UCL
DEPARTMENT OF CHEMICAL ENGINEERING

CENG0038 Chemical Engineering Research Project

SYSTEMS-BASED APPROACHES FOR TIME SERIES FORECASTING

by

Riyad HASAN

Student's registration number: 17014126

Supervisor(s):	Dr V Charitopoulos
PDRA/PhD contact	N/A

A thesis submitted to the University of London for the degree of MEng Engineering (Chemical)

Department of Chemical Engineering

University College London

January 2021

Declaration

I, Riyad Hasan, confirm that the work presented in this thesis is my own.

Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Word count: 9,999

(excluding the title page, table of contents, references, tables, figures and appendices)

Has the written report been submitted on Moodle?	Yes
Have relevant source codes and/or raw data been submitted on Moodle?	Yes
Has the lab space used been cleaned up (if applicable)?	N/A

Riyad Hasan		25/01/21
Student’s Name	Student’s Signature	Date

Abstract

New forecasting methods such as support vector regression (SVR) are showing promise in various research but have yet to displace traditional methods that are used in industry such as the autoregressive integrated moving average (ARIMA) model. This paper aims to investigate the strengths and weaknesses of SVR and test its viability for customer demand and wind time series forecasting versus ARIMA. An SVR algorithm was developed for GAMS utilising linear and nonlinear programming followed by a recursive regression function to calculate the forecast, whilst python code was used for an ARIMA model with automated parameter optimisation. The SVR model was optimised in terms of its parameters and kernel functions, proving to be very successful by achieving a MAPE of 9.77% in customer demand forecasting using a polynomial kernel function, outperforming ARIMA. However, all models showed weaker performance for volatile wind data, exemplifying the importance of training data for time series forecasting.

Acknowledgments

I would like to extend a special thanks and gratitude to my supervisor Dr Vasileios Charitopoulos for the invaluable guidance and support throughout the entirety of this project. I am truly grateful to have been provided with this opportunity to learn from you and to produce this paper.

I also want to thank my research partner, Mohammed. It was truly rewarding to work alongside you in this challenging but fun experience.

Finally, I could not have completed this project without the love and support from my friends, Rafiz, Mohammed, Eunsung, Azmi, Radi, Shaks; and my family. Thank you for keeping me sane and upbeat during this project and these recent times, and a special thanks to my parents for supporting and motivating me as far I can remember. Love you guys.

Nomenclature

All nomenclature used in this paper is listed below, unless specified separately.

Abbreviation	Description
AIC	Akaike information constant
MAPE	Mean absolute percentage error
RMSE	Root mean squared error

Symbol	Description
c	Constant for kernel functions
C	Regularisation hyperparameter for SVR
k	Constant for ε formula for SVR
K	Kernel function
L	Lagrangian function
m	Periodicity for ARIMA model
N	Number of training data points for SVR
p	Polynomial degree for polynomial kernel function
(p, d, q)	Nonseasonal ARIMA parameters
(P, D, Q)	Seasonal ARIMA parameters
t	Time
$\boldsymbol{w}$	Slope vector for SVR
$\boldsymbol{x_t}$	Input vector for SVR
y_t	Customer demand
$y_{t,actual}$	Actual customer demand
$\bar{y}_t$	Mean customer demand
z	Attributes of the input vector

Greek Symbols	Description
β	Parameter for SVR regression function
ε	Width of tube / error tolerance for SVR dual problem
λ_t, λ_t^*	Lagrange multipliers / dual variables for SVR dual problem
μ_t, μ_t^*	Lagrange multipliers for SVR dual problem
ξ_t, ξ_t^*	Slack variables for SVR dual problem
σ	Standard deviation

Contents

1. Introduction

Forecasting is a process where past and present data are analysed to make informed predictions of future trends. It aids companies in decision-making and planning regarding sales, production, pricing and demand. It is also used to promote renewables, being used in solar and wind data predictions[1]. Recently, the rise in privatisation and competition in conventionally controlled markets has led to an increased reliance on customer demand as a metric to drive growth. By quantifying and understanding customer demand, companies from various industries can minimise uncertainty in many stages including production and anticipate/react to market changes and volatility. Indeed, overproduction or underproduction can incur massive financial losses, meanwhile successfully navigating the market and responding effectively can net equally massive financial gains. Furthermore, with the need to transition from fossil fuels to renewables due to climate change, planning for solar and wind energy capacities is vital. Therefore, forecasting, including customer demand and wind forecasting, has emerged as an important tool for companies in chemical engineering industries.

Traditional forecasting approaches include statistical models such as exponential smoothing (ES) and autoregressive integrated moving average (ARIMA)[2].

ES was developed in the late 1950s as a time series forecasting technique for univariate data whereby the prediction of the forecast is calculated from an exponentially weighted moving average of the past observations from the dataset[3]. Thus, the older the data, the less weight it has in the prediction. ES has derivatives/variants to account for trends and seasonality, such as Holt-Winters triple ES[4].

The Box-Jenkins method applies ARIMA models for time series forecasting. This was developed in the 1960s and combined autoregressive and moving average components with differencing[5,6].

In the 1990s, with the rise in available processing power, artificial neural networks (ANNs) were developed. The constituent mathematical models of ANN are inspired by the biological neural networks of the human brain and are more complex than prior forecasting approaches, being able to capture nonlinear, complex behaviour that conventional methods could not. However, the increased complexity is oftentimes problematic for already complicated applications such as chemical or process engineering due to the number of variables that can be considered.

Support vector regression (SVR) is another recent model developed from Vapnik's support vector machine (SVM) framework[7]. SVR avoids ANN's complexity as comparatively less parameters are needed, whilst retaining a high degree of flexibility for use in various applications provided sufficient data is available.

The motivation of this paper is to review, test and evaluate the performance of SVR, a newer model, and ARIMA, a conventional model, with comparative analysis to determine their viability in two applications: customer demand and wind capacity forecasting. The paper includes a literature review which will examine the state-of-the-art, followed by methodologies and problem formulation, finalised with a presentation and evaluation of results with recommendations for future work.

2. Project Objectives

This paper will aim to build upon the expansive research in customer demand and wind forecasting, focussing on the use of SVR, a newer method, and ARIMA, a traditional method, to forecast two different time series.

First, a literature review will be conducted to evaluate various state-of-the-art and traditional forecasting models, with a focus on chemical engineering, wind data and customer demand applications.

Since SVR does not yet have a formal framework for implementation, an approach for SVR to be used on GAMS will be developed, to expand on SVR research. An ARIMA model made using python code with pre-packaged libraries will also be tested for comparative purposes to evaluate the performance of a state-of-the-art model versus a conventional model.

The two datasets being considered are a customer demand series with a clear trend and seasonality patterns, and a wind capacity series which is more volatile with no distinct trends or patterns. This will ensure that the forecasts are for relevant fields to chemical engineering whilst also testing the versatility and ability of the models in two different conditions.

The results will be interpreted and discussed, and conclusions will be made to determine if SVR is suitable for industrial applications or if mature techniques are still preferable.

The scope of the study is centred on SVR and ARIMA in two relevant time series. While there are various other models and time series, which shall be covered in the literature review, they are extensive and can be revisited in future studies. This scope will still allow for a comprehensive and useful analysis of two important forecasting models in two relevant fields of study.

3. Literature Review

In this section, various forecasting techniques, primarily statistical and computer intelligence models, will be covered and their suitability for forecasting will be discussed, along with their application in relevant scenarios.

Forecasting has been expansively researched in literature. The bulk of research is on conventional models such as ES and ARIMA, since they are the most developed and traditional models. This is followed by relatively newer SVR and ANNs and research into novel, specialised hybrid models. Due to its usefulness, forecasting has numerous research applications including customer demand, solar and wind data, weather, electricity prices, population growth and more. As mentioned in the project objectives and scope, this review will focus primarily on sales/demand and wind data as they are two pertinent examples for chemical engineering.

Firstly, ES, a traditional and established method, will be discussed.

3.1 Exponential Smoothing

ES specifically works by having an exponentially weighted moving average of historical data, resulting in older observations having less weight. The simple form of ES is given by the following equation[6]:

$$s_t = \alpha x_t + (1-\alpha)s_{t-1} \tag{3.1.1}$$

Where s_t is the smoothed value and a weighted average of x_t, the current observation, and s_{t-1}, the previous smoothed value. The weighting is controlled by the parameter α for $0 \leq \alpha \leq 1$, where the larger the value of α, the less weighting past observations have. Derivatives of ES include sophisticated models that consider trend and seasonality components, such as Gardner's smoothing, Brown's quadratic smoothing and others[8].

Generally, results from literature showed that ES forecasting models have merit but are limited in their applications. Studies by Schnaars and Makridakis et al. showed that these models, in particular single ES, struggle with longer time horizons around 5 years, but are acceptable for one year forecasts[9]. This is expected as these models do not consider trend. Indeed, the longer the tome horizon, the worse the predictions of simple models. However, models such as Holt or Holt-Winters, which are more sophisticated and complex models that consider seasonality and trends, perform expectedly better for longer time horizons than simpler models according to the analysis of results of a forecasting competition by Makridakis[10].

An exception is the THETA model, developed by Assimakopoulos and Nikolopoulos[11]. Hyndman and Billah, and Makridakis and Hibon's analysis of the M3-competition showed that THETA maintained its excellent performance consistently for all time horizons (monthly to yearly) and accuracy measures[12,13]. However, THETA was not tested for hourly or daily data and as such it remains to be seen whether it would be viable for short-term forecasts[3]. Indeed, ES is generally unpopular for short-term forecasting. In their review, Weron found only one article using ES for short-term electricity price forecasting[3].

Another observation is that the performance of ES depends greatly on the type of data. Makridakis et al. show that simple methods perform better with micro data than the sophisticated, complex models while it is the opposite for macro data[10]. Moreover, noisy data, meaning data with randomness, can affect the performance of sophisticated models, where more noise diminishes the impact of sophisticated models versus simple models[10]. This is amplified for short-term data (hourly-daily) where there is greater significance/impact of randomness versus longer term horizons such as quarterly or yearly data. Therefore, for customer demand and electricity price forecasting, where hourly and daily forecasts are important, sophisticated methods will not necessarily outperform simple methods. Indeed, an article by Cruz et al. showed that double seasonal ES outperformed the more sophisticated ARIMA, however both were worse than ANN and dynamic regression methods[1].

The aforementioned issue with noise, along with other factors, demonstrate that a sophisticated model does not necessarily outperform or provide a more accurate forecast than a simpler one, a conclusion shared by Makridakis and Hibon's analysis of the M3-competition[13]. Reviews including Schnaars and Gardner's, show that combination methods, meaning the arithmetic average of multiple methods, outperform the individual methods that were combined. Makridakis and Hibon show that a combination of Single, Holt and Dampened ES obtains a higher prediction accuracy (PA) for all tested time horizons than the individual constituent methods, the combination had a mean absolute percentage error (MAPE) of 8.8% while the individual methods achieved 9.5%, 9% and 8.9% respectively[13].

Therefore, the forecasting accuracy of ES depends greatly on a multitude of factors including the time horizon, type of data and series and whether a simple or sophisticated model is employed. As such, any company who plan to use ES must ensure that they select the specific model that would maximise the forecasting accuracy for their time series. This degree of variability is contrary to recent methods such as SVR and ANN since these are driven by training data and as such excel for a variety of applications, so long as sufficient past data is available.

3.2 ARIMA

Likewise with ES, ARIMA models are a developed and popular approach for time series forecasting[6]. ARIMA models are based on ARMA (autoregressive moving average) models, the latter of which makes the assumption that the time series is stationary i.e. a series which does not depend on the time it is observed[3]. As such, the ARMA model was limited in its application as it would not be applicable for non-stationary time series[3,5]. Box and Jenkins pioneered using differencing, a method to transform a non-stationary time series to stationary by eliminating trends and seasonality, from which the ARIMA model was formulated[5].

The ARIMA model is a combination of an autoregressive model, moving average model and differencing, hence the common notation of ARIMA(p,d,q) where p is order of autoregressive component, q is order of moving average component, d is degree of differencing. The non-seasonal model is described by Eq. 3.2.1, obtained from section 8.4 of *Forecasting: Principles and Practice* by Athanasopoulos and Hyndman[6].

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{3.2.1}$$

Where y'_t is the differenced series, ε_t is noise, and ϕ, θ are assigned coefficients.

The (p,d,q) can have different values hence the non-seasonal model is described as the ARIMA(p,d,q) model.

In cases where differencing is not sufficient to make a series stationary, ARIMA has a seasonal variant, known as SARIMA[3]. This is done by adding a seasonal component called $(P,D,Q)_m$ to the ARIMA model, where m is the number of observations in the seasonal pattern, or the periodicity[3,6].

ARIMA has seen numerous applications in research including forecasting of customer demand and electricity prices[14].

A case study by Permatasari et al. showed that ARIMA(1,1,0) achieved a MAPE of 3.52% in a newspaper sales forecast[15]. Mircetic et al. found that SARIMA models were successful in forecasting demand for a beverage company, with the best model being $\text{SARIMA}(5,0,1)(1,0,0)_{52}$ achieving a 14.18% MAPE[16]. Significantly, the best models included autoregressive (p,P) and moving average (q,Q) parts but no differencing or seasonal components showing their series had a low degree of seasonality[16]. Indeed, the models with these components performed worse than the basic naïve method[16]. This demonstrates that complex models are not necessarily better, and the benefits of the additional complexities are only realised for appropriate time series. Indeed, a study by Lagarto et al found that ARIMA models outperformed naïve method and other methods in Iberian electricity price forecasting[17]. This strengthens the observation that there is no best model, rather they can excel or struggle depending on the data and its properties.

A notable finding was that ARIMA models where each hour is modelled in a separate forecast showed better performance than a SARIMA model which was specified for the entire time period[18,19]. A similar conclusion is reported by Cuaresma et al. whereby ARMA and ARIMA models showed superior performance for electricity price forecasting when each hour in the time series was modelled separately, ahead of modelling the whole time series[20]. This suggests that standard AR-type models perform better for shorter forecasting horizons than their seasonal counterparts.

Hybrid models have also been used. Che and Wang showed that SVRARIMA and NNARIMA had the lowest error criterions in electricity price forecasting, outperforming singular models[21]. This increased performance is due to the hybrids' ability to distinguish linear and non-linear patterns, which singular models cannot do as effectively[21].

AR time series (ARIMA, SARIMA etc.) do not consider exogenous factors, rather they focus solely on past data. However, this is limited for series where exogenous factors are important. For example, weather conditions are an exogenous factor that can influence customer demand or electricity price[3,22]. ARIMAX and SARIMAX are generalisations of ARIMA and SARIMA respectively that consider exogenous factors[3,23].

In two studies by Nogales et al. the ARX-type model has greater capability for electricity price forecasting than traditional ARIMA, with an average weekly MAPE of 3% in California and 5% in Spain in their 2003 study[19,24]. SARIMAX was used in electric load forecasting with a root mean squared error (RMSE) of 1.32-2.80%.[23] In a PV-generation study, the consideration of exogenous factors meant SARIMAX(3,1,2)(3,1,2)$_{24}$ was the best model beating SARIMA and ANNs with an RMSE of 10.93%[25]. A food sales forecast demonstrated SARIMAX outperformed the traditional SARIMA[26]. These studies showcase ARX-types' versatility and strong performance.

However, ARX-types require extra data which may not be available, making them more specialised generalisations of ARIMA.

Generally, and as noted in the review by Cruz et al., ARIMA models show better performance in shorter time horizons but are consistently outclassed in other scenarios[1]. Furthermore, Weron concludes that ARIMA models are successful in relaxed time series but struggle capturing spikes[3]. As such, ARIMA has limited viability, and in its deficient areas, other models should be prioritised. Hybrid models address this drawback, and further research and development of hybrids is recommended for their potential industrial usage.

3.3 Support vector regression

SVR is classified as a computer intelligence model and is a recent development proposed in 1996 by Vapnik et al., rooted in the Vapnik-Chervonenkis theory[27]. SVR is relatively simple compared to other recent forecasting methods, such an ANN, as it requires less parameters[22]. Indeed, a major advantage of SVR is that the dimensionality of its inputs has no impact on the complexity of the problem. This, coupled with SVR being a data-driven approach allows its use for applications rich with historical data that SVR can utilise effectively, such as stock pricing and customer demand[22,28]. However, this flexibility means that selection of parameters is troublesome since each assumption and selection must be justified from theoretical and even empirical perspectives[3]. Currently, grid-searches and Cherkassky and Ma's method are used for parameter selection, where parameters are selected heuristically as there exists no formal or industrially accepted method[29]. This poses problems as the interpretation of heuristics can differ, meaning parameter selection has major impacts on forecasting quality. This is a consequence of SVR's recency, and more research is needed before SVR becomes an industrial preference for forecasting.

SVR works via an optimisation problem proposed by Vapnik[7]:

$$\min \frac{1}{2}\|w\|^2 + C\sum_{t=1}^{N}(\xi_t + \xi_t^*) \tag{3.3.1}$$

Subject to:

$y_t - \boldsymbol{w}\boldsymbol{x_t} - \beta \leq \varepsilon + \xi_t$ for $t = 1, \dots, N$

$-y_t + \boldsymbol{w}\boldsymbol{x_t} + \beta \leq \varepsilon + \xi_t^*$ for $t = 1, \dots, N$

$\xi_t \geq 0$ for $t = 1, \dots, N$ and $\xi_t^* \geq 0$ for $t = 1, \dots, N$

This is called the primal problem. The optimisation works by controlling the first and second term, which represent model complexity and error tolerance of the model respectively. It must be sufficiently complex to ensure the model is not too flat/smooth or close to the training data. A visualisation is shown below.

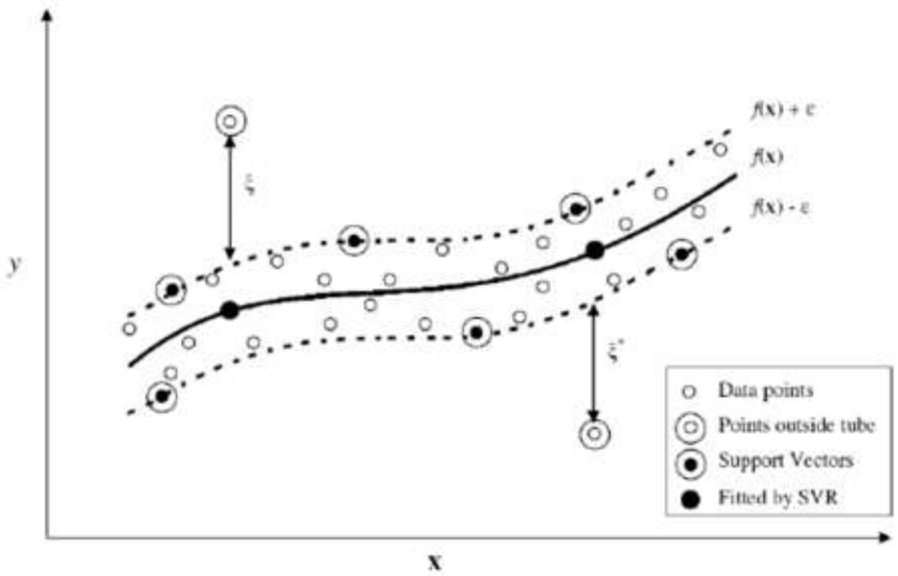

Figure 3.1 - Visualisation of SVR[30]

Research has found that solving the dual counterpart of the primal problem will provide the same solution whilst minimising the difficulty and computational demand of solving the problem[22.31,32,33].

Further details on the dual problem, including its formulation and parameters will be covered in the methodology.

The kernel trick is used in the dual problem. This allows the nonlinearization of the model by mapping the points of the input space into a feature space. While mapping onto a higher feature space would necessitate increased computational effort, the kernel function achieves the same result whilst reducing this computational complexity[22,27,32].

The most used kernels in SVR are the linear, polynomial and gaussian radial basis function (RBF) kernels due to their simplicity and strong performance. Other kernels include the chi-square and sigmoid kernel. For instance, studies performed by Soukissian et al. and Samal et al. have found success of the chi-square kernel in wind data predictions[34,35]. The sigmoid function is commonly used in ANNs but has started being used for SVR. Indeed, a study by Ye et al. found that the sigmoid kernel achieved similar results with polynomial and gaussian RBF on the prediction of fuel cell performance attenuation using SVR[36]. Krishnamurthy et al. discovered that SVR using a sigmoid kernel estimated values of Lyapunov exponents with an average deviation of ± 0.098 and a maximum deviation less than ± 0.2 when compared to expected values[37]. This was compared with SVR using linear kernels which achieved an average deviation of ± 0.38 and a maximum deviation of ± 0.83, demonstrating the potential of the less popular sigmoid kernel as it obtained superior results versus the common linear kernel.

However, SVR using the chi-square and sigmoid kernels must be researched in more SVR applications before determining their viability, as current research is limited.

The kernel functions will be covered further in the methodology.

Regarding results, SVR has achieved high forecasting performance in numerous uses. In a power prediction case study, SVR predicted results with only a 1.03% error[28]. For customer demand forecasting, PAs of 95.22%, 95.24% and 93.29% were achieved for three case studies[22]. Sansom et al. compared SVR methods with other methods such as ANNs for electrical price forecasting, and found that while both produced similar, accurate results, SVR required less training, computational demand and produced consistent forecasts[38]. SVR is also employed in hybrid models.

For instance, Che and Wang's SVRARIMA model produced better electricity price forecasts than existing approaches including ANNs and ARIMA[21]. An interesting note is that across all model types, hybrid or composite approaches often outperform singular models, stressing the need for their further research.

Of all the methods, SVR is the least mature technology with the least research, evidenced by its lack of industrial/commercial application, making SVR a germane method to expand research upon.

3.4 Artificial neural networks

ANNs have achieved significant attention in forecasting and all academia as they show promising results for a myriad of applications. In particular, it is of great interest in forecasting due to its ability to represent complex, nonlinear behaviour and to identify seemingly indiscernible patterns, something which conventional methods struggle with[39].

ANNs are a network of nodes, or 'neurons' that are arranged in layers with unidirectional connections between nodes of each layer. The first layer (input layer) is composed of predictors and the final layer gives the output. An intermediate hidden layer can be introduced which transforms the ANN to a nonlinear one (see Figure 3.2). This class of ANN is called a nonlinear multilayer perceptron (MLP)[3]. Each predictor is assigned a coefficient or 'weight' that is calibrated/tuned based on training data by learning algorithms that minimise a specified cost function[6,40].

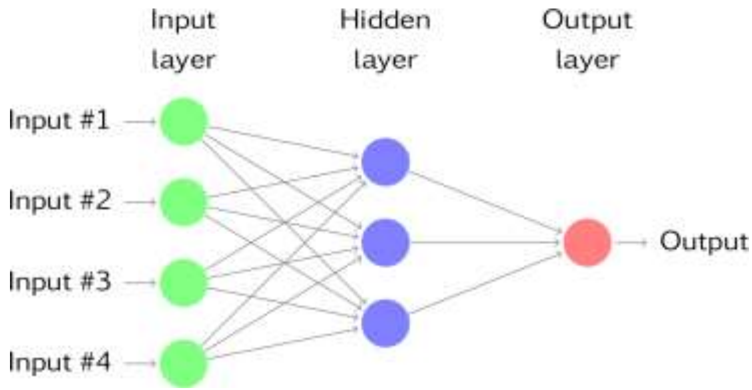

Figure 3.2 – Neural network with a hidden layer

In feed-forward networks there are no loops, where the nodal outputs in a preceding layer are the nodal inputs in the following layer. Feed-forward networks, specifically nonlinear MLP, are favoured for forecasting[41] Indeed, Cruz et al., Chen et al., Gareta et al., Mandal et al. and Yamin et al. employ MLP ANNs for electricity price forecasting[1,42,43,44,45]. Several of these including Cruz et al. compare MLP ANNs to ES and ARIMA methods finding that ANNs perform better.

MLP ANNs have also been used in hybrids and as a comparison benchmark to more sophisticated ANNs. Gonzalez et al. proposed a successful MLP input/output hidden Markov model hybrid that yielded accurate results and information about the Spanish electricity market[46].

ANNs have seen similar success in sales and customer demand forecasting. Lo et al. found that ANNs with an accompanying genetic algorithm for weights optimisation, accurately and rapidly forecasted sales for a restaurant in Taipei City, Taiwan[47]. Alon et al. conducted a comparative study of ANNs and traditional methods such as ARIMA and Winters ES. For two forecasting horizons and two forecasting periods, ANN achieved the lowest average MAPE of 1.5%, beating ARIMA (1.67%) and Winters (2.19%)[48]. Further analysis revealed that ANNs could more effectively capture dynamic nonlinear seasonal patterns/trends and their interactions. For instance, they suggested ANNs yield superior forecasts during periods of economic turbulence and large fluctuations[48]. This is supported by Kang who found that while ARIMA models achieve similar MAPE to ANNs, they have greater error when trends/seasonality are present in data[49].

While empirical findings are inconclusive regarding whether ANNs or traditional methods are better for yearly forecasts, ANNs are favoured for immediate or short-term horizons. Indeed, reviews by Donkor et al and Weron share the conclusion that newer models, particularly ANNs, are more proficient at short-term forecasting irrespective of the dataset/series, hence why ANNs are rising as an industrial option for short-term forecasts[3,50].

Studies by Ghiassi et al., Bougadis et al., Jain et al. and Firat et al., found that ANNs had better performance for water demand forecasting for several forecasting horizons versus conventional approaches[51,52,53,54].

A criticism of ANN, highlighted by Lasek et al., is that its architecture setup and optimisation for forecasting is often time-consuming and costly[40]. A feed-forward or recurrent network must be decided, along with the number of hidden layers and nodes in each layer. Furthermore, the training algorithm to calibrate weights must be selected or formulated[40]. Therefore, if simpler methods can be used to generate a reliable and accurate forecast, this can be more cost-effective and advantageous than using ANN. Additionally, for complex applications such as process industries that have many possible variables, ANNs may lead to excessive parameter usage. This criticism was highlighted by Bhat and McAvoy as using ANN for already complicated problems can lead to needlessly complex application and overparameterization[55]. This could cause significant time wastage and economic ramifications.

3.5 Conclusion of Literature Review

This literature review demonstrated that forecasting is a useful process with important real-life applications, thus commanding an extensive body of research. Most research focusses on conventional methods such as ARIMA and ES, which show good results but limited viability for a wide range of applications. However, they are technologically mature and have an accepted, established framework for implementation, making them widely used in industry.

Whereas, data driven methods such as ANN and SVR are newer and seen as more sophisticated models utilising the benefits of computational advancements. As such, research has shown they generally achieve better results and are more versatile than conventional techniques. However, they are less mature techniques meaning they are less widely used in industrial applications where proven methods with an accepted, formal framework are favoured. This is particularly true for SVR, where literature has shown there is no clear best method for parameter selection.

Hybrid models have shown very promising results, but they are markedly less developed than singular models, being used in minimal specialised research applications. As such, considerably more research is required before usage in real-world forecasting.

The motivation of this paper is to research the performance of SVR on two datasets and comparing the results to the performance of a conventional technique in ARIMA. The literature review supports this motivation, as further research into SVR and its performance is necessary before it is common in industry. ANNs were also identified as a strong method, so future work and research into their application is recommended.

4. Methodology

4.1 Problem Formulation – SVR

As mentioned in the review, the dual problem is formulated from the primal problem and utilises kernelization. In doing so, the parameters and computational demand is reduced, without compromising the solution. Firstly, a Lagrangian function is constructed from the primal objective function[22,56].

$$L = \frac{1}{2}\|w\|^2 + C\sum_{t=1}^{N}(\xi_t + \xi_t^*) - \sum_{t=1}^{N}\lambda_t\big(\varepsilon + \xi_t - y_t + (\boldsymbol{w}\boldsymbol{x_t} + \beta)\big) - \sum_{t=1}^{N}\lambda_t^*\big(\varepsilon + \xi_t^* - y_t + (\boldsymbol{w}\boldsymbol{x_t} + \beta)\big) - \sum_{t=1}^{N}(\mu_t\xi_t + \mu_t^*\xi_t^*) \tag{4.1.1}$$

Where $\lambda_t, \lambda_t^*, \mu_t, \mu_t^*$ are Lagrange multipliers and L is the Lagrangian.

According to the saddle point condition, the partial derivatives of L with respect to each primal variable w, ξ_t, ξ_t^*, β must vanish at the optimal conditions[56]. Therefore:

$$\partial_w L = w - \sum_{t=1}^{N}(\lambda_t - \lambda_t^*)\boldsymbol{x_t} = 0 \tag{4.1.2}$$

$$\partial_\beta L = \sum_{t=1}^{N}(\lambda_t - \lambda_t^*) = 0 \tag{4.1.3}$$

$$\partial_{\xi_t} L = C - \lambda_t - \mu_t = 0 \quad , \quad \partial_{\xi_t^*} L = C - \lambda_t^* - \mu_t^* = 0 \tag{4.1.4}$$

Substituting Equations 4.1.2, 4.1.3 and 4.1.4 into 4.1.1 yields the dual optimisation problem.

$$\max \frac{1}{2}\sum_{t'=1}^{N}\sum_{t=1}^{N}(\lambda_{t'} - \lambda_{t'}^*)(\lambda_t - \lambda_t^*)\cdot \boldsymbol{x_{t'}}\boldsymbol{x_t} - \varepsilon\sum_{t=1}^{N}(\lambda_t + \lambda_t^*) + \sum_{t=1}^{N}y_t(\lambda_t - \lambda_t^*) \tag{4.1.5}$$

Subject to: $\sum_{t=1}^{N}(\lambda_t - \lambda_t^*)$ and $\lambda_t, \lambda_t^* \in [0, C]$, where λ_t, λ_t^* are the dual variables.

Next, the variable β and slack variables ξ_t, ξ_t^* are found via differentiation of the primal objective function[22].

$$\min \sum_{t=1}^{N}(\xi_t + \xi_t^*) \tag{4.1.6}$$

Subject to:

$$y_t - \textstyle\sum_{t'=1}^{N}(\lambda_{t'} - \lambda_{t'}^*)\cdot \boldsymbol{x_{t'}}\boldsymbol{x_t} - \beta \le \varepsilon + \xi_t \, , -y_t + \sum_{t'=1}^{N}(\lambda_{t'} - \lambda_{t'}^*)\cdot \boldsymbol{x_{t'}}\boldsymbol{x_t} - \beta \le \varepsilon + \xi_t^*$$

and $\xi_t, \xi_t^* \ge 0$

Now, kernelization can be employed where the inner product of the input vectors $\boldsymbol{x_{t'}}\boldsymbol{x_t}$ is mapped onto a higher feature space, from which the kernel function is introduced in its place. The final forms of the dual problem are now displayed, and these will be inputted into GAMS

$$\max \frac{1}{2}\sum_{t'=1}^{N}\sum_{t=1}^{N}(\lambda_{t'} - \lambda_{t'}^*)(\lambda_t - \lambda_t^*)\cdot K(x_t, x_{t'}) - \varepsilon\sum_{t=1}^{N}(\lambda_t + \lambda_t^*) + \sum_{t=1}^{N}y_t(\lambda_t - \lambda_t^*) \tag{4.1.7}$$

Subject to:

$\sum_{t=1}^{N}(\lambda_t - \lambda_t^*) = 0$ and $\lambda_t, \lambda_t^* \in [0, C]$, where λ_t, λ_t^* are the dual variables.

$$\min \sum_{t=1}^{N} (\xi_t + \xi_t^*) \tag{4.1.8}$$

Subject to:

$$y_t - \sum_{t'=1}^{N}(\lambda_{t'} - \lambda_{t'}^*) \cdot K(x_t, x_{t'}) - \beta \leq \varepsilon + \xi_t \text{ , } -y_t + \sum_{t'=1}^{N}(\lambda_{t'} - \lambda_{t'}^*) \cdot K(x_t, x_{t'}) - \beta \leq \varepsilon + \xi_t^*$$

and $\xi_t, \xi_t^* \geq 0$

Finally, the regression function from which the predictions are calculated is formed.

$$y_t = \sum_{t=1}^{N} (\lambda_{t'} - \lambda_{t'}^*) \cdot K(x_{t'}, x_t) + \beta \tag{4.1.9}$$

These equations are inputted into GAMS, where Eq. 4.1.7 is solved via nonlinear programming (NLP) giving the values of dual variables λ_t, λ_t^*. These values are fixed and treated as parameters for Eq. 4.1.8, which is solved via linear programming (LP) yielding the values of β and slack variables ξ_t, ξ_t^*. Finally, Eq. 4.1.9 is used to recursively calculate the predictions for a specified forecasting horizon.

4.2 Problem formulation – ARIMA

The ARIMA models are built by open-source python code on Jupyter Notebook, utilising pre-packaged modules/libraries including Pandas, NumPy and statsmodels.

As discussed in the review, there are variants/generalisations of ARIMA that are used depending on the characteristics of the time series. When a series exhibits seasonality, SARIMA or a SARIMA variant is recommended[26]. Otherwise, a standard ARIMA model will be used as a comparison benchmark for SVR.

To determine ARIMA parameters, composed of nonseasonal (p,d,q) and seasonal (P,D,Q) elements, an automated grid search is used with the aim of minimising the Akaike Information Criterion (AIC). AIC is a value estimating the information lost when a statistical model, like ARIMA, models a series[57]. As such, by selecting parameters minimising AIC, the best-fitting model will be obtained[57].

The parameter m (periodicity) for SARIMA/SARIMAX is assigned a value depending on the data properties. For instance, if the series has a seasonality pattern every seven days, then m=7.

The SARIMAX model code also creates a standardised residual plot, to determine whether the series' seasonality pattern has been eliminated or minimised to a sufficient degree

The standard ARIMA model code works similarly. It will feature automated parameter optimisation via minimisation of AIC, but forgoes the seasonal parameters $(P,D,Q)_m$. From this, the best-fitting model is selected from which a forecast is generated. The P-value will also be reported, where a sufficiently low value means the series is stationary and ARIMA is applicable.

The SARIMAX and ARIMA models will be compared, to demonstrate whether the seasonal component of SARIMAX actually improves model accuracy, or whether the standard ARIMA is sufficient for seasonal datasets.

4.3 Parameter Selection – SVR

The parameters must be selected prior to SVR modelling. As mentioned, parameter selection is vital as they directly influence the forecast. For SVR, only four parameters need selection, which is less than most forecasting methods.

The parameters are:

- C (regularisation factor)
- ε (error tolerance)
- K (kernel function)
- number of past demand attributes to be included

C has an important effect on the model fit, in turn affecting the forecasting performance. If C is too high, there is greater weight to minimisation of error, causing overfitting. If C is too low, there is lesser weight to minimisation of error, causing underfitting as the model is unable to accurately describe the training data learns from[31,58]. Thus, C is essentially a trade-off between model accuracy and model smoothness/generalisation[58]. Indeed, analysis by Ito and Nakano (see Figure 4.1) shows the effect of C variation, where if C is too small, there is a negligible penalty meaning the regression is too flat/smooth[32]. On the contrary, if C is too large, there is an excessively strict penalty, with SVR fitting too close to training data[32].

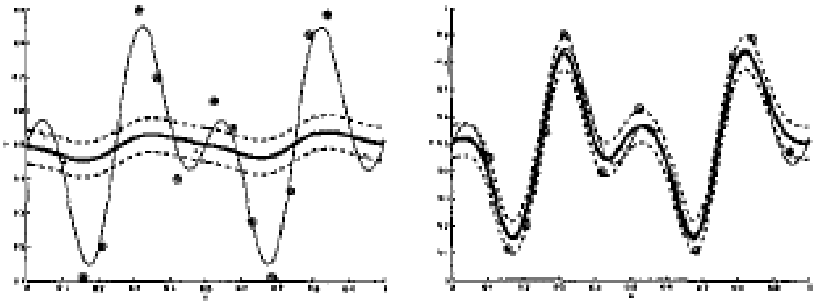

Figure 4.1 – Effect of C on SVR. Left graph has small C and right graph has large C

Currently, there is no established method for selection of this hyperparameter[22,58]. While grid-searches are often used, they are computationally complex and expensive[59]. As such, the Cherkassky and Ma heuristic rule is used as it shows good performance whilst minimising computational complexity[22,59]. It is defined below:

$$C = \max\left(\bar{y}_t + 3\sigma_{y_t}, \qquad \bar{y}_t - 3\sigma_{y_t}\right) \tag{4.3.1}$$

Where $\bar{y}_t$ is the mean of training data and σ_{y_t} is the standard deviation of training data.

To determine the effect of C on the forecasting performance, a sensitivity analysis will be performed where C will be varied by an order of magnitude of 1.

ε is error tolerance, determining the number of training datapoints that are outside the ε-sensitive tube, meaning ε determines the degree to which error and noise of training data is minimised. If ε is too small, then the tube becomes too thin, missing most of the training datapoints. If ε is too large, then the tube includes too many points meaning the regression will be less meaningful as it will be affected by noise. This is illustrated in Figure 4.2.

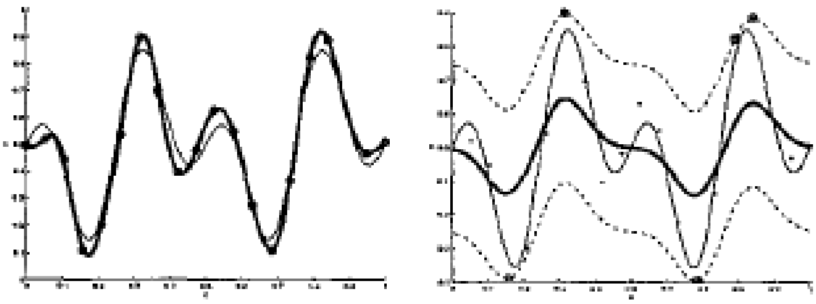

Figure 4.2 – Effect of ε on SVR. Left graph has small ε and right graph has large ε

For this study, a heuristic rule is used where the value of ε is taken as one order of magnitude lower than $\bar{y}_t$[22,29].

$$\varepsilon = \bar{y}_t / k \tag{4.3.2}$$

Where k is a constant between [10,30]. For the purposes of determining the impact of ε on forecasting quality, k shall be varied at values 10, 20 and 30.

Many kernel functions can be used. The most common kernels used for SVM/SVR are linear and polynomial kernels. This study will also analyse the application of a novel kernel function in chemical engineering in the chi-square kernel. While the gaussian RBF is often used, the popular linear and polynomial and novel chi-square kernels are chosen such that this study expands upon the depth and breadth of SVR and kernels research.

The positive-definite chi-square kernel is written as follows:

$$K(x_t, x_{t'}) = \sum_{i=1}^{N} \frac{2x_{t,i}x_{t',i}}{x_{t,i}x_{t',i}} \tag{4.3.3}$$

The other kernels are defined below.

Linear Kernel:

$$K(x_t, x_{t'}) = x_t \cdot x_{t'} + c \tag{4.3.4}$$

Where c is a constant ($c = 1$).

Polynomial kernel:

$$K(x_t, x_{t'}) = (x_t \cdot x_{t'} + c)^p \tag{4.3.5}$$

Where c is a constant ($c = 1$), and p is the polynomial degree ($p = 2$).

The forecasting performance for each kernel will be quantitatively and qualitatively evaluated to determine the most appropriate kernel for each series. Comparisons will be made to literature, with the expectation that the polynomial kernel will have the best performance, due to its success in research, and that the chi-square kernel will be the weakest due to its scarcity in SVR literature.

The number of past demand attributes included in the input vector is dependent on the dataset's seasonality. As such, it will differ greatly for short-term to long-term time series. The selected values will be discussed in the results section where both time series will be analysed.

4.4 Evaluation of Forecasting

This section will cover the comparison and assessments criteria for the results. Firstly, the MAPE criterion will be used. This is a common statistical measure to simply and quantitatively assess forecasting performance in terms of accuracy. The formula is reported below.

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{N}\left|\frac{y_{t,actual} - y_t}{y_{t,actual}}\right| \cdot 100 \tag{4.4.1}$$

Where n is number of forecasted points, y_t is the predicted point, $y_{t,actual}$ is the corresponding actual point.

Furthermore, each forecast will be plotted to qualitatively evaluate the forecasting performance, since this will allow one to assess the shape/pattern of the forecast compared to the actual series.

In addition, comparisons between SVR and ARIMA for each dataset will be made, to determine their strengths and weakness versus each other.

5. Results

This section presents the results for two time series: customer demand and wind capacity.

5.1 Customer sales dataset

The daily sales/demand time series is visualised below.

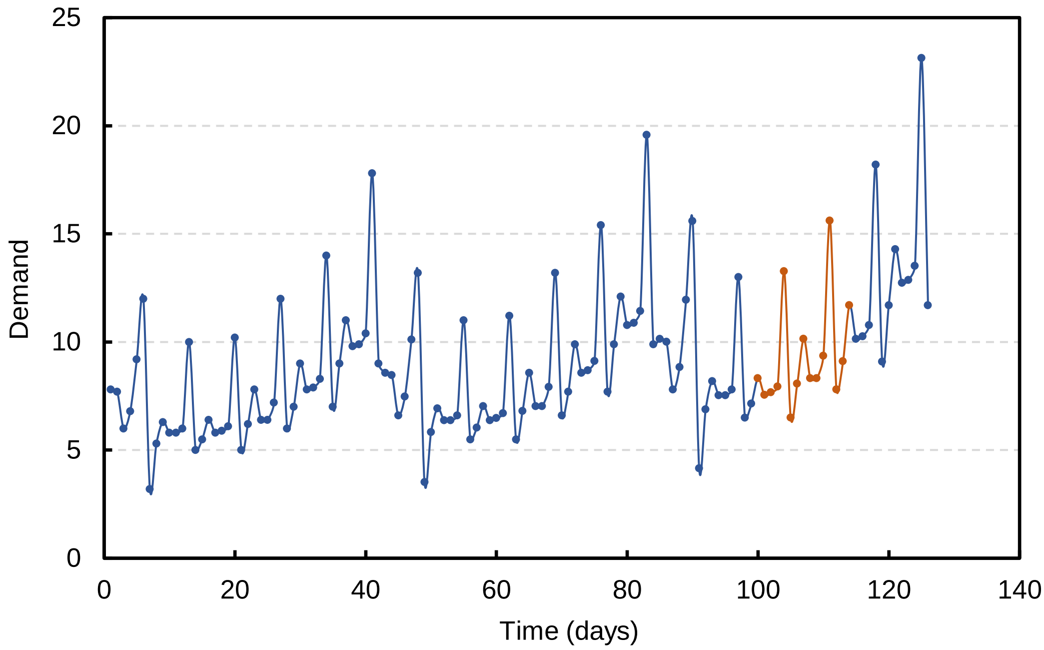

Figure 5.1 - Customer sales time series

Figure 5.1 shows the series has seasonality and trends. One trend is the gradual increase in sales followed by a large peak and a large drop the following day. This pattern repeats itself every week, meaning the series exhibits seasonality. Another trend is a general increase in sales over the entire 126-day series.

Thus, this dataset is considered predictable with 'friendly' training data meaning it is expected for the forecasts to achieve a high performance.

The orange portion is the 2-week section that will be forecasted, meaning the preceding 100 datapoints will be the training data. This allows the forecast to be compared to the actual data to assess its performance and accuracy.

From the parameter selection formulae in the methodology, the required SVR parameters are determined. Sample calculations for hyperparameters C and ε, using Eq. 4.3.1 and 4.3.2 respectively, are shown.

$$C = \max\left(\bar{y}_t + 3\sigma_{y_t}, \quad \bar{y}_t - 3\sigma_{y_t}\right)$$

$$C = \max(8.38 + 3 \cdot 2.86, \quad 8.38 + 3 \cdot 2.86) = \max(16.96, -0.196)$$

$$\therefore C = 16.96 \approx 17.0$$

$$\varepsilon = \bar{y}_t / k$$

Initially, let k=20, but this will be varied in the sensitivity analysis to analyse the effect of ε on the forecast.

$$\varepsilon = 8.38/20$$

$$\therefore \varepsilon = 0.419$$

The past demand attributes to be included will equal 14. This is because it is recommended to choose a value that is a multiple of the seasonality pattern exhibited by the series[22]. From observation it was determined that the pattern repeated itself every 7 days, so 14 is a suitable number for past demand attributes meaning that the calculations for each forecasted point will be based on the past 14 datapoints.

The SVR parameters for this time series are tabulated in Table 5.1.

Table 5.1. SVR parameters for customer demand time series.

Parameter	Value
C	17.0
ε	0.419
Past demand attributes	14

Using these parameters, SVR is used to generate a forecast. Firstly, SVR using a linear kernel is considered.

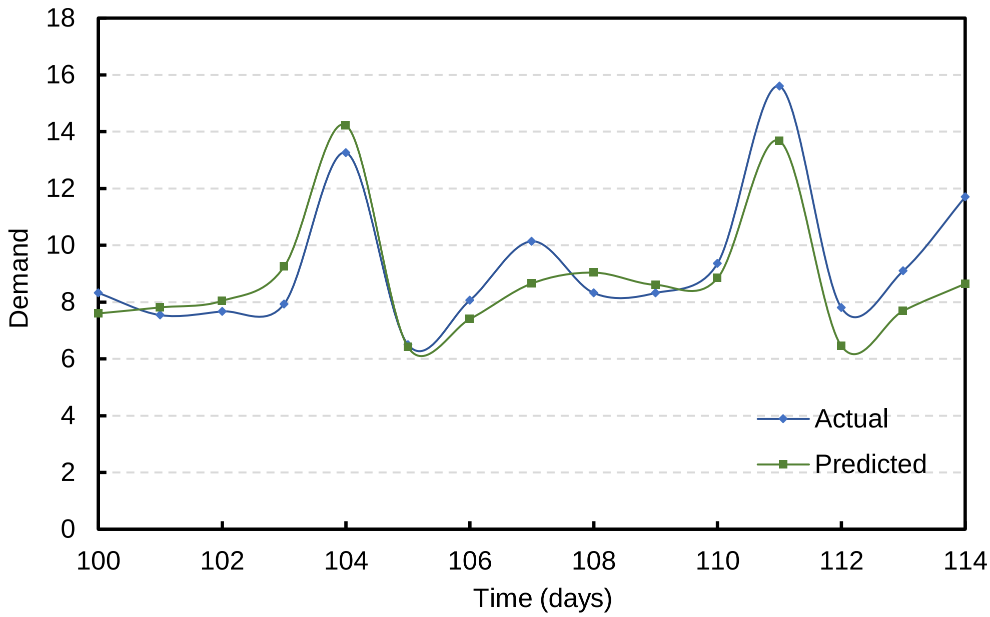

Figure 5.2 - SVR customer demand forecast using linear kernel.

From observation of Figure 5.2, the forecast is successful. The model can capture the pattern of the series, albeit not perfectly. Indeed, it fails to capture the more intricate pattern between 106-110 days, and is quite wayward in the last 2 days, where if the forecast was extrapolated the MAPE would worsen. The model also captures the underlying overall trend of an increase in sales over time, showing its ability to capture distinct and indistinct trends.

Interestingly, it slightly overpredicts the first peak (104 days) and slightly underpredicts the second peak (111 days). However, it can still closely match these turbulent periods whereas techniques, like ARIMA, often struggle closely following extreme peaks in a series[3].

MAPE is calculated using Eq. 4.4.1, with a sample calculation below.

$$\text{MAPE} = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{y_{t,actual} - y_t}{y_{t,actual}}\right| \cdot 100 = \frac{1}{14} \cdot 1.54 \cdot 100 = 10.99\% \approx 11.0\%$$

The linear kernel model has a MAPE of 10.99% corresponding to a PA of 89.01%. This shows that the forecast performed very well but it could be improved via optimisation of parameters or selecting a more sophisticated kernel function. Indeed, it was not expected for a linear kernel to have the best results as it is the simplest kernel function.

Secondly, SVR using a polynomial kernel function will be considered.

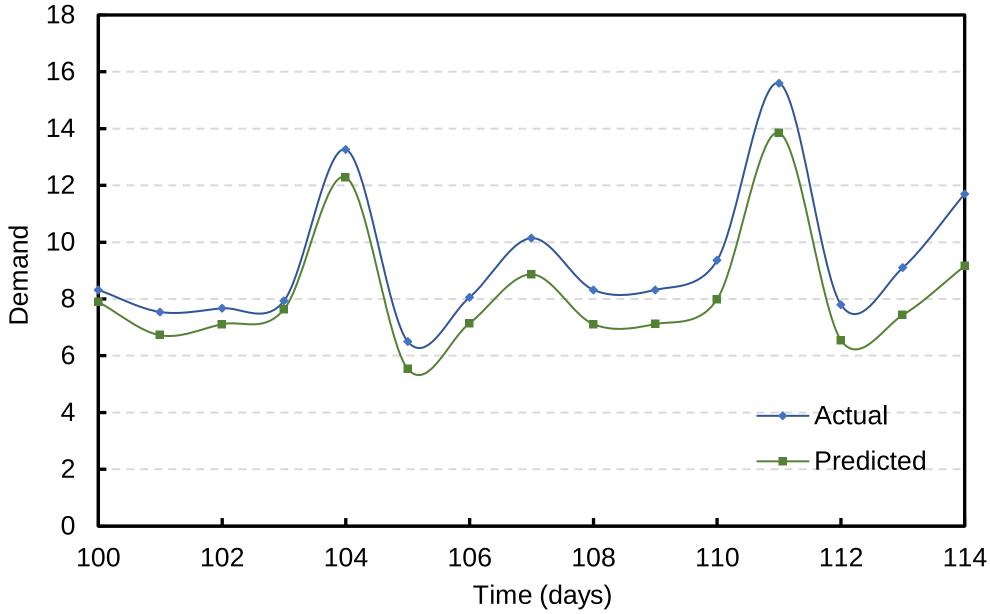

Figure 5.3 - SVR customer demand forecast using polynomial kernel.

Figure 5.3 shows the polynomial kernel captures the distinct trends well. While the forecast is not wholly accurate, as most datapoints are underpredicted, it follows the series' patterns closely. Indeed, where the linear kernel failed to follow the intricacy of the series between 106-100 days, the polynomial kernel captures it. This was expected since the polynomial kernel is more complex than its linear counterpart. Similarly, with the linear kernel, the polynomial kernel can capture the spiky periods and general increase over time.

However, as mentioned, all data points are underpredicted. This could be due to some data bias, meaning that varying and optimising the SVR parameters could address this improving the forecast accuracy. Recalling the regression function that recursively calculates the prediction (Eq. 4.1.9), it can be inferred that β controls the position of the forecast. Therefore, if β changes, the forecast can shift which may increase the accuracy. This modification will be investigated in the sensitivity analysis.

The MAPE was found to equal 13.1% with a PA of 86.9%. This is lower than SVR using linear kernel, meaning the linear kernel is quantitatively more accurate. However, as discussed, the polynomial kernel captures the series' shape and pattern better, so variation of its SVR parameters could significantly improve its MAPE.

This illustrates that MAPE alone is not a conclusive measure of forecasting performance, since MAPE suggests that the linear kernel was superior, but visual inspection of both models shows that the polynomial kernel follows the pattern/shape more closely and just needs further optimisation to improve its accuracy.

The novel chi-square kernel approach is now modelled.

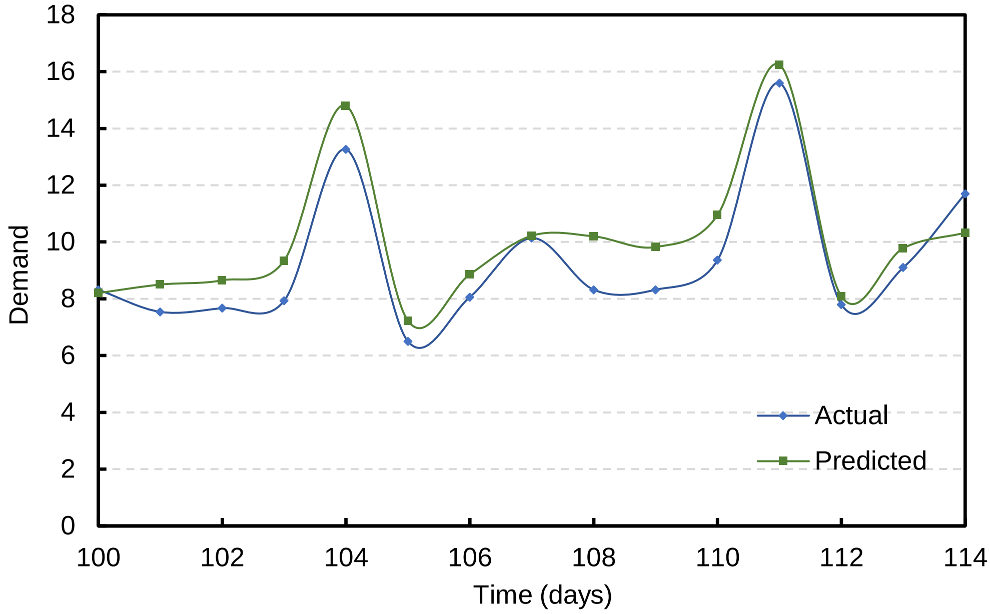

Figure 5.4 - SVR customer demand forecast using chi-square kernel.

Figure 5.4 illustrates the chi-square kernel is more erratic in its prediction. Again, as with previous examples, it successfully follows the explicit and implicit trends of the series. Indeed, it accurately follows the extreme peaks at 104 and 111 days. However, at times the forecast doesn't follow the pattern well. Notably the period between 106-110 days, which the linear kernel also had some trouble with, and the last period of 112-114 days where it has a significantly different gradient and shape to the actual series.

The MAPE determined to be 11.6% with a PA of 88.4%. From an accuracy perspective, the chi-square kernel was level with the other kernels, showing SVR was generally successful.

However, visual inspection illustrated that the linear and especially the chi-square kernel SVR forecasts were weaker than polynomial kernel in their ability to capture the pattern/shape of the actual series. As such, the former two have less potential to be improved via parameter optimisation, whereas the polynomial kernel has the most potential, as parameter optimisation can address biases in the data and can determine the optimal value for β to shift the forecast upwards.

Next, an ARIMA model was formed to produce a forecast with the same forecasting horizon, to compare the performance of SVR to that of a more mature and established technique.

The ARIMA model was generated via open-source python code using pre-packaged modules/libraries which automates the parameter optimisation and produces a resulting forecast. Due to seasonality present in the customer sales dataset, a seasonal ARIMA variant is needed to eliminate the seasonality elements, with SARIMAX chosen as it had the best performance in the review.

The series' periodicity is seven days, hence m=7. The code calculates AIC for all possible SARIMAX model variations, with the aim of minimising AIC to determine the best-fitting model.

The code found that SARIMAX(3, 1, 1)(3, 1, 1)$_7$ was the best-fitting model, minimising AIC to 219.5.

The standardised residual of the SARIMAX model is plotted to show the the seasonality was eliminated to transform the series from non-stationary to stationary, as required by ARIMA models. It is evident the residuals do not show seasonality (see Figure 5.5).

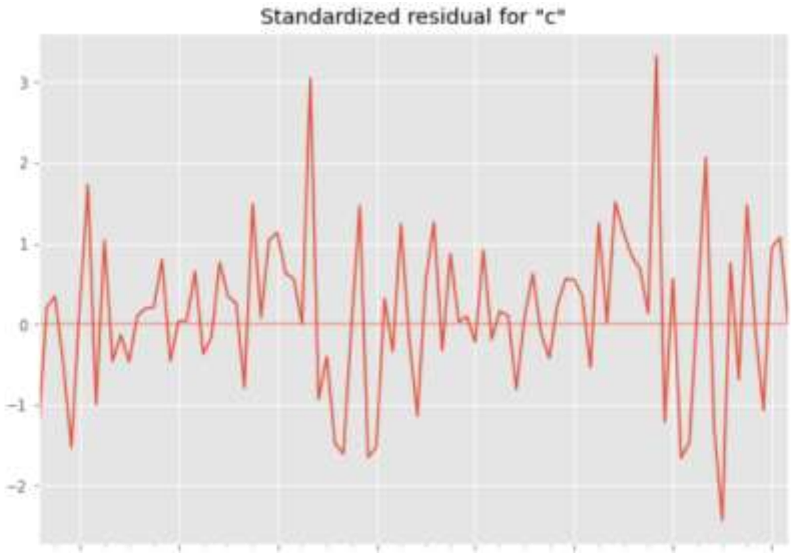

Figure 5.5 - Standardised residual plot generated by SARIMAX model via python code.

Finally, the SARIMAX model generated a forecast with the results visualised below.

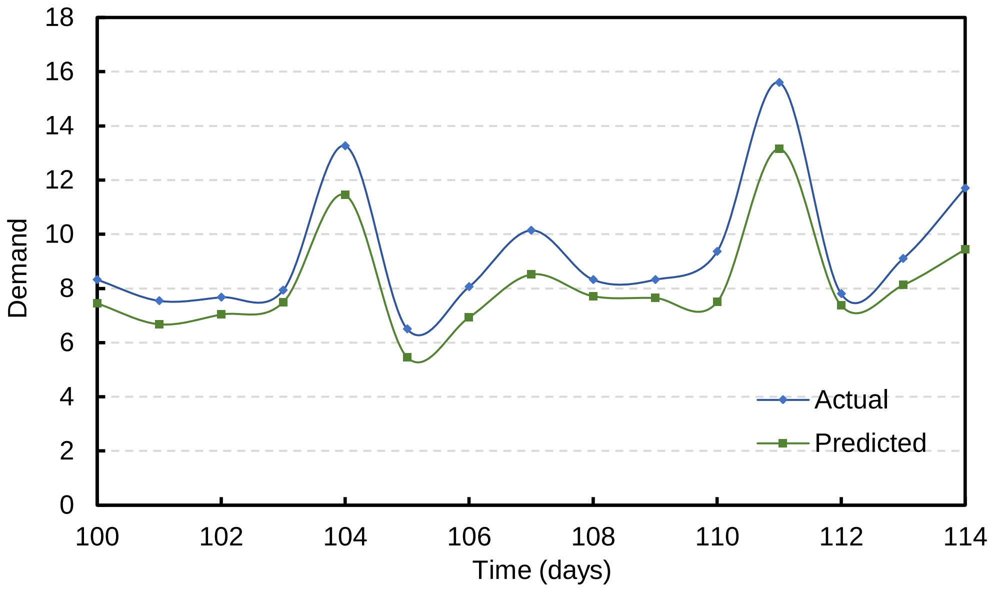

Figure 5.6 - Customer demand forecast using SARIMAX model via python code.

Firstly, the ARIMA model captures the general trend and shape of the series on a similar level to SVR models and similarly underpredicts all points as with the polynomial kernel model.

However, observation of Figure 5.6 shows that it struggles with the large peaks at 104 and 110-111 days, with these discrepancies chiefly exacerbating its forecasting quality (MAPE). While all SVR models were able to closely match the actual series during these peaks, the SARIMAX model struggles to capture the extremity of these datapoints. Based on literature findings, ARIMA models generally struggle to closely follow extreme time periods, here the peaks on days 104 and 111, but can capture relaxed time periods[3]. Therefore, the performance and shape of the SARIMAX model is expected, and Figure 5.6 serves as an apt depiction of the finding.

The MAPE for the SARIMAX model was 13.0% giving a PA of 87.0%. This shows the forecast was a similar, high performance to the SVR models in terms of accuracy. However, like with the MAPE of the polynomial kernel, MAPE does not reflect the true performance of the model. Regarding the forecast shape, the SARIMAX model is below the SVR model due to its struggle to capture the extremity of the two peaks in the series. Thus, for this series which had clear seasonality and trends, the results show SVR creates a slightly better forecast than ARIMA.

For comparative purposes, a standard ARIMA model was also tested despite the clear seasonality. This is to assess whether the additional seasonal component of SARIMAX improves forecasting performance. The code, utilising Pandas, NumPy and pmdarima libraries, deduced that ARIMA(2,1,2) minimising AIC to 544.2, was the best-fitting model. Worse performance is expected, as the series has strong seasonality which differencing may not be able to eliminate.

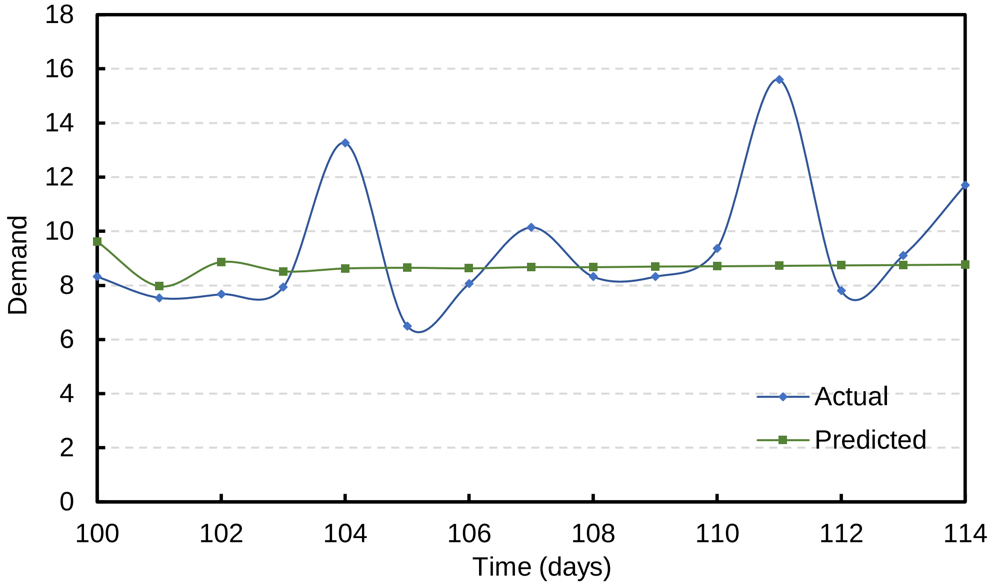

Figure 5.7 - Customer demand forecast using ARIMA model via python code.

Figure 5.7 proves the expectations. While ARIMA captures the general level, it is unable to capture the general trend of the series, contrary to all other models, yielding the lowest MAPE of 16.7%. This is because ARIMA was unable to transform this non-stationary series with differencing. Indeed, the P-value was 0.26 which is too high for ARIMA, indicating the series is highly seasonal and a seasonal component is necessary to transform the series to stationary.

When comparing Figure 5.7 to 5.6, it demonstrates the importance of seasonal ARIMA. The addition of a seasonal component allowed the transformation of non-stationary to stationary, facilitating a superior forecasting performance.

The results are reported below.

Table 5.2. Results for customer demand forecasts.

Method	MAPE	Prediction accuracy
SVR linear kernel	11.0	89.0
SVR polynomial kernel	13.1	86.9
SVR chi-square kernel	11.6	88.4
SARIMAX	13.0	87.0
Average	12.2	87.8

The results and discussion showed all forecast qualities were similar but SVR was slightly better overall with an average MAPE of 87.8%. The figures illustrated polynomial kernel SVR most closely captured the shape/pattern of the actual series. Consequently, the sensitivity analysis will focus on the polynomial kernel to eliminate any data bias and improve the MAPE and overall quality of the forecast.

5.2 Wind capacity dataset

The previous section focussed on a less noisy dataset with clearly identifiable seasonality and trends. As such, the training data is higher quality, allowing models to be trained and produce better forecasts more easily.

This section focusses on a volatile, noisy dataset with no distinct seasonality or trends. Though, there may still be underlying, nonlinear trends the models may or may not learn. This will be a useful test to evaluate each methods' ability to decipher a 'difficult' dataset, allowing for comparison of results from the 'easy' dataset.

The time series here is wind capacity data from London, United Kingdom. The plot below shows the wind capacities recorded from 16/12/2019 00:00 to 31/12/2019 23:00 in 1-hour intervals, giving a total of 384 datapoints.

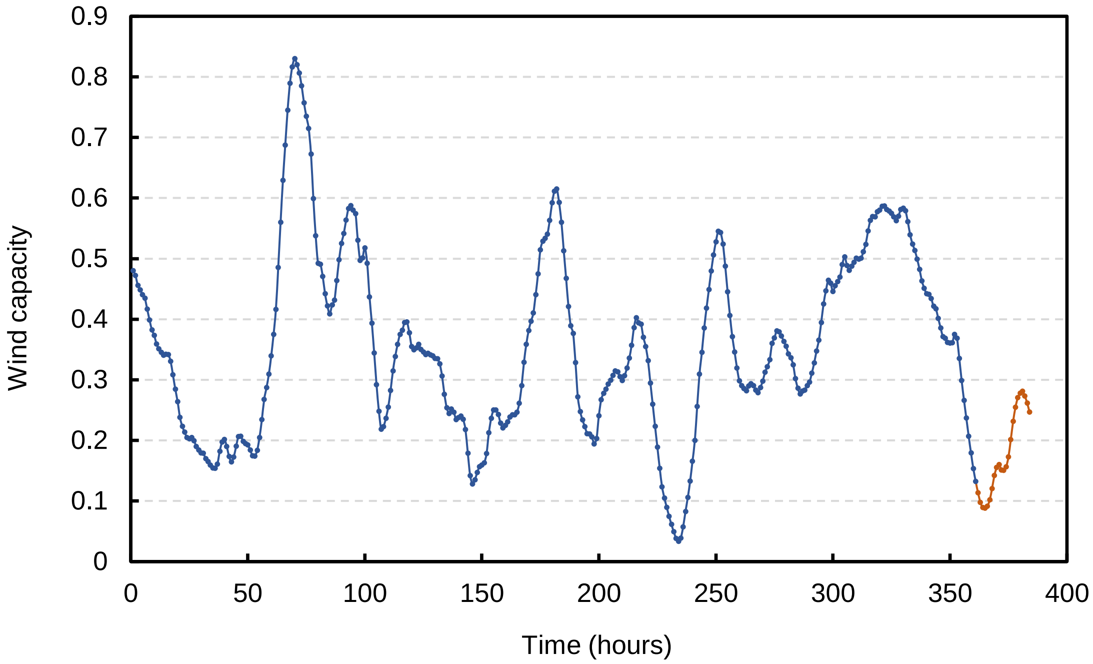

Figure 5.8 - Wind capacity time series

Looking at Figure 5.8, this is an evidently volatile dataset with no discernible trends or seasonality. Also, this is a spiky dataset with numerous moderate to extreme peaks and troughs. Therefore, the ARIMA model may struggle due to its known issue of underpredicting extreme peaks, which was seen in the previous section. While SVR was able to closely predict spikes, this dataset is a lot more random and unpredictable, so it will be a worthwhile test for the SVR models to observe if they can follow the shapes and spikes of this graph despite its lack of predictability.

The orange portion (23 hours from 31/12/19 00:00 to 31/12/19 13:00) will be forecasted and compared to the actual series. Therefore, there are over 300 training data points which could help offset the volatility and unpredictability of this series. However, more training data does not necessarily mean a better forecast because if the training data quality is low, such as having lots of noise, then the model will not efficiently learn from it, meaning the number of training points is less important.

Furthermore, this dataset will test the ability of the SVR and ARIMA models to distinguish between noisy data and meaningful data.

The SVR parameters for the wind time series are shown below.

Table 5.3. SVR parameters for wind capacity time series.

Parameter	Value
C	1.018
ε	0.0217
Past demand attributes	24

Firstly, the chi-square kernel is considered due to the use of chi-square in wind data forecasting research[34,35].

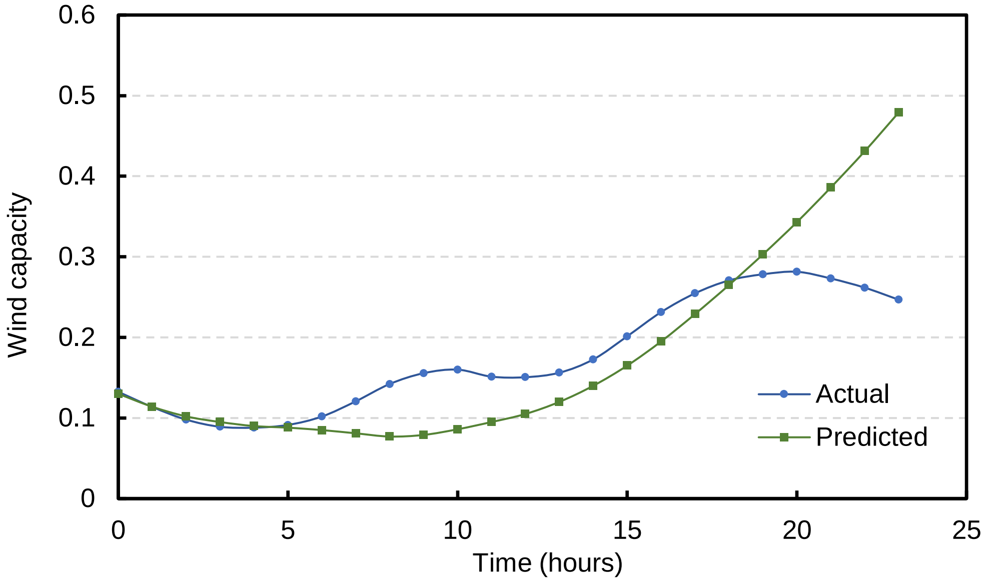

Figure 5.9 - SVR wind capacity forecast using chi-square kernel.

This graph shows that the chi-square kernel matches the general trend of the series of an increase in wind capacity over time. However, like its performance for the customer sales series, it fails to closely capture the intricate parts of the series. For instance, while it captures the trend of the period between 5-10 hours, it fails to capture the minor fluctuations in this period. Furthermore, it underpredicts near the middle before overpredicting at the end, where it continues to rise despite the actual series reaching a peak before declining.

The MAPE for this example is 24.8%. This shows the forecast was accurate overall but had deficiencies which are confirmed in the graph.

For this section, the linear kernel was omitted as it was determined in the previous section along with literature that polynomial is better. The wind capacity forecast using polynomial kernel function is displayed below.

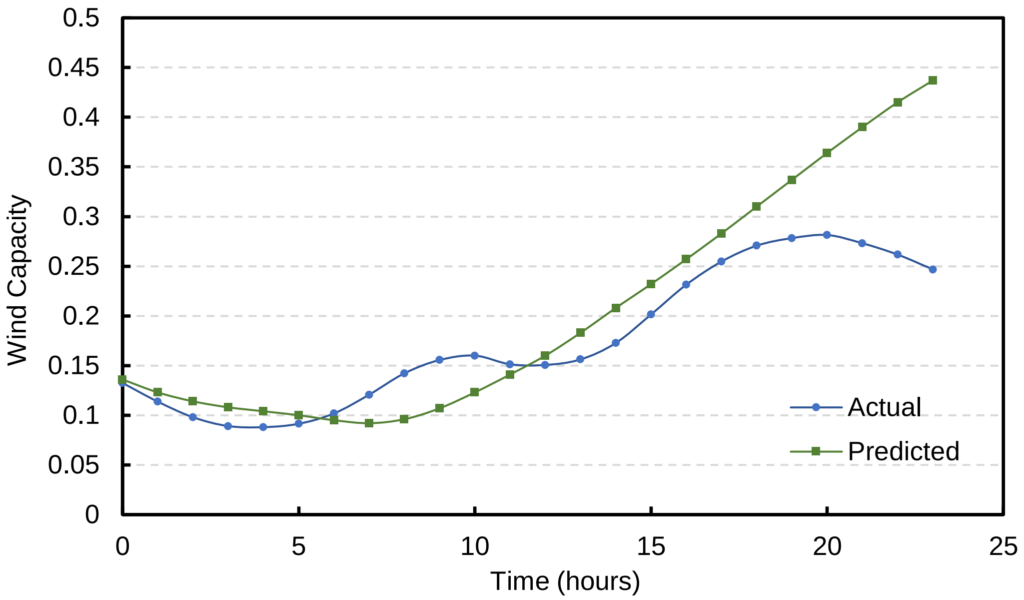

Figure 5.10 - SVR wind capacity forecast using polynomial kernel.

As with chi-square kernel and the customer demand forecasts, this example follows the overall trend of the time series. However, the performance of the polynomial kernel here is significantly worse than for the customer demand forecast. It fails to properly capture the small fluctuations and peaks throughout this series and often time the prediction points are far from their corresponding actual datapoints, particularly after 20 hours.

The MAPE was found to be 21.9% which is worse than its MAPE for the customer demand forecast. This was also seen in the chi-square kernel demonstrating that SVR generally struggled with this dataset. The reason for this is due to the volatility and unpredictability of the dataset. This made sure that despite the quantity of training datapoints, the data driven SVR algorithm found it difficult to learn from the dataset which was reflected in the forecasting quality. A forecast is only as good as its training data meaning this time series was perhaps too difficult for a rudimentary SVR setup.

Both SVR examples also seemed to have a small lag with the series. Indeed between 5-10 hours, the predicted points rise around 2 hours after the actual rise. And at 20-24 hours, the actual series peaks and declines while the forecasts continue rising and near their peak. This similarity indicated that SVR in general struggled with the dataset. The results are displayed in Table 5.4.

Next, an ARIMA model was created with python code utilising the pmdarima library. This code struggled with seasonal data but is expected to be better with the stationary wind data. From analysis of the series, the P-value was found to equal 0.00014 indicating the series is indeed stationary. The best-fitting model was ARIMA(2,0,1) with the minimised AIC equalling -2435.5. The forecast using this model is displayed in Figure 5.11.

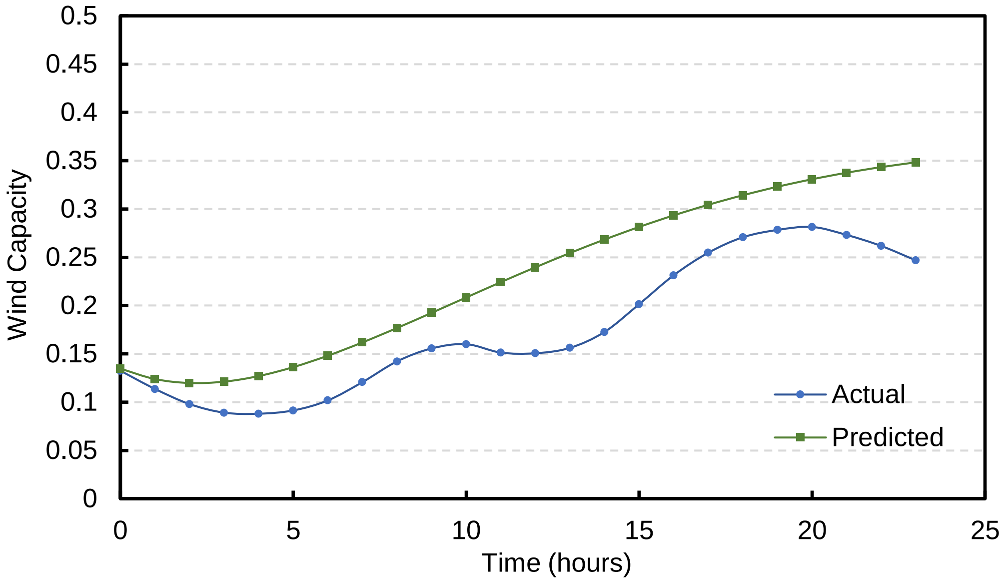

Figure 5.11 – Wind capacity forecast using ARIMA model via python code.

Figure 5.11 shows that the ARIMA model was very smooth and captured the general trend of the series best, but as with the SVR forecasts, it fails to capture the subtle fluctuations of the series. Notably, unlike SVR, the ARIMA forecast does not continue to rise at the end and does not show any apparent lag. This is expected as the literature review confirmed ARIMA does not overpredict during extreme time periods, with its smoothness for this forecast versus the SVRs sharp rise after 20 hours highlighting this. However, the overall model was too smooth and overpredicted all points, so it did not describe the training data in enough detail. Indeed, it did not follow certain sections (0-5 hours) as closely as SVR. This meant ARIMA achieved a MAPE of 32.3%, meaning all models had lacklustre performances for the wind series, indicating that training data is of paramount importance in influencing forecasting quality. That said, ARIMA still achieved the closest shape, which is expected due to SVR's reliance on high-quality historical data.

Furthermore, in comparison to the non-stationary customer demand ARIMA forecast, it was expectedly better here with stationary data, being able to capture the approximate shape/pattern of the series.

Table 5.4. Results for customer demand forecasts.

Method	MAPE	Prediction accuracy
SVR polynomial kernel	21.9	78.1
SVR chi-square kernel	24.8	75.2
ARIMA	32.4	67.6
Average	26.4	73.6

The forecasts of these two datasets revealed three important findings. First, both models were successful in capturing the shape of the customer demand time series, but SVR was slightly better due to its superiority in following the series closely, especially during extreme time periods with large peaks.

Secondly, the polynomial kernel was the best performing kernel across both datasets. This was expected as it amongst the most popular kernels used for SVR and it is more complex than its linear counterpart. The chi-square is rarely used in SVR and it did not show any notable performance in this study.

Thirdly, the training data quality heavily influences the forecasting quality. Where the customer sales series had predictable data with clear seasonality, all models were successful with an average PA of 87.8%, capturing the series' periodic patterns and trends. On the contrary, where the wind capacity series had no clear pattern or seasonality with lots of noise in the data, the forecasts were able to capture the general trend, but were less accurate with an average PA of 73.6%, failing to capture specific characteristics of the series. This is especially true for the data driven SVR which is heavily dependent on good historical data. ARIMA had the closest shape in the wind series, which was expected as it does better with stationary data, meanwhile SVR is data-driven and struggled closely capturing noisy, volatile data.

5.3 Sensitivity analysis

In this section, the hyperparameters C and ε will be varied to assess their effects on forecasting quality. While the kernel function counts as a hyperparameter, this has already been varied in the previous section, with results showing the polynomial kernel was the outstanding kernel function across both datasets. Furthermore, the ARIMA parameters were already optimised meaning they are already close to their optimal values, and further optimisation of ARIMA parameters will produce less meaningful results/changes to their forecast than SVR parameter optimisation. Hence, the sensitivity analysis shall focus on the polynomial kernel SVR forecasts for both series.

Firstly, the customer demand forecast will be analysed. As mentioned earlier, the polynomial kernel very closely matched the shape of the customer sales series including the muted fluctuations. However, it underpredicted all points by an almost equal amount suggesting that the forecast is shifted due to biases in the data. Thus, only ε will be varied as this will control the value of β, which determines the position of the graph via the regression equation. C needs not be varied as the model fitting is already close to the actual series. Figure 5.12 shows the effects of ε on the forecast.

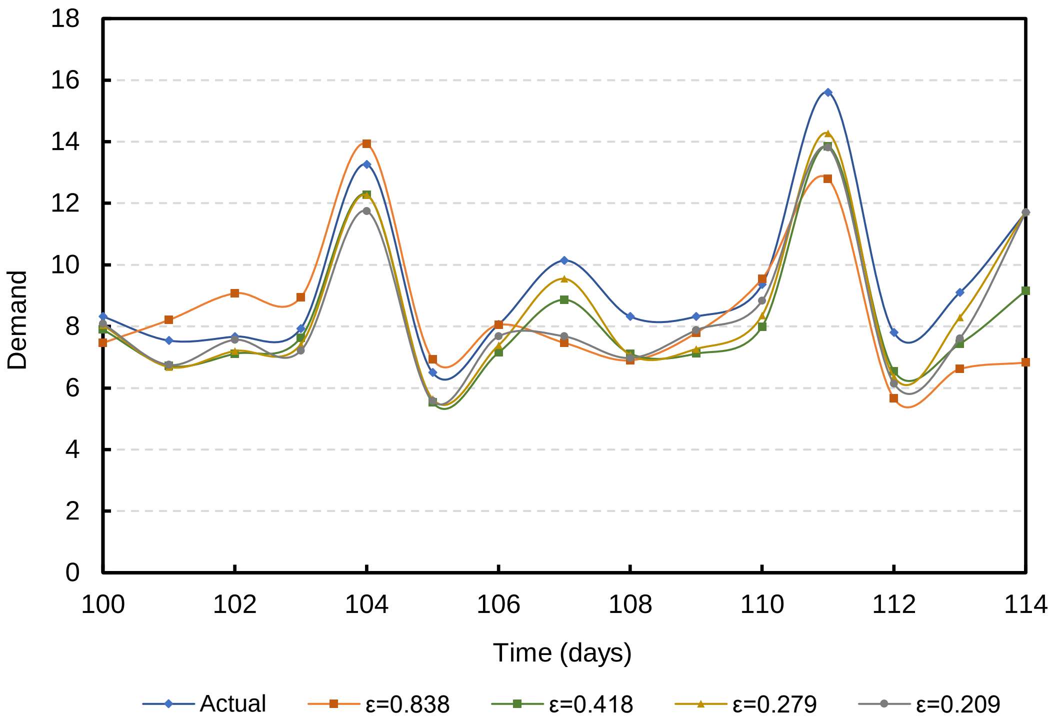

Figure 5.12 - Effect of varying ε on polynomial kernel SVR customer demand forecast

Figure 5.12 clearly demonstrates the influence of ε on the forecast quality. When k=20 and ε=0.838, ε is higher meaning the ε-sensitive tube includes more points and noise for the recursive regression function. As such, the forecast is more erratic and captures the shape of the actual series less efficiently than that of the original value of ε, presumably due to noisy data now being included which was not before.

When k=30 meaning ε=0.279, ε is smaller meaning the tube includes less points. While this can be negative for forecasting performance as it results in fewer training data, it also means the regression is more resistant to noise. Indeed, the forecast when ε=0.279 is marginally shifted upwards compared the original ε, suggesting that the original value perhaps included too much noise or caused a bias in the data than shifted the graph down.

While Eq. 4.3.2 for ε is for k=[10,30], k=40 will be investigated to see what happens with an even smaller value for ε. When k=40 (ε=0.209), the graph is visually less accurate, albeit only a small difference. This suggests that ε is now too small and thus eliminates too many datapoints to be considered, exacerbating the forecasting performance. Thus the values of k=30 and ε=0.279 are the most optimal value to maximise the forecasting performance of the polynomial kernel SVR.

Notably, the variation in ε also changed β, where a lower ε resulted in a lower β. However, the difference was minimal meaning it did not have as much effect on the forecast compared to other factors.

The MAPE values for each variation are tabulated below, showing that the smaller values of ε are more accurate with ε=0.279 being the most optimal value to reduce data biases, decreasing MAPE by a difference of 3.33, bringing the PA to 90.2%. This shows that the variation of ε has helped optimise this forecast.

Table 5.5 MAPE values when ε is varied for polynomial kernel SVR customer demand forecast

Value of ε	Value of k	Value of β	MAPE	Prediction accuracy
0.838	10	4.00	16.3	83.7
0.418	20	3.58	13.1	86.9
0.279	30	3.35	9.77	90.2
0.209	40	3.22	11.0	89.0

Next, the wind capacity series will be analysed. Again, only the polynomial kernel SVR is analysed as it showed the best overall performance for both datasets, in terms of its accuracy (MAPE) and shape of forecast.

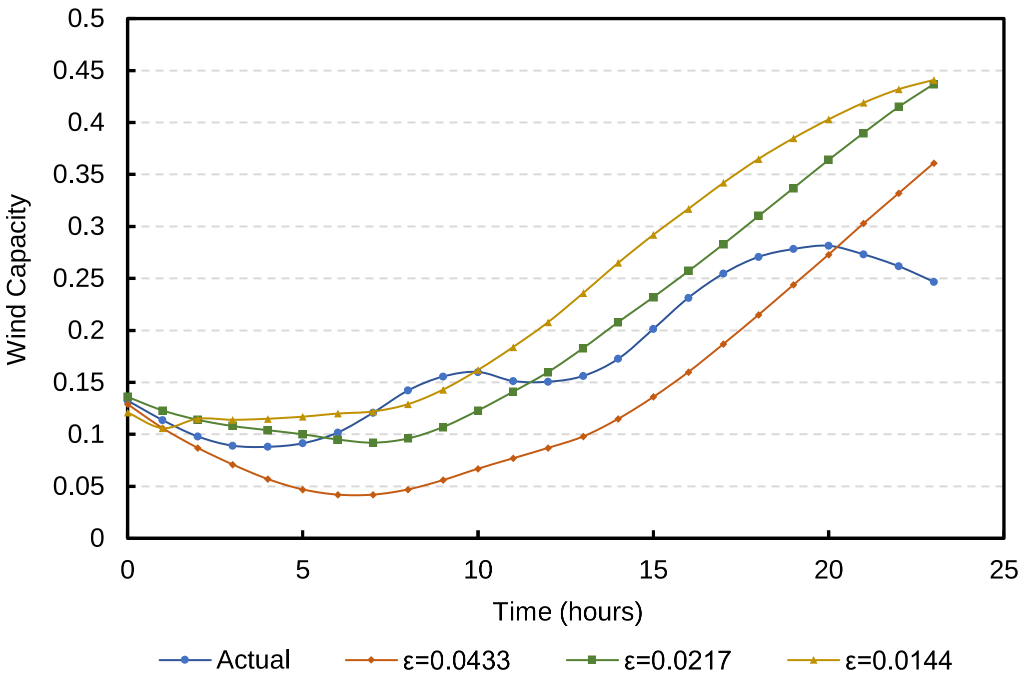

Figure 5.13 - Effect of varying ε on polynomial kernel SVR wind capacity forecast

The numerical results for each variation are displayed below.

Table 5.6 MAPE values when ε is varied for polynomial kernel SVR wind forecast

Value of ε	Value of k	Value of β	MAPE	Prediction accuracy
0.0433	10	0.110	33.7	66.3
0.0217	20	0.045	21.9	78.1
0.0144	30	0.032	31.3	68.7

The table and figure show that the position of the SVR wind forecast is more sensitive to changes in ε than the SVR customer demand forecast. Indeed, Figure 5.13 shows that as ε varies, the forecast shape remains similar however the position shifts upwards as ε gets larger. Notably, when ε=0.0144, the forecast generally underpredicts for much of the series and when ε=0.0433, the forecast overpredicts. As such, k=20 meaning ε=0.0217 is the ideal spot where the forecast does not excessively underpredict or overpredict. Indeed, Table 5.6 shows the MAPE for ε=0.0217 is best, while for the large and small variations, the MAPE increases by 11.8 and 9.4 respectively.

Table 5.6 also shows that β changes more significantly with ε in the wind series than in the customer sales series. This explains why there is a larger change in the positions of each forecast versus the customer sales series where deviations in ε only changed β slightly, resulting in the forecasts being shifted less.

Finally, since the shapes/patterns of the forecast did not change or become closer to the actual series, it indicates that the deviations in ε had a marginal effect on the training data to be included in the regression. Thus, the wind series already had a large degree of noise, observed in Figure 5.8, such that variations in ε did not substantially affect the amount of noise or quality of data that SVR learned from to create a forecast. On the contrary, for the sensitivity analysis of the customer demand forecast, the deviations in ε had a greater, albeit still small, effect on the amount of noise and quality of data that SVR used to learn from, evidenced by the improvement in MAPE when ε was made smaller.

Next, C (regularisation factor) for the polynomial kernel SVR wind forecast was varied, with results reported in Table 5.7.

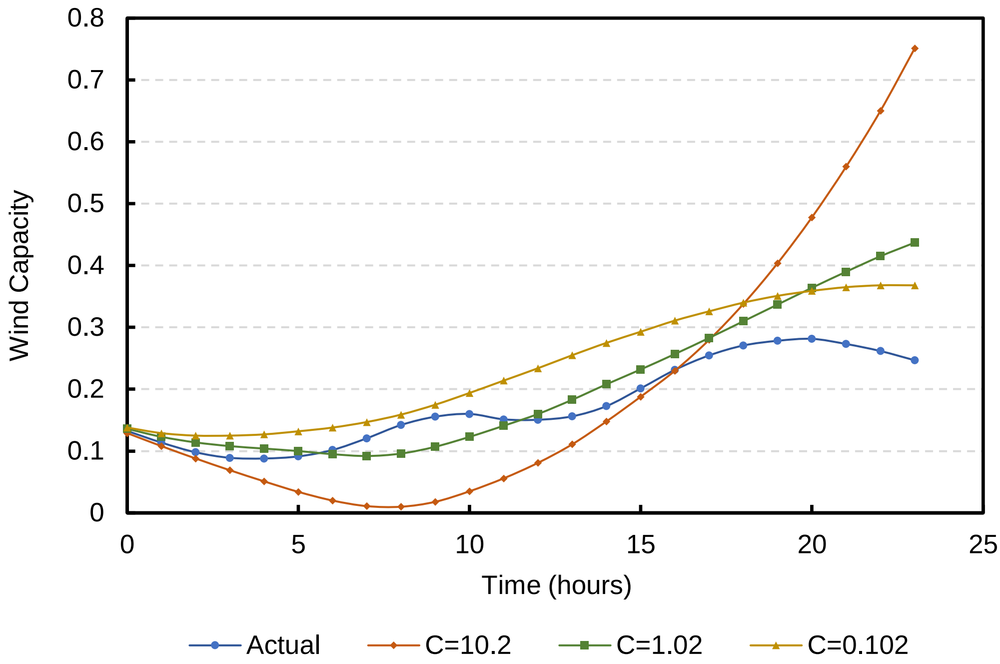

Figure 5.14 - Effect of varying C on polynomial kernel SVR wind capacity forecast

Table 5.7 MAPE values when C is varied for polynomial kernel SVR wind forecast

Value of C	MAPE	Prediction accuracy
10.2	56.0	44.0
1.02	21.9	78.1
0.102	33.6	66.4

The results are expected. When C is large (C=10.2), the forecast becomes more unbalanced and sensitive. Where there are minor fluctuations in the actual series, the forecast now overestimates their change; this is apparent at 0-15 hours and from 20 hours. Indeed, when C is excessively large, the model loses its smoothness and too much weight is prescribed to minimisation of error leading to over-fitting, which is observed in Figure 5.14. Because of this, the MAPE is 56.0% showing that C=10.2 is far from an optimal value, causing deficient performance.

Contrarily, when C is smaller (C=0.102), the forecast has a different shape. It follows the same general trend as the actual series, but less erratically and wildly. Instead, it is a lot smoother. This is again expected, as when C is small, the model will follow the training data less stringently making the forecast smoother, and if C is too low, then underfitting can occur. While there is some underfitting as the forecast is quite off with a MAPE of 33.6%, it still has better performance than when C=10.2, meaning that the optimal value is closer to a smaller value of C than a large C. Table 5.7 also reveals the initial value of C (C=1.02), calculated from the Cherkassky and Ma heuristic rule, has the best performance proving the rule is an effective, simple way of determining an acceptable value for the hyperparameter[29].

However, all variations still struggle to capture the exact shape and intricacies of the wind series. As discussed, the reason for this is the quality of training data. When SVR parameters are varied, it affects how SVR learns from and treats the data; it does not affect the data itself. This means that if the data is noisy and low-quality, such as the wind data, then variation of SVR parameters will not substantially improve the forecast as the forecast still depends on the data. Thus, before producing models to forecast a given time series, it must be ensured that the data is of good quality otherwise it can be considered futile as a forecast is only as good as its training data, especially for a data driven model like SVR.

6. Limitations and considerations for future research

The first limitation is the SVR parameter optimisation. There is yet to be a formal framework for SVR parameter selection, so there are many methods to choose from. The most popular in research are cross-validation and grid search procedures[58,60]. However, these are computationally complex and intensive, for instance the Bayesian criteria having a complexity of $O(N^3)$ where N is the number of training datapoints[61]. Thus, simpler heuristic rules were used for hyperparameter selection, providing less optimal parameters values[29]. Essentially this is a trade-off between optimal parameter values and computational demand. Thus, this is a limitation since more optimal parameter values could have been chosen providing a superior forecast. Therefore, when revisiting this topic in the future, cross-validation or grid search procedures can be used. It would be of interest to evaluate the trade-off of whether the added computational intensity is worth the optimised forecasting performance.

Secondly, ANNs could be investigated in future research. The literature review revealed that ANNs show strong performance across a variety of implementations, at the cost of difficult parameter selection and optimisation, and general computational complexity. Furthermore, ANNs can be troublesome for complex processes, such as for chemical engineering applications, due to the parameter selection and deciding how many parameters to include[55]. This is again a computational demand issue. Therefore, similarly with grid search parameter optimisation, ANNs can be researched in future work to determine whether their performance warrants their increased complexity versus models such as SVR or ARIMA.

Hybrid models are another area of interest but fall outside the scope of this study. While they show exceptional results compared to singular models in literature, they have seen minimal usage due to their specialisation and need more formal development before seeing commercial usage. Thus, future research into their performance and viability can be an area of interest

7. Conclusion

This paper successfully examined the implementation and performance of SVR forecasting with relevant comparisons made to ARIMA models. A literature review was conducted to assess the role and importance of forecasting in chemical engineering and other industries, finding that while SVR showed good results in literature, its industrial use was limited with conventional models such as ARIMA being dominantly used. An SVR approach on GAMS utilising the dual problem, kernel trick and recursive regression function was formulated due to its relative simplicity, while python code was used for the ARIMA model, with the aim of comparing results to assess each model's performance for two time series: a predictable seasonal customer demand dataset and an unpredictable, volatile wind capacity series. Furthermore, SVR hyperparameters including C, ε and kernel functions were varied to evaluate their impact on forecasting quality. It was found that SVR using a polynomial kernel function achieved the best performance across both time series. For the customer demand series, it was able to closely capture the trends and patterns of the actual series achieving a low 9.77% MAPE, while ARIMA was less accurate, struggling with the extreme peak periods achieving a MAPE of 13%. For the wind capacity series, both models faltered as the dataset was volatile and noisy, showcasing that a forecast is heavily dependent on the data it learns from. However, SVR using a polynomial kernel achieved the best result with a 21.9% MAPE, proving it was the top overall model configuration in this study, highlighting its potential as a computationally simple but effective model. However, further research should be performed in developing a formal, established framework for SVR parameter selection before it can be considered an industrially preferred method, along with comparative studies with other state-of-the-art models such as ANNs and hybrids.

References

1. Cruz, A., Muñoz, A., Zamora, J. L. & Espínola, R. The effect of wind generation and weekday on Spanish electricity spot price forecasting. *Electr. Power Syst. Res.* **81**, 1924–1935 (2011).
2. Lasek, A., Cercone, N. & Saunders, J. Smart restaurants: survey on customer demand and sales forecasting. in *Smart Cities and Homes* (eds. Obaidat, M. & Nicopolitidis, P.) 361–386 (Elsevier, 2016). doi:10.1016/B978-0-12-803454-5.00017-1.
3. Weron, R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *Int. J. Forecast.* **30**, 1030–1081 (2014).
4. Gardner, E. S. Exponential smoothing: The state of the art. *J. Forecast.* **4**, 1–28 (1985).
5. Geurts, M., Box, G. E. P. & Jenkins, G. M. Time Series Analysis: Forecasting and Control. *J. Mark. Res.* **14**, 269 (1977).
6. Hyndman, R. J. & Athanasopoulos, G. *Forecasting: principles and practice*. (OTexts, 2018).
7. Vapnik, V. *Statistical Learning Theory*. (Wiley, 1998).
8. Gardner, E. S. Exponential smoothing: The state of the art—Part II. *Int. J. Forecast.* **22**, 637–666 (2006).
9. Schnaars, S. P. A comparison of extrapolation models on yearly sales forecasts. *Int. J. Forecast.* **2**, 71–85 (1986).
10. Makridakis, S. *et al.* The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *J. Forecast.* **1**, 111–153 (1982).
11. Assimakopoulos, V. & Nikolopoulos, K. The theta model: a decomposition approach to forecasting. *Int. J. Forecast.* **16**, 521–530 (2000).
12. Hyndman, R. J. & Billah, B. Unmasking the Theta method. *Int. J. Forecast.* **19**, 287–290 (2003).
13. Makridakis, S. & Hibon, M. The M3-Competition: results, conclusions and implications. *Int. J. Forecast.* **16**, 451–476 (2000).
14. Fattah, J., Ezzine, L., Aman, Z., El Moussami, H. & Lachhab, A. Forecasting of demand using ARIMA model. *Int. J. Eng. Bus. Manag.* **10**, 184797901880867 (2018).
15. Permatasari, C. I., Sutopo, W. & Hisjam, M. Sales forecasting newspaper with ARIMA: A case study. in 030017 (2018). doi:10.1063/1.5024076.
16. Mircetic, D., Nikolicic, S., Maslaric, M., Ralevic, N. & Debelic, B. Development of S-ARIMA Model for Forecasting Demand in a Beverage Supply Chain. *Open Eng.* **6**, (2016).

17. Lagarto, J., de Sousa, J., Martins, A. & Ferrao, P. Price forecasting in the day-ahead Iberian electricity market using a conjectural variations ARIMA model. in *2012 9th International Conference on the European Energy Market* 1–7 (IEEE, 2012). doi:10.1109/EEM.2012.6254734.

18. Weron, R. & Misiorek, A. Forecasting Spot Electricity Prices with Time Series Models. *Int. Conf. "The Eur. Electr. Mark. EEM-05"* 8 (2005).

19. Contreras, J., Espinola, R., Nogales, F. J. & Conejo, A. J. ARIMA models to predict next-day electricity prices. *IEEE Trans. Power Syst.* **18**, 1014–1020 (2003).

20. Crespo Cuaresma, J., Hlouskova, J., Kossmeier, S. & Obersteiner, M. Forecasting electricity spot-prices using linear univariate time-series models. *Appl. Energy* **77**, 87–106 (2004).

21. Che, J. & Wang, J. Short-term electricity prices forecasting based on support vector regression and Auto-regressive integrated moving average modeling. *Energy Convers. Manag.* **51**, 1911–1917 (2010).

22. Levis, A. A. & Papageorgiou, L. G. Customer demand forecasting via support vector regression analysis. *Chem. Eng. Res. Des.* **83**, 1009–1018 (2005).

23. Tarsitano, A. & Amerise, I. L. Short-term load forecasting using a two-stage sarimax model. *Energy* **133**, 108–114 (2017).

24. Nogales, F. J. & Conejo, A. J. Electricity price forecasting through transfer function models. *J. Oper. Res. Soc.* **57**, 350–356 (2006).

25. Vagropoulos, S. I., Chouliaras, G. I., Kardakos, E. G., Simoglou, C. K. & Bakirtzis, A. G. Comparison of SARIMAX, SARIMA, modified SARIMA and ANN-based models for short-term PV generation forecasting. in *2016 IEEE International Energy Conference (ENERGYCON)* 1–6 (IEEE, 2016). doi:10.1109/ENERGYCON.2016.7514029.

26. Arunraj, N. S., Ahrens, D. & Fernandes, M. Application of SARIMAX Model to Forecast Daily Sales in Food Retail Industry. *Int. J. Oper. Res. Inf. Syst.* **7**, 1–21 (2016).

27. Drucker, H., Surges, C. J. C., Kaufman, L., Smola, A. & Vapnik, V. Support vector regression machines. *Adv. Neural Inf. Process. Syst.* **1**, 155–161 (1997).

28. Hao, W. & Yu, S. Support Vector Regression for Financial Time Series Forecasting. in *Knowledge Enterprise: Intelligent Strategies in Product Design, Manufacturing, and Management* 825–830 (Springer US). doi:10.1007/0-387-34403-9_115.

29. Cherkassky, V. & Ma, Y. Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks* **17**, 113–126 (2004).

30. Lahiri, S. K. & Ghanta, K. C. The Support Vector Regression with the parameter tuning assisted by a differential evolution technique: Study of the critical velocity of a slurry

flow in a pipeline. *Chem. Ind. Chem. Eng. Q.* **14**, 191–203 (2008).

31. Awad, M. & Khanna, R. Support Vector Regression. in *Efficient Learning Machines* 67–80 (Apress, 2015). doi:10.1007/978-1-4302-5990-9_4.

32. Ito, K. & Nakano, R. Optimizing Support Vector regression hyperparameters based on cross-validation. in *Proceedings of the International Joint Conference on Neural Networks, 2003.* vol. 3 2077–2082 (IEEE).

33. Smola, A., Schölkopf, B. & Müller, K.-R. Convex Cost Functions for Support Vector Regression. in 99–104 (1998). doi:10.1007/978-1-4471-1599-1_10.

34. Soukissian, T. H. & Karathanasi, F. E. On the selection of bivariate parametric models for wind data. *Appl. Energy* **188**, 280–304 (2017).

35. Samal, R. K. & Tripathy, M. Estimating wind speed probability distribution based on measured data at Burla in Odisha, India. *Energy Sources, Part A Recover. Util. Environ. Eff.* **41**, 918–930 (2019).

36. Ye, H., Ma, X., Yang, T. & Hou, Y. Comparative Study on the Performance Prediction of Fuel Cell Using Support Vector Machine with Different Kernel Functions. in 337–351 (2020). doi:10.1007/978-981-13-9718-9_25.

37. Krishnamurthy, K., Manoharan, S. C. & Swaminathan, R. A Jacobian approach for calculating the Lyapunov exponents of short time series using support vector regression. *J. Ambient Intell. Humaniz. Comput.* **11**, 3329–3335 (2020).

38. Sansom, D. C., Downs, T. & Saha, T. K. Evaluation of support vector machine based forecasting tool in electricity price forecasting for Australian national electricity market participants. *J. Electr. Electron. Eng. Aust.* **22**, 227–233 (2002).

39. Tealab, A. Time series forecasting using artificial neural networks methodologies: A systematic review. *Futur. Comput. Informatics J.* **3**, 334–340 (2018).

40. Lasek, A., Cercone, N. & Saunders, J. Restaurant Sales and Customer Demand Forecasting: Literature Survey and Categorization of Methods. in 479–491 (2016). doi:10.1007/978-3-319-33681-7_40.

41. Jain, A. K., Jianchang Mao & Mohiuddin, K. M. Artificial neural networks: a tutorial. *Computer (Long. Beach. Calif).* **29**, 31–44 (1996).

42. Chen, X. *et al.* Electricity Price Forecasting With Extreme Learning Machine and Bootstrapping. *IEEE Trans. Power Syst.* **27**, 2055–2062 (2012).

43. Gareta, R., Romeo, L. M. & Gil, A. Forecasting of electricity prices with neural networks. *Energy Convers. Manag.* **47**, 1770–1778 (2006).

44. Mandal, P., Senjyu, T. & Funabashi, T. Neural networks approach to forecast several hour ahead electricity prices and loads in deregulated market. *Energy Convers. Manag.* **47**, 2128–2142 (2006).

45. YAMIN, H., SHAHIDEHPOUR, S. & LI, Z. Adaptive short-term electricity price forecasting using artificial neural networks in the restructured power markets. *Int. J. Electr. Power Energy Syst.* **26**, 571–581 (2004).

46. MateoGonzalez, A., MunozSanRoque, A. & Garcia-Gonzalez, J. Modeling and Forecasting Electricity Prices with Input/Output Hidden Markov Models. *IEEE Trans. Power Syst.* **20**, 13–24 (2005).

47. -, P. C., -, C. Lo, -, H. C. & -, Y. L. A Study of Applying Artificial Neural Network and Genetic Algorithm in. *J. Converg. Inf. Technol.* **6**, 352–362 (2011).

48. Alon, I., Qi, M. & Sadowski, R. J. Forecasting aggregate retail sales: *J. Retail. Consum. Serv.* **8**, 147–156 (2001).

49. Kang, S. . An Investigation of the Use of Feedforward Neural Networks for Forecasting. (Kent State University, 1991).

50. Donkor, E. A., Mazzuchi, T. A., Soyer, R. & Alan Roberson, J. Urban Water Demand Forecasting: Review of Methods and Models. *J. Water Resour. Plan. Manag.* **140**, 146–159 (2014).

51. Ghiassi, M., Zimbra, D. K. & Saidane, H. Urban Water Demand Forecasting with a Dynamic Artificial Neural Network Model. *J. Water Resour. Plan. Manag.* **134**, 138–146 (2008).

52. Bougadis, J., Adamowski, K. & Diduch, R. Short-term municipal water demand forecasting. *Hydrol. Process.* **19**, 137–148 (2005).

53. Senthil Kumar, A. R., Sudheer, K. P., Jain, S. K. & Agarwal, P. K. Rainfall-runoff modelling using artificial neural networks: comparison of network types. *Hydrol. Process.* **19**, 1277–1291 (2005).

54. Firat, M., Yurdusev, M. A. & Turan, M. E. Evaluation of Artificial Neural Network Techniques for Municipal Water Consumption Modeling. *Water Resour. Manag.* **23**, 617–632 (2009).

55. Bhat, N. V. & McAvoy, T. J. determining model structure for neural models by network stripping. *Comput. Chem. Eng.* **16**, 271–281 (1992).

56. Smola, A. J. & Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **14**, 199–222 (2004).

57. Chujai, P., Kerdprasop, N. & Kerdprasop, K. Time series analysis of household electric consumption with ARIMA and ARMA models. *Lect. Notes Eng. Comput. Sci.* **2202**, 295–300 (2013).

58. Rubio, G., Pomares, H., Rojas, I. & Herrera, L. J. A heuristic method for parameter selection in LS-SVM: Application to time series prediction. *Int. J. Forecast.* **27**, 725–739 (2011).

59. Lima, A. R., Cannon, A. J. & Hsieh, W. W. Nonlinear regression in environmental sciences by support vector machines combined with evolutionary strategy. *Comput. Geosci.* **50**, 136–144 (2013).

60. An, S., Liu, W. & Venkatesh, S. Fast cross-validation algorithms for least squares support vector machine and kernel ridge regression. *Pattern Recognit.* **40**, 2154–2162 (2007).

61. Van Gestel, T. *et al.* Financial time series prediction using least squares support vector machines within the evidence framework. *IEEE Trans. Neural Networks* **12**, 809–821 (2001).