Attachment 1.2

Energy & Customer Numbers Forecast Methodology

Revised proposed access arrangement information

15 November 2022

Access Arrangement (AA) for the period
1 July 2023 to 30 June 2027

EDM 61987240

Energy & Customer Numbers Forecast Methodology

Connections, energy, demand and reactive power

3 November 2022

Western Power

363 Wellington Street
Perth WA 6000
GPO Box L921 Perth WA 6842

Document Information

Title	Energy & Customer Numbers Forecast Methodology
Subtitle	Connections, energy, demand and reactive power

Authorisation

	Title	Name	Date
Authors	Senior Data Science Analyst Data Science Team Leader		8 October 2020
Reviewer	Principal Analyst		15 February 2021
Approver:	Customer Insights and Predictive Analytics Manager Acting for Head of Business Intelligence and Data Analytics		3 November 2022

Document History

Rev No	Date	EDM Version	Amended by	Details of amendment

Review Details

Review Period:	2021/22
NEXT Review Due:	2023/24

Contents

List of tables

List of figures

Abbreviations

Term	Definition
Annual average demand	Average electricity demand over the course of one year, expressed in megawatt (MW). In this report, annual average demand refers to the annual average measured based on average demand during five-minute intervals.
Annual peak (or maximum) demand	The highest five-minute average electricity demand of any five-minute interval during one year, usually expressed in MW.
Block load	Block loads are typically large industrial customers whose electricity demand is large. When connected to the Western Power network, they create a large change in electricity consumption and demand for the relevant part of the network. The forecast method accounts for the influence of these large users using an adjustment to the forecast produced for small and medium customers. Once added to demand growth for small and medium size customers, a block load introduces an often-permanent step-change into an otherwise smooth trend.
Coincident and non-coincident demand	See section 0 on page 33 for the definitions.
Day-time	Day-time is defined as the day hours from 8am to 6pm.
kWh	Kilowatt-hour is a basic measuring unit of electric energy equal to one kilowatt of power supplied to or taken from an electric circuit steadily for one hour. One kilowatt-hour can power ten 100 watt light bulbs for one hour.
Load factor	The ratio of annual average demand to annual peak demand. This is a partial indicator of network utilisation.
MVA	MVA stands for Mega-Volt-Amps and is a measure of apparent power. If the total load requirement is 1,000 volts and 5,000 amps (1,000 x 5,000 = 5,000,000 VA) it can be expressed as 5 MVA. Apparent power takes into consideration both the resistive and reactive load.
MW	Megawatt is a measure of the active component of electrical demand and represents the capacity to deliver energy.
MWh	Megawatt-hour is a measure of energy delivered. For example, one MWh of electricity can power ten thousand 100 watt light bulbs for one hour.
NMI	NMI or the National Metering Identifier is a unique identifier that identifies a supply or connection point and is assigned by the providing distributor.
Night-time	Night-time is defined as the hours from 6pm to 8am.

Term	Definition
PoE	Probability of exceedance (PoE) is a statistical measure that describes the probability that a particular value will be met or exceeded – e.g., a PoE 10 forecast is expected to be exceeded 10 per cent of the time i.e. one year in 10. And a PoE 20 forecast is expected to be exceeded 20 per cent of the time i.e. one year in five.
Serial correlation	Serial correlation is the relationship between a given variable and itself over various time intervals. Serial correlations are often found in repeating patterns, when the level of a variable effects its future level.
Summer	Summer refers to the period from 1 December to 31 March (inclusive).
SWIN	South West Interconnected Network is all the transmission and distribution components of the electricity system comprising the Western Power Network and other transmission and distribution assets owned and operated by other parties.
SWIS	South West Interconnected System is the entire electricity system covering the south west of Western Australia. It comprises the Western Power Network, other transmission and distribution assets owned and operated by other parties than Western Power and all generators.
Western Power Network	Is the transmission and distribution element of the SWIN that is owned and operated by Western Power.
Winter	Winter refers to the period from 1 May to 31 August (inclusive).

Document References

Doc #	Title of document

1. Introduction

Electricity demand drives the amount of investment in Western Power network and many operational decisions for the network. Every year Western Power develops forecasts of peak demand and energy consumption to assist with its planning processes. These plans inform decisions to ensure the network will continue to deliver a safe, reliable and efficient supply of electricity to customers.

Western Power's forecast models cover three planning horizons:

Short-term forecasts covering hours up to several days: these forecasts assist with the operation of the network.

Medium-term forecasts covering forecast horizons up to 10 years: these forecasts drive network plans and development strategies and are used in the setting of network tariffs as part of the Access Arrangements mechanism.

Long-term forecasts covering forecast periods extending to 50 years: these forecasts are primarily 'what-if' scenarios that explain plausible alternative futures based on several assumptions including customer connections growth, changes to consumer behaviour and energy services, and availability of alternative energy technologies. These scenarios are used in the investment decision making process.

This report covers the method used for the development of medium-term forecasts. Within Western Power the Regulation and Investment Assurance, Finance Treasury and Risk, and Grid Transformation functions require the preparation of forecasts to assist with:

- the review of the annual price list as part of the access arrangement. Customer connection, energy export and maximum demand forecasts determine expected revenues.
- the budgeting of maintenance projects.
- Network capacity assessment and network planning. Forecasts allow Western Power to assess growth requirements.

1.1 Forecasting approach

Western Power develops the medium-term forecast models with a segmented 'middle-out' approach. Electricity customers have different consumption and demand patterns. To suitably account for the effect of consumption drivers, forecast models segment customers into four main categories based on their tariff classes: residential, small business, medium business, and large business. Western Power also produces forecast at lower hierarchy levels. Forecasts at different hierarchy levels are reconciled to ensure they provide consistent results – forecasts produced at the bottom of the hierarchy when aggregated provide results consistent with those produced at higher levels.

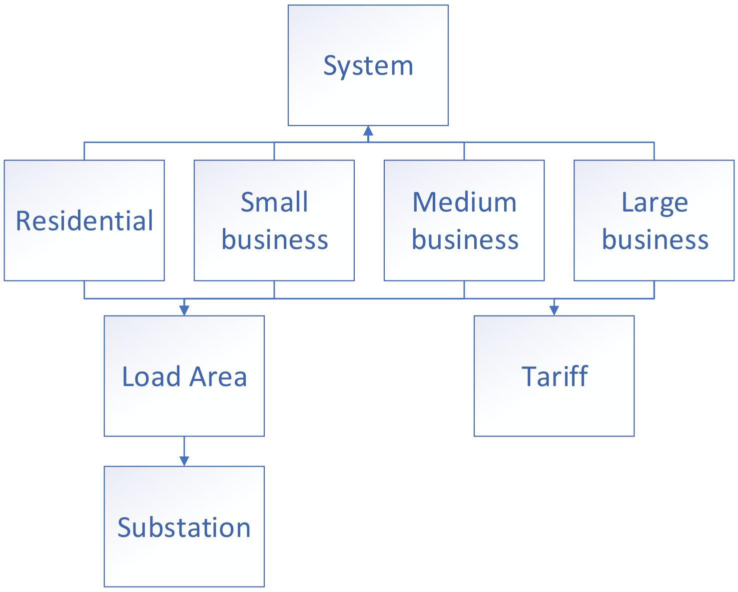
Figure 1.1 depicts the forecast hierarchy. At the lowest levels, the forecast of energy consumption categorises customers based on their connection to each substation and tariff. Tariff categories are consistent with those specified in the Western Power's Access Arrangements.[1] For residential customers this comprises forecasts produced for customers with time-of-use and anytime energy tariffs. For Business

[1] Most recent tariffs are available in the 2017/18 Price List, Amended proposed access arrangement, 28 February 2019, (available online).

HV and Business LV customers forecasts are produced based on contract maximum demand, metered demand and energy-based tariffs

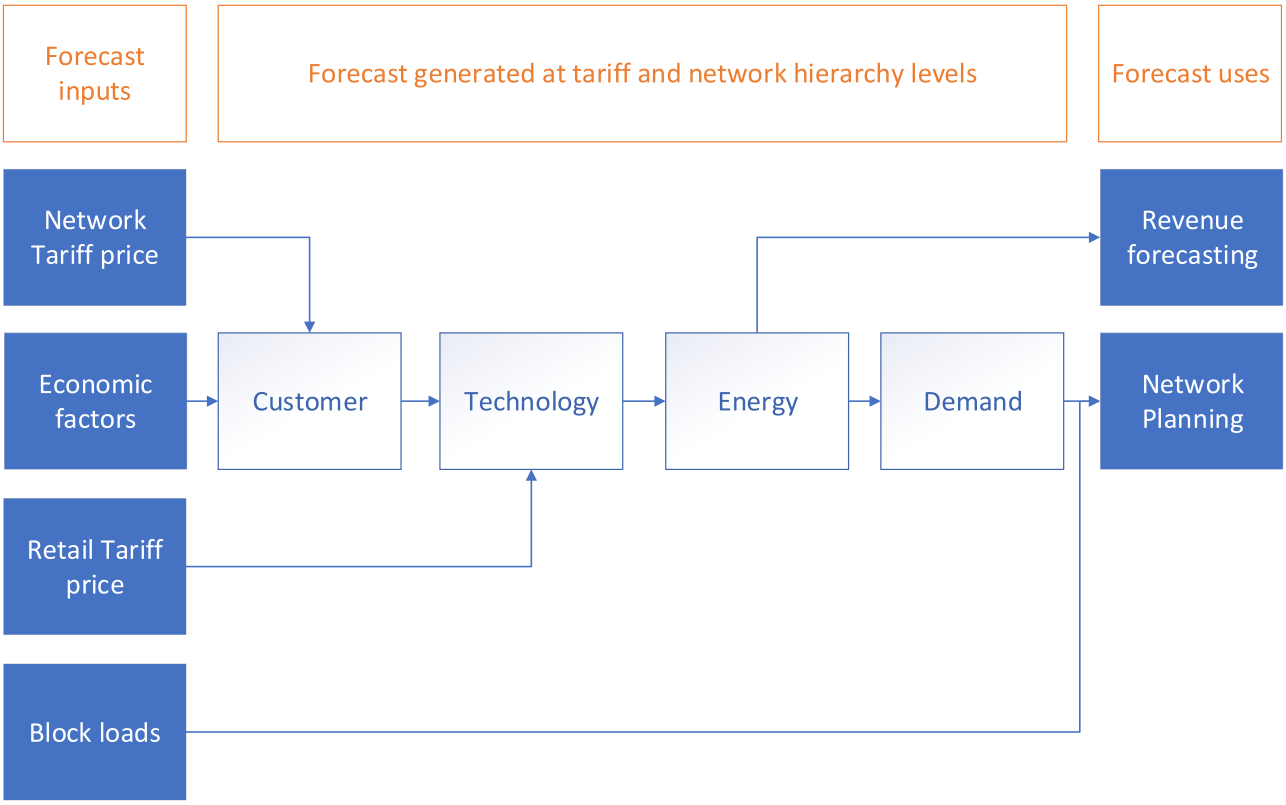
Forecasts of minimum and maximum demand are produced at system, load area, and substation levels only. For short- and medium-term forecast horizons, demand forecasts at tariff levels are not typically required. They also cannot be produced with a reasonable level of accuracy, because for many customers demand is not currently measured frequently.

Figure 1.1. Forecast hierarchy overview



Many inter-related changes in energy supply technologies, consumer behaviour and consumption patterns underpin changes to electricity demand and consumption. Detailed middle-out forecast models use separate forecasts of main underlying variables that drive the consumption of electricity. When combined, these underlying forecasts provide total electricity demand and consumption on the Western Power network. This approach can suitably and transparently incorporate the effect of factors that drive changes to electricity demand and consumption. Figure 1.2 depicts the overall bottom up approach to forecasting demand and consumption.

Figure 1.2. Overall approach to forecasting demand and energy consumption



The method starts with a forecast of customer connections based on regional economic factors and tariff churn based on relative network tariff prices. It then produces a forecast of PV uptake based on connection growth and retail tariff prices, which explain all movements in uptake over the last decade. Then a forecast of energy exported from Western Power network to customers is produced using several statistical models. Explanatory variables for energy forecasts include the results of two underlying forecasts:

- Forecast of customer connections
- Forecast of energy imports from behind-the-meter solar photovoltaic panels

Subsequently Western Power uses a systematic approach to determine the potential 'block loads' that are to be included in the energy consumption and demand forecasts. The method considers the effect of block loads – new customers or changes that represent a material increase or decrease in both energy and load demand – using a post-modelling adjustment process. A block load introduces an often-permanent step-change into demand and energy forecasts. Such step-changes occur infrequently and are not easily accommodated by most statistical methods.

The rapid uptake of solar PV installations and energy efficiency improvements have had a material effect on both customers' demand for network-supplied electricity and the amount of energy exported from the network to customers. Forecast models should suitably account for both the amount and timing of energy generated by solar PV installations. The revised method provides for internal consistency between the forecast of maximum demand and those for the number of connections, solar PV capacity installed and energy exports to customers.

The forecast process also includes forecasts of minimum demand. Increased installation of rooftop and large-scale solar photovoltaic generators is increasingly contributing to challenges for the operation of the network when demand in the network is low – particularly when the sun is shining and generation from solar photovoltaic generators is large, and customers’ electricity demand is low.

This document is organised as follows:

Section 2 briefly articulates the principles guiding main decisions in constructing forecasts

Section 3 explains the data preparation and quality assurance process

Section 4 provides a conceptual overview of the structure and organization of the underlying load growth forecast.

Section 5 describes the methods used to forecast customer connections, energy and Solar PV forecasts.

Section 5 explains how the method used the extreme value theory to create maximum (and minimum) demand forecasts.

Section 6 explains the process for forecasting energy and demand for block loads

Section 7 explains the method for forecasting reactive power in the network

Section 9 explains the forecasting process and reporting

2. Forecasting principles

Western Power has established principles to guide the development of forecast models. These principles guide choices about how the forecasts are done, particularly where there are trade-offs in developing them. For example, there are trade-offs between simplicity and comprehensiveness. More comprehensive forecasts might improve the accuracy of models but typically entail a greater level of complexity and increase costs. While simple models can reduce administration costs and improve forecasting process transparency, they may not provide reasonable results in all applications.

Western Power considers three primary principles in its forecasting: accuracy, transparency and evidence-based decisions. Western Power strives to deliver forecasts that are reasonably accurate and unbiased, transparent and repeatable, and evidence-based and data-driven.

In doing so it identifies and incorporates main factors driving forecasts and makes use of the best available data. The forecasting process checks validity of forecasts by running statistical tests and ensures consistency of forecasts at different levels of aggregation. For the medium-term forecasts particularly, the method ensures consistency between energy consumption, customer count, imported energy from solar PVs, and maximum and minimum demand forecasts.

2.1 Application of principles in practice

Western Power continually improves the quality of its forecasts to ensure forecasts do not contain any systematic bias. A primary source of improvement is the diagnosis of existing forecasts using observed data. Western Power monitors the accuracy of past forecasts using pre-defined accuracy benchmarks; those forecasts not meeting forecast error benchmarks are diagnosed for possible causes of inaccuracy. Any evidence of bias is incorporated in adjustments made in the design of new forecast models or the type and quality of the data relied on to calibrate forecast models.

The forecasting method aims for transparency and repeatability by employing good industry practices including development of work instructions, record keeping and documentation. The forecast process includes clear work instructions that adequately describe each task performed in producing the forecasts.

The process also clearly identifies the source of information and maintains adequate records of input data used. Model development process uses standard practices and techniques – e.g. separation of source data, intermediate calculations and final output – and computer scripts are presented in a readable style that avoid use of hard coded values in the body of scripts.

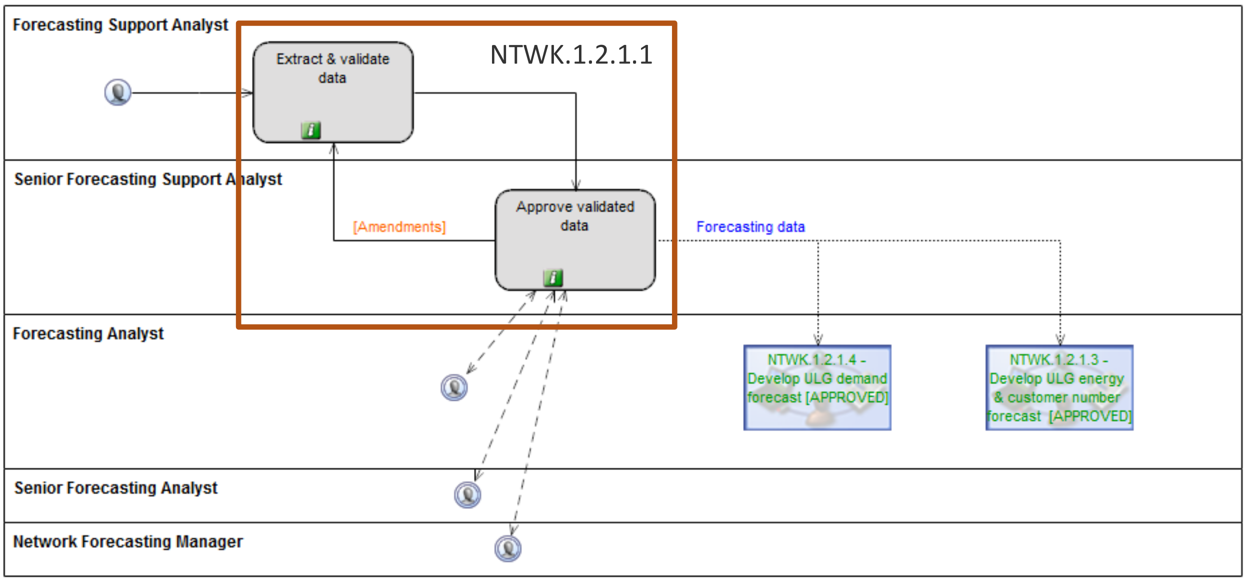
Western Power assesses the credibility of data to ensure it can be relied on for forecasting purposes. This assessment starts with identifying the source of data and an assessment of the credibility of the data sources. The relevance of data is then determined using objective and pre-defined processes to ensure low-quality data does not create any biases in the forecasts.

3. Data preparation and quality assurance

Western Power undertakes a thorough data preparation and validation process using several comparisons, tests and root cause analyses.

The data preparation and quality assurance process is covered by NTWK.1.2.1.1 'Validate forecasting data' and describes the data validation process in the Process Library, as shown in Figure 3.1. The process produces data that feeds into NTWK.1.2.1.3 'Develop underlying load growth energy and customer numbers forecasts' and NTWK.1.2.1.4 'Develop underlying load growth forecasts.

Figure 3.1. Responsibilities for extracting and validating data



The main steps in the process are:

- Extract and validate data
- Approve validated data.

Data validation is a test-driven process. The first test is to establish expectations about the data and then examine the data to determine if a data extract matches those expectations. If the data extract satisfies a specific test, it is accepted without further investigation. Otherwise, the extract is investigated to explain why the data fails to satisfy the test. On failure of a test, Western Power contacts relevant subject matter experts within the business for advice and assistance in investigations and developing remedies.

Tests against established benchmarks

The first round of tests is based on comparisons of a data extract with the last validated data set for the overlapping part of data. Where there is no overlapping history, other measures are implemented based on:

- comparisons with an alternative credible data source
- more detailed benchmarks using past data properties. For example, queries are developed to assess if the latest data exhibits the same properties as the earlier validated data, and
- any established rules based on logical reasoning.

The second round of tests is based on determining the plausibility of the data. This involves statistical profiling the first and second order differences of each five-minute interval (which are a measure of ramping) and comparing them with the distribution of all movements to determine the likelihood that any given minimum or maximum reading was the result of an error or switching event. The results are calibrated to minimise the number of false positives by guaranteeing that no data points from the system minimum and system maximum days are excluded. The rationale being that these days are known to be extreme, and therefore that they are likely to be the maximum days of ramping.

3.1 Tolerance for imperfection

The data that Western Power relies on for establishing forecasts may contain imperfections – for example, due to measurement errors – as with any data set containing measurements. Historic data extracts are validated to ensure that they are sufficiently clean to produce forecasts that are likely to remain within acceptable accuracy ranges. Although imperfections in data sets can create bias in forecasts, the cost of removing imperfections can be prohibitively large. The validation process should suitably account for additional costs of improving the quality of data.

As a broad guideline, the forecasting process deems an imperfect data set as valid for forecasting purposes if:

- It can be demonstrated that the forecasting methods are unlikely to be biased as a result of the inherent data imperfections.
- The imperfections are known and can be effectively mitigated before the data is relied on for forecasting purposes.

In practice, the established heuristic is acceptance of the data where there is less than a two per cent variance with established benchmarks. This is known as the materiality test. Where this test is not satisfied, a root cause analysis is conducted, as explained in section3.2

3.2 Root cause analysis

As the term suggests, root cause analysis identifies the root causes of faults. There is a distinction between a causal factor and a root cause. The defining attribute is that once a root cause has been removed, the fault ceases to exist. In the context of energy volume and connection numbers data, a double count in a query script is a good example of a root cause. Effective root cause analysis:

- is performed systematically
- is backed up by evidence, typically specimens illustrating a source of the fault
- has an adequate description of each problem
- ensures recommended corrective actions are undertaken.

Problems are identified when a formal comparative benchmark test fails. Comparative tests are established in a top-down order. An example of a comparative test relating to reconciliation of Balcatta zone substation data is presented below.

Box 1. Comparative benchmark test example

Measurement error is a primary risk when forecasting demand. Examples of measurement error include: occasional SCADA sensors fail; incorrect calibration and reassignment of measurement tags with a lag in updates of database records.

For these reasons, Western Power regularly and systematically validates and corrects data before use in forecasting processes. Determining whether measurements can be considered to be correct involves a series of cross-checks across SCADA measurements of electricity flows entering, transitioning and exiting zone substations. In addition, Western Power cross-checks SCADA and metering data by zone substation.

This process identified that approximately 2,000 Balcatta zone customers were incorrectly configured in the metering data. Correcting this led to a material re-evaluation of trend across Balcatta zone substations with respect to the neighbouring zone substations.

Comparative benchmark testing involves comparing latest data extracts to previous extracts and cross-matched against similar extracts from other sources (e.g. extracts from metering and customer care and billing system can be cross-matched against SCADA[2] data for a load area level).[3] The test fails if the average difference is outside a prescribed tolerance (e.g. more than two per cent). On failure, the following activities are undertaken:

1. Determine if a data correction has been implemented since the last extract. If so, document the previously unidentified problem with the previous extract.
2. If there is no correction implemented:

Examine the time series of energy values to determine if there has been a step-change in energy numbers at a point in time. If this is identified, check that this is genuine, e.g. confirm that there was a new customer that used that energy.

Compare connection numbers. If this test passes, then there must be erroneous meter readings.

If this test fails, then there is likely to be a problem with meter counts. Follow up with a test of the Data Generating Process. This assumes that the latest data extract will exhibit the same data properties as previous extracts based on any one of the following:

- Box plot overlay of the latest data points. Latest monthly observations should fall within the “box”.
- Similar seasonal pattern as defined by the SAS DECOMP procedure. Compare the seasonal patterns available in the seasonal component (SAS SC Component).
- Similar correlations and Augmented Dickey Fuller test results to those as previously established.

2 The SCADA (Supervisory Control and Data Acquisition) system is Western Power's system for managing the electricity network, both day to day and in emergencies. The system monitors and records network data.

3 This involves using data retrieved from NetCIS, which is Western Power’s core customer care and billing system. NetCIS stores most of the company’s customer information and interactions and is used daily throughout the organisation. It is the system used to calculate and invoice network access charges and bill these charges to electricity retailers.

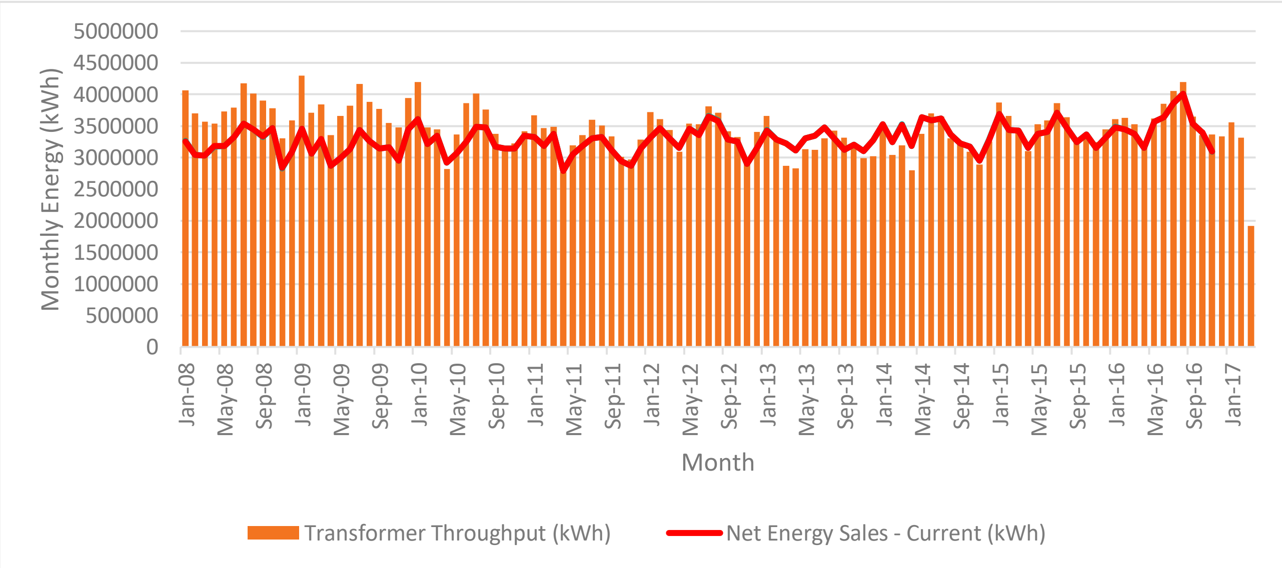
3.3 Data validation and quality checking process

Western Power follows good industry practice and has implemented a robust data validation and quality checking process. The Senior Forecasting Analyst checks to ensure that:

- data validation process has been followed by the analysts preparing the data
- data either conforms to established benchmarks within prescribed tolerances or there is an evidence-based reason for identified anomalies.

Below is an example of the data validation process used in 2017, using the Wundowie substation as an example.

Figure 3.2. Wundowie (WUN) substation data from the validation process



In this example, the net energy sales suitably align with the substation transformer throughput, having regard to distribution losses and metering inaccuracies. The alignment demonstrated in Wundowie substation data provides a high degree of confidence in the metering and SCADA data for all forecasting including customer connections, solar PV, energy and demand.

3.4 Forecasting Approach

Forecasting approach and input data Western Power applies time series statistical models to most of the commercial and residential customer connection and energy and solar PV forecasts. Many of these forecasts are produced using automatic functions of the forecasting software platform. The forecast process also allows for manual construction of some residential and commercial forecasts where required, typically as a correction to an automatically generated forecast. For instance, Western Power manually adjusts forecasts for industrial customers when advised of changes in demand, because these forecasts are typically flat (i.e. have no growth).

Western Power's medium-term forecasts do not include all possible anticipated technological innovations (e.g. electric vehicles and batteries); however, they are being monitored and may be included in future forecasts if received evidence suggests that inclusion would be prudent.

3.5 Time series statistics methods

All models for customer connection, energy and solar PV forecasts are created using SAS and R as discussed in process NTWK.1.2.1.6 Develop demand forecast. These models employ three broad styles of time series forecast models:

- Autoregressive integrated moving average (ARIMA)
- Unobserved components models
- Multivariate regression.[4]

These tools use best practice forecast diagnostic and model building processes.

The forecasts are produced in a hierarchy, which permits the use of both space and time dimensions to maximise model flexibility, resulting in improved precision in model coefficients. In addition, the forecasts are reconciled so that the forecast sub-groups add up to the total.

In 2016 and 2019, several new tariffs were introduced with customers reallocated from other tariffs. For example, in 2015 residential customers who owned solar photovoltaic power systems were allocated to RT1 tariff. In 2016, these customers were reallocated to the tariff RT13. The reallocation of energy volume and customer numbers to the new tariffs impedes comparison of tariff-based customer numbers and energy volumes between forecast reports. To overcome this issue an approximate reconstruction of the previous tariff structures can be produced for the purposes of comparison.

While most of the forecasts are produced automatically, forecast models can (and have been) manually constructed and selected. SAS and R provide a wide array of diagnostic test results that are relatively easy to interpret but do require a high degree of expertise to use effectively.

3.6 Econometric forecasts

Given that a large number and wide variety of statistical models are used to produce the forecasts, it is not practical to describe each forecast model applicable to each tariff. Instead, this report describes the models employed within broader stylistic structures and themes.

---

4 Multivariate regression is a technique that estimates a single regression model with more than one outcome variable.

3.6.1 Benefits of employing reduced form models

In the past Western Power produced forecasts of customer counts and energy consumption based on long-term structural models. Those models provided direct estimates of the effect of economic variables on energy and demand forecasts. They, for example, modelled customer connections and energy consumption based on responses to variation in electricity tariffs and income or economic activity.

While such models are highly desirable for long-term business planning, directly estimated structural models often perform poorly as short- or medium-term forecast models. Several statistical issues contribute to their poor performance including incorrectly specified dynamics and insufficient variation in explanatory variables, such as tariffs.

To overcome such problems, Western Power has employed time series methods. Time-series models are reduced-form models as opposed to structural models.[5] That means, for example, that the estimated regression coefficients are not economic parameters such as long-run price and income elasticities, as is the case for structural models.

The benefit of employing reduced form models is that data-driven diagnostic and model building methods capture the short-run dynamics contained in the data. For example, most of the monthly energy volume series exhibit a high degree of serial correlation. This means that ARIMA models, which exploit serial correlation, produce reasonably accurate short-term forecasts.

3.6.2 Underlying drivers of electricity demand

This section provides a description of the external (i.e. independent) variables included in the forecast training data set[6]. Selection of these variables is justified by economic or demonstrated statistical relevance. Other variables could also have been included but their historical values or forecasts are not available.

The following categories apply to the selected external variables shown in Table 1:

- economic activity: variables that measure the level of activity in the economy, such as gross regional product, regional final demand and consumer price index.
- Electricity retail price: volumetric and daily components of the electricity price
- substitution of network electricity: to capture any influence of alternatives to network delivered electricity.

Table 1. External variable description

Category	Variable	Description
Economic	Regional final demand	SWIS regional final demand forecasts (in million $) (annual, monthly and percentage change) prepared by BIS Oxford Economics.
Economic	Gross regional product	SWIS gross regional product forecasts (in million $) (Annual, Monthly and percentage change) prepared by BIS Oxford Economics.

[5] See James D. Hamilton (1994), Time Series Analysis, Princeton University Press, pp. 244–246

[6] The training data set, also known as the estimating data set, is the data used to calculate forecast model parameters

Category	Variable	Description
Price	Tariff A1	Synergy residential retail tariff, variable (in $/kWh) and fix (in $/day) parts of the tariff. Forecast tariffs sourced from budget papers. Tariffs beyond the budget paper forecast horizon were assumed to remain constant in real terms.
Price	Tariff L1	Synergy business retail tariff, variable (in $/kWh) and fix (in $/day) parts of the tariff. Forecast tariffs sourced from budget papers. Tariffs beyond the budget paper forecast horizon were assumed to remain constant in real terms.
Substitution	PV count	Count of customers with a bidirectional network tariff
Substitution	PV capacity	Sum of PV inverter capacity (in MW)

As indicated above there are many variables included in the estimating data set. Many of these variables are highly correlated, so most of Western Power's forecast models only include a small subset of these variables based on a balance of goodness of fit and forecasting accuracy criteria.

The variables and associated data are from published documents by Western Australian Government agencies, and Australian Bureau of Statistics. Western Power engaged BIS Oxford Economics[7] to provide estimates and forecasts of gross regional product and final demand based on the area covered by the Western Power network.

Gross regional product is a broad measure of economic activity that typically accounts for a small component of monthly variation in electricity demand. Nevertheless, it will have an influence on the medium- to long-term trend in electricity demand since electricity is one of the inputs for productive economic activity.

Note that other economic measures of overseas demand for Western Australia's exports (such as exchange rates, the terms of trade, commodity prices etc.) are also likely to be influential on electricity demand. However, the relatively short time series of electricity demand impedes precise estimation of the impact on electricity demand. Moreover, the high volatility of these series and the absence of credible long-term forecasts for these additional economic variables limit the suitability of their use.

Electricity prices have an inverse relationship with electricity demand, or at least network delivered electricity demand. Assuming fixed customers' budgets in the short-term, a higher price (i.e. higher electricity tariffs) should have a persistent dampening effect on electricity demand. This is an important factor for long-term forecasting.

An issue limiting the usefulness of the tariff is limited variation in prices. Typically, consumer electricity prices update just once a year. With just nine years of time series data, it is difficult to estimate a statistically precise relationship between electricity prices and the demand for network delivered electricity.

[7] BIS Oxford Economics "SWIS REGION & WESTRN AUSTRALIA ECONOMIC FORECASTS TO FY2031". SWIS Region & WA Forecasts - 4.11.21.docx (http://edm.westernpower.com.au/otcs/cs.exe?func=ll&objaction=overview&objid=61773829)

3.7 Prospect of further substitution via new technology

As of 2017, speculation about a mass adoption of other network competing technologies was intensified. At the time it was expected that a rapid adoption of battery storage systems could materially decrease electricity demand delivered through the network.

Given those expectations, in 2017 Western Power developed a model to assess customers' incentive to adopt solar PV and battery storage systems for its forecasts at the time. The model considered factors affecting the adoption of battery systems including the present value of the cost of a battery system in compare to network delivered electricity in several scenarios.[8]

Results showed that that the average customer did not have the incentive to install batteries, but that many customers had the incentive to partially load defect in the next five to 10 years. For its medium-term forecasts in 2017, Western Power assumed substantial battery uptake using the results of the adoption model.

Western Power has been actively monitoring developments in battery costs and installations and conducted further assessments and revisited its forecasts based on observed trends in battery uptakes since 2017. Western Power has estimated that the effect of residential battery installations on peak demand during the forecast horizon is small.

Western Power also drew on forecasts produced by CSIRO and AEMO. AEMO's estimates of battery storage uptake and relevant effect on peak demand were consistent with those assessed by Western Power. AEMO estimated that batteries' influence on peak demand in the SWIS will be approximately 9 MW in 2024/25.[9]

Western Power estimates show that residential customers' individual contribution to peak demand is approximately 800 watts. In 2019, CSIRO developed forecasts of behind-the-meter battery storage systems for residential and commercial sectors in Western Australia.[10] CSIRO forecasted approximately 5,000 residential battery installations by 2025. These installations are likely to reduce peak demand in the SWIS by four megawatts.

Given the expected small effect of battery storage systems during the forecast horizon, Western Power's medium-term forecasts in 2020 did not make any adjustment for the effect of battery storage systems.

The 2019 forecasts did not include any adjustment for the uptake of electric vehicles. In its central scenario, CSIRO forecasted that electric vehicles reach parity with the upfront cost of internal combustion vehicles in 2030.[11] Before then, the effect of electric vehicles on energy consumption and demand was small.[12]

---

8 For details of the model results refer to Connections, Energy and Demand Forecast Methodology, Access Arrangement Supplementary, 2 October 2017 (available online).

9 AEMO, 2019. 2019 Electricity Statement of Opportunities, A report for the Wholesale Electricity Market, p.32 (available online).

10 CSIRO, 2019. Projections for small scale embedded energy technologies, Report to AEMO (available online).

11 Ibid, p. 52.

12 Based on CSIRO's forecasts, AEMO estimated that by 2024/25 electric vehicles will increase peak demand in the SWIS between 1.1 MW and 17.2 MW (and network energy consumption between 8.7 GWh and 122.0 GWh). Refer to AEMO,2019. 2019 Electricity Statement of Opportunities, A report for the Wholesale Electricity Market, p. 34 (available online).

4. Customer connections, solar PV and energy forecast methods

The method that produces the forecast of energy exports from the network is based on three separate trends: number of customer connections, adoption of solar PVs and energy imports from solar PVs. This allows the model to reliably incorporate the effect of socio-economic and technological forces that result in highly dynamic and evolving energy consumption patterns. These forecasts are further used for forecasting maximum and minimum demand.

4.1 Number of connections forecast

The forecast includes gross regional product, gross regional demand and regional population in the set of predictors to create a monthly forecast of the number of connections. Number of connections comprises counts of National Metering Identifier (NMI) and connection counts for streetlights and unmetered supplies.

4.1.1 Number of connections forecast by network hierarchy

The forecast of connections by network hierarchy is defined in order by customer type and supply area, as depicted in section 1, Figure 1.1 . The forecast also reconciles forecasts generated for each hierarchy level to ensure they suitably add up to higher or lower levels. Without reconciliation forecasts at the top level of hierarchy would not equal the sum of the forecasts at the middle level or at the bottom level. By reconciling the differences, the model uses the information contained in the hierarchical structure that might improve the forecasts.

From a range of candidate time series model, SAS produces connection forecasts primarily using the following time series statistics methods:

- Auto-Regressive Integrated Moving Average (ARIMA) method with external regressors as well as Vector Auto-Regressive (VAR) methods
- Unobserved Components Models

For each time series in the hierarchy, SAS performs diagnoses of the time series using time series analysis techniques. It then creates a list of candidate model specifications based on the diagnostics. The analyst then selects the most appropriate model specification and set of predictors based on either in-sample or holdout sample evaluation using a model selection criterion.

SAS reconciles the model forecasts to form consistent reconciled forecasts across the defined hierarchy. The reconciliation method also evaluates the forecast using in-sample analysis and provides for out-of-sample analysis of forecast performance.

Design of customer connections forecast

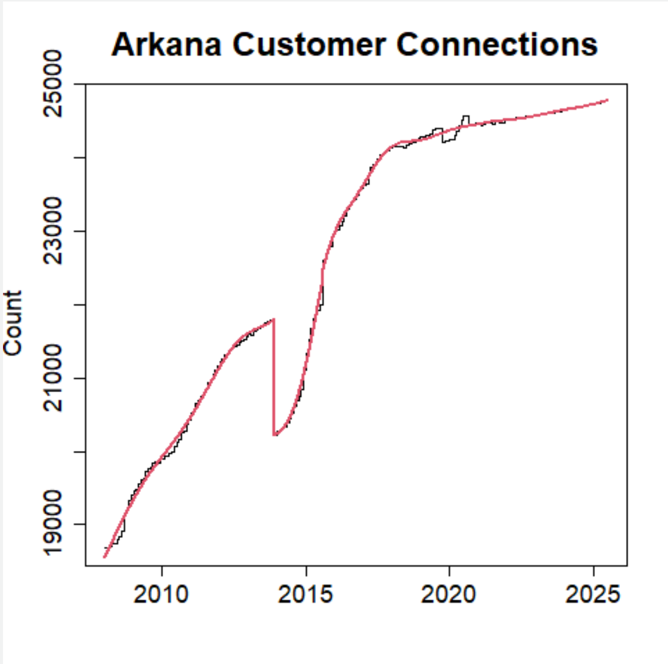
The forecast of number of connections is set up based on regional population and economic predictors in SAS.

The forecast is set up based on current geographical coverage of substations at the time of generating forecast. This ensures that changes to electrical configuration of the network will not distort the forecast of customer counts. The model applies historical changes to network configuration after forecasting connection counts, which will be incorporated in minimum and maximum demand forecasts.

This is presented using Arkana substation forecast in Figure 4.1. The historical connection count included in the sample shows the monthly number of connections covered by the Arkana substation. In late 2013, Western Power transferred some of the connections in Arkana substation to the newly commissioned Balcatta substation. The connection count forecast for the Arkana substation, however, uses the most recent coverage configuration of the substation to ensure that the economic and population predictors can suitably explain changes to connection counts. The forecasts generated by this design also represent the forecast of connection counts if the current network configuration remains unchanged throughout the forecast horizon.

Historical changes to the configuration of network are reflected in minimum and maximum demand forecasts. This historical adjustment is important because minimum and maximum forecasts should incorporate these changes in the fitting of extreme demand models. For the Arkana substation this adjustment is presented in Figure 4.1. The Extreme Value Theory (EVT) model uses the trend in connection counts incorporating historical electrical reconfigurations.

Figure 4.1. Arkana substation number of connections forecast



In Figure 4.1, the black line shows history and forecast connection counts and the red line shows a piecewise smoothing model fitted to capture trend in connection counts.

4.1.2 Customer connections forecast by tariff

A separate forecast of connections per tariff is produced to minimise the amount of work required to create the revenue forecasts, these numbers are reconciled to the system total to guarantee consistency between the outputs.

An Unobserved Components Model is used to create a forecast of the underlying growth in each tariff and customer type combination. The combination is important because some tariffs, such as AER/RT1, cannot neatly be defined as either residential or small business due to the presence of home businesses, schools, and charities which are allowed onto the tariff. These non-residential connections have their own distinct trends which are important to forecast independently. Unidirectional and bidirectional tariffs are merged at this stage because the underlying growth of connections in each tariff appears to be independent of the decision to install solar photovoltaic systems.

A second model of movement between tariffs is then used to forecast overall changes in connection numbers by tariff. Movement between tariffs is the result of two independent drivers – installation of solar photovoltaic systems and cost minimisation by electricity retailers. Historic uptake of solar photovoltaic systems has been closely linked to changes in retail tariff prices, such as Feed In Tariff (2011), Renewable Energy Buyback Scheme (2012), above inflation increases in network tariff prices (2017), and potentially Distributed Energy Buyback Scheme (2020). Moving a connection to a different network tariff is a cost minimisation strategy by retailers because they are under no obligation to pass on pricing signals from network tariffs to customers. Historic movements have often reflected either price gaps or long-term changes to customer behaviour that create an incentive. Movement has been high in recent years due to the introduction of several new tariffs in July 2019, and the changes to prices of these new tariffs in July 2020. The model of these movements is a linear regression based on network and retail tariff prices, which account for the relative arbitrage opportunity.

The forecast of Streetlight connections is performed as a linear regression against overall connections on the network because the number of streetlights is closely correlated with new developments. The forecast of Unmetered Supply is performed using an ARIMA model because the growth in this segment appears to be uncorrelated with any forecastable explanatory data.

4.2 Forecasting solar PV capacity

Producing reliable long-term forecasts for the number of solar PV installations is important to developing accurate forecasts for electricity consumption and demand. Although the mass adoption of solar PV installations is a relatively recent phenomenon, the rate of adoption has had a material demand-reducing impact. Given its importance, Western Power conducted several investigations into forecasting methods for solar PV capacity and counts.

Solar PV Capacity (kVA) is the sum of photovoltaic (PV) inverter capacities for solar panels that are installed on customer's premises and connected downstream from the substation. The PV capacity has been demonstrated as a key driver that influences energy consumption and peak demand. This year, the primary focus was the development of a data-driven approach to forecasting PV capacity for each substation as explained in more detail below.

4.2.1 PV capacity forecast

Monthly PV capacity forecast was developed for four system-tariff sectors (Residential, Small Business, Medium Business, and Large Business) per substation. To proceed, for a given tariff sector of a substation, a linear regression model was first fitted to the increment of PV capacity and driving factors including customer connections, tariff and service changes from the observed data from January 2008 to June 2020. Then forecast of PV capacity after June 2020 is obtained by the sum of the previous month PV capacity and the forecasted monthly increment of PV capacity estimated by the regression model.

Regression model applied to the increment of PV capacity $\Delta Res\_PVC_{i,t+1} =: Res\_PVC_{i,t+1} - Res_{PVC_{i,t}}$ of residential sector from month t to month $t+1$:

$$\Delta Res\_PVC_{i,t+1} = \beta_0 + \beta_1 Res\_NMI_{i,t+1} + \beta_2 TariffA1_{t+1} + \beta_3 TariffA1\_SC_{t+1} + \epsilon_{i,t+1}$$

The forecast of $Res\_PVC_{i,t+1}$ can be obtained by

$$Res\_PVC_{i,t+1} = Res\_PVC_{i,t} + \hat{\beta}_0 + \hat{\beta}_1 Res\_NMI_{i,t+1} + \hat{\beta}_2 TariffA1_{t+1} + \hat{\beta}_3 TariffA1\_SC_{t+1}$$

Regression model applied to the increment of PV capacity of (Small, Medium and Large) business:

$$\Delta Bus\_PVC_{i,t+1} = \gamma_0 + \gamma_0 Bus\_NMI_{i,t+1} + \beta_1 TariffL1_{t+1} + \beta_2 TariffL1\_SC_{t+1} + \epsilon_{i,t+1}$$

The forecast of $Bus\_PVC_{i,t+1}$ can be obtained by

$$Bus\_PVC_{i,t+1} = Bus\_PVC_{i,t} + \Delta Bus\_PVC_{i,t+1}$$

Variables and their descriptions are detailed in Table 2.

Table 2. Historical and forecast input data source for PV capacity modelling

Variable	Description	Unit
$Res\_PVC_{i,t+1}$	Monthly residential PV capacity for substation i in month $t+1$	KVA
$\Delta Res\_PVC_{i,t}$	Monthly increment of residential PV capacity for substation i in month t	KVA
$Bus\_PVC_{i,t+1}$	Monthly business PV capacity for substation i in month $t+1$ month	KVA
$\Delta Bus\_PVC_{i,t}$	Monthly increment of PV capacity for substation i in month t	KVA
$Res\_NMI_{i,t}$	Monthly residential connections (including both historical and forecasted values) for substation i in t month from January 2008 to June 2025	Count
$Bus\_NMI_{i,t}$	Monthly (Small, Medium and Lager) business connections (including both historical and forecasted values) for substation i in t month from January 2008 to June 2025	Count
$TariffA1_t$	Residential tariff in t month from January 2008 to June 2025, source from Energy Operators (Electricity Generation and Retail Corporation) (Charges) By-laws 2006[13]	cent/kWh
$TariffA1\_SC_t$	Residential service charge in t month from January 2008 to June 2025, source from Energy Operators (Electricity Generation and Retail Corporation) (Charges) By-laws 2006	cent/day
$TariffL1_t$	Business tariff in t month from January 2008 to June 2025, source from Energy Operators (Electricity Generation and Retail Corporation) (Charges) By-laws 2006	cent/kWh
$TariffA1\_SC_t$	Business service charge in t month from January 2008 to June 2025, source from Energy Operators (Electricity Generation and Retail Corporation) (Charges) By-laws 2006	cent/day

4.3 Energy forecasts

Energy forecast model is developed in SAS and produces separate forecasts for exported energy from the grid and imported energy from solar photovoltaic panels.

The model produces monthly forecasts at different hierarchy levels comprising tariff type, customer segment, and substation levels. It also reconciles forecasts at different hierarchy levels, as explained in section 4.1.1

Energy exported to different customer segments is driven by different factors and the model suitably accounts for such differences by splitting the customers to different segments and/or producing forecasts for each segment separately. For instance, the model develops separate forecasts for large business users.

13 Energy Operators (Electricity Generation and Retail Corporation) (Charges) By-laws 2006, (available online).

The forecast of large business users is further adjusted by information available to Western Power at the time of producing forecast. This is explained in more details in section 6.

4.3.1 Import energy forecast

The model of import energy from medium business, small business, and residential connects assumes that the source of generation is solar photovoltaics. It uses a linear regression model to predict the ratio of energy imported per MVA of installation, which is the multiplied by the forecast capacity to estimate future solar imports. It is a purely autoregressive model because this ratio has been quite stable over time.

A separate model also estimates large business customers' energy import from their embedded generators by an autoregressive model of historical import energy. These customers cannot be assumed to have solar photovoltaic systems and are generally observed to have stable consumption patterns.

4.3.2 Export energy forecasts

This model produces separate forecasts of energy exports for all customer segments and all tariffs on an average per connection basis. It develops several unobserved component models and seasonal decomposition models using different combinations of hyperparameters and explanatory variables comprising solar photovoltaic capacity and customer connections. The model provides diagnostics for model selection and results are inspected by the user to select desirable models.

Customer segment is a relatively coarse grouping, so forecasts produced at this level may not provide a robust estimate of future energy exports for each tariff. To remedy this problem, the model adjusts the tariff level export energy forecasts to ensure they are consistent with the more robust forecasts produced at customer segment level. The model adjusts tariff level forecasts by the ratio of energy exports from customer segment level forecast to energy exports from tariff level forecasts.

4.3.3 Energy export forecast for streetlights and unmetered supplies

The forecast of Streetlight connections is performed as a linear regression against overall connections on the network because the number of streetlights is closely correlated with new developments. The forecast of Unmetered Supply is performed using an ARIMA model because the growth in this segment appears to be uncorrelated with any forecastable explanatory data

5. Maximum and minimum demand forecasts

Western Power improved its method for forecasting maximum demand. The revised method provides for the internal consistency between the forecast of maximum demand and those for the number of connections, energy exports to customers and solar PV capacity installed.

To improve its forecasts, Western Power has developed a novel method for forecasting substation maximum and minimum demands using the extreme value theory. The new method has been peer-reviewed and published in the international journal *IEEE Transactions on Power Systems.*[14]

The new forecast method has two innovative features:

- it ensures an internal consistency between demand, customer count, energy consumption and solar PV forecasts. Maximum demand is modelled as a function of trends in three common factors already required by utilities, including customer count, energy consumption, and installed photovoltaic capacity.
- it is robust to changes to network configurations – e.g., substation transfers.

The new forecast method also provides forecasts of minimum demand. The increased penetration of behind-the-meter solar PV systems has created challenges for the operation of the network when customers' demand is met by energy imported from solar panels during daytime and when demand for network delivered electricity is low during mild weather conditions. It is important that network development and operation plans suitably consider changes in minimum demand.

5.1 Introduction

The forecast method produces substation maximum and minimum demand forecasts using extreme value theory.

Extreme value theory

- Extreme value theory is a branch of statistics dealing with the extreme deviations from the median of probability distributions. It seeks to assess, from a given ordered sample of a given random variable, the probability of events that are more extreme than any previously observed
- Extreme value analysis is widely used in many disciplines, such as structural engineering, finance, earth sciences, traffic prediction, and geological engineering. However, appropriate statistical tools based on extreme value theory have rarely been used to analyse annual maximum electricity demand
- Western Power has developed a novel method for forecasting substation maximum and minimum demands using the extreme value theory

Due to variability, forecast maximum demand is expressed at three probability levels of 10, 50 and 90 per cent probability of exceedance (PoE), rather than a point forecast. For any given season or year, PoE10, PoE50 and PoE90 are defined as below:

[14] Y. Li and B. Jones, "The Use of Extreme Value Theory for Forecasting Long-Term Substation Maximum Electricity Demand," in *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 128-139, Jan. 2020, doi: 10.1109/TPWRS.2019.2930113.

- PoE10 or 10 per cent PoE maximum demand value is expected to be exceeded, on average, one year in 10.
- PoE50 or 50 per cent PoE maximum demand value is expected to be exceeded, on average, one year in two.
- PoE90 or 90 per cent PoE maximum demand value is expected to be exceeded, on average, nine year in 10.

Despite the extreme nature of annual maximum demand, the statistical theory of extreme values has only rarely been applied. Network operators typically complete energy consumption and maximum demand forecasts separately through two different processes, leading to inconsistent results. The recent uptake of solar PV systems and changes to energy services have driven changes to system demand. In many instances, Western Power's previous approach to demand forecasts could no longer produce reliable forecasts.

The previous demand forecast method applied a load factor to monthly average demand for calculating maximum demand forecasts.[15] For instance, for the Arkana substation the previous forecast underestimated demand despite a larger forecast prediction interval, when compared to the new method. None of the observed peak demands between 2015 and 2018 fell in the forecast prediction interval. The forecast produced using the EVT, however, could cover two out of four observed annual peak demands within a relatively narrow prediction interval.

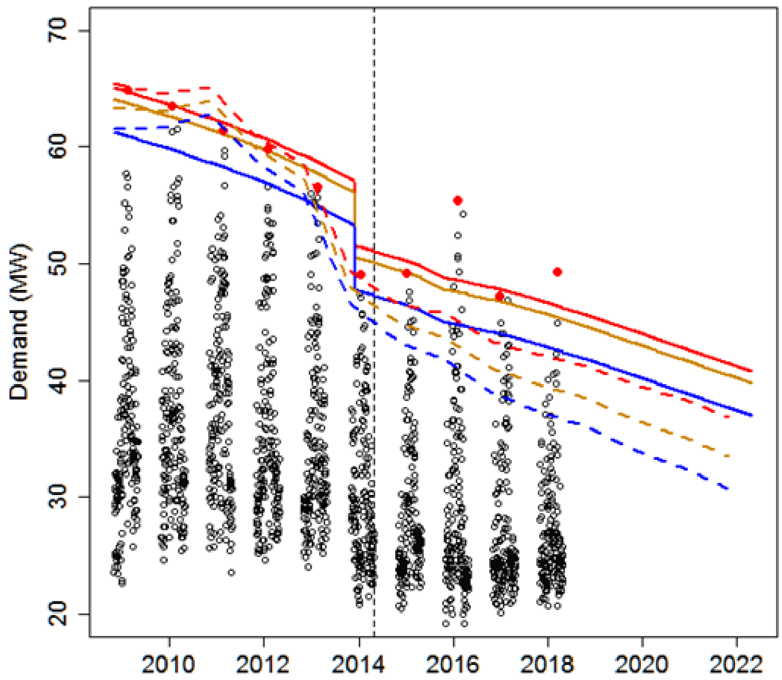


Figure 5. A comparison of forecasts produced using the EVT and the previous load factor adjustment methods.[16]

15 For further details about the application of load factor in the calculation of maximum demand refer to Connections, Energy and Demand Forecast Methodology, Access Arrangement Supplementary, pp.22–27 (https://www.erawa.com.au/cproot/18941/2/WPAA4%20-%20ProposedAAattachment7.3.1%20-%20Connections%20Energy%20and%20Demand%20Forecast%20Methodology.pdf)

16 Models are calibrated with data in the training period from 2008 to 2014, the end of training data marked as the vertical dash line. The PoE10, PoE50 and PoE90 estimated from the point process models are marked as red, orange, and blue solid lines and those based on the load factor forecast approach are masked as dashed lines, respectively.

The new peak demand forecast model uses a point process model from the extreme value theory. It forecasts substation maximum (or minimum) demand as a function of customer count, average demand, and installed photovoltaic capacity trends.

Point process model

- Point process model from extreme value theory combines modelling the occurrence of the extreme demand over a high threshold and the intensity of extreme demand as a two-dimensional Poisson point process in the sense that
  - The occurrence of extreme demand is assumed to have a Poisson distribution.
  - The intensity parameter of extreme demand is driven by a generalized extreme value distribution
- The parameters of point process model can be parameterised as generalised extreme value distributions which are driven by energy consumption, customer connection and solar capacity trends.

As the generalized extreme value distribution governs the behaviours of block maxima (annual maximum demand), maximum demand can be estimated for different quantiles, or i.e. PoE10, PoE50 and PoE90, as required by planning standards.

5.2 Data preparation

5.2.1 Energy consumption, customer count and solar PV capacity data

Electricity demand on Western Power's network is measured and aggregated to five-minute averages for the purposes of forecasting. Data preparation process for both maximum and minimum demand further calculates daily maximum and minimum data based on the five-minute demand data.

The point process model uses forecasts of customer connections, energy, and solar PV capacity forecasts explained in section 4 to explain the variation in the peak demand. Customer connections is the number of unique customers connected to the existing network (i.e. counts of National Metering Identifier (NMI) and connection counts for streetlights and unmetered supplies). Over time, this series will change as new customers connect the existing network or as network configuration changes, switching customers from one substation to another. For the forecasts developed in 2020, monthly customer connection series from January 2008 to June 2025 is available: part history from January 2008 to June 2020 and part forecast from July 2020 to June 2025. The customer count series is then made daily and minor inconsistencies smoothed by fitting a piecewise smooth trend[17] in order to capture changes in customer connections. Seasonality is less critical in customer connection, and the trend is used simply to smooth minor variation (e.g. Figure 5.1a).

Energy consumption forecasts, both exports (energy to customers) and imports (energy to the network) including energy generated from solar PVs), are prepared by network tariff and monthly average demand series are constructed using the following equation:

[17] Y. Li and B. Jones, "The Use of Extreme Value Theory for Forecasting Long-Term Substation Maximum Electricity Demand," in *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 128-139, Jan. 2020, doi: 10.1109/TPWRS.2019.2930113.

$$Energy\ comsumption = sum\ of\ export\ energy - sum\ of\ import\ energy$$

For the forecasts developed in 2020, the monthly energy consumption (kWh) series from January 2008 to June 2025 is available: part history from January 2008 to June 2020 and part forecast from July 2020 to June 2025 (based on forecasts explained in section 4.3). The available monthly series is then allocated evenly to each day within a month. The trend in average demand at the substation is then obtained by fitting a piecewise smooth trend such that all seasonality and weather impacts are removed. The piecewise smooth trend was used to catch up step changes when there are network transfers at the substation (e.g. Figure 5.1b).

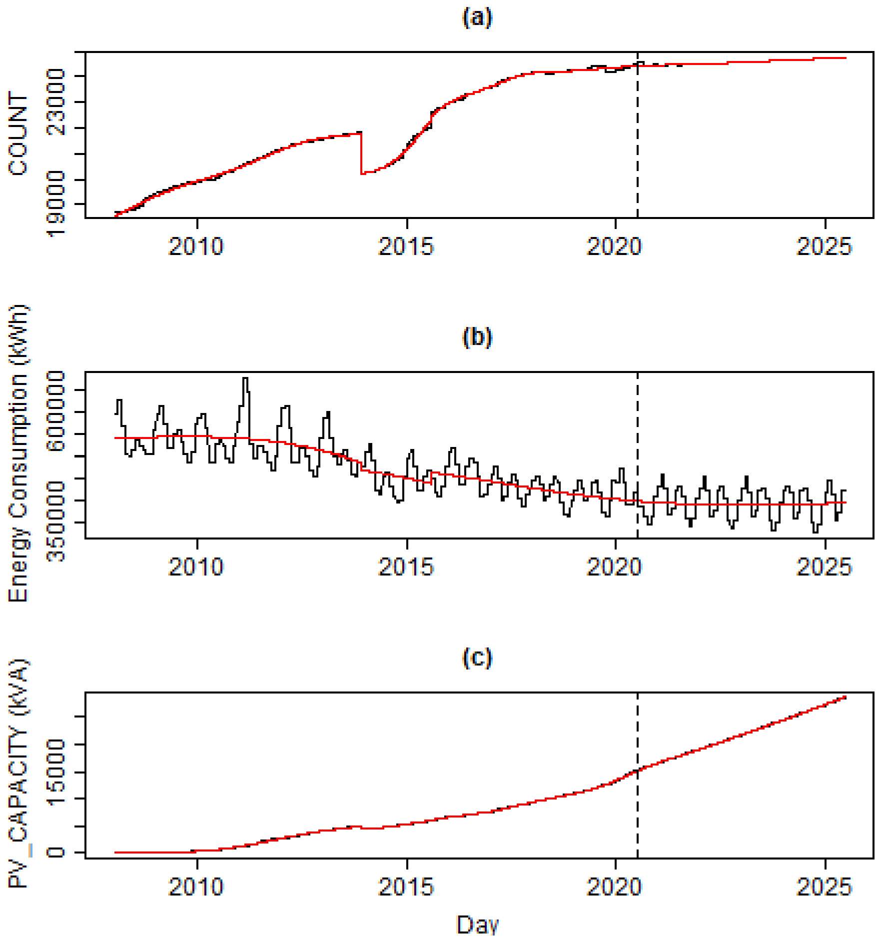


Figure 5.1 Time series of observed (black) and fitted trends (red) for customer count, panel (a), energy consumption, panel (b), and solar PV capacity, panel (c), for Arkana substation.[18]

PV Capacity (kVA) is the sum of photovoltaic (PV) inverter capacities for solar panels that are installed on customer's premises and connected downstream from the substation. Over time, this series will increase as new solar panels are installed but may vary due to network transfers that switch customers and therefore PVs from one substation to another. For the forecasts developed in 2020, a monthly data series January 2008 to June 2025 is available: part history from January 2008 to June 2020 and part forecast from July 2020 to June 2025. The PV capacity series is then made daily and minor inconsistencies smoothed by fitting

18 The vertical dashed line is the end of period for observed data on 30 June 2020.

a piecewise smooth trend. Like the customer count, the seasonality is less critical and, therefore, used simply to smooth minor variation (e.g. Figure 5.1c).

5.2.2 Outlier detection and removal

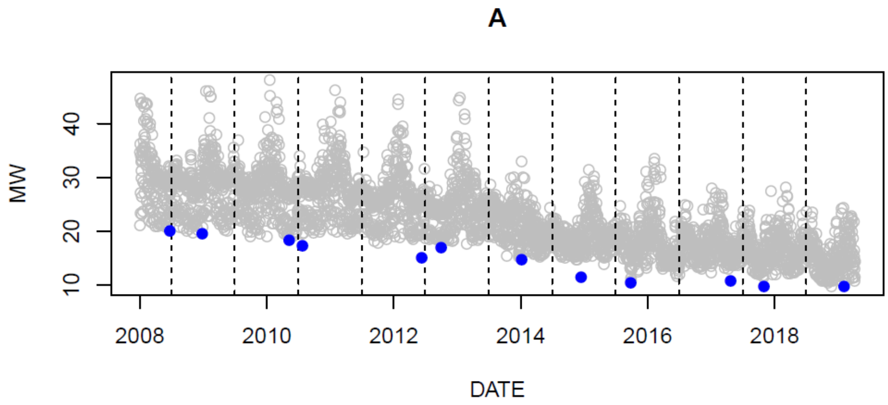
Outlier detection procedures are used to remove outliers in daily load caused by any measurement errors or outages.[19] The detection process for a given substation is based on a four-step process:

1. Calculate sample standard deviation of daily maximum (or minimum) load for each summer or winter
2. Calculate the distance between two consecutive observed daily maximum (or minimum) loads
3. Check the distance between two consecutive observed daily maximum loads to ensure the distance is not greater than three times the sample standard deviation for the summer or winter. Otherwise, the data is shortlisted for a further test in step 4. For the minimum demand time series, a similar test is applied based on a different distance threshold. If the distance between two consecutive daytime[20] or night-time minimum demand data points exceeds the standard deviation of the sample, shortlist the data for the test in step 4.
4. The resulting list of potential outliers are further assessed based on other data to determine the cause of the deviation. For example, if the deviation is caused by faults, major event days and outages, they are removed from the analysis. The data validated to be outliers is removed and does not affect the extreme demand modelling process.

Figure 5.2 shows a comparison of time series of day-time daily minimum demand before (upper panel) and after (bottom panel) the data cleaning process for Arkana. The day-time daily minimum demand is more realistic by removing the outliers. Similarly, as shown in Figure 5.3, by applying the above data cleaning process to substation Mandurah (MH), the time series of daily day-time minimum demand is obtained by removing outliers (e.g., blue dot in 2008).

[19] Outliers are observations in a data set that are substantially different from the bulk of the data. An outlier may be due to variability in the measurement or it may indicate experimental error; the latter may be excluded from the data set.

[20] The difference in standard deviation threshold is empirical, i.e. based on the differences in observed volatility and skewness in the minima and maxima data.



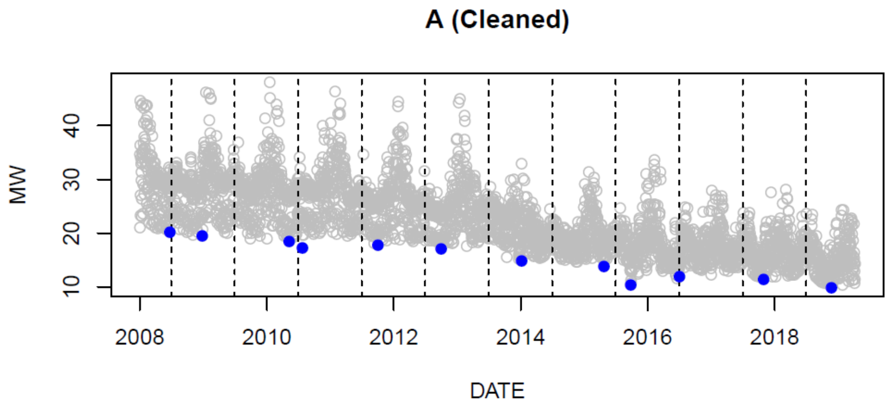


Figure 5.2 Data cleaning process for Arkana substation day-time minimum demand. The blue dots denote financial year minimum demand based on raw (top panel) and cleaned (bottom panel) data.

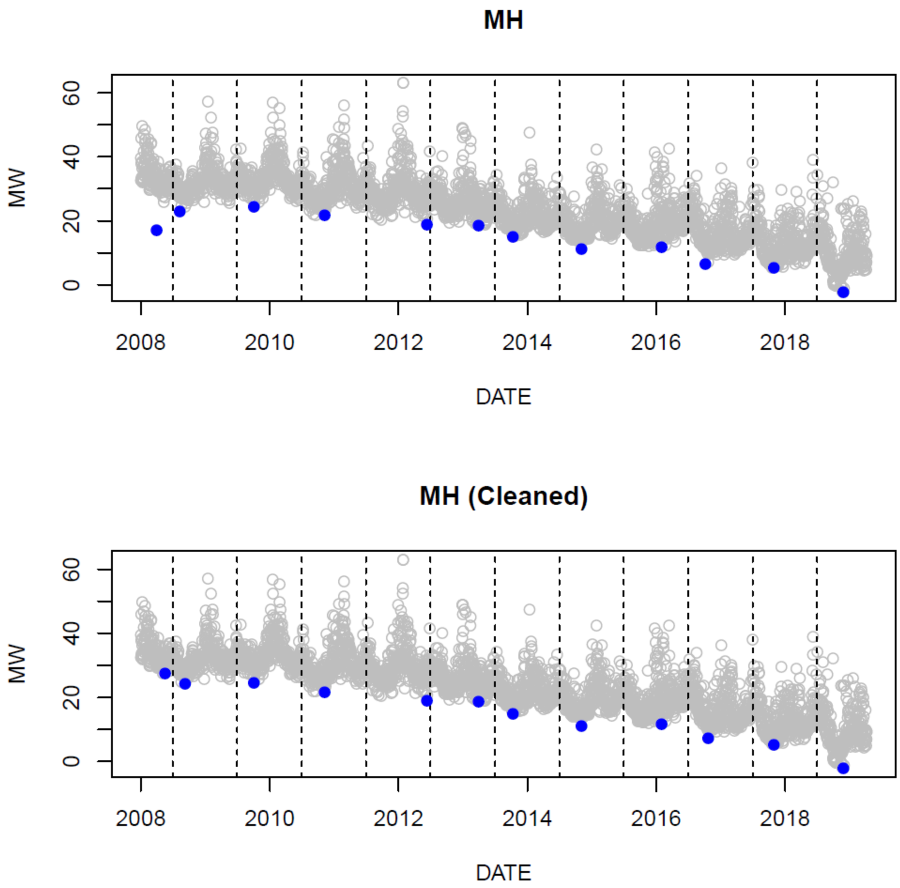


Figure 5.3 Data cleaning process for Mandurah (MH) substation day-time minimum demand. The blue dots denote financial year minimum demand based on raw (top panel) and cleaned (bottom panel) data.

5.3 Statistics of Extremes for peak demand forecast

The appropriate statistical tools for analysing extremes are well developed in statistics literature[21]. The popular approach is using the generalized extreme value (GEV) distribution to model block maxima (i.e., annual or seasonal maxima, or equivalent minima).

Generalized extreme value

Let $X_1, \dots, X_n$ be an independent and identically distributed sequence of n random variables from a distribution F, and the block maximum $M_n = max\{X_1, \dots, X_n\}$. Suppose there exists constants, $a_n > 0$ and b_n, such that,

$$\Pr\left\{\frac{M_n - a_n}{b_n} \le x\right\} = F(a_n + b_n x)^n \to G(x) \text{ as } n \to \infty$$

Then, the distribution of G is a generalised extreme value (GEV); that is,

$$G(x;\mu,\sigma,\xi) = \exp\left\{-\left[1+\xi\frac{x-\mu}{\sigma}\right]^{-1/\xi}\right\}$$

where $1 + \xi\frac{x-\mu}{\sigma} > 0$. The parameters μ, $\sigma > 0$ and ξ denote the location, scale, and shape parameters, respectively.

However, directly modelling the block (annual) maxima with the GEV distribution can be problematic when the data are only available for a few years. This is the case of the electricity peak demand forecast at Western Power, where there are only 12 years (2008-2020) historical data available.

5.3.1 Point process model

The point process model has several advantages over the block maxima directly modelled by a GEV distribution:

- Point process model uses more data (i.e., all data over a high threshold, as explained in detail below) resulting in more reliable results than those based on a direct fit of the GEV distribution to annual extreme demands.
- The point process model can be formulated as function of trends in customer count, average demand and solar PV installed capacity.

A point process models the occurrence of maximum (or minimum) demand as random events distributed with a two-dimensional Poisson (point) process. This process provides a stochastic rule for the occurrence and position of point events – in this context occurrence and magnitude of extremely large demand exceeding a high threshold. From the model, the probability of a certain number of events over a selected threshold within a specified period could be calculated. A two-dimensional process is used to describe the position, or the magnitude, of extreme demand events.

This two-dimensional point process is illustrated in the below figure. The process is illustrated over a time interval of duration T and all observations above threshold u are recorded. These points are annotated on

[21] S. Coles, *An Introduction to Statistical Modelling of Extreme Values*. London, U.K.: Springer, 2001.

a two-dimensional scatter plot. For a set A shown in Figure 5.4, the count of observations in the set A is assumed to be Poisson with an intensity parameter driven by an extreme value distribution:

$$\Lambda(A) = (t_2 - t_1)\left(1 + \xi \frac{y - \mu}{\sigma}\right)^{-1/\xi}$$

where μ is a location parameter, σ is a scale parameter and ξ is a shape parameter for a GEV distribution.

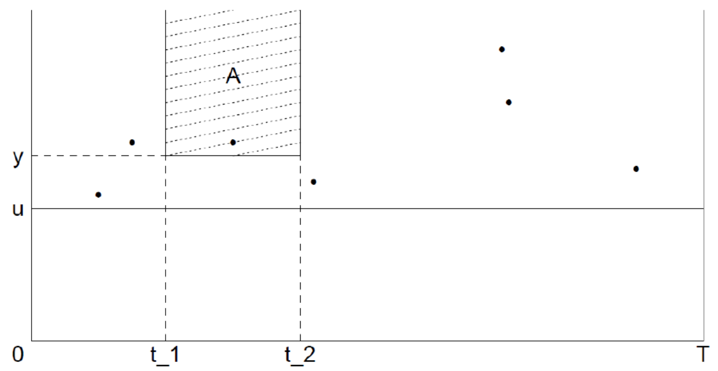


Figure 5.4. Diagram showing a two-dimensional point process for exceedances over a threshold u[22]

The intensity of the Poisson process is determined by three parameters (μ, σ, ξ), which can be demonstrated that they are equal to the location, scale and shape parameters of a GEV distribution.

In the developed model location and scale parameters are assumed to be dependent on customer connection counts, energy consumption and solar PV capacity trends, and the shape parameter ξ remains same. Table 3 lists 12 candidate point process models (A to L) used to fit to the substation data. Model A is a stationary model with no influence from other covariates. Models B to L are nonstationary models. Each nonstationary model consists of a GEV distribution with the location parameter assumed to be linearly influenced by trends in the customer count, Z_{1t} , energy consumption, Z_{2t}, and PV capacity, Z_{3t}. The scale parameter is affected by the PV capacity trend as an exponential form in models H to L to ensure the scale parameter is greater than zero.

[22] Smith R., 2013. Statistics of extremes, with applications in environment, insurance and finance, Department of Statistics, University of North Carolina Chapel Hill, NC 27599-3260, USA.

Table 3. Candidate point process models used to fit the substation data.

Model	Notation	Location	Scale	Shape	Covariate
A	mu0	$\mu(t)=\mu$	$\sigma(t)=\sigma$	$\xi(t)=\xi$	Stationary
B	mu1	$\mu(t)=\mu_0+\mu_1 Z_{1t}$	$\sigma(t)=\sigma$	$\xi(t)=\xi$	Z_{1t}
C	mu2	$\mu(t)=\mu_0+\mu_1 Z_{2t}$	$\sigma(t)=\sigma$	$\xi(t)=\xi$	Z_{2t}
D	mu12	$\mu(t)=\mu_0+\mu_1 Z_{1t}+\mu_2 Z_{2t}$	$\sigma(t)=\sigma$	$\xi(t)=\xi$	Z_{1t} & Z_{2t}
E	mu13	$\mu(t)=\mu_0+\mu_1 Z_{1t}+\mu_3 Z_{3t}$	$\sigma(t)=\sigma$	$\xi(t)=\xi$	Z_{1t} & Z_{3t}
F	mu23	$\mu(t)=\mu_0+\mu_2 Z_{2t}+\mu_3 Z_{3t}$	$\sigma(t)=\sigma$	$\xi(t)=\xi$	Z_{2t} & Z_{3t}
G	mu123	$\mu(t)=\mu_0+\mu_1 Z_{1t}+\mu_2 Z_{2t}+\mu_3 Z_{3t}$	$\sigma(t)=\sigma$	$\xi(t)=\xi$	Z_{1t} , Z_{2t} & Z_{3t}
H	mu2_sig3	$\mu(t)=\mu_0+\mu_2 Z_{2t}$	$\sigma(t)=exp(\sigma_0+\sigma_1 Z_{3t})$	$\xi(t)=\xi$	Z_{2t} & Z_{3t}
I	mu12_sig3	$\mu(t)=\mu_0+\mu_1 Z_{1t}+\mu_2 Z_{2t}$	$\sigma(t)=exp(\sigma_0+\sigma_1 Z_{3t})$	$\xi(t)=\xi$	Z_{1t} , Z_{2t} & Z_{3t}
J	mu13_sig3	$\mu(t)=\mu_0+\mu_1 Z_{1t}+\mu_3 Z_{3t}$	$\sigma(t)=exp(\sigma_0+\sigma_1 Z_{3t})$	$\xi(t)=\xi$	Z_{1t} & Z_{3t}
K	mu23_sig3	$\mu(t)=\mu_0+\mu_2 Z_{2t}+\mu_3 Z_{3t}$	$\sigma(t)=exp(\sigma_0+\sigma_1 Z_{3t})$	$\xi(t)=\xi$	Z_{2t} & Z_{3t}
L	mu123_sig3	$\mu(t)=\mu_0+\mu_1 Z_{1t}+\mu_2 Z_{2t}+\mu_3 Z_{3t}$	$\sigma(t)=exp(\sigma_0+\sigma_1 Z_{3t})$	$\xi(t)=\xi$	Z_{1t} , Z_{2t} & Z_{3t}

5.3.2 Model selection

The maximum likelihood estimation technique[23] is used to estimate GEV parameters in the point process models. Having all models in Table 3 fitted to data, the model selection is conducted based on the following processes:

i. selecting a nonstationary point process model with the minimum Akaike's information criteria (AIC) value from all candidate nonstationary models; and

ii. conducting the likelihood test to assess the significant difference between the selected nonstationary model against the stationary model A.

iii. checking if the selected model is a valid one. If not, select the model with immediate smaller AIC and conduct the likelihood test to assess the significant difference between the selected nonstationary model against the stationary model A.

23 S. Coles, *An introduction to statistical modelling of extreme values*. London, UK; Springer, 2001.

Valid point process model for maximum demand

- A valid model must have a positive and statistically significant coefficient for the customer connection and energy consumption, and a negative and statistically significant coefficient for PV capacity when they are fitted into the location parameters of the GEV distribution. This is because the more customer connection, the higher maximum demand; the maximum demand and energy consumption are moving together (a significant positive correlation between them which can be demonstrated by system total and substation data), and the PV generation has negative offset influence on the maximum demand.
- A valid model must have a positive and statistically significant coefficient for the PV capacity if it is chosen to fit into the scale parameter of the GEV distribution because it is expected the PV influence on the scale parameter causes the deviation increase of peak demand.

5.3.3 Forecast probability of exceedance

Forecasts of annual maximum demand in terms of a given probability of exceedance p follow the $(1-p)$th time varying quantiles of the non-stationary point process model (or 'effective' return level), which would reduce to a conventional return level (with return period $1/p$). The $(1-p)$th time varying quantiles can be given by:

$$PoE\left(p,t\right)=\mu(t)+\frac{\sigma(t)}{\xi(t)}\left\{\left[-\ln\left(1-p\right)\right]^{-\xi}-1\right\}$$

Hence, forecasts of PoE10(t), PoE50(t) and PoE90(t) at time t are given by the above equation with p =0.10, 0.50 and 0.90, respectively.

Western Power produces system maximum demand, coincident and noncoincident maximum demands as well as day-time minimum demand at all substations and selected feeders by the PoE10(t), PoE50(t) and PoE90(t) at time t up to 30 June 2025 based on the selected point process model.

5.3.4 Minimum demand forecast

Because the block minimum $\widetilde{M}_n$=min $\{X_1, X_2,..., X_n\} = -$max $\{-X_1, -X_2,..., -X_n\}$, it follows that GEV distribution can be also used to model the behaviour of the block minimum $\widetilde{M}_n$ (e.g. annual day-time minimum). That is, when the sequence $X_1, \dots, X_n$ are recorded as the sequence of daily minimum load, the methods for maximum demand forecasts can be applied to the negative sequence of random variables $-X_1, -X_2,..., -X_n$ in order to get a point process model from EVT to forecast the annual minimum demand.

5.3.5 Quantile regression

A similar model specification for maximum and minimum demand forecasts could be achieved with the use of quantile regression, with quantiles of a distribution being directly modelled as functions of covariates. Unlike linear regression model which specifies the change in the conditional mean of the dependent variable associated with a change in the covariates, the quantile regression model (QRM) specifies changes in the conditional quantile of the dependent variable associated with a change in the covariates. Quantile regression has recently been used for short-term load forecasts[24] .

This year, Western Power also developed quantile regression modelling (QRM) approach to medium term load forecast on the basis that QR is easy to implement and computationally fast, compared to the point process models from EVT.

A general linear QRM can be expressed as:

$$Q_x\left(q \mid Z_t\right) = \beta_0^{(q)} + \beta_1^{(q)} Z_{1t} + \cdots + \beta_k^{(q)} Z_{kt} + \varepsilon_t^{(q)} = \mathbf{Z}_t \boldsymbol{\beta}^{(q)} + \varepsilon_t^{(q)} \quad (1)$$

where $Q_x(q|.)$ is the conditional q-th quantile of the daily electric load distribution (X_t), $\mathbf{Z}_t = [1, Z_{1t}, \ldots, Z_{kt}]$ are regressors (covariates) , k is he number of covariates and $\boldsymbol{\beta}^{(q)} = \left\{\beta_0^{(q)}, \beta_1^{(q)}, \ldots, \beta_k^{(q)}\right\}$ is a vector of parameters for quantile q. The parameters are estimated by minimizing the loss function for a particular q-th quantile:

$$\underset{\boldsymbol{\beta}^{(q)}}{\arg\min} \left[\sum_{\{t: x_t >= Z_t \beta^{(q)}\}} q \left| y_t - \mathbf{Z}_t \boldsymbol{\beta}^{(q)} \right| + \sum_{\{t: x_t < Z_t \beta^{(q)}\}} (1-q) \left| y_t - \mathbf{Z}_t \boldsymbol{\beta}^{(q)} \right| \right] \quad (2)$$

where x_t is the actual load of X_t, $t = 1, 2, \ldots$ denote the time day in the study period.

Forecasts of annual maximum (or minimum) demand in terms of a given probability of exceedance p, the $q = (1 - p)$th time varying quantiles of the valid quantile regression model can be given by

$$PoE(q,t) = \hat{\beta}_0^{(q)} + \hat{\beta}_1^{(q)} Z_{1t} + \cdots + \hat{\beta}_k^{(q)} Z_{kt} \quad (3)$$

where $\hat{\beta}_0^{(q)}$, $\hat{\beta}_1^{(q)}$,..., $\hat{\beta}_k^{(q)}$ are estimated parameters.

The application of QRM to maximum and minimum demand forecasts

With the formulation (1), we can link q-th quantile of daily electricity demand to three (i.e., $k = 3$) drivers of trends in the customer count Z_{1t}, energy consumption Z_{2t} and PV capacity Z_{3t}. All seven QRM models with various covariate combinations are listed in Table 4.

For a given quantile q, the model selection from those in Table 4 is based on the following process:

i. selecting a QRM with the minimum Akaike's information criteria (AIC) value from all candidate quantile regression models;

ii. checking if the selected model is a valid model, which must have a positive and statistically significant coefficient for the customer connection and average demand if these terms are in the model, and a negative and statistically significant coefficient for PV capacity if that term is in the model. If not, select the model with immediate smaller AIC.

[24] Y. Wang, N. Zhang, Y. Tan, T. Hong, D. D. Krischen, and C. Kang, "Combing probability load forecasts," *IEEE Trans. Smart Grid*, vol. 10, pp. 3664–3674, 2019.

After the choice of the valid QRM, it can be used to estimate the quantile for a given probability of exceedance p by (3).

Since the quantile regression models in Table 4 are fitted to daily load data, estimated quantiles by (3) with respective p =0.10, 0.50 and 0.90 are the 0.90-th, 0.50-th and 0.10-th quantile of daily load, not the quantiles (POE10, POE50 and POE90) for annual maximum or minimum demand. As such, the POE forecasts of summer, winter maximum and minimum demand based on the valid QRM are obtained following the process:

1. Find summer maximum demand in the historical period from year 2008 to 2020
2. Calculate quantiles $q_{0.10}, q_{0.50}, q_{0.90}$ of summer maxima at the 0.90, 0.50 and 0.10 levels
3. Forecasts of PoE10(t), PoE50(t) and POE90(t) of annual maximum/minimum demand are obtained by using (3) with q equals to quantiles $q_{0.90}, q_{0.50}, q_{0.10}$, respectively. That is,

$$PoE10(t) = \hat{\beta}_0^{(q_{0.90})} + \hat{\beta}_1^{(q_{0.90})} Z_{1t} + \cdots + \hat{\beta}_k^{(q_{0.90})} Z_{kt} \quad (4)$$

$$PoE50(t) = \hat{\beta}_0^{(q_{0.50})} + \hat{\beta}_1^{(q_{0.50})} Z_{1t} + \cdots + \hat{\beta}_k^{(q_{0.50})} Z_{kt} \quad (5)$$

$$PoE90(t) = \hat{\beta}_0^{(q_{0.10})} + \hat{\beta}_1^{(q_{0.10})} Z_{1t} + \cdots + \hat{\beta}_k^{(q_{0.10})} Z_{kt} \quad (6)$$

4. Adjust $q_{0.10}, q_{0.50}, q_{0.90}$ slightly and refit the valid QRM to make the estimated POE values to more realistic in case it is needed.

Table 4. Candidate quantile regression models used to fit the substation data

Model	Notation	Model specification	Covariate
I	M1	$Q_y(q\|Z_{1t}, Z_{2t}, Z_{3t}) = \beta_0^{(q)} + \beta_1^{(q)} Z_{1t}$	Z_{1t}
II	M2	$Q_y(q\|Z_{1t}, Z_{2t}, Z_{3t}) = \beta_0^{(q)} + \beta_2^{(q)} Z_{2t}$	Z_{2t}
III	M3	$Q_y(q\|Z_{1t}, Z_{2t}, Z_{3t}) = \beta_0^{(q)} + \beta_3^{(q)} Z_{3t}$	Z_{3t}
IV	M12	$Q_y(q\|Z_{1t}, Z_{2t}, Z_{3t}) = \beta_0^{(q)} + \beta_1^{(q)} Z_{1t} + \beta_2^{(q)} Z_{2t}$	Z_{1t} & Z_{2t}
V	M23	$Q_y(q\|Z_{1t}, Z_{2t}, Z_{3t}) = \beta_0^{(q)} + \beta_2^{(q)} Z_{2t} + \beta_3^{(q)} Z_{3t}$	Z_{2t} & Z_{3t}
VI	M13	$Q_y(q\|Z_{1t}, Z_{2t}, Z_{3t}) = \beta_0^{(q)} + \beta_1^{(q)} Z_{1t} + \beta_3^{(q)} Z_{3t}$	Z_{1t} & Z_{3t}
VII	M123	$Q_y(q\|Z_{1t}, Z_{2t}, Z_{3t}) = \beta_0^{(q)} + \beta_1^{(q)} Z_{1t} + \beta_2^{(q)} Z_{2t} + \beta_3^{(q)} Z_{3t}$	Z_{1t}, Z_{2t} & Z_{3t}

5.4 System total maximum and minimum demand forecast

This section provides the results of system total maximum and minimum demand forecasts using EVT and QRM models (Section 5.3) based on three drivers of trends in the customer count Z_{1t}, average demand Z_{2t} and PV capacity Z_{3t} (Figure 5.5). The aim is to provide the evidence that both EVT and QRM can provide realistic maximum and minimum demands, and good interpretation of how customer connection, energy consumption and PV capacity impact system total maximum and minimum demands.

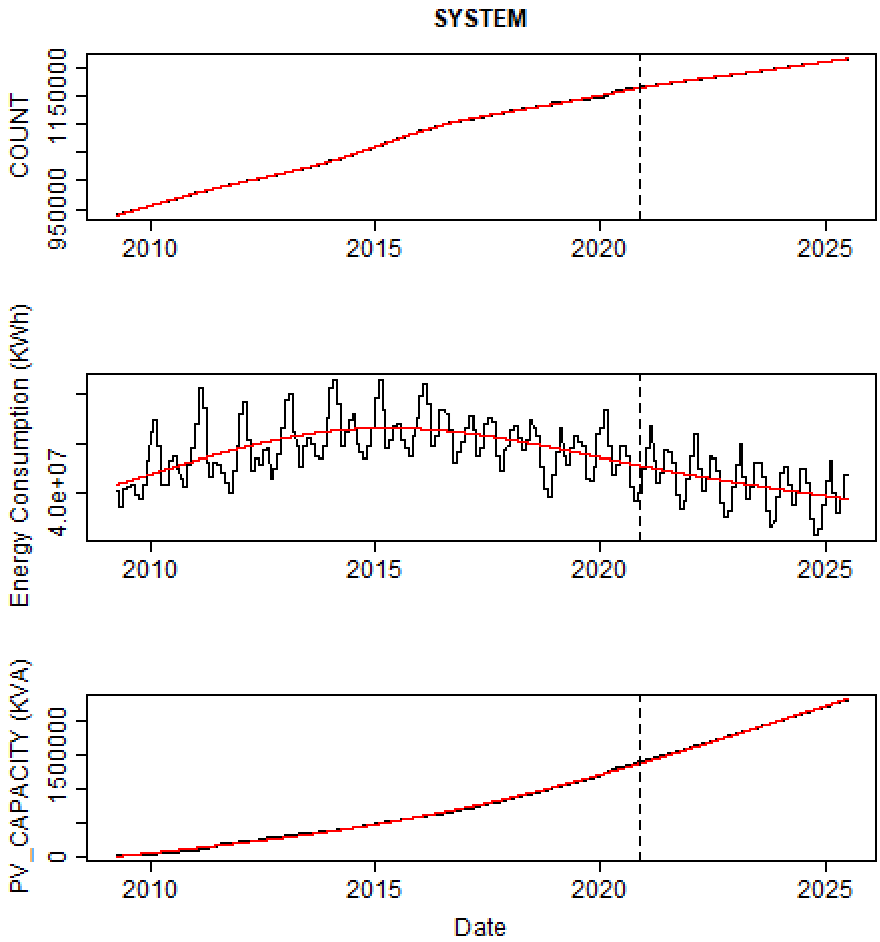


Figure 5.5 Daily time series of observed (black) and trends (red) in customer count (top panel), energy consumption (middle panel) and PV capacity (bottom panel) for the system. The vertical dash line is location of date on 30 June 2020.

Key messages from three drivers:

- Customer connection trend has been seen to increase in historical period (1 January 2008 to 30 June 2020) and it has been forecasted to be increasing up to 1.215 million by 20 June 2025
- Energy consumption trend has increased since 2009 until 2016 and it has been decreasing since then and is forecasted to keep decreasing to 39439201 (kWh) by 30 June 2020.
- PV capacity trend has been seen to increase in historical period and is forecasted to keep increasing to 2324928kVA by 30 June 2025.

5.4.1 System summer maximum demand forecast

System summer (November to April) system maximum demand forecasts in terms of PoE10, PoE50 and PoE90 at day t, based on the fitted EVT model, are

$$\mathrm{POE}(t,p)=3937+6.55\times10^{-7}Z_{2t}-\frac{\exp(4.95+2.74\times10^{-7}Z_{3t})}{0.384}\left\{\left[-\ln\left(1-p\right)\right]^{0.384}-1\right\}$$

with $p = 0.90, 0.50$ and 0.10, respectively. Figure 5.6 shows the corresponding estimated PoE10, PoE50 and PoE90 curves by using daily values of the consumption trend Z_{2t} and the PV capacity trend Z_{3t} (Figure 5.5) from the observed period from 1 November 2009 to 30 April 2020 and the forecasting period from 1 November 2020 to 30 April 2025. The observed summer maximum demand is spread evenly over the POE50 level. Maximum demand in 2016 hot summer is above the POE10 level and below the POE90 level in 2019 summer, implying a skill for probabilistic summer maximum demand based on the preferred model.

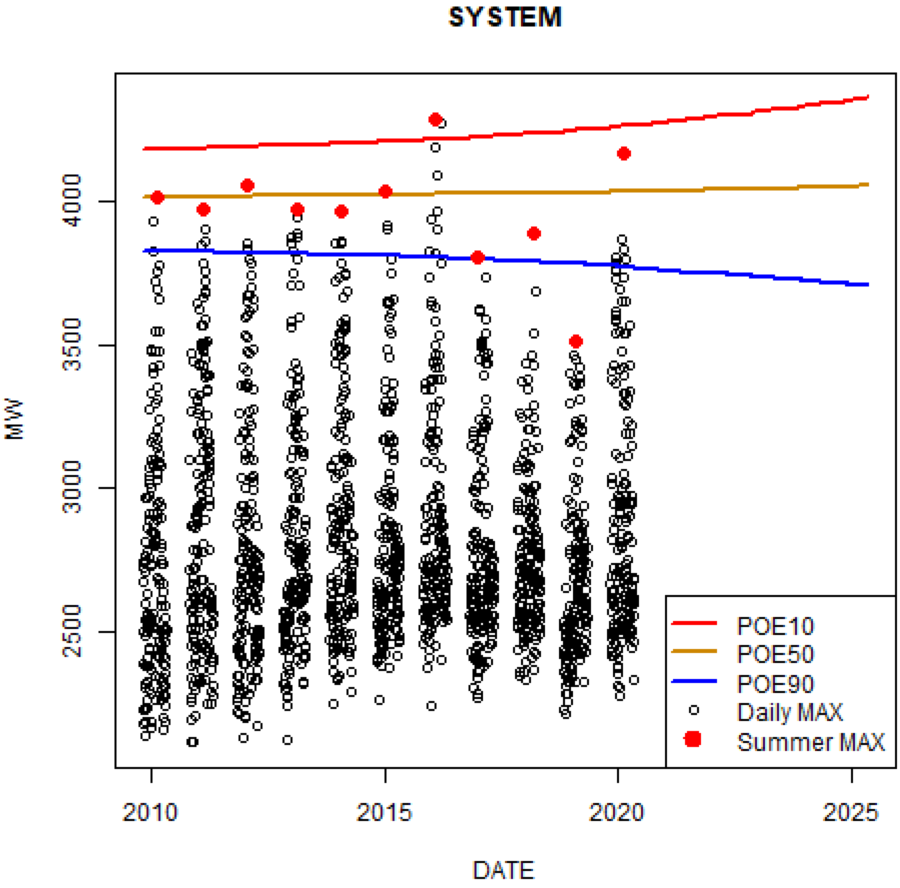


Figure 5.6 System summer maximum demand forecast via EVT model.

The above preferred model also provides an interoperation of how the changes in the energy consumption trend Z_{2t} and the PV capacity trend Z_{3t} influences the system summer maximum demand. For instance, for a give probability p ($= 0.90, 0.50$ and 0.10) the POEs forecasts have positive linear relationship Z_{2t} and positive relationship in exponential with the PV capacity trend Z_{3t}. As such, the slightly increasing POE50 forecast with large uncertainty characterised by increasing POE10 and decreasing POE90 forecasts is due to the combined influence of the decreasing energy consumption trend and increasing PV capacity in summer over the forecasted period from 1 November 2020 to 30 April 2025.

5.4.2 System winter maximum demand forecast

System winter (May to October) maximum demand forecasts in terms of PoE10, PoE50 and PoE90 at day t, based on the fitted QRM model, are:

$$\text{PoE90}(t) = 428 + 0.00123 Z_{1t} + 3.53 \times 10^{-5} Z_{2t}$$

$$\text{PoE50}(t) = 306 + 0.00153 Z_{1t} + 3.29 \times 10^{-5} Z_{2t}$$

$$\text{PoE10}(t) = 416 + 0.00241 Z_{1t} + 1.23 \times 10^{-5} Z_{2t}$$

Figure 5.7 shows the corresponding estimated PoE10, PoE50 and PoE90 curves by using daily values of the customer connection trend Z_{1t} and consumption trend Z_{2t} (Figure 5.5).

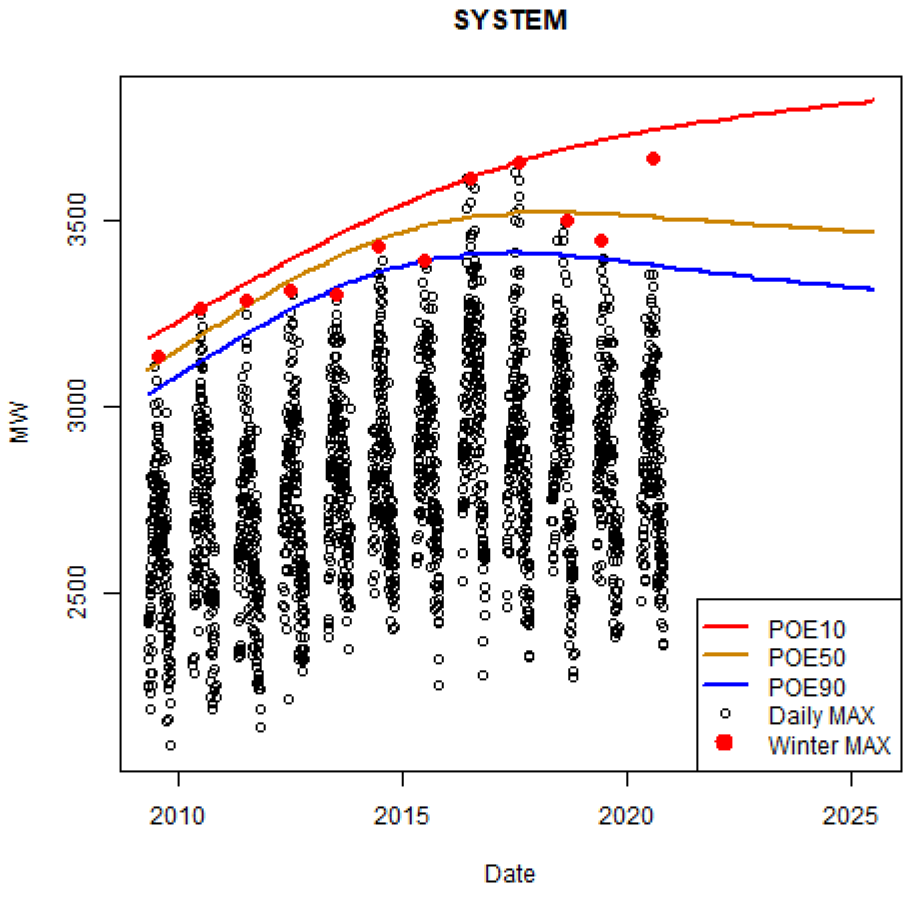


Figure 5.7 System winter maximum demand forecast via quantile regression models.

The observed winter maximum demand is spread evenly over the POE50 level. Three winter maximum demands in 2017, 2018 and 2020 fall between the forecasted POE50 and POE10 levels as expected, implying a skill for probabilistic summer maximum demand based on the preferred model. The increasing POE10 forecast and decreasing POE50 and POE90 forecasts are due to the combined influence of the increasing trend in the customer connection trend Z_{1t} and decreasing consumption trend Z_{2t} through the preferred QRM. Note that the PV capacity trend Z_{3t} is not included in the selected model for system winter maximum demand, implying that the impacts of PVs are not significantly in winter.

5.4.3 System daytime minimum demand forecast

System daytime (6am to 6pm) annual minimum demand forecasts in terms of PoE10, PoE50 and PoE90 at day t, based on the fitted EVT model, are

$$\text{POE}(t,p) = -1.31 + 6.56\times10^{-5} Z_{2t} - 1.92\times10^{-4} Z_{3t} - \frac{\exp(2.33+5.57\times10^{-7} Z_{3t})}{0.148}\left\{\left[-\ln(1-p)\right]^{0.148} - 1\right\}$$

with $p = 0.90, 0.50$ and 0.10, respectively.

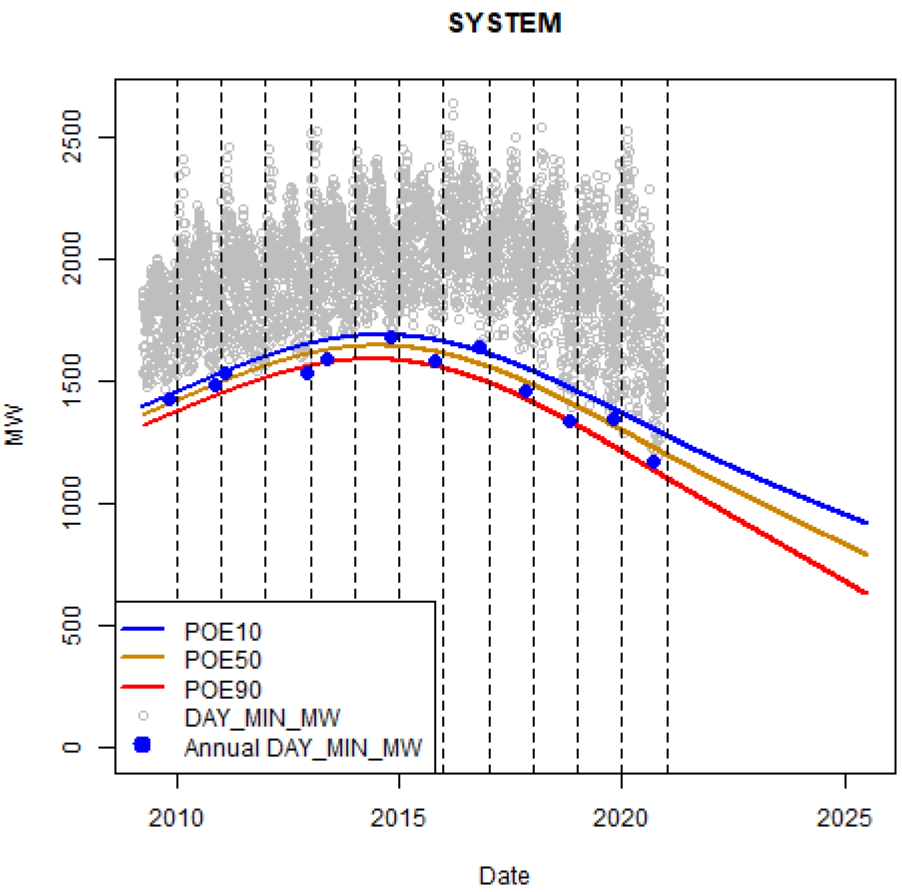


Figure 5.8 System daytime minimum demand forecasts based on the fitted EVT model. Vertical lines are

Figure 5.8 shows the corresponding estimated PoE10, PoE50 and PoE90 curves by using daily values of the consumption trend Z_{2t} and the PV capacity trend Z_{3t} (Figure 5.5) from the observed period from 1 January 2009 to 30 June 2020 and the forecasting period from 1 July 2020 to 30 June 2025. The observed daytime minimum demand spreads evenly over the POE50 level and falls within forecasted POE90 and POE10 levels, implying a skill for probabilistic summer maximum demand based on the preferred model.

The above preferred model also provides an interoperation of how the changes in the energy consumption trend Z_{2t} and the PV capacity trend Z_{3t} influences the system daytime maximum demand. For instance, for a give probability p ($= 0.90, 0.50$ and 0.10) the POEs forecasts have positive linear relationship with energy consumption trend Z_{2t}, the forecasted POE levels decrease 6.56×10^{-5} MW per kWh. The influence of the PV capacity trend Z_{3t} can be interpreted as two components: the linearly negative impact with the POEs declined 1.92×10^{-4} MW per kVA increase in Z_{3t} and the exponentially positive influence with 5.57×10^{-7} MW per kVA increase in Z_{3t}. As such, the forecasted PoE levels are declined with large uncertainty characterised by expending difference in POE10 and POE90 forecasts because of the combined influence of the decreasing energy consumption trend and increasing PV capacity in summer over the forecasted period from 1 January 2020 to 30 June 2025. The forecasted POE50=705MW with POE90=636MW and POE10=922 MW by 30 June 2025.

5.5 Substation maximum and minimum demand forecasts

Zone substation maximum demand contains non-coincident maximum and coincident demand measures by summer and winter maximum demand. Zone substation non-coincident maximum demand measures the season maximum demand at a zone substation. Zone substation coincident demand refers to demand at each zone substation at the time of the occurrence of whole system maximum demand.[25] Typically, non-coincident and coincident maximum demands occur at different times and may occur on different days.

Zone substation minimum demand refers to the annual daytime (6am to 6pm) minimum demand at a zone substation.

Zone substation maximum and minimum demand forecasts are developed by EVT and QRM models (Section 6.3) fitted to daily data based on three drivers of trends in the customer count Z_{1t}, average demand Z_{2t} and PV capacity Z_{3t}.

In the following sections, we demonstrate that both EVT and QRM can provide realistic substation maximum and minimum demands, and good interpretation of how customer connection, energy consumption and PV capacity through a case study of maximum and minimum demand forecasts at substation Arkana (A).

Figure 5.9 shows time series of observed (black) and fitted trends (red) for customer count, panel (a), energy consumption, panel (b), and solar PV capacity, panel (c), for Arkana substation. It is evident that the fitted trend (red carves) has captured the step changes in customer counter, energy consumption and PV capacity due to the customer transfers on 1 December 2013 and 1 August 2015.

25 Western Power measures whole system demand based on total supply from market generators at transmission and distribution network plus the remaining power supply from the large customers' embedded generators after supplying their own load.

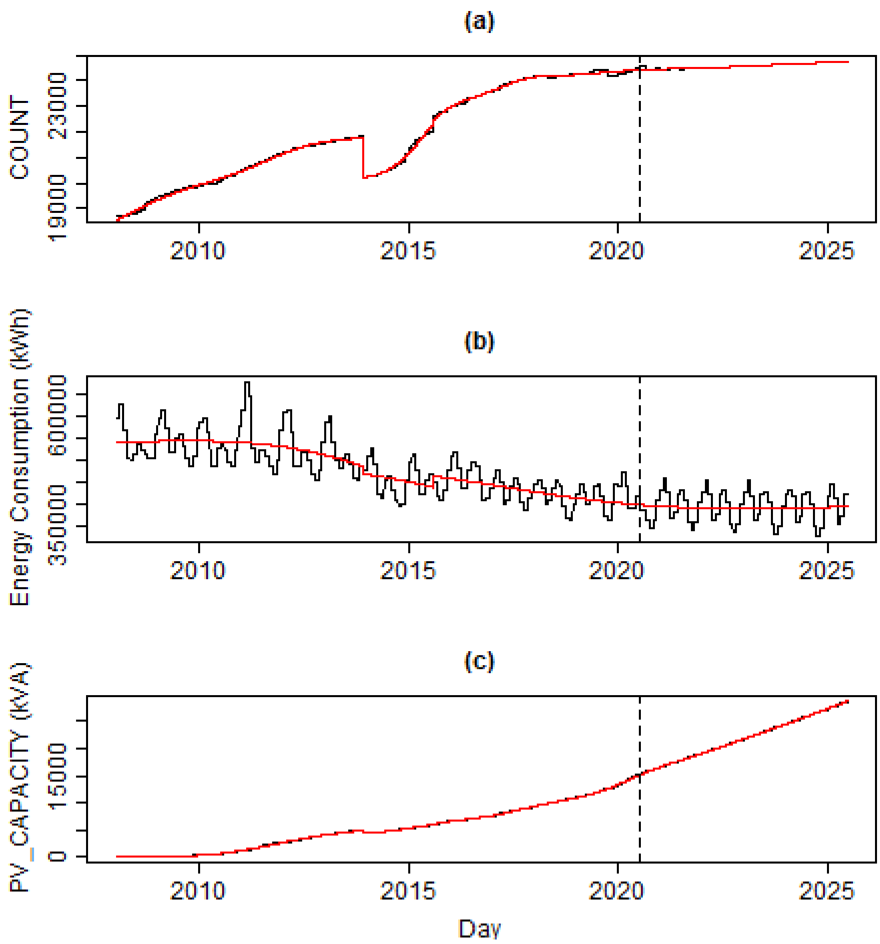


Figure 5.9 Time series of observed (black) and fitted trends (red) for customer count, panel (a), energy consumption, panel (b), and solar PV capacity, panel (c), for Arkana substation.[26]

5.5.1 Non-coincident maximum demand forecast

Arkana summer non-coincident maximum demand forecasts in terms of PoE10, PoE50 and PoE90 at day t, based on the fitted EVT model, are

$$\mathrm{POE}(t,p)=1.69+1.07\times10^{-4}Z_{2t}-\frac{\exp(-0.57+7.40\times10^{-5}Z_{3t})}{0.048}\left\{\left[-\ln(1-p)\right]^{0.048}-1\right\}$$

with $p = 0.90, 0.50$ and 0.10, respectively.

Figure 5.10 shows the corresponding estimated PoE10, PoE50 and PoE90 curves by using daily values of the consumption trend Z_{2t} and the PV capacity trend Z_{3t} (Figure 5.9) from the observed period from 1 November 2009 to 30 April 2020 and the forecasting period from 1 November 2020 to 30 April 2025. A striking feature is that the fitted POE10, POE50, and POE90 values all show step changes corresponding to those in both Z_{2t} and Z_{3t} on 1 December 2013 and 1 August 2015. The observed summer maximum demand is spread evenly over the POE50 level, implying a skill for probabilistic summer maximum demand based on the preferred model.

[26] The vertical dashed line is the end of period for observed data on 30 June 2020.

The preferred model also provides an interoperation of how the changes in the energy consumption trend Z_{2t} and the PV capacity trend Z_{3t} influences the Arkana summer maximum demand. For instance, for a give probability p $(= 0.90, 0.50$ and $0.10)$ the POEs forecasts have positive linear relationship Z_{2t} and positive relationship in exponential with the PV capacity trend Z_{3t}. As such, the slightly stabilizing POE50 forecast with uncertainty characterised by increasing POE10 and decreasing POE90 forecasts is due to the combined influence of the decreasing energy consumption trend and increasing PV capacity in summer over the forecasted period from 1 November 2020 to 30 April 2025.

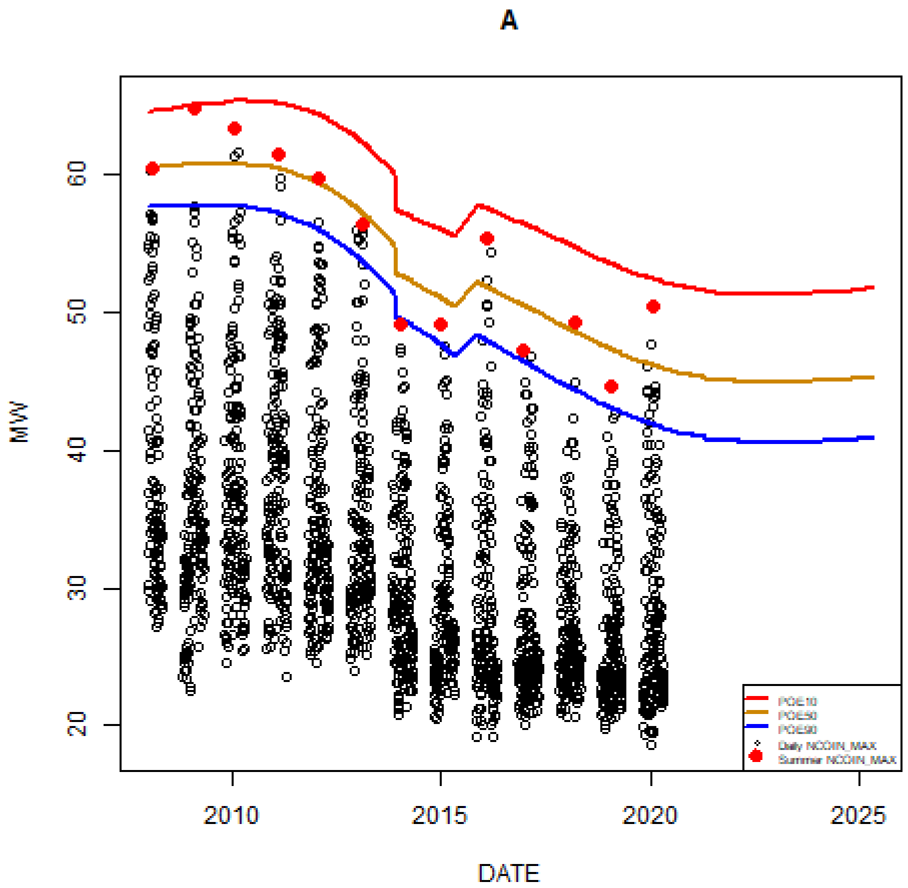


Figure 5.10 Observed daily maximum demand (back dots) and forecasted PoE10 (red), PoE50 (orange) and PoE90 (blue) for summer maximum demand at Arkana (A) substation by EVT model.

Arkana winter non-coincident maximum demand forecasts in terms of PoE10, PoE50 and PoE90 at day t, based on the fitted EVT model, are

$$\mathrm{POE}(t, p) = 17.72 + 4.82 \times 10^{-5} Z_{2t} - 8.33\left\{\left[-\ln(1-p)\right]^{0.185} - 1\right\}$$

with $p = 0.90, 0.50$ and 0.10, respectively.

As shown in Figure 5.11, the fitted POE10, POE50, and POE90 values all show step changes corresponding to those in energy consumption trend Z_{2t} on 1 December 2013 and 1 August 2015. The observed winter maximum demand is spread evenly over the POE50 level and with one winter maximum demand above the

POE10 in 2020 and one below POE90 in 2014, implying a skill for probabilistic winter maximum demand based on the preferred model. Note that the PV capacity trend Z_{3t} is not included in the selected model for Arkana winter maximum demand, implying that the impacts of PVs are not significantly in winter at substation Arkana.

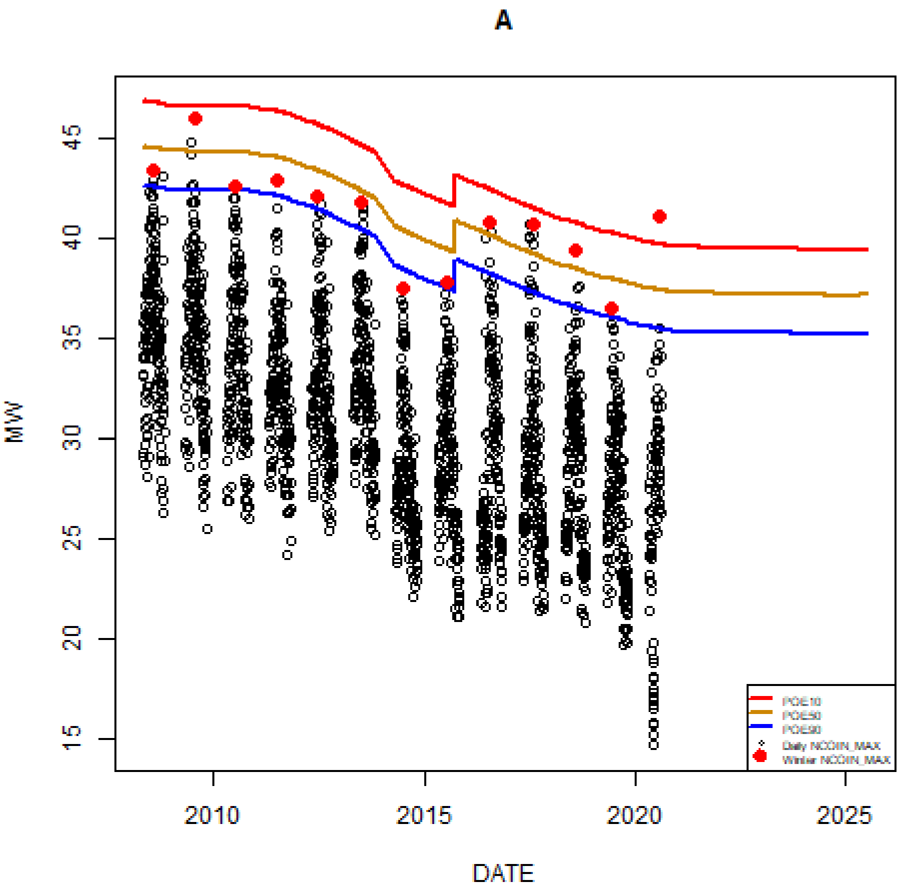


Figure 5.11 Observed daily maximum demand (back dots) and forecasted PoE10 (red), PoE50 (orange) and PoE90 (blue) for winter maximum demand at Arkana (A) substation by EVT model.

5.5.2 Coincident maximum demand forecast

Substation coincident incident maximum demand forecasts are developed by season using the QRM approach as described in section 6.3.5.

Substation Arkana winter coincident maximum demand forecasts in terms of PoE10, PoE50 and PoE90 at day t, based on the fitted QRM model, are:

$$\mathrm{PoE90}(t) = -23.95 + 0.00075 Z_{1t} + 0.000131 Z_{2t}$$

$$\mathrm{PoE50}(t) = -47.43 + 0.00146 Z_{1t} + 0.000138 Z_{2t}$$

$$\mathrm{PoE10}(t) = -68.08 + 0.00205 Z_{1t} + 0.000164 Z_{2t}$$

Figure 5.12 shows the corresponding estimated PoE10, PoE50 and PoE90 curves by using daily values of the customer count trend Z_{1t} and consumption trend Z_{2t} (Figure 5.9) from the observed period from 1 November 2009 to 30 April 2020 and the forecasting period from 1 November 2020 to 30 April 2025. A striking feature is that the fitted POE10, POE50, and POE90 values all show step changes corresponding to those in both Z_{1t} and Z_{2t} on 1 December 2013 and 1 August 2015. The observed summer coincident maximum demand is spread evenly over the POE50 level, and the uncertainty of winter coincident maximum is also corrected captured by the fitted POE10 and POE90 levels, implying a skill for probabilistic summer maximum demand based on the selected QRM model.

The stabilized POE10, POE50 and POE90 forecasts are due to the combined influence of the increasing trend in the customer connection trend Z_{1t} and decreasing consumption trend Z_{2t} through the preferred QRM. Note that the PV capacity trend Z_{3t} is not included in the selected QRM, implying that the impacts of PVs are not significantly on summer coincident maximum demand at substation Arkana.

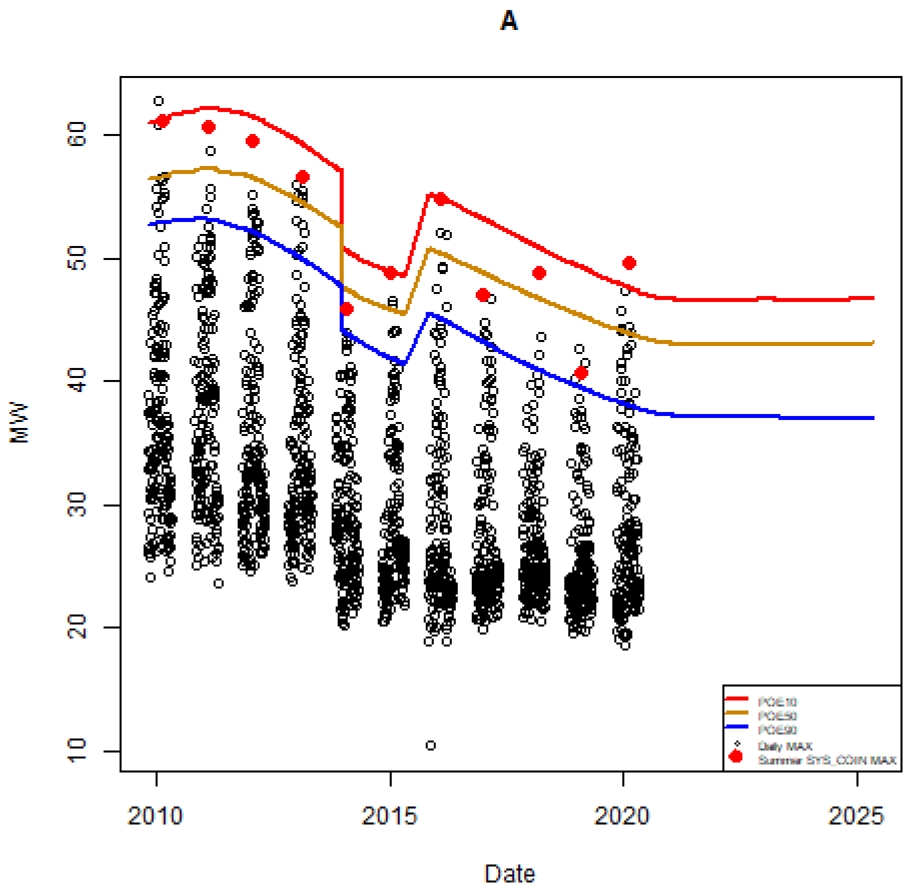


Figure 5.12 Observed summer coincident daily maximum demand (back dots) and forecasted PoE10 (red), PoE50 (orange) and PoE90 (blue) for summer coincident maximum demand at Arkana (A) substation by quantile regression model.

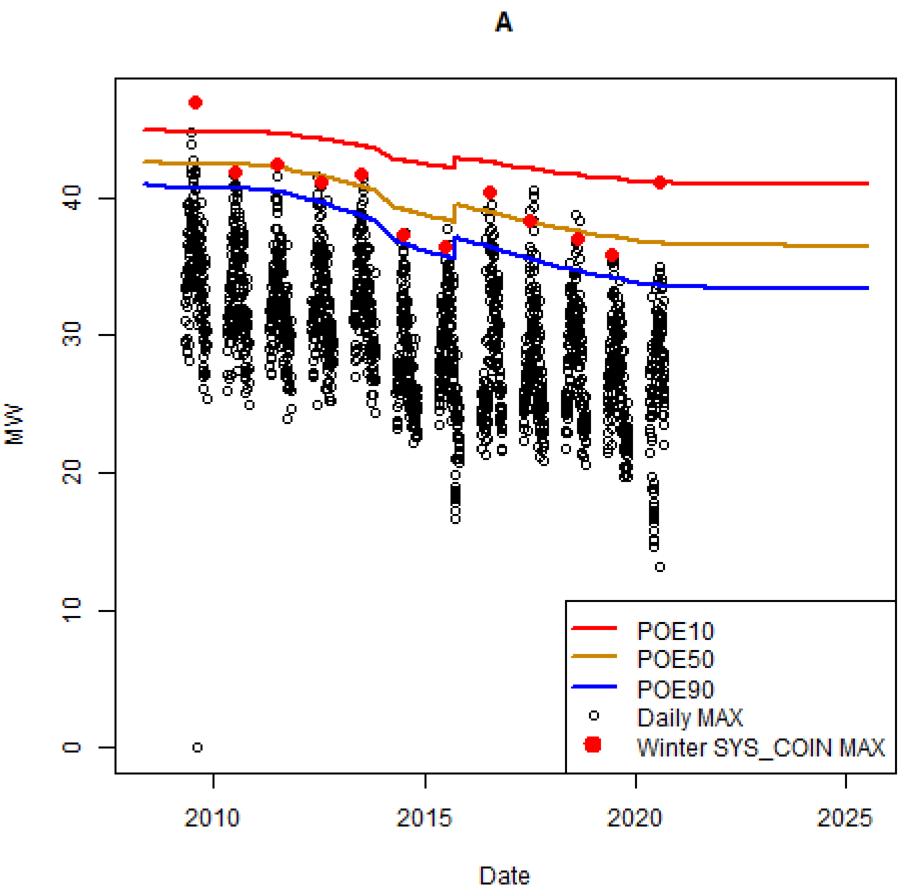


Figure 5.13 Observed winter coincident daily maximum demand (back dots) and forecasted PoE10 (red), PoE50 (orange) and PoE90 (blue) for winter coincident maximum demand at Arkana (A) substation by quantile regression model.

Arkana winter coincident maximum demand forecasts in terms of PoE10, PoE50 and PoE90 at day t, based on the fitted QRM model, are

$$\text{PoE90}(t) = 14.10 + 4.92 \times 10^{-5} Z_{2t}$$

$$\text{PoE50}(t) = 20.98 + 3.96 \times 10^{-5} Z_{2t}$$

$$\text{PoE10}(t) = 30.87 + 2.57 \times 10^{-5} Z_{2t}$$

As shown in Figure 5.13, Figure 5.11, the fitted POE10, POE50, and POE90 values all show step changes corresponding to those in energy consumption trend Z_{2t} on 1 December 2013 and 1 August 2015. The observed winter coincident maximum demand is spread evenly over the POE50 level and the is also corrected captured by the fitted POE10 and POE90 levels, implying a skill for probabilistic winter maximum demand based on the selected QRM model. Note that the PV capacity trend Z_{3t} is not included in the selected QRM model for Arkana winter coincident maximum demand, implying that the impacts of PVs are not significantly on winter coincident maximum demand at substation Arkana.

5.5.3 Day-time minimum demand forecast

Substation Arkana daytime (6am to 6pm) annual minimum demand forecasts in terms of PoE10, PoE50 and PoE90 at day t, based on the fitted EVT model, are

$$\mathrm{POE}(t,p) = -4.16 + 2.80\times10^{-5} Z_{2t} - 6.09\times10^{-4} Z_{3t} + 2.68\left\{\left[-\ln\left(1-p\right)\right]^{0.2} - 1\right\}$$

with $p = 0.90, 0.50$ and 0.10, respectively.

Figure 5.14 shows the corresponding estimated PoE10, PoE50 and PoE90 curves by using daily values of the consumption trend Z_{2t} and the PV capacity trend Z_{3t} (Figure 5.5) from the observed period from 1 January 2009 to 30 June 2020 and the forecasting period from 1 July 2020 to 30 June 2025. The observed daytime minimum demand spreads evenly over the POE50 level with uncertainty captured by the fitted POE10 and POE90 levels, implying a skill for probabilistic daytime maximum demand based on the preferred EVT model.

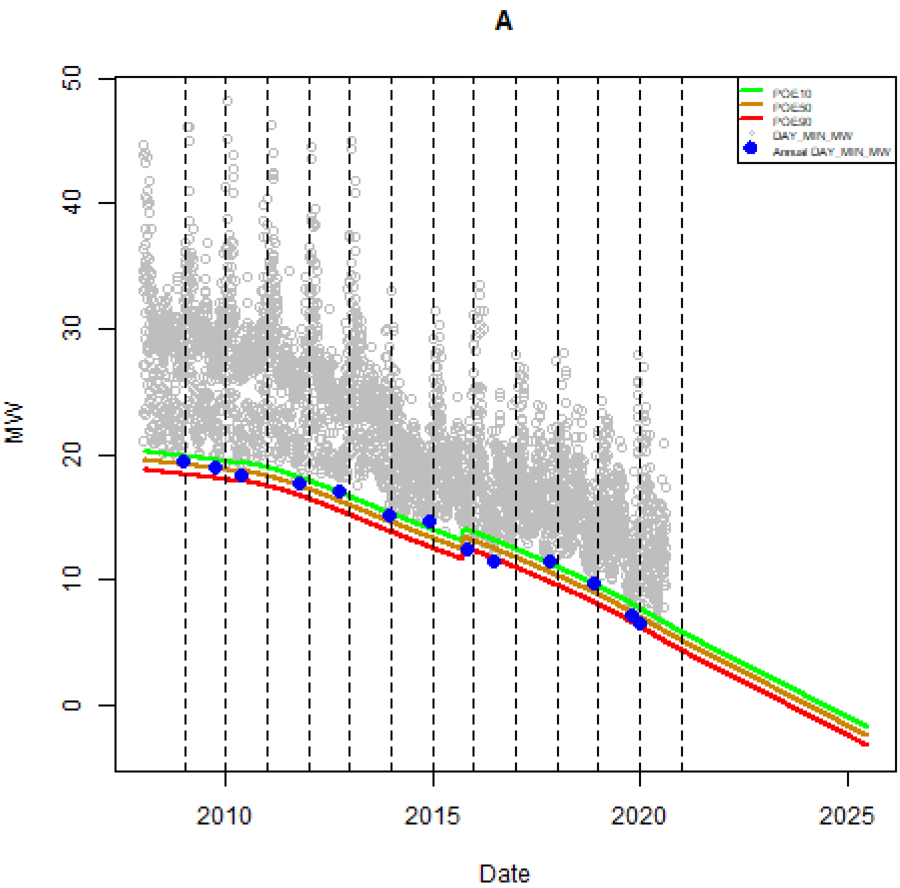


Figure 5.14 Observed daily day-time minimum demand (grey dots), annual day-time minimum demand (blue dots) and forecasted PoE10 (green), PoE50 (orange) and PoE90 (green) for winter maximum demand at Arkana (A) substation by EVT model. Vertical dash lines are locations of date 31 December each year.

The forecasted PoE50 is declined with small uncertainty characterised by the small difference in POE10 and POE90 forecasts because of the combined influence of the decreasing energy consumption trend and increasing PV capacity trend. As shown in Figure 5.14, Arkana daytime minimum demand has been declined in historical period and it is forecasted to below 0 MW by 30 June 2025.

5.5.4 Transmission substation maximum and minimum demand forecasts

A substation on the transmission network is typically customer owned. Customers connected directly to a transmission substation are mainly large industrial types with usage profiles that are commonly stationary over time. For these substations, extreme demand forecasts are expected to be relatively stationary, or changing with the average demand only. This implies the EVT models fitted to a given transmission substation are expected to be either stationary (of the type mu0) or dependent on the trend in average demand only, which drives the location parameter of the model.

Figure 5.15 shows coincident and noncoincident maximum demand forecasts in summer and winter for the Australian Fused Materials (AFM) substation. As shown in the figure, summer and winter coincident maximum demands (red dots) are not the same as substation maximum in observed years (top panels). The AFM coincident maximum demand in both summer and winter is forecasted to be stable during the five-year forecast period with large difference between PoE10 and PoE90 bands.

The location parameter of the model is dependent on the trend in average demand, implying that the summer forecast is driven by a flat trend in average demand. The observed that shows that the daily coincident maximum demand at AFM approached to zero MW when summer demand on the transmission system reached to the maximum in summers 2010, 2011 and 2012 (or winters 2008, 2010, 2011, 2012, 2014 and 2015).

The AFM non-coincident maximum demand in both summer and winter is forecasted to be stationary in the five-year forecasting period and with a large difference (~2MW) between PoE10 and PoE90 bands. The EVT model is parameterised as a stationary GEV, which implies that PoEs of the AMF noncoincident maximum in summer and winter do not change during the five-year forecast horizon.

Figure 5.16 shows day-time minimum demand forecasts at the AFM substation. The PoE10, PoE50 and PoE90 of the day-time minimum demand at AFM are forecasted to be stationary in the five-year forecasting period. As the EVT model is parameterised as a stationary GEV, the day-time minimum demand is forecast to be stationary. The small uncertainty in terms of small difference between PoE10 and PoE90 is consistent with the well-known fact that the minimum demand has small variability for a given substation.

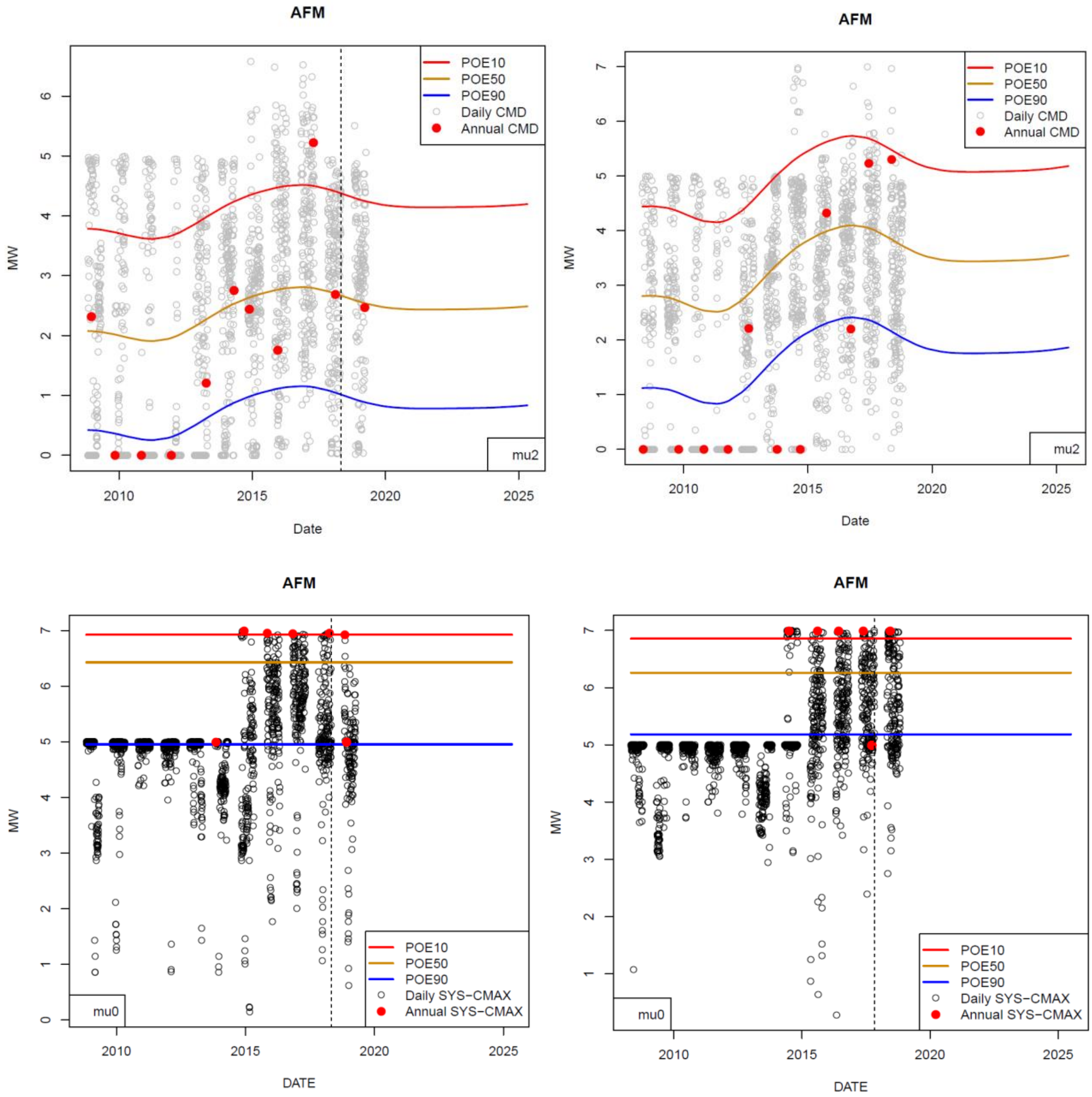


Figure 5.15 Observed daily coincident maximum demand (grey dots) and forecasted PoE bands for Australian Fused Materials (AFM) substation in summer (top left) and winter (top right) from 2008 to 2025. The lower panels are observed daily noncoincident maximum demand and forecasted PoE bands for AFM substation in summer (lower left) and winter (lower right). The vertical line represents the end of the observed summer training data on 30 April 2018 and winter training data on 31 October 2017.

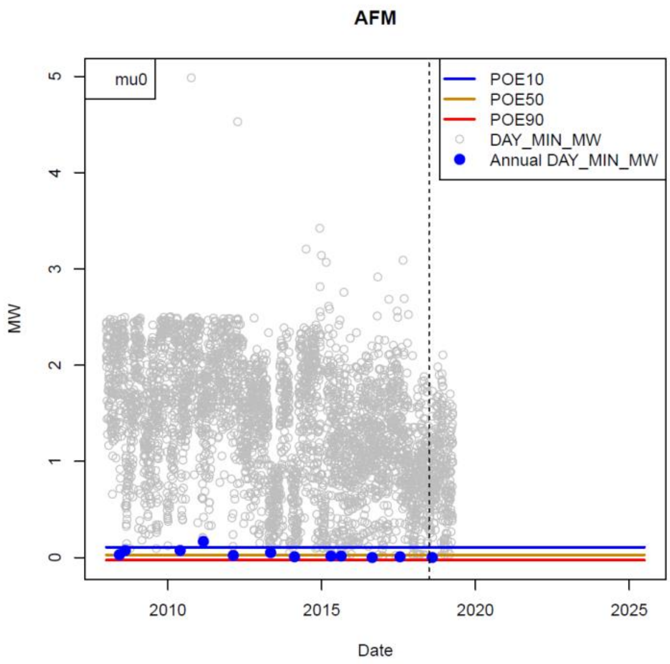


Figure 5.16 Observed daily day-time minimum demand (grey dots) and forecasted POE10 (red curve), POE50 (orange curve) and POE90 (blue curve) of annual noncoincident minimum demand forecast at Tx substation Australian Fused Materials (AFM) from 2008 to 2025. The vertical line represents the location at end of the observed training data on 30 June 2018.

6. Block load forecast

Western Power connects many new customers every year. Most of these connections are small energy users that are suitably included in the base energy and load demand forecasts and do not need to be accounted for separately. Less frequently, Western Power is required to connect new customers that represent a material increase in both energy and demand. Western Power must also cater for existing customers that have expansion or contraction plans that could result in a material increase or decrease in demand above or below the base forecasts. These new and existing customer loads are collectively referred to as block loads.

Once added to the base forecasts, a block load introduces an often-permanent step-change into the maximum demand of an otherwise smooth trend. Block loads minimum demand, for some substations, are managed via a ratio based on similar industry types. Such step-changes occur infrequently and are not easily accommodated by most statistical methods.

Western Power uses a systematic approach to determine the potential block loads that are to be included in the energy and demand forecasts. This approach is intended to be consistent, systematic, consultative, evidence-based and is applied to qualifying customer applications to determine if they satisfy a set of pre-defined conditions.

The 'likelihood to proceed' is considered the most critical of these conditions. Western Power receives many applications for connection to the network. Only a few applications proceed that result in significant new loads on the network. The forecast process runs an assessment and only includes those loads that are reasonably likely to proceed. This assessment requires an investigation and assessment of each potential block load's information that is retrieved from multiple internal and external sources. The determination of likelihood to proceed often involves discussions with both internal and external stakeholders. For example, consultation occurs between Western Power's Business Intelligence and Data Analytics function (BIDA) and the representatives form AEMO, Western Power's Grid Transformation and Customer Service functions.

The assessment of block loads produces a list of potential loads. These loads have been assessed with a high likelihood to proceed within the five-year forecast period and predict an energy and/or load demand potential that will materially affect the network and the corresponding base energy and load demand forecasts.

6.1 Preparation of block load forecasts

Western Power assesses applications for connection to the distribution and transmission network. All distribution connection applications that meet the conditions of a 'non-competing application'[27] are automatically cleared and approved for firm-access connection to the network and are subsequently regarded as part of the 'natural load growth' that is included in the developed base energy and demand forecasts. All competing applications are subject to a clearance and approval (to connect) process and are logged accordingly in the register load connections tracker.

27 For further details refer to https://westernpower.com.au/faqs/connect-to-the-network/new-connection/what-is-a-non-competing-application-threshold/ . All connection applications that are eligible for network tariffs RT1-RT6 (or equivalent e.g. RT13-RT16) and have a total load expectation not exceeding 1.5MVA for their National Metering Identifier (NMI) are considered part of the 'natural load growth'. These connection applications are deemed as complying and are catered for by the developed base energy and load demand forecasts.

All competing applications in the register load connections tracker are further refined to exclude those in progress, not cleared or for feasibility study only. The assessment also excludes small connection applications, which have an immaterial effect on network demand. [28]

The forecasting processes also uses a list of potential large transmission connected projects that have a high 'likelihood to proceed' within the five-year forecast period. This list provides detailed information about each large transmission connected project. The forecast process conducts an independent assessment of the list, which includes a crosscheck with the information that is stored in Western Power's customer relationship management application. [29] Any identified anomalies are discussed directly with the Customer Project Development team and the appropriate representative from the customer service function.

[28] All applications with an undiversified demand less than five megawatt are considered small connections and are excluded. The undiversified demand refers to a customer's requested maximum demand, expressed in megawatt, for the potential load. The actual effect of the connection on the network is influenced by the relation between demand profile of existing loads and the new connection. The effect of new connection on network demand is typically less than the undiversified load demand and is referred to as diversified demand.

[29] Salesforce was Western Power's customer relationship management application prior to the end of July 2019. It contained relevant customer and load information for potential major transmission connected projects. It was decommissioned and replaced with a new customer management system from August 2019. Salesforce was referenced for the block load forecast that was developed and used in 2019 medium-term forecasts.

7. Reactive power forecast

Reactive power[30] influences the operation of power system in four ways:

- Some loads consume reactive power, so it must be provided by some source.
- The transmission and distribution lines consume reactive power, so it must be provided by some source. Lines also provide some reactive power due to their capacitance, which offsets their consumption of reactive power.
- The flow of reactive power from supplies to the sinks causes additional heating of the lines and voltage drops in the network.
- The generation of reactive power can limit the generation of real power.[31]

The flow of reactive power on the network is managed by synchronous generators, reactors and capacitors. Sufficient reactive power is needed to be provided to loads and lines in the network, but excessive reactive power causes excess heating and voltage drops. If not managed suitably, voltage drops can result in voltage collapse and instability of the network.

Forecast of reactive power is required to assist Western Power with the operation and planning of the network. The method produces four reactive power forecasts for each substation. It comprises two forecasts for reactive power coincident with substation minimum and maximum demand that is coincident with system minimum and maximum demand and two other forecasts coincident with substation minimum and maximum demand.

The data generating process for reactive power forecast is complex because charging current is a localised measure. Western Power does not measure the movement of power across the network but the current on each section of the network contributes to the overall reactive power supply. Also, individual network planning decisions (e.g. which type of underground cable to install) and operational decisions (e.g. which switches should be changes to balance the network) influence reactive power injection. Finally, individual customer choice (e.g. use of pool pumps and washing machines) influences reactive power absorption. The supply and absorption of reactive power by transmission and distribution lines have a complex nature. Several variables influence the balance of supply and demand for reactive power, including line capacitance, reactance, voltage level, current and length. From a broad perspective the uncompensated[32] reactive power throughput on the network can be depicted using the below equation:

$$Reactive\ power = network\ charging\ capacity \times charging\ current - customer\ absorption$$

Much of the data required to directly estimate the supply and consumption of reactive power on the network is difficult to obtain. For example, the network charging capacity is partially a function of the type

[30] Reactive power has its origin in the phase shift between sinusoidal voltage and current waveforms. When current wave to a device lags the voltage wave, it consumes reactive power.

For a detailed discussion refer to Sauer P.W. (2005) Reactive Power and Voltage Control Issues in Electric Power Systems. In: Chow J.H., Wu F.F., Momoh J. (eds) Applied Mathematics for Restructured Electric Power Systems. Power Electronics and Power Systems. Springer, Boston, MA.

[32] Uncompensated reactive power refers to the amount of reactive power that would have travelled over a particular section of network if the local capacitor banks and/or reactors were not running. It is a measure of the underlying reactive power supply and absorption before being influenced by network operation controls.

and length of conductors, which change due to changes to network design, augmentations and the replacement of overhead network with underground cables.[33]

Instead of a fundamental model, the forecast employs a partial-linear model using proxy values for the drivers of reactive power:

$$Q_p = \beta_0^p + \beta_1^p NMI + f^p(D) + \epsilon^p$$

where,

p is the estimated p^{th} quantile, and $0 < p < 1$ indicates the proportion of the population scores (e.g., daily reactive power Q) below the quantile p,

Q_p is the p^{th} quantile level of reactive power (Q), expressed in mega Volt-Amps-Reactive (MVAr). We forecast the p^{th} quantile associated with the POE10, POE50 and POE90 levels,

NMI is the customer count measured by the number of NMIs (proxy for the network length)

D is coincident demand, expressed in MW (approximation for charging current)

$f^p(D)$ is the smooth function of D, given p. It is data-driven relationship in the model, which can be linear and non-linear

ϵ^p is residual error, often assumed to be a Gaussian process.

The model is static in nature by assuming that there will be no significant changes to the relation between explanatory variables and reactive power. A model is fit separately to each substation to account for the local characteristics of the downstream network.

The model uses NMI count as a reasonable approximation for growth in the network length, solar PV capacity and average energy consumption as reasonable approximations for customer absorption of reactive power, and coincident demand as a reasonable approximation for charging current.

Reactive power has a quadratic relation with load and line charging current.[34] With increased penetration of rooftop solar PV in the distribution system, the flow of power can reverse, and the model should account for this change in the flow of power. The linear model developed above cannot suitably account for this effect. A separate estimation was conducted to estimate the line charging current of the distribution network at each feeder and substation. This provided an estimate of a floor to limit the decrease of reactive power on feeders and substations.

Estimation of line charging current

Line charging current is a function of the conductor type, length and voltage level. A linear regression was conducted to model reactive power using the length of several conductor and cable types and voltage level as explanatory variables. The model results estimated the minimum reactive power coincident with substation or feeder minimum demand periods.

---

33 Underground cables have high capacitance and generally supply reactive power to the network.

34 An improvement to the forecast of reactive power in the future can be provided by using squared demand in the forecast of reactive power.

The forecast of reactive power will be improved in several ways. Some of the proposed future improvements to the model are:

- Using squared demand in the forecast to better reflect the functional form of charging current
- Including an estimation of the changes to reactive power due to changes in voltage
- Including a breakdown of NMI growth by subdivision and infill, and
- Including an analysis of planned changes to engineering standards.[35]

[35] Engineering standards, such as Volt/Var compensation, are likely to have a significant influence on reactive power requirements. It is not yet possible for Western Power to precisely model the effect of these changes, because the voltage data at each customer connection is not available.

Appendix A

A.1 Project references

Project Wiki page

Details about the implementation of forecasts is provided on 'Customer Technology Energy Demand' Wiki page:

http://wiki/display/BIDA/Customer+Technology+Energy+Demand

Forecast results

Forecast results are presented on Qlik dashboards:

https://qlik.ads.westernpower.com.au/sense/app/b8ce8f91-46c6-4087-9efd-6278f27afcd3

Results are also summarised in the EDM document 49890576:

http://edm.westernpower.com.au/otcs/cs.exe/Overview/50074955

Project implementation

The project is implemented in SAS Forecast Studio, SAS Enterprise Guide and R.

SAS Forecast Studio was used to create customer number forecasts. The project is titled CUSTED19_CUSTOMER_COUNT_NOL3.

Energy forecasts were conducted in SAS Enterprise Guide. The project folder is titled CusTED2019:

SAS Path: Western Power\Business Intel & Data Analytics\CusTED19

Maximum and minimum demand forecasts were implemented in R. The project folder is:

\\ho-fs1\SHARE\Business Intelligence and Data Analytics\02_Analytics_Projects\CUSTED19

All codes implemented in R and relevant markdown files were packaged and placed on the EDM (#50830691):

http://edm.westernpower.com.au/otcs/cs.exe?func=ll&objaction=overview&objid=50830691