What do verification scores tell us?

Anna Ghelli, ECMWF

anna.ghelli@ecmwf.int

ECMWF

Outline

- Verification: Basic Concepts
- Verification: description of the true status of the atmosphere
- Scores: their formulation and what they measure
- Conclusions

Verification: basic concepts

- ***Definition***: forecast verification is the process of assessing the quality of a forecast.
- ***Purpose***:
  - Administrative: monitoring performance
  - Scientific: identify and correct model flaws, forecast improvements
  - Economic: improved decision making
- ***Type***: verification can be
  - Qualitative : it will answer to questions like "does my forecast look right?“
  - Quantitative: it will answer to questions like: "how accurate was my forecast?".

Forecast quality versus forecast value

- A forecast has high **QUALITY** if it predicts the observed conditions well according to some objective or subjective criteria.

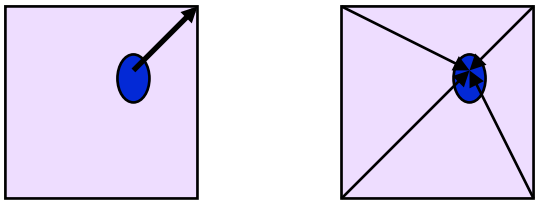
Quality but no value

- A forecast has **VALUE** if it helps the user to make a better decision.

Value but no quality

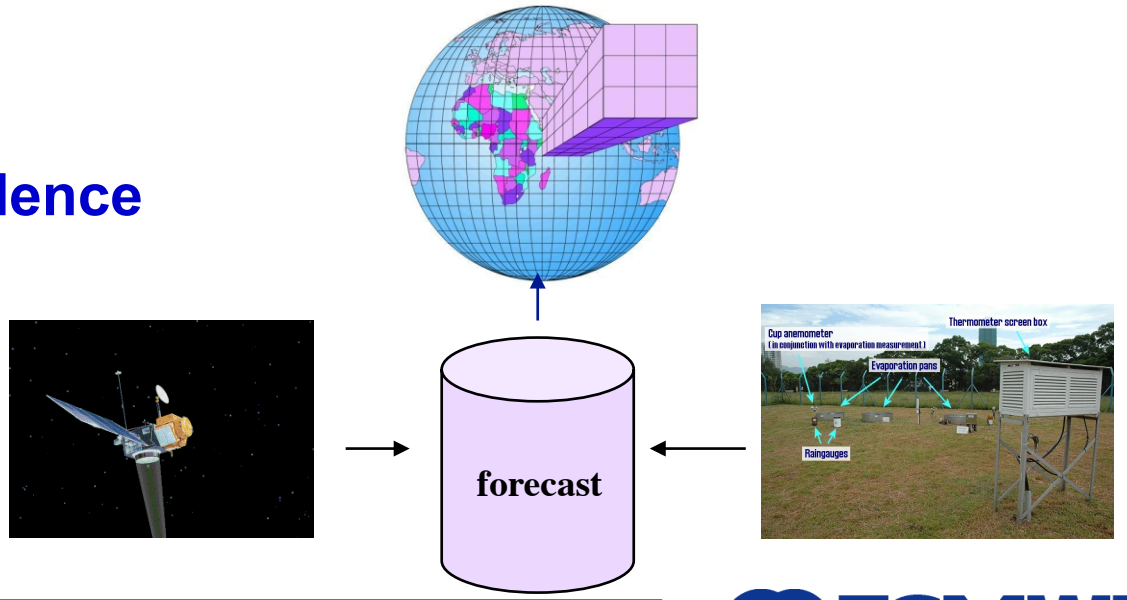
Observations and analysis

- Forecast and observation **matching**



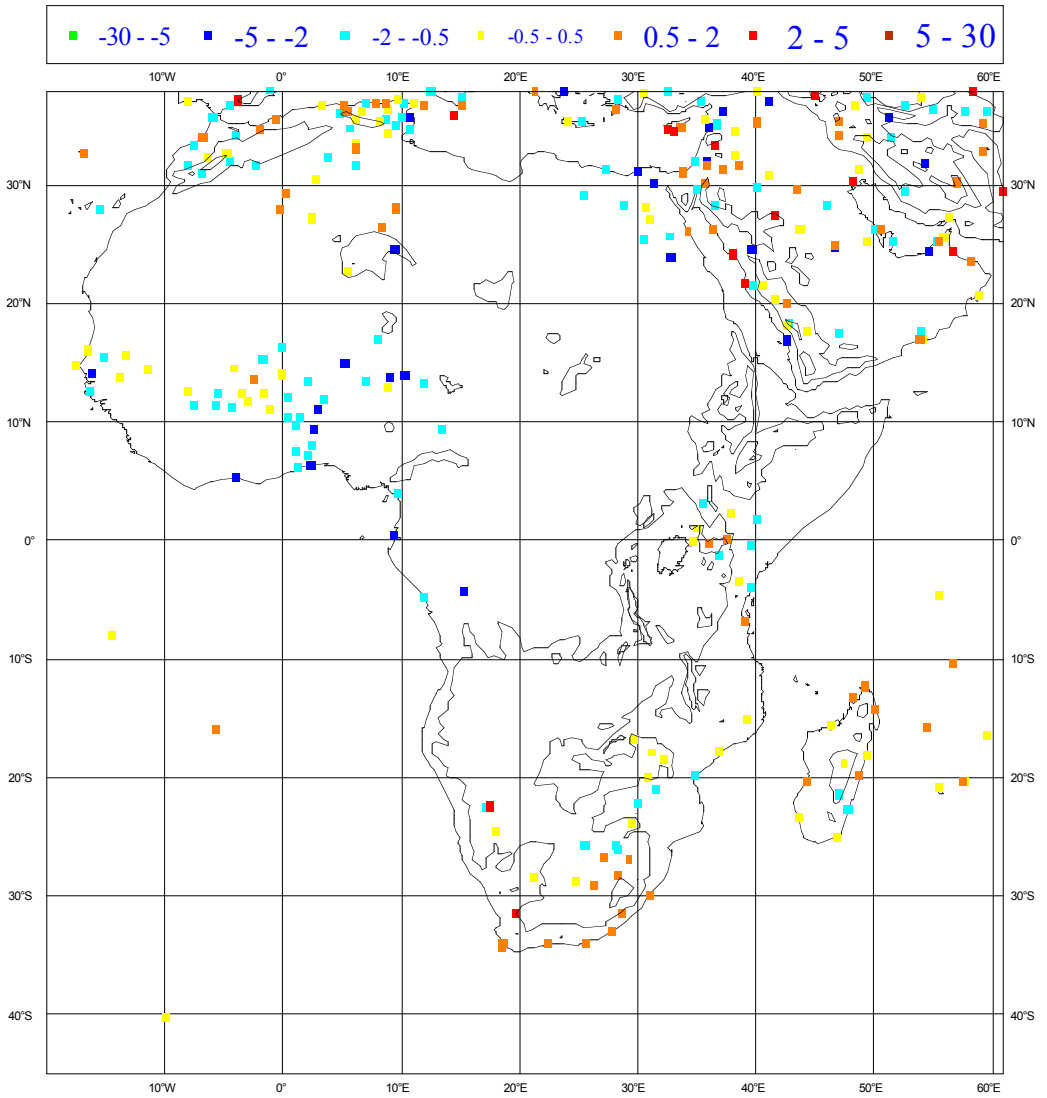
- Analysis:
  - **independence**

- Observations:
  - **coverage**
  - **error**
  - **“representativeness”**



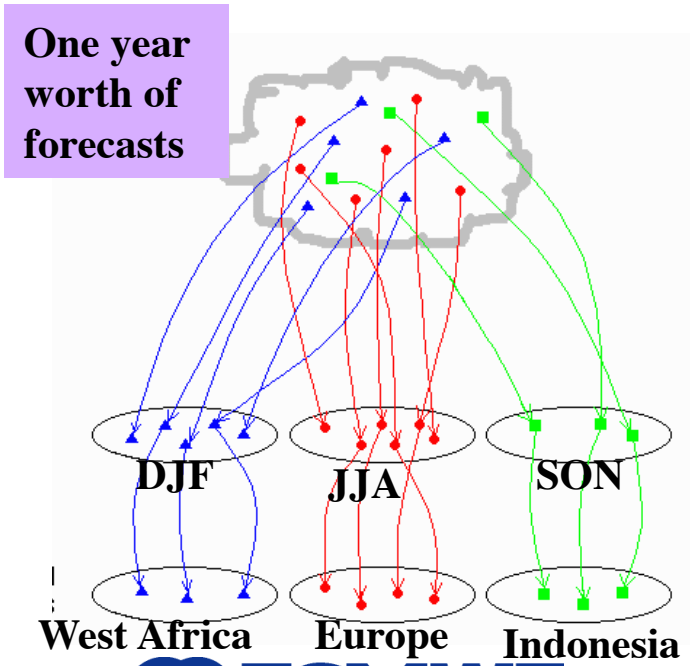
ECMWF

Observations -- Coverage



Pooling versus stratifying results

- Samples (forecast/observation pairs) may be **pooled** over time and/or space.
  - Mask differences in forecast performance
  - Biased toward the most commonly sampled regime (i.e. days with no severe weather).
- Forecast can be **stratified** into quasi-homogeneous subsets
  - Be aware of subsets sample size!



Scores: formulation

- *Root Mean Square Error:*

$$E = \sqrt{\overline{(fc - an)^2}}$$

Measures accuracy
Range: 0 to infinity perfect score = 0

- *Bias:*

$$BIAS = \overline{FC - OBS}$$

Measures bias
Range: -infinity to +infinity perfect score = 0

- *Mean Absolute Error :*

$$MAE = \overline{|FC - OBS|}$$

Measures accuracy
Range: 0 to infinity perfect score = 0

- *Anomaly Correlation:*

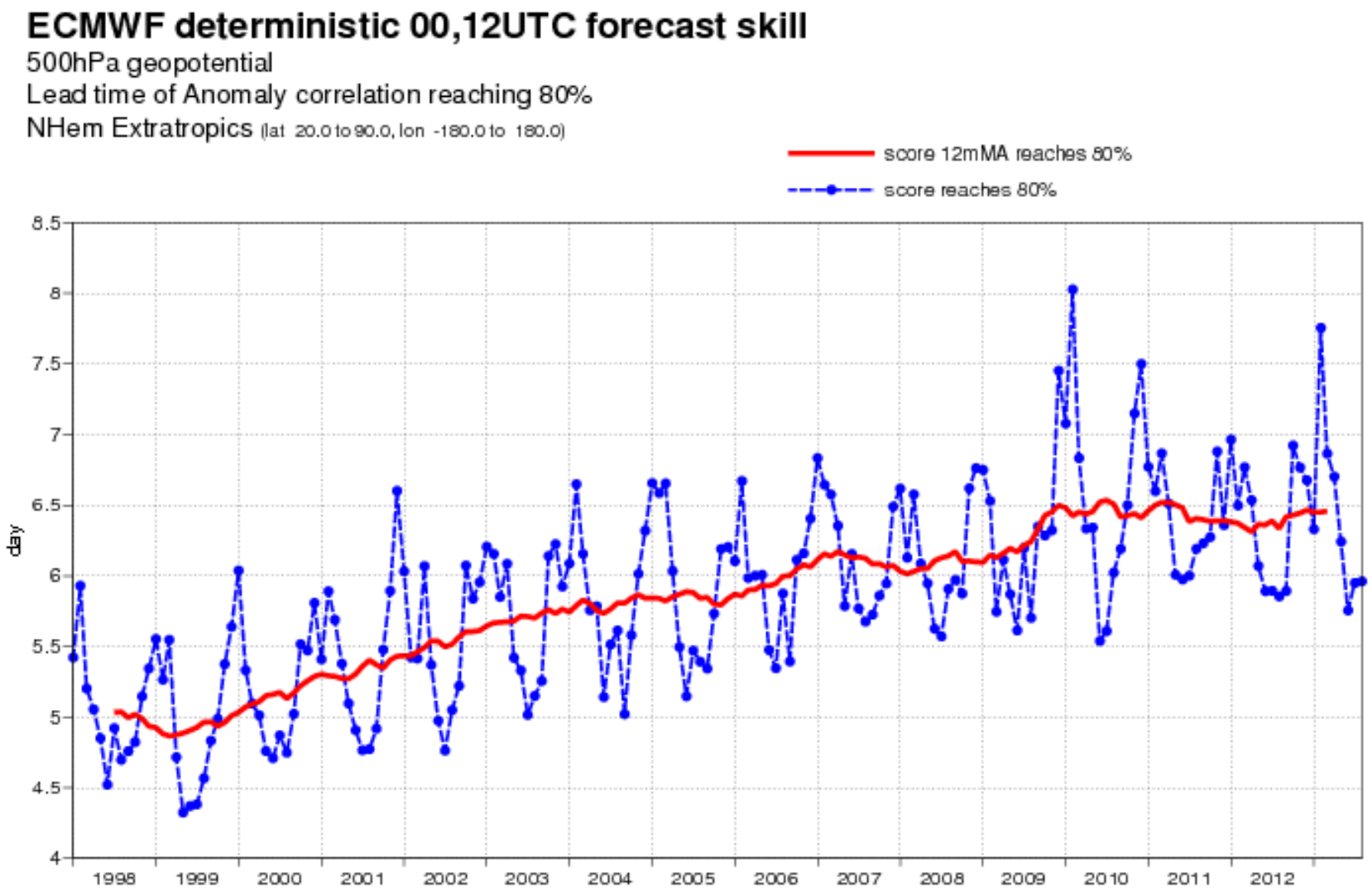
$$ACC = \frac{\overline{(fc - c)(an - c)}}{\sqrt{A_{fc} A_{an}}}$$

$$A_{fc} = \overline{(fc - c)^2}$$

$$A_{an} = \overline{(an - c)^2}$$

Measures accuracy
Range: -100% to 100%
perfect score = 100%

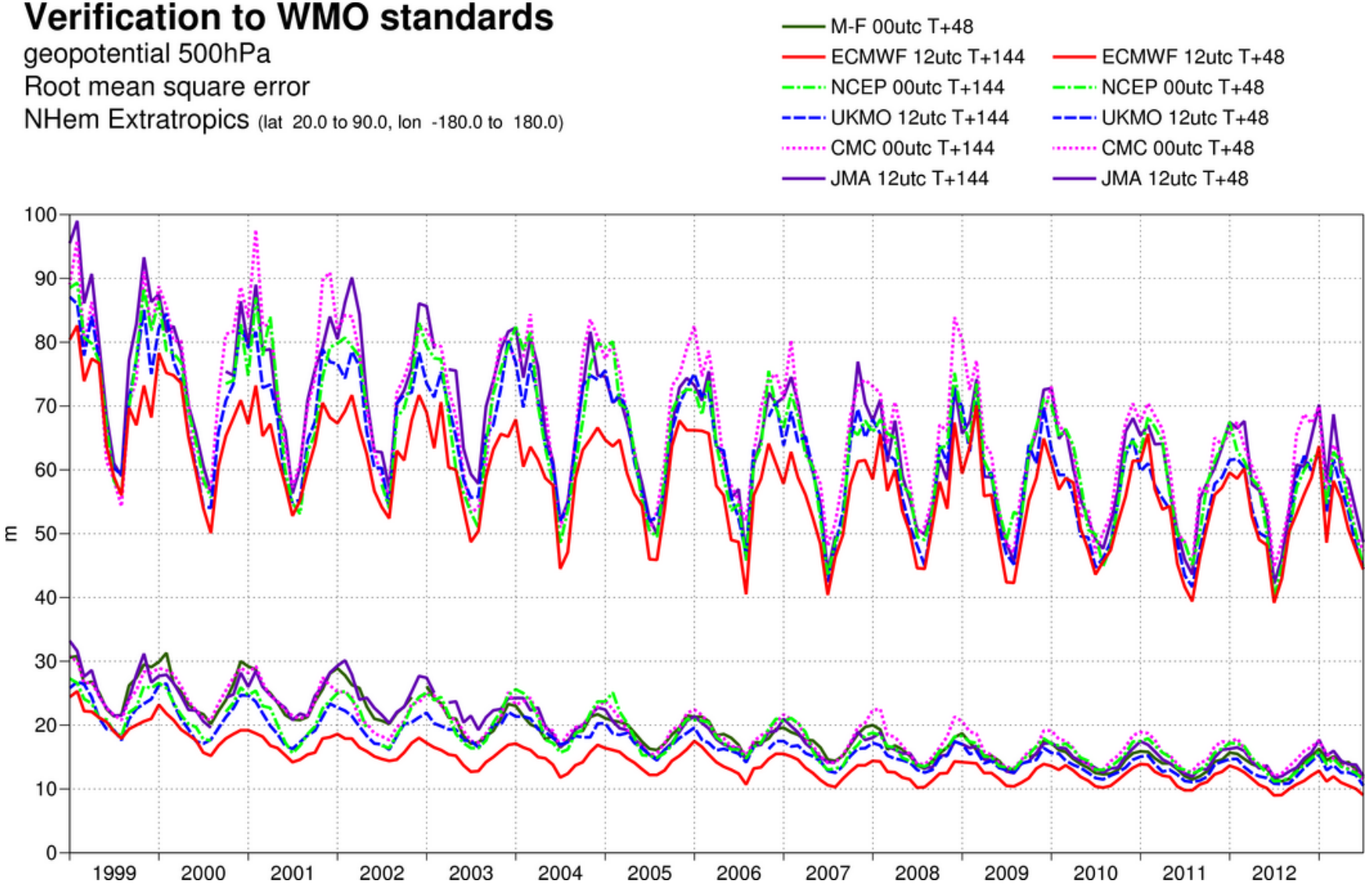
Time series Acc=80% N hemisphere



ECMWF

Verification to WMO standards

geopotential 500hPa
Root mean square error
NHem Extratropics (lat 20.0 to 90.0, lon -180.0 to 180.0)



RMSE - behaviour of the overall error, size of the overall error
No information on phase and amplitude components of error

What is the Event?

- For categorical and probabilistic forecasts, one must be clear about the "event" being forecast
  - Location or area for which forecast is valid
  - Time range over which it is valid
  - Definition of category
- And now, what is defined as a correct forecast?
  - The event is forecast, and is observed – anywhere in the area? "At least one observation" in the area is a hit if a valid warning is out
  - No report is taken to mean no severe weather in the domain; proxies are allowed
  - The troublesome "d"

Contingency tables

Frequency Bias

$$FBI = B = \frac{(a+b)}{(a+c)}$$

Hit Rate

$$H = POD = \frac{a}{(a+c)}$$

False Alarm Rate

$$= \frac{b}{(b+d)}$$

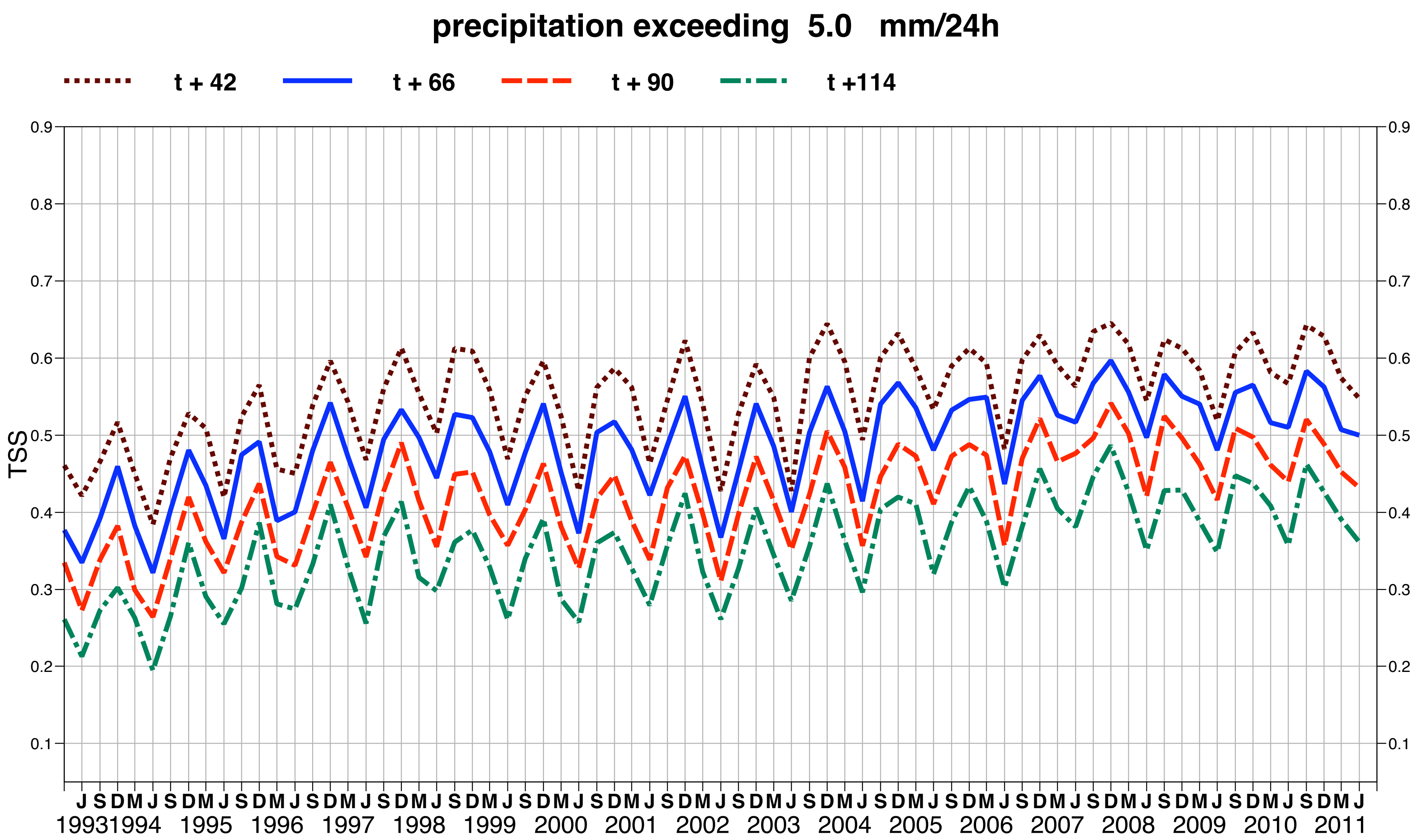
Event forecast	Event observed		
	Yes	No	Marginal total
Yes	Hit	False alarm	Fc Yes
No	Miss	Correct non-event	Fc No
Marginal total	Obs Yes	Obs No	Sum total

Event forecast	Event observed		
	Yes	No	Marginal total
Yes	a	b	a + b
No	c	d	c + d
Marginal total	a + c	b + d	a + b + c + d =n

True Skill Score (also known as Pierce's Skill Score)

$$TSS = PSS = \frac{ad - bc}{(a+c)(b+d)}$$

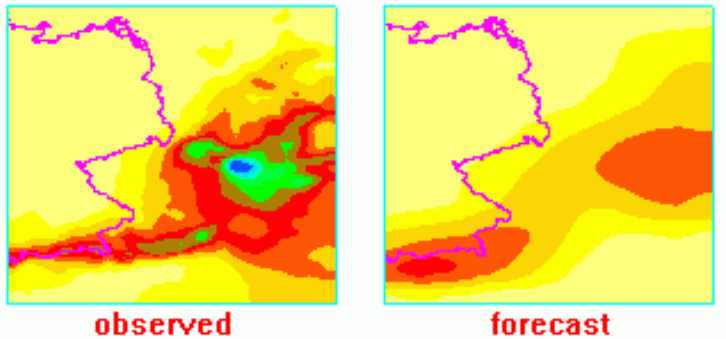
Contingency tables – True Skill Score



Scores: what they can/cannot offer

- Overall measures of skill (accuracy, bias)
- Smooth forecasts -> best performance (in general)
- Minimal diagnostic information
- Cannot answer the following questions:

  - What went wrong? What was right?
  - Does the forecast looks realistic?
  - How can I improve the forecast?
  - How can I use the forecast to make a decision?



Conclusions

- Verification is part of the forecasting process
- Traditional scores measure overall skills
- Advantages and disadvantages when using observations and/or analyses to define the ‘true’ status of the atmosphere

Detailed verifications help improving models . An improved model means:

- **Better forecasts** for extreme events
- **Augment credibility** of forecasters / Met Services
- **Better decision** making