Improved energy demand forecasting using rooftop photovoltaic power data

Paul Cronin[1], Rahul Kumar[1], Raphael Muema[1], Patrice Santiago[1], and Mathew Traini[1]

[1] Department of Mathematics and Statistics, University of New South Wales, Sydney, 2052, Australia.

Abstract. The rapid growth in small-scale rooftop photovoltaic (PV) installations in Australia represents a paradox in the sun-soaked country's transition to renewable energy. Solar energy is emission-free, renewable, and forecast to supply a majority of Australia's peak energy demands by 2030. However, PV output is instantaneously affected by weather conditions and is therefore highly variable. This presents significant challenges for the accurate forecasting of total residual energy demand that must be met by the grid, which is critical for ensuring reliable and secure operation of the national electricity supply system.

To address this issue, this report explores the value of including rooftop-PV generation data in forecasts of energy demand in the Australian electricity market. Australian state-level aggregated rooftop-PV generation data was obtained and integrated into linear regression, temporal fusion transformer (TFT) and neural basis expansion analysis for time series (N-BEATS) models of day-ahead total energy demand forecasting. Inclusion of PV data improved forecast accuracy mean average percentage error (MAPE) of NSW January energy demand by 0.70%, 4.22% and 0.73% for the regression, N-BEATS and TFT models respectively, versus the same models trained without PV data. The linear regression model, using an optimised time-decay weighting matrix, produced the overall most accurate forecasts of energy demand on the NSW summer test set (MAPE: 3.01%). These findings demonstrate the importance of considering rooftop-PV production in forecasting total energy demand, in the context of both classical statistical and modern deep-learning methods.

1 Introduction

Unlike most commodities electrical energy cannot be easily stockpiled, and so its production must be closely balanced with instantaneous demand. Accurate energy demand forecasts are critical for achieving this balance, as well as for longer-term electrical grid infrastructure and resilience planning. In Australia, the Australian Energy Market Operator (AEMO) is the statutory body responsible for overseeing the operation and security of the interconnected electrical grids of the states NSW, Victoria, Queensland, South Australia and Tasmania [1], collectively referred to as the National Electricity Market (NEM) [2]. AEMO produces a series of short- and long-term energy demand

forecasts, which are used by NEM electricity producers to efficiently schedule generator operation and purchases of raw inputs to meet upcoming energy requirements. Energy demand forecasts must be highly accurate, as under- or overgeneration of electricity by NEM producers carries significant economic, societal, and environmental costs [3].

Not all electrical demand is met by the output of NEM-connected power plants centrally coordinated by AEMO. Australia has the highest per-capita installed capacity of small-scale rooftop photovoltaic (PV) solar panels in the world [4]. Because of these high uptake rates, together with Australia's naturally high solar irradiance [5], rooftop PV is forecast to deliver up to 77% of peak NEM demand by 2026 [6]. Rooftop PV generation (an example of "behind-the-meter" energy production [7]) directly substitutes a portion of energy demand which would otherwise be supplied from the grid. However, PV output is highly variable and dependent on minute-to-minute changes in weather. This means that sudden, large drops in behind-the-meter generation due to unpredictable cloud coverage may result in equally sudden variance in grid energy demand and stability [8]–[10].

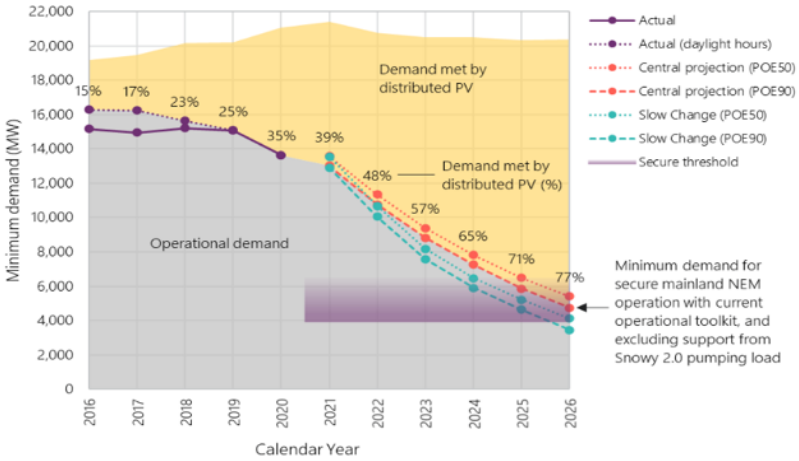


Fig. 1. Long-term forecasts of NEM operational demand, distributed PV production and minimum demand thresholds for secure NEM grid operation. (Adapted from [6])

AEMO has identified Australia's high rooftop-PV output capacity and its inherent variability as creating significant challenges for existing energy demand forecast models [6], particularly for short-term predictions which NEM stakeholders rely upon for day-to-day operational management. The rise of Australian behind-the-meter capacity is accelerating already falling energy demand trends, and demand is forecast to drop below minimum critical levels required to ensure stability of the NEM grid by 2026 [6], [11] (**Fig. 1**).

The pivotal role of rooftop PV generation in the future of Australia's electrical grid requires an evolution in how energy demand forecasting is conceptualised. "Net energy

demand" refers to the fusion of household energy requirements met by behind-the-meter generation, together with any residual unmet need that must be instantaneously delivered by the electrical grid [12]. By including the contribution of PV and other behind-the-meter technologies, net energy demand forecast models are expected to produce more accurate and meaningful predictions, particularly in solar-dense markets such as Australia [13]. Although significant rooftop PV capacity and associated real time online smart-metering technology has been installed globally in the last decade [14], net energy forecasting is still an emerging area, with most published studies remaining focused on classical grid-demand models based on predictors such as day-of-week and temperature.

Net energy forecasting studies to date have ranged from individual households [13], private commercial buildings [15], university campuses [16], 75-house residential neighbourhoods [17], to large population-level models integrating aggregated PV penetration rates with whole-grid demand load forecasts [18], [19] (**Table 1**).

Table 1. Summary of PV-aware and net energy demand forecasting studies

Forecast target	**Model types**	**Data sources**	**Citation**
Individual and aggregated household grid energy demand and net energy demand	Online and batch-trained Long short-term memory (LSTM)	65 Australian household smart-meter energy demand and PV output measurements	[13]
Household net energy demand	Multiple linear regression, multi-layer perceptron (MLP)	Simulated PV output using British weather forecasts	[19]
Very short-term (15 min) small neighbourhood (75 houses) net energy demand	ARIMAX, MLP	Household smart meter direct measurements; weather station solar irradiance data.	[17]
Net energy demand (24 hours) of a single commercial building	Hybrid MLP-marine predators algorithm	Building environmental sensors, smart meters, private thermal and PV output measurments	[15]
PV farm energy production and net energy demand (24 hours ahead) for university campus	ARIMAX, random forest, MLP	Direct PV-farm energy outputs; Building smart-meters	[16]

Most studies have focused on net energy demand of individual buildings or relatively small neighbourhoods, a scale at which fluctuations in the local micro-climate cause large variability in PV output measurements. Despite this confounding effect, inclusion of PV data resulted in more accurate estimates of underlying energy demand in most

cases, especially during the middle of the day when the contribution of PV generation was at a maximum.

Modelling approaches in the nascent field of net energy demand prediction broadly reflect established methods in time-series prediction. Parametric statistical approaches, such as multiple regression or ARIMAX are commonly used for establishment of baseline metrics (**Table 1**). These methods have the advantage of relatively low computational burden for training and tuning, and yield explainable, easily interpretable parameters and predictions. Artificial neural network (ANN) methods such as MLPs and LSTMs commonly eclipsed the accuracy of statistical methods for PV-aware net energy demand forecasting. However, such architectures no longer represent state-of-the-art in time-series modelling, with newer designs such as Temporal Fusion Transformers (TFT) and neural basis expansion analysis for interpretable time series (N-BEATS) having recently established benchmark accuracies in the M4 time-series forecasting competition [20]. The net energy demand and PV-aware forecasting task has created an opportunity to exploit the TFT and N-BEAT architectures in a novel domain. The following section reviews the characteristics of these newer approaches, and reviews their recent use in energy demand and production forecasting.

1.1 Temporal Fusion Transformers (TFT)

Temporal Fusion Transformers (TFT) [21] are attention based Deep Neural Network (DNN) architectures for the prediction of variables-of-interest at multiple future time steps (multi-horizon forecasting), achieving high performance while enabling new forms of interpretability. Practical multi-horizon forecasting applications commonly have access to a variety of data sources (**Fig. 2**), including known information about the future (e.g. upcoming holiday dates), other exogenous time series (e.g. historical temperature data), and static metadata (e.g. location of the weather station). Such heterogeneity of data sources, together with little prior knowledge of how they interact, makes multi-horizon time series forecasting particularly challenging [21].

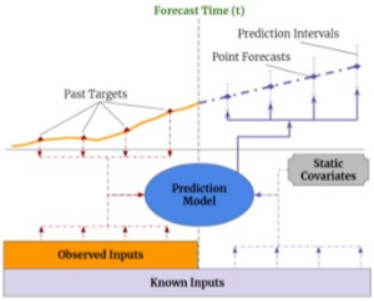


Fig. 2. Illustration of multi-horizon forecasting with static covariates, past-observed and apriori-known future time-dependent inputs (adapted from Lim et al., 2021).

The major constituents of TFTs (**Fig. 3**) are 1) gating mechanisms, 2) variable selection networks, 3) static covariate encoders, 4) temporal processing and 5) prediction intervals. TFT inputs static metadata, time-varying past inputs and time-varying *a priori* known future inputs. Variable selection is used for judicious selection of the most salient features based on the input. Gated information is added as a residual input, followed by normalization. Gated residual network (GRN) blocks enable efficient information flow with skip connections and gating layers. Time-dependent processing is based on LSTMs for local processing, and multi-head attention for integrating information from any time step.

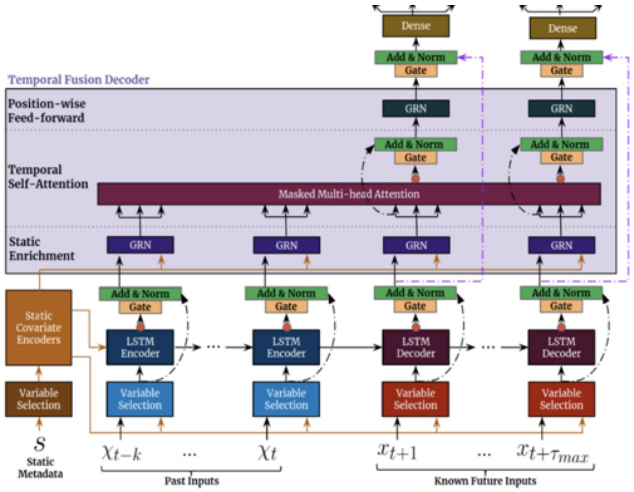


Fig. 3. Major constituents of the Temporal Fusion Transformer architecture (adapted from [22])

TFTs outperformed comparable methods such as DeepAR, ARIMA, and LSTM Seq2Sep on standard time-series benchmark datasets, with 50% and 90% quantile losses at least 7% lower than the next best model (**Fig. 4**) [22].

Model	Electricity	Traffic	Volatility	Retail
ARIMA	0.154 (+180%)	0.223 (+135%)	-	-
ETS	0.102 (+85%)	0.236 (+148%)	-	-
DeepAR	0.075 (+36%)	0.161 (+69%)	0.050 (+28%)	0.574 (+62%)
Seq2Seq	0.067 (+22%)	0.105 (+11%)	0.042 (+7%)	0.411 (+16%)
MQRNN	0.077 (+40%)	0.117 (+23%)	0.042 (+7%)	0.379 (+7%)
TFT	**0.055**	**0.095**	**0.039**	**0.354**

P50 quantile losses (lower is better) for TFT vs. alternative models.

Fig. 4. Summary of TFT model performance across standard benchmark datasets versus other deep learning architectures (adapted from [22])

Although relatively new, TFTs have demonstrated promising performance in the field of energy demand forecasting, matching the accuracy of a highly optimized LSTM [23], and ranking among the highest performing models for predicting continuous hourly building energy demand [24].

1.2 Neural basis expansion analysis for interpretable time series (N-BEATS)

Neural basis expansion analysis for interpretable time series (N-BEATS) is another recently developed deep learning architecture, specifically designed to address existing shortcomings of neural networks for univariate time-series prediction [20].

N-BEATS models take several time steps as their input (lookback period, **Fig. 5A**), which is passed into the basic subunit of N-BEATS (the 'block', **Fig. 5B**), consisting of a fully-connected multi-layered network with ReLU activation. Block outputs are passed through two separate fully connected layers, with waveform generator functions producing a future forecast and an historical backcast (corresponding to the best-fit estimate of the original input, **Fig. 5B**). Blocks are arranged into 'stacks' (**Fig. 5C**) with a double-residual arrangement of skip connections. This arrangement allows the backcast prediction from one block to be subtracted from the input of the next block, thus modelling the residual error. Multiple stacks are combined (**Fig. 5D**), and their individual outputs aggregated to produce a final forecast horizon of a pre-specified number of time steps (**Fig. 5E**). This pure deep-learning architecture, similar to an unrolled recurrent neural network, allows N-BEATS to solve non-linear time series problems while maintaining relatively fast training times compared to other architectures.

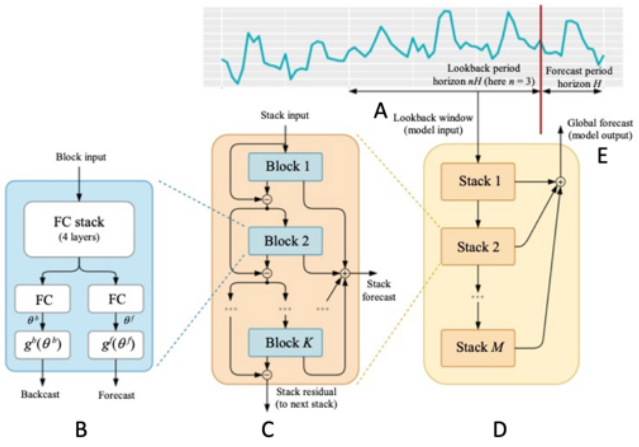


Fig. 5. N-BEATS architectural overview (adapted from [20])

N-BEATS models have been successfully applied to energy demand and production models over a range of forecasting intervals. Very short-term prediction of wind farm power output from 15 different countries with N-BEATS achieved benchmark accuracy

performance on average [25], as did as medium-term forecasting of energy load for 35 different countries [26]. Recently, the original N-BEATS architecture has been extended to allow addition of exogenous predictors during training. Inclusion of day-ahead energy load and wind power forecast variables significantly enhanced the accuracy of short term (24 hours ahead) prediction of energy demand on the standard Nordic Price NP dataset, versus the original N-BEATS design [20]. Support for exogenous time-dependent predictors has been added to open-source deep learning libraries such as PyTorch and darts [27], providing support essential for testing the hypothesis that inclusion of exogenous PV power data will increase energy demand prediction accuracy.

1.3 Aims of this report

Given the intertwined nature of grid demand and rooftop-PV production, and the challenges this creates for accurate energy demand forecasting outlined above, it is proposed that:

Integration of rooftop-PV power data into Australian energy demand forecast models will improve their accuracy.

This hypothesis will be tested according to the research plan:

1. Obtain Australian energy demand and rooftop-PV power data.
2. Prepare baseline, statistical and state-of-the-art deep learning models to forecast energy demand, with or without integration of rooftop-PV data.
3. Assess if integration of rooftop-PV data improves model forecast accuracy.

2 Materials and Methods

2.1 Core software and libraries

All data ingestion, cleaning, EDA, and neural network (N-BEATS/TFT) model building was performed using Python. Jupyter notebooks (.ipynb files) were used for presentation of code output and graphs for EDA. Linear regression models were developed using R. Supporting libraries for each language are summarised in Appendix Table 1. A listing of all notebooks developed to support this project, and their Github URLs is compiled in Appendix Table 2.

2.2 Datasets

Sponsor supplied datasets – Energy demand, forecasts and temperatures.
Data files containing were provided by the project sponsor over the 11 year period of 2010 to 2021 at 30 minute intervals. The first dataset is the Forecast Demand, a data set containing the AEMO forecast made for energy demand at the timestamp for the 30 minute interval periods till 4AM 48 hrs from the time stamp. The second dataset is the

Total Demand, a data set containing the historical energy demand data for the time period (**Table 2**). The final dataset being the Temperature dataset, which contains historical temperature for the time period. Separate datasets are provided for the states NSW, QLD, SA and VIC. Temperatures are recorded at a single location for each state, which is representative of that state's largest geographical source of energy demand.

Table 2. Field definitions and descriptions of demand, temperature and forecast data files supplied by the project sponsor. Based on standard definitions used by AEMO [28]

Field	Type	Description
Total Demand		
DATETIME	Datetime	Estimation half-hour interval (time ending)
TOTALDEMAND	Float	Recorded Energy Demand (MW) at interval end
REGIONID	String	Region (state) identifier
Temperature		
LOCATION	String	Location where Temperature was taken
DATETIME	Datetime	Estimation half-hour interval (time ending)
TEMPERATURE	Float	Temperature Recording
Forecast Demand		
PREDISPATCHSEQID	Integer	Identification Number
REGIONID	String	Region (state) identifier
PERIODID	Integer	Periodic Id Reference number for each half-hour interval
FORECASTDEMAND	Float	Estimated Forecast Energy Demand (MW) at interval end
LASTCHANGED	Datetime	Forecast Estimation half-hour interval (time ending)
DATETIME	Datetime	Estimation half-hour interval (time ending)

As each dataset was provided separately, the primary step was to ensure the data was combined into a single entity for EDA. As the data was provided in csv formats, the DATETIME column had to be converted to datetime format. Initially the Demand dataset was merged with the temperature dataset using outer join on STATE (a generated column that identified each state) and DATETIME to ensure even unmatched rows of both data sets remain. This yields a dataset that has 842,820 entries and 5 columns. Similarly, the Forecast data was first aggregated by mean for each DATETIME entry. This was similarly joined using an outer join on STATE and DATETIME. This resulted in a dataset that also contained 842,820 entries and 6 columns.

Aggregated Australian state rooftop PV production data

AEMO operates the Australian Solar Energy Forecasting System (ASEFS) [29], which generates half-hourly estimates of actual total power production for both large solar power farms (output $\geq$ 30MW), and smaller rooftop scale systems (output $\leq$ 100kW) for the states NSW, QLD, VIC, SA and TAS. The latest iteration of this system

(ASEFS phase 2) produces two separate datasets of half-hourly PV production estimates. One is based on satellite imaging data and installed rooftop PV data from the Clean Energy Regulator (CER) [30] (referred to as 'SATELLITE'). The second is based on collation of an Australia-wide sample of daily solar analytics and domestic smart-meter readings, combined with up-scaling based on CER installed PV capacity data (referred to as 'MEASUREMENT'). As the 'MEASUREMENT' methodology offers a more direct estimation of total PV power production, only this dataset was considered for this study.

PV data availability, date ranges and format

AEMO makes all ASEFS data freely available for download from its website [31]. ASEFS phase 2 half-hourly PV data obtained for this project spans the period 9:30AM 6th March 2018 to 3:30PM 5th March 2022. Time and date intervals in the ASEFS PV estimates are exactly harmonised with those of the AEMO electricity demand and forecast data used in this project, ensuring reliable time-matching of datasets. The data is distributed by AEMO as multiple plain-text .CSV files with fields defined as in **Table 3**, compressed either singly or multiply into .ZIP archives, and stored in a web-server directory hierarchy arranged by date. Dates covered in each .CSV file vary depending on the age of the data, with older historical data being aggregated into larger monthly summary files.

Table 3. Field definitions and descriptions of AEMO ASEFS actual PV power estimates

Field	Type	Description
INTERVAL_DATETIME	Datetime	Estimation half-hour interval (time ending)
REGIONID	Text	Region (state) identifier
POWER	Float	Estimated PV power output (MW) at interval end
TYPE	Text	One of DAILY, MEASUREMENT or SATELLITE

PV data processing

The Unix 'wget' utility was used to automatically download all ASEFS phase 2 PV data .ZIP files to the local workstation, files were unzipped, and individual .CSV files concatenated into a single large .CSV with the Unix shell commands 'unzip', 'head' and 'tail' piped together. The Python notebook 'rooftop_pv.ipynb' (Appendix Table 2) loads the concatenated .CSV files into a Pandas dataframe, captures rows containing only the 'MEASUREMENT'-type observations, sorts by DATETIME in ascending order, and writes an output file in .CSV format.

As the original data was distributed among several thousand separate .CSV files in multiple directories on the AEMO web server, particular attention was paid to identifying missing ranges of data. Where these were identified (using standard Pandas functions) missing source .CSV files were manually found, downloaded from AEMO, and re-processed using the 'rooftop_pv.ipynb' notebook. After processing, no missing

POWER values were found to exist in the 9:30am 6th March 2018 – 3:30pm 5th March 2021 dataset.

2.3 Missing Data

Certain temperature observations in the client-supplied dataset were made at times not synchronised with the fixed half-hourly intervals of the AEMO energy demand, forecast and PV. Other observations at standard intervals were missing. As total energy demand is the key predictive target in this project, temperature data not perfectly matching a date and hourly/half-hourly time interval of a TOTALDEMAND record was discarded from the analysis. Where possible the MeteoStat meteorological database was used to obtain other missing values [32]. The Python MeteoStat API was used to determine if a temperature reading was available at each specific datetime, and at the same geographical site at which temperatures were originally observed (implemented in the 'temp_cleaner_multistate.ipynb' Python notebook, Appendix Table 2). All remaining missing temperature values consisted of short runs of 3-4 contiguous datetime periods. These missing values were filled by roll-forward filling of the most recent known temperature value via standard Pandas functions.

2.4 New Features

Several additional features to assist with EDA or model training were engineered. Features corresponding to day-of-week, month and quarter-of-year were derived from the DATETIME field using standard Pandas functions. Season was determined according to [33]. Boolean fields for day/night and public holiday status were derived from sunrise and sunset data for each state [34] and using the Python 'holidays' package [35], respectively.

2.5 Designation of final core project dataset

As the central hypothesis of this study was to examine the effect of including PV data as an additional covariate, it was determined that only data which had a full matching set of PV 'POWER' values would be considered in the scope of the report. This encompassed the range 9:30AM 6th March 2018 to 3:30PM 5th March 2021 inclusive for the states NSW, VIC, QLD and SA (see 'PV data availability, date ranges and format'). This was designated the 'core' dataset and used for all EDA. For modelling, the short timeline of this project necessitated a more focused approach. *Therefore, all models were trained and assessed using only the NSW subset of the core dataset.*

2.6 Training and testing strategy

Roll-forward validation testing

Conventional random shuffling and splitting of data for cross-validation destroys the inherent temporal ordering of time-series data, and was considered unsuitable for this

study. Roll-forward validation (also known as "time cross-validation") testing was used for final evaluation of all model types. This is where historical data up to and inclusive of the current time T was used to train the model, and then unseen demand at time T+1 is forecast. The training window is then expanded by an interval, the model re-trained, a further prediction made, and so on (**Fig 6**).

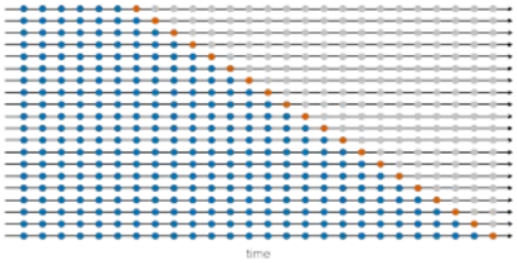


Fig. 6. Graphical depiction of roll-forward testing methodology. Blue dots indicate timepoints used for training up to current time T, orange dots indicate the T+1 forecast target. Grey dots represent timeseries data yet to be forecast or trained upon. Figure reproduced from [36].

Selection of training and testing sets

A split of the entire PV-supplemented dataset available for training all models was defined in the range 9:30AM 6th March 2018 to 11:30PM 31st December 2020. A hold-out testing split was defined between 12:00AM 1st January and 18th March 2021. As roll-forward retraining validation is computationally expensive for certain model types (especially deep learning models), a minimum test subset was chosen from the larger testing split. 1st-31st January 2021 was selected, as this one-month range contains several maximum and minimum outlier events in the full test set, which may be useful for 'stress-testing' and identifying model shortcomings. These include two public holidays (1st and 26th January), days with maximum and minimum daily demand (25th and 1st January respectively), largest intra-day change in demand (January 26th), maximum and minimum temperatures (26th and 21st January respectively), and the largest intra-day temperature change (26th January).

N-BEATS validation and testing

For the N-BEATS neural network model, validation after parameter tuning was performed by historical backforecasting. Backforecasting is where historical forecasts that would have been made by the model at previous timepoints in the training set are calculated (implemented using the darts *historical_forecasts()* method [37]). Historical backforecasting was performed on the last 12 months of the training set (1st January – 31st Deember 2020) and used to calculate mean average percent error (MAPE) and root mean-squared error (RMSE) scores for training error analysis. Hold-out testing of N-BEATS models was performed using a roll-forward expanding window testing strategy (**Fig 6**). While this testing strategy closely mimics the situation that may be encountered when a timeseries forecasting model is 'in production' (ie: the model is regularly re-

trained as new time data becomes available to improve the accuracy of the next forecast), repeated N-BEATS re-training is computationally expensive. Therefore the roll-forward testing window was restricted to the minimum test subset 1st January to 31st January 2021. In addition to calculating overall accuracy measures such as RMSE and MAPE, detailed day-by-day error plots were also produced to understand where the model was most and least successful.

2.7 Naïve baseline model

A naïve drift model was constructed to act as a comparison baseline, which for any given forecast date and time, simply reports the TOTALDEMAND value from exactly 24 hours (48 time intervals) prior.

2.8 Linear regression models

Two linear regression models were developed, which would allow for:

1. Explaining the historical AEMO data
2. Generation of synthetic data to explore future trends
3. Short-term (half-hour) forecasts

These two models were a simple linear model (LM) [38][39] and a generalized linear model (GLM) [40].

Linear model (LM)

PV data was integrated into the linear models by combining it with the dependent variable TOTALDEMAND, to generate a new target variable. Some new target variable terms will be introduced here: 'GeneratedSupply' is defined as the original state-wide TOTALDEMAND data supplied from the project sponsor. 'ActualDemand' is defined as the 'GeneratedSupply' value (ie: TOTALDEMAND), plus the state-wide aggregated PV 'POWER' value at the same time point. 'ActualDemand' is a net-energy demand metric [13], and represents a close proxy to the actual underlying total energy demands of the state. As such, it is expected to not be affected by changes in solar irradiance throughout the day and therefore have reduced overall variance versus 'GeneratedSupply'. Hence the hypothesis is that modelling 'ActualDemand' (ie: incorporating PV 'POWER') will result in a more accurate long term and short-range modelling.

The linear model developed is a log-linear model as shown in Equation 1:

$$\log(Demand) = \beta_0 + \beta_1 Time + \beta_2 Time^2 + \beta_3 Time^3 + \beta_4 Temp + \beta_5 Temp^2 + \beta_6 Temp^3 + \beta_7 DayIndex + \beta_8 DayName + \beta_9 Season + \beta_{10} PublicHoliday \quad (1)$$

where *Demand* can be either *GeneratedSupply* or *ActualDemand*. *Time* is the time-step index from 1st January 2020. *Temp* is the temperature measured in degrees Celsius.

DayIndex is the categorical variable for each of the 48 half-hourly intervals during the day. *DayName* is the categorical variable for each of the 7 days of the week. *Season* is the categorical variable for the four annual seasons, and *PublicHoliday* is a binary categorical indicating if the day is a public holiday.

2.9 Considerations on the use of 'future' versus 'present' data for forecasting

In discussions with the project sponsor, it was implied that some future (ie: T+1) data might be legitimately used, such as the temperature domestic PV demand at the future forecast timepoint. However, for a fair comparison between the existing AEMO forecast data and the models presented here, it was decided that no future data would be used to make forecasts.

This simple linear model was tuned in slightly different ways for exploring the different aspects of the project. For example, when explaining what factors are historically important, there was no preferential weighting of the data, but an exponential weighting of the data was used for short term (half-hour) roll-forward forecasting – see Equation 2:

$$w_i = exp\left[{(i - N)}/{\tau}\right] \qquad (2)$$

This allowed more recent historical data to be more heavily weighted when performing short-term (half-hour) roll-forward forecasts. To determine the optimal weighting decay rate tau, an experiment was undertaken. For different values of the decay rate tau, the MAPE was measured and plotted, and an optimal decay rate for forecasting half-hour ahead of 256 time periods (~5 days) was determined (**Fig. 7**)

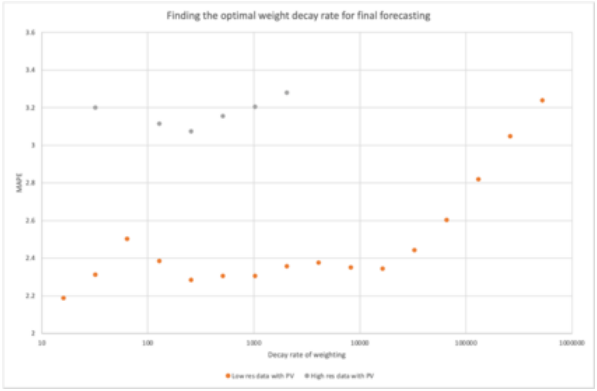


Fig. 7. Weighting decay rate versus MAPE Experiment to determine optimal decay rate of weighting for forecasting

As future information is not being used, it was necessary to predict covariates such as temperature. A further experiment was performed to determine which was the best

method of forward temperature prediction, and how many time periods were optimal. Average, linear, quadratic and cubic covariate prediction models were explored, and the optimal was found to be a linear predictor using only 10 time-periods (equivalent to 5 hours). In this manner, the various covariates were determined for the roll-forward analysis (**Fig. 8**).

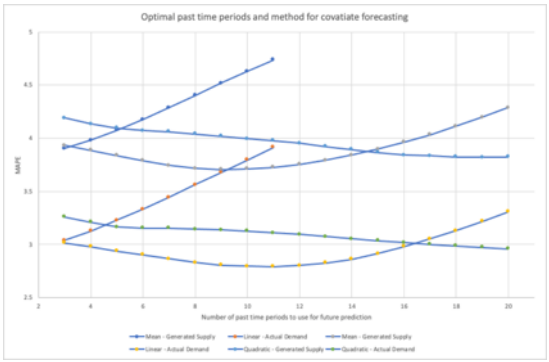


Fig. 8. Optimal model for temperature prediction in the next time slot. Optimal method was found to be a linear predictor from the last 5 hours (10 half-hour measurements).

2.10 Generalised Linear Model

The Generalised Linear Model was defined in Equation 3:

$$Demand \sim (Date + Date^2 + Date^3 + Temp + Temp^2 + Temp^3) \times (DayIndex + DayName + Season + PublicHoliday) \quad (3)$$

Each of the target and predictor variable definitions are the same as the linear model. The non-linear expression of date and temperature arises from the EDA-derived insight that energy demand increases in summer (with rising temperature), but also in winter (to a lesser extent). A third order temperature model was found to best fit these results. Likewise, the long-term trend of the demand also exhibited a non-linear effect. A trend-seasonality [41] measurement was undertaken during the data exploration period, indicating a good fit with a third order trend.

2.11 Neural Basis Expansion Analysis for Interpretable Time Series (N-BEATS) model

The Python Darts library implementation of N-BEATS was adopted [27], which provides a user-friendly wrapper around the PyTorch Lightning API. Due to time constraints, parameter optimisation was focused only on tuning the input chunk length ('lookback period', **Fig. 5**). This parameter controls the number of time periods that N-BEATS uses as its lookback period for training. Lookback periods corresponding to 3 days (144 half-hour intervals), 1 week (336), 2 weeks (672), 1 month (1440) and 2

months (2880) were examined, with a fixed forecast period of 1 day (ie: 48 intervals predicted at each step). Model training accuracy was evaluated by backforecasting over the most recent 1 year of the training data and calculating RMSE and MAPE. The number of training epochs (3, 5 or 10) was also evaluated in this manner.

N-BEATS models were trained using TOTALDEMAND as the target, and with a common set of past covariates consisting of day-of-the-week, month, and public holidays. To identify the effect of PV data on forecast accuracy, the rooftop-PV POWER variable was included as an additional covariate and a separate 'PV-aware' model trained and saved. A full listing of hyperparameters is contained in Appendix Table 3.

2.12 Temporal Fusion Transformer (TFT) model

The pytorch forecasting Temporal Fusion Transformer was adopted ([21]), which also provided a user-friendly wrapper around the PyTorch Lightning implementation. Data used for training the TFT model was the NSW subset of the core dataset (from April 2018 to December 2021). The training was done on a month's worth of data and prediction was done for the last day of the month (48 intervals). This simple model was chosen as the focus of the analysis and was the comparison of the results after inclusion of PV data.

The TFT model was trained using TOTALDEMAND as the target and weekday, season, day type and holiday as the time-varying known categories, while quarter, month, day, temperature and power as the time-varying known reals. For the model including PV data, the POWER variable was also included. Both non-PV and PV models were trained for a maximum of 200 epochs with early stopping enabled when the validation loss stopped improving. Optimal hyperparameters were obtained using the optimise_hyperparameters function from the TFT library (Appendix Table 3).

3 Exploratory Data Analysis

3.1 Distribution and outlier analysis of core data

The core dataset from 9:30AM 6th March 2018 to 3:30PM 5th March 2021 contained no missing values for the key variables TOTALDEMAND, POWER or TEMPERATURE. The distribution of TOTALDEMAND and TEMPERATURE were approximately normal for all states, although a right-skew in demand is evident for NSW, QLD and VIC. The distribution of PV POWER is heavily biased toward zero values, reflecting the majority of daily hours where there is insufficient or no sunlight present to drive solar production (**Fig. 9**).

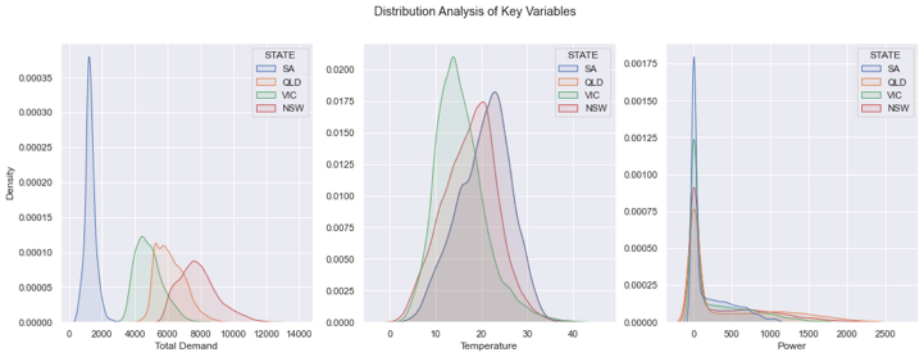
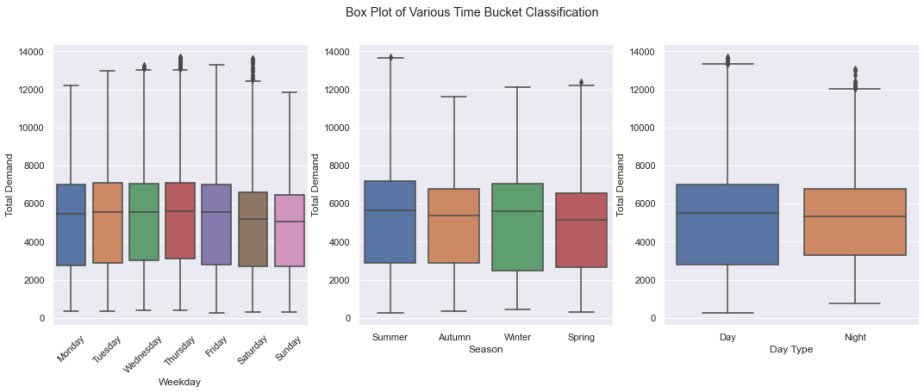


Fig. 9. Kernel density estimates of sample distributions of A: Total demand B: Temperature and C: Photovoltaic power for the states NSW, QLD, SA and VIC over the period March 2018-2021

Based on a 1.5 times interquartile range threshold, some outliers in Total demand are apparent (**Fig. 10**).

Fig. 10. Boxplots of TOTALDEMAND versus day-of-week, season and day type. Whiskers indicate 1.5 x interquartile range.



3.2 Correlation analysis

Calculating a Variable Correlation Heatmap can be a great indicator of potential variables useful in modelling. It can be seen that Forecast Demand has a 1 correlation (**Fig. 11**); however, this wouldn't be very useful in forecasting as it would be equivalent of knowing the answer the model needs to forecast prior to forecasting. The next highest correlated variable is Power, which supports the argument that understanding the power generation of Photovoltaic Cells would help understand the total demand generated. Finally, the next correlated variable would be the Temperature, however based on the

value this can be assumed to be completely random, however a deep dive into the relation between Temperature and Total Demand is done in the following section. All other time derived variables are shown to be uncorrelated.

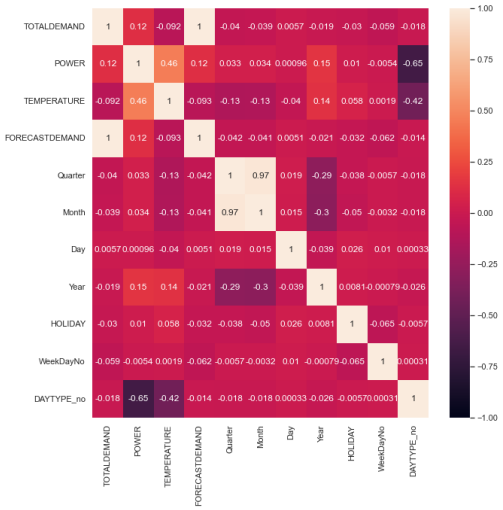


Fig. 11. Correlation heatmap of variables with TOTALDEMAND

3.3 Temperature and Total Demand analysis

Overall daily energy demand has a flat-to-decreasing trend, despite increasing levels of population and economic activity in Australia over the same period (**Fig. 12**). Distinctive seasonal patterns in total energy demand are apparent, with bi-annual peaks occurring mid-winter and mid-summer for all states except Queensland. Winter demand in Queensland remains relatively flat compared to the other states, potentially due to a comparatively lower requirement for heating due to warmer average winter temperatures. Notably, variance in maximum daily energy demand is greater during the summer demand peak versus winter. This may reflect sharp spikes in energy demand due to increased cooling system use on extremely hot days.

Consistent with this hypothesis, several temperature spikes appear perfectly aligned with corresponding peaks in daily demand, with examples of this phenomenon present in the plots for all states. Taken together, this EDA suggests a strong relationship between extremes of temperature and energy demand which might be exploitable in a forecasting model.

Despite this apparent trend, there is little overall correlation between demand and temperature (Pearson correlation coefficient -0.092). Stratifying demand by state and season revealed the expected positive correlation during warmer months (particularly in NSW), which became weakly negative in cooler months (**Fig. 13**). This suggests that temperature may be a useful predictor of demand, but only when considered in conjunction with month and state.

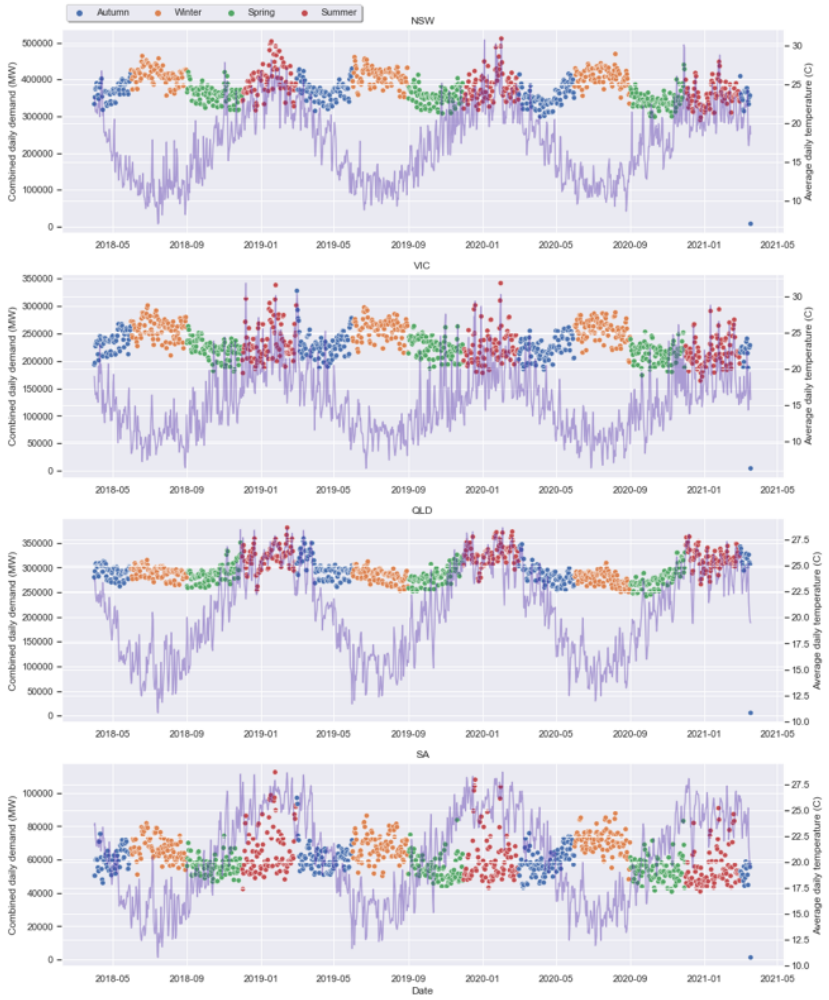


Fig. 12. Combined daily state-wide demand (MW; scatterplot) and average daily temperatures (ºC, purple lineplot) for NSW, VIC, QLD and SA (March 2018-2021).

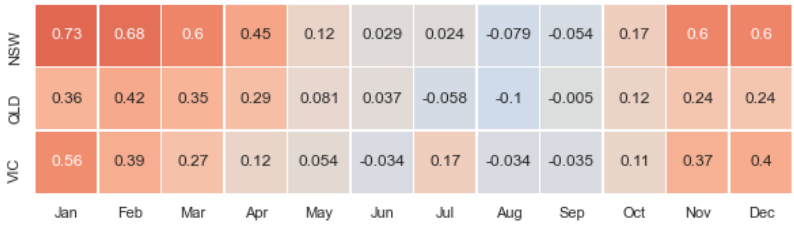


Fig. 13. Heatmap of Pearson correlation coefficients TOTALDEMAND and TEMPERATURE (stratified by state and month).

3.4 Rooftop PV power EDA

Rooftop PV power production for all states shows an overall increasing trend in maximum daily output, reflecting the growing prevalence of PV installations in Australia. A clear seasonality is apparent, with maximum PV energy production in summer and minimal in winter (**Fig. 14**).

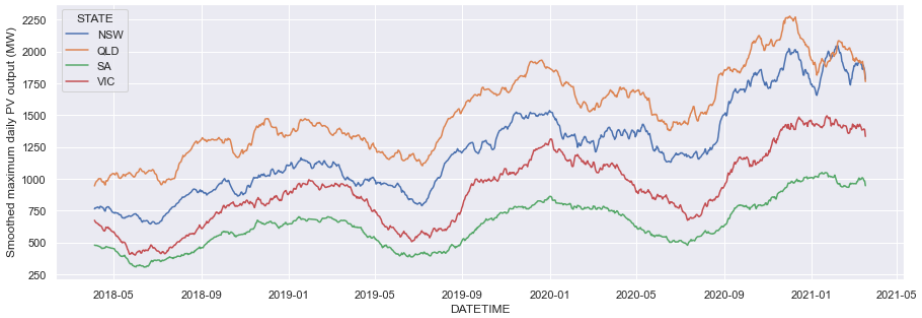


Fig. 14. Maximum daily state-wide photovoltaic output (MW) for NSW, QLD, SA and VIC (March 2018-2021). Daily maximums smoothed by rolling 30-day average function

The average daily PV power production profile follows a highly symmetrical bell-shaped curve, with maximum output occurring between 11:30AM and 12PM in NSW and QLD, with the peak offset for VIC and SA due to longitudinal and timezone differences respectively (**Fig. 15**).

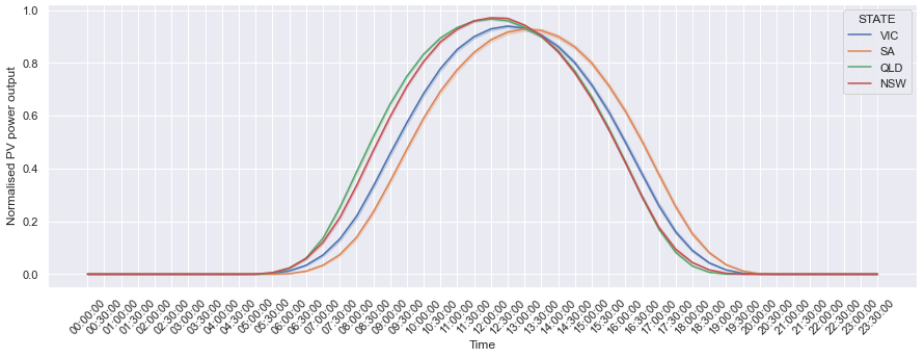


Fig. 15. Daily state-wide average PV output profile for NSW, QLD, SA and VIC (March 2018-2021). Outputs at each timepoint for each day were normalized against the daily maximum to account for seasonal and state differences. Shading indicates 95% CI interval of mean.

In contrast the daily grid demand profile follows a distinctive bimodal distribution with separate morning (~6:30-9:00AM) and evening (~5:30-7:30PM) peaks, creating a demand "valley" throughout the afternoon (**Fig. 16**).

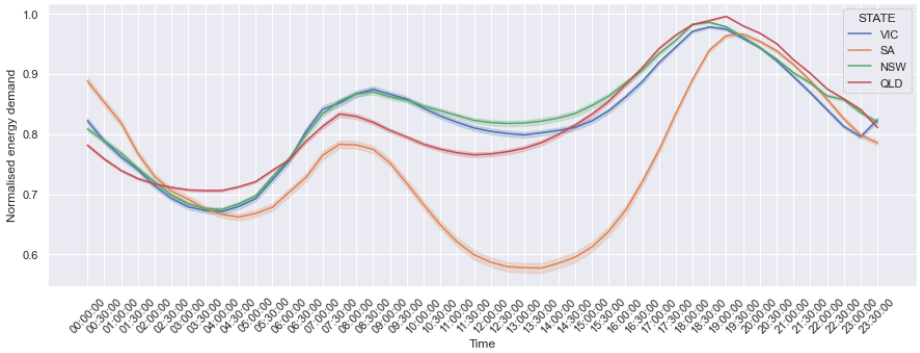


Fig. 16. Daily state-wide average energy demand profile for NSW, QLD, SA and VIC (March 2018-2021). Demands at each timepoint at each day were normalised against individual daily maximums to account for seasonal variation. Shading indicates 95% confidence intervals of the mean.

This "valley" has a close inverse correlation with PV output power, suggesting that rooftop solar generation may account for this phenomenon. Indeed, when grid demand and PV output are combined (representing an approximation of total actual power consumption for the state ie: the 'ActualDemand' term in Section 2.8), a plateau in net energy demand/ActualDemand (ie: TOTALDEMAND+POWER) is observed from the early morning to the early evening (**Fig. 17**), a pattern which is consistent across all states in the dataset (**Fig. 18**). This observation highlights the profound influence that rooftop PV production has on grid demand across all four Australian states measured and suggests that incorporation of PV data into forecasting models may have a beneficial effect in reducing the unpredictability of forecasting daytime grid demand.

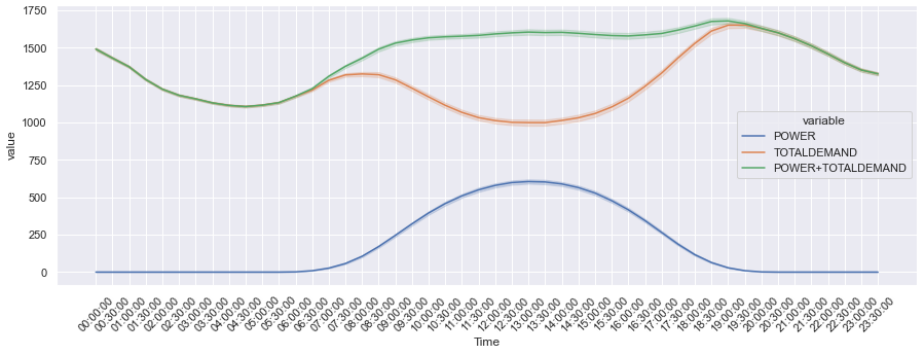


Fig. 17. Daily average energy demand profile (TOTALDEMAND), rooftop PV output (POWER), and these profiles summed (POWER+TOTALDEMAND) (MW) for South Australia (March 2018-2021). Shading indicates 95% confidence intervals of the mean.

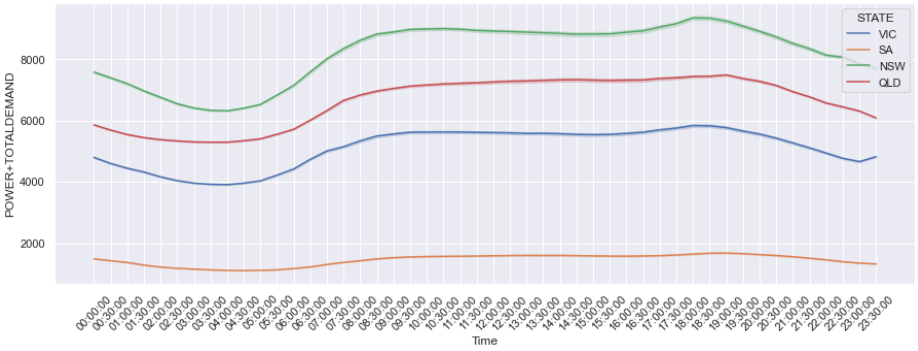


Fig. 18. Daily combined average of grid demand and rooftop PV output (ie: POWER+TOTALDEMAND) for NSW, QLD, SA and VIC (March 2018-2021). Shading indicates 95% CI boundaries of the mean.

4 Analysis of Results

4.1 Reporting of results

Each model type was evaluated as described in the Materials and Methods, using a roll-forward testing approach on the hold-out test set between 1st-31st January 2021 for the NSW TOTALDEMAND data. MAPE and RMSE were calculated for each model (**Table *4***). Graphical summaries of the test-set forecasts (**Fig. 20**), distribution of prediction residuals (**Fig. 21**) and prediction residuals versus time of day for all models (**Fig. 22**) are presented in composite. Analysis of results specific to each model type are discussed in sections 4.2-4.5 below.

4.2 Naïve drift baseline models

Predictions of energy demand for all 30 minute intervals were made on the hold-out test set from 1st to 31st January 2021 using the naïve drift baseline model (which simply repeats the energy demand from 24 hours prior to the forecast timepoint). The baseline MAPE and RMSE were established as 6.744 and 787.658, respectively (**Table *4***).

4.3 Linear models

Prediction performance

Optimal time-decay and polynomial transformation of predictor variables for the log-linear (LM) and generalized linear (GLM) models as described in sections 2.8 and 2.10, respectively. Roll-forward test accuracy on the hold-out NSW 1st-31st January 2021 for the no-PV model (ie: predicting 'GeneratedDemand') was superior to the naïve drift baseline. Notably, inclusion of PV data in the LM (predicting 'ActualDemand', then subtracting 'POWER') further improved test accuracy (MAPE: 3.012) (**Table *4***).

Roll-forward test accuracy of both GLM models also exceeded the performance of the naïve baseline. Consistent with results for the LMs, inclusion of PV data improved GLM accuracy versus not (MAPE 4.636 versus 5.039). Outright test accuracy for both GLM models was worse than the LMs (**Table *4***).

Table 4. Roll-forward test accuracy on the 1st-31st January 2021 hold-out set for all models described in this report. Accuracy reported as both MAPE and RMSE.

	No PV		PV	
Model	**MAPE**	**RMSE**	**MAPE**	**RMSE**
AEMO	2.183	221.199	-	-
Naïve drift	6.744	787.658	-	-
LM	3.711	439.601	3.012	377.457
GLM	5.039	538.299	4.636	499.879
N-BEATS	12.960	1288.857	8.740	927.413
TFT	14.213	1397.900	13.480	1417.669

Visualisation of predictions for both LM and GLM-based models over the entire test period showed a generally good fit to actual energy demand (**Fig. 20**). There was a major underprediction error made by all models in the evening of January 25th, which does not occur anywhere else in the test set and is responsible for the largest residual outliers. Future development should focus on identifying the cause of this anomalous prediction.

Prediction residuals for LM and GLM models were normally distributed and centered on a mean of zero (**Fig. 21**), indicating low prediction bias and fulfilling a key assumption of linear regression. Similarly, there were no specific biases detected in a plot of prediction residuals versus prediction time-of-day (**Fig. 22**).

Analysis of regression coefficients

A key benefit of linear regression models is their generation of interpretable coefficients. These can be used to estimate historical and future trends, by varying inputs into predictor variables of interest while holding all others constant. As an example, the effect on energy demand of different seasons, days-of-week or occurrence of public holidays was modelled by varying the corresponding categorical variables (**Table 5**). These effects are expressed as percentage change in energy demand relative to an arbitrary reference day: a non-public holiday Sunday during summer.

Table 5. Effect on energy demand predicted by the linear model by varying season, public holiday or day-of-week. Effect expressed as percentage change versus a reference day of a summer Sunday.

Categorical variable	Value	No PV (%)	PV (%)
Seasons	Autumn	-4.57	-4.57

	Winter	4.63	4.37
	Spring	-6.03	-4.72
Public Holiday		-9.60	-9.01
Day	Monday	9.33	8.66
	Tuesday	10.51	9.96
	Wednesday	10.39	9.67
	Thursday	10.36	9.64
	Friday	10.16	9.38
	Saturday	2.79	2.51

The linear model faithfully recapitulates seasonal, public holiday and day-of-week demand trends observed in EDA, which are independent of PV generation. Similar modelling can also be performed by varying the time-series and continuous variables. The linear model contains a third-order relationship term between energy demand and time. By varying the time index (ie: date), a falling trend from summer 2018-2019 onwards can be visualised for both actual demand (inclusive of PV), and residual grid demand (generated supply) (**Fig. 19A**). The increasing gap between these two curves reflects the growing share of PV in the net energy mix and highlights the value of this linear model for prediction of critical long-term metrics, such as forecasting minimum NEM grid demand thresholds [6]. Varying temperature input allows reproduction of the non-linear relationship between itself and energy (**Fig. 19B**), where energy demand spikes upwards at temperatures above 30ºC, and to a lesser extent on colder days below 15 ºC.

Demand variation for a typical summer week was also modelled by setting the season variable and varying the time index through a week. Known morning and afternoon demand peaks are apparent, as is the diminished mid-day dip in actual demand, due to rooftop PV supplementation of household energy requirements during peak solar irradiance (**Fig. 19C**).

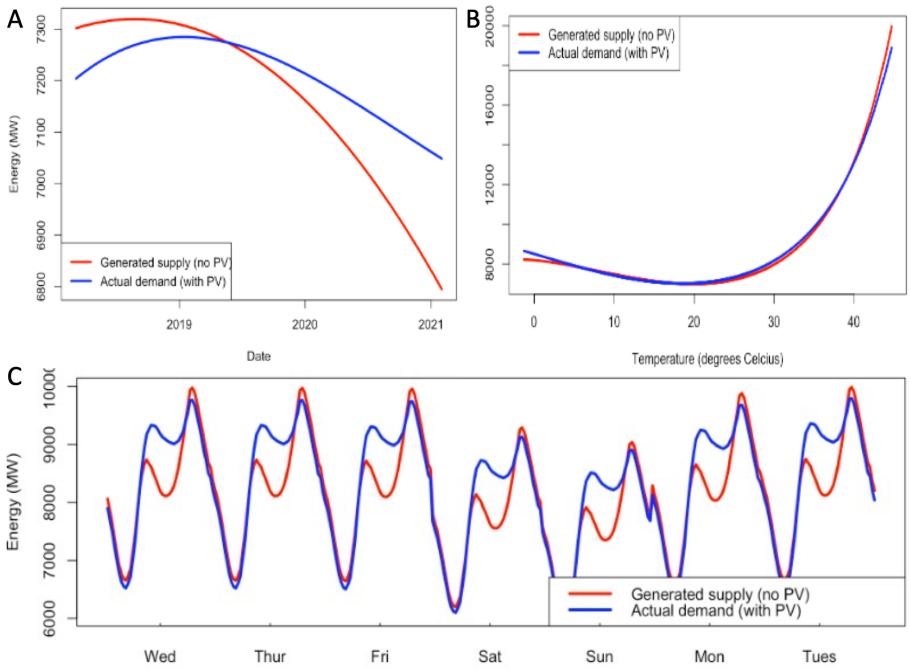


Fig. 19. Forecasts of GeneratedSupply (no PV) and ActualDemand (with PV) from the linear model generated by varying predictors **A:** Date over the year range 2018-2021, **B**: Temperature over the range 0-40ºC and **C:** Day over the range Wed-Tues for a week in NSW summer.

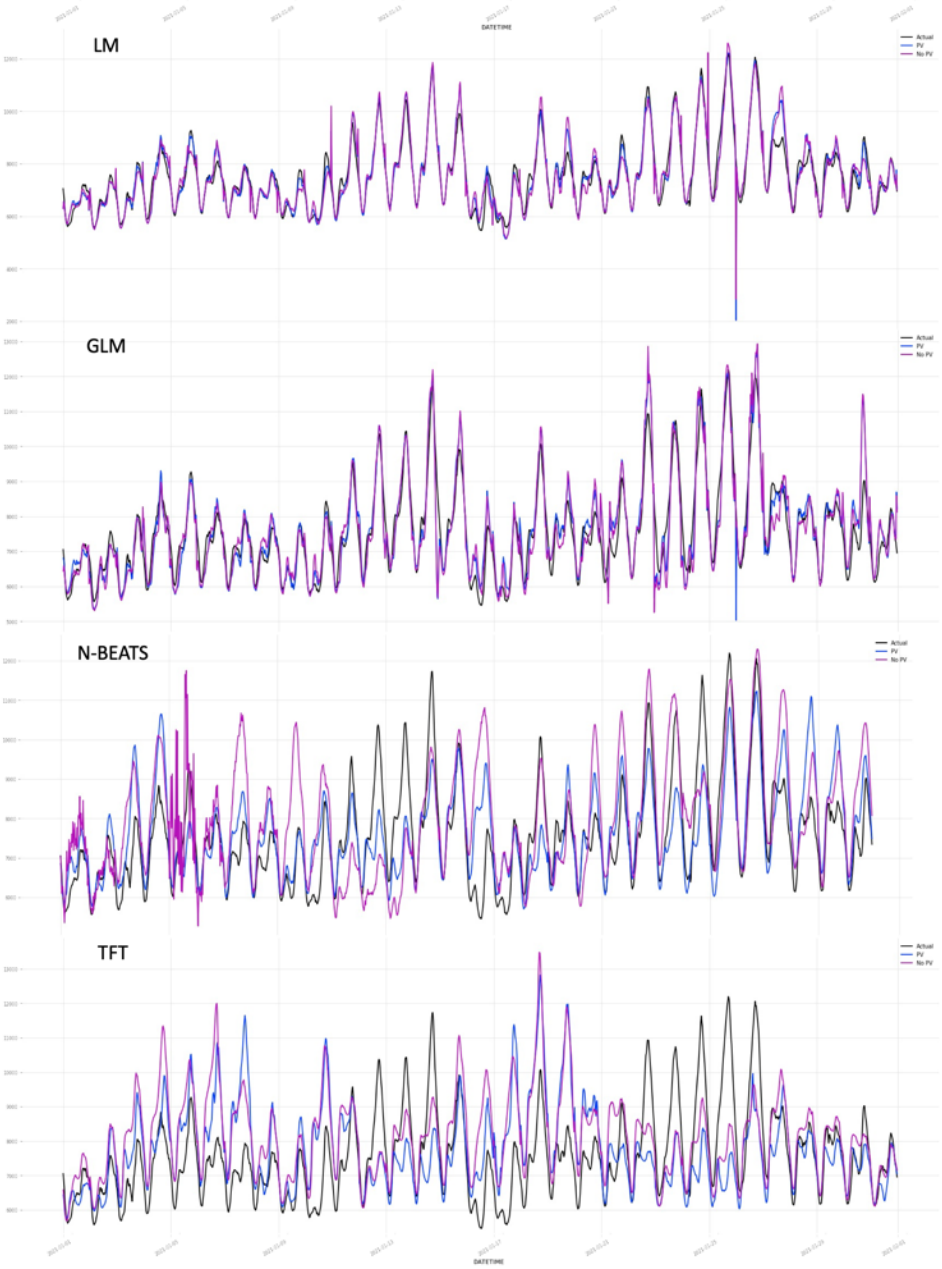


Fig. 20. Forecast total energy demand ('TOTALDEMAND') (MW) for the 1st-31st January 2021 hold-out test set, for models trained without (purple) or with (blue) PV data, versus actual values (black).

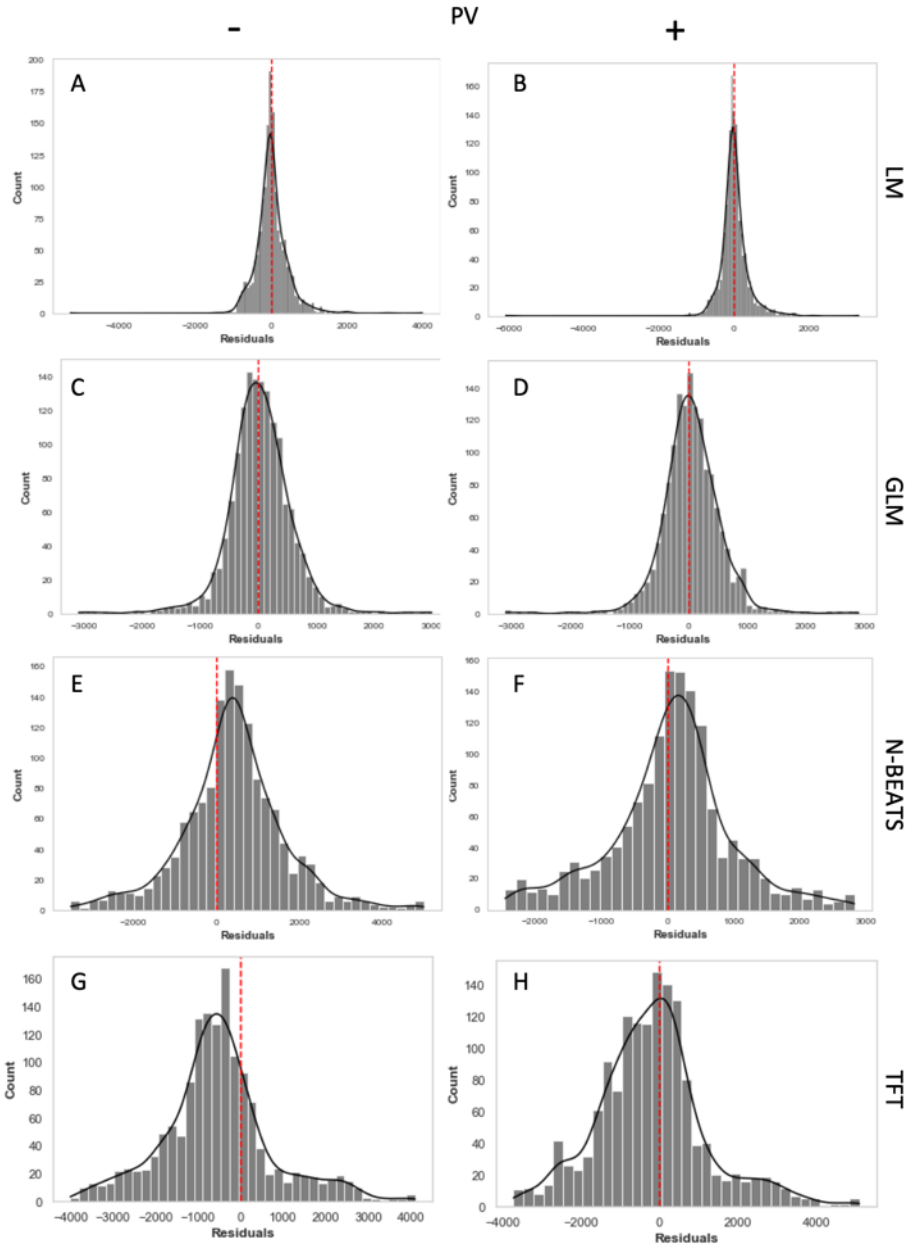


Fig. 21. Histograms of prediction residuals (predicted – actual) for models trained without (-) or with (+) PV data. Zero residual error indicated by dashed red line.

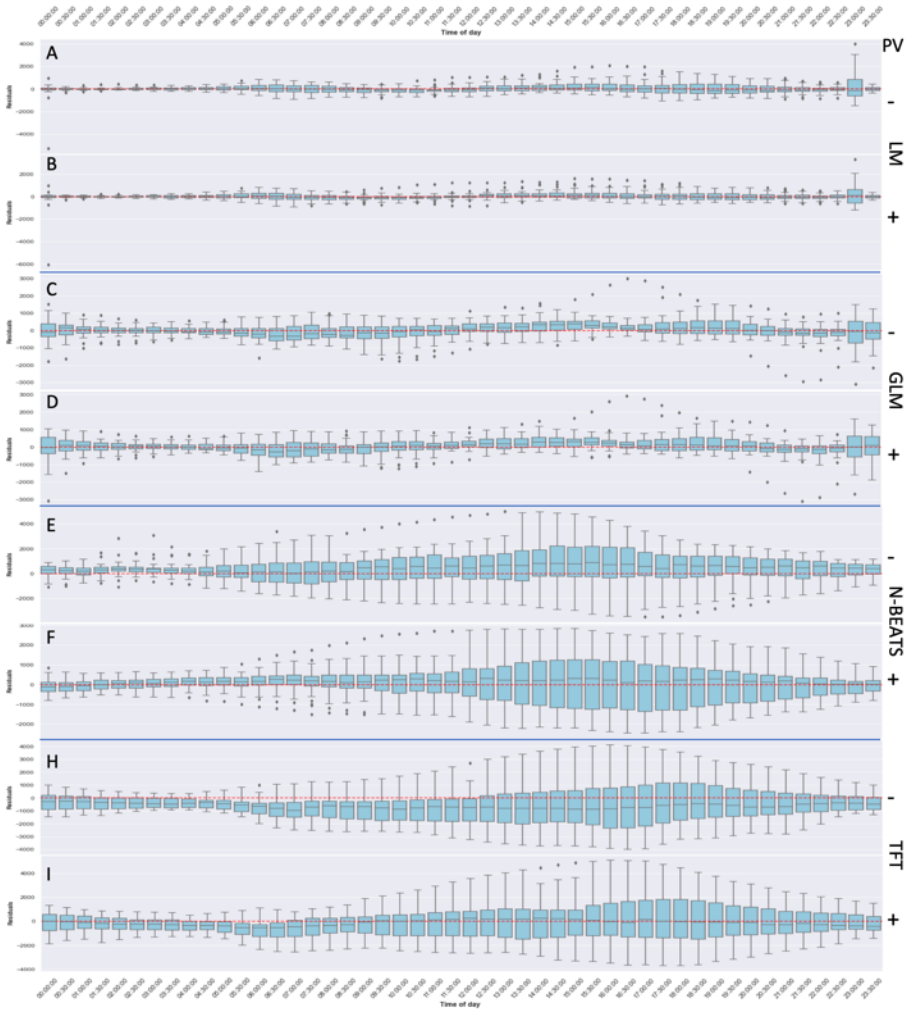


Fig. 22. Boxplots of prediction residuals (prediction – actual) versus time of day for models trained without (-) or with (+) PV data. Whiskers indicate 1.5 x IQR, dots are outliers beyond this.

4.4 Neural network models with and without PV data

N-BEATS

An optimal lookback period of 2 months was identified, using backforecast RMSE over the entirety of 2020 data for NSW as the evaluation metric. This longer lookback window presumably allowed the model to learn greater seasonal variation, albeit at the cost of significantly increased training time compared with smaller lookback windows.

Backforecast training-set accuracy by the non-PV supplemented model over the last 12 months of the NSW training set was worse than the naïve drift model (**Table 6**), indicating a relatively poor fit to the training data by the N-BEATS model. Inclusion of the PV 'POWER' covariate in N-BEATS training markedly improved backforecast training-set accuracy, exceeding the naïve model. However, roll-forward testing on the 1st-31st January 2021 NSW hold-out test set produced poor accuracies for both PV and non-PV models, which were both eclipsed even by the naïve drift model. Despite this disappointing test performance, inclusion of the PV covariate still improved test set accuracy relative to the non-PV model.

Table 6. Model backforecast accuracy over the January 1st - December 31st 2020 portion of the training set (NSW data only).

Model	MAPE	RMSE
AEMO	2.045	230.126
Naïve drift	5.890	694.460
N-BEATS non-PV	6.946	766.310
N-BEATS PV	4.265	452.281

While test-set residuals were normally distributed (**Fig. 21 E, F**), there was a tendency for N-BEATS models to over-predict demand, especially the non-PV model. The best test-set performance for both models was obtained in the early hours of the morning. Residuals were the greatest after 12:00PM for both models, with this effect most prominent for the PV-aware N-BEATS model (**Fig. 22 E, F**).

Individual analysis of predictions on 'outlier' days in the January 2021 test set (**Fig. 20**) revealed reasonable performance over the days January 25th-26th, a period containing the highest demand, intra-day demand change, temperature, intra-day temperature change in the extended NSW test set, as well as a public holiday.

Poor test performance while achieving good backforecasting accuracy on the training set suggest both N-BEATS models are overfitting the training set. This could be addressed by reducing the number of epochs, the size of the lookback window or the number of block layers. Despite these issues, these results demonstrate that supplementation of an N-BEATS model with the additional covariate of PV power production improves prediction accuracy of total energy demand on the NSW dataset.

TFT

During training of the TFT model, validation error was reduced by inclusion of PV data (**Table 7**). Addition of a PV covariate in rollforward testing on the holdout 1st-31st January 2021 NSW test-set also yielded improvements in prediction accuracy, versus a TFT model trained without this data (**Table *4***). However, in absolute terms both TFT networks were outperformed by all other models, including the naïve drift baseline model. This underperformance reflects the minimal fine-tuning of hyperparameters of input time-series length which was performed due to time constraints.

Table 7. TFT backforecast accuracy over the January 1st-December 31st 2020 portion of the training set (NSW data only)

Model	MAPE	RMSE
TFT non-PV	3.593	37.111
TFT PV	3.227	33.365

Residual analysis (**Fig. 21**. **G, H**) revealed an underprediction bias for the non-PV TFT model, causing potential costly demand shortfalls for the energy operator. In contrast, PV-aware prediction residuals were closer to normally distributed around zero, suggesting that inclusion of PV data in the TFT model reduces overall bias in future predictions.

Individual analysis of predictions on 'outlier' days in the January 2021 test set, revealed better performance after the January 26th public holiday (**Fig. 20**). This may be explained as it is the end of the holiday period and fewer variations on demand will be expected. Comparing validation dataset and test dataset accuracy, both TFT model have likely overfit the training dataset. Future improvement of the TFT model would be invest further time hyperparameter tuning, and establishing an optimal number of training epochs to reduce overfitting.

5 Discussion

Integration of aggregated state-wide rooftop-PV power data into energy demand forecast models for NSW consistently improved their test accuracy. This improvement was observed for both statistical (linear and generalized linear models) and deep learning approaches (N-BEATS and TFT). This finding is consistent with previous studies of net energy demand forecasting [15], [16] and demonstrates the importance of including behind-the-meter and solar information in the Australian context with its record-high levels of PV penetration.

The use of PV data aggregated from entire states has insulated this study from the high variability associated with neighbourhood- or household-scale solar studies [42], with the obvious trade-off of only enabling whole-state forecasts of demand to be made. It would be of interest to re-train our models (especially the more accurate linear regression models) using PV data obtained from individual households, and determine if this more tailored, yet variable view of PV generation will produce high accuracy forecasts of localized demand.

Linear regression models produced the most accurate forecasts of energy demand over the NSW 1st-31st January 2021 test set. Residuals were normally distributed and relatively free from bias across all forecast periods of the day, satisfying a key assumption of linear regression models. Comparatively, deep learning approaches (N-BEATS and TFT) performed poorly, with average prediction below that of the baseline naïve drift forecast, even with inclusion of PV data. That both N-BEATS and TFT demonstrated good training set backforecast performance suggests that overfitting was a significant problem.

The short timeframe of this project restricted opportunities for thorough N-BEATS and TFT hyperparameter tuning, a process critical for development of high performing deep learning models. Further development could address apparent training set overfitting by incorporating regularization approaches, such as ensembling. Ensembling been previously used with N-BEATS models to improve their accuracy, by combining predictions from several separate models trained across a range of lookback time windows [26].

Both PV and non-PV N-BEATS models in this report displayed a prediction bias, tending to over-predict energy demand on average in the NSW January 2021 test set (particularly for afternoon timepoints) (**Fig. 22E, F**). This bias might be addressed by inclusion of a loss function which asymmetrically penalises errors, such as pinball-symmetric-MAPE (pinball-sMAPE). Pinball-sMAPE has been successfully applied in N-BEATS models of energy demand and production forecasting [25], [26] to reduce undesirable prediction biases. This loss function is not yet implemented in the PyTorch library, and time constraints precluded a custom function being developed for inclusion in our N-BEATS models.

The relatively strong performance of the regression models highlights the practical utility of classical linear regression methods for forecasting, especially in time constrained projects. Ordinary-least-squares linear regression requires a fraction of the computational time required for training deep learning models, enabling a faster tuning-testing iteration cycle, and allowing tuning of a linear model which approaches the accuracy of the AEMO forecasts on the January 2021 NSW test set when supplemented with PV data.

The linear regression models described in this report provide not only accurate forecasts of energy demand, but also access to coefficients directly interpretable by domain experts. An example is given in **Table 5**, where relative changes in energy demand due to changing predictor variables such as season, public holiday or day-of-week can be modelled. The ability to generate 'what if' forecasts of various future energy demand scenarios is highly desirable for AEMO and NEM stakeholders [6]. An example of a future application of the linear model described here would be generation and visualization of specific scenarios of interest to AEMO, such as the impact of a one-degree increase in average daily temperature on winter energy demand.

Related to this is the 'inference mode' feature of the N-BEATS model, which fixes the waveform generator functions to those suitable for determining seasonality and trend for the training data [20] . A future extension of this work would be to take advantage of this N-BEATS feature to produce additional interpretable parameters to supplement those obtained in the regression analysis (after suitable tuning to improve accuracy).

In conclusion, this report shows that inclusion of aggregated rooftop-PV data, either as a covariate in a deep learning model, or incorporated into a net-energy demand target variable for linear regression, improves the accuracy of energy demand forecasts, strongly suggesting its further use in future related projects.

6 References

[1] "Australian Energy Market Operator." https://aemo.com.au (accessed Apr. 12, 2022).

[2] "National Electricity Market." https://www.energy.gov.au/government-priorities/energy-markets/national-electricity-market-nem (accessed Apr. 12, 2022).

[3] H. Zareipour, C. A. Cañizares, and K. Bhattacharya, "Economic impact of electricity market price forecasting errors: A demand-wide analysis," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 254–262, Feb. 2010, doi: 10.1109/TPWRS.2009.2030380.

[4] International Energy Agency, "Trends in Photovoltaic Applications 2020," 2021. Accessed: Apr. 09, 2022. [Online]. Available: https://iea-pvps.org/wp-content/uploads/2022/01/IEA-PVPS-Trends-report-2021-3.pdf

[5] W. S. Chandler, C. H. Whitlock, and P. W. Stackhouse, "NASA climatological data for renewable energy assessment," *Journal of Solar Energy Engineering, Transactions of the ASME*, vol. 126, no. 3, 2004, doi: 10.1115/1.1748466.

[6] Australian Energy Market Operator, "Electricity Statement of Opportunities A report for the National Electricity Market Important notice," 2021. https://aemo.com.au/-/media/files/electricity/nem/planning_and_forecasting/nem_esoo/2021/2021-nem-esoo.pdf?la=en

[7] P. Paudyal *et al.*, "The Impact of Behind-the-Meter Heterogeneous Distributed Energy Resources on Distribution Grids," in *Conference Record of the IEEE Photovoltaic Specialists Conference*, 2020, vol. 2020-June. doi: 10.1109/PVSC45281.2020.9300626.

[8] A. Woyte, V. van Thong, R. Belmans, and J. Nijs, "Voltage fluctuations on distribution level introduced by photovoltaic systems," *IEEE Transactions on Energy Conversion*, vol. 21, no. 1, 2006, doi: 10.1109/TEC.2005.845454.

[9] R. Hudson and G. Heilscher, "PV grid integration - System management issues and utility concerns," in *Energy Procedia*, 2012, vol. 25. doi: 10.1016/j.egypro.2012.07.012.

[10] A. Alshahrani, S. Omer, Y. Su, E. Mohamed, and S. Alotaibi, "The technical challenges facing the integration of small-scale and large-scale PV systems into the grid: A critical review," *Electronics (Switzerland)*, vol. 8, no. 12. 2019. doi: 10.3390/electronics8121443.

[11] S. Rahman *et al.*, "Analysis of Power Grid Voltage Stability with High Penetration of Solar PV Systems," *IEEE Transactions on Industry Applications*, vol. 57, no. 3, pp. 2245–2257, May 2021, doi: 10.1109/TIA.2021.3066326.

[12] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond," *International Journal of Forecasting*, vol. 32, no. 3. 2016. doi: 10.1016/j.ijforecast.2016.02.001.

[13] S. E. Razavi, A. Arefi, G. Ledwich, G. Nourbakhsh, D. B. Smith, and M. Minakshi, "From Load to Net Energy Forecasting: Short-Term Residential

Forecasting for the Blend of Load and PV behind the Meter," *IEEE Access*, vol. 8, pp. 224343–224353, 2020, doi: 10.1109/ACCESS.2020.3044307.

[14] J. Leiva, A. Palacios, and J. A. Aguado, "Smart metering trends, implications and necessities: A policy review," *Renewable and Sustainable Energy Reviews*, vol. 55. 2016. doi: 10.1016/j.rser.2015.11.002.

[15] O. Ajayi and R. Heymann, "Data centre day-ahead energy demand prediction and energy dispatch with solar PV integration," *Energy Reports*, vol. 7, pp. 3760–3774, Nov. 2021, doi: 10.1016/j.egyr.2021.06.062.

[16] C. Rollins, "Forecasting Energy Demand & Peak Load Days with the Inclusion of Solar Energy Production," Rochester Institute of Technology, Rochester, 2020.

[17] P. Kobylinski, M. Wierzbowski, and K. Piotrowski, "High-resolution net load forecasting for micro-neighbourhoods with high penetration of renewable energy sources," *International Journal of Electrical Power and Energy Systems*, vol. 117, May 2020, doi: 10.1016/j.ijepes.2019.105635.

[18] M. Sun, C. Feng, and J. Zhang, "Factoring behind-the-meter solar into load forecasting: Case studies under extreme weather," 2020. doi: 10.1109/ISGT45199.2020.9087791.

[19] Z. A. Khan and Di. Jayaweera, "Smart Meter Data Based Load Forecasting and Demand Side Management in Distribution Networks with Embedded PV Systems," *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2019.2962150.

[20] B. N. Oreshkin, D. Carpov, N. Chapados, and Y. Bengio, "N-BEATS: Neural basis expansion analysis for interpretable time series forecasting," *CoRR*, vol. abs/1905.10437, May 2019, [Online]. Available: http://arxiv.org/abs/1905.10437

[21] B. Lim, S. Arık, N. Loeff, and T. Pfister, "Temporal Fusion Transformers for interpretable multi-horizon time series forecasting," *International Journal of Forecasting*, vol. 37, no. 4, 2021, doi: 10.1016/j.ijforecast.2021.03.012.

[22] S. Arik, "Interpretable Deep Learning for Time Series Forecasting," Dec. 13, 2021. https://ai.googleblog.com/2021/12/interpretable-deep-learning-for-time.html (accessed Apr. 08, 2022).

[23] J. F. Torres, F. Martínez-Álvarez, and A. Troncoso, "A deep LSTM network for the Spanish electricity consumption forecasting," *Neural Computing and Applications*, 2022, doi: 10.1007/s00521-021-06773-2.

[24] D. Mariano-Hernández *et al.*, "A data-driven forecasting strategy to predict continuous hourly energy demand in smart buildings," *Applied Sciences (Switzerland)*, vol. 11, no. 17, Sep. 2021, doi: 10.3390/app11177886.

[25] D. Putz, M. Gumhalter, and H. Auer, "A novel approach to multi-horizon wind power forecasting based on deep neural architecture," *Renewable Energy*, vol. 178, pp. 494–505, 2021, [Online]. Available: https://www.elementai.com/

[26] B. N. Oreshkin, G. Dudek, P. Pełka, and E. Turkina, "N-BEATS neural network for mid-term electricity load forecasting," Sep. 2020, [Online]. Available: http://arxiv.org/abs/2009.11961

[27] J. Herzen *et al.*, "Darts: User-Friendly Modern Machine Learning for Time Series," Oct. 2021, [Online]. Available: http://arxiv.org/abs/2110.03224

[28] Australian Energy Market Operator, “Demand Terms in EMMS Data Model,” 2021. Accessed: Mar. 13, 2022. [Online]. Available: https://www.aemo.com.au/-/media/files/electricity/nem/security_and_reliability/dispatch/policy_and_process/demand-terms-in-emms-data-model.pdf?la=en

[29] Australian Energy Market Operator, “Australian Solar Energy Forecasting System,” 2021. https://aemo.com.au/en/energy-systems/electricity/national-electricity-market-nem/nem-forecasting-and-planning/operational-forecasting/solar-and-wind-energy-forecasting/australian-solar-energy-forecasting-system (accessed Mar. 14, 2022).

[30] Australian Government: Clean Energy Regulator, “Clean Energy Regulator Home Page.” http://www.cleanenergyregulator.gov.au (accessed Apr. 14, 2022).

[31] Australian Energy Market Operator, “Market Data NEMWEB.” https://www.aemo.com.au/energy-systems/electricity/national-electricity-market-nem/data-nem/market-data-nemweb#rooftop-pv-actual (accessed Apr. 12, 2022).

[32] “MeteoStat Home Page.” https://meteostat.net/en/ (accessed Apr. 16, 2022).

[33] Bureau of Meteorology, “Climate Glossary.” http://www.bom.gov.au/climate/glossary/seasons.shtml (accessed Apr. 02, 2022).

[34] Geoscience Australia, “Geodetic Calculators.” https://geodesyapps.ga.gov.au/sunrise (accessed Apr. 02, 2022).

[35] “Python ‘holidays’ 0.13 documentation.” https://pypi.org/project/holidays/ (accessed Apr. 14, 2022).

[36] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed. oTexts, 2021.

[37] “Darts N-BEATS API documentation.” https://unit8co.github.io/darts/generated_api/darts.models.forecasting.nbeats.html (accessed Apr. 12, 2022).

[38] P. Laurinec, “Forecast double seasonal time series with multiple linear regression in R,” *https://petolau.github.io/Forecast-double-seasonal-time-series-with-multiple-linear-regression-in-R/*, Dec. 03, 2016.

[39] B. Beranger, “Regression Analysis for Data Scientists,” *https://moodle.telt.unsw.edu.au/course/view.php?id=65691&section=8*, Jan. 01, 2022.

[40] A. Dobson, *An Introduction to Generalized Linear Models, Fourth Edition*, 2nd ed. Chapman and Hall/CRC, 2018. doi: 10.1201/9781315182780.

[41] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed. oTexts, 2021.

[42] D. W. van der Meer, J. Munkhammar, and J. Widén, “Probabilistic forecasting of solar power, electricity consumption and net load: Investigating the effect of seasons, aggregation and penetration on prediction intervals,” *Solar Energy*, vol. 171, pp. 397–413, Sep. 2018, doi: 10.1016/j.solener.2018.06.103.

7 Appendix

Table 1. List of supporting libraries used in Python and R software

Library	Description
Python	
Pandas	Dataframe manipulation and analysis
Seaborn	Graphing and plotting
Numpy	Core numerical array handling
darts	General time series data handling and N-BEATS modelling
PyTorch	TFT neural network model
Meteostat	Interface to the MeteoStat meteorological database
R	
ggplot2	Graphing and plotting
lubridate	Date and time handling
plotly	Graphing and plotting
datawizard	Data wrangling and cleaning

Table 2. GitHub repository details for software supporting this study

Description	Github URL
MeteoStat lookup of missing temperature values	https://github.com/the-rahul-kumar/UNSW-Group-C-Project/blob/main/src/Mat/temp_cleaner_multistate.ipynb
Processing of AEMO PV data	https://github.com/the-rahul-kumar/UNSW-Group-C-Project/blob/main/src/Mat/rooftop_pv.ipynb
Merge of demand and PV datasets	https://github.com/the-rahul-kumar/UNSW-Group-C-Project/blob/main/src/Mat/demand_pv_combiner.ipynb
Analysis of model predictions and errors	https://github.com/the-rahul-kumar/UNSW-Group-C-Project/blob/main/src/Mat/model_error_analysis.ipynb
EDA and AEMO forecast error analysis	https://github.com/the-rahul-kumar/UNSW-Group-C-Project/blob/main/src/Mat/PV_AEMO_erroranalysis_EDA.ipynb
TFT model (no PV data)	https://github.com/the-rahul-kumar/UNSW-Group-C-Project/blob/main/src/Rapha/final/ TFT - no PV.ipynb
TFT model (with PV data)	https://github.com/the-rahul-kumar/UNSW-Group-C-Project/blob/main/src/Rapha/final/ TFT - PV.ipynb
EDA	https://github.com/the-rahul-kumar/UNSW-Group-C-Project/blob/main/Rahul-Workings/EDA.ipynb
Linear models	https://github.com/the-rahul-kumar/UNSW-Group-C-Project/blob/main/src/Paul/LinearModels/LinearModelsWithAndWithoutPV.Rmd
Data set combiner	https://github.com/the-rahul-kumar/UNSW-Group-C-Project/blob/main/Rahul-Workings/Data-Set-Combiner.ipynb

Table 3. Hyperparameters of N-BEATS and TFT models

Model	Python Library	Architecture and Hyper-Parameters
N-BEATS	Darts	• num_stacks = 30, num_blocks = 1, num_layers = 4, layer_widths = 256 • loss_fn = Adam, learning_rate = 0.001, loss metric = RMSE • expansion_coefficient_dim = 5 • past_covariates (non-PV models)= [‘Weekday’,’Month’,’HOLIDAY’]

		• past_covariate (PV models) = [‘Week-day’,’Month’,’HOLIDAY’, ‘POWER’]
TFT	PyTorch	• hidden_size = 93, lstm_layers = 2, attention_head_size = 2 • drop rate = 0.12, batch_size = 128, hidden_continuous_size=33 • log interval = 10, learning_rate = 0.0378, • gradient_clip_val = 0.977, reduce_on_plateau_patience = 4 • time_varying_known_categoricals = ["Weekday", "Season", "DAYTYPE", "HOLIDAY"] • time_varying_known_reals = ["Quarter", "time_idx", "Month", "Day", "TEMPERATURE","POWER"], • The loss function used was Quantile Loss with seven quantiles. The model with the best epoch in the loss function was selected.