Input variable selection for time series prediction with neural networks– an evaluation of visual, autocorrelation and spectral analysis for varying seasonality

Sven F. Crone[1] and Nikolaos Kourentzes[1]

1- Lancaster University Management School, Dept. of Management Science
Lancaster, LA1 4YX, United Kingdom

Abstract. The identification and selection of adequate input variables and lag structures without domain knowledge represents one the core challenges in modeling neural networks for time series prediction. Although a number of linear methods have been established in statistics and engineering, they provide limited insights for nonlinear patterns and time series without equidistant observations and shifting seasonal patterns of varying length, leading to model misspecification. This paper describes a heuristic process and stepwise refinement of competing approaches for model identification for multilayer perceptrons in predicting the ESTSP'07 forecasting competition time series.

1. Introduction

Artificial neural networks (NN) have found increasing consideration in forecasting theory, leading to successful applications in time series prediction and explanatory forecasting [1]. Despite their theoretical capabilities, NN have not been able to confirm their potential in forecasting competitions against established statistical methods, such as ARIMA or Exponential Smoothing [2]. As NN offer many degrees of freedom in the modelling process, from the selection of activation functions, adequate network topologies of input, hidden and output nodes, learning algorithms etc. their valid and reliable use is often considered as much an art as science. Previous research indicates, that the parsimonious identification of input variables and lags to forecast an unknown data generating process without domain knowledge poses a key problem in model specification [1, 3]. This becomes particularly important, as complex time series components may include deterministic or stochastic trends, cycles and seasonality, interacting in a linear or nonlinear model with pulses, level shifts, structural breaks and different distributions of noise. Although a number of statistical methods have been developed to support the identification of linear dependencies, their use in nonlinear prediction has not been investigated in detail.

Therefore a structured evaluation of different methodologies to specify the input vector of NN in time series forecasting is required. This paper contributes to the discussion, presenting an analysis of different methodologies of input variable identification through an empirical simulation on the ESTSP forecasting competition time series. This paper is organized as follows. First, we briefly introduce NN in the context of time series forecasting and various methodologies for input variable

identification. Section III presents the experimental design and the results obtained. Finally, we provide conclusions and future work in section IV.

2. Methods

2.1 Forecasting with Multilayer Perceptrons

Forecasting with NNs provides many degrees of freedom in determining the model form and input variables to predict a dependent variable $\hat{y}$. Through specification of the input vector of n lagged realisations of only the dependent variable y a feedforward NN can be configured for time series forecasting as $\hat{y}_{t+1} = f\left(y_t, y_{t-1}, ..., y_{t-n+1}\right)$, or by including i explanatory variables x_i of metric or nominal scale for causal forecasting, estimating a functional relationship of the form $\hat{y} = f\left(x_1, x_2, ..., x_z\right)$. By extending the model form through lagged realisations of the independent variables $x_{i,t-n}$ and dependent variable y_{t-n} more general dynamic regression and autoregressive (AR) transfer function models may be estimated. To further extend the autoregressive model forms of feedforward architectures, recurrent architectures allow incorporation of moving average components (MA) of past model errors in analogy to the ARIMA-Methodology of Box and Jenkins [4]. Due to the large degrees of freedom in modelling NN for forecasting, we present a brief introduction to specifying feedforward NN for time series modelling; a general discussion is given in [5, 6]. Forecasting time series with NN is generally based on modelling a network in analogy to an non-linear autoregressive AR(p) model using a feed-forward Multilayer Perceptron (MLP) [1]. The architecture of a MLP of arbitrary topology is displayed in figure 1.

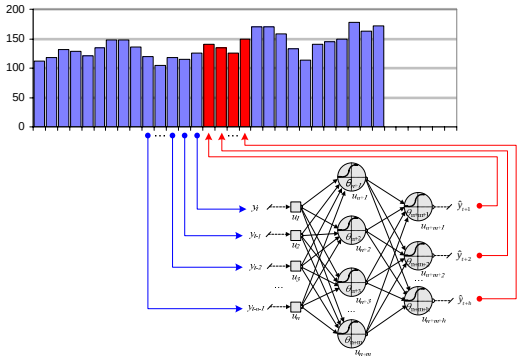

Fig. 1: Autoregressive MLP for time series forecasting.

In time series prediction, at a point in time t, a one-step ahead forecast $\hat{y}_{t+1}$ is computed using $p=n$ observations $y_t, y_{t-1}, \ldots, y_{t-n+1}$ from n preceding points in time t, $t-1$, $t-2$, …, $t-n+1$, with n denoting the number of input units of the ANN. Data is presented to the MLP as a sliding window over the time series observations. The task of the MLP is to model the underlying generator of the data during training, so that a valid forecast is made when the trained ANN network is subsequently presented with a new input vector value [5].

The network paradigm of MLP offers extensive degrees of freedom in modeling for prediction tasks. Structuring the degrees of freedom, each expert must decide upon the selection and sampling of datasets, the degrees of data preprocessing, the static architectural properties, the signal processing within nodes and the learning algorithm in order to achieve the design goal, characterized through the objective function or error function. For a detailed discussion of these issues and the ability of NN to forecast univariate time series, the reader is referred to [1]. The specification of the input vector has been identified as being particularly crucial to achieving valid and reliable results, and will be examined in the next section.

2.2 Input Variable Selection for Multilayer Perceptron predictions

The identification of relevant input variables and variable lags aims at capturing the relevant components of the data generating process in a parsimonious form. In time series modeling, it is closely related with identifying the underlying time series components of trend and seasonality and capturing their deterministic behavior in lags of the dependent variable. A simple visual analysis of the time series components frequently fails to reveal the complex interactions of autoregressive and moving average components, multiple overlying and interacting seasonalities and nonlinear patterns. Several methodologies have been developed for input variables selection of the significant lags in forecasting, originating from linear statistics and engineering. However, currently no uniformly accepted approach exists to identify linear or nonlinear input variables [1].

Seasonality is frequently identified following the Box-Jenkins methodology of linear statistics [4] as a mixture of autoregressive and moving average components. The specification of a parsimonious input vector requires a stepwise analysis of the patterns in the plotted autocorrelation function (ACF) and partial autocorrelation function (PACF) to identify statistically significant autoregressive lags of the dependent variable and of moving average lags of the errors of past predictions. The iterative methodology is frequently employed in identifying significant lags for NN forecasting, following Lachtermacher and Fuller [7]. As in detrending, no consensus exits on whether a time series with identified sesonality should be deseasonalised first to enhance the accuracy of NN predictions [3, 8, 9] or seasonality be incorporated as AR- and MA-components in the NN structure [10-13]. Earlier studies in MLP modeling claim that an analysis of the AR-terms purely from PACF-analysis is sufficient to identify the relevant lags of the time series [14]. However, an AR-analysis can only reveal linear correlations within the time series structure, but not of linear moving average components that require the use of recurrent NN architectures. In addition, ACF and PACF analysis allow no identification of nonlinear interdependencies [1]

In addition, spectral analysis (SA) may provide additional information on the linear autoregressive structure of multiple seasonalities with overlaying periodicities in comparison to an ACF - & PACF-analysis [15], albeit losing information on the potential moving average structure. SA expresses a time series as a number of overlaid sine and the cosine functions of different length or frequency. It identifies the correlation in a periodigram using Fast Fourier Transforms (FFT), which plot the power spectral density versus the frequency of the signal to identify frequencies of

high power as an indication of a strong periodicity. To recode the power spectrum as lags instead of frequencies, we plot the horizontal axis of the periodogram as $n/2$ lags of the time series. This allows a direct association of the power and lags, since the power is expressed in non-continuous terms and directly associated with a specified lag. Significant spikes in the periodigram identify interrelations as input lags for the NN model, which will allow the network to learn and extrapolate the overlaying periodicities. Consequently, SA can be employed in analogy to the ARIMA-methodology to identify periodicities and lags in the time series.

3. Experimental Design

3.1 Exploratory Data Analysis

A single time series of 875 observations, displayed in fig. 1, was provided for the forecasting competition of the 2007 European Time Series Symposium (ESTSP).

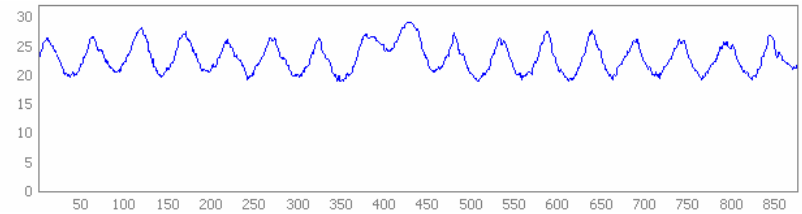

Fig. 2: ESTP2007: Competition time series

The ESTSP competition evaluates the forecasting accuracy on a single time series from a single time origin using the mean squared error (MSE) on the next 15 and the next 50 observations. No domain knowledge was provided to aid the identification of a suitable input vector or network architecture, making the selection of input variables one of the core problems in the competition task. Several different modelling approaches were evaluated, including visual analysis, Autocorrelation analysis and Spectral analysis using FFT.

A visual analysis of the time series reveals a non-trended, seasonal structure of approximately 52 observations with a high signal to noise ratio and a single seasonal outlier that promises easy approximation and extrapolation with a deterministic sine function and exogenous outlier correction. A further visual analysis of the repeating sine pattern displays the repetitive structure in a seasonal diagram, overlying each season containing 52 observations as separate time series displayed in fig. 2. The seasonal diagram confirms a general seasonal structure and two outliers. Although the series seems to obey a 52 observation seasonal length, the peaks of multiple series do not correlate adequately as visualised by the horizontal shift of the series. In contrast to a vertical variation caused by the inherent randomness of the series this indicates an inconsistent or shifting seasonal pattern. The length of 16.8 seasons in the complete series also suggests a seasonality of varying length, rather then an incomplete time series with 9 missing observations. To evaluate this a simple benchmark using a 52 period seasonality in a t-52 lag structure is modelled and evaluated in MLP_{NAIVE} .

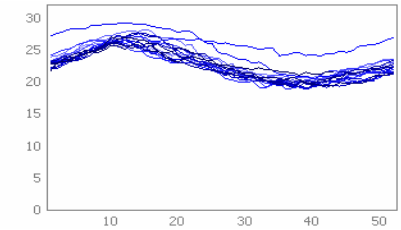

Fig. 3: Seasonal diagram of a 52 observations length

Furthermore, the time series was annotated in a descriptive analysis, to analyse the properties of the seasonal structure of the time series in further detail. Fig. 3 shows the time series plot overlayed with a sine of 52 observations period length.

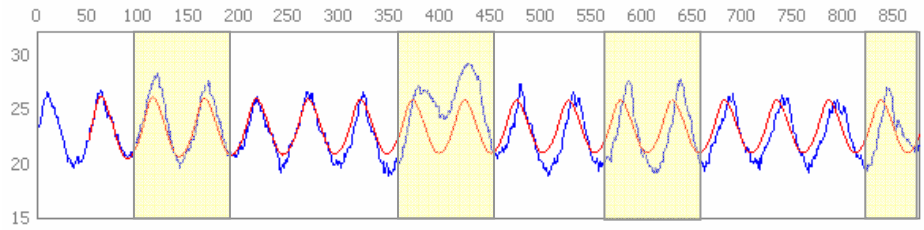

Fig. 4: Time series overlaid by a 52 observations period sine.

Two anomalous seasonal patterns are evident from observation 350 to 450 and in the last season from 825 to 875. Before each anomalous pattern 7 normal seasons are identifiable, with a further pattern of 2 seasonal peaks with increased magnitude every 2 and 3 seasonal patterns apart, indicated as shaded seasons in fig. 3. Although this may suggest a structural pattern in the data generating process, and that an important deviation from the more general model form may be expected for the final ex ante forecasted seasonality, too limited evidence of 1 full cycle is provided. Hence we must consider leaving the anomalies as part of the general model structure or modelling them as outliers using binary dummy variables for NN predictions.

Additionally, a comparison of the time series in fig.3 and a sine function with constant seasonality of 52 observations reveals a varying length of the seasonal patterns. To quantify the pattern of varying seasonal length we estimate the number of observations between each seasonal maxima and minima, applying a 9 period moving average to smooth out randomness. The variation of seasonal lengths in Table 1 shows an average length of 52.06 observations between minima with a standard deviation of 2.59 observations and a significant range of up to 10 observations.

Season	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Length	52	52	51	51	53	51	55	53	50	51	56	49	50	53	58	48

Table 1: Varying seasonal lengths of the time series.

The average length of the seasonality of 52.06 is supported by the visual analysis of the seasonal pattern. However, it biases the identification through ACF and PACF analysis as well as the SA periodigrams, as the temporal interdependencies vary along the time series. Also, the varying seasonal length provides problems as

conventional MLP models assume an AR(p)-process with a deterministic seasonality in an input vector of fixed length. However, the varying seasonality appears to be not entirely stochastic, as a plot of the seasonal lengths in Fig. 4 suggests a regular pattern that may allow exploitation to predict the seasonal length of the forecasted period through the model form or an explanatory variable.

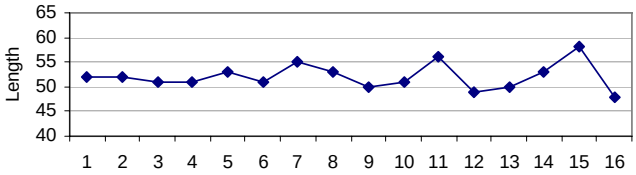

Fig. 5: Length of the time series “sine blocks”

In contrast, the autocorrelation analysis following the Box-Jenkins methodology reveals contradicting information on the more complex structure of the time series. The analysis of the ACF and PACF patterns provided in figure 5a and 5b allow an iterative identification of the significant seasonal or shorter lags using single or seasonal differencing.

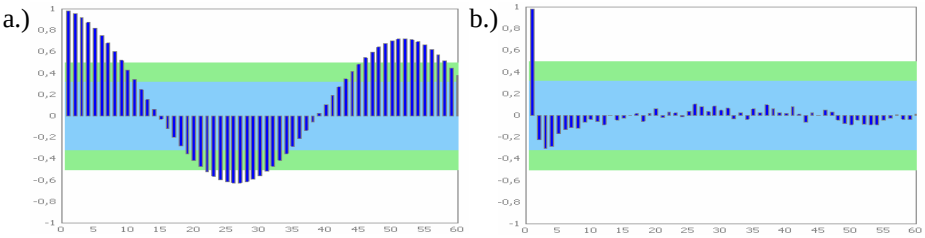

Fig. 6: ACF plot (a.) and PACF plot (b.) of the ESTSP competition time series

The information on the seasonal structure derived from fig. 5 is ambiguous. The ACF plot in fig. 5a reveals a significant seasonal autoregressive process in a decaying, sinusoid pattern of the ACF of a length shorten then 52 periods. In contrast, in the PACF of figure 3b only the first lag is found to be statistically significant at a 0.95 level, and no significant lags are identified around the 26th or 52nd lag. Hence we can not conclude a statistically significant linear seasonal autoregressive process of length 52 from the ACF analysis, despite the series visual appearance. In addition, no moving average process is identified either. An augmented Dickey-Fuller unit root test confirms the stationary form of the time series; hence further differencing provides to no additional information. Consequently, the Box-Jenkins methodology does not allow valid and reliable identification of the model form of this time series, which will later be reflected in the poor performance of the MLP_{ACF} candidate models created using the input vector identified by the ACF-Analysis.

To further analyse the periodicity of the time series a spectral analysis was conducted to reveal information that the autocorrelation analysis may have missed. A variation of a periodogram shows the first 60 lags instead of the frequency along the horizontal axis in fig. 6, as the remaining lags were found to be insignificant.

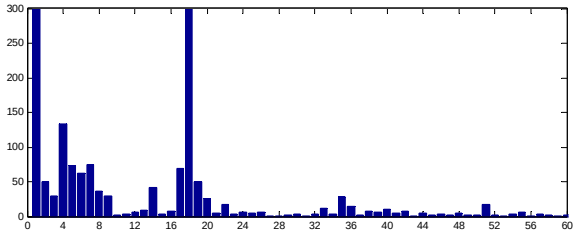

Fig. 7: Periodogram expressed in lags for the first 60 lags.

The periodigram in fig. 6. identifies the 1st and the 18th lag as highly significant, plus a set of additional lags {2, 3, 4, 6, 7, 8, 9, 14, 17, 19, 20, 22, 34, 35, 51} to be of lesser significance. Interestingly, the 52nd lag is again insignificant despite the visual appearance of a 52-observation seasonality. In contrast, lag 51 and preceding lags are found to be significant, which contradicts the analysis in the seasonal diagram in fig. 2 and the ACF and PACF cycles of 26 observations in figure 3a. As different approaches of data exploration lead to different input vectors we consider both lag-groups as candidate models MLP_{FFT} in the later evaluation to build MLP forecasts.

3.2 **Artificial Neural Network Models**

We create a NN model for each of the three candidate methods of data exploration. First, based upon the visual analysis of the seasonal diagram an input vector containing only the last seasonal lag of the dependent variable y_{t-51} is created, named MLP_{NAIVE} according to the seasonal Naïve forecasting method [16] which serves as a benchmark. In addition, two NN candidate architectures using the input vector identified by the Box-Jenkins methodology of autocorrelation analysis are created, using lags of $\{y_t, y_{t-1}, y_{t-2}\}$ named MLP_{BJ-1} and using $\{y_t, y_{t-1}, y_{t-2}, y_{t-51}\}$ named MLP_{BJ-2} including the plausible ACF information found statistically insignificant in the PACF function. Using the spectral analysis and FFT to determine input lags, three distinct candidate models were created, using additional information on the time variation of the seasonality and the outlier seasons. First, a basic MLP_{FFT} was created using an input vector of the significant lags as identified by the SA $\{y_t, y_{t-1}, \ldots, y_{t-8}, y_{t-13}, y_{t-16}, \ldots, y_{t-19}, y_{t-21}, y_{t-33}, y_{t-34}, y_{t-50}\}$. In addition, a MLP using only the highly significant lags of $\{y_t, y_{t-17}\}$ is evaluated but discarded due to significantly inferior results.

In order explicitly model the varying length of he seasonal cycles in the time series an explanatory variable is created to encode the seasonal length. Using the number of observations between consecutive minima in table 1 the relative position of each observation in each season was calculated. We divided an arbitraty number of 100 by the number of observations per season, creating a time series of {2, 4, 6, …, 100} for a season with 50 observations, {1.851, 3.703, 5.555, …, 100} for a season with 54 observations etc. Essentially, this created a temporal mapping that translated the relative position of each observation in a season of varying length onto a stationary level. A MLP_{TEMP} using the temporal encoding as an explanatory time series x_t was created, using only the explanatory variable $\{x_{t+1}\}$ in $t+1$ as an input. In addition, a topology using the lags identified from the FFT was created utilising the lags only for the dependent time series $\{y_t, y_{t-1}, \ldots, y_{t-8}, y_{t-13}, y_{t-16}, \ldots, y_{t-19}, y_{t-21}, y_{t-33}, y_{t-34}, y_{t-50}\}$ and $\{x_{t+1}\}$ for $MLP_{TEMP-FFT-1}$, using the FFT lags only on the explanatory

time series of the temporal encoding $\{x_t, x_{t-1}, \ldots, x_{t-8}, x_{t-13}, x_{t-16}, \ldots, x_{t-19}, x_{t-21}, x_{t-33}, x_{t-34}, x_{t-50}\}$ for $MLP_{TEMP\text{-}FFT\text{-}2}$, and using the FFT lags on both the time series y_t and the temporal encoding x_t for $MLP_{TEMP\text{-}FFT\text{-}3}$.

In order to eliminate the impact of the two abnormal seasonal profiles in the mid section of the time series a binary dummy variable $z_t \in \{0, 1\}$ was created with the value '1' for the two abnormal seasons and '0' otherwise. An input vector using only contemporaneous realizations of the explanatory variables for temporal encoding x_t and the time series of the binary dummies z_t was created $\{x_{t+1}, z_{t+1}\}$ for $MLP_{BIN\text{-}TEMP}$. In addition, corresponding topologies using the identified lags from FFT analysis were created for only the dependent variable y_t as $MLP_{BIN\text{-}TEMP\text{-}FFT\text{-}1}$, using the FFT lags for both time series of the dependent variable y_t and time mapping x_t as $MLP_{BIN\text{-}TEMP\text{-}FFT\text{-}2}$ and a topology using the FFT lags only for the explanatory series for time mapping as $MLP_{BIN\text{-}TEMP\text{-}FFT\text{-}3}$. Finally, we created a topology using the FFT lags on all three time series y_t, x_t and z_t as $MLP_{BIN\text{-}TEMP\text{-}FFT\text{-}4}$.

For the comparative analysis of alternative input vectors prior to the final predictions the time series was sequentially split into 60% observations for training, 20% for validation and 20% for out of sample testing. All data was linearly scaled into the interval of -0.6 to 0.6 to avoid saturation effects of the activation functions. As no indication for a MA-process that would require recurrent topologies could be determined from the ACF & PACF data analysis, we limited our evaluation to feedforward architectures of MLP. All MLPs architecture contained a single output node for iterative one-sep ahead forecasts up to 50 steps into the future, $\hat{y}_{t+1}, \hat{y}_{t+2}, ..., \hat{y}_{t+50}$. For each MLP candidate, we evaluate topologies with 1 … 20 hidden nodes in steps of 4 for a single and two hidden layers. Each network was initialised 20 times with randomised starting weights to account for local minima. It was then trained for 1000 epochs on minimising the final evaluation criteria MSE using the backpropagation algorithm with an initial learning rate of η=0.5 that was decreased by 1% every epoch. Training was terminated using early stopping if the MLP did not decrease the MSE over 0.1% in 100 epochs. A composite error of 30% training MSE and 70% validation MSE was used to avoid overfitting effects on the validation set in early stopping. We select the network topology and initialisation with the lowest composite early stopping error and evaluate its accuracy. All MLP models were calculated using the software BISlab Intelligent Forecaster (IF).

4. Experimental Results

The experimental results provided in table 2 give an overview of the criteria used to specify the input vector for the dependent variable y_t, the explanatory variable mapping temporal seasonal lengths x_t and the binary variable for outlier mapping z_t.

Although the simple approach of a seasonal naïve model MLP_{NAIVE} demonstrated adequate accuracy on validation and test data, a visual inspection of the predictions showed unsatisfactory results of extrapolating only a simple sine pattern, without replication of the outliers or the shifting seasonality. However, in accordance with established practice in forecasting a naïve approach may serve as a parsimonious benchmark to compare potential improvements of more complex model forms.

	y_t Predictor Variable	x_t Time Mapping	z_t Outlier Coding	MSE Train	MSE Valid	MSE Train &Valid	MSE Test
MLP_{NAIVE}	t-51	-	-	3.95	0.86	1.79	0.75
MLP_{BJ-1}	BJ-1	-	-	6.24	6.20	6.21	4.75
MLP_{BJ-2}	BJ-2	-	-	5.12	4.39	4.61	2.92
MLP_{FFT}	FFT	-	-	1.70	0.81	1.08	2.12
MLP_{TEMP}	-	t+1	-	2.23	0.46	0.99	0.55
$MLP_{TEMP-FFT-1}$	FFT	t+1	-	1.38	0.68	0.89	0.86
$MLP_{TEMP-FFT-2}$	-	FFT	-	1.53	0.64	0.97	0.72
$MLP_{TEMP-FFT-3}$	FFT	FFT	-	1.60	0.70	0.91	3.92
$MLP_{BIN-TEMP}$	-	t+1	t+1	1.32	0.43	0.70	0.52
$MLP_{BIN-TEMP-FFT-1}$	FFT	t+1	t+1	0.60	0.60	0.60	0.33
$MLP_{BIN-TEMP-FFT-2}$	FFT	FFT	t+1	0.44	0.55	0.51	0.68
$MLP_{BIN-TEMP-FFT-3}$	-	FFT	t+1	1.21	0.45	0.68	0.61
$MLP_{BIN-TEMP-FFT-4}$	FFT	FFT	FFT	0.41	0.70	0.55	0.59

Table 2: MLP candidate inputs and MSE on training, validation and test set

Both MLP_{BJ-1} and MLP_{BJ-2} using the Box-Jenkins methodology for input vector specification failed to approximate the shifting seasonality or the anomalies in the training set. As a consequence, the provided only a smooth sine curve with dampening magnitude on the validation and test set, leading to higher MSE then the MLP_{NAIVE} benchmark on all data subsets. Due to the nature of the varying seasonality, the specification of alternative lag structures did not increase accuracy either.

Similarly the MLP_{FFT} using SA and FFT to identify potential multiple overlying seasonalities failed to generalize on the test set, although providing significantly better results in approximating the pattern in sample. as indicated by the lower in sample errors on training and validation set. Again, the MLP were unable to capture the anomalous observations and the shifting seasonal length across different initializations and topologies, justifying a different modeling approach in providing additional information on seasonal length through explanatory variables.

The MLP_{TEMP} topologies using the temporal encoding x_t as a causal variable in t+1 reduced MSE in sample and out of sample, supporting the importance of external coding of shifting seasonal lengths. The forecasts showed a repeating sine-pattern of varying seasonality, closely resembling the observed time series frequencies. However, the MLP_{TEMP} failed to capture some subtle repetitive patterns that previous models using FFT lags had been able to approximate. In contrast, the $MLP_{TEMP-FFT-1}$ utilizing the lags identified from SA on the time series of the dependent variables $\{y_t, y_{t-1}, \ldots, y_{t-8}, y_{t-13}, y_{t-16}, \ldots, y_{t-19}, y_{t-21}, y_{t-33}, y_{t-34}, y_{t-50}\}$ plus a temporal coding showed little error improvement in comparison to MLP_{TEMP}. However, providing the FFT lags only on the temporal variable $\{x_t, x_{t-1}, \ldots, x_{t-8}, x_{t-13}, x_{t-16}, \ldots, x_{t-19}, x_{t-21}, x_{t-33}, x_{t-34}, x_{t-50}\}$ for $MLP_{TEMP-FFT-2}$ allowed a closer approximation of different periodicities and a significant increase in accuracy on the hold out data of the test set.

Despite reduced errors the seasonal anomaly observed in the time series could not be explained and negatively affected the accuracy of the approximation in sample. As only a single anomaly could be observed and no MLP model had shown the capability of approximating it as part of the data generating process, the lack of further evidence suggested an exclusion of these outliers from model building using a binary variable in addition to the previous models of temporal encoding and FFT lags. The the binary outlier variable in $MLP_{BIN-TEMP}$ enhanced the in sample approximation

and reduced the training MSE significantly. In addition, it further reduced the errors on the validation and test set in comparison to the previous topology of MLP_{TEMP}. A use of the dynamic FFT lags on the three variables *y*, *x* and *z* resulted in the selection of $MLP_{BIN\text{-}TEMP\text{-}FFT\text{-}2}$ with the lowest composite error of in sample approximation and out of sample generalization for the final forecasts. Fig. 7 illustrates the models iterative t+1, … t+50 step ahead prediction of multiple overlaying 50 period ahead forecasts originating from each point of the time series. The graph shows that the MLP has adequately learned the pattern on training and validation set, including the abnormal seasonality coded as outliers, except the last seasonal pattern also possibly containing an outlier.

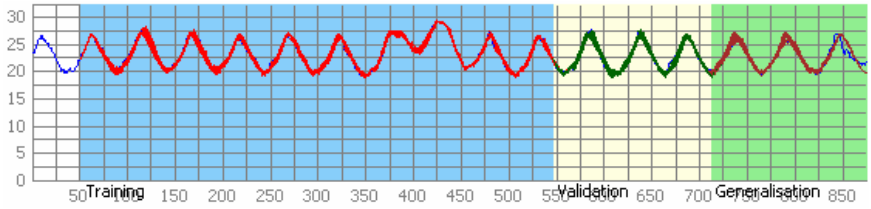

Fig. 8: 50 periods forecasts originating from each point of the time series

The selected $MLP_{BIN\text{-}TEMP\text{-}FFT\text{-}2}$ utilises the 16 lags identified by the FFT for the dependent variable y_t, the explanatory variable of temporal coding x_t and a single explanatory dummy variable to encode the outlier z_{t+1}, constructing an input vector of 37 variables. The MLP uses two hidden layers of 20 nodes each with a logistic activation function and a single output node with the identify function. The model is used to compute the final forecasts 50 steps ahead, as shown in fig. 8.

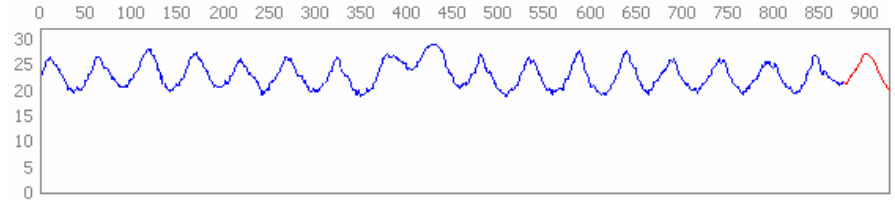

Fig. 9: The time series is plotted with the forecasts

The final ex ante forecasts required the prediction of the temporal explanatory variable beyond the provided dataset. A decision upon the position of the minimum of the last observable season and the expected length of the next seasonal cycle outside the provided data was based upon the regularity in seasonal cycle length observed in fig. 4 and the ESTSP competition objectives. Hence the next seasonal cycle was expected to be 50 observations for the final ex ante forecasts.

5. Conclusions

We evaluate a number of conventional methodologies of visual inspection, autocorrelation analysis and spectral analysis to specify significant input variables for NN prediction on the ESTSP competition time series. Due to the particular nature of

the series, containing a seasonal pattern with varying length and anomalous observations, the conventional approaches fail to specify adequate input variable lags.

To compensate for this we propose a dynamic causal modelling approach, coding the shifting seasonal cycle length and the outliers in explanatory variables, utilising the same temporal lag structure as identified in the original time series using spectral analysis. Although ACF & PACF analysis as well as SA fail to identify the input lags, they are frequently applied in NN modelling where they have a proven track record in identifying seasonal patterns of constant cycle length. Hence the results provided here should not be generalised beyond the single time series. In comparison, SA based upon FFT periodigrams demonstrate a better performance in extracting more information regarding periodic effects from this time series. However, this may again prove misleading for moving average processes which require identification in ACF-plots and subsequent modelling with recurrent NN.

Fur future research, a systematic evaluation of methodologies for input lag identification is required, extending the analysis to multiple time series, multiple time origins to increase generalisation and to unbiased error metrics, avoiding over penalisation of high deviations and outliers from squared error measures.

References

[1] G. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks: The state of the art," *International Journal of Forecasting*, vol. 14, pp. 35-62, 1998.

[2] S. Makridakis and M. Hibon, "The M3-Competition: results, conclusions and implications," *International Journal Of Forecasting*, vol. 16, pp. 451-476, 2000.

[3] T. Hill, M. O'Connor, and W. Remus, "Neural network models for time series forecasts," *Management Science*, vol. 42, pp. 1082-1092, 1996.

[4] G. E. P. Box and G. M. Jenkins, *Time series analysis: forecasting and control*. San Francisco: Holden-Day, 1970.

[5] C. M. Bishop, *Neural networks for pattern recognition*. Oxford: Oxford University Press, 1995.

[6] S. S. Haykin, *Neural networks: a comprehensive foundation*, 2nd ed. Upper Saddle River, N.J.: Prentice Hall, 1999.

[7] G. Lachtermacher and J. D. Fuller, "Backpropagation in time-series forecasting," *Journal of Forecasting*, vol. 14, pp. 381, 1995.

[8] M. Nelson, T. Hill, W. Remus, and M. O'Connor, "Time series forecasting using neural networks: Should the data be deseasonalized first?" *Journal of Forecasting*, vol. 18, pp. 359-367, 1999.

[9] G. P. Zhang and M. Qi, "Neural network forecasting for seasonal and trend time series," *European Journal Of Operational Research*, vol. 160, pp. 501-514, 2005.

[10] L. Zhou, F. Collopy, and M. Kennedy, "The Problem of Neural Networks in Business Forecasting - An Attempt to Reproduc th Hill, O'Connor and Remus Study," CWR, Cleveland 2003 2003.

[11] S. F. Crone, J. Guajardo, and R. Weber, "The impact of Data Preprocessing on Support Vector Regression and Artificial Neural Networks in Time Series Forecasting," presented at World Congress in Computational Intelligence, WCCI'06, Vancouver, Canada, 2006.

[12] S. F. Crone, J. Guajardo, and R. Weber, "A study on the ability of Support Vector Regression and Neural Networks to Forecast Basic Time Series Patterns," presented at IFIP World Congress in Computation, WCC'06, Santiago, Chile, 2006.

[13] S. F. Crone, S. Lessmann, and S. Pietsch, "An empirical Evaluation of Support Vector Regression versus Artificial Neural Networks to Forecast basic Time Series Patterns," presented at World Congress in Computational Intelligence, WCCI'06, Vancouver, Canada, 2006.

[14] Z. Y. Tang and P. A. Fishwick, "Feed-forward Neural Nets as Models for Time Series Forecasting," *ORSA Journal on Computing*, vol. 5, pp. 374-386, 1993.

[15] S. M. Kay and S. L. Marple, "Spectrum Analysis - A Modern Perspective," *Proceedings Of The Ieee*, vol. 69, pp. 1380-1419, 1981.

[16] S. G. Makridakis, S. C. Wheelwright, and R. J. Hyndman, *Forecasting: methods and applications*. New York: Wiley, 1998.