Modelling and Forecasting of Glaucomatous Visual Fields Using Genetic Algorithms

Stephen Swift, Xiaohui Liu
Department of Computer Science
Birkbeck College, University of London
Malet Street, London, WC1E 7HX, UK
[swifty,hui]@dcs.bbk.ac.uk
+44 (0)171 631 6723/6711

Abstract

The prediction of visual field deterioration in patients who are suffering from normal tension glaucoma plays an important role in the management of the disease. The Vector Auto-Regressive (VAR) process appears to be an appropriate way of modelling the multivariate time series data from the visual fields. However, standard parameterisation techniques such as the Yule-Walker equations for building a VAR model place a restriction on the minimum length of time series observations. In this paper genetic algorithms are suggested as a way of finding the order and estimating the parameters for the VAR process. To evaluate the effectiveness of this approach, the VAR process in S-Plus, the Holt-Winters forecasting method, and a pure noise model are applied to the same set of visual field data.

1. Introduction

Glaucoma is the name given to a family of eye conditions [4]. The common trait of these conditions is a functional abnormality in the optic nerve, leading to loss of visual field. The prediction of visual field deterioration in patients who are suffering from glaucoma plays an important role in the management, treatment and control of the diseases progress. For example, if the deterioration is slowing down, it might be appropriate to reduce the medication; or if the deterioration is speeding up, an increase in medication might be needed or surgery might be necessary.

The Vector Auto-Regressive process [7] appears to be an appropriate way of modelling the multivariate time series data from the patient's visual fields. For the VAR process to be of use, the order must be identified and the associated parameters must be estimated, for example using the standard method of solving the appropriate set of Yule-Walker equations. However this technique places constraints on the minimum number of time series observations in the dataset.

In this paper we describe a Genetic Algorithm (GA) [6] which is used to overcome these problems by learning both the order and corresponding parameters. This method is compared with the conventional VAR method used within S-Plus [12], the Holt-Winters forecasting method [2], and a pure noise model. Each model is judged by how well it performs on the short term forecasting [1] of a dataset collected from the visual field tests of normal tension glaucoma sufferers.

Section 2 describes the nature of the visual field data and the Vector Auto-Regressive process. Section 3 describes the VARGA method, which uses a GA to parameterise and finds the order of a VAR process. Section 4 describes the evaluation method and presents the results. Finally section 5 draws conclusions from the research and suggests future work.

2. Background

2.1 Visual Field Data

The dataset is a section of Normal Tension Glaucoma visual field data [4]. Vision loss is usually only part of the visual field, however untreated glaucoma can lead to blindness. A patient's visual field can be seen initially as a circle, containing values ranging from zero representing no vision to 60 representing perfect vision. Visual field tests are performed on a clinical machine. The particular test used with this dataset examines 76 points in each eye (see Figure 1).

Current theory [3, 5] states that deterioration of the visual field can be highly correlated if two points lie on the same nerve fibre bundle. Figure 1 shows the test point locations used for the visual field dataset; the number corresponds to a mathematical mapping of the points (x,y) co-ordinate to a single value, which has been omitted. For the purpose of this paper, a selection of nine points (those shaded in grey in Figure 1) is considered, corresponding to one nerve fibre bundle (number 5 as listed in [5]). Bundle 5 has been chosen for two reasons. The first is that the nine points are the largest in number for any of the nerve fibre bundles. The second is that glaucoma damage tends to originate from the blind spot, and then move through these points to the visual periphery, thus usually affecting nerve fibre bundle 5.

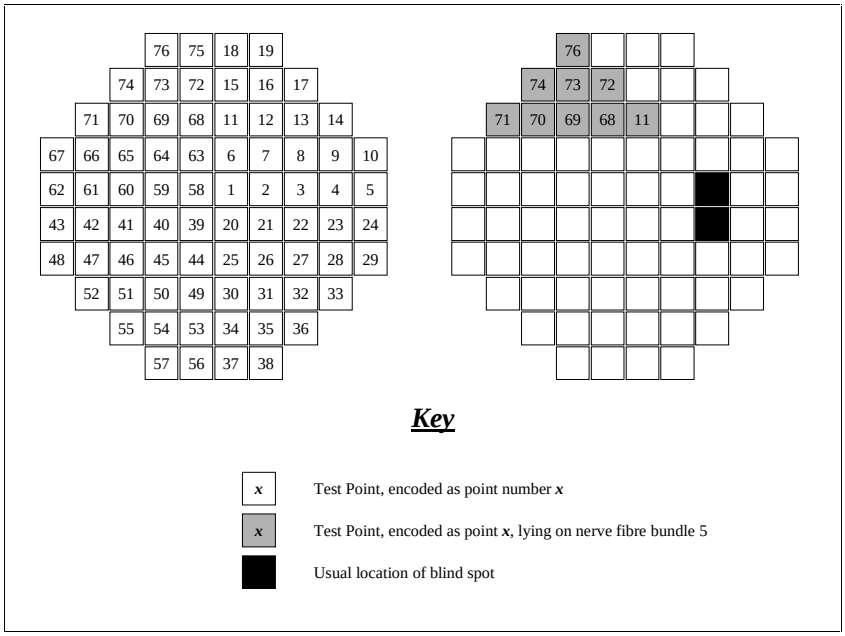


Figure 1. Visual Field Data Points for the Right Eye

To show that a multivariate model is appropriate, it is necessary to show that the variables being modelled have a strong interdependency. The correlations between all points, with a time lag of up to five time units, have been calculated by using the visual field records of the patients. A simple average was taken to give a single correlation value. Table 1 shows these for all of the points, and then for nerve fibre bundle five. The numerical triple (*x*,*y*,*lag*) represents the correlation between the variables *x* and *y* at time lag *lag*. Pearson's Correlation Coefficient [11] is commonly used in time series data. Clearly it can be seen from the table that the points within bundle 5 have a higher correlation with each other than with other points.

Points	Maximum	Minimum	Average	Variance
All	0.654 (75,76,0)	0.186 (27,26,2)	0.343	0.003
Bundle 5	0.636 (70,69,0)	0.301 (68,11,2)	0.388	0.003

Table 1. Pearson's Correlation Coefficient Comparison

For this dataset, there is no missing data. It is assumed that each test is spaced evenly in time, i.e. the time gap between subsequent tests is a constant. The data itself is a continuous variable. The dataset contains information on 82 patient's right eyes tested approximately every six months for between five and 22 years. Therefore, the length of time series corresponding to some of the patients' visual field tests can be rather short. All patients had been diagnosed and are undergoing treatment for Normal Tension Glaucoma, and were representative of the population.

2.2 The VAR Process

A VAR process of order P, written VAR(P), is defined in equation 1.

$$\underline{x}(t) = \sum_{i=1}^{p} \mathrm{A}_i \cdot \underline{x}(t-i) + \underline{\varepsilon}(t) \qquad (1)$$

Where $\underline{x}(t)$ is the next data vector of size K (the number of variables in the model), A_i is a $K \times K$ coefficient matrix at time lag i, and $\underline{\varepsilon}(t)$ is a K length noise vector at time t (usually Gaussian) with zero mean. The value of each element in A_i is usually a real number in the range ± 1. To use equation 1 for prediction purposes the parameter matrices A_i must be estimated from the data.

Two commonly used techniques for estimating the parameter matrices are the Yule-Walker and Maximum Likelihood methods. With the Yule-Walker method, there is a restriction on the minimum length of the time series; this method is used within S-Plus. With the Maximum Likelihood method, the distribution of the $\underline{x}(t)$ must be known and unfortunately this is not the case for the visual field data. The data probably does not fall into any standard distribution since the visual field values always lie between zero and 60.

3. The VARGA Method

This section describes VARGA, a GA designed to find the order and associated parameter matrices for a VAR(P) process best suited to fitting an individual patient's visual field. The level of accuracy for the GA (the *fitness* function) is defined in equations 2 and 3.

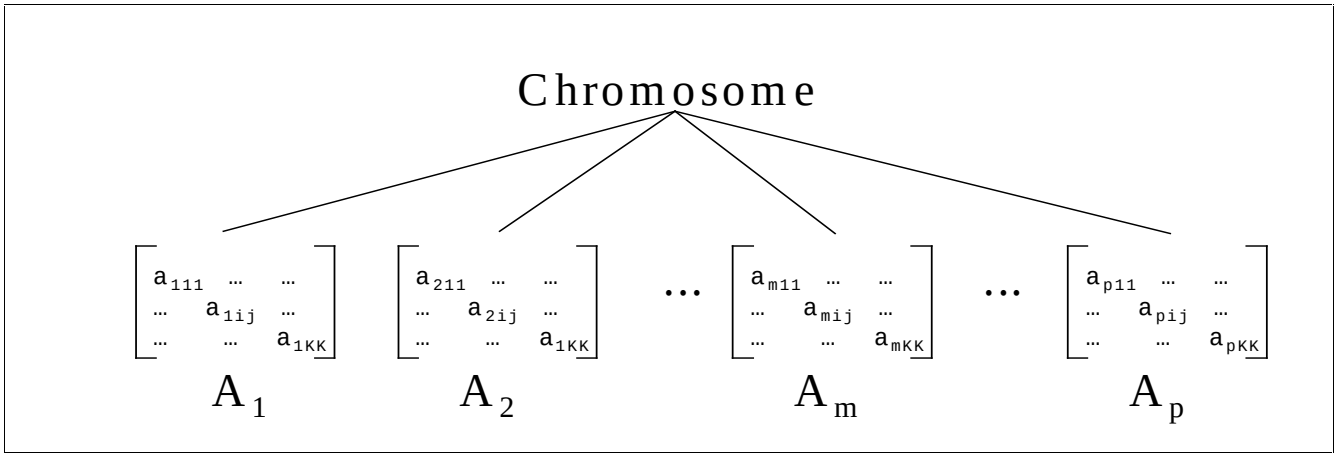


Figure 2. Chromosome Representation for VARGA method

$$\hat{\varepsilon}(t) = \underline{x}(t) - \sum_{i=1}^{p} \hat{\mathrm{A}}_i \cdot \underline{x}(t-i) \qquad (2)$$

$$\varepsilon = \sum_{j=1}^{K} \left| \hat{\varepsilon}_j(t) \right| \qquad (3)$$

Where $\hat{\varepsilon}(t)$ is the estimation of the noise vector, $\hat{\varepsilon}_j(t)$ is the *jth* element of $\hat{\varepsilon}(t)$, $\hat{\mathrm{A}}_i$ is the estimation of the *ith* parameter matrix, and ε is a scalar that represents the level of noise. All other variables are defined in section 2.2. The model with the smallest ε value is deemed the best for forecasting since it is assumed that the best estimation for any unobserved noise vector is the zero vector. For the visual field application the maximum allowable P (MAXP) for VARGA is set to eight since the smallest time series in the dataset is of length ten.

The chromosome representation is a list of ($K \times K$) matrices, whose elements are integers ranging between [0..20000). A simple scaling is done to map each value between ±1. Each matrix's order in the list corresponds to the equivalent coefficient matrix for the VAR process being represented. The visual field data is mean-adjusted before being used in this method. This chromosome representation is shown in Figure 2.

The VARGA algorithm essentially follows the standard Holland genetic algorithm [6], however, crossover is different and there are two mutation operators. VARGA is described as follows (the notation *U(x,y)* will be used to represent a random integer uniformly distributed between *x* and *y*).

1) Create *Population* random chromosomes of order U(1,MAXP)
2) Sort population ascending according to fitness (equation 3)
3) For g = 1 to *Generations* do
4) Crossover population
5) Mutate population's genes
6) Mutate population's order
7) Sort population in ascending order according to fitness (equation 3)
8) Select the new population
9) Next g
10) The best VAR process is the chromosome from the final population with the smallest fitness score (equation 3)

The **Crossover** operator is as follows ($x[a_{ijk}]$ refers to the *jth*,*kth* element of the *ith* parameter matrix of chromosome *x* (Figure 2). This is also referred to as a *Gene*).

1) Randomly select *Crossover* proportion of the population for breeding
2) Randomly pair up the breeding stock
3) For each parent pair c, d do
4) x = c, y = d
5) i = U(1,order of x)
6) j = U(1,order of y)
7) m = U(1,K)
8) n = U(1,K)
9) $x[a_{irs}]=d[a_{jrs}]$ $\forall$r,s where (r≤m or s≤n)
10) $y[a_{jrs}]=c[a_{irs}]$ $\forall$r,s where (r≤m or s≤n)
11) Add x, y back to the population
12) Continue

Gene Mutation is as follows:

1) Each gene of every chromosome has a *Gene Mutation* chance of mutating
2) For each gene that mutates do
3) v = [(gene value + U(1,20000)) modulo 20000]
4) Gene value = v
5) Continue

Order Mutation is as follows:

1) Each chromosome has an *Order Mutation* chance of mutating
2) For each chromosome that mutates do
3) w = U(1,2)
4) if w = 1 then delete the last coefficient matrix
5) if w = 2 add a new random coefficient matrix to the end of the list
6) Continue

Selection of the new population is exactly the same as the Roulette Wheel method [8], however the reciprocal of the fitness score (equation 3) of each chromosome is used. This is because the score represents the residual noise; the lower the score, the better the model being represented.

For the VARGA method, the parameters for the GA are listed in Table 2.

Feature	Value	Comment
Population	10	Constant
Generations	5000	Crossover will not be so effective since only a portion of each chromosome is crossed over
Selection	Roulette Wheel	The best is always carried forward
Order Mutation Rate	5%	If the order = 1 then add a coefficient matrix, if the order = MAXP then delete
Gene Mutation Rate	0.5%	After crossover, including the parents; the population best is not mutated
Crossover Rate	100%	Percentage of population allowed to breed, uniform and one point
Chromosome Size	Order×81	Order×K×K
Gene Type	0 to 19999 Integer	
Fitness	Real	Positive, nearest to zero the better. Defined in equation 3

Table 2. Genetic Algorithm Parameters for VARGA

4. Evaluation

In this section the models found using the VARGA method are compared with those produced by the conventional way of finding a VAR process, i.e. the solution of the Yule-Walker equations using S-Plus. To provide more insight into the accuracy of the VARGA method, it is further compared with the results from two other techniques: Holt-Winters forecasting and the noise model.

4.1 VAR in S-Plus

S-Plus has an easy-to-use function for finding the best-fit VAR(P) process for a given dataset. Each patient's visual field results give a model that is rated according to equation 3. Since S-Plus uses *"Whittles Recursion"* [13], a limit on the minimum length *N* of a time series with *K* variables is constrained by inequality 4.

$$N \geq K(P+1) \qquad (4)$$

4.2 Holt-Winters Forecasting Method

Despite the fact that the dataset is multivariate, it is worth treating it as univariate to see if the assumptions about point clustering (by nerve fibre bundles) are accurate. The Holt-Winters (HW) forecasting method [2] is a simple way of predicting the next value in a univariate time-series. For the visual field dataset, it is assumed that there is no seasonal effect, and that we are only interested in one step ahead forecasting. The HW Method is defined as follows:

$$L_t = \alpha X_t + (1-\alpha)(L_{t-1} + T_{t-1}) \qquad (5)$$

$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1} \qquad (6)$$

$$\hat{X}_t = L_t + T_t \qquad (7)$$

Equation 5 defines the mean level at time *t*, equation 6 defines the trend at time *t*, and equation 7 defines the forecast of the dependant variable at time *t*, given values of *X* from *0..t-1*. All the variables are scalars. The method needs the values of α, γ, L_0 and T_0 to be defined (the *starting* values), so that the subsequent values of $L_t, T_t \& \hat{X}_t$ can be calculated. According to [2] α and γ lie between zero and one. However with L_0 and T_0, they are usually determined as a function of the observed values of *X*, and often are limited by the maximum and minimum values. To allow for some variation from this rule, they are assumed to lie within ±100 for this application. There are various ways of finding the values of these parameters as suggested in [2], but for simplicity sake, a genetic algorithm will be used to estimate them. A standard Holland GA was used, with the modifications and parameter values listed in Table 3. The *fitness* for the HW method is rated in a similar way to a VAR process, but the residuals from a one step ahead forecast are summated for each point. A visual field case is treated as nine univariate forecasts. Each chromosome consists of twelve genes which are integers ranging between zero and 99. Three genes represent each parameter. The value of the three genes is then scaled accordingly, i.e. for parameters α, γ, between zero and one, and for L_0, T_0 between ±100.

Feature	Value	Comment
Population	50	Constant
Generations	1000	Crossover will not be so effective since the Chromosome size is quite small
Selection	Roulette Wheel	See VARGA below, the best is always carried forward
Mutation Rate	0.5%	After Crossover, including parents; if a gene mutates, add U(0,99) then modulo by 100, this is detailed within the section on VARGA
Crossover Rate	80%	Percentage of Population allowed to breed, uniform and one point
Chromosome Size	12	
Gene Type	0 to 99 Integer	
Fitness	-ve, Real	Nearest to zero the better, calculated as above

Table 3. Genetic Algorithm Parameters for the HW method

4.3 The Noise Model

The noise model is defined in equation 8. Note that equation 8 is equivalent to equation 1 with P=0, i.e. a VAR(0) process. A forecast for the noise model is defined in equation 9. Any method that provides a forecast worse than the noise method is a very poor method.

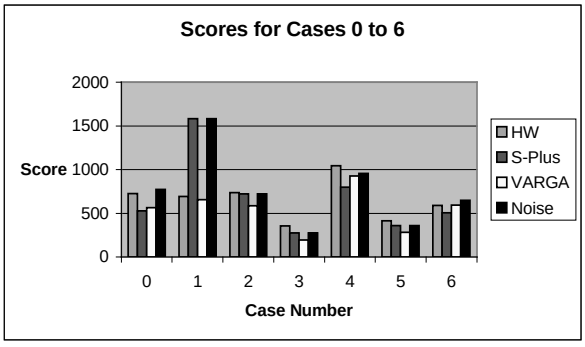
$$\underline{x}(t) = \underline{\varepsilon}(t) \tag{8}$$

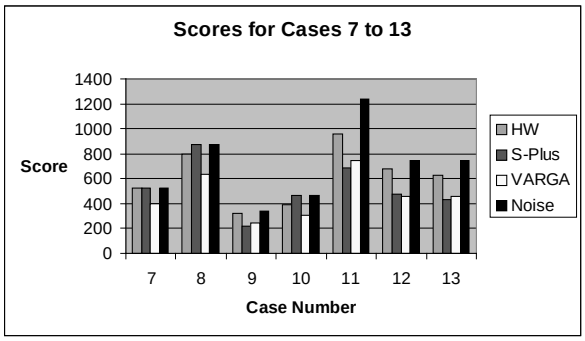
$$\underline{x}(t) = \underline{0} \tag{9}$$

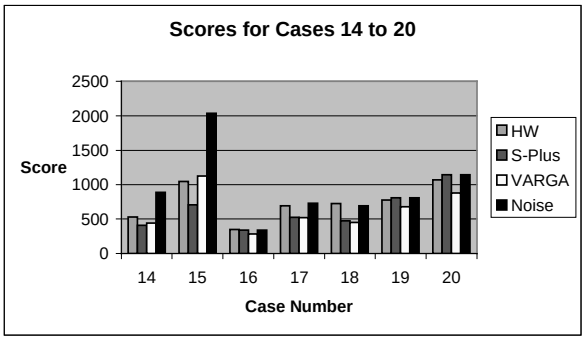
4.4 Experimental Results

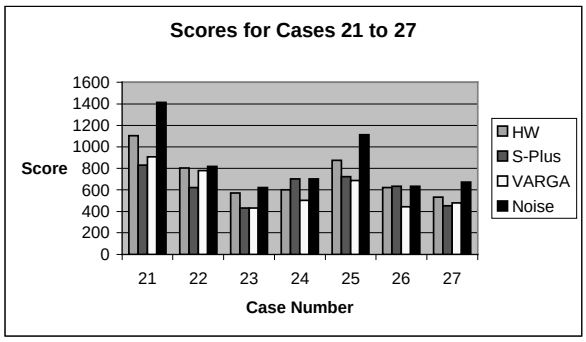
This section describes the results of the experiments. The restriction described in equation 4 meant that the dataset for the experiments had to be reduced significantly. If an order (P) of at least one is under consideration and since there are nine variables (K), the time series length (N) must be at least 18. This restriction reduced the dataset from 82 patients to 28.

VARGA was run once on each of the patients, and Figures 3 to 6 display the results for four methods over the 28 patients' visual field tests.









Figures 3-6. Test Results

Method	Order (*number* of *order*)	Average Score
VARGA	26 of 1, 2 of 2	559.82
S-Plus	12 of 0, 14 of 1, 1 of 2, 1 of 3	616.12
HW	N/A	683.79
Noise	28 of 0	816.53

Table 4. Results Summary

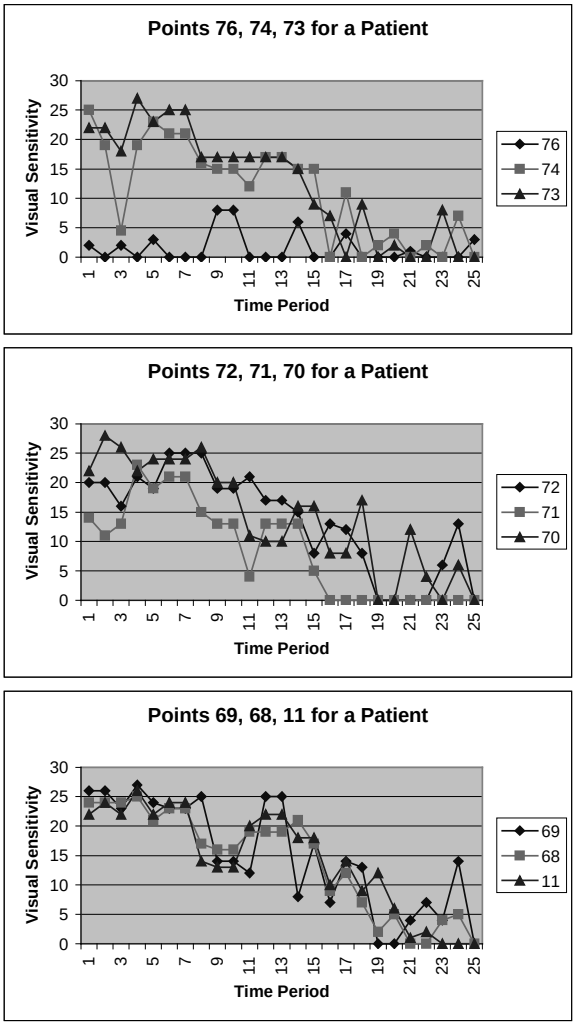
Table 4 summaries these results, and lists the order of the best models. From this table the following can be observed. Firstly, VARGA has the best performance, followed by S-Plus, Holt-Winters, and the noise process. Secondly, VARGA has a more consistent set of results for the order than the VAR process in S-Plus. Thirdly, VARGA located models of only order one and two whilst the order ranged between zero and three with S-Plus.

It could be argued that the VARGA method is biased towards finding a low order model, since the search space increases each time the order of the model increases. However tests have been run where the VARGA method is forced to search for a VAR(3) process (thus behaving like a conventional GA) where the number of generations is increased to compensate for the number of variables being found. The results here still showed that a VAR(1) or VAR(2) still fit the data better. From this finding, further work could be done to more accurately fit a VAR(1) or VAR(2) process, reducing the need to search for the model order.

It is curious that S-Plus was unable to find a VAR(1) process in the cases where it found instead a VAR(0) process. These can be seen in Figures 3-6 where the score of S-Plus is the same as the noise model, e.g. case 1 and case 20. If these are ignored, S-Plus marginally becomes the best.

With the HW method, the results show that a multivariate method is more accurate than a series of univariate models for the visual field dataset. To give this method more credence, a full search for the parameters (this takes a very long time) has shown that the method locates them very accurately (within 1% of their actual values to four decimal places). The poor performance is a result of the data being multivariate.

It is worth noting that for the patient case where the difference between the errors for the VAR process found by VARGA and the noise process is the greatest (case 1), the error (one step forecast errors) is reduced by only 58.5%. This could indicate that the visual field data has a very large noise term. This is confounded when the actual data is viewed. The visual field history for patient case 1 is given in Figures 7 to 9.



Figures 7-9. A Patient's Visual Field

As would be expected, visual field sensitivity deteriorates down to zero. However the graph seems to show that some of the points get better, which is impossible with the disease glaucoma. This can be explained by the fact that each visual field point corresponds to an area on the retina, and that the exact retina cells being tested for damage are not always the same when the point undergoes another test. This is due to there being a limit on the accuracy of the test machine. Hence the data seems to contain a large element of noise. These could be treated as outliers, but how these could be dealt with is a difficult problem, especially with time series data.

5. Concluding Remarks

We have presented a method for learning a Vector Auto-Regressive process from a given set of multivariate time series. This is achieved through a new representation and associated crossover and mutation operators for a genetic algorithm. The results clearly show that the VARGA model provides a better method for fitting a VAR process than the conventional statistical methods.

As demonstrated with the visual field data, the VARGA model can be applied to multivariate time series datasets where there are a small number of observations. This gives the method a wider range of applications than the standard statistical methods (e.g. than with the Yule-Walker equations). Although there have been other GA-based methods to find the order and parameters for time series models [9, 10], these methods are applied to *univariate* time series data. VARGA has been found to be a promising method for modelling *multivariate* time series data.

Future work will include the following:

1) The relationships between variables can be generalised into groups through clustering. A GA can be used for this, hence a competitive co-evolution strategy could be implemented involving VARGA to find a selection of VAR models for the whole visual field.

2) The investigation of other multivariate time series models (e.g. exponential) since the visual field data points seem to settle down to zero at a non-linear rate.

3) Since the proportion of noise is quite large, methods for placing confidence limits on any forecast will be investigated.

6. Acknowledgements

The authors wish to thank the project Sponsors; the Moorfields Eye Hospital in London and the Engineering and Physical Sciences Research Council, UK. We would also like to thank Allan Tucker for his help and advice.

References

[1] C. Chatfield, "What is the 'best' method of forecasting?", Journal of Applied Statistics, Vol. 15, No. 1, 1988, pp. 19-38

[2] C. Chatfield, M. Yar, "Holt-Winters forecasting: some practical issues", The Statistician, Vol. 37, 1988, pp. 129-140

[3] D. Crabb, F. Fitzke, A. McNaught, R. Hitchings, "A Profile of the Spatial Dependence of Pointwise Sensitivity Across The Glaucomatous Visual Field", Perimetry Update, 1996/1997, pp. 301-310.

[4] M. J. Haley (editor), "The Field Analyzer Primer", Allergan Humphrey, 1987.

[5] A. Heijl, A. Lindgren, G. Lindgren, "Inter-Point Correlations of Deviations of Threshold Values in Normal and Glaucomatous Visual Fields", Perimetry Update, 1988/89, pp. 177-183.

[6] J. H. Holland, Adaptation in Natural and Artificial Systems, Ann Arbor, MI: The University of Michigan Press, 1975.

[7] H. Lütkepohl, "Introduction to Multivariate Time Series Analysis", Springer-Verlag, 2nd edition, 1993

[8] Z. Michalewicz, "Genetic Algorithms + Data Structures = Evolution Programs", Springer, 3rd edition, 1996

[9] S. Rolf, J. Sprave, W. Urfer, "Model Identification and Parameter Estimation of ARMA Models by means of Evolutionary Algorithms", University of Dortmund, 1997

[10] Z. Shi, H. Aoyama, "Estimation of the Exponential Auto-Regressive Time Series Model by using the Genetic Algorithm", Journal of Sound and Vibration, Vol. 205, 1997, pp. 309-321

[11] G. Snedecor, W. Cochran, "Statistical Methods", Iowa State University Press, 6th edition, 1967.

[12] W. N. Venables, B. D. Ripley, "Modern Applied Statistics with S-Plus", Springer-Verlag, 1996

[13] P. Whittle, "Prediction and Regulation", Basil Blackwell, 2nd edition, 1984