Comparing COSMO Models 3.19 vs. 3.5 on the 30h Precipitation Forecast by Parallel Run during July - September 2006 over Europe.

Jan Wacław Parfiniewicz

IMGW Poland

1 Introduction

During July-September 2006 two versions of COSMO-Model: 3.5 and 3.19 (both in terrain following coordinates) have been tested in parallel. Altogether 90 precipitation forecasts from parallel runs starting at 00 UTC for 30h have been completed and compared against daily totals GTS SYNOP data over Europe (the runs 20060701 and 20060729 have been dropped out). The models were run with diagnostic precipitation scheme version as it has been set into the operational IMGW model LM_3.5 14 km grid version (taken as reference), which is stably running there since about 2 years.

2 Standard statistics

The averaged results of the statistical indices used for quantification of precipitation forecasts LM_3.5 vs. LM_3.19 are the following (PC - proportion correct, HIT - probability of correct rain signal, FAL - false alarm, BIAS - systematic error, HSS - Heidke skill score, PSS - Peirce skill score, ETS - Equitable threat score, CHI - the measure of cross relationship with χ^2 distribution, DEV - standard deviation error, and, CC - cross correlation coefficient):

LM:	PC	HIT	FAL	BIA	HSS	PSS	ETS	CHI	CC	DEV
3.5	0.81	0.75	0.20	1.26	0.52	0.55	0.37	0.30	0.50	5.16
3.19	0.81	0.74	0.20	1.21	0.52	0.54	0.36	0.30	0.50	5.28

Table 1: Standard statistics: averaged values

An example of ETS diagram is presented in Fig. 1. The vertical lines indicate dates when significant errors occurred (in red for LM_3.5, in green for LM_3.19).

As standard statistics are not sensitive enough to assess model behavior in extreme cases, it is necessary to work with all spectrum of errors.

3 Comparing overall 30h forecast errors distributions: LM_3.5 vs. LM_3.19

Distribution of errors in both models was based on more than 50000 sample points alltogether. In Fig. 2 two diagrams of precipitation errors are presented; left: monotonically arranged (ranked) by self - order from smallest (negative) to largest (positive); right: by classes of errors' magnitude. As the errors of order $< -2, 2 >$ (the thick band for numbers between $\sim$ 5000 - 44000 on right panel) are negligible (however it may comprise cases of forecasts with severe precipitation), the further analysis will concentrate on separating extreme cases from dominating "dust" by zooming right and left parts of the diagram.

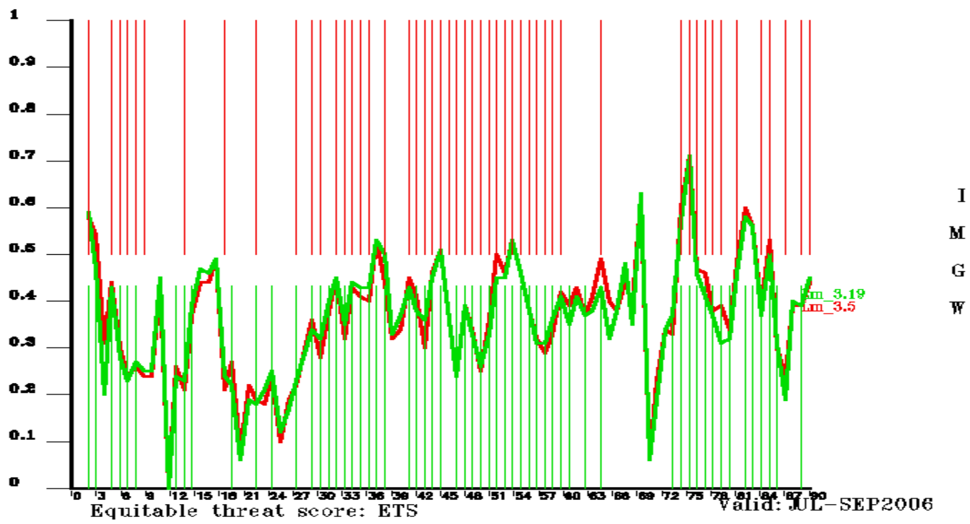
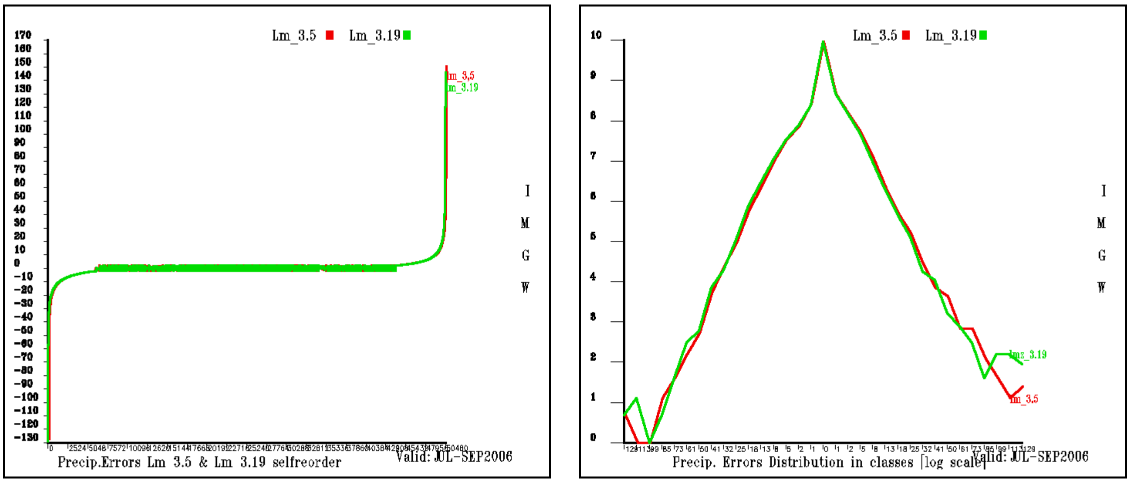


Figure 1: ETS diagram



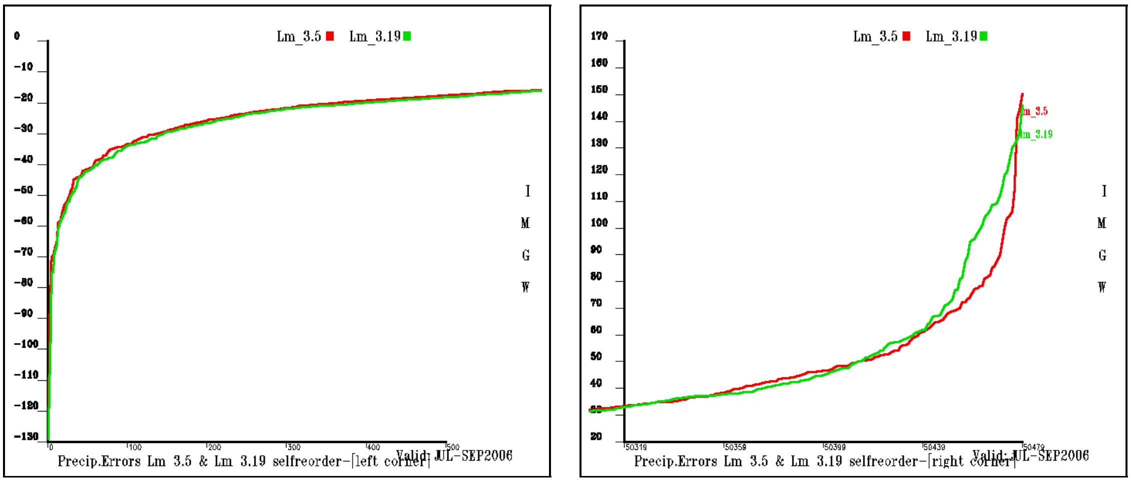
(a) Errors by selforder

(b) Errors by classes

Figure 2: Overall 30h forecast precipitation errors distributions

In Fig. 3 investigation of the two border distributions is presented in the following way: left for values under predicted ones (dry branch) and right for overshoots (wet branch).

The further way to proceed is clear: select doubtful, significant cases and attempt to develop new model version that would clarify situation (or enlarge range of clarity). A control list of 162 point-cases errors > 40 [mm/day] was selected, which gives altogether 56 dates to monitor model quality progression (the administration of such a list is a separate challenge). Each of such cases has its own metric and graphical illustration. An example of the extremely imperfect LM_3.19 forecast is the 8th of July, when daily total error for Chemnitz, Germany (10577) was 121 [mm/day], while for LM_3.5 "only" 53 [mm/day] (Fig. 4). The illustration shows influence of singular error signal onto water accumulation over a given station. When taking accumulated water amount for an enough long period of aggregation as an ultimate indicator of forecast quality, one may see how huge forecast collapses are then smoothly compensated by a number of less significant events.



(a) Errors by selforder: left side (b) Errors by selforder: right side

Figure 3: Investigation of the two border distributions

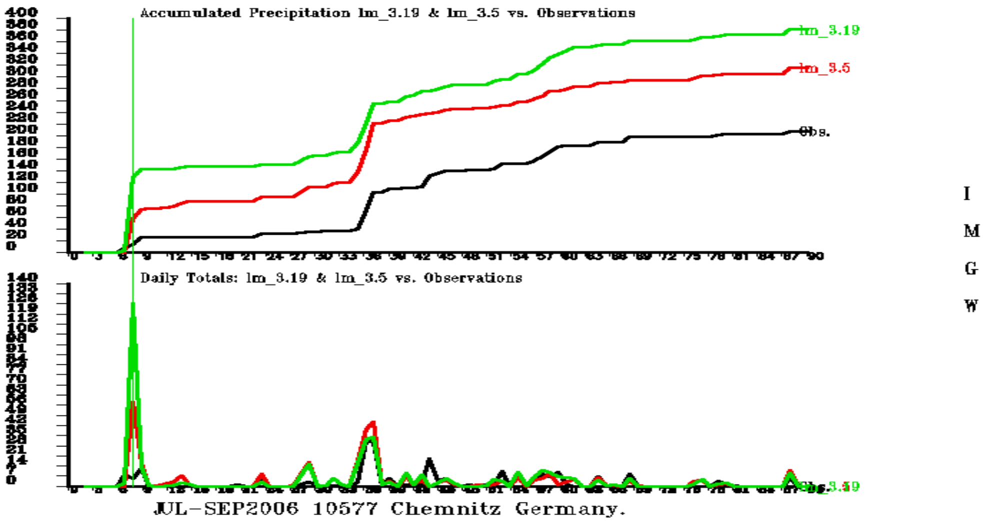


Figure 4: Accumulated precipitation and daily totals: Chemnitz Germany (10577)

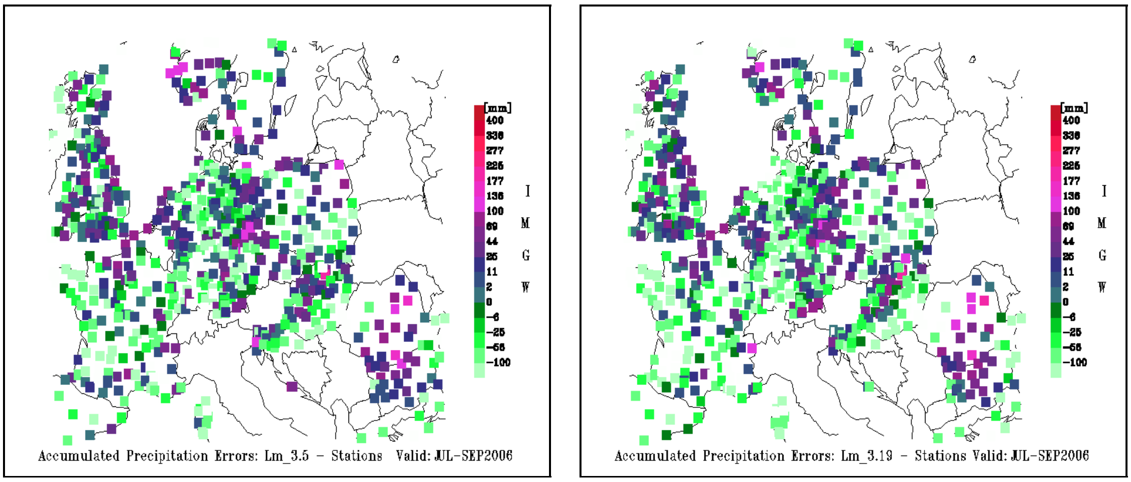
4 Comparing accumulated precipitation water errors over Europe

a) Distribution of errors over Europe : LM_3.5 vs. LM_3.19. (Fig. 5)

b) Relative Predominance LM_3.19 over LM_3.5 regarding accumulated water (Fig. 6). To asses each particular relationship between errors over given geographical point, the simple predominance factor RP is proposed:

$$RP = (\text{Relative Predominance } (|\text{DelLM\_3.5}|)/(|\text{DelLM\_3.5}| + |\text{DelLM\_3.19}|)$$

where DelLM_(given version) means difference between 30 h forecast of daily total (i.e. 30h -06h model forecasts of total precipitation) and its realization over given station. If the number of errors in LM_3.5 is large, the RP of LM_3.19 approaches the value 1. Calculated RP of LM_3.19 of accumulated water allows for ultimate geographical review of errors over Europe (Fig. 6).



(a) Lm_3.5 errors (b) Lm_3.19 errors

Figure 5: Comparing Accumulated Precipitation Water Errors over Europe

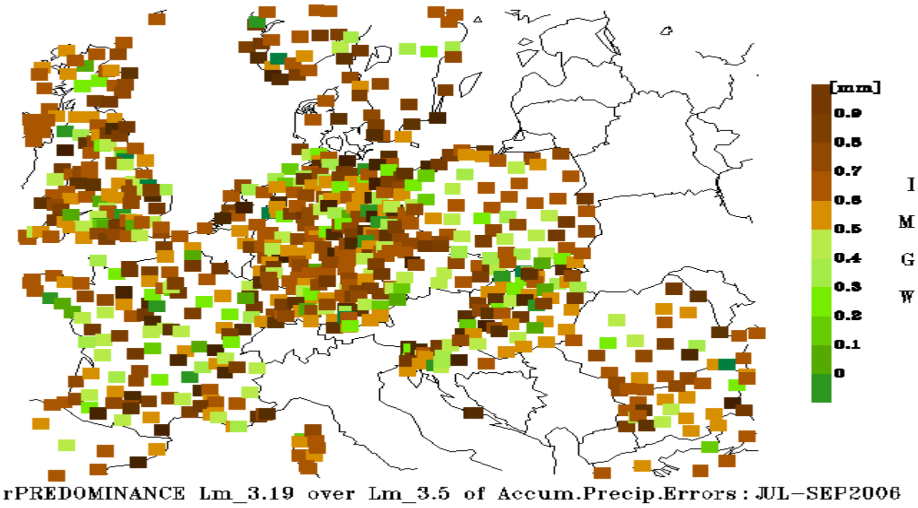


Figure 6: Relative Predominance LM_3.19 over LM_3.5 regarding accumulated water

5 Summary

Despite there is no evidence that Total_Precipitation LM_3.19 30h forecast errors are significantly smaller than in the LM_3.5 version, the accumulated water errors during the analyzed season indicate clear predominance of the newer version. This result, obtained for summer season, indicates that configuration applied at IMGW till now, should be changed to account for new developments - relevant to finer mesh size (2.8 km), the CLM developments and changes in the new multi-layer soil model. Similarly, the diagnostic precipitation parametrisation should be re-tested against the prognostic one in the near future.

Acknowledgments
I appreciate the input to this work by my valued colleagues: Uli Schättler, Marek Lazanowicz, Jürgen Steppeler.