Intercomparison of the ensembles of the TIGGE-LAM archive: first results at ARPA-SIMC

Andrea Montani, C. Marsigli, T. Paccagnella

ARPA Emilia-Romagna Servizio IdroMeteoClima, Bologna, Italy

EWGLAM Meeting
Belgrade, 5 October 2015

arpa agenzia regionale prevenzione e ambiente dell'emilia-romagna

CONSORTIUM FOR SMALL SCALE MODELING COSMO

Outline

- **TIGGE-LAM archive**:
  - Introduction
  - Data providers
- **Performance of TIGGE-LAM ensembles**:
  - Methodology of verification
  - 2-metre temperature spread/skill
  - Probabilistic verification of precipitation
  - Single model vs multi-model approach
- **Conclusions and plans**

About TIGGE-LAM

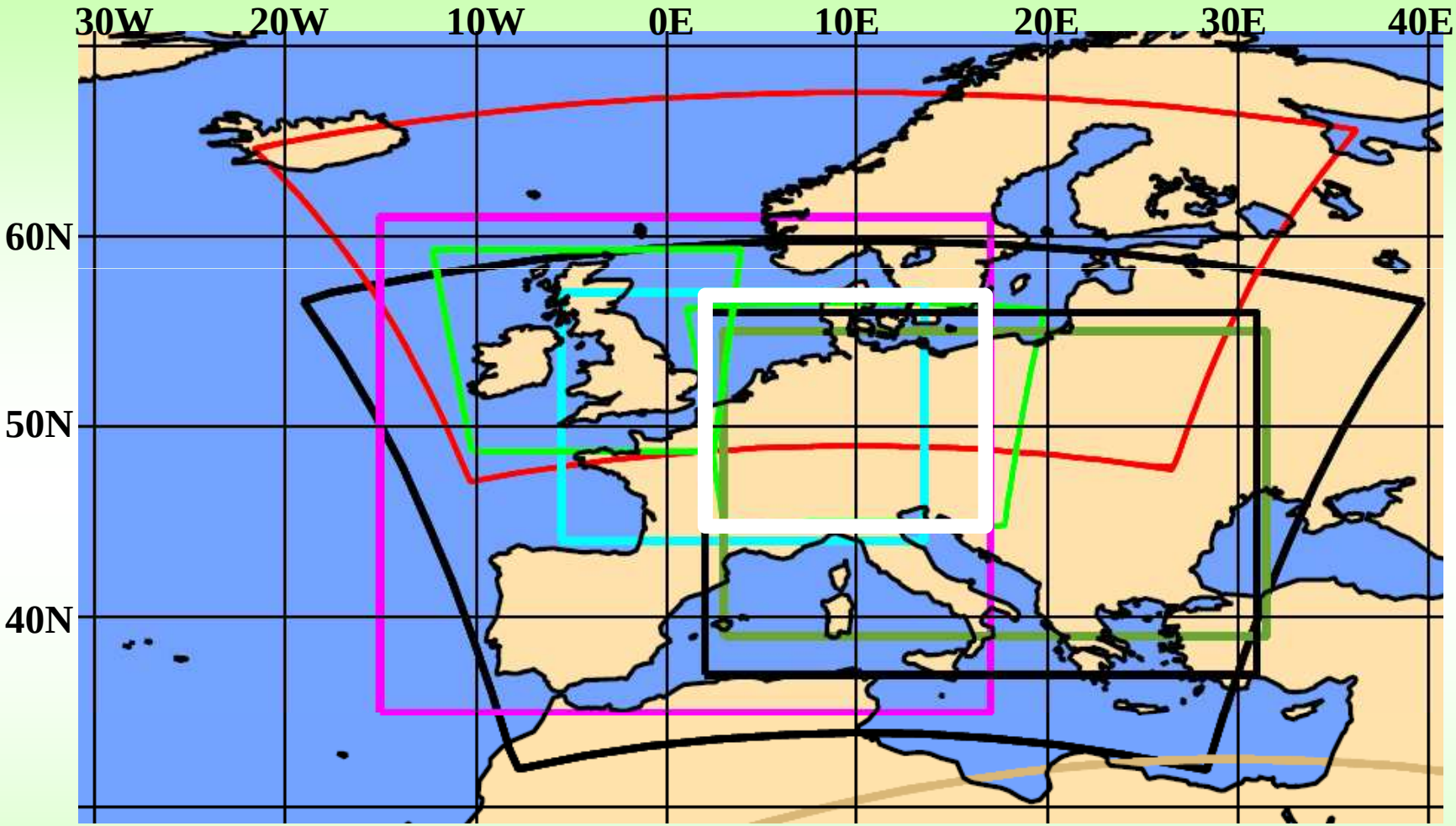
TIGGE-LAM is an extension of the THORPEX Interactive Grand Global Ensemble (TIGGE) to include weather forecasts from limited area model (LAM) ensembles.

Archive of some parameters by a set of European limited-area ensemble systems running on an operational basis with the following specification of the input data:

- Data format: WMO-GRIB2.
- Time step frequency: 3h (cumulated parameters will be not archived at step 0).
- Grid: original model grid.
- High-priority Parameters: 10u, 10v, cape, cin, mslp, **2t**, 2d, **tp**, lsp, 10fg3, orography, land-sea mask.

Currently, 7 systems populate the TIGGE-LAM archive, hosted at ECMWF.

TIGGE-LAM domains



A common overlap region for the 7 systems hardly exists!

➔ choose a verification domain (**45.5-56N, 3-17E**) covered by 5 systems (4 conv param, 1 conv permitting).

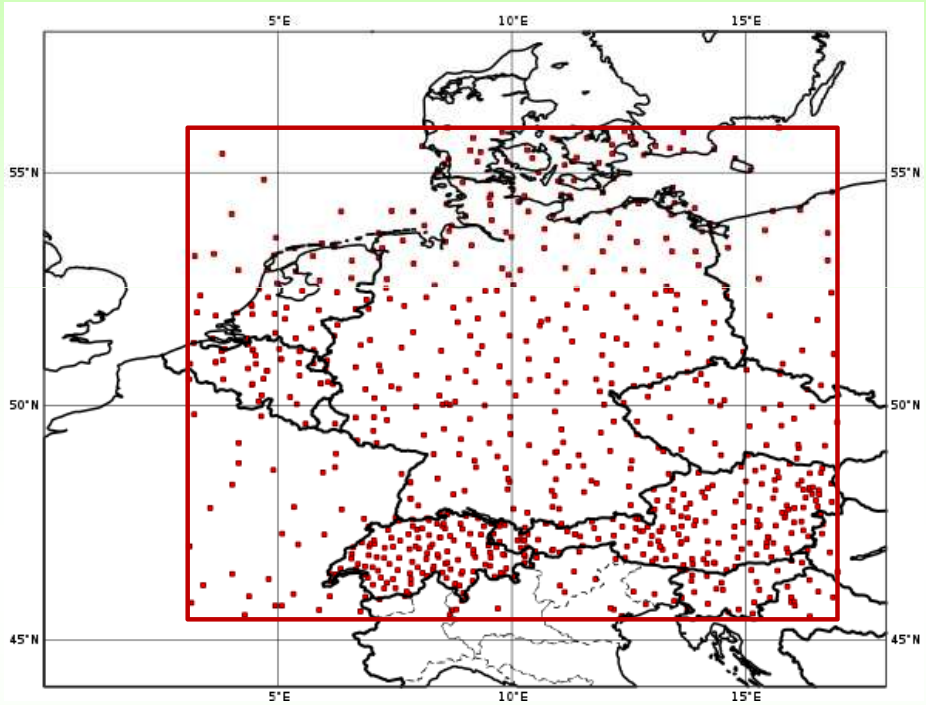
TIGGE-LAM data providers

(more info under https://software.ecmwf.int/wiki/display/TIGL/Home)

System name (organisation, country)	Ensemble size	Resolution	Forecast length (h)	Boundary conditions	Model runs (UTC)
ALADIN-LAEF (ZAMG, Austria)	**16+1**	**~15 km x 37 ML**	**72**	**ECMWF ENS**	**00,12**
ALADIN-HUNEPS (HMS, Hungary)	**10+1**	**~11 km x 49 ML**	**60**	**M-F PEARP**	**18**
COSMO-DE-EPS (DWD, Germany)	**20+0**	**~2.8 km x 50 ML**	**27**	**GFS, IFS, ICON, GSM**	**00,06,12,18**
COSMO-LEPS (ARPA-ER for COSMO, Italy)	**16+0**	**~7 km x 40 ML**	**132**	**ECMWF ENS**	**00,12**
PEARP (M-F, France)	**34+1**	**~25 km x 90 ML**	**54**	**M-F PEARP**	**06,18**
DMI-HIRLAM (DMI, Denmark)	24+1	~5.5 km x 40 ML	64	ECMWF ENS	00,06,12,18
MOGREPS (UKMO, UK)	11+1	~2.2 km x 70 ML	36	MOGREPS global	03,09,15,21

4 convection parameterised, 1 convection permitting

Evaluation of TIGGE-LAM systems



- **COSMO-DE-EPS (20 members, 2.8 km)**
- **COSMO-LEPS (16 members, 7 km)**
- **ALADIN-LAEF (17 members, 15 km)**
- **ALADIN-HUN (11 members, 11 km)**
- **PEARP (35 members, 25km)**

variables: 6h cumulated **precipitation** (00-06, 06-12, 12-18, 18-24UTC) and 2-metre **temperature**;

period : 1 September 2014 to 30 November 2014;

region: 45.5-56N, 3E-17E,

method: nearest grid point (T2m forecasts are corrected according to the height difference between model grid-point and station);

obs: synop reports (about 722/day);

forecasts: from fc+0h to fc+72h;

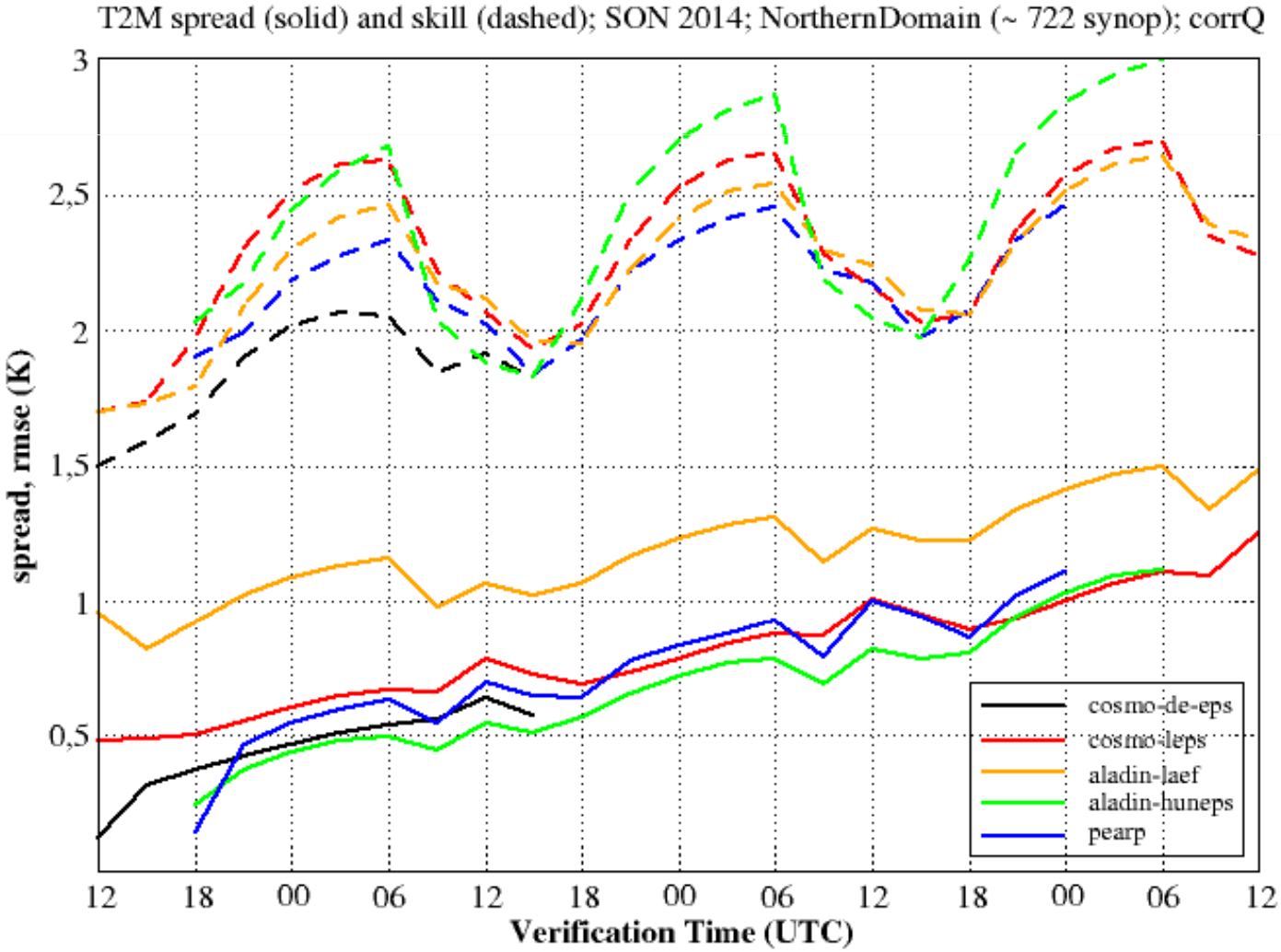
thresholds: 1, 5, 10, 15, 25, 50 mm/6h;

Scores: ROC area, BSS, RPSS, Outliers, spread/skill, bias,...

T2m: spread-skill for the individual systems

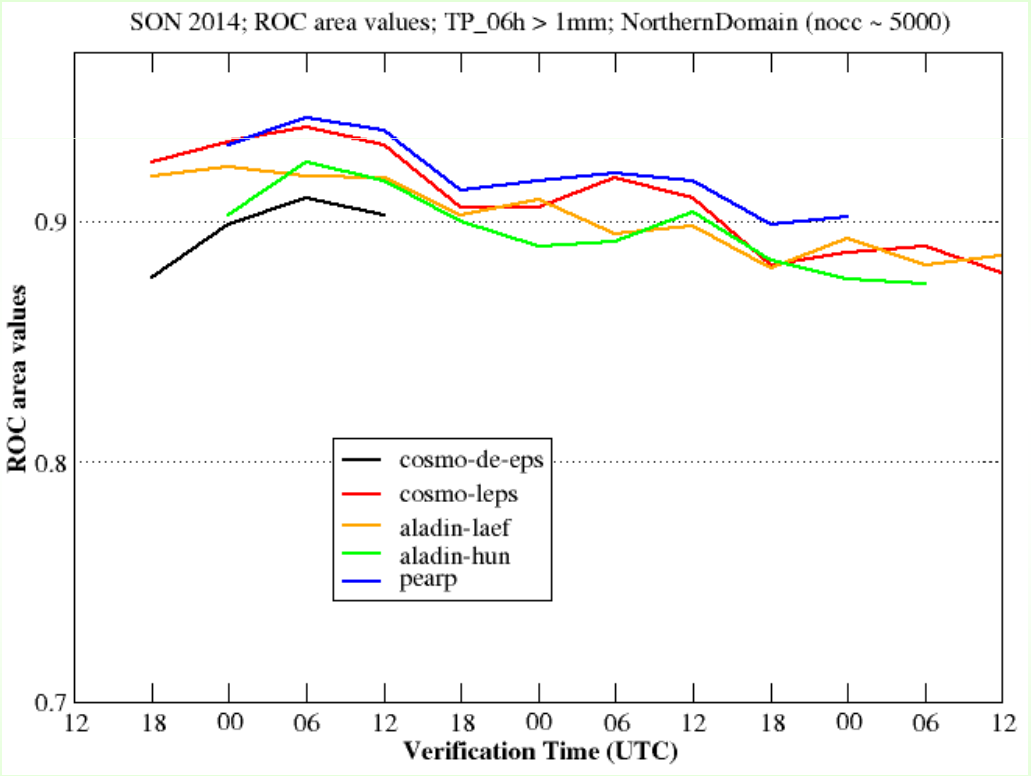
- On average, the spread among the ensemble members should match the skill of the ensemble mean.
- Large spread → lower predictability → larger ensemble–mean errors.

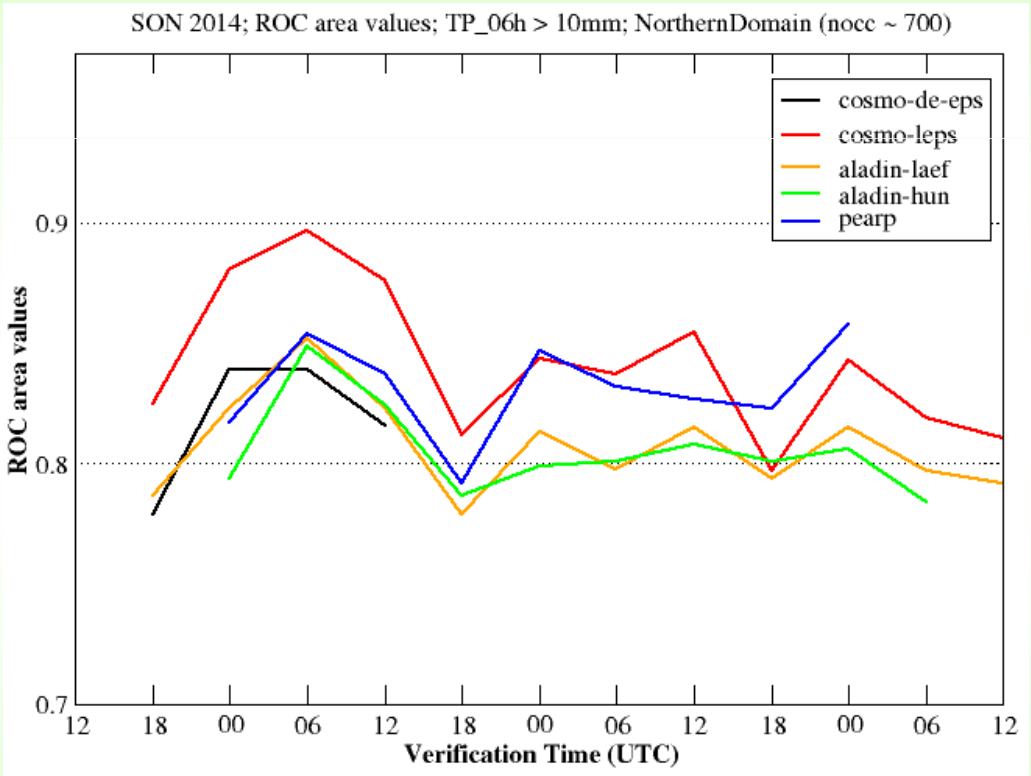
- Added value of high-resolution (lower errors in **COSMO-DE-EPS**).
- All systems are under-dispersive (about one half of what "should" be); **ALADIN-LAEF** is slightly more dispersive than the others.
- Daily cycle of rmse errors (larger errors in the morning) are very similar for all systems and only partly followed by spread behaviour.



TotPrec_6h: ROC area values

- Area under the curve in the HIT rate vs FAR diagram; the higher, the better …
- Valuable forecast systems have ROC area values > 0.6.
- **Consider two events: 6-hour precipitation exceeding 1 and 10 mm.**

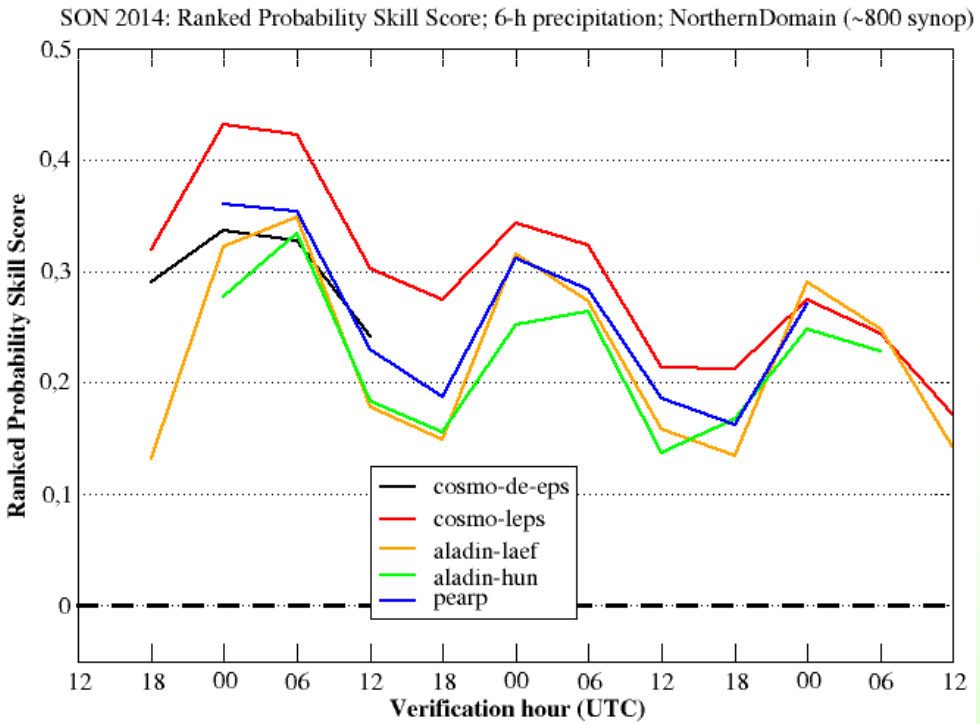


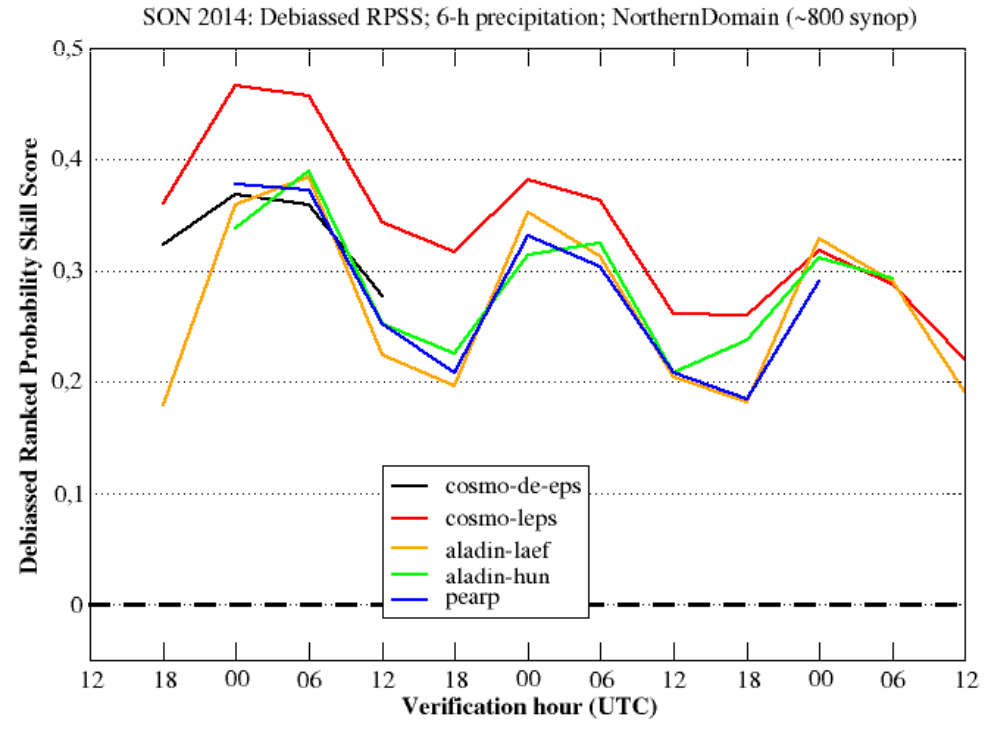


- Good performance by all systems (above 0.8) for both thresholds.
- For the lower threshold, good results by **PEARP**, despite the lower resolution.
- For the 10 mm threshold, **COSMO-LEPS** outperforms the other systems in the short range.

TotPrec_6h: Ranked Probability Skill Score

- RPSS: it is a sort of BSS "cumulated" over all thresholds. RPSS is written as 1-RPS/RPS_{ref}. **Sample climate** is the reference system. RPS is the extension of the Brier Score to the multi-event situation.
- RPSS depends on the ensemble size N and penalises small ensemble sizes.
- Consider also debiased RPSS: **$RPSS_D = 1 - (RPS/(RPS_{ref} + RPS_{ref}/N))$**; useful systems have RPSS > 0.





- In either cases, good performance of **COSMO-based** ensembles.
- Daily cycle of the score is evident for all systems, despite initialisation, perturbations, nesting strategy.
- Higher skill of the systems at predicting night-time precipitation.

Combination of TIGGE-LAM systems

- Reinterpolate fields on a common 0.1x0.1 regular lat/lon grid (do NOT include COSMO-DE-EPS).
- Generate a large-size (varying with forecast range) multi-model ensemble system.

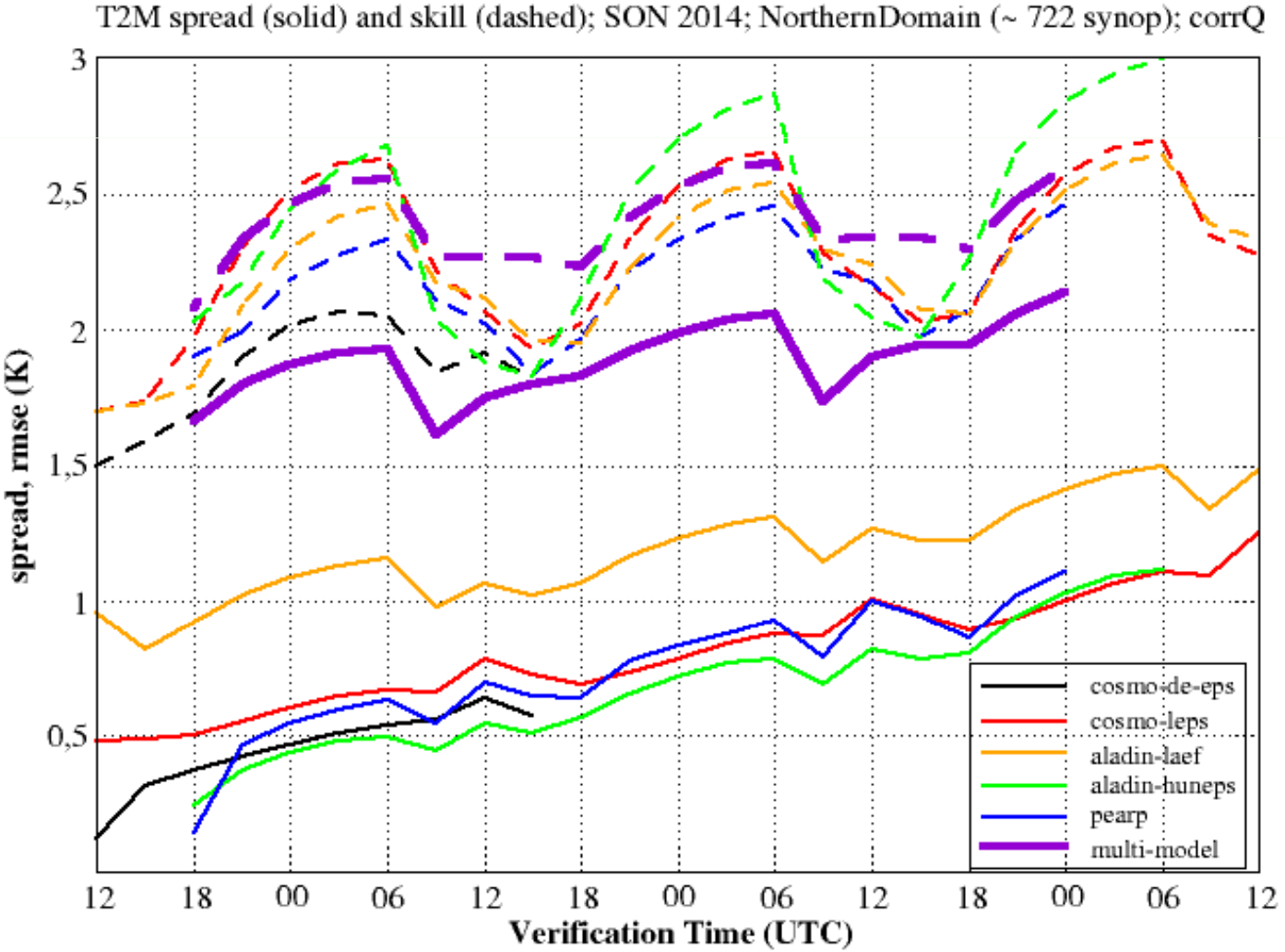
- *COSMO-DE-EPS (20 members, 2.8 km)*
- **COSMO-LEPS (16 members, 7 km)**
- **ALADIN-LAEF (17 members, 15 km)**
- **ALADIN-HUN (11 members, 11 km)**
- **PEARP (35 members, 25km)**
- **MultiModel (up to 79 members, ~10 km)**

T2m: spread-skill (MultiModel)

- On average, the spread among the ensemble members should match the skill of the ensemble mean.
- Large spread → lower predictability → larger ensemble–mean errors.

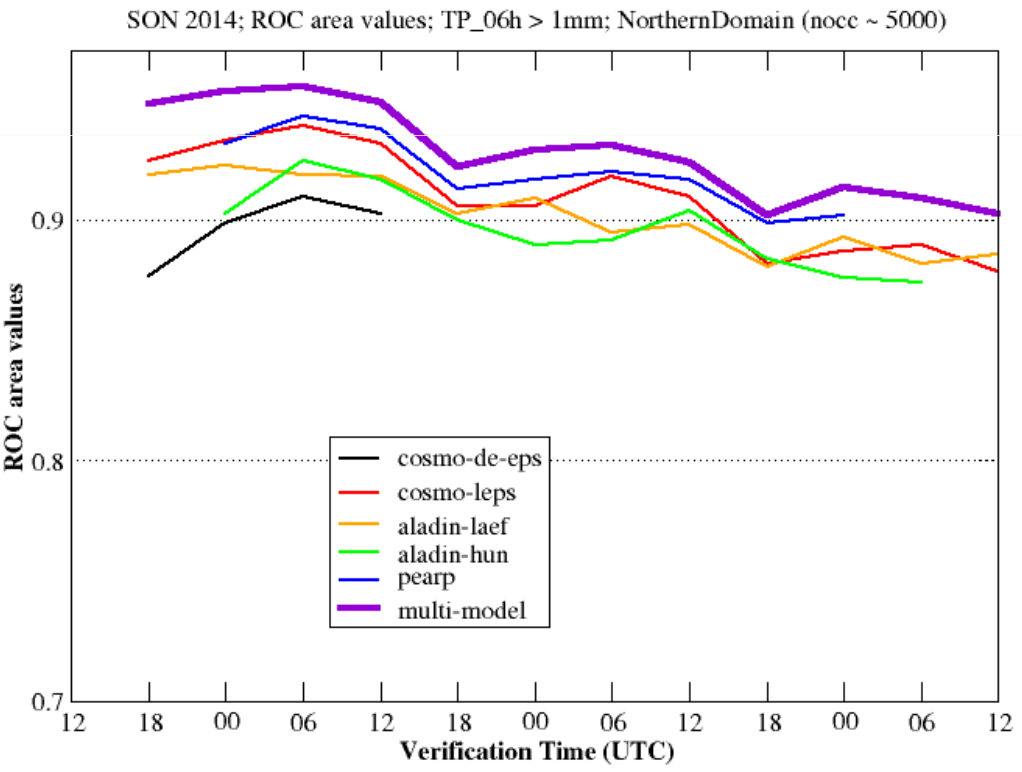
In the **multi-model** ensemble:

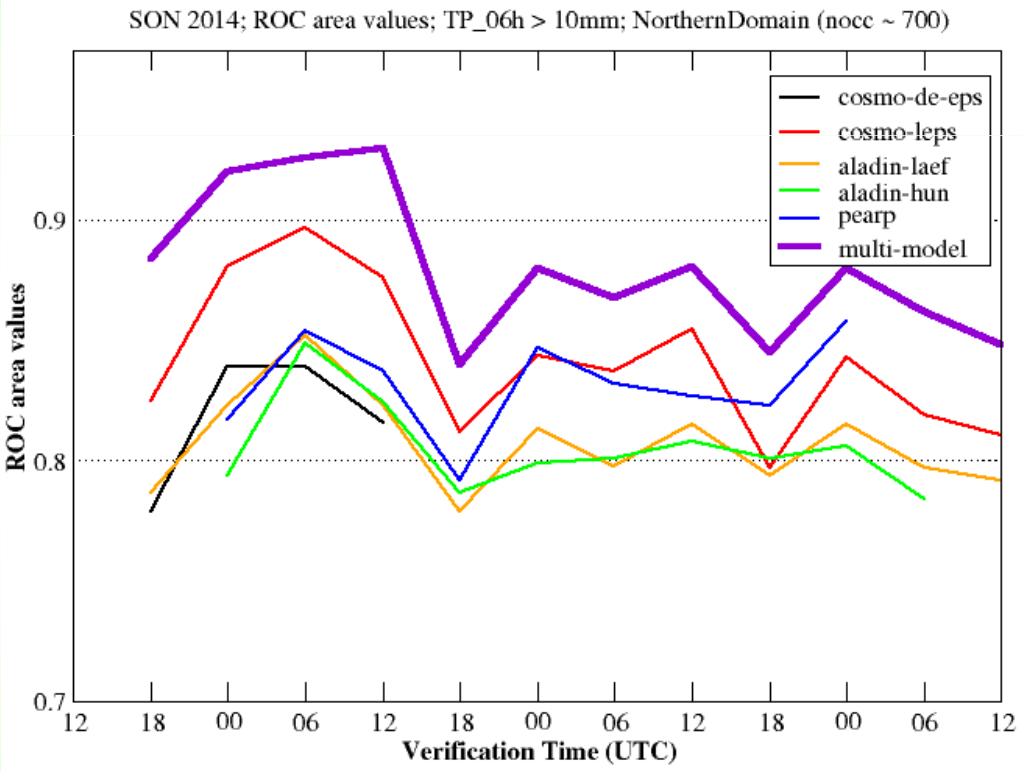
- clear increase of ensemble spread for all forecast ranges without great loss of predictability,
- the spread-skill relation is almost correct,
- the daily cycle of rmse errors is better followed by spread behaviour.



TotPrec_6h: ROC area values (MultiModel)

- Area under the curve in the HIT rate vs FAR diagram; the higher, the better …
- Valuable forecast systems have ROC area values > 0.6.
- **Consider two events: 6-hour precipitation exceeding 1 and 10 mm.**



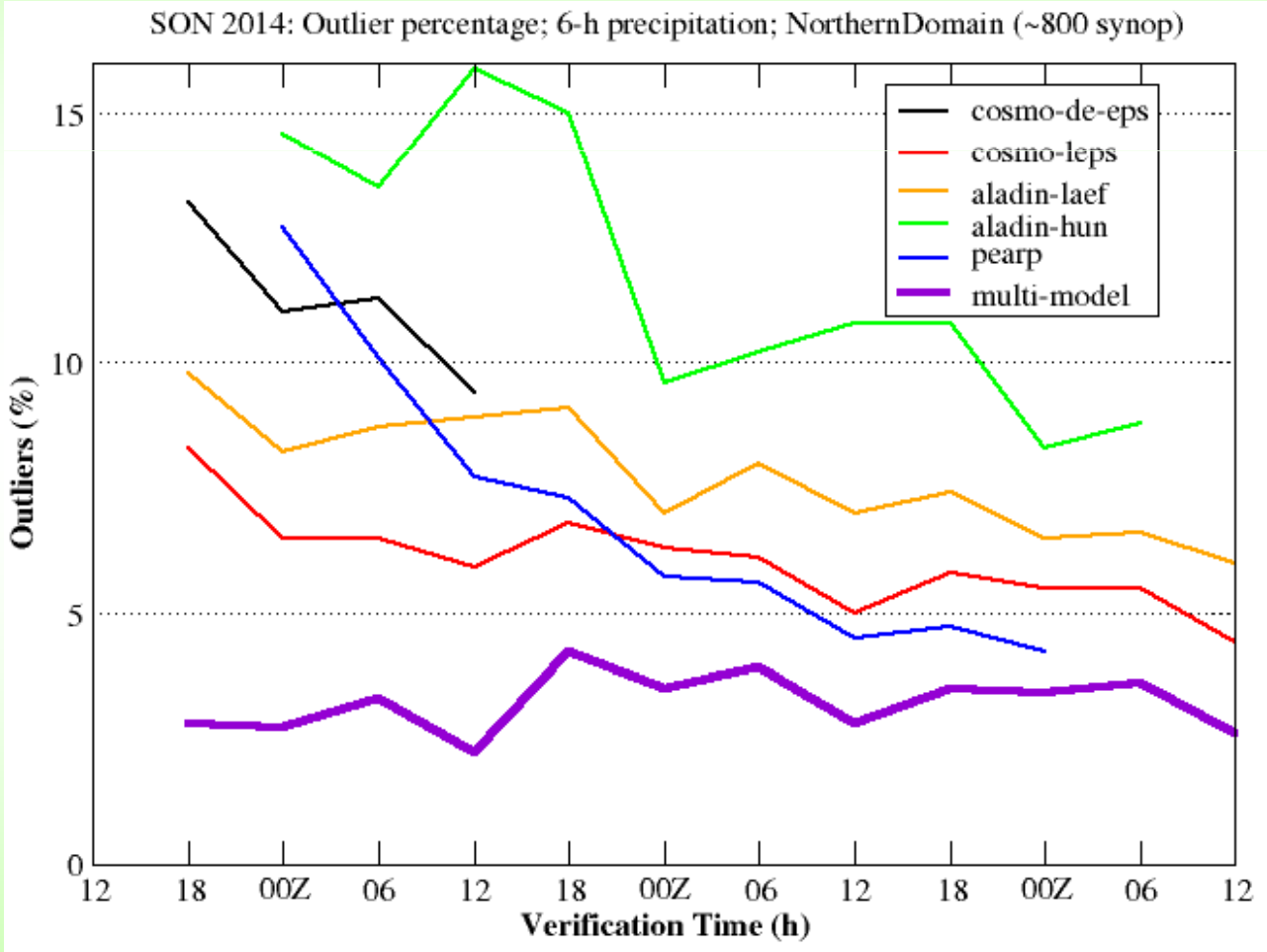


- Positive impact of the **multi-model** for all forecast ranges.
- The added value turns out to be more evident for the higher threshold.
- The same results are confirmed also by other scores (RPSS, Outliers, ...)

Outliers (MultiModel)

- How many times the analysis is out of the forecast interval spanned by the ensemble members.
- … the lower the better …

- Very different behaviour by the individual ensembles (related to ensemble size, perturbation strategy).
- Lowest percentages by **COSMO-LEPS** and **PEARP**.
- Very clear added value of the **multi-model** ensemble, especially in the short range.



Conclusions and plans

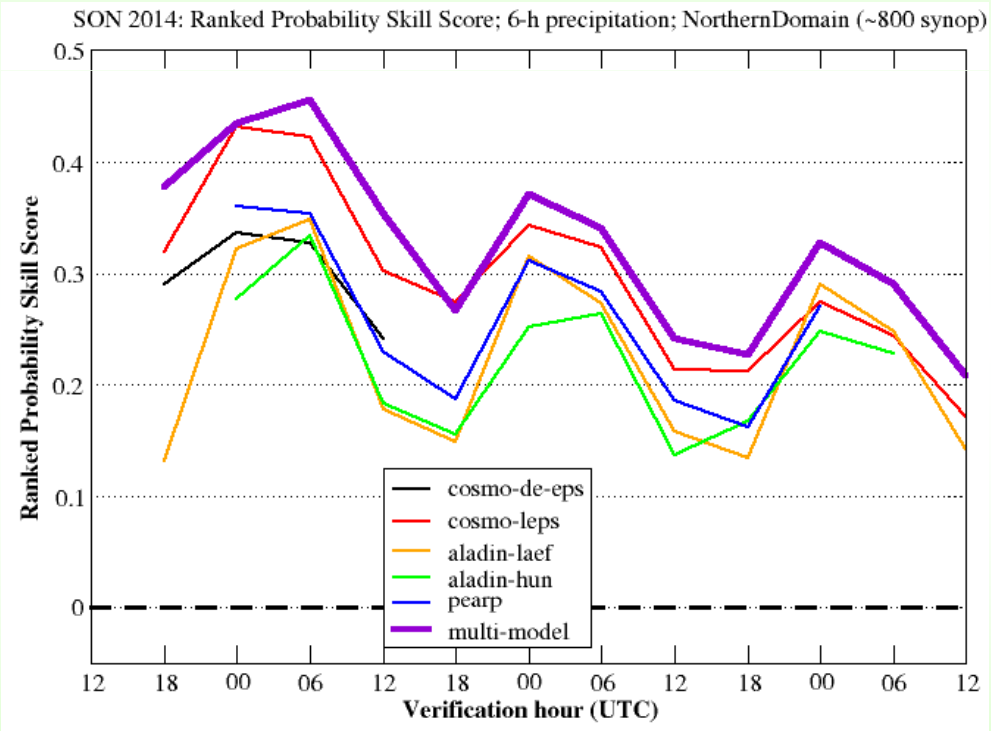
- Access to TIGGE-LAM archive is free (!), fast and simple.
- Great potential of TIGGE-LAM archive for case-study investigations and research purposes.
- Verification of 2-metre temperature:
  - **lack of ensemble spread for all systems;** added value of higher resolution.
- Probabilistic verification of 6-hour precipitation:
  - good performance of COSMO-based and PEARP ensembles,
- **Positive impact of a multi-model approach** on several probabilistic scores for both temperature and precipitation (more evident for heavier precipitation events and short ranges).
- Calibrate the individual systems before combination, assess the statistical significance of the results, explore the availability of high-resolution verification networks, compare against TIGGE global, ...

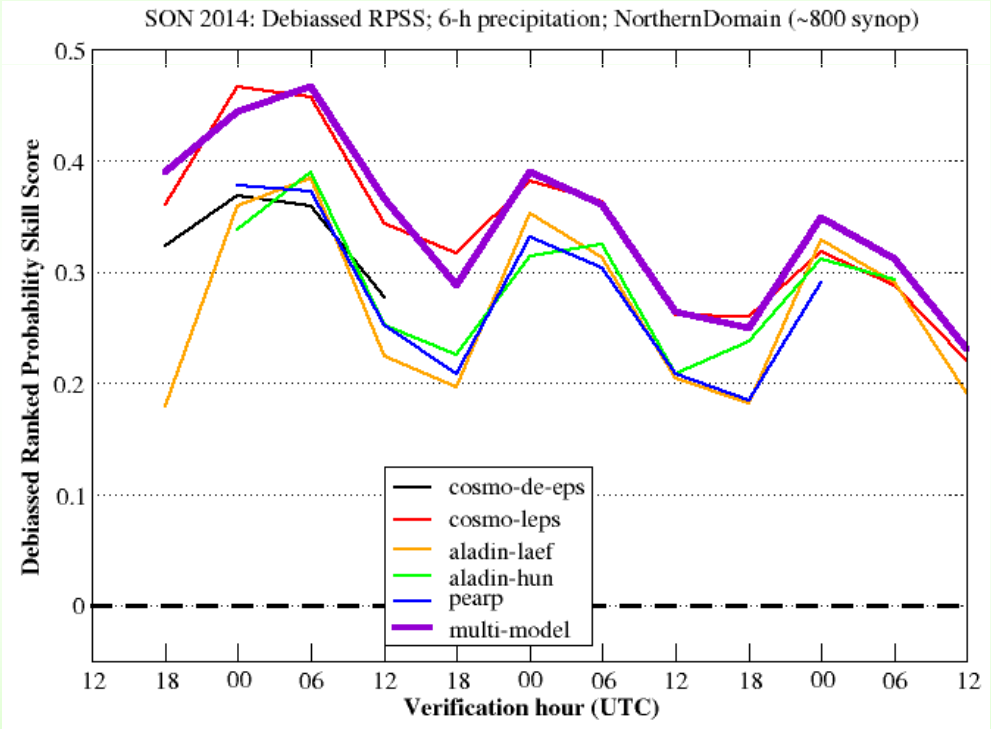
Thanks for your attention !

Extra slides

TotPrec_6h: Ranked Probability Skill Score (MultiModel)

- RPSS: it is a sort of BSS "cumulated" over all thresholds. RPSS is written as 1-RPS/RPS_{ref}. **Sample climate** is the reference system. RPS is the extension of the Brier Score to the multi-event situation.
- RPSS depends on the ensemble size N and penalises small ensemble sizes.
- Consider debiased RPSS: $RPSS_D = 1 - (RPS/(RPS_{ref} + RPS_{ref}/N))$; useful systems have for RPSS > 0.





- Higher skill of the **multi-model** ensemble is less marked, but still evident at all forecast ranges.

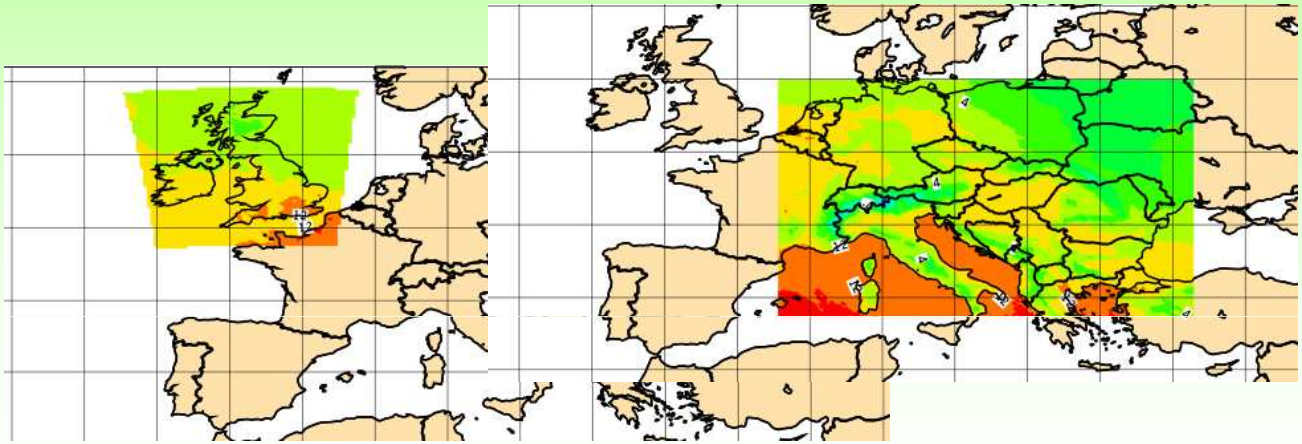
TIGGE-LAM data providers

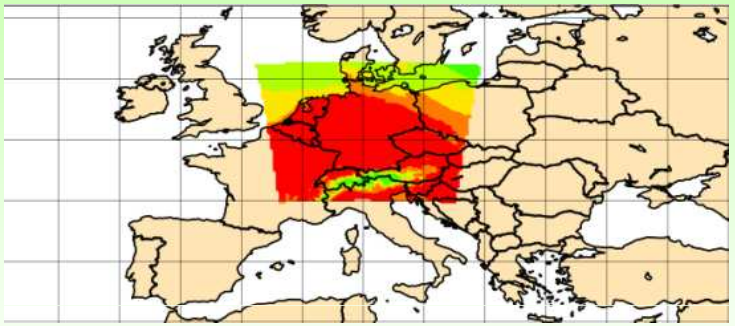
(more details under https://software.ecmwf.int/wiki/display/TIGL/Home)

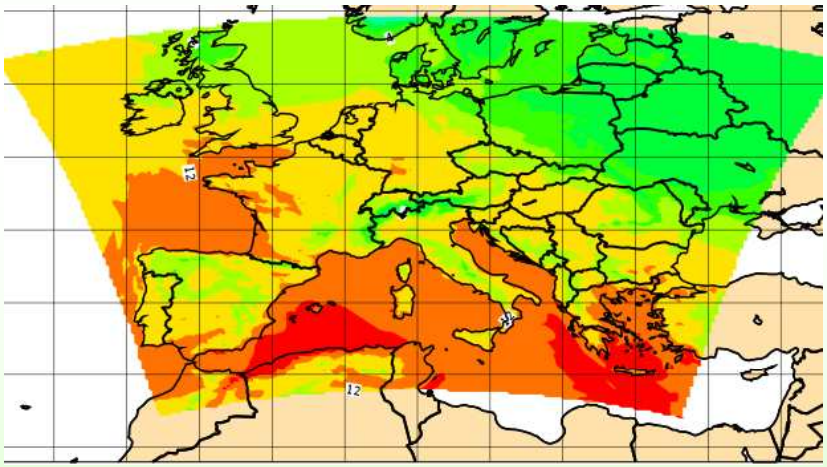
System name (organisation, country)	Ensemble size	Resolution	Forecast length (h)	Boundary conditions	Model runs (UTC)
ALADIN-LAEF (ZAMG, Austria)	**16+1**	**~15 km x 37 ML**	**72**	**ECMWF ENS**	**00,12**
ALADIN-HUNEPS (HMS, Hungary)	**10+1**	**~11 km x 49 ML**	**60**	**M-F PEARP**	**18**
COSMO-DE-EPS (DWD, Germany)	**20+0**	**~2.8 km x 50 ML**	**27**	**GFS, IFS, ICON, GSM**	**00,06,12,18**
COSMO-LEPS (ARPA-ER for COSMO, Italy)	**16+0**	**~7 km x 40 ML**	**132**	**ECMWF ENS**	**00,12**
PEARP (M-F, France)	**34+1**	**~25 km x 90 ML**	**54**	**M-F PEARP**	**06,18**
DMI-HIRLAM (DMI, Denmark)	24+1	~5.5 km x 40 ML	64	ECMWF ENS	00,06,12,18
MOGREPS (UKMO, UK)	11+1	~2.2 km x 70 ML	36	MOGREPS global	03,09,15,21

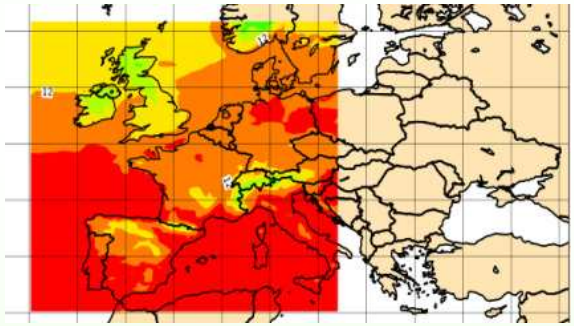
5 convection parameterised, 2 convection permitting

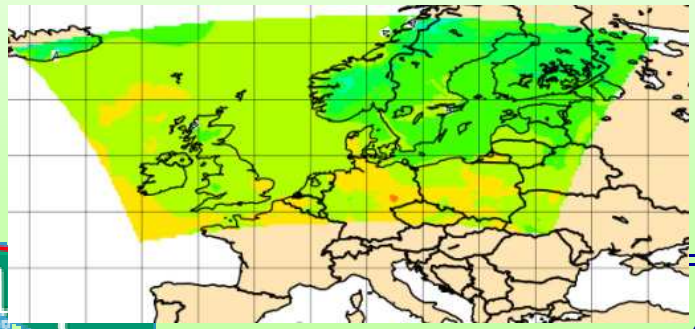
About the different domains

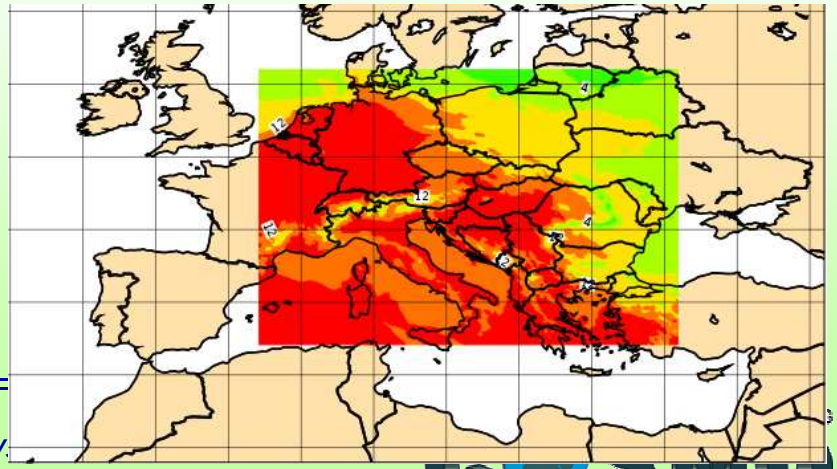




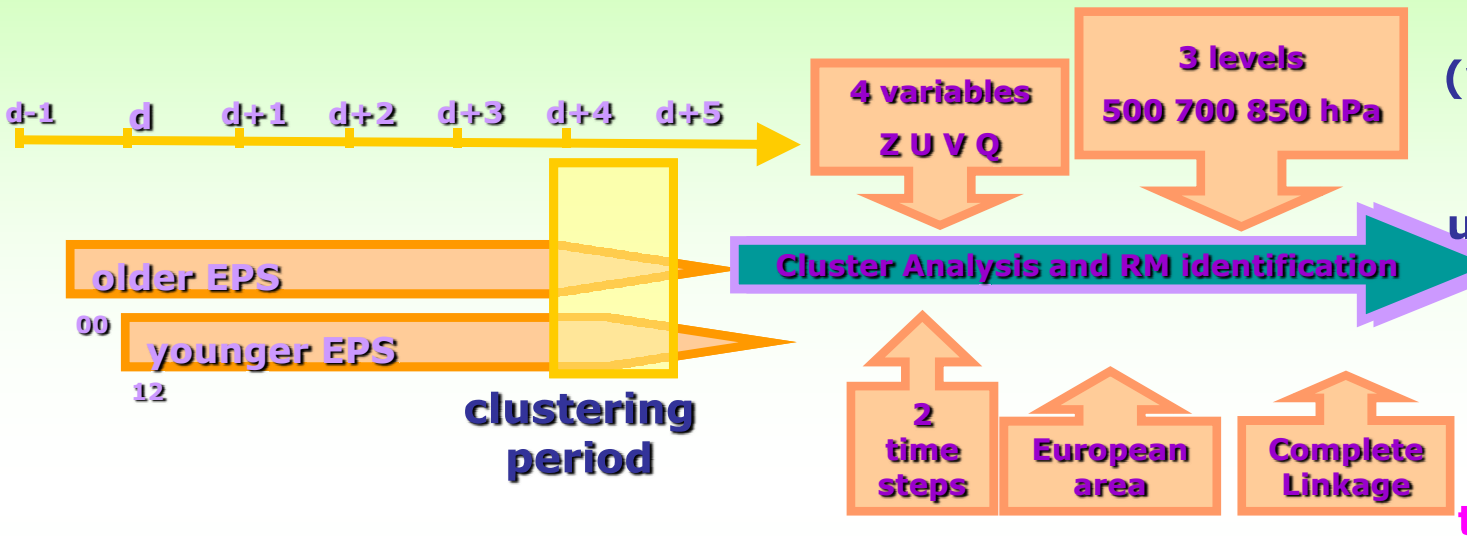








COSMO-LEPS suite @ ECMWF: present status



16 Representative Members driving the 16 COSMO integrations (weighted according to the cluster populations)

using either Tiedtke or IFS-Bechtold convection scheme (members 1-8 T, members 9-16 IFS)

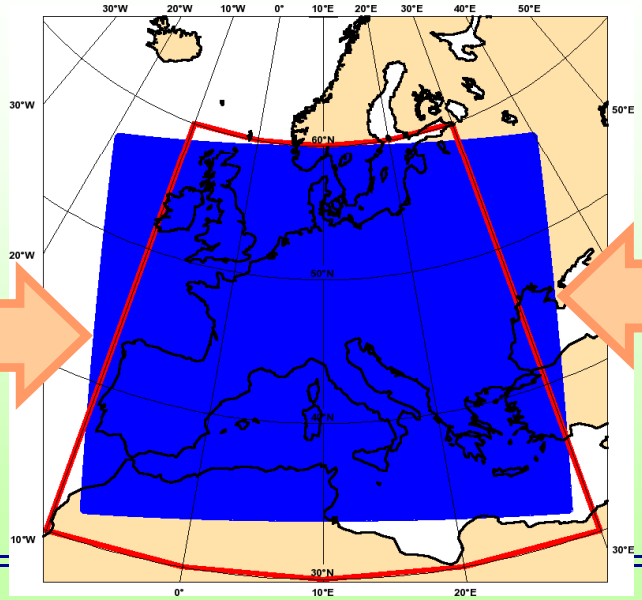
+

perturbations in turbulence scheme and in physical parameterisations

+

soil initial conditions from COSMO-EU

COSMO-LEPS clustering area

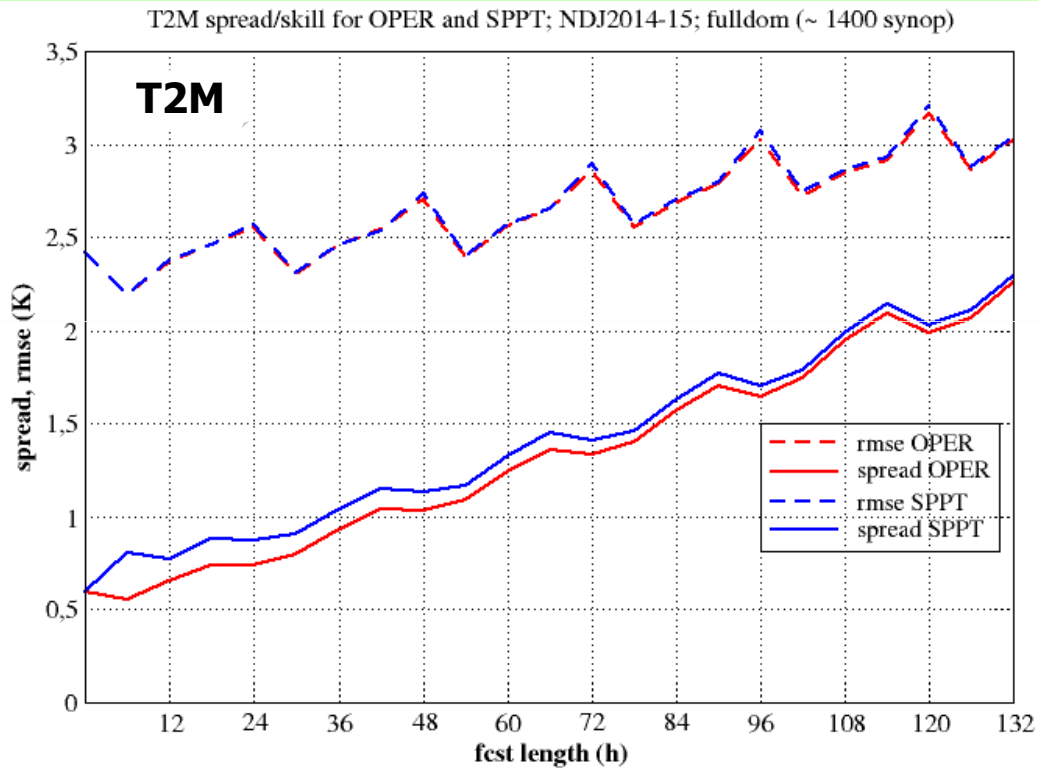


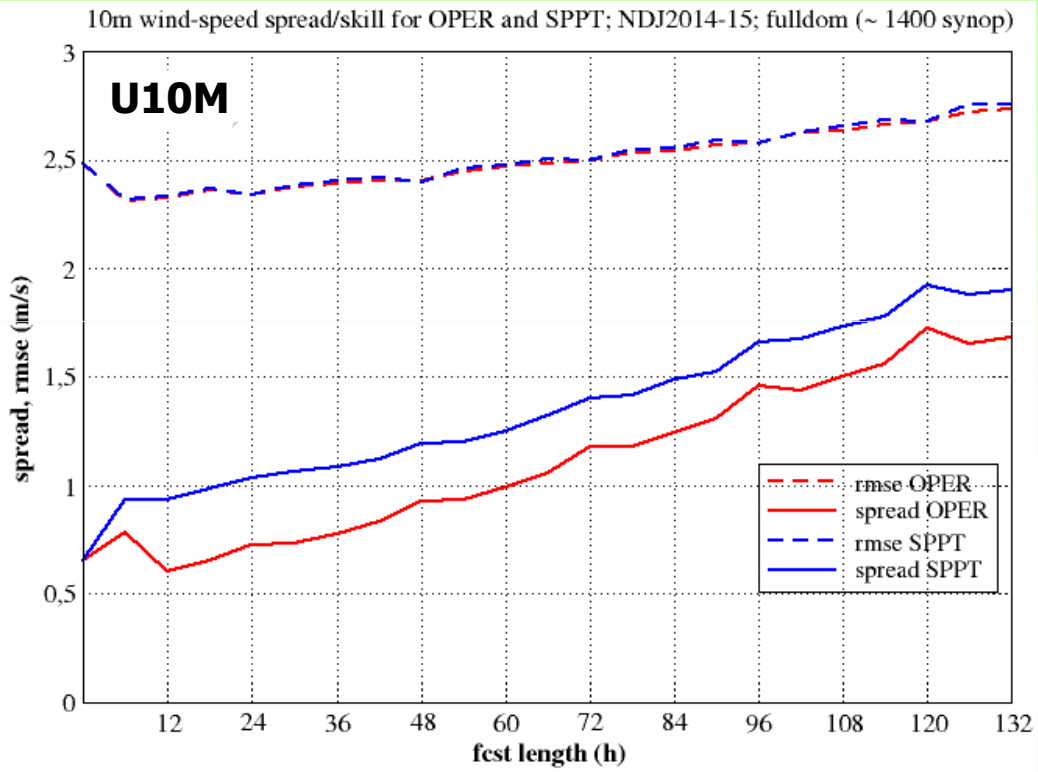
COSMO-LEPS Integration Domain

- **suite runs as a "time-critical application" managed by ARPA-SIMC; runs at both 00 and 12TC;**
- **Δx ~ 7 km; 40 ML; fc+132h;**
- **COSM0 v5.0 since Feb 2014;**
- **computer time (50 million BUs for 2015) provided by the COSMO partners which are ECMWF member states.**

A.Montani

SPPT: spread/skill for T2m and WSPEED10m

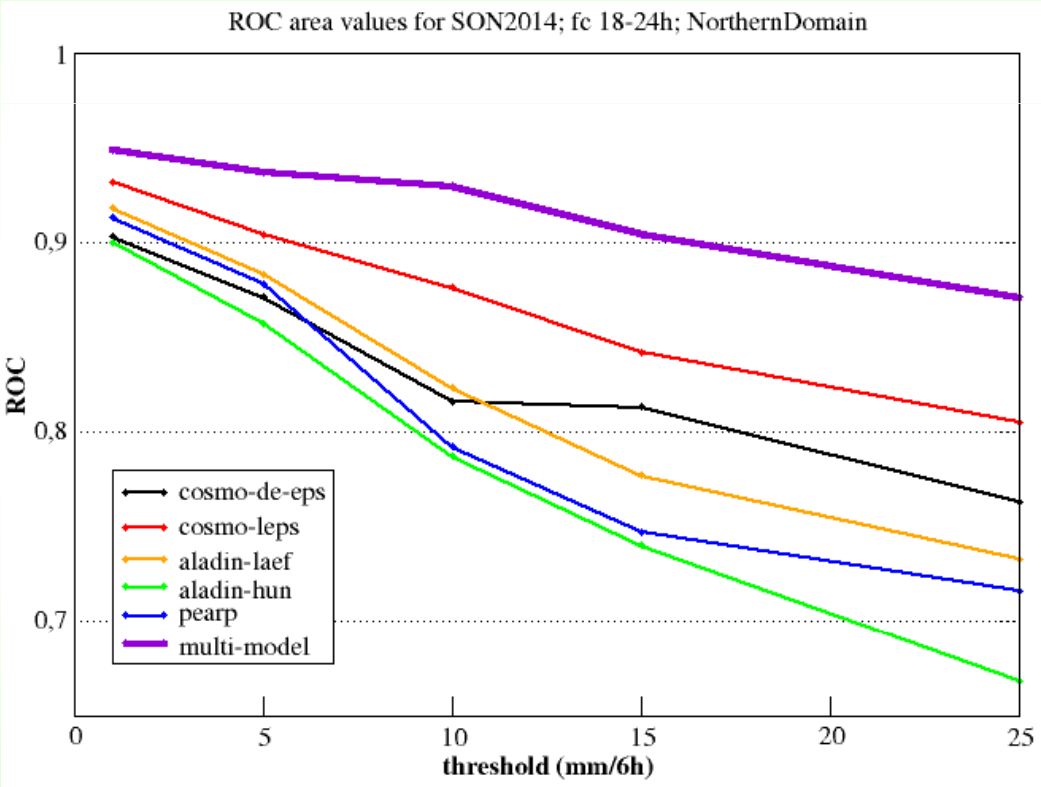


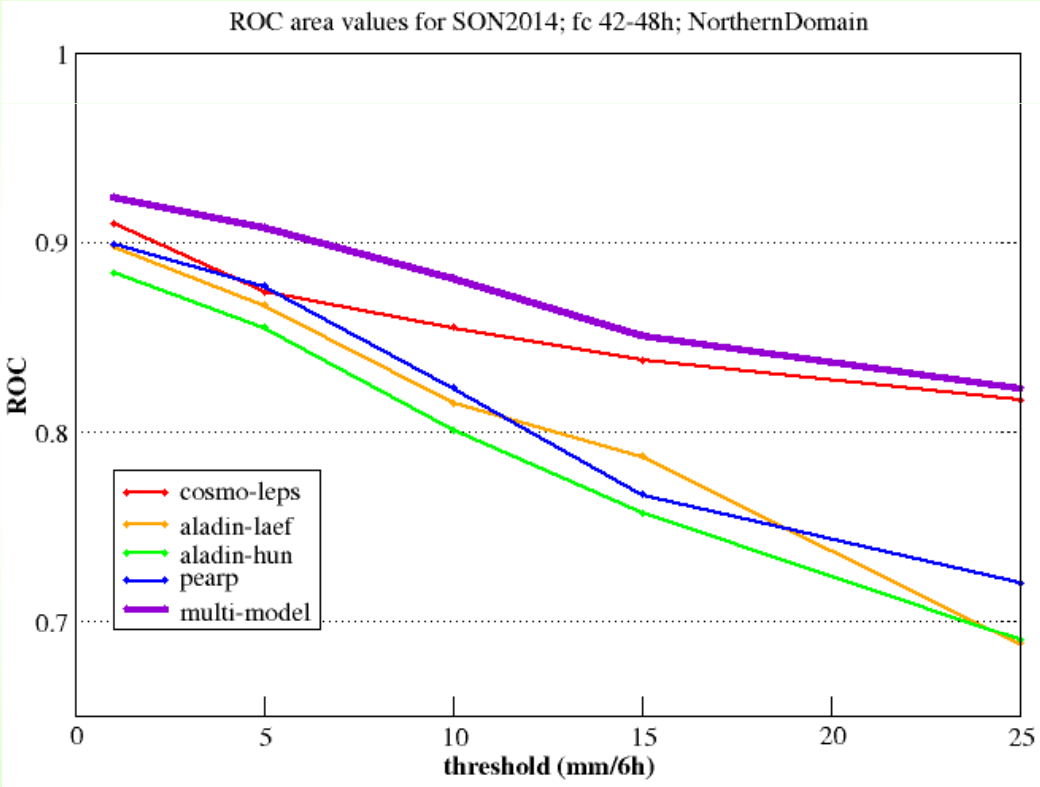


- Larger spread for COSMO-LEPS **with SPPT**, especially for wind-speed.
- In either cases, lack of spread in the short range.
- Limited impact (if any) on forecast skill of the ensemble mean.

TotPrec_6h: ROC area values vs threshold (MultiModel)

- **Fixed fcst ranges** (18-24h and 42-48h): consider the performance of the system for increasing thresholds.
- Need to take into account the different statistics for the different events: fewer observations are recorded (5000 → 90) as the threshold value increases.





- For low thresholds, similar skill for all systems (good performance by **COSMO-LEPS**).
- Positive impact of the **multi-model** is evident for all thresholds and especially in the short range.

TIGGE-LAM ensemble datasets for the prediction of heavy precipitation events: first results at ARPA-SIMC

Andrea Montani, C. Marsigli, T. Paccagnella

ARPA Emilia-Romagna Servizio IdroMeteoClima, Bologna, Italy

EWGLAM Meeting
Belgrade, 5 October 2015

arpa agenzia regionale prevenzione e ambiente dell'emilia-romagna

CONSORTIUM FOR SMALL SCALE MODELING COSMO