

AN ARTIFICIAL INTELLIGENCE METHODOLOGY FOR THE ADAPTATION OF AGRICULTURAL MODELS

Gianni Jacucci, Mark Foy and Carl Uhrik

*Laboratorio di Ingegneria Informatica (LII)
Dipartimento di Informatica e Studi Aziendali (DISA)
Università degli Studi di Trento
Via F. Zeni 8
I-38086 Rovereto (TN) Italy*

Abstract: An AI adaptation methodology designed to assist in transporting agricultural models between regions is presented. Models frequently need adaptation when transported because models developed in one region often do not produce valid results when used in a different region. The methodology prescribes the linkage of a genetic algorithm to a model. This makes the model more robust because it is able to *adapt* to the region in which it is being used. This methodology has been implemented within a DSS, and preliminary testing indicates this methodology has the ability to allow agricultural models developed in one area to be effectively utilized in other regions.

Keywords: Agriculture, Adaptation, Genetic algorithms, Intelligence, Methodology, Models, Parameter optimization.

1. INTRODUCTION

In the domain of agriculture, the utilization of already developed models in a broad area is often hindered by one or more factors. One frequent factor which impedes transportation is model inaccuracy. For example, when models that perform well in one region, are transported to be used in a different region, they often do not give accurate output (such as, recommendations, results, and/or indicators) in their new environment (i.e., when they are run in a new region).

This is one of the major difficulties of *model technology transfer*. To address this difficulty, an artificial intelligence (AI) methodology is proposed. At the heart of this methodology is a genetic algorithm (GA) (an AI search technique) which is linked to the agricultural model engine (e.g., a risk assessment, previsionsal, or crop growth model). The general component created by this combinational methodology will here be called an 'Agricultural Model-GA' or an AGMOD-GA.

The following sections will describe the overall structure and elements of this methodology, the generic component created by following this methodology, and discuss the application of this methodology.

2. DESCRIPTION OF THE AI METHODOLOGY FOR ADAPTING AGRICULTURAL MODELS

The theory of this adaptation methodology is that by utilizing historical data from a particular region, a model's parameter settings can be *adapted* so that the new parameters allow the model to work well in the particular region. This adaptation is done by trying to match the model parameter settings to the particular region. To find matching model parameter settings, intelligent search is performed which utilizes historical data as part of the objective function.

Overall, by following this methodology, a component will be created which can search for good model parameter settings such that when the given model is applied and run at the location in question, the output values given will be consistent with the historical outcome data; moreover it is hoped that this will also allow the model to be generally used in this region, producing accurate output values on data which it has not seen. The component that performs this search/adaptation can be called an expert system component since it intelligently modifies and adjusts a model to work in a new location in the same way an expert would modify and adjust a model.

Additionally, it should be emphasized that this methodology is particularly appealing because it is not a strictly empirical or analytical, but both. That is, this methodology does not perform a search to *fit the*

historical data from a particular location into an empirical algorithm; rather it performs the search in a larger context, *fitting the model parameter settings* to a particular location. Therefore, the resulting instantiation of the adapted/localized agricultural model (with the new parameter settings inside) is as good (or as bad) as the original model; consequently, if the model is biologically significant (e.g., if it simulates biological events) then this is not lost by this adaptation methodology since the model is used in the same form (i.e., the structure of the model is left intact), only the model parameter settings are changed.

3. ELEMENTS OF THIS AI METHODOLOGY

In general, this methodology prescribes the utilization of:

- (i) historical situation data,
- (ii) historical outcome data,
- (iii) the agricultural model, and
- (iv) an intelligent search method (in this case, a genetic algorithm, also called a GA, which is an artificial intelligence search technique).

Historical situation data is the basic data required by the model in question. In the domain of agricultural models, this often includes meteorological data since this is frequently an important input to the model. In this methodology, the more historical situation data that is available, the better.

The presence of historical outcome data plays a large part in how accurately a model will be adapted using this methodology. This is due to the basic fact that models accept situations and compute outcomes. The historical outcome data will be used to fit the model parameter settings to the new region in question. Therefore, when constructing a component using the methodology described here, it must be possible to match model outputs to some combination of outcomes and/or events in the real-world (and there must be one-to-one correspondence). For risk assessment models, historical outcome data regards the occurrence of fungus or pest problems in past years (epidemiological data); or for crop growth models, historical outcome data regards crop yield in past years.

In this methodology, the agricultural model (i.e., the engine or core of this model) is fundamental because it will be used to obtain evaluations of how well particular model parameter settings work in the given region (i.e., with the given data). In particular, the intelligent search method will repeatedly call upon this model engine as it constructs new model parameter settings that need to have their worth evaluated. This methodology has the capability to address many types of agricultural models: risk assessment models, damage prediction models, crop growth simulation models, etc.

The intelligent search method is an important part of this component because, in this particular domain of agricultural models, knowledge of the domain is often hard to codify (i.e., 'rules of thumb' are vague and difficult to construct), and the selection of an intelligent search method can help to alleviate this difficulty. This is due to the fact that intelligent search methods do not rely on 'rules of thumb', rather, rules are not required and these methods can actually facilitate the user in identifying 'rules of thumb'.

The selection of the actual intelligent search method to be employed was made among the following possible methods: hill-climbing, simulated annealing, and genetic algorithms (GAs). In the end, GAs were selected as the most desirable method based on the arguments given in Goldberg (1989), Grefenstette (1985, 1987), and Schaffer (1989).

4. AGRICULTURAL MODEL-GA (AGMOD-GA)

To allow a GA to search the space of an agricultural model's parameters, the agricultural model is linked to a GA, and the GA uses the model as the evaluation function. Furthermore, the model uses the historical situation data (as this is necessary to run the model in the given historical years), and the GA additionally uses the historical outcome data (discussed earlier) in combination with the output of the model. Whenever the GA wants to evaluate one instance of model parameter settings, the agricultural model is called, and the final outcome is returned through an objective function to the GA so that a fitness can be computed. The resulting general component created by employing this combinational methodology can be called an 'Agricultural Model-GA' or an AGMOD-GA.

Figure 1 illustrates an agricultural model and a GA linked to form an AGMOD-GA.

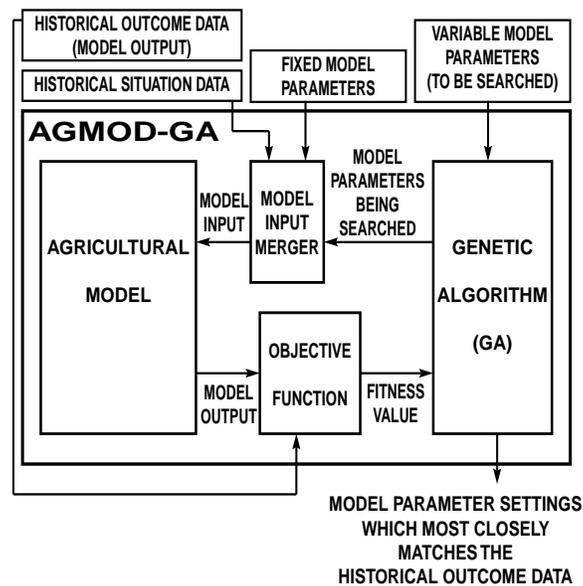


Figure 1. Structure of an AGMOD-GA

The function of the AGMOD-GA is to find near-optimal model parameter settings for the given desired behaviour (i.e., matching the given historical outcome data). There are three main steps involved in the execution of a typical AGMOD-GA. First, the agricultural model and the GA are initialized. With the simple GA implemented in this case, an entirely random initial set (i.e., population) of parameter settings is generated. This has the effect of starting the search in a number of different random points in the space. A collection of random starting points does not have a negative effect on GA performance because a GA searches from many different point at the same time, not just from one point. The second step in a typical AGMOD-GA is the fitness computation (i.e., valuation of each population member's worth). This involves taking each GA population member and executing one or more model executions using the model parameter settings represented by this member. These model executions utilize the user-provided historical situation data, with one execution initiated for each one of these sets of data. The outcomes from these model executions are compared against the user-provided historical outcome data (which is the target). The further the model outputs are from the historical outcome data, the lower the fitness, and inversely, the closer the model outputs are, the higher the fitness. This fitness evaluation step is executed many times because new population members are continually being generated by the GA. Fitness evaluation is usually continued until the GA has converged on a suitable optimal or quasi-optimal solution. The last step is the evolution of the GA population. This involves applying operations to the population members. The three operators used in a typical GAs are reproduction, crossover, and mutation. They act by treating the GA bit strings (which represent model parameters) in a way analogous to the evolution of chromosomes in genetics (Goldberg, 1989).

5. APPLICATION OF THE METHODOLOGY

An example of applying the above described methodology to create a *real* AGMOD-GA component is given. To put this example in context, the project under which this methodology was developed (Project SYBIL) and the DSS which utilizes this methodology (one of the SYBIL DSSes which focuses on grapes and apples) are outlined.

5.1. Description of Project SYBIL

EC Project SYBIL (consisting of five partners from four countries) involves the implementation of computerized decision support systems (DSSes) to assist farmers in intelligently governing their crops such that environmental impact is reduced and economic returns are increased. Existing agro-meteorological computer models from multiple sources are integrated into the portable, user-friendly DSSes designed to assess the risk of a crop to pest and fungus damage. By evaluating this risk, the farmer has the option to apply pesticides and fungicides only when needed and avoid using these, often environmentally damaging, chemicals blindly on a regular basis or when the risk of pest and fungus damage is small. This evaluation has the potential to save the farmer both time and money because expensive chemicals will not be applied when they do not benefit the crop.

5.2. DSS Description

The SYBIL DSS described here (which includes the AI methodology discussed) is targeted to grape and apple growers. Figure 2 displays the first screen of this DSS. The model focused on for illustrating the adaptation methodology described here will be the P.R.O. model for grapes.

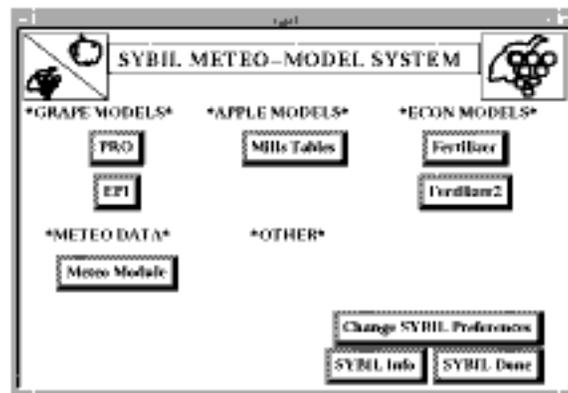


Figure 2. The System's Main Screen

5.3. Description and Origin of the P.R.O. Model

The P.R.O. (**Plasmopara Risikoproggnose Oppenheim** or **Plasmopara Risk Oppenheim**) model for grapes was the first model selected for the application of the adaptation/localization methodology discussed above. This analytical model is a biological life cycle model that simulates the infection and growth of downy mildew (viz. *Plasmopara viticola*, also called peronospora) on grape vines based on meteorological conditions. The model was developed in Rheinhessen, Germany by Dr. Georg K. Hill (Hill 93). It was designed to help growers determine when it is necessary to spray grape vines against peronospora.

The model had been used by multiple Rheinhessen region farmers with good results; that is, the information provided to the farmer has assisted in the making of intelligent decisions about when to perform the first spray of the season against peronospora. The goal is to overcome the habit (which is not based on temporal information) of performing the first spray early in the season (possibly around May), which is often before it is necessary. This goal is approached by using the P.R.O. model to produce interpreted-operational-temporal information (i.e., useful up-to-date information about the status of the peronospora growth), then examining this information, and deciding if it necessary to spray at the current moment, or if spraying can be delayed (possibly many weeks beyond when growers would traditionally perform the first spray) because the grape vines are not currently at risk to being damaged by peronospora. In the common cases where spraying can in fact be delayed beyond when an agriculturalist would normally spray, the overall number of interventions and amounts of chemicals sprayed on the crop are reduced. Agriculturalist using this model in the region around Rheinhessen have been able to save between one and four spray applications per year, with an average saving of two (Hill 93).

5.4. Results of Transporting the P.R.O. Model

After deciding that the P.R.O. model was a good choice for inclusion into the DSS (and therefore a good choice for trying to transfer this model between countries), an instantiation of the model (with only small changes so that the model would accept other types of meteorological data) was programmed into the DSS, and test runs were made with various data from other regions (e.g., Würzburg, Germany and Trentino, Italy). Upon running these tests, it was found that the output values (which in the case of the P.R.O. model are: the primary infection date, the end of the incubation period, a list of special night occurrences, and a recommended spray date), were inaccurate in the new regions. That is, the P.R.O. model outputs were rejected by agricultural experts based on their historical epidemiological data (more generally, their historical outcome data) and general knowledge of when epidemiological events occur in their regions.

For example, Table 1 displays the results from running the P.R.O. model with data from an area inside the Trentino region of Italy. As this table shows, the dates produced by the original P.R.O. model using original model parameter settings (i.e., model parameter settings selected by Dr. Hill for the Rheinhessen area) (these dates shown in the column titled "Case 1") for data coming from Trentino, only approached the

dates known to be correct from observations done by agricultural experts in Trentino (these dates shown in the column titled "Actual Dates") for the primary infection dates (rows titled "Prim Inf 19xx"). For the recommended spray dates (rows titled "Rec Spray 19xx"), the model could not even produce estimates of this date (with this data from Trentino) for two out of the three years in which actual dates were available for comparison. Therefore, the model in this state is of little use to agriculturalists in Trentino since it is generally not able to produce an accurate estimate of the recommended spray date. Additionally, the difficulties observed in this case also held true for data taken from other regions, so overall the P.R.O. model was problematic because it did not give accurate output when run in regions external to where it was developed.

Table 1. Results from running the P.R.O. model; Case 1 uses data from the Trentino region of Italy with original parameter settings used by Dr. Hill in Rheinhessen, Germany; actual dates in Trentino were only available for three years; date 12/31 is used to indicate the model never reached this date; difference is in absolute days

	Actual Dates	Case 1	Difference between Case 1 and Actual
Prim Inf 1988		5/3	
Rec Spray 1988		12/31	
Prim Inf 1989		5/13	
Rec Spray 1989		7/3	
Prim Inf 1990	5/24	6/5	12
Rec Spray 1990	6/7	12/31	207
Prim Inf 1991	5/11	5/11	0
Rec Spray 1991	6/7	12/31	207
Prim Inf 1992	5/5	5/23	18
Rec Spray 1992	6/1	7/10	39
Prim Inf 1993		4/30	
Rec Spray 1993		12/31	
Avg. Difference			80.50

In addressing this problem, it is proposed that this model could be adapted to local conditions and that the problems stem from the fact that the model parameter settings were custom tailored to the region where it was developed (in this case, Rheinhessen, Germany). In addition, it is desirable to keep the same basic overall structure of the P.R.O. model because it had been proven valid and useful in the past. Therefore, the above described methodology was applied and an instantiation of an AGMOD-GA for the P.R.O. model was created, calling the new component the PRO-GA.

Table 2. Results from running the P.R.O. model; Case 1 is the same as shown in Table 1; Case 2 uses data from the Trentino region of Italy with parameter settings found by the PRO-GA using all available historical data; date 12/31 is used to indicate the model never reached this date; differences are in absolute days

	Actual Dates	Case 1	Difference between Case 1 and Actual	Case 2	Difference between Case 2 and Actual
Prim Inf 1988		5/3		5/3	
Rec Spray 1988		12/31		5/11	
Prim Inf 1989		5/13		4/19	
Rec Spray 1989		7/3		5/13	
Prim Inf 1990	5/24	6/5	12	5/24	0
Rec Spray 1990	6/7	12/31	207	6/6	1
Prim Inf 1991	5/11	5/11	0	5/12	1
Rec Spray 1991	6/7	12/31	207	6/7	0
Prim Inf 1992	5/5	5/23	18	5/4	1
Rec Spray 1992	6/1	7/10	39	5/30	2
Prim Inf 1993		4/30		4/30	
Rec Spray 1993		12/31		6/23	
Avg. Difference			80.50		0.83

By running this PRO-GA component with data from a particular region, new model parameter setting values can be found which should allow the model to give accurate output values (i.e., recommendations, results, and/or indicators) for the region in question. Upon running the PRO-GA with three years (1990, 1991, and 1992) of historical situation data (in the case of the P.R.O. model, meteorological data) and historical outcome data (in this case, epidemiological data describing primary infection dates and

recommended spray dates) from an area inside the Trentino region of Italy, new model parameter setting values were derived. Table 2 shows the results of using these new model parameter settings inside the P.R.O. model, again running with data from Trentino.

As expected, this table shows that the model parameter settings derived by the PRO-GA component allow the P.R.O. model to function with significantly increased accuracy on the Trentino data. The absolute average difference in days (i.e., the absolute average of how far off the model output is from actual dates) (in the row titled "Avg. Difference") between actual dates (column titled "Actual Dates") and dates produced by the running the P.R.O. model (column titled "Case 2") drops from 80.50 days to 0.83 days when new parameter settings derived by the PRO-GA are used instead of the original parameter settings. Generally, an average difference (i.e., an accuracy error) of 0.83 days is not significant or a critical inaccuracy, and this variation is well within the tolerable limits.

Because this adaptation is performed with all available historical sets of data, this type of behaviour is generally expected since accuracy is tested on the same set of historical model data (in this case, meteorological data) that was used for adapting. This is still an important result because it shows that the model can be fit to the entire set of data, and that it is possible to find parameter settings that will give accurate results over many years. On the other hand, a better verification of this methodology is to adapt the model using one subset of historical model data, and then test the accuracy of the adapted model (i.e., the model with the new parameter settings) by running the model on a different subset of historical data.

Unfortunately, the currently available historical outcome (viz. epidemiological) data is very limited (three years of data from Trentino); therefore only a few small subsets of historical model data can be formed (in this case, only three interesting subsets: (1991, 1992), (1990, 1992) and (1990, 1991)). Therefore, it is not possible to test newly derived parameter settings as extensively as desired, but three tests have been performed, the results of which are shown in Table 3.

Table 3. Results from running the P.R.O. model; Case 1 is the same as shown in Table 1; Cases 3, 4, and 5 use data from the Trentino region of Italy with parameter settings found by the PRO-GA using three different subsets of the available historical data; Case 3 used only 1991 and 1992 data; Case 4 used only 1990 and 1992 data; Case 5 used only 1990 and 1991 data; bold results indicate results from data sets which were not used in adaptation; date 12/31 is used to indicate the model never reached this date

	Actual Dates	Case 1	Case 3	Case 4	Case 5
Prim Inf 1988		5/3	5/3	5/3	5/3
Rec Spray 1988		12/31	5/11	5/11	5/11
Prim Inf 1989		5/13	4/19	4/19	4/19
Rec Spray 1989		7/3	5/13	5/13	5/13
Prim Inf 1990	5/24	6/5	5/24	5/24	5/24
Rec Spray 1990	6/7	12/31	6/4	6/6	6/6
Prim Inf 1991	5/11	5/11	5/12	5/12	5/12
Rec Spray 1991	6/7	12/31	6/7	6/17	6/7
Prim Inf 1992	5/5	5/23	5/4	5/4	5/4
Rec Spray 1992	6/1	7/10	5/30	5/30	5/23
Prim Inf 1993		4/30	4/30	4/30	4/30
Rec Spray 1993		12/31	6/23	6/23	6/23

Even with the limited amount of historical data, these results are still interesting and significant. For example, when adapting the model using historical outcome data from 1990 and 1991 (shown in the last column of the table) the average difference between actual results and P.R.O. model produced results using the new parameter settings is only 2.00 (not shown in table, but computed separately). This is a lower accuracy than the average difference shown in Table 2 using an adaptation with all three historical data sets, but it is believed still to be quite adequate.

It is believed that if an increased amount of historical outcome data was available, and this was used in the adaptation process, that the adaptation should become even more robust, increasing the probability that accurate results are produced in the future when the model is running in real-time, giving output values to the agriculturalist for use in making intelligent crop management decision.

6. CONCLUSION

The artificial intelligence (AI) methodology discussed addresses *model technology transfer* (i.e., the moving of functional and useful agricultural models that are developed in one location to a new location so they can be used in this new location). In particular, it addresses one of the major difficulties within this area, namely, *model accuracy*; that is, it addresses the instance when a useful model is transported from a

region where it is functioning accurately (viz. producing accurate recommendations, results, and/or indicators) to a new region where it subsequently does not function accurately.

The methodology employs four main elements, with a genetic algorithm (GA) at the center. By employing this AI component in conjunction with the engine of an agricultural model and historical data, model parameter settings can be adapted to new locations, allowing the model to give accurate results when run in the new location. Specifically, the module created by this methodology can be applied to *localize* models by deriving new model parameter settings that can be employed in the particular location to give good suggestions/decision support.

With the assumption that model technology transfer is an advantageous action (refer to (Jacucci, et al., 1994) for an elaboration of advantages and disadvantages in transporting models between regions), this AI methodology has been found to efficiently address this issue and improves the current state-of-the-art in model technology transfer.

This has been shown through an example which describes the utilization of this methodology within one of the decision support systems (DSSes) developed under EC Project SYBIL. In particular, this DSS was designed to provide temporal information to assist grape and apple agriculturalists in the management of crops with respect to controlling fungus and pests. Specifically, this methodology has been applied to an instantiation of the P.R.O. model that is programmed into one of the SYBIL DSSes. This model addresses the infection and growth of *Plasmopara viticola* on grape vines, and has the capability to provide information to a farmer so that decisions regarding when to apply fungicides are made more intelligently. Due to difficulties in transporting this model to run in regions outside where it was developed, our methodology to adapt model parameter settings was employed. The testing of new model parameter settings produced by this *adaptation* showed that this methodology has great potential to localize model parameter settings, and this should assist in achieving the goal of making sound models more widely available.

ACKNOWLEDGEMENTS

Funding for the SYBIL project (PL 900615) is provided by the EC CAMAR Program (Competitiveness in Agriculture). The authors wish to thank Dr. Val Reilly, SYBIL project supervisor from the European Commission, for his support and encouragement.

Furthermore, the authors would like to thank all EC Project SYBIL partners: Technische Universität München-Weihenstephan (Technical University of Munich) (TUM) (D), Bayerischen Staatsministerium für Ernährung, Landwirtschaft und Forsten (BALIS) (D), Association De Coordination Technique Agricole (ACTA) (F), and Danish Institute of Plant and Soil Science (DIPSS) (DK). In particular, thanks go EC Project SYBIL's coordinator Dr. Johann Bergermeier at BALIS in Munich, Germany and Dr. J.V. Herrmann at Bayerische Landesanstalt für Weinbau und Gartenbau in Würzburg, Germany for their assistance with the P.R.O. model.

REFERENCES

- Goldberg, D.E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, Reading, MA, USA.
- Grefenstette, J.J., ed. (1985). *Proceedings of an International Conference on Genetic Algorithms and Their Applications*. U.S. Navy Center for Applied Research in Artificial Intelligence, USA.
- Grefenstette, J.J., ed. (1987). *Proceedings of the Second International Conference on Genetic Algorithms*. L. Erlbaum Associates, USA.
- Hill, G.K., Breth, K., Spies, S. (1993). The Application of the P.R.O. - Simulator for Minimizing of *Plasmopara* Sprays in the Frame of an Integrated Control Project in Rheinhessen/Germany. *Vitic. Enol. Sci.*, **48**, 176-183.
- Jacucci, G., Foy, M., and Uhrig, C. (1994). SYBIL DSS:Localization of Agricultural Risk Assessment Models. In: *Proceedings of Decision Support 2001 (DSS2001) (The 17th Annual Geographic Information Seminar & The Resource Technology '94 Symposium)*, American Society of Photogrammetry and Remote Sensing.
- Schaffer, J.D., ed. (1989). *Proceedings of the Third International Conference on Genetic Algorithms*. Morgan Kaufmann Publishers, Inc., USA.