Next Generation Flow Field Forecasting

Kyle A. Caudle * Patrick Fleming † Michael Frey ‡ Noah Brubaker §

Abstract
A statistical learning methodology, called flow field forecasting, was introduced in 2011 for forecasting a univariate time series. Flow field forecasting draws information from a flow field interpolated over a history space of the observed time series to incrementally build a forecast. We extend flow field forecasting to the setting of multiple data streams. Also, flow field forecasting as originally formulated assumed that an appropriate set of predictors was known for the history space. Two methods are presented for automatically selecting the history structure to use for each increment of the forecast. These results are preliminary in nature.

Key Words: times series, forecasting, multivariate data streams

1. Introduction

A new general purpose method for forecasting univariate time series data called flow field forecasting (FFF) was introduced in 2011 [5]. FFF was introduced to address various limitations of traditional forecasting methods (Box-Jenkins ARIMA [2], exponential smoothing [1, 3, 4], artificial neural networks [7, 8]) . Specificially, the impetus for FFF was to create forecasting software with the following features:

- No expert guidance needed
- Ability to accept non-uniformly spaced observations
- Fast/Computationally efficient
- Error bounds on forecast estimates
- Ability to exploit parallel data

The basic premise of flow field forecasting is that past associations between history and change are predictive of changes associated with current histories/future changes. Flow field forecasting is a three-step framework. The aim of step I in flow field forecasting is to extract associations from the given data record. Flow field histories are short sequences of process levels and level changes. The associated changes are simply those changes that were observed to follow immediately from each history. Fig. 1 is a visualization of histories and subsequent level changes.

In step II of flow field forecasting, the set of associations—histories with changes (called the data skeleton) is used to construct an interpolator so that, given any current history, an associated change can be interpolated from the changes in the data skeleton. Interpolation can be accomplished by different available methods; to perform step II we use Gaussian Process Regression (GPR) [9]. The two panels in Fig. 1 each illustrate an immediate future change d_* to be interpolated from a current history $\mathbf{h}_*$. In each panel the changes labeled d_1 and d_2 should weigh most heavily in the interpolation of d_* because the histories $\mathbf{h}_1$ and

*South Dakota School of Mines and Technology, Rapid City, SD 57701, USA
†South Dakota School of Mines and Technology, Rapid City, SD 57701, USA
‡Bucknell University, Lewisburg, PA 17837
§South Dakota School of Mines and Technology, Rapid City, SD 57701, USA

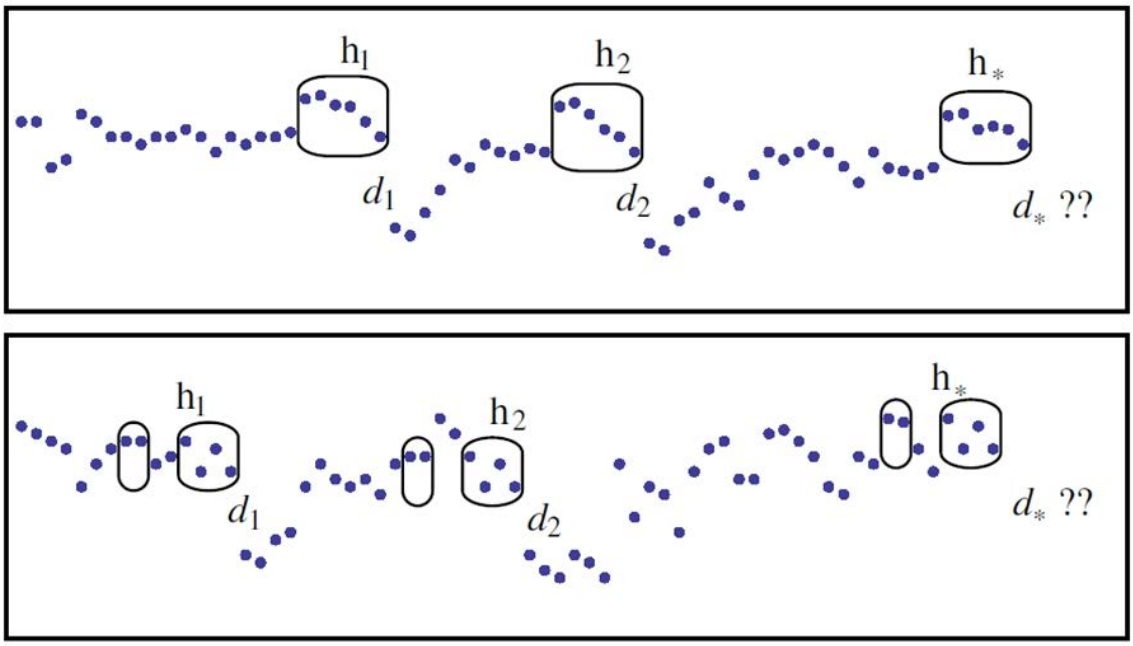

Figure 1: Past histories $\mathbf{h}_1, \mathbf{h}_2$ and their associated changes d_1, d_2. These associations are used to interpolate the change d_* associated with the current history $\mathbf{h}_*$. The histories in the top panel have consecutive components. The histories in the bottom panel have non-consecutive components.

$\mathbf{h}_2$ are most similar to the current history $\mathbf{h}_*$. The top panel of Fig. 1 shows histories based on consecutive points in the data skeleton; the bottom panel makes the point that histories need not be based on consecutive points. The choice of structure for the history—the predictors to include—is a flexibility of flow field forecasting. A purpose of this paper is to present two approaches for automatically selecting the history structure. Step III of flow field forecasting uses the interpolater to incrementally construct a forecast of the future path of the process.

2. Flow Field Forecasting

In our 2013 paper [6] we showed for a baseline data model that FFF performs competitively relative to traditional forecasting methods. The strength of FFF lies in the fact that when a distinctive history recurs, FFF substantially outperforms these other forecasting methods. Figure 2 shows the comparison results for FFF when a distinctive history recurs. A more detailed discussion of the time series models used to produce Fig. 2 can be found in [6]. Of note is that these univariate results are similar to the bivariate time series results presented in section 5 of this paper.

Flow field forecasting takes its name from the fact that the GPR in the second step predicts a change—a velocity, in other words—so that the GPR's response surface is a flow (velocity) field. It is appropriate to predict process change rather than the process level itself because, in the simple case of GPR without a

mean model, GPR forecasts default to zero in sectors of its response surface away from observed data. In other words, using GPR to predict process change means that, in the absence of past histories similar to the current history, the GPR predicts zero change. This means that FFF for univariate time series has the built-in assumption that predicting no change is the best (most "unbiased") choice in the absence of relevant data. As we will show in section 5, predicting no change may not be the best option when forecasting in more than one dimension.

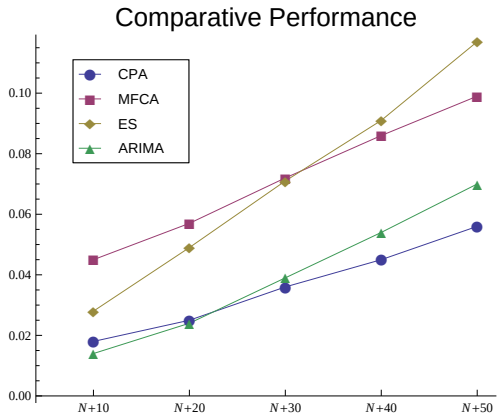

Figure 2: Comparison results for a time series where a distinctive history recurs (N=1500).

3. Bivariate Forecasting

The next generation of flow field forecasting begins with forecasting a bivariate time series of n observations (x_i, y_i) at each observation time t_i $(i = 1, 2, ..., n)$. To begin, we assume that the data collected has been preprocessed so that it is standardized and the process noise has been estimated. For discussion of how this might be done using penalized spline regression, see [6].

As stated earlier, the basic premise of flow field forecasting is that past associations between history and change are predictive of changes associated with current histories/future changes. Therefore, we seek to identify histories "closest" to the current history and then base our forecast on what occurred in the past when these closest histories were observed. In this section, we describe two methodologies for associating the current history with previous histories. Each method has advantages and disadvantages. We expect that continuing investigation will yield a forecasting procedure that capitalizes on the strengths of both.

3.1 Closest Point Approach

The closest point approach (CPA) starts with a large set of candidate predictors. This user-selected set is based variously on qualitative knowledge of the given time series behavior and available computing resources. The candidate set of predictors may include any combination of current and previous (x, y) values as well as current and previous first (or possibly higher) order derivatives of x and y. The goal is a set of

candidate predictors both sufficiently rich to capture the structure of the history space and sufficiently parsimonious to not over-tax available computing resources.

Given a set of p candidate predictors, there are $q = 2^p - 1$ non-trivial candidate subsets. Each candidate subset represents a potential history structure. The goal of CPA is to identify the history structure (i.e. candidate subset) that yields the smallest distance from each historical point and the current point. To accomplish this, we compute the following distance matrix.

$$D = \begin{bmatrix} ||C-P_1||_{H1} & ||C-P_1||_{H2} & \dots & ||C-P_1||_{Hq} \\ ||C-P_2||_{H1} & ||C-P_2||_{H2} & \dots & ||C-P_2||_{Hq} \\ \vdots & \vdots & \vdots & \vdots \\ ||C-P_i||_{H1} & ||C-P_i||_{H2} & \dots & ||C-P_i||_{Hq} \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

Where $||C-P_i||_{Hj}$ represents the distance between the current point (C) and point P_i under history structure H_j.

We would like to determine the minimum distance in our distance matrix. A problem arises however, because the history structures have different dimensions. We proceed by calculating the minimum of each column.

$$P^*_{H_j} = \text{argmin}_{P_i} \|C - P_i\|_{H_j} \tag{1}$$

In order to compare each of these minimum values, we standardize them by subtracting the column mean and then dividing by the column standard deviation:

$$Q_{H_j} = \frac{P^*_{H_j} - \overline{D}_j}{sd(D_j)} \tag{2}$$

Standardizing is necessary because we need a method for comparing distances across different dimensional history structures. After standardizing, we can now compare distances between history structures of different dimensions. The closest point is identified by the smallest Q_{H_j} in equation (2). By inspection of equation (2) we see that minimum value of Q_{H_j} provides us both the closest point and the history structure that achieves that smallest distance. We use the closest point to forecast the next value of (x, y).

The CPA algorithm is statistically equivalent to adding a penalty to the distance when comparing two different dimensional history structures. Suppose we have a history structure of dimension j whose minimum standardized distance is Q_{H_j} and we want to make a comparison with a history of dimension k whose minimum standardized distance is Q_{H_k}. Let

$$Q'_{H_j} = \frac{P^*_{H_j}}{sd(D_j)} \text{ and } Q'_{H_k} = \frac{P^*_{H_k}}{sd(D_k)}$$

Next, we check to see if $Q'_{H_j} + \Pi_{jk} < Q'_{H_k}$, where

$$\Pi_{jk} = \frac{\overline{D}_k}{sd(D_k)} - \frac{\overline{D}_j}{sd(D_j)}.$$

This shows that comparing standardized minimimum distances effectively introduces a penalty term Π_{jk} to account for different dimensions.

3.2 Mean Flow Certainty Approach

The Gaussian process regression flow field estimates in step II are interpolations of these flow field values. Gaussian process regression readily yields standard errors for its estimates. In [5] we define the mean flow certainty, which expresses (through the variance) the reduction in uncertainty that would otherwise result from a knot step into the future. The mean flow certainty reflects the overall usefulness of the information in the interpolated flow field for the forecast. If the region where we are trying to forecast is poorly represented in the history space, then the mean flow certainty will be close to zero. Alternatively, if the region where we are trying to forecast has histories that are similar to the forecast region than the mean flow certainty will be close to one.

We use the mean flow certainty, along with the mean prediction error, to determine which history structure has the greatest potential to produce an accurate forecast. We start by building the full history space (H). The full history space may be expressed as an $(n-d+1)\times 4d$ matrix.

$$H = \begin{bmatrix} \mathbf{h_d} \\ \mathbf{h_{d+1}} \\ \vdots \\ \mathbf{h_n} \end{bmatrix} \tag{3}$$

where,

$$\mathbf{h_i} = \begin{bmatrix} x_i & x_{i-1} & \dots & x_{i-d+1} & y_i & y_{i-1} & \dots & y_{i-d+1} & \Delta x_i & \Delta X_{i-1} & \dots & \Delta x_{i-d+1} & \Delta y_i & \Delta y_{i-1} & \dots & \Delta y_{i-d+1} \end{bmatrix}$$

Using all but the last 5 data values, we test all possible subsets of our full history space. For each subset of H as a history space, we use standard flow field forecasting in order to forecast the last five data values. We then calculate the mean prediction error for the five withheld data values and the mean flow certainty for each of these subsets. The prediction strength (PS) is a measure of these two measures.

$$\text{PS} = \text{MFC} \times \exp(-\text{MPE}). \tag{4}$$

The best history structure will be the one whose prediction strength is closest to 1.

We conclude this section by noting benefits and drawbacks of this approach. First, MFCA is very effective at identifying a change in the data pattern and then adding weight to the slopes of past histories which have a similar data pattern. This weighted average of past histories has proven to be an effective method of forecasting. Next, the bivariate nature of the history structure allows the software to identify the relationships between x and y values. If we performed two univariate forecasts, one for x and one for y, this would not be the case. As an example, suppose the underlying structure of the data can be represented by a system of differential equations in which the change in y is heavily influence by the position of x. Performing two univariate forecasts would be not be nearly as effective as our method. And finally, MFCA requires minimum input from the user. Only maximum history depth (d) and history structure size (n) needs to be chosen.

During our testing, we have noticed the previously discussed methods have some failings. First, choosing the best history structure is computationally expensive. Next, if the data is streaming and the history space is growing, the history structure will need to be updated periodically to compensate for changing structures in the data patterns. Finally, the GPR process also requires a correlation length in each of the predictors. Finding the optimum values is neither theoretically nor computationally trivial.

3.3 Hybrid Approach

A major drawback of CPA is that if an incorrect point is chosen, the resulting forecast will be poor. To reduce the likelihood of this, we propose a hybrid approach. Our hybrid approach "buys insurance" against incorrectly picking the closest point by expanding consideration to a small set of closest points.

Beginning with a small set of closest points, in our hybrid approach we use GPR to interpolate amongst just the closest points. The GPR automatically down weights those points which match less well. The benefit of this is that while the closest point still gets great weight, it does not get all the weight. In this way we protect ourselves from a poorly chosen closest point. The disadvantage, of course, is that if the chosen closest point is indeed the closest point, our forecast is slightly degraded because we are not giving it full weight. Further study is needed to better understand how many points to include and what the change in forecast would be in the situation where the closest point was chosen in the first place.

4. Simulation Study

We now describe an initial, restricted simulation study. We note that, while the results presented here are limited, they provide evidence that our forecasting method is a viable alternative to the more traditional forecasting methods. For the simulation study, we generate 30 realizations from the following parametric function.

$$x(t) = \frac{1}{8}\left[t + \frac{1}{2} * \cos(3*t) + \frac{1}{10} * \sin(10*t) + N(0, 0.05)\right]$$

and

$$y(t) = \frac{1}{8}\left[t + 3 * \sin(t) + \frac{1}{10}\cos(10*t) + N(0, 0.05)\right]$$

One such realization is pictured in figure 3.

We compare CPA and MFCA to ARIMA, Exponential Smoothing and Neural Network forecasts. Each time series was comprised of 1550 data points. We withheld the last 50 points for testing, so the forecast model in each case was built on the first 1500 points. The forecast error was determined by calculating the L_2 norm for each forecast value.

$$\text{Forecast Error} = \frac{1}{30}\sum_{i=1}^{30}\sqrt{(x_i - x_t)^2 + (y_i - y_t)^2} \tag{5}$$

Where (x_t, y_t) is the true value at forecast time t.

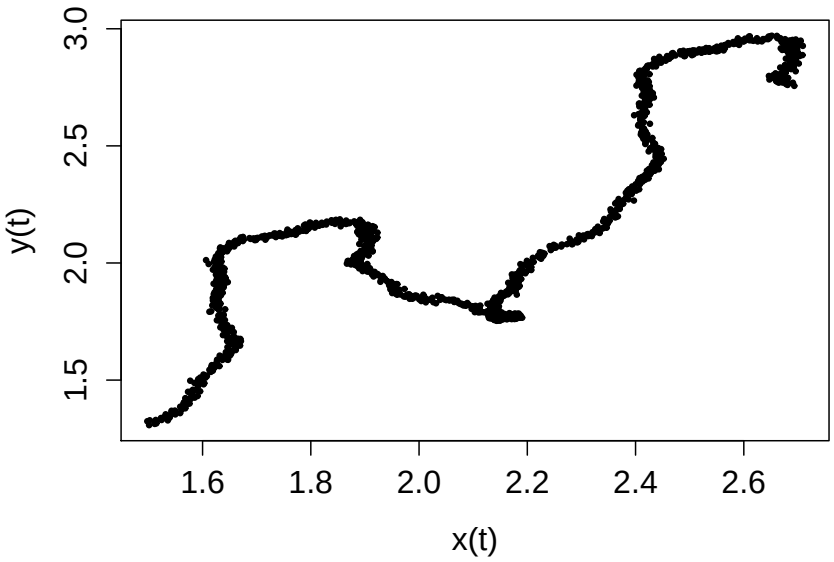

Figure 3: One realization of a data record with N = 1500 uniformly spaced observations. Additive noise is normally distributed with mean zero, standard deviation 0.05.

Figure 4 shows the flow field comparisons with ARIMA and exponential smoothing. The neural network results were an order of magnitude larger so for visualization purposes were not included in the graph. The actual values are included in table 1.

Table 1: Comparison Results

Time	CPA	MFCA	Esmooth	ARIMA	NNet
1510	0.018	0.045	0.028	0.014	0.618
1520	0.025	0.057	0.049	0.024	0.660
1530	0.036	0.072	0.071	0.039	0.656
1540	0.045	0.086	0.091	0.054	0.672
1550	0.056	0.099	0.117	0.070	0.698

The data for this simulation study was selected because it represents a situation where a distinctive history recurs. Because flow field forecasting is looking for patterns in the history, flow field forecasting was able to detect this recurring history, whereas the other traditional methods were not able to detect this change. We also note that the results are consistent with the univariate study that was performed in reference [5] (figure 2); that is, flow field forecasting is competitive in the short range, but decisively superior at longer range forecasting, as shown in both figures 2 and 4.

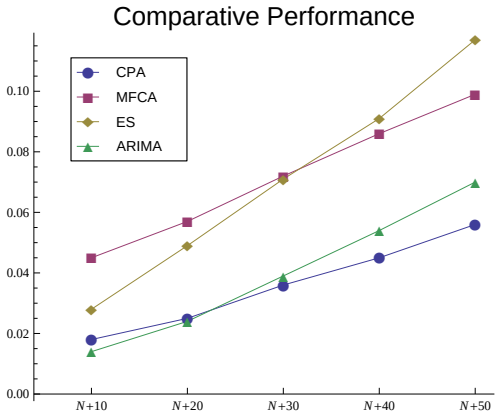

Figure 4: Comparative results for parametric data model (N=1500).

5. Future Work

Flow field forecasting (FFF), as originally proposed, is a three-step process for forecasting a univariate time series. The first step uses a semi-parametric regression to: 1) identify and strip away additive IID (white) noise and 2) adjust the time series observations so that they are uniformly spaced. This first step also performs a data reduction role where this is a need. In FFF's first step the size of the spacing between derived observations is a matter of choice, and in making this choice the relationship among observation spacing, the mean magnitude of incremental change, and the desired time of the forecast into the future needs to be considered. In our present study of flow field forecasting, we assume that this first step has been accomplished so that we have uniformly spaced observations with little or no additive IID noise and so that the observations are spaced to a degree commensurate with the amount of change observed in the data record.

The second step in FFF uses Gaussian process regression (GPR) to predict the change in the time series at the current time. GPR makes this prediction using previous changes and their associated histories as data. The choice of GPR in the second step has advantages: it has well-understood properties, it yields standard errors with its predictions, it is relatively easily coded, and it runs efficiently. Other techniques such as regression trees, though, could possibly be used. Using GPR in the second step, the third step of FFF simply adds the GPR change prediction to the current observation to construct a forecast of the time series level at the first future point. These second and third steps can be applied iteratively to predict successive changes and forecast the time series level at subsequent points (with increasing error going forward.)

FFF's built-in unbiasedness assumption is open to question in multivariate FFF. Consider forecasting the dynamics of particle moving in a physical space of two dimensions. The particle may exhibit a net drift in a certain direction and we may want to preserve that drift in our forecast even in the absence of relevant data. This is readily accomplished by using GPR to forecast the "change in change"—acceleration, in other words. GPR, and other regression methods, can certainly do this; we can always associate both an acceleration and a history to each of our observations and use this association to perform a regression. GPR to predict acceleration will default to a zero-acceleration forecast in the absence of relevant data, thereby preserving the particle velocity. Newton's second law tells us the acceleration is proportional to force, so

we use the name force field forecasting to describe the use of GPR to forecast acceleration and make time series forecasts. Force field forecasting assumes that a process in motion tends to remain in motion. This assumption may or may not be more natural and appropriate than FFF's choice to zero out the velocity in the absence of data.

We arrived at the notion of force field forecasting through a consideration of the default prediction inherent in GPR, and the actual name "force field forecasting" is based on little more than a superficial appeal to Newton's second law. In the classical mechanical situation, though, of a particle with unit mass moving under force, the GPR response surface in force field forecasting is indeed a force field. Even outside this limited classical mechanical setting, it may be useful to think in these terms. For example, if we consider that we are using a GPR to interpolate a force field we might suppose—simply to lend structure to the problem—that the force field is conservative. In any conservative (or path-independent) force field, the field integral along any closed path is zero. Given a GPR-interpolated force field, this zero-integral condition can be tested, and unless the interpolated force field sufficiently violates this condition, we might require that the interpolated force field be re-computed to meet this condition. Time series forecasting done in this fashion can be termed conservative force field forecasting. Associated with any conservative force field is a potential energy, so conservative force field forecasting is from this perspective a matter of estimating a potential energy function, which is a task much simpler than estimating a (vector) force field.

We remark, finally, that our preceding considerations of variants of flow field forecasting grew out of our choice of GPR in step 2. However, these observations apply as well with other procedures, such as regression trees, used in place of GPR. A regression tree basically splits the history space (with dimension N equal to the number of predictors) into roughly 2N regions and assigns a response to each region. Depending on the size of N, this can become computationally intensive. To make a forecast, though, we do not need the entire regression tree response surface; we only need the response in the region containing the current history. Finding one region of a regression tree appears computationally much easier than constructing (growing) the whole tree. In fact, the order of the calculation changes from exponential-in-N to just linear-in-N. Linear-in-N would allow us to work with very rich history structures, with the best structure chosen automatically by the regression tree. Also, the regression tree prediction comes with a standard error. Additionally, we can easily incorporate different subspace bases into the development (growth) of the tree. This is an important point. To clarify this, consider the joint time evolution of electronic pressure P, temperature T, and magnetization M at a point in a plasma. This is a trivariate process evolving in a three-dimensional space with three distinct subspaces defined by the process's three components P, T, and M. Contrast this example with the example of a particle with a varying temperature moving in a plane. This is also a trivariate process evolving in a three-dimensional space, the process components being location (X,Y), and temperature T. In contrast to the first example, this second example has only two distinct subspaces, location and temperature. The nominal X,Y basis of the location subspace is not privileged and does not define distinct subspaces for the purpose of choosing predictors; we might as well use X+Y and X–Y as our basis and use these to guide our choice of predictors to include in our history. The possibility of multidimensional subspaces in multivariate time series forecasting is an important way in which the multivariate situation differs from univariate forecasting. Regression tree forecasting can potentially robustly discover good subspace bases for selecting history predictors. We hope to report multivariate time series forecasting results with regression trees in the near future.

6. Final Remarks

This paper presents initial steps toward extending FFF beyond its original formulation in two important ways: from univariate to multivariate time series forecasting and toward automatic history structure selection. Neither of these two efforts is trivial and neither is fully accomplished here. With standard general-purpose forecasting tools, multivariate time series forecasting typically devolves in practice to treating the data as parallel univariate time series and making forecasts component-by-component, with no advantage taken of information carried jointly. FFF, by contrast, conceptually readily extends to accept multivariate time series. This adaptation is based on separate interpolations for each multivariate component, with each interpolation based on predictors pooled from *all* the component series. This gives each interpolation full access to both component and joint information relevant to the interpolation. Pooling predictors, though, linearly increases their potential number for each interpolation and exponentially increases the dimensionality of the potential history structures.

In the introduction of FFF for univariate time series, little attention was paid to automatic history structure selection. Selection in the univariate case is manageable on a case-by-case basis because only a relatively few predictors are likely to be important in a given problem, and these are relatively easily found by exhaustive search. So, in the univariate case automatic selection is desirable but not necessary. For multivariate time series, on the other hand, automatic selection is both desirable and, increasingly, essential. Thus, multivariate series and automatic history structure selection present a conjoined challenge. We have proposed and demonstrated two solutions to this challenge, our CPA and MFCA methods, each showing some promise at this initial stage of study. We only investigated bivariate series here and we expect that some aspects of our lines of approach will be superseded by better ideas. In any case, our initial results show promise that FFF can eventually robustly treat multivariate time series in full multivariate form, doing so by efficient and effective automatic history structure selection.

Our immediate goal is to finalize a method for automatically selecting the history structure and also to determine which method (i.e. forecasting change or the change in change) is most suitable for multivariate data. We plan to report these findings elsewhere.

Acknowledgments

This publication [article] results from research supported by the Naval Postgraduate School Assistance Grant N00244-15-1-0052 awarded by the NAVSUP Fleet Logistics Center San Diego (NAVSUP FLC San Diego). The views expressed in written materials or publications and/or made by speakers, moderators, and presenters, do not necessarily reflect the official policies, or organizations imply endorsements by the U.S. Government.

References

[1] G.E.P. Box. Understanding exponential smoothing—a simple way to forecast sales and inventory. *Quality Engineering*, 3(4):561–566, 1991.

[2] G.E.P. Box, G.M. Jenkins, and G.C. Reinsel. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, Hoboken, NJ, 4th edition, 2008.

[3] R.G. Brown. *Statistical Forecasting for Inventory Control*. McGraw-Hill, New York, 1959.

[4] R.G. Brown. *Smoothing, Forecasting, and Prediction*. Prentice Hall, Englewood Cliffs, NJ, 1963.

[5] Michael R. Frey and Kyle A. Caudle. Introducing flow field forecasting. In *Proceedings of the 10th International Conference on Machine Learning and Applications*, IEEE, 2011.

[6] Michael R. Frey and Kyle A. Caudle. Flow field forecasting for univariate time series. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 6:506–518, 2013.

[7] S. Haykin. *Neural Networks and Learning Machines*. Pearson, New York, 3rd edition, 2009.

[8] T. Hill, L. Marquez, M. O'Connor, and W. Remus. Artificial neural network models for forecasting and decision making. *International Journal of Forecasting*, 10:5–15, 1994.

[9] C.E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA, 2006.