Project 2: Verification of precipitation forecasts over SE Asia

Introduction. This project offers the opportunity to compare and contrast precipitation forecasts from three operational global models, for the SE Asian region. The models are the ECMWF global model, the Japanese global model (GSM) and the US global forecast model (GFS)

Goal: To evaluate and compare the characteristics of the forecasts from the three models.

Data: Six months (July to Dec. 2017) of observations and corresponding forecasts of 24 precipitation accumulations for six stations in SE Asia, for 5 forecast projections, 24,48,72,96, and 120h

Analysis suggestions: Precipitation accumulations over specific time lengths are often distributed like a gamma distribution with a long low-probability tail for the higher amounts, and are bounded at 0. Some users of forecasts tend to be most interested in the rarer occurrences of high amounts (because of flood risks), while others may be interested in the ability of the forecast to distinguish rainy situations from dry situations. These differences in user interest, along with the non-gaussian distribution, suggest that useful verification can best be achieved using a categorical form of the variable, along with contingency table verification metrics.

Suggested Steps: 1. Look at the data: Plot the observations against each of the model forecasts, visually compare the plots and check for suspicious data values.

2. Determine which thresholds will be possible to evaluate, based on a visual assessment of the sample sizes for the higher observed amounts. Choose thresholds from the set 1,5,10,20,30, and 50 mm. Include the higher thresholds only if there are more than 5 occurrences in the dataset.

3. Choose appropriate tools, carry out the calculations, to obtain the contingency tables for each threshold,

4. Analyse results and prepare presentation.
-examine the three contingency tables for each threshold chosen. Are they balanced (misses more or less the same as the number of false alarms)?
To discern the different aspects of the performance of the three models, compute and interpret score results obtained from the three contingency tables.

Results should be supported by bootstrapped confidence intervals wherever possible.

If time permits, the data could be stratified by station to see if there are any significant differences in performance over the regions

Possible questions to answer: How do the models differ in their prediction of precipitation extremes. Are there biases? How does the bias change over the range of the observed amounts? Can you identify a “winner” with clearly superior performance for all thresholds, or does it vary? Is the same model superior for all projections, or does the best model change for different projections?