Time Dependent Neural Network Models For Detecting Changes Of State In Earth and Planetary Processes

Julio J. Valdés
National Research Council
Institute for Information Technology
M50, 1200 Montreal Rd. Ottawa, ON K1A 0R6
Canada
E-mail: julio.valdes@nrc-cnrc.gc.ca

Graeme Bonham-Carter
Geological Survey of Canada
601 Booth St.
Ottawa, Ontario K1A 0E8
Canada
E-mail: gbonhamc@NRCan.gc.ca

***Abstract*— This paper explores a computational intelligence approach to the problem of detecting internal changes in time dependent processes described by heterogeneous, multivariate time series with imprecise data and missing values. Processes are approximated by collections of time-dependent nonlinear AR models represented by a special kind of neuro-fuzzy neural networks. Grid and high throughput computing model-mining procedures using neuro-fuzzy networks and genetic algorithms, generate collections of models composed by sets of time lag terms from the time series, as well as prediction functions represented by neuro-fuzzy networks. The composition of the models and their prediction capabilities, allows the identification of changes in the internal structure of the process. These changes are associated with the alternation of steady and transient states, zones with abnormal behavior, instability, and other situations. This approach is general, and its potential is revealed by experiments using paleoclimate and solar data.**

I. Introduction

Time series data in the geosciences present difficult problems for analysis for a variety of reasons. The data are often incomplete (missing observations), imprecise (noisy) and heterogeneous (mixed scales of measurement), making conventional time series approaches unsatisfactory. Neural network models using similarity-based heterogeneous neurons and systematic analysis of complex lags offer an approach that is robust and sensitive. The ability to predict such changes of state (long or short term) has many important applications to natural systems. In complex or poorly known processes, knowledge discovery designed to uncover the underlying structure of the physical process is crucial, especially for revealing patterns and time dependencies, detecting abnormal behavior, instabilities, changes of state, deriving prediction criteria, and constructing forecasting procedures.

The paper discusses the use of a computational intelligence approach for model discovery and model-change detection in multivariate time processes with different kinds of variables, missing data and uncertainty. It is a hybrid approach to time-dependent model discovery, based on a combination of neural networks and evolutionary algorithms (genetic algorithms). A model of a time-dependent target variable is understood as a prediction function of its value at a given time, generally non-linear. The arguments of this function are past values of some or all of the variables involved in the multivariate process (the dependency pattern). The mathematical description is that of a non-linear multivariate autoregressive (AR) model when both the arguments, and the prediction function, are continuously changing with time. In non-linear dynamics and chaos theory there AR-type of dependencies, and genetic algorithms are used for finding proper lag terms as well. However, in the presented approach the nature of the time series is more general (composed of heterogeneous data), and chaos is not assumed.

In the present approach, the prediction functions are represented by hybrid neural networks using heterogeneous neurons in the first hidden layer, which accept as input heterogeneous, fuzzy and missing data. Instead of trying to find a global model for the whole multivariate process, like the conventional approaches, the discovery process proceeds as a continuous exploration along the multivariate series. The method finds sets of non-linear models for a target signal at time-intervals. The overall dependencies between the multivariate heterogeneous time series are characterized as probability distributions and error-cost functions over the sets of time lags of the ensemble of discovered dependency patterns. These distributions are represented as images (spectra), and are combined with the prediction-error curves associated with the discovered models. Their joint interpretation allows the segmentation of the multivariate process, and stable and transient states can be recognized.

From the methodological point of view, this approach can be considered as an abstract, conceptual filtering of multivariate time series of general character, which transforms the original heterogeneous, imprecise, and incomplete collection into a time series of models (with time-varying dependency patterns, and time-varying neural networks).

Applications of this approach to two datasets are presented here to demonstrate the potential of the approach: oxygen isotope data from ice cores, and sunspot cycles.

A. Heterogeneous Domains and Multivariate Time Series

A formal approach for describing heterogeneous information in general observational problems was given in [17], and for constructing neuron models in [13], [14], and [2]. Different information sources are associated with the attributes, relations and functions, and these sources are associated with the nature of what is observed (e.g. point measurements, signals, documents, images, etc). They are described by mathematical sets of the appropriate kind called source sets (Ψ_i), constructed according to the nature of the information source to represent (e.g. point measurements of continuous variables by subsets of the reals in the appropriate ranges, structural information by directed graphs, etc). Source sets also account for incomplete information. A heterogeneous domain is a Cartesian product of a collection of source sets: $\hat{H} = \Psi_1 \times \cdots \times \Psi_n$, where $n > 0$ is the number of information sources to consider. For example, consider a domain where objects are described by attributes like continuous crisp quantities, discrete features, fuzzy features, time-series, images, and graphs (missing values are allowed). Individually, they can be represented as Cartesian products of subsets of real numbers($\hat{R}$), nominal ($\hat{N}$) or ordinal sets($\hat{O}$), fuzzy sets($\hat{F}$), set of images ($\hat{I}$) , set of time series ($\hat{S}$) and sets of graphs ($\hat{G}$), respectively, all properly extended for accepting missing values. Thus, the heterogeneous, time dependent domain is $\hat{H}^n(t) = \hat{N}^{n_N}(t) \times \hat{O}^{n_O}(t) \times \hat{R}^{n_R}(t) \times \hat{F}^{n_F}(t) \times \hat{I}^{n_I}(t) \times \hat{S}^{n_S}(t) \times \hat{G}^{n_G}(t)$, where n_N is the number of nominal sets, n_O of ordinal sets, n_R of real-valued sets , n_F of fuzzy sets , n_I of image-valued sets, n_S of time-series sets, and n_G of graph-valued sets, respectively ($n = n_N + n_O + n_R + n_F + n_I + n_S + n_G$). A multivariate, heterogeneous time series is shown in Fig-1.

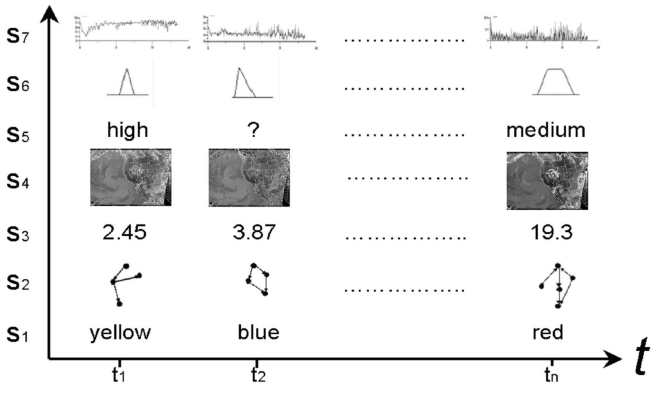

Fig. 1. An example of a heterogeneous, time-dependent multivariate process. Each row is a time series of a different type: nominal, graph, ratio, image, ordinal, fuzzy, time-series. Attributes may have missing values (?). The sampling interval is assumed to be the same (and synchronized) for each of the individual series.

B. Model Mining with Heterogeneous Neurons and Hybrid Neural Networks

The purpose of model mining in heterogeneous, multivariate, time varying processes is to discover dependency models. A model expresses the relationship between values of a previously selected time series (the target), and a subset of the past values of the entire set of series. Different classes of functional models could be considered, in particular, a generalized non-linear auto-regressive (AR) model like the one given by Equation-1.

$$S_T(t) = \mathbf{F}\begin{pmatrix} S_1(t-\tau_{1,1}), \cdots, S_1(t-\tau_{1,p_1}), \\ S_2(t-\tau_{2,1}), \cdots, S_2(t-\tau_{2,p_2}), \\ \cdots \\ S_n(t-\tau_{n,1}), \cdots, S_n(t-\tau_{n,p_n}) \end{pmatrix} \quad (1)$$

where $S_T(t)$ is the target signal at time t, S_i is the i-th time series, n is the total number of signals, p_i is the number of time lag terms from signal i influencing $S_T(t)$, $\tau_{i,k}$ is the k-th lag term corresponding to signal i ($k \in [1, p_i]$), and $\mathbf{F}$ is the unknown function describing the process.

The classical approaches in time series consider mostly univariate, homogeneous (real-valued), time series, without missing values [3], [12], [11], [9]. Conventional multivariate approaches are complex and have difficulties in handling heterogeneity, imprecision and incompleteness.

A hybrid soft-computing algorithm for approaching this kind of problems using heterogeneous neural networks and genetic algorithms has been given elsewhere [15], and it goes along other approaches within the neural network field [10]. This approach requires the simultaneous determination of: (i) the number of required lags for each series, (ii) the particular lags within each series carrying the dependency information, and (iii) the prediction function. A requirement on function F is to minimize a suitable prediction error, usually the root mean squared error (RMS Error). The procedure is based on: (a) exploration of a subset of the model space with a genetic algorithm, and (b) use of a similarity-based neuro-fuzzy system representation for the unknown prediction function. As mentioned, statistical or other classical approaches either have difficulties handling these kinds of situations or cannot handle them at all. The size of the model space is immense, and grows exponentially as the value of the maximal lag included in the model increases. Considering only 10 time series and the first 20 time lags, the search space contains about 10^{60} models. The prediction function F is represented by a hybrid neural network with a hidden layer composed by heterogeneous neurons (h-neurons). A heterogeneous neuron is a general mapping $h : \hat{H} \times \hat{H} \rightarrow Y$, where $\hat{H}$ is a heterogeneous domain, and Y is an arbitrary set. If $x, w \in \hat{H}$, and $y \in Y$, then $y = h(x, w)$. A particular class of h-neurons is obtained when Y is the real interval [0,1] and h is given by a composition of a similarity function s [4], and an isotone automorphism $g : [0, 1] \rightarrow [0, 1]$ (usually a non-linear function). In this case the h-neuron is given by $h = g \circ s$, and called a similarity-based neuron (s-neuron) [13] (Fig-2).

This neuron model is flexible (heterogeneous data with missing values are its natural input, without the need of data type transformation or imputation of missing values), and it is robust. Networks using this neuron also have the general function approximation property [2]. The s-neuron can be

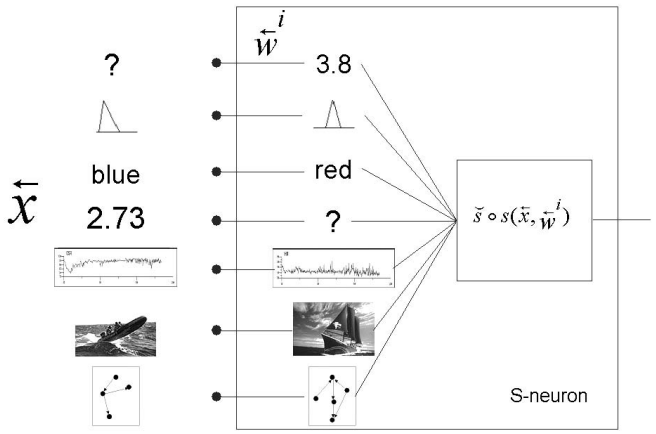

Fig. 2. . A similarity-based heterogeneous neuron. Both the input and the neuron weights are objects from a heterogeneous domain (? is a missing value). The output is a similarity value.

coupled with classical neurons (aggregation function given by the scalar product, and activation given by the sigmoid or hyperbolic tangent), forming hybrid neural networks. In the particular case of homogeneous, real-valued source sets, fast training algorithms can be used, as proposed in [18]. An implementation of this model discovery approach as a parallel computing algorithm (Fig-3), was reported in [16].

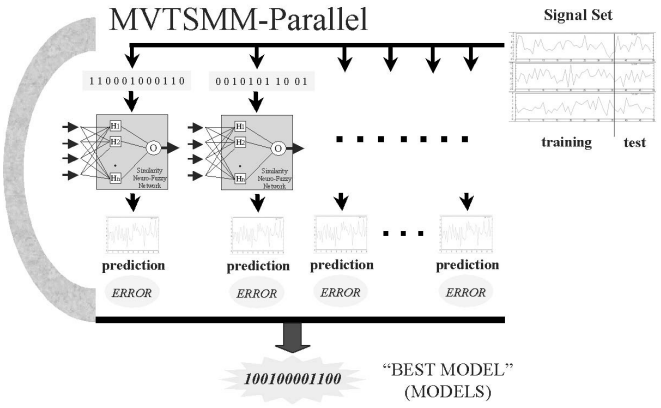

Fig. 3. Multivariate Time Series Model Miner System (MVTSMM). The arc is a parallel genetic algorithm evolving populations of similarity-based hybrid neural networks. The binary strings encode dependency patterns for the target signal. For each, a hybrid neural network is constructed and trained with a fast algorithm. The network represents the prediction function, and is applied to an independent test set. The best models (those with their RMSError under a preset threshold), are collected.

C. Models as a Function of Time

Evolving through a sequence of changes of state, where stable states of different nature are separated by the corresponding transient states, is a situation typical from most physical systems, including earth and planetary processes. In these cases, the relationship between the variables involved in the process changes with time, and so will do the models describing them. Accordingly, it is important to find the points along the time series, where the changes are taking place. A statistical approach for their estimation in the case of real-valued, univariate time series is given in [1]. It is based on numerous assumptions, which are difficult to consider in the case of multivariate, heterogeneous, imprecise and incomplete series. In any case, regardless of the type of models used for describing the physical process, they are no longer going to be the same as the process evolves. In other words, the models are going to be a function of time, and in the present approach it means that either the dependency pattern, the prediction function (the neural network), or both, are going to change with time. Theoretical and practical considerations for a computational intelligence approach to time varying model discovery, are given in [19], where general nonlinear AR models as expressed by eq-1 are interpreted as random variables. Of particular importance is the concept of lag probability function (lpf), given by: $\mathcal{L}^p_\epsilon(t, \tau_{p,q}) = P(\tau_{p,q} \in \hat{\mathcal{M}}_{\epsilon_T})(t)$, where t is a given time, $\tau_{p,q}$ is the p-time lag term from time series q, as defined in (1), ϵ_T is a prediction error threshold, $\hat{\mathcal{M}}$ is the set of all discovered models, and $\hat{\mathcal{M}}_{\epsilon_T} \in \hat{\mathcal{M}}$ is the subset of those whose prediction error is upper bounded by ϵ_T. The lpf expresses the probability with which a particular time lag will appear in ϵ_T-good discovered models (those with their prediction errors upper-bounded by the given error threshold). The rationale is that if a given lag appears systematically within the dependency pattern of models with good performance, then that lag carries important dependency information from the point of view of predicting the process at the time t.

Another view of the importance of the lag terms of eq-1 considering $\hat{\mathcal{M}}$ as a whole, is given by the *weighted lag importance function* (wlif):

$$\mathcal{L}^w(t, \tau_{p,q}) = \frac{\sum_{i=1}^{card(\hat{\mathcal{M}})} \mu(\tau_{p,q}, \hat{\mathcal{M}}_i(t))/E(\hat{\mathcal{M}}_i(t))}{\sum_{i=1}^{card(\hat{\mathcal{M}})}(1/E(\hat{\mathcal{M}}_i)(t))} \quad (2)$$

where $card(\hat{\mathcal{M}})$ is the number of discovered models, $\hat{\mathcal{M}}_i(t) \in \hat{\mathcal{M}}$ is the i-th model found at time t, $\mu(\tau_{p,q}, \hat{\mathcal{M}}_i(t))$ is the boolean membership function of $\tau_{p,q}$ w.r.t. $\hat{\mathcal{M}}_i(t)$, and $E(\hat{\mathcal{M}}_i(t))$ is a prediction error measure on $\hat{\mathcal{M}}$. The wlif transforms into a lpf (a frequency), if all of the models in $\hat{\mathcal{M}}$ are ϵ_T-good, and it is a more general measure, as it works with all of the models and not only with a specific subset of them. The wlif can be represented as a image spectrum in the same way as it is done with the lpf [19].

In a stable process the set of important lags, would be approximately the same with time. On another hand, the models discovered at any time t, are the result of a search in the space of such networks. Clearly, the set of discoverable/discovered models will be conditioned by the nature and parameters of the search, and by the nature of the underlying physical process. Therefore, differences between the probability distribution of the prediction error associated with the sets of k-best models found for two different times t_1, t_2, would be also an indicator of model change, provided that both sets of models were obtained under the same search conditions.

From the point of view of the dependency patterns (the inputs of the neural network), a procedure for exploring model changes [19] would consist of: *i)* specify a time frame of a given length, *ii)* fix the set of parameters of the model mining algorithm, *iii)* collect the set of dependency patterns, and of neuro-fuzzy networks representing the corresponding prediction functions, and *iv)* compute the lag probability spectrum or the weighted lag importance spectrum, and the mean prediction error, for all time frames. In addition, differences in the probability distribution of the prediction error (for the same sets of models described by the lag probability spectra), would also serve as indicator of change.

II. Examples from Earth and Planetary Processes

With the purpose of start a systematic investigation of the behavior and properties of the proposed methodology, time series from the following natural processes were used:

- Sunspot numbers
- Greenland ice-core data

For simplicity, only univariate, homogeneous, real-valued, non-fuzzy, and non-incomplete time series data were chosen. They are particular cases of the kind covered by the presented approach, but the algorithms applied were those of the general case. Further studies will use more complex kinds of data.

A. Experimental settings

The algorithm was applied using the following parameters: number of responsive neurons in the hidden layer=3, 5, 7, 9, similarity function for the neuron model derived from the Euclidean distance (d) as $s = 1/(1+d)$, number of generations=100, 200, 300, 500, population size=50, 100, 200, roulette-wheel selection, single-point crossover, crossover probability=0.6, 0.7, 0.8, 0.9, mutation probability=0.01, 0.025, 0.05, and elitism. Sliding time frames of size 101 were set, exploring models up a maximum depth of 15, and 20 time lags (10 and 20 for the ice core data). Three different seeds were used for the random number generators. Within each frame, the first 75 values were used as training, and the remaining as test. For each of the data sets 3456 experiments were performed. Each produces an amount of models equal to the population size of the genetic algorithm for each sliding time window (sometimes only the 10-best were kept). Data processing was performed using distributed and grid computing facilities at the National Research Council Canada involving a variable set of approximately 70 computers.

B. Sunspot numbers

Sunspot cycles are well studied, because of their importance in affecting solar radiation received at the Earth's surface, and their potential for affecting global climate. The data used are the annual mean sunspot numbers in the period 1749-2003, as reported by the National Geophysical Data Center (USA, *www.ngdc.noaa.gov/stp/SOLAR/SSN/ssn.html*).

In this case 155 different time locations were investigated (from 1799 to 1953). For each time, 109440 models were kept, for a total of 16963200 models computed using a maximum lag depth of 15. The behavior of the RMS error for all of these models is shown in Fig-4. It is a multimodal distribution suggesting a mixture of different populations, with the first spawning from $[8, 25]$, the second from $[25, 34]$, and the third from $[35, 51]$. There is a positive skewness (with a mode around 12), many times larger than the second most frequent mode (at around 43), associated with the population of error values in the interval $[34, 51]$.

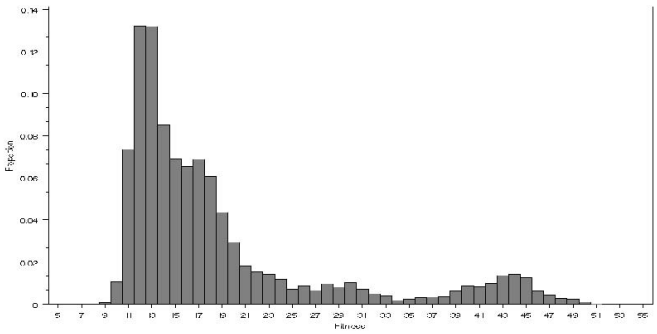

Fig. 4. RMS Error distribution for the 16963200 models found for the sunspot numbers series, with maximum time lag=15.

The kind of prediction obtained with models in the lower end error range is illustrated in Fig-5, corresponding to a model given by eq-3 (RMS Error = 8.897).

$$S(t) = \mathbf{F}\left(\begin{array}{l} S(t-1), S(t-2), S(t-4), S(t-5), \\ S(t-7), S(t-11), S(t-18) \end{array} \right) \quad (3)$$

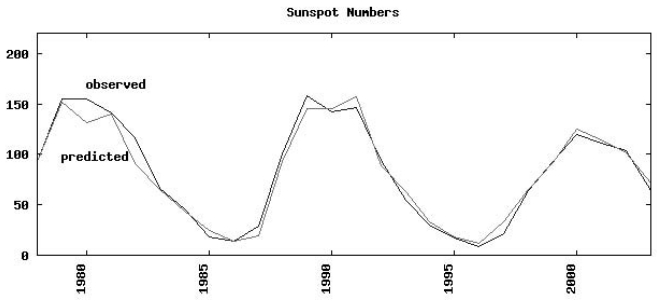

Fig. 5. Sunspot numbers data between 1978 and 2003, with the observed and the predicted values corresponding to the model explained in eq-3 (RMS error = 8.897).

Besides the good predictive behavior of the discovered model, it is interesting to note the presence of a time lag of value 11, which coincides with the classical cycle length of 11.1 years determined by the Swiss astronomer Wolf in 1848.

The weighted lag importance spectra (eq-2) corresponding to the maximum lags of 15 and 20 years, as well as the mean functions of the root mean prediction errors for all models as a function of time, are shown in Fig-6. The spectra are shown in the form of images. Brightness is proportional to the importance of a given lag, and each spectrum is normalized by the value of its most important lag.

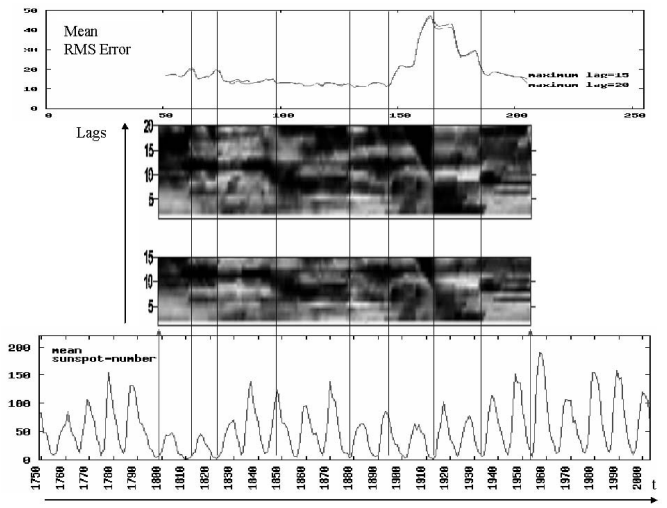

Fig. 6. Sunspot numbers data between 1749 and 2003, with the weighted lag importance spectra (for maximum lags of 15 and 20), and the mean RMS prediction error. 16963200 models were investigated. In the spectrum, brightness is proportional to lag importance. Vertical lines indicate approximate spectrum landmarks where changes in model composition are observed. They correspond well with the behavior of the error function.

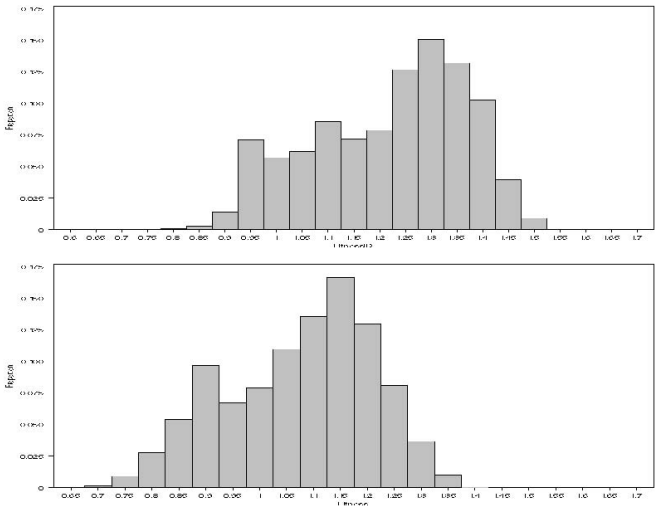

Fig. 7. RMS Error distributions for the Delta O18 data, for maximum time lags of 10 (top), and 20 (bottom). 1987200 models were computed for each maximum lag.

The mean error functions for the models with maximum lags of 15 and 20 coincide almost completely. This indicates that looking for dependencies farther than 15 years does not lead to better models. Moreover, there is a clear time period ($[1893, 1913]$ approximately), in which the prediction quality falls approximately 4 times.

The spectra for 15 and 20 maximum lags show basically the same information in terms of lag patterns. From them, approximate time landmarks at $\{1812, 1822, 1847, 1878, 1893, 1913, 1932\}$ can be recognized, and they are depicted with vertical lines in Fig-6. They correspond well with the main features of the error function.

C. Greenland ice-core data

Ice cores from various parts of the world have been studied intensively, and Greenland cores have received particular attention because they have the potential to record changes in temperature from oxygen isotope values over an extended period of history. The data used are annual averages (May -April) of delta 18-O (per mil), in the period 1761-1975 ([5], [6], [7], [8]), from the World Data Center for Paleoclimatology (Boulder, USA, www.ngdc.noaa.gov/paleo/icecore/greenland/gisp/campcentury /campc_data.html). The Delta O18 is the change of the ratio of oxygen isotopes O18/O16 and is often used as a measure of atmospheric temperature.

According to the experimental settings, and the time period covered by this data, 115 different time locations were investigated (from 1811 to 1925). For each time, 17280 models were kept, for a total of 1987200 models computed using a maximum lag depth of 10 (the same holds for 20). The behavior of the RMS error for all models is shown in Fig-7.

The error distributions for the models with a maximum time lags of 10 and 20 have the same type of skewness and are multimodal, but they differ in their statistical descriptors (mean= 1.218, standard dev.= 0.149, minimum = 0.819, maximum= 1.538, median = 1.250 for maximum lag = 10, vs, mean= 1.075, standard dev.= 0.135, minimum = 0.665, maximum= 1.410, median = 1.099 for maximum lag = 20). The multimodal character of the error distributions indicates a different behavior of the process with time, as the same search controlling parameters were used on each case.

The weighted lag importance spectra corresponding to the maximum lags of 10 and 20 years, as well as the mean functions of the root mean prediction errors for all models as a function of time, are shown in Fig-8.

Changes in lag distribution with time for models computed to a maximum lag of 10 years, are shown with vertical lines placed at the time points where the changes in the pattern of lag distribution are observed. They correspond approximately to years $\{1827, 1843, 1867, 1877, 1894, 1903, 1914\}$. The behavior of the mean RMS error with time for 10-years maximum lag models, also exhibits variations whose landmarks correspond quite well with those suggested by the spectrum. When a deeper search is considering by looking at models with lag terms up to 20 years, the behavior of the mean prediction error is very similar w.r.t 10 years, but the errors are in all cases better. The spectrum shows that the most important lags are actually those located approximately between $12 - 15$ years in the past. This indicates that when values of Delta O18 at times greater than 10 years before the current time are considered, the prediction accuracy of the models increases systematically. It may also suggests that even deeper maximum lags should be considered in further studies, also focussing on the meaning of the time landmarks found. Most of the spectral landmarks coincide with those found independently for the sunspot numbers data, and in

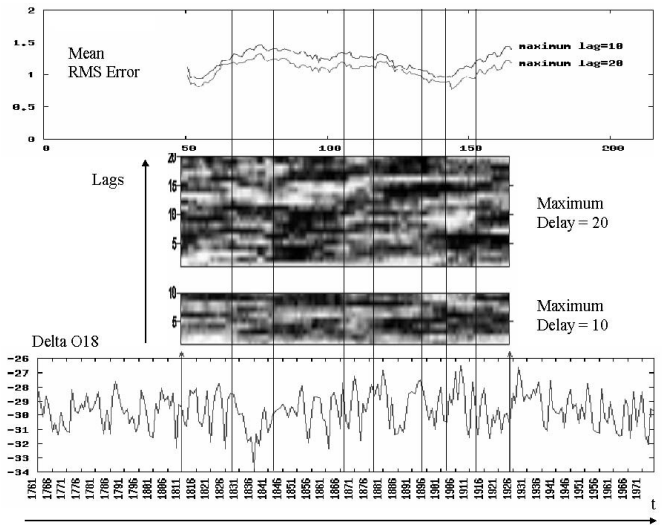

Fig. 8. Delta O-18 data from Greenland (Camp Century), with the weighted lag importance spectra, and the mean RMS prediction error. In the two spectra, brightness is proportional to lag importance. Vertical lines indicate approximate spectral landmarks where changes in model composition are observed. They correspond well with the behavior of the mean error functions.

particular, two of them $\{1894, 1914\}$ delimiting regions of sharp changes, are almost identical. This would reinforce the idea of a relationship between the two processes (mentioned elsewhere), which should be further investigated.

III. Conclusions

The application of the proposed approach to the detection of changes of state in time dependent processes through model discovery, allowed the detection of regions exhibiting differential behavior in terms of the sets of past values of the process, and the quality of their associated prediction models. The use of computational intelligence tools in a grid computing environment proved to be a very effective way for model mining applied to complex problems, like those studied by earth and planetary sciences. The mining process lead to good prediction models, as illustrated by the sunspot-numbers case. In particular the results obtained for paleotemperature, and solar activity data suggest a relation between the two processes in terms of the changes of state observed in them independently, which should be investigated further.

Acknowledgment

The authors would like to thank: Ratilal Haria, Gabriel Mateescu, and Roger Impey (National Research Council Canada, NRC) for their assistance in using high performance computing facilities, and to Luis M. Boleda for his help during the data processing stages. This work is part of the ADAM and WILD Miner projects of the Group of Integrated Reasoning of the Institute for Information Technology (NRC).

References

[1] J. Antoch and M. Huškova, Estimators of Changes. *Asymptotics, Nonparametrics, and Time Series (S. Gosh ed)*. Marcel Dekker Inc, New York, 1999, pp. 533–578.

[2] Ll. Belanche, *Heterogeneous neural networks: Theory and applications*. PhD Thesis, Department of Languages and Informatic Systems, Polytechnic University of Catalonia, Barcelona, Spain, July 2000.

[3] G. Box and G. Jenkins, *Time Series Analysis, Forecasting and Control*, Prentice Hall, 1976.

[4] J.L. Chandon and S. Pinson,*Analyse typologique. Thorie et applications* Masson, Paris, 1981.

[5] W. Dansgaard, S.J. Johnsen, J. Moller, and C.C. Langway, Jr. One thousand centuries of climatic record from Camp Century on the Greenland ice sheet. *Science* vol. 166, 1969 pp. 377–381.

[6] L.B. Hansen, and C.C. Langway, Jr. Deep core drilling in ice and core analysis at Camp Century, Greenland, 1961-1966. *CRREL Spec. Report 126, no. 5*, 1966, pp. 207–208.

[7] S.J. Johnsen, W. Dansgaard, H.B. Clausen, and C.C. Langway, Jr. Climatic oscillations 1200-2000 A.D. *Nature* vol. 227, 1970, pp. 482–483.

[8] S.J. Johnsen, W. Dansgaard, H.B. Clausen, and C.C. Langway, Jr. Oxygen isotope profiles through the Antarctic and Greenland ice sheets. *Nature* vol. 235, 1972 pp. 429–434.

[9] T. Masters, *Neural, Novel & Hybrid Algorithms for Time Series Prediction*, John Wiley & Sons, 1995.

[10] V. Petridis and A. Kehagias, *Predictive Neural Networks. Applications to Time Series*. Kluver 1998.

[11] A. Pole, M. West, J. Harrison, *Applied Bayesian Forecasting and Time Series Analysis.*, CRC Press, 1994.

[12] G. C.. Reinsel, *Elements of Multivariate Time Series Analysis.*, Springer Verlag, 1993.

[13] J. J. Valdés, and R., García, A model for heterogeneous neurons and its use in configuring neural networks for classification problems. *Proceedings of the Int. Work-Conf. on Artificial and Natural Neural Networks, IWANN'97. Biological and Artificial Computation: From Neuroscience to Technology*. *Lecture Notes in Computer Science 1240, Springer-Verlag*, Lanzarote, Spain, 1997, pp 237–246.

[14] J. J. Valdés, Ll. Belanche, R. Alquézar, Fuzzy Heterogeneous Neurons for Imprecise Classification Problems. *International Journal of Intelligent Systems* vol. 15, no. 3, 2000, pp. .

[15] J. J. Valdés, Time Series Models Discovery with Similarity-Based Neuro-Fuzzy Networks and Evolutionary Algorithms. *Proc. IEEE World Congress on Computational Intelligence WCCI'2002*. Hawaii, USA, 2002.

[16] J. J. Valdés and G. Mateescu, Time Series Models Discovery with Similarity-Based Neuro-Fuzzy Networks and Genetic Algorithms: A Parallel Implementation. *Third. Int. Conf. on Rough Sets and Current Trends in Computing RSCTC 2002*. Malvern, PA, USA. *Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence Series)* LNCS 2475, Springer-Verlag, 2002, pp. 279–288.

[17] J. J. Valdés, Similarity-based heterogeneous neurons in the context of general observational models *Neural Network World* vol. 5, 2002, pp. 499–508.

[18] J. J. Valdés and A.J. Barton , Multivariate Time Series Model Discovery with Similarity-Based Neuro-Fuzzy Networks and Genetic Algorithms. *Proceedings of the IEEE, INNS, IJCNN 2003 International Joint Conference on Neural Networks*, Portland, Oregon, July 20-24, 2003, pp. 1945–1950.

[19] J. J. Valdés, Detecting Changes of State in Heterogeneous Dynamic Processes with Time-dependent Models: A Soft-Computing Approach. *Proceedings of the IASTED International Joint Conference on Artificial Intelligence and Soft Computing, ASC'2004*, Marbella, Spain. ACTA Press, Anaheim, USA, 2004, pp. 155–160.