General description of CANARI analysis software

François Bouyssel

with inputs from F. Taillefer, C. Soci, A. Horanyi, J. Jerman, S. Ivatek-Sahdan, …

Plan

- Brief history of CANARI
- Optimal Interpolation
- Background and observation error covariances
- Selection of observations
- Quality control of observations
- Namelist parameters
- Use of CANARI « DIAGPACK » for mesoscale PBL analysis
- Exemple of tuning CLS analyses
- Conclusions

CANARI acronym

Code
d'**A**nalyse
Nécessaire à
ARPEGE pour ses
Rejets et son
Initialisation

Code for the
Analysis
Necessary for
ARPEGE for its
Rejects and its
Initialisation

Brief history of CANARI

- 1988: Decision at MF to develop a global analysis based on Optimal Interpolation (OI) « CANARI »
- 1992: ARPEGE operational : T79 L15 C1.0 (cycle10!) with CANARI analysis
- 1993: CANARI adaptation to LAM (ALADIN)
- 1996: CANARI operational in Marocco
- 1997 : CANARI replaced by 3D-VAR at MF (CANARI quality control and surface analysis kept)

Brief history of CANARI

- 1998 : ISBA operational in ARPEGE and ALADIN with a surface analysis for soil moisture and soil temperature

- 1999 : Adaptation of CANARI to add more flexibility (tunings, statistical model for CLS) : « DIAGPACK »

  Operational in Hungary (1999), France (2001), …

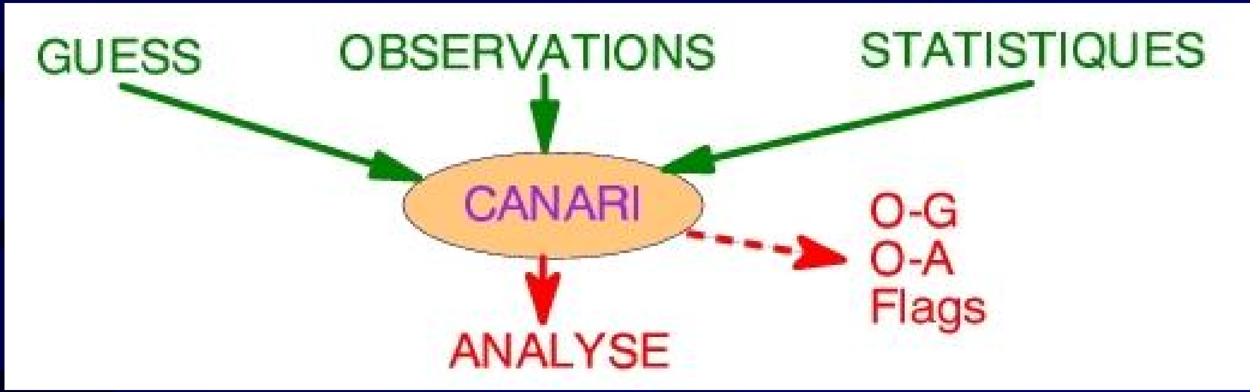
- 2001 : Use of Observation Data Base (ODB) in CANARI

- 2001 & 2003 : Improvments of soil moisture analysis

Generalities

Objective analysis :

- Produce an atmospheric state as close as possible to the reality and at the same time dynamically consistent, taking into account all the available information : observations, model, physical constraints, climatology



Applications of CANARI :

- Quality Control of observations
- Verification of model forecast
- Data assimilation (initial state of a forecast model)
- Nowcasting type of analyses (named CANARI-DIAGPACK)

Optimal Interpolation : basic theory

Based on Best Linear Unbiased Estimation (BLUE) :

$$X^A = X^G + \underbrace{BH^T\left(HBH^T + R\right)^{-1}}_{K}\left(Y - HX^G\right)$$

with X^A: analysed state vector

X^G: background state vector

Y: observation vector

H: observation operator (model space to observation space)

B: background error covariance matrix

R: observation error covariance matrix

K: gain matrix

Optimal Interpolation : basic theory

$$K = BH^T\left(HBH^T + R\right)^{-1}$$

- Matrix inversion => selection of observations (most informative ones)

$$HBH^T = \left(\overline{(x_i^G - x_i^T)(x_j^G - x_j^T)}\right)_{1\le i\le N;1\le j\le N}$$

$$BH^T = \left(\overline{(x_i^G - x_i^T)(x_A^G - x_A^T)}\right)_{1\le i\le N}$$

(N : nb of obs)

- Computation of HBH^T and BH^T: definition of background error covariances

- Computation of R

Optimal Interpolation : basic theory

- According to the type of observations in x_i^o, the analysis can be:
  - 3D multivariate in: U, V, T, Ps
  - 3D univariate in: RH
  - 2D univariate for CLS fields

- The analysis is performed:
  - for the variables of the forecast model,
  - at the model grid-point,
  - on the model levels

Background and observation error covariances (hypothesis, caracteristics)

- Guess and observations are supposed unbiased

$$\overline{(x_i^G - x_i^T)} = 0 \text{ and } \overline{(x_i^O - x_i^T)} = 0$$

- Observation errors are supposed non correlated (R diagonal matrix)

$$\mathrm{cov}(x_i^O, x_j^O) = \overline{(x_i^O - x_i^T)(x_j^O - x_j^T)} = 0$$

- Guess errors and observation errors are supposed non correlated

$$\mathrm{cov}(x_i^G, x_j^O) = \overline{(x_i^G - x_i^T)(x_j^O - x_j^T)} = 0$$

Background and observation error covariances (hypotheses, caracteristics)

- Homogeneity, isotropy and separability hypotheses for background error correlations :

$$\mathrm{cov}(x_i^G, x_j^G) = \sigma_{xi}^G \sigma_{xj}^G cor(x_i^G, x_j^G)$$
$$= \sigma_{xi}^G \sigma_{xj}^G corh(x_i^G, x_j^G).corv(x_i^G, x_j^G)$$
$$= \sigma_{xi}^G \sigma_{xj}^G f(r).g(z)$$

$$f(r) = \exp\left(-\frac{1}{2}\left(\frac{r}{d}\right)^2\right) \text{ and } \mathrm{g(z)} = \frac{1}{1 + pz^2}$$

Distance between i and j points

Characteristic parameters

ln(Pi/Pj)

Background and observation error covariances (hypothesis, caracteristics)

- Following variables are use to define the statistical model: geopotential (ϕ), streamfunction (ψ), potential velocity (χ) and specific humidity (q) with hydrostatic relation for temperature (T) and Helmotz relation for wind (V)

- The statistical model is determined by :

  - standard errors σ_{ϕ}, σ_{HU}
  - characteristic horizontal lengths “a” for ϕ,ψ, “b” for χ, “c” for q
  - characteristic vertical lengths “k” for ϕ, ψ and χ, “l” for q
  - coefficients μ related to geostrophism, ν to divergence
  - slow variations with latitude and altitude of statistical parameters
  - dependency to the stretching factor (for ARPEGE)

Background and observation error covariances (hypothesis, caracteristics)

- For the boundary layer parameters (T_{2m}, HU_{2m}, U_{10m}, V_{10m}), snow, SST specific statistical models are defined.

- There are no cross-correlation between these different parameters.

- On the vertical the auto-correlation is always one (the analysis is done on height surface), but to allow the use of boundary layer parameters in upperair analysis, we define a vertical correlation between U/V/T and $U_{10m}/V_{10m}/T_{2m}$ with a characteristic parameter height to define that limit into the boundary layer the impact of a surface observation

Observations in CANARI

- *OBSERVATION:* ensemble of measured parameters with a given type of instrument at a moment of time (ex: SYNOP, TEMP)

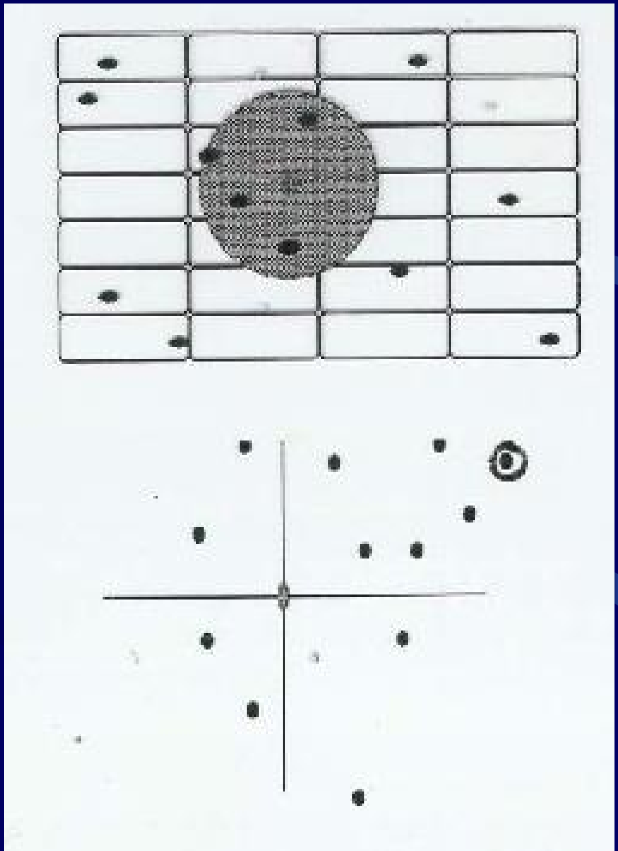
- *DATA:* a measured parameter at a given level and certain moment of time (ex: T at 850hPa)

- 10 types of observations classified:
  - SYNOP: Ps, 2m T and Rh, 10m Wind, Prec, Snow depth (SST if possible)
  - AIREP: P (or Z), Wind, T
  - SATOB: P, Wind, T - from geostationnary satellite imagery
  - DRIBU: Ps, 2m T, 10m Wind, SST
  - TEMP: P, Wind, T, Q
  - PILOT: Wind with the corresponding Z, (sometimes 10m Wind)
  - SATEM: Q, T retrieved from radiances

Selection of the Observations (I)

STEP 1: Geographical selection

- searching the observations in a cylinder around the point to analyse;
- computing the distance from observations to the point of the analysis and selection of the nearest N observations according with their type;
- selection of the M nearest observations for each type and for every quadrant of the circle.

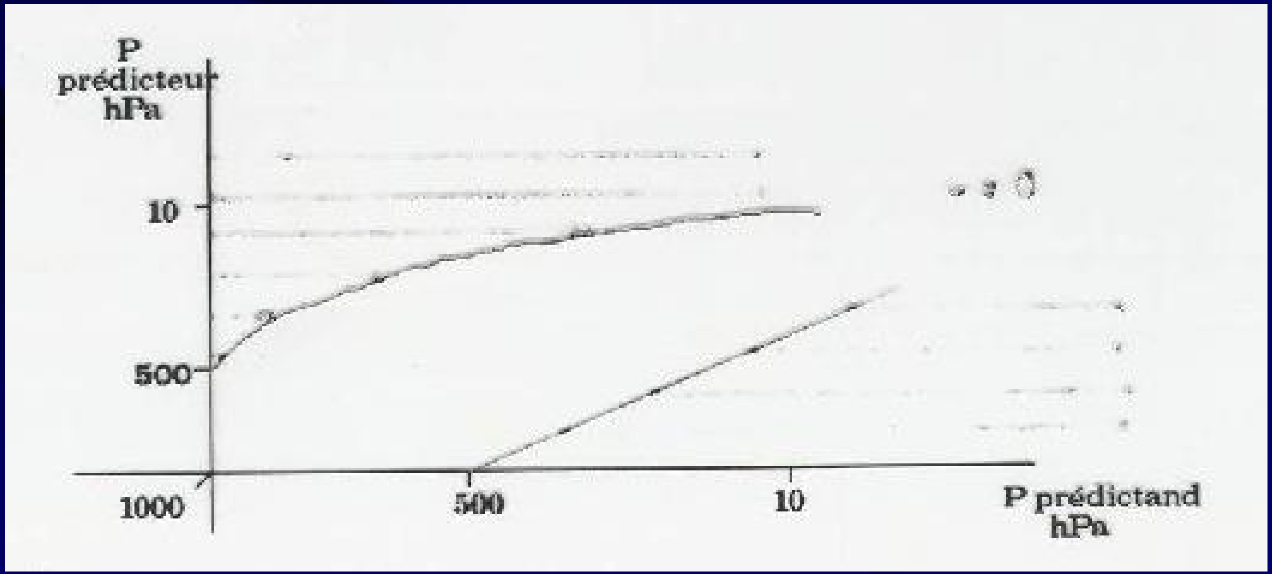


Selection of the Observations (II)

STEP 2: Statistical selection

- Phase 1:
  - selection of the parameters kept after STEP 1
  - eliminating the redundant parameters on the vertical (ΔP min)

- Phase 2: For every vertical point:
  - selection of the parameters located within a ΔP region
  - selection of the best correlated predictors



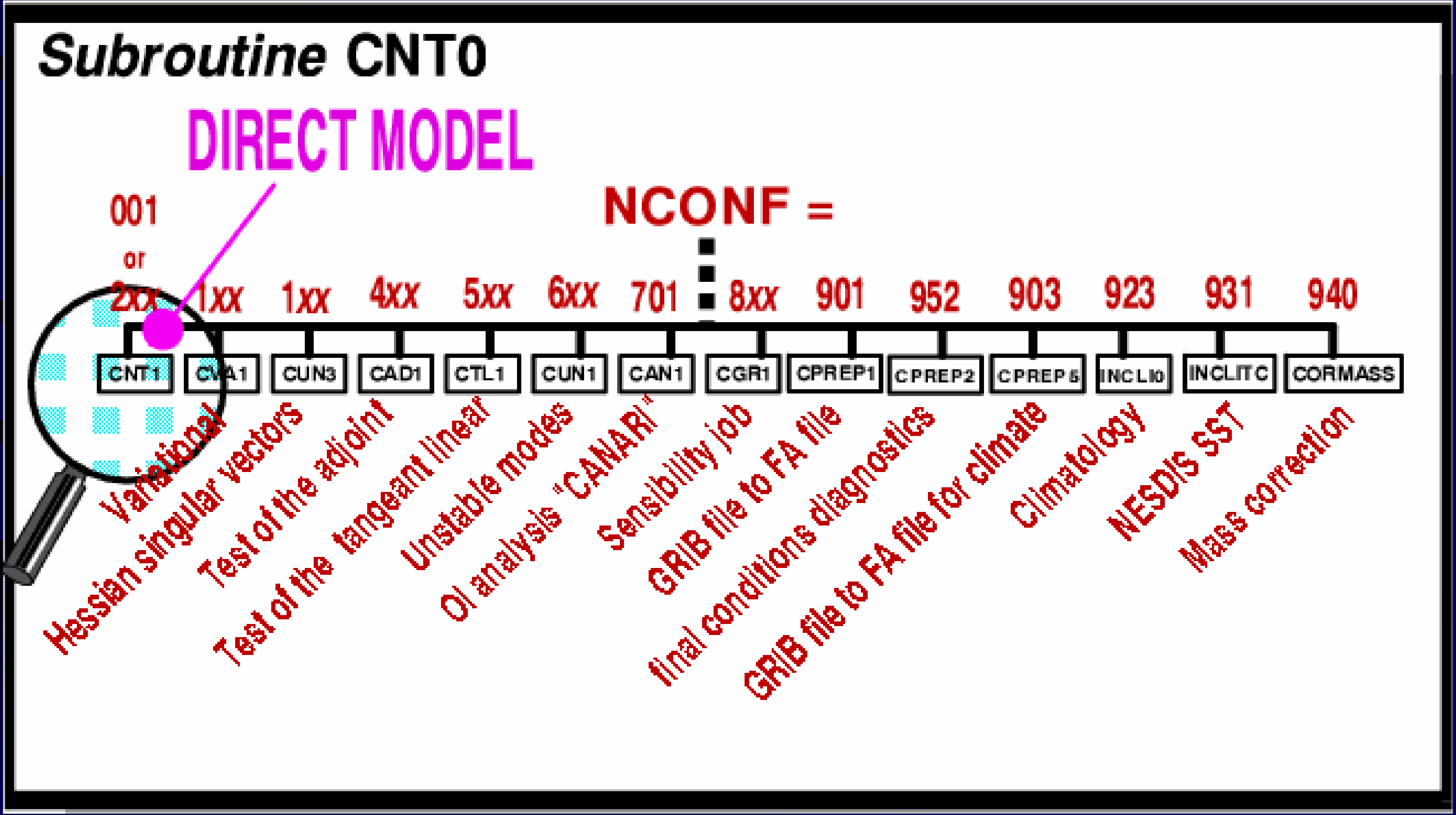
Quality Control of the Observations

- STEP 1: FIRST GUESS CHECK
  - (O – G) compared with standard deviation error $(\sigma_o^2 + \sigma_b^2)^{1/2}$
  - MARKS:
    - 5 - good
    - 3 - doubtful
    - 2 - bad
    - 1 - eliminated

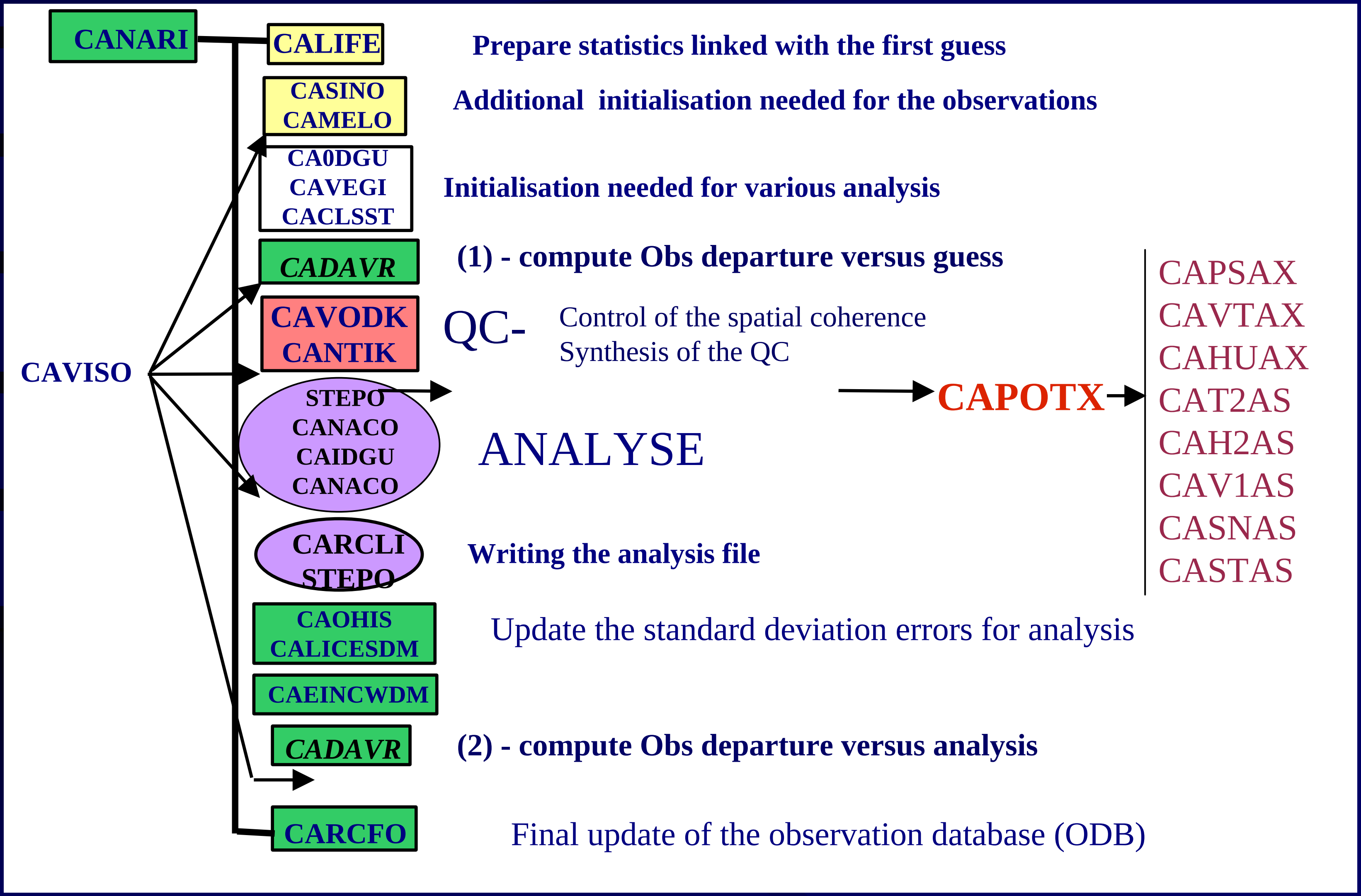
- STEP 2: SPATIAL COHERENCE
  - (O – A) compared with standard deviation error $(\sigma_o^2 + \sigma_a^2)^{1/2}$
  - MARKS:
    - 5 - good
    - 3 - doubtful
    - 2 - bad

- STEP 3: SYNTHESIS OF STEP 1 & STEP2
  - the result from STEP 2 is prevalent when there is no doubt; otherwise the result from STEP 1 become crucial.

Configuration 701



Code description



Various analyses

STEPO → CAPOTX →

	ANALYSIS	PREDICTORS
CAPSAX	Ps	- U, V, Z, T, U10m, V10m
CAVTAX	U, V, T	- Z, T, U, V, layer thickness
CAHUAX	RH	- RH on the level and layer
CAT2AS	T2m	- T2m, T
CAH2AS	RH2m	- RH2m, RH
CAV1AS	U10m, V10m	- U10m, V10m, U, V
(CASNAS	Snow cover	- RR flux, Snow quantity)
CASTAS	SST	- SST
CACSTS	Soil moisture and température	

Namelist parameters

NACTEX : controls the different steps of the analysis

LAEOMF	: calculation O-G
LAEOMN	: calculation O-A
LAECHK	: spatial quality control
LAEPDS	: Ps analysis
LAEUVT	: U, V, T upperair analysis
LAEHUM	: RH upperair analysis
LAET2M	: T2m analysis
LAEH2M	: H2m analysis
LAEV1M	: U10m, V10m analysis
LAESNM	: snow analysis
LAEICS	: soil moisture and soil temperature analysis
LAESTA	: saving of the analysis error statistics
LAERFO	: updating ODB
RCLIMCA	: relaxation coeff for the land surface fields
RCLIMSST	: relaxation coeff for the SST field
NSSTLIS	: use of the NCEP SST in the relaxation field

etc ...

Namelist parameters

NACTAN: defines the analysis area

LANMASK=T : analysis reduced on a geographical domain
ALATNB, ALATSB, ALONWB, ALONEB : domain limits

NACOBS: sets up some observations related variables

OROLIM	: max observation altitude for a SYNOP
ORODIF	: max difference allowed between SYNOP and model heights

NADOCK: defines the observations selection criteria

NMXGQA	: maximum number of observations by quadrant
QDSTRA	: maximum distance for the horizontal selection
QDSTVA	: maximum distance for the vertical selection
MINMA	: predictors number by predictand
QCORMIN	: minimum correlation for the selection by predictand
QDELPI	: minimum distance between 2 selected levels of one observation

NAMCOK: list of the rejection thresholds for the quality control various steps

Namelist parameters

NALORI: contains the coefficient of the function used to take into account the stretching of the grid in the estimation of the correlations (ARPEGE)

NAIMPO: controls some observations related prints

NAM_CANAPE: definition of background error statistics

REF_STAT(.,1)	: pressure of the N levels
REF_STAT(.,2)	: geopotential error standard deviation
REF_STAT(.,3)	: temperature error standard deviation
REF_STAT(.,4)	: wind error standard deviation
REF_STAT(.,5)	: relative humidity error standard deviation
REF_STAT(.,6)	: vertical lengthscale
REF_STAT(.,7)	: horizontal lengthscale
REF_PHUD	: ratio of the horizontal lengthscales for divergence and geopotential
REF_PHHU	: ratio of the horizontal lengthscales for RH and geopotential

REF_COEFN, REF_COEFT, REF_COEFS: dependency of $\sigma\phi$ to latitude

etc ...

Namelist parameters

NAM_CANAPE:

REF_S_SST	: standard error deviation for SST
REF_S_SN	: standard error deviation for Snow
REF_S_T2	: standard error deviation for T2m
REF_S_H2	: standard error deviation for H2m
REF_S_V1	: standard error deviation for U10m, V10m
REF_A_SST	: horizontal lenghtscale for SST
REF_A_SN	: horizontal lenghtscale for Snow
REF_A_T2	: horizontal lenghtscale for T2m
REF_A_H2	: horizontal lenghtscale for H2m
REF_A_VOR1	: horizontal lenghtscale for 10m wind vorticity
REF_A_DIV1	: horizontal lenghtscale for 10m wind divergence
REF_AP_SN	: reference vertical lengthscale for the snow
REF_NU_BL	: ageostrophism coefficient in the boundary layer
REF_KP_BL	: vertical extent coefficient for the boundary layer

CANARI « DIAGPACK »

- IDEA: to be able to analyse some mesoscale features even if it is not possible to keep them in subsequent forecast
- HOW: via detailed analyses of boundary layer fields (high data density at the surface)
- Driving signal for processes depending on boundary layer (e.g. convection, phase of precipitation, ...)

⇒More flexibility in CANARI analysis (more namelist parameters, separation of surface statistical model to upperair)

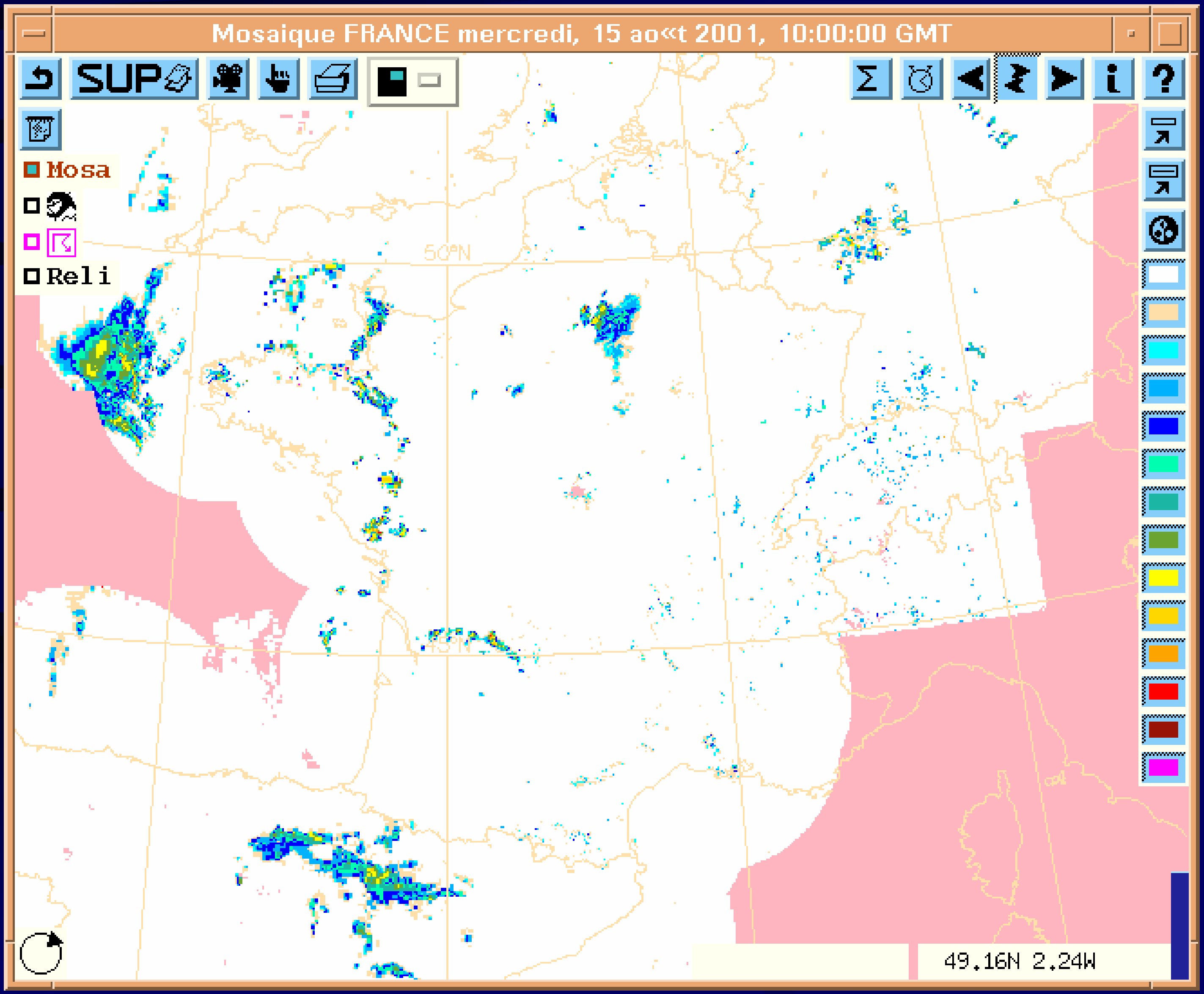
- Operational hourly mesoscale analysis over France of T_{2m}, H_{2m}, V_{10m}, U,V, T, RH at 10km horizontal resolution based essentially on CLS observations (T_{2m}, H_{2m}, V_{10m}, P_s)
- Specific tunings:

```
REF_S_T2  = 3.0,    REF_A_T2   =  40000.,
REF_S_H2  = 0.20,   REF_A_H2   =  40000.,
REF_S_V1  = 5.,     REF_A_VOR1=  60000., REF_A_DIV1= 50000.,
Etc ...
```

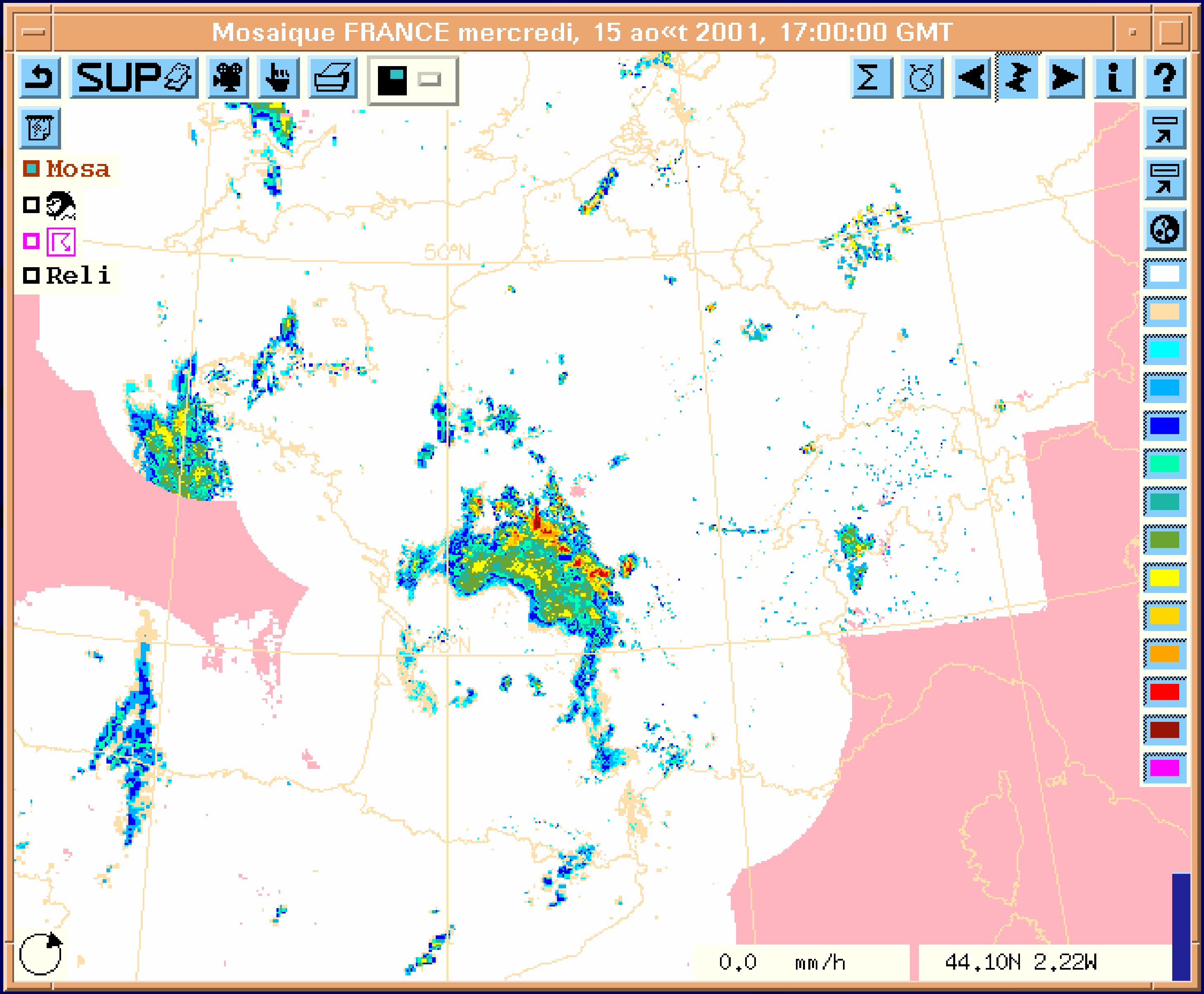
General appreciation by forecasters

- Quality of CANARI « DIAGPACK » analysis as interpolator:
  - limitations over sea and in mountain areas
- Benefit of mesoscale analyses:
  - adding value against ARPEGE and ALADIN analyses
  - adding value against pointing observations
- Interesting to follow the ALADIN forecasting system
- Benefit of analysed convective diagnostics (CAPE, MOCON) not demonstrated

Radar 15/08/2001 animation 10h-23h



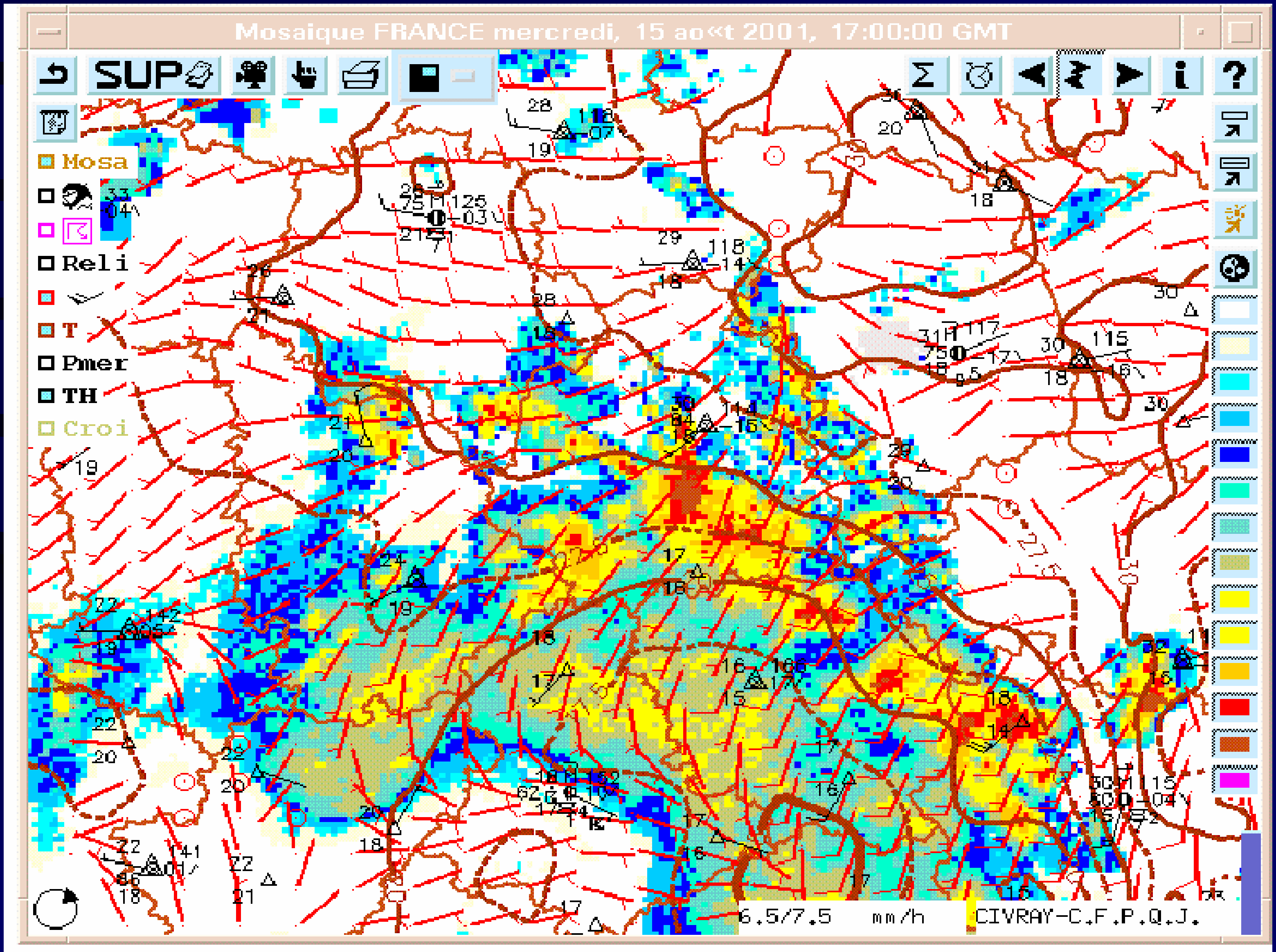
- 17 H



Mosaique FRANCE mercredi, 15 ao«t 2001, 17:00:00 GMT
SUP
Mosa
Reli
50°N
45°N
0.0 mm/h
44.10N 2.22W

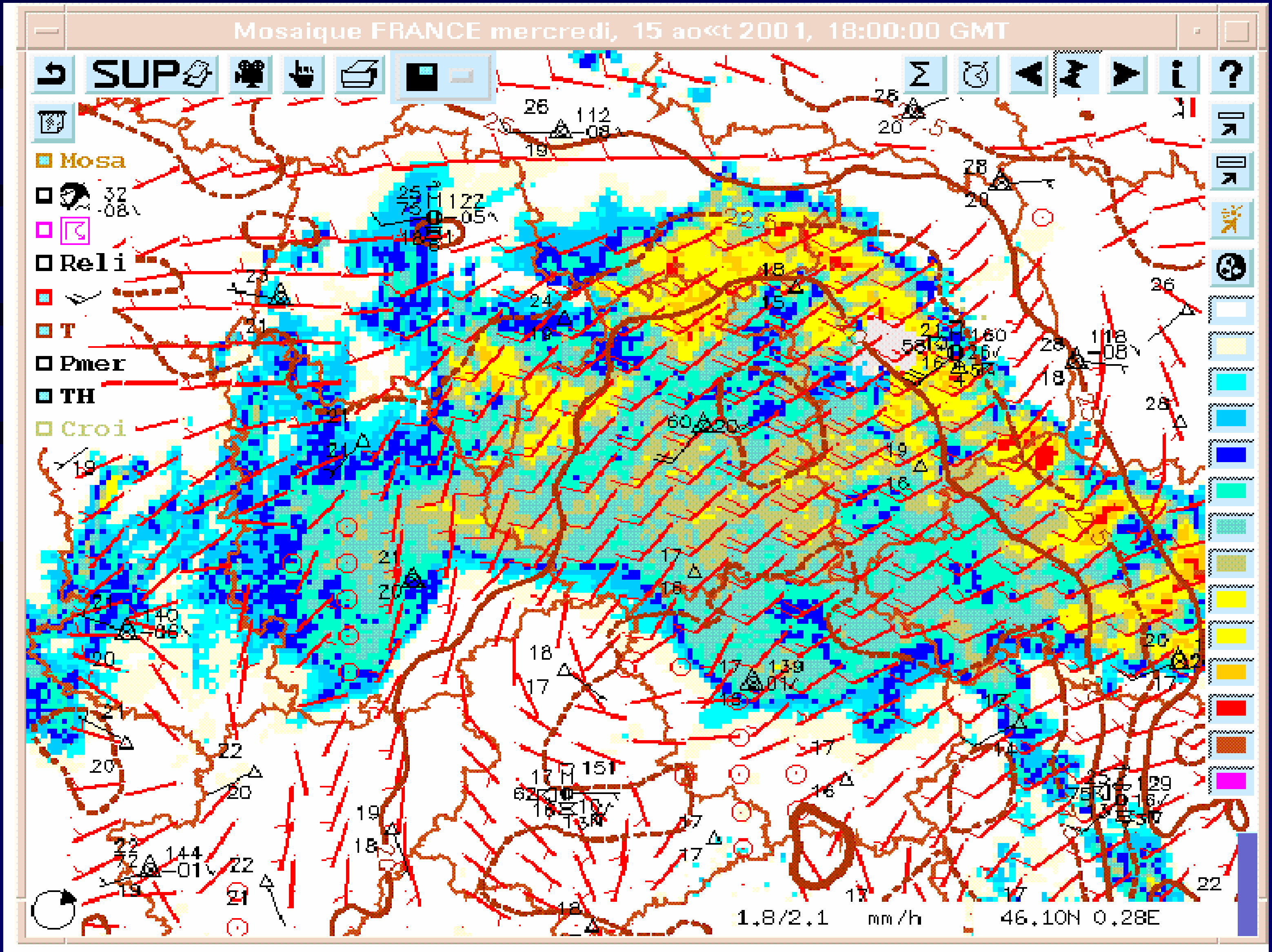
15/08/2001 17h

10m Wind and 2m Temperature:



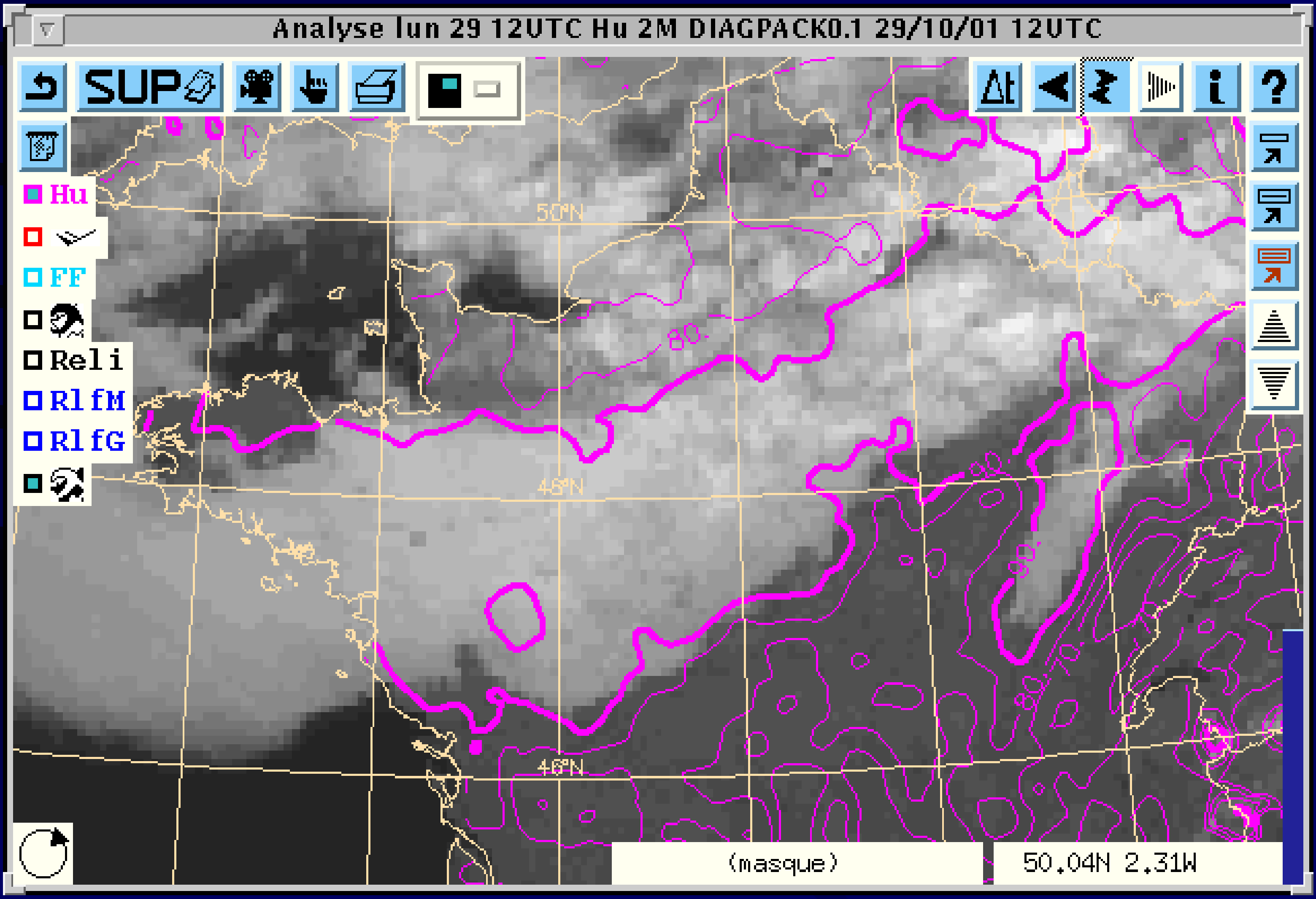
15/08/2001 18h

10m Wind and 2m Temperature:



29/10/2001 à 12 h

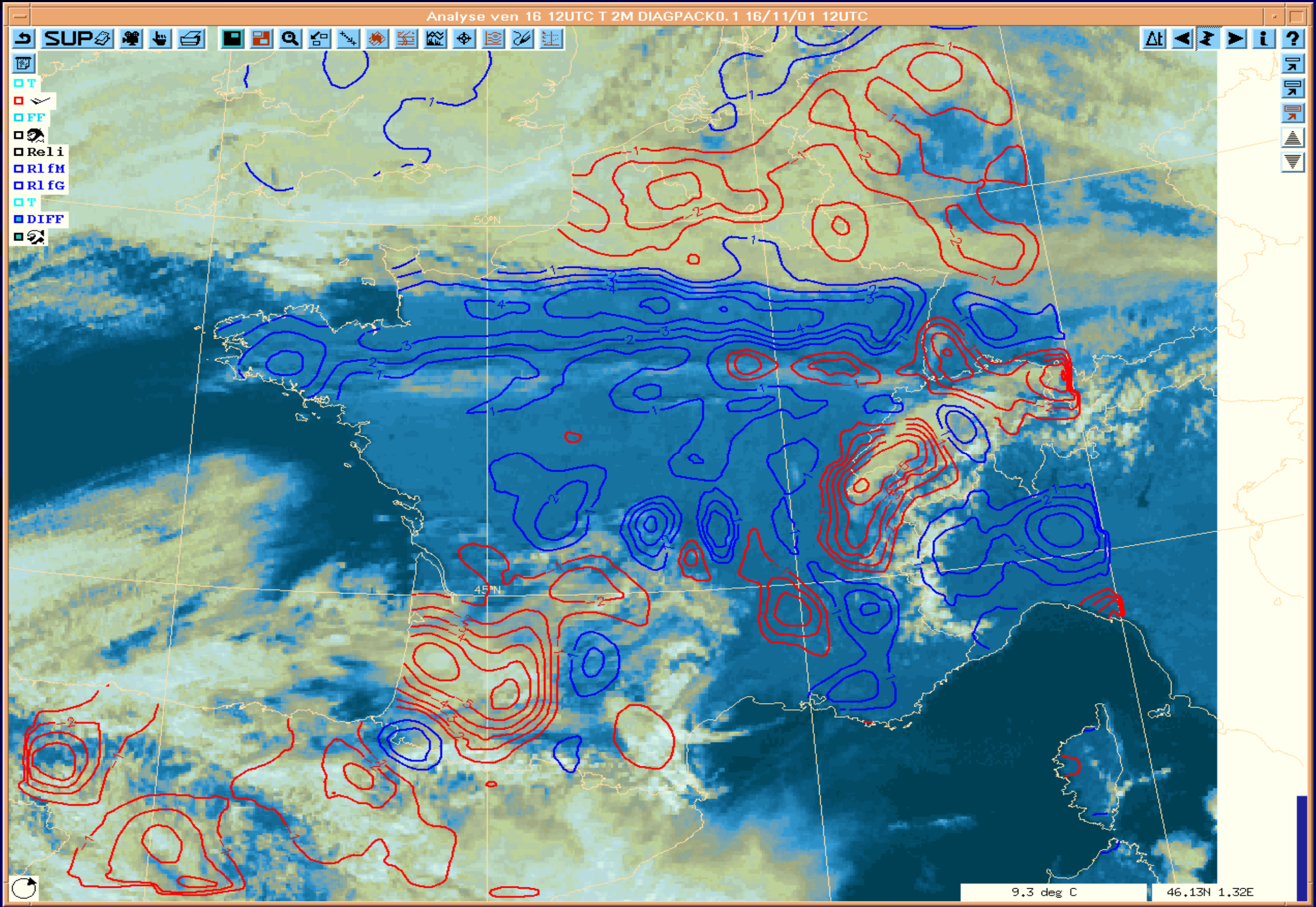
HUMIDITE + Visible METEOSAT:



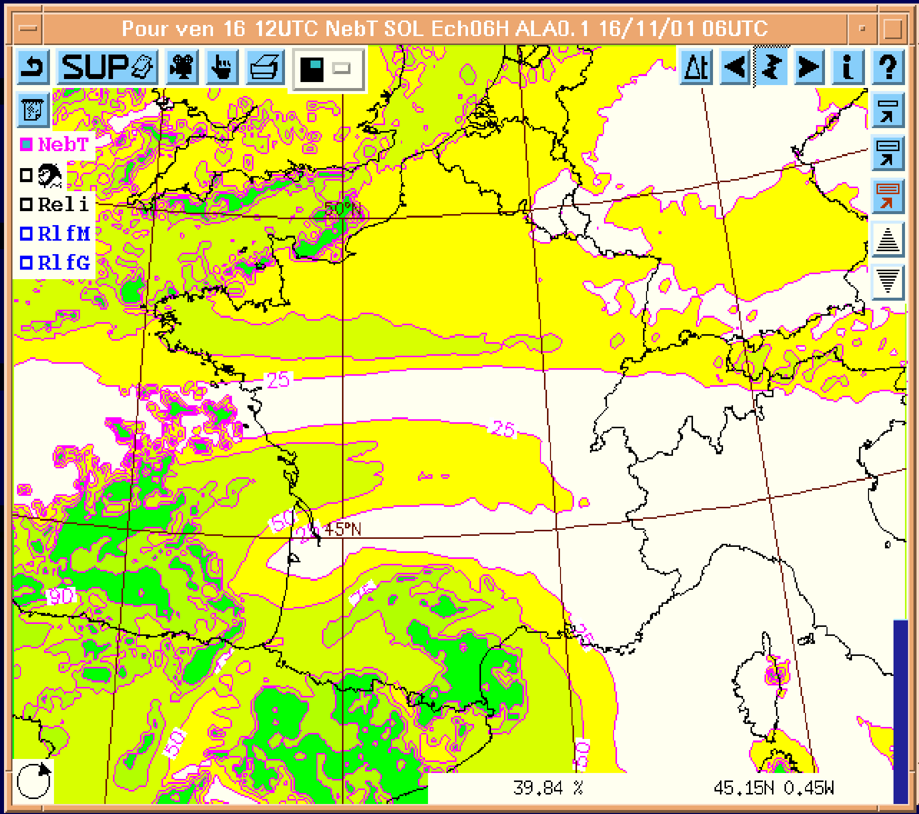
16/11/2001 à 12 h

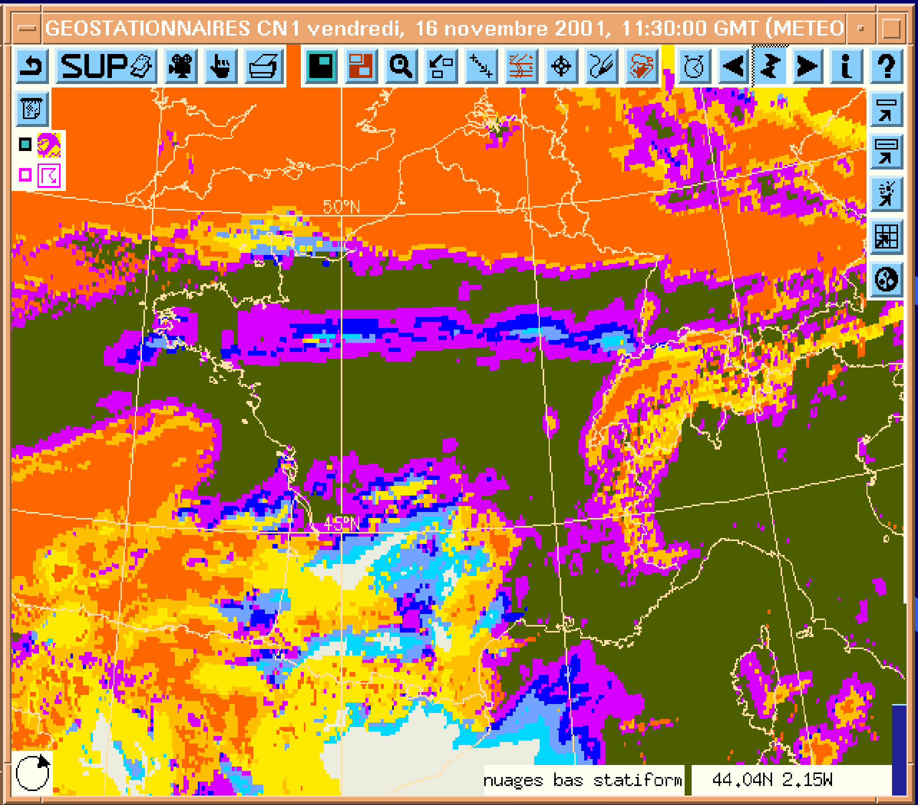
- 2m Temperature / Clouds (visible METEOSAT)

Difference on T2m : Analyse – Guess



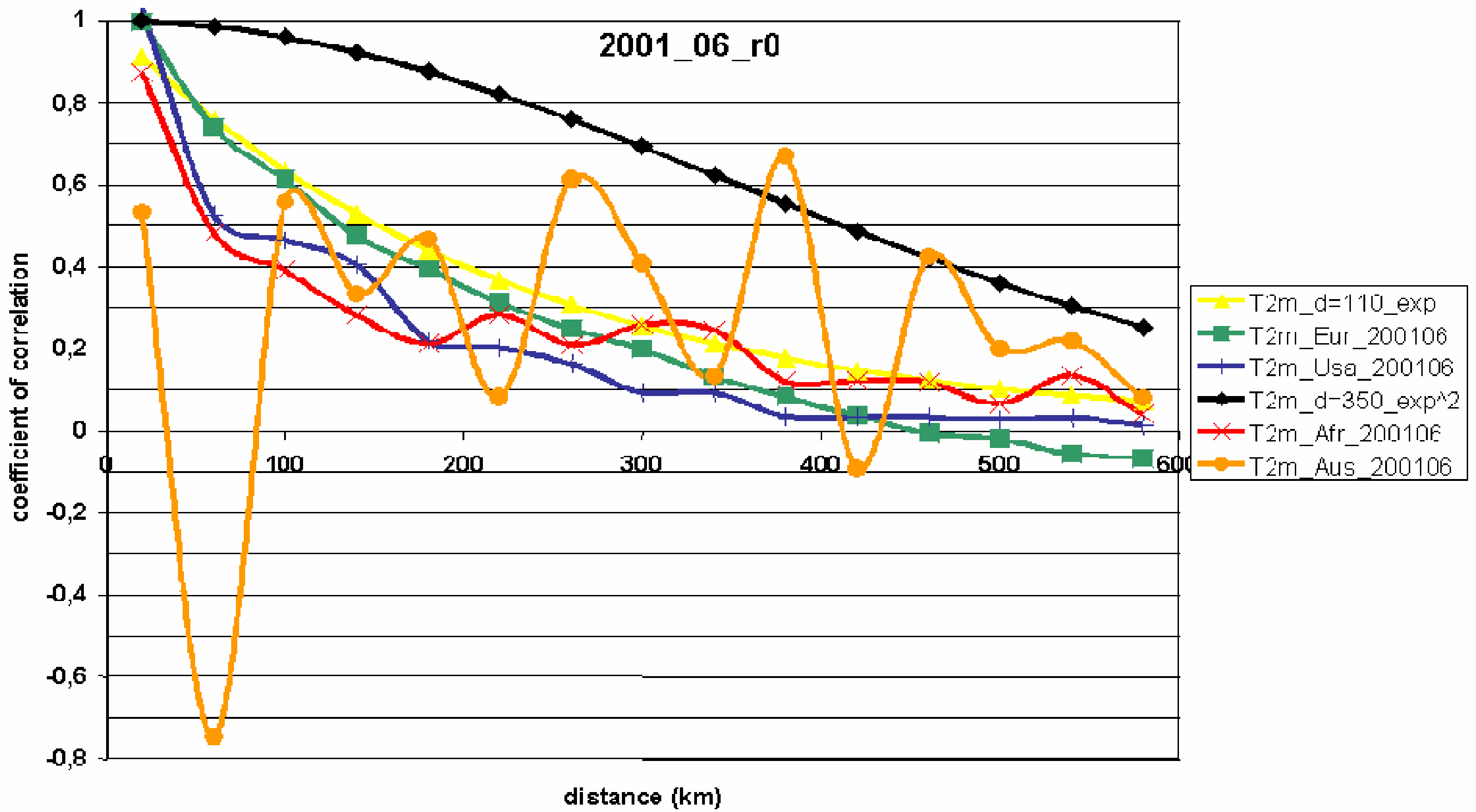
16/11/2001 à 12 h





Exemple of tuning CLS analyses

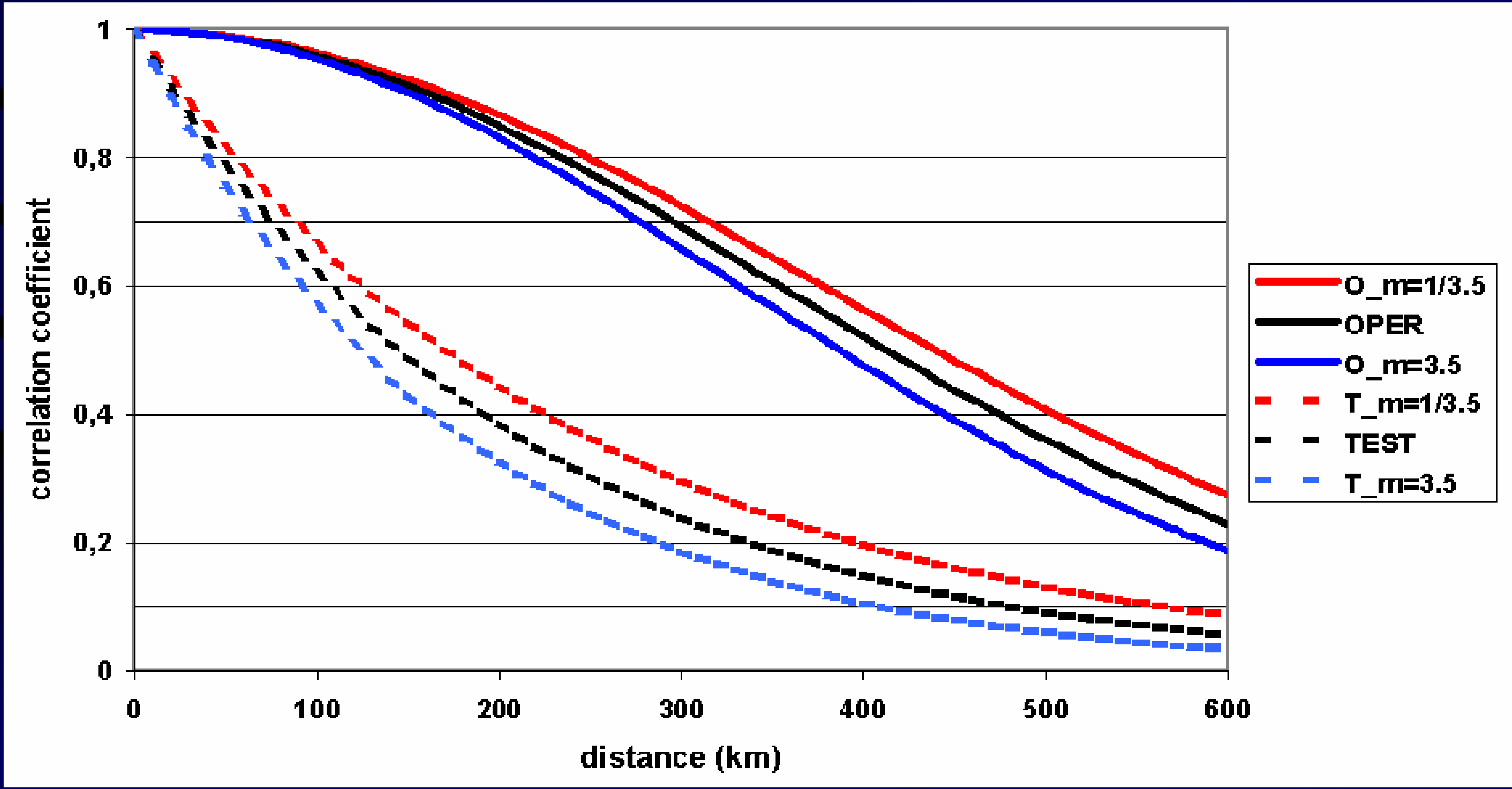
Computing background error statistics for T2m and H2m using (O-G) statistics at the observation location:



Picture 1. Coefficient of correlation dependency to distance between points for different domains, black line represents operational coefficient of correlation, and yellow the new definition

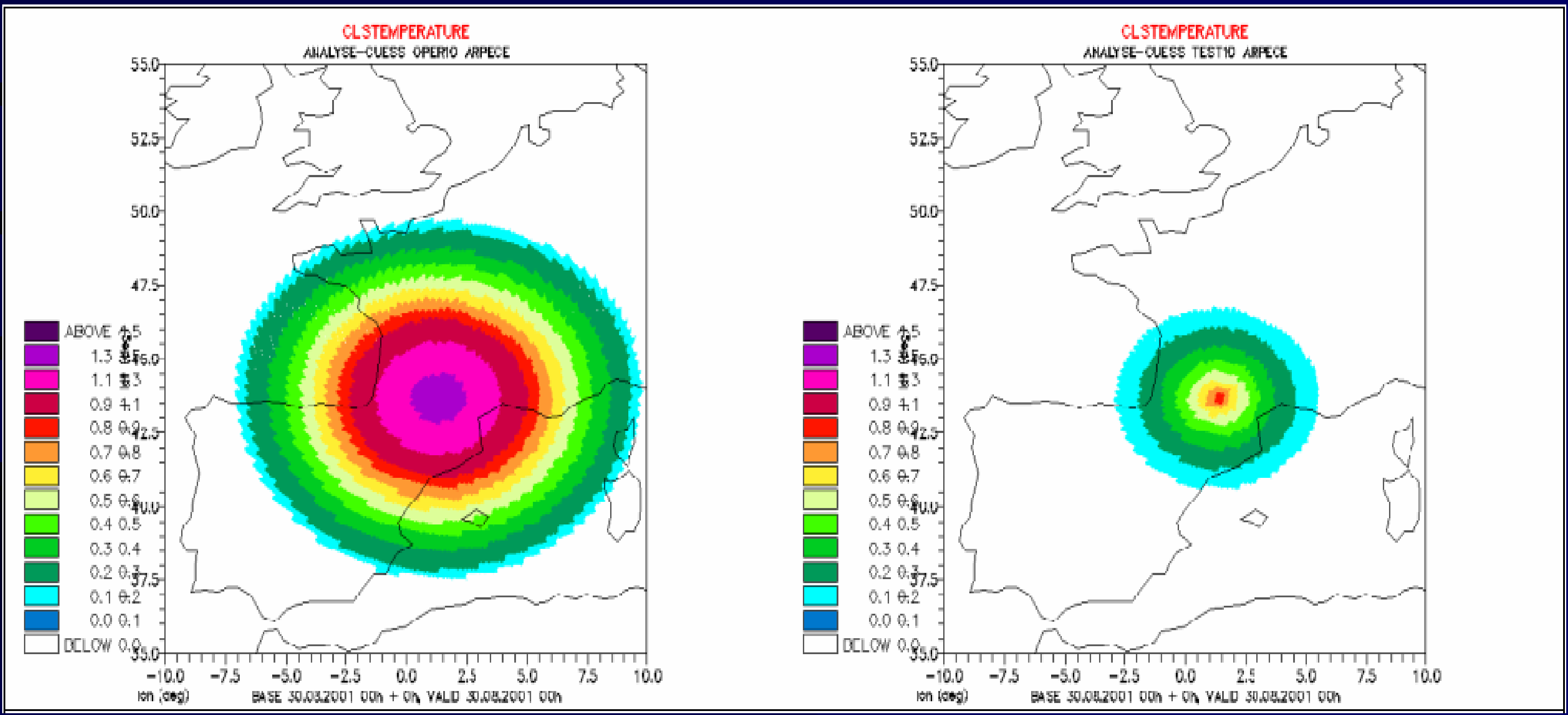
Exemple of tuning CLS analysis

Definition of a new horizontal correlation function (LCORRF):



Tuning CLS analysis

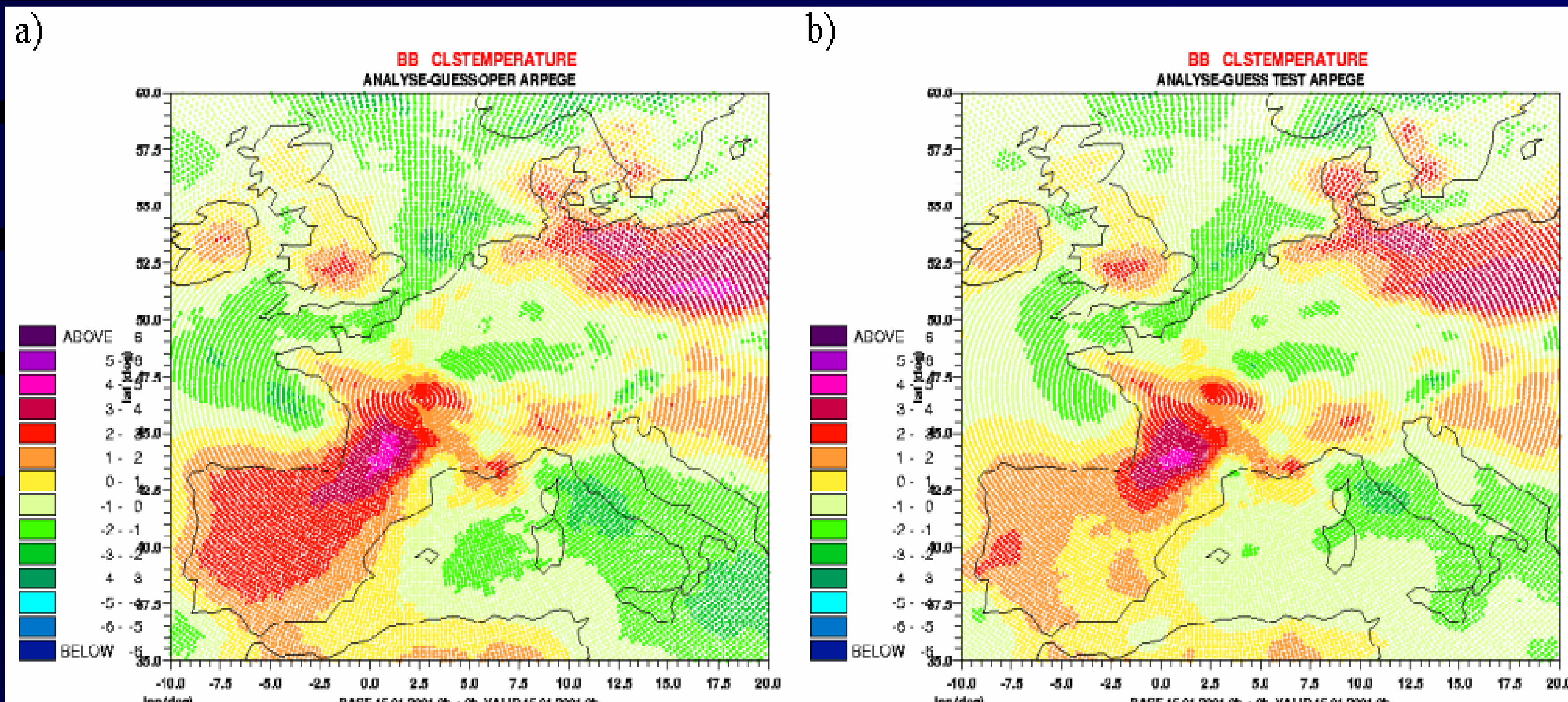
Analysis increment on T2m for a single T2m observation



OLD **NEW**

Tuning CLS analysis

Exemple of analysis increment on T2m



OLD **NEW**

Conclusions

CANARI analysis:

- Strenghts:

  Optimal Interpolation algorithm, good observation quality control, quite simple to use, relatively modular, uses ODB, operational and is part of the official code source ARP/IFS

- Limitations:

  Optimal Interpolation (linear observation operator, instantaneous analysis, selection of observations), statistical model relatively simple (homogeneity, isotropy, separability), no assimilation of satellite raw radiances