Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA
Federal Office of Meteorology and Climatology MeteoSwiss

Forecast verification

4th VALUE Training School

Jonas Bhend, **Sven Kotlarski**

Forecast verification is

the process of comparing forecasts with relevant observations to assess the forecast *quality* (not *value*).

Outline

1. Projections and predictions
2. Rationale for forecast verification
3. How to verify forecasts
    1. Types of forecasts
    2. Aspects of forecast quality and scores
    3. Skill
4. Seasonal forecasting
    1. Success stories
    2. Verification at MeteoSwiss
5. The easyVerification R package

Glossary

take care:
Sometimes «forecast» is used as the genral term for predictions and projections!

Climate prediction
A climate prediction or climate forecast is the result of an attempt to produce (starting from a particular state of the climate system) an estimate of the actual evolution of the climate in the future

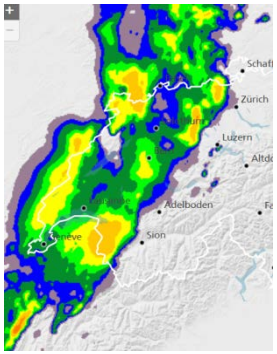
VALUE, climate downscaling

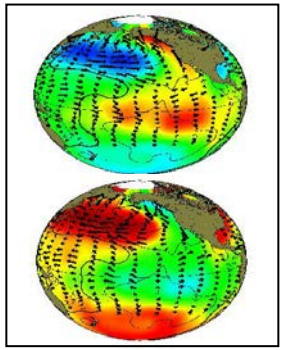
Projection
A projection is a potential future evolution of a quantity or set of quantities, often computed with the aid of a model. Unlike predictions, projections are conditional on assumptions concerning, for example, future socioeconomic and technological developments that may or may not be realized.

Important commonalities!

From forecasts to projections

A question of **time scale!**





NWP **monthly** **seasonal** **decadal** **multi-decadal**

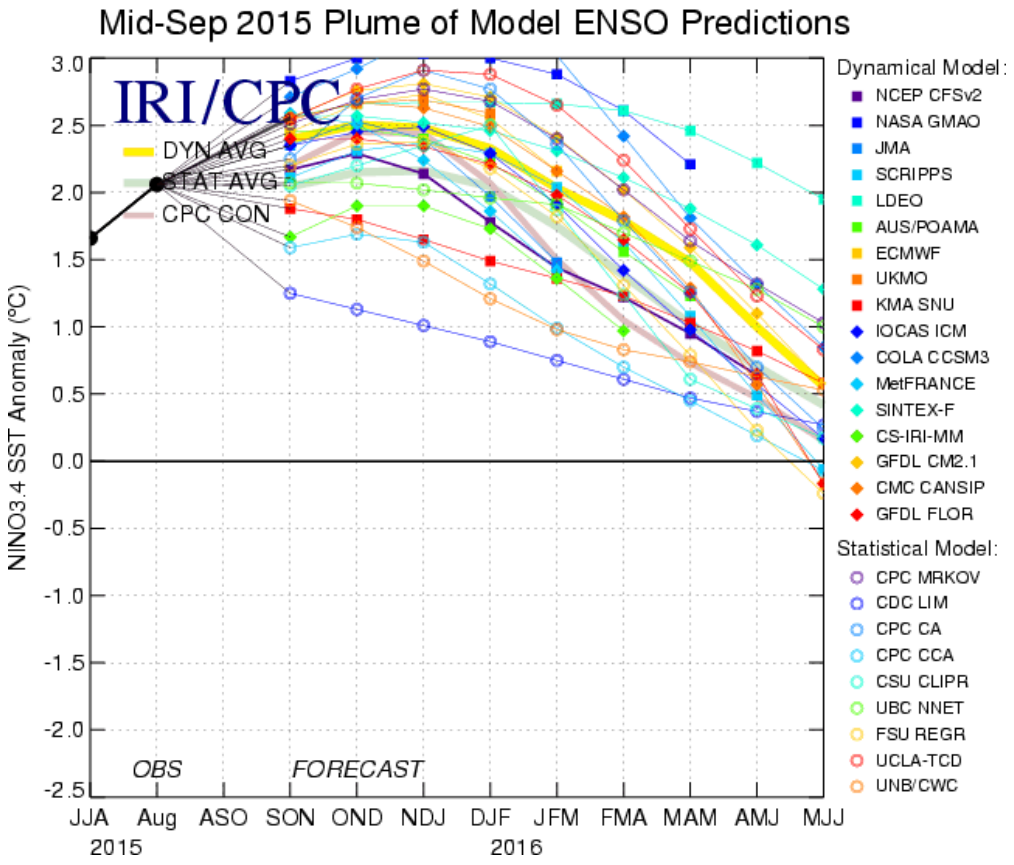
Initial condition predictability

Boundary forcing predictability

Prediction / forecast
Initialized with observed state

Projection
initialized with plausible state

Seasonal forecasting systems

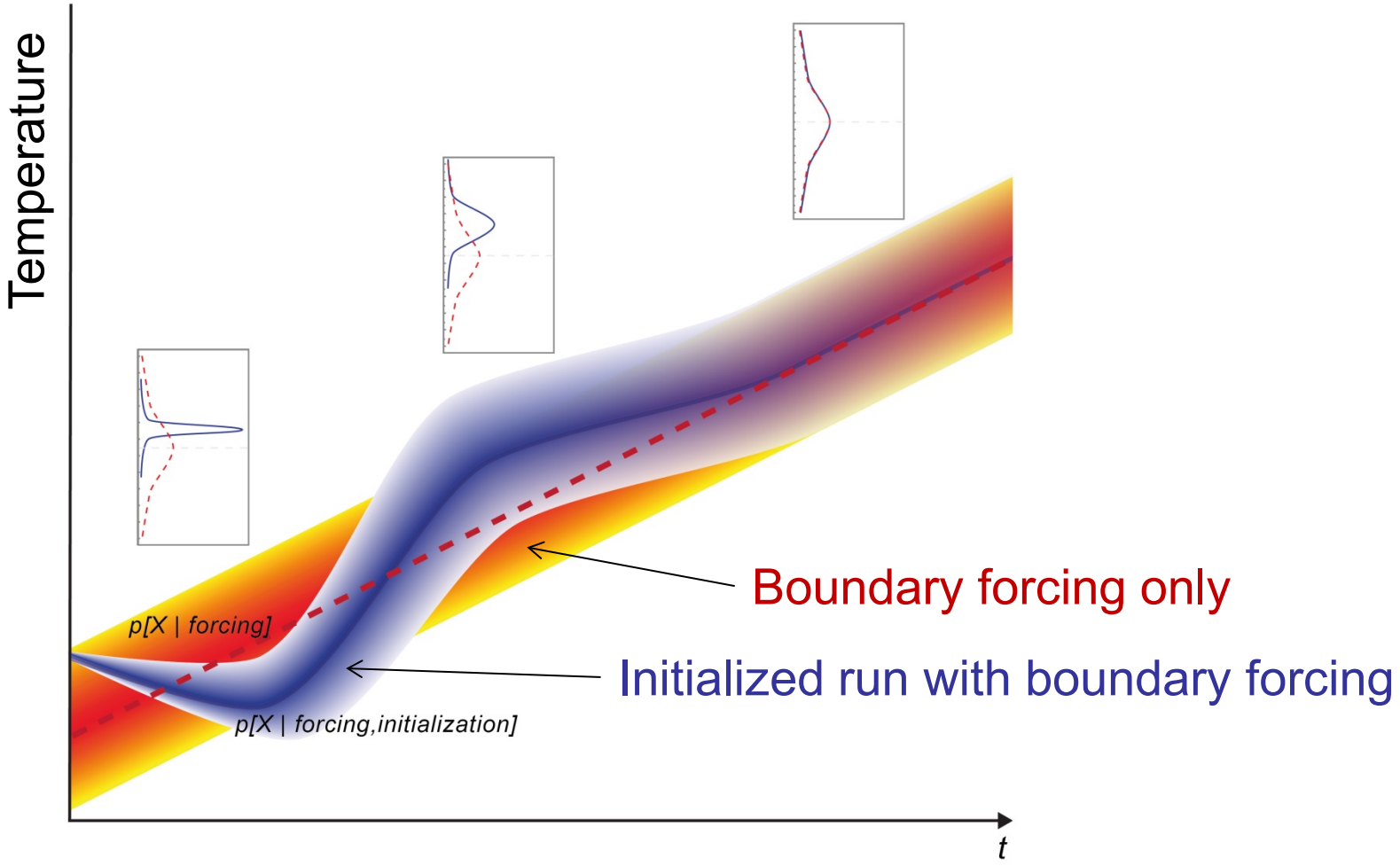


Seasonal forecasts are operationally produced using **statistical** and **dynamical** models

Dynamical models usually are closely related to either NWP (ECMWF) or climate models (GFDL)

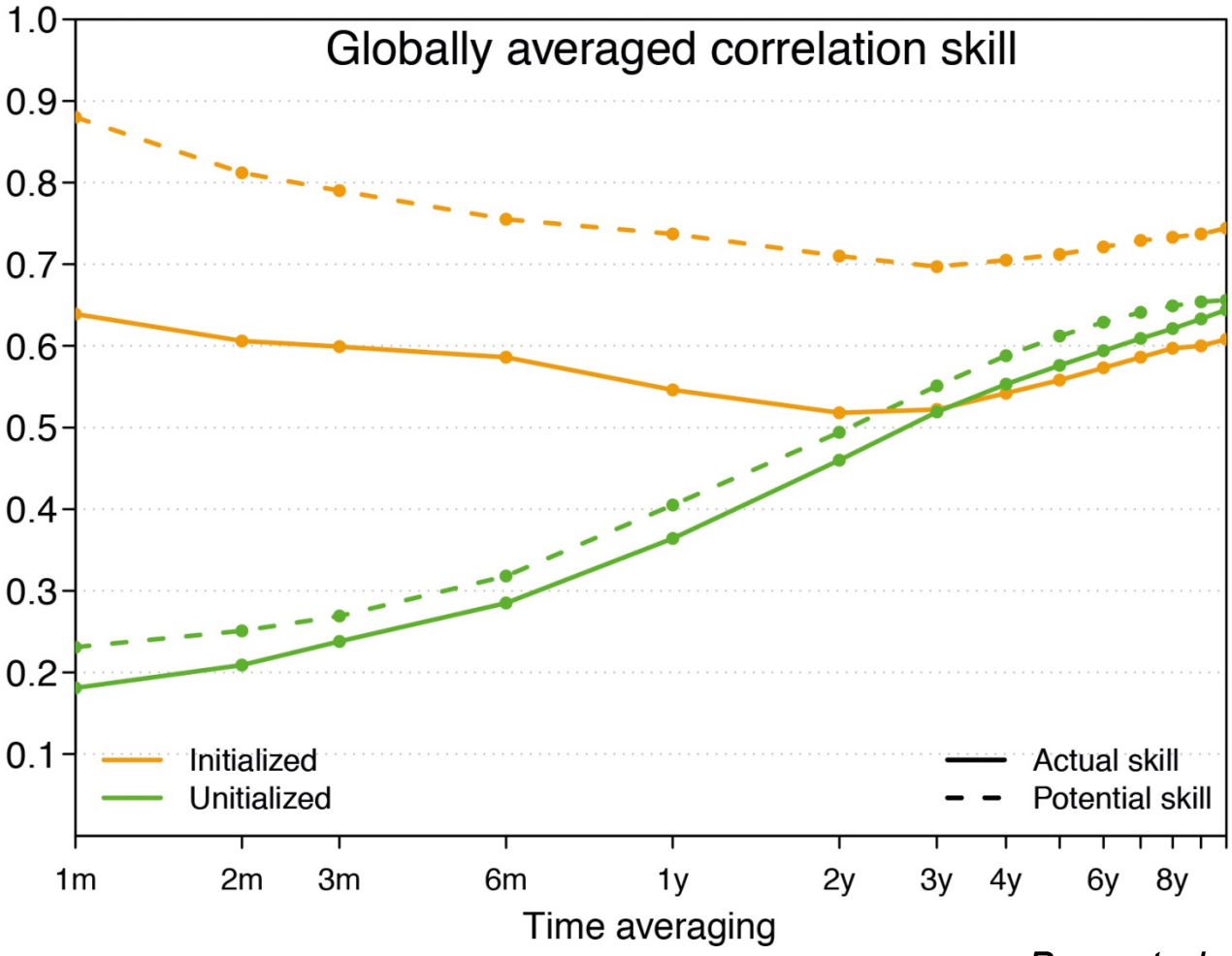
The European model, ECMWF System 4, corresponds to a previous version of the IFS (frozen because of hindcasts)

Predictability



Branstator and Teng, 2010; IPCC WG1

Predictability



Boer et al. 2013; IPCC WG1

Types of forecasts

Nature of forecasts

Deterministic: Temperature tomorrow,18°C

Probabilistic: The probability of tomorrow's temperatures exceeding 18°C is 60% -> **Model ensembles!**

Specificity of forecasts

Dichotomous (yes/no): Tomorrow it will rain

Multi-category: The rainfall tomorrow will be above average

Continuous: 15mm of rain

Time series, spatial distribution, spatio-temporal distribution?

hands-on

Why verify?

- **Administration**
  - track performance of forecasting system
  - ideally one metric to summarize forecast performance

- **Science**
  - understand predictability of forecast (and real) systems
  - plethora of verification metrics

- **Economy**
  - assess benefit of using forecasts in decision-making
  - verification metrics tailored to user needs

Attributes of forecast quality

(e.g., Murphy 1993; www.cosmo-model.org)

Bias – Overall (average) error in the forecasts, i.e. the correspondence between the mean forecast and mean observation.

Association - The strength of the linear relationship between the forecasts and observations (for example correlation coefficient)

Accuracy – Average degree of correspondence between an individual forecast and observations. The difference between the forecast and the observation is the *error* (e.g., RMSE). The lower the errors, the greater the accuracy.

Skill - the **relative** accuracy of the forecast over some reference forecast. The reference forecast is generally an unskilled forecast such as random chance, persistence (defined as the most recent set of observations, "persistence" implies no change in condition), or climatology.

Reliability – Measure of how closely the forecast probabilities correspond to the conditional frequency of occurrence of an event (PDFs).

Attributes of forecast quality

Resolution - measure of how well the observations are “sorted” among the different forecasts. Even if the forecasts are wrong, the forecast system has resolution if it can successfully separate one type of outcome from another.

Sharpness - Degree of “spread” or variability in the forecasts. While probability forecasts vary between 0 and 1, perfect forecasts only include the two end points, 0 and 1. Sharper forecasts will tend toward values close to 0 and 1. Sharpness is a property of the forecast only, and like resolution, a forecast can have this attribute even if it's wrong (in this case it would have poor reliability).

Discrimination - Measure of how well the forecasts discriminate between events and non-events. Ideally, the distribution of forecasts in situations when the forecast event occurs should differ from the corresponding distribution in situations when the event does not occur.

Uncertainty - The variability of the observations. The greater the uncertainty, the more difficult the forecast will tend to be.

Selection from zoo of metrics

Attributes of forecast quality (Murphy, 1993)	Deterministic	Ensemble forecasts (probabilistic)
Bias	Mean error	Mean error (of ensemble mean)
Association	Correlation	Correlation (with ensemble mean)
Accuracy	Mean square error, Mean absolute error	Continuous rank probability score, Ignorance score
Reliability	Reliability diagram	Reliability diagram, Spread to error ratio, Rank histogram
Resolution	ROC area	ROC area
Sharpness	Variance of forecasts	Ensemble spread
Discrimination	Generalized discrimination score	Generalized discrimination score

Skill of forecasts

(to measure the relative accuracy)

score, such as RMSE or Brier score (the larger the worse)

Relative accuracy of the forecasting system (A_{fcst}) compared with the accuracy of a reference system (A_{ref})

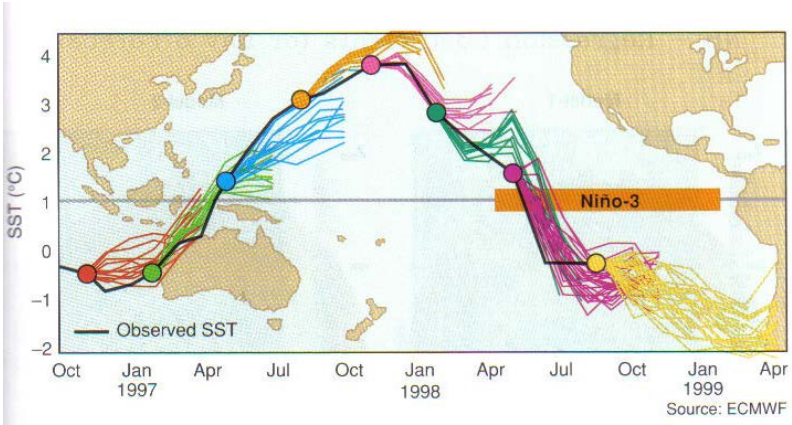
$$AS = 1 - \frac{A_{fcst}}{A_{ref}}$$

- Forecast has skill: $AS > 0$
- Forecast as accurate as reference (no skill): $AS = 0$
- Forecast worse than reference: $AS < 0$

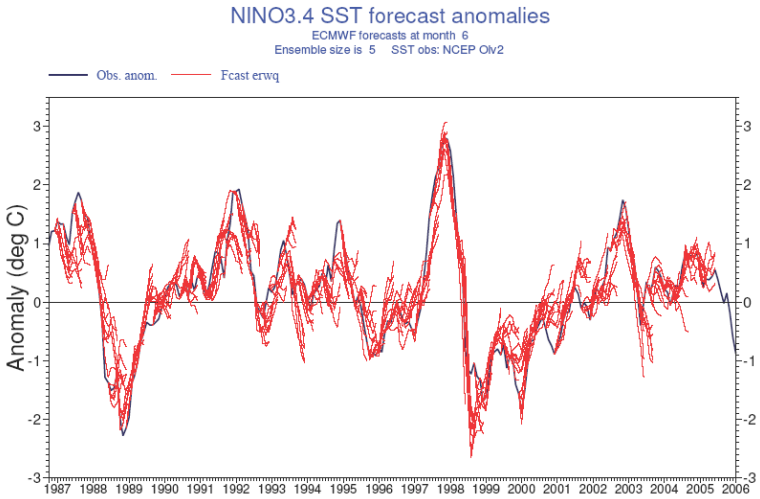
Climatological or persistence forecasts are often used as reference
→ Compared to climatology, a transient GCM run has skill

Skill can be used to compare forecasting systems
→ Accuracy of **downscaled** data vs. GCM data

How well can we predict ENSO one or two seasons ahead ?

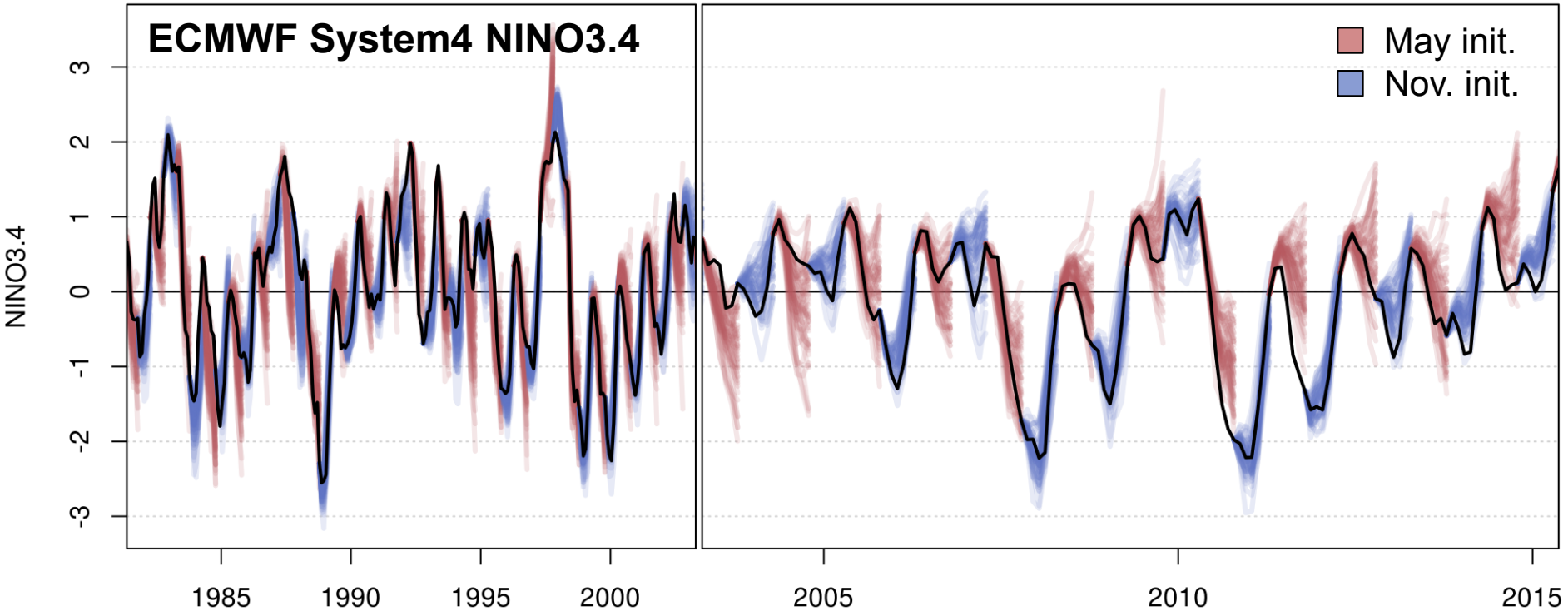


The success story: Dynamical El Niño forecast 97/98



What is the overall forecast skill? Look to at least 20 years or longer ECMWF System 3: El Niño

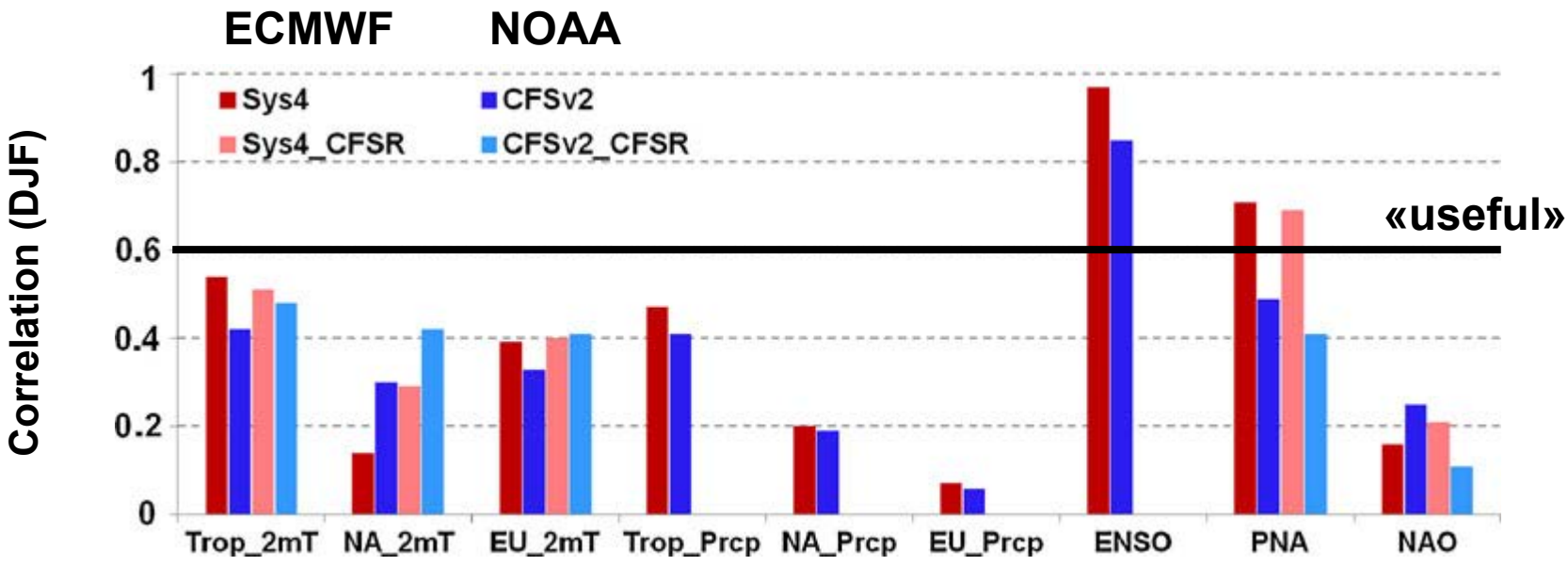
How well can we predict ENSO one or two seasons ahead ?



- **Skill especially of winter forecasts (El Nino is predictable!)**
- **No large improvement since 1990s**

Forecast skill of current seasonal forecasting systems

Correlation of 3-month mean (DJF), re-forecasts initialized 1st Nov

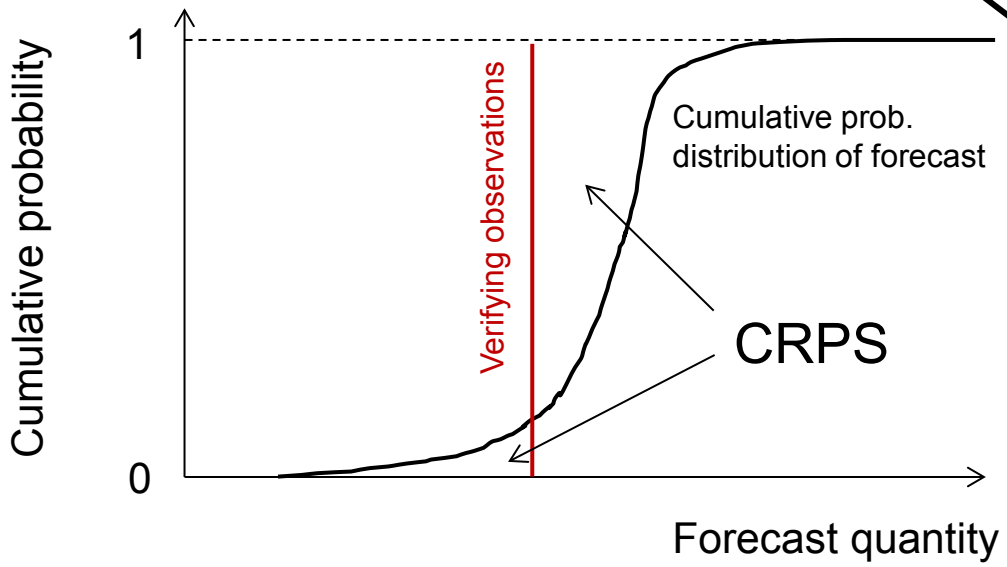


Kim et al., 2012

Verification and recalibration at MeteoSwiss

The Continuous Ranked Probability Score (CRPS)

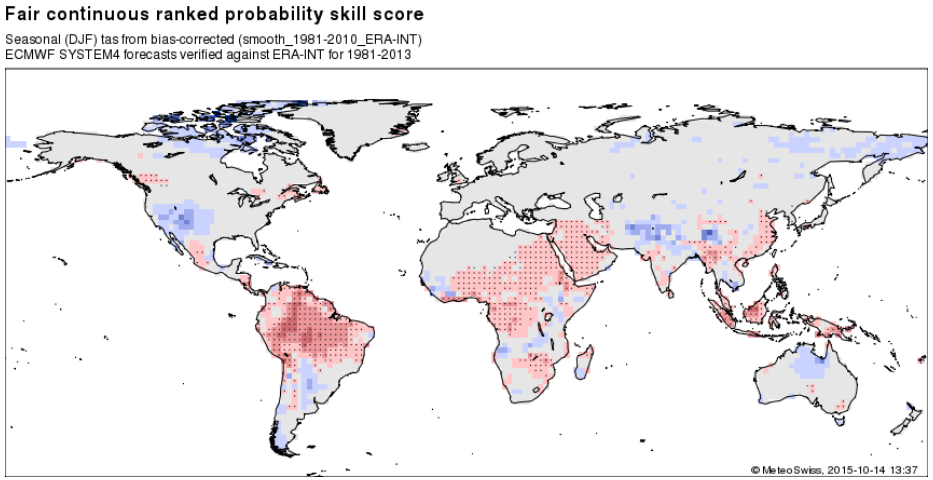
- Mean absolute error for ensemble forecasts
- Sensitive to both bias and reliability



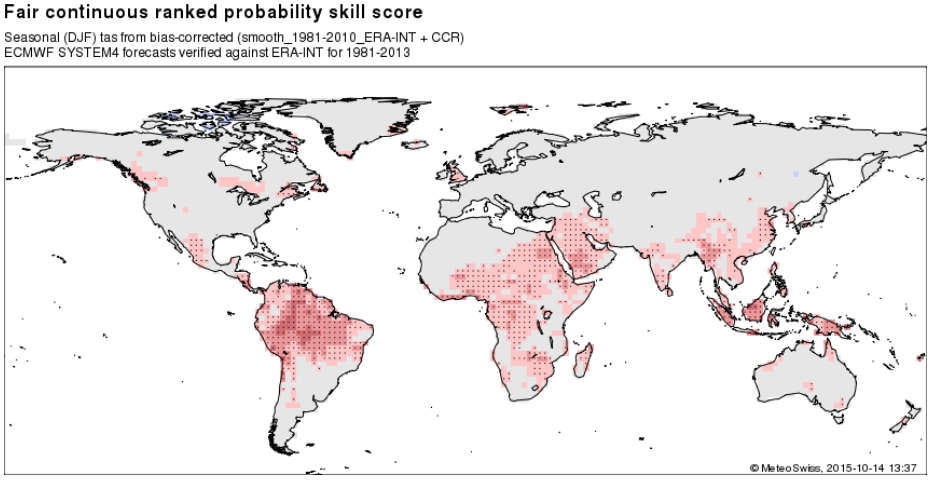
reliable forecast: forecasted probabilities are in agreement with observed probabilities

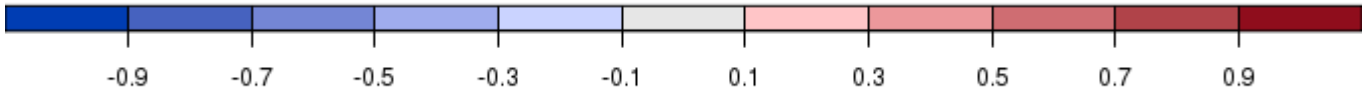
CRPS for calibrated (bias corrected) and recalibrated forecasts

Calibrated forecast



Recalibrated forecast (Weigel et al. 2009)





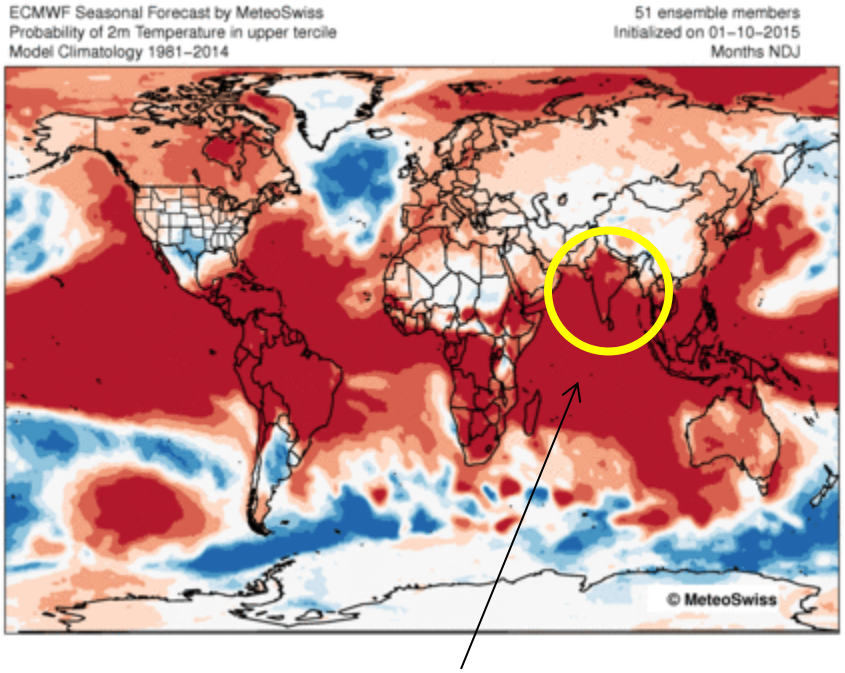
Worse than climatological forecast

Better than climatological forecast

Weigel, et al. (2009). Seasonal ensemble forecasts: Are recalibrated single models better than multimodels? *Monthly Weather Review*

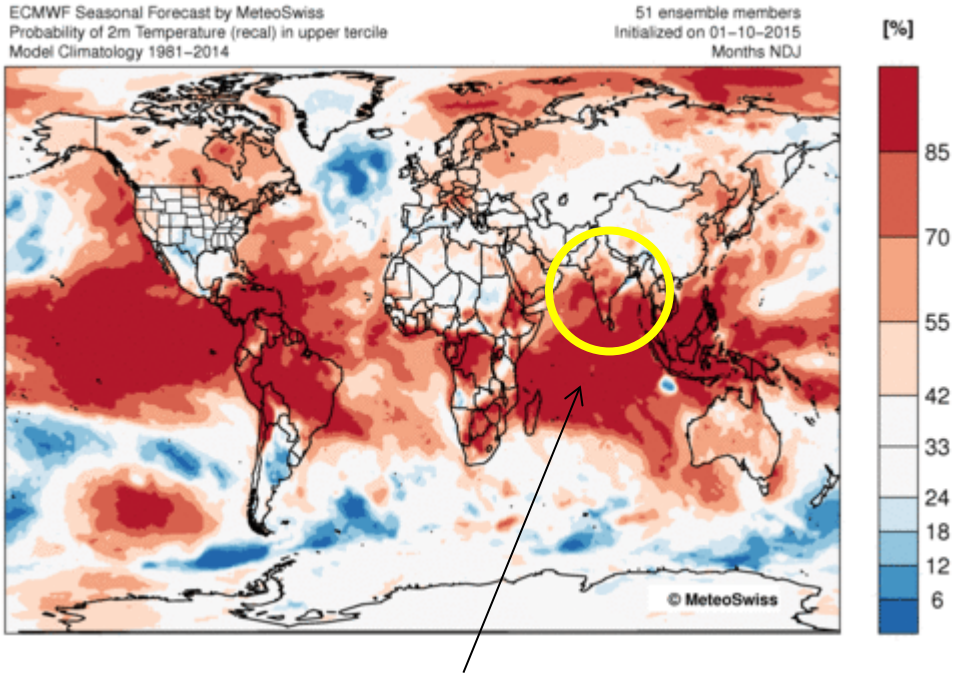
Examples of raw and recalibrated forecasts

Raw forecast



Raw forecast is certain about warming in India

Recalibrated forecast



Recalibrated forecast is much less certain

Conclusions

- Forecast quality is multi-faceted
- Forecast skill depends on:
  - Variable
  - Lead time
  - Region
  - Time of year
  - Spatio-temporal aggregation
  -

easyVerification

R package for verification of ensemble forecasts

Design goals:

- **Easy to use**
  - One wrapper to apply verification functions to large datasets
  - **Operational application**: Supply ensembles of continuous forecasts and observations, conversion to categories etc. handled internally
- **Flexible**
  - Can use third-party verification code (e.g. SPECS, user)
  - Supports a variety of array-based data structures
- **Convenience and flexibility over speed**
  - Vectorization used where possible but not extensively
  - Multicore parallelized execution available on *NIX systems

easyVerification

Scores and skill scores

Deterministic (ens. mean)	Probabilistic
Correlation	2AFC
Mean error (bias)	ROC area*
Mean absolute error	Spread to error ratio
Mean squared error	(fair) CRPS*
	(fair) RPS*
	Dressed Ignorance, CRPS

in easyVerification
from SpecsVerification
* with significance

new:
- normalization by ensemble size
- significance tests

easyVerification

example use – technical detail

verification functions:

- vector of obs.
- matrix of forecasts
- output vector, scalar, or list

veriApply, the workhorse:

- reformat inputs and outputs
- convert inputs to required data format internally (e.g. category forecasts)
- reference forecast for skill scores

```
library(easyVerification)
```

```
## Loading required package: SpecsVerification
## Loading required package: RCurl
## Loading required package: bitops
## WARNING: Your current version of easyVerification is not up-to-date
## Get the latest version 0.1.4.1 using
install_github("MeteoSwiss/easyVerification")
```

Forecast ensemble with 100 spatial instances, 30 forecasts and 15 ensemble members

```
fcst <- array(rnorm(100*30*15), c(100, 30, 15))
```

Verifying observations

```
obs <- array(rnorm(100*30), c(100, 30))
```

Compute CRPSS

```
fo.crpss <- veriApply("FairCrpss", fcst=fcst, obs=obs)
```

Compute RPSS for tercile forecasts

```
fo.rpss <- veriApply("FairRpss", fcst=fcst, obs=obs, prob=c(1/3,2/3))
```

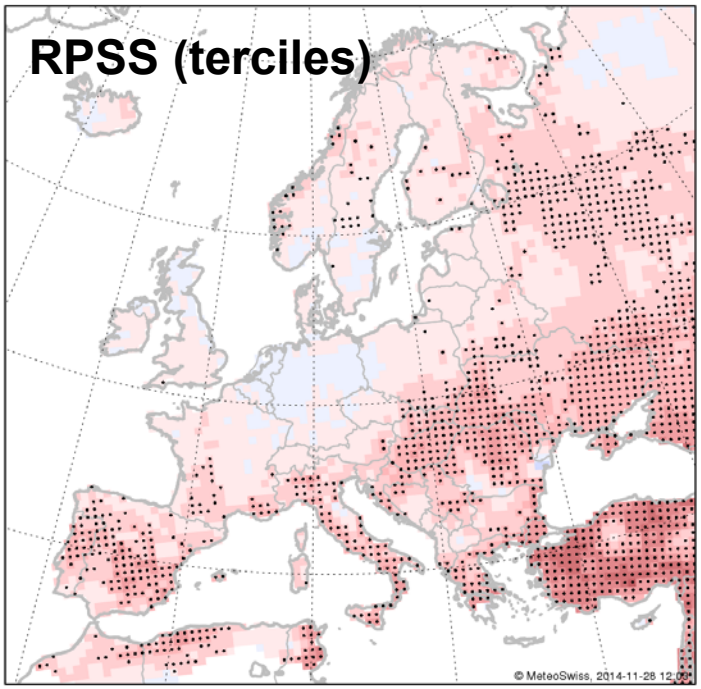
Compute CRPSS against user-defined reference

```
fo.crpss2 <- veriApply("FairCrpss", fcst=fcst, obs=obs, fcst.ref=fcst2)
```

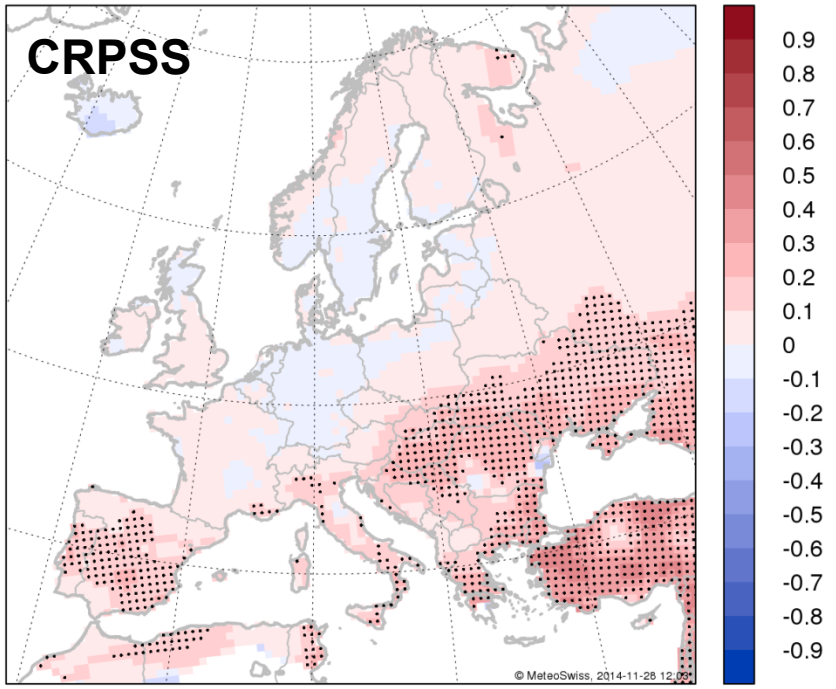
easyVerification

Examples

Ranked probability skill score for JJA tas from bias-corrected (none + CCR) ECMWF SYSTEM4 forecasts (May initialisation) verified against E-OBS for 1981-2012



Continuous ranked probability skill score for JJA tas from bias-corrected (none + CCR) ECMWF SYSTEM4 forecasts (May initialisation) verified against E-OBS for 1981-2012



easyVerification

Further documentation

Install easyVerification

```
install.packages("easyVerification")
```

Access further documentation in vignettes

```
## general usage of easyVerification
vignette("easyVerification")

## download forecasts from ECOMS-UDG and
## verify with easyVerification
vignette("ecoms_forecast_verification")
```

Get help on package and functions

```
help(package="easyVerification")
```

http://github.com/meteoswiss/easyverification