Chapter 2

A new Monte Carlo method

2.1 Explanation of statistical concepts

2.1.1 Statistical Probability Density Functions and their application to the problem.

At this time a few statistical concepts should be explained. The standard way to present data about the probability of a certain variable having a given value is with a Probability Density Function, or PDF. A PDF is a function with a variable, such as temperature, on the horizontal and the expected probability of that variable for a specific differential interval on the vertical. The most well known of these is the Gaussian or Normal probability distribution.

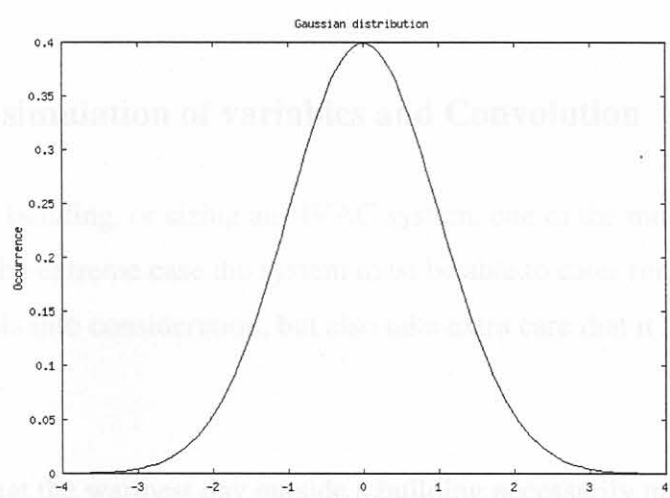

Figure 2.1.1-1

Figure 2.1.1-1 shows a Gaussian distribution, with a mean at 0, and a standard deviation of 1. If we want to find out the probability that an event might occur between two values of the variable, we find the area under the curve between those two values of the variable.

The weather data we have is in the form of hourly values. To generate PDF's from such data the following procedure was followed.

Bins were created for each hour, starting at -5 °C for temperature, with a one degree interval up to 45°C. For radiation bins started at -0.05kW/m^2 with a 0.05kW/m^2 interval, up to 2.45kW/m^2. Should a reading then be greater than the lower value of the bin, or smaller or equal to the higher value, that bin is incremented by one. A radiation reading of 0kW/m^2 would therefore be placed in the bin -0.05 to 0kW/m^2. This will ensure that any non-zero reading of radiation, no matter how small, will not be taken as zero.

As is the standard practice, the PDF's are then normalized by dividing each value with the area under the curve for that hour. The result is then written out a file that contains not only a PDF for each of the twenty-four hours per day, but also the bin information. This eases later use of the file.

2.1.2 Separate simulation of variables and Convolution

When designing a building, or sizing an HVAC system, one of the most important considerations is the extreme case the system must be able to cater for. We therefore need not only to take this into consideration, but also take extra care that it is adequately described.

We cannot state that the warmest day outside a building necessarily produce the warmest inside climate. "Almost invariably the hottest days are not those with the greatest amounts of solar radiation."(Chartered Institution of Building Services Engineers, 1986) This brings us to a problem, what days should we select for the extreme case? Should we base it on the outside temperature or radiation? We cannot base it on the sol-air temperature at this stage, that requires information on the color of the building, and we do not have that information beforehand.

We also cannot simply use the extreme temperature together with the extreme radiation, since this is not a realistic or possible weather day. The only output variable we are interested in at the moment is the statistics of the inside temperature. If we can see the effect of each variable on its own, and then find the combined effect, this would be perfect. But the combined effect cannot be found by normal addition.

This brings us the process of convolution. To quote "If two quantities can assume values of x describable by frequency distributions $P_1(x)$ and $P_2(x)$, respectively, then their sum is described by a frequency distribution P(x) given by $P(x) = P_1(x) * P_2(x)$", where * indicates the standard convolution integral. (Bracewell, R.N., 1978) This is true if the two distributions are independent from one another.

The convolution integral is given by Numerical Recipes in C (Press, W.H. *et al*, 1992) for functions g and h:

$$g * h \equiv \int_{-\infty}^{+\infty} g(\tau)h(t-\tau)d\tau \qquad \text{Eq 2.1.2-1}$$

Since we will work with discrete distributions, we will use the discrete form, again from Numerical Recipes in C:

$$(r * s)_j \equiv \sum_{k=-M/2+1}^{M/2} s_{j-k} r_k \qquad \text{Eq 2.1.2-2}$$

By using the convolution integral we can combine the output statistics of both input variables. The problem is that the weather variables we are working with have cross-correlation's, as explained earlier. For the time being we will assume that they can be neglected, and later test the results to find the error made.

2.2 Development of a new Monte Carlo technique

2.2.1 Introduction to Monte Carlo techniques

The simplest way to compute any problem with random input will be to have simulations for all possible inputs. Computationally, this would be prohibitively expensive. The other option is to randomly select possible values for the input variables. If we assume that the output we get from this is a representative sample from the whole population, we can calculate the statistics of the output with much less effort. This is known as a Monte Carlo method.

A True Monte Carlo Method would entail full random selection, out of all possible values. It was quickly realized that by being selective about the selection of data, the amount of simulations could be reduced even further. Care must be taken, however, not to influence the outcome too much. At the end, the results will have to be tested to prove that the selection process did not bias the result.

While trying to find a Monte Carlo method we must always remember the 3D nature of our input. We have two input PDF's for each hour, one for outside temperature and one for outside radiation, or assembled for a day as in Figure 2.2.1-1.

What we are looking for is a method to map these distributions of the outside variables to the inside temperature distribution. For this we need to select a number of days of 24 hour outside data so that a deterministic program can be used to simulate them. From the output an approximation must then be found of the distribution of the output variable.

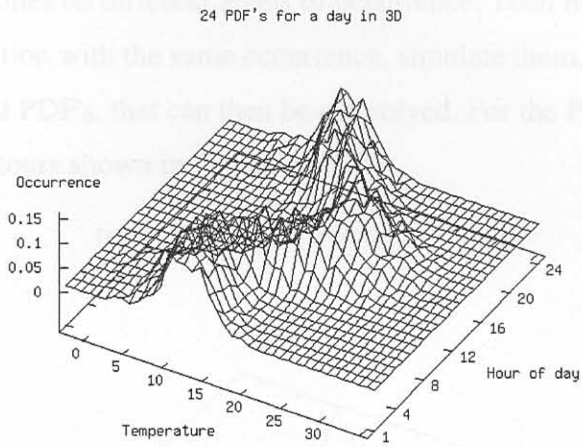

Figure 2.2.1-1

2.2.2 Method of Contour lines

The first possible method comes from data as supplied by CIBSE (CIBSE, 1986). They give banded weather data for different locations. This data is set up by taking the range of radiation data, dividing it by ten, and then giving the proportion of the month that falls within each of the 10 bands. The coincident max, mean and min temperatures are then given for each band. Data is also available where the banding was done with the temperature data, and coincident radiation is given.

Clearly this is a start, but fundamental problems arise if the data must be used for a statistical method. First, we will have to decide on a set, either banded according to radiation or to temperature. This then may not include that day that the building under consideration will be at its most extreme internal climate. Also the data is given on a daily value, with little possibility for accurate hourly predictions. This is because the data is primarily for the calculation of simplified monthly energy usage.

With the theory of convolution as discussed earlier this idea can be adjusted slightly: Use the PDF's as we have it for the external weather, considering radiation and temperature apart. Draw contour lines on different levels of occurrence. Then make up days of hourly temperatures or radiation with the same occurrence, simulate them, and assemble the results for the internal PDF's, that can then be convolved. For the PDF's of Figure 2.2.1-1 this will give the contours shown in Figure 2.2.2-1:

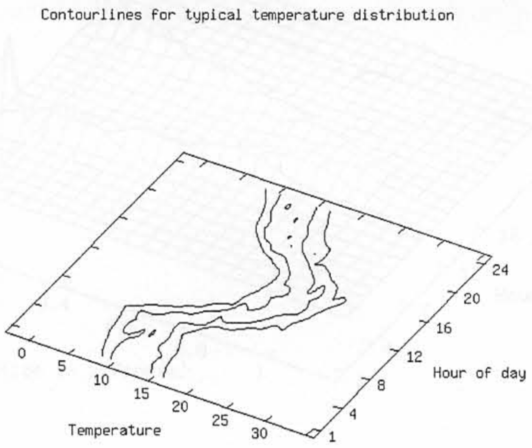

Figure 2.2.2-1

Problems arise here because of the fact that the distributions during the night, because of a smaller range, normally have a higher maximum occurrence. If a contour line is set on one of these values of maximum occurrence, there is no corresponding value during the daytime. If all contour lines are chosen to ensure they fall below this point, the maximum occurrence values of the night time is not accounted for.

This problem is most noticeable with radiation. There all the nighttime values are 0, giving an impulse function for a PDF. Even if only hours with radiation are considered, there are still major problems with the differences in range and maximum value of occurrence. Figure 2.2.1-1 and Figure 2.2.2-1 represent typical temperature data. Figure 2.2.2-2 shows typical radiation data, and figure 2.2.2-3 show the contours of such data.

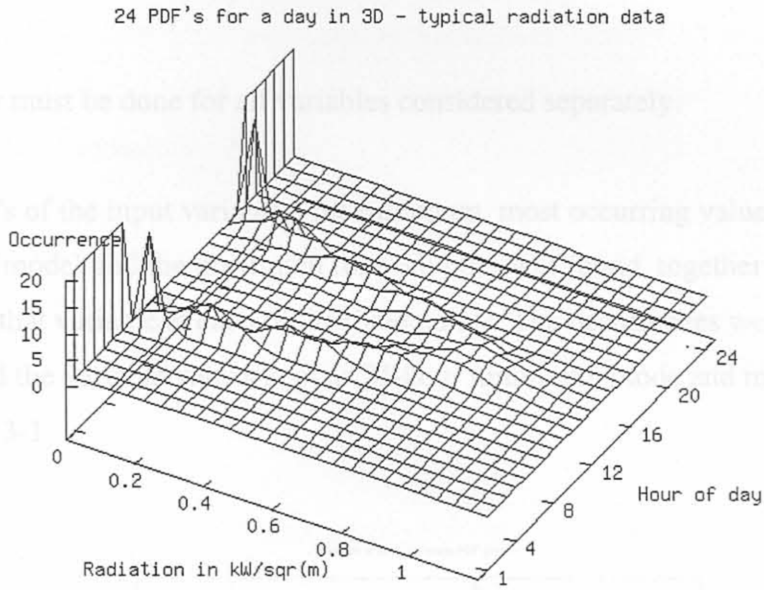

Figure 2.2.2-2

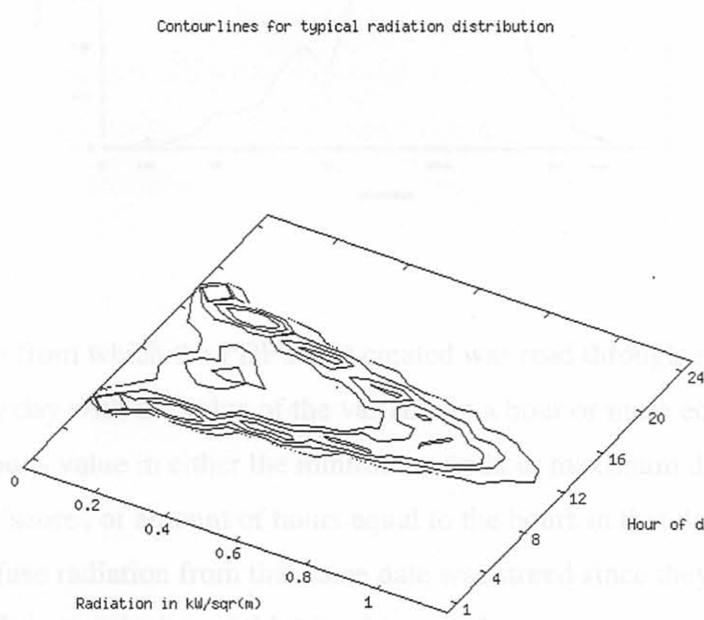

Figure 2.2.2-3

Clearly this will not work.

2.2.3 Method of minimum, mode and maximum

All that follow must be done for all variables considered separately.

From the PDF's of the input variables, the minimum, most occurring value of the variable (known as the mode) and the maximum for each hour was found, together with the occurrence of that variable at each of the three points. The occurrences were stored separately, and the variable values kept as 24-hour minimum, mode and maximum days. See figure 2.2.3-1

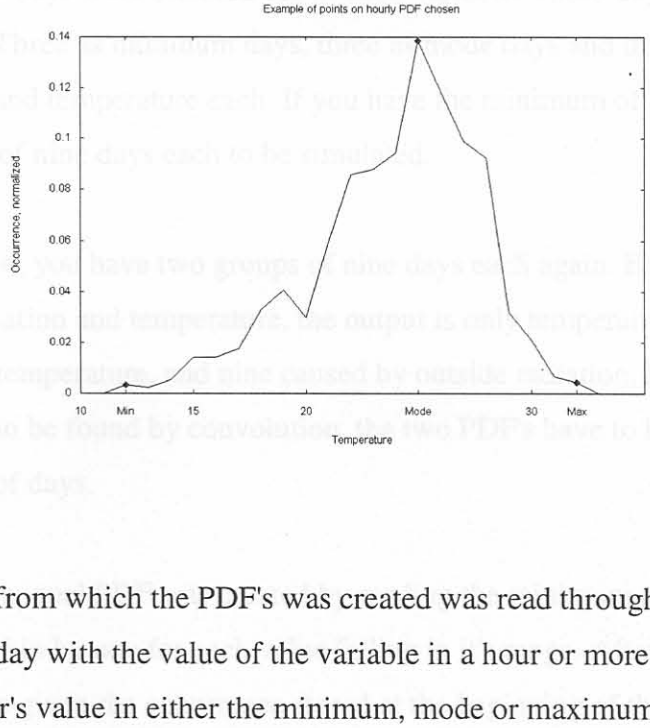

Figure 2.2.3-1

Now the raw data from which the PDF's was created was read through again. This time the data from any day with the value of the variable in a hour or more equal to the corresponding hour's value in either the minimum, mode or maximum day was stored, together with it's 'score', or amount of hours equal to the hours in that day. With global radiation, the diffuse radiation from that same date was stored since they are considered fully dependent (i.e. as a single variable) for this exercise.

Next the top three scores were found from the stored data, together with the amount of times they occurred. In other words, lets say that the top three scores found was 15, 13 and 10, then the amount of days those scores occurred will be counted. From this the best three days must be chosen. If the top score, in this case 15, occurred 3 times, or the top 2 scores together occurred 3 times, or the top 3 scores together occurred 3 times, then no random selection is necessary.

However, if the top score occurred more than three times, three days with that score was randomly chosen. If the top score occurred two times and the second one more than one time, one day of the second score was randomly chosen. The rest of the logic is self-evident.

In this manner nine days were selected out of each variable. These days are real, made up of measured data. Three as minimum days, three as mode days and three as maximum days for radiation and temperature each. If you have the minimum of two variables, this leaves two groups of nine days each to be simulated.

After the simulation, you have two groups of nine days each again. But where the input variables were radiation and temperature, the output is only temperature, nine days caused by outside temperature, and nine caused by outside radiation. Before the combined effect can be found by convolution, the two PDF's have to be reconstructed out of the two groups of days.

First a three dimensional PDF was created by reading the minimum days and incrementing each bin by one for each value falling in it's range. After that the value occurring most was given the occurrence stored at the beginning of the exercise. The same was done for the mode and maximum days.

The PDF was completed by calculating the values of the empty bins by assuming a straight line from the first value in the PDF to the highest occurrence, and another line down from the highest occurrence to the last value of the PDF. The shape of the PDF is thus a simple triangle.

Because the output variable values was given the same occurrence as the input value, it was in effect assumed that the effect of the process on a single variable would be to stretch the PDF by lengthening or shortening the x-axis, but not changing the shape in the y direction. The y values would only be influenced by the re-normalization necessary. This assumption can alternatively be stated that if the input variable is changed in a certain way, the way the output variable change would be constant, no matter where in the range of the input variable the change occurred.

At this stage you are left with two PDF's for the inside variable as result of the each of the input variables. After convolution the result is then again a single PDF, the combined effect.

References

BRACEWELL, R.N., 1978, *The Fourier transform and its applications*, McGraw-Hill Kogakusha LTD, Tokyo, 0-07-007013-X.

CHARTERED INSTITUTION OF BUILDING SERVICES ENGINEERS, 1986, *CIBSE GUIDE A*, A2, CIBSE, 0 900953 29 2.

PRESS, W.H., TEUKOLSKY, S.A., VETTERLING, W.T. AND FLANNERY, B.P, 1992, *Numerical Recipes in C*, Cambridge university press, Cambridge, 0-521-43108-5.