The Creation of Weather Data Files to Fulfil the Simulation Needs of New Zealand

Mike Donn*, Robert Amor**

** School of Architecture, Victoria University of Wellington*[1]

*** Computer Science Department, University of Auckland*[2]

This paper discusses the creation of standard weather data files covering the whole of New Zealand. As part of this process we describe the development of a method for selecting extreme weeks of data. This provides the simulationist with a set of weather patterns which test a building's performance under the one-percentile most demanding conditions to be found at a particular location.

1. Introduction

Up until 1990, only four files of hourly weather data suitable for use in thermal simulation of building performance existed for four locations in New Zealand. Created in 1977, they were error checked from four year periods of actual weather records available at the time. Much of the research effort expended in their creation was devoted to ensuring the accuracy of the measured solar radiation data. In 1990 a comprehensive set of weather data files (including TRY format data) was created from raw weather data for 22 locations in New Zealand and the South Pacific. This research concentrated on exploiting the potential for reliable statistical analysis in the multi-year data.

This paper describes: the process used to create these weather data files; problems with the development of such files using published methods; alternative methods of simulating the missing data found in raw weather data files; and the development of one percentile extreme weeks of weather data to test bounds in the operation of buildings simulated with weather data.

Determining the most "average" year of weather data for a site is a relatively simple procedure (ASHRAE Handbook 1981) given many years of weather data to examine. Problems emerge after the data have been selected. Often the selected data will have periods of missing information for several of the measured attributes. The occur at times when the measurements were not made due to equipment malfunctions, or when the human observer was indisposed. The simulationist may also find that the data collected in each hourly observation may not include some parameters of the weather data which are necessary for simulation work, such as radiation data. This is especially the case for the earlier years of weather data collection.

There are many methods of determining missing weather data for short periods of time, and standard methods of deriving missing parameters from existing data. The problems arise when these methods use information or parameters which were not collected for the data sets at hand, as was the case with this work. For example, to use the available procedures for calculating radiation data for New Zealand weather data it was necessary to find clearness numbers for all the sites. With a definition of clearness numbers (Threlkeld and Jordan 1958) and the help of the New Zealand Meteorological Service (NZMS), values of clearness numbers were selected to suit the known weather patterns in various New Zealand locations.

Where selected data files contained patches of missing data it was necessary to simulate the missing data before their use in a simulation tool, which expects all data to exist. As various parameters of the weather data are inter-related it was possible to calculate values for some weather

1: PO Box 600, Wellington, New Zealand;
Ph: +64 4 495 5031; Fax: +64 4 495 5233
e-mail: michael.donn@vuw.ac.nz
2: Private Bag 92019, Auckland, New Zealand; Ph +64 9 ; Fax: +64 9
e-mail: trebor@cs.auckland.ac.nz

data from other parameters (ASHRAE Handbook 1981). Where this was not possible it was necessary to simulate a probable value. Various methods of generating data using sinusoidal curves were examined (Leslie and Trethowen 1977), but the resulting smooth and regular curves were discouraging and after discussions with the NZMS a procedure was developed based on weather profiles for various months, calculated on the total amount of data held for each site (Dr S Reid personal communication 1990).

As the planned use of this data was in design, and given the amount of data held for each site, a procedure was developed to provide data that could be used to examine a buildings' performance in extreme weather conditions. The procedure determined single weeks of data which are the one-percentile extremes for a given set of parameters (van der Werff et al. 1990). This procedure allowed the definition of the one-percentile extremes for the following weather patterns: hot and windy; hot and calm; cool and windy; cool calm and sunny; cool calm and cloudy. For the sites with many years of collected weather data it was also possible to determine warm and cool years of data using maximum summer and minimum winter temperatures in association with the cooling or heating degree-day values for each year.

The outcome of this project is a base set of TRY weather data for 15 New Zealand and South Pacific locations and a collection of extreme weeks of weather data for the full 22 sites of weather data held at the School of Architecture. This data is a valuable resource for anyone wanting to perform simulations of buildings in any of the major climate areas of New Zealand and the South Pacific.

2. Analysis of Weather Files

The original weather data files provided for New Zealand locations have various short-comings which make them less than useable for simulation purposes. They have missing periods of data, various attributes of the weather not collected at various sites and errors in the files which have to be corrected before use with a simulation program. The following sections describe the problems that were encountered and the solutions proposed.

2.1 Missing Data

Most of the weather files have periods ranging from an hour through to several days where there is missing data for one or more of the weather data attributes recorded each hour. This occurs where there is a breakdown in the machinery performing the monitoring, or in the earlier years of collected data where the human observer was indisposed.

For most sites the bare minimum range of weather variables were recorded. In these cases there was often no possibility of deriving a missing value through dependencies on other values. For some values required by simulation programs, it was possible to derive a new value from existing data e.g dew-point temperatures. However, this still left large amounts of missing data in the weather files.

Initial attempts at calculating values for missing data in the weather files were through the generation of sinusoidal patterns that followed the daily variations in temperature fairly closely. These patterns were scaled to match the start and end of the missing data, and worked well. The problem with this sort of fit, was that it was very smooth, and the sinusoidal curve did not approximate the temperature swings through the day. Moreover, the introduction of these curves had a significant effect on the average values for each month.

After consultation with Steve Reid from the Meteorological office (Reid 1990), an algorithm was developed to calculate missing data. For each month of every year for which data was available a temperature profile for that month was created which consisted of the hourly average temperatures for that month. This profile was scaled to the first valid value before the missing data and an offset was added to each successive hour so the profile would match the first valid value after the stretch of missing data. This technique produced very realistic weather curves for periods with missing data.

2.2 Spikes

From analysis of plots of the weather data, it was apparent that there were some erroneous data in the raw weather data files, namely spikes. Analysing the attributes of information collected it was apparent that it would only be possible to detect spikes in temperature data, which is about the only attribute which is not subject to large fast swings in its values.

A spike was defined to be a period of three hours where the temperature would rise and then fall three or more degrees, or vice versa. To fix this error the spike value would be changed to the average of the two points before and after the spike. This removed most of the major spikes.

2.3 Unknown Data

There are several instances where the weather files required by simulation programs require values for attributes not recorded by the New Zealand meteorological service.

In common with many other countries, the greatest problem of unknown data in New Zealand was radiation data. This information was not collected in New Zealand until a few years ago, and even now is not recorded for every site in New Zealand. This lack of radiation data was weighed against the large number of years of data available in the data sets. It was decided that the potential benefits from statistical analysis of the data were greater than the loss of precision arising from calculated radiation data. As the data had mostly been collected for hydro-electric planning purposes, it contained a surfeit of information on cloud cover. Using the weather processor in the thermal simulation package DOE2, radiation data values were generated from the available values. This process required input of the clearness number for each site.

After consulting the literature to find a definition of clearness numbers (refs) it was found that this measurement was devised and measured mainly in the USA. It is based on various measurements of the atmosphere which are not widely available in New Zealand, such as the depth of precipitable water in the air. From estimates of the depth of precipitable water in the air and comparing the estimates to USA data it was concluded that New Zealand clearness numbers were most likely to follow the numbers found along the west coast of the USA. The only exceptions to this would be in Christchurch where the winter smog would give lower clearness numbers in winter, and sites at high altitudes, like Alexandra or Waiouru which would be clearer than locations along the west coast of the USA. This yielded a list of monthly clearness numbers for sites in New Zealand (see Table 1).

Site	NZ	Chch	High
Jan	1.03	1.02	1.08
Feb	1.02	0.98	1.07
Mar	1.00	0.95	1.05
Apr	0.98	0.92	1.03
May	0.97	0.88	1.02
Jun	0.95	0.85	1.00
Jul	0.97	0.88	1.02
Aug	0.98	0.92	1.03
Sep	1.00	0.95	1.05
Oct	1.02	0.98	1.07
Nov	1.03	1.02	1.08
Dec	1.05	1.05	1.10

Table 1. Clearness numbers for New Zealand locations.

Another attribute which was often not recorded as required was ground temperatures. In some cases ground temperatures were not recorded for some of the sites. In this, case the closest location that recorded ground temperatures was selected. For some sites ground temperatures were only collected at 0.3 metres depth, and not the 1 metre depth we required. In these cases the 0.3 metre values were used. For some sites ground temperature records were started at a later date than some of the years data selected. In these cases, later year's ground temperatures were used for the earlier years data.

2.4 Format Translations

In many instances the value recorded for an attribute was in a different form from that required by the simulation tool weather processor. For instance, unique station numbers are required for the TRY format weather files, and the station numbers for New Zealand weather station range between 0 and 1000 which is below the acceptable range for TRY station numbers. To solve this simply, 10,000 was added to all New Zealand station numbers to obtain a number for use in the TRY format.

A more serious mismatch occurs between descriptions of the weather type and cloud type at a given observation time. For these instances a mapping was created to map between the codes required by one representation of a weather file and the codes required for another representation. Table 2 gives an example of two different representations for cloud types.

Value	NZMS	TRY
0	Cirrus	Clear
1	Cirrocumulus	Fog
2	Cirrostratus	Stratus
3	Altocumulus	Strato- cumulus
4	Altostratus	Cumulus
5	Nimbostratus	Cumulonimbus
6	Stratocumulus	Altostratus
7	Stratus	Altocumulus
8	Cumulus	Cirrus
9	Cumulonimbus	Cirrostratus

Table 2. Representation of cloud types.

2.5 Few years to analyse

Some of the weather data sites available have spans of only a few years. For all cases where there is less than 20 years data there is not enough data to calculate average, hot and cold years. The sites affected by this problem were Campbell Island, Chatham Island, Raoul Island, Rarotonga, Nasouri, Alexandra and Hamilton.

3. Extreme Weeks

The development of extreme weeks of data was stimulated by the need to analyse the response of a building to weather patterns which are known to be at the outside level of what can reasonably be expected for a building in a certain location. Analysis of a building using a TRY data file (or other standard data files) can only produce average results for any given period, and performing an analysis of a building using a particularly hot or cold week of data from one of these files is not statistically valid.

Five types of weather patterns were sought from the weather data files. These correspond to situations which may have the greatest effect on a building as opposed to a standard year's data. The five weather patterns sought were:

- hot and windy
- hot and calm
- cool and windy
- cool, calm and sunny
- cool, calm and no sun

For this project the one-percentile extreme weeks of each of these weather types were required. As there are two or more variables involved in each of these weather patterns, and they are not directly additive (eg temperature in degrees C and wind speed in m/s) each of the variables had to be normalised before the selection process. The process of normalising a variable is performed by dividing the raw variable values by the mean value of that variable over the period being analysed. These normalised variable values are unitless and can be added to other normalised variables and compared. So for each of the five weather types the appropriate normalised variables are examined and then the weekly average of these combinations is calculated and then ranked in ascending or descending order, depending upon the weather pattern required.

From these ranked lists of averaged weekly normalised variables the appropriate percentile week is selected. For example, the one-percentile week for a weather data set of 25 years, which has 1300 weeks of data, is taken as the thirteenth item in the ranked list. When looking at the thirteenth item in the ranked list a quick hand check is made that the week chosen is extreme in all the variables which make up its sum, and not just very extreme in one of the variables. If the selection is composed of one very extreme variable then the next week is chosen from the list. For example, Table 3 shows the selection of the one-percentile extreme week in the listing of the cool and windy weather pattern at Kelburn, which has an average temperature of 12.5^{oC} and a monthly average wind of 345km/day.

Rank	Norm	Temp	Wind
1	-0.13	9.2	21.5
2	-0.12	7.0	19.3
3	-0.06	5.6	17.3
4	0.02	5.8	16.5
5	0.11	6.9	16.6
6	0.13	8.4	17.8
7	0.15	7.2	16.5
8	0.15	8.2	17.3
9	0.16	8.3	17.3
10	0.16	10.1	18.9
11	0.17	6.9	15.8
12	0.20	9.1	17.6
13	0.21	8.0	16.5

Table 3. Cool and windy week selection.

This shows quite clearly the balancing that the normalisation gives between the low temperatures and high wind speeds, and while neither variable is very extreme they are both significantly distant from the average for a whole week.

4. Hot and Cold Years

In the same way that an extreme week of weather data is useful to test a building at the bounds of expected weather at a site, hot and cold years of weather data are useful to simulate the differences in energy usage and loads on a building in comparison to a standard year.

To select weather data for a hot year the maximum temperatures for the two hottest months in summer (January and February) were found for all years. For each of these months a ranked list of the maximum monthly temperature was created. To select a year, we first discard the hottest year from each month (as extremes) and then seek the next hottest years which appear in both month's lists. The most extreme of these matched years is selected as the hot year for this site. As a final check that this selected year is indeed a good year it must lie in the top five cooling degree day years of either January or February. For example, the selection of a hot year for Kelburn is shown in Tables 4 and 5. From Table 4, 1975 and 1982 are eliminated as being the top two extreme years, and the next closest match for the two months is in 1971, which is ranked third in January and second in February. Looking at the degree-day cooling chart in Table 5 it can be seen that 1971 is counted in the top 5 degree day cooling years for both January and February. So 1971 is selected as the hot year for Kelburn.

Jan	Temp	Feb	Temp
75	30	82	28
70	28	71	27
71	27	74	27
65	26	79	26
62	26	66	26
81	26	70	26
78	26	78	26

Table 4. Maximum temperatures in summer

The same procedure is followed for the selection of a cold year, except using the minimum temperatures for the two coldest months, July and August, and checking the selected year to see if it is in the top five heating degree day years for July and August.

Jan	DD Cool	Feb	DD Cool
75	121	71	36
70	39	79	23
78	25	82	16
65	18	66	14
71	18	74	13

Table 5. Degree Day Cooling for summer

As this selection process works by comparing years of data it is only undertaken where there is a large number of years of weather data available for a particular site. For this analysis this equated to 25 or more years of weather data.

5. Conclusion

This paper described the processes applied to a large collection of raw weather data files to create weather files for simulation systems. As well as following the necessary process for the selection of a standard years data from each of the sites weather data we developed procedures for identifying extreme weather patters from the weather files. Five weather patterns which can have a great affect on a buildings performance were identified, and through the use of normalised and averaged weather data variables a ranking of the most extreme weeks of the particular weather patterns was created. For our purposes we used one-percentile weeks of data from these lists of extreme weeks, though other percentiles could be chosen for other situations. We also defined a simple method to identify a year with a hot summer or a cold winter from a large number of years of weather data.

References

ASHRAE Handbook 1981 "Fundamentals", ASHRAE Inc.

Threlkeld, K.L. and Jordan, R.C. 1958 "Direct Solar Radiation Available on Clear Days", ASHRAE Transactions, Vol. 64, 45-68.

Leslie, S.F. and Trethowen, H.A. 1977 "A Computer File of New Zealand Climate", Technical Report 21, Building Research Association of New Zealand, Porirua, New Zealand.

van der Werff, I., Amor, R. and Donn, M. 1990 "Standard Data Files for computer thermal simulation of solar low energy non-residential buildings", Energy Research Group Report CRP 53, School of Architecture, Victoria University, Wellington, New Zealand.