

Papers published in *Hydrology and Earth System Sciences Discussions* are under open-access review for the journal *Hydrology and Earth System Sciences*

Spatially explicit seasonal forecasting using fuzzy spatiotemporal clustering of long-term daily rainfall and temperature data

M. B. Plain, B. Minasny, A. B. McBratney, and R. W. Vervoort

Faculty of Agriculture, Food, and Natural Resources, The University of Sydney, Australia

Received: 2 January 2008 – Accepted: 3 March 2008 – Published: 14 May 2008

Correspondence to: B. Minasny (b.minasny@usyd.edu.au)

Published by Copernicus Publications on behalf of the European Geosciences Union.

1159

Abstract

A major limitation of statistical forecasts for specific weather station sites is that they are not spatial in the true sense. And while spatial predictions have been studied, their results have indicated a lack of seasonality. Global Circulation Models (GCMs) are
5 spatial, but their spatial resolution is rather coarse. Here we propose spatially explicit seasonal forecasting, based on the Fuzzy Classification of long-term (40 years) daily rainfall and temperature data to create climate memberships over time and location. Data were obtained from weather stations across south-east Australia, covering sub-tropical to arid climate zones. Class memberships were used to produce seasonal
10 predictions using correlations with climate drivers and a regression rules approach. Therefore, this model includes both local climate feedback and the continental drivers. The developed seasonal forecasting model predicts rainfall and temperature reasonably accurately. The final 6-month forecast for average maximum temperature and rainfall produced relative errors of 0.89 and 0.56 and Pearson correlation coefficients
15 of 0.83 and 0.82, respectively.

1 Introduction

Climate change due to human impacts has become a major issue for the 21st century. Climate variability and climate change have significant consequences for climate dependent systems, such as agriculture, introducing uncertainty into management strategies
20 (Hansen and Sivakumar, 2006). Seasonal climate forecasts can improve decisions at the management level, especially in agriculture, natural resource management and in relation to climate change (Meinke and Stone, 2005).

Much of the current research into seasonal forecasts is based on the global set of climate anomalies referred to as El Niño-Southern Oscillation (ENSO) and the related
25 sea surface temperature indicators, which are a major source of inter-annual climate variability (Cane, 2005; Verdon and Franks, 2005), particularly in Australia.

1160

The use of classification and interpolation techniques for developing climate prediction models is relatively new. The majority of the literature on these topics is in other areas of geophysical science. For instance, the classification and interpolation of data sets has been widely used in the field of soil science (McBratney and DeGrujter, 1992). Interpolation of climate data using geostatistical or other techniques has been attempted (Hunter and Meentemeyer, 2005). For example, Stahl et al. (2006) review a range of papers using interpolation techniques and apply twelve different techniques, including kriging, to minimum and maximum temperature in British Columbia. The most often used predictor in linear regression and universal kriging of climate data is elevation, but the actual lapse rate with elevation (the slope of the rainfall or temperature relationship with elevation) changes with time within a year and possibly decadal and this makes prediction difficult (Stahl et al., 2006).

There are further shortcomings in the conventional methodologies, such as the inability to account for climate-forcing factors (Goovaerts, 2000). Advances in cluster analysis also have provided applications in the area of climate research for mapping climate zones and clustering climatic variables (McBratney and Moore, 1985; Unal et al., 2003). However, hard clustering is often too restrictive due to the continuous nature of climate variables and limits further improvement (Fovell and Fovell, 1993). The method of fuzzy classification has been developed to provide a more realistic and flexible description of continuous variables (McBratney and Moore, 1985; McBratney et al., 1992), by allowing data to exist in more than one class rather than having abrupt class boundaries. Application of fuzzy techniques in the area of climate research has improved simulation and predictive systems (Bardossy et al., 2001). Fuzzy rules prediction systems have particularly been used for downscaling GCM's (Bardossy et al., 2001; Ozelkan et al., 1998; Wetterhall et al., 2006), but fuzzy classification has not been used extensively for classifying the local weather. The closest to this research in terms of classification is the work by Annas et al. (2007), who used a neuro-fuzzy classification method on past rainfall data to predict variations in daily tropical rainfall. Their work used a short rainfall series (8 years) and concentrated on predicting daily

1161

rainfall (Annas et al., 2007).

A coupled approach using both fuzzy classification and spatial interpolation of local weather data may improve previous studies in seasonal forecasting in this area. The concept behind this is that seasonal climate is often described as being similar to another point in space or time. Fuzzy classification of historical climate data will give the average seasonal climate for a specific time and location as a function of the past weather, set within the climate parameters. The resulting classified memberships may be correlated to the main climate drivers through a lag relationship to predict into the future. This will not attempt to predict the weather or a precipitation time series (e.g. Bardossy et al., 2001), but rather a seasonal outlook, where the season is assumed to be an intermediate stage between the weather and the climate.

This means that this work assumes that the future weather at a location is not a totally random event, but is a function of several components and relationships. The first of these are the past and present climate at the location, local moisture feedback and memory (Entekhabi et al., 1992; Seneviratne et al., 2006) and broader influences, such as Southern Oscillation Index (SOI) (Verdon and Franks, 2005), all of which are temporal components. In addition there are the spatial components related to the geographical location (Hunter and Meentemeyer, 2005). Weather and climate generally indicates trends in space, such as cross correlations in local rainfall, relations to elevation (Stahl et al., 2006) and the length of a dry spell. The final component is the random component. This method therefore, includes both the statistical characteristics in time and space and the physical drivers.

The objectives of this paper are:

- To use fuzzy classification as a means to cluster climate data over time and location.
- To use climate data classes to identify trends in seasonal variability.
- Determine whether class memberships can be used to predict the seasonal climate using correlations with climate drivers.

1162

2 Materials and methods

A summary of the methods used in this study is given in Fig. 1.

2.1 Data acquisition and manipulation

Climate data was acquired from the Bureau of Meteorology (BOM) for 107 weather stations from south-east Australia, with daily temperature and rainfall records from 1967 to 2006. The data consisted of the station name and number, location coordinates, maximum temperature, minimum temperature and rainfall.

A major problem when dealing with historical climate data sets is filling the missing values (Coulibaly and Evora, 2007). A considerable number of stations had missing data and the classification technique requires continuous data. Any stations with greater than 10% missing values were excluded from the data matrix. For the remaining 75 stations the following strategy was applied: if only a single point was missing the average of the surrounding values was used; if more than one value was missing in a sequence, the temperature data was filled using a cubic spline and the rainfall data was assumed to be zero.

2.2 Data classification

The climate data was classified using fuzzy k means algorithm, which requires a time series for each data point. Each weekly time point in the data matrix can be classified based on the values of the variables for a designated time period following that point (i.e. temperature, rainfall). Two matrices were created for the analysis, based on a 6 month weekly time series and a 1 year weekly time series. The rows of the resulting matrices were re-sampled at 2 weekly intervals for the 6 months time series and at monthly intervals for the 1 year time series to reduce the data sizes. An inverse correlation matrix was computed for each data matrix to determine the inter-correlation between weather stations.

1163

Program FuzME 3.0 (Fuzzy k -Means with Extragrades) (Minasny and McBratney, 2002) was used to partition a set of data into k classes so it minimises the distance within the cluster. The objective function is (Minasny and McBratney, 2002):

$$J(\mathbf{M}, C) = \sum_{i=1}^n \sum_{j=1}^k m_{ij}^{\phi} d_{ij}^2 \quad (1)$$

where ϕ is the fuzzy exponent determining the degree of fuzziness and d is the component of the distance matrix, calculated as:

$$d_{ij}^2 = (x_i - c_j)^T A (x_i - c_j) \quad (2)$$

where A is the distance norm matrix, here we use the inverse of variance-covariance matrix (Mahalanobis distance). The minimization of the objective function J provides the solution for membership matrix \mathbf{M} and the centroid C as follows:

$$m_{ij} = \frac{d_{ij}^{2/(\phi-1)}}{\sum_{j=1}^k d_{ij}^{2/(\phi-1)}} \quad i = 1, 2, \dots, n; j = 1, \dots, k \quad (3)$$

$$C_j = \frac{\sum_{i=1}^n m_{ij}^{\phi} x_i}{\sum_{i=1}^n m_{ij}^{\phi}} \quad j = 1, 2, \dots, k \quad (4)$$

The summary of the Modified Partition Entropies (MPE) and the Fuzzy Partitioning Index (FPI) was used to determine the best number of classes for clustering the data (McBratney and Moore, 1985). Each of the classes represents a centroid for temperature and rainfall across the classification period and shows the seasonal trends. For the selected number of classes, this produces a classification membership for each point in time over the 40 years for every station.

1164

2.3 Spatial and temporal data interpolation

The memberships for the unknown locations between weather stations were derived using spatial interpolation. The location of the 75 stations was projected onto the Albers projection system, a specialised coordinate system for large areas, using ArcMap 9.2 (ESRI, 2006). A 10-km grid was created to cover the study area, which contained 14 620 points. The program VESPER 1.6 (Variogram Estimation and Spatial Prediction plus ERror) (Minasny et al., 2002) was used to spatially interpolate the 75 station memberships using kriging to create 14 620 memberships on a grid. The interpolation of individual class memberships uses the method of ordinary kriging of log-ratio transformed memberships with a non-linear back transformation (McBratney et al., 1992). This method is appropriate for the interpolation of compositional data resulting from Fuzzy Classification and produces similar interpolations to Compositional Kriging (Walvoort and de Gruijter, 2001). The data were transformed using the symmetrical log-ratio transformation as:

$$t_i = \ln \left(\frac{m_i + \eta}{\left(\prod_{j=1}^k (m_j + \eta) \right)^{\frac{1}{k}}} \right) \quad i = 1, \dots, k \quad (5)$$

with inverse transformation:

$$m_i = \left(\frac{\exp(t_i)}{\sum_{j=1}^k \exp(t_j)} - \frac{\eta}{\sum_{j=1}^k \eta} \right) \left(1 + \sum_{j=1}^k \eta \right) \quad i = 1, \dots, k \quad (6)$$

where η is one half of the smallest membership other than zero (0.00005), m is the membership coefficient, t is the transformed membership coefficient and k is the num-

1165

ber of classes. Multiple interpolations over a time period were used to identify the spatial and temporal trends in the class memberships.

2.4 Model parameterisation and predictions

Seasonal prediction models were made using Regression Rules in CUBIST 2.04 (Rulequest Research, 2007) for the classification memberships for each station over the 40 year period. CUBIST constructs a regression tree, where the prediction contains linear regression models rather than discrete values. CUBIST creates a set of rules, where each rule has an associated multivariate linear model. Whenever a situation matches a rule's conditions, the associated model is used to calculate the predicted value.

Seasonal predictions were made for 1 month, 3 months and 6 months ahead for the observed average maximum temperature and total rainfall. To make a truly predictive model requires a physical explanation of why the observed climate variability in a given climate zone occurs. This requires an analysis of whether the identified season types are related to any large or small scale drivers of climate variability and their attributive indicators. The Southern Oscillation Index (SOI) and elevation parameters were included in the model to improve accuracy of predictions, as these are known to strongly influence the rainfall in south-eastern Australia (Hammer et al., 1996; Nicholls, 2004; Verdon and Franks, 2005). SOI data was obtained from the Bureau of Meteorology. Lag periods were created for SOI for 2, 4, 6, 8, 10 and 12 months. Regression rules models were conducted for the multiple predicting periods using SOI alone and SOI with lag periods. Elevation data was obtained from SRTM Digital Elevation Model (DEM), projected on the Albers projection system, for each station and each of the 14 620 points on the 10 km grid.

Predictions using the regression rules were conducted for the 6 month time series matrix (Model 1) and the 1 year time series matrix (Model 2). The regression models using station memberships and interpolated memberships were analysed for average maximum temperature and rainfall for the 3 predicting periods and for SOI and SOI lag periods. Predictions were made for stations and interpolated points for the best

1166

predicting models, selected based on the relative errors and correlation coefficients.

2.5 Model prediction accuracy

The predicted seasonal climate at the individual weather stations was tested against the real climate data for accuracy and reliability of the prediction. 70% of the data was used to train & build the regression model, while 30% was used for testing the accuracy of prediction. This form of model testing is known as hindcast skill. Values were calculated for the average absolute error and relative absolute error using the equations:

$$\text{Ave.Err.} = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \quad (7)$$

and

$$\text{Rel.Err.} = \frac{\text{Ave.Err.}}{\frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|} \quad (8)$$

where, y is the predicted value, x is the observed value, \bar{x} is the mean observed value and n is the number of measurements. Relative error is the ratio of the average error magnitude to the error magnitude that would result from always predicting the mean value. A useful model should be less than one and close to zero. A value of one indicates the model is simply predicting the mean value. Additionally, the model accuracy was compared against Rainman 4.3 (Clewett et al., 2003).

1167

3 Results

3.1 Fuzzy classification

An assessment of the FPI, MPE and the class distribution found 6 classes to provide the best description of the 6 month data set (McBratney and Moore, 1985). In contrast, 7 classes were found to be the best for the 1 year data set. The extra class in the 1 year time series classification probably reflects the additional information of covering a full set of seasons. The temperature centroids indicate clear seasonal trends (Fig. 2), which will be picked up in the station memberships over time. In contrast, the rainfall centroids indicate much less seasonality. The 1 year time series model was created to test whether a longer time series for the data classification would improve the climate class and produce greater accuracy in the seasonal prediction. A similar trend to the first model is evident, but the plotting period is twice as long (Fig. 2). However, for the 1 year time series, different rainfall classes do reflect such features as autumn dominated rainfall (class A), spring dominated rainfall (class G), and winter dominated rainfall (class B).

The memberships for the 75 stations were summed over the 40 year period (Fig. 3). These memberships identify the dominant class at a specific location over time. If the classification of historical climate is accurate the major climate zones present in the south-east of Australia should be identifiable. In principle three dominant zones in the station memberships can be identified (Fig. 3), which are a coastal zone and ranges with high rainfall, a far western dry zone and a southern zone with high rainfall and lower temperatures.

The total station memberships can only indicate the overall climate of a location. Plotting the memberships over time highlights the dominant classes, seasonal trends and periodic fluctuations or deviations from the usual pattern. The memberships for 2 stations are shown in Fig. 4. Station 59030 (South West Rocks) is located on the NSW north coast. Station 45017 (Thargomindah Post Office) is located in the south-west of Queensland.

1168

Station 59030 membership is dominated mostly by classes D and A representing the highest rainfall. There is a strong seasonal trend present in these classes. Remaining classes show periodical influences indicating significant climatic deviation from the norm. Station 45017 membership is dominated mostly by classes C, E and F representing low levels of rainfall or high temperatures. Seasonal trends can also be seen in this data. It would be expected that similar seasonal trends are present in the memberships for all stations, with varying dominance of the 6 classes.

3.1.1 Spatial interpolation and mapping

To display the station memberships graphically a spatial interpolation is required to obtain class memberships for all unknown locations. To illustrate the ability of the model to identify seasonal trends spatial interpolations at 2 month intervals for 2 years beginning in January of 1967 were derived, and projected to create digital images of memberships displayed over space and time (Fig. 5).

Each row in Fig. 5 shows the class memberships at a specific time period, and the areas of dark brown indicate class dominance. Locations with high rainfall can be identified, such as the north coast of NSW in January of 1967. Locations with low rainfall can also be identified, such as July of 1967 when almost the entire area was dry. Seasonal trends in temperature are also evident. During the summer months the majority of the area is class C dominant indicating warmer temperatures. During the winter months class B has more dominance, especially in areas with high elevation and southern regions, indicating lower temperatures. These spatiotemporal trends in the station memberships are the basis of the seasonal prediction.

3.2 Regression model parameterisation and seasonal prediction

The incorporation of SOI and elevation data significantly improved upon the original regression models containing only class memberships and station locations (Table 1). Model accuracy, as measured by the relative error, for the maximum temperature de-

1169

creased with greater prediction time, while accuracy for total rainfall increased with prediction time for the 6 month time series. This is probably due to the fact that the 6 month time series did not capture the full seasonality, as this decrease in accuracy is not visible when using the 1 year time series. Models with SOI lag period variables show improved accuracy on models relying on current SOI alone. In both cases the model fit is highly correlated. The SOI lag periods provide a greater improvement for rainfall models. This is explained by the greater influence SOI has on rainfall than on temperature (Nicholls, 2006).

Regression rules output from CUBIST also provided the attribute usage in the model and the prediction rules. A larger percentage of attribute usage indicates a greater ability for the attribute to describe climate variability and use in climate prediction. Northings, eastings and elevation attributes had the highest usage in all models. The usage for the class attributes was variable amongst the models. An analysis of the various SOI lag periods was unable to identify a dominant period in the model attribute usage.

The final predictive models for maximum temperature and rainfall used a 6 month predicting period with SOI lag variables included. The 6 month time series model for maximum temperature had a relative error of 0.42 and a correlation coefficient of 0.88 (Fig. 6), and the final model for rainfall produced a relative error of 0.54 and a correlation coefficient of 0.83 (Fig. 6). In contrast, the final 1 year time series model for maximum temperature had a relative error of 0.22 and a correlation coefficient of 0.97, while for rainfall the values were a relative error of 0.56 and correlation coefficient of 0.81 (Fig. 6 and Table 1) These regression models indicate that the use of memberships, SOI and elevation variables have a strong ability to predict seasonal climate variability. Overall, the memberships provide a good fit for the predicted temperature and rainfall in the models.

The final four models for the 6 month and 1 year time series were used to create a prediction for the current seasonal climate. Predictions were made based on the station memberships at the end of 2006 and the interpolated memberships at the end

1170

of 2006. Predictions of interpolated points were used to create a spatial map of the predicted seasonal climate for the beginning of 2007 (Fig. 7). Predictions at individual stations were used to test model accuracy (Fig. 8).

The temperature and rainfall predictions show a high degree of seasonal climate variability across the south-east of Australia. There is evidence in the predicted maps that the regression rules model has produced some inconsistent artefacts in the prediction. Firstly, there is a band of seasonal climate in South Australia over the Flinders Ranges that deviates highly from the surrounding temperature and rainfall (Fig. 7). The prediction indicates an area of low temperature and high rainfall, which is not consistent with the real data. Lower temperatures would be expected at the higher elevations, however the real increase in rainfall over the Flinders Ranges is only minimal and the prediction is unrealistic. This is probably due to the lack of data on rainfall at high elevations, which will therefore decrease the accuracy of the elevation predictor (Stahl et al., 2006). Secondly, there are several sharp divides between bands of temperature and rainfall values. These are present as horizontal or vertical lines in the prediction maps. It is likely this is due to the influence of location coordinates in areas where there are too few stations in the prediction. Therefore, the prediction could be improved with the use of larger numbers of weather stations. A visual comparison of the predicted and recorded temperature and rainfall maps (Fig. 8) identifies a strong correlation in the spatial trends. The temperature prediction appears to have picked up the important influences on spatial variation for the seasonal forecast. However, the model over predicts the magnitude of temperature over the whole prediction zone. The rainfall model has predicted the recorded data well, with almost all spatial trends accounted for. However, the recorded rainfall map shows a small zone of unusually high rainfall on the east coast of NSW, to the north of Sydney, which the model did not pick up. Such events are extremely difficult to predict.

1171

3.3 Model prediction accuracy

The locational accuracy of the predicted climate data was tested for both the 6 month time series and 1 year time series models using the average error, relative error and correlation coefficient (Table 2). The temperature prediction for Model 1 has a relative error of 0.89 and Model 2 has a relative error of 1.07. Therefore, both models have a high degree of error, with Model 2 doing slightly worse than representing the mean. However, given that both models for the temperature prediction had strong correlation coefficients, it indicates the seasonal temperature trends are predicted correctly, but there is a degree of error in the exact spatial prediction. Model 1 rainfall prediction has a relative error of 0.56 and Model 2 rainfall prediction has a relative error of 0.64. These errors indicate a much higher degree of accuracy in the seasonal climate forecast. Model 1 and Model 2 both produced strong correlation coefficients of 0.82 and 0.72 respectively, meaning the forecast has accurately accounted for a large amount of the seasonal trends and variability. Based on these results, Model 1 produced a higher degree of accuracy in the seasonal climate forecast than Model 2.

4 Discussion

As described, Model 1 used a 6 month time series, with 40 years of data subset at 2 weekly intervals. Model 2 used a 1 year time series, with 40 years of data subset at monthly intervals. Therefore, Model 1 was actually using only half of the original weekly data set and Model 2 was using only a quarter of the original weekly data set. The added temperature and rainfall information in the density of the series used in Model 1 explains the difference in predictive capabilities between the models. The reason for reducing the data set was due to computer power limitations when dealing with large data sets. Single classifications take up several weeks making multiple classifications extremely time consuming.

It is interesting to note that the high correlations in the regression rules models build

1172

(i.e. Table 1), do not necessarily translate in better predictions. The 1 year time series had better correlations and lower relative errors in the model building, but is less effective in predictive mode. An additional reason could be that the current weather is more autocorrelated to the weather over a 6 month period than over a 1 year period.

5 Testing the skill of forecast systems generally uses the hindcast test. The current seasonal climate prediction, produced using the proposed model, was tested against a forecast for the same season produced using Rainman. Rainman (Clewett et al., 2003) is a commercially available statistical seasonal climate forecasting software package. Rainman creates seasonal forecasts for specific weather station sites based on his-
10 torical data and the influence of SST and SOI phases. The test compared a random sample of 8 weather stations for January to March 2007. Rainman only has the ability to produce rainfall forecasts for a particular weather station for 3 months in advance (Table 3).

The Rainman prediction gives the chance of receiving a certain amount of rainfall
15 using a 70% probability. For the sample weather stations Model 2 predicts with the least accuracy. Model 1 and Rainman both produced good correlations with the actual data, indicating the predictions have replicated the seasonal trends. Model 1 and Rainman produced relative errors of 0.49 and 0.6 respectively, again indicating Model 1 is predicting with a greater level of accuracy. Rainman's model accuracy is highly re-
20 liant on the phases of SOI and SST (Clewett et al., 2003). For the current prediction Rainman's statistical tests indicated there was no significant relationship with the SOI phases, resulting in a reduction in the prediction accuracy. The proposed model provides advantages by placing more reliance on the current climate state at a specific location, and thus including both land use feedbacks and the physical drivers. There-
25 fore, the seasonal climate forecasts provide greater accuracy over a model such as Rainman when the SOI phase is not significantly influencing rainfall and temperature across the south-east of Australia. An expansion of the model suggested in this research could include other climatic drivers such as the IPO, DMI and IDO (Verdon and Franks, 2005).

1173

The advantage of the statistical forecasting method proposed in this study, is the balance between local weather station forecasts and more continental GCMs. Fuzzy Classification effectively models climate over space and time and describes the sea-
5 sonal trends and cross-correlations between known weather stations. It predicts over a large area, while maintaining a fine local spatial prediction scale, in this case 10 km. This is satisfactory for most seasonal climate forecast applications, as seasonal climate variability below this scale is minimal. This work thus starts from a different end compared to earlier work which focused on fuzzy classification of the climate drivers to predict daily rainfall time series (Ozelkan et al., 1998). A combination of this work
10 with such classifications is a further research opportunity. It also includes within its spatial predictions the seasonal variability (through the fuzzy classification), which was missing in earlier work on interpolation (Hunter and Meentemeyer, 2005; Stahl et al., 2006).

A major limitation of this method is the number of weather stations available across
15 the study area with long and complete historical records of temperature and rainfall, and this is noted by all authors working on interpolating climate data (e.g. Stahl et al., 2006). Spatial interpolation, when used appropriately, is a useful technique for analysing geophysical data obtained from point sources. However, there remains a degree of error in the interpolated values, especially with greater distance between point
20 sources. Increasing the number of weather stations used in the analysis decreases the dependence on the spatial interpolation method and would improve the accuracy of the final seasonal prediction. However, increasing the amount of data increases the demand on computing power, which was a limitation in the proposed model in the study, but with current advances in computing power might not be a major limitation.

25 To make this forecasting method useful requires the implementation of a system to conduct periodic forecasts for the prediction period specified. For example, a system may be developed to make updated forecasts every 2 weeks. This system would incorporate 3 modules. The Fuzzy Classification module already contains the clustered data for 40 years of rainfall and temperature, with the specified number of classes and

1174

class centroids. This module could be updated with the current data and reclassified to create new memberships. This process would not be overly time consuming as only a few iterations would be required to recalculate the class centroids and produce new memberships. A second module would be required to create a spatial interpolation of the most recent weather station memberships. The regression rules module would only be required to be updated on a 3 to 6 monthly basis, as short periods of data only have a small influence on the overall regression model. Therefore, the current interpolated memberships can be applied directly to the regression rules model to obtain the predicted rainfall and temperature values for up to 6 months in advance. An automated system could be developed to combine these 3 modules and produce seasonal climate forecasts on a regular basis.

5 Conclusion

The proposed method used fuzzy classification to create memberships of climate data over space and time. These memberships provide the basis for the seasonal climate forecast and are the only requirement for predicting rainfall and temperature. The inclusion of location and elevation variables in the model allows for spatial predictive abilities and improves on prediction accuracy. The inclusion of SOI in the model provides a relationship between the continental climate driver and the observed climate variability. SOI has no spatial application in the model and only improves the overall forecast accuracy. The proposed model provides an advantage over current seasonal climate forecast models through its spatially explicit predictive abilities, and improves on earlier spatial interpolation by including seasonality.

This seasonal climate forecasting model presented here is a prototype and hence developmental. Further research required for this model includes: (1) Developing and testing forecast models for predicting 1 and 3 months in advance and conducting hind-cast tests for a large sample of forecasts for the 1 month, 3 months and 6 months prediction models to accurately measure predictive skill; (2) Developing a model based

1175

on daily rainfall and temperature using the entire 40 years of data including further climatic drivers; (3) Developing an automated system that could conduct seasonal climate forecasts on a regular basis. A simple and user friendly forecasting model would provide benefits for a diverse range of applications; (4) Determining the models applications in detecting and attributing climate change and enhanced climate variability, with focus on moisture recycling and feedback.

Acknowledgements. The authors would like to thank the Australian Bureau of Meteorology for supplying rainfall and temperature data. We would like to thank N. Odgers for assistance with digital elevation modelling and ArcMap procedures.

10 References

- Annas, S., Kanai, T., and Koyama, S.: Assessing daily tropical rainfall variations using neuro-fuzzy classification model, *ecological informatics*, 2, 159–166, 2007.
- Bardossy, A., Stehlik, J., and Caspary, H. J.: Generating of areal precipitation series in the upper neckar catchment, *Phys. Chem. Earth, Part B: Hydrology, Oceans and Atmosphere*, 26, 683–687, 2001.
- Cane, M. A.: The evolution of el nino, past and future, *Earth Planet. Sci. Lett.*, 230, 227–240, 2005.
- Clewett, J. F., Smith, P. G., Partridge, I. J., George, D. A., and Peacock, A.: Rainman streamflow version 4.3: A comprehensive climate and streamflow analysis package on CD to assess, seasonal forecasts and manage climate risk, in: Department of Primary Industries, Queensland, Toowoomba, Queensland, Australia, 2003.
- Coulibaly, P. and Evora, N. D.: Comparison of neural network methods for infilling missing daily weather records, *J. Hydrol.*, 341, 27–41, 2007.
- Entekhabi, D., Rodriguez-Iturbe, I., and Bras, R. L.: Variability in large-scale water balance with land surface-atmosphere interaction, *J. Climate*, 5, 798–813, 1992.
- ESRI: Arcmap 9.2, Environmental Systems Research Institute Inc, Redlands, CA, USA, 2006.
- Fovell, R. G. and Fovell, M. Y. C.: Climate zones of the conterminous united-states defined using cluster-analysis, *J. Climate*, 6, 2103–2135, 1993.

1176

- Goovaerts, P.: Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall, *J. Hydrol.*, 228, 113–129, 2000.
- Hammer, G. L., Holzworth, D. P., and Stone, R.: The value of skill in seasonal climate forecasting to wheat crop management in a region with high climatic variability, *Australian J. Agric. Res.*, 47, 717–737, 1996.
- Hansen, J. W. and Sivakumar, M. V. K.: Advances in applying climate prediction to agriculture, *Clim. Res.*, 33, 1–2, 2006.
- Hunter, R. D. and Meentemeyer, R. K.: Climatologically aided mapping of daily precipitation and temperature, *J. Appl. Meteorol.*, 44, 1501–1510, 2005.
- McBratney, A. B. and Moore, A. W.: Application of fuzzy-sets to climatic classification, *Agric. For. Meteorol.*, 35, 165–185, 1985.
- McBratney, A. B. and DeGrujter, J. J.: A continuum approach to soil classification by modified fuzzy k-means with extragrades, *J. Soil Sci.*, 43, 159–175, 1992.
- McBratney, A. B., DeGrujter, J. J., and Brus, D. J.: Spatial prediction and mapping of continuous soil classes, *Geoderma*, 54, 39–64, 1992.
- Meinke, H. and Stone, R.: Seasonal and inter-annual climate forecasting: The new tool for increasing preparedness to climate variability and change in agricultural planning and operations, *Climatic Change*, 70, 221–253, 2005.
- Minasny, B. and McBratney, A. B.: Fuzme version 3.0, Australian Centre for Precision Agriculture, in: The University of Sydney, Australia, <http://www.usyd.edu.au/su/agric/acpa>, 2002.
- Minasny, B., McBratney, A. B., and Whelan, B. M.: Vesper version 1.6., Australian Centre for Precision Agriculture, The University of Sydney, Australia, 2002.
- Nicholls, N.: The changing nature of Australian droughts, *Clim. Change*, 63, 323–336, 2004.
- Nicholls, N.: Detecting and attributing Australian climate change: A review, *Australian Meteorological Magazine*, 55, 199–211, 2006.
- Ozelkan, E. C., Galambosi, A., Duckstein, L., and Bardossy, A.: A multi-objective fuzzy classification of large scale atmospheric circulation patterns for precipitation modeling, *Appl. Math. Comput.*, 91, 127–142, 1998.
- Rulequest Research: Cubist version 2.04, Rulequest Research Pty Ltd, St. Ives, NSW, Australia, 2007.
- Seneviratne, S. I., Luthi, D., Litschi, M., and Schar, C.: Land-atmosphere coupling and climate change in Europe, *Nature*, 443, 205–209, 2006.
- Stahl, K., Moore, R. D., Floyer, J. A., Asplin, M. G., and McKendry, I. G.: Comparison of

- approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density, *Agric. For. Meteorol.*, 139, 224–236, 2006.
- Unal, Y., Kindap, T., and Karaca, M.: Redefining the climate zones of Turkey using cluster analysis, *Int. J. Climatol.*, 23, 1045–1055, 2003.
- Verdon, D. C. and Franks, S. W.: Indian ocean sea surface temperature variability and winter rainfall: Eastern Australia, *Water Resour. Res.*, 41, W09413, doi:10.1029/2004WR003845, 2005.
- Walvoort, D. J. J. and de Grujter, J. J.: Compositional kriging: A spatial interpolation method for compositional data, *Math. Geol.*, V33, 951–966, 2001.
- Wetterhall, F., Bardossy, A., Chen, D., Halldin, S., and Xu, C.-Y.: Daily precipitation-downscaling techniques in three Chinese regions, *Water Resour. Res.*, 42, W11423, doi:10.1029/2005WR004573, 2006.

Table 1. Regression rules results for the classification for 6 month time series and 6 classes and the 1 year time series and 7 classes.

Predict	Months	Current SOI			SOI lag periods		
		Average Error	Relative Error	Corr. Coeff.	Average Error	Relative Error	Corr. Coeff.
Regression Rules Model Results for 6 Classes, 6 month							
Maximum Temperature (°C)	1	2.05	0.33	0.93	1.99	0.32	0.93
	3	1.95	0.35	0.92	1.89	0.34	0.92
	6	1.68	0.43	0.86	1.62	0.42	0.88
Total Rainfall (mm)	1	25.90	0.8	0.55	24.32	0.75	0.62
	3	50.09	0.69	0.71	45.95	0.63	0.76
	6	74.86	0.6	0.79	67.63	0.54	0.83
Regression Tree Model Results for 7 Classes, 1 year							
Maximum Temperature (°C)	1	1.596	0.27	0.95	1.547	0.26	0.95
	3	1.261	0.23	0.96	1.201	0.22	0.97
	6	0.961	0.23	0.97	0.921	0.22	0.97
Total Rainfall (mm)	1	26.055	0.79	0.56	24.921	0.76	0.6
	3	50.644	0.68	0.7	47.577	0.64	0.75
	6	74.130	0.58	0.8	70.483	0.56	0.81

1179

Table 2. Accuracy coefficients for testing model accuracy of the final seasonal climate predictions.

Model	Model Prediction	Accuracy		
		Average Error	Relative Error	Correlation Coefficient
Model 1: 6 month weekly time series	Maximum Temperature (°C)	2.12	0.89	0.83
	Rainfall (mm)	78.60	0.56	0.82
Model 2: 1 year weekly time series	Average Maximum Temperature (°C)	2.03	1.07	0.82
	Rainfall (mm)	89.60	0.64	0.72

1180

Table 3. Three month current seasonal predictions for the two proposed models and Rainman compared against the actual data. Relative errors and correlation coefficients between the model and the actual data are given.

Comparison of Proposed Models and Rainman Predictions				
Station	Model 1 Predicted Rainfall	Model 2 Predicted Rainfall	Rainman (based on 70% probability)	Actual Rainfall
Marree Composite	51	38	23	58
Tibooburra Post Office	113	43	60	152
Cobar MO	114	60	50	42
Gunnedah Composite	182	121	130	192
Sydney Airport AMO	266	207	240	223
Canberra Airport	139	105	120	141
Mildura Airport	64	53	32	75
Rutherglen Research	112	127	70	68
Accuracy Test				
Relative Error	0.49	0.75	0.60	
Correlation Coefficient	0.85	0.68	0.86	

All rainfall units are in millimetres (mm).

1181

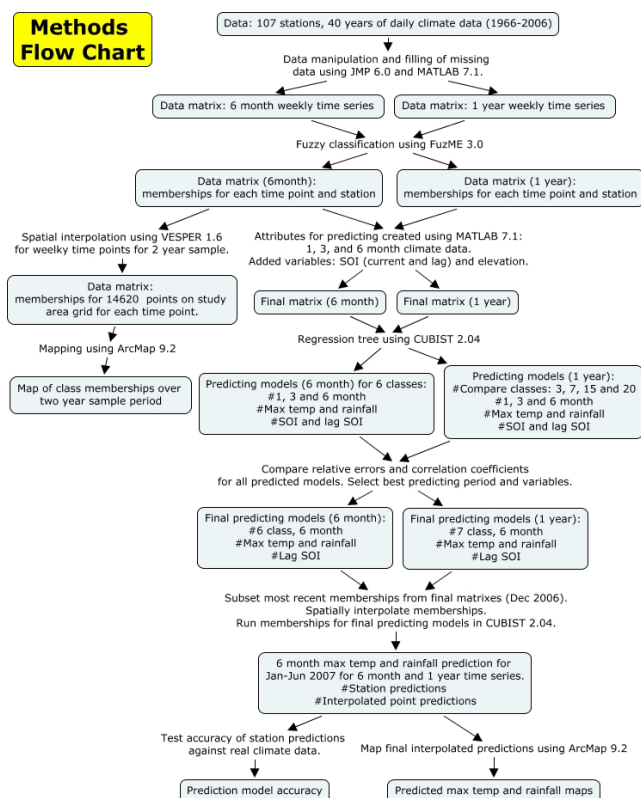


Fig. 1. Flow chart of the methods used in formulating a seasonal forecasting model. Boxes describe the products at each step and the joining text describes the processes used.

1182

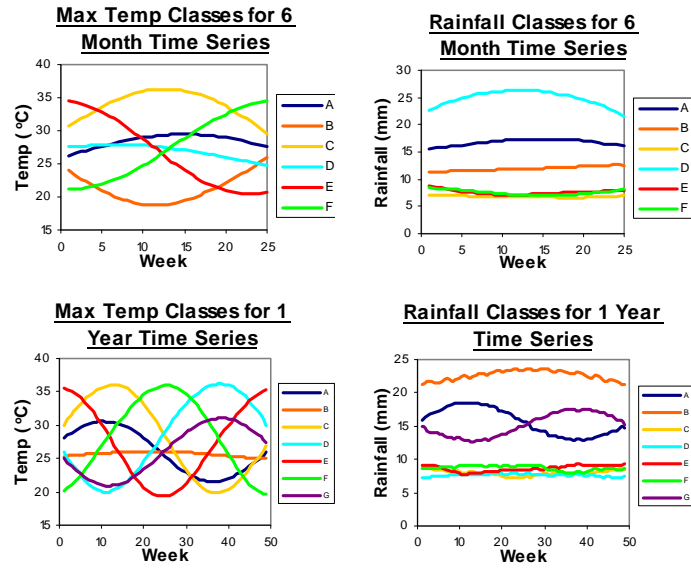


Fig. 2. Maximum temperature and rainfall classes obtained from the classification using a 6 month time series (top) and a 1 year time series (bottom). A ,B, C,...G represent the centroids of the fuzzy classes.

1183

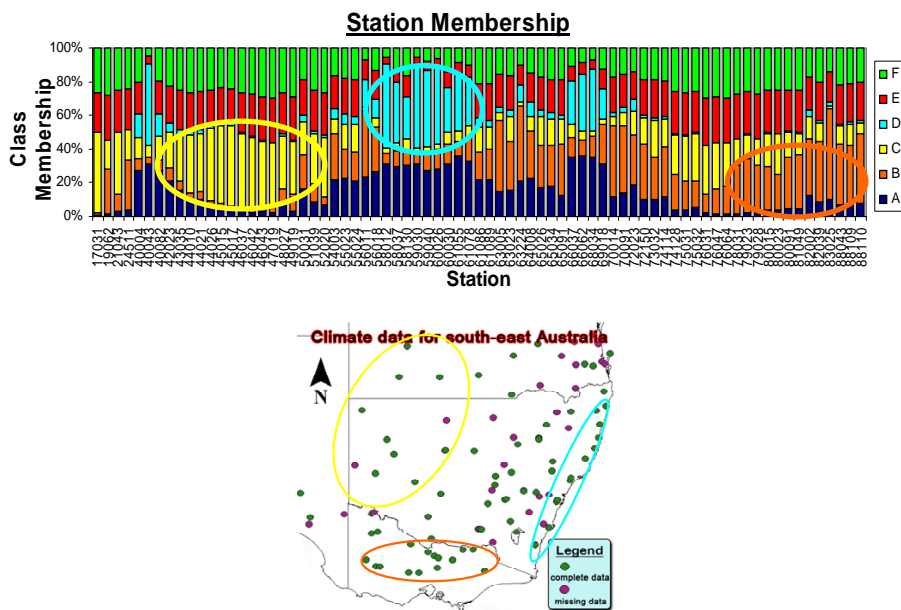


Fig. 3. (a) Total station memberships for 40 years of climate data based on the classification using a 6 month weekly time series and for 6 classes. Three major climatic zones are identified in the memberships. **(b)** Map of the location of weather stations showing 3 major climatic zones as determined by the total station memberships.

1184

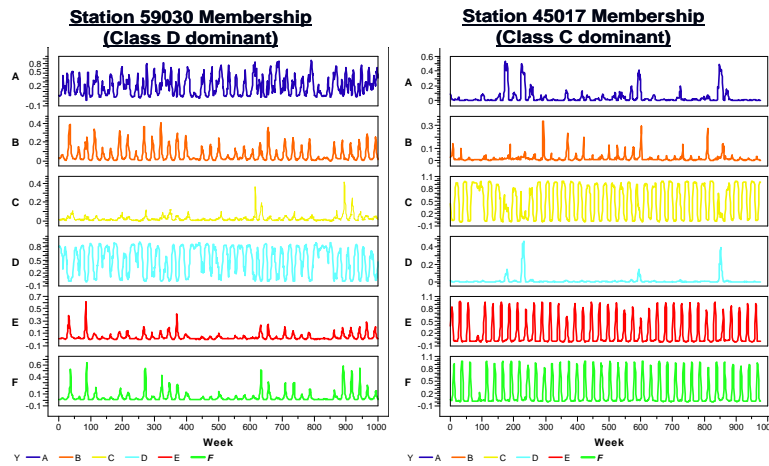


Fig. 4. Individual station memberships for 40 years climate data based on classification using a 6 month weekly time series. Example of a station with class D dominant (high rainfall), and a station with class C dominant (low rainfall). Class memberships are plotted separately showing trends over time.

1185

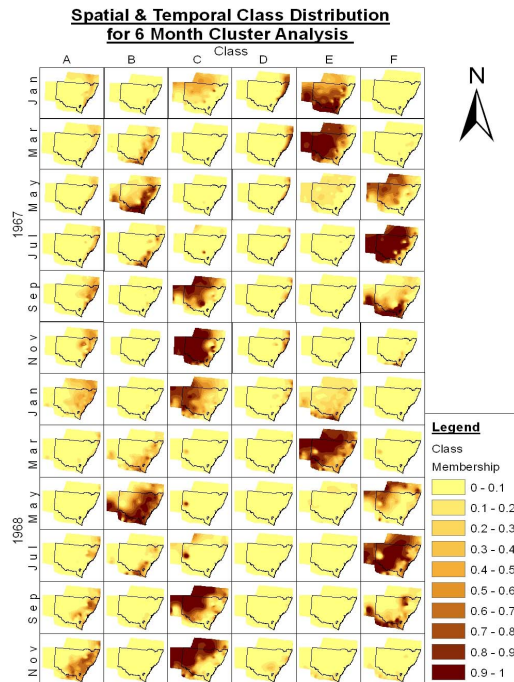


Fig. 5. Spatially interpolated class memberships for the cluster analysis using 6 classes (A, B, C...F). Classes are shown on the horizontal axis with time on the vertical axis. The map is geographically orientated to the Albers projections system with the NSW boundaries displayed.

1186

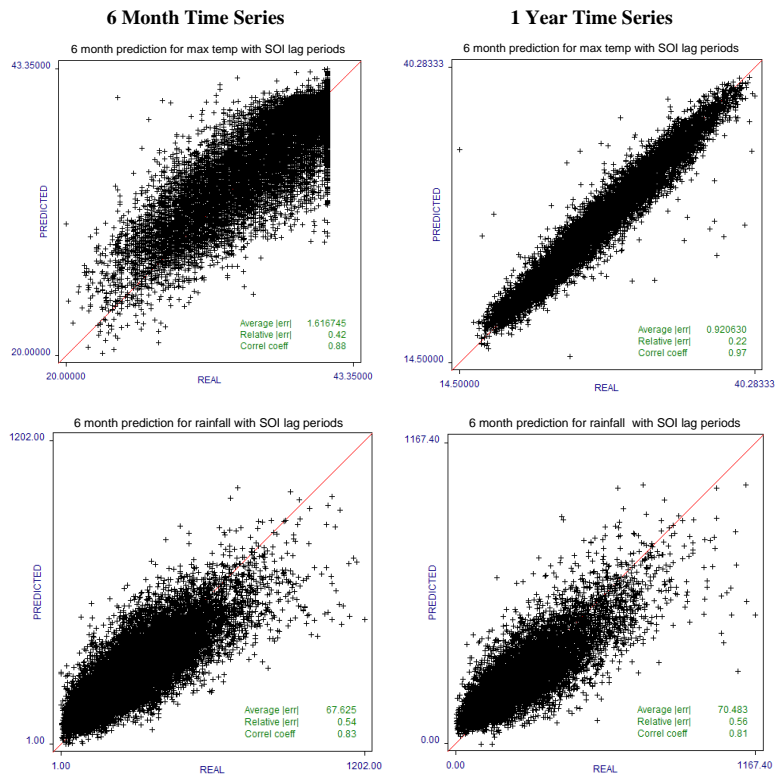


Fig. 6. Model 1 scatter plots for regression rules models predicting 6 months maximum temperature and rainfall from station memberships and using SOI lag periods for 6 month time series (left panels) and for 1 year time series (right panels).

1187

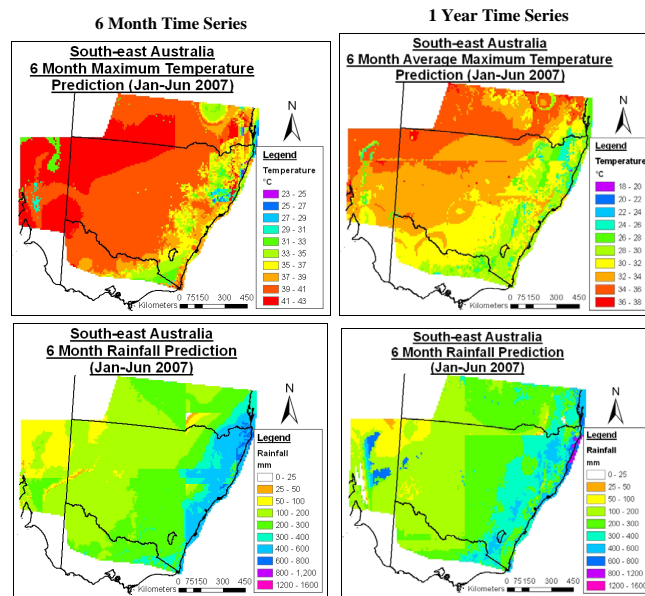


Fig. 7. Six month maximum temperature and rainfall prediction maps for January to June 2007. Prediction based on classification using 6 month time series (left panels) and 1 year time series (right panels).

1188

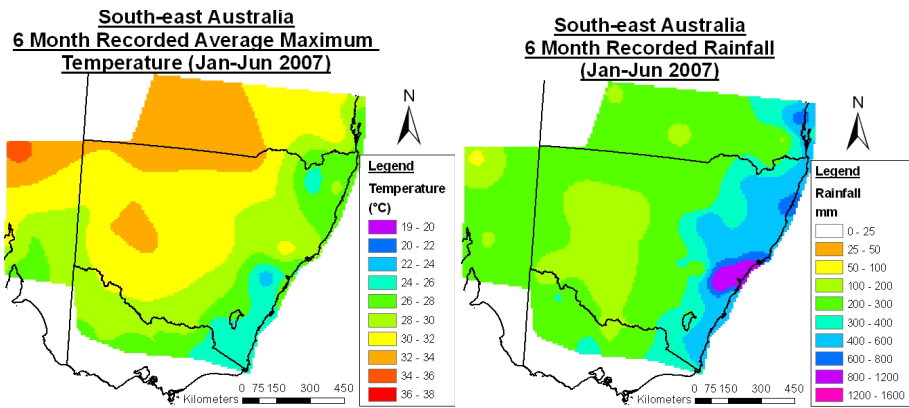


Fig. 8. Six months recorded maximum average temperature and rainfall maps for January to June 2007. Data obtained from the Bureau of Meteorology.