

Plant disease models. Critical issues in development and use *

A. Susi , A. Perini and E. Olivetti ^a

^a*ITC – IRST, Via Sommarive 18 - Loc. Pantè
I-38050 Povo, Trento - Italy
{susi,perini,olivetti}@irst.itc.it*

Abstract: Models which simulate the evolution of a plant disease during the season give important information to assess the seriousness of the situation. This activity precedes the choice of an appropriate action to be implemented for reducing the economical damage. Having effective models is a critical issue in modern agriculture, especially in Integrated Protection and Organic Farming, which rest on a set of plant disease management practices with low environmental impact.

Considerable effort has gone in the study of models for the simulation of plant diseases evolution. Phenology models, population and epidemiological models have been developed for several, diffused diseases. Problems are still open, especially when the aim is that of including these models into decision support systems at use of producers and agronomists. Phenology and population models have to be developed, choosing the most promising techniques. Moreover, requirements such as that of providing justification to the user of the results computed by a model or making the user aware of the accuracy of the model results, become critical.

In this paper we focus on models that address practical plant disease management issues and use mathematical techniques or Artificial Intelligence techniques (especially Machine Learning techniques). We describe relevant examples for each approach pointing out how they deal with critical issues such as adapting a model to different geographical area, or validating and maintaining the model on a long period.

Keywords: Integrated Production; Plant Disease Models; Machine Learning.

1 INTRODUCTION

Current approaches in agriculture, such as Integrated Protection (IP) and Organic Farming (OF), aim at favoring the set up of an agricultural management model characterized by a reduced environmental impact. Concerning plant disease management, these approaches admit a tolerance threshold under which the crop damage can be acceptable. So, the problem of disease management is stated in the following terms: find the most appropriate action, for instance a chemical which is effective at a precise stage of the malady, or a natural defense technique, that maintains the damage under the tolerance threshold.

Critical skills for choosing the best remedy action are early diagnosis, expert knowledge on the behav-

ior of the pathogenic in the specific geographical area and the ability to forecast the disease evolution. In this context, having effective models which can simulate the evolution of a disease during the season is an important issue.

Our interest in plant disease models is motivated by an ongoing project devoted to the development of decision support systems for plant disease management at use of producers and agronomists (the technicians of the local agricultural advisory service).

In particular, we are interested in understanding how models that are currently used by entomologist and agronomists (who participates to the project) can be complemented by using machine learning (ML) techniques (Mitchell [1997]), which proved to be useful in characterizing non-linear and discontinuous phenomena.

We started our analysis considering the following

*The work presented in the paper is partially funded by the Italian Ministry of Scientific and Technological Research.

questions:

- which kind of models are currently used and in which steps in the decision making process of plant disease management are they used?
- how can we choose the suitable modeling technique for a specific disease aspect we need to characterize? That is when should we apply deductive reasoning techniques (e.g. mathematic, logic based) and when inductive reasoning techniques (e.g. ML techniques)?
- what are the main issues to be faced when taking a model from the lab to the field?
- are issues of model maintenance relevant? How have them been addressed?

We discuss here preliminary results of this analysis. We shall note that making a complete state of the art of plant disease modeling techniques and of their applications is far beyond the scope of the paper.

The paper is structured as follows. Section 2 describes the main types of disease models used in practical disease management. Section 3 reports examples of mathematical models. Section 4 discusses ML application examples. Section 5 discusses the questions stated above, in the light of the cited experiences. Finally, conclusions and future work are presented in Section 6.

2 MODELS USED IN PLANT DISEASES MANAGEMENT

The following two types of models are most commonly used in agricultural practices:

Phenology models. They assume that the development process of the organism of interest can be modeled through a sequence of state (the phenology stages of the organisms) transitions (Baumgärtner and Baronio [1988]). The quantity computed by the model is the number of individuals in each stage at a certain time, or the probability of observing the raising of a certain stage in the development of an organism. For instance for the Codling Moth, a relevant pest for apple, the phenology model consider the following main stages: egg, larva, pupa, moth (adult) as described in Butturini et al. [1992]. Each stage can be further detailed in sub-stages, for instance red-ring egg, black-head egg, yellow egg. The model can give information about the life-cycle trend of an insect and about the set of conditions

(environmental variables) for which critical events, such as a transition to a certain stage or relevant increase in the number of individuals in a certain stage, can appear. These indications are useful for advisors and producers to take a specific defense action in the orchards.

Population models. They describe demographic processes and interactions between different organisms. The single organism can be described using a phenology model. These kind of models address the behavior of symbiotic phenomena as well as of antagonist organisms (Madden et al. [2000]). The development and use of population models is considered as the appropriate approach for obtaining adequate insight into demographic processes and for designing management strategies. These models allow to take advantage of the effects of a prey-predator model in the behavior of two populations in the real environment. This is particularly useful in the Integrated Protection approach, where a natural defense techniques which exploits the presence of antagonist organisms, needs to be defined.

3 MATHEMATICAL MODELS

In the following we analyze two phenology models used in our region, which exploits mathematical techniques.

3.1 The Relative Infection Measure (RIM) model

The RIM model is a mathematical method able to simulate the development phases of the *Venturia Inaequalis*, an apple fungus present in many European Countries (Trapman [1994]). The model uses differential equation systems. The target of the model is to describe the infection risk during spring time. In particular the model is based on two different sub-models that are able to detect the development of two different stages for the fungus: the ascospore maturity (determining the quality and the quantity of the risk); the germination and the penetration of the fungus in the fruit (determining the effective damage for the orchards). Both stages are strictly related to the temperature, the rain, the photo period for a given latitude and the quantity of mature ascospores in the environment, a parameter not considered by previous models (e.g. Mills and LaPlante [1951]). The output of the entire process is the severity risk index for a given period of time.

Including the model into an application. The math-

emathical model has been included in the RIMpro, a software system, experimented and used in different European Countries. The system is able to compute the risk degree related to an infection, given the meteorological values and the presence of ascospore ready to pass from the first stage (ascospore maturity) to the second and most dangerous stage (the germination and penetration). The tuning of the RIMpro respect to a specific geographical area can be synthesized in two main steps. Initially it uses a general model based on two sets of parameter: a first set of fixed parameters that are part of the differential equations structure, and a second set of modifiable parameters useful when the model has to be adapted to a particular area, describing environmental variables like temperature, rain, humidity, presence of ascospore in a given period of time. During the practical use, the set of possible values for the modifiable parameters are identified taking into account the experience and the data coming from the entomologists and from the agronomists working in a given area. The use of the system in different environments produces information about the set of fixed parameters giving a validation of the mathematical model structure. Moreover, the feedback from the users contributes also to the software system improvement from the point of view of the model usability and integration with other software systems devoted to the automatic acquisition of the model input parameters values (e.g. the acquisition and selection of the meteorological values).

3.2 The Codling Moth phenology model

Insects are *heterothermic* organisms so their growth is strictly related to the environmental temperature. This property can be exploited for building models that simulate the development stage of an insect. We consider here the pest called Codling Moth (*Cydia Pomonella*, lepidoptera: Tortricidae) an insect species present in Europe and America which represents one of the most dangerous disease for the apple orchards. It generates damages both in the quality and quantity of the production. The individual evolution can be characterized by four major stages: egg, larva, pupa, adult. Butturini et al. [1992] proposes a phenology model based on the *distributed delay model* (Severini et al. [1990]). The model is built using the results of the observations made on a population grown in a controlled environment (e.g. the laboratory). The analysis allows to find the parameter function of a system of differential equations which describes the evolution of every stage of the Codling Moth life cycle.

Including the model into an application. The major difficulty in the use of this model is its need of measurements of environmental variables like rain, wind, humidity and daylight intensity which can be significantly different from one area to an adjacent one. At the present it has been validated only in a limited area in Emilia-Romagna, in Italy, a territory with homogeneous environmental characteristics. The model is still in use by experts only who identify critical situations and possibly alert agronomists and producers. related to the indications given by the researcher to the agricultural advisors; it is very difficult for the advisors' network to directly use indication given by the model and to adapt it to their own area.

4 MACHINE LEARNING MODELS

From a preliminary analysis of the literature on Artificial Intelligence applications to problems in agriculture we found that machine learning techniques are used mainly for classification purposes. We have analyzed a group of interesting works ranging from pest control to plant protection, from the identification of potential risk areas, to the prediction of economic loss in agricultural products commerce. In the following, we mention two applications for which different aspects related to the transition of a model from the lab to the orchard, have been faced. We give also a preliminary description of an application we are currently developing.

4.1 The rangeland grasshoppers management advising system (CARMA)

The CARMA system (CAsE-based Range Management Advisor) is a case-based reasoning system able to predict an expert judgment on the forage consumption due to grasshoppers in a rangeland ecosystem (Branting et al. [1997]), a relevant problem in the western United States. Rangeland grasshoppers infestations can be treated with insecticides, but in many situations the costs of the insecticides application exceed the value of the forage saved. Determining the most cost-efficient response to a grasshoppers infestation requires predicting the forage savings that would ensue from each response and comparing the saving to the cost of the response itself. Grasshoppers population dynamics are poorly understood, so the CARMA system was built upon a protocol analysis of the problem solving done by several experts: entomologists and pest experts, able to provide useful recommendations. The recommendations were supported by explanations given in terms of causal, economic and

pragmatic factors, including a numerical estimate of forage consumed and cost-benefit analysis of the various treatment options.

The structure of the CARMA system of case-based and model-based that cooperates. Initially, a case matching between the current situation (in input) and a historical database describing some environmental and grasshoppers species features, is performed. The result of the matching procedure is then adapted by the model-based subsystem using three model-based adaptation methods: temporal projection, feature adaptation and critical period adaptation. The best results of matching and adaptation are used to compute the probability of future infestation and for an economic analysis about current-year savings and long-term savings to assess cost-efficient responses. Each resulting advice is showed using natural language explanation based on conventional templates.

Including the model into an application. The CARMA system has been used in 23 Wyoming counties and trained by a selected group of local experts, this is required by the matching learning algorithm. In Wyoming counties the CARMA system is used by ranchers (non-expert end users) through a configurable user interface. No feedback is necessary to improve CARMA's performance; the system is trained only in the starting phase.

4.2 The plant protection advisory system (PRO_PLANT)

The PRO_PLANT system, described in Visser et al. [1994], is a protection advisory system whose main goal is to reduce the application rate and frequency of pesticides in farming. The problems on which the system focuses is that of reducing the contamination of ground, maintaining the same economic returns. The system exploits scientific knowledge on phytomedicine and phytopathology as well as practical experiences of plant-protection advisors and farmers. The specific problems addressed by the system, concern cereals diseases (fungi) and herbicides selection. PRO_PLANT includes many subsystems: a database system, a problem solver, a user interface and a data import system. In particular the problem solver is a knowledge-based system made of three modules each one specialized on a single pest: fungicide system, insecticide system, herbicide system. Every module has to perform complex database inquiries to compose the advice which consists of a suitable mixture of pesticides. The herbicide system in particular has been implemented with the help of a neural network to over-

come the problem of a time consuming search into the databases. The use of neural network which can not give motivations for the outcomes, raised the problem of providing an explanation of how and why a decision was determined, in order to prove the reliability of the advice. In a subsequent work (Visser et al. [1996], Visser et al. [1998]) used decision tree techniques for performing rules extraction from neural network parameters.

Including the model into an application. The PRO_PLANT advisory system has been intensively tested in the period between '91 to '95, by plant protection advisors and farmers all over Germany and other European countries. The usage outside Germany has poses some difficulties due to the lack of weather data (or of their availability in different formats) and to the different laws about pesticides. The database subsystem is automatically updated on weather data by automatic weather station and other databases; allowed pesticides data are given by registration offices and field information is collected on site by the user.

4.3 PICO project

PICO is a project currently ongoing at our institute, devoted to the development of a decision support system for apple pest management (see Perini et al. [2002]). The aim of the system is twofold. On one side the entomologist (the expert) is given support while developing a pest model using Machine Learning techniques, given a set of biological and meteo data collected on different orchards, in different seasons. On the other side it supports the final user (agronomist and producer) in tuning the resulting model to the specific environmental characteristics of the territory under his own control. In particular we aim at offer the users the possibility to follow the changes that occur in the insects behavior during the time also supporting the experts during the process of selecting the most relevant characteristics that influence the life of the insects, using some Machine Learning techniques like decision tree, that, given a set of parameters and their values for a particular territory in a certain period of time, are able to produce sets of rules induced by these data.

Figure 1 shows a sketchy view of the PICO architecture. The system is based on a core module that contains the Machine Learning routines (eg. decision tree algorithms) that allows to analyze the data. As a result of this process some rules, describing the behavior of the pest, are produced. They can

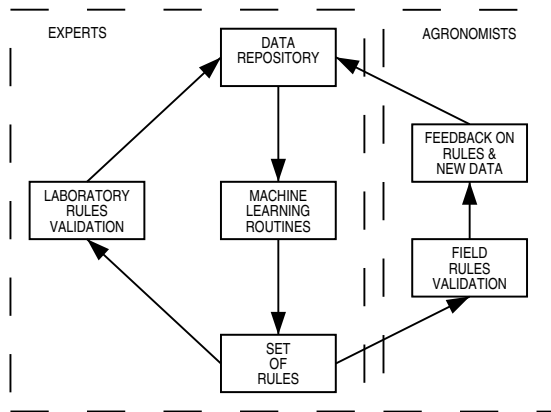


Figure 1: a sketchy view of the PICO system.

be tested in laboratory and in the real environment, by the expert and by the agronomist. As new real data are available from the agronomists they can be included as a new set of cases that are useful to validate and/or adjust the set of rules previously retrieved. Moreover, the set of retrieved rules can be used to define an adaptation methodology for mathematical phenology models such as those described in Section 3.

5 DISCUSSING OUR STARTING QUESTIONS

In the following we examine the questions stated in the introduction, referring to the mathematical models and ML systems described in the previous sections.

Plant disease models in use. The usefulness of disease models in practical activities such as performing infection forecast and identifying warning situations, as well as in higher level activities such as gathering biological data, is commonly recognized, see for instance Rossi et al. [1999].

In Section 2, we presented examples of phenology models used in our country for seasonal infection forecast and alarming. In particular, we noticed that the RIM model is currently being used directly by technicians of our region (see RIMpro application) for the management of the *Venturia Inequalis* fungus, while the phenology model for Codling Moth, at the current stage, is still to be used by experts for notifying alerts during the season.

Modeling techniques, pro and cons. Mathematical models seems to be effective for simulating the development of an organism or of a population of organisms in a controlled environment (the lab), but they require to be carefully adapted to real environ-

ment, taking into account possible territorial heterogeneity, and this can require a considerable effort. In this regard we have mentioned the effort in tuning the RIMpro system to the Trentino area and the experience, which is currently going on, of adapting the Codling Moth model of the Emilia-Romagna region to the Trentino area.

A possible way to simplify the adaptation activity is the one proposed in CARMA, where mathematical modeling and machine learning techniques cooperate. The approach used in CARMA exploits case-based reasoning for finding appropriate adaptation strategies that will guide the model-based adaptation.

Machine learning techniques can be effective when performance issues are to be addressed. In fact, after a lengthy learning period (bootstrap phase), usually machine learning algorithms perform efficiently in computing the outcomes. In the PRO_PLANT system, for example, a neural network is preferred to standard query construction and database retrieval.

In general, ML techniques seems to be often preferred when highly non-linear and discontinuous phenomena need to be modeled. They introduce rough approximation but are still able to distinguish between different macro-behavior of the system. In situations of low degree of non-linearity and without discontinuities, pure mathematical models can give results with a better quality respect to the machine learning based models.

From the lab to the orchard. From the analysis of the experiences presented in sections 3 and 4, we found that two main sources of complexities should be considered when including a model into an application, namely: the definition of mechanisms for making the model easily adaptable to different environments and the inclusion of mechanisms for obtaining justifications of the results produced by the models at use of non-expert people (e.g. producers and technicians). For instance, in PRO_PLANT, rules extraction techniques were used to justify the neural network outcomes to the user.

Maintenance issues. Several specific problems described in the previous sections can be brought back to the more general issue of maintenance. For instance the approach followed for adapting the RIMpro system to the Trentino area: the RIM model is tuned respect to environmental variables according to a yearly revision of the model based on the results of the comparison of the model forecast and the real data measured during the season.

Different aspects of maintenance have to be faced in the case of systems using data which are periodically updated. For example in the PRO_PLANT system, the database module gets weather information everyday automatically and needs to be feed of pesticides updates (new products or law changes) twice a year. In the PICO project we are designing mechanisms for including new seasonal data that should improve the previously extracted rules.

6 CONCLUSION AND FUTURE WORK

The paper presented preliminary results of an analysis of plant disease models, such as phenology and population models, that exploit mathematical techniques. The aim was that of identifying critical issues in model development and in their inclusion into decision support systems at use of technicians and producers dealing with seasonal disease management.

Among our objectives: to understand the opportunity of complementing mathematical models with machine learning based models; to examine other experiences where issues such as making model results understandable and reliable by non-experts or model maintainability have been faced.

The results of this analysis will influence the design and development of a decision support system for technicians of the advisory service of our region when dealing with the management of critical pests for apple.

REFERENCES

- Baumgärtner, J. and P. Baronio. Modello fenologico di volo di *Loboesia botrana* Den. & Schiff. (Lep.Tortricidae) relativo alla situazione ambientale della Emilia-Romagna. *Bollettino dell'Istituto di Entomologia "G. Grandi"*, 43: 157–170, 1988. in Italian.
- Branting, L. K., J. D. Hastings, and J. A. Lockwood. Integrating Cases and Models for Prediction in Biological Systems. *AI Applications*, 11(1):29–48, 1997.
- Butturini, A., R. Tiso, and E. De Berardinis. Influenza della temperatura sullo sviluppo di *Cydia pomonella* (L.)(Lepidoptera: Tortricidae). *Bollettino dell'Istituto di Entomologia "G. Grandi"*, 47:123–134, 1992. in Italian.
- Madden, L. V., M. J. Jeger, and F. van de Bosch. A Theoretical Assessment of the effects of Vector-Virus Mechanism on Plant Virus Disease Epidemics. *Phytopathology*, 90(6):576–594, 2000.
- Mills, W. D. and A. A. LaPlante. Diseases and Insects in the Orchard. In *Cornell Extension Bulletin*, volume 711, pages 21–27, 1951.
- Mitchell, T. *Machine Learning*. McGraw Hill, 1997.
- Perini, A., A. Susi, and F. Giunchiglia. Designing Coordination among Human and Software Agents. Technical report, ITC-IRST, january 2002.
- Rossi, V., P. Battilan, P. Racca, and S. Giosue. Supporti forniti dai modelli di simulazione nella gestione fitosanitaria delle colture. *Frustula Entomologica*, 22:71–76, 1999.
- Severini, M., J. Baumgärtner, and M. Ricci. Theory and practice of parameter estimation of distributed delay models for insect and plant phenologies. In *Meteorology and Environmental Sciences*, pages 674–719. World Scientific, 1990.
- Trapman, M. Development and evaluation of a simulation model for ascospore infections of *Venturia inaequalis*. *Norwegian Journal of Agricultural Sciences*, 17:55–67, 1994.
- Visser, U., R. Nayak, and M. Wong. Rule extraction from trained neural networks and connectionist knowledge representation for the determination of pesticide mixtures. In *Proceedings of the Australian Conference of Neural Networks (ACNN 98)*, pages 138–142, Brisbane, Australia, February 1998.
- Visser, U., A. Tickle, R. Hayward, and R. Andrews. Rule-Extraction from trained neural networks: Different techniques for the determination of herbicides for the plant protection advisory system PRO_PLANT. In *Rules and Networks*, Proceedings of the Rule Extraction From Trained Artificial Neural Networks Workshop, Brighton (UK), pages 133–139, Brisbane, April 1996. Society for the Study of Artificial Intelligence and Simulation of Behavior, QUT.
- Visser, U., U. Voges, and U. Streit. Integration of AI-, Database- and Telecommunication-Techniques for the Plant Protection Expert System PRO PLANT. In *Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*, volume 7 of *International Conference of IEA/AIE*, pages 367–374. Gordon and Breach Science, June 1994.