Global Ensemble Prediction Systems

principles, use and limitations

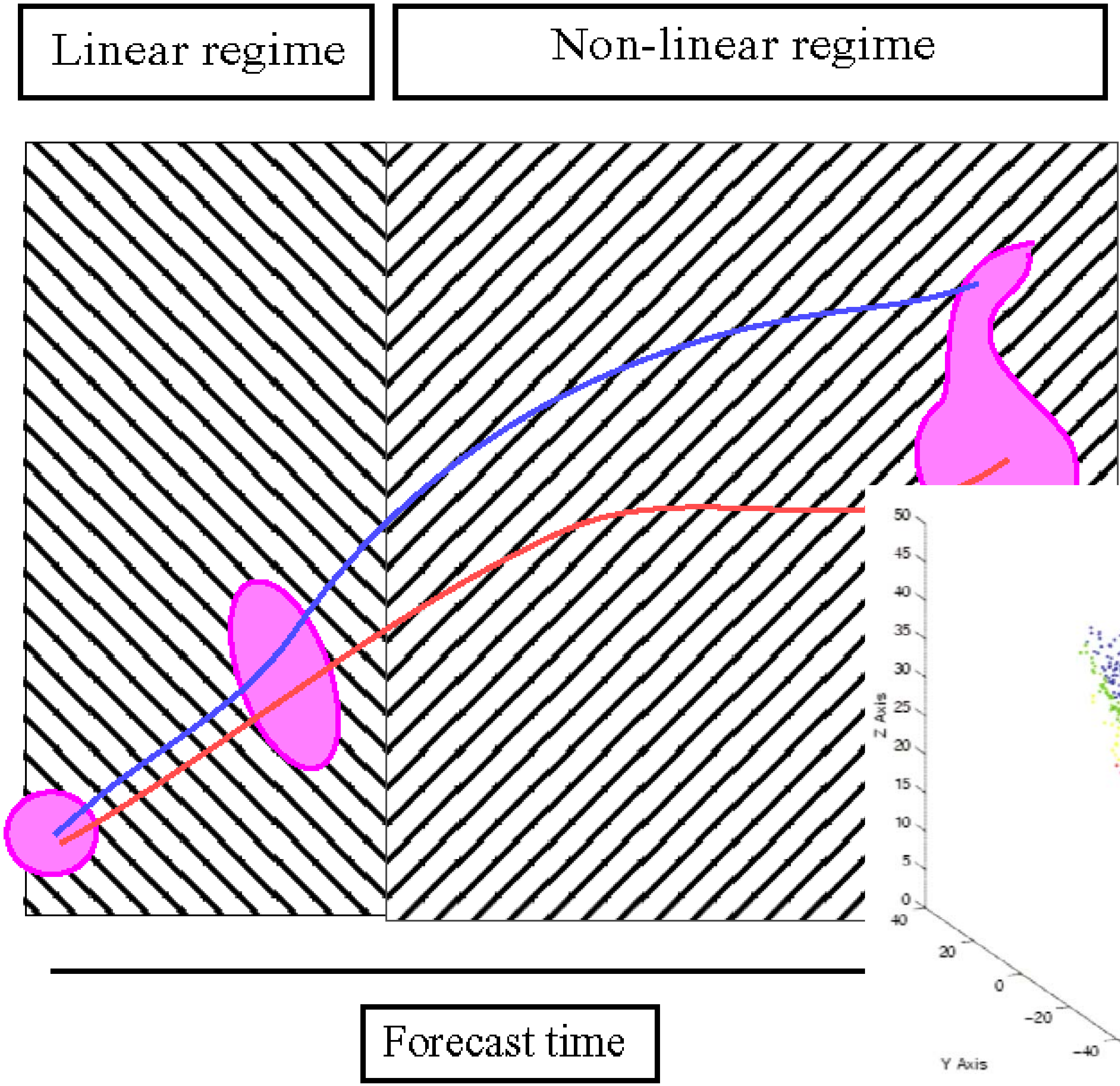
Per Undén

- Reasons for uncertainty
- Different EPS methods
- Comparisons
- Quality of EPS
- Use of EPS
- Limitations and problems

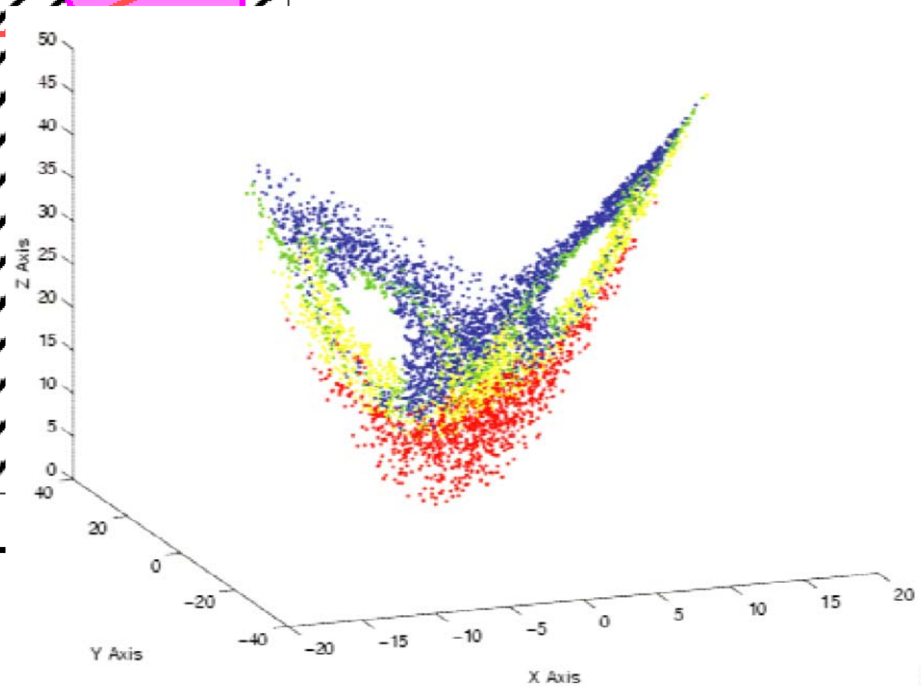
Uncertainty

- Chaotic atmospheric system
  - Scale dependent – small scales grow quickly, medium and large scales retain predictability up to 1-2 weeks

Linear – non-linear – Chaos (Lorenz)

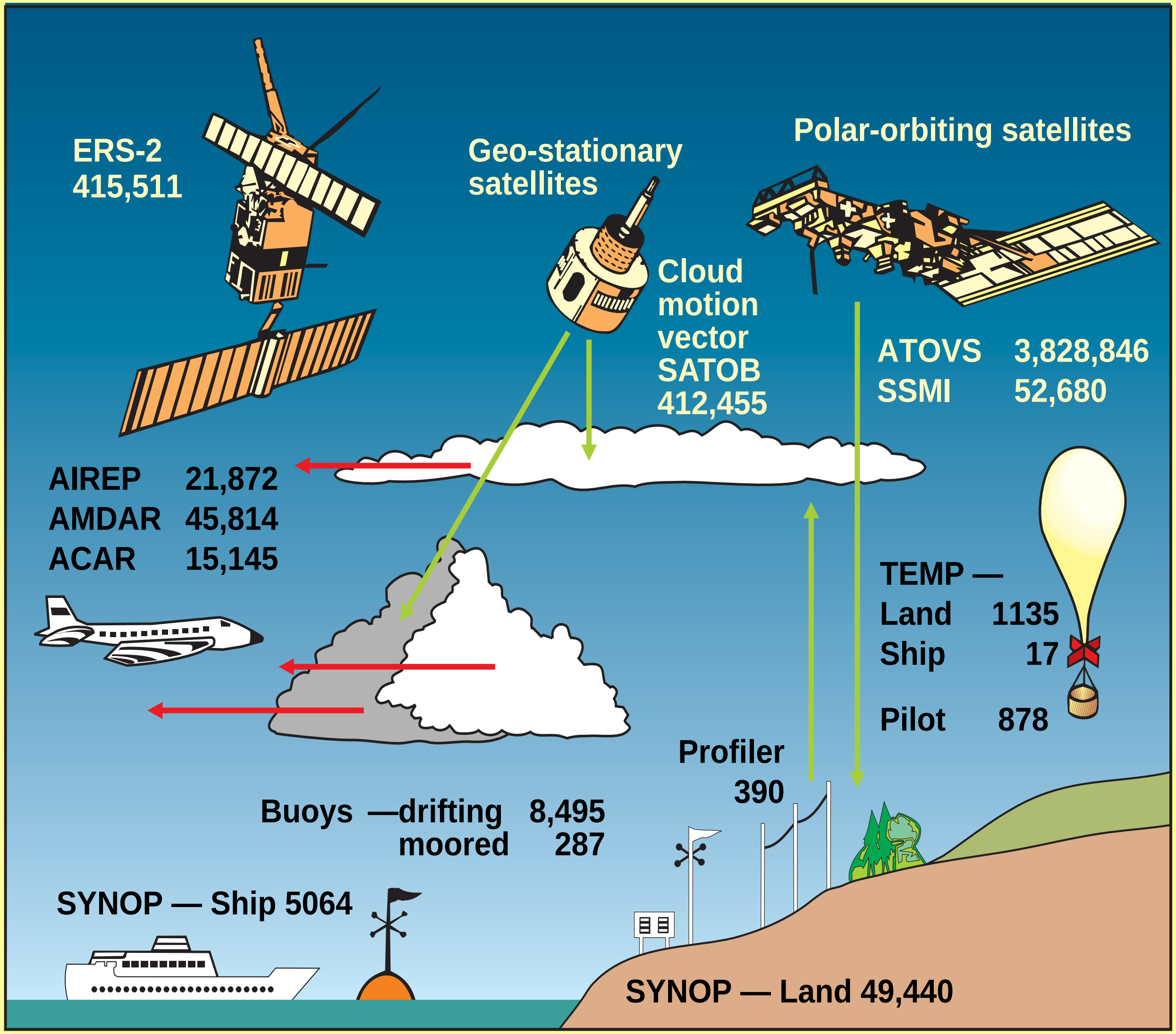


$$\dot{X} = -\sigma X + \sigma Y$$
$$\dot{Y} = -XY + rX - Y$$
$$\dot{Z} = XY - bZ$$



Uncertainty II & III

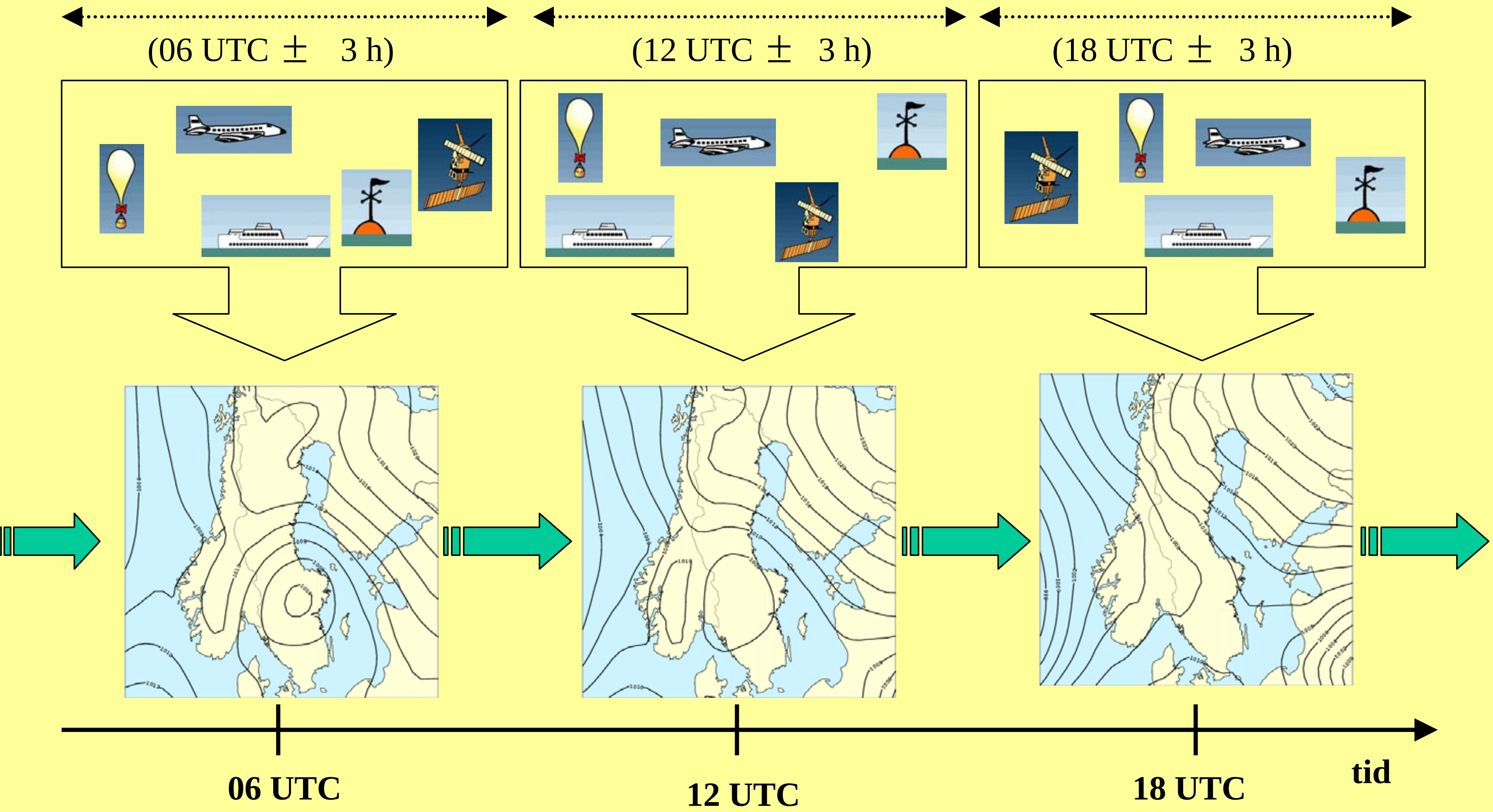
- Dynamical Forecast Models contain approximations
  - Quite accurate fluid dynamics
  - Solved in finite differences / truncated spectrum
  - Unresolved “physics” parameterised
- Initial state is known only within some accuracy
  - Instrument errors or indirect measurements, representativity
  - Observation paucity – limits in areas, levels, variables
  - Data assimilation methods include approximations
  - Data assimilation affected by the above (model and chaos)



Data sources for the ECMWF Meteorological Operational System (EMOS).

The numbers refer to all data items received over a 24 hour period in March 2000.

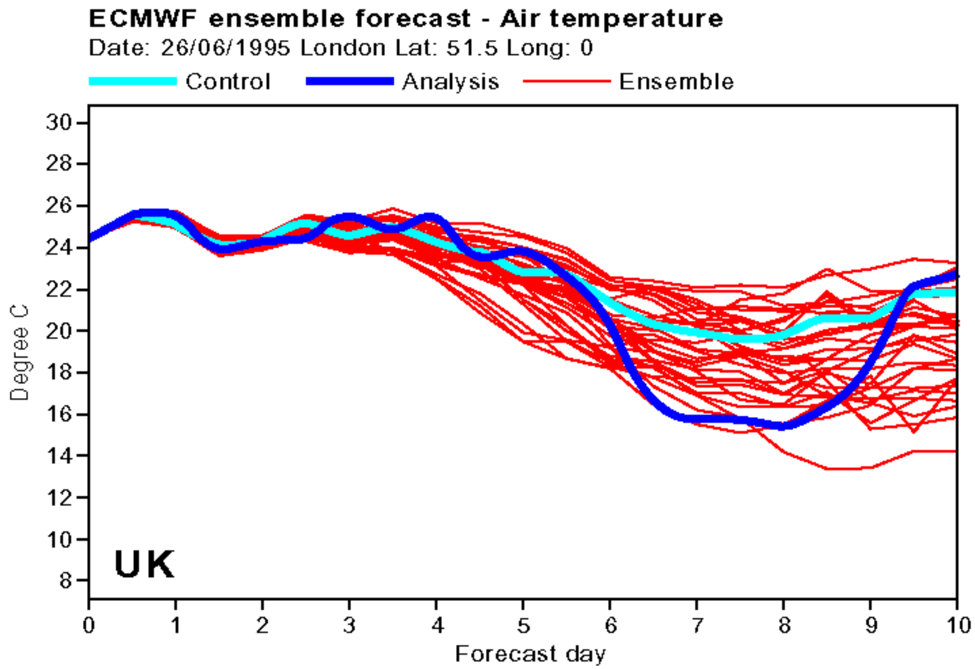
dataassimilation

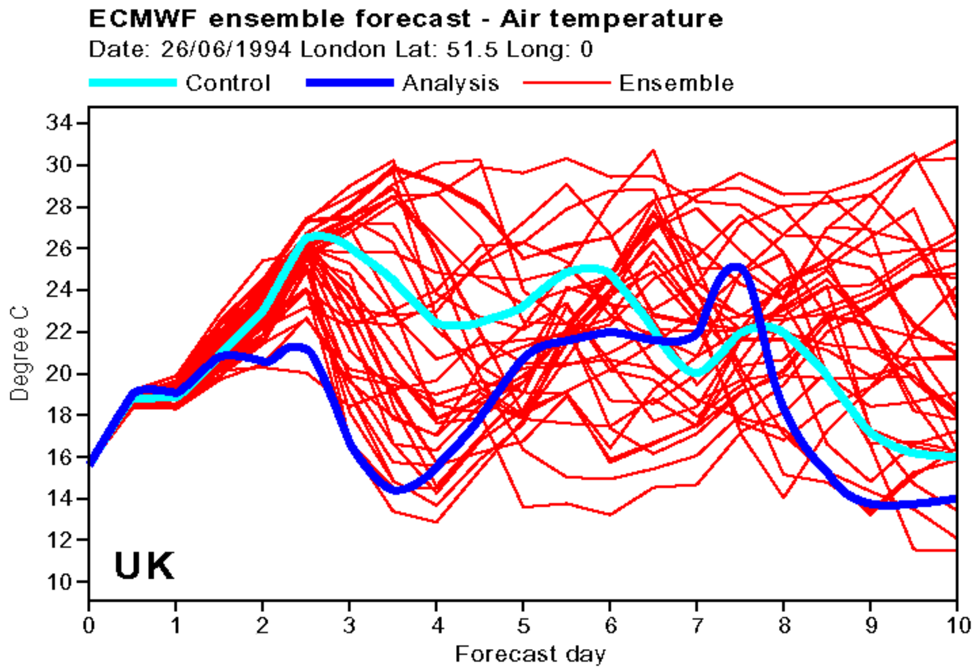


The task of the (meteorological) data assimilation is to extract the largest amount of a ***useful*** information from observations taking into account ***prior*** information about the model state describing the atmosphere

Methods which have been used to merge a background state (the prior) and observations in a way consistent with the estimated accuracy of the each type of information

- The optimum interpolation scheme (OI) (Eliasen 1954, Gandin 1963) (*uses the minimum squared error criteria*);
- The 3D-Variational data assimilation (3D-Var)(Parrish, Derber,1992) (*uses the maximization of the posterior pdf*)
- The 4D-Variational data assimilation (4D-Var) (Le Dimet, Talagrand,1986; Courtier et.al 1994) (*uses the maximization of the posterior pdf*)
- The recent trials with the Kalman filter approach (The implementation of the ensemble Kalman filter looks promising.) (Evenssen, 1994, 2003,Anderson and Anderson, 1999, Houtekamer et.al, 1995, Hamill and Whitaker, 2002).
- Developing of ensemble filtering for nonlinear models based on the particle filter approach(Kim et.al, 2003, Leeuwen 2003)





Spread in two different cases (for London)

How to deal with uncertainty

- Higher order models (error covariances, Liouville)
  - Unrealistic (square cost)
  - Need to know initial uncertainty
  - Theoretical tool in low order models
- Monte Carlo methods
  - Sampling the forecast PDF
  - Estimating skill ?
  - Starting from initial PDF
  - Limited number of realisations
  - Ortogonality for efficiency desired

Ensemble methods

- Poor man's ensemble
  - Available from different models
  - Difficult to interpret and not optimal
  - Difficult to use in production
- Lagged average forecasting
  - Already available forecasts from 6, 12, 24 h .. Earlier
  - At full resolution
  - At no extra cost
  - From same model and easy to use

Both these are **low order** sampling

Ensemble methods II

- Singular vectors
  + Leading eigenvectors for optimal growth of errors
  + Good sampling of different directions
  + Represent errors in the future
  – Expensive to compute
  – Many samples but at low resolution (SV and forecasts)
  – Perturbed around most likely state (=> each worse)
  – Optimised at 48 h – not good for short range
  – Do not show really extreme events – thresholds – index
  – Ex. ECMWF, Reading

I. Theory of singular vectors

.1 Mathematical background :

Non-linear primitive equations : (1) $\frac{dX}{dt}=A(X)$.

Let x be a perturbation of the state vector X :

(2) $\frac{d(X+x)}{dt}=A(X+x)$.

A first order expansion of A in the vicinity of X gives : $A(X+x)\approx A(X)+\mathsf{L}x$

Perturbation forecast equation : (3) $\frac{dx}{dt}=\mathsf{L}x$.

Is integrated in time : (4) $x(t)=\mathsf{M}x_0$,

where M stands for the tangent linear model integrated from t_0 to t.

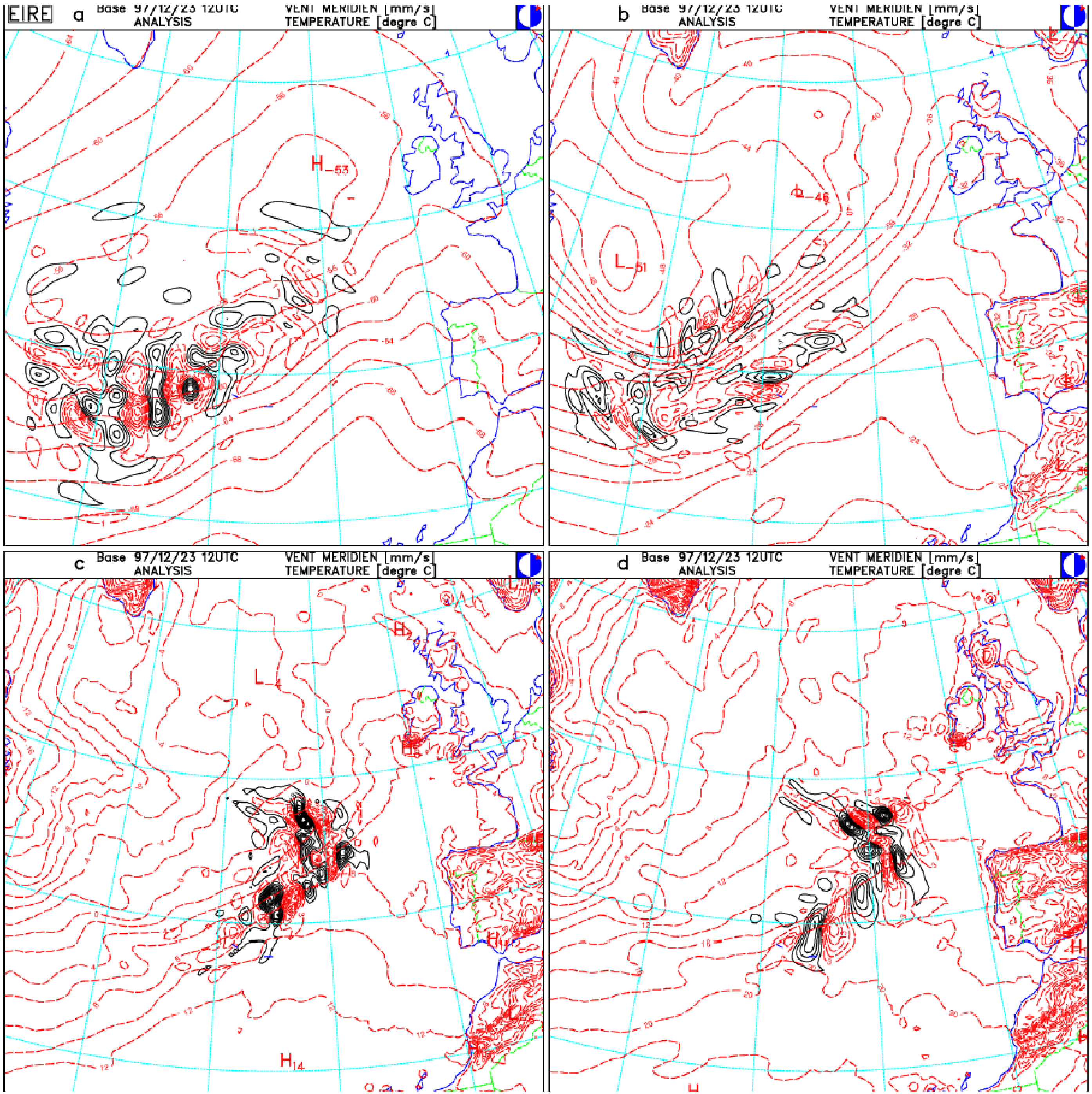
amplification of the perturbation from t_0 to t : $\sqrt{\dfrac{<\mathrm{M}x_0, \mathrm{M}x_0>}{<x_0, x_0>}}$

and introducing the adjoint model : $\sqrt{\dfrac{<\mathrm{M}^*\mathrm{M}x_0, x_0>}{<x_0, x_0>}}$.

PS : different scalar products can be used at the numerator and at the denominator

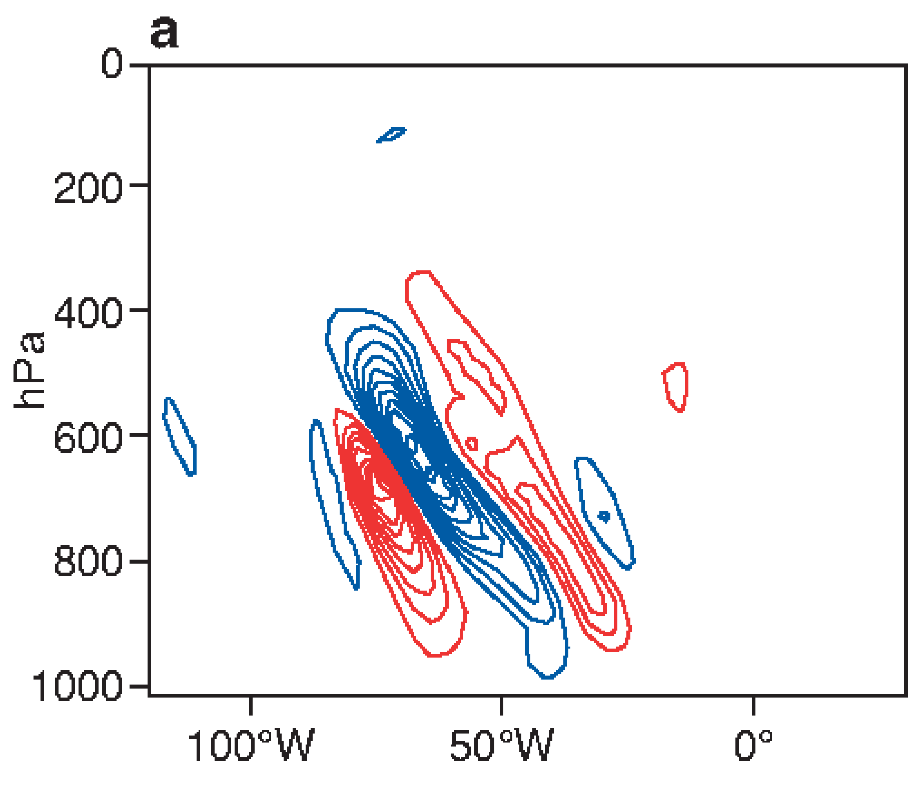
The eigenvectors of M*M are called the singular vectors of M,

And the eigenvalues of M*M are the singular values of M.



Structures for the meridional wind (v') superimposed on the

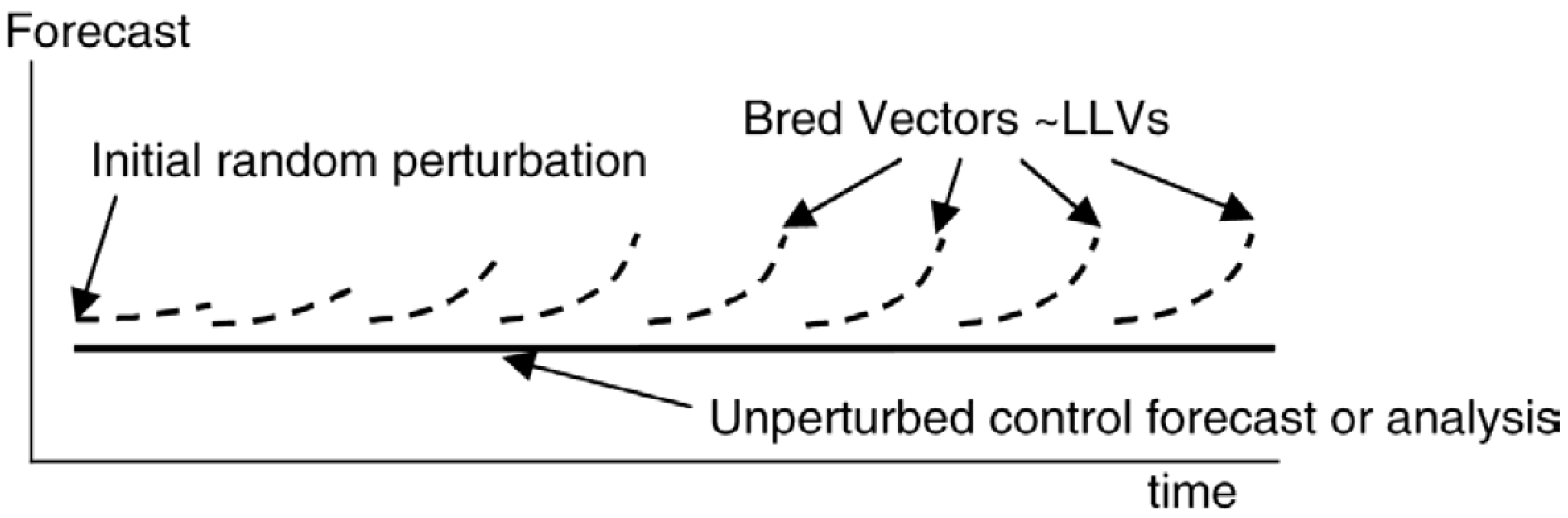
Singular vector EPS



Ensemble methods III

- Breeding methods
  + Perturbed observations and a few parallel assimilations
  + Differences grow in organised way but need scaling
  + Cheap to compute
  – Good for short range
  – Represent errors in the past
  – Not so theoretically founded
  – Ex. NCEP Washington
- Ensemble assimilation
  – Perturbed observations in many parallel assimilations
  – Sampling of covariances in Data Assimilation

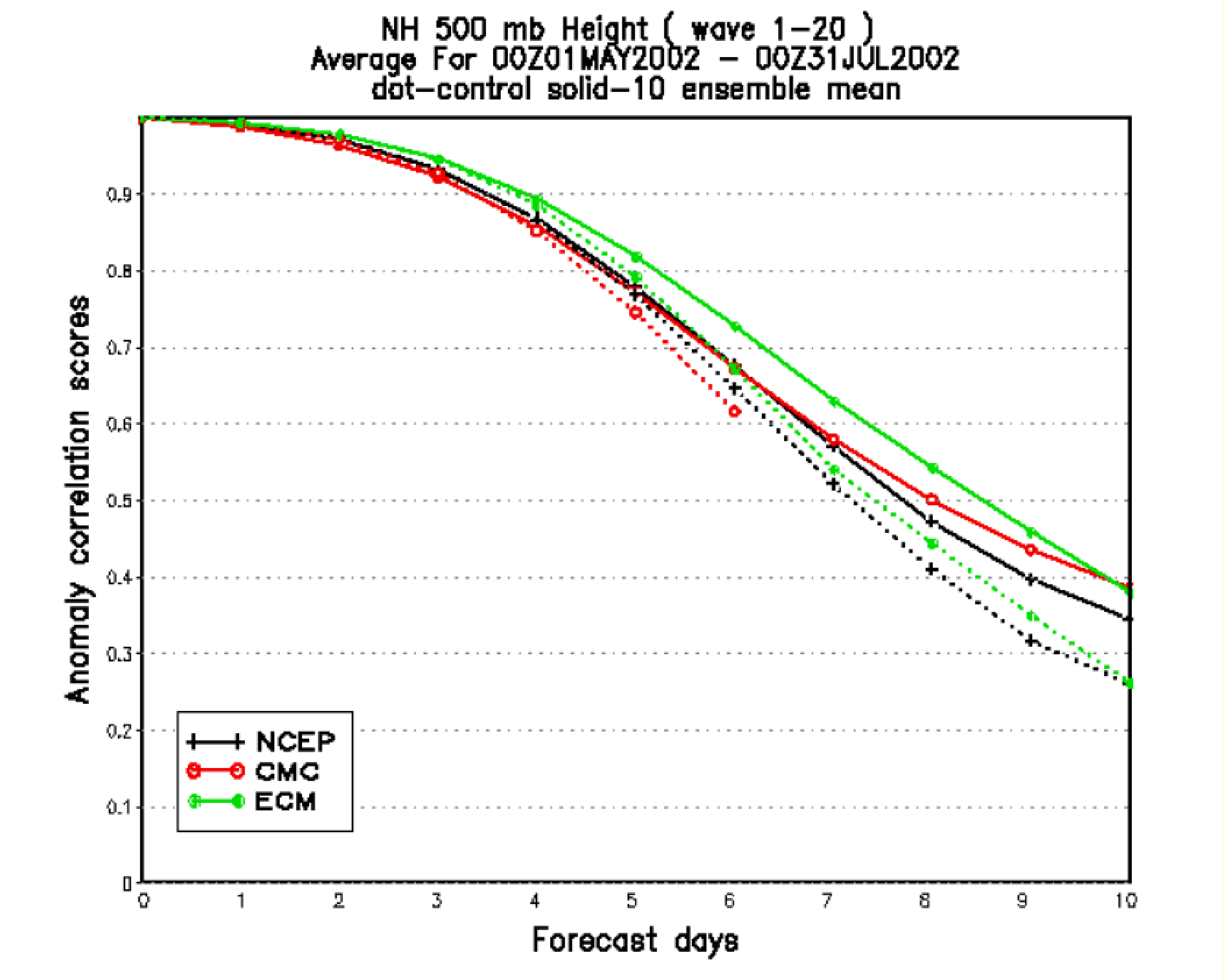
Breeding principle



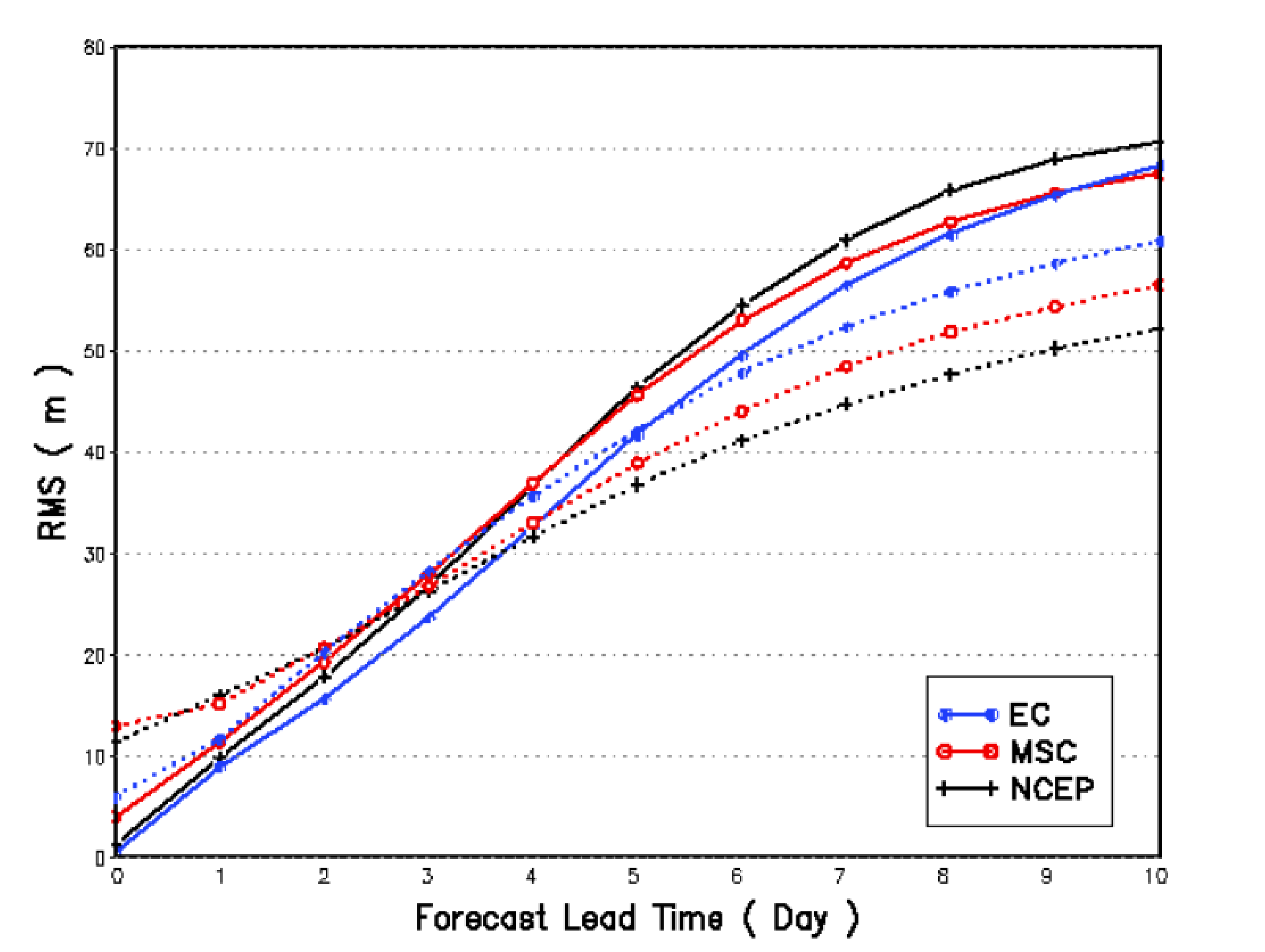
Quality of Ensemble forecast

- Ensemble mean error
- Correct spread – related to skill ?
- How many outliers or not
- Reliability – correct PDFs
- Resolution – many probabilities
- Operating characteristics – Hit rate – false alarm
- Cost/Loss value

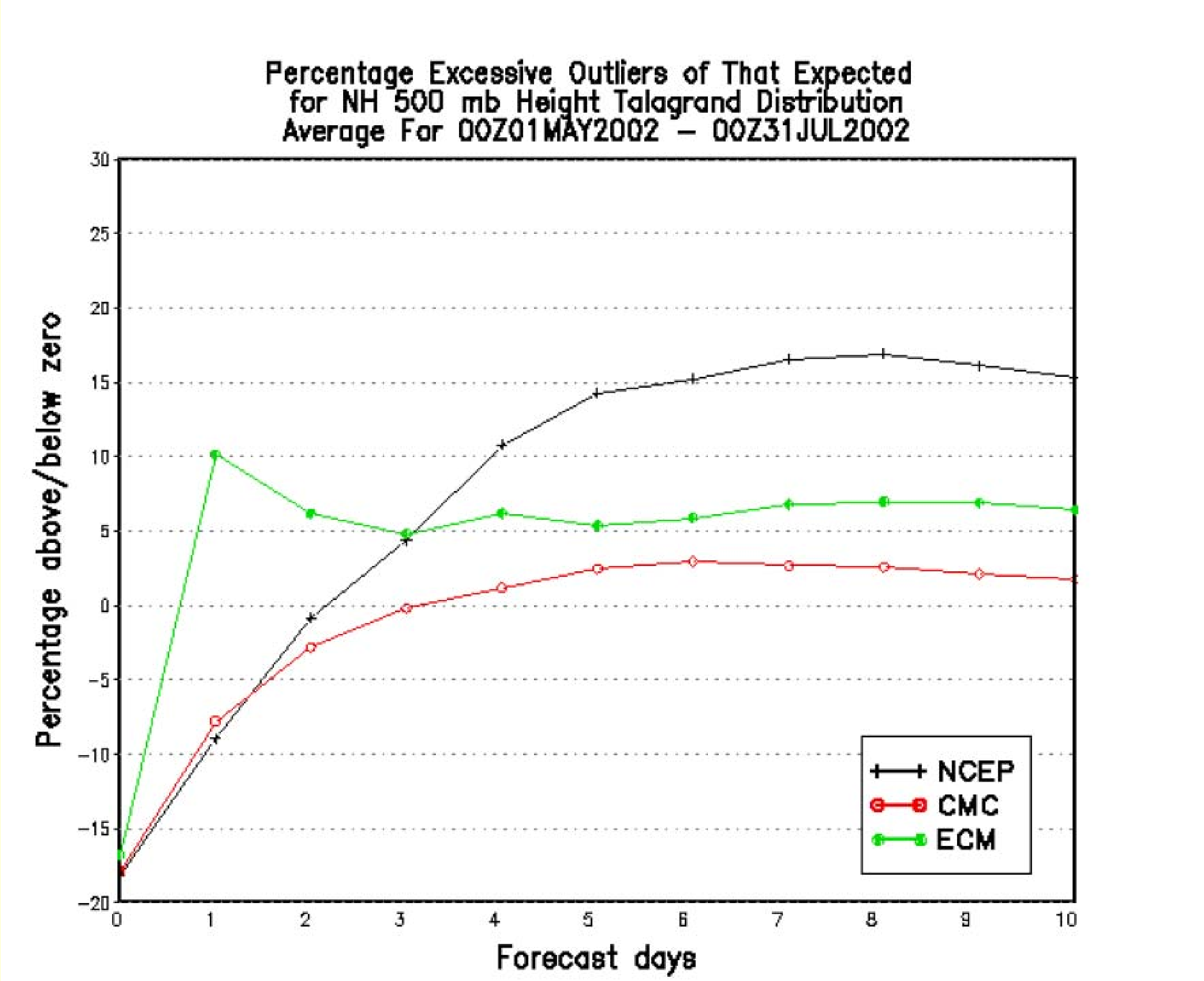
Quality of Ensemble Mean



Spread-skill relationship



Outliers – extremes not represented (or ?)

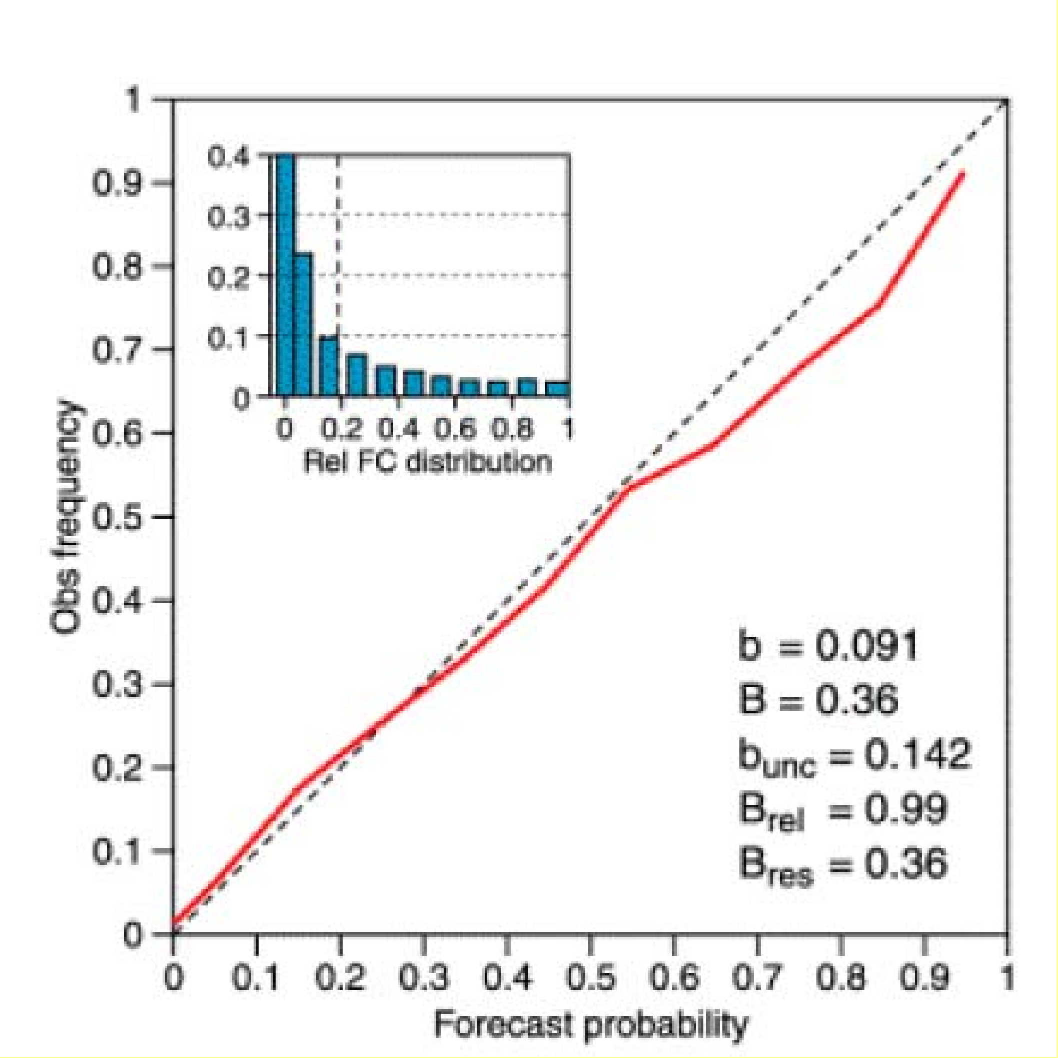


Reliability

realistic probabilities on average

Resolution

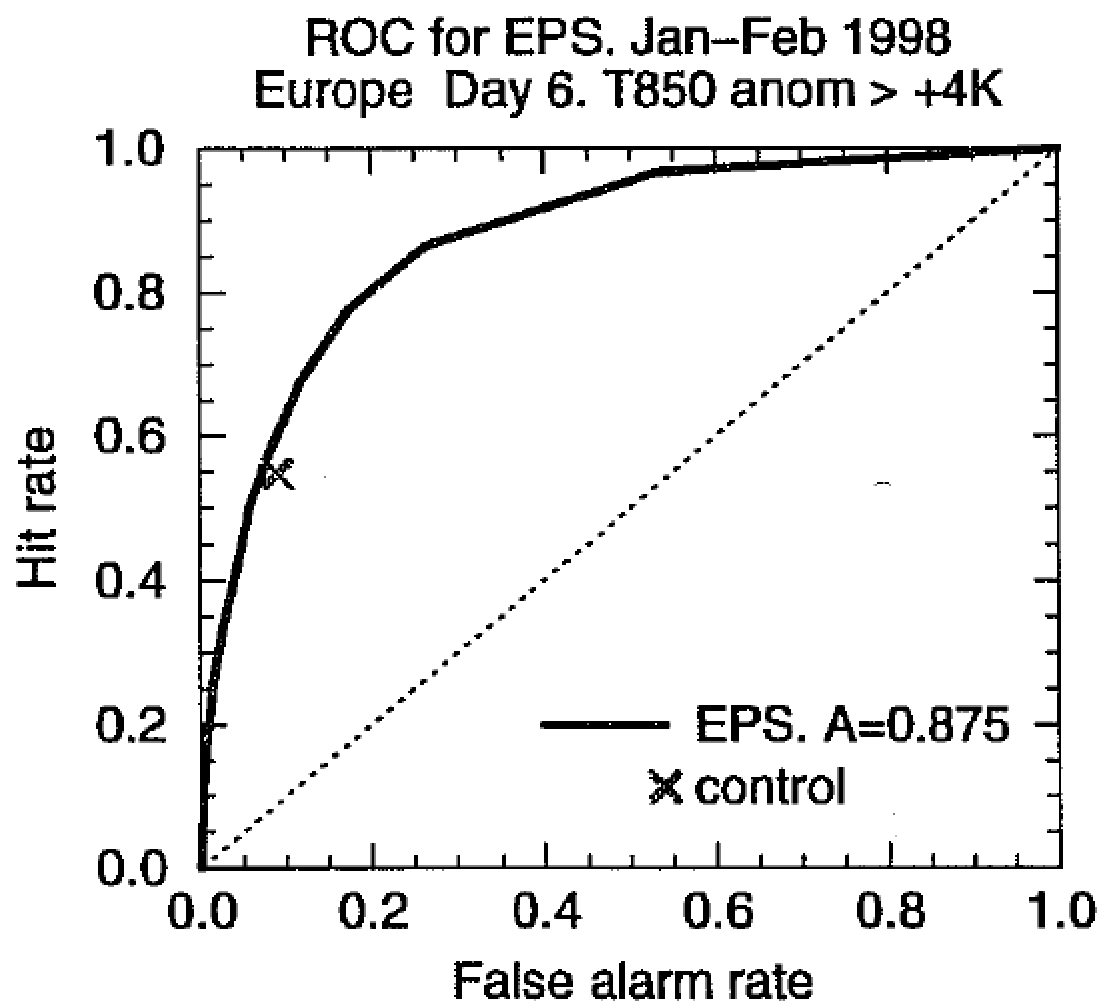
Equitable distribution of probabilities



Relative Operating Characteristics

TABLE 2. CONTINGENCY TABLE FOR FORECAST AND OCCURRENCE OF BINARY EVENT

		observed
forecast	no	
	yes	



Hit Rate = d / (a+b)

False Alarm Rate = c / (a+c)

EPS resolves 50 p:s

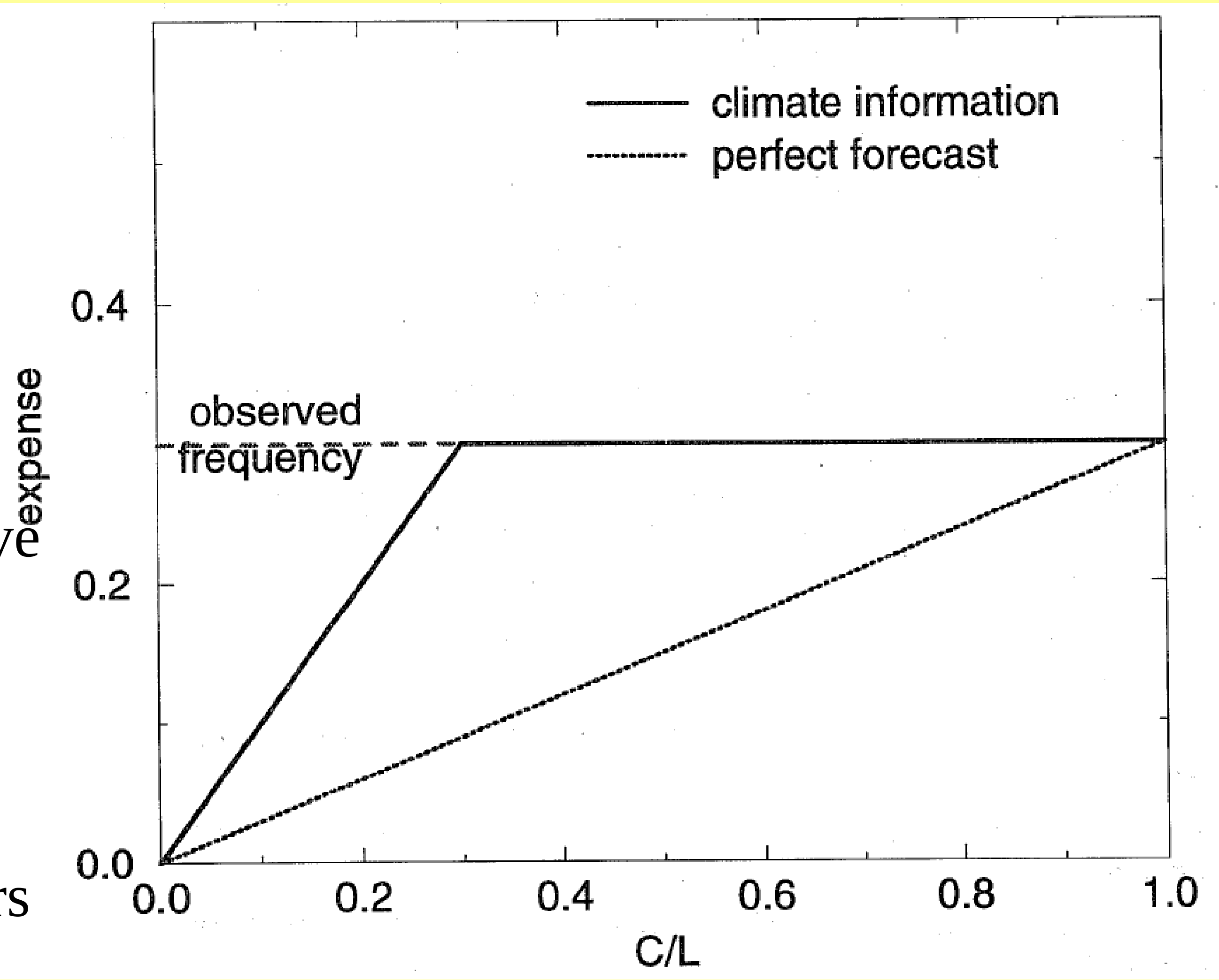
Determ. only 1

Cost/Loss ration and Value

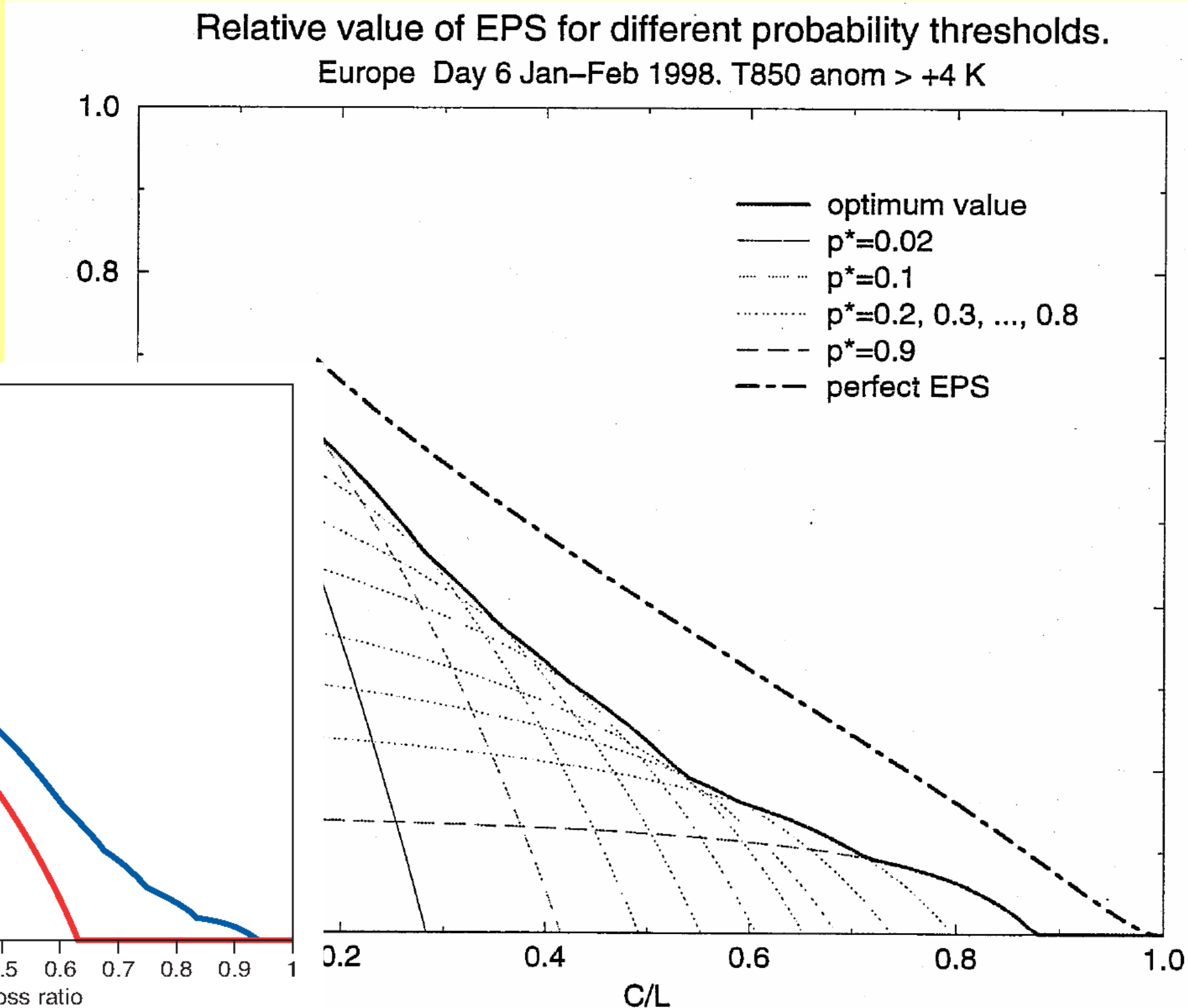
Expense = obs * L
Doing nothing

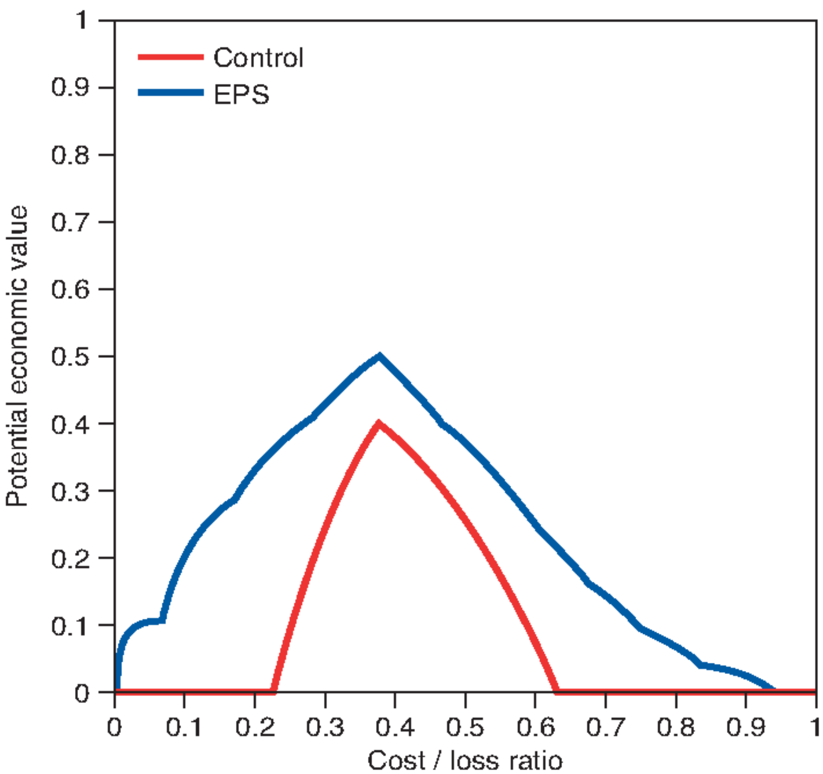
Expense = C/L if act
when C/L < obs
Expense= obs*L
Doing nothing if above
C/L

Expense= obs * C
Perfect forecast
Act when event occurs



Cost/loss - value

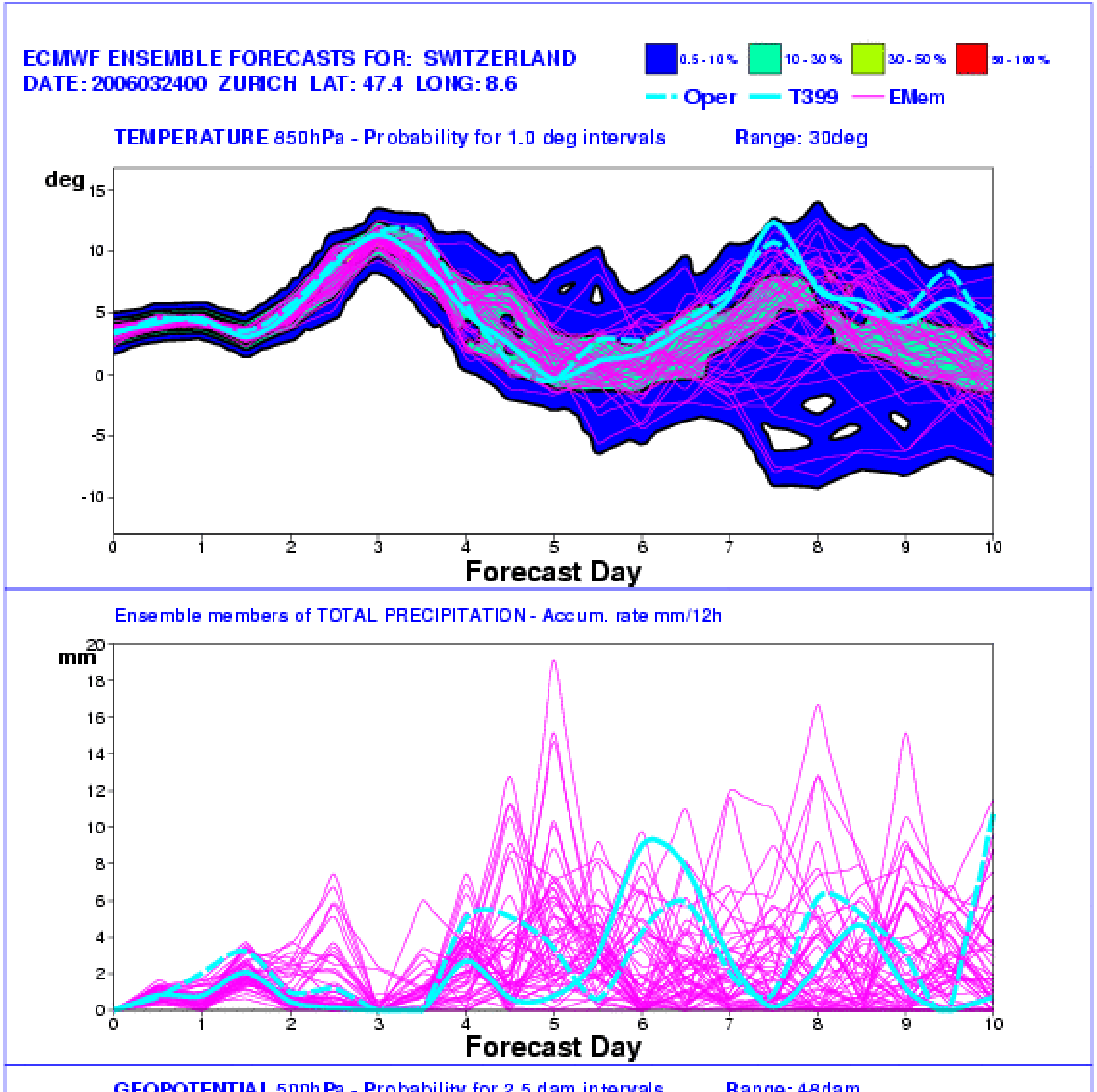


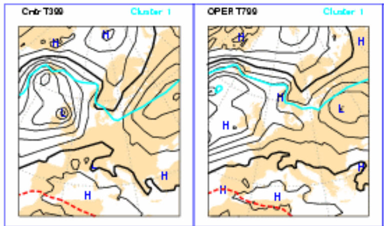


Use of EPS I

- Uncertainty of deterministic forecast
  - Spread - error relationship limited
  - Spread around **erroneous** forecast - not nature
- The likely evolution – ensemble mean
  - Useful product and still essential features
  - No details – but they are unpredictable
  - No extreme values
- Probability distribution
  - Classes limited by number of samples
  - Extreme values outside of the PDFs
- Probabilities of event x > a etc.

Spagetti plots of 51 EPS forecasts

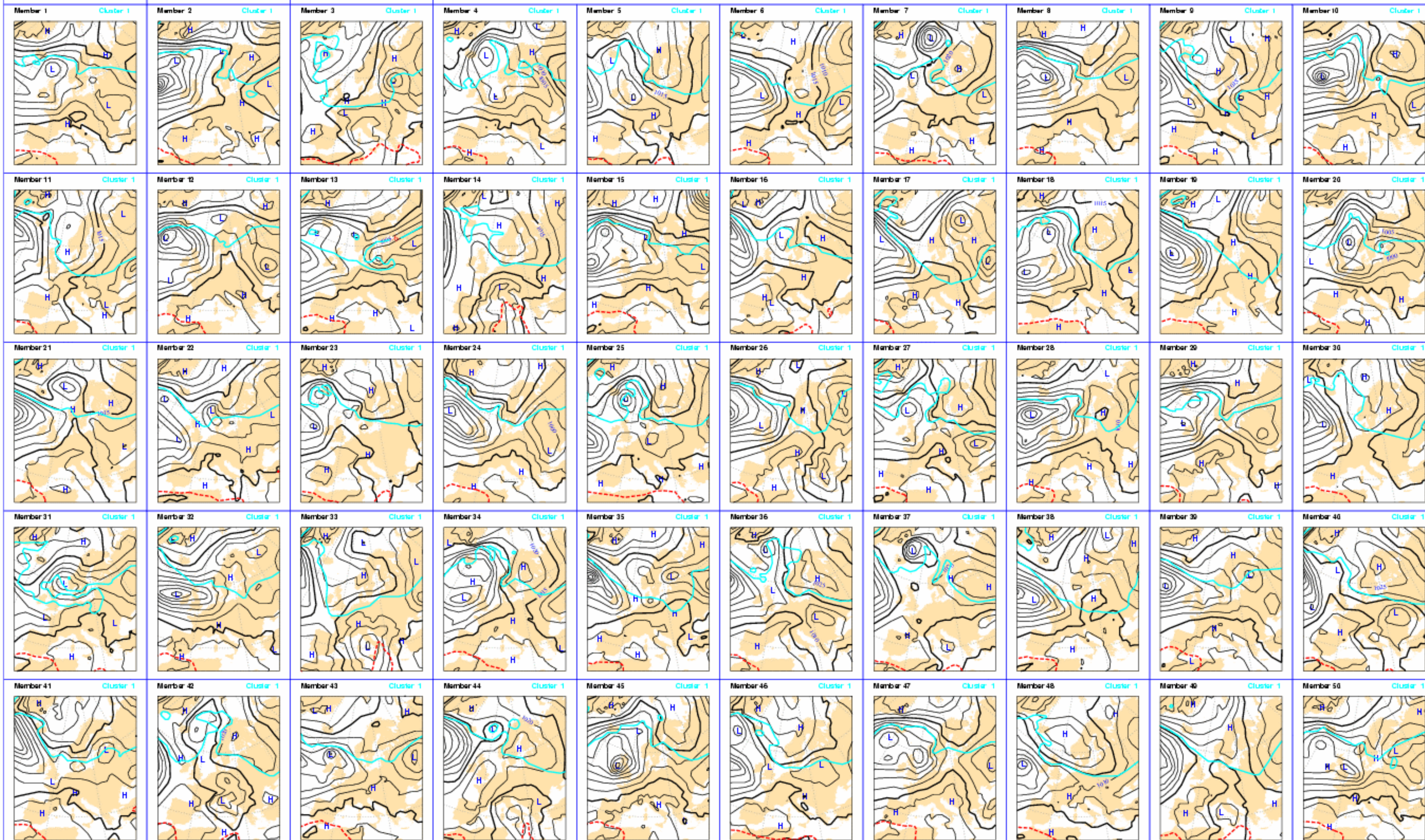




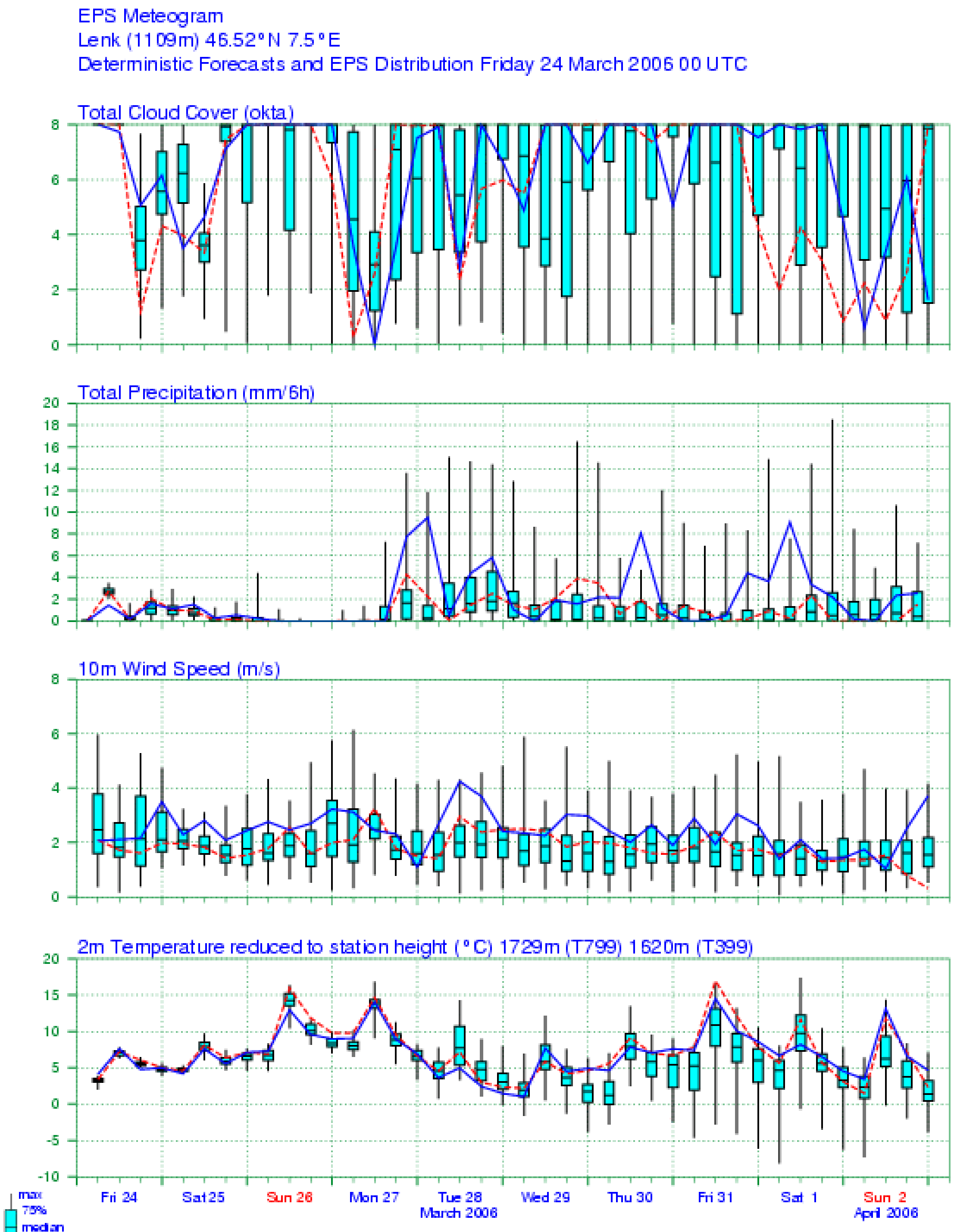
ECMWF ENSEMBLE FORECASTS

Friday 24 March 2006 00UTC ECMWF Forecast t+156 VT: Thursday 30 March 2006 12UTC Surface: mean sea level pressure

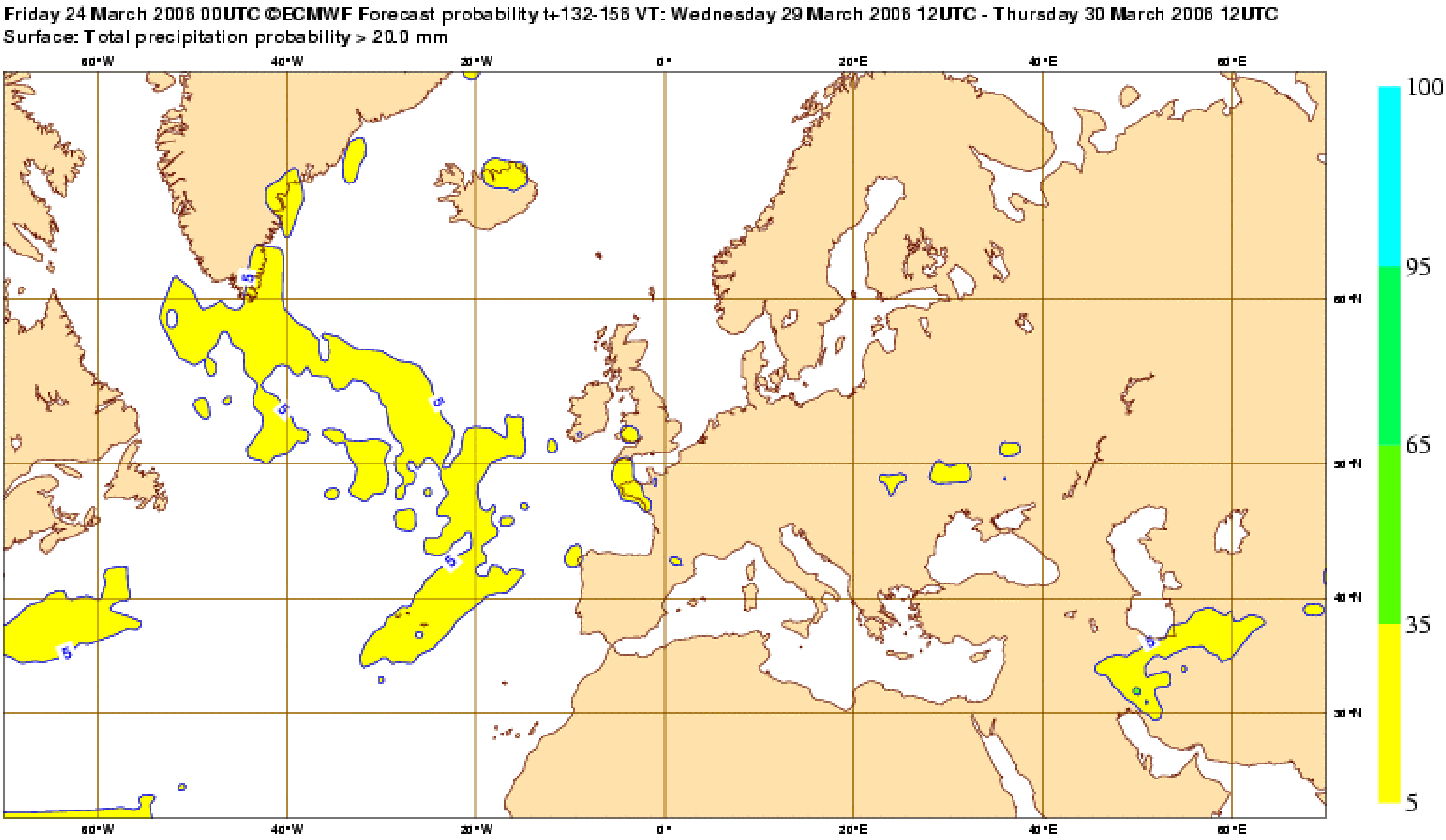
MSLP (contour every 5hPa) and Temperature at 850hPa (only -6 and 16 isolines are plotted)



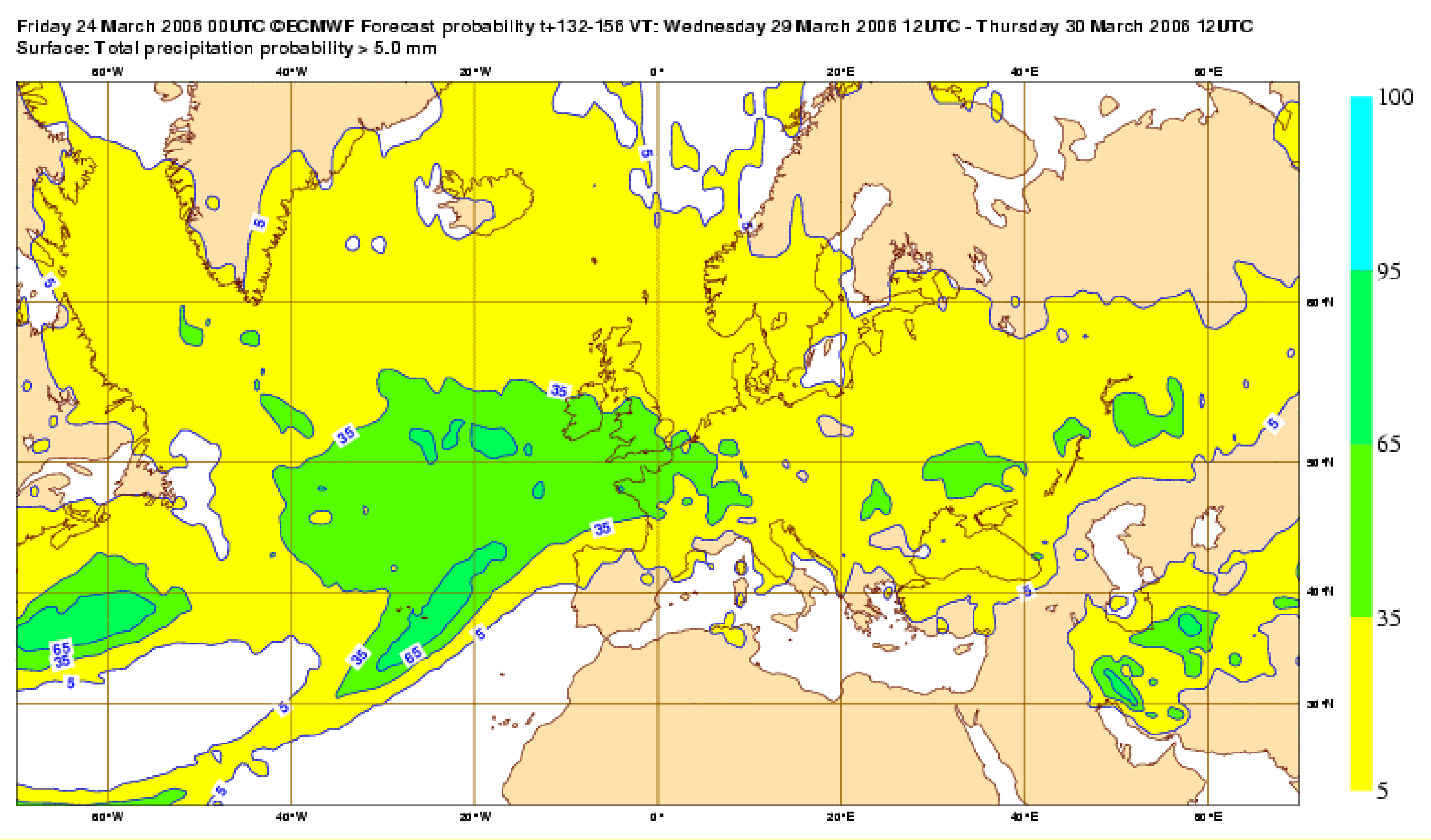
52 forecasts for one location with error bars



Probability >20 mm / 24 h



Probability >5 mm / 24 h



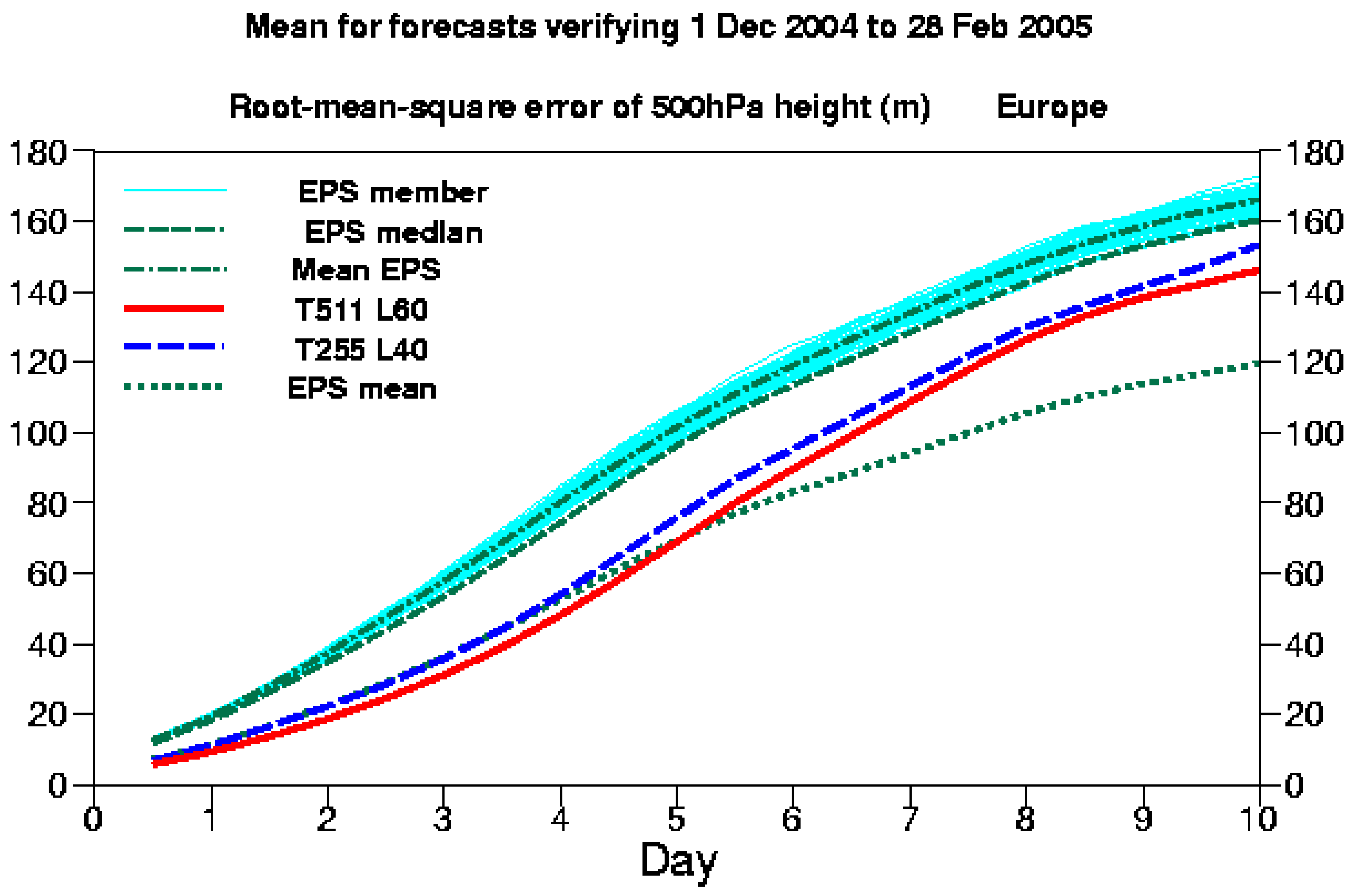
Use of EPS II

- Extreme forecast index
  - To address extreme values not represented
- Clustering techniques
  - Low number of “alternatives”
  - Limited success and debatable
- Decision making cost/loss ratio
  - Advanced used of probabilities
  - Customer oriented
- Boundary conditions for LAMs
  - Note that the LAM results are very dependent on global forcing

EPS problems I

- ECMWF size of perturbations
  - 1.5 day problem worse
  - Necessary for spread
  - Difficulty in interpretation of each member
- Severe weather, hurricane “Gudrun” 8 January 05
  - Only 1-3 members at +72 to +132 hours
  - When deterministic forecast got it +60 hours EPS too

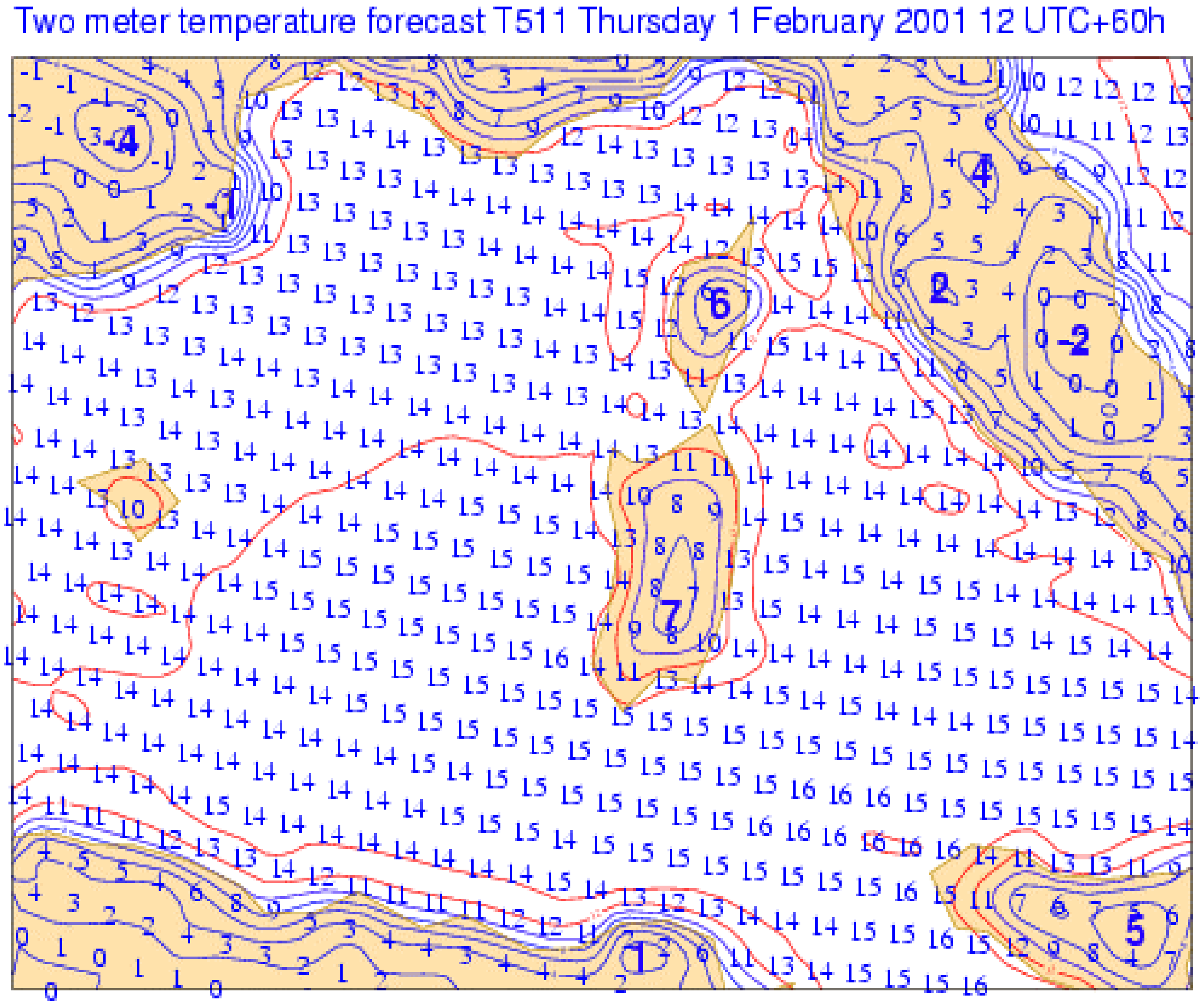
ECMWF EPS members, control (T255) and deterministic (T511). Larger errors in EPS

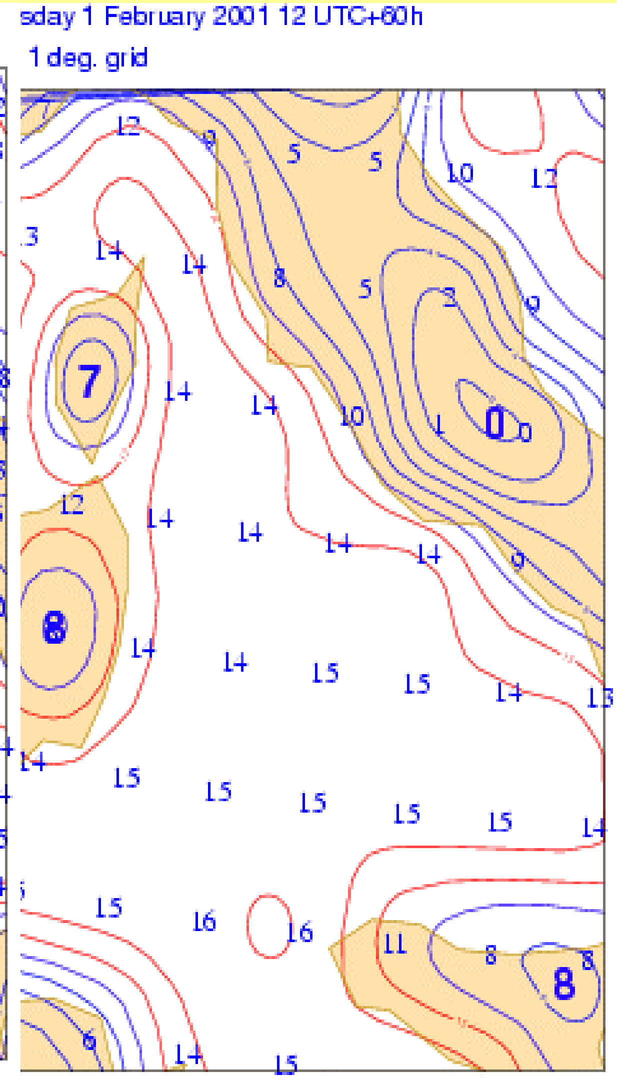


EPS problems II

- Spread – skill relationship limited
- Extreme values often outside the PDF of the EPS
  - Extreme forecast index (threshold)
- Optimisation time in Singular Vector EPS limits short range use

Spatial resolution is lower in EPS
(ex. 45 km <-> 111 km)

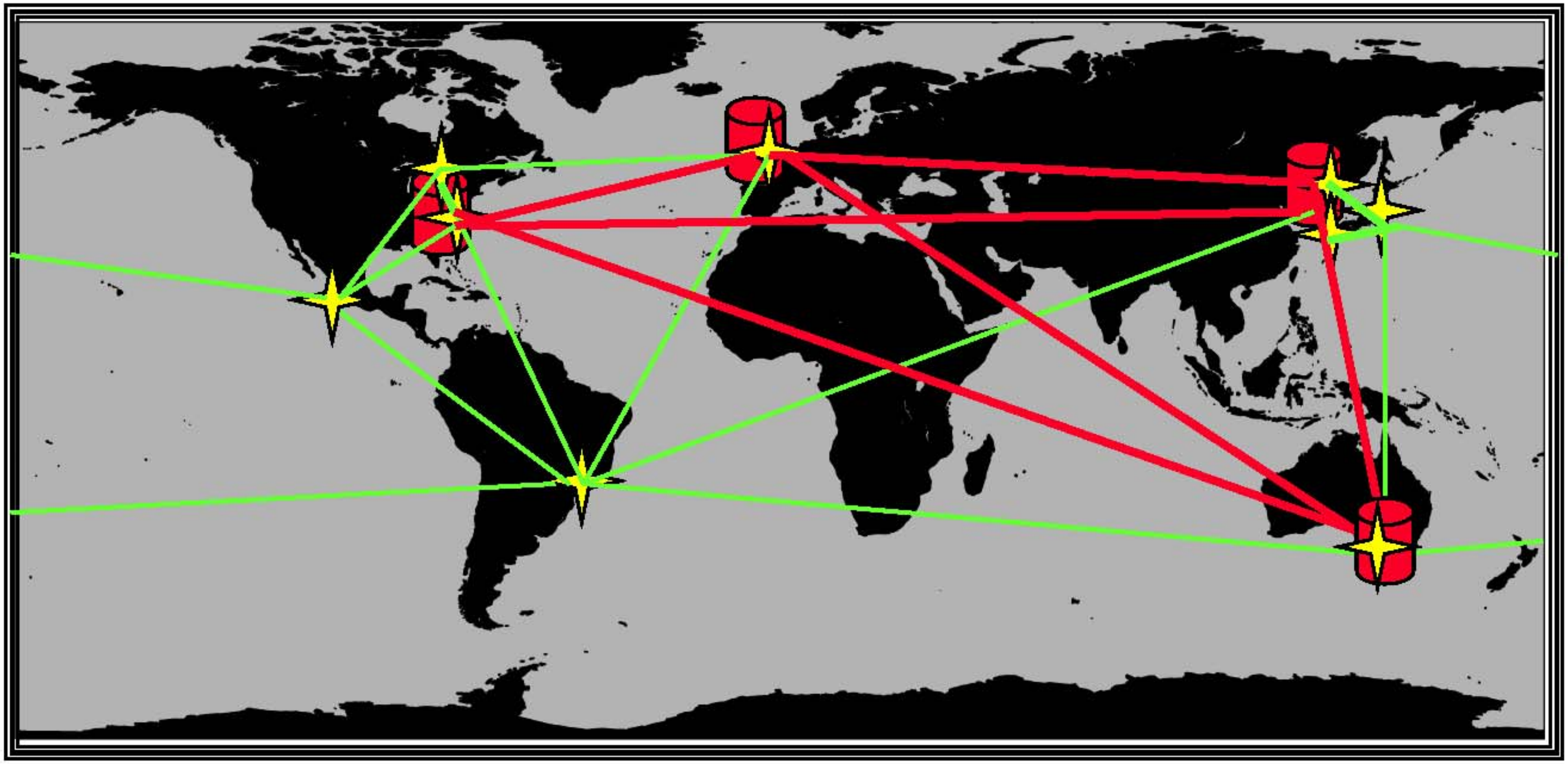




Outlook

TIGGE could lead to a MUMMA-GEPS

TIGGE could lead to a **Multi-Model, Multi-Analysis Global Ensemble Prediction System (MUMMA-GEPS)**, with N production centers (yellow stars) and few data-hubs (red) connected by high-speed, high-capacity communication lines.



Flood applications can help to value a MUMMA-GEPS

The value of the MUMMA-GEPS could be assessed by linking TIGGE with the European Flood Alert System (EFAS) and the Hydrological Ensemble Prediction Experiment (HEPEX).

Observations	Data assimilation	Global M~al modelling	Local M~al modelling	Local H~al modelling	Quality control
•Weather data •Hydrological data •Observation perturbation methods • …	•3/4D-Var Data Assimilation Systems •Ensemble Data Assimilation • …	•Global perturbation methods •Model design •Model resolution •Ensemble Size	•Nested Limited Area Models •Local perturbation methods • …	•Post-processing methods •Combination methods •Rainfall, runoff models •…	•Verification methods •Users' value •Verification data • …