This is the author-created version of the following work:

Sexton, J., Everingham, Y., Skocaj, D., Biggs, J., Thorburn, P., and Schroeder, B. (2017) ***Identification of climatological sub-regions within the Tully mill area.*** **In: Proceedings of the 39th Annual Conference of the Australian Society of Sugar Cane Technologists, pp. 342-350. From: ASSCT 2017: 39th Annual Conference of the Australian Society of Sugar Cane Technologists, 3-5 May 2017, Cairns, QLD, Australia.**

AG35

IDENTIFICATION OF CLIMATOLOGICAL SUB-REGIONS WITHIN THE TULLY MILL AREA

By

J SEXTON[1], Y EVERINGHAM[1], D SKOCAJ[2], J BIGGS[2], P THORBURN[3], B SCHROEDER[4]

[1]*James Cook University*
[2] *Sugar Research Australia*
[3]*Commonwealth Scientific and Industry Research Organization*
[4]*University of Southern Queensland*

justin.sexton@my.jcu.edu.au
KEYWORDS: Clustering, K-means, Climate, Nitrogen, Agriculture

Abstract

IDENTIFYING optimal nitrogen application rates that reduce nitrogen loss without adversely reducing yields would benefit growers and the environment. In order to identify optimal nitrogen application rates throughout the Tully mill area, it is important to identify sub-regions that share similar topographical, soil, farm management, productivity or climatological attributes. While current SIX EASY STEPS nitrogen guidelines enable a hierarchy of district, soil, block and crop nitrogen requirements for sugarcane, it would be beneficial for management zones to also take spatial climate variability information into account. Unfortunately, spatial climate variability within a region, is generally not considered when developing nitrogen management practices. The objective of this paper was to identify sub-regions within the Tully mill area based on climatological attributes as a first step towards better informing nitrogen management decisions. Rainfall, radiation and temperature data were obtained on a 0.05 by 0.05˚ grid (approximately 5 km by 5 km) for sugarcane-growing areas within the Tully Mill region. A K-means clustering algorithm was then used to cluster these grid cells into distinct sub-regions based on seasonal or annual climate data. Two distinct sub-regions were identified based on total annual rainfall and annual average daily radiation data. These sub-regions were identified as a northern and southern sub-region, divided roughly along the Tully River. The northern sub-region was characterised by lower radiation, lower temperatures and higher rainfall than the southern sub-region. Crop simulation models will now be able to use this knowledge to assess if nitrogen management plans should vary between the two sub-regions in Tully.

Introduction

Climate is an important component for planning farming, harvesting, transport, milling and marketing activities in the Australian sugar industry. Climate varies considerably from region to region in the Australian sugar industry. For example, the cooler temperatures in New South Wales is linked with two-year-old crop cycles in this region (Hughes and Muchow, 1998). In drier regions like Bundaberg and the Burdekin, irrigation management is a priority issue (Everingham *et al.*, 2008; Flynn, 2015). However, in wetter regions such as Tully in the Wet Tropics where the average annual rainfall exceeds 4000 mm, priority issues include harvest management scheduling and denitrification due to anaerobic conditions associated with water logging (Skocaj *et al.*, 2013).

Whilst climate will vary substantially from region to region, it can also vary within a region. One well known example of this is in the Herbert sugarcane growing region. Garside

(2013) found that rainfall in November had a lesser effect on cane yields in some Herbert districts than others. Specifically, in the southern Ingham Line district cane and sugar yields were less affected by seasonal rainfall (Garside *et al.*, 2014). In his review of productivity trends in the Herbert, Garside (2013) recommended that district and sub-district rainfall differences be considered in developing management practices.

Other studies have also recommended developing harvest management practices based on spatial rainfall variability within a region. Sexton *et al.* (2015) examined the impact of climate change on harvest management using climate projections for 2050 within 7 sugarcane-growing regions in Australia. For the Herbert region Sexton *et al.* (2015) found that the projected number of unharvestable days increased in northern districts, but decreased in the southern districts. These authors further suggested that adaptation strategies should take such factors into consideration. For example, preferential scheduling of harvests would be beneficial in areas with a greater risk of increased days when harvests would not be possible due to wet conditions. Prior to Sexton *et al.* (2015), Higgins and Muchow (2003) used mathematical optimisation methods to propose a new harvest management system that varied spatially within regions. They found that industry profitability could be increased by up to A$157/ha for the Invicta mill if harvest scheduling was managed based on sugar accumulation profiles on a farm by farm basis. Further, they found that profitability for individual districts within a region could vary depending on local rainfall patterns throughout the growing season.

Although the impact of climate variability within a region on harvest management has been explored, the impact on nitrogen (N) fertilizer management decisions remains relatively unexplored. For example, the SIX EASY STEPS N guidelines are currently based on a soil specific nitrogen mineralization index and district yield potentials (Schroeder *et al.*, 2010) but does not consider climatic differences within a region. Recently, Skocaj (2015) has shown that optimal N application rates can differ due to temporal climate variability, suggesting that N fertilizer rates could be reduced in years with a high spring-summer rainfall in the Tully region. It stands to reason then that spatial climate variability may also provide guidance for N fertilizer management decisions.

While the SIX EASY STEPS N guidelines enable a hierarchy of district, soil, block and crop N requirements for sugarcane, it would be beneficial for management zones to also take into consideration spatial climate variability within a region. The objective of this paper was to identify climate zones within the Tully region as a first step towards developing N management guidelines that are informed by local spatial climate variability.

Materials and methods

Data

The Tully mill region was identified as spanning from latitude 17.75˚ S to 18.30˚ S and from longitude 145.65˚ E to 146.10˚ E (Figure 1 (a)). Sugarcane-growing areas in the Tully mill region were identified on an 0.05 by 0.05˚ (approximately 5 x 5 km) grid (Figure 1 (b)). In total 52, 0.05 by 0.05˚ sugarcane grid cells were identified.

Daily climate data for each 0.05 by 0.05˚ grid cell were obtained from the Scientific Information for Land Owners (SILO), Data drill data base (Jeffrey *et al.*, 2001). This database contains daily climate data spatially interpolated from climate station records. Daily rainfall, temperature and solar radiation data were obtained for the 40 year period from 1975 to 2014.

For each sugarcane-growing grid cell, mean daily maximum temperature, mean daily minimum temperature and mean daily radiation as well as total rainfall were calculated for summer (Dec – Feb), autumn (Mar – May), winter (Jun – Aug), spring (Sep – Nov) and annual (Jan – Dec). For each sugarcane growing grid cell, the median of the 40 years of data was calculated for use in the spatial clustering analysis. This gave 20 climate based variables for use in the analysis (five 'seasons' (summer, autumn, winter, spring, annual) * four climate variables (total rainfall, mean max temperature, mean minimum temperature, mean radiation)) each with 52 observations (0.05 by 0.05˚ sugarcane grid cells).

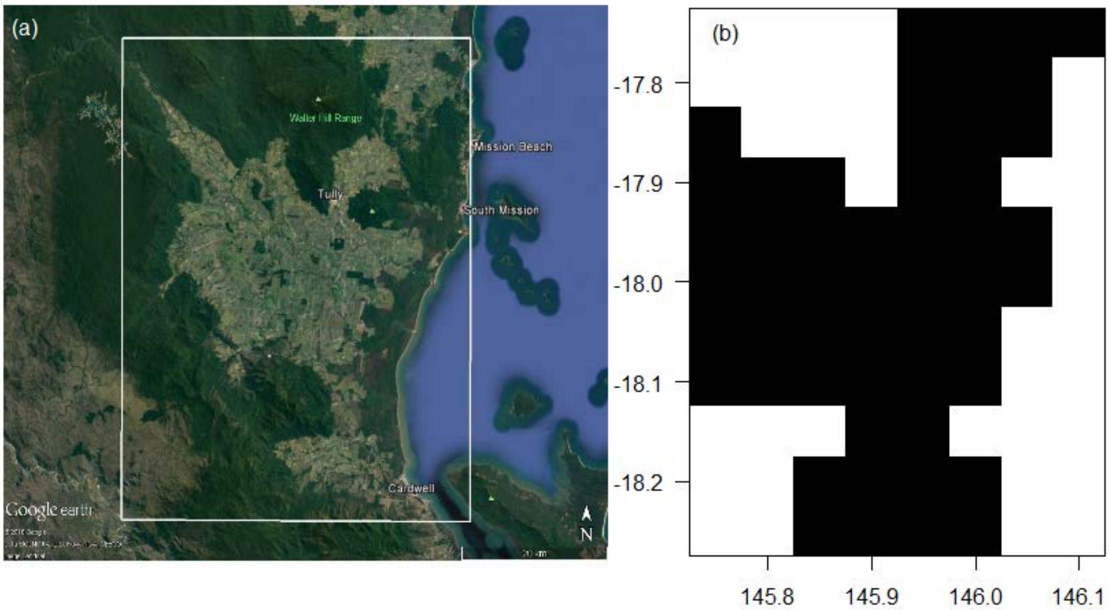

Fig. 1 — A map of the Tully mill area (a) and sugarcane-growing land in the Tully mill area identified on a 0.05 by 0.05˚ grid (b).

Cluster analysis

K-means clustering (Hartigan and Wong, 1979) was used to cluster grid cells based on the 20 possible predictor variables. K-means clustering is method of finding K clusters of observations in a set of unlabelled data (Hastie *et al.*, 2013). The number of cluster centers (K; means) must be chosen by the user. The K-means algorithm then moves the cluster centers iteratively to minimize the within cluster variance, usually using the Euclidean distance (a least squares approach). The iterative procedure can be described in two steps (Hastie *et al.*, 2013):

1. For each center identify the subset of points that are closer to that center than any other.
2. Compute the mean for each predictor variable in each cluster, with the new vector of means becoming the new cluster center.

Steps 1 and 2 are repeated until the cluster centers do not change.

While several K-means algorithms can be used, the Hartigan-Wong algorithm performs well and has the advantage of ensuring that no single observation switching clusters will decrease the variance further (Hastie *et al.*, 2013). K-means has been used in defining rainfall zones in the East Asia monsoon region (Awan *et al.*, 2015) as well as identifying in-field management zones for crops such as cotton in the United States of America (Boydell and McBratney, 2002) and sugarcane in Australia (Cupitt and Whelan, 2001).

The clustering algorithm “kmeans” in the R “stats” package (R Core Team, 2014), was applied to build 10 cluster models (Table 1). Models were built using all climate variables for each season (*e.g.* summer rainfall, summer radiation, summer maximum temperature, summer minimum temperature) and annually (*e.g.* annual total rainfall). Models were then built using only rainfall and radiation variables together as these are believed to have the most influence on yields in the Tully mill area.

Table 1 — Design matrix of K-means algorithms analysed in this report. Numbers refer to the total number of predictor variables given to the K-means algorithm.

Season	All climate variables	Rainfall and Radiation variables
Summer	4	2
Autumn	4	2
Winter	4	2
Spring	4	2
Annual	4	2

There are many different indices that can be used to identify the "best" number of clusters. For each analysis of the design matrix (Table 1) the best number of clusters was found using the "NbClust" package (Charrad *et al.*, 2014) in the R statistical analysis program. The NbClust package provides 30 indices for determining the ideal number of clusters. The final recommended number of clusters is the mode number of clusters recommended by all of the indices.

The NbClust analysis requires the specification of the distance measure (set to Euclidean by default), minimum and maximum number of clusters (set to two and 10 respectively), method of clustering (K-means) and indices to be calculated. The best number of clusters was recorded for each analysis. The "kmeans" function (R Core Team, 2014), in the R statistical package was then used to perform each cluster analysis from Table 1. For all analyses predictor variables were standardized using the "scale" function in the R statistical analysis program. Scaling the data removed any effect of difference in scale between the climate variables. The resulting clusters were plotted spatially. Graphical analysis of the modelled clusters was used to check the sensibility of the modelled clusters and identify the most appropriate cluster model. The differences between the climatic variables for each sub-region identified by the final cluster model were then explored using boxplots. Significant differences in climate variables between clusters were identified using a Kruskal-Wallis test. Significance was tested at the 0.05 level.

Results and discussion

Identifying number of clusters

When all climate variables were considered, three was identifed as the best number of clusters for each season except spring. Using spring data, a two cluster solution was identified as the best number of clusters (Table 2). When only radiation and rainfall variables were considered, data from individual seasons were split between three clusters when using summer or spring data and two clusters when using autumn or winter data. The low number of clusters chosen is advantageous as it is easier to interpret two or three larger spatial clusters than to interpret a large number of small spatial clusters.

Table 2 — Recommended number of clusters from NbClust.

Season	All Climate Variables	Rainfall and Radiation Variables
Summer	3	3
Autumn	3	2
Winter	3	2
Spring	2	3
Annual	3	2

Spatial plot of clusters

Clustering seasonally on all climate variables

Spatial patterns of the three cluster solutions for summer (Figure 2 (a)), autumn (Figure 2 (b)), winter (Figure 2 (c)) and annual (Figure 2 (e)) were relatively consistent. Clusters 1 and 2 were spatially continuous representing a north-eastern cluster (blue) and south-western cluster (red). The third cluster (green) consisted of the same six grid cells in each season. By comparison with Fig 1. It is clear these grid cells were located on the edge of local mountains and may represent higher elevations compared to the other clusters. The same six grid cells formed the second cluster (red) for spring data (Figure 2 (c)).

Comparison of individual climate variables between the clusters suggested that the north-eastern clusters tended to be characterized by higher rainfall and lower radiation compared to the south-western clusters (data not shown). The clusters consisting of the six mountainous grid cells tended to be characterized by lower maximum and minimum temperature variables than either the north-eastern or south-western clusters (data not shown). The interpretation of the small mountainous clusters should be considered with care. Due to the interpolated nature of the climate data used in the analysis (Jeffrey *et al.*, 2001), and the relatively course grid size, these cells may be representing climate features of a higher elevation than any crop is actually being grown.

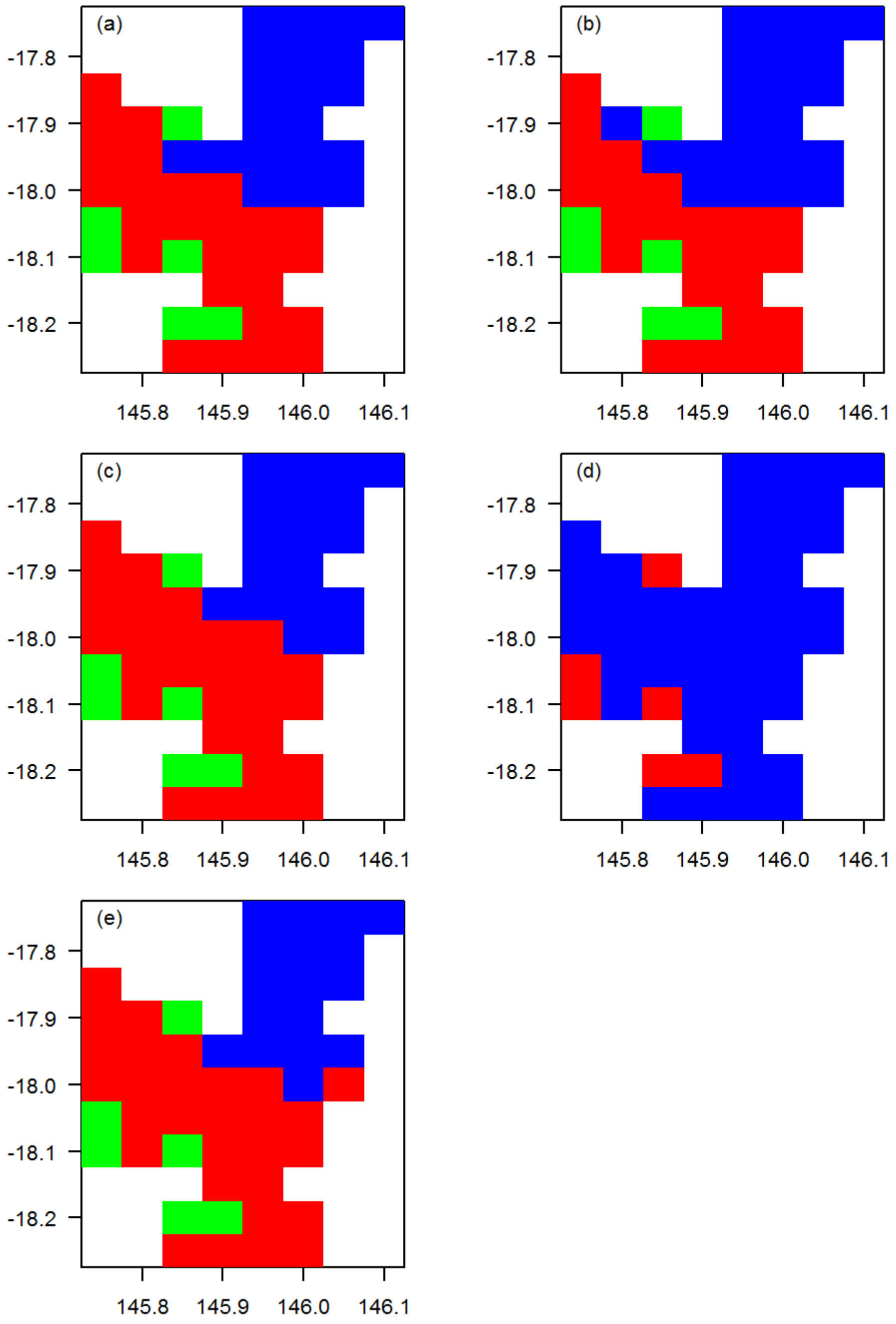

Fig. 2 — Spatial clusters in the Tully mill area for (a) summer, (b) autumn (c) winter, (d) spring and (e) annual. Clusters were based on all climate variables, using the 'best' number of clusters identified.

<u>Clustering on rainfall and radiation variables</u>

FigureSpatial patterns of the two cluster solutions for autumn (Figure 3 (b)), winter (Figure 3 (c)) and annual (Figure 3 (e)) were consistent. Clusters 1 and 2 were spatially continuous representing a north-eastern cluster (blue) and south-western cluster (red).

The three cluster solution for spring (Figure 3 (d)), had a more direct north/south divide between the two primary clusters than in other seasons. The small third spring cluster (green) consisted of the same six mountainous grid cells as the three cluster solutions using all climate variables (Figure 2). In comparison, the three cluster solution for summer rainfall and radiation data (Figure 3 (a)) consisted of three more evenly divided clusters. Comparison of the rainfall and radiation data suggested that rainfall decreased from the northern most to southern most cluster while radiation increased (data not shown).

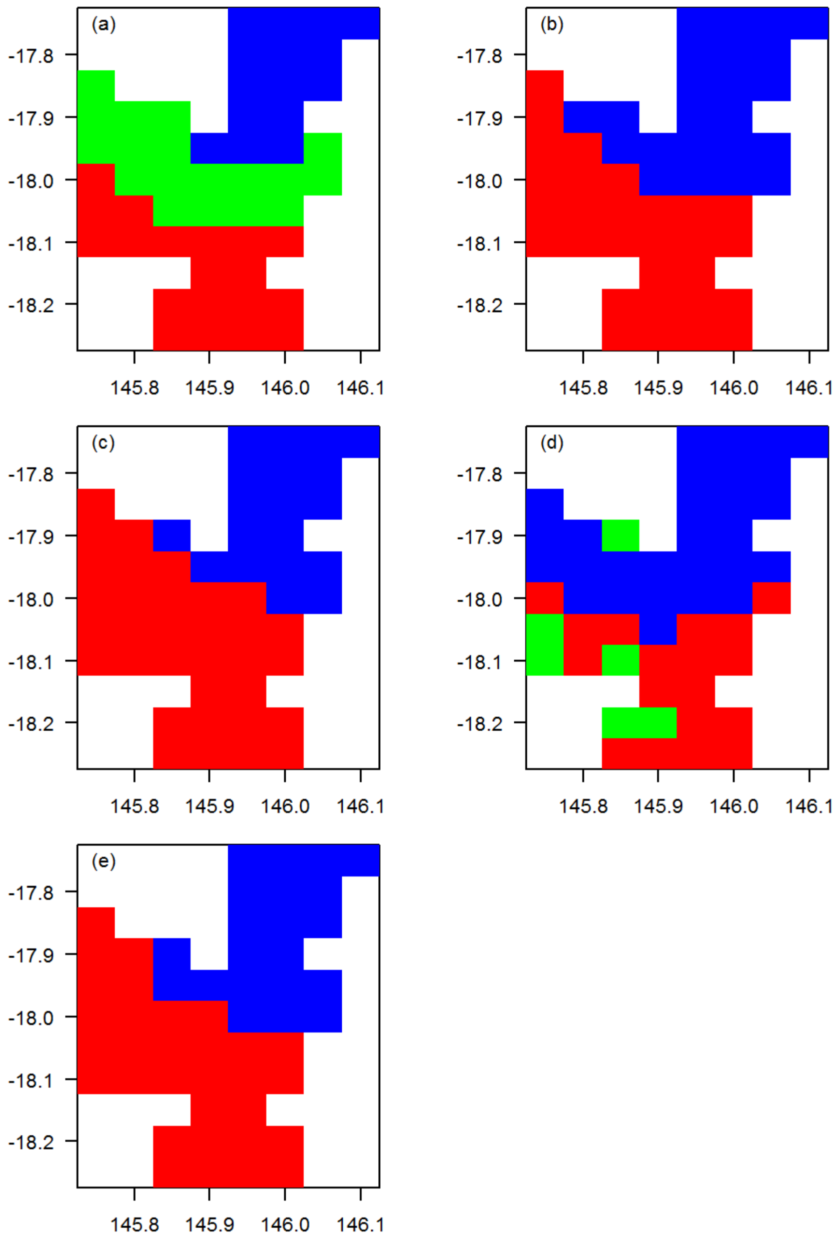

Fig. 3 — Spatial clusters in the Tully mill area using rainfall and radiation climate variables for (a) summer, (b) autumn, (c) winter, (d) spring and (e) annual climate data. Clusters represent the 'best' number of clusters for each season (Table 2).

Analysis of final clusters

The final clusters were chosen based on the annual rainfall and radiation data model as shown in Figure 3 (e). The final model was used to describe two clusters based on the suggestion of the NbClust algorithm (Table 2) and expert industry opinion. Spatially, the Tully region was split into southern and northern climatic sub-regions (Figure 4 (a)). Geographically, these sub-regions split roughly along the Tully River.

The southern sub-region (red) was characterised by higher radiation and lower rainfall than the northern sub-region (blue) with an obvious difference in the distribution of annual rainfall, radiation and maximum temperature between the two sub-regions (Figure 4). However, there was little graphical evidence for a difference in annual minimum temperature. Results from the Kruskal-Wallis test confirmed that differences in annual rainfall, radiation and maximum temperatures were significant at the 0.05 level, while there was no significant difference in annual minimum temperature. Therefore, the northern sub-region can be described as the area north-east of the Tully River consisting of higher rainfall, lower radiation and lower maximum temperatures. The southern sub-region can then be described as the area south of the Tully River consisting of lower rainfall and higher radiation and maximum temperatures. This agreed with local industry knowledge.

Boxplots of annual maximum (Figure 3 (d)) and minimum (Figure 3 (e)) highlighted a small number of grid cells with particularly low temperature values (points). These 'outliers' were identified as the six grid cells located closer to the mountain ranges. The same six grid cells that make up the smaller cluster of three cluster solutions such as those shown in Figure 2 and Figure 3 and the two cluster solution for spring in Figure 2 (c). This suggests that if the identified two cluster solution were used in decision support tools, analysis for growers close to the mountain ranges should be considered carefully. Unfortunately, one of the limitations of the K-means clustering approach is the mutually exclusive nature of the clusters. As real climate patterns are unlikely to have sharp divisions between clusters locations, future research could consider less constrictive clustering algorithms. For example, fuzzy K-means has been used in defining climatic zones in recent research (Bharath and Srinivas, 2015).

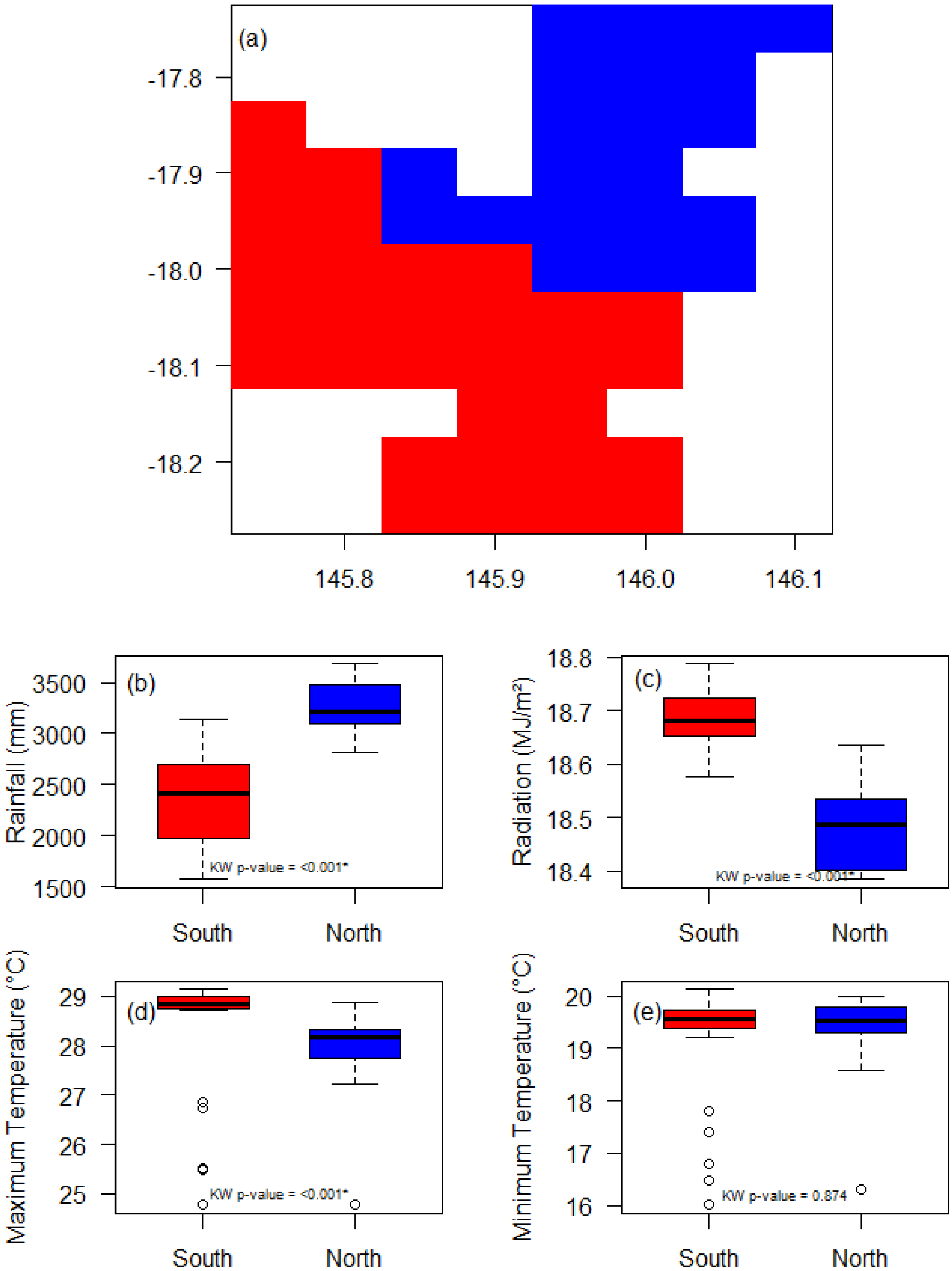

Figure 4—Climatic sub-regions in the Tully mill area using annual rainfall and radiation climate data. (a) Spatial plot showing the northern (blue) and southern (red) sub-regions. Figures (b) to (e) contain boxplots of median (1970:2015) annual climate variables for all 0.5 by 0.5˚ grid cells in the southern (red) and northern (blue) sub-regions. Boxplots show: (b) Total rainfall, (c) average daily radiation (d) average daily minimum temperature and (e) average daily minimum temperature. Differences in rainfall, radiation and maximum temperature were significant at the 0.05 level based on the KW test.

Given that rainfall in the Tully region has been found to influence optimum nitrogen levels (Skocaj, 2015), growers in lower rainfall areas may be able to consider applying lower rates of nitrogen fertilizer than growers in higher rainfall areas. The two cluster solution identified should therefore be considered when developing new fertilizer best management practices for the Tully region. Similar analysis could be applied to other sugarcane growing regions such as the Herbert, where it is likely that climatic sub-regions exist. The analysis described in this paper could easily be extended to other primary industries.

Conclusions

To identify optimal N application rates throughout the Tully mill area it is important to identify sub-regions that share similar climate attributes. Two distinct sub-regions were identified based on total annual rainfall and radiation data together. These sub-regions were identified spatially as a southern and northern sub-region, divided roughly by the Tully River. The southern sub-region was characterised by lower rainfall and higher radiation compared to the northern sub-region agreeing with local knowledge. Sub-regions within the Tully mill area could be further refined by including other attributes such as soil or topography. Future work will assess if crop model simulations indicate different N management strategies for each sub-region. The research outlined in this paper, will provide a valuable starting point for these simulations to be conducted. Although this paper has focused on identifing climate zones to improve nitrogen management, knowledge of regions that differ in climate can also be useful for improving harvesting management and identifying varieties that are better suited for different climates. The methodology outlined in this paper is easily extended to other regions in the Australian and international sugarcane growing regions.

Acknowledgments

This research was supported by funding from Sugar Research Australia and the Department of Environment and Heritage Protection.

REFERENCES

Awan JA, Bae D-H, Kim K-J (2015) Identification and trend analysis of homogeneous rainfall zones over the East Asia monsoon region. *International Journal of Climatology* **35**, 1422-1433. doi: http://dx.doi.org/10.1002/joc.4066.

Bharath R, Srinivas VV (2015) Delineation of homogeneous hydrometeorological regions using wavelet-based global fuzzy cluster analysis. *International Journal of Climatology* **35**, 4707-4727. doi: http://dx.doi.org/10.1002/joc.4318.

Boydell B, McBratney AB (2002) Identifying potential within-field management zones from cotton-yield estimates. *Precision Agriculture* **3**, 9-23. doi: http://dx.doi.org/10.1023/A:1013318002609.

Charrad M, Ghazzali N, Boiteau V, Niknafs A (2014) NbClust: An R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software* **61**, 36. doi: http://dx.doi.org/10.18637/jss.v061.i06.

Cupitt J, Whelan B (2001) Determining potential within-field crop management zones. In 'Proceedings of the 3rd European conference on precision agriculture. pp. 7-12.

Everingham Y, Baillie C, Inman-Bamber G, Baillie J (2008) Forecasting water allocations for Bundaberg sugarcane farmers. *Climate Research* **36**, 231-239. doi: http://dx.doi.org/10.3354/cr00743

Flynn J (2015) Data driven irrigation boosts Bundaberg's potential. *Australian Canegrower* 19-20. Available: http://search.informit.com.au/documentSummary;dn=505032183397984;res=IELHSS [Accessed 14 Nov 16].

Garside AL (2013) Review of productivity trends in the Herbert sugarcane growing region. Herbert Cane Productivity Services Limited Publication 1/2013.

Garside AL, Di Bella LP, Sefton M, Wood AW (2014) Review of productivity trends in the Herbert sugarcane growing region. *Proceedings of the Australian Society for Sugar Cane Technologists* **36**, 53-61.

Hartigan JA, Wong MA (1979) Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **28**, 100-108. doi: 10.2307/2346830.
Hastie T, Tibshirani R, Friedman JH (2013) 'The elements of statistical learning.' (Springer: New York)
Higgins AJ, Muchow RC (2003) Assessing the potential benefits of alternative cane supply arrangements in the Australian sugar industry. *Agricultural Systems* **76**, 623-638. doi: http://dx.doi.org/10.1016/S0308-521X(02)00031-8.
Hughes RM, Muchow RC (1998) Contribution of first and second year growth to the yield of two-year cane in NSW. *Proceedings of the Australian Society of Sugar Cane Technologists* **20**, 280-287.
Jeffrey SJ, Carter JO, Moodie KB, Beswick AR (2001) Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environmental Modelling & Software* **16**, 309-330. doi: http://dx.doi.org/10.1016/S1364-8152(01)00008-1.
R Core Team (2014) 'R: A language and environment for statistical computing.' (R Foundation for Statistical Computing: Vienna, Austria)
Schroeder B, Hurney AP, Wood AW, Moody PW, Allsopp PG (2010) Concepts and value of the nitrogen guidlines contained in the Australian sugar industry's "Six Easy Steps" nutrient management program. *Proceedings of the International Society of Sugar Cane Technologists* **27**, 26.
Sexton J, Everingham Y, Timbal B (2015) Harvest disruption projections for the Australian sugar industry. *International Journal of Climate Change Strategies and Management* **7**, 41-57. doi: http://dx.doi.org/10.1108/IJCCSM-03-2013-0018.
Skocaj D (2015) Improving sugarcane nitrogen management in the Wet Tropics using seasonal climate forecasting. Research thesis, James Cook University.
Skocaj D, Everingham Y, Schroeder B (2013) Nitrogen management guidelines for sugarcane production in Australia: Can these be modified for wet tropical conditions using seasonal climate forecasting? *Springer Science Reviews* **1**, 51-71. doi: http://dx.doi.org/10.1007/s40362-013-0004-9.