Observations and their importance in the verification process: view of the Joint Working Group on Forecast Verification Research (JWGFVR)

Anna Ghelli

Thanks to: Simon Mason, Laurie Wilson, Barbara Casati, Barbara Brown, Beth Ebert, Joel Stein, Pertti Nurmi

Who are we?

Barbara **Brown** (NCAR)
Beth **Ebert** (BOM)
Anna **Ghelli** (ECMWF)
Marion **Mittermaier** (UK MetOffice)
Joel **Stein** (Meteo France)
Clive **Wilson** (UK MetOffice)
Barbara **Casati** (Ouranos)
Harold **Brooks** (NOAA)
Martin **Goeber** (DWD)
Pertti **Nurmi** (FMI)
David **Stephenson** (Uni. Exeter)
Laurie **Wilson** (Env. Canada)

Vision:

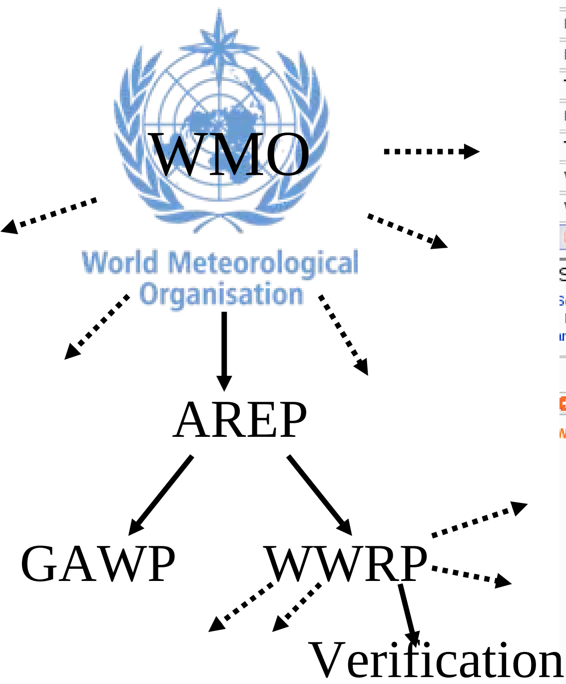
To promote best practices and understanding of verification methods

Activities:

- Verification guidance and support for WMO Forecast Demonstration Projects
- Participation in activities of other WMO groups
- Documentation of recommended methods for specific application
- Education

Where in WMO?

http://www.wmo.int/pages/prog/arep/wwrp/new/Forecast_Verification.html



History and achievements:

2002 Birth

2002 1st workshop (Boulder, Colorado, USA)

2004 2nd workshop (Montreal, Canada)

2007 3rd workshop (ECMWF, Reading, UK) the workshop included for the first time tutorials

2008 Special issue of Meteorological Applications vol. 15 no. 1 with papers from the 3rd international workshop.

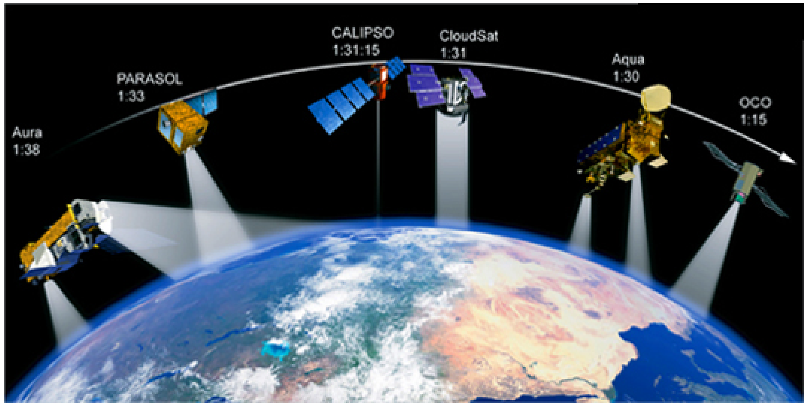
2009 Publication WMO/TD-No. 1485 “Recommendations for verification of QPF”

2009 4th workshop (FMI, Helsinki, Finland), tutorials were run as integral part of the workshop

3rd International Workshop on Verification Methods, ECMWF, Reading, January 2007

outline

- Quality control
- Observation uncertainty: how to account for it
- Observation dataset independency
- The role of analyses in verification
- Conclusions



Quality control

Quality control

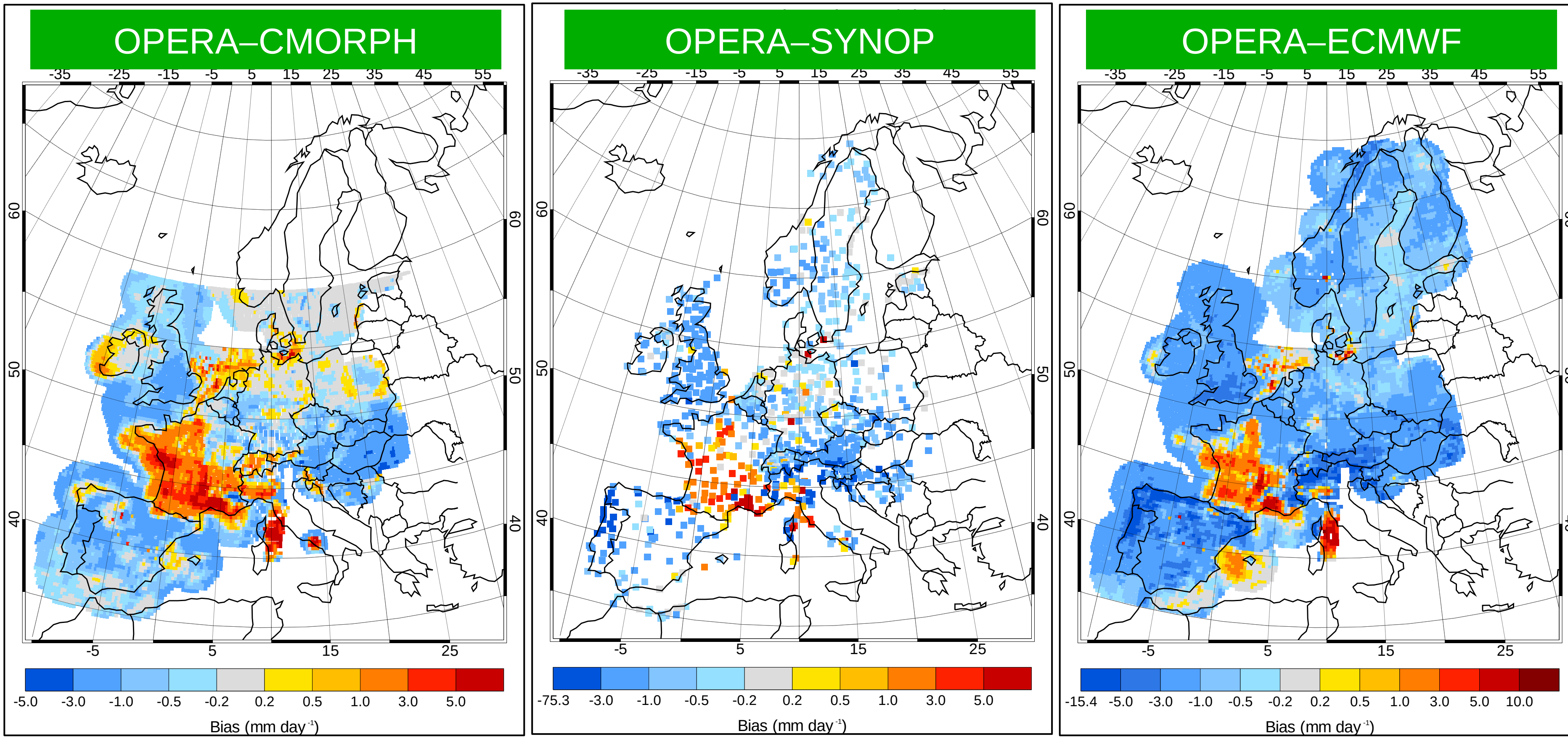
- Remove gross errors
- Remove instruments and reporting errors
- Remove biases

Properties

- Standardized procedures
- Model independent

(P. Lopez, ECMWF,Tech. Memo. 569, 2008)

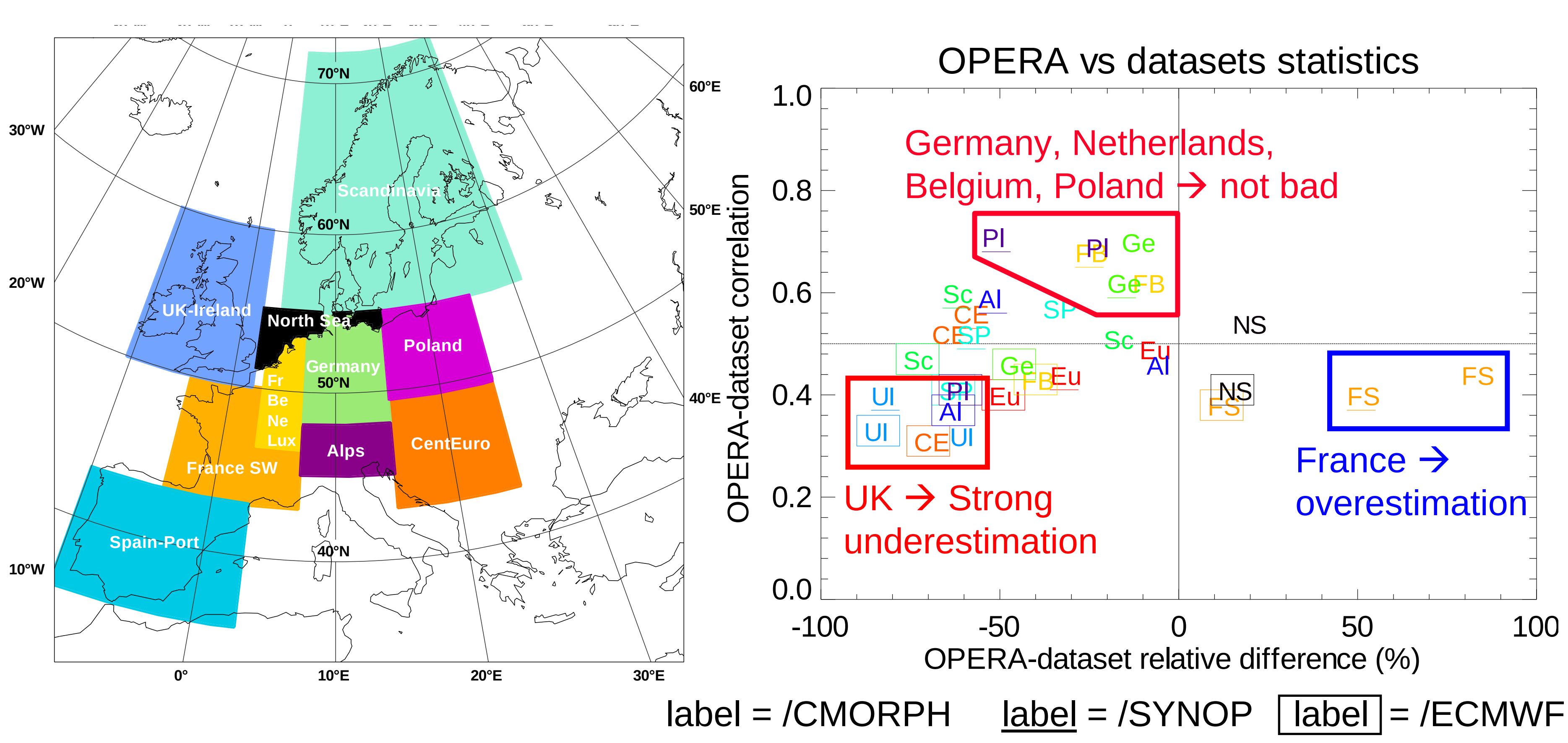
Mean differences OPERA–dataset (mm day^{-1})
Period: 10 April – 8 June 2008 (60 days)



-- Comparisons with other datasets
-- Common procedures

(P. Lopez, ECMWF,Tech. Memo. 569, 2008)

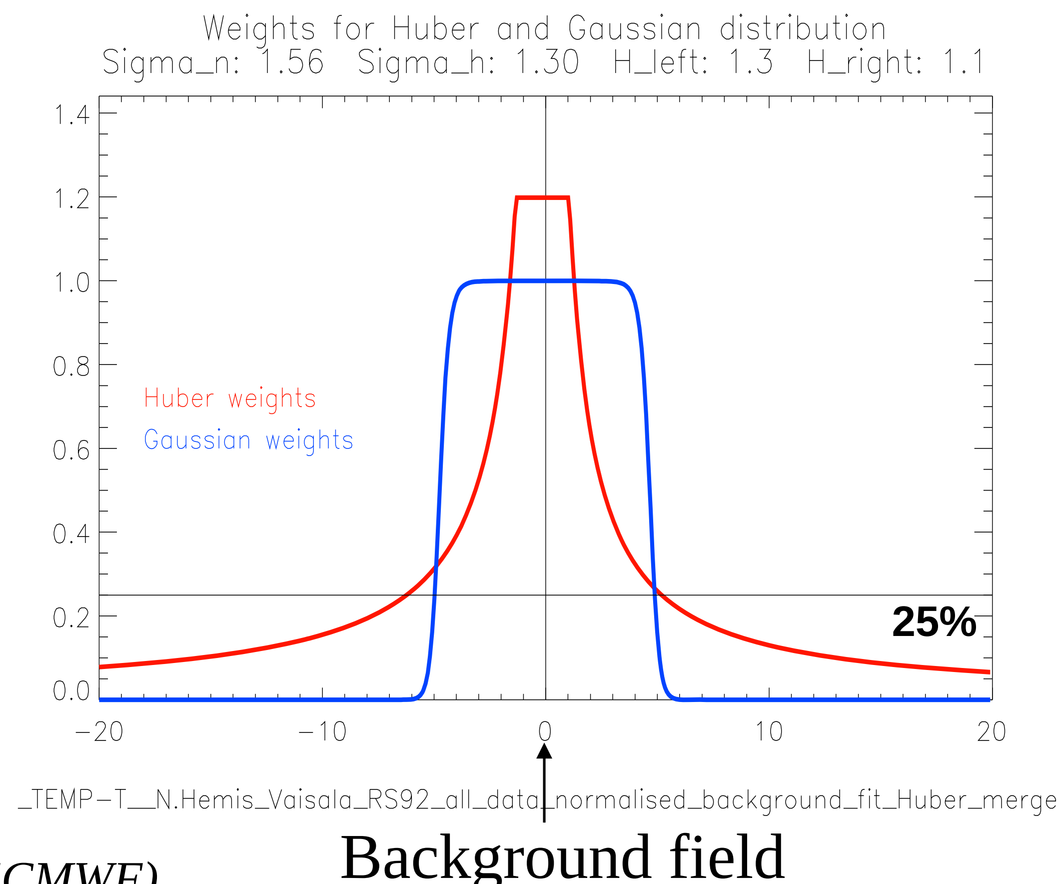
OPERA-dataset mean correlation vs OPERA–dataset mean difference (various domains)
Period: 10 April – 8 June 2008 (60 days)



Comparing two norms: Huber (red) Gaussian (blue)

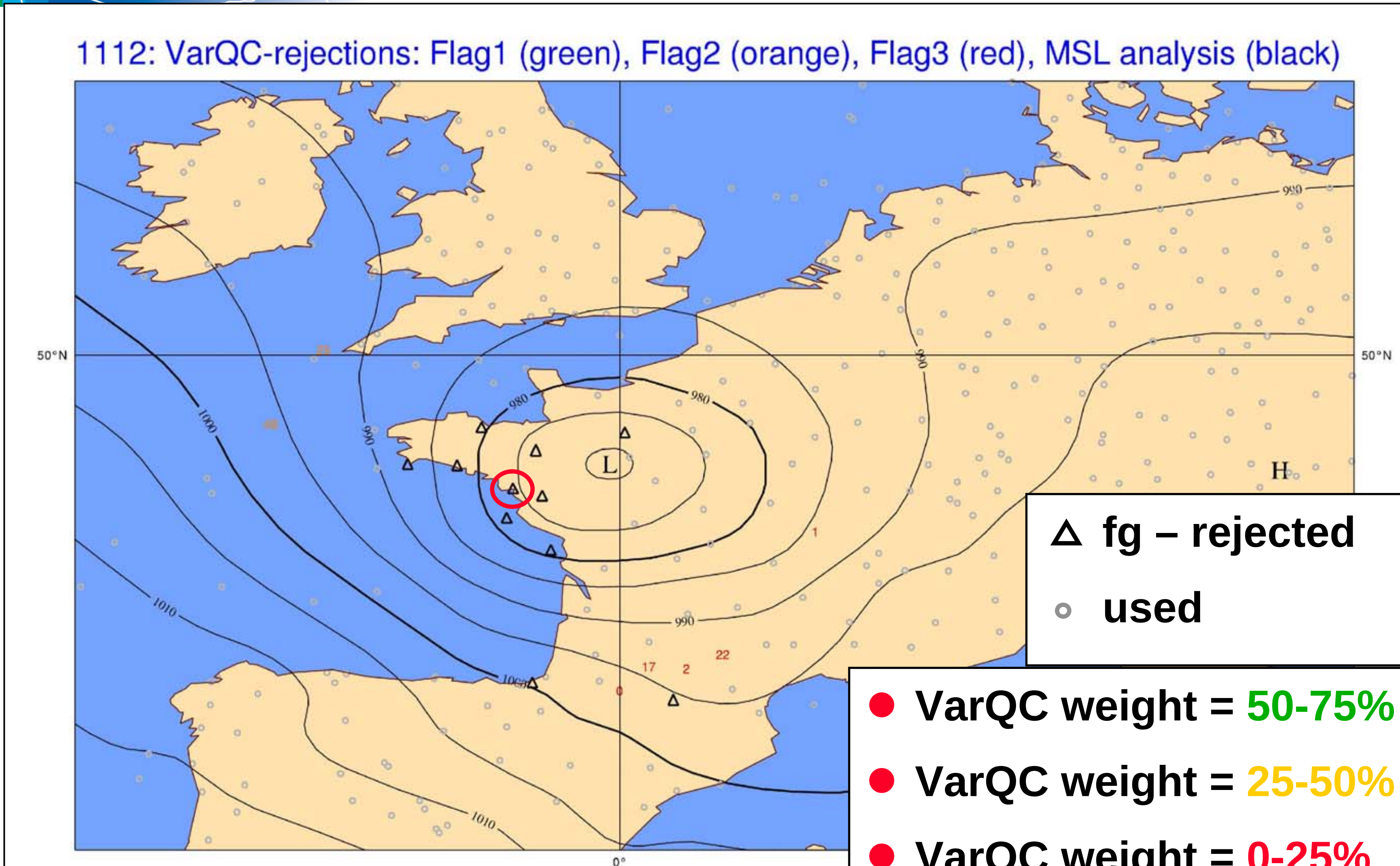
- **More weight in the middle of the distribution**
- **More weight on the edges of the distribution**
- **More influence of data with large departures**

  -Weights: 0 – 25%



(Lars Isaksen and Cristina Tavolato, ECMWF)

French Storm



△ fg – rejected

∘ used

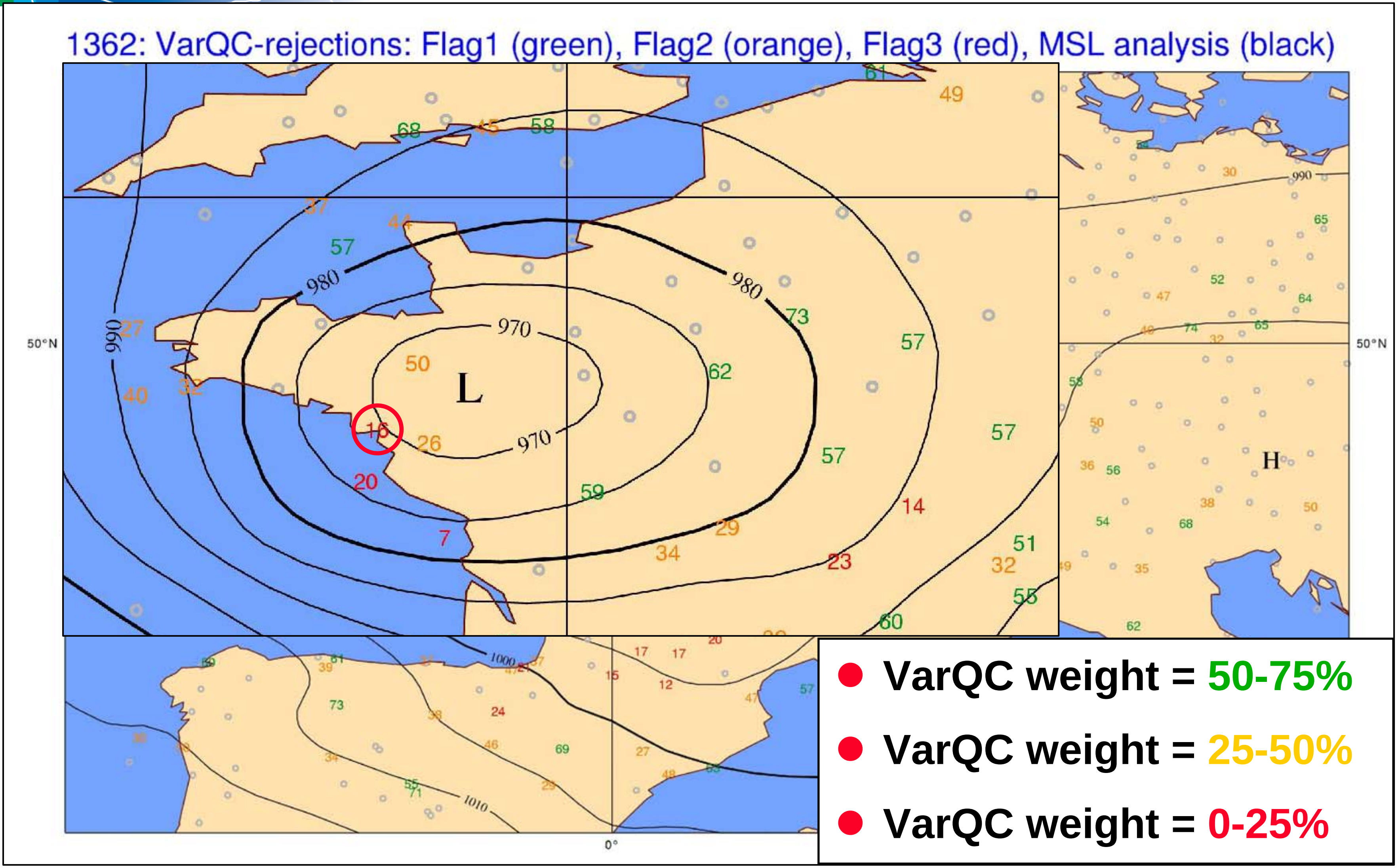
● VarQC weight = 50-75%

● VarQC weight = 25-50%

● VarQC weight = 0-25%

JWGFVR French Storm

-- not model independent



- VarQC weight = 50-75%
- VarQC weight = 25-50%
- VarQC weight = 0-25%

ECMWF

Uncertainty in observations

Sources of Uncertainty

- Observation error
- “Under-sampling” of station data
- Interpolation (time and space)
- Analysis errors
- ....

How can we cope with observational uncertainty?

JWGFVR

Verification using RADAR and rain gauges

- QPF in pre-specified area ➔ River / Lake catchment
- Three independent components addressing the quality

		-2	...	0	...	+2
♦ Structure	- S -	-2 object too small or too peaked	...	0 Perfect	...	+2 objects too large or too flat
♦ Amplitude	- A -	-2 Averaged QPF under-estimated	...	0 Perfect	...	+2 Averaged QPF over-estimated
♦ Location	- L -	-2	...	0 Perfect	...	+2 wrong location of Total Center of Mass (TCM) and / or of objects relative to TCM

For a perfect forecast: S = A = L = 0

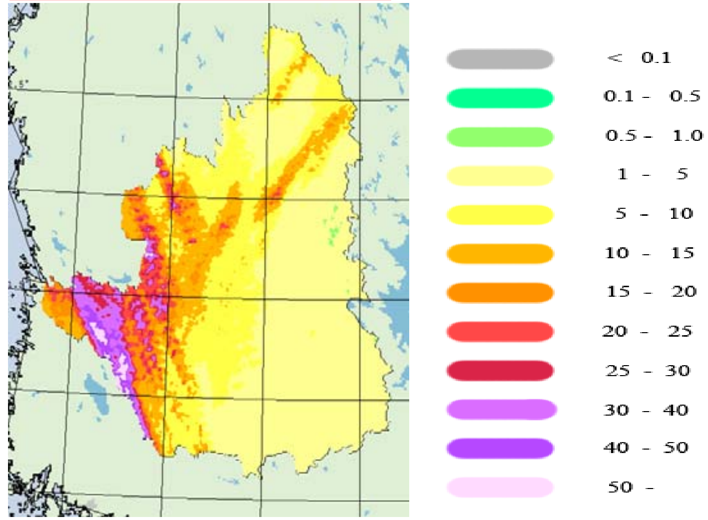
Wernli, Paulat, Hagen, Frei, 2008 (MWR, 136, 4470-4487)

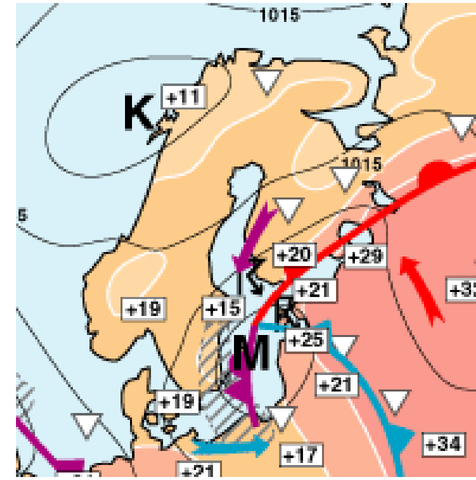
ECMWF

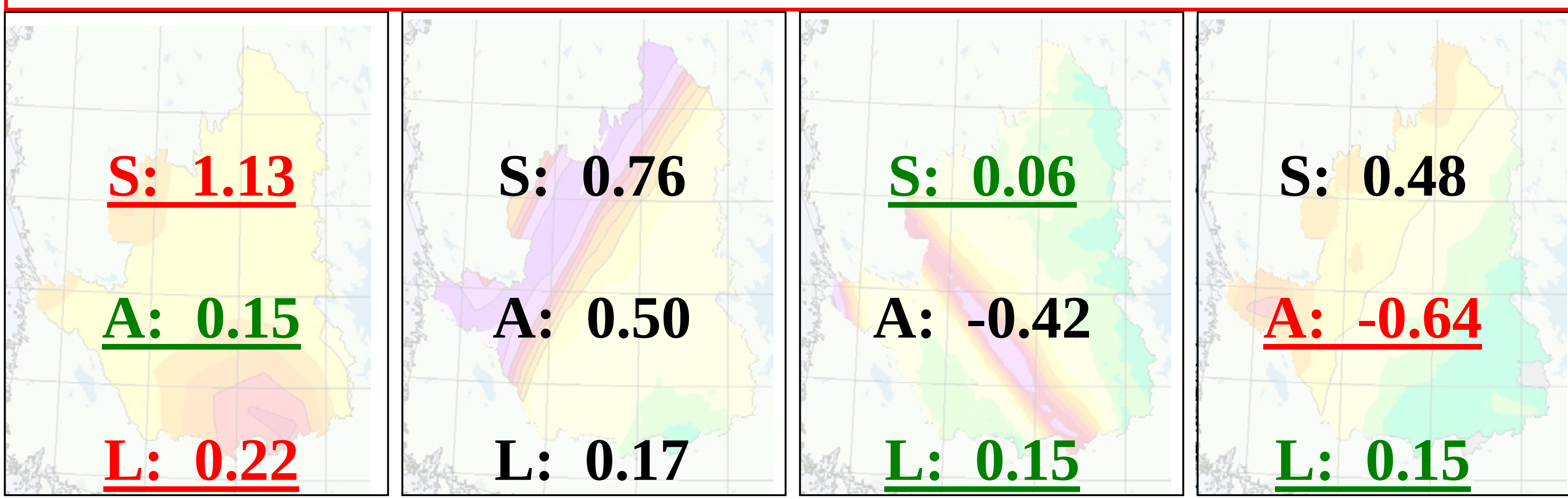
Medium-size catchment

17 August 2008;
24 hour accumulated precipitation

Radar ~ 1 km





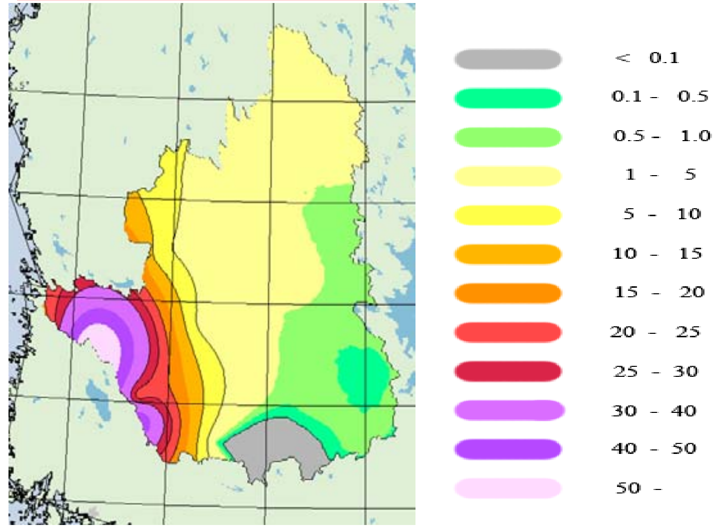


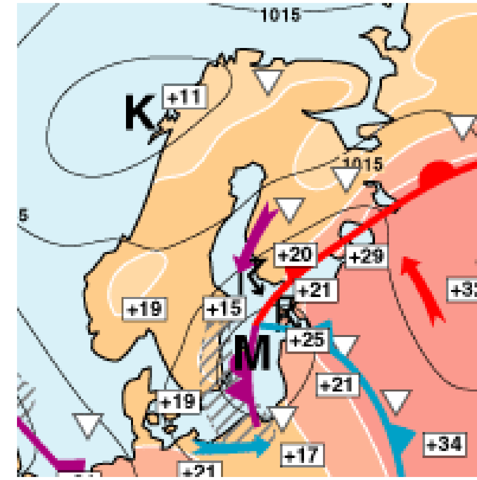
ECMWF ~ 25 km	HIR_RCR ~ 16 km	HIR_MB71 ~ 7.5 km	MET_Edit ~ 15 km
S: 1.13	S: 0.76	S: 0.06	S: 0.48
A: 0.15	A: 0.50	A: -0.42	A: -0.64
L: 0.22	L: 0.17	L: 0.15	L: 0.15

Medium-size catchment

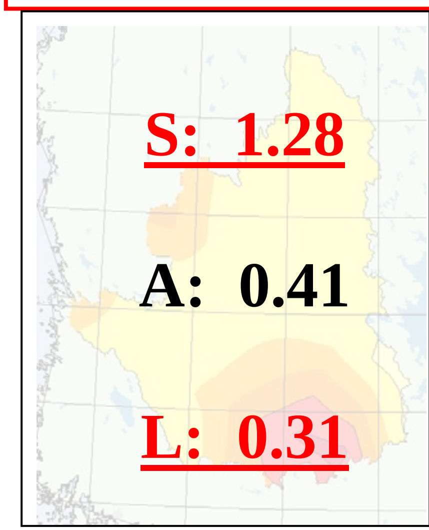
17 August 2008
24 hour accumulated precipitation

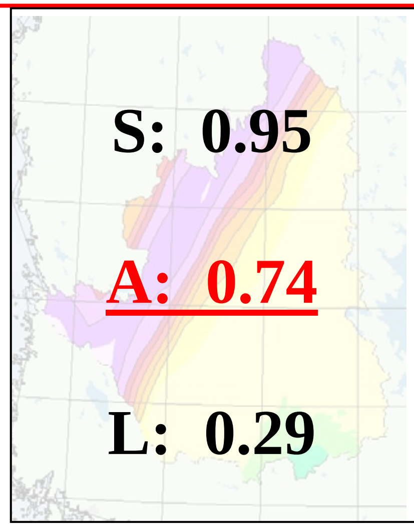
Gauge

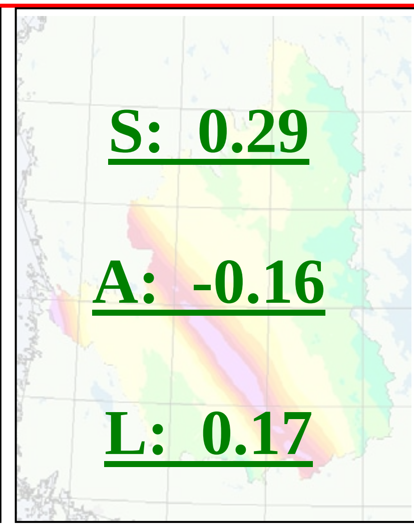


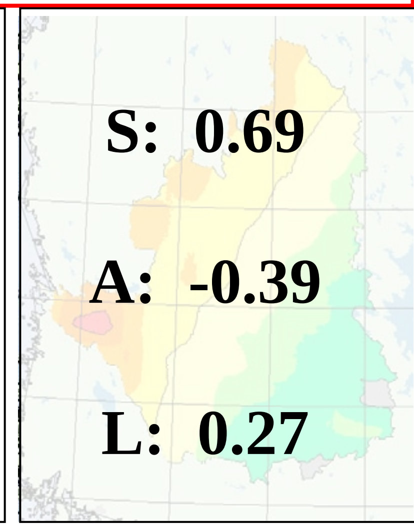


ECMWF ~ 25 km	HIR_RCR ~ 16 km	HIR_MB71 ~ 7.5 km	MET_Edit ~ 15 km





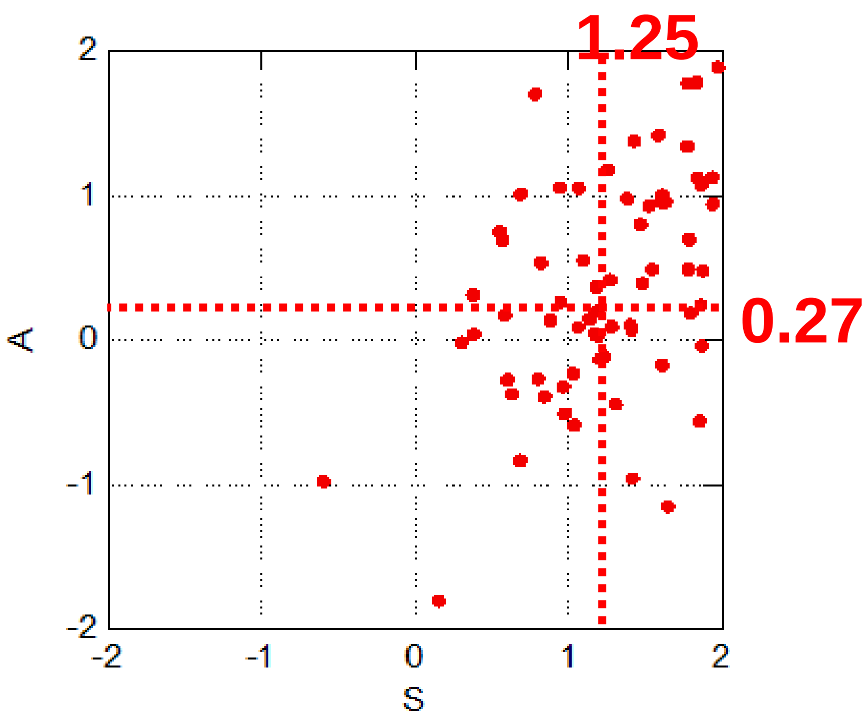




-- Which is the truth?
-- Observation uncertainty

Largest/flattest precip objects ref. winter/spring

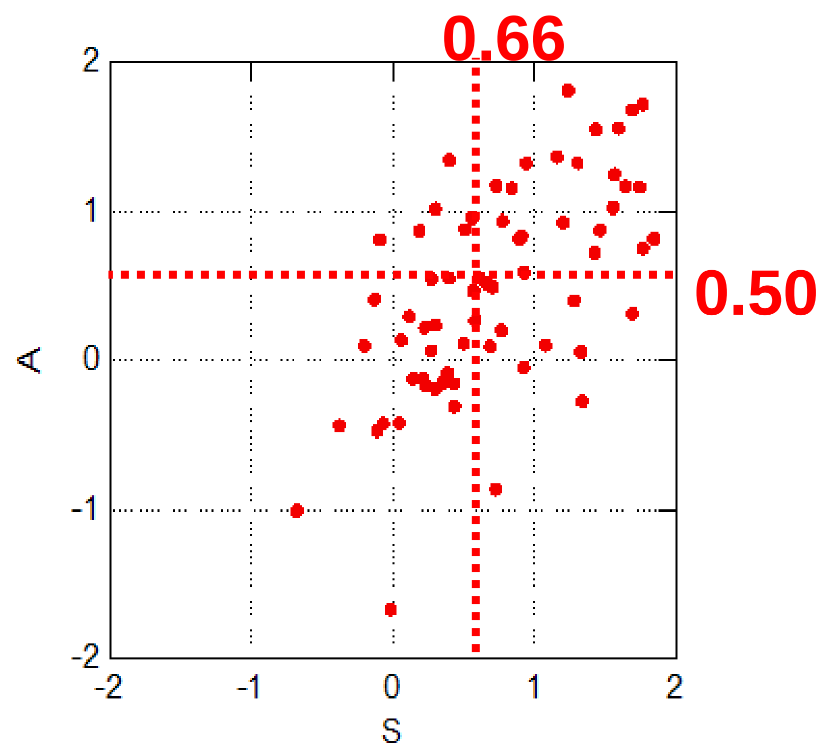
Amplitude somewhat overestimated



ECMWF vs RADAR

Too large/flat precip objects, on average

Amplitude strongly overestimated



ECMWF vs GAUGES

Pertti Nurmi, FMI

Direct approaches for coping with observational uncertainty

Compare forecast error to known observation error

- If forecast error is smaller, then
  - A good forecast

- If forecast error is larger, then
  - A bad forecast

- Bowler (2008)

  -Methods for reconstructing contingency table statistics, taking into account errors in classification of observations

- Ciach and Krajewski (1999)

  -Decomposition of RMSE into components due to “true” forecast errors and observation errors

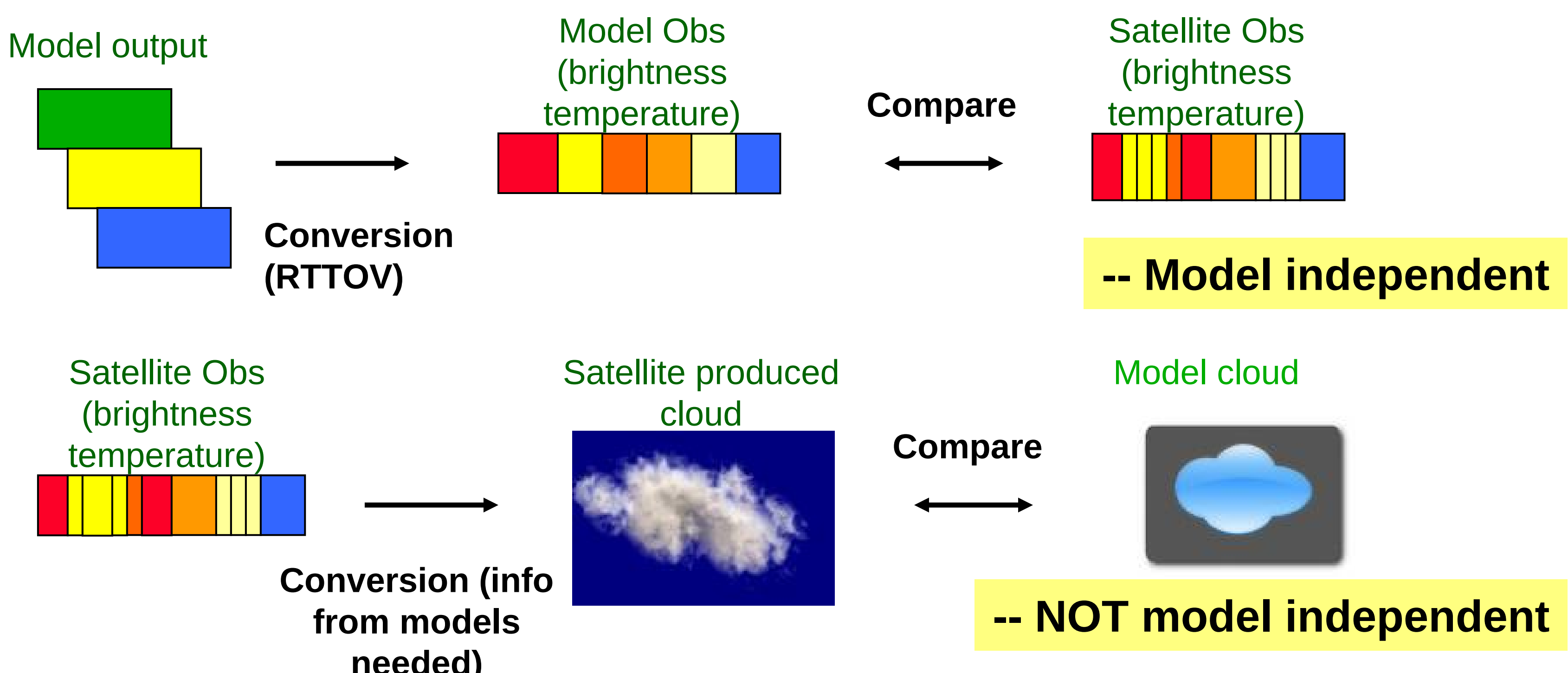
Direct approaches for taking uncertainty into account

- Candille and Talagrand (2008)
  - Treat observations as probabilities (new Brier score decomposition)
- Briggs et al. (2005)
  - Incorporating mis-classification errors using a “gold standard”
- Casati (2008)
  - Wavelet reconstruction
- Roberts and Lean (2008)
  - Perturb pixels in the observed field to obtain error bars
- Hamill (2001)
  - Rank histogram perturbations
- Mittermaier (2008)
  - Incorporation of uncertainty in radar-rainfall estimates

The matching game: Strive for an independent dataset

Approaches:

- Model to observations ➔ model output is manipulated to become comparable to observations
- Observations to model ➔ observations are manipulated to become comparable to model output

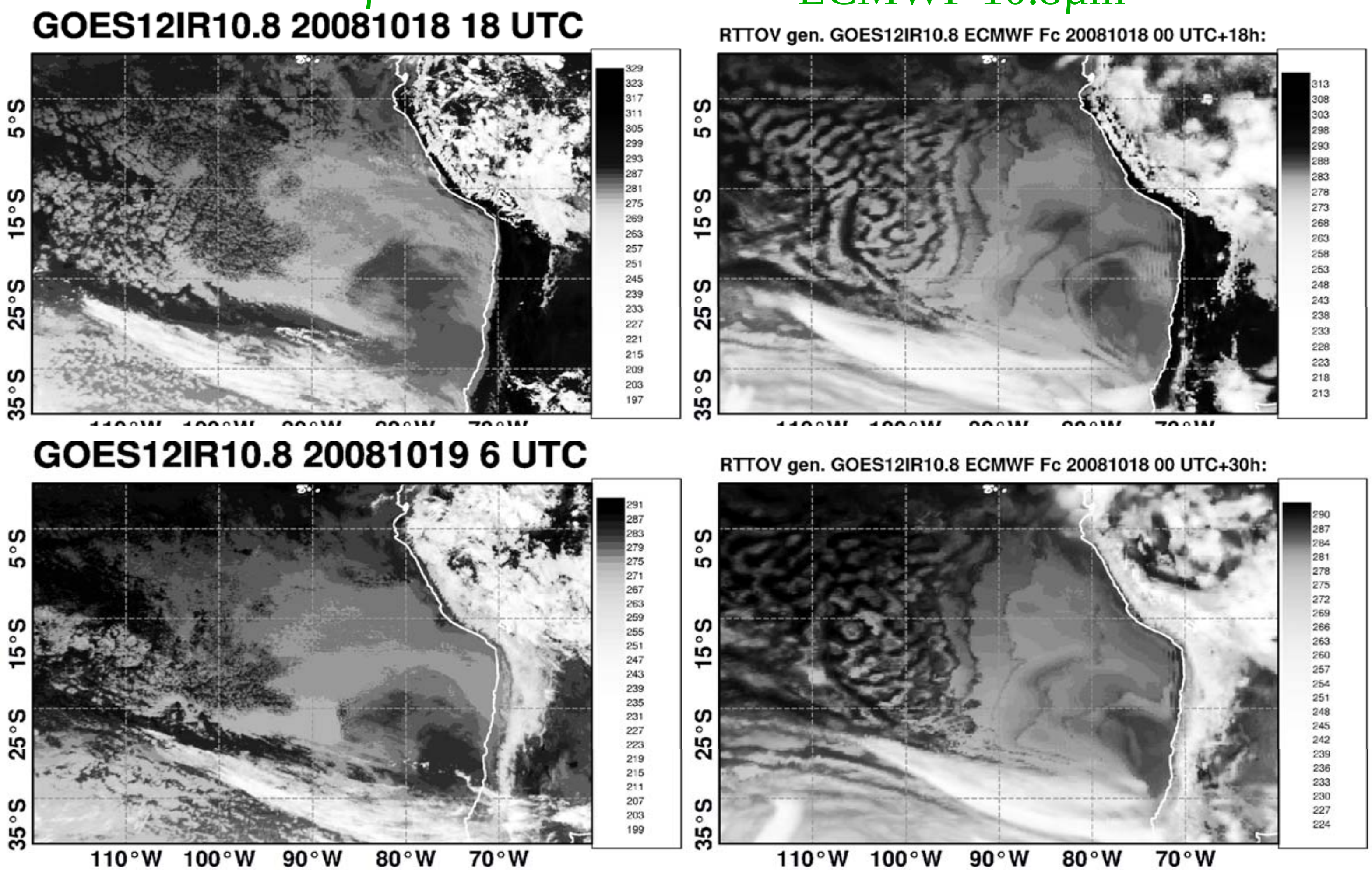


The matching game

VOCALS field experiment off Chile

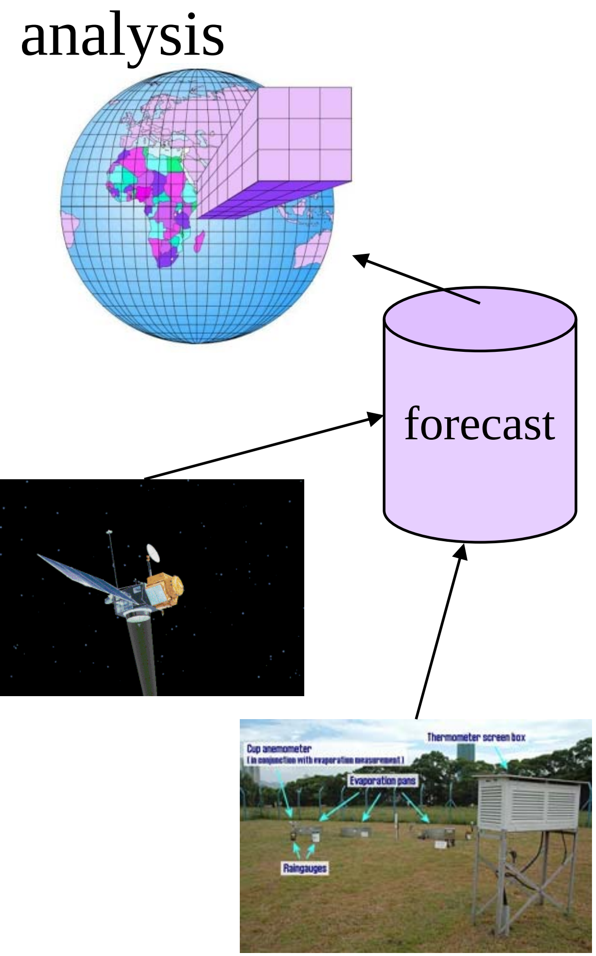
GOES12 10.8µm

ECMWF 10.8µm



The role of the analysis in verification

- **Analyses are model dependent**
  - **Allows to use a number of different type of sensors to provide a coherent analysis for the model ➔ this out-weight the drawback of model contamination**
  - **Good if used for specific purposes e.g. when performance needs to be assessed for scales that the model can resolve and for comparison of same model (operational vs. experimental suite)**
  - **Multi-analysis against observations scores better than single analysis**
  - **Use of randomly drawn analyses for comparative verification of multiple models.**



What do we need for verification purposes?