MEAT & LIVESTOCK AUSTRALIA

On farm

RISK ASSESSMENT ON THE OCCURRENCE OF EXCESSIVE HEAT LOAD FOR THE MAJOR FEEDLOT REGIONS OF AUSTRALIA: SOFTWARE REPORT

Project number FLOT.312
Final Report prepared for MLA by:

KATESTONE ENVIRONMENTAL
Address

SEPTEMBER 2002

Meat and Livestock Australia Ltd
Locked Bag 991
North Sydney NSW 2059

ISBN: 9781 7419 1 5716

Published by Meat & Livestock Australia Limited
January 2003

FLOT.312 – RISK ASSESSMENT ON THE OCCURRENCE OF EXCESSIVE HEAT LOAD FOR THE MAJOR FEEDLOT REGIONS OF AUSTRALIA: SOFTWARE REPORT

FROM KATESTONE ENVIRONMENTAL TO MLA

September 2002

Table of contents

List of Tables

List of Figures

1. INTRODUCTION

The objective of this consultancy was to determine the probability distribution of heat load for each major feedlot region of Australia and give a risk rating. This has been done by assessment of long-term records of meteorological parameters at the available monitoring stations to determine the frequency of high THI conditions.

The second objective of this consultancy was to develop a guideline and assessment tool to allow feedlot operators to assess the potential risk of excessive heat load conditions occurring at their specific location. It was envisaged that this would be done through a decision support system that interrogates the user for site-specific information and uses the historical database or particular feedlot database of meteorological information to determine the potential risk. At the present stage of the project the software is available that present the risk at existing long-term weather station sites.

This report describes the operation and theoretical background of the risk assessment software developed by Katestone Environmental for the Meat & Livestock association under contract Flot 312. The purpose of the software is to allow persons intending to establish a feedlot in a particular area to determine the frequency of high heat load conditions at the site that could potentially affect cattle health or operational efficiency. This allows assessment of the risk of particularly adverse conditions, based on a long-term meteorological database for representative locations near the region of interest.

A pilot study of the southeast Queensland region was undertaken in 2002. Software was developed that allowed users to select locations of interest on maps provided via a web site. The software currently presents data for locations near long-term weather station sites operated by the Bureau of Meteorology (BoM) (these are typically Automatic weather stations or AWS). The feasibility of calculations for intermediate sites has been established, but this feature has not been incorporated into the current software.

2. CHANGE OF SCOPE

The original scope and design of the risk assessment software was based on the availability of detailed meteorological information for the selected feedlot regions. These data were thought to be available as a similar heat risk assessment was undertaken for the dairy cow industry. Inquiries have shown that the databases are either not adequate for calculating the required detailed information on an hourly basis as a risk measure or are not freely available. The project requires at least some stations in each feedlot region to have long-term databases of regular measurements (e.g. 2-24 times per day) of the key parameters. This then resulted in the need for assessing alternative methods for the generation of a risk assessment software package.

The following report details the beta version of the software for a south east Queensland region which presents detailed information on heat load risk at sites with long term monitoring stations. Methods for interpretation between the known sites have been investigated and are presented in the attached appendices.

3. HEAT STRESS INDICES

Environmental indices have been developed, such as the temperature-humidity index (THI) to help determine the thermal loads on animals. The level of the THI can indicate to a feedlot operator the potential risk to the animals and the requirements for mitigatory actions. The THI can also be used in conjunction with a time factor to determine a THI-hour or the amount of time the THI is above (or below) a give critical value. Other indices that include wind and radiation variables are likely to provide a better definition of the environmentally related risks for feedlot selection and design.

The index used in this assessment to assess the risk of excessive heat load conditions is the THI. The equations adopted for this study are given by:

$$THI = 0.8\,T + RH\,(T-14.4) + 46.4 \qquad \text{(Eqn 1)}$$

and

THI = T + 0.36 DP +41.2 (Eqn 2)

where:

T = screen temperature in °Celsius at a hight of 1.2 m above ground level

RH = relative humidity expressed as a value between 0 and 1 measured at 1.2 m above ground level

DP = dew point temperature in °Celsius at a hight of 1.2 m above ground level

Relative humidity is a function of temperature and dewpoint, and both equations give very similar numerical values. Experience has shown the range of THI values that correlate well with observed cattle comfort (with typical response levels) are listed in the Comfort Index THI Range presented in Table 1. Certain cattle breeds are more/less susceptible to heat loads and acclimatisation is also a factor, so this classification may require modification in some situations.

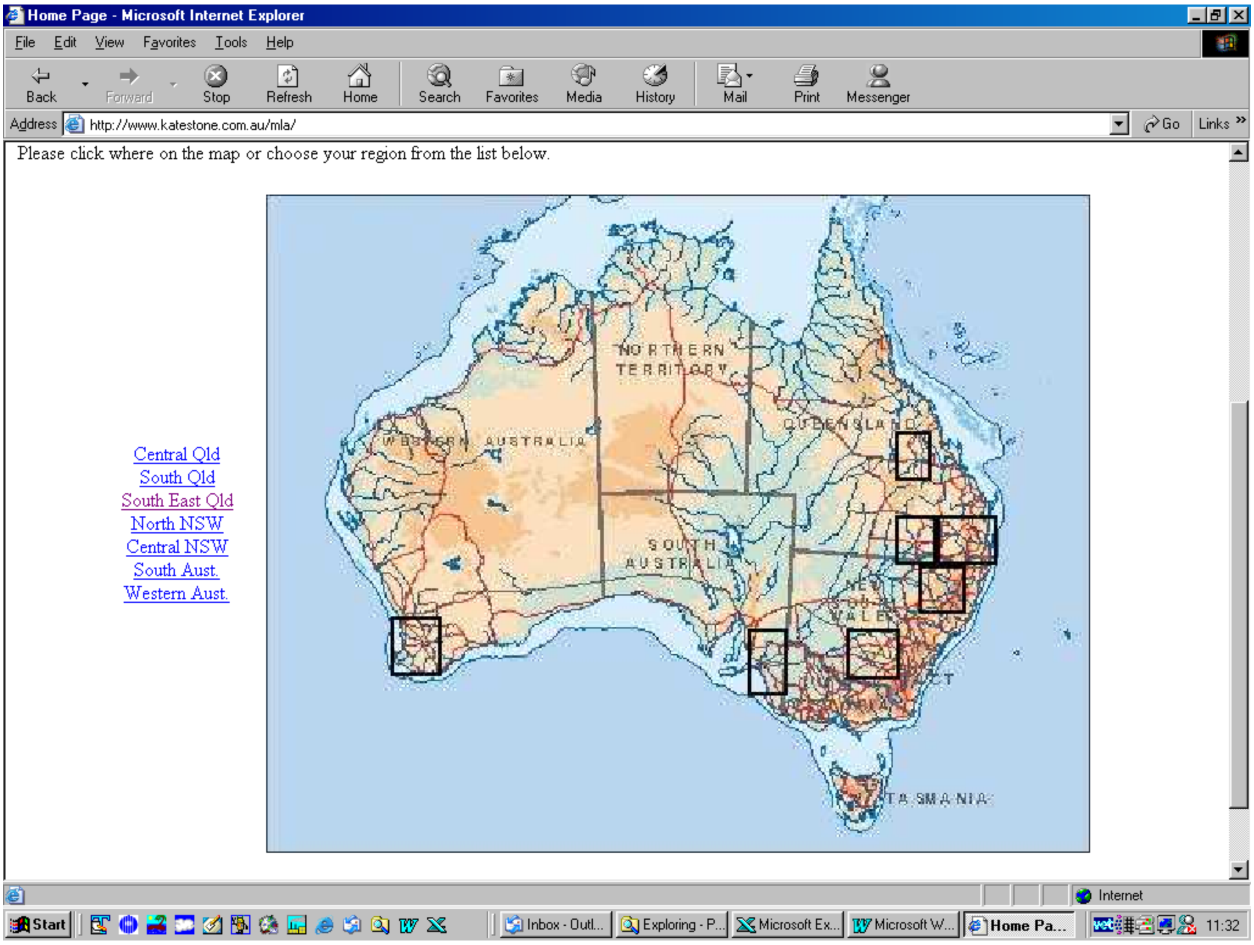
Table 1: Comfort Index THI Range

Range	**THI**
Alert	73 – 79
Danger	79 – 84
Emergency	84 – 90
Crisis	90 +

4. RISK SOFTWARE

Microsoft Windows™ compatible software was developed in the form of an Internet web site that can be accessed by any computer having an appropriate Internet connection and browser. The user is prompted to click on one of a list of regions or on a map to identify the approximate location of the site of interest (Figure 1).

Figure 1: Selection of region from the front screen of the web site



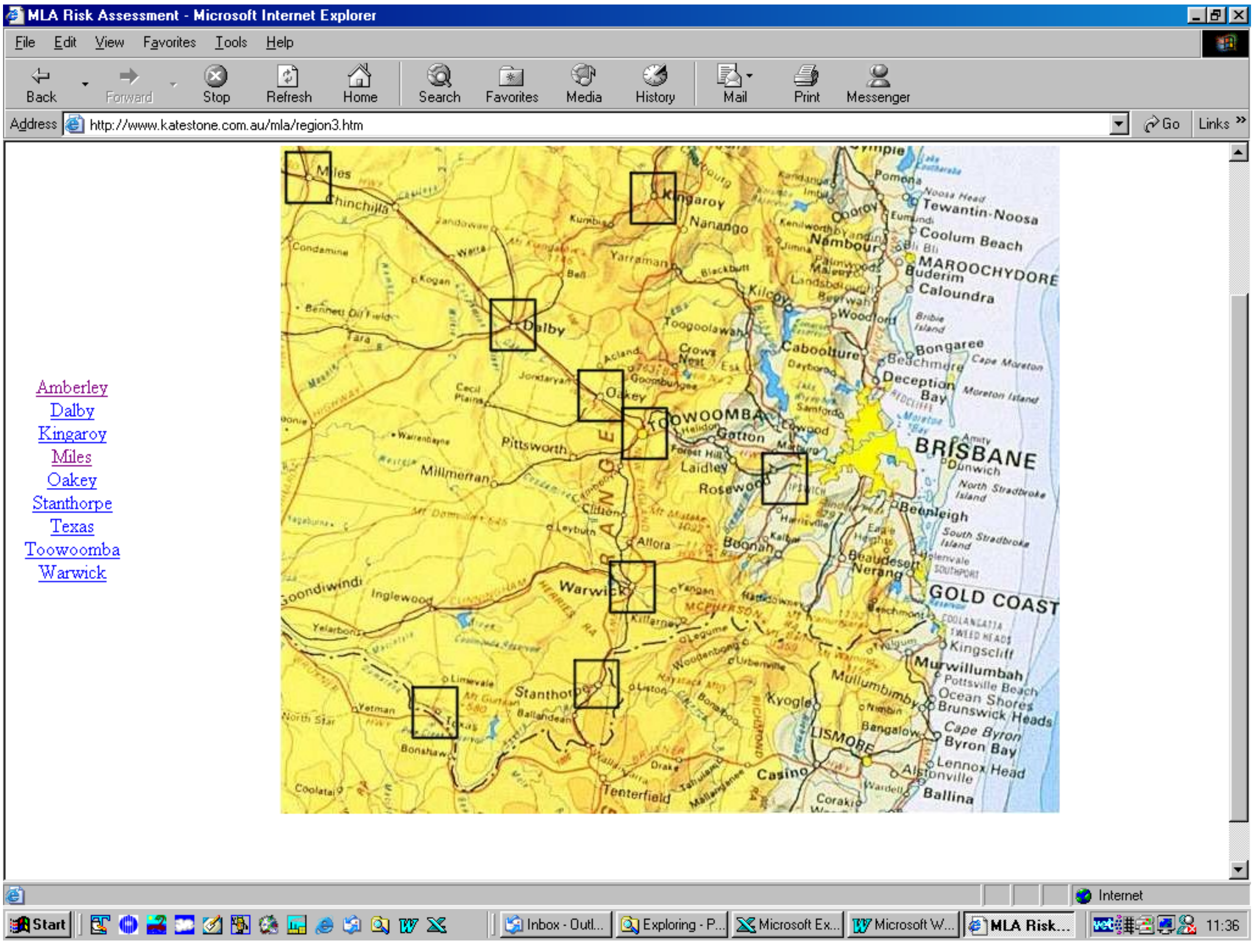
The software prompts for the location of the site of interest:

View the THI Graph?

Please click on the map where your station is or choose your nearest AWS site from the list below.

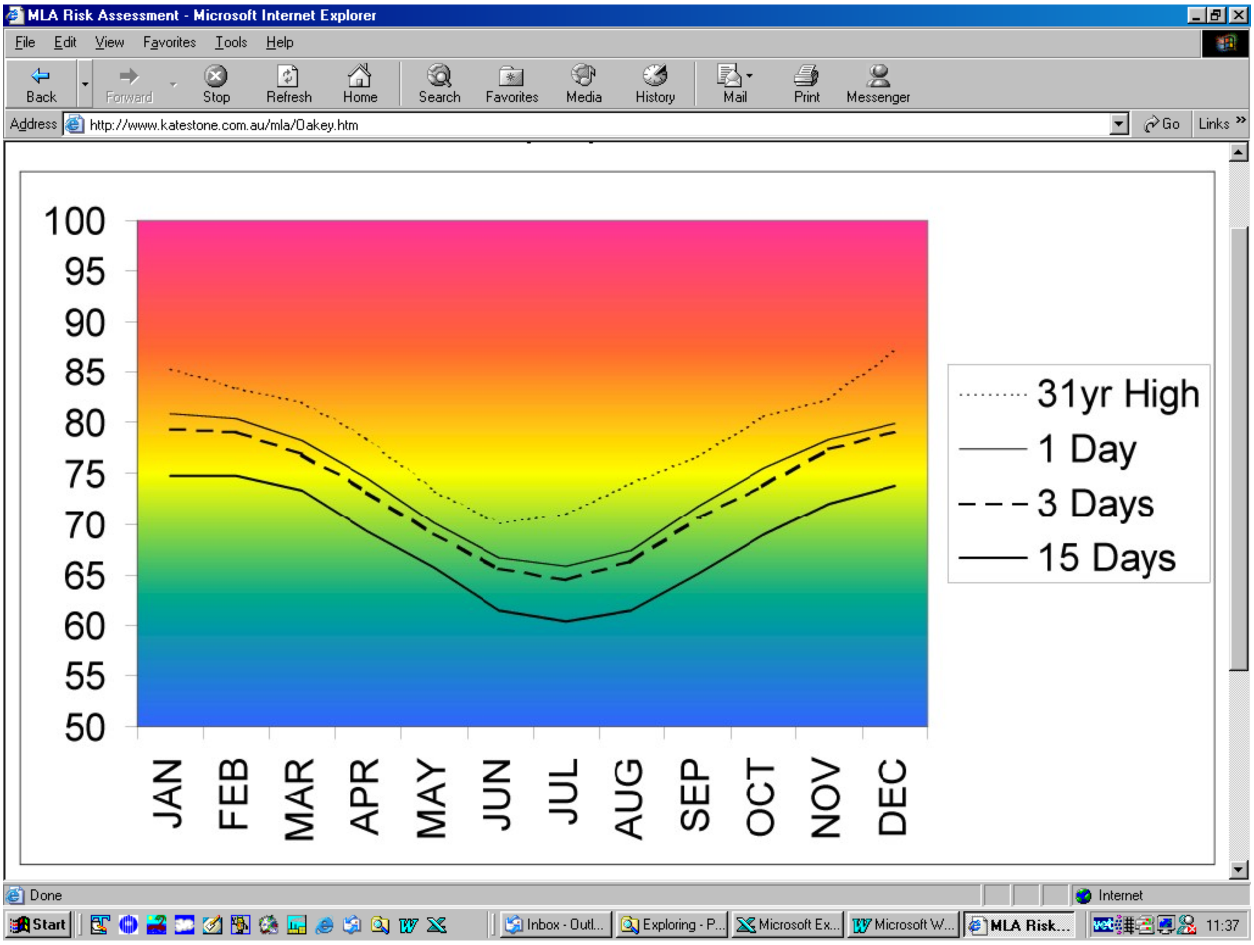
An image of the screen following the prompt is provided in Figure 2.

Figure 2: Site selection from the web site screen



An example of the output provided following selection of the Oakey site is provided in Figure 3.

Figure 3: Example 'Maximum THI' screen for Oakey



The above graph shows the monthly average frequency of the daily maximum THI calculated from Equation (2). Lines on this graph show the daily maximum THI likely to be exceeded for 1, 3 or 15 days per month. Thus, in February, for example, it is expected that 15 days will have a maximum THI that will exceed 75, whilst only 1 day is likely to exceed this value in April. The figure also indicates the maximum THI recorded for that month over the entire dataset (presented as a 31 year high for Oakey).

Further information is available from the software as shown in Tables 2-4.

Table 2 gives the expected number of days that will reach or exceed the THI level for any month. The number in brackets gives the maximum number of days that reached or exceeded the THI level for that month in the past 31 years. For example in February an average of 2.6 days per year are in the Danger zone with a maximum of 9 days per year recorded in this zone over the entire monitoring period.

Table 2: Web site output example for 'Expected THI Days' for Oakey, with maximum number give in brackets

Month	Alert	Danger	Emergency	Crisis
Jan	18.0 (29)	3.1 (11)	0.1 (2)	0
Feb	17.2 (28)	2.6 (9)	0	0
Mar	14.4 (30)	0.7 (7)	0	0
Apr	2.7 (11)	0	0	0
May	0.1 (3)	0	0	0
Jun	0	0	0	0
Jul	0	0	0	0
Aug	0 (1)	0	0	0
Sep	0.8 (3)	0	0	0
Oct	3.7 (11)	0.1 (1)	0	0
Nov	10.2 (22)	0.9 (4)	0	0
Dec	15.0 (27)	2.7 (12)	0 (1)	0

Table 3 shows the expected number of heatwaves, or days in a row that will reach or exceed the THI range. If more than one event is expected to occur per year, this number is expressed as *x per yr* otherwise it is expressed as *x in z yrs.* Thus, from the above table we can expect 1-2 two-day events per year with the THI greater than 79, and also 1 six-day event in every 31 years.

Table 3: Web site output example for 'Heatwaves' for Oakey

Days in a Row	2	3	4	5	6	7	8+
Danger	1-2 per yr	1 in 2 yrs	1 in 5 yrs	2 in 15 yrs	1 in 31 yrs	1 in 31 yrs	0
Emergency	1 in 31 yrs	0	0	0	0	0	0
Crisis	0	0	0	0	0	0	0

Table 4 gives the percentage probability of a certain THI range succeeding another. The THI level headings above columns 2-6 represent today's THI while the probability (%) of tomorrow's THI matching each of the levels in column 1 is represented by the numerical value in each row. For example, if a 'Danger' day occurred today, the probability that it will be followed by a 'Normal' day tomorrow is 4%, by an 'Alert' day 50%, by a 'Danger' day 44% and an 'Emergency' day 1%.

Table 4: Web site output example for 'Successive Types' for Oakey

Tomorrow	Today				
	Normal	**Alert**	**Danger**	**Emergency**	**Crisis**
Normal	92	25	4	0	0
Alert	8	68	50	0	0
Danger	0.1	8	44	100	0
Emergency	0	0.1	1	0	0
Crisis	0	0	0	0	0

5. WEATHER DATA INTERPOLATION AND GENERATION

A literature review was conducted of methodologies for the spatial and temporal interpolation of sparse data sets (such as that collected for the Darling Downs). A philosophy has been developed to assist in producing probability distributions of heat stress indices throughout the region, based on existing data, terrain and land use categories.

Various interpolation schemes have been tested to determine the optimum approach to estimating THI values at locations for which historical meteorological readings suitable for calculation of THI were not available.

5.1 Literature review

With the development of geographic information systems handling dynamic environmental data, there has been increased interest in producing high resolution maps of environmental measurements from limited measurements. Geostatistical approaches have long attempted spatial interpolation via methods such as:

- Optimal interpolation techniques such as kriging (i.e. best linear unbiased prediction) see Cressie (1996).
- Multivariate modelling.
- Empirical orthogonal functions in which a meteorological field is represented as a linear combination of component patterns.
- Wavelet and fractal characteristic analysis.
- Semivariogram models.
- Neural network processes.

Environmental processes often have strong time dependence, resulting in the need to investigate stochastic multivariate space-times models using techniques such as:

- Time-series modelling;
- Kalman filtering;

- Monte Carlo simulation techniques;

- Classification models with transition state approaches.

There are several examples of successful interpolation of temperature and rainfall data. Brazilian researchers have produced good maps of countrywide temperature from 140 observation points by representing the hourly temperature as a function of elevation, land cover and distance from the coast. British mathematicians have produced wind energy maps from detailed information from a limited number of observing stations by modelling long-memory correlations. Australian researchers (Hutchinson et al, 1995) have used stochastic space-time models to produce detailed maps of five-day rainfall and other parameters.

Such models are difficult to apply to evaluating the spatial variability of extreme values of THI indices if only 7-10 observation points of varying data integrity are available. Extreme temperatures are likely to be very dependent on elevation and distance to the nearest coastline (for penetration of cooling seabreeze flows). Humidity is likely to be strongly influenced by land use and closeness to water bodies. Heat stress is alleviated by moderate windspeeds – winds are very strongly influenced by local topography and vegetation type.

Hence, for example, producing hourly estimates of THI over a 100 x 100 km region at a resolution of 1 km (i.e. 10,000 points) from observations at 10 locations is fraught with potential inaccuracies unless ways are found to estimate time-varying correlations between values of the key meteorological parameters between a given grid point and the nearest observational points.

Synoptic cluster analysis may be useful to reduce the dimensionality of the problem. Correlation structures on days with similar synoptic weather patterns might be expected to have a similar diurnal variation. A cluster anaysis was conducted for a southeast Queensland region and results are presented in Appendix A.

Various simple methods that may be robust include:

- Distance-weighted (e.g. inverse square) interpolation schemes for elevation-corrected temperatures, THI etc.

- Least squares regression models for dependence on elevation, inland distance and surface roughness.

- Simplified orthogonal function analysis.

The performance of several methods for predicting THI at any location in Region 3 was investigated. The methods included Kriging, Inverse Weighted Distance (first and second order) and Laplacian Interpolation. The reader is referred to Appendix B for a description of these techniques. It was found that there was little difference between the various interpolation approaches. Further details of the statistical and interpolation methods are summarised in the following section.

5.2 Testing of Interpolation methods

It is desirable for future versions of the software to be able to provide data for sites where no local AWS data are available. The possibility of providing useful frequencies of THI for intermediate points was tested by calculating THI values for an intermediate location using various interpolation schemes.

The region on which the interpolations were performed was Region 3, which covers the southeast corner of Queensland. The region covers an approximately square area measuring roughly 250 km x 250 km with Miles on the northwestern corner, Amberley close to the mid-eastern side and Texas on the mid-southern boundary. The area includes about a dozen meteorological station sites, nine of which were used in this investigation. These included Toowoomba, Amberley, Oakey, Kingaroy, Warwick, Stanthorpe, Miles, Dalby and Texas. The data file containing the meteorological parameters spanned the period starting in 1992 and ending in 2002. The details of each site are shown in Table 5.

Table 5: AWS sites used for interpolation investigations

Site	Site code	Latitude	Longitude	East	North	Altitude	Distance from coast
1	TOO	-27.54	151.92	392868	6953010	675	120
2	AMB	-27.63	152.71	471527	6943986	27	55
3	OAK	-27.4	151.74	375427	6968798	406	140
4	KNG	-26.55	151.85	385446	7063056	442	130
5	WAR	-28.21	152.1	411679	6879379	475	140
6	STA	-28.62	151.95	397353	6833837	872	160
7	MLS	-26.66	150.18	219303	7048287	305	300
8	DAL	-27.17	151.27	328295	6994161	345	180
9	TXS	-28.85	151.17	321481	6807432	284	275

The performance was tested by the "leave one out" method. Here the THI at one of the monitoring sites was predicted but the data from this site were not used in any calculation. The predicted THI values were then compared with observed values by various methods. The results of the success of the methods investigated in detailed in the following section.

5.2.1 Results

All of the above methods were originally designed for predicting parameter values on a two-dimensional plane (i.e. no height dependence). For this exercise, these were modified to take height into account the change in temperature with height.

The THI is strongly dependent on temperature and to a lesser degree on relative humidity, consequently, the behaviour of these was investigated. It was found that the temperature scatter plots for any two sites showed behaviour similar to the THI plots, a result which was not unexpected. The dew point scatter plot showed very large scatter, and the mixing ratio, which was determined from the dew point, also mimicked the dew point plots.

Figures 4 and 5 show the comparison of the predicted THI at Dalby using the second order inverse weighted distance interpolation scheme and actual THI values. The scatter plot shows general agreement with an R^2 value of 0.937.

Figure 4: Scatter plot showing actual Vs predicted THI at Dalby using the second order inverse weighted distance interpolation scheme

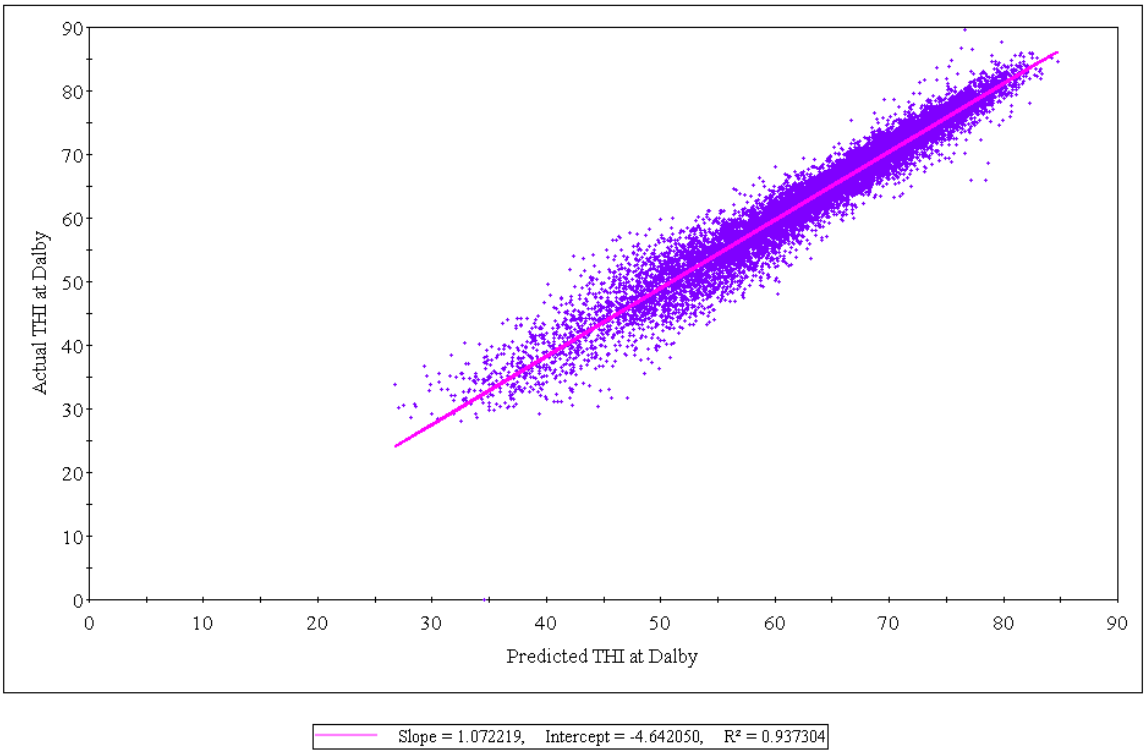
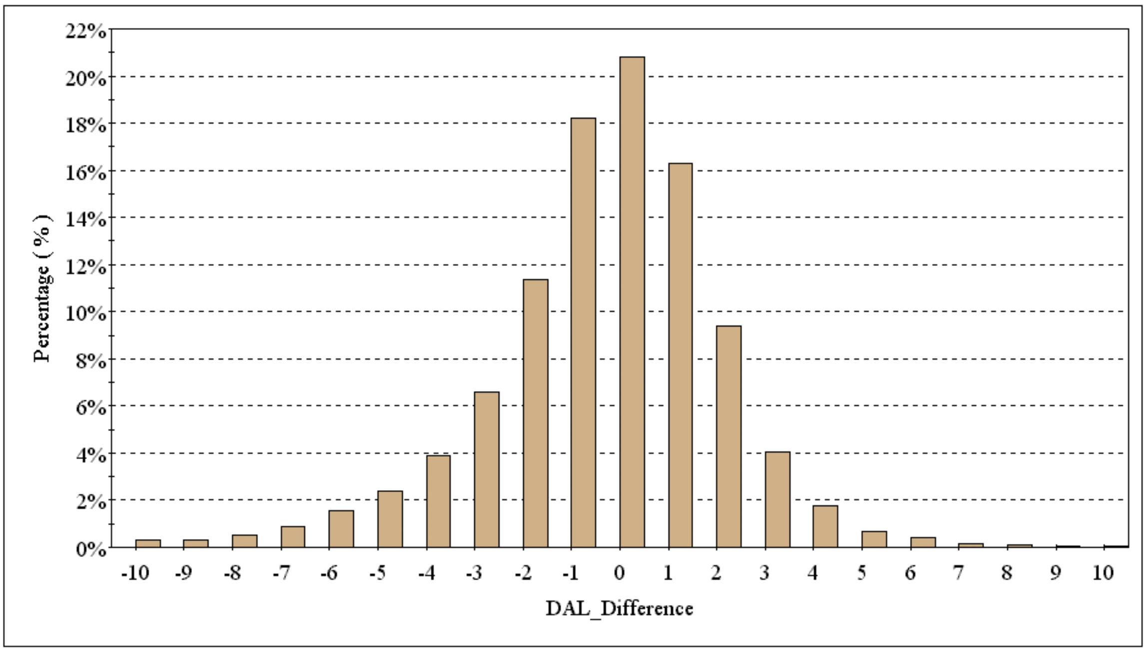


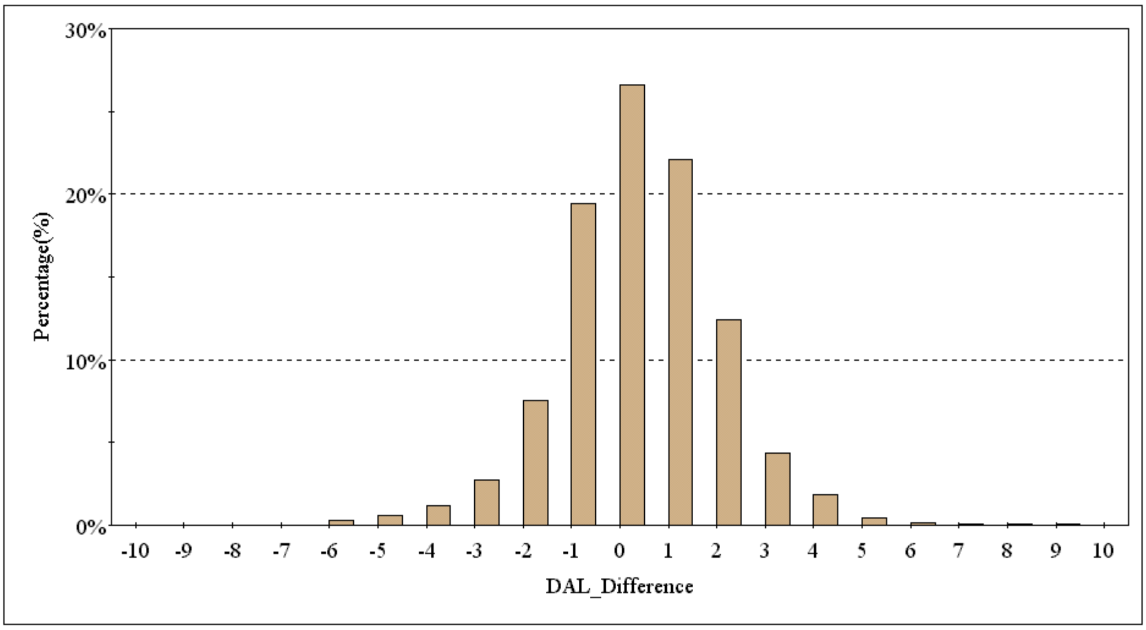
Figure 5 presents the frequency of difference between the actual THI and the predicted THI at Dalby using the second order inverse weighted distance interpolation scheme. This first figure is for all data points and the second figure shows only the results for 3pm. The increased accuracy for the 3pm only data period is probably due to the increased data from more locations for this time. For all data periods the percentage of predicted THI within 2°C of the actual THI is approximately 75%, this increases to almost 90% for the 3pm data period only. Figure 5(c) shows the accuracy of the predictions for high THI days (i.e. THI greater than 75) and indicates that the method has a tendency to under-predict the high events, with approximately 85% of predicted THI values within 2°C.

Figure 5: Histogram showing difference between actual and predicted THI at Dalby using the second order inverse weighted distance interpolation scheme for (a) all data points and (b) 3pm only (c) THI greater than 75

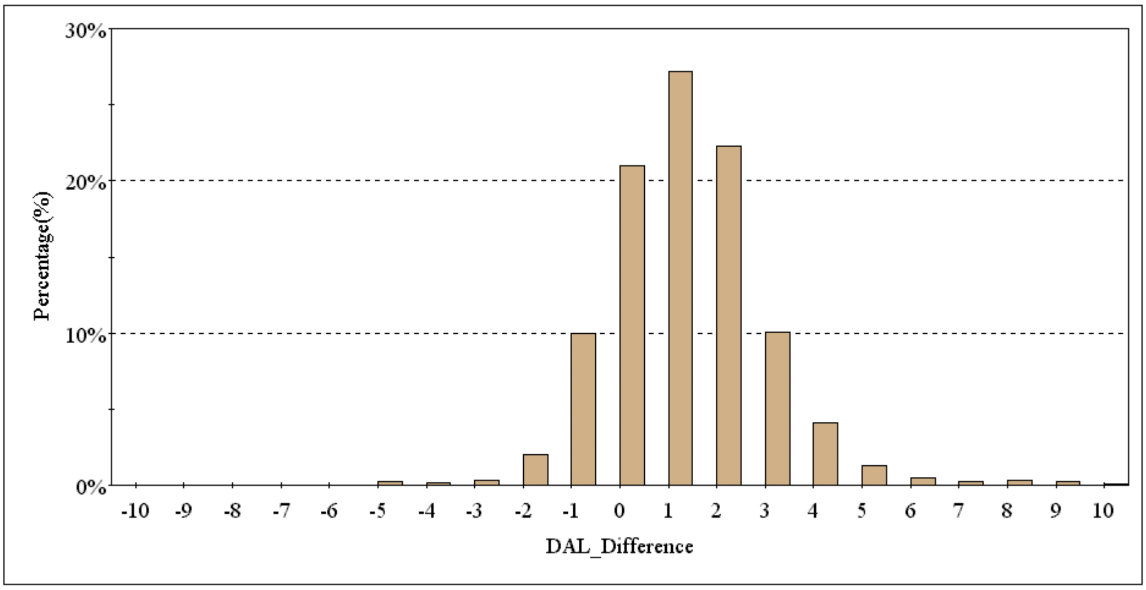
(a)



(b)



(c)



Similar results were obtained for the kriging method.

An alternative to determining the correlation between the predicted and actual values hour by hour (as presented above) is the inspection of the overall distribution at a location. As this assessment is not interested in predicting the actual THI on a given hour but more interested in the percentage occurrence of high events, this type of analysis is of value.

Figures 6 and 7 present the frequency distribution of the predicted THI values at two locations in Region 3 based on data from all other sites in the region. Various methods are presented with the simple height correction applied to the closest site giving the most encouraging results (i.e. the closer to the 1-1 diagonal line the better the predictions). A comparison is also presented for the closest site, uncorrected for height, which indicates that use of data from the closest sites is generally a good method for assessing the risk at a new location. Most methods seem to under-predict the frequency of high THI events; this is critical as the frequency of extreme events is probably the most important risk feature that requires prediction. However, it should be noted that the figures presented show more of the extreme events with the highest 4 data points corresponding to an event which may occur less than once per year and the tenth data point corresponding to an event that could occur 10 times per year. The accuracy of all methods at this points is quite good.

Figure 6: Comparison of frequency distribution of THI values at Dalby using various methods

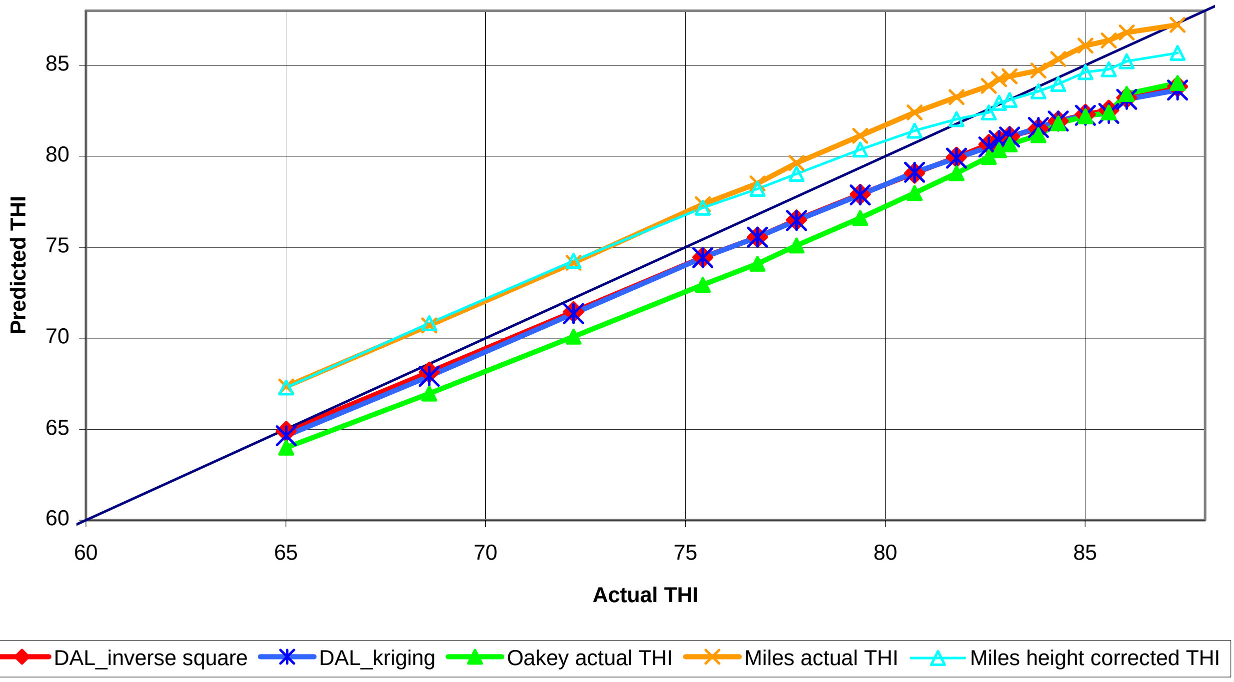
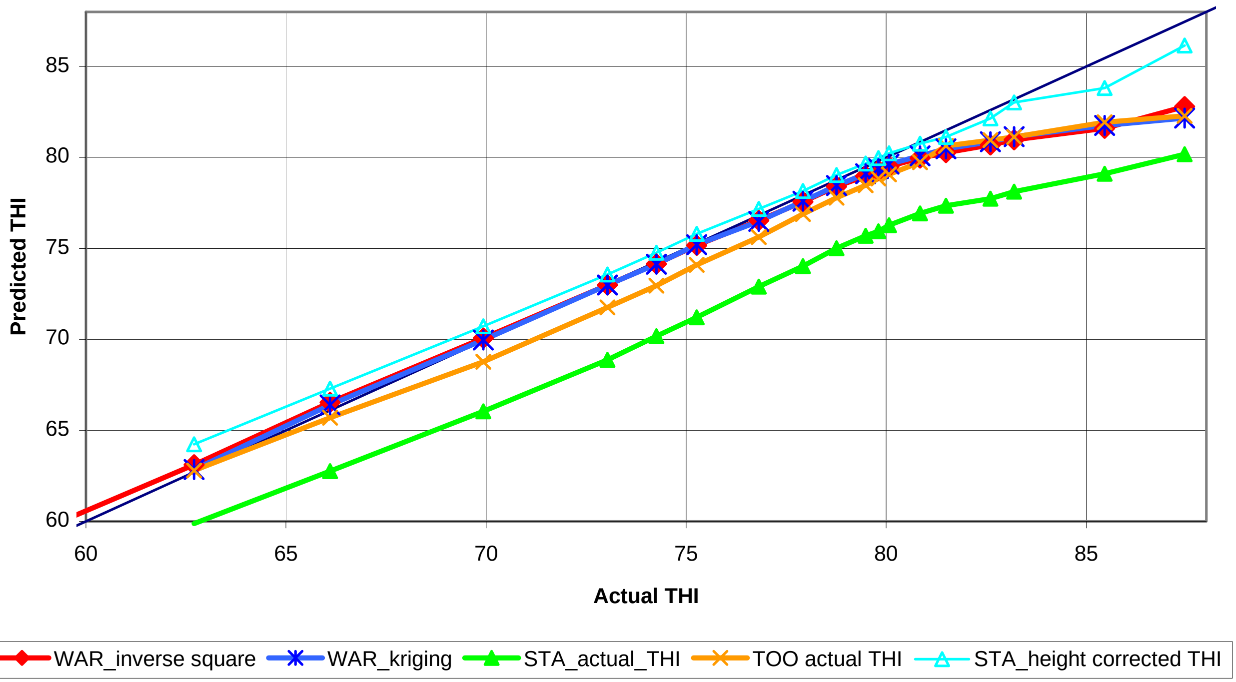


Figure 7: Comparison of frequency distribution of THI values at Warwick using various methods



It appears that all of the above methods are capable of predicting the general THI behaviour, but there is a degree of variability among the sites that can result in uncertainties in individual measurements (note that this could also be due to missing data from a number of sites). However, it can be seen from the analysis of the cumulative frequency distribution that the frequency of extreme events is predicted with a reasonably high degree of accuracy. Only the highest percentiles (less than one event per year) were possibly under-predicted. This methodology is limited to some degree by the availability of long term weather data within a region, the more sparse the data the more difficult and less accurate the interpolation.

5.3 Numerical approaches to determining spatial variability of meteorological information

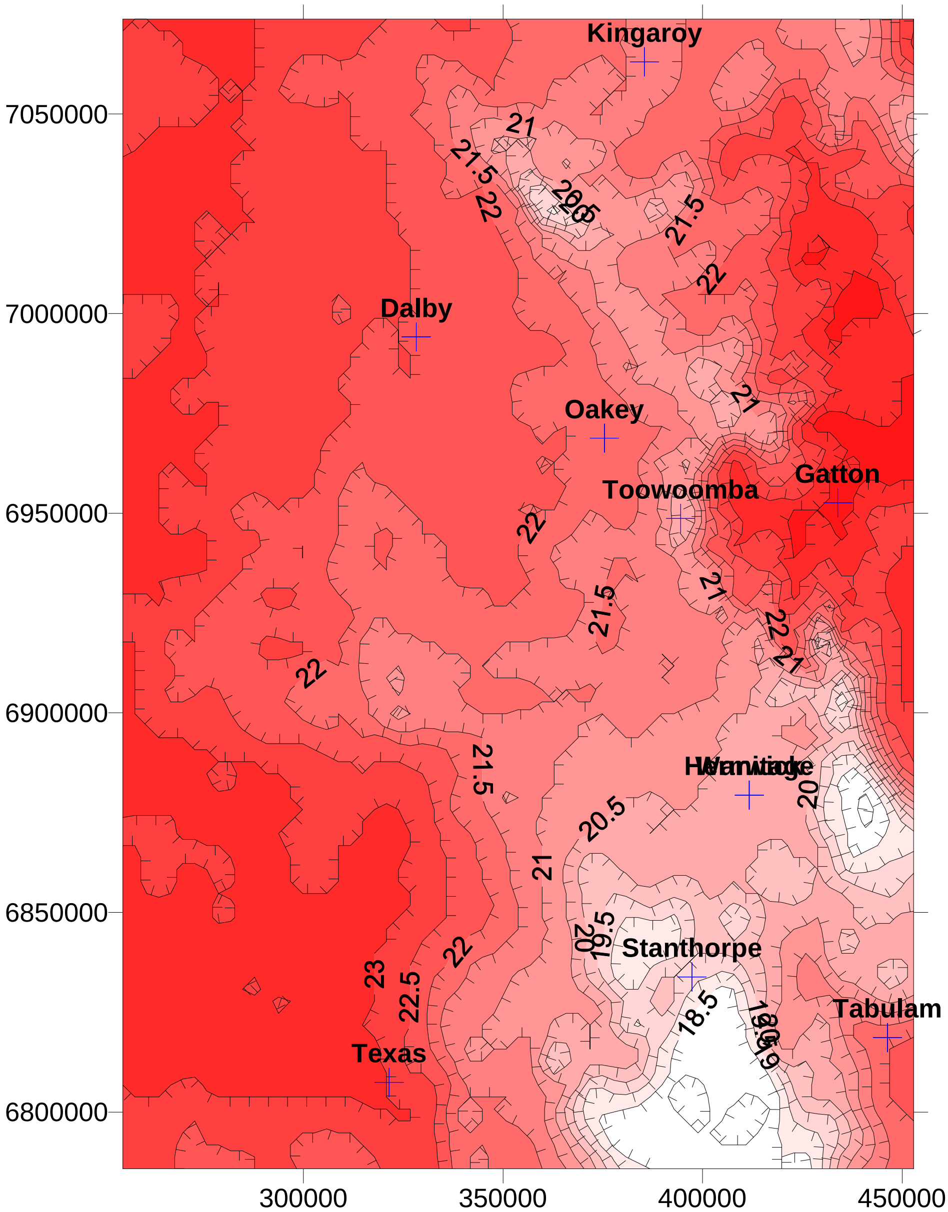
The spatial variability may be better assessed by using direct numerical modelling schemes that incorporate data assimilation schemes to incorporate the available measurements. The model results then incorporate known general synoptic conditions, detailed terrain and land use information as well as the production of non-measured parameters (e.g. heat fluxes, black globe temperatures, cloud cover and hourly rainfall).

For the project, a test case has been undertaken using the latest version of the CSIRO numerical model developed for air quality predictions (TAPM Version 2, Hurley et al (2002)). This model uses a nested modelling approach, can deal with variable synoptic conditions and has a surface data assimilation scheme. The model can be used to generate 1-4 year datasets over the feedlot regions, but is relatively computationally intensive and can take considerable time to generate a long term dataset over a large area. There is also difficulty in trying to use the scheme to track extreme conditions that may have a high degree of interannual variability (as concludes from the cluster analysis results in Appendix A to be very important).

Figure 8 presents an example of the predicted variation in average temperature over Region 3 as predicted by TAPM for a three month period.

For this project a test case was run for a short time period and a comparison of the TAPM model predictions were made with the available data for Region 3. The results were encouraging but were not considered good enough to warrant further analysis until better agreement could be made with temperature predictions. There is as yet little published work on the use of hybrid schemes that couple the best characteristics of numerical models and statistical/time series evaluation of long datasets and considerable research would be required to investigate this approach further.

Figure 8: Contour of average temperature (°C) over the region as predicted by TAPM Version 2 for the period from January 2000 - 14th March 2000.



6. CONCLUSIONS

Software has been developed that allows any user with Internet access to obtain risk information on excessive heat load conditions for feedlot sites in southeast Queensland located near Bureau AWS sites.

Analysis of historical information indicates that interpolated THI values for locations without long-term meteorological data can readily be calculated by relatively simple methods. However, the accuracy of these methods is only marginally better than a simple height correction factor applied to the closest available long-term weather station data or looking to the results from the closest station.

7. REFERENCES

Cressie NAC (1996), Comment on “An approach to statistical spatial-temporal modelling of meteorological fields” by M.S. Handcock and J.R. Wallis, Journal of the American Statistical Society, 89, pp 379-382.

Hurley P, (2002), “The air pollution model (TAPM) Version 2, User Manual”, CSIRO Atmospheric Research Internal Paper No. 25.

Hutchinson MF (1995), “Stochastic space-time weather models from ground based data”, Agricultural and Forest Meteorology, 73, 237-264.

8. APPENDIX A - STATISTICAL ANALYSIS OF REGION 3 DATA – CLUSTERING

A cluster analysis was undertaken for this assessment to identify any correlation between high THI days and meteorological conditions. This could then be used to simplify the analysis of excessive heat load risk over a long time period.

8.1 Methodology

Clustering is a technique that is used to simplify a long term data set of a range of parameters, in this case meteorological parameters, into identifiable groups or cluster types. This then allows an interpretation of the data in terms of “like” days to show pattern in the information and for this analysis allows the identification of day types that are susceptible to high THI values.

For Region 3 the data for the period from 1972 onwards from the following sites was obtained and validated:

- Toowoomba;
- Amberley;
- Oakey;
- Kingaroy;
- Dalby;
- Warwick;
- Stanthorpe;
- Miles;
- Tabulam;
- Texas;
- Hermitage and
- Gatton.

A clustering technique was used to automatically partition the meteorological data into subsets with the least difference between the variables chosen to describe the weather conditions. The clustering methodology requires all parameters analysed to be present in the source file. Any missing records will result in that day not being incorporated in the clustering. This factor limits the number of fields to be clustered together to those for which a large proportion of data is available. Therefore, from the available data two cluster files were made to incorporate the following characteristics:

- Most amount of meteorological fields but fewer records to cluster and;
- Most amount of records to cluster by minimising meteorological fields.

The two cluster files were assembled with sites, fields and hours selected to give the largest range, validity and availability of data.

Cluster file one (September 1973 to January 2001) comprised:

- Toowoomba 9 am and 3 pm measurements for wind speed, wind direction, temperature and dew point temperature.
- Dalby 9 am and 3 pm temperature.
- Gatton 9 am temperature and dew point temperature.
- Hermitage 9 am temperature.
- Kingaroy 9 am and 3 pm temperature and dew point temperature.
- Miles 9 am and 3 pm temperature and dew point temperature.
- Oakey 9 am and 3 pm measurements for wind speed, wind direction, temperature and dew point temperature.
- Tabulam 9 am temperature.
- Texas 9 am temperature.
- Amberley 9 am and 3 pm measurements for pressure, wind speed, wind direction, temperature and dew point temperature.

Cluster file two (September 1973 to September 2001) comprised:

- Toowoomba 9 am and 3 pm measurements for wind speed, wind direction and temperature.
- Dalby 9am and 3 pm temperature.
- Gatton 9 am temperature and dew point temperature.
- Hermitage 9 am temperature.
- Miles 9 am and 3 pm temperature and dew point temperature.
- Oakey 9 am measurements for wind speed, wind direction, temperature and dew point temperature.
- Tabulam 9 am temperature.
- Amberley 9 am and 3 pm measurements for pressure, wind speed, wind direction, temperature and dew point temperature.

This selection of fields allowed 4,577 valid days to be clustered for cluster file 1 and 7,199 valid days to be clustered for cluster file 2.

The clustering process allocates each valid record (or day) to 1 of 20 clusters, and records the means and standard deviations of each variable in each cluster for further reference. A visual check of the file found 20 clusters to have a relatively equal division between clusters as well as a few clusters to which only limited days have been allocated (due to extreme meteorological conditions and therefore rarity of the event).

8.2 Clustering Results - Cluster file 1

The THI value for the Toowoomba, Miles and Oakey sites at 9 am and 3 pm were calculated and added to both cluster files. For this cluster file only a few cluster types were found to be important for high THI. Figures 9 and 10 show a percentage of each cluster type occurring for THI above 75 at Oakey at 9 am (Figure 9) and Toowoomba at 3 pm (Figure 10). These figures both show cluster type 1 as being the most important for high THI. Other cluster types found to also be significant were clusters 3, 12 and 20. The characteristics of cluster type 1 were identified to have 3 pm temperatures above 30°C at all sites with wind speeds to be less than 3 m/s from the north-west to north-east.

Figure 9: Percentage of cluster types occurring for Oakey at 9 am when THI is above 75.

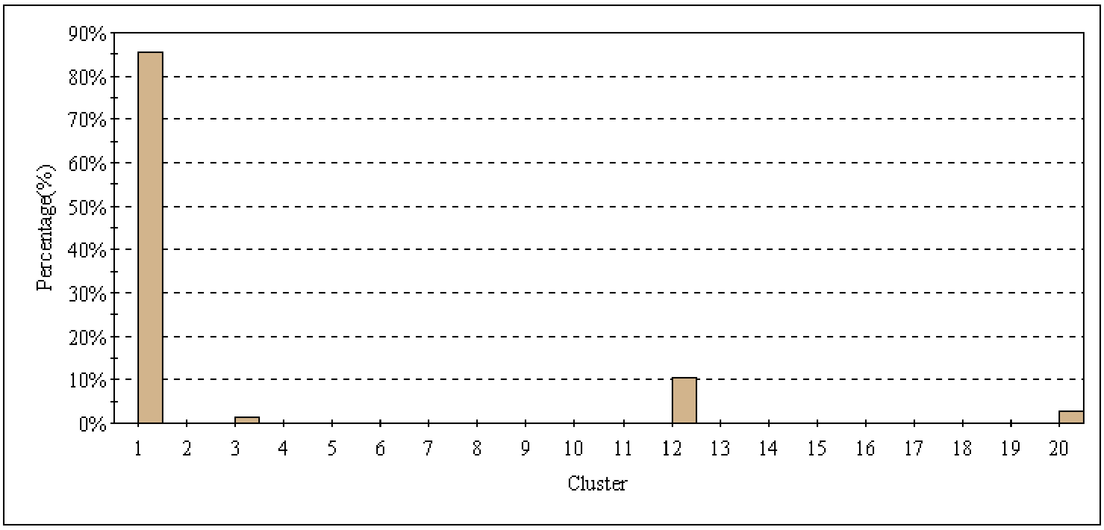


Figure 10: Percentage of cluster types occurring for Toowoomba at 3 pm when THI is above 75

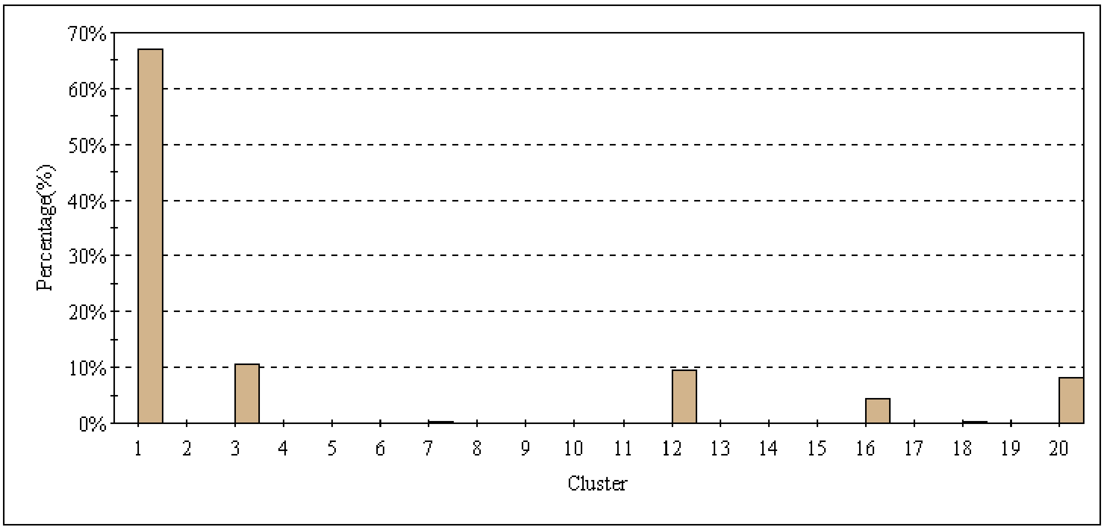
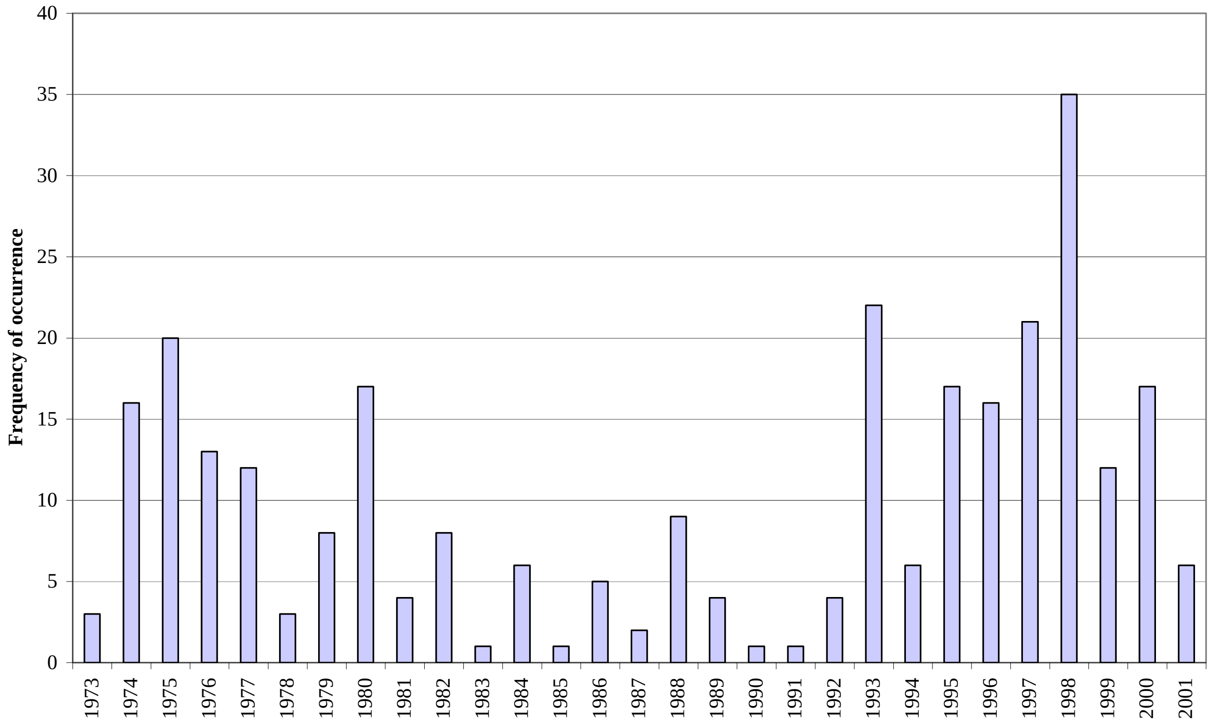


Figure 11 shows the frequency of occurrence of cluster type 1 for each year. As can be seen in this figure that inter-annual variability is significant with the frequency ranging between 1 and 35 days per year. 1998 was found to have the highest occurrence of cluster type 1, however, it must be stressed that there may be a large amount of missing data throughout the data period.

Analysis of the successive cluster types identified that cluster type 1 was most likely to occur again the next day. The second most likely cluster type to occur after cluster type 1 was cluster type 3.

Figure 11: Frequency of occurrence of cluster type 1 per year.



8.3 Clustering Results - Cluster file 2

As above, THI was calculated for Toowoomba, Miles and Oakey at 9 am and 3 pm. In cluster file 2 only a couple of cluster types were found to be important for high THI. Figures 12 and 13 show a percentage of each cluster type occurring for THI above 75 at Oakey at 9 am and Toowoomba at 3 pm. These figures both show cluster types 7 and 16 as being the most important for high THI.

Figure 12: Percentage of cluster types occurring for Oakey at 9 am when THI is above 75.

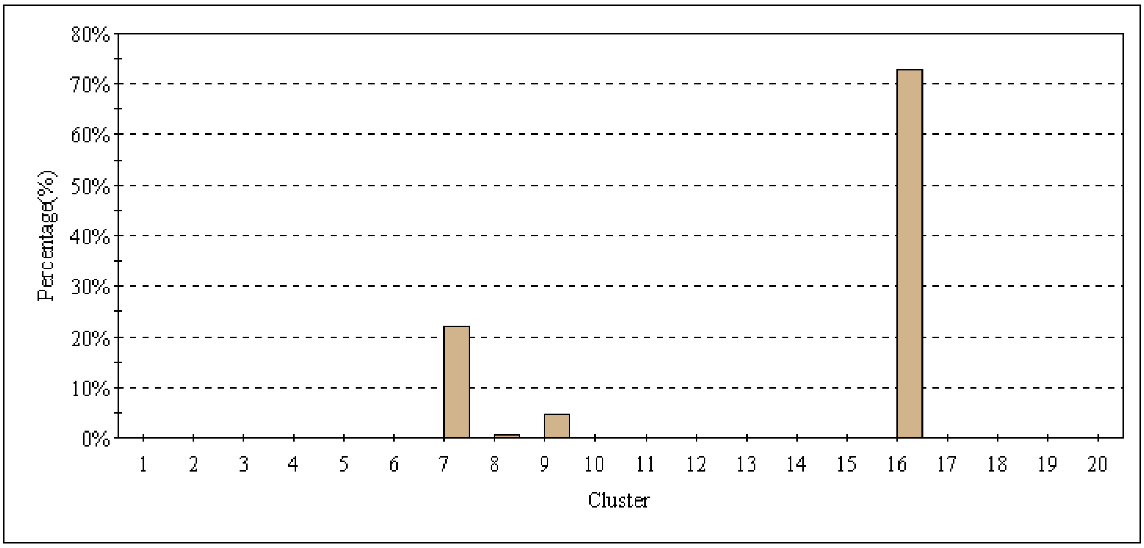
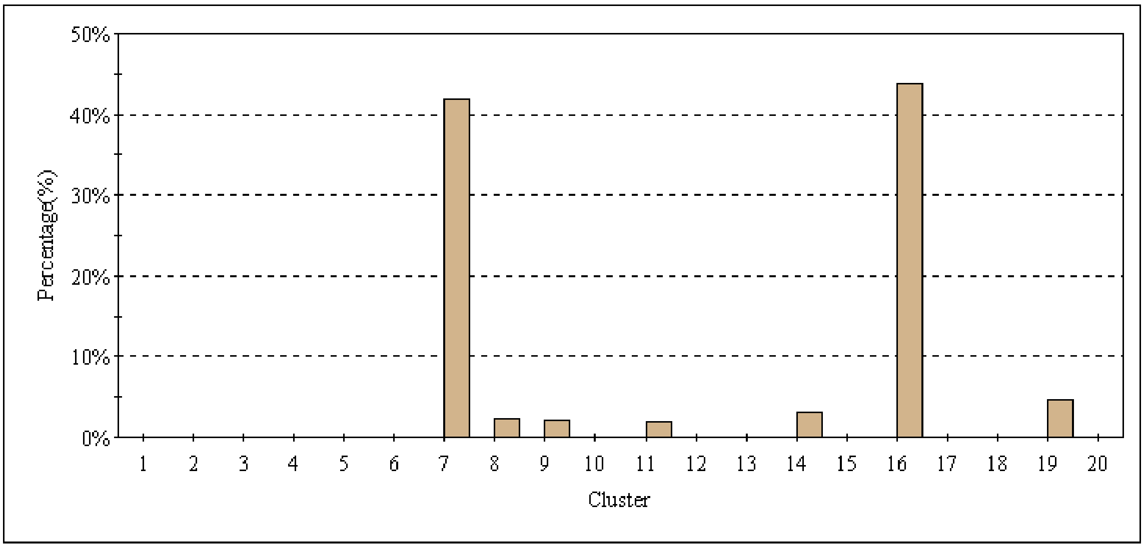


Figure 13: Percentage of cluster types occurring for Toowoomba at 3 pm when THI is above 75.



The characteristics of cluster type 7 were for 3 pm temperatures above 27°C at all sites with wind speeds to be less than 5 m/s at Toowoomba and 3 m/s at Oakey from the north-east to east-north-east. Cluster type 16 was identified as having 3 pm temperatures over 30° at all sites and winds less than 4 m/s from the west to north-westerly sector.

Figures 14 and 15 show the frequency of occurrence of cluster type 7 and 16 respectively. For cluster type 7, 1998 was shown to have recorded the highest occurrence. Cluster type 16 showed the most frequent occurrence was during 1997. It is stressed again that large amounts of data may be missing throughout the dataset.

Analysis of the successive cluster types found that when cluster type 7 occurred, it was most likely to occur again the next day compared to any other cluster type. The second most likely cluster type to follow cluster type 7 was cluster type 16.

Cluster type 16 was also identified as being most likely to follow itself followed by cluster type 7.

Figure 14: Frequency of occurrence of cluster type 7 per year.

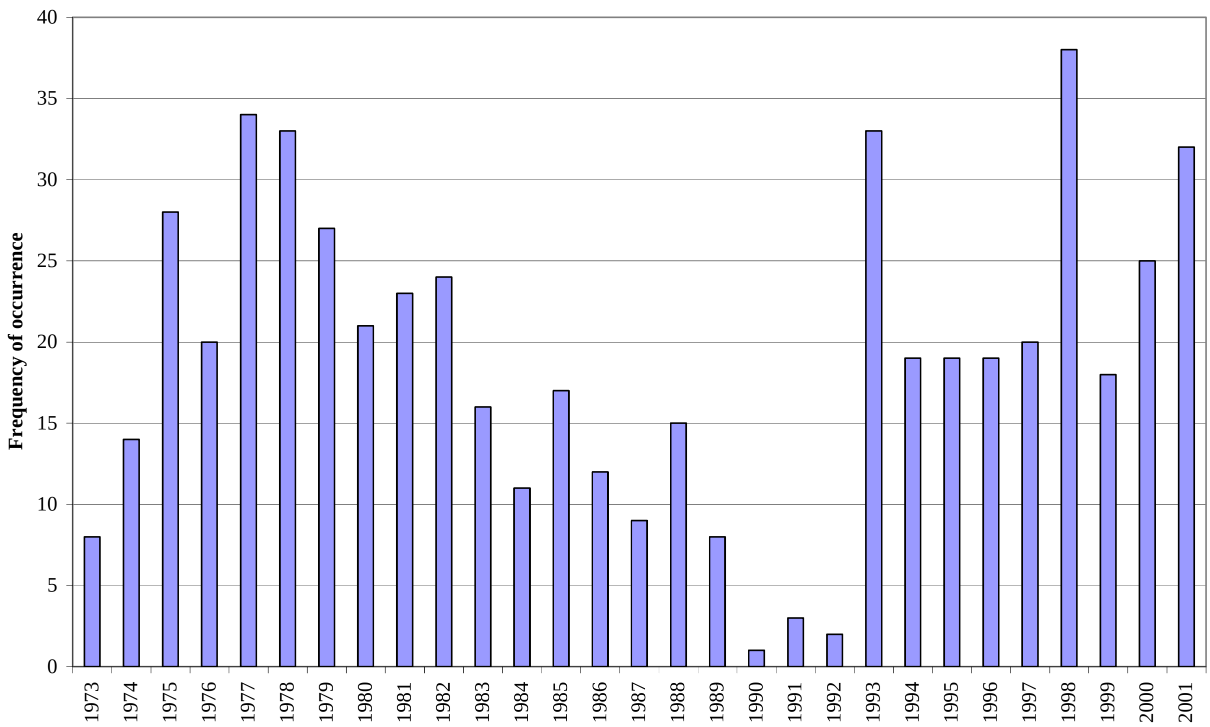
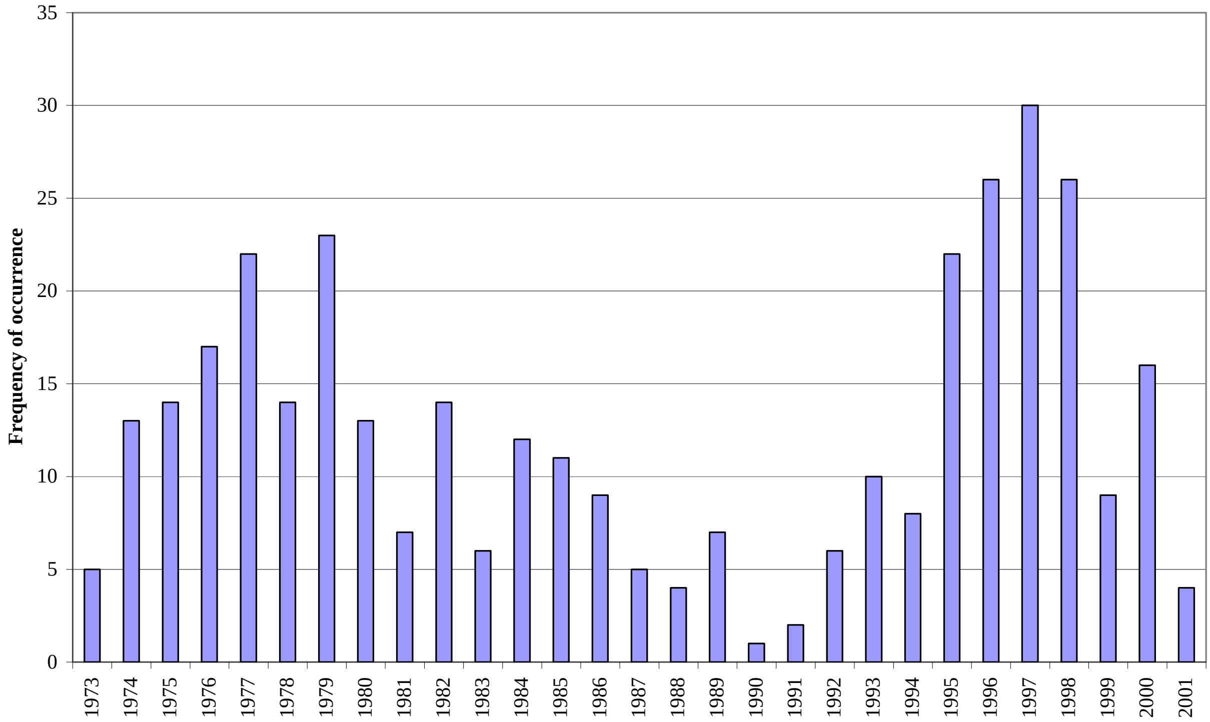


Figure 15: Frequency of occurrence of cluster type 16 per year.



8.4 Summary

Both clustering results have shown that high THI days have been identified as having only a few different characteristics such as high temperatures and light winds from the west to north-westerly sector and can be characterised by only a few cluster types.

The cluster types identified as important for high THI days were found to most likely occur for a couple of days in a row.

9. APPENDIX B - DESCRIPTION OF INTERPOLATION METHODS

The first two methods, Inverse Distance Weighting and Kriging, calculate the THI at a given (unknown) site as a weighted average of the THIs at all other (known) sites. The distinction between these two methods lies in how the weights are determined. The third technique, Laplacian Interpolation, uses an interpolation technique using the observed THIs at known locations as anchor points and calculates the missing values between these points.

9.1 Inverse Distance Weighting

Here the weight given to each site is either the inverse of the distance between the known and unknown sites (first order) or the inverse square of the distance (second order). The weights are then normalised so that their sum is unity.

9.2 Kriging

Kriging is a geo-statistical technique for spatial interpolation named after a South African mining engineer named D. G. Krige who developed the technique in an attempt to predict ore reserves more accurately. Over the past several decades kriging has become a fundamental tool in the field of geostatistics. It has since found use in a variety of other fields.

In calculating the weights, first a table of the covariances (ie the square of the difference between THI values) between all pairs of sites as a function of distance is constructed. Next an analytic function is fitted to the data in this table. Common functions are linear and exponential functions. The covariance for an unknown site relative to each known site is determined using this function. The weights are determined using the rule that if the calculated covariance is large, the weighting is small. Again the weights are normalised so that they sum to unity.

9.3 Laplacian Interpolation

This iterative technique is used to interpolate values on a grid when only a small number of values on the grid (i.e. the known sites) are available. It is used in interpolating quantities such as temperature, electric potential and concentrations resulting from diffusive processes. The technique is relatively simple – each grid point is updated by replacing it with the average of its four neighbours. The known/observed points are used in the calculations but are not updated. Since it is an iterative process, the computation terminates when subsequent iterations fail to alter the grid values by a pre-specified amount. When large grids are used and high accuracy is required this technique becomes prohibitively long.

9.4 References

Many references describing the above techniques are available in the literature and on the WEB, with the WEB references being much more easily accessible.

Kriging: http://www.uiowa.edu/~geog/health/Interp/kriging.html

Inverse distance weighting: http://www.uiowa.edu/~geog/health/Interp/inv.html

Laplacian interpolation: http://www.eece.unm.edu/faculty/humphrie/cpa/chap04.pdf