MANUAL FOR ESKOM DISTRIBUTION PRE-ELECTRIFICATION TOOL (DPET)

2014/11/10

Authors:

Dr S HEUNIS (schalk.heunis@gmail.com)
M DEKENAH (marcus@mdekenah.co.za)

TABLE OF CONTENTS

1 ABOUT DISTRIBUTION PRE-ELECTRIFICATION TOOL

Eskom Distribution Pre-Electrification Tool (Distribution PET) is a software application for predicting domestic consumer ADMD (with Herman Beta Parameters), Consumption and the load profile for a group of typically 60 or more consumers.

Distribution PET has been derived after extensive research by the authors, using data collected during the course of the NRS Load Research project since1995.

The results of load predictions have been confirmed to follow very closely what was actually measured from groups of consumers in the field.

The prediction is “driven” with information about gross household income for the target community. Therefore a diligent assessment of income levels is essential.

<u>Boundaries of operation</u>
At this time, the application has the following boundaries:
Average household income 100-25 000 (in 2014 Rands)
Project time electrified in range1-15 years (inclusive)
Geographical boundaries: The tool is only designed to operate inside the boundaries of South Africa.

<u>Period of validity</u>
The current version of the application is designed to auto-update itself whenever there is new data or models available, over the internet. Typically data and profile models are reviewed every 3-4 years.

Please use the contacts given in the “about” page if you have any queries.

2 SYSTEM REQUIREMENTS

Operating system: Windows XP, Vista and Windows 7.
Processor: Intel Pentium 4 or later
Free disk space: 577 MB (typical)
Ram: 512 MB

This application uses HTML5.0 technology.

A small chrome browser is installed on the client, with all necessary program code (Java), lookup data (ie load profiles) and GIS Map data.

The code and data is kept updated via internet link whenever newer versions are available or more detailed map data is requested when the user zooms beyond the level of the local cache. Such updates only happen at run-time, if there is network connectivity.

Network connectivity is otherwise not essential for software to operate.

3 WORKING WITH THE SOFTWARE

3.1 Installation

The Distribution PET computer program is distributed as a download from
https://di.enerweb.co.za/dpet/installers/

The install file is roughly 80 MB in size.

Use the normal installation procedure for Windows applications should be followed to install the program.

3.2 Overview: Parts of the application

The application contains the following parts:

- **Input**: Handles *input*, *save*, *recall*, *import* and *print* of prediction reports
- **Summary**: Displays the tabular output of a prediction (year, energy, ADMD,and Alpha/Beta values.
- **Energy & Demand**: Shows time-graphs of ADMD & Consumption
- **Beta parameters**: Displays the probability distribution of consumers at the peak load in the forecast.
- **Profile prediction**: Produces hourly profiles from the predicted consumption
- **About:** Acknowledgements and Contacts (in case if you have any problems).

3.3 How to generate a prediction

Predictions are driven by climatic severity, average income of the community, and time-since-electrified.

The following steps should be followed:

- Select a geographic location of the site (by clicking on the map)
- Select an average income for the community (from the list)
- Click Calculate

Optional inputs:

- Risk level
- Floor area: For higher income consumers the floor area of the dwelling is significant and may be optionally specified – the norm for that income level is also shown. Further optional inputs include
- Free basic electricity (number of units per month)
- Percentage electricity theft
- Income growth per year (in excess/shortfall of normal inflation)

These inputs are used to adjust the estimated consumption of the evaluated consumer group.

Note: All other fields (ie Project name, description etc) are for information only, and may be used to make the prediction more traceable. The additional text information is saved/ loaded with the prediction, and will appear on any electronic reporting documents (Pdf) created.

Results of a prediction are shown on the **Summary**, **Energy**, **Demand**, and **Beta-parameter** and **Profile** tabs.

Once a prediction is acceptable, it can be saved. A prediction report can be previewed & printed, or saved for later.

3.4 Managing a prediction

Some input variables may be adjusted to manage a prediction.

Year in which income was collected: Income information should always be stated in terms of the year that it was assessed. No adjustment by the user is necessary. All income information is adjusted internally to a base-year by the software, using the Consumer Price Index (CPI).

Circuit-breaker size: Circuit breaker size is suggested by the software as a default, which depends upon the level of the ADMD. This can be over-ridden however. If the user selects a CB rating which is too small for the predicted level of ADMD, then an error-message will show*: "nuisance tripping* of consumers will probably occur".

Risk level: The model is very sensitive to error in the estimated income and since this is a statistical model, it also contains modeling uncertainty. This parameter can be used to adjust the estimates for these uncertainties. A value of 50% means no adjustment is made and 90% means for 9 out of 10 cases the actual measurements will be less than the predicted measurements. Users should be guided by policy.

Floor Area: This variable can be adjusted to adjust the predictions for larger or smaller than normal dwellings. *This is only relevant for higher consumption consumers* where the largest driver for consumption is not household income, but the floor area of the dwelling that is serviced by the electricity supply.

BEST free units: BEST is the Basic Electricity Support Tariff, this tariff includes a component with a number of free electricity units which is supplied free of charge. The number of free units can be entered to adjust predictions.

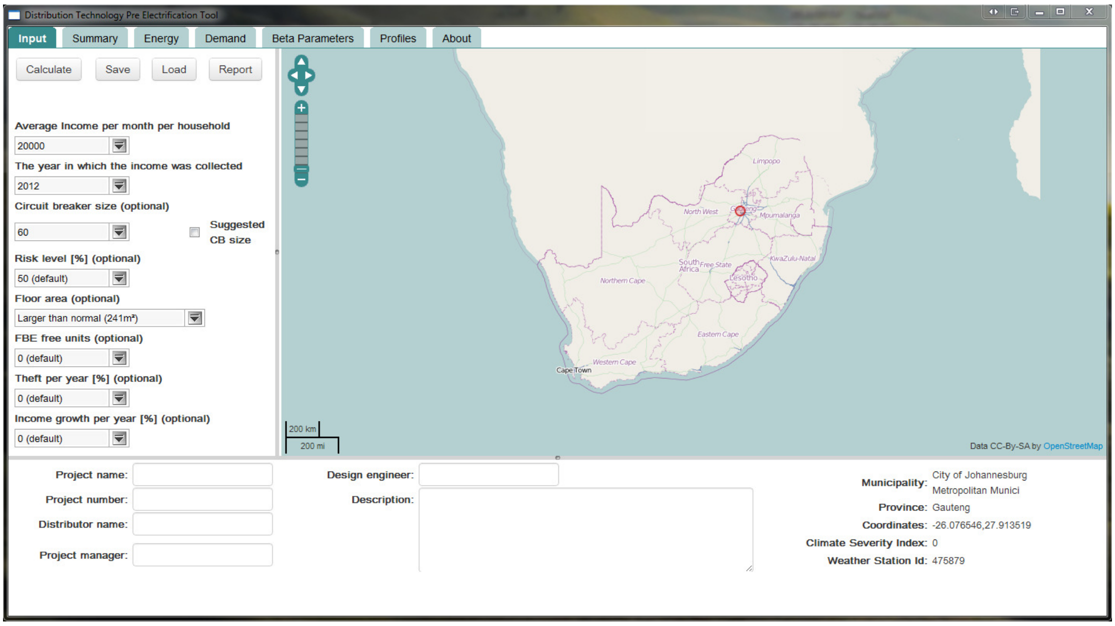

Figure 1 Input screen for predictions, showing fields as described above

The procedure for managing predictions is simple:
Alter any of the above values – NB: if you are not sure about a value leave the default!
Re-calculate the prediction and review the results.
Save or generate pdf of the prediction when satisfied.

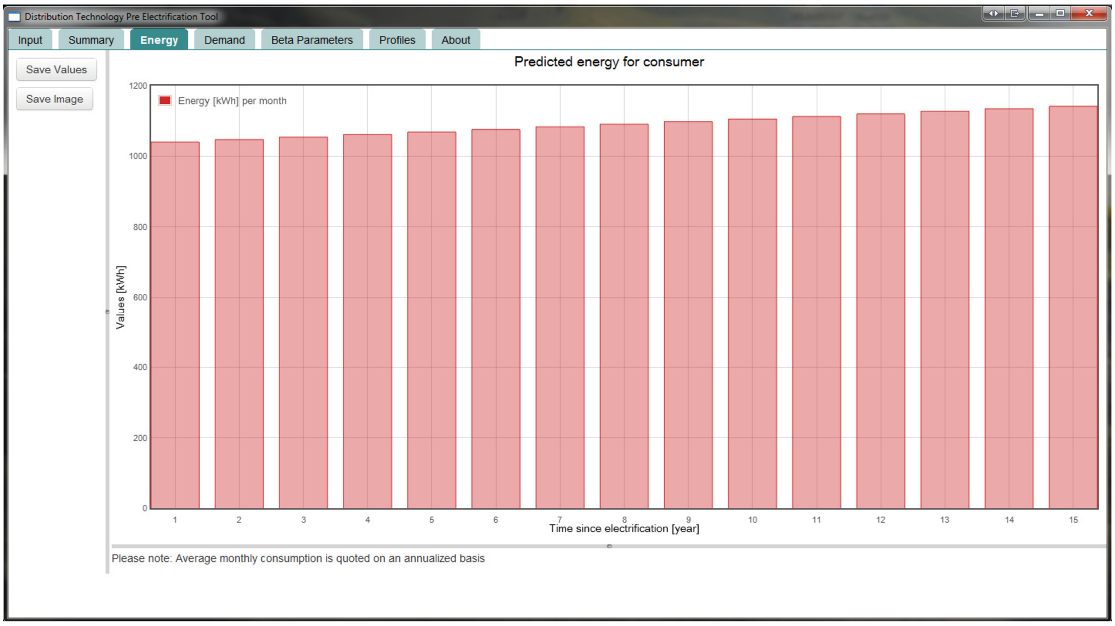

Figure 2 Example of predicted energy for consumer

The following quantities are predicted:

- Average consumption per consumer in kWh
- After diversity maximum demand in kVA (ADMD)
- Standard deviation of individual household currents around the ADMD
- Alpha, beta and circuit breaker size (if not specified), these quantities are used in the Herman Beta calculations of reticulation voltage drops.
- Average and standard deviation profiles in kVA – for more details see below.

Saving of predicted information:

- “Save values” will create a CSV export file
- “Save image” will create a “png” export file

3.4.1 Profile Prediction

A load profile prediction module is included as part of the software.

The profile model uses the prediction of consumption to estimate hourly profile shape and uncertainty.

The profile model is estimated for following specific weather stations:

- Cape Town
- Port Elizabeth
- Durban
- Bloemfontein
- Johannesburg

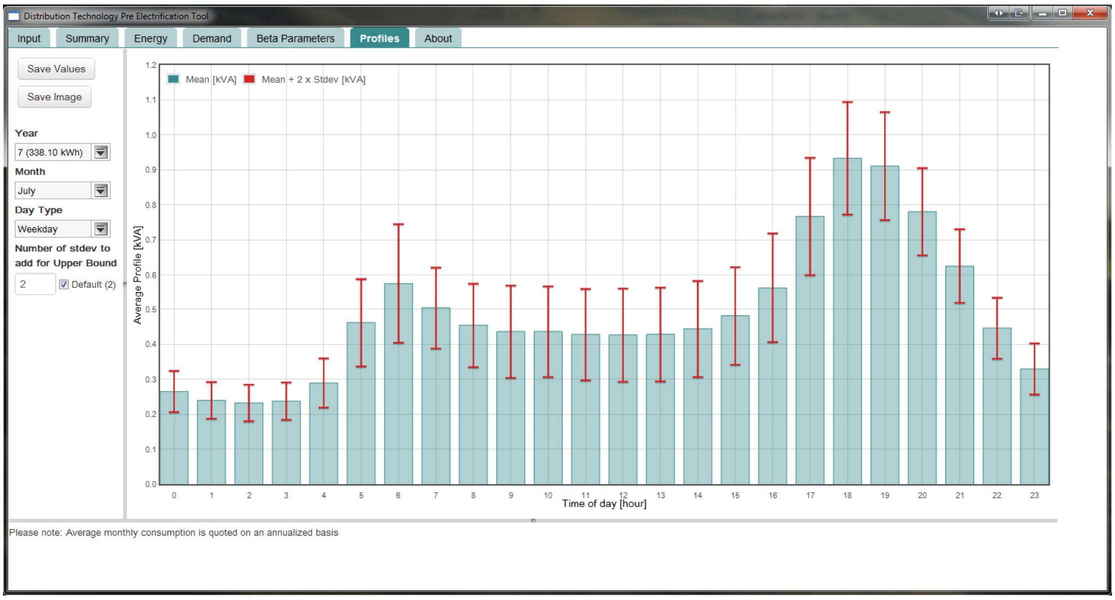

Figure 3 Example of predicted group profile expressed per consumer

For the profile prediction two quantities are estimated, the average profile and the standard deviation around the average. A group level aggregate profile is predicted and it should be noted that profiles from individual consumers in the group may vary drastically from this mean.

The predicted group standard deviation expresses the daily movement from the mean aggregate profile, so the average would be the mean value for all weekdays (as an example) and the standard deviation would be indicative of how the aggregate profile may very per day. Note that the standard deviation is not indicative of the variation within the group.

The predicted profile can be exported to the clipboard for further analysis and calculations in spreadsheet or statistical packages.

4 GATHERING INCOME SURVEY DATA

The Pre-Electrification Household-income Questionnaire is included in appendix II.

This section covers advice on how to conduct the income surveys:

- what preparations must be undertaken.
- how to execute surveys.
- evaluation and checking of the results.

It is not easy to collect accurate income information from domestic consumers, partly because the information is considered "private", and subject to sudden short-term changes at a household level.

For these reasons, it is recommended that:

- professional market-researchers are used to carry out the surveys, and
- previous experience by the *staff* of the research organisations in this area must be very carefully evaluated.
- Care should be taken to ensure that field-workers are adequately trained & suitably briefed.

4.1 Preparation phase

- Community contact must be made
- interviewers must be selected and trained, and
- the "who and how" of sample design should be decided.

4.1.1 Community contacts

If there are community channels, use them to ensure that the survey is supported, *but don't create the expectation that electricity will follow automatically.*

4.1.2 Selection & Training

4.1.2.1 Selection

The surveyors should be picked specifically to fit into the local environment. They should be of the same color and language group as the sample population. Surveyors should be literate, meticulous, and outgoing.

It is preferred that enough surveyors are used to work through the community in a day, because fore-warning in the community can distort results.

4.1.2.2 Training of interviewers

Surveyors must be adequately trained.

Role-playing is an effective method of training, since it gives each surveyor a chance to practice and to be on the receiving end. Surveyors who cannot role-play successfully should not be used on the community.

Completeness of information is important for this questionnaire, and surveyor's should be briefed as to what actions to take in case of refusals, don't-knows and no-one-homes.

4.1.3 Sample selection

4.1.3.1 Sample size

The following is to be done to ensure the reliability of the information:

Sample size:

NUMBER OF HOUSE HOLDS	Minimum SAMPLE SIZE
LESS THAN 200	40 (52)*
LESS THAN 500	60 (78)
LESS THAN 1000	80 (104)

* The number in brackets is 30% more than the Minimum. This is the recommended sample size to be given to the surveyor teams. This should ensure that the minimum is reached in spite of people not-at-home, refusals, don't-knows and badly filled in forms.

4.1.3.2 Selection of households

The methodology for generating the list of households to be surveyed is the following:

- Obtain a list of the households in the area.
- Determine the ratio (n) of total number of households to number to be sampled
- Select the house holds as per every n th in the list

E.g. :
List of households

1. Mr. Moroga
Address
2. Mr. Geldenhuys
Address
etc.

E.g. Total house holds = 473
Sample size = 78 from the table above.
Ratio = 473/ 78 = 6**.** 06
Every 6th household should therefore be selected**:**
1, 2 , 3, 4, 5**, 6**, 7, 8, 9, 10, 11, **12**, 13, 14, 15, 16, 17, **18**, 19, 20, 21, 22, 23, **24**, 25, 26, ...460, 461, **462**, 463, 464, 465, 466, 367, **468**, 469, 470, 471, 472, 473

This will result in 78 questionnaires to be filled in (i.e. 18 more than the minimum required). The few extra will serve as a backup for people that may not be at home to achieve the minimum required.

4.2 Execution of surveys

During the execution, the supervisor should equip each surveyor with his call-list, giving sites to survey. He should aim to finish the survey within a day.

Consideration should be given to the not-at-home households; it may be the reason for not-at-home is that they are at work. This may bias the sample if these are excluded from the sample.

Weekends are a good time for surveying, but care should be taken to ascertain whom the permanent residents are at the address, and assess only them (See appendix II for definition of a household).

4.3 Evaluation of results & back-checking

The field-supervisor is responsible for quality assurance in the field.
This takes the form of three types of check:

- Validity check 1: All returned survey-forms should be checked visually by the supervisor for correctness. The supervisor should spend considerable effort in the early stages of the survey to ensure there are no systematic problems with interpretation of questions by the field-workers.
- Validity check 2: About 10% of the forms should be back-checked to the consumers to ensure that they agree with their response. (i.e. the response was not invented by the field-worker). Any surveys which are "back-checked" in this manner, should be marked clearly.
- Sufficiency: The field-supervisor should check that *enough correct* surveys are returned.

5 EXPLANATION OF WARNING MESSAGES INCORPORATED IN SOFTWARE

Several warning messages have been incorporated into the software in order to avoid help the user avoid common problem areas.

Warnings concerning circuit breaker size:

- *'The circuit breaker might cause nuisance tripping for some consumers'* -This message is shown when the user has chosen a CB size smaller than the Admd plus 3 standard deviations of modeled dispersion of consumers at the time of the peak.

- 'The circuit breaker size is too small and will cause nuisance tripping for most consumers' -This message is shown when the beta parameters cannot be computed because the CB size is too small.

6 APPENDIX I: OVERVIEW OF THEORY BEHIND THE PRE-ELECTRIFICATION TOOL

6.1 An overview of the energy, ADMD and profile prediction model

Measurements of energy, demand and profile data from domestic consumers were analysed and developed into a series of sub-models for energy, ADMD and profiles for the Distribution Pre-electrification tool (i.e. Distribution PET). This appendix describes the development of the predictor in the following parts:

a) **The Energy/demand model**

The cause-effect of energy-use in residential consumers was modeled using data collected by the NRS Load Research (LR) project. The following were identified as primary influences:

- Average household income
- Project time electrified
- Floor Area
- Climatic Severity
- Theft
- Free Basic Electricity

b) **Consumption modeling**

Considerable past research has shown that aggregate wealth and time-electrified[1] have a very strong statistical connection to consumption.

Available source data for consumption modeling was adjusted to bring all income data to the same year.

Household load data were used to estimate average load per household (a proxy for consumption).

Data-sparsity in some cases inhibited the modeling process.

Some assumptions were used to deal with the data area where no data was available (i.e Income = 0 implies consumption = 0).

A local regression (or loess) technique was found to most satisfactorily model the curvilinear consumption relationship, (based upon data available at this stage). The prediction model explains *97% of the variation*[2] between the predictors (i.e. average household-income & time-with-electricity) and our measure of average household load.

c) **ADMD modeling**

The relationship between energy and demand was measured and a prediction model which explains *98% of the variation*[3] between the predictors (average household energy measure, climatic-severity) and the measured ADMD was derived.
This means that a prediction of average household load can be used to predict the maximum demand (i.e. the *ADMD*), given knowledge of the regional *climatic severity* of an electrification project.

d) **Model for consumer-dispersion during peak**

[1] Time for which the electrification project has had electricity.
[2] $R^2 = 0.96$, Standard error = 0.48 A
[3] $R^2 = 0.98$, Standard error = 0.93 A

A model for dispersion of consumers at the time of the peak was developed using data on the NRS LR database. The model explains more than 91% of the relationship in the data between dispersion of consumers (quantified as Standard deviation) at the time of the peak load, and the level of the peak load.

e) Boundaries of the prediction model

The model is reasonably accurate *within the boundaries* as set out by the source data, *outside* these boundaries the model may not perform equally well. The boundaries of the model should therefore be clearly defined and can be summarized as follows:

Average household income [R/hh/mth (2012)] : R 100-25 000
Time since electrification [years] : 1-15

A sample design for future data collection of residential consumer was compiled based on the requirements from various groups in Eskom and will improve the representation of the model and extend these boundaries.

6.2 Form of the prediction model

A two-component model was derived after much testing and verification of relationships which displayed promise. This is shown in the figure below:

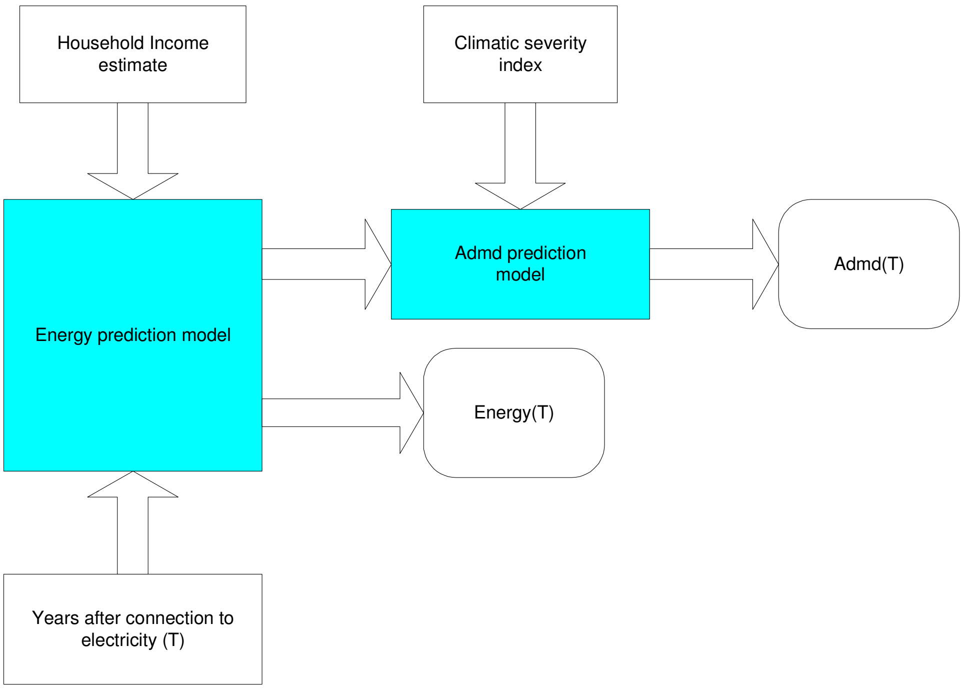

Figure 4: De-coupled load model

Development of the model evolved into this de-coupled form over some time. This form has the advantage that its components can be refined separately, using different data processing and statistical modeling techniques.

Similarly a decoupled model for load profiles was derived:

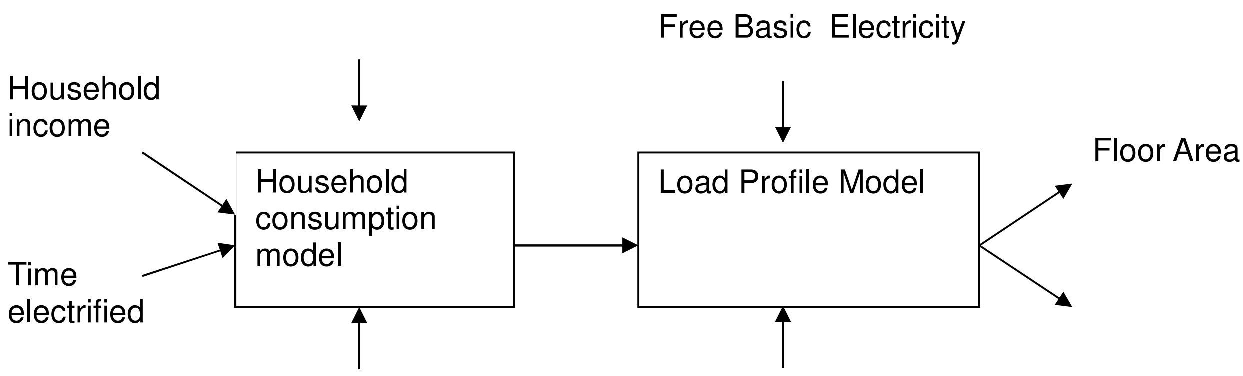

Figure 5 Decoupled load profile model

6.3 The energy prediction sub-model

Considerable past research has been conducted during statistical review of data collected for the NRS LR Project (1998 - 2011). During the course of these studies, more than 40 different kinds of sociodemographic variables collected from consumers were tested against their associated load, in terms of both statistical significance and usefulness.

These researches showed aggregate wealth and time-with-electricity have a very strong statistical connection to consumption, and are also practical variables for construction of a consumption model.

The causality behind these variables and consumption is clear.

Only appliances cause consumption, and only if they are operated by consumers, according to their habits.

Appliance ownership is strongly linked to disposable income, which in turn is related to household income. The appliance acquisition - time curve is strongly influenced by household income as first order driver. Appliance ownership and appliance usage is not necessarily related, but the assumption is that consumers with more appliances use on average more energy. The relationship may not necessarily be linear.

The following factors may also influence appliance ownership in some circumstances:

- Household income, expenses and disposable income
- Time since electrification
- Availability and cost of alternative fuels
- Circuit breaker size (load limiting)
- Appliance availability
- Infrastructure (water availability)
- Size of the dwelling (multiple appliance ownership)
- Number of occupants per dwelling

The majority of these circumstances are directly or indirectly driven by the general level of wealth in a community.

The model does not explain the effect of rural living, political change, technology options, cultural specific events, temperature etc. At this time (i.e. after modeling), *these unknowns represent about 15% of our total understanding* about the relationship between the predictors and our field-measurements acquired during the course of the NRS LR Project.

6.3.1 Source data adjustment

Load and sociodemographic data collected from the field was conditioned prior to model-building. Income data was adjusted to year 2012 values. Average household demand was estimated on an annual basis.

6.3.1.1 Household income adjustment

The household income was collected over several years and had to be normalized to a single year before a model can fitted. Household income was inflated using the consumer price index (CPI) to 2012 levels. The derived figures are an indication of the purchasing power or income at 2012 prices.

The collected household income, the collection year, the CPI and the inflated income is attached in the addenda at the end of this document.

6.3.1.2 ADD adjustment for part-of-year measurements

Average Daily Demand (i.e. ADD) for a specific year (time-with-electricity) is an approximation load very similar to energy. ADD is expressed in terms of Amps/household.

The data collection process was not always sustained for an entire year and the ADD needs to be adjusted from a partial-year to a full-year equivalent.

This adjustment was achieved by observing the relationship between the yearly ADD and monthly ADD's. Correction factors were calculated and the yearly ADD estimated, based upon the available data, for each household. Annual township ADD's were calculated as the average of the annual consumer ADD's. The correction factors and the measured ADD's in each month, for each project are included in the addenda.

6.3.1.3 "No-income : no-consumption" assumption

In some regions of the predictor space, most notably the "very poor" zone, sparsity was still a big problem, with no data available.

The model was bounded by assuming that if a consumer has no disposable income, then he will not consume any electricity. This assumption does not hold when theft occurs[4], but is valid for revenue prediction.

This assumption does provide a useful lower bound for the model and is implemented by adding a series of zero income-ADD data points over the prediction period.

The final data set after adjustments is included in the addenda.

6.3.2 The energy model

The selection of an appropriate mathematical model to represent the relationship between monthly household income, time-since-electrification and ADD was done by comparing the significance of various models.

[4] In the case of uncontrolled theft, the marginal price of electricity units is nil. The level of consumption is then probably a function of what utility the consumer requires, and the appliance penetration.

Many different models were tried. A linear model explained about 80% of the relationship. The linear model does not have enough flexibility to capture the change in all of the consumer factors as Income and Time-with-electricity changed.

The linear model was improved by adding a non-linear component. Various non-linear models exist and could be applied. *Instead of imposing a non-linear function on the model, the data points can be used to shape the fitted surface.* This is achieved using a local (or Loess) regression.

6.3.2.1 Floor Area as additional effect

During the 2005 analysis, it was noted that the consumption for some sites were significantly lower than expected. This was attributed to the relatively small household floor area, when compared to the household income.

The following figure shows the marginal effects of floorarea and household income when a model with the following formula was fitted:

$$ADD_{Year} = lo(FloorArea) + lo(Income) + TimeElectrified$$

where

FloorArea	is the average floorarea per household
Income	is the inflation adjusted income per household per month
TimeElectrfied	is the average time electrified for the community
lo	is the loess operator

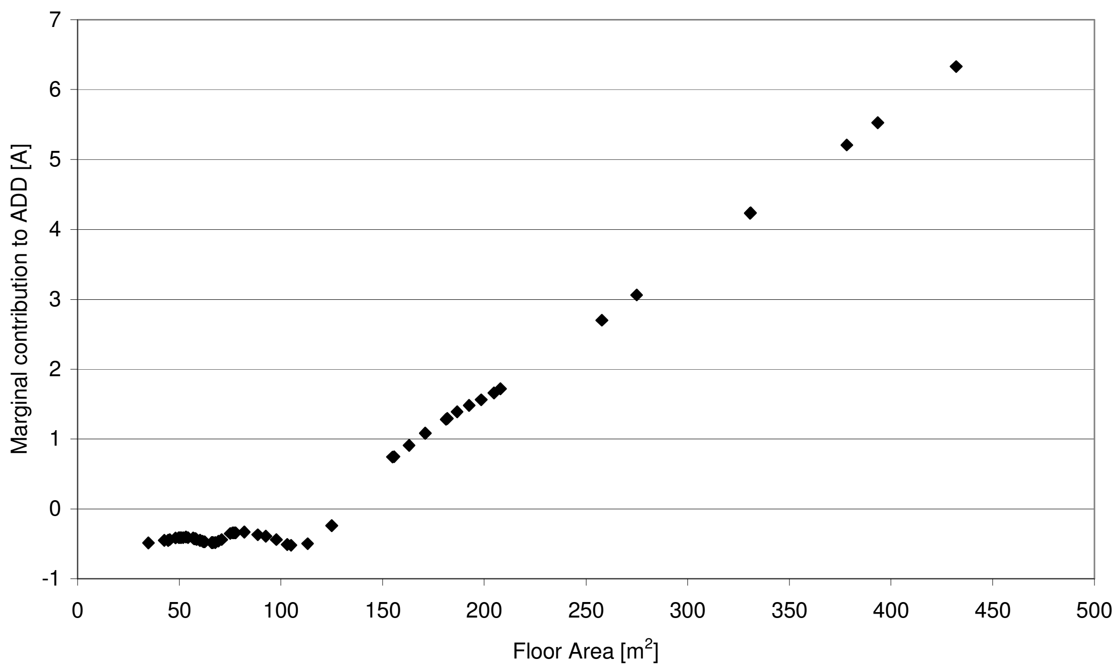

Figure 6 Marginal contribution of floor area to ADD [A]

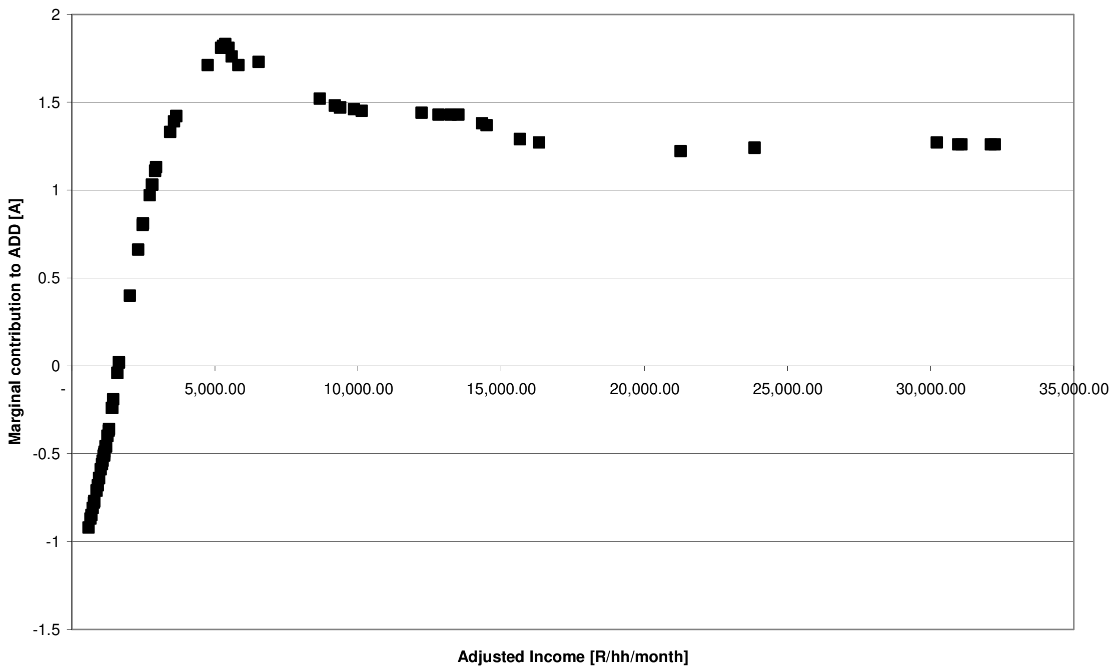

Figure 7 Marginal contribution of Inflation adjusted income (2004 Rands) to ADD [A]

It appears that the responsiveness to either income or floorarea changes at around R5000/hh/m or 100m^2.

Note that the marginal contribution of income is very high at low income – a small change in income therefore causes a significant change in consumption. This sensitivity combined with the difficulty of estimating household income accurately, causes uncertainty in the predictions (see section 20).

Note that the ADD is not sensitive to floor area when the floor area is less than 100m^2. Floor Area and household income is however highly correlated for higher income communities. Household income can therefore still be used to estimate household consumption even in higher income communities, but the response is non-linear

A linear regression model was fitted for the relationship between floor area and household income for sites with a floor area greater than 120 m^2. The model explains 88% of the variation in the source data:

$$Income = 146.2\ FloorArea - 7000$$

where

FloorArea is the average floorarea per household
Income is the inflation adjusted income per household per month

The household income for sites with an income greater than 7000 per household per month was adjusted using this relationship to cater for the impact of household floor area.

6.3.2.2 Comparison between 2010 and 2012 Models

The model was originally fitted in 2001 and was updated in 2005, and 2010 with the latest LR data. The following figure shows a comparison of the error per site using the 2010 model and the 2012 model.

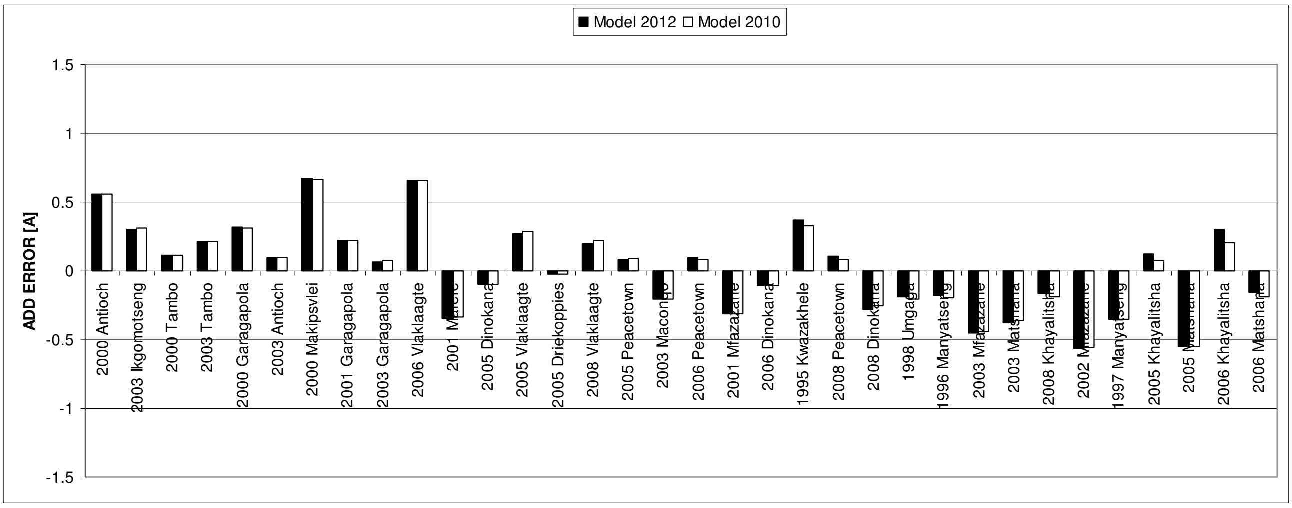

Figure 8: Comparison of error in ADD prediction, using the 2010 Model or the 2012 Model for sites in the LR database[5]. Ordered according to consumption up to 2006 Matshana.

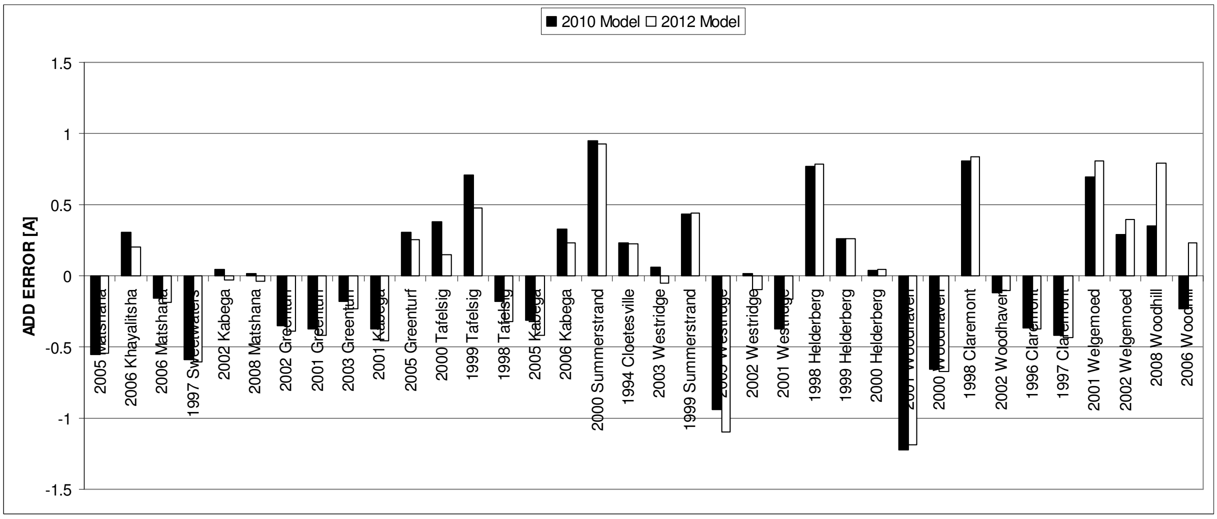

Figure 9: (Continued) Comparison of per unit error in ADD prediction, using the 2010 Model or the 2012 Model for sites in the LR database[6]. Ordered according to consumption from 2006 Matshana.

The model has a standard error of 0.48A and an R^2 of 0.96. The model was fitted with a span of 0.7, which means each local neighborhood contains about 70% of the total data points[7].

A number of outliers were detected:

[5] 1 Amp of ADD is roughly equivalent to consumption of 168 kWh/month.
[6] 1 Amp of ADD is roughly equivalent to consumption of 168 kWh/month.
[7] "Splus 4 - Guide to statistics", Data analysis products division, Mathsoft, Seattle, 1997, p227.

Outlier Name	Discussion
1999 Sanctuary	Possible data problems when compared with the 2000 observations at Sanctuary which fits the model very closely
2001 Woodhaven	Possible data problems or growth constraints – the 2003 observations at Woodhaven fits the model very closely.
2000 Makiepsvlei	Consumption lower than income predicts – income could be overstated or consumption constrained
2000 Antioch	Consumption lower than income predicts – possibly a sample problem (see Makiepsvlei 2000) or a slow adoption of new technology, i.e. need time to acquire appliances – not available locally
2003 Ikgomotseng	Consumption lower than income predicts – possibly a slow adoption of new technology, i.e. need time to acquire appliances – not available locally

6.3.2.3 Model uncertainty and risk adjustment of predictions

The following figure compares the standard error [pu] in the model fit and the standard error in estimating average household income [pu]. Both measures are expressed as a function of the average household income.

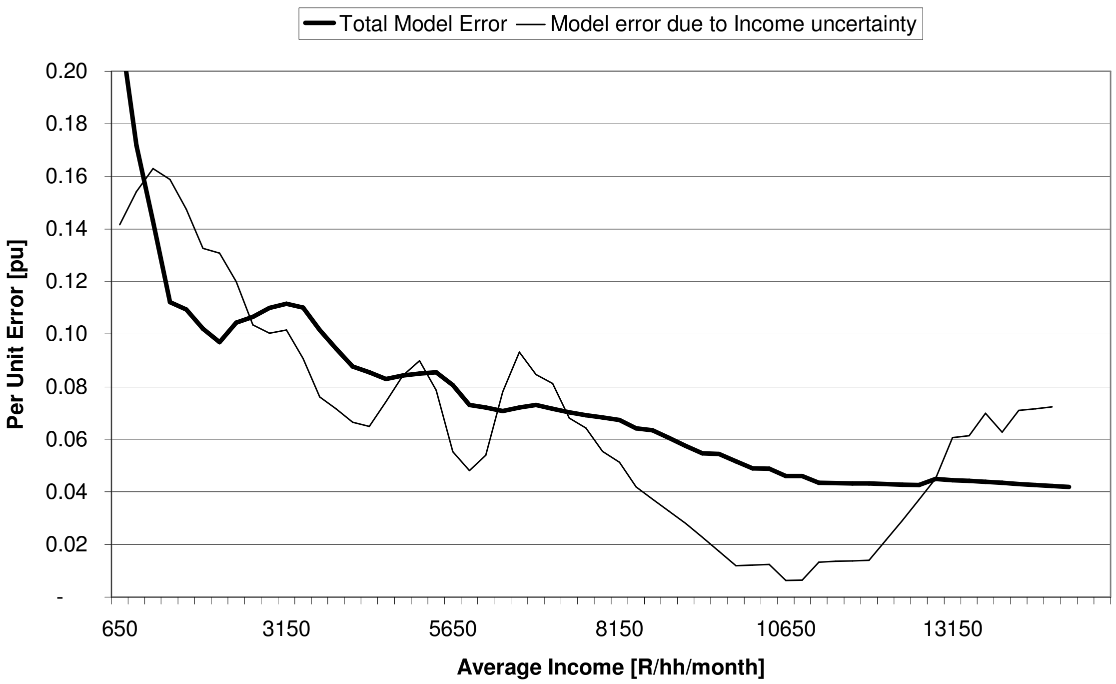

Figure 10 A comparison of the total model error [pu] and model error due to estimating average household income [pu]. Both measures are expressed as a function of the average household income.

Due to the high marginal contribution of income to ADD for low income communities, the per unit prediction error is elevated for these communities.

The load predictions can be risk - adjusted to reflect the uncertainty in the model to cater for the effect of errors in estimating the average income per household per month. The amount of risk adjustment, expressed as probability, needs to be determined.

6.3.2.4 BEST- Basic Electricity Support Tariff

DISTRIBUTION PET use the result of a study that was done to estimate the impact of a Basic Electricity Support Tariff. The tariff has a component with a number of free units and the effect of these free units were estimated using a price elasticity approach.

The effect of BEST is included using a model based on price elasticity of energy demand:

$$ADD_{extra} = -0.151\log(ADD) + 1.1$$

Where

ADD is the estimated average daily demand
ADD_{extra} is the additional load due to the free issues

The assumption is that the pickup rate for free units is 100% which differs from experiment sites conducted during 2002 and 2003.

6.4 The ADMD prediction sub-model

The amount of electrical energy used at a specific time of day is influenced by:

- Appliances that the consumers own,
- How consumer's use these appliances (i.e.habit) and
- How external drivers (like time of day and the weather) correlate adjacent consumers.

Factors which affect the size of a demand peak
A statistical approach to the definition of the peak load defines the magnitude of a system peak as the 100th percentile of the demand. This is a function of variance in the system load profile, and the mean value of the load profile (i.e. the average demand or energy).

The mean load of a system load profile is mostly a function of the mean load of the individuals (i.e. average demand or consumption).

The variance in a *system load profile* is mostly a function of the variance of the individual load profiles, and the correlation[8] between the loads of individual households.

The variance of *individual consumer loads* is a function largely of the types of appliance present in the house, and how they are used (i.e consumer habit). Variance of household loads is limited by the presence of current-limiting devices, and accentuated by climatic severity.

The *individual consumer loads* are independent, but are very much correlated with time-of-day, and local weather conditions.
The following factors tend to accentuate correlation:

- Strong local tradition (i.e. the Sunday Lunch of the Western Cape)
- Cold local weather

[8] Correlation is measured in terms of the degree to which profiles "move with each other". Coincidence is an attempt to measure correlation.

- Geographic proximity (i.e. a combination of the above)

In summary, any factor that increases the energy consumption of households in a community, the variance of household load profiles and their inter-correlation will tend to increase the measured peak demand of the community.

Model for peak demand

The effect of appliance ownership and consumer habit under the influence of external drivers is recorded in the consumer data collected onto the NRS LR database[9]. This assumption allows us to model a relationship between average load of a group of consumers and their maximum demand simply as a straight line relationship.

Measurements from the groups listed in 6.7.3 were used to fit a linear regression with reasonable goodness-of-fit, which explains about 97% of the relationship between average load and the measured ADMD[10].

$$ADMD = 0.08 ADD.CSI + 2.69 ADD + 0.40$$

Where

ADD	is the estimated average daily demand
ADMD	is the After Diversity Maximum Demand
CSI	is the climatic severity index (see below)

The effect of weather (or rather climatic severity) causes a difference in maximum demand of communities in different localities with similar energy. The difference is statistically significant and may be quantified by a Climatic Severity Index (CSI), based on the climatic regions in South Africa.

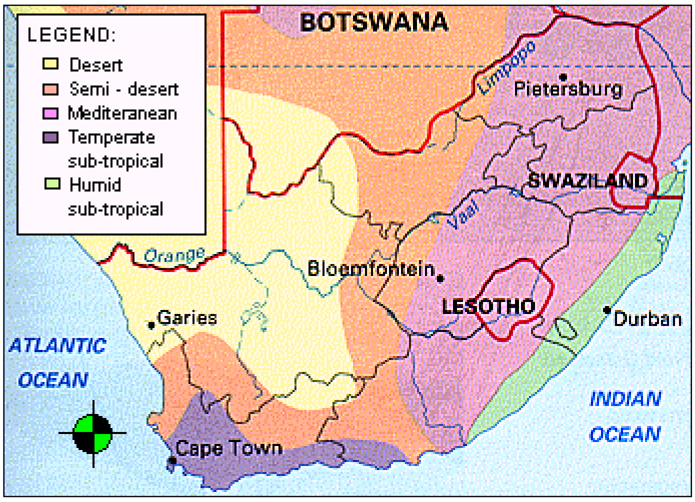

Figure 11: Climatic regions in South Africa[11]

[9] Consumer's habit is difficult to quantify. In the model described later, we assume that communities with similar income have similar habits, *in the absence* of a model of consumer behavior. Consumer habits *are* influenced by regional or cultural events/traditions and this assumption needs to be investigated as part of *future research.*

[10] The transfer function of the model, evaluated at any particular ordinate, is very similar to *annual load factor,* based upon measures of current only.

Table 1 gives a hand-scored climatic severity index (CSI) for each region.

Table 1: Climatic severity index for different regions in South Africa

Climatic region	CSI
Desert	0.5*
Semi Desert	0.5*
Mediterranean	1
Temperate sub-tropical	0
Humid sub-tropical	-1

*= Interpolations used. No data has yet been collected from these areas.

.

[11] Source: “Macmillan New Secondary School Atlas for South Africa”, 1996, ISBN 0 7978 0539 7. MacMillan Boleswa, p23.

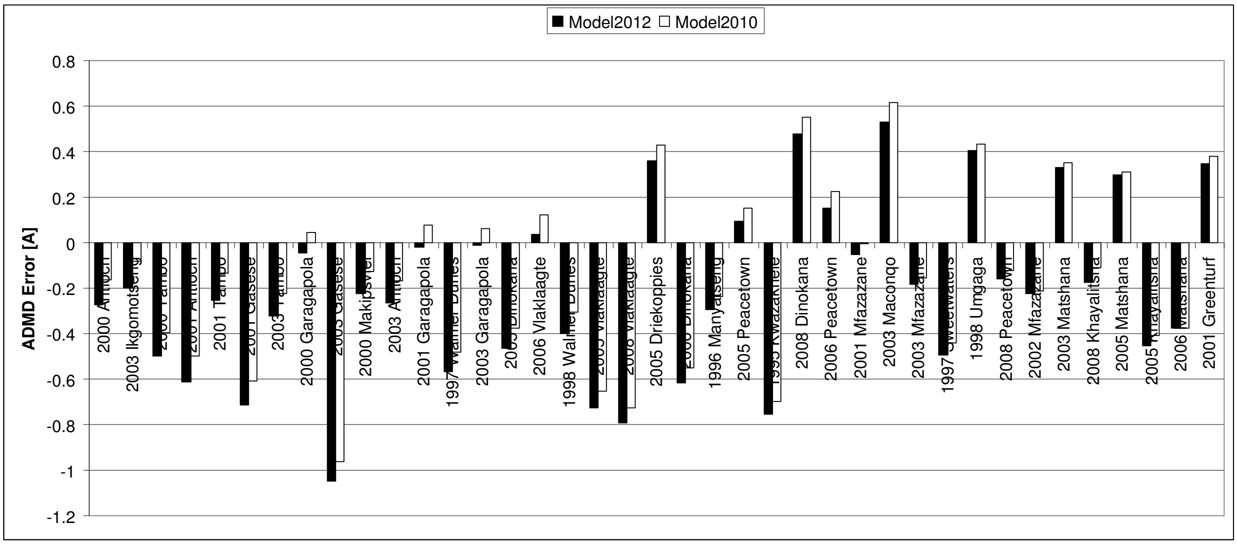

Figure 12: A comparison of the measured ADMD and the ADMD predicted (from ADD and CSI) for various South African sites. Source: NRS LR Database 1994-2008 inclusive. Ordered by consumption, up to 2001 Greenturf

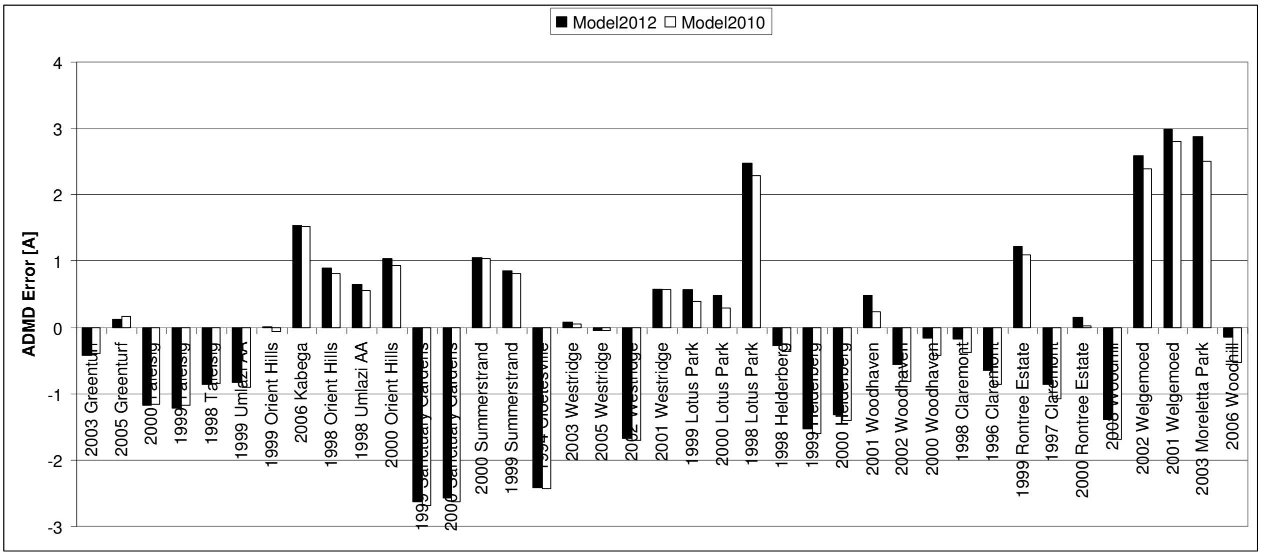

Figure 13: (Continued) A comparison of the measured ADMD and the ADMD predicted (from ADD and CSI) for various South African sites. Source: NRS LR Database 1994-2008 inclusive. Ordered by consumption, from 2003 Greenturf

Investigation of the lowest consumers (i.e. poor regions) indicates that the CSI component is less relevant in areas where consumers do not have access to space heating-type appliances. However the available South African sample from these areas is climatically biased at present. Choice of new load-monitoring sites to reduce bias of the SA rural & poor sector is an area for further research.

The actual data used for the fit is attached in the addenda later in this document.

<u>Model for dispersion of consumer-loads at the time of the peak</u>

Dispersion of consumers at the time of the peak is measured in terms of the Standard deviation of the group at the time of the peak. Standard deviation at the time of peak is most usefully related to the level of the Admd.

A linear model was developed to track the relationship between the Variance of consumers at the time of the peak and the Admd.

All domestic urban domestic load research data collated onto the NRS LR database projects to Year-end 2008 was used to form this relationship:

$$\sigma^2 = 21.4 \cdot ADD + 3.17 \cdot ADD \cdot CSI$$

Where:
σ^2 = Variance of household loads about the ADMD at the time of the peak [A^2]
ADD = Average Daily Demand [Amp]
CSI = Climatic Severity Index

This model explains 91% of the relationship in the data between dispersion of consumers at the time of the peak load, and the level of the peak load (R^2=0.89, SE=4.58).

The performance of this model is shown in Figure 15.

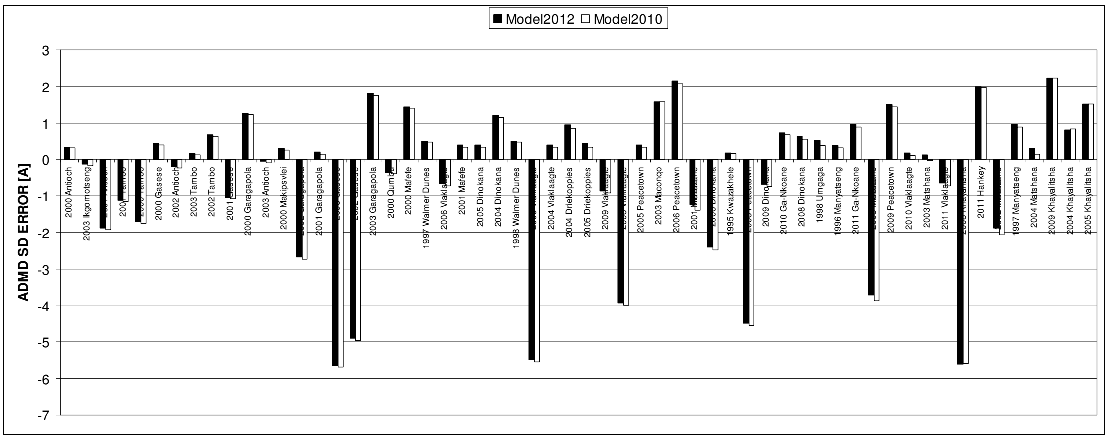

Figure 14: A comparison of the measured Standard Deviation at the time of the Peak and the predicted values (from ADD and CSI) for various South African sites. Ordered by standard deviation, up to 2005 Khayalitsha

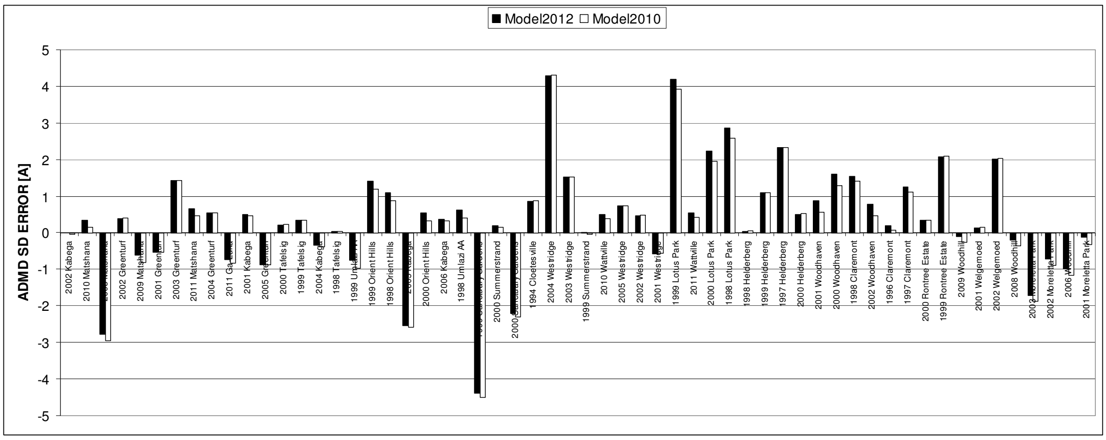

Figure 15 (Continued) A comparison of the measured Standard Deviation at the time of the Peak and the predicted values (from ADD and CSI) for various South African sites. Ordered by standard deviation, upwards from 2002 Kabega

6.5 The profile prediction sub-model

Figure 16 shows the relationship between the profile sub model and the consumption model. In order to produce an estimate of the average hourly profile per month for a specific week day type, the profile sub-model uses

- Consumption
- Temperature
- Daylight

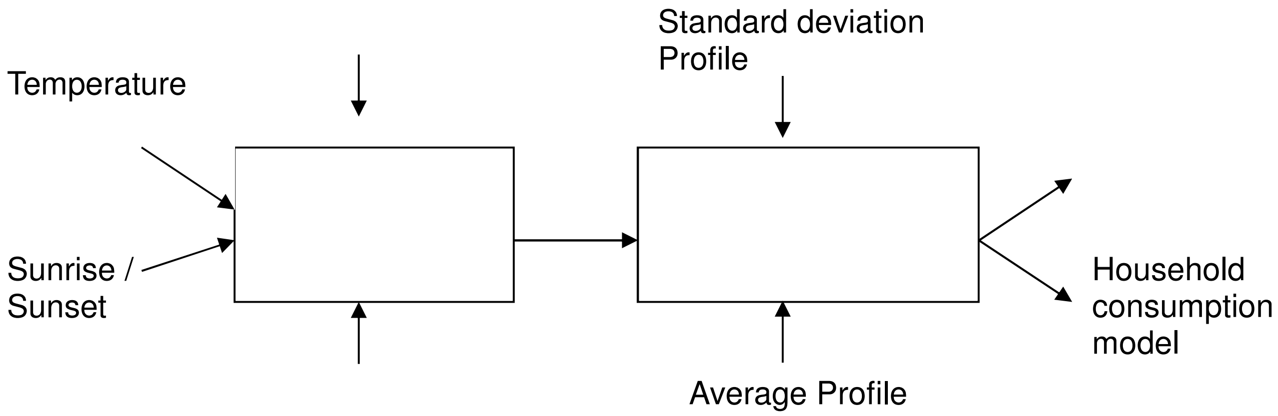

Figure 16 Decoupled load profile model

The sub model has separate regression models for each:

- Consumption class (less than 2 A ADD and greater than 2 A ADD)
- Month class (May-August, December-January, February-April + September-November)
- Weekday class (Weekdays, Saturday and Sunday)
- Hour of the day (0-23)

In total that means 432 linear regression models which are fitted individually and the results combined to produce the profile model.

Figure 17 and Figure 18 summarizes the R^2 and standard errors of these 432 linear regression models. The R^2 is greater than 70% for all models with a mean of 90% and the standard error less than 1.1A for all models with a mean of 0.34A.

A sample design for the collection of load data from residential consumers in South Africa has been compiled and will be used to collect further data in order to enhance this model and improve its performance.

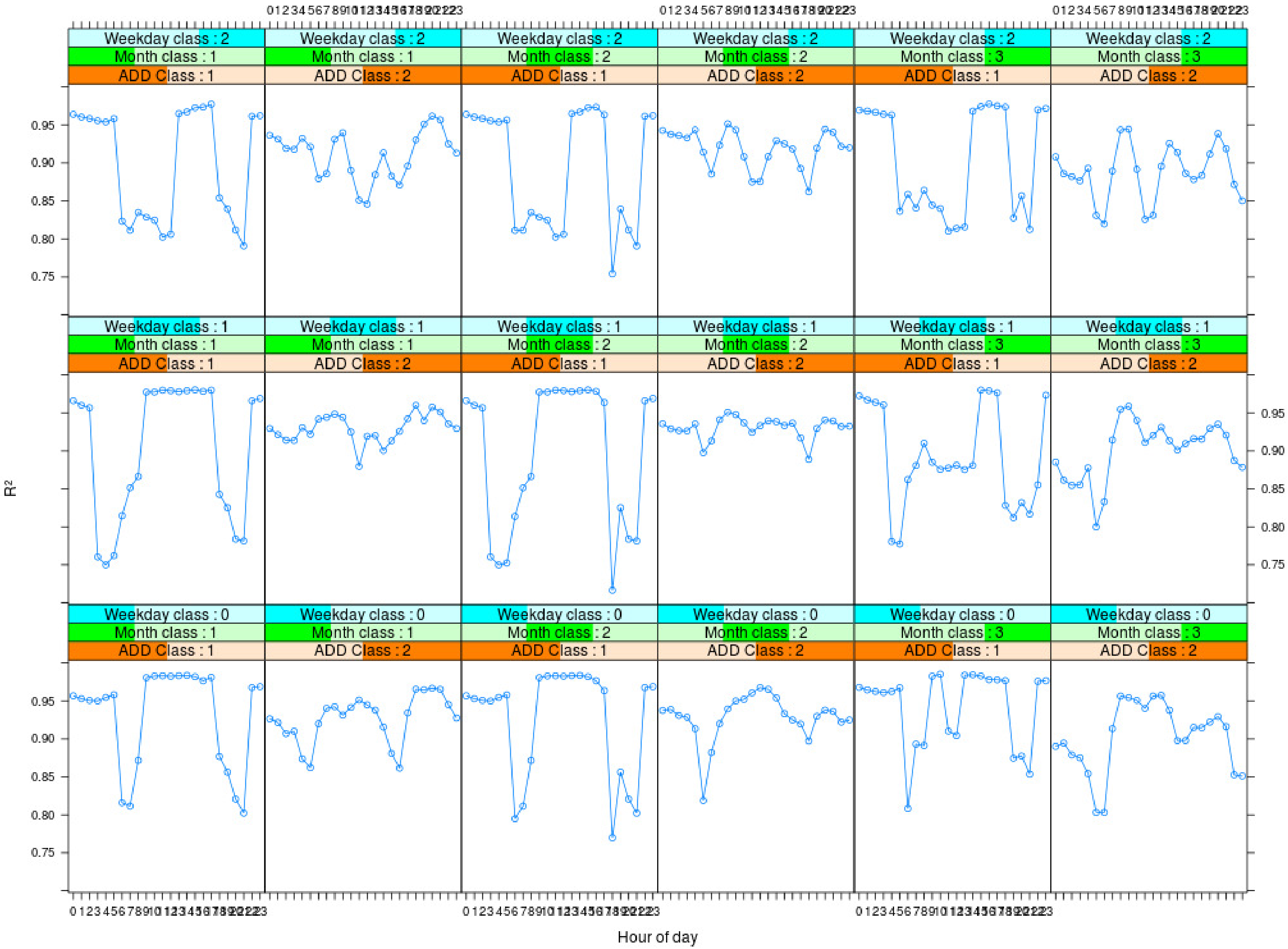

Figure 17 R^2 of the 432 linear regression models used to estimate the profile shape

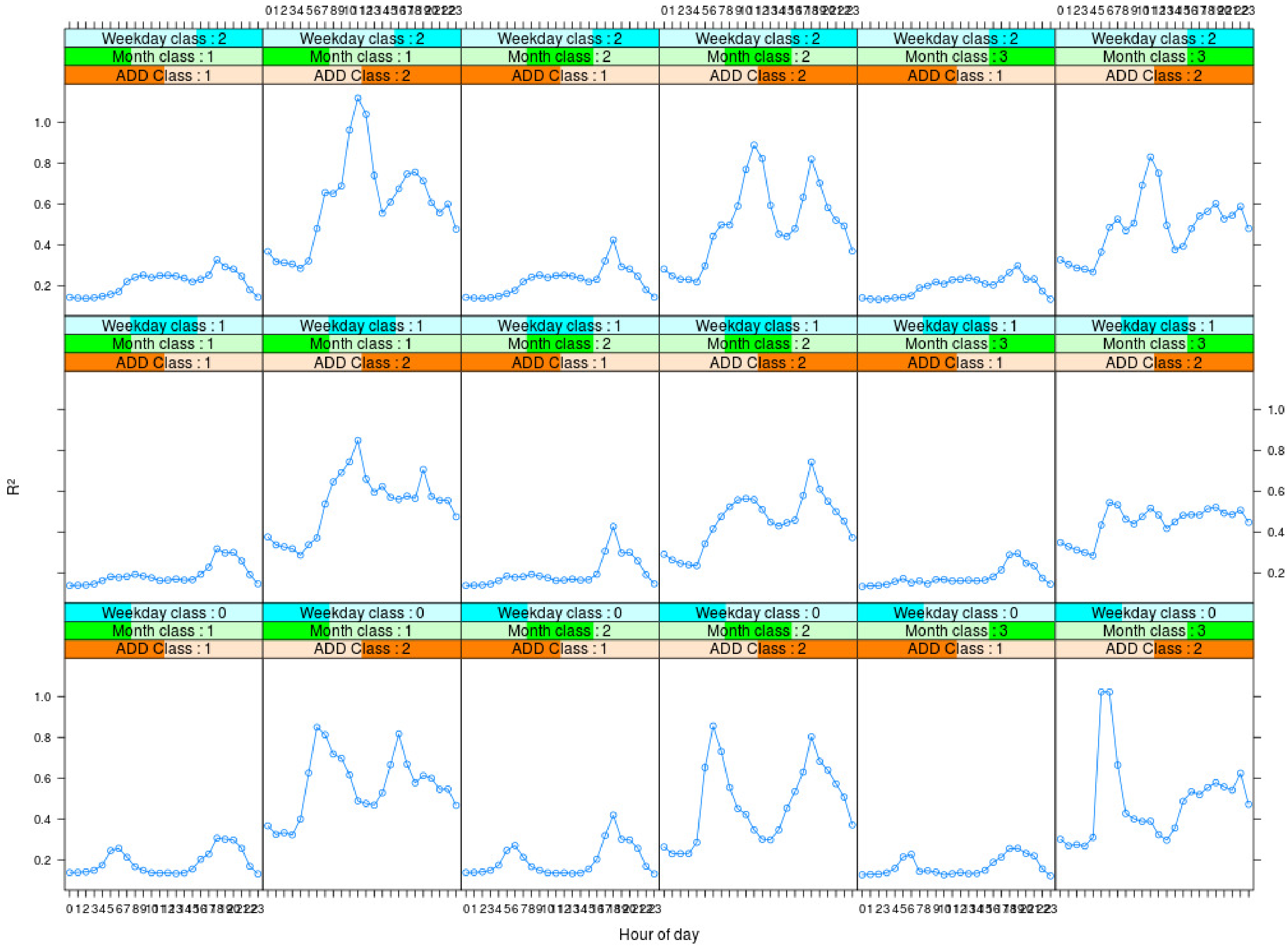

Figure 18 Standard error of the 432 linear regression models used to estimate the profile shape

6.6 Boundaries of the prediction model

The model operates within assumptions as discussed in this report. Contradictory or unrealistic predictions will result if any of these assumptions are violated.

6.6.1 Range limits

The prediction model has been developed using real load readings from the field, collected from a number of localities and over several years. The prediction model has been shown to be reasonably accurate *within the boundaries* set out by the source data.

Outside these boundaries the model may not perform equally well.

The boundaries of the model are summarized as follows:
Average household income [R/hh/mth (2012l)]: R 50-25 000
Time since electrification [years]: 1-15

6.7 *Addenda*

6.7.1 Inflation-adjustment of household-income data

Income data associated with past measurements from consumers was inflated to 2004 prices using the following formula:

$$AdjIncome = Income\frac{CPI_{12004}}{CPI_{year}}$$

with values from the following table[12]:

Year	CPI
1996	77.70
1997	84.40
1998	90.20
1999	94.90
2000	100.00
2001	106.55
2002	116.34
2003	123.08
2004	124.84
2005	129.02
2006	134.95
2007	144.58
2008	161.27
2009	171.86
2010	179.72
2011	187.75
2012	197.37
2013	210.36
2014	223.39
2015	236.03
2016	248.34
2017	261.72
2018	276.39
2019	292.23

Inflation estimates from 2014 onwards were supplied by Dr Johannes Jordaan (economic modelling solutions) based upon macro-economic economic models for SA, 2014/11/06.

[12] Statistical Release P0141.4, “Prices : Consumer price index base: 1995”, Central Statistical Service South Africa.

6.7.2 Estimated yearly ADD from Monthly ADD

Monthly ADD adjustment to yearly ADD – correction factors

Month	<2.5 A ADD	>2.5 A ADD
1.00	1.05	1.11
2.00	1.06	1.08
3.00	1.04	1.09
4.00	1.01	1.09
5.00	0.97	0.99
6.00	0.96	0.88
7.00	0.97	0.86
8.00	0.97	0.92
9.00	0.97	0.99
10.00	1.00	1.05
11.00	0.97	1.09
12.00	1.04	1.15

The factors in this table are interpreted as the amount by which the average consumption in any month exceeds the average consumption for the year.

6.7.3 The source data set and the results of modelsNotes:

Group name	ADD	Adj Income	Time elec	Floor area	ADMD	ADMD SD	CSI	BEST	ADD Excl	Theft	ADMD Excl	Profiles Excl	Comments
1994 Cloetesville	3.48	9,704	10	51	12.71	8.35	1	n	n	n	n	n	
1995 Kwazakhele	1.32	3,575	4	45	4.56	5.19	0.5	n	n	n	n	n	
1996 Claremont	5.88	10,474	27	94	16.59	10.23	0	n	n	n	n	n	Claremont 1996 floorarea changed to 1997, value was 96 vs 171
1996 Manyatseng	1.44	2,374	8	59	3.75	5.26	0	n	n	n	n	n	
1997 Claremont	5.97	9,329	28	171	17.33	10.40	0	n	n	n	n	n	
1997 Helderberg	5.13	8,841	12	193	18.42	8.77	1	n	y	n	y	n	Income difference year to year
1997 Manyatseng	1.69	2,504	9	75	4.38	4.97	0	n	n	n	n	y	
1997 Sweetwaters	1.99	2,815	5	66	4.56	5.30	-0.5	n	n	n	n	n	
1997 Walmer Dunes	0.90	1,986	3	77	3.27	3.99	0.5	n	n	n	n	n	
1998 Claremont	5.71	9,344	29	205	16.08	9.47	0	n	n	n	n	n	
1998 Helderberg	4.84	16,584	13	199	13.81	10.78	1	n	n	n	n	n	
1998 Lotus Park	4.46	5,220	21	82	9.52	6.29	-1	n	n	n	n	n	
1998 Orient Hills	3.02	2,908	18	48	7.28	6.33	-1	n	n	n	n	n	
1998 Tafelsig	2.65	4,406	17	93	8.49	8.04	1	n	n	n	n	n	
1998 Umgaga	1.43	2,511	4	59	3.68	4.52	-1	n	n	n	n	n	
1998 Umlazi AA	3.20	4,992	18	76	7.56	6.09	-1	n	n	n	n	n	
1998 Walmer Dunes	0.99	2,099	4	57	3.42	4.14	0.5	n	n	n	n	n	
1999 Helderberg	4.98	18,030	14	182	15.33	9.92	1	n	n	n	n	n	
1999 Lotus Park	4.20	6,362	22	122	11.09	4.59	-1	n	n	n	n	n	Changed FloorArea to 1998 value, used to be 122.23
1999 Orient Hills	2.96	4,121	19	58	8.02	5.92	-1	n	n	n	n	n	
1999 Rontree Estate	6.34	38,473	25	275	15.57	12.00	1	n	y	n	n	n	Holiday homes
1999 Sanctuary Gardens	3.25	22,591	1	77	11.55	7.61	0	n	n	n	n	n	
1999 Summerstrand	3.62	24,137	3	155	8.74	8.85	0.5	n	n	n	n	n	
1999 Tafelsig	2.64	6,500	18	67	8.75	8.18	1	n	n	n	n	n	
1999 Umlazi AA	2.89	4,814	19	71	8.61	8.00	-1	n	n	n	n	n	

Group name	ADD	Adj Income	Time elec	Floor area	ADMD	ADMD SD	CSI	BEST	ADD Excl	Theft	ADMD Excl	Profiles Excl	Comments
2000 Antioch	0.30	1,698	4	63	1.49	2.02	0	n	n	n	n	n	
2000 Garagapola	0.62	1,864	4	89	2.05	2.30	0	n	n	n	n	n	
2000 Gasese	0.51	2,164	4	105	1.95	3.17	0	n	y	n	n	y	Income differs significantly from 2001 Gasese
2000 Helderberg	5.38	19,491	15	187	16.19	11.65	1	n	n	n	n	n	
2000 Lotus Park	4.38	7,570	23	78	11.21	5.97	-1	n	n	n	n	n	
2000 Mafefe	0.84	1,872	0	67	1.87	2.33	0	n	y	n	n	y	Income differs significantly from 2001 Mafefe
2000 Makipsvlei	0.74	2,803	6	125	2.33	3.30	0	n	n	n	n	n	Floor area suspiciously high
2000 Orient Hills	3.07	4,483	20	43	7.18	6.94	-1	n	n	n	n	n	
2000 Qumbu	0.82	2,373	5	86	2.30	3.60	0.5	n	y	n	n	y	
2000 Rontree Estate	6.33	38,025	26	258	17.45	12.83	1	n	y	n	n	n	Holiday homes
2000 Sanctuary Gardens	3.38	22,853	2	113	11.65	10.63	0	n	n	n	n	n	
2000 Summerstrand	3.34	26,025	4	163	8.20	7.54	0.5	n	n	n	n	n	
2000 Tafelsig	2.55	5,400	19	57	8.54	7.46	1	n	n	n	n	n	
2000 Tambo	0.50	1,223	4	60	1.89	2.92	0	n	n	n	n	n	
2000 Woodhaven	5.65	23,113	9	181	14.93	8.86	-1	n	n	n	n	n	
2001 Antioch	0.41	1,525	5	50	2.15	4.66	0	n	n	n	n	n	
2001 Garagapola	0.74	1,889	5	66	2.31	3.56	0	n	n	n	n	n	
2001 Gasese	0.58	1,093	5	50	2.47	4.10	0	n	n	n	n	n	
2001 Greenturf	2.15	3,937	2	69	5.94	5.34	1	n	n	n	n	n	
2001 Kabega	2.43	15,608	7	78	6.66	7.82	0.5	n	n	n	n	y	
2001 Mafefe	0.94	1,174	5	57	2.41	3.31	0	n	n	n	y	y	2.5 A connection
2001 Mfazazane	1.27	1,772	8	45	3.81	5.86	-1	n	n	n	n	n	
2001 Moreletta Park	10.26	50,787	16	432	29.03	17.62	0	n	n	n	y	y	Low customer count in Winter
2001 Tambo	0.47	1,040	5	45	1.84	2.47	0	n	n	n	n	n	
2001 Welgemoed	7.99	47,806	23	331	18.71	13.76	1	n	n	n	n	n	
2001 Westridge	3.86	8,262	27	98	10.34	9.71	1	n	n	n	n	n	
2001 Woodhaven	5.56	24,749	9	156	13.90	8.77	-1	n	n	n	n	n	

Group name	ADD	Adj Income	Time elec	Floor area	ADMD	ADMD SD	CSI	BEST	ADD Excl	Theft	ADMD Excl	Profiles Excl	Comments
2002 Antioch	0.55	1,515	6	55	2.21	3.28	0	y	n	n	n	n	
2002 Garagapola	0.74	1,708	6	84	3.03	6.41	0	y	n	n	n	n	
2002 Gasese	0.77	1,216	6	57	2.39	7.88	0	y	n	n	n	n	
2002 Greenturf	2.14	3,921	3	62	6.66	6.73	1	n	n	n	n	n	
2002 Kabega	2.03	14,533	8	78	6.35	7.01	0.5	n	n	n	n	y	
2002 Mfazazane	1.69	2,073	9	53	4.43	7.04	-1	n	n	n	n	n	
2002 Moreletta Park	9.49	49,156	17	378	25.71	14.46	0	n	n	n	y	y	Low customer count in Winter
2002 Tambo	0.57	1,235	6	51	2.39	3.33	0	y	n	n	n	n	
2002 Welgemoed	8.46	48,369	24	331	18.85	12.37	1	n	n	n	n	n	
2002 Westridge	3.77	8,844	28	105	12.32	9.13	1	n	n	n	n	n	
2002 Woodhaven	5.79	21,345	10	208	15.31	9.52	-1	n	n	n	n	n	
2003 Antioch	0.66	1,413	7	61	2.50	3.31	0	n	n	n	n	n	
2003 Garagapola	0.81	1,660	7	77	2.69	2.97	0	n	n	n	n	n	
2003 Gasese	0.76	933	7	52	3.14	8.78	0	n	n	n	n	n	
2003 Greenturf	2.25	4,613	4	62	7.06	6.00	1	n	n	n	n	n	
2003 Ikgomotseng	0.34	1,253	5	35	1.52	2.68	0	n	n	n	n	n	
2003 Maconqo	1.20	1,893	7	62	3.56	4.13	1	n	n	n	n	n	
2003 Matshana	1.53	1,990	12	60	4.02	5.06	-1	n	n	n	n	n	
2003 Mfazazane	1.46	1,802	10	68	4.16	7.77	-1	n	n	n	n	n	
2003 Moreletta Park	9.38	50,996	18	393	21.07	15.44	0	n	n	n	n	n	
2003 Tambo	0.56	1,434	8	53	2.22	3.32	0	n	n	n	n	n	
2003 Westridge	3.62	8,790	29	103	10.34	8.30	1	n	n	n	n	n	
2004 Dinokana	0.96	1,456	6	60	3.30	4.17	0	n	y	n	n	y	
2004 Driekoppies	1.07	1,929	3	38	2.51	3.43	-0.5	n	y	n	n	y	
2004 Greenturf	2.40	4,152	6	68	6.73	6.55	1	n	y	n	n	y	
2004 Kabega	2.64	16,509	10	83	7.32	7.31	0.5	n	y	n	n	y	
2004 Khayalitsha	1.82	2,944	15	44	5.47	5.12	1	n	y	n	n	y	
2004 Matshana	1.70	2,431	13	55	4.22	4.41	-1	n	y	n	n	y	
2004 Vlaklaagte	1.06	2,235	10	68	2.82	4.22	0	n	y	n	n	y	

Group name	ADD	Adj Income	Time elec	Floor area	ADMD	ADMD SD	CSI	BEST	ADD Excl	Theft	ADMD Excl	Profiles Excl	Comments
2004 Westridge	3.50	9,997	30	105	8.74	5.21	1	n	n	n	y	y	Outlier compared with 2005
2005 Dinokana	0.95	1,602	7	67	3.26	4.13	0	n	n	n	n	n	
2005 Driekoppies	1.07	2,105	4	61	2.76	2.93	-0.5	n	n	n	n	n	
2005 Greenturf	2.43	6,142	7	60	7.22	7.50	1	n	n	n	n	n	
2005 Kabega	3.04	17,055	11	87	8.46	10.56	0.5	n	n	n	n	y	
2005 Khayalitsha	1.87	3,379	17	43	5.97	5.06	1	n	n	n	n	n	
2005 Matshana	1.89	2,250	14	54	4.88	5.90	-1	n	n	n	n	n	
2005 Peacetown	1.18	2,146	13	63	3.58	4.47	0	n	n	n	n	n	
2005 Vlaklaagte	1.00	2,258	11	70	3.78	9.74	0	n	n	n	n	n	
2005 Westridge	3.76	10,825	31	87	10.50	8.68	1	n	n	n	n	n	
2006 Dinokana	1.28	2,227	8	70	3.96	7.33	0	n	n	n	n	n	
2006 Kabega	3.09	18,302	12	101	6.86	7.53	0.5	n	n	n	n	n	
2006 Khayalitsha	1.92	3,842	18	45	7.72	12.24	1	n	n	n	y	n	Outlier compared with 2008
2006 Matshana	1.97	3,128	15	54	5.76	6.15	-1	n	n	n	n	n	
2006 Peacetown	1.24	2,166	16	59	3.66	4.41	0	n	n	n	n	n	
2006 Vlaklaagte	0.91	2,810	12	74	2.84	4.11	0	n	n	n	n	n	
2006 Woodhill	9.89	86,728	7	499	24.49	15.81	0	n	n	n	n	n	
2008 Dinokana	1.40	2,106	8	78	3.31	4.62	0	n	n	n	n	n	
2008 Khayalitsha	1.61	2,314	19	49	5.29	8.66	1	n	n	n	n	n	
2008 Matshana	2.14	3,877	15	65	6.88	8.87	-1	n	n	n	n	y	
2008 Peacetown	1.33	2,303	17	70	4.37	9.05	0	n	n	n	n	n	
2008 Vlaklaagte	1.14	2,344	12	84	3.93	8.19	0	n	n	n	n	n	
2008 Woodhill	8.52	119,885	8	438	22.04	14.36	0	n	n	n	n	n	
2009 Dinokana	1.35	2,389	9	79	3.61	4.41	0	n	n	n	n	n	
2009 Khayalitsha	1.74	3,720	20	49	4.12	5.11	1	n	n	n	n	n	
2009 Matshana	2.14	2,871	16	58	5.32	6.64	-1	n	n	n	n	n	
2009 Peacetown	1.48	2,765	18	62	3.18	4.03	0	n	n	n	n	n	
2009 Vlaklaagte	1.12	2,386	13	76	3.41	5.16	0	n	n	n	n	n	
2009 Woodhill	7.56	115,478	9	446	22.06	13.80	0	n	n	n	n	n	

Group name	ADD	Adj Income	Time elec	Floor area	ADMD	ADMD SD	CSI	BEST	ADD Excl	Theft	ADMD Excl	Profiles Excl	Comments
2010 Ga-Nkoane	1.36	2,568	9	78	3.36	4.33	0	n	n	n	n	n	
2010 Matshana	2.11	3,964	15	69	4.73	8.11	-1	n	n	n	n	n	
2010 Vlaklaagte	1.51	2,185	15	89	4.52	5.39	0	n	n	n	n	n	
2010 Wattville	3.71	4,279	50	73	9.66	8.44	0	n	n	y	n	y	
2011 Ga-Luka	2.41	5,618	15	87	8.31	10.42	0	n	n	n	n	n	
2011 Ga-Nkoane	1.44	2,167	8	79	4.44	4.95	0	n	n	n	n	n	
2011 Hankey	1.62	2,759	15	57	4.57	4.15	0.5	n	n	n	n	n	
2011 Matshana	2.31	2,887	18	67	7.72	8.77	-1	n	n	n	n	n	
2011 Vlaklaagte	1.57	3,345	23	101	4.84	5.94	0	n	n	n	n	n	
2011 Wattville	4.25	3,888	51	71	13.67	8.84	0	n	n	y	n	y	

1. Climatic severity was estimated by hand-scoring against the different climatic regions in South Africa. Results show that a rough estimate of this parameter is sufficient.

2. ADD-Measured was calculated by applying the ADD correction factors to the ADD of individual consumers within their respective groups.

7 APPENDIX II: DISTRIBUTION PET SURVEY QUESTIONNAIRE

The survey questionnaire consists of 3 components:

- Briefing to surveyors
- Survey form
- Income show-card

7.1 Briefing to surveyors

The following briefing should be given to field workers before they go out into the field:

1. Always approach the respondents in a friendly manner. Explain exactly what you want and why you are doing the research.

2. Always explain that the information is confidential and will not be divulged to anyone. The information will be used for research purposes only. It is useful to use a letter printed on a letterhead, to prove that your motives are just that.

3. Never get into an argument with any of the respondents. If you have explained your purpose and the respondent still refuses, thank him/her for their time and discontinue the interview.

4. NEVER enter into political or religious discussions even when asked. This is the most dangerous discussion you can have in a situation like this.

5. Always ask the questions on the questionnaire and write down the answers as they give them to you. Never assume anything and if a respondent declines to answer a specific question, leave it blank and write a comment on the side.

6. When you approach a respondent, be professional. Remember you are working for a company and you need to adhere to their code of conduct.

7. You approach a respondent with the sole purpose of completing the questionnaire. Do not deviate from the questionnaire for a social chat. The respondent can become tired and give substandard information. Respondent fatigue starts after approximately 20 - 30 minutes. Sometimes even sooner.

8. If you need to translate any information into another language, make sure you give the correct facts in that language.

9. Do not agree or disagree with opinions expressed by the repondent. This tends to bias their response.

10. Complete the questionnaire and always thank the respondent for their time. Remember to write anything that you are not sure of on the questionnaire. It makes the analysis so much easier.

11. Definition of a household: A group of people who regularly cook and share meals in one place.

Remember that some of your work WILL be verified by someone else, so do it right the first time.

7.2 PET Survey-form

PRE-ELECTRIFICATION RESEARCH STUDY - 1999

Name of respondent: ________________________________
Stand #: ___________________________
Extension no: ___________________________
Area: ___________________________________
Tel no: ___________________________________
Interviewer Name: ____
Form-check OK: _______ **Back-check OK:** _______

1. DEMOGRAPHIC INFORMATION

1.1 Language preferred: ________________________________

1.2 Household size (indicate number of males and females for each category).

Age	*Male*	*Female*
<16		
16 – 24		
25 – 34		
35 – 49		
50+		

1.3 Working status (indicate for each individual living in the household).

	Head	*Spouse*	*>16 yrs*	*< 16 yrs*
Full time				
Part time				
Unemployed/student				
Pension				
Self employed				

2. INCOME INFORMATION

a (I) Do you or does anyone else get or receive any money from family or friends outside this household?

Yes	No

(ii) If "yes", how much money do they get per month? (Show income card and record number from card if respondent does not know).

R	
No.	
Midpoint	

b (I) Does this household make any money from the sale of agricultural produce (e.g. sugar cane, meat, poultry, maize, vegetables, etc.) ?

Yes	No

(ii) If, "yes", ask how much is made from the sale of these products per month.

R	
No.	
Midpoint	

c (I) Does this household get any money from Any other sources like child maintenance, Small business etc. ?

Yes	No

(ii) If, "yes", ask how much money is received per month?

R	
No.	
Midpoint	

d Of all the adults in the household, how many Earn money from salary/wage/ pension?

e How much do these adults earn per month in total? (before any deductions)

R	
No.	
Midpoint	

Interviewer to sum the amounts from a-e above. Write the value in total column.

Total ___________

Thank the respondent and close the interview.

7.3 PET Income show-card

The income show card is used in conjunction with the survey questionnaire.
The Income-ranges are specially designed to control sample error and should be reproduced exactly. Enlarge this card slightly to A4 size and ensure each field worker takes one into the field.

INCOME RANGE (RANDS PER MONTH)

Number	From	To
1	0	99
2	100	199
3	200	299
4	300	399
5	400	499
6	500	599
7	600	699
8	700	799
9	800	899
10	900	999
11	1000	1099
12	1100	1199
13	1200	1399
14	1400	1599
15	1600	1999
16	2000	2499
17	2500	2999
18	3000	3999
19	4000	4999
20	5000	5999
21	6000	6999
22	7000	7999
23	8000	8999
24	9000	9999
25	10000	10999
26	11000	11999
27	12000	13999
28	14000	15999
29	16000	17999
30	18000+	

Source: NRS LR Project 1999