2014 International Conference on Intelligent Agriculture
IPCBEE vol.63 (2014) © (2014) IACSIT Press, Singapore
DOI: 10.7763/IPCBEE. 2014. V63. 3

Correlation between Factors for Crop Grow towards Modeling a Complex System

M.G. Lugo-Ibarra, Dora-Luz Flores, Claudia M. Gómez, and Graciela Guerra-Rivas [+]

Autonomous University of Baja California

Abstract. We examined the correlations between the factors leading to the development of crops, techniques and methods considered for predicting pests and diseases among other situations. Climatic changes and human influence have deteriorated crop growth. They have caused economic losses and decreases in product quality; besides the excessive use of fertilizers and other products have greatly affected the development of the field and even human health. There are a number of prediction models for different situations Agricultural sector, some of them are: expert systems and tools to support decision making, predicting crop growth, beginning of planting, pests, diseases, economic costs among others. It is complex to characterize all the factors involved in crop growth due to the extensive amount of data included and heterogeneity between them. However it is possible to design predictive models using strategies that have been used elsewhere (methods, techniques, hybrid and selection of factors). This article shows a compilation of recent research, including factors such as (water, soil and weather conditions, etc.), methods and techniques used in predictive models (evolutionary algorithms, neural networks and mathematical methods, etc.). This is part of the theoretical fame work of the work in progress for the development of a predictive model of pests and diseases in open field crops and protected crops.

Keywords: Pest and diseases, crop, complex system, forecasting.

1. Introduction

The agricultural sector is considered as one of the primary activities in the world. The demand for the various products in this sector, establishes a framework to increase research and technological developments. The main areas of interest are product quality, increased production, pest and disease control, resource management (water, soil, climate, etc.), crop monitoring, and infrastructure, including the human factor and its relationship to the environment has meant a major factor study. In recent years, there has been a lot of developments in technology, both hardware and software, arising mainly as a support for decision-making in the agricultural sector. In this context there is a great number of methods and techniques used to develop these models or applications that offer services such as; prediction of pests and diseases crop performance, control on the beginning of planting or harvesting, control of density index of crop, pest, diseases and weeds; concentration of mineral resources (water and soil), among others.

The heterogeneity of the factors, the variability of the parameters and the large number of tools (methods and techniques) that exist to solve the different situations that arise in crops, promote the level of complexity of prediction models for this reason this article shows the main factors affecting crops, and the methods and techniques that have been used for prediction. This framework is part of the work in progress to develop a predictive model of pests and diseases.

This article is composed as follows: The first part shows the reader predictors that have been used so far.

1.1. Crop phonological stages and weather conditions

[+] Corresponding author: M.G. Lugo-Ibarra. Tel:+52 646 174 43 33
E-mail: lugo.maria@uabc.edu.mx

Contribution Cultivation: There are a variety of crops around the world. Each crop developed in different stages or phases of growth. These growth stages are conditioned by four environmental or climatic factors: temperature, relative humidity, light and carbon dioxide level. These factors combine to varying degrees for plants to perform their functions, and the absence of any of them ceases its metabolism [1]. The phenological stages are defined as periodic phenomena of living beings and their relationship to environmental conditions. These stages are very variable, i.e., there is a different number of phonological for each live crop. However, it is possible to associate the stages and organize them into small groups. Thus it will be relatively easy to identify the stages of the crop and to conduct tests for the development of the plant.

Environmental: The concentration of greenhouse gases has increased in recent decades, eroding the quality and size of the product of crops [2]. Changes in weather conditions greatly affect crops, both in the open and under protected cultivation.

Water and soil: The advent of the Green Revolution increased the development of synthetic fertilizers, pesticides and irrigation control in crops. However, these techniques caused contamination in water and soil, affecting product quality. From this arises the integrated pest management (IPM) and disease (MID), whose purpose is providing harmony between users and products used for the handling and care of crops.

Pests and diseases: The organisms causing crop losses have the feature to regulate their body temperature, and are exposed to climate change, so their rate of development depends on them. For this reason when the insects cause damage to crops and exceed the economic injury level (EIL), an unstable system is produced triggering environments for the development of pests and plant diseases.

1.2. About the techniques and methods

Models have now been developed able to classify and quantify parameters for the different effects of the factors on the crop. According to the variable of interest in the prediction, the models can be mechanical, empirical [3], qualitative and / or quantitative. The mechanic is characterized by greater explanatory power, statistical methods used to describe the relationships between environmental variables and increasing the intensity of the disease. This generally based on field experiments, where production conditions are simulated and experimental results can be used to direct the construction-controlled model of interest. Empirical models are characterized by the cause / effect, however can be used to make inferences about the underlying biology of the studio system.

Both types of modeling, mechanical and empirical, are used to identify parts of a cycle of the disease requiring further exploration based on experiments. Likewise, there are currently almost as many mechanical and empirical models, which are successful. They consider the weather as significant factor in the model, because of its close correlation with crop performance [1, 4].

To perform the assessment of a model, it is necessary to replicate plot level investigations to determine the performance of the model-based decisions and compare them with the corresponding in practice. There are a number of predictive models using various techniques and / or methods to support decision making to control the factors around the crop. Among which are evolutionary algorithms, neural networks, statistical techniques [5], hybrid [6-7] among others (Table 1).

1.3. Importance of the study of various factors in open field and protected crops

This occurs in both crops protected as open countryside. Although protected crops have gained great influence in recent decades in the market due to obtaining greater control over the cultivation, the parameters of the factors considered for the prediction of any event around the crop, both in protected crops as are generally concentrations of pollen, humidity [1], open field, solar radiation, climate and microclimate [5], relationship between pathogens and plant (temperature, water and soil nutrients and light effect) nitrate concentration in ground water [6], land topography, soil quality, water, [8], grasses (relationship between agriculture and livestock), CO2 [9], air velocity and moisture absorption, resource deterioration factor derived from human and animals [10], use of pesticides and / or fertilizers, reproductive cycles of pests [11], pest resistance in crops [12], type of region, pest and disease resistance due to the application of pesticides and herbicides, [13], natural crop enemies depending on the season [14], increased level of NDE economic damage due to the relationship between the incidences of pests depending on phenological stages [15], from a large number of parameters associated with each factor (Table 1).

1.4. Data analysis techniques and methods

The amount of data generated situation is clear, when it is required to analyze, based on more than one feature sets with large numbers of abstract objects (data p-dimensional). Cluster analysis of data is also known as analysis clustering (cluster analysis) of segmentation, unsupervised learning, learning without a teacher (within the pattern recognition), numerical taxonomy (biology and ecology), typology (in social sciences) and partitioning (in graph theory), shows the intrinsic organization of subsets, consisting of data with similar features [16-21]. Based on several authors summarize the process of data clustering as a general methodology and a powerful conceptual framework and algorithm for the analysis and interpretation of data (see Table 1).

Meanwhile [18] believes that the groups mentioned above should have the following properties: homogeneity within the cluster, that is, the data belonging to a cluster should be as similar as possible. Heterogeneity between clusters, i.e. data belonging to different cluster should be as different as possible.

Thus the grouping process data, it may be determined as a set of observations of a physical process, wherein each n measured features forming an n-dimensional vector $xk = [xk1, \ldots xk2, XKN]$ T, xk E R. Making a set of observations is identified by $x = \{x\ k \mid k = 1,2, \ldots N\}$ and is represented by a matrix of N X n.

The above is disclosed for a scenario, where the goal would be to find the family of C classes that reveal hidden in the data structure through the cluster centers represented by a vector $vi = \{vn1, vn2, \ldots vnc\}$ for each element of the vector vi, are the values of each of the variables measured in physical processes and are characterized to vector type of each cluster.

To facilitate the identification of both behavior patterns as the distance between each factor and / or variable, methods and / or techniques may be applied such as similarity measures, useful for data pertaining to isolated or compact clusters, Hamming (Manhattan) and Euclidean Mahalanobis [11], for geometric distance calculations between p-dimensional [22], distance Minkowski, for highly correlated data, different variances and a different range [23].

As mentioned above, the distance relationship between each factor and / or variable is of utmost importance to make predictions for this reason and based on the literature review (Table 1) showing the main predictive factors such as: cultivation, resources (water, soil), pests and diseases, in addition to the human factor. Which were used to predict various situations around the crop will be used to define the layout of the prediction model of pests and diseases.

2. Study material

It is necessary to form clusters of the factors affecting crops, with the characteristics of homogeneity within them and heterogeneity outside of them. The following describes the methodology proposed to initiate this process.

In process optimization, statistical experimental design is efficiency, economy and scientific objectivity in the use of resources of a production system, providing techniques, strategies and methods leading to better processes operating conditions [24].

The Taguchi Method (TM) [25] is a special variant of the classic experimental design, whose purpose is to determine the optimal value of levels of the process factors, so that the quality features established from being minimally affected by the variability transmit those factors difficult to control called noise factors, achieving high quality processes with lower costs of development and operation in this context, using Orthogonal arrangements (OA) in parametric design, ensure reproducibility and scaling system designed [25-26].

Here are parametric design stages, [27] the approach and the composition of the system for the design of predictive model of pests and diseases in crops.

Step 1: Determine the quality characteristics with the possibility of achieving up to 98% accuracy or methods used in predictive models [7]. Determining the response variable will be reflected by the percentage of prediction. The importance of this lies in the extraction and analysis of a large amount of data to provide accurate and interpretable models [28]. On the other hand there are the different factors to develop various models of prediction Agricultural sector applications.

Step 2: Identify noise factors and define operating conditions. The development of the experiment will be explained in two sessions, the corresponding to open field and protected cultivation. In both the human factor is definitely a noise factor.

Step 3: Identify the control factors and operating conditions. Factors: soil, water, crop, pests, diseases and field (protected crop or open field), representing control factors (Tables 1 & 2).

Step 4: Designing the experimental matrix and the procedure for data analysis. For the analysis of the effect of the parameters and set the model prediction is used an AO L18 Orthogonal Arrangement (March 1 ** 2 ** 7) x L27 (3 ** 13). Experiments in both open field and protected cultivation will be performed. In protected cultivation noise factors will be only the human factor, while open field weather conditions will be included. For both main factors are ranging from 7-18, with variants of field type.

Step 5: To develop the experiment, analyze data generated and determine the optimal level of the control factors. This will be conducted using predictive tools established by other studies (Table 1), taking into account the characteristics of selected crop and techniques and / or optimal methods for this.

Table 1. Study of parameters for the prediction models

AUTHOR	METHOD	ANALYSIS FOR…
[29]	Bayes, Likelihood	Disease Prediction
[30]	ELISA (Enzyme-Linked ImmunoSorbent Assay)	Density and distribution of pathogens
[31]	DSS (Decision Support System)	Calculation pre-sowing
[32]	IPM (Integrated Pest Management)	Potential pest entomology
[33]	Statistical	Effects of pests and natural enemies of crop
[34]	CGE (Computable General Equilibrium)	Economic effects
[35]	HVAC (Heating Ventilation and Air-Conditioning)	Relationship Between external and internal factors of protected cultivation for prediction of different parameters
[36]	Mathematical methods, model ALOMYSYS	Comparison between empirical and mechanical models for the prediction of pests and diseases
[37]	ANN (Artificial Networks), MPL (Multilayer Perceptrons)	Relationship between meteorological factors and migration of Betula pollen
[38]	ELISA (Enzyme-linked immunosorbent assay)	Factors affecting the growth of cucumber
[39]	Precision Agriculture, Data Mining	Prediction of pests and diseases
[40]	ANN (Artificial Networks), MPL (Multilayer Perceptrons)	Crop yield
[41]	Mathematical methods, ELI KDD	Economic injury level phenological stages of the crop
[42]	ANN (Artificial Networks), mathematical models: Midilli, Newton, Page, Henderson and Pabis, Logarithmic, Tow term, Approximation of diffusion	Variations between the values of soil moisture and drying

3. Results and conclusions

The crops are surrounded by a complex and variable existence of parameters (factors and variables). There are models, applications and techniques used to predict the behavior of crops in particular regions. These combine to analyze data obtained from crops and to subsequently be treated by algorithms for precision and accuracy in predictions of any event around the crop, mainly used to support decision-making.

By combining the methods, mechanical and empirical, mathematical - statistical, predictive-evolutionary algorithms, artificial neural networks etc., provides an optimal response to each prediction process for decision decision-making. Establishing the factors that have the greatest impact on cultivation development, in addition to the methods-techniques algorithms, used for a given situation, facilitating the design of predictive models for crops.

Table 2. List of study parameters of the pest prediction models.

PARAMETERS	UNIT OF MEASURE	OPERATING CONDITIONS		
		LOW	MEDIUM	HIGH
Solid Type		Clayey	Sandy	Hard
Water Quality	pH	V1: basic	V2: Medium	V3: Acid
Crop Type	Product Type	Vegetable (tomato)	Spices	Fruits
Pest Type	Family			
Diseases Type	Family			
Field for growing Type	Field	Open field		Greenhouse
Soil Temperature	℃	T1: Low Range	T2: Medium Range	T3: High Range
Ambient Temperature	℃	T1: Low Range	T2: Medium Range	T3: High Range
Soil Moisture	% Relative humidity	T1: Low Range	T2: Medium Range	T3: High Range
Ambient Moisture	% Relative humidity	T1: Low Range	T2: Medium Range	T3: High Range

4. References

[1] F. Shaxson and R. Baber. (2005). Optimización de la humedad de suelo para la produccion vegetal, el significado de la prosperidad del suelo. Organización de las Naciones Unidas para la Agricultura y la Alimentación. ftp://ftp.fao.org/agl/agll/docs/sb79s.pdf.

[2] J. E. Yuen, & Hughes, G. (2002a). Bayesian analysis of plant disease prediction. *American Journal of Agricultural Economics*, **68**: 407–412.

[3] T. Challinor, A.Wheeler, and T. O. J. Slingo. (2007). Development of a Combined Crop and Climate Forecasting System for Seasonal to Decadal Predictions, 31–40.

[4] M. H. Zhou, H. B. Li, and W. Wu, (2010). Development of a web-based geographic information system for crop pests and diseases management. 2010 2nd IEEE International Conference on Information Management and Engineering, 70–73. doi:10.1109/ICIME.2010.5477696.

[5] W.W. Guo, and H. Xue. (2011). An incorporative statistic and neural approach for crop yield modelling and forecasting. *Neural Computing and Applications*, **21**(1), 109–117. doi:10.1007/s00521-011-0636-0.

[6] A. Motevali, A. Abbaszadeh, G. H. Najafi, S. Minaei, and B. Ghobadian, (2012). Drying of jujube (Zizyphus jujube mill) fruit : comparison of prediction from mathematical models and artificial neural networks. *Autralian Joournal of Crop Science*, **6**(2):210–218.

[7] P. Eredics and T. P. Dobrowiecki. (2010b). Hybrid knowledge modeling for an intelligent greenhouse. IEEE 8th International Symposium on Intelligent Systems and Informatics, 459–463. doi:10.1109/SISY.2010.5647334.

[8] M. Puc. (2012). Artificial neural network model of the relationship between Betula pollen and meteorological factors in Szczecin (Poland).pdf. Int J Biometeorol, 395–401.

[9] K. Al-Mahallawi, J. Mania, A. Hani, and I. Shahrour. (2011). Using of neural networks for the prediction of nitrate groundwater contamination in rural and agricultural areas. Environmental Earth Sciences, 65(3), 917–928. doi:10.1007/s12665-011-1134-5.

[10] T. Li and J. Yang. (2010). The early warning and prediction method of flea beetle based on maximum likelihood algorithm ensembles. 2010 Sixth International Conference on Natural Computation, (Icnc), 1901–1905. doi:10.1109/ICNC.2010.5584642.

[11] Pedrycz Witold, (2005). Knowledege-Based Clustering: From Data to Information Granules, Wiley-Interscience, U.S.A.

[12] W. G. Dilantha, R. Ramarathnam and S. Nakkeeran. (2009). Integrated Pest Management: Innovation-Development Process. (R. Peshin & A. K. Dhawan, Eds.), 131–162, Dordrecht: Springer Netherlands. doi:10.1007/978-1-4020-8992-3.

[13] J. W. Mulumba, R. Nankya, J. Adokorach, C. Kiwuka, C. Fadda, P. De Santis, and D. I. Jarvis. (2012). A risk-minimizing argument for traditional crop varietal diversity use to reduce pest and disease damage in agricultural ecosystems of Uganda. Agriculture, Ecosystems & Environment, 157, 70–86. doi:10.1016/j.agee.2012.02.012

[14] N. J. Bone, L. J. Thomson, P. M. Ridland, P. Cole, and A. a. Hoffmann. (2009). Cover crops in Victorian apple orchards: Effects on production, natural enemies and pests across a season. *Crop Protection*, **28(**8): 675–683. doi:10.1016/j.cropro.2009.03.021.

[15] M. Głazek. (2009). Occurrence of Eyespot on Winter Wheat in the Central-Southern Region of Poland. *Journal of Plant Protection Research*, **49**(4): 426–433. doi:10.2478/v10045-009-0068-3.

[16] J. Abonyi and B. Feil. (2007). Cluster Analysis for Data Mining and System Identification, Birkhäuser Verlang AG, Germany.

[17] B. Sing-Tze, (2002). Pattern Recognition and Image Preprocessing, Second Edition, Marcel Dekker, Incorporation, USA.

[18] F. Höppner, F. Klawonn, R. Kruse and T. Runkler. (2000). Fuzzy Cluster Analysis Methods for Classification, Data Analysis and Image Recognition, John Wilet & Sons LTD, Toronto.

[19] W. Pedrycz. (2005). Knowledege-Based Clustering: From Data to Information Granules, Wiley-Interscience, U.S.A.

[20] A. Schenker, H. Bunke, M. Last and A. Kandel. (2006). "Polynomial Time Complexity Graph Distance Computation for Web Content Mining", en Data Complexity in Pattern Recognitions. Basu Mitra and Ho Tin Lam, Editores. Springer-Verlang, USA.

[21] S. Theodoridis and K. Koutroumbas. (2003). Pattern Recognition, Second Edition, Elsevier Academic Press, USA.

[22] J. Myatt Glenn. (2007). Making Senses of Data: A practical Guide to Exploratory Data Analysis and Data Mining, Wiley-Interscience, U.S.A.

[23] C. Vercellis. (2009). Business Intelligence: Data Mining and Optimization for Decision Making, John Wiley and Sons Ltd., United Kingdom.

[24] Montgomery. Editorial Limusa S.A. De C.V., Jun 30, (2005) - Technology & Engineering - 686 pages.

[25] Tagucchi, (1987). System of Experimental Desing UNIPUB Kraus International Publications, New York.

[26] Park, S. H. (1996). Robust Design and Analysis for Quality Engineering. Springer. ISBN 0412556200, 9780412556203.

[27] Reh, S. (2000). Quality Based Design with Probabilistic methods. ANSYS Solutions, 2(2) (2000), pp 20-22.

[28] Cano, R., Herrera, F. & Lozano, M. (2005). Extracción de modelos predictivos e interpretables en conjuntos de datos tamaño grande mediante la selección de conjuntos de entrenamiento. Granada España. Taller Nacional de minería de datos y Aprendizaje. 145-152.

[29] Yuen, J. E., & Hughes, G. (2002a). Bayesian analysis of plant disease prediction. *American Journal of Agricultural Economics*, **68**: 407–412.

[30] Kyuchukova, M. A., Gabler, J., Grosch, R., Crops, O., & Diagnostics, P. (2006b). Evaluation of a method for quantification of Pythium aphanidermatum in cucumber roots at different temperatures and inoculum densities. *Jornal of Plant Diseases and Protection*, **113**(3): 113–119.

[31] Guo, W.W., & Xue, H. (2011). An incorporative statistic and neural approach for crop yield modelling and forecasting. *Neural Computing and Applications*, 2011, **21**(1): 109–117. doi:10.1007/s00521-011-0636-0.

[32] Dilantha, W. G., Ramarathnam, R., & Nakkeeran, S. (2009). Integrated Pest Management: Innovation-Development Process. (R. Peshin & A. K. Dhawan, Eds.) (pp. 131–162). Dordrecht: Springer Netherlands. doi:10.1007/978-1-4020-8992-3.

[33] Bone, N. J., Thomson, L. J., Ridland, P. M., Cole, P., & Hoffmann, A. A. (2009). Cover crops in Victorian apple orchards: Effects on production, natural enemies and pests across a season. Crop Protection, 28(8), 675–683. doi:10.1016/j.cropro.2009.03.021.

[34] Xie Jie (2010). Greenhouse Efects on theWorld Agriculture. In Mechanic Automation and Control Engineering (MACE), 2010 International Conference on, 5–8.

[35] P. Eredics, D., Fantinou, A., & Lykouressis, D. (2011). Enhancing pest control in annual crops by conservation of predatory Heteroptera. Biological Control, 59(1), 13–21. doi:10.1016/j.biocontrol.2011.03.014.

[36] Colbach, N. (2012). Modelling cropping system effects on crop pest dynamics How to compromise between process analysis and decision aid.

[37] Malgorzata, P. (2012). Artificial neural network model of the relationship between Betula pollen and meteorological factors in Szczecin (Poland).pdf. Int J Biometeorol. 395–401.

[38] Panova, G. G., Heißner, A., Grosch, R., & Kläring, H. P. (2012). Pythium aphanidermatum May Reduce Cucumber Growth without Affecting Leaf Photosynthesis. *Journal of Phytopathology*, **160**(1): 37–40. doi:10.1111/j.1439-0434.2011.01849.

[39] Tripathy, a. K., Adinarayana, J., Sudharsan, D., Merchant, S. N., Desai, U. B., Vijayalakshmi, K., Raji Reddy, D., (2011). Data mining and wireless sensor network for agriculture pest/disease predictions. 2011 World Congress on Information and Communication Technologies, 1229–1234. doi:10.1109/WICT.2011.6141424.

[40] Selvaraj, K., Chander, S., & Sujithra, M. (2012). Determination of multiple-species economic injury levels for rice insect pests. Crop Protection, 32, 150–160. doi:10.1016/j.cropro.2011.10.019.

[41] Motevali, A., Abbaszadeh, A., Najafi, G. H., Minaei, S., & Ghobadian, B. (2012). Drying of jujube (Zizyphus jujube mill) fruit : comparison of prediction from mathematical models and artificial neural networks. *Autralian Joournal of Crop Science*, **6**(2):210–218.

[42] Mulumba, J. W., Nankya, R., Adokorach, J., Kiwuka, C., Fadda, C., De Santis, P., & Jarvis, D. I. (2012). A risk-minimizing argument for traditional crop varietal diversity use to reduce pest and disease damage in agricultural ecosystems of Uganda. Agriculture, Ecosystems & Environment, 157, 70–86. doi:10.1016/j.agee.2012.02.012.