Cold Water Temperature in Melbourne 1994-2013, preliminary statistical analysis.

Project: Water-Energy-Carbon Links in Households and Cities: A New Paradigm

Adam Grace, Thomas Taimre, Steven Kenway, Julijana Bors

The University of Queensland
Water-Energy-Carbon Group
2 December 2014

Document Control

Date	Revision	Prepared By	Reviewed By	Approved By
15/01/2014	Draft	A. Grace	T. Taimre, S. Kenway	S. Kenway
15/05/2014	Final Draft	S. Kenway	Project team	S.Kenway
02/12/2014	Final	S. Kenway	Yarra Valley Water, City West Water, South East Water	S. Kenway

Abstract

In this report, we present a preliminary statistical analysis of cold water temperature data from part of the suburban Melbourne water system. Accurate characterisation of cold water temperature is an important variable for estimating water-related energy, water quality management and other reasons including efficiency analysis of hot water systems.

The data consisted of approximately 55,000 cold water temperature measurements over the period from 1994 to 2013, recorded by Yarra Valley Water. We consider possible reasons for changes in cold water temperature over time, including an apparent increase in water temperatures from pre-2004 to post-2006. The effect of sample selection criteria is specifically considered. Where relevant, scientific hypotheses are tested statistically as a first step to understanding these causes.

The research demonstrates that there is no *consistent, constant, year-on-year* increase in temperature over time. Secondly, it established that both the location and time of day for cold water temperature records has changed significantly through time. Thirdly, no evidence was found that these changes in sampling methodology contributed to warmer recorded temperatures over time.

The current water temperature samples are collected as a component of a distributed water quality monitoring program. However the dataset also has high values related to management of water-related energy in cities due to the strong influence of water temperature. While the current dataset is understood to be of the highest available resolution for an urban system, it is possible that its value could be further improved from the perspective of managing energy consumption in cities. Redesigning the current sampling program could be undertaken to minimise controllable sampling variation. Such work could lead to a data set of greater utility and insight for the purpose of understanding and managing water-related energy, from the existing on-going sampling effort.

We suggest that a subsequent investigation involving structural and meteorological data (identified as relevant but not able to be considered in this report) would be informative, and could elucidate in greater detail, primary factors of influence on water temperature.

This report, together with analysis of cold water temperatures in Melbourne (Bors et al 2014) contributes greater, though initial, understanding of variability in water temperature in Melbourne and resultant influences on water-related energy associated with water heating. This report contributes towards a statistically rigorous methodology for characterising water temperature in a suburban water network, and evaluating changes through time.

This work has been undertaken as a component of a collaborative research project funded by the Smart Water Fund and the Australian Research Council. The overall project is aimed at understanding and managing the connection between residential water and energy use, related costs and greenhouse gas emissions.

Table of Contents

List of Figures

List of Tables

Abbreviations

iid	Identically and independently
f	A given distribution from which samples or measurements are taken

Acknowledgements

The authors acknowledge the support of the Smart Water Fund including Jessica Yeung, Project Coordinator; Yarra Valley Water: including Francis Pamminger, (Manager, Research & Innovation) Peter Roberts (Demand Management) and Asoka Jayaratne, (Water Quality Specialist); South East Water: (Sam Johnson, Greenhouse and Energy Specialist); City West Water: (John Fawcett, Manager, Business Resource Efficiency); The University of Queensland (Prof Paul Lant, Amanda Binks, Research Engineer). The support of the Australian Research Council is also acknowledged.

1. Introduction

The temperature of water temperature has considerable influence on household energy use (Kenway et al., 2013, Kenway et al., 2014). Knowledge of water temperature is also important for a range of other reasons including management of distributed water quality and efficiency analysis of hot water systems. It is therefore important to quantitatively characterise the temperature of cold water in water systems.as well as being important to the management of water quality. Cold water temperature, which is the temperature of water in the mains lines supplying households and industry. It, directly influences the levels of energy consumption in households when water is heated (eg with energy fuelled appliances).

Water-related energy consumption in Australian cities accounts for approximately 6800 kWh (or 9%) of total primary energy per person (0.78 kW/person) (Kenway et al., 2011). Residential households account for about 40% of the effect. This amounts to 2040 kWh per person and year (0.23 kW per person) of primary energy use. Consequently, water-related energy in households in Australia contributes to approximately 4% of national primary energy use.

Water temperature in the distribution network can significantly influence household energy use and related greenhouse gas emissions. Preliminary modelling of five households in Melbourne indicates that 5-20 kWh/hh.d (15-40% of total household energy use) is influenced in some way by water (Binks et al., 2014). Preliminary sensitivity analysis in the households shows that a 10% change in the temperature of cold water (a change of about 2°C), can influence 0.3-0.7 kWh/hh.d energy use. This is equivalent to around 3-15% of water-related energy use, or 3-5% of total household energy use (Kenway et al., 2014).

In this report we present a statistical analysis of cold water temperature data from part of the suburban Melbourne water system. We consider possible reasons for changes in cold water temperature over time, including an apparent increase in water temperatures from pre-2004 to post-2006. The effect of sample selection criteria is specifically evaluated. This analysis was initiated following a characterisation of water temperatures in Melbourne undertaken for the purpose of improving estimates of household water-related energy use (Bors and Kenway, 2014).

This work has been undertaken as a component of a collaborative research project funded by the Smart Water Fund and the Australian Research Council. Overall, the project aims to understand connections between residential water and energy use, related costs and greenhouse gas emissions. It has objectives as:
1) Understand water and energy connections in individual households;
2) Characterise "household types" with regard to water-related energy;
3) Understand city scale water-related energy & greenhouse gas emissions; and
4) Identify opportunities to reduce water-related energy;
This particular report is intended to help unpack some of the statistical challenges of achieving these objectives, by addressing one relatively well-characterised and influential variable.

2. Data Overview

This investigation was undertaken on a data set provided by Yarra Valley Water and South East Water. The data detailed approximately 55,000 cold water temperatures at specific times and specific locations in suburban Melbourne from 1994 to 2013. Additional data sets for elsewhere in Melbourne (Seqwater and City West Water) were also investigated (Refer also to Bors 2014), were however these were generally sampled less frequently and consequently not used in this detailed analysis.

Figure 1 plots the cold water temperature data over the recorded time period, clearly displaying seasonality. Table 1 states the total number of cold water temperature records (sample sizes) for each month. There are two clear changes in the cold water temperature records that occur around 1998 and 2005. There is much more data reported in the final time period (2005-2013), which corresponds to years after the second reporting change. The sample sizes per month, aggregated over all the years, are of similar size, and in particular there does not seem to be sample size bias toward any particular month (or season).

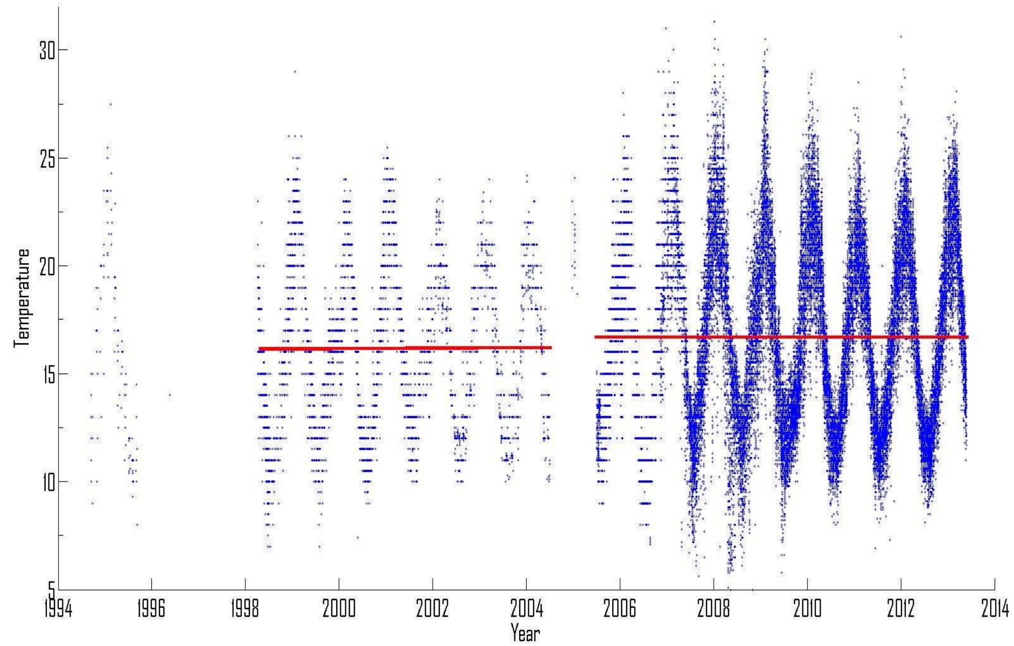

Figure 1: Cold water temperature over time from Yarra Valley Water data. The average aggregate monthly temperature for the contiguous time-periods (1998-2004) and (2005-2013) are shown in red.

There is also an apparent increase in the monthly temperatures of the period 2005-2013 when compared with those from 1998-2004. This can be seen directly by comparing the mean cold water temperature of each month, aggregated over the corresponding two time periods. Table 2 gives these means for the Yarra Valley Water data (in degrees Celsius). Similarly, Table 3 gives these means for the South East Water data, which had measurements from 2008 to 2013. The average aggregated monthly means are illustrated for the Yarra Valley Water data in Figure 1 as red lines superimposed for the two time periods. Yarra Valley Water had averages of 16.2°C (1998-2004) and 16.6°C (2005-2013). South East Water had an average of 16.5°C (2008-2013). While a temperature change of this magnitude has minimal significance, it is clear from Figure 1 that the pattern of sampling is also quite distinct in the periods from pre-2004 to post-2006. The change in average aggregated monthly mean temperature observed in the Yarra Valley Water data is discussed in more detail in Section 4. Seasonal variability does have consequence due to the impact on energy demands for water heating.

Table 1: Samples sizes for each month in Yarra Valley Water data.

Monthly Sample Sizes (Yarra Valley Water)												
	Jan	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
1994	0	0	0	0	0	0	0	0	28	7	9	9
1995	11	8	11	10	8	11	11	10	7	0	0	0
1996	0	0	0	0	2	0	0	0	0	0	0	0
1997	0	0	0	0	0	0	0	0	0	0	0	0
1998	0	0	0	436	576	643	199	47	51	61	96	140
1999	144	67	96	80	86	101	69	76	81	77	79	94
2000	81	86	82	79	86	78	85	96	83	100	86	83
2001	92	83	79	80	94	84	86	81	38	40	45	41
2002	37	41	40	41	39	39	39	38	39	40	38	39
2003	86	76	78	78	76	78	94	58	68	70	68	68
2004	84	66	68	64	68	68	0	0	0	0	0	24
2005	13	0	0	0	0	0	429	431	455	421	422	458
2006	432	429	454	427	426	456	415	459	415	456	442	405
2007	372	392	444	418	505	341	526	576	452	523	501	487
2008	512	514	491	517	497	489	485	460	488	507	450	560
2009	463	449	496	495	463	513	507	476	495	577	574	551
2010	532	514	609	525	489	502	476	486	497	469	501	496
2011	513	501	515	494	506	513	511	519	528	517	508	506
2012	508	476	515	483	489	485	497	607	442	507	502	446
2013	542	444	460	486	390	0	0	0	0	0	0	0

The data in December 2004 and January 2005 was omitted for most of this report as it does not clearly belong to any main reporting period (There is also an apparent increase in the monthly temperatures of the period 2005-2013 when compared with those from 1998-2004. This can be seen directly by comparing the mean cold water temperature of each month, aggregated over the corresponding two time periods. Table 2 gives these means for the Yarra Valley Water data (in degrees Celsius). Similarly, Table 3 gives these means for the South East Water data, which had measurements from 2008 to 2013. The average aggregated monthly means are illustrated for the Yarra Valley Water data in Figure 1 as red lines superimposed for the two time periods. Yarra Valley Water had averages of 16.2°C (1998-2004) and 16.6°C (2005-2013). South East Water had an average of 16.5°C (2008-2013). While a temperature change of this magnitude has minimal significance, it is clear from Figure 1 that the pattern of sampling is also quite distinct in the periods from pre-2004 to post-2006. The change in average aggregated monthly mean temperature observed in the Yarra Valley Water data is discussed in more detail in Section 4. Seasonal variability does have consequence due to the impact on energy demands for water heating.
Table 1). Moreover, the data from 1994-1996 was not the primary focus of this report due to its relatively low sample size and large temporal separation from the bulk of the data.

Table 2: Aggregated mean recorded temperature per month in Yarra Valley Water data, over two reporting periods (1998-2004) and (2005-2013). The later time period had higher aggregated monthly mean temperature values for most months.

Monthly Mean Temperature for Yarra Valley Water (°C)												
	Jan	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
1998-2004	20.25	21.07	20.25	18.09	15.50	12.77	11.47	11.79	13.18	14.35	16.96	18.63
2005-2013	20.93	21.92	20.54	18.20	15.03	12.70	11.66	11.99	13.44	15.57	17.97	19.69

Table 3: Aggregated mean recorded temperature per month in South East Water data, together with their corresponding sample sizes.

Monthly Mean Temperature South East Water (°C)												
	Jan	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
2008-2013	20.64	20.85	20.31	17.99	15.50	13.11	12.42	12.44	13.3	15.09	17.61	18.99
Sample size	170	149	164	161	169	156	128	133	159	163	162	144

3. Possible Causes of Temperature Increase

The increase in cold water temperature between the two time periods found in the Yarra Valley Water data must have a satisfactory explanation. The possible causes can be categorised as follows:

a) **Meteorological / Environmental** – the water temperature was affected by natural environmental factors, such as changes in air temperature over time.

b) **The Water "System" (i.e. structural causes)** – the water temperature was affected by changes in the water delivery system. For example, the water was sourced from a dam at different depths, or a different region, or key pipe infrastructure changed over time. Alternatively, it is possible that differences exist in the depth of the pipes in the area around the sample points, and/or that the volume of water going through the network is different. For example in 1998-2004 the average annual volume was 188 GL and in 2005-2013 the average volume was 146 GL (Peter Roberts, Yarra Valley water *pers. comm. May 2014*).

c) **Sample Program Design (selection of sites and times)** – the households were sampled differently in the two time periods. For example, apartment blocks may have been sampled more frequently, or the cold water temperature recordings could have been later in the day, during one of the time periods than the other.

d) **Sample Collection Procedure (data measurement)** – the method for measuring and recording temperature may change over time, resulting in different recorded data for similar water temperatures. For example, changes in the time allotted to flushing water through the pipes before measurement, or the recording equipment could influence the recorded data.

The first two categories are *environmental or physical*, whereas the last two are *statistical* and directly related to the sampling methodology. Table 4 outlines preliminary (not exhaustive) possible reasons in each category for the noted temperature increase.

Preliminary examination of three of the sample program (selection) factors is undertaken in Section 4. The other factors which may have contributed to the recorded rise in average cold water temperature in the Yarra Valley Water data are not examined further in this report, with the exception of a brief investigation into air temperature.

Approximately 5,000 data points representing maximum air temperatures over the period 2000-2013 were collected from the Viewbank weather station in Melbourne. This weather station is inside the region which accesses Yarra Valley Water's resources. Figure 2 shows the mean monthly air temperatures. There is clear seasonal variation over the time period.

Table 4: A partial list of potential causes for the noted temperature increase

Possibilities Examined In This Report	Possibilities Not Examined in This Report
a) Meteorological / Environmental	
• Different weather conditions – The weather could have changed significantly over twenty years, thereby influencing cold water temperature.	• Different terrain around pipes or dam infrastructure.
b) The Water System (structural)	
	• Different water sourcing method at dam. • Different water delivery infrastructure from the dam.
c) Sample Program Design (sample selection)	
• Cold water temperature increases over time – Time (year) could be a predictor variable. • Time of day that measuring occurs – there is a consistent correlation with time of day and temperature. That is, higher temperatures are more likely to have been measured in the afternoon. • Location of selected households– Different households have different ambient temperatures influenced by local climatic conditions or aspect. The households chosen for recording should be selected according to a set of criteria which ensure representative sampling (such as random allocation) and such selection should not change over time.	• Time-delayed temperature effects (or the influence of (previous) weather conditions, for example if rain periods influence sampling in particular regions. • Different sampling of water sources. • Different sampling of water delivery systems.
d) Sample Collection Procedure (data measurement)	
	• Different measuring methodology. • Different measuring equipment.

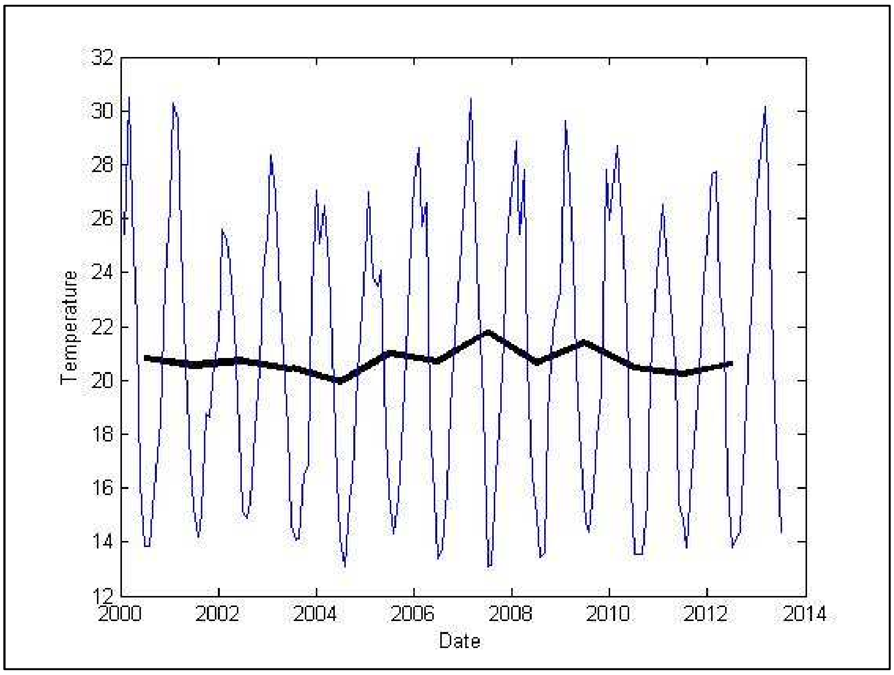

Figure 2: Viewbank weather station mean monthly maximum air temperatures (blue, thin) and mean yearly maximum air temperatures (black, thick).

After initial inspection and analysis of the data we concluded there was no *clear, strong* evidence that explained the observed increase in cold water temperature, despite the notable increase in average yearly maximum air temperature after 2004. Despite this finding there is clearly a very strong relationship between air temperature and water temperature. Consequently, air temperature is likely to be a strong factor in "actual" water temperature. However, such changes can only be detected if sampling methodologies between the two time-periods are consistent.

4. Analysis of Sample Selection

In order to investigate possible sample selection reasons for an increase in cold water temperature over time, we first construct a statistical framework for the data. We assume that the sample of approximately 55,000 data records X1, X2, X3, … were drawn from a common distribution, f. We assume that each data point, consisting of a temperature, date, time, and location (e.g. X = (17.4, 15/6/2006, 14:30, 15 Maple Street – Brunswick), was drawn independently and identically ("iid") from this common distribution f.[1]

Potential shortcomings of this framework are:

1) Data values could be incorrect (slightly inaccurate recordings).
2) Data is not independently distributed (some data may be correlated).
3) Data is not identically distributed or may come from different sources (there may be multiple distributions f1, f2, f3, for data records over time).

The specific statistical investigations carried out in sections 4.1, 4.2 and 4.3, each deal slightly differently with this underlying framework. These differences are explained in the relevant sections hereafter.

4.1. Cold Water Temperature Increases over Time

Figure 1 and Table 2 show an increase in average aggregate monthly temperature between the two contiguous time-periods (1998-2004) and (2005-2013). In order to test whether there is a consistent, constant, year-on-year increase in cold water temperature in addition to the large seasonal changes, we construct a simple linear model for the mean cold water temperature of each month, in which time is a predictor variable. This model is fitted to all of the Yarra Valley Water data, as well as separately to data from the two time periods of interest. If there were a consistent, constant, year-on-year increase, then the model fits in the two time periods should be similar to each-other, as well as similar to the overall model fit.

The means of each month were calculated for the entire recorded period, 1994-2013. These mean values were then plotted chronologically for each month. Simple linear regression was used to represent the apparent increase in cold water temperature over time. Lines of best fit were calculated for each month, using the model

Equation 1: **Mean monthly temperature** $(\boldsymbol{M}) = \boldsymbol{A} * \boldsymbol{t} + \boldsymbol{B}$

where A and B are constants and t is the date of the measurement (within the range [1994, 2014]).

[1] Iid means independently and identically. This common statistical term essentially means that the sampling method remains the same from sample to sample, and sample outcomes do not influence the sampling method. f is a given distribution (the distribution from which the house, date, time, temperature is sampled).

Figure 3 shows the mean values plotted over time for four months representative of the four seasons, with their corresponding fitted linear regression lines. Table 5 lists the coefficients for the regression model along with their corresponding 95% confidence intervals. Note that A is the parameter of interest, which represents the estimated constant year-on-year temperature increase for a particular month (in degrees Celsius), and for simplicity we have not included B values. As expected, most of the twelve (monthly) gradient values (A) were positive, indicating that mean monthly temperature *recordings* have been increasing over the last 20 years. Most months show a slight, but statistically significant, increase in temperature over time. This is indicated by a positive value for the lower 95% bound for the slope coefficient (A). This however does not establish time as a causative factor.

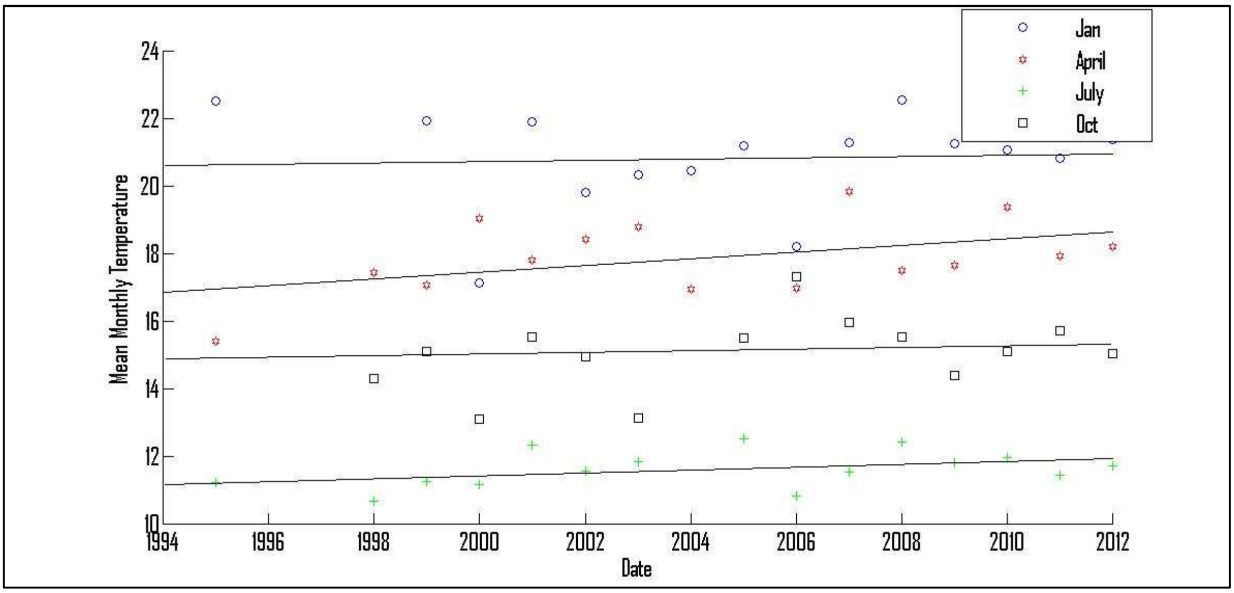

Figure 3: Monthly mean values for Jan, April, July, and Oct, with fitted linear regression lines.

Table 5: Regression coefficients for the slope of a linear regression model of monthly cold water temperature (over the entire period) together with corresponding 95% confidence intervals.

	Total Period (1994-2013)		
	Slope	**- 95% CI**	**+**
Jan	0.0231	0.0227	0.0235
Feb	-0.0041	-0.0044	-0.0039
Mar	0.0546	0.0544	0.0548
Apr	0.0877	0.0875	0.0880
May	-0.2077	0.2084	-0.2070
Jun	-0.0046	-0.0047	-0.0044
Jul	0.0411	0.0409	0.0412
Aug	0.0658	0.0656	0.0659
Sep	0.0663	0.0662	0.0664
Oct	0.0207	0.0204	0.0210
Nov	0.0745	0.0743	0.0747
Dec	0.0655	0.0652	0.0658

In order to establish consistency over the entire time period (1994-2013), a similar analysis was performed for each of the two smaller time periods of interest, (1998-2004) and (2005-2013). The corresponding results of the linear regression model fits are presented in Table 6 and Table 7.

By examining the estimated coefficients in Table 5, Table 6 and Table 7, it is apparent that the year-on-year temperature change for each month is not consistent across different time periods. Moreover, from Table 6 and Table 7, neither of the smaller time periods provides consistent evidence of a year-on-year increase in cold water temperature. In fact these two periods have a slight decreasing trend.

This inconsistency implies that the linear regression model is not appropriate for the Yarra Valley Water data. An alternative model consistent with the data is a (roughly) constant step increase in recorded temperature that occurs between the two time periods (where sampling methodology appears to have changed).

Table 6 and Table 7: Regression coefficients for the slope of a linear regression model of monthly cold water temperature over both time periods together with corresponding 95% confidence intervals. Many months have a decreasing temperature trend for both time periods.

	First Period (1998-2004)		
	Slope	**- 95% CI**	**+**
Jan	-0.0001	-0.0010	0.0008
Feb	-0.4908	-0.4911	-0.4905
Mar	-0.3099	-0.3101	-0.3096
Apr	0.0505	0.0501	0.0508
May	-0.0435	-0.0438	-0.0432
Jun	0.0141	0.0139	0.0143
Jul	0.2259	0.2257	0.2261
Aug	0.1518	0.1516	0.1520
Sep	0.0957	0.0955	0.0960
Oct	-0.1083	-0.1088	-0.1077
Nov	0.0099	0.0097	0.0102
Dec	0.3812	0.3805	0.3819

	Second Period(2005-2013)		
	Slope	**- 95% CI**	**+**
Jan	0.1250	0.1245	0.1254
Feb	-0.0634	-0.0636	-0.0631
Mar	-0.0588	-0.0592	-0.0585
Apr	0.0449	0.0445	0.0453
May	0.0144	0.0139	0.0148
Jun	0.0895	0.0891	0.0898
Jul	-0.0236	-0.0238	-0.0233
Aug	0.0101	0.0099	0.0102
Sep	-0.0457	-0.0459	-0.0456
Oct	-0.1807	-0.1810	-0.1803
Nov	-0.0378	-0.0381	-0.0375
Dec	0.0064	0.0060	0.0068

4.2. Time of Day That Measuring Occurs

The time of day that the temperature was measured is present for most data records. From investigation these records, we suspect there are mild problems with the accuracy of these times (for example records at unusually early or late times). However, the influence of these records is not critical on the overall consistent positive correlation between cold water temperature and the time of day that the temperature was recorded. That is, higher temperatures are more likely to have been measured in the afternoon rather than the early morning. Figure 4 below shows that this correlation is similar over eight years for a single month (April). A similar investigation for all months was performed, and they were found to display similar correlation behaviour. There is a lot of variation in the data so the correlation is not obviously strong from examining plots. However, it is more apparent upon examining the correlation coefficients for each month, which are collected in Table 8. These correlation coefficients are also known in statistical literature as “R squared” values. Here, they quantify how much of the random variation in the cold water temperature can be attributed (linearly) to the time of day. The positive R squared values across all months and years illustrates that this correlation is inherent to the data, despite varying meteorological / environmental conditions, locations, and possible rough reporting accuracy. Note that Table 8 only reports monthly values in the more recent time block (2005-2013) due to issues with data records in the earlier time block (1998-2004).

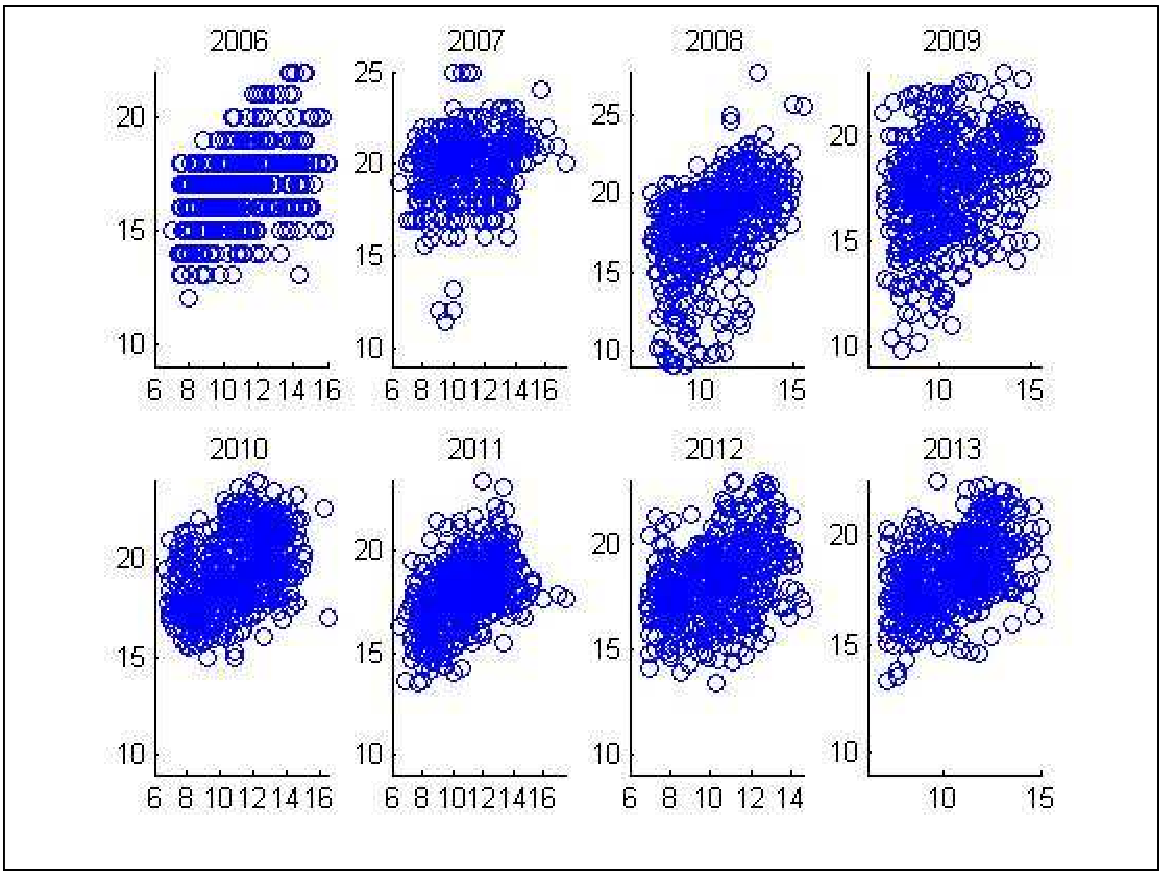

Figure 4: Scatter plots showing temperature (Y axis, in °C) versus time of recording (X axis, on a 24 hour scale) for the month of April from 2006 to 2013. Note there is a consistent, but slight, positive correlation.

Table 8: R squared values representing the correlation between temperature and time of day temperature was recorded. All values are positive and of similar magnitude.

Correlation Coefficient Values (R squared)*												
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2005							0.0809	0.1076	0.0576	0.0476	0.1155	0.1983
2006	0.1092	0.1726	0.2751	0.1631	0.1269	0.0502	0.1629	0.0439	0.0244	0.1014	0.1231	0.1240
2007	0.1642	0.0154	0.0001	0.0606	0.1542	0.1684	0.1648	0.2560	0.1667	0.1730	0.1480	0.1782
2008	0.1809	0.0000	0.2428	0.2299	0.1606	0.0166	0.1419	0.1500	0.1006	0.0606	0.0762	0.0456
2009	0.0079	0.1337	0.1841	0.0920	0.2406	0.1166	0.2589	0.1858	0.0887	0.1127	0.0871	0.1035
2010	0.0950	0.1284	0.1509	0.1492	0.1136	0.1825	0.2102	0.1906	0.1650	0.1800	0.1317	0.1688
2011	0.1159	0.1828	0.1679	0.2155	0.1627	0.2391	0.2053	0.3203	0.2606	0.1028	0.0821	0.1770
2012	0.0476	0.1563	0.0303	0.1547	0.1154	0.1626	0.2079	0.1647	0.1939	0.1811	0.0682	0.1192
2013	0.1514	0.1025	0.0783	0.2062	0.1936							

**Blank cells had no data*

Based on the observed positive correlation between cold water temperature and time of day, it is natural to check whether the increase in average temperature of the records from (1998-2004) to

(2005-2013) could be due to a change in the time of day at which data is recorded. In other words, we would like to see if records from (2005-2013) have a higher proportion of later recording times than the records from (1998-2004), which would contribute to higher recorded temperatures in (2005-2013).

In order to achieve this, in each of the two time periods of interest, we compare the proportion of records that are sampled after a certain time of day (e.g. midday). Figure 5 shows the time-of-day distributions for both time periods. Surprisingly, the period (1998-2004) has a greater proportion of temperature records from later in the day, and due to the great number of records for each time period, we can readily conclude that this difference is real (that is, not a sampling anomaly).

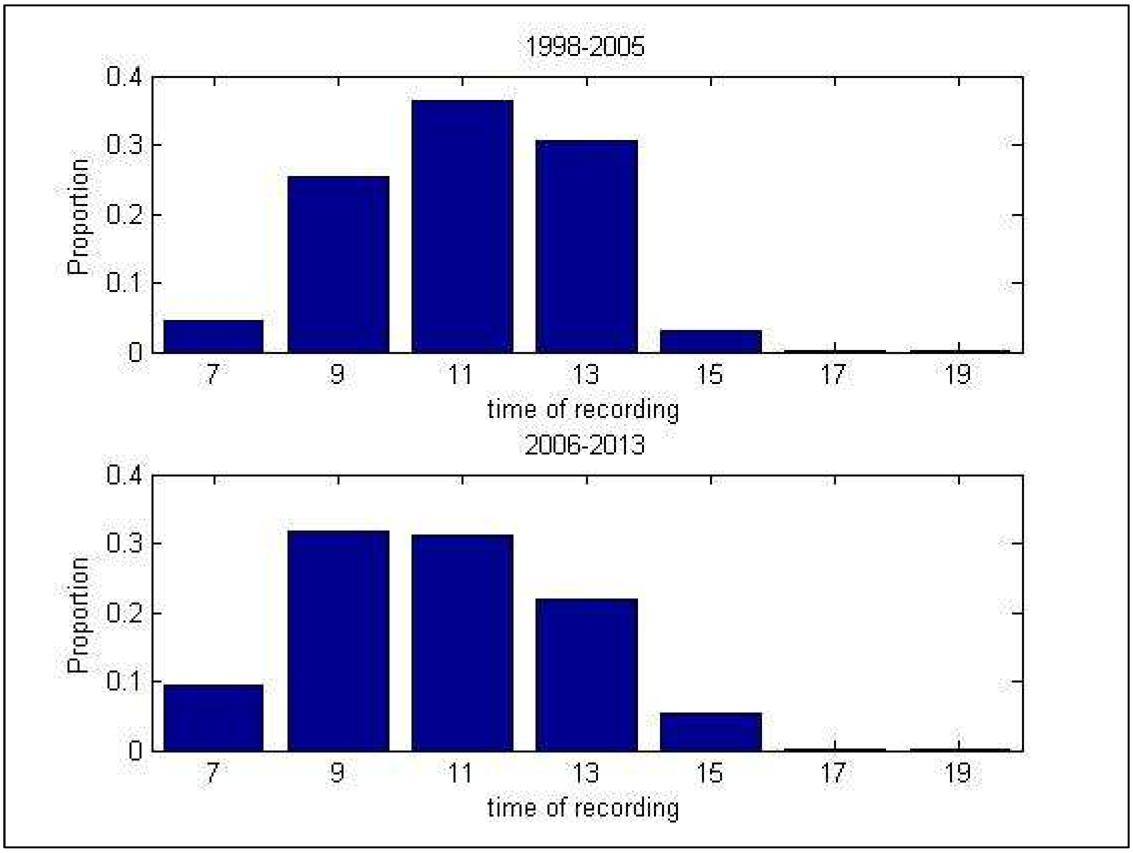

Figure 5: Distribution of times of cold water temperature records, for the two time periods (1998-2004) and (2005-2013). Each entry on the X axis represents two-hour windows, centred at the indicated 24-hour time (i.e. 9 represents times between 8am and 10am).

4.3. The Location of Households That Are Selected

It is well known that households from different locations have different cold water temperatures. The households chosen for recording should be selected according to a set of criteria which ensure representative sampling (such as random allocation) and such selection methodology should not change over time. We want to test whether the later time period has a higher proportion of locations which are typically warmer than the earlier time period. We achieve this in a similar way to the test in Section 4.2.

In order to describe the distribution of locations, the data was divided into 33 regions, each representing a geographical cluster of suburbs. The percentages and total counts of households' records from each region were calculated, forming an empirical distribution of location. Table 9 presents this calculated data, and also reports the mean temperature measurements for each region in both of the time periods of interest. Note that the mean temperatures vary greatly between the two periods. This may be due to differing proportions of measurements taken in different seasons. The final column of Table 9 presents the change in sampling percentage, suggesting differences in location distributions between the two time periods.

The null hypothesis is that the distribution of locations in which the temperature is measured is the same for both time periods, or symbolically,

H0: f1=f2,

where f1 and f2 represent the distribution for the earlier and later time blocks respectively. The alternative hypothesis is that the distribution of locations differs, or symbolically,

HA: f1 ≠ f2.

The Pearson chi-squared test can be performed to test how likely the suggestive change in distributions from Table 9 is simply due to chance. The test statistic is calculated by comparing the number of expected samples to the number of observed samples, in each region. High test statistics indicate it is unlikely such differing in sampling would occur by chance. If the null hypothesis holds, the test statistic is approximately chi-squared distributed, with "number of regions minus 1" degrees freedom. Symbolically, the test statistic is computed as

Equation 2: $$\chi^2 = \sum_{r=1}^{33} \frac{(O_r - E_r)^2}{E_r}$$

where r indexes the regions, and O and E are the observed and expected number of data records in a region, respectively.

The computed test statistic is extremely large (approximately 7,000, with 32 degrees freedom) for this data, resulting in a p-value of much less than one in a million. Therefore, we can safely reject the null hypothesis and conclude that the two sampling distributions are not the same. In other words, we conclude that the sampling methodology (with respect to location) has changed between the period (1998-2004) and the period (2005-2013). Note, however, that this is not evidence that this change in sampling actually contributed to the warmer temperature recordings in the second period.

Table 9 lists the regions ordered from highest to lowest in terms of mean cold water temperature over the second period (2005-2013). The warmer regions may appear to have a larger proportion of sampled households during the second time period. However, when the list is ordered similarly according to mean temperatures for the period 1998-2004, there appears to be a greater proportion of households in the second time period sampled from colder regions. In light of this, despite the conclusion that the distribution of locations changed between the two periods, it cannot be inferred that this change contributed to the higher recorded temperatures during the second time period.

Problems with this analysis are the accuracy of data, and the variation in the mean temperatures at each location due to both seasonal variation and specific day weather and time variation. However the difference in sampling appears obvious and strong regardless. A similar analysis is outlined in the appendix. This analysis also illustrates the difference in sampling between the two periods.

Table 9: Differences in sampling frequency, proportion and mean temperature for over 33 regions in the two time periods of interest*.

Region ID	1998-2004			2005-2013			Difference in Selection Probability
	Temperature	Selection Probability	Total	Temperature	Selection Probability	Total	
10	15.74	0.025457	177	17.867	0.048035	2205	0.022578
55	15.896	0.010931	76	17.779	0.017667	811	0.0067368
9	16.058	0.05494	382	17.657	0.061803	2837	0.0068626
7	16.19	0.046599	324	17.612	0.048928	2246	0.0023296
53	16.23	0.074213	516	17.547	0.069035	3169	-0.0051772
61	16.195	0.025025	174	17.459	0.041478	1904	0.016453
14	15.511	0.062707	436	17.125	0.059799	2745	-0.002908
50	15.459	0.028477	198	16.903	0.044419	2039	0.015942
59	16.334	0.060118	418	16.879	0.00464	213	-0.055478
18	15.927	0.017834	124	16.868	0.021676	995	0.0038416
17	16.121	0.031353	218	16.857	0.031174	1431	-0.0001796
20	16.415	0.024594	171	16.855	0.01599	734	-0.0086038
60	15.642	0.054221	377	16.839	0.061454	2821	0.0072331
52	15.662	0.044873	312	16.82	0.060125	2760	0.015253
12	15.927	0.022149	154	16.798	0.038973	1789	0.016824
16	16.509	0.006616	46	16.741	0.017253	792	0.010638
71				16.634	0.015576	715	0.015576
25	16.297	0.064001	445	16.584	0.019388	890	-0.044613
19	15.898	0.025025	174	16.51	0.017362	797	-0.0076628
51	15.714	0.042284	294	16.37	0.062151	2853	0.019868
49	15.785	0.048324	336	16.224	0.029017	1332	-0.019307
23	16.061	0.01654	115	16.185	0.016033	736	-0.00050616
48	15.588	0.041996	292	16.167	0.024726	1135	-0.017271
24	15.953	0.017259	120	15.988	0.016665	765	-0.00059352
56	15.801	0.016971	118	15.937	0.01697	779	-8.93E-07
22	15.723	0.018985	132	15.432	0.016861	774	-0.0021233
26	15.508	0.013376	93	15.231	0.017929	823	0.0045532
72	15.003	0.004315	30	15.115	0.016643	764	0.012329
21	16.471	0.009924	69	15.078	0.01758	807	0.0076564
46	14.587	0.021573	150	14.831	0.019301	886	-0.0022723
74	14.5	0.000863	6	14.616	0.016121	740	0.015258
47	14.787	0.066446	462	14.554	0.018996	872	-0.04745
73	15	0.002014	14	14.44	0.01623	745	0.014216

**Blank cells had no data*

5. Conclusion

This report investigates a potential increase in average recorded cold water temperatures in suburban Melbourne in Yarra Valley Water and South East Water data. A number of potential causes are identified and categorised. Analysis was carried out on three of these, where data was available. This included: (a) a year-by-year increase in temperature; (b) changes to sampling; and (c) influence of the change in sampling.

Firstly, it is demonstrated that there is no *consistent, constant, year-on-year* increase in temperature over time. Secondly, it is established that both the location and time of day for cold water temperature records has changed significantly over time in the data analysed. Thirdly, this investigation did not uncover any evidence that these changes in sampling methodology contributed to warmer recorded temperatures over time.

Through the course of this work it became clear that the current water temperature data set has been collected for water-quality management purposes. However, the dataset has wide potential value beyond this original objective, potentially including: (a) understanding the performance of hot water systems; (b) managing physical and biological processes in the water network; (c) understanding wastewater temperatures within the sewer network and related heat-recovery implications; and (d) understanding the influence of regional or local changes through time (Refer also to Bors 2014). By redesigning the current sampling program to minimise controllable sampling variation, this set of data will gain even higher value, thereby derive greater utility and insight from the on-going sampling effort.

We suggest that a subsequent comprehensive investigation involving structural and meteorological data would be informative, and could elucidate in greater detail, primary factors of influence on water temperature.

This report, together with analysis of cold water temperatures in Melbourne (Bors et al 2014) contributes to a much greater understanding of variability in water temperature in Melbourne and resultant influences on water-related energy. Specifically, this report builds towards a statistically rigorous methodology for characterising water temperature in a suburban water network, and evaluating changes through time.

References

AS/NZS 2008. AS/NZS 4234:2008 : Heated water systems - Calculation of energy consumption. Sydney, Wellington: SAI Global Limited.

BINKS, A., KENWAY, S. J., LANT, P. & PAMMINGER, F. 2014. Detailed characterisation of water-related energy use in households. *In:* AUSTRALIAN WATER ASSOCIATION (ed.) *Ozwater 2014.* Brisbane: Australian Water Association,.

BORS, J. & KENWAY, S. 2014. Water Temperature in Melbourne and Implications for Household Energy Use. Melbourne: Smart Water Fund.

KENWAY, S. J., BINKS, A., BORS, J., PAMMINGER, F., LANT, P., HEAD, B., TAIMRE, T., GRACE, A., FAWCETT, J., JOHNSON, S., YEUNG, J., SCHEIDEGGER, R. & BADER, H. P. 2014. Understanding and managing water-related energy use in Australian Households. *Water,* April**,** 184-188.

KENWAY, S. J., LANT, P. & PRIESTLEY, A. 2011. Quantifying the links between water and energy in cities. *Journal of Water and Climate Change,* 2**,** 247-259.

KENWAY, S. J., SCHEIDEGGER, R., LARSEN, T. A., LANT, P. & BADER, H. P. 2013. Water-related energy in households: a model designed to understand the current state and simulate possible measures. *Energy and Buildings,* 58**,** 378-389.

Appendix A

For each of the two time periods, we calculate and plot (in Figures 6 and 7), the proportion of each measurement location (households, relative to the total sample for this time period) against their corresponding mean temperature values. The later time period clearly has a higher rate (than the earlier period) of sampling and recording temperatures at locations which have relatively high mean temperature values.

A similar analysis was performed for regions, each containing a cluster of suburbs. Figures 8 and 9 illustrate that particular regions with higher mean temperatures are more likely to be selected in the later time period, as is seen from Figures 6 and 7.

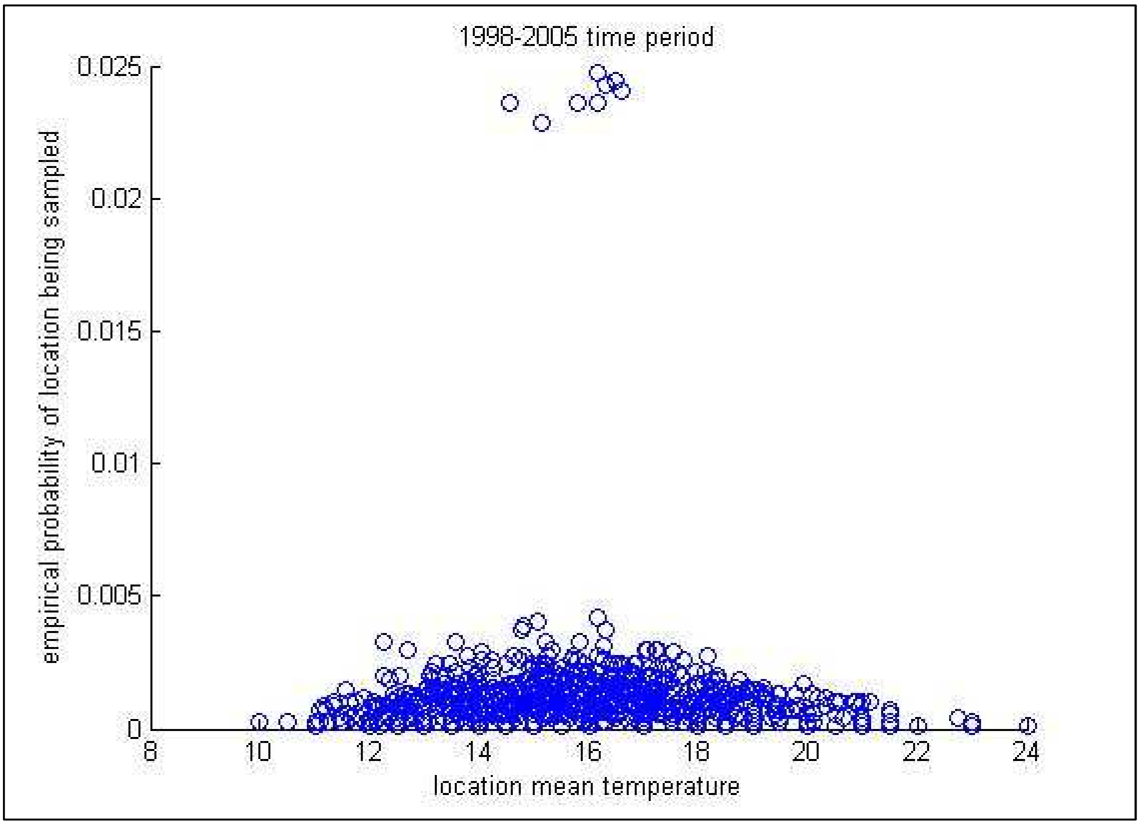

Figure A - 1: Scatter plot of the probability of sampling specific locations (households) and their mean values.

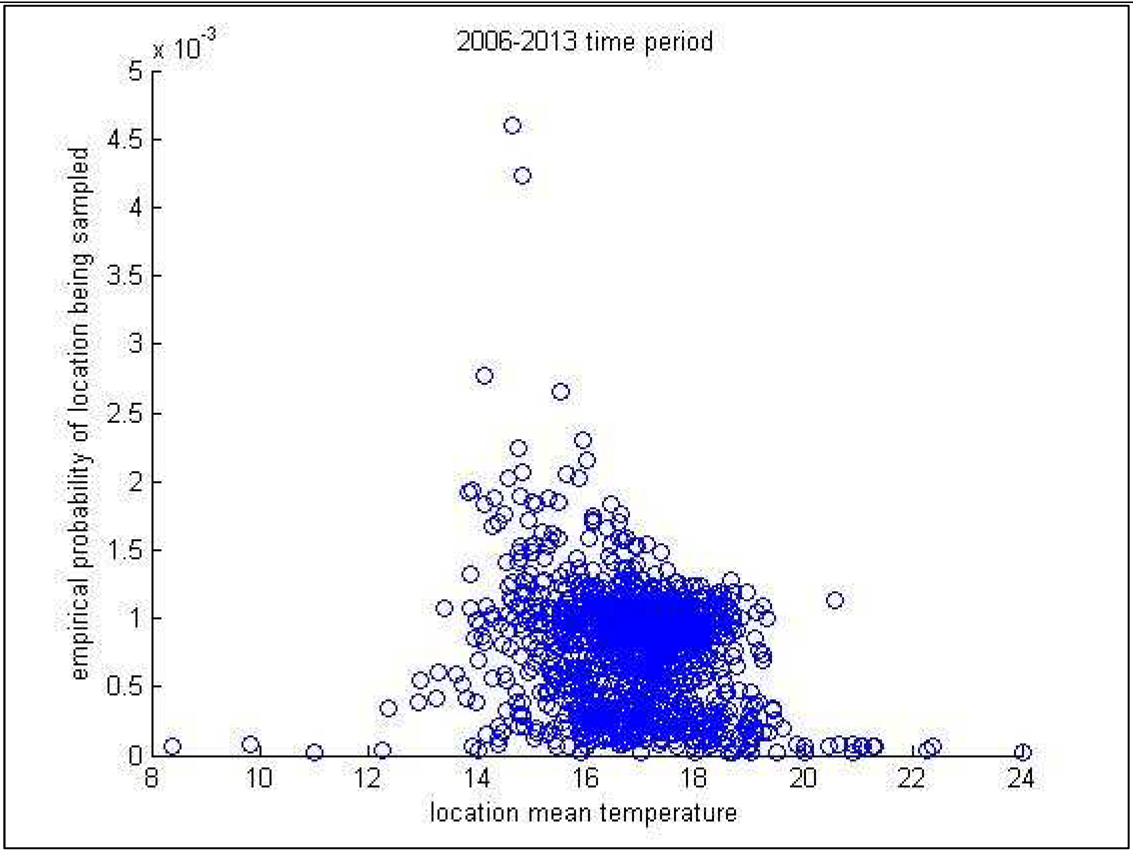

Figure A - 2: Scatter plot of the probability of sampling specific locations (households) and their mean values.

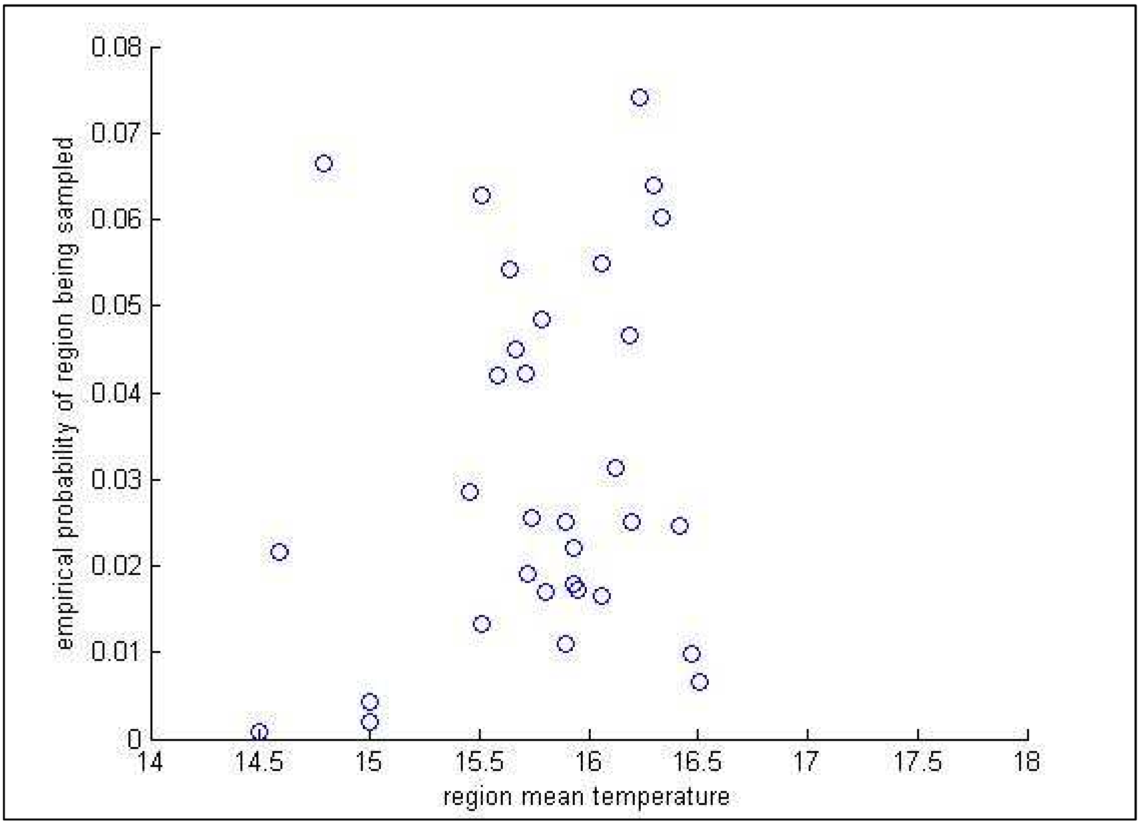

Figure A - 3: Scatter plot of the probability of sampling specific regions (clusters of suburbs) and their mean values.

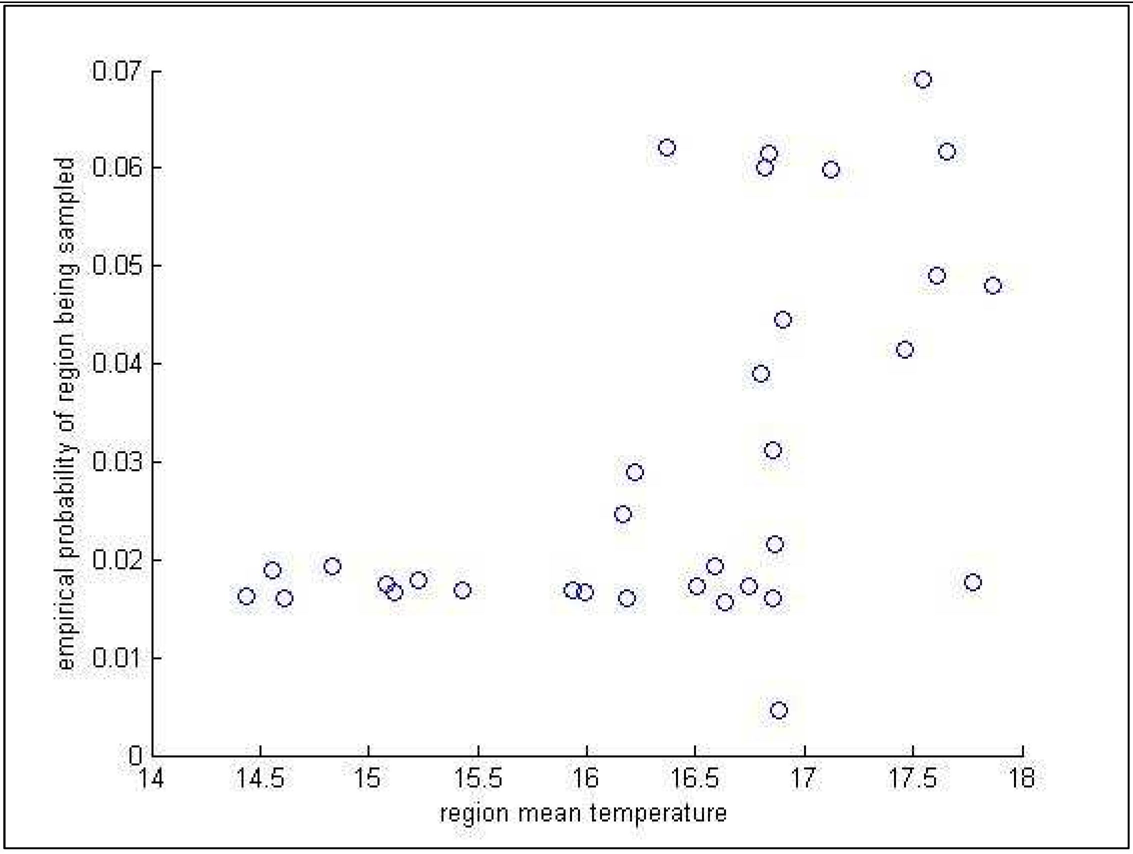

Figure A - 4: Scatter plot of the probability of sampling specific regions (clusters of suburbs) and their mean values.