Initial Review of the NZECS Residual Supply Mix Methodology

Prepared by Energy Link

for

New Zealand Energy Certificate System

December 2019

Quality Assurance Information

Name	NZECS RSM method review Dec-19 v2.docx
File reference	E-NZECS-1298
Issue Status	Final
Issue Date	13 December 2019
Client	New Zealand Energy Certificate System

Definitions

The following abbreviations and acronyms are used in this report.

Code	Electricity Industry Participation Code
Dx	Distribution
EA	Electricity Authority
EG	Embedded generation. Also includes distributed generation.
EMS	Energy Market Services, a division of Transpower
GIP	Grid injection point
GXP	Grid exit point
Location factor	The ratio of two spot prices
OATIS	Open Access Transmission Information System for gas
RSF	Residual supply factor
RSM	Residual supply mix
RTP	Real-time pricing
Tx	Transmission

Contents

DEFINITIONS II
1 INTRODUCTION 1
2 SUMMARY 1
3 THE RSM METHODOLOGY 2
3.1 TOP-DOWN APPROACH 5
3.2 BOTTOM-UP APPROACH 6
3.3 RECOMMENDATION 8
4 30-MINUTE RECONCILIATION 8
5 SUPPLIER-SPECIFIC RSM 9
6 TRANSMISSION LOSS FACTORS IN THE RSM 9
7 DISTRIBUTION LOSS FACTORS IN THE RSM 14
8 HANDLING EMBEDDED GENERATION 14

1 Introduction

NZECS engaged Energy Link to review the process introduction document describing how the Residual Supply Mix (RSM) and Residual Supply Factor (RSF) are calculated, including consideration of any weaknesses or opportunities for improvement.

NZECS provided the following document describing the process of determining the RSM and RSF:

[R1] *New Zealand Energy Certificate System: Residual supply mix process introduction, Version 1.0, NZECS, November 2019*

In addition, a series of five questions, listed below, were included in the scope, relating to potential future developments, for Energy Link to comment on:

QUESTION 1. How could a 30-min reconciliation window be operated?

QUESTION 2. How could an implicit supplier-specific residual supply factor be derived?

QUESTION 3. What is the best method for determining a point-to-point transmission loss factor?

QUESTION 4. What is the best method for determining network specific loss factors from publicly available information?

QUESTION 5. How could we estimate the half-hourly output of embedded generation facilities in New Zealand?

Unless otherwise stated, any dollar values in this report are exclusive of GST.

2 Summary

Our review of the methodology shows that, since losses are always greater than zero, if sufficient certificates are issued in a period, one potential consequence is that the RSM will become negative. This suggests that either the methodology needs to be modified to account for losses at the point where certificates are issued, or the total certificates that can be issued must be limited to a value which is sufficiently small to ensure that the RSM can never go negative.

The RSM and RSF can be calculated using either a top-down or a bottom-up approach, the difference being the datasets that are primarily used. The top-down is simpler than the bottom-up approach but relies on having an accurate total annual demand figure from MBIE. The MBIE data is available quarterly, but final data for a calendar year is only available many months after the year.

The top-down approach builds up the generation mix using a combination of EA and Transpower datasets, with estimates of embedded generation attained by reference to the MBIE data. Unless long delays can be tolerated, allowing the top-down approach to be used, then the bottom-up approach will have to be used. To get the most accurate estimates for the larger EG plant, it will be necessary to work with both the EA and the SCADA datasets.

The reconciliation period can be reduced down to 30 minutes, but this would require use of the bottom-up approach and Transpower's SCADA data, updated every half hour.

The bottom-up approach works with the output of individual plant, and so facilitates the calculation of supplier-specific RSM and RSF.

Distribution losses can be calculated using loss factors published by distributors, and is already used in the monthly reconciliation of the spot market, coming under the auspices of the Code.

Transmission losses, however, are more challenging to work with, and any method to differentiate these losses on a physical basis will require a substantial investment in software and process in order to produce anything useful. However, there is a method using published spot prices that, with further investigation, potentially provides a proxy for the Tx losses between any arbitrary pair of nodes, one with generation and one with demand.

Taken overall, we believe the approach outlined in [R1] is reasonable when implemented using the bottom-up approach.

3 The RSM Methodology

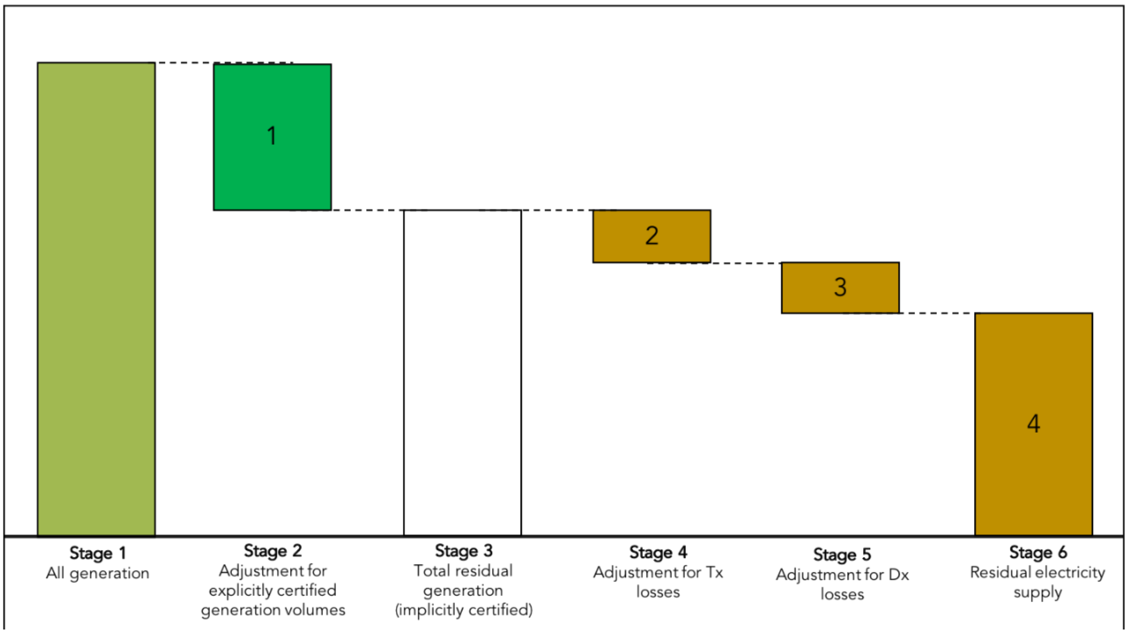
Section 6 of [R1] shows the following figure along with an explanation of how the RSM and RSF are calculated. In principle, the starting point at Stage 1 is all generation in New Zealand for the relevant period, initially one year.

Certificates are issued against generation held in a registry maintained by NZECS. The name of the generator, details of its emissions, if any, the number of MWh represented by the certificate, and other details, are recorded against each certificate.

Certificates may be issued by the respective generator to consumers, and the generation represented by these certificates is the green area shown at Stage 2 in the figure.

A certificate is issued against a specified renewable generator and effectively represents a 'claim' on the generation, in the ratio of 1 MWh of generation to 1 MWh thus certified. In other words, the point of certification is at the point of metering closest to the generator output terminals.

Figure 1 - Diagrammatic explanation of RSM components



The white region shown at Stage 3 represents all generation that is not certified during the period – the residual generation - which can be further broken down into transmission (Tx) losses (shown at Stage 4), distribution (Dx) losses (shown at Stage 5) and then all other generation (shown at Stage 6).

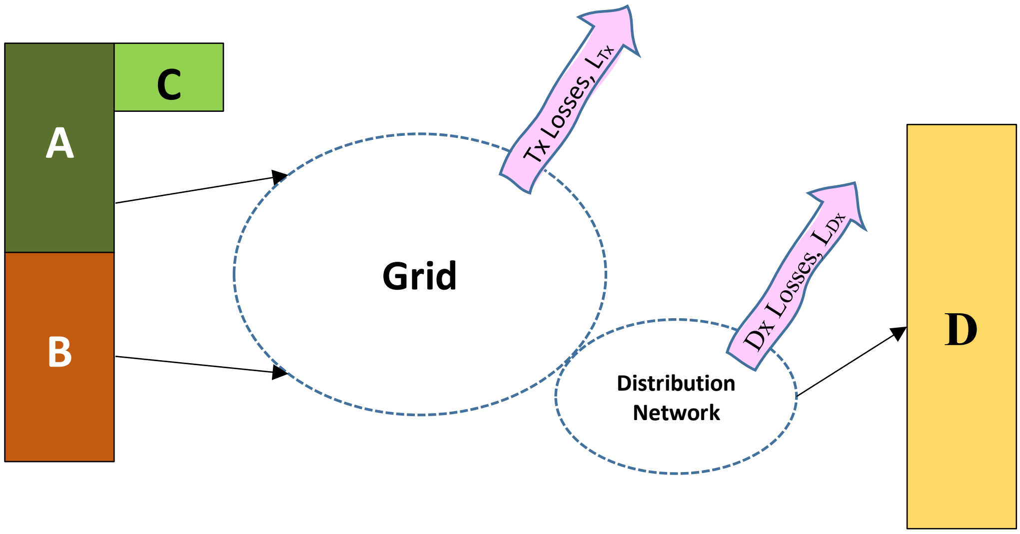
The RSF is the total emissions from the residual generation at Stage 3 divided by the total generation making up the residual generation at Stage 2.

Under this methodology, certified generation is effectively lossless.

At Stages 4 and 5, the Tx and Dx losses are deducted, leaving the residual electricity consumption (shown as residual electricity supply) at Stage 6. There Tx and Dx losses are assumed to homogeneous and to have the same RSF as the total residual generation.

The following figure shows a simple example where the supply mix consists of two generators, one renewable (generator A, output G_A) and one thermal (generator B, output G_B), with total metered demand, D.

Figure 2 – Physical Supply and Demand



The total generation for the period in Figure 2 is $G = G_A + G_B$ and we can write a formula which relates generation, losses and demand:

$$D = G - L_{Tx} - L_{Dx} \tag{1}$$

Now suppose that certificates totaling C are issued and redeemed against A, then G – C is the total physical energy for the remaining supply. If we subtract C from both sides of (1) then we get

$$D - C = G - C - L_{Tx} - L_{Dx} \tag{2}$$

But the righthand side of (2) is the total output of the RSM, shown in Stage 6 of Figure 1 so for the RSM as an energy measure, we can write

$$RSM = D - C \tag{3}$$

Since losses are always greater than zero, (1) tells us that $D < G$. If sufficient certificates are issued in a period, one potential consequence is that the RSM will become negative. This suggests that either the methodology needs to be modified to account for losses at the point where certificates are issued, or the total certificates that can be issued must be limited to a value which is sufficiently small to ensure that the RSM can never go negative, e.g. to around 90% of total generation.

Equations (1) and (2) also show us that there are at least two methods by which the RSM and RSF can be calculated, which we'll call the top-down approach and the bottom-up approach.

The RSF is ratio of the emissions of generator A and B, less emissions attached to certificates, C, to the total generation at Stage 3 in Figure 1, and is given by

$$RSF = \frac{(G_A - C)\times\varepsilon_A + G_B\times\varepsilon_B}{G_A + G_B} = \frac{(G_A\times\varepsilon_A + G_B\times\varepsilon_B) - C\times\varepsilon_A}{G} = \frac{E - C\times\varepsilon_B}{G} \tag{4}$$

where ε_A and ε_B are the emission factors for A and B, respectively, in units of kg/MWh[1], E is the total emissions from generation, and G is the total generation output.

3.1 Top-down Approach

Equations (2) and (3) suggest that the simplest approach to calculating RSM and RSF is to use total demand, D, total emissions, E, total certificates, C, and emission factor ε_A which applies to the certificates. In reality, certificates will be issued against more than one generator, so C actually becomes the sum of all emissions applying to certificates, as shown below:

$$C = \sum_{i=1}^{N} C_i \times \varepsilon_i \quad (5)$$

where C_i is the generation attached to certificate i and ε_i is the relevant emission factor.

Pros

- The main advantage of this approach is its simplicity, as NZECS only need to manage the data associated with certificates that are issued and redeemed.
- Losses do not need to be calculated.
- This approach also implicitly manages the issue of EG, on the assumption that the authority that calculates the total demand and the certificates, has access to data about EG.
- The data are available from MBIE and MfE and so have a high degree of credibility.
- Generation into and sales from the spot market are also available from NZX, in its role as the spot market Clearing Manager, and this data is compiled using a robust process prescribed by the Code.

Cons

- The method relies on having accurate data from MBIE and MfE. We have recently discovered issues with MBIE's demand data: they were missing returns from some smaller retailers, leading to an underestimate of total demand. It appears that the missing data is added over time, presumably as it arrives, but this means a prolonged delay in accessing the data.
- The data may not be available until long after the end of each year. In the case of demand, MBIE published the final figures for calendar 2018 in October 2019. In the case of total emissions from the electricity sector, these were issued in August 2019 for the 2018 calendar year.
- The Clearing Manger data for spot purchases is at GXP level, and so it includes Dx losses.

[1] Emission factors are available in g/kWh for some plant, for example geothermal. g/kWh, tonnes/GJ and kg/MWh all have the same value. For thermal generation, the factors may be available directly, or they can be calculated using the emission factors for the relevant fuel, scaled up the heat rate. For example, gas has emissions of 0.053 tonne per GJ of gas input energy. If a station is 50% efficient, then the heat rate is usually defined as $\frac{3600}{efficiency}$ = 7,200 GJ of fuel input per GWh of electricity output. The emission factor for plant output is 7,200 × 0.053 = 381.6 kg/MWh.

MBIE demand data is available quarterly, generally in the last month of each quarter for the immediately preceding quarter, so is delayed by about 2.5 months. This data, however, is updated progressively as gaps in the data are filled, and even after the final data for the previous calendar year is published in October, there can still be small adjustments made to the demand.

3.2 Bottom-up Approach

This approach is the one outlined in [R1] and it builds up the total generation using data for each grid-connected generator and for EG where the data is available. In this approach, the RSM is given by

$$RSM = G_A + G_B - C \tag{6}$$

Equation (5) looks simple because there are only two generators shown in Figure 2 but in the real market there are in excess of 200 generators when EG is included.

The calculation of the RSF is the same in principle, as shown in the first part of (3), but once again involves a large number of generators making up the total emissions, *E*.

Pros

- Output data for grid-connected generation[2] is available from EMS half hourly, and even down to the 10-minute level.
- Similar data is also available from the EA but it is posted about fifteen days into the month for the immediately preceding month.

Cons

- Data for EG is not readily available, neither in detail nor in aggregate.
- Neither the SCADA nor the EA data is fully complete for grid-connected generation.
- The EA and SCADA datasets do not match exactly, even for the same generator, as one would expect them to.

The EA data is in kWh and is posted for an entire month about fifteen days into the next month. The data includes EG data for Ngawha, Kaimai, Paerau and Tararua stages 1 and 2, none of which are provided in SCADA.

Data that appears to be missing from the EA output is Cobb, Te Ahi O Maui, Kawerau Onepu, Teviot Waipori (at the Halfway Bush GXP) and Mill Creek windfarm.

Mahinerangi windfarm is missing from both the EA dataset and the SCADA dataset.

The EA data has Stratford Peaker and TCC generation added together, which is problematic since these stations have different heat rates.

[2] Known as SCADA data, as it is collected via Transpower Supervisory Control and Data Acquisition system.

Energy Link uses a bottom-up approach to calculate emissions each week for inclusion in our Energy Trendz Weekly free report, and for other purposes. This analysis works at the plant level, with plant-specific emission factors.

The table below shows the data published by MIBE with a direct comparison to Energy Link's data, primarily based on SCADA data. Of note; the coal emissions from MIBE are higher than the Energy Link figure as well as total emissions and the total generation figure. In calculating the data below, we ignore EG, although it would be possible to use the EA data to add average values back in.

Table 1 – Energy Link versus MBIE Data

	Units	2018	2018 ELL
Electricity generation implied emission factor	kt $CO_{2\text{-}e}$/GWh	0.10	0.1014
Electricity consumption implied emission factor	kt CO2-e/GWh	0.11	
Total Emissions	kt CO2-e	4,241.19	4,157.00
Combustion Emissions	kt CO2-e	3,495.77	3,409.00
Biomass Emissions	kt CO2-e	0.12	
Coal Emissions	kt CO2-e	928.07	775.00
Gas Emissions	kt CO2-e	2,562.65	2,634.00
Geothermal Emissions	kt CO2-e	745.42	743
Liquid Fuels Emissions	kt CO2-e	4.93	
Annual Generation	GWh	4,2900.89	41,017.00
Annual Consumption	GWh	3,9432.70	

The major source of emissions from the electricity sector are the Rankine[3] units at Huntly which run on a mix of coal and gas. We noted that Genesis Energy disclosed the average coal:gas ratio for FY19 (88:12) and FY18 (63:37), but this data is available four months into the next FY, assuming the company continues to disclose it.

To estimate the emissions from these units, Energy Link uses SCADA generation data and gas off-take data from the Maui Information Exchange section of the Open Access Transmission Information System (OATIS) for gas, as shown below. The gas delivered to the Huntly gas offtake point supplies the Rankine units, the e3p combined cycle gas turbine (a.k.a. Huntly unit 5) and the P40 open cycle gas turbine (a.k.a. Huntly unit 6). Note that 'eff.' is used to denote the efficiency of the relevant plant.

1. All_Daily_Gas_Huntly_GJ = Daily offtake from Huntly Power Station 4002993 (Measured Qty) as recorded by OATIS in GJ/day
2. All_Huntly_Gas_MWh = All_Daily_Gas_Huntly_GJ / 3.6
3. E3p_Gas_MWh = E3p_Generation_MWh / eff.
4. P40_Gas_MWh = P40_Generation_MWh / eff.
5. Huntly_Remain_Gas_MWh = All_Huntly_Gas_MWh – E3p_Gas_MWh – P40_Gas_MWh
6. Rankine_Gas_Generation_MWh = Minimum (Huntly_Remain_Gas_MWh × eff., Rankine_Generation_MWh)
7. Rankine_Coal_Generation_MWh = Rankine_Generation_MWh – Rankine_Gas_Generation_MWh
8. The Rankine_Coal_Generation_MWh is then used to calculate the emissions from the coal burn.

[3] Steam turbine.

SCADA data also includes demand data, and a further source of demand data is the SPD demand data that is released each day by the EA: this is the demand data that is used in the SPD market-clearing software that dispatches generation and calculates spot prices each and every trading period. The SPD data is usually available with a delay of around 12 hours, but is difficult to extract. However, Energy Link extracts this data and can supply it for a modest sum, if required. SPD demand data does include windfarms but does not include EG.

As can be seen from the above, using the bottom-up approach requires a thorough understanding of the datasets being used, and probably requires use of both the EA and SCADA datasets, the latter coming at a cost of approximately $460 per month[4].

EG data is also missing in the bottom-up approach, but estimates of this can be made using MBIE data, either on a quarterly or annual basis.

The reality is that there is no perfect set of generation and demand data, and any and every approach requires careful analysis of data, and regular checks to ensure that changes have not gone unnoticed.

3.3 Recommendation

Unless long delays can be tolerated, allowing the top-down approach to be used, then the bottom-up approach will have to be used. To get the most accurate estimates for the larger EG plant, it will be necessary to work with both the EA and the SCADA datasets.

4 30-minute Reconciliation

A move to half-hourly reconciliation would require the use of SCADA data, augmented with estimates for EG added from the EA generation dataset. In principle, this should be relatively straightforward, assuming that the annual reconciliation also uses SCADA data. A move to daily reconciliation prior to half hourly might make for a smoother transition.

The SCADA data is updated every half hour, but does appear to be delayed by up to an hour. 10-minute SCADA is also available at a price, and we assume this would be updated more often. We recommend approaching EMS direct to determine whether the half hourly SCADA generation data can be obtained shortly after the end of each half hour.

The steps would be:

1. at the end of the half hour (or as soon as the data is available for the half hour), download the SCADA generation and demand from EMS' web site[5];
2. perform the RSM and RSF calculations;
3. publish the results and store for future use.

[4] Should be confirmed with EMS.
[5] EMS' web site for SCADA data is at www.em6.co.nz

5 Supplier-specific RSM

One thing the top-down approach cannot do, is to estimate supplier-specific RSMs and RSFs. Using the bottom-up approach, would facilitate this, but it would need to be based on plant-level data and emissions, so that a supplier's total and residual generation and emissions can be calculated using plant-specific emission factors[6].

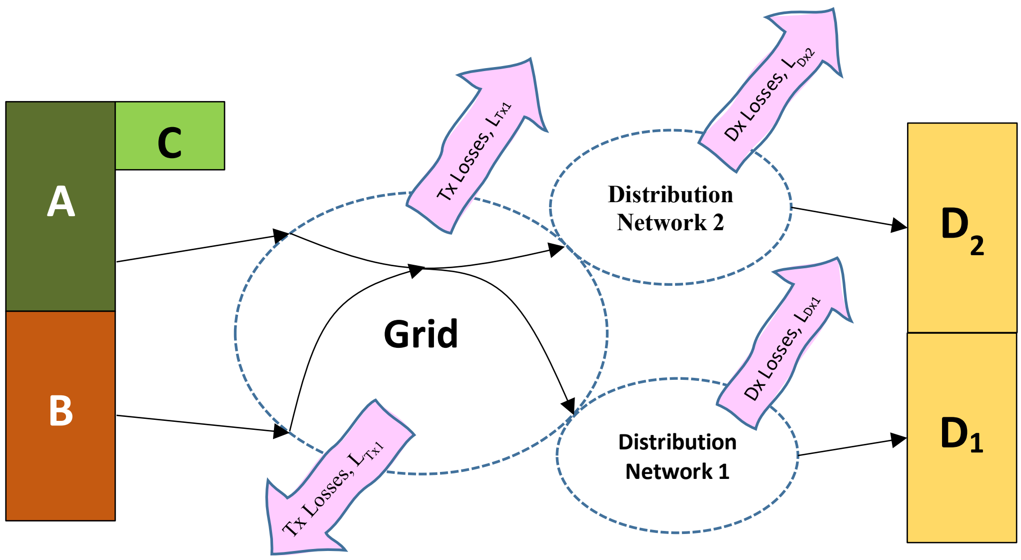
The methodology is a simple extension to the bottom-up approach, just focusing on one supplier at a time. If it were deemed necessary to include hedge and other contracts between a retailer and a generator, this would be considerably more difficult. A prime example is the swaption contract between Meridian Energy and Genesis Energy, which allows Meridian to call on a hedge contract if spot prices are high while its generation is less than its retail exposure: Meridian pays a fixed premium to Genesis, which Genesis applies to help keep its thermal plant in the market. In effect, Meridian is paying Genesis to keep thermal plant running.

The details of hedge and other supply contracts are not all in the public domain, so it would be necessary to obtain details direct from suppliers. A lot of hedging is also done on the ASX futures market, which does not identify counterparties, so there is no way of knowing what emissions are associated with the futures contracts held by a supplier.

6 Transmission Loss Factors in the RSM

Figure 3 below shows physical supply with demand connected via two distribution networks, and demand D_1 and D_2, assuming that we can separate out the losses attributable to each network at grid level.

Figure 3 – Physical Supply with Tx and Dx Losses



We can expand (2) as shown below

$$D_1 + D_2 = G - L_{Tx1} - L_{Dx1} - L_{Tx2} - L_{Dx2} \quad (7)$$

[6] As opposed to using technology or fuel-type emission factors published by government.

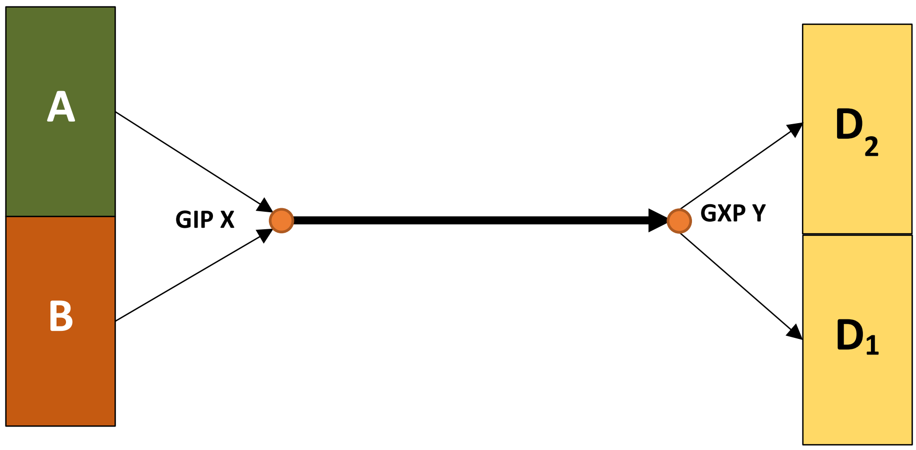
Whereas the total grid losses L_{Tx} are easily calculated by subtracting total grid offtake from total injection onto the grid, the losses $L_{Tx1} - L_{Dx1} - L_{Tx2} - L_{Dx2}$ are not: to calculate the losses relating to each distribution network requires the use of a piece of software called a 'power flow' model which has grid injection, demand, and the grid configuration as inputs.

The grid consists of around 850 transmission lines and transformers, and the data for the New Zealand grid is available, but it changes by half hour as lines are taken out of service for maintenance. If one were to deploy a power flow model to calculate grid losses by local network, then one approach would be to run the model with and without the customer's demand, and the losses relating to that customer would be the difference in losses between the two scenarios, with and without the customer.

This would be a significant undertaking, and even if it were done on an annual basis, would still require running these scenarios for all half hours in the year.

Figure 4 shows a simplified grid with just one transmission line running from GIP X to GXP Y and carrying all generation from generators A and B, to consumers on local networks 1 and 2 with total demand D_1 and D_2.

Figure 4 – Simple Power Flow Example



The losses on the transmission line are a function of the power flowing in the line, which in this example is $D_1 + D_2$ (in MW) and a physical property of the line called 'impedance'. The impedance has two parts, but we can simplify the impedance down to a single property called the 'resistance' of the line, R, and use this to calculate the losses, L_{Tx}.

$$L_{Tx} = \frac{R\times(D_1+D_2)^2}{100} = \frac{R\times {D_1}^2}{100} + \frac{R\times D_1 \times D_2}{100} + \frac{R\times {D_2}^2}{100} \tag{8}$$

If we set $D_2 = 0$ and calculate the losses for D_1 alone, then we get

$$L_{Tx1} = \frac{R\times {D_1}^2}{100} \tag{9}$$

and if we set $D_1 = 0$ and calculate the losses for D_2 alone, then we get

$$L_{Tx2} = \frac{R \times D_2^{\,2}}{100} \tag{10}$$

and we see immediately that $L_{Tx} > L_{Tx1} + L_{Tx2}$ (except when either D_1 or D_2 is zero)[7].

If one were to use a power flow model to calculate losses by subtracting demand, then the total losses thus obtained would not sum to the actual total losses, which further suggests that running power flows to calculate transmission losses is not worth the effort.

Referring again to Figure 4, we can also calculate the 'marginal losses' on the transmission line as ML_{Tx}, where

$$ML_{Tx} = \frac{2 \times R \times (D_1 + D_2)}{100} \tag{11}$$

and it is a feature of our spot market that the spot price at each grid node (GIP or GXP), in each half hour, is usually a function of the marginal losses at the relevant node, and the marginal offer in the half hour.

In our simple example, one of the generators A or B will be on the margin, and suppose the price of its marginal offer is $S/MWh, then the spot price at the GIP will be $S/MWh and at the GXP it will be $S \times (1 + ML_{Tx})$. The price calculations are more complex on the real grid, but the impact of marginal losses is still the same, except that the marginal losses are a function of the power flowing on more than one line.

So, in principle, one could calculate an RSM and RSF for each local network connected to the grid, in each half hour, using the spot prices published each day for the previous day. When Real-time Pricing (RTP) is introduced in 2022, spot prices will be published at the end of each and every half hour.

However, spot prices are also impacted by reserves and by network constraints, so they do not always truly reflect the impact of marginal losses, which means that calculations based on spot prices could produce results that are not in line with marginal losses, and this could continue for prolonger periods in some circumstances. For example, the extended HVDC outage scheduled for Q1 of 2020 will see large price differences between the two islands which are created, not by marginal losses, but by the need to provide costly reserves to cover the risk of the HVDC link suffering an unplanned full outage.

Spot prices could be recalculated ignoring reserves and network constraints, but this would require at least as much effort as using the full power flow by half hour.

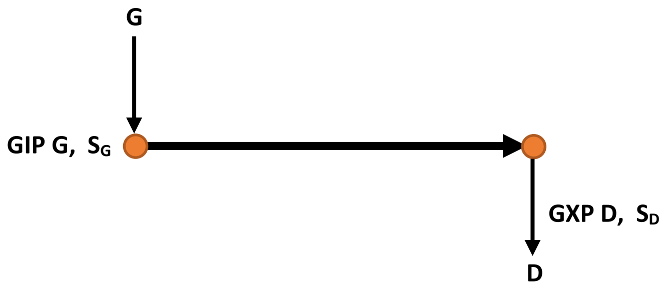
All things considered, there does not appear to be a practical way forward for calculating Tx losses down to the network or customer level.

[7] This occurs because the losses are not a linear function of power flows, but proportional to the square of the power flows. Another way of stating this is that the "marginal losses are greater than the average losses".

6.1 Point-point Proxy Losses

However, as noted in section 6 above, spot prices are a function of marginal Tx losses. Suppose we have generation at GXP G with spot price S_G, and demand at GXP D with spot price S_D, as shown below.

Figure 5 – Spot Prices



Then we can show that

$$\frac{S_D}{S_G} = \frac{(1+ML_D)+N_D+R_D}{(1+ML_G)+N_G+R_G} \tag{12}$$

where ML_D and ML_G are the marginal Tx losses at nodes D and G, respectively, and the N and R are network and reserve terms which may or may not be zero in any given half hourly trading period of the spot market.

The N's are present when there is network congestion, which occurs when a Tx line or Tx constraint reaches its limit, for example during a spring washer event. But most of the time the N's are zero.

The R's are greater than zero when the HVDC link has a non-zero reserve cost associated with it. For example, during the extended period in early 2020 when the HVDC link has only one pole operating while the other pole is being serviced, there are long periods each day when all spot prices in the North Is have an R value which is greater than zero, often in the \$10's of dollars per MWh. Once the HVDC link is fully back in service, however, the R values will fall back to zero, or close to it, most of the time.

Assuming for a moment that we can ignore the N and R values in equation (12), we have a formula relating the marginal losses at the two nodes to their spot prices, now treating generation as negative demand[8]:

$$\frac{S_D}{S_G} = \frac{(1+ML_D)}{(1-ML_G)} \tag{13}$$

The ratio $\frac{S_D}{S_G}$ is known as the 'location factor' of node D relative to node G, and when the N's and R's are zero this is purely a function of marginal Tx losses.

[8] Which means the sign of ML_G becomes negative.

If a consumer at GXP D were looking to factor Tx losses into a decision about certificates at GIP G, then they would be interested in the Tx losses between nodes G and D. The marginal Tx losses tell us how the total Tx losses changed when the last MW of demand was added at a node:

$$ML = \frac{\Delta L}{\Delta D} \tag{13}$$

where L is the total Tx loss during the trading period, ΔL is the increase in total losses due to the last MW of demand at GXP D, and if we set $\Delta D = 1$ then we have

$$\frac{S_D}{S_G} = \frac{(1+\Delta L_D)}{(1-\Delta L_G)} \tag{14}$$

What the consumer is really interested in, referring to Figure 5 above, is the loses supplying GXP D when the supplying generator is at GIP G. So if we call these losses ΔL_{GD} then we could write

$$S_D = S_G \times (1 + \Delta L_{GD}) \tag{15}$$

then from (14) and (15) we get

$$\Delta L_{GD} = \frac{S_D}{S_G} - 1 \tag{16}$$

which says that the marginal losses between GXP D and GIP G are equal to the location factor of node D relative to node G, minus 1. Depending on the two prices, the losses could be positive, but they could also be negative if, for example, GXP D is close to a large generator and GIP G is in a large load centre.

Equation (16) could be used to calculate a proxy for the Tx losses between any pair of nodes, with the following caveats:

1. to obtain the location factor for a period such as a month or year, the two prices should first be averaged for the period, then divided through[9];
2. there will be periods when the location factors are not just functions of marginal losses, but also of reserves and possibly spring washer effect, although the impact of this will be reduced to an extent by averaging over longer periods;
3. marginal losses are greater than average losses, as discussed in section 6, which means that this proxy method overestimates the losses between pairs of nodes.

Notwithstanding the caveats above, the advantage of this proxy method is that it is easy to calculate because the spot prices are readily available. Currently they are usually available the following day but when RTP is introduced then they will be available at the end of each trading period. With further investigation into the impact of network congestion and reserves on location factors, and the degree of overestimation of the

[9] Instead of calculating the half hourly locations factors and then averaging these over the period.

losses[10], this method may suffice as a proxy for a full power flow analysis of the physical losses across the grid.

7 Distribution Loss Factors in the RSM

While Tx losses are problematic, there is already a simple method for calculating Dx losses: distributors must publish loss factors that can be applied right down to the half hourly, customer-specific level. These loss factors are used in the monthly spot market settlement process[11] to calculate how much electricity left the grid at the relevant GXP to supply each customer. If LF is the loss factor, then in any given half hour

$$Grid\_offtake\_at\_GXP = Meter\_reading \times LF \tag{11}$$

The EA has a guide for calculating these loss factors, so there is likely to be a degree of consistency between distributors, although the granularity of loss factors varies considerably amongst networks and customers on networks.

Given that these loss factors are already in use, readily available, and under the eye of the EA, we recommend the use of the published loss factors for the purpose of determining customer-specific Dx Losses.

8 Handling Embedded Generation

EG is covered in earlier sections, but these are the two key points:

1. in principle, MBIE's quarterly and annual demand data includes the impact of EG;
2. some larger EG is contained in the EA and SCADA datasets, but both datasets have gaps.

In practice, for a bottom-up approach, some of the gaps can be filled by using both the SCADA and EA datasets, with the remaining gap being filled by taking the difference between MBIE's final generation total and the total generation obtained using the bottom-up approach. This would provide an average amount of EG to use in the bottom-up calculations.

There is additional information about installed capacities for EG at the EA's EMI portal, but this does not include the output: the data includes total installed capacity, average installed capacity, EG by region, and the number of EG installations, all by 'fuel type'.

[10] For example, when the total quantity of certificates is such that all generation is certified, then it might be that the total proxy Tx losses exceed the actual Tx losses, in which case scaling would be required on the proxy losses.

[11] The process is called Reconciliation and is defined under the Code.