Winter 2023

Evaluating the Potential of Ensemble Learning for One Day-Ahead Forecasting of Power System Demand in Ireland

Karol Skowronski
CCT College Dublin

Evaluating the Potential of Ensemble Learning for One Day-Ahead Forecasting of Power System Demand in Ireland

Karol Skowronski

A Thesis Submitted in Partial Fulfilment

of the requirements for the

Degree of

Master of Science in Data Analytics

September 2023

Supervisor: Vikas Tomer

Abstract

Accurate One Day-Ahead Demand Forecasting (ODADF) is crucial for electrical network reliability, the environment, and trading markets. While individual models face challenges in achieving accurate predictions, ensemble learning models have emerged as potential solution. They have achieved success in ODADF in several countries; however, there has been no research conducted for the Irish power system. Therefore, research objectives were formed, to develop a framework of ensemble learning models, evaluate their performance, and examine their potential for ODADF in Ireland, to fill the gap. Experimentation, and CRISP-DM were selected as primary research methodology, and project management framework, respectively. The development of the framework considered a balance between performance and computational complexity of the configurations. Three stacking approaches were considered, such as classifiers and regressors as meta-learners, and heuristic rules. Various potential base-learners were considered, and two methods of supervised problem creation, based on Similar Day (SD) and Moving Window (MW) approaches, were proposed to enhance their pattern recognition in data. The cause-and-effect relationship between ensemble configurations and performance metrics for ODADF in Ireland was established, and the integration method emerged as the primary causal variable.

The research methodology was divided into three phases, such as data preparation, experimentation with ensembles architectures, and validation of results. Data preparation included temporal features extraction, Daylight-Saving Time removal, and replacement of missing data and outliers. The results were validated by performance metrics, visual comparison to SDs from neighbouring weeks, and distributions before and after the processing. Investigations into lagged weekly and daily demand, and window size were performed for SD and MW approaches, respectively. Following investigation into correlation between lagged weather variables and demand; temperature, relative humidity and wind speed, lagged by 39-hours were selected as exogenous features. As weather data was distributed locally, three approaches for representative stations were proposed. Scaling of time series, and encoding of temporal features to cyclical and vector formats, were found beneficial to ODADF by correlation study and distributions comparison. Feature selection was performed separately for SD and MW approaches. Given that data from year 2020 was found to be an outlier, datasets were split primarily into training and testing datasets, covering years 2014-2019, and 2021-2022, respectively. Experimentation with base-learners and three integration methods was performed. Training and testing datasets were further split into training and validation subsets, covering years 2014-2018 and 2019, and 2021-2022, respectively. Bayesian optimisation with 10-fold cross-validation was selected for hyperparameters tuning. Potential base-learners were tuned, trained and evaluated on training datasets, and the twenty most promising ones were selected as base-learners. They showed fluctuations in their MAPE across different days of the week, months and hours. Potential ensembles were tuned, trained and evaluated on base-learners' predictions for years 2015-2019 and training datasets, respectively. In the validation phase, base-learners and classification-based ensembles with hyperparameters inferred from previous phase, were refitted on unseen data, and the base-learners' predictions for years 2021-2022, respectively, and evaluated on year 2022.

The results proved the high potential of classification-based ensembles for ODADF in Ireland. Ensembles of twenty base-learners, with SVM and MLP classifiers as meta-learners, stood out as the most effective solution for ODADF in Ireland. They both achieved the lowest MAPE 1.91%, which was 11.2% improvement in comparison to the best base-learner, SVM (SD) registering MAPE 2.15%. While introduction of SD and MW approaches amplified the diversity of the base-learners' predictions, incorporating virtual weather stations benefited the performance of classification-based ensembles. They not only harnessed the combined strengths but also mitigated the potential inconsistencies found in individual base-learners, achieving predictions aligned with the distribution of actual demand. Finally, while this research addressed the gap in knowledge, further work, using wider variety of base-learners and their integration methods, is needed to comprehensively bridge this gap.

GitHub: https://github.com/karol-skowronski/Capstone **Wordcount: 17,383**

Contents

Abbreviations

ACF	Autocorrelation (Seasonality Test)
ADF	Augmented Dickey-Fuller (Stationarity Test)
API	Application Programming Interface
ANN	Artificial Neural Networks
AR	Auto-Regression
ARIMA	Auto-Regressive Integrated Moving Average
BL	Base-Learner
BSR	Backward Selection (Elimination) Regression
CNN	Convolutional Neural Networks
CSV	Comma-Separated Values
DES	Double Exponential Smoothing
DL	Deep Learning
DST	Daylight-Saving Time
DT	Decision Trees
EL	Ensemble Learning
ETS	Exponential Smoothing
FSR	Forward Selection Regression
GBM	Gradient Boosting Machines
HWES	Holt-Winters Exponential Smoothing
JSON	JavaScript Object Notation
k-NN	k-Nearest Neighbours
KDE	Kernel Density Estimate
KPSS	Kwiatkowski-Phillips-Schmidt-Shin (Stationarity Test)
LR	Linear Regression
LSTM	Logh Short-Term Memory
MA	Moving Average
MAPE	Mean Average Percentage Error (Performance Metric)
ML	Machine Learning
MLP	Multi-Layer Perceptron
MLR	Multi-Linear Regression
MSE	Mean Square Error (Performance Metric)
MSTL	Multiple Seasonal Pattern Decomposision
MW	Moving Window
NBEATS	Neural Basis Expansion Analysis for Time Series
ODADF	One Day-Ahead (Power System) Demand Forecasting
OHE	One-Hot Encoder
PACF	Partial Autocorrelation (Seasonality Test)
R^2	Coefficient of Determination
RES	Renewable Energy Sources
RF	Random Forest
RMSE	Root Mean Square Error (Performance Metric)
RNN	Recurrent Neural Networks
SD	Similar Day
SES	Single Exponential Smoothing
STLF	Short-Time Load Forecasting
SVM	Support Vector Machines
TBATS	Trigonometric Seasonal decomposition of Time series
Virtual_Lasso	Virtual Weather Station created using Lasso Regression
Virtual_LR	Virtual Weather Station created using Linear Regression

Figures

Tables

1. Introduction

1.1. Background

Short-Time Load Forecasting (STLF) refers to the prediction of power demand within a relatively short time horizon, typically spanning hours to a few days ahead. One Day-Ahead Demand Forecasting (ODADF) is a particular case of STLF, where the hourly mean of power system demand is predicted for the following day. Accurate ODADF is crucial for electrical network reliability, the environment, and trading markets. Instantaneous balance, between generation and demand of electricity, must be maintained by system operators at all times. This is the responsibility of the **EirGrid** in the Republic of Ireland, who operates the electricity grid, including interconnection to neighbouring grids, and running the wholesale electricity market. Implications of imbalance are negative, both technically and financially. Many state-of-the-art models struggle to achieve accurate predictions in the constantly changing conditions, in which the power system operates. The growth in weather-dependent renewable sources (RES), coupled with climate change, the advent of contemporary appliances, and changes in customer habits, challenge the system operation. While the work of conventional power plants can be controlled at system level, RESs are solely dependent on weather conditions. Moreover, low-scale RES generation, connected to the grid at distribution level, while being outside of the system operator's control, decreases customer power demand from the grid. Furthermore, the increasing number of electric vehicles, combined with changes in customer habits due to remote working, alters the demand patterns.

The dynamic conditions of power system operations challenge many state-of-the-art STLF methods, ODADF in particular (Table 1.1). Traditional models are unable to capture non-linear patterns, and are sensitive to initial conditions, outliers and abrupt changes in data. Auto-regressive models assume linear relationship between past and future observations, require stationary time series, and cannot capture sudden changes in patterns. Machine learning models suffer from **system depended phenomenon**, lack of interpretation, and require challenging selection of appropriate features and hyperparameters. Deep learning models require selection of appropriate architecture, hyperparameters, and training strategies, which is challenging and significantly impacts their performance. They have higher standard deviation of accuracy and sensitivity to reduction of training data compared to machine learning models. Moreover, they require large amount of training data and high computational resources. Hybrid models require complex process of selecting and integrating individual models to attain efficiency. Moreover, determining optimal combination of weights or rules of integration requires careful experimentation and validation, and hyperparameter tuning is complex. They need more data and are more computationally expensive than their components.

Authors (year)	Dataset (size)	Model	Reported Results		Limitations
			Metric	Value	
Sethi and Kleiss (2020)	US (1 year)	HWES	MAPE	4.60%	Unable to capture non-linear patterns. Need careful selection of parameters.
Shahare et al (2023)	Metropolitan Area (5 years)	SES	MAE	1.400	Sensitive to initial conditions, outliers and abrupt changes in data.
Dudek (2015)	Poland (2 years)	ARIMA	MAPE	2.42%	Assumes linear relationship between past and future observations.
Al-Musaylh et al (2018)	Australia (5 years)			2.16%	Require stationary time series.
Sethi and Kleissl (2020)	US (1 year)	ARIMA SARIMA		3.80% 3.20%	Unable to capture sudden changes in patterns in data.
Dudek (2015)	Poland (2 years)	MLR	MAPE	2.81%	Long training time.
Tsekouras, Kanellos and Mastorakis (2015)	Greece (1-7 years)	MLP		1.78%	Not possible to generalise findings due to system depended phenomenon.
Al-Musaylh et al (2018)	Australia (5 years)	MARS SVR		1.27% 1.71%	Increased memory requirements
Foster (2020)	N. Ireland (7 years)	MLR MLP		2.53% 2.62%	Challenging selection of appropriate features and hyperparameters, which significantly impacts model performance.
Foster (2020)	New York (6 years)			4.34% 3.71%	
Kathirgamanathan et al (2022)	New England (6 years)	MLP		4.14%	High computational complexity of SVR and MLP models.
Osowski et al (2022)	Poland (6 years)	SVR MLP RBF		2.17% 2.08% 2.26%	Difficult to understand underlying relationship between the input features and predicted values.
Shahare et al (2023)	M. Area (5 years)	SVR MLP	MAE	0.088 0.079	Increased memory requirements.
Han et al (2019)	China (4 years)	CNN LSTM	MAPE	3.30% 2.70%	Sensitive to reduction of data compared to MLP models, with LSTM being most affected.
Sethi and Kleissl (2020)	US (1 year)	LSTM		2.10%	
Foster (2020)	N. Ireland (7 years)	LSTM CNN		3.65% 3.02%	Challenging selection of appropriate architecture, hyperparameters, and training strategies, which significantly impacts their performance.
Foster (2020)	New York (6 years)			4.12% 3.71%	
Osowski et al (2022)	Poland (6 years)	LSTM		1.63%	Require large amount of training data and high computational resources
Kathigamanathan et al (2022)	New England (6 years)	CNN LSTM		3.91% 3.89%	High standard deviation of accuracy and sensitivity to reduction of data, with LSTM being most affected.
Dudek (2015)	Poland (2 years)	LR+DT	MAPE	2.53%	Challenging selection and integration of individual models.
Dudek (2016)	Poland (4 years)	SD+PCR SD+PLSR		1.34% 1.35%	Determining optimal combination of weights or rules of integration requires careful experimentation and validation.
Dudek (2016)	Australia (4 years)			2.83% 3.00%	
He (2017)	China (3 years)	CNN+LSTM		1.41%	Complex hyperparameter tuning. Computationally expensive.
Shahare et al (2023)	M. Area (5 years)		MAE	0.029	Interpretation more challenging than their components.

Table 1.1. Summary of recent research in one day-ahead power demand forecasting by traditional, auto-regressive, machine learning, deep learning and hybrid models, published after year 2014.

The decreased accuracy of ODADF negatively impacts system reliability, the environment, and translates to financial losses of system operators. Their customers, however, are interested in clean energy, delivered to their premises uninterruptedly, and at low cost. The most promising solution to increase ODADF accuracy is to use an ensemble of individual predictors. Motivated by the importance and currency of the problem, as well as the gap in the research for Ireland, by doing this project I hope to establish, that an optimal ensemble learning approach can not only be successfully applied for ODADF in Ireland, but also combine the strengths, and offset the limitations of both, single and hybrid models. Fortunately, ensemble learning has been applied successfully to increase performance of forecasting in many areas. Moreover, that approach has already been proven to work for STLF and ODADF in several countries. However, there is no research in that area for the power system in Ireland. Hence, there is an opportunity for this research to fill that gap.

1.2. Research Objectives

The primary aim of this research is to experiment with ensemble learning models' architectures in order to evaluate their potential to ODADF in Ireland. Therefore, the research objectives are:

• **To develop a framework of ensemble learning models for ODADF in Ireland.** This includes pre-selection of variety of machine and deep learning models, including those able to incorporate temporal and weather factors, by evaluating their performance, and comparing their metrics. Subsequently, a framework of ensemble learning models that combines predictions of preselected base-learners and their integrations techniques will be developed for ODADF in Ireland. The framework will provide a foundation for the primary research.

• **To examine the performance of ensembles to ODADF in Ireland by comparing their metrics.** This includes experimentation with selected architectures of ensemble models and measurement of influence of their configuration to ODADF performance metrics. Each configuration will be trained, validated and tested using real-world data, and its performance metrics will be automatically recorded as function of day of the week, month and hour. The performance dataset will enable the examination of cause-and-effect relationships between variables and allow further examination of the potential in ensemble learning architectures to ODADF in Ireland.

• **To evaluate the potential of ensemble learning application to ODADF in Ireland.** This includes evaluation of performance metrics of ensemble learning models against each other and their base-learners, which will enable the examination of their potential to ODADF in Ireland. Subsequently, based on findings of the research, the most optimal ensemble learning architectures will be determined and validated on unseen data. Finally, limitation of the research and recommendations for future research and development will be provided.

1.3. Outline of the Thesis

The remainder of this thesis is organised as follows. Chapter 2 presents an overview of the literature. Chapter 3 covers methodology, including project management framework, limitations and ethical considerations of this research. Chapter 4 describes power demand and weather factors data collection, initial data extraction, exploration, and detection of distortions, missing values and outliers in time series. Chapter 5 covers power demand data preparation, including temporal features extraction, daylight-saving time distortion removal, missing data and outliers' replacement, exploratory data analysis, followed by definition of supervised learning problems from time series, using similar day and moving window approaches. Chapter 6 presents multivariate exploratory data analysis, correlations between lagged weather variables and power demand, finding representative weather stations, feature selection, followed by supervised learning problems creation. Chapters 7-9 present results of experimentation, including base-learners modelling and evaluation in Chapter 7, modelling and evaluation of various ensembles in Chapter 8, and validation of classification-based ensembles' potential to ODADF in Ireland in Chapter 9. Chapter 10 concludes the thesis by summarising the research and its limitations, and providing recommendations for future investigation.

2. Literature Review

One Day-Ahead Demand Forecasting (ODADF) is the most prominent application of data analytics in the electricity sector, crucial for system reliability and trading markets (Scheidt *et al.*, 2020). Instantaneous balance between supply and demand of electricity must be maintained by power system operators at all time (Hong and Fan, 2016). Implications of imbalance are negative both technically and financially. While overestimation leads to starting excessive number of generators, underestimation forces purchasing energy abroad at higher price (Foster, 2020). Therefore, accurate ODADF is essential to support the decision making processes of market participants and system operators, especially in the relatively small and isolated system with increasing penetration of Renewable Energy Sources (RES) in Ireland (EirGrid Group, 2022).

Researchers have been working to improve the accuracy of both STLF and ODADF, and so far, invented many state-of-the-art methods (Ahmad *et al.*, 2022). Nevertheless, the conditions of power system operation are constantly changing. Main factors, such as growth in weather-dependant RES, new technologies used by customers, climate change, and changes in customer habits, challenge the system operation. Hence, existing models have been found to struggle achieving accurate predictions in the new reality (Foster, 2020). In recent years, approaches based on statistical, machine and deep learning algorithms have been very popular in that task (Osowski *et al.*, 2022). There is an ongoing debate on superiority of one method over another. The most promising solution, however, is to combine individual solutions into an ensemble of predictors. The following literature review aims to identify the most promising models to serve as base-learners for ODADF in Ireland, highlighting their benefits and limitations. Moreover, it seeks to determine exogenous factors affecting the demand. To achieve the aim, it was divided into five main sections as follows: single and complex models in STLF, forecasting competitions, issues with comparison of findings, and conclusion. To assess the performance of various forecasting models, Mean Average Percentage Error (MAPE) is used as the evaluation metric, as it has been widely accepted in the literature (Hyndman and Koehler, 2006).

2.1. Single Models in Short-Term Demand Forecasting

The first section of the literature review explored the main single methodologies developed and applied to STLF, including traditional, autoregressive, machine and deep learning models.

2.1.1. From Naïve, through Moving Average to Exponential Smoothing Models

The naïve forecast is the simplest method, where the future values repeat the last observation. Therefore, ignoring any patterns in time series, it produces horizontal line. To overcome pure memorisation, Moving Average (MA) takes the mean on the latest number of steps as the forecast. A Seasonal (S) naïve model looks back a number of steps for each forecast, blindly mimicking the last seasonal cycle. Exponential Smoothing (ETS) combines intuition of naïve models, adding weighted average, where the weights decrease exponentially with the distance into the history. The most known are three models: Single (SES), Double (DES) and Holt-Winters Exponential Smoothing (HWES). They have been widely used for time series forecasting since 1950s (Joseph, 2022).

For example, Taylor and McSharry (2007) used HWES to half-hourly ODADF for twenty-one european countries, based on years 1988-1993, split 5:1, achieving MAPE 2.40%. Study by Sethi and Kleissl (2020) used HWES to hourly ODADF for US, based on year 2017, split 3:1, achieving MAPE 4.60%. Another study by Shahare *et al.* (2023) used SES to hourly ODADF for a metropolitan area, based on years 2017-2021, split 3:1, achieving MAE of 1.40.

Two of above three studies used HWES and reported MAPE, while the other used SES model and reported MAE for ODADF. Additionally, all studies were carried in different regions and timeframes. Therefore, it was not possible to compare them directly. Generally, ETS were relatively simple and computationally efficient algorithms, responsive to recent changes. Despite their ability to capture trend and seasonality components, they exhibited several limitations. ETS made assumption of stationarity and was unable to capture complex non-linear patterns in the data. However, STLF exhibits non-stationary characteristics, such as seasonality and trend, which adversely impacted the performance of ETS models (Hyndman and Athanasopoulos, 2018). Furthermore, they were sensitive to initial conditions, outliers, and abrupt changes in the data (Chen and Liu, 1993), and required careful selection of smoothing parameters (Ord, Koehler and Snyder, 1997). Moreover, ETS simplicity limited their performance relative to advanced techniques (Al-Musaylh *et al.*, 2018).

2.1.2. Autoregressive Models

In addition to ETS, family of Auto-Regressive (AR) models stood the test of time and are still one of the most popular classical methods of forecasting. Their primary advantage was the ability to capture temporal dependencies within the data. While ETS was modelled around trend and seasonality, AR models rely on autocorrelation (Joseph, 2022). Three variants of AR are widely used in STLF. The first one, Auto-Regressive Integrated Moving Average (ARIMA) model, combines AR and moving average (MA) components with differencing to model non-stationary (Hyndman and Athanasopoulos, 2018). The AR and MA components capture the dependency between current and preceding observations, and the observations and a residual error, respectively (Brockwell and Davis, 2002). The integrated (I) component applies differencing to make the time series stationary (Shumway and Stoffer, 2017). Two other variants are SARIMA and SARIMAX, where ARIMA model is extended by Seasonal (S) component (Chatfield, 2015), and external factors (X) into the model (Tashman, 2000).

For example, Taylor and McSharry (2007) used AR and ARIMA to half-hourly ODADF for twenty-one european countries, based on years 1988-1993, split 5:1, achieving MAPE 3.30% and 2.80%, respectively, which were both worse than the HWES MAPE 2.40% from previous section. Study by Dudek (2015) used ARIMA to hourly ODADF for Poland, based on two years, split 12:1, achieving MAPE 2.42%. Study by Al-Musaylh et al. (2018) used ARIMA to hourly ODADF for Australia, based on years 2012-2016, split 4:1, achieving MAPE 2.16%. Another study (Sethi and Kleissl, 2020) used ARIMA and SARIMA to hourly ODADF for US, based on year 2017, split 3:1, achieving MAPE 3.80% and 3.20%, respectively, both better than HWES MAPE 4.60% from previous section.

Above studies employed AR, ARIMA and SARIMA and reported their MAPE for ODADF in various regions and timeframes. The results highlighted the context-dependent performance of those models. Moreover, although ability of SARIMAX to successfully capture temporal dependencies, handling trends and seasonality, and incorporating exogenous variables in STLF, they had their specific limitations. They assumed linear relationship between past and future observations (Zhang, 2003), and required the stationary time series. Moreover, they could not capture sudden changes in patterns (Enders, 2014). To address above limitations, researchers have proposed various extensions and modifications to AR models, as well as hybrid approach, combining them with other techniques.

2.1.3. Machine Learning Models

Machine Learning (ML) models gained significant attention due to their ability to capture complex patterns and non-linear relationships within data, making them suitable for STLF. This subsection discussed some of the most widely used ML models, including Linear Regression (LR), Multi-Linear Regression (MLR), Support Vector Regressor (SVR), Decision Trees (DT), k-Nearest Neighbors (k-NN), and Multi-Layer Perceptron (MLP). SVR can handle nonlinear relationships in data through the use of kernel functions (Vapnik, 2013). DT recursively split the input data into subsets, based on feature values, creating a tree-like structure with nodes representing decisions and branches the outcomes (Quinlan, 2014). The k-NN algorithm is a non-parametric ML method predicting new instance based on the outputs of its k-nearest neighbours in the training data, considering specific distance metric (Altman, 1992). MLP, being Artificial Neural Network (ANN) comprise interconnected neurons organised in layers. They can learn non-linear relationships in data through the process of adjusting weights and biases during training (Haykin, 2010).

For example, Hinman and Hickey (2009) used twenty-four LR models to hourly ODADF for US, based on years 2004-2009, split 5:1, achieving MAPE 3.75%, better than HWES MAPE 4.60% (Sethi and Kleissl, 2020). Study by Hong *et al.* (2010) used MLR to hourly ODADF for US, based on years 2005-2008, split 3:1, achieving MAPE 4.56%, similar to HWES MAPE 4.60% (Sethi and Kleissl, 2020). Study by Fan and Hyndman (2012) used MLR to half-hourly ODADF for Australia, based on years 2004-2009, split 5:1, achieving MAPE 1.68%, better than ARIMA MAPE 2.16% (Al-Musaylh *et al.*, 2018). Study by Ceperic, Ceperic and Baric (2013) used SVR to hourly ODADF for New England, based on years 2003-2006, split 3:1, achieving MAPE 1.31%, and for US, based on years 1988-1992, split 4:1, achieving MAPE 2.05%. Study by Dudek (2015) used MLR to hourly ODADF for Poland, based on two years, split 12:1, achieving MAPE 2.81%, worse than ARIMA MAPE 2.42%. Study by Tsekouras, Kanellos and Mastorakis (2015) used MLP to hourly ODADF for Greece, based on years in range from one to seven, split 9:1, achieving MAPE 1.78%. Researchers claimed that it was not possible to generalise their findings to other STLF case studies due to every case being distinct. That phenomenon of neural networks had already been discovered by Lu, Wu and Vemuri (1993) and named as **system dependency**. Study by Al-Musaylh et al. (2018) used Multivariate Adaptive Regression Splines (MARS), and SVR to hourly ODADF for Australia, based on years 2012-2016, split 4:1, achieving MAPE 1.27% and 1.71%, respectively, both beating ARIMA MAPE 2.16%. Foster (2020) used MLR and MLP to hourly ODADF for Northern Ireland, based on years 2013-2019, achieving MAPE 2.53% and 2.62%, respectively; and for New York, based on years 2012-2017, achieving MAPE 4.34% and 3.71%, respectively. MLP performed worse for Northern Ireland, and better for New York than MLR model. Study by Kathirgamanathan *et al.* (2022) used MLP to forecast hourly demand for New England, based on years 2004-2009, split 5:1, achieving MAPE 4.14%, worse than SVR MAPE 2.05% (Ceperic, Ceperic and Baric, 2013). Study by Osowski *et al.* (2022) used seven ML models to hourly ODADF for Poland, based on years 2014-2019, with split 7:3. The first three models, being SVR with Gaussian kernel, MLP and Radial Basis Function (RBF) network, achieved MAPE 2.27%, 2.08% and 2.26%, respectively. Those models were doubled employing autoencoder, a deep neural solution that reduces dimensionality of input data, achieving MAPE 2.13%, 1.96% and 1.81%, respectively. The seventh model was Self-Organising Kohonen Network Application (SOKNA), achieving MAPE 2.34%. RBF with autoencoder achieved the best result. All models outperformed ARIMA MAPE 2.42% (Dudek, 2015). Another study (Shahare *et al.*, 2023) used SVR and MLP to hourly ODADF in metropolitan area, based on years 2017-2021, split 3:1, achieving MAE 0.088 and 0.079, respectively, both better than SES MAE 1.40.

Above studies employed various ML algorithms for ODADF for different countries and timeframes. Despite their ability to capture complex patterns and non-linear relationships in the data and promising results of ML models in STLF, they had experienced challenges and limitations. One of the major challenges was the selection of appropriate features and hyperparameters, which significantly impacted model performance (Zhang, Eddy Patuwo and Y. Hu, 1998). Additionally, it was difficult to understand the underlying relationship between the input features and predicted values using SVR and MLP (Vellido, Martín-Guerrero and Lisboa, 2012). Finally, the computational complexity of SVR and MLP models was high, leading to longer training time and increased memory requirements (Goodfellow, Bengio and Courville, 2016).

2.1.4. Deep Learning Models

Deep Learning (DL) models have gained significant attention in recent years due to their ability to learn complex and abstract patterns from large volumes of data. This subsection discussed Long Short-Term Memory (LSTM) as implementation of Recurrent Neural Network (RNN), and Convolutional Neural Network (CNN). RNN are type of ANN designed to handle sequential data by incorporating feedback connections, allowing them to maintain hidden states and capture temporal dependencies in data (Elman, 1990). LSTM are advanced type of RNN specifically designed to address the vanishing gradient problem that occurs when training RNN on long sequences (Hochreiter and Schmidhuber, 1997). CNN are type of ANN designed to primarily handle images, utilising convolutional layers, reducing the need for manual feature extraction, followed by dense layers (LeCun, Bengio and Hinton, 2015). Nevertheless, they have been successfully implemented to STLF, ODADF in particular.

For example, Sethi and Kleissl (2020) used LSTM to hourly ODADF for US, based on year 2017, split 3:1, achieving MAPE 2.10%, better than HWES, ARIMA and SARIMA MAPE 4.60%, 3.80% and 3.20%, respectively. Study by Han *et al.* (2019) used CNN and LSTM for hourly ODADF in China, based on years 2014-2017, split 4:1, achieving 3.3% and 2.7%, respectively. Foster (2020) used LSTM and CNN to hourly ODADF for Northern Ireland, based on years 2013-2019, achieving MAPE 3.65% and 3.02%, respectively, and for New York, based on years 2012-2017, achieving MAPE 4.12% and 3.71%, respectively. For Northern Ireland, DL performed worse than MLR and better than MLP MAPE 2.53% and 4.34%, respectively. For New York LSTM performed worse, and CNN performed better than MLR and equally to MLP MAPE 4.34% and 3.71%, respectively. Study by Osowski *et al.* (2022) used LSTM to hourly ODADF for Poland, based on years 2014-2019, split 7:3, achieving MAPE 1.63%, better than all seven ML models from previous section MAPE from 1.81% to 2.34%. Study by Kathirgamanathan *et al.* (2022) used CNN and LSTM to hourly ODADF for New England, based on years 2004-2009, split 5:1, achieving MAPE 3.91% and 3.89%, respectively. Researchers found that DL models had better accuracy but higher standard deviation compared to MLP models, with LSTM being the most affected. Reduction of training data degraded the prediction and increased standard deviation for all models, with DL being more sensitive than MLP. Another study by Shahare *et al.* (2023) used LSTM and CNN to hourly ODADF for metropolitan area, based on years 2017-2021, split 3:1, achieving MAE 0.051 and 0.067, respectively, better than SES, SVR and MLP MAE 1.40, 0.088 and 0.079, respectively.

Above studies employed LSTM and CNN algorithms to ODADF for different countries and timeframes. CNN and LSTM performed equally well as MLP but with higher standard deviation, with LSTM being the most affected and sensitive to reduction in data. Therefore, despite the promising results of DL models in STLF, they had their challenges and limitations. The major challenge was the selection of appropriate architecture, hyperparameters, and training strategies, which significantly impacted their performance (Géron, 2019). Furthermore, DL required large amount of training data and high computational resources, which might be prohibitive in some applications (Zhang *et al.*, 2018). Moreover, DL are considered **black boxes**, making it difficult to interpret the learned relationship between the input features and predicted values (Chollet, 2021).

2.2. Complex Methodologies in Short-Term Demand Forecasting

All single-method algorithms have their limitations in achieving high accuracy in ODADF. Nevertheless, there are two solutions to overcome them: hybrids and ensembles, where multiple single algorithms work together to complement and augment each other. The second section of the literature review explored those complex methodologies developed and applied to STLF.

2.2.1. Hybrid Models

Hybrid models typically involve combining two or more different techniques to take advantage of the strengths of each component while mitigating their weaknesses. Some common hybrid models involved combining statistical methods with ML or DL techniques, integrating feature selection and dimensionality reduction techniques with DL, as well as combining different DL models.

For example, Song *et al.* (2006) used hybrid of Fuzzy LR and ETS to hourly ODADF for South Korea, based on year 1996, split 13:1, yielding MAPE 0.97%, 1.31% and 1.50% for spring, summer and autumn weeks, respectively. Study by Dudek (2015) used hybrid of LR and DT to hourly ODADF for Poland, based on two years, achieving MAPE 2.53%, better than MLR but worse than ARIMA MAPE 2.81% and 2.42%, respectively. In following study, Dudek (2016) used hybrid of Similar Day (SD) approach and Principal Component Regression (PCR) or Partial Least Squares Regression (PLSR) for hourly ODADF in Poland, based on four years, split 3:1, achieving MAPE 1.34% and 1.35%, respectively, outperforming all previous models built by Dudek for Poland. For Australia, hybrid models achieved MAPE 2.83% and 3.00%, respectively, which were worse than MARS and SVR MAPE 1.27% and 1.71%, respectively (Al-Musaylh *et al.*, 2018). Study by He (2017) used hybrid CNN-LSTM to hourly ODADF for China, based on years 2010-2012, split 10:1, achieving MAPE 1.41%, beating LR and SVR MAPE 2.76% and 1.72%, respectively. The result was also better than CNN and LSTM MAPE 3.3% and 2.7%, respectively (Han *et al.,* 2019). Another study by Shahare *et al.* (2023) used hybrid CNN-LSTM to hourly ODADF for metropolitan area, based on years 2017-2021, split 3:1, achieving MAE of 0.029, outperforming SES, SVR, MLP, LSTM and CNN MAE 0.051-1.400.

Above studies employed various hybrid approaches to ODADF. Hybrid models leveraged the complementary strengths of different forecasting methods, resulting in improved robustness, stability and prediction accuracy. However, they faced some challenges and limitations. One of the main challenges was selecting and combining the individual forecasting models effectively. Determining the optimal combination weights or rules for integrating the forecasts from different models required careful experimentation and validation. Tuning their parameters was more complex and both data and computationally expensive than their components. Moreover, interpretation and understanding of the hybrid model was even more challenging than individual models.

2.2.2. Ensemble Learning Models

Ensemble learning models, being a set of base-learners integrated by bagging, boosting, stacking or other technique, combine diverse single or complex models' predictions to create more accurate and robust forecast, leveraging the strengths of individual models while compensating for their weaknesses. Bagging, which stands for bootstrap aggregating, is a technique that generates multiple base models by training them on different subsets of the training data. Boosting iteratively trains base models on the same dataset but with different sample weights, encouraging the model to focus on harder examples. Stacking combines the outputs of base-learners using a meta-learner. The base-learners are trained on the original dataset, while the meta-learner is trained on the outputs of the base-learners. Furthermore, there are pre-made ensemble learning models available in standard modules, such as Random Forrest (RF) and Gradient Boosting Machines (GBM). RF is the simplest example of ensemble that constructs multiple DTs during training and combines their predictions to improve accuracy and reduce overfitting (Breiman, 2001). GBM is a sequential ensemble learning technique that combines the output of multiple weak learners in a weighted manner, to form a strong learner (Friedman, 2001), where the model grows in a stage-wise fashion, and its performance improves over iterations (Natekin and Knoll, 2013).

For example Dudek (2015) used RF to hourly ODADF for Poland, based on two years, split 12:1, achieving MAPE 1.84%. RF performed better than ARIMA, MLR and hybrid of LR-DT MAPE 2.42%, 2.81% and 2.53%, respectively. However, it performed worse than hybrid of SD approach and PCR or

PLSR MAPE 1.34% and 1.35%, respectively (Dudek, 2016). Study by Fan, Chen and Lee (2009) used ensemble of four bagged MLP to hourly ODADF for US, based on years 2005-2007, split 2:1, achieving MAPE 2.06%, while best base-learner achieved MAPE 2.71%. Study by Khwaja *et al.* (2015) used two ensembles of MLPs for New England, based on years 2004-2009, split 2:1, achieving MAPE 1.74% and 1.66% for thirty bagged and twenty-five boosted ensembles, respectively, while the best base-learner achieved MAPE 2.74%. Study by Grmanová *et al.* (2016) used ensemble of eleven stacked models for hourly ODADF in Slovakia, based on years 2013-2015, split 5:1, achieving MAPE 1.3%, while the best base-learner, HWES achieved MAPE 1.5%. Base-learners included MLP, SVR, MLP and AR models, each optimised based on its structure. Foster (2020) used RF and ensemble of 14 models to hourly ODADF for Northern Ireland, based on years 2013- 2019, achieving MAPE 2.68% and 2.63%, respectively, while the best base-learner, CNN achieved MAPE 3.02%. For New York, based on years 2012-2017, the models achieved MAPE 4.51% and 3.20%, respectively, while the best base-learner, MLP achieved MAPE 3.75%. All ensembles performed better than MLR, MLP, LSTM and CNN from previous sections. Study by Kathirgamanathan *et al.* (2022) used six ensembles of various models to hourly ODADF for New England, based on years 2004-2009, split 4:1. First three models were bagged MLP, CNN and LSTM, achieving MAPE of 4.06%, 3.52% and 3.54, respectively. Other three were boosted MLP, CNN and LSTM, achieving MAPE of 3.00%, 3.00% and 3.03%, respectively. Bagging and boosting improved the consistency and accuracy of base-learners, respectively. However, boosting improved both for LSTM. Reduction of training data degraded the prediction and increased its standard deviation for all models. Nevertheless, boosting and bagging was able to compensate for reduction of data. Another study by Osowski *et al.* (2022) used six ensembles of various base models and various combining functions to hourly ODADF for Poland, based on years 2014-2019, split 7:3. The first four ensembles consisted of MLP, RBF and SVR models integrated using Ordinary Averaging, Principal Component Analysis (PCA), Independent Component Analysis (ICA), and autoencoder, respectively, achieving MAPE 1.83%, 1.81%, 1.73% and 1.61%, respectively. The fifth ensemble was based on Wavelet Decomposition and application of SVR and achieved MAPE 1.54%. The sixth ensemble was composed of five integrated LSTM predictors and achieved MAPE 1.53%. The results revealed high advantage of using variety of individual predictors integrated into ensemble. Additionally, each ensemble performed better than any base-learner of which ensemble was composed of.

Above studies employed various ensemble architectures to ODADF. By combining the strengths of different models and reducing the impact of individual model weaknesses, ensembles could produce more accurate and robust predictions than their base-learners separately. Moreover, they were effective in reducing overfitting, handling noise and outliers, and help to generalise better to unseen data. Furthermore, they mitigated the impact of individual model errors or biases, resulting in more consistent and reliable predictions. However, they faced some challenges and limitations. Selecting the appropriate base-learners and determining the optimal combination methods was the largest challenge. It required careful experimentation, validation, and fine-tuning to achieve the best performance. Moreover, combination of multiple models reduced interpretability. Finally, they required more computational power and memory compared to individual models, which could limit their applicability in resource-constrained or real-time forecasting scenarios. Therefore, future research should focus on exploring the potential of ensemble learning (Wu *et al.*, 2021) and transformers (Vaswani *et al.*, 2017) in STLF, which have shown promising results in other domains. That effort would contribute to the development of more efficient, reliable, and sustainable power systems, better equipped to handle the challenges of the future.

2.3. Forecasting Competitions and Rankings

Public competitions, although not strictly considered a research, provide valuable insights into the benefits, challenges and limitations associated with different methods used to solve a particular problem. Participants from various backgrounds propose their solutions, revealing potential advantages and drawbacks of different techniques. While public competitions may not strictly adhere to the research methodologies commonly followed in academic studies, the findings and knowledge gained from winning solutions can be leveraged to develop new methodologies and refine existing models. Time series forecasting competitions can be traced back to 1970's and have promoted the development of forecasting research and applications (Hyndman, 2020).

In the STLF area, a series of three Global Energy Forecasting Competitions took place in years 2012, 2014, and 2017 (Hong, Pinson and Fan, 2014; Hong et al., 2016; Hong, Xie and Black, 2019). However, their results might have become outdated. Nevertheless, in 2022 (Farrokhabadi *et al.*, 2022) published paper with results of the most recent competition, organized by their authors in year 2020-21. Participants were tasked to perform ODADF in a municipal area in a post-COVID paradigm. Dataset, covering a period of four years 2017-2021, had been split arbitrarily and released daily in batches. Nearly three-hundred participants were expected to send their forecast within twenty-four hours. Finally, nine of them achieved average performance metrics above benchmark model. The first place claimed ensemble of seventy-two base models, including AR, LR, Generalised Additive Models (GAM), RF, RF for GAM residuals, MLP, and Kalman Filter Adaptation. Base model were combined using Online Robust Aggregation of Experts, described in (Obst, De Vilmarest and Goude, 2021). The second place took hybrid model, combining Similar Days (SD) approach using day type and peak temperature, with adjustment based on peak load forecast recent profiles from days of the same type. The third place secured ensemble of 674 models, including decomposed ETS, AR, GAM, and Lasso. Base models were combined by smoothed Bernstein Online Aggregation. The fourth place took ensemble of three other ensemble learning models, such as RF, GBM, and XGBoost. They were combined using weighted average based on recent performance. The fifth place took Deep Residual Networks model.

In summary, ensemble learning models took the first, third, and fourth place, hybrid model took the second place and Deep Learning model took the fifth place on the podium. Neither MAPE nor other common performance metrics were not given in the paper. Instead, a MAE Skill Score was calculated based on ratio of MAE for benchmark model and evaluated model, then decreased by one. On that scale the models from the first to fifth place got the skill score of approximately 28%, 25%, 24%, 20% and 19%, respectively.

2.4. Exogenous Input Variables

Historical power system demand is the main critical input variable of all forecasting models. There are solutions utilising solely that variable, called univariate AR models, which assume that other factors impacting demand change gradually, and become feature which is captured in the data series itself (Taylor and McSharry, 2007). However, the vast majority of researchers found that inclusion of calendar and weather factors improves the performance of forecasting models.

2.4.1. Calendar Variables

The majority of researchers extracted the calendar variable and used it in pair with historical demand as input vector to their models. To represent the annual pattern, a common choice was to use cyclical encoding, such as sine and cosine waves to optimally represent the seasonality of the time series. The same approach was used to represent daily pattern but not the weekly pattern, as the variation in load profile was not sufficiently cyclic over a week (Foster, 2020). For example, that approach was utilised in MLP in (Bakirtzls *et al.*, 1996; Tsekouras, Kanellos and Mastorakis, 2015). Additionally, time of the day was also represented using a different number for each hour as in MLP in (Park *et al.*, 1991). To represent days of the week, a common choice was to use One-Hot Encoding. For example, that approach was employed in MLP in (Tsekouras, Kanellos and Mastorakis, 2015), and in GRU in (Yu *et al.*, 2019). Moreover, holidays were commonly marked to differentiate them from working days by binary variable, as in (Jiao *et al.*, 2018; Yu *et al.*, 2019).

2.4.2. Weather Variables

Weather variables are the meteorological measurements that describe various aspects of atmospheric conditions, such as temperature, relative humidity, wind speed, precipitation, atmospheric pressure, solar radiation and cloud cover. Their impact on STLF has been well researched but their selection varied between researchers. Some used only temperature, while others use more or even all variables with dimensionality reduction techniques. Additionally, Fidalgo and Matos (2007) claimed that weather variables should reflect the characteristics of the specific country. Therefore, each forecasting model must be finely tuned to the country in which it is being used.

Temperature has been the most apparent variable which impacts power demand. Hong and Shahidehpour (2015) claimed that temperature alone explains at least 70% of the variation in demand. However, customer do not react immediately to a change in temperature (Bolzern, Fronza and Brusasca, 1982). Additionally, use of air-conditioning affects the demand in summer (Hong and Shahidehpour, 2015). Therefore, researchers using linear models wanted to capture the non-linearity. For example, Hinman and Hickey (2009) used squared value of daily temperature average in MLR model. Study by Wang, Liu and Hong (2016) addressed the phenomenon of the demand dependency on the recent temperature pattern, and described it as **recency effect**. Using MLR model, Hinman and Hickey found that the optimal performance when lagged temperature from the previous twelve hours and two daily moving average temperatures for the past two days were included. Another study by Pardo, Meneu and Valor (2002) examined introduction of heating (HCD) and cooling degree days (CCD) as input to the model to represent very high or very low temperature, above or below which the building did not require heating or cooling, respectively (Day, 2006). Many sources used HCD of 15.5^{o}C and did not use CCD for UK (Day, 2006; Beggs, 2012). However, Foster (2020) found that complex temperature variables were not superior to simple lagged ones for Northern Ireland.

Relative humidity and wind speed have been the subsequent two most influential variables identified in research. They had impact on how people feel the temperature (Steadman, 1984). Xie *et al.* (2018) expanded the algorithm build by Wang, Liu and Hong (2016) by relative humidity variable and further increased the model performance. Wind speed had double impact on load variation. Higher wind speeds affected the chill index, increasing demand (Taylor and Buizza, 2003) and at the same time impacted RES generation, decreasing the demand from system sources (Foster, 2020).

2.5. Issues with Comparison of Research Results

The number of papers on STLF increases every year and hence the number of proposed techniques. Unfortunately, most of the approaches could not be replicated or validated and many issues have been observed and reported in literature. The first one was related to dataset on which models are evaluated. Many studies were based on unique and non-publicly available data (Hong *et al.*, 2020). Furthermore, more than half of publications were evaluated on single dataset and most of them demonstrate superiority of their proposed approach over others based on above (Scheidt *et al.*, 2020). Moreover, the length of analysed period significantly affected the quality of forecasting (Czapaj, Kamiński and Sołtysik, 2022). The length of used time series and their graduation differed substantially between studies and ranged from one month to several years and from one minute to one hour, respectively (Scheidt *et al.*, 2020). The second problem was related to lack of consensus regarding evaluation metrics (Ahmad *et al.*, 2022). They were often inadequate or arbitrarily selected (Hong *et al.*, 2020). RMSE and MAE were easy to interpret but made comparison between datasets complicated (Scheidt *et al.*, 2020). Additionally, Czapaj, Kamiński and Sołtysik (2022) claimed that the forecasting effectiveness described solely by the lowest MAPE was ambiguous. The third problem was related to risk of bias. Some well-known datasets might give researchers unfair advantage to their proposed method (Hong *et al.*, 2020). Additionally, truncation of the dataset might be used to increase accuracy (Czapaj, Kamiński and Sołtysik, 2022). The last problem was avoidance of direct comparison with established and state-of-the-art models. Therefore, while there has been no universal one-fits-all solution, comparison of research findings was even more complicated due to many factors.

2.6. Conclusion

This literature review provided a comprehensive overview of various approaches to STLF abroad, focusing on single and complex methodologies, exogenous variables influencing demand and issues with comparison of research results. It served as a robust foundation for the research objectives, which include the development and evaluation of ensemble learning models, as well as revealing their potential in ODADF in Ireland. While ensemble learning has not yet been studied in the context of Ireland's power system demand forecasting, this approach holds significant promise and could address the current knowledge gap in this area.

The diverse range of reviewed sources, employing variety of methodologies from traditional ones to current advancements, offered a rich, multifaceted perspectives on the field, which provided vital information that directly informed the first research objective. The strengths and weaknesses of various methodologies, under specific conditions reported in research, was invaluable in guiding the selection of the base-learners and integration methods to develop the ensemble learning framework for this research. The significance of incorporating historical power demand data, calendar variables, and weather variables in forecasting models and various strategies employed by researchers in the field for incorporating these variables, informed the implementation of the framework.

As for the second research objective, the literature review underscored the importance of evaluating the performance of ensemble learning models by comparing their metrics based on a consistent and effective approach. Insights gleaned from real-world applications informed the experimental design for this research.

As for the third research objective, the literature review shed light on the challenges and limitations in the field, such as issues with lack of comparisons with established models. These insights were instrumental in the exploration of ensemble learning models' potential to ODADF in Ireland.

Finally, this literature review had not only deepened the understanding of the field, but also significantly shaped the direction of the research objectives. The rich information and critical evaluations gleaned from the sources illuminated the methodologies, variables, and practical considerations that required focus. The review thus provided an invaluable basis for the design and development of an optimal ensemble learning configurations to ODADF and significantly informed future research and development recommendations.

3. Research Methodology

3.1. Concise Methodology Framework

The primary research methodology of this research was experimentation. The aim was to determine the cause-and-effect relationship between ensemble learning configurations and power demand forecasting performance metrics, in order to find the best architectures for One Day-Ahead Demand Forecasting (ODADF) in Ireland. The methodology framework was divided into three major phases, such as data preparation, experimentation with ensemble learning architectures, including base-learners and integration methods, and validation of results (Figure 3.1). Figure 3.2 depicted structured overview of the research methodology as high-level pseudocode.

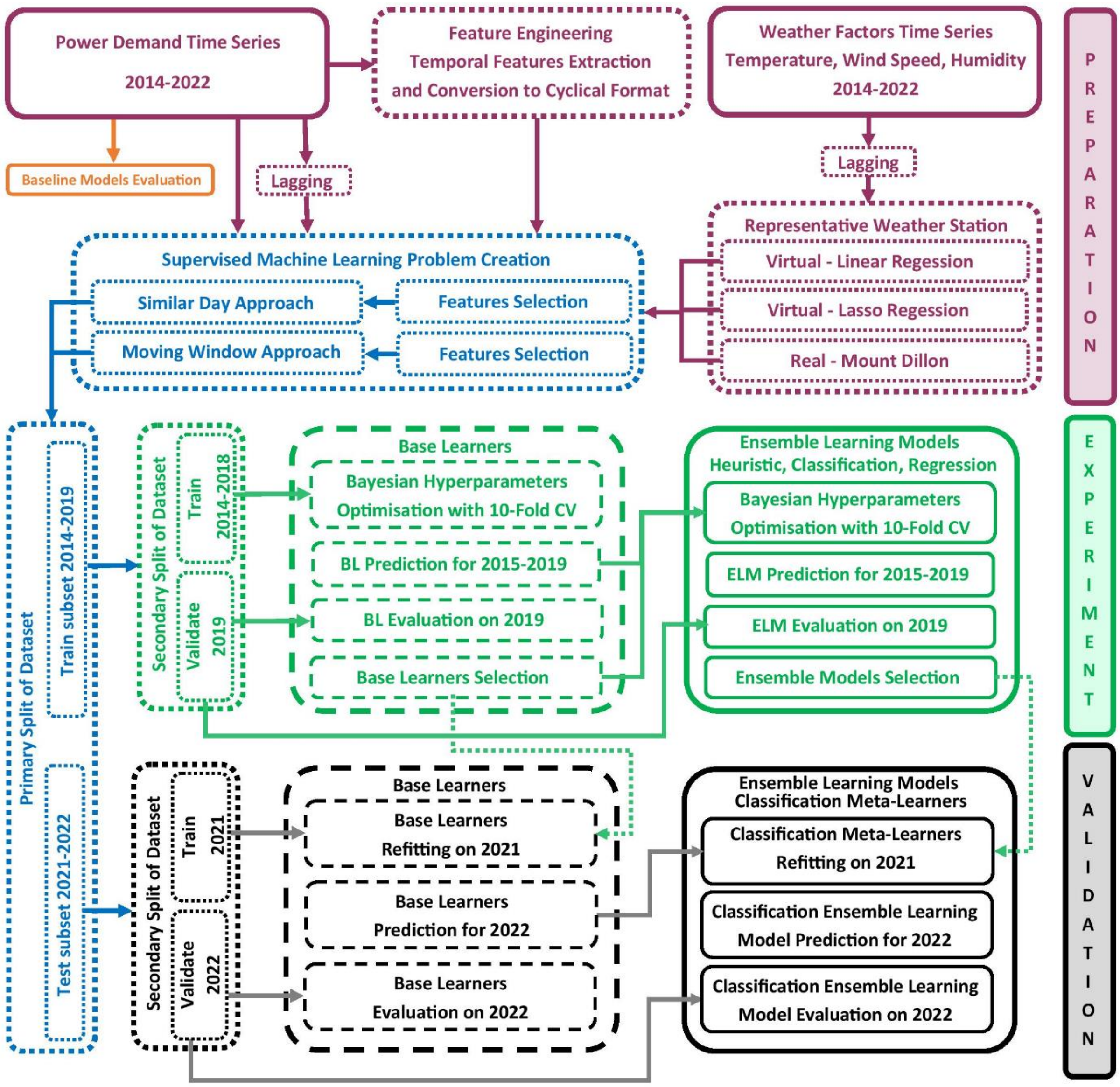


Figure 3.1. Schematic of the research methodology framework.

```
// Preparation Phase
    LOAD dataset of power demand time series for years 2014:2022
    SELECT 22 automatic weather stations
    FOR weather station IN automatic weather stations:
        LOAD weather dataset for weather station time series
        SELECT [name] FROM weather dataset AS station name
        SELECT [temperature, humidity, wind speed] FROM weather dataset WHERE 2014 ≤ year ≤ 2022 AS selection
        APPEND [station name, selection] TO weather features
    SAVE dataset of weather features to CSV file
// Cleaning and Processing the Data Phase
    CLEAN power demand time series
    RESAMPLE power demand time series
    SCALE power demand time series
    EXTRACT temporal features from power demand time series timestamps
    CONVERT temporal features of power demand to cyclical and vector formats
    LOAD dataset of weather features
    FOR weather factor IN [temperature, relative humidity, wind speed]:
        FOR approach IN [linear regression, lasso regression]:
            CREATE representative virtual weather station
        SELECT representative real weather station
    FOR approach IN [similar day, moving window]:
        DEFINE supervised learning problem
        IF approach IS similar day:
            SELECT temporal features, weather station, weather factors and lagged power demand
        ELSE: // moving window
            SELECT temporal features, weather station, weather factors, and window size
        CREATE supervised learning problem dataset
    SPLIT supervised learning problem datasets into training datasets 2014:2019 and testing datasets 2021:2022
    SELECT power demand FROM training datasets WHERE 2014 ≤ year ≤ 2019
    SELECT baseline models
    FOR baseline model IN baseline models:
        MAKE baseline models predictions for year 2019
        EVALUATE baseline models ON year 2019
// Experimentation Phase - Base-learners of Ensembles
    SPLIT training datasets into training subsets 2014:2018 and validation subsets 2019
    SELECT potential base-learners along with associated hyperparameters and their ranges
    FOR base-leaner IN potential base-learners:
        TUNE hyperparameters of potential base-learner 2014:2018 using Bayesian optimization with 10-Fold CV
        TRAIN potential base-learners on training subsets
        MAKE base-learners predictions for years 2015:2019
        EVALUATE potential base-learner ON validation subsets
    SELECT 20 top-performing potential base-learners as base-learners
// Experimentation Phase - Meta-Learners of Ensembles
    SELECT potential meta-learners of ensembles along with associated hyperparameters and their ranges
    FOR meta-learner IN [heuristic rules, classifier and regressor]:
        IF meta-learner <> heuristic rules:
            TUNE hyperparameters of meta-learners 2015:2018 using Bayesian optimization with 10-Fold CV
            TRAIN potential meta-learners using 20 base-learners predictions for years 2015:2018
        MAKE ensemble predictions for years 2015:2019
        EVALUATE potential meta-learners ON validation subsets
    SELECT classification meta-learner of ensembles for validation phase
// Validation Phase
    SPLIT testing datasets into training subset 2021 and validation subset 2022
    FOR base-learner IN base-learners:
        REFIT base-learners on training subset 2021
        MAKE base-learners predictions for years 2021:2022
        EVALUATE base-learners on unseen validation subset 2022
    REFIT classification meta-learner of ensembles using 20 base-learners predictions for year 2021
    MAKE classification-based ensembles predictions for year 2022
    EVALUATE classification-based ensembles ON validation subset 2022
```

Figure 3.2. High-level pseudocode as structured overview of research methodology.

The first part of the preparation phase identified two populations, in the context of utilising secondary datasets into primary research. The first one was the whole dataset containing total power demand measurements from Ireland, available from EirGrid (2023). The second one consisted of weather factors measurements from weather stations in Ireland, available from MetÉireann (2023). Non-probability sampling and judgment sampling were used as sampling method and type, respectively. Quantitative research approach was used, which involved pre-processing, integration of data, and conducting Exploratory Data Analysis (EDA). Time series of power demand, collected from national system operator, was cleaned, resampled, scaled, and enriched by extraction of temporal features, and subsequently converted to cyclical and vector formats. Time series of weather factors, collected from twenty-two weather stations of the national meteorological service, were trimmed to cover temperature, relative humidity, and wind speed for years 2014-2022, and then cleaned. Unlike demand, where the data was given for the entire country, weather data was distributed locally. Given that inclusion of twenty-two weather stations into research was not feasible, three approaches to find representative stations were proposed, such as two virtual weather stations created by Linear and Lasso Regressions, and the most important real one.

In the second part of preparation phase, baseline models were examined, and supervised learning problems were defined, for Similar Day (SD) and Moving Window (MW) approaches. As one-day ahead forecasting is in practice performed at least several hours in advance, current-day partially available data had to be removed from consideration, to avoid **double forecasting phenomenon**. Investigation into feature importance was performed to select the best ones, considering all three representative weather stations, utilising methods such as Forward and Backward Selection (Elimination) Regression, Lasso Elimination Regression, Random Forest and Permutation Feature Importance, separately for SD and MW approaches. Subsequently, supervised learning problem datasets were created for both approaches. Finally, datasets were split into training and testing datasets, covering years 2014-2019, and 2021-2022, respectively.

The first part of the experimentation phase identified the population of interest, in the context of the primary research, as the entire set of possible ensembles configurations, including all combinations of selected base-learners and techniques of integration, with their respective hyperparameters. Non-probability sampling and judgment sampling were used as sampling method and type, respectively. The base-learners belonged to supervised machine and deep learning algorithms designed to solve regression problems, possibly able to incorporate exogenous factors, such as calendar and weather features. Ensemble learning models were compilation of base-learners with their hyperparameters integrated by three stacking techniques, such as heuristic rule, classification, and regression-based methods.

In the second part of experimentation, training datasets were further split into training and validation subsets, covering years 2014-2018, and 2019, respectively. Then, Bayesian optimisation was used for hyperparameters tuning of potential base-learners, using 10-fold cross-validation, and predictions were made for years 2015-2019. Potential base-learners were evaluated on year 2019, and twenty of the most promising ones were selected as base-learners. Finally, predictions of base-learners became the training dataset for meta-learners of the ensembles. Heuristic rule-based ensembles were used as the baseline models. Bayesian optimisation was used for hyperparameters tuning of classification-based and regression-based meta-learners, using 10-fold cross-validation, and predictions were made for years 2015-2019. Potential meta-learners were evaluated on data from year 2019, and classification-based ensembles were selected for further investigation.

In validation phase, testing datasets were further split into training and validation subsets, covering years 2021 and 2022, respectively. Then, the twenty base-learners, with hyperparameters inferred from the second stage, were refitted and evaluated on unseen data from 2021 and 2022, respectively. Subsequently, predictions of base-learners became the training dataset for meta-learners of the ensembles. Finally, the potential of classification-based ensembles was validated on data from year 2022.

3.2. Project Management Framework

CRISP-DM, a proven and widely acceptable methodology in data science projects (Vorhies, 2016), ensures that business objectives remain at the centre of the project (Saltz, Shamshurin and Crowston, 2017). It is free, neutral in regards of application, industry and tools used, and approaches the life cycle of the project from both, an application-focused and technical perspectives (Negro, 2021). Therefore, CRISP-DM, adapted to the requirement of this research, was selected as project management framework for this research (Figure 3.3).

3.3. Business Understanding

The background, research objectives, and literature review covered in Chapters 1 and 2, respectively, formed the business understanding phase. Accurate ODADF was found crucial for electrical network reliability, the environment, and trading markets. Problem with achieving accurate predictions by individual and hybrid models in the dynamic conditions of the power system operation, had been identified. Finally, ensemble learning approach was identified as possible solution to overcome the limitations of those methods.

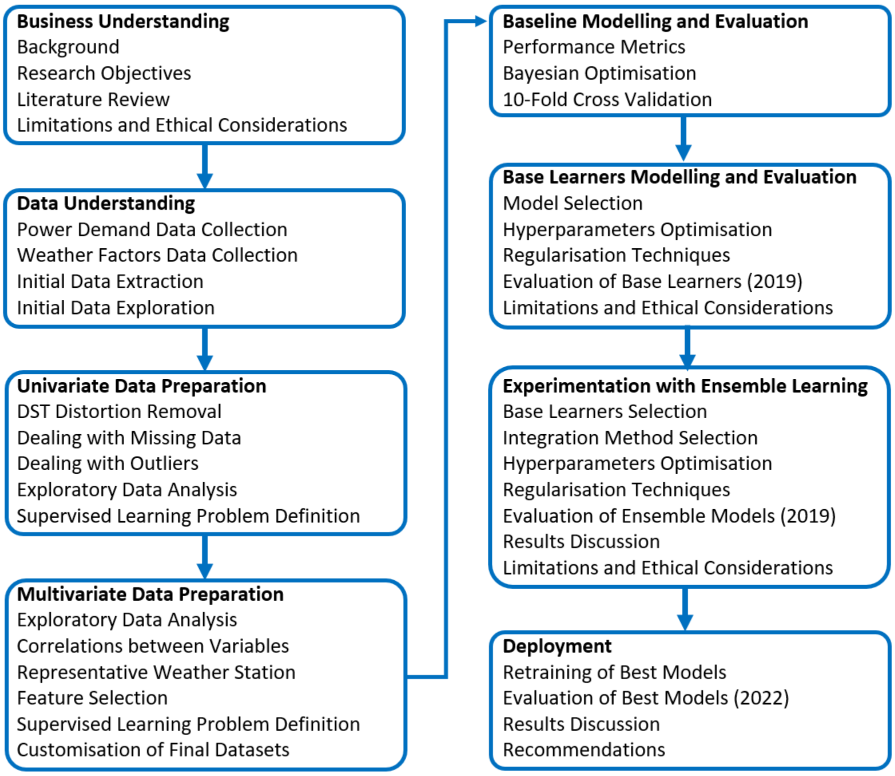


Figure 3.3. CRISP-DM Project Management Framework. Adapted from Chapman et al. (2000)

3.4. Data Understanding

Time series of power demand in Ireland was collected from the national system operator EirGrid (2023), through their Application Programming Interface (API). It covered fifteen-minutes average power demand in Ireland, for the available period from 18/07/2013 to 10/06/2023. Full years 2014-2022 were used in order to allow proper pattern recognition. Time series of Irish weather factors were collected from repository of national meteorological service MetEireann (2023). Unlike demand, where the data was given for the entire country, weather data was distributed locally. From all observing weather stations in Ireland, twenty-two automatic ones, recording data for temperature, wind speed and relative humidity every minute, subsequently hourly aggregated and validated by meteorologists, were selected. Given that the original dataset covered more variables for much longer period than needed, datasets were trimmed to cover years 2014-2022 and three selected weather factors only, to accelerate their further processing. Additionally, coordinates of weather stations were extracted from datasets, and presented on the map of Ireland, using **Folium** module and information contained in **JSON** dataset, collected from Free Software Foundation (FSF, 2023). Initial EDA was performed to detect distortions, missing timestamps, missing data and outliers in time series.

3.5. Univariate Data Preparation

The univariate data was enriched by extraction of day of the week, holidays observed on weekdays, day of the year and hour from timestamps. Detected distortions, missing data and outliers were dealt with, using data analytics methods. Firstly, time series affected by Daylight-Saving Time (DST) distortion were shifted back by one hour to revert their continuity, and the one-hour gaps were filled using second-degree Polynomial Regression interpolation. The new data was validated using Coefficient of Determination R^2 and RMSE metrics, as well as visual comparison to SDs from up to five neighbouring weeks before the affected day.

Single and continuous missing values were filled using Linear Regression (LR) interpolation of the neighbouring timestamps, and SDs from neighbouring weeks, respectively. Flat parts of time series, with the same value repeating more than six times (1.5 hour) and spikes, with rapid value change by more than 300MW, were treated as outliers, and replaced using LR interpolation of SDs from neighbouring weeks. Validation of the cleaning process was performed by comparing the distributions of power demand before and after the process.

EDA was performed on cleaned dataset, including descriptive statistics, and visualisation of raw and normalised time series, to find patterns in time series, as well as similarities and differences between days of the week. Time series was then resampled to one-hour interval by averaging the values, for consistency with current research. Inferential statistic was performed on cleaned dataset, including Shapiro-Wilk normality tests, Mann-Kendall trend tests, Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) stationarity tests, auto-correlation (ACF) and partial auto-correlation (PACF) tests, on original and first-order differenced time series. Additionally, Multiple Seasonal Patterns Decomposition (MSTL), as proposed by Bandura, Hyndman and Bergmeir (2021), was performed on of hourly time series to inform further research. Subsequently, supervised learning problems were defined for SD and MW approaches, utilising historical weekly lags and continuous historical demand values, respectively.

For SD approach, an investigation into weekly lagged power demand was performed to find how an individual or range of individual values could be utilised for ODADF. Then, LR was considered for adjusted predictions for one week, one year (52 weeks), and a range 1-52 weeks ago. Similarly, investigation into daily lagged demand was performed. Subsequently, research into MW size was conducted for MW approach.

Power demand values were scaled by dividing by the global maximum value, effectively placing them in a range between 0 and 1. This approach, while similar to the **MinMax** scaler, avoided values near or equal to zero. Such scaling was found beneficial for modelling and evaluating models, particularly for the MAPE metric calculation. The day of the week, represented as categorical values from Monday to Sunday, was converted into a numerical vector using a One-Hot Encoder. In this encoding, a particular column was set to 1, while the others were filled with zeros. Observed holidays, being categorical values Yes or No, were converted to numerical values 1 or 0, respectively. Day of the year and hour of the day, being numerical values 1-365 (366 in leap year) and 0-23, respectively, were converted to a cyclical format using sine and cosine functions. The conversion approach was then validated by correlation study.

3.6. Multivariate Data Preparation

Time series of weather factors, affected by DST, were shifted back by one hour to restore their original continuity, and the one-hour gaps were filled using their monthly mean values. Wind speed, given in knots, nautical-miles-per-hour, was converted to kilometres-per-hour, by multiplication by 1.852. Then, correlation study between actual power demand and lagged weather factors was performed. For each weather factor, three approaches for representative weather station were proposed, such as two virtual weather stations created by Linear and Lasso Regression, and the most important real weather station, revealed by Lasso Regression. Results were validated by correlation study between actual demand and lagged weather factors for each solution, and by comparing individual distributions and correlation plots. Finally, weather factors were scaled using the **MinMax** scaler, effectively converting their values to a range from 0 to 1.

The final supervised learning problem datasets were created for SD and MW approaches. For the SD approach, weekly lags with a correlation above 90% were considered. An investigation into feature importance was carried out to select the best features. Several methods were utilized:

- Forward Selection Regression: the variable that, upon addition, provided the most significant improvement to the fit was permanently added to the selection,
- Backward Selection (Elimination) Regression: the variable whose removal resulted in the least increase in error was permanently eliminated from the selection,
- Lasso Features Selection (Elimination) Regression: variables were removed from the selection by increasing the value of hyperparameter alpha,
- Random Forest Feature Importance. This method leveraged the inherent structure of random forests by assessing how frequently a feature was used to split the data and its impact on reducing impurity. Important features were those that frequently improved the purity of the node.
- Permutation Feature Importance. This method evaluated feature importance by measuring the decrease in model accuracy when the feature's values were randomly shuffled. A significant drop in performance indicated that the feature was important.

For MW approach, an investigation into feature importance was performed to select the best ones for the univariate size of the MW. Then, investigation into MW size and selected features was performed recurrently to select features and size which achieved the lowest MAPE.

The whole dataset covered period from year 2014 to 2022. However, it was observed that from 27/03/2020 to 26/07/2020, during various phases of Covid-19 lockdown in Ireland, the pattern of power demand was seriously disrupted by regulations at national level. As the research did not concentrate on examining that period specifically, but rather focused on forecasting of power demand under normal conditions, an approach was proposed to train all models on data in years 2014-2018 and validate the results on year 2019. Then, all models with hyperparameters tuned on data in years 2014-2018 were refitted and tested on years 2021 and 2022, respectively. That way, data from year 2020 was not considered in models training and evaluation, and the disrupted period did not influence the results of this research. That approach aligned with the optimal splitting considerations by Roshan (2022). Finally, given that different models had unique data shape requirements, the final datasets were tailored accordingly.

Several performance metrics were reported in this research, including:

- Mean Square Error (MSE): calculated as the quadratic mean of the differences between predicted and observed values. It served as the function that optimizers aimed to minimize,
- Root Mean Square Error (RMSE): derived from the square root of the MSE,
- Mean Average Error (MAE): determined by dividing the sum of absolute errors by the sample size,
- Mean Average Percentage Error (MAPE): the mean difference between predicted and observed values relative to the observed values. Furthermore, MAPE was broken down by day of the week (including observed holidays), month, and hour to facilitate a detailed evaluation.

While Grid-search and Random-search are well-known techniques for hyperparameters tuning, they face significant limitations. Grid-search is computationally intensive due to exhaustive combinations and is highly sensitive to the range and count of pre-selected hyperparameters. On the other hand, Random-search, while efficient, might overlook optimal values due to its random selection process. In contrast, Bayesian optimisation offered a more efficient and data-driven approach. It worked efficiently with both discrete and continuous variables, beginning with random hyperparameters sampling and then finding the best hyperparameters using a Gaussian process, iterating until the maximum number of trials was achieved. Significantly, Bayesian optimisation seamlessly integrated model's internal hyperparameters with external ones, such as name of weather station, for example. Therefore, for this project, Bayesian optimisation was selected for hyperparameters tuning, using **hyperopt** module. To mitigate the risk of overfitting and to ensure a reliable hyperparameter performance estimate, 10-fold cross-validation was employed. This approach aided in evaluating both the model's bias and variance, providing insights into its generalization capabilities on unseen data (Burkov, 2019).

Finally, the simplest baseline models were considered, such as naïve last year value, historical years mean, and last year mean, which repeated a single value for the whole next year. Then, similar day - one week, and one year (52 weeks) ago - were examined. Finally, the state-of-the-art univariate time series models, such as TBATS (De Livera, Hyndman and Snyder, 2011), NBEATS (Oreshkin et al, 2020), and NHITS (Challu et al, 2023) were taken into consideration. However, the last one was not employed in this research due to lack of available documentation.

3.7. Modelling and Evaluation of Base-Learners

According to research in ensemble learning models (Sollich and Krogh, 1996; Kuncheva and Whitaker, 2003), they tend to yield better results when there is a significant diversity among the base-learners. Therefore, successful implementations promote diversity among the models they combine (Adeva, Beresi and Calvo, 2005; Brown et al, 2005). Based on that, and the information gathered through literature review, ML models, such as LR, Ridge, Lasso, SVM, GBM and MLP Regressors were considered as potential base-learners, using SD and MW approaches. Additionally, DL models, such as CNN and LSTM were examined as base-learners using SD approach. However, it was not feasible to include them, based on MW approach, within the timeframe of the research. Hyperparameters tuning was performed, for all base-learners and their selected hyperparameters, using data from years 2014-2018 and validated, using 10-fold cross-validation, on data from year 2019. Kolmogorov-Smirnov tests were performed to compare predictions' and actual demand's distributions.

3.8. Experimentation with Ensemble Learning Models

Various architectures of ensembles, that incorporated twenty base-learners, were examined. Integration by stacking was represented by heuristic rule-based ensembles, as well as by classification, and regression-based meta-learners of ensembles. Their performance as a function of the day of the week, month and hour were examined, followed by Kolmogorov-Smirnov tests and the percentage share of base-learners in the best predictions by year and the day type.

Firstly, heuristic rule-based ensembles were considered, including mean, median, and base-learners with lowest MAPE one day, one week, and one year (52 weeks) ago. Secondly, classification-based ensembles - on explanatory variables only - were examined. Classifiers, such as Logistic Regression, SVM, GBM and MLP, were assessed as meta-learners. The product of probabilities, of base-learners being the best predictors for particular hour, day type, day of the year, weather factors, and their predictions, were taken as ensemble predictions. Finally, regression-based ensembles, based on limited and unlimited number of base-learners' predictions, were considered. Regressors such as LR, SVM, GBM and MLP were examined as regression meta-learners, their predictions were evaluated on data from year 2019, and tested with Kolmogorov-Smirnov tests. Then, ensemble learning models' performance as a function of the day of the week, month, and hour, were performed.

3.9. Deployment

Classification-based ensembles, as the top performers, were refitted and tested on unseen data. Firstly, the twenty base-learners, with hyperparameters inferred from previous phase of experimentation, were refitted and evaluated on unseen data from years 2021-2022, respectively. Subsequently, classification-based meta-learners, with hyperparameters inferred from previous phase of experimentation, were refitted and evaluated on the base-learners' predictions, and data from year 2022, respectively. Then, base-learners' and classification-based ensembles' performance as a function of the day of the week, month and hour were performed, followed by the percentage share of base-learners in best predictions by year and day type study. Finally, the results and limitations of the research were discussed, and recommendations for further investigation were provided.

3.10 Tools and Technologies

All computations in this project were performed on Lenovo Legion 5 Pro with AMD-Ryzen7-5800H processor, 64GB RAM, running Windows 10. Python and its libraries implemented in Jupyter Notebook were used for this project. Version control system Git helped to track and manage changes to files on local computer, while repository hosting GitHub helped to secure them in the cloud.

3.11. Limitations and Ethical Considerations

This section addressed the ethical considerations inherent in the Data Analytics Project as foundation for responsible research conduct, ensuring the protection and well-being of individuals and organisations, both involved in the research and affected by its results, following principles for dissertations (Resnik, 2005, 2018; Shamoo and Resnik, 2009; Bryman and Bell, 2015; Saunders, Lewis and Thornhill, 2015; Adams and McGuire, 2022).

The whole research was performed on personal computer with original operating system, word processing software, open-source distribution of Python programming language and its relevant scientific packages and libraries. Secondary datasets for power demand and weather factors are publicly available, with Open Data and Creative Commons Licenses, respectively, subject their acknowledgment, while retaining ownership of the data.

In regards to the primary research, bias was inherent in non-probability, judgment sampling. Moreover, due to time constraints of the research, a balance between performance and computational complexity of the configuration was considered. To reduce bias and ensure that chosen sample was truly representative for the entire population, a wide range of models, with different characteristics was selected. To ensure identical experimental environment, each ensemble was trained and validated on the same training datasets. The results from sample of units inferred to the performance of all ensembles in population. Therefore, the best performing model from the sample was as close as possible to, and most likely, the best performing model in the whole population of interest.

To build an ensemble, suitable individual predictors and ensemble techniques were pre-selected. As the project was conducted within twelve-weeks period, a limitation in architecture of the ensemble was recognised and accepted, with aim to develop a balanced ensemble model, where potential benefits in accuracy were weighed against computational effort and complex model building. Therefore, DL models were not incorporated for MW approach, and Bayesian optimisation of hyperparameters tuning was restricted to twenty trials. Besides, hybrid models were not considered for base-learners. Moreover, early-stopping was not integrated into the MLP, LSTM, and CNN models because it was found to be incompatible with Bayesian optimisation, causing early stopped trials to fail. Instead, the number of epochs was set as a hyperparameter to be optimised. Furthermore, while three classes of stacking integration methods were examined in this research, neither bagging or boosting ensembles were explored due to time-constraints of this research.

4. Data Understanding

4.1. Data Collection

Time series of power demand in Ireland was collected from the national system operator EirGrid (2023), Smart Grid Dashboard, through their Application Programming Interface (API) (Figure 4.1). Time series of Irish weather factors were collected, as **CSV** files, from website repository of national meteorological service MetEireann (2023). Unlike demand, where the data was given for the entire country, weather data was distributed locally. From all observing weather stations in Ireland, twenty-two automatic ones, recording temperature, wind speed and relative humidity every minute, then hourly aggregated and validated by meteorologists, were selected.

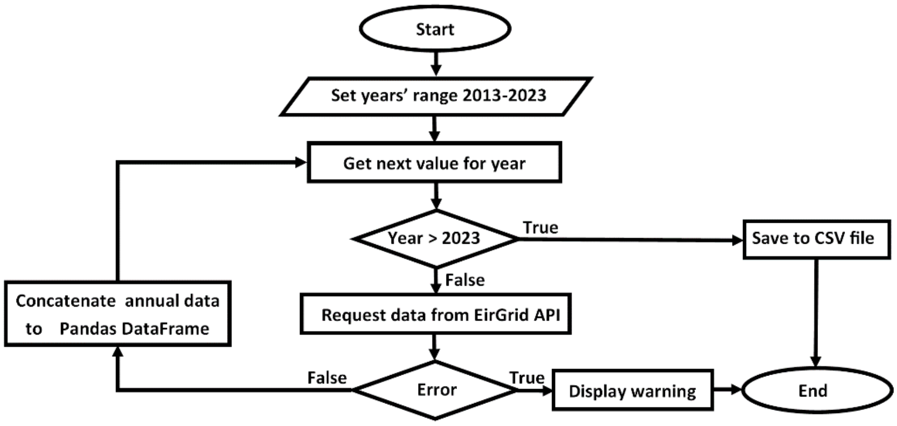


Figure 4.1. Schematic of data collection for univariate time series of power demand in Ireland.

4.2. Initial Data Extraction and Exploration

4.2.1. Time Series of Power Demand in Ireland

Time series of fifteen-minutes average power demand for the whole Republic of Ireland covered the period from 18/07/2013 to 10/06/2023. Full years from 2014 to 2022 were used in order to allow proper pattern recognition. The data was enriched by extraction of temporal features, such as day of the week, holiday observed on weekdays, day of the year and hour, from timestamps, using **calendar** and **holidays** modules (Figure 4.2).

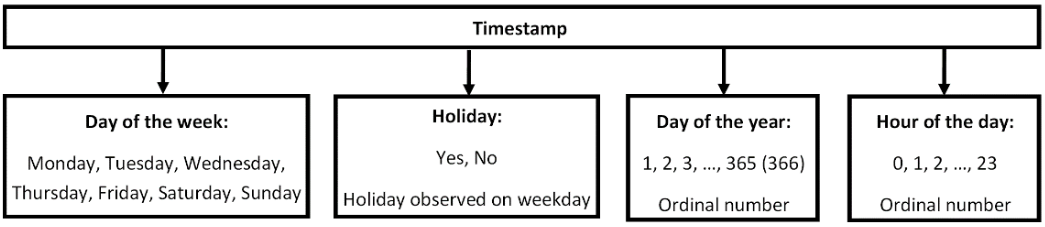


Figure 4.2. Schematic of data collection for univariate time series.

4.2.2. Time Series of Weather Factors in Ireland

Time series of weather factors in Ireland covered one-hour average of various factors for many decades. Temperature, relative humidity and wind speed were found the most important in recent research. Therefore, to accelerate further data reading and processing, time series were trimmed to cover years from 2014 to 2022 and above weather factors only (Figure 4.3). Additionally, names and coordinates of weather stations, were extracted from those datasets, and presented on the map of Ireland (Figure 4.4), using **Folium** module and geographical information contained in **JSON** dataset, collected from Free Software Foundation (FSF, 2023).

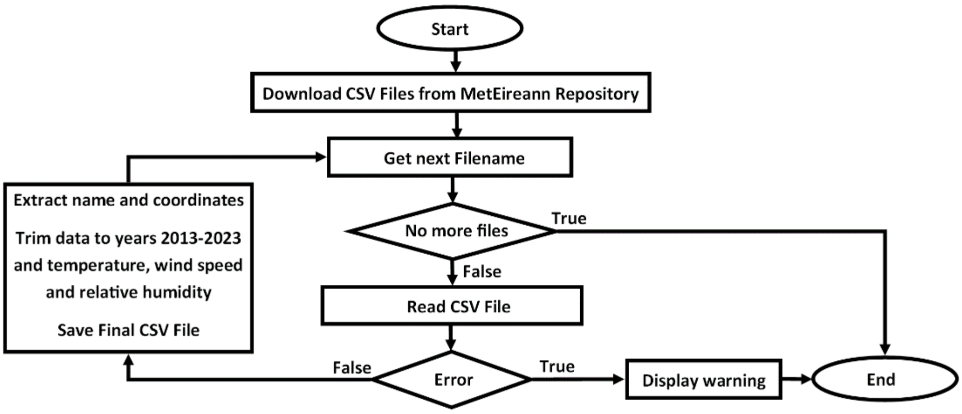


Figure 4.3. Schematic of data collection for weather factors time series in Ireland.

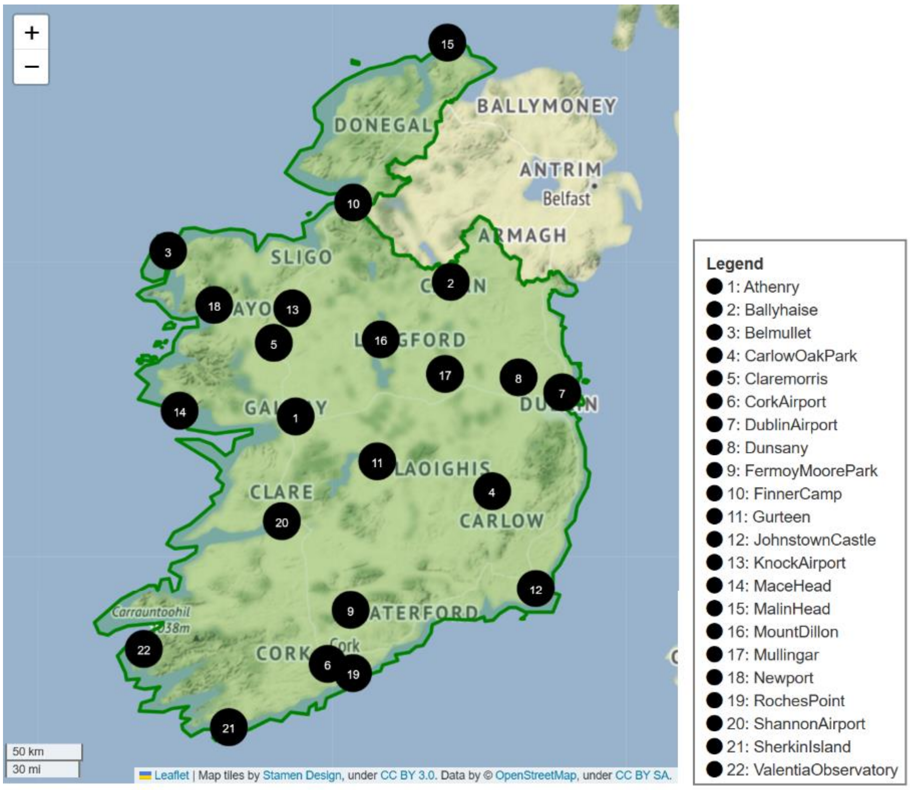


Figure 4.4. Location of 22 automatic observing weather stations in Ireland. Generated with Folium module and geographical information contained in JSON dataset collected from Free Software Foundation (FSF, 2023).

4.2.3. Distortions, Missing Values and Outliers Detection in Power Demand Time Series

While time series had no missing timestamps, ten days with missing values of 147 among 346,852 all values (less than 0.04%), were detected. Additionally, Daylight Saving Time (DST) introduced artificial distortion by setting the clock by one hour forward and back in the spring and autumn, respectively. That shifted time series by one-hour between time-change days, and created one-hour gap and overlap every year, on the time-change day, in autumn and spring, respectively. That changed the pattern in time series (Figure 4.5), and needs to be addressed. Dates of time-changes were established using **pytz** module.

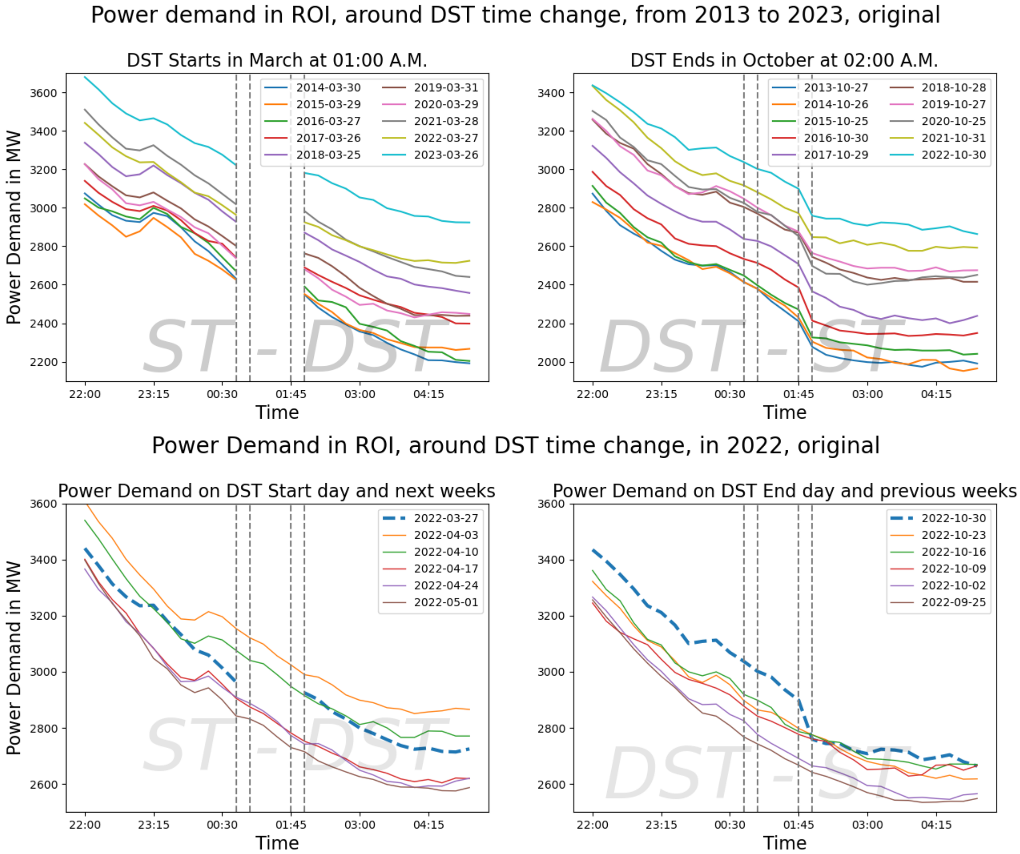


Figure 4.5. Daylight Saving Time (DST) distortion creates one-hour gap and one-hour overlap every year, on the time-change day in autumn and spring, respectively (see years 2013-2023 on the top row) what changes pattern in the time series (see year 2022, similar days up to 5 weeks after/before the time-change day on the bottom row).

Within the ten days with missing values, five days missed single value and five days missed continuous values. Furthermore, twenty-one days with outliers were detected. Flat parts of time series, where the same value repeating more than six times (1.5 hour) occurred in twelve days (Figure 4.6). Spikes in time series, where the value changed rapidly, by more than 300MW in a single step, occurred in nine days (Figure 4.7). They were treated as outliers, and need to be replaced.

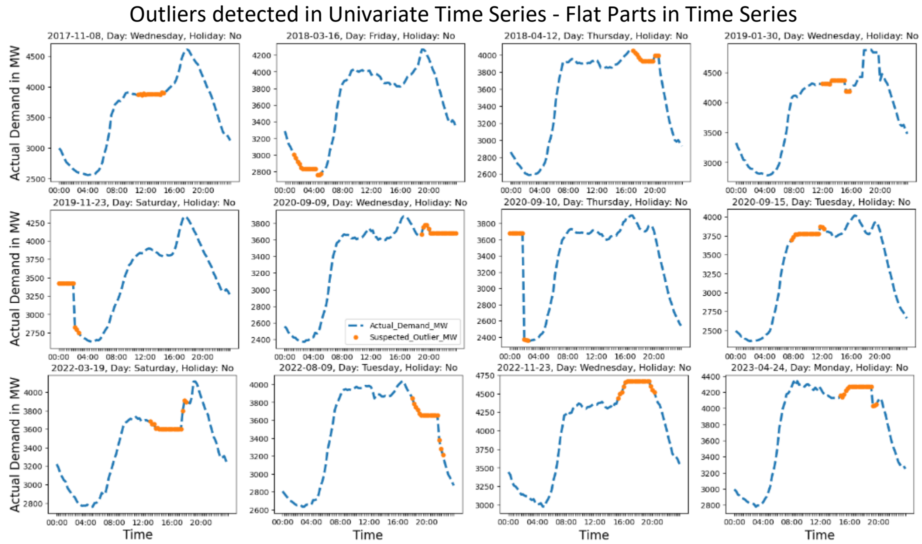


Figure 4.6. Outliers detected in twelve days as flat parts of time series, where the same value repeats more than six times.

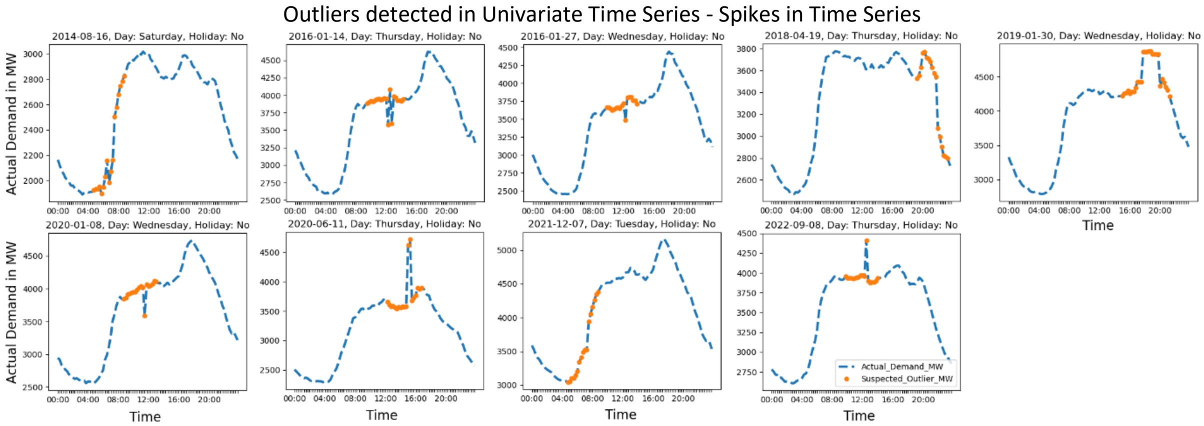


Figure 4.7. Outliers detected in nine days as spikes, where the value changed rapidly, by more than 300MW per single step.

4.2.4. Daylight-Saving Time Distortion in Weather Factors Time Series in Ireland

DST distortion in weather factors time series was found, and needs to be addressed.

4.3. Conclusion

In this chapter, sources and selection of datasets were identified, and raw data was collected for the fifteen-minutes averages of power demand, and one-hour averages of weather factors from twenty-two weather stations in Ireland. Extraction of variables, such as day of the week, holidays observed on weekdays, day of the year and hour, enriched the data by temporal features. Three most important weather factors, such as temperature, relative humidity and wind speed were selected. Initial EDA detected missing values and outliers in time series, including ten days with missing values and twenty-one days with outliers. Moreover, DST distortion was identified in both, power demand and weather factors time series. They need to be addressed to allow proper pattern recognition. Finally, data was trimmed to cover full years 2014-2022.

5. Data Preparation: Time Series of Power Demand in Ireland

5.1. Data Cleaning

5.1.1. Daylight-Saving Time Distortions Removal

Time series, affected by DST, was shifted back by one hour to restore their original continuity, moving the one-hour gap from spring to autumn (Figure 5.1).

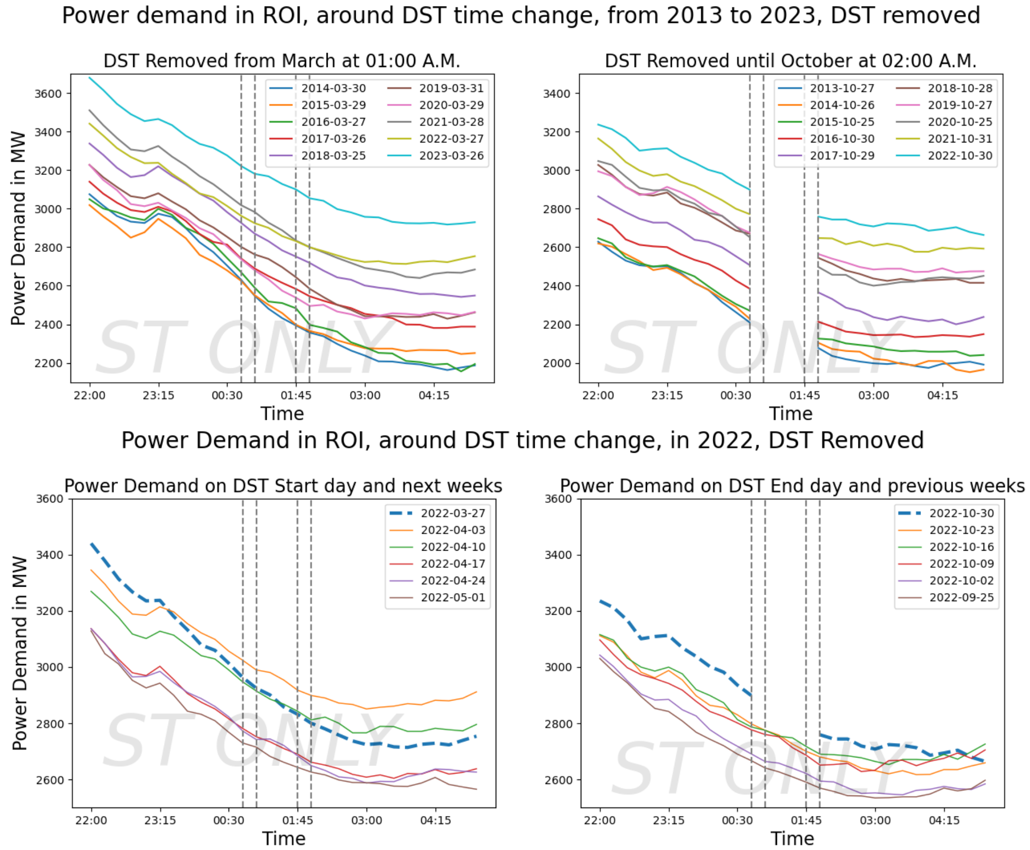


Figure 5.1. DST distortion removal restores original continuity (see years 2013-2023 on the top) and moves the one-hour gap from spring to autumn (see year 2022, days up to 5 weeks after/before the time-change day on the bottom).

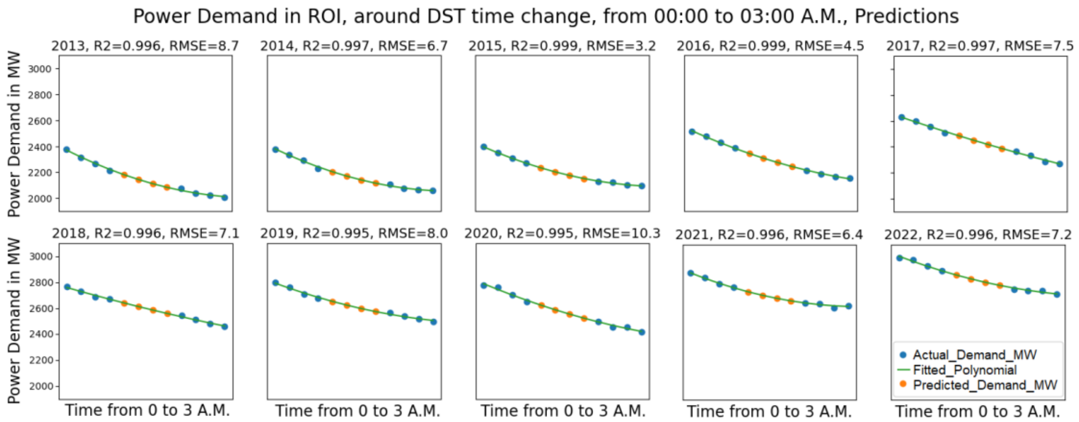


Figure 5.2. The one-hour gap connected with DST removal was filled using Polynomial Regression of second-degree interpolation. Metrics, such as Coefficient of Determination R^2 and RMSE validated the result.

The one-hour gap was then filled using second-degree Polynomial Regression interpolation. New data was validated by R^2 and RMSE metrics (Figure 5.2), as well as visual comparison to Similar Days (SD) up to five weeks before that day (Figure 5.3).

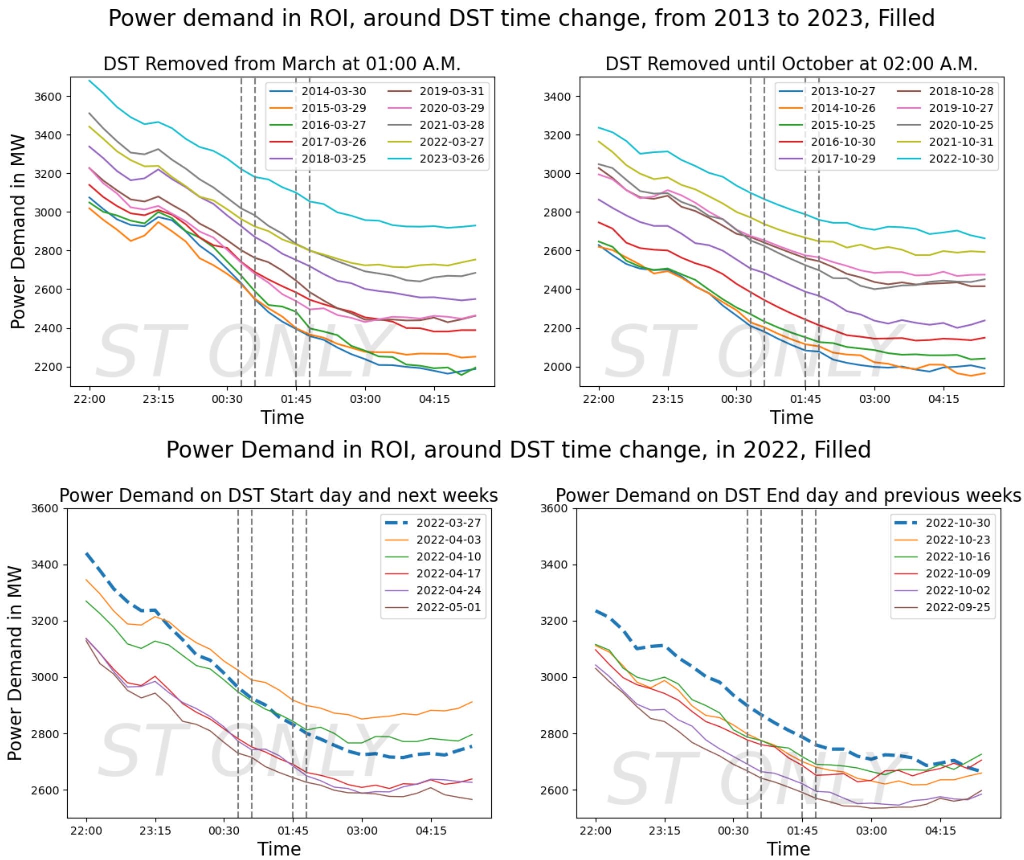


Figure 5.3. Daylight Saving Time (DST) distortion removal and one-hour gap filling using Polynomial Regression of second-degree interpolation restores their original continuity (see years 2013-2023 on the top row and year 2022, similar days up to 5 weeks after/before the time-change day on the bottom row).

5.1.2. Dealing with Missing Data

Single and continuous missing values were filled using Linear Regression (LR) interpolation of neighbouring timestamps (Figure 5.4), and SDs of neighbouring weeks (Figure 5.5), respectively.

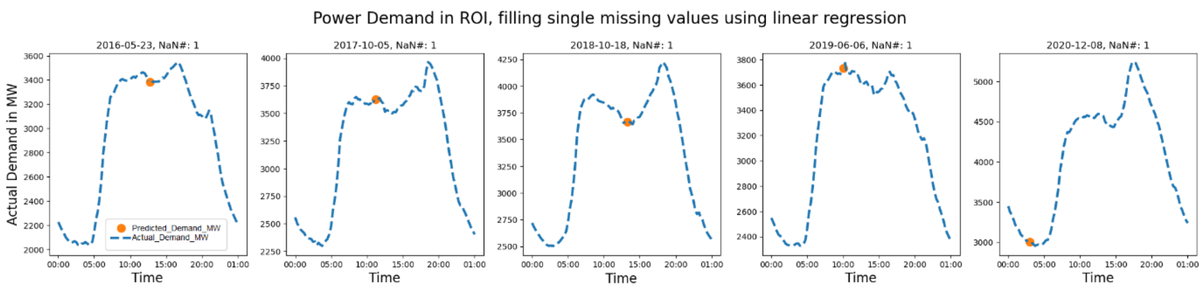


Figure 5.4. Filling single missing values for five days using Linear Regression interpolation.

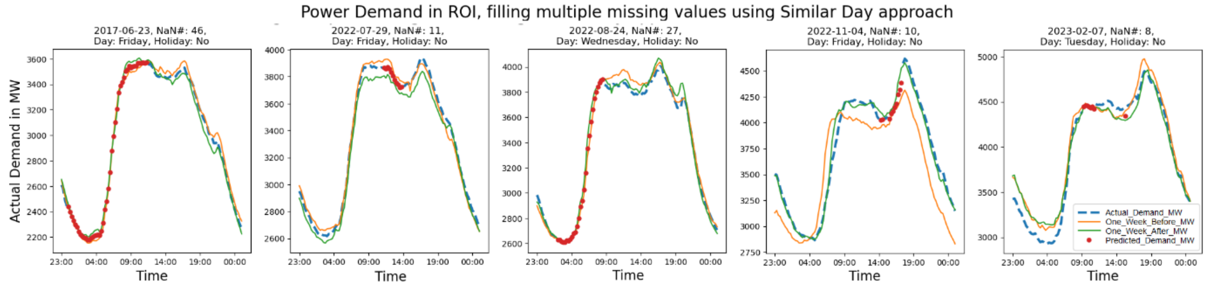


Figure 5.5. Filling continuous missing values for five days using Linear Regression interpolation of similar days.

5.1.3. Dealing with Outliers

Outliers in time series, being flat parts and spikes, detected in previous chapter, were replaced using LR interpolation of SDs from neighbouring weeks (Figures 5.6-5.7).

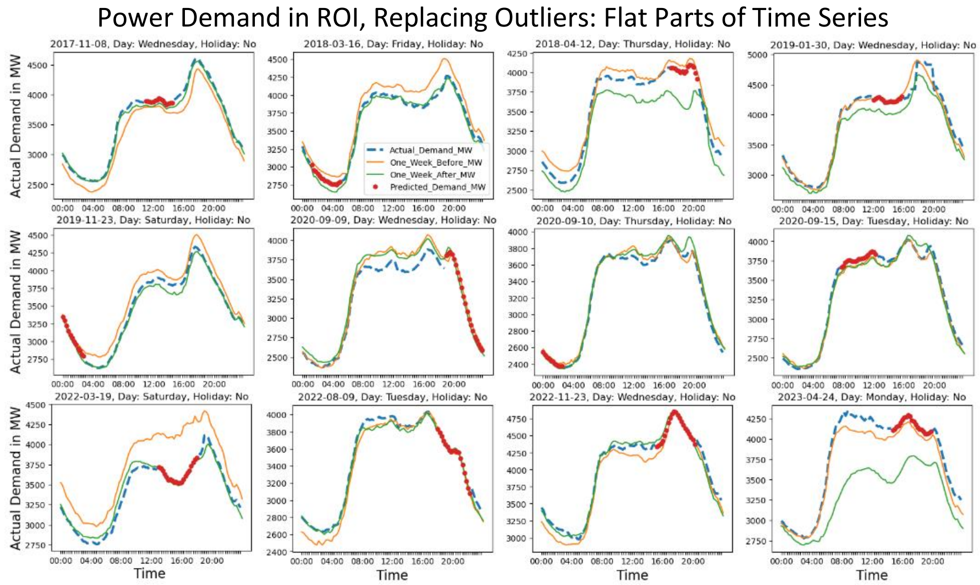


Figure 5.6. Outliers: flat parts replaced using Linear Regression interpolation of similar days one week before and after.

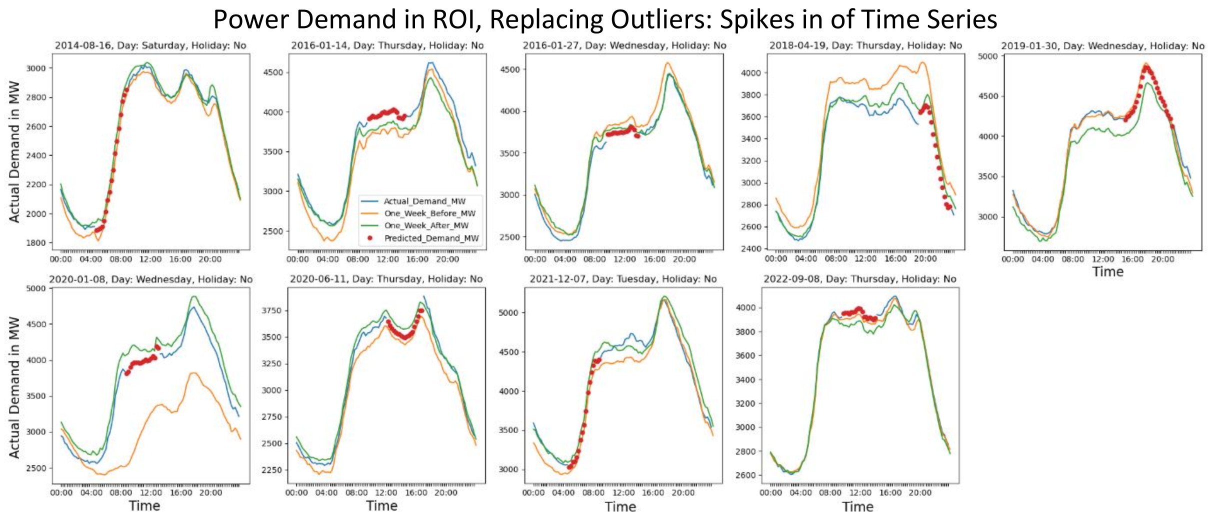


Figure 5.7. Outliers: spikes replaced using Linear Regression interpolation of similar days one week before and after.

5.1.4. Summary and Validation of the Cleaning Process

The count of replaced values in the dataset was shown on Figure 5.8. Validation of the cleaning process was performed by comparison of the power demand time series distributions before and after the process (Figure 5.9). Finally, summary of the cleaning process was given in Table 5.1.

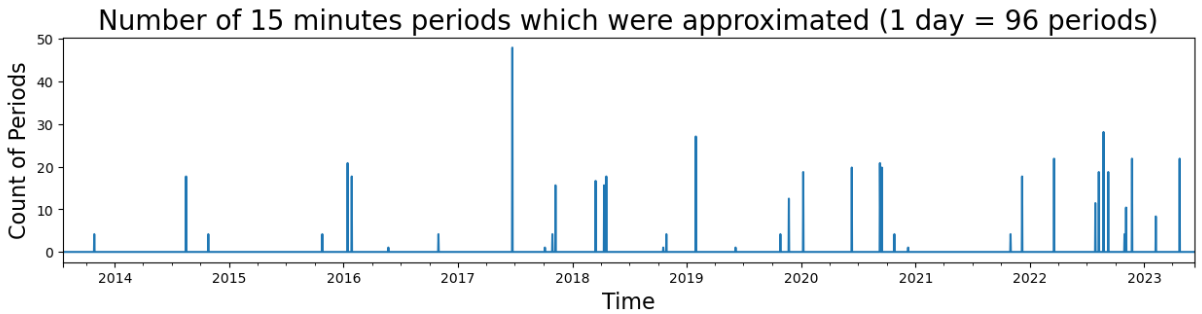


Figure 5.8. The percentage of filled values in the dataset.

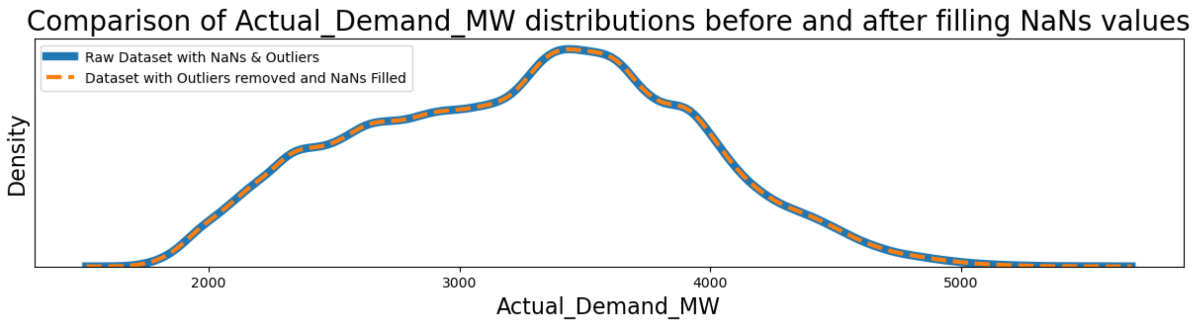


Figure 5.9. Comparison of distributions of Univariate Time Series of Power Demand in Ireland distribution before and after the cleaning process of DST removal, filling NaN's and replacing outliers.

Univariate Data Cleaning Process		No of days affected	Verification of solution	
Problem type	**Solution**		**Primary**	**Final**
DST distortion	Shifting	10	Visual	Comparison of data distributions before and after the cleaning process
DST missing data	Second-degree Polynomial Regression interpolation		R^2, RMSE, visual	
Single missing data	Linear Regression of neighbouring values	5	Visual: similar days up to 5 weeks before and/or after the day	
Continuous missing data	Linear Regression of similar days 1-week before and after the day	5		
Outliers: repeated data		12		
Outliers: spikes		9		

Table 5.1. Summary of the data cleaning process of Univariate Time Series of Power Demand in Ireland.

5.2. Exploratory Data Analysis: Power Demand Time Series

Descriptive statistics, including annual mean, standard deviation, minimum, median, and maximum values, for years 2014-2022 was given in Table 5.2. Year 2020, affected by Covid-19 lockdown, was marked in grey. Distribution of observed holidays in Ireland by day of week for period from July 2013 to June 2023 was given in Table 5.3. The majority of observed holidays in Republic of Ireland occurred on Mondays.

Year	2014	2015	2016	2017	2018	2019	2020	2021	2022
Mean	2492	3034	3092	3167	3299	3319	3338	3529	3606
б	599	591	593	597	604	606	638	611	607
Min	1664	1759	1882	1929	2037	2052	2004	2269	2429
Median	3012	3102	3157	3237	3365	3390	3348	3591	3650
Max	4613	4704	4768	4939	4916	5015	5357	5363	5527

Table 5.2. Descriptive statistics of power demand in Ireland for years 2014-2022

Day of Week	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
# Holidays	61	6	6	6	8	0	0
# Total	516	516	516	516	517	516	516
% Holidays	11.8	1.2	1.2	1.2	1.5	0	0

Table 5.3. Distribution of observed holidays in Ireland by day of week for period from July 2013 to June 2023.

It was observed that from 27/03/2020 to 26/07/2020, during various phases of Covid-19 lockdown in Ireland, the pattern of power demand was significantly disrupted as consequence of regulations introduced at national level (Figure 5.10).

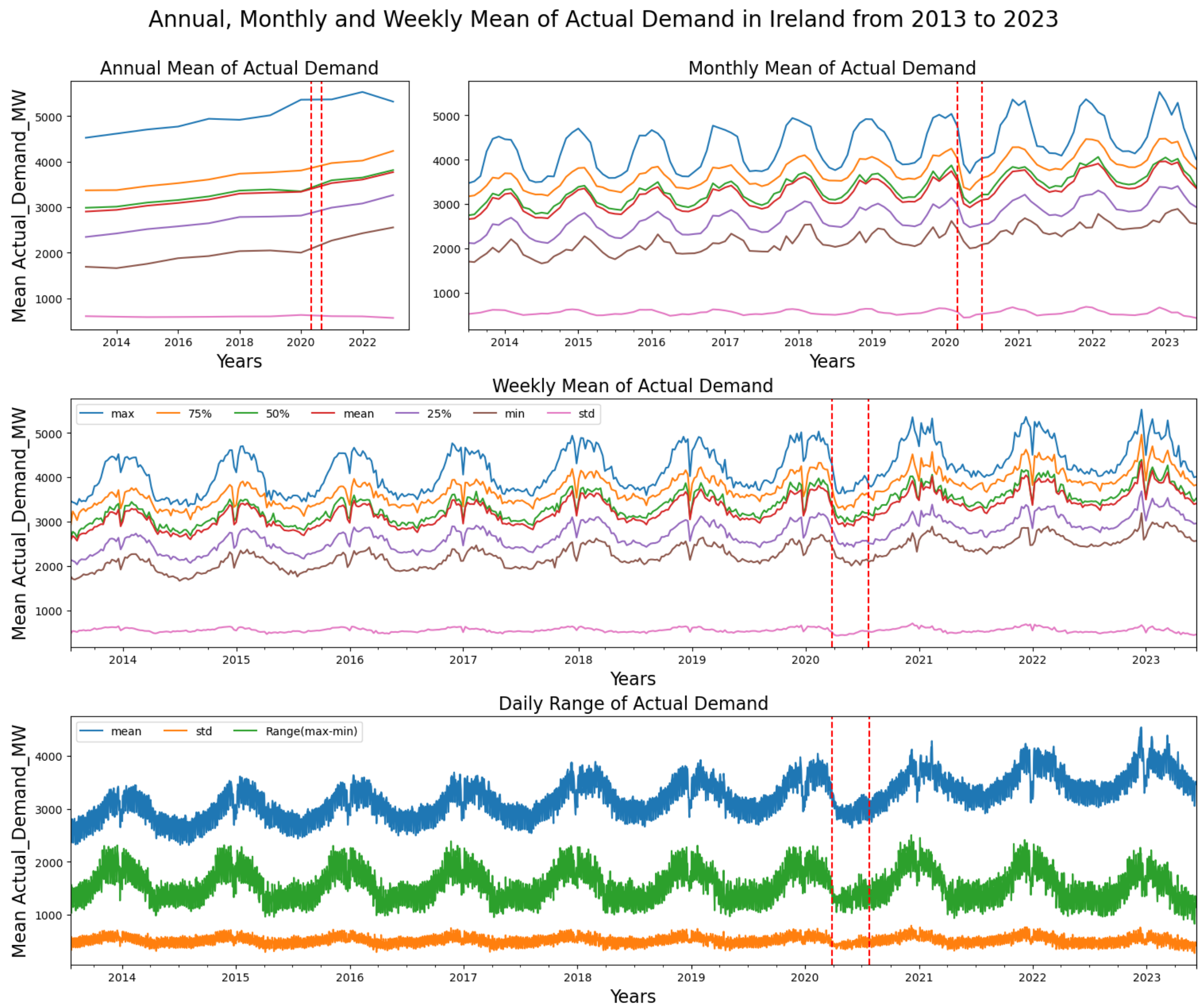


Figure 5.10. Visualisation of univariate time series: annual, monthly and weekly mean of power demand, and daily range of power demand in Ireland. Covid-19 lockdown period marked as dashed red vertical lines.

The annual boxplot of power demand in Ireland from 2013 to 2023 showed the median value increasing every year, while the interquartile range remaining similar. The monthly boxplot suggested annual seasonality, with lowest and highest values in summer and winter, respectively (Figure 5.11).

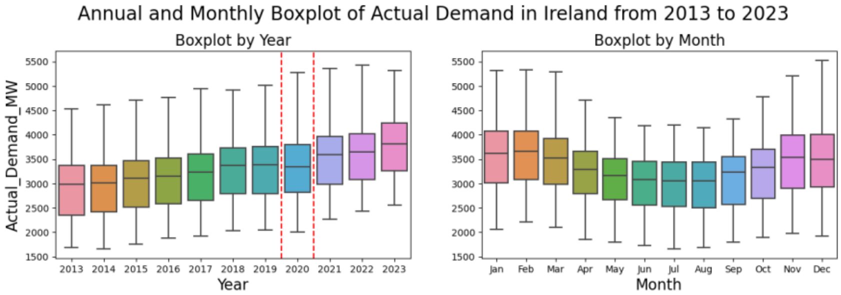


Figure 5.11. Annual (on the left) and monthly (on the right) boxplot of power demand in Ireland. Covid-19 lockdown period marked as dashed red lines.

Furthermore, the monthly boxplot for full years 2014-2022 (Figure 5.12) and daily mean of power demand (Figure 5.13) confirmed that year 2020 might be indeed considered as an outlier.

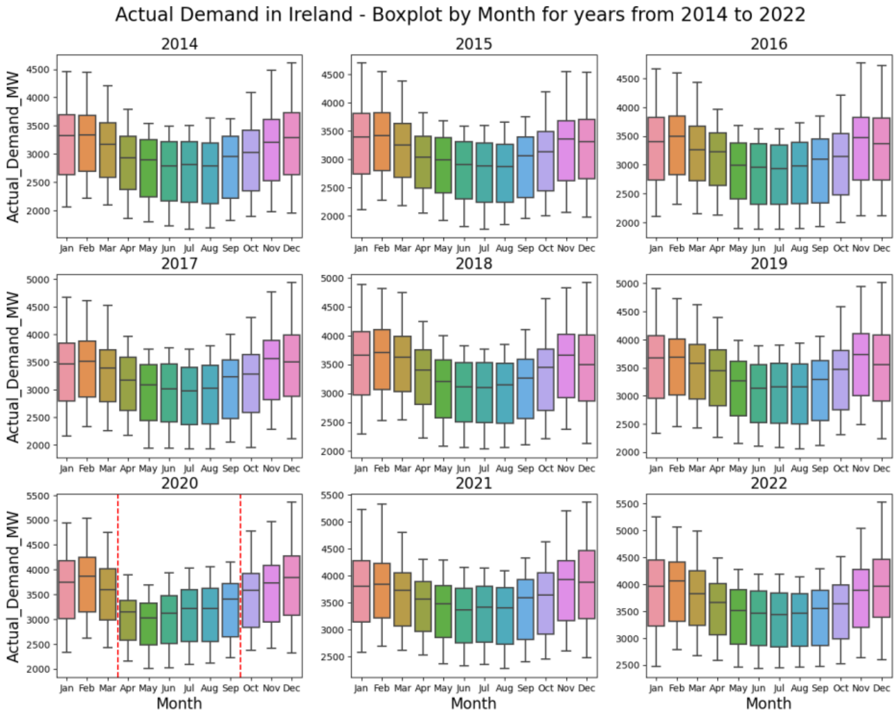


Figure 5.12. Monthly boxplots of power demand in Ireland by years. Covid-19 lockdown period marked as dashed red lines.

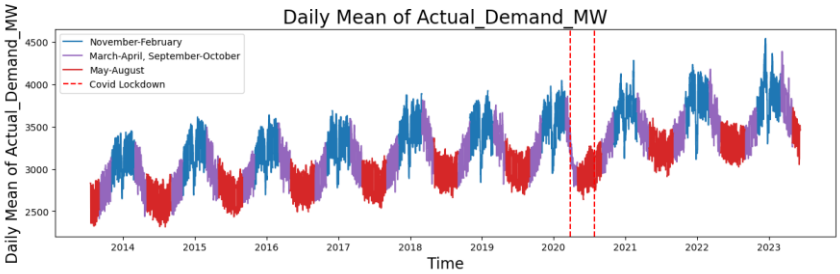


Figure 5.13. Daily mean of power demand in Ireland, with colour-marked seasons. Covid-19 lockdown period marked as dashed red vertical lines.

To find similarities and differences between days of the week, power demand was normalised by its daily mean values and visualised as average normalised demand for years 2014-2021 and 2022 (Figure 5.14). It was noticed that days of the week and holiday on weekday could be grouped. Tuesdays shared similar pattern with Wednesdays and Thursdays, and Sunday shared similar pattern with holidays on weekdays (Figure 5.15).

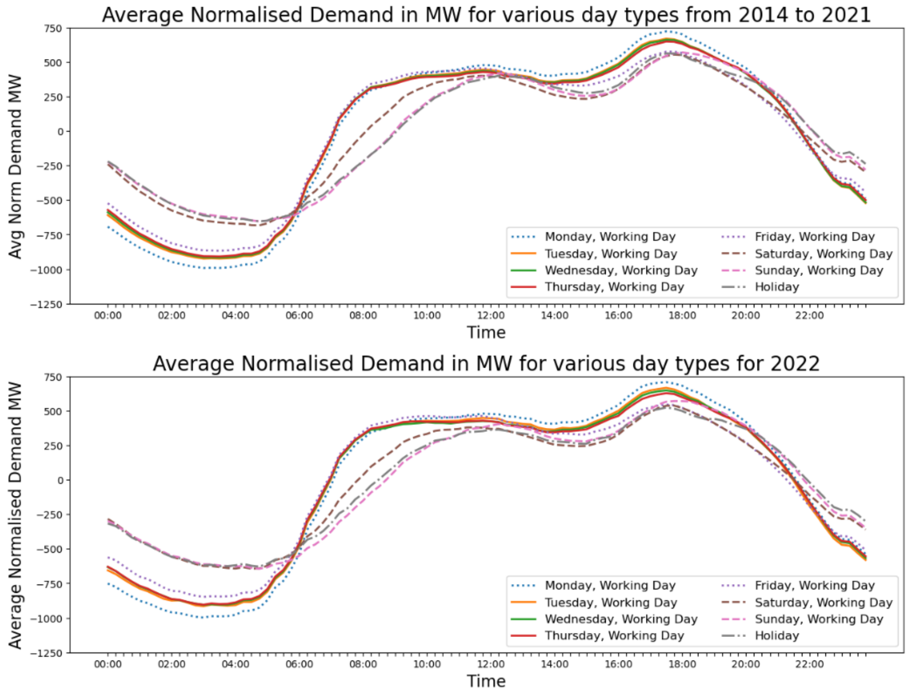


Figure 5.14. Average normalised demand in Ireland, for various day types from 2014 to 2021 (top) and 2022 (bottom).

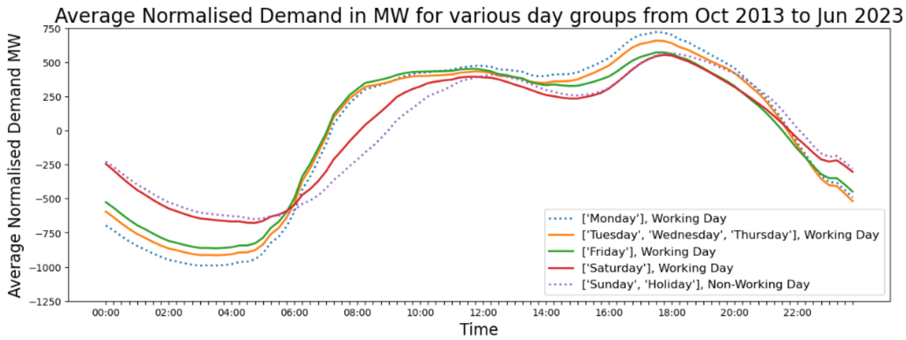


Figure 5.15. Average normalised power demand in Ireland, for grouped day types from 2013 to 2023.

To investigate the pattern further, full normalised dataset was plotted for each of the groups individually, with colour-marked seasons. The Tuesdays-Wednesdays-Thursdays group was plotted as grey silhouette for other groups. The evening peak was the highest on Mondays in winter, and the lowest on Sundays and holidays on weekdays in summer. The night minimum was the lowest on Mondays in winter, and the highest on Saturdays, Sundays and holidays on weekdays. Furthermore, morning demand rose later in winter than in summer, and was shifted forward for Saturdays, Sundays and holidays on weekdays (Figure 5.16).

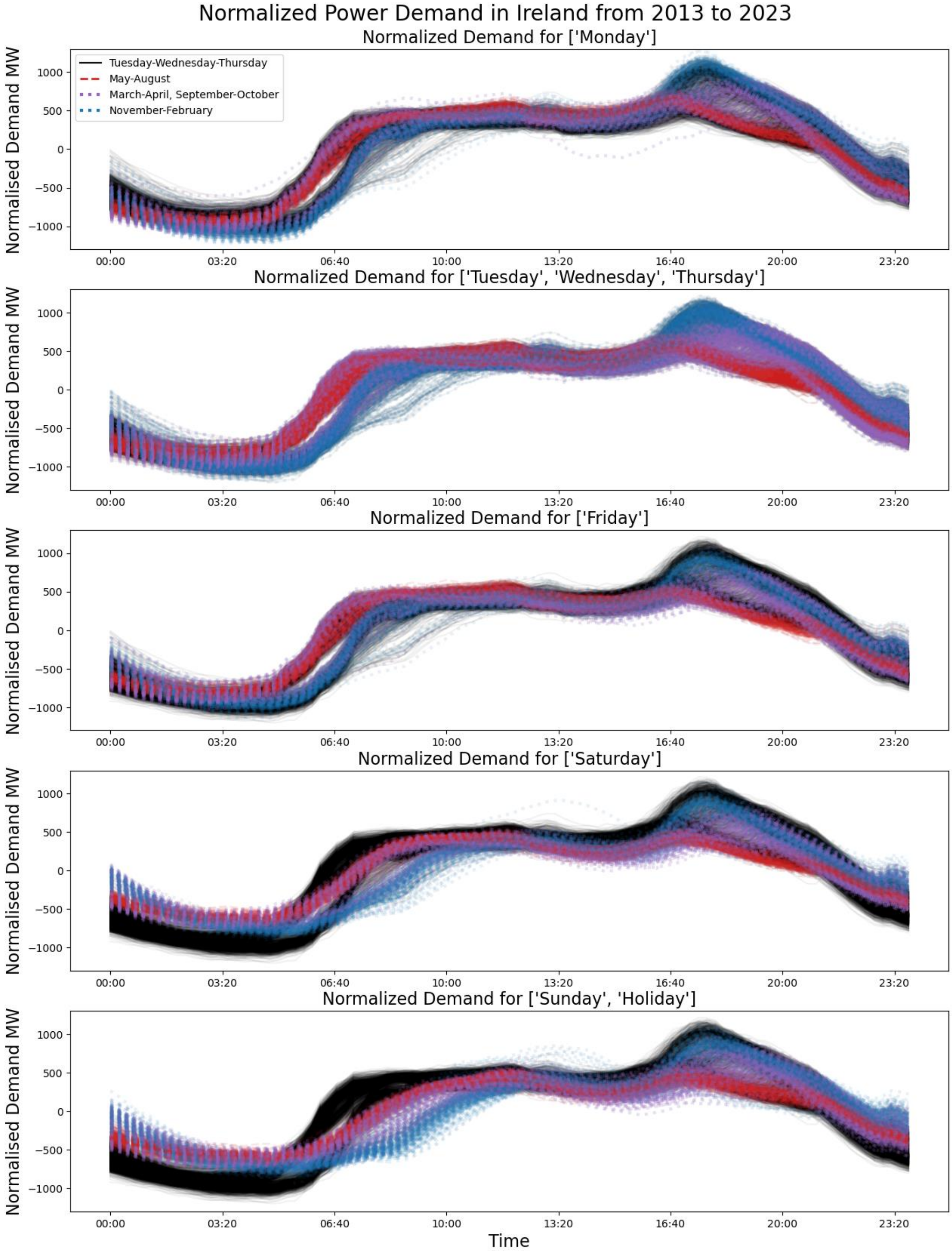


Figure 5.16. Normalised power demand in Ireland, for grouped day types from 2013 to 2023, with marked seasons.

The boxplot of normalised dataset for each fifteen-minute period and all groups individually (Figure 5.17), confirmed observations from previous paragraph. Moreover, it clearly depicted higher values of interquartile range of normalised demand in the mornings and evenings.

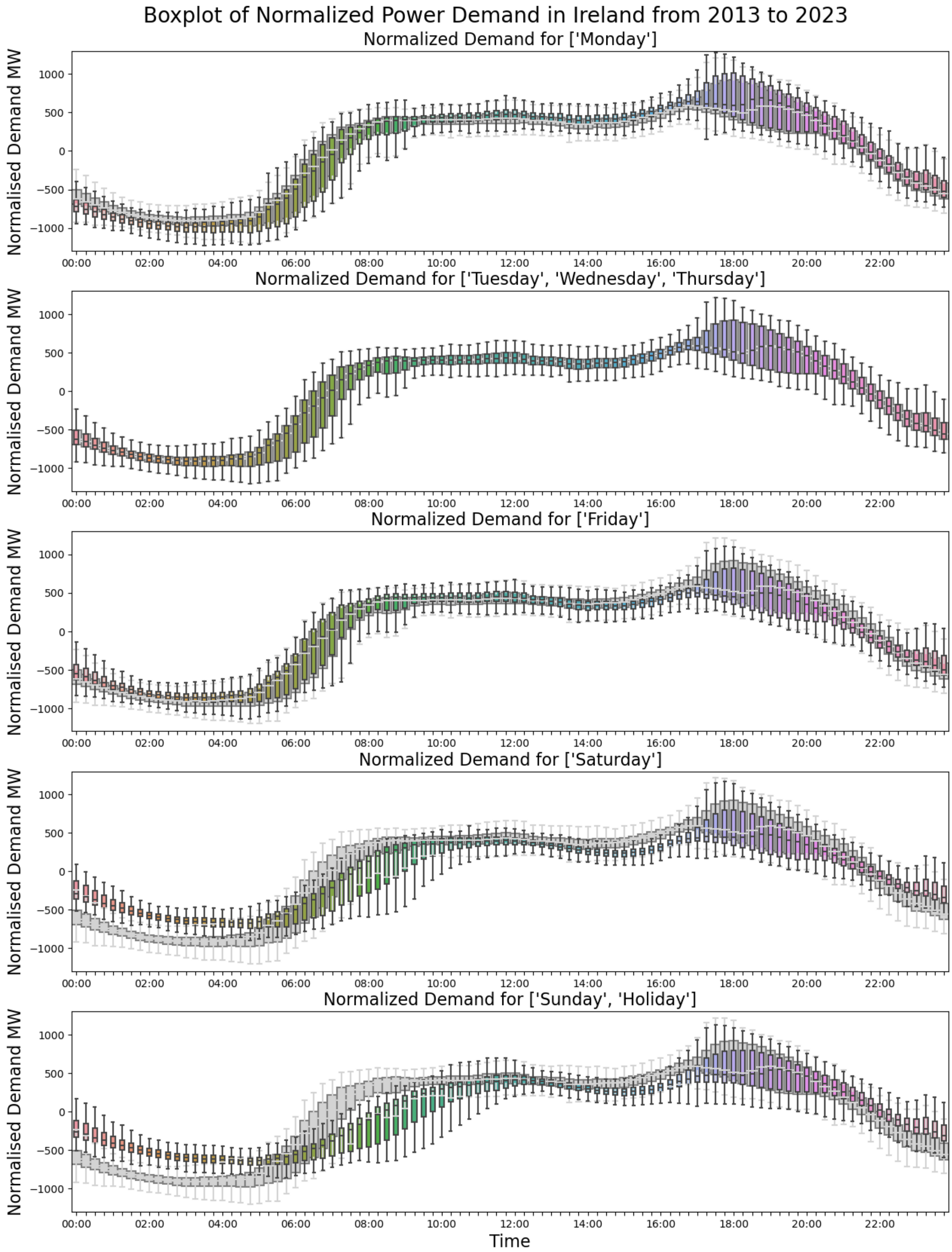


Figure 5.17. Boxplot of normalised power demand in Ireland, for grouped day types from 2013 to 2023, with Tuesday-Wednesday-Friday group shadowed in other graphs.

Density histograms of power demand and its daily mean were similar for working days but shifted towards lower values for Saturdays, Sundays and holidays on weekdays. Moreover, normalised power demand showed narrower range on those days in comparison to working weekdays (Figure 5.18).

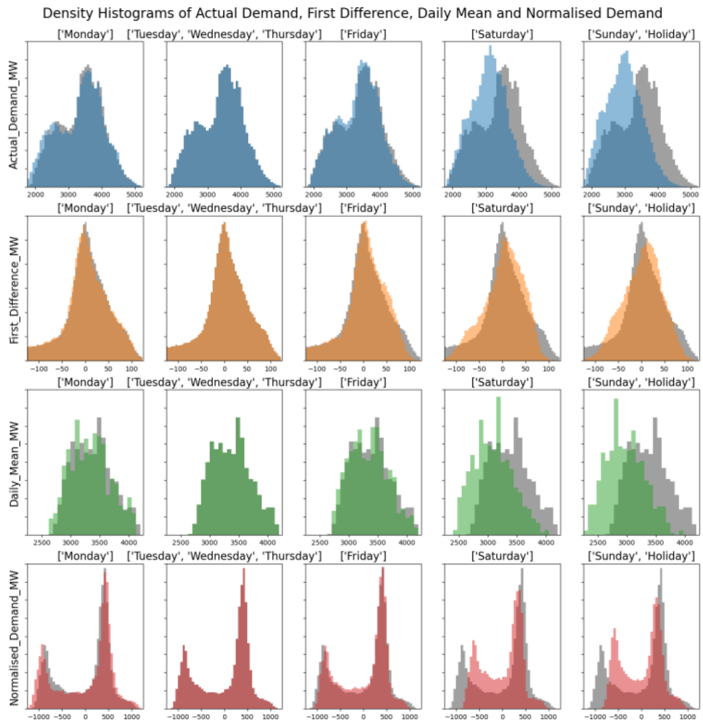


Figure 5.18. Density histograms of original and first-order differenced power demand, daily mean, and normalised power demand in Ireland, for grouped day types, with Tuesday-Wednesday-Friday group shadowed in other graphs.

The original fifteen-minutes time series was resampled to one-hour and daily time series, by averaging the values, for consistency with current research, and further analysis, respectively.

Shapiro-Wilk normality tests for power demand, daily mean of power demand and normalised power demand, with null hypotheses that hourly and daily time series were normally distributed revealed, that there was enough evidence to reject the null hypothesis and accept the alternative hypothesis, that tested time series were not normally distributed. Above was consistent with density histograms and Q-Q plots on Figure 5.19.

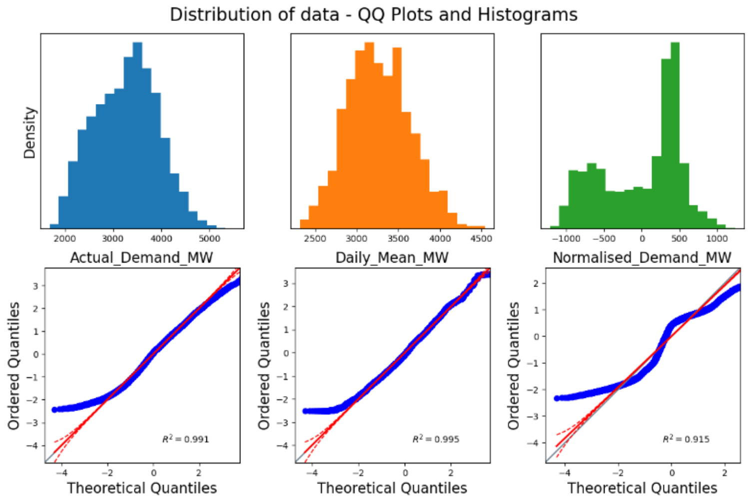


Figure 5.19. Density histograms and Q-Q plots of power demand, daily mean, and normalised power demand in Ireland, for grouped day types from 2013 to 2023, with Tuesday-Wednesday-Friday group shadowed in other graphs.

Mann-Kendall trend tests for power demand, daily mean of power demand and normalised power demand, with null hypotheses that there was no trend in hourly and daily time series revealed, that there was enough evidence to reject the null hypothesis and accept the alternative hypothesis, that there was trend in tested time series, with a positive slope value.

ADF stationarity tests for power demand, with null hypothesis that hourly and daily time series were not stationary revealed, that there was not enough evidence to reject the null hypothesis. However, KPSS stationarity test, with null hypothesis that the time series were stationary revealed that there was enough evidence to reject the null hypothesis and accept the alternative hypothesis that both time series were not stationary. Therefore, based on Statsmodels (2023), both time series were difference stationary. Performing tests for first-order differenced demands confirmed that first-order differenced time series were stationary.

ACF and PACF plots (Figure 5.20) were used for seasonality tests, by determining the significance of the lags. The blue shaded area signifies the 95% confidence intervals, calculated under the null hypothesis that the data are independently distributed. The blue dots are values of ACF and PACF for particular lag. Values beyond blue shaded area would indicate they are statistically significantly different from zero. Despite visible daily and weekly patterns in ACF, and annual pattern in PACF, in first-order differenced daily time series, all values lie within the 95% confidence intervals. Therefore, there was no statistically significant seasonality in either of them.

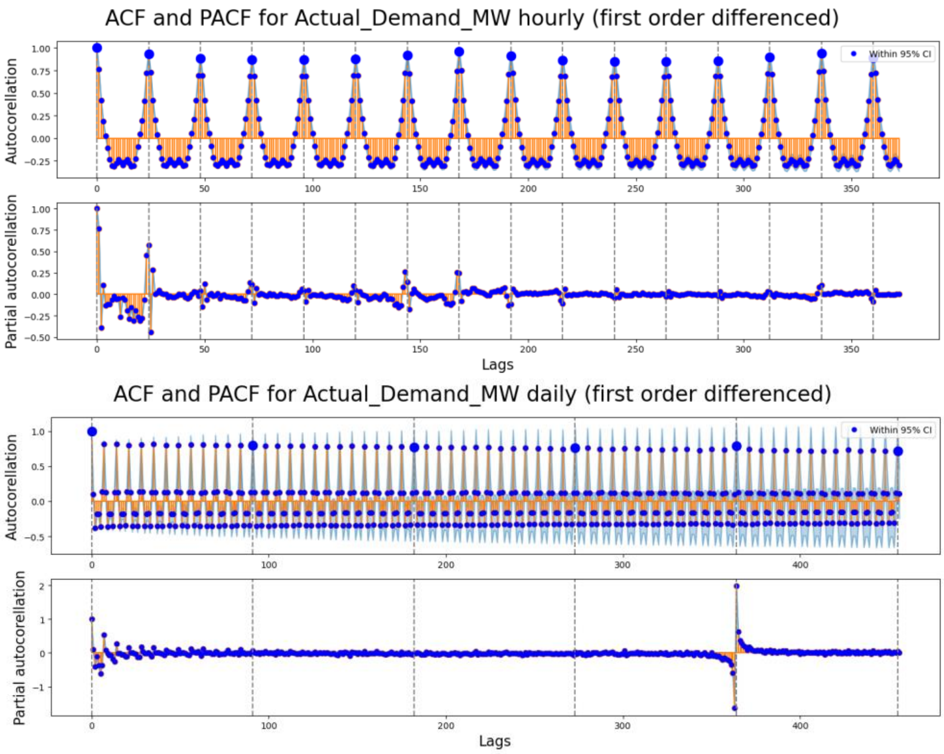


Figure 5.20. ACF and PACF plots for first-order differenced hourly (top) and daily (bottom) power demand time series.

The classical decomposition was not able to break down power demand correctly, due to existence of multi-seasonal patterns in the time series. Therefore, Multiple Seasonal Patterns Decomposition (MSTL), proposed by Bandura, Hyndman and Bergmeir (2021) was performed on hourly time series (Figure 5.21). The first component illustrated trend, with values 2900-3700MW, where the value experienced a steady increase over time, followed by a pronounced acceleration in 2018, which was succeeded by a more moderate incline in 2019. Covid-19 disturbed the 2020 trend significantly. Finally, the trend reverted to its initial steady increase, continuing to rise at a gradual rate. Further three components depicted daily, weekly and annual seasonality. The daily pattern, with values -1000-1000MW, began with a nightly low, rose to a flat midday level, then climbed to an evening peak, and finally decreased at night, completing the cycle. The weekly pattern, with values -750-500MW, caught weekdays and weekends, and the annual pattern varied -750-500MW, leaving residuals in range -250-250MW. That confirmed daily, weekly and annual seasonality in time series.

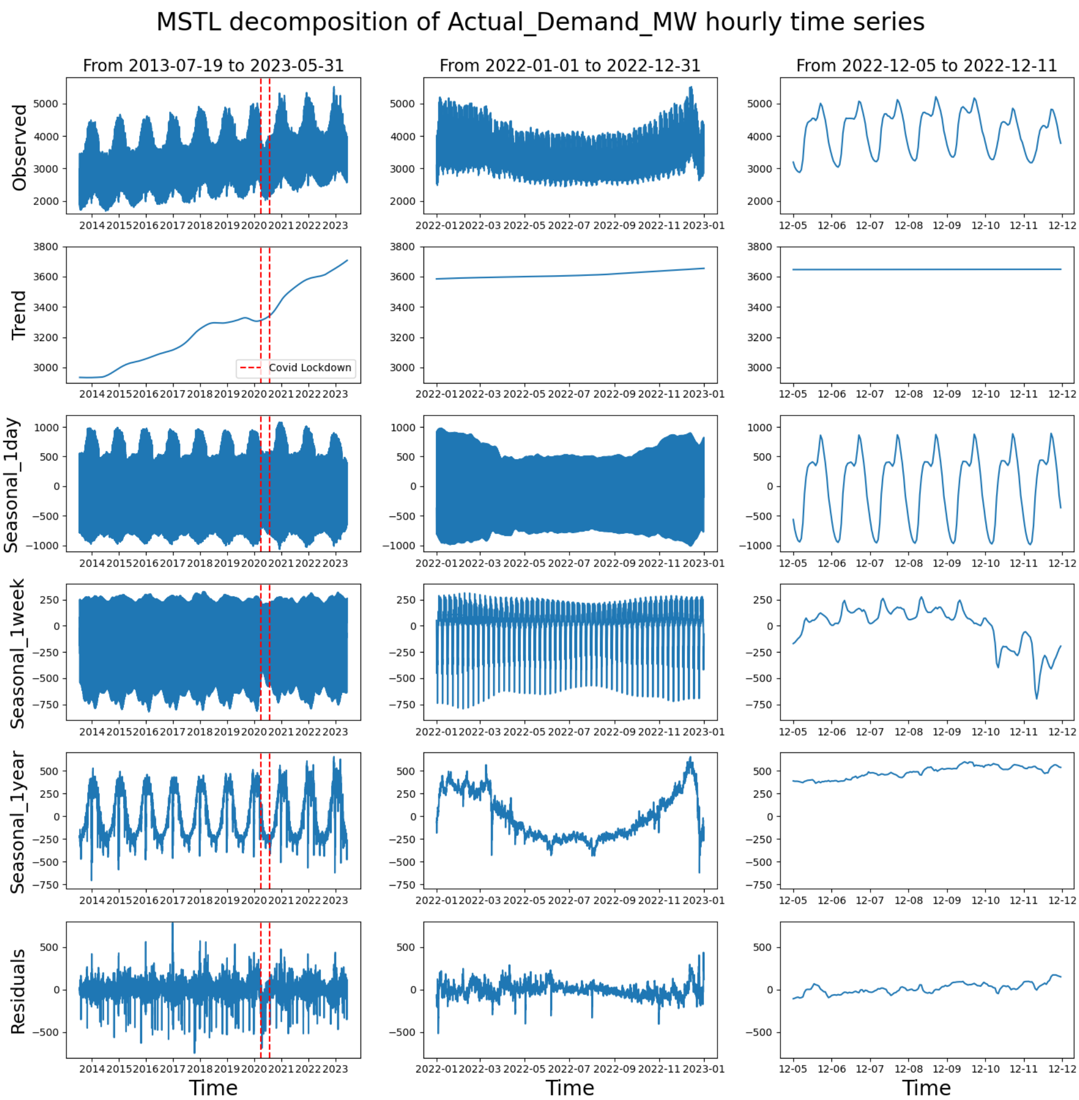


Figure 5.21. MSTL decomposition of hourly power demand daily time series. Covid-19 lockdown period marked as dashed red lines. Columns present years 2013-2023, 2022, and one week in December 2022.

5.3. Investigation into Weekly Lagged Power Demand Time Series

To assess the potential use of weekly lagged time series in ODADF, an investigation was carried out to identify the values with the highest correlation for the SD approach (Figure 5.22). It revealed that lagging the demand by range 1-5 and 47-57 weeks yielded a correlation between variables exceeding 90%.

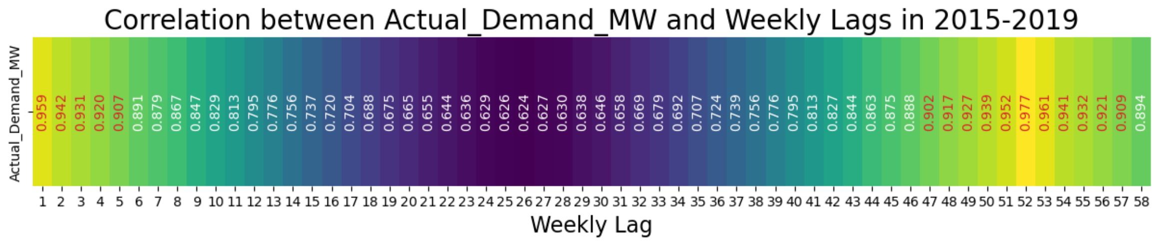


Figure 5.22. Correlation between actual and weekly lagged power demand in Ireland.

To estimate the influence of a single k-weekly, and a range from 1 to k-weekly lagged time series on ODADF MAPE, Linear Regression was used. Using adjusted values for one week ago or one year (52 weeks) ago, gave MAPE in range 3.16-3.57% and 2.64-4.45%, respectively, with years 2020-2021 being the most affected by Covid-19 disturbances for the latter. As the value of k in range of 1 to k-weekly lagged values increased, the MAPE steadily decreased. A noticeable plunge in the MAPE was observed around value k=52, reaching MAPE in range 2.11-2.63%, after which the decline continued but at a slower pace (Figure 5.23).

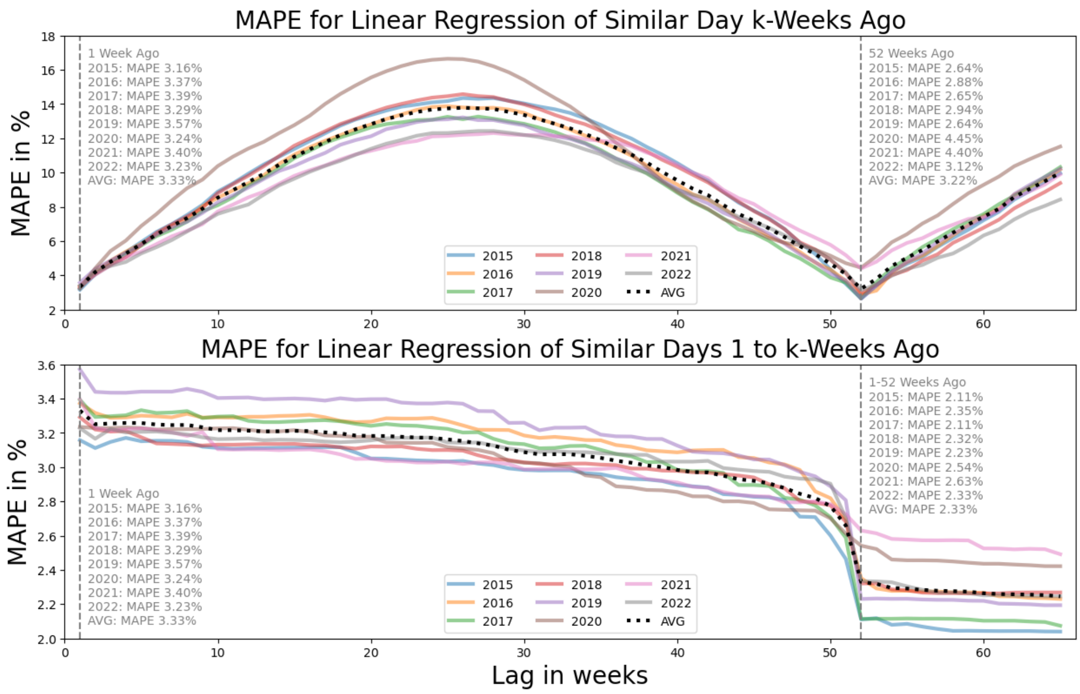


Figure 5.23. Influence of single k-weekly, and a range from 1 to k-weekly lagged time series on MAPE of one day-ahead forecasting of power demand in Ireland, using linear regression.

5.4. Investigation into Daily Lagged Power Demand Time Series

To assess the potential benefits of using daily lagged time series over weekly ones in ODADF, an investigation was undertaken to identify the values with the highest correlation (Figure 5.24).

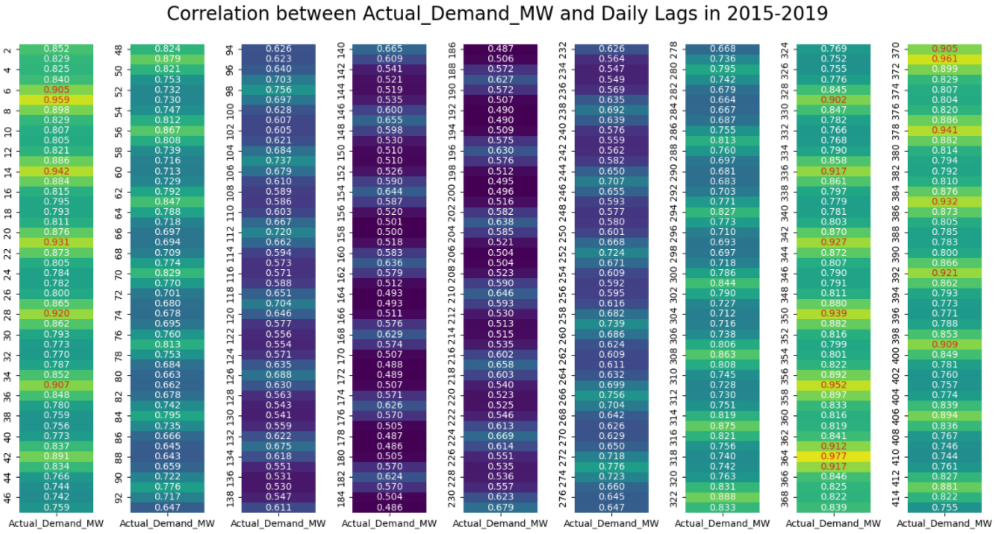


Figure 5.24. Correlation between actual and daily lagged power demand in Ireland.

To estimate the influence of single k-daily, and a range from 2 to k-daily lagged time series on ODADF MAPE, Linear Regression was used. While MAPE from k-days ago demonstrated weekly seasonality, it decreased steadily for the range of 2 to k daily lagged values, with increasing value of k. A noticeable plunge in the MAPE was observed around value k=7, reaching MAPE in range 3.12-3.58%, after which the decline continued but at a slower pace (Figure 5.25). Therefore, using daily lagged time series would not be more beneficial than the weekly ones.

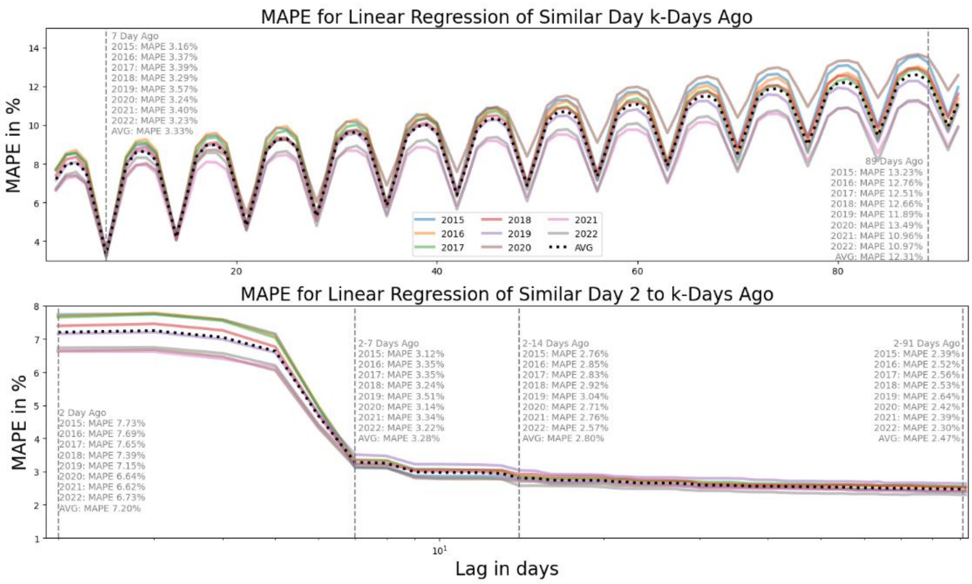


Figure 5.25. Influence of single k-weekly, and a range from 1 to k-weekly lagged time series on MAPE of one day-ahead forecasting of power demand in Ireland, using linear regression.

5.5. Investigation into Moving Window Size of Power Demand Time Series

To determine the optimal window size for ODADF using Moving Window (MW) approach, an investigation was carried out using Linear Regression (Figure 5.26). MAPE decreased steadily with increasing value of k. A noticeable plunge in the MAPE was observed around value k=7 and k=14, reaching MAPE in range 2.51-2.80%, and 2.24-2.59%, respectively. Then, the decline continued but at a slower pace, reaching minimum around k=36, achieving MAPE in range 2.19-2.50%. Further increasing of k value increased MAPE (Figure 5.26).

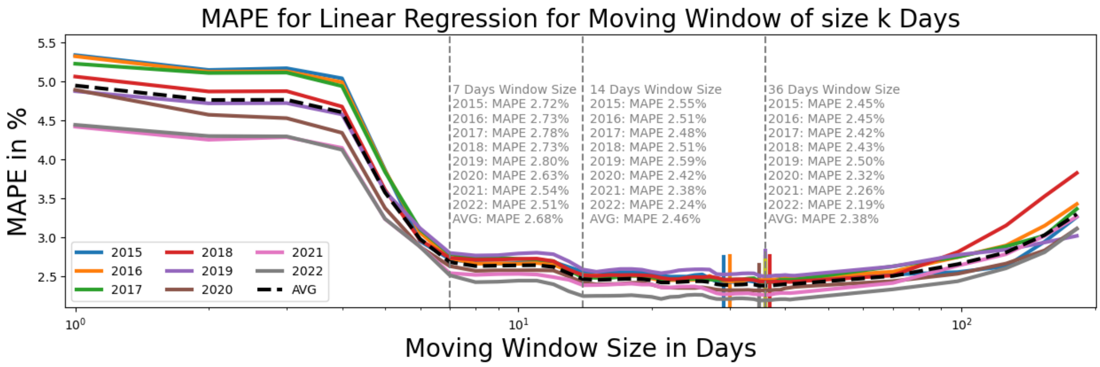


Figure 5.26. Influence of window size on MAPE of forecasting of power demand in Ireland, using linear regression.

5.6. Scaling and Encoding the Data for Modelling

Power demand and temporal features were scaled and encoded (Table 5.4) as described in Chapter 3. The conversion approach was validated by correlation between scaled power demand and encoded features (Figure 5.27).

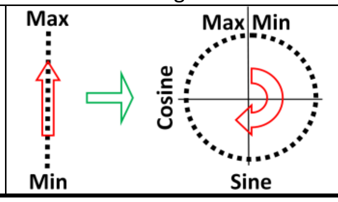
Variable	Original Range	Conversion		Converted Range
Power Demand	1684.25 to 5517.75	Scaling by division by 5517.75		0.3052 to 1.0000
Day of the Week	Monday to Sunday	One-Hot Encoding		[1, 0, 0, 0, 0, 0, 0]
Observed Holiday	Yes or No	Numerical Encoding		1 or 0
Day of the Year	1 to 365 (366)	Cyclical Encoding		[-1 to 1, -1 to 1]
Hour of the Day	0 to 23			

Table 5.4. Scaling and encoding the data for modelling.

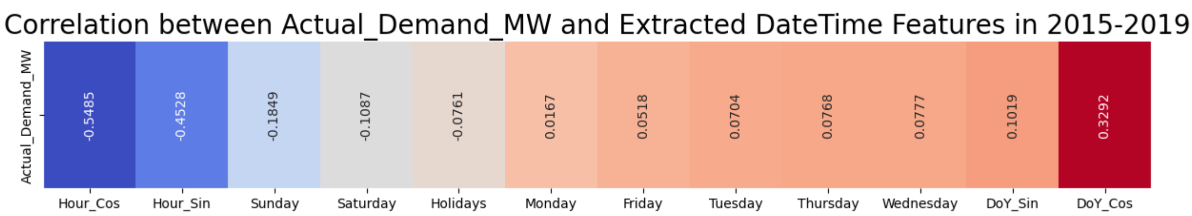


Figure 5.27. Correlation between scaled power demand and extracted and encoded temporal features

5.7. Evaluation of Baseline Models

Baseline models, described in Chapter 3, were considered for ODADF as a starting point and reference for comparison with more complex models developed in experimentation phase of the research. The results, being MAPE per day type, month, and hour, as well as visualisation of baseline models predictions against actual demand for year 2019, were shown on Figures 5.28-5.29.

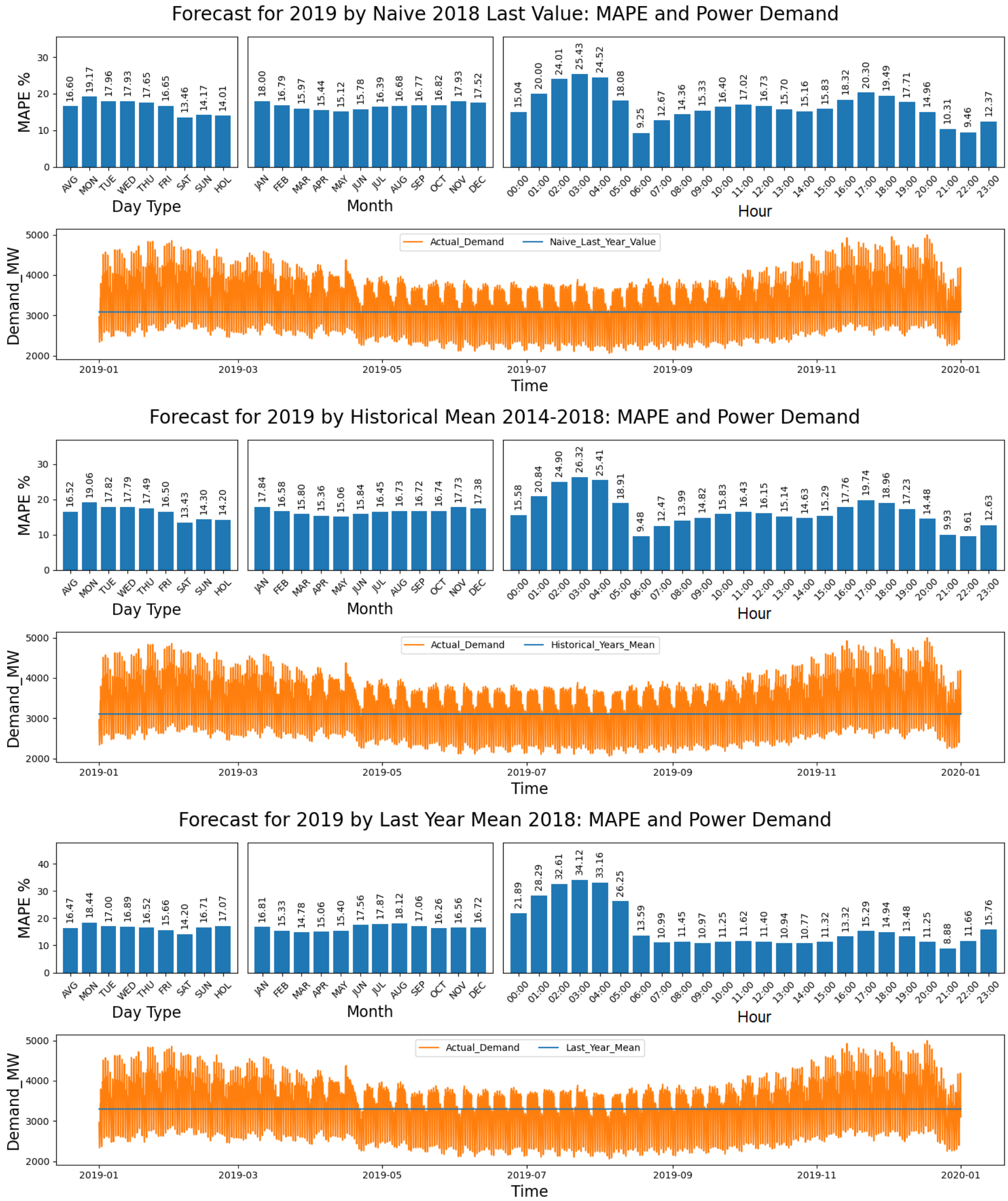


Figure 5.28. Forecast of power demand in Ireland for year 2019 by baseline models (1-3).

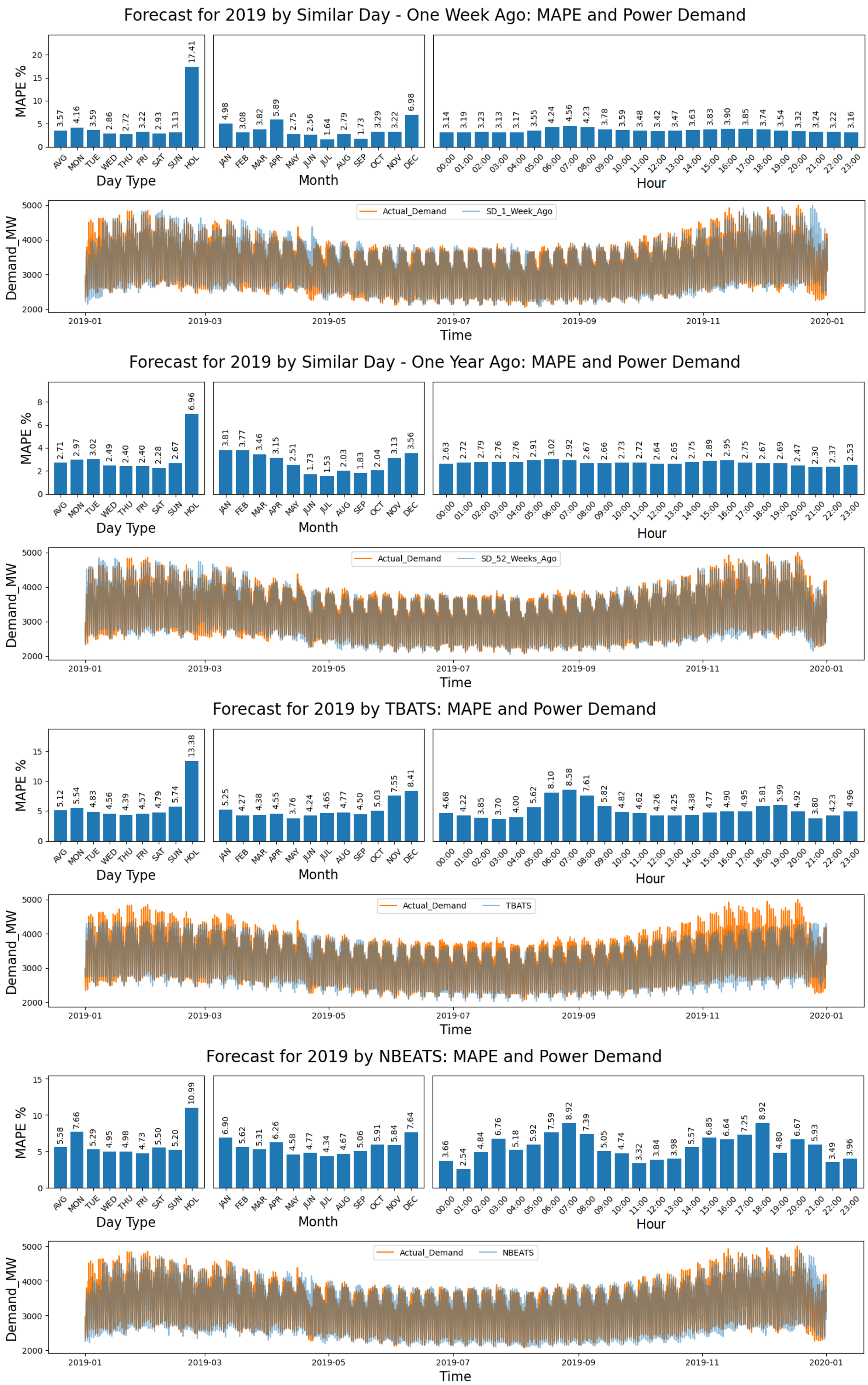


Figure 6.17. Forecast of power demand in Ireland for year 2019 by baseline models (4-7).

5.8. Defining Supervised Learning Problems for Similar Day and Moving Window Methods

Supervised learning problems were defined from power demand time series and temporal features by SD approach, utilising historical weekly lags of the power demand (Table 5.5), and MW approach, utilising continuous historical demand values (Table 5.6). In practice, ODADF is conducted several hours in advance. Therefore, to avoid the **double forecasting phenomenon**, partially available current-day data was excluded from consideration.

Record Time Index In Hours	Independent Variable X					Dependent Variable Y
	Temporal Variables				Lagged Power Demand Values	Power Demand Y
	Holidays H	Day of Week DoW	Day of Year DoY	Hour of Day HoD		
$t_K=t_0+k\cdot h$	$H(t_K)$	$DoW(t_K)$	$DoY(t_K)$	$HoD(t_K)$	$[Y(t_K-n_1\cdot 7\cdot 24\cdot h, Y(t_K-n_2\cdot 7\cdot 24\cdot h, \ldots, Y(t_K-n_L\cdot 7\cdot 24\cdot h]$ where: k=1, 2, … (record index) L: number of weekly lags t_K: timestamp k in hours $n_1, n_2, \ldots n_L$: number of lags in weeks	$Y(t_K)$

Table 5.5. Supervised learning problem definition for similar days approach (univariate time series of power demand in Ireland and temporal features).

Record Time Index In Hours	Independent Variable X					Dependent Variable Y
	Temporal Variables				Moving Window of Power Demand Values	Power Demand Y
	Holidays H	Day of Week DoW	Day of Year DoY	Hour of Day HoD		
$t_K=t_0+k\cdot h$	$H(t_K)$	$DoW(t_K)$	$DoY(t_K)$	$HoD(t_K)$	$[Y(t_K-(24\cdot(1+w_S)-1)\cdot h,$ $Y(t_K-(24\cdot(1+w_S)-2)\cdot h,$ $\ldots, Y(t_K-(24\cdot(1+w_S)-w_S)\cdot h]$ where: k=1, 2, … (record index) w_S: window size in days current day not considered	$Y(t_K)$

Table 5.6. Supervised learning problem definition for moving window approach (univariate time series of power demand in Ireland and temporal features).

5.9. Conclusion

In this chapter, supervised learning problems for SD and MW approaches were defined, and potential weekly lags and window size were identified, respectively.

Firstly, DST distortion was removed from power demand time series, and missing values and outliers were replaced using appropriate methods, which were proved to be valid by comparison of distributions from power demand datasets before and after the process.

Secondly, descriptive statistics revealed possible daily, weekly and annual seasonality of power demand in Ireland, disrupted by Covid-19 lockdown in Ireland in 2020. Furthermore, similarities and differences between daily patterns were identified. Subsequently, inferential hypotheses tests revealed that power demand time series was not normally distributed, had positive trend, was first-order difference stationary and, despite visible daily and weekly patterns, its seasonality was not statistically significant. Therefore, Auto-Regressive models were not expected to perform well for ODADF in Ireland.

Then, baseline models were used as starting point and reference for more complex models developed in experimentation phase of the research, achieving 90.7-553.9MW, 128.0-647.2MW, and 2.71-16.60% for MAE, RMSE, and MAPE, respectively.

Finally, power demand and temporal features were scaled and encoded. The day of the year and hour of the day variables were converted to cyclical format, to highlight their seasonality, while the day of the week variable was encoded to sparse vector format. The validity of that approach was confirmed through a correlation study.

6. Multivariate Data Preparation

6.1. Weather Factors' Data Cleaning - Daylight-Saving Time Distortions Removal

Weather factors time series, affected by DST, were shifted back by one hour to restore their original continuity, and one-hour gaps in autumn were filled using monthly mean values.

6.2. Exploratory Data Analysis: Weather Factors' Time Series

Descriptive statistic for three weather factors from twenty-two weather stations in Ireland, for years 2014-2022 was performed for all, and given for three exemplary weather stations in Table 6.1. Considering the variations in the values of respective factors among weather stations, and to mitigate the challenges posed by high dimensionality of data, potential solutions for identifying representative stations were outlined in section 6.4.

Weather station	Athenry			Shannon Airport			Valentia Observatory		
	Temperature °C	Relative Humidity %	Wind Speed KMH	Temperature °C	Relative Humidity %	Wind Speed KMH	Temperature °C	Relative Humidity %	Wind Speed KMH
Mean	10.0	83.5	13.4	10.7	81.9	16.6	11.3	80.6	17.8
б	4.9	11.6	7.4	4.8	12.3	9.7	3.9	11.4	10.4
Min	-8.0	27.0	0.0	-4.8	24.0	0.0	-4.9	22.0	0.0
Median	10.1	87.0	13.0	10.8	85.0	14.8	11.3	82.0	16.7
Max	30.1	100.0	61.1	31.5	100.0	96.3	27.9	100.0	88.9

Table 6.1. Descriptive statistics of weather factors from three exemplary weather stations in Ireland for years 2014-2022

6.3. Correlation Between Lagged Weather Factors and Power Demand

To evaluate correlations between weather factors and power demand time series, full datasets were plotted individually for power demand, temperature, relative humidity and wind speed from all twenty-two weather stations, with colour-coded seasons (Figure 6.1). They revealed that all weather factors were correlated with power demand, with temperature and wind speed showing a negative correlation, while relative humidity demonstrated a positive correlation.

To estimate correlations between lagged weather factors and power demand time series, correlation plots for each factor against the value of lag in hours were plotted, for each weather station (Figure 6.2). All factors exhibited daily periodicity, with peak values occurring at a lag of $15+24 \cdot k$ hours, for positive integer k values. Taking into consideration, that data from previous day is not fully available at the time of forecasting, the first possible lag value of 39-hours was selected for further analysis. For lagged temperature and wind speed, the highest correlation value was observed with data from Mount Dillon, whereas for lagged relative humidity, Gurteen's data exhibited the highest correlation.

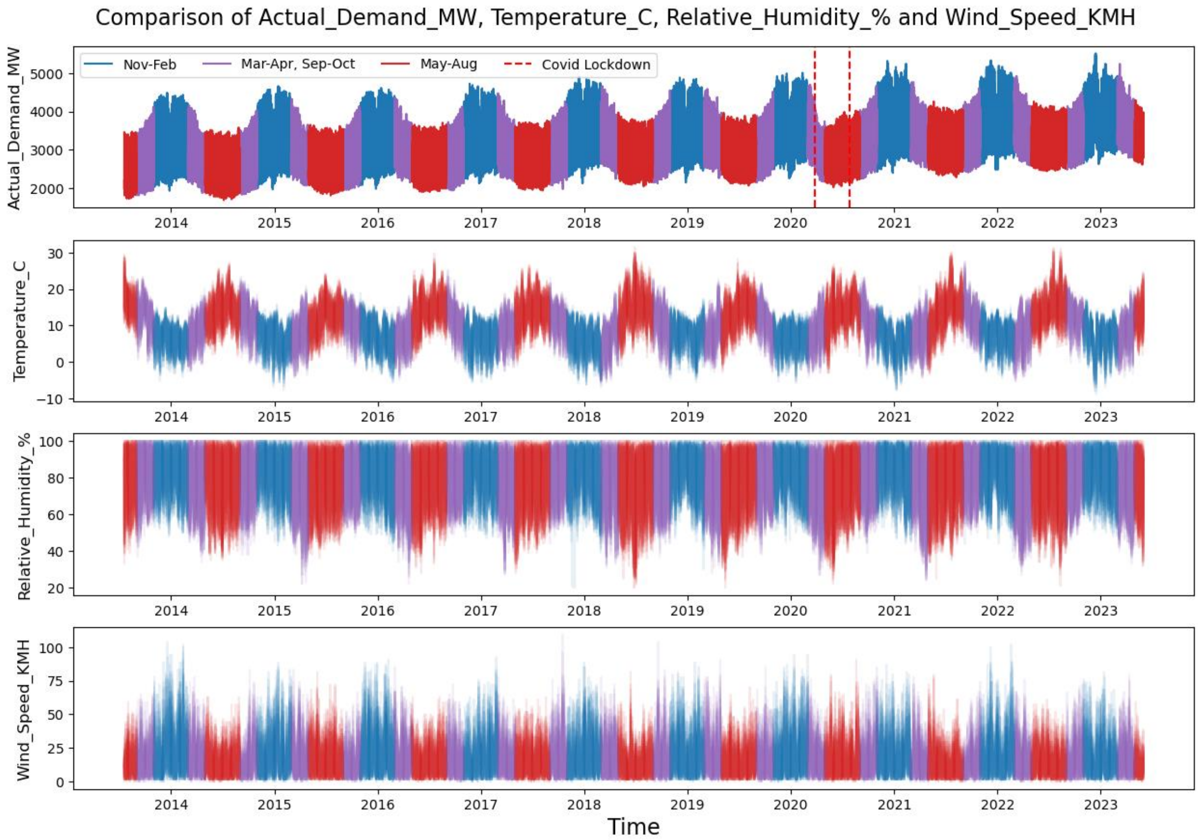


Figure 6.1. Comparison of power demand, temperature, relative humidity and wind speed time series in Ireland 2013-2023

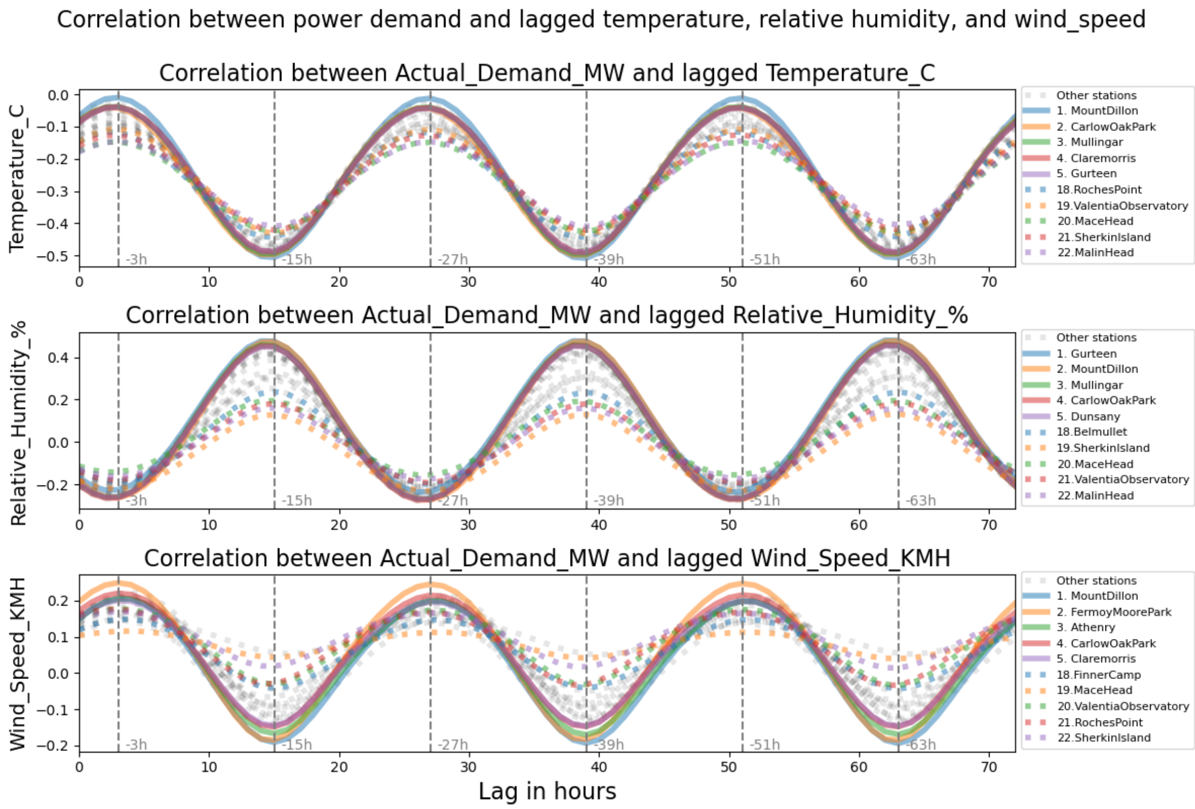


Figure 6.2. Correlation between power demand and lagged weather factors, such as temperature, relative humidity and wind speed time series in Ireland from 2013 to 2023

6.4. Representative Weather Stations

Given the infeasibility of incorporating data from twenty-two weather stations into the research, a study was undertaken to identify representative stations for each weather factor. Three approaches were proposed, such as virtual stations created by Linear and Lasso Regression, as well as the most important single weather station revealed by Lasso Regression and correlation study.

6.4.1. Virtual Weather Station Created by Linear Regression of all Real Stations

Linear Regression was used to create the first virtual weather station, where individual weather factors was the sole explanatory variable for power demand. Absolute values of coefficients (Figure 6.3) were normalised using **MinMax** scaler, and then used to predict respective factors (Figure 6.4).

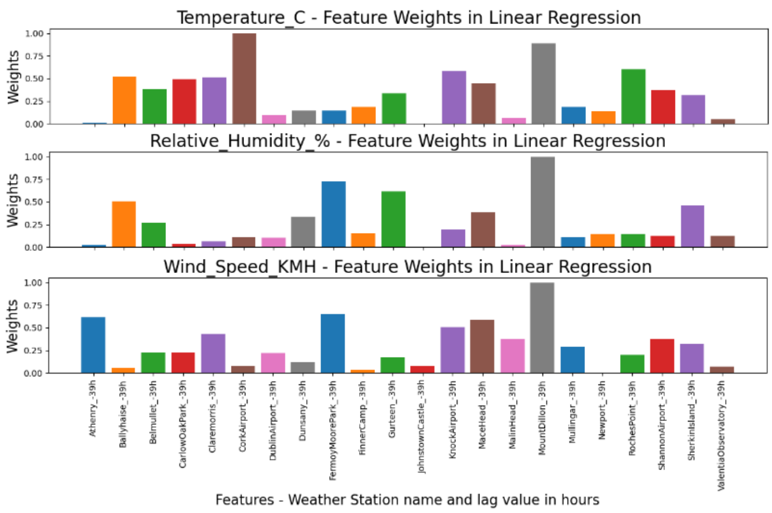


Figure 6.3. Normalised coefficients of linear regression used to create virtual weather station

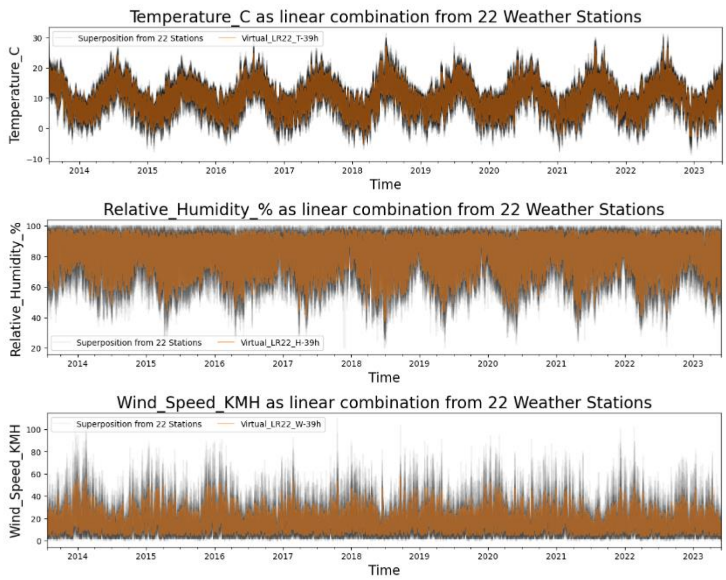


Figure 6.4. Temperature, relative humidity and wind speed from virtual weather station created with linear regression of respective factors from all twenty-two real weather stations in Ireland.

6.4.2. Virtual Weather Station Created by Lasso Regression of Selected Real Stations

Lasso Regression was used to create the second virtual weather station. Value of hyperparameter alpha was increased until the number of non-zero coefficients reached specified values (Figure 6.5). Absolute values of coefficients were normalised and used for prediction of factors.

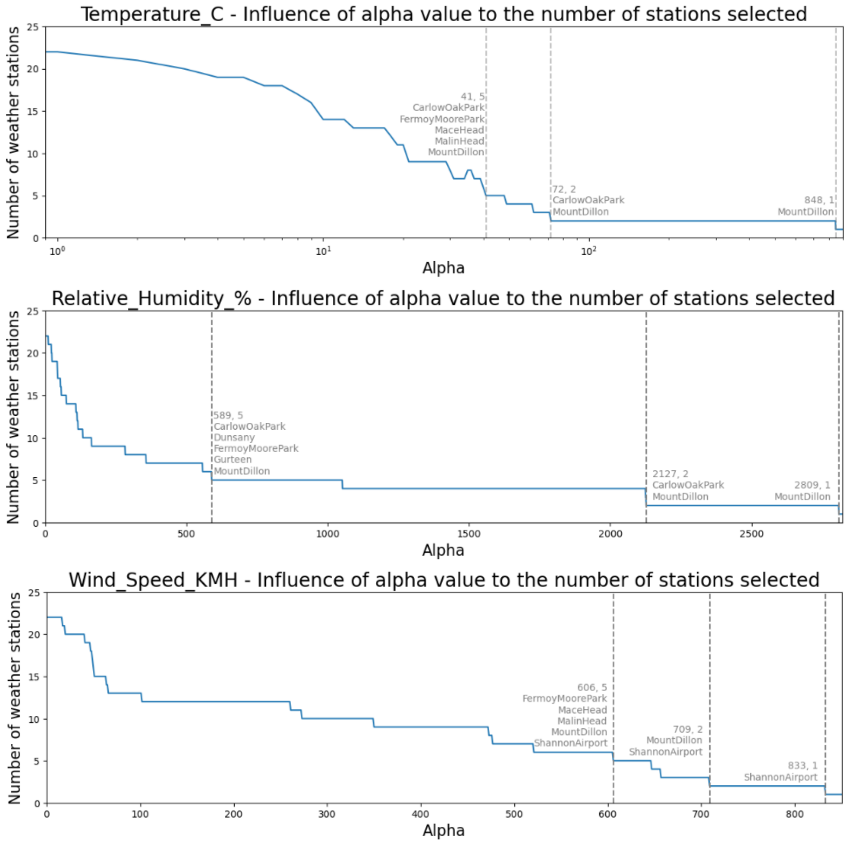


Figure 6.5. Influence of hyperparameter alpha on the number of non-zero coefficients of lasso regression for temperature, relative humidity and wind speed.

6.4.3. Single Real Weather Station and Validation of Results

Validation of the results was performed by correlation study between actual demand and lagged weather factors for each solution (Figure 6.6), comparison of individual distributions (Figure 6.7), and correlation plots (Figure 6.8). Taking into consideration results from Section 6.4, Mount Dillon was selected as single real weather station, which could represent weather factors for the whole Republic of Ireland. Table 6.2 summarises weather factors from all representative weather stations.

Weather Stations	Temperature ºC	Relative Humidity %	Wind Speed KMH
Virtual Linear Regression	Virtual_LR_22_T	Virtual_LR_22_H	Virtual_LR_22_W
Virtual Lasso Regression	Virtual_Lasso2_T	Virtual_Lasso5_H	Virtual_Lasso2_W
Single Real Station	MountDillon_T	MountDillon_H	MountDillon_W

Table 6.2. Summary of representative weather stations in Ireland for power demand forecasting

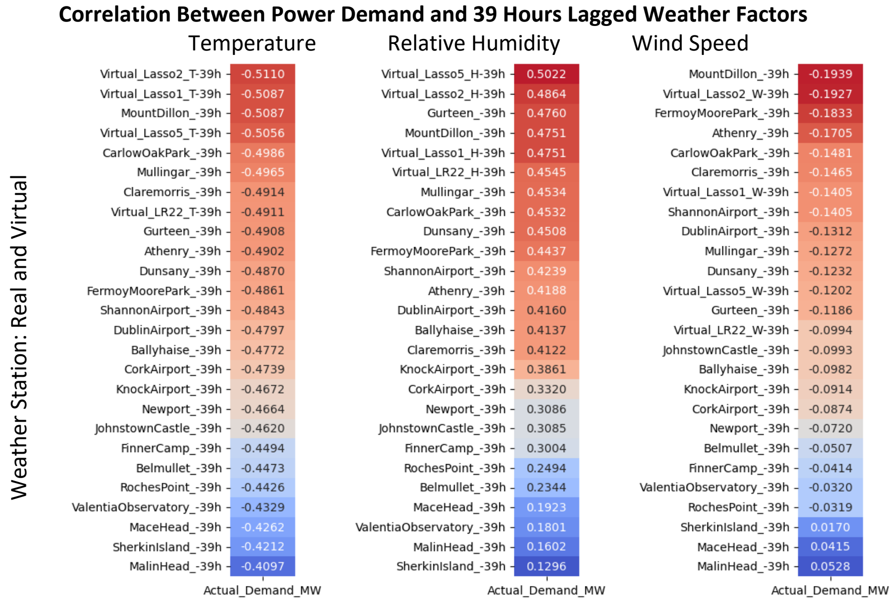


Figure 6.6. Correlation between power demand and 39-hours lagged weather factors.

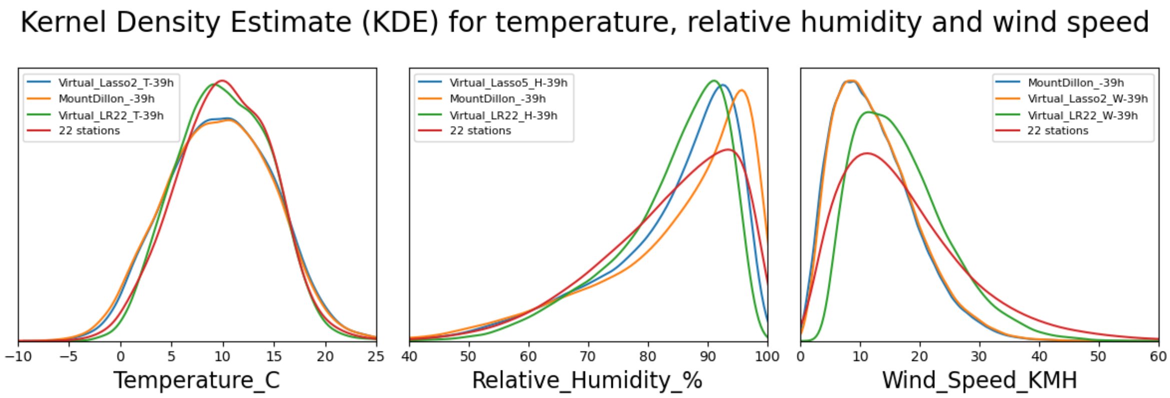


Figure 6.7. Kernel Density Estimate for temperature, relative humidity and wind speed from virtual weather stations created using linear and lasso regression, single real weather station and all twenty-two weather stations.

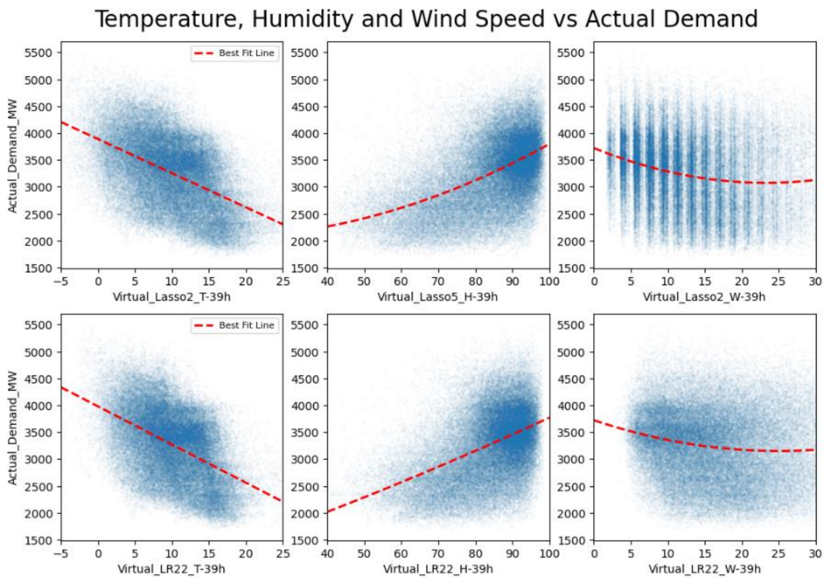


Figure 6.8. Correlations between power demand and temperature, relative humidity and wind speed from virtual weather stations created using linear and lasso regression.

Figure 6.9 depicted the time series of temperature, relative humidity, and wind speed from the approaches achieving the highest correlation with power demand.

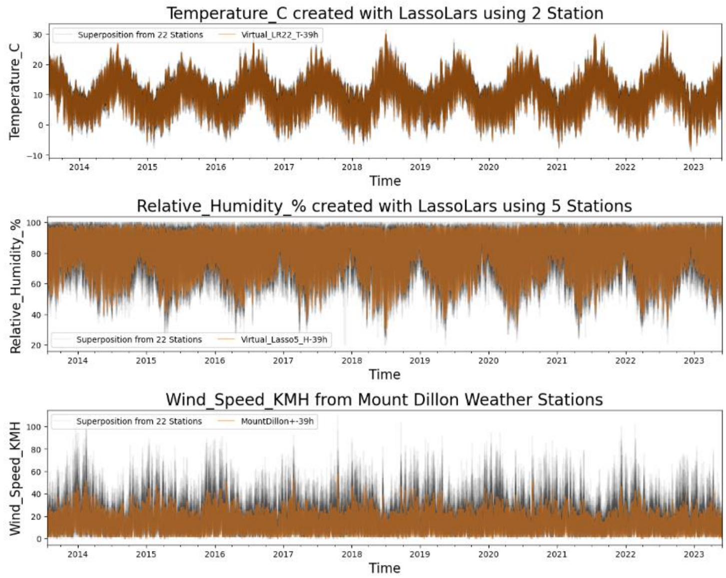


Figure 6.9. Temperature and relative humidity from virtual weather station created with lasso regression of two and five real weather stations, respectively. Wind speed from Mount Dillon real weather station in Ireland.

6.5. Feature Selection for Potential Base-learners

6.5.1. Feature Selection for Similar Day Approach

Weekly lags with correlation above 90%, temporal and weather's features were considered to create supervised learning problem for SD approach. Investigation into feature importance was performed to select the best ones, utilising methods such as Forward (Figure 6.10) and Backward (Figure 6.11) Selection Regression, Lasso Features Selection Regression (Figure 6.12).

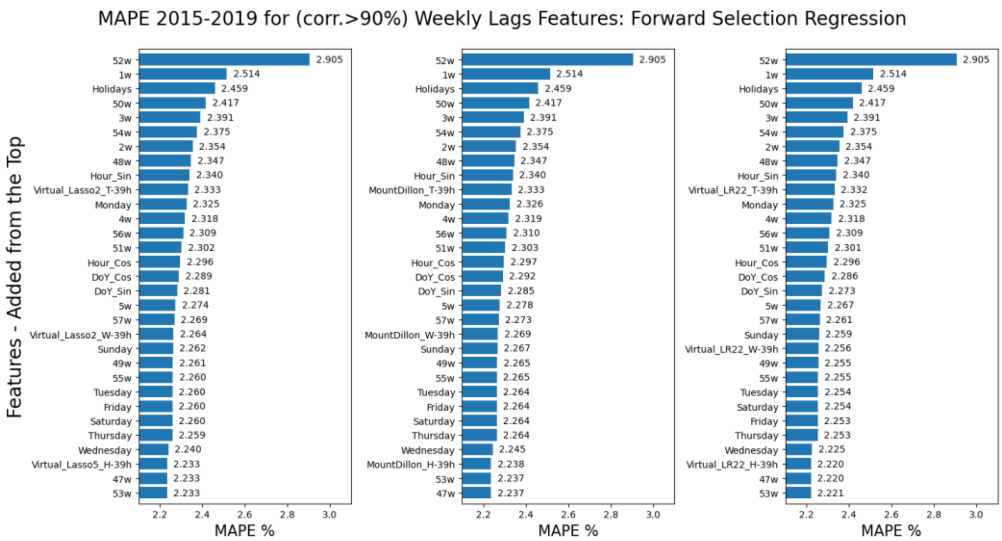


Figure 6.10. Features selection for similar days approach using forward selection regression.

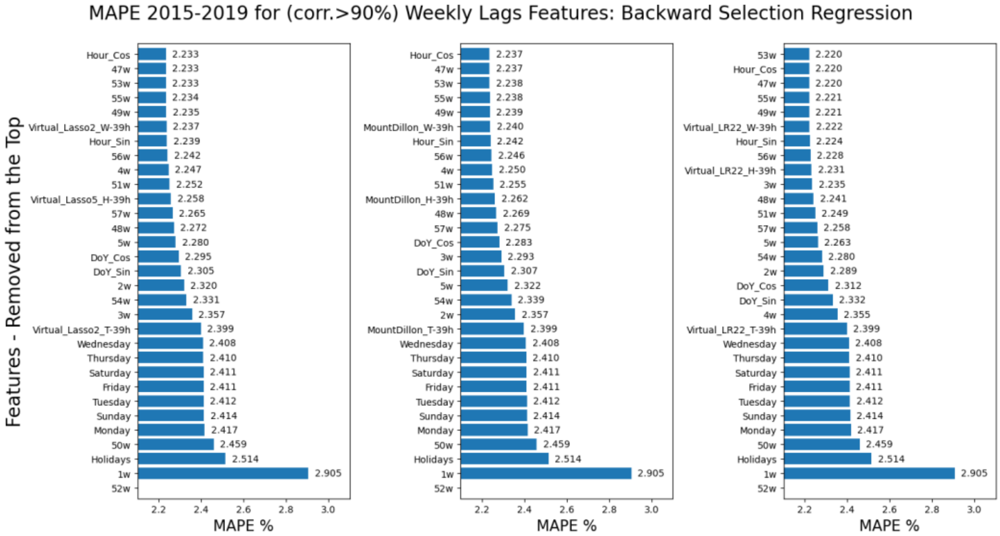


Figure 6.11. Features selection for similar days approach using backward selection (elimination) regression.

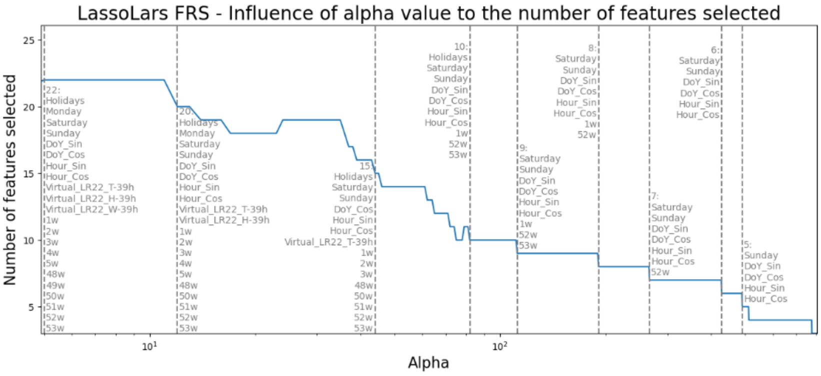


Figure 6.12. Features selection for similar days approach using lasso selection (elimination) regression.

Given that the inclusion of weather factors from virtual weather station created using Linear Regression yielded the best results, they were further investigated for SD approach using Random Forest and Permutation Feature Importance (Figure 6.13).

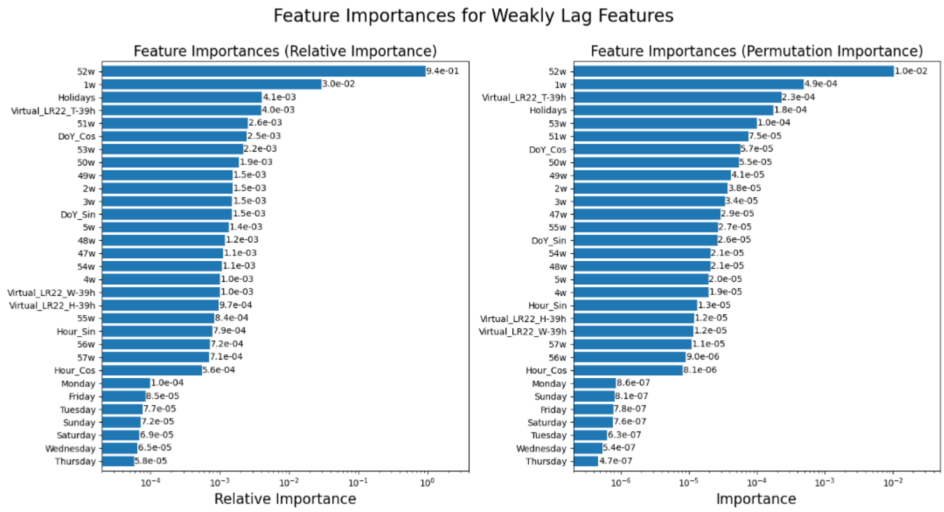


Figure 6.13. Random Forest Feature Importance and Permutation Feature Importance for similar days approach.

Taking above results into consideration, the final set of twenty-nine features for SD approach was selected to include: historical power demand 1-5, 48-52, and 54-57 weeks ago, holidays, days of the week, day of the year, hour, as well as temperature, relative humidity and wind speed, lagged by 39-hours, from virtual weather station created with Linear Regression.

6.5.2. Feature Selection for Moving Window Approach

Investigation into feature importance for window sizes of seven, fourteen and thirty-five days (the new minimum MAPE) was performed (Figure 6.14), for three representative weather stations, to select features achieving the optimal value of MAPE for MW approach (Figure 6.15).

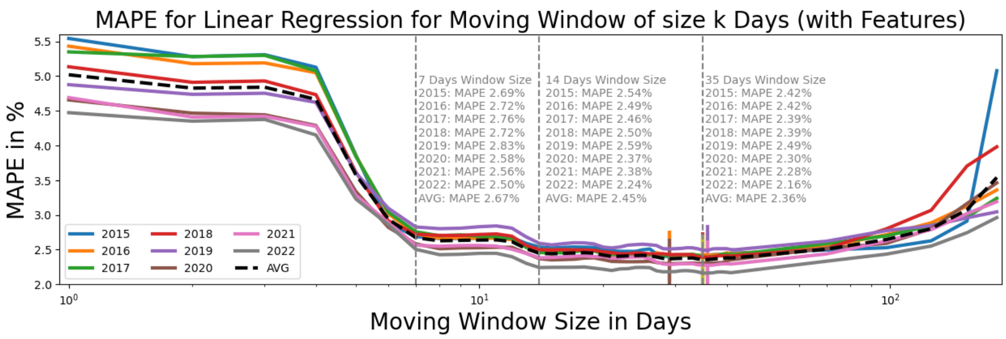


Figure 6.14. Influence of window size on MAPE of forecasting of power demand in Ireland, using linear regression.

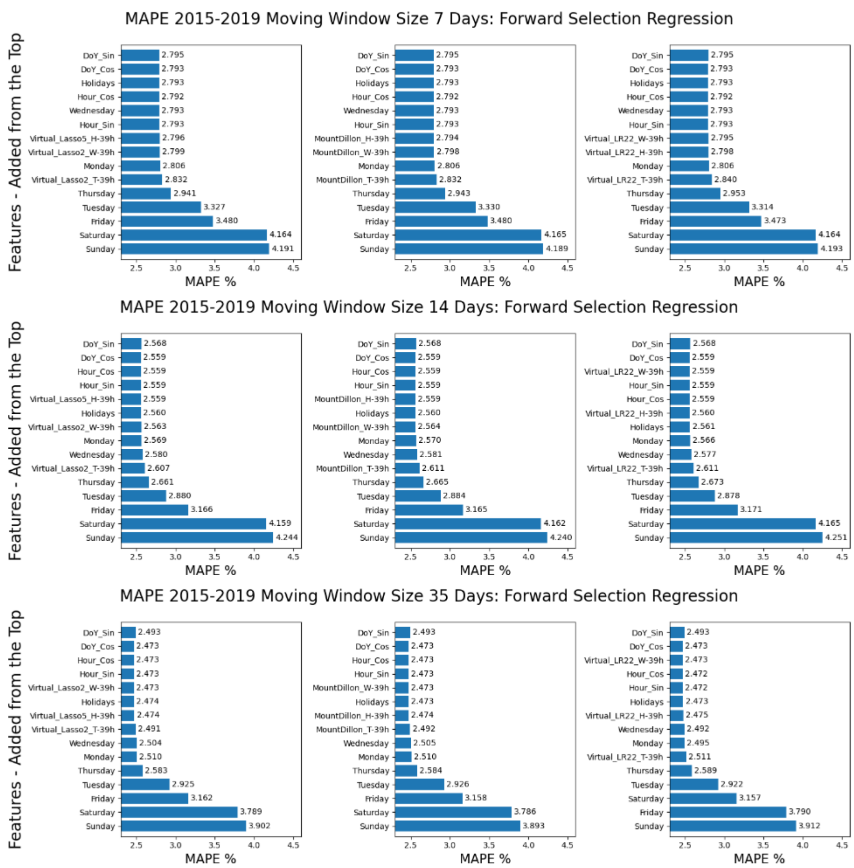


Figure 6.15. Features selection for moving window approach using forward selection regression.

Taking above results into consideration, the final set of features for MW approach was selected to include: moving window of historical power demand with size of thirty-five days, holidays, day of the year, hour, as well as temperature, relative humidity and wind speed, lagged by 39-hours, from virtual weather station created with Lasso Regression.

6.6. Redefining Supervised Learning Problems by Including Weather Factors

Supervised learning problems, defined in section 5.7, were amended to include features selected in section 6.5 for SD (Table 6.3), and MW approaches (Table 6.4).

Record Time Index In Hours	**Independent Variable X**			**Dependent Variable Y**
	Temporal Variables	**Weather Variables**		
	Holidays, Day of Week, Day of Year, Hour of Day **H, DoW, DoY, HoD**	**Temperature, Relative Humidity, Wind Speed** **T, RH, WS**	**Lagged Power Demand Values**	**Power Demand Y**
$t_K=t_0+k\cdot h$	$[H(t_K)$, $DoW(t_K)$, $DoY(t_K)$, $HoD(t_K)]$	$[T(t_K-39\cdot h)$, $RH(t_K-39\cdot h)$, $WS(t_K-39\cdot h)]$ Virtual weather station created using Linear Regression	$[Y(t_K-n_1\cdot 7\cdot 24\cdot h)$, $Y(t_K-n_2\cdot 7\cdot 24\cdot h)$, …, $Y(t_K-n_L\cdot 7\cdot 24\cdot h)]$ k=1, 2, … (record index) L: number of weekly lags t_K: timestamp k in hours n_1, n_2, … n_L: 1…5, 48…52, 54…57 (weeks)	$Y(t_K)$

Table 6.3. Supervised learning problem definition for similar days approach, including weather factors

Record Time Index In Hours	**Independent Variable X**			**Dependent Variable Y**
	Temporal Variables	**Weather Variables**		
	Holidays, Day of Year, Hour of Day **H, DoY, HoD**	**Temperature, Relative Humidity, Wind Speed** **T, RH, WS**	**Moving Window of Power Demand Values**	**Power Demand Y**
$t_K=t_0+k\cdot h$	$[H(t_K)$, $DoY(t_K)$, $HoD(t_K)]$	$[T(t_K-39\cdot h)$, $RH(t_K-39\cdot h)$, $WS(t_K-39\cdot h)]$ Virtual weather station created using Lasso Regression	$[Y(t_K-(24\cdot(1+w_S)-1)\cdot h)$, $Y(t_K-(24\cdot(1+w_S)-2)\cdot h)$, …, $Y(t_K-(24\cdot(1+w_S)-w_S)\cdot h)]$ where: k=1, 2, … (record index) w_S=35 (window size in days) current day not considered	$Y(t_K)$

Table 6.4. Supervised learning problem definition for moving window approach, including weather factors.

6.7. Conclusion

In this chapter, the final sets of features were separately selected for SD and MW approaches, and the supervised learning problems were amended, accordingly. DST distortion was removed from weather factors time series, and correlation between lagged factors and power demand was evaluated. Representative weather stations for each weather factors were found to be valid by correlation study. For SD approach, the LR-based virtual station excelled, while Lasso Regression was optimal for the MW method.

7. Experimentation with Potential Base-Learners of Ensemble Learning Models

7.1. Architectures and Hyperparameters Tuning of Base-Learners

Architectures of considered Deep Learning (DL) models were shown in Table 7.1. Hyperparameters of all potential base-learners with their values ranges, and their best values found by Bayesian optimisation with 10-fold cross-validation were shown in Table 7.2.

Group of layers	Model architecture				Inclusion
	LSTM Model		CNN Model		
	Description	Remarks	Description	Remarks	
1	LSTM	return_seq=False for last layer	Convolutional1D BatchNormalisation MaxPooling1D Dropout	filters x1 - kernel_size cnn_dropout	Obligatory
2	LSTM		Convolutional1D BatchNormalisation MaxPooling1D Dropout	filters x2/4, respectively - kernel_size cnn_dropout	Optional
3	LSTM				
4	Dense BatchNormalisation Dropout Dense	dense_units - dense_dropout 1 unit	Flatten Dense BatchNormalisation Dropout Dense	- dense_units - dense_dropout 1 unit	Obligatory

Table 7.1. Deep learning potential base-learners' architectures.

Regressor	Data Shape	Hyperparameters				
		Name	Range of Values	Remarks	Best Values	
					Similar Day	Moving Window
Linear	2D	fit_intercept	False, True	-	False	False
Ridge		fit_intercept alpha	False, True 10^{-9}…10^{3}	- logarithmic scale	False $2.17 \cdot 10^{-4}$	True $4.82 \cdot 10^{-4}$
Lasso					True $1.38 \cdot 10^{-4}$	True $2.58 \cdot 10^{-4}$
SVM		kernel C gamma epsilon	linear, rbf 10^{-6}…10^{3} 10^{-9}…10^{1} 10^{-9}…10^{1}	- logarithmic scale logarithmic scale logarithmic scale	linear $2.23 \cdot 10^{-1}$ $9.10 \cdot 10^{-4}$ $1.99 \cdot 10^{-4}$	linear 1.51 $2.41 \cdot 10^{-3}$ $3.30 \cdot 10^{-3}$
GBM		n_estimators max_depth learning_rate	10…250 10…250 10^{-9}…10^{-1}	- - logarithmic scale	90 100 $3.64 \cdot 10^{-1}$	210 60 4.69.10^{-2}
MLP		hidden_layer_sizes learning_rate batch_size max_iter	(d_1,)…(d_1, d_2) constant, invscaling, adaptive 1, 24, 72, 168 25, 50, 100	d_1, d_2 = 1…1024 - - -	(238, 47) constant 24 100	(249, 18) invscaling 24 25
LSTM	3D	no_lstm_layers lstm_units dense_units dense_dropout batch_size learning_rate epochs	1, 2, 3 1…1024 1…1024 0…0.5 1, 24, 72, 168 10^{-9}…10^{-1} 25, 50, 100	- - - - - logarithmic scale -	1 43 58 $3.68 \cdot 10^{-2}$ 24 $2.91 \cdot 10^{-1}$ 25	not performed for moving window approach
CNN		no_cnn_layers filters kernel_size cnn_dropout dense_units dense_dropout batch_size learning_rate epochs	1, 2, 3 1…256 2…7 0…0.5 1…1024 0…0.5 1, 24, 72, 168 10^{-9}…10^{-1} 25, 50, 100	- 1st layer - - - - - logarithmic scale -	1 20 3 $9.98 \cdot 10^{-2}$ 236 $7.60 \cdot 10^{-4}$ 24 $1.86 \cdot 10^{-3}$ 50	

Table 7.2. Base-learners hyperparameters and their values selection, and their best values found by Bayesian optimisation.

7.2. Evaluation of Similar Day Approach Based Base-Learners

7.2.1. Similar Day Approach Baseline Base-Learners

Linear Regression was used for SD - one week, one year (52 weeks) and range 1-52 weeks ago - to get adjusted ODADF in Ireland in year 2019. Their evaluation was shown on Figure 7.1.

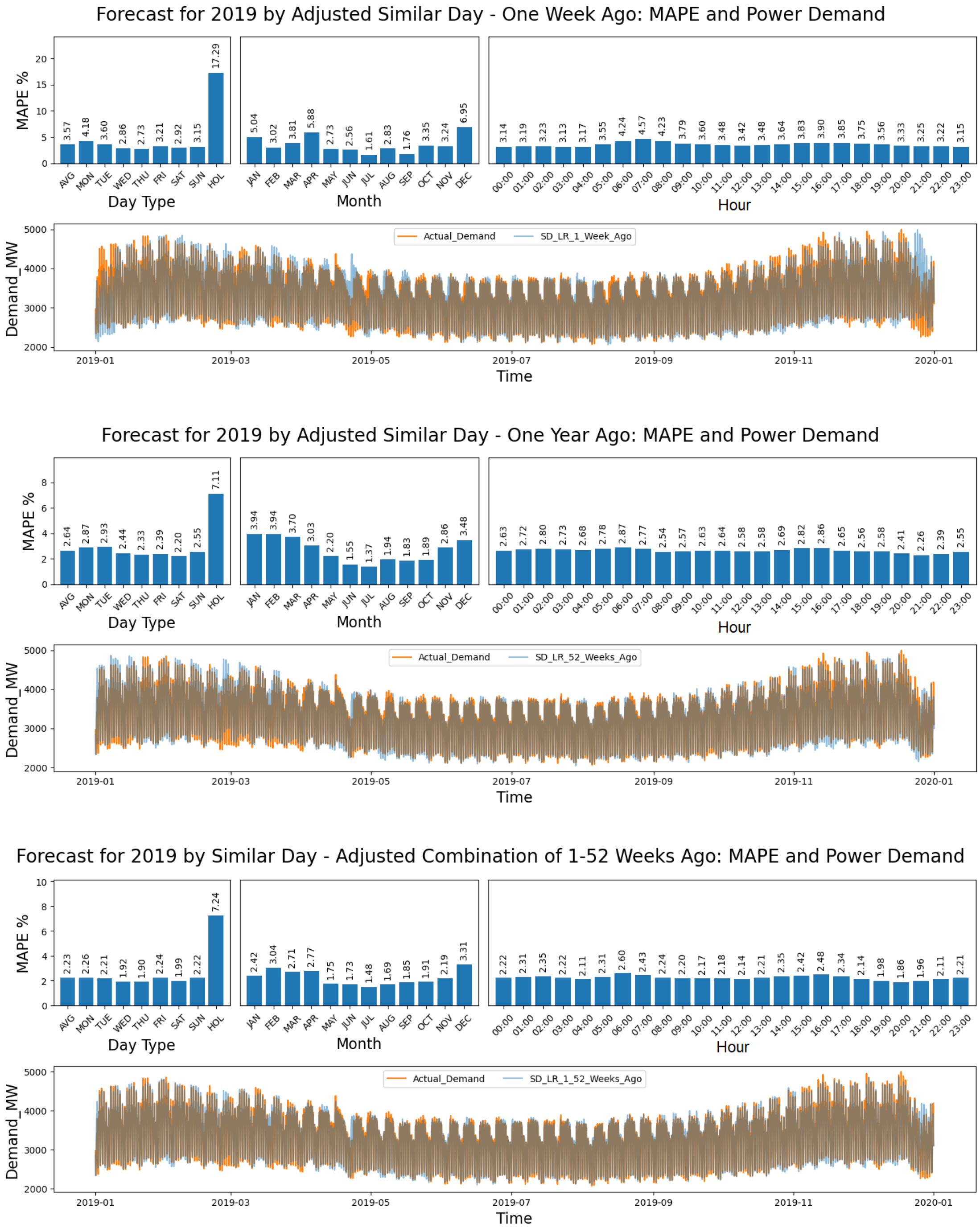


Figure 7.1. Forecast of power demand in Ireland for year 2019 by adjusted similar days one week, one year (52 weeks) and range 1-52 weeks ago, using linear regression.

7.2.2. Similar Day with Temporal and Weather Features Approach Base-Learners

Evaluation of machine and deep learning models, described in previous section, for ODADF in Ireland in year 2019, was shown on Figures 7.2-7.3.

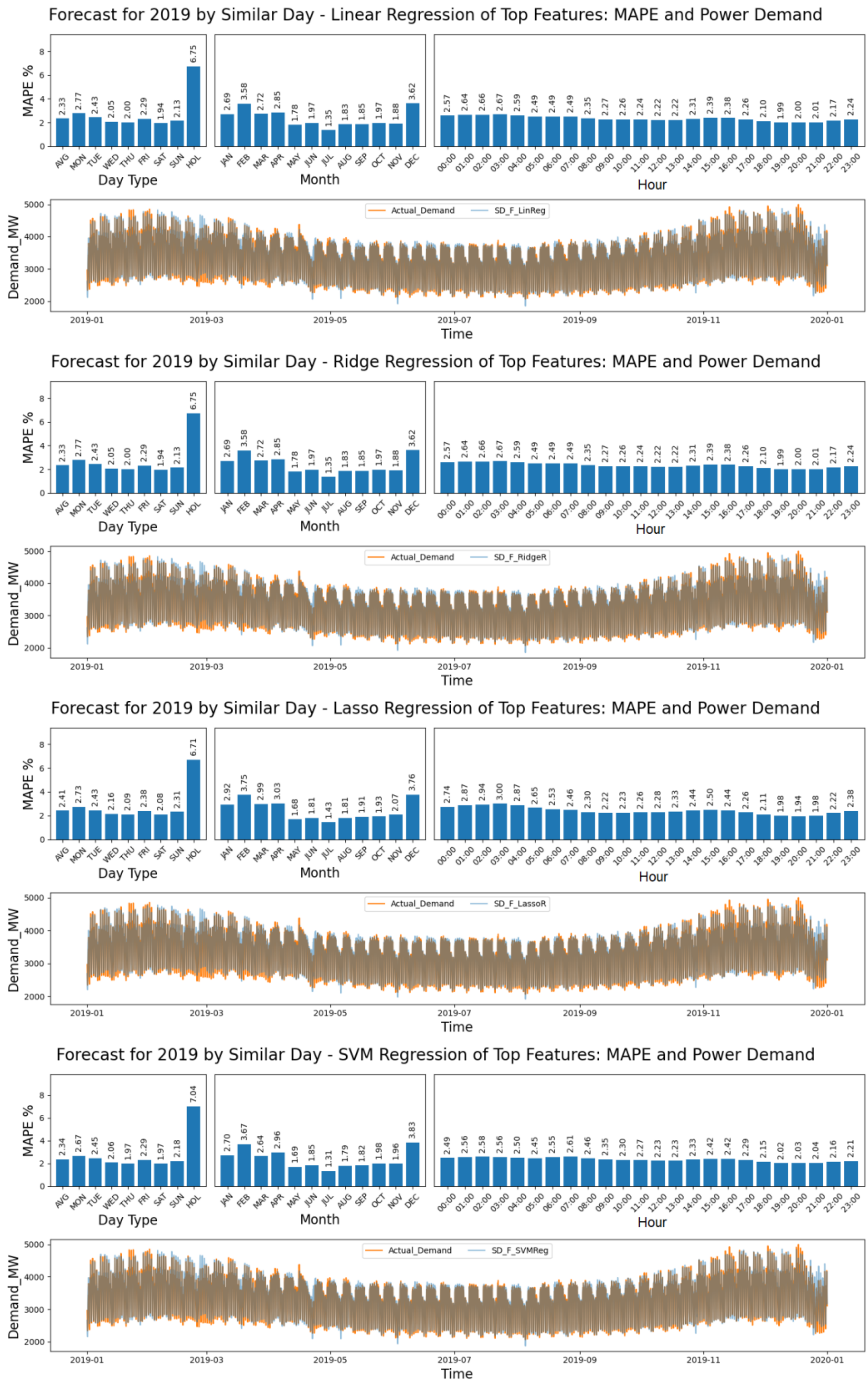


Figure 7.2. Forecast of power demand in Ireland for year 2019 by linear, ridge, lasso and SVM regression.

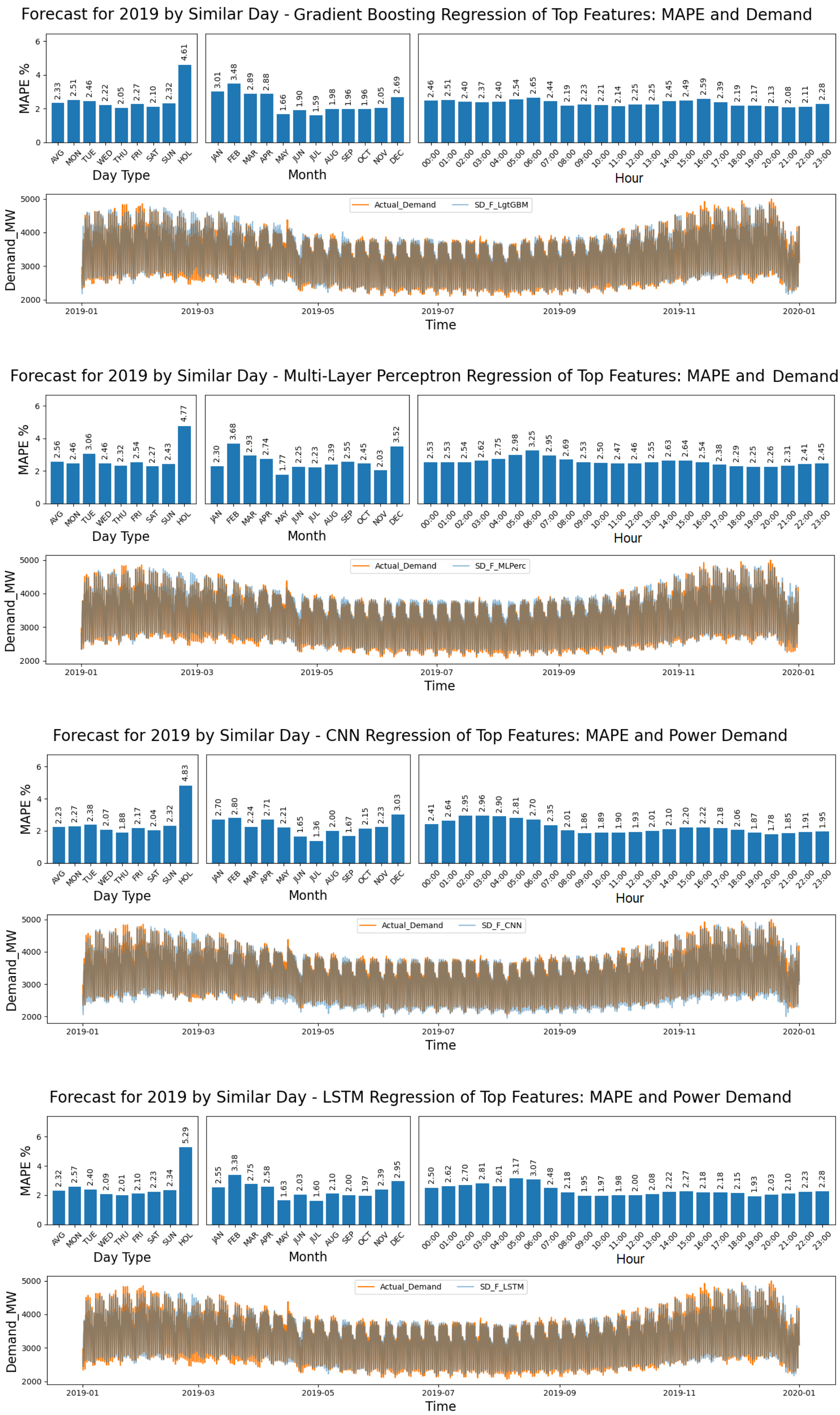


Figure 7.3. Forecast of power demand in Ireland for year 2019 by GBM, MLP, CNN and LSTM regression.

7.3. Evaluation of Moving Window Approach Based Base-Learners

7.3.1. Moving Window Approach Baseline Base-Learners

Evaluation of machine learning models for three sizes of univariate MW, described in Section 6.5.2, for ODADF in Ireland in year 2019, was shown on Figure 7.4.

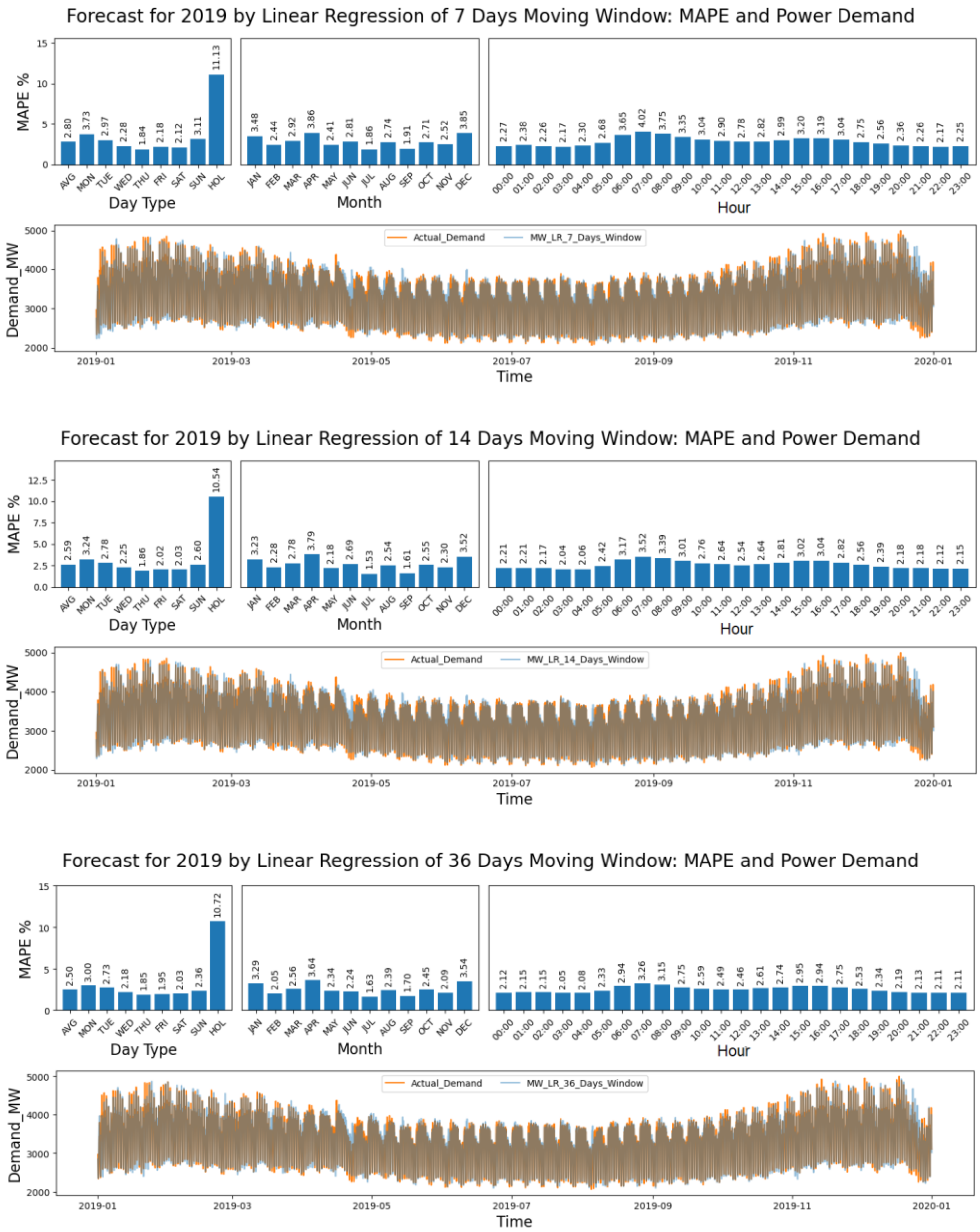


Figure 7.4. Forecast of power demand for year 2019 by linear regression for moving window size of 7, 14 and 35 days.

7.3.2. Moving Window with Temporal and Weather Features Approach Base-Learners

Evaluation of machine learning models for MW approach with temporal and weather features, described in previous section, for ODADF in Ireland in year 2019, was shown on Figures 7.5-7.6.

Figure 7.5. Forecast of power demand in Ireland for year 2019 by linear, ridge, lasso and SVM regression.

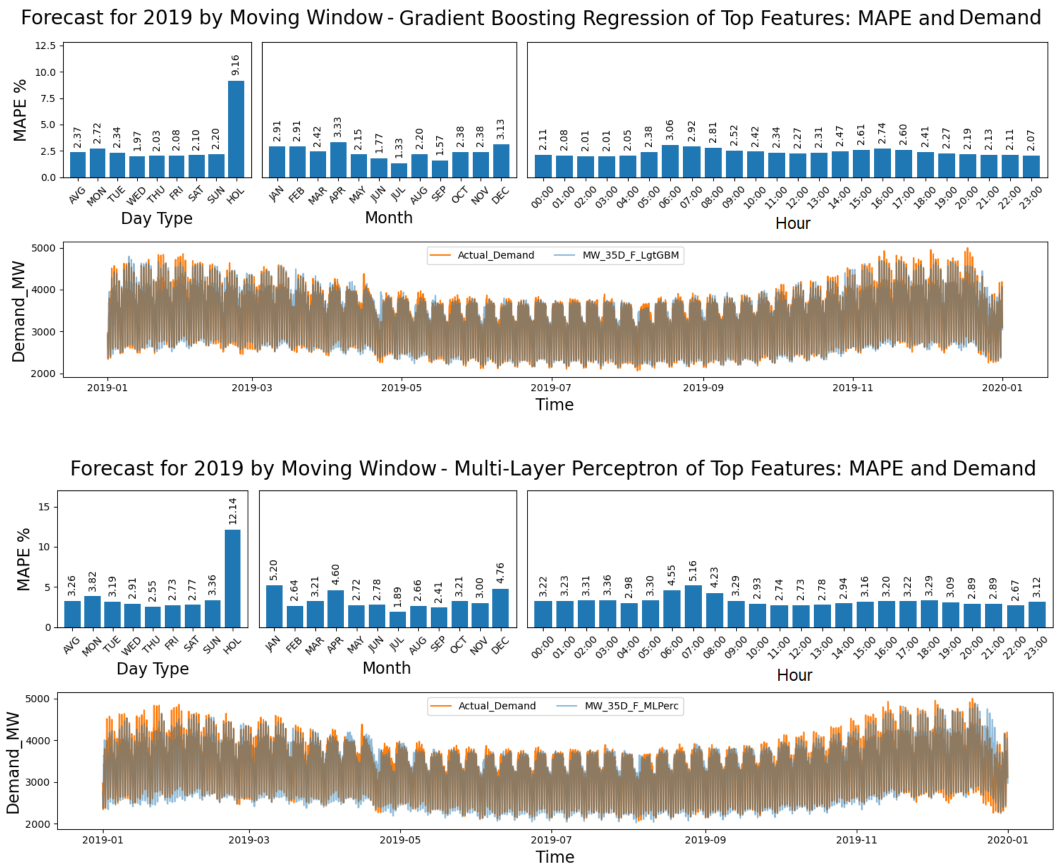


Figure 7.6. Forecast of power demand in Ireland for year 2019 by GBM and MLP regression.

7.4. Final Selection of Base-Learners

Following analysis of potential base-learners' results from sections 7.1-7.3, twenty out of twenty-seven models (Figure 7.7, Table 7.3), achieving the lowest MAPE for ODADF in year 2019, were selected as base-learners for further experimentation with ensembles. Ranking of the twenty base-learners as function of day, month and hour, based on MAPE for year 2019 was shown on Figure 7.8.

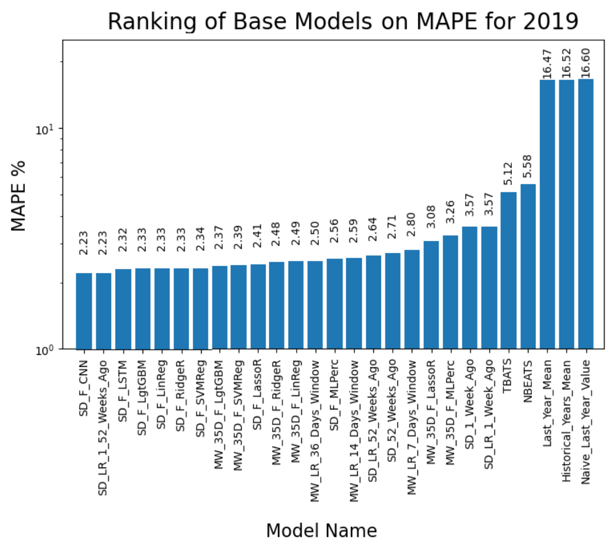


Figure 7.7. Ranking of base-learners on MAPE for 2019.

Model		Performance metrics for year 2019		
Name	Description (WF=weather features)	MAE in MW	RMSE in MW	MAPE in %
SD_F_CNN	SD-based CNN with WF	72.8	97.3	2.226
SD_LR_1_52_Weeks_Ago	SD-based LR of 1-52 weekly lagged demand	73.8	103.9	2.231
SD_F_LSTM	SD-based LSTM with WF	75.6	100.4	2.320
SD_F_LgtGBM	SD-based GBM with WF	77.0	104.0	2.330
SD_F_LinReg	SD-based LR with WF	76.4	105.6	2.334
SD_F_RidgeR	SD-based Ridge Regression with WF	76.4	105.6	2.334
SD_F_SVMReg	SD-based SVM with WF	76.9	107.6	2.342
MW_35D_F_LgtGBM	MW-based GMB with WF	79.4	114.2	2.370
MW_35D_F_SVMReg	MW-based SVM with WF	79.3	121.8	2.387
SD_F_LassoR	SD-based Lasso Regression with WF	78.6	109.1	2.414
MW_35D_F_RidgeR	MW-based Ridge Regression with WF	82.7	120.8	2.484
MW_35D_F_LinReg	MW-based LR with WF	83.0	122.0	2.491
MW_LR_36_Days_Window	MW-based LR, window size 36 days	83.3	122.2	2.497
SD_F_MLPerc	SD-based MLP with WF	83.6	109.1	2.563
MW_LR_14_Days_Window	MW-based LR, window size 14 days	86.3	127.1	2.585
SD_LR_52_Weeks_Ago	SD-based LR of 52 weeks ago	88.2	126.7	2.637
SD_52_Weeks_Ago	SD 52 weeks ago	90.7	128.0	2.706
MW_LR_7_Days_Window	MW-based LR, window size 7 days	93.1	139.0	2.798
MW_35D_F_LassoR	MW-based Lasso Regression with WF	102.8	153.0	3.079
MW_35D_F_MLPerc	MW-based MLP with WF	108.1	153.3	3.262
SD_1_Week_Ago	SD 1 week ago	118.5	190.0	3.567
SD_LR_1_Week_Ago	SD-based LR of 1 week ago	118.7	190.0	3.571
TBATS	Trigonometric seasonal decomposition of time series	172.8	227.0	5.119
NBEATS	Neural basis expansion analysis for time series	188.5	246.9	5.576
Last_Year_Mean	Mean value of previous year	514.8	603.6	16.468
Historical_Years_Mean	Mean value of previous years	547.8	639.6	16.522
Naïve_Last_Year_Value	Last value seen in previous year	553.9	647.2	16.601

Table 7.3. Performance metrics for potential base-learners forecasting power demand in Ireland for year 2019. Twenty base-learners, selected for further experimentation with meta-learners of ensembles, were marked in grey.

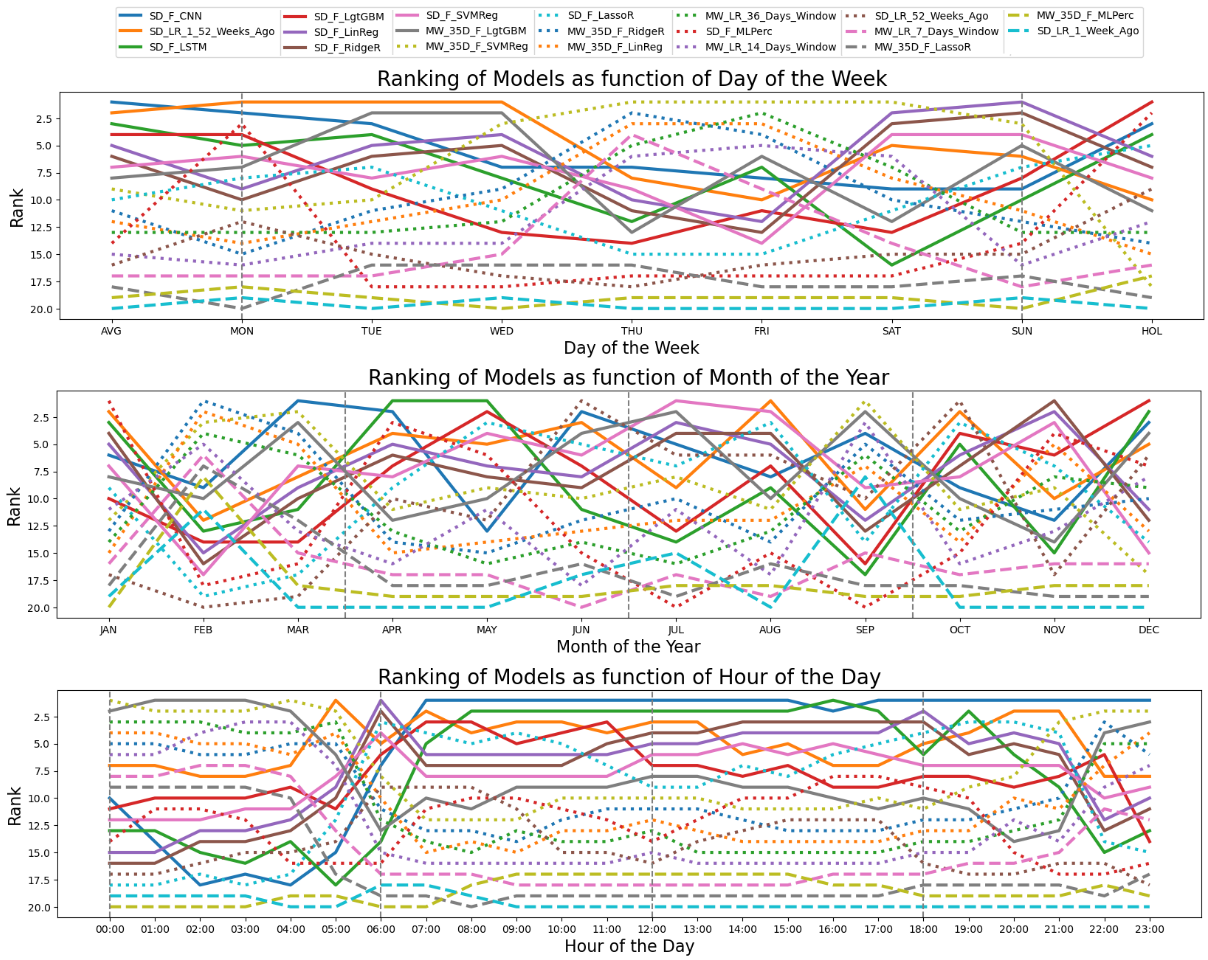


Figure 7.8. Ranking of 20 base-learners as function of day, month and hour, based on MAPE for year 2019.

Comparison of distributions, forecasted by base-learners and actual power demands, for year 2019 was shown on Figure 7.9. As power demand was not normally distributed, non-parametric Kolmogorov-Smirnov test was used to investigate whether forecasted and actual demand, in pairs, were from the same distribution, with the null hypothesis that two samples were from the same distribution. Results were shown in Table 7.4.

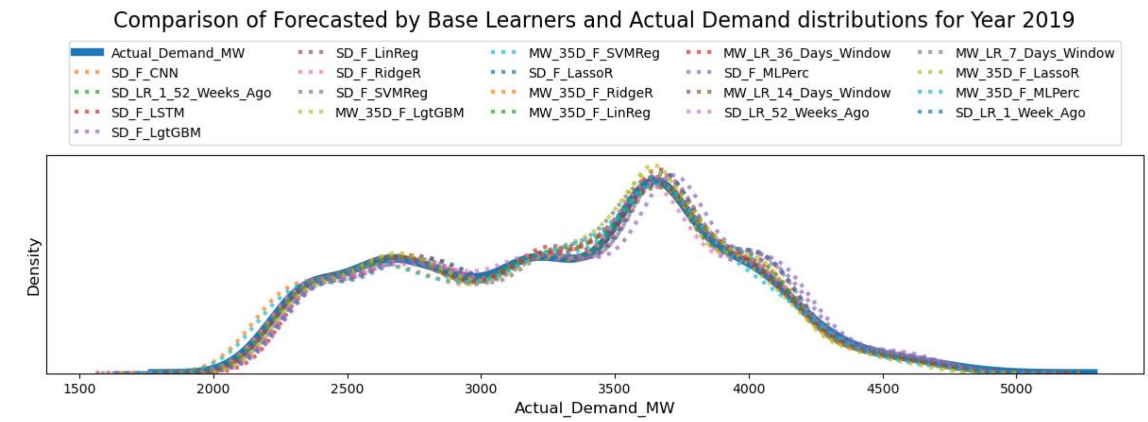


Figure 7.9. Comparison of forecasted by base-learners and actual power demand distributions for Year 2019.

Samples of predictions of below base-learners and actual demand in 2019	
were from the same distributions	**were from different distributions**
SD_LR_1_52_Weeks_Ago	SD_F_CNN
MW_35D_F_LgtGBM	SD_F_LSTM
MW_35D_F_SVMReg	SD_F_LgtGBM
MW_35D_F_RidgeR	SD_F_LinReg
MW_35D_F_LinReg	SD_F_RidgeR
MW_LR_36_Days_Window	SD_F_SVMReg
MW_LR_14_Days_Window	SD_F_LassoR
SD_LR_52_Weeks_Ago	SD_F_MLPerc
MW_LR_7_Days_Window	MW_35D_F_LassoR
SD_LR_1_Week_Ago	MW_35D_F_MLPerc

Table 7.4. Results of Kolmogorov-Smirnov test for predictions of base-learners and actual power demand in 2019.

7.5. Percentage Share of Base-Learners in the Best Hourly Predictions for Years 2015-2019

The percentage share of base-learners in the best hourly predictions was shown by year (Figure 7.10), and by day type (Figure 7.11).

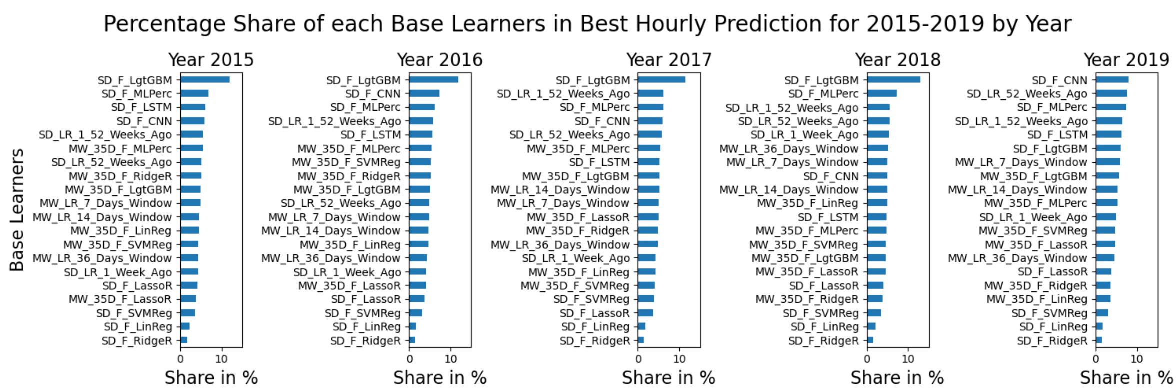


Figure 7.10. Percentage share of base-learners in the best hourly prediction for years 2015-2019 by year.

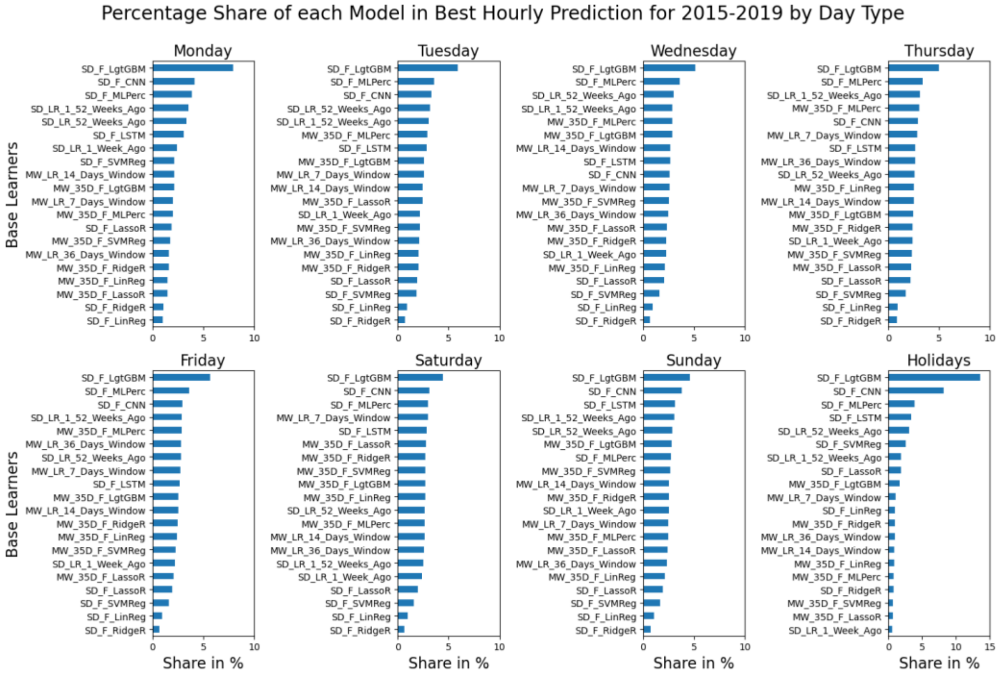


Figure 7.11. Percentage share of base-learners in the best hourly prediction for years 2015-2019 by day type.

7.6. Discussion Regarding Results of the Experimentation with Base-Learners

Experimentation with potential base-learners revealed significant variations in the performance of models. For example, MAE, RMSE and MAPE varied 72.8-118.8MW, 97.3-190.0MW, and 2.226-3.571%, respectively. While base-learners trained on SD dataset performed better on average metrics than those trained on MW dataset, all models showed fluctuations in their MAPE across different days of the week, months and hours. Therefore, there was no single best or worst model in all categories.

The lowest MAPE was observed on Thursdays and Saturdays, with GBM (MW) model performing the best. However, almost all models underperformed on Mondays and holidays occurring on weekdays. While similar day CNN model achieved the lowers average MAPE, other models, such as Linear Regression of 1-52 weeks ago, SVM (MW), Linear Regression (SD), GBM (SD), shared the first place for Monday to Wednesday, Thursday to Saturday, Sunday, and holidays occurring on weekdays, respectively. Base-learners predicted specific days of the week better than others, based on the approach of dataset creation. MW-based models generally performed better on Thursdays, Fridays, and Sundays. Conversely, SD-based models had superior performance on the other days.

Examining monthly performance revealed that the lowest MAPE was observed in July, with SVM (SD) emerging as the top performer. On the other hand, December, April, and January yielded the least favourable results. The best-performing models varied from month to month, and no single model consistently underperformed across all months. Notably, LSTM (SD) was the only model to excel in two consecutive months: April and May. Furthermore, while MW-based models generally achieved better results for February, March, September and November, the SD-based models had superior performance in the other months.

Upon analysing hourly performance, it was evident that the MAPE was at its lowest at 9 a.m. and 8 p.m., with CNN (SD) standing out as the best performer during these hours. Conversely, the period from 5 a.m. to 9 a.m. consistently showed less-promising results, with 7 a.m. identified as the most challenging hour. The model rankings experienced notable shifts between 4 a.m. to 7 a.m. and again from 8 p.m. to 11 p.m. Between 10 p.m. and 4 a.m., GBM (MW) took the lead, closely followed by SVM (MW). At 5 a.m., the Linear Regression of 1-52 weeks ago was the top-performing, with SVM (MW) right behind. By 6 a.m., LR (SD) became the model of choice, with Ridge (SD) as the runner-up. Lastly, from 7 a.m. to 9 p.m., CNN (SD) led the pack, but between 7 a.m. to 5 p.m., LSTM (SD) and GBM (SD) closely followed it, with the Lasso Regression (SD) and SVM (MW) also making a significant impact.

Finally, while comparison of predictions by base-learners and actual power demand in 2019 distributions looked similar visually, inferential statistics tests revealed that, majority of moving window base-learners’ predictions, were from the same, and majority of similar day models, were from different than actual demand distribution.

7.7. Conclusion

The diverse performance of various base-learner models across different time frames, be it days of the week, months, or specific hours, underscored the value of leveraging a variety of models when constructing ensemble systems. No single model consistently excelled or faltered in every scenario. This variability reinforced that a diverse set of models could complement each other's strengths and mitigate each other's weaknesses, potentially leading to a more robust and accurate ensemble prediction in ODADF.

8. Experimentation with Integration Methods of Base-Learners into Ensemble Learning Models

8.1. Architectures of Ensembles

Three stacking approaches to obtain ensemble predictions were investigated: heuristic rules, optimal model classification-based predictors, and regression-based base-learners' predictions.

Heuristic rules were baseline of ensemble modelling. Five following techniques were used:

- mean and median of all base-learners' predictions,
- base-learner with lowest MAPE at the same time one day, one week, and one year (52 weeks) ago was selected to make the prediction.

Classification-based ensembles used classifier as meta-learner to predict which base-learner provided the best estimate for the current time instant based on explanatory variables, such as temporal and weather factors. The output of meta-learner was then interpreted as probability that each base-learner yields the best prediction. Classifiers such as logistic regression, SVM, GBM and MLP were considered as classification meta-learners. The product of probability of base-learners and their prediction value was then taken as ensemble prediction of ODADF in Ireland in 2019.

Finally, regression-based ensembles, considering limited and unlimited number of base-learners' predictions, with LR, SVM, GBM and MLP as meta-learners were examined. Base-learners' predictions were used as the independent variable for meta-learners to forecast ODADF in Ireland for 2019.

8.2. Hyperparameters Tuning of Ensembles' Meta-Learners

Architectures and hyperparameters of considered meta-learners of ensembles, with their values ranges, and their best values found by Bayesian optimisation with 10-fold cross-validation were shown in Tables 8.1-8.2.

Classifying Meta-learner	Data Shape	**Hyperparameters**			
		Name	**Range of Values**	**Remarks**	**Best Values**
Logistic Regression Classifier	2D	weather_station	linear, lasso, Mount Dillon	-	linear regression
		fit_intercept	False, True	-	True
		C	$10^{-6}…10^{3}$	logarithmic scale	$2.88 \cdot 10^{-3}$
SVM Classifier		weather_station	linear, lasso, Mount Dillon	-	linear regression
		kernel	linear, rbf	-	rbf
		C	$10^{-6}…10^{3}$	logarithmic scale	$7.52 \cdot 10^{-1}$
		gamma	$10^{-9}…10^{1}$	logarithmic scale	$6.05 \cdot 10^{-2}$
GBM Classifier		weather_station	linear, lasso, Mount Dillon	-	lasso regression
		n_estimators	10…500	-	130
		max_depth	10…500	-	260
		learning_rate	$10^{-9}…10^{-1}$	logarithmic scale	$3.08 \cdot 10^{-4}$
MLP Classifier		weather_station	linear, lasso, Mount Dillon	-	lasso regression
		hidden_layer_sizes	$(d_1,)…(d_1, d_2)$	$d_1, d_2 = 1…1024$	(2, 49)
		learning_rate	constant, invscaling, adaptive	-	invscaling
		batch_size	1, 24, 72, 168	-	24
		max_iter	25, 50, 100	-	25

Table 8.1. Classifying meta-learners' hyperparameters and their values selection, and their best values.

Regressor Meta-learner	Data Shape	Hyperparameters				
		Name	**Range of Values**	**Remarks**	**Best Values**	
					Selected	**All**
Linear	2D	fit_intercept	False, True	-	True [(a)]	True
SVM Regressor		kernel	linear, rbf	-	linear [(b)]	rbf
		C	$10^{-6}...10^{3}$	logarithmic scale	$5.22 \cdot 10^{-1}$	$5.61 \cdot 10^{-1}$
		gamma	$10^{-9}...10^{1}$	logarithmic scale	$7.59 \cdot 10^{-1}$	$3.34 \cdot 10^{-1}$
		epsilon	$10^{-9}...10^{1}$	logarithmic scale	$1.70 \cdot 10^{-3}$	$1.82 \cdot 10^{-3}$
GBM Regression		n_estimators	10...500	-	230 [(c)]	290
		max_depth	10...500	-	440	500
		learning_rate	$10^{-9}...10^{-1}$	logarithmic scale	$4 \cdot 17.10^{-2}$	$3.55 \cdot 10^{-2}$
MLP Regression		hidden_layer_sizes	$(d_{1},)...(d_{1}, d_{2})$	$d_1, d_2 = 1...1024$	(198, 48) [(d)]	(250, 21)
		learning_rate	constant, invscaling, adaptive	-	invscaling	adaptive
		batch_size	1, 24, 72, 168	-	24	24
		max_iter	25, 50, 100	-	25	100

A combination of up to 6 base-learners was selected in the first approach:
(a) SD_F_SVMReg, SD_F_LgtGBM, SD_F_MLPerc, MW_LR_14_Days_Window, MW_35D_F_LinReg, MW_35D_F_LassoR
(b) SD_F_LinReg, SD_F_LassoR, SD_F_LgtGBM, MW_LR_14_Days_Window, MW_35D_F_RidgeR, MW_35D_F_LassoR
(c) SD_F_LassoR, SD_F_SVMReg, SD_F_LgtGBM, SD_F_CNN, MW_LR_14_Days_Window, MW_35D_F_SVMReg
(d) SD_F_LinReg, SD_F_LassoR, SD_F_LgtGBM, SD_F_MLPerc, MW_LR_7_Days_Window, MW_35D_F_LassoR

Table 8.2. Regression meta-learners' hyperparameters, their values selection, and their best values.

8.3. Evaluation of Heuristic Rule-Based Ensembles

Evaluation of heuristic rule-based ensembles, described in Chapter 3 and Section 8.1, for ODADF in Ireland in year 2019, was shown on Figures 8.1-8.2.

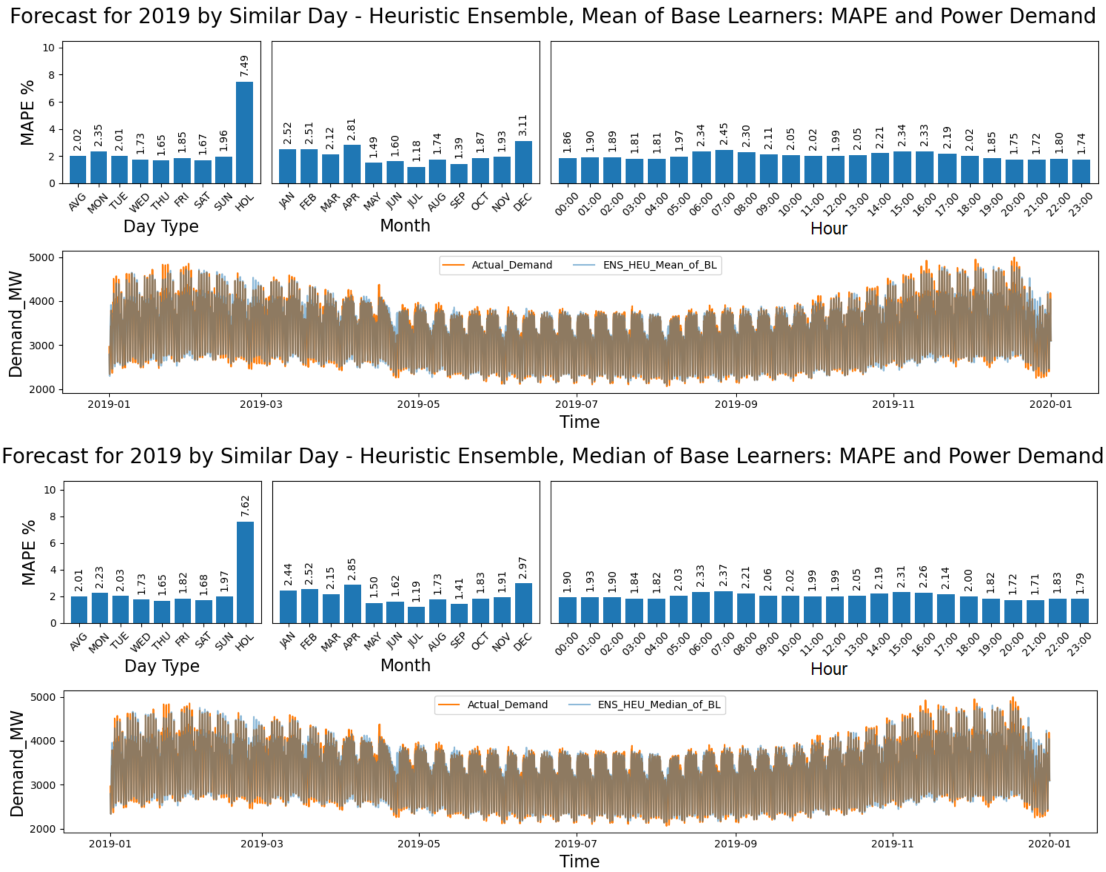


Figure 8.1. Forecast of power demand in Ireland for year 2019 by heuristic rule-based ensembles (1-2).

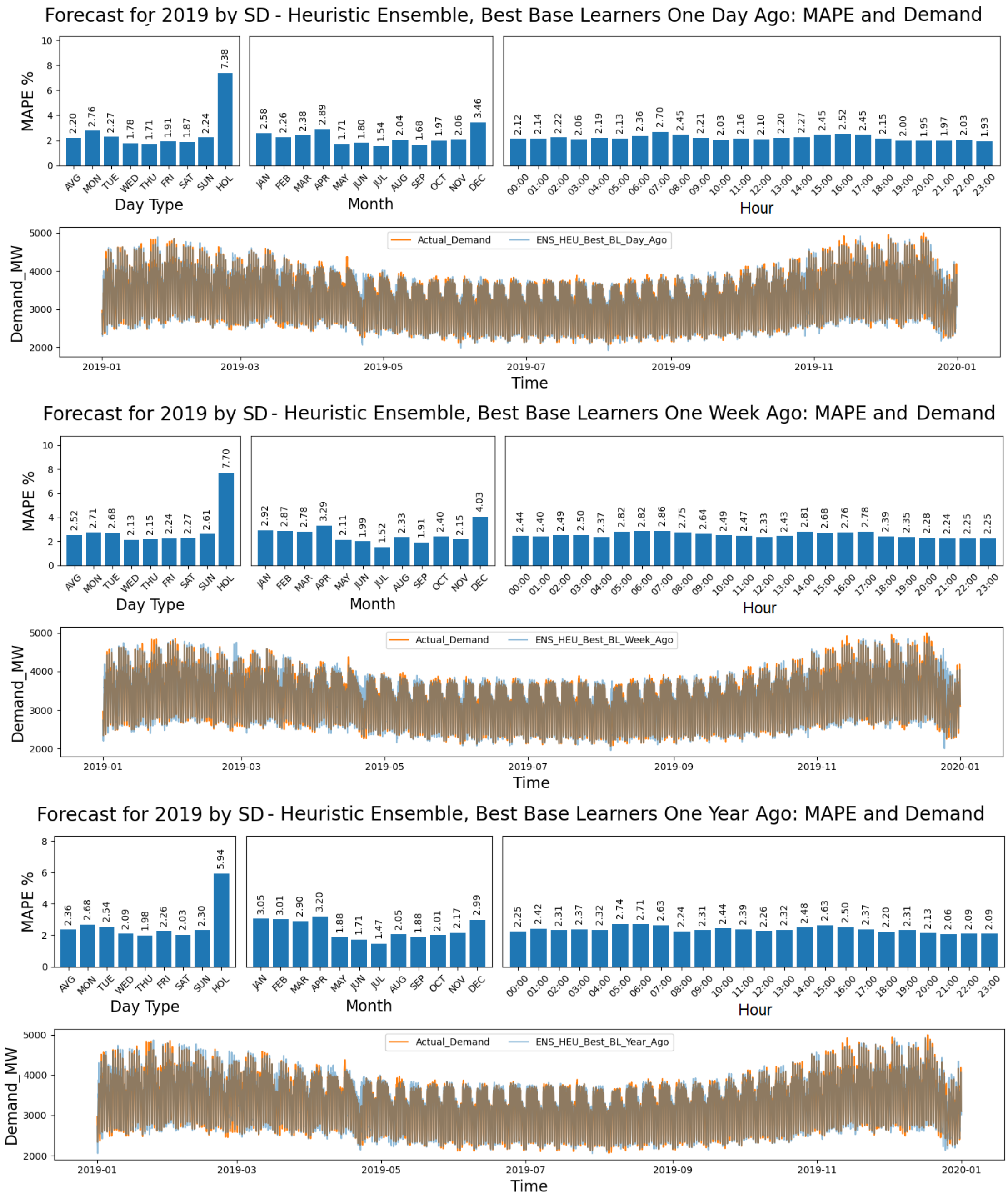


Figure 8.2. Forecast of power demand in Ireland for year 2019 by heuristic ensembles (3-5).

8.4. Evaluation of Classification-Based Ensembles

Evaluation of classification-based ensembles, described in Chapter 3 and Section 8.2, for ODADF in Ireland in year 2019, was shown on Figure 8.3.

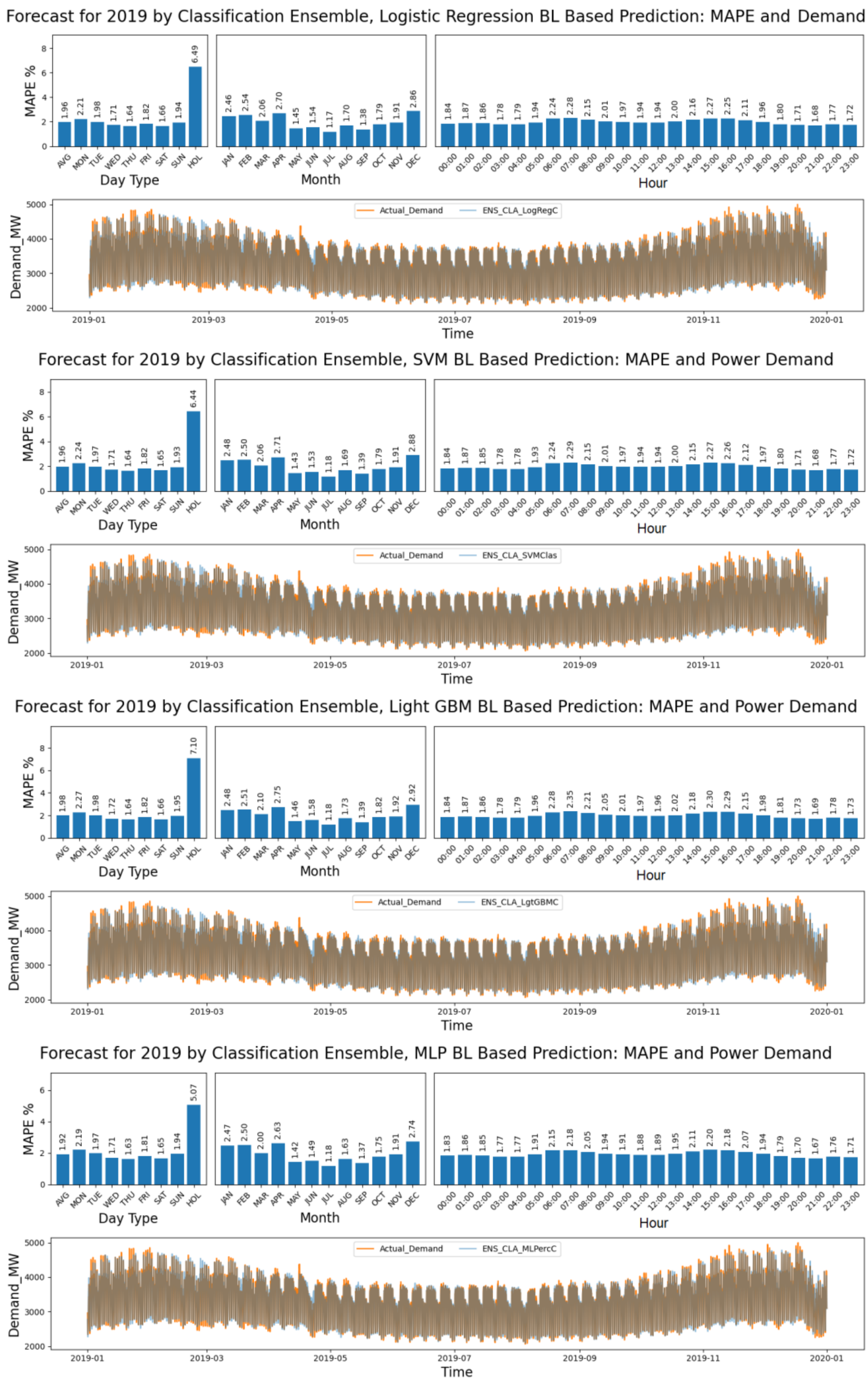


Figure 8.3. Forecast of power demand in Ireland for year 2019 by classification-based ensembles.

8.5. Evaluation of Regression-Based Ensembles

Evaluation of regression-based ensembles, described in Chapter 3 and Section 8.2, for ODADF in Ireland in year 2019, was shown on Figure 8.4.

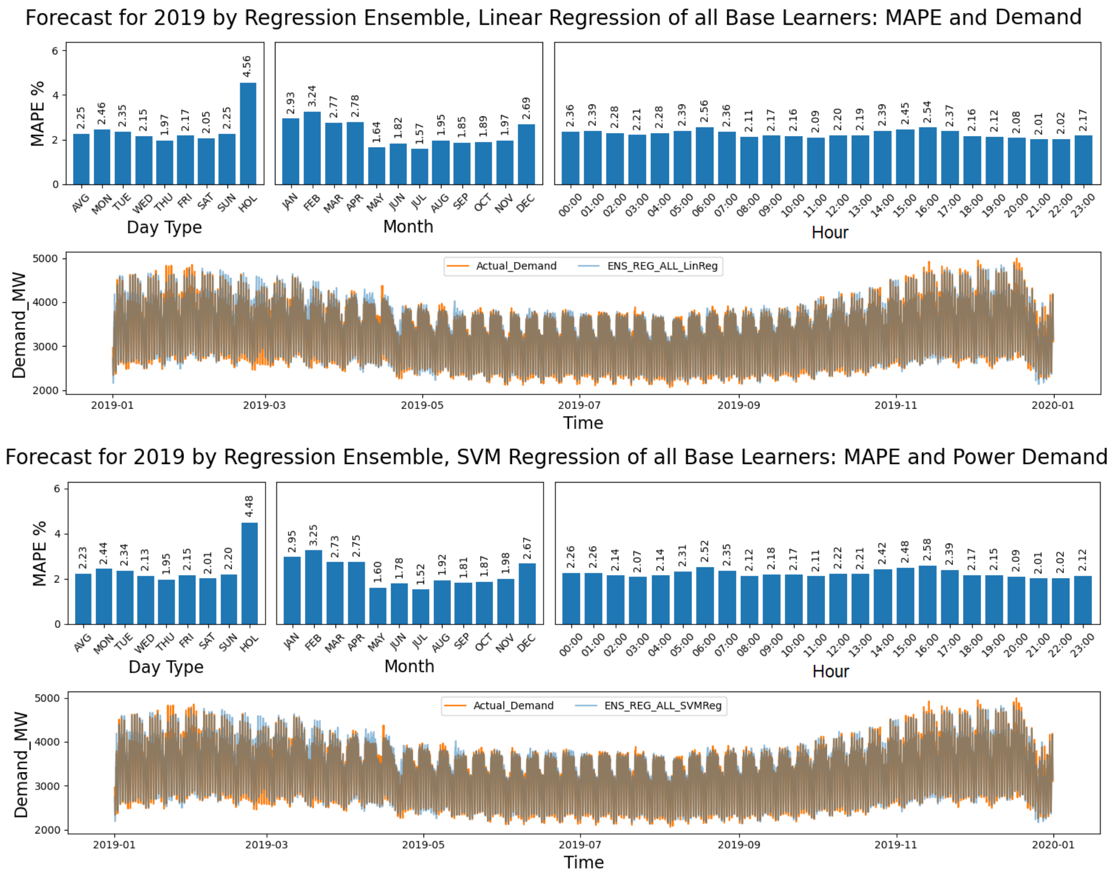


Figure 7.10. Forecast of power demand in Ireland for year 2019 by regression-based ensembles.

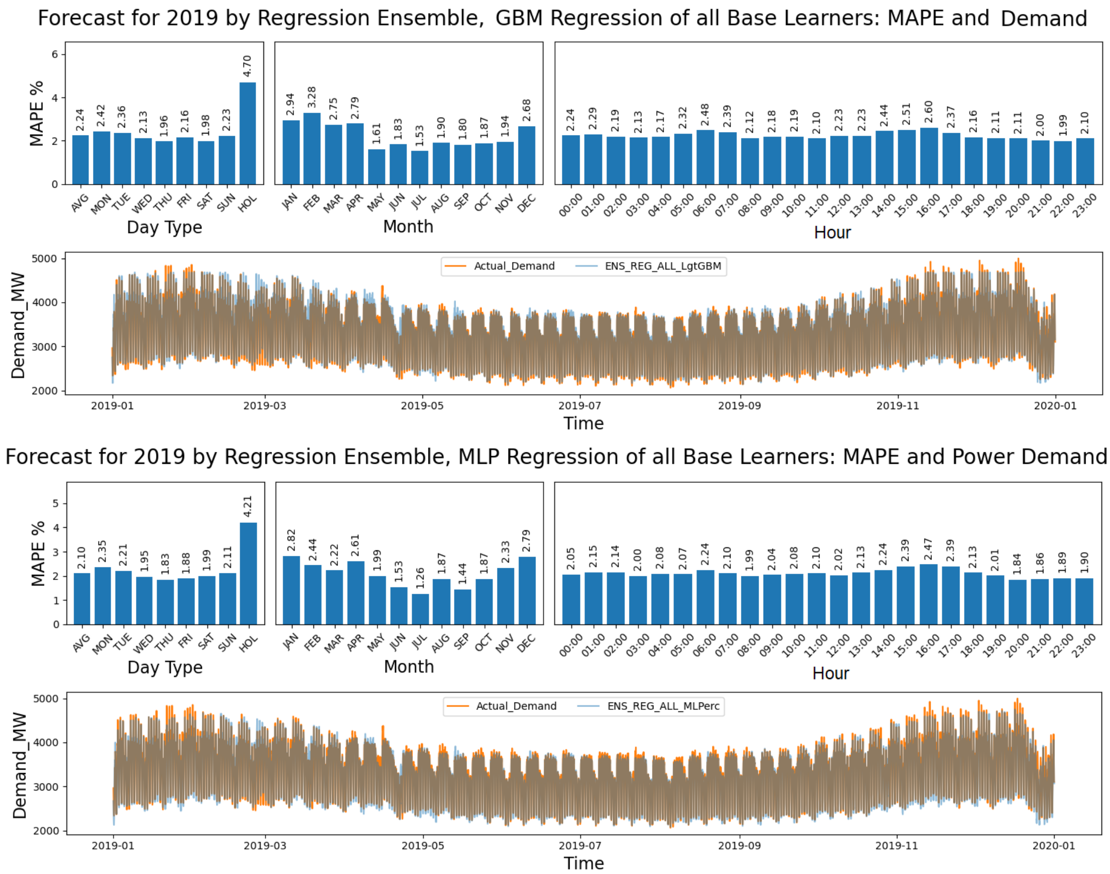


Figure 8.4. Forecast of power demand in Ireland for year 2019 by regression-based ensembles.

8.6. Ranking of Base-Learners and Ensembles for Year 2019

Ranking of base-learners and ensembles from Sections 8.3-8.5, based on average MAPE for ODADF in Ireland in year 2019, was presented on Figure 8.5. More performance metrics achieved by ensemble learning models were presented in Table 8.3. The classification-based ensembles were the winners of the study.

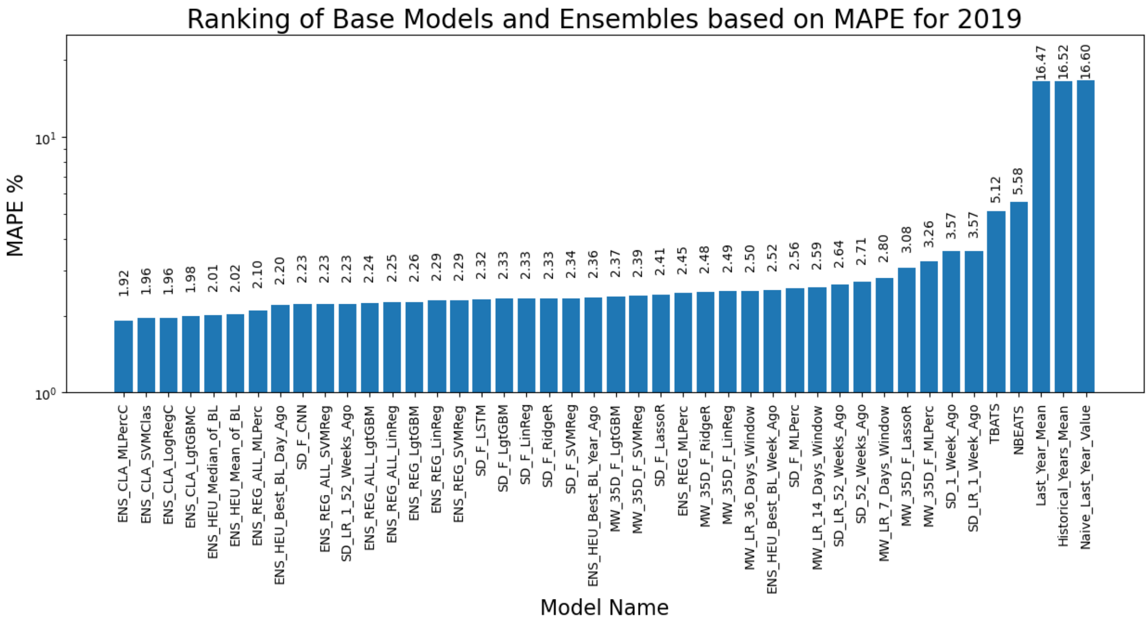


Figure 8.5. Ranking of potential base-learners and ensembles based on average MAPE for one day-ahead power demand in year 2019 in Ireland.

Model		PERFORMANCE METRICS FOR YEAR 2019		
Name	**Description**	**MAE in MW**	**RMSE in MW**	**MAPE in %**
ENS_CLA_MLPerc	Ensemble with MLP as classification meta-learner	64.1	89.1	1.919
ENS_CLA_SVMClas	Ensemble with SVM as classification meta-learner	65.3	92.3	1.960
ENS_CLA_LogRegC	Ensemble with Logistic Regression as meta-learner	65.3	92.1	1.960
ENS_CLA_LgtGBMC	Ensemble with GBM as classification meta-learner	66.1	94.1	1.985
ENS_HEU_Median_of_BL	Heuristic ensemble, median of 20 BL predictions	66.7	95.0	2.009
ENS_HEU_Mean_of_BL	Heuristic ensemble, mean of 20 BL predictions	67.3	96.8	2.020
ENS_REG_ALL_MLPerc	Ensemble with MLP as regression meta-learner of 20 BL	70.5	95.9	2.096
ENS_HEU_Best_BL_Day_Ago	Heuristic ensemble, best predictors one day ago	72.8	105.1	2.199
ENS_REG_ALL_SVMReg	Ensemble with SVM as regression meta-learner of 20 BL	74.2	101.0	2.228
ENS_REG_ALL_LgtGBM	Ensemble with GBM as regression meta-learner of 20 BL	74.3	101.4	2.235
ENS_REG_ALL_LinReg	Ensemble with LR as regression meta-learner of 20 BL	74.6	101.1	2.253
ENS_REG_LgtGBM	Ensemble with GBM as regression meta-learner of 6 BL	75.1	102.3	2.261
ENS_REG_LinReg	Ensemble with LR as regression meta-learner of 6 BL	75.7	102.4	2.289
ENS_REG_SVMReg	Ensemble with SVM as regression meta-learner of 6 BL	75.7	102.2	2.290
ENS_HEU_Best_BL_Year_Ago	Heuristic ensemble, best predictors one year ago	78.2	108.3	2.357
ENS_REG_MLPerc	Ensemble with MLP as regression meta-learner of 6 BL	81.1	107.9	2.446
ENS_HEU_Best_BL_Week_Ago	Heuristic ensemble, best predictors one week ago	83.6	119.6	2.524

Table 8.3. Performance metrics for potential ensembles forecasting power demand in Ireland for year 2019. Classification-based meta-learners of ensembles, selected for validation were marked in grey.

Figure 8.6 depicted base-learners and ensembles performance as a function of the day of the week, month, and hour.

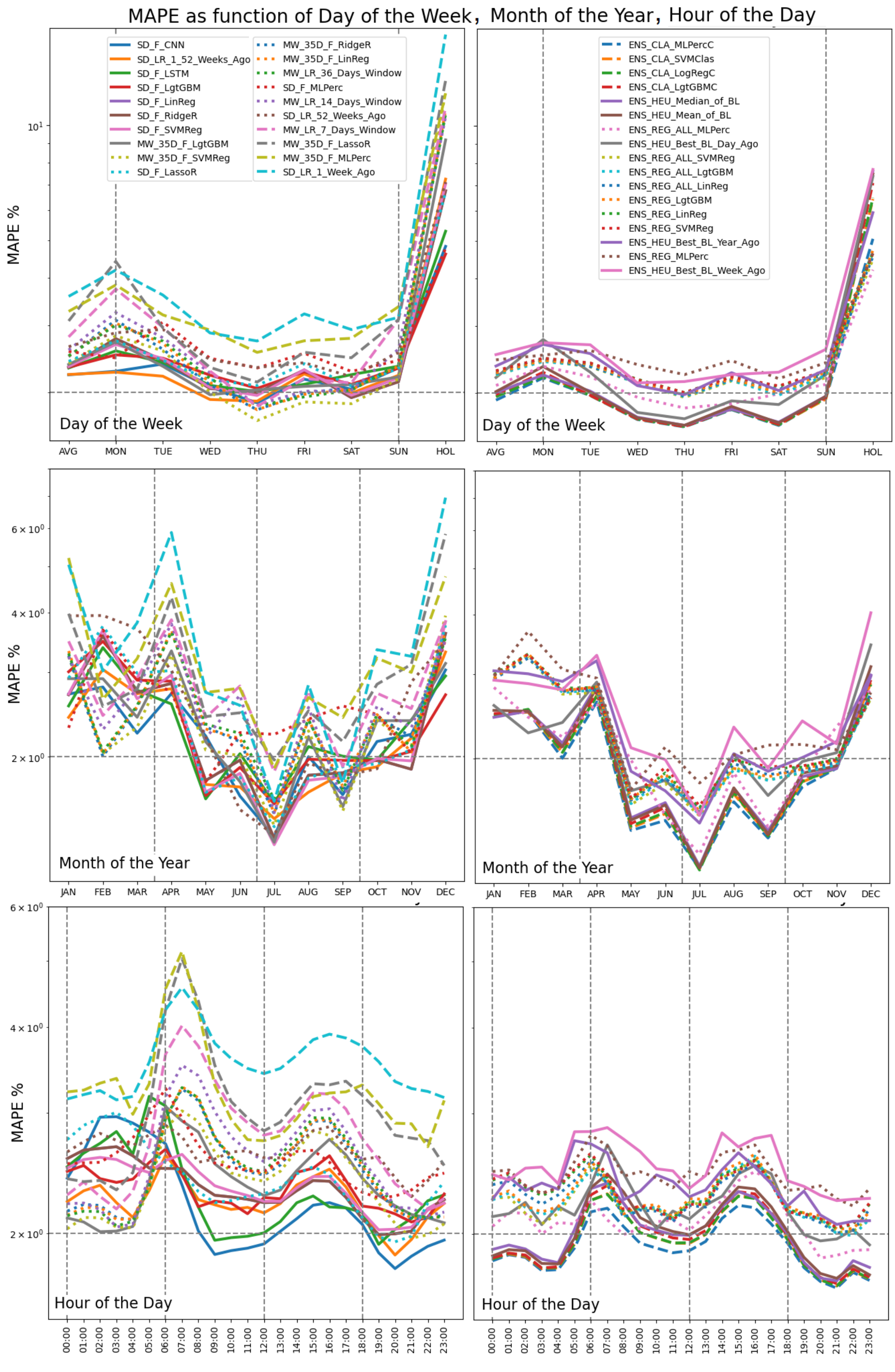


Figure 8.6. MAPE for predictions for 2019 by base-learners (left column) and ensemble learning models (right column) as function of day of the week (top row), month of the years (middle row) and hour of the day (bottom row).

Comparison of distributions, forecasted by ensembles and actual power demand, for year 2019 was presented on Figure 8.7. As power demand was not normally distributed, non-parametric Kolmogorov-Smirnov test was used to investigate whether forecasted and actual demand, in pairs, were from the same distribution (Table 8.4).

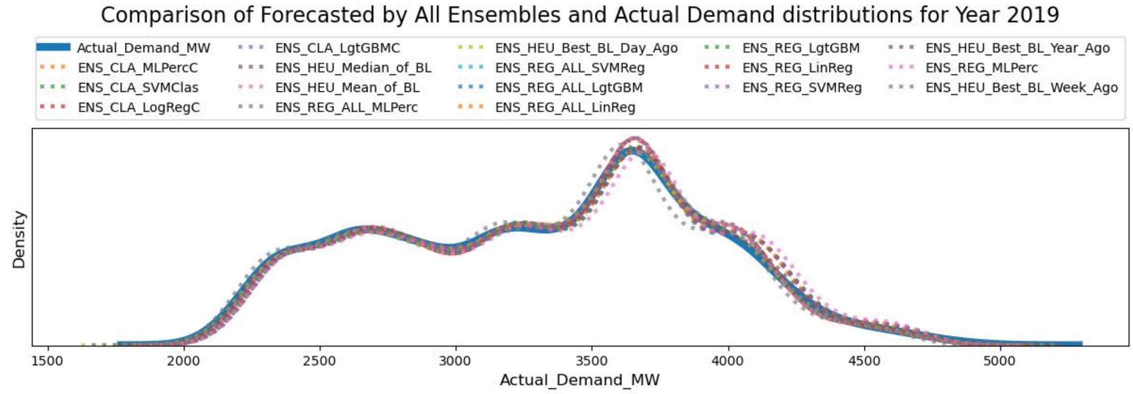


Figure 8.7. Comparison of forecasted by ensembles and actual power demand distributions for Year 2019.

Samples of predictions of below ensembles and actual demand in 2019	
were from the same distributions	**were from different distributions**
ENS_CLA_MLPercC	ENS_REG_ALL_MLPerc
ENS_CLA_SVMClas	ENS_REG_ALL_SVMReg
ENS_CLA_LogRegC	ENS_REG_ALL_LgtGBM
ENS_CLA_LgtGBMC	ENS_REG_ALL_LinReg
ENS_HEU_Median_of_BL	ENS_REG_LgtGBM
ENS_HEU_Mean_of_BL	ENS_REG_LinReg
ENS_HEU_Best_BL_Day_Ago	ENS_REG_SVMReg
ENS_HEU_Best_BL_Year_Ago	ENS_REG_MLPerc
ENS_HEU_Best_BL_Week_Ago	

Table 8.4. Results of Kolmogorov-Smirnov test for predictions of ensembles and actual power demand in 2019.

8.7. One Day-Ahead Power Demand Forecasting in Ireland for Year 2019

To visualise the potential of ensemble learning models in ODADF in Ireland, time series of predictions by the best ensembles, such as classification-based ensemble with MLP and SVM as meta-learners were compared to the prediction of the best base-learner CNN (SD) and real power demand in Ireland in 2019 on Figure 8.8.

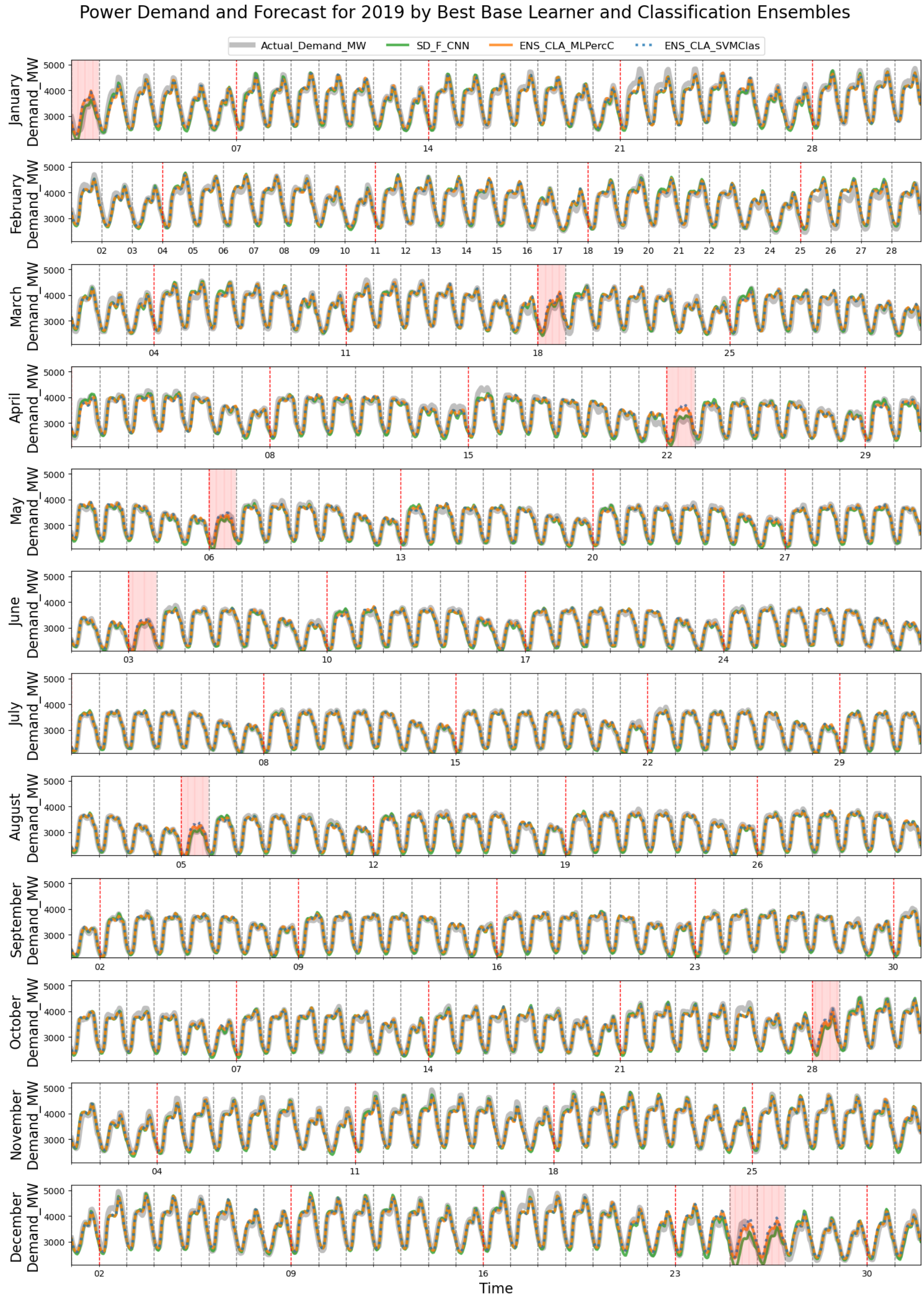


Figure 8.8. One day-ahead power demand forecasting for year 2019 using best base-learner two best ensembles. Holidays on weekdays are marked in pink.

8.8. Discussion Regarding Results of the Experimentation with Architectures of Ensembles

Experimentation with meta-learners of ensembles revealed their more consistent performance compared to standalone base-learners. For example, MAE, RMSE and MAPE varies 64.1-83.6MW, 89.1-119.6MW, and 1.92-2.52%, respectively. Regression-based ensembles that integrated twenty base-learner predictions exhibited superior performance compared to those restricted to a limited number of predictions. This suggested that ensemble models possess the capability to effectively handle and manage their input.

All classification-based ensembles outperformed other meta-learners and best base-learners. Ensemble of twenty base-learners, with MLP classifier as meta-learner, achieved the lowest MAPE 1.92%, which was 17.2% improvement in comparison to the best base-learner, CNN (SD) MAPE 2.32%.

Ensemble models utilizing SVM, Logistic Regression, and GBM as meta-learners closely trailed the leading model, registering MAPE scores of 1.96%, 1.96%, and 1.99%, respectively. Three other ensembles - two heuristic rule-based ones using median and mean, and a regression-based ensemble with MLP regressor incorporating twenty base-learner predictions - also surpassed the performance of the best base-learner. Nevertheless, other ensembles did not achieve this benchmark. Even so, the ensemble with the lowest performance recorded MAPE of 2.52%, which was not only proximate to the best base-learner but is also significantly better than the poorest-performing base-learner, which had a MAPE of 3.57%. Interestingly, classification-based ensembles using Logistic Regression and SVM as meta-learners performed better with weather factors from the virtual weather station created with Linear Regression. Meanwhile, ensembles utilizing GBM and MLP favoured the Lasso Regression.

Finally, while comparison of predictions by ensembles and actual power demand in 2019 distributions looked similar visually, inferential statistics tests revealed that, predictions of classification and heuristic rule-based models, aligned with the distribution of the actual demand, while those of regression-based ensembles models deviated from actual demand distribution.

8.7. Conclusion

The primary research established a cause-and-effect relationship between ensemble configurations and ODADF performance metrics. While the performance of ensemble learning models was negatively impacted by limiting the number of base-learners' predictions, changing the integration approach, while keeping their number constant, had significant influence on the overall performance. Therefore, the integration method of base-learners emerged as the primary causal variable.

Additionally, the data preparation phase for base-learners was found as important as designing the ensembles’ architectures. The introduction of the SD and MW approaches amplified the variety of the base-learners' predictions. Subsequently, incorporating diverse base-learners, with varied performance across different timeframes, proved successful in ensemble constructions. Ensemble systems not only harnessed the combined strengths but also mitigated the potential inconsistencies found in individual base-learners. This was evident from the notably reduced performance variations among ensembles in comparison to base-learners. The superiority of regression-based ensembles, especially those that incorporated the full spectrum of twenty base-learner predictions, indicated the adaptability and self-management capabilities of ensemble models.

Classification-based ensembles, marked significant advancements over other meta-learners and leading base-learners. Interestingly, even the lesser-performing ensembles not only approached the proficiency of the best base-learner but also outstripped the performance of the least effective base-learner. Moreover, while half of the base-learner predictions deviated from the distribution of actual demand, predictions from classification and heuristic rule-based models aligned with the distribution of the actual demand. Furthermore, while ensembles using Logistic Regression and SVM as meta-learners performed better with weather factors from the virtual weather station created with Linear Regression, ensembles utilizing GBM and MLP favoured the Lasso Regression. Finally, performance of classification-based ensembles for ODADF in Ireland must be validated on unseen data from years 2021-2022.

9. Validation of the Experimentation's Results

Classification-based ensembles, as the best approach revealed in Chapter 8, were refitted and tested on unseen data. All base-learners, with hyperparameters tuned on training dataset, were fitted on data from year 2021, and validated on data from year 2022. Subsequently, classification meta-learners, with hyperparameters tuned on training dataset, were fitted on base-learners predictions for years 2021-2022, and validated on data from year 2022.

9.1. Refitting and Evaluation of Base-Learners

Evaluation of twenty base-learners of classification-based ensembles, for ODADF in Ireland in year 2022, was shown on Figures 9.1-9.6.

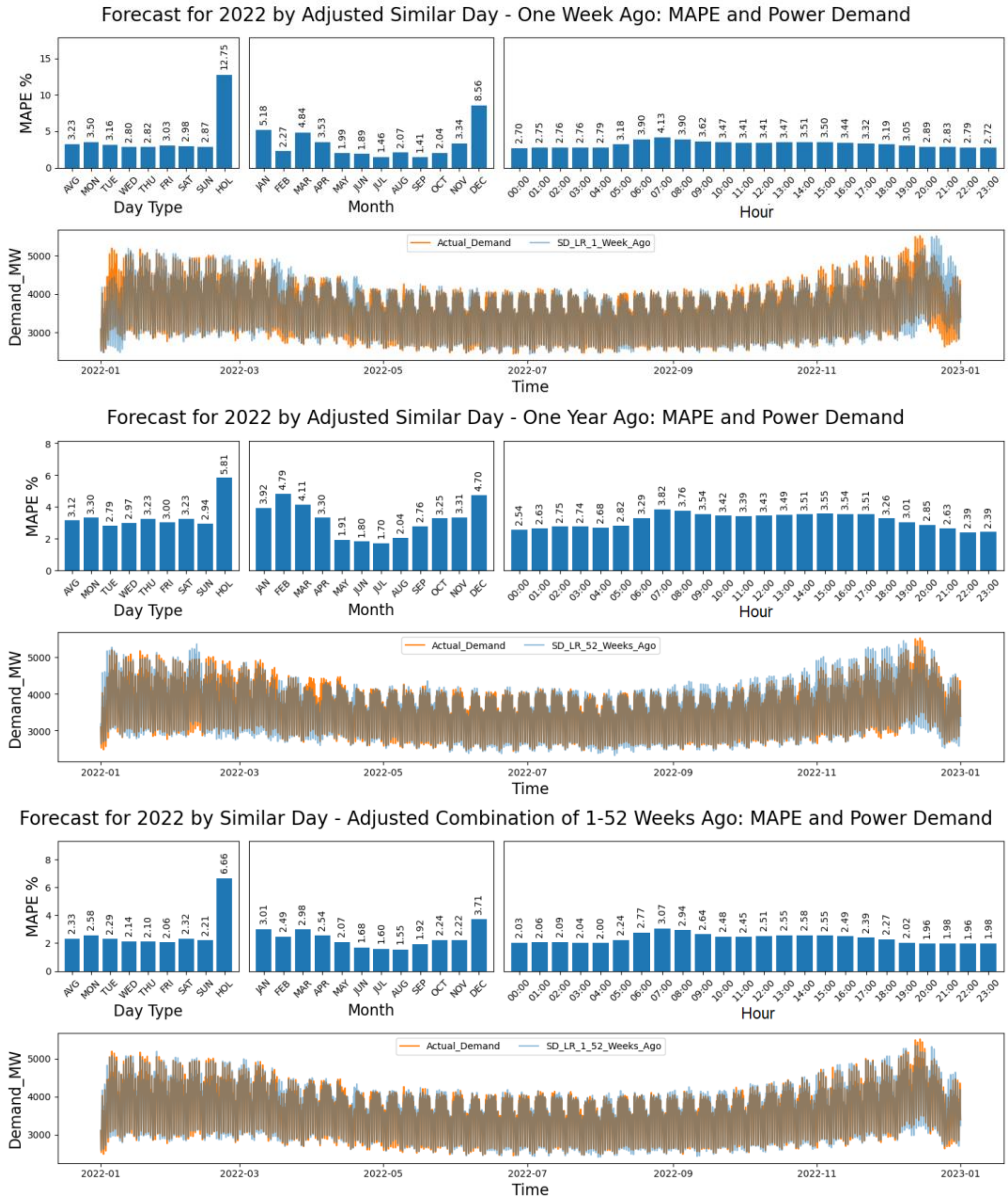


Figure 9.1. Forecast of power demand in Ireland for year 2022 by base-learners (1-3).

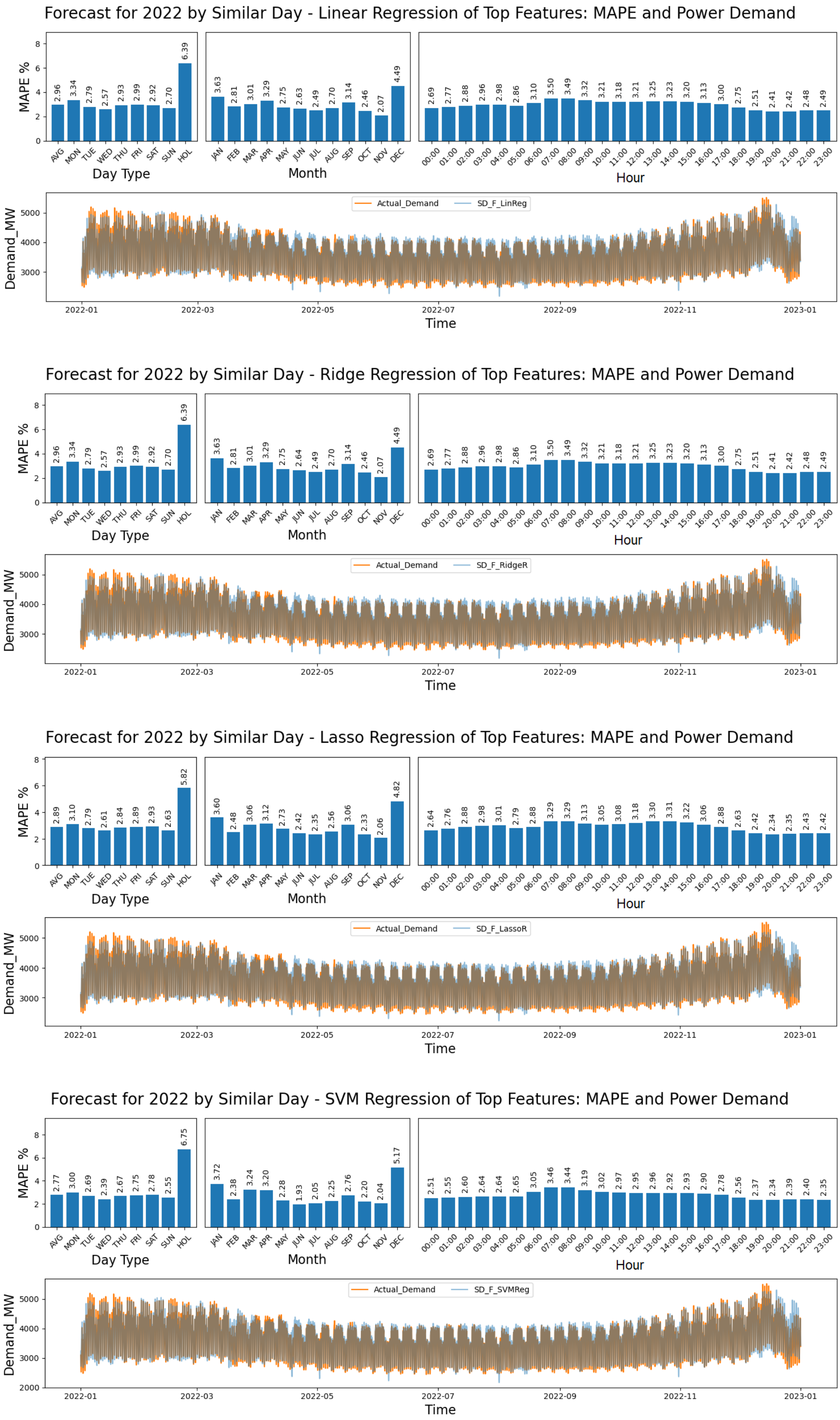


Figure 9.2. Forecast of power demand in Ireland for year 2022 by base-learners (4-7).

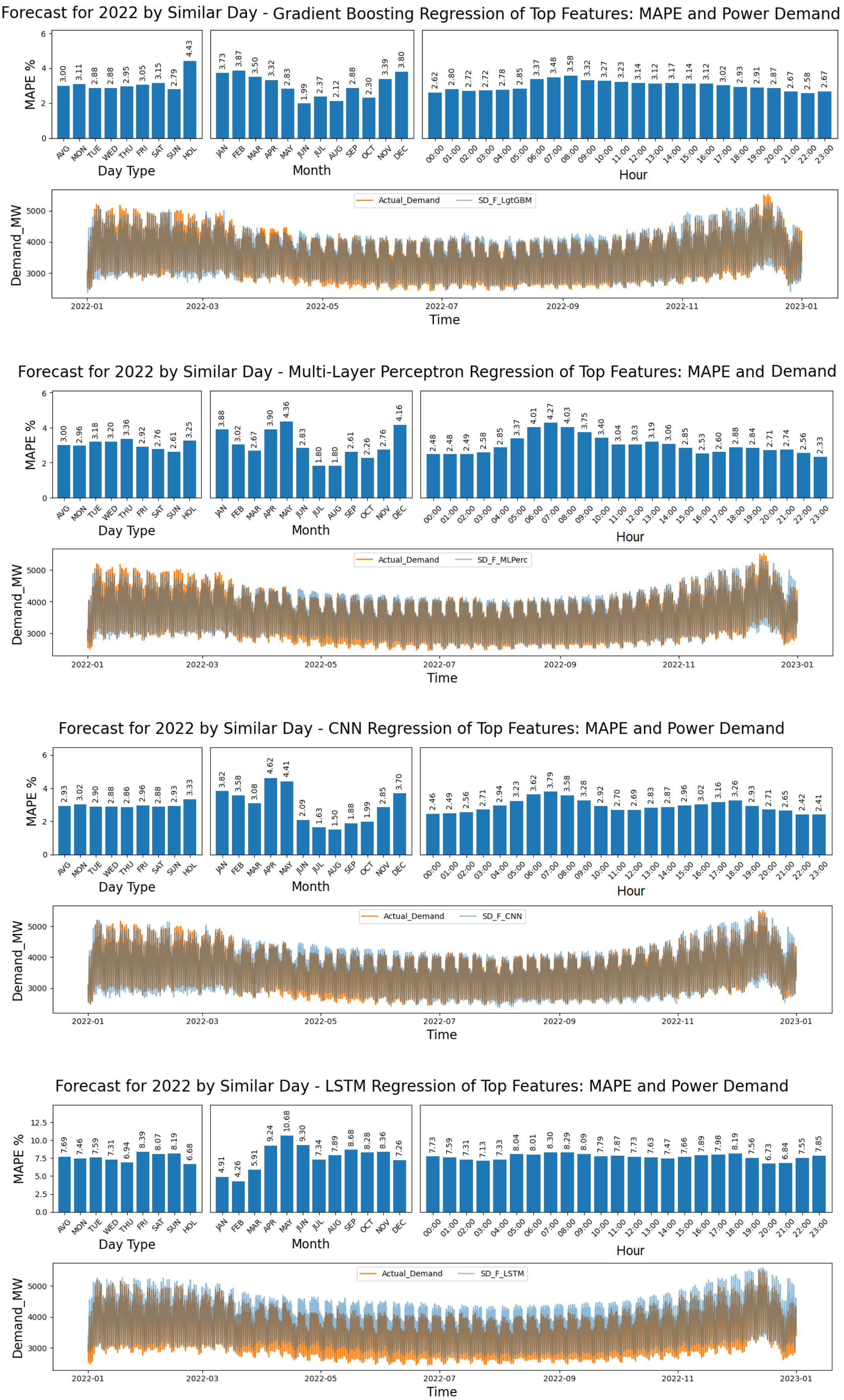


Figure 9.3. Forecast of power demand in Ireland for year 2022 by base-learners (8-11).

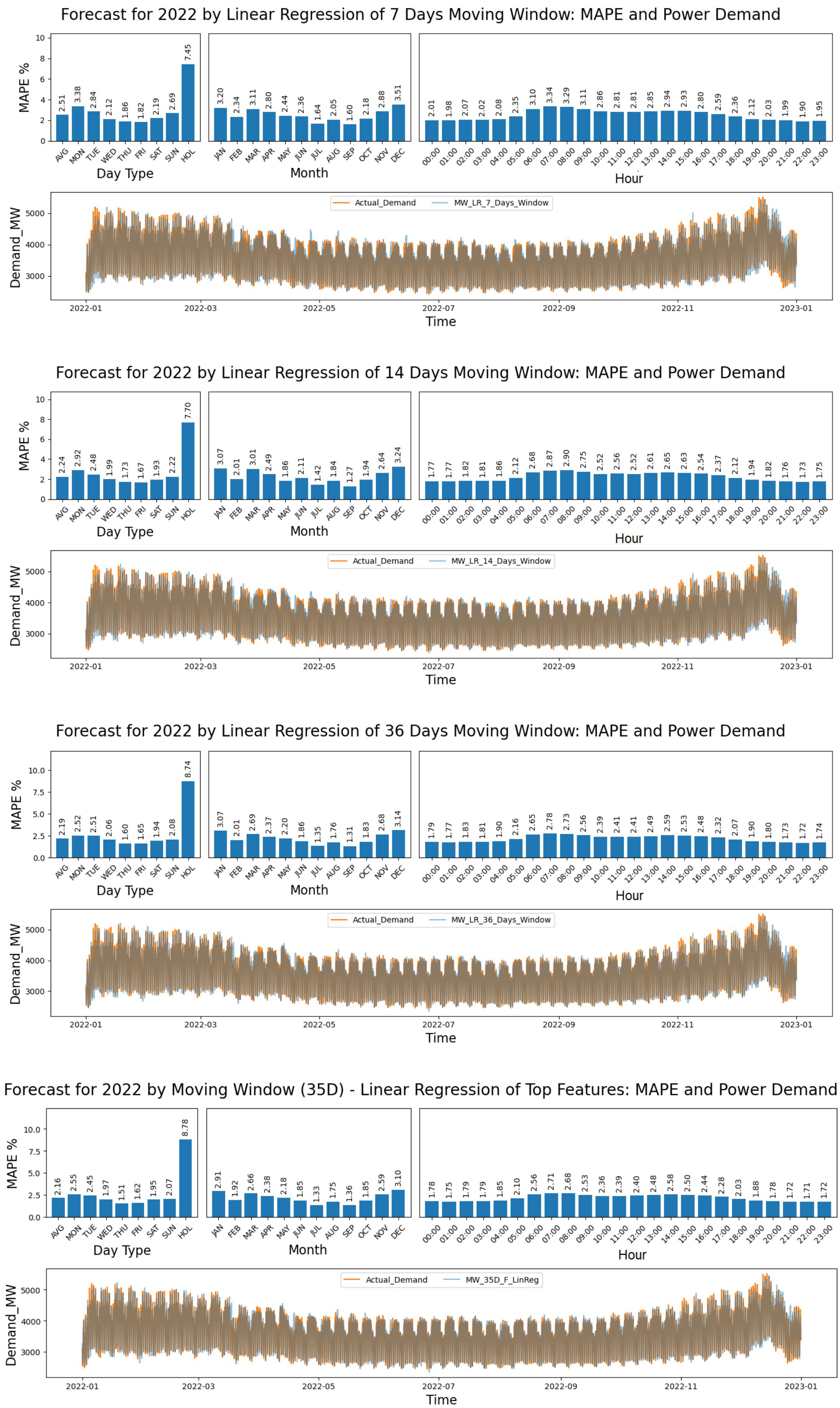


Figure 9.4. Forecast of power demand in Ireland for year 2022 by base-learners (12-15).

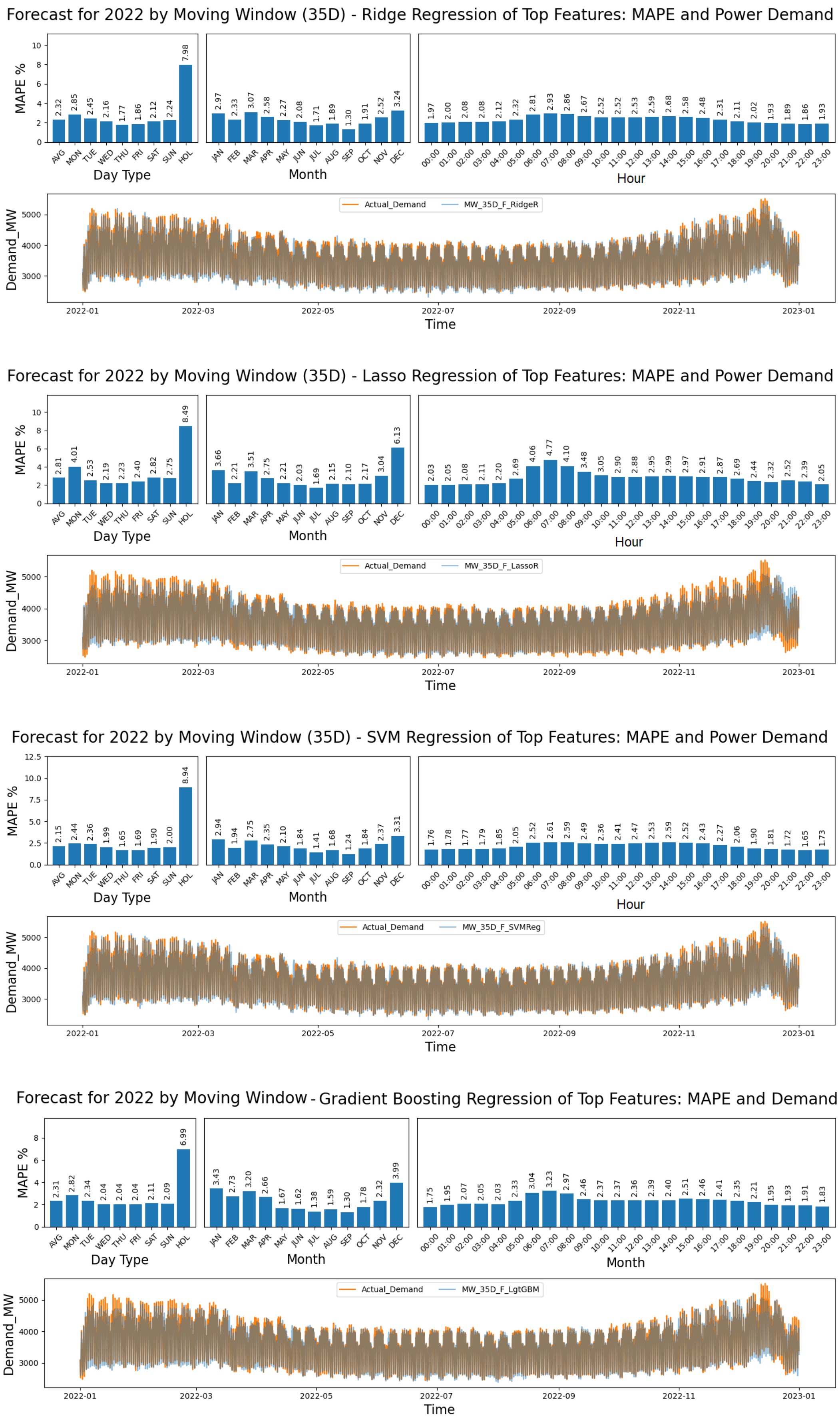


Figure 9.5. Forecast of power demand in Ireland for year 2022 by base-learners (16-19).

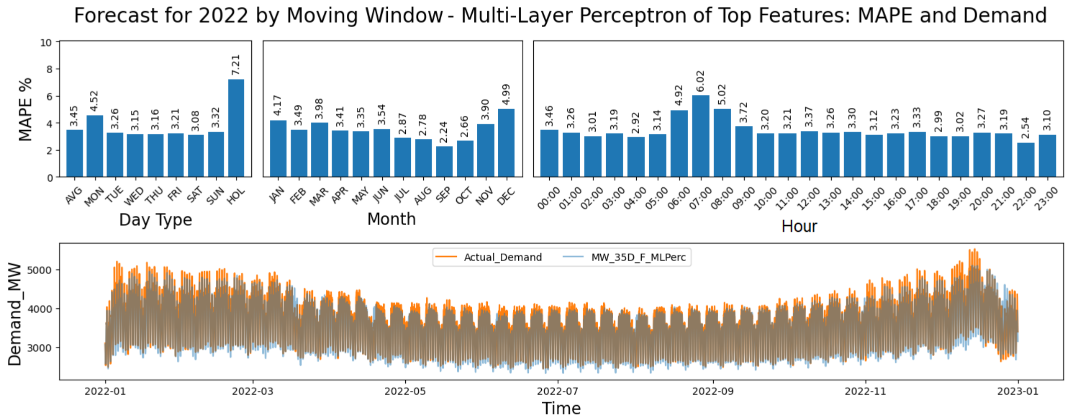


Figure 9.6. Forecast of power demand in Ireland for year 2022 by base-learners (20).

9.2. Ranking of Base-Learners for Year 2022

Ranking of the twenty base-learners as function of day, month, and hour, based on MAPE for ODADF in Ireland in year 2019 was shown on Figure 9.7.

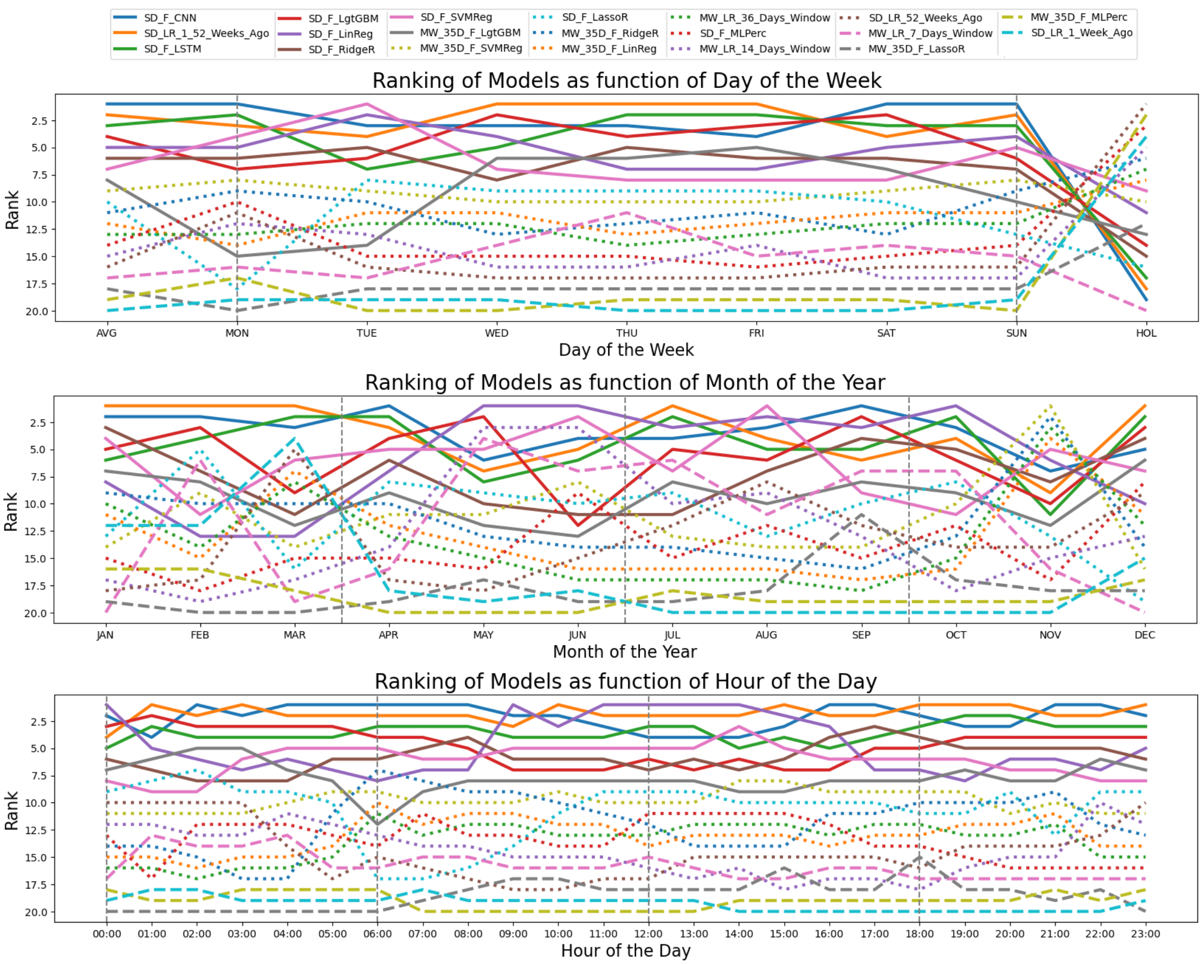


Figure 9.7. Ranking of 20 base-learners as function of day, month and hour, based on MAPE for year 2022.

Comparison of distributions, forecasted by base-learners and actual power demand, for year 2022 was shown on Figure 9.8. Non-parametric Kolmogorov-Smirnov test was used to investigate whether forecasted and actual demand, in pairs, were from the same distribution (Table 9.2).

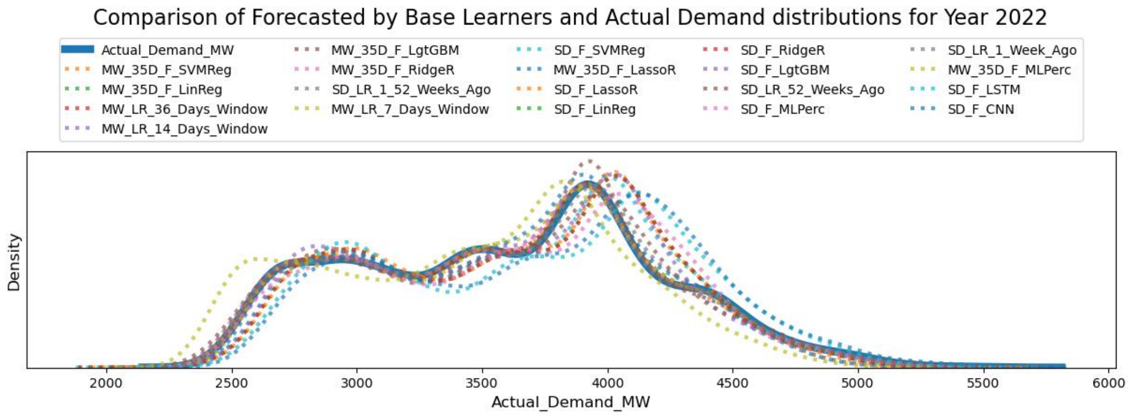


Figure 9.8. Comparison of forecasted by base-learners and actual power demand distributions for Year 2022.

Samples of predictions of below base-learners and actual demand in 2022	
were from the same distributions	**were from different distributions**
SD_LR_1_52_Weeks_Ago	SD_F_CNN
MW_35D_F_SVMReg	SD_F_LSTM
MW_LR_36_Days_Window	SD_F_LgtGBM
MW_35D_F_LinReg	SD_F_LinReg
MW_LR_14_Days_Window	SD_F_RidgeR
MW_LR_7_Days_Window	SD_F_SVMReg
SD_LR_1_Week_Ago	MW_35D_F_LgtGBM
	SD_F_LassoR
	MW_35D_F_RidgeR
	SD_F_MLPerc
	SD_LR_52_Weeks_Ago
	MW_35D_F_LassoR
	MW_35D_F_MLPerc

Table 9.2. Results of Kolmogorov-Smirnov test for predictions of base-learners and actual power demand in 2022.

9.3. Percentage Share of Base-Learners in the Best Hourly Predictions for Year 2022

The percentage share of base-learners in the best hourly prediction was shown by year (Figure 9.9), and by day type (Figure 9.10).

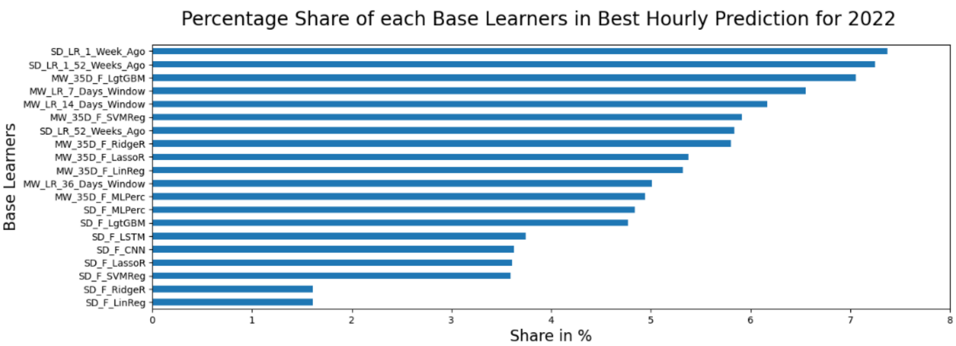


Figure 9.9. Percentage share of base-learners in the best hourly prediction for years 2015-2019 by year.

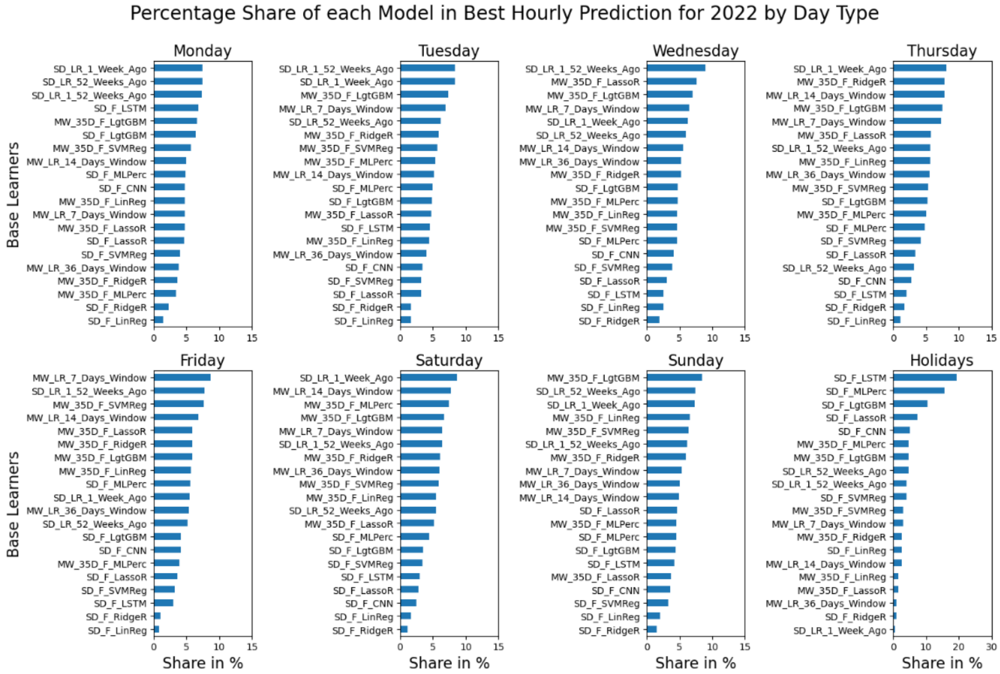


Figure 9.10. Percentage share of base-learners in the best hourly prediction for years 2015-2019 by day type.

9.4. Refitting and Evaluation of Classification-Based Ensembles

Evaluation of classification-based ensembles, for ODADF in Ireland in year 2022, was shown on Figures 9.11-9.12.

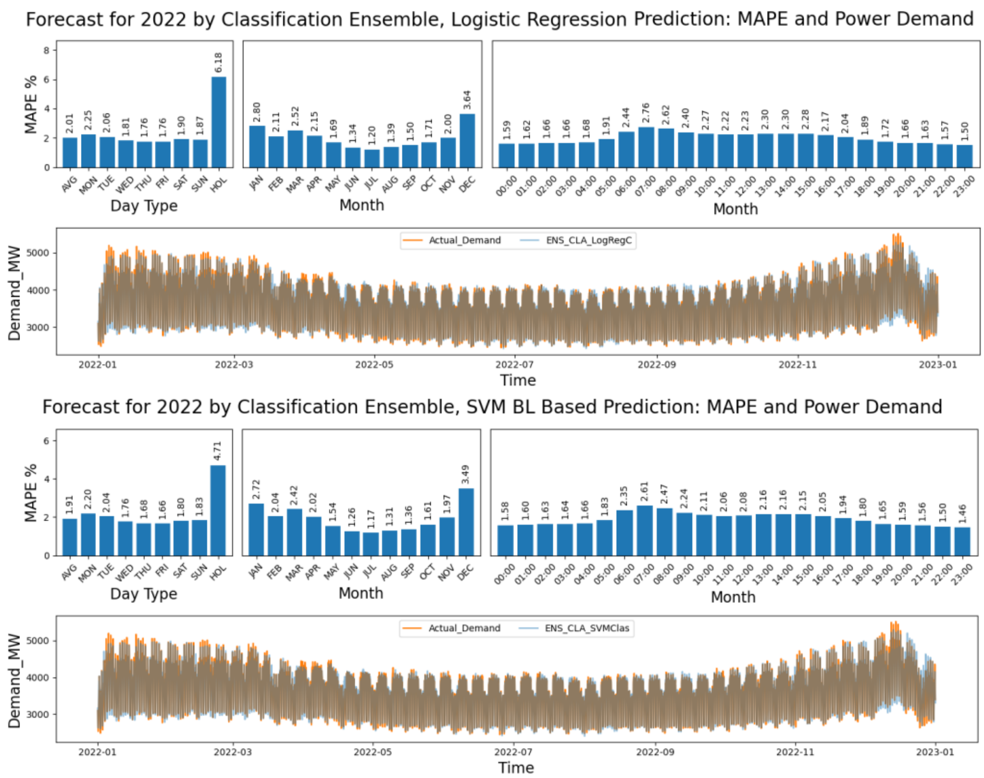


Figure 9.11. Forecast of power demand in Ireland for year 2022 by classification-based ensembles (1-2).

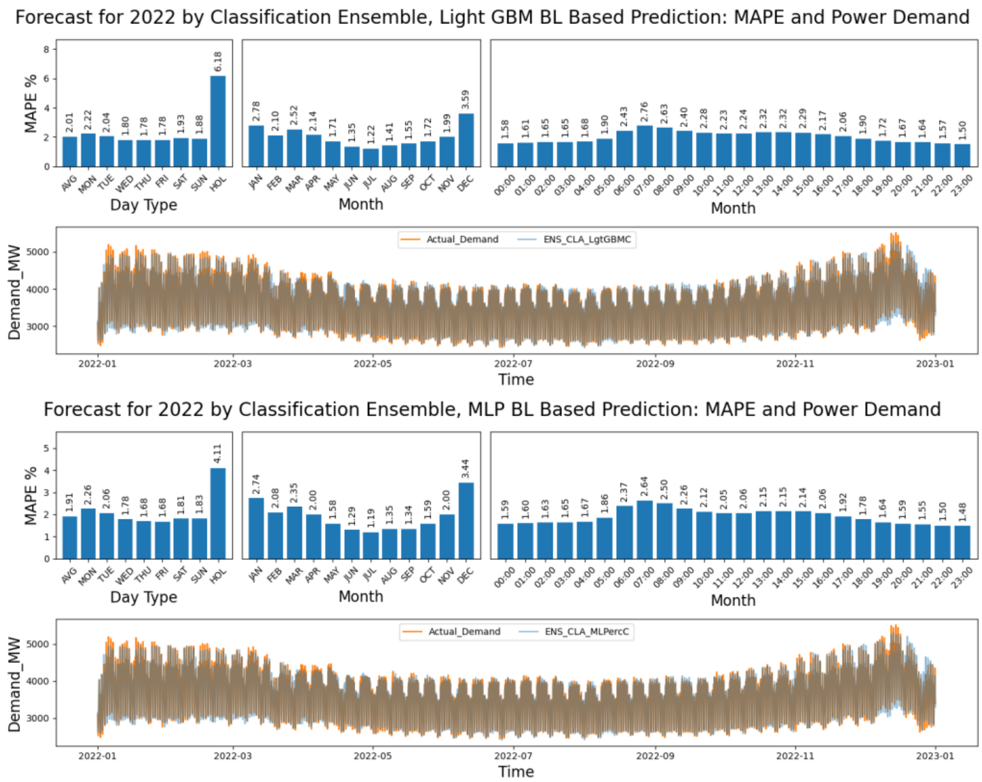


Figure 9.12. Forecast of power demand in Ireland for year 2022 by classification-based ensembles (3-4).

9.5. Validation Ranking of Base-learners and Classification-Based Ensembles for Year 2022

Ranking of base-learners and ensembles from Sections 9.1-9.4, based on average MAPE for ODADF in Ireland in year 2022, was shown on Figure 9.13. More performance metrics achieved by the best twenty ensembles and base-learners were presented in Table 9.3. The classification-based ensembles were the winners of the study, again.

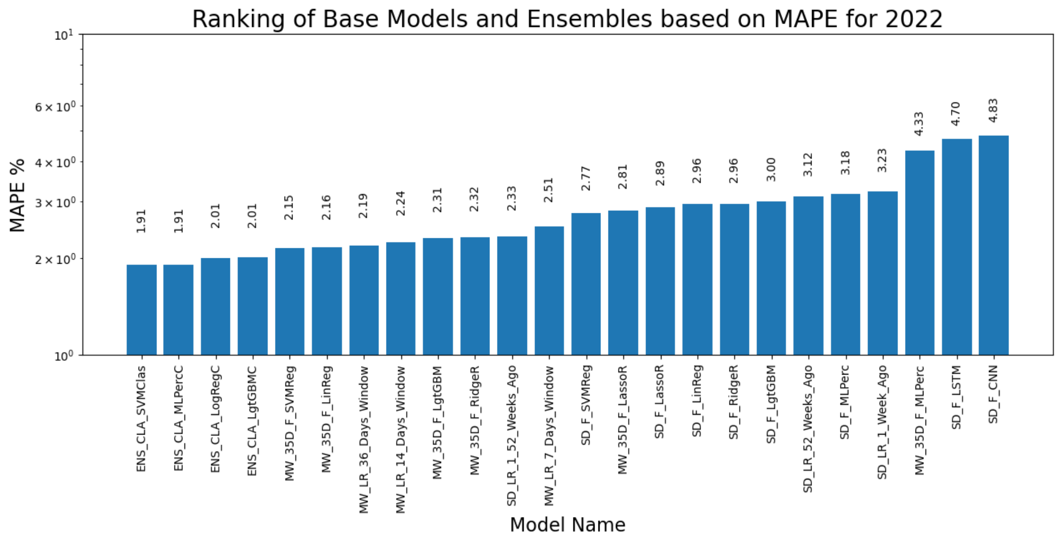


Figure 9.13. Ranking of base-learners and ensembles based on average MAPE for one day-ahead power demand in year 2022 in Ireland.

Model	PERFORMANCE METRICS FOR YEAR 2022		
	MAE in MW	RMSE in MW	MAPE in %
ENS_CLA_SVMClas	70.3	100.3	1.911
ENS_CLA_MLPercC	70.5	99.9	1.914
ENS_CLA_LogRegC	73.9	105.8	2.006
ENS_CLA_LgtGBMC	74.0	105.5	2.009
MW_35D_F_SVMReg	79.0	117.1	2.152
MW_35D_F_LinReg	78.9	115.7	2.159
MW_LR_36_Days_Window	80.2	116.9	2.191
MW_LR_14_Days_Window	82.4	121.2	2.244
MW_35D_F_LgtGBM	85.6	124.3	2.306
MW_35D_F_RidgeR	85.1	119.5	2.324
SD_LR_1_52_Weeks_Ago	85.5	117.6	2.335
MW_LR_7_Days_Window	92.0	134.0	2.511
SD_F_SVMReg	100.1	138.2	2.774
MW_35D_F_LassoR	103.7	150.7	2.813
SD_F_LassoR	103.9	136.7	2.888
SD_F_LinReg	106.4	139.4	2.959
SD_F_RidgeR	106.4	139.4	2.959
SD_F_LgtGBM	109.9	143.5	3.003
SD_LR_52_Weeks_Ago	116.3	162.5	3.123
SD_F_MLPerc	116.4	152.5	3.176
SD_LR_1_Week_Ago	119.5	196.9	3.228
MW_35D_F_MLPerc	155.7	192.6	4.327
SD_F_LSTM	169.1	205.8	4.701
SD_F_CNN	173.1	205.6	4.831

Table 9.3. Performance metrics for twenty best base-learners and classification-based ensembles forecasting power demand in Ireland for year 2022. Classification-based ensembles, selected for validation, were marked in grey.

Figure 9.14 depicted base-learners and classification-based ensembles' performance as a function of the day of the week, month, and hour.

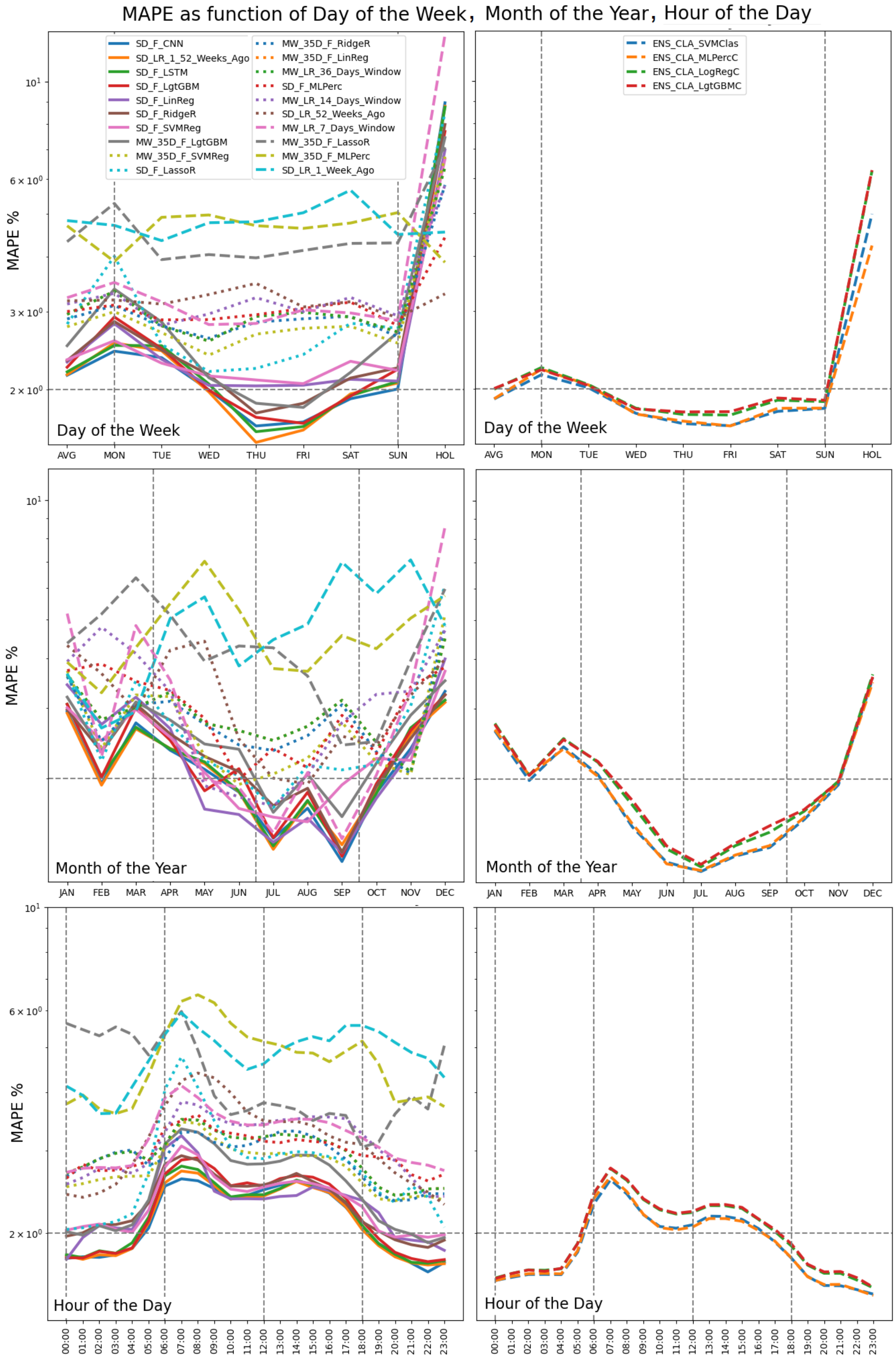


Figure 9.14. MAPE for predictions for 2022 by base-learners (left column) and ensemble learning models (right column) as function of day of the week (top row), month of the years (middle row) and hour of the day (bottom row).

Comparison of distributions, forecasted by ensembles and actual power demand, for year 2022 was shown on Figure 9.15. Non-parametric Kolmogorov-Smirnov test was used to investigate whether forecasted and actual demand, in pairs, were from the same distribution (Table 9.4).

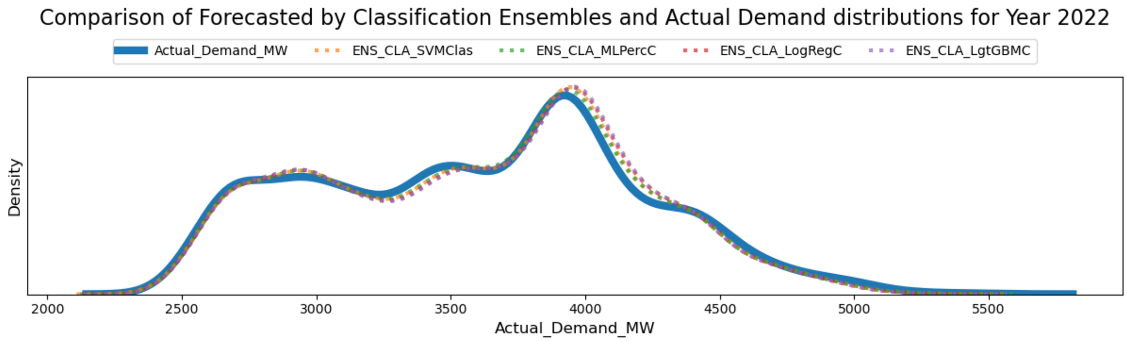


Figure 9.15. Comparison of forecasted by ensembles and actual power demand distributions for Year 2019.

Samples from predictions of below ensembles and actual demand in 2019	
were from the same distributions	**were from different distributions**
ENS_CLA_MLPercC	ENS_CLA_LogRegC
ENS_CLA_SVMClas	ENS_CLA_LgtGBMC

Table 9.4. Results of Kolmogorov-Smirnov test for predictions of ensembles and actual power demand in 2022.

9.6. One Day-Ahead Forecasting of Power Demand in Ireland for Year 2022

To visualise the potential of ensemble learning models in ODADF in Ireland, time series of predictions by the best ensembles, such as classification-based ensemble with MLP and SVM as meta-learners were compared to the prediction of the best base-learner SVM (MW) and real power demand in Ireland in 2022 on Figure 9.16.

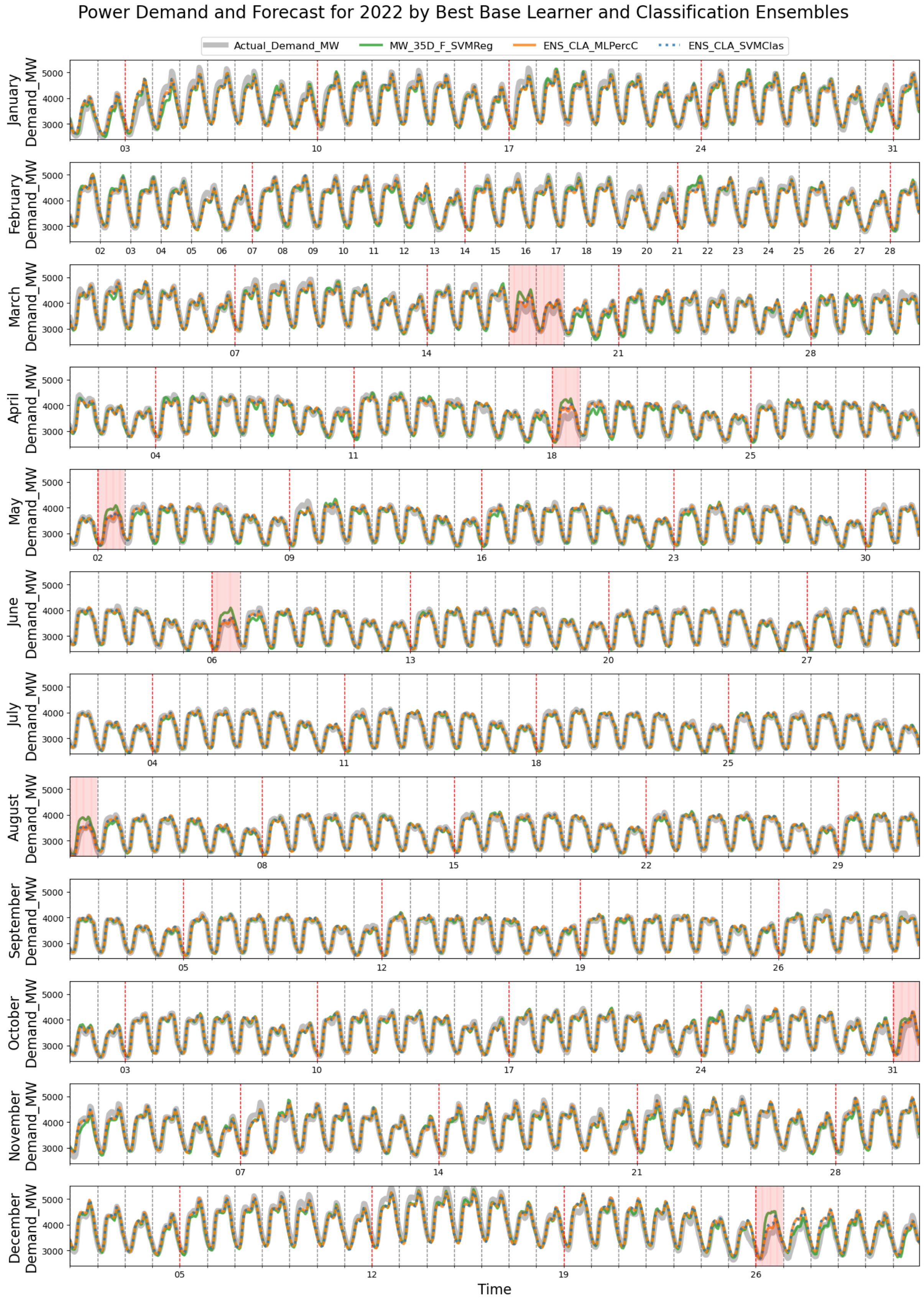


Figure 9.16. One day-ahead power demand forecasting for year 2019 using best base-learner two best ensembles. Holidays on weekdays were marked in pink.

9.7. Discussion Regarding Validation of Experimentation Results

The validation of experimentation with classification-based ensembles verified that they exhibited notably less performance variation compared to base-learners. For example, MAE, RMSE and MAPE varied 70.3-74.0MW, 100.3-105.5MW, and 1.91-2.01%, respectively. For comparison, MAE, RMSE and MAPE for twenty base-learners varied 79.0-173.1MW, 117.1-205.6MW, and 2.15-4.83%, respectively.

All classification-based ensembles outperformed the best base-learner. Ensembles of twenty base-learners, with SVM and MLP classifiers as meta-learner, achieved the lowest MAPE 1.91%, which was 11.2% improvement in comparison to the best base-learner, SVM (SD) MAPE 2.15%. Ensemble models utilizing Logistic Regression, and GBM as meta-learners closely trailed the leading model, registering MAPE scores of 2.01%.

Finally, while comparison of predictions by ensembles and actual power demand in 2019 distributions looked similar visually, inferential statistics tests revealed that, predictions of two winning classification-based models, utilising MLP and SVM as meta-learners, aligned with the distribution of the actual demand, and those utilising Logistic Regression and GBM as meta-learners, deviated from the distribution of actual demand.

9.8. Conclusion

The efficacy of classification-based ensembles, which integrated a variety of diverse base-learners, was validated on previously unseen data. Ensembles not only harnessed the combined strengths but also mitigated the potential inconsistencies found in individual base-learners. Furthermore, while predictions of 65% base-learners deviated from the distribution of actual demand, predictions of classification-based models, utilising MLP and SVM as meta-learners, aligned with the distribution of the actual demand. That underscored the value of leveraging a variety of models when constructing ensemble systems, which led to more robust and accurate ensemble predictions.

Finally, classification-based ensembles stood out as the most effective solution in ODADF. Specifically, these models employed meta-learners like MLP and SVM. Their role was two-fold:

- predicting which base-learners were likely the best predictors based on factors like the day of the week, holiday on weekday, day of the year, hour, and lagged by 39-hours weather factors,
- providing ensemble predictions by multiplying the probabilities of base-learners with their respective forecasts.

10. Research Conclusion: Summary, Limitations, and Recommendations

Introduction and literature review were covered in Chapters 1-2, respectively. Chapters 3-6 covered the preparation phase. Experimentations with base-learners and meta-learners of ensembles were covered in Chapters 7-8, respectively, followed by validation of the results in Chapter 9.

In Chapter 1, accurate One Day-Ahead Demand Forecasting (ODADF) was found crucial for electrical network reliability, the environment, and trading markets. Problem with achieving accurate predictions by individual models, as well as possible solution, the use of ensembles, had been both identified. Finally, the research objectives were formed, to develop a framework of ensemble learning models, evaluate their performance, and examine their potential for ODADF in Ireland.

In Chapter 2, comprehensive overview of the state-of-the-art methodologies to Short-Term Load Forecasting (STLF), revealed gap in the knowledge which needs to be filled. No research, examining implementation of ensemble learning models, either to ODADF or STLF in Ireland, was found. Nevertheless, strengths and weaknesses of single and hybrid models, reported in papers, were invaluable in guiding the selection of base-learners and integration methods to develop ensemble learning framework. Furthermore, the significance of incorporating historical demand, calendar data, weather variables and their encoding, informed the development and implementation of the framework. Moreover, insights gleaned from real-world applications as well as challenges and limitations in the field, informed the experimental design for this research.

In Chapter 3, the methodology framework of research was developed, and CRISP-DM, adapted to requirements of the research, was selected as project management framework. Experimentation was selected as primary research methodology, and the population of interest, sampling method and type, as well as quantitative research approach were identified as appropriate. The development of ensembles' framework considered a balance between performance and computational complexity of the configurations. Three stacking approaches were considered, such as classifiers and regressors as meta-learners, and heuristic rules. A variety of machine and deep learning regressors for potential base-learners was considered. Moreover, two methods of supervised problem creation, based on Similar Day (SD) and Moving Window (MW) approaches, were proposed to increase the ability of base-learners to find various patterns in data. To evaluate the performance of ensembles by comparing their metrics, strict experimentation setting was established, where all models were trained and evaluated under the same conditions. To examine the potential of ensembles to ODADF in Ireland, ensembles and base-learners, were compared using various performance metrics as a function of day type, month and hour. Testing the best solutions on unseen data, and performing inferential statistical tests, were used to validate the results. Bayesian optimisation with 10-fold cross-validation was selected for hyperparameters tuning. Project management section covered detailed descriptions of all steps performed in the thesis. Finally, limitations and ethical considerations were described.

In Chapter 4, sources and selection of datasets were identified, and raw data for power demand and weather factors was collected. Initial Exploratory Data Exploration (EDA) detected Daylight-Saving Time (DST) distortions, missing values and outliers in time series. Extraction of day of the week, holidays occurring on weekdays, day of the year, and hour, enhanced the data by temporal features. Finally, temperature, relative humidity and wind speed were selected as exogenous variables, and data was trimmed to cover years 2014-2022.

In Chapter 5, power demand data preparation for modelling was performed, including DST distortion removal, and replacement of missing data and outliers. New data was validated by performance metrics, visual comparison to SDs from neighbouring weeks, and distributions before and after the cleaning process. EDA, including descriptive and inferential statistics, revealed patterns, trends and seasonality in power demand, which were seriously disrupted by Covid-19 lockdown restrictions in Ireland in 2020. Subsequently, baseline models were examined on data from year 2019, as starting point and reference for comparison with more complex models. Investigations into weekly and daily lags, and window size were performed for SD and MW approaches, respectively. Scaling of power demand, and encoding of temporal features to cyclical and sparse vector formats, were found valid and beneficial to ODADF in Ireland by correlation study. Finally, supervised learning problems were defined separately for SD and MW approaches.

In Chapter 6, multivariate data preparation for modelling was performed, including DST distortion removal from weather factors, and investigation into correlation between lagged weather variables and power demand. Temperature, relative humidity and wind speed, lagged by 39-hours were selected as potential weather features. As weather data was distributed locally, three approaches to find representative stations were proposed, such as virtual weather stations created by Linear and Lasso Regression, as well as the most important real one. Weather factors were scaled and results were validated by correlation study and distributions comparison. Investigation into feature importance was performed to select the best ones, considering all three representative weather stations, utilising various methods, separately for SD and MW approaches. Finally, supervised learning datasets, utilising selected features, were created for SD and MW approaches, employing weather factors created with Linear and Lasso Regression, respectively. Given that data from year 2020 was found to be indeed an outlier, datasets were split primarily into training and testing sets, covering years 2014-2019, and 2021-2022, respectively.

In Chapter 7, experimentation with wide range of potential base-learners was performed. Training datasets were further split into training and validation subsets, covering years 2014-2018, and 2019, respectively. Then, Bayesian optimisation with 10-fold cross-validation was used for base-learners hyperparameters tuning, and predictions were made for years 2015-2019. Potential base-learners were evaluated on year 2019, and the twenty most promising ones were selected as base-learners. Both, similar day-based and moving window-based models found their place in the selection. All base-learners showed fluctuations in their MAPE across different days of the week, months and hours. That variability was found potentially beneficial for ensemble learning models.

In Chapter 8, experimentation with three potential integration methods, identified in Chapter 3, was performed, incorporating predictions of twenty base-learners from Chapter 7 as training data for the ensembles. Heuristic rule-based ensembles were used as the baseline models. Bayesian optimisation with 10-fold cross-validation was used for hyperparameters tuning of classifiers and regressors as meta-learners, and predictions by ensembles were made for years 2015-2019. Potential ensembles were evaluated on year 2019, and classification-based ensemble learning models, as the winners of study, were selected for further investigation.

The experimentation phase established a cause-and-effect relationship between ensemble configurations and performance metrics of ODADF in Ireland. Firstly, limiting the number of base-learners adversely affected ensembles' performance, which highlighted the capability of ensembles to effectively handle and manage their input. Concurrently, while keeping the same number of base-learner predictions as an independent variable, changing the integration approach had significant influence on their overall performance. As a result, the integration method emerged as the primary causal variable. Additionally, the importance of the data preparation phase was on par with that of designing the ensemble architectures. The introduction of SD and MW approaches amplified the diversity of the base-learners' predictions. This enhanced diversity underscored the benefits of incorporating a varied range of base-learners, ultimately benefiting the performance of ensemble systems for ODADF in Ireland.

Ensemble learning models not only harnessed the combined strengths but also mitigated the potential inconsistencies found in individual base-learners. Even the lesser-performing ensembles not only approached the proficiency of the best base-learners but also outstripped the performance of the least effective base-learner. Additionally, incorporating virtual representative weather stations was found beneficial to performance of classification-based ensemble learning models. While ensembles using Logistic Regression and SVM as meta-learners performed better with weather factors from the virtual weather station created with Linear Regression, those ones, which utilized GBM and

MLP favoured the Lasso Regression. Interestingly, those two virtual weather stations were favoured over the most important real weather station in Mount Dillon both, by SD and MW base-learners, as well as by meta-learners of classification-based ensembles. That proves the benefit of their introduction to this research. Moreover, while half of the base-learners predictions deviated from the distribution of actual demand, predictions of classification-based and heuristic rule-based models aligned with the distribution of the actual demand. That underscored the value of leveraging a variety of models when constructing ensemble systems, which led to their more robust and accurate predictions.

In Chapter 9, validation of experimentation results was performed. Testing datasets were further split into training and validation subsets, covering years 2021 and 2022, respectively. Then, the twenty base-learners, with hyperparameters inferred from Chapter 7 were refitted and evaluated on unseen data from years 2021-2022, respectively. Subsequently, classification-based meta-learners, with hyperparameters inferred from Chapter 8 were refitted and evaluated on the base-learners' predictions and data for year 2022, respectively.

The results proved the high potential of classification-based ensembles for ODADF in Ireland. Ensembles of twenty base-learners, with SVM and MLP classifiers as meta-learners, achieved the lowest MAPE 1.91%, which was 11.2% improvement in comparison to the best base-learner, SVM (SD) registering MAPE 2.15%. Furthermore, while predictions of 65% base-learners deviated from the distribution of actual demand, predictions of above ensembles aligned with the distribution of the actual demand. Therefore, classification-based ensembles with SVM and MLP classifiers as meta-learners stood out as the most effective solution for ODADF, and their high potential was revealed, recognised and validated on unseen data.

All research objectives were fully addressed in the thesis. The framework of ensemble learning models for ODADF in Ireland was developed in Chapters 3, 7-8. Following experimentation with selected architectures of ensemble learning models, their performance was evaluated by comparing their metrics, and the cause-and-effect relationship between architecture of ensembles and ODADF performance was established in Chapters 7-8. Finally, the potential of classification-based ensemble learning application to ODADF in Ireland was revealed and validated in Chapters 8-9, respectively.

While this research addressed the gap in knowledge, identified in Chapter 2, demonstrating the potential of ensemble learning models for ODADF in Ireland, further work is needed to comprehensively bridge this gap. Given that the project was conducted with the use of personal computer and within twelve weeks period, limitations in architecture of the ensembles were recognised and accepted, with the aim to develop a balanced ensemble model, where potential benefit in accuracy was weighted against computational effort and complex model building. Therefore, DL was not incorporated for MW approach, and Bayesian optimisation of hyperparameters tuning was restricted to twenty trials. Besides, hybrid models were not considered as base-learners. Moreover, early-stopping was not integrated into the MLP, LSTM, and CNN models because it was found incompatible with Bayesian optimisation, causing early-stopped trials to fail. Nevertheless, the number of epochs was set as a hyperparameter to be optimised. Furthermore, while heuristic rule, classifiers and regressors from stacking integration methods were examined in this research, bagging and boosting ensembles were not explored due to time-constraints. Lastly, although feature selection was conducted separately for the SD and MW approaches using a variety of methods, the final decision was based on the performance of the LR model, and applied to all other models.

Future research could be performed to evaluate the potential of using wider variety of base-learners for stacking meta-learners, as well as other integration methods for ODADF in Ireland. Firstly, it would be recommended to evaluate DL models for MW approach, and increase the number of trials in Bayesian optimisation from twenty to at least one-hundred. Secondly, inclusion of hybrid models could enhance the variety of base-learners. Thirdly, integration of early-stopping into Bayesian optimisation could reduce the time needed for hyperparameters tuning. Lastly, conducting feature selection individually for each model could enhance the performance of both base-learners and ensemble learning models.

References:

Adams, K.A. and McGuire, E.K. (2022) Research Methods, Statistics, and Applications. SAGE Publications.

Adeva, J.J.G., Beresi, U.C., Calvo, R.A. (2005). Accuracy and Diversity in Ensembles of Text Categorisers. CLEI Journal. 8 (2): 1:1–1:12. Available at https://doi.org/10.19153/cleiej.8.2.1.

Ahmad, N., Ghadi, Y., Adnan, M. and Ali, M. (2022) 'Load Forecasting Techniques for Power System: Research Challenges and Survey', IEEE Access, 10, pp. 71054–71090. Available at: https://doi.org/ 10.1109/ACCESS.2022.3187839.

Al-Musaylh, M.S., Deo, R.C., Adamowski, J.F. and Li, Y. (2018) 'Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia', Advanced Engineering Informatics, 35, pp. 1–16. Available at: https://doi.org/10.1016/ j.aei.2017.11.002.

Altman, N.S. (1992) 'An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression', The American Statistician, 46(3), pp. 175–185. Available at: https://doi.org/10.1080/ 00031305.1992.10475879.

Bakirtzls, A.G., Petridls, V., Klartzis, S.J., Alexladls, M.C. and Malssls, A.H. (1996) 'A neural network short term load forecasting model for the Greek power system', IEEE Transactions on Power Systems, 11(2), pp. 858–863. Available at: https://doi.org/10.1109/59.496166.

Bandura, K., Hyndman, R.J. and Bergmeir C. (2021) MSTL: A Seasonal-Trend Decomposition Algorithm for Time Series with Multiple Seasonal Patterns. Available at: https://doi.org/10.48550/ arXiv.2107.13462.

Beggs, C. (2012) Energy: Management, Supply and Conservation. Routledge.

Bolzern, P., Fronza, G. and Brusasca, G. (1982) 'Temperature Effects on the Winter Daily Electric Load', Journal of Applied Meteorology (1962-1982), 21(2), pp. 241–243.

Breiman, L. (2001) 'Random Forests', Machine Learning, 45(1), pp. 5–32. Available at: https://doi.org/10.1023/ A:1010933404324.

Brockwell, P.J. and Davis, R.A. (eds) (2002) 'State-Space Models', in Introduction to Time Series and Forecasting. New York, NY: Springer New York (Springer Texts in Statistics), pp. 259–316. Available at: https://doi.org/10.1007/0-387-21657-X_8.

Brown, G., Wyatt, J., Harris, R. and Yao, X. (2005). Diversity creation methods: a survey and categorisation. Information Fusion, 6(1), pp.5-20, 2005. Available at: https://doi.org/10.1.1.421.349

Bryman, A. and Bell, E. (2015) Business Research Methods. Oxford University Press.

Burkov, A. (2019) The Hundred-Page Machine Learning Book. Scientific Research Publishing.

Ceperic, E., Ceperic, V. and Baric, A. (2013) 'A Strategy for Short-Term Load Forecasting by Support Vector Regression Machines', IEEE Transactions on Power Systems, 28(4), pp. 4356–4364. Available at: https://doi.org/10.1109/ TPWRS.2013.2269803.

Challu, C., Olivares, K. G., Oreshkin, B. N., Garza Ramirez, F., Mergenthaler Canseco, M. and Dubrawski, A. (2023) NHITS: Neural Hierarchical Interpolation for Time Series Forecasting. Proceedings of the AAAI Conference on Artificial Intelligence, 37(6), pp. 6989-6997. Available at: https://doi.org/ 10.1609/aaai.v37i6.25854.

Chatfield, C. (2015) 'The Analysis of Time Series Forecasting: An Introduction'. CRC Press Boca Raton.

Chen, C. and Liu, L.-M. (1993) 'Joint Estimation of Model Parameters and Outlier Effects in Time Series', Journal of the American Statistical Association, 88(421), pp. 284–297. Available at: https://doi.org/10.1080/01621459.1993.10594321.

Chollet, F. (2021) Deep Learning with Python, Second Edition. Simon and Schuster.

Czapaj, R., Kamiński, J. and Sołtysik, M. (2022) 'A Review of Auto-Regressive Methods Applications to Short-Term Demand Forecasting in Power Systems', Energies, 15(18), p. 6729. Available at: https://doi.org/10.3390/en15186729.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T.P., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. Available at: https://api.semanticscholar.org/ CorpusID:59777418. (Accessed 02/08/2023)

Day, T. (2006). Degree days: Theory and application. TCIoBSE, London, 106, 2006. Available at: https://www.cibse.org/knowledge-research/knowledge-portal/technical-memorandum-41-degree-days-theory-and-application-2006-pdf (Accessed: 13/05/2023).

De Livera, A.M., Hyndman, R.J., & Snyder, R. D. (2011), Forecasting time series with complex seasonal patterns using exponential smoothing, Journal of the American Statistical Association, 106(496), 1513-1527. Available at: https://doi.org/10.1198/jasa.2011.tm09771.

Dudek, G. (2015) 'Short-Term Load Forecasting Using Random Forests', in D. Filev, J. Jabłkowski, J. Kacprzyk, M. Krawczak, I. Popchev, L. Rutkowski, V. Sgurev, E. Sotirova, P. Szynkarczyk, and S. Zadrozny (eds) Intelligent Systems'2014. Cham: Springer International Publishing (Advances in Intelligent Systems and Computing), pp. 821–828. Available at: https://doi.org/10.1007/978-3-319-11310-4_71.

Dudek, G. (2016) 'Pattern-based local linear regression models for short-term load forecasting', Electric Power Systems Research, 130, pp. 139–147. Available at: https://doi.org/10.1016/j.epsr.2015.09.001.

EirGrid Group (2022) All-Island Generation Capacity Statement 2022-2031, EirGrid Group. Available at: https:// www.eirgridgroup.com/library/ (Accessed: 23 April 2023).

EirGrid Group (2023) Explore the Smart Grid Dashboard, Smart Grid Dashboard. Available at: https:// smartgriddashboard.com/ (Accessed: 12 May 2023).

Elman, J.L. (1990) 'Finding Structure in Time', Cognitive Science, 14(2), pp. 179–211. Available at: https://doi.org/ 10.1207/s15516709cog1402_1.

Enders, W. (2014) Applied Econometric Time Series, 4th Edition | Wiley, Wiley.com. Available at: https://www.wiley.com/en-us/Applied+Econometric+Time+Series%2C+4th+Edition-p-9781118808566 (Accessed: 22/07/2023).

Fan, S., Chen, L. and Lee, W.-J. (2009) 'Short-Term Load Forecasting Using Comprehensive Combination Based on Multimeteorological Information', IEEE Transactions on Industry Applications, 45(4), pp. 1460–1466. Available at: https://doi.org/10.1109/TIA.2009.2023571.

Fan, S. and Hyndman, R.J. (2012) 'Short-Term Load Forecasting Based on a Semi-Parametric Additive Model', IEEE Transactions on Power Systems, 27(1), pp. 134–141. Available at: https://doi.org/10.1109/TPWRS.2011.2162082.

Farrokhabadi, M., Browell, J., Wang, Y., Makonin, S., Su, W. and Zareipour, H. (2022) 'Day-Ahead Electricity Demand Forecasting Competition: Post-COVID Paradigm', IEEE Open Access Journal of Power and Energy, 9, pp. 185–191. Available at: https://doi.org/10.1109/OAJPE.2022.3161101.

Fidalgo, J.N. and Matos, M.A. (2007) 'Forecasting Portugal Global Load with Artificial Neural Networks', in J.M. De Sá, L.A. Alexandre, W. Duch, and D. Mandic (eds) Artificial Neural Networks – ICANN 2007. Berlin, Heidelberg: Springer Berlin Heidelberg (Lecture Notes in Computer Science), pp. 728–737. Available at: https://doi.org/10.1007/978-3-540-74695-9_75.

Foster, J. (2020) Electric load forecasting with increased embedded renewable generation. Ph.D. Queen's University Belfast. Available at: https://pure.qub.ac.uk/en/theses/electric-load-forecasting-with-increased-embedded-renewable-generation(437a7dd7-20f3-44bc-884a-7f8aae42c65b).html (Accessed: 22/07/2023).

Friedman, J.H. (2001) 'Greedy Function Approximation: A Gradient Boosting Machine', The Annals of Statistics, 29(5), pp. 1189–1232.

FSF (2023). Free Software Foundation. Available at: https://fsf.org. (Accessed: 05/07/2023)

Géron, A. (2019) Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, Inc.

Goodfellow, I., Bengio, Y. and Courville, A. (2016) Deep Learning. MIT Press.

Grmanová, G., Laurinec, P., Rozinajová, V., Ezzeddine, A.B., Lucká, M., Lacko, P., Vrablecová, P. and Návrat, P. (2016) 'Incremental Ensemble Learning for Electricity Load Forecasting', Acta Polytechnica Hungarica, 13(2). Available at: https://doi.org/10.12700/APH.13.2.2016.2.6.

Han, L., Peng, Y., Li, Y., Yong, B., Zhou, Q. and Shu, L. (2019) 'Enhanced Deep Networks for Short-Term and Medium-Term Load Forecasting', IEEE Access, 7, pp. 4045–4055. Available at: https://doi.org/10.1109/ACCESS.2018.2888978.

Haykin, S.O. (2010) Neural Networks and Learning Machines: International Edition. Pearson Education, Limited.

He, W. (2017) 'Load Forecasting via Deep Neural Networks', Procedia Computer Science, 122, pp. 308–314. Available at: https://doi.org/10.1016/j.procs.2017.11.374.

Hinman, J. and Hickey, E. (2009) 'MODELING AND FORECASTING SHORT-TERM ELECTRICITY LOAD USING REGRESSION ANALYSIS', Journal of Institute for Regulatory Policy Studies [Preprint]. Available at: https://irps.illinoisstate.edu/downloads/research/documents/LoadForecastingHinman-HickeyFall2009.pdf.

Hochreiter, S. and Schmidhuber, J. (1997) 'Long Short-Term Memory', Neural Computation, 9(8), pp. 1735–1780. Available at: https://doi.org/10.1162/neco.1997.9.8.1735.

Hong, T. and Fan, S. (2016) 'Probabilistic electric load forecasting: A tutorial review', International Journal of Forecasting, 32(3), pp. 914–938. Available at: https://doi.org/10.1016/j.ijforecast.2015.11.011.

Hong, T., Pinson, P. and Fan, S. (2014) 'Global Energy Forecasting Competition 2012', International Journal of Forecasting, 30(2), pp. 357–363. Available at: https://doi.org/10.1016/j.ijforecast.2013.07.001.

Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A. and Hyndman, R.J. (2016) 'Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond', International Journal of Forecasting, 32(3), pp. 896–913. Available at: https://doi.org/10.1016/j.ijforecast.2016.02.001.

Hong, T., Pinson, P., Wang, Y., Weron, R., Yang, D. and Zareipour, H. (2020) 'Energy Forecasting: A Review and Outlook', IEEE Open Access Journal of Power and Energy, 7, pp. 376–388. Available at: https://doi.org/10.1109/OAJPE.2020.3029979.

Hong, T. and Shahidehpour, M. (2015) 'Load Forecasting Case Study', EISPC and NARUC, US, Department of Energy [Preprint]. Available at: https://pubs.naruc.org/pub.cfm?id=536E10A7-2354-D714-5191-A8AAFE45D626.

Hong, T., Xie, J. and Black, J. (2019) 'Global energy forecasting competition 2017: Hierarchical probabilistic load forecasting', International Journal of Forecasting, 35(4), pp. 1389–1399. Available at: https://doi.org/10.1016/j.ijforecast.2019.02.006.

Hyndman, R.J. (2020) 'A brief history of forecasting competitions', International Journal of Forecasting, 36(1), pp. 7–14. Available at: https://doi.org/10.1016/j.ijforecast.2019.03.015.

Hyndman, R.J. and Athanasopoulos, G. (2018) Forecasting: principles and practice. OTexts.

Hyndman, R.J. and Koehler, A.B. (2006) 'Another look at measures of forecast accuracy', International Journal of Forecasting, 22(4), pp. 679–688. Available at: https://doi.org/10.1016/j.ijforecast.2006.03.001.

Jiao, R., Zhang, T., Jiang, Y. and He, H. (2018) 'Short-Term Non-Residential Load Forecasting Based on Multiple Sequences LSTM Recurrent Neural Network', IEEE Access, 6, pp. 59438–59448. Available at: https://doi.org/10.1109/ACCESS.2018.2873712.

Joseph, M. (2022) Modern Time Series Forecasting with Python: Explore industry-ready time series forecasting using modern machine learning and deep learning. Packt Publishing Ltd.

Kathirgamanathan, A., Patel, A., Khwaja, A.S., Venkatesh, B. and Anpalagan, A. (2022) 'Performance comparison of single and ensemble CNN, LSTM and traditional ANN models for short-term electricity load forecasting', The Journal of Engineering, 2022(5), pp. 550–565. Available at: https://doi.org/10.1049/tje2.12132.

Khwaja, A.S., Naeem, M., Anpalagan, A., Venetsanopoulos, A. and Venkatesh, B. (2015) 'Improved short-term load forecasting using bagged neural networks', Electric Power Systems Research, 125, pp. 109–115. Available at: https://doi.org/10.1016/j.epsr.2015.03.027.

Kuncheva, L. and Whitaker, C. (2003). Measures of diversity in classifier ensembles. Machine Learning, 51, pp. 181-207. Available from: https://link.springer.com/content/pdf/10.1023/A:1022859003006.pdf (Accessed: 18/08/2023)

LeCun, Y., Bengio, Y. and Hinton, G. (2015) 'Deep learning', Nature, 521(7553), pp. 436–444. Available at: https://doi.org/10.1038/nature14539.

Lu, C.-N., Wu, H.-T. and Vemuri, S. (1993) 'Neural network based short term load forecasting', IEEE Transactions on Power Systems, 8(1), pp. 336–342. Available at: https://doi.org/10.1109/59.221223.

Met Éireann (2023) Historical Data - Met Éireann - The Irish Meteorological Service. Available at: https://www.met.ie/climate/available-data/historical-data (Accessed: 12 May 2023).

Natekin, A. and Knoll, A. (2013) 'Gradient boosting machines, a tutorial', Frontiers in Neurorobotics, 7. Available at: https://doi.org/10.3389/fnbot.2013.00021.

Negro, A. (2021). Graph-powered machine learning. Shelter Island: Manning Publications Co.

Obst, D., De Vilmarest, J. and Goude, Y. (2021) 'Adaptive Methods for Short-Term Electricity Load Forecasting During COVID-19 Lockdown in France', IEEE Transactions on Power Systems, 36(5), pp. 4754–4763. Available at: https://doi.org/10.1109/TPWRS.2021.3067551.

Ord, J.K., Koehler, A.B. and Snyder, R.D. (1997) 'Estimation and Prediction for a Class of Dynamic Nonlinear Statistical Models', Journal of the American Statistical Association, 92(440), pp. 1621–1629. Available at: https://doi.org/10.1080/01621459.1997.10473684.

Oreshkin, B.N., Carpov, D. Chapados, N. and Bengio, Y. (2020). N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. ICLR 2020 Conference Paper. Available at: https://doi.org/10.48550/arXiv.1905.10437.

Osowski, S., Szmurlo, R., Siwek, K. and Ciechulski, T. (2022) 'Neural Approaches to Short-Time Load Forecasting in Power Systems—A Comparative Study', Energies, 15(9), p. 3265. Available at: https://doi.org/10.3390/en15093265.

Pardo, A., Meneu, V. and Valor, E. (2002) 'Temperature and seasonality influences on Spanish electricity load', Energy Economics, 24(1), pp. 55–70. Available at: https://doi.org/10.1016/S0140-9883(01)00082-2.

Park, D.C., El-Sharkawi, M.A., Marks, R.J., Atlas, L.E. and Damborg, M.J. (1991) 'Electric load forecasting using an artificial neural network', IEEE Transactions on Power Systems, 6(2), pp. 442–449. Available at: https://doi.org/10.1109/59.76685.

Roshan, J.V. (2022). Optimal ratio for data splitting. Statistical analysis and data mining. The ASA Data Science Journal, 15(4). Available at: https://doi.org/10.1002/sam.11583.

Quinlan, J.R. (2014) C4.5: Programs for Machine Learning. Elsevier.

Resnik, D.B. (2005) The Ethics of Science: An Introduction. Routledge.

Resnik, D.B. (2018) The Ethics of Research with Human Subjects: Protecting People, Advancing Science, Promoting Trust. Springer.

Saltz, J.S., Shamshurin, I. and Crowson, K. (2017). Comparing data science project management methodologies via a controlled experiment. Proceedings of the 50th Hawaii International Conference on System Sciences. Available at: https://doi.org/10125/41273.

Saunders, M., Lewis, P. and Thornhill, A. (2015) Research Methods for Business Students. Pearson Education.

Scheidt, F.V., Medinová, H., Ludwig, N., Richter, B., Staudt, P. and Weinhardt, C. (2020) 'Data analytics in the electricity sector – A quantitative and qualitative literature review', Energy and AI, 1, p. 100009. Available at: https://doi.org/10.1016/j.egyai.2020.100009.

Sethi, R. and Kleissl, J. (2020) 'Comparison of Short-Term Load Forecasting Techniques', in 2020 IEEE Conference on Technologies for Sustainability (SusTech). 2020 IEEE Conference on Technologies for Sustainability (SusTech), Santa Ana, CA, USA: IEEE, pp. 1–6. Available at: https://doi.org/10.1109/SusTech47890.2020.9150490.

Shahare, K., Mitra, A., Naware, D., Keshri, R. and Suryawanshi, H.M. (2023) 'Performance analysis and comparison of various techniques for short-term load forecasting', Energy Reports, 9, pp. 799–808. Available at: https://doi.org/10.1016/j.egyr.2022.11.086.

Shamoo, A.E. and Resnik, D.B. (2009) Responsible Conduct of Research. Oxford University Press.

Sollich, P. and Krogh, A. (1996). Learning with ensembles: How overfitting can be useful. Advances in Neural Information Processing Systems, volume 8, pp. 190-196, 1996. Available at: https://proceedings.neurips.cc/paper/1995/file/1019c8091693ef5c5f55970346633f92-Paper.pdf (Accessed 18/08/2023).

Statsmodels (2023). Stationarity and differencing (ADF/KPSS). Available at: https://www.statsmodels.org/stable/examples/notebook/generated/stationarity_detrending_adf_kpss.html (Accessed 06 July 2023)

Shumway, R.H. and Stoffer, D.S. (2017) Time Series Analysis and Its Applications: With R Examples. Cham: Springer International Publishing (Springer Texts in Statistics). Available at: https://doi.org/10.1007/978-3-319-52452-8.

Song, K.-B., Ha, S.-K., Park, J.-W., Kweon, D.-J. and Kim, K.-H. (2006) 'Hybrid Load Forecasting Method With Analysis of Temperature Sensitivities', IEEE Transactions on Power Systems, 21(2), pp. 869–876. Available at: https://doi.org/10.1109/TPWRS.2006.873099.

Steadman, R.G. (1984) 'A Universal Scale of Apparent Temperature', Journal of Climate and Applied Meteorology, 23(12), pp. 1674–1687. Available at: https://doi.org/10.1175/1520-0450(1984)023<1674:AUSOAT>2.0.CO;2.

Tao Hong, Min Gui, Baran, M.E. and Willis, H.L. (2010) 'Modeling and forecasting hourly electric load by multiple linear regression with interactions', in IEEE PES General Meeting. Energy Society General Meeting, Minneapolis, MN: IEEE, pp. 1–8. Available at: https://doi.org/10.1109/PES.2010.5589959.

Tashman, L.J. (2000) 'Out-of-sample tests of forecasting accuracy: an analysis and review', International Journal of Forecasting, 16(4), pp. 437–450. Available at: https://doi.org/10.1016/S0169-2070(00)00065-0.

Taylor, J.W. and Buizza, R. (2003) 'Using weather ensemble predictions in electricity demand forecasting', International Journal of Forecasting, 19(1), pp. 57–70. Available at: https://doi.org/10.1016/S0169-2070(01)00123-6.

Taylor, J.W. and McSharry, P.E. (2007) 'Short-Term Load Forecasting Methods: An Evaluation Based on European Data', IEEE Transactions on Power Systems, 22(4), pp. 2213–2219. Available at: https://doi.org/10.1109/TPWRS.2007.907583.

Tsekouras, G.J., Kanellos, F.D. and Mastorakis, N. (2015) 'Short Term Load Forecasting in Electric Power Systems with Artificial Neural Networks', in Nikos Mastorakis, A. Bulucea, and G. Tsekouras (eds) Computational Problems in Science and Engineering. Cham: Springer International Publishing (Lecture Notes in Electrical Engineering), pp. 19–58. Available at: https://doi.org/10.1007/978-3-319-15765-8_2.

Vapnik, V. (2013) The Nature of Statistical Learning Theory. Springer Science & Business Media.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I. (2017) 'Attention is All you Need', Advances in Neural Information Processing Systems, 30. Available at: https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html (Accessed: 23 April 2023).

Vellido, A., Martín-Guerrero, J.D. and Lisboa, P.J. (2012) 'Making machine learning models interpretable.', in ESANN. Citeseer, pp. 163–172.

Vorhies, W. (2016). CRISP-DM - a standard methodology to ensure a good outcome. Available at: https://www.datasciencecentral.com/crisp-dm-a-standard-methodology-to-ensure-a-good-outcome. (Accessed 02/09/2023)

Wang, P., Liu, B. and Hong, T. (2016) 'Electric load forecasting with recency effect: A big data approach', International Journal of Forecasting, 32(3), pp. 585–597. Available at: https://doi.org/10.1016/j.ijforecast.2015.09.006.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C. and Yu, P.S. (2021) 'A Comprehensive Survey on Graph Neural Networks', IEEE Transactions on Neural Networks and Learning Systems, 32(1), pp. 4–24. Available at: https://doi.org/10.1109/TNNLS.2020.2978386.

Xie, J., Chen, Y., Hong, T. and Laing, T.D. (2018) 'Relative Humidity for Load Forecasting Models', IEEE Transactions on Smart Grid, 9(1), pp. 191–198. Available at: https://doi.org/10.1109/TSG.2016.2547964.

Yu, Z., Niu, Z., Tang, W. and Wu, Q. (2019) ‘Deep Learning for Daily Peak Load Forecasting–A Novel Gated Recurrent Neural Network Combining Dynamic Time Warping’, IEEE Access, 7, pp. 17184–17194. Available at: https://doi.org/10.1109/ACCESS.2019.2895604.

Zhang, G., Patuwo, E.B. and Hu, Y.M. (1998) ‘Forecasting with artificial neural networks: The state of the art’, International Journal of Forecasting, 14(1), pp. 35–62. Available at: https://doi.org/10.1016/S0169-2070(97)00044-7.

Zhang, G.P. (2003) ‘Time series forecasting using a hybrid ARIMA and neural network model’, Neurocomputing, 50, pp. 159–175. Available at: https://doi.org/10.1016/S0925-2312(01)00702-0.

Zhang, Q., Yang, L.T., Chen, Z. and Li, P. (2018) ‘A survey on deep learning for big data’, Information Fusion, 42, pp. 146–157. Available at: https://doi.org/10.1016/j.inffus.2017.10.006.