Journal of Energy in Southern Africa 26(4): 48–58
DOI: http://dx.doi.org/10.17159/2413-3051/2014/v25i4a2238

Application of multiple regression analysis to forecasting South Africa's electricity demand

Renee Koen

Jennifer Holloway

Decision Support and Systems Analysis Research Group, Built Environment Unit, CSIR, Pretoria, South Africa

Abstract

In a developing country such as South Africa, understanding the expected future demand for electricity is very important in various planning contexts. It is specifically important to understand how expected scenarios regarding population or economic growth can be translated into corresponding future electricity usage patterns. This paper discusses a methodology for forecasting long-term electricity demand that was specifically developed for applying to such scenarios. The methodology uses a series of multiple regression models to quantify historical patterns of electricity usage per sector in relation to patterns observed in certain economic and demographic variables, and uses these relationships to derive expected future electricity usage patterns. The methodology has been used successfully to derive forecasts used for strategic planning within a private company as well as to provide forecasts to aid planning in the public sector. This paper discusses the development of the modelling methodology, provides details regarding the extensive data collection and validation processes followed during the model development, and reports on the relevant model fit statistics. The paper also shows that the forecasting methodology has to some extent been able to match the actual patterns, and therefore concludes that the methodology can be used to support planning by translating changes relating to economic and demographic growth, for a range of scenarios, into a corresponding electricity demand. The methodology therefore fills a particular gap within the South African long-term electricity forecasting domain.

Keywords: long-term forecasting, South African electricity demand

1. Introduction

In a developing country such as South Africa, understanding future patterns of electricity usage is very important in various planning contexts. The future national demand for electricity is an important consideration for electricity providers who need to plan to have sufficient and secure supply of electricity (Imtiaz *et al.*, 2006 and Soontornrangson *et al.*, 2003). Investment in electricity generation capacity, whether using fossil-based or renewable energy sources, is largely motivated through the anticipated long-term need for electricity (Ulutaş, 2005; Doriana; Franssen and Simbeck, 2006). South Africa has a growing population which creates an increasing need for housing and services, as well as a need to expand economic activity both to accommodate new entrants into the labour market and to address current high unemployment levels. Therefore, virtually all types of national and local planning, public or private, requires consideration of the implications on future electricity needs in order to establish whether there is sufficient electricity supply capacity in the country to support future plans.

This paper discusses a methodology for forecasting long-term electricity demand that was initially used for assisting with strategic planning for the South African branch of a multi-national company. The initial modelling objective expressed by the company was to be able to determine potential fluctuations in the future national demand for electricity, i.e. the amount of electricity required from the national grid, in order to assess the impact of that on their own business plans. This required a methodology that could determine the effect of possible changes in various political, demographic or economic patterns on the future national electricity consumption patterns. Therefore, forecasts using extrapolation of past trends would potentially not suffice, and a type of scenario-based forecasting was foreseen to be more appropriate.

A methodology was developed that satisfied these needs expressed by the company and that produced scenario forecasts that could be successfully incorporated into their strategic planning processes. During the development of this methodology, various electricity forecasting studies published locally and internationally were consulted, but it was found that a scenario-based methodology using multiple regression models to forecast electricity demand in various electricity usage sectors had not been applied before. Furthermore, the extensive collection of public domain data and the interrogation process applied in order to create a usable dataset out of the various information sources had not been found in any other study. In addition, the South African electricity demand and supply patterns, and the driving forces behind them, are different from that of other countries. Therefore, this methodology could be viewed as unique, both locally and internationally.

The methodology described in this paper was applied successfully within the company it was developed for, and the same basic methodology has subsequently been used to support planning of electricity supply needs within the public sector.

This paper first discusses the objectives of the forecasting and then provides more details with regard to the data collection and validation, as well as the modelling methodologies used. This is followed by descriptions of the models used and the forecasts derived from the models. The paper concludes with a discussion of the chosen methodology as it compares to approaches used in other studies, as well as comments about the usefulness of the methodology.

2. Forecasting objectives

The main objective to be met by the forecasts, as expressed by the initial client, was to estimate the future demand for electricity from certain expected changes in the national economy and demography. This meant that a forecasting model (or set of models) had to be developed that would be able to translate aspects such as economic growth or decline into subsequent growth or decline in electricity usage. The focus was therefore placed on the development of a model(s) that could quantify historical patterns of relationships between electricity usage and the relevant economic and demographic variables, using data available in the public domain. Ultimately, the objective was to use such a model(s) to derive future electricity usage patterns from these quantified relationships once expected future values for the relevant economic or demographic variables had been estimated.

Since these objectives required the quantification of historical patterns as a basis for future forecasts, a statistical modelling and forecasting approach seemed appropriate. However, in a statistical modelling approach it would not only be important to consider the correlation between electricity usage and the variables used to predict electricity usage, but also to consider the correlation of the predictor variables with each other. Such correlation between predictor variables is called multicollinearity or near-linear dependence (Montgomery, Peck and Vining, 2006). Including variables that are highly correlated with each other as predictors in the same model can be problematic when that model has to be used for forecasting. This is especially true for scenario forecasting, since there are no guarantees that the historical relationships between variables would be maintained when creating future scenario inputs. For instance, one may want to purposefully create a scenario in which variables do not follow the same patterns as in the past. If a model exhibits high multi-collinearity, violating such relationships in the created scenario inputs could then invalidate the model's outputs. Therefore, an important consideration of the methodology development was to ensure that the models would be developed in a way that they were statistically valid, i.e. that multi-collinearity between the predictor variables used to estimate electricity demand would be managed correctly.

3. Forecasting methodology overview

In developing a statistical model(s) to use as a basis for scenario forecasting, an attempt was first made to use predictor variables to forecast future annual demand for electricity at a national level. Data on various external factors, such as Gross Domestic Product (GDP), population, electrification of households, major industrial projects (using start-up and shut-down dates) and climate variables, as well as relevant derivations and transformations of these variables, were collected and analysed. A particular problem experienced with trying to forecast the total national demand, however, was the consistent problem of multi-collinearity that was measured in any model that contained GDP and population, or transformations of these two variables. This made it difficult to develop an appropriate model that would support forecasts from scenarios that contained both of these variables.

Instead, the approach was adapted by breaking the total electricity consumption up into sectors of electricity usage, forecasting the consumption per sector and then combining these sector forecasts into a total annual forecasted demand for the country. Losses also had to be estimated and incorporated into the forecasted total. In this manner, each sector could have its own set of drivers (predictors) that were appropriate for the electricity consumption in that sector, and sectoral models could be derived that had acceptable levels of multi-collinearity. Consultation with different experts in the field of electricity consumption forecasting also confirmed that forecasts via different electricity usage sectors give better results than directly forecasting total consumption values. The main challenge was to find reliable historical values for sector consumption to use as a basis for these forecasts, together with historical values of potential predictor variables. The problems encountered during data collection and verification are discussed in more detail in section 4, but it may be summarised by saying that reliable data on electricity consumption per sector is very difficult to find.

Multiple regression modelling was chosen as the forecasting technique for each sector as this has been noted to be the most appropriate statistical technique for long-term forecasting (Makridakis, Wheelwright, and Hyndman, 1998). There have been previous reported studies where multiple regression has been used for long-term forecasting of electricity consumption in other countries, either for total consumption (Bianco, Manca and Nardini, 2009; Mohamed and Bodger, 2005; Egelioglu, Mohamad, and Guven, 2001) or for a specific sector (Al-Ghandoora *et al*, 2008). These models used different drivers for forecasting, as appropriate within the specific country, but employed the same regression technique as discussed in this paper. Details on the regression models are provided in section 5.

4. Electricity demand and 'driver' data

A very important component of the regression modelling involved the collection of appropriate data for the relevant variables required.

Data on national electricity consumption in South Africa from 1978 to 2010 was obtained from Statistics South Africa (Stats SA), from the series of monthly publications under the Statistical Release series P4141 – Generation and consumption of electricity, available from the website (StatsSAweb). The specific time series used is defined as the 'electricity available for distribution in South Africa'. Although this data is not broken down into electricity usage sectors, it was a valuable dataset to use to reconcile other datasets containing sector data. In order to obtain an adequate data series on electricity consumption per sector, it was necessary to obtain data from various sources as no single source could give data per sector from 1972 until present. When the forecasting methodology was initially developed, and subsequently updated, the following potential sources of sector data were identified:

- Statistics South Africa (some sectors, 1964–1984) (StatsSAweb)
- Rand Afrikaans University (RAU) (now University of Johannesburg) (1972-1991) (Cooper and Kotze, 1992)
- Department of Minerals and Energy (DME), 1998 Digest (now Department of Energy) (1990-1997) (Cooper, 1998)
- Department of Minerals and Energy (DME), 2000 Digest (now Department of Energy) (1992-2000) (DIGEST2002)
- Department of Minerals and Energy (DME), 2006 Digest (now Department of Energy) (1992-2004) (DIGEST2006)
- Department of Minerals and Energy (DME) Energy Balance Spreadsheets (now Department of Energy) (1992-2009) (ENERGYBALANCES)
- National Energy Regulator of South Africa (NERSA), Electricity Supply Statistics of South Africa (1996-2006) (ESS)
- South Africa Energy Statistics (Vol. 1 and Vol. 2) (1950-1993) (SAES1, SAES2)
- Eskom Annual Reports (1997-2010) (Eskom)
- Eskom Statistical Yearbooks (1957-1996) (Eskom Year Book)

The various sources had to be checked against each other and against the Stats SA national totals. Although the total consumption figures were roughly consistent between the various sources, there were large discrepancies for many of the data points (years) at the individual sector level. This is partly due to the mismatch between the number of sectors used in each source, but also to the inconsistency between sector definitions used in the different sources for some sectors. In order to use and compare data from the different sources, the sectors had to be aligned to one common set of sectors, and for this purpose, the sectors used by NERSA, and also reported by Eskom, were used as the standard. Although the Eskom Annual Reports provide a reflection of the electricity distribution by sectors, these sectors are broken down only for Eskom's direct customers and not for the electricity that is redistributed by the municipalities. Eskom has a large sector for redistributors that is listed in their reports, but this electricity consumption cannot be broken down by Eskom into the typical economic activity sectors. However, for the purposes of our sector forecasting and in the absence of suitable data from other sources, certain sector data was estimated using Eskom's electricity consumption per sector together with the historical estimates of Eskom's percentage share in the sector, as reported by NERSA.

The data discrepancies between the different data sources were considered and potential reasons were sought for these discrepancies. Where the reasons for discrepancies could not be ascertained, representatives from the data sources were contacted in order to obtain clarity on definitions and to gain understanding as to reasons for differences between sources. Examples of data issues that arose from the discussions include:

- The NER (now NERSA) data was collected from Eskom and municipalities, but Eskom has a different financial year to that of the municipalities, with the result that the data from the two sources are not aggregated over the same periods for the year being reported.
- Municipalities did not use the NER categories on their own systems, and therefore every year they had to match their data to the categories provided by NER before submitting the data, often leading to the same municipal client being classified differently in different years. This was particularly true of the commerce and manufacturing sectors.
- The Platinum mining sector was initially not included under mining in all sources, but was classified by some as industrial based on the platinum processing plants found on mining sites.
- Data from DME (now Department of Energy) was not always consistent across years when certain users were first classified into a 'non-specified other' category and later allocated to industry, commerce and residential. If a data version was published before this re-allocation, the sector data would be incorrectly reflected in the published version.
- In one or two sources, changes were made to definitions without adjusting the data 'backwards' to match the definition change.

After an extended period of data checks and consultation, the most reliable data series to use for each sector was selected, mostly using a combination of sources. All the recommended sector data series, once confirmed, were added together, distribution and transmission losses were added, and when checked against the Statistics SA national consumption figures the recommended total was found, for the majority of the years, to be within 1% of the total national electricity consumption, with only a few years differing by an amount close to 2.5% of the total. Note that the NERSA and Eskom data had to be adjusted for those years where the financial year did not coincide with the calendar year, with the former being the time period used for reporting and the latter being the time period used for our analysis. Adjustments were also made to the NERSA data to align it better with data published in Eskom's annual reports.

The graphs in Figures 1–5 provide an illustration

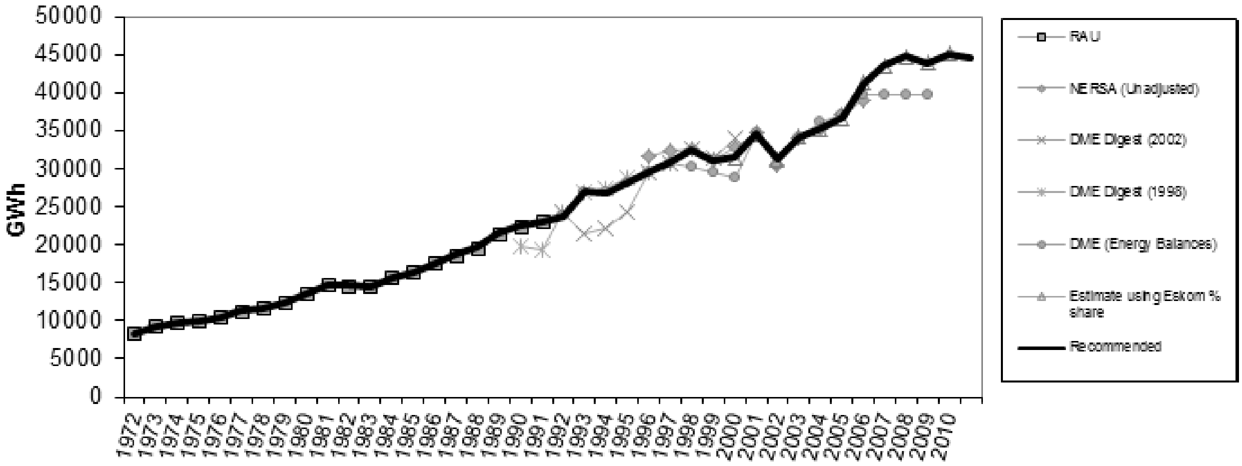

Figure 1: Comparing agricultural sector data between different sources

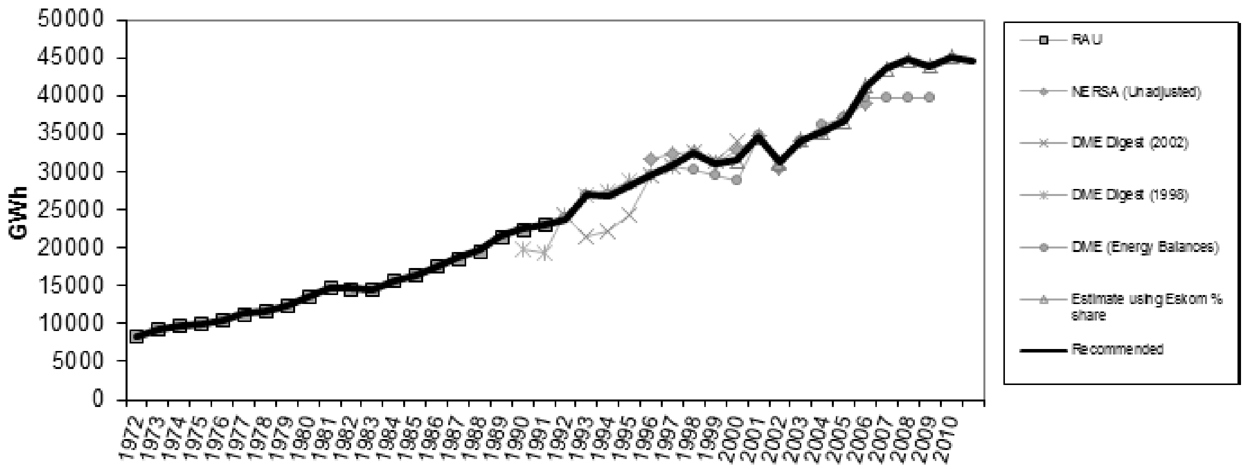

Figure 2: Comparing domestic sector data between different sources

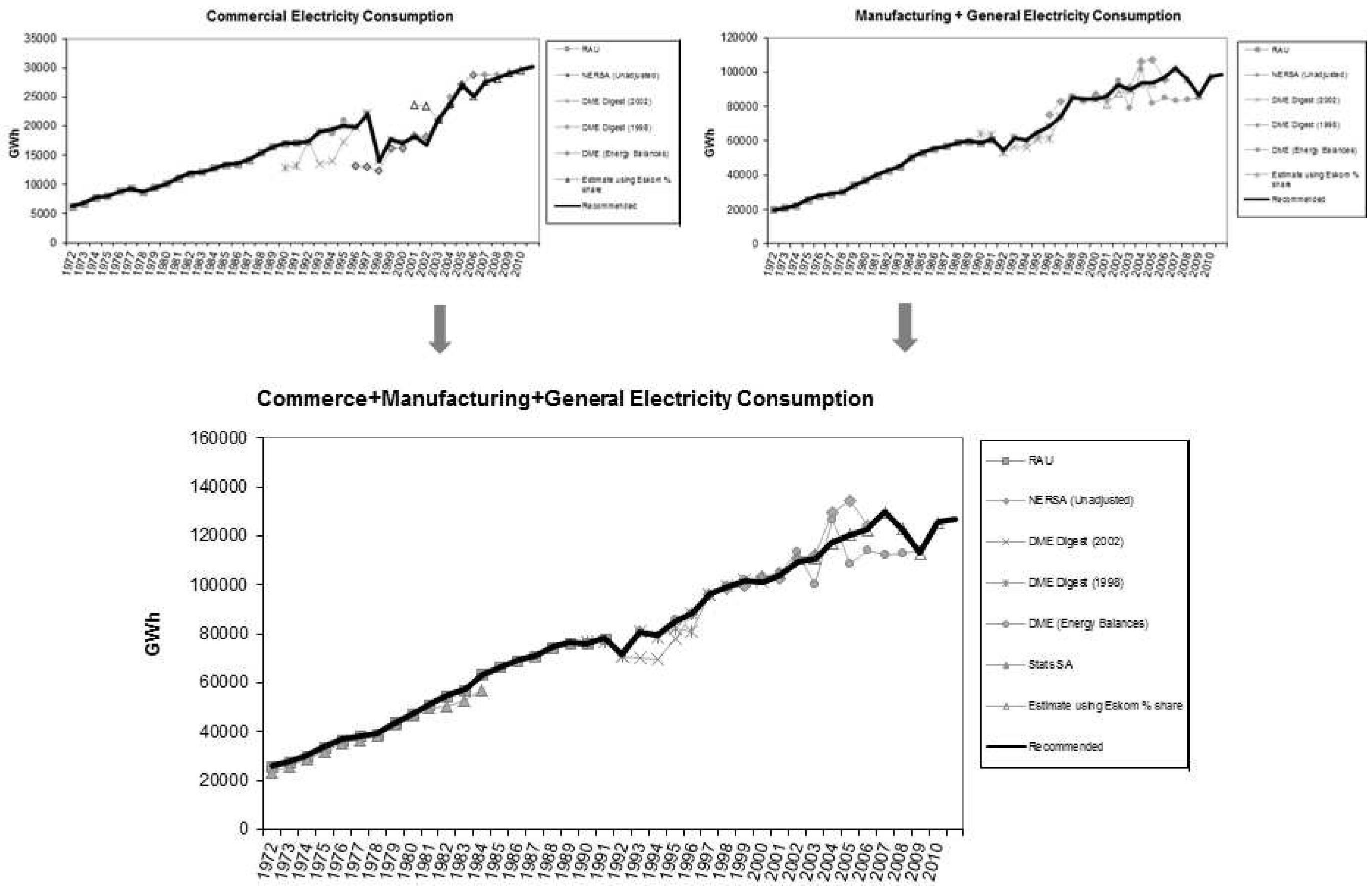

Figure 3: Combining and comparing commercial, manufacturing and general sector data between different sources

of the various data sources consulted, and the relative differences between them. The thick black line indicates the data pattern that was considered to be a reliable estimate for the sector and these patterns were therefore used as a basis for the forecasting.

Although most sources provide 'Commerce' and 'Manufacturing' (also referred to as 'Industrial') sectors, definitions differed widely between them, and even between different years of the same source. Consultation with representatives of the data sources also confirmed that differentiating between commerce and manufacturing within municipal customers was problematic and could change from year to year. Furthermore, most sources contain a 'general' category, and the definition of this category was also found to be inconsistent between sources. However, Figure 3 shows that when data on 'commerce', 'manufacturing' and 'general' sectors were combined for each of the various sources, the differences between the sources were reduced.

It can therefore be seen that the collection and selection of appropriate data for electricity consumption per sector from public domain sources was not a trivial task. However, it was considered necessary to develop a consistent and reliable set of historical data on which to apply the chosen methodology.

Data on predictor variables were also collected, but there were fewer sources for these, and sources generally had consistent patterns. Therefore, this data collection process is not discussed in as much detail as the electricity sector consumption data. Predictor data could be sourced from Statistics South Africa and the Reserve Bank of South Africa, though the electronic data download facilities on their websites (StatsSA2, SARB) and also from the Chamber of Mines (CoMines). Data on rail freight ton-kms was previously obtained from Spoornet, but has since 2003 been difficult to obtain.

5. Model development

An important step in the model development was to select a range of appropriate economic and demographic variables that could potentially affect electricity usage in the sectors, or could be proxies for the patterns observed in the sectors' electricity usage. The next step was then to collect the historical data for these variables, as discussed in the previous section. Potential variables included population figures, GDP or Gross Value Added (GVA) values per sector, mining production volumes, and so on. By investigating the strength of the statistical relationship between each of the potential predictor

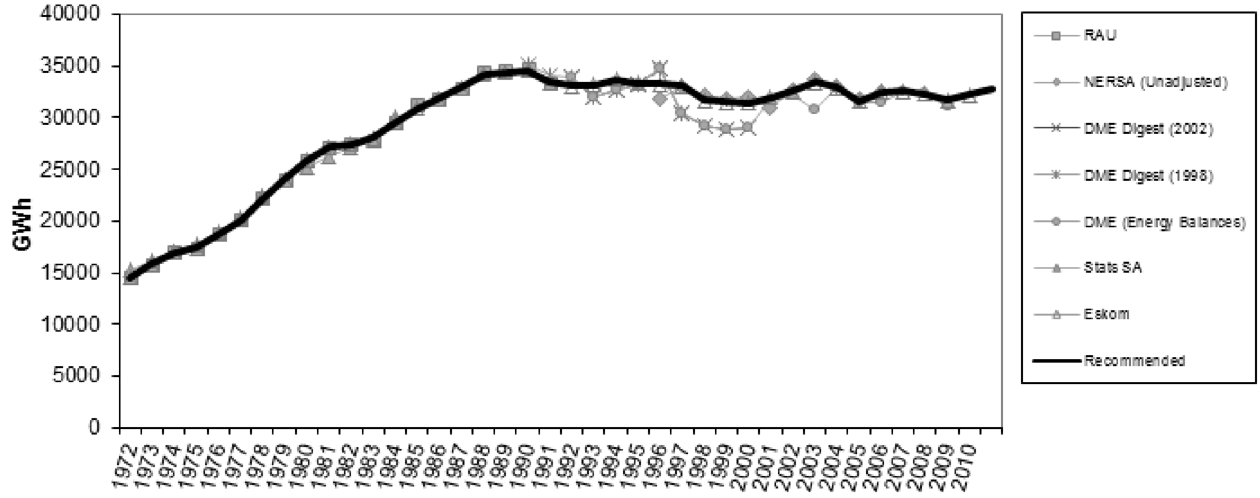

Figure 4: Comparing mining sector data between different sources

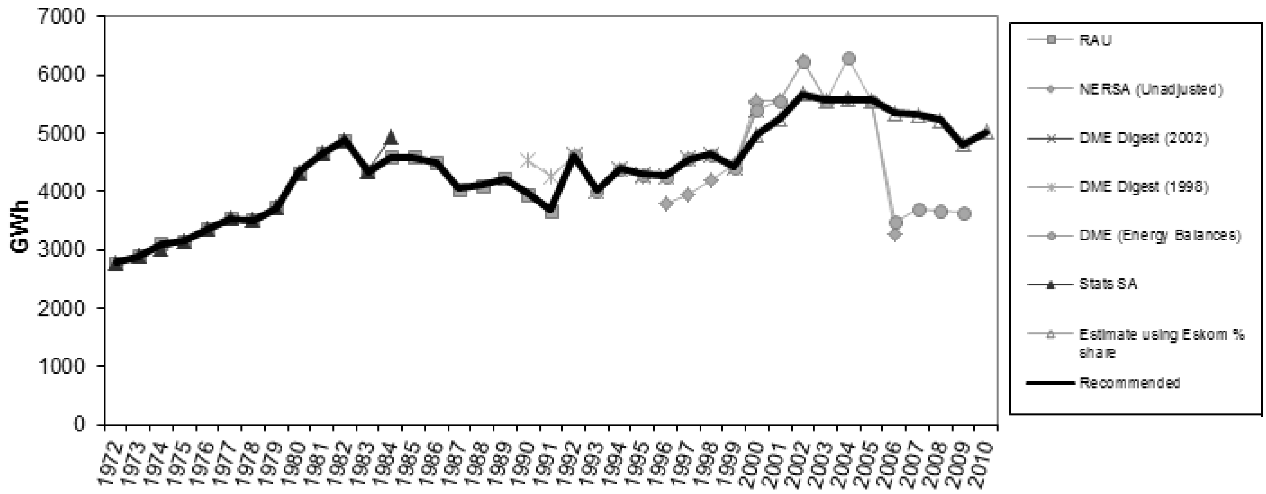

Figure 5: Comparing transport sector data between different sectors

variables and the electricity usage per sector, a smaller subset of these variables could be identified for inclusion in the final set of sector models.

The methodology followed to derive the final forecasts of total electricity consumption consequently involved an aggregation of several regression forecasts, with each sector having its own regression model. Scenarios were used to quantify the future values of various predictor variables that were identified during the regression modelling phase. Each scenario produced its own set of sector forecasts that could be added together and adjusted for estimated losses, on both distribution and transmission, to create forecasts for the total annual demand at a national level. The advantage of being able to visualise the electricity forecasts for each economic sector in a scenario, in addition to the total electricity forecast, is that one can assess the relevance and compatibility of the models and their outputs to the scenario descriptions.

In order to determine the statistical validity of the various regression models used for each electricity usage sector, the following factors were considered:

- The model had to be a statistically acceptable quantification of the relationships in the historical data, which meant that the included predictor variables had to be as few as possible but had to provide a good overall description of the electricity usage values over the period of historical data available. The goodness of fit was measured with the R^2 (correlation coefficient) and adjusted R^2 measures: the higher the R^2, the better the fit. (Note that relationships were assumed to be linear, and if non-linear relationships were found a relevant transformation, such as a logarithmic transformation, were applied to linearise the relationship).
- Residual patterns for the various model options were also considered. Residuals are defined as the difference between the values predicted from the model for a particular year and the electricity usage actually measured in that year. Very large residuals or residuals that seem to show a pattern that is not random could be an indication that the model does not fit well or that an important predictor variable was not included in the model.

- Models had to be selected in which the predictor variables showed low levels of multi-collinearity. This is measured with the condition index value – the lower the condition index, the better. A condition index of between 5 and 10 indicate low levels of multi-collinearity, while a condition index of between 30 and 100 indicate moderate levels of multi-collinearity in a model. (See the classic reference on multi-collinearity and the condition index (Belsey, Kuh and Welsch, 1980), or a discussion of the impact of multi-collinearity on regression models (Montgomery, Peck and Vining, 2006) for more information).
- The regression coefficients associated with each predictor variable had to be statistically significant. A 90% significance level was used for allowing variables to enter into any of the sectoral regression models.
- Each sectoral model should to some extent accommodate factors that are believed to be influencers of electricity consumption in that sector, i.e. be logically defensible.

Unfortunately, it is very seldom that a model will be equally good on all these criteria. Often, trade-offs had to be made. For instance, it may be that one model does not have the largest R^2 value, but has a lower condition index than another model.

For each of the individual electricity consumption sectors, a 'best' model was selected by applying an iterative procedure. A first potential model would be developed using stepwise regression, and then the model would be assessed against the above criteria. If the model was not found to be a good balance of all the criteria, changes would be made to the input variables and the process would be repeated and the resulting models would be re-assessed. Table 1 provides an indication of the predictor variables used in the initial version of the models, for the various electricity demand sectors, and typical values for the associated goodness of fit measures obtained.

Finally, although regression was chosen as the most appropriate statistical technique for this type of forecasting, it should be emphasised that the results from a regression model must be interpreted purely as a statistically proven inter-relationship. A significant relationship between the predictor variables and the variable they predict does not necessarily establish a cause and effect relationship between them. Also, note that the regression models were developed very specifically for the South African situation and were based on historical data collected from public domain sources for the South African situation.

6. Developing forecasts

After applying the sectoral regression modelling approach to the recommended sector data to obtain sector models, for example, those listed in Table 1, these models were then applied to a set of scenarios, comprising of forecasted values for the relevant predictor variables, to produce a set of electricity demand forecasts. Such a set of forecasts were initially produced using this methodology in 2003, using historical data up to 2002, for the identified client. Subsequently, updated forecasts were produced for the same client with the same methodology in two further studies, using historical data up to 2004, and then later using data up to 2006. The forecasts were incorporated by the client into various internal strategic planning processes and were based on scenarios that were matched to their strategic planning needs and questions.

The same methodology was later applied, using updated historical sector data and different scenario values as inputs, to provide a set of forecasts to compare to those prepared by Eskom for use in the Integrated Resource Plan developed in 2010 (IRP2). Since the project team that developed the forecasts using this methodology were from the CSIR, the set of forecasts developed for the IRP2 using this methodology was called the 'CSIR model' forecasts, to distinguish them from the ones produced by Eskom. Figure 6 indicates how the forecasts from this methodology, for the three scenarios used within the IRP ('High', Moderate' and 'Low'), compared to the actual values observed subsequent to the publication of the forecasts. Note that the long-term forecasts used in the IRP forecasts and the details regarding the scenarios can be obtained from the website of the Department of Energy .

Table 1: Summary of initial regression models used per sector

Electricity sector	*Predictor variables*	*Adjusted R^2*	*Condition index*
Agriculture	Final Consumption Expenditure by Households (also called Private Consumption Expenditure)	$R^2 = 0.96$	N/A if only 1 variable in model
Transport	Rail freight ton-kms and GDP	adjusted $R^2 = 0.76$	CI = 26.5
Domestic	Final Consumption Expenditure by Households (also called Private Consumption Expenditure)	$R^2 = 0.96$	N/A if only 1 variable in model
Commerce & manufacturing	Population and GDP	adjusted $R^2 = 0.98$	CI = 69.2
Mining	Platinum production volume index, Coal production volume index, Gold ore milled	adjusted $R^2 = 0.91$	CI = 44.7

Figure *6*: Forecasted values obtained for different scenarios compared to actual values observed

The forecasting methodology involved applying the quantified (statistical) relationships, identified from historical patterns, to different future scenarios relating to expected economic and demographic changes, in order to obtain the resulting forecasts for electricity usage. Since the forecasts are directly linked to scenarios for GDP and other predictor variables, the forecasted electricity values need to be interpreted within the context of these scenarios. For example, if a particular scenario was compiled using GDP growth figures that were very far from the actual grow patterns observed, then the electricity forecasts generated for this scenario may also be unrealistically high.

Therefore, although the 'true test' of any forecasts relates to how far the forecasted values were from actual recorded values, the problem with comparing the forecasts provided in Figure 6 with the actual values provided in the same figure is that the scenario values underlying the forecasts may differ from the actual recorded economic and demographic values. While scenario thinking can support planning and forecasting well, there are certain pitfalls to avoid when generating scenarios (nkal, Sayim, and Gönül, 2013).

Therefore, to provide a better comparison of 'forecasts' to 'actuals', figure 7 needs to be considered. In Figure 7, the model that was developed in 2007 (using historical data up to 2006), as well as the model developed in 2010 for the IRP2 (from historical data updated to 2009), were both applied to the predictor variables recorded in 2007 – 2012 to develop predicted values for national electricity consumption for these years. These predicted values were then compared to the actual electricity consumption published by Stats SA (StatsSAweb).

Figure 7 shows visually that the forecasting methodology produced forecasts close to those of the actual patterns. The forecasted values from the 2007 model differed by less than 0.5% from the actual 2007 value, by 3% from the actual 2008 value, and between 4 and 7% from the actual patterns in 2009 – 2011. The IRP2 model produced forecasts that differed by less than 2% from the actual values in 2010, less than 0.5% in 2011, and was 4.4% higher in 2012 than the actual recorded value. The 2007 model therefore clearly predicted the actual values well up to 2008, but did not manage to anticipate the 'dip' over the 2008 – 2010 period which could mainly be ascribed to the effects of the global recession (Eberhard, 2011). The effect of the recession was not reflected sufficiently in the combination of drivers incorporated into the 2007 model and therefore did not successfully translate into a dip in electricity demand.

In the IRP2 model, a decision was made to use the manufacturing production volume index rather than the GDP as a predictor in the Commerce and Manufacturing sector model, and this change resulted in a much better overall fit. Note that this sector accounts for about 50% of the overall electricity usage, and so forecasting errors in this sector affects the accuracy of the overall forecast substantially. The IRP2 model therefore gave much better predictions over the 2008 – 2011 period, but overestimated demand in 2012. Subsequent studies has shown that the decrease in 2012 may have been due to improvements in energy efficiency, which in the manufacturing context equates to less electricity being used to produce the same manufacturing vol-

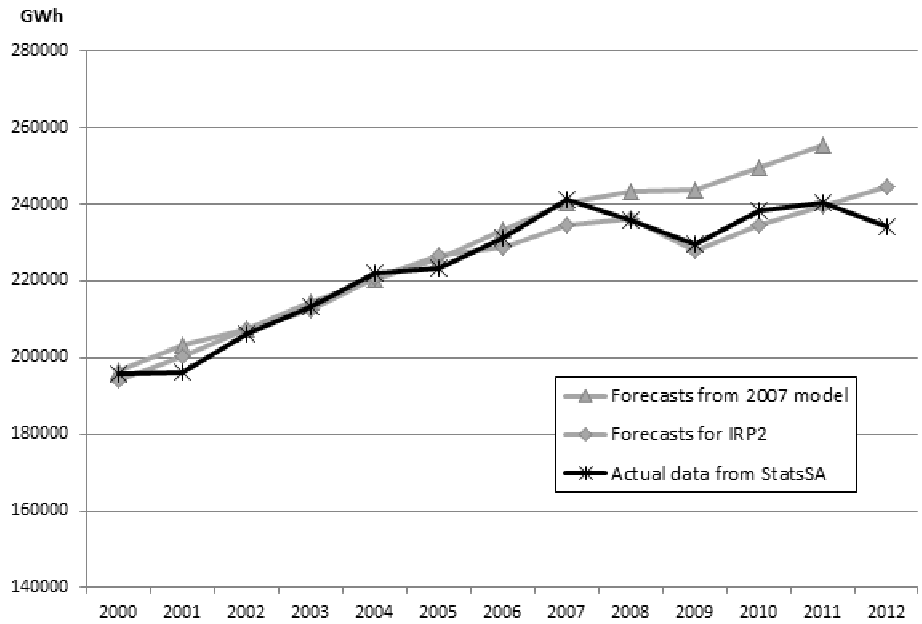

Figure 7: Comparing forecasts developed from methodology using actual predictors to actual electricity demand values measured

umes. Since these patterns have only started to appear since 2011, sufficient data is not yet available to quantify electricity efficiency improvements, but current research by the team is focusing on introducing an 'intensity factor' into some of the sectoral models.

It can therefore be said that the methodology described in this paper does provide a useful translation of economic and demographic scenarios into electricity consumption forecasts. It is, however, still necessary to continuously update these models in order to incorporate the most recent electricity usage behaviour, particularly in light of the increasing constraints on the South African energy system.

7. Concluding remarks

The methodology discussed in this paper differs from other reported studies that produced forecasts for long-term electricity demand in South Africa (Prinsloo, 2009; Van Wyk and Fourie, 2009; Inglesi, 2010). This methodology forecasts electricity within all usage sectors and then aggregates the sector forecasts to a national total, while some of the other studies (Van Wyk and Fourie, 2009; Inglesi, 2010; Inglesi and Pouris, 2010) directly forecast demand at a national level. Forecasts produced by Eskom, such as those that are provided in the IRP 2010 documentation (DOE) or done internally within the organisation (Prinsloo, 2009), are also based on data at sector level, but as the major supplier of electricity they have access to a great deal of customer-specific data, both quantitative and qualitative, from which they can provide detailed sectoral forecasts. However, such customer-specific data is confidential and not available in the public domain. The methodology described in this paper therefore fills a particular gap since it uses only publicly available data by combining data reported in the public domain by Eskom with other public domain sources, and it provides a breakdown of the forecasts per usage sector. In addition, this methodology is particularly well suited to supporting scenario-based forecasting since it forecasts the expected electricity demand based on only a few key economic and demographic indicators, thereby being able to produce a new set of forecasts for a new scenario very quickly.

This methodology is also purely data-driven, and the predictor variables are selected into the models on the basis of the strength of the statistical relationships inherent in the historical data. Although the choice of variables that are tested for inclusion in the models are based on a logical understanding of each sector, this methodology does not incorporate any theories regarding the underlying causes of patterns, which would distinguish it from econometric studies focusing on the effect of a particular pre-chosen variable(s) of interest on electricity consumption (Wolde-Rufael, 2006; Ziramba, 2008; Odhiambo, 2009; Amusa, Amusa, and Mabugu, 2009; Inglesi-Lotz, 2011). In particular, a number of these studies investigate the effect of price on electricity demand, either at the national level (Inglesi, 2010; Inglesi and Pouris, 2010; Amusa, Amusa, and Mabugu, 2009; and Inglesi-Lotz, 2011) or within electricity usage sectors (Ziramba, 2008, Inglesi-Lotz and Blignaut, 2011). The results of the studies differ, but in general seem to conclude that in the late 1990s up to the mid-2000s, real income (as defined by GDP) has had a significant effect on electricity demand, while price elasticity has not been a significant predictor.

Our chosen data-driven methodology has the disadvantage of not being able to model the effect of variables that did not play a statistically significant role in the historical data, or of causal factors which could not be quantified. The continual media interest in the Eskom price hikes in recent years may indicate the need to study the effect of the price increases on demand, and it may also be argued that constraints on the supply of electricity have since 2007 (and could in future) constrain demand and even limit economic expansion. However, in the absence of reliable, sustainable sources of publicly available information on many aspects that drive current and future demand for electricity in South Africa, the statistical models proposed in this paper, which implicitly take into account the combined effect of a number of aspects in an overall pattern, could be seen to be an appropriate way to make use of the data that is available.

A concern that has to be raised, however, is that NERSA has not released any data on electricity demand per sector since 2006, and the Department of Energy has only released limited information on electricity usage per sector within the Energy Balances documents (DOEStats). However, the information contained in the Energy Balances documents seem to be questionable, since the totals do not correspond to the national electricity consumption figures released by Statistics South Africa. Although the International Energy Agency (IEA) has published electricity usage per sector since 2006, and lists the Department of Energy, along with Eskom, as sources for the statistics they have published on South Africa, they also indicate that they used their own internal estimates to determine their published data. The total production figures published by the IEA correspond to that of Statistics South Africa, but the IEA sectoral breakdowns differ from those contained in the Energy Balances documents. If data is not available in the public domain regarding electricity usage per sector it will be difficult to use the forecasting methodology described in this paper in future and it may also affect other types of energy modelling, such as energy efficiency studies.

In conclusion, the forecasting methodology presented in this paper has a strong scientific basis, and is also suitable for providing support to future strategic planning. Although any forecasts derived from historical patterns may have certain limitations, when used within the correct context, namely as one input into a high-level strategic planning process, they can provide valuable insight to support planning for potential long-term patterns of electricity demand. If reliable data on sectoral usage of electricity is consistently available in future, this methodology could offer value to a range of such strategic planning processes.

References

Al-Ghandoora, A., Al-Hintib, I., Jaberc, J.O., and Sawalha, S.A., (2008). Electricity consumption and associated GHG emissions of the Jordanian industrial sector: Empirical analysis and future projection, *Energy Policy*, Vol. 36: p. 258-267.

Amusa, H., Amusa, K., and Mabugu, R., (2009). Aggregate demand for electricity in South Africa: An analysis using the bounds testing approach to cointegration, *Energy Policy*, Vol. 37: p.4167 – 4175.

Belsey, D.A., Kuh, E., and Welsch, R.E., (1980). Regression Diagnostics, (Chapter 3, specifically p. 104, 105, 112), John Wiley & Sons.

Bianco, V., Manca, O., and Nardini, S. (2009). Electricity consumption forecasting in Italy using linear regression models, *Energy*, Vol. 34: p. 1413–1421.

CoMines Mining production data is available from the Chamber of Mines website at http://www.bullion.org.za/content/?pid=71&pagename=Facts+and+Figures [accessed July 2013].

Cooper, C.J., (1998). Digest of South African Energy Statistics 1998, Department of Minerals and Energy (DME).

Cooper, C.J. and Kotze, D.J., (1992). *Energy Projections for South Africa*, Vol. 8, Institute for Energy Studies, Rand Afrikaans University (RAU), (Appendix A).

DIGEST2002, Digest of South African Energy Statistics 2002, Department of Minerals and Energy (DME).

DIGEST2006, Digest of South African Energy Statistics 2006, Department of Minerals and Energy (DME).

DOE, Link to document available on http://www.doe-irp.co.za/ [accessed June 2013].

DOEStats Available via the website of the Department of Energy http://www.energy.gov.za/files/energyStats_frame.html, [Accessed June 2013].

Doriana, J.P., Franssen, H.T., and Simbeck, D.R. (2006). Global challenges in energy, *Energy Policy*, Volume 34: p. 1984–1991.

Eberhard, A. (2011). The Future of South African Coal: Market, Investment and Policy Challenges, Working Paper #100, Program on Energy and Sustainable Development, Stanford University, available from http://iis-db.stanford.edu/pubs/23082/WP_100_Eberhard_Future_of_South_African_Coal.pdf [accessed September 2014].

Egelioglu, F. Mohamad, A.A., and Guven, H. (2001). Economic variables and electricity consumption in Northern Cyprus, *Energy*, Vol. 26: p. 355 – 362.

ENERGYBALANCES, Energy Balance Spreadsheets, Department of Minerals and Energy (DME).

Eskom, Eskom Annual Reports, 1990-2001, Eskom.

ESKOMYEARBOOK Eskom Statistical Yearbook, 1996, Eskom.

ESS, Electricity Supply Statistics, 1996, 1998-2001, National Electricity Regulator (NER).

Imtiaz, A.K., Mariun, N.B., Amran M. R., M., Saleem, M., Wahab, N.I.A., Mohibullah, (2006). Evaluation and Forecasting of Long Term Electricity Consumption Demand for Malaysia by Statistical Analysis, First International Power and Energy Conference, (PECon 2006) Proceedings , Art. No. 4154502, p. 257-261.

Inglesi, R., (2010). Aggregate electricity demand in South Africa: Conditional forecasts to 2030, *Applied Energy*, Volume 87, p. 197–204.

Inglesi, R. and Pouris, A., (2010). Forecasting electricity demand in South Africa: A critique of Eskom's projections, *South African Journal of Science*, Vol. 106: p. 50-53.

Inglesi-Lotz, R. (2011). The evolution of price elasticity of electricity demand in South Africa: A Kalman filter application, *Energy Policy*, Vol. 39, p.3690 – 3696.

Inglesi-Lotz, R., and Blignaut, J.N. (2011). Estimating the price elasticity of demand for electricity by sector in South Africa, *South African Journal of Economic and Management Sciences*, Vol. 14 (4): p. 449 – 465.

Makridakis, S., Wheelwright, S.C. and Hyndman, R.J. (1998). Forecasting: methods and applications (3rd Edition), John Wiley & Sons Inc., p. 521.

Montgomery, D.C., Peck, E.A. and Vining, G.G., (2006). Introduction to Linear Regression (4th Edition), John Wiley & Sons Inc., p. 109 – 111 and Chapter 11.

Odhiambo, N.M. (2009). Electricity consumption and economic growth in South Africa: A trivariate causality test, *Energy Economics*, Vol. 31: p. 635–640.

nkal, D., Sayim, K.Zk, and Gönül, M.S. (2013). Scenarios as channels of forecast advice, *Technological Forecasting and Social Change*, Vol. 80 (4): p. 772–788.

Prinsloo, J. D. (2009). Long term electricity demand forecasting in Eskom (1987 to mid-2008), Available as part of the proceedings of the First Annual Forecasting Conference 'Forecasting in an ever changing environment', held in Midrand on 17 & 18 November 2009.

SAES1, South African Energy Statistics 1950-1989, No 1, 1990, National Energy Council (NEC), (Chapter 4).

SAES2, South African Energy Statistics 1950-1993 No 2, 1995, Department of Minerals and Energy Affairs (DMEA) and Eskom Marketing Intelligence, (Chapter 4).

SARB, The online statistical query facility of the Reserve Bank of South Africa is available at http://www.resbank.co.za/Research/Statistics/Pages/OnlineDownloadFacility.aspx [accessed July 2013].

Soontornrangson, W., Evans, D.G., Fuller, R.J., and Mohamed, Z., Bodger, P., (2005). Forecasting electricity consumption in New Zealand using economic and demographic variables, Energy, Vol. 30: p. 1833 – 1843.

StatsSAweb, Available via the automatic download facility of StatsSA, at http://www.statssa.gov.za/publications/statsdownload.asp?ppn=P4141&SCH=5562, [accessed July 2013].

StatsSA2, Interactive data download facility of Statistics SA available from http://www.statssa.gov.za/timeseriesdata/main_timeseriesdata.asp [accessed July 2013].

Stewart, D.F. (2003). Scenario planning for electricity supply, *Energy Policy*, Volume 31: p. 1647–1659.

Ulutaş¸ B. H. (2005). Determination of the appropriate energy policy for Turkey, *Energy*, Vol. 30: p. 1146–1161.

Van Wyk, L., and Fourie, E.T., (2009). Forecasting South Africa's electricity demand, available as part of the proceedings of the First Annual Forecasting Conference 'Forecasting in an ever changing environment', held in Midrand on 17 & 18 November 2009.

Wolde-Rufael, Y., (2006). Electricity consumption and economic growth: a time series experience for 17 African countries, *Energy Policy*, Vol. 34, p. 1106 – 1114.

Ziramba, E. (2008), The demand for residential electricity in South Africa, *Energy Policy*, 2008, Volume 36, p. 3460– 3466

Received 8 November 2013; revised 1 October 2014