Appendix
Monash Time Series Forecasting Archive

Rakshitha Godahewa
Monash University
Melbourne, Australia
rakshitha.godahewa@monash.edu

Christoph Bergmeir
Monash University
Melbourne, Australia
christoph.bergmeir@monash.edu

Geoffrey I. Webb
Monash University
Melbourne, Australia
geoff.webb@monash.edu

Rob J. Hyndman
Monash University
Melbourne, Australia
rob.hyndman@monash.edu

Pablo Montero-Manso
University of Sydney
Australia
pmontm@gmail.com

A Data records

Our archive contains 26 time series datasets. Out of them, 20 datasets contain multiple related time series to facilitate the evaluation of global time series forecasting models (Section A.6). The remaining 6 datasets contain single very long time series (Section A.7). As a large amount of data oftentimes renders machine learning methods feasible compared with traditional statistical modelling, and we are not aware of good and systematic benchmark data in this space either, these series are included in our repository as well.

A.1 Data collection procedure

All these datasets were already publicly available in different platforms with different data formats. Hence, we did not have to manually collect the observations. The original sources of all datasets are mentioned in the datasets descriptions (Sections A.6 and A.7). Out of the 26 datasets, 8 originate from competition platforms, 3 from a prior research conducted by Lai et al. [1], 6 are taken from R packages, 1 is from the Kaggle platform [2], and 1 is taken from a Johns Hopkins repository [3]. The other datasets have been extracted from corresponding domain specific platforms.

After extracting datasets, we analysed them individually to identify the datasets containing series with different frequencies and missing observations. Nine datasets contain time series belonging to different frequencies and the archive contains a separate dataset per each frequency. Seven of the datasets have series with missing values. The archive contains 2 versions of each of these, one with and one without missing values. In the latter case, the missing values have been replaced by using an appropriate imputation technique as explained in Sections A.6 and A.7. Finally, we obtain 50 datasets with the above explained different versions.

The 50 datasets are then converted into .tsf format which is a new format we introduce to store time series datasets as explained in the Section 2.1 of our original manuscript. The .tsf files are zipped and uploaded into our datasets archive available at `https://zenodo.org/communities/forecasting` where the other researchers can directly download them for further research use. Code to load the datasets in .tsf format into R and Python is available in our github repository available at `https://github.com/rakshitha123/TSForecasting`.

Submitted to the 35th Conference on Neural Information Processing Systems (NeurIPS 2021) Track on Datasets and Benchmarks. Do not distribute.

A.2 Intended use of datasets

All datasets in our repository are intended to use only for research purposes and to evaluate the performance of new forecasting algorithms.

A.3 Hosting, licensing, maintenance and preservation

All datasets are permanently available, and they are hosted and maintained at `https://zenodo.org/communities/forecasting` where the researchers can directly download the datasets. All datasets are under Creative Commons Attribution 4.0 International license where the users can modify, distribute and use the datasets as long as they credit the authors for the downloaded datasets from the repository.

Furthermore, a summary of datasets, links to download all datasets, the extracted features of each time series of all datasets, benchmark results of the datasets across 8 error metrics and links to all implementations related to the archive are hosted and maintained in our website at `https://forecastingdata.org/`.

We also encourage other researchers to contribute time series datasets to our repository either by directly uploading them into the archive and/or by contacting the authors via email.

A.4 Code availability and reproducibility of results

All implementations related to the forecasting archive including code for loading the datasets in our .tsf format into the R and Python environments, code for feature calculations and evaluation of baseline forecasting models are publicly available at `https://github.com/rakshitha123/TSForecasting`.

We ensure that both feature analysis and benchmark results are reproducible. The instructions on executing the related experiments are available in our github repository. The implementations are there in both R and Python programming languages and they do not require any paid dependencies.

A.5 Author statement

We, all authors bear all responsibility in case of violation of datasets rights. We have checked the licensing of all datasets and have only uploaded the publicly shareable datasets to our repository. We have also mentioned the original sources of all datasets in our website, `https://forecastingdata.org/`.

If there are any copyright issues of the datasets, please contact the authors via email.

The next sections explain the datasets in our repository in detail.

A.6 Time series datasets

This section describes the benchmark datasets that have a sufficient number of series from a particular frequency. The datasets may contain different categories in terms of domain and frequency.

A.6.1 M1 dataset

The M1 competition dataset [4] contains 1001 time series with 3 different frequencies: yearly, quarterly, and monthly as shown in Table 1. The series belong to 7 different domains: macro 1, macro 2, micro 1, micro 2, micro 3, industry, and demographic.

Research work which uses this dataset includes:

- Forecasting with artificial neural networks: the state of the art [5]
- Time series forecasting using a hybrid ARIMA and neural network model [6]
- Automatic time series forecasting: the forecast package for R [7]
- Exponential Smoothing: the state of the art [8]

Table 1: Summary of M1 dataset

Frequency	No: of Series	Min. Length	Max. Length	Forecast Horizon
Yearly	181	15	58	6
Quarterly	203	18	114	8
Monthly	617	48	150	18
Total	1001			

- Neural network forecasting for seasonal and trend time series [9]

The DOI links to access and download the datasets are as follows:

- Yearly dataset: `http://doi.org/10.5281/zenodo.4656193`
- Quarterly dataset: `http://doi.org/10.5281/zenodo.4656154`
- Monthly dataset: `http://doi.org/10.5281/zenodo.4656159`

A.6.2 M3 dataset

The M3 competition dataset [10] contains 3003 time series of various frequencies including yearly, quarterly, and monthly, as shown in Table 2. The series belong to 6 different domains: demographic, micro, macro, industry, finance, and other.

Table 2: Summary of M3 dataset

Frequency	No: of Series	Min. Length	Max. Length	Forecast Horizon
Yearly	645	20	47	6
Quarterly	756	24	72	8
Monthly	1428	66	144	18
Other	174	71	104	8
Total	3003			

Research work which uses this dataset includes:

- The theta model: a decomposition approach to forecasting [11]
- Recurrent neural networks for time series forecasting: current status and future directions [12]
- Ensembles of localised models for time series forecasting [13]
- Out-of-sample tests of forecasting accuracy: an analysis and review [14]
- Metrics for evaluating performance of prognostic techniques [15]
- Temporal link prediction using matrix and tensor factorizations [16]
- Forecasting time series with complex seasonal patterns using exponential smoothing [17]
- Evaluating forecasting methods [18]
- Exponential smoothing with a damped multiplicative trend [19]

The DOI links to access and download the datasets are as follows:

- Yearly dataset: `http://doi.org/10.5281/zenodo.4656222`
- Quarterly dataset: `http://doi.org/10.5281/zenodo.4656262`
- Monthly dataset: `http://doi.org/10.5281/zenodo.4656298`
- Other dataset: `http://doi.org/10.5281/zenodo.4656335`

Table 3: Summary of M4 dataset

Frequency	No: of Series	Min. Length	Max. Length	Forecast Horizon
Yearly	23000	19	841	6
Quarterly	24000	24	874	8
Monthly	48000	60	2812	18
Weekly	359	93	2610	13
Daily	4227	107	9933	14
Hourly	414	748	1008	48
Total	100000			

A.6.3 M4 dataset

The M4 competition dataset [20, 21] contains 100,000 time series with 6 different frequencies: yearly, quarterly, monthly, weekly, daily, and hourly, as shown in Table 3. The series belong to 6 different domains: demographic, micro, macro, industry, finance, and other, similar to the M3 forecasting competition. This dataset contains a subset of series available at ForeDeCk [22].

Research work which uses this dataset includes:

- A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting [23]
- FFORMA: Feature-based Forecast Model Averaging [24]
- Ensembles of localised models for time series forecasting [13]
- Recurrent neural networks for time series forecasting: current status and future directions [12]
- LSTM-MSNet: leveraging forecasts on sets of related time series with multiple seasonal patterns [25]
- Are forecasting competitions data representative of the reality? [26]
- Averaging probability forecasts: back to the future [27]
- A strong baseline for weekly time series forecasting [28]

The DOI links to access and download the datasets are as follows:

- Yearly dataset: `http://doi.org/10.5281/zenodo.4656379`
- Quarterly dataset: `http://doi.org/10.5281/zenodo.4656410`
- Monthly dataset: `http://doi.org/10.5281/zenodo.4656480`
- Weekly dataset: `http://doi.org/10.5281/zenodo.4656522`
- Daily dataset: `http://doi.org/10.5281/zenodo.4656548`
- Hourly dataset: `http://doi.org/10.5281/zenodo.4656589`

A.6.4 Tourism dataset

This dataset originates from a Kaggle competition [29, 30] and contains 1311 tourism related time series with 3 different frequencies: yearly, quarterly, and monthly as shown in Table 4.

Table 4: Summary of tourism dataset

Frequency	No: of Series	Min. Length	Max. Length	Forecast Horizon
Yearly	518	11	47	4
Quarterly	427	30	130	8
Monthly	366	91	333	24
Total	1311			

Research work which uses this dataset includes:

- Recurrent neural networks for time series forecasting: current status and future directions [12]
- A meta-analysis of international tourism demand forecasting and implications for practice [31]
- Improving forecasting by estimating time series structural components across multiple frequencies [32]
- Forecasting tourist arrivals using time-varying parameter structural time series models [33]
- Forecasting monthly and quarterly time series using STL decomposition [34]
- A novel approach to model selection in tourism demand modeling [35]

The DOI links to access and download the datasets are as follows:

- Yearly dataset: `http://doi.org/10.5281/zenodo.4656103`
- Quarterly dataset: `http://doi.org/10.5281/zenodo.4656093`
- Monthly dataset: `http://doi.org/10.5281/zenodo.4656096`

A.6.5 NN5 dataset

This dataset contains 111 time series of daily cash withdrawals from Automated Teller Machines (ATM) in the UK, and was used in the NN5 forecasting competition [36]. The forecast horizon considered in the competition was 56. The original dataset contains missing values. Our repository contains two versions of the dataset: the original version with missing values and a modified version where the missing values have been replaced using a median substitution where a missing value on a particular day is replaced by the median across all the same days of the week along the whole series as in Hewamalage et al. [12]. Furthermore, Godahewa et al. [28] use the weekly aggregated version of this dataset for their experiments related to proposing a baseline model for weekly forecasting. The aggregated weekly version of this dataset is also available in our repository. Research work which uses this dataset includes:

- Recurrent neural networks for time series forecasting: current status and future directions [12]
- A strong baseline for weekly time series forecasting [28]
- Forecasting across time series databases using recurrent neural networks on groups of similar series: a clustering approach [37]
- Forecast combinations of computational intelligence and linear models for the NN5 time series forecasting competition [38]
- Forecasting the NN5 time series with hybrid models [39]
- Multiple-output modeling for multi-step-ahead time series forecasting [40]
- Recursive multi-step time Series forecasting by perturbing data [41]
- Benchmarking of classical and machine-learning algorithms (with special emphasis on bagging and boosting approaches) for time series forecasting [42]

The DOI links to access and download the datasets are as follows:

- Daily dataset with missing values: `http://doi.org/10.5281/zenodo.4656110`
- Daily dataset without missing values: `http://doi.org/10.5281/zenodo.4656117`
- Weekly dataset: `http://doi.org/10.5281/zenodo.4656125`

A.6.6 CIF 2016 dataset

The dataset from the Computational Intelligence in Forecasting (CIF) 2016 forecasting competition contains 72 monthly time series. Out of those, 24 series originate from the banking sector, and the remaining 48 series are artificially generated. There are 2 forecast horizons considered in the competition where 57 series have a forecasting horizon of 12 and the remaining 15 series consider the forecast horizon as 6 [43]. Research work which uses this dataset includes:

- Recurrent neural networks for time series forecasting: current status and future directions [12]
- Ensembles of localised models for time series forecasting [13]
- Forecasting across time series databases using recurrent neural networks on groups of similar series: a clustering approach [37]
- Improving time series forecasting: an approach combining bootstrap aggregation, clusters and exponential smoothing [44]
- Time series clustering using numerical and fuzzy representations [45]
- An automatic calibration framework applied on a metaheuristic fuzzy model for the CIF competition [46]

The DOI link to access and download the dataset is `http://doi.org/10.5281/zenodo.4656042`.

A.6.7 Kaggle web traffic dataset

This dataset contains 145063 daily time series representing the number of hits or web traffic for a set of Wikipedia pages from 01/07/2015 to 10/09/2017 used by the Kaggle web traffic forecasting competition [47]. The forecast horizon considered in the competition was 59. As the original dataset contains missing values, we include both the original dataset in our repository and an imputed version. This dataset is intermittent and hence, we impute missing values with zeros. Furthermore, Godahewa et al. [28] use the weekly aggregated version of this dataset containing the first 1000 series. Our repository also contains this aggregated weekly version of the dataset for all series. The missing values of the original dataset were imputed before the aggregation. Research work which uses this dataset includes:

- Recurrent neural networks for time series forecasting: current status and future directions [12]
- Ensembles of localised models for time series forecasting [13]
- A strong baseline for weekly time series forecasting [28]
- Web traffic prediction of Wikipedia pages [48]
- Improving time series forecasting using mathematical and deep learning models [49]
- Foundations of sequence-to-sequence modeling for time series [50]

The DOI links to access and download the datasets are as follows:

- Daily dataset with missing values: `http://doi.org/10.5281/zenodo.4656080`
- Daily dataset without missing values: `http://doi.org/10.5281/zenodo.4656075`
- Weekly dataset: `http://doi.org/10.5281/zenodo.4656664`

A.6.8 Solar dataset

This dataset contains 137 time series representing the solar power production recorded every 10 minutes in the state of Alabama in 2006. It was used by Lai et al. [1], and originally extracted from Solar [51]. Furthermore, Godahewa et al. [28] use an aggregated version of this dataset containing weekly solar power production records. The aggregated weekly version of this dataset is also available in our repository.

The DOI links to access and download the datasets are as follows:

- 10 minutes dataset: `http://doi.org/10.5281/zenodo.4656144`
- Weekly dataset: `http://doi.org/10.5281/zenodo.4656151`

A.6.9 Electricity dataset

This dataset represents the hourly electricity consumption of 321 clients from 2012 to 2014 in kilowatt (kW). It was used by Lai et al. [1], and originally extracted from UCI [52]. Our repository also contains an aggregated version of this dataset representing the weekly electricity consumption values.

The DOI links to access and download the datasets are as follows:

- Hourly dataset: http://doi.org/10.5281/zenodo.4656140
- Weekly dataset: http://doi.org/10.5281/zenodo.4656141

A.6.10 London smart meters dataset

This dataset contains 5560 half-hourly time series that represent the energy consumption readings of London households in kWh from November 2011 to February 2014 [53]. The series are categorized into 112 blocks in the original dataset. The series in our repository are in the same order (from block 0 to block 111) as they are in the original dataset. The original dataset contains missing values and we impute them using the last observation carried forward (LOCF) method. Our repository contains both versions: the original version with missing values and the modified version where the missing values have been replaced. Research work which uses this dataset includes:

- Predicting electricity consumption using deep recurrent neural networks [54]
- A single scalable LSTM model for short-term forecasting of disaggregated electricity loads [55]
- Deep learning based short-term load forecasting for urban areas [56]
- Smart grid energy management using RNN-LSTM: a deep learning-based approach [57]

The DOI links to access and download the datasets are as follows:

- Dataset with missing values: http://doi.org/10.5281/zenodo.4656072
- Dataset without missing values: http://doi.org/10.5281/zenodo.4656091

A.6.11 Wind farms dataset

This dataset contains very long minutely time series representing the wind power production of 339 wind farms in Australia. It was extracted from the Australian Energy Market Operator (AEMO) online platform [58]. The series in this dataset range from 01/08/2019 to 31/07/2020. The original dataset contains missing values where some series contain missing data for more than seven consecutive days. Our repository contains both the original version of the dataset and a version where the missing values have been replaced by zeros.

The DOI links to access and download the datasets are as follows:

- Dataset with missing values: http://doi.org/10.5281/zenodo.4654909
- Dataset without missing values: http://doi.org/10.5281/zenodo.4654858

A.6.12 Car parts dataset

This dataset contains 2674 intermittent monthly time series showing car parts sales from January 1998 to March 2002. It was extracted from the R package expsmooth [59]. The package contains this dataset as *“carparts"*. As the original dataset contains missing values, we include the original version of the dataset in the repository as well as a version where the missing values have been replaced with zeros, as the series are intermittent. Research work which uses this dataset includes:

- Principles and algorithms for forecasting groups of time series: locality and globality [60]

The DOI links to access and download the datasets are as follows:

- Dataset with missing values: http://doi.org/10.5281/zenodo.4656022
- Dataset without missing values: http://doi.org/10.5281/zenodo.4656021

A.6.13 Dominick dataset

This dataset contains 115704 weekly time series representing the profit of individual stock keeping units (SKU) from a retailer.

It was extracted from the Kilts Center, University of Chicago Booth School of Business online platform [61]. This platform also contains daily store-level sales data on more than 3500 products collected from Dominick's Finer Foods, a large American retail chain in the Chicago area, for approximately 9 years. The data are provided in different categories such as customer counts, store-specific demographics and sales products. Research work which uses this dataset includes:

- Principles and algorithms for forecasting groups of time series: locality and globality [60]
- The value of competitive information in forecasting FMCG retail product sales and the variable selection problem [62]
- Beer snobs do exist: estimation of beer demand by type [63]
- Downsizing and supersizing: how changes in product attributes influence consumer preferences [64]
- Reference prices, costs, and nominal rigidities [65]
- Sales and monetary policy [66]

The DOI link to access and download the dataset is `http://doi.org/10.5281/zenodo.4654802`.

A.6.14 FRED-MD dataset

This dataset contains 107 monthly time series showing a set of macro-economic indicators from the Federal Reserve Bank [67] starting from 01/01/1959. It was extracted from the FRED-MD database. The series are differenced and log-transformed as suggested in the literature. Research work which uses this dataset includes:

- Principles and algorithms for forecasting groups of time series: locality and globality [60]

The DOI link to access and download the dataset is `http://doi.org/10.5281/zenodo.4654833`.

A.6.15 San Francisco traffic dataset

This dataset contains 862 hourly time series showing the road occupancy rates on San Francisco Bay area freeways from 2015 to 2016. It was used by Lai et al. [1], and originally extracted from Caltrans [68]. Godahewa et al. [28] use a weekly aggregated version of this dataset, which is also available in our repository.

The DOI links to access and download the datasets are as follows:

- Hourly dataset: `http://doi.org/10.5281/zenodo.4656132`
- Weekly dataset: `http://doi.org/10.5281/zenodo.4656135`

A.6.16 Melbourne pedestrian counts dataset

This dataset contains hourly pedestrian counts captured from 66 sensors in Melbourne city starting from May 2009 [69]. The original data are updated on a monthly basis when the new observations become available. The dataset in our repository contains pedestrian counts up to 30/04/2020. Research work which uses this dataset includes:

- Enhancing pedestrian mobility in smart cities using big data [70]
- Visualising Melbourne pedestrian count [71]
- PedaViz: visualising hour-level pedestrian activity [72]

The DOI link to access and download the dataset is `http://doi.org/10.5281/zenodo.4656626`.

A.6.17 Hospital dataset

This dataset contains 767 monthly time series showing the patient counts related to medical products from January 2000 to December 2006. It was extracted from the R package expsmooth [59]. The package contains this dataset as *"hospital"*. Research work which uses this dataset includes:

- Principles and algorithms for forecasting groups of time series: locality and globality [60]

The DOI link to access and download the dataset is `http://doi.org/10.5281/zenodo.4656014`.

A.6.18 COVID deaths dataset

This dataset contains 266 daily time series that represent the total COVID-19 deaths in a set of countries and states from 22/01/2020 to 20/08/2020. It was extracted from the Johns Hopkins repository [3, 73]. The original data are updated on a daily basis when the new observations become available.

The DOI link to access and download the dataset is `http://doi.org/10.5281/zenodo.4656009`.

A.6.19 KDD cup 2018 dataset

This competition dataset contains long hourly time series representing the air quality levels in 59 stations in 2 cities, Beijing (35 stations) and London (24 stations) from 01/01/2017 to 31/03/2018 [74]. The dataset represents the air quality in multiple measurements such as $PM2.5$, $PM10$, NO_2, CO, O_3 and SO_2 levels.

Our repository dataset contains 270 hourly time series which have been categorized using city, station name, and air quality measurement.

As the original dataset contains missing values, we include both the original dataset and an imputed version in our repository. We impute leading missing values with zeros and the remaining missing values using the LOCF method. Research work which uses this dataset includes:

- AccuAir: winning solution to air quality prediction for KDD cup 2018 [75]

The DOI links to access and download the datasets are as follows:

- Dataset with missing values: `http://doi.org/10.5281/zenodo.4656719`
- Dataset without missing values: `http://doi.org/10.5281/zenodo.4656756`

A.6.20 Weather dataset

This dataset contains 3010 daily time series representing the variations of four weather variables: rain, minimum temperature, maximum temperature and solar radiation, measured at weather stations in Australia. The series were extracted from the R package *bomrang* [76]. Research work which uses this dataset includes:

- Principles and algorithms for forecasting groups of time series: locality and globality [60]

The DOI link to access and download the dataset is `http://doi.org/10.5281/zenodo.4654822`.

A.7 Single long time series datasets

This section describes the benchmark datasets which have single time series with a large amount of data points.

A.7.1 Sunspot dataset

The original data source contains a single very long daily time series of sunspot numbers from 01/01/1818 until the present [77]. Furthermore, it also contains monthly mean total sunspot numbers (starting from 1749), 13-month smoothed monthly total sunspot numbers (starting from 1749), yearly mean total sunspot numbers (starting from 1700), daily hemispheric sunspot numbers (starting from 1992), monthly mean hemispheric sunspot numbers (starting from 1992), 13-month smoothed monthly hemispheric sunspot numbers (starting from 1992), and yearly mean total sunspot numbers (starting from 1610). The original datasets are updated as new observations become available.

Our repository contains the single daily time series representing the sunspot numbers from 08/01/1818 to 31/05/2020. As the dataset contains missing values, we include an LOCF-imputed version besides it in the repository. Research work which uses this dataset includes:

- Re-evaluation of predictive models in light of new data: sunspot number version 2.0 [78]
- Correlation between sunspot number and ca II K emission index [79]
- Dynamics of sunspot series on time scales from days to years: correlation of sunspot births, variable lifetimes, and evolution of the high-frequency spectral component [80]
- Long term sunspot cycle phase coherence with periodic phase disruptions [81]

The DOI links to access and download the datasets are as follows:

- Dataset with missing values: `http://doi.org/10.5281/zenodo.4654773`
- Dataset without missing values: `http://doi.org/10.5281/zenodo.4654722`

A.7.2 Saugeen river flow dataset

This dataset contains a single very long time series representing the daily mean flow of the Saugeen River at Walkerton in cubic meters per second from 01/01/1915 to 31/12/1979. The length of this time series is 23,741. It was extracted from the R package, *deseasonalize* [82]. The package contains this dataset as *"SaugeenDay"*.

Research work which uses this dataset includes:

- Telescope: an automatic feature extraction and transformation approach for time series forecasting on a level-playing field [83]

The DOI link to access and download the dataset is `http://doi.org/10.5281/zenodo.4656058`.

A.7.3 US births dataset

This dataset contains a single very long daily time series representing the number of births in the US from 01/01/1969 to 31/12/1988. The length of this time series is 7,305. It was extracted from the R package, *mosaicData* [84]. The package contains this dataset as *"Births"*. Research work which uses this dataset includes:

- Telescope: an automatic feature extraction and transformation approach for time series forecasting on a level-playing field [83]

The DOI link to access and download the dataset is `http://doi.org/10.5281/zenodo.4656049`.

A.7.4 Electricity demand dataset

This dataset contains a single very long time series representing the half hourly electricity demand for Victoria, Australia in 2014. The length of this time series is 17,520. It was extracted from the R package, *fpp2* [85]. The package contains this dataset as *"elecdemand"*. The temperatures corresponding with each demand value are also available in the original dataset. Research work which uses this dataset includes:

- Telescope: an automatic feature extraction and transformation approach for time series forecasting on a level-playing field [83]

The DOI link to access and download the dataset is `http://doi.org/10.5281/zenodo.4656069`.

A.7.5 Solar power dataset

This dataset contains a single very long time series representing the solar power production of an Australian wind farm recorded per each 4 seconds starting from 01/08/2019. It was extracted from the AEMO online platform [58]. The length of this time series is 7,397,222.

The DOI link to access and download the dataset is `http://doi.org/10.5281/zenodo.4656027`.

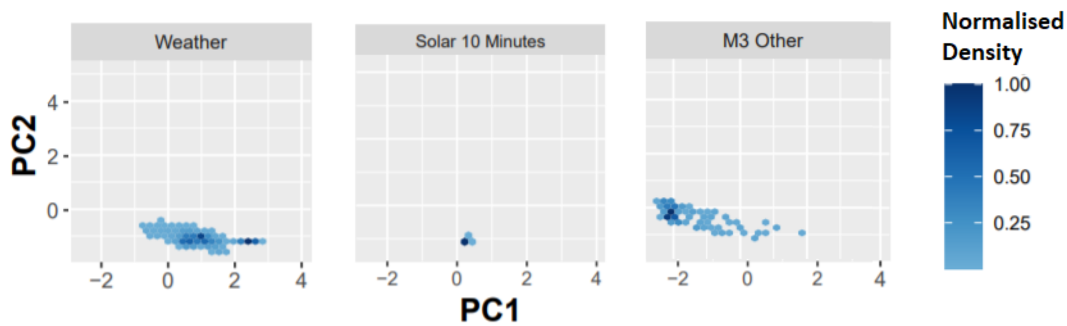
A.7.6 Wind power dataset

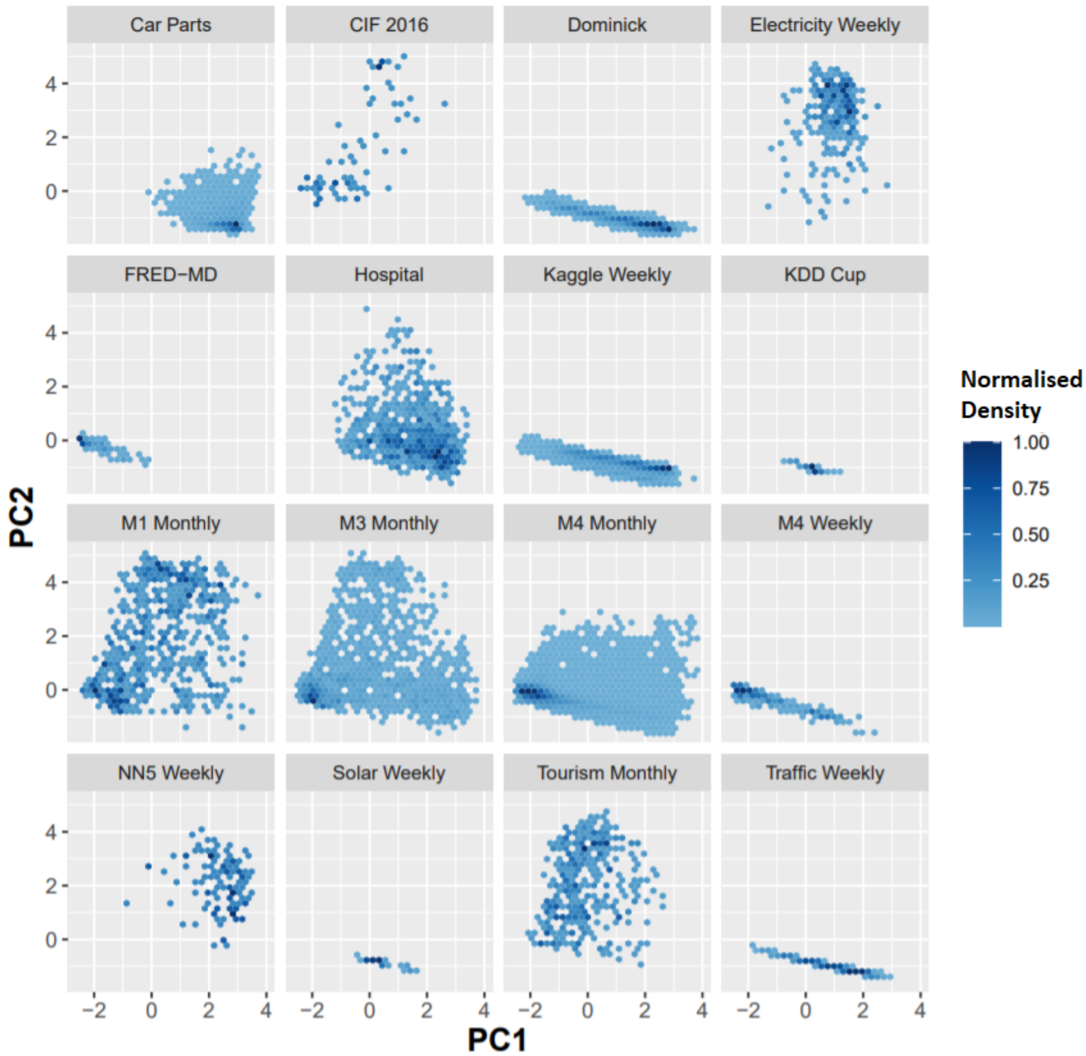
This dataset contains a single very long time series representing the wind power production of an Australian wind farm recorded per each 4 seconds starting from 01/08/2019. It was extracted from the AEMO online platform [58]. The length of this time series is 7,397,147.

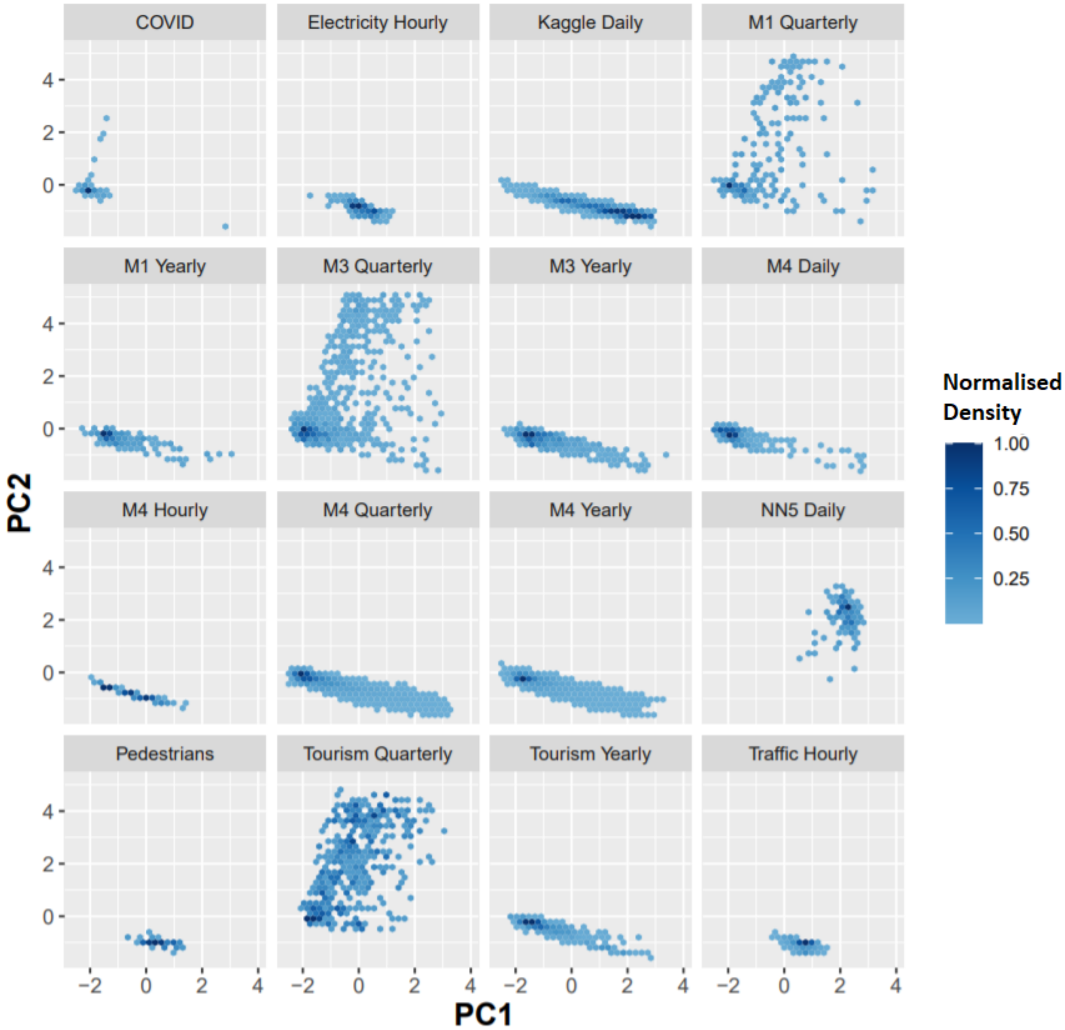
The DOI link to access and download the dataset is `http://doi.org/10.5281/zenodo.4656032`.

B Feature plots

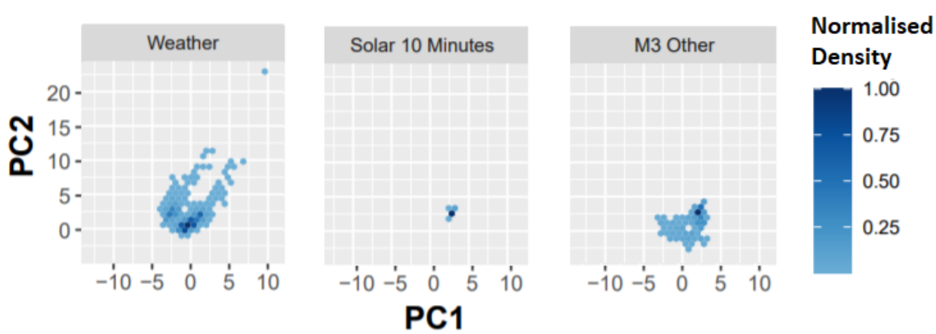
The following hexbin plots show the normalised density values of the low-dimensional features space generated by PCA for the datasets in our archive across 4 tsfeatures: ACF1, trend, entropy and seasonal strength, and the Box-Cox transformation parameter, lambda. The dark and light hexbins denote the high and low density areas respectively.

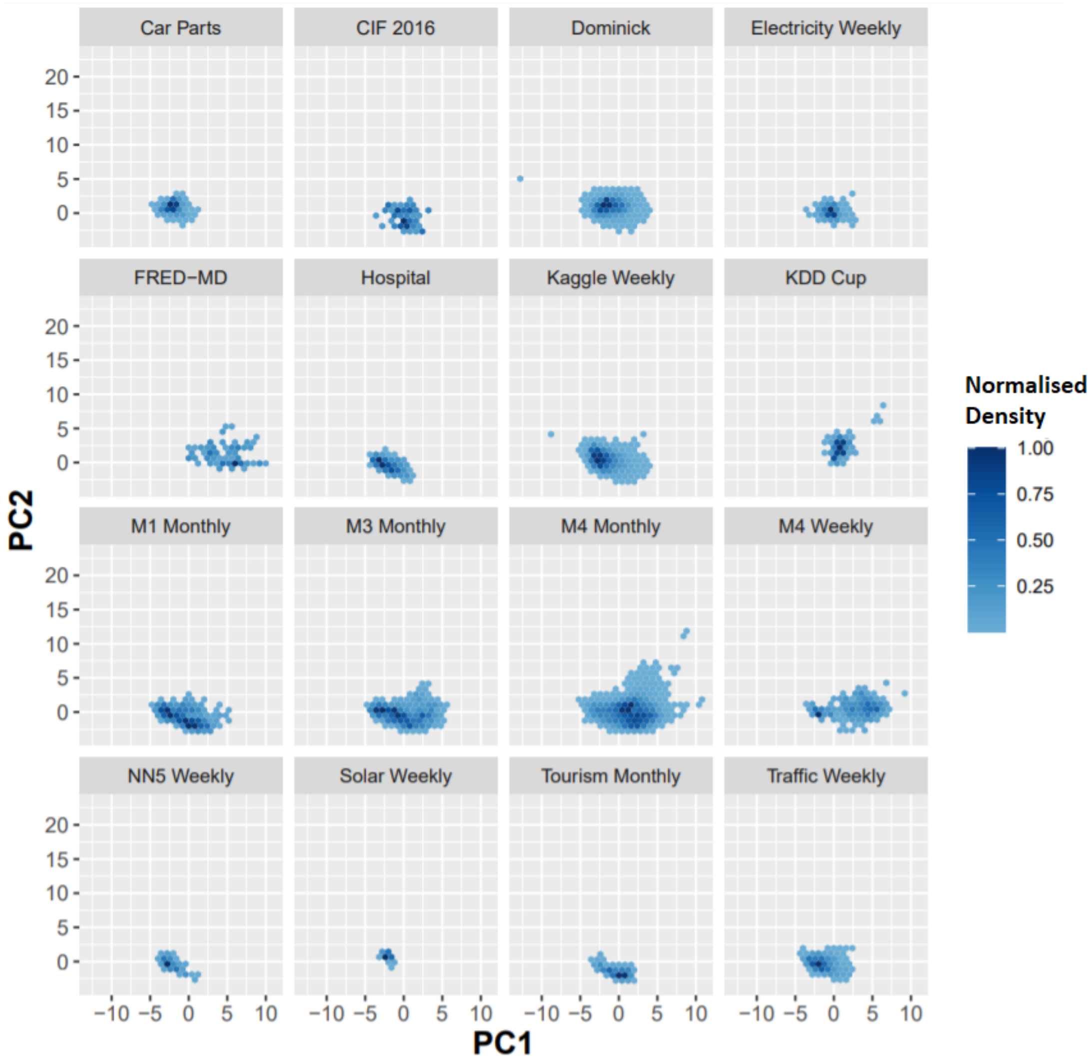


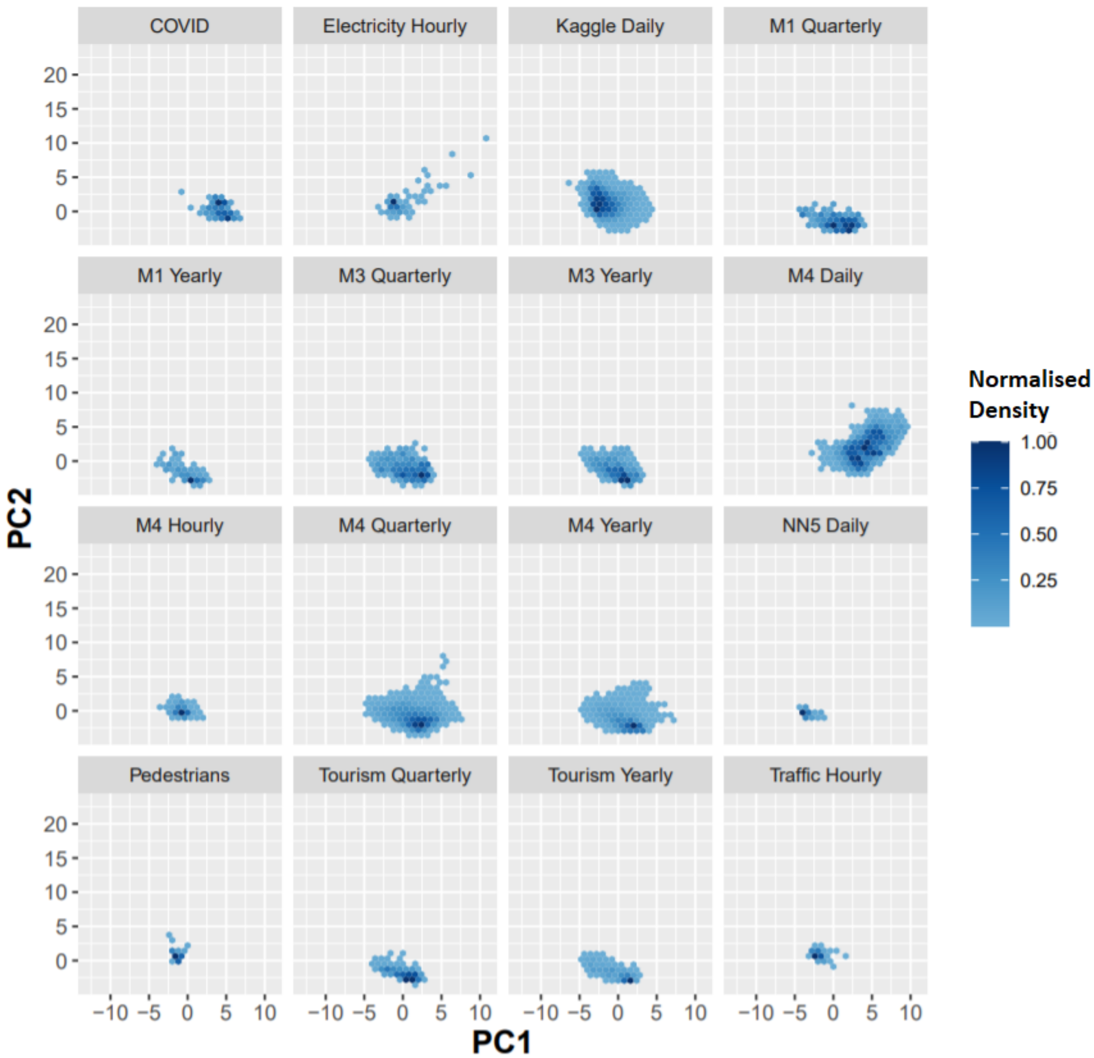

Car Parts
CIF 2016
Dominick
Electricity Weekly
FRED−MD
Hospital
Kaggle Weekly
KDD Cup
M1 Monthly
M3 Monthly
M4 Monthly
M4 Weekly
NN5 Weekly
Solar Weekly
Tourism Monthly
Traffic Weekly
PC1
PC2
Normalised Density
1.00
0.75
0.50
0.25



The following hexbin plots show the normalised density values of the low-dimensional features space generated by PCA for the datasets in our archive across the catch22 features. The dark and light hexbins denote the high and low density areas respectively.




Car Parts
CIF 2016
Dominick
Electricity Weekly
FRED−MD
Hospital
Kaggle Weekly
KDD Cup
M1 Monthly
M3 Monthly
M4 Monthly
M4 Weekly
NN5 Weekly
Solar Weekly
Tourism Monthly
Traffic Weekly
PC2
PC1
Normalised Density
1.00
0.75
0.50
0.25



C Baseline results

Tables 5, 6, 7, 8, 9, 10, 11, 12, 13 and 14 respectively show the mean MASE, median MASE, mean sMAPE, median sMAPE, mean msMAPE, median msMAPE, mean MAE, median MAE, mean RMSE and median RMSE results of SES, Theta, ETS, ARIMA, TBATS, DHR-ARIMA and PR models on all datasets.

Table 5: Mean MASE results

Dataset	SES	Theta	ETS	ARIMA	TBATS	DHR-ARIMA	PR
NN5 Daily	1.521	0.885	0.865	1.013	-	-	1.263
NN5 Weekly	0.903	0.885	-	-	0.872	0.887	0.854
CIF 2016	1.291	0.997	0.841	0.929	-	-	1.019
US Births	4.343	2.138	1.529	1.917	-	-	2.094
Saugeen River Flow	1.426	1.425	2.036	1.485	-	-	1.674
Elecdemand	1.126	1.125	-	-	1.272	0.902	1.153
Kaggle Daily	0.924	0.928	1.231	0.890	-	-	-
Kaggle Weekly	0.698	0.694	-	-	0.622	0.815	1.021
Tourism Yearly	3.253	3.015	3.395	3.775	-	-	3.516
Tourism Quarterly	3.210	1.661	1.592	1.782	-	-	1.643
Tourism Monthly	3.306	1.649	1.526	1.589	-	-	1.678
Traffic Hourly	1.922	1.922	-	-	2.482	2.535	1.281
Traffic Weekly	1.116	1.121	-	-	1.148	1.191	1.122
Electricity Hourly	4.544	4.545	-	-	3.690	4.602	2.912
Electricity Weekly	1.536	1.476	-	-	0.792	0.878	0.916
Solar 10 Minutes	1.451	1.452	-	-	3.936	1.034	1.451
Solar Weekly	1.215	1.224	-	-	0.916	0.848	1.053
Sunspot	0.128	0.128	0.128	0.067	-	-	0.099
M1 Yearly	4.938	4.191	3.771	4.479	-	-	4.588
M1 Quarterly	1.929	1.702	1.658	1.787	-	-	1.892
M1 Monthly	1.379	1.091	1.074	1.164	-	-	1.123
M3 Yearly	3.167	2.774	2.860	3.417	-	-	3.223
M3 Quarterly	1.417	1.117	1.170	1.240	-	-	1.248
M3 Monthly	1.091	0.864	0.865	0.873	-	-	1.010
M3 Other	3.089	2.271	1.814	1.831	-	-	2.655
M4 Yearly	3.981	3.375	3.444	3.876	-	-	3.625
M4 Quarterly	1.417	1.231	1.161	1.228	-	-	1.316
M4 Monthly	1.150	0.970	0.948	0.962	-	-	1.080
M4 Weekly	0.587	0.546	-	-	0.504	0.550	0.481
M4 Daily	1.154	1.153	1.239	1.179	-	-	1.162
M4 Hourly	11.607	11.524	-	-	2.663	13.557	1.662
Pedestrian Counts	0.957	0.958	-	-	1.297	3.947	0.256
KDD Cup	1.645	1.646	-	-	1.394	1.982	1.265
Carparts	0.897	0.914	0.925	0.926	-	-	0.755
Hospital	0.813	0.761	0.765	0.787	-	-	0.782
Covid Deaths	7.776	7.793	5.326	6.117	-	-	8.731
FRED-MD	0.617	0.698	0.468	0.533	-	-	8.827
Dominick	0.582	0.610	-	-	72721475.060	0.796	0.980
Weather	0.677	0.749	0.702	0.746	-	-	3.046

Table 6: Median MASE results

Dataset	SES	Theta	ETS	ARIMA	TBATS	DHR-ARIMA	PR
NN5 Daily	1.482	0.838	0.809	0.926	-	-	1.224
NN5 Weekly	0.781	0.805	-	-	0.827	0.769	0.781
CIF 2016	0.862	0.662	0.532	0.559	-	-	0.746
US Births	4.343	2.138	1.529	1.917	-	-	2.094
Saugeen River Flow	1.426	1.425	2.036	1.485	-	-	1.674
Elecdemand	1.126	1.125	-	-	1.272	0.902	1.153
Kaggle Daily	0.539	0.548	0.667	0.528	-	-	-
Kaggle Weekly	0.432	0.418	-	-	0.330	0.529	0.573
Tourism Yearly	2.442	2.360	2.373	2.719	-	-	2.356
Tourism Quarterly	2.309	1.348	1.275	1.388	-	-	1.361
Tourism Monthly	2.336	1.382	1.276	1.337	-	-	1.484
Traffic Hourly	1.817	1.817	-	-	1.380	2.365	1.228
Traffic Weekly	0.973	0.983	-	-	0.996	1.035	0.980
Electricity Hourly	4.766	4.766	-	-	2.300	4.630	2.878
Electricity Weekly	1.341	1.303	-	-	0.705	0.798	0.842
Solar 10 Minutes	1.403	1.404	-	-	2.431	1.029	1.403
Solar Weekly	1.231	1.241	-	-	0.894	0.861	1.063
Sunspot	0.128	0.128	0.128	0.067	-	-	0.099
M1 Yearly	3.772	3.155	2.324	2.127	-	-	2.847
M1 Quarterly	1.417	1.264	1.196	1.171	-	-	1.376
M1 Monthly	1.167	0.885	0.851	0.894	-	-	0.947
M3 Yearly	2.261	1.985	1.907	2.003	-	-	2.267
M3 Quarterly	1.073	0.831	0.855	0.917	-	-	0.902
M3 Monthly	0.861	0.721	0.712	0.704	-	-	0.825
M3 Other	2.771	1.896	1.489	1.418	-	-	2.067
M4 Yearly	2.940	2.312	2.329	2.753	-	-	2.568
M4 Quarterly	1.142	0.973	0.886	0.925	-	-	1.038
M4 Monthly	0.867	0.763	0.736	0.727	-	-	0.844
M4 Weekly	0.441	0.416	-	-	0.365	0.382	0.392
M4 Daily	0.862	0.861	0.859	0.867	-	-	0.868
M4 Hourly	3.685	3.688	-	-	1.873	3.507	1.010
Pedestrian Counts	0.604	0.605	-	-	1.004	4.125	0.128
KDD Cup	1.357	1.357	-	-	1.246	1.744	1.035
Carparts	0.562	0.482	0.562	0.600	-	-	0.375
Hospital	0.745	0.723	0.731	0.733	-	-	0.740
Covid Deaths	1.554	2.192	0.614	0.982	-	-	5.313
FRED-MD	0.430	0.407	0.385	0.355	-	-	8.458
Dominick	0.194	0.208	-	-	0.535	0.453	0.000
Weather	0.618	0.624	0.643	0.687	-	-	1.755

Table 7: Mean sMAPE results

Dataset	SES	Theta	ETS	ARIMA	TBATS	DHR-ARIMA	PR
NN5 Daily	35.50	22.01	21.57	26.01	-	-	30.30
NN5 Weekly	12.24	11.96	-	-	11.63	11.84	11.45
CIF 2016	14.95	13.05	12.18	11.70	-	-	12.33
US Births	11.77	5.82	4.05	5.17	-	-	5.75
Saugeen River Flow	36.03	36.01	67.60	37.58	-	-	45.37
Elecdemand	10.99	10.98	-	-	11.85	8.53	10.51
Kaggle Daily	-	-	-	-	-	-	-
Kaggle Weekly	-	-	-	-	-	122.64	-
Tourism Yearly	34.14	31.96	36.56	33.44	-	-	46.94
Tourism Quarterly	27.41	15.37	15.07	16.58	-	-	15.86
Tourism Monthly	36.39	19.90	19.02	19.73	-	-	21.11
Traffic Hourly	-	82.44	-	-	70.59	92.58	-
Traffic Weekly	12.49	12.56	-	-	12.88	13.54	12.55
Electricity Hourly	-	-	-	-	40.47	-	-
Electricity Weekly	-	14.58	-	-	-	10.86	-
Solar 10 Minutes	200.00	200.00	-	-	165.81	85.81	199.99
Solar Weekly	24.59	24.76	-	-	19.05	17.87	21.65
Sunspot	196.19	196.19	196.19	194.29	-	-	195.56
M1 Yearly	23.10	20.17	18.61	19.47	-	-	18.79
M1 Quarterly	18.10	16.35	17.47	16.62	-	-	16.67
M1 Monthly	17.43	16.53	15.05	15.65	-	-	15.20
M3 Yearly	17.76	16.76	17.00	18.84	-	-	17.13
M3 Quarterly	10.90	9.20	9.68	10.24	-	-	9.77
M3 Monthly	16.22	13.86	14.14	14.24	-	-	15.17
M3 Other	6.28	4.92	4.37	4.35	-	-	5.32
M4 Yearly	16.40	14.56	15.36	16.03	-	-	14.53
M4 Quarterly	11.08	10.31	10.29	10.52	-	-	10.84
M4 Monthly	14.38	13.01	13.53	13.08	-	-	13.74
M4 Weekly	9.01	7.83	-	-	7.30	7.94	7.43
M4 Daily	3.05	3.07	3.13	3.01	-	-	3.06
M4 Hourly	42.95	42.98	-	-	28.12	35.99	11.68
Pedestrian Counts	123.96	124.19	-	-	119.96	138.67	41.10
KDD Cup	62.20	62.31	-	-	56.37	86.13	50.73
Carparts	-	-	-	-	-	-	-
Hospital	17.98	17.31	17.50	17.83	-	-	17.60
Covid Deaths	-	-	-	-	-	-	-
FRED-MD	10.65	13.41	10.33	11.36	-	-	33.21
Dominick	-	-	-	-	-	-	-
Weather	62.16	68.24	62.85	-	-	-	-

Table 8: Median sMAPE results

Dataset	SES	Theta	ETS	ARIMA	TBATS	DHR-ARIMA	PR
NN5 Daily	34.68	20.56	20.35	22.80	-	-	28.81
NN5 Weekly	10.95	10.96	-	-	10.97	11.08	10.50
CIF 2016	11.40	7.95	6.58	7.69	-	-	8.43
US Births	11.77	5.82	4.05	5.17	-	-	5.75
Saugeen River Flow	36.03	36.01	67.60	37.58	-	-	45.37
Elecdemand	10.99	10.98	-	-	11.85	8.53	10.51
Kaggle Daily	-	-	-	-	-	-	-
Kaggle Weekly	-	-	-	-	-	117.72	-
Tourism Yearly	18.81	16.83	19.20	22.66	-	-	16.88
Tourism Quarterly	22.48	13.17	12.89	13.13	-	-	13.33
Tourism Monthly	30.24	17.40	17.16	18.01	-	-	18.47
Traffic Hourly	-	74.21	-	-	55.69	86.56	-
Traffic Weekly	9.70	9.75	-	-	10.05	10.54	9.75
Electricity Hourly	-	-	-	-	23.23	-	-
Electricity Weekly	-	11.72	-	-	-	6.97	-
Solar 10 Minutes	200.00	200.00	-	-	161.88	85.73	200.00
Solar Weekly	24.76	24.90	-	-	18.36	17.64	21.77
Sunspot	196.19	196.19	196.19	194.29	-	-	195.56
M1 Yearly	17.33	14.74	13.01	11.99	-	-	13.49
M1 Quarterly	11.24	8.63	8.40	9.66	-	-	10.13
M1 Monthly	14.35	11.18	10.82	11.51	-	-	11.88
M3 Yearly	12.44	11.54	11.52	12.37	-	-	12.92
M3 Quarterly	6.74	5.23	5.53	6.36	-	-	5.73
M3 Monthly	10.71	9.25	9.13	9.01	-	-	10.40
M3 Other	4.62	2.88	2.23	2.19	-	-	3.42
M4 Yearly	11.41	9.23	8.97	10.20	-	-	9.49
M4 Quarterly	6.94	6.06	5.61	5.80	-	-	6.34
M4 Monthly	8.38	7.24	7.00	7.13	-	-	8.20
M4 Weekly	5.17	5.19	-	-	4.81	5.10	4.99
M4 Daily	1.99	2.01	1.99	2.01	-	-	2.00
M4 Hourly	19.88	19.79	-	-	6.55	32.18	5.80
Pedestrian Counts	123.48	124.69	-	-	118.71	142.81	36.80
KDD Cup	60.40	60.57	-	-	53.99	85.87	52.82
Carparts	-	-	-	-	-	-	-
Hospital	16.58	15.91	16.13	16.77	-	-	16.14
Covid Deaths	-	-	-	-	-	-	-
FRED-MD	1.61	1.53	1.54	1.57	-	-	29.14
Dominick	-	-	-	-	-	-	-
Weather	23.71	23.83	25.36	-	-	-	-

Table 9: Mean msMAPE results

Dataset	SES	Theta	ETS	ARIMA	TBATS	DHR-ARIMA	PR
NN5 Daily	35.38	21.93	21.49	25.91	-	-	30.20
NN5 Weekly	12.24	11.96	-	-	11.62	11.83	11.45
CIF 2016	14.94	13.04	12.18	11.69	-	-	12.32
US Births	11.77	5.82	4.05	5.17	-	-	5.75
Saugeen River Flow	35.99	35.97	67.50	37.55	-	-	45.32
Elecdemand	10.85	10.84	-	-	11.70	8.42	10.38
Kaggle Daily	45.87	47.98	57.94	44.39	-	-	-
Kaggle Weekly	45.10	47.72	-	-	40.88	65.55	72.93
Tourism Yearly	34.10	31.93	36.52	33.39	-	-	46.92
Tourism Quarterly	27.41	15.37	15.07	16.58	-	-	15.86
Tourism Monthly	36.39	19.89	19.02	19.73	-	-	21.11
Traffic Hourly	8.73	8.73	-	-	12.58	11.72	5.97
Traffic Weekly	12.40	12.48	-	-	12.80	13.45	12.46
Electricity Hourly	44.39	44.94	-	-	40.15	43.78	30.00
Electricity Weekly	14.17	14.58	-	-	8.50	10.86	9.98
Solar 10 Minutes	65.07	65.75	-	-	154.38	30.20	65.09
Solar Weekly	24.59	24.76	-	-	19.05	17.87	21.65
Sunspot	192.36	192.36	192.36	172.80	-	-	190.14
M1 Yearly	23.08	20.16	18.60	19.46	-	-	18.77
M1 Quarterly	18.07	16.33	17.41	16.59	-	-	16.64
M1 Monthly	17.12	15.53	14.64	15.26	-	-	14.85
M3 Yearly	17.76	16.76	17.00	18.84	-	-	17.13
M3 Quarterly	10.90	9.20	9.68	10.24	-	-	9.77
M3 Monthly	16.22	13.86	14.14	14.24	-	-	15.17
M3 Other	6.28	4.92	4.37	4.35	-	-	5.32
M4 Yearly	16.40	14.56	15.36	16.03	-	-	14.53
M4 Quarterly	11.08	10.31	10.29	10.52	-	-	10.83
M4 Monthly	14.38	13.01	13.52	13.08	-	-	13.73
M4 Weekly	9.01	7.83	-	-	7.30	7.94	7.43
M4 Daily	3.04	3.07	3.13	3.01	-	-	3.06
M4 Hourly	42.92	42.94	-	-	28.10	35.94	11.67
Pedestrian Counts	121.39	122.08	-	-	119.76	138.58	40.29
KDD Cup	61.68	61.80	-	-	55.91	85.32	50.33
Carparts	64.88	59.27	65.76	65.61	-	-	43.23
Hospital	17.94	17.27	17.46	17.79	-	-	17.56
Covid Deaths	15.35	15.57	8.64	9.26	-	-	18.34
FRED-MD	8.72	9.72	8.40	7.98	-	-	30.77
Dominick	72.94	114.89	-	-	112.76	136.72	68.44
Weather	50.85	56.19	51.47	57.98	-	-	106.01

Table 10: Median msMAPE results

Dataset	SES	Theta	ETS	ARIMA	TBATS	DHR-ARIMA	PR
NN5 Daily	34.57	20.51	20.31	22.72	-	-	28.70
NN5 Weekly	10.94	10.96	-	-	10.97	11.08	10.50
CIF 2016	11.40	7.95	6.58	7.69	-	-	8.43
US Births	11.77	5.82	4.05	5.17	-	-	5.75
Saugeen River Flow	35.99	35.97	67.50	37.55	-	-	45.32
Elecdemand	10.85	10.84	-	-	11.70	8.42	10.38
Kaggle Daily	37.02	37.69	46.19	35.16	-	-	-
Kaggle Weekly	32.52	33.01	-	-	29.31	46.41	73.32
Tourism Yearly	18.77	16.83	19.04	22.57	-	-	16.88
Tourism Quarterly	22.48	13.17	12.89	13.13	-	-	13.33
Tourism Monthly	30.24	17.40	17.16	18.00	-	-	18.47
Traffic Hourly	8.26	8.26	-	-	7.58	10.66	5.43
Traffic Weekly	9.66	9.71	-	-	10.01	10.48	9.69
Electricity Hourly	42.08	42.20	-	-	23.22	38.30	24.78
Electricity Weekly	12.22	11.72	-	-	6.17	6.97	7.41
Solar 10 Minutes	64.68	65.26	-	-	152.54	30.26	64.68
Solar Weekly	24.76	24.90	-	-	18.36	17.64	21.77
Sunspot	192.36	192.36	192.36	172.80	-	-	190.14
M1 Yearly	17.14	14.74	12.99	11.99	-	-	13.49
M1 Quarterly	11.23	8.63	8.40	9.66	-	-	10.13
M1 Monthly	14.24	11.18	10.80	11.48	-	-	11.79
M3 Yearly	12.44	11.54	11.52	12.37	-	-	12.92
M3 Quarterly	6.74	5.23	5.53	6.36	-	-	5.73
M3 Monthly	10.71	9.25	9.13	9.01	-	-	10.39
M3 Other	4.62	2.88	2.23	2.19	-	-	3.42
M4 Yearly	11.41	9.23	8.97	10.20	-	-	9.49
M4 Quarterly	6.94	6.06	5.61	5.80	-	-	6.34
M4 Monthly	8.38	7.24	7.00	7.13	-	-	8.20
M4 Weekly	5.17	5.19	-	-	4.81	5.10	4.99
M4 Daily	1.99	2.01	1.99	2.01	-	-	2.00
M4 Hourly	19.86	19.75	-	-	6.55	32.08	5.80
Pedestrian Counts	121.41	124.44	-	-	118.52	142.71	36.46
KDD Cup	60.18	60.23	-	-	53.98	84.90	52.77
Carparts	45.45	45.45	46.18	46.18	-	-	30.30
Hospital	16.57	15.90	16.11	16.75	-	-	16.12
Covid Deaths	2.68	6.10	1.38	1.94	-	-	16.99
FRED-MD	1.58	1.53	1.54	1.57	-	-	29.09
Dominick	41.90	130.16	-	-	137.87	164.14	0.00
Weather	22.26	22.31	23.59	23.00	-	-	119.15

Table 11: Mean MAE results

Dataset	SES	Theta	ETS	ARIMA	TBATS	DHR-ARIMA	PR
NN5 Daily	6.63	3.80	3.72	4.41	-	-	5.47
NN5 Weekly	15.66	15.30	-	-	14.98	15.38	14.94
CIF 2016	581875.97	714818.58	642421.42	469059.49	-	-	563205.57
US Births	1192.20	586.93	419.73	526.33	-	-	574.93
Saugeen River Flow	21.50	21.49	30.69	22.38	-	-	25.24
Elecdemand	0.42	0.42	-	-	0.47	0.33	0.43
Kaggle Daily	363.43	358.73	403.23	340.36	-	-	-
Kaggle Weekly	2337.11	2373.98	-	-	2241.84	3115.03	4051.75
Tourism Yearly	95579.23	90653.60	94818.89	95033.24	-	-	82682.97
Tourism Quarterly	15014.19	7656.49	8925.52	10475.47	-	-	9092.58
Tourism Monthly	5302.10	2069.96	2004.51	2536.77	-	-	2187.28
Traffic Hourly	0.03	0.03	-	-	0.04	0.04	0.02
Traffic Weekly	1.12	1.13	-	-	1.17	1.22	1.13
Electricity Hourly	845.97	846.03	-	-	574.30	868.20	537.38
Electricity Weekly	74149.18	74111.14	-	-	24347.24	28457.18	44882.52
Solar 10 Minutes	3.28	3.29	-	-	8.77	2.37	3.28
Solar Weekly	1202.39	1210.83	-	-	908.65	839.88	1044.98
Sunspot	4.93	4.93	4.93	2.57	-	-	3.83
M1 Yearly	171353.41	152799.26	146110.11	145608.87	-	-	134246.38
M1 Quarterly	2206.27	1981.96	2088.15	2191.10	-	-	1630.38
M1 Monthly	2259.04	2166.18	1905.28	2080.13	-	-	2088.25
M3 Yearly	1022.27	957.40	1031.40	1416.31	-	-	1018.48
M3 Quarterly	571.96	486.31	513.06	559.40	-	-	519.30
M3 Monthly	743.41	623.71	626.46	654.80	-	-	692.97
M3 Other	277.83	215.35	194.98	193.02	-	-	234.43
M4 Yearly	1009.06	890.51	920.66	1067.16	-	-	875.76
M4 Quarterly	622.57	574.34	573.19	604.51	-	-	610.51
M4 Monthly	625.24	563.58	582.60	575.36	-	-	596.19
M4 Weekly	336.82	333.32	-	-	296.15	321.61	293.21
M4 Daily	178.27	178.86	193.26	179.67	-	-	181.92
M4 Hourly	1218.06	1220.97	-	-	386.27	1310.85	257.39
Pedestrian Counts	170.87	170.94	-	-	222.38	635.16	44.18
KDD Cup	42.04	42.06	-	-	39.20	52.20	36.85
Carparts	0.55	0.53	0.56	0.56	-	-	0.41
Hospital	21.76	18.54	17.97	19.60	-	-	19.24
Covid Deaths	353.71	321.32	85.59	85.77	-	-	347.98
FRED-MD	2798.22	3492.84	2041.42	2957.11	-	-	8921.94
Dominick	5.70	5.86	-	-	6.95	7.10	8.19
Weather	2.24	2.51	2.35	2.45	-	-	8.17

Table 12: Median MAE results

Dataset	SES	Theta	ETS	ARIMA	TBATS	DHR-ARIMA	PR
NN5 Daily	5.94	3.55	3.48	3.85	-	-	5.06
NN5 Weekly	14.18	13.90	-	-	13.73	14.82	12.84
CIF 2016	107.09	103.39	70.43	80.66	-	-	95.13
US Births	1192.20	586.93	419.73	526.33	-	-	574.93
Saugeen River Flow	21.50	21.49	30.69	22.38	-	-	25.24
Elecdemand	0.42	0.42	-	-	0.47	0.33	0.43
Kaggle Daily	51.05	51.64	69.27	46.27	-	-	-
Kaggle Weekly	357.12	355.50	-	-	278.00	609.62	494.38
Tourism Yearly	4312.77	4085.98	4271.06	4623.59	-	-	4340.90
Tourism Quarterly	1921.00	1114.30	1003.24	1047.01	-	-	992.12
Tourism Monthly	967.57	478.45	457.04	462.53	-	-	474.72
Traffic Hourly	0.02	0.02	-	-	0.02	0.03	0.02
Traffic Weekly	0.92	0.92	-	-	0.94	0.98	0.93
Electricity Hourly	210.20	210.20	-	-	127.05	215.60	137.88
Electricity Weekly	10983.75	10447.12	-	-	6149.88	6789.75	7090.88
Solar 10 Minutes	2.92	2.92	-	-	5.64	2.13	2.92
Solar Weekly	1091.23	1103.20	-	-	780.04	760.63	942.23
Sunspot	4.93	4.93	4.93	2.57	-	-	3.83
M1 Yearly	379.28	255.75	191.24	179.98	-	-	245.67
M1 Quarterly	22.30	19.55	19.59	16.23	-	-	19.19
M1 Monthly	45.33	38.23	38.51	40.54	-	-	37.37
M3 Yearly	703.33	660.49	641.07	701.32	-	-	711.86
M3 Quarterly	371.95	294.16	304.53	333.74	-	-	325.44
M3 Monthly	517.09	420.80	408.92	412.47	-	-	479.18
M3 Other	164.13	104.93	83.64	77.02	-	-	127.12
M4 Yearly	529.96	428.94	427.24	493.19	-	-	456.65
M4 Quarterly	318.93	274.24	250.82	262.40	-	-	295.64
M4 Monthly	291.89	249.73	244.21	243.12	-	-	280.83
M4 Weekly	219.63	210.47	-	-	163.68	188.39	176.01
M4 Daily	92.14	91.85	92.16	92.18	-	-	92.28
M4 Hourly	49.20	49.21	-	-	33.77	30.75	14.21
Pedestrian Counts	67.40	67.52	-	-	131.83	448.38	17.02
KDD Cup	27.75	27.74	-	-	22.25	31.06	18.00
Carparts	0.33	0.25	0.33	0.33	-	-	0.25
Hospital	6.67	6.67	6.67	6.83	-	-	6.67
Covid Deaths	2.23	4.42	1.65	1.78	-	-	6.77
FRED-MD	1.89	1.94	2.35	2.73	-	-	41.36
Dominick	0.89	1.25	-	-	5.05	3.66	0.00
Weather	2.17	2.18	2.27	2.23	-	-	7.89

Table 13: Mean RMSE results

Dataset	SES	Theta	ETS	ARIMA	TBATS	DHR-ARIMA	PR
NN5 Daily	8.23	5.28	5.22	6.05	-	-	7.26
NN5 Weekly	18.82	18.65	-	-	18.53	18.55	18.62
CIF 2016	657112.42	804654.19	722397.37	526395.02	-	-	648890.31
US Births	1369.50	735.51	607.20	705.51	-	-	732.09
Saugeen River Flow	39.79	39.79	50.39	43.23	-	-	47.70
Elecdemand	0.51	0.51	-	-	0.59	0.43	0.53
Kaggle Daily	590.11	583.32	650.43	595.43	-	-	-
Kaggle Weekly	2970.78	3012.39	-	-	2951.87	3777.28	4750.26
Tourism Yearly	106665.20	99914.21	104700.51	106082.60	-	-	89645.61
Tourism Quarterly	17270.57	9254.63	10812.34	12564.77	-	-	11746.85
Tourism Monthly	7039.35	2701.96	2542.96	3132.40	-	-	2739.43
Traffic Hourly	0.04	0.04	-	-	0.05	0.04	0.03
Traffic Weekly	1.51	1.53	-	-	1.53	1.54	1.50
Electricity Hourly	1026.29	1026.36	-	-	743.35	1082.44	689.85
Electricity Weekly	77067.87	76935.58	-	-	28039.73	32594.81	47802.08
Solar 10 Minutes	7.23	7.23	-	-	10.71	5.55	7.23
Solar Weekly	1331.26	1341.55	-	-	1049.01	967.87	1168.18
Sunspot	4.95	4.95	4.95	2.96	-	-	3.95
M1 Yearly	193829.49	171458.07	167739.02	175343.75	-	-	152038.68
M1 Quarterly	2545.73	2282.65	2408.47	2538.45	-	-	1909.31
M1 Monthly	2725.83	2564.88	2263.96	2450.61	-	-	2478.88
M3 Yearly	1172.85	1106.05	1189.21	1662.17	-	-	1181.81
M3 Quarterly	670.56	567.70	598.73	650.76	-	-	605.50
M3 Monthly	893.88	753.99	755.26	790.76	-	-	830.04
M3 Other	309.68	242.13	224.08	220.77	-	-	262.31
M4 Yearly	1154.49	1020.48	1052.12	1230.35	-	-	1000.18
M4 Quarterly	732.82	673.15	674.27	709.99	-	-	711.93
M4 Monthly	755.45	683.72	705.70	702.06	-	-	720.46
M4 Weekly	412.60	405.17	-	-	356.74	386.30	350.29
M4 Daily	209.75	210.37	229.97	212.64	-	-	213.01
M4 Hourly	1476.81	1483.70	-	-	469.87	1563.05	312.98
Pedestrian Counts	228.14	228.20	-	-	261.25	820.28	61.84
KDD Cup	73.81	73.83	-	-	71.21	82.66	68.20
Carparts	0.78	0.78	0.80	0.81	-	-	0.73
Hospital	26.55	22.59	22.02	23.68	-	-	23.48
Covid Deaths	403.41	370.14	102.08	100.46	-	-	394.07
FRED-MD	3103.00	3898.72	2341.72	3312.46	-	-	9736.93
Dominick	6.48	6.74	-	-	8.12	7.96	9.44
Weather	2.85	3.27	2.96	3.07	-	-	9.08

Table 14: Median RMSE results

Dataset	SES	Theta	ETS	ARIMA	TBATS	DHR-ARIMA	PR
NN5 Daily	7.46	4.95	4.86	5.42	-	-	6.80
NN5 Weekly	17.52	16.82	-	-	16.99	17.49	16.26
CIF 2016	129.06	118.29	85.77	103.14	-	-	109.09
US Births	1369.50	735.51	607.20	705.51	-	-	732.09
Saugeen River Flow	39.79	39.79	50.39	43.23	-	-	47.70
Elecdemand	0.51	0.51	-	-	0.59	0.43	0.53
Kaggle Daily	74.58	75.16	98.97	68.13	-	-	-
Kaggle Weekly	424.02	429.71	-	-	346.60	707.60	576.40
Tourism Yearly	4718.37	4615.95	4626.74	5174.76	-	-	4717.10
Tourism Quarterly	2295.67	1392.89	1207.24	1196.05	-	-	1184.48
Tourism Monthly	1250.26	675.10	598.88	603.66	-	-	596.26
Traffic Hourly	0.03	0.03	-	-	0.03	0.04	0.02
Traffic Weekly	1.20	1.22	-	-	1.21	1.21	1.19
Electricity Hourly	256.22	256.22	-	-	181.79	275.52	171.57
Electricity Weekly	12460.16	11805.76	-	-	7278.04	8268.55	8237.57
Solar 10 Minutes	6.59	6.60	-	-	7.47	5.05	6.59
Solar Weekly	1193.90	1214.27	-	-	885.59	878.01	1016.25
Sunspot	4.95	4.95	4.95	2.96	-	-	3.95
M1 Yearly	416.37	323.31	230.39	207.82	-	-	304.77
M1 Quarterly	24.46	22.81	21.86	20.23	-	-	22.53
M1 Monthly	54.67	46.40	44.39	47.11	-	-	45.35
M3 Yearly	803.71	740.10	758.62	814.68	-	-	824.55
M3 Quarterly	436.25	355.79	368.91	405.87	-	-	378.31
M3 Monthly	633.56	516.79	495.97	497.97	-	-	582.04
M3 Other	182.17	120.84	99.97	92.60	-	-	144.46
M4 Yearly	610.38	497.80	494.90	567.70	-	-	525.42
M4 Quarterly	378.29	322.60	297.17	310.08	-	-	346.99
M4 Monthly	348.59	299.02	293.25	292.51	-	-	333.30
M4 Weekly	262.04	242.14	-	-	197.26	224.55	223.12
M4 Daily	108.04	108.55	108.77	108.40	-	-	108.48
M4 Hourly	61.40	61.58	-	-	42.90	42.93	19.89
Pedestrian Counts	88.65	88.76	-	-	155.94	627.43	22.09
KDD Cup	30.66	30.64	-	-	25.60	40.08	22.37
Carparts	0.71	0.65	0.71	0.71	-	-	0.58
Hospital	8.26	8.20	8.25	8.45	-	-	8.25
Covid Deaths	3.09	5.29	2.21	2.16	-	-	8.28
FRED-MD	2.31	2.36	2.70	3.49	-	-	45.18
Dominick	0.93	1.32	-	-	5.88	4.00	0.00
Weather	2.67	2.68	2.76	2.76	-	-	8.74

D Execution times

Table 15 shows the execution times corresponding with the traditional univariate forecasting models and the PR models across all datasets. Note that the *Univariate Models* column show the addition of ETS, ARIMA, SES and Theta execution times for yearly, quarterly, monthly and daily datasets, whereas it shows the addition of TBATS, DHR-ARIMA, SES and Theta execution times for weekly, hourly, half hourly and 10 minutes datasets.

The experiments are run on an Intel(R) Core(TM) i7-8700 processor (3.2GHz) and 65GB of main memory.

Table 15: Execution times of univariate and PR models

Dataset	Univariate Models	PR Model
NN5 Daily	16.11 mins	1.23 secs
NN5 Weekly	5.92 mins	0.79 secs
CIF 2016	1.85 mins	0.48 secs
US Births	12.15 secs	0.06 secs
Saugeen River Flow	24.22 secs	0.14 secs
Electricity demand	1.83 mins	8.02 secs
Kaggle Daily	8.79 days	-
Kaggle Weekly	4.12 days	1.04 days
Tourism Yearly	37.07 secs	3.05 secs
Tourism Quarterly	4.05 mins	2.70 secs
Tourism Monthly	26.78 mins	3.30 secs
Traffic Hourly	3.01 days	15.74 mins
Traffic Weekly	8.81 mins	8.43 secs
Electricity Hourly	1.78 days	3.07 mins
Electricity Weekly	12.25 mins	3.08 secs
Solar 10 Minutes	21.99 hours	2.66 mins
Solar Weekly	47.86 secs	1.70 secs
Sunspot	3.74 mins	0.18 secs
M1 Yearly	13.20 secs	1.13 secs
M1 Quarterly	51.19 secs	2.56 secs
M1 Monthly	18.01 mins	4.06 secs
M3 Yearly	46.56 secs	3.78 secs
M3 Quarterly	3.01 mins	4.61 secs
M3 Monthly	56.97 mins	13.45 secs
M3 Other	7.81 secs	0.06 secs
M4 Yearly	30.62 mins	5.93 mins
M4 Quarterly	2.59 hours	18.81 mins
M4 Monthly	1.49 days	4.78 hours
M4 Weekly	42.84 mins	26.89 secs
M4 Daily	3.85 hours	15.31 mins
M4 Hourly	50.21 mins	57.47 secs
Pedestrian Counts	19.07 hours	7.18 mins
KDD Cup	5.06 hours	9.93 mins
Car Parts	21.54 mins	37.54 secs
Hospital	20.31 mins	9.69 secs
COVID Deaths	20.03 mins	1.94 secs
FRED-MD	7.51 mins	1.62 secs
Dominick	1.12 days	8.42 hours
Weather	1.34 days	51.55 mins

References

[1] G. Lai, W.-C. Chang, Y. Yang, and H. Liu. Modeling long- and short-term temporal patterns with deep neural networks. In *The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval*, page 95–104, New York, NY, USA, 2018.

[2] Kaggle. `https://www.kaggle.com/`, 2019.

[3] Center for Systems Science and Engineering at Johns Hopkins University. COVID-19 data repository. `https://github.com/CSSEGISandData/COVID-19`, 2020.

[4] S. Makridakis, A. Andersen, R. F. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, and R. L. Winkler. The accuracy of extrapolation (time series) methods: results of a forecasting competition. *Journal of Forecasting*, 1(2):111–153, 1982.

[5] G. Zhang, B. E. Patuwo, and M. Y. Hu. Forecasting with artificial neural networks:: the state of the art. *International Journal of Forecasting*, 14(1):35–62, 1998.

[6] G. P. Zhang. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159 – 175, 2003. ISSN 0925-2312.

[7] R. J. Hyndman and Y. Khandakar. Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 27(3):1–22, 2008. URL `http://www.jstatsoft.org/v27/i03`.

[8] E. S. Gardner. Exponential smoothing: the state of the art. *Journal of Forecasting*, 4(1):1–28, 1985.

[9] G. P. Zhang and M. Qi. Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, 160(2):501 – 514, 2005. ISSN 0377-2217.

[10] S. Makridakis and M. Hibon. The M3-competition: results, conclusions and implications. *International Journal of Forecasting*, 16(4):451–476, 2000.

[11] V. Assimakopoulos and K. Nikolopoulos. The theta model: a decomposition approach to forecasting. *International Journal of Forecasting*, 16(4):521–530, 2000.

[12] H. Hewamalage, C. Bergmeir, and K. Bandara. Recurrent neural networks for time series forecasting: current status and future directions. *International Journal of Forecasting*, 2021.

[13] R. Godahewa, K. Bandara, G. I. Webb, S. Smyl, and C. Bergmeir. Ensembles of localised models for time series forecasting. https://arxiv.org/abs/2012.15059, 2020.

[14] L. J. Tashman. Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*, 16(4):437 – 450, 2000. ISSN 0169-2070.

[15] A. Saxena, J. Celaya, E. Balaban, K. Goebel, B. Saha, S. Saha, and M. Schwabacher. Metrics for evaluating performance of prognostic techniques. In *2008 International Conference on Prognostics and Health Management*, pages 1–17, 2008.

[16] D. M. Dunlavy, T. G. Kolda, and E. Acar. Temporal link prediction using matrix and tensor factorizations. *ACM Trans. Knowl. Discov. Data*, 5(2), February 2011. ISSN 1556-4681.

[17] A. M. De Livera, R. J. Hyndman, and R. D. Snyder. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 106(496):1513–1527, 2011.

[18] J. S. Armstrong. *Evaluating Forecasting Methods*, pages 443–472. Springer US, Boston, MA, 2001. ISBN 978-0-306-47630-3.

[19] J. W. Taylor. Exponential smoothing with a damped multiplicative trend. *International Journal of Forecasting*, 19(4):715–725, 2003.

[20] S. Makridakis, E. Spiliotis, and V. Assimakopoulos. The M4 competition: results, findings, conclusion and way forward. *International Journal of Forecasting*, 34(4):802–808, 2018.

[21] S. Makridakis, E. Spiliotis, and V. Assimakopoulos. The M4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1):54 – 74, 2020. ISSN 0169-2070.

[22] Forecasting & Strategy Unit. Foredeck. `http://fsudataset.com/`, 2019.

[23] S. Smyl. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting*, 36(1):75–85, 2020.

[24] P. Montero-Manso, G. Athanasopoulos, R. J. Hyndman, and T. S. Talagala. FFORMA: feature-based forecast model averaging. *International Journal of Forecasting*, 36(1):86–92, 2020.

[25] K. Bandara, C. Bergmeir, and H. Hewamalage. LSTM-MSNet: leveraging forecasts on sets of related time series with multiple seasonal patterns. *IEEE Transactions on Neural Networks and Learning Systems*, 2019.

[26] E. Spiliotis, A. Kouloumos, V. Assimakopoulos, and S. Makridakis. Are forecasting competitions data representative of the reality? *International Journal of Forecasting*, 36(1):37 – 53, 2020. ISSN 0169-2070.

[27] R. L. Winkler, Y. Grushka-Cockayne, K. C. Lichtendahl, and V. R. R. Jose. Averaging probability forecasts: back to the future. 2018.

[28] R. Godahewa, C. Bergmeir, G. I. Webb, and P. Montero-Manso. A strong baseline for weekly time series forecasting. https://arxiv.org/abs/2010.08158, 2020.

[29] G. Athanasopoulos, R. J. Hyndman, H. Song, and D. C. Wu. The tourism forecasting competition. *International Journal of Forecasting*, 27(3):822–844, 2011.

[30] P. Ellis. Tcomp: data from the 2010 tourism forecasting competition. `https://cran.r-project.org/web/packages/Tcomp`, 2018.

[31] B. Peng, H. Song, and G. I. Crouch. A meta-analysis of international tourism demand forecasting and implications for practice. *Tourism Management*, 45:181 – 193, 2014. ISSN 0261-5177.

[32] N. Kourentzes, F. Petropoulos, and J. R. Trapero. Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting*, 30(2):291 – 302, 2014. ISSN 0169-2070.

[33] H. Song, G. Li, S. F. Witt, and G. Athanasopoulos. Forecasting tourist arrivals using time-varying parameter structural time series models. *International Journal of Forecasting*, 27(3): 855 – 869, 2011. ISSN 0169-2070.

[34] M. Theodosiou. Forecasting monthly and quarterly time series using STL decomposition. *International Journal of Forecasting*, 27(4):1178 – 1195, 2011. ISSN 0169-2070.

[35] M. Akin. A novel approach to model selection in tourism demand modeling. *Tourism Management*, 48:64 – 72, 2015. ISSN 0261-5177.

[36] S. Ben Taieb, G. Bontempi, A. F. Atiya, and A. Sorjamaa. A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. *Expert Systems with Applications*, 39(8):7067 – 7083, 2012.

[37] K. Bandara, C. Bergmeir, and S. Smyl. Forecasting across time series databases using recurrent neural networks on groups of similar series: a clustering approach. *Expert Syst. Appl.*, 140: 112896, 2020. ISSN 0957-4174.

[38] R. R. Andrawis, A. F. Atiya, and H. El-Shishiny. Forecast combinations of computational intelligence and linear models for the NN5 time series forecasting competition. *International Journal of Forecasting*, 27(3):672 – 688, 2011. ISSN 0169-2070.

[39] J. D. Wichard. Forecasting the NN5 time series with hybrid models. *International Journal of Forecasting*, 27(3):700 – 707, 2011. ISSN 0169-2070.

[40] S. Ben Taieb, A. Sorjamaa, and G. Bontempi. Multiple-output modeling for multi-step-ahead time series forecasting. *Neurocomputing*, 73(10):1950 – 1957, 2010. ISSN 0925-2312.

[41] S. Ben Taieb and G. Bontempi. Recursive multi-step time series forecasting by perturbing data. In *2011 IEEE 11th International Conference on Data Mining*, pages 695–704, 2011.

[42] U. Pritzsche. *Benchmarking of Classical and Machine-Learning Algorithms (with special emphasis on Bagging and Boosting Approaches) for Time Series Forecasting*. PhD thesis, 2015.

[43] M. Štěpnička and M. Burda. On the results and observations of the time series forecasting competition CIF 2016. In *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–6, 2017.

[44] T. M. Dantas and F. L. C. Oliveira. Improving time series forecasting: an approach combining bootstrap aggregation, clusters and exponential smoothing. *International Journal of Forecasting*, 34(4):748 – 761, 2018. ISSN 0169-2070.

[45] T. Afanasieva, N. Yarushkina, and I. Sibirev. Time series clustering using numerical and fuzzy representations. In *2017 Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems (IFSA-SCIS)*, pages 1–7, 2017.

[46] V. N. Coelho, I. M. Coelho, I. R. Meneghini, M. J. F. Souza, and F. G. Guimarães. An automatic calibration framework applied on a metaheuristic fuzzy model for the cif competition. In *2016 International Joint Conference on Neural Networks (IJCNN)*, pages 1507–1514, 2016.

[47] Google. Web traffic time series forecasting, 2017. URL `https://www.kaggle.com/c/web-traffic-time-series-forecasting`.

[48] N. Petluri and E. Al-Masri. Web traffic prediction of wikipedia pages. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 5427–5429, 2018.

[49] M. Gupta, A. Asthana, N. Joshi, and P. Mehndiratta. Improving time series forecasting using mathematical and deep learning models. In A. Mondal, H. Gupta, J. Srivastava, P. K. Reddy, and D. V. L. N. Somayajulu, editors, *Big Data Analytics*, pages 115–125, Cham, 2018. Springer International Publishing. ISBN 978-3-030-04780-1.

[50] Z. Mariet and V. Kuznetsov. Foundations of sequence-to-sequence modeling for time series. In K. Chaudhuri and M. Sugiyama, editors, *Proceedings of Machine Learning Research*, volume 89 of *Proceedings of Machine Learning Research*, pages 408–417. PMLR, April 2019.

[51] Solar. Solar power data for integration studies, national renewable energy laboratory, 2020. URL `https://www.nrel.gov/grid/solar-power-data.html`.

[52] UCI. Electricityloaddiagrams20112014 data set, UCI machine learning repository, 2020. URL `https://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014`.

[53] D. Jean-Michel. Smart meter data from London area, 2019. URL `https://www.kaggle.com/jeanmidev/smart-meters-in-london`.

[54] A. Nugaliyadde, U. V. Somaratne, and K. W. Wong. Predicting electricity consumption using deep recurrent neural networks. *ArXiv*, abs/1909.08182, 2019.

[55] A. M. Alonso, F. J. Nogales, and C. Ruiz. A single scalable lstm model for short-term forecasting of disaggregated electricity loads. *ArXiv*, abs/1910.06640, 2019.

[56] M. Maksut, A. Karbozov, M. Myrzaliyeva, H. S. V. S. K. Nunna, P. K. Jamwal, and S. Doolla. Deep learning based short- term load forecasting for urban areas. In *2019 IEEE Industry Applications Society Annual Meeting*, pages 1–6, 2019.

[57] D. Kaur, R. Kumar, N. Kumar, and M. Guizani. Smart grid energy management using RNN-LSTM: a deep learning-based approach. In *2019 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6, 2019.

[58] AEMO. Market data nemweb. `http://www.nemweb.com.au/`, 2020.

[59] R. J. Hyndman. expsmooth: data sets from forecasting with exponential smoothing. `https://cran.r-project.org/web/packages/expsmooth`, 2015.

[60] P. Montero-Manso and R. J. Hyndman. Principles and algorithms for forecasting groups of time series: locality and globality. *International Journal of Forecasting*, 2021. URL `https://arxiv.org/abs/2008.00444`. to appear.

[61] James M. Kilts Center. Dominick's dataset. `https://www.chicagobooth.edu/research/kilts/datasets/dominicks`, 2020.

[62] T. Huang, R. Fildes, and D. Soopramanien. The value of competitive information in forecasting fmcg retail product sales and the variable selection problem. *European Journal of Operational Research*, 237(2):738 – 748, 2014. ISSN 0377-2217.

[63] D. Toro-González, J. J. McCluskey, and R. Mittelhammer. Beer snobs do exist: estimation of beer demand by type. *Journal of Agricultural and Resource Economics*, 39(2):174–187, 2014.

[64] A. Jami and H. Mishra. Downsizing and supersizing: how changes in product attributes influence consumer preferences. *Journal of Behavioral Decision Making*, 27(4):301–315, January 2014. ISSN 0894-3257.

[65] M. Eichenbaum, N. Jaimovich, and S. Rebelo. Reference prices, costs, and nominal rigidities. *American Economic Review*, 101(1):234–262, February 2011.

[66] B. Guimaraes and K. D. Sheedy. Sales and monetary policy. *American Economic Review*, 101 (2):844–76, April 2011.

[67] M. W. McCracken and S. Ng. FRED-MD: a monthly database for macroeconomic research. *Journal of Business and Economic Statistics*, 34(4):574–589, 2016.

[68] Caltrans. Caltrans performance measurement system, California department of transportation, 2020. URL `http://pems.dot.ca.gov`.

[69] City of Melbourne. Pedestrian counting system – 2009 to present (counts per hour), 2017. URL `https://data.melbourne.vic.gov.au/Transport/Pedestrian-Counting-System-2009-to-Present-counts-/b2ak-trbp`.

[70] E. Carter, P. Adam, D. Tsakis, S. Shaw, R. Watson, and P. Ryan. Enhancing pedestrian mobility in smart cities using big data. *Journal of Management Analytics*, 7(2):173–188, 2020.

[71] H. O. Obie, C. Chua, I. Avazpour, M. Abdelrazek, and J. Grundy. Visualising melbourne pedestrian count. In *2017 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pages 343–344, 2017.

[72] H. O. Obie, C. Chua, I. Avazpour, M. Abdelrazek, J. Grundy, and T. Bednarz. Pedaviz: visualising hour-level pedestrian activity. In *Proceedings of the 11th International Symposium on Visual Information Communication and Interaction*, VINCI '18, page 9–16, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450365017.

[73] E. Dong, H. Du., and L. Gardner. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Inf Dis.*, 20(5):533–534, 2020.

[74] KDD2018. Kdd cup 2018, 2018. URL `https://www.kdd.org/kdd2018/kdd-cup`.

[75] Z. Luo, J. Huang, K. Hu, X. Li, and P. Zhang. AccuAir: winning solution to air quality prediction for kdd cup 2018. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '19, page 1842–1850, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450362016.

[76] A. H. Sparks, J. Carroll, J. Goldie, D. Marchiori, P. Melloy, M. Padgham, H. Parsonage, and K. Pembleton. *bomrang: Australian Government Bureau of Meteorology (BOM) Data Client*, 2020. URL `https://CRAN.R-project.org/package=bomrang`.

[77] Sunspot. Sunspot number version 2.0: new data and conventions, 2015. URL `http://www.sidc.be/silso/newdataset`.

[78] A. Gkana and L. Zachilas. Re-evaluation of predictive models in light of new data: sunspot number version 2.0. *Solar Physics*, 291, August 2016.

[79] L. Bertello, A. Pevtsov, A. Tlatov, and J. Singh. Correlation between sunspot number and ca ii k emission index. *Solar Physics*, 291, June 2016.

[80] A. Shapoval, J. Le Mouel, M. Shnirman, and V. Courtillot. Dynamics of sunspot series on time scales from days to years: correlation of sunspot births, variable lifetimes, and evolution of the high-frequency spectral component. *Journal of Geophysical Research: Space Physics*, 122, December 2017.

[81] G. Pease and G. Glenn. Long term sunspot cycle phase coherence with periodic phase disruptions. October 2016.

[82] A. I. McLeod and H. Gweon. Optimal deseasonalization for monthly and daily geophysical time series. *Journal of Environmental Statistics*, 4(11), 2013. URL `http://www.jenvstat.org/v04/i11`.

[83] A. Bauer, M. Züfle, N. Herbst, S. Kounev, and V. Curtef. Telescope: an automatic feature extraction and transformation approach for time series forecasting on a level-playing field. In *2020 IEEE 36th International Conference on Data Engineering (ICDE)*, pages 1902–1905, 2020.

[84] R. Pruim, D. Kaplan, and N. Horton. *mosaicData: project MOSAIC Data Sets*, 2020. URL `https://CRAN.R-project.org/package=mosaicData`.

[85] R. J. Hyndman. *fpp2: data for "Forecasting: Principles and Practice" (2nd Edition)*, 2018. URL `https://CRAN.R-project.org/package=fpp2`.