Icon Water Limited

# Attachment E1: Water volumes forecasts

**November 2014**

**Information return to the Independent Competition and Regulatory Commission biennial recalibration of prices for water and sewerage services**

Icon Water Limited is an unlisted public company that owns and operates the water and sewerage assets and business in the ACT. The company is owned by the ACT Government and has two voting shareholders: the Chief Minister and Deputy Chief Minister of the ACT.

Until 30th October 2014, Icon Water Limited was known as ACTEW Corporation Limited. Supporting documentation produced prior to that day will refer to ACTEW Corporation Limited or ACTEW Water. Icon Water Limited will continue to use the brand name "ACTEW Water" until further notice.

# Contents

# Summary

The Breusch-Ward demand forecasting model proposed in ACTEW Water's April 2013 submission remains ACTEW Water's preferred forecasting model. However, given this model was not accepted by the Commission in its Final Decision, ACTEW Water has prepared a different model for the purpose of the biennial recalibration to meet the Commission's preferences. In particular, ACTEW has calibrated a model to the period following the removal of temporary water restrictions (the last 3.5 years).

It is necessary to revise the Breusch-Ward model predictors because evaporation is no longer regularly recorded at Canberra Airport. ACTEW Water has replaced Canberra Airport evaporation with three climate variables that were all found to improve the statistical significance of the model:

- Burrinjuck Dam evaporation;
- Canberra Airport maximum humidity; and
- Canberra Airport maximum temperature.

The model predicts demand accurately across the short 3.5 year calibration period (Figure 1). The daily r-squared value for 91 day aggregated data is 0.975 for the multivariate linear model.

Given the small amount of calibration data, the model is unlikely to perform as well outside the calibration period, especially for climate conditions that have not been experienced in the calibration period. However, some model uncertainty must be accepted at present until a significant period of stable consumption data has been observed.

There is also a significant amount of uncertainty generated by climate. Using historical climate conditions observed since 1965, the model estimates that total annual water releases can vary between 40.7 GL/year and 57.2 GL/year as a direct result of climate conditions.

The average releases volume across all historical climate conditions is 47.4 GL/year. This is higher than the Commission's determination for 2013/14 and 2014/15 (44.5 GL/year), but is slightly below actual consumption in these years. It is also a little lower, but similar to the forecast in ACTEW Water's April 2013 submission.

# 1 Methodology

## 1.1 The 2012 Breusch-Ward model

The Breusch-Ward model used in the 2012 regulatory submission forms the starting point for the updated model. Ignoring the autoregressive component, the model is as follows:[1]

$$releases = 149.6868 - 39.6261.stage0 - 19.4883.stage2 + 7.1937.stage0.s + 0.9470.cumevap \\ - 2.4526.cumrain2.stage3.s - 6.9164cumrain2$$

where releases are the total daily water demand

*stage0* is set to 1 when any water restrictions apply, including permanent water conservation measures and set to 0 at other times

*stage2* is set to 1 when stage 2 or more severe water restrictions apply and set to 0 at other times

*stage3* is set to 1 when stage 3 or more severe water restrictions apply and set to 0 at other times

*s* is a seasonality index $0.5 \cos 2\pi \frac{days\ since\ 21/01/1989}{365.25} + 0.5$

*cumevap* is an evaporation index based on the past 7 days evaporation at Canberra airport $\sum_{n=0}^{6} evap_n(1 + 4e^{-1.1n})$

*cumrain2* is the square root of a rainfall index based on the past 21 days rainfall at Canberra airport $\sqrt{\sum_{n=0}^{20} rain_n(0.25 + 1.2e^{-0.2n})}$

## 1.2 Updating the model to post-restrictions data

The above model is calibrated to data from 2001 to 2011.[2] For the biennial review, the model is to be calibrated to the post water restrictions period, 2010 to 2014. Two issues arise when this is performed:

- The water restriction level is constant, so all the terms related to water restrictions drop out of the model, leaving only the seasonality, evaporation, rainfall and constant terms.
- Canberra airport evaporation is only available for 62% of days in the calibration period and has not been recorded since 5/06/2013. It had previously been recorded nearly continuously since 1967.

Given current high storage levels, restrictions are unlikely in the medium term forecast period, so it is not necessary to include the impact of water restrictions in the model.

[1] Dr Michael Ward & Dr Trevor Breusch, *Building a forecasting model of ACT water use*, June 2012
[2] Dr Michael Ward & Dr Trevor Breusch, *Building a forecasting model of ACT water use*, June 2012

However, the lack of continuously recorded evaporation data makes inclusion of this term in the model unviable. It is necessary to replace the evaporation term with other relevant variable(s) that have a long climate record. A long record is necessary for forecasting so that the model can be run with a reasonable cross-section of climate conditions.

The following variables were trialled by fitting a multivariate linear model to the rainfall index *cumrain2*, the seasonality index and an index created from the new variable:

- Burrinjuck Dam evaporation
- Canberra Airport maximum temperature
- Canberra Airport minimum temperature
- Canberra Airport maximum humidity
- Canberra Airport average humidity
- Canberra Airport minimum humidity

Of these variables, Burrinjuck Dam evaporation, Canberra Airport maximum temperature and Canberra airport maximum humidity were found to produce the strongest models and were selected for further evaluation.

# 2 Building the model

## 2.1 Fitting a model to the data

To match the billing cycle, both releases and predictor variable data are aggregated over 91 days before calibrating the model, as described in section 4.1 of Ward & Breusch's model description.[3]

Burrinjuck evaporation, Canberra airport temperature and Canberra airport humidity indices were developed using a similar form to those used by Ward & Breusch for rainfall and evaporation. The *cumrain2* rainfall index was also evaluated by removing the square root, changing the number of days and altering the 1.2 and -0.2 terms, however none of these changes improved the model. The resulting model calibrated to the post restrictions period with all variables aggregated over 91 days is as follows:

$$releases = 590.9 - 29.4901.s + 0.1585.cumBurrinjuckevap - 11.2498.cumhumidity2 \\ + 0.1083cumtemp - 3.706.cumrain2$$

Where releases are the total daily water demand

s is a seasonality index $0.5 \cos 2\pi \frac{days\ since\ 21/01/1989}{365.25} + 0.5$

*cumBurrinjuckevap* is the sum of the past 28 days evaporation at Burrinjuck Dam $\sum_{n=0}^{27} evap_n$

*cumhumidity2* is the square root of a humidity index based on the past 7 days maximum humidity at Canberra airport $\sqrt{\sum_{n=0}^{6} maxhumidity_n (1 + 10e^{-n})}$

*cumtemp* is the sum of the past 28 days maximum temperature at Canberra Airport $\sum_{n=0}^{27} maxtemperature_n$

*cumrain2* is the square root of a rainfall index based on the past 21 days rainfall at Canberra airport $\sqrt{\sum_{n=0}^{20} rain_n (0.25 + 1.2e^{-0.2n})}$

Examining the sign of the variables, demand shows a positive relationship with evaporation and temperature and a negative correlation with humidity and rainfall, as expected. The negative relationship with the seasonality index is counter-intuitive, but given that the variables are all related to season (especially temperature and evaporation), the seasonality component of the model is really an adjustment to the weather based model rather than a direct driver of demand.

The seasonality index and the rainfall index, *cumrain2*, are taken directly from the original Breusch-Ward model. Appendix A describes the selection of the form of the other indices.

## 2.2 Testing for significance

Clearly there is a risk that the model may be overfitted. Where the Breusch-Ward model used only evaporation and rainfall as climate variables, this model uses four climate variables. An overfitted model performs well against the calibration set but is less likely to perform well outside the calibration set because it contains variables that are less statistically significant.

[3] Dr Michael Ward & Dr Trevor Breusch, *Building a forecasting model of ACT water use*, June 2012.

The model was built in the R statistical software[4] to test the significance of the variables. The R output in Table 1 shows that all variables are significant. To confirm this, R was used to fit a linear multivariate model to the data using stepwise regression with both forward selection and backward elimination. Both approaches selected all variables in the resulting model. The full R output is shown in Appendix B. As a final test, the Bayesian Information Criterion (BIC) was calculated for the five variable model, a four variable model without the *cumBurrinjuckevap* term and a three variable model without the *cumBurrinjuckevap* and *s* terms. The BIC was clearly lowest for the full five variable model. Therefore, the model described above is proposed for adoption for ACTEW's biennial recalibration submission.

**Table 1 – R Output for New Model**

| | *Estimate* | *Std. Error* | *t value* | *Pr(>\|t\|)* | *Significant* |
|---|---|---|---|---|---|
| (Intercept) | 590.87517 | 20.210638 | 29.236 | < 2e-16 | *** |
| cumhumidity2 | -11.249752 | 0.44431 | -25.32 | < 2e-16 | *** |
| cumBurrinjuckevap | 0.158461 | 0.022126 | 7.162 | 1.33E-12 | *** |
| cumtemp | 0.10831 | 0.004842 | 22.371 | < 2e-16 | *** |
| cumrain2 | -3.706119 | 0.134132 | -27.63 | < 2e-16 | *** |
| s | -29.490089 | 2.147854 | -13.73 | < 2e-16 | *** |

## 2.3 Model performance

The model performs very well across the calibration period, as indicated by an R-squared value of 0.975. Figure 1 compares the model output to the observed values for 91-day aggregated total water demand. The model predicts the summer peaks well in both the extremely wet years (2010/11 and 2011/12) and normal years (2012/13 and 2013/14). Apart from the first winter (2011), there is a tendency to underestimate the winter demand and bottom a little too early.

[4] R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2013.

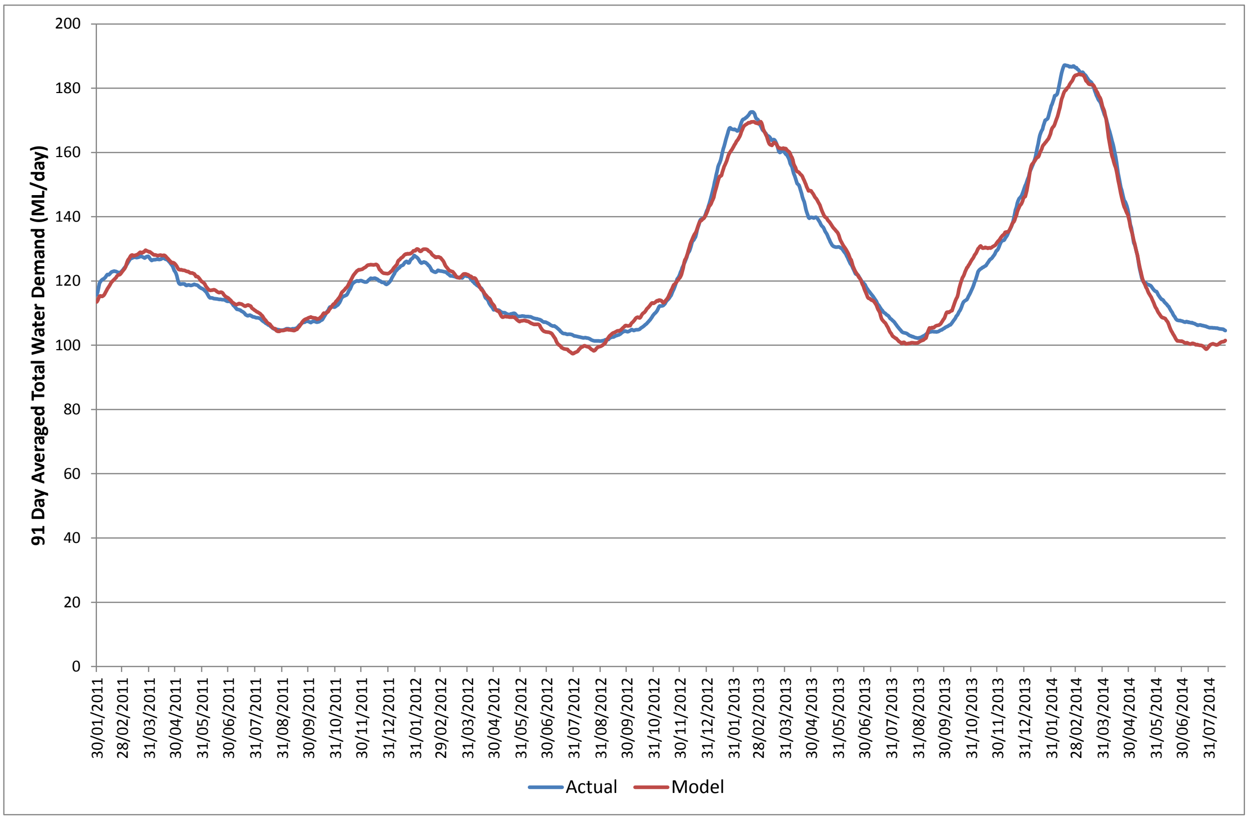



**Figure 1 – Comparison between Model Output and Observed 91-day Total Water Demand**

## 2.4 Model uncertainty

The model is unlikely to perform as well outside the calibration period as within the period. When the model is calibrated to only 3.5 years of data it is unlikely to fully capture the effects of the full range of likely predictor levels. It is also possible that water demand behaviour could change again following an extended period without water restrictions. Like the original Breusch-Ward model, the model contains no allowance for population growth, implicitly assuming that population growth is offset by demand reductions. Population growth was found to be not statistically significant by Ward & Breusch.[5] While this is a reasonable assumption for the calibration period and the short-term future, in the longer term population growth may increase demand. The model is therefore not suitable for longer term modelling.

While demand forecasting is currently somewhat uncertain this is not a deficiency in the model, but a function of the recent changes in water use behaviour. Even with a perfect model it is not possible to forecast medium term future weather with any accuracy, which leads to significant uncertainty in demand.

## 2.5 Autoregression

Autoregression is not included in the model because:

- With only a few years of calibration data it is not likely to be statistically significant.
- The static model errors are small, so it would have little impact.

[5] Dr Michael Ward & Dr Trevor Breusch, *Building a forecasting model of ACT water use*, June 2012

# 3 Forecast demand

## 3.1 Using period of record climate to forecast demand

When forecasting demand it is necessary to use a broader climate set than the short set of calibration data. It is possible to extend the climate data back to 1/01/1965, so the period from 1965 to present has been used for climate inputs to the model. This provides 49 estimates of annual total water demand that can be used for ACTEW Water's biennial recalibration forecast. The average, median, minimum and maximum of these 49 estimates are shown in Table 2, while Figure 2 shows how the 49 estimates are distributed.

**Table 2 – Summary of Forecast Demand (GL p.a.)**

| | |
|---|---|
| Average | 47.4 |
| Median | 47.0 |
| Minimum | 40.7 |
| Maximum | 57.2 |

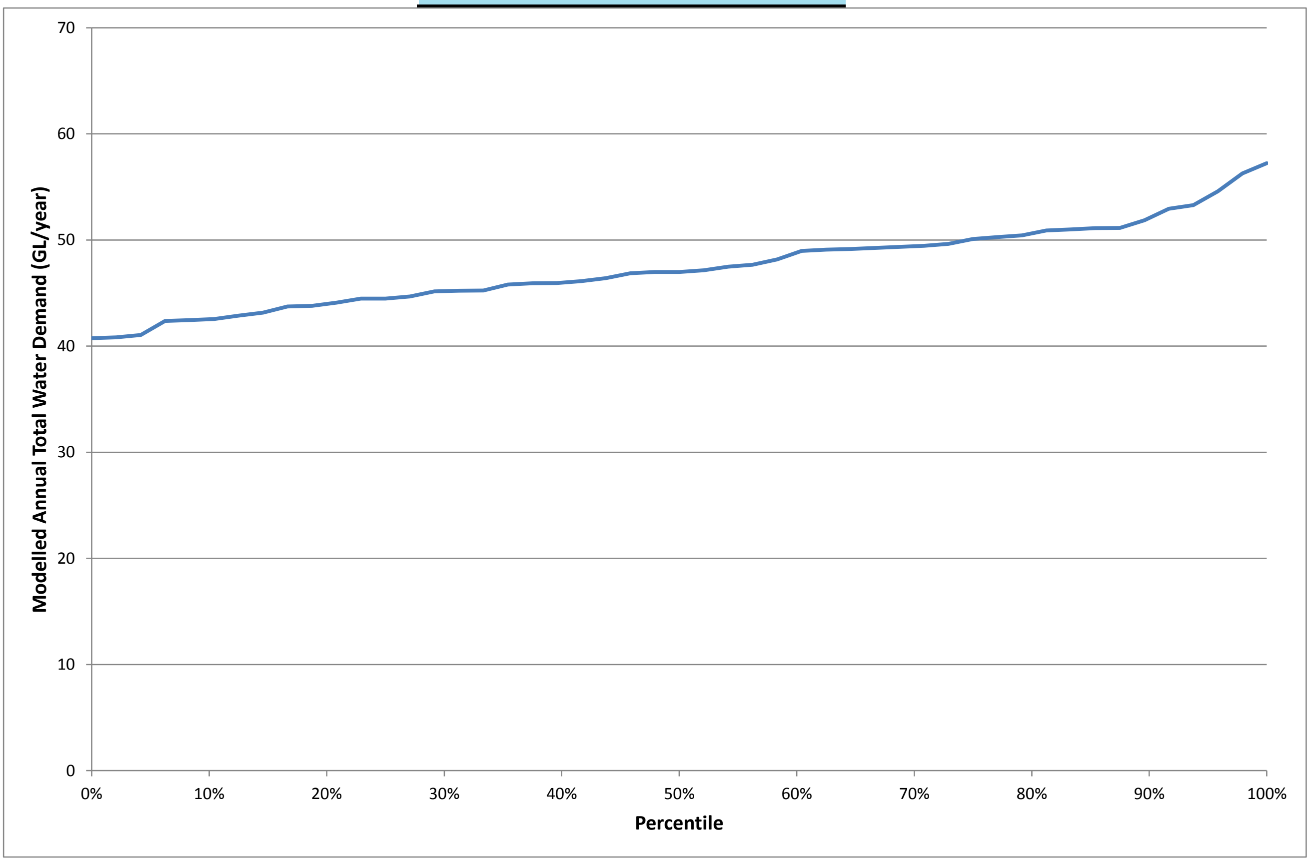



**Figure 2 – Variation in Forecast Demand with Climate**

## 3.2 Comparing the demand output to previous models and observations

Table 3 compares the average model output to recent demand forecasts and observed data. This comparison is reassuring because:

- The forecast total water demand of 47.4 GL/year sits above the ICRC determination for 2013/14 and 2014/15 of 44.5 GL/year and the 4 year average of 44.8 GL/year. This is expected, since the four year average is skewed by two extremely wet years. In the model, 2010/11 has the lowest consumption of the 49 years, while 2011/12 has the sixth lowest consumption.
- The forecast total water demand is reasonably close to the last two years (2012/13 and 2013/14), which experienced relatively normal climate conditions.
- Although a different method is used, the forecast water demand is reasonably close to ACTEW Water's forecast submitted in the response to the draft determination, which is also close to the observed demand.

**Table 3 – Summary of Modelled and Observed Demand**

| | *Total water demand (GL/year)* |
|---|---|
| Updated Model | 47.4 |
| ICRC Determination | 44.5[6] |
| ACTEW Water 2012 Regulatory Submission | 52.5[7] |
| ACTEW Water 2013 Response to Draft Determination | 47.6 in 2013/14 to 49.6 in 2017/18[8] |
| 2010/11 Actual | 40.9 |
| 2011/12 Actual | 41.6 |
| 2012/13 Actual | 47.8 |
| 2013/14 Actual | 48.7 |
| 4 Year Average | 44.8 |

[6] ICRC, *Final Report: Regulated Water and Sewerage Services*, p. 117, June 2013

[7] ACTEW, *ACTEW main submission to the Independent Competition and Regulatory Commission*, p. 79, July 2012

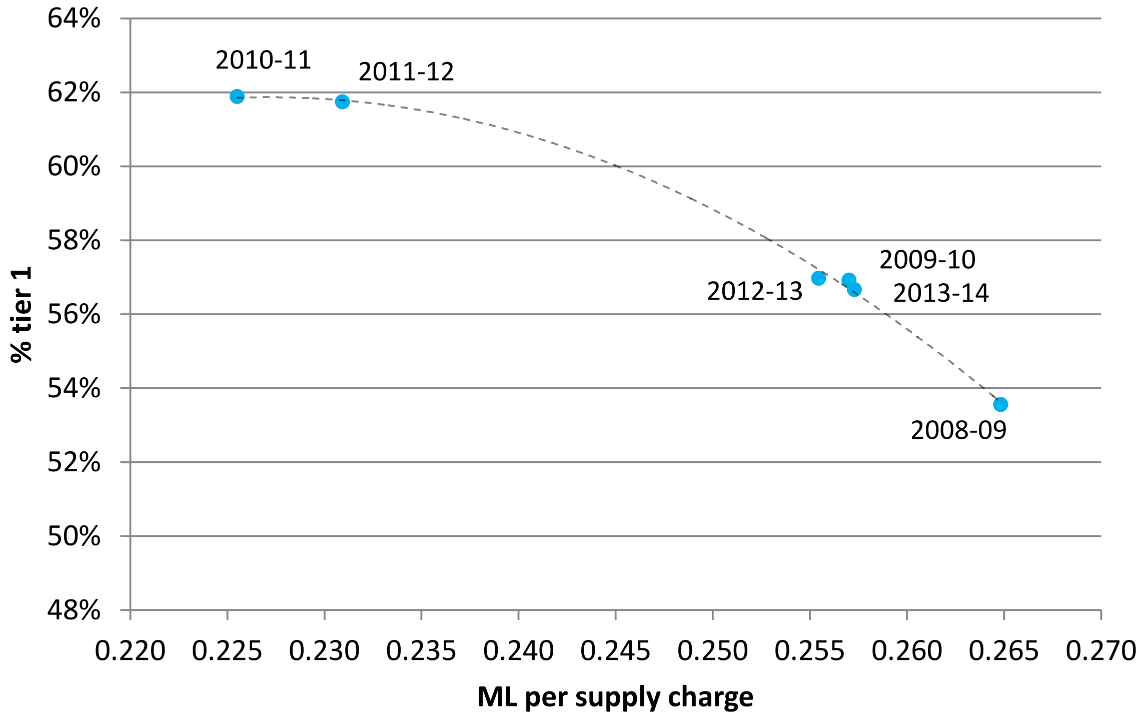
[8] ACTEW, *Response to the Draft Report Regulated Water and Sewerage Services*, p.66, April 2013

# 4 Forecast sales volumes by price tier

Total forecast water sales are calculated by multiplying the forecast releases outlined in the previous section by 85 per cent – the observed ratio of ACT billings to releases volumes in the calibration period. This calculation gives an annual sales forecast of 40.3 GL per annum.

The proportion of consumption charged at tier 1 prices was estimated based on the observed relationship with average consumption per customer in the historical data. The historical observations of these two variables show a strong relationship (see Figure 3).

**Figure 3: Historical relationship between proportion of sales charged at tier 1 prices and average consumption**




ACTEW Water estimated this relationship as:

$$\textit{Propn tier 1} = 0.75264 - 0.00811\ e^{0.012254 \times \text{total annual consumption (ML) / number of annual supply charges}}$$

Using this relationship, ACTEW Water estimated tier 1 and tier 2 consumption for each year 2014-15 to 2017-18 for each of the 49 weather years in the conditioning data. The proposed billed consumption forecast for each year 2014-15 to 2017-18 (set out in Table 4) are calculated as the average of these 49 scenarios.

**Table 4: ACTEW Water's revised billed consumption forecast (ML)**

| *Year* | *Tier 1* | *Tier 2* | *Total* |
|---|---|---|---|
| 2014-15 | 23,745 | 16,558 | 40,302 |
| 2015-16 | 24,238 | 16,064 | 40,302 |
| 2016-17 | 24,683 | 15,619 | 40,302 |
| 2017-18 | 25,086 | 15,216 | 40,302 |

# Appendix A – Selecting the form of the predictor variables

The predictor variables *cumBurrinjuckevap*, *cumhumidity* and *cumtemp* all take the form $\sum_{n=0}^{t} x_n(a + be^{cn})$. t, a, b and c must be determined. This section examines the sensitivity of the final model fit in order to defend the selection of the forms used in the model.

## cumBurrinjuckevap

The highest r-squared is generated with b equal to approximately 1 and c approximately -0.4. However, this provides no real improvement over a straight sum of the Burrinjuck evaporation, so for simplicity a and b are set to zero. Setting a to 0, b to 1 and c to -0.4 was also trialled and reduced the model r-squared.

| | | b | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| | 0 | 0.975015 | 0.975015 | 0.975015 | 0.975015 | 0.975015 | 0.975015 | 0.975015 | 0.975015 | 0.975015 | 0.975015 | 0.975015 |
| | -0.2 | 0.975015 | 0.975028 | 0.975011 | 0.974984 | 0.974956 | 0.974929 | 0.974906 | 0.974885 | 0.974866 | 0.97485 | 0.974836 |
| | -0.4 | 0.975015 | 0.975031 | 0.975025 | 0.975008 | 0.974984 | 0.974959 | 0.974933 | 0.974908 | 0.974885 | 0.974863 | 0.974843 |
| | -0.6 | 0.975015 | 0.97503 | 0.97503 | 0.97502 | 0.975003 | 0.974984 | 0.974962 | 0.97494 | 0.974919 | 0.974898 | 0.974878 |
| | -0.8 | 0.975015 | 0.975028 | 0.975031 | 0.975025 | 0.975013 | 0.974998 | 0.97498 | 0.974961 | 0.974942 | 0.974923 | 0.974904 |
| c | -1 | 0.975015 | 0.975027 | 0.975031 | 0.975027 | 0.975018 | 0.975006 | 0.974991 | 0.974975 | 0.974957 | 0.97494 | 0.974923 |
| | -1.2 | 0.975015 | 0.975027 | 0.975031 | 0.975028 | 0.975021 | 0.975011 | 0.974998 | 0.974984 | 0.974968 | 0.974952 | 0.974936 |
| | -1.4 | 0.975015 | 0.975026 | 0.97503 | 0.975029 | 0.975023 | 0.975014 | 0.975003 | 0.97499 | 0.974975 | 0.97496 | 0.974945 |
| | -1.6 | 0.975015 | 0.975026 | 0.97503 | 0.975029 | 0.975024 | 0.975016 | 0.975006 | 0.974994 | 0.974981 | 0.974967 | 0.974952 |
| | -1.8 | 0.975015 | 0.975025 | 0.97503 | 0.97503 | 0.975025 | 0.975018 | 0.975009 | 0.974997 | 0.974985 | 0.974971 | 0.974957 |
| | -2 | 0.975015 | 0.975025 | 0.97503 | 0.97503 | 0.975026 | 0.975019 | 0.97501 | 0.975 | 0.974988 | 0.974975 | 0.974961 |

The sensitivity to the number of days in the summation is shown below. 27 (four weeks) was selected to achieve the highest r-squared.

| | Days | | | |
|---|---|---|---|---|
| | 6 | 13 | 20 | 27 |
| Model $R^2$ | 0.9746 | 0.9748 | 0.9749 | 0.9750 |

## cumhumidity

The highest r-squared is generated with c equal to -0.8 and a large value of b. However, setting b to 10 and c to -1 produces the same r-squared to four decimal places, so this was selected. Setting a to zero also slightly improves the r-squared value (to 0.975028) but, given that this did not provide a meaningful improvement this was not implemented in the model for consistency with the other variables.

| | | b | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0 | 1 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 |
| | 0 | 0.974671 | 0.974671 | 0.974671 | 0.974671 | 0.974671 | 0.974671 | 0.974671 | 0.974671 | 0.974671 | 0.974671 | 0.974671 |
| | -0.2 | 0.974671 | 0.97476 | 0.974857 | 0.974868 | 0.974872 | 0.974875 | 0.974876 | 0.974877 | 0.974878 | 0.974878 | 0.974878 |
| | -0.4 | 0.974671 | 0.974796 | 0.974954 | 0.974973 | 0.97498 | 0.974983 | 0.974985 | 0.974987 | 0.974988 | 0.974988 | 0.974989 |
| | -0.6 | 0.974671 | 0.974808 | 0.974996 | 0.975017 | 0.975023 | 0.975027 | 0.975029 | 0.97503 | 0.975031 | 0.975032 | 0.975032 |
| | -0.8 | 0.974671 | 0.97481 | 0.975011 | 0.97503 | 0.975035 | 0.975037 | 0.975038 | 0.975039 | 0.975039 | 0.975039 | 0.975039 |
| c | -1 | 0.974671 | 0.974809 | 0.975015 | 0.975029 | 0.975031 | 0.975032 | 0.975031 | 0.975031 | 0.975031 | 0.975031 | 0.975031 |
| | -1.2 | 0.974671 | 0.974807 | 0.975014 | 0.975024 | 0.975023 | 0.975021 | 0.975019 | 0.975018 | 0.975017 | 0.975016 | 0.975015 |
| | -1.4 | 0.974671 | 0.974805 | 0.975011 | 0.975016 | 0.975012 | 0.975008 | 0.975005 | 0.975003 | 0.975001 | 0.975 | 0.974998 |
| | -1.6 | 0.974671 | 0.974803 | 0.975008 | 0.975009 | 0.975002 | 0.974996 | 0.974992 | 0.974989 | 0.974986 | 0.974984 | 0.974982 |
| | -1.8 | 0.974671 | 0.974801 | 0.975005 | 0.975002 | 0.974993 | 0.974985 | 0.97498 | 0.974975 | 0.974972 | 0.97497 | 0.974967 |
| | -2 | 0.974671 | 0.9748 | 0.975002 | 0.974996 | 0.974984 | 0.974976 | 0.974969 | 0.974964 | 0.97496 | 0.974957 | 0.974954 |

The sensitivity to the number of days in the summation is shown below. 6 (one week) was selected to achieve the highest r-squared.

| | Days | | | |
|---|---|---|---|---|
| | 6 | 13 | 20 | 27 |
| Model $R^2$ | 0.9750 | 0.9748 | 0.9743 | 0.9731 |

## cumtemp

A simple summation of the temperature over four weeks provided the highest model r-squared, as shown below.

| | | b | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| | 0 | 0.975015 | 0.975015 | 0.975015 | 0.975015 | 0.975015 | 0.975015 | 0.975015 | 0.975015 | 0.975015 | 0.975015 | 0.975015 |
| | -0.2 | 0.975015 | 0.974444 | 0.973865 | 0.973328 | 0.97285 | 0.972432 | 0.972067 | 0.971751 | 0.971474 | 0.971232 | 0.971019 |
| | -0.4 | 0.975015 | 0.974626 | 0.974193 | 0.973746 | 0.973304 | 0.972877 | 0.972475 | 0.972099 | 0.971752 | 0.971432 | 0.971139 |
| | -0.6 | 0.975015 | 0.974718 | 0.974385 | 0.974032 | 0.97367 | 0.973309 | 0.972954 | 0.972612 | 0.972284 | 0.971972 | 0.971678 |
| | -0.8 | 0.975015 | 0.974767 | 0.974491 | 0.974196 | 0.973889 | 0.973577 | 0.973266 | 0.972959 | 0.972659 | 0.972369 | 0.972091 |
| c | -1 | 0.975015 | 0.974797 | 0.974556 | 0.974297 | 0.974026 | 0.973749 | 0.973469 | 0.97319 | 0.972915 | 0.972646 | 0.972384 |
| | -1.2 | 0.975015 | 0.974817 | 0.974598 | 0.974364 | 0.974118 | 0.973865 | 0.973608 | 0.97335 | 0.973094 | 0.972841 | 0.972594 |
| | -1.4 | 0.975015 | 0.974831 | 0.974628 | 0.97441 | 0.974182 | 0.973946 | 0.973706 | 0.973464 | 0.973222 | 0.972983 | 0.972747 |
| | -1.6 | 0.975015 | 0.974841 | 0.974649 | 0.974444 | 0.974229 | 0.974006 | 0.973778 | 0.973548 | 0.973317 | 0.973088 | 0.972861 |
| | -1.8 | 0.975015 | 0.974848 | 0.974665 | 0.974469 | 0.974263 | 0.97405 | 0.973832 | 0.973611 | 0.973389 | 0.973167 | 0.972948 |
| | -2 | 0.975015 | 0.974853 | 0.974677 | 0.974488 | 0.974289 | 0.974084 | 0.973873 | 0.973659 | 0.973443 | 0.973228 | 0.973014 |

The sensitivity to the number of days in the summation is shown below. 27 (four weeks) was selected to achieve the highest r-squared.

| | Days | | | |
|---|---|---|---|---|
| | 6 | 13 | 20 | 27 |
| Model $R^2$ | 0.9669 | 0.9699 | 0.9730 | 0.9750 |

# Appendix B – R Output

```
R version 3.0.2 (2013-09-25) -- "Frisbee Sailing"
Copyright (C) 2013 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> setwd("B:/2014-15/2014-08-DemandModels")
> mydata=read.csv("Book2.csv")
> str(mydata)
'data.frame':   1298 obs. of  6 variables:
 $ Releases    : num  116 117 118 119 120 ...
 $ cumhumidity2: num  46 46 46 46 46 ...
 $ cumevap     : num  111 112 113 113 114 ...
 $ cumtemp     : num  668 671 674 677 680 ...
 $ cumrain2    : num  6.46 6.42 6.37 6.32 6.35 ...
 $ season      : num  0.865 0.869 0.873 0.878 0.882 ...
> model1 = lm(Releases ~ cumhumidity2 + cumevap + cumtemp + cumrain2 + season,data=mydata)
> summary(model1)

Call:
lm(formula = Releases ~ cumhumidity2 + cumevap + cumtemp + cumrain2 +
    season, data = mydata)

Residuals:
    Min      1Q  Median      3Q     Max
-9.8117 -2.1394 -0.5038  2.2907  8.9333

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  590.875170  20.210638  29.236  < 2e-16 ***
cumhumidity2 -11.249752   0.444310 -25.320  < 2e-16 ***
cumevap        0.158461   0.022126   7.162 1.33e-12 ***
```

```
cumtemp       0.108310   0.004842  22.371  < 2e-16 ***
cumrain2     -3.706119   0.134132 -27.630  < 2e-16 ***
season      -29.490089   2.147854 -13.730  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.494 on 1292 degrees of freedom
Multiple R-squared:  0.975,     Adjusted R-squared:  0.9749
F-statistic: 1.008e+04 on 5 and 1292 DF,  p-value: < 2.2e-16

> step(model1, direction="backward")
Start:  AIC=3253.47
Releases ~ cumhumidity2 + cumevap + cumtemp + cumrain2 + season

               Df Sum of Sq   RSS    AIC
<none>                      15770 3253.5
- cumevap       1     626.1 16396 3302.0
- season        1    2301.0 18071 3428.3
- cumtemp       1    6108.6 21879 3676.4
- cumhumidity2  1    7824.9 23595 3774.5
- cumrain2      1    9318.4 25088 3854.1

Call:
lm(formula = Releases ~ cumhumidity2 + cumevap + cumtemp + cumrain2 +
    season, data = mydata)

Coefficients:
 (Intercept)  cumhumidity2       cumevap       cumtemp      cumrain2        season
    590.8752      -11.2498        0.1585        0.1083       -3.7061      -29.4901

> step(model1, direction="forward")
Start:  AIC=3253.47
Releases ~ cumhumidity2 + cumevap + cumtemp + cumrain2 + season


Call:
lm(formula = Releases ~ cumhumidity2 + cumevap + cumtemp + cumrain2 +
    season, data = mydata)

Coefficients:
 (Intercept)  cumhumidity2       cumevap       cumtemp      cumrain2        season
    590.8752      -11.2498        0.1585        0.1083       -3.7061      -29.4901
```