The Pseudotemporal Bootstrap for Predicting Glaucoma From Cross-Sectional Visual Field Data

Allan Tucker and David Garway-Heath

Abstract—Progressive loss of the field of vision is characteristic of a number of eye diseases such as glaucoma, a leading cause of irreversible blindness in the world. Recently, there has been an explosion in the amount of data being stored on patients who suffer from visual deterioration, including visual field (VF) test, retinal image, and frequent intraocular pressure measurements. Like the progression of many biological and medical processes, VF progression is inherently temporal in nature. However, many datasets associated with the study of such processes are often cross sectional and the time dimension is not measured due to the expensive nature of such studies. In this paper, we address this issue by developing a method to build artificial time series, which we call *pseudo time series* from cross-sectional data. This involves building trajectories through all of the data that can then, in turn, be used to build temporal models for forecasting (which would otherwise be impossible without longitudinal data). Glaucoma, like many diseases, is a family of conditions and it is, therefore, likely that there will be a number of key trajectories that are important in understanding the disease. In order to deal with such situations, we extend the idea of pseudo time series by using resampling techniques to build multiple sequences prior to model building. This approach naturally handles outliers and multiple possible disease trajectories. We demonstrate some key properties of our approach on synthetic data and present very promising results on VF data for predicting glaucoma.

Index Terms—Bootstrapping, cross section, data analysis, glaucoma, time series.

I. INTRODUCTION

PROGRESSIVE loss of the field of vision is characteristic of a number of eye diseases such as glaucoma, a leading cause of irreversible blindness in the world. Recently, there has been an explosion in the amount of data being stored on patients who suffer from visual deterioration, including visual field (VF) tests, retinal images [1], and perhaps, in the not too distant future, even continuous intraocular pressure measurements [2]. The aim now is to extract as much information as possible from these data in order to address fundamental questions still open

within the glaucoma research community. The VF test assesses the sensitivity of the retina to light. It is typically measured by automated perimetry, a technique in which the subject views a dim background as brighter spots of light are shone onto the background at various locations in a regular grid pattern. The brightness at which the subject sees the spots of light is related to the retinal sensitivity. There are many diseases and conditions that affect the VF, the most common being the optic neuropathy "glaucoma," neurological, and retinal diseases. Early detection of glaucoma as well as other conditions and diseases that cause visual impairment is invaluable as early intervention can slow VF deterioration.

Like the progression of many biological and medical processes, VF progression is inherently temporal in nature. While many studies involve learning computational and statistical time-series models of progression from longitudinal data, such as [3] and [4], many datasets associated with the study of such processes are often cross sectional and the time dimension is not measured due to the expensive nature of such studies across a large population and the slowly progressive nature of the condition requiring many years of data collection.

In this paper, we address this issue by developing a method to build artificial time series, which we call *pseudo time series* from cross-sectional data. This involves building trajectories through all of the data that can then, in turn, be used to build temporal models. Results demonstrate how these models can be used to perform temporal analyses, such as classification and forecasting, that would otherwise be impossible without longitudinal data. Glaucoma, like many diseases, is a family of conditions, and therefore, it is likely that there will be a number of key trajectories that are important in understanding the disease. In order to deal with such situations, we extend the idea of pseudo time series by using resampling techniques to build multiple sequences prior to model building. This approach naturally deals with outliers and multiple disease trajectories. In medical terms, the use of pseudo time series for modeling multiple trajectories will assist clinicians in forecasting future values of variables such as VF sensitivity and to determine on which trajectory a patient is likely to be traveling in order to make the appropriate intervention.

The paper is organized as follows. Section II introduces pseudo time series for cross-sectional data as well as the temporal bootstrap that deals with outliers and multiple trajectories. In Section III, our methods are applied to simulated data in order to explore some of the characteristics of the associated models. We then explore learning these models from a cross-sectional VF dataset that are tested on an independent glaucoma longitudinal study. Finally, we present our conclusions in Section IV.
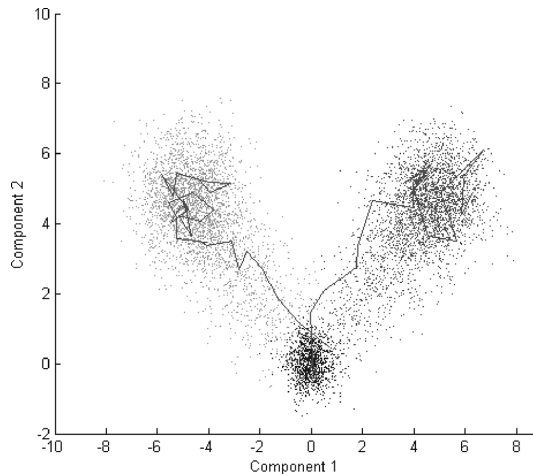
Fig. 1.   Full MTS dataset from $N = 250$ with two sample MTS lines plotted using the first two components after multidimensional scaling.

II. METHODS

In this section, we describe the two datasets that are used to explore pseudo time series. We also describe two methods for building pseudotemporal models: the pseudotemporal model, which builds an ordering across all cross-sectional data, and the pseudotemporal bootstrap, which builds a number of sequences over the data and can handle outliers and multiple disease trajectories.

A. Synthetic Data

Multiple time-series data are generated using an autoregressive hidden Markov model (ARHMM) with three discrete states in order to determine the trajectory of each time series. An ARHMM was used as it manages to capture the relatively smooth transition from healthy to disease states through autoregressive dependencies. This smooth transition is typical in VF data (and indeed many other medical data such as cancer). The three states represent a starting healthy region and two diseased regions. The observed variables are Gaussian and conditioned upon the hidden state and the same variable at the previous time point. We set the length of each multivariate time series (MTS) T to be 30, as this reflects common longitudinal studies in the biological and medical literature. The hidden variable always starts in the healthy state and has a probability distribution that determines the probability of change to one of the two disease states. This probability is set to 0.1 and the probability of remaining in the healthy state is 0.8. Once a disease state is reached, it cannot change (as the probability of remaining in this state is 1.0). These parameters ensure realistic progression of states, as seen in VF data, that has been labeled according to an established field test classification algorithm [5].

Throughout the paper, we use multidimensional scaling to plot data. As a result, axes represent the first two components of this scaling and allow us to visualize higher dimensions. The ARHMM data are plotted in Fig. 1 using multidimensional scaling given the Euclidean distance between data. Also plotted are some sample time-series trajectories that originate in

TABLE I
BREAKDOWN OF THE DATASETS

Dataset	CS Data	Longitudinal Data
# VF variables	54	54
# NFB vars	6	6
# VF tests	162	587
# Patients	162	23
Pos:Neg class ratio	78:84	229:358
Mean(SD) Series Length	NA	25.1(6.8)

the healthy region and then travel to one of the two diseased states. We sample a single point from each time series in order to simulate a cross-sectional dataset. We then use these data to explore how well we can recreate the original time-series model using our pseudo-time-series approach on the sampled cross-sectional data only. This allows us to perform classification and forecasting using the original full time-series data as a test set. These datasets along with the code that generates them (in MATLAB in conjunction with the Bayes Net Toolbox) are available at *http://people.brunel.ac.uk/~cssrajt/ARHMMdat*.

We explore the effect of cross-section sample size on model prediction accuracy and explore the smoothing property of the temporal bootstrap.

B. VF Datasets

We also use two VF datasets to test our approach. One dataset involves a large cross-sectional study of VF tests on 162 people, representing an expanded dataset that was used to evaluate the classification accuracy of an optic nerve head imaging device [6]. In brief, there were 84 healthy subjects and 78 patients with early glaucomatous VF loss. A full medical history was taken and detailed ocular examination performed. Subjects underwent Humphrey VF testing with the 24-2 program. The VF data for each subject are classified into one of two classes: healthy or glaucomatous based upon an established classification algorithm for the field test [5]. We also have data from a longitudinal study of 23 ocular hypertensive patients (who eventually develop reproducible glaucomatous VF loss) from a longitudinal study at Moorfields Eye Hospital. A total of 255 patients with ocular hypertension (raised intraocular pressure, a major risk factor for glaucoma) volunteered to take part in a randomized placebo-controlled trial of treatment to prevent the onset of glaucoma [7]. Of these, a number developed reproducible VFs loss, as judged by the same classification algorithm, over a median period of six years. Subjects had several repeated clinical visits (approximately every six months). We use this as an unseen dataset to evaluate the learned pseudotemporal models. In other words, we ask "can we use the cross-sectional data to build models to forecast and classify future states of the genuine temporal data?". Table I summarizes the two VF datasets. For both datasets, we calculate the average sensitivity of groups of VF points based upon their relationship to *nerve fibre bundle* (NFB) that carries the information from the retina to the visual cortex, via the optic nerve head. There is a well-established relationship between VF test locations on the retina and the NFB distribution [1]. Fig. 2 shows how the NFBs are distributed over the VF points where "X" denotes the blindspot—these measurements are ignored (taken from [8]). As a result, both datasets contain six NFB variables and one class variable.
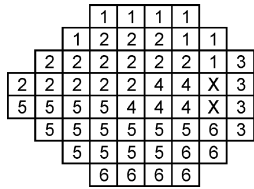
Fig. 2.   Allocation of VF points to NFBs.

C. Preliminaries

In order to explain the algorithms, we first set out some notation. Let N be the number of variables in a dataset, D_T be a time-series dataset containing T measurements, D_T^i represent the ith time point in D_T, D_{CS} be a cross sectional dataset containing L cases, and O be some ordering of the cases in D_{CS}. We define a pseudo time series to be the pair $\langle D_{CS}, O \rangle$ such that the cross-sectional data are placed into an ordering that can be treated as a temporal or developmental series prior to model building. In addition, let a dataset be assigned two or more classes in C where one is deemed to be the starting class, C_1 or the "healthy" class and the others as the "diseased" classes, $C_i, i > 1$. Using the pseudotemporal ordering, O, of the data samples in D_{CS}, we can build a time-series model of the dataset. The idea is that the inclusion of this ordering should both improve model performance and enable the ability to forecast future states. The hidden Markov model [9] will be used to model the pseudo time series from both approaches. Parameters are learned from the pseudo time series and used to forecast future class states or variable values.

We now describe two algorithms for building pseudo-time-series models, one that involves using PQ trees to efficiently build sequences via partial orderings, and the other that uses resampling techniques to build numerous pseudo time series for data that have multiple possible disease states and that can deal with outliers.

D. Pseudo Multivariate Time-Series Model Construction

The first method for generating a pseudo multivariate time-series model (PMTS) is based on fitting a partial path through the entire cross-sectional dataset D_{CS}. This path is designed to minimize some distance metric between data and is found using a PQ-tree approach, as outlined in [10] in combination with a simple hill-climbing procedure. Here, we use Euclidean distance.

The PQ-tree approach is used simply to add efficiency to the search as the possible number of orderings can be very large. PQ-trees are a graph-structure device that can represent a partial ordering of points, and indicate which parts of the ordering are well supported (Q nodes) and which parts contain more uncertainty (P nodes). While the children of a P node can be put into any order, children of a Q node may be reversed in order but may not otherwise be reordered. To construct an ordering from data, first, a distance matrix is generated between all the variables and this is used to build a minimum spanning tree. The diameter path of the tree is used as the main Q node of the PQ tree—the backbone of the reconstructed ordering.

Algorithm 1 Pseudo temporal model construction

1: Input: Cross Sectional Data D_{CS}, $startpoint$, $endpoint$
2: Standardise the data to be zero-centred with a standard deviation of 1.0
3: Construct weighted graph and minimum spanning tree of D_{CS}
4: Construct PQ tree from minimum spanning tree using the algorithm in [11]
5: Derive O from PQ-tree using a hill-climb search on P-nodes to minimise sequence length whilst fixing the $startpoint$ and $endpoint$
6: Construct pseudo time-series using O and D_{CS}
7: Train temporal model on pseudo time-series
8: Output: PMTS Temporal Model

Branches of the diameter path are added as P and Q nodes to the main Q node. Therefore, the constructed PQ tree represents a partial ordering of the data samples. The full algorithm can be found in [10].

We convert the partial ordering represented by the PQ tree into a full ordering, O using a hill-climb method to further minimize the distance within the PQ-tree constraints. We constrain this search to find paths with a fixed, user-defined, start- and endpoint representing the most extreme healthy and diseased cases in the cross-sectional data.

Once a PMTS is constructed from the full ordering O, a model-fitting stage is carried out to build a temporal model for forecasting. The whole process for creating a pseudotemporal model using this approach is described in Algorithm 1.

The parameters that affect the quality of the PMTS model in Algorithm 1 are the size L and the number of disease states C in the dataset D_{CS}. Fitting one trajectory through a dataset with many disease states will result in an unrealistic trajectory and, therefore, poor temporal models. Also, if the $startpoint$ and $endpoint$ are badly chosen then this too will result in poor trajectories. In many datasets, these points are relatively easy to select using multidimensional scaling methods.

E. Pseudotemporal Bootstrap Model Construction (TBS)

We now introduce another way of generating pseudo time series that can naturally deal with multiple trajectories and outliers in the data. It is inspired by Efron's bootstrap [11], and hence the name, and attempts to deal with the natural variability between the trajectories of disease within a population by repeatedly resampling from D_{CS} and fitting shortest paths between the "healthy" C_1 and "diseased" classes $C_i, i > 1$. Each of these paths is used as an ordering O to generate a different pseudo time series, which a time-series model can then be trained on. The method used here for finding the shortest path is the Floyd–Warshall algorithm [12]. The TBS algorithm is described in Algorithm 2.

Building temporal models using the TBS approach in Algorithm 2 avoids the issues with selecting good startpoints and endpoints that is inherent with the PMTS in Algorithm 1. As long as the data have suitably labeled classes C representing diseased and healthy cases, the algorithm should build realistic trajectories. However, as we show in the next section, the size

---

Algorithm 2 Pseudo temporal bootstrap model construction

1: Input: Cross Sectional Data D_{CS}, Class Labels C
2: Standardise the data to be zero-centred with a standard deviation of 1.0
3: Construct weighted graph
4: **for** all samples in m **do**
5:      Randomly sample L data and their allocated classes from D_{CS}
6:      Randomly select a $startpoint$ in C_1 and an $endpoint$ in $C_i > 1$
7:      Retrieve the shortest path between $startpoint$ and $endpoint$ from the sample subset of the weighted graph
8:      Randomly assign a $startpoint$ in C_1 and $endpoint$ in $C_{i>1}$ and generate a time-series containing the elements in the shortest path between $startpoint$ and $endpoint$
9: **end for**
10: Train the temporal model from the m sampled time-series
11: Output: TBS Temporal Model

---

of the $D_C S$ sample, L, has a dramatic effect, while the variance within the data has less impact.

III. EXPERIMENTS AND RESULTS

The experiments involve looking at two datasets: first, a synthetic one that involves two disease states. This is in order to explore the properties of the proposed TBS algorithm in dealing with multiple trajectories. Second, real-world cross-sectional data and longitudinal data from VF tests of both healthy and glaucomatous patients. These are explored using both the PMTS and the TBS algorithms to see which is the most appropriate and whether there are multiple trajectories that can reveal characteristics of glaucoma.

A. Synthetic Data

First, we look at the synthetic data generated from the ARHMM. Fig. 3(a) shows some sample trajectories learned using the TBS algorithm and Fig. 3(b) shows the trajectory learned using the PMTS algorithm. It is clear that PMTS is not capable of learning multiple trajectories, so for the remainder of this section on synthetic data, we focus on the TBS model, which appears to build realistic looking trajectories into both diseased regions. We generate the data as described earlier and build an ARHMM model from the cross-sectional data using the TBS algorithm. The model is then used to make a one-step-ahead forecast of the original MTS data as well as one-step-ahead classification of the next state. The whole process is then repeated 100 times generating different MTS and cross-sectional data. These results are used to calculate accuracies and confidence limits for different sample sizes L.

Table II (top) shows the outcome of forecasting the unseen MTS data using the models learned using the TBS algorithm for different sample sizes from 50 to 500 by measuring the sum squared error (SSE) of forecasts for the different models. It also includes the results of learning a model from the original MTS data for comparison. Fig. 4 shows the mean values graphically for clarity. It appears that for smaller samples of cross-sectional data ($L = 100$ or less), the quality of forecasts for the model are substantially lower than the ideal model learned from the
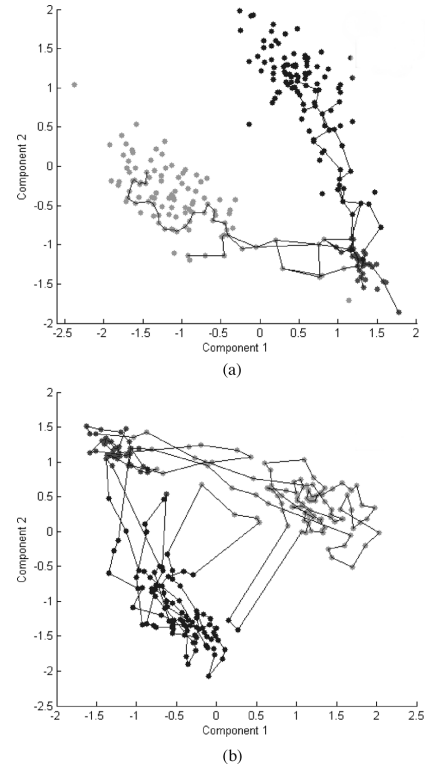

Fig. 3. (a) Sampled cross-sectional dataset from the original MTS with example TBS plotted. (b) PMTS on the sampled cross-section data.

TABLE II
(TOP) MEAN FORECAST SSE AND 95% CONFIDENCE FOR ARHMM DATA, AND (BOTTOM) MEAN CLASSIFICATION FORECAST ACCURACY AND 95% CONFIDENCE FOR ARHMM DATA

L	Full MTS	TBS	MTS smoothed
50	0.129 ± 0.039	0.251 ± 0.228	0.095 ± 0.055
100	0.125 ± 0.023	0.158 ± 0.121	0.086 ± 0.012
250	0.126 ± 0.013	0.079 ± 0.034	0.083 ± 0.012
500	0.125 ± 0.015	0.067 ± 0.023	0.084 ± 0.014

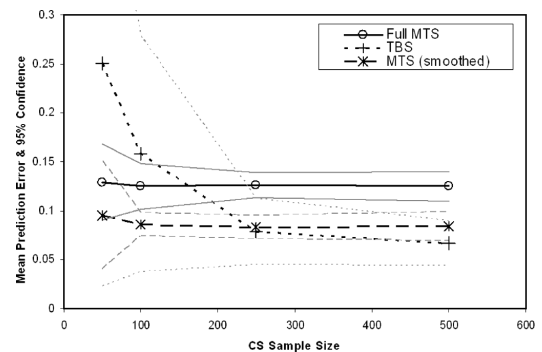| L	Full MTS	TBS	MTS smoothed
50	0.907 ± 0.047	0.897 ± 0.092	0.903 ± 0.055
100	0.905 ± 0.045	0.912 ± 0.044	0.905 ± 0.048
250	0.905 ± 0.046	0.910 ± 0.046	0.905 ± 0.048
500	0.905 ± 0.046	0.912 ± 0.044	0.904 ± 0.049

Fig. 4. Mean SSE with 95% confidence limits (in gray) on HMM learned from original MTS (both smoothed and unsmoothed) and TBS from cross-sectional data for various samples size L.

full MTS data. However, as the sample size increases the error rates decrease dramatically. In fact, for sample sizes of 250 or greater, the TBS model appears to outperform the original data.

TABLE III
HMM LEARNED FROM FULL MTS (LEFT) AND FROM TBS (RIGHT) FOR
CLASSIFICATION OF THE ARHMM DATA—CONFUSION MATRIX FOR
ONE-STEP-AHEAD CLASSIFICATION (THREE CLASSES)

Obs	1	2	3	Tot	1	2	3	Tot
Pred	-	-	-	-	-	-	-	-
1	1082	187	204	1473	1068	161	183	1412
2	0	2647	0	2647	1	2687	2	2690
3	0	33	3097	3130	13	19	3116	3148
Tot	1082	2867	3301	7250	1082	2867	3301	7250

This unexpected result is potentially due to the way the TBS algorithm selects trajectories with small distance between each datum. As a result, the TBS models are generally "smoother" than the original MTS because outliers are included with less frequency than other data. We test this hypothesis by applying a standard convolution smoothing algorithm to the original MTS data and then running the forecast experiments on this. These results are also documented in Table II (top) and Fig. 4, and show that the accuracies of the model learning using the TBS algorithm do indeed converge much closer to the smoothed data model as sample size increases. Looking at the confidence limits of the different models, it seems that the TBS also reduces the variation as sample size increases. However, the confidence intervals are always wider than the smoothed MTS models. We found the smoothing effect of TBS to be more pronounced if we increased the variance in the original ARHMM-generating process so that the data appeared noisier, i.e., the TBS improvement over the full MTS was even more pronounced but both had higher errors in general.

We next explore how well the TBS model performs at classifying the next state in the MTS data. Classification is performed by treating each consecutive pair of time points as inputs into the system with no class information. The model must then forecast the class at the second time point. Table II (bottom) shows the mean classification accuracy and 95% confidence limits. It is clear that the sample size has much less effect on the accuracy than on the forecasts, with all methods performing well for samples sizes of 50 or more. However, the confidence plots reveal that the TBS is far more variable when the sample size L is 50, while at $L >= 100$, both the mean values and the confidence limits converge.

Looking at the classification accuracy in more detail, Table III (left) shows the confusion matrix of the three classified states using the model learned from the full MTS data and Table III (right) shows the confusion matrix of the TBS model, both for $L = 250$. It is interesting to note that while the overall accuracy is approximately the same, the TBS model does less well on predicting healthy states (here class 1), whereas the errors for the full MTS are only found to be between the two diseased states (classes 2 and 3). Nevertheless, the TBS model still performs extremely well in general at forecasting future class values across all classes.

To summarize, our experiments indicate that if there is a suitably large enough sample of cross-sectional data, then the true underlying MTS model can indeed be learned with a high degree of accuracy. This applies to both the classification and forecasting tasks, though the sample size was more influential on forecast accuracy. In addition, the TBS algorithm performed
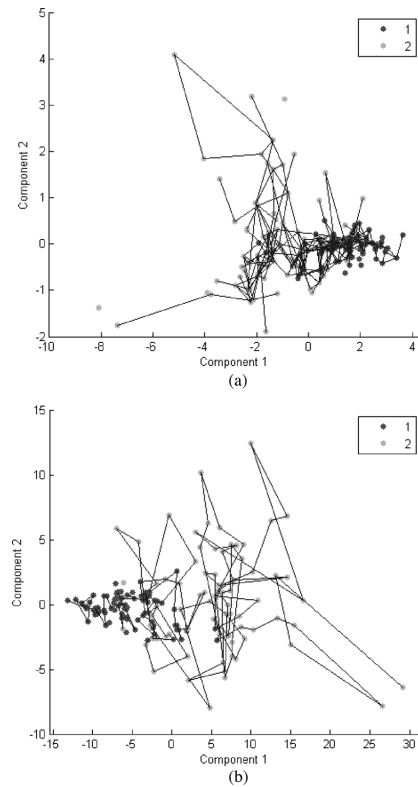

Fig. 5. VF cross-sectional data with (a) TBS and (b) PMTS (1 represents healthy patients and 2 represents glaucomatous according to AGIS score [5]).

well while simultaneously filtering the data, making it particularly useful in dealing with outliers. We now explore how both the PMTS and the TBS algorithm perform on real-world VF data.

B. VF Data

The cross-sectional data with sample TBS trajectories are plotted in Fig. 5(a) and the PMTS trajectory on the cross-sectional data is plotted in Fig. 5(b). It appears that there is a general progression from healthy to diseased that the PMTS trajectory captures [from left to right in Fig. 5(b)]. However, the TBS trajectories also seem to capture two distinct regions of disease state (in the top left and bottom left of Fig. 5(a)—note that the multidimensional scaling has resulted in progression moving from right to left in this figure). We now compare the ability of the two algorithms to forecast and classify future states, and see if TBS does indeed capture distinct diseased regions.

Fig. 6 shows box–whisker plots of the SSE for one-step-ahead forecast using the PMTS model, the TBS model and the model learned from the longitudinal data. Here, one step is based upon the longitudinal VF data that were collected approximately every six months. We also include the SSE scored using the previous NFB value as the forecast state ("PrevDatum") in order to provide a reference—this is used as clinicians generally consider there to be slow progression between tests. The median SSE over all patients is 0.284 for the full MTS data, 0.326 for the TBS model, 0.525 for the PMTS model, and 0.393 for the PrevDatum approach. It appears that the TBS model captures
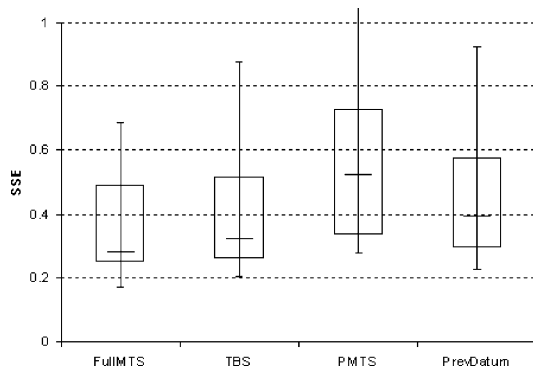
Fig. 6. Box-whisker plots for SSE of each forecast over all NFBs of the 23 patients using models learned from the original MTS, the TBS and PMTS on the cross-sectional data, and a straw-man approach based upon the previous datum.

Fig. 7. Sample data from the VF cross-sectional data based at the extreme end of two trajectories. NFBs are as described in Fig. 2.

some of the dynamics of the disease process in that it improves the accuracy of forecasts of the longitudinal data compared with the baseline PrevDatum (the PMTS performs relatively poorly). The TBS model seems to perform significantly better than both PrevDatum and PMTS on a number of individual patients, most likely due to the TBS capturing the two distinct trajectories.

The SSE for one-step-ahead forecasts for each NFB varied between 0.319 (NFB4) and 0.811 (NFB3) with a mean of 0.551. This shows that some NFBs are more easily predicted than others, possibly due to varying noise and VF sampling density. Scatter plots (not shown) seem to imply that the accuracy is better for higher values of sensitivity that makes sense because, as sensitivity decreases, there is known to be increased noise within the VF.

As with the ARHMM data, classification is performed by treating each consecutive pair of time points as inputs into the system with no class information. The model aims to forecast the class at the second time point. It appears that for classification, the PMTS model performed as well as the TBS model (the area under receiver operating characteristic (ROC) curve is 0.817 and 0.813, respectively), both of which are almost as good as the model learned from the longitudinal data (0.826). Obviously, the key classification is the "point of conversion," and for these data, both TBS and PMTS manage to achieve an accuracy of 0.739.

Finally, we explore the meaning of the discovered trajectories in a little more detail with respect to glaucoma. Looking at the most extreme points according to the multidimensional scaling plots of the TBS trajectories in Fig. 5(a), we can explore some of the characteristics of the two disease states. Fig. 7 shows the values of the NFBs at the extremes of each disease trajectory as well as the data at the most "healthy" point in the cross-sectional data. It seems that one trajectory is characterized by only two NFBs deteriorating in sensitivity (NFBs 1 and 2) whereas the other seems to be characterized by a more general decrease in sensitivity and a particular collapse in NFB 3. A well-known feature of glaucoma is called the "nasal step" whereby early damage is seen in the nasal region represented by the distal regions of NFBs 2 and 5 (see Fig. 2). These findings have been supported by clustering approaches on nontemporal data [13].
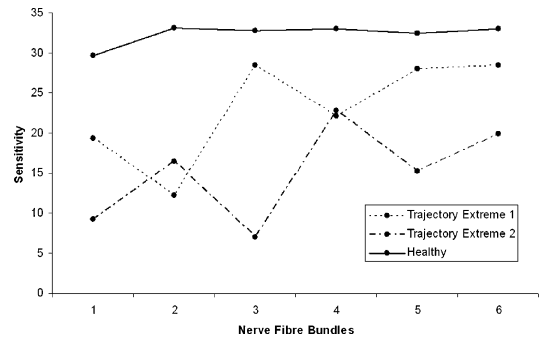
Interestingly, it appears from this figure that the two trajectories result in different patterns of VF loss: NFBs 1 and 2 (in the upper arcuate region) in extreme 1, and more generalized loss including NFBs 3, 5 and 6, in extreme 2.

To summarize the VF data results, it appears that given enough data, both the PMTS and TBS models are capable of learning the temporal behavior of the disease well enough to classify the next state when tested on independent longitudinal data. However, the TBS model shows superior performance to the PMTS model when it comes to forecasting values of NFB variables. The TBS is also shown to be able to automatically discover disease states that represent well-known characteristics of glaucomatous VF loss.

IV. CONCLUSIONS AND FUTURE WORK

In this paper, we have introduced a method to construct approximate temporal models from cross-sectional data without any time information. We adopt an ordering approach using a shortest-path algorithm on multiple samples of data to generate different trajectories through the data. We test this on synthetic data and real-world medical data from VF tests with very promising results. This sort of analysis will potentially enable clinicians to forecast future values of important variables related to disease and to determine on which trajectory a patient is likely to be traveling along. This can greatly improve the choice of appropriate intervention.

A related area of research, known as panel analysis [14], involves trying to build models along both the temporal dimension and the population dimension from panel studies. Another line of research known as *pooling* has explored combining cross-sectional data with time-series data [15]. Fitting trends through data [16] is a common approach and is related in some ways to the idea of identifying a trajectory. Another related area of research is sequence reconstruction. This involves trying to find the best order for a particular set of data. Methods include the traveling-salesman-problem approach that aims to minimize the distance between each datum [17], and more recently, the use of PQ trees has been explored to encode partial orderings in order to account for uncertainty in the data due to elements such as noise [10] (which we adopt in one of our algorithms here).

The research in this paper opens up a multitude of other directions for investigation.

1) Exploring the algorithm on other biological and medical datasets. For example, microarray studies of diseases with slow progression over many years such as Becker muscular dystrophies [18]. Since expression data come from muscle biopsies, there are no longitudinal data from one patient, but it would be extremely valuable to have an idea of how the disease progresses at a molecular level.

2) Incorporating other types of data into the glaucoma models to see what effect they have on the trajectories. For example, in early glaucoma, there is frequently poor agreement between retinal image data and VF data concerning the presence of glaucoma [19], [20]. It could be that the initial manifestation or progression of glaucoma varies between eyes with certain identifiable characteristics. This could create several new trajectories that should be captured by our approach. Also, the exploration of the effects of medication upon trajectory could be explored within this context.

3) There are numerous approaches to dealing with irregular time series [21]. In this paper, we do not address the fact that the "disease-process time difference" between each datum is unknown. It would be interesting to see if distances such as Euclidean could be used in conjunction with irregular time-series methods to interpolate values.

Finally, a note of caution should be made concerning the use of cohort studies where subjects may differ through some non-disease-related factor prevalent at different points in time. Care must be taken in these cases and data must be interpreted appropriately. We intend to carry out further tests to explore the performance of our approach on nonprogressing patients (essentially controls). It is extremely difficult to get hold of longitudinal data for controls and one method could involve generating pseudo time series of healthy patients where the series would "walk" around the healthy region.

ACKNOWLEDGMENT

REFERENCES

[1] N. G. Strouthidis, V. Vinciotti, A. Tucker, S. K. Gardiner, D. P. Crabb, and D. F. Garway Heath, "Structure and function in glaucoma; the relationship between a functional visual field map and an anatomical retinal map," *Invest. Ophthalmol. Vis. Sci.*, vol. 47, no. 12, pp. 5356–5362, 2006.

[2] P. Gouws, E. B. Moss, G. E. Trope, and C. R. Ethier, "Continuous intraocular pressure (iop) measurement during glaucoma drainage device implantation," *J. Glaucoma*, vol. 16, no. 3, pp. 329–333, 2007.

[3] S. Swift and X. Liu, "Predicting glaucomatous visual field deterioration through short multivariate time series modelling," *Artif. Intell. Med.*, vol. 24, no. 1, pp. 5–24, 2002.

[4] A. Tucker, V. Vinciotti, D. Garway Heath, and X. Liu, "A spatio-temporal Bayesian network classifier for understanding visual field deterioration," *Artif. Intell. Med.*, vol. 34, no. 2, pp. 163–177, 2005.

[5] AGIS, "Advanced glaucoma intervention study. 2. visual field test scoring and reliability," *Ophthalmology*, vol. 101, no. 8, pp. 1445–1455, 1994.

[6] G. Wollstein, D. F. Garway Heath, and R. A. Hitchings, "Identification of early glaucoma cases with the scanning laser ophthalmoscope," *Ophthalmology*, vol. 105, pp. 1557–1563, 1998.

[7] D. Kamal, D. Garway Heath, S. Ruben, F. O'Sullivan, C. Bunce, A. Viswanathan, W. Franks, and R. Hitchings, "Results of the betaxolol versus placebo treatment trial in ocular hypertension," *Graefes Arch. Clin. Exp. Ophthalmol.*, vol. 241, pp. 196–203, 2003.

[8] D. F. Garway Heath, F. Fitzke, and R. A. Hitchings, "Mapping the visual field to the optic disc," *Ophthalmology*, vol. 107, pp. 1809–1815, 2000.

[9] L. R. Rabiner, "A tutorial on HMM and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[10] P. M. Magwene, P. Lizardi, and J. Kim, "Reconstructing the temporal ordering of biological samples using microarray data," *Bioinformatics*, vol. 19, no. 7, pp. 842–850, 2003.

[11] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap (Monographs on Statistics and Applied Probability)*. Boca Raton, FL: CRC Press, 1993.

[12] R. W. Floyd, "Algorithm 97: Shortest path," *Commun. ACM*, vol. 5, no. 6, p. 345, 1962.

[13] P. A. Sample *et al.*, "Using unsupervised learning with variational Bayesian mixture of factor analysis to identify patterns of glaucomatous visual field defects," *Invest. Ophthalmol. Vis. Sci.*, vol. 45, pp. 2596–2605, 2004.

[14] D. Kasprzyk, G. Duncan, G. Kalton, and M. P. Singh, *Panel Surveys*. New York: Wiley, 1989.

[15] R. Fanfani, "Pooling time-series and cross-section data: A review," *Eur. Rev. Agric. Econ.*, vol. 2, no. 1, pp. 63–85, 1974.

[16] M. Bianchi, M. Boyle, and D. Hollingsworth, "A comparison of methods for trend estimation," *Appl. Econ. Lett.*, vol. 6, no. 2, pp. 103–109, 1999.

[17] S. S. Skiena, *Traveling Salesman Problem*. Berlin, Germany: Springer-Verlag, 1997, pp. 319–322.

[18] M. Matsuo, "Duchenne/Becker muscular dystrophy: from molecular diagnosis to gene therapy," *Brain Dev.*, vol. 18, no. 3, pp. 167–172, 1996.

[19] F. A. Medeiros, L. M. Zangwill, C. Bowd, and R. N. Weinreb, "Comparison of the gdx vcc scanning laser polarimeter, hrt ii confocal scanning laser ophthalmoscope, and stratus OCT optical coherence tomograph for the detection of glaucoma," *Arch. Ophthalmol.*, vol. 122, pp. 827–837, 2004.

[20] N. G. Strouthidis, A. Scott, N. M. Peter, and D. F. Garway Heath, "Optic disc and visual field progression in ocular hypertensive subjects: Detection rates, specificity, and agreement," *Invest. Ophthalmol. Vis. Sci.*, vol. 47, pp. 2904–2910, 2006.

[21] X.-T. Zhang, W. Zhang, X. Xiong, Q.-W. Wang, and C.-Y. Li, "A model-based clustering for time-series with irregular interval," in *Proc. IEEE Int. Conf. Mach. Learning Cybern.*, 2004, vol. 5, pp. 2883–2888.

Allan Tucker received the B.Sc. degree in cognitive science from the University of Sheffield, Sheffield, U.K., in 1996, and the Ph.D. degree in computer science from Birkbeck College, University of London, London, U.K., in 2001.

For four years, he was a Postdoctoral Fellow at the Centre of Intelligent Data Analysis, Brunel University, Uxbridge, U.K. He is currently a Research Lecturer with the School of Information Systems, Computing and Mathematics, Brunel University. His current research interests include machine learning, data mining, Bayesian networks, bioinformatics, and medical informatics.

David Garway-Heath received the M.D. degree on structure/function correlations in glaucoma from the University of London, London, U.K., in 2001.

He studied medicine at St. Thomas' Hospital, University of London, London, U.K., and undertook his Residency in ophthalmology at Moorfields Eye Hospital, London. He is currently a Consultant Ophthalmologist at Moorfields Eye Hospital, and Glaucoma Theme Lead at the National Institute for Health Research (NIHR) Biomedical Research Centre for Ophthalmology, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London. He is also Honorary Visiting Professor at the City University, London. His current research interests include optic nerve and retinal imaging, psychophysics, tonometry, ocular biomechanics, and risk factors for glaucoma.