Royal Netherlands
Meteorological Institute
Ministry of Infrastructure and the Environment

Introduction to empirical models
(including relevant statistical concepts)

MedCOF Training Workshop
Madrid, 26-30 October 2015

Jonathan Eden
Royal Netherlands Meteorological Institute (KNMI)

Outline

1. Introduction
2. Teleconnections and the basis empirical seasonal prediction
3. Methods in empirical seasonal prediction
4. Empirical seasonal prediction in practice
5. Summary and outlook

Outline

1. Introduction
2. Teleconnections and the basis empirical seasonal prediction
3. Methods in empirical seasonal prediction
4. Empirical seasonal prediction in practice
5. Outlook

What is empirical prediction?

- "Empirical" – based on experience
- Using existing global data sets to construct hypotheses and prediction methods.
  - What do observations tell us about the workings of the climate system?
  - What is the maximum value we can extract from real world data?

- Eliciting statistical relationships to represent known physical processes in the climate system.
- Usually, the relationship between regional-scale anomalies in a target variable (the "predictand") and climate phenomena (the "predictors").

"Empirical" or "statistical"?

- Terms often used interchangeably:
- van den Dool (2007) asks:
  - Are we driven by intuition? (empirical)
  - Or to apply or develop a statistical methodology? (statistical)

- In seasonal prediction, it's usually the former…
  - The first step is always to understand physical relationships.
  - Then to find a statistical model that best represents these relationships.

Dynamical prediction… and its limitations

- Operational seasonal forecasting now a regular activity.
- Dynamical forecast systems remain the most important tool in producing predictions.
- However…
  - Development is inherently complex.
  - Forecast generation is computationally demanding.
  - Errors and biases limit model skill.

A place for empirical prediction

- An alternative to dynamical prediction systems.
- May serve as a **baseline** for dynamical models.
- Used to improve forecasts by limiting the effects of dynamical model biases.
  - Particularly so in regions where dynamical systems have known weaknesses.

A place for empirical prediction

- An alternative to dynamical prediction systems.
- May serve as a **baseline** for dynamical models.
- Used to improve forecasts by limiting the effects of dynamical model biases.
  - Particularly so in regions where dynamical systems have known weaknesses.
- However…

Different development…

…different output…

…systematic comparison difficult…

(more later…)

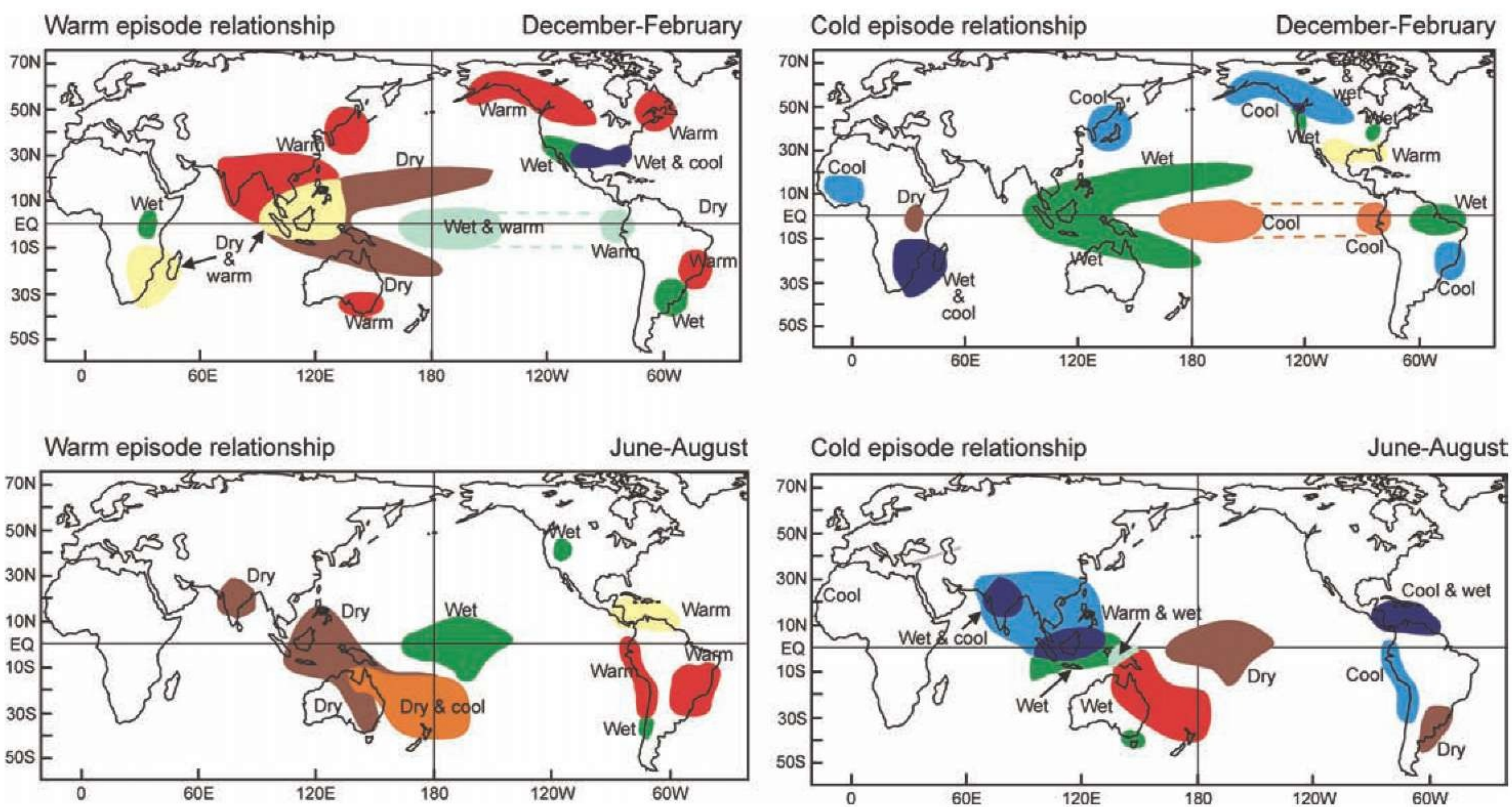
Outline

1. Introduction
2. Teleconnections and the basis empirical seasonal prediction
3. Methods in empirical seasonal prediction
4. Empirical seasonal prediction in practice
5. Summary and outlook

Teleconnections

- Atmospheric circulation exhibits sustainable variability on time scales from a few days to several years or even centuries.
- Recurring and persistent patterns of pressure and circulation anomalies across large vast geographical areas - teleconnections.
- Teleconnections typically form the basis for an empirical seasonal prediction model.
  - When we have existing knowledge of a teleconnection…
  - …we can develop an empirical model to represent it.

Known ENSO teleconnections



Lagged teleconnections

- Teleconnections often spoken about in simultaneous terms… which is not much use for prediction.
- In reality, an anomaly in the climate system may take days/ weeks/months to propagate.
- Regional climate anomalies associated with ENSO, for instance, may be predictable many months in advance.
- Empirical models often incorporate some lag – finding a link between the predictand and the predictor at some lead time.

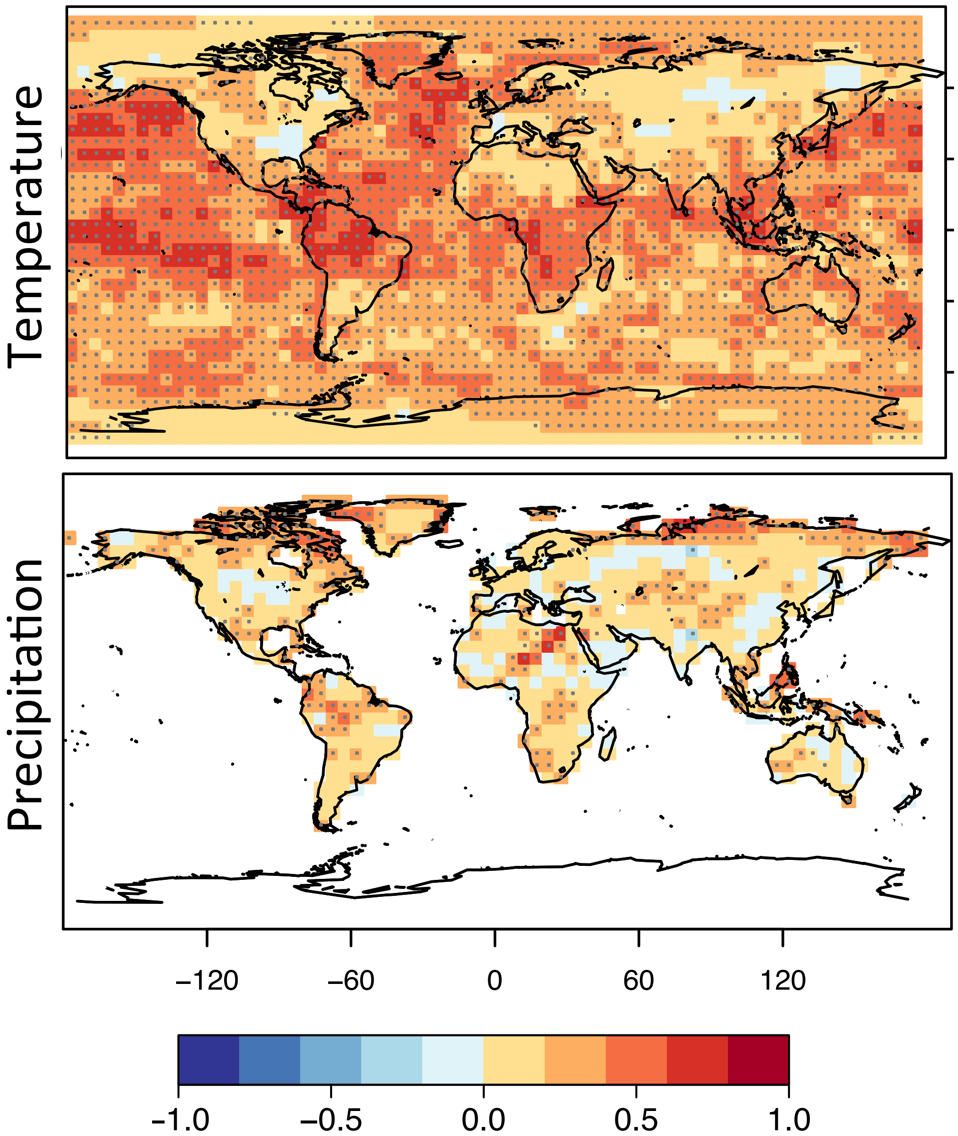
Outline

1. Introduction
2. Teleconnections and the basis empirical seasonal prediction
3. Methods in empirical seasonal prediction
4. Empirical seasonal prediction in practice
5. Summary and outlook

Persistence

- Empirical prediction at its most simple.
- **Persistence**: observations of a given variable at some lead time are taken as the forecast for that variable.
- Such forecasts have frequently performed better at short lead times than those simply prescribed by climatology.
- Often persistence is used as a reference method for other statistical methods.

Persistence – what is the best we can do?



Correlation between MAM and DJF climate anomalies (1961-2013).

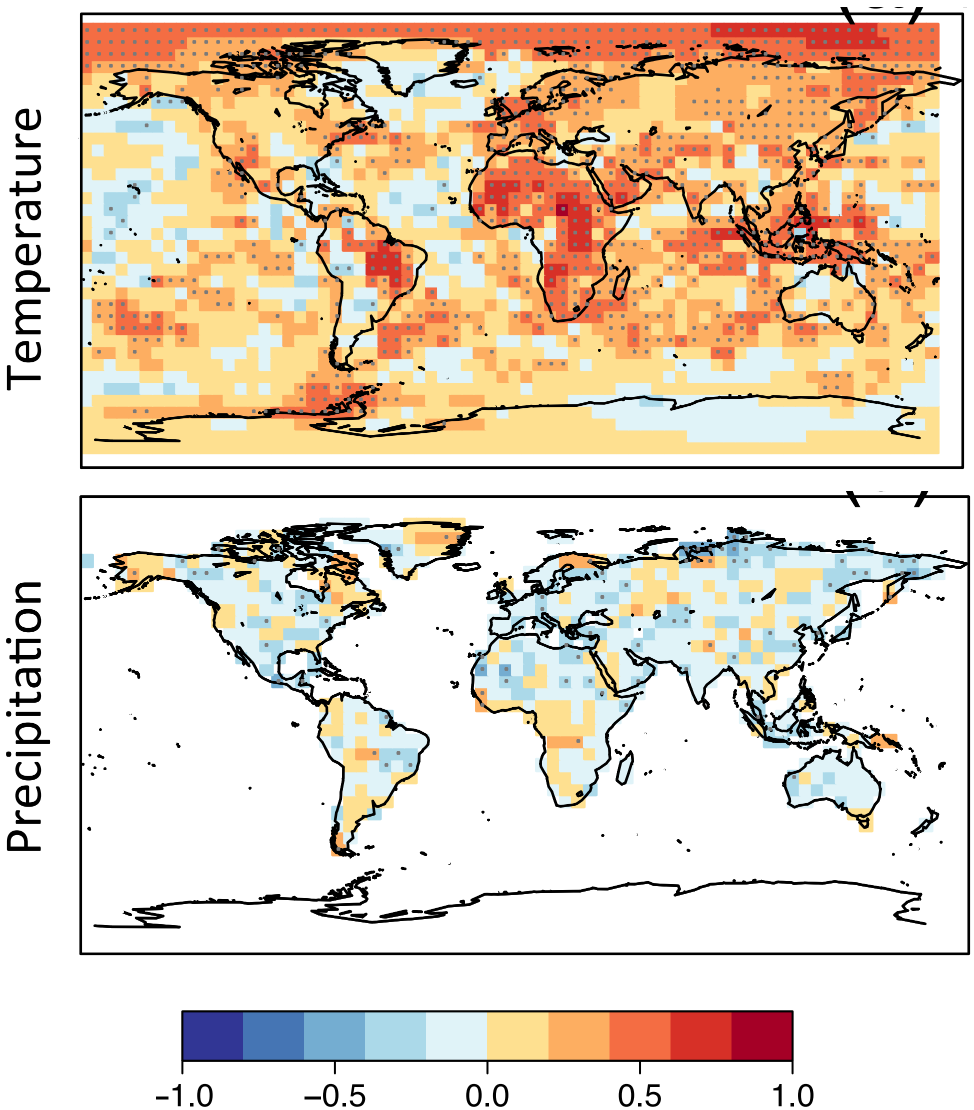
- Persistence usually skillful for temperature in the tropics... less so in other regions.
- For precipitation, skill is generally limited.

Linear regression

- Stronger skill usually comes from relating the temporal sequence of predictor and predictand events in the observed record by simple **linear regression**.
- For instance, a linear regression model may describe the relationship between soil moisture during March and temperature during April.
  - Dry soil -> decreased evaporation -> increased temperature.
  - Forecast for time *t* requires regression coefficients applied to soil moisture at time *t-1*.

- Other possibilities include making use of the predictive skill associated with long-term climate change.

Linear regression – global warming signal



Much discussion recently on the merits of using global warming signal as a source of skill.

Correlation between observed and predicted MAM climate (1961-2013) using greenhouse gas concentration as a predictor in a linear regression model.

- Useful, potentially for temperature.
- Precipitation dominated by climate system internal variability… alternative predictors needed!

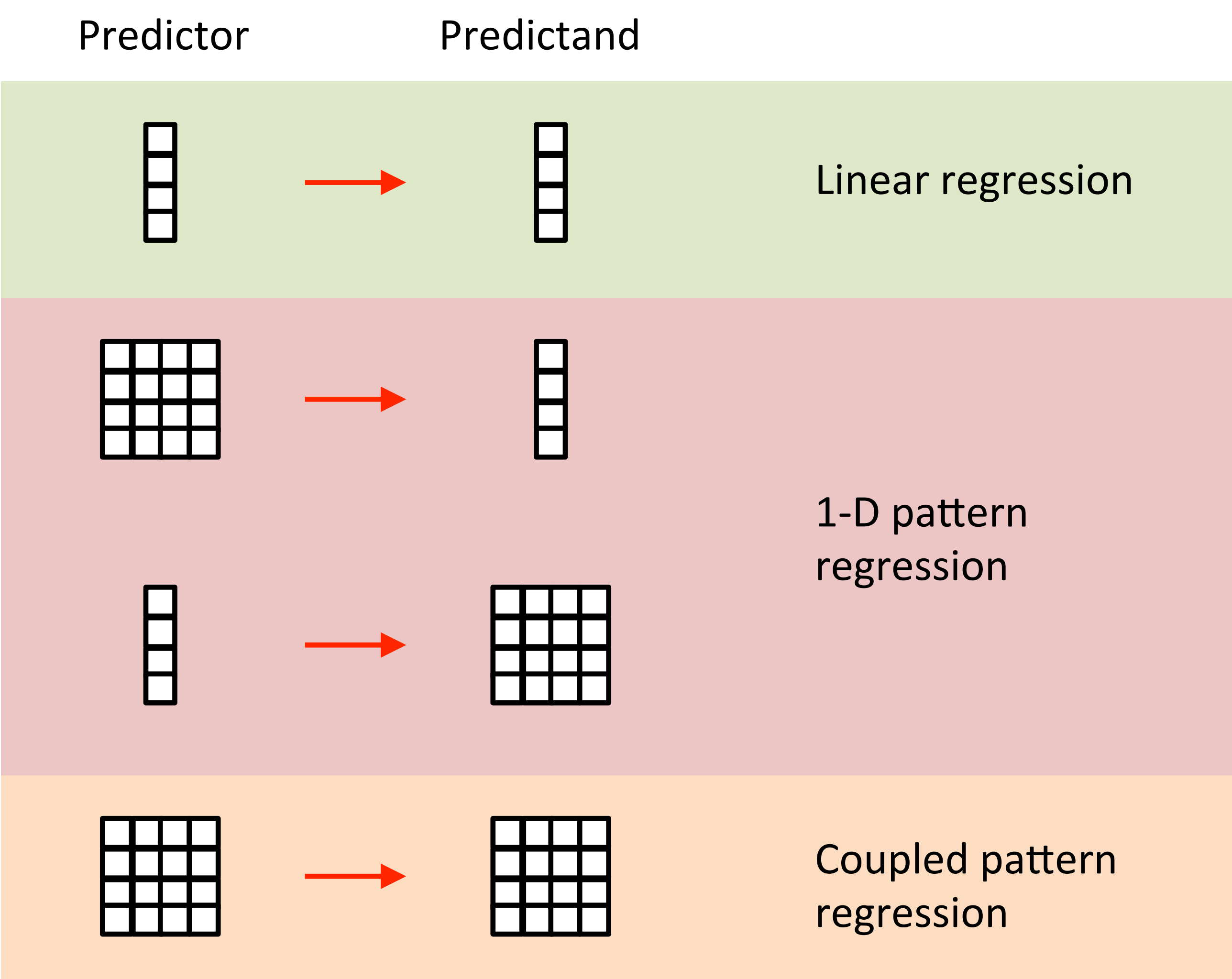
Non-local linear regression

- In seeking to represent teleconnections, “non local” linear regression is more suitable.
- For example, what is the linear relationship between European winter temperatures and Pacific SST from the preceding autumn.
- Often, we use climate indices to describe climate anomalies in remote regions:
  - ENSO: NINO3.4, NINO4.
  - Pacific Decadal Oscillation (PDO).
  - Indian Ocean Dipole (IOD)…. etc.

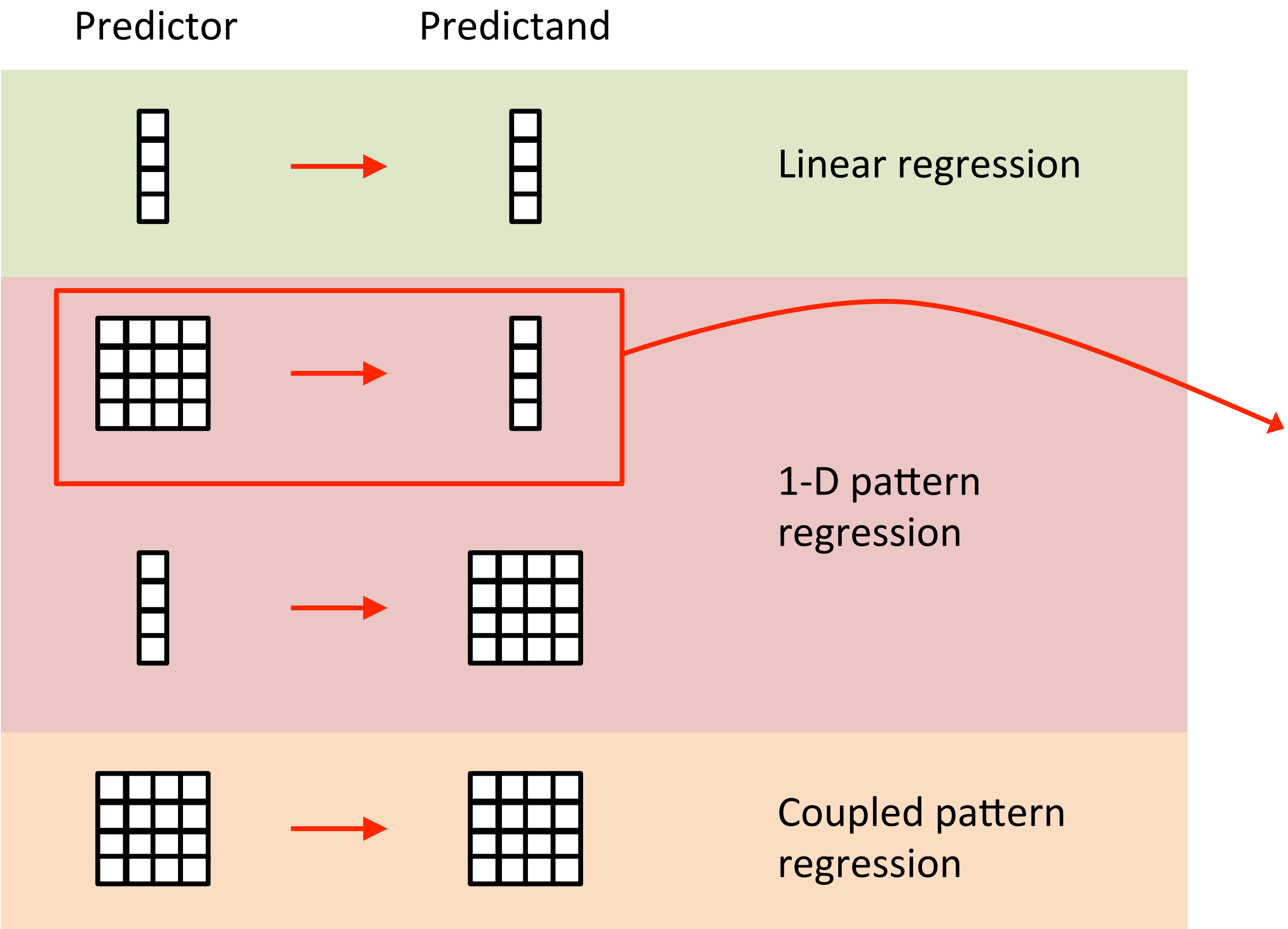
Pattern-based regression

- More complex methods seek to find patterns of variability in spatial fields of climate data.
- EOF analysis is common method of reducing data into several 'modes'.
- The linking of a time series of spatial patterns with, either, a time series at a given location or, alternatively, another time series of patterns.
- For examlple, the time series of one or mode EOF modes may be used to fit a (multiple) linear regression model to estimate the predictand.

Pattern-based regression



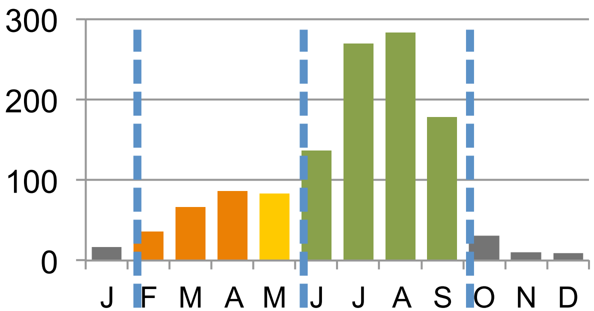
Pattern-based regression

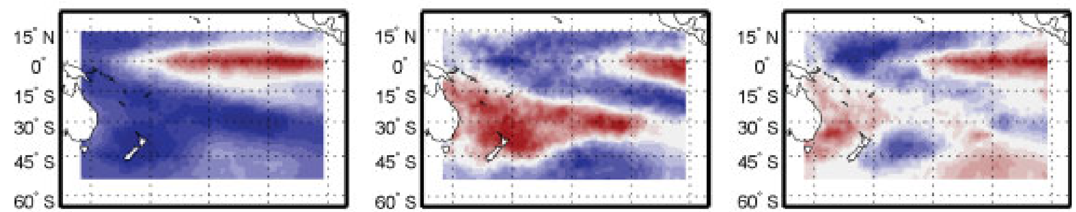


Example:
Forecasting spring rains in Ethiopia

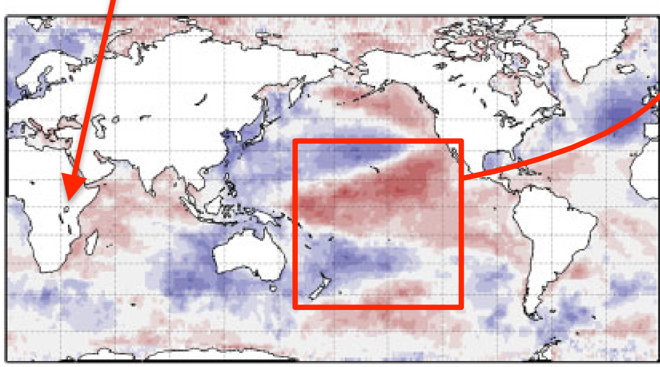
Pattern-based regression – example

Forecasting spring rains in Ethiopia

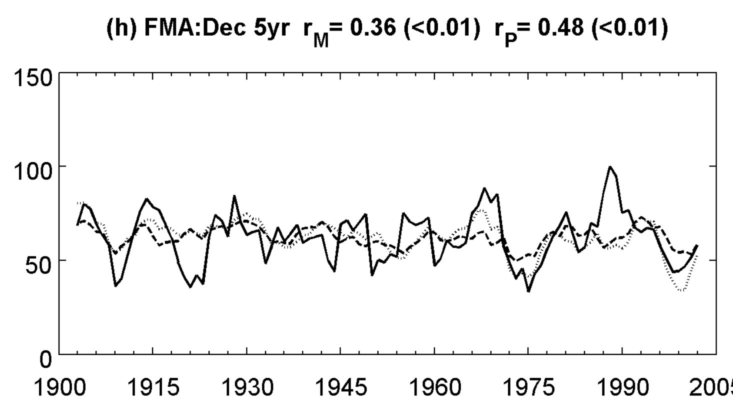




Leading EOFs used to estimate Ethiopia spring precip.

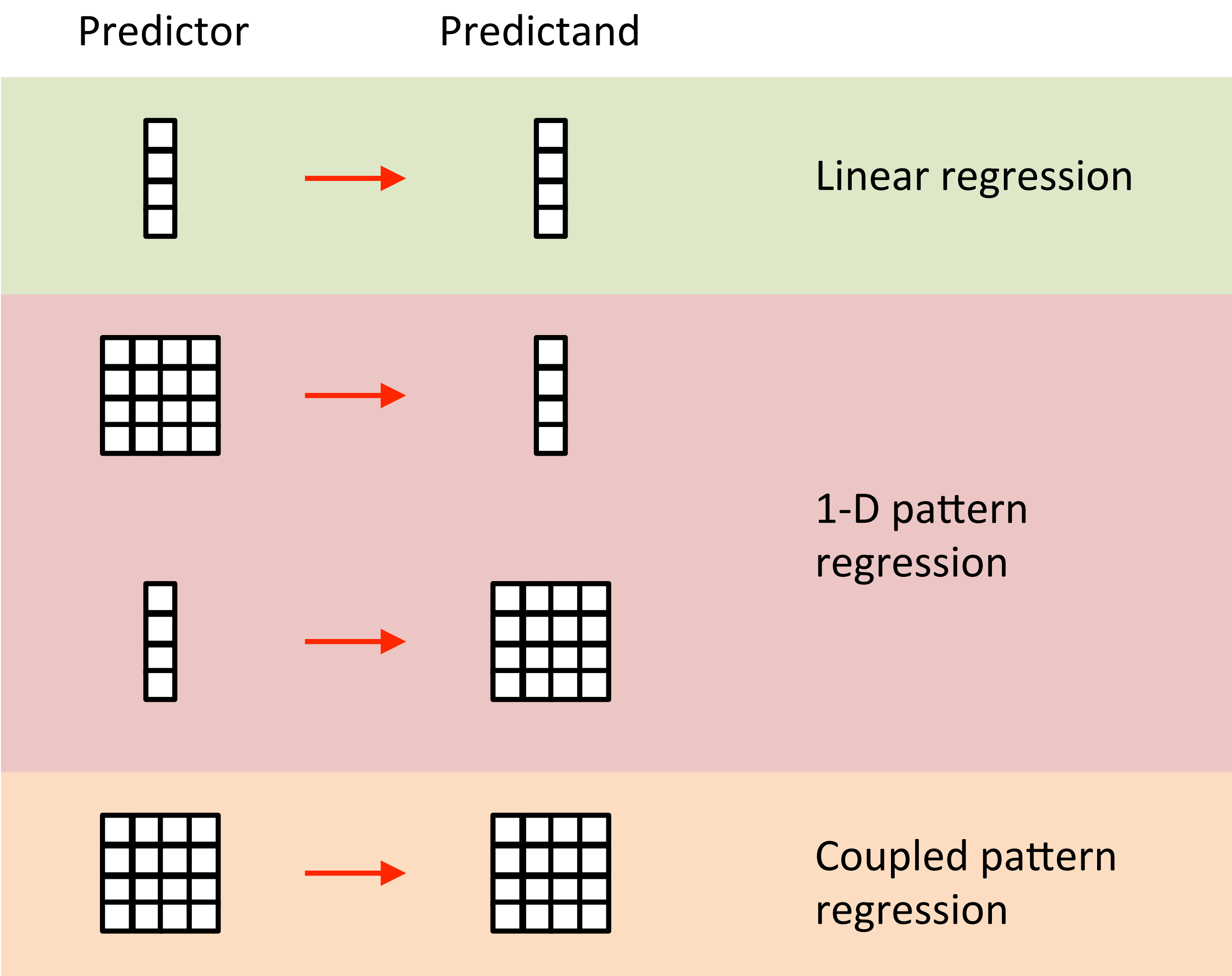


Correlation between Ethopia precip and SST (1901-2010).

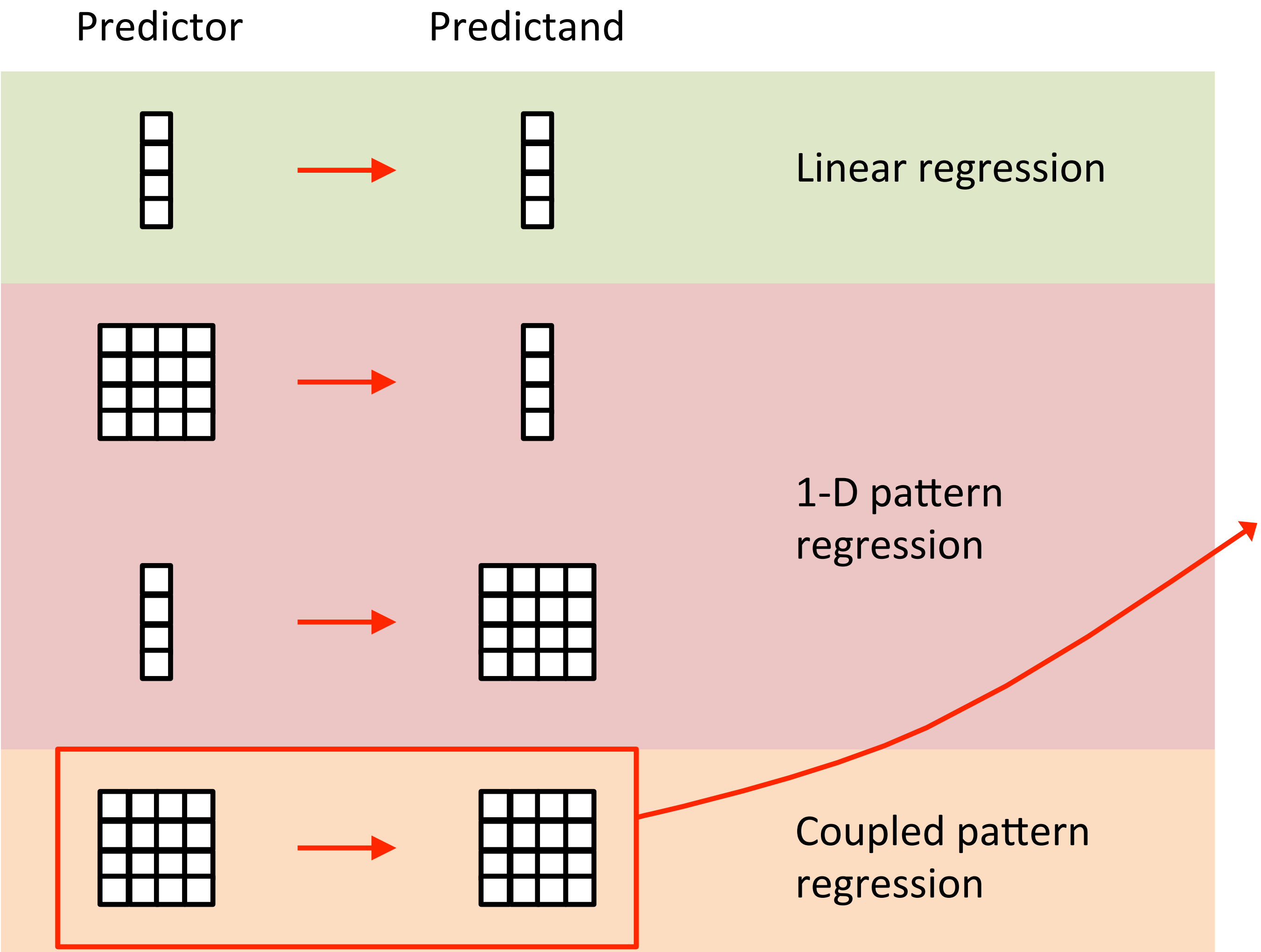


Eden *et al*. (2014) Int. J. Climatol.

Pattern-based regression – coupled patterns



Pattern-based regression – coupled patterns



Methods include:

- Canonical correlation analysis (CCA)
- Maximum covariance analysis (MCA).
- Linear inverse models (LIM)

Much literature on the merits of each!

Generally more appropriate over limited region where meteorology is well-understood.

Towards probabilistic prediction…

- **Deterministic** vs **probabilistic** prediction.
- Whereas an *individual* deterministic forecast can be often be judged right or wrong, this is not possible for probabilistic forecasts.
- For comparison with probabilistic output from dynamical model ensembles, empirical prediction must also be probabilistic.

- More in the next section…

Outline

1. Introduction
2. Teleconnections and the basis empirical seasonal prediction
3. Methods in empirical seasonal prediction
4. Empirical seasonal prediction in practice
5. Summary and outlook

Development of a global empirical prediction system

SPECS (Seasonal-to-decadal climate Prediction for the improvement of European Climate Services)

Work package 5.1: Building a prototype empirical prediction system for surface air temperature and precipitation:

- Seasonal prediction (led by KNMI)
- Decadal prediction (led by University of Reading)

Development of a global empirical prediction system

SPECS
Seasonal-to-decadal climate Prediction for the improvement of European Climate Services

Design brief:

- Global applicability with emphasis on Europe
- Exploit the long-term trend as an important source of skill
- Probabilistic output; act as a benchmark for dynamical forecast systems as well as a forecast system in its own right.

System built on lagged multiple linear regression.

- Seasonal (three-month) forecasts produced at one month lead time.
- Predictors: CO_2 equivalent; indices describing modes of variability, locally-varying predictors.
- Predictor selection based on physical principles to fullest extent.

Predictor selection and model fitting

CO_2 equivalent

Predictors describing internal variability

$$x = \alpha + \beta \boxed{C} + \sum_{i=1}^{n} (\Phi_i \boxed{F_i}) + \epsilon$$

Predictor selection and model fitting

CO_2 equivalent

Predictors describing internal variability

$$x = \alpha + \beta \boxed{C} + \sum_{i=1}^{n} (\Phi_i \boxed{F_i}) + \epsilon$$

- Nino3.4
- PDO
- AMO
- QBO
- IOD
- Local SST
- Persistence

Predictor selection and model fitting

CO_2 equivalent

Predictors describing internal variability

$$x = \alpha + \beta \boxed{C} + \sum_{i=1}^{n} (\Phi_i \boxed{F_i}) + \epsilon$$

Predictor selection

- Nino3.4
- PDO
- AMO
- QBO
- IOD
- Local SST
- Persistence

Predictor selection and model fitting

CO_2 equivalent

Predictors describing internal variability

$$x = \alpha + \beta \boxed{C} + \sum_{i=1}^{n} (\Phi_i \boxed{F_i}) + \epsilon$$

Predictor selection

- Nino3.4
- PDO
- AMO
- ~~QBO~~
- ~~IOD~~
- ~~Local SST~~
- Persistence

Predictor selection and model fitting

CO_2 equivalent

Predictors describing internal variability

$$x = \alpha + \beta \boxed{C} + \sum_{i=1}^{n} (\Phi_i \boxed{F_i}) + \epsilon$$

Nino3.4

PDO

AMO

~~QBO~~

~~IOD~~

~~Local SST~~

Persistence

Predictor selection

$$x = \alpha + \beta C + \sum_{i=1}^{k} (\Phi_i \boxed{F_i^S}) + \epsilon$$

Predictor selection and model fitting

CO_2 equivalent

Predictors describing internal variability

$$x = \alpha + \beta \boxed{C} + \sum_{i=1}^{n} (\Phi_i \boxed{F_i}) + \epsilon$$

Nino3.4

PDO

AMO

~~QBO~~

~~IOD~~

~~Local SST~~

Persistence

Predictor selection

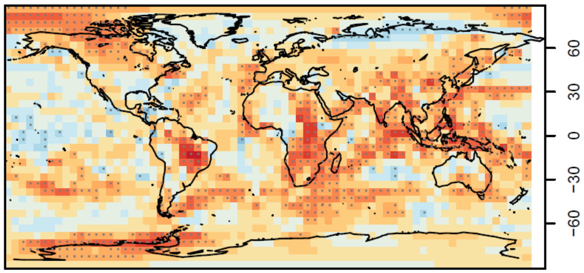
$$x = \alpha + \beta C + \sum_{i=1}^{k} (\Phi_i \boxed{F_i^S}) + \boxed{\epsilon}$$

Residuals from model fit randomly sampled and added to estimate to produce forecast ensemble

Forecast generation and verification

- Hindcasts produced for 1961-2013 compared with observations.
- Data from all previous years used to generate hindcast.
- System design facilitates framework used in verification of dynamical system output.
- Deterministic output:
  - Correlation
  - RMSE
- Probabilistic output:
  - Verification skill scores, including continuous rank probability skill score (CRPS) (R package: SpecsVerification)
  - Skill scores generated with respect to a reference forecast, generated from randomly sampling the climatology (sample size = 51).

Predictor selection

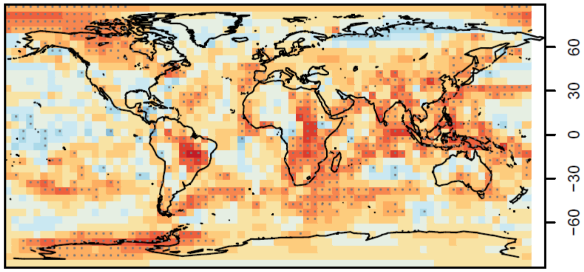


CO2EQV

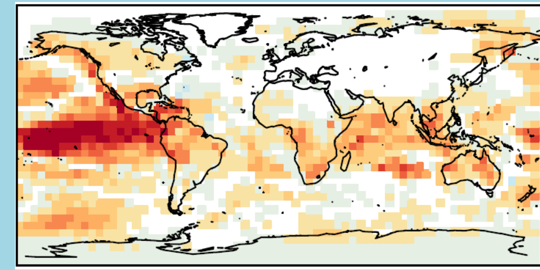
Correlation between observations predictions for DJF surface air temperature (1961-2014).

Linear regression model with one predictor: CO_2-equivalent (CO2EQV).

Predictor selection



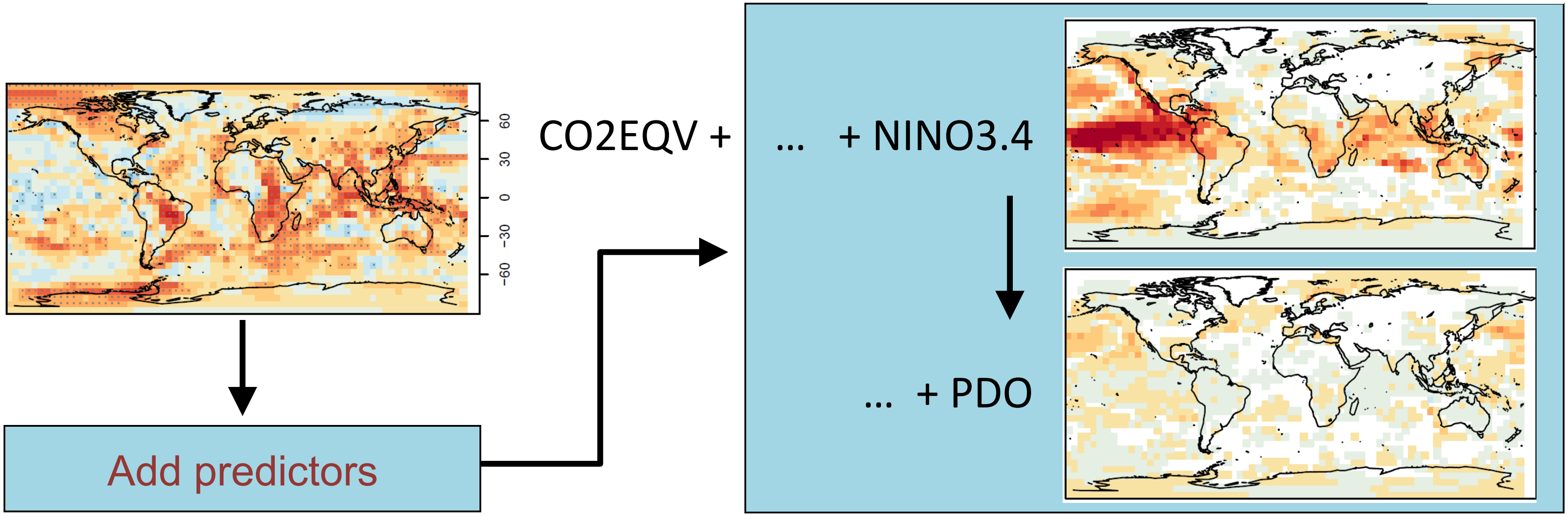
CO2EQV + … + NINO3.4



Add predictors

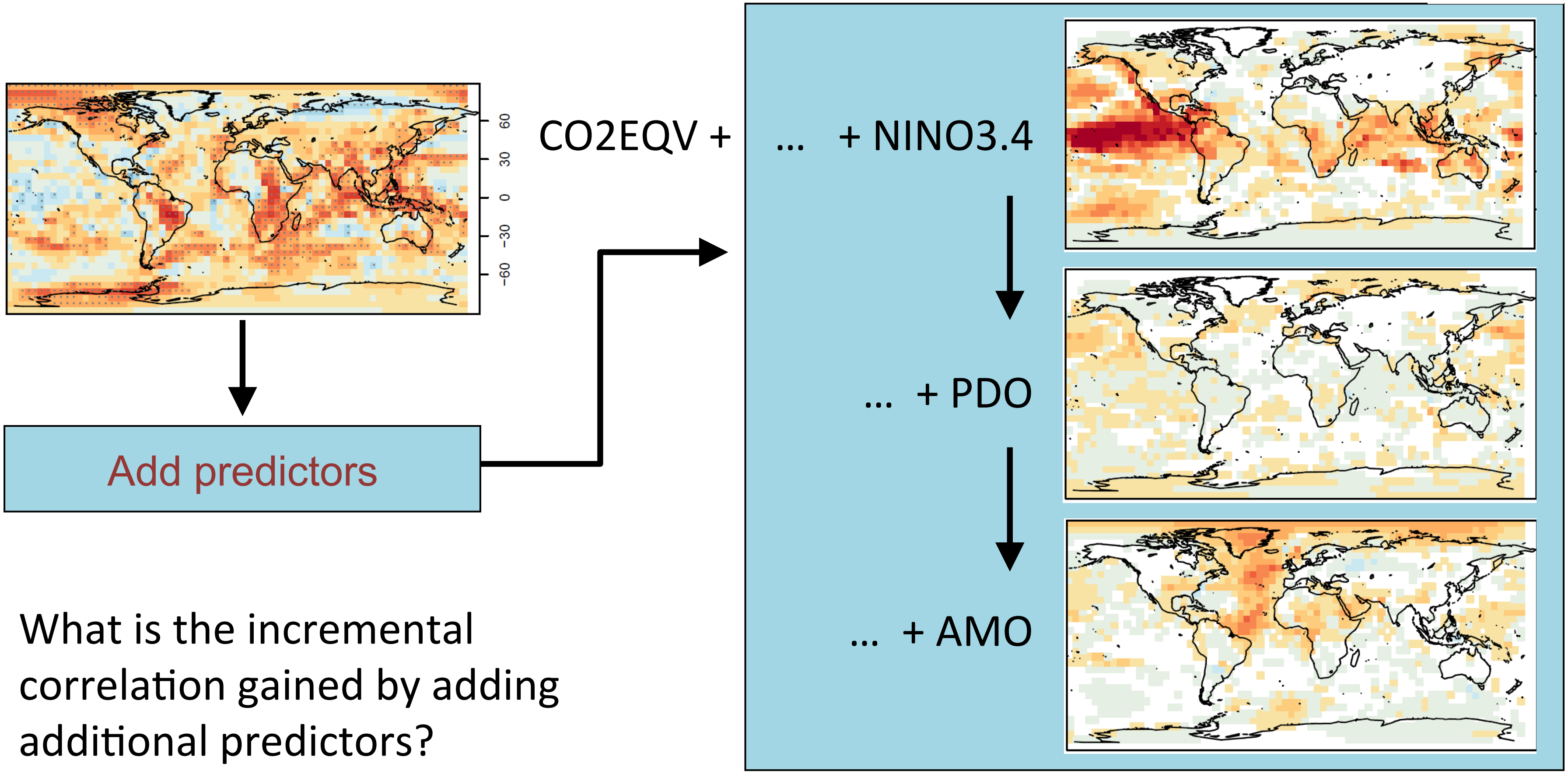
What is the incremental correlation gained by adding additional predictors?

Predictor selection



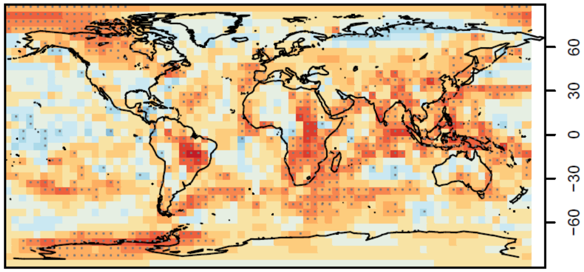
What is the incremental correlation gained by adding additional predictors?

Predictor selection



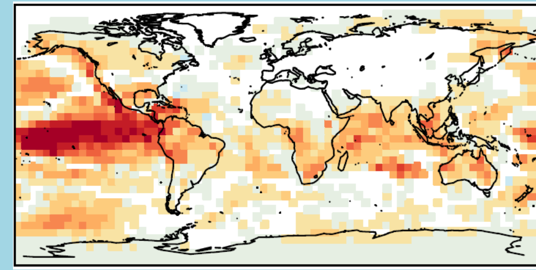
What is the incremental correlation gained by adding additional predictors?

Predictor selection

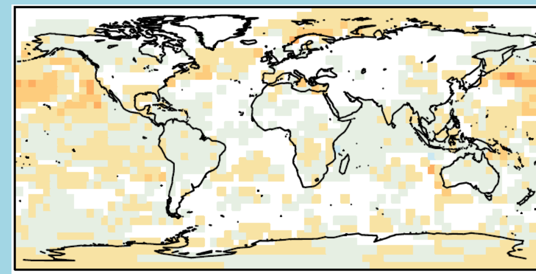


Add predictors

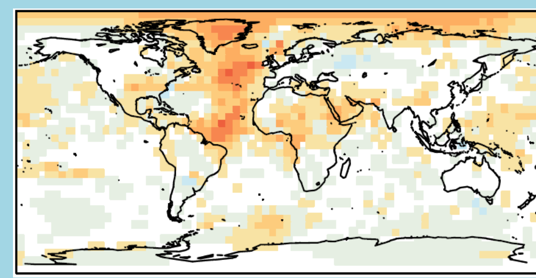
CO2EQV + … + NINO3.4



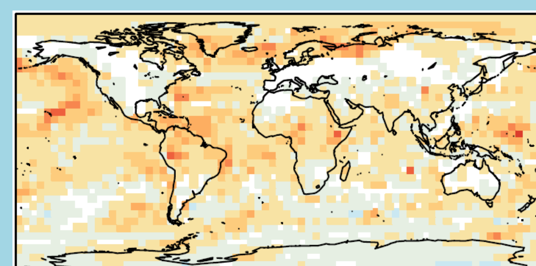
… + PDO



… + AMO

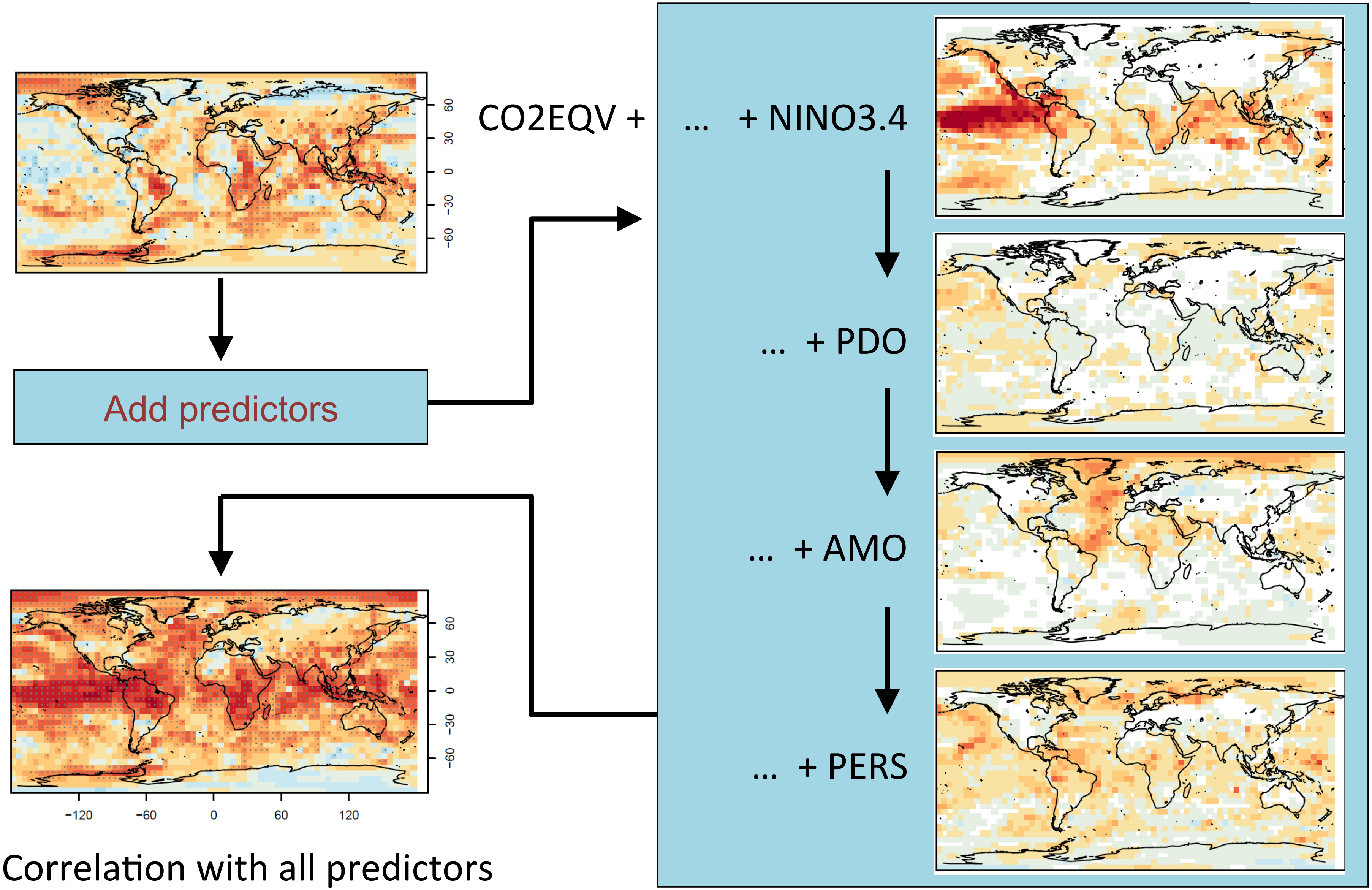


… + PERS



What is the incremental correlation gained by adding additional predictors?

Predictor selection

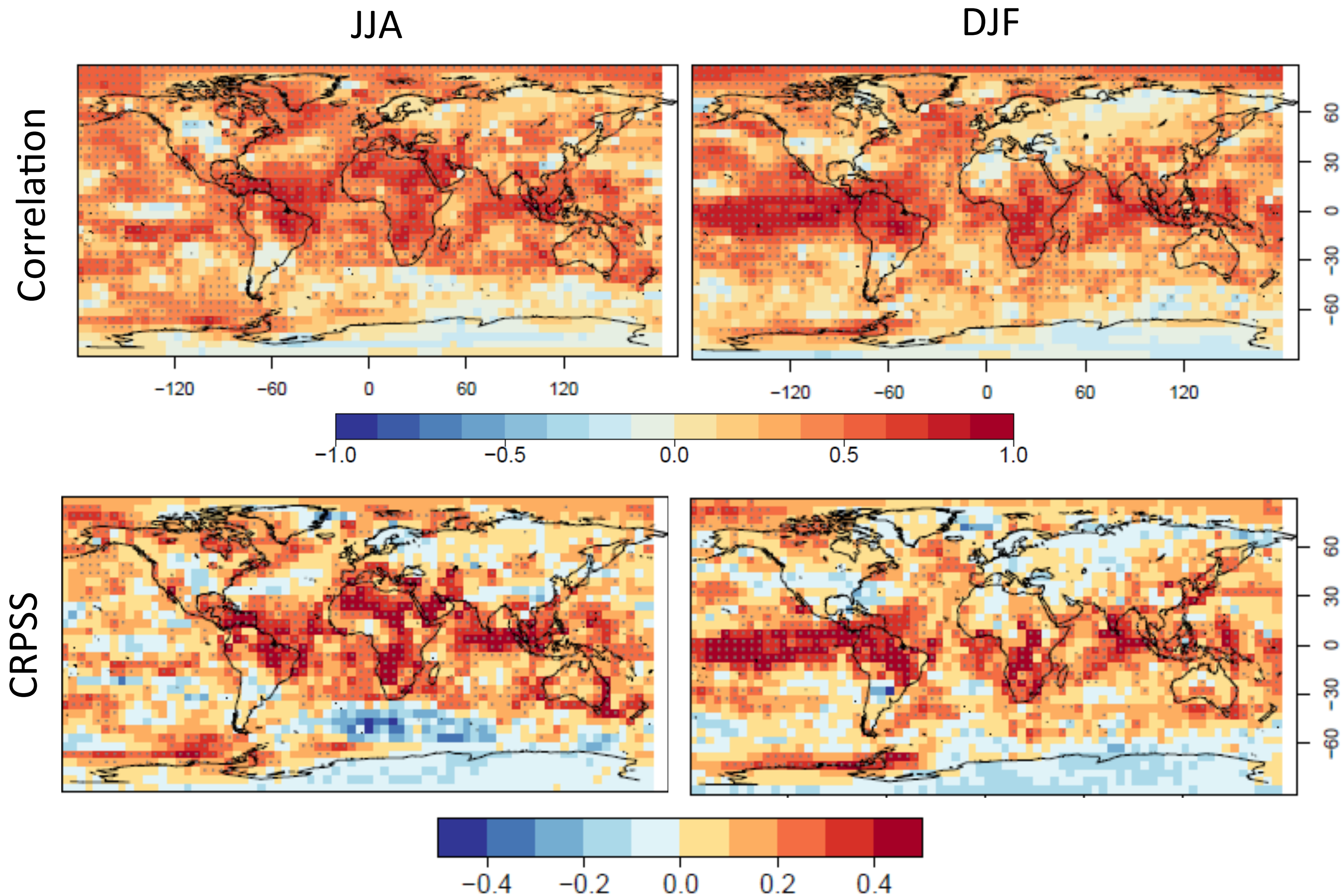


Correlation with all predictors

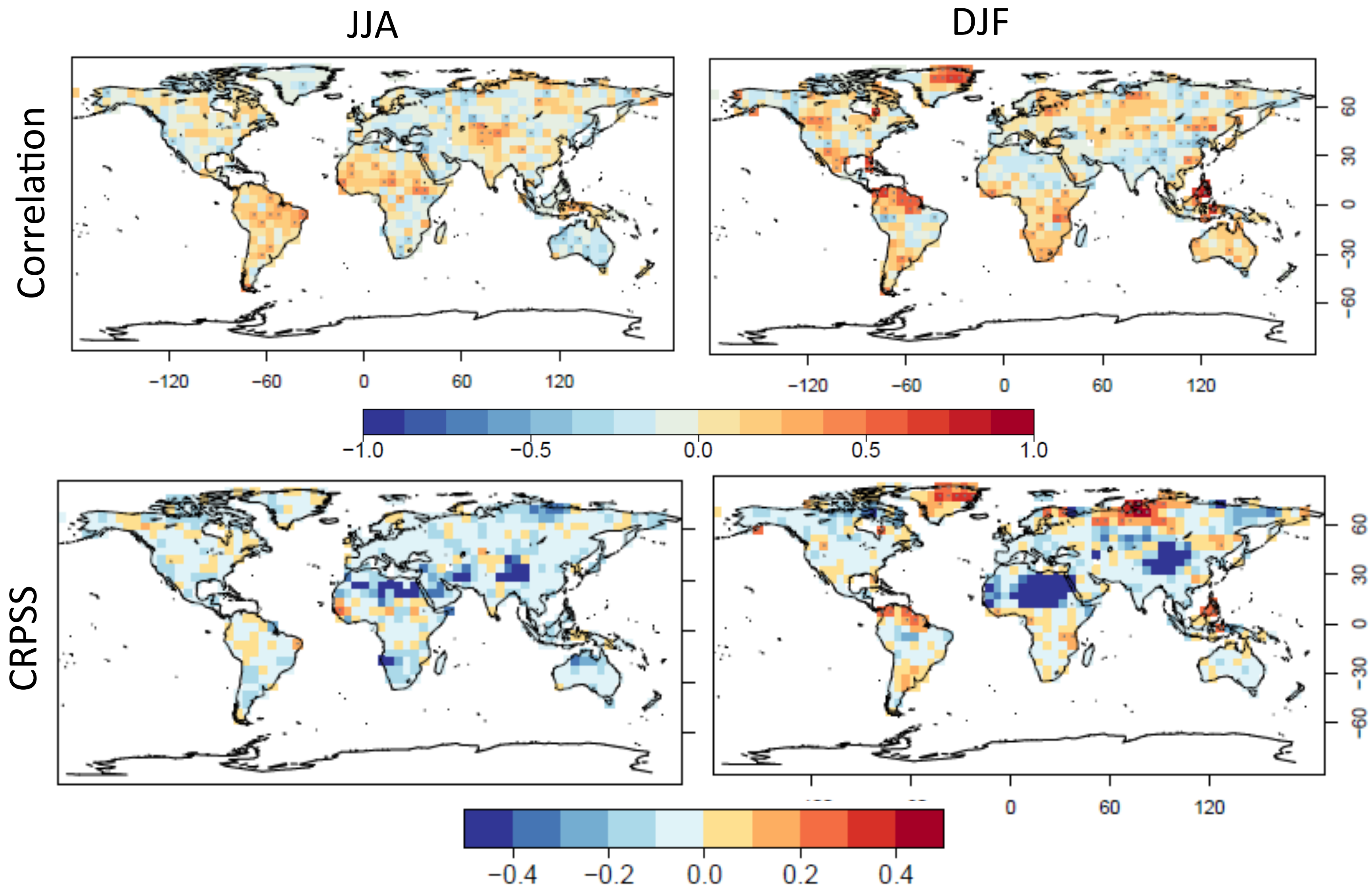
Forecast generation and verification

- Hindcasts produced for 1961-2013 compared with observations.
- Data from all previous years used to generate hindcast.
- System design facilitates framework used in verification of dynamical system output.
- Deterministic output:
  - Correlation
  - RMSE
- Probabilistic output:
  - Verification skill scores, including continuous rank probability skill score (CRPS) (R package: SpecsVerification)
  - Skill scores generated with respect to a reference forecast, generated from randomly sampling the climatology (sample size = 51).

Forecast skill: surface air temperature



Forecast skill: precipitation



Prediction system: summary

Prediction system produces good skill in many regions

SAT: Interannual variability is well-represented throughout the tropics and in a number of extra-tropical regions:

- Parts of Europe, particularly during spring and summer
- Southern Africa, eastern Australia
- Spatial for correlation and CRPS patterns broadly similar

PRECIP: Few areas of notable skill are found

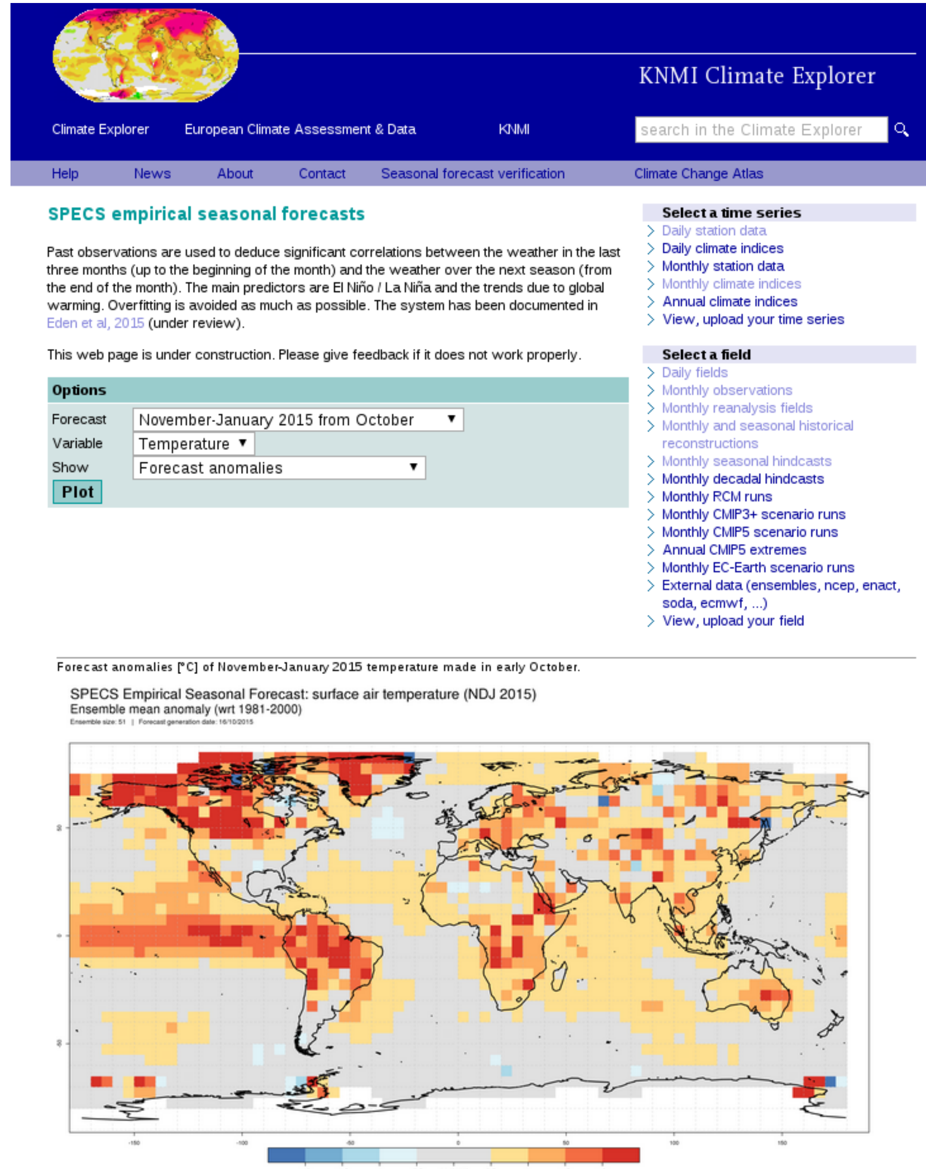
- Correlation in regions with known ENSO teleconnection is strong.
- Probabilistically, system does not perform better than the climatological ensemble throughout most of the world

Forecast dissemination

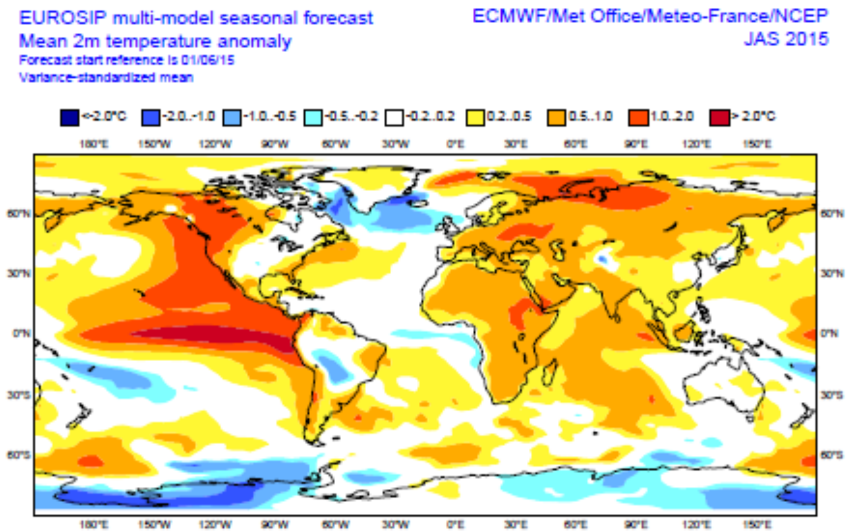
- Forecasts made as soon as predictor data is made available.
  - In practice, between 5th and 10th of each month.
  - e.g. Forecast for November, made between 5-10th October using predictor data from July-August-September (i.e., one month lead).
- Forecasts and verification statistics made available on KNMI Climate Explorer.

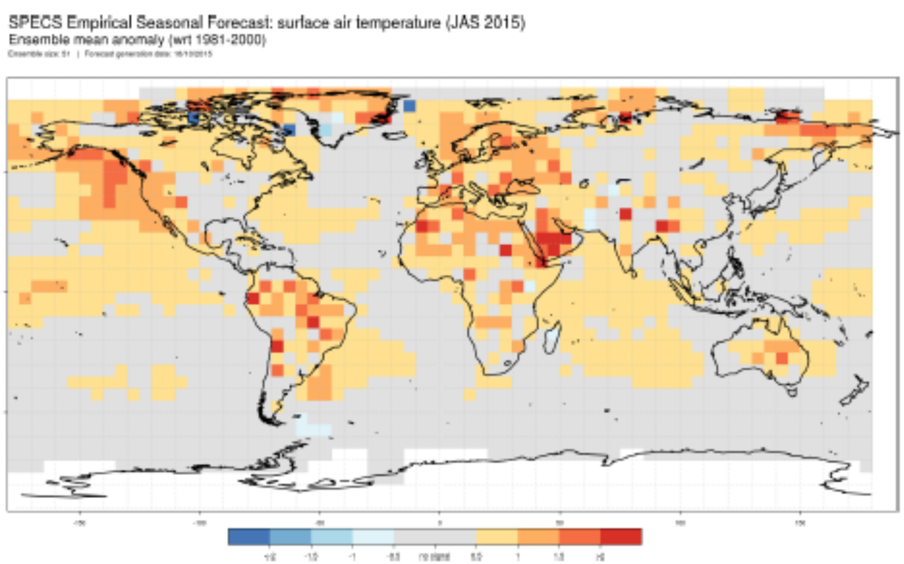
  climexp.knmi.nl/spes.cgi

- Forecast/hindcast netcdf files available for download.

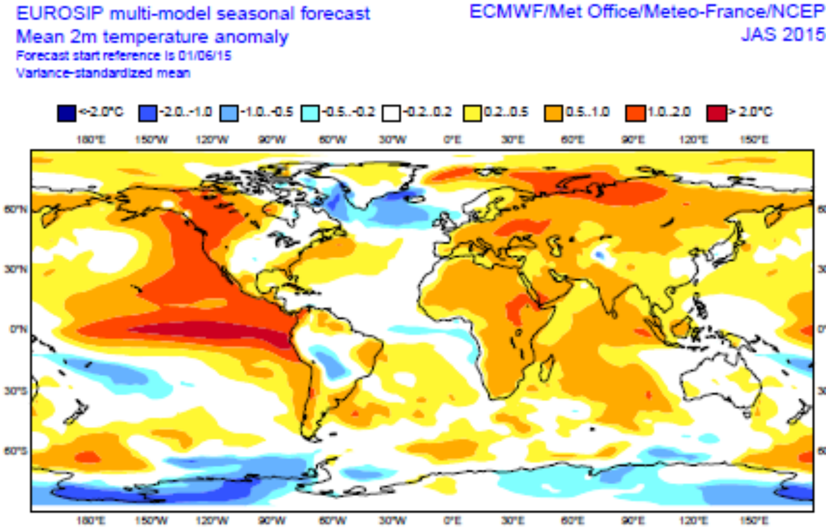


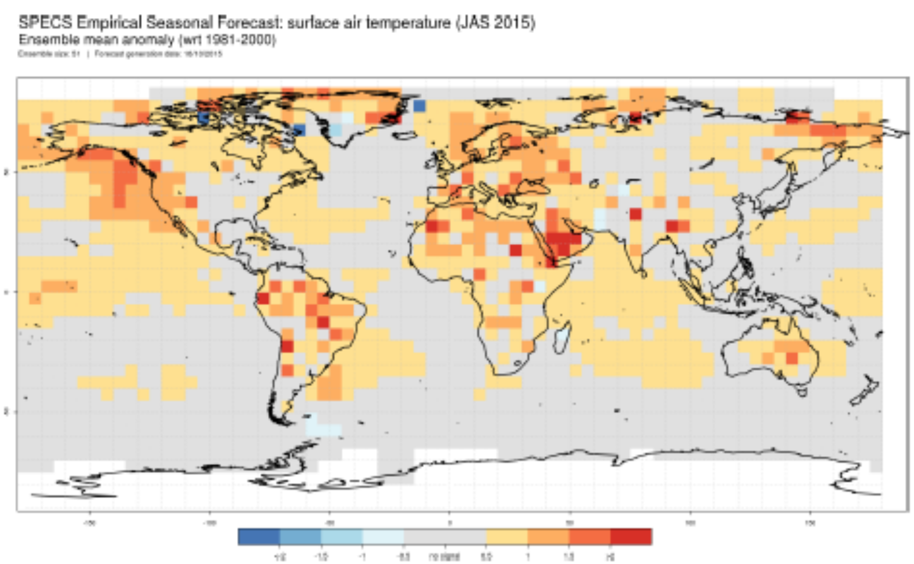
Comparison with dynamical forecasts

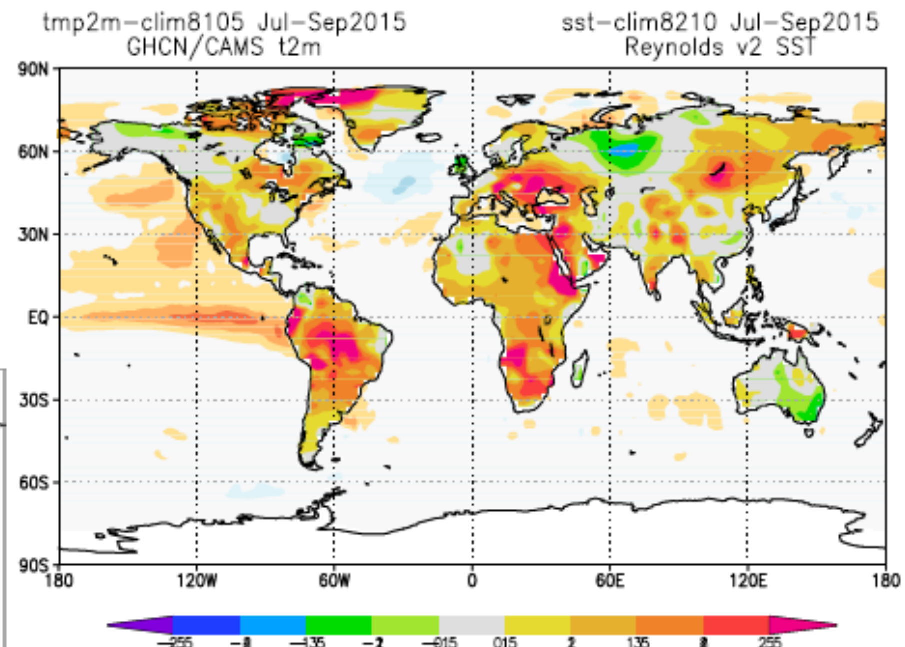




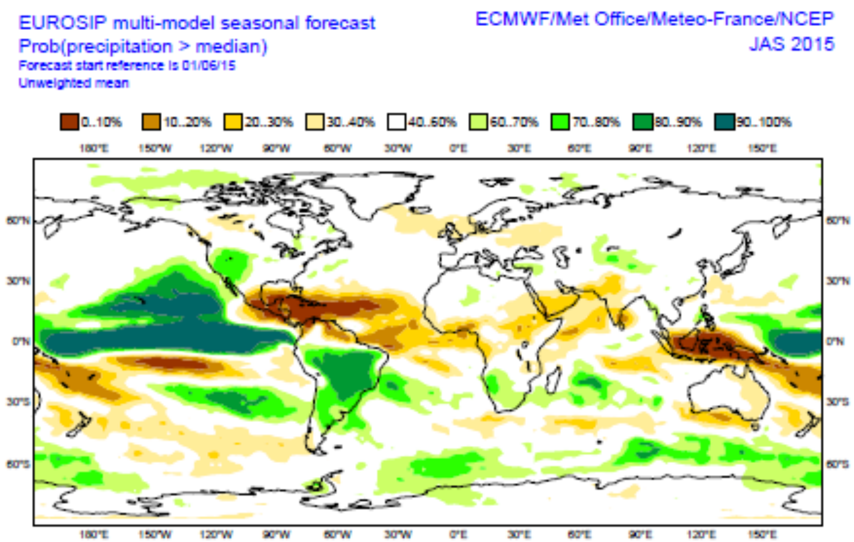
Comparison with dynamical forecasts

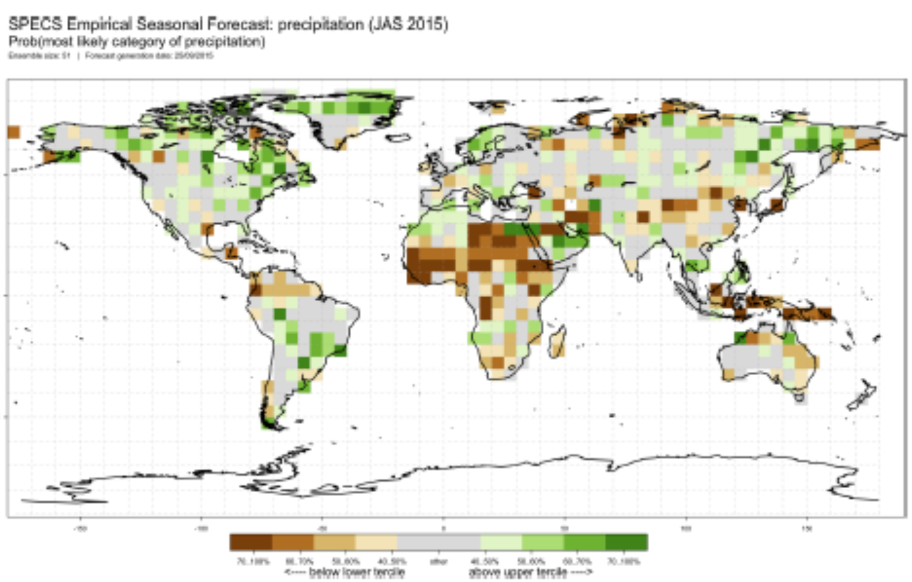




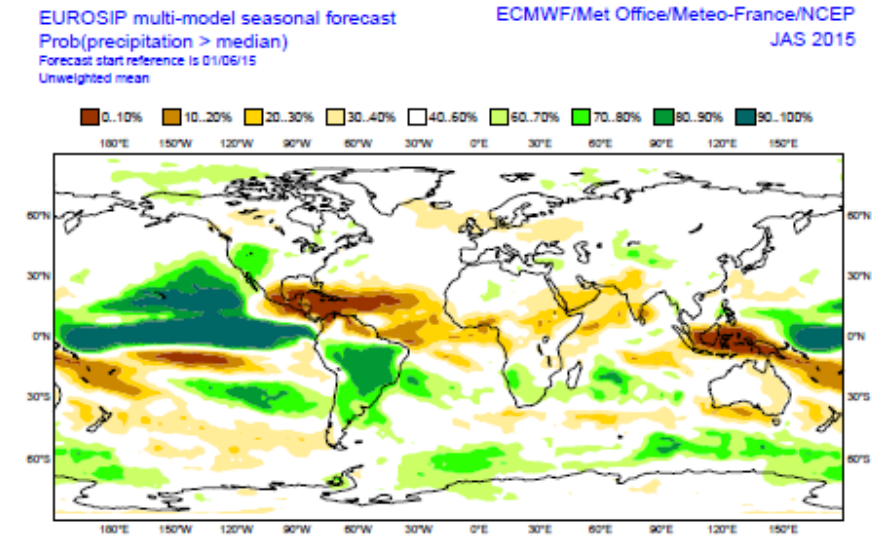


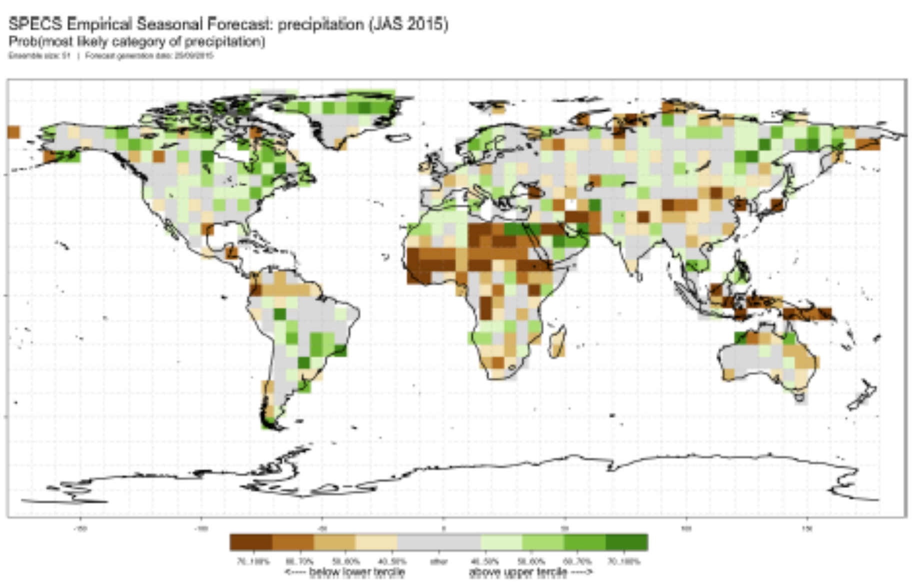
Comparison with dynamical forecasts

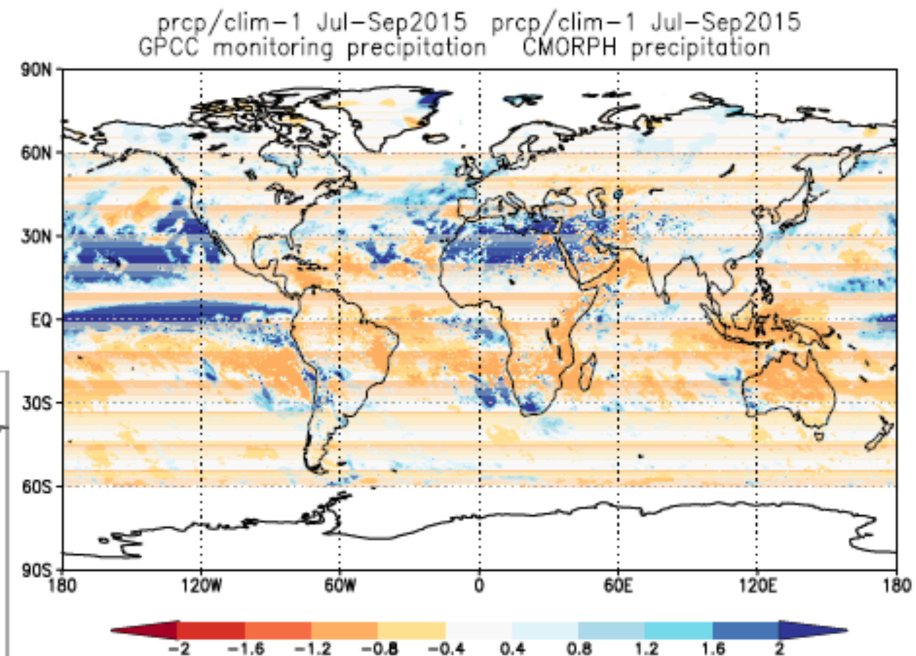




Comparison with dynamical forecasts







Further development

- Prediction system presented and validated in Eden et al. (2015), Geosci. Model Dev. (in review)
- Further model development based on feedback from users.
  - Alternative methods of ensemble generation.
  - Additional forecast information to made available via Climate Explorer.
  - Higher resolution forecasts forthcoming.

- A model for decadal prediction built with the same principles has also been developed (Suckling *et al*., Clim. Dynamics., submitted).

Outline

1. Introduction
2. Teleconnections and the basis empirical seasonal prediction
3. Methods in empirical seasonal prediction
4. Empirical seasonal prediction in practice
5. Summary and outlook

Summary and outlook

- Dynamical models are process-based and can resolve non-linear terms – an inherent advantage over empirical models that are (usually) linear.
- However, when the processes involved in seasonal prediction are also linear, dynamical methods may have no clear advantage.
- Thus, empirical methods continue to play an important role:
  - As an often credible alternative to dynamical models.
  - As a benchmark for aiding the development of dynamical models.

- Is there scope for greater communication and collaboration among dynamical and empirical forecast-makers?

Thank you

jonathan.eden@knmi.nl

@Jonathan_Eden

Selected references

Doblas-Reyes, F. J. et al. 2013. Seasonal climate predictability and forecasting: status and prospects, Wiley Interdisciplinary Reviews-Climate Change, 4, 245–268, doi:10.1002/wcc.217.

Dool, H., van den. 2007. Empirical methods in short-term climate prediction, Oxford University Press.

Eden, J. M. et al. 2015. A global empirical system for probabilistic seasonal climate prediction, Geosci. Model Dev. Discuss., 8, 3941-3970, doi:10.5194/gmdd-8-3941-2015.

Eden, J. M. et al. 2014. Pacific SST influence on spring precipitation in Addis Ababa, Ethiopia. Int. J. Climatol., 34: 1223–1235. doi: 10.1002/joc.3759

Oldenborgh, G. J. van et al. 2005. Did the ECMWF seasonal forecast model outperform statistical ENSO forecast models over the last 15 years?, J. Clim., 18, 3240–3249.