# Load forecasting spatial maximum demand 2020-25

January 2019

Part of the Energy Queensland Group

## Methodology

This methodology was used to develop accurate forecasts of Spatial Maximum Demand for summer and winter within the two networks of Energy Queensland. The three key objectives in establishing the spatial maximum demand forecasting process include establishing:

- robust, reliable, and repeatable techniques
- transparent methods
- econometrically based models, that is, normalised for weather, and utilising externally verifiable data as a source.

## Methods

The bottom-up spatial forecast methodology allows the capture of underlying characteristics of the areas serviced by individual substations. This approach allows for variations across the diverse environments based on observed and planned local developments (local population growth, housing developments etc.) that lead to different growth patterns across the different areas. Although substation level demand forecasts are based on inherently "noisy" load metering data compared to the aggregated system level, it is necessary to capture the growth rates of individual substations relative to one another.

This paper notes that forecasting at the "spatial" zone substation level presents the following issues:

- time series need to be adjusted for network transfers and switching events
- greater degrees of randomness which is not easily explained
- difficulties incorporating the impacts of macro level drivers

ACIL Tasman (2010) argue that these factors mean that while it is possible to get the relativities between individual zone substations reasonably correct, the aggregation of individual zone substation forecasts is likely to result in forecasts that deviate substantially from those that would be derived from a top-down system level forecast. The trimming process allows examination and amelioration of these variations.

Implementation of spatial maximum demand forecasting is a complex and difficult exercise. A high level of sophistication and data is necessary to establish a methodology that satisfies operational requirements and addresses the issues raised by external and independent consultants. In response to these earlier recommendations, Ergon Energy and Energex adopted and developed the Substation Investment Forecasting Tool (SIFT).

## SIFT Process

SIFT produces a long term, 10-year forecast of zone substation demand. Any adjustments made to the extrapolated maximum demand accommodate both confirmed and anticipated developments as well as other known local factors. From this, bulk supply and other external stakeholder forecasts, and connection point demand forecasts are prepared by aggregating the loss-adjusted forecast of the attached zone substations.

After performing a reconciliation between a separately created top-level system forecast and the aggregated zone substation forecasts, a trim factor is applied selectively to bring the forecast results, as near as is practicable, in line with measured results.

Currently SIFT requires both maximum demand growth rates and temperature corrected demand information to complete the forecast and these are provided as major parameters within the SIFT automation process.

These include:

- Probability of Exceedance (PoE) data, 10PoE, 50PoE, and 90PoE for each network element.
- Growth data, data that describes the predicted growth at the zone substation level over the life of the forecast.

**Temperature dependence**

The relationship between winter daily maximum demand and temperature has a different relationship to that expected in summer.

As winter tends to be characterised by colder temperatures, demand increases as temperatures decline.

Alternately, summer tends to be characterised by warmer temperatures, demand increases as temperatures increase.

**Calendar days**

Treatment for determining demand modelling currently includes the public holidays using separate variables.

**Model method details**

The following describes the model of both summer and winter models at the spatial level. The methodology based on multivariate analysis uses multiple linear regressions. This involves fitting a multi-dimensional linear function through the daily summer and winter maximum demand data to minimise the total squared errors between the fitted function and the observed data. The model structure needs to pick up the impacts of the underlying drivers. The summer and winter maximum demand model takes the base form:

**Equation 1**

$$Load = c + \beta_1 \times Max\ temp + \beta_2 \times Min\ temp + e$$

This equation explains the season daily maximum demand by a constant term together with the drivers of daily maximum and minimum temperature. Temperature related demand variation is captured by inclusion of the daily maximum and minimum temperatures. Any variations in the daily peaks not captured by temperature are soaked up by the error term (e).

Temperature is the main driver for the current model, as other econometric and demographic details are either not available or too volatile to be included as a robust set of inputs into the model.

The specification above can vary depending on what data is included in the regression model; for example – if weekends and public holidays are included but not accounted for explicitly, the lower peak demands associated with them will diminish the precision of the model in its accuracy. To avoid this occurring, the effect of the lower demand on weekends is resolved through the inclusion of specific variables which take a value of '1' on that particular day and '0' on other days. The specific variable will shift the regression line downwards on weekends and holidays to reflect the lower levels of peak demand on those days.

Weather normalisation (or weather correction) is a key aspect of the spatial demand forecasting methodology. Electricity demand is highly sensitive to weather due to temperature sensitive loads. High temperature conditions in summer result in high peak demands, while cold temperatures in winter result in peak demands. Any comparison of historical electricity loads over time requires adjustment of these loads to standardised weather conditions. In this model, actual demand is standardised to 10%, 50%, and 90% Probability of Exceedance levels (PoE).

**Monte Carlo approach**

The maximum demand data set is used in the model containing drivers. A forecast is possible only with knowledge of future temperatures. As it is not possible to generate an accurate forecast when the period is greater than a few weeks, the technique applied is one using a distribution of historical temperatures.

The forecast is determined using all previous temperatures. This is now a distribution in future maximum demand values, due to each past temperature being a contributor.

As this is now a distribution in future maximum demand values (due to each past temperature being a contributor) then calculation of best estimate of the distribution, if symmetric, or the 50 PoE value, is from the distribution to provide the forecast.

This method of generating a maximum demand distribution is the Monte Carlo simulation and is one of the standard mechanisms for temperature correction.

Current and historical points in time provide the 'In sample' 50 PoE values for maximum demand needed for a continuous PoE time series for each season.

**Development of the summer and winter models**

The basis for the choice of model parameter is derived from of a broad range of potential key drivers representing exogenous (external variables) variables that influence the dependent maximum demand. Consequently, a sense of the likely size and direction of model coefficients can be determined. Model validation is required and any validation should confirm established relationships fit with known theory.

**Model Validation**

The main methods of model validation include:

- assessment of the goodness of fit of the regression
- statistical significance of the explanatory variables

A statistically significant result is one which, if an observed effect were to occur, would be sufficiently large that it would rarely occur by chance. Each estimated coefficient in the regression models has an associated t-statistic and associated p value[1].

One poorly understood key issue is that a statistically significant result does not necessarily imply that the inclusion of a particular variable will have a sizeable impact on the model outcomes.

Firstly, and often in large sample sizes, the statistically significant results identified are of little or no economic consequence. Secondly, a statistically significant result also has some chance of not being significant. At the 5% significance level, 1 in 20 significant results will in fact be insignificant. Thirdly, the longevity of the variable requires judgement to ensure the robustness of its significance and relevance over the forecasted timeframes.

The most commonly used measure of the goodness of fit of the regression model to the observed data is $R^2$. An $R^2$ coefficient of determination is a statistical measure of how well, the regression approximates the real data points. An $R^2$ of 1.0 indicates that the regression perfectly fits the data, while the rating of an $R^2$ above 0.7 is 'reasonably good.'

Other measures include analysis of the residuals for complete specification review.[2] However, for reasons of stationarity simplicity and in-sample PoE calculation, only $R^2$ and p values were analysed.

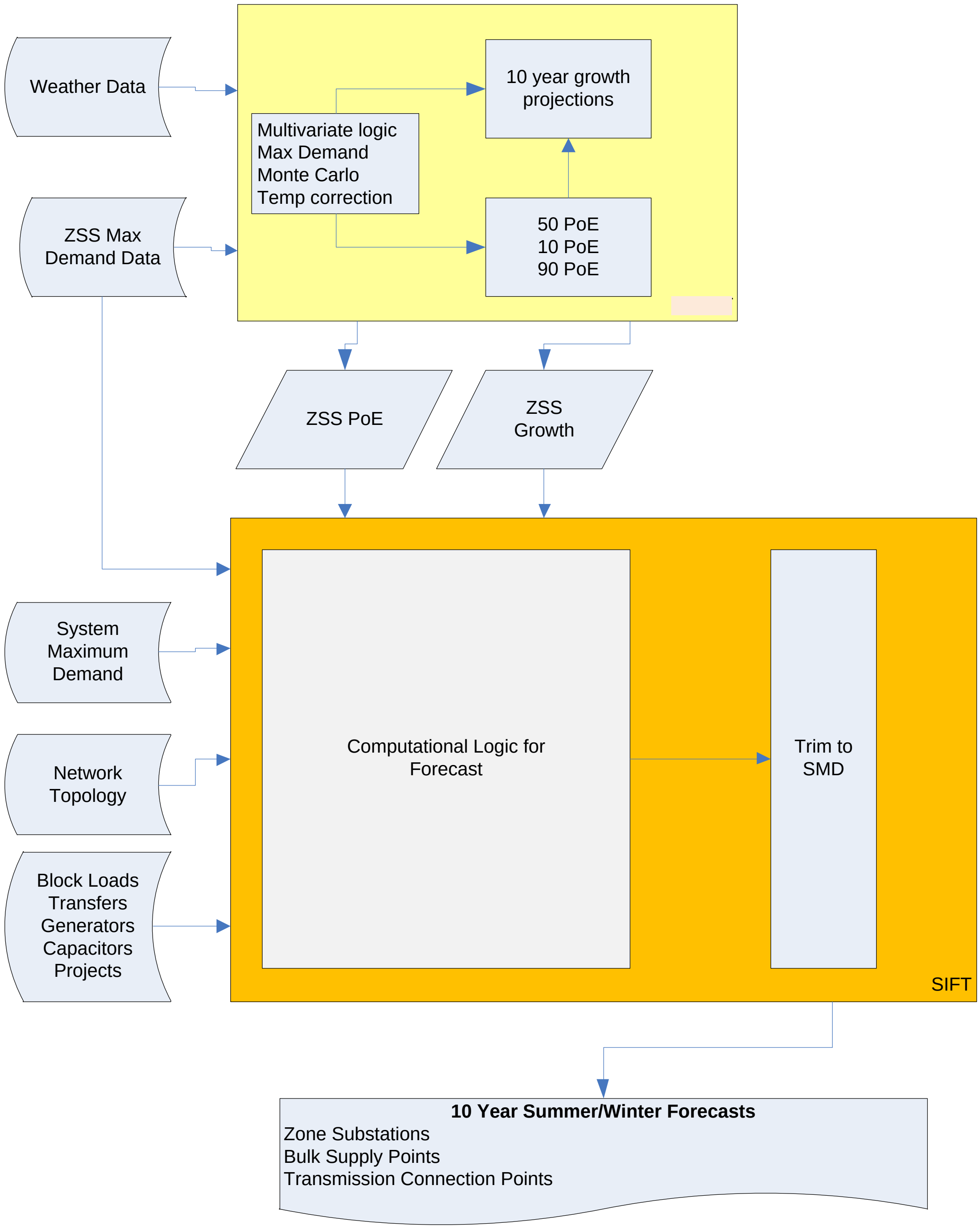
[1] If the estimated p value is less than 0.01 then that coefficient is statistically significant at the 1% significance level. A p-value that is less than 0.05 is significant at the 5% level of significance. The lower the observed p value on a coefficient the greater the probability that a meaningful relationship exists between the dependent variable and the explanatory variable concerned.

[2] Akaike Information Criterion (AIC statistic) together with the Schwartz Criterion (SC). Both the Akaike Information Criterion (AIC) and Schwartz Criterion (SC) are model selection tools designed to choose between sets of potential models. The criteria penalises models which have poorer R2 which is measured by the sum of squared errors and also includes a penalty for each additional explanatory variable included in the model. Out of several competing models the one with the lowest AIC and SC are considered to be the best. Other tests include heteroskedasticity, serial correlation and multiple colinearity

The forecast process chart describes the high-level process flows for developing summer and winter spatial maximum demand forecasts.

The process for developing summer and winter forecasts are identical. The variation between summer and winter forecasts arises from using different data sets for demand and temperature as well as minor changes in the algorithm to account for different calendar days and holiday periods.

The process uses a series of external data sources and system forecast loaded into SIFT along with a range of other network data. SIFT is used to develop the spatial forecast.



**Data**

The main criteria used to determine data series suitability for use in the modelling methodology include:

- reputability of the data source
- reliability of the data source
- completeness (no or few missing values)
- suitably long time series
- accuracy of the data
- continued availability into the future.

**Maximum Demand**

An effective methodology requires daily observation of both summer and winter maximum demand for a number of years in order to capture a series of data that shows the potential range of output, with variations in input. However, because we are considering a 'stationary' multivariate approach each season has its own set of coefficients that are determined for each temperature variable (where several seasons are used, non-stationary effects are accounted for using rebasing techniques). Metering data provides the demand data used in the process.

**Weather**

The modelling process requires the use of a suitable weather series to relate daily movements in Spatial Maximum Demand to weather variation. Currently the models use only daily minimum and maximum temperature data. Other weather data elements may be utilised in future.

The methodology uses weather data both as part of the regression model to relate spatial maximum demand to weather drivers, as well as part of the long run weather series used to derive the 10% PoE and 50% PoE demands.

The source of the weather time series is the Bureau of Meteorology (BOM)[3.] A nearby weather station provides source data for each zone substation.

[3] Weather data was sourced from the Bureau of Meteorology, Climate Data Online website. http://www.bom.gov.au/climate/data/.

The following derivation and equations used to determine the coefficients of the demand equation and the coefficients of the regression used to determine the degree of fit between the model and the actual demand and temperature statistics.

**Coefficient determinations of the multivariate peak demand equation**

Writing the multivariate equation more generally as

$$y_i = (\beta_0 + \sum_{j=1}^{k} (\beta_j x_{ji}) + \varepsilon)$$

We can use standard minimisation of residuals to obtain methods to evaluate the multivariate coefficients by the following:

Obtaining the squared of the residual

$$\sum_{i=1}^{n} [y_i - (\beta_0 + \sum_{j=1}^{k} (\beta_j x_{ji}))]^2$$

And minimising

$$\frac{\partial}{\partial \beta_0} \sum_{i=1}^{n} [y_i - (\beta_0 + \sum_{j=1}^{k} (\beta_j x_{ji}))]^2 = 0$$

$$\frac{\partial}{\partial \beta_1} \sum_{i=1}^{n} [y_i - (\beta_0 + \sum_{j=1}^{k} (\beta_j x_{ji}))]^2 = 0$$

$$\frac{\partial}{\partial \beta_2} \sum_{i=1}^{n} [y_i - (\beta_0 + \sum_{j=1}^{k} (\beta_j x_{ji}))]^2 = 0$$

$$\frac{\partial}{\partial \beta_k} \sum_{i=1}^{n} [y_i - (\beta_0 + \sum_{j=1}^{k} (\beta_j x_{ji}))]^2 = 0$$

So differentiating we obtain

$$-\sum_{i=1}^{n} [y_i - (\beta_0 + \sum_{j=1}^{k} (\beta_j x_{ji}))] = 0$$

$$-\sum_{i=1}^{n} x_{1i} [y_i - (\beta_0 + \sum_{j=1}^{k} (\beta_j x_{ji}))] = 0$$

$$-\sum_{i=1}^{n} x_{2i} [y_i - (\beta_0 + \sum_{j=1}^{k} (\beta_j x_{ji}))] = 0$$

.

$$.-\sum_{i=1}^{n} x_{ki} [y_i - (\beta_0 + \sum_{j=1}^{k} (\beta_j x_{ji}))] = 0$$

Rearranging becomes

$$\sum_{i=1}^{n} \beta_0 \quad + (\sum_{i=1}^{n} \quad \sum_{j=1}^{k} (\beta_j x_{ji}))] = \sum_{i=1}^{n} y_i$$

$\sum_{i=1}^{n} x_{1i}\beta_0 \quad + (\sum_{i=1}^{n} \quad \sum_{j=1}^{k} (\beta_j x_{ji}))] = \sum_{i=1}^{n} x_{1i}\, y_i$

$\sum_{i=1}^{n} x_{2i} \quad + (\sum_{i=1}^{n} \quad \sum_{j=1}^{k} (\beta_j x_{ji}))] = \sum_{i=1}^{n} x_{2i}\, y_i$

.

$.\sum_{i=1}^{n} x_{ki} \quad + (\sum_{i=1}^{n} \quad \sum_{j=1}^{k} (\beta_j x_{ji}))] = \sum_{i=1}^{n} x_{ki}\, y_i$

In matrix form, we can represent this as

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{2i} & \sum_{i=1}^{n} x_{ki} \\ \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} {x_{1i}}^2 & \sum_{i=1}^{n} x_{1i}x_{ki} & \sum_{i=1}^{n} x_{1i}x_{ki} \\ \sum_{i=1}^{n} x_{ki} & \cdots \sum_{i=1}^{n} x_{ki}x_{ki} & \sum_{i=1}^{n} x_{ki}x_{ki} & \sum_{i=1}^{n} {x_{ki}}^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_j \\ . \\ . \\ . \\ \beta_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} \quad y_i \\ \sum_{i=1}^{n} x_{1i}\, y_i \\ \sum_{i=1}^{n} x_{2i}\, y_i \\ . \\ . \\ . \\ \sum_{i=1}^{n} x_{ki}\, y_i \end{bmatrix}$$

Which is of the form

$$AC = \mathrm{H}$$

$$C = A^{-1}\mathrm{H}$$

AND

$$A^{-1} = \begin{bmatrix} n & \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{2i} & \sum_{i=1}^{n} x_{ki} \\ \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} {x_{1i}}^2 & \sum_{i=1}^{n} x_{1i}x_{ki} & \sum_{i=1}^{n} x_{1i}x_{ki} \\ \sum_{i=1}^{n} x_{ki} & \cdots \sum_{i=1}^{n} x_{ki}x_{ki} & \sum_{i=1}^{n} x_{ki}x_{ki} & \sum_{i=1}^{n} {x_{ki}}^2 \end{bmatrix}^{-1}$$

The Coefficients are given by

$$\begin{bmatrix}\beta_0\\ \beta_1\\ \beta_j\\ \cdot\\ \cdot\\ \cdot\\ \beta_k\end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{2i} & \sum_{i=1}^{n} x_{ki} \\ \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} {x_{1i}}^2 & \sum_{i=1}^{n} x_{1i}x_{ki} & \sum_{i=1}^{n} x_{1i}x_{ki} \\ \sum_{i=1}^{n} x_{ki} & \cdots \sum_{i=1}^{n} x_{ki}x_{ki} & \sum_{i=1}^{n} x_{ki}x_{ki} & \sum_{i=1}^{n} {x_{ki}}^2 \end{bmatrix}^{-1} \begin{bmatrix}\sum_{i=1}^{n} y_i\\ \sum_{i=1}^{n} x_{1i}y_i\\ \sum_{i=1}^{n} x_{2i}y_i\\ \cdot\\ \cdot\\ \sum_{i=1}^{n} x_{ki}y_i\end{bmatrix}$$

So in summary, to evaluate the coefficients of the maximum demand given by

$$y_i = (\beta_0 + \sum_{j=1}^{k} (\beta_j x_{ji}) + \varepsilon)$$

The coefficients are given by

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{2i} & \sum_{i=1}^{n} x_{ki} \\ \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} {x_{1i}}^2 & \sum_{i=1}^{n} x_{1i}x_{ki} & \sum_{i=1}^{n} x_{1i}x_{ki} \\ \sum_{i=1}^{n} x_{ki} & \cdots \sum_{i=1}^{n} x_{ki}x_{ki} & \sum_{i=1}^{n} x_{ki}x_{ki} & \sum_{i=1}^{n} {x_{ki}}^2 \end{bmatrix}^{-1} \begin{bmatrix}\sum_{i=1}^{n} y_i\\ \sum_{i=1}^{n} x_{1i}y_i\\ \sum_{i=1}^{n} x_{2i}y_i\\ \cdot\\ \cdot\\ \cdot\\ \sum_{i=1}^{n} x_{ki}y_i\end{bmatrix}$$

**Regression Coefficients**

The regression coefficients determine the degree of ‘fit’ of the model to the data. We proceed as follows:

The actual spatial peak demand is given by

$$y_i = (\beta_0 + \sum_{j=1}^{k} (\beta_j x_{ji}) + \varepsilon_i)$$

And the spatial peak demand model is

$$\hat{y_i} = (\hat{\beta_o} + \sum_{j=1}^{k} \left(\hat{\beta_j} \quad x_{ji}\right) + \varepsilon_i)$$

Then R-squared is given by

$$R^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 / \sum_{i=1}^{n} (_i \hat{y_i} - \bar{y})^2$$

And

The variance is given by

$$\sigma^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 / n - l$$

# Annex A - SIFT Meshing Calculations

The following diagram contains the definition of the SIFT meshing calculation:

## Radial Network

**Abbreviations**

- BSP – bulk supply point
- FC – forecast
- ‘L’ – local
- LF – loss factor
- MW – mega watt
- MW0 – mega watt at year zero
- MWx – mega watt at
- ‘S’ – system
- Reconciled – unreconciled value x reconciliation factor
- MVAR0 - MVAR at year zero
- MVARx – MVAR at forecast point
- ZS – zone substation
- ZSr – zone substation with a set “trim” flag
- ZSu – zone substation without a set “trim” flag

### Forecast Equations

$$\text{BSP reconciled coincident MW FC} = \sum \text{unreconciled coincident ZS MW} \times \text{Reconciliation Factor} \times \text{BSP MW LF}$$

$$\text{BSP reconciled peak MW FC} = \frac{\text{BSP reconciled coincidence MWx}}{\text{BSP MW CF}}$$

$$\text{BSP reconciled coincident uncompensated MVAR FC} = (\sum(\text{unreconciled coincident uncompensated ZS MVAR} \times \text{Reconciliation Factor} \times \text{BSP MVAR LF})) - (\sum \text{FC ZS reconciled coincident compensation})$$

[If the first part of thes equations = 0, then use 1]

$$\text{BSP reconciled peak uncompensated MVAR FC} = \left( \frac{(\sum(\text{unreconciled coincident uncompensated ZS MVAR} \times \text{Reconciliation Factor} \times \text{BSP MVAR LF}))}{\text{BSP MW CF}} \right) - (\sum \text{FC ZS reconciled coincident compensation})$$

$$\text{ZS reconciled coincident MW FC} = \text{ZS coincident MWx} \times \text{reconciliation factor}$$

$$\text{ZS reconciled peak MW FC} = \frac{\text{ZS coincident MWx} \times \text{reconciliation factor}}{\text{ZS CF}}$$

$$\text{ZS reconciled coincident MVAR FC} = \text{ZS coincident FC MVAR} \times \text{reconciliation factor}$$

$$\text{ZS reconciled peak MVAR FC} = \left( \sqrt{\left( \frac{\text{ZS unreconciled coincident MWx / ZS CF}}{\text{ZS PPF}} \right)^2 - \left( \frac{\text{ZS unreconciled coincident MWx}}{\text{ZS CF}} \right)^2} \right) \times \text{reconciliation factor}$$

$$\text{ZS unreconciled peak MVAR FC} = \left( \sqrt{\left( \frac{\text{ZS unreconciled coincident MWx / ZS CF}}{\text{ZS PPF}} \right)^2 - \left( \frac{\text{ZS unreconciled coincident MWx}}{\text{ZS CF}} \right)^2} \right)$$

### Radial Network Coincidence Factors (CF)

$$\text{Bulk Supply Point CF year 0 MW} = \frac{\text{BSP coincident MW0}}{\text{BSP peak MW0}}$$

$$\text{Bulk Supply Point CF year X MW (year 1-10)} = \frac{(\text{sum (ZS unreconciled coincidence MWx) / sum (ZS unreconciled peak MWx)})}{(\text{BSP peak MW0 / sum (ZS peak MW0)})}$$

### Reconciliation (Trim) Factors

$$\text{System Reconciliation Factor} = \frac{(\text{system MW FC} - \sum \text{ZSu unreconciled coincidence MWx} \times \text{system MW LF})}{\sum \text{ZSr unreconciled coincidence MWx} \times \text{system MW LF}}$$

### Loss Factors (LF)

$$\text{System MW Loss Factor} = \frac{\text{Energex or Ergon adjusted coincident MW0}}{\sum \text{ZS adjusted coincident MW0}}$$

$$\text{System MVAR Loss Factor} = \frac{\text{Energex or Ergon adjusted MVAR0} + \sum(\text{BSP compensation MVAR0}) + \sum(\text{ZS compensation MVAR0})}{\sum (\text{adjusted coincident uncompensated ZS MVAR0})}$$

$$\text{BSP MW Loss Factor} = \frac{\text{Adjusted coincident BSP MW0}}{\sum \text{attached coincident adjusted ZS MW0}}$$

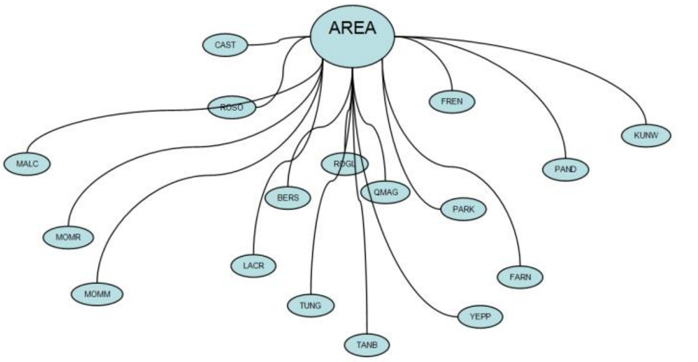
$$\text{BSP MVAR Loss Factor} = \frac{\text{Adjusted coincident uncompensated BSP MVAR0} + \sum(\text{attached coincident compensation ZS year 0})}{\sum (\text{adjusted coincident uncompensated ZS MVAR0})}$$

### Peak Power Factors

$$\text{Peak Power Factor (ZS PF) for 4 values (SD,SN,WD,WN)} = \text{Avg} \left( \frac{\text{ZS Adjusted peak MW}}{\left( \sqrt{(\text{ZS Adjusted peak MW})^2 + (\text{ZS Adjusted UnCompensated peak MVAR})^2} \right)} \right) \text{ for each period over all available load history}$$

**Abbreviations**
BSP – bulk supply point
FC – forecast
LF – loss factor
MBSPr – mesh bulk supply with a set “trim” flag
MBSPu – mesh bulk supply without a set “trim” flag
MW – mega watt
MW0 – mega watt at year zero
MWx – mega watt at forecast point
Reconciled – unreconciled value X reconciliation factor
MVAR0 - MVAR at year zero
MVARx – MVAR at forecast point
SS – switching station
ZS – zone substation
ZSr – zone substation with a set “trim” flag
ZSu – zone substation without a set “trim” flag

# Mesh Area Forecasting Equations




## Mesh Area Loss Factors (LF)

$$\text{Area MW Loss Factor} = \frac{\sum \text{attached coincident adjusted MBSP MW0}}{\sum \text{attached coincident adjusted ZS MW0}}$$

$$\text{Area MVAR Loss Factor} = \frac{\sum(\text{attached adjusted coincident uncompensated MBSP MVAR0}) + \sum(\text{attached coincident compensation ZS year 0})}{\sum (\text{adjusted coincident uncompensated ZS MVAR0})}$$

## System Reconciliation (Trim) Factors

$$\text{System Reconciliation Factor} = \frac{(\text{system MW FC} - \sum \text{ZSu unreconciled coincident MWx x system MW LF})}{\sum \text{ZSr unreconciled coincident MWx x system MW LF}}$$

## Area Forecast Equations

Area unreconciled coincident MW FC = ∑ZSr&u unreconciled coincident MW x Area MW LF

Area reconciled coincident MW FC = [(∑ZSr unreconciled coincident MW x System Reconciliation Factor) + (∑ZSu unreconciled coincident MW)] x Area MW LF

Area reconciled coincident uncompensated MVAR FC = (∑(unreconciled coincident uncompensated ZS MVAR x System Reconciliation Factor x Area MVAR LF)) – (∑FC ZS & SS reconciled coincident compensation)

[If the first part of these equations = 0, then use 1]

[Compensation is the sum of the secondary capacitors at the connected mesh zone substations and switching stations]

Area reconciled coincident compensated MVAR FC = (Area reconciled coincident uncompensated MVARx) - (∑(MBSP secondary compensation + ZS & SS primary compensation))

## Area Reconciliation Factor

$$\text{Area Reconciliation Factor} = \frac{\text{Area reconciled coincident MWx} - \sum \text{MBSPu unreconciled coincident MWx}}{(\sum \text{MBSPr unreconciled coincident MWx}) \text{ x System Reconciliation Factor}}$$