Using Bayesian Model Averaging to calibrate short-range forecasts from a multi-model ensemble

DANIEL SANTOS-MUÑOZ , ALFONS CALLADO, JOSE A. GARCIA-MOYA, CARLOS SANTOS AND JUAN SIMARRO.

Predictability Group

Spanish Meteorological Institute (INM). 28040 Madrid. Spain

BMA INTRO

- BMA is a statistical method for postprocessing ensembles based on standard method for combining predictive distributions from different sources.

- The BMA predictive PDF of any quantity of interest is a weighted average of PDFs centered on the individual bias-corrected forecasts, where the weights are equal to posterior probabilities of the models generating the forecasts and reflect the models' relative contributions to predictive skill over the training period.

Raftery et al, 2005

BMA FUNDAMENTALS

$$p(parameter) = \sum_{k=1}^{K} p(parameter \mid f_k)\, p(f_k \mid parameter^{Training})$$

$p(parameter \mid f_k)$ is the forecast PDF based on f_k alone.

$p(f_k \mid parameter^{Training})$ is the posterior probability of f_k being correct given the training data => Reflects how well f_k fits the training data.

$$\sum_{k=1}^{K} p(f_k \mid parameter^{Training}) = 1 \Rightarrow w_k$$

can be viewed as weights

BMA FUNDAMENTALS

- If the conditional PDF of the *parameter* can be approximated to a normal distribution centred at a linear function of the forecast

$$\boxed{parameter \mid f_k \approx N(a_k + b_k f_k, \sigma_k^2)}$$

- We need to estimate

$$\boxed{a_k \quad b_k \quad w_k \quad \sigma_k^2}$$

$\boxed{a_k\, b_k}$ => Estimate by means simple linear regression of $parameter_{s\,t}$ on f_{kst} for the training data => a_k b_k simple BIAS correction. (Glahn and Lowry 1972; Carter et al. 1989).

BMA FUNDAMENTALS

$$\boxed{w_k \quad \sigma_k^2}$$

=> Estimate by maximum likelihood (Fisher,1922) of the training data.

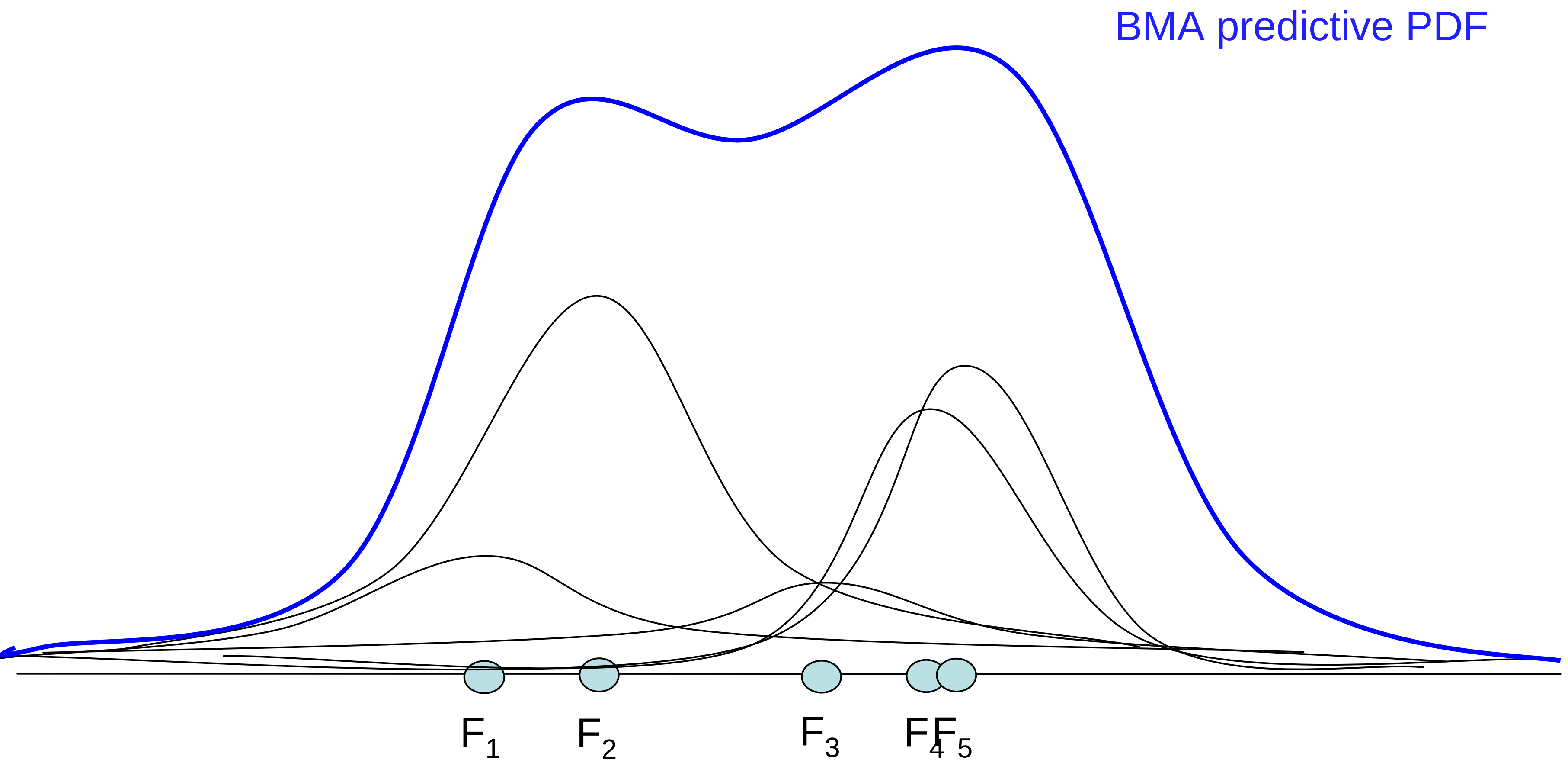
- Assuming independence of forecast errors in space and time, the log-likelihood function is

$$\boxed{l(w_1....w_k,\sigma_k^2) = \sum_{s,t} \log\left(\sum_{k=1}^{K} w_k g_k(parameter_{st} \mid f_{kst})\right)}$$

where the summation is over *s* and *t.*

- Maximization of log-likelihood function by means expectation-maximization (EM) algorithm. (Dempster et al. 1977; McLachlan and Krishnan1997).
- The σ estimate is refined to optimize the continuous ranked probability score (CRPS), searching numerically over a range of values of centered at the maximum likelihood estimate.

BMA FUNDAMENTALS



BMA FUNDAMENTALS

- The BMA predictive variance can be written as (Raftery,1993)

$$\boxed{Var(parameter_{st} \mid f_{1st} \ldots f_{kst}) = \sum_{k=1}^{K} w_k \left((a_k + b_k f_{kst}) - \sum_{i=1}^{K} w_i (a_i + b_i f_{ist}) \right)^2 + \sigma^2}$$

- Predictive Variance =

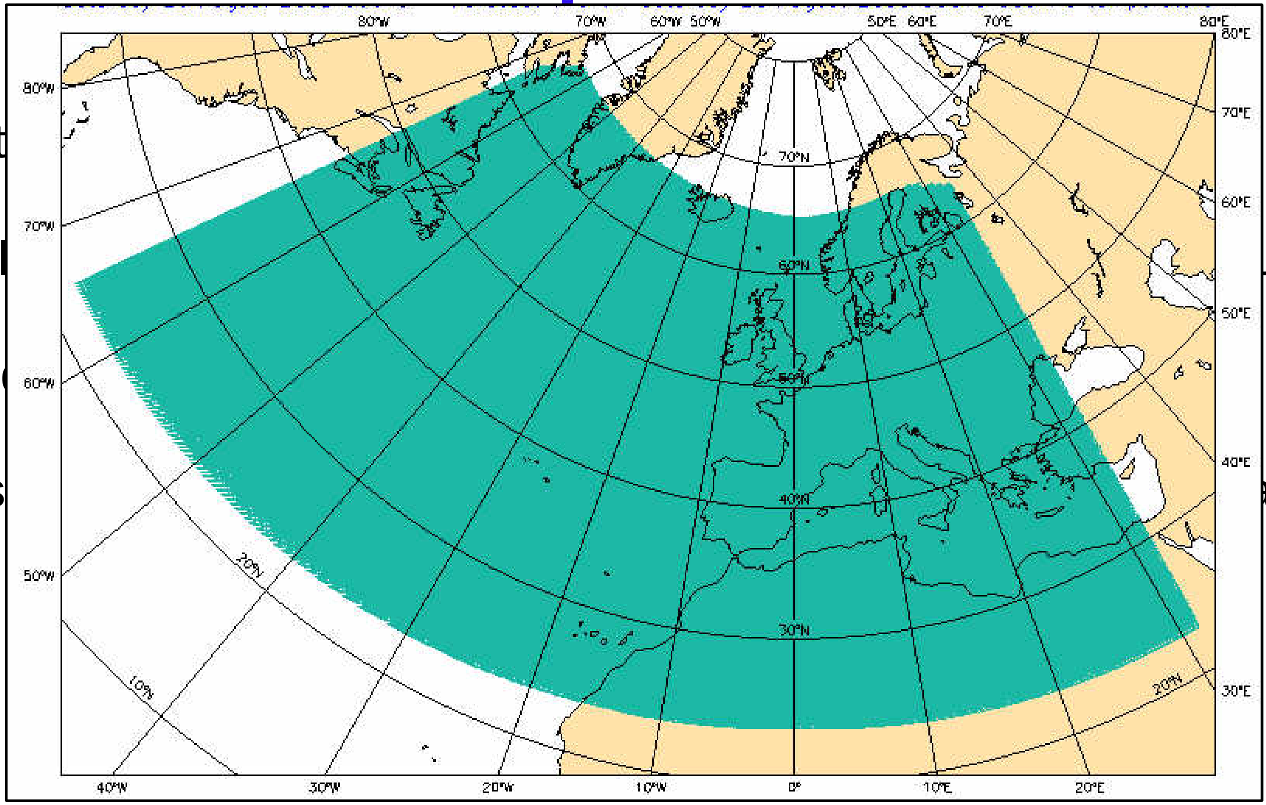
  Between-Forecast Variance + Within-Forecast Variance

  Between-Forecast Variance is the ensemble spread

  $\sigma^2 = \sum_{k=1}^{K} w_k \sigma_k^2$ measures the expected uncertainty conditional on one of forecast being the "best" locally in time=> adds more spread if the ensemble is underdispersive.

SREPS DESCRIPTION

- SREPS
  integrat
- 5 model ... 2 hours
- The mo
- Full des ... -Moya et al.) and



BMA EXPERIMENTS

- BMA calibration:

  500 hPa Temperature (T500)
  500 hPa Geopotencial (Z500)
  10m Wind speed (S10m)

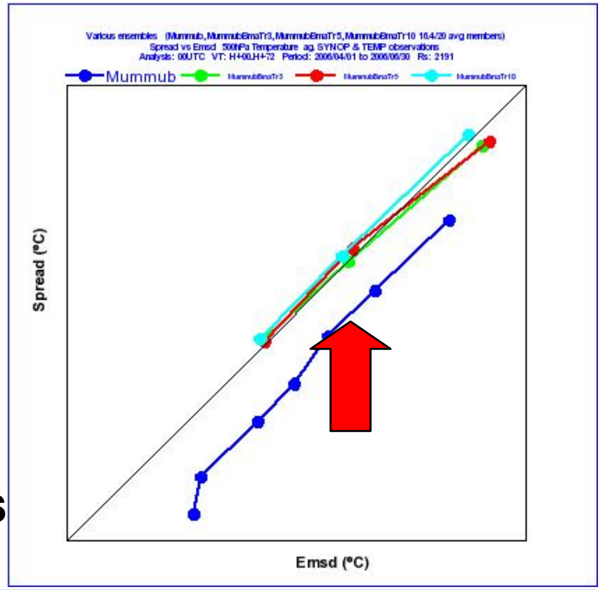
  - 3, 5 and 10 days of training period
  - 3 months of calibration (April, May and June of 2006) for Z500 and T500 and 1 month for S10m (April 2006)
  - 24, 48 and 72 hours forecast for Z500 and T500 and 24 and 72 hours forecast for S10m

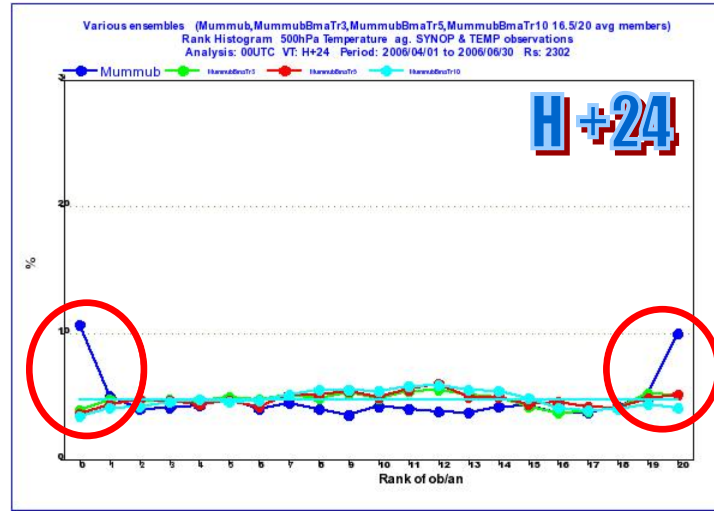
- BMA calibration using TEMP and SYNOP observations over whole area

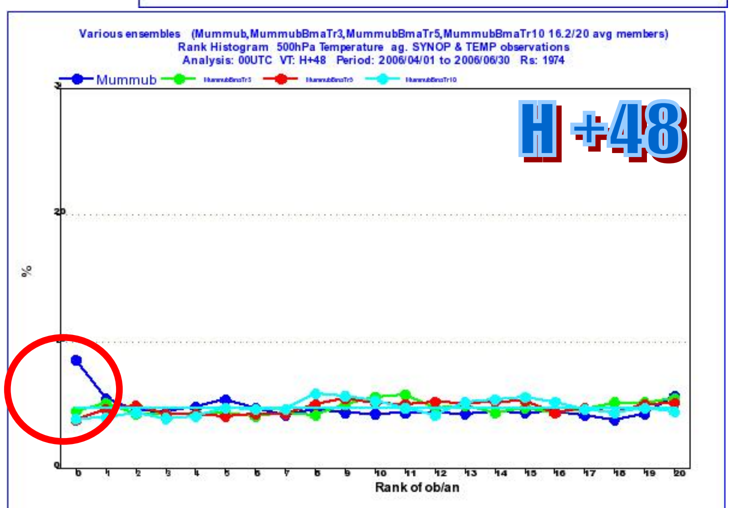
RESULTS

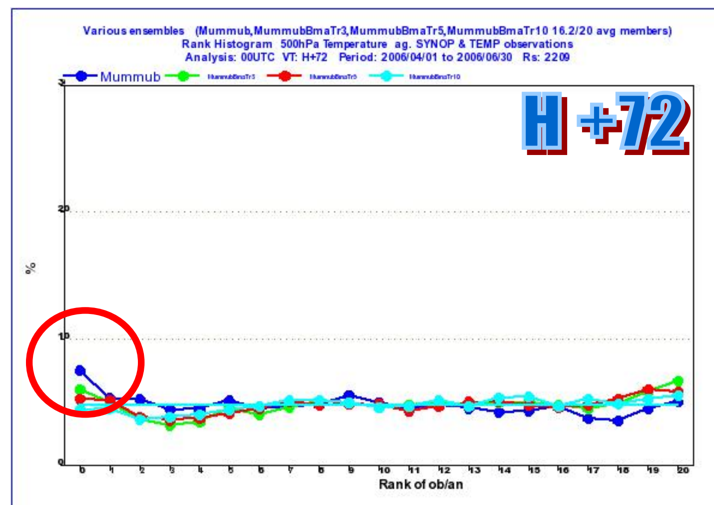
T 500

- MULTIMODEL
- BMA 3 T. DAYS
- BMA 5 T. DAYS
- BMA 10 T. DAYS





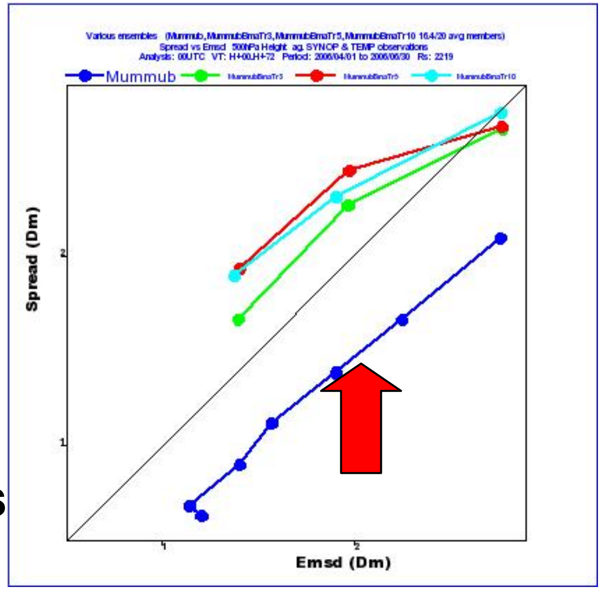


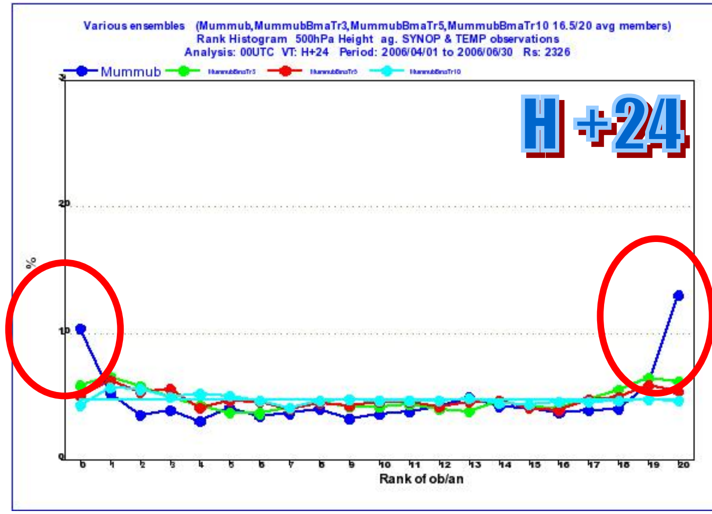


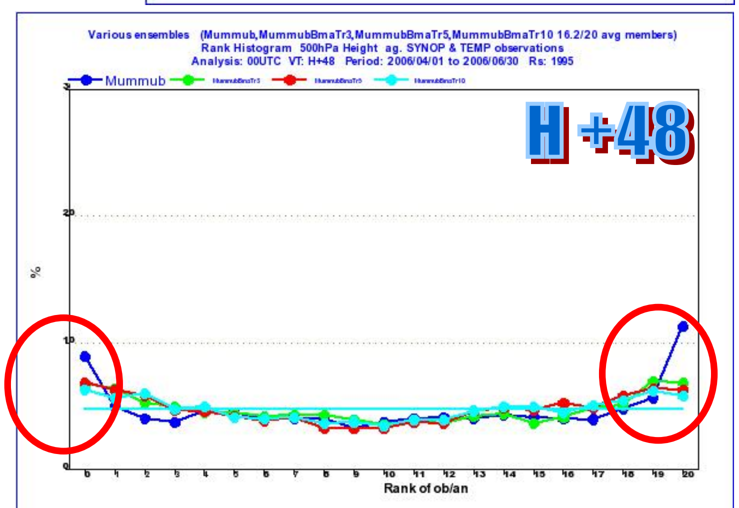
RESULTS

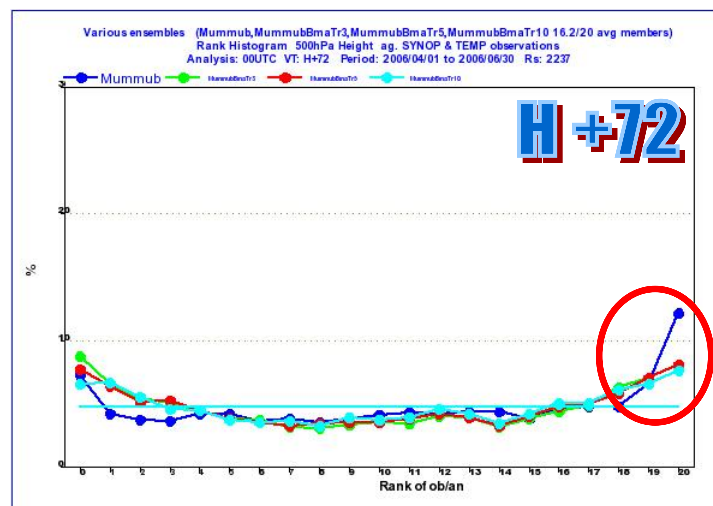
Z500

- MULTIMODEL
- BMA 3 T. DAYS
- BMA 5 T. DAYS
- BMA 10 T. DAYS





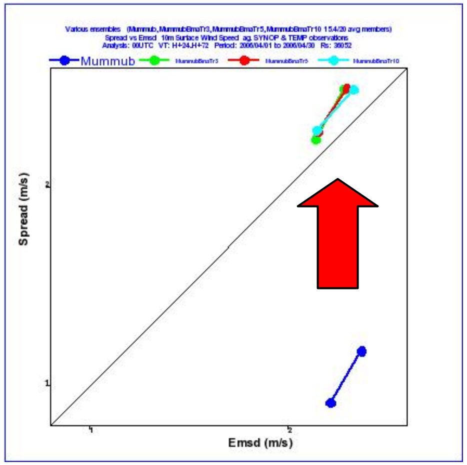


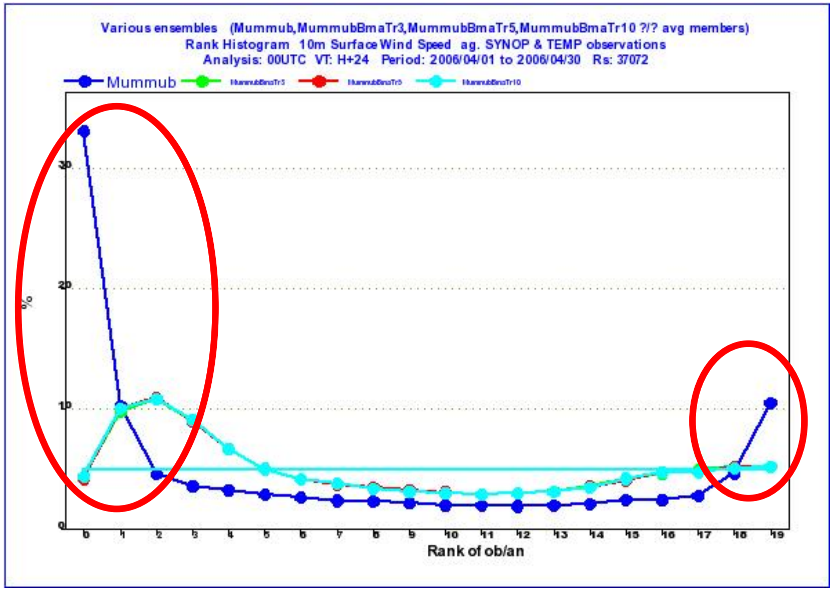


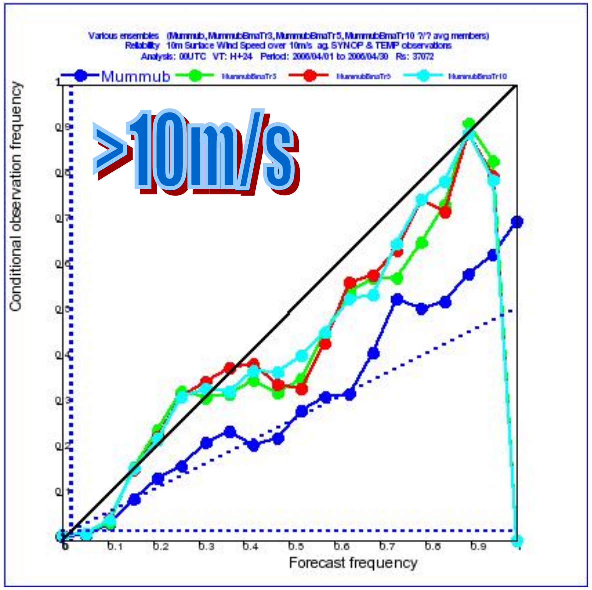
RESULTS

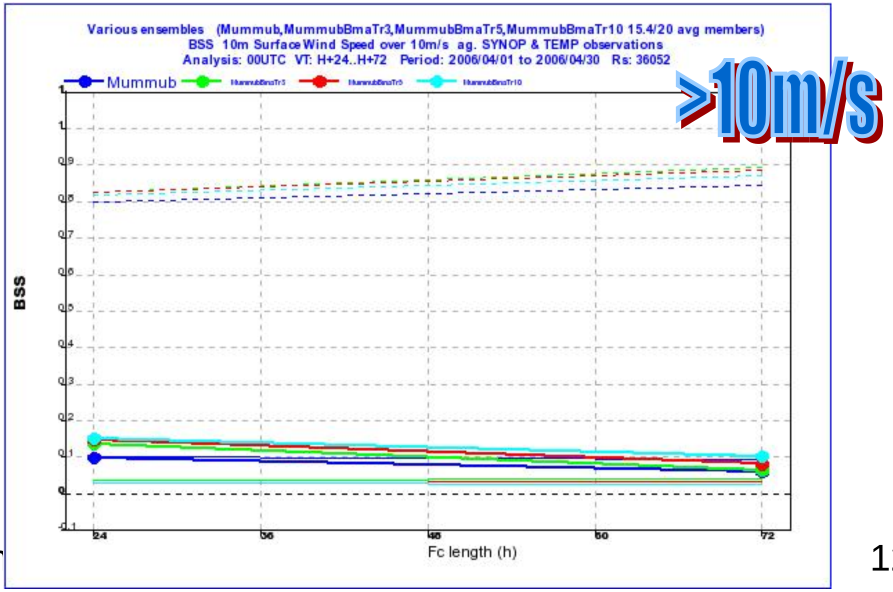
S10m

- MULTIMODEL
- BMA 3 T. DAYS
- BMA 5 T. DAYS
- BMA 10 T. DAYS



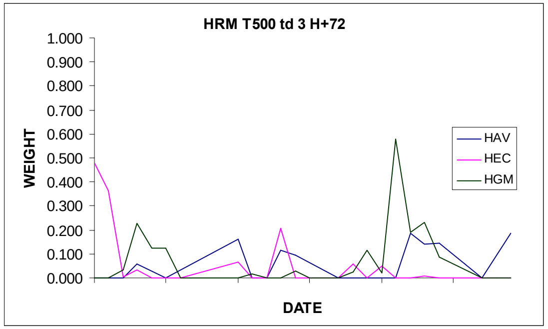


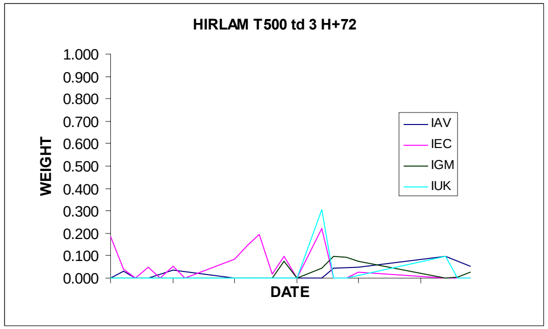


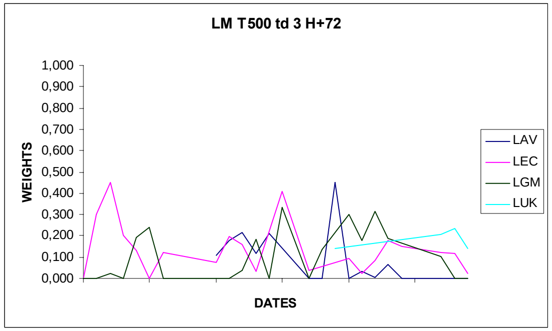


WEIGHTS TIME SERIES

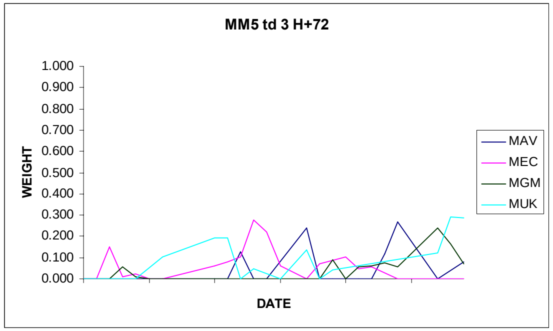
T 500 td 3 H +72

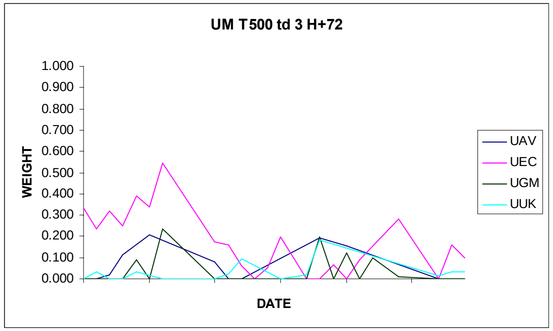






HRM HIRLAM LOKAL MODEL





MM5 UM

CONCLUSIONS

- As Gneiting and Raftery have published in Science (2005): “We anticipate notable improvements in probabilistic forecast skill through the continued development of multimodel, multi–initial condition ensemble systems and advanced, grid-based statistical postprocessing techniques. “
- Better calibrated ensemble has been obtained after BMA for Z500 and T500. First calibration results exhibit a great improvement in spread-skill calibration as well as better rank histograms.
- The preliminary calibrated ensemble for S10m shows a good spread-skill relationship, reduction of outliers in rank histograms, better reliability diagrams and brier skill scores than multimodel.
- Testing longer training periods seems to be necessary. However, if there is not a clear improvement in verification scores by using these longer training periods, then the shortest training period seems to be suitable for calibration of short range forecasts.

FUTURE WORK

- BMA calibration of surface parameters with quasi-normal conditional PDFs: T2m, Pmsl and S10m.

- Research on optimum length of the training periods for surface parameters.

- BMA calibration of accumulated precipitation using the cube root of precipitation amount and the conditional PDF (Slaughter, JM et al.,2006):

$$p(parameter \mid f_k) = \frac{1}{\beta_k^{\alpha_k}\Gamma(\alpha_k)} y^{\alpha_k - 1} \exp(y/\beta_k)$$

REFERENCES

- Carter, G. M., J. P. Dallavalle, and H. R. Glahn, 1989: Statistical forecasts based on the National Meteorological Center's numerical weather prediction system. *Wea. Forecasting,* **4,** 401–412.
- Dempster, A. P., N. M. Laird, and D. B. Rubin, 1977: Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc.*, **39B,** 1–39.
- Fisher, R. A., 1922: On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London,* **222A,** 309–368.
- Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.,* **11,** 1202–1211.
- Gneiting, T. and Raftery, A.E. (2005). Weather forecasting with ensemble methods. *Science*, 310, 248-249.
- McLachlan, G. J., and T. Krishnan, 1997: *The EM Algorithm and Extensions.* Wiley, 274 pp.
- Raftery, A.E., Gneiting, T., Balabdaoui, F. and Polakowski, M. (2005). Using Bayesian Model Averaging to Calibrate Forecast Ensembles. *Monthly Weather Review*, 133, 1155-1174.

REFERENCES

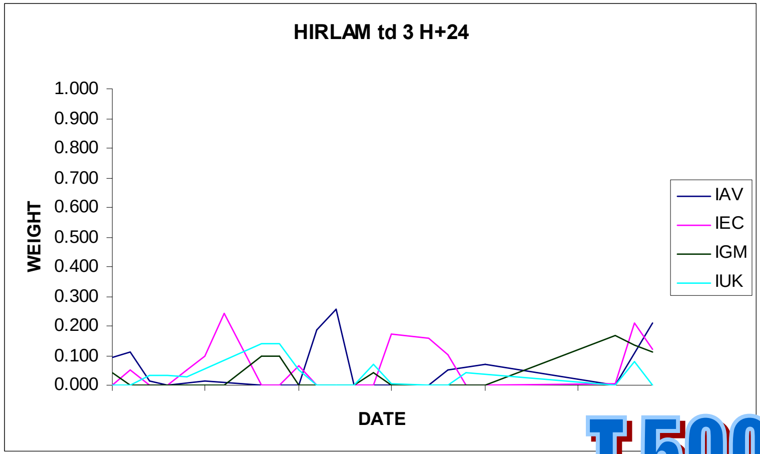
- Raftery, A.E., Painter, I. and Volinsky, C.T. (2005). BMA: An R package for Bayesian Model Averaging. *R News*, volume 5, number 2, 2-8.
- Richard A. Becker, John M. Chambers and Allan R. Wilks (1988), *The New S Language.* Chapman & Hall, New York. This book is often called the "*Blue Book*".
- Slaughter, J.M., Adrian E. Raftery and Tilmann Gneiting (2006), Probabilistic Quantitative Precipitation Forecasting Using Bayesian Model Averaging. Technical Report no. 496 Dpt. Statistics University of Washington.

BMA LINKS

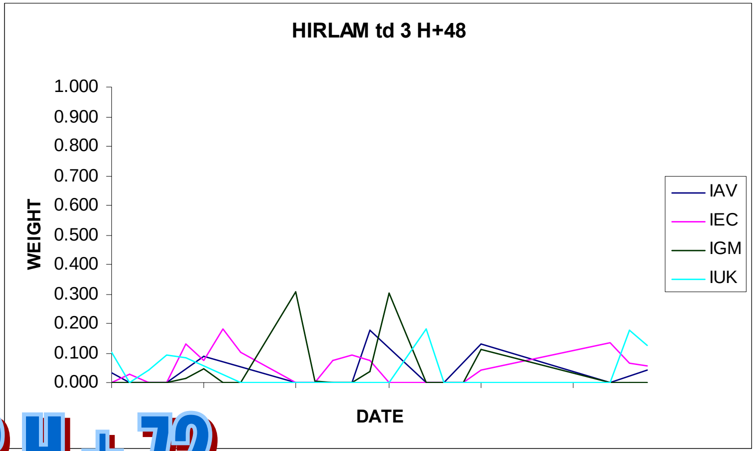
- http://www.stat.washington.edu/raftery/
- http://bma.apl.washington.edu/
- http://www.r-project.org/
- http://www.research.att.com/~volinsky/bma.html

WEIGHTS TIME SERIES

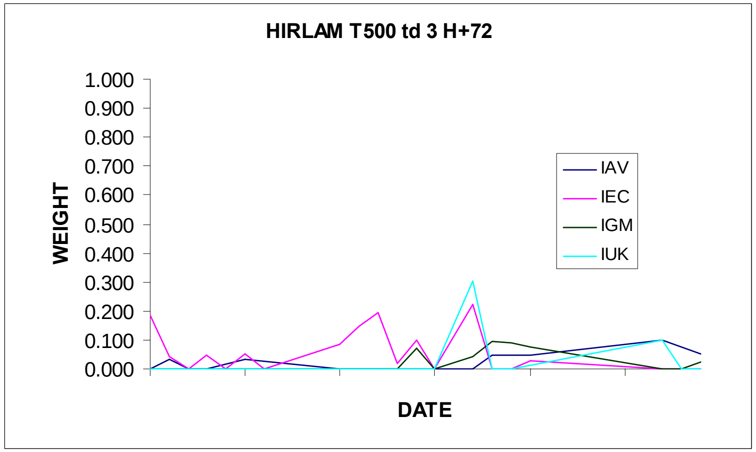
T500 td 3 H+24



T500 td 3 H+48



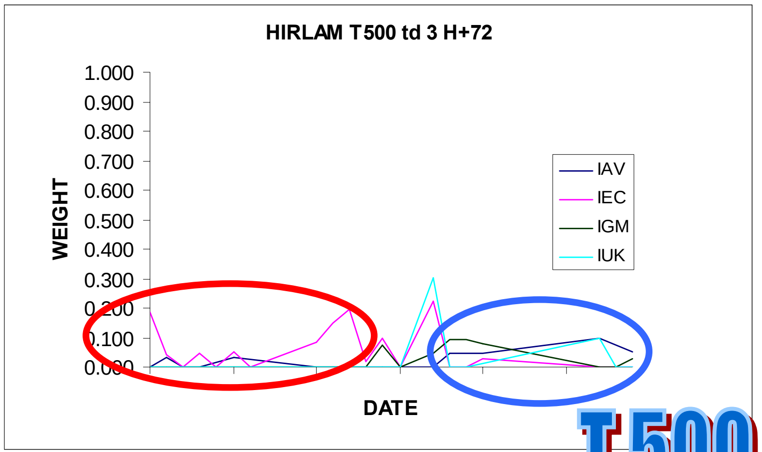
T500 td 3 H+72



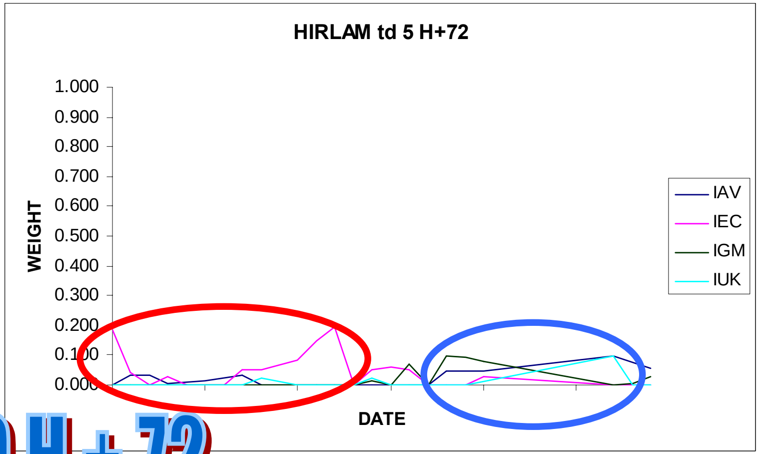
HIRLAM

WEIGHTS TIME SERIES

T500 td 3 H+72

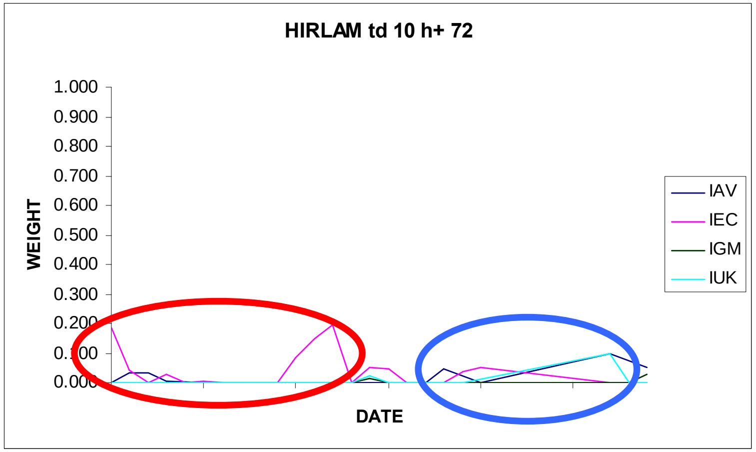


T500 td 5 H+72



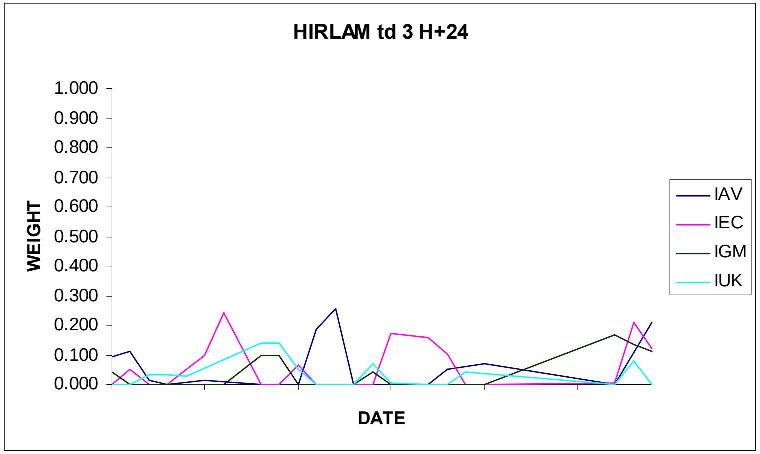
T500 td 10 H+72

HIRLAM

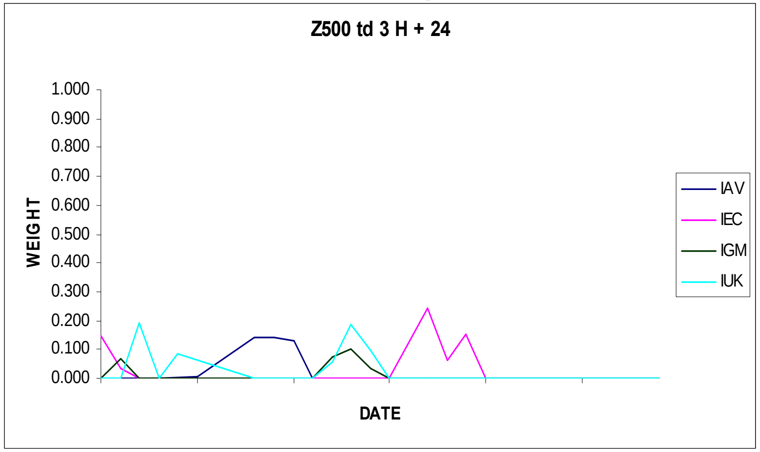


WEIGHTS TIME SERIES

T 500 td 3 H + 24



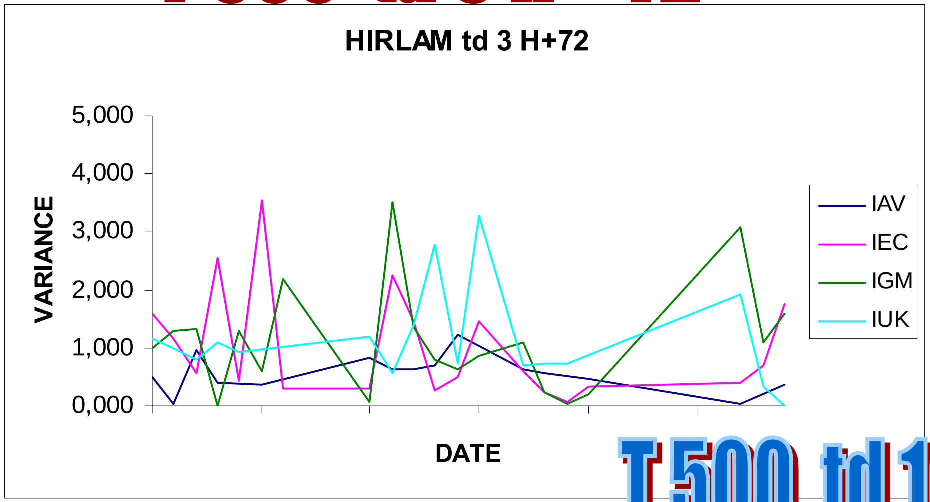
Z500 td 3 H + 24



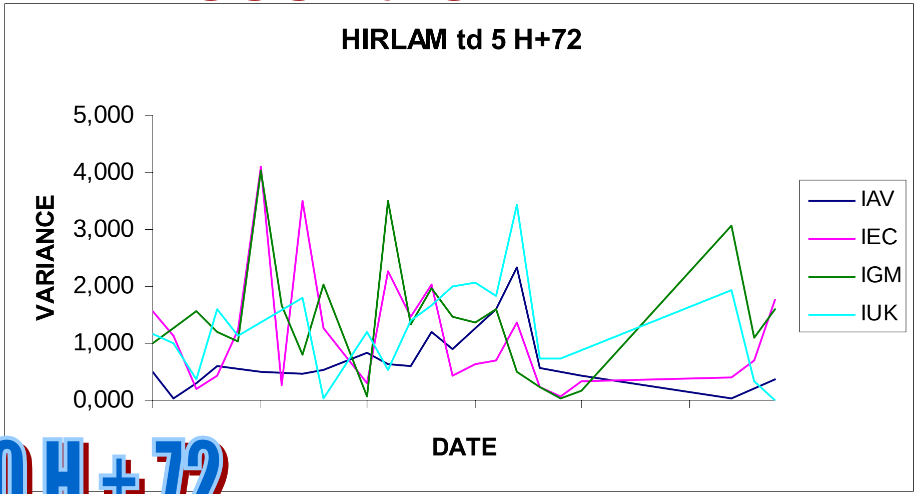
HIRLAM

VARIANCE TIME SERIES

T 500 td 3 H + 72

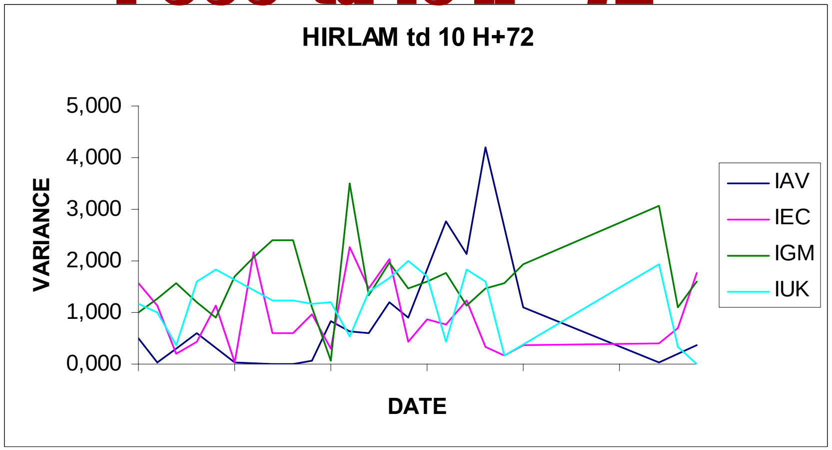


T 500 td 5 H + 72



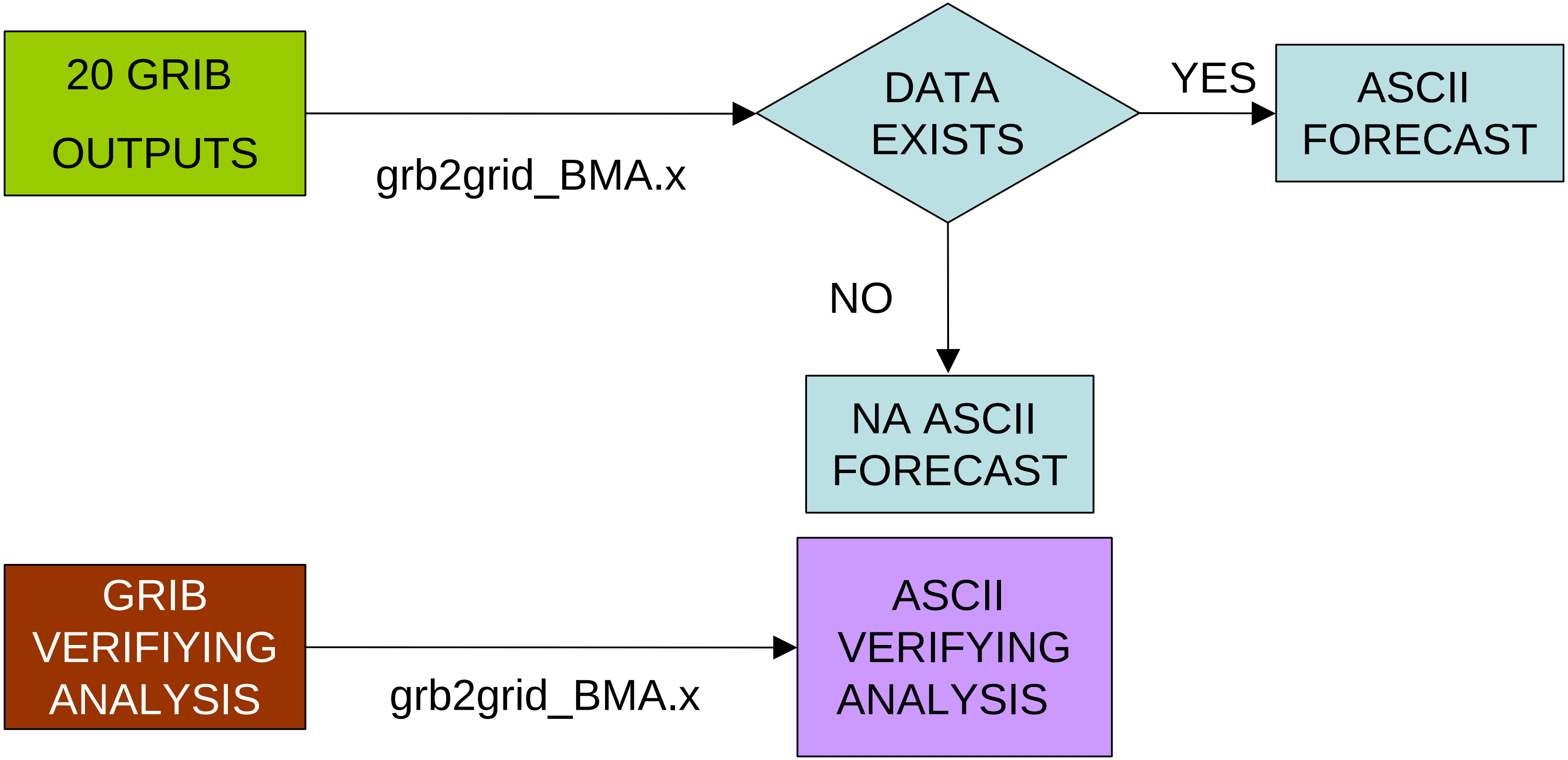
T 500 td 10 H + 72

HIRLAM



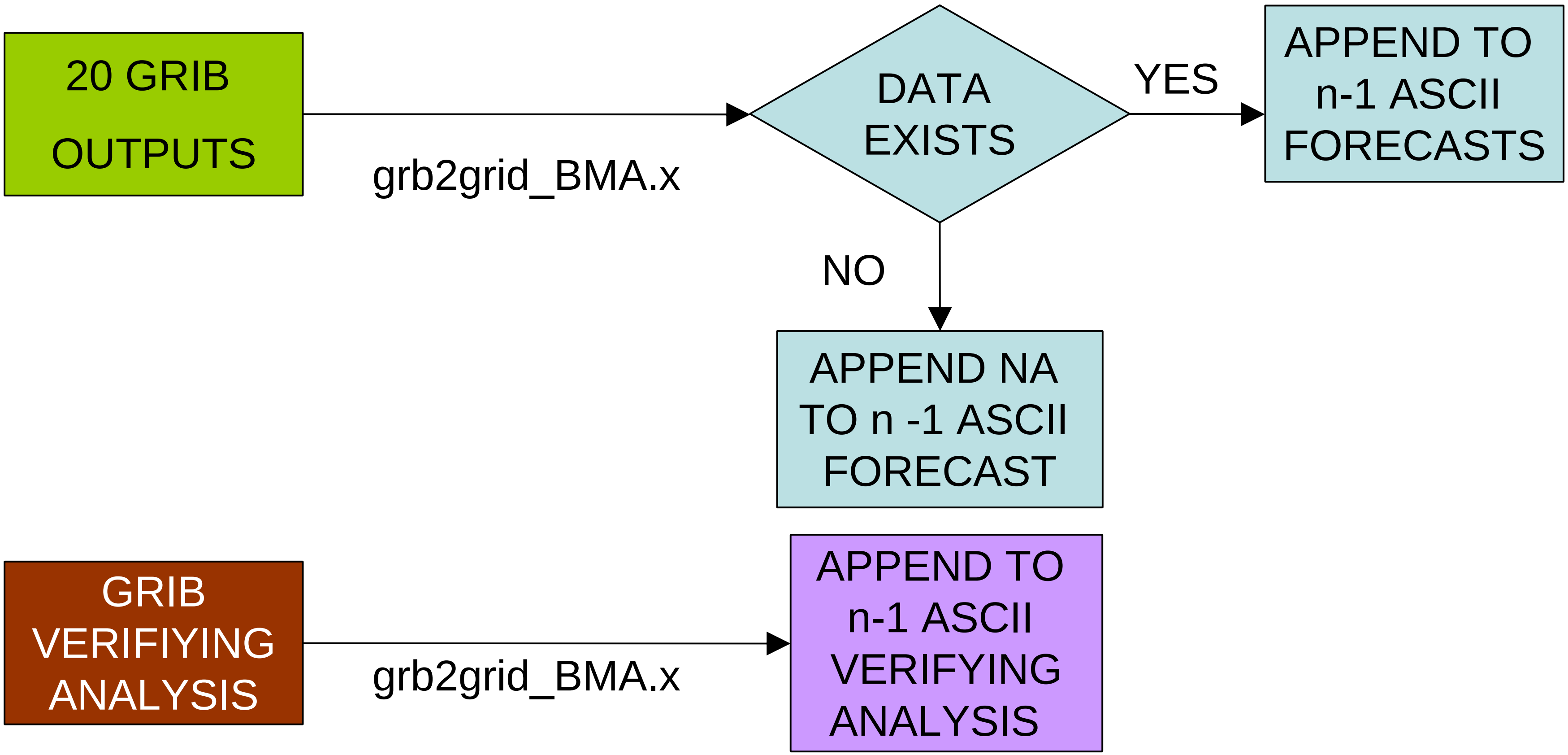
BMA IMPLEMENTATION

- TRAINING DATA GENERATION
  - 1st training day



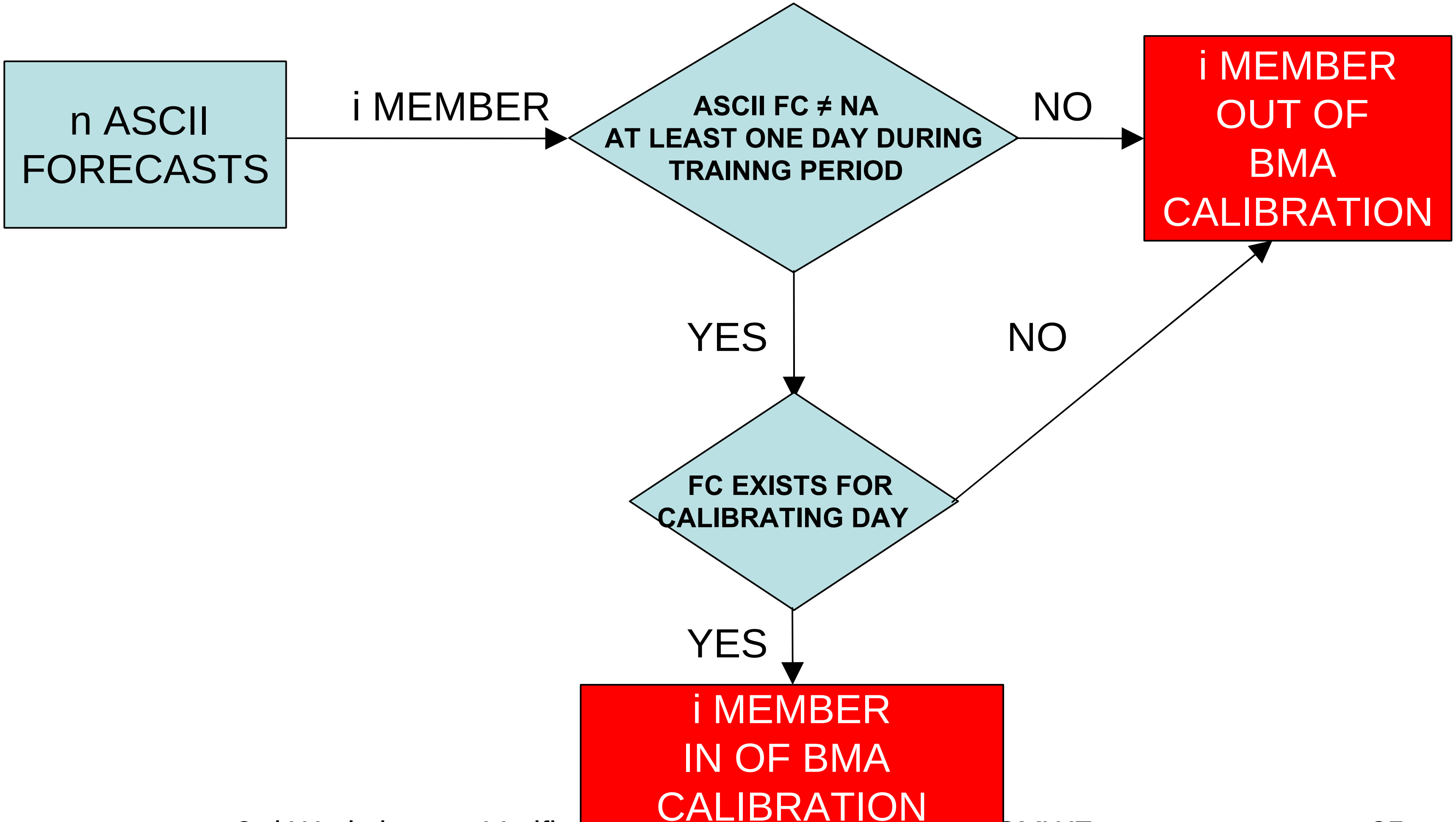
BMA IMPLEMENTATION

- TRAINING DATA GENERATION
  - Nth training day



BMA IMPLEMENTATION

- Ensemble BMA over R



BMA IMPLEMENTATION

- The BMA has been implemented over R (www.r-project.org ; Richard et al.1998) and the ensembleBMA package (http://cran.r-project.org/src/contrib/Descriptions/ensembleBMA.html, Raftery et.al. 2005)

- Our SREPS is multi-model multi-initial conditions with 72 hours forecast integrations twice a day (00 & 12 UTC).
- 5 models x4 boundary conditions =>20 member ensemble every 24 hours
- The models outputs are codified in GRIB