Review of Transpower's electricity demand forecasting methods

Professor Rob J Hyndman

B.Sc. (Hons), Ph.D., A.Stat.

Contact details:

Telephone: 0458 903 204

Email: robjhyndman@gmail.com

Web: robjhyndman.com

ABN: 99 118 173 027

Report for

Transpower New Zealand Ltd

22 September 2011

Contents

Background

This report provides a brief review of three documents:

1. Transpower's national and regional demand forecast document (May 2011);
2. NZIER's review of Transpower's document (June 2011);
3. an emailed description of Transpower's GXP-level demand forecast (September 2011).

This is a methodological review. I have not attempted to review the implementation of the forecasting methods employed, nor to reconstruct the forecasts using the methods used by Transpower. Instead I have looked at the description of the process and methods used in order to produce the forecasts, and compared these with the best practice implemented elsewhere and as reported in the academic literature.

1 Transpower's national and regional demand forecasts

The Transpower forecasting process is described in "Draft demand forecast" (May 2011)[1]. In general, the process is satisfactory, uses sound statistical principles and appears to achieve reasonable results for both point forecasts and estimates of forecast uncertainty. My comments are mostly restricted to areas where I think improvement is possible, or clarification is required.

Design principles and purpose

1. The Transpower forecasts are used in grid planning as inputs to models that simulate the power system. Consequently, it is important that the forecasts represent the probability distribution of future demand. This is partially allowed for by Transpower in that the mean and 90th percentiles of the future demand distributions are computed. While this information will be useful for some purposes, a more robust approach is to simulate the full future demand distributions. The simulation-based approach used by Transpower makes it possible to estimate the full distribution, so perhaps this is already done but not reported.

2. The report suggests (p6) that simple load duration curve projections can be produced by scaling an historical reference curve. There is not enough information in this statement for me to be sure that it is an appropriate approach. The scaling factor would need to take account of changes in economic activity, population growth, responses to climate change, etc. Further, the curve may change shape as a result of these factors, so a simple scaling approach may not be appropriate. I would prefer to see a fully stochastic model of demand being used in order to produce future demand distributions.

3. Electricity demand is subject to a range of uncertainties and contributing factors. The electricity demand experienced each year will vary widely depending upon prevailing weather conditions (and the timing of those conditions), economic activity, population growth, changing customer behaviour, responses to changing climate and electricity prices, changes in technology and measures to reduce carbon intensity, as well as the general randomness inherent in individual usage.

   The Transpower forecasts do not attempt to combine all these factors into one model. Instead, they use a combination of several forecasting models, thus providing an "ensemble approach" to obtaining forecasts. The ensemble approach is widely used in other contexts and is an appropriate way to incorporate model uncertainty as well as to improve forecast accuracy.

4. The effect of demand predictors will vary across the year, and so it is appropriate to produce separate models for winter, summer and shoulder seasons. This is standard practice in other contexts (e.g., Hyndman and Fan, 2010). There is a trade-off to be made between having a large enough period for model estimation, and a small enough period for the model to be relevant for the entire period. In Australia, I have used periods of about 3–4 months of each year for model estimation purposes. This seems consistent with the Transpower approach.

[1] Downloaded from
http://www.gridnewzealand.co.nz/f4847,54542693/transpower-demand-forecast-May-2011.pdf

5. Transpower uses 14 years of data to produce thirty years of forecasts. As a general rule, I teach my students never to forecast ahead more than half the length of the historical data. I realise that often longer-horizon forecasts are demanded, and that there is sometimes little choice other than to produce them, but I would have liked to see much more in the way of uncertainty statements and qualifications for the very long-term forecasts. Beyond about seven years, the forecasts are probably sufficiently unreliable as to be misleading rather than helpful.

6. While it appears that half-hourly demand data are available, this data is not used directly. Instead, annual energy demand is modelled, and then the results are converted into seasonal and regional averages, peaks and troughs. In some cases, this approach is probably preferable, especially when the data quality or history with regions is limited. However, if the half-hourly national demand data is available, then why not model it directly? There are several well-developed models for modelling half-hourly demand directly (e.g., McSharry et al., 2005; Weron, 2006; Hyndman and Fan, 2010)

Methodology

7. Transpower uses four high level forecasting models (HLFMs) which use very different types of information. There is considerable academic literature (e.g., Clemen, 1989; Timmermann, 2006) on the value of combining forecasts, especially when the forecasts are generated using different types of models, and the Transpower approach conforms to best practice in this respect.

8. The Transpower report does not provide sufficient information for the process to be replicated. For example, in the randomization of GDP and population, there is no indication of the probability distributions used or their parameters. I understand the purpose of this document was probably not to fully specify the forecasting process, but it would be useful if the algorithm was fully specified somewhere, and not just in the computer code.

9. The econometric HLFM is similar to that used in Australia for forecasting annual electricity demand, including the use of an offset (Tiway energy demand) and a log link function. However, one difference is in how the high correlation between P_y and G_y is handled. The solution adopted (i.e., fixing the coefficient of $\log(P_y)$ to 0.5) is one way to overcome the problems associated with the high correlation. Another, that may be worth considering in the future, is to model per-capita energy demand instead, and include per-capita GDP as a covariate. This removes the effect of population, and allows the effect of GDP to be handled separately.

10. I am concerned that the econometric HLFM does not include any weather-based covariates. In many parts of the world, temperature variation is the biggest contributor to variation in electricity demand. It would be possible to include cooling degree days and heating degree days into the model to allow for these effects. In addition, the existing model does not allow for climate change. By including temperature-based covariates in the model, some scenarios for climate change may be incorporated. As currently implemented, there seems no way to explore the possible impacts of climate change on electricity demand.

11. The procedure for producing peak forecasts from energy forecasts seems very ad hoc. Further, the assumption of a constant ratio between peak demand and energy demand is not justified, and probably inaccurate. In Western Australia, for example, this ratio has been increasing over the last ten years.

12. The endogenous HLFM assumes a simple linear time trend. This is a very strong assumption that has not been checked as far as I can see. A more realistic endogenous model would be a random walk with drift. It would give a similar time trend, but with much more uncertainty in the forecasts.

13. I am not surprised that temperature adjustment was not effective. Temperatures have a partially cumulative effect on energy usage due to thermal inertia in buildings. For example, a hot day in Australia has a much greater effect on electricity demand if it has been preceded by a hot week than if it has been preceded by a cool week. So simple temperature adjustment is not usually effective. Instead, the complicated relationship between temperatures and demand needs to be modelled, allowing for lagged effects and nonlinear relationships (e.g., Hyndman and Fan, 2010).

14. The use of judgemental forecast methods (the "Ad Hoc HLFM") requires a systematic process, such as the Delphi process (Rowe and Wright, 1999), in order to ensure the forecasts have a sound basis and are replicable (to some extent). It is not clear that the process used for the ad hoc HLFM is sufficiently systematic to be considered a reliable judgemental forecast.

15. The distribution of forecasts from the Ad Hoc HLFM is tri-modal (see Figure on p30) because the possible growth rates are far apart compared to the added noise. A better solution would be to have a distribution of growth rates based on the judgemental assessments. For example, if L, M and H represent the low, medium and high "world views" in a particular future year, then the demand could be modelled using a triangular distribution with density given by

$$f(d) = \begin{cases} 0 & d < L; \\ \frac{2(d-L)}{(H-L)(M-L)} & L \le d < M; \\ \frac{2(H-d)}{(H-L)(H-m)} & M \le d < H; \\ 0 & d \ge H. \end{cases}$$

This would avoid the trimodal distribution, while still allowing the judgemental forecasts to be taken into account.

16. My only concern with the process of allocation outlined in Section A.10 is the use of judgemental standard deviations to add uncertainty. I suggest some effort is made to document the process used and to find ways of making this more objective. I am not doubting the choices made, but a reliable forecasting system should be replicable and testable. The more ad hoc and judgemental decisions, the harder it is to test if the system is working appropriately.

17. The change in summer/winter ratio over time (top of p.33) is probably due to changing electricity usage in meeting heating and cooling needs. For example, an increase in air-conditioning will lead to a relative increase in the summer peak, and more efficient heating will lead to a relative decrease in the winter peak.

Validation and evaluation

18. The Transpower forecasts do not seem to be subject to rigorous and regular validation, apart from the one-year backcast outlined in Section C.3. It would be difficult to fully validate the forecasts given the number of judgemental decisions that are made. However, this is an important part of forecasting and provides an assurnace that the processes are producing reliable and realistic forecasts.

    A fully automated forecasting procedure, such as that outlined in Hyndman and Fan (2010), can be validated by applying it to any previous year and checking the forecasts against actual demand. This can be done for each year in the historical data, and the results provide a measure of the validity of the forecasts. This is particularly important for peak demand such as a P90 value, where, by definition, the result is unusual. If there are fifteen years of historical data available, then it is reasonably likely that in one of those years, demand will have exceeded the P90 value. If several years are above the predicted P90 values, then the forecasts of P90 are probably too low. This sort of validation is important to ensure the forecasting algorithm is working properly, and to build trust in the forecasts by users.

19. I suggest some efforts be made to reduce the judgemental aspects of the algorithm, and to develop a rigorous validation of the results by comparing forecasts against many previous years, not just one year.

2 NZIER's review of Transpower's electricity demand model

NZIER provided a review of the Transpower Electricity demand model in June 2011 in the document "Electricity demand model review" by Yang and O'Connor[2]. In general, NZIER provide some very useful advice and comments. In most cases, I have not provided comments where the NZIER review is well argued and helpful. I have tended to concentrate on the small number of points where I have minor disagreement, or where some additional comments may be useful. This should not detract from the high quality review provided by NZIER.

20. p.3. The comments regarding ensemble forecasting seem to reflect weather forecasting practices rather than what is done by Transpower. The Transpower ensemble consists of four very different models, and future sample paths are simulated by randomizing inputs, error terms, and the model itself. This is different from weather forecasting ensembles where one large model is used with perturbed initial conditions.

21. p.6. The use of a 10% POE (not 90% POE), corresponding to the 90th percentile, is standard in the electricity industry and represents a 10% probability of exceedance. So the 10% POE forecasts of the peak demand gives a series of values that will only be exceeded on average once every ten years.

    This does not correspond to an 80% *confidence* interval, but to an 80% *prediction* interval, which is common in forecasting (e.g., Makridakis et al., 1998). The usual 95% confidence interval arose out of a statistical testing paradigm rather than a prediction paradigm, and I can see no reason why it should be used here.

22. p.6. I think the authors meant "prediction interval" when they wrote "confidence interval".

23. p.6. Using multiple models is not a substitute for forecast validation. All the models might be incorrect but similar. The only effective way to validate forecasts is to compare the forecasts against actual values. This can be done by withholding some historical data (e.g., one year) and try to predict that data with a model fitted to the remaining data.

24. p.7 (second last bullet point). The expected path of *each* model is similar or identical to the point forecasts from that model. However, the expected path of the ensemble takes account of *all* models. There is abundant evidence of the value of combining point forecasts (e.g., Timmermann, 2006), especially when they involve a variety of models.

25. p.7 (second last bullet point). There is also no assumption of symmetric variability here. The econometric HLFM, for example, assumes multiplicative log-Normal errors (equivalently Normal errors on the log scale) which are far from symmetric.

26. p.7 (last bullet point). I don't understand the comment on the impact of randomization on total peak usage, and its connection with negatively correlated variables.

27. p.13. I think the authors meant "prediction interval" when they wrote "confidence interval".

[2] Downloaded from http://www.gridnewzealand.co.nz/f4847,54542833/nzier-review-of-transpower-forecast.pdf.

28. p.13. It is not clear what model was used to determine the variance for the GDP and population forecasts, but it sounds like a random walk model was used to estimate the variance, and the GDP forecasts from Statistics NZ and NZIER were used for the point forecasts. Assuming I have guessed correctly, this seems appropriate, but I would have liked it specified more precisely. On the other hand, the size of the prediction intervals shown in the figures look too narrow to have come from such a model, so I'm not sure what has been done.

29. pp.13–14. For the regression models for peak demand usage, tests of the statistical significance of covariates are not the best way to choose predictors in a model. Such tests were designed for a purpose other than variable selection. Instead, cross-validation or an approximation to it such as the AIC is much better (Harrell, Jr, 2001). It is possible to have variables that are statistically significant but make the forecasts worse, and variables that are not significant, but which would make the forecasts more accurate.

30. p.16. I do not think that the smaller variation is due only to the exclusion of model variance. Even for a single model, the variance shown appears to be too small. I suspect this has something to do with the assumed variances for future values of GDP.

3 Transpower's GXP-level demand forecasts

Transpower provides some GXP-level demand forecasts for transmission planning purposes. A description of the forecasting process was sent to me in an email from Brian Bull on 6 September 2011. Not enough information was provided for me to know the details of the methodology. Nevertheless a few comments are provided.

31. The use of a reference load duration curve in order to estimate the load shape at a GXP is prone to error. Different GXPs will have different load shapes, but this approach assumes they are all the same. The fact that some local lines companies are known to switch substantial amounts of load between two or more nearby GXPs demonstrates the problems that can arise. If load shapes are important, then I suggest they be studied directly and not via reference curves.

32. The validation of last year's forecasts against this year's demand is encouraged. One of the best ways of learning how to improve a forecasting system is to identify where the largest errors have been in previous forecasting exercises.

33. The use of a Matlab script is *much* better than using spreadsheet calculations. I strongly encourage the use of scripting for all aspects of the forecasting process. Not only does it make the process replicable, it also helps eliminate errors, makes it possible for other people to see precisely what has been done, allows similar methods to be tested on different data sets, and so on. I use R for a similar purpose. The tool is not so important as the process, and Matlab is well-suited for this type of work.

34. I am surprised that even simple statistical methods cannot be used at the GXP level. One technique may be to study the proportion of load within a region that is due to a particular GXP over time. These proportions are often easier to forecast than the GXP-level loads. If there are step changes present, then only the most recent proportions may be applicable. Alternatively, it may be possible to model average GXP load over time, and to model the relationship between peak and average demand within a region. Average loads are much better behaved than peak loads, making them easier to model.

35. There is a necessary level of ad hoc judgements required in forecasting at a highly disaggregated level where local conditions have a large effect on the local demand. So I am not concerned by the subjective nature of these forecasts. Nevertheless, I would continue to explore ways of using more objective and statistical methods where possible. A useful process is to compare what the statistical forecasts would have predicted compared to what was actually predicted, once the data in the forecast period are available. Provided the judgemental methods continue to do better, on average, than the statistical methods, they should continue to be used. At the same time, ongoing efforts can be made to try to develop statistical methods that will eventually out-perform the judgemental forecasts.

References

Clemen, R. (1989) Combining forecasts: a review and annotated bibliography with discussion, *International Journal of Forecasting*, **5**, 559–608.

Harrell, Jr, F. E. (2001) *Regression modelling strategies: with applications to linear models, logistic regression, and survival analysis*, Springer, New York.

Hyndman, R. J. and S. Fan (2010) Density forecasting for long-term peak electricity demand, *IEEE Transactions on Power Systems*, **25**(2), 1142–1153.

Makridakis, S. G., S. C. Wheelwright and R. J. Hyndman (1998) *Forecasting: methods and applications*, John Wiley & Sons, New York, 3rd ed.

McSharry, P. E., S. Bouwman and G. Bloemhof (2005) Probabilistic forecasts of the magnitude and timing of peak electricity demand, *IEEE Transactions on Power Systems*, **20**, 1166–1172.

Rowe, G. and G. Wright (1999) The Delphi technique as a forecasting tool: issues and analysis, *International Journal of Forecasting*, **15**, 353–375.

Timmermann, A. G. (2006) Forecast combinations, in G. Elliott, C. W. J. Granger and A. G. Timmermann (eds.) *Handbook of Economic Forecasting*, chap. 4, pp. 135–196, Elsevier.

Weron, R. (2006) *Modeling and forecasting electricity loads and prices: a statistical approach*, John Wiley & Sons, Chichester, UK.