APPENDIX A

Analysis of the performance of multivariate models inevitably involves a choice of the statistics used to assess the goodness of fit of model predictions (P_j) or simulations against observations (O_j) or measurements. We wish to point out that difference or error measures have gradually replaced correlation- and skill-based indices which are not consistently linked to model accuracy. In a key paper that serves as an important point of reference for such data versus model comparisons, Willmott et al (1985) outlined a set of difference measures that can be used to evaluate the operational performance of a wide spectrum of geophysical models, regardless of whether the model predictions are manifested as scalars, directions, or vectors. In particular, it was demonstrated that both model accuracy and precision can be meaningfully estimated by analyzing and reporting the following statistics:

- the prediction mean (μ_P) and standard deviation (s_P)
- the observation mean (μ_O) and standard deviation (s_O)
- the mean bias (b) calculated from the differences $D_j = P_j - O_j$ using $b = \sum_{j=1}^{N} D_j / N$
- the mean absolute error (MAE) and root mean-squared error (RMSE) calculated from the differences $D_j = P_j - O_j$ using $E^{1/\gamma} = \left[\frac{\sum_{j=1}^{N}|D_j|^{\gamma}}{N}\right]^{1/\gamma}$ such that MAE corresponds to γ=1 and RMSE correponds to γ=2
- the index of agreement (d_2) and the modified index of agreement (d_1) calculated from the differences $D_j = P_j - O_j$ using $d_\gamma = 1 - \frac{\sum_{j=1}^{N}|D_j|^{\gamma}}{\sum_{j=1}^{N}(|P_j - \mu_P| + |O_j - \mu_O|)^{\gamma}}$ such that d_1 corresponds to γ=1 and d_2 correponds to γ=2
- the systematic average error: $RMSE_s = \sqrt{\frac{\sum_{j=1}^{N}|\hat{P}_j - O_j|^2}{N}}$ and the unsystematic average error: $RMSE_u = \sqrt{\frac{\sum_{j=1}^{N}|\hat{P}_j - P_j|^2}{N}}$ calculated by decomposing the RMSE using $\hat{P}$, an OLS estimate of P obtained by regressing P on O
- the coefficient of determination R^2 as a measure of the amount of variance captured by the model.

At the level of exploratory statistics, the average values of predictions and observations are given by the means μ_P and μ_O and their variability by the standard deviations s_P and s_O. The *MAE* and *RMSE* provide complimentary measures of the size of the average difference between predictions and observations. The $RMSE_s$ is a measure of the size of the linear bias between model predictions and observations which is complimented by the mean bias (b) which measures the direction of this linear bias (i.e. under- or over-estimation). The average error measures: *b*, *MAE*, *RMSE* and $RMSE_s$ all have the same units as P and O but the former two depend on linear differences between P and O while the latter two (and also the $RMSE_u$ which is a measure of model precision) depend on squared differences between P and O meaning that they are more stringent. It is important to note that d_1 and d_2 are unit-independent measures of model performance and take values in the interval [0,1] with the former being calculated from linear differences and the latter being calculated from squared differences between P and O.