverse pathosystems (Savary et al. 2006), further confirming the value of the concept.

A hierarchical cluster analysis of variety responses performed on a subset of data (CMGT3, i.e., reduced level of chemical intensification, including limited pesticide use) is shown in Fig. 5. Three groups, VARGROUP-C, -B, and -A, are generated, corresponding to 17, 12, and 16 wheat varieties, respectively. VARGROUP-C corresponds to moderate BR, high FHB, low-moderate PM, high STB; whereas VARGROUP-B corresponds to moderate-low BR, low YR, low FHB, low PM, and moderate STB; while VARGROUP-A corresponds to moderate-low BR, low-moderate YR, moderate FHB, low PM, lower STB. The characteristics of the groups generated in the present study quite differ from a similar analysis reported by Loyce et al. (2008), which contrasted three

groups: (a) susceptible to BR and STB; (b) susceptible to YR and moderately resistant to BR and STB; and (c) moderately susceptible to STB, and resistant to BR and YR. Three main reasons explain the differences between the two analyses: (1) the analysis reported by Loyce et al. (2008) involved data from regular varietal field trials that were specifically designed to assess susceptibility, whereas the analysis of Fig. 5 reflects measurements in a network of field experiments where input levels (including fungicides; see Fig. 7, CMGT3), varied; (2) the analysis reported by Loyce et al. (2008) concerns data from 2001 to 2002, i.e. was conducted before the beginning of the experimental period (2003–2010) reported here; and (3) the analysis discussed in the present study involves two diseases (powdery mildew and fusarium head blight), which were not considered by Loyce et al. (2008), in addition to brown rust, yellow rust and septoria tritici blotch. Our results do suggest varieties that experienced lower disease levels throughout the period; this is the case for brown rust, yellow rust, and fusarium head blight in VARGROUP-B; and for powdery mildew and septoria tritici blotch in VARGROUP-A. The result concerning BR is consistent with available information on wheat resistances. Results concerning YR and PM are to be considered cautiously, the overall disease levels during the period having been often so low. The results concerning FHB, on the other hand, may lead to useful hypotheses with respect to host plant resistances regarding this disease, since the levels of disease have been fairly high in some site x year combinations.

Levels of crop management

Principal component analysis on the amounts (frequency) of chemical inputs (Fig. 7) leads to considering the successive levels of crop management, CMGT1 to 4, much more as transitions towards a reduction of chemicals (fertilizer, pesticides, growth regulators) rather than pre-set, fixed levels. While these levels of crop management may very well correspond to clearly categorized yield targets, they certainly have been associated with progressively variable levels of disease control. This further strengthens the approach followed here (through categorization of disease levels, multivariate analyses, and logistic regression), since a linear model, where crop management would be seen as a fixed effect (Loyce et al. 2008), would not have been appropriate.

Associations between factors and diseases

Table 2 documents the strength of associations between factors (climatic years, crop management, variety groups) and binarized disease levels. The association of diseases with years is usually very strong (chi-square values ranging from 87 to 1402, Table 2), indicating the very strong effect of climatic years on occurrence of epidemics. The strength of disease associations with variety groups comes next, with variable chi-square values ranging from 6.41 to 321. The association of disease levels with crop management level comes last, with comparatively smaller but significant chi-square values (21 to 82). Tests involving disease syndromes (CLUSDISEASE) provide a summary of the magnitude of the association of wheat diseases with factors, from very large with years (1005), to large with variety groups (134), and smaller with crop management (32).

Plant disease epidemics may be considered in two broad phases, (1) disease establishment in the host population, and (2) disease intensification (Lenné and Jeger 1994). Conversion of disease data in a binary form leads the analysis to collapse these two phases into one, and to consider plant disease in one dimension only, which was referred to as ‘epidemic’, as opposed to ‘non-epidemic’ when disease does not occur, or when epidemic fails to intensify. The occurrence of favourable climatic conditions throughout a cropping season (Year) and the deployment of a susceptible variety (VARGROUP) enables both disease establishment and intensification, thus explaining why chi-square values associated with Year and VARGROUP (Table 2) can be so large. By contrast, crop management generally only affects disease intensification, conditional to disease establishment; thus the comparatively smaller chi-square values associated with CMGT. This explains why climatic years, and why variety groups, correspond to chi-square values (Table 2) that are, respectively, one or two orders of magnitude, and one order of magnitude, larger than values corresponding to crop management. Another element of explanation of the low effect of crop management on disease is that chemical intensification from CMGT4 to CMGT1 corresponds to pesticide intensification, which is expected to reduce disease intensity, and nitrogen intensification, which in general favours biotrophic pathogen such as rusts and powdery mildew (e.g., de Wit 1992), as well as STB (e.g., Leitch and Jenkins 1995). The shifts in crop management

components therefore have opposite effects on disease, which partly cancel each other.

Multiple correspondence analysis

Multiple correspondence analysis (Fig. 8) generates a synthesis of paired associations using a chi-square metric (chi-square tests, Table 2), with three domains. A first domain is that of regular occurrence of FHB, STB, PM, and YR epidemics and of fairly frequent BR epidemics, where varieties of VARGROUP-C predominate. A second domain corresponds to possible PM, YR, and BR epidemics, associated with VARGROUP-A. A third domain is associated with infrequent disease epidemics in general, and is associated with varieties of VARGROUP-B. As in earlier analysis, the positioning of PM and YR is difficult, owing to the relative low frequency of epidemics of these two diseases. The two axes of Fig. 8 only account for very small fractions of the inertias of CMGT1, CMGT2, CMGT3, and CMGT4, respectively. The four crop management treatments are located at the origin of axes, and thus could not be represented in Fig. 8. This suggests that climatic years, first, varieties, second, and crop management, last, contributed to the occurrence of wheat disease epidemics, individually or collectively. This concurs with the examination of chi-square values.

Epidemics and non-epidemics, logistic regressions, and risk factors

The overall result of conversion of disease intensities in a binary form is shown in Fig. 6. Distributions of epidemics vs. non-epidemics indicate differing patterns depending on diseases, with several years of frequent epidemics in the case of BR, FHB, and STB, and a few years when few epidemics occurred in the case of YR and PM. Year 2007, when a pathogen population shift took place across France (Goyeau and Lannou 2011; Goyeau et al. 2012) and presumably a strong BR epidemic occurred in western Europe (Berry et al. 2010), is visible in the binarized BR data (Fig. 6a), illustrating that the conversion of disease data in a binary form retained important features of recorded epidemics. Figure 6 suggests that while predictions of STB epidemics may be rather straightforward, that of BR and FHB is more challenging, and that of YR, could be very hard, based on the data set used. This is in line with the frequency of epidemics of the various diseases.

Logistic regressions confirm this. Quantifying the relative role of factors in determining the likelihood of epidemics is the objective of logistic regression of epidemic occurrences. Logistic regression enables the development of models where the occurrence of epidemics can be related to a number of variables (VARGROUP, CMGT, and YEAR). These variables are addressed in logistic regression as risk factors, represented by a parameter and the associated odds ratio (Harrell 2001). The models summarized in Table 4 indicate that, irrespective of the disease, a given climatic year always can be a significant predictor of epidemics. This is also true for VARGROUP and for CMGT. This series of analyses illustrates the value of a risk factor approach (Savary et al. 2011), where the occurrence of disease epidemics is considered as the outcome of a series of predictors, climatic years, crop management, and variety groups.

A risk factor approach can provide insights in the main elements influencing crop health in large data sets. Such an analysis for instance enabled to characterize and hierarchize risk factors associated with production situation in rice from a large survey data set in several hundreds of Asian farmers' fields (Savary et al. 2011). In this risk factor analysis of rice health in Asia, production situation components were further decomposed, indicating (1) their respective importance with respect to rice crop health, and (2) that the individual effect of any production situation component was one or two orders of magnitude smaller than that of any production situation considered as a whole (Savary et al. 2011). Table 4 provides an analogous information for wheat in France, with the effects of climatic years, variety groups, and crop management levels. The analysis reported here cannot enter the lower hierarchy of components of crop management (fertilizer inputs, pesticide use, seeding rate, growth regulators), because these were involved in each experiment as locally-defined and fixed combinations. Nevertheless, the five models derived for each disease provide an overview of the risk factors associated with occurrence of epidemics of wheat diseases. Beyond the analysis of this particular data set, this approach might well be worth considering for the strategic management of wheat diseases in Europe.

Conclusion

Better documentation of disease levels is a necessary first step to further improve disease management,

including in the important case of winter wheat in France and in Europe where pesticide use must be reduced, and the potential of genetic diversity fully enhanced. This article focused on disease levels, their combinations, and their determinants, in a network of wheat field experiments involving a range of climatic years, crop management, and varieties in France.

The different analyses indicate that climatic years, wheat varieties, and crop management, in this decreasing order, define disease risk in the multiple wheat-diseases pathosystem in this network of experiments. The comparatively lower effect of CMGT on diseases may be related to three groups of reasons. A first reason is that CMGT effects were assessed on disease information which had been simplified into a binary form, epidemic vs. non-epidemic. As a result, analyses did not address the intensification of epidemics, which can strongly be influenced by crop management. A second reason is the confounded pesticide and fertilizer effects across CMGT levels. A third reason is that only a few management components are included in the data set used. Management components with potentially important effects on diseases such as crop rotation, crop establishment date, or landscape diversity (Palti 1981; Wiese 1982; Zadoks 1993; Savary 2014) were not considered here.

Among several other questions, the nature of climatic years requires further investigation, which will be addressed in a companion article. Questions pertaining to the succession of effect of crop management and varieties, first on disease levels, and second on yield performances, further require new analyses that account for the very large variability observed in these data. This can be achieved through formal meta-analyses, which also will be reported in a companion article.

Acknowledgments This research was supported partly by PEBiP – “Analyse stratégique des relations Pratiques - Environnement - Bioagresseurs - Pertes de récoltes”, funded by the French ministry of agriculture and fisheries. We also thank the Blé Rustiques Network (INRA, ARVALIS, Chambres d'Agriculture, and CIVAM) for making data available.

References

- Avelino, J., Willocquet, L., & Savary, S. (2004). Effects of crop management patterns on coffee rust epidemics. *Plant Pathology*, 53, 541–547.
- Avelino, J., Zelaya, H., Merlo, A., Pineda, A., Ordonez, M., & Savary, S. (2006). The intensity of a coffee rust epidemic is

- dependent on production situations. *Ecological Modelling*, 197, 431–447.
- Benzécri, J. P. (1973). *L'Analyse des Données, Tome 2* (632 p). Paris: L'Analyse des Correspondances. Dunod.
- Berry, P. M., Kindred, D. R., Olesen, J. E., Jorgensen, L. N., & Paveley, N. D. (2010). Quantifying the effect of interactions between disease control, nitrogen supply and land use change on the greenhouse gas emissions associated with wheat production. *Plant Pathology*, 59, 753–763.
- Bouchard, C., Bernicot, M. H., Félix, I., Guérin, O., Loyce, C., Omon, B., et al. (2008). Associer des itinéraires techniques de niveaux d'intrants variés à des variétés rustiques de blé tendre: évaluation économique, environnementale et énergétique. *Courrier de l'Environnement*, 55, 53–77.
- Brancourt-Hulmel, M., Doussinault, G., Lecomte, S., Bérard, P., Le Buane, B., & Trotter, M. (2003). Genetic improvement of agronomic traits of winter wheat cultivars released in France from 1946 to 1992. *Crop Science*, 43, 37–45.
- De Wit, C. T. (1992). Resource use efficiency in agriculture. *Agricultural Systems*, 40, 125–151.
- De Wolf, E. D., Madden, L. V., & Lipps, P. E. (2003). Risk assessment models for wheat Fusarium head blight epidemics based on within season weather data. *Phytopathology*, 93, 428–435.
- Doussinault, G. (1998). La sélection pour des variétés adaptées à une agriculture durable. *Dossiers Environnement*, 16, 39–44.
- Esker, P. E., Harri, J., Dixon, P., & Nutter Jr., F. W. (2006). Comparison of models for forecasting of Stewart's disease of corn in Iowa. *Plant Disease*, 90, 1353–1357.
- Félix, I., Loyce, C., Bouchard, C., Meynard, J. M., Bernicot, M. H., Rolland, B., et al. (2002). Associer des variétés rustiques à des niveaux d'intrants réduits: intérêts économiques et perspectives agronomiques. *Perspectives Agricoles*, 279, 30–35.
- Goyeau, H., & Lannou, C. (2011). Specific resistance to leaf rust expressed at the seedling stage in cultivars grown in France from 1983 to 2007. *Euphytica*, 178, 45–62.
- Goyeau, H., Berder, J., Czerepak, C., Lanen, C., & Lannou, C. (2012). Low diversity and fast evolution in the population of *Puccinia triticina* causing durum wheat leaf rust in France from 1999 to 2009, as revealed by an adapted differential set. *Plant Pathology*, 61, 761–772.
- Greenacre, M. J. (1984). *Theory and Applications of Correspondence Analysis*. London: Academic Press.
- Harrell Jr., F. E. (2001). *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer Verlag.
- Hau, B., & Kranz, J. (1990). Mathematics and statistics for analysis in epidemiology. In J. Kranz (Ed.), *Epidemics of plant diseases. Mathematical analysis and modeling* (Second ed.). Berlin, Heidelberg, New York: Springer Verlag.
- Jennings, P., Coates, M. E., Walsh, K., Turner, J. A., & Nicholson, P. (2004). Determination of deoxynivalenol- and nivalenol-producing chemotypes of *Fusarium graminearum* isolated from wheat crops in England and Wales. *Plant Pathology*, 53, 643–652.
- Lames, F. C., & McCulloch, C. E. (1990). Multivariate analysis in ecology and systematics: panacea or Pandora's box? *Annual Review of Ecology and Systematics*, 21, 129–166.
- Leitch, M. H., & Jenkins, P. D. (1995). Influence of nitrogen on the development of Septoria epidemics in winter. *Journal of Agricultural Science*, 124, 361–368.
- Lenné, J. M., & Jeger, M. J. (1994). Evaluation of plant pathogens in complex ecosystems. In J. P. Blakeman & B. Williamson (Eds.), *Ecology of plant pathogens* (pp. 63–77). Wallingford: CABI.
- Loyce, C., Meynard, J. M., Bouchard, C., Rolland, B., Lonnet, P., Bataillon, P., et al. (2008). Interaction between cultivar and crop management effects on winter wheat diseases, lodging, and yield. *Crop Protection*, 27, 1131–1142.
- Loyce, C., Meynard, J. M., Bouchard, C., Rolland, B., Lonnet, P., Bataillon, P., et al. (2012). Growing winter wheat cultivars under different management intensities in France: a multicriteria assessment based on economic, energetic and environmental indicators. *Field Crops Research*, 125, 167–178.
- Nielsen, L. K., Jensen, J. D., Nielsen, G. C., Jensen, J. E., Spliid, N. H., Thomsen, I. K., et al. (2011). Fusarium head blight of cereals in Denmark: Species complex and related mycotoxins. *Phytopathology*, 101, 960–969.
- Palti, J. (1981). *Cultural Practices and Infectious Crop Diseases*. Berlin, Heidelberg, New York: Springer Verlag.
- Reddy, C. S., Laha, G. S., Prasad, M. S., Krishnaveni, D., Castilla, N. P., Nelson, A., et al. (2011). Characterizing multiple linkages between individual diseases, crop health syndromes, germplasm deployment, and rice production situations in India. *Field Crops Research*, 120, 241–253.
- Rolland, B., Bouchard, C., Loyce, C., Meynard, J.-M., Guyomard, H., Lonnet, P., et al. (2003). Des itinéraires techniques à bas niveaux d'intrants pour des variétés rustiques de blé tendre: une alternative pour concilier économie et environnement. *Courrier de l'environnement de l'INRA*, 49, 47–62.
- Savary, S. (2014). The roots of crop health: cropping practices and disease management. *Food Security*, 6, 819–831.
- Savary, S., Elazegui, F. A., Moody, K., & Teng, P. S. (1994). Characterization of rice cropping practices and multiple pest systems in the Philippines. *Agricultural Systems*, 46, 385–408.
- Savary, S., Madden, L. V., Zadoks, J. C., & Klein-Gebbinck, H. W. (1995). Use of categorical information and correspondence analysis in plant disease epidemiology. In *Advances in Botanical Research incorporating Advances in Plant Pathology*, 21, 213–240. London: Academic Press Ltd..
- Savary, S., Teng, P. S., Willocquet, L., & Nutter Jr., F. W. (2006). Quantification and modeling of crop losses, a review of purposes. *Annual Review of Phytopathology*, 44, 89–112.
- Savary, S., Mila, A., Willocquet, L., Esker, P. D., Carisse, O., & McRoberts, N. (2011). Risk factors for crop health under global change and agricultural shifts: a framework of analyses using rice in tropical and subtropical Asia as a model. *Phytopathology*, 101, 696–709.
- Wiese, M. V. (1982). Crop management by comprehensive appraisal of yield determining variables. *Annual Review of Phytopathology*, 20, 419–432.
- Wilkinson, L., Engelman, E., Corter, J., & Coward, M. (2007). Cluster analysis, pp. 65–124: SYSTAT Statistics 12, Vol 1. San Jose, CA.

- Willoquet, L., Aubertot, J. N., Lebard, S., Robert, C., Lannou, C., & Savary, S. (2008). WHEATPEST, a production situation - based simulation model of yield losses caused by multiple injuries for wheat in Europe. *Field Crops Research*, *107*, 12–28.
- Zadoks, J. C. (1993). Cultural methods. In: Zadoks, J.C. (Ed), pp. 163–170: Modern crop protection: development and perspectives. Wageningen Pers, Wageningen.
- Zadoks, J. C., Chang, T. T., & Konzak, C. F. (1974). A decimal code for the growth stages of cereals. *Weed Research*, *14*, 415–421.
- Zhang, X. Y., Loyce, C., Meynard, J. M., & Savary, S. (2006). Characterization of multiple disease systems and cultivar susceptibilities for the analysis of yield losses in winter wheat. *Crop Protection*, *25*, 1013–1023.