11 Predictability Measures

11.1 Measures of model-derived predictability

Before we formally introduce the concepts of signal and noise in the seasonal prediction context, let us first consider an example of a seasonal prediction in Fig. 74.

Let x_{ij} be a model variable (e.g. near-surface temperature or rainfall) at a certain gridpoint at a discretized time $i = 1, ..., N$ for the ensemble member $j = 1, ..., M$. The *noise* in weather and climate is then usually defined as the variance of the deviations from the ensemble mean

$$NO = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} (x_{ij} - [x]_i)^2 \quad , \tag{164}$$

where $[x]_i$ is the ensemble mean

$$[x]_i = \frac{1}{M} \sum_{j=1}^{M} x_{ij} \quad . \tag{165}$$

We can obviously define these values at every gridpoint of the model and thus define maps of e.g. NO. The meaning of this definition becomes clear if we imagine that the results of an ensemble of 2 (or more) simulations are identical. In this case the noise NO is zero. The ensemble mean of an ensemble is supposed to define the signal, thus

$$SI = \frac{1}{N} \sum_{i=1}^{N} \left([x]_i - \overline{[x]} \right)^2 \quad , \tag{166}$$

where $\overline{[x]}$ is the time mean of the ensemble mean

$$\overline{[x]} = \frac{1}{N} \sum_{i=1}^{N} [x]_i \quad . \tag{167}$$

It should be noted that the ensemble mean usually contains a noise residual, particularly if the ensemble size is small, and provides therefore a *biased* estimate of the signal. Imagine we have just 2 simulations, then it is clear that the mean of these 2 simulations will not be effective to identify the signal (unless they are identical, and therefore the noise is zero). There are ways to correct/improve this, but we will not deal with this here. Discuss this using ensemble Nino3.4 forecasts! With this the *signal-to-noise* ratio $S - N$ is simply

$$S - N = \frac{SI}{NO} \quad . \tag{168}$$

In principle, $S - N$ can become infinity, if NO is zero. In practise this is usually not the case, although values can become quite large (in which case we are lucky!!!). A useful threshold to be considered could be $S-N=1$, for which the signal has the same variance as the noise, indicating some predictability. In practical applications to important variables such as surface temperatures, precipitation or geopotential height in seasonal forecasts, $S-N$ turns out to be typically small for precipitation and other variables outside the tropical Pacific regions (see Fig. 75).

A useful transformation of the $S-N$ is called *theoretical limit of predictability*

$$R_{limit} = \sqrt{\frac{SI/NO}{SI/NO+1}} = \sqrt{\frac{SI}{SI+NO}} = \sqrt{\frac{SI}{T}} \quad , \tag{169}$$

where $T = SI + NO$ is the *total* variance. By definition R_{limit} lies between 0 and 1. It may be interpreted as maximum expected correlation skill if we were to correlate the ensemble mean with the observations (correlation skill will be introduced in the next section). Zero means there is no predictability, 1 means there is perfect predictability. A value of $S-N$ of 1 translates into a value of R_{limit} of about 0.7. There are many more indicators of predictability, that are related to information theory (e.g. relative entropy), but we will restrict ourselfs here to just the basic ones. Fig. 76 shows the R_{limit} for seasonal mean (September-to-November) precipitation over land points. As we can see, unfortunately the seasonal mean theoretical limit of Predictability is typically low over land points. Another global assessment of R_{limit} for surface temperature, mean sea level pressure and precipitation for the DJF season is shown in Fig. 77

Another, purely model derived predictability measure is the *potential correlation skill*. The idea is to calculate some kind of mean correlation derived from the model that may be compared with the correlation of an ensemble mean with an *observation*. We may treat every single ensemble member of an ensemble of realizations as observation (it contains internal and forced variability components). We can correlate each ensemble member with an ensemble mean of the *remaining* simulations. For example, we use x_{i1} and correlate this with the ensemble mean

$$[x]_i^{no_1} = \frac{1}{M-1} \sum_{j=2}^{M} x_{ij} \quad . \tag{170}$$

This is because if x_{i1} were included in the ensemble mean calculation then we get trivial correlations due to this. Then we can calculate M correlation coefficients

$$\rho_j = \frac{\frac{1}{N}\sum_{i=1}^{N}\left([x]_i^{no_j} - \overline{[x]}^{no_j}\right)(x_{ij} - \overline{x}_j)}{\sigma_{[x]^{no_j}}\sigma_{x_j}} \tag{171}$$

where the standard deviations in time are defined as

$$\sigma_{x_j} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{ij} - \overline{x}_j)^2} \tag{172}$$

and

$$\sigma_{\overline{x}^{no_j}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left([x]_i^{no_j} - \overline{[x]_i^{no_j}}\right)^2} \quad . \tag{173}$$

Now one should average the correlations ρ_j. However, we should not just average these correlations since they are limited by [-1,1]. Instead, we may average them after applying a $FischerZ$-transformation

$$r_j = \frac{1}{2}\ln\left(\frac{1+\rho_j}{1-\rho_j}\right) \quad , \tag{174}$$

to get

$$[r] = \frac{1}{M}\sum_{j=1}^{M} r_j \quad . \tag{175}$$

After this, in order to get the average correlation, we have to transform back using the inverse transformation

$$[\rho] = \frac{e^{2[r]}-1}{e^{2[r]}+1} \quad . \tag{176}$$

Note that, in practise, this potential correlation skill $[\rho]$ is very similar numerically to the theoretical limit of predictability, R_{limit}. The reason for this is that the square of a correlation is the *explained variance*, which applied to our case is the explained variance fraction by the ensemble mean or by the signal. An example if a Potential Correlation Skill calculation is shown in Fig. 78.

11.2 Predictability from comparison with observations

Of course, in order to investigate the 'goodness' or *skill* of a model simulation, we should compare the model output with observations. There are a number of quality measures of a model simulation. The most basic one, perhaps, is the *bias* of the model. If the model has several realizations, that is, ensemble members, the (time-mean) bias is best evaluated by comparing the ensemble mean of a field, $[x]_i$, with the corresponding observation, ox_i

$$bias = \frac{1}{N}\sum_{i=1}^{N}([x]_i - ox_i) \quad . \tag{177}$$

If an ensemble is not available, but just a single simulation, the single simulation can be used to assess the bias. One of the most commonly used measures of error that takes variability into account is the *Root-mean-square error (RMSE)*

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}([x]_i - ox_i)^2} \quad . \tag{178}$$

Again, a single realization may be also used to evaluate the RMSE of a model. The problem with the $RMSE$ is that we need to know what is an acceptable value of it, which can be quite tricky. In a forecast for a week, is a $RMSE$ of 1 K acceptable? A more generically comparable predictability measure is the *Correlation Skill*

$$\rho = \frac{\frac{1}{N}\sum_{i=1}^{N}(ox_i - \overline{ox})([x]_i - \overline{[x]})}{\sigma_{ox}\sigma_{[x]}} \quad . \tag{179}$$

Such a measure may be somewhat simpler to evaluate as is has values between -1 and 1 and we may simply pick a generic threshold of, e.g. 0.5, for any variable. On the other hand, physically me may prefer the $RMSE$ error measure (for example if your model has variations that are a factor of 1000 smaller than the observations, the Correlation Skill may still be 1, but the RMSE would be large). In order to better evaluate the $RMSE$ we should compare it with some kind of trivial forecast without real skill. Such a forecast could be a climatological forecast or a persistence forecast. The *Brier Skill Score* uses this idea to define skillful forecasts

$$BS = 1 - \frac{RMSE^2}{RMSE_{cl}^2} \quad , \tag{180}$$

where $RSME_{cl}$ is the base-line $RMSE$ of, for example a climatological forecast. A forecast is skillfull (compared to the base-line forecast) is the BS is positive.

An advantage of the Correlation Skill measure, ρ, is that we can directly compare it with the model-derived potential correlation skill $[\rho]$ (Eq. 176) or the theoretical limit of predictability, R_{limit} (Eq. 169). Fig. 79 shows an example of real seasonal prediction skill from a multimodel ensemble of seasonal *hindcasts*

11.3 Some other useful and simple techniques

A very useful technique in climate research is the *regression* analysis. Assume we want (in a model or in observations) investigate what is the influence of ENSO (or any other phenomenon) on rainfall. Assuming that we can characterize ENSO by a single time-series (i.e. the Nino3.4 SST index), then the influence of ENSO on rainfall at a certain location may be determined by a linear regression of rainfall at time i, r_i onto the Nino3.4 index (I_i)

$$r_i = a + bI_i \quad . \tag{181}$$

In the linear regression we are looking now for the coefficients a and b that minimize the sum of the squared differences between the linear model, Eq. (181), and the observational (or numerical model) counterpart ro_i

$$\epsilon = \frac{1}{N}\sum_{i=1}^{N}(ro_i - r_i)^2 \tag{182}$$

The theory of the linear regression tells us now how to determine (exercise!) the coefficients a and b (see Fig. 80). The mostly used coefficient b can be expressed as

$$b = \frac{\frac{1}{N}\sum_{i=1}^{N}(ro_i - \overline{ro})(I_i - \overline{I})}{\sigma_I^2} \quad . \tag{183}$$

Thus b can be interpreted as covariance between the index I_i and ro_i, divided by the variance of I_i.

Note that this formulae can be evaluated at every gridpoint, leading to a map of regression coefficients. The map of the coefficients b would tell us what is the typical linear *response* to a 1 K Nino3.4 SST anomaly in global rainfall (the dimension is mm/day per K). Also note that sometimes a different scaling is used

$$b^* = \frac{\frac{1}{N}\sum_{i=1}^{N}(ro_i - \overline{ro})(I_i - \overline{I})}{\sigma_I} \quad , \tag{184}$$

so that the dimension of b^* is simply mm/day. This would be the result if we had taken from the beginning a normalized index $I_i^* = (I_i - \overline{I})/\sigma_I$ which has standard deviation 1 by definition. Thus b^* is simply the covariance between the normalized index I_i and ro_i. This may be interpreted as the response to a *normalized* index I_i or the response to one standard deviation of the regression coefficient. One may interpret this also as a *composite* map based on linear regression. If we have an ensemble of simulations we could, for example, compare the ENSO regression map of the ensemble mean onto global rainfall with the observed regression map and try to identify if there are errors in the ENSO *teleconnections*.

A comparision between observations and an AGCM for the b^* regression coefficients for the Nino3.4 index regression onto winter (December-to-February) mean precipitation is shown in Fig. (81).

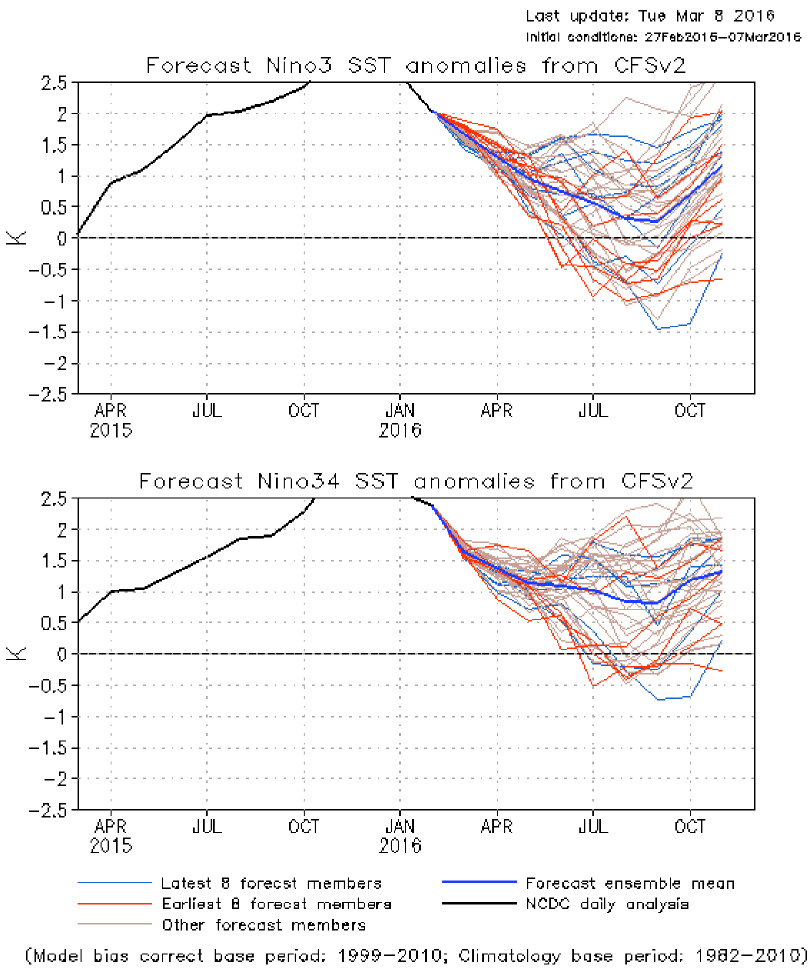
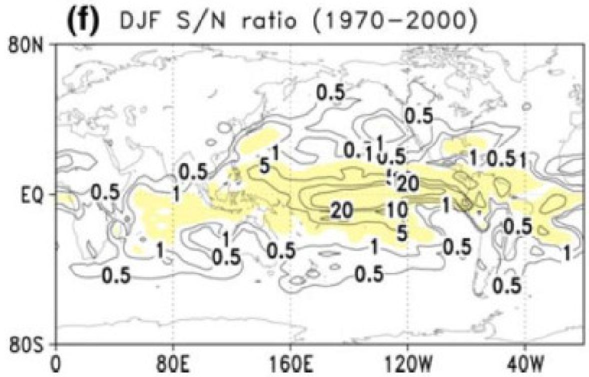


Figure 74: An exampe of Nino3 and Nino3.4 SST index prediction to illustrate the Signal-to-Noise ratio concept.



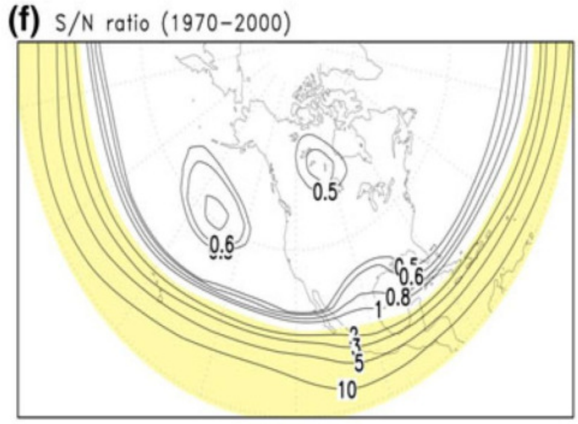


Figure 75: Signal-to-Noise ratio for winter (December-to-February) mean derived from an AGCM ensemble. Upper panel: Precipitation, lower pane: 200 hPa geopotential height. From paper: Ehsan et al., 2013: A quantitative assessment of changes in seasonal potential predictability for the 20th century. *Clim Dyn*, **41**, 2697-2709, doi: 10.1007/s00382-013-1874-x.

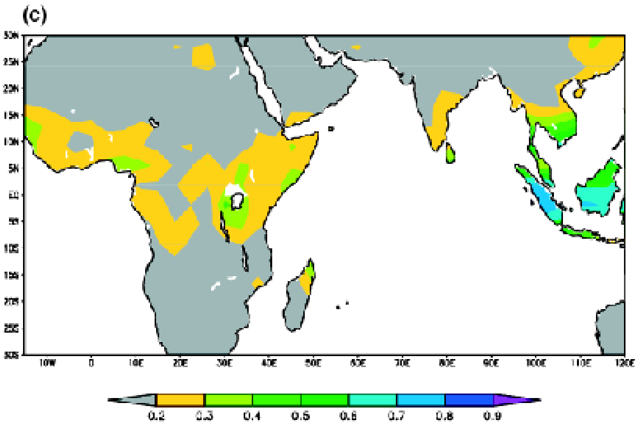


Figure 76: Theoretical limit of predictability, R_{limit} for September-to-November mean precipitation derived from an AGCM ensemble. From paper: Bahaga et al., 2015: Potential predictability of the sea-surface forced Equatorial East African short rains interannual variability in the 20th century. *Q. J. R. Meteorol. Soc.*, **141**, 16-26, doi: 10.1002/qj2338.

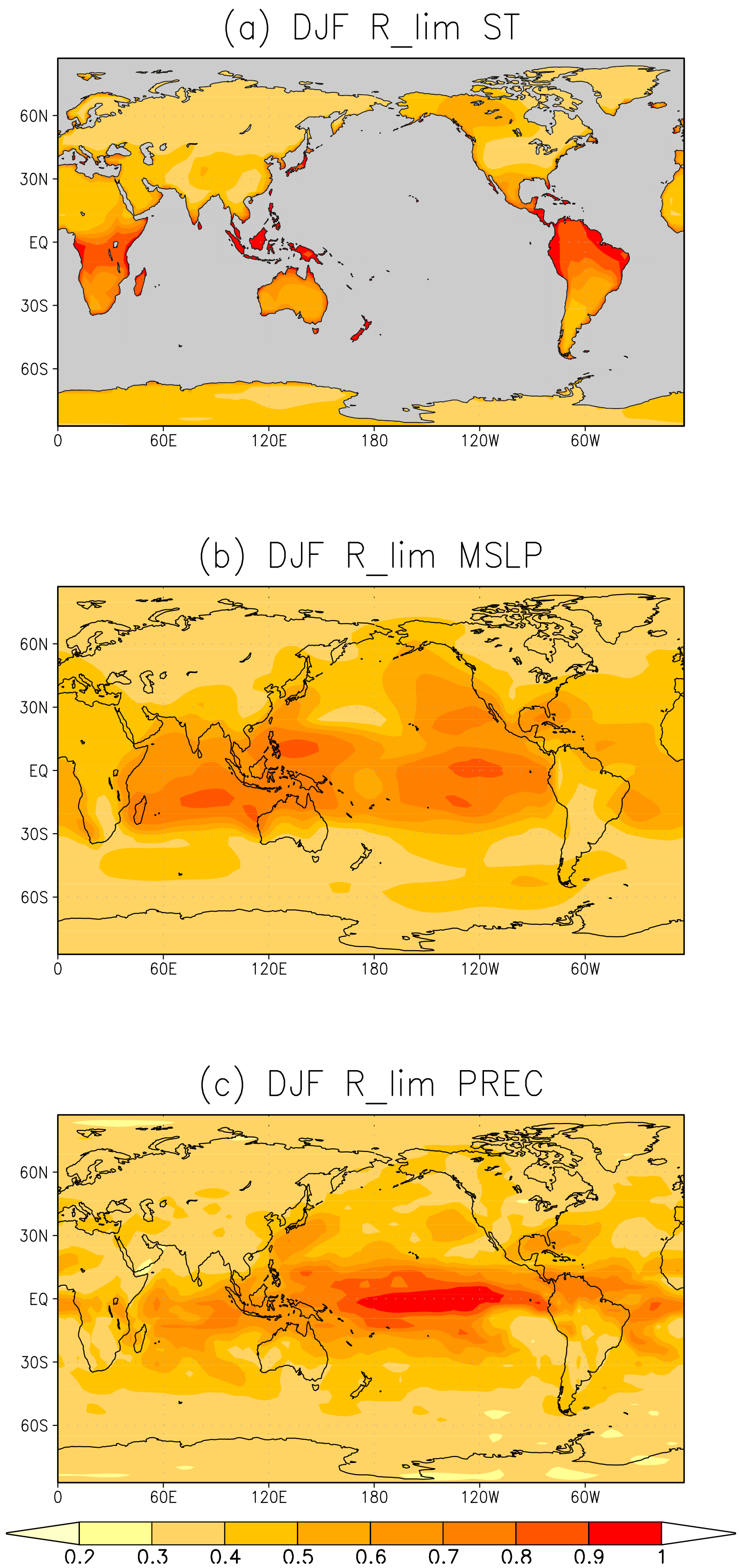


Figure 77: Theoretical limit of predictability, R_{limit} for December-to-Februar mean a) surface temperature, b) mean sea level pressure and c) precipitation for the DJF season derived from an AGCM ensemble. From Thesis: Bianca Mezzina, 2016: Seasonal influences of SST variability on European climate. University of Trieste Master thesis.

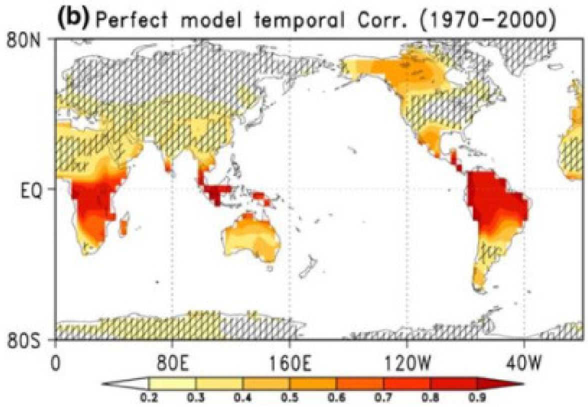


Figure 78: Potential Correlation Skill of near-surface temperature for winter (December-to-February) mean derived from an AGCM ensmemble. From paper: Ehsan et al., 2013: A quantitative assessment of changes in seasonal potential predictability for the 20th century. *Clim Dyn*, **41**, 2697-2709, doi: 10.1007/s00382-013-1874-x.

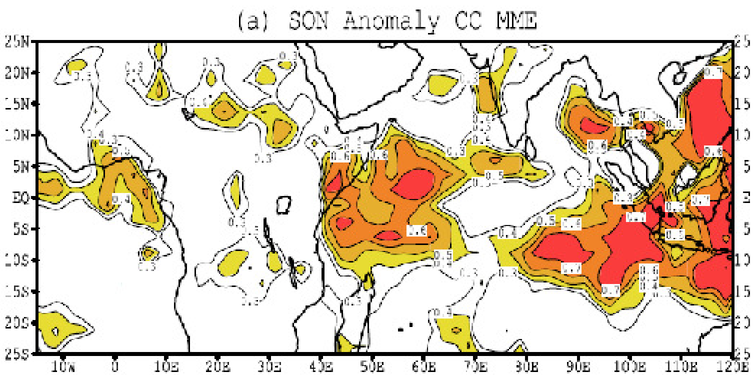


Figure 79: Correlation Skill of seasonal mean (September-toNovember) precipitation from a multimodel seasonal forecast enssemble. From paper: Bahaga et al., 2015: Assesment of prediction and predictability od short rains over equatorial East Africa using a multi-model ensemble. *Theor. Appl. Climatol.* , doi: 10.1007/s00704-014-1370-1.

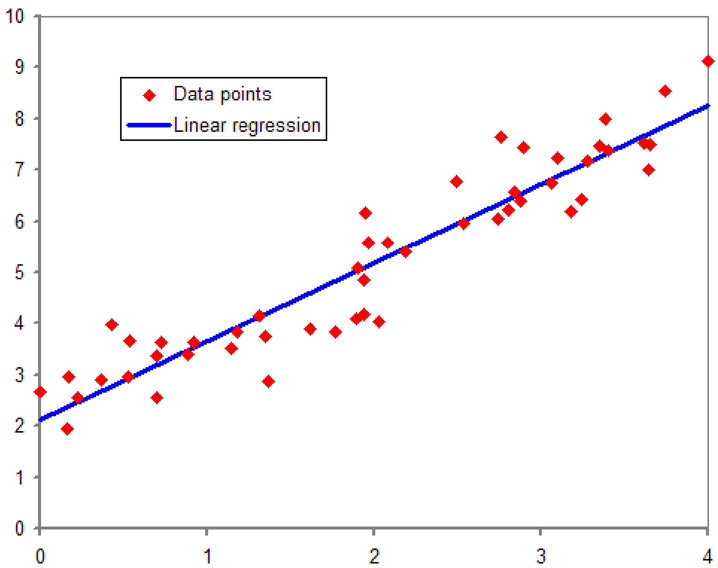


Figure 80: Linear regression method.

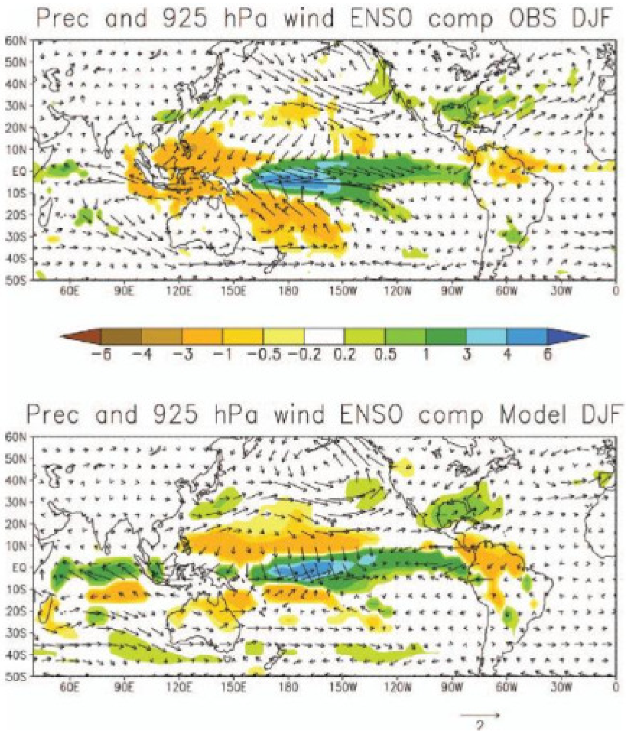


Figure 81: Regression coefficients b^* of the Nino3.4 index onto winter (Dececmber-to-February) mean precipitation. a) Obseravtinos, b) AGCM. Units are mm/day. From paper: Kucharski et al., 2015: On the need of intermediate complexity General Circulation Models: A “SPEEDY” Example. *BAMS*, **94**, 25-30, DOI: 10.1175/BAMS-D-11-00238.1.