18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

The use of TRMM satellite data as a predictor in the spatial interpolation of daily precipitation over Australia

Oke, A.M.C.[1], **Frost, A. J.** [1], **Beesley, C. A.**[1]

[1]*Water Division, Australian Bureau of Meteorology*
Email: a.oke@bom.gov.au

Abstract: The accurate estimation of gridded daily precipitation is critical to hydrological modelling and water resource assessment. High resolution precipitation datasets based on gauge rainfall are the primary input to spatially distributed rainfall-runoff models and water balance calculations (e.g. Chiew et al., 2008). The spatial heterogeneity of rainfall variability is currently not captured adequately by gauge based interpolation methods. The errors in gridded rainfall fields have the potential to significantly bias model calibrations and water balances calculations. Sources of data other than gauge rainfall, such as satellite derived fields, radar based rainfall observations, and climatological fields from numerical weather prediction models, can be used as a predictor to improve interpolated rainfall fields. However, to date mainly gauge based interpolation of observations are used for modelling purposes because satellite/model based observed precipitation surfaces have relatively high errors in the rainfall distribution to which runoff is very sensitive.. To the authors' knowledge no assimilation of observed and modelled precipitation data (gauged or from satellite) has been undertaken for the Australian continent. This paper investigates the additional value of the inclusion of the Tropical Rainfall Measuring Mission (TRMM) 3B42 satellite based precipitation product to gauge based precipitation estimation.

Three simple geostatistical methods are applied to investigate the use of satellite based rainfall in gauge based interpolation: a) ordinary kriging (OK) – used as the baseline gauge based method; and b) cokriging (CK) and c) simple kriging with locally varying mean (SKlm) for incorporating additional information. The three methods are compared using cross validation statistics (including mean error, mean absolute error and root mean squared error).

Incorporating satellite based rainfall estimates into the interpolation of daily rainfall does not increase the overall accuracy. However, in data sparse areas some increase in accuracy is observed. The poorer performance of the methods incorporating satellite data is attributed to the relationship between satellite derived rainfall and gauge rainfall which is highly variable (Figure 1). In particular, strong spatially consistent negative biases are found for coastal regions (greater the 0.5mm daily) and strong positive biases for high altitude regions (greater than 0.5mm daily). These biases and methods to deal with them should be considered in future research.

Keywords: *Daily rainfall, kriging, cross validation, interpolation, satellite.*

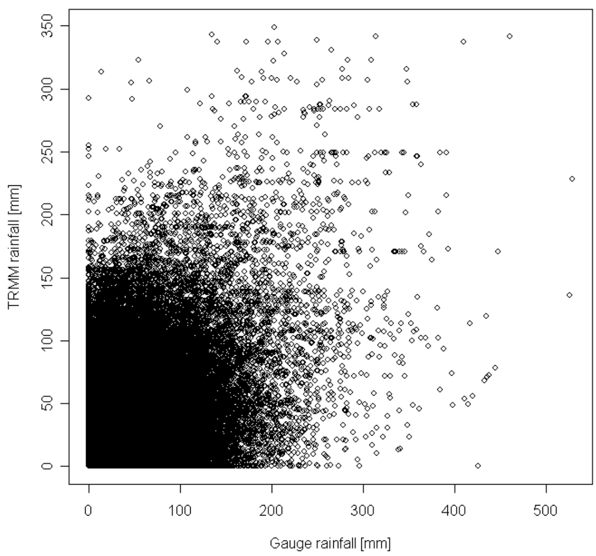

Figure 1. **Observed gauge versus TRMM daily rainfall**

1. INTRODUCTION

Two products are currently publicly available that contain archived Australia-wide gridded (0.05° by 0.05°) gauge-based daily rainfall: the SILO product (www.longpaddock.qld.gov.au/silo) produced by the Queensland Environmental Protection Agency; and the product produced by the Bureau of Meteorology (BoM) as part of the Australian Water Availability Project (www.bom.gov.au/jsp/awap)(BAWAP). Beesley et al. (2009) compared these products using cross validation statistics and found them to be similar. In both data sets, high altitude runoff generating areas are negatively biased and neither capture the North-South gradient of variability, echoing the cross validation results found for SILO (Jeffrey *et al.*, 2001) and BAWAP (Jones et al., 2007). The current methods are not designed for the skewed zero bounded nature of daily rainfall and do not adequately account for altitudinal and location effects. Moreover, the spatial heterogeneity of rainfall distribution across Australia is not adequately accounted for. This paper investigates an approach to potentially improve spatially interpolated surfaces through the incorporation of other covariates.

Several alternative data sources to gauge based precipitation measurements are available that offer opportunities to enhance the quality of rainfall surfaces. Gridded atmospheric data and terrain characteristics can be used as covariates in the spatial prediction of precipitation to provide extra information and spatial detail (Hutchinson, 1998; Kyriakidis et al., 2001). Remotely sensed/satellite based observations offer a potentially wide range of observational fields that can improve the gauge based interpolated surfaces. Indeed, algorithms exist for the derivation of gridded precipitation rainfall fields from satellite observations. Ebert et al. (2007) developed a real-time validation system for several satellite derived gridded rainfall surfaces across Australia http://www.bom.gov.au/bmrc/SatRainVal/validation-intercomparison.html. Renzullo (2008) has constructed daily precipitation surfaces for Australia from the post real-time TRMM 3B42 rainfall product. This product has been used in streamflow and flood modelling studies where real-time gauge data is sparse (eg. Hazarika et al. (2007); Su et al. (2008)). However, to date satellite based rainfall has not been combined with gauge based rainfall estimation on a national scale. This paper describes a simple first attempt at using satellite based precipitation as an additional source of data through the application of kriging techniques.

2. DATA AND METHODOLOGY

2.1. Data

The BoM's daily rainfall archive contains over 17,000 stations and extends back prior to 1900 for many sites. This study focuses on the approximately 6,400 stations operational at some point during 2001-2007. To enable comparison with the existing methods of SILO and BAWAP the stations are restricted to those used in the cross validation of these methods in Beesley et al. (2009). The TRMM satellite-derived daily precipitation estimates covering Australia were constructed from the post real-time TRMM Multi-Satellite Precipitation Analysis (TMPA) product, 3B42 (Huffman et al., 2007). These data are instantaneous precipitation rates at 0.25° x 0.25° degree grid resolution time-stamped at 3-hourly intervals. The daily rainfall is produced based on local time and represents an accumulation to 9 am produced by temporally interpolating between the three hourly rate measurements (Renzullo, 2008). Twenty-four hour (9am – 9am) accumulation periods were used separately for each area of Australia under differing time-zones such that the satellite daily rainfall matches the period over which the gauge rainfall was measured (Renzullo, 2008).

2.2. Geostatistical Methods

Kriging forms the basis of the three geostatistical methods trialled to improve gridded rainfall interpolation. Kriging estimates values at ungauged sites based on known values at georeferenced locations, weighted according to a variogram model. The parameters of the variogram model can be estimated or assumed. Kriging is based on the concept that sampled values at closely spaced locations are more similar than those further apart (Isaaks and Srivastava, 1989). A variogram represents this spatial continuity by describing variation within a dataset as a function of the distance separating the samples within it. A variogram model describes the line of best fit to the variogram points.

The first method applied within this study, ordinary kriging (OK), uses only the gauge data to interpolate the rainfall surface. It is the most common form of kriging where the weights are determined from variogram model through minimizing the estimation variance but ensuring the unbiasedness of the estimator (Goovaerts, 2000). This method assumes that the field is stationary and has a constant unknown mean, although reasonable results can typically be achieved even when these assumptions are violated (Cressie, 1993). Cokriging (CK) is the multivariate extension of ordinary kriging that uses a cross-semivariogram in weighting the interpolation accounting for dependence between variables. It minimizes the variance of the estimation error by exploiting the cross-correlation between several variables (Isaaks and Srivastava, 1989). To take advantage of the spatial coverage of the TRMM data the full grid was used instead of just the collocated points. The use of the full grid requires the production of a non co-located cross-semivariogram where the sill and the nugget at zero forms the covariance between the two variables.

SKlm essentially replaces the unknown stationary mean in the OK system with a known varying mean derived from a covariate (Goovaerts, 2000). The local means are defined by the relation between the gauge and TRMM rainfall records. Collocated points are extracted at the gauge station sites to define the linear relationship between the rain gauge data and the TRMM. This relationship is used to predict precipitation at each of the 0.05° grid surface points. A variogram is then calculated for the station point residuals and used along with the regression model to generate a daily rainfall surface. A constant linear relationship between the gauge and TRMM rainfall data is defined based on the regression equation generated for the full period of 2001-2007 with all stations. The linear modelling yielded a regression coefficient of 0.6542772 and intercept of 0.213172. It is the definition of this global regression model that distinguishes this method from Kriging with External Drift where the regression coefficients vary with time. The global regression equation is defined to avoid problems associated with large areas of no rainfall disrupting the method when run with a defined neighbourhood. A defined neighbourhood of influence on the variogram estimation is used to reduce processing time but it can result in the application of the regression analysis to areas where the covariate is constant (0 mm) and therefore no longer independent from the intercept.

All analyses were completed using the freeware statistical package R (R Development Core Team, 2008) using the gstat package (Pebesma, 2004) for the spatial modelling. For all three kriging techniques an exponential model was fitted to the transformed rainfall data for each day using a weighted least squares fit to the experimental variogram - consequently the variogram model parameters vary over time. Precipitation (gauge and TRMM) data were transformed using a square root , as is typically required for the appropriate application of such techniques to rainfall data to produce approximately constant variability in residuals for all levels of rainfall (Hutchinson, 1998).

2.3. Cross validation

The different geostatistical interpolation methods are compared in terms of daily cross validation statistics and the spatial characteristics of the cross-validation errors. Cross validation refers to the repeated process by which one or more observations are omitted from the analysis and prediction. The difference between the predicted value and the observed value at each location is used to assess the accuracy of the interpolation. In this analysis 10% of the data was omitted in each repetition (also known as 10-fold cross validation). The cross validation statistics calculated are mean error (ME - also referred to as bias), mean absolute error (MAE) and root mean square error (RMSE). These statistics are calculated here according to the following equations:

$$ME = \frac{1}{n}\sum_{i=1}^{n}(E_i - O_i) \quad (1)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|(E_i - O_i)\right| \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(E_i - O_i)^2} \quad (3)$$

Table 1. Cross validation statistics for the period of 2001-07.

Method	ME (mm)	MAE (mm)	RMSE (mm)
OK stations only	-0.26	0.73	3.33
CK with TRMM	-0.26	0.78	3.46
SKlm with TRMM	-0.19	0.86	3.59

where E_i is the interpolated (estimated) value at a station on a particular day, O_i is the observed and n is the total number of records in the analysis.

3. RESULTS AND DISCUSSION

The summary error statistics are presented in Table 1 and represent the cross validation of over 16 million records in each method run for the 2001-2007 analysis period. The errors in the three methods are very similar despite the inclusion of the TRMM data in the CK and SKlm. The consistently negative ME indicates that all three methods underestimate the gauge rainfall in their prediction. A negative bias is a common feature of weighted average interpolation techniques as the interpolated value can never be greater than the surrounding observed values. However, this contradicts values recorded for previous gauge based cross validation analysis of SILO and BAWAP rainfall series (Jeffrey et al., 2001, Jones et al., 2007, Beesley et al. 2009) which find that the mean error bias values are approximately zero. These studies report zero overall bias because for small/large rainfall events the model is positively/negatively biased (Beesley et al., 2009). However, as a square root transformation is used here, the positive biases for small events are reduced relatively to the negative biases for large rainfall events. OK has slightly higher overall accuracy according to the MAE and RMSE summary statistics (Table 1). Of the two methods that use the TRMM in the interpolation, CK is more accurate. The error statistics are slightly greater than those calculated for the cross validation supplied for SILO in Beesley et al. (2009), and approximately the same as those for the BAWAP methodology, although the studies are not strictly comparable as differing subsets of data were used.

The error level in the daily rainfall interpolation techniques is not constant in time due to the heterogeneity of rainfall event type in space and time. Figure 2 presents the number of records for each day and the temporal variation in the national daily ME, and RMSE statistics for the three methodologies. It clearly shows that the error in all techniques is greater in the southern hemisphere summer, associated with a higher prevalence of localized convective events (Figure 2). The smaller error in winter is primarily due to the reduced daily totals and therefore reduced potential for large errors. The number of records available for analysis on each day reduces from over 6600 in 2004 to less than 6200 in 2007. The increased errors later in the record are partly due to the reduction in the gauge network. This point emphasizes the need to incorporate other measures of precipitation estimation (e.g. satellite and radar). It also highlights the benefits of the Water Act (2007) regulations in compelling water organizations to supply any observed rainfall records to the Bureau of Meteorology.

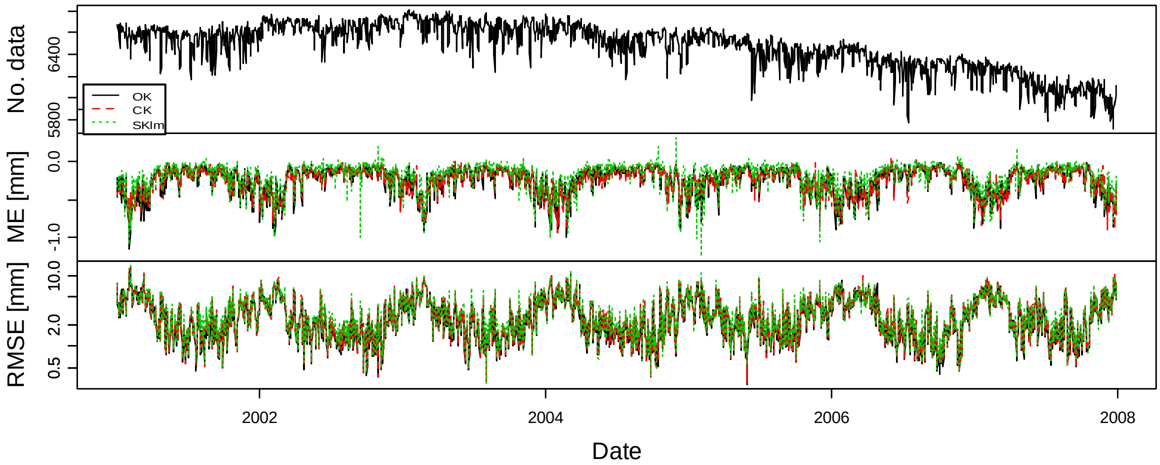

Figure 2. Daily time series of a) the number of data; and, the national daily average b) ME, and c) RMSE calculated for 2001-2007 for the three kriging methodologies.

The spatial trends of the error in the interpolation methodologies are represented in the national maps of ME and RMSE calculated using OK (Figure 3). The overall trends are similar between methodologies and, as such, only OK is displayed. The specific differences between the estimates will be analysed in the following section. The negative bias (ME) is spatially dominant across most of Australia and is more negative in Northern Australia. This pattern suggests that higher rainfall levels present in Northern Australia are generally underestimated by these interpolation methods. The exception to this trend is the small area on the North Coast of Queensland, where the methods both over- and underestimate for different stations in the area. The areas of high rainfall in Southern Australia (and especially around the inland side of the Great Dividing Range and Western Tasmania) are generally overestimated. These areas are typically marked by increases in altitude.

Analysis of the spatial distribution in RMSE in Figure 3b (when coupled with the results for the ME of Figure 3a) reveals that the underestimation in daily rainfall in Northern Australia is the largest source of error in the interpolation methods. The RMSE shows a North-South gradient with local average RMSE in excess of 5 mm. A small section of the Queensland coast shows the highest concentration of error. This area around Tully (arguably Australia's wettest town), has an annual rainfall in excess of 4000 mm. The sharp rise in the mountainous ranges in this area cause orographic rainfall from systems approaching from the ocean. The OK method does not include a mechanism to account for such orographic lifting. It, therefore, tends to underestimate rainfall on the windward side of the mountains. The opposite occurs on the leeward side of the mountain ranges. A rainshadow effect results in a reduction of rainfall and in this case OK overestimates precipitation (Figure 3a). Similar effects can be observed in the mountains in Central and Western Tasmania. The regular frontal systems produce the majority of rainfall approaching from the West.

The relatively high number of convective events in sub-tropical Northern Australia during summer causes an underestimation in variability. Due to the sparse gauge network in these areas, coupled with the small scale convection, these systems are often only captured by one station and the omission of stations in the cross validation process can result in substantial errors.

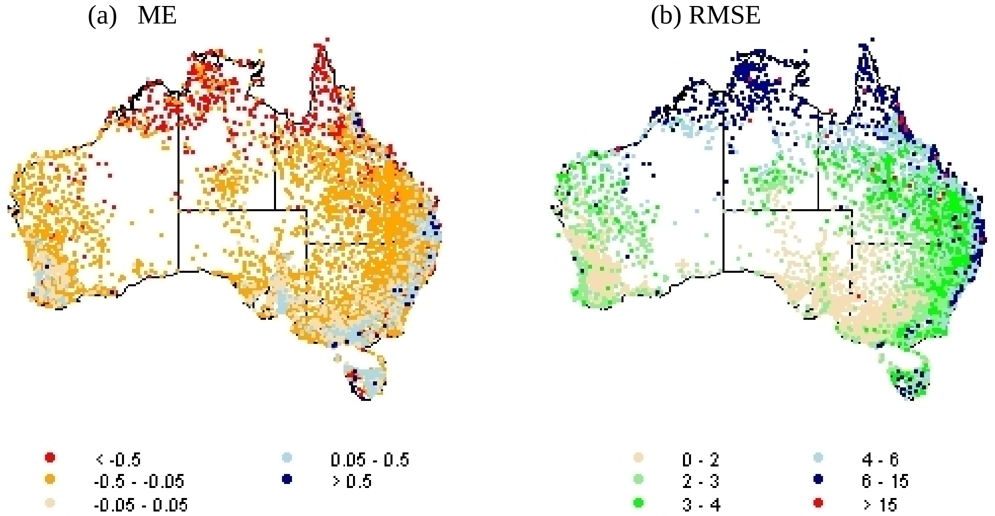

Figure 2. Spatial distribution of: a) the bias (ME) and b) RMSE for the OK interpolation method.

The difference between the two methods incorporating satellite data and OK, are presented for ME (Figure 4) and RMSE (Figure 5). Notably with regard to ME, a relatively lower/higher amount of rainfall is predicted in the coastal fringes for CK/SKlm compared to OK. Also, higher altitudinal areas tend to have more rainfall for the CK/SKlm. Typically OK shows a slight negative bias in the coastal fringes (due to orographic rainfall), which may imply that CK better captures the orographic effect. This consequently translates to other high altitude areas (such as the Great Dividing Range). The RMSE in the SKlm is considerably higher than the OK in the coastal regions of Southern Australia, where the gauge density is high. The SKlm (and the CK) does, however, perform better than the OK in Northern Australia where the gauges are sparse. Of the two methods incorporating the TRMM data, SKlm generally performs better in data sparser areas, but produces much higher errors in high rainfall coastal areas with a high gauge density,.

Figure 1 plots observed gauge versus satellite precipitation. The scatter reveals large differences occur between gauge and satellite based precipitation. This somewhat explains why CK/SKlm produced poorer results than OK. These differences are to some extent expected due to the difference in spatial scale of the two datasets. With this in mind, Figure 6 shows the spatial distribution of ME and RMSE, which compares the TRMM data to gauge values. Spatially coherent strong negative biases are observed for the majority of coastal areas (greater than -0.5mm daily) this is coupled with a positive bias (greater than 0.5mm daily) for inland high altitude areas. These results are consistent with the bias in the results for SKlm and CK. The positive bias of the satellite data for inland areas is also consistent with results reported by Janowiak et al. (2004) for central US and Su et al. (2008) for high rainfall events in the La Plata Basin, South America. The negative bias in coastal regions may be due to the fact that TRMM estimated precipitation does not capture the high resolution local rain events which are produced by uplift in coastal regions. In addition, the temporal scale of these events may not be captured by the 3hr TRMM time step. Another factor contributing to the coastal bias is the way the TRMM estimates fail to capture low level precipitation over land. Due to the high land emissivity in microwave frequencies only the scattering signal from ice can be used to detect rain over land. Wintertime coastal stratiform rain often has no ice signal and so goes undetected by microwave algorithms (E. Ebert pers. comm. 2009). The over estimation of rainfall in the high altitude regions could be due to increased clouds which does not result in precipitation. These clouds could be located on the leeward side of higher altitude areas where rainfall occurs.

4. CONCLUSIONS

The inclusion of satellite based precipitation estimates into gauge based interpolation methods is successfully trialled in this paper. However, the methods including satellite data (CK and SKlm) do not improve the overall error in the interpolation of daily rainfall compared to OK using gauge data alone. Analysis of the spatial variation in performance of the three methods reveals that the inclusion of the satellite does potentially improve spatial prediction of rainfall in data poor areas. The inclusion of the TRMM data increases the accuracy in North-West Australia. The poor relationship between gauge and TRMM precipitation causes large errors over most of Eastern and Southern Australia. The use of a single regression equation to define

(a) CK-OK **(b) SKlm-OK**

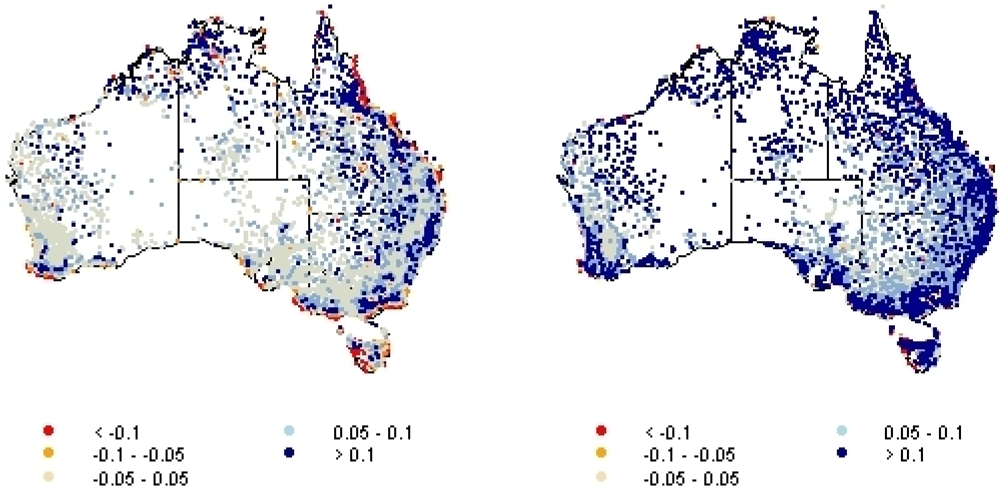

Figure 4. Difference between the RMSE for the CK/SKlm compared to OK

(a) CK-OK **(b) SKlm-OK**

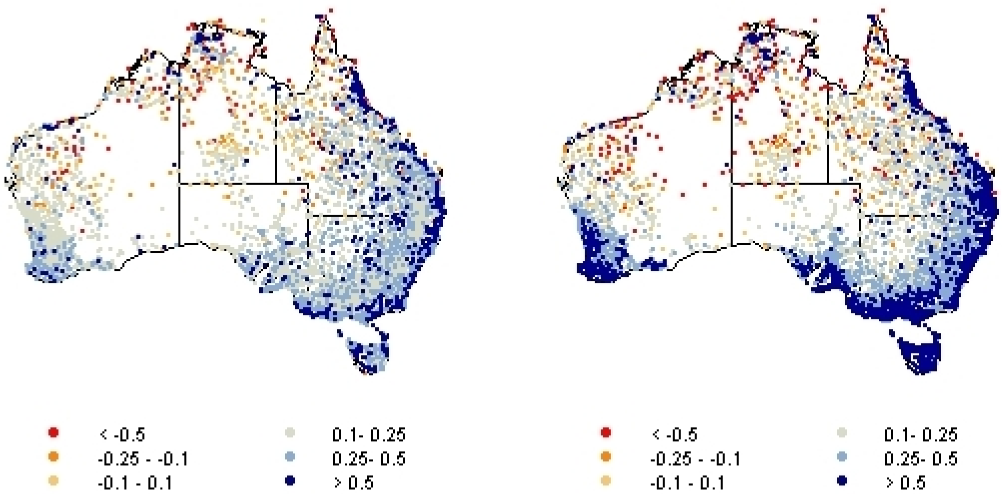

Figure 5. Difference between the RMSE for the CK/SKlm compared to OK

(a) ME **(b) RMSE**

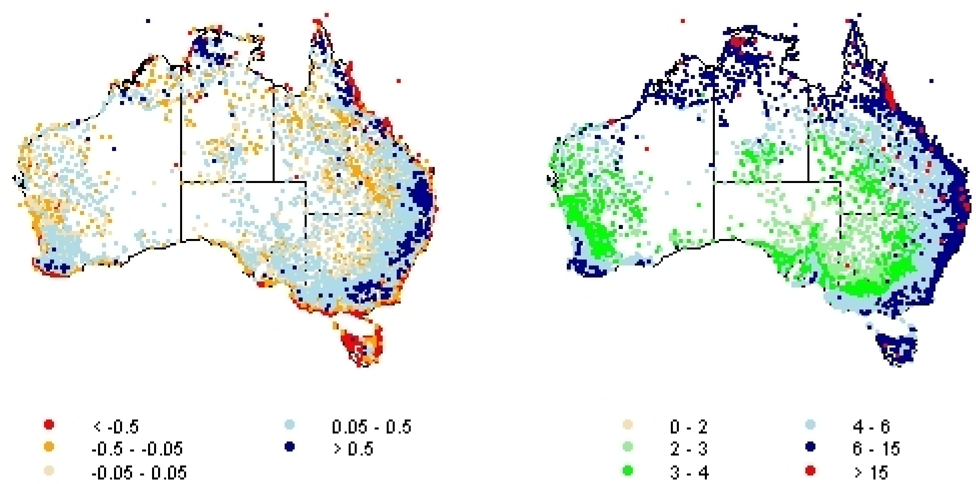

Figure 6. TRMM daily precipitation a) ME and b) RMSE compared to gauge values

the relationship between TRMM and the gauge data, where temporal and spatial variation in this relationship exists, may also explain the limited improvement provided by the inclusion of TRMM with SKlm.

Further analysis of the TRMM and observed gauge data revealed that there were significant negative biases in coastal precipitation estimates for the TRMM rainfall, and significant positive biases for inland higher altitude areas. The TRMM data does not capture the coastal showers due to the short duration and low level nature of these systems. The low spatial resolution of TRMM (0.25°) does not capture orographic effects that produce localised storms and rain shadows. These factors contribute to the weakening of the relationship between the gauge and the TRMM data and increase the error in the SKlm and CK methods of interpolation.

For future applications the inclusion of spatial biases found in the TRMM data and the development of methods to deal with these biases is suggested. This could possibly be carried out by using a spatially varying regression equation for SKlm (Kamarianakis et al., 2008) or by using bias removal on the TRMM fields before CK application The data could also be temporally stratified into summer tropical wet-season and mid-latitude winter before completing regression modelling for SKlm. Further, the use of alternative spatial fields (radar and numerical weather prediction models) is suggested to be included in future applications, either through simple interpolation techniques or rather by using model/data assimilation techniques.

5. ACKNOWLEDGMENTS

The authors wish to express their thanks to Dr Luigi Renzullo from CSIRO Land and Water for supplying the TRMM 9am Australian state aligned data used within this study. The authors would also like to thank Dr Alan Seed from the BoM for helpful discussions regarding interpolation techniques and coffee.

6. REFERENCES

Beesley, C., Frost, A.J. and Zajaczkowski, J. (2009) A comparison of the BAWAP and SILO spatially interpolated daily rainfall datasets, Water Division, Bureau of Meteorology, 32 pp

Chiew, F.H.S. et al. (2008) Rainfall-runoff modelling across the Murray-Darling Basin, A report to the Australian Government from the CSIRO Murray-Darling Basin Sustainable Yields Project Water for a Healthy Country Flagship. CSIRO, 70 pp

Cressie, N. (1993) Statistics for Spatial Data. John Wiley and Sons, 900 pp.

Ebert, E.E., Janowiak, J.E. and Kidd, C. (2007). Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bulletin of the American Meteorological Society*, 88, 47-64.

Goovaerts, P. (2000) Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology*, 228, 113-129.

Hazarika, M.K., Kafle T. P., Sharma R., Karki S., Shrestha R. M. and Samarakoon L. (2007). Statistical approach to discharge prediction for flood forecasts using TRMM data. Proceedings of the 5th Annual Mekong Flood Forum, Ho Chi Minh City, Vietnam, May 17-18.

Huffman, G.J., Adler, R.F., Bolvin, D.T.. Gu, G., Nelkin, E.J., Bowman, K.P., Hong, Y., Stocker, E.F., Wolff, D.B. (2007) The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales.*Journal of Hydrometeorology*,8,38-55.

Hutchinson, M.F. (1998). Interpolation of rainfall data with thin plate smoothing splines: II analysis of topographic dependence. *Journal of Geographic Information and Decision Analysis*, 2, 168-185.

Isaaks, E.H. and Srivastava, R.M. (1989). Applied geostatistics. Oxford University Press, New York, 561 pp.

Janowiak, J.E., Xie, P., Joyce, R.J., Chen, M. and Yarosh, Y. (2004). Validation of satellite-derived rainfall estimates and numerical model forecasts of precipitation over the United States. Proceedings of the 2nd International Precipitation Working Group Workshop, Monterey, CA, U.S.

Jones, D.A., Wang, W. and Fawcett, R. (2007). Climate Data for the Australian Water Availability Project. *Final Milestone Report*.

Kamarianakis, Y., Feidas, H., Kokolatos, G., Chrysoulakis, N. and Karatzias, V. (2008). Evaluating remotely sensed rainfall estimates using nonlinear mixed models and geographically weighted regression. *Environmental Modelling & Software*, 23(12), 1438-1447.

Kyriakidis, P.C., Kim, J. and Miller, N.L. (2001). Geostatistical mapping of precipitation from rain gauge data using atmospheric and terrain characteristics. *Journal of Applied Meteorology*, 40, 1855-1877.

Pebesma, E. (2004). Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*,30,683-91.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Renzullo, L.J. (2008). Considerations for the blending of multiple precipitation data sets for hydrological applications. 4th International Precipitation Working Group workshop, Beijing, China, October, 2008.

Su, F., Hong, Y. and Lettenmaier, D.P. (2008). Evaluation of TRMM Multisatellite Precipitation Analysis (TMPA) and Its Utility in Hydrologic Prediction in the La Plata Basin. *Journal of Hydrometeorology*, 9, 62-640.