18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

Visualisation techniques for modelling environmental influences in viticulture

Owusu-Banahene, W. [1], **Shanmuganathan, S** [2] **and P. Sallis** [2]

[1] *University of Ghana, Legon, Accra, Department of Computer Engineering. E-mail: banahene@ug.edu.gh*

[2] *Geoinformatics Research Centre, School of Computing and Mathematical Sciences, Auckland University of Technology, New Zealand.*

Abstract: The quality of wine is in its making but significantly also depends on the grape ripening conditions, such as environmental, weather and growing factors, as well as grape varietal characteristics. Some centuries-old Mediterranean grape growing tradition has helped the winemaking industry pool resources in an effort to gain further understanding of the interrelationships between genetic (or "cultivars") and environmental factors, such as climate and soil (or "terrior"), The industry seems to be successful in using historic, in some cases, years old data to improve cultivation practices or "vintages" with an ultimate aim of producing finer wine. This paper outlines past research on enhancing data depiction methodologies generally used to represent environmental influence factors in viticulture along with some recent approaches, with special emphasis on data clustering techniques applied to analysing multivariate data sets using standard pixel representations and also some dependency relationship visualisation methodologies based on self-organising map (SOM) techniques. Literature reviewed reveals the limited research in the visualisation of environmental data in viticulture. The paper then elaborates upon a system design with some approaches being investigated for the implementation of novel visualisation techniques that could facilitate knowledge extraction from data on environmental influence factors in viticulture. The two main sets of approaches being referred to as neural net (NN) and non NN. The NN approach encompasses various visualisation techniques that utilises computational neural networks, such as SOM. Those in the non NN category are novel ways that are in complimentary to those in the NN with capabilities to overcome the limitations of the former, by creating a synergetic effect. An example of a possible non NN visualisation could be MIR-max and other algorithmic techniques. The system proposed herein attempts to integrate the two broad techniques into one system to enhance the visualization of environmental influence factors in viticulture especially for use in assessing any environmental impact with novel data visualisation techniques for decision making and in determining "*what makes a good year for wine*". The system would facilitate visualisation techniques over a distributed system, such as the Internet, and provide functions with user interface useful to both experts and novices in viticulture.

Keywords: *computational neural networks, data depiction methodologies, environmental influence factors, novel visualisations, viticulture*

1. INTRODUCTION

Although the quality of wine ultimately relies upon the skills of the winemaker (Smith, 2002), the amounts of sugar, aroma, flavour forming protein components in the fruit (grapes) used for producing wine as well determine the quality, especially wine taste and flavour. There are numerous environmental factors that influence grape quality during fruit ripening, together with the growing practices that attend them. Smith (2002), as well as many other studies, describe the main factors affecting vines, their development and annual cycle of growth as well as the yield produced each year, under the following headings:

1) *Physical Environment* (e.g. landform, altitude, slope, aspect, natural and built features, soils, nutrients, water and drainage),
2) *Natural Phenomena* (e.g. climate, heat summations, seasonal variations, longer term cycles and fluctuations, weather and hazards); and
3) *Viticulture & Vineyard Management* (e.g. selection of sites, design and layout of vineyards, varieties and rootstocks, infrastructures and services such as irrigation).

This paper is intended to investigate into some major issues in visualising environmental data in viticulture and then to describe an approach to enhance a set of combinational data depiction methodologies generally used to represent environmental influence factors in viticulture, mostly data clustering techniques especially applied to analysing multivariate data sets using standard pixel representations and also some dependency relationship visualisation methodologies with self-organising map (SOM) techniques. Literature reviewed thus far for this work revealed that not much work has been done in the visualisation of environmental data in viticulture. Hence, this paper seeks to investigate into novel visualisation techniques applied in other disciplines to see how best environmental influence/ impact data relating to viticulture could be presented to the decision making management. The approach presented herein seeks to add novel features to visualisation techniques currently in use and furthermore investigate into how best these novel and useful features could be incorporated into software tools developed for non experts using prototypes. An integrated data visualisation module with artificial neural networks coupled with other non-neural net based techniques is added to the system to overcome the major issues in integrating data from different sources/ database systems. SOM, and other concepts will be reused as components. With that introduction the paper is organized as follows: Section 2 presents a review of techniques currently employed in modelling environmental factors. In section 3, we introduce some key research questions relating to areas of visualization techniques which this paper would attempt to propose possible answers to. Section 4 discusses our proposal to modelling environmental influence factors using novel visualization techniques. Expected outcomes are discussed in section 6. Section 7 presents concluding remarks.

2. REVIEW OF VISUALISATION TECHNIQUES USED FOR MODELLING ENVIRONMENTAL FACTORS

In the last two decades, research into the development of information management systems using geo-coded data and computational intelligence on-site in vineyards has intensified in order to gain a more complete understanding of regional factors and their influences on crop growth and yield. Techniques for identifying the complex interactions among these factors range from modelling the effects of the environment, such as climate, atmosphere, soil type, terrain, minerals, transition metals and irrigation of saline water on the development of rootstock, to crop growth and berry ripening rates, protein conversion in different varieties and their composition, within and among vineyards (Shanmuganathan et al.,2008). The following is a review on some recent work that could be described as ways of novel visualisation.

In (Shanmuganathan et al.,2007a), the authors described SOMs as an excellent tool for visualising multivariate data sets on low (usually 1 or 2) dimensional displays otherwise found to be difficult with standard statistical methods. Standard statistical methods are good at studying simple statistics (i.e., mean, standard deviation and so forth) of low dimensional data sets. The paper illustrated a useful means for visualising multivariate data sets using SOM based approaches along with simple statistical methods, the former as applied to industrial process dynamics modeling with proven success. The multiple regression technique explained in the paper as well produced satisfactory results in analysing the trends within natural habitats.

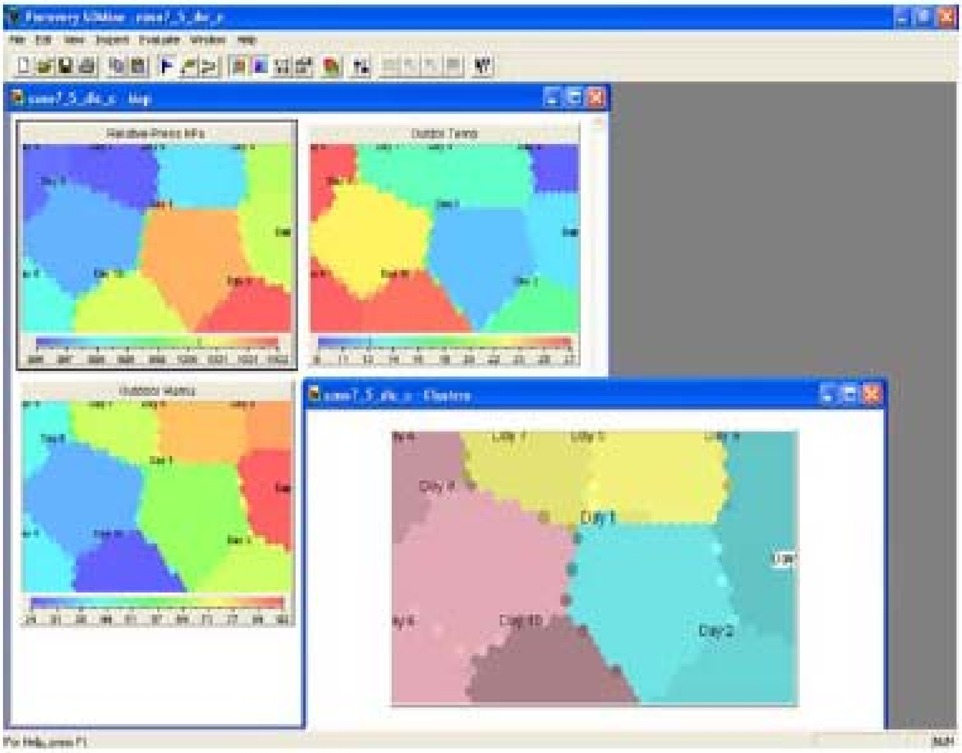


Figure 1. SOM of Climate Data (Sallis etal., 2008,p5)

In (Sallis et al.;2008), researchers illustrated an approach for analysing imprecise data in free-text format on human sensory perception relating to wine taste and quality integrated with precise data on grapevine growth factors. For the main project called **Eno Humanas,** the authors proposed a data synthesizer with several software modules, one of them is a weather data modelling module, which collects climate and environmental data in real time for anlaysis using computational neural network modelling methods and geographic information systems to depict weather and growth factors being studied. In the same project, (Sallis et al.;2008) the authors used a commercial software package, called Viscovery (www.eudaptics.com), as a tool to interpret variable dependencies within the raw data and to depict the project results. The software tool used for the analysis uses self organising map (SOM) techniques, first proposed by Kohonen**Error! Reference source not found.**. The weather prediction model elaborated upon in the paper uses meteorological data collected in real time especially, for the project and feeds the weather data as input to this software and produces a colourful display, a jigsaw-like illustration of value densities and data dependencies as shown in figure 1 from the original paper (Sallis et al.;2008]

In (Obach,2001) ,the authors modelled the total number of individuals of selected water insects based on a 30-year data set of population dynamics and environmental variables (discharge, temperature, precipitation, abundance of parental generation) in a small stream in central Germany. In this study Kohonen's SOMs were used for exploratory analysis of the data; the methods used in the study could be classified as data visualisation, outlier detection, hypothesis generation, and the detection of basic patterns in the raw data. Furthermore, they applied linear and general regression neural networks, modified multi-layer perceptrons, and radial basis function networks combined with a SOM (RBFSOM) and successfully predicted the annual abundance of selected species from environmental variables. Results were projected onto three-dimensional plots for easy visualisation purposes. RBFSOMs were used to denominate and visualise local and general model accuracy. Results were interpreted on the basis of known species traits. The authors further stated that in addition to the error measures discussed in the paper, several options existed for visualising the data and resulting models. They displayed similarities among prototypes either as rectangular maps (3×2) or as 'chains' (6×1) and concluded that the advantage with the 'chains' over two-dimensional SOMs as the wider variability of prototype vectors, despite the given low number of nodes.

The authors of (Walley and O'Connor,2001), described a novel pattern recognition system (MIR-max) that was developed to facilitate the construction of a river pollution diagnostic system for the British Environment Agency. MIR-max is a non-neural self-organising map based on information theory, which, unlike Kohonen's Self-organised map (SOM), separates the processes of clustering and ordering. A novel feature of MIR-max is that it permits the disaggregation of the classes in the output map, thus allowing exceptional classes to separate from their neighbours. They concluded that MIR-max has considerable potential for use in the visualisation and interpretation of multivariate ecological data.

Another investigation into the use of SOM for the classification of river quality using biological and environmental data (Walley et al.,1998 cited in Walley and O'Connor 2001;Walley et al.,1992cited in Walley and O'Connor,2001) produced promising results, but led the authors to conclude that the SOM algorithm as not ideally suited to that particular data set. In comparative tests of (Walley and O'Connor**Error! Reference source not found.** based on biological data derived from river quality surveys, MIR-max was shown to outperform SOM and GTM (Generative Topographic Mapping)(Bishop et al,1998) with respect to clustering and global ordering. Unlike SOM and GTM, MIR-max separates the clustering and ordering processes. The clustering process is based on information theory, whilst the ordering process is based upon the maximization of the correlation between corresponding distances in data space and output space. The separation of the two processes enables MIR-max to offer some powerful new features, such as the disaggregation of classes into 'natural' clusters in output space and the ability to order the classes on either a local or global scale (Walley and O'Connor**Error! Reference source not found.**. These papers reviewed so far revealed that SOM based algorithms alone might not be ideal for visualising environmental data.

However coupling them with other conventional computing algorithms could provide significant boost for use with a wide range of data. This would also create a synergy where non-neural algorithms could augment neural net applications, such as SOM.

3. RESEARCH QUESTIONS

In view of the above NN and non NN applications, this research is directed towards investigating into techniques that could enhance the visualisation of environmental data in viticulture and looks at all possible answers to resolving the following research questions based on the work so far discussed:

What are the major issues in visualising environmental data in viticulture?

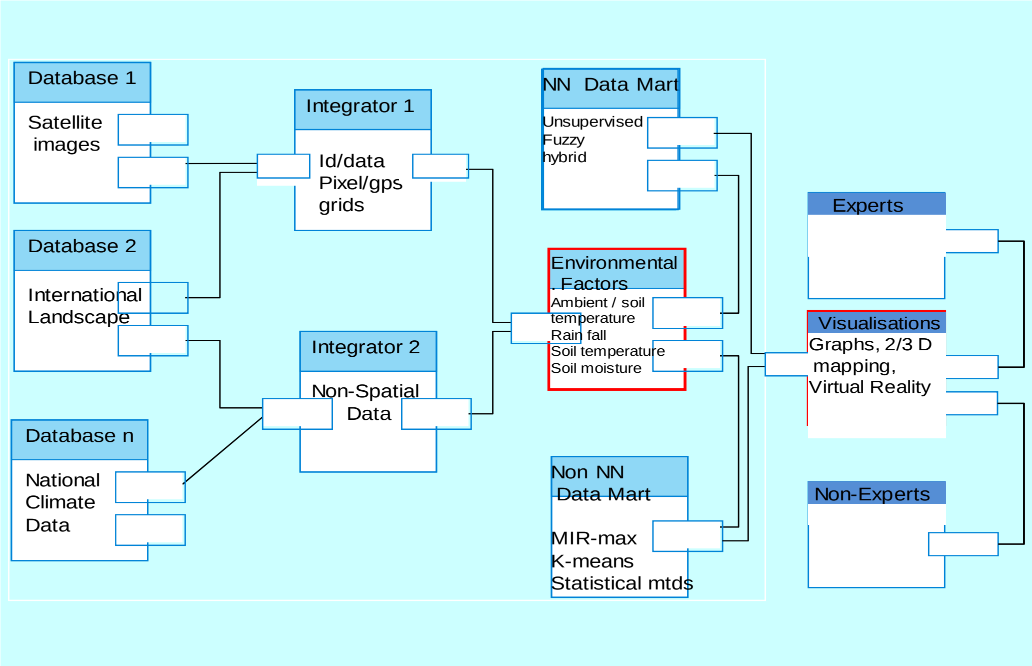
- *What novel features could be added to current visualisation techniques in determining what makes a good year for wine?*
- *How can these visualisation features be made useful to non experts in viticulture?*
- *Which kind of software architecture could best facilitate the visualisation of environmental data in viticulture?*

4. METHODOLOGY: MODELLING APPROACH USING VISUALISATION TECHNIQUES

The proposed modelling approach attempts to explore novel visualizations from two main categories or groups namely NN and non NN. NN approach encompasses various visualisation approaches that utilises computational neural networks, such as SOM. Those in the non NN category are novel ways that are in complimentary to those in the NN being explored to overcome the limitations of the former creating a synergetic effect. An example of a possible non NN visualisation could be MIR-max and other algorithmic techniques such as K-means. The main goal of the system is to integrate the two broad techniques into one system to enhance the visualization of environmental influence factors in viticulture. Figure 2 shows a proposed system for enhancing the visualization of environmental influence factors in viticulture. This approach seeks to develop an integrated visualization system that uses a distributed system to access data sources via a network. It is expected to provide levels of visualisation for users (experts and non-experts alike) depending on the goals they intend to achieve, i.e., identify, locate, distinguish, categorize, cluster, rank, compare, associate or correlate data (Wehrend and Lewis 1990 cited in Batagelj and Mrvar,1997).

The proposed system (as shown in Figure 2) is designed to use a distributed object approach to implement client-server system (Sommerville,2004,). In this approach each database would be encapsulated as a distributed object with an interface that provides read only access to its data. Integrator objects would each be linked to specific types of data –spatial or non spatial, and they collect information from all the databases. Wireless sensor data gathered from vineyards in these countries (New Zealand, Chile, Uruguay, Argentina,Japan and USA) will be analysed to produce a better visualisation to see any similarities/dissimilarities on climate change variability across the globe and the resulting grapevine responses. Each integrator object is linked to a Core Database . The Core Database (Core DB) contains reliable non-redundant collection of data, on the environmental influence factors, parts of which are made available for queries through NN and Non NN data marts designed for specific purposes and groups of users (Experts and Non Experts) . Visualisation objects would interact with data mart to produce visualisations for experts or non experts on the environmental influence factors. The system will have two functions, one to allow experts to choose the NN and statistical methods for analysing the data and the other for novice and when selected the latter will run different sets of pre-defined methods (NNs and other) producing results and methods used along with rationale for the selection of the methods.

Similar to Sommerville (2004), the system proposed herein (Figure 2). consists of a distributed object architecture with a data mining tool that looks for relationships between the data stored in a number of databases. The reasons for selecting a distributed object approach is that firstly, databases could be added to such a system without major disruptions and the database objects could provide a simplified interface that controls access to the data. Secondly, this system allows for the integration of new visualisation techniques just by adding new data marts.



NN= neural net techniques, Non NN= non neural net techniques

Figure 2. Proposed System for Visualisations of Environmental Factors in Viticulture

5. OUTCOME AND CONCLUSIONS

It is envisaged that this approach would lead to the development of novel visualisations to model environmental influence factors in viticulture. It would provide a system for assessing any environmental impact with novel data visualisation techniques for decision making and in determining what makes a good year for wine. The system would facilitate visualisation techniques over a distributed system such as the Internet and provide user interfaces useful to both experts and non experts in viticulture. The novelty of the proposed system lies in the combined use of these as well as statistical methods and more in visualising the results obtained from collective analysis of data from disparate sources (databases).This system would contribute to developing novel visualisation techniques for analyzing environmental data incorporating at different spatial and temporal scales.

6. REFRENCES

Batagelj V. and Mrvar A., (1997), Visualization of Multivariate Data Using 3D and VR . Indo-French

Workshop on Symbolic Data Analysis and its Applications. Paris XI - Dauphine, Vol 1., pp 66-76.

Bishop, C.M., Svense´n, M., and Williams, C.K.I., (1998). GTM: Generative topographic mapping. Neural Comput. 10 (1), 215–234. In Walley W. J., O'Connor M. A., Unsupervised pattern recognition for the interpretation of ecological data. A publication of Elsevier Science B.V. Ecological Modelling 146, 2001 pp 219–230

Jones, G.V., M.A. White, O.R. Cooper, and K. Storchmann4, Climate Change and Global Wine Quality.*Climatic Change*, (2005). **73(3)**, p. 319-343. As quoted in Shanmuganathan, S., P. Sallis., L. Pavesi, and M. C. Jarur Muñoz, Computational intelligence and geo-informatics in viticulture, in proceedings of the Second Asia International Conference on Modelling & Simulation being held in Kuala Lumpur, Malaysia 13 - 15 May 2008

Kohonen, T.,(1997), Self-Organizing Maps. Series in Information Sciences, Vol. 30. Springer, Heidelberg. 2nd Ed.

Obach M.,Wagner R., Werner H., Schmidt H-H.,(2001), Modelling population dynamics of aquatic insects withartificial neural networks. Elsevier Science B.V. *Ecological Modelling* 146, 2001 pp 207–217

Pereira, G.E., J.-P. Gaudillere, C.V. Leeuwena, G. Hilbert, M.e. Maucourt, C. Deborde, A. Moingb, and D. Rolin, (2005),HNMR metabolite fingerprints of grape berry: Comparison of vintage and soil effects in Bordeaux grapevine growing areas. Analytica Chimica Acta, 2005. **563**: p. 346-352. As quoted in Shanmuganathan, S., P.

Redpath R. and Srinivasan B.(2003),. Criteria for a Comparative Study of Visualization Techniques in DataMining.In Proceedings of the IEEE 3rd International Conference on Intelligent Systems Design and Application, Tulsa,USA,(ISDA 2003) Springer-Verlag, Berlin

Sallis., L. Pavesi, and M. C. Jarur Muñoz, Computational intelligence and geo-informatics in viticulture, in proceedings of the Second Asia International Conference on Modelling & Simulation being held in Kuala Lumpur, Malaysia 13 - 15 May 2008 (in press).

Sallis, P.J., Shanmuganathan, S., Pavesi, and L., and Muñoz, M.C.J.,(2008), A system architecture for collaborative environmental modelling research. The 2008 International Symposium on Collaborative Technologies and Systems (CTS 2008), Eds., Waleed W. Samari and William McQuay, A publication of the IEEE, New Jersey, USA., Irvine, California,May 19-23 2008 pp 39-47

Shanmuganathan, S., P. Sallis., L. Pavesi, and M. C. Jarur Muñoz,(2008) Computational intelligence and geo-informatics in viticulture, in proceedings of the Second Asia International Conference on Modelling & Simulation being held in Kuala Lumpur, Malaysia 13 - 15 May 2008 (in press).

Shanmuganathan, S. and Sallis, P. J., and Claster, W.,(2007) Statistical Methods in Ecological Dynamics Modelling. In Oxley, L. and Kulasiri, D. (eds) MODSIM 2007 International Congress on Modelling and Simulation Modelling and Simulation Society of Australia and New Zealand, December 2007, pp. 2967-2973.

Shanmuganathan, S., Cassim, M. and Ukita, K.,(2007) Environmental awareness education at Ritsumeikan's primary, secondary and tertiary institutions in Japan, in proceedings of the 6th Global Conference on Environmental Justice and Global Citizenship 2007 held by Inter-Disciplinary Net at Mansfield College, Oxford, UK, 2-5 July 2007.

Smith L (2002): Site Selection for Establishment & Management of Vineyards.Paper Presented at SIRC 2002 – The 14th Annual Colloquium of the Spatial Information Research Centre University of Otago, Dunedin, New Zealand. December 3-5th 2002

Sommerville I., (2004). *Software Engineering*. Pearson Education Limited, Essex. 7th Ed, pp265-283

Walley W. J., and O'Connor M. A., (2001), Unsupervised pattern recognition for the interpretation of ecological data. Elsevier Science B.V. *Ecological Modelling* 146, 2001 pp 219–230

Walley, W.J., Fontama, V.N.,and Martin, R.W., (1998), Applications of artificial intelligence in river quality surveys. R&D Technical Report E52. Environment Agency, Bristol. In: Walley W. J., and O'Connor M. A.,Unsupervised pattern recognition for the interpretation of ecological data. Elsevier Science B.V. *Ecological Modelling* 146, 2001 pp 219–230

Walley, W.J., Hawkes, H.A.,and Boyd, M., (1992), Application of Bayesian inference to river water quality surveillance. As quoted in Walley W. J., O'Connor M. A., Unsupervised pattern recognition for the interpretation of ecological data. Elsevier Science B.V. *Ecological Modelling* 146, 2001 pp 219–230

Wehrend S., Lewis C.(1997),: A Problem-Oriented Classification of Visualization techniques. In Proceedings of IEEE Visualization'90, 1990, p. 139-143. As quoted in Batagelj V. and Mrvar A. (1997). Visualization of Multivariate Data Using 3D and VR . Indo-French Workshop on Symbolic Data Analysis and its Applications. Paris XI - Dauphine, Vol 1., pp 66-76.

White, R.E., Soils for Fine Wines.,(2004),: Oxford University Press. 312. As quoted in Shanmuganathan, S., P. Sallis., L. Pavesi, and M. C. Jarur Muñoz, Computational intelligence and geo-informatics in viticulture, in proceedings of the Second Asia International Conference on Modelling & Simulation being held in Kuala Lumpur, Malaysia 13 - 15 May 2008 (in press).

Wilson, J. E. (1998). Terroir: The Role of Geology, Climate and Culture in the Making of French Wine. Iin Smith L (2002): Site Selection for Establishment & Management of Vineyards.Paper The 14th Annual Colloquium of the Spatial Information Research Centre University of Otago, Dunedin, New Zealand. December 3-5th 2002.