Forecast quality assessment: Making skill and bias information meaningful to the users

Antje Weisheimer

University of Oxford

ECMWF

EUPORIAS Deliverable D12.3 "**Report summarising users' needs for S2D predictions**"

led by **Marta Bruno Soares and Suraje Dessai**
University of Leeds, UK

based on 80 in-depth interviews with EUPORIAS stakeholders and a European survey of users' needs of European government and private organisations across various sectors

see also
Bruno Soares, M. and Dessai, S. (2016). Barriers and enablers to the use of seasonal climate forecasts amongst organisations in Europe. *Climatic Change*, DOI: 10.1007/s10584

"Reasons for not using seasonal forecasts were mainly associated to their **lack of reliability** …"

Examples

Water sector

"They work globally so their needs for information can vary but all year round forecasts and/or information on wet season/dry season or winter/summer months would be valuable to them.

They would also be interested in forecasts with more than 1 year predictions lead time **provided these were reliable**."

Flood risk and flood management

government organisation at the national level with more than 10,000 employees

"They currently don't use seasonal forecasts as *"(…) there's a **lack of confidence** in the existing products [and] (…) what it would mean for our business planning and processes*."

However, **if these were to become more reliable** in the future, there would be a potential to use this information to help them understand the total winter and summer rainfall."

Energy sector

"Seasonal forecasts could be useful for their long-term planning but these would **have to be more reliable**."

Examples

Insurance sector

”If seasonal forecasts become **more reliable** in the future, the organisation could use this information in their annual budgets or actuarial studies.”

Health sector

*“Although the organisation already uses seasonal forecasts as qualitative information they would potentially use it to manage their warning system **if the reliability was higher.**”*

Agricultural sector

“They are aware of the seasonal forecast available from the NMHS website but don’t use it *“( ) because this is **not enough reliable** and predictable*”.”

Forestry sector

“There is potential interest in seasonal forecasts **providing these were reliable**.”

Transport sector

“(…) the forecast never seems to be able to tell us you know, last year was a classic. We were really planning for drought up until 1 May and then we had the wettest summer on record it just wasn’t seen to be coming.”

The lack of reliability is the main barrier for not using the seasonal forecasts.

Unreliable forecasts can be dangerously misleading and should not be used for decision making.

How to communicate the forecasts, their reliability and uncertainties?

End-users vary in expertise:

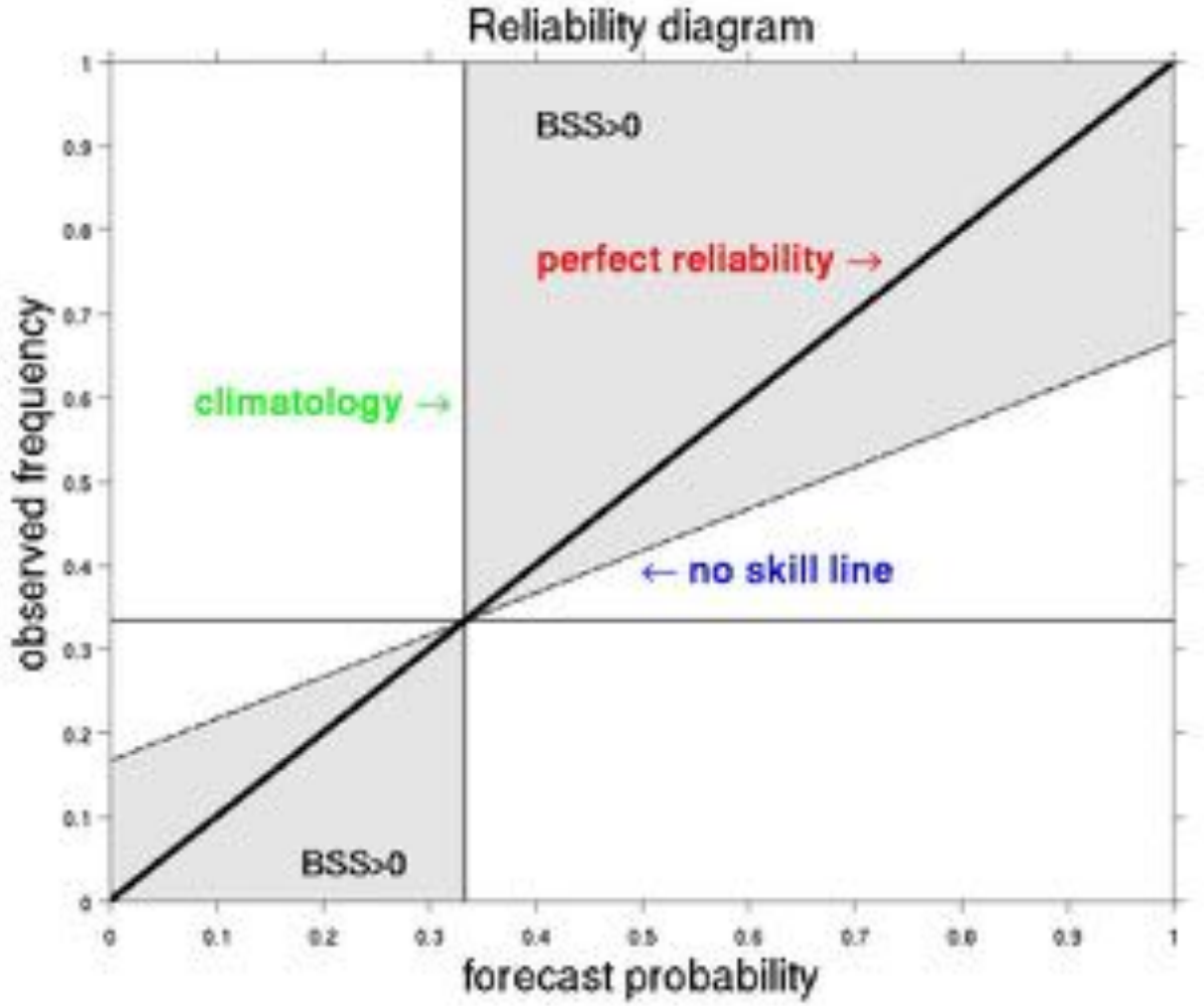
- expert users → e.g. tercile plots, bubble plots
- less experienced users → evaluative categories and simple text

→ see work by Andrea Taylor (Uni Leeds) and others in EUPORIAS

Reliability = correspondence between forecast probability and observed frequency of an event, given the forecast

E.g. Suppose an event *E* has a forecast probability of 70%.

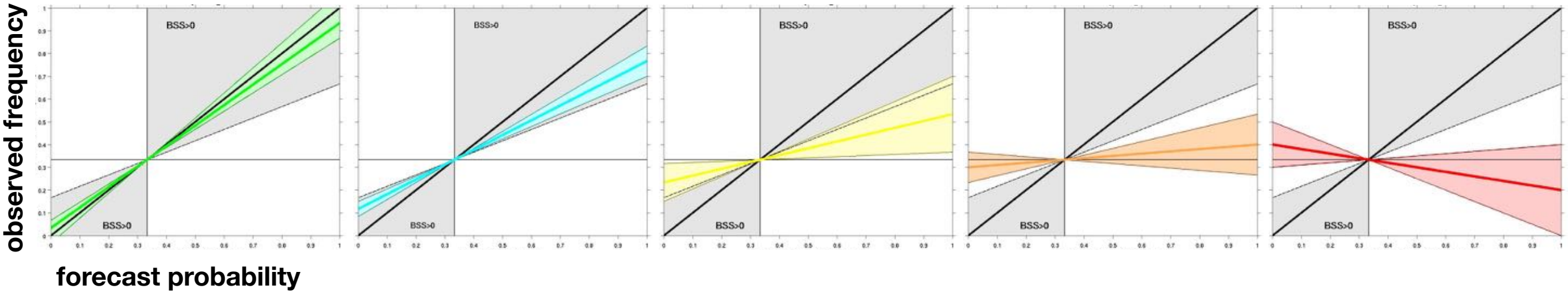
The forecasting system is said to be **reliable** if the observed frequency of *E* is, within its uncertainty ranges, also 70%.



Reliability = correspondence between forecast probability and observed frequency of an event, given the forecast

E.g. Suppose an event *E* has a forecast probability of 70%.
The forecasting system is said to be reliable if the observed frequency of *E* is, within its uncertainty ranges, also 70%.

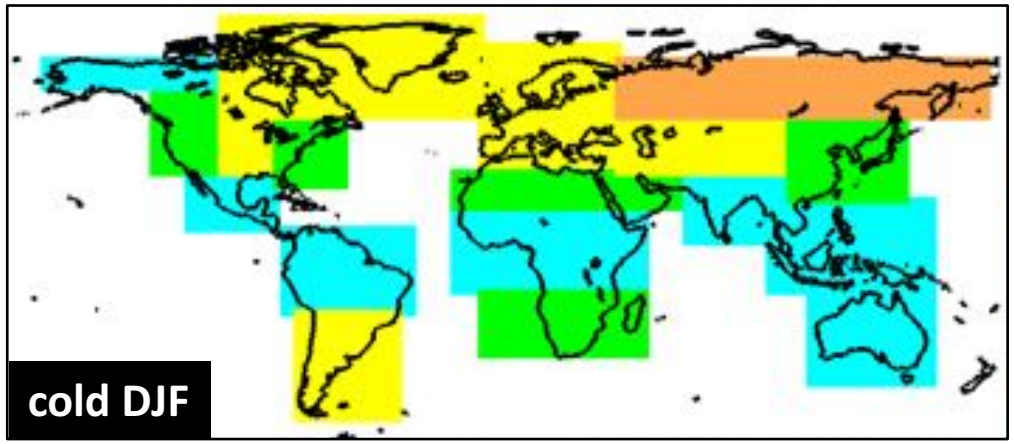
5 categories of reliability

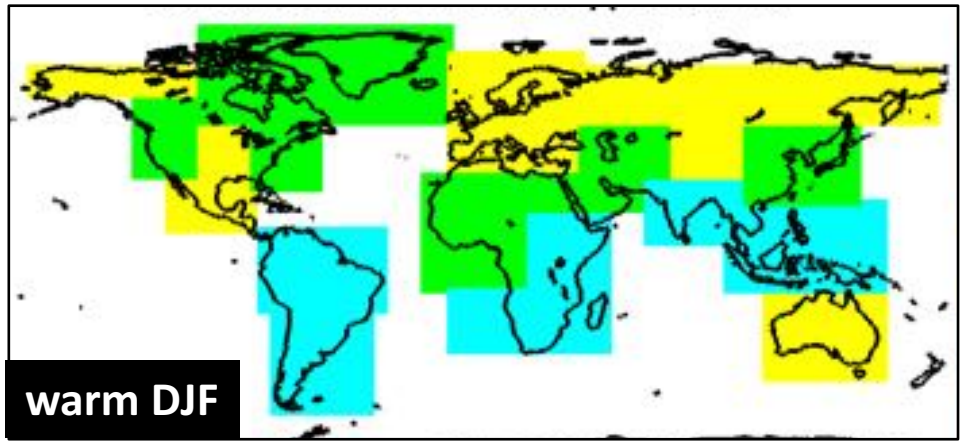


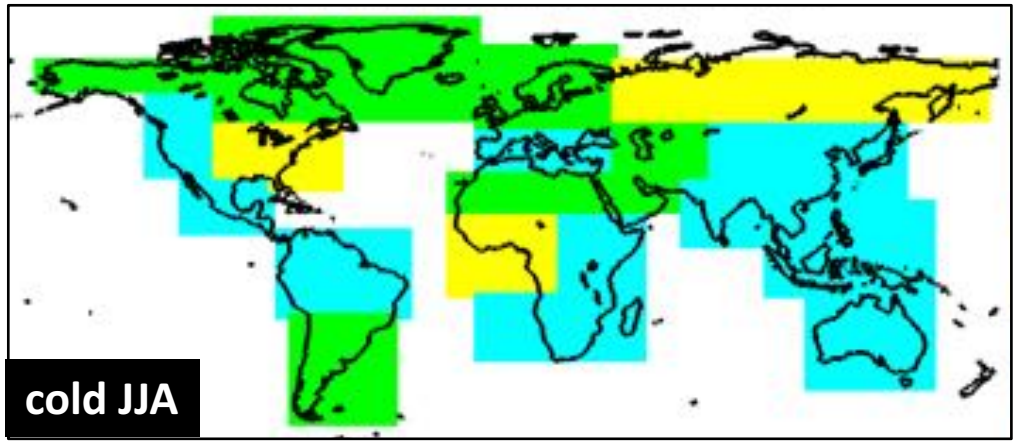
perfect **still useful** **marginally useful** **not useful** **dangerously useless**

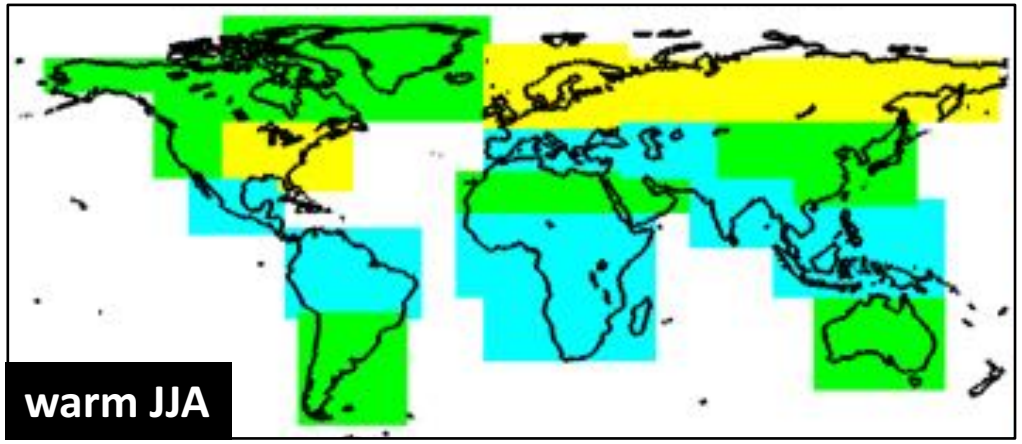
Weisheimer & Palmer (2014, JRS Interface)

Reliability of ECMWF's seasonal forecasts of temperature





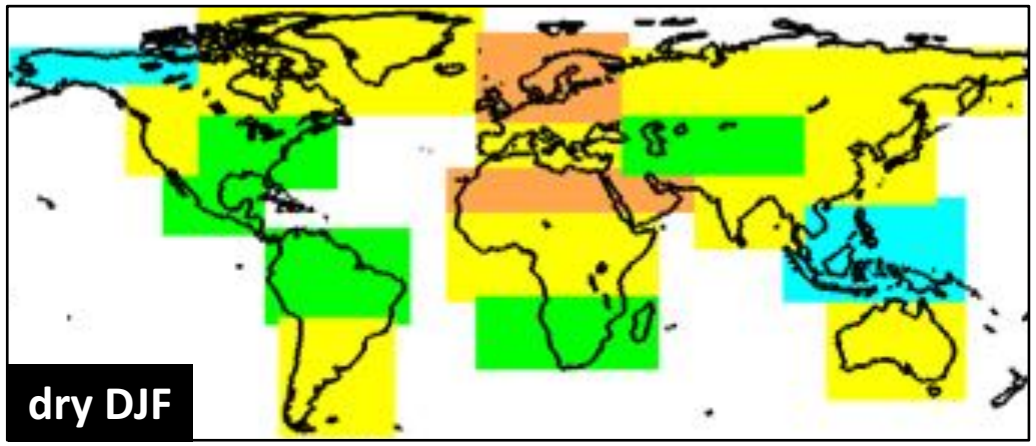


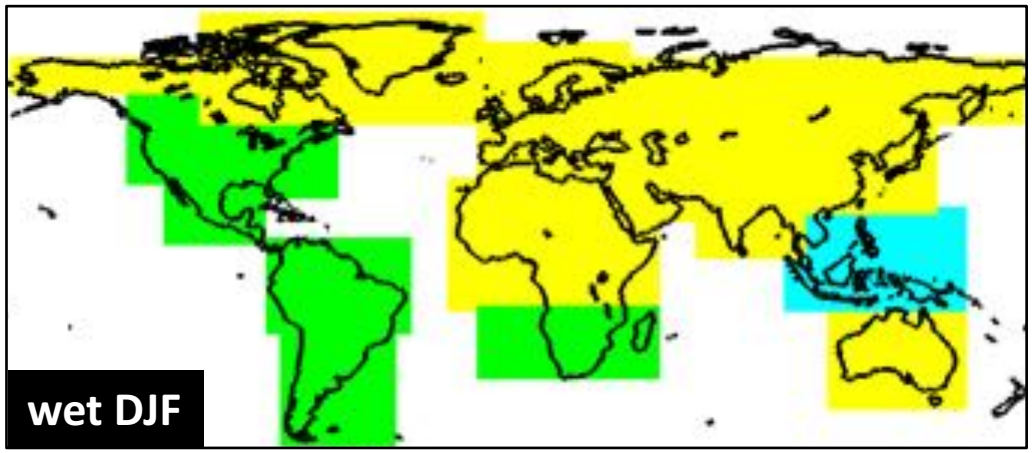


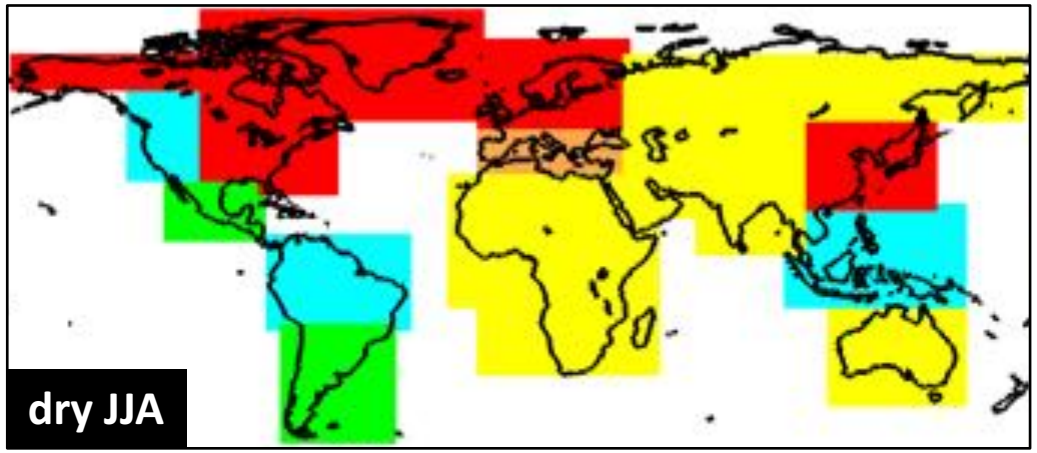
perfect | still useful | marginally useful | not useful | dangerously useless

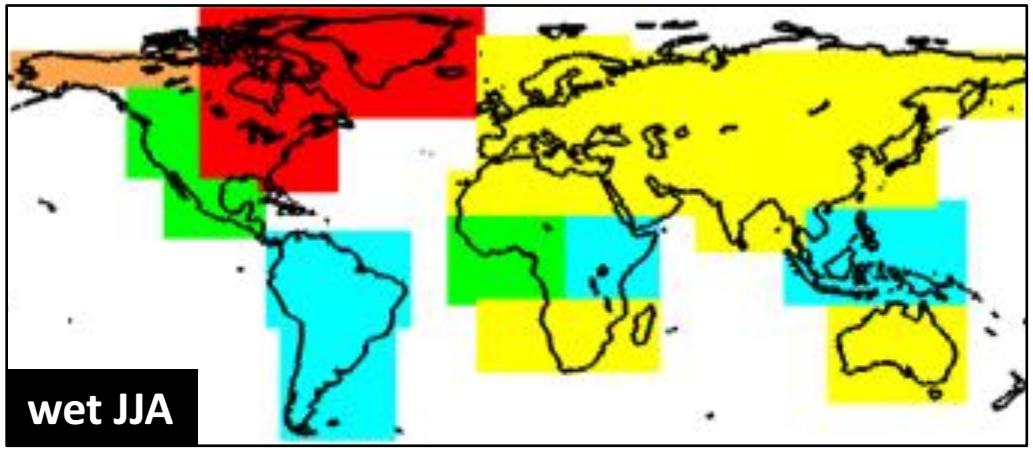
Weisheimer & Palmer (2014, JRS Interface)

Reliability of ECMWF's seasonal forecasts of precipitation









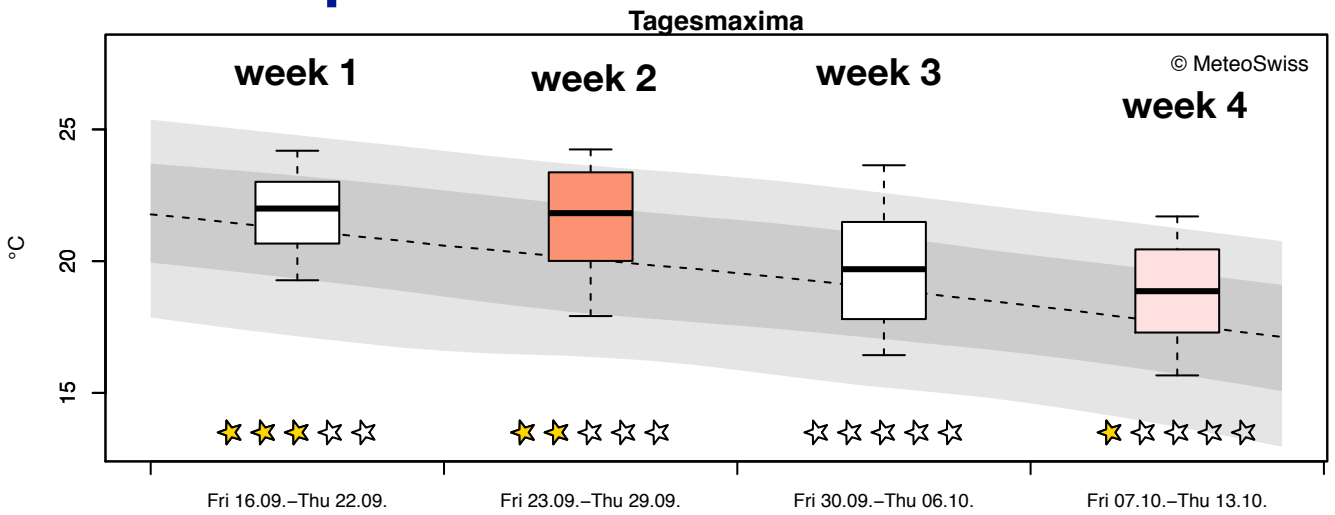
perfect | still useful | marginally useful | not useful | dangerously useless

Weisheimer & Palmer (2014, JRS Interface)

Christoph Spirig, Jonas Bhend and Mark Liniger (MeteoSwiss): Visualisation of operational probabilistic forecast and hindcast skill

TEST: Monthly forecasts: forecast from 12/09/2016

tercile plots



Beobachtungen 1996–2015
- 10–90% Quantil
- 25–75% Quantil
- Mittelwert

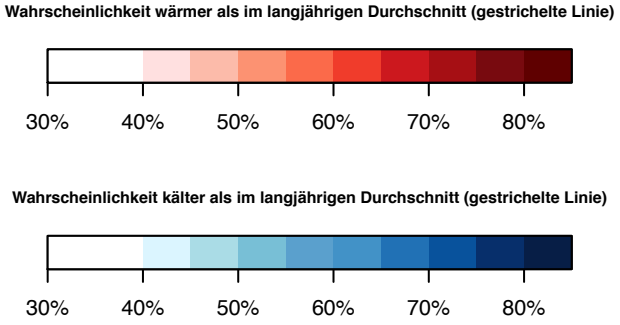
climatology

Vorhersagen
Whiskers: 10–90% Quantil
Box: 25–75% Quantil

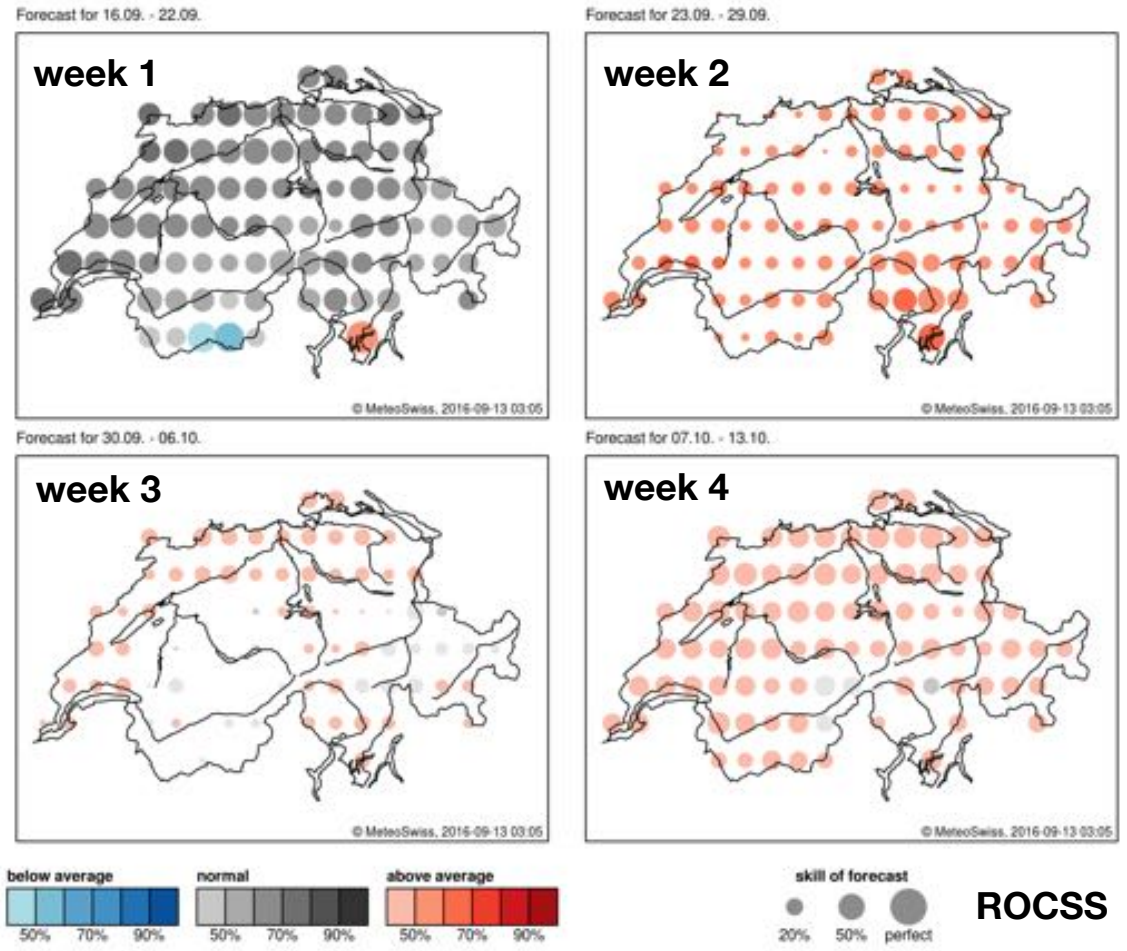
Vorhersagegüte in den letzten 20 Jahren
0: nicht besser als Raten
5: sehr gut

RPSS

forecast probabilities



bubble plots



ROCSS

Emma Suckling (Uni Reading):
Simple format using evaluative categories and text to communicate sub-seasonal forecasts for energy trading

	Temperature	Wind speed	Cloud cover
Europe			
UK			
France			
Spain			
Germany			
Northern Germany			

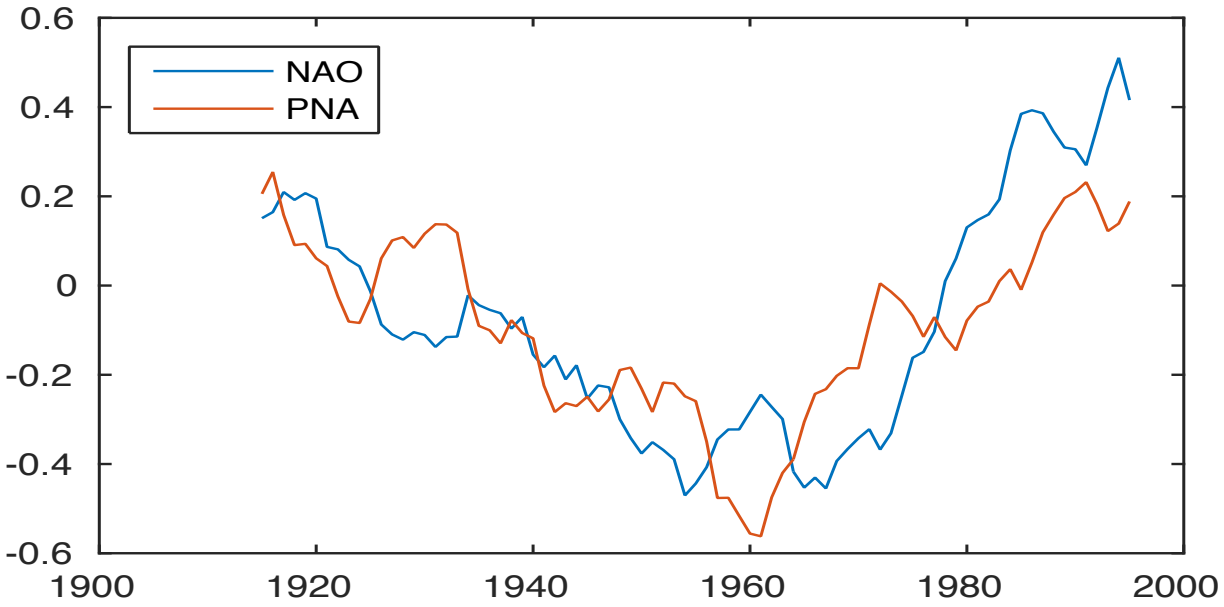
Classification	Description	Interpretation
Significant skill	The CRPS and ACS scores show statistically significant skill, while RMSE shows improvement over climatology	We have demonstrated that there is forecast skill
Moderate skill	CRPS, ACS and RMSE all demonstrate positive skill, however not all metrics may demonstrate statistically significant improvements over climatology	There is likely to be skill but more data is required for confirmation
Positive skill	More than one metric shows positive skill, however it is not statistically significant	There may be some skill but further investigation into the details is necessary
Low/no skill	All metrics return negligible or negative skill relative to climatology	There is no obvious indication of skill in this property

What can estimates of past forecast skill tell us about the performance of our forecasting systems in the future?

Is skill based on ~30 years of hindcasts a guarantee for success in the future?

Example: NAO forecasts

- **NAO variability on interannual to multi-decadal time scales**



What can estimates of past forecast skill tell us about the performance of our forecasting systems in the future?

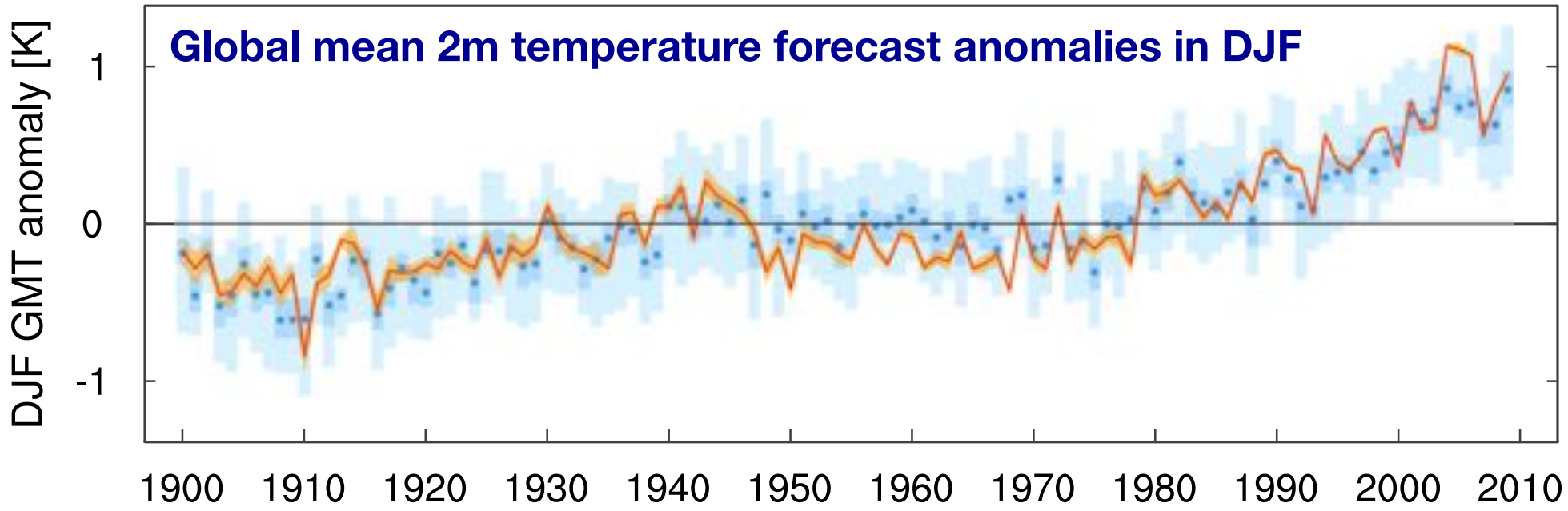
Is skill based on ~30 years of hindcasts a guarantee for success in the future?

Example: NAO forecasts

- **NAO variability on interannual to multi-decadal time scales**
- **Skilful interannual predictions of the winter NAO during recent decades (NAO was predominantly in its positive phase)**
- **Would our forecasts be equally good if the NAO was/will be in a different phase of multi-decadal variability, e.g. negative regime?**
- **If not, why?**

Atmospheric seasonal hindcasts of the 20th Century (ASF-20C)

- **ECMWF atmospheric model (recent version), T_L255L91**
- **Hindcast period: 1900 – 2010**
- **Initial data: ERA-20C, HadISST for prescribed SST and sea-ice**
- **51-member ensemble**

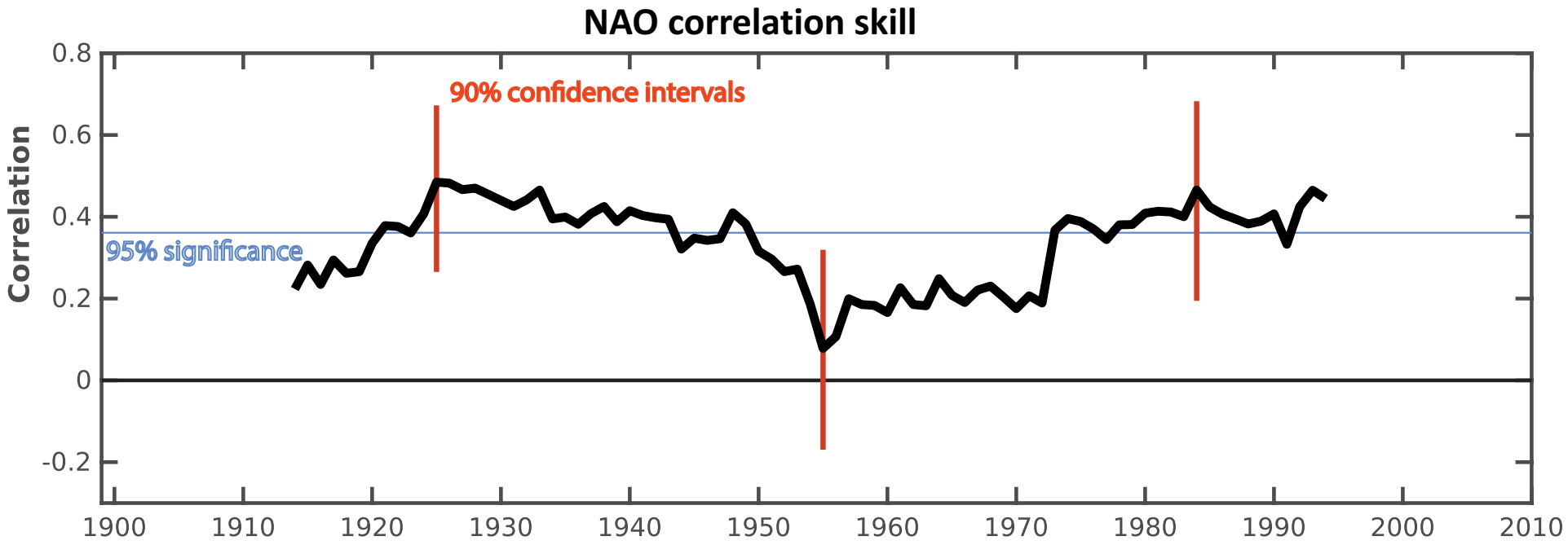


DJF global mean 2m temperature in ERA-20C (red) and the re-forecast ensemble of ASF-20C (blue). Uncertainty estimates from the reanalysis and the re-forecast ensemble are shown in orange (full range of the 10-member ensemble) and with blue shades (light blue: full range; darker blue: interquartile 25%-75% range; blue dots: ensemble median), respectively.

Weisheimer et al. (2017), QJRMS

Multi-decadal variability of NAO forecast skill

- estimates from 30-year moving windows -



Anomaly correlation coefficient of the DJF NAO index between the ensemble mean and ERA-20C computed for moving 30-year windows by one year. Values are plotted at the 15th year of each window. The horizontal line indicates the *t*- test 95% significance level of the correlations and the red vertical bars show 90% confidence intervals estimated from bootstrap re-sampling (1000 times) with replacement for three representative periods.

Weisheimer et al. (2017), QJRMS

- **Positive and significant skill in predicting the interannual NAO variations for DJF over the entire period**
- **Distinct multi-decadal variability of winter NAO forecast skill**
- **Asymmetry in predictive skill of NAO phases**
- **Non-stationarity of signal-to-noise behaviour**

→ **Mid-Century period stands out as an important period on which to test the performance of future seasonal forecast systems.**