

# Market Performance Quarterly Review

---

January-March 2022

# Contents

1	Purpose of this report	3
2	Highlights	3
3	Atmospheric Conditions	3
4	Demand	4
5	Retail	5
6	Wholesale	7
7	Forward Market	17
8	Deep Dive: Exploring NZ's Electricity Generation with Association Rule Mining	19

## Figures

Figure 1:	Composite La Nina summer pressure and wind patterns over four events	4
Figure 2:	Daily Load, March 2022 Quarter vs Historic Average.	5
Figure 3:	Changes in retailer market share	6
Figure 4:	ICP switches by type 2021/22	7
Figure 5:	Average Daily Wholesale Spot Prices	8
Figure 6:	Daily national storage and daily average spot price	8
Figure 7:	Half hourly wind generation	9
Figure 8:	Half hourly thermal and peaker generation	9
Figure 9:	Weekly generation by fuel breakdown	10
Figure 10:	Average Daily Generation by Fuel Type	11
Figure 11:	Controlled National Hydro Storage 2021/2022	12
Figure 12:	Major Lake Storage v mean, 10 <sup>th</sup> -90 <sup>th</sup> percentile	13
Figure 13:	Daily Gas Production and Consumption 2021/2022	14
Figure 14:	Ahuroa Gas Storage by Month, ending March 2022	15
Figure 15:	Daily Spot Gas Prices	16
Figure 16:	Estimated fuel usage at Huntly 2021/22	16
Figure 17:	Future Prices	18
Figure 18:	New Zealand Carbon Unit Spot Price	19

# 1 Purpose of this report

- 1.1 This document covers a broad range of topics in the electricity market. It is published quarterly to provide visibility of the regular monitoring undertaken by the Electricity Authority (Authority).
- 1.2 This report also includes:
  - (a) Exploring NZ's Electricity Generation with Association Rule Mining.

## 2 Highlights

- 2.1 Weekly load was higher than the historical average<sup>1</sup> in January. From mid-February onwards, however, weekly load was lower than average. These trends were likely due to unseasonably high temperatures associated with La Niña, possibly combined with Covid-related behaviour changes - as the country entered the Red traffic light setting.
- 2.2 Market share of larger retailers this quarter marginally decreased, while for small-medium sized retailers it slightly increased, with Electric Kiwi having the greatest growth. Trader switches increased, whilst move-in switches remained fairly constant. Trader switches may have increased due to the April price increases that some retailers made. Move-in switches are driven by how often people are changing where they live.
- 2.3 Wholesale electricity prices varied significantly, influenced largely by swings in hydro storage. The average spot price for the quarter was \$173/MWh, which is over \$100/MWh greater than the December 2021 quarter. High prices were experienced in January as the country experienced low inflows, resulting in some daily average prices over \$300/MWh, whilst heavy rain in February brought prices close to zero. Other factors influencing prices were variable wind generation, higher coal usage and a HVDC outage.
- 2.4 Wind generation was lower this quarter compared to last, however, lower wind generation does typically occur during a La Niña summer.
- 2.5 Nationally hydro storage decreased over the quarter, which is common during a La Niña summer. Below average inflows at Lakes Te Anau and Manapouri constrained their hydro generation, as they entered their low operating range in late January.
- 2.6 Gas spot prices increased substantially compared to last quarter, reaching almost \$50/GJ on 23 March.
- 2.7 Forward prices increased over the quarter, as more participants priced in additional risks related to winter 2022, which include possible continued low hydro inflows, upcoming gas and generation outages, and increasing thermal fuel costs.

During the March ETS auction 10.5 million NZUs were sold, after the triggering of the cost containment reverse at \$70/unit, which added an additional 5.7 million NZUs.

## 3 Atmospheric Conditions

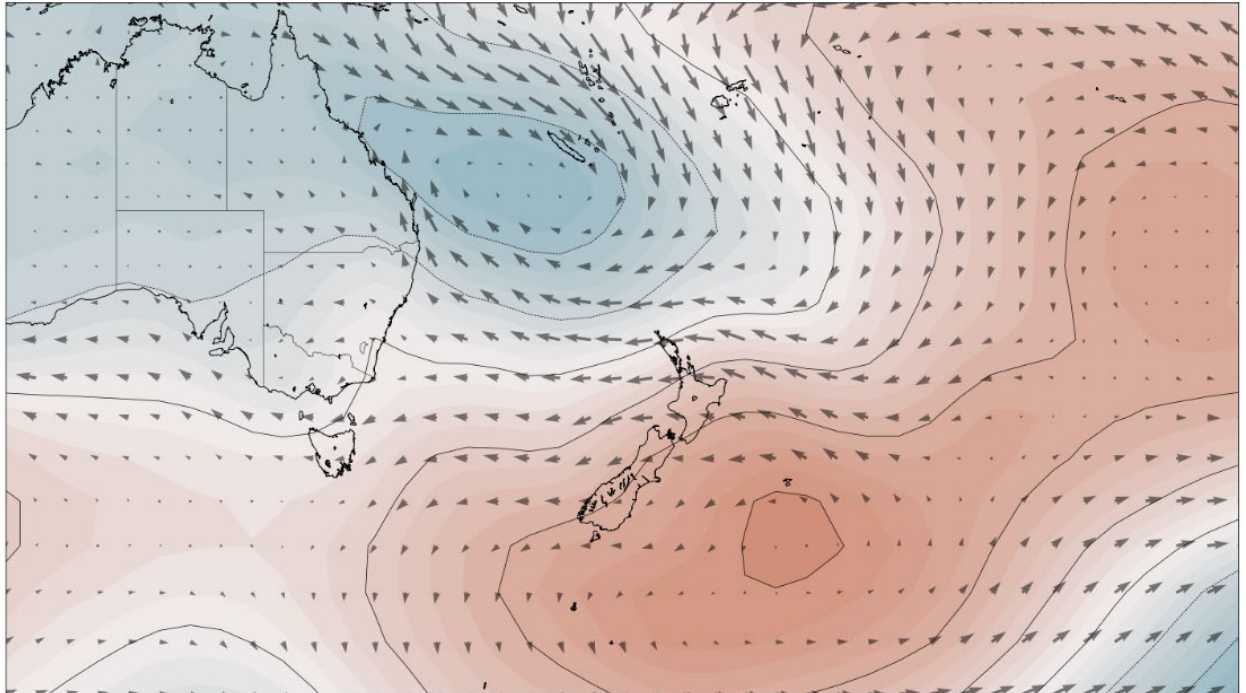
- 3.1 Since November 2021, the La Niña phase of the EL Niño Southern Oscillation (ENSO), has been influencing weather in New Zealand. During La Niña, New Zealand typically experiences moist, rainy conditions in the north and east of the North Island and reduced rainfall for the lower and western South Island - resulting in reduced inflows for the Southern hydro lakes. The country will also tend to experience unusually high sea

---

<sup>1</sup> Averaged over January to March for 2017-2021

surface temperatures and air pressure, causing higher temperatures, and generally less generation from wind. A composite of a summer pressure and wind pattern during four different La Niña events is shown in Figure 1. Areas of above normal pressure are shaded in red while areas of below normal pressure are shaded in blue. The wind direction is indicated by the arrows.

**Figure 1: Composite La Nina summer pressure and wind patterns over four events<sup>2</sup>**



## 4 Demand

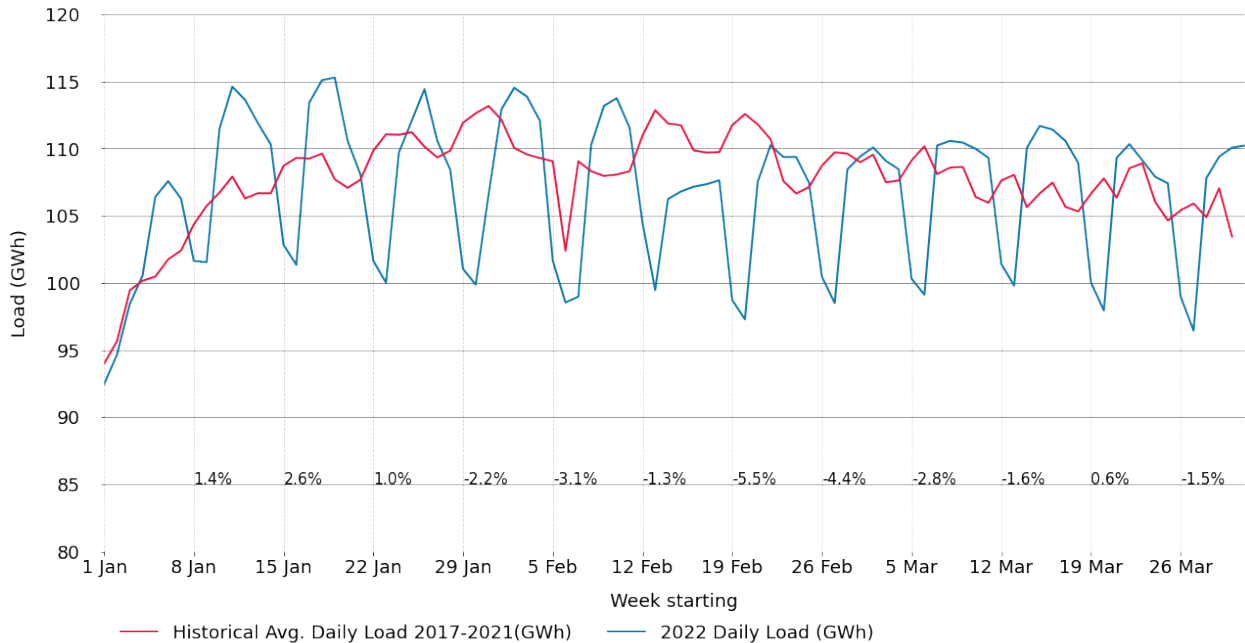
- 4.1 In mid-January, the detection of the Covid-19 variant Omicron saw New Zealand move into the “Red” “traffic light”, under the COVID-19 Protection Framework. In this setting, people were encouraged to work from home, and entry status into businesses, like hospitality, was dependent on having a vaccine pass. Unlike the previous alert level systems, many businesses remained open and people were able to travel freely.
- 4.2 Figure 2 shows total daily national demand for the 2022 March quarter against the average daily national demand for the 2017 - 2021 March quarters. Note that the historical average is based on the date, so the average for any day will likely include weekends (which have different demand patterns from week days). Annotations display the weekly percentage difference between the 2022 load and the 2016-2021 historical average load.
- 4.3 After the Christmas/New Year holiday period, weekly load for the quarter peaked at 658 GWh, and was 2.6 per cent greater than the historical average. Weekly demand began decreasing from mid-February, and was up to 5.5 per cent less than the historical average. This is when daily Covid cases in New Zealand began totalling over 1000 per day. Weekly load remained lower than January levels throughout March.
- 4.4 High load in January may be due to the exceedingly dry and hot conditions, resulting in a higher usage of air-conditioning in the North and irrigation in the South. This may have

<sup>2</sup> [niwa.co.nz/climate/information-and-resources/el-nino](https://niwa.co.nz/climate/information-and-resources/el-nino)

been compounded by a split workforce, with some at home and some in the office. Lower than average load in February and March may also be due to the warmer temperatures, which were 0.5°C<sup>3</sup> and 1.3°C<sup>4</sup> warmer than 1981-2010 average respectively. This could have resulted in less use of heating.

- 4.5 Reconciled demand for each month January, February and March 2022 was 3,302 GWh, 2,987 GWh, and 3,315 GWh respectively.

**Figure 2: Daily Load, March 2022 Quarter vs Historic Average.**



## 5 Retail

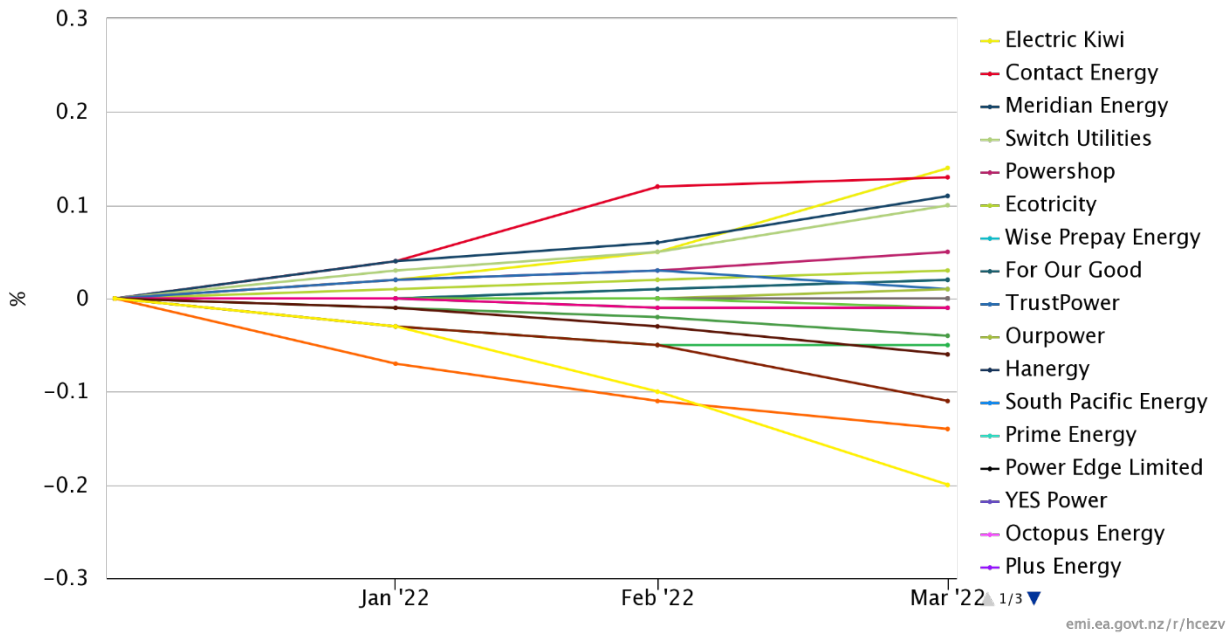
- 5.1 Over the March 2022 quarter, collective market share of the five largest retailers, being Contact, Genesis, Mercury, Meridian and Trustpower<sup>5</sup>, decreased minimally by 0.11 per cent, to 84.06 per cent.
- 5.2 On 31 March 2022 the five largest retailers held 1,875,354 ICPs between them, gaining 1,310 ICPs over the quarter. Small-medium sized retailers held 355,498 ICPs between them, gaining 3,019 ICPs.
- 5.3 Figure 3 shows the changes in market share of each retailer from 1 January to 31 March 2022. Electric Kiwi and Contact gained 0.14 and 0.12 per cent of market share respectively. The retailer with the greatest decline in market share was Mercury, losing 0.2 per cent. Genesis also lost 0.14 per cent.
- 5.4 The largest regional change in ICPs came from Contact’s gains in Auckland, with an increase of 2,635 ICPs over the quarter.

<sup>3</sup> [Climate Summary for February 2022 | NIWA](#)

<sup>4</sup> [Climate Summary for March 2022 | NIWA](#)

<sup>5</sup> As of 1 May 2022 Trustpower sold part of its business to Mercury and changed its name to Manawa Energy. In future Electricity Authority reports it will not be bundled with the other large retailers as it has sold its retail assets to Mercury.

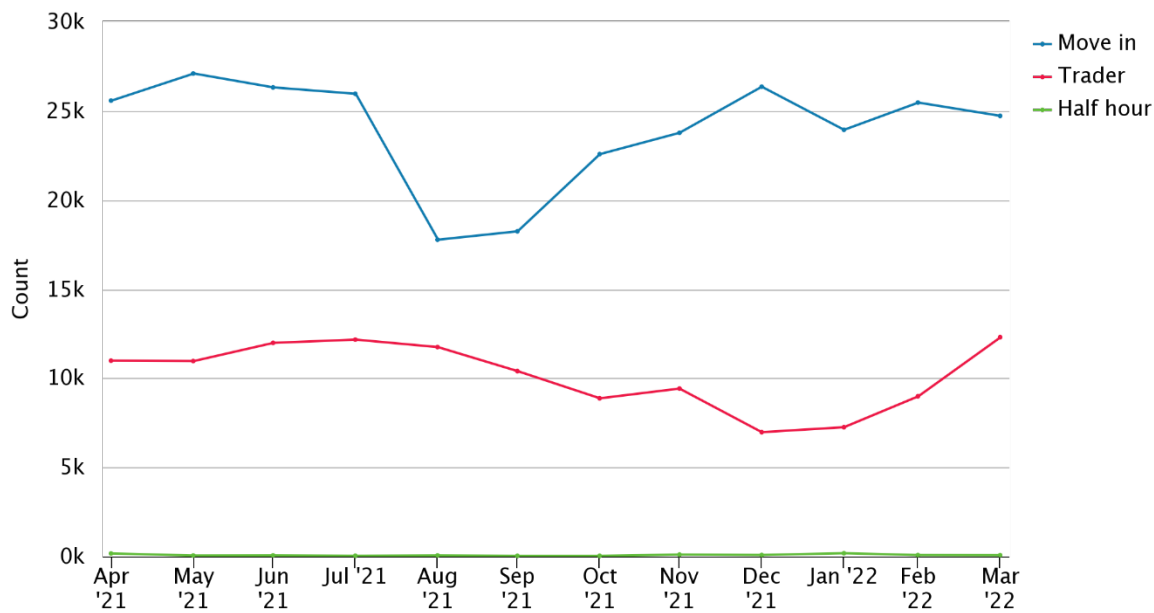
**Figure 3: Changes in retailer market share<sup>6</sup>**



- 5.5 Figure 4 shows the number of electricity connections (ICPs) that have changed electricity suppliers from 1 April 2021 to 31 March 2022 categorised by type 'move in', 'trader' or 'half hour'. Move in switches are switches where the customer *does not have* an electricity provider contract with a trader, whereas trader switches are switches where the customer *does have* an existing contract with a trader, and the customer obtains a new contract with a different trader.
- 5.6 Over the quarter trader switches almost doubled, from 6,949 per month in December 2021 to 12,823 per month in March 2022. Move in switches on the other hand remained constant, at roughly 25,000 per month.
- 5.7 Compared to the previous five years total trader switches in March were higher than at any point in 2021, and move-in switches are roughly the same as March 2021. The increase in trader switches could be due to April the price increases made by some retailers, causing consumers to shop around.

<sup>6</sup> Please note that not all traders fit in the key, please go to [emi.govt.nz/r/hcezv](http://emi.govt.nz/r/hcezv) to view key with all traders

**Figure 4: ICP switches by type 2021/22**



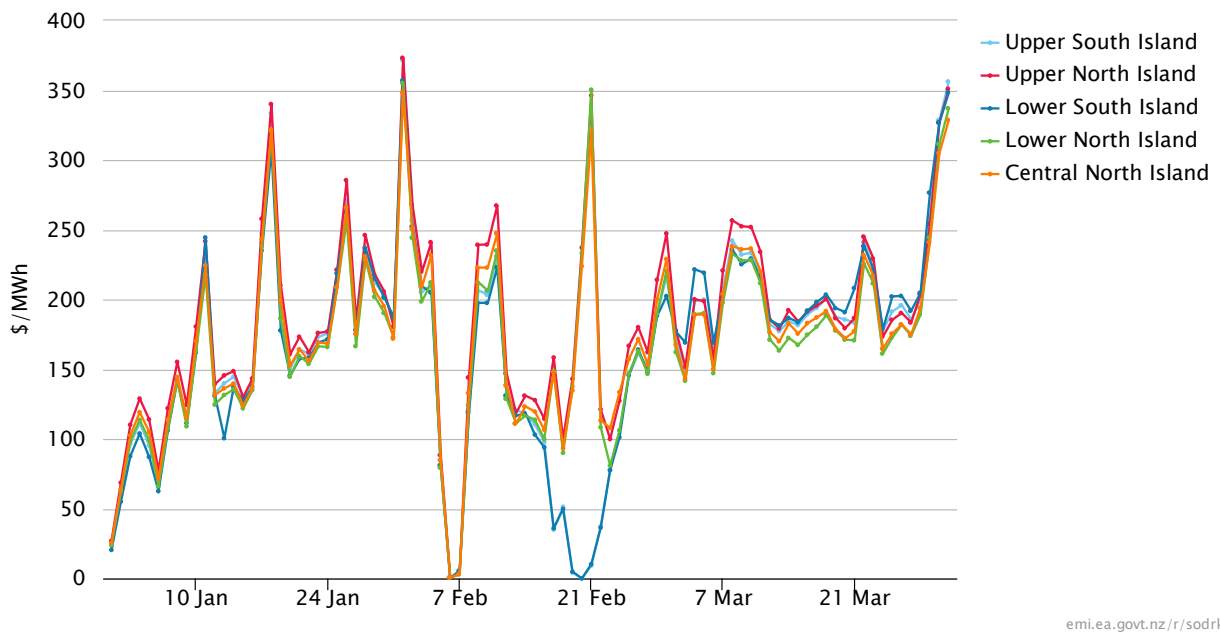
emi.ea.govt.nz/r/4rhun

5.8 All three conditions of sale were met for the sale of Trustpower’s retail business to Mercury with the High Court’s decision in mid-December to approve the proposed restructuring of Tauranga Energy Consumer Trust’s (TECT) trust deed. The transaction has been completed in early May 2022.

## 6 Wholesale

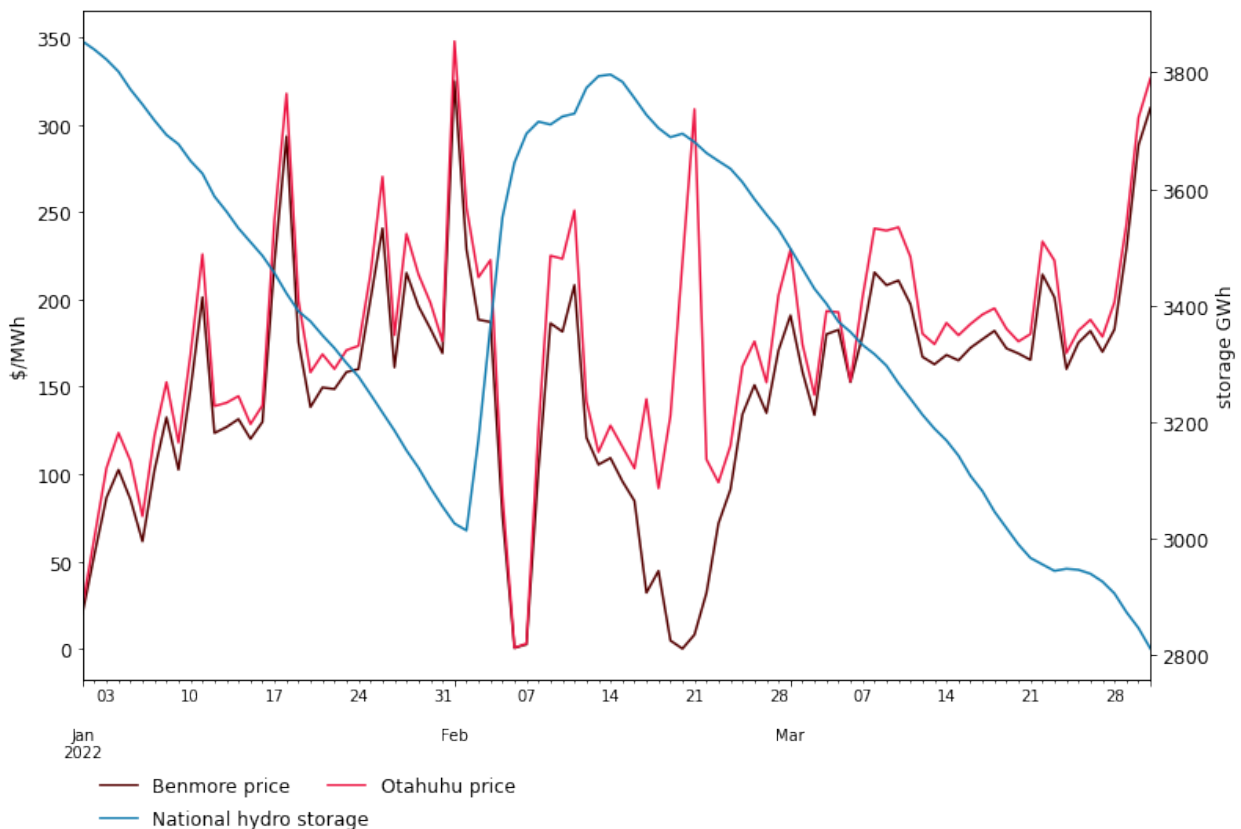
- 6.1 Wholesale electricity spot prices fluctuated during the March 2022 quarter. Daily average spot prices, from 1 January to 31 March 2022, are shown below in Figure 5. Spot prices averaged \$173.6/ MWh across all nodes, which is over \$100 more than the \$68.54/MWh average for last quarter.
- 6.2 The highest spot price for the quarter was \$761/MWh. This occurred on 1 March and coincided with a Rankine unit tripping at Huntly, which resulted in extra generation being required from hydro and peakers.
- 6.3 Spot prices this quarter have been impacted by fluctuating hydro storage. Prices rose in January as the country experienced low inflows. In contrast, prices fell to close to zero during a large rainfall event in early February. Prices increased in March as hydro storage again decreased, see Figure 6.
- 6.4 Outages of note for the quarter include a HVDC outage, which caused price separation between 16 – 22 February. There were other large, planned outages at Huntly, TCC and Manapouri. Some of these outages were likely scheduled to coordinate with transmission outages, such as the Clutha Upper Waitaki Lines project in the lower South Island which began in 2020 and was completed in April 2022.

**Figure 5: Average Daily Wholesale Spot Prices**



emi.ea.govt.nz/r/sodrk

**Figure 6: Daily national storage and daily average spot price**

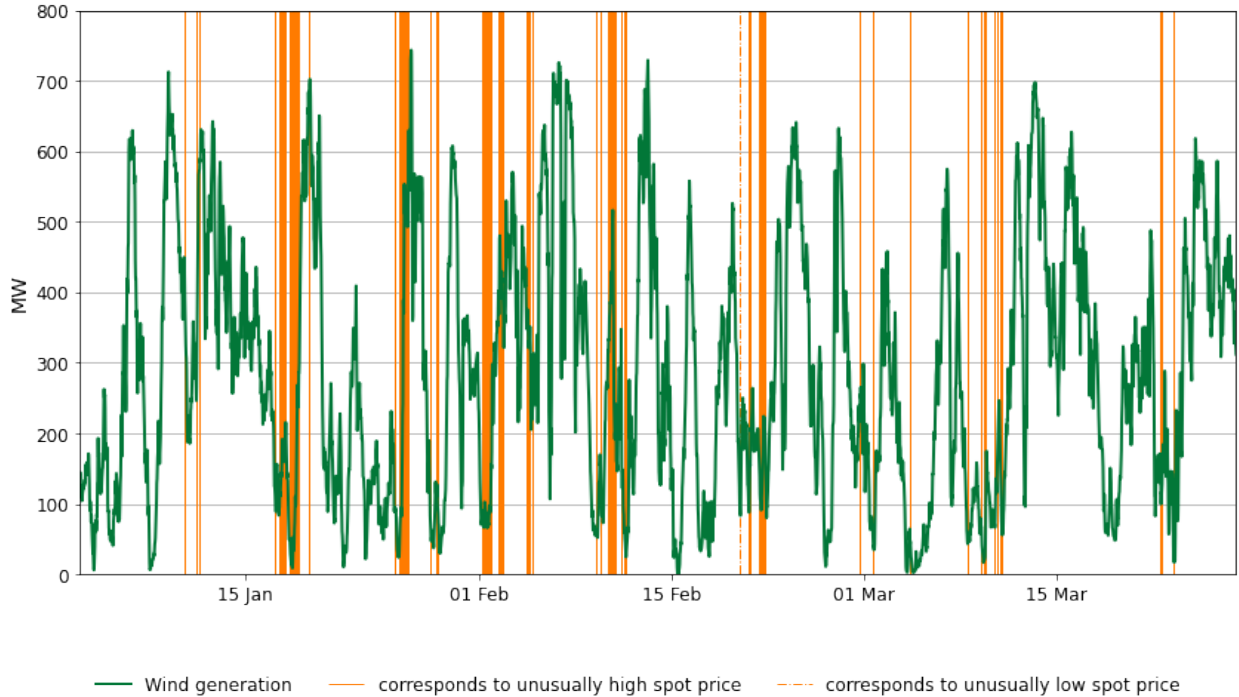


6.5 Generally, outside of unusual circumstances, spot prices increased when grid demand increased, wind generation decreased, and gas and coal powered generation increased. The relationship between wind generation and prices can be seen in Figure 7, which shows national half hourly wind generation for with quarter and instances when

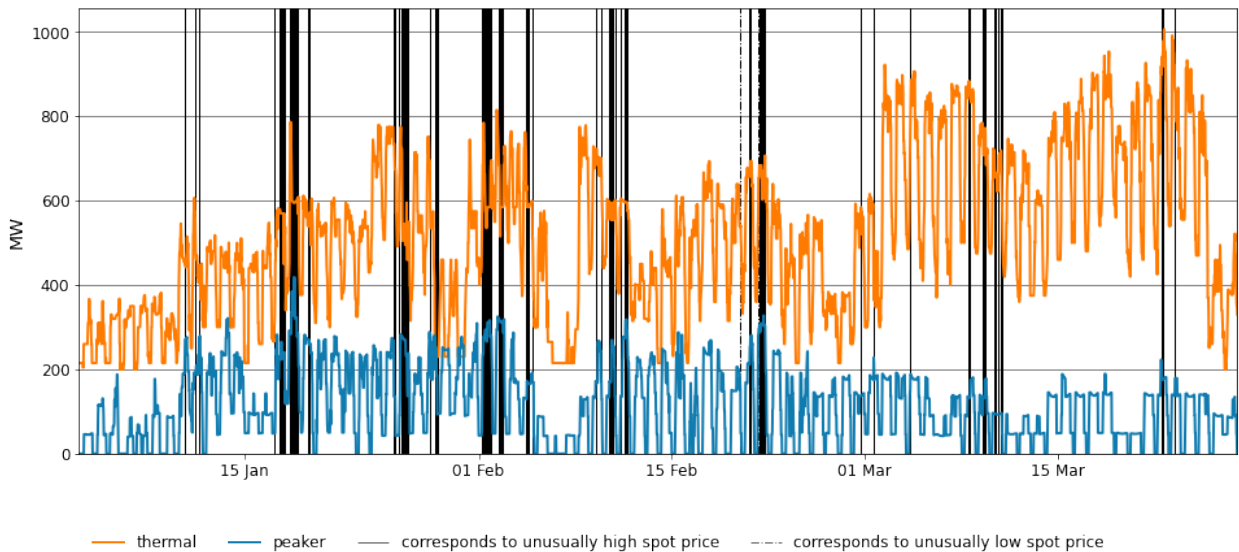


unusually high prices occurred. Most high prices<sup>7</sup> occurred when wind generation was low. Thermal generation offers set high prices for higher tranche offers due to a constrained gas supply keeping the cost of thermal generation high, see the relationship of high thermal usage often corresponding to high spot prices in Figure 8. Also, decreasing hydro storage pushed prices up, as water has been priced high to conserve. Figure 6 shows the daily averaged spot price against the national hydro storage.

**Figure 7: Half hourly wind generation**



**Figure 8: Half hourly thermal and peaker generation**



<sup>7</sup>Classified as when a price is above the 90<sup>th</sup> percentile of historical prices from 1997-2021

6.6 Generation from renewable sources for the quarter averaged just over ~85 per cent of total generation. The slight decrease from last quarter is due to reduced hydro generation, with the shortfall being picked up by extra thermal. Weekly variation can be seen in Figure 9.

**Figure 9: Weekly generation by fuel breakdown**

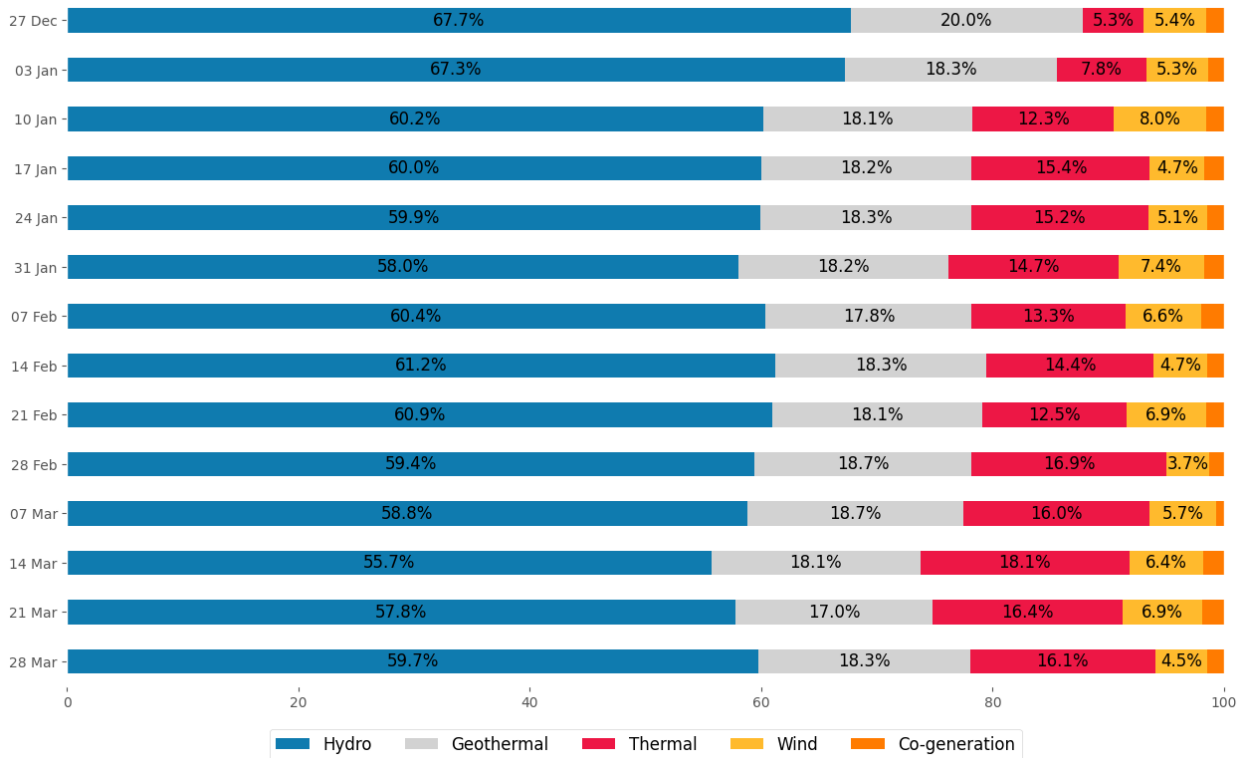
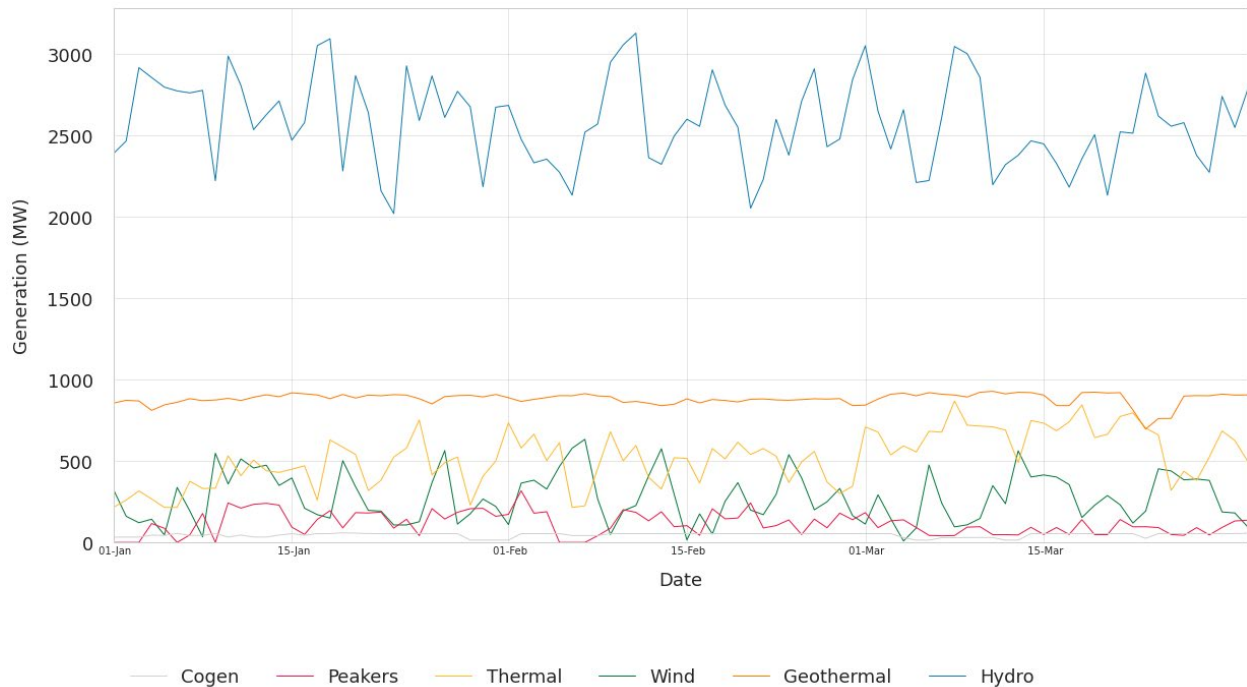


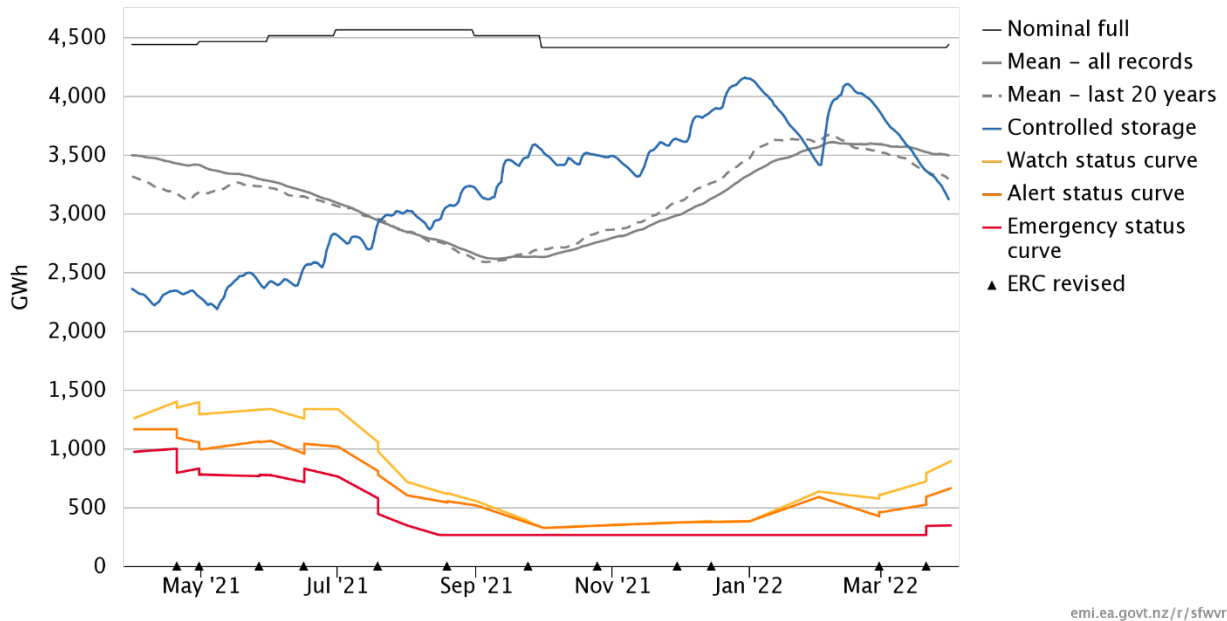
Figure 10 shows the percentage of daily generation for the quarter by fuel type. Wind generation had a half hourly generation averaging 256 MW for the quarter, 5.9 per cent of total generation. Half hourly hydro generation averaged 2,600 MW, 60 per cent of total generation, and half hourly thermal generation averaged 624 MW, 14.2 per cent of total generation.

**Figure 10: Average Daily Generation by Fuel Type**



- 6.7 Hydro generation was constrained by a decrease in hydro inflows and storage. Figure 11 shows total national controlled hydro storage up to 31 March 2022. Over the March quarter hydro storage decreased by 1,028 GWh, from 4,4148 GWh on 1 January 2022 to 3,120 GWh on 31 March 2022.
- 6.8 On 1 January 2022 hydro storage was at its highest point for the quarter, and for the year thus far, at 125 per cent of historical mean (3,329 GWh) and 94 per cent of nominal full (4,412 GWh). On March 31 storage was at its lowest, at 89 per cent of the historical mean (3495 GWh) and 70% of nominal full (4,437 GWh)
- 6.9 Hydro inflows were low this quarter, especially in the deep south. During the quarter, total national inflows were 5,121 GWh, which is 24 per cent less than the historical average of 6,767 GWh.
- 6.10 Inflows at Lake Manapouri were exceptionally low, at 137 GWh, 65 per cent less than the historical average of 393 GWh. Inflows at Lake Te Anau totalled 373 GWh, 59 per cent less than the historical average of 909 GWh.

**Figure 11: Controlled National Hydro Storage 2021/2022**

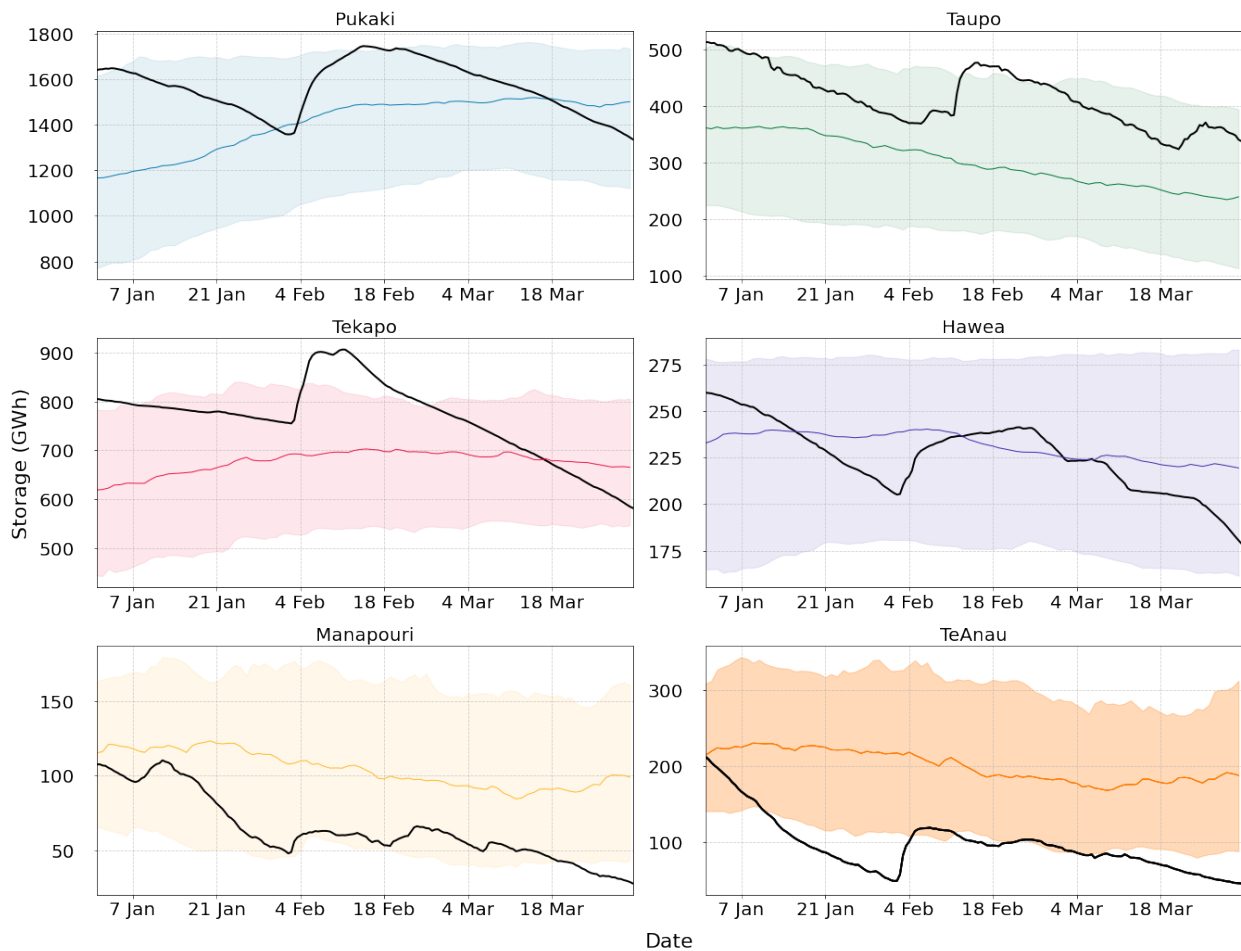


emi.ea.govt.nz/r/sfwr

Figure 12 shows the storage of major catchments Lakes Pukaki, Taupo, Tekapo, Hawea, Manapouri and Te Anau for the quarter against their historical means and 10th-90th percentiles based on data from 1926-2022. Storage at Pukaki, Taupo and Tekapo was healthy for the first half of the quarter, with all lakes above their historical mean. By mid-February, however, all three began declined due to low inflows. By the end of March, only Taupo was above its historical average. Storage at Hawea dropped below its historical average in mid-January, and was briefly above it again in mid-February. By late March, however, Hawea was reaching its 10<sup>th</sup> percentile.

- 6.11 Lake Manapouri levels began dropping to their historical 10th percentile in late January. Meridian hit Manapouri's low operating range by 30 January 2022. While Manapouri is in its low operating range draw-down cannot exceed 5 cm per day. This constrained generation from Manapouri to ~6-6.5 GWh per day. To manage the situation Meridian withdrew Manapouri capacity from the market to maintain full control over drawdown of the lake.
- 6.12 During February there was spill at Tekapo due to large inflows combined with an 80 MW outage at Tekapo B, which occurred from September 2021 – mid February 2022.

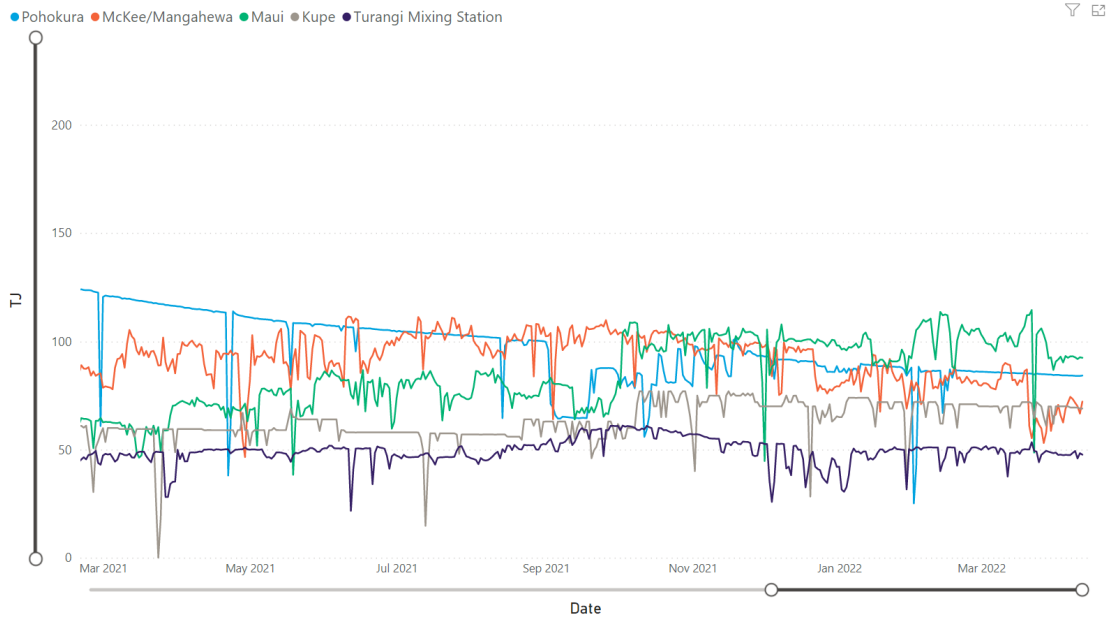
**Figure 12: Major Lake Storage v mean, 10<sup>th</sup>-90<sup>th</sup> percentile**



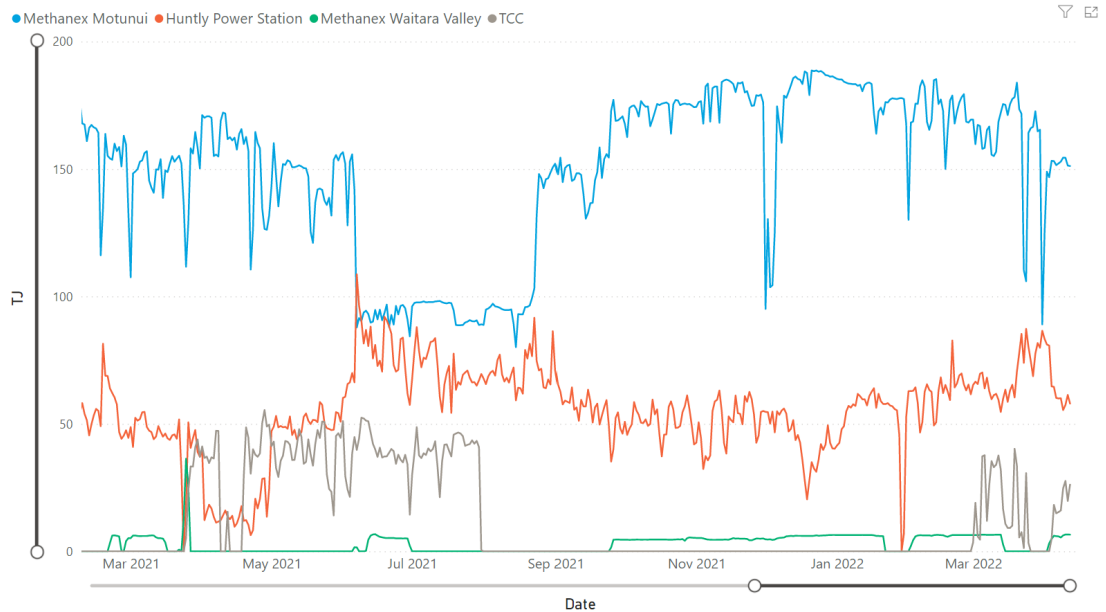
- 6.13 Figure 13 shows gas production by major fields and gas consumption by major users for the previous year.
- 6.14 Total gas production for the March quarter decreased by 23 TJ/day, from 370 TJ/day on 1 January 2022 to 347 TJ/day on 31 March 2022.
- 6.15 Maui was the largest producing gas field during the quarter, producing an average of 100 TJ/day before an unexpected outage on 23 March. Output at Maui then reduced, and was averaging 93 TJ/day between 23-31 March. Output at McKee averaged 82 TJ/day, before a planned outage on March 21, after which it averaged 60 TJ/day. Kupe appears to have begun declining again, after a period of relative plateau. Pohokura continues its long-established decline in output, which has fallen from ~245 TJ/day in 2018, to just 80-85 TJ/day in 2022.
- 6.16 Consumption by major gas user Methanex decreased slightly, from 185 TJ/day on January 1 to 165 TJ/day on March 30. There were days where consumption suddenly dropped by more than 30 TJ/day, these include February 1, March 23-24, and March 31. Some of these coincide with higher gas usage at Huntly and or TCC.
- 6.17 Huntly gas usage steadily increased throughout the quarter as more baseload demand was being met with thermal as hydro generation decreased. Consumption was 41 TJ/day on January 1 and 86 TJ/day on March 31.

**Figure 13: Daily Gas Production and Consumption<sup>8</sup> 2021/2022**

**Daily Gas Production by Major Fields**



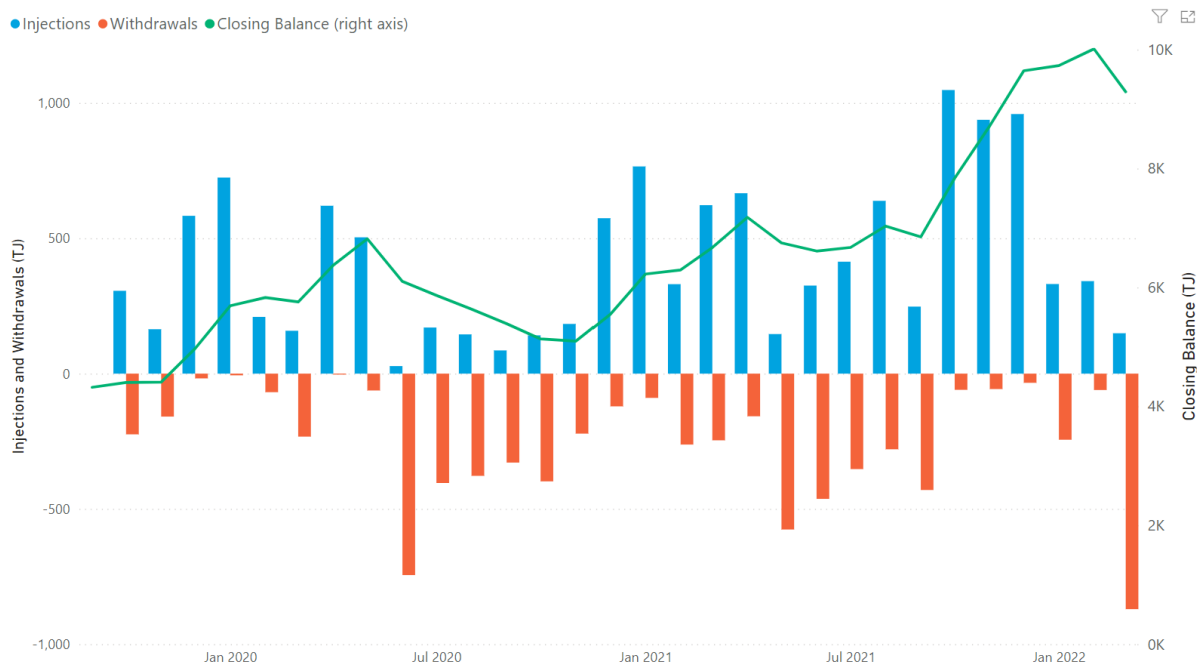
**Daily Consumption by Largest Users**



6.18 Despite the availability of surplus gas earlier in summer, there were low injection volumes into the Ahuroa underground gas storage (AGS) facility. Performance issues at AGS were confirmed by a footnote in Contact’s 1H FY22 report which noted an unexpected and unexplained increase in pressure at AGS. With gas production still relatively constrained the effects on the market have been muted, though this may change following drilling at Maui B. Storage at the end of March was close to 9 PJ, see Figure 14, which is almost double the amount from March 2021.

<sup>8</sup> <https://www.gasindustry.co.nz/about-the-industry/gas-industry-information-portal/gas-production-and-major-consumption-charts/>

**Figure 14: Ahuroa Gas Storage by Month, ending March 2022<sup>9</sup>**

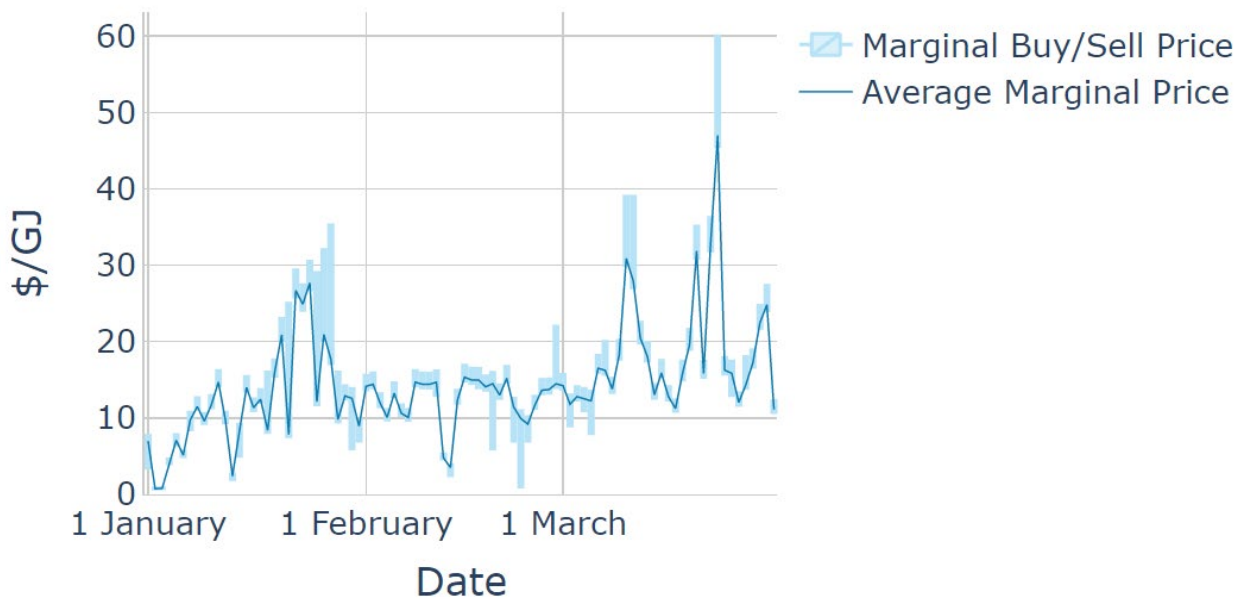


- 6.19 Figure 15 shows the Maui pipeline average marginal price (AMP) for the March quarter. Pricing data was taken from BGIX<sup>10</sup> (Balancing Gas Information Exchange) which is used here as a proxy for gas spot prices.
- 6.20 Gas spot prices increased in the March quarter, following their relatively steady and low prices seen in Q4 2021, roughly \$10/GJ. Prices rose to almost \$30/GJ in late January, but sank back down to \$10-15/GJ in February. Prices increased again in March, peaking at ~\$50/GJ on March 23.
- 6.21 Coal prices similarly rose, with Indonesian coal (the coal Genesis imports) reaching over NZ \$600/tonne at some points of the quarter, almost four times the price in 2020. This was driven mostly by global sanctions on Russian energy resources, including gas, oil and coal, as a result of the War in Ukraine. This increased the opportunity cost of running the Huntly Rankines to over ~\$200/MWh.
- 6.22 The latest coal numbers from Genesis put Huntly's coal stockpile at 817,000 tonnes at the end of the March 2022 quarter, a decrease of 18,000 tonnes from December 2021. This is, however, 526,000 tonnes more than the March 2021 quarter. Note thermal generation offers tend to reflect the opportunity cost of using thermal fuel with offers based on current market prices rather than costs incurred.
- 6.23 Notably, coal use this quarter at Huntly increased substantially from Q4 2021, as seen in Figure 16.

<sup>9</sup> <https://www.gasindustry.co.nz/data/gas-storage/>

<sup>10</sup> [BGIX - Balancing Gas Information Exchange](#)

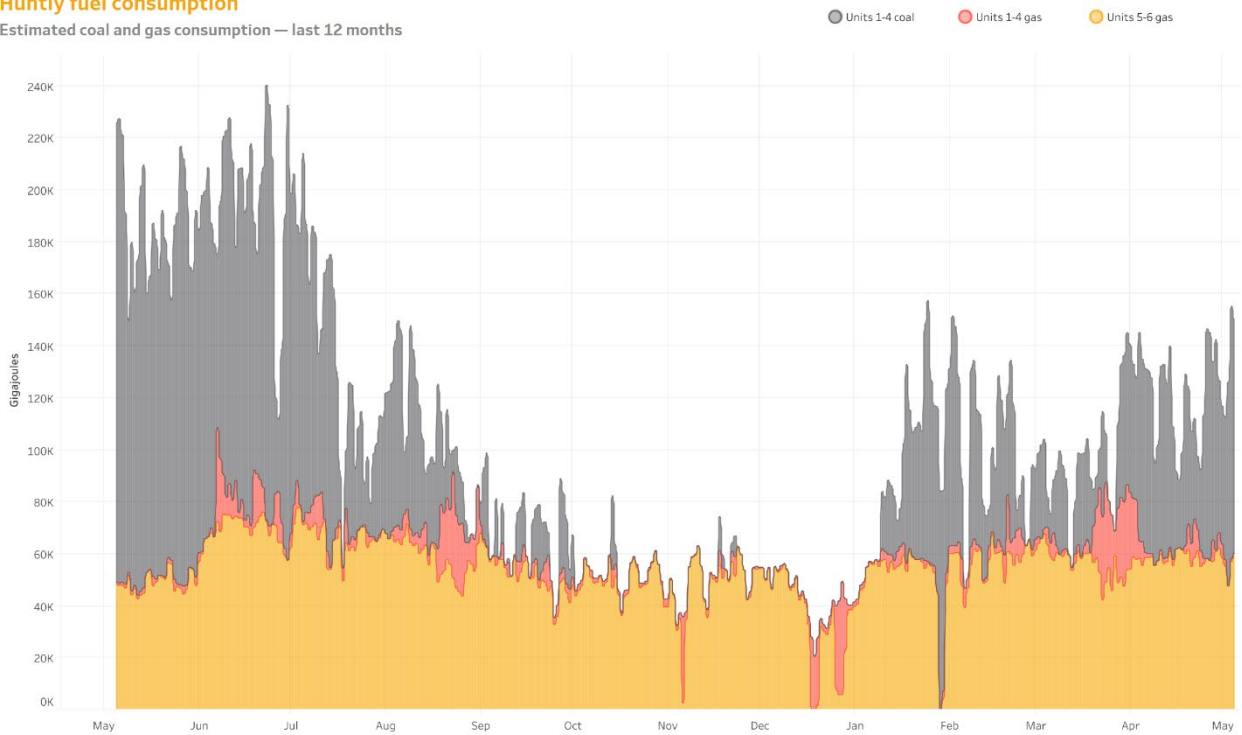
**Figure 15: Daily Spot Gas Prices**



**Figure 16: Estimated fuel usage at Huntly 2021/22**

**Huntly fuel consumption**

Estimated coal and gas consumption — last 12 months





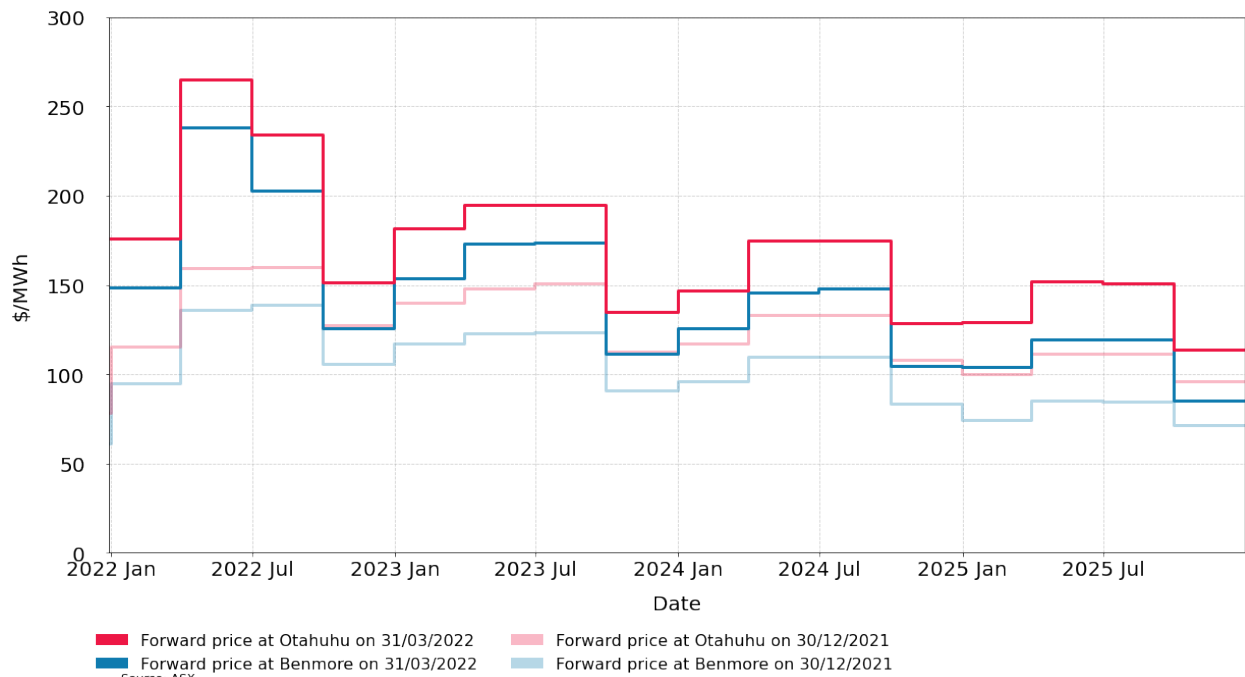
## 7 Forward Market

- 7.1 The ASX forward price curve provides a view of future wholesale spot prices. Figure 17 shows forward prices for Otahuhu and Benmore at the beginning of the quarter and at the end of the quarter to illustrate how forward prices have changed.
- 7.2 Short term forward prices (June 2022 and October 2022 quarters) rose by around ~\$100/MWh and ~\$80/MWh respectively over the quarter, sitting between \$200/MWh and \$260/MWh on 31 March 2022. Long term forward prices rose by roughly \$50/MWh.
- 7.3 These rises were fueled by high coal prices, high carbon prices, a prolonged trend of low hydro inflows, skewed North and South Island hydro storage, generation outages and gas outages, as well La Niña. Collectively, forward prices are reflecting all these factors as participants factor in cost and risk.
- 7.4 Forward prices in mid-2022 were priced at around ~\$230/MWh, indicating that the market is factoring in higher than usual supply risk for winter 2022. Below average gas production and potential low future inflows at the time increased market participant concerns that the same conditions that caused high prices in 2021 could repeat in 2022.
- 7.5 The concern for low potential future inflows came from NIWA's January 2022 - March 2022 climate outlook<sup>11</sup> which reported La Niña conditions continuing in the equatorial Pacific during Autumn. During La Niña rainfall in the lower and western South Island is reduced which usually results in below average inflows and storage levels in South Island catchments.
- 7.6 On top of current limited gas production, short term forward prices also factor in the risk gas outages scheduled in early 2022, at the Māui and Mangahewa, which could pose to thermal generation capacity. Maui is scheduled to undergo maintenance for 31 days from 7 May 2022, undergoing a full field outage from 14 May 2022. Concerns were that should these outages extend or return at less than full capacity thermal generation availability would be reduced when it is most needed right before the peak demand winter period.

---

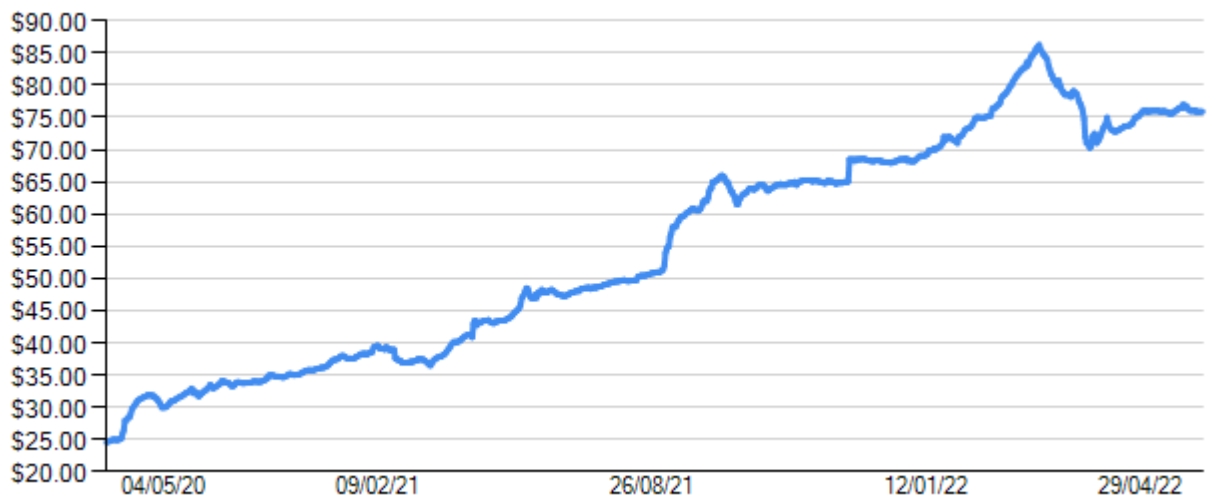
<sup>11</sup> [Seasonal climate outlook January 2022 - March 2022 | NIWA](#)

**Figure 17: Future Prices**



- 7.7 Long term forward prices compared to five years ago are noticeably higher. One of the factors that appears to play a role in the rise in forward prices is the increasing price of carbon which increases the costs of running thermal generation.
- 7.8 Figure 18 shows spot NZ carbon unit prices since the first ETS auction began in 2020, taken from CommTrade’s (a platform for buying and selling NZ ETS carbon credits owned by Jarden Securities Limited) website.
- 7.9 Overall carbon prices rose by 81.3 per cent in 2021. During the March ETS auction 10.5 million NZUs were sold, after the triggering of the cost containment reserve at \$70/unit, which added an additional 5.7 million NZUs.
- 7.10 How the price of carbon feeds into thermal generation costs is that approximately ~40 per cent of the price of carbon is added to the final cost of generation for a CCGT (thermal plant) and ~50-60 per cent for OCGTs (peaker plants) when run on gas and ~100 per cent when run on coal. For example, when carbon is at \$70/tonne an additional ~\$28/MWh to ~\$35/MWh would be added to the running cost of gas fired generation thermal generation and ~\$70/MWh would be added to coal fired generation.
- 7.11 The gas outlook for 2023 should be greatly improved compared to 2022, with gas production increases expected at multiple fields. Some of the expected drops in gas prices, however, are likely to be offset by continued increases in carbon prices with the cap for ETS auctions in 2023 set to \$78.40/tonne and \$87.81/tonne in 2024. It is not clear that generators are facing gas supply constraints this winter, so the increased supply may simply mean more production by Methanex.
- 7.12 With current renewable generation sources such as wind and solar too intermittent and capacity too low to provide the baseload and firming capacity that thermal generation currently does thermal plants are expected to continue to influence the market for a few more years to come.

**Figure 18: New Zealand Carbon Unit Spot Price<sup>12</sup>**



## 8 Deep Dive: Exploring NZ's Electricity Generation with Association Rule Mining

- 8.1 Included in this report is analysis which uses the machine learning algorithm Apriori to find relationships between different types of electricity generation in New Zealand. The Apriori algorithm identifies relationships or association rules between different datasets and evaluates the strength of these rules using statistical methods.
- 8.2 This study analysed half hourly generation data from North Island thermal stations at Huntly, Taranaki, and Whirinaki; North Island hydro generation; North Island wind farms in Bunnythorpe (which includes the Tararua region); and South Island hydro at the major lakes (Waitaki, Manapouri, Clutha) from 1 January 2016 – 21 December 2021, and temperature data from Auckland, Wellington and Christchurch.
- 8.3 The analysis found many association rules, many of which are commonly known, such as the economic relationship between gas prices and thermal generation. The Authority plans follow up work on the output from this modelling to investigate the different associations that exist in the spot market.

<sup>12</sup> <https://www.commtrade.co.nz/>

# Exploring NZ's Electricity Generation with Association Rule Mining

---

5 April 2022



## Executive summary

This report is an overview of how machine learning algorithms can be used to find the relationships between different types of electricity generation in New Zealand. The Authority is undertaking this work to explore the interactions between generators as they compete in the spot market. In doing so we hope to understand the competitive process of the spot market more thoroughly.

# Contents

Executive summary	ii
1 Introduction	1
2 Method	1
Apriori in a nutshell	1
Association Rules	2
Support	2
Confidence	3
Lift	3
Conviction	3
Apriori algorithm	4
3 Datasets	4
GIPs	4
Weather data	5
Gas price	6
4 Market basket analysis vs GIP generation	7
5 Results	12
Overview of results	13
Further discussion	16
6 Conclusions, future work	18
Future work	18
References	20

# 1 Introduction

- 1.1 We use the Apriori algorithm for association rule mining (Agrawal *et al.*, 1993; Agrawal & Ramakrishnan, 1994) to identify the underlying relationships between GIPs across the country. The algorithm is often used in market basket analysis (i.e. shopping habits of customers). However with some modifications, we repurpose the algorithm to explore the relationships between different types (thermal, wind, hydro) and quantities (MW) of electricity generation.
- 1.2 We build a model that considers wind generation in Tararua and Bunnythorpe; thermal generation at Huntly and Taranaki; hydro generation at Manapouri, Clutha, and Waitaki; average marginal gas price; and temperature in Auckland and Wellington. We use Python's **mlxtend** package (Raschka, 2018) to see how these variables relate to each other. We use half hourly data that covers the six year period between 2016 and 2021. We identify association rules which describe the relationships between different types of generation and Auckland/Wellington temperatures.

# 2 Method

- 2.1 Association rule learning is a rule-based machine learning method for discovering relationships between variables in large databases (Tan, 2019; Bramer, 2020). The Apriori algorithm is one example of the method. It identifies relationships or `\textit{association rules}` between different itemsets and evaluates the strength of these rules using statistical methods. It is intended as a proposition of the type

$$\text{antecedent} \Rightarrow \text{consequent}$$

for example,

$$\text{wind generation is low} \Rightarrow \text{Huntly runs at full capacity.}$$

- 2.2 This rule is not intended to be deterministic, and therefore a probability is associated with it. Such *probabilistic* association rules assign a probability to the consequence once the condition has been fulfilled.
- 2.3 The next section is an overview of the mechanisms behind Apriori. Although we use market basket analysis as an example, the same principles can be appropriated to study electricity generation after performing transformations on the dataset.

## Apriori in a nutshell

- 2.4 Apriori is an algorithm for frequent itemset mining and association rule learning for boolean datasets. The aim is to determine rules that highlight general trends in the dataset. To make the algorithm simple to understand, assume we have a list of items in a shopping basket for different customers.
- 2.5 Mathematically, there are two quantities to consider:

$$J = \{j_1, j_2, \dots, j_n\}, \tag{1}$$

a set of  $n$  different items. For the shopping example,  $J$  can show all available products in a shop: bread, milk, apples, etc. Secondly,

$$D = \{t_1, t_2, \dots, t_m\}, \tag{2}$$

is a set of  $m$  transactions. Each transaction  $t_m$  is equivalent to a customer receipt and consists of a subset of items from set  $J$  (i.e. a list of products from the shop). From these lists, we develop *rules* that describe the strong associations between variables.

2.6 A rule is defined as an implication of the form

$$X \Rightarrow Y \quad (3)$$

where

$$X, Y \subseteq J \quad (4)$$

and  $X$  and  $Y$  are disjoint itemsets

$$X \cap Y = \emptyset. \quad (5)$$

2.7  $X$  and  $Y$  are subsets of  $J$  (eq. (4)) but do not have common elements (eq. (5)). In the shopping example, this means  $X$  and  $Y$  are lists of one or more products in the shop but the lists cannot have the same products.

2.8 Here is an example: A customer's receipt  $D$  contains the products  $J = \{\text{bread, rice, pasta, milk, toothpaste}\}$ . We create arbitrary itemsets  $X = \{\text{bread, rice, milk}\}$  and  $Y = \{\text{pasta}\}$ . Hence one possible *rule* is  $\{\text{bread, rice}\} \Rightarrow \{\text{pasta}\}$ . This means a customer who buys bread and rice (*antecedents*) is likely to buy pasta (*consequent*).

## Association Rules

2.9 Now the question is, how do we measure the *strength* of a rule? In other words, is it reliable enough to be used as an association rule? The following metrics allow us to determine the most meaningful and significant rules:

- support (Agrawal *et al.*, 1993),
- confidence (Agrawal *et al.*, 1993),
- lift (Brin *et al.*, 1997),
- conviction (Brin *et al.*, 1997).

## Support

2.10 **Support** of itemset  $X$  is a proportion of total transactions that contain  $X$ ,

$$\text{support}(X) = \frac{\text{no. transactions where } X \text{ appears}}{\text{total no. transactions}}. \quad (6)$$

2.11 Similarly,

$$\text{support}(X \Rightarrow Y) = \frac{\text{no. transactions where both } X \text{ and } Y \text{ appear}}{\text{total no. transactions}}. \quad (7)$$

2.12 Support is an important measure because rules that have low support may occur simply by chance. Hence this quantity helps to identify the rules worth considering for further analysis.

2.13 If we only want to consider the itemsets that occur at least 50 times out of a total of 1000 transactions, we set the minimum support level of the algorithm to be 0.05. The idea is to focus computational resources to search for associations between frequently bought items while ignoring the infrequent ones.



- 2.14 If an itemset has a low support, we do not have enough information on the relationship between its items and hence no conclusions can be drawn from such a rule.

### Confidence

- 2.15 **Confidence** measures the reliability of the inference made by a rule. By definition, it is the conditional probability of the occurrence of consequent given the antecedent,

$$\text{confidence}(X \Rightarrow Y) = \frac{\text{no. transactions with both } X \text{ and } Y}{\text{total no. transactions with } X} = \frac{\text{support}(X \cup Y)}{\text{support}(X)}. \quad (8)$$

where  $X \cup Y$  is the union of itemsets  $X$  and  $Y$ . It means that every time a customer purchases  $X$ , there is a probability (eq. (8)) they will also purchase  $Y$ . It provides an estimate of the conditional probability of  $Y$  on  $X$ , i.e.

$$P(Y|X) = \frac{P(X \cup Y)}{P(X)} = \text{confidence}(X \Rightarrow Y). \quad (9)$$

- 2.16 A high confidence value does not necessarily give an accurate picture if the consequent appears frequently in the dataset. The confidence would be high regardless of the antecedents. If we know intuitively that two itemsets have weak association a high confidence value could be misleading. This is why we use the *lift* metric.

### Lift

- 2.17 **Lift** calculates a conditional probability of  $Y$  on  $X$  (like for confidence) but also considers the frequency of consequent  $Y$ ,

$$\text{lift}(X \Rightarrow Y) = \frac{\text{support}(X \cup Y)}{\text{support}(X) \cdot \text{support}(Y)} \begin{cases} = 1 \\ > 1 \\ < 1 \end{cases} \quad (10)$$

- 2.18 There are three possible cases:

- $\text{lift}(X \Rightarrow Y) = 1$  implies that the probability of  $X$  and  $Y$  being bought together is statistically insignificant. The purchases of these itemsets are independent of each other, and there is no statistically proven relationship.
- $\text{lift}(X \Rightarrow Y) > 1$  implies that  $X$  and  $Y$  are more likely to be bought together (positive dependence). The higher the lift value, the stronger the dependence. The itemsets are complements to each other.
- Similarly,  $\text{lift}(X \Rightarrow Y) < 1$  implies that  $X$  and  $Y$  are less likely to be bought together (negative dependence). The lower the lift value, the lower the dependence. The itemsets are substitutes of each other.

### Conviction

- 2.19 **Conviction** is another way to measure association:

$$\text{conviction}(X \Rightarrow Y) = \frac{1 - \text{support}(Y)}{1 - \text{confidence}(Y)} \begin{cases} = 1 \\ > 1 \end{cases} \quad (11)$$

- 2.20 The possible cases are:

- $\text{conviction}(X \Rightarrow Y) = 1$  means that  $X$  and  $Y$  are not associated.
- $\text{conviction}(X \Rightarrow Y) > 1$  implies a relationship exists.

2.21 The higher the conviction, the stronger the relationship. Generally, high confidence( $X \Rightarrow Y$ ) with low support( $Y$ ) produces high conviction. When confidence( $X \Rightarrow Y$ ) = 1, then conviction( $X \Rightarrow Y$ ) is infinity. This means the dataset shows  $X$  is *always* bought with  $Y$ , and confidence is 100 %.

2.22 Compared to lift,

$$\text{lift}(X \Rightarrow Y) = \text{lift}(Y \Rightarrow X), \quad (12)$$

conviction is a directed measurement,

$$\text{conviction}(X \Rightarrow Y) \neq \text{conviction}(Y \Rightarrow X). \quad (13)$$

and indicates a directional relationship between  $X$  and  $Y$ .

### Apriori algorithm

2.23 The Apriori algorithm is straightforward. It follows the steps:

- (i) Calculate support for itemsets of size 1.
- (ii) Apply the minimum support threshold and discard itemsets that do not meet the threshold.
- (iii) Move on to itemsets of size 2 and repeat steps above.
- (iv) Continue the same process until no additional itemsets satisfying the minimum threshold can be found.

2.24 Finally, arguably the most important step is using domain-specific knowledge to analyse the results. We use our judgement to decide whether the algorithm outputs a sensible association rule.

2.25 The algorithm only relies on numerical inputs and knows nothing about macroeconomics or NZ's electricity system. The inferences made by association rules do not necessarily imply causality -- instead they suggest strong co-occurrence relationships between items in the antecedent and consequent.

## 3 Datasets

3.1 All generation data in this report is from EA's database. As a first look, we will consider North Island thermal stations at Huntly and Taranaki; North Island wind farms in Bunnythorpe (which includes the Tararua region); and South Island hydro generation at the major lakes (Waitaki, Manapouri, Clutha).

3.2 We also consider the temperature at major population centres in Auckland and Wellington.

3.3 We take the total half hourly generation (MW) and average temperature<sup>1</sup> (Kelvin).

### GIPs

3.4 We use data that covers the period from 1 Jan 2016 to 31 Dec 2021. This consists of roughly  $6 \times 365 \times 48 = 105,000$  time points. We consider the GIPs

- H = Huntly thermal generation: total generation at nodes HLY2201 HLY1, HLY2, HLY3, HLY5, HLY6.

---

<sup>1</sup> Source: Weather Underground.

- T = Taranaki thermal generation: MKE1101 MKE1, JRD1101 JRD0, SFD2201 SFD21, SFD2201 SFD22, SFD2201 SPL0, KPA1101 KPA1, HWA1102 WAA1.
- B = Bunnythorpe wind generation: BPE0331 TWF0, WDV1101 TAP0, TWC2201 NZW0, TWC2201 TWF0, LTN0331 TWF0, LTN2201 TUR0.
- W = Waitaki hydro generation
  - Tekapo: TKA0111 TKA1, TKB2201 TKB1;
  - Ohau: OHA2201 OHA0, OHB2201 OHB0, OHC2201 OHC0;
  - Benmore: BEN2202 BEN0;
  - Aviemore: AVI2201 AVI0;
  - Waitaki: WTK0111 WTK0.
- C = Clutha hydro generation
  - Roxburgh: ROX1101 ROX0, ROX2201 ROX0;
  - Clyde: CYD2201 CYD0.
- M = Manapouri hydro generation: MAN2201 MAN0.

3.5 For brevity, we assign labels H, T, B, W, C, M to indicate the total generation of selected GIPs that are grouped by geography and fuel type.

### Weather data

3.6 Apparent temperature  $T_A$  measures what an observer feels (Steadman, 1984). It considers four environmental factors: wind, temperature, humidity, and radiation from the sun:

$$T_A = T_M + 0.348E - 0.70 \left( w_s - \frac{Q}{w_s + 10} \right) - 4.25 \quad (14)$$

with dry bulb (measured) temperature  $T_M$  (°C); wind speed  $w_s$  (m/s) at an elevation of 10m; net radiation absorbed per unit area of body surface  $Q$  (W/m<sup>2</sup>); and water vapour pressure (humidity),

$$E = \frac{\text{RH}}{100} 6.105 e^{17.27T_M/(237.7+T_M)} \quad (15)$$

in hPa where RH is relative humidity (%).

3.7 For simplicity, we assume that the observer is outdoors in the shade so  $Q = 0$ .

3.8 Instead of considering three environmental factors separately, the apparent temperature is a more compact way to estimate how hot or cold it feels in major population centres (Auckland, Wellington) and therefore the likelihood that many people will switch on air conditioners and heaters.

3.9 In summary, we consider two more parameters,

- X = Auckland apparent temperature
- Y = Wellington apparent temperature

3.10 Note it is more convenient to work in units of Kelvin (K) instead of degrees Celsius (°C). This means the temperatures  $T_K$  are always positive,

$$T_K = T_{\circ C} + 273 \tag{16}$$

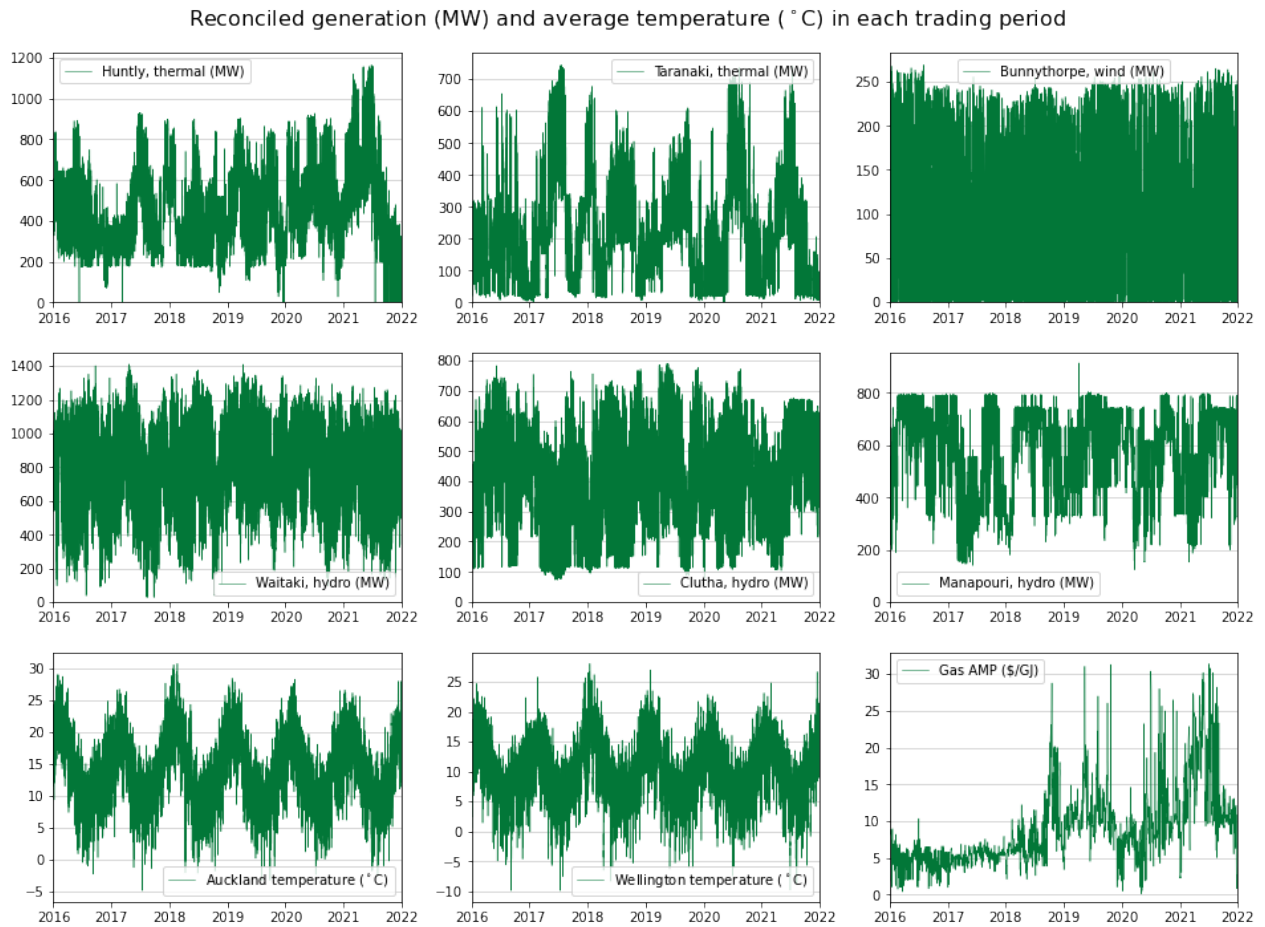
and will make it easier to perform discretisation in the next section.

## Gas price

3.11 We also consider average marginal gas price<sup>2</sup> (\$/GJ).

**Figure 1**

Raw data.



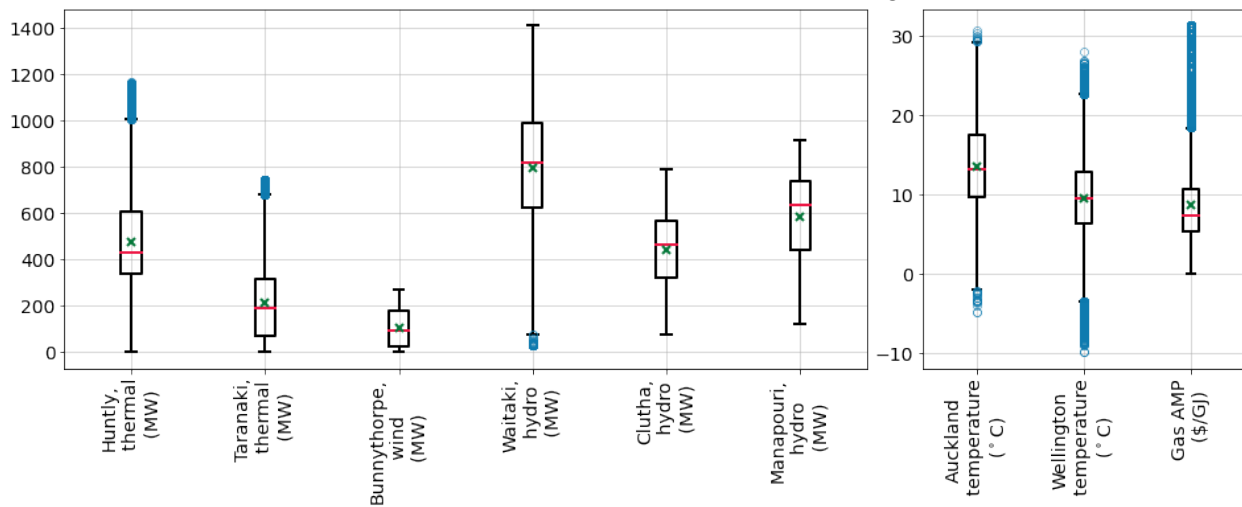
Source: Electricity Authority, Weather Underground, BGIX.

3.12 For visualisation purposes, Figure 1 and Figure 2 show the generation, temperature and gas price in 2016-2021.

<sup>2</sup> Source: BGIX.

**Figure 2**

Box plot of parameters. Red line is median, blue circles are outliers, green x marker is mean.

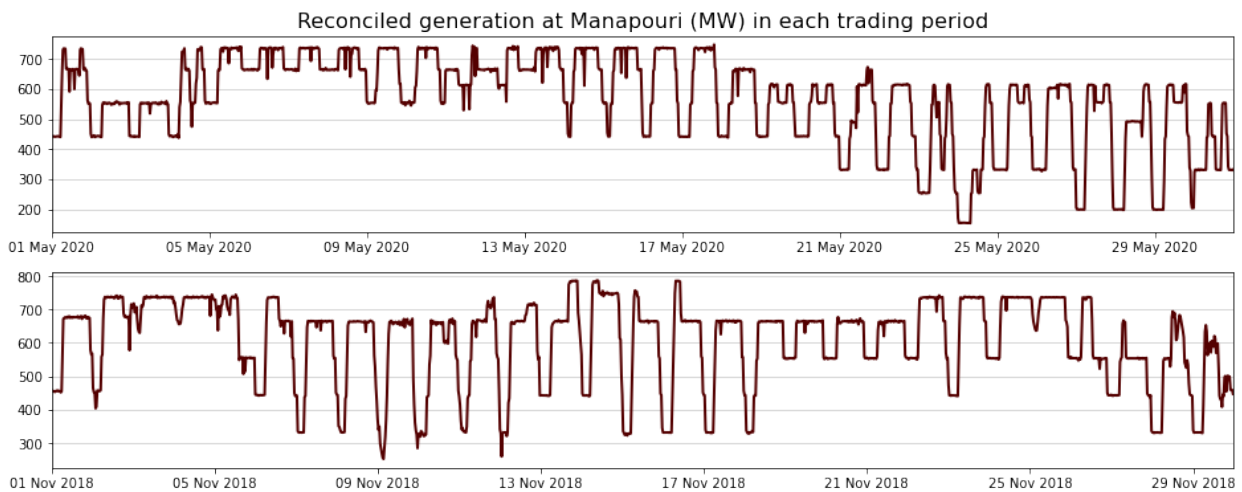


Source: Electricity Authority

3.13 Manapouri supplies electricity to the Tiwai aluminium smelter, so we expect generation to have flat peaks (Figure 3) during operating hours. Figure 1 also suggests that when Manapouri generates over 600 MW, it is running at full capacity.

**Figure 3**

Manapouri hydro generation in May 2020 and Nov 2018.



Source: Electricity Authority

## 4 Market basket analysis vs GIP generation

- 4.1 At this point, it's fair to ask, how can association rule mining be used to understand the relationship between different generation types?
- 4.2 The answer lies in *discretisation*, i.e. binning the continuous data, and *one-hot encoding*.
- 4.3 The first five rows of raw data are

**Table 1**

Date	Huntly (H)	Taranaki (T)	Bunnythorpe (B)	Waitaki (W)	Clutha (C)	Manapouri (M)	Auckland (X)	Wellington (Y)	Gas (G)
2016-01-01 00:00:00	510.83	31.42	164.21	948.59	253.27	255.46	290.44	288.13	8.8
2016-01-01 00:30:00	493.16	31.36	154.82	954.82	153.96	256.34	289.41	286.37	8.8
2016-01-01 01:00:00	504.50	31.04	132.71	872.26	156.89	255.01	289.10	286.68	8.8
2016-01-01 01:30:00	512.86	31.89	112.31	820.13	158.44	255.57	289.41	287.82	8.8
2016-01-01 02:00:00	500.45	31.28	91.78	800.56	154.42	255.23	288.79	287.01	8.8

Source: Electricity Authority

- 4.4 *Discretisation* is often used for handling continuous attributes. It groups the adjacent values of a continuous attribute into a finite number of intervals.
- 4.5 From inspecting Figure 1, it appears sensible to bin the thermal and hydro data in 100 MW intervals, wind in 50 MW intervals, Auckland and Wellington temperatures in 5 K intervals, and gas price in \$5 /GJ intervals.
- 4.6 For simplicity, we assign labels to the bins<sup>3</sup> in Table 2 and Table 3.

**Table 2**

Bins for generation and gas price

Thermal/hydro		Wind		Gas	
Bin (MW)	Label	Bin (MW)	Label	Bin (\$/GJ)	Label
0	0	0	0	0	0
(0, 100]	1	(0, 50]	1	(0, 5]	1
(100, 200]	2	(50, 100]	2	(5, 10]	2
...	...	...	...	...	...
(100(k-1), 100k]	k	(50(k-1), 50k]	k	(5(k-1), 5k]	k

**Table 3**

Bins for temperature

Bin (Kelvin, K)	Bin (Celsius, °C)	Label
(0, 5]		0
(5, 10]		1
...	...	...

<sup>3</sup> Interval notation (100, 200] means 100 MW < generation ≤ 200 MW.

Bin (Kelvin, K)	Bin (Celsius, °C)	Label
	(2, 7]	55
	(7, 12]	56
	(12, 17]	57
	(17, 22]	58
	(22, 27]	59
	(27, 32]	60
$(5k, 5(k + 1)]$	$(5k - 273, 5(k + 1) - 273]$	$k$

Source: Electricity Authority

and transform the raw data into Table 4

**Table 4**

Binned data

Date	H, binned	T, binned	B, binned	W, binned	C, binned	M, binned	X, binned	Y, binned	G, binned
2016-01-01 00:00:00	6	1	4	10	3	3	59	58	2
2016-01-01 00:30:00	5	1	4	10	2	3	58	58	2
2016-01-01 01:00:00	6	1	3	9	2	3	58	58	2
2016-01-01 01:30:00	6	1	3	9	2	3	58	58	2
2016-01-01 02:00:00	6	1	2	9	2	3	58	58	2

Source: Electricity Authority

where the labels represent the different sites of interest (H, T, B, W, C, M = generation; X, Y = weather; G = gas price) and the subscripts are bin names.

- 4.7 We can think of the table as containing categorical features. Although the bin labels are assigned numerical values (0, 1, 2, ...,  $k$ ), there is no specific order of preference and all labels should have equal importance.
- 4.8 But using numerical values adds bias to the model -- the temperature bins have much larger label values than the other data. The model would give greater weight (higher preference) to the temperature variables.
- 4.9 This calls for a binary True/False or 1/0 style of categorising, namely the *one-hot encoding* technique in Table 5.

**Table 5**

One-hot encoding of binned data (Table 4).

Date	H <sub>0</sub>	H <sub>1</sub>	H <sub>2</sub>	H <sub>3</sub>	H <sub>4</sub>	H <sub>5</sub>	H <sub>6</sub>	H <sub>7</sub>	...	T <sub>0</sub>	T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>	T <sub>5</sub>	T <sub>6</sub>	...	B <sub>0</sub>	...
2016-01-01 00:00:00	0	0	0	0	0	0	1	0	...	0	1	0	0	0	0	0	...	0	...

Date	H <sub>0</sub>	H <sub>1</sub>	H <sub>2</sub>	H <sub>3</sub>	H <sub>4</sub>	H <sub>5</sub>	H <sub>6</sub>	H <sub>7</sub>	...	T <sub>0</sub>	T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>	T <sub>5</sub>	T <sub>6</sub>	...	B <sub>0</sub>	...
2016-01-01 00:30:00	0	0	0	0	0	1	0	0	...	0	1	0	0	0	0	0	...	0	...
2016-01-01 01:00:00	0	0	0	0	0	0	1	0	...	0	1	0	0	0	0	0	...	0	...
2016-01-01 01:30:00	0	0	0	0	0	0	1	0	...	0	1	0	0	0	0	0	...	0	...
2016-01-01 02:00:00	0	0	0	0	0	0	1	0	...	0	1	0	0	0	0	0	...	0	...

Source: Electricity Authority

where the labels represent the different sites of interest (generation or weather) and subscripts are bin names.

- 4.10 The binary format shows that each row and column corresponds to a transaction and item respectively. In our example, the transaction is analogous to the types and quantities of generation and temperature of a specific trading period.
- 4.11 An item is treated as a binary variable whose value is one if the item is present in a transaction and zero otherwise. Because the presence of an item in a transaction is often considered more important than its absence, it is an *asymmetric* binary variable.
- 4.12 We add some more columns to represent time features,

$$q = \begin{cases} 1, & \text{weekend} \\ 0, & \text{not weekend} \end{cases} \quad (17)$$

$$r = \begin{cases} 1, & \text{NZ public holiday} \\ 0, & \text{not NZ public holiday} \end{cases} \quad (18)$$

- 4.13 Instead of defining time by trading period (1 to 48), we let  $t_k$  represent four-hourly periods where

$$k = \begin{cases} 0, & 00:00 \text{ to } 03:59 \\ 1, & 04:00 \text{ to } 07:59 \\ 2, & 08:00 \text{ to } 11:59 \\ 3, & 12:00 \text{ to } 15:59 \\ 4, & 16:00 \text{ to } 19:59 \\ 5, & 20:00 \text{ to } 23:59 \end{cases} \quad (19)$$

so that each bin contains eight consecutive trading periods. This reduces the number of table columns as well as giving an idea of the peak morning and evening demand.

- 4.14 Finally, we define the year  $y_k$  where

$$k = \begin{cases} 16, & 1 \text{ Jan } 2016 \text{ to } 31 \text{ Dec } 2016 \\ 17, & 1 \text{ Jan } 2017 \text{ to } 31 \text{ Dec } 2017 \\ 18, & 1 \text{ Jan } 2018 \text{ to } 31 \text{ Dec } 2018 \\ 19, & 1 \text{ Jan } 2019 \text{ to } 31 \text{ Dec } 2019 \\ 20, & 1 \text{ Jan } 2020 \text{ to } 31 \text{ Dec } 2020 \\ 21, & 1 \text{ Jan } 2021 \text{ to } 31 \text{ Dec } 2021 \end{cases} \quad (20)$$

- 4.15 The truncated Table 5 shows that in the first row (1 Jan 2016, midnight), Huntly units generated 500-600 MW ( $H_6 = 1$ ) and Taranaki generated less than 100 MW ( $T_1 = 1$ ). It is not shown here but Bunnythorpe generated 150-200 MW ( $B_4 = 1$ ), Waitaki generated 900-1000 MW ( $W_{10} = 1$ ), and Clutha and Manapouri each generated 200-300 MW ( $C_3 =$



$M_3 = 1$ ). Auckland temperature was 22 to 27 °C ( $X_{59} = 1$ ) and Wellington was 17 to 22 °C ( $Y_{58} = 1$ ). Gas price on 1 Jan was \$5-10 /GJ ( $G_2 = 1$ ). That day was a public holiday ( $r = 1$ ) and weekday ( $q = 0$ ).

- 4.16 The table size depends on the number of degrees of freedom. In this case, the table has in total  
 [(22 thermal) + (7 wind) + (32 hydro) + (17 temperature) + (7 gas) + (14 time)] columns  
 x (6 years) x (365 days) x (48 trading periods) ~ 10 million cells.
- 4.17 This is important to consider, as large datasets take more computational time and resources to process. As a rule of thumb, "large" datasets contain >100,000 cells.
- 4.18 The table is analogous to the market basket example. Instead of studying the products that customers buy, we consider the location, type, and amount of generation (and temperature range). For example,  $H_6$  means that Huntly thermal generation was 500-600 MW.  $H_6$  is analogous to one specific item in a customer's shopping basket.
- 4.19 One major difference is table structure. In this dataset, for every trading period there is always one True value for each of: H, T, B, W, C, M, X, Y, G,  $t$ ,  $y$ . The probability that it will appear in an itemset of antecedents or consequents depends on the number of bins (Table 6).

**Table 6**

No. bins indicates probability  $P$  that bin  $k$  of parameter  $x$  will appear in association rule.

Parameter ( $x$ )	No. bins ( $k$ )	$P(x_k) \sim 1/k$
H (Huntly)	13	0.08
T (Taranaki)	9	0.11
B (Bunnythorpe)	7	0.14
W (Waitaki)	15	0.07
C (Clutha)	8	0.125
M (Manapouri)	9	0.11
X (Auckland)	8	0.125
Y (Wellington)	9	0.11
G (gas price)	7	0.14
$q$ (weekend)	-	$2/7 \sim 0.28$
$r$ (public holiday)	-	$12/365 \sim 0.03$
$t$ (time period)	6	0.17
$y$ (year)	6	0.17

Source: Electricity Authority

- 4.20 For example, if all variables were *independent* then the rule

$$[H_6, B_4, q, X_{59}] \Rightarrow [C_3]$$

would have the following properties. The *antecedent support* is the probability of all items being grouped together in the antecedent  $[H_6, B_4, q, X_{59}]$ ,

$$P(H_6) \times P(B_4) \times P(q) \times P(X_{59}) \simeq 4 \times 10^{-4}. \quad (21)$$

Similarly the *consequent support* is

$$P(C_3) = 0.125. \quad (22)$$

4.21 Hence the *support* of the rule is

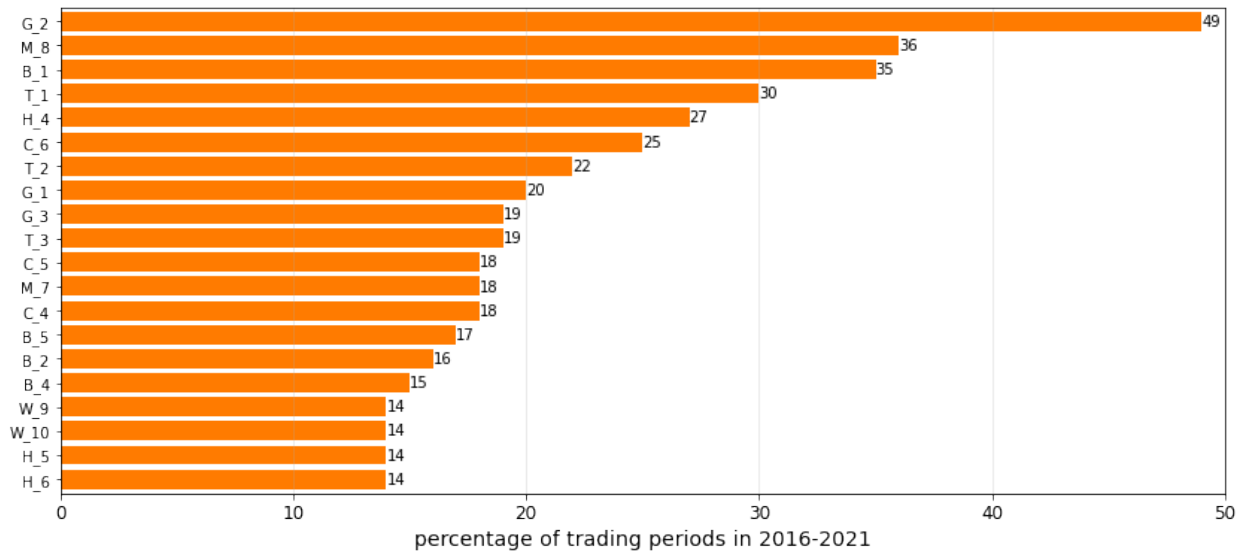
$$4 \times 10^{-4} \times 0.125 \approx 5 \times 10^{-5}, \quad (23)$$

i.e. a 0.005 % chance that the association rule would occur in a given trading period. If these variables were *dependent*, the rule's support measure would be higher.

4.22 Figure 4 shows a histogram of the one-hot encoded data. The highest number of occurrences is  $G_2$  (49 %). This means that gas price was \$5-10 /GJ for about 35 months of the six year period. The second highest was  $M_8$  (36 %), where Manapouri generated 700-800 MW for 38,000 trading periods or 26 months. In comparison, there were roughly 22 months where Taranaki ( $T_1$ ) generated under 100 MW.

**Figure 4**

Fraction of the total number of trading periods in 1 Jan 2016 to 31 Dec 2021 period. Top 20 instances shown.



Source: Electricity Authority

## 5 Results

5.1 The Apriori algorithm requires us to define the minimum support. As mentioned above, this measure can be exploited by the algorithm to efficiently search for the most relevant association rules.

5.2 Because of the inherent nature of the dataset, most table elements are zero -- by definition a *sparse matrix*. This means the minimum support level should be small. As a first test, we try with 0.01.

5.3 For a table with  $d = 99$  columns, there are up to  $2^d \sim 6 \times 10^{29}$  possible itemsets<sup>4</sup>. The total number of possible association rules is (Tan, 2019)

<sup>4</sup> To give a sense of scale, there are roughly  $5 \times 10^{30}$  bacterial cells on Earth.

$$\sum_{\ell=1}^{d-1} \left[ \binom{d}{\ell} \times \sum_{j=1}^{d-\ell} \binom{d-\ell}{j} \right] = 3^d - 2^{d+1} + 1 \sim 10^{47}, \quad (2421)$$

i.e. 100 quattuordecillion<sup>5</sup>, with binomial coefficients

$$\binom{d}{\ell} = \frac{d!}{\ell! (d-\ell)!} \quad (25)$$

- 5.4 In our dataset, for every trading period there is always one True value for each of: H, T, B, W, C, M, X, Y, G,  $t$ ,  $y$ . This would decrease the total number of possible itemsets ( $10^{29}$ ) and association rules ( $10^{47}$ ) by several orders of magnitude.
- 5.5 The following table of results is organised according to decreasing confidence. To reduce the number of rows, we filter out the most relevant results using the following conditions:
- Ignore all association rules where both antecedents and consequents contain any uncontrolled variables, i.e. wind generation (B), temperature (X, Y), and time features ( $q$ ,  $r$ ,  $t$ ,  $y$ ).
  - Ignore all rules where consequents contain only uncontrolled variables.
  - Only consider trustworthy rules where confidence is greater than 0.6.
- 5.6 Returning to Figure 4, because the gas prices were \$5-10 /GJ for a large number of days ( $G_2$ ), we should not be surprised if many of the association rules contained  $G_2$ . This also applies to  $M_8$  (Manapouri, 700-800 MW) and  $B_1$  (Bunnythorpe, < 100 MW). The full list of results (available upon request) shows that a large proportion of rules contain these parameters.
- 5.7 Because Manapouri's generation is routed to the Tiwai smelter, we can assume that their daily demand will be approximately the same regardless of how much the other hydro and thermal stations are generating. As a result, any association rules containing Manapouri ( $M_k$ ) as a consequent should be taken with a grain of salt.

## Overview of results

- 5.8 The results table contains over 400 association rules so we discuss a subset (Table 7), ordered by decreasing confidence.

**Table 7**

Subset of results.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	conviction
0	[H <sub>3</sub> , M <sub>8</sub> , y <sub>21</sub> ]	[T <sub>1</sub> ]	0.0132	0.3039	0.0132	<b>0.9971</b>	3.2806	238.9253
1	[H <sub>3</sub> , y <sub>21</sub> , G <sub>3</sub> ]	[T <sub>1</sub> ]	0.0128	0.3039	0.0126	<b>0.9835</b>	3.2358	42.1121
3	[y <sub>18</sub> , X <sub>60</sub> ]	[G <sub>2</sub> ]	0.0137	0.4980	0.0126	<b>0.9232</b>	1.8541	6.5404
11	[y <sub>18</sub> , X <sub>59</sub> , B <sub>1</sub> ]	[G <sub>2</sub> ]	0.0148	0.4980	0.0126	<b>0.8505</b>	1.7079	3.3572
21	[X <sub>59</sub> , H <sub>3</sub> , Y <sub>58</sub> ]	[T <sub>1</sub> ]	0.0134	0.3039	0.0110	<b>0.8179</b>	2.6911	3.8229
33	[G <sub>1</sub> , y <sub>16</sub> , H <sub>4</sub> , C <sub>6</sub> ]	[M <sub>8</sub> ]	0.0131	0.3618	0.0104	<b>0.7972</b>	2.2033	3.1468

<sup>5</sup> There are  $5 \times 10^{46}$  possible legal chess positions.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	conviction
41	[q, H <sub>3</sub> , M <sub>8</sub> ]	[T <sub>1</sub> ]	0.0162	0.3039	0.0127	<b>0.7862</b>	2.5868	3.2561
73	[y <sub>16</sub> , t <sub>3</sub> , C <sub>6</sub> ]	[M <sub>8</sub> ]	0.0143	0.3618	0.0106	<b>0.7426</b>	2.0523	2.4792
77	[C <sub>7</sub> , B <sub>5</sub> ]	[M <sub>8</sub> ]	0.0280	0.3618	0.0207	<b>0.7390</b>	2.0424	2.4451
159	[W <sub>9</sub> , C <sub>7</sub> ]	[M <sub>8</sub> ]	0.0206	0.3618	0.0141	<b>0.6849</b>	1.8929	2.0253

Source: Electricity Authority

### 5.9 Row 0 shows the association rule

$$[H_3, M_8, y_{21}] \Rightarrow [T_1].$$

The antecedents are H<sub>3</sub> (Huntly generated 200-300 MW), M<sub>8</sub> (Manapouri generated 700-800 MW) and y<sub>21</sub> (this rule applies to trading periods in 2021). The consequent is T<sub>1</sub> (Taranaki generated less than 100 MW).

5.10 The lift is 3.3 which suggests a strong positive dependence between antecedents and consequent. If Huntly and Manapouri generated 200-300 MW and 700-800 MW respectively during a trading period in 2021, then it was 3.3 times as likely that Taranaki would have generated under 100 MW (compared to a trading period where Huntly and Manapouri didn't generate this amount).

5.11 A confidence value of 0.997 means that Taranaki would have almost certainly generated less than 100 MW during a trading period in 2021 if Huntly and Manapouri had generated 200-300 MW and 700-800 MW respectively.

5.12 The extremely high conviction (240) confirms the existence of a relationship between the antecedents and consequent.

5.13 From inspecting Figure 1 and Figure 3, it also seems likely that Manapouri would have run at or near full capacity. Figure 1 suggests that this rule could more likely have occurred during the second half of 2021 - this was when Huntly generation had a 10-90 % percentile range of 200-670 MW. In the first half of the year, Huntly had a range of 490-980 MW.

5.14 Similarly, row 1 shows the association rule

$$[H_3, y_{21}, G_3] \Rightarrow [T_1].$$

5.15 If Huntly generated 200-300 MW and gas price was \$10-15 /GJ in 2021, the confidence of 0.98 suggests that there was a 98 % chance that Taranaki would have generated less than 100 MW.

5.16 Like the previous rule, a lift of 3.2 suggests a strong positive dependence between antecedents and consequent. The antecedent and consequent support of this association rule are 0.0128 and 0.3039 respectively. They show the probability that all items are grouped together in the antecedent or consequent.

5.17 To compare, if these items were completely independent then the probability of them being in the antecedent itemlist would have antecedent support:

$$P(H_3) \times P(y_{21}) \times P(G_3) \approx 0.002 < 0.0128. \quad (26)$$

Similarly the consequent support would be

$$P(T_1) \approx 0.111 < 0.3039. \quad (27)$$

Hence if all items were independent, the support of the rule would be

$$0.002 \times 0.111 \approx 2 \times 10^{-4}, \quad (28)$$

i.e. a 0.02 % chance that the association rule would occur in a given trading period. If these variables were *dependent*, the support measure would be 0.013 or 1.3 %, according to Table 7.

5.18 Row 41

$$[q, H_3, M_8] \Rightarrow [T_1]$$

suggests that if Huntly and Manapouri generated 200-300 MW ( $H_3$ ) and 700-800 MW ( $M_8$ ) during a weekend ( $q = 1$ ) trading period, there was a 79 % chance that Taranaki would have generated under 100 MW.

5.19 Rows 3, 11, and 21 show association rules when Auckland is 22-27 °C ( $X_{59}$ ) and 27-32 °C ( $X_{60}$ ).

5.20 Row 3

$$[y_{18}, X_{60}] \Rightarrow [G_2]$$

suggests that when Auckland was 27-32 °C ( $X_{60}$ ) in 2018, there was a 92 % chance (confidence = 0.92) that gas price would be \$5-10 /GJ ( $G_2$ ). From inspecting Figure 1, it's likely that this would have occurred during the summer heatwave in Jan-Feb 2018. The lift of 1.85 also indicates a strong positive dependence between the antecedents and consequent.

5.21 In comparison, row 11

$$[y_{18}, X_{59}, B_1] \Rightarrow [G_2]$$

suggests that when there was little wind generation at Bunnythorpe ( $B_1$ , less than 50 MW) and Auckland was 22-27 °C ( $X_{59}$ ) in 2018, there was an 85 % chance (confidence = 0.85) that gas price would be \$5-10 /GJ ( $G_2$ ).

5.22 Row 21

$$[X_{59}, H_3, Y_{58}] \Rightarrow [T_1]$$

shows that when Huntly generated 200-300 MW and Auckland and Wellington were 22-27 °C and 17-22 °C respectively, there was an 82 % chance (confidence = 0.82) that Taranaki would have generated less than 100 MW.

5.23 These rules may seem counterintuitive - in hot weather we would expect higher generation and higher gas prices due to increased demand from air conditioners. However the *absence* of certain items in the itemsets (like strong hydro generation at Waitaki/Clutha) means that such events *may* have occurred. The presence of an item in the itemset is considered more important than its absence, so we cannot say for certain Waitaki or Clutha's level of generation.

5.24 Rows 33, 73, 77, and 159 show strong hydro generation at Clutha ( $C_6$  and  $C_7$ ) in the antecedent itemset with Manapouri generating 700-800 MW as a consequent.

5.25 Row 33

$$[G_1, y_{16}, H_4, C_6] \Rightarrow [M_8]$$

shows that in 2016, if gas price was under \$5 /GJ, Huntly generated 300-400 MW, and Clutha generated 500-600 MW, then there was an 80 % chance (confidence = 0.8) that

Manapouri generated 700-800 MW. The lift indicates there is a strong positive dependence between the antecedents and consequent.

5.26 To compare, row 73

$$[y_{16}, t_3, C_6] \Rightarrow [M_8]$$

shows that during a trading period between noon and 3pm ( $t_3$ ), if Clutha generated 500-600 MW ( $C_6$ ), there was a 74 % chance that Manapouri would generate 700-800 MW during their operating hours.

5.27 Similarly, row 77

$$[C_7, B_5] \Rightarrow [M_8]$$

suggests that given strong wind generation at Bunnythorpe ( $B_5$ , 200-250 MW) and strong hydro at Clutha ( $C_7$ , 500-600 MW), there was a 74 % chance that Manapouri would generate 700-800 MW in the same trading period.

5.28 Row 159

$$[W_9, C_7] \Rightarrow [M_8]$$

shows the relationship between hydro generation at Waitaki, Clutha and Manapouri. Because lift is a non-directed measurement, we can say that

$$\text{lift}([W_9, C_7] \Rightarrow [M_8]) = \text{lift}([M_8] \Rightarrow [W_9, C_7]) = 1.9.$$

5.29 The lift for row 159's rule is 1.9, so there is a strong positive dependence between Waitaki/Clutha generation and Manapouri. If Manapouri generated 700-800 MW, it was 1.9 times as likely that Waitaki ( $W_9$ ) and Clutha ( $C_7$ ) generated 800-900 MW and 600-700 MW respectively (compared to a trading period where Manapouri didn't generate 700-800 MW).

## Further discussion

5.30 The following tables show the top 10 rules ordered by decreasing support, confidence, lift, and conviction.

**Table 8**

Top 10 rules, decreasing support.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	conviction
304	[ $y_{17}$ ]	[ $G_2$ ]	0.1681	0.4980	<b>0.1077</b>	0.6404	1.2862	1.3963
366	[ $y_{18}$ ]	[ $G_2$ ]	0.1659	0.4980	<b>0.1044</b>	0.6295	1.2643	1.3552
257	[ $C_7$ ]	[ $M_8$ ]	0.1377	0.3618	<b>0.0896</b>	0.6506	1.7980	1.8263
466	[ $H_3$ ]	[ $T_1$ ]	0.1245	0.3039	<b>0.0765</b>	0.6145	2.0219	1.8058
505	[ $G_1, y_{16}$ ]	[ $M_8$ ]	0.0841	0.3618	<b>0.0512</b>	0.6091	1.6833	1.6325
401	[ $T_1, C_6$ ]	[ $M_8$ ]	0.0807	0.3618	<b>0.0505</b>	0.6251	1.7277	1.7023
92	[ $y_{17}, B_1$ ]	[ $G_2$ ]	0.0666	0.4980	<b>0.0484</b>	0.7271	1.4602	1.8397
378	[ $H_4, C_6$ ]	[ $M_8$ ]	0.0748	0.3618	<b>0.0470</b>	0.6278	1.7351	1.7147
212	[ $y_{18}, M_8$ ]	[ $G_2$ ]	0.0671	0.4980	<b>0.0445</b>	0.6632	1.3319	1.4907
456	[ $Y_{57}, y_{16}$ ]	[ $M_8$ ]	0.0709	0.3618	<b>0.0437</b>	0.6162	1.7031	1.6629

Source: Electricity Authority

**Table 9**

Top 10 rules, decreasing confidence.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	conviction
0	[H <sub>3</sub> , M <sub>8</sub> , y <sub>21</sub> ]	[T <sub>1</sub> ]	0.0132	0.3039	0.0132	<b>0.9971</b>	3.2806	238.9253
1	[H <sub>3</sub> , y <sub>21</sub> , G <sub>3</sub> ]	[T <sub>1</sub> ]	0.0128	0.3039	0.0126	<b>0.9835</b>	3.2358	42.1121
2	[H <sub>3</sub> , y <sub>21</sub> ]	[T <sub>1</sub> ]	0.0235	0.3039	0.0229	<b>0.9738</b>	3.2041	26.6028
3	[y <sub>18</sub> , X <sub>60</sub> ]	[G <sub>2</sub> ]	0.0137	0.4980	0.0126	<b>0.9232</b>	1.8541	6.5404
4	[X <sub>56</sub> , y <sub>17</sub> ]	[G <sub>2</sub> ]	0.0118	0.4980	0.0108	<b>0.9176</b>	1.8428	6.0941
6	[y <sub>17</sub> , H <sub>7</sub> ]	[G <sub>2</sub> ]	0.0167	0.4980	0.0149	<b>0.8956</b>	1.7986	4.8097
7	[H <sub>4</sub> , M <sub>8</sub> , y <sub>21</sub> ]	[T <sub>1</sub> ]	0.0143	0.3039	0.0125	<b>0.8763</b>	2.8831	5.6253
8	[r, G <sub>1</sub> ]	[T <sub>1</sub> ]	0.0129	0.3039	0.0113	<b>0.8757</b>	2.8814	5.6019
9	[y <sub>18</sub> , Y <sub>59</sub> ]	[G <sub>2</sub> ]	0.0183	0.4980	0.0160	<b>0.8709</b>	1.7489	3.8878
11	[y <sub>18</sub> , X <sub>59</sub> , B <sub>1</sub> ]	[G <sub>2</sub> ]	0.0148	0.4980	0.0126	<b>0.8505</b>	1.7079	3.3572

Source: Electricity Authority

**Table 10**

Top 10 rules, decreasing lift.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	conviction
172	[H <sub>4</sub> , y <sub>16</sub> , X <sub>58</sub> ]	[G <sub>1</sub> ]	0.0213	0.2036	0.0145	0.6801	<b>3.3403</b>	2.4895
200	[H <sub>4</sub> , Y <sub>57</sub> , y <sub>16</sub> , M <sub>8</sub> ]	[G <sub>1</sub> ]	0.0151	0.2036	0.0101	0.6694	<b>3.2880</b>	2.4092
0	[H <sub>3</sub> , M <sub>8</sub> , y <sub>21</sub> ]	[T <sub>1</sub> ]	0.0132	0.3039	0.0132	0.9971	<b>3.2806</b>	238.9253
1	[H <sub>3</sub> , y <sub>21</sub> , G <sub>3</sub> ]	[T <sub>1</sub> ]	0.0128	0.3039	0.0126	0.9835	<b>3.2358</b>	42.1121
2	[H <sub>3</sub> , y <sub>21</sub> ]	[T <sub>1</sub> ]	0.0235	0.3039	0.0229	0.9738	<b>3.2041</b>	26.6028
509	[H <sub>4</sub> , y <sub>21</sub> ]	[G <sub>3</sub> ]	0.0251	0.1936	0.0153	0.6087	<b>3.1437</b>	2.0610
512	[H <sub>4</sub> , T <sub>1</sub> , y <sub>21</sub> ]	[G <sub>3</sub> ]	0.0209	0.1936	0.0127	0.6079	<b>3.1395</b>	2.0567
319	[H <sub>4</sub> , X <sub>59</sub> , y <sub>17</sub> ]	[G <sub>1</sub> ]	0.0191	0.2036	0.0122	0.6389	<b>3.1380</b>	2.2054
484	[H <sub>4</sub> , Y <sub>57</sub> , y <sub>16</sub> ]	[G <sub>1</sub> ]	0.0272	0.2036	0.0167	0.6125	<b>3.0085</b>	2.0554
500	[X <sub>59</sub> , y <sub>16</sub> , M <sub>8</sub> ]	[G <sub>1</sub> ]	0.0240	0.2036	0.0146	0.6096	<b>2.9939</b>	2.0397

Source: Electricity Authority

**Table 11**

Top 10 rules, decreasing conviction

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	conviction
0	[H <sub>3</sub> , M <sub>8</sub> , y <sub>21</sub> ]	[T <sub>1</sub> ]	0.0132	0.3039	0.0132	0.9971	3.2806	<b>238.9253</b>
1	[H <sub>3</sub> , y <sub>21</sub> , G <sub>3</sub> ]	[T <sub>1</sub> ]	0.0128	0.3039	0.0126	0.9835	3.2358	<b>42.1121</b>
2	[H <sub>3</sub> , y <sub>21</sub> ]	[T <sub>1</sub> ]	0.0235	0.3039	0.0229	0.9738	3.2041	<b>26.6028</b>
3	[y <sub>18</sub> , X <sub>60</sub> ]	[G <sub>2</sub> ]	0.0137	0.4980	0.0126	0.9232	1.8541	<b>6.5404</b>

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	conviction
4	[X <sub>56</sub> , Y <sub>17</sub> ]	[G <sub>2</sub> ]	0.0118	0.4980	0.0108	0.9176	1.8428	<b>6.0941</b>
7	[H <sub>4</sub> , M <sub>8</sub> , Y <sub>21</sub> ]	[T <sub>1</sub> ]	0.0143	0.3039	0.0125	0.8763	2.8831	<b>5.6253</b>
8	[r, G <sub>1</sub> ]	[T <sub>1</sub> ]	0.0129	0.3039	0.0113	0.8757	2.8814	<b>5.6019</b>
6	[Y <sub>17</sub> , H <sub>7</sub> ]	[G <sub>2</sub> ]	0.0167	0.4980	0.0149	0.8956	1.7986	<b>4.8097</b>
14	[H <sub>4</sub> , Y <sub>21</sub> ]	[T <sub>1</sub> ]	0.0251	0.3039	0.0209	0.8316	2.7362	<b>4.1336</b>
15	[H <sub>3</sub> , M <sub>8</sub> , Y <sub>58</sub> ]	[T <sub>1</sub> ]	0.0122	0.3039	0.0102	0.8315	2.7358	<b>4.1309</b>

Source: Electricity Authority

- 5.31 We now take the top 100 rules ordered by support, confidence, lift and conviction, then find a common set of rules from each subset (Table 12).

**Table 12**

Intersection of top 100 rules ordered by support, confidence, lift and conviction.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	conviction
2	[H <sub>3</sub> , Y <sub>21</sub> ]	[T <sub>1</sub> ]	0.0235	0.3039	0.0229	<b>0.9738</b>	3.2041	26.6028
14	[H <sub>4</sub> , Y <sub>21</sub> ]	[T <sub>1</sub> ]	0.0251	0.3039	0.0209	<b>0.8316</b>	2.7362	4.1336
24	[H <sub>3</sub> , Y <sub>58</sub> ]	[T <sub>1</sub> ]	0.0295	0.3039	0.0239	<b>0.8105</b>	2.6667	3.6732
36	[X <sub>59</sub> , H <sub>3</sub> ]	[T <sub>1</sub> ]	0.0261	0.3039	0.0207	<b>0.7947</b>	2.6147	3.3902
44	[G <sub>1</sub> , Y <sub>16</sub> , C <sub>6</sub> ]	[M <sub>8</sub> ]	0.0307	0.3618	0.0239	<b>0.7799</b>	2.1554	2.8991

Source: Electricity Authority

- 5.32 These association rules suggest that there is a strong relationship (positive dependence) between thermal generation, year, and temperature in the major population centres. There also appears to be a relationship between gas price and hydro generation in 2016.

## 6 Conclusions, future work

- 6.1 This report is a first attempt at using the association rule mining algorithm Apriori to find relationships between GIPs (NI thermal, SI hydro, Bunnythorpe wind generation), gas price, and apparent temperature in major population centres.
- 6.2 There is evidence of causality due to strong co-occurrent relationships between Auckland/Wellington temperatures, gas prices, hydro and thermal generation.
- 6.3 There is a strong positive dependence between Waitaki, Clutha and Manapouri hydro generation, as well as between Huntly and Taranaki thermal units.

### Future work

- 6.4 The following is a list of ideas and avenues for further investigation.
- Instead of focussing on different sites and generation types, it would be interesting to explore the relationships between different gentailers (Mercury, Meridian, Contact, Genesis).



- Find other ways to bin the data (e.g. different sized bins for gas price). Explore different constraints for the algorithm like minimum support level and minimum confidence.
- Investigate the relationships of predispach prices and offer stacks at different nodes.

## References

- R. Agrawal *et al.*, Mining association rules between sets of items in large databases, in *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, (Association for Computing Machinery, New York, 1993), pp. 207-216.
- R. Agrawal and S. Ramakrishnan, Fast algorithms for mining association rules, in *Proceedings of the 20th International Conference on Very Large Data Bases*, (1994), pp. 487-499.
- M. Bramer, *Principles of Data Mining*, (Springer, 2020), 4th ed.
- S. Brin *et al.*, Dynamic itemset counting and implication rules for market basket data, in *Proceedings of the 1997 ACM SIGMOD International Conference on Management of Data*, (1997), pp. 255-264.
- S. Raschka, MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack, *J. Open Source Softw.*, **3**(24), 638 (2018).
- R. G. Steadman, A Universal Scale of Apparent Temperature, *J. Clim. Appl. Meteorol.* **23**(12):1674-1687 (1984).
- P.-N. Tan *et al.*, *Introduction to Data Mining*, (Pearson, 2019), 2nd ed.