The SuperEnsemble technique

Daniele Cane, Massimo Milelli

VI COSMO Meeting
Milano, September 22-24 2004

Overview

- Multimodel Ensemble and SuperEnsemble Theory
- Multimodel results
- A comparison with Kalman filter results
- Conclusions

We use the following operational runs of the 0.0625° resolution version of LM (00UTC runs)

Local Area Model Italy (UGM, ARPA-SIM, ARPA Piemonte) (nud00)
Lokal Modell (Deutscher Wetterdienst) (lkd00)
aLpine Model (MeteoSwiss) (alm00)

We evaluate the model performances with respect to our regional high resolution network.

Here presented are the results of 2m temperature forecasts, compared with the measurements of 201 stations, divided in altitude classes (<700 m, 700-1500 m, >1500 m).

Multimodel Theory

As suggested by the name, the Multimodel SuperEnsemble method requires several model outputs, which are weighted with an adequate set of weights calculated during the so-called training period.

The simple ensemble method with bias-corrected or biased data respectively, is given by

$$S = \frac{1}{N}\sum_{i=1}^{N}\left(F_i - \overline{F_i}\right) \qquad (1) \qquad \text{or} \qquad S = \frac{1}{N}\sum_{i=1}^{N}\left(F_i - \overline{O}\right) \qquad (2)$$

The conventional superensemble forecast (Krishnamurti et. al., 2000) constructed with bias-corrected data is given by

$$S = \overline{O} + \sum_{i=1}^{N} a_i\left(F_i - \overline{F_i}\right) \qquad (3)$$

[illegible]

The calculation of the parameters a_i is given by the minimisation of the mean square deviation

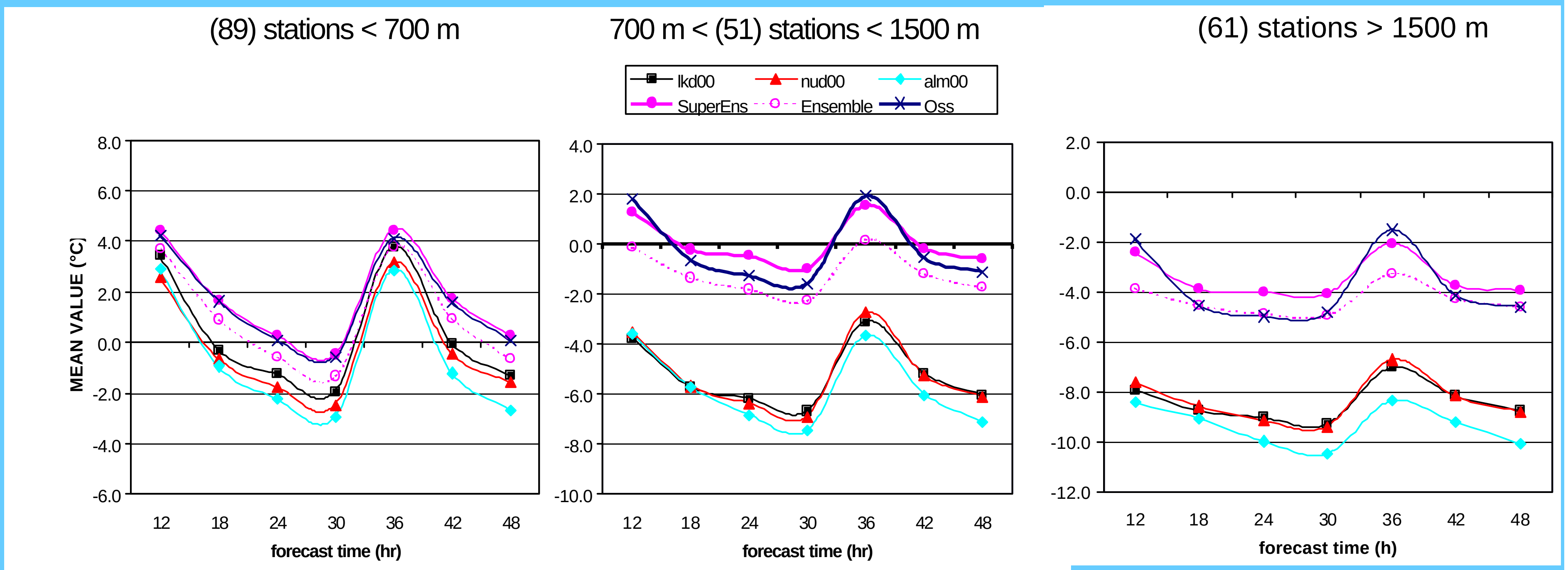
$$G = \sum_{k=1}^{T} \left(S_k - O_k \right)^2$$

by derivation $\left(\frac{\partial G}{\partial a_i} = 0 \right)$ we obtain a set of N equations, where N is the number of models involved:

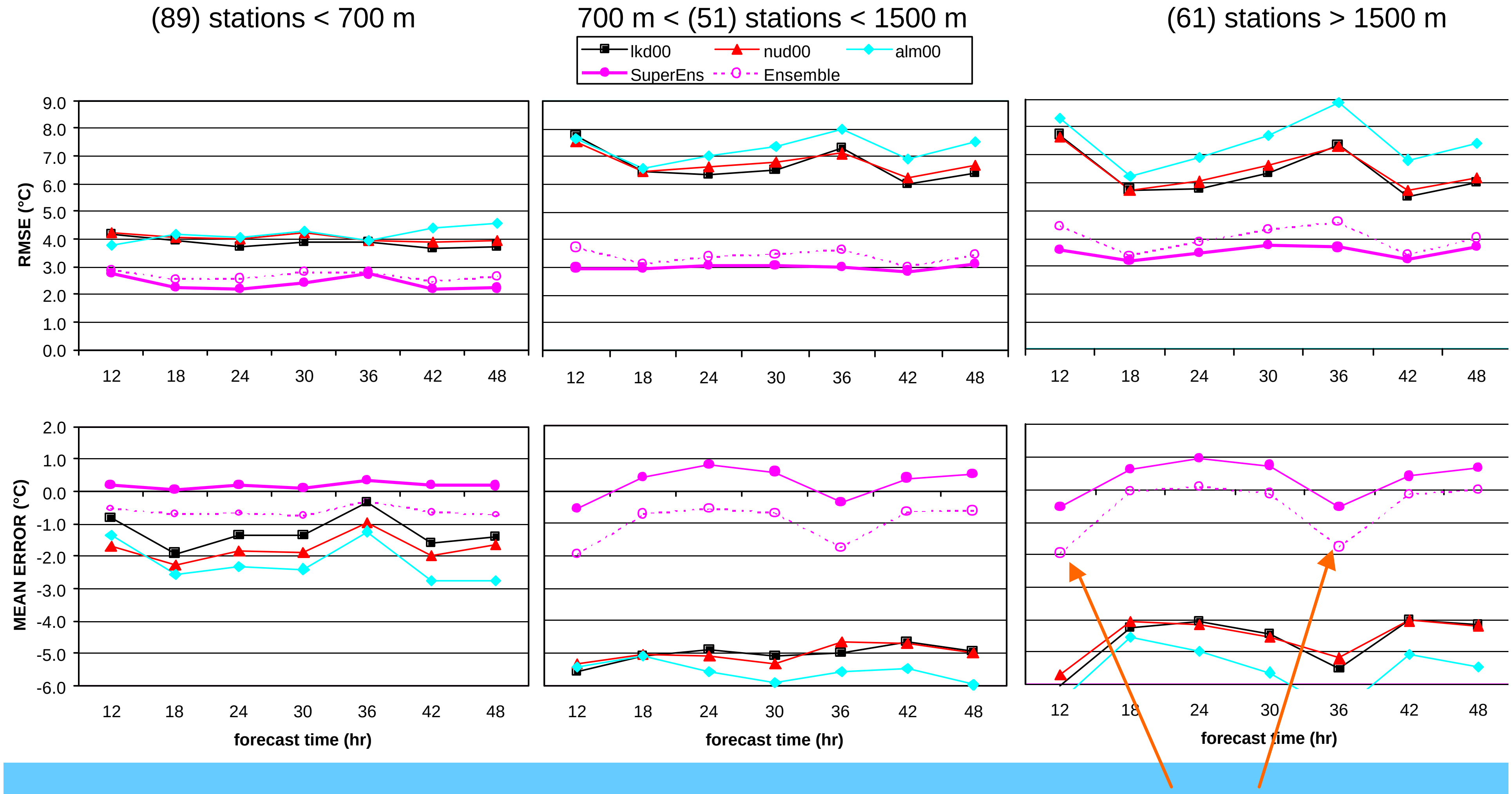
$$\begin{pmatrix} \sum_{k=1}^{T} \left(F_{1_k} - \overline{F_1}\right)^2 & \sum_{k=1}^{T} \left(F_{1_k} - \overline{F_1}\right)\left(F_{2_k} - \overline{F_2}\right) & \cdots & \sum_{k=1}^{T} \left(F_{1_k} - \overline{F_1}\right)\left(F_{N_k} - \overline{F_N}\right) \\ \sum_{k=1}^{T} \left(F_{2_k} - \overline{F_2}\right)\left(F_{1_k} - \overline{F_1}\right) & \sum_{k=1}^{T} \left(F_{2_k} - \overline{F_2}\right)^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ \sum_{k=1}^{T} \left(F_{N_k} - \overline{F_N}\right)\left(F_{1_k} - \overline{F_1}\right) & \cdots & \cdots & \sum_{k=1}^{T} \left(F_{N_k} - \overline{F_N}\right)^2 \end{pmatrix} \bullet \begin{pmatrix} a_1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ a_N \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{T} \left(F_{1_k} - \overline{F_1}\right)\left(O_k - \overline{O}\right) \\ \vdots \\ \vdots \\ \sum_{k=1}^{T} \left(F_{N_k} - \overline{F_N}\right)\left(O_k - \overline{O}\right) \end{pmatrix}$$

We then solve these equations using Gauss-Jordan method.

Multimodel post-processing results: JANUARY 2004

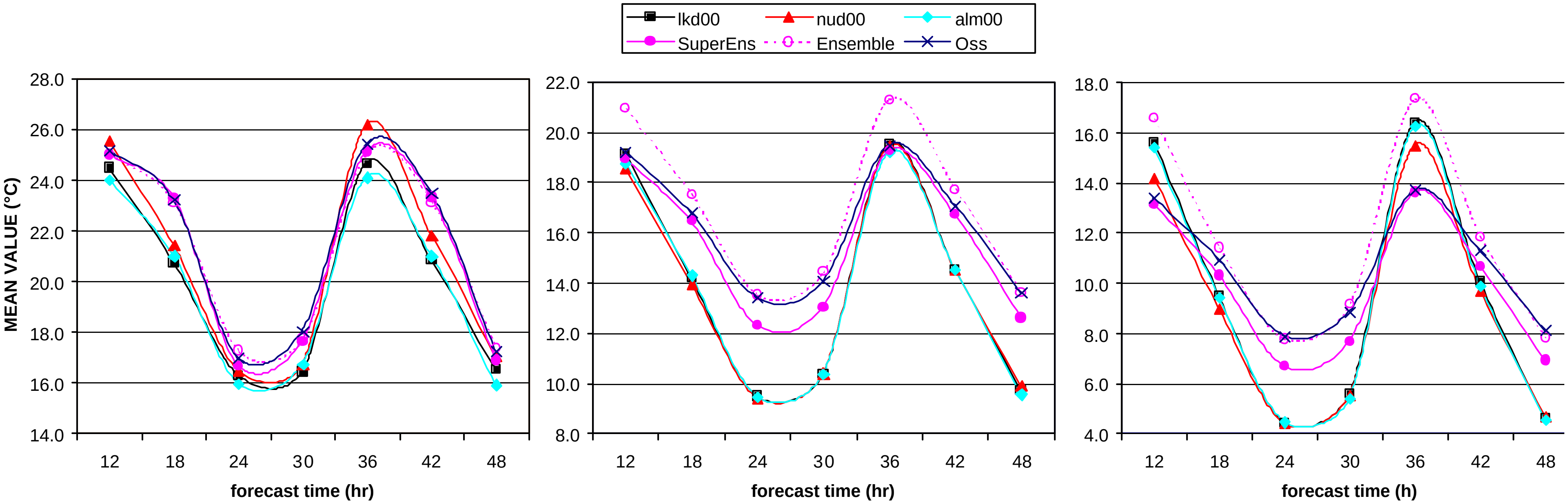


Multimodel post-processing results: JANUARY 2004

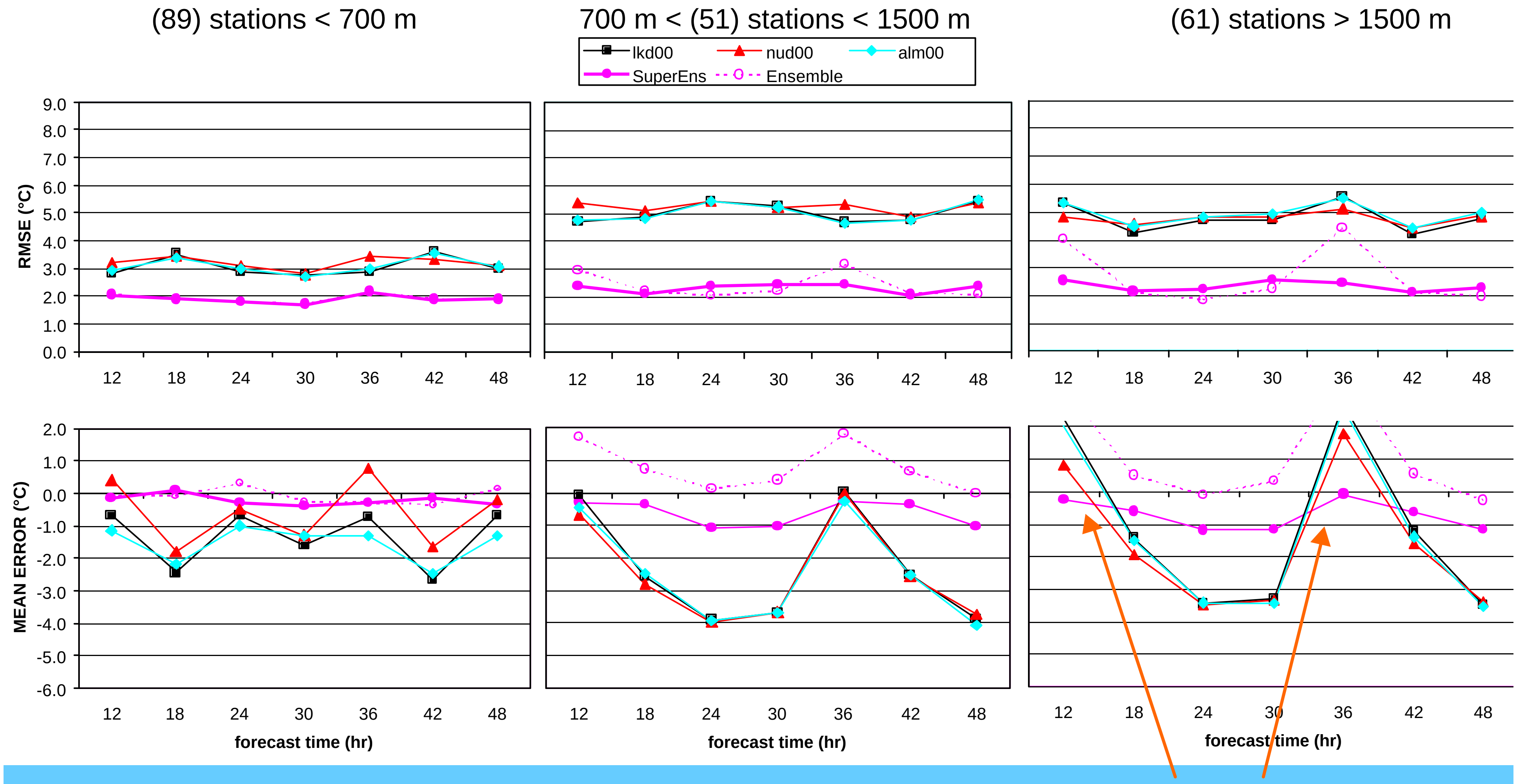


Multimodel post-processing results: JUNE 2004

(89) stations < 700 m 700 m < (51) stations < 1500 m (61) stations > 1500 m



Multimodel post-processing results: JUNE 2004



A comparison with Kalman filter results

A version of the Kalman filter (Kalman, 1960) post-processing on the direct output of ECMWF run at 12UTC is currently used every day for the 2m temperature forecasts over the entire Piedmont, but we register a degradation of the predicted values with increasing height of the weather stations. The same problem is evident also in the filtering of the limited area model outputs.

The reason for this degradation can be found in the strong variability of the performances of the models day by day in the alpine region. In fact Kalman filter is very good in reducing the strong systematic errors, but fails when the difference between predicted and observed values varies strongly from one day to the next one.

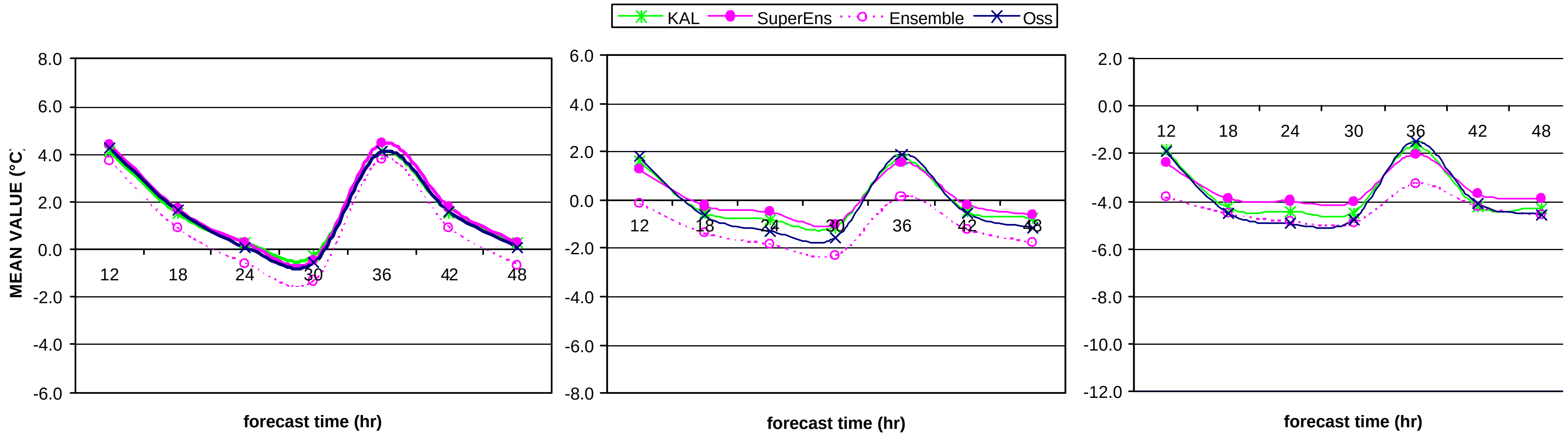
Here we present the comparison between a test run of Kalman filter post-processing method on LM and the Multimodel method results.

Confrontation between Kalman filter (based on LM-DWD) and Multimodel results: JANUARY 2004

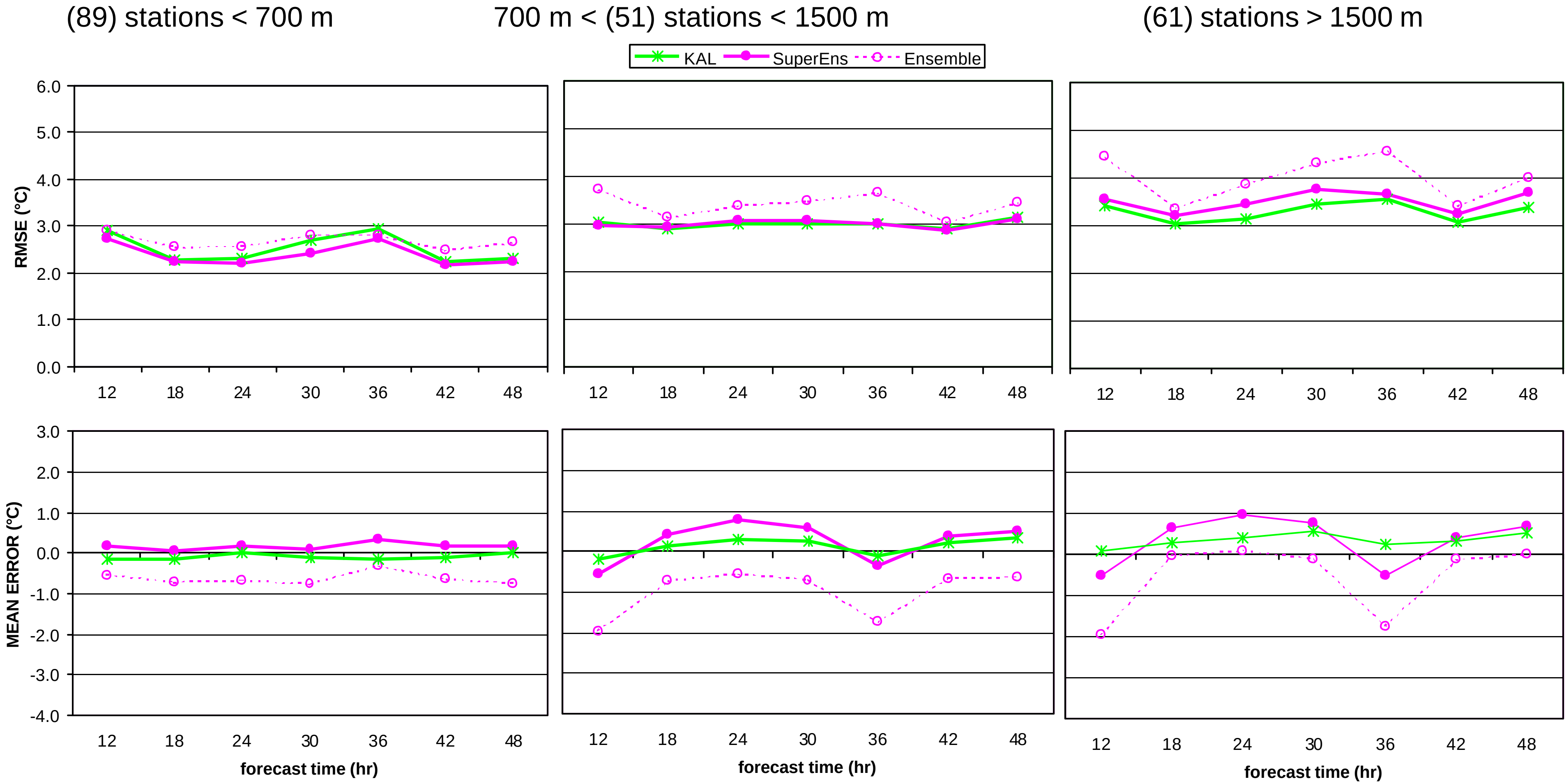
(89) stations < 700 m

700 m < (51) stations < 1500 m

(61) stations > 1500 m

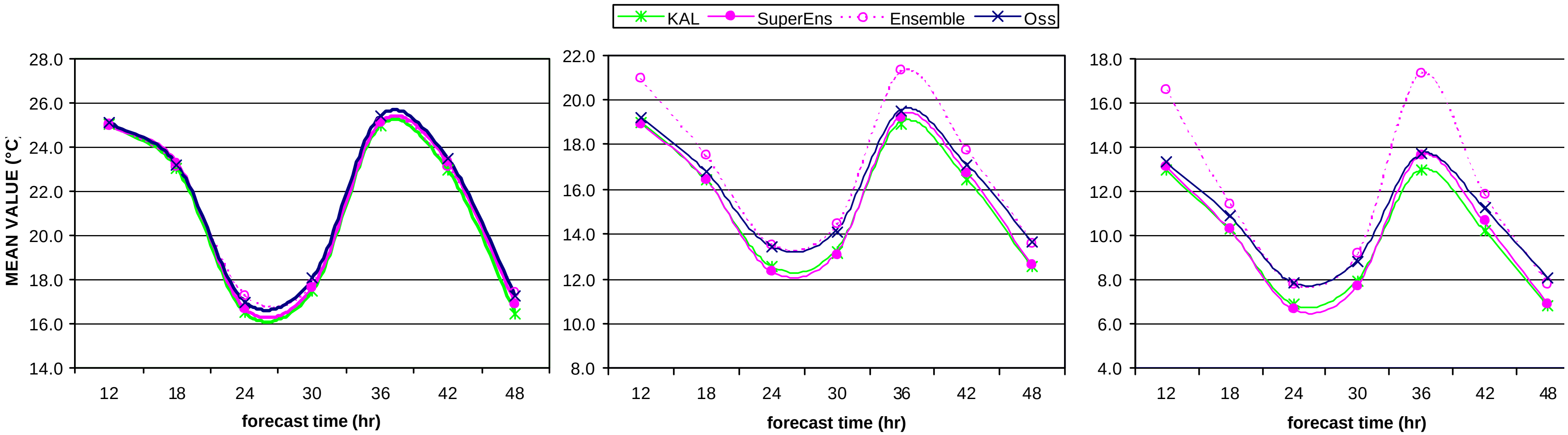


Confrontation between Kalman filter (based on LM-DWD) and Multimodel results: JANUARY 2004



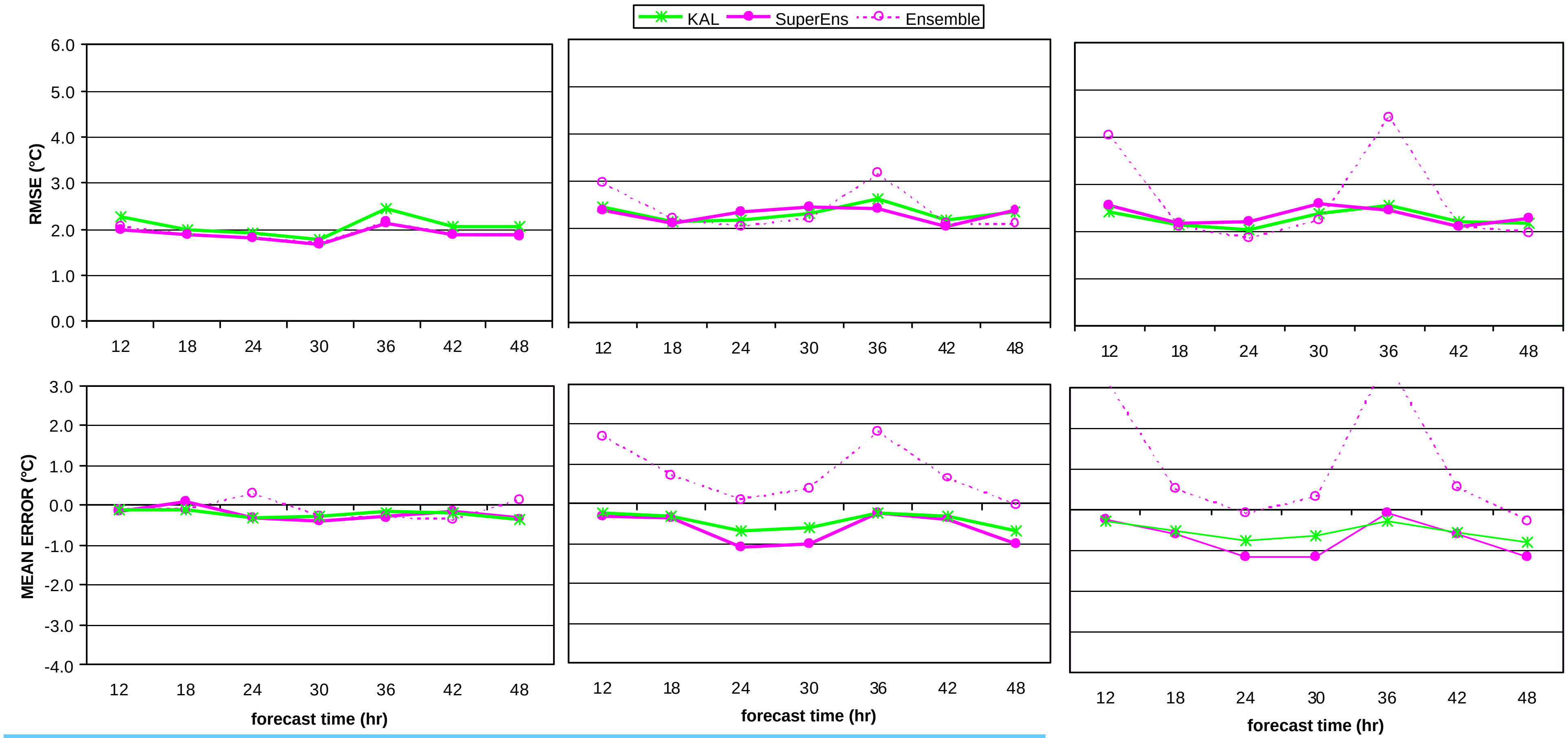
Confrontation between Kalman filter (based on LM-DWD) and Multimodel results: JUNE 2004

(89) stations < 700 m | 700 m < (51) stations < 1500 m | (61) stations > 1500 m



Confrontation between Kalman filter (based on LM-DWD) and Multimodel results: JUNE 2004

(89) stations < 700 m | 700 m < (51) stations < 1500 m | (61) stations > 1500 m



Conclusions

- Direct Model Output 2m temperature forecasts show a notable degradation in the Alpine region
- The Multimodel technique is tested for the first time on limited area models in the Alpine area
- The Multimodel improves the forecasts in high mountains locations, both in bias and RMSE and its performances are similar to those from Kalman filter
- The Kalman filter is a much more complex technique and it is not suitable for all kind of variables

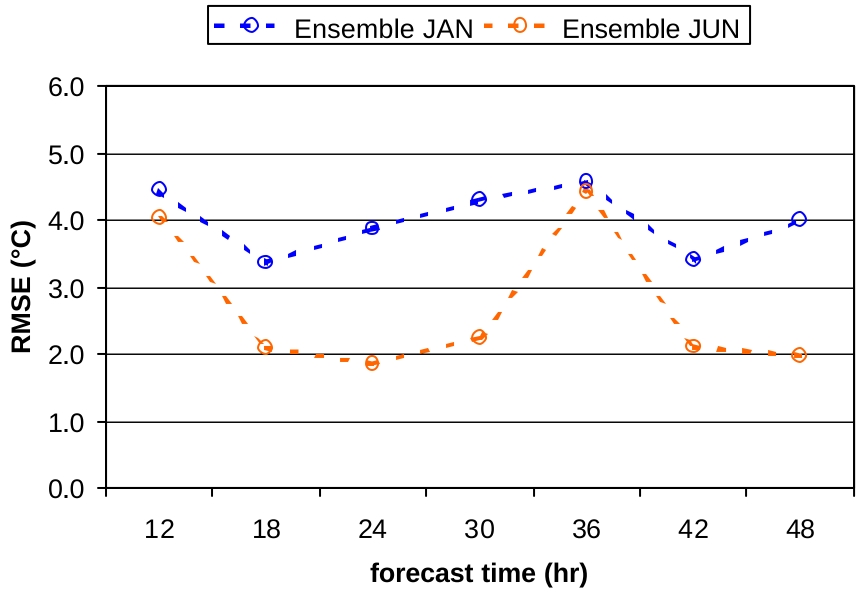
Work in progress...

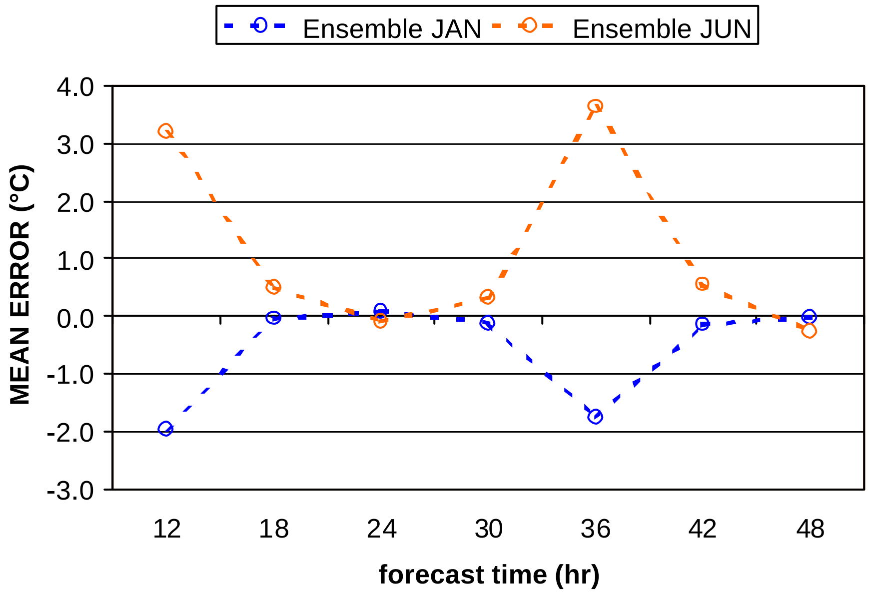
- Extension of the Multimodel method to other parameters: humidity, precipitations…..

Concluding remarks

2m temperature RMSE for higher stations, as shown by Multimodel Ensemble, is worse during daytime hours (+12 hr and +36 hr forecasts), and the bias is opposite during winter (models underestimate the observed values) with respect to summer (models overestimate) ⟶ **snow on the ground !** ⟶ **Heat fluxes ?**

(61) stations > 1500 m





References

- Kalman R. E., *Journal of Basic Engineering, 82 (Series D): 35-45*, 1960
- Krishnamurti T.N. *et al., Science, 285, 1548-1550*, 1999
- Krishnamurti, T. N. *et al., J. Climat*e, *13, 4196-4216*, 2000
- COSMO Newsletter, 3, 2003

Acknowledgements

We wish to thank the Deutscher Wetterdienst and MeteoSwiss for providing the models outputs for this research work.