Validation du POD

Contents

A l'issue du recueil d'expertise et de la modélisation, la question se posait de la validation de Mildium.

6.1 Simuler Mildium pour valider le recueil de connaissances

Dans un premier temps, il a fallu clarifier les objectifs et le type de validation que l'on pouvait effectuer sur ce modèle. C'est à cette occasion que l'usage de modèles biotechniques a été définitivement exclu de ma démarche. En effet, l'opportunité de réaliser un modèle biotechnique du multi-pathosystèmes *E. Necator + P. Viticola + V. Vinifera* a été analysée. L'usage d'un modèle biotechnique pour la validation du POD aurait eu l'avantage de permettre l'évaluation de la qualité agronomique du POD, notamment d'en étudier la sensibilité au scénario climatique. L'unique modèle biotechnique candidat a été développé au sein du projet ADD-Vin (Deola et al., 2007; Bazoche et al., 2008). Il s'agit du modèle « bio-économique » déjà évoqué

Le recours à l'expertise pour développer les stratégies nous est imposé par manque de connaissances scientifiques. Il en va de même pour le modèle « bio-économique » qui a été construit sur base bibliographique avec un paramétrage expert. Les résultats ne donnent accès qu'à des indications qualitatives.

Une hypothétique tentative de validation du POD Mildium avec ce modèle aurait donc consisté à confronter une expertise à une autre expertise. Il m'a donc paru inopérant de tenter ce type de validation. Ceci n'invalide pas l'intérêt d'un travail de simulation sur Mildium interfacé au modèle bio-économique à titre exploratoire.

L'analyse a permis de conclure que l'objectif de la validation du modèle était de montrer que le modèle simule l'expertise, c'est à dire qu'il est possible de reproduire les décisions prises par les experts au cours des expérimentations menées avant l'établissement du « Mildium formel ». Nous avons donc conçu et mis en œuvre une méthodologie de validation dans cet objectif.

Le troisième article (« *Working paper* »[a]) qui constitue le corps de ce chapitre présente mon travail de validation de la conformité des décisions prises par le modèle « Mildium formel » (Mf) noté GrapeMilDeWS avec celles prises par les experts mettant en œuvre Mildium noté « Mildium expert » (Mx) en Français et original expert GrapeMilDeWS (*oeGrapeMilDeWS*) en Anglais. Les premiers résultats agronomiques sont également présentés afin de montrer l'efficacité pratique d'un POD.

6.2 Matériel et Méthodes

La section 6.4.2 présente d'une part le dispositif expérimental bordelais et d'autre part la méthodologie de validation du modèle qui se décompose de la manière suivante :

- construction des scénarios d'entrée du simulateur
- stratégies de simulation en l'absence de modèle biotechnique
- les méthodes de comparaisons et d'analyse des différences « simulations *vs* expérimentations ».

a. décliné le 09/10/2008 par l'un des éditeurs d'*Agricultural Systems* comme hors du domaine de la revue, sans évaluation par des juges arbitres

NB. les données des scénarios de la simulation ne sont pas données dans l'article mais en annexe E.

6.3 Présentation de l'article

L'article est structuré de manière traditionnelle avec une partie « matériels et méthodes » et une partie consacrée aux « résultats ». Dans chacune de ces parties, on présente dans un premier temps, les expérimentations de Mildium au champ qui ont été menées à Bordeaux sur quatre parcelles pendant les saisons 2005 et 2006 (protocole expérimental section 6.4.2.1 ; performance agronomique section 6.4.3.1).

Remarque : *Afin d'ajuster les seuils de décision des différents indicateurs parcellaires (M, ILM, O, Og), chaque parcelle a fait l'objet d'une hypothèse alternative au cours de chacune des deux saisons étudiées (parcelle coupée en deux). Au final, il y a 15 historiques de décisions disponibles (une parcelle n'ayant pas été coupée en deux en 2005).*

Le second temps de chaque partie présente la validation du modèle formel. En section 6.4.2.2, on montre comment, par simulation, on cherche à reproduire les décisions prises par les experts avec le modèle Mf. On reconstruit les scénarios climatiques et les scénarios d'évènements auxquels les experts ont été soumis pendant les deux saisons étudiées à partir des données conservées dans les historiques d'expérimentations. Les données pour ces scénarios sont disponibles en annexe E. La dernière sous-partie de « matériels et méthodes » précise comment est évaluée la similitude des décisions entre les simulations et les expérimentations. Cette comparaison se fait en trois temps. On compare d'abord le nombre de traitements ordonnés. On compare ensuite les décisions en terme de choix et de dates (date d'observation, décision d'application d'un traitement optionnel et date de cette décision). Enfin, on observe au sein des séquences réelles et des séquences simulées, la distribution des écarts relatifs entre deux décisions (par exemple le temps entre la première évaluation et la date de décision du traitement anti-mildiou à l'étape 1), voir section 6.4.3.2 & section 6.4.3.3.

Chacune de ces trois analyses permet de tirer des enseignements sur la qualité du recueil de connaissances et de la modélisation (85% de décisions [b] sont identiques avec cependant des différences temporelles) mais également des enseignements sur la manière dont les experts ont mené leurs expérimentations qui renseigne sur certaines pratiques non exprimées pendant le recueil de connaissances (section 6.4.3.4).

La discussion reprend les résultats de la validation et aborde les points suivants : la façon dont les experts ont su mobiliser leurs connaissances scientifiques pour concevoir Mildium ; l'intérêt qu'il y a à gérer deux maladies dans une même stratégie ; les difficultés liées aux observations, particulièrement en ce qui concerne le suivi de la phénologie ; et dernier point de la discussion, l'acceptabilité de la stratégie dans une perspective de transfert.

b. Il y a 15 situations de décision dans Mildium qui sont comparées selon 16 scénarios soit au total 240 décisions simulées vs 240 décisions expérimentales.

6.4 Article « GrapeMilDeWS (Part.2) Expérimentation d'un POD PIC contre le Mildiou et l'Oïdium de la vigne »

"Working paper" décliné le 09/10/2008 par l'un des éditeurs d'*Agricultural Systems* comme hors du domaine de la revue, sans évaluation par des juges arbitres .

GrapeMilDeWS (part.2) experimenting an integrated pest management (IPM) Decision Process against grapevine powdery and downy mildews

Bertrand Léger[$*] and Olivier Naud[*] Michel Clerjeau[•] Véronique Bellon-Maurel[*] Laurent Deliére[$] Philippe Cartolaro[$] Lionel Delbac[$]
[*]Cemagref - UMR ITAP - BP 5095 34196 Montpellier Cedex 5
[$]INRA - UMR Santé Végétale -BP 81 33883 Villenave d'Ornon Cedex
[•]ENITA Bordeaux - 1, cours du Général de Gaulle CS 40201 33175 Gradignan Cedex

Abstract
GrapeMilDeWS is an expert based strategy for the integrated pest management of the vineyard (*Vitis vinifera*), at plot scale, against two cryptogamic diseases: powdery and downy mildews (*Erysiphe necator* and *Plasmopara viticola* resp.). It aims at reducing pesticides use on low epidemics years while always securing production output. GrapeMilDeWS was experimented for two years on 4 plots (≈ 0.5ha each) on 2 sites near Bordeaux. The results show that GrapeMilDeWS yields to about 50% reduction in the number of fungicides treatments and satisfactory quality and quantities at harvest. A formal and exhaustive Decision Workflow System (DeWS) model of GrapeMilDeWS was elicited from the expert designers (phytopathologists and crop protection engineers). The model is detailed in Part.1 as we consider this model as the best available description for transferring GrapeMilDeWS' knowledge. This assumption had to be assessed by checking that the model makes decisions consistent with the experts' ones during experiments. Such assessment is provided here, together with results about performance of GrapeMilDeWS. The model is shown highly consistent with experts' decisions (85% identical decisions). Yet, there are some temporal discrepancies which are analysed and discussed.

6.4.1 Introduction

With only 3% of the land use in France, viticulture is responsible for 20% of the pesticide treatments (Aubertot et al., 2005). According to (ASK, 2000) powdery mildew and downy mildew (respectively caused by *Erysiphe necator* and *Plasmopara viticola*) are considered by French vine growers as the two most important diseases, leading to the highest number of treatments. The common preventive crop protection strategies are indeed efficient. Yet, on most years, a number of these applications are unnecessary. Therefore, it is desirable for a sustainable viticulture to find means to avoid these.

We advocated in GrapeMilDeWS (part.1) (Léger et al., 2008a), that, by monitoring epidemics at the plot level, together with some local bioclimatic information about downy mildew, it is possible to manage crop protection over the whole season and to apply treatments only when required. Based on crop protection expertise, a team of grapevine pathologist experts designed a first decision workflow, which was a collection of guidelines and tables, to implement these principles. The team experimented its design, which

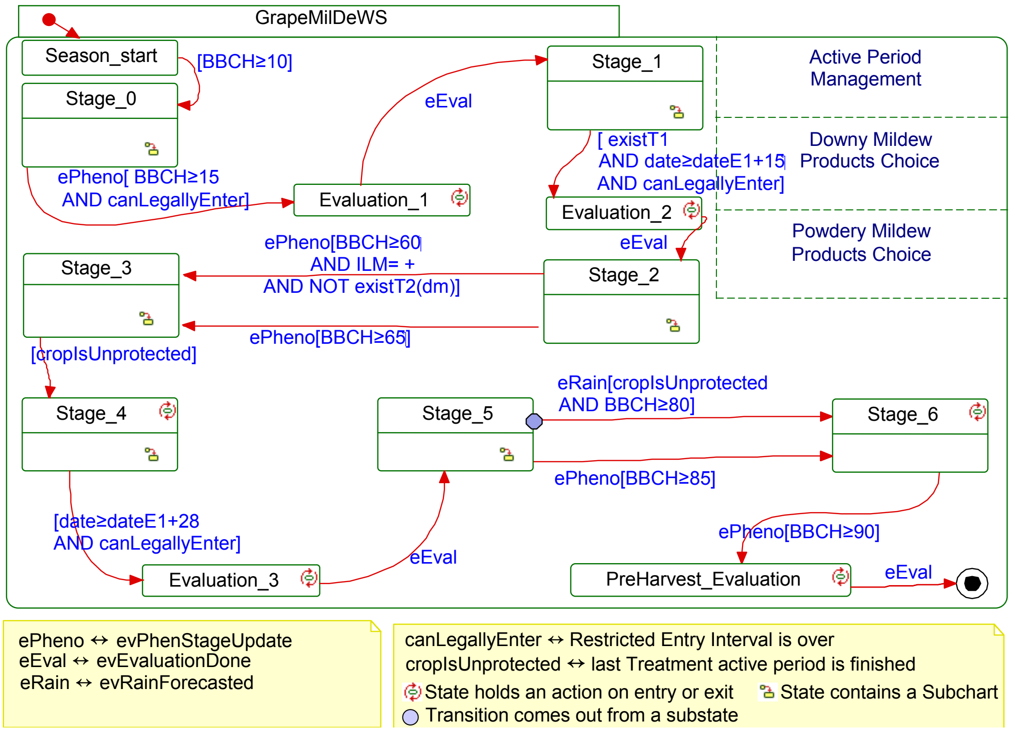


Figure 6.1: GrapeMilDeWS main process view

we will hereafter refer to as *oeGrapeMilDeWS* (for original expert GrapeMilDeWS), for two years. In the meantime, the whole expertise was elicited as a formal model in Statechart(Léger and Naud, 2007). Statechart (Harel, 1987; Harel and Kugler, 2004) is a formal language which is both graphical and mathematically sound (i.e. it can be compiled). Such a formal model was meant to: (i) to clarify the design; (ii) as a transfer tool to communicate the experts' ideas with little ambiguity and (iii) as a simulator for computer virtual experiments.

Theoretical details of the GrapeMilDeWS model (for Grapevine powdery and downy Mildew Decision Workflow System) are presented in part.1. This model is based on a succession of decision stages where decision variables, updated from plot diseases observations, are combined with phenological stage and climatic data to decide the applications of fungicides and to specify their timing. The observations in the field are represented in the model as evaluation states which interleave with the decision stages. Figure 6.1 gives the general organisation of the decision workflow.

In this second article, we show that GrapeMilDeWS provides an efficient crop protection with a reduced number of fungicide treatments. We also show that the formalised GrapeMilDeWS contains the necessary information to reproduce the experts' decisions.

First, the experimental setup and results of two years (2005 and 2006) of field experiments are presented. Second, these results are discussed in details through a confrontation of the decisions made by the experts during the experiments and the ones provided by the formal model elicited from the expert knowledge. Finally, we discuss our results and question the transferability of GrapeMilDeWS decision procedure.

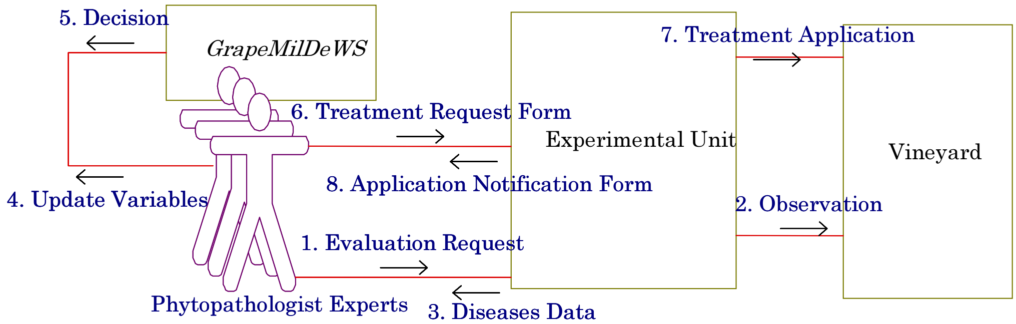


Figure 6.2: *oeGrapeMilDeWS'* communication flow

6.4.2 Materials and Methods

This first section is organised as follows: first, we describe the experimental protocol designed to assess *oeGrapeMilDeWS* as an efficient crop protection strategy; then we present our methodology for the validation of the elicited formal model (GrapeMilDeWS).

6.4.2.1 Relevance of the experts' *oeGrapeMilDeWS* as a crop protection strategy through a field experiment approach.

Experimental setup The first design of *oeGrapeMilDeWS* was informal, in the sense that the original documentation for guidelines was insufficient to capture all the knowledge the experts actually use while experimenting (Léger and Naud, 2007). All experiments presented here were conducted with this informal decision process description. *oeGrapeMilDeWS* was tested in 2005 and 2006 on four plots in two experimental vineyards of the French national institute of agronomical research (INRA), near Bordeaux.

The first experimental site is located in the village of Latresne south east of the city of Bordeaux. It produces a "premier cote de Bordeaux" registered designation of origin (RDO) wine. The second site is in the city of Villenave d'Ornon (south of Bordeaux). It produces a "Pessac Léognan" RDO wine. In each case, two plots were selected, with two cultivars: Merlot and Cabernet Sauvignon.

The two Latresne (Lat.) plots were planted in 1997. Each one has an area of 0.28ha and a density of 5050 stocks/ha. Two trellis systems are used: half the stocks is managed with double Guyot and the other half uses double cordon de Royat.

On the Villenave d'Ornon (V.O.) site, both plots have double Guyot trellis systems. There, Merlot (M) and Cabernet Sauvignon (CS) were planted respectively in 1991 and 1993, at a density of 6250 stocks/ha. The respective areas are 0.25ha and 0.35ha. [c]

Two types of actors were involved into the experimental implementation of *oeGrapeMilDeWS* : (i) the phytopathologist experts who made the decisions and (ii) the INRA experimental unit staff who executed the decisions. Figure 6.2 summarizes the communication between these actors.

The decision variables defined in *oeGrapeMilDeWS'* are described in GrapeMilDeWS part.1, section 3.3. Three variables account for epidemic pressure at the plot scale: M for downy mildew on leaves, O and Og for powdery mildew respectively on leaves and clusters. Their values result from a sampling protocol which is given hereafter. ILM

c. Map of the experimental sites on the internet: http://tinyurl.com/2zpgpo.

accounts for downy mildew pressure at a bioclimatic region scale (about 20km radius around the plot). Rain forecasts are also used in the decision process.

Using these variables, the experts decided the opportunity of a fungicide application. The decisions were then transmitted as treatment orders to the experimental unit staff in charge of the two vineyards. Traceability was ensured by the use of a formalised procedure using "*Treatment Request Forms*" (TRF) transmitted for orders and "*Application Notification Forms*" (ANF) sent in return after the applications were carried out.

A TRF was composed of a target disease, a proposed product (or class of products) and an application period. The "application period" may be expressed as a phenological stage, a time interval (usually a week) or the date of a forecasted rainfall as a deadline. The experimental unit then adjusted the exact application date according to its operational constraints (availability of the human resources as well as mechanical ones, regulations, weather and soil conditions).

	Powdery mildew			Downy mildew		
Assessment	-	+	++	-	+	++
E0 (M)	-	-	-	0%	> 0%	-
E1 (M-O)	0-2%	2-10%	> 10%	0%	0-10%	> 10%
E2 (M-O)	0-20%	-	> 20%	0-10%	10-50%	> 50%
E3 (M-Og)	0-20%	> 20%	-	0-10% or $E3-E2=0$	10-50% and $E3-E2>0$	> 50% and $E3-E2>0$

Table 6.1: Thresholds used in the 2005 & 2006 experiments for each evaluation and disease.

The evaluations of the level of diseases were done using a 10% regular sampling of the plot's stocks. All evaluations w.r.t. downy mildew consisted in counting the number of infected leaves per stock. For powdery mildew, the observations were done on a sub sample of leaves or bunches, taken from the stocks of the sample: (i) in the first evaluation, 4 leaves were chosen at the shoot base; (ii) in the second evaluation, 6 leaves were sampled in the fruiting zone; (iii) in the third evaluation, 5 bunches per stock were observed. For each disease three classes of contamination were built (thresholds given in Table 6.1.)

The decisions and actions are compiled in a matrix M_e with the decisions (i.e. treatment requests) and actions (i.e. treatments and evaluations) in columns and the vintage/plot individuals in rows. The matrix data are the date of the Decisions/Actions (D/A) expressed in day count since January the $1st$ (day cardinality). Each disease is individuated for treatment requests and treatments. The columns are ordered according to the sequence of evaluations and stages given by the guidelines (see x-axis in figure 6.6a). In order to further study the time lags in the decision process, we derive the matrix $M_{e_{relative}}$ from M_e, as the relative time interval between the relevant D/A (i.e. between two columns).

Criteria for crop protection performance analysis *oeGrapeMilDeWS'* performances are estimated through three criteria:

– severity at harvest time: (i) severity of the powdery mildew epidemic on bunches and (ii) for both diseases, severity on the leaves.

- production level of the plot
- maturity of the berries at harvest (probable alcoholic content, total acidity) estimated on a sample of 200 berries per plot.

oeGrapeMilDeWS is considered to have provided a satisfactory crop protection if the latter two criteria are within the RDO's requirements, guarantying that the product keeps its quality label. As for the first criterion, crop protection is satisfactory if less than 5% severity is found on the bunches.

Note that, for both years, each plot was cut in halves (with the exception of V.O. CS in 2005). This was to assess, at a given time, the need and efficiency of an optional treatment and thus evaluate the level of the decision thresholds. At the end of both seasons, no statistical difference between the 2 half-plots was ever found. This allowed to choose the thresholds that led to the lowest number of treatments. Only the results with lowest number of treatment are presented in section 6.4.3.1.

6.4.2.2 Running Simulations of the GrapeMilDeWS elicited model

In a second step, this section will consider GrapeMilDeWS w.r.t. actual decisions made during the experiments. We examine hereafter if the elicited Statechart model, GrapeMilDeWS, could be used by someone else than its designers (i.e. the phytopathologist experts who also experimented it). The hypothesis is tested through a comparison of output decisions of the simulated model and actual decisions made during the field experiments.

Using the same input data, we compared the output of the simulated model against that of the experiments in the sense of decisions made and actions performed during the season. In no way did we attempt to estimate by simulation the crop protection quality that could have been achieved had we used the elicited model. That hypothetical goal would require a yield loss function for the combined powdery and downy mildews. Our expertise based approach is justified because such function is yet out of reach. Comparisons include each couple

$$Expe(plot, vintage) \ vs. \ Sim(plot, vintage).$$

The aim is twofold. By comparing the output of each modalities of experimented *oeGrapeMilDeWS* (both TRF and ANF are considered) to the simulated GrapeMilDeWS, we can assess the fitness of the model to the experiments. Then looking deeper in the sequences, we identify and interpret temporal discrepancies or even contradictions in the reasoning logic which may either come from errors in the experiments or miss-modelling.

To do, so we need to go through the following steps. First, the data sources are to be identified (section 6.4.2.2). Then, scenarios are built to run the GrapeMilDeWS simulator (section 6.4.2.2). Finally, the output sequences of D/A from simulation are compared to those from the experiments (section 6.4.3.2. The comparison methodology is presented in section 6.4.2.2).

Data sources In order to compare simulations and experimental outputs, simulations require input data which are consistent with the experimental conditions.

The required data for the simulation input files, throughout the whole crop protection season (May 1^{st} to mid August), are the disease level variables, the local downy mildew information, the rain forecasts, the phenological development.

The field evaluation records provide the dates and diseases levels used to build the update scenarios for variables M, O and Og.

The values for the ILM variable are interpreted from the plant protection service advisory bulletins. From the beginning of the season up to flowering (BBCH 10 to 65) [d], ILM turns positive at the first sighting of downy mildew symptoms. After flowering, it follows the risk assessment level of the plant protection service: recommendations to tighten the protection are interpreted as '+' and recommendations to wait before renewing treatment as '0'.

The triggering events for downy mildew treatments are not rainfalls but rain forecasts. However, during the experiment, not all rain forecasts events were considered significant. e.g. when the contamination risk associated to a rain event was considered low, on the basis of the meteorological bulletin, the event was discarded. The rules for building the simulator's input files will be presented in the next section.

The phenological development of the plots conditions the triggers for the first evaluation "E1", the third treatment "T3" and the last treatment "T6". Yet, the experts made a choice to have phenology loosely evaluated as precise evaluations are highly costly. Even though there are models to estimate the phenological stage of Grapevine (Garcia de Cortazar, 2006), we were not able to find the "function" used by the experts when they "loosely estimated" phenology. Therefore, we used the experts estimates, found in the TRF to build each plot vintage phenological scenario.

In section 6.4.2.1 we considered for the evaluation of performances only the halved plots with less treatments. When considering the fitness of the simulations of GrapeMilDeWS' model to its experimental counterpart, we have compared, for a given plot and year, the simulation run with both experimental half plot modalities. We meant to check if the simulation laid between the modality with minimum number of treatments and the one with maximum number of treatments.

Setting up simulations We have defined, above, the input scenarios as a succession of climatic, phenological and epidemic data. The following section details how these input scenarios are interpreted by the simulator to emulate the communication patterns between GrapeMilDeWS and its environment. The communication of variable values is triggered by events. The simulation time step is of one day.

Appropriate events are generated when ILM changes.

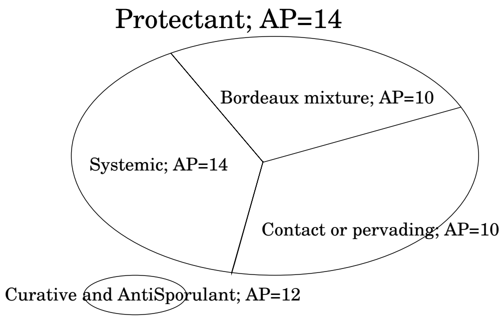
The case of rain forecasts is more complex, as the rain forecast records on the TRF are quite vague. Rain forecast events are sent daily to the simulator from the date of the initial forecast until the rainfall date, or, if that information was not recorded on the TRF, for three consecutive days past the initial forecast.

The only available data for diseases scenarios consist in the values (M, O and Og) at each evaluation date. As we cannot model the evolution of the disease accurately, we consider that whenever an evaluation is ordered, the value input into GrapeMilDeWS' simulator is the value from the experiment recorded at the same evaluation state, regardless of their respective calendar date. Consistently, phenology is fed into the system when significant information is available, either because a precise estimate of the phenology was recorded on the TRF, or because an action depending on a phenological stage

d. BBCH is a universal growth stages scale of mono and dicotyledonous plants. The adaptation to *Vitis vinifera* is due to Lorenz et al. (1995).

has been carried out. In that latter case, we inferred the phenology from the action. For example, if E1 was done on May the 4th then phenology "must have been" 15 on that day, as E1 is planned at *[BBCH≥15]* (i.e. more than five leaves on the shoots).

The active period (AP) of the fungicides is used in the simulator, to drive many decisions for the timing of the application and in some cases to spare optional treatments.



(a) downy mildew fungicide classes selection rule hierarchy.

Product Name	Active Period Length
SBI 1	14
SBI 2	12
Quinoxifen	14
Strobilurin	14

(b) powdery mildew fungicide classes selection rule.

Figure 6.3: Fungicide classes and associated active periods (AP) as modeled in GrapeMilDeWS

The following configuration parameters were used in all simulations: (i) the assessment of the distance from/to an event has a close/far threshold set to 3 (days); (ii) the active periods are presented in figure 6.3; (iii) the post treatment restricted entry interval (REI) is set to 1 (day); and (iv) the delay during which the "plot will soon be unprotected" before the end of the active period, is set to 3.

Simulation results are sequences of decisions and actions. We represent these sequences in a dataset M_s with the same format as M_e presented on page 165. Similarly, we derive $M_{s_{relative}}$ from M_s in the same way $M_{e_{relative}}$ was derived from M_e.

Study of the decision/action similarities between experimental and simulated sequences
The comparison is based on the two matrices (M_e and M_s). We compute the new matrix $M_{e-s} = M_e - M_s$.

The *total mismatches* are the cells $M_{e-s}(i,j)$ defined as: $|M_e(i,j) - M_s(i,j)| \geq 100$. Indeed, the crop protection season in our Bordeaux area starts usually early May (day card. is 122). If a D/A (Decision/Action) does not exist (cell value for no D/A is 0), the absolute difference will be over 100. The *perfect fits*, on the opposite, are cells defined as: $|M_e(i,j) - M_s(i,j)| = 0$. Indeed, if the same value is present on both M_e and M_s the difference will be nil. Finally, *temporal discrepancies* are all the remaining cells from M_{e-s} with values in $]-100,0[\cup]0,100[$.

With each of these three categories, we build the contingency table presented in figure 6.6(a).

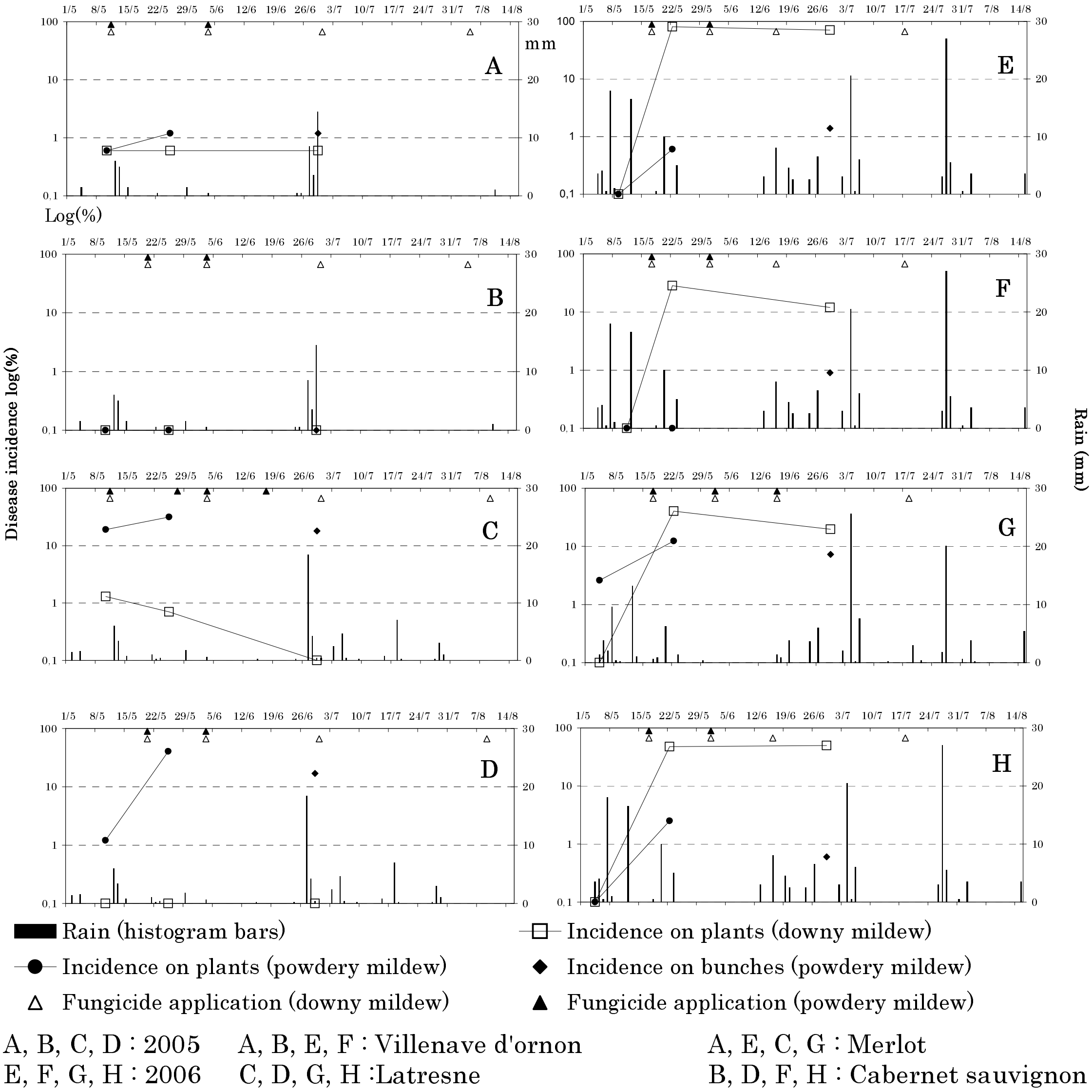


Figure 6.4: 2005 and 2006 crop protection seasons for each experimental plots, with on site precipitations, disease incidence at evaluation and treatment dates

6.4.3 Results

In this section, experimental results are first given. Then it is shown how the elicited model, once simulated, fits the experimental decision. The last part of this section forays in the detail of the decision making, comparing the timing of the decision from experiments with that of the simulations.

6.4.3.1 Performance obtained by experts experimenting *oeGrapeMilDeWS*

The main facts about crop protection in 2005 and 2006 for each of the 4 plots are given in figure 6.4.

2005

Analysis of crop protection in regard to powdery mildew. Concerning powdery mildew at Villenave d'Ornon, (figure 6.4 charts A and B) during the second evaluation (E2 on 25/5), no symptom was found on CS and only 0.6% of the stocks were attacked on merlot. Therefore only the two mandatory treatments were performed (resp.: 11/5 and 3/6 figure 6.4(A); 19/5 and 13/6 figure 6.4(B)). At Latresne, (figure 6.4(C)) 19% of the Merlot displayed powdery mildew symptoms after the first evaluation (E1 on 10/5) and after E2 the incidence reached 31% (25/5). Thus, in addition to the mandatory treatments, two optional ones were ordered during `Stage_2` and `Stage_4` (resp. 25/5 and 17/06)(Decision stages are presented on figure 6.1). On bunches, the incidence at the third evaluation (E3) was 18% (28/6). At the same site, on CS, 1.2% and 41% of the stocks were contaminated at E1 and E2 respectively. Thus only the two mandatory treatments were carried out on that plot.

Analysis of crop protection in regard to downy mildew. Only 3 symptoms of downy mildew, were found over all sites on Merlot at E1. No development of the disease was recorded during the next evaluation. Two optional treatments were applied, the first one upon discovery of the symptoms in conjunction with a rain forecast (11/5 on M. and 19/5 on CS). The second optional treatment (applied at `Stage_5` on 30/6) should not have been done as no increase in the attacked stocks frequency was witnessed. That application was decided by the experimental unit, because its managers considered the risks to be too high after a heavy precipitation event.

Overall quality At harvest, the estimated level of diseases on the leaves and bunches were very low (see table 6.2). The measured yield was above the authorized quota of the respective RDOs. The estimated alcoholic content, and total acidity were above RDO's requirements, thus compatible with the production of Bordeaux quality wine. Therefore the production objectives were met and the crop protection provided by *oeGrapeMilDeWS* was considered satisfactory.

2006

Analysis of crop protection in regard to powdery mildew. No symptom of powdery mildew was observed at E1 on 3 ou of 4 plots (2 at Villenave d' Ornon and CS at Latresne i.e. figure 6.4(E,F,H)). E2 having shown less than 20% of attacked stocks, and E3 less than 20% of the bunches, only the two mandatory treatments were carried out on these three plots. On the fourth plot (see figure 6.4(G)), 2.6% of the stocks were attacked at E1, 12.4% at E2 and 7.3% of the bunches showed powdery mildew symptoms at E3. The optional treatment T4 was thus carried out (16/6), whereas T5 could be spared as the situation seemed to have improved at E3 (29/6).

Analysis of crop protection in regard to downy mildew. No symptom of downy mildew was observed at E1. An optional treatment (T1) was nevertheless carried out (coupled with a mandatory powdery mildew treatment on 17/5), following the discovery of spots on some others plots in the vineyard estate (ILM turned to '+'). The application was carried out before a forecasted rainfall. This same treatment was done at all sites, 5 days before E2 (22/5). At E2, high levels of downy mildew were found at all sites: from 28 to

81% of attacked stocks depending on the plot. No optional treatment T2 was done as the vineyards were protected by T1. High levels of downy mildew at E2 mechanically implied application of an optional treatment T4 on all sites (16/6), in addition to T3 (31/5 at V.O. and 1/6 at Lat.). The mandatory treatment T6 was carried out between the 16th and the 18th of July. It was anticipated upon request of the vineyard management.

Overall quality The alteration of harvest due to the downy mildew was limited (less than 5% severity) and did not impact on yield below the production quotas authorized by the RDO. The foliage was affected by the downy mildew, mainly on the site of Villenave d' Ornon where defoliations were observed. Nevertheless, these defoliations were late and the potential degrees obtained largely higher than the minima authorized in the Registered Designation of Origin concerned.

Year	Site	Cultivar	Disease	Nb of fungicide application	Disease severity at harvest (%)		Défoliation (%)	Yield (hl/ha)	Harvest quality	
					bunches	leaves			Probable alcohol degree (% vol)	Acidity (gH2SO4/l)
2005	V.O.	M	Dm	4	0	0	0	62	13.3	3.7
			Pm	2	0.01	-				
	V.O.	CS	Dm	4	0	0	0	53	12.6	3.8
			Pm	2	0	-				
	Lat.	M	Dm	4	0	0	0	66	13.1	3.8
			Pm	4	0.53	-				
	Lat.	CS	Dm	4	0	0	0	71	12.4	4.5
			Pm	2	0.66	-				
2006	V.O.	M	Dm	4	4.0	28.0	9.3	52	12.6	3.8
			Pm	2	0.1	20.0				
	V.O.	CS	Dm	4		25.0	15.0	46	11.8	5.0
			Pm	2	<0.1	-				
	Lat.	M	Dm	4	1.2	6.3	0.6	62	13.9	4.6
			Pm	3	0.4	21.9				
	Lat.	CS	Dm	4	0.6	7.0	1.0	59	12.4	5.3
			Pm	2	<0.1	-				

Dm : Downy mildew V.O. : Villenave d'Ornon M : Merlot
Pm : Powdery mildew Lat. : Latresne CS : Cabernet sauvignon

Table 6.2: Quality results at harvest. The number of fungicide treatments corresponds to the minimal application half plots.

Conclusion of the experiments: Overall quality of the oeGrapeMilDeWS decision workflow On these two first years of experiments, we have shown that *oeGrapeMilDeWS* was compatible with grape and wine production standards. All plot yields were above the RDO target yield: 45 and 50 hl/ha in Pessac Léognan and 1ère côtes de Bordeaux respectively. The probable alcoholic degrees are also above the requirement: 10 and 10.5 alcoholic degrees respectively, as shown on table 6.2. More important, these results were achieved with only 2 to 3 powdery mildew treatments and 4 downy mildew treatments. These figures could have been even lower as unjustified extra treatments were applied by the vineyard management.

With the data from the evaluations, the experts have been able to decide whether to apply or skip a number of applications. It was shown above that the experts implement-

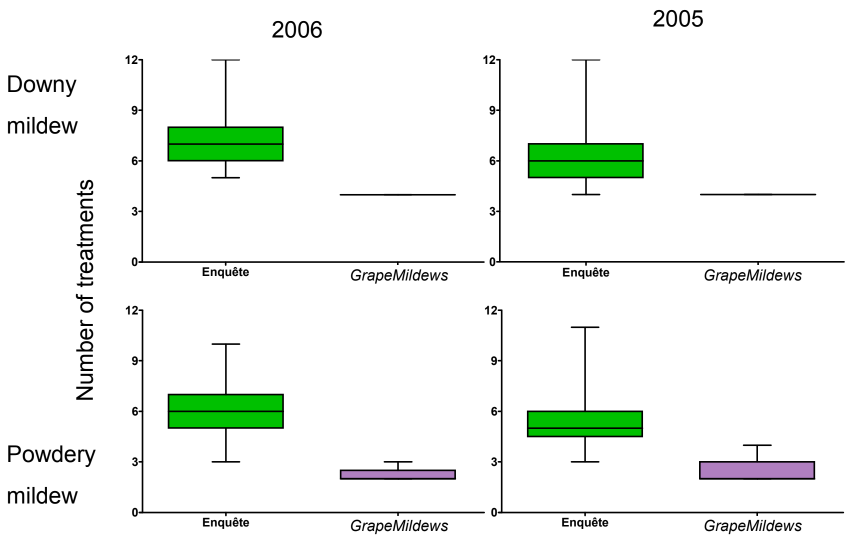


Figure 6.5: Number of treatments applied with *oeGrapeMilDeWS* compared to the declared number of treatments made in the Bordeaux region, during the years 2005 and 2006 for both powdery mildew and downy mildew. Sample size for the common practices: 103 vine growers from four Bordeaux RDO's *(Source: unpublished survey for the ANR project "ADD Vin et Environnement Durable", Bordeaux, INRA/Santé Végétale, 2006. See ann. A).*

ing *oeGrapeMilDeWS* achieved the crop protections goals. In all cases the production satisfied the RDO requirements and less than 5% severities were measured regarding each disease. Similar results were obtained during the 2007 campaign which in France was characterized by a heavy downy mildew epidemic. Unfortunately, the data were not available at the date this paper was written.

In all cases: 2005, 2006 (and 2007), the number of treatments was reduced to about half the median number of treatments done in the Bordeaux region (see figure 6.5). Note that the number of treatments recommended by *oeGrapeMilDeWS* is not absolutely lower than the lowest examples in the sampled population of vine growers. That is consistent with our objectives, to propose an IPM inspired crop protection strategy that reduces the number of treatments and yet keeps yield loss risks at an *acceptable level*.

6.4.3.2 Analysis of the similarity of the experimental and simulated decision workflows

The objective of comparing D/A sequences from GrapeMilDeWS and its expert counterpart *oeGrapeMilDeWS* is to assess the fitness between the experts' behaviour when implementing *in vivo* *oeGrapeMilDeWS* and the behaviour induced from the elicited specifications of such crop protection decision workflow (i.e. GrapeMilDeWS). The comparison will highlight two effects: quality of the modelling and conformity of the expert experimental behaviour to the decision workflow they described.

We will first look at the number of treatments requested in each of the simulations and experiments sequences. Then we study the fitness of GrapeMilDeWS' behaviour to *oeGrapeMilDeWS'* outputs. Fitness is analysed according to two dimensions: similarity and timing of the D/A. In a second subsection, the distributions of time intervals between a couple of D/A are studied and explained for both the experiments and the simulations.

Number of treatments In GrapeMilDeWS, the sum of treatments against both targeted diseases, can vary between 4 and 11. Table 6.3 shows that the two years experiment we have carried out stand in between (minima and maxima averages are 6.4 and 7.4). The simulations led to similar number of treatments: 6.6 treatments per plot on average. These satisfying simulation results stand between the experiments minima and maxima.

	Nbr of Treatments		
plot vintage	Sim	minima Expe	maxima Expe
2005 Lat. CS	7	6	7
2005 Lat. M	7	8	9
2005 V.O. CS	7	6	
2005 V.O. M	5	6	7
2006 Lat. CS	6	6	7
2006 Lat. M	7	7	8
2006 V.O. CS	7	6	7
2006 V.O. M	7	6	7

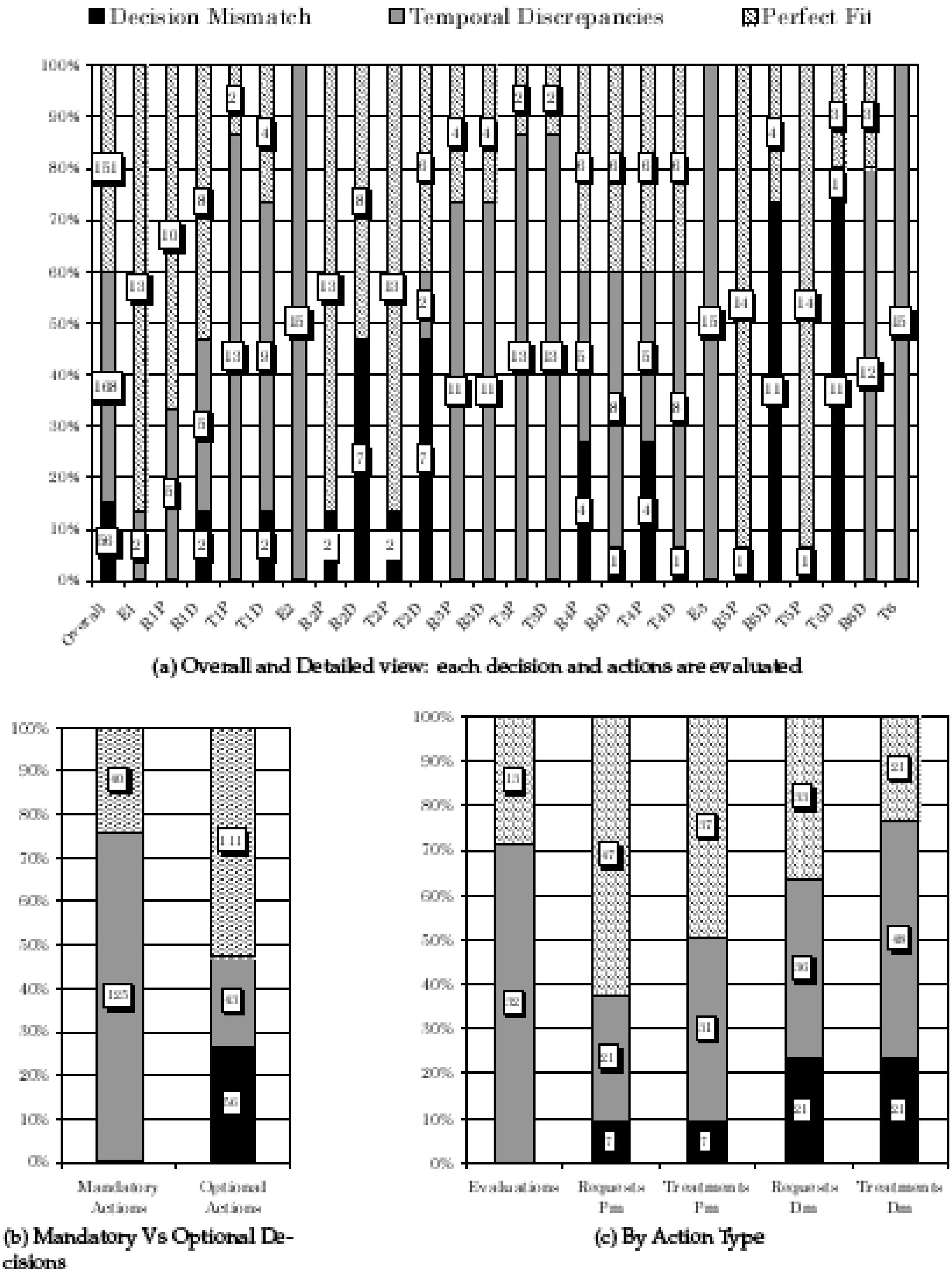
Table 6.3: Number of treatments per plot-vintage. Sim for simulation results; minima Expe for the half-plot-vintage with minimum number of treatments (idem Tab.6.2); maxima Expe for the second half with one more treatment.

Similarities of the Decision/Action dates between experimental and simulated workflows Thanks to the difference matrix M_{e-s} (presented in section 6.4.2.2), we check the way the simulations fit the experiments. When $M_{e-s}(i, j) = 0$, we have "perfect fit": there is less than one day difference between simulation and experiment dates.

Figure 6.6 presents the contingency tables from M_{e-s}, of perfect fits, temporal discrepancies and mismatches. Each column accounts for one action or decision. Overall, as shown on figure 6.6(a), there are 14.9% complete mismatches and 40.3% perfect fits. Figure 6.6(b) and (c) show that there are more perfect fits with the optional actions than with the mandatory ones (10.6% vs. 29.6% overall D/A). This satisfactory result is explained by the fact that "non occurrence" of a D/A on both the experiment and the simulation is a perfect fit. Yet, when D/As are absent there can only be perfect fits or complete mismatches, but on first approach we consider both perfect fits and temporal discrepancies as satisfactory.

Decisions regarding powdery mildew led to fewer errors than those concerning downy mildew (figure 6.6(c)). In table 6.4, it can be seen that complete mismatch occur for T2D and T5D for a number of plots while simulation is very close to experiment for T2P and T5P. For T2P, differences can be related to processing of rain forecasts, and cancelling of late T2D in anticipation of stage 3 in the experiments. For T5D, it happens that experts decided an extra-treatment which should not have occurred.

Figure 6.7 summarizes the distributions of the temporal discrepancies, viewed through the Tuckey quartile box plot representation. We used the R statistics software (R-Development-Core-Team, 2008) to generate all box and whiskers figures.



(a) Overall and Detailed view: each decision and actions are evaluated

(b) Mandatory Vs Optional Decisions

(c) By Action Type

E	for Evaluations	R	for treatment Requests	T	for Treatment applications
{1: 6}	refers to the Stage at which the decision for such action is taken				
{P,D}	the target disease: P for powdery mildew and D for downy mildew				

Figure 6.6: Proportions of perfect fits, temporal discrepancies and complete mismatches between the Experimental and Simulated GrapeMilDeWS. Labels represent the number of cases in each category.

Optional Treatment	Sim		minima Expe		maxima Expe	
name	2005	2006	2005	2006	2005	2006
T1D	4	2	4	2	3	4
T2P	0	0	1	0	1	0
T2D	3	2	0	2	0	0
T4P	2	2	1	1	3	1
T4D	1	4	0	4	0	4
T5P	0	1	0	0	1	0
T5D	0	0	4	0	3	4

Table 6.4: Distribution of optional treatments over all plots per vintage, treatment and GrapeMilDeWS implementation type: Sim. for simulation results; minima Expe. for the half-plot-vintage with minimum number of treatments over the season (idem Tab.6.2); maxima Expe. for the second half with more treatments.

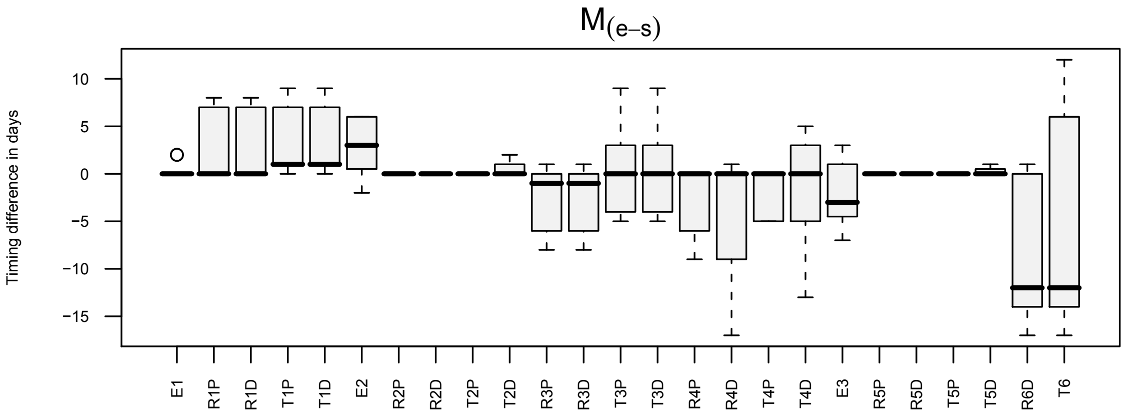


Figure 6.7: Tuckey quartile box plot of time differences from $M_{(e-s)} = M_e(i,j) - M_s(i,j)$ for each action or each decision

With this view, the directions of the discrepancies are clearly visible. Values in $M_{(e-s)}$ are positive when the simulator acted earlier than was recorded in the experiment, for example E2 was done on May 25th (day 145) in the experiment and yet "done" on May 19th (day 139) with the simulator; and vice versa the difference is negative when the simulator was "late".

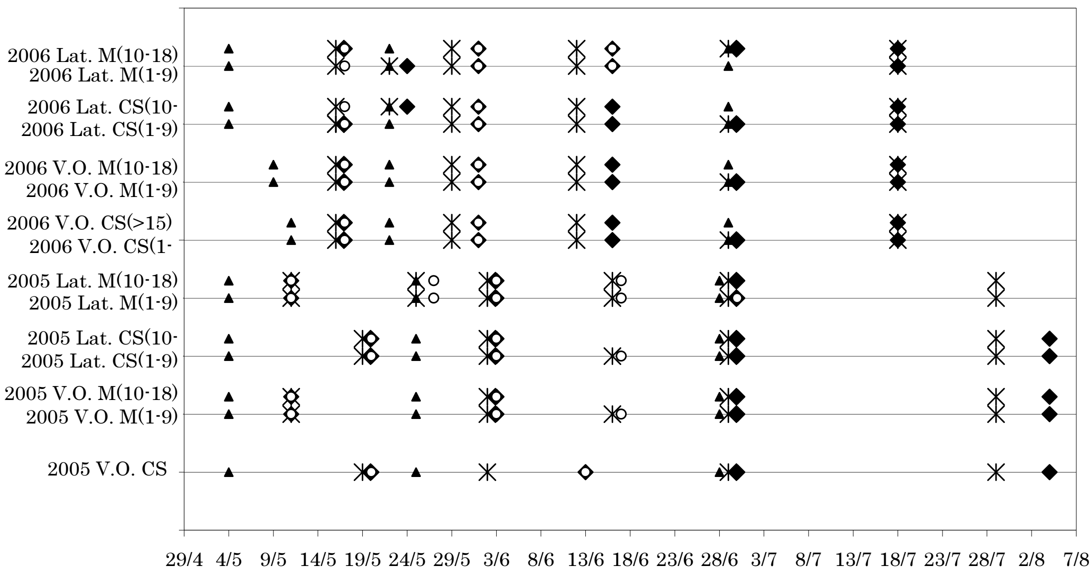
The last two boxes, on figure 6.7, concern the last treatment T6. Their high values are artefacts caused by the early application of the downy mildew treatment in 2006 on the initiative of the vineyard manager (see section 6.4.2.1). Once these artefacts have been excluded, it clearly appears that, whatever the decision, the time gap between experiments and simulations is stable. The average is -0.25 day over the whole season. This comforts the idea that no lag is accumulated in the process. Indeed, the simulation scenarios for phenology, taken from the experiment, had a resynchronisation effect which prevented temporal discrepancies to accumulate unduly between simulation and experiments.

Overall 84% of the decisions are identical between the simulator's and the experts'. Considering the fact that the model was purely declarative, that we did not exclude the experimental artefacts, we are satisfied with the quality of the model in its ability to make the same decisions as the experts. Yet 53.9% of these correct decisions are made at with a timing error. However, once the T6 artefacts have been removed, the average discrepancy is less than one day ("perfect fit" included).

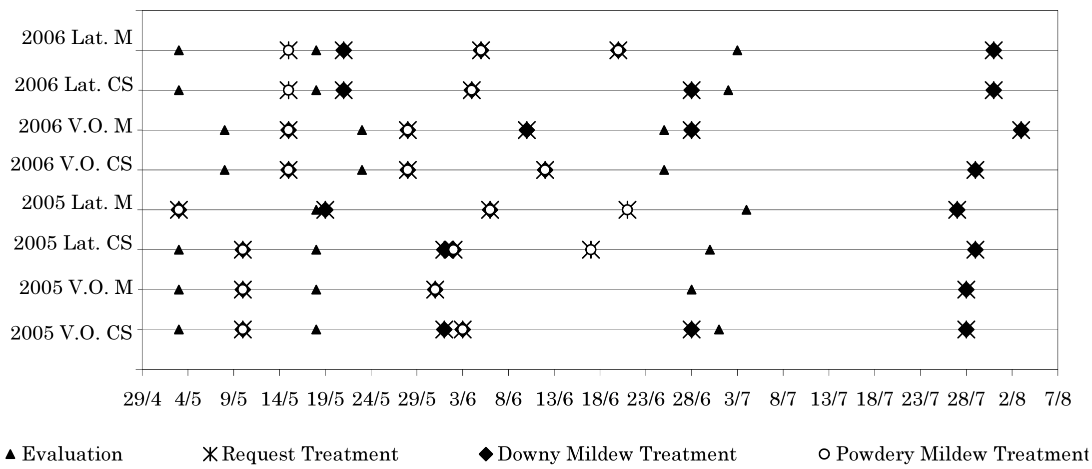
6.4.3.3 Analysis of the Decision/Action sequences in experiment and simulations

The behaviour of GrapeMilDeWS is analysed hereunder. Reasons for some of the major temporal discrepancies are given, and some patterns in the executions of the experiments over the set of 4 plots are identified, in order to exhibit potential modelling problems. We will proceed in two steps, first looking at the whole sequences of decisions and actions and secondly, analysing the time interval between couples of D/A.

Figure 6.8 is given for reference as raw data. We will use figures 6.9 and 6.10 to support our analysis over these data.



(a) Decisions/actions from the experiments



(b) Decisions/actions from the simulations

Figure 6.8: Sequences of decisions and actions for each (half) plot and year. (symbols may overlap if decisions are taken on the same day)

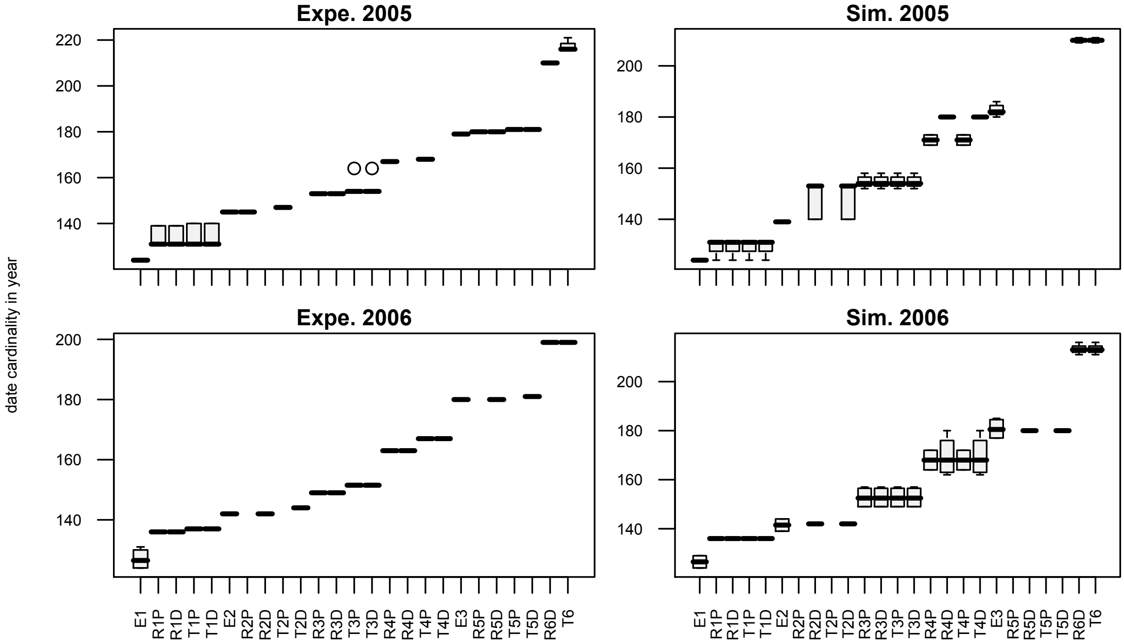


Figure 6.9: Temporal distributions of Experiments and Simulations, decisions/actions for all vintages. Zero excluded (non decisions are not presented here)

Sequences of decisions and actions Simulations (figure 6.8(b)) are synchronised at the beginning of the season with the first evaluations of the experiment (figure 6.8(a)) based on the experts assessment of phenology. The phenological stage conditions also ensure synchrony between the experiments and the simulations at `Stage_3` and `Stage_6`'s entry.

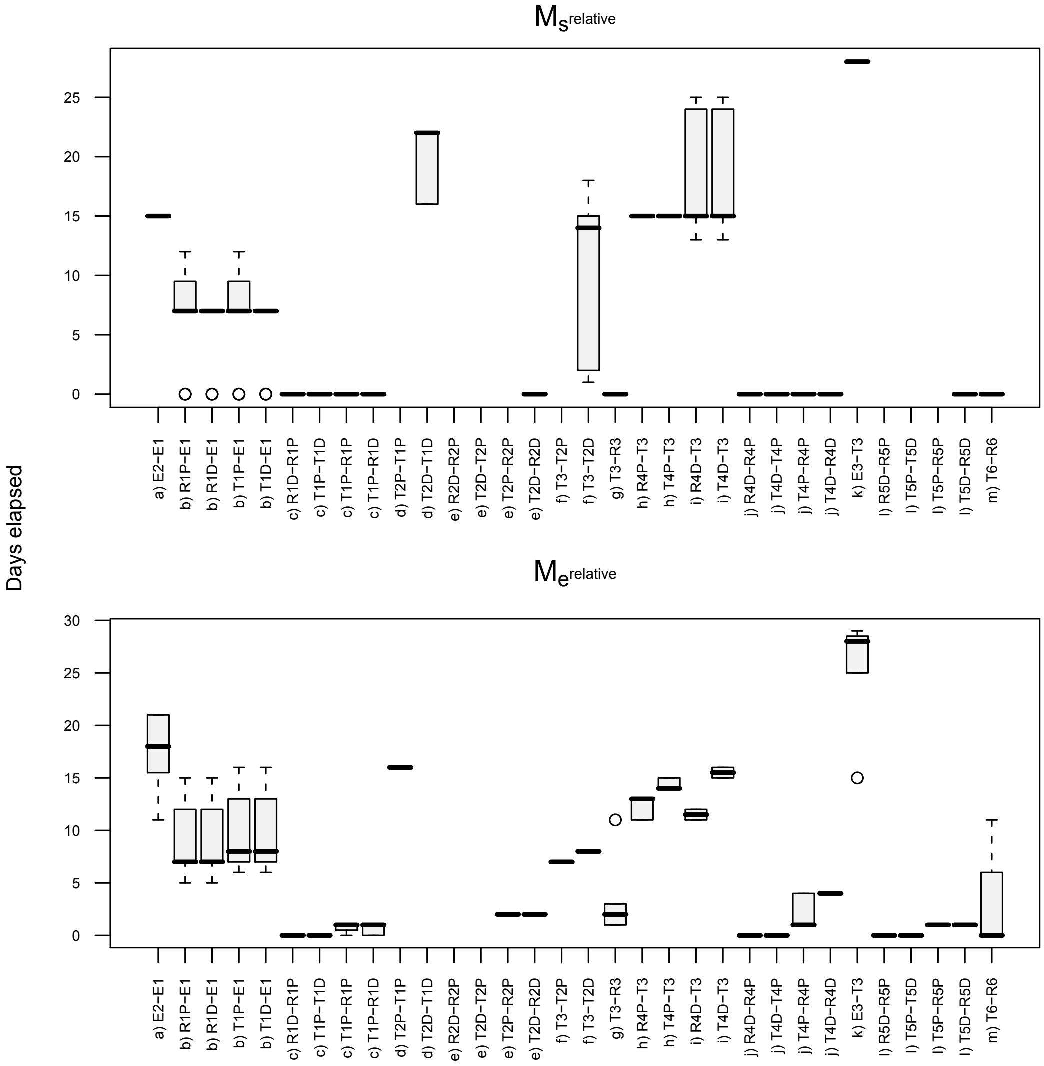
However, at first glance, figure 6.8 shows quite different patterns in the sequences. Even though both decisions systems were "initialised" between May 4th and May 11th, there is clearly more synchronicity between plots in the experiments sequences (a) than in the simulation ones (b).

Figure 6.9, which gives the temporal distributions of each D/A is another illustration of this point. For instance, in 2006, all experimental decisions were made on the same day for all plots whereas the simulations display some variability. The experts seem to have executed the *oeGrapeMilDeWS* process simultaneously for their four plots whereas the simulations led to decisions that were totally independent one from the other.

Indeed each simulation of a plot-vintage couple was a separate execution of the simulator. The simulator has no operational constraints, therefore all decisions were executed on the exact same day any order was issued; whereas the experts could not abstract themselves from the operational constraints of the vineyard management the way the simulator can.

time intervals The next section exposes how synchrony of the experts' decisions was achieved. We analyse the distributions of $M_{s_{relative}}$ and $M_{e_{relative}}$, identifying time constraints that were broken and those that were respected.

Both diagrams on figure 6.10, display relative times between pairs of events, from $M_{s_{relative}}$ and $M_{e_{relative}}$ respectively. Working with relative time intervals allows us to check constraints such as "does $E2 - E1 = 15$?" (i.e. is the 15 days rule between evaluation 1 and



E	for Evaluations	R	for treatment Requests	T	for Treatment applications
$\{1:6\}$	refers to the Stage at which the decision for such action is taken				
$\{P,D\}$	the target disease P for powdery mildew and D for downy mildew				

Figure 6.10: Tuckey Box plot for the relative differences between pairs of D/A. From bottom to top, experiment ($M_{e_{relative}}$) and simulation distributions ($M_{s_{relative}}$). The Y-axis represents days elapsed between two D/A. N.B. Dots represent outliers and whiskers are not significant due to sample size

evaluation 2 respected? See figure 6.1). Time intervals also permit to control if powdery mildew treatments and downy mildew treatments have been separated (ex. T1P-T1D). Finally, they also measure the lag between the decision and the execution (ex. T6-R6).

Note that many values have had to be excluded as they contained missing events in one term of the difference and would have artificially created high values. As we are working with few data, the box plot should be interpreted as intervals of variations rather than a summary of some theoretic distributions.

The points that have been studied on figure 6.10 are explicited below. Each item of the list (c,e,g,j,l,m) is about time lag between decision and the actual corresponding action (spraying) within a stage, which we will call *spraying delay*. The second list (a,b,d,f,h,i,k) is about inter-stage conformance to temporal constraints.

c) Spraying delays at `Stage_1`.

The delay between the decision and its execution in the experiments was never longer than 1 day. It is also possible to test the disjunction of the optional downy mildew treatment from the mandatory powdery mildew treatment. The experiments and the simulations were consistent with applying joint treatment in all tested situations.

e) Spraying delays at `Stage_2`.

No mixed treatment was done. It was not observed any significant delay between request and application of a treatment.

g) Spraying delays at `Stage_3`.

In the experiment the delay spans from 1 day to 3 days with the exception of plot V.O. CS in 2005 which was treated 11 day after the request was sent out. In that later case, the treatment was requested at the same time as for other plots (June 2^{nd}), but the experts requested that plot to be treated later, because its phenology was less advanced.

j) Spraying delays at `Stage_4`.

Only one plot in the experiment was treated with a delay above one day targeting powdery mildew, but all plots concerned by a downy mildew application were treated with a 4 day delay.

l) Spraying delays at `Stage_5`.

No disjunction between applications, no significant delay was observed.

m) Spraying delays at `Stage_6`.

The delay span from 0 to 11 days. Indeed the last mandatory treatment was done in urgency and with no delay in 2006 (see section 6.4.3.1)). On the other hand, the 2005 applications seem to have been greatly anticipated. For instance, plot Lat. M was treated 11 days after the order was sent and the other plots have been treated 6 days after the request.

We shall now study the inter-stage delays.

a) Time span between E1 and E2.

The simulations ($M_{s_{relative}}$) strictly respects the `Stage_1` exit condition, i.e. E2 was always carried out 15 days past E1. In the experiment($M_{e_{relative}}$), the length of E2-E1 varies between 11 days and 21 days. This seems due to a 'batch' implementation of the procedure: the experiments display total synchrony for all plots at E1 and

E2 in 2005 and a clear re-synchronisation of all plots at E2 in 2006. Consequently, no plot was evaluated at precisely 15 days interval.

b) Time span between the first evaluation E1 and the request or the application of the first treatment.
The simulations' decisions were consistent with those from the experiments: generally 5 days between E1 and R1 and/or T1 [e](see "**c**"). Rain and downy mildew symptoms detected in the region (i.e. ILM='+') triggered treatments T1D and T1P.

d) Time span variability between T1 and T2.
There is no T2D in the experiments but the simulator results show a great heterogeneity in the timing of that application, from 16 days and up to 22 days. This denotes GrapeMilDeWS' flexibility. It can either provided "a perfect tilling protection" (the active period of T1 lasted 14 days) or have the treatments positioned just in time.

f) Time lag between treatments at `Stage_2` and T3.
In the experiments, about one week (7, 8 days) separates T2 (P and/or D) from T3 (P and D, mandatory), whereas the simulations exhibit a large variability: between 1 and 18 days. The reason for such variability is to be found in the simulator ordering "two treatments on downy mildew, two days in a row". The first at `Stage_2` the second at `Stage_3`(see section 6.4.3.4)

h) Time span between the application of T3 and the request or treatment against *powdery mildew* at `Stage_4`.
The simulations, enforced the "one active period" constraint between the treatments made at `stage_3` and the entry into `Stage_4`: 15 days between T3 and R4P. Considering the experimental results, the 4th treatment request was anticipated (with as little as only 11 days between T3 and R4P). The request was issued while still in `Stage_3`. However, the treatments were carried out in `Stage_4`: there was never less than 14 days between T3 and T4P in the experiments.

i) Time span between the application of T3 and the request or treatment against *downy mildew* at `Stage_4`.
This treatment was applied solely in the 2006 experiments. The request was *anticipated*. For simulations, the delays span from 13 to 25 days. The shortest delay was due to a curative fungicide selected at T3 (12 days AP). The longest delay: 25 days, is explained by a moderate downy mildew risk situation where a rain forecast was awaited and which eventually triggered the treatment request.

k) Time span between T3 and the third evaluation.
The GrapeMilDeWS model states that E3 should be done 28 days after T3 (i.e. about two active periods later). In the experiments, the delays between T3 and E3 span from 15 to 29 days. In 2005, the delay was of 25 days with the exception of V.O. CS which only has a 15 day delay because it was treated later than the other plots against T3 (i.e. the outlying point at tick '**k**' on figure 6.10 chart $M_{e_{relative}}$). The expert resynchronised this plot with the other by shortening the $E3 - T3$ delay to 15 days for that plot.

e. any request and/or any treatment at `Stage_1`.

6.4.3.4 Discussion about the causes of discrepancies between experiments and simulations

Overall our analysis of the results from the previous section points out three phenomena: (i) The batch mode management of the experiments, (ii) anticipations and (iii) existence of artefacts from both the experiments and the simulations.

The batch mode management: synchronisation of the experiments' actions Our understanding is that the experts have taken into account operational constraints. With figure 6.9 it clearly appears that experimental decisions and actions for all four plots were made at the same time and similarly, carried out in synchrony. Although the decisions may be different i.e. not all plot will be treated, all those that will, will be treated at the same time.

Anticipations As shown above, the experts have a tendency to synchronise the decisions on all four plots. This is carried out by anticipating the decision making for the plots that are late (see **k** E3-T3 on page 180). Yet synchronisation was not the only reason for anticipations. It seems that the expert anticipated some decisions because they could foresee the future necessary actions (see "**h**" on page 179). For instance, as soon as T3 was done, it was possible to make most `Stage_4`'s decisions regarding which treatments would be necessary. Note that only ILM may be updated at `Stage_4`, yet it is not always involved in the decision making (see Part.1 section 4.8 Léger et al., 2008a). This anticipation capacity is not provided by GrapeMilDeWS' statechart. As a reactive system modelling language, Statechart is ill equipped to program such introspective behaviour.

Artefacts The detailed analysis of the results below points out a few artefacts both in the experiments and the simulations.

Under certain conditions, the simulator could recommend to have two treatments applications done in less than two days. On both CS plots in 2005 (see figure 6.8(b))the second treatment was done between June 1^{st} and June 3^{rd}, date of the third treatment (day 153 to 155)). GrapeMilDeWS' statechart "could not foresee" the upcoming realisation of the `stage_3` entry condition.

Nevertheless, under more realistic conditions, the above scenario should not happen. Indeed, there is a security mechanism against late treatment at that stage[f]. This mechanism could not have worked in the simulations as we used partial phenological data which are not continuous.

The frequency of such unwanted sequences would need to be evaluated. Yet under realistic conditions, which would be the case if GrapeMilDeWS was used embedded in a decision support system (DSS), we feel confident about the fact that the above combinations would be rare.

We do not consider as an artefact the non-respect of temporal constraint in the experiments caused by the batch mode synchronisations or by anticipations. However, we identified as an experimental artefact the application of treatments at `Stage_5` in 2005, which could not be justified according to GrapeMilDeWS principles. Similarly,

f. When `Stage_2` is active and no T2D was carried out and ILM is '+' then `Stage_3` should be entered at the very beginning of flowering instead of waiting until mid flowering (BBCH 60 and 65 respectively).

the early `Stage_6` treatments applied in 2006 were considered as artefacts. Indeed, they were decided by the vineyard management who feared epidemic development in July because the previous downy mildew treatment had been done at least one month before. confronted with forecasts of heavy rainstorms, it was not possible to convince these operational managers, who are responsible for the economic output of the vineyard, not to act on what they felt was a highly risky position.

6.4.4 Discussion

The discussion is organised as follows. First, we study how knowledge has been leveraged by the experts to design GrapeMilDeWS. Second, we show how joining the management of both downy mildew and powdery mildew is an innovation. Third, we discuss the difficulties of monitoring. And finally, the acceptability of our crop protection strategy is discussed.

6.4.4.1 GrapeMilDeWS and expert knowledge

In this section, we show how GrapeMilDeWS' design efficiently takes advantage of the available epidemiological knowledge to protect the vineyard with less sprayings than common practices.

Against powdery mildew: the aim is to control the early stages of the epidemic and to break its dynamic before the exponential replication stage. This goal is achieved by assessing the existence of an epidemic in the plots when five leaves are unfolded (BBCH 15). The probability of not spotting *Erysiphe necator*'s primary infections this early in the season is quite important. Therefore, a 10% sampling rate is used, and immediate spraying action is taken above a threshold frequency of only 2% of the stocks infected. The second evaluation regarding powdery mildew aims at correcting an underestimation of the risk during the first evaluation.

Past the flowering period (T3), the yield loss risk induced by powdery mildew decreases rapidly. In our 2005 and 2006 experiments, during the cluster closure and ripening periods, no treatment was applied. This way of doing is very different from the common practices but gave, so far, satisfactory results. On that very point, the simulations seem to indicate that the formal model may be too conservative (see table 6.4) (i.e. the experts have been more daring in real life than when specifying the formal GrapeMilDeWS model during the elicitation).

Against downy mildew: a common heuristic to protect grapevine against *Plasmopara viticola*, in Bordeaux, is to make sure that the vines are protected before rainfall events. GrapeMilDeWS' principles are to take no chance when the epidemic is widely spread (see '++' thresholds on table 6.1) and monitor the rainfalls when the epidemic pressure is moderate. This is achieved by starting the crop protection when the first symptoms are spotted in the plot or if there is risk in the vicinity (ILM='+'), and then to renew the treatments when either the attack is fierce (i.e. M='++') or rain is forecasted. These are widespread principles against downy mildew (for example (CA11, 1999) or (Bleyer et al., 2007)).

The innovative aspects here, are much more methodological than epidemiologic. By giving the strategy an explicit form, we implement the common heuristic in a way that prevents GrapeMilDeWS' user to simply state the common heuristic (have the vineyard protected before rainfalls) and then implement it with a systematic two weeks renewal frequency under the justification that rainfalls are very frequent in Bordeaux. With this strategy, the application decision needs to be justified, "documented".

6.4.4.2 Managing two diseases in the same crop protection strategy

The common practice is to apply mixed fungicide treatments against powdery and downy mildew most often at a fixed renewal rate (10 to 15 days depending on the fungicides used). GrapeMilDeWS' approach combines the rational decision making against two pathogens which implies possible desynchronisation of the treatments.

By positioning two mandatory treatments for each target disease, including one targeting both, it ensures a minimal protection at key periods. The number of mandatory treatments could be theoretically lower, but GrapeMilDeWS would require extra monitoring to reach the same compromise between pest control and yield loss risks.

As the pathogens development conditions for powdery mildew and downy mildew differ, the expertise consisted in making the necessary arbitrations, setting priorities and taking into account the operational constraint of the vineyard management. Indeed, to be acceptable by growers, GrapeMilDeWS should avoid whenever possible disjoining the application of treatments if both are needed.

These arbitrations led to the stages structure. For instance `Stage_4`'s logic is made so that there can be only one application, which implies that if upon entering that stage, a powdery mildew treatment alone was ordered, then it is a strategic choice not to have a second application against downy mildew whatever the upcoming conditions. These measured risks are acceptable at that period as downy mildew needs to be low for the sole powdery mildew application to be decided.

The counterpart from having the possibility to spare one or the other or both treatments all along the season (except at `Stage_3`), is that there must be re-synchronisation, from time to time, of powdery and downy mildew treatments, in order to keep the process manageable. Thus, some treatments may be renewed before the end of their active period. On the other hand, on years with moderate epidemics GrapeMilDeWS spares treatments.

The evolution of priorities during the season led the experts to the 7 stages architecture for their reasoning. That architecture in turn led us, when preparing the formalisation, to view the experts IPM crop protection solution as a decision workflow, where each stage follows an independent decision flow. Statechart became a good candidate for efficiently representing the whole system in a compact, yet coherent manner, although alternative process formalisms could be evaluated to represent decision workflows (Activitychart, (Eshuis and Wieringa, 2002) or Petri nets (van der Aalst et al., 1994), among others).

6.4.4.3 The burning issue of monitoring

Monitoring is the heart of the decision workflow. It implies acceptance to live with the disease and control it. Yet, evaluations are very costly. This is the reason why they

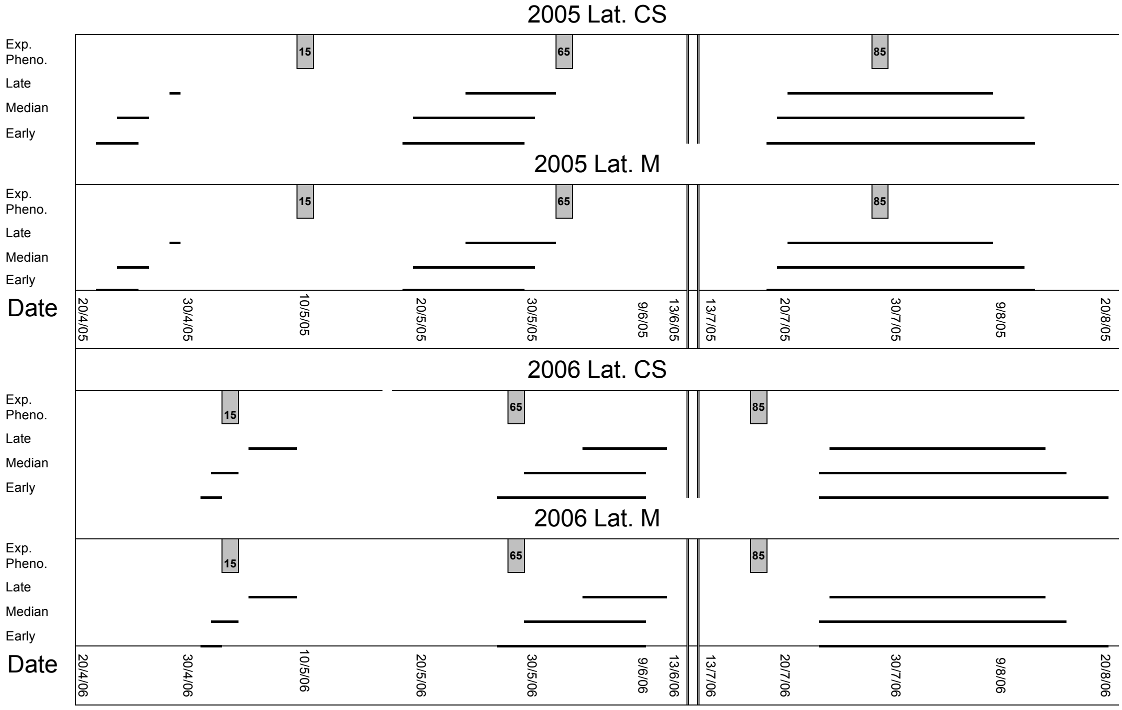


Figure 6.11: Comparison of experts' phenological stage estimates in 2005 and 2006 Vs three, sum of daily avg. temp. ≥ 10°C + Cultivar + Site, multi-linear regression models. Early was fitted over the early organs. Median represent central tendency half the plot organs are above a given stage, half are below, Late was fitted with most immature organs. With each models, we have represented the time duration of the following phenological stage intervals : [15,16[;[60,65];[80;85]. In the gray boxes the expert phenological estimates for E1, R3 and R6.

are limited to three. We argue that this is enough to take all pest management decisions for the whole season and achieve satisfactory results.

Other works we have carried out with vine growers have shown that most of them observe each of their vineyards (60% of the surveyed growers) [g]. Yet, these observations are rarely carried out with a systematic methodology and decision oriented objectives (only 35% declare alleviating or strengthening their crop protection program). GrapeMilDeWS proposes to replace these observations that check the crop protection's efficiency (61% of the answers), with observations that provide a picture of the epidemics which is precise enough to allow the grower to decide and act on them.

Furthermore, the epidemics levels' evaluations are positioned according to an estimation of the phenological stage. Figure 6.11 represents the experts' estimate of the phenological stages with phenological models based on cumulated temperatures.

The models have been fitted for each site and each cultivar over a dataset of systematic phenological evaluations from 2001 to 2005. Models "Early" represent the early organs at the plot scale; "Median" models indicate that, at a given value of the model, 50% of the organs in the plot are in or below that given stage; as for "Late" models, they represent the development of the most immature organs. The methodology for this modelling is

g. The statistics presented are part of an unpublished survey done by INRA/Santé végétale in 2006 for the ANR project "ADD Vin et Envirronement Durable". The sample was composed of 103 vine growers from four Bordeaux RDOs.

presented in (Léger, 2008). The fitted data were collected into the Eichhorn and Lorenz scale (Eichhorn and Lorenz, 1977), which does not differentiate the foliar dynamic from the fruit one. The work from Coombe (1995) was used to establish a function from the Eichhorn and Lorenz scale to the BBCH scale.

The problem here is to define what the experts consider when estimating phenological stages. For instance, figure 6.11 shows that the experts' actions based on their phenological estimates are systematically late in 2005 compared to the fitted models, yet the opposite becomes true the following year.

One of the difficulties lies in how the focus is changed from monitoring the leaf development early in the season, to observing the fruits from pre flowering up to harvest. Understanding the way the experts switch from foliar to fruiting phenological monitoring would help in modelling GrapeMilDeWS and thus standardise its use. Indeed, the phenological stage estimate should be as disambiguated as possible in order to make better comparisons between experiments. However, that standardization should not sacrifice the operationality of GrapeMilDeWS.

Further work with GrapeMilDeWS should also quantify the sensitivity of the model to the experts' interpretation of rain forecasts.

One of our future work is to design a powdery mildew equivalent to downy mildew's ILM. Indeed, spotting early powdery mildew symptoms requires some skills. Having a procedure to pool information on the beginning of the powdery mildew epidemics each year would be of great help. So far, in France, this information is not provided by the plant protection services.

6.4.4.4 Acceptability and risk management

Acceptability was indeed one of the concerns of the expert designers when they chose to have three mandatory applications, when they chose to limit the number of observations to three, and when they designed the decision system as a decision workflow broken down in 7 stages.

The design choices were mostly focused on acceptability, making sure that the process is economically and logistically feasible. However, on both years, experimental artefacts have been exhibited (at `Stage_5` in 2005 and `Stage_6` in 2006) due to the vineyard managers not perceiving the epidemic risk in the same way the phytopathologist experts did. The reason for that is the very long period between `Stage_3` and `Stage_6` during which no mandatory treatment is required. There is usually more than a month between the third evaluation E3 and the ripening period at which the last treatment T6 should be applied. Under low epidemic conditions at E3, a treatment at `Stage_5` can be spared, but then there is necessarily a period during which the crop is not protected . From our experience, such scenarios are perceived as highly risky by the growers. In the advent of heavy rains, the experts were unable to coerce them into inaction. We can agree with the vineyard management, that there is a lack of consensus amongst professionals on the fact that under low epidemic pressure at cluster closure (the phenological stage at which the third evaluation is done), downy mildew could be controlled with only one treatment positioned at mid ripening.

6.4.5 Conclusion

An IPM crop protection annual strategy targeting two of the most aggressive pathogens of vineyard : *Plasmopara viticola* and *Erysiphe necator*, has been presented extensively in part.1 (Léger et al., 2008a) using a graphical yet mathematically sound formalism. This second paper, aims at assessing the quality of the GrapeMilDeWS' formal model.

Designed by expert phytopathologists, GrapeMilDeWS was evaluated for two years on four real scale plots. In the meantime, the model was elicited from the experts. In order to assess the model quality as a crop protection tool, we have proceeded in two times.

First, the crop protection experiments were presented and commented. Then we assessed the ability of the formal model GrapeMilDeWS used as a simulator, to take the same decisions as the experts did, had it been submitted to the same environmental conditions.

The experts when applying *oeGrapeMilDeWS* are able to protect the plots against both powdery and downy mildew, while achieving the Registered Designation of Origin target yield and quality (i.e. sugar concentrations). These results were achieved with roughly half the number of treatments applied by the Bordeaux region vine growers, and yet less than 5% severity at harvest for powdery mildew on the cluster and similar results for downy mildew on the leaves.

When evaluating if the formal model would have behaved as the experts, we used a comparison methodology structured in two steps. First, the overall fitness of the model's decisions to the experts' decision was evaluated. The second step, was much more critical, looking deep into both the simulations and the experiments to identify the causes of decision mismatches and temporal discrepancies.

The fitness evaluation was quite satisfactory since 85.1% of the decisions taken by the simulator match those of the experts. However, 44.8% out of 85.1% of the decisions were not made precisely at the same time; yet, no drift was found overall between the simulations and the experiments.

Furthermore, the discrepancies found were explained by three main factors. Indeed, when experimenting and devising GrapeMilDeWS, the expert designers were not always aware of their own resource management, and thus did not report on it when elicited. Comparison showed that when experimenting, the experts managed the plots through *batches of decisions* for operational reasons. Operational constraints were also the reason for the experts to regularly *re-synchronize* the four plots. The experts were also shown to use *anticipations* on some decisions in order to have the applications positioned in such a way that the crop would be better protected.

This later point is just impossible for the current implementation of the simulator to achieve, as it requires introspection and breaking temporal constraints. The formal model was elicited from the experts without accounting for anticipation and the chosen modelling technique should be adapted in the light of these findings.

Finally, the vineyard staffs were also found responsible of some of the discrepancies. GrapeMilDeWS being an experimental tool, the operational management felt on some occasions too risky not to protect the crop. These experimental artefacts are interesting in that they question the acceptability of the solution.

In the next step, we will improve the formalism in order to satisfactorily represent the sequentiality of decisions in a more versatile way than Statechart did, yet formal enough to be considered a program. Timed Statechart should be investigated in this regard (Kesten and Pnueli, 1992; Graf et al., 2003).

The experiments have continued and been extended to professional vineyards. Transfer is our target for the future years. In order to achieve that goal, we will focus on evaluating GrapeMilDeWS' behaviour when run by non expert growers with the help of a prototype Decision Support System (DSS). But the more challenging work towards providing a DSS solution based on GrapeMilDeWS is in the operational research field (Naud et al., 2007). This research would allow to manage the whole vineyard estate through GrapeMilDeWS and still take the decisions at the plot scale. Another direction for our research is to find ways to lower the costs of observations through efficient observation generalisation mechanism. And finally; there is need for investigation on which organizational structures should be best for the sharing of the epidemic information between neighbouring growers especially where the land ownership is scattered as is the case in most French vineyards.

6.5 Discussion du Chapitre

La discussion de l'article est prolongée par cette section. D'abord on montre pourquoi cette validation était nécessaire dans le processus de formalisation du POD Mildium. Puis on évoque les points critiques de la méthode de conception mise en œuvre par l'unité Santé Végétale et comment la formalisation permettrait d'améliorer l'expérimentation. Enfin on évalue la généricité de la méthode de validation présentée dans le précédent article.

6.5.1 Méthode de validation

On a cherché à montrer que le modèle « Mildium formel » (Mf) simule (au sens de la définition 9) le POD « Mildium expert » (Mx). On a considéré Mx comme une boîte noire[h], mais dont une partie de son comportement peut être observée à partir des enregistrements issus de l'expérimentation. Il s'agit en effet de valider que Mf peut reproduire les comportements déjà observés[i]. On recherche une preuve que la formalisation a bien assuré la reprise des acquis du passé.

A l'issue de cette validation on conclut que le modèle formel a valeur de référence, d'étalon. En effet, tout en conservant les structures et principes généraux du concept POD Mildium, le recueil de connaissances a permis d'objectiver l'expertise (c'est à dire de clarifier et de transcrire les choix de conception) mais la modélisation a aussi permis de couvrir l'ensemble des situations possibles de manière exhaustive. Cela a demandé aux experts de gérer des cas de figure non prévus et donc de créer de nouvelles réponses, de prendre position.

La validation a permis de montrer que, malgré les transformations issues du recueil, il n'y a pas eu de trahison du système de décision expérimenté. En montrant que Mf décide ce que Mx décidait, il devient possible de revendiquer pour le modèle l'efficacité agronomique obtenue par les experts au cours des expérimentations.

6.5.2 Résultat de la validation : les points critiques

Cette validation a également permis de mettre en évidence certains points critiques, en particulier le problème de l'estimation humaine de certaines entrées du modèle. Le cas de la phénologie est discuté ci-après.

Les résultats montrent que l'estimateur de phénologie ne peut être considéré comme une horloge absolue. Les experts se basent en effet sur leurs observations pour faire une estimation du stade parcellaire, mais ils considèrent l'hétérogénéité intra-parcellaire comme un estimateur des degrés de liberté temporelle dont ils disposent lorsqu'ils estiment l'urgence d'une action. Or la quantification des degrés de liberté temporelle à chaque étape de Mildium est l'une des questions importantes auxquelles il faut répondre

h. On a fait l'hypothèse que l'expertise tenait dans une boîte noire. Certes, avant la formalisation, les principes de conception de Mx avaient été énoncé, mais « l'objet » dans l'esprit des experts, n'avait pas de contour précis. En effet les concepteurs pouvaient à tout instant adapter le système pour résoudre une nouvelle difficulté.

i. Rappelons que la relation de simulation n'impose pas que le comportement de Mf se limite aux comportements observés de la Mx.

pour pouvoir en garantir l'efficacité sur le terrain. Les expérimentations menées en 2008 dans plusieurs vignobles méridionaux, par différentes équipes ont encore fait ressortir l'importance d'une mesure de référence de la phénologie. Sans cette référence quantitative, il n'est pas possible d'expliquer les écarts de réactivité des différents opérateurs.

A l'avenir, l'installation d'une station météorologique mesurant les températures dans chaque parcelle expérimentale, serait un moyen simple de régler ce problème, en y associant des modèles de développement phénologique (ceux-ci sont disponibles pour les principaux cépages, cf. Garcia de Cortazar, 2006). Cet outillage pourra être utilisé comme étalon, soit pour chercher à définir la fonction d'estimation de la phénologie utilisée par les experts, ou plus simplement, comme estimation directe de la phénologie et servant de base à la prise de décision.

6.5.3 Généricité de la méthode de validation

A la question de savoir si la méthode développée ici pourra être réutilisée, on peut apporter deux réponses :

- Dans le cadre de la validation du recueil d'un nouveau POD qui reproduirait la séquence « expérimentation, recueil, validation », il est trivial de dire que la mise en œuvre sera possible d'autant qu'on a montré à travers l'exemple de la phénologie comment s'abstraire avantageusement d'une variable de synchronisation pour finalement ne conserver que les séries d'événements et les mises à jour des variables de décision.
- Dans le cadre de la poursuite des expérimentations de Mildium (projet Systèmes Décisionnels pour une Réduction des Traitements phytosanitaires sur vigne Appel à Projets Protection Vigne (A2PV)), il s'agira non plus de valider un modèle vis à vis de comportements constatés, mais de vérifier la conformité des expérimentations vis à vis des contraintes imposées par le POD.
  Les analyses pourront être menées dans un premier temps de manière identique à celles présentées dans l'article précédent. Par la suite, il sera possible, grâce à la massification des données, d'effectuer des classifications sur la base de région, de cépage, et de tenter d'identifier des sous-populations de comportement. La catégorisation deviendra probablement incontournable : la distribution de chaque délai tendant vers la normalité avec la massification des expérimentations.
  Pour pouvoir exploiter les données collectées dans diverses régions sur des expérimentations conduites par différents acteurs, l'étalonnage par la phénologie devient nécessaire si l'on veut pouvoir analyser les séquences de Décision/Action au regard de la qualité de la réponse agronomique. La phénologie permettant de faire le lien entre le temps calendaire, le climat et les invariants biologiques de la culture.
  L'analyse par comparaison aux simulations devra être favorisée dans la mesure où l'on travaille dans différentes régions et que les scénarios climatiques ne seront pas identiques. La comparaison avec la simulation permet de ne considérer que la conformité de l'exécution. Ces analyses devraient permettre d'identifier les contraintes de réactivité aux étapes clé du POD Mildium.