CONFRONTING MODEL WITH DATA

prepared by F. Bouttier, Aug 2003

Problem : I have a forecast model, how good is it ?

– motivation and methodology
– model climate
– predictive potential
– case studies
– assimilation
– different kinds of data
– mathematics of verification

MOTIVATION AND METHODOLOGY

What is a 'good' model ? It produces data that is consistent with independent verification data.

– It depends on the application of interest,
– we want good-looking publications and talks,
– BUT, safety & sanity also require that the model be physically sound i.e. not just tuned to some specific data.
– statistical models are ok if calibration and validation data are independent.
– a model may fit a dataset and be completely unphysical.

The conditions of usability of the model must be known.

Model validation usually involves generalization hypotheses : period, level, types of weather, photochemical regime...

These hypotheses often fail in some cases (eruptions, long-term climate drift...).

METHODOLOGY : EXAMPLES

- variables : model prognostic ones and well-observed ones
- whole model grid seen as horizontal fields, cross-sections
- timeseries and Hovmöller diagrams, (back-)trajectories
- run the model over 10-50 times the system's characteristic times
- compromise between model run length and number of independent cases
- be critical of data quality (data monitoring and cross-checks) and model postprocessing (e.g. reinterpolation, smoothing)

MATHEMATICS OF VERIFICATION

Mathematically, verification = comparison of a distribution of model output p(m) to a distribution of verifying data p(d).

- usual approach : study distribution of errors p(m-d) (good for continuous parameters with homogeneous errors)
- importance of statistical significance testing (Fischer, Student and Mann-Whitney tests) on large datasets with small variations.
- other approach : study joint distribution of model and data p(m,d)=p(m/d)p(d)=p(d/m)p(m) (=data stratification vs model or data)
- usually studied by contingency tables on discrete events : False Alarm Rate, Hit Rate, etc.
- scores on contingency tables : False Alarm Rate, Hit Rate, Equitable Threat Score, Critical Success Index...

contingency table

	Event is observed	Event is not observed
Event is forecast	Hit	False Alarm
Event is not forecast	Miss	Correct Rejection

- FAR=FA/(FA+CR) false alarm rate
- HR=H/(H+M) hit rate

MATHEMATICS (2) : PROBABILISTIC ASPECTS

If uncertainties cannot be neglected (in model and/or data), comparisons have to be done in terms of PDFs.

Approach 1 : general probability score (Continuous Ranked Probabilistic Score = rms difference of model and data cumulative PDFs). Rarely used. Most useful : histograms of predicted vs. observed frequency of each event = "climate of the PDFs".

Approach 2 : consider probabilities of crossing predefined thresholds as discrete events. Build contingency table for each choice of threshold.

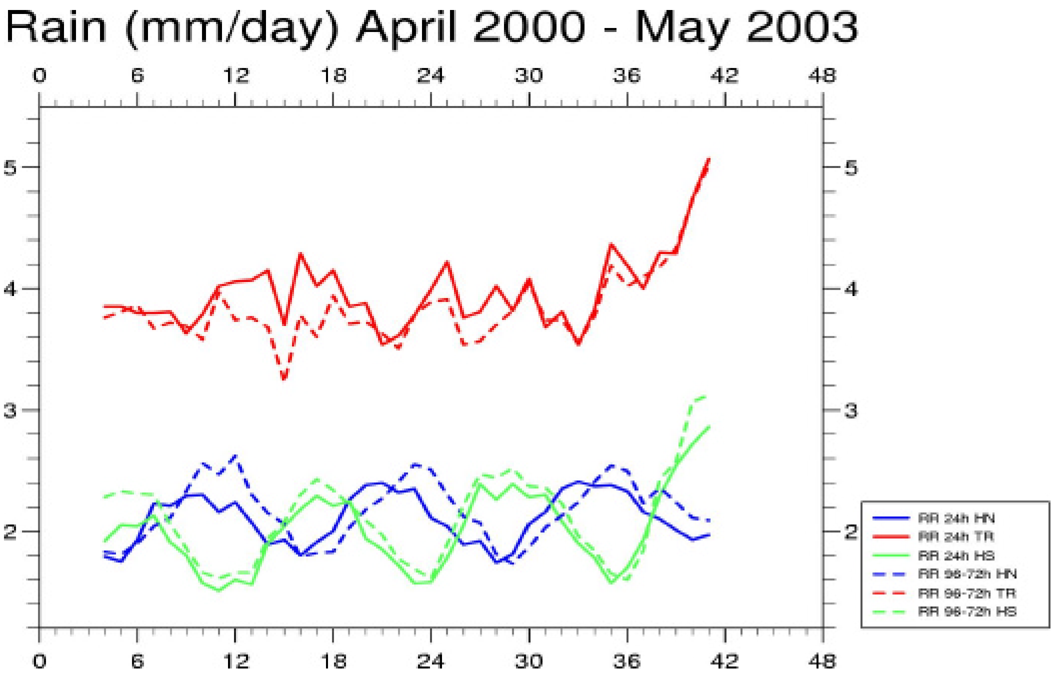
- the (FAR,HR)=f(threshold) is the Relative Operating Curve. Shows whether the events are over- or under-predicted.
- economic value after definition of 'Cost' of an event prediction and 'Loss' of failing to predict it. The Relative Value diagram is the 'money' saved by the model=f(C/L) vs. climate.

TESTING THE MODEL 'CLIMATE'

Question : Is the average model behaviour reasonable ?

- check broad features of model output statistics : average and rms departures vs the same statistics from data.
- plot the model output and data and look for systematic pathologies (numerical model noise, drift, subjective patterns, etc.)
- stratify statistics against main causes of variability : latitude, level, season, model forecast range.
- be careful when stratifying against state variables (introduces statistical biases even if model is perfect)
- let the model run for a long time (systematic errors will be more apparent)
- check the model tendencies : are there large non-physical cancellations ?
- check the distribution of extreme values (tricky)
- check scale dependencies : spectral or wavelet analysis.
- check boundary conditions (imposed but maybe wrong !)

example: climate of model rain=f(range)

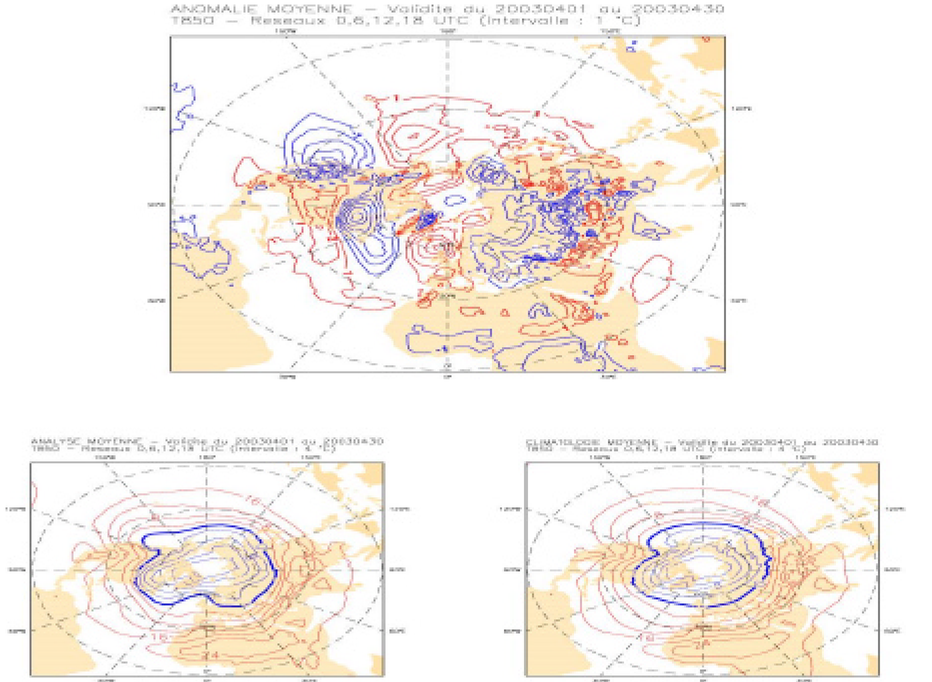


TESTING THE MODEL CLIMATE : EXAMPLES

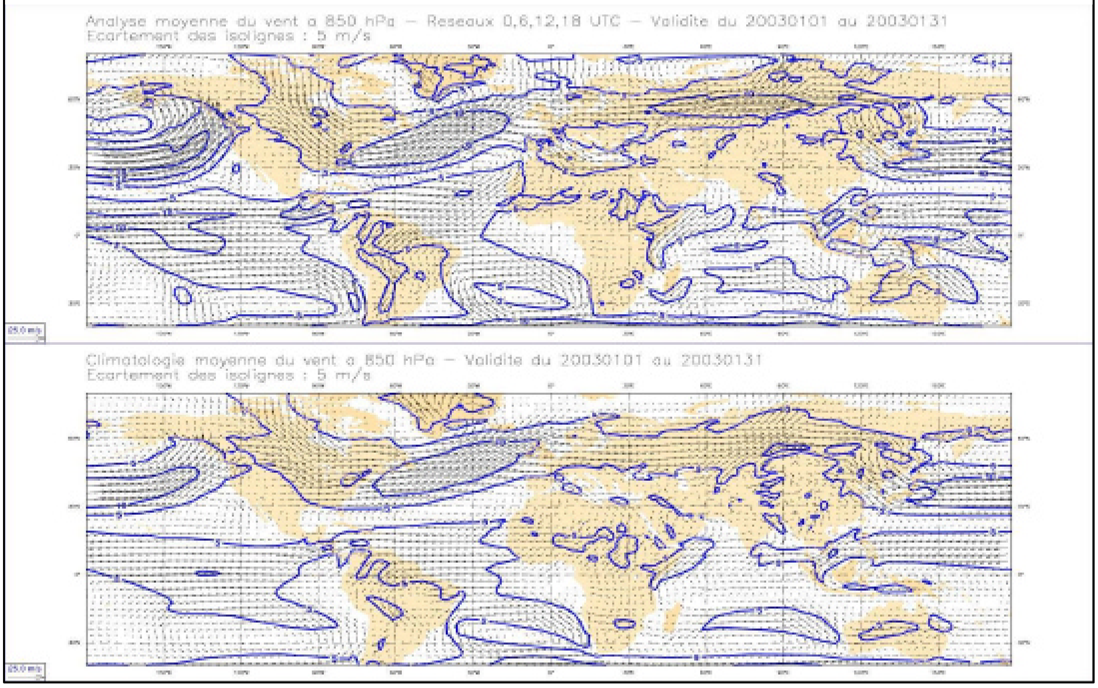
using global atmospheric models :

- vertical N/S cross-sections of state variables and tendencies
- average circulation and its variance
- timeseries of kinetic energy, enstrophy, mass, angular momentum (are there leaks ?)
- frequency of main circulation patterns and waves (blocking, QBO, MJO, El Nino, sudden warmings...)

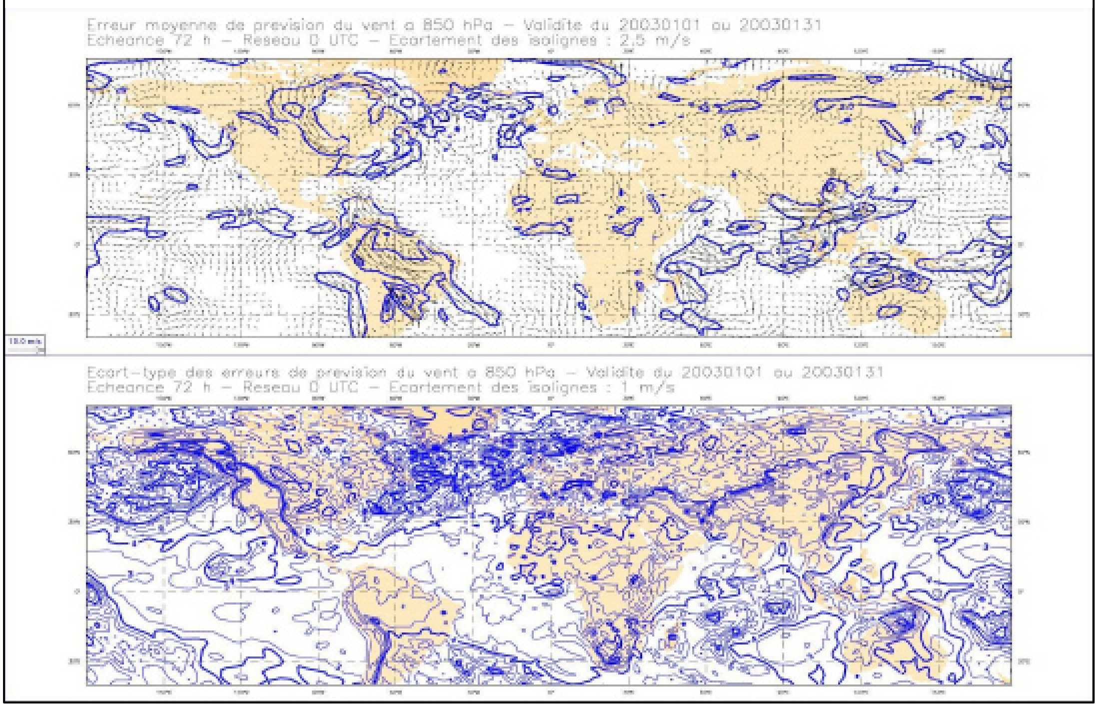
model temperature climate



model wind climate



model wind errors

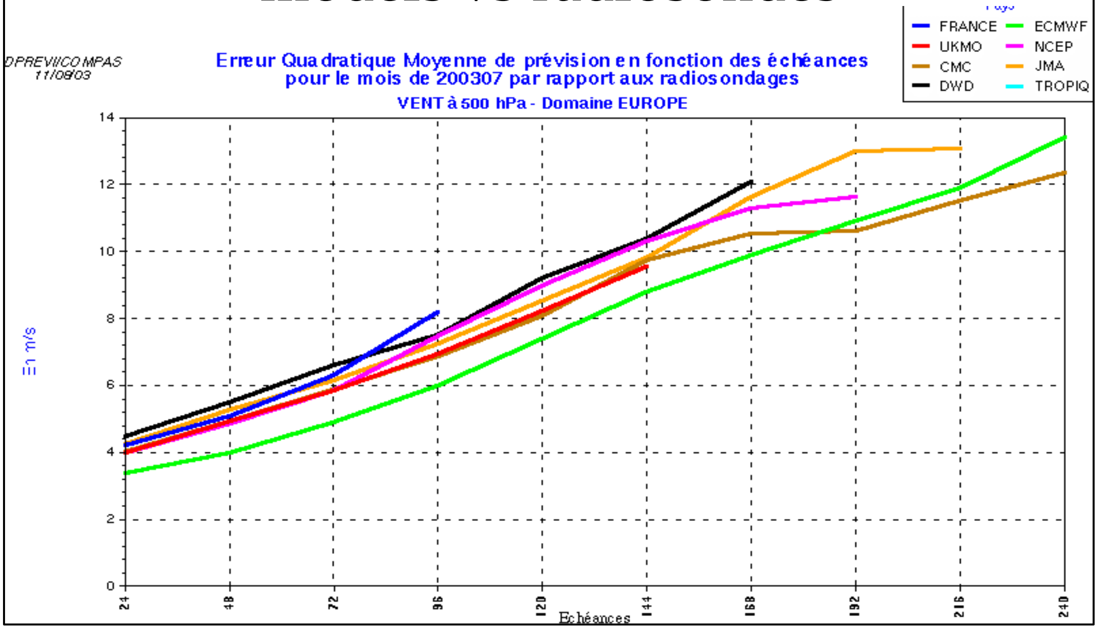


MEASURING PREDICTIVE POTENTIAL

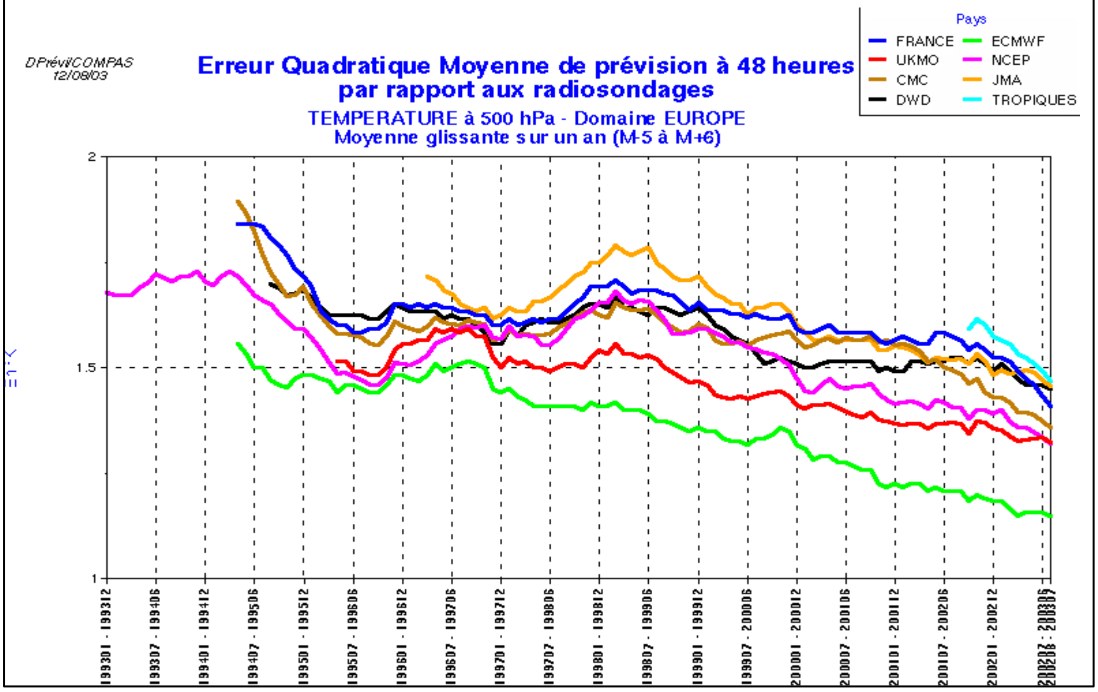
Does the model beat a trivial climate- or persistence-based prediction ?

- requires verifying data (observations, own analysis, reference analysis)
- rms error over each field = f(forecast range, forecast index). Gives an advantage to models that underpredict features.
- anomaly error correlation = correlation of fields of differences wrt. a fixed reference e.g. climate anomalies. Needs structured, big enough errors.
- tendency correlation = correlation of the forecasted/observed evolutions.
- need to separate systematic errors (= biases, perhaps case-dependent) from the random errors.
- validating a model change requires many forecasts to obtain a statistically stable signal in chaotic models. A model physical improvement rarely improves all variables in every forecast.
- validating using observations is tricky in data-poor areas.
- the really useful predicted features may be difficult to assess using scores = apply subjective validation (extreme events, simulated satellite images, etc) or geometrical data analysis (pattern matching, contour extraction)

forecast scores = f(range)
models vs radiosondes



forecast scores, evolution

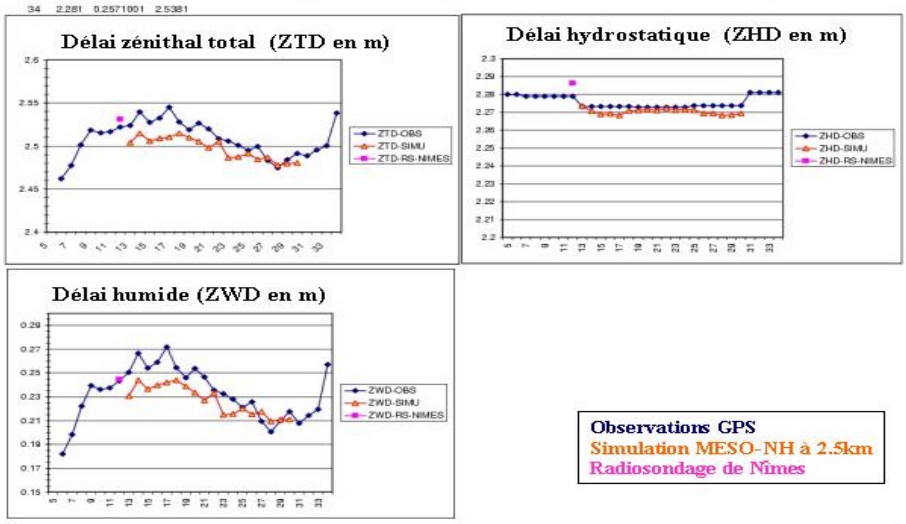


CASE STUDIES

- complementary to average statistical measures of quality
- concentrate on a few events of interest : special event or special data (field experiments, Special Observing Periods)
- thorough evaluation of all aspects of the model
- compare with results from other teams or models on same case (very useful to locate model weaknesses)
- usually targeted towards some specific processes to help their modelling

monitoring using an observable (GPS)

Comparaison des délais zénithaux pour la station de Châteaurenard
pour l'épisode du 8/9 septembre 2002

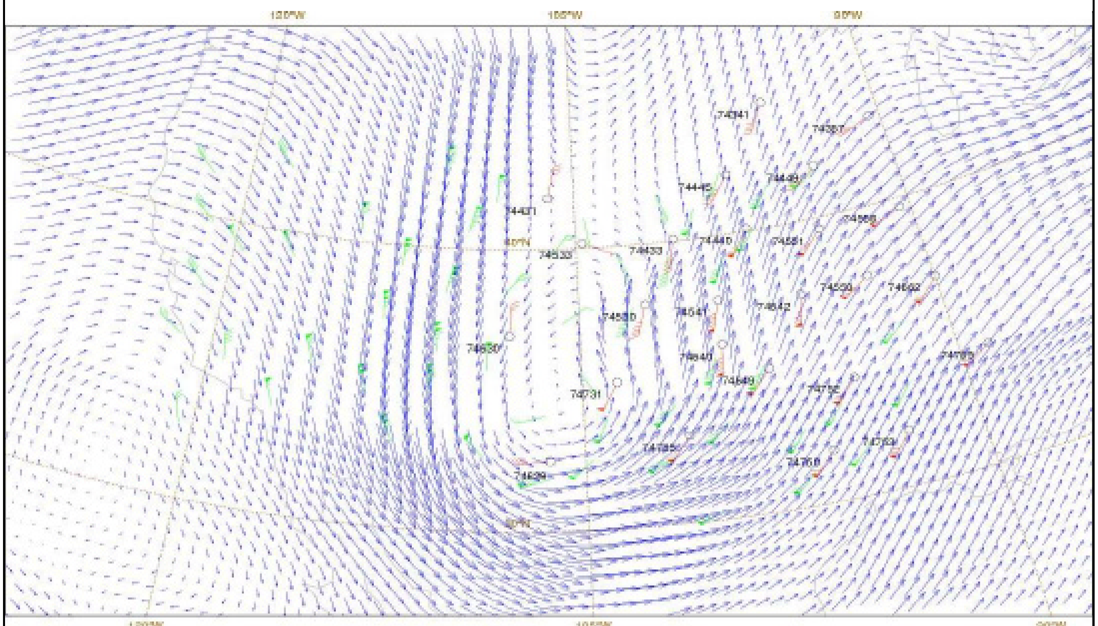


METEO FRANCE

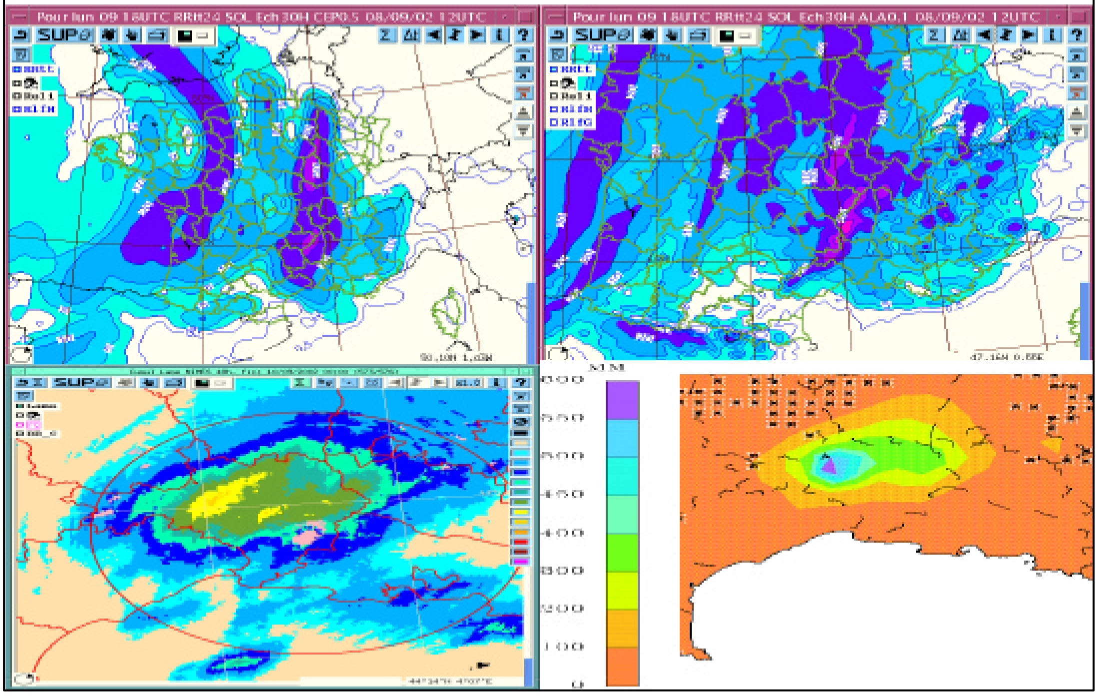
H. Brenot, V. Ducrocq, A. Walpersdorf

LGIT Grenoble

basic obs/model case study (ECMWF wind, aircraft, profilers)



flooding case study: 2 models, radar, raingauges



VERIFICATION OF ASSIMILATION SYSTEM (1)

when the model part of a data assimilation that initializes the forecasts.

Forecast verification : a good assimilation should produce good forecasts

- altering the model will alter the analyses (usually in a chaotic way, for a long time)
- initial steps of the forecasts will have a nonzero departure from the data (observations or reference analysis) except when verifying against own analysis
- verification against own analysis gives an advantage to consistent forecasts : perfect if no obs is used, usually degraded if more obs are used !

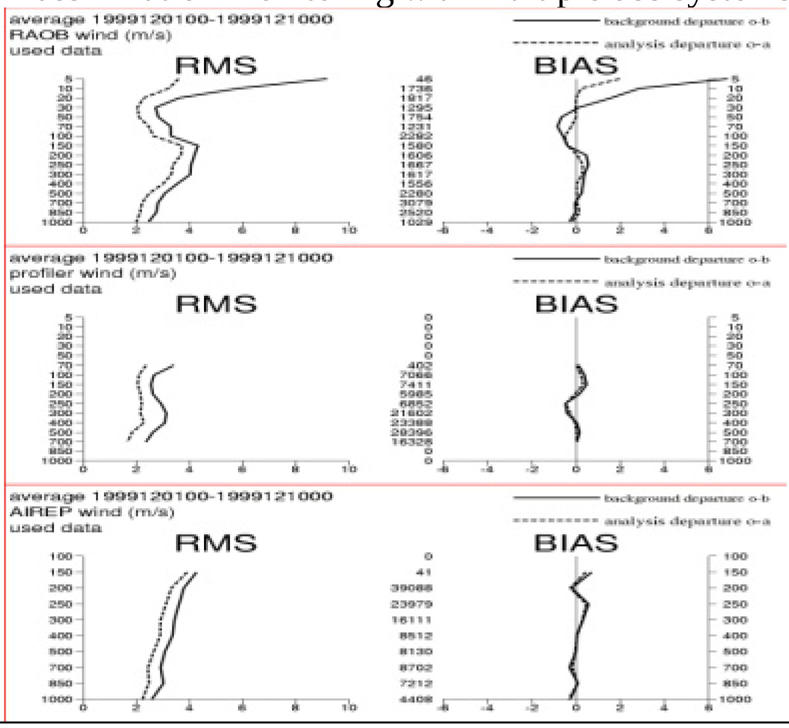
VERIFICATION OF ASSIMILATION SYSTEM (2)

Analysis quality is directly influenced by the model.

Analysis verification : a good analysis should be a good representation of nature.

– the analysis should be reasonably close to observations (within the obs error)
– BUT a better fit does not imply a more realistic analysis or a better forecast.
– more reliable measure : fit of backgrounds to observations (=very short-range forecast errors)
– measured in variational analysis by the observation term Jo
– ideal measure : cross-validation = fit of analysis to non-used data
– other aspect : non-physical transients ('spin-up') in short-range forecasts.
– analysis errors can be investigated using forecast error inversion techniques (adjoint sensitivity, error tracking, pseudo-inverse model)

assimilation monitoring with multiple obs systems



DIFFERENT KINDS OF OBSERVATIONS

Observations are the most objective data, but sometimes complex to use.

- Conventional routine observations (e.g. World Weather Watch, Global Ocean Observing System operational networks, Satellites with operational status) : generally easy to access in near-real time from data centres.
- Acquisition times may be long (preprocessing and transmission)
- There are data problems : need for monitoring (check timeseries and general consistency with model) and rejection of suspect obs.
- Raw remote-sensed data : require expertise, care and substantial preprocessing. Need to talk with a community of instrument experts.
- Retrievals of physical parameters from remote-sensed data can be useful but often dependent on nontrivial hypotheses and arbitrary extra information = complex error characteristics.
- Research data and special observations : require some specific work but often useful to validate specific aspects of the model.
- Imagery data : not suited for quantitative use, but essential for subjective validation.
- Computer animation or 3D graphics can help pinpoint features with a complex structure.

References :

- Atger, F., 1999, *Mon. Wea. Rev.*, 127, 1941-1953.
- Mason, I., 1982, *Aust. Met. Mag.*, 30, 291-303.
- Murphy, 1977, *Mon. Wea. Rev.*, 105, 803-816.
- recommendations of the World Meteorological Organization for verifications techniques.