Experiments with stochastic perturbation of physical tendencies in COSMO-Ru2-EPS

DMITRY ALFEROV AND ELENA ASTAKHOVA

Hydrometcenter of Russia, Roshydromet, Moscow, Russia

Abstract

The experiments with the scheme of stochastic perturbation of physical tendencies (SPPT) were carried out using the COSMO-Ru2-EPS ensemble prediction system. Several SPPT settings were tested. Both case studies and probabilistic verifications of forecast monthly series were performed.

It was found that SPPT could be useful for precipitation forecasts improving the description of the rain location and start, increasing the ensemble spread in the areas of uncertain forecasts, and slightly improving the probabilistic scores.

SPPT does not add value to 2-m temperature forecasts but results in a better description of the 2-m temperature distribution. It is possible to improve the skill of temperature forecasts by varying the SPPT settings.

1 Introduction

Ensemble forecasting is a common method for predicting the future state of the atmosphere and the probability of this state. The well-known problem of ensembles is their insufficient spread.

The RMSE of prognostic realizations with respect to the ensemble mean (the ensemble spread) and the RMSE of the ensemble mean with respect to observations should demonstrate a similar growth with forecast lead-time, but it is often not so.

To increase the ensemble spread and to get its adequate growth in time, it is necessary to allow for forecast uncertainties following not only from errors in our knowledge of the initial atmospheric state (that is, from possible errors in initial and lateral boundary conditions) but also from the model imperfections as well as from errors in surface boundary conditions.

In this paper, we examine how the implementation of the scheme of stochastic perturbation of physical tendencies (SPPT) to the COSMO-Ru2-EPS system affected the ensemble spread and performance.

2 Experiment setup

In our experiments, we used the COSMO-Ru2-EPS system that had been previously developed within the framework of the CORSO Priority project (Rivin, Rozinkina, 2011). The system provided a dynamical downscaling of COSMO-S14-EPS, the Italian ensemble prediction system for the Sochi-2014 Olympics.

In turn, COSMO-S14-EPS was a clone of COSMO-LEPS (Montani et al., 2011) moved to the Sochi region. The systems are sketched in Fig. 1 and described in detail in (Montani et al., 2013, 2014).

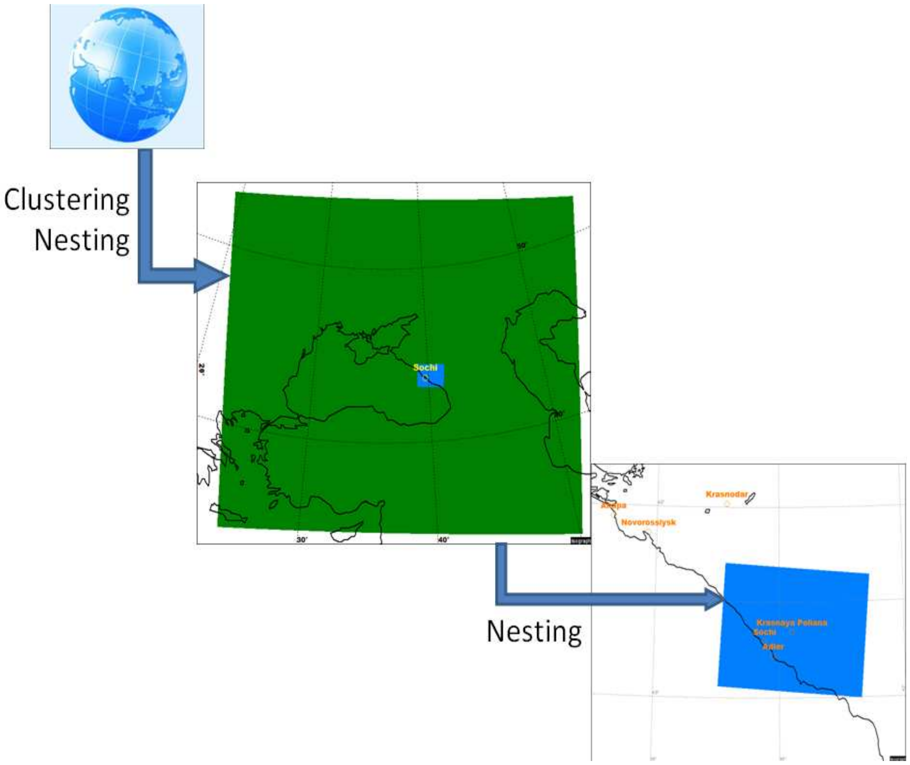


Figure 1: Ensemble nesting for Sochi. The integration domains for COSMO-S14-EPS and COSMO-Ru2-EPS are colored blue.

Both COSMO-S14-EPS and COSMO-Ru2-EPS ran operationally during the Olympic Games 2014 providing probabilistic products to Sochi forecasters. All observations and forecasts issued during the Olympics are stored in a special TIGGE-LAM styled archive (Astakhova et al., 2016) thus facilitating further research.

In this study we extracted the operational COSMO-Ru2-EPS forecasts for February 2014 starting at 00 and 12 UTC from the archive and used them as a reference experiment hereafter referred to as noSPPT. Some details of the operational runs are summarized in Table 1.

Table 1: COSMO-Ru2-EPS settings for the operational Olympic runs (**noSPPT** experiment)

Model	COSMO model version 4.22
Forecast area	Sochi region (see Fig. 1)
Grid step	2.2 km
Number of levels	50
Initial& boundary conditions	Taken from COSMO-S14-EPS (COSMO-LEP relocated to the Sochi region; see Fig.1)
Membership	10
Forecast length	48h
Output time step	1h
Physical perturbations	No perturbations (no SPPT scheme included)

After the Olympic Games, additional experiments were carried out with COSMO-Ru2-EPS with the aim to test the SPPT scheme and to assess its effect on the forecast spread and skill. The model resolution, the integration domain, the forecast length, the ensemble size, as well as initial and boundary conditions were the same as in the reference experiment **noSPPT**.The period from February 1 to February 28, 2014 was considered.

The SPPT scheme (Buizza et al., 1999) has been implemented to the COSMO model v.5.1. However, due to the courtesy of L. Torrisi and C. Schraff, who provided the necessary software, we could start the experiments prior to the official release of version 5.1. Therefore, the first experiments with the SPPT scheme at Roshydromet were performed with version 5.0 of the COSMO model complemented by some additional modules. Later, after the SPPT scheme had been introduced to the official COSMO code and model version 5.1 had been released, we changed to this version in our experiments. Have in mind that version 5.1 didn't differ much from version 5.0 with additional modules.

There are several parameters in the SPPT scheme that govern the perturbation size and their spatiotemporal correlations. A full description of SPPT settings can be found in COSMO User's Guide (Schaettler et al., 2014). The goal of our experiments was not only to test SPPT with its recommended parameters but also to understand to which degree the variations of these parameters (the SPPT setting) influence the results. We tried the following parameters, defining several aspects of random number field generation:

- the random number coarse grid distances **dlat_rn** and **dlon_rn**;
- the type of distribution of random numbers **lgauss_rn**;
- the standard deviation of the Gaussian distribution of random numbers **stdv_rn**;
- the upper limit imposed to the absolute value of random numbers **range_rn**;
- the parameter showing whether the random numbers are interpolated in space **lhorint_rn** and time **ltimeint_rn**;
- number of random number patterns with different correlation scales **npattern_rn**;
- time increment for drawing new random number field **hinc_rn**.

We also tried to vary the parameter **itype_qxpert_rn**, showing which hydrometeor tendencies are perturbed, and the parameter **itype_qxlim_rn**, determining the type of reduction/removal of the perturbation in case of negative or supersaturated values of specific water vapor content or negative other water-content related characteristics.

The list of experiments and the corresponding SPPT settings are given in Fig.2. COSMO model v. 5.1 was used in all experiments except for the experiment **SPPTtest** which was run with COSMO model v.5.0. Note that the reference experiment **noSPPT** was based on COSMO model v.4.22.

Both case studies and verification of monthly series of forecasts were carried out. The results are presented in the next sections.

3 Case studies

The main attention was given to the ability of COSMO-Ru2-EPS to predict precipitation and 2-m temperature over the mountain area. Two cases were analyzed, both from the list of interesting events prepared by the Olympic forecasters and recommended for thorough analysis (see Astakhova et al., 2016). The results of experiments **SPPTtest** and **noSPPT** were considered.

<table>
<tr><th rowspan="2">Perturbed parameters</th><th colspan="7">Experiment name and SPPT settings</th></tr>
<tr><th>noSPPT</th><th>SPPTtest</th><th>SPPTphys</th><th>SPPTintphys</th><th>SPPT_W</th><th>SPPT_W +phys</th><th>SPPT_W +intphys</th></tr>
<tr><td>lgauss_rn</td><td rowspan="10"><i>N/A</i></td><td colspan="6"><i>.TRUE.</i>
(Gaussian distribution)</td></tr>
<tr><td>hinc_rn (hours)</td><td colspan="6"><i>6</i></td></tr>
<tr><td>dlat_rn, dlon_rn (degrees)</td><td colspan="6"><i>5</i></td></tr>
<tr><td>stdv_rn</td><td colspan="3"><i>0.4</i></td><td colspan="3"><i>1.0</i></td></tr>
<tr><td>range_rn</td><td colspan="3"><i>0.8</i></td><td colspan="3"><i>0.9</i></td></tr>
<tr><td>lhorint_rn</td><td colspan="2" rowspan="2"><i>.FALSE.</i>
(no interpolation of perturbations)</td><td rowspan="2"><i>.TRUE.</i>
(interpolated perturbations)</td><td colspan="2" rowspan="2"><i>.FALSE.</i>
(no interpolation of perturbations)</td><td rowspan="2"><i>.TRUE.</i>
(interpolated perturbations)</td></tr>
<tr><td>ltimeint_rn</td></tr>
<tr><td>itype_qxpert_rn</td><td><i>0</i>
(q^v only)</td><td colspan="5"><i>2</i>
(q^v, q^c, q^i, q^r, q^s, q^g)</td></tr>
<tr><td>itype_qxlim_rn</td><td><i>0</i>
(no limitations imposed)</td><td colspan="2"><i>1</i>
(do not perturb tt T- and tt q^v tendencies if new q^v are negative or supersaturated)</td><td><i>0</i>
(no limitations imposed)</td><td colspan="2"><i>1</i>
(do not perturb tt T- and tt q^v tendencies if new q^v are negative or supersaturated)</td></tr>
<tr><td>npattern_rn</td><td colspan="6"><i>1</i></td></tr>
</table>

Figure 2: The list of experiments and the corresponding SPPT settings.

The first case was the tropospheric Foehn event on February 7, 2014. It was characterized by higher than usual 2-m temperature with very weak diurnal variations, low humidity, and east and southeast winds at 1500–2300 m. The rise of atmospheric temperature at about 1500 m above the sea level was poorly predicted by most models from many countries participating in the FROST-2014 project (the WWRP RDP/FDP project devoted to the Sochi Olympics, Kiktev et al., 2014, 2017).

The ensemble spread fields obtained for this case with and without SPPT were compared. The interesting thing found in the difference of the spread fields was that it depended on orography. Figure 3 demonstrates the difference of 2-m temperature spread in 30-h forecasts with and without SPPT (experiments **SPPTtest** and **noSPPT**) (top) and the model orography (bottom). The correlation of the fields is obvious.

The maximum increase of the spread due to SPPT introduction was found over high mountains, the spread over low areas (including sea) was also big. Meanwhile, at middle altitudes, SPPT somewhere even decreased the ensemble spread.

The strongest increase in the ensemble spread at high altitudes along with the fact of poor temperature forecasts above 1500 m in this case can be considered as a positive effect of SPPT introduction (areas of higher spread coincided with the areas of less skillful forecast).

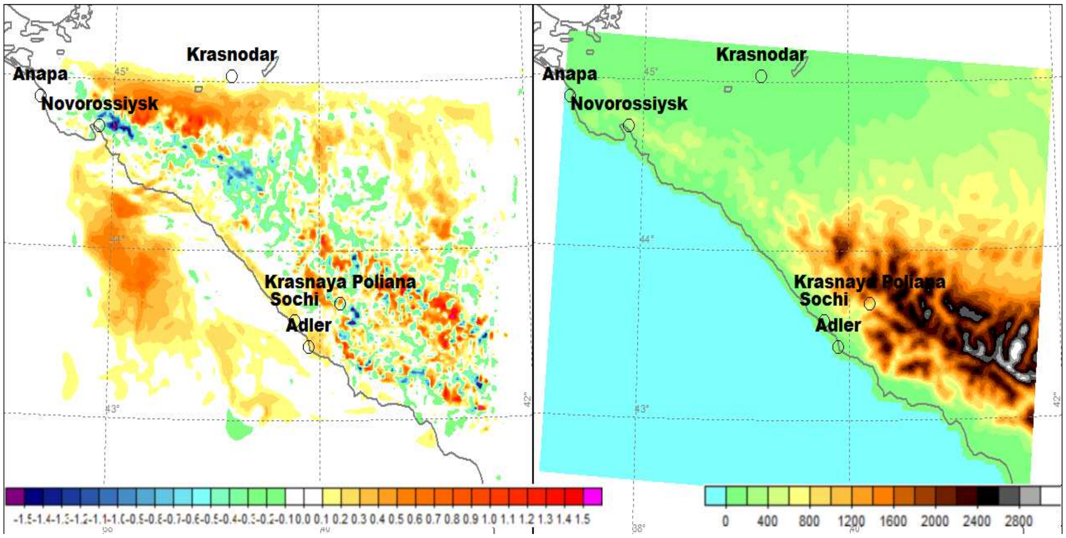


Figure 3: Left panel: The difference of 2-m temperature ensemble spread in experiments with and without SPPT (**SPPTtest** minus **noSPPT**). 30-h forecast starting at 00 UTC on February 6, 2014. Right panel: model orography.

We also considered a heavy precipitation event on February 18, 2014. The fields of predicted probabilities of the rain occurrence (rain exceeding 0.1 mm in 3 h) and of intense precipitation (more than 10 mm of rain in 3 h) in experiments **SPPTtest** and **noSPPT** were compared to METEOSAT data (not shown).

The comparison demonstrated that the system with SPPT was more skillful in predicting the time when it started raining. Also less false heavy rain areas and more actual peaks were predicted in the **SPPTtest** experiment. However, the location of maximum precipitation was better described in the **noSPPT** experiment.

4 Verification results

We used the results of **SPPTtest** and **noSPPT** experiments as well as the results of five more experiments with various SPPT settings (see Fig.2) in the verification exercise. The considered period was 1–28 February 2014. The forecasts were issued twice a day starting from 00 and 12 UTC analyses; the forecast length was 48 hours. No separation by the initial forecast time was made, thus we used a series of 56 forecasts in computations.

The verification was performed for three meteorological fields: 3-hour total precipitation sum (R_{sum}), 2-m air temperature (T_{2m}) and 10-m wind speed. The following three ensemble forecast scores were considered: the Brier score (BS), the Brier skill score (BSS) and the area under the ROC curve (ROCA). (It's worth reminding here that the perfect scores are BS=0, BSS=1, ROCA=1).

The verification was made against observations of 31 meteorological stations in the Sochi region (see Figure.

4). R-based utilities developed and kindly provided by A. Muravev were applied.

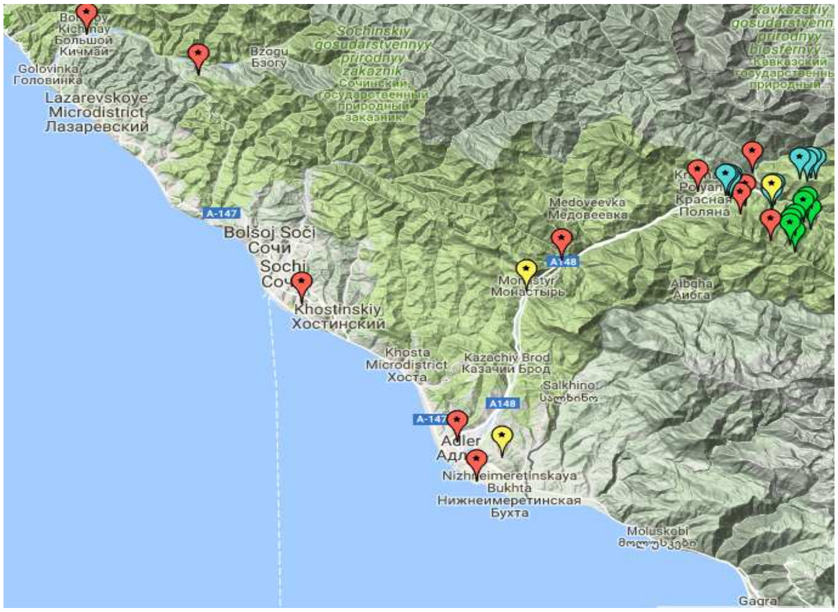


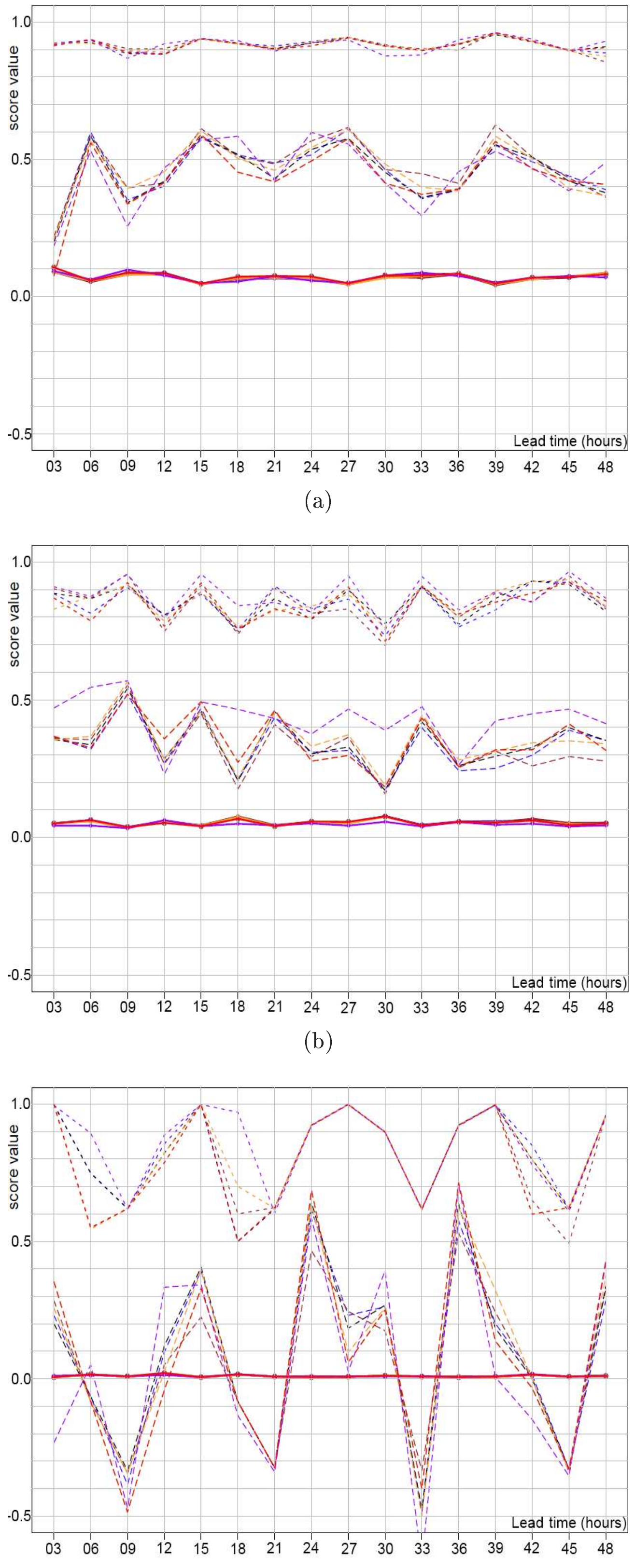
Figure 4: Stations used for verification (see the FROST-2014 project website http://frost2014.meteoinfo.ru/ for details)

.

The resulting scores are presented in Figures.5-6.

It was nice to see that the introduction of SPPT did not result in the precipitation forecast degradation. Figure 5 demonstrates BS, BSS and ROCA as functions of forecast lead-time for the events “3-h precipitation is greater than 0.1 mm/3h, 1 mm/3h, and 5 mm/3h” for all experiments listed in Fig 2.

The scores for different experiments are very close. However, for higher thresholds ($R_{sum} > 1$ mm/3h and $R_{sum} > 5$ mm/3h) the **SPPTtest** experiment gives the best results. Note that intense precipitation (R_sum $>$ 5 mm/3h) is predicted badly in all experiments (BSS is low, even below zero for some lead-times). It is probably related to insufficient statistics, such events were rather rare during the period considered.



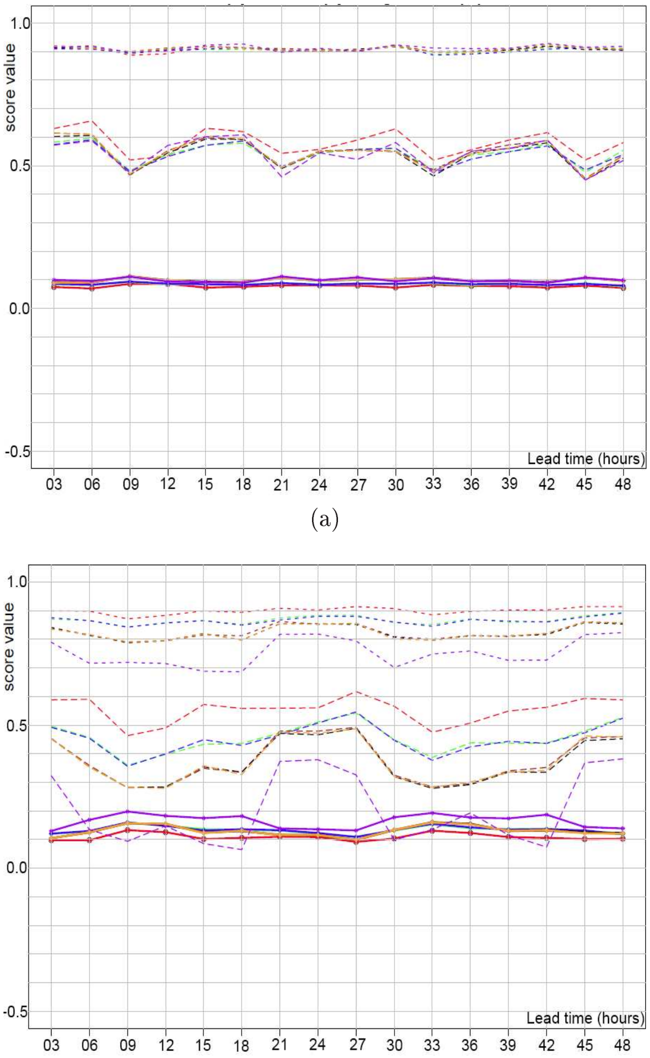
(a)

(b)

(c)

Figure 5: Verification scores as functions of forecast lead-time for the events "3-h precipitation (R_{sum}) is greater than 0.1 mm/3h (a) , 1 mm/3h (b) , and 5 mm/3h (c) " for all experiments listed in Fig.2. Solid line: BS, long-dashed line: BSS, dashed line: ROCA. Red lines: noSPPT, purple: SPPTtest, orange: SPPTphys, black: SPPTintphys, green: SPPT_W, brown: SPPT_W+phys, blue: SPPT_W+intphys. February 2014; 31 stations.

The results were not so encouraging for 2-m temperature forecasts. Figure 6 demonstrates the verification scores for the two events “2-m temperature is above 0°C” and “2-m temperature is above 5°C”. For the first event, BS and ROCA are very similar for all experiments, while BSS is slightly better for **noSPPT**. However, for the second event (panel b), the situation changes significantly. The scores range much between the experiments and the great diversity of results gives a chance to analyze the effect of different SPPT settings. The experiment **noSPPT** is clearly the best for all lead times. In contrast to precipitation forecasts, the 2-m temperature predictions are the worst for **SPPTtest** (violet in the plots). Analyzing the curves, we can conclude that interpolation of perturbed values in space and time did not affect the scores noticeably. Most likely it is associated with too coarse perturbation grid (compared to the model grid) used in the experiments. Also only a small effect followed from varying itype_qxlim_rn def, which defined the type of reduction/removal of the perturbation in case of negative or supersaturated values of water vapor content or negative other water-content related characteristics. The scores additionally suggest that not only specific water vapor tendencies but all hydrometeor tendencies should be perturbed.



(a)

(b)

Figure 6: Verification scores as functions of forecast lead-time for the events “2-m temperature is above 0°C” (a) and “2-m temperature is above 5°C” (b) for all experiments listed in Fig.2. Solid line: BS, long-dashed line: BSS, dashed line: ROCA. Red lines: noSPPT, purple: SPPTtest, orange: SPPTphys, black: SPPTintphys, green: SPPT_W, brown: SPPT_W+phys, blue: SPPT_W+intphys. February 2014; 31 stations.

The importance of perturbing all humidity tendencies is confirmed by the skill of ensemble mean forecasts obtained in different experiments. In Fig. 7 the mean error (ME), the mean absolute error (MAE) and the root-mean-square error (RMSE) of 2-m temperature ensemble mean forecasts at Krasnaya Poliana station are presented as functions of lead-time for all the experiments.

Here we again see the prevalence of **noSPPT** experiment in RMSE, MAE, and ME. **SPPTtest** experiment, in which only specific water vapor tendencies were perturbed, gave the largest errors. Perturbing all hydrometeor tendencies helps to improve the scores.

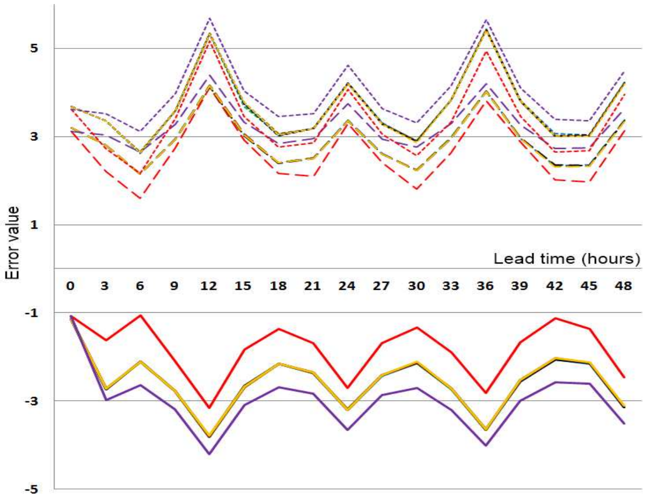


Figure 7: Error graphs for T_{2m} ensemble mean forecasts at Krasnaya Poliana station. Solid line: mean error, long-dashed: mean absolute error, dashed: RMS error. Red lines: noSPPT, purple: SPPTtest, orange: SPPT-phys, black: SPPT_intphys, green: SPPT_W, brown: SPPT_W+phys, blue: SPPT_W+intphys. February 2014.

To complete the analysis, we decided to examine distributions of observed and predicted temperatures. Figure 8 demonstrates the temperature distribution histograms at Krasnaya Poliana for experiments **noSPPT** and **SPPTtest** (the distributions for experiments with other SPPT settings were alike). The eyeball analysis shows that SPPT seems to make the representation of temperature distribution more accurate.

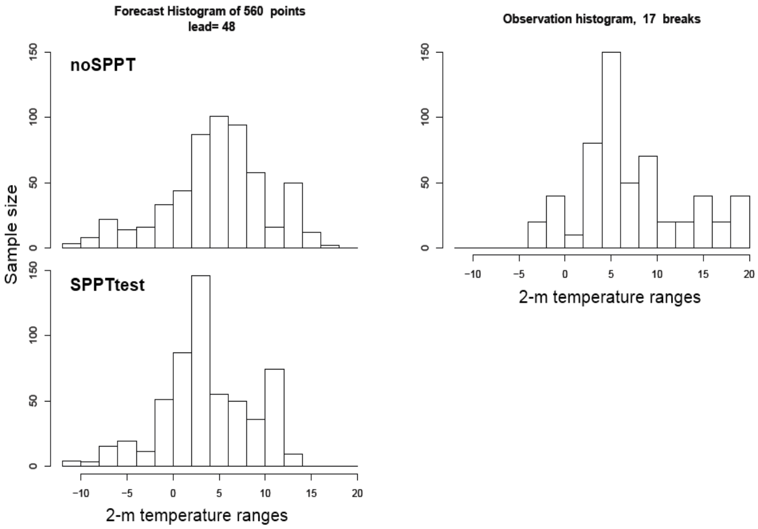


Figure 8: Comparison of T_{2m} distribution histograms for **noSPPT** and **SPPTtest** 48-h forecasts and for observations. February 2014.

Wind speed forecast scores were rather poor both with and without SPPT. SPPT did not make significant difference. Therefore we do not present them here.

5 Conclusions

The experiments with the scheme of stochastic perturbation of physical tendencies (SPPT) were performed using the COSMO-Ru2-EPS ensemble prediction system. The initial and boundary conditions for the runs were provided by COSMO-S14-EPS, the Italian ensemble prediction system developed within the framework of the WWRP FDP/RDP project FROST-2014. The period 1–28 February 2014 was considered. The operational forecasts issued during the Sochi Olympic Games 2014 were used as a reference. Several SPPT settings were tested. Both case studies and probabilistic verifications of forecast series were performed.

Case studies demonstrated that SPPT could be useful for precipitation forecasts improving the description of the rain location and start. The analysis of 2-m temperature predictions in the tropospheric Foehn case revealed the correlation between the T_{2m} ensemble spread and the model orography. Also the coincidence between high-spread areas and the areas of less skillful forecast was found.

The probabilistic verification was performed for the monthly series of COSMO-Ru2-EPS forecasts (56 in total). Some positive effect of using SPPT was found for precipitation forecasts, especially for the event "3-h precipitation is greater than 1 mm". Variations in the SPPT settings did not influence the results much. As for the 2-m temperature forecasts, SPPT does not improve their skill. The verification scores showed rather large difference between experiments with various SPPT settings. Judging by Brier score, the Brier skill score and the area under the ROC curve, the experiment without SPPT gave the best temperature forecasts.

At the same time, the eyeball analysis shows that introduction of SPPT makes the predicted temperature distribution more realistic. Therefore, SPPT did not add value to temperature forecasts, but can sometimes improve the representation of distribution. It is possible to improve the T_{2m} forecast by varying the SPPT settings. For example, perturbing all hydrometeor tendencies in most cases leads to better results than perturbing only specific water content tendency. Also increasing the range of standard deviation for the Gaussian distribution of random numbers and using the higher upper limit imposed to the absolute value of random numbers positively contributed to the results.

Acknowledgments

The authors are grateful to Lucio Torrisi and Christoph Schraff for providing additional SPPT modules for COSMO model version 5.0, Anatoly Muravev for verification software, and Andrea Montani for providing initial-boundary conditions. The study was made within the COTEKINO and SPRED COSMO priority projects.

References

[1] Astakhova, E., A. Montani, D. Kiktev, A. Smirnov: 2016: COSMO-based ensemble forecasting for Sochi-2014 Olympics: archiving the results. COSMO Newsletter, No. 16, **P.40-45**. Available online at $http : //cosmo - model.org/content/model/documentation/newsLetters/newsLetter16/cnl16_06.pdf$

[2] Buizza, R., Miller M., Palmer T., 1999: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. Q. J. R. Met. Soc. ,vol. 125, **P.2887–2908**.

[3] Kiktev, D. B., Astakhova E. D., Zaripov R. B., Murav'ev A. V., Smirnov A. V, and Tsyrulnikov M.D., 2015: FROST-2014 project and meteorological support of the Sochi-2014 Olympics. Russian meteorology and hydrology, vol.40, iss.8, **P.504-512**. DOI 10.3103/S1068373915080051.

[4] Kiktev, D., P. Joe, G. Isaac, A. Montani, I. Frogner, P. Nurmi, B. Bica, J. Milbrandt, M. Tsyrulnikov, E. Astakhova, A. Bundel, S. Belair, M. Pyle, A. Muravyev, G. Rivin, I. Rozinkina, T. Paccagnella, Y. Wang, J. Reid, T. Nipen, and K. Ahn, 2017: FROST-2014: The Sochi Winter Olympics International Project. Bull. Amer. Meteor. Soc. DOI:10.1175/BAMS-D-15-00307.1, in press.

[5] Montani, A., Cesari D., Marsigli C., Paccagnella T., 2011: Seven years of activity in the field of mesoscale ensemble forecasting by the COSMO-LEPS system: main achievements and open challenges. Tellus, 63A, **P.605-624**. DOI: 10.1111/j.1600-0870.2010.00499.x

[6] Montani, A., C. Marsigli, and T. Paccagnella, 2013: Development of a COSMO-based limited-area ensemble system for the 2014 Winter Olympic Games. COSMO Newsletter, No. 13,**P.93–99**. Available online at $http : //cosmo - model.org/content/model/documentation/newsLetters/newsLetter13/cnl13_12.pdf$.

[7] Montani, A., D. Alferov, E. Astakhova, C. Marsigli, and T. Paccagnella, 2014: Ensemble forecasting for Sochi-2014 Olympics: the COSMO-based ensemble prediction systems. COSMO Newsletter, No. 14, **P.88–94**. Available online at $http://cosmo-model.org/content/model/documentation/newsLetters/newsLetter14/cnl14_10.pdf$.

[8] Rivin G. and Rozinkina I., 2011: Priority Project "CORSO": Consolidation of Operation and Research results for the Sochi Olympic Games. Available online at $http://cosmo-model.org/content/tasks/pastProjects/corso/default.htm$

[9] Schaettler U., Doms G., Schraff C. A., 2014: Description of the Nonhydrostatic Regional COSMO-Model. Part VII: User's Guide. Offenbach. November 2014. **P.211**.