Multistep-ahead Prediction: A Comparison of Analytical and Algorithmic Approaches*

Fouad Bahrpeyma, Mark Roantree, and Andrew McCarren

Insight Centre for Data Analytics, School of Computing, Dublin City University, Dublin 9, Ireland.
Fouad.Bahrpeyma2@mail.dcu.ie, mark.roantree@dcu.ie, andrew.mccarren@dcu.ie

Abstract. Most approaches to forecasting time series data employ one-step-ahead prediction approaches. However, recently there has been focus on multi-step-ahead prediction approaches. These approaches demonstrate enhanced prediction capabilities. However, multi-step-ahead prediction increases the complexity of the prediction process in comparison to one-step-ahead approaches. Typically, studies in the examination of multi-step ahead methods have addressed issues such as the increased complexity, inaccuracy, uncertainty, and error variance on the prediction horizon, and have been deployed in various domains such as finance, economics, agriculture and hydrology. When determining which algorithm to use in a time series analyses, the approach is to analyze the series for numerous characteristics and features, such as heteroscedasticity, auto-correlation, seasonality and stationarity. In this work, a comparative analysis of 20 different time series datasets is presented and a demonstration of the complexity in deciding which approach to use is given. The study investigates some of the main prediction approaches such as ARIMA (Autoregressive integrated moving average), NN (Neural Network), RNN (Recurrent neural network) and SVR (Support vector regression), which focus on the recursive prediction strategy and compare them to a new approach known as MRFA (Multi-Resolution Forecast Aggregation).

1 Introduction

A time series data-set can typically be described as a series of values or quantities that are ordered in a time sequenced fashion, often with equal intervals between them [4]. In the past time series analysis was typically, but not exclusively applied to traditional econometric problems [15]. However, with the advent of Big data sources such as sensor technology, temporal and streaming data, one can certainly consider these data sources as time series problems. Generally, when predicting forward using such data, one can either implement one-step ahead prediction (OSAP) or multi-step-ahead prediction (MSAP) approaches

* This work is supported by Science Foundation Ireland under grant number SFI/12/RC/2289

[3]. Both approaches have advantages and disadvantages when making predictions. OSAP methods typically suffer from expanding error variance when going beyond the short term prediction horizon, while MSAP methods can be difficult to construct, as one is using complex RNN (Recurrent neural network) or LSTM (Long short term memory) algorithms, and determining an appropriate number of hidden nodes and layers is less than trivial. Additionally, each time series has hidden attributes such as heteroscedasticity, auto-correlation, seasonality and stationarity. These attributes can be measured using various metrics that have been constructed over the years [5, 21]. Certain fields of interest such as the Agri or Finance sectors generate large amounts of univariate time series data. Traditionally, one would consider these metrics when applying analytical methods such as ARIMA or ARCH (Auto regressive conditional heteroscedasticity). However, when choosing modern machine learning applications in time series prediction problems these characteristics are rarely considered.

Motivation. When determining which approach to use in a time series prediction problem, one is often faced with the significant task of selecting one method/algorithm from a large selection of algorithms that can be derived for example, from the ARIMA, SVR, NN, RNN and LSTM families. By creating a list of defined criteria that characterize a time series one can compare a new time series with work previously carried out using the meta characteristics that were generated in the historical analysis. This then allows researchers to choose an algorithm that predicts more accurately in the required prediction horizon, and with minimum computational and research effort.

Contribution. In this paper, an extensive set of time series metrics and a wide portfolio of OSAP and MSAP methods on disparate data streams and sources are implemented to guide the analyst in choosing the most appropriate algorithm. A detailed analysis and classification of each approach is given, which in turn, provides a meta analysis for each series.

Paper Structure. The paper is structured as follows: in Section 2, we discuss related research in time series analysis; In Section 3, a description of the time series characteristic metrics are outlined; Section 4 describes the classification evaluation procedure and provides an analysis of the disparate data sources; and finally, in Section 5, we present our conclusions.

2 Related Research

Various measures are implemented in time series analytical approaches [8]. Assessing the characteristics of a time series have prevalently focused on measuring the series for auto-correlation, seasonality, hetroscedasticity, non-linearity, volatility and non-stationarity [15]. In this paper we intend to extend the analyses with a few more measures to conduct a study on their presence with respect to the performance of time series prediction.

Time series prediction has long been practiced using the ARIMA family. The ARIMA family fundamentally relies on two main components, an Autoregressive (AR) model and a Moving Average (MA) model [4]. From the machine learning point of view, time series prediction problems are fundamentally sequential supervised learning problems, and need to be converted to classical supervised learning, in order to solve them using machine learning techniques. NNs are the most popular ML technique used for prediction purposes and were inspired from the human brain's neural inference system [6]. RNNs by incorporating feedback connections extend the memory of NNs with improved performance in particular for time series prediction [18]. LSTM is an extension of RNNs with an enhanced ability in capturing long term memory from the given time series [12]. SVR is the extension of the well known SVM classifiers used for regression purposes [9].

Multi-step ahead prediction has predominantly been implemented using the principles of either the direct or the recursive strategies. The main disadvantage of the direct strategy is that serial correlations are discarded while the recursive strategy suffers from accumulation of errors. The recursive strategy has generally been seen as the more popular method as it attempts to use the inherent memory of a time series by using its natural serial correlation. Thus, in this research we focus on the recursive strategy. In the recursive strategy [19], an OSAP model (which can be a machine learning regression model) is run multiple times over the given prediction horizon. For a signal y, the output of the OSAP model $f(.)$ is calculated by Eq. 1:

$$y_{t+1} = f(y_t, ..., y_{t-d+1}) \tag{1}$$

Where d is the number of input lags. In Eq. 1, $f(.)$ can be modeled using NN, RNN, LSTM or SVR as the OSAP method to implement the recursive strategy. There is also another method called Multi-Resolution Forecast Aggregation (MRFA) [2] that follows the principles of the recursive strategy to implement MSAP but not exactly through Eq. 1. MRFA incorporates a concept known as the Resolution of Impact (ROI) and analyses the impact of local patterns on the future of the signal; resolution refers to the length of the horizon for which this impact is studied. Using MRFA, the value of the signal in the desired prediction horizon is predicted at multiple resolutions using separate forecast models. The predicted values are then aggregated to gain the final forecast.

3 Time series characteristics and metrics

In traditional econometric modelling, an analysis to determine the characteristics of the time series is undertaken. This will generally help the analyst in determining which algorithm or analytical tool to use. However, most studies only rely on a few basic characteristics such as stationarity and auto-correlation, which cannot sufficiently describe time series properties. In this section, some prediction metrics that help determine the most appropriate algorithm are presented in an attempt to provide a more informative analysis.

3.1 Gaussianity & Sampling frequency

Gaussianity is a measure of how similar the time series (or the fluctuations/error) is to a normal distribution, and can be measured via the D'Agostino's K-squared test [10]. The frequency at which the system is sampled contributes to the complexity of the time series. Typically, the complexity of a low frequency signal is higher than that of high frequency one. In a low frequency signal, each sample is predominantly an aggregation over a period of time. This aggregation, in particular, tends to increase the Gaussianity of the data, which consequently enhances the complexity of the modeling process [7]. In contrast, in high frequencies, noise or temporal anomalies appear more salient in the formation of the signal. Various methods will show differing performances when dealing with such noise effects [5].

3.2 Trend & Seasonality

Trend is a pattern in a time series which can be caused by high level factors such as socio-economic or political forces. Depending on the data, the contribution of trend to a time series can be additive or multiplicative [4]. Seasonality occurs as regular patterns or fluctuations in a time series, [5], and can be caused by the influence of factors such as season, month, or day of the week [4]. After eliminating trend and seasonality in the time series, the remaining components are usually referred to as white noise or irregular component [8].
Like trend, seasonality may be additive or multiplicative:

$$\begin{cases} I_t + T_t + S_t \rightarrow additive, \\ I_t \times T_t \times S_t \rightarrow Multiplicative. \end{cases} \tag{2}$$

where y_t, I_t, T_t, and S_t are the original time series, irregular component, trend, and seasonality, respectively. One approach to detect seasonality is to analyze sample auto-correlation function (ACF) and partial auto-correlation function plots (PACF) [1].

3.3 ACF and PACF plots

Sample auto-correlation function (ACF) and partial auto-correlation function (PACF) help provide metrics in assessing the level of auto-correlation, seasonality and to a lesser extent the levels of non-stationarity. The ACF plot represents lagged correlations in a time series in terms of correlation coefficients. For the time series y of length T, the ACF is computed as [1]:

$$r_T = \frac{c_\tau}{c_0}, \tau = 0, ..., T-1 \tag{3}$$

where c_0 is the sample variance, and is the empirical auto-covariance at lag τ:

$$c_\tau = \frac{1}{T} \sum_{t=\tau}^{T-1} (y_t - \bar{y})(y_{t-T} - \bar{y}) \tag{4}$$

The PACF plot illustrates the partial correlation coefficients between the given time series and its lags. PACF is the estimated lag-h coefficient in an AR model containing h lags adjusting for the lags between time point 1 and lag h [1], i.e.:

$$\begin{cases} \phi_{1,1}(y) = corr(y_{t+1}, y_t) = \rho(1) \\ \phi_{h,h}(y) = corr(y_{t+h} - \hat{y}_{t+h}, y_t - \hat{y}_t); h \geq 2 \end{cases}, \tag{5}$$

where corr(x,y) is the correlation function, defined as:

$$corr(x,y) = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)}\sqrt{n\sum y_i^2 - (\sum y_i)}} \tag{6}$$

3.4 Stationarity

Stationarity is a favourable feature in time series data and can be described as having a steady mean, variance and auto-correlation over the time span of the series [1]. As a result, some mathematical transformations have been used to transform time series into a stationary data-set. Differencing is the most common approach used to reduce the effects of non-stationarity, [4]. However, real world time series tend to exhibit non-stationarity even after differencing. In such cases, other data pre-processing techniques such as de-trending or de-seasonalizing may be required, [8]. Non-stationarity can be identified using unit-root tests. Some popular unit-root tests are the Augmented Dickey–Fuller (ADF) test [11] and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test [17].

3.5 Conditional Heteroskedasticity

Heteroscedasticity refers to the presence of non-constant variance in a time series over time, which can be conditional or unconditional. The phenomenon is challenging in the unconditional form, where the time series is characterized by non-constant volatility such that future periods of high and low volatility cannot be identified [1]. The Autoregressive conditional heteroscedasticity (ARCH) family can be used to model such phenomena when conditional hetroscedasticity (CH) exists. CH can also be evaluated by applying Ljung–Box Q test to the squared residual series [16].

3.6 Long range dependence and Sample entropy

Long-range dependence (LRD) or long term memory is a feature that is associated with the predictability of a time series. LRD measures the statistical dependence between two differing time points. LRD is present if the statistical dependence declines at a slower than an exponential rate of decay. The Hurst exponent [13] and the Detrended Fluctuation Analysis (DFA) coefficients [14], for example, are two metrics that evaluate a time series for long range dependence. Sample entropy is a metric to assess the complexity of time-series data. For a signal y with sample size N and tolerance r, sample entropy is the negative logarithm of the conditional probability that a sub-series of length m matches point-wise with the next point with tolerance (distance less than) r [20].

4 Evaluation

20 different time series were chosen for analysis in this paper and were taken from disparate monthly, weekly, daily and hourly data sources. The monthly series were lake Erie levels (1921-1970), monthly milk production pounds or `m_m_p` (1962-1974), and the number of employed persons in Australia (1978-1991). The weekly time series were two Irish beef prices (2011-2018), Irish pig prices, 2007-2016, German pig prices (2008-2016), Canadian barley (2008-2016), and German feed barley (2008-2016). The daily time series are foreign exchange rates (1979-1998), minimum temperatures in Melbourne (1981-1990), total female births in California (1959-1959), and mean temperature of the fisher River (1988-1991), bike share variables (2011-2012) and births in USA between 1978 and 2015. Hourly data was taken from one series measuring hourly carbon dioxide emissions.

In order to characterize each dataset, differing tests and metrics were applied. ACF and PACF plots were analyzed to confirm the existence of seasonality. CH was examined using the Ljung-Box Q test on the squared residual series. Also, the Hurst exponent and DFA were conducted on the data to characterize non-stationarity in the series. Table 1 shows these characteristics as well as the frequency, sample size and type/status of non-stationarity for each of the original time series. Gaussianity is evaluated using D'Agostino's K^2 test and denoted by H for high, L for low, V for very low, and 0 for no Gaussianity.

In Table 1 which shows the results for 8-steps-ahead prediction, m_l_e_l, m_m_p, and m_n_o_e indicate lake Erie water levels, monthly milk production pounds, and the level of employment in Australia, respectively (Monthly data). Also, H_Um_3P and S_Um_2P indicate two Irish beef prices, IreCent is the Irish pig price, GerCent is the German pig prices, CanBar is Canadian barley prices and GerFBar is the German feed barley prices (weekly data). Daily foreign exchange rates are denoted by d_f_ex_r, daily minimum temperatures in Melbourne by d_m_t, daily total female births in California by d_t_f_b, mean daily temperature of fisher River near Dallas by m_d_t_f, daily bike share variables by d_b_sh1, d_b_sh2 and d_b_sh4, a US economic series by Ec_unem, births in US in 1978 and 2015 by US_B_78 and US_B_15, (Daily data) and hourly carbon dioxide emission by LNOxEm.

4.1 Results

In Table. 1, for each series the method with the minimum prediction error is reported as the preferred method. The prediction error is calculated as the average normalised mean square error. For each series the order of the $SARIMA(p, d, q)$ model was obtained using the exact maximum likelihood using a Kalman filter. A detailed discussion on optimal models is beyond the scope of this paper. However, details on model parameters selection can be found in [23, 22]. The experimental results based the application of the methods discussed are outlined in Table 1.

Table 1. Experimental results

Frequency	Time series	Gaussianity	Sample size	Seasonality	Stationarity	CH	Hurst exponent	DFA	Sample entropy	Type	Min error
M	m_l_e_l	L	600	12	N	Y	0.59	1.17	0.88	Non-stationary	MRFA
M	m_m_p	L	156	12	N	N	0.69	1.52	0.68	Brownian noise	MRFA
M	m_n_o_e	V	759	12	N	Y	0.97	1.28	0.29	Non-stationary	NN
W	H_Um_3P	H	357	52	N	Y	0.93	1.79	0.41	Non-stationary	MRFA
W	S_Um_2P	L	357	52	N	Y	0.92	1.68	0.60	Non-stationary	SVR
W	IreCent	V	512	52	N	Y	0.92	1.86	0.32	Non-stationary	MRFA
W	GerCent	V	468	52	N	Y	0.86	1.58	0.53	Non-stationary	NN
W	CanBar	V	468	52	N	N	0.88	1.64	0.20	Non-stationary	LSTM
W	GerFBar	V	468	52	N	N	0.86	1.63	0.25	Non-stationary	NN
D	d_f_ex_r	V	4774	365	N	Y	0.95	1.52	0.05	Brownian noise	SVR
D	d_m_t	V	3650	365	Y	N	0.90	1.08	1.62	Pink noise	ARIMA
D	d_t_f_b	V	365	-	Y	Y	0.61	0.69	2.20	Stationary	NN
D	m_d_t_f	V	1461	365	N	N	0.88	1.24	0.73	Non-stationary	MRFA
D	d_b_sh4	V	731	-	N	Y	0.51	0.66	2.05	White noise	ARIMA
D	d_b_sh2	V	731	-	N	N	0.72	1.05	0.88	Pink noise	MRFA
D	d_b_sh1	V	731	-	N	N	0.77	1.07	0.87	Pink noise	MRFA
D	Ec_unem	V	574	-	N	N	1.00	1.58	0.23	Brownian noise	MRFA
D	US_B_15	V	365	-	N	N	0.09	0.08	0.65	RNN	RNN
D	US_B_78	V	365	-	N	N	0.23	0.29	1.21	Anti-correlated	LSTM
H	LNOxEm	0	8081	24	Y	N	0.39	0.37	0.45	Anti-correlated	ARIMA

Sample entropy of values close to zero indicated high levels of self-similarities, and thus higher probability of predictability. For US_B_78, d_b_sh4, d_t_f_b, and d_m_t the sample entropies were relatively high and the best performance were exhibited by ARIMA, NN and LSTM. It can be seen that best performance for Brownian noise was MRFA (two series) and SVR (one series), and for pink noise series, MRFA (two series) and ARIMA (one series).

However, for the Brownian noise, SVR had the best performance on d_f_ex_r where the sample entropy was extremely small, implying that the series was highly self-similar and predictable. Also, for the Brownian noise, ARIMA came out best on d_m_t where its sample entropy was high, meaning that the complexity is high and thus the conclusion in this case is unreliable. It can be seen that ARIMA and NN outperformed the other methods when applied to stationary time series. It suggests that ARIMA and NN which are characterized by relatively less complex structures (compared to RNN, LSTM and MRFA) are predominantly suitable for series which have a steady mean, variance and auto-correlation. However, for two out of the three stationary time series, the sample entropy was high, suggesting that stationarity can be as important as self-similarity in predictability of time series data. The results show that when Gaussianity is present MRFA was the preferred option.

For series characterized by CH, NN (m_n_o_e, GerCent, and d_t_f_b), MRFA (m_l_e_l, H_Um_3P and IreCent), SVR (S_Um_2P and d_f_ex_r) and ARIMA (d_b_sh4) preformed best. However, the series on which ARIMA was the preferred candidate, was characterized by substantial degree of White noise (Hurst exponent $\simeq 0.5$).

4.2 Analysis

It is apparent that differing methods performance are a function of the characteristics of the time series in question. MRFA does seem to posses more robust characteristics as it is either the preferred method or was one of the high performers on the majority of the series examined with a preferred candidate score of 50%. In particular it was the preferred method on 80% of the series when applied to series with either Brownian or pink noise and scored 44% when applied to data with non-stationary. MRFA, NN and LSTM all appeared to perform well when non-stationarity or conditional hetroscedasticity existed within the data. Additionally, ARIMA also performed well with non hetroscedastic data, and is much easier method to implement in comparison to machine learning approaches such as MRFA, NN, RNN or LSTM. Fig 1 outlines the performance of all the methods attempted.

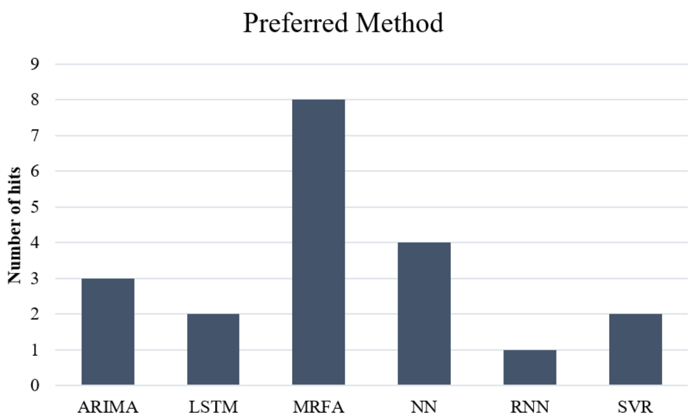

Fig. 1. Method versus preferred method count

5 Conclusions

This paper proposes an analysis that will assist researchers in the selection of the most appropriate prediction model based on historical meta analysis of time series characteristics. A number of important characteristics were analyzed including frequency, sample size, Gaussianity, long-term memory (Hurst exponent and Detrended Fluctuation Analysis), non-stationarity, and conditional heteroscedasticity. Experimental results were obtained on 20 series with varying characteristics including anti-correlated series, pink noise, Brownian noise, stationary and non-stationary series. The evaluation demonstrates that the performance of a prediction method, in particular machine learning techniques, is a function of the characteristics of the given time series. Using a meta analysis of historical time series analysis as a basis for the selection of a time series approach will reduce the effort required within each specific application. Additionally, increasing the size of the Meta analysis database substantially would have obvious benefits to researchers in areas such as the Agri or Finance sectors.

References

1. James Aweya. *Sensitivity Methods for Congestion Control in Computer Networks.* PhD thesis, Ottawa, Ont., Canada, Canada, 1999. AAINQ48085.
2. Fouad Bahrpeyma, Mark Roantree, and Andrew McCarren. Multi-resolution forecast aggregation for time series in agri datasets. In *Proceedings of the 25th Irish Conference on Artificial Intelligence and Cognitive Science, Dublin, Ireland, December 7 - 8, 2017.*, pages 193–205, 2017.
3. Gianluca Bontempi, Souhaib Ben Taieb, and Yann-Aël Le Borgne. *Machine learning strategies for time series forecasting*, pages 62–77. Springer, 2013.
4. George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control.* John Wiley and Sons, 2015.
5. Peter J Brockwell and Richard A Davis. *Introduction to time series and forecasting.* springer, 2016.
6. Antony Browne. *Neural network analysis, architectures and applications.* CRC Press, 1997.
7. Ana Christina Cadavid, John K Lawrence, and Alexander Ruzmaikin. Principal components and independent component analysis of solar and space data. In *Solar Image Analysis and Visualization*, pages 37–51. Springer, 2007.
8. Chris Chatfield. *The analysis of time series: an introduction.* CRC press, 2016.
9. Haibo Chu, Jiahua Wei, Tiejian Li, and Kun Jia. Application of support vector regression for mid-and long-term runoff forecasting in "yellow river headwater" region. *Procedia Engineering*, 154:1251–1257, 2016.
10. Gregory W Corder and Dale I Foreman. *Nonparametric statistics: A step-by-step approach.* John Wiley & Sons, 2014.
11. David A Dickey and Wayne A Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431, 1979.
12. Thomas Fischer and Christopher Krauss. Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 2017.

13. Harold Edwin Hurst. Long term storage capacity of reservoirs. *ASCE Transactions*, 116(776):770–808, 1951.
14. Jan W Kantelhardt, Eva Koscielny-Bunde, Henio HA Rego, Shlomo Havlin, and Armin Bunde. Detecting long-range correlations with detrended fluctuation analysis. *Physica A: Statistical Mechanics and its Applications*, 295(3-4):441–454, 2001.
15. Evžen Kočenda and Alexandr Černỳ. *Elements of time series econometrics: An applied approach.* Charles University in Prague, Karolinum Press, 2015.
16. Evžen Kočenda and Alexandr Černý. *Elements of time series econometrics: An applied approach.* Charles University in Prague, Karolinum Press, 2015.
17. Denis Kwiatkowski, Peter CB Phillips, Peter Schmidt, and Yongcheol Shin. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, 54(1-3):159–178, 1992.
18. Danilo P Mandic and Jonathon A Chambers. *Recurrent neural networks for prediction: learning algorithms, architectures and stability.* Wiley Online Library, 2001.
19. Alexander G Parlos, Omar T Rais, and Amir F Atiya. Multi-step-ahead prediction using dynamic recurrent neural networks. *Neural networks*, 13(7):765–786, 2000.
20. Joshua S Richman, Douglas E Lake, and J Randall Moorman. Sample entropy. In *Methods in enzymology*, volume 384, pages 172–184. Elsevier, 2004.
21. Abdol S Soofi and Liangyue Cao. *Modelling and forecasting financial data: techniques of nonlinear dynamics*, volume 2. Springer Science and Business Media, 2012.
22. Wei-li Xiong and Bao-guo Xu. Study on optimization of svr parameters selection based on pso [j]. *Journal of System Simulation*, 9:017, 2006.
23. Gioqinang Zhang and Michael Y Hu. Neural network forecasting of the british pound/us dollar exchange rate. *Omega*, 26(4):495–506, 1998.