SPECIAL PROJECT PROGRESS REPORT

Progress Reports should be 2 to 10 pages in length, depending on importance of the project. All the following mandatory information needs to be provided.

Reporting year	2017
Project Title:	Investigation of case studies during Sochi Olympic Games using COSMO-based ensemble prediction systems.
Computer Project Account:	SPCOLEPS
Principal Investigator(s):	Montani Andrea
Affiliation:	Arpae-SIMC
Name of ECMWF scientist(s) collaborating to the project (if applicable)	
Start date of the project:	2015
Expected end date:	2017

Computer resources allocated/used for the current year and the previous one
(if applicable)
Please answer for all project resources

		Previous year		**Current year**	
		Allocated	Used	Allocated	Used
High Performance Computing Facility	(units)	1.000.000	547.000	1.000.000	126.000
Data storage capacity	(Gbytes)	50	40	50	2

Summary of project objectives

(10 lines max)

As for 2016, the overall aims are twofold:

1. to investigate the performance of the COSMO-S14-EPS system during the Winter Olympics 2014, providing the ensemble fields to fill the gaps in the FROST archive and varying the configurations of the ensembles;
2. in the framework of mesoVICT project, to investigate the skill of the COSMO-based ensemble systems for old case studies, occurred in 2007 in Europe and where high-density observations were available (namely COPS-DPHASE observational dataset).

Summary of problems encountered (if any)

(20 lines max)

Summary of results of the current year (from July of previous year to June of current year)

This section should comprise 1 to 8 pages and can be replaced by a short summary plus an existing scientific report on the project

The Billing Units of the project were used for 2 aims:

1. to perform reruns of ECMWF ENS and of COSMO-S14-EPS to fill the gaps in the FROST archive and enable a proper intercomparison among the ensemble systems participating to the FROST campaign;
2. to perform reruns of ECMWF ENS so as to provide both initial and boundary conditions to drive limited-area ensemble forecasts based on COSMO model for a number of mesoVICT case studies.

Since part 2) is still ongoing, the attached report (taken from the publication listed below) describes the activity of part 1).

The report is SCI-REPORT_spcoleps_2017.pdf

List of publications/reports from the project with complete references

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.

Summary of plans for the continuation of the project

(10 lines max)

This project ends at the end of 2017.

It is planned to start a new special project, which aims at assessing the skill of COSMO-based deterministic and ensemble systems as a function of the scheme used for parameterised convection.

The DOI for this manuscript is doi: 10.1175/BAMS-D-15-00307.1

The final published version of this manuscript will replace the preliminary version at the above DOI once it is available.

If you would like to cite this EOR in a separate work, please use the following full citation:

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.

FROST-2014: The Sochi Winter Olympics International Project

by Dmitry Kiktev[1], Paul Joe[2], George A. Isaac[2], Andrea Montani[3], Inger-Lise Frogner[4], Pertti Nurmi[5], Benedikt Bica[6], Jason Milbrandt[7], Michael Tsyrulnikov[1], Elena Astakhova[1], Anastasia Bundel[1], Stéphane Bélair[7], Matthew Pyle[8], Anatoly Muravyev[1], Gdaly Rivin[1], Inna Rozinkina[1], Tiziana Paccagnella[3], Yong Wang[6], Janti Reid[2], Thomas Nipen[4], Kwang-Deuk Ahn[9]

1 - Hydrometcentre of Russia, Moscow, Russia

2 - Environment and Climate Change Canada, Toronto, Ontario, Canada

3 - Regional Agency for Prevention, Environment and Energy in the Emilia-Romagna region, Italy

4 - MET Norway, Oslo, Norway

5 - Finnish Meteorological Institute, Helsinki, Finland

6 - Central Institute for Meteorology and Geodynamics, Vienna, Austria

7 - Environment and Climate Change Canada, Dorval, Que., Canada

8 - National Centers for Environmental Prediction, College Park, Maryland, USA

9 - National Institute for Meteorological Sciences, Seogwipo, Jeju-do, Korea

CORRESPONDING AUTHOR: Dmitry Kiktev,

Hydrometcentre of Russia, 11-13, Bolshoi Predtechensky st., Moscow, 123242, Russia

E-mail: kiktev@mecom.ru

Capsule

Six nowcasting systems, nine deterministic mesoscale numerical weather prediction models, and six ensemble prediction systems took part in the FROST-2014 project.

Abstract

The WMO WWRP project FROST-2014 (FROST - Forecast and Research in the Olympic Sochi Testbed) was targeted at the advancement and demonstration of state-of-the art nowcasting and short-range forecasting systems for winter conditions in mountainous terrain. The project field campaign was held during the 2014 XXII Olympic and XI Paralympic Winter Games and preceding test events in Sochi. An enhanced network of in-situ and remote sensing observations supported weather predictions and their verification. Six nowcasting systems (model-based, radar tracking, and combined nowcasting systems), nine deterministic mesoscale numerical weather prediction models (with grid spacings down to 250 m), and six ensemble prediction systems (including two ones with explicitly simulated deep convection) participated in FROST-2014. The project provided forecast input for the meteorological support of the Sochi Olympic Games. The FROST-2014 archive of winter weather observations and forecasts is a valuable information resource for mesoscale predictability studies as well as for development and validation of nowcasting and forecasting systems in complex terrain. The resulting innovative technologies, exchange of experience and professional developments contributed to the success of the Olympics and left a post-Olympic legacy.

INTRODUCTION. The Olympic Games are one of the most successful social inventions made in the ancient Greece - like democracy, academia, or theater. As thousands of years ago, the modern Olympics bring people together from across the world for peaceful competitions and invaluable human interactions. Meteorologists have not stayed aside from these events. Since 2000, a number of meteorological projects have been organized in connection with the Olympic Games (Keenan et al. 2003; Wilson et al. 2010; Duan et al. 2012; Isaac et al. 2014; Golding et al. 2014). Most of them were conducted under the umbrella of the WMO World Weather Research Programme (WWRP) as Forecast Demonstration Projects (FDPs) and/or Research and Development Projects (RDPs). FDPs implement scientifically established technologies in practice and demonstrate their capabilities. RDPs aim to advance the meteorology and develop new forecasting methods and technologies. Both provide excellent opportunities for meteorologists from many countries to showcase and further develop their forecast technologies, compare capabilities of different prediction systems, take advantage of an enhanced observation coverage in the area of the Olympic Games, and last but not least, provide operational meteorological support of sport events.

The RDP/FDP FROST-2014 (Forecast and Research in the Olympic Sochi Testbed) was associated with the 2014 XXII Winter Olympic and XI Paralympic Games (henceforth, the Games) held in Sochi, Russia, from 7 to 23 February and from 7 to 16 March 2014, respectively. FROST-2014 (Kiktev et al. 2015a, 2015b) dealt with winter complex terrain forecasting ranging from nowcasting to short-range numerical weather prediction (NWP). Recently, a new RDP/FDP was initiated in connection with the 2018 Pyeong-Chang Winter Olympics.

This paper provides a general overview of the FROST-2014 project, outlines its achievements in nowcasting, short-range deterministic and ensemble forecasting, presents some assessments of the automated project forecasts performance vs. manual forecasts, and concludes with a summary of lessons learned and legacy left.

OLYMPIC DEMANDS AND WEATHER CHALLENGES. Timely provision of high-quality meteorological forecasts is very important to organizers, participants, and spectators of Olympic events because unfavorable weather conditions can lead to delays or even cancellations of open-air competitions. The general logistics of the Olympic infrastructure is also weather sensitive. The Sochi Olympic venues were divided into two clusters: a coastal cluster for indoor ice sport competitions and a mountain cluster for snow sport outdoor events. The latter was located in the Krasnaya Polyana township about 45 km away from the coast (see Fig.1). Sport activities in the mountain cluster were especially weather-dependent.

Weather in the mountains is notoriously capricious. In Sochi, this is exacerbated by the proximity of the Black Sea, a source of heat and moisture. Sharp weather contrasts and high variability are typical for the region. In winter, severe weather conditions include heavy precipitation, freezing rain, fog, strong wind. The nearby Achishkho Ridge (10-15 km to the north-west of Krasnaya Polyana) experiences annual precipitation up to 4.5 m and is the wettest place in Russia. In winter, daily snowfall as large as 92 cm and snow intensities up to 30 cm/h have been registered in the mountain cluster area. Conversely, sometimes the presence of snow might be under threat, affecting snow sports. For example, a strong heat wave in mid-February 2014 with maximum temperatures up to 19°C in Krasnaya Polyana affected the snow cover and was a serious concern for the slope managers. Table 1 presents other interesting weather situations and challenges worth further analysis.

In the context of Olympic Games, high impact weather (HIW) is not necessarily restricted to common severe weather events. Due to the specificity of snow sports, HIW also includes transitions of meteorological variables through sport-specific decision-making thresholds, e.g., there are wind speed restrictions for ski jumping, visibility limitations for biathlon and mountain skiing. Accurate prediction of these sport-specific HIW conditions was as important and challenging as skillful traditional weather forecasts.

PROJECT SCOPE, GOALS, AND PARTICIPANTS. The main attention in FROST-2014 was given to nowcasting and high-resolution short-range numerical prediction, both deterministic and ensemble, of winter weather over complex terrain. The project goals were:

1. Development of a comprehensive information resource of alpine winter weather observations and forecasts.
2. Development of nowcasting systems, mesoscale deterministic and ensemble forecasting systems for winter weather conditions in complex terrain with focus on HIW phenomena.
3. Operational meteorological support of the Games.
4. Improvement of understanding of regional HIW phenomena physics/mechanisms.
5. Evaluation of the developed forecasting systems and assessing benefits of their use (verification and societal impacts).

The list of participating institutions and consortia is given in Table 2.

METEOROLOGICAL OBSERVATION NETWORK. The observational network in the region of Sochi was substantially expanded before the Games. Thirty eight automatic weather stations (Fig. 1) were installed. In addition to temperature, humidity, atmospheric pressure, liquid precipitation, wind speed and direction, some of these stations measured solid precipitation intensity and amount (15 stations), visibility (21 stations), cloud base height (11 stations), radiation balance (6 stations), and snow cover parameters (19 stations). The network strategy was that each sport venue had one basic station and up to five supplementary stations with a reduced list of observed parameters. The primary sampling interval was 10 min. At five stations it was enhanced to 1 min. In addition, a high vertical resolution radiosonde was launched in Sochi daily at 0, 6, 12, and 18 UTC.

A Vaisala C-Band Doppler dual polarization radar WRM200 was installed on Akhun mountain (Fig. 1) at an altitude of 680 m above the sea level. This position was chosen to ensure optimal surveillance coverage and to monitor cloud and precipitation systems approaching the Olympic venues from the Black Sea. In winter 2013/2014, data from two C-Band Doppler radars (located in Samsun and Trabzon), and two X-Band radars (located in Simferopol and Donetsk) were kindly provided by the Turkish Meteorological Service and the Ukrainian State Air Traffic Service, respectively. The latter four radars were invaluable as they provided upstream coverage over the sea (Fig. 2). For the first time a nearly complete radar coverage of the Black Sea was produced. These data were supplemented by measurements from a RPG-HATPRO temperature and humidity profiler, a Scintec LAP3000 sodar, an ATTEX MTP-5 temperature profiler and two METEK micro rain radars (Fig. 1). These instruments were helpful for monitoring of low atmospheric layers in the valleys shaded from the Akhun radar by the mountains (Fig. 2).

The Sochi observations also included images from seven webcams, and snow surveys by local avalanche-protection troops. The real-time observation data were available to the FROST-2014 participants via Internet from the project server (see section "FROST-2014 ARCHIVE").

FORECAST VERIFICATION AND INTERCOMPARISON SETUP. The verification setup for the FROST-2014 weather prediction systems has been introduced in Murav'ev et al. (2013, 2015) and Nurmi et al. (2014, 2015). Predictions were compared with near-surface station data for a period from 15 January 2014 to 18 March 2014, if not indicated otherwise. Some nowcasting systems produced predictions at the observation locations, whereas other forecasting systems provided gridded fields. For gridded predictions, observations were compared with the closest grid points without vertical adjustment and not accounting for slope orientation. As the models' computational grids were different, some models were in a more favorable position for some stations. This effect could be significant in complex terrain. To reduce the resulting noise in the verification scores, an aggregation for groups of similar stations was performed. The verification results are presented in the later sections.

NOWCASTING SYSTEMS. The FROST-2014 participants provided the project with various kinds of prognostic information, that was made available to the Sochi forecasters' team for elaboration of official forecasts and meteorological support of the Olympic events (see section "MANUAL AND AUTOMATED

FORECASTS”). Six nowcasting systems contributed to the project (Table 3). They are briefly characterized as follows.

ABOM (see Bailey et al. 2014) produces nowcasts combining observations, observation trends and trends from a single NWP model, while INTW (Huang et al. 2012, 2014a, 2014b) provides integrated nowcasts from blending observation data and weighted forecasts from several NWP models. Both systems use observations from the previous six hours to train the algorithms and generate an improved point forecast. In FROST-2014,ABOM and INTW predicted 2-m temperature (T_{2m}), relative humidity (RH_{2m}), 10-m wind, and visibility for selected points at 10 min intervals for the first hour and then either hourly (ABOM) or every 10 min (INTW) for up to eight hours. INTW used model output from GEM-1, GEM-0.25, COSMO-Ru2, COSMO-Ru7, and WRF-ARW-NIMS models (Table 4) to produce the integrated forecast. ABOM produced nowcasts based on each of these models. Both systems employ the visibility prediction algorithm described in Boudala and Isaac (2009), and Boudala et al. (2012) using nowcast RH_{2m} to help overcome the model humidity errors.

CARDS is a radar-processing nowcasting system based on Lagrangian radar data extrapolation. During the Games, CARDS mosaicked the data from Akhun, Trabzon, Samsun, Donetsk, and Simferopol radars every 15 min. Point forecasts of precipitation were produced using a cross-correlation nowcast technique (Bellon and Austin 1978). The uncertainty in the precipitation intensity was conveyed by back-trajectory and estimating the upstream intensity along the mean of the track and the maximum intensity in the swath of ±8°. This has proved to be highly reliable (Ebert et al. 2004) and easily interpreted.

INCA (Haiden et al. 2011) is a gridded analysis and nowcasting system that uses different kinds of observation and model forecast data. The FROST-2014 INCA domain was 180x140 km. The system predicted precipitation and precipitation type with 10 min resolution. Wind speed and direction, T_{2m}, RH_{2m}, dew-point, ground temperature, freezing level, and snow line were predicted with hourly resolution. The INCA nowcasting fields were merged into the NWP fields with a linearly decreasing weighting factor. The analysis background and model forecasts were provided by ALARO (Wang et al. 2011) with a physics package designed for a horizontal grid spacing of around 5 km. For precipitation, the analysis background was derived from the Akhun radar.

The JOINT system generated nowcasts and short-range forecasts at station locations as weighted NWP multi-model means adjusted to the latest observations. The system aggregated all the latest deterministic model forecasts available from the project participants (Table 4), and also employed the Lagged Average Forecasting (Hoffman and Kalnay 1983) adding several overlapping consecutive model forecasts from earlier analyses. For the Games period, JOINT was implemented only for continuous meteorological variables, not including precipitation.

The MeteoExpert system combined several nowcasting tools including a radar-processing component and a numerical model of atmospheric boundary layer fed by observations and external NWP background. Cross-correlation tracking, averaged Doppler velocity, and prognostic wind at 700 hPa level were combined to estimate precipitation advection. Site-specific 4-h forecasts of T_{2m}, RH_{2m}, dew-point temperature, wind, precipitation intensity, cloud base height, and visibility were provided by the system with a 10-min update (Bazlova 2014).

Most nowcasting systems have been developed for prediction of summer convective phenomena and for regions with relatively flat topography. Experience in winter nowcasting in mountains has been very modest. SNOW-V10 (Science of Nowcasting Olympic Weather for Vancouver-2010) was the first winter

Olympic nowcasting project in complex terrain conducted under the WWRP that involved international researchers (Isaac et al. 2014). Several model-based nowcasting approaches tested in SNOW-V10 were adopted to the Sochi testbed. Testing of the systems in the different environments disclosed some local specificity in their behaviour. For example, during the Vancouver-2010 Olympics most cases with reduced visibility were associated with snowfall. By contrast, in Sochi, low visibility was mostly caused by fog or low clouds. Due to considerable errors in the numerical predictions of humidity, visibility reductions in fog were predicted less successfully than visibility reductions in precipitation.

Figure 3 displays Mean Absolute Errors (MAE) of the point-specific NWP-based nowcasts of T_{2m} and RH_{2m} (INCA and CARDS are not shown in the figure as INCA is not a point-based system, and CARDS does not predict the considered variables). The persistence forecasts were still competitive as compared to the more sophisticated techniques. For T_{2m} persistence was overtaken by the model-based systems only after 2 to 3 h. The nowcasts for T_{2m} were more successful than for RH_{2m}, which was probably caused by the better skill of temperature NWP contributions relative to the model humidity input to the nowcasting systems. After 1 h, the lowest MAEs of T_{2m} predictions were demonstrated by JOINT. For RH_{2m} INTW performed better than the other systems.

In mountainous regions, the Lagrangian radar echo extrapolation does not properly capture the orographic effects. The orographic impact on precipitation fields is complex and depends on the speed of incoming flow and the stratification of the atmosphere (e.g., Medina et al. 2005). This impact is manifested in the general increase of precipitation on windward and weakening on lee side slopes. Some preliminary assessments of the orographic forcing on precipitation intensity were obtained from a series of Akhun radar precipitation rate fields (not shown). Cross-correlation tracking of reflectivity fields at 1.5-km height above the radar with 5-min update and 1-km horizontal resolution was used to generate about five thousand nowcasts for the 2013 winter season. Reflectivity was converted to precipitation rate using the Marshall-Palmer relationship (Marshall and Palmer 1948; Marshall and Gunn 1952). However, quantifying systematic differences between the precipitation intensity in upstream areas and at the forecast locations has been inconclusive. Challenges include objective identification and separation of orographic enhancement from other phenomena, proper conversion of reflectivity to precipitation rate considering precipitation type, extrapolation to the surface, and determining the upstream location and precipitation value. Nevertheless, the CARDS radar nowcasting products (90-min point predictions of precipitation intensity) proved very useful by the Sochi Olympics forecasters for intensity, start and cessation times. The strong point of the radar approach with respect to the NWP-based nowcasting is the more accurate initial locations of meteorological features. Further work on the intercomparison of the radar and NWP-based nowcasts is ongoing.

An inherent part of nowcasting is diagnosis of weather phenomena. In particular, precipitation type is of special interest for winter sport events. EC modified a radar dual-polarization algorithm (Park et al. 2009) for the C-Band Akhun radar and compared it to the Vaisala hydrometeor classification algorithm (Liu and Chandrasekar 2000). For rain, present weather detectors (PWD-20 and PWD-22 by Vaisala) at different weather stations within the mountain cluster showed that the EC algorithm compared better than the Vaisala algorithm for rain (with occurrence rate of 82% and 40%, respectively, vs. the observed 90% of the rain detection) at 500-750 m above sea level. The EC algorithm overestimated wet snow over rain in the bottom part of the melting layer and underestimated wet snow at the top (Fig. 4). The Vaisala algorithm tended to produce deeper layers of wet snow where the EC algorithm reported graupel and dry snow (Reid et al., 2014). The main difference between the two algorithms is the determination of the height of the melting level.

DETERMINISTIC NUMERICAL WEATHER FORECASTING. Nine deterministic NWP systems contributed to the project (Table 4). Their descriptions can be found in Baldauf et al. (2011), Rivin et al. (2015), Milbrandt et al. (2016), Niemelä et al. (2014), and Janjic and Gall (2012).

Figures 5 and 6 give an impression of the general performance of the 1-km deterministic forecasting systems in the mountain cluster. More specific validation results are reported in Murav'ev et al. (2013, 2015). Figure 5 shows the MAEs for T_{2m}, RH_{2m}, 10-m wind direction and speed as functions of lead time. Figure 6 presents the verification statistics for 1-h precipitation in terms of the Equitable Threat Score (ETS) (WMO 2008). Both MAE and ETS are *pointwise* scores here and thus can suffer from the double penalty problem (If an observed event is misplaced with respect to its predicted location then this forecast is penalized twice: at both the actual and the predicted locations). Verification results with spatial methods are intended to be published in follow-up papers.

Forecast error growth. In Fig. 5 any visible forecast error growth with the lead time is hardly visible. For some of the models this error evolution was compared with the error growth in flat terrain. Over flatlands, the initial MAE was usually lower than in the mountains and the error growth was more pronounced (not shown). This difference might be important for some practical purposes, e.g., the Lagged Average Forecasting may appear more efficient in complex terrain than in flat terrain. A number of studies revealed the similar forecast error evolution in complex terrain (Colman et al. 2013). It is conjectured (Anthes et al. 1985) that at least in some cases physical forcing at the land surface, such as mountains, may contribute to extended atmospheric predictability. Some mechanisms behind this effect were investigated by Vukicevic and Errico (1990).

Inter-model differences. From Figs. 5 and 6, one can see that the performance of a model with respect to the other models depends on the predicted variable. The NEMS/NMMB model manifested the best T_{2m} MAEs and good precipitation scores, while it had the worst RH_{2m} and wind speed MAEs among the 1-km models when averaged over all runs. HARMONIE Arome performed very well for wind, however, its T_{2m} and precipitation scores were poor. Precipitation was better forecasted by GEM-1, but its wind direction MAEs were the largest. In most cases, the scores of COSMO were in between the other models and never the worst. These inter-model differences can be caused by multiple reasons. For example, in case of the T_{2m} and RH_{2m} forecast scores it can be linked to distinctions in the employed land surface models, different vertical resolutions in the lower boundary layer etc. The differences in the wind scores can be attributed to differences in the model orographies and roughness parameters. A more focused experimental setup is needed to identify the sources of individual distinctive features of model behaviour more confidently.

Aggregation of the verification scores over all forecast start times masks some features in model behaviour. More details can be drawn from Figs. 15 and 16 (which are primarily devoted to the comparison of the numerical schemes with the human forecasts in the section "MANUAL AND AUTOMATED FORECASTS") for forecasts started from 1200 UTC. Specifically, the diurnal cycle of T_{2m} MAE was different for various models: with the daytime maximum for COSMO-Ru7, COSMO-Ru2, NEMS/NMMB and INCA (which transited to ALARO forecasts at these lead times) and daytime minimum for WRF-ARW-NIMS and HARMONIE Arome. The odd behaviour of HARMONIE Arome (poor T_{2m} scores at night and the best ones at daytime) was investigated in Niemelä et al. (2014). It appeared that the large nighttime errors were mostly caused by the CANOPY turbulence scheme (Masson and Seity, 2009). Without it, the temperature had a more moderate underestimation of 1-2°C. For precipitation (Fig. 16), all the models exhibited poorer forecasts at daytime than at night.

Data assimilation. There were several efforts to benefit from data assimilation for deterministic NWP in FROST-2014:

- HARMONIE Arome used 3D-Var data assimilation for upper air quantities and optimum interpolation for surface variables. Only observations from regular (i.e., not including stations from the enhanced Olympic Sochi network) near-surface stations, radiosondes, and aircraft observations were utilized. The background error statistics were created by using an ensemble method (Niemelä et al. 2014).

- The nudging scheme (Schraff 1997) was implemented to assimilate near-surface data and radiosondes with the COSMO model at resolutions 7 and 2.2 km.

- A limited area 3D-Var was developed at Roshydromet to assimilate near-surface, radiosonde, aircraft, and satellite wind data in the COSMO-Ru2 model.

The attempts to use data assimilation with COSMO-Ru2 and COSMO-Ru7 did not result in substantial forecast improvements in the Sochi testbed. This can be interpreted as follows. First, in a small domain, the information from initial conditions is quickly swept out from the domain being largely replaced by information propagated from the lateral boundary conditions; as a result, data assimilation in limited area applications is in general not as beneficial as it is on the global scale. Second, land surface data assimilation, which affects the important surface forcing, was lacking in these experiments. Third, many more observations (radar and satellite) are needed to impact a model with tens of millions of degrees of freedom. Particularly this concerns the vast upstream areas of the Black Sea that are poorly covered with contact observations.

Role of resolution. Both COSMO-Ru and GEM systems were available at three different horizontal grid spacings (Table 4). This made it possible to evaluate the effect of the horizontal grid spacing on the quality of forecasts (Figures 7 and 8). The MAE and the Extremal Dependence Index (EDI) (Ferro et al. 2011) were selected as verification metrics. The EDI was recommended as a good estimator of forecast accuracy for all thresholds, and for rare events, in particular. It is positively oriented (the higher the better) and ranges from -1 to 1 with 0 corresponding to the level of random forecast. Note that in Figs. 7 and 8, the number of model runs per day was significantly different for COSMO-Ru and GEM (see caption to Fig. 7). This may explain the flatter curves for COSMO-Ru compared to GEM models, where the larger variability in the scores might be attributed to the diurnal cycle effects.

The near-surface forecast errors partly originate from the differences between the actual and model orographies. With smaller horizontal grid spacings, these errors are expected to be reduced. Indeed, the refinement of the COSMO model resolution from 7 to 2.2 km was beneficial for T_{2m}, RH_{2m} and 10-m wind direction forecasts (Fig.7). The further refinement of the COSMO-Ru model horizontal grid from 2.2 to 1.1 km appeared to be positive mainly for wind speed. For the GEM model, the improvement at higher resolution is clear for T_{2m}. Transition to 250 m grid spacing was also quite beneficial for nighttime wind direction, but made the wind speed forecast worse. In some cases, the effect of resolution enhancement was less evident.

A low visibility event. One of the most serious weather impacts on the Games was caused by the low clouds and related visibility reduction in the mountain cluster during 16-17 February. The biathlon men's mass-start was postponed from 16 to 17 February and further to 18 February, and the snowboard qualification was postponed from 17 to 18 February. Both the long-lasting visibility reduction due to fog on 16 February and subsequent window of relatively good visibility in the afternoon of 17 February (before the next visibility reduction due to heavy snowfall) were captured in the official forecast bulletin issued daily at 15 h.

Figure 9 shows COSMO-Ru1 and COSMO-Ru2 forecasts starting at 06 UTC 16 February, along with observations. In Fig. 9 one can see the growth of RH_{2m} on 16 February (the onset of the event), then reaching 100% RH_{2m} for about 24 hours (fog) with subsequent decrease in the late afternoon of 17 February (the good visibility window). It is remarkable that all the phases of the event were reasonably well predicted by both COSMO-Ru versions (Shatunova et al. 2015) in terms of relative humidity (COSMO-Ru does not predict visibility directly). This numerical guidance was very helpful in elaboration of the official forecast of this HIW event on 16 February, and the planned women's biathlon mass-start was held during the predicted window of good visibility on 17 February.

Along with the traditional meteorological variables, some project models predicted less common variables, such as visibility, cloud base height, and reflectivity. Fig. 10 illustrates direct visibility forecasts for the same event by three versions of GEM model with different grid spacings. It is interesting to note that forecast by GEM-0.25 from 00 UTC on 16 February was the most successful. It realistically reproduced the timing of the sharp visibility reduction on 16 February (although the duration of low visibility period was underestimated).

ENSEMBLE PREDICTION. The FROST-2014 ensemble prediction systems (EPS) are listed in Table 5. Two convection-permitting systems (i.e., systems with explicitly simulated deep convection), COSMO-Ru2-EPS and HarmonEPS, were tested in research mode while the coarser resolution EPSs were operational. All forecasts were issued twice a day, starting from 00 and 12 UTC with the exception for the HIRLAM systems that started at 06 and 18 UTC. The detailed information about the systems can be found in Frogner et al. (2016), Du et al. (2014), Iversen et al. (2011), Montani et al. (2013, 2014), and Wang et al. (2011). The Games area was within the operational domains of ALADIN-LAEF and GLAMEPS, whereas the other systems were specifically set up for FROST-2014.

The EPSs generated a set of probabilistic products, including ensemble mean and ensemble standard deviation for several near-surface and upper-air variables, probability of exceeding a specified threshold, as well as ensemble meteograms for selected points. Additionally, pointwise calibrated and hourly updated GLAMEPS forecasts were produced. At the time of the Games, GLAMEPS had been operational for several years, and the development of calibrated forecasts had reached a level where it could be provided as part of the FDP. For HarmonEPS it was the first attempt to run the system in real time, and calibration was not part of it. HarmonEPS was calibrated after the Games, and this is documented in Frogner et al. (2016). The impact of calibration on the skill of COSMO-based ensembles will be investigated in forthcoming studies. The ensemble products were systematically presented at the FROST-2014 site and widely applied and appreciated by the Sochi forecasters.

After the Games the project research was mainly focused on possible advantages of high-resolution convection-permitting and multi-model ensembles as well as on the effects of calibration. Figure 11 presents the Continuous Ranked Probability Score (CRPS; the lower the better) (WMO 2008) for ECMWF

EPS, GLAMEPS, calibrated GLAMEPS, and HarmonEPS forecasts, three systems having quite different resolutions. While ECMWF EPS and GLAMEPS had a comparable number of ensemble members (51 and 54, respectively), HarmonEPS had only 13 members. The most striking feature in Fig. 11 is the effect of calibration producing much better scores for temperature and wind, and slightly better for precipitation for most lead times. Running an EPS is expensive, while calibration is much cheaper in terms of computational cost and thus appears to be a highly beneficial approach.

Other developments of HarmonEPS after the Games were calibration and an enrichment of the ensemble. Besides 13 AROME-based members, another 13 ALARO model members were added. Figure 12 shows CRPS for the original HarmonEPS and its extended version (labeled as "multi-physics"). There is a clear effect of the ensemble extension leading to better CRPS, which can be explained by the increased diversity in the ensemble and, thus, its higher representativeness.Figure 12 also includes calibrated HarmonEPS and a calibrated subset of GLAMEPS based on 26 members only, that is, the same number of members as in the extended HarmonEPS. Like for GLAMEPS, calibration was beneficial for HarmonEPS, and calibrated HarmonEPS scored better than the calibrated GLAMEPS with the same number of members, indicating that the finer resolution calibrated HarmonEPS has the higher potential than thecalibrated GLAMEPS for predicting winter weather. For details, see the dedicated paper on the HIRLAM contribution to FROST-2014 (Frogner et al., 2016).

Figure 13 illustrates the potential of multi-model approach using the FROST-2014 EPSs. The areas under the relative operating characteristic (ROC) curves for individual EPSs and their combined multi-model ensemble are shown. The scores for convection-parameterized (left) and convection-permitting (right) EPSs are given as functions of forecast lead time for 6-h precipitation exceeding 1 mm. All FROST-

2014 EPSs exhibited quite high and, on average, comparable ROC values. It can be noticed that the scores of the multi-model ensemble are consistently higher than those of its constituents for all forecast ranges, indicating a better ability of the system to predict this type of events. For more details, see (Montani et al. 2016).

The role of spatial resolution for EPS performance is demonstrated in Figure 14. Here, the debiased RPSS (Ranked Probability Skill Score) was selected as it makes ensembles with differing sizes comparable (Weigel et al. 2007). In general, the higher-resolution ensembles with an explicit treatment of convection performed better than the convection-parameterized systems (COSMO-Ru2 and HarmonEPS vs. COSMO-Ru7 and GLAMEPS, respectively).

Before the Games, the majority of the local forecasters had a very limited practice in use of ensemble forecast products. The Games experience facilitated the gradual embedding of the probabilistic thinking into their working practices and formed a new need for this kind of numerical guidance. The probabilistic information tended to be more actively used by the forecasters for the second and third forecast days, while the deterministic predictions were preferred for the shorter forecast ranges. In some situations (particularly in the case of previously mentioned low visibility event) the information on forecast uncertainty was conveyed to the sport managers for support of the decision making.

FROST-2014 ARCHIVE. A special server with data storage was dedicated to the FROST-2014 project at the Hydrometcentre of Russia. All the participants were provided with access to operational meteorological observations and used them to run and verify their forecasts for the Sochi region. The FROST-2014 contributors computed the forecasts at their home institutes in real-time and uploaded the results to the server via Internet. On the project website http://frost2014.meteoinfo.ru, the forecasters and the project participants could get the data in digital and graphical formats and also use additional online tools for forecast verification and comparison.

The most intense data collection period was during the cold season of 2013/2014. However, some of the forecast and observation records are 2-3 years long. Automatic weather station data, regional SYNOP observations, radar graphical products and raw data (volume files), vertical profiler data, images from web-cameras, upper-air sounding data, project automated forecasts, official forecast bulletins and some additional information are available to the meteorological scientific community via the project server.

MANUAL AND AUTOMATED FORECASTS. FROST-2014 was an 'end-to-end' project. Its operational forecasts were used by the Olympic Forecasting team gathered from the whole Roshydromet for meteorological support of the Games. List of the models and products expanded significantly in 2013 and even shortly before the Games. This diversity of forecast data was both a great help and a challenge. Sometimes the numerical guidance was misleading. Occasionally, the automated forecasting systems experienced difficulties in predicting the timing of weather events. Difficulties in the prediction of the presence/absence and amount of precipitation tended to grow under conditions of low-gradient fields. Substantial errors were noticed in relative humidity, wind direction and maximum speed forecasts. The visibility and cloud-base forecasts should still be considered experimental despite their capability of producing a useful signal.

Time and practical experience were needed for the forecasters to adapt to the new products. Forecasters tend to use familiar products in their operational work. The most popular were products whose regular delivery started well before the Games and which were introduced to the forecasters during the pre-Olympic trainings in 2010-2013. Transfer of experience of FROST-2014 experts from EC and COSMO lecturing at the training courses helped a lot in building the forecasters' confidence in the new forecasting products.

Under the operational time constraints, usually the forecasters did not have enough time to review and analyze all the available products. To compress this information and to facilitate preparation of the required hourly forecast updates for the information system of the Games, an automatic generation of a forecast first guess was employed using multi-model blended forecasts of the JOINT system. A special web-interface was developed for the forecasters to correct this first guess, if necessary. This FROST-2014 data feed to the Olympic information system can be considered as one of the strongest project societal impacts. The performance of combined multi-model products was on average at the level of the best forecasts of individual forecasting systems and sometimes even exceeded it, especially during the first forecast hours.

FROST-2014 provided a good opportunity to compare performance of the manual forecasts with the automated ones being used as numerical guidance. One of the regular official products for the Games was the Forecast Bulletin for the mountain cluster of Olympic sport venues. This bulletin was issued at around 1500 Local Time (LT) and covered a 24-h period starting from 2200 LT, that is, with a 7-h lead time. The nearest models start time for comparison of automated forecasts with the Forecast Bulletins is 12 UTC (1600 LT). Some results of the inter-comparison between the official and automated forecasts with hourly temporal resolution for the period from 1 November 2013 to 23 February 2014 are presented in Figures 15 and 16.

Figures 15, 16, and similar results on winds and visibility (not shown) demonstrate the following

- Automated temperature forecasts, especially blended multi-model forecasts, were competitive to manual forecasts;

- For wind speed and visibility, the human forecasts demonstrated the psychological biases towards higher speed and lower visibility (the phenomenon of overforecasting hazardous events by human forecasters is discussed, e.g., by Doswell (2004));

- For precipitation, the manual forecasts did add value to model forecasts.

SUMMARY AND CONCLUSIONS. Weather forecasts were crucial for the efficient conduct of the Sochi Olympic Games. This information was essential for sport teams, organizers, broadcasters, spectators, and general public. Itaffected decisions of sport managers and was the reason for a number of changes in the Games schedule. FROST-2014 nowcasts and NWP guidance data were used by the forecasters for meteorological support of the Games, and thus contributed to the success of these events.

Implementation of the project strengthened the numerical guidance for the Olympic weather services with new state-of-the-art forecast products. A series of training sessions, including ones with participation of the project international experts greatly helped in capacity building of the forecasters. The multi-model JOINT forecasts served as a first guess for the forecasters in their production of hourly prognostic updates requested by the International Olympic Committee. Involvement in the project had an important educational value for the local forecasters.

Despite the diversity of available state-of-the-art forecast data, the project experience shows that the tested systems were insufficient on their own for meteorological support of such a high-profile event and that the role of a human forecaster was still crucial. A post-event survey among the forecasters showed their great interest in new prediction technologies resulting from FROST-2014. The survey also highlighted some lessons learnt, e.g., a diversity of available prognostic products makes their form and usability very important to forecasters.

The high-resolution data assimilation in the Sochi testbed was mostly limited to assimilation of non-satellite and non-radar observations. More extensive assimilation of remote sensing data and updating land surface fields is important for further forecast improvements in complex terrain.

The NWP systems demonstrated some benefits of transition from several kilometers to one kilometer and down to sub-kilometer grid spacing. A number of NWP post-processing techniques (in particular, ABOM, INTW, JOINT and calibrated GLAMEPS) were implemented for further refinement of the project numerical forecasts down to the individual Olympic venues and proved themselves quite efficient under conditions of complex terrain. Model-based nowcasts of continuous variables were informative and helpful, but sometimes struggled to beat persistence. Radar nowcasting was limited by the problem of Lagrangian echo extrapolations in complex terrain but the forecasters found the CARDS products useful. The acquired experience facilitated implementation of a number of new methods and products into operations in the post-Olympic period (e.g., radar data assimilation, new NWP postprocessing techniques with rapid forecast update, spatial verification methods etc).

All the forecasting systems exhibited their strengths and weaknesses. It is quite difficult to single out an unambiguous winner among the systems that participated in the field campaign, because the results of this rating vary substantially depending on location, meteorological variable, forecast lead time, and other factors. The same applies to the ensemble prediction systems. A more robust outcome is that, as with over flat areas, the multi-model forecasts were consistently more informative than the forecasts of individual systems. However, there were significant differences in skill for particular cases and variables. These differences might come from many sources: data assimilation schemes, types and numbers of assimilated data, driving global models, configurations of nested limited area models, and other details. A more rigorous unified experimental setup, e.g., with common global driving model and boundary conditions, is needed in this respect for more in-depth diagnostic studies and inter-comparisons of the forecasting systems. In general, the FROST-2014 NWP systems were state of the art, so the Sochi testbed verifications may be considered as characteristic of current NWP capabilities in mountain conditions.

Only a few systematic inter-comparisons of multiple mesoscale forecasting systems in mountains are known due to the lack of appropriate observations and coordinated forecasting activities. In this respect the Sochi testbed provided a valuable information resource for development of forecasting systems and research of mesoscale predictability in complex terrain. Despite the limitations of the observational network in the Sochi region, the content and density of these Olympic testbed observations substantially surpassed the normal operational networks. The observations, project forecasts, likewise official forecast bulletins are available to the meteorological scientific community via the FROST-2014 Internet-server.

Another page in the history of the Olympics is closed. However, for FROST-2014 this is not the end of the story. The project participants continue processing and analyzing the field campaign data. A series of papers is under preparation to shape the project legacy.

ACKNOWLEDGMENTS. The authors acknowledge the guidance of the WWRP and its Working Groups, especially the Nowcasting and Mesoscale Research and Forecast Verification Research Working Groups, to facilitate and promote this work. Roshydromet thanks the COSMO consortium and, in particular, DWD and MeteoSwiss for their help with the COSMO system. The authors would like to thank Slobodan Nickovic, Nanette Lomarda and Alexander Baklanov from the World Weather Research Division, WMO; Anna Glazer, Ruping Mo, Ivan Heckman, Monica Bailey, Laura Huang, David Hudak, Sudesh Boodoo, Norman Donaldson from EC; Sami Niemelä, Sigbritt Näsman, Ari Aaltonen, Matias Brockmann and Mikko Partio from FMI; John Bremnes from Met Norway, Kai Sattler from DMI; Alexander Kann, Jasmina Hadzimustafic, Florian Weidle, Martin Suklitsch from ZAMG; Nikolai Bocharnikov and Tatyana Bazlova from IRAM; Valery Lukyanov, Radomir Zaripov, Alexander Smirnov, Denis Blinov, Marina Shatunova, Dmitry Alferov, Alexander and Yury Melnichuk, Arkady Koldaev, and the Olympic Forecasting team of Roshydromet.

References

- Anthes, R. A., Y.-H. Kuo, D. Baumhefner, R. M. Errico, and T. W. Bettge, 1985: Predictability of mesoscale atmospheric motions. *Adv. Geophys.*, **28B**, 159–202.
- Astakhova, E. D., A. Montani, and D. Yu. Alferov, 2015: Ensemble forecasts for the Sochi-2014 Olympic Games. *Russ. Meteor. Hydrol.*, **40**, 531-539, doi: 10.3103/S1068373915080051.
- Bailey, M. E., G. A. Isaac, I. Gultepe, I. Heckman, and J. Reid, 2014: Adaptive blending of model and observations for automated short range forecasting: examples from the Vancouver 2010 Olympic and Paralympic Winter Games. *Pure Appl. Geophys.*,**171**, 257–276, doi: 10.1007/s00024-012-0553-x.
- Baldauf, M., A. Seifert, J. Förstner, D. Majewski, M. Raschendorfer, and T. Reinhardt, 2011: Operational Convective-Scale Numerical Weather Prediction with the COSMO Model: Description and Sensitivities. *Mon. Wea. Rev.*, **139**, 3887-3905.
- Boudala, F. S. and G.A. Isaac, 2009: Parameterization of visibility in snow: Application in Numerical Weather Prediction models. *J. Geophys. Res.*, **114**, D19202, doi:10.1029/2008JD011130.
- Boudala, F. S., G. A. Isaac, R. Crawford, and J. Reid, 2012: Parameterization of runway visual range as a function of visibility: Implications for numerical weather prediction models. *J. Atmos. Oceanic Technol.*, **29**, 177-191.
- Bazlova, T., 2014: Weather stations for Sochi 2014. *Meteorol. Technol. Int.*, August 2014, 142-145.

- Bellon, A., and G. L. Austin, 1978: The Evaluation of Two Years of Real-Time Operation of a Short-Term Precipitation Forecasting Procedure (SHARP). *J. Appl. Meteor.*, **17**, 1778–1787, doi: 10.1175/1520-0450(1978)017<1778:TEOTYO>2.0.CO;2.
- Colman, B., K. Cook, and B. J. Snyder, 2013: Numerical Weather Prediction and Weather Forecasting in Complex Terrain. *Mountain Weather Research and Forecasting: Recent Progress and Current Challenges*, Springer, 655-692.
- Doswell III, C. A., 2004: Weather forecasting by humans - heuristics and decision making. *Wea. Forecasting*, **19**(6), 1115-1126.
- Du, J., G. DiMego, B. Zhou, D. Jovic, B. Ferrier, B. Yang, and S. Benjamin, 2014: NCEP Regional Ensembles: Evolving toward hourly-updated convection-allowing scale and storm-scale predictions within a unified regional modeling system. *22nd Conf. on Numerical Weather Prediction and 26th Conf. on Weather Analysis and Forecasting*, Atlanta, GA, Amer. Meteor. Soc., J1.4.
- Duan, Y., and Coauthors, 2012: An Overview of the Beijing, 2008: Olympics Research and Development Project (B08RDP). *Bull. Amer. Meteor. Soc.*, **93**, 381–403, doi:10.1175/BAMS-D-11-00115.1
- Ebert, E. E., L. J. Wilson, B. G. Brown, P. Nurmi, H. E. Brooks, J. Bally, and M. Jaeneke, 2004: Verification of nowcasts from the WWRP Sydney 2000 Forecast Demonstration Project, *Wea. Forecasting*, **19**, 73-96, doi: 10.1175/1520-0434(2004)019<0073:VONFTW>2.0.CO;2.
- Golding, B. W., and Coauthors, 2014: Forecasting capabilities for the London 2012 Olympics. *Bull. Amer. Meteor. Soc.*, **95**, 883–896, doi: 10.1175/BAMS-D-13-00102.1.

● Ferro, C. A. T., and D. B. Stephenson, 2011: Extremal dependence indices: improved verification measures for deterministic forecasts of rare binary events, *Wea. Forecasting*, **26**, 699-713, doi: 10.1175/WAF-D-10-05030.1

● Frogner, I.-L., T. Nipen, A. Singleton, J. B. Bremnes, and O. Vignes, 2016: Ensemble prediction with different spatial resolution for the 2014 Sochi Winter Olympic games - the effects of calibration and multi-model approaches. *Wea. Forecasting,* in press.

● Haiden, T., A. Kann, C. Wittmann, G. Pistotnik, B. Bica, and C. Gruber, 2011: The Integrated Nowcasting through Comprehensive Analysis (INCA) system and its validation over the Eastern Alpine region. *Wea.Forecasting,* **26**, 166-183, doi: 10.1175/2010WAF2222451.1.

● Hoffman, R. N., and E. Kalnay, 1983: Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus,* **35A**, 100-118.

● Huang, L. X., G. A. Isaac, and G. Sheng, 2012: Integrating NWP forecasts and observation data to improve nowcasting accuracy. *Wea. Forecasting*, **27**, 938-953, doi: 10.1175/WAF-D-11-00125.1.

● Huang, L. X., G. A. Isaac, and G. Sheng, 2014a: A new integrated weighted model in SNOW-V10: Verification of continuous variables. *Pure Appl. Geophys.*,**171**, 277–287, doi: 10.1007/s00024-012-0548-7.

● Huang, L. X., G. A. Isaac, and G. Sheng, 2014b: A new integrated weighted model in SNOW-V10: Verification of categorical variables. *Pure Appl. Geophys.*,**171**, 289–302, doi: 10.1007/s00024-012-0549-6.

● Iversen, T., A. Deckmyn, C. Santos, K. Sattler, J. B. Bremnes, H. Feddersen, and I.-L. Frogner, 2011: Evaluation of "GLAMEPS" - a proposed multi-model EPS for short range forecasting. *Tellus*, **63A**, 513-530, doi: 10.1111/j.1600-0870.2010.00507.x.

● Isaac, G. A., and Coauthors, 2014: Science of nowcasting Olympic weather for Vancouver 2010 (SNOW-V10): A World Weather Research Programme project. *Pure Appl. Geophys.*, **171**, 1-24, doi: 10.1007/s00024-012-0579-0.

● Janjic, Z., and R. L. Gall., 2012: Scientific documentation of the NCEP Nonhydrostatic Multiscale Model on the B grid (NMMB). Part 1 Dynamics. NCAR Tech. Note. NCAR/TN-489.+STR, 75 pp, doi:10.5065/D6WH2MZX.

● Joe, P., M. Falla, P. Van Rijn, L. Stamadianos, T. Falla, D. Magosse, L. Ing, and J. Dobson, 2003: Radar data processing for severe weather in the National Radar Project of Canada, *21st Conf. on Severe Local Storms*, San Antonio, TX, Amer. Meteor. Soc., P4.13.

● Kann, A., G. Pistotnik, and B. Bica, 2012: INCA-CE: A Central European initiative in nowcasting severe weather and its applications, *Adv. Sci. Res.*, **8**, 67-75, doi:10.5194/asr-8-67-2012.

● Keenan, T., and Coauthors, 2003: The Sydney 2000 World Weather Research Programme forecast demonstration project. Overview and current status. *Bull. Amer. Meteor. Soc.*, **83**, 1631–1643, doi: 10.1175/BAMS-84-8-1041.

● Kiktev, D. B., E. D. Astakhova, R. B. Zaripov, A. V. Murav'ev, A. V. Smirnov, and M. D. Tsyrulnikov, 2015a: FROST-2014 project and meteorological support of the Sochi-2014 Olympics. *Russ. Meteor. Hydrol.*, **40**, 504-512, doi: 10.3103/S1068373915080051.

● Kiktev, D. B., E. D. Astakhova, R. B. Zaripov, A. V. Murav'ev, A. V. Smirnov, and M. D. Tsyrulnikov, 2015b: Erratum to: "FROST-2014 project and meteorological support of the Sochi-2014 Olympics". *Russ. Meteor. Hydrol.*, **40**, 844–845, doi: 10.3103/S1068373915120109.

- Liu, H., and V. Chandrasekar, 2000: Classification of hydrometeors based on polarimetric radar measurements: Development of fuzzy logic and neuro-fuzzy systems, and in situ verification. *J. Atmos. Oceanic Technol.*, **17**, 140-164, doi: 10.1175/1520-0426(2000)017<0140:COHBOP>2.0.CO;2
- Masson, V., and Y. Seity, 2009: Including atmospheric layers in vegetation and urban offline surface schemes. *J. Appl. Meteor. Climatol.*, **48**, 1377–1397, doi: 10.1175/2009JAMC1866.1.
- Marshall, J. S., and K. L. S. Gunn, 1952: The measurement of snow parameters by radar. *J. Meteor.*, **9**, 322-327, doi: 10.1175/1520-0469(1952)009<0322:MOSPBR>2.0.CO;2.
- Marshall, J. S., and W. McK. Palmer, 1948: The distribution of raindrops with size. *J. Meteor.*, **5**, 165-166, doi: 10.1175/1520-0469(1948)005<0165:TDORWS>2.0.CO;2.
- Medina, S., B. F. Smull, R. A. Houze Jr., and M. Steiner, 2005: Cross-barrier flow during orographic precipitation events: results from MAP and IMPROVE. *J. Atmos. Sci.*, **62**, 3580–3598, doi: 10.1175/JAS3554.1.
- Milbrandt, J. A., S. Belair, M. Faucher, M. Vallee, M. A. Carrera, and A. Glazer, 2016: The Pan-Canadian high resolution (2.5-km) deterministic prediction system. Submitted to *Wea. Forecasting*.
- 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, 93–99. [Available online at http://cosmo-model.org/content/model/documentation/newsLetters/newsLetter13/cnl13_12.pdf]
- 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, 88–94. [Available online at http://cosmo-model.org/content/model/documentation/newsLetters/newsLetter14/cnl14_10.pdf].

• Montani, A., D. Alferov, E. Astakhova, I.-L. Frogner, M. Pile, and F. Weidle, 2016: Intercomparisons of the FROST-2014 ensemble forecasts: predictive capability and performance of multi-model ensembles. Submitted to *Wea. Forecasting*.

• Murav'ev, A. V., D. B. Kiktev, A. Yu. Bundel, T. G. Dmitrieva, and A. V. Smirnov, 2015: Verification of high-impact weather event forecasts for the region of the Sochi-2014 Olympic Games. Part I: Deterministic forecasts during the test period. *Russ. Meteor. Hydrol.*, **40**, 584-597, doi:10.3103/S1068373915090034.

• Murav'ev, A. V, A. Yu. Bundel', D. B. Kiktev, and A. V. Smirnov, 2013: Verification of mesoscale forecasts in the 2014 Olympic Games region. Part II: Preliminary results of diagnostic evaluation of quality and calibration of the forecasts by the COSMO-RU2 model. *Russ. Meteor. Hydrol.*, **38**, 797–807, doi:10.3103/S1068373913120017.

• Niemelä, S., S. Näsman, and P. Nurmi, 2014: FROST-2014 - Performance of HARMONIE 1km during Sochi Olympics. *ALADIN-HIRLAM Newsletter*, No. 3, 79-86 [Available online at http://www.hirlam.org/index.php/hirlam-documentation/cat_view/77-hirlam-official-publications/285-aladin-hirlam-newsletters].

• Nurmi, P., M. Brockmann, and S. Näsman, 2014: Forecast verification framework and some early results of the Sochi 2014 Winter Olympics. *WWOSC*, Montreal, Canada, WMO. [Available online at https://www.wmo.int/pages/prog/arep/wwrp/new/wwosc/documents/WWOSC_SCI-PS124.02_Nurmi_Monday-am.pdf].

• Nurmi, P., and Coauthors, 2015: The framework of the WMO/WWRP FROST-2014 forecast verification setup and activities. *15th EMS Annual Meeting / 12th European Conference on Application of*

Meteorology (ECAM), Sofia, Bulgaria, EMS2015-577. [Available online at http://meetingorganizer.copernicus.org/EMS2015/EMS2015-577.pdf].

- Park, H. S., A. V. Ryzhkov, D. S. Zrnic, and K. E. Kim, 2009: The hydrometeor classification algorithm for the polarimetric WSR-88D: Description and application to an MCS, *Wea. Forecasting*, **24**, 730-48, doi: 10.1175/2008WAF2222205.1.
- Reid, J., D. Hudak, S. Boodoo, N. Donaldson, P. Joe, D. Kiktev, and A. Melnichuk, 2014: Dual-Polarization radar particle classification results during the Sochi Olympic Games. *ERAD2014-The 8th European Conf. on Radar in Meteorology and Hydrology*, Garmisch-Partenkirchen, Germany [Available online at http://www.pa.op.dlr.de/erad2014/programme/ExtendedAbstracts/151_Reid.pdf].
- Rivin, G. S., and Coauthors, 2015: The COSMO-Ru system of nonhydrostatic mesoscale short-range weather forecasting of the Hydrometcenter of Russia: The second stage of implementation and development. *Russ. Meteor. Hydrol.*, **40**, 400-410, doi: 10.3103/S1068373915060060
- Schraff, C., 1997: Mesoscale data assimilation and prediction of low stratus in the Alpine region. *Meteorol. Atmos. Phys*., **64**, 21-50.
- Shatunova, M. V., G. S. Rivin, and I. A. Rozinkina, 2015: Visibility forecasting for February 16–18, 2014 for the region of the Sochi-2014 Olympic Games using the high-resolution COSMO-Ru1 model. *Russ. Meteor. Hydrol*., **40**, 523-530, doi: 10.3103/S106837391508004X.
- Vukicevic, T., and R. M. Errico, 1990: The Influence of Artificial and Physical Factors upon Predictability Estimates Using a Complex Limited-Area Model. *Mon. Wea. Rev.*, **118**, 1460-1482.
- Wang, Y., and Coauthors, 2011: The Central European limited-area ensemble forecasting system: ALADIN-LAEF. *Quart. J. Roy. Meteor. Soc.*,**137**, 483–502, doi: 10.1002/qj.751.

- Weigel, A. P., M. A. Liniger, and C. Appenzeller, 2007: The discrete Brier and ranked probability skill scores. *Mon. Wea. Rev.*, **135**, 118–124, doi: 10.1175/MWR3280.1.
- Wilson, J. W., Y. Feng, M. Chen, and R. D. Roberts, 2010: Nowcasting challenges during the Beijing Olympics: Successes, failures, and implications for future nowcasting systems. *Wea. Forecasting*, **25**, 1691–1714, doi:10.1175/2010WAF2222417.1.
- WMO, 2008: Recommendations for the verification and intercomparison of QPFS and PQPFS from operational NWP models. WWRP 2009-1, WMO-TD/No. 1485, World Meteorological Organization.[Available online at https://www.wmo.int/pages/prog/arep/wwrp/new/documents/WWRP2009-1_web_CD.pdf].

Table 1. List of the most interesting weather cases during the Sochi Games.

Case	**Meteorological Phenomenon**	**Models' behaviour**	**Impact on competitions**
7 Feb	Tropospheric Foehn	Most models underestimated temperatures above 1700 m by 1.4...3.7°C	
11-12 Feb	Precipitation dissipation	Precipitation in the Mountain Cluster predicted by majority of the systems, but not observed actually	
15 Feb		Poor maximum wind speed forecast by most models at Krasnaya Polyana (underestimation by 3.5…7 m/s)	
16-17 Feb	Low visibility	Only high-resolution models were useful	Postponed competitions at the Biathlon sport venue and Extreme Park
18 Feb	Cold front	Good precipitation forecast by most models	
22 Feb	Foehn	Most models underestimated temperature by 2.4...4.4°C (most markedly at 1500 m)	
11 Mar	Cold front. Low visibility	Poor forecasts of temperature maximum by most models (maximum temperature was forecasted at noon, whereas in reality it occurred in the morning)	Postponed skiing events at the mountain skiing venue
13 Mar	Low-gradient field	Poor precipitation forecast by most models above 1500 m	
17 Mar	Cold front	Poor maximum wind speed forecast (underestimation) by most models above 1500 m	

666 **Table 2. FROST-2014 participants.**

Participating institutions	**Consortium / overarching organization**	**Country**
Central Institute for Meteorology and Geodynamics (ZAMG)	High Resolution Numerical Weather Prediction Project **A**ire **L**imitee, **A**daptation dynamique, **D**eveloppement **I**nter**N**ational (ALADIN)	Austria
Environment and Climate Change Canada (ECCC, hereinafter referred to as EC)		Canada
Federal Service for HydroMeteorology and Environmental Monitoring (Roshydromet)	**CO**nsortium for **S**mall-scale **MO**deling (COSMO)	Russia
Finnish Meteorological Institute (FMI)	**HI**gh **R**esolution **L**imited **A**rea **M**odel (HIRLAM)	Finland
Hydro-Meteo-Climate Service of the Environmental Agency of Emilia-Romagna (ARPA/SIMC)	**CO**nsortium for **S**mall-scale **MO**deling (COSMO)	Italy
National Centers for Environmental Prediction (NCEP)	**N**ational **O**ceanic and **A**tmospheric **A**dministration (NOAA)	USA
National Institute for Meteorological Sciences (NIMS)	**K**orean **M**eteorological **A**dministration (KMA)	Republic of Korea
Met Norway	**HI**gh **R**esolution **L**imited **A**rea **M**odel (HIRLAM)	Norway

Table 3. FROST-2014 nowcasting systems.

System	Organization / institute	Country
ABOM (**A**daptive **B**lending of **O**bservations and **M**odel)	EC	Canada
CARDS (**CA**nadian **R**adar **D**ecision **S**upport system)	EC	Canada
INCA (**I**ntegrated **N**owcasting through **C**omprehensive **A**nalysis)	ZAMG	Austria
INTW (**INT**egrated **W**eighted forecasts)	EC	Canada
JOINT	Roshydromet/ Hydrometcentre of Russia	Russia
MeteoExpert	Institute of Radar Meteorology (IRAM)	Russia

Table 4. FROST-2014 deterministic forecasting systems.

System name /consortium /institution	**NumForc /ForcLen /OutFreq**	**Model resolution/ grid type**	**Lateral boundary conditions**	**Initial conditions**	**Boundary Layer /Convection /Land-surface /Radiation schemes**
COSMO-Ru7 /COSMO /Roshydromet	4/ 78h /3h	7 km L40 / Rot Lat-Lon	GME 20 km L60	GME 20 km L60	TKE at level 2.5 /Tiedke for COSMO-Ru-7; reduced Tiedtke scheme for shallow convection only for COSMO-Ru2,1 /TERRA-ML /Ritter-Geleyn
COSMO-Ru2 /COSMO /Roshydromet	4/ 48h /1h	2.2 km L50 / Rot Lat-Lon	COSMO-Ru7	COSMO-Ru7+ nudging	
COSMO-Ru1 /COSMO /Roshydromet	4/ 36h /1h	1.1 km L50 / Rot Lat-Lon	COSMO-Ru2	COSMO-Ru2+ nudging	
NEMS/NMMB //NCEP	2-4/ 24h / 0.5h	1 km L40 / Rot Lat-Lon	GFS T574L64	Down-scaled from a global (GFS) analysis	Mellor-Yamada-Janjic level 2.5 /Betts-Miller-Janjic at 10% "strength"/ NOAH/ RRTM
GEM-2.5// EC	1/27h / 1h	2.5 km L57 / Lat-Lon	GEM (global grid) run at 25 km grid spacing provided initial and boundary conditions for the first nested GEM 10 km (LAM grid) then subsequently for GEM 2.5 km, 1 km and 250 m (LAM grids).		Moist TKE / Kuo-transient shallow convection scheme/ ISBA / Li-Barker radiation scheme
GEM-1// EC	1/25h / 1h	1 km L57 / Lat-Lon			
GEM-0.25// EC	1/24h / 1h	0.25 km L57/Lat-Lon			
HARMONIE Arome /HIRLAM / FMI	4/36h / 1h	1 km L65 / Lambert	ECMWF model, hourly, one-way nesting	3D-Var for upper air and OI for surface and soil	1D prognostic TKE with a diagnostic mixing length / EDFM for dry thermals and non-precipitating shallow cumuli /SURFEX/RRTM
WRF-ARW-NIMS//KMA	2/ 48h/ 1h	2 km L40 / Lambert	ECMWF model, T127	Global ECMWF analysis	YSU/no/ modified NOAH /RRTM

Note: NumForc is the number of forecasts per day, ForcLen denotes the forecast length, OutFreq is the frequency of output information, (Rot) Lat-Lon means the (rotated) latitude-longitude grid, Lambert stands for the Lambert projection, L is the number of vertical levels, OI is the optimum interpolation, 3D-Var is the three-dimensional variational assimilation. PBL is the planetary boundary layer, TKE means the PBL parameterization with an equation for turbulent kinetic energy prognosis, YSU is the Yonsei University PBL scheme. ISBA, NOAH, SURFEX, TERRA-ML are the land surface models, TRRTM is the T rapid radiative transfer model, EDFM stands for the eddy-diffusivity mass-flux scheme, GME, GEM, and GFS are the global numerical weather prediction models (operational in the German Weather Service, EC, and NCEP, respectively), ECMWF is the European Center for Medium-Range Weather Forecasts, LAM is the limited area model.

Table 5. FROST-2014 ensemble prediction systems.

System name /consortium /center	ForcLen/ OutFreq	Model resolution/ grid type	Ensemble size	Driving system/ representation of forecast uncertainty
COSMO-S14-EPS /COSMO /ARPA-SIMC	72 h/ 3 h	7 km L40 / Rot Lat-Lon	10	ECMWF EPS/ multi-physics
GLAMEPS /HIRLAM-ALADIN /MET Norway	54 h/ 3 h	11 km L37-91 /Rot Lat-Lon	54	ECMWF EPS/ different models, stochastic physics, surface data assimilation for all forecasts
ALADIN-LAEF /ALADIN /ZAMG	72h / 3h	11 km L45 interpolated to Lat-Lon 7 km	17	ECMWF EPS + regional perturbations/ T_{2m} and RH_{2m} assimilation, multi-physics, W_s and T_s perturbations
NMMB-EPS //NCEP	72 h/ 3h	7 km L60 / Lat-Lon	7	GEFS/ multi-physics
COSMO-Ru2-EPS /COSMO /Roshydromet	48 h/ 1 h	2.2 km L50 / Rot Lat-Lon	10	COSMO-S14-EPS
HarmonEPS /HIRLAM /MET Norway	36 h/ 1 h	2.5 km L65 / Lambert	13	ECMWF EPS/ 3D-Var for the control forecast, surface data assimilation for all forecasts

Note**:** W_s and T_s are soil moisture and surface temperature. Multi-physics denotes application of different parameterization schemes and/or their parameters. GEFS is the NCEP global ensemble forecast system. For other notations see Table 4.

Figure captions

Figure 1. The Sochi Olympic area on the global map (a), the magnified map with locations of the meteorological equipment (b), and the mountain cluster with the stations and five sport venues (c). Symbols' meanings: the red bulbs designate the automatic meteorological stations, the radar icon is the Doppler radar, green bulbs are the micro rain radars, blue bulbs are the temperature profilers, and the yellow bulb is the wind profiler.

Figure 2. Example of the radar reflectivity composite for region of the Games. Akhun, Trabzon, Samsun, Donetsk and Simferopol radar coverages are shown by circles.

Figure 3. MAE of point-specific forecasts aggregated over stations at the sport venues of the mountain cluster. ABOM for COSMO-Ru2 is the ABOM system based on COSMO-Ru2 forecasts. Aggregation period: from 15 January to 18 March 2014, averaged over hourly runs.

Figure 4: (Top) EC and (Bottom) Vaisala particle classifications for 1.1º scan on February 26 1755 UTC. The EC shows more rain than the Vaisala classification (red tones) and the opposite for wet snow (blue tones).

Figure 5. MAE of 1-km resolution model forecasts. The scores are aggregated over all model runs (COSMO-Ru1 and HARMONIE Arome: 0000, 0600, 1200, and 1800 UTC; GEM-1: 2300 UTC; NEMS/NMMB: 0000 and 1200 UTC) and over 22 stations in the mountain cluster. Here and in Figs. 6, 7, and 8 the period is from 15 January to 18 March 2014.

Figure 6. As in Fig. 5 but for the Equitable Threat Score of 1-h precipitation > 1 mm (the higher the better).

Figure 7. The role of the horizontal grid spacing for COSMO (left) and GEM (right) model families. Score: MAE. The scores are aggregated over all model runs (COSMO: 0000, 0600, 1200, and 1800 UTC; GEM-2.5: 2100 UTC, GEM-1: 2300 UTC, GEM-0.25: 0000 UTC) and over 22 stations in the mountain cluster.

Figure 8. As in Fig. 7 but for 1-h precipitation occurrence forecasts. Score: Extremal Dependence Index (the higher the better).

Figure 9. The RH_{2m} forecasts by COSMO-Ru2 and COSMO-Ru1 from 0600 UTC 16 February 2014 and corresponding observations for the low visibility event at the Biathlon stadium.

Figure 10. The visibility forecasts by GEM-2.5 (from 2100 UTC 15 February), GEM-1 (from 2300 UTC 15 February), GEM-0.25 (from 0000 UTC 16 February) and corresponding observations for the low visibility event at the Biathlon stadium. A model prediction of 100 km indicates unlimited visibility. The PWD sensors can report a maximum of 20 km visibility.

Figure 11. CRPS for ECMWF EPS, GLAMEPS, calibrated GLAMEPS, and HarmonEPS (the lower the better). Top: T_{2m}; Middle: 10-m wind speed; Bottom: 3-h precipitation.

Figure 12. CRPS for 10-m wind speed forecasts for HarmonEPS, extended HarmonEPS with two sub-ensembles, calibrated HarmonEPS, and calibrated GLAMEPS based on 26 members only.

Figure 13. Area under the ROC curve (the higher the better) for forecasts of the event "6-h accumulated precipitation is above 1 mm" aggregated over the stations of the mountain cluster for convection-parameterized (left panel) and convection-permitting (right panel) EPSs as well as for the corresponding multi-model ensembles. Note that about 200 occurrences of the above event were observed during the verification period.

Figure 14. Debiased RPSS (the higher the better) for 6-h accumulated precipitation forecast by two convection-parameterized (COSMO-S14-EPS and GLAMEPS) and two convection-permitting EPSs (COSMO-Ru2-EPS and HarmonEPS), aggregated over the stations of the mountain cluster.

Figure 15. The skill of official and model forecasts as a function of the lead time. MAE of T_{2m} aggregated over the mountain cluster (heights of about 600, 1000, 1500, and 2000 m), period from 1 November 2013 to 23 February 2014 (HARMONIE Arome - from 9 December 2013, WRF-ARW-NIMS - from 23 December 2013), official forecasts issued at 1100 UTC, the models started at 1200 UTC. After 24 h lead time, the HARMONIE Arome forecasts were issued with 6-h step, that's whence the blue dot at 30 h lead time on the plot.

Figure 16. Same as Fig. 15, but for the Extremal Dependence Index (EDI, the higher the better) of 1-h precipitation occurrence.

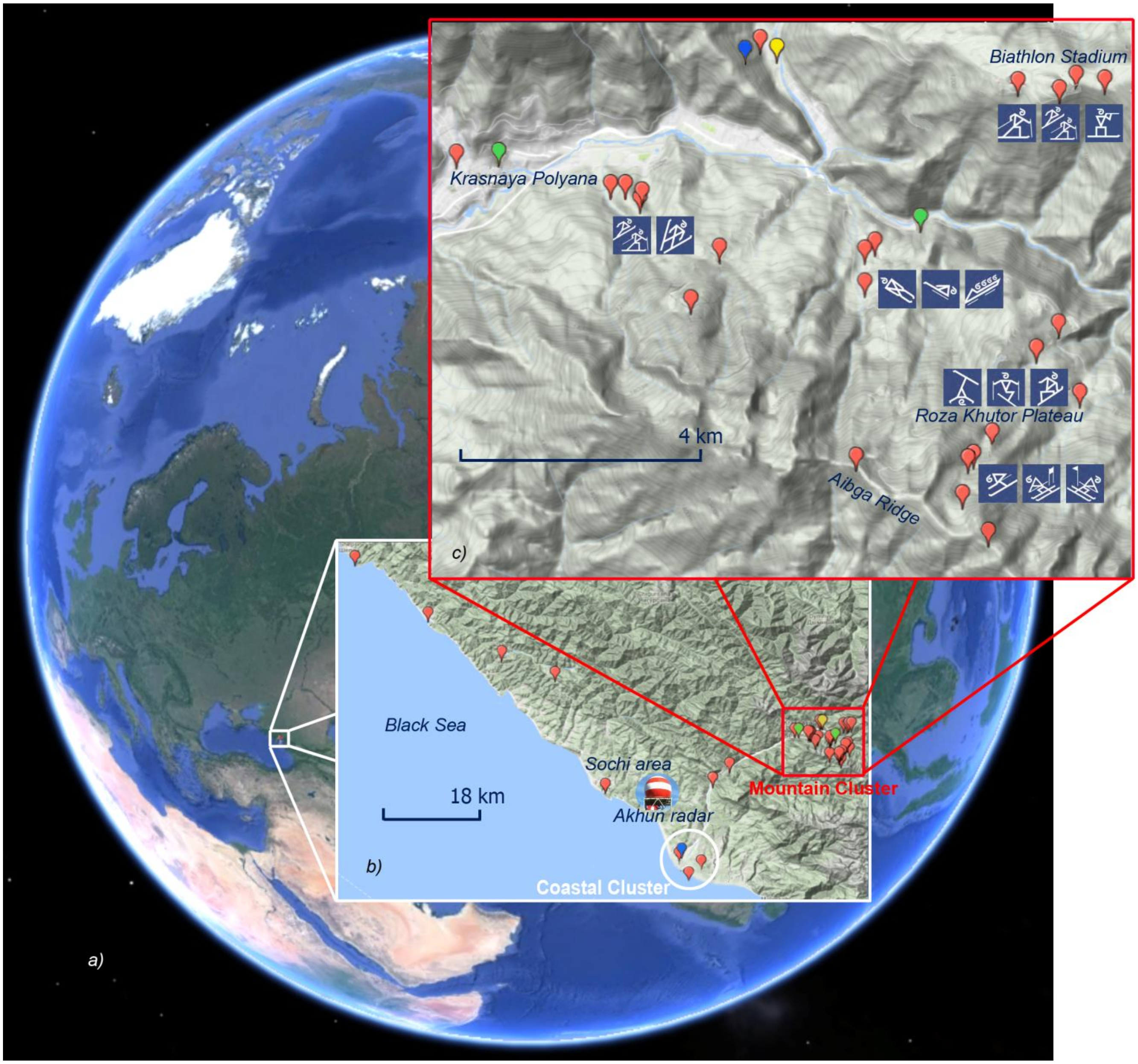

Figure 1. The Sochi Olympic area on the global map (a), the magnified map with locations of the meteorological equipment (b), and the mountain cluster with the stations and five sport venues (c). Symbols' meanings: the red bulbs designate the automatic meteorological stations, the radar icon is the Doppler radar, green bulbs are the micro rain radars, blue bulbs are the temperature profilers, and the yellow bulb is the wind profiler.

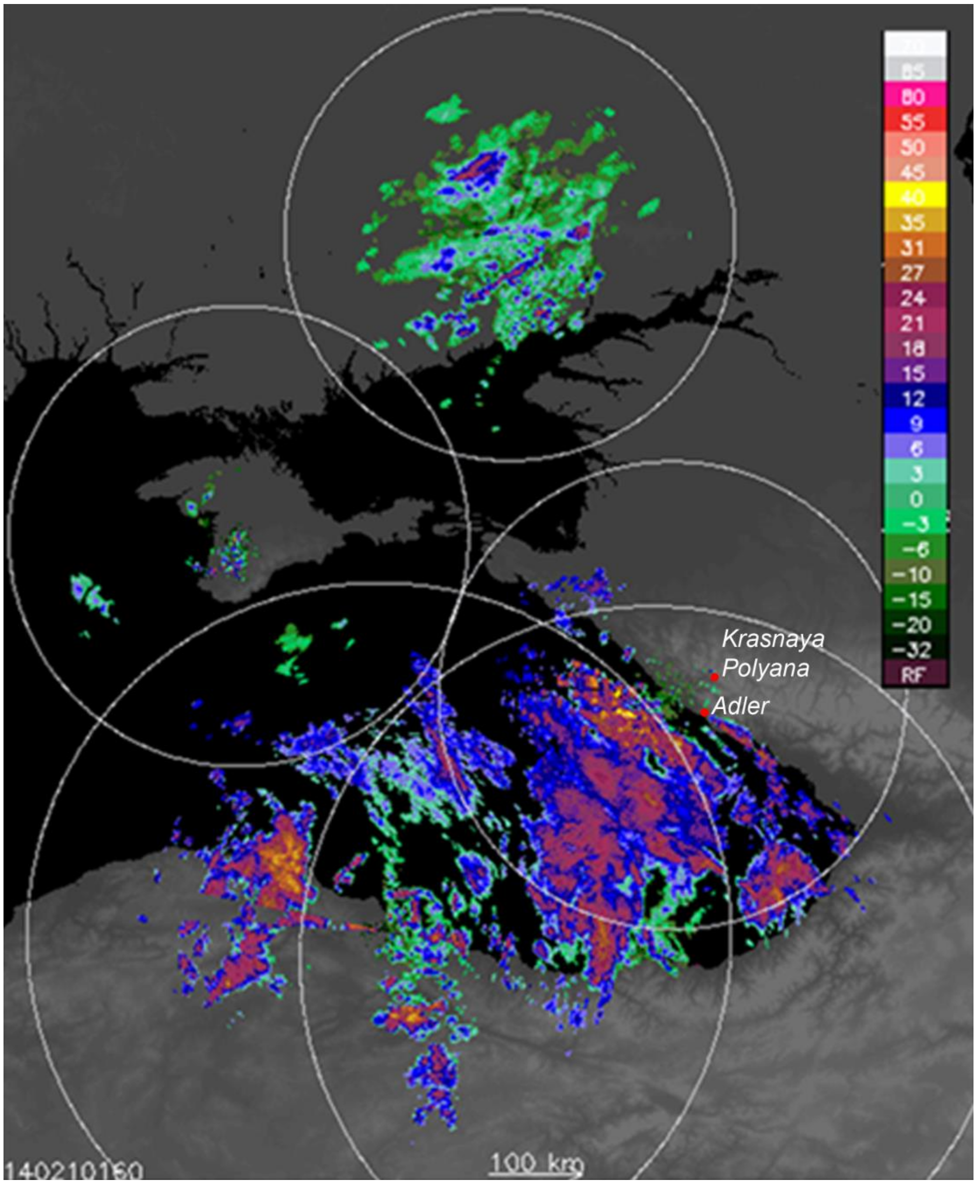

Figure 2. Example of the radar reflectivity composite for region of the Games. Akhun, Trabzon, Samsun, Donetsk and Simferopol radar coverages are shown by circles.

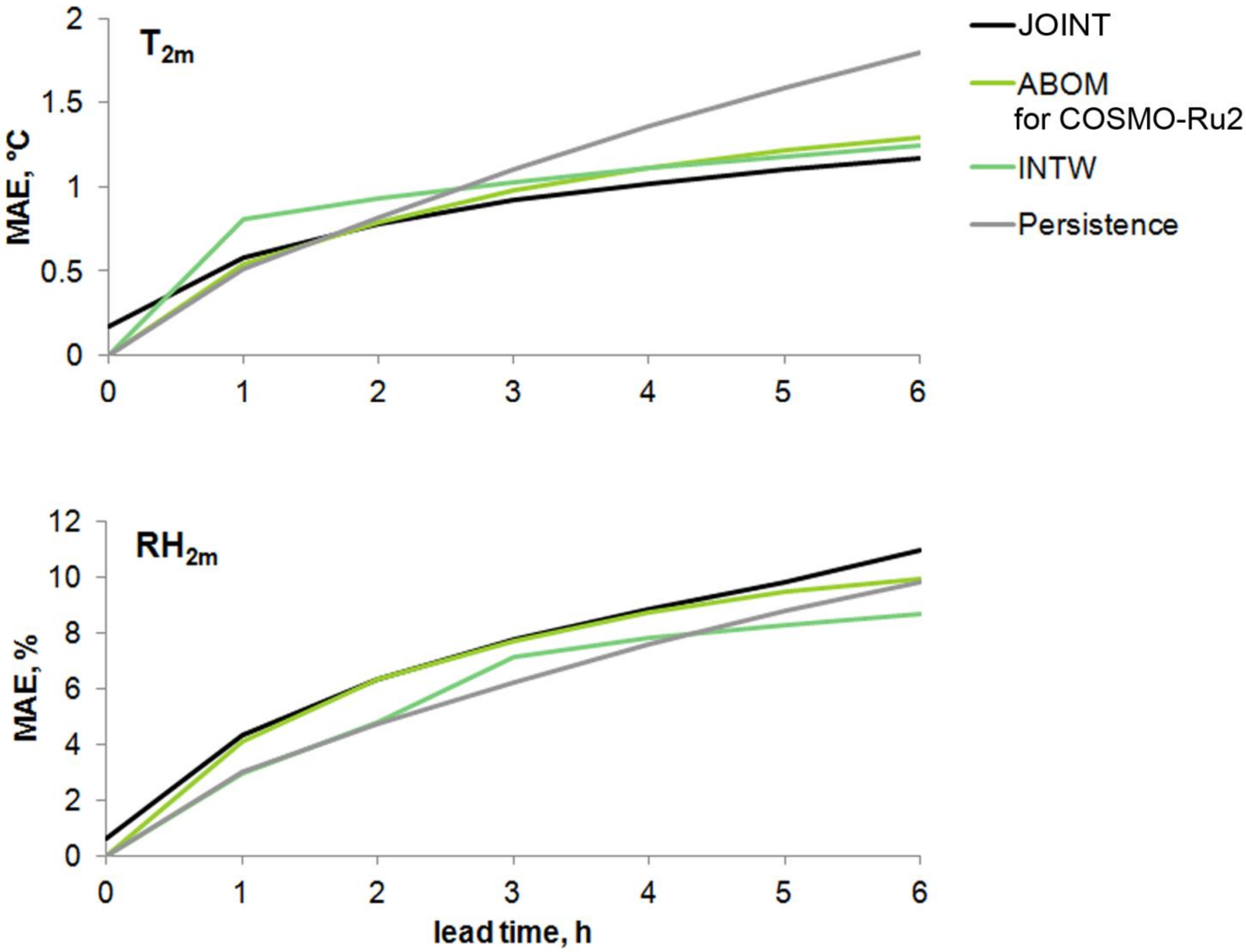

Figure 3. MAE of point-specific forecasts aggregated over stations at the sport venues of the mountain cluster. ABOM for COSMO-Ru2 is the ABOM based on COSMO-Ru2 forecasts. Aggregation period: from 15 January to 18 March 2014, averaged over hourly runs.

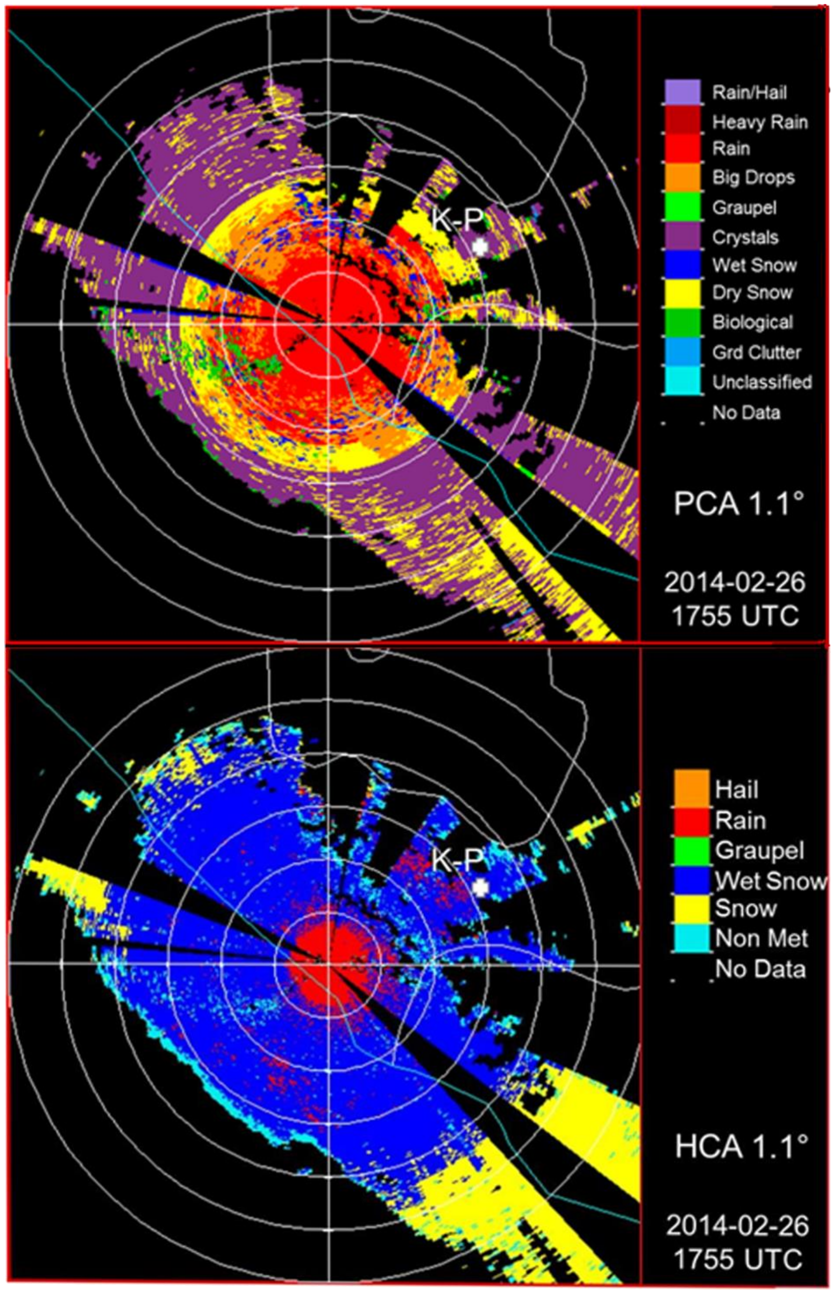

Figure 4: (Top) EC and (Bottom) Vaisala particle classifications for 1.1º scan on February 26 1755 UTC. The EC shows more rain than the Vaisala classification (red tones) and the opposite for wet snow (blue tones).

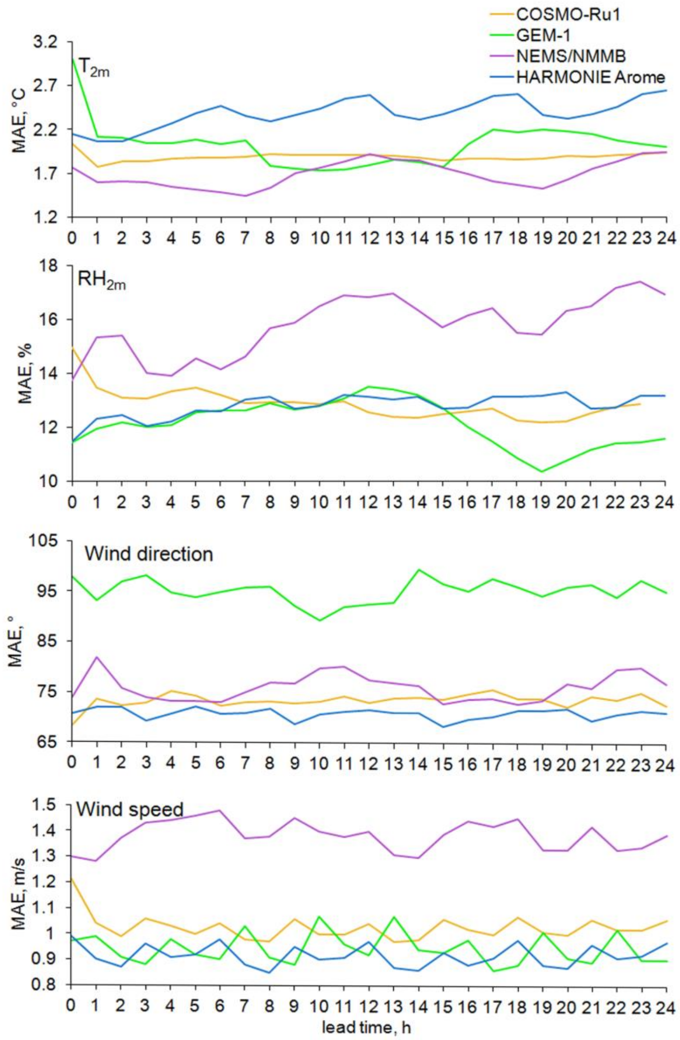

Figure 5. MAE of 1-km resolution model forecasts. The scores are aggregated over all model runs (COSMO-Ru1 and HARMONIE Arome: 0000, 0600, 1200, and 1800 UTC; GEM-1: 2300 UTC; NEMS/NMMB: 0000 and 1200 UTC) and over 22 stations in the mountain cluster. Here and in Figs. 6, 7, and 8 the period is from 15 January to 18 March 2014.

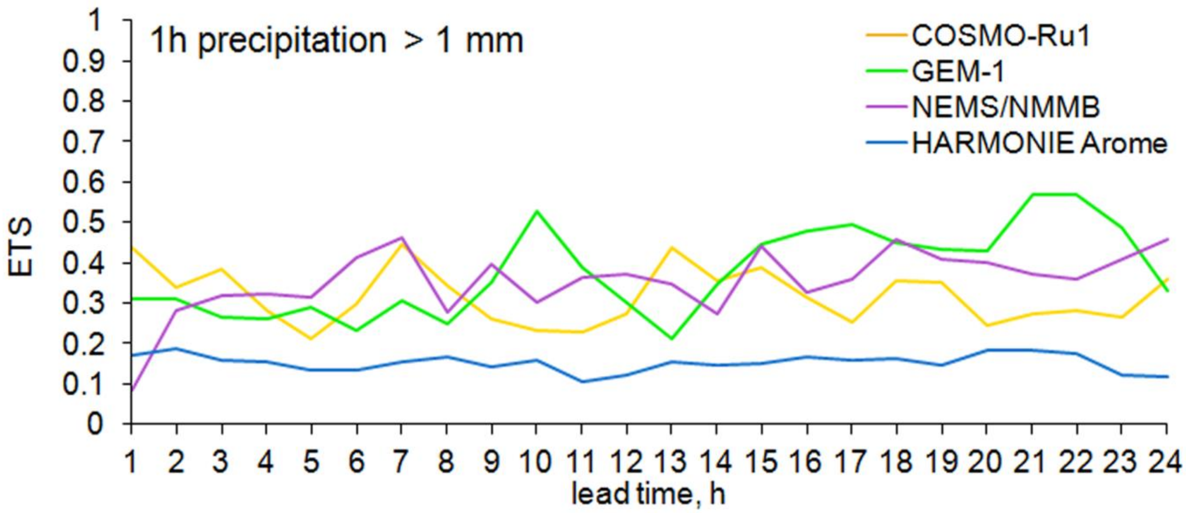

Figure 6. As in Fig. 5 but for the Equitable Threat Score of 1-h precipitation > 1 mm (the higher the better).

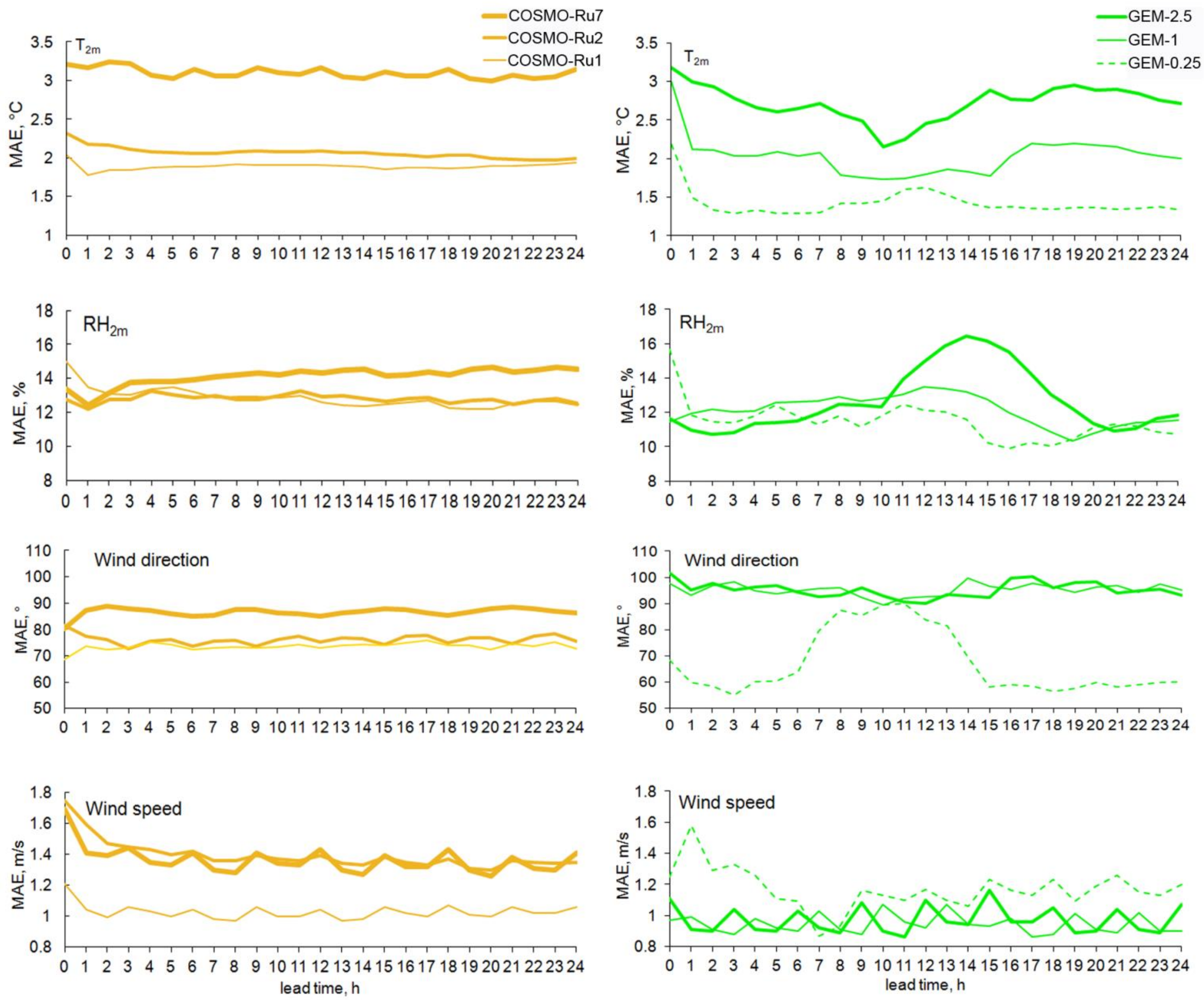

Figure 7. The role of the horizontal grid spacing for COSMO (left) and GEM (right) model families. Score: MAE. The scores are aggregated over all model runs (COSMO: 0000, 0600, 1200, and 1800 UTC; GEM-2.5: 2100 UTC, GEM-1: 2300 UTC, GEM-0.25: 0000 UTC) and over 22 stations in the mountain cluster.

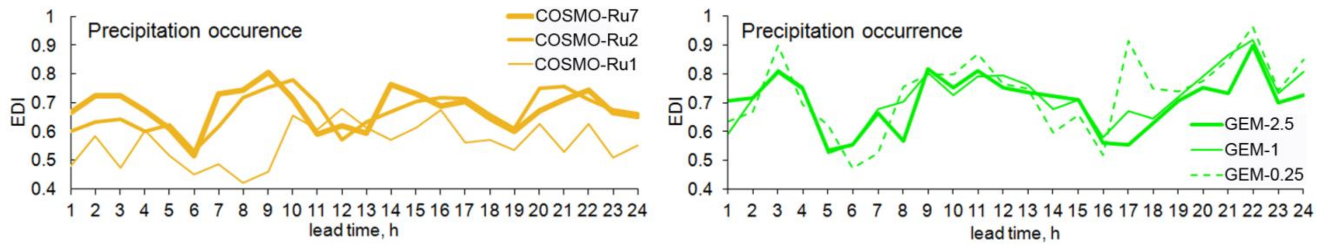

Figure 8. As in Fig. 7 but for 1-h precipitation occurrence forecasts. Score: Extremal Dependence Index (the higher the better)

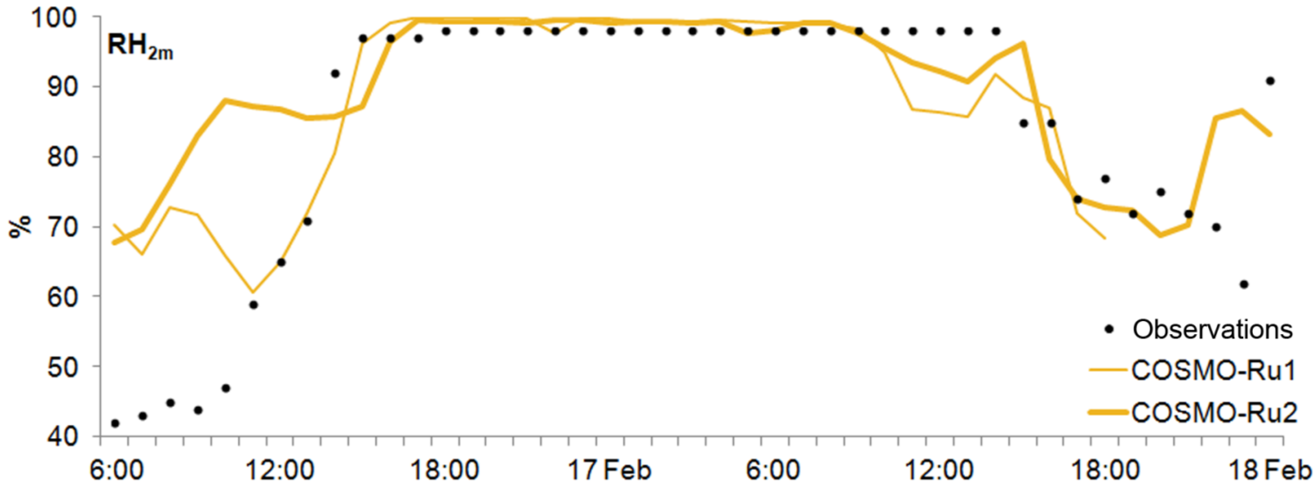

Figure 9. The RH_{2m} forecasts by COSMO-Ru2 and COSMO-Ru1 from 0600 UTC 16 February 2014 and corresponding observations for the low visibility event at the Biathlon stadium.

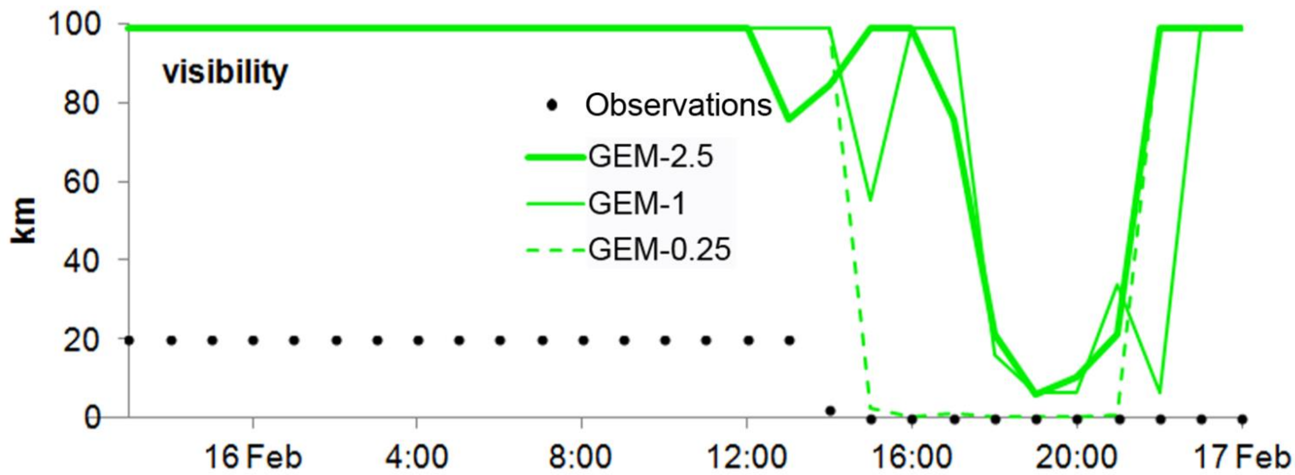

Figure 10. The visibility forecasts by GEM-2.5 (from 2100 UTC 15 February), GEM-1 (from 2300 UTC 15 February), GEM-0.25 (from 0000 UTC 16 February) and corresponding observations for the low visibility event at the Biathlon stadium. A model prediction of 100 km indicates unlimited visibility. The PWD sensors can report a maximum of 20 km visibility.

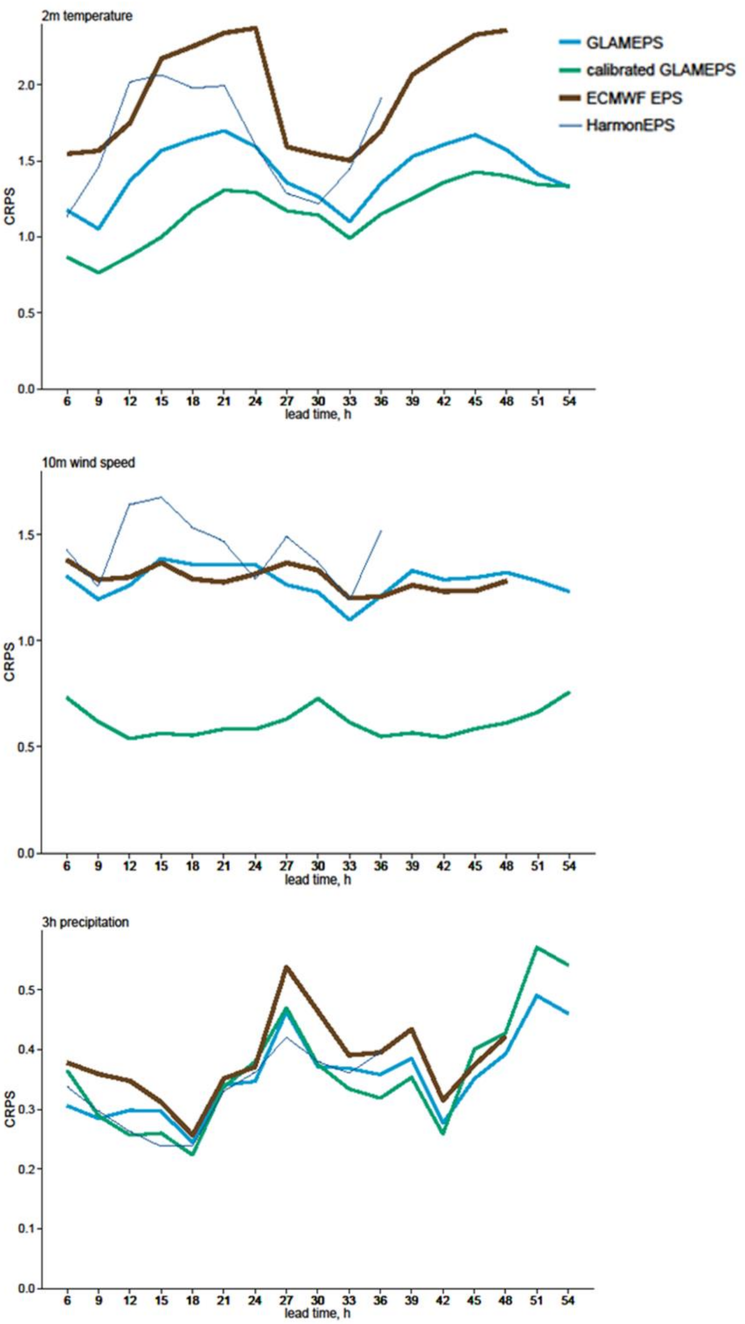

Figure 11. CRPS for ECMWF EPS, GLAMEPS, calibrated GLAMEPS, and HarmonEPS (the lower the better). Top: T_{2m}; Middle: 10-m wind speed; Bottom: 3-h precipitation.

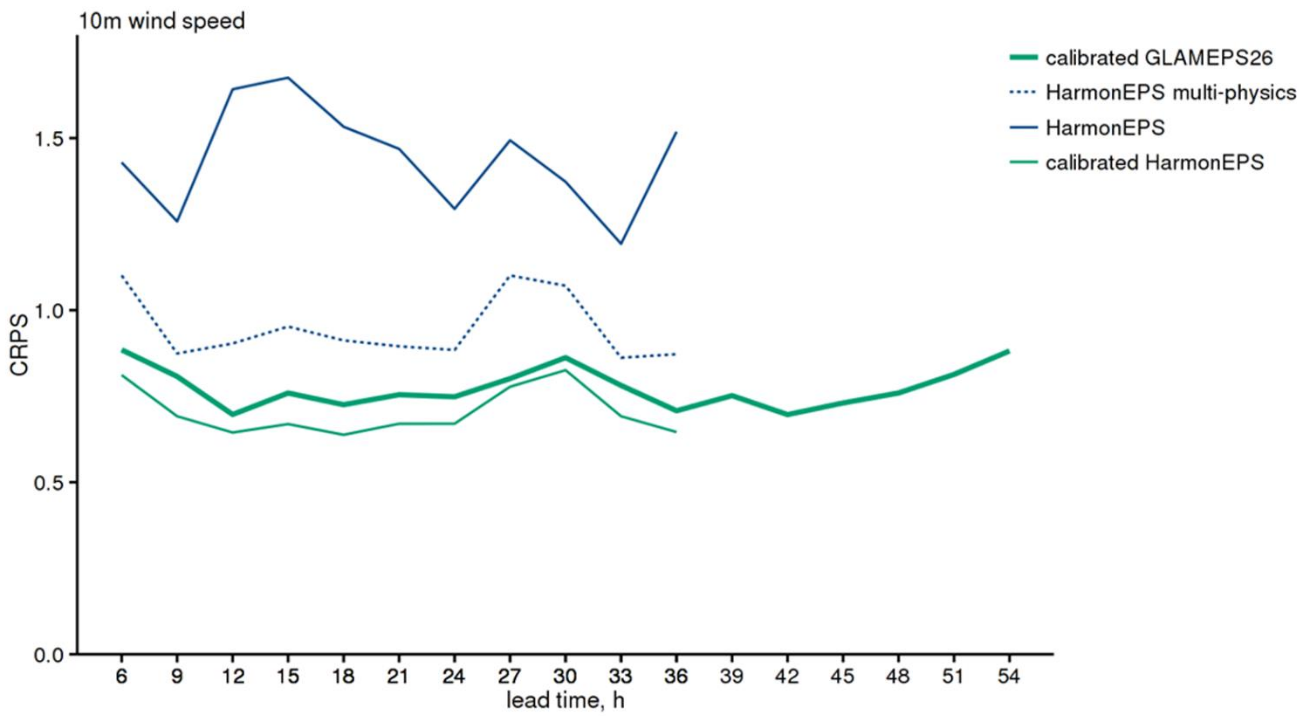

Figure 12. CRPS for 10-m wind speed forecasts for HarmonEPS, extended HarmonEPS with two sub-ensembles, calibrated HarmonEPS, and calibrated GLAMEPS based on 26 members only.

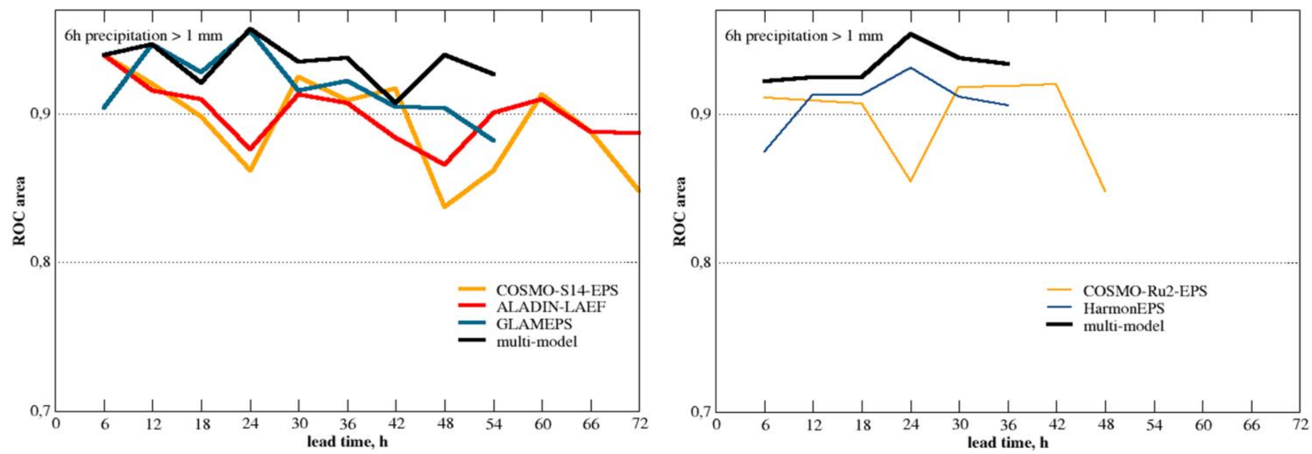

Figure 13. Area under the ROC curve (the higher the better) for forecasts of the event "6-h accumulated precipitation is above 1 mm" aggregated over the stations of the mountain cluster for convection-parameterized (left panel) and convection-permitting (right panel) EPSs as well as for the corresponding multi-model ensembles. Note that about 200 occurrences of the above event were observed during the verification period.

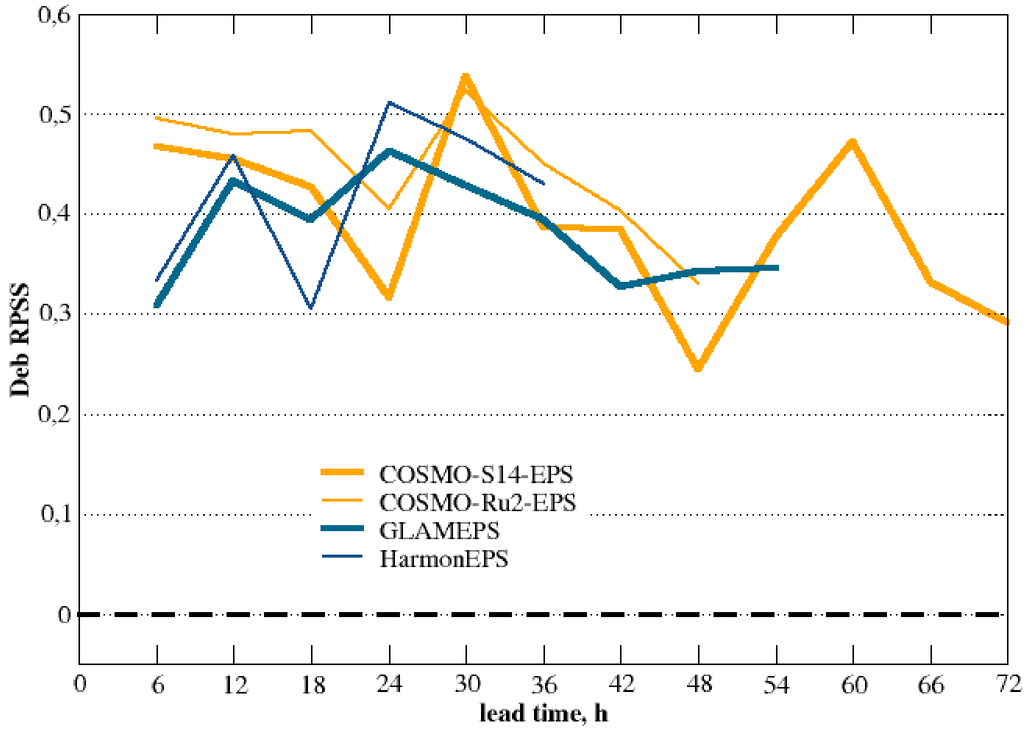

Figure 14. Debiased RPSS (the higher the better) for 6-h accumulated precipitation forecast by two convection-parameterized (COSMO-S14-EPS and GLAMEPS) and two convection-permitting EPSs (COSMO-Ru2-EPS and HarmonEPS), aggregated over the stations of the mountain cluster.

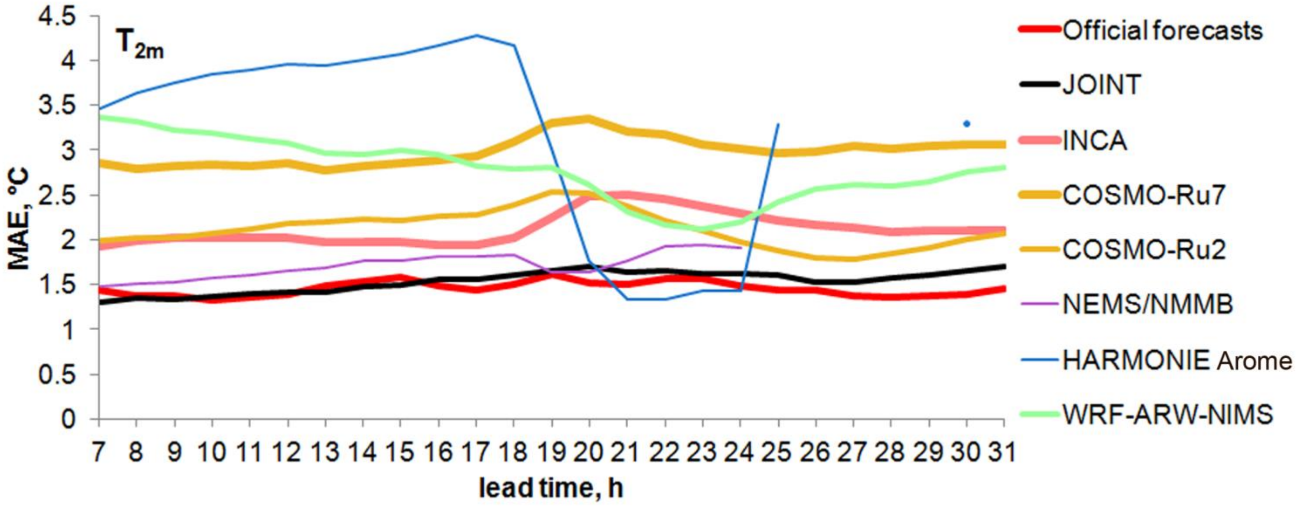

Figure 15. The skill of official and model forecasts as a function of the lead time. MAE of T_{2m} aggregated over the mountain cluster (heights of about 600, 1000, 1500, and 2000 m), period from 1 November 2013 to 23 February 2014 (HARMONIE Arome - from 9 December 2013, WRF-ARW-NIMS - from 23 December 2013), official forecasts issued at 1100 UTC, the models started at 1200 UTC. After 24 h lead time, the HARMONIE Arome forecasts were issued with 6-h step, that's whence the blue dot at 30 h lead time on the plot.

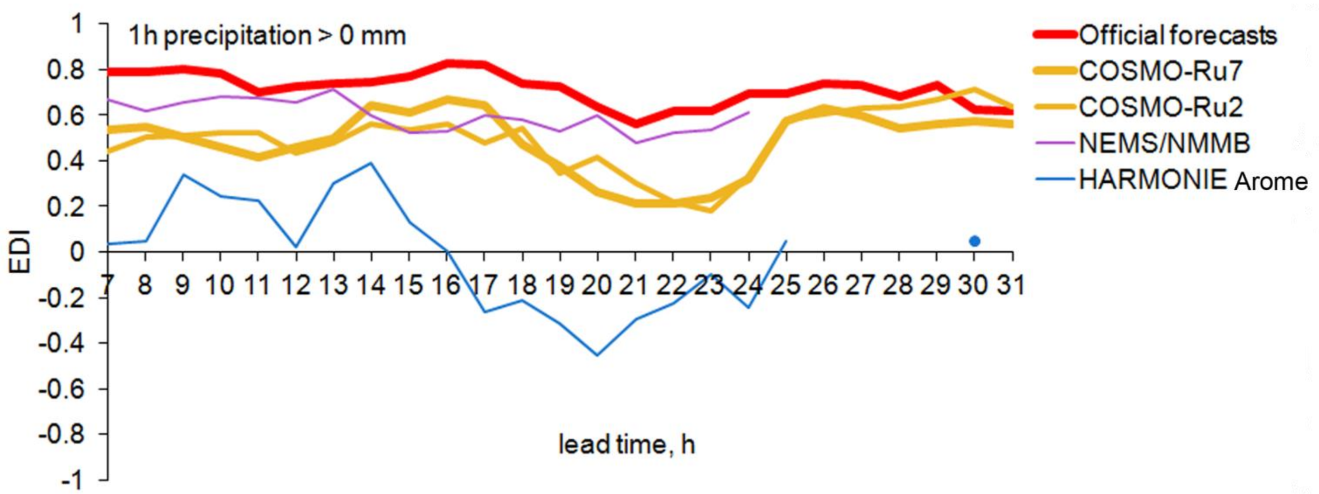

Figure 16. Same as Fig. 15, but for the Extremal Dependence Index (EDI, the higher the better) of 1-h precipitation occurrence.