IADIS International Journal on Computer Science and Information Systems
Vol. 15, No. 2, pp. 148-160
ISSN: 1646-3692

A HIDDEN MARKOV CHAIN APPROACH TO CROP YIELD FORECASTING

Jean Samarone de Almeida Ferreira[1], Ana Paula Lüdtke Ferreira[1] and Naylor Bastiani Perez[2]
[1]*Universidade Federal do Pampa, Brazil*
[2]*Embrapa Pecuária Sul, Brazil*

ABSTRACT

Prediction of harvest yield is an important and challenging problem. Attempts to solve this problem rely usually rely on regression techniques highly dependent on local factors. This paper presents a hidden Markov model approach for forecasting weight production. The model can deal with any culture or provided data. Results show that the model can capture both spatial and temporal harvest variability. Model analysis can help determine causes of variability, differently from regression or more straightforward Markov chain approaches. The resulting structure can benefit from statistical techniques for model tuning and model fitting.

KEYWORDS

Precision Agriculture, Stochastic Processes, Probabilistic Inference

1. INTRODUCTION

Crop productivity depends on a diversity of factors ranging from the production environment to plants' genetic characteristics (Franchini et al., 2016). The production environment refers to soil properties, water availability, terrain elevation, slope, and orientation, among other traits. It is possible to modify some of the production environment by human management activities. Land management can be costly, so it is mandatory to understand the impact changes will have on the final profit. To make matters more complicated, carrying out the same way planting, management, and harvest are done will not change the fact that the final production results will present both spatial and temporal variability. Discovering the main factors associated with production variability assures two significant developments: the definition of management practices aimed at maximizing the results of the harvest, and the correct prediction of crop results, given the existing conditions in the actual production environment.

Looking for the leading causes of production variability is an active topic of research within the Agricultural Sciences. The reviewed literature shows that experiment findings depend strongly on the available data, usually specific to each culture and location. So, it is often not possible to extrapolate the results to other places or different types of cultures. The findings referring to the primary causes of productivity variation cover almost all agronomic and meteorological variables: air temperature (Franchini et al., 2016), water deficit (Guedes-Filho, 2009), topography (Kravchenko and Bullock, 2000), cation exchange capacity (Usowicz and Lipiec, 2017), and so forth. The difficulty in analyzing the variables that intervene in the production results seems to be intrinsic to the problem: Vieira and Gonzalez (2003) suggest that the factors that cause the variability in the production results themselves also vary with time.

Precision in forecasting crop productivity has essential economic, environmental, and social impacts. Even a partial solution to that problem is desirable if an optimal one is not available. To be useful, a crop yield forecasting model must deal with both spatial and temporal variability and also provide information to the producer to take management actions even during the planting season. The methods used to define the causes of variability in production results are, for the most part, based on correlation and regression techniques as seen in Cai et al. (2013) and Miller, Singer, Nilsen (1998) or geostatistical methods as discussed in Al-omran et al. (2013), Mattioni, Schuch and Villela (2011) and Acosta et al. (2019). However, correlation cannot capture causation or any other non-linear relationships (Murphy, 2012). Geostatistics requires interpolation of data points that may not accurately represent actual soil characteristics and cannot capture crop development's temporal aspects.

A stochastic process is a probabilistic chain of events or situations where future events in the chain depend on the previous ones (Taha, 2007). Markov chains are a special kind of stochastic process suitable to represent time-evolving processes where the next system state is determined uniquely from the past state by a fixed set of computed conditional probabilities. Unlike most statistical and other forecasting models, Markov chains are characterized by less rigorous assumptions and provide more information about causality than logistic regression, neural networks, and other non-explanatory inference models. Named after the Russian mathematician Andrei Andreyevich Markov (1856-1922), Markov chains have various applications, ranging from – but not limited to – finance, marketing, biology, text processing, and pattern recognition.

Markov chains are an often encountered modelling technique for ecological and environmental systems (Paegelow and Olmedo, 2008). Diniz (1984) highlighted the potential for applying Markovian principles and information technology resources to agricultural transition and transformation studies. Matis et al. (1985) and Matis, Birkett, Boudreaux (1989) present different ways of using Markov chains to predict crop yields and establish the methodology's generality and practicality that involves this type of model. Thirunavukkarasu (2015) assessed the trend (increase or decrease) of barley production in India, using a Markov model, which only depends on the past data set instead of a more rigorous set of assumptions. Growth rate analyses are widely used to study different agricultural products' behaviour and are usually estimated using parametric models and time series.

Although its general use in a wide range of applications, Markov chains have a simple definition based on states without any internal structure. Modelling variable relations – binary or otherwise – within the framework of Markov chains can be very challenging. Markov chains are simply a set of states with transitions from one state to another. Each state transition in a Markov chain is labelled with the conditional probability of the system make it at some point.

However, farmers are more interested in how the environment affects production, to which a structure consisting solely of states cannot adequately respond. The question "what is the most likely harvest yield for this plantation, based on current data?" determines an inference problem from a set of conditions to a final result. The search for causal effects between agricultural variables and crop yield demands a richer framework. It is not enough to know the system change probability from one state representing yield change to another. It is also necessary to establish the likelihood of how system variables modifications affect that change.

A *hidden Markov model* provides such a framework. Hidden Markov models separate *observable* variables (attributes of the production environment) from *hidden* variables (production results), which is the inference goal. One can address three problems within that framework: i) the likelihood or probability evaluation problem, ii) the decoding or optimal state sequence problem, and iii) the learning or parameter estimation problem (Rabiner and Juang, 1986). We want to determine the most likely harvest yield, given actual production data. So, the second problem is of interest. The Viterbi algorithm (Viterbi, 1967) presents an efficient optimal recursive solution to this problem (Forney, 1973).

This work focuses on developing a flexible forecasting model for predicting harvest results, capable of incorporating new information as they become available, considering it is a characteristic of agricultural production systems. To the extent of our knowledge, supported by relating literature review, it is a novel approach.

The rest of the text is structured as follows: Sec. 2 presents the technological framework used to build the model and their algorithms; Sec. 3 presents and discusses the most significant findings; finally, Sec. 4 concludes and discusses some further developments.

2. MATERIAL AND METHODS

Conceptually, this work can be categorized as applied research, consisting of the following steps: data collection and preprocessing; model design, implementation, and subsequent simulation and analysis. The dataset comes from two sources: the meteorological data originates from the Brazilian National Institute of Meteorology database; the soil and production data come directly from the experimental research land owned by EMBRAPA South Livestock.

Gathered data was imported and subsequently preprocessed with the Geographic Information System QGIS (QGIS, 2019). Data residing in Microsoft Excel™ datasheets were imported to LibreOffice Calc, reorganized, exported to ASCII files, and later used as input for QGIS (QGIS, 2019) and R Studio (R Core Team, 2019). Productivity data were in a Shapefile (.shp) format, generated by Stara TOPPER 4500 (STARA, 2011) precision agriculture controller, which saves the actual production weight collected during harvest. The shapefile contains polygons associated with a local database using DBF files, imported to layered QGIS files. A GPS built into the controller generates altitude data.

Meteorological data were retrieved from the INMET database (INMET, 2018a), referring to Bagé/RS, Brazil (OMM 83980, latitude -31:305661º, longitude -54:119352º, altitude 245.66m). Notice that there was no meteorological station inside the property. The actual station is about 12km straight away; therefore, that information is just an approximation. The chosen variables are the accumulated precipitation in 24 hours, solar irradiation per time unit, expected water deficit and surplus for soy cultures, measured in millimetres. The latter was estimated using the decision support system for agriculture SISDAGRO (INMET, 2018b).

Soil penetration resistance data were collected in the field by EMBRAPA technicians using the soil compactness electronic meter penetroLOG, from Falker (FALKER, 2018). The full dataset corresponds to measurements in 50 sites, with 4 to 8 samples taken in each locale. Each measure registers soil penetration resistance between 1 and 40 centimetres, totalizing 40 measurements per sample.

Soil chemical fertility data also correspond to 50 georeferenced samples, with 22 variables quantified. We only used 5 of them to build the model, as shown in Table 1.

Table 1. Soil measurements variables

Name	Measurement unit	Description
Clay	%	amount of clay in the soil
P	mg/L	amount phosphor in the soil
K	mg/L	amount potassium in the soil
Organic matter	%	organic material percentage
CTC	cmolc/dm3	cation exchange capacity with pH 7

Meteorological information was the same for the whole area. Fifty (50) points across the terrain provided the rest of the data. Each of those points has productivity, soil, climate, and altimetry data across different harvests. Figure 1 shows the distribution of data points across the area.

Data discretization of numerical values followed the guidelines presented in the literature. If there was no available definition of appropriate class values, a frequency distribution was computed and partitioned into classes. The class frequency is the number of elements belonging to each class (Assis; Arruda; Pereira, 1996). Spiegel and Stephens (2008) recommended defining the total amplitude by calculating the difference between the highest and lowest data set value. Afterwards, the amplitude value was divided by three, according to a criterion of the same class size and number of pre-fixed classes. The final step was to determine class frequencies and class intervals to feed into the model. Table 2 presents all the model variables and their respective classes and value limits for each class.

Model development and programming used the R programming language (R Core Team, 2019) within the R Studio. The model implementation used the following R packages: TraMineR, gmodels, sf, tmap, readODS, gstat, sp, raster, and rgdal. The functions in the R package markovchain (Spedicato, 2019) implement conditional probabilities calculations, transition matrices definitions, matrix multiplication, and other elementary functions. Package seqHMM (Helske e Helske 2019a, 2019b) provides the methods for data organization, model building, analysis, and testing; the same package provides an implementation of the Viterbi algorithm. The model analysis procedure used the Bayesian Information Criterion, or simply BIC, which measures overfitting (Robles et al., 2012).

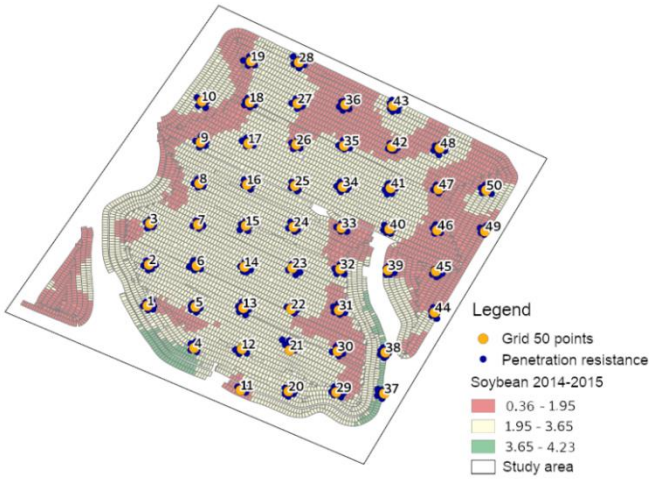

Figure 1. Data points distribution across the field of study

The algorithms developed can present model results as lists, sequential graphs, percentage graphs, spatialized maps in the form of points, and interpolated maps with the Inverse Distance Weighting or IDW method (Bivand, Pebesma e Gómez-Rubio, 2008). The only culture used in the model was soy, which can be replaced by any other culture without changing the model structure. A straightforward modification in the model permits to tackle several cultures at once.

Table 2. Variable classes and value limits

Name	Low	Medium	High
Production weight (ton)	< 1.9	1.9-3.5	> 3.5
Precipitation (mm)	<= 671.4	-	> 671.4
Insolation (h)	<= 1,226.9	-	> 1,226.9
RP (Kpa)	<= 2,000	-	> 2,000
Water deficit (mm)	<= 164.1	-	> 165.1
Water surplus (mm)	<= 488.5	-	> 488.5
Altitude (m)	<= 243.5	-	> 243.5
Organic matter (%)	<= 2.5	-	> 2.5
Phosphor	<= 12	-	> 12
Potassium (CTC <= 15)	< 40	40-60	> 60
Potassium (CTC >15)	<=90	-	> 90

3. THE HIDDEN MARKOV MODEL

The production yield forecasting system was model as a finite hidden Markov chain. The hidden states are the production yield; the visible states correspond to each observable variable: precipitation (Prec), insolation (Insol), water deficit (Defi), water surplus (Exce), soil compaction (RP), altitude (Alti), organic material (MO), phosphorous (P), and potassium (K). Production yield has three classes: low, medium, and high. Figure 2 depicts the model organization.

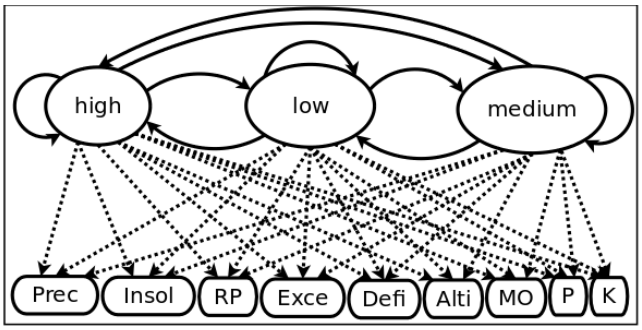

Figure 2. Production yield forecasting model organization

Formally, a hidden Markov chain is a tuple M = (S, V, A, B, π), where S is a finite set of (hidden) states; V is a set of observable variables; A is the transition matrix where each element a(i,j) in A is the conditional probability p(Sj|Si), or the probability of system M goes to state Sj given that its current state is Si; B is the observation probability matrix, where each element b(j,k) in B is the conditional probability p(Vk|Sj) or the probability of observing values Vk given that the system is in state Sj; and π is a probability mass function, where π(s) indicates the likelihood of system M starts in state s. The system M = (S, V, A, B, π) is sometimes written as M = (A, B, π) since it is possible to infer the first by the latter (Fink, 2008; Jurafsky e Martin, 2018; Rabiner e Juang, 1986). The decodification problem can be formally stated as follows: given a model M and a sequence of observations O = o1, o2,…, on, whats is the most probable sequence of hidden states occurring? Each observation consists of one value for each visible variable in the model. The Viterbi algorithm gives an efficient recursive optimal solution for this problem (Viterbi, 1967; Forney, 1973).

The data gathered from four years of production fed the model. The marginal and the initial probabilities were directly calculated from the production yield in each year. Figures 3a, 3b, 3c, and 3d show the productivity yield in each harvest, categorized by produced weight (low, medium, high). Notice that productivity is not uniform along the area. Some regions were consistently more fruitful than others in all periods, while others presented high variation rates along the time.

Markov chains have the property that the final results are independent of the initial conditions, given enough system transitions over time. It means the system stabilizes around general condition probabilities. The actual production and measure data, taken from the 50 points shown in Figure 1, served as input to the model. A matrix with 50 lines and 4 columns codes the production data. Each element in that matrix gives a production yield from a specific year. Figure 4 shows the system transition matrix and the system emission (or observation) matrix and its elements. The transition matrix elements refer to the probability of a specific (high, medium, low) yield given last year's (high, medium, low) production result. The emission matrix E informs the yield's probabilities (high, medium, low) given the set of observed variable classes. There is an emission matrix for each of the visible variables: precipitation, insolation, water deficit, water surplus, soil compaction, altitude, organic material, phosphorous, and potassium. As shown in Figure 4, emission matrices have three lines, one for each possible production yield. The number of columns in an emission matrix depends on how many value classes the variable has.

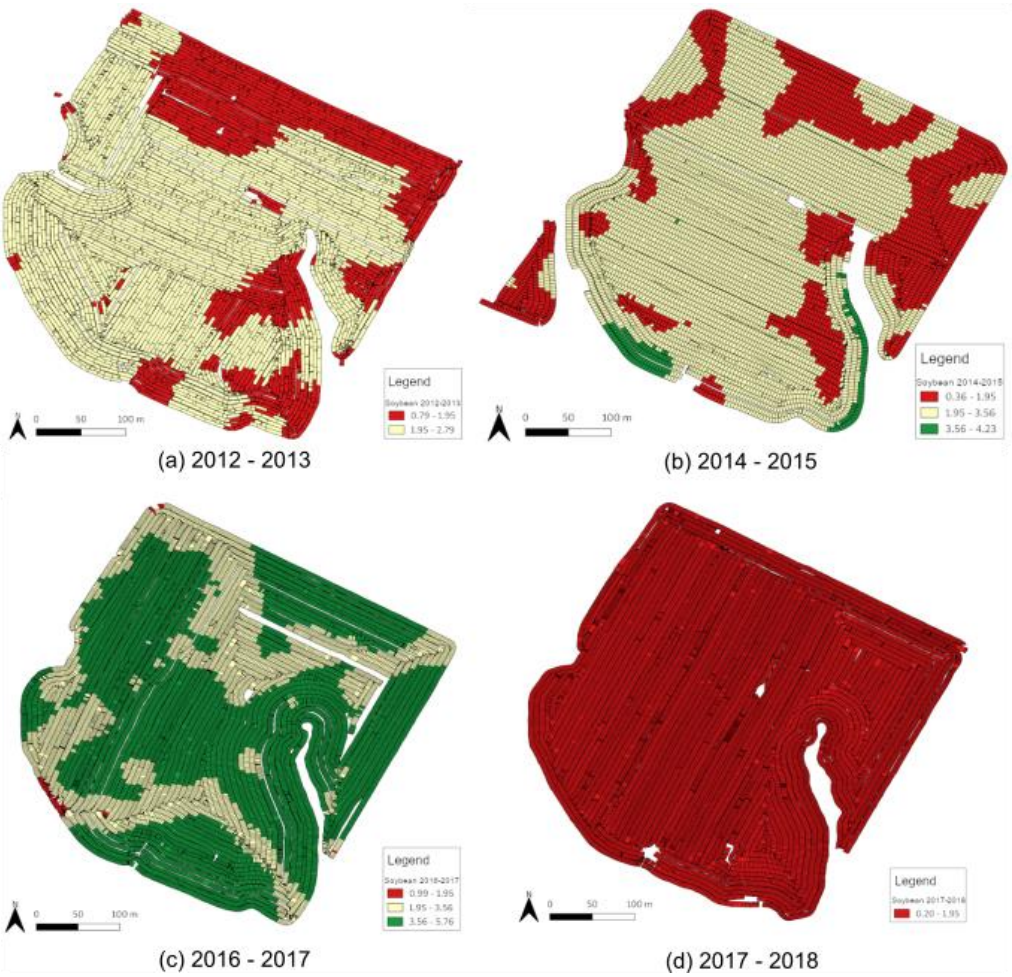

Figure 3. Production yield in four different harvests

$$M = \begin{pmatrix} p(h|h) & p(h|m) & p(h|l) \\ p(m|h) & p(m|m) & p(m|l) \\ p(l|h) & p(l|m) & p(l|l) \end{pmatrix} \quad E = \begin{pmatrix} p(a|o1) & \dots & p(a|on) \\ p(m|o1) & \dots & p(m|on) \\ p(l|o1) & \dots & p(l|on) \end{pmatrix}$$

Figure 4. Probability transition matrix and emission probability matrix

The straightforward extension of the model structure allows for addressing any number of variables. Since the model is discrete, it is necessary to group continuous values into arbitrary classes, depending only on the concrete problem. There is no need to establish any assumptions about how variables influence the final yield. The computed conditional probabilities allow the inference of how observations affect results. It is also possible to build a sequence of models to

predict the outcome given a timeline of events. Time intervals can be years, months, days, growth phase, or any other chosen interval. The only existent restriction is data availability for a frequentist approach to probability. A Bayesian approach to initialize the model is also possible. Model tuning is possible as new data arrives. Sec. 3 presents the results obtained by the model described herein, and Sec. 4 discusses further improvements to our model.

4. RESULTS AND DISCUSSION

Data were available for four harvests, with two of them having a medium yield, one having a low output, and one having a high result. The supplied data allowed the calculation of marginal and conditional probabilities. However, the patent limitations on the sample size prevent the achievement of general conclusions about causal relationships. Figure 5 depicts the generated hidden Markov model, in the form of a graph, with the nine variables described in Table 2. Graph nodes represent the three hidden states (low, medium, and high yield). Each node is partitioned accordingly to the probabilities of observing different classes of each visible variable. Graph edges show the transition probabilities. Initial probabilities appear under the nodes. Observation states with probability lower than 1% appear as a single node sector (coloured white). The graph also omits the states with an estimated zero probability. Transitions probabilities values are round off for legibility.

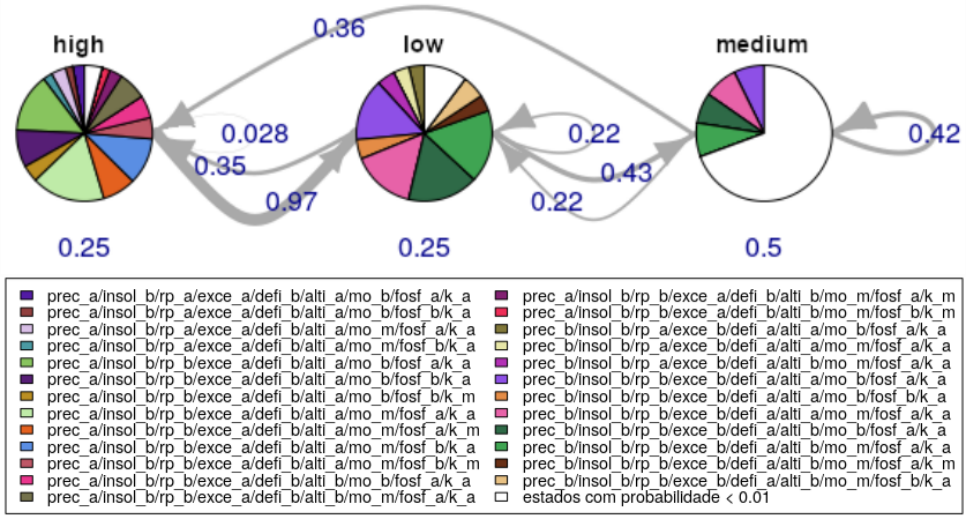

Figure 5. The complete hidden Markov model

The model testing used both factual and simulated data. The four harvests data set served to build the model and to analyze its results. Investigation of model strength and adequacy used simulated data. QGIS preprocessed all georeferenced data before their input into the model, coded in the R programming language. Figure 6 presents the most probable harvest yield evolution along the years, given observation data. The graphics in Figure 6 (right) reads as follows: results point to the first harvest having high yield with 48% chance, no chance for low output, and 52% chance of medium yield; productivity yield for the second harvest has a

probability of 12% for high yield, 50% of low yield, and 38% for medium yield; the third harvest have 80% chance of high yield, 12% of low yield, and 8% medium yield; the fourth and last harvest has a chance of high, low and medium yield of, respectively, 6%, 80%, and 14%. The image on the left side of Figure 6 shows the expected outcome per data point. The pictures displayed in Figure 6 allows the assessment of both spatial and temporal yield variability within the field.

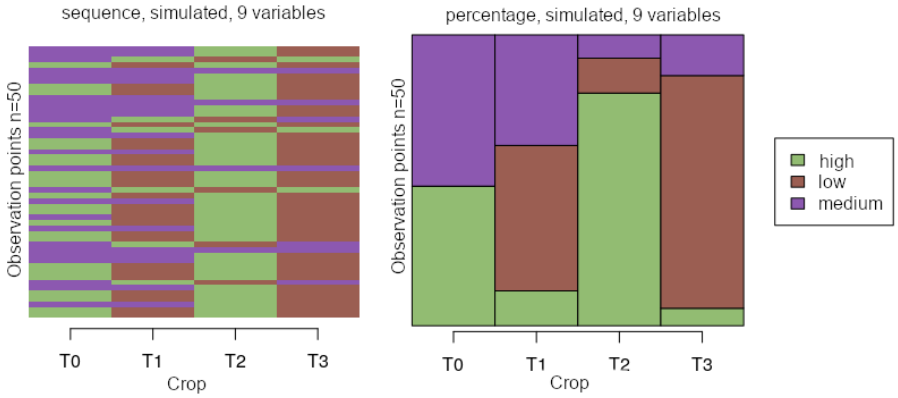

Figure 6. Most probable harvest yield

Figure 7 presents the expected productivity by region in each of the four harvests (indicated as T0, T1, T2, and T3). Interpolated data from the 50 information sources points appear where no data exist. The map identifies the original regions that serve as data sources for yield weight, soil samples, and soil penetration resistance. QGIS, or any other geographical information system, can generate a map like this from the model. Farmers or other non-specialists more easily read charts and other graphic presentations.

Emission matrices, built from observed variables concerning production yield, can estimate a particular variable's relative impact on the outcome. Figure 8 shows the computed emission matrices for precipitation and potassium in the soil. Both matrices present data ordained from high to low, where the lines refer to productivity, and the columns refer to the variable classes, also sort from higher to lower. Although the amount of data is not enough to draw definitive conclusions about causality relations, it still provides information. For instance, precipitation values correlate almost directly with the final weight result; potassium, on the other hand, does not appear to have any influence on the harvest outcome.

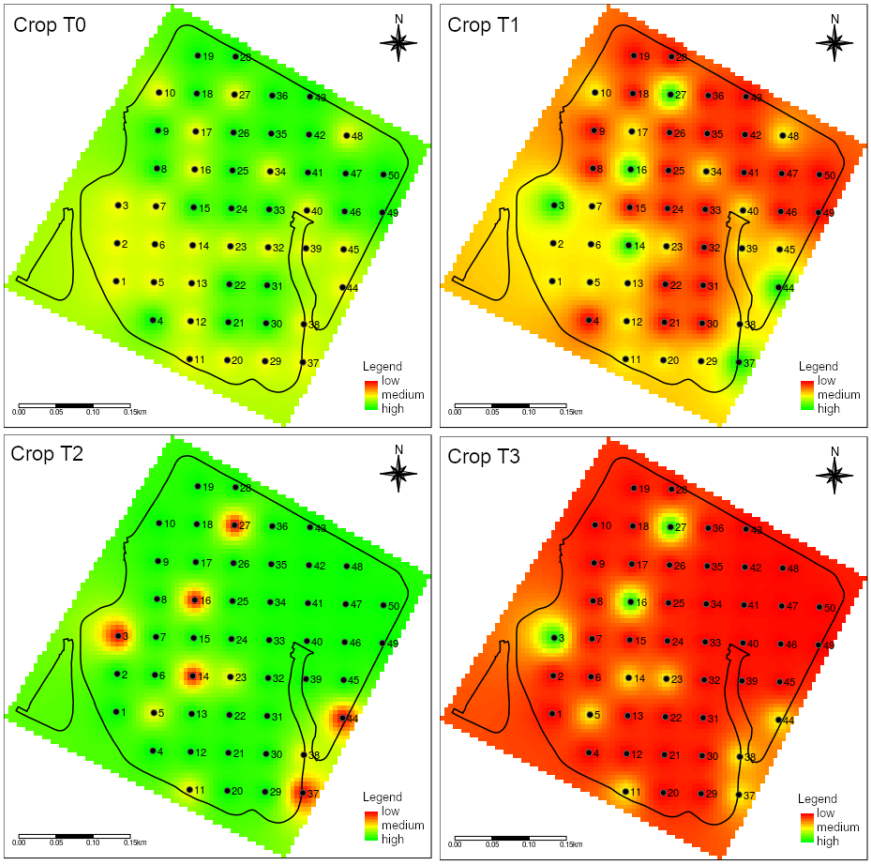

Figure 7. Expected productivity by region

$$Prec = \begin{pmatrix} 0.9722 & 0.0278 \\ 0.5256 & 0.4744 \\ 0.1733 & 0.8267 \end{pmatrix} \quad K = \begin{pmatrix} 0.6111 & 0.2778 & 0.1111 \\ 0.6410 & 0.3077 & 0.0513 \\ 0.7733 & 0.1733 & 0.0533 \end{pmatrix}$$

Figure 8. Emission matrices values

The model testing proceeded to investigate what was the best set of variables to fit the model. According to the Bayesian Information Criterion (BIC), the best model to accurately represent the data with minimum overfitting has six variables (1207.326) followed by a five variable model (BIC 1398.319); the complete nine variables model has the worst result (BIC 2090.064). The emission matrices analysis had already pointed out that result since the model relies on variables with no influence on the actual product yield.

5. CONCLUSION

Harvest yield prediction is a relevant and challenging problem for farmers in particular and for the society in general. Currently, most approaches to solve this problem rely on regression techniques that cannot establish causal relationships. Furthermore, factors influencing production variability can themselves change with time.

Markov models are a special kind of stochastic process with a short-term memory of events. Hidden Markov chains are a special kind of Markov models that separate visible variables (the information on the production system) from the hidden variables (the harvest outcome). There are no assumptions to build a model other than some statistical relationship between the observed elements and the result. Additionally, Markov chains capture the temporal process evolution in a very straightforward manner.

This work presented a hidden Markov model to predict crop yield results. The model uses nine soil and meteorological variables as the visible ones: accumulated precipitation, accumulated insolation, accumulated water deficit, accumulated water surplus, soil compaction, altitude, organic material percentage, amount of phosphorous, and amount of potassium. The hidden variable is the harvest weight production.

The model received data from four harvests. The sample size was not sufficient to draw general conclusions about relations between variables. However, there was enough information to establish the pertinence of the technique. Emission matrices permit investigating causal relations while the transition system model can infer future states based on past ones. The hidden transition system allows the inference of probable future yields given present observable conditions. The same model technique applies to any culture or available production system data. If a new measurement becomes available, it suffices to calculate the new emission matrix to extend the model. We can also attach multiple models to a region, to investigate variability in a more precise way.

We used the Bayesian Information Criterion (BIC) to evaluate models built with simulated data. The simulated data was necessary because the actual data was too scarce to reveal any of the method's strengths when applied to the model. The BIC analysis has shown the pertinence of statistical techniques to refine and tune hidden Markov chain models.

REFERENCES

Acosta, J. J. B. et al., 2019. Variabilidade espacial da produtividade, perdas na colheita e lucratividade da cultura de soja. *Revista Agrogeoambiental*, IFSULDEMINAS, Vol. 10, pp. 27-46, ISSN 2316-1817.

Al-Omran, A. et al., 2013. Spatial variability for some properties of the wastewater irrigated soils. *Journal of the Saudi Society of Agricultural Sciences*, Vol. 12, n. 2, pp. 167-175, ISSN 1658-077X.

Assis, F. N.; Arruda, H. V. de; Pereira, A. R., 1996. *Aplicações de Estatística à Climatologia - Teoria e Prática*. 8th edition, Editora Universitária UFPEL, Pelotas.

Bivand, R. S.; Pebesma, E.; Gómez-Rubio, V., 2008. *Applied Spatial Data Analysis with R*. NY: Springer, New York.

Cai, R. et al., 2013. Using a climate index to measure crop yield response. *Journal of Agricultural and Applied Economics*, Vol. 45, n. 4, pp. 719-737.

Diniz, J. A. F., 1984. *Geografia da Agricultura*. Difel, São Paulo.

Ferreira, J. S. A., 2019. Predição da variabilidade espacial da produtividade agrícola com modelos ocultos de Markov. Master Thesis. Programa de Pós-graduação em Computação Aplicada, Universidade Federal do Pampa.

Fink, G. A., 2008. *Markov models for pattern recognition - from theory to applications*. 8th. ed., Springer-Verlag, Londres.

Forney, G. D., 1973. The Viterbi algorithm. *Proceedings of IEEE*, Vol. 61, n. 3, pp. 268-278.

Franchini, J. C. et al, 2016. Variabilidade espacial e temporal da produção de soja no Paraná e definição de ambientes de produção. Embrapa Soja, Londrina, Brasil. ISSN 2176-2937.

Guedes-Filho, O., 2009. Variabilidade espacial e temporal de mapas de colheita e atributos do solo em um sistema de semeadura direta. Master Thesis, Instituto Agronômico, Campinas, São Paulo.

Helske, J.; Helske, S., 2019. seqHMM: mixture hidden Markov models for social sequence data and other multivariate, multichannel categorical time series. R package version 1.0.14. [online] Available at: https://www.rdocumentation.org/packages/seqHMM/versions/1.0.14. (Accessed: 02 november 2020).

Helske, S.; Helske, J., 2019. Mixture hidden Markov models for sequence data: the seqHMM package in R. *Journal of Statistical Software*, vol. 88, n. 3, pp. 1-32.

INMET, 2018. Banco de Dados Meteorológicos do INMET. Available at: http://www.inmet.gov.br/portal/index.php?r=bdmep/bdmep. (Accessed: 25 august 2018).

INMET, 2018. SISDAGRO. Available at: http://sisdagro.inmet.gov.br/sisdagro/app/index. (Accessed: 25 march 2018).

Jurafsky, D.; Martin, J. H., 2018. Speech and Language Processing. [online] Available at: https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf (Accessed: 12 november 2020).

Kravchenko, A. N., and Bullock, D. G., 2000. Correlation of corn and soybean grain yield with topography and soil properties. *Agronomy Journal*, American Society of Agronomy, Vol. 92, n. 1, pp. 75-83, ISSN 0002-1962.

Matis, J. et al., 1985. A Markov chain approach to crop yield forecasting. *Agricultural Systems*, Vol. 18, n. 3, pp. 171-187, ISSN 0308-521X.

Matis, J. H.; Birkett, T.; Boudreaux, D., 1989. An application of the Markov chain approach to forecasting cotton yields from surveys. *Agricultural Systems*, Vol. 29, n. 4, pp. 357-370.

Mattioni, N. M.; Schuch, L. O. B.; Villela, F. A., 2011. Variabilidade espacial da produtividade e da qualidade das sementes de soja em um campo de produção. *Revista Brasileira de Sementes*, Vol. 33, pp. 608- 615, ISSN 0101-3122.

Miller, M. P.; Singer, M. J.; Nielsen, D. R., 1998. Spatial variability of wheat yield and soil properties on complex hill. *Soil Science Society of American Journal*, Vol. 52, pp. 1133-1141, ISSN 0361-5995.

Murphy, K. P., 2012. *Machine Learning – A Probabilistic Perspective*. MIT Press, Cambridge, USA.

Paegelow, M. and Olmedo, M. T. C. 2008. *Modelling Environmental Dynamics - Advances in Geomatic Solutions*. Berlin Heidelberg: Springer-Verlag. (Environmental Science).

QGIS. QGIS Geographic Information System. 2019. [Online] Available at: http://qgis.osgeo.org. (Accessed: 05 november 2020).

Rabiner, L. R. and Juang, B. H. 1986. An introduction to hidden Markov models. *IEEE ASSP Magazine* 3, pp. 4-16.

Robles, B. et al., 2012. Methods to choose the best hidden Markov model topology for improving maintenance policy. In: *International Conference of Modeling, Optimization and Simulation*. Bordeaux, France.

R Core Team., 2019. R: A Language and Environment for Statistical Computing.Vienna, Austria: R Foundation for Statistical Computing. [Online] Available at: http://www.R-project.org/. (Accessed: 15 november 2020).

Spedicato, G. A., 2019. Package markovchain. Vienna, Austria, [Online] Available at: https://cran.r-project.org/web/packages/markovchain/markovchain.pdf. (Accessed: 15 november 2020).

Spiegel, M. R.; Stephens, L. J., 2008. Theory and Problems of Statistics. 4th. ed., McGRAW-HILL, NewYork.

STARA, 2011. Manual de Instruções TOPPER 4500VT.

Taha, H. A. 2007. *Operations Research: An Introduction*. 8th ed., Pearson Prentice Hall, New Jersey.

Thirunavukkarasu, M., 2015. Stochastic modeling in agricultural production. PhD Thesis, Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu, India.

Vieira, S. R.; Gonzales, A. P., 2003. Analysis of the spatial variability of crop yield and soil properties in small agricultural plots. *Bragantia*, Vol. 62, pp. 127-138, ISSN 0006-8705.

Viterbi, A. J., 1967. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, Vol. 13, n. 2 pp.260-269.

Usowicz, B.; Lipiec, J., 2017. Spatial variability of soil properties and cereal yield in a cultivated field on sandy soil. *Soil and Tillage Research*, Vol. 174, pp. 241-250, ISSN 0167-1987.