TREE BASED MODELING, PREDICTION AND ANALYSIS OF CHAOTIC TIME SERIES

Olivier Michel, [□] *Alfred Hero,* [◇] *Anne-Emmanuelle Badel* [□] *and Patrick Flandrin* [□]

[□] Laboratoire de Physique (URA 1325 CNRS)
Ecole Normale Supérieure de Lyon
46 allée d'Italie, 69364 Lyon Cedex 07, France

e-mail : omichel@physique.ens-lyon.fr
phone : (33) 72 72 83 78, fax : (33) 72 72 80 80.

[◇] Department of Electrical Engineering and Computer Science
University of Michigan
Ann Arbor, MI 48109-2122, USA

Tree based models were first introduced as a non-parametric exploratory data analysis technique for non-additive statistical models [1]. The tree-based model represents the data in a hierarchical structure where the leaves of the tree induce a non-uniform partition of the data space. Each leaf can be labeled by a scalar or vector value of a one-step predictor, a non-linear response variable, or a multi-variable quantizer output. Once the tree was grown, it can be used for classification, non linear prediction or quantization. In this paper, the quantized prediction that may be derived from the tree is emphasized. Furthermore, it is shown how the prediction error of a 1-step forward tree based predictor gives rise to some estimation of the "best" embedding dimension, when applied to chaotic time series analysis.

We use the Takens time delay embedding method to construct a phase space for the signal which captures the linear or non-linear dynamics of any finite dimensional state model. We apply recursive tree growing techniques to specify an optimal tiling of the phase space which represents the best piecewise constant approximation to the joint probability density function. The partitioning is accomplished by adding or deleting branches (nodes) of the tree according to a maximum entropy principle: to test that the joint distribution is approximately uniform within any candidate partition we compare the conditional entropy of the data points in the candidate partition to the maximum conditional entropy.

The attractive features of the tree based approach are the following: no parametric model is required, unlike the likelihood approaches, however if one is available it can be incorporated into the tree structure as a constraint. Furthermore, since tree model is constructed on the quantiles of the joint distributions, it is stable even in the case of heavy tailed densities, unlike higher order statistical approaches. Finally, unlike either likelihood or moment based models, the tree structured signal model is invariant with respect to monotonic non-linear transforms of the data.

1. DESCRIPTION OF TREE-BASED APPROACH

Let $X_k,\ k = 1, \ldots, N$ be an observed data sample from some ergodic random process $\{X_k\}$. The objective is to find a good partition of the p-dimensional phase space that allows accurate representation of the joint histograms of order p by piecewise uniform joint densities. To this end, we perform recursive partitioning, from coarse to fine partitions, by the following: for a given partition at level l, a rectangle in the p-dimensional phase space called a parent node, we develop a 2^p-ary partition at level $l+1$, called children-nodes, using a min-max splitting rule described below. Then the sample distributions of data points over the set of children-nodes are each tested against the uniform distribution (null hypothesis H_0) via the Chi-square goodness of fit test. If the null hypothesis is rejected, the split is memorized, along with the resulting subsets, and 2^p parent nodes at level $l+1$ are created. Otherwise, the splitting procedure is stopped and the parent node at level l is declared a terminal node. The set of terminal nodes are called the leaves of the tree. See Fig. 1 for an illustration of the tree growing procedure.

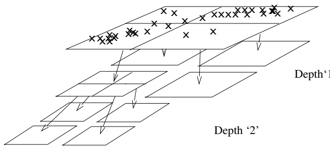

Figure 1: For a $p = 2$ dimensional phase space embedding, the root-node is split into 4 subsets and the distribution is found be non-uniform. Among the derived subsets, only the one depicted by the lower left corner square was found to be non-uniform and split further.

The splitting rule : At level l we realize a 2^p partition of the phase space creating at level $l+1$ 2^p cells where p is the phase space dimension and test the uniformity of the distribution (null hypothesis H_0) with the Chi-square test. If the distribution is uniform then the splitting is stopped and it is a terminal node else this subset is memorized, 2^p parent nodes are created at level $l+1$ and the iteration of the partitioning with these new parent nodes is continued. We must choose the threshold on each axis for realizing the splitting: it will be the marginal median such that half of the projected sample falls to the right and half to the left of the threshold. It is shown that this splitt maximises the entropy of the resulting partition, for the case of statisticaly independent coordinate of the state vector. By this way, the maximum of the marginal histograms's error is also minimized. The process is stopped when there are no enough points to have a test with enough significance or when the distribution is uniform.

The final partition of the phase-space is described by the set of leaves of the tree, together with the probability for a realization of state vector to be found within each of these leaves.

2. APPLICATION TO PREDICTION

Let $X_{n+1} = F_{X_n}(X_n) + \varepsilon_n$ be the sampled form of a p-dimensional dynamical system equation. F_{X_n} depicts the dynamical behavior of the system when the state vector has value X_n, that is F may be state-dependent. ε can be regarded as a realization of an observation noise or as a deterministic state perturbation. We first derive a tree based on a "learning set" $X_n, 1 \leq n \leq N$ which creates cells in the phase space $\mathcal{R}^p$ within which the distribution of the realizations of the state vector was estimated to be uniform (up to the accuracy of the Chi-squared test. A predicted value is associated with each subset or cell of the partition,

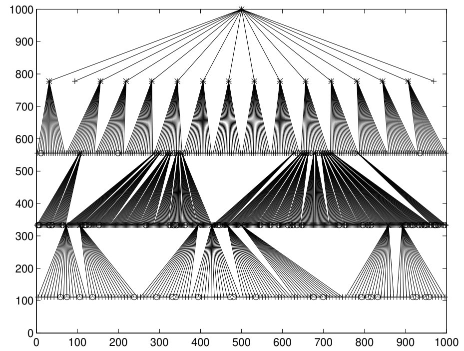

Figure 2: Tree estimated from a 4-d representation of the experimental system

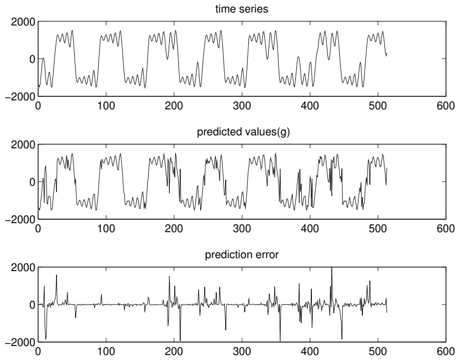

Figure 3: Tree based 1 step forward prediction and prediction error for the experimental system

e.g. x_{n+1}, according to the max-entropy (conditional MAP over the subset) criterion, thus creating an optimal tree-structured quantized predictor. Figure (2) shows the tree obtained from a 4-dimensional embedding of an experimental chaotic circuit ([6]). 1-step prediction is performed as described above, from the MAP estimate of X_{n+1} within each cell. The result is presented in figure(3). Figures (4) and (5) are the results obtained for the well-known Lorentz system, for a 3-dimensional reconstructed phase space. The choice of the embedding delay is discussed in [[4]].

This prediction method is compared with the nearest neighbors's prediction method [[7]].

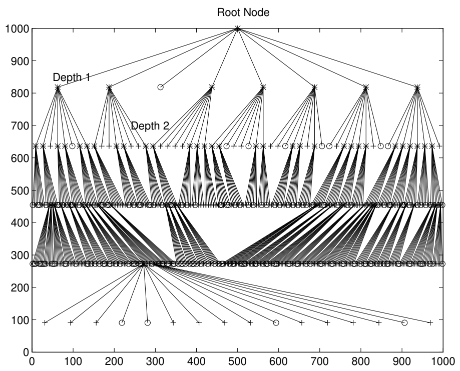

Figure 4: Tree estimated from a 3-d representation of the Lorentz's system

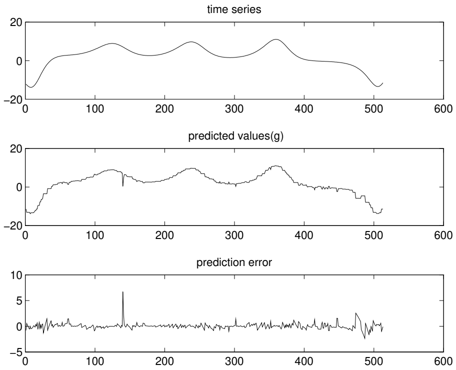

Figure 5: Tree based 1 step forward prediction and prediction error for the Lorentz's system

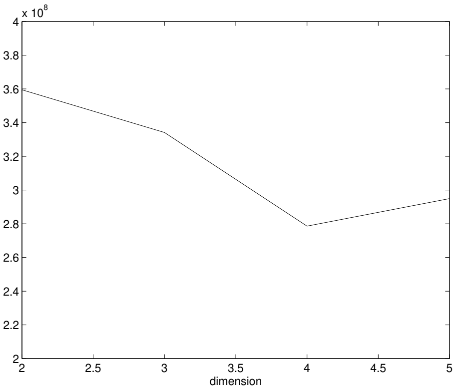

Figure 6: Tree based 1 step forward prediction error for the experimental system

3. TREE BASED ESTIMATION OF AN "OPTIMAL" EMBEDDING DIMENSION

Tree-based prediction was preformed for different embedding dimensions. The reconstruction delay wasthe same for each prediction experiment, chosen to be equal to the first minimum of the auto-correlation function. The study of the error variance as a function of the embedding dimension exhibits a minimum, which indicates some "optimality" (with respect to that measure) of a given dimension. All trees were estimated from a learning sequence of 8192 points. The results are compared with the error given by the nearest neighbor prediction method, performed from the same learning sequence and for the same embedding dimension. Figures (6) (7) (8) and (9) show the prediction error variance as a function of the reconstruction dimension, for both the experimental double-scroll system and the Lorentz's system, for the tree-based and neirest neighbor approaches. These results are discussed, together with the effect on the length of the learning sequence over the accuracy of the prediction.

REFERENCES

[1] J.N. Sonquist and J.N. Morgan, "The detection of interaction effects," Monograph 35, Survey Research Center, Institute for Social Research, University of Michigan, 1964.

[2] L. Breiman, J.H.Friedman, R.A. Olshen, C.J. Stone, "Classification and Regression Trees," Wadsworth Advanced Books and Software, 1984.

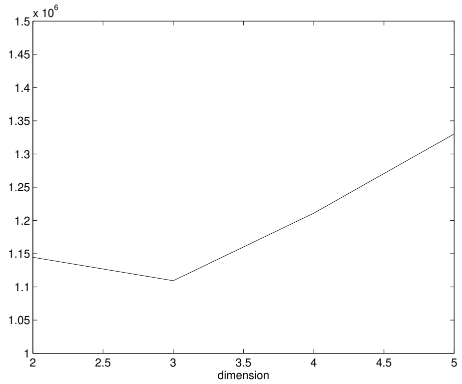

Figure 7: Nearest neighbors's 1 step forward prediction error for the experimental system

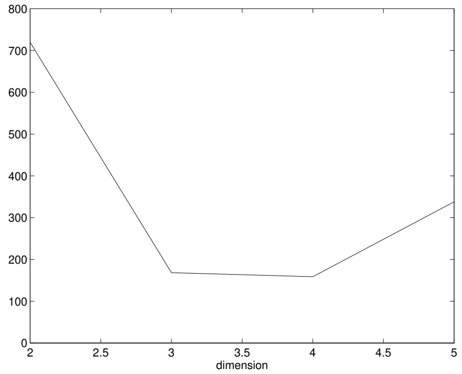

Figure 8: Tree based 1 step forward prediction error for the Lorentz's system

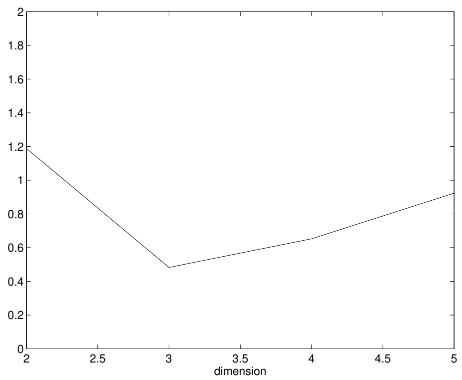

Figure 9: Nearest neighbors's 1 step forward prediction error for the Lorentz's system

[3] P.A. Chou, T. Lookabaugh, R.M. Gray, "Optimal Pruning with Applications to Tree-Structures Source Coding and Modeling," *IEEE Trans on Inf. theory*, vol.35, No.2, 1989, pp 299-315.

[4] J.P. Eckmann and D. Ruelle, "Ergodic Theory of Chaos and Strange Attractors," *Rev. Mod. Phys.*, Vol. 57, No. 3, pp. 617–656, 1985.

[5] A.M. Fraser, "Information and Entropy in Strange Attractors," *IEEE Trans on Inf. theory*, vol.35, No.2, 1989, pp 245-262.

[6] T.P. Weldon, "An Inductorless Double-Scroll Chaotic Circuit," *Am. J. Phys.*, vol.58,No.10, pp. 936-941, 1990.

[7] J.D. Farmer, J.J. Sidorowitch, "Exploiting Chaos to Predict the Future and Reduce Noise," *Evolution, Learning and Cognition*, ed. Lee, Y.C. (World Scientific).