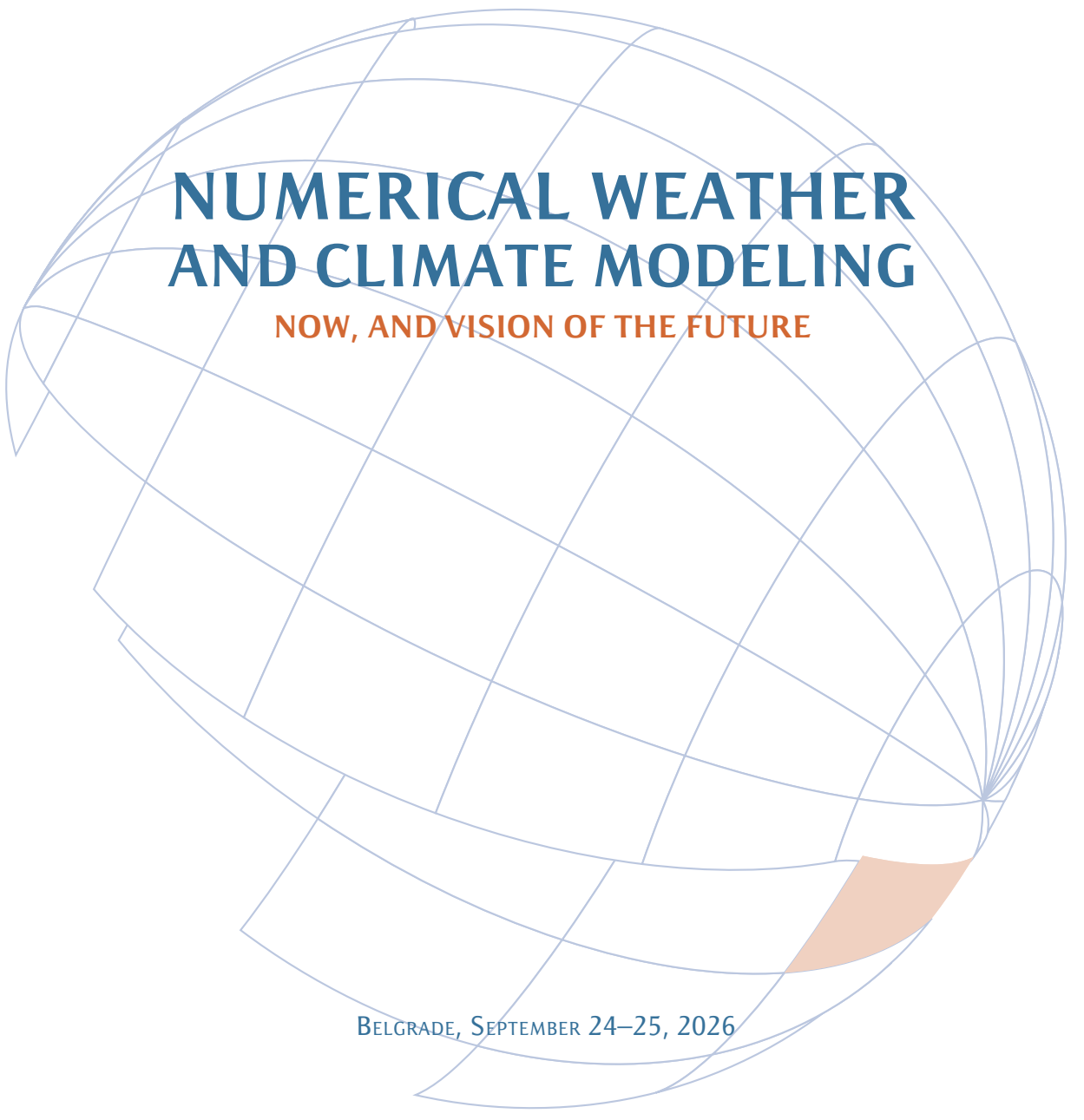




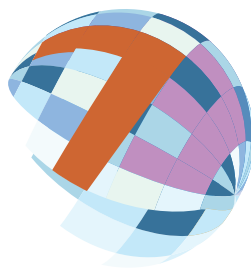
SERBIAN ACADEMY OF SCIENCES AND ARTS

A large, stylized globe made of thin blue lines representing latitude and longitude. A small portion of the globe at the bottom right is shaded in a light orange color.

NUMERICAL WEATHER AND CLIMATE MODELING

NOW, AND VISION OF THE FUTURE

BELGRADE, SEPTEMBER 24–25, 2026



ORGANIZING COMMITTEE:

Fedor Mesinger

Zoran Knežević

Serbian Academy of Science and Arts, Serbia

MEMBERS:

Jugoslav Nikolić

Katarina Veljović Koračin

Ana Vuković Vimić

Sin Chan Chou

Miodrag Rančić

Jürgen Steppeler

Numerical Weather and Climate Modeling: Now, and Vision of the Future

Belgrade, September 24, 2026

10:00–10:10 Opening addresses

10.10–10.40 **What cut-cell Eta features additional to its intersecting the topography are responsible for its successful performance?**

Fedor Mesinger, Jorge L Gomes , Dušan Jović ,
Katarina Veljović Koraćin

10.40–11.10 **Eta Model at all scales**

Sin Chan Chou, Jorge Luís Gomes, André A. Lyra,
Isabel D. Pilotto, Daniela Carneiro Rodrigues, Daniel Andrés
Rodriguez, Priscila da Silva Tavares, Gustavo Sueiro Medeiros,
Diego de Andrade Campos, Diego José Chagas, Fedor Mesinger

11.10-11.30 Coffee break

11.30–12.00 **Graph Machine-Learning Data Assimilation on a Global Icosahedral Grid: Accuracy, Physical Constraints, and Observation-Specific Control**

Miodrag Rančić

12.00-12.30 **Multigrid Graph Filter for Unstructured Finite-Volume Meshes**

Miodrag Rančić, R. James Purser, and Manuel S. F. V. De Pondecá

12.30–13.00 **AI-Driven Optimisation of an Adaptive Mesh Framework in the Fluidity–Atmosphere Model**

Jinxi Li, P. Gan, F. Fang, X. Wu, J. Zhu, Z. Wang, M. Zhu, X. Zou

13:00–14:00 Lunch break

14.00–14.30 Ensemble forecasting of radiational fog over northern France using a non-hydrostatic Eta model

Dragana Mujković, Katarina Veljović Koračin, Darko Koračin

14.30–15.00 Sparse L₂ Galerkin irregular meshes of high approximation order and their simple implementation in arbitrary existing models using the method of geometric files

Jürgen Steppeler

15.00–15.30 A Unified Solver for Global and Regional Atmospheric Simulations

Xindong Peng, Siyuan Chen, Xiaohan Li, Yifan Zhao, Jun Chang

15.30-16.00 Poster session and Coffee break

Diagnosis and Numerical Modeling of an Explosive Cyclone over the Northwestern Pacific

Baitang Sun, Pengyuan Li, Shuqin Zhang, Jingtian Guo, and Gang Fu



Numerical Weather and Climate Modeling: Now, and Vision of the Future

Belgrade, September 25, 2026

10.00-10.30 SL-AV global atmosphere model: participation in WP-MIP project and further model development

Mikhail Tolstykh, Rostislav Fadeev, Vladimir Shashkin, Gordey Goyman, Radomir Zaripov, Ekaterina Biryucheva

10.30-10.50 Coffee break

10.50-11.20 Rainfall-Informed SWMM–LSTM Surrogate Modelling for Storm Sewer Network Flow Forecasting within a Digital Twin Framework

Haoyue Wang, Miloš Stanić, Jasmina Moskovljević, Čedo Maksimović

11.20-11.50 Coupling numerical precipitation forecasts with flood-oriented Local Digital Twins: The Kolubara-Tamnava case

Jelena Batica, Luka Stojadinović, Marija Milovanović, and Nikola Milivojević

11.50-12.50 Lunch break

12.50-13.20 Numerical Modeling Study of “Millipede Clouds”

Gang Fu, Zhongshi Wang, Xinyu Zhang, Weikang Sun, Pengyuan Li

13.20-13.50 Rain Gauge–Radar Merging Improves Short-Term Rainfall Nowcasts and the Sewer Overflow Forecasts

Surya Prayoga and Čedo Maksimović

13.50-14.20 Combining machine learning and physically based models in the forecasting of extreme weather events

Ivan Ristić, Ivana Kordić



ABSTRACTS

What cut-cell Eta features additional to its intersecting the topography are responsible for its successful performance?

Fedor Mesinger¹, Jorge L Gomes², Dusan Jovic³,
Katarina Veljović Koračin⁴

¹ Serbian Academy of Sciences and Arts, Belgrade, Serbia (fedor.mesinger@gmail.com)

² National Institute for Space Studies (INPE), Cachoeira Paulista, SP, Brazil
(jorgeluisgomes@gmail.com)

³ Lynker at NCEP Environmental Modeling Center (EMC), College Park, MD, USA
(dusanjovic@gmail.com)

⁴ Faculty of Physics, University of Belgrade, Belgrade, Serbia (katarina@ff.bg.ac.rs)

INTRODUCTION

Primarily due to the problem identified by Gallus and Klemp (2000), at EMC the Eta model was essentially frozen in 2002. However, after several years dedicated to its planned replacement by NMM/GSI,

the results of the four+ month EMC “parallel” test in 2006 showed verification scores favoring the Eta. Note, e.g., Fig. 4 of Mesinger, Veljovic (2017):

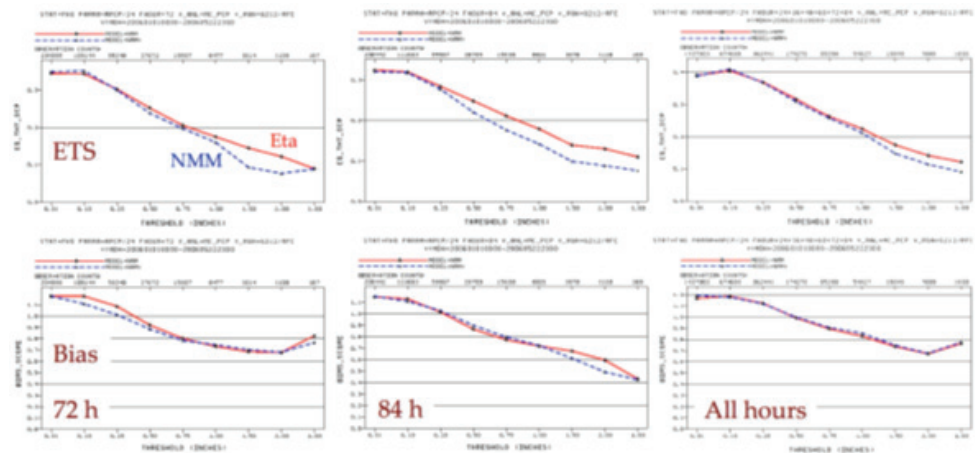


Fig. 4 24-h precipitation Equitable Threat Scores (upper panels) and Bias Scores (lower panels) of the Eta model/EDAS (red) and WRF-NMM/GSI (blue), of the 1 January–22 May 2006 parallel, run at 12-km grid spacing. 24-h precipitation thresholds are increasing from 0.01 to 3 in/24 h along the abscissas of the plots. Verifications at 72 h

(left), 84 h (middle), and combined 24, 36, 48, 60, 72, and 84 h (right). Note that if biases of two models are the same, the model with higher ETS is more accurate in placing its precipitation (e.g., Hamill 1999; Mesinger 2008). After DiMego (2006)

Nevertheless, the Gallus–Klemp problem needed to be addressed. Since Gallus and Klemp identified step-corners to be the reason of the problem, this was accomplished by changing from step

topography to “sloping steps,” now generally referred to as cut-cells, with no step-corners. The change implemented is illustrated by the following 2D schematic:

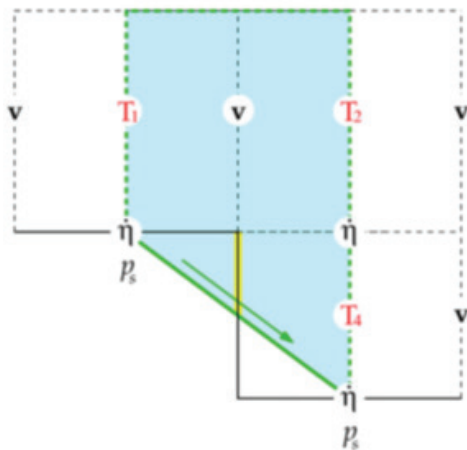


Figure 3 Schematic illustrating the “sloping steps” Eta discretization. The vertical topography side below the velocity (v) cell in the middle is replaced by a sloping side going from the left to the right surface pressure (ps) points. Symbols T denote temperature cells, η denotes the eta coordinate, with dot ($\cdot$) above a symbol standing for the time derivative. From Mesinger et al. (2012).

Irrespective of this the Eta continued to perform well. In an ensemble experiment the Eta members driven by ECMWF (EC) members using for the first 10 days about the same resolution achieved better scores for 250 hPa winds stronger than 45 m s^{-1} than their EC drivers.

A puzzling byproduct of that experiment was that the Eta ensemble switched to sigma still achieved these wind scores better than their driver members, although to a lesser extent. Unique features of the Eta contributing to this advantage include its “fairly well-posed” lateral boundary conditions (LBCs), introduction of the finite-volume vertical and slantwise advection, and the use of realistic topography without smoothing. The difference of the ETSa scores (Equitable Threat Score adjusted to unit bias,

Mesinger 2008) of the Eta are presented here or displayed following the introduction of these features.

LATERAL BOUNDARY CONDITIONS (LBCs)

Over the years, two experiments were performed with the limited area Eta ensembles driven by ECMWF (EC later on) ensembles, in which the Eta members for the most part showed more accurate placement of upper tropospheric jet streams east of the Rockies than their driver EC members. The first of those, reported in Mesinger and Veljovic (2017, 2020) resulted in mostly more accurate verifications of the upper tropospheric winds stronger than 45 m s^{-1} than those of the driver EC members. Note that to achieve this the Eta members need to do better than just overcome the errors due to interpolations needed to arrive to the nested Eta lateral boundary values, since these are required for the Eta LBC scheme.

The second of those ensemble tests is the six-member ensemble reported on and analyzed by Mesinger and Veljovic (2013), expanding on the earlier three-member test of Veljovic et al. (2010). But the problem with the 2013 analysis is that the only verification scores of all three ensembles, the Eta, the Eta using relaxation LBCs, and the EC driver ensemble, were shown on a single line plot as a function of time, so that with three zig-zag lines, often crossing and/or on top of each other, it was not easy to see how long, or how many times, one ensemble was more accurate than another. Therefore, we have generated *one* curve, of the difference of the Equitable Threat Score adjusted to

unit bias of the Eta and the Eta using relaxation LBCs. This difference is shown in the following figure:

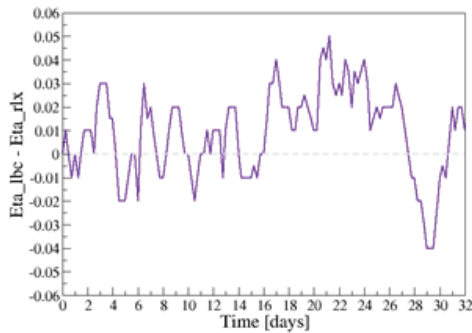


Figure. The difference of the ETs scores of the Eta ensemble members, and the same ensemble except for using the relaxation LBCs. Valid times are 32 days January 1 through February 1, 2009. The average of the 128 values shown is 0.009. For additional detail see Veljovic et al. (2010) and Mesinger and Veljovic (2013).

FINITE-DIFFERENCE VS FINITE-VOLUME VERTICAL AND SLANTWISE ADVECTION

A problem that received a fair amount of attention is the design of “high ac-

curacy” schemes, even though apparently no improved skill has ever been demonstrated from using such schemes in “prediction” NWP models. Note the final sentences of Li et al. (2025). What seems not to have been noticed is that column-wise physical parameterizations generate grid point to grid point noise, which invalidates Taylor series expansion of difference schemes. Therefore, no benefit from “higher accuracy” schemes should not have been a surprise.

A case of a downslope windstorm, so called “zonda”, east of the Andes, showed a benefit we had from replacing a finite difference scheme, by finite-volume vertical and slantwise schemes. The finite-volume schemes, replacing a centered Lorenz-Arakawa advections, achieved a stronger and a more accurate placement of the foehn-type warming east of the Andes, shown in the right compared to the left plot of the figure (from Mesinger 2023):

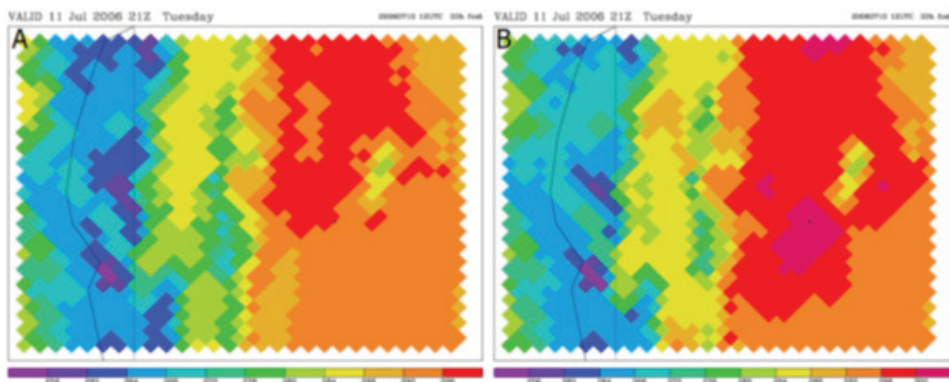


Figure 6 Forecast lowest cell temperatures (K) at 33 h of the case discussed in Section 9 of (Mesinger et al. 2012): A. Result obtained using Equation 3 for both the slantwise and the vertical advection; B. Result with these advections replaced by the finite-volume versions. The roughly vertical line on the left sides of the plots is the Chile-Argentina border, while the straight line is the 70°W meridian. The small cross to the right of the centers of plots shows the place of the San Juan station.

THE TOPOGRAPHY

The sophisticated topography, without smoothing, also achieved consistency with physics parameterizations by removing reliance on Taylor-series type accuracy, not requiring any grid-box to grid-box smoothness. To design topography first silhouette and mean values for topography cells were obtained. Silhouette values were defined so as to depend on the lowest wind directions. Next, the Laplacian of the 3x3 mean value was calculated.

If it is positive, the mean value is taken as the grid-box value. Finally, the silhouette values are checked not to have chosen a gap existing in the 3x3 mean values in any of the four directions. If so, the silhouette value is replaced by the mean value.

The W-E topography in this way obtained crossing the Andes at the latitude of the San Juan station is shown in the next figure. Note the size of the cliffs, and no noise in the 33-h forecast with isotherms that were not smoothed.

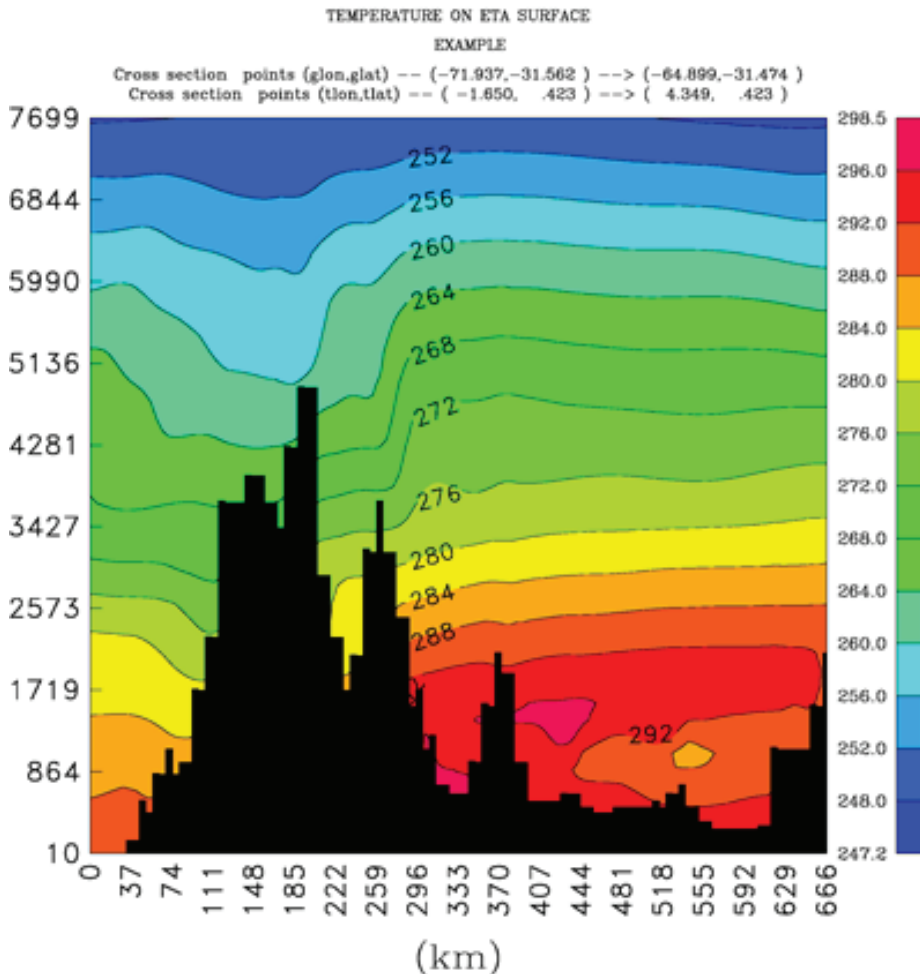


Figure 7. Eta 8-km topography, crossing the Andes at the San Juan station. Elevation, m, is on the left side of the plot, and isotherms are denoted in degrees K. (From Mesinger 2023)

REFERENCES:

- DiMego G., (2006) WRF-NMM and GSI Analysis to replace Eta Model and 3DVar in NAM Decision Brief. <http://nomads.ncdc.noaa.gov/docs/WRF-NMM-replacing-Eta-in-NAM.pdf>
- Gallus Jr., W.A. & Klemp, J.B. 2000, 'Behavior of flow over step orography', *Mon. Weather Rev.*, 128, 4, 1153-64, doi:10.1175/1520-0493(2000)128%3C1153:BOFOSO%3E2.0.CO;2.
- Li, J., J. Steppeler, P. Gan, F. Fang, and L. Li, (2025) Numerical weather prediction in the time of exaflop computers. *Bull. Amer. Meteorol. Soc.*, 106, E1087-E1097, <https://tpchange.de/meetings/4310-2/>.
- Mesinger, F., (2008) Bias adjusted precipitation threat scores. *Adv. Geosci.*, 16, 137-143, <https://doi.org/10.5194/adgeo-16-137-2008>.
- Mesinger, F., (2023) Cut-cell Eta model: History and challenges overcome. *Anuário do Instituto de Geociências*, 46, 56300, https://doi.org/10.11137/1982-3908_2023_46_56300.
- Mesinger, F. and K. Veljovic, (2013) Limited area NWP and regional climate modeling: a test of the relaxation vs Eta lateral boundary conditions', *Meteorol. Atmos. Phys.*, 119, 1-2, 1-16, doi:10.1007/s00703-012-0217-5.
- Mesinger, F., and K. Veljovic, (2017) Eta vs. sigma: Review of past results, Gallus-Klemp test, and large-scale wind skill in ensemble experiments. *Meteorol. Atmos. Phys.*, 129, 573-593.
- Mesinger, F. and K. Veljovic, (2020) Topography in weather and climate models: Lessons from cut-cell Eta vs. European Centre for Medium-Range Weather Forecasts experiments, *J. Meteorol. Soc. Japan*, 98, 5, 881-900, doi:10.2151/jmsj.2020-050.
- Mesinger F., S.C. Chou, J. Gomes, D. Jovic, P. Bastos, J.F. Bustamante, L. Lazic, A.A. Lyra, S. Morelli, I. Ristic, and K. Veljovic (2012) An upgraded version of the Eta model. *Meteorol. Atmos. Phys.*, 116:63-79. doi:10.1007/s00703-012-0182-z
- Veljovic K, B. Rajkovic, M.J. Fennessy, E.L. Altshuler, and F. Mesinger (2010) Regional climate modeling: should one attempt improving on the large scales? Lateral boundary condition scheme: any impact? *Meteorol. Zeitschrift*, 19:237-246, doi:10.1127/0941-2948/2010/0460

Eta Model at all scales

Sin Chan Chou, Jorge Luís Gomes, André A. Lyra, Isabel D. Pilotto, Daniela Carneiro Rodrigues, Daniel Andrés Rodríguez, Priscila da Silva Tavares, Gustavo Sueiro Medeiros, Diego de Andrade Campos, Diego José Chagas, Fedor Mesinger

National Institute for Space Research (INPE), São Paulo, Brazil (chou.chan@inpe.br)

Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil

Serbian Academy of Science and Arts (SASA), Belgrade, Serbia

1. INTRODUCTION

The model was originally developed by the Belgrade University in collaboration with the Hydrometeorological Institute of the former Yugoslavia (Mesinger et al., 1988). It was operationally used in US NCEP (Black, 1994). It was later installed at the Brazilian National Institute for Space Research (INPE), at its Center for Weather Forecasting and Climate Studies (CPTEC/INPE), from where it has provided early weather forecasts over South America since December 1996 (Chou, 1996). So, 2026 is the 30th anniversary of the model.

The model has the unique feature of the vertical eta coordinate (Mesinger, 1984), whose surfaces are approximately horizontal (Fig.1). This is an advantage for the region, given the steep, complex terrain of the Andes Cordillera. This feature has proven capable of capturing storms driven by foehn-type winds (Seluchi et al., 2003; Antico et al., 2021). The model has been modified to adapt to local South American climate conditions, particularly for Brazilian conditions.

The objective of this article is to describe the developments of the Eta Model, par-

ticularly over South America. Here, we describe the additions to the model after Mesinger et al. (2012; 2016).

In the next sections, the developments of the latest version will be described in Section 2, and Eta model applications of various temporal and spatial scales will be described.

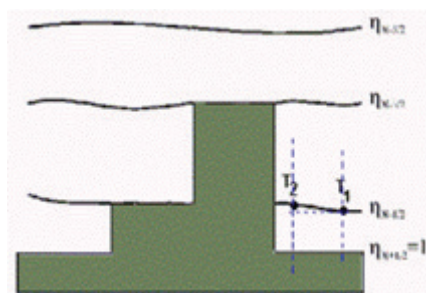


Figure 1 - The eta coordinate and the step topography representation. T1 and T2 are temperatures at two grid points.

2. MODEL DEVELOPMENTS

2.1 Tropical forests with deeper roots

The original Noah land-surface scheme (Ek et al., 2002) used 4 soil layers, reaching a depth of 2 m. However, trees in tropical forests have deeper root systems.

The current version has 8 soil layers and extends to a depth of 12 meters (Pilotto et al., 2023). Not much change is noticed during the rainy season, but this modification improves evapotranspiration and precipitation in the Amazon during the dry season.

2.2 Inclusion of Brazilian biomes

A type of tropical forest typical in Brazil is the Mata Atlantica. Although it can be considered a tropical forest, it has some distinct characteristics from the Amazon forest. Mata Atlantica trees have roots 5 m deep, whereas roots in the Amazon forest now reach 12 m.

Another unique vegetation of the Brazilian biome is the caatinga. It is a whitish vegetation that survives in the semi-arid areas of Northeast Brazil. Caatinga can be dense or sparse. Both types of caatinga are distinguished in the model. Figure 2 shows the updated land cover map at 20-km resolution; it combines the original ESA vegetation map (ESA, 2017) and the inclusion of Brazilian biomes.

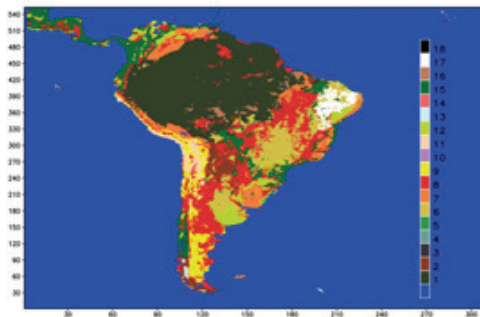


Figure 2 - Vegetation cover with new biomes included.

2.3 Radiation scheme with convective water load

The version of the Eta model using the RRTM radiation scheme has been

shown to reduce Eta model bias in incoming solar radiation at the surface (SWi; Campos et al., 2024) under both clear and cloudy sky conditions. However, we noticed that during deep convective activity, the incoming solar radiation at the surface was hardly reduced. The inclusion of water load produced by the deep convection produced SWi more consistent with observations (Campos et al., 2024)

2.4 Higher model top

The model top has been raised to 1 hPa. The modification has shown greater impact with longer integration, although it has revealed cold biases at upper levels, particularly in the stratosphere.

2.5 Restart functionality

For continuous multidecadal integration, the model requires restart functionality that maintains the precision of the calculations. This is implemented, and precision is guaranteed.

3. APPLICATIONS AT ALL SCALES

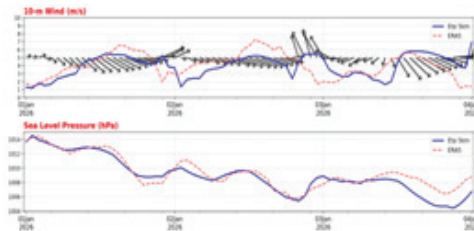
Due to demands from different socioeconomic sectors in Brazil, the model has been applied at different spatial resolutions and for different lead times, starting from 1 km up to about 20 km, from 3-day forecasts up to over a century period for projections of climate change.

3.1 Weather forecasts

For weather forecasts, the model is run at 1-km resolution to capture the detailed topography and the coastline between the two major Brazilian cities: Rio de Janeiro and São Paulo. The model produces daily 3-day forecasts (Fig. 3) whose outputs drive a dispersion model. It is a

short-range forecast. The model is run in convection-permitting non-hydrostatic mode, in a single deterministic run. The forecast is nested into the 8-km Eta medium-range forecast. The topography in the region is complex, with mountains extending near the coastline and elevations reaching about 800 m, a valley towards the interior, and another mountain chain at higher elevations, around 2000 m. A direction-dependent form drag scheme operates near the surface.

a)



b)

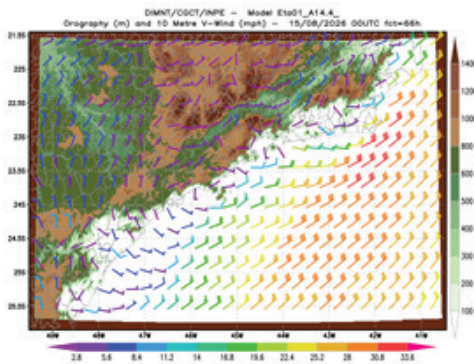


Figure 3 - 10-m winds and mean sea level pressure hourly forecasts at 1-km resolution from 000UTC 01 to 04 January 2026 for São Paulo (a). Wind barbs around the Rio de Janeiro and São Paulo coast, forecast verifying at 1800 UTC, 3 January 2026. Color shading represents topographic elevation (m).

For the medium-range forecasts, the model is set up at 8-km resolution, hydrostatic, over South America and runs out to 11 days, driven by GFS model forecasts. The extension of the run beyond 10 days is done to discard the first forecast day. As the initial conditions are taken from GFS forecasts at 0.25° resolution, the Eta model's pressure and winds are adjusted during the first 12 hours. The medium-range forecast is useful for energy and agriculture operations and planning.

The model has shown skill in forecasting explosive cyclones that reach the coast of southern Brazil in 10-day forecasts. Figure 4 shows the change in mean sea level pressure over the last 24 hours, captured at a 210-h lead time for the event. Evaluation of these forecasts is conducted continuously.

a)

b)

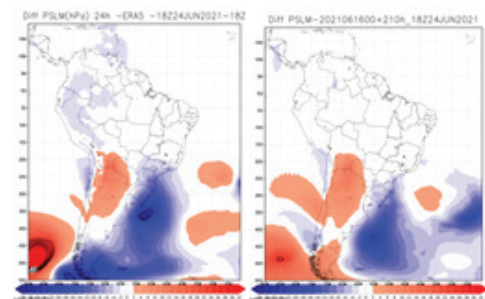


Figure 4 - 24-h Mean Sea Level Pressure change (hPa) on 18Z 24 June 2021 from ERA5 (left) and 210-h Eta 8-km forecast (right).

3.2 Seasonal forecasts

Since a severe drought in 2001 over central Brazil, the Eta model has produced Seasonal forecasts, extending them up to 4,5 months using 5 members (Chou et al.,

2005; 2020), driven by CPTEC's AGCM. Initially at 40-km resolution and later at 20-km resolution, evaluations of the hindcasts have shown the added value of the Eta model seasonal forecasts over the driver AGCM forecasts, especially during the rainy season in Brazil (Chou et al., 2020).

Recently, to extend the forecast lead time, the Eta model was nested within the NCEP CFS forecasts and began producing a 5-member ensemble with a 6-month lead time. Due to computer power limitations, model seasonal climatology was constructed from a 10-year hindcast. The outputs are shown as 1-month, 3-month, and 6-month anomalies of precipitation and 2-m temperature. In addition, these outputs are bias-corrected based on empirical quantile mapping, following Bárdossy and Pegram (2020). These seasonal forecasts were evaluated in Tavares et al. (2025).

Figure 5 shows the average precipitation anomaly for the May-June-July 2026 season, based on a forecast initialized in

February; therefore, it is a season 3 lead time. Each season comprises the mean of 3 months. The forecast is compared against the observations (Fig. 5b)

3.3 Climate change projections

Demands to produce higher-resolution projections of climate change for impact and adaptation studies over South America have motivated the development of the functionalities for multidecadal continuous runs.

The changes from weather forecasts to climate change studies required that the model be able to update sea surface temperature, vegetation greenness, and greenhouse gas concentrations during multidecadal integrations. In addition, some variables were required to operate in double precision. The model shows numerical stability during the continuous multidecadal runs. The model simulations driven by CMIP3 and CMIP5 global models have been evaluated for the reference periods (Pesquero et al., 2010; Chou et al., 2012; 2014). The mod-

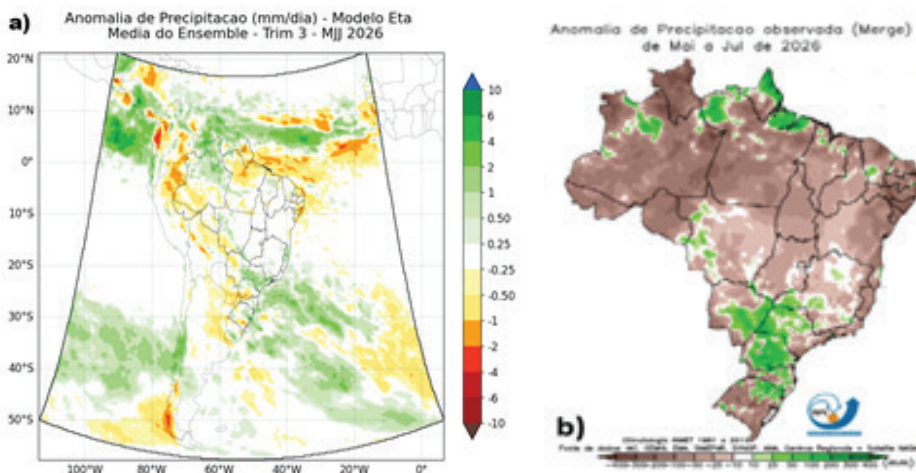


Figure 5 - Precipitation anomaly (mm/day) for May-June-July (MJJ) season 2026: (a) Eta ensemble mean forecast - season 3, and (b) observation.

el has been shown to reproduce the major climate features over South America across different seasons. Projections of future climate change showed reduced precipitation in the Amazon and central Brazil and increased precipitation in southern Brazil. 2-m air temperature increases across the entire continent, but stronger in the central part. Consecutive dry days, which is an extreme climate indicator, are becoming longer despite an increase in the 95th percentile of precipitation (Marengo et al., 2012; Chou et al., 2014b). Later, the non-hydrostatic version was implemented to run high-resolution simulations and study impacts in urban areas at 5-km resolution (Lyra et al., 2018). Figure 6 shows the changes in precipitation in three timeslices projected by the Eta Model nested in 4 global climate models and two levels of RCP4.5 and RCP8.5 scenarios. From the ensemble of 8 projections, the upper bound (top row) and the lower bound

(middle row) of the changes show the broad spread, with values ranging from negative to positive. The bottom row shows the prevailing sign taken from the ensemble of 8 runs. The areas covered by red (blue) dots are where most projections indicate a reduction (increase) in precipitation amount. The reduction prevails in the central part of Brazil, and precipitation increases are projected in southern Brazil.

3.4 Paleoclimate

A numerical model capable of producing future projections of climate change should also be able to run for the past climate. The Berger parameterization (Berger, 1978) was used to calculate orbital parameters: eccentricity, obliquity, and precession. The Regional Eta Model was nested into CPTEC's global climate model, which provided the initial and boundary conditions for the regional model. CPTEC's global climate model

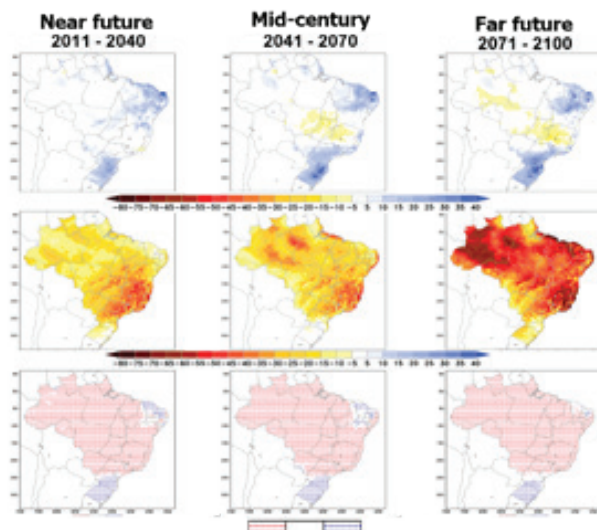


Figure 6 - Projections of precipitation changes (mm) in three timeslices. Upper (top row) and lower (middle row) bounds of the range of changes. Red (blue) dots indicate the most probable negative (positive) signs of the changes (in the bottom row).

also adopted the Berger parameterization. The Eta running in paleoclimate mode (Marchi et al., 2022) was evaluated against proxies of the mid-Holocene (6k).

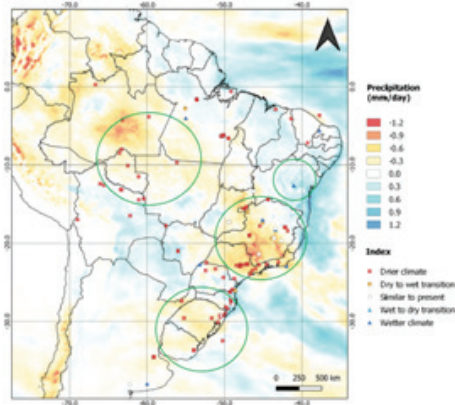


Figure 7 - Differences between 6k and 0k simulations from the Eta Model simulations. Square, triangle, and circle symbols represent dry, wet, and neutral climates, respectively, derived from proxies (Gorestein et al., 2022). Large green circles highlight areas of model agreement with proxies. Adapted from Marchi et al. (2022).

4. FINAL REMARKS

The model is an open-source code available from <https://github.com/EtaModel/>. From the original Fortran, it has been updated to modern Fortran code. The version described here is named Eta2 (Chou et al., 2026). Various developments have been introduced to the model, and the implementation into the unified version is ongoing. Some of these developments include the dynamical vegetation scheme (Lyra, 2015) for studies of Amazon forest tipping points; the lightning parameterization (Moura, 2021); the full Noah-MP (Pilotto et al., 2023); and lake model Flake coupling to the Eta model (Afonso and Chou, 2023). These developments are mostly oriented toward climate change impact and adaptation studies.

ACKNOWLEDGEMENTS

The authors would like to thank the Brazilian research sponsors CNPq, for the support from grants PQ 307252/ 2026-4 and 406591/2022-9, and CAPES, for grant 88887.688971/2022-00.

REFERENCES

- Afonso, E.O.; Chou, S.C. (2023). Modeling the Effects of Local Atmospheric Conditions on the Thermodynamics of Sobradinho Lake, Northeast Brazil. *Climate* 11, 208. <https://doi.org/10.3390/cli1100208>
- Antico, P.L.; Chou, S.C.; Seluchi, M.E.; Medeiros, G.S. (2021). Foehn-like wind in the mountains of southeastern Brazil as seen by the Eta model simulation. *Revista Brasileira de Meteorologia*, v. 36, n. 1, p. 79–86 DOI: 10.1590/0102-77863540074.
- Bárdossy, A., & Pegram, G. (2011). Downscaling precipitation using regional climate models and circulation patterns toward hydrology. *Water Resources Research*, 47(4), 2010WR009689. <https://doi.org/10.1029/2010WR009689>.
- Berger, A.L. (1978). Long-term variations of daily insolation and Quaternary climatic changes. *Journal of the Atmospheric Sciences*, 35(12): 2362-2367. Doi:10.1175/1520-0469(1978)035<2362:LTVODI>2.0.CO;2

- Campos, D.A., Chou, S.C., Bottino, M.J., Gomes, J.L. & Lyra, A. (2024). Inclusion of the radiative effect of deep convective clouds in the Eta model simulations. *Quarterly Journal of the Royal Meteorological Society*, 1–12. Doi:10.1002/qj.4673
- Chou, S.C. (1996) Modelo Regional Eta. *Climanálise* (São José dos Campos), Cachoeira Paulista, SP, v. 1, p. 27- 29, n.ED ES-PECIA.
- Chou, S.C., Bustamante, J., Gomes, J. L., (2005) Evaluation of Eta Model seasonal precipitation forecasts over South America. *Nonlinear Processes in Geophysics*, v.12, 537-555. doi:10.5194/npg-12-537-2005
- Chou S.C., Marengo JA, Lyra A, Sueiro G, Pesquero J, Alves LM, Kay G, Betts R, Chagas D, Gomes JL, Bustamante J, Tavares P (2012) Downscaling of South America present climate driven by 4-member HadCM3 runs. *Climate Dynamics*. Vol. 38 n. 3-4, 635-653. DOI 10.1007/s00382-011-1002-8
- Chou, S.C, Lyra, A.A., Mourão, C., Dereczynski, C., Pilotto, I., Gomes, J., Bustamante, J., Tavares, P., Silva, A., Rodrigues, D., Campos, D., Chagas, D., Sueiro, G., Siqueira, G., Nobre, P. and Marengo, J. (2014a) Evaluation of the Eta Simulations Nested in Three Global Climate Models. *American Journal of Climate Change*, 3, 438-454. doi: 10.4236/ajcc.2014.35039
- Chou, S.C, Lyra, A, Mourão, C., Dereczynski, C., Pilotto, I., Gomes, J., Bustamante, J., Tavares, P., Silva, A., Rodrigues, D., Campo, D., Chagas, D., Sueiro, G., Siqueira, G. and Marengo, J. (2014b) Assessment of Climate Change over South America under RCP 4.5 and 8.5 Downscaling Scenarios. *American Journal of Climate Change*, 3, 512-527. doi: 10.4236/ajcc.2014.35043
- Chou, S.C.; Dereczynski, C.P.; Gomes, J.L.; Pesquero, L.F.; Ávila, A.M.H.; Resende, N.C.; Alves, L.F.; Ruiz-Cárdenas, R.; Souza, C.R.; Bustamante, J.F.F. (2020). Ten-year hindcasts of Eta seasonal forecasts. *An Acad Bras Cienc* (2020) 92(3): e20181242 DOI 10.1590/0001-3765202020181242
- Chou et al., (2026). Climate Extremes in South America: observations and simulation by the Eta Regional Climate Model. Submitted to the *International Journal of Climatology*.
- Ek, M. B.; Mitchell, K. E.; Lin, Y.; Rogers, E.; Grunman, P.; Koren, V.; Gayano, G.; And Tarpley, J. D. (2003). Implementation of Noah Land Model advances in the NCEP operational Eta Model. *Journal of Geophysical Research*, v. 108 (D22), p. 8851-8867. doi:10.1029/2002JD003296
- ESA. Land Cover CCI Product User Guide Version 2. Tech. Rep. (2017). Available at: maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf
- Gomes, J.L.; Chou, S.C.; Mesinger, F.; Lyra, A.A. Rodrigues, D.C, Rodriguez, D.A., Campos, D.A., Chagas, D.J., Pilotto, I., Tavares, P.S., Lucci, L.T., Veljovic, K., Laureanti, N.C., Mantovani Jr, J.A., Lamosa, D., Dourado, M.S.C. (2023) Manual Modelo Eta - versão 1.4.2. INPE. São José dos Campos, SP. 138pp. <http://urlib.net/8JMKD3MGP3W34T/48G6PU5>. ISBN 978-65-89159-08-7
- Gorenstein, I.; Prado, L.F.; Bianchini, P.R.; Wainer, I.; Griffiths, M.L.; Pausata, F.S.R.; Yokoyama, E. (2022) A fully calibrated and updated mid-Holocene climate reconstruction for Eastern South America. *Quaternary Science Reviews*, 292: 107646. doi:10.1016/j.quascirev.2022.107646
- Lyra, A., Tavares, P., Chou, S.C., Sueiro, G., Dereczynski, C.P., Sondermann, M., Silva, A., Marengo, J., Giarolla, A. (2018). Climate change projections over three metropolitan regions in Southeast Brazil using the non-hydrostatic Eta regional climate model at 5-km resolution. The-

- or Appl Climatol 132, 663–682 (2018). doi:10.1007/S00704-017-2110-0
- Lyra, A.A (2015) A Vulnerability Study of Amazon Biome to Climate Projections. Thesis in Earth System Science. National Institute for Space Research.129p. <http://urlib.net/8JMKD3MGP3W34P/3K3A-TEL>.
- Lyra, A. A.; Chou, S. C.; Sampaio, G. O. (2016) Sensitivity of the Amazon biome to high-resolution climate change projections. *Acta Amazônica* (online), v. 46, p. 175-188. doi:10.1590/1809-4392201502225
- Marchi, A.C. de, Melo, M.L.D. de, Lyra, A.A., Kubota, P.Y., Chou, S.C., Rosas, P. (2022). Mid-Holocene Climate Simulations over Brazil Using the Eta Regional Paleoclimate Model. *Derbyana*, 43, e775. doi:10.14295/derb.v43.775
- Marengo, J. A.; Chou, S. C; Kay G.; Alves, L.; Pesquero, J. F Soares, W.R; Santos, D.C.; Lyra, A. A.; Sueiro, G.; Betts, R.; Chagas, D. J.; Gomes, J. L.; Bustamante, J. F.; Tavares, P. (2012). Development of regional future climate change scenarios in South America using the Eta CPTEC/HadCM3 climate change projections: Climatology and regional analyses for the Amazon, São Francisco and the Parana River Basins. *Climate Dynamics*. Vol 38, n 9-10, 1829-1848. DOI 10.1007/s00382-011-1155-5
- Mesinger, F. A blocking technique for representation of mountains in atmospheric models. *Rivista di Meteorologia Aeronomica*, v. 44, p. 195–202, 1984.
- Mesinger, F.; Janjić, Z.I.N.; Gavrilov, D.; Deaven. D.G. (1988). The step-mountain coordinate: Model description and performance for cases of Alpine lee cyclogenesis and for a case of Appalachian redevelopment. *Monthly Weather Review*, v. 116, n. 7, p. 1493-1518.
- Mesinger F, Chou SC, Gomes JL, Jovic D, Bastos P, Bustamante JF, Lazic L, Lyra AA, Morelli S, Ristic I, Veljovic K. (2012) An upgraded version of the Eta model. *Meteorology and Atmospheric Physics*. 116 (3), 63-79. Doi:10.1007/s00703-012-0182-z
- Mesinger F, Veljovic K, Chou S.C., Gomes J.L., and Lyra A.A. (2016). The Eta Model: Design, Use, and Added Value, Topics in Climate Modeling, Theodore Hromadka and Prasada Rao (Ed.), InTech, DOI: 10.5772/64956
- Moura, J.D.O. (2020). Parameterization of Atmospheric Electric Discharges and The Effects on Rain and NOx Production. Thesis in Meteorology. National Institute for Space Research. 114p. <http://urlib.net/8JMKD3MGP3W34R/43J-CLGB>
- Pesquero, J.F., Chou, S.C., Nobre, C., Marengo, J.A. (2010). Climate downscaling over South America for 1961-1970 using the Eta Model. *Theoretical and Applied Climatology*. Vol 99, 1-2, p.75–93. DOI: 10.1007/s00704-009-0123-z
- Pilotto, I.L., Rodriguez, D.A., Chou, S.C., Garofolo, L., Gomes, J.L. (2023). Impacts of the land use and land-cover changes on local hydroclimate in southwestern Amazon. *Clim Dyn* 61, 5597–5612. doi:10.1007/s00382-023-06872-x
- Seluchi, M. E., Norte, F.A., Satyamurty, P., & Chou, S.C. (2003). Analysis of Three Situations of the Foehn Effect over the Andes (Zonda Wind) Using the Eta–CPTEC Regional Model. *Wea. Forecasting*, 18, 481–501 doi:10.1175/1520-0434(2003)18<481:AOTSOT>2.0.CO;2
- Tavares, P.S, Rodrigues, D.C., Chou, S.C., Laureanti, N.C., Medeiros, G.S., Gomes, J.L., Vergasta, L.A., Correia, F.W.S. (2025) Dynamical-statistical downscaling of seasonal hindcasts of temperature and precipitation over South America. *Brazilian Journal of Water Resources*. RBRH, v. 30, e7, 2025.doi:10.1590/2318-0331.302520240073

Graph Machine-Learning Data Assimilation on a Global Icosahedral Grid: Accuracy, Physical Constraints, and Observation-Specific Control

Miodrag Rancic

Lynker at NOAA/NWS/OMD, College Park, Maryland, USA (miodrag.rancic@noaa.gov)

1. INTRODUCTION

Machine-learning (ML) approaches to data assimilation (DA) are increasingly being investigated as alternatives or complements to conventional variational and ensemble methods. A learned analysis operator can map background fields and observations directly to an analysis increment and, after training, can be evaluated without solving a new optimization problem during each analysis cycle. A central question, however, is whether such an operator can deliver more than an accurate statistical reconstruction. Practical DA systems must also account for physical consistency, the spatial structure of analysis increments, and the relative influence assigned to observations.

A controlled shallow-water system provides a useful framework for examining these questions because the true state, background, and synthetic observations are all known. The present work uses a global finite-volume shallow-water model defined on a hexagonal-pentagonal icosahedral grid. The nearly uniform global mesh avoids the polar singularity of latitude-longitude coordinates and, at

the same time, naturally defines the computational graph used by a Graph Neural Network (GNN). The numerical model and the ML analysis operator therefore operate on the same native topology, without interpolation to an auxiliary structured grid. This correspondence is particularly attractive for extending graph-based forecasting and analysis systems toward global atmospheric applications.

The objective of the present study is threefold: first, to establish a graph-based ML-DA operator for simultaneous analysis of shallow-water height and horizontal velocity; second, to investigate how selected physical properties can be controlled explicitly; and third, to examine whether the influence of selected observations can be controlled as an independent conditioning input to the learned analysis.

2. GRAPH ML-DA FRAMEWORK

The shallow-water state is represented as $\mathbf{x} = (h, u, v)^T$, and synthetic backgrounds are generated by adding prescribed perturbations to known truth fields. Obser-

variations are sampled from the truth and perturbed by prescribed observation errors. Rather than predicting the complete analysis state, the GNN learns the increment required to correct the background,

$$\delta \mathbf{x}^{\text{ML}} = F_{\theta}(\mathbf{x}^b, \mathbf{y}), \quad \mathbf{x}^a = \mathbf{x}^b + \delta \mathbf{x}^{\text{ML}},$$

where F_{θ} denotes the learned nonlinear analysis operator.

Each finite-volume cell is represented by a graph node and neighboring cells are connected according to the native mesh

connectivity. Node features contain the background state, observation innovations, observation masks, and geometric information. Successive message-passing layers propagate observational information through the graph, while residual connections stabilize training. The baseline loss is the mean squared difference between predicted and target normalized analysis increments. This formulation is first tested for univariate height analysis and subsequently extended to the multivariate (h, u, v) state. The overall supervised-learning procedure is summarized in Fig. 1.

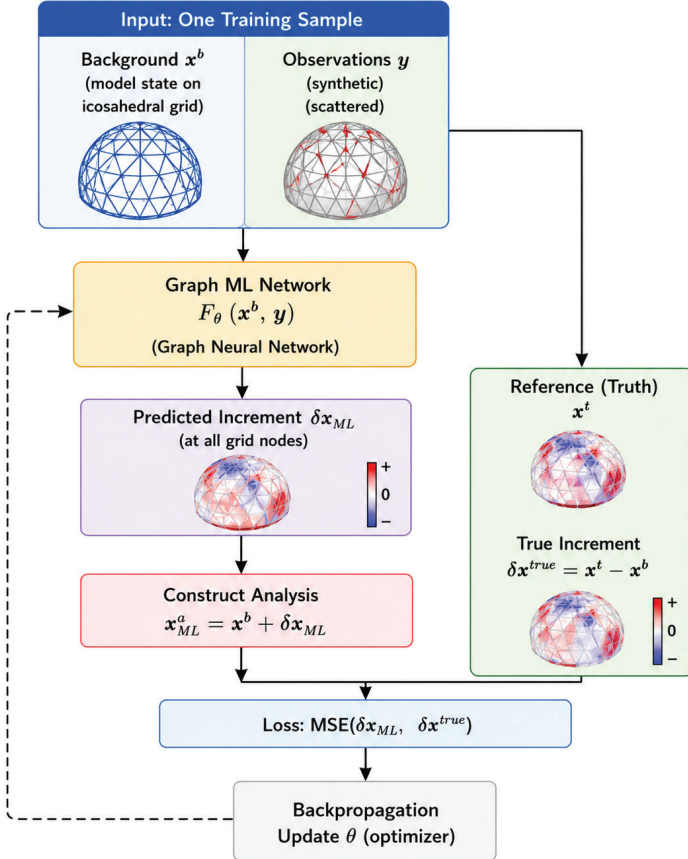


Fig. 1. Graph ML-DA workflow for learning analysis increments from background fields and observations.

3. BASELINE ANALYSIS PERFORMANCE

The univariate experiment provides an initial proof of concept. Over 20 independent test cases, the aggregate height RMSE decreases from 46.54 m in the background to 6.33 m in the ML analysis, corresponding to an 86.4% reduction. The learned correction is spatially coherent, extends beyond individual observation locations through graph message passing, and demonstrates no pronounced imprint of the underlying hexagonal–pentagonal mesh.

The more demanding multivariate experiments test a single learned operator

across localized Gaussian vortices and dipoles, legacy test configurations, and planetary-scale Rossby–Haurwitz flows. Across the complete independent test set, the background RMSE is reduced from 46.77 to 16.76 m for height, from 2.955 to 0.950 m s^{-1} for zonal velocity, and from 3.381 to 1.237 m s^{-1} for meridional velocity. These correspond to error reductions of 64.2%, 67.8%, and 63.4%, respectively. The Rossby–Haurwitz cases remain the most demanding because of their planetary-scale structure and strong rotational flow, yet consistently show substantial improvements in all three prognostic variables.

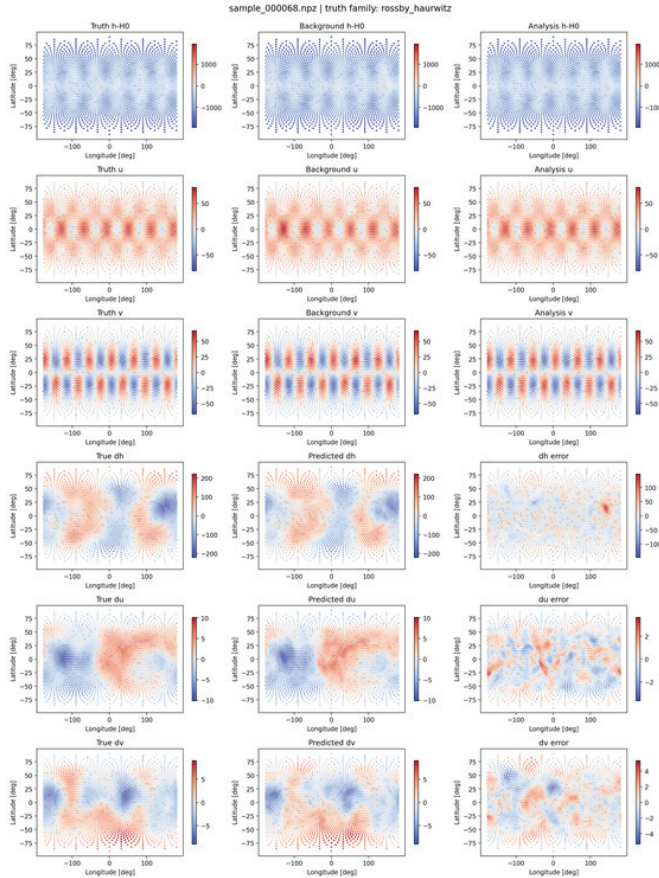


Fig. 2. Representative multivariate Graph ML-DA analysis for a Rossby–Haurwitz flow.

4. PHYSICAL CONSTRAINTS AND OBSERVATION-SPECIFIC CONTROL

Accurate reconstruction of the target increment does not automatically guarantee desired physical properties. This is demonstrated clearly by the mass-conservation experiment. Although the unconstrained multivariate analysis substantially reduces conventional RMSE, it introduces an RMS area-weighted global-mean height change of 3.664 m. Introducing a soft mass penalty into the training objective reduces this inconsistency only modestly, and stronger penalty weights eventually degrade conventional analysis accuracy.

Because global mass conservation is a linear integral condition on the height increment, it can instead be imposed exactly. The predicted height increment is projected onto the mass-conserving subspace by subtracting its area-weighted global mean. This reduces the RMS mass change from 3.664 m to 3.92×10^{-13} m, i.e. numerical roundoff, while the height, zonal-wind, and meridional-wind RMSE values improve to 15.292 m, 0.911 m s^{-1} , and 1.187 m s^{-1} , respectively. Thus, exact conservation is achieved without sacrificing analysis accuracy and, in this experiment, even slightly improves it.

A complementary soft dynamical constraint is introduced using an edge-based balance residual consistent with

the finite-volume geometry. The residual compares the local pressure-gradient and Coriolis terms in the analysis with those of the reference truth, rather than forcing the state toward exact geostrophic balance. With exact mass conservation already imposed, increasing the balance-loss weight reduces the dynamical imbalance while leaving conventional analysis skill nearly unchanged. For the selected weight $\lambda_{\text{bal}}=1$, the balance-error reduction improves from 52.12% in the exact-mass baseline to 58.72%, while the analysis improvements remain 66.82%, 69.61%, and 65.32% for h , u , and v . Exact global mass conservation is retained throughout.

The framework also permits explicit control of observational influence. The spatial response to increasing observation priority is illustrated in Fig. 3. Observation-specific priority is supplied as an additional conditioning input rather than being interpreted as a change in observation-error variance. Once trained, the observation priority can be varied during inference without retraining the network. In a representative test, increasing the priority of a selected height observation from $p=1$ to $p=10$ reduces the local analysis-minus-observation residual from 3.88 to 0.17 m. The induced response remains strongly localized: the analyzed height changes by about -4.05

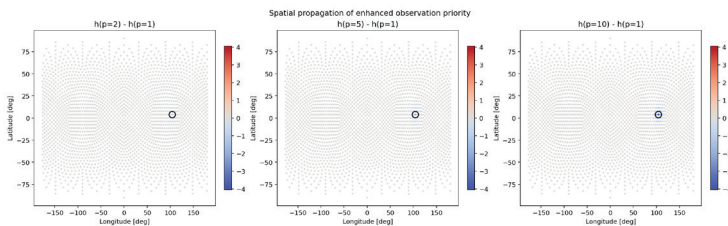


Fig. 3. Localized analysis response to increasing priority of a selected observation.

m at the selected observation, whereas the global area-weighted RMS difference from the standard analysis is only about 0.13 m. Global RMSE remains essentially unchanged, exact mass conservation is preserved, and the balance diagnostic is practically unaffected.

Table 1. Summary of the principal Graph ML-DA results.

Development	Principal result
Univariate height	RMSE reduced from 46.54 to 6.33 m (86.4%).
Multivariate (h, u, v)	Error reductions of 64.2%, 67.8%, and 63.4%, respectively.
Exact mass conservation	RMS mass change reduced from 3.664 m to 3.92×10^{-13} m.
Dynamical balance	Balance-error reduction increased from 52.1% to 58.7%.
Observation priority	Local residual reduced from 3.88 to 0.17 m for $p = 1 \rightarrow 10$.

5 CONCLUSIONS

The experiments demonstrate that a graph-based ML analysis operator can simultaneously provide substantial mul-

tivariate error reduction on a global unstructured grid and support explicit control of selected analysis properties.

The main methodological result is that different forms of information can be incorporated according to their mathematical character: exact projection for algebraic constraints such as global mass conservation, additional loss terms for properties more naturally enforced softly, and conditioning inputs for application-dependent analysis preferences such as observation priority.

The global shallow-water system provides a controlled environment in which these components can be isolated and understood before application to more comprehensive atmospheric systems. Natural next steps include graph-spectral or scale-selective control of analysis increments, explicit multiscale covariance guidance using a Multigrid Graph Filter, more general observation embedding, and extension from the present single-time analysis operator toward four-dimensional ML data assimilation in which learned analysis and forecast operators are coupled through time.

REFERENCES

- Battaglia, P.W. et al. (2018) Relational inductive biases, deep learning, and graph networks, arXiv:1806.01261.
- Cheng, S. et al. (2023) Machine learning with data assimilation and uncertainty quantification for dynamical systems: A review, IEEE/CAA J. Autom. Sinica, 10, 1361–1387.
- Kipf, T.N., Welling, M. (2017) Semi-supervised classification with graph convolutional networks, Proc. ICLR 2017.
- Ringler, T., Thuburn, J., Klemp, J.B., Skamarock, W.C. (2010) A unified approach to energy conservation and potential vorticity dynamics for arbitrarily structured C-grids, J. Comput. Phys., 229, 3065–3090.
- Thuburn, J. (2012) Horizontal discretisations for global weather and climate prediction models, J. Comput. Phys., 231, 4327–4345.
- Williamson, D.L., Drake, J.B., Hack, J.J., Jakob, R., Swarztrauber, P.N. (1992) A standard test set for numerical approximations to the shallow water equations in spherical geometry, J. Comput. Phys., 102, 211–224.

Multigrid Graph Filter for Unstructured Finite-Volume Meshes

Miodrag Rancic, R. James Purser, and Manuel S. F. V. De Pondecá

Lynker at NOAA/NWS/OMD, College Park, Maryland, USA (miodrag.rancic@noaa.gov)

1 INTRODUCTION

Background-error covariance modeling is a fundamental component of atmospheric and oceanic data assimilation. Recursive filters, diffusion operators, spectral methods, and related approaches provide efficient representations of spatial correlations, but modern atmospheric models increasingly employ unstructured finite-volume meshes for which traditional coordinate-dependent filtering formulations are not naturally defined. The Multigrid Beta Filter (MGBF) addresses a broad range of covariance scales on structured grids through a hierarchy of local compact-support filtering operations. Extending this multilevel covariance philosophy to arbitrary unstructured meshes requires a formulation based on mesh connectivity rather than coordinate directions.

The present work develops a Graph Filter (GF) and its multilevel extension, the Multigrid Graph Filter (MGGF), directly on the graph associated with a finite-volume mesh. Each finite-volume cell is represented by a graph vertex and neighboring cells sharing a common boundary are connected by graph edges. The graph therefore represents the native mesh topology itself rather than an auxiliary

computational structure. This formulation is coordinate independent, treats hexagonal and pentagonal cells through the same neighborhood operations, and provides a natural route toward covariance modeling on general unstructured finite-volume meshes.

The objective is to preserve the essential covariance properties of the original MGBF while removing its dependence on structured-grid geometry. Particular attention is given to conservation, self-adjointness, localization, approximate isotropy, multiscale covariance representation, and robustness near pentagonal topological defects. The formulation is first developed and tested on recursively refined dual-icosahedral meshes and is then applied without modification to an independently constructed irregular spherical Voronoi hierarchy.

2 GRAPH FILTER ON AN UNSTRUCTURED MESH

Let the computational mesh be represented by an undirected graph $G=(V,E)$, where graph vertices correspond to finite-volume cells and graph edges

connect neighboring cells. Local mesh connectivity defines the adjacency and degree operators and, through them, the graph Laplacian L . The single-level Graph Filter is written as

$$B_{\text{GF}} = (I + \beta L)^{-1},$$

where the parameter β controls the effective smoothing radius. Filtering is therefore performed through exchanges between neighboring graph vertices rather than along coordinate directions (Fig. 1).

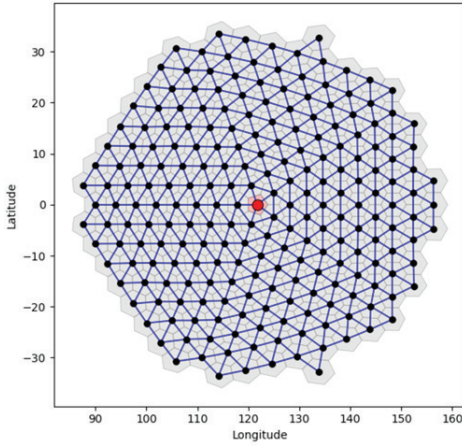


Fig. 1. Finite-volume mesh and its graph representation. Cell centers define graph vertices and neighboring cells are connected by graph edges.

The resulting covariance operator is positive definite and self-adjoint with respect to the area-weighted inner product. Conservative normalization preserves constant fields and the area-integrated quantity, while the graph construction naturally accommodates variations in local vertex degree. On the dual-icosahedral mesh, this permits the same filter operator to be applied to the 2550 hexagonal cells and the twelve pentagonal cells required by spherical topology.

3 MULTIGRID GRAPH FILTER

A single graph generation represents covariance efficiently over a limited range of spatial scales. To obtain a multiscale operator, the MGGF introduces a hierarchy of progressively coarser graph generations,

$$G_0 \rightarrow G_1 \rightarrow G_2,$$

with 2562, 642, and 162 vertices in the demonstration considered here. Each coarse graph vertex represents a connected aggregate of neighboring fine-grid vertices. Parent values are formed as area-weighted averages of their children, so the integrated quantity is preserved exactly during aggregation.

Communication between successive generations is performed by conservative restriction (upsend) and prolongation (downsend) operators. The prolongation is constructed as the area-weighted adjoint of restriction, providing the symmetry required for a self-adjoint multilevel covariance operator. Unlike classical multigrid solvers, the coarse graph generations are not introduced to accelerate convergence of an iterative solution. Instead, each generation acts as an independent covariance model associated with a characteristic range of spatial scales.

The complete MGGF combines filtered contributions from all graph generations after conservative transfer to the finest graph. In schematic form,

$$B_{\text{MGGF}} = \frac{1}{W_0 + W_1 + W_2} \left[W_0 B_0 + W_1 P_{01} B_1 R_{10} + W_2 P_{01} P_{12} B_2 R_{21} R_{10} \right].$$

Fine generations provide localized covariance structure, while progressively coarser generations introduce regional and planetary-scale correlations. Their weighted superposition therefore provides a scale-dependent representation of background-error covariance within a single graph-based multilevel operator.

4 NUMERICAL DEMONSTRATION

Impulse-response experiments on the recursively refined dual-icosahedral hierarchy verify the principal properties of the formulation. At the finest level, the GF produces positive, localized, compactly supported, and rapidly decaying covariance structures. Responses centered on representative hexagonal and pentagonal cells are very similar (Fig. 2), showing that the unavoidable fivefold topological defects have only a weak influence on the radial covariance structure.

Global diagnostics obtained by centering impulses at every graph vertex confirm this behavior over the complete mesh. The mean anisotropy is 1.0684, with a maximum of 1.1161, while the mean relative mass-conservation error is 3.0×10^{-15} . The effective support contains approximately 125 cells on average. These statistics demonstrate that the single-level GF remains nearly isotropic and conservative throughout the mesh despite the presence of the twelve pentagonal cells.

The multilevel impulse response (Fig. 3) demonstrates the intended separation of covariance scales. The G_0 contribution remains localized, the lifted G_1 contribution introduces broader regional structure, and the lifted G_2 contribution provides planetary-scale correlations. Their weighted superposition forms the final MGGF response while retaining the positive and conservative character of the single-level filter.

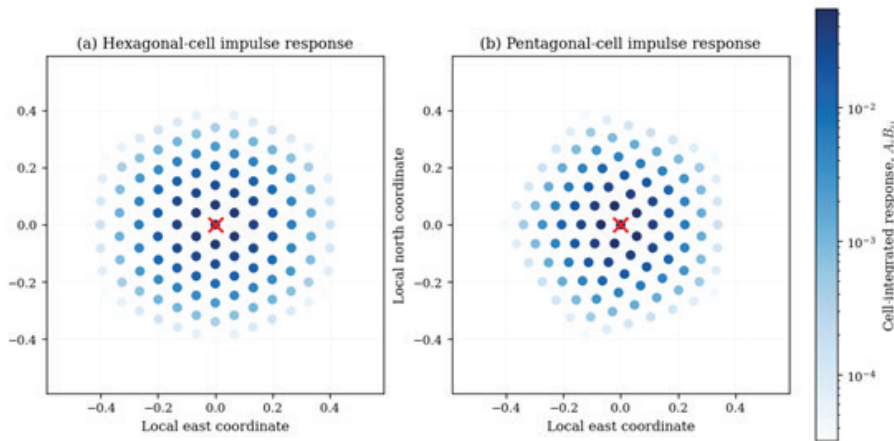


Fig. 2. Local Graph Filter impulse responses centered on (a) a hexagonal and (b) a pentagonal cell.

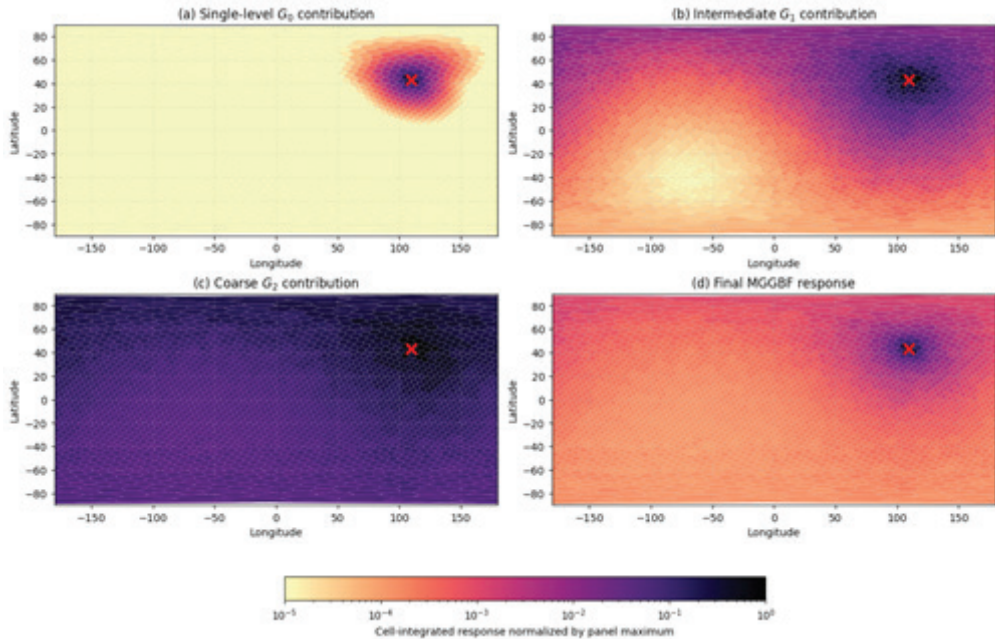


Fig. 3. Multiscale contributions of the three graph generations and the resulting MGGF response.

5 GENERALIZATION TO AN IRREGULAR VORONOI HIERARCHY

To determine whether the formulation depends on the recursively refined icosahedral construction, the same MGGF is applied to an independently generated irregular spherical Voronoi hierarchy (Fig. 4). The finest graph again contains 2562 cells, while conservative graph aggregation produces intermediate and coarse generations containing 642 and 162 parent vertices. Transfer-operator tests preserve constants and area-integrated quantities to machine precision and satisfy area-weighted adjointness to relative errors of order 10^{-15} – 10^{-14} .

The covariance diagnostics (Table 1) closely reproduce those obtained on the recursively refined hierarchy. The effective covariance radii of the lifted G_0 , G_1 ,

and G_2 contributions are approximately 786, 3352, and 8417 km, respectively, while the final MGGF radius is approximately 3699 km. The final response has anisotropy 1.036, a centroid displacement of 63 km, and an integrated response equal to unity to numerical precision. No negative response values are detected, and the graph filters remain self-adjoint to relative errors of order 10^{-15} .

Table 1. Selected MGGF diagnostics on the Irregular Voronoi hierarchy.

Response	r_{eff} (km)	Anisotropy	Integral
G_0	786	1.128	1.000000
G_1	3352	1.064	1.000000
G_2	8417	1.026	1.000000
MGGF	3699	1.036	1.000000

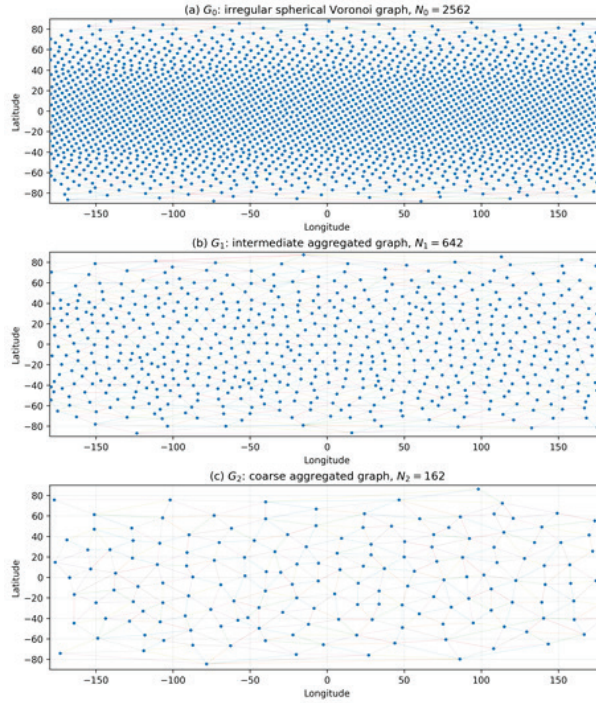


Fig. 4. Irregular spherical Voronoi graph hierarchy used to test generality of the MGGF formulation.

These results provide direct numerical evidence that the essential covariance properties arise from graph connectivity, conservative aggregation, and adjoint inter-generation transfer operators rather than from a particular mesh construction strategy. The MGGF is therefore applicable to a broader class of graph-based finite-volume discretizations, including irregular and variable-resolution Voronoi-type meshes.

6 CONCLUSIONS

The Multigrid Graph Filter reformulates the multilevel covariance principles of the MGBF entirely in terms of graph connectivity and conservative inter-generation communication. The resulting op-

erator preserves the essential properties required for covariance modeling: localization, conservation, self-adjointness, approximate isotropy, and representation of a broad range of spatial scales. Numerical experiments demonstrate stable behavior both on a recursively refined hexagonal-pentagonal icosahedral hierarchy and on an independently constructed irregular spherical Voronoi hierarchy.

The principal methodological result is that multilevel covariance modeling does not require structured indexing or globally defined coordinate directions. Provided that the finite-volume mesh admits a graph representation, successively coarser graphs can be constructed through conservative aggregation, and

adjacent generations can communicate through conservative restriction and area-weighted adjoint prolongation, the same multilevel formulation can be applied without dependence on a particular mesh geometry. This establishes a

general graph-based foundation for covariance modeling on modern unstructured atmospheric models and provides a natural connection to graph-based machine-learning methodologies.

REFERENCES

- Bannister, R. N., 2008: A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances. *Quart. J. Roy. Meteor. Soc.*, 134, 1951–1970.
- Shuman, D. I., S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, 2013: The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Process. Mag.*, 30, 83–98.
- Skamarock, W. C., and coauthors, 2012: A multiscale nonhydrostatic atmospheric model using centroidal Voronoi tessellations and C-grid staggering. *Mon. Wea. Rev.*, 140, 3090–3105.

AI-Driven Optimisation of an Adaptive Mesh Framework in the Fluidity–Atmosphere Model

J. Li¹, P. Gan^{2,3}, F. Fang⁴, X. Wu², J. Zhu¹, Z. Wang¹, M. Zhu¹, X. Zou⁵

¹ State Key Laboratory of Atmospheric Environment and Extreme Meteorology, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, 100029, China

² School of Atmospheric Sciences, Chengdu University of Information Technology, Chengdu, 610225, China

³ Institute of Earth System Science, Faculty of Geographical Science, Beijing Normal University, Beijing, 100875, China

⁴ Applied Modelling and Computation Group, Department of Earth Science and Engineering, Imperial College London, London, SW7 2AZ, United Kingdom

⁵ CW3E, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California, 92093, United States

ABSTRACT

High-resolution numerical weather prediction faces the dual challenge of high computational cost and the need to represent multi-scale processes. Dynamically adaptive meshes provide a promising approach—increasing resolution in regions of interest while improving computational efficiency. However, their application in atmospheric models is limited by expensive three-dimensional (3D) mesh generation and frequent remeshing. To address these limitations, this study develops an AI-driven adaptive mesh framework within the Fluidity–Atmosphere (FA) model. An LSTM-based model for predicting the Riemann metric tensor effectively replaces original mesh regulation in FA model, achieving accurate metric prediction and reduces mesh elements and computational costs by approximately 40%–50% in three-dimensional simulations, while preserving fine-scale features such as fluid interfaces and vortical structures. Additionally, a

hybrid CNN–ANN model (HyMeshAI) enables end-to-end prediction of both mesh density and nodal positions, showing strong agreement with FA benchmarks and improving mesh quality by 30%~40%, and accurately capturing fine-scale structures such as Kelvin–Helmholtz instabilities. The AI-enhanced adaptive mesh framework is further applied to mountain wave simulation over the northeastern Tibetan Plateau, the AI-driven mesh refinement effectively eliminates numerical noise near steep topography, reduces the total number of nodes, improves computational efficiency by approximately 20%, and accurately captures wave evolution consistent with theoretical predictions. This work establishes an efficient and high-fidelity framework for adaptive mesh atmospheric modelling, providing a approach for next-generation numerical weather prediction systems.

Ensemble forecasting of radiational fog over northern France using a non-hydrostatic Eta model

Dragana Mujković¹, Katarina Veljović Koračin¹, Darko Koračin²

¹ Institute of Meteorology, Faculty of Physics, University of Belgrade, Belgrade, Serbia
katarina@ff.bg.ac.rs

² University of Split, Faculty of Science, Split, Croatia

INTRODUCTION

As a result of a broad anticyclone over northern France in December 2016, widespread fog occurred and persisted for about two weeks. During that period, Météo France conducted a comprehensive field program, “SIRTA” (Haefelin et al. 2009; Dupont et al. 2016), including tower and remote sensing measurements focused on fog and low-visibility conditions. The observed data from the SIRTA, Villacoublay, Orly, and Trappes stations show high correlation coefficients between station pairs for surface pressure, air temperature, and relative humidity (above 0.91), as well as for wind speed (above 0.61). This lower correlation in wind direction was expected under anticyclonic conditions with weak winds, where local terrain effects dominate directional variability.

According to the SIRTA measurements, the morning of December 17 was calm, with drizzle, followed by clear skies and high humidity. Based on SIRTA visibility measurements, the first appearance of fog was recorded in the afternoon of December 17 (between 1714 and 1833 UTC, with a couple of short breaks of a few minutes). Later, at 2322 UTC, a fog episode started that lasted almost continuously through December 18 and December 19 up to 0716 UTC (with one

longer break from 0221 to 0523 UTC on December 19). The vertical structure of the fog reached a height of 500 m, with a maximum liquid water content (LWC) of 0.35 g/m³ at the top of the layer, according to SYRSOC data. Ground-level meteorological parameters indicated 2-m air temperatures between 2°C and 5°C, while 10-m wind speeds were extremely low, remaining below 3 m s⁻¹. The fog lifted into stratus on December 19 at 0716 UTC, when winds shifted from northerly to easterly and increased to 4 m s⁻¹.

The main objective of the study was to investigate whether, and to what extent, the ensemble modeling approach can improve forecasting of radiational fog with respect to its onset, evolution, and dissipation, based on statistical scores, and to evaluate predicted visibility against measurements.

METHOD

The data from the SIRTA field campaign offer great opportunities for numerical experiments focusing on fog onset, evolution, and dissipation. In particular, uncertainty and errors in initial and boundary conditions motivate the use of ensemble forecasting, especially for simulating fog, which is a complex atmos-

pheric phenomenon. We used a non-hydrostatic Eta model, structured with the original Eta grid, a cut-cell grid, and unique domain interaction (Mesinger et al. 2012). The model has demonstrated the ability to successfully simulate regional and mesoscale phenomena such as jet-stream evolution (Veljovic et al., 2010; Mesinger and Veljovic, 2017, 2020), extreme precipitation (Marjanovic et al. 2017), climate evolution (Djurdjevic and Rajkovic, 2010), tropical processes (Chou et al. 2002) among others.

The horizontal and vertical resolution of the global driver model was 18 km and 64 vertical levels. The Eta model was run at 5 km resolution in both horizontal directions and with 45 layers in the vertical up to 25 hPa level, on a domain centered over Eastern France and covering approximately 1700 km $\times$ 1245 km in size. The lowest model level above ground at the SIRTa point during the event was located at approximately 23 m. The LBC time frequency was 6 hours. Time step for dynamics was 12 s, while the physics package was called every 25 time steps.

The main prognostic variable for detecting fog in the model results is the liquid water mixing ratio at the lowest model level, which was available from hourly outputs and converted into LWC. Comparison of the observed data with a single model grid point can introduce additional error in the verification, particularly for phenomena such as fog, which are highly variable in space. Therefore, the verification was extended to include the nearest grid point and four surrounding grid points, resulting in an effective ensemble size of 5 $\times$ 31 members. This approach is justified by the high correlation

between neighboring stations and reduces the sensitivity of the results to the exact gridpoint positioning relative to the observation site.

To assess the deterministic performance of the 72-h Eta ensemble-mean forecast for LWC, four standard statistical metrics are computed using the hourly model outputs against the SIRTa site observations: Mean Error, Mean Absolute Error, Root Mean Square Error, and the Pearson correlation coefficient. To assess the uncertainty and probabilistic spread within the forecast at each time step, the ensemble variance is used. The contingency-table counts of hits, false alarms, misses, and correct negatives, as well as the probabilistic scores, are calculated for a range of LWC thresholds. The statistical tests also include calculation of the Brier score and Talagrand diagrams. With this approach, a transition is made from deterministic to probabilistic verification. Instead of evaluating the ensemble purely as a single deterministic outcome (such as the ensemble mean), deterministic scores are calculated for each individual member. With these 155 separate scores plotted as a boxplot, the distribution, spread, and uncertainty of the forecast performance are captured. This effectively shifts the validation framework from traditional single-value verification to probabilistic verification that covers the entire ensemble space.

RESULTS

In the first 19 hours of the forecast, the ensemble members show negligible LWC values, indicating a consistent absence of fog (Figure 1). After 20 hours of lead time, however, a substantial increase in

ensemble spread becomes apparent. The ensemble mean (red line) smoothly captures the event, peaking between 35 and 54 hours before dropping sharply back to near-zero values by hour 62. The ensemble mean shows LWC values similar to the observations during the first ten hours of the fog event. For the remainder of the fog period, the observations are systematically lower than the ensemble mean and the control members. While

some ensemble members successfully predict the timing of both the onset and dissipation of the fog, reflected in the reduction of LWC, the ensemble system clearly overestimates the magnitude of LWC. After a lead time of 47 hours, the five control members occupy the upper bound of the ensemble spread, further highlighting that the unperturbed simulation overpredicts LWC compared with the observations.

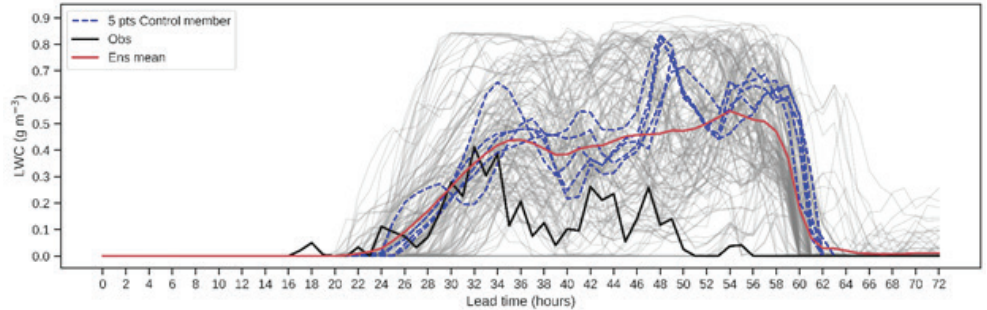


Fig.1 Time series of perturbed ensemble forecasts (gray lines) and control members (blue dashed lines) and observations (solid black line) for liquid water content (LWC, g m^{-3}) at five selected points, over the 72-hour forecast period. The solid red line indicates the ensemble mean. The ensemble consists of 155 members. The Eta Model initial time was 00 UTC 17 December 2016.

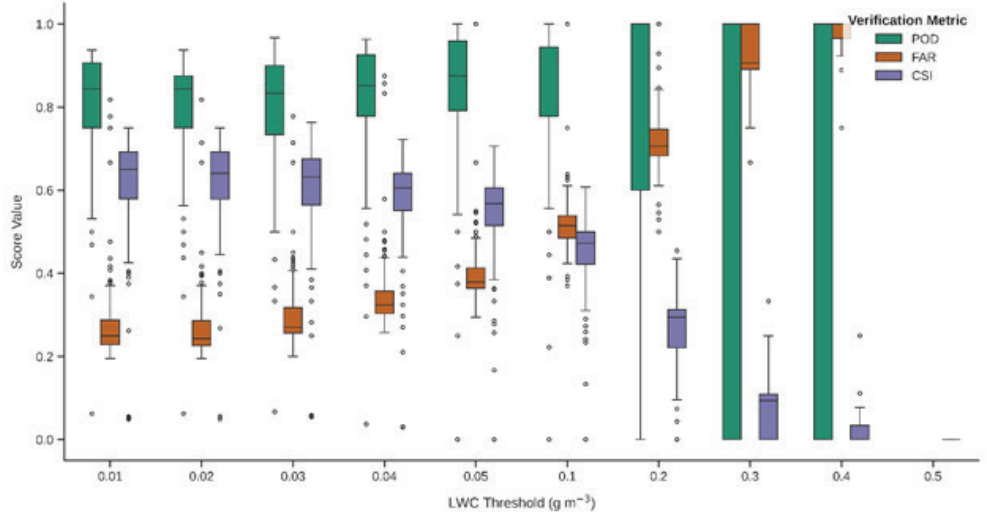


Fig. 2 Verification metrics of the Eta ensemble members (155) as a function of the LWC threshold (g m^{-3}): POD - Probability of Detection (green), FAR - False Alarm Ratio (orange), CSI - Critical Success Index (purple).

Verification using the deterministic scores POD, FAR, and CSI across the 155 members for the Eta model's ability to correctly forecast LWC (Figure 2) indicates that the model has skill at low thresholds, within the typical fog criteria range (0.02 to 0.05 g m⁻³). The high probability of detection (POD medians between 0.80 and 0.88), combined with relatively low false alarm ratios (FAR), yields a highly satisfactory Critical Success Index (CSI around 0.55–0.65). This confirms that individual ensemble members are highly reliable for operational fog detection. At extreme thresholds (0.3 to 0.5 g m⁻³), the deterministic metrics become unreliable because such high LWC events are extremely rare in the verification sample.

For the range of typical fog thresholds for LWC, the ensemble shows distinct pre-

dictive skill, as its Brier score, shown in Figure 3, remains consistent (about 0.16) and significantly below the sample reference (about 0.22–0.25). This wide gap between the curves indicates the ensemble's capacity to capture predictable patterns, sharply outperforming the sample reference, which is prone to high variance and uncertainty at lower thresholds. However, performance shifts noticeably above the threshold of 0.05 g m⁻³. In the LWC threshold range from 0.05 to 0.20 g m⁻³, the score lines diverge: the reference score rapidly decreases as event frequency declines, while the ensemble Brier score rises to a peak of roughly 0.22. The Talagrand diagram (not shown) exhibits an asymmetric U-shape, heavily dominated by a pronounced peak in the highest rank bin.

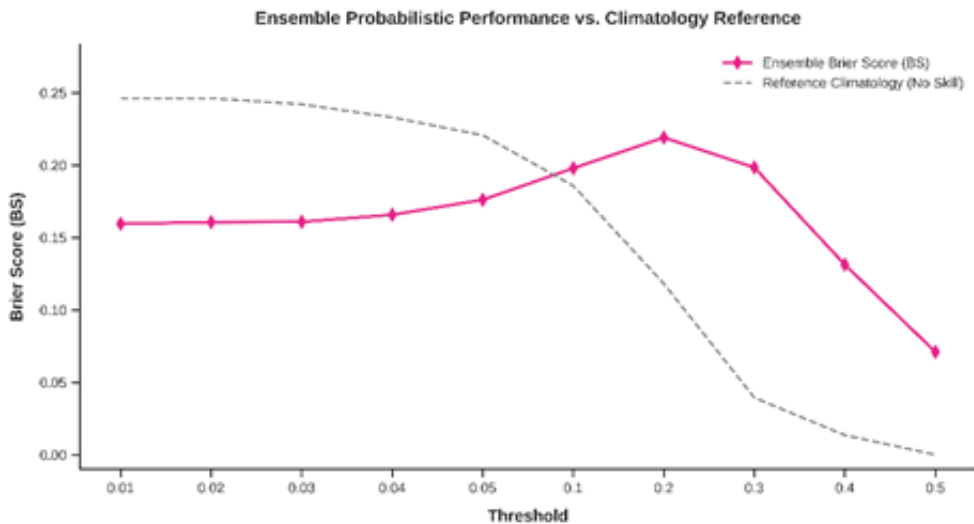


Fig. 3 The Brier Score (BS) for the Eta ensemble (pink solid line with diamond markers) alongside the sample reference baseline (grey dashed line) across the selected LWC (g m⁻³) thresholds. The ensemble consists of 155 members.

As an additional experiment, the Ensemble Kalman Filter (EnKF) was applied to assimilate the available LWC observations hourly into the ensemble. The EnKF analysis followed the observations more closely than the original ensemble, showing the adjustment of the modeled LWC state toward the observed evolution of the fog.

Acknowledgement

This study is supported by the Ministry of Science, Innovation, and Technological Development of the Republic of Serbia (Grants 451-03-34/2026-03/200162) and by the Program PRIZMA, Extreme Weather Events in Serbia - Analysis, Modeling and Impacts (EXTREMES), Grant/Award 7389.

REFERENCES

- Dupont, J.C., M. Haeffelin, S. Stolaki, and T. Elias, 2016: Analysis of Dynamical and Thermal Processes Driving Fog and Quasi-Fog Life Cycles Using the 2010–2013 ParisFog Dataset. *Pure Appl. Geophys.* 173, 1337–1358.
- Djurdjevic, V., and B. Rajkovic, 2010: Development of the EBU-POM coupled regional climate model and results from climate change experiments, In: *Advances in Environmental Modeling and Measurements*, Editors: T. D. Mihajlovic and Lalic B., Nova Publishers.
- Haeffelin, M., T. Bergot, T. Elias, R. Tardif, D. Carrer, P. Chazette, M. Colomb, P. Drobinski, E. Dupont, J.C. Dupont, L. Gomes, L. Musson-Genon, C. Pietras, A. Plana-Fattori, A. Protat, J. Rangognio, J.C. Raut, S. Rémy, D. Richard, J. Sciare, and X. Zhang, 2009: PARISFOG: Shedding new light on fog physical processes. *Bull. Am. Meteorol. Soc.* 91: 767–783.
- Marjanović, D., K. Veljović, and M. Zarić, 2017: Forecasts of extreme precipitation in the western Balkans in May 2014: model skill and sensitivity to the vertical co-ordinate. *Meteorol. Appl.* 24:387–396. <https://doi.org/10.1002/met>.
- Mesinger, F., and K. Veljovic, 2017: Eta vs. sigma: Review of past results, Gallus-Klemp test, and large-scale wind skill in ensemble experiments. *Meteorol. Atmos. Phys.*, 129, 573–593.
- Mesinger, F., and K. Veljovic, 2020: Topography in weather and climate models: Lessons from cut-cell Eta vs. European Centre for Medium-Range Weather Forecasts experiments, *J. Meteorol. Soc. Japan*, 98, 5, 881-900, doi:10.2151/jmsj.2020-050.
- Mesinger, F., S. C. Chou, J. Gomes, D. Jovic, P. Bastos, J. F. Bustamante, L. Lazic, A. A. Lyra, S. Morelli, I. Ristic, and K. Veljovic, 2012: An upgraded version of the Eta model. *Meteorol. Atmos. Phys.*, 116, 63–79.
- Chou, S. C., C. A. S. Tanajura, Y. Xue, and C. A. Nobre, 2002: Validation of the coupled Eta/SSiB model over South America, *J. Geophys. Res.*, 107(D20), 8088, doi:10.1029/2000JD000270.
- Veljovic, K., B. Rajkovic, M. J. Fennessy, E. L. Altshuler, and F. Mesinger, 2010: Regional climate modeling: Should one attempt improving on the large scales? Lateral boundary condition scheme: Any impact? *Meteor. Z.*, 19, 237–246.

Sparse L_Galerkin irregular meshes of high approximation order and their simple implementation in arbitrary existing models using the method of geometric files

J. Steppeler

MOW&more and University Bonn

ABSTRACT

The Local Galerkin method o2o3 has been defined before and simple numerical tests have been done, such as von Neumann analysis (see ‘Mathematics of the Weather’, Springer Nature 2020, referenced as STLI in future) . The subject of this paper is to define this method for arbitrarily irregular grids and apply it to realistic numerical models (models with physics, realistic orography and initial data). It is suggested to do this by introducing this method into existing meteorological forecast models, such as MPAS, ICON or ETA. This model is called the target model. Such introductions of new methods have been done before, for example with the Semi-Lagrange method. These were always mayor projects lasting several years. The difficulty with introducing new numerical methods has been, that in the past it was considered necessary that the numerical specialist learns the target model and this regularly turned out to be a time consuming process. For example the introduction of the icosahedral grid into the DWD model GME took 3 years and this at the time was considered a very fast implementation. Here a simpler method is proposed, called the method of geometric files. The discretization of the L-Galerkin method can be written as modified centered difference equations. The modification consists in the introduction of weights

for each grid points. For regular grids the weights become 1 and the original centered difference scheme is recovered. For irregular grids a set of weights for each grid point is needed and this is called the geometric file. This means that each grid and each numerical resolution needs a corresponding geometric files. For the tests given in STLI these geometric files were provided by the company MOW&more and for the intended low resolution experiments the geometric files are planned to be published. While the geometric files depend on the grid, they do not depend on the form of the equations, such as advection or conservation form of the equations. This means that advection of pollutants with pre-computed velocity and a full Noer Stokes model of the atmosphere will need the same set of geometric files. As toy models shallow water tests with hexagonal grids on the sphere and 2-d cut cell discretisations are planned. Hexagonal grids with Cut cells for resolutions t21 and 250 km grids are planned and a target model for this discretization is needed to obtain realistic weather forecasts. The geometric files will allow the implementation within many target models, but the CAS (Chinese Academy of Sciences) fluidity team is intent on implementing this for at least one target model.

A Unified Solver for Global and Regional Atmospheric Simulations

Xindong Peng¹, Siyuan Chen^{1,2}, Xiaohan Li³, Yifan Zhao⁴, Jun Chang⁵

¹ CMA Earth System Modeling and Prediction Centre, Beijing, China (pengxd@cma.gov.cn)

² Department of Atmospheric and Oceanic Sciences, Fudan University, Shanghai, China

³ Nanjing University of Information Science and Technology, Nanjing, China

⁴ Institute of Urban Meteorology, China Meteorological Administration, Beijing, China

⁵ Disaster Prevention Center, Sichuan Provincial Meteorological Service, Chengdu, China

1. GENERAL DESCRIPTION

A unified numerical solver for both global and regional atmospheric simulations is developed on the Yin–Yang grid (Fig.1, Chen et al. 2023), which is named Yin–Yang-grid unified model for the atmosphere (YUNMA), at China Meteorological Administration. The composite feature of the Yin–Yang grid makes it ready to use the solver as a limited-area model with only one component grid and a global model with both components activated. Local refinement by multi-level grid nesting is straightforward in either Yin or Yang grid. Full consideration of the three-dimensional Coriolis forc-

ing in the Navier–Stokes equation helps the free rotation of local coordinate on sphere, which facilitates application of the limited-area model configuration in polar region. Semi-implicit semi-Lagrangian solver of the atmospheric core doesn't ensure dry-air mass conservation, and a global dry-air mass fixer is implemented into the YUNMA model by scaling the pressure with a varying global-mass ratio to the init. The mass adjustment doesn't disturb the dynamical and thermo-dynamical balance among wind, temperature and pressure at all because the three-dimensional pressure gradients are preserved. Conservation of the water vapour and cloud hydrometeors are ensured in the tracer transport processes by using a piecewise rational-function method (PRM) advection scheme with positivity-preserving feature (Xiao and Peng 2004). Three-dimensional hydrostatic atmosphere is introduced as a reference state to reduce computational error with the hydrostatic subtracting method, which is much better than the one-dimensional isothermal reference profile. Parallel computation is performed on rectangular subareas with very simple area division in the Yin or Yang zone with point-to-point data com-

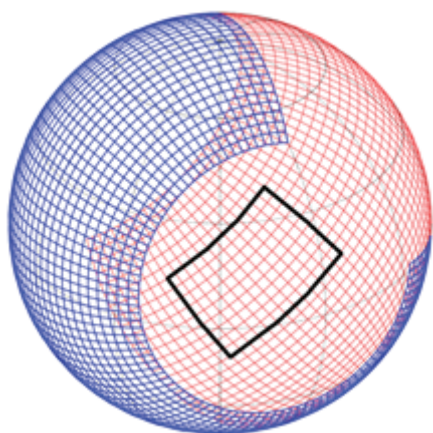


Fig.1 The rotated Yin–Yang grid and local refinement with grid nesting in the Yang zone

munications among processors (Chen et al. 2023).

2. RESULTS AND MODEL PERFORMANCE

The average parallelizing efficiency of the YUNMA dynamical core is of 90.65% in a 1600-processor computation. Convergence rate of the dynamical core is about order 2.3 in an idealized balanced flow test, as shown in Fig.2. Simulation of the baroclinic wave in a nesting grid refined the wave structure, especially the sharpness of the horizontal temperature gradient. The use of three-dimensional reference atmosphere significantly improved the wave propagation and its intensity (Zhao et al. 2024), which is comparable with the simulation by the spectral model ECHAM5. The one-dimensional reference isothermal atmosphere, however, tended to slow down the wave propagation and weaken the vorticity intensity because of the large error associate with the prognostic quantities. The semi-implicit average coefficient is set to be off-centred for stable integration because of the spurious orographic resonance.

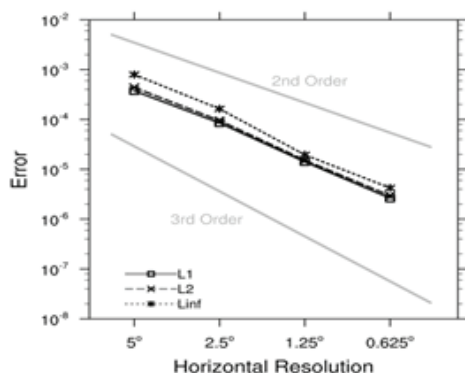


Fig. 2 Convergence rate of the YUNMA dynamical core tested with the three-dimensional balanced flow experiment (Jablonowski et al., 2008)

By using a physical parameterization package including the RRTMG long and short-wave radiation scheme, the MRF planetary boundary layer scheme, the new simplified Arakawa-Schubert cumulus scheme (NSAS), the double-moment microphysics scheme developed at CMA Earth System Modeling and Prediction Centre, and the Common Land Model, a real case simulation of the heavy precipitation of a cold vortex in North China with multi-level nesting YUNMA model showed clearly the enhanced rainfall in high-resolution model (Fig.3). The fitted terrain at boundaries of the nested grids successfully suppressing the boundary grid imprinting, and simulated smooth propagation of the rain bands and weather systems. The multi-level nest provides an efficient tool to improve the simulation with relatively cheap cost.

3. PHYSICAL SCHEME REFINEMENT

Diurnal cycle of the simulated precipitation is found quite different to the observations. Modifications of the trigger and closure conditions in the NSAS scheme greatly improved the diurnal cycle of local precipitation, especially in South China. Additional scale-aware arrangement of the MYNN2.5 boundary-layer scheme, with non-local turbulence being considered, obviously predicts reasonable spatial and temporal distribution of sea and land fogs.

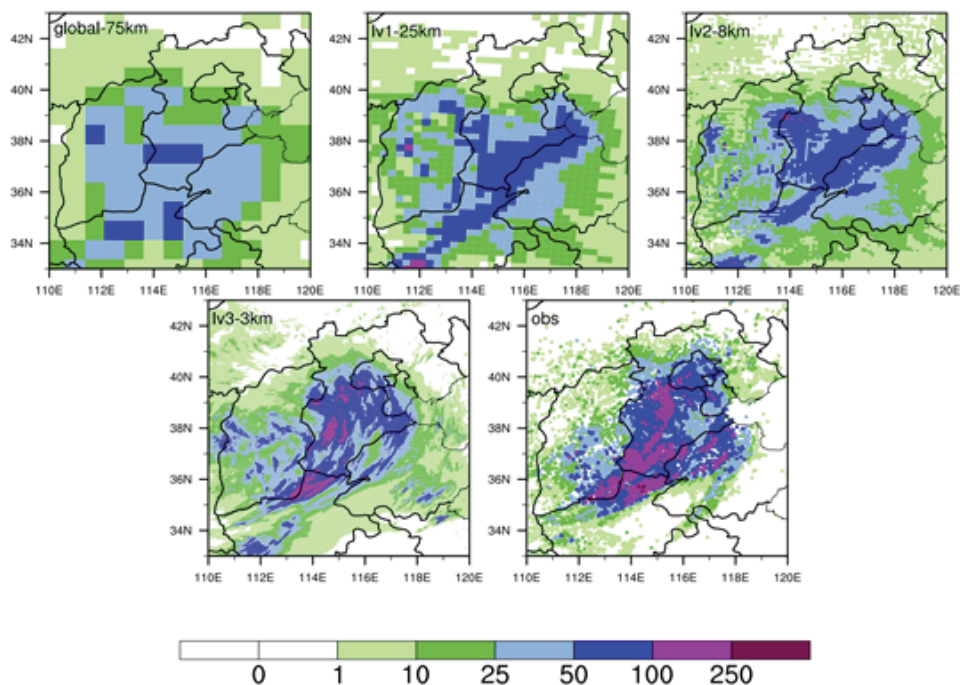


Fig. 3 Simulated precipitation with the multi-level nesting YUNMA model, i.e. 0.75° global (top left), 0.25° level-1 nesting (top middle), 0.83° level-2 nesting (top right) and 0.24° level-3 nesting grid (bottom left), in comparison with the surface observation (bottom right)

4. CONCLUSIONS AND DISCUSSIONS

The flexible global and limit-area non-hydrostatic model YUNMA is largely improved in both numerical and computing performances since last conference at Shanghai 2018. The coordinate rotation and local refinement of grid nesting computation help to improve the model accuracy and flexibility. The nesting-grid

simulation greatly improved the convective precipitation and also long-distance propagation of tropical cyclones because of the better uniformed grid distribution. Improved physical processes, such as the MYNN2.5 boundary-layer scheme and cumulus parameterization, benefit the detailed diurnal cycle of precipitation and fog simulation.

REFERENCES

- Chen, S., Y. Zhao, X. Peng, and X. H. Li, 2023, A Global-regional Unified Atmospheric Dynamic Core on the Yin-Yang Grid, *Mon. Wea. Rev.*, 151, 967-987, DOI: 10.1175/MWR-D-22-0079.1
- Xiao, F. and X. Peng, 2004, A convexity preserving scheme for conservative advection transport, *J. Comput. Phys.*, 198, 389-402.
- Zhao, Y. F., X. D. Peng, D. H. Chen, et al., 2024: Impact of a three-dimensional reference state in a global semi-implicit semi-Lagrangian non-hydrostatic atmospheric model on Yin-Yang grids. *J. Meteor. Res.*, 38(5), 901-922, doi: 10.1007/s13351-024-4022-0

Diagnosis and Numerical Modeling of an Explosive Cyclone over the Northwestern Pacific

*Sun Baitang^{1,2}, Li Pengyuan², Zhang Shuqin³,
Guo Jingtian⁴, and Fu Gang²*

¹ Laixi Meteorological Bureau, Qingdao 266622, China

² College of Oceanic and Atmospheric Sciences, Ocean University of China, Qingdao 266100, China / Corresponding author: pengyuanli@ouc.edu.cn

³ College of Ocean and Meteorology, Guangdong Ocean University, Zhanjiang 524008, China

⁴ North Marine Forecasting Center, Ministry of Natural Resources, Qingdao 266100, China

ABSTRACT

An explosive cyclone that took place over the Northwestern Pacific from 12 UTC 18 to 18 UTC 21 November 2007 was investigated. The synoptic situations and structure of this cyclone were documented by using the $1^\circ \times 1^\circ$ final analysis data of the National Center for Environmental Prediction. This cyclone developed explosively around 18 UTC 19 and reached its maximum deepening rate (MDR, 1.3 Bergeron) around 06 UTC 20 November 2007. At its MDR moment, the surface cyclone center was located in the downstream of the upper-level trough and northern entrance zone of the upper-level jet. The diagnosis using Zwack-Okossi equation suggested that cyclonic-vorticity advection and warm air advection acted to deepen this cyclone, while adiabatic cooling suppressed its development. In an investigation of this cyclone development,

numerical sensitivity results obtained by using the Weather and Research Forecasting model showed that the latent heat release in the lower level had less contribution, whereas the surface sensible and latent fluxes played important roles. With a warmer ocean surface, the cyclone tended to intensify. Two topography tests were designed to examine the mountain influences on the development of this cyclone: removing a mountain and doubling the height of a mountain. Results show that the Changbai Mountains suppressed the development of the cyclone by preventing the southern moisture air from invading the inland. Without the moisture air, no latent heat release occurs when this cyclone passes over the Changbai Mountains.

Key words:

explosive cyclone; Changbai Mountains; Zwack-Okossi equation; WRF model

SL-AV global atmosphere model: participation in WP-MIP project and further model development

*Mikhail Tolstykh^{1,2,3}, Rostislav Fadeev^{1,2}, Vladimir Shashkin^{1,2},
Gordey Goyman^{1,2}, Radomir Zaripov², Ekaterina Biryucheva²*

¹ Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences, Moscow, Russia

² Hydrometeorological Research Centre of Russia, Moscow, Russia

³ email: mtolstykh@mail.ru

1. INTRODUCTION

Recent advances in AI-based medium-range and subseasonal weather prediction motivated WMO WP-MIP project comparing different kind of weather prediction models (McTaggart et al, 2026). The preliminary WP-MIP results pose the inevitable question on future of 'classical' hydrodynamical models of the atmosphere in medium-range prediction. Moreover, recent introduction of AI-based AIFS SUBS subseasonal prediction model at ECMWF demonstrate the advantage of AI-based model at subseasonal time scale (Schloer et al, 2026). However, today it is somewhat difficult to imagine a further expansion of AI-model to seasonal and longer time scale, as the training dataset is limited in time to the satellite era started in 1979. So it makes sense to further develop and improve hydrodynamical models. Hybrid atmosphere models also show promising results.

2. SL-AV MODEL

SL-AV is the global atmosphere model applied for the operational medium and long-range weather forecast at Hydrometeorological Research Center of Russia

(Tolstykh et al 2015). SL-AV is the model acronym (semi-Lagrangian, based on Absolute-Vorticity equation). The model is developed at Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences (INM RAS) in cooperation with the Hydrometeorological Research Centre of Russia (HMCR). The dynamical core of this model uses the semi-implicit semi-Lagrangian time-integration algorithm, the vorticity-divergence formulation and fourth-order finite differences for approximation of non-advective terms of the equations (Tolstykh et al, 2017). Another feature of SL-AV is the possibility to use variable resolution in latitude.

3. IDENTIFYING THE PROBLEMS

We present some WP-MIP results for the global SL-AV atmosphere model forecasts with approximately 0.1 degrees horizontal resolution using both own initial data (OIC) and the same ECMWF initial data as all participating groups do (SIC).

Three areas having the potential for improvements in model scores are identified. The first one is the initial data quali-

ty and is currently beyond of the scope of SL-AV model developers. The shortwave radiation and deep convection parameterizations look as first ones needed the replacement or improvements.

terizations with ecRad package are presented. These are especially visible in the global precipitation field. The probable ways of deep convection parameterization improvement are discussed.

4. RECENT RESULTS AND FURTHER WORK

The results of replacement of CLIRAD SW and RRTMG LW radiation param-

REFERENCES

- McTaggart-Cowan R., Magnusson L., Polichtchouk I., Ackerley D., Koehler M., Casati B., Jan-Huey Chen, Hudson D., Ujiie M., Amir Aziz N., Bonavita M., Ben Bouallegue Z., de Burgh-Day C., Chamberland S., Kyounngmi Cho, Coelho C., Fadeev R., Fuentes M., Garcia Franco J., Gilbert C., Guimaraes B., Harris C., Harrold M., Husain S., James M., Kaltenbaugh A., Koch M., Kubota P., Eun-Hee Lee, Chen Li, Wei Li, Weiwei Li, Lledo L., Loveday N., Lussana C., Maalick Z., Mateyisi M., McGovern A., van der Merwe K., Miller, Mittermaier J., Mladek R., Newman K., Neves A., Pill J., Potthast R., Pradhan M., Rath S., Richardson D., Separovic L., Reboita M., Skok G., Srivastava A., Tolstykh M., Zhuo Wang, Woodham B., Fangling Yang, Zaripov R., Gan Zhang, Hongyan Zhu (2026) WP-MIP: An Artificial Intelligence, Hybrid, and Physically Based Model Intercomparison Project for Weather Prediction, accepted to Bull. Amer. Meteor. Soc. Doi: 10.48550/arXiv.2604.16643
- Schloer J., Tietsche S., Roberts C., Zampieri L., Lang S., Mertes G., Jones G., Chantry M., Vitart F. (2026) AIFS-SUBS: Extending Data-Driven Forecasting to Sub-Seasonal Timescales. doi: 10.48550/arXiv.2607.05100
- Tolstykh M., Shashkin V., Fadeev R., Goyman G. (2017) Vorticity-divergence semi-Lagrangian global atmospheric model SL-AV20: dynamical core. Geosci. Model Dev. 10, 1961–1983 (2017), doi:10.5194/gmd-10-1961-2017.
- Tolstykh, M., Fadeev R., Shashkin V., Zaripov R., Travova S., Goyman G. (2024) The SLAV072L96 model for long-range meteorological forecasts. Russian Meteorol. and Hydrol., 49, 576–586, doi:10.3103/S1068373924070021.

Rainfall-Informed SWMM–LSTM Surrogate Modelling for Storm Sewer Network Flow Forecasting within a Digital Twin Framework

Haoyue Wang¹, Miloš Stanić², Jasmina Moskovljević² and Čedo Maksimović¹

¹ Department of Civil and Environmental Engineering, Imperial, London, United Kingdom (hw2225@ic.ac.uk and c.maksimovic@imperial.ac.uk)

² Faculty of Civil Engineering, University of Belgrade, Belgrade, Serbia (mstanic@grf.bg.ac.rs and jasmina.moskovljevic@gmail.com)

1. INTRODUCTION

Urban drainage systems can respond rapidly to intense rainfall because urbanisation increases impervious cover, reduces infiltration and accelerates runoff conveyance. In small urban catchments, substantial hydraulic changes can therefore occur over timescales of minutes, creating a need for short-term forecasts of flow, peak timing and storage loading (Fletcher et al., 2013). These forecasts are preferably linked with (initial and boundary conditions taken from) larger scale NWP (Numerical Weather Prediction) model. Initial trials of linking these two types of models are performed within the RainGain EU project¹ and other projects in collaboration of Imperial College London (current name Imperial) and ENPC Paris (Gires et al., 2012; Wang et al., 2013). Physically based drainage models such as the U.S. Environmental Protection Agency Storm Water Management Model (EPA SWMM) can represent rainfall-runoff generation, conduit flow, storage, pumps and dynamic routing through drainage networks

(Rossman and Simon, 2022). However, repeated full-network simulation can be inconvenient when rapid and frequently updated forecasts are required (Piadeh et al., 2022). This study addresses these issues.

Machine-learning (ML) surrogates provide a complementary approach by learning selected responses from physically based simulations and then reproducing them rapidly (Garzón et al., 2022; Zahura et al., 2020). Long Short-Term Memory (LSTM) networks are well suited to analysis of sequential hydrological problems because they can retain information from previous forcing and system states (Hochreiter and Schmidhuber, 1997; Kratzert et al., 2018). Within a Digital Twin context, such a surrogate can be treated as one forecasting service connected to a physically based representation and other data services rather than as a replacement for the whole physical model (Pedersen et al., 2021).

The information required by a forecasting model changes with lead time. Recent hydraulic conditions may contain enough information for a very short

¹ <https://hmc.enpc.fr/portfolio-archive/raingain/>

time forecast, whereas a longer forecast increasingly depends on rainfall that occurs after the forecast issue time. A history-only model cannot distinguish between two futures that begin from similar hydraulic conditions but experience different subsequent rainfall. This study therefore developed and evaluated a rainfall-informed SWMM–LSTM surrogate for two-hour inflow prediction using the existing Blok 13 EPA SWMM model for New Belgrade, provided by the research team of the Faculty of Civil Engineering, University of Belgrade, led by Prof. Miloš Stanić and Jasmina Moskovljević. Flow through conduit COND303 to the CRPILISTE storage and pumping-station system was analysed. The work also compared internal chronological testing with independent synthetic-event verification to assess generalisation beyond the controlled rainfall profiles used during this development.

2. METHODOLOGY

The study used the existing Blok 13 EPA SWMM model supplied by the Faculty of Civil Engineering, University of Belgrade. The schematic of the network is presented in Fig 1.

In this network, COND303 is a 13.80 m long circular conduit connecting node J301 directly to the CRPILISTE storage unit. Its simulated flow therefore represents inflow entering the storage and pumping-station system rather than discharge through an individual pump. The surrogate prediction task was deliberately restricted to this management-relevant downstream response.

Synthetic design rainfall was generated



Fig 1. Schematic of the Block 13 case study drainage network. (Source: author no. 1, based on the supplied Block 13 SWMM model provided by the FCE-UB.)

from Belgrade design-rainfall information provided by Maksimović and Jandrić (2025), developed by Plavšić et al. (2018). Six return periods (2, 5, 10, 20, 50 and 100 years) and four storm durations (30, 60, 120 and 240 min) were combined with four simplified temporal rainfall distributions (storm profiles): constant, early-peaking, middle-peaking and late-peaking (Fig 2).

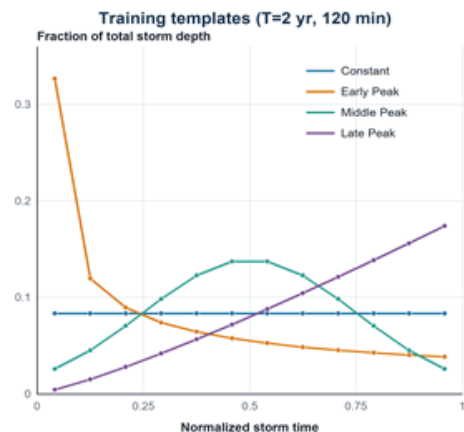


Fig. 2. Hypothetical temporal rainfall distributions used in the study: constant, early-peaking, middle-peaking and late-peaking profiles (Source: author no. 1).

This produced 96 storm templates at 10-min temporal resolution. Historical hourly precipitation from ERA5 reanalysis was used only to characterise inter-event dry periods (Hersbach et al., 2020; Zippenfenig, 2023). The storm library and empirical dry-period samples were combined into a continuous two-year virtual design and training sequences, not a reconstruction of observed Belgrade historic storms.

The synthetic rainfall sequences were then applied to the available Block 13 model. EPA SWMM transformed the rainfall forcing into hydraulic time series, from which rainfall intensity, CRPILISTE water depth and COND303 flow were aligned at a common 10-min learning interval. COND303 flow was the supervised target. Figure 3 summarises the complete workflow from rainfall generation to internal testing and frozen-model verification.

The common LSTM history consisted of six consecutive 10-min states, giving a 60-min lookahead. Each state contained rainfall intensity, CRPILISTE depth and COND303 flow. A 10-min history-only baseline first predicted one future flow value. The same input basis was then extended to direct multi-step forecasting, where 12 future values from +10 to +120 min were generated simultaneously. Because the outputs were direct rather than recursive, deterioration with lead time could not be attributed to feeding earlier LSTM predictions back into later forecasts.

Two diagnostic preprocessing experiments tested whether long dry periods were responsible for poor multi-step performance. One limited post-event

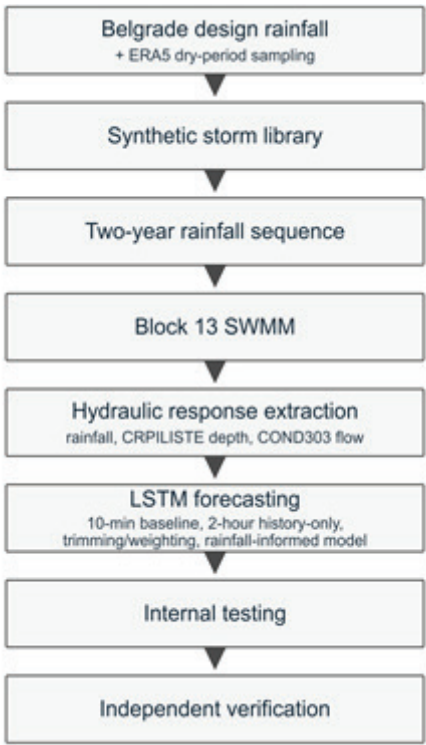


Fig. 3. Overall SWMM-LSTM surrogate-modelling workflow (Source: author no. 1).

dry periods in the training data to 24h, while the other combined trimming with increased weight on hydraulically active, high-rainfall and high-flow samples. These tests changed training emphasis but did not add new forecast information.

The retained two-hour model added a second input branch containing the 12 rainfall values over the forecast horizon. No future flow, water depth or target-derived feature was supplied. The historical branch used one LSTM layer with 64 units followed by 0.10 dropout; its representation was concatenated with the future-rainfall vector and passed through a dense layer with 32 ReLU units (Nair and Hinton, 2010) to

a 12-unit output. The model was trained using the Adam optimiser (Kingma and Ba, 2015) with a learning rate of 0.001, a batch size of 256 and a maximum of 60 epochs. Mean squared error was used as the loss function. Early stopping monitored validation loss with a patience of seven epochs and restored the best-performing weights. ReduceLROnPlateau (Keras, 2026) reduced the learning rate by a factor of 0.5 after four epochs without improvement, with a minimum learning rate of 1×10^{-6} . The continuous dataset was divided chronologically into approximately 70% training, 15% validation and 15% internal testing, with scalers fitted using training data only.

Because the 12 future rainfall values were taken directly from the known synthetic forcing, this configuration was a perfect-future-rainfall experiment. It quantified an idealised upper bound on the information value of forecast-period rainfall rather than the expected skill of an operational meteorological forecast. Model skill was evaluated using RMSE, MAE and R^2 on inverse-transformed flow in L/s.

After model development, the retained rainfall-informed LSTM and all training-fitted preprocessing parameters were frozen. Twelve additional irregular synthetic extreme rainfall events were generated independently, covering 60-, 90- and 120-min durations and four interpolated depth-equivalent return periods. Their 10-min rainfall structures were screened against reconstructed training templates using a cosine-similarity threshold. Each accepted event was simulated with a fresh SWMM working copy, and rolling two-hour forecasts were issued every 10

min without retraining or scaler refitting. This produced 1,440 forecast cases across the 12 events.

3. RESULTS

The 10-min history-only model reproduced the SWMM-derived COND303 inflow with high accuracy on the internal chronological test set, achieving RMSE = 18.07 L/s, MAE = 5.53 L/s and $R^2 = 0.9944$. It substantially outperformed a persistence benchmark, for which RMSE was 101.36 L/s and R^2 was 0.825. Recent rainfall, CRPILISTE depth and COND303 flow therefore contained sufficient information for very-short-term prediction.

Extending the same history-only information to a direct two-hour forecast produced a large deterioration. When all 12 horizons were pooled, RMSE increased to 180.01 L/s, MAE to 34.93 L/s and R^2 fell to 0.4484. At +10 min the model still achieved $R^2 = 0.989$, but R^2 declined to 0.371 at +60 min and 0.151 at +120 min. For active-flow targets, R^2 became negative at the longest lead times. Dry-period trimming and active-flow weighting did not restore performance, confirming that sample imbalance was not the main limitation.

Adding forecast-horizon rainfall changed the result substantially. Across the 12 lead times, the rainfall-informed model achieved RMSE = 24.30 L/s, MAE = 4.45 L/s and $R^2 = 0.9900$. At +120 min, RMSE fell from 223.37 L/s for the history-only model to 26.73 L/s, while R^2 increased from 0.151 to 0.988. The reduction in +120-min RMSE was approximately 88%, and MAE decreased

by approximately 91.5%. Figure 4 shows that the separation between the two input configurations grows strongly with lead time.

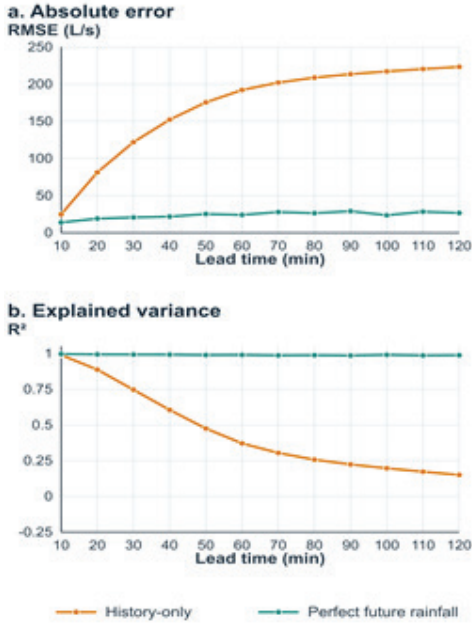


Fig. 4. Internal history-only and perfect-future-rainfall performance across forecast lead times (Source: author no. 1).

The frozen rainfall-informed model was then evaluated on the 12 independent-generated irregular synthetic events. Across all 1,440 rolling forecast cases, RMSE was 411.60 L/s, MAE was 278.29 L/s, R^2 was 0.8482 and mean bias was +75.21 L/s. Event-level R^2 ranged from 0.7243 to 0.9046. These results indicate useful but materially reduced generalisation relative to the internal rainfall-informed test.

The external hydrographs also revealed limitations that were less apparent from pooled internal statistics. Errors were concentrated around rising limbs and peaks, with premature predicted re-

sponse in some cases, smoothing of sharp peaks, variable peak-timing error and a tendency toward positive cumulative-volume bias. Bias was positive at most forecast horizons and increased at some of the longest lead times. The contrast between the internal R^2 of 0.9900 and the independent-event R^2 of 0.8482 demonstrates a material generalisation gap for the same frozen rainfall-informed model.

4. DISCUSSION

The experiments show that the main two-hour forecasting limitation was an information problem rather than a simple model-capacity or sample-balance problem. At very short lead times, current flow and storage depth contain hydraulic memory of rainfall that has already entered and propagated through the drainage system. Over a longer horizon, however, an increasing part of the response is controlled by rainfall occurring after the forecast origin. Two cases with nearly identical history can therefore diverge strongly if subsequent rainfall differs.

This explains why trimming and weighting did not recover long-horizon skill. Those procedures changed the frequency or optimisation weight of certain training samples, but they did not tell the model whether an intense rainfall burst would occur 30, 60 or 100 min after forecast issue. In contrast, supplying rainfall over the forecast horizon introduced the missing external forcing. The large improvement to $R^2 = 0.9900$ therefore shows that future rainfall was a central information source for two-hour prediction in this rapidly responding urban drainage system.

The perfect-rainfall result is nevertheless an upper-bound experiment. In an operational system, future rainfall would need to come from radar nowcasting, numerical weather prediction or a combined product, all of which contain timing, intensity and spatial uncertainty. Short-term urban drainage studies have shown that the usefulness of rainfall forecasts depends on lead time and forecast source (Thorndahl et al., 2013). The present experiment isolates the ability of the LSTM to use forecast-period rainfall when that forcing is known; it does not quantify the additional degradation that would arise from meteorological forecast error.

The independent verification provides an equally important result. A chronological split prevented direct temporal overlap between training and internal testing, but the rainfall in both subsets was generated from the same four controlled temporal-profile families. The irregular external events were therefore a stronger test of structural generalisation. The decrease from internal $R^2 = 0.9900$ to external $R^2 = 0.8482$ shows that conventional internal testing substantially overstated performance on less familiar rainfall structures, even though the frozen model retained meaningful predictive skill without retraining.

For pumping-station applications, hydrograph shape is important in addition to aggregate goodness-of-fit. Peak magnitude, peak timing and forecast-window volume influence different operational decisions. A forecast can underestimate the sharpest instantaneous peak while still overestimating cumulative volume if it rises too early or remains elevated for

too long. The positive external bias and variable peak-timing behaviour therefore identify practical aspects that should be improved before deployment.

Within a wider Digital Twin architecture, the retained LSTM is best interpreted as one rapid forecasting service. SWMM or sensor-assisted state estimation could provide current hydraulic conditions, a rainfall nowcast could provide the future forcing, and the surrogate could return two-hour predictions of pumping-station inflow. A complete operational implementation would additionally require field observations, automated data transfer, quality control, state updating and forecast communication. The present study isolates and tests the forecasting component before those wider integrations are attempted.

5. CONCLUSIONS

A rainfall-informed SWMM–LSTM surrogate was developed for two-hour prediction of inflow to the CRPILISTE storage and pumping-station system in the Block 13 drainage model. Four main conclusions follow. First, recent rainfall and hydraulic conditions were sufficient for highly accurate 10-min prediction, with $RMSE = 18.07$ L/s and $R^2 = 0.9944$. Second, the same history-only information was insufficient across the full two-hour horizon, and direct forecasting performance deteriorated strongly with lead time.

Third, supplying known rainfall over the forecast horizon restored high internal skill, producing pooled $RMSE = 24.30$ L/s and $R^2 = 0.9900$. This identifies forecast-period rainfall as a critical input for

longer-range inflow prediction. Fourth, independent verification on 12 irregular synthetic extreme rainfall events revealed a substantial generalisation gap: the frozen model retained useful skill ($R^2 = 0.8482$) but showed greater error in peak magnitude, timing and accumulated volume.

The study therefore demonstrates the feasibility of using a rainfall-informed

LSTM as a rapid surrogate for a selected SWMM hydraulic response while also showing why independent event testing is necessary. The next stage should replace perfect future rainfall with realistic rainfall forecasts, incorporate observed hydraulic data and expand the diversity of withheld extreme-event structures before the surrogate is considered for operational Digital Twin deployment.

REFERENCES

- Fletcher, T., Andrieu, H. and Hamel, P. (2013) Understanding, management and modelling of urban hydrology and its consequences for receiving waters: A state of the art, *Advances in Water Resources*, 51, 261–279.
- Garzón, A., et al. (2022) Machine Learning-Based Surrogate Modeling for Urban Water Networks: Review and Future Research Directions, *Water Resources Research*, 58.
- Gires, A., Schertzer, D., Tchiguirinskaia, I., Lovejoy, S., Onof, C., Maksimovic, C., Simoes, N., 2012. Impact of small-scale rainfall uncertainty on urban discharge forecasts. In *Weather Radar and Hydrology*, Eds: Moore, J., Cole, S., Illingworth, A.: IAHS Publ. 351. p 400–406
- Hersbach, H., et al. (2020) The ERA5 global reanalysis, *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049.
- Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory, *Neural Computation*, 9(8), 1735–1780.
- Keras (2026) ReduceLROnPlateau. Keras API documentation. Available at: https://keras.io/api/callbacks/reduce_lr_on_plateau/ [Accessed 7 September 2026].
- Kingma, D. P. and Ba, J. (2015) Adam: A Method for Stochastic Optimization, *International Conference on Learning Representations (ICLR)*.
- Kratzert, F., et al. (2018) Rainfall-Runoff modelling using Long-Short-Term-Memory (LSTM) networks, *Hydrology and Earth System Sciences Discussions*, 1–26.
- Maksimović, Č. and Jandrić, B. (2025) Advanced Urban Drainage: Lesson 2 Parts 1 and 2, *Urban Water and Green Infrastructure: Integrated Planning and Management*, Imperial College London.
- Nair, V. and Hinton, G. E. (2010) Rectified Linear Units Improve Restricted Boltzmann Machines, *Proceedings of the 27th International Conference on Machine Learning*, pp. 807–814.
- Pedersen, A., et al. (2021) Living and Prototyping Digital Twins for Urban Water Systems: Towards Multi-Purpose Value Creation Using Models and Sensors, *Water*, 13.

- Piadeh, F., Behzadian, K. and Alani, A. M. (2022) A critical review of real-time modelling of flood forecasting in urban drainage systems, *Journal of Hydrology*, 607, 127476.
- Plavšić, J., Jelušić, D. and Jevtić, D. (2018) Analiza efekata planiranih akumulacija za ublažavanje velikih voda na slivu Topčiderske reke (Study of the effects of planned retention ponds for flood mitigation in the Topčider River catchment), *Vodoprivreda*, 50(1–3), pp. 47–58.
- Rossman, L. A. and Simon, M. A. (2022) Storm Water Management Model User's Manual Version 5.2, U.S. Environmental Protection Agency, EPA/600/R-22/030.
- Thorndahl, S., et al. (2013) Comparison of short-term rainfall forecasts for model-based flow prediction in urban drainage systems, *Water Science and Technology*, 68(2), 472–478.
- Wang, L.-P., Ochoa-Rodríguez, S., Simoes, N., Onof, C. & Maksimović, Č. (2013). Radar-raingauge data combination techniques: a revision and analysis of their suitability for urban hydrology. *Water Science & Technology*, 68 (4), 737-747.
- Zahura, F., et al. (2020) Training Machine Learning Surrogate Models From a High-Fidelity Physics-Based Model: Application for Real-Time Street-Scale Flood Prediction in an Urban Coastal Community, *Water Resources Research*, 56.
- Zippenfenig, P. (2023) Open-Meteo.com Weather API Zenodo. <https://doi.org/10.5281/zenodo.7970649>.

Coupling numerical precipitation forecasts with flood-oriented Local Digital Twins: The Kolubara-Tamnava case

*Jelena Batica, Luka Stojadinović, Marija Milovanović,
and Nikola Milivojević*

Jaroslav Cerni Water Institute LLC Belgrade, Belgrade, Serbia (jelena.batica@jcerni.rs)

1. INTRODUCTION

Extreme rainfall, fluvial flooding, and flash-flood propagation in the Kolubara-Tamnava system require modelling approaches that connect meteorological forcing with hydrological, hydraulic, and impact-based representations across multiple spatial scales. This extended abstract presents a flood-oriented Local Digital Twin framework for the Kolubara-Tamnava basin, designed to couple numerical precipitation forecasts with hydrological-hydrodynamic modelling and decision-relevant flood indicators, with specific attention to the Obrenovac area and exposed critical infrastructure. Rather than treating numerical weather prediction only as a stand-alone forecasting product, the framework positions forecast rainfall as a dynamic forcing input whose uncertainty propagates through the Digital Twin and affects inundation timing, flood depth, flood extent, and operationally relevant impact indicators. This framing is consistent with current flood digital twin literature, which increasingly emphasises integrated, impact-oriented workflows rather than isolated model applications (Hlal et al., 2025; Kaynak et al., 2025).

The operational stakes are well illustrat-

ed by the floods of May 2014, when record rainfall over the Sava and Kolubara basins inundated large parts of city of Obrenovac and forced the evacuation of tens of thousands of residents. The case is rooted in a multiscale modelling concept in which the Kolubara basin is treated as the main fluvial demonstration domain and the Tamnava catchment as a representative sub-basin for more rapid runoff and flash-flood dynamics. Such a configuration allows the same modelling architecture to address basin-scale rainfall generation, tributary response, river flood-wave routing, urban flood-plain dynamics, and consequences for district-scale critical assets.

Similar multilevel architectures have been identified as central to flood digital twins, especially where real-time data, hydraulic models, and infrastructure-oriented decision layers must be connected within a single analytical system (Hlal et al., 2025; Lo et al., 2026). In that sense, the Kolubara-Tamnava system offers a suitable bridge between numerical weather and climate modelling on the one hand and practical flood-risk reduction and climate-resilience planning on the other.

2. MULTISCALE DIGITAL TWIN CONCEPT

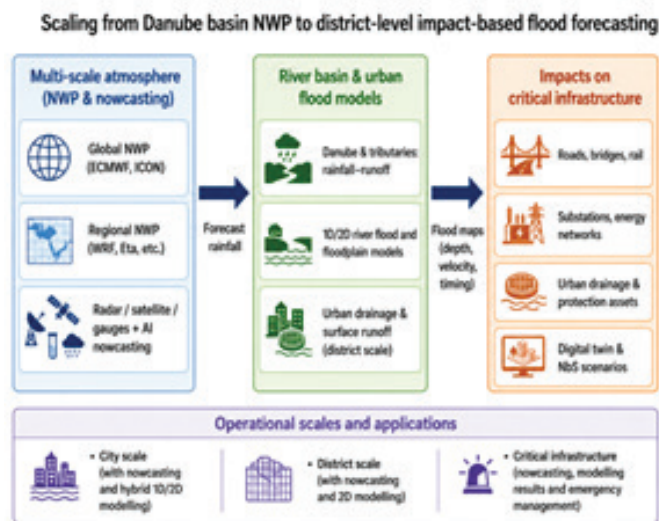


Figure 1: Scaling from basin-scale numerical weather prediction and nowcasting to district-level, impact-based flood forecasting and critical-infrastructure assessment in the Danube-Kolubara system

Figure 1 schematically summarises the general multi-scale chain from global and regional numerical weather prediction, radar- and gauge-supported nowcasting, and basin-scale precipitation forcing toward river-basin, floodplain, and district-level flood models, and ultimately toward impact assessment for critical infrastructure.



Figure 2: Multiscale spatial and modelling nesting from the Danube-Sava-Kolubara-Tamnava basin to the Obrenovac flood footprint and district critical infrastructure within the Local Digital Twin framework

Figure 2 presents the specific nesting logic used in the Kolubara-Tamnava concept, from the Danube basin and Sava corridor down to the Kolubara basin, including Tamnava, the Obrenovac flood footprint, and district-level critical infrastructure, along with the corresponding modelling levels. Taken together, these figures support the central argument that the performance of meteorological inputs cannot be adequately assessed without examining how they drive downstream flood responses in a coupled modelling environment, a point increasingly emphasised in recent impact-based flood modelling studies (Kaynak et al., 2025).

3. FEWS “KOLUBARA” AND DATA-MODEL BASIS

The data and model basis for this framework are unusually strong. The Kolubara basin is already supported by an opera-

tional early warning system, FEWS “Kolubara”, developed through a feasibility and concept phase followed by implementation phases centred around a dedicated flood management control centre operated by the Public Water Management Company “Srbijavode”(Jaroslav Černi Water Institute, 2020; Jaroslav Černi Water Institute, 2024). The system combines an acquisition server, a central server, and a computational server, together with user tools for monitoring hydro-meteorological conditions, analysing and forecasting hydrological and hydraulic variables, operating 2D model components, and managing assimilation services.

The modelling chain rests on a basin-wide hydrological and hydraulic concept defined in the feasibility study for FEWS “Kolubara” (Jaroslav Černi Water Institute, 2020). The hydraulic component is a HEC-RAS 1D unsteady model covering the major watercourses and designated flood-prone reaches of the basin, with a total modelled river length of approximately 410.7 km, including the Kolubara main stem, the Tamnava and Ub rivers, and significant tributaries such as the Ljig, Toplica, Ribnica, Turija, Beljanica, and Peštan. The model configuration explicitly represents levees and other line flood-protection structures relevant for operational flood defence. Because the HEC-RAS platform supports combined 1D and 2D simulation, flood-prone urban reaches can be resolved with 2D inundation domains while the basin-wide system retains the efficiency of 1D routing for early-warning purposes. The feasibility study already envisaged preliminary calibration of both the hydrological and hydraulic models, with further cali-

bration foreseen in subsequent phases of system development.

Implementation followed this concept in stages. The system was first piloted on the Tamnava catchment and later extended to the remaining parts of the Kolubara basin: a supplementary network of 19 additional rain gauge stations and 7 hydrological stations was installed, and a coupled hydrological-hydraulic model was implemented for the non-Tamnava part of the basin and linked to the existing Tamnava model. The software components were expanded accordingly, with automatic data ingestion from the new stations, updated quality-control modules, an extended rainfall-runoff and flood-routing model domain, and the incorporation of relevant flood-protection structures into the post-processing of model outputs (Jaroslav Černi Water Institute, 2024).

4. FORECAST COUPLING AND DIGITAL TWIN FUNCTIONALITY

The proposed Local Digital Twin therefore builds on an already implemented operational early warning system and focuses on coupling numerical precipitation forecasts and ensemble products to this existing hydro-meteorological and modelling infrastructure, rather than starting from a purely conceptual prototype. In this context, the term Digital Twin is used deliberately rather than as a relabelling of the existing early warning system: the framework is intended to maintain a continuously updated digital representation of the basin state through data assimilation, to support two-way interaction between observations and models, and to enable on-demand what-

if scenario analysis, properties that recent critical assessments identify as distinguishing digital twins from conventional forecasting chains (Lo et al., 2026).

In practical terms, the Digital Twin can ingest precipitation forecasts from regional numerical weather prediction systems, including Eta-type and other regional products, translate them into basin-scale forcing fields, and propagate the resulting runoff and river response toward floodplain and urban impact layers. Candidate inputs include operational regional products from the Republic Hydrometeorological Service of Serbia and ECMWF high-resolution and ensemble forecasts. Lead time is a central constraint in this design. Whereas the Kolubara main stem can be served by day-ahead regional forecasts, the Tarnava sub-basin's response time is measured in hours. It lies at the limit of convective-scale forecast skill, which is the principal argument for strengthening the radar-supported nowcasting component.

The same framework can also accommodate future radar-based nowcasting or assimilation improvements, which are already being considered as a natural continuation of current work on radar-supported meteorological inputs.

This is consistent with broader developments in flood digital twins, where radar, sensors, and ensemble products are increasingly used to strengthen short-lead-time decision support (Hlal et al., 2025; MohammadZadeh et al., 2025). The conceptual basis for this coupling also aligns with the systems perspective of Blue-Green Solutions and integrated urban water planning proposed by Božović et al. (2017), particularly where flood mitigation, infrastructure performance, and

adaptation measures must be considered jointly.

5. FLOOD-ORIENTED FORECAST EVALUATION AND ENSEMBLES

Methodologically, the key issue is not only whether forecast rainfall matches observations at the station level, but whether forecast errors matter in ways that are hydrologically and operationally significant. For that reason, the proposed framework places emphasis on flood-oriented forecast evaluation. Numerical precipitation forecasts are first transformed to the modelling scales relevant for the basin and sub-basin domains through temporal aggregation and spatial matching, after which they are used as forcing in the hydrological-hydrodynamic chain represented by the Digital Twin. Forecast performance can then be evaluated at two connected levels: first, through conventional meteorological indicators such as bias, RMSE, event detection skill, and ensemble spread; and second, through impact-oriented indicators such as the timing of flood-wave arrival, simulated peak water levels, inundation depth, inundation extent, and sensitivity of critical infrastructure exposure. To attribute impact-level errors to the meteorological forcing rather than to the hydrological and hydraulic models themselves, simulations driven by observed precipitation are used as the reference against which forecast-driven runs are compared, allowing forcing error and model error to be separated. This distinction is consistent with current thinking that forecast evaluation should move beyond generic meteorological verification toward consequence-aware and decision-oriented analysis (Kaynak et al., 2025).

This distinction is important because acceptable meteorological skill does not automatically imply acceptable flood-prediction skill. A forecast that modestly underestimates basin-average rainfall, delays the precipitation maximum, or misplaces the spatial core of a convective event may still appear reasonable in standard verification terms, yet produce substantial differences in routed flood volumes and inundation patterns in a vulnerable reach such as Obrenovac. Conversely, an ensemble system may contain members that better capture the relevant hydrological trigger even when the deterministic forecast does not, which supports the use of probabilistic thinking in flood-oriented Digital Twins. For an impact-based warning framework, these distinctions are more meaningful than purely point-wise rainfall verification because emergency decisions depend on flood consequences, not on meteorological scores alone. Propagating ensembles through the hydraulic chain in operational time is nevertheless computationally demanding, particularly where 2D inundation domains are involved; the framework therefore anticipates pragmatic strategies such as reduced ensemble subsets, pre-computed scenario libraries, and surrogate or machine-learning emulators for the most expensive model components.

6. OUTLOOK: CLIMATE STRESS TESTING AND FUTURE WORK

At the present stage, the work should be understood as a structured case for coupling forecast verification with flood-impact modelling rather than as a finished,

exhaustive statistical assessment. The available material already supports a robust methodological narrative: a nested modelling system exists; archived numerical forecasts and observations exist; an operationally relevant hydraulic and 2D-capable modelling application exists; and there is a clear path toward integrating these elements into a Local Digital Twin for flood early warning and climate-risk analysis.

A complete set of forecast-skill statistics is still a work in progress, but the current stage is sufficient to define the framework, justify the coupled approach, and demonstrate why flood-oriented evaluation is necessary.

Preliminary insights from the Kolubara-Tamnava concept suggest that the sensitivity of flood outputs to meteorological forcing is high enough to justify this integrated framing. In flood-prone and infrastructure-exposed areas such as Obrenovac, relatively moderate errors in forecast rainfall intensity or timing can alter the simulated arrival time of the flood wave, maximum inundation depth, and the spatial footprint of floodplain activation. Such sensitivity is especially relevant in a system that combines main-river flooding, tributary response, and urban-scale consequences, because uncertainty can amplify as the signal moves from basin rainfall to local impact. This is precisely the type of cross-scale propagation that a Local Digital Twin is intended to capture and communicate more effectively than isolated model components (Hlal et al., 2025).

The climate-change dimension adds another level of relevance. The existing application of current-climate and

RCP8.5-type conditions in the Obrenovac protection study shows that the same modelling structure can be used not only for real-time or near-real-time forecasting, but also for stress-testing the system under future extremes (Jaroslav Černi Water Institute, 2024). Although RCP8.5 is increasingly regarded as a high-end rather than a central trajectory, it remains useful here precisely as an upper-bound stress test, and extending the analysis to SSP-based scenarios is a natural refinement. In a Digital Twin context, this means that the framework can support both operational and planning functions: event-based forecasting, preparedness exercises, infrastructure sensitivity analysis, and exploration of adaptation or nature-based solution scenarios.

This dual use is especially valuable for river basins where flood management must increasingly combine short-term warning with long-term climate adaptation, and where blue-green measures are assessed as system components rather than isolated interventions (Božović et al., 2017).

From the perspective of the conference theme, the Kolubara-Tamnava case offers a practical interpretation of “now, and vision of the future”. The “now” is represented by established FEWS infrastructure, numerical weather prediction systems, existing hydraulic modelling practice, and dense station observations. The “vision of the future” is represented

by tighter coupling between forecasts and impact models, increased use of ensemble and radar-supported inputs, and the evolution of these components into operational Local Digital Twins capable of supporting district-level flood decisions and infrastructure-oriented warnings. In that sense, the contribution does not argue for replacing physically based modelling, but for making it more actionable through better integration across scales, models, and decision layers.

The Kolubara-Tamnava Local Digital Twin case therefore offers a scientifically credible and operationally relevant test-bed for coupling numerical precipitation forecasts with hydrological-hydrodynamic flood modelling. Its value lies not only in the availability of multiple data streams and model components, but in the possibility of tracing forecast uncertainty all the way to district-scale impacts on critical infrastructure. Future work should concentrate on systematic forecast verification against the observed hydro-meteorological archive, ensemble-based characterization of uncertainty, and the incorporation of radar-supported nowcasting and post-processing methods into the same flood-oriented Digital Twin framework. Such a pathway would strengthen both scientific understanding of forecast skill and the operational usefulness of numerical weather and climate modelling in a flood-risk management setting.

REFERENCES

- Božović, R., Maksimović, Č., Mijić, A., Smith, K. M., Suter, I., & van Reeuijk, M. V. (2017). Blue Green Solutions: A Systems Approach to Sustainable and Cost Efficient Urban Development. EIT Climate-KIC.

- Hlal, M., Munyaka, J.-C. B., Chenal, J., Azmi, R., Diop, E. B., Bounabi, M., Ebnou Abdem, S. A., Adou Sidi Almouctar, M., & Adraoui, M. (2025). Digital Twin Technology for Urban Flood Risk Management: A Systematic Review of Remote Sensing Applications and Early Warning Systems. *Remote Sensing*, 17(17), 3104. <https://doi.org/10.3390/rs17173104>.
- Jaroslav Černi Water Institute. (2020). FEWS “Kolubara”: Feasibility Study and Concept for the Development and Implementation of an Early Warning System for the Kolubara Basin. Internal report, Belgrade.
- Jaroslav Černi Water Institute. (2024). FEWS “Kolubara”: Phase B.2 – Technical Documentation of the Control Centre and Coupled Hydrological–Hydraulic Model Implementation. Internal report, Belgrade.
- Kaynak, S., Kaynak, B., Mermer, O., & Demir, I. (2025). City-scale digital twin framework for flood impact analysis: Integrating urban infrastructure and real-time data analytics. *Urban Climate*, Article 102640. <https://doi.org/10.1016/j.uclim.2025.102640>.
- Lo, T. H., Malmström, A., & Juhola, S. (2026). Critical assessment on digital twins implementation for flood risk management and governance: A systematic literature review. *International Journal of Disaster Risk Reduction*, in press.
- MohammadZadeh, F., Eghbalian, H., Gheibi, M., Yeganeh-Khaksar, R., Ghazikhani, A., & Behzadian, K. (2025). A Digital Twin Framework for Real-Time Flood Monitoring and Multidimensional Prediction: A case study in Iran. *EGU General Assembly 2025*, abstract EGU25-20713. <https://doi.org/10.5194/egusphere-egu25-20713>

Numerical Modeling Study of “Millipede Clouds”

Gang Fu, Zhongshi Wang, Xinyu Zhang, Weikang Sun, Pengyuan Li*

Department of Marine Meteorology, Ocean University of China

*E-mail: fugangouc@qq.com; fugang@ouc.edu.cn

ABSTRACT

“Millipede Clouds” are a rarely documented type of stratocumulus characterized by a long, meandering central axis (hundreds to thousands of kilometers) with numerous well-organized radial cloud arms extending tens of kilometers on both sides, resembling the legs of a millipede (Fig. 1). Based on MODIS and GOES satellite imagery as well as ERA5 reanalysis data from 2012 to 2021, we systematically analyzed the morphological features, global spatial distribution, seasonal variation, vertical structure, and

favorable thermodynamic environments of “Millipede Clouds” (Fu et al., 2023). Results show that these clouds mainly occur over subtropical oceans, with the high-frequency over the Eastern Pacific of the Southern Hemisphere, and a prominent seasonal peak during June–August. As typical low-level stratocumulus, they have cloud-top heights of 800–1600 m (800–950 hPa) and cloud-top temperatures of 270–290 K, and are closely associated with a low-level air temperature inversion layer (Fu et al., 2023). To explore

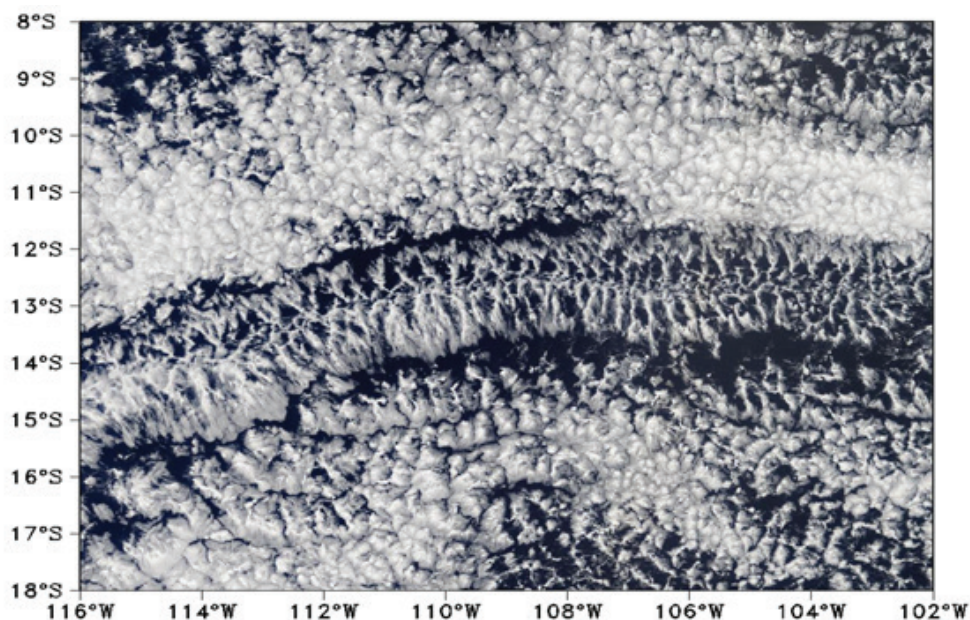


Figure 1. Satellite image at 17:30 UTC 3 September 2017

their mechanisms, a numerical modeling study was conducted using the Weather Research and Forecasting (WRF) model with a two-nested domain configuration (6-km and 2-km horizontal resolutions, 55 vertical levels). Nine numerical sensitivity experiments were designed with various combinations of cloud micro-physics and atmospheric boundary layer (ABL) schemes (Table 1). But only six numerical experiments (A-F) can be run without stopping. By comparing simulated cloud water mixing ratio and spatial structure with ERA5 reanalysis data and satellite imagery, we verified that the NSSL two-moment micro-physics scheme coupled with the Yonsei Uni-

versity (YSU) ABL scheme might reproduce the morphological characteristics, horizontal distribution, and vertical moisture structure of the studied “Millipede Cloud” case. It is confirmed that low-level moisture convergence, stable atmospheric stratification, and weak vertical mixing are the key dynamic and thermodynamic conditions for the formation of “Millipede Clouds”. This study provides the first systematic observational and numerical modeling investigation of “Millipede Clouds”. However, their complex formation mechanisms need more numerical experiments for the in-depth understanding of such marine cloud systems.

Table 1. Specifications of various schemes of WRF modeling

Schemes	Micro-physics	Atmospheric Boundary Layer	Long-wave Radiation	Short-wave Radiation	Run
A	NSSL 2-moment scheme (Mansell et al., 2010)	Yonsei University scheme (Hong et al., 2006)	RRTM scheme (Iacono et al., 2008)	Dudhia scheme (Dudhia, 1989)	√
B	NSSL 2-moment scheme (Mansell et al., 2010)	Mellor-Yamada-Janjic scheme (Janjic, 1994)	RRTM scheme (Iacono et al., 2008)	Dudhia scheme (Dudhia, 1989)	√
C	NSSL 2-moment scheme (Mansell et al., 2010)	Shin-Hong scheme (Shin and Hong, 2015)	RRTM scheme (Iacono et al., 2008)	Dudhia scheme (Dudhia, 1989)	√
D	Milbrandt-Yau double-moment 7-classscheme (Milbrandt and Yau, 2005a, b)	Yonsei University scheme (Hong et al., 2006)	RRTM scheme (Iacono et al., 2008)	Dudhia scheme (Dudhia, 1989)	√
E	Milbrandt-Yau double-moment 7-classscheme (Milbrandt and Yau, 2005a, b)	Mellor-Yamada-Janjic scheme (Janjic, 1994)	RRTM scheme (Iacono et al., 2008)	Dudhia scheme (Dudhia, 1989)	√
F	Milbrandt-Yau double-moment 7-classscheme (Milbrandt and Yau, 2005a, b)	Yonsei University scheme (Hong et al., 2006)	RRTM scheme (Iacono et al., 2008)	Dudhia scheme (Dudhia, 1989)	√
G	HUJI spectral bin micro-physics (Khain et al., 2004)	Yonsei University scheme (Hong et al., 2006)	RRTM scheme (Iacono et al., 2008)	Dudhia scheme (Dudhia, 1989)	×
H	HUJI spectral bin micro-physics (Khain et al., 2004)	Mellor-Yamada-Janjic scheme (Janjic, 1994)	RRTM scheme (Iacono et al., 2008)	Dudhia scheme (Dudhia, 1989)	×
I	HUJI spectral bin micro-physics (Khain et al., 2004)	Yonsei University scheme (Hong et al., 2006)	RRTM scheme (Iacono et al., 2008)	Dudhia scheme (Dudhia, 1989)	×

REFERENCE:

Fu, G., Wang, Z.-S., Zhang, X.-Y., Sun, W.-K., Li, P.-Y., & Wang, R.-Y. (2023). First observational perspectives of “Millipede Clouds” over the Eastern Pacific Ocean. *Geophysical Research Letters*, 50, e2023GL106265. <https://doi.org/10.1029/2023GL106265>.

Keywords: Millipede Clouds; Stratocumulus; WRF Modeling; Satellite Observation; Cloud Morphology; Air Temperature Inversion; Atmospheric Boundary Layer scheme.

Rain Gauge–Radar Merging Improves Short-Term Rainfall Nowcasts and the Sewer Overflow Forecasts

Surya Prayoga and Čedo Maksimović

Department of Civil and Environmental Engineering, Imperial College London, London, United Kingdom (surya.prayoga25@imperial.ac.uk and c.maksimovic@imperial.ac.uk)

1. INTRODUCTION

Urban catchments respond to rainfall on timescales of minutes, and their runoff is shaped as much by where rain falls as by how much. Fine-scale rainfall forecasting (nowcasting) and the urban flooding it drives are preferably linked to larger-scale Numerical Weather Prediction (NWP) models, from which initial and boundary conditions are taken. Initial trials of linking these two model types have been performed within the RainGain EU project¹ and in other collaborations between Imperial College London and ENPC Paris (Gires et al., 2012; Wang et al., 2013).

Berne et al. (2004) derived minimum required resolutions of 1 minute and 1.5 km below 10 hectares, and Ochoa-Rodríguez et al. (2015) found resolution-driven peak discharge errors falling from up to 250% around 1 hectare to roughly 50% near 800 hectares. Resolving rainfall at that scale requires weather radar, preferably X-band for its finer resolution, but radar carries systematic bias relative to ground measurement while gauges are accurate yet point-based. Merging combines the two at each grid location in proportion to their relative uncertainties, for better calibration of

the radar signal even in real time (Todini, 2001; Wang, 2012). Whether the improvement is worth having depends on what happens downstream, and the evidence there is not encouraging: a portable X-band radar deployed over Central London in 2013 under RainGain, itself gauge-adjusted following Todini (2001), showed no significant advantage over the coarser national product when fed into a semi-distributed catchment model (Ochoa-Rodríguez et al., 2014). The merge was applied; a decade on, the benefits are yet to arrive.

The open question is therefore not whether merging improves rainfall estimation, which is well established (Ochoa-Rodríguez et al., 2019), but whether the improvement survives transmission. This paper traces a merged product through a complete chain on a sensor-instrumented sewer network: hydraulic simulation, machine learning emulation, and extrapolation nowcasting validated against both rain gauges and simulated overflow. Two features make the test demanding. The delivered radar product is already gauge-corrected, so the merge competes against an operational correction rather than an uncorrected field. In this case study the rain gauge network comprises two to three

¹ <https://hmc.enpc.fr/portfolio-archive/raingain/>

instruments, below the sparsest configuration previously examined in this research line. Two results are of wider interest than the site-specific numbers: the advantage transmits through physical routing and into forecast skill but reverses at the overflow outfalls when the hydraulic model is replaced by a statistical emulator, and the residual error in forecast-driven overflow prediction is governed by timing rather than magnitude.

2. STUDY SITE AND DATA

The Bellinge catchment drains 172 hectares of predominantly residential development south-west of Odense, Denmark, through a combined sewer network. The dataset and its calibrated SWMM model are documented by Pedersen et al. (2021) and distributed openly, and the model is used here without structural modification. It discretises the catchment into 713 sub-catchments of median area 0.3 hectares, placing the network inside the regime where Berne et al. (2004) predict rainfall representation to matter most. Two of nine outfalls discharge combined sewer overflow to surface water. Nine monitored structures carry fifteen

channels of level, flow, pump status and gate position at 1-to-2-minute resolution over a decade, which is what permits hydraulic consequences to be compared against observation rather than inferred.

Two tipping-bucket gauges operated across the full record; a third operated only from 2017 to 2018, which constrains how many storms support genuine three-point interpolation. A local X-band radar provided estimates on a 58-cell grid from March 2012 to October 2020. The delivered product is not raw: radar estimates are compared against approximately ten regional gauges and a single multiplicative factor applied to every cell in the domain, recalculated every 5 to 30 minutes, so the correction is dynamic in time but uniform in space (Borup et al., 2016). A single gauge in one corner therefore corrects the radar everywhere equally, and the comparison in this paper is between two gauge-corrected products differing in whether the correction may vary from cell to cell. Forty storms spanning June 2012 to August 2020, with depths from 8.6 to 40.2 mm, were selected from the gauge 5425 event catalogue and screened against X-band data density; the same set is used at every stage

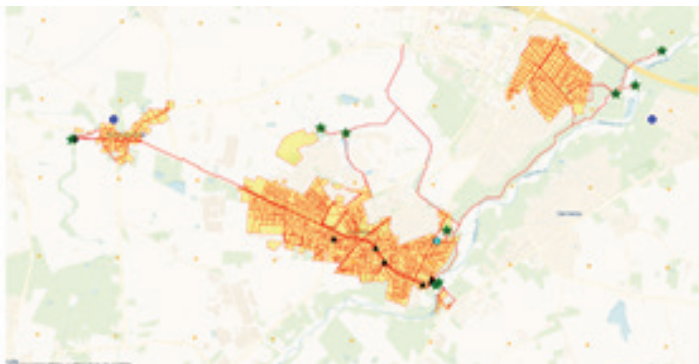


Fig.1. Bellinge catchment: sewer network, rain gauges, X-band radar grid, in-sewer sensors, and CSO outfalls (basemap: OpenStreetMap contributors, CARTO).

3. METHODS

3.1 SPATIALLY DISTRIBUTED BAYESIAN MERGE

The merge follows Todini (2001). Gauge observations are block-kriged onto the radar grid to give Z_{BK} , the radar field Z_{RD} is retained as an independent estimate, and the two are combined through a Kalman gain weighted by relative error variance, evaluated at every one-minute timestep.

$$Z_{BAY,t} = Z_{RD,t} + K (Z_{BK,t} - Z_{RD,t}) \quad (1)$$
$$K = V_{RD} / (V_{RD} + V_{RG}) \quad (2)$$

Two adaptations are forced by the gauge network. Fitting a three-parameter semi variogram requires observations across a range of separation distances: Ochoa-Rodríguez (2017) obtains these from gauge pairs and Wang (2012) from gauge covariance, but two to three gauges yield at most three pairs. The variogram is therefore derived from the radar field's own spatial structure following Haberlandt (2007), using the climatological radar-pixel-pair procedure of Ochoa-Rodríguez (2017). An exponential model fitted to 317,376 pooled pairs converged at a nugget of 0.18, a sill of 3.24 and a range of 12,297 m. Second, a period-specific rather than climatological gain is used, justified by the reliability decline reported in Section 4.1 and following the precedent for non-stationary bias correction of Seo and Breidenbach (2002).

The error variances entering the gain are estimated from each field's realised error against a withheld gauge observation, averaged within period, giving K

$= 0.903$ before mid-2016 and $K = 0.692$ thereafter. Performance is assessed by leave-one-gauge-out cross-validation, withholding each gauge in turn and comparing both reconstructed and radar estimates against the withheld observation at its nearest radar cell, giving 90 trials. The stratification is material: where only one gauge remains, kriging degenerates to a spatially uniform field, so only the 30 trials retaining two gauges test genuine spatial interpolation.

3.2 HYDRAULIC MODEL

Four storms serve as calibration events, three reproducing the set used by Pedersen et al. (2021) and a fourth added to cover the Brændekilde structures. Agreement with observed level is assessed jointly on Nash-Sutcliffe efficiency, percent volume error, percent peak error and peak timing error, with bands from Moriasi et al. (2007) and tolerances from CIWEM Urban Drainage Group (2021); reporting the four together is what distinguishes genuine failure from metric artefacts on low-variance and pump-cycled targets.

Both products were run across all 40 storms, giving 80 simulations. One constraint bounds every downstream result: the model reads rainfall from two gauges, and all 713 sub-catchments are assigned to one of the two, so the 58-cell merged field is sampled at two points before entering the hydraulic model. Any difference surviving that collapse is a lower bound on what fully distributed input might produce.

3.3 NOWCASTING AND HYDRAULIC VALIDATION

Forecasts are produced by Lagrangian extrapolation. Motion is estimated by Lucas–Kanade optical flow from a rolling window of three consecutive one-minute frames and advected semi-Lagrangian, both implemented through pysteps, an open-source Python framework for radar-based precipitation nowcasting that provides interchangeable motion-estimation and extrapolation modules (Pulkkinen et al., 2019), to fifteen one-minute lead times. Four constructions are evaluated on identical storms and issue times, differing only in the rainfall field supplied:

- Track A uses the delivered product;
- Track B the merged field;
- Hybrid takes motion from the delivered product but advects merged intensities;
- Blend averages the two forecasts.

A shared start frame is imposed across all four, without which the tracks forecast different parts of the same storm. Skill is assessed against gauge observation at the two radar cells containing gauges, as the median of per-storm mean absolute error resolved by lead time, and as the proportion of pairs within fixed tolerances.

Rainfall skill does not by itself establish operational value. A first validation design supplied only the fifteen minutes of forecast rainfall to SWMM as a complete storm input; it generated overflow in one storm of 39. Overflow on this network occurs between roughly 1,830 and 2,530 minutes after onset, because the sewer must fill through cumulative loading from the whole catchment before any

spill reaches an outfall, and fifteen minutes of rainfall into an empty network cannot supply that loading however accurate the forecast. The final design instead simulates the full storm duration using observed gauge rainfall throughout, substituting the merged forecast only for the fifteen minutes it covers, and runs a paired fully observed simulation as reference. The difference isolates how forecast error within one short window propagates into full-storm overflow (Leitão et al., 2010). Three storms representative of the structural groupings in the storm set were simulated over 32 to 43 hours each, giving six storm and outfall comparisons.

4. RESULTS

4.1 RADAR RELIABILITY DECLINE

Comparing radar against gauge-interpolated depth storm by storm produced an unanticipated result. Across the full population of matched observations, the ratio of radar depth to gauge depth declines systematically with date, correlating at $r = -0.848$ (Fig.2). Period-averaged ratios over a balanced sample of 20 storms either side fall from 0.926 before 2016 to 0.414 after, so the delivered

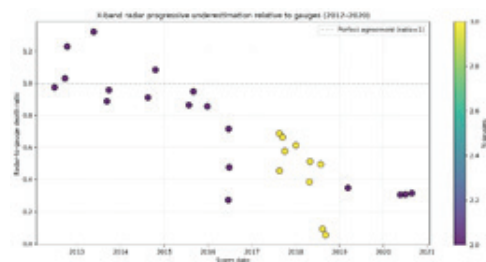


Fig.2. Radar-to-gauge depth ratio by storm date across the 40-storm working set, coloured by number of gauges reporting.

product agrees with the gauges to within about 8% in the early record and underestimates by roughly 60% in the late record. Two storms depart far enough to constitute a separate phenomenon, at ratios of 0.092 and 0.055 in August and September 2018, with the radar detecting rainfall normally on surrounding days.

The decline is not documented in the dataset’s accompanying publication, which reports continuous X-band availability with no flagged downtime but states that radar power was not checked for erroneous data (Pedersen et al., 2021). It is not an artefact of this analysis: the ratio is computed before any merging, holds across the whole matched record rather than only the selected storms, and persists despite the delivered product already being gauge-adjusted. Its cause cannot be determined without maintenance records and raw reflectivity, neither of which is distributed.

The finding matters beyond this study, because Bellinge is distributed as an open benchmark and has been used as such. Within this study it justifies the period-specific gain used at every subsequent stage, and it means roughly half the merge’s aggregate advantage compensates for a deteriorating record rather than resolving spatial variability.

4.2 MERGE ACCURACY

Leave-one-gauge-out cross-validation places the merged field closer to the withheld gauge than the delivered product in every stratum tested (Table 1). Overall, the merge wins 70% of 90 trials at 5.98 mm root mean square error against 16.11 mm. The strata differ in

what they demonstrate. In the 60 trials from two-gauge storms only one gauge remains after withholding, so kriging degenerates to a uniform field; the merge still wins 63% of these, which shows robustness rather than spatial skill. The 30 trials retaining two gauges test genuine interpolation, and there the win rate is 83.3% with the delivered product’s error slightly more than doubled, 9.22 mm against 4.47 mm.

Table 1. Leave-one-gauge-out cross-validation results, stratified by period and gauge availability.

Stratification	n	Z_BAY RMSE (mm)	Radar RMSE (mm)	Z_BAY win rate
Overall	90	5.98	16.11	70.0%
Pre-2016	40	5.41	21.33	60.0%
Post-2016	50	6.41	10.15	78.0%
2-gauge (degenerate)	60	6.61	18.62	63.3%
3-gauge (genuine kriging)	30	4.47	9.22	83.3%

Two qualifications travel with it. All ten three-gauge storms fall after 2016, so the genuine-interpolation subset is confounded with the worst period of the reliability decline. And on that same subset the block-kriged gauge field alone achieves 3.90 mm, slightly ahead of the merge, since a gain of 0.692 retains a radar contribution that is a liability where interpolation is well constrained. The kriged field is not a general alternative, because on the other thirty storms it reduces to a single gauge reading.

4.3 FORECAST SKILL AGAINST GAUGES

The merge advantage transfers to extrapolation nowcasting. Forecasts built on the merged field fall within 0.5 mm/h

of the gauge in 19.3% of pooled forecast-observation pairs, within 1.0 mm/h in 38.7% and within 2.0 mm/h in 52.2%, against 14.5%, 21.1% and 37.0% for the delivered product (Fig.3). The margin is widest at the intermediate tolerance, where the merged track is very nearly twice as accurate. Hybrid follows closely at 17.9%, 37.9% and 50.2%.

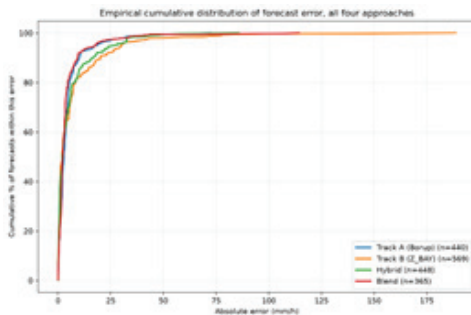


Fig.3. Empirical cumulative distribution of absolute forecast error for the four extrapolation constructions, pooled across storms and lead times.

Lead-time-resolved skill supports the same conclusion less cleanly. Taking the median of per-storm mean absolute error at each of the fifteen lead times, the merged track is best at seven, Blend at six and Hybrid at two, while the delivered product is best at none. The merged track is nonetheless worse than the delivered product in 38.0% of the 187 individual storm and lead-time pairs, and sample size falls from 36 storms at one minute to 11 at fifteen.

The merged field carries one weakness. Implied storm translation speeds exceed a physically implausible 150 km/h in 8 of 39 storms under the merged track against 1 of 39 under the delivered product, because the merge alters intensities

at gauge-influenced cells in ways that distort the optical flow solution. Taking motion from the delivered product while retaining merged intensities restores the failure rate to 1 of 39, confirming the instability is a property of the merged field's texture rather than of the extrapolation method.

4.4 PROPAGATION INTO OVERFLOW PREDICTION

Substituting the fifteen-minute merged forecast into an otherwise fully observed full-duration simulation, and comparing against a paired fully observed run, gives a median peak overflow flow error of 3.6% and a median volume error of 5.9% across the six storm and outfall comparisons. Individual peak errors range from 0.01% to 24.7%, the weakest being 17 August 2014 at G71U10F, so the median describes typical rather than bounded performance.

One limitation should be stated plainly. The comparison is between a merge-driven forecast and a fully observed reference, not between merge-driven and delivered-product-driven forecasts. It establishes that the merged forecast reproduces overflow behaviour closely, and that forecast error within a fifteen-minute window propagates to single-digit median error at the outfall, but not that the merged input outperforms the delivered input at this stage.

4.5 TIMING AS THE RESIDUAL MECHANISM

Hydrograph inspection identifies the mechanism behind the spread, and it is not what the aggregate figures imply.

On 17 August 2014 the forecast-driven hydrograph at G71U10F peaks approximately eight minutes earlier than the observed one, consistent with the extrapolation advecting the rainfall field forward (Fig.4). Whether that offset produces a large or negligible error depends on the steepness of the true hydrograph's rising limb where it lands. On that storm the limb rises from near zero to over $1.0 \text{ m}^3/\text{s}$ within roughly 20 to 25 minutes, so a modest timing error converts into a large flow discrepancy despite comparable peak magnitude. On 17 June 2016 an offset of similar order intersected a gentler limb and produced an error of 0.5%.

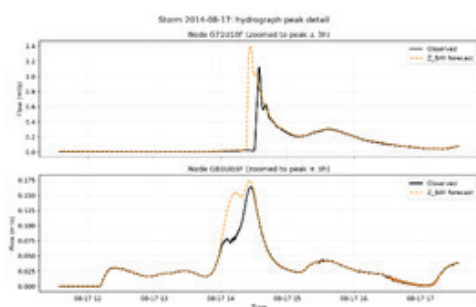


Fig.4. Observed and forecast-driven overflow hydrographs at G71U10F, 17 August 2014, showing the forecast timing lead against the rising limb

The implication is directional. If error is dominated by timing rather than magnitude, effort spent improving motion-field precision will yield more than effort spent improving intensity accuracy, which is the opposite of where merging contributes. It also explains why Hybrid is more valuable than its accuracy statistics suggest, since its purpose is stabilising the motion field. The eight-minute figure is a single-storm observation and should not be treated as a calibrated constant.

5. DISCUSSION AND CONCLUSIONS

Merge quality transmits through the chain at three of the four interfaces tested. It survives hydraulic routing, where driving the calibrated model with each product across all 40 storms gives a median 34.5% higher overflow volume under the merged input, with the advantage materialising specifically when the delivered product fails; given that Section 4.1 shows such failure becoming more common across the record, that is operationally meaningful rather than a null result. It survives into extrapolation nowcasting and into forecast-driven overflow prediction. It does not transmit into statistical emulation. An antecedent-depth LSTM following Palmitessa et al. (2021), retrained on delivered-product simulations under otherwise identical conditions, favours the merged input at 11 of 12 junction nodes but reverses at both overflow outfalls in all four, reproducing across three independent training runs; the mechanism supported by positive evidence is target variance, since merge-driven outfall series carry higher standard deviation than delivered-driven ones. A more accurate rainfall field produces a harder-to-learn signal at those two points. Rainfall to hydraulics is deterministic physical routing; hydraulics to emulation is statistical learning, where properties of the signal other than its accuracy govern how faithfully it is reproduced. The uncertainty accounting reinforces this, with emulation contributing 66.0% against 36.9% for the merge and 13.1% for the hydraulic model.

Three limitations bound these conclusions. The merged field is sampled at two gauge locations before entering the

hydraulic model, so every result is a lower bound on what fully distributed input might achieve; the hydraulic forecast validation rests on three storms; and the genuine-interpolation subset is confounded with the period of steepest radar decline, so spatial value and drift

correction cannot be separated cleanly. The most tractable next step is to refit the mass-balance emulator on delivered-product simulations (Mahmoodian et al., 2018), which would distinguish a property of the data from a capacity limitation of the learner.

REFERENCES

- Berne, A., Delrieu, G., Creutin, J.-D. & Obled, C. (2004) Temporal and spatial resolution of rainfall measurements required for urban hydrology. *Journal of Hydrology*. 299 (3–4), 166–179.
- Borup, M., Grum, M., Linde, J.J. & Mikelsen, P.S. (2016) Dynamic gauge adjustment of high-resolution X-band radar data for convective rain storms: model-based evaluation against measured combined sewer overflow. *Journal of Hydrology*. 539, 687–699.
- CIWEM Urban Drainage Group. (2021) Integrated Urban Drainage Modeling Guide. Version 2.01. Allitt, R. & Pearson, K. (eds.) London, Chartered Institution of Water and Environmental Management.
- Gires, A., Schertzer, D., Tchiguirinskaia, I., Lovejoy, S., Onof, C., Maksimović, Č. & Simões, N. (2012) Impact of small-scale rainfall uncertainty on urban discharge forecasts. In: Moore, R.J., Cole, S.J. & Illingworth, A.J. (eds.) *Weather Radar and Hydrology*. IAHS Publ. 351, 400–406.
- Haberlandt, U. (2007) Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. *Journal of Hydrology*. 332 (1–2), 144–157.
- Leitão, J.P., Simões, N.E., Maksimović, Č., Ferreira, F., Prodanović, D., Matos, J.S. & Sá Marques, A. (2010) Real-time forecasting urban drainage models: full or simplified networks? *Water Science and Technology*. 62 (9), 2106–2114.
- Mahmoodian, M., Carbajal, J.P., Bellos, V., Leopold, U., Schutz, G. & Clemens, F. (2018) A hybrid surrogate modeling strategy for simplification of detailed urban drainage simulators. *Water Resources Management*. 32 (15), 5241–5256.
- Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D. & Veith, T.L. (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*. 50 (3), 885–900.
- Ochoa-Rodríguez, S. (2017) Rainfall estimates for urban drainage modelling: an investigation into resolution requirements and radar-rain gauge data merging at the required resolutions. PhD thesis. Imperial College London.
- Ochoa-Rodríguez, S., Wang, L.-P., Pfeifer, M., Li, X., Onof, C., Maksimović, Č., Pool, M. & Didszun, J. (2014) X-band radar monitoring campaign

- in Central London: an experiment to assess and improve the applicability of a low-cost portable X-band radar to urban hydrology. In: 2014 International Weather Radar and Hydrology Symposium, 7–10 April 2014, Washington D.C., US.
- Ochoa-Rodríguez, S., Wang, L.-P., Gires, A., Pina, R.D., Reinoso-Rondinel, R., Bruni, G., Ichiba, A., Gaitan, S., Cristiano, E., van Assel, J., Kroll, S., Murlà-Tuyls, D., Tisserand, B., Schertzer, D., Tchiguirinskaia, I., Onof, C., Willems, P. & ten Veldhuis, M.-C. (2015) Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: a multi-catchment investigation. *Journal of Hydrology*. 531, 389–407.
- Ochoa-Rodríguez, S., Wang, L.-P., Willems, P. & Onof, C. (2019) A review of radar-rain gauge data merging methods and their potential for urban hydrological applications. *Water Resources Research*. 55 (8), 6356–6391.
- Palmitessa, R., Mikkelsen, P.S., Borup, M. & Law, A.W.K. (2021) Soft sensing of water depth in combined sewers using LSTM neural networks with missing observations. *Journal of Hydro-environment Research*. 38, 106–116.
- Pedersen, A.N., Wied Pedersen, J., Viguera-Rodríguez, A., Brink-Kjær, A., Borup, M. & Mikkelsen, P.S. (2021) The Belling data set: open data and models for community-wide urban drainage systems research. *Earth System Science Data*. 13 (10), 4779–4798.
- Pulkkinen, S., Nerini, D., Pérez Hortal, A.A., Velasco-Forero, C., Seed, A., Germann, U. & Foresti, L. (2019) Py-steps: an open-source Python library for probabilistic precipitation nowcasting (v1.0). *Geoscientific Model Development*. 12 (10), 4185–4219.
- Seo, D.-J. & Breidenbach, J.P. (2002) Real-time correction of spatially nonuniform bias in radar rainfall data using rain gauge measurements. *Journal of Hydrometeorology*. 3 (2), 93–111.
- Todini, E. (2001) A Bayesian technique for conditioning radar precipitation estimates to rain-gauge measurements. *Hydrology and Earth System Sciences*. 5 (2), 187–199.
- Wang, L.-P. (2012) Improved rainfall downscaling for real-time urban pluvial flood forecasting. PhD thesis. Imperial College London.
- Wang, L.-P., Ochoa-Rodríguez, S., Simões, N., Onof, C. & Maksimović, Č. (2013) Radar-raingauge data combination techniques: a revision and analysis of their suitability for urban hydrology. *Water Science and Technology*. 68 (4), 737–747.

Combining machine learning and physically based models in the forecasting of extreme weather events

Ivan Ristic, Ivana Kordic

Weather2, Belgrade, Serbia (ivanr@weather2.rs)

Weather2, Belgrade, Serbia (ivana.kordic@weather2.rs)

ABSTRACT

In the last few years data driven models trained on reanalysis data, such as GraphCast, Pangu-Weather, FourCastNet and the Artificial Intelligence Forecasting System (AIFS) of ECMWF, have become a part of everyday forecasting practice. Once such a model is trained, a global forecast is produced in minutes on a single graphics card instead of hours on a supercomputer, and in the medium range the scores for the large scale flow are comparable with those of the physically based global models, and better for a number of fields. AIFS has been operational at ECMWF since February 2025, the ensemble version since July 2025, and its output is freely available as open data.

These models, however, are not a replacement for the standard approach. They are distributed on a coarse grid (about 0.25 degrees) and few pressure levels, they do not contain hydrometeors and other fields that limited area models need, their fields are smoother, and extremes tend to be damped, because the training minimises an error norm. What they add is speed, low cost and a good description of the large scale flow, exactly what is needed at the lateral boundaries of a high resolution model in which deep convection is explicit.

The machine learning data are used here

in the IRIE model, a further development of the Eta model. In the IRIE AIFS configuration it is driven by two sources at the same time: AIFS and ECMWF HRES fields are used together as initial and lateral boundary conditions. HRES provides the higher horizontal and vertical resolution and the variables that the machine learning model does not produce, while AIFS contributes its own, independent description of the large scale flow. Their preparation requires vertical interpolation to the model coordinate, consistency of the mass and wind fields along the boundaries, and generation of cloud water, cloud ice and snow by the cloud scheme of the model itself.

A second configuration uses AIFS ENS together with HRES data as initial and lateral boundary conditions. In this way the ensemble supplies the range of possible synoptic scenarios, while the intensity and the structure of the convection are produced by IRIE itself. Results of this configuration will be presented as well.

The WRF model is used in a different way, without machine learning data. It is run in a time shifted configuration, in which successive initialisations are combined into a single continuous forecast series, with the Ristic microphysics and convection schemes. It has already

shown good results for extreme weather events and serves here as the physically based counterpart of the machine learning driven runs.

Both models are integrated into domains covering the Central Mediterranean and the Balkans. The selected events include supercell storms with large hail and damaging wind over Italy and North Macedonia, and systems that produced flash floods in the region. The forecasts are verified against radar, satellite and station data. The timing and position of the convective systems, and the ability of the models to produce a supercell structure, are the main criteria.

Preliminary results indicate that the combination of AIFS (AIFS ENS) and HRES fields gives a realistic large-scale environment for short range forecasts of extreme weather in IRIE, while the smoother gradients of the machine learning fields influence the fine structure of the convection. Together with the WRF results, they suggest that machine learning and physically based models are best used as complementary tools. Validation of these test cases and the advantages and limitations of each approach will be presented.

