SUPPLEMENT

ADDITIONAL APPLICATIONS OF THE D-PHASE DATASETS

BY MATHIAS W. ROTACH, PAOLO AMBROSETTI, FELIX AMENT, CHRISTOF APPENZELLER, MARCO ARPAGAUS, HANS-STEFAN BAUER, ANDREAS BEHRENDT, FRANÇOIS BOUTTIER, ANDREA BUZZI, MATTEO CORAZZA, SILVIO DAVOLIO, MICHAEL DENHARD, MANFRED DORNINGER, LIONEL FONTANNAZ, JACQUELINE FRICK, FELIX FUNDEL, URS GERMANN, THERESA GORGAS, CHRISTOPH HEGG, ALESSANDRO HERING, CHRISTIAN KEIL, MARK A. LINIGER, CHIARA MARSIGLI, RON MCTAGGART-COWAN, ANDREA MONTAINI, KEN MYLNE, ROBERTO RANZI, EVELYNE RICHARD, ANDREA ROSSA, DANIEL SANTOS-MUÑOZ, CHRISTOPH SCHÄR, YANN SEITY, MICHAEL STAUDINGER, MARCO STOLL, HANS VOLKERT, ANDRE WALSER, YONG WANG, JOHANNES WERHAHN, VOLKER WULFMEYER, AND MASSIMILIANO ZAPPA

In the main paper, four selected applications of Demonstration of Probabilistic Hydrological and Atmospheric Simulation of Flood Events in the Alpine region (D-PHASE) products and procedures are presented for high-resolution numerical modeling, ensemble forecasting, nowcasting, and model analysis. Here, the reader can find additional examples on how the D-PHASE model ensemble was used during the D-PHASE Operations Period (DOP). The examples explain the calibration of an Ensemble Prediction System (EPS) using the reforecasting technique, assimilation of global positioning system (GPS) data, and the implementation of a high-resolution EPS [Micro-PEPS (MPEPS)]. More applications can be found in Arpagaus et al. (2009) and Zappa et al. (2008).

USING REFORECASTS TO IMPROVE COSMO-LEPS. The Limited Area Prediction System (LEPS) is the local ensemble prediction system of the Consortium of Small Scale Modeling (COSMO) that is used to predict rare events at several operational centers (Montani et al. 2003; Walser et al. 2006). Here, we show that calibrating the forecast with reforecasts strongly improves the forecast skill. "Reforecasts" basically provide a model climatology; that is, the currently operational model version is used to produce a sufficiently large number of past realizations of the date in question (as they would have been forecasted had the current model version been in operation at the time). The improvement of calibration through reforecasting is mainly by enhancing the forecasts reliability (see, e.g., Hamill et al. 2006). A

calibration draws the forecasted probability toward the observed event frequency (Wilks 2006). Several calibration techniques (e.g., analog techniques, model output statistics) have been described (e.g., Hamill et al. 2004; Rajagopalan et al. 2002; Hamill and Whitaker 2006), and all of the techniques require observations covering the period and domain of the reforecast, which is a strong limitation for most model output parameters and areas. An alternative approach is the extreme forecast index (EFI) of the European Centre for Medium-Range Forecasts (ECMWF; Lalaurette 2003; Zsótér 2006). The EFI measures the extremity of the ensemble forecast relative to model climate. It can be calculated for every model output parameter and location, and it potentially corrects for systematic model errors. However, it is ambiguous because it combines properties of the forecast and climate distribution function in one number. A more direct approach for a probabilistic measure of extreme forecasts using reforecasts, thus calibrating the forecast, is presented here for the example of 24-h precipitation sums.

COSMO-LEPS is a 16-member EPS with an approximately 10-km grid size and 40 levels over Europe, with a 132-h lead time, driven by selected members from the ECMWF EPS. One-member reforecasts were generated from 1971 to 2000 with undisturbed boundary conditions. On each day a 42-h forecast was initialized at 1200 UTC using initial and boundary conditions from the 40-yr ECMWF Re-Analysis (ERA-40). To calibrate a forecast with reforecasts from the actual season, a monthly subset ±14 days around the actual date was used. For each grid point, this yields a total of 870 (30 years × 29 days) days for calibration.

From the model, climatology return levels (RL) for different return periods (RPs) were estimated. For events likely to happen several times within the time covered by the reforecasts, RLs were estimated from quantiles from the model climate. The fraction of forecast members exceeding a RL then gives the probabilistic, calibrated information on the severity of an event. Assuming that the recurrence times of forecasted and observed events are equal, expressing the forecast as RPs improves its reliability without requiring observations.

Forecasted probabilistic RPs are presented as 2D plots for specific lead times and return periods (Fig. S1). They are similar to the usual EPS products by showing the probability to exceed some threshold. Here, the threshold is an RP derived from the model climate, that is, more reliable but not in absolute terms. If warning levels at meteorological offices are
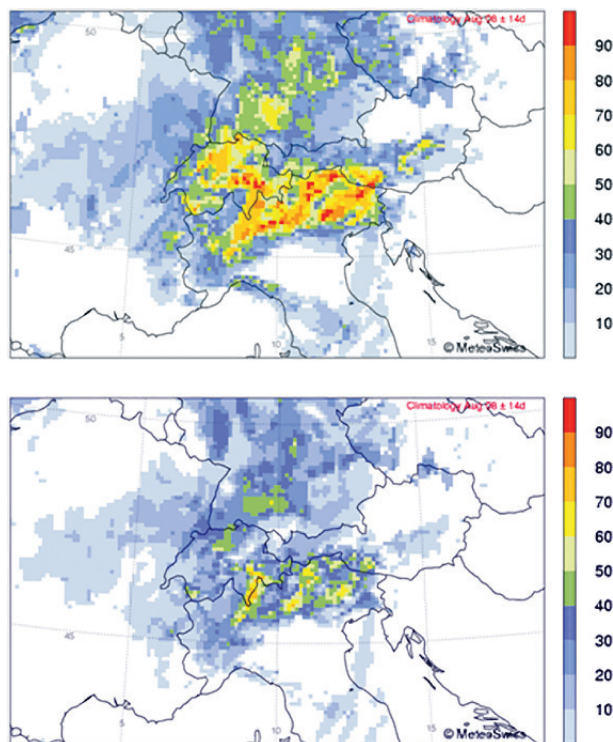

Fig. S1. Calibrated **COSMO-LEPS 24-h rainfall forecast probabilities (%) to exceed events with recurrence times of (top) twice per Aug and (bottom) every second Aug for the flooding event of 8 Aug 2007 in the region north of the Alps.**

based on RPs of events, then this product can easily be adapted to show the probability to reach a warning level based on a calibrated forecast system.

The new warning product was verified during the DOP using observations over Switzerland. Twenty-four-hour rainfall was interpolated to the COSMO-LEPS grid according to Frei and Schär (1998) and Frei et al. (2006). The probability to exceed an RP was calculated analogously to the warning product. For verification, the categorical debiased Brier skill score (BSSD) was used for dichotomous events (Weigel et al. 2007). In comparison to the uncalibrated probability to exceed, an RP (estimated from the observed, not the model, climate) shows that the model calibration strongly improves skill scores (Fig. S2) mainly as a result of improved reliability. For the events and period verified here, forecast quality is 1–2 days better. Less frequent events seem to profit more from calibration. Even forecasts with initially no significant skill are skillful after calibration. The presented method is likely to improve the forecast skill over the entire model domain without requiring observations. These results should encourage the operational implementation of reforecasts.
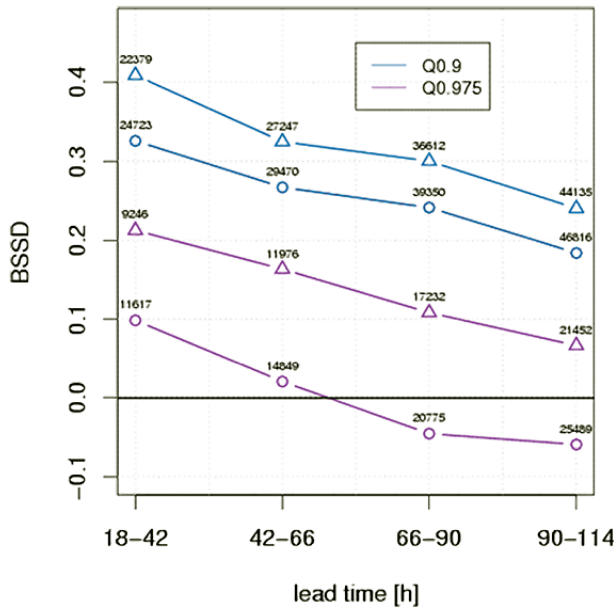
Fig. S2. BSSD for the uncalibrated (circles) and calibrated (triangles) COSMO-LEPS 24-h precipitation forecast during the DOP. Event thresholds are 10 days (blue, 90% quantile) and 40 days (purple, 97.5% quantile). Numbers indicate the number of forecasted events.

OPERATIONAL ASSIMILATION OF GPS SLANT PATH DELAYS USING THE MM5 4DVAR SYSTEM DURING D-PHASE AND COPS.

The importance of an accurate initialization of water vapor for numerical weather prediction has been underlined (e.g., Weckwerth et al. 2004; Wulfmeyer et al. 2006; Kawabata et al. 2007). Although the number of observations used in data assimilation is constantly rising, large data gaps are present, especially for quantities of the hydrological cycle on the mesoscale. Systematic initialization errors are, apart from deficiencies in the physical parameterizations, thought to be among the main reasons for forecast errors in general and errors in quantitative precipitation forecasts (QPFs).

With recent advances in GPS, ground-based GPS receivers have become important in providing water vapor measurements operationally at low cost (e.g., Bender et al. 2008). GPS slant path observations contain information concerning the water vapor field with high temporal resolution and spatial coverage under all weather conditions.

The present study uses the fifth-generation Pennsylvania State University–National Center for Atmospheric Research Mesoscale Model (MM5) four-dimensional variational analysis (4DVAR) scheme (Ruggiero et al. 2002). It requires adjoints of the used parameterizations, which are only available for simpler

schemes. Therefore, simplified physics were used for the 4DVAR assimilation on a coarse domain (18-km resolution). A sophisticated physics package was then used for the forward simulations in three two-way interactive nesting steps with 18-, 6-, and 2-km horizontal resolutions and 36 vertical levels up to 100 hPa. In the innermost domain, the convection parameterization was switched off. A forward operator for GPS slant path delay data and its adjoint were implemented into the MM5 4DVAR, and a real-time forecast system was set up in the framework of D-PHASE and the Convective and Orographically-Induced Precipitation Study (COPS). It provided two forecasts each day: one driven by ECMWF forecasts only and one modified by the 4DVAR. GPS data were provided by the GFZ Helmholtz Centre Potsdam and Met Office.

Figure S3 shows a first statistical analysis of the mean diurnal cycles of precipitation in the COPS domain. The 4DVAR has a wet bias in the beginning of the free forecast after a 3-h data assimilation window. This is due to two reasons. First, the only parallelized convection scheme, for which an adjoint is available (i.e., that of Anthes–Kuo), was used in the 4DVAR. This tends to overestimate precipitation (Grell et al. 1995). Second, horizontal advection was calculated along sigma levels, leading to erroneous temperature and moisture transports in mountainous regions, thus triggering convection (e.g., Zängl 2002; Zus et al. 2008). However, the longer the forecast range, the closer the 4DVAR simulation is to the observed precipitation and becomes superior to the control simulation after about 3-h free forecast. This indicates a positive effect of the assimilation.

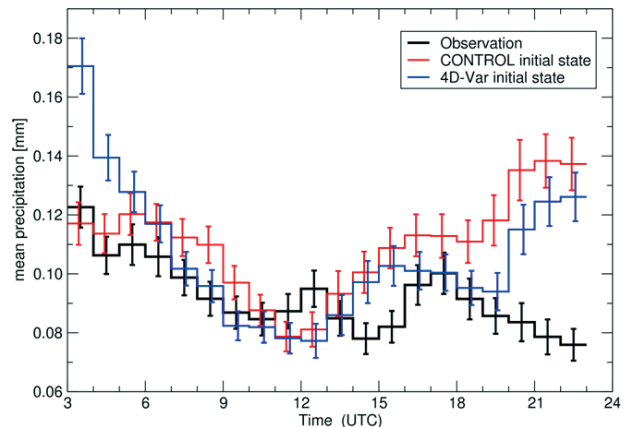As a consequence of these findings, solutions for both problems mentioned above have already been

Fig. S3. Mean daily cycles of precipitation (mm) in the COPS domain for Aug–Oct 2007: Observations (black), MM5 control simulation (red), and MM5 forecasts after 4DVAR assimilation (blue).

implemented. The Anthes–Kuo scheme is replaced by the Grell scheme and a simple horizontal diffusion along z levels is introduced. The first case study simulations indicate that the wet bias in the beginning of the forecast is removed (Zus et al. 2008).

THE MICRO-PEPS: HIGH-RESOLUTION POOR MAN'S ENSEMBLE.
One major goal of the D-PHASE DOP has been to operate very high-resolution models at the convection-permitting scale. Because forecasting on smaller spatial and temporal scales becomes increasingly influenced by, for example, stochastic physical processes, information on forecast uncertainty from a very high-resolution EPS would be desirable. One obvious approach in the absence of a true very high-resolution EPS is the construction of a "poor man's ensemble" (Denhard and Trepte 2006) by grouping some of the very high-resolution D-PHASE models into a multimodel prediction system. This MPEPS included COSMO-2, COSMO-DE, AROME, CMCGEMH, and ISACMOL2 (Table 1 from the main text). Although the former two models generated forecasts every 3 h, the other models ran less frequently. To increase the ensemble size, lagged average forecasts were added to the MPEPS. It was assumed that forecasts initialized up to 6 h earlier responded to similar synoptic conditions and might add value to a convection-permitting ensemble. If t is the initialization time of the ensemble forecast, then three runs each of the models COSMO-2 and COSMO-DE (t, $t – 3h$, and $t – 6h$) and the single deterministic forecasts from AROME, CMCGEMH, and ISACMOL2, which were provided within the 6-h lagged average time window, made up the MPEPS. Note that the ensemble size of MPEPS was sometimes smaller than nine members as a result of the availability of its constituents. Four runs at 0000, 0600, 1200, and 1800 UTC were performed daily.

The ensemble mean fields were determined for all variables, as defined in the D-PHASE implementation plan (including 2-m temperature,

10-m wind, and accumulated precipitation). In addition, alerts were submitted to the Visualization Platform (VP), and exceeding probabilities for precipitation and snow were calculated and displayed. Figure S4 illustrates the type of product one can obtain from the MPEPS. Because of the high resolution of the contributing models, maps of exceeding probabilities (e.g., for 3-h accumulated precipitation) show extremely fine structure. Detailed analysis of the results will have to show to what extent this fine structure is indeed more realistic. For interpreting MPEPS results, note that ensemble size varies across the domain as a result of different domains of contributing models and that the incorporation of lagged forecasts limited the forecasting horizon.

The MPEPS design for D-PHASE was a special configuration that may not necessarily yield the most valuable probabilistic forecasts. Rather, it tried to extract the maximum information from the many convection-permitting models available to explore the prospects for ensemble techniques at this scale. The MPEPS may be recalculated at any time using the output of its components stored in the data archive. It performs such "reanalyses" for selected cases or periods to
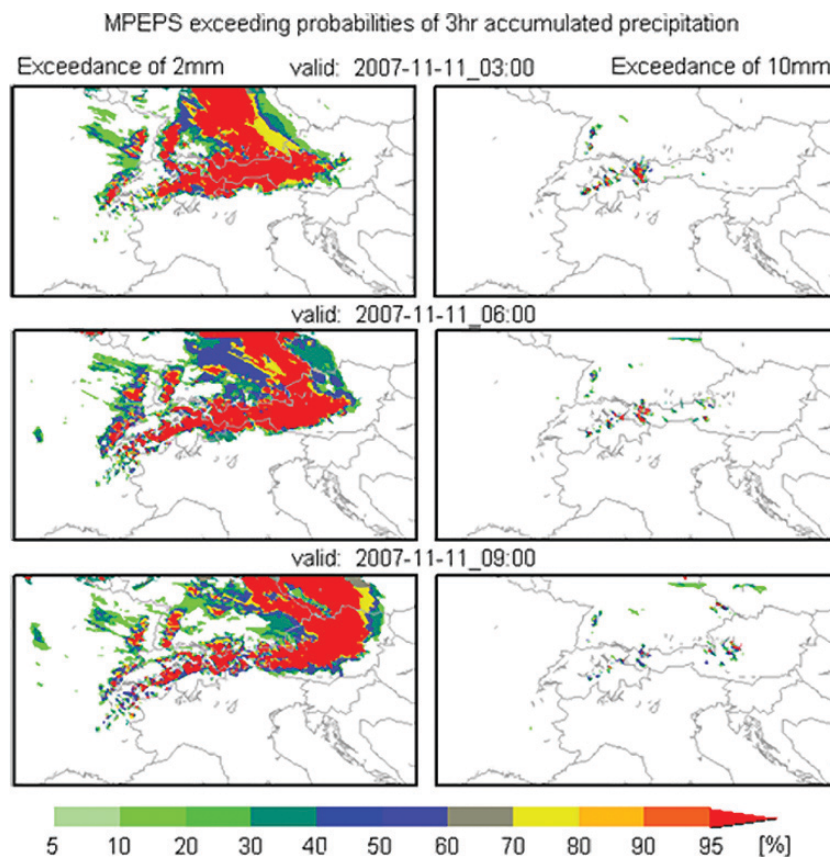

Fig. S4. Probability of exceeding (left) 2 mm in 3-h accumulated precipitation and (right) 10 mm for different validation times from the MPEPS initialized at 0000 UTC 11 Nov 2007.

optimize the poor man's ensemble approach, based on verification of its forecast skill, especially as compared to other ensemble prediction systems.

REFERENCES

Arpagaus, M., and Coauthors, 2009: MAP D-PHASE: Demonstrating forecast capabilities for flood events in the Alpine region. MeteoSwiss Publication Series 78, 75 pp. [Available online at www.meteoschweiz.admin.ch/web/de/forschung/publikationen/meteoschweiz_publikationen/veroeffentlichungen.html.]

Bender, M., G. Dick, J. Wickert, T. Schmidt, S. Song, G. Gendt, M. Ge, and M. Rothacher, 2008: Validation of GPS slant delays using water vapour radiometers and weather models. *Meteor. Z.,* **17,** 807–812.

Denhard, M., and S. Trepte, 2006: Calibration of the European multi-model ensemble SRNWP-PEPS. *Proc. Second THORPEX Int. Science Symp.,* WMO Tech. Doc. 1355, WWRP/THORPEX 7, Landshut, Germany, THORPEX, 56–57.

Frei, C., and C. Schär, 1998: A precipitation climatology of the Alps from high-resolution rain-gauge observations. *Int. J. Climatol.,* **18,** 873–900.

——, R. Schöll, S. Fukutome, J. Schmidli, and P. L. Vidale, 2006: Future change of precipitation extremes in Europe: Intercomparison of scenarios from regional climate models. *J. Geophys. Res.,* **111,** D06105, doi:10.1029/2005JD005965.

Grell, G. A., J. Dudhia, and R. R. Stauffer, 1995: A description of the fifth-generation Penn State/NCAR mesoscale model (MM5). NCAR Tech. Note NCAR/TN-398+STR, 122 pp.

Hamill, T. M., and J. S. Whitaker, 2006: Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and application. *Mon. Wea. Rev.,* **134,** 3209–3229.

——, ——, and X. Wei, 2004: Ensemble reforecasting: Improving medium-range forecast skill using retrospective forecasts. *Mon. Wea. Rev.,* **132,** 1434–1447.

——, ——, and S. L. Mullen, 2006: Reforecasts: An important dataset for improving weather predictions. *Bull. Amer. Meteor. Soc.,* **87,** 33–46.

Kawabata, T., H. Seko, K. Saito, T. Kuroda, K. Tamiya, T. Tsuyuki, Y. Honda, and Y. Wakazuki, 2007: An assimilation and forecasting experiment of the Nerima heavy rainfall with a cloud-resolving nonhydrostatic 4-dimensional variational data assimilation system. *J. Meteor. Soc., Japan,* **85,** 255–276.

Keenan, T., and Coauthors, 2003: The Sydney 2000 World Weather Research Programme Forecast Demonstration Project: Overview and current status. *Bull. Amer. Meteor. Soc.,* **84,** 1041–1054.

Lalaurette, F., 2003: Early detection of abnormal weather conditions using a probabilistic extreme forecast index. *Quart. J. Roy. Meteor. Soc.,* **129,** 3037–3057.

Montani, A., and Coauthors, 2003: Operational limited-area ensemble forecasts based on the "Lokal Modell." *ECMWF Newsletter,* No. 98, ECMWF, Reading, United Kingdom, 2–7.

Rajagopalan, B., U. Lall, and S. E. Zebiak, 2002: Categorical climate forecasts through regularization and optimal combination of multiple GCM ensembles. *Mon. Wea. Rev.,* **130,** 1792–1811.

Ruggiero, F., G. Modica, T. Nehrkorn, M. Cerniglia, J. Michalakes, and X. Zou, 2001: A MM5-based four-dimensional variational analysis system developed for distributed memory computers. *Proc. 14th Conf. on Numerical Weather Prediction,* Fort Lauderdale, FL, Amer. Meteor. Soc., J168–J171.

Walser, A., M. Arpagaus, C. Appenzeller, and M. Leutbecher, 2006: The impact of moist singular vectors and horizontal resolution on short-range limited-area ensemble forecasts for two European winter storms. *Mon. Wea. Rev.,* **134,** 2877–2887.

Weckwerth, T. M., and Coauthors, 2004: An overview of the International H_2O Project (IHOP_2002) and some preliminary highlights. *Bull. Amer. Meteor. Soc.,* **85,** 253–277.

Weigel, A. P., M. A. Liniger, and C. Appenzeller, 2007: The discrete Brier and ranked probability skill scores. *Mon. Wea. Rev.,* **135,** 118–124.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences.* 2nd ed., International Geophysical Series, Vol. 91, Academic Press, 627 pp.

Wulfmeyer, V., H.-S. Bauer, M. Grzeschik, A. Behrendt, F. Vandenberghe, E. V. Browell, S. Ismail, and R. Ferrare, 2006: Four-dimensional variational assimilation of water vapor differential absorption lidar data: The first case study within IHOP_2002. *Mon. Wea. Rev.,* **134,** 209–230.

Zängl, G., 2002: An improved method for computing horizontal diffusion in a sigma-coordinate model and its application to simulations over mountainous topography. *Mon. Wea. Rev.,* **130,** 1423–1432.

Zappa, M., and Coauthors, 2008: MAP D-PHASE: Real-time demonstration of hydrological ensemble prediction systems. *Atmos. Sci. Lett.,* **9,** 80–87, doi:10.1002/asl.183.

Zsótér, E., 2006: Recent developments in extreme weather forecasting. *ECMWF Newsletter,* No. 107, ECMWF, Reading, United Kingdom, 8–17.

Zus, F., M. Grzeschik, H. S. Bauer, V. Wulfmeyer, G. Dick, and M. Bender, 2008: Development and optimization of the IPM MM5 GPS slant path 4DVAR System. *Meteor. Z.,* **17,** 867–885.