

928

Verification of global and regional NWP models over South America

Ramón de Elía¹, Thomas Haiden², Cynthia Matsudo¹, Federico Otero³, Estíbaliz Gascón², Lucía Castro⁴, Linus Magnusson², Alejandro Godoy¹, Caio Coelho⁵, Hernán Bechis^{7,8}, Barbara Casati⁶, Daniela D'Amen¹, Ariane Frasson⁹, Santiago Galgano¹, Marisol Osman^{7,8,10}, Soledad Osoreo¹, Silvina Righetti^{1,7}, Paola Salio^{7,8,10}, Marcos Saucedo¹, Yanina García Skabar¹, Pablo Spennemann^{1,9}, José Luis Stella¹, Nils Wedi²

¹ Servicio Meteorológico Nacional, Buenos Aires, Argentina

² European Centre for Medium Range Forecasting, Reading, UK

³ IANIGLA, Mendoza, Argentina

⁴ Formerly at Servicio Meteorológico Nacional, Buenos Aires, Argentina

⁵ CPTEC, INPE, São Paulo, Brazil

⁶ Environment and Climate Change Canada, Montreal, Canada

⁷ Universidad de Buenos Aires, DCAO, Buenos Aires, Argentina

⁸ CONICET – UBA, CIMA, Buenos Aires, Argentina

⁹ Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

¹⁰ CNRS – IRD – CONICET – UBA. IFAECI

June 2025

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:

<http://www.ecmwf.int/publications/>

Contact: library@ecmwf.int

© Copyright 2025

European Centre for Medium Range Weather Forecasts
Shinfield Park, Reading, Berkshire RG2 9AX, England

Literary and scientific copyrights belong to ECMWF and are reserved in all countries. The content of this document is available for use under a Creative Commons Attribution 4.0 International Public License. See the terms at <https://creativecommons.org/licenses/by/4.0/>.

The information within this publication is given in good faith and considered to be true, but ECMWF accepts no liability for error, omission and for loss or damage arising from its use.

Abstract

South America has received less attention from the global numerical weather prediction community than many other regions. This is reflected, for example, in its absence as a region in the standardised WMO verification score exchange. As a result, global modelling centres may have missed out on interesting scientific challenges, and potential insights, in a unique testing ground in one of the few land masses in the Southern Hemisphere. To address this situation, a pilot project was launched in early 2024 by the Working Group on Numerical Experimentation (WGNE) and the Joint Working Group on Forecast Verification Research (JWGFVR), partly in response to the requirements of the United Nations Early Warnings for All initiative (EW4All). This pilot project, led by the Argentine National Meteorological Service (SMN) and the European Centre for Medium Range Forecasting (ECMWF), aims to explore the performance of global and regional forecasts over South America and their potential as early warning tools. For the sake of simplicity, we have focused mainly on the territory of Argentina and on deterministic operational forecasts. This Technical Memorandum presents the first results of the pilot project, focusing on two perspectives: a general statistical verification of the ECMWF Integrated Forecasting System (IFS) control operational forecast, and an in-depth evaluation of model performance in three individual High Impact Weather (HIW) events which, due to their characteristics, are both of societal relevance and posing a challenge to models. Within the first perspective, temporally and/or spatially aggregated summary verification statistics are produced for upper air variables by comparing forecasts with their own analysis, and for surface variables by comparing surface forecast fields with SYNOP station data. In addition, the IFS deterministic operational forecast is compared with other global models and with a limited area model over the South America region. Within the second perspective, three individual case studies covering HIW events with very different characteristics are analysed: a convective storm causing severe wind damage and loss of life, a strong Zonda wind (Foehn) on the lee side of the Andes, and a prolonged heat wave.

Plain Language Summary

In the verification of global weather forecast models, South America has traditionally received less attention than other continents. As a result, information about forecasting skill in the region has been limited, and model developers may have missed out on potential insights provided by atmospheric phenomena in the region. This report summarizes results of an initiative started in 2024 by the Working Group on Numerical Experimentation (WGNE) and the Joint Working Group on Forecast Verification Research (JWGFVR) and led by the Argentine National Meteorological Service (SMN) and the European Centre for Medium Range Forecasting (ECMWF) which attempts to close this gap. In the first part of the report, statistical verification results for South America are shown for verification of the large-scale flow against the model's own analysis and near-surface parameters against SYNOP observations. While the main focus is the performance of the ECMWF model, some model intercomparisons with other global and regional models are presented. In the second part of the report, three case studies of high-impact weather in Argentina are analysed: a severe thunderstorm event, a strong Zonda wind (Foehn) event, and a prolonged heat wave.

1 Introduction

Over the past two decades, several South American countries have developed significant numerical weather prediction capabilities, with a particular emphasis on high-resolution limited-area models. According to official information from the World Meteorological Organization (WMO) (see [WIPPS dashboard](#), selecting “NWP models” and “Region III”) eight countries operate limited area models over South America, two of which are ensembles. This focus on limited area models aims to better predict weather hazards by complementing existing forecasts from global modelling centres.

Ongoing regional modelling activities –both operational and quasi-operational in weather services, institutes and universities– are supported by verification studies, typically comparing regional model output with the driving global model, verifying against surface observations (since surface fields are usually the most strongly impacted by enhanced resolution). These studies help to establish the added value of the regional simulation over the global model, which is generally found in the prediction of phenomena affected by small-scale convection and steeper topography where convection-permitting high-resolution models have an advantage.

Despite continued local progress in modelling expertise and operational use of limited area models, South America has received less systematic attention from the global numerical weather prediction community, and this is reflected in its absence as a region for verification in the standardised WMO verification score exchange (see Figure 1).

This information gap, which has historical and technical reasons, has had a negative impact on local forecasters' operational activities, which have had to rely more strongly than in other regions on subjective assessments of the strengths and weaknesses of individual global models. At the same time, by skipping South America, global modelling centres are missing an interesting testing ground in one of the few land masses in the Southern Hemisphere. The Andes mountain range, due to its scale and orientation, significantly affects the large-scale dynamics and gives rise to regional phenomena such as the Zonda wind.

To address this situation, a pilot project was launched in early 2024 by the Working Group on Numerical Experimentation (WGNE) and the Joint Working Group on Forecast Verification Research (JWGFVR), partly in response to the requirements of the Early Warning for All (EW4All) initiative (WMO 2022). This pilot project, led by the Argentine National Meteorological Service (SMN) and European Centre for Medium Range Forecasting (ECMWF), aims to explore the performance of global and regional forecasts in South America and their potential as early warning tools. For reasons of simplicity, we have focused mainly on the Argentinean territory and deterministic operational forecasts.

Argentina is characterized by a variety of climates due to its latitudinal range (from subtropical to sub-Antarctic), its long coastline with the Atlantic on the east, and the presence of the Andes mountain range on the west, second only in height to the Himalayas. The northeastern climate is warm and humid, and prone to severe convection, making it one of the thunderstorm hotspots of the world (see Zipster et al. 2006). Heat waves are among the most challenging hazards for the population particularly in the big cities and they are increasing in frequency (see for example Suli et al. 2023). Weather forecasting remains a challenge due the sparse surface and upper air

observations network and the isolation of the southern landmass of South America between two vast oceans.

This Technical Memorandum presents first results of this pilot project, focusing on two perspectives: a general statistical verification of the ECMWF control operational forecast, and a more detailed evaluation of model performance in three individual High Impact Weather (HIW) events which, due to their characteristics, are both of societal relevance and an NWP modelling challenge. Section 2 presents the models, data, and domains used in the evaluation. Section 3 presents the aggregated summary verification statistics, computed for upper air variables by comparing forecasts with their own analysis, and for the surface by comparing surface forecast fields with SYNOP data. In addition, a comparison is shown between the ECMWF deterministic operational forecast and a limited-area model over the South America region.

Section 4 presents the evaluation of three case studies with very different characteristics: a convective storm that caused severe damage and loss of life, a strong Zonda wind (Foehn) at the lee side of the Andes, and a prolonged heat wave. Section 5 concludes by highlighting the importance of this first step of the project both for the region and for the global forecasting community.

2 Domains, models, observations and methodology

Evaluations carried out in this study involved several models, global as well as regional, different domains over South America, and sets of observations.

2.1 Evaluation domains

Despite the absence of South America as an official WMO domain for verification (Figure 1), ECMWF has been verifying forecasts over the region for internal use for 20 years (Figure 2). This long-standing verification domain divides the South America region into two sub-domains, leaving uncovered the north-eastern part of the continent. At this stage of the pilot project it was decided to focus mostly on the southern sub-domain when presenting domain-average verification results and case studies.

In parallel with the activities discussed here, the WMO Task Team on Reviewing NWP Standardized Verification (TT-NWPSV, <https://community.wmo.int/en/meetings/tt-nwpsv-17>) proposed to extend the coverage of global model verification and score exchange to new domains in Africa and South America. Figure 2b shows the proposed domains for the South American region, separating the predominantly tropical from the extratropical. Part of the results shown here focus on the southern Extratropical South America domain. The set of domains of interest in this text is summarized in Table 1.

Table 1: List of different domains for which verification results of global models are discussed in this report.

Domain name	Defined by	Shown in
Southern South America	ECMWF	Figure 2 (left)
Tropical South America	TT-NWPSV	Figure 2 (right)
Extratropical South America	TT-NWPSV	Figure 2 (right)
Southern Hemisphere	WMO	Figure 1 (left)
Tropics	WMO	Figure 1 (left)

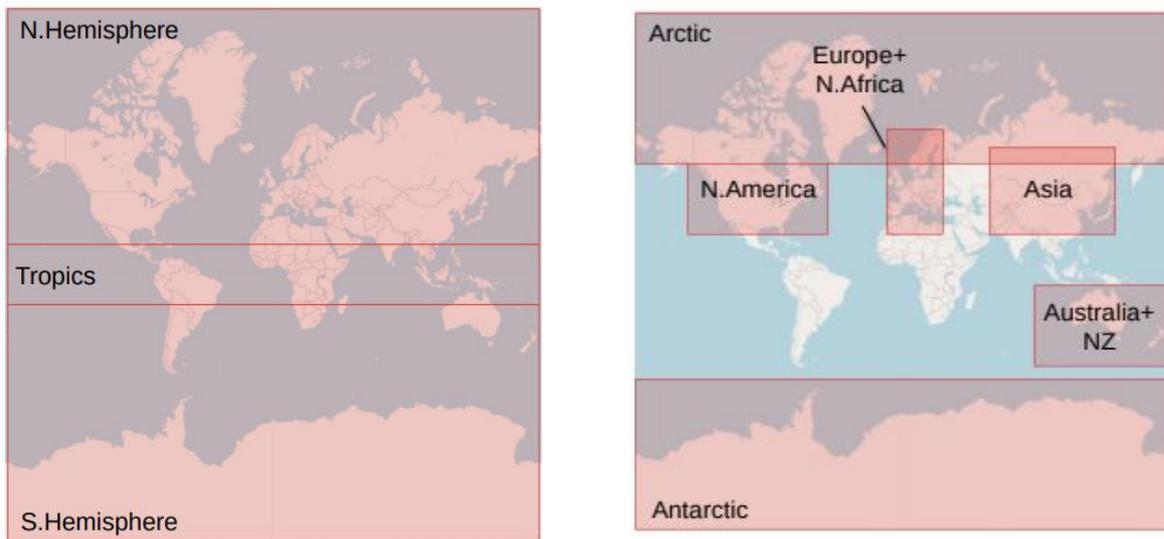


Figure 1: Areas where global deterministic models are verified following a standard procedure defined by the Manual on the WMO Integrated Processing and Prediction System. These areas are still valid at the time of this study (see WMO 2023, Appendix 2.2.34).

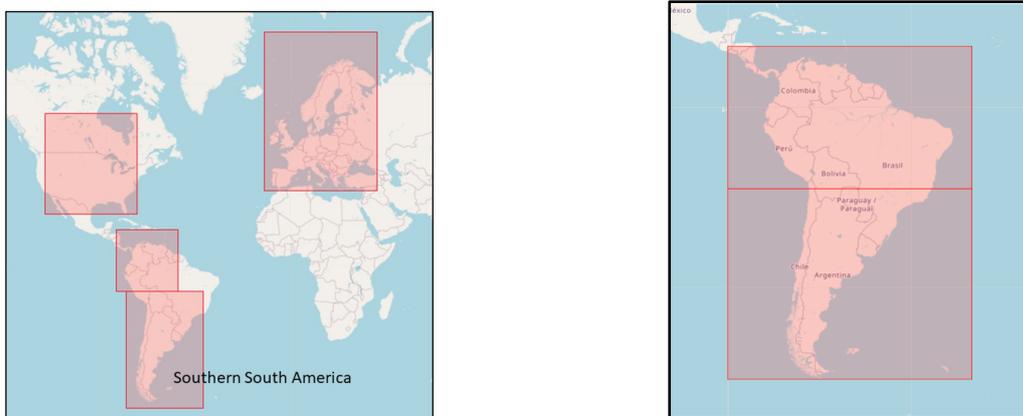


Figure 2: Left, some of the domains within which the ECMWF performs regular forecast verification. Right, new domains proposed for South America by the WMO’s Task Team on Reviewing NWP Standardized Verification (TT-NWPSV, <https://community.wmo.int/en/meetings/tt-nwpsv-17>).

2.2 Models and observations

The main model evaluated in this work is the ECMWF Integrated Forecasting System (IFS). To provide a general perspective on IFS performance, results from other World Meteorological Centres' global models are presented as well. For the summary verification statistics, results are shown comparing operational forecasts and analyses provided to ECMWF from the following global modelling centers: the Canadian Meteorological Center (CMC), the National Centers for Environmental Prediction (NCEP), and the UK Met Office (UKMO).

For two of the case studies, comparison to higher resolution models using two different versions of the WRF are used, as well as km-scale forecasts run with higher-resolution IFS configurations used as testing ground for the Weather-induced Extremes Digital Twin in the European Commission's Destination Earth (DestinE) initiative (see Wedi et al, 2025).

2.2.1 Global forecasts

The IFS is run in different configurations, and here the 15-day high-resolution control forecast (CTRL, ex-HRES), the 15-day ensemble forecast (ENS), which are now both at about 9 km resolution, and the 45-day sub-seasonal forecast at about 36 km resolution are used (see ECMWF, 2022). The version of CTRL used for the case studies is Cycle 48r1, and the verification results are based on archived data produced by the operational versions at the time. In addition, experimental versions of IFS for km-scale resolution forecasts were also used for a case study. The specifics of each forecast are discussed in the case study section.

The operational forecasts from other global modelling centers are the following: the Global Deterministic Prediction System (GDPS) from the Canadian Meteorological Centre (CMC-GEM; see McTaggart-Cowan et al. 2019); the Global Forecast System (GFS) from the National Centers for Environmental Prediction (NCEP, see NCEP 2015); and the Unified Model (UM) from the UK Met Office (UKMO, see Walters et al. 2019).

The 16th World Meteorological Organisation Congress (2011) designated ECMWF as the Lead Centre for Deterministic NWP Verification (LC-DNV). In order to provide consistent verification information the WMO Commission on Basic Systems (now INFCOM) has defined detailed procedures for the production and exchange of a standard set of verification scores for deterministic NWP forecasts produced by WMO centres (see WMO 2023). The scores are exchanged between the participating centres through the LC-DNV, whose role is to facilitate this standardised verification, to ensure the routine exchange of the required verification results between centres and to provide consistent comparisons of these results (for more information see <https://confluence.ecmwf.int/display/WLD>). For upper-air forecasts, results are computed on a standard $1.5^{\circ} \times 1.5^{\circ}$ grid. In this work only verification against each centre's own analysis is carried out. For surface fields, after a basic quality control, verification is performed both against own analysis and against SYNOP data exchanged through the Global Telecommunication System (now WMO Information System 2.0). For 2-metre temperature, standard corrections are applied for differences between model and true orography using a constant lapse rate assumption provided the magnitude of the correction is less than 4°C (see section A.3 of Haiden et al. 2021).

2.2.2 Regional forecast system from the SMN

The operational forecasting system from the National Meteorological Service of Argentina (SAP.SMN for its acronym in Spanish) has deterministic and ensemble components (SAP.SMN-DET and SAP.SMN-ENS respectively). It uses the Advanced Research WRF (ARW) version 4.0 dynamic core WRF model developed by the National Center for Atmospheric Research (NCAR, Skamarock et al., 2019). The WRF is a compressible, non-hydrostatic mesoscale model that allows a variable integration time step to explicitly solve the microphysics. The domain covers all of Argentina with a Lambert conformal projection with a horizontal resolution of 4 km, with 45 vertical levels (top 10hPa). Four initializations are performed per day at 00, 06, 12 and 18 UTC and outputs are generated hourly up to 72 hours for the deterministic scheme and up to 48 hours for the ensemble. The initial and boundary conditions (CIs/CBs) are taken from the Global Forecast System (GFS, see NCEP 2015) analyses and forecasts produced by the National Center for Environmental Prediction (NCEP). For more detail see (Matsudo and Garcia Skabar, 2023). Only results for SAP.SMN-DET are presented here.

2.2.3 High-resolution hindcast from the IANIGLA

The Instituto Argentino de Nivología, Glaciología y Ciencias Ambientales (IANIGLA) runs WRF forecasts exclusively to study past events. In this case, the model was used with 60 vertical levels and three successive one-way nestings with horizontal resolutions of 18 km, 6 km, and 2 km. The final domain covers the Cuyo region of Argentina, centered on the city of Mendoza (where one of the HIW events is analyzed, see Section 4.2). The model is initialized for spin-up purposes 12 hours before the start of the event (21st of July 2023 at 06 UTC). The boundary conditions are the ECMWF ERA5 reanalysis (Hersbach et al., 2020) renewed every six hours with a horizontal resolution of 0.25° . The boundary layer parameterization is the Mellor–Yamada–Nakanishi–Niino (MYNN)-Eddy Diffusivity-Mass Flux (EDMF) scheme (Nakanishi and Niino 2009, Olson et al. 2019), and the microphysics parameterization is the 1-moment WSM6 scheme (Hong and Lim, 2006).

3 General verification results

3.1 Verification of global models against their own analysis

This section focuses on the verification of the CTRL forecast (ex-HRES) for upper air variables in the southern South America domain (see Figure 2a). To put the verification in context, the results are compared with those obtained in other ECMWF domains (see Figure 1) and with the performance of other global models over South America (CMC, NCEP, UKMO).

3.1.1 850 hPa Temperature

Figure 3 shows the anomaly correlation of the forecast temperature at 850 hPa during 2024 for four domains: Europe, North America, the Southern Hemisphere and southern South America (Figure 2, left panel). During the first two days, the correlation in the two southern domains appears to decrease more rapidly compared to the two northern domains. However, later the South America curve converges to similar values as North America. Given the poor observation

network over the Southern Pacific upstream South America, it is curious that this domain retains the highest level of correlation at 10 days. This position in the ranking, however, has considerable interannual variability as will be shown in Figure 5.

Generally, the anomaly correlation of upper-air variables decreases towards the equator in the short range and increases towards the equator in the later medium range. Therefore, the latitudinal range covered by a given domain determines to some extent the shape of curves such as shown in Figure 3, with the domain of South America being closer to the equator than the one of North America

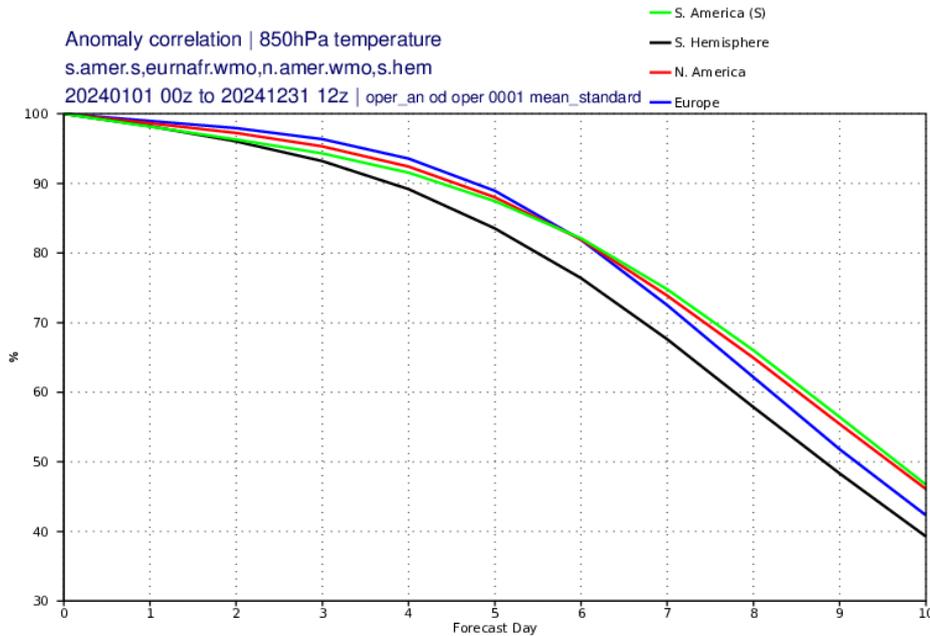


Figure 3: Anomaly correlation for 850 hPa temperature CTRL (ex-HRES) forecasts for the year 2024 in different domains. The North American, European, and Southern Hemisphere domains are shown in Figure 1, while the southern South America domain is shown in Figure 2 (left panel).

Figure 4 shows the Root Mean Square Error (RMSE) of the forecast for the 850 hPa temperature during 2024. After day five, and consistent with Figure 3, the domain over South America shows the smallest RMSE. Both figures suggest that the overall error of the CTRL (ex-HRES) upper-air forecast in the southern South America domain is comparable to, or even smaller, than that of other extratropical domains, and smaller than the average error over the southern extratropics as a whole. It should be noted that the RMSE of 850 hPa temperature scales with its variance, which generally increases towards higher latitudes and exhibits longitudinal variations due to land-sea contrasts and topography. For example, the higher RMSE in North America compared to Europe is due to more extreme temperature contrasts and associated cold-air outbreaks and warm air intrusions over the North American continent. South America has an RMSE which is similar to North America in the short range, but subsequent error growth in the medium range is considerably smaller, and even below that of Europe.

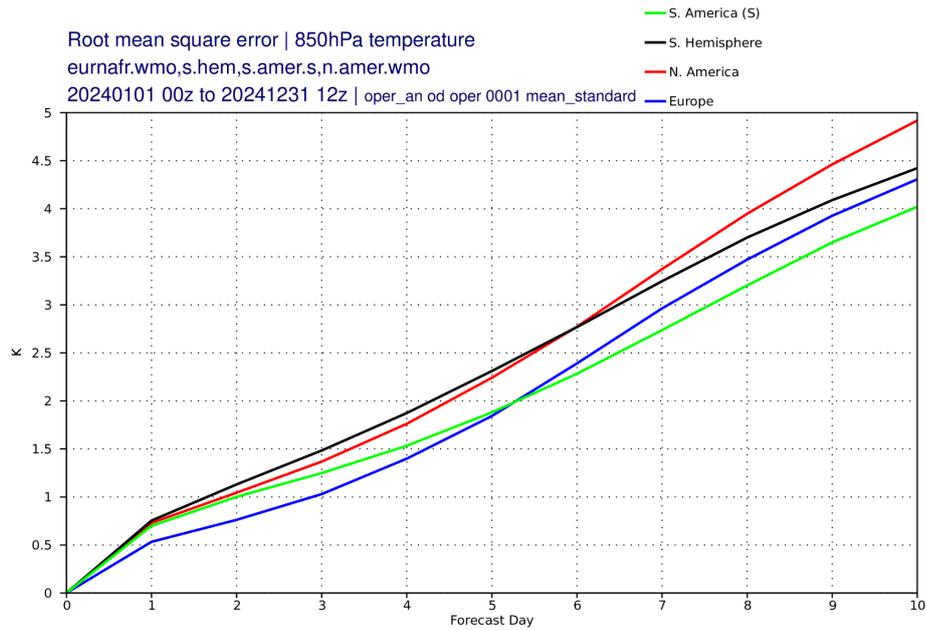


Figure 4: RMSE for 850-hPa temperature CTRL (ex-HRES) forecasts for the year 2024 in different domains. The North American, European, and Southern Hemisphere domains are shown in Figure 1, while the southern South America domain is shown in Figure 2 (left panel).

Figures 5 and 6 show the evolution of the CTRL (ex-HRES) day-5 forecast anomaly correlation and RMSE (presented in Figures 3 and 4) over the last 20 years. The southern South America domain reveals improvements comparable to those of the other domains. The results show strong interannual variations, and the three regional domains exchange ranking positions over time. The increase over the last 20 years of about 10% in units of anomaly correlation at day 5 is equivalent to a lead time gain of about two days (cf. Figure 3).

The RMSE in Figure 6 allows for a similar interpretation of the degree of improvement, but here the lower RMSE is undeniably the southern South America domain. Again, this can be related to the reduced variability of temperature at 850 hPa towards lower latitudes. In terms of the relative reduction of RMSE over time, the different domains behave similarly. Over the last 10 years the South America domain does not show a net improvement in terms of RMSE (it did show improvement in terms of anomaly correlation, see Figure 5), whereas the other three domains do. This appears to be related to an increasing warm bias at 850 hPa over parts of the South American continent (similar to an increasing warm bias seen in Central Africa) and is currently being investigated at ECMWF.

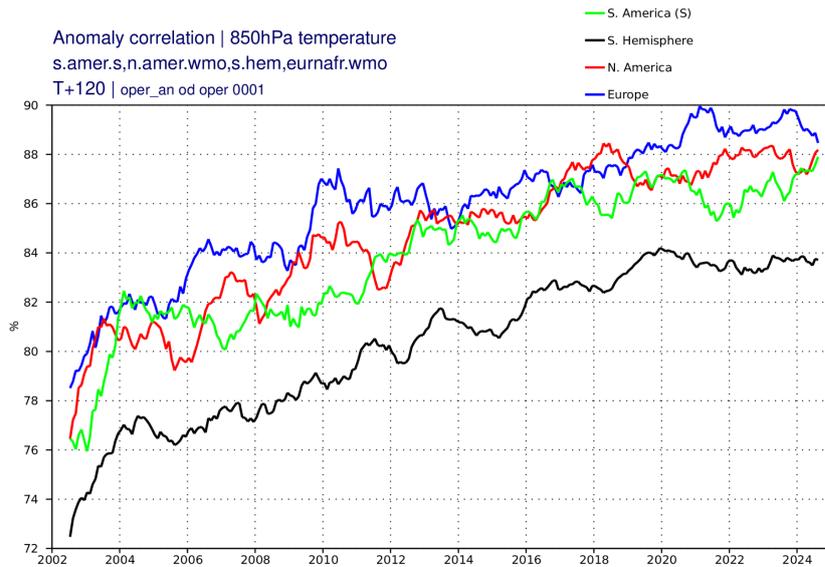


Figure 5: Evolution of the CTRL (ex-HRES) 850-hPa temperature day-5 forecast anomaly correlation (presented in Figure 2) over the last 20 years for the North American, European, and Southern Hemisphere domains (shown in Figure 1) and the southern South America domain (shown in Figure 2, left panel). Values are 12-month running averages.

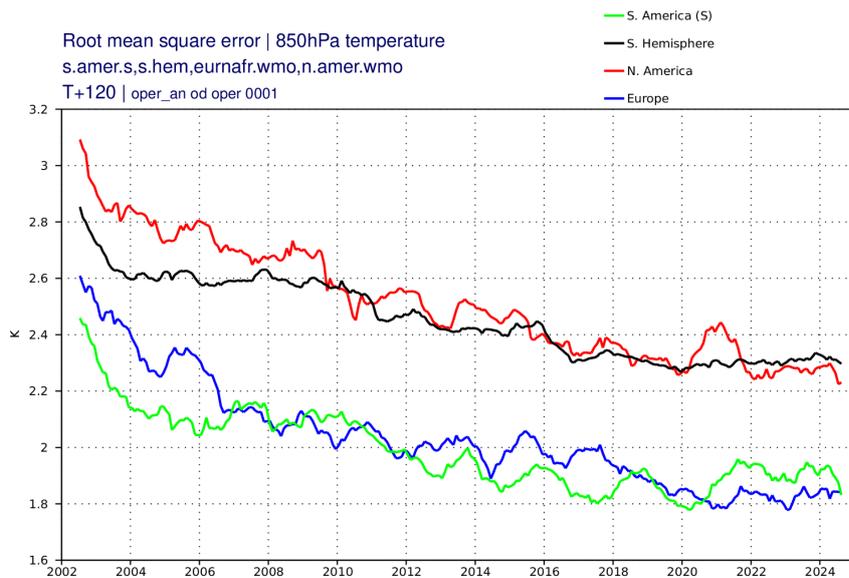


Figure 6: Evolution of the CTRL (ex-HRES) 850-hPa temperature day-5 forecast RMSE (presented in Figure 4) over the last 20 years for the North American, European, and Southern Hemisphere domains (shown in Figure 1) and the southern South America domain (shown in Figure 2, left panel). Values are 12-month running averages.

Figures 7 and 8 compare the verification results for 2024 of the CTRL (ex-HRES) over the southern South America domain with those of other global centers —the Canadian Meteorological Centre (CMC), the National Centers for Environmental Prediction (NCEP), the UK Met Office (UKMO)— for different forecast lead times. Figure 7 shows the 850 hPa temperature anomaly correlation between forecasts and observations for different forecast horizons, while Figure 8 presents the RMSE. In both cases it can be seen that the global models perform rather similarly. The difference between some of the models is more pronounced for

the RMSE (Figure 8), where the ECMWF forecast leads other forecasts by an amount equivalent to 0.5-1 forecast days, hinting at smaller biases of the ECMWF forecast in this domain.

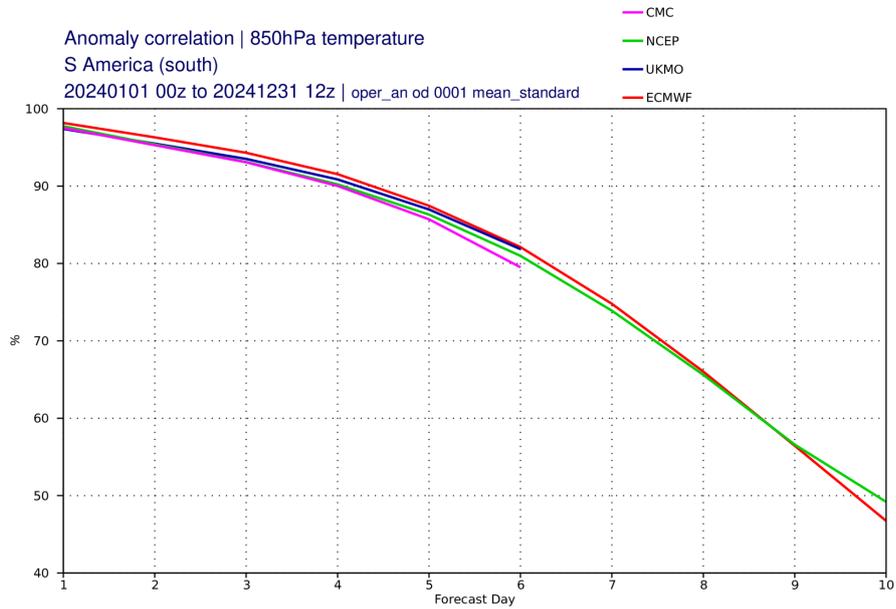


Figure 7: Anomaly correlation for 850-hPa temperature forecasts for the year 2024 in the southern South America domain for forecasts of four global forecast centers: the Canadian Meteorological Centre (CMC), the National Centers for Environmental Prediction (NCEP), the UK Met Office (UKMO) and the ECMWF. UKMO and CMC scores are only available until day 6. All models are verified against their own analyses.

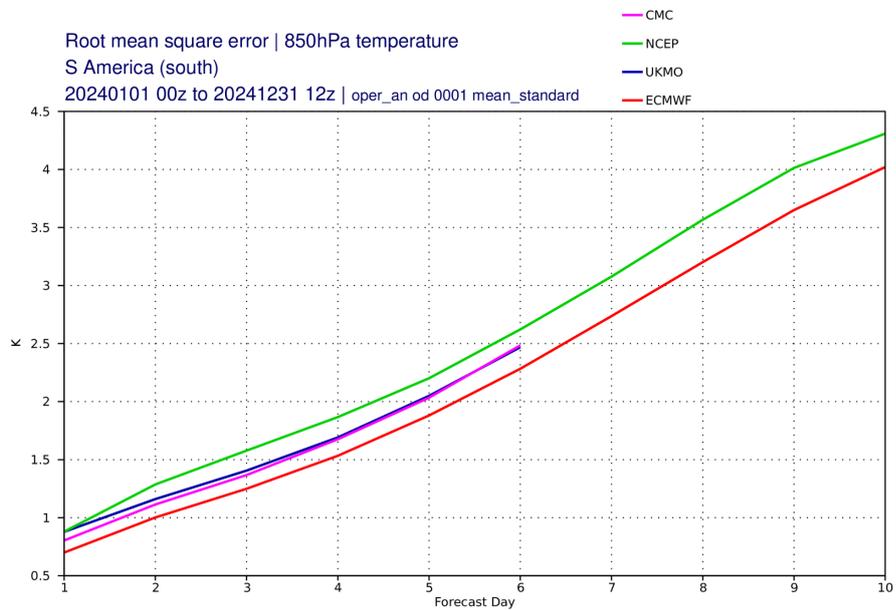


Figure 8: RMSE for 850-hPa temperature forecasts for the year 2024 in the southern South America domain for forecasts from four global forecast centres: the Canadian Meteorological Centre (CMC), the National Centers for Environmental Prediction (NCEP), the UK Met Office (UKMO) and the ECMWF. UKMO and CMC scores are only available until day 6. All models are verified against their own analyses.

3.1.2 2-metre temperature and 10-metre wind speed

In addition to the upper air variables, ECMWF also verifies the southern South American domain for surface variables against its own analysis. Compared to the verification against SYNOP, this has the disadvantage that systematic errors present in both the analysis and the forecast are not detected. On the plus side, the results are not affected by regional differences in sampling, thus making results over different domains more directly comparable. They are also not affected by the scale mismatch (also called representativeness mismatch) between the point observation and the grid-scale forecast.

Figure 9 illustrates the evolution of the RMSE of the 2-metre temperature forecast over the last 20 years, which is similar to the RMSE at 850 hPa shown in Figure 6. The southern South American domain and the tropical domain have the smallest RMSE, reflecting lower values of baroclinicity in the region. Over the last 20 years, a reduction of about 0.5 degrees has been achieved on average but there has been little visible improvement in more recent years.

Figure 10 shows the evolution of the RMSE of 10-metre wind speed over the last 20 years: unlike the previous cases, southern South America exhibits verification results comparable to Europe, but significantly worse than North America. Further investigations are needed to clarify if this simply reflects on average more windy conditions in the South American domain.

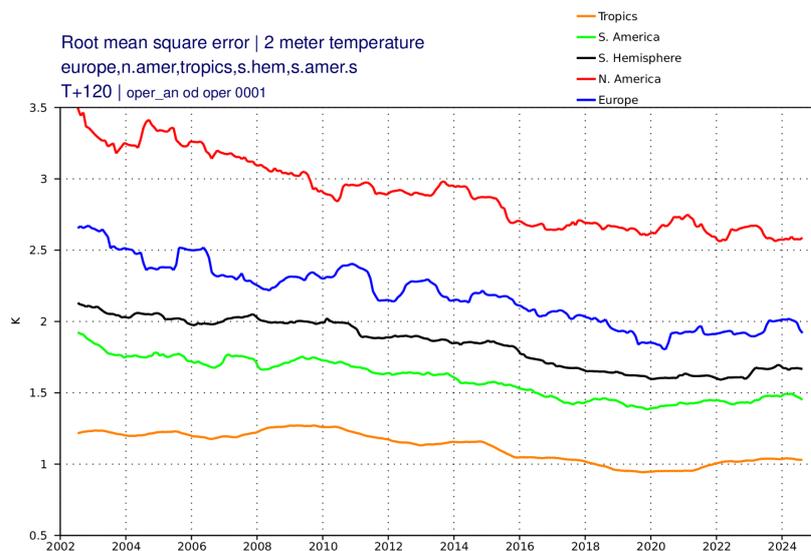


Figure 9: Evolution of the CTRL (ex-HRES) 2-metre temperature day-5 forecast RMSE for the last 20 years over the North American, European, Tropics, and Southern Hemisphere domains (shown in Figure 1) and the southern South America domain (shown in Figure 2, left panel).

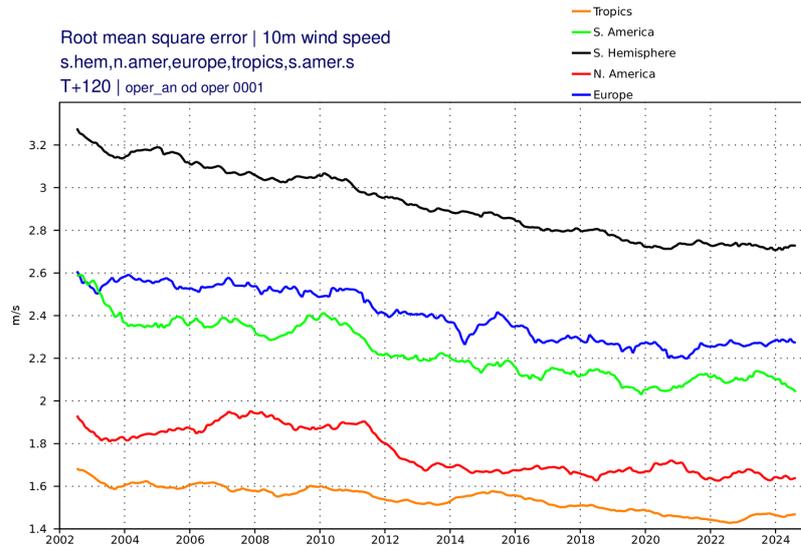


Figure 10: Evolution of the CTRL (ex-HRES) 10-metre wind speed day-5 forecast RMSE for the last 20 years over the North American, European, Tropics, and Southern Hemisphere domains (shown in Figure 1) and the southern South America domain (shown in Figure 2, left panel).

3.2 Verification against surface observations

This section focuses on verification with surface synoptic observations (SYNOP). Unlike the previous section, here domain-average results are shown over the extratropical South American domain proposed by the TT-NWPSV (see Figure 2, right panel).

3.2.1 2-metre temperature

Figure 11 shows the RMSE and ME (bias) of the CTRL (ex-HRES) 2-metre temperature forecast for the year 2024 for a 72 h forecast. Most of the southern plains (eastern Argentina, Uruguay, Paraguay and southern and northeastern Brazil) have a bias of less than 1 K magnitude. In more central areas, where the bias is slightly larger (1-2 K), it is usually positive. Stations with negative bias are few, and mostly along the Andes as well as in some places in Brazil. The RMSE increases from values between 1 and 1.5 K in northeastern Brazil to values around 3 K in the more central and extra-tropical areas of South America. These values are comparable to those of non-alpine southern Europe (not shown). Interestingly, the Andes do not appear as an area of increased RMSE beyond the values found in central areas.

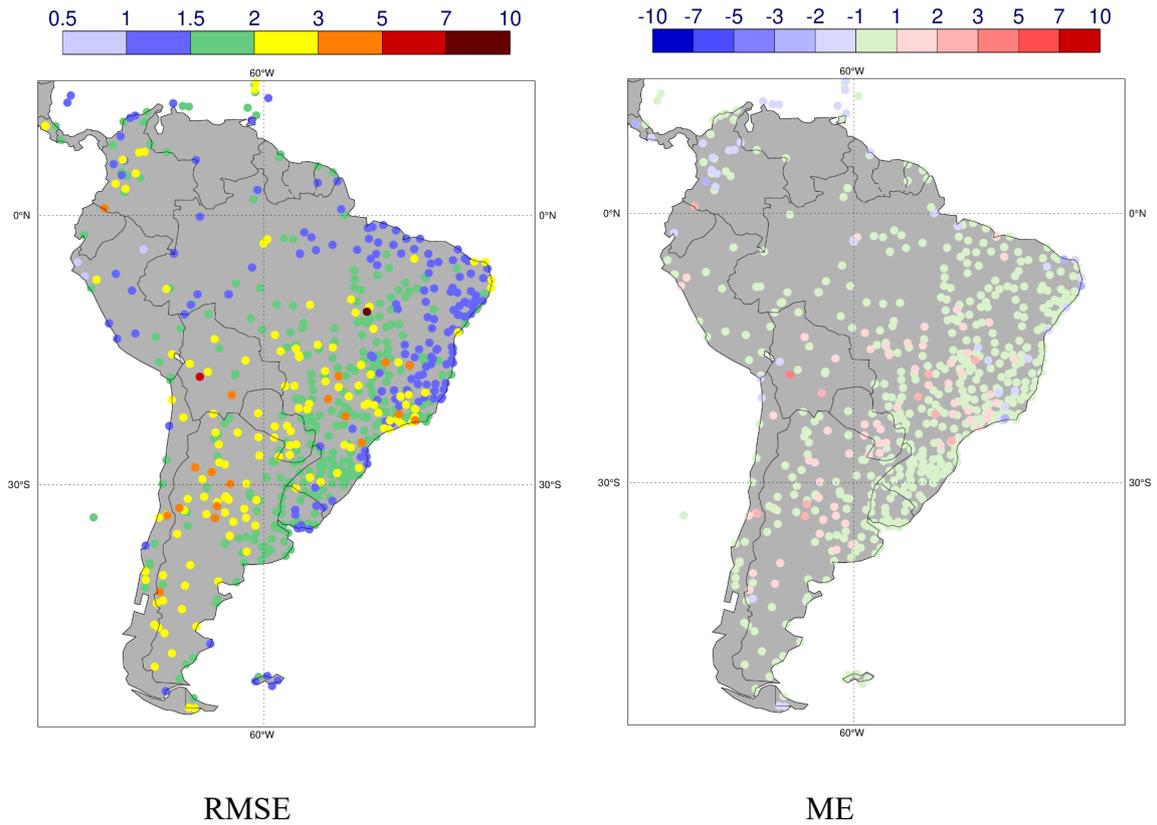


Figure 11: Error of the 72h CTRL (ex-HRES) 2-metre temperature forecast for the year 2024 measured against SYNOP observations. The left side shows the Root Mean Square Error, while the right shows the Mean Error. Results correspond to the 12 UTC run, units are Kelvin.

Figure 12 illustrates the time series of 2-metre temperature RMSE (blue) and ME (green) for the extratropical South American domain (Figure 2b), as well as the number of meteorological stations available for the calculation (grey bars). A seasonal variation can be observed for both scores, while the 12-month average RMSE remains unchanged overall and the ME decreases from 0.6 to 0.4 K over the study period. The number of available weather stations has increased considerably, mainly due to changes in Brazil.

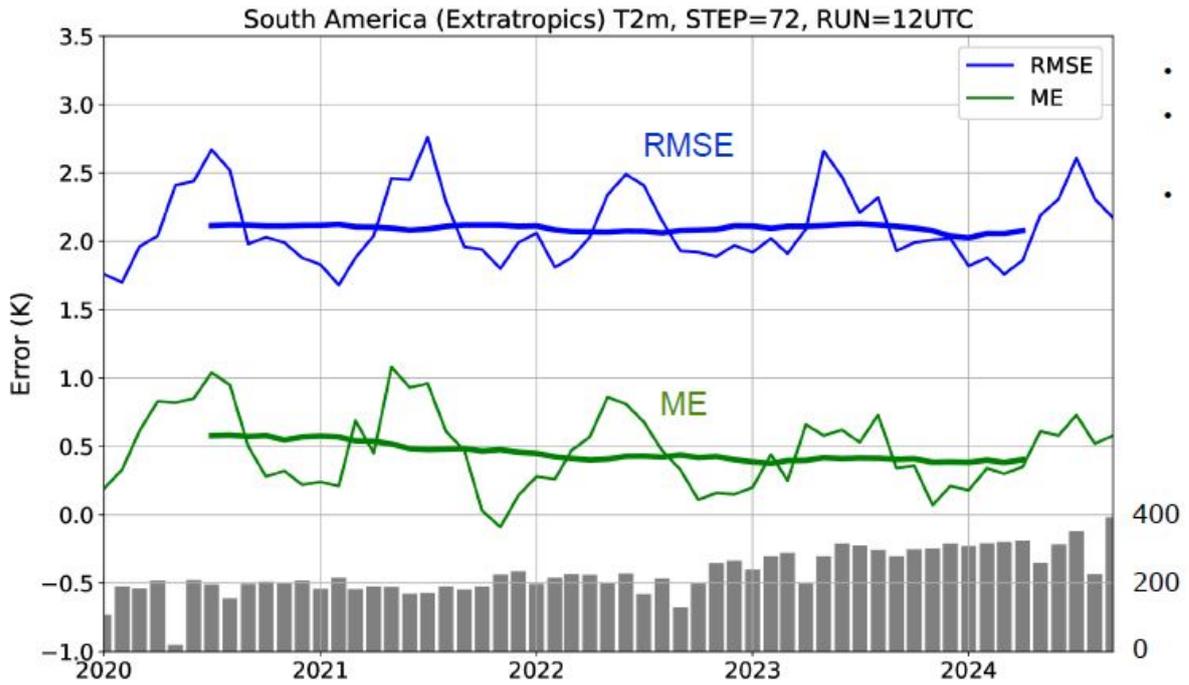


Figure 12: Time evolution of domain-averaged RMSE (thin blue curve) and ME (thin green curve). Thick curves present a 12-month running mean. Grey bars with scale on the right side display the number of meteorological stations available for verification.

Figure 13 shows forecast error (RMSE and ME) as a function of lead time for different seasons (JJA 2024 and DJF 2024-25). Results for the extratropical South American domain are located on the left side, while on the right results from Europe are shown for comparative purposes. While the RMSE values are overall relatively similar between South America and Europe, the seasonal dependence is different. In Europe, summer and winter scores are close to each other, while in South America the winter RMSE is significantly larger than the summer one.

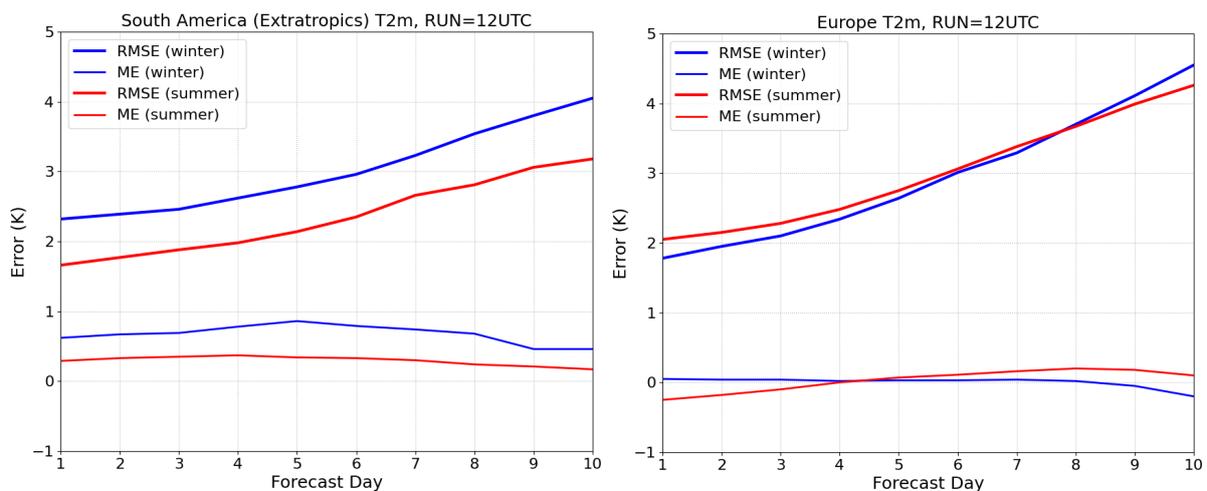


Figure 13: Forecast error (RMSE and ME) of 2-metre temperature as a function of lead time for different seasons. The left panel shows results for the extratropical South America domain shown in Figure 2, the right panel shows results for the European domain shown in Figure 2. The analysed periods are JJA 2024 (winter in South America and summer in Europe) and DJF 2024-25 (summer in South America and winter in Europe).

Another difference is the error growth beyond Day 6-7: in Europe it remains steep, while in South America it flattens, so that at Day 10 the RMSE in South America is lower than in Europe in both seasons. It would need to be further investigated if this is a typical feature or the result of the lower-than-usual medium-range predictability seen in Europe in the winter 2024-25. The domain-averaged ME is larger, and more clearly positive, in South America. However, the domain-averaged ME hides regionally varying biases, which are of both signs in Europe, so that compensation occurs. It also needs to be kept in mind that this verification is for 12 UTC which represents a different local time, by a few hours, on the two continents.

3.2.2 24-h precipitation

Figure 14 presents the forecast error for 24-h precipitation for different lead times for the Stable Equitable Error in Probability Space score (SEEPS, Rodwell et al. 2010) for extratropical South America (left) and for Europe (right) as defined by ECMWF, in Figure 2 (left panel). The analysed periods are JJA 2024 (summer in NH and winter in SH) and DJF 2024-25 (winter in NH and summer in SH). Results are rather similar between continents for winter precipitation. For summer precipitation, South America displays worse performance at shorter lead times, and slightly better performance from day 6 onwards. A flatter lead-time dependence of precipitation forecast performance is indicative of a bigger contribution coming from convective (as opposed to frontal) events, however the difference between the curves is not very large, so some of it may be due to the specific weather patterns in the two periods.

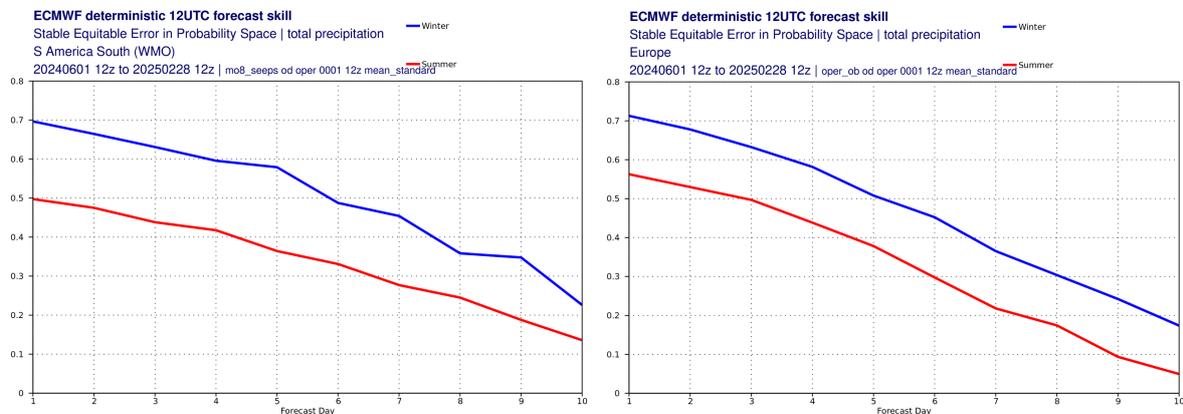


Figure 14: Forecast performance for 24h precipitation as a function of lead time for the Stable Equitable Error in Probability Space (SEEPS) score in extratropical South America (left) and in Europe (right), with both domains as shown in Figure 2. The analysed periods are JJA 2024 (winter in South America and summer in Europe) and DJF 2024-25 (summer in South America and winter in Europe).

3.2.3 2-metre dewpoint temperature

Figure 15 shows the RMSE and ME for the 2-metre dewpoint temperature 72h CTRL (ex-HRES) forecast. There is a tendency for a dry bias towards the east of the continent and the west (Andes) and a slight tendency towards too wet in central Argentina. The lowest RMSEs are found near the coasts, likely due to the ocean being a reliable source of advected humidity. Highest RMSEs are found in the Andes, partly due to the more difficult forecasting in complex

terrain but also due to the nonlinearity of the relationship between dewpoint and relative humidity. In dry conditions, a small error in relative (or specific) humidity translates into a larger dewpoint error.

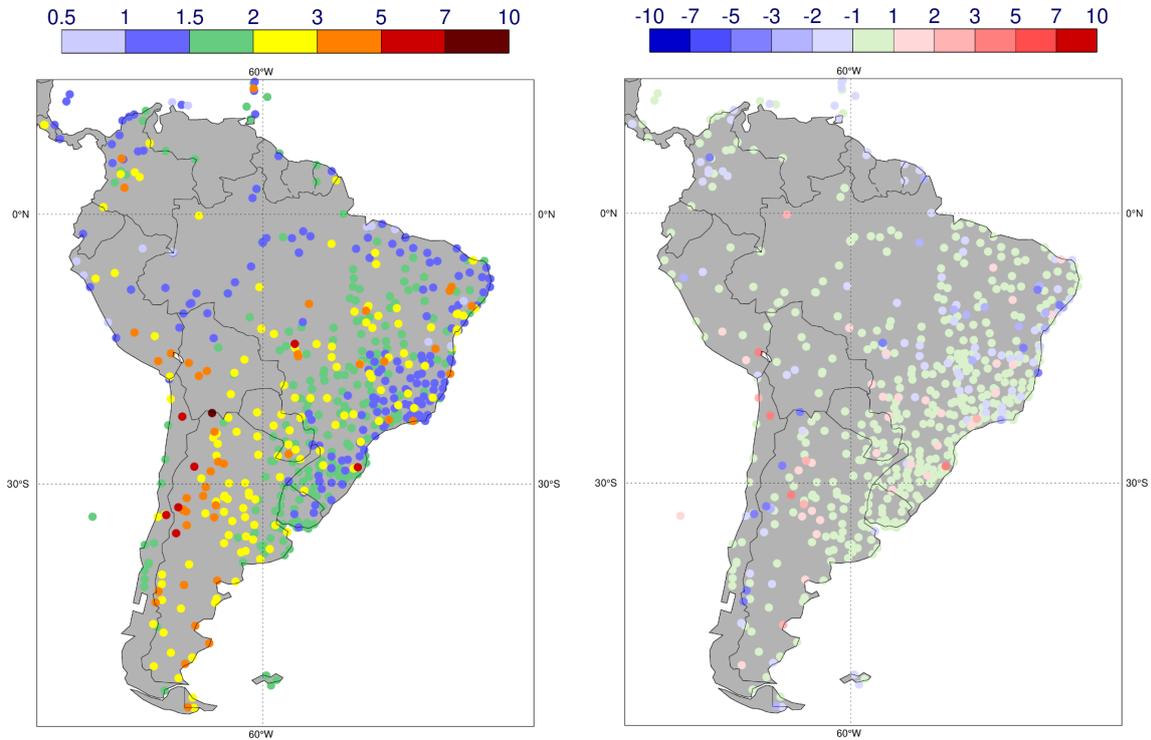


Figure 15: Error of the 72h CTRL (ex-HRES) dewpoint forecast for the year 2024 measured against SYNOP observations. The left side shows the Root Mean Square Error, while the right shows the Mean Error. Results correspond to the 12 UTC run, units are Kelvin.

3.2.4 10-metre wind speed

Figure 16 illustrates the RMSE and ME for 10-metre wind speed 72h CTRL (ex-HRES) forecast over South America. There is a near zero bias in the center of the continent while it becomes positive (forecast wind exceeds observations) towards the east, particularly in Brazil. Contrarily, Chile and parts of Argentina show a slight negative bias (observed winds stronger than predicted). This is consistent with results from Europe, where an underestimation of 10 m wind speed is found for mountainous areas. Also in terms of error magnitude, results shown here are comparable to those over Europe, which are generally for RMSEs around 2.0 to 2.5 m/s and for MEs around 0.5 m/s (see Bouallègue et al. 2023).

In summary, the comparative evaluation of forecast performance for surface parameters in extratropical South America and Europe shows a similar magnitude of errors when verified against SYNOP observations. There are regionally varying biases which should be further investigated.

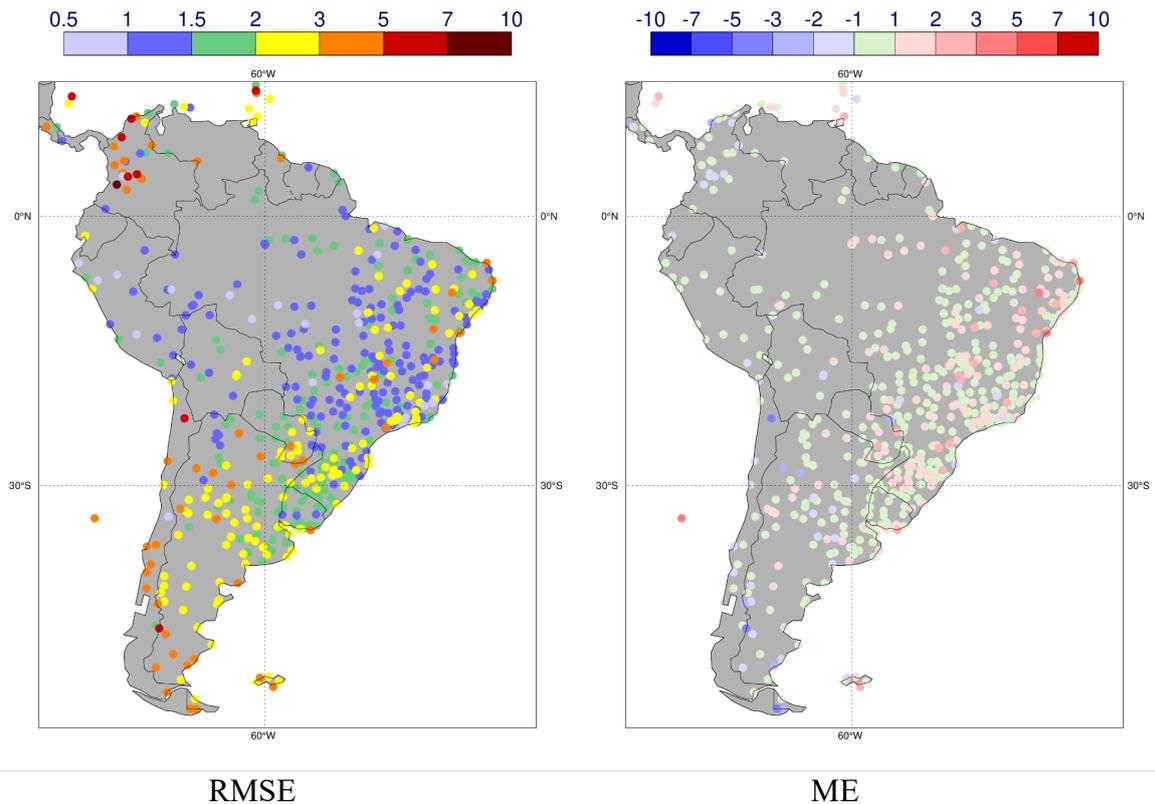


Figure 16: Error of the 72h CTRL (ex-HRES) 10-metre wind speed forecast for the year 2024. The left side shows the Root Mean Square Error, and the right the Mean Error between forecast and SYNOP observations. Results correspond to the 12 UTC run, units are m/s.

3.3 Comparison between global models and SAP.SMN-DET for surface variables

This section continues with the verification of forecasts against surface stations but here the CTRL (ex-HRES) is compared with two models, the GFS, and the limited-area SAP.SMN-DET over Argentina's continental territory. The results focus on 101 surface stations for the five-year period from January 2020 to December 2024. The forecast initialization time at 06 UTC is chosen for its relevance to local forecasters and because it skips the model spin-up as the forecasts are valid at 12 UTC (particularly 24-hour precipitation). Below we illustrate the main features of this comparison for 2-metre temperature, 24-hour precipitation and 10-metre wind speed. For 2-metre temperature and 10-metre wind speed, the results are presented in two ways, considering raw and calibrated data. Calibrations are performed on the GFS and the SAP.SMN-DET to improve model guidance for SMN forecasters who take calibrated values from these two models as default forecasts (see Righetti et al. 2024 and Cutraro et al. 2020). Although a comparison with raw CTRL (ex-HRES) forecasts may not constitute a “fair” comparison, it does illustrate the conditions under which SMN forecasters use them, as calibrations for CTRL (ex-HRES) forecasts are not carried out locally.

3.3.1 2-metre temperature

Figure 17 shows the ME (bias) and RMSE of the 2-metre temperature forecasts against Argentine surface stations for CTRL (ex-HRES), GFS and SAP.SMN-DET. The left panel uses model raw forecast data while the right panel uses calibrated data for the GFS and for the SAP.SMN-DET.

The raw data show that the three models have a noticeable diurnal cycle in the bias, with the CTRL (ex-HRES) having a negative bias near the maximum temperature hour, and a positive bias near the minimum hour, but a very small net bias. Both the GFS and the SAP.SMN-DET show a diurnal cycle of similar amplitude but shifted in phase and with a net positive bias. The CTRL (ex-HRES) shows a clear advantage in RMSE. The right panel shows that despite the calibration in both the GFS and the SAP.SMN-DET, the raw output of the CTRL (ex-HRES) remains competitive in RMSE.

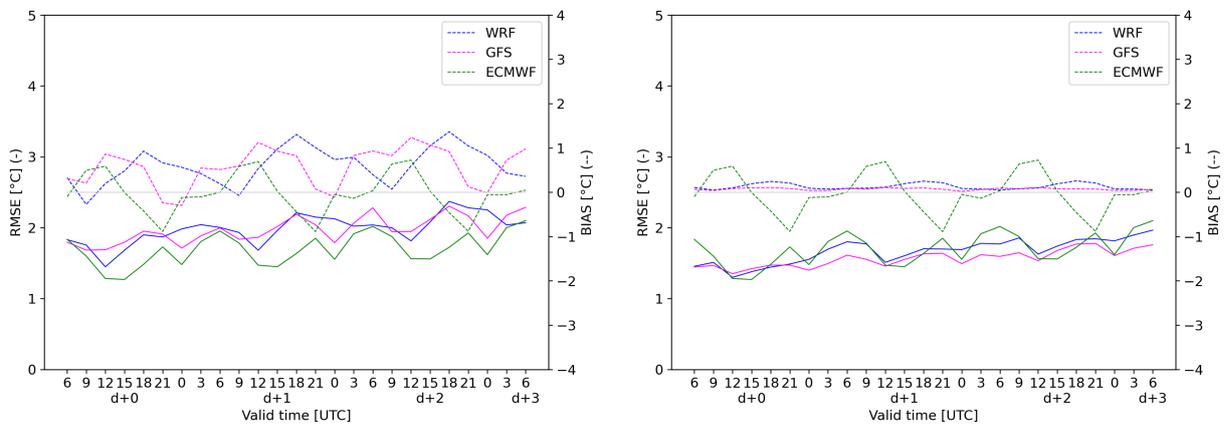


Figure 17: ME (bias) (dashed lines) and RMSE (solid lines) for 2-metre temperature forecast against Argentine weather stations for CTRL (ECMWF in legend), GFS and SAP.SMN-DET (WRF in legend). The left panel uses the model’s raw data while the right panel considers calibrated data for GFS and for the SAP.SMN-DET.

3.3.2 24-h precipitation

The Taylor diagram shown in Figure 18 compares 24-hour precipitation forecasts at 30 and 54 hours for the CTRL (ex-HRES), GFS and SAP.SMN-DET. The results show that for both lead times the CTRL (ex-HRES) obtains the same correlation as the GFS but a more realistic variance (note, however, that CTRL reaches a higher RMSE due to its larger variance, probably because of the “double penalty” effect). The SAP.SMN-DET forecast has a slightly better variance compared to that of the CTRL (ex-HRES) but with lower correlation.

Figure 19 shows the distribution of precipitation volume as a function of precipitation rate for 24-h precipitation forecasts at 30 hours (left panel) and 54 hours (right panel) for each model (colour lines) and observations (grey bars). The SAP.SMN-DET fits with the observations quite well, while the two global models underestimate large values of precipitation rate. For lower precipitation rates the observed distribution is somewhat noisy, hence making it difficult to make a robust comparison. Of the three models, it is the GFS that forecasts the largest amount of precipitation at low precipitation rates.

The box within each panel shows the total amounts of forecast and observed 24-h precipitation as well as their relative value using the observations as a reference. Both the GFS and the WRF underestimate the total, while the CTRL (ex-HRES) overestimates it. All models show an increase in the total 24-h precipitation for the longest lead time, which for the CTRL (ex-HRES) means an increase in the overestimation to almost 30%.

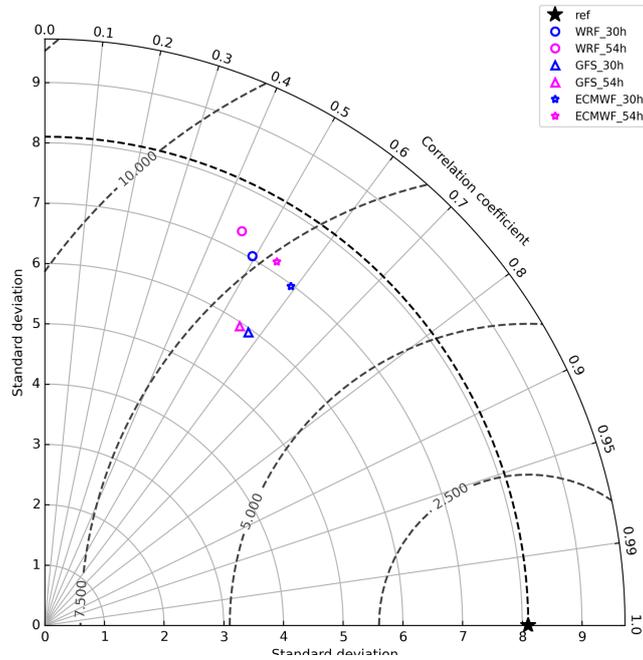


Figure 18: Taylor diagram showing 24-precipitation forecasts at 30 hours (blue) and 54 hours (pink) for the CTRL (ECMWF in legend, star), GFS (triangle) and SAP.SMN-DET (WRF in legend, circle). The reference observed value is marked with a large black star. Radial distances indicate the standard deviation, angles represent Pearson's correlation, and sections of circumference centred in the black star denote isolines of equal RMSE values.

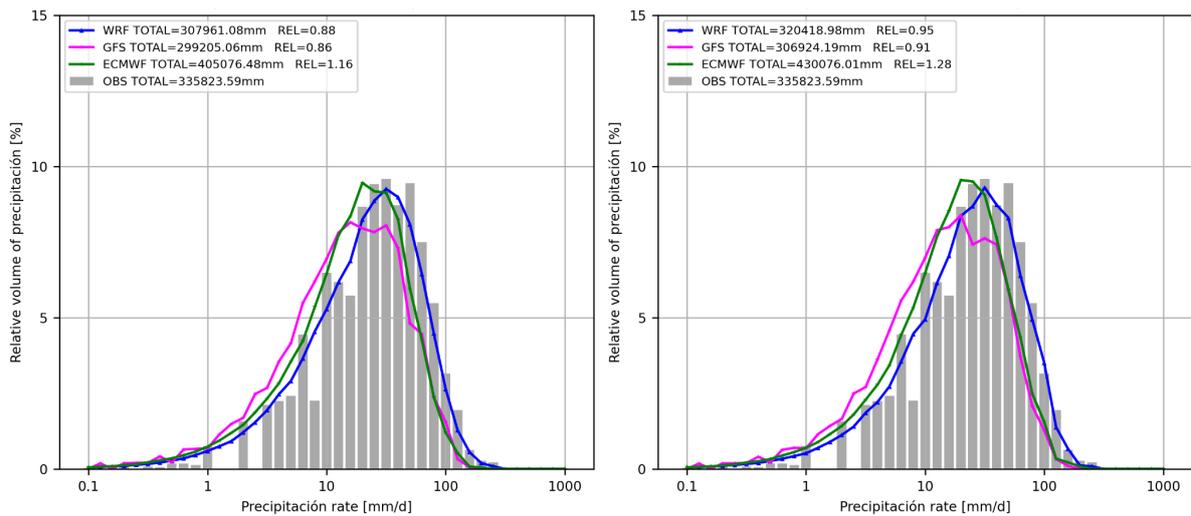


Figure 19: Distribution of precipitation volume as function of precipitation rate for 24-h precipitation forecasts at 30 hours (left panel) and 54 hours (right panel) for each model (colour lines), CTRL (ECMWF in legend), GFS and SAP.SMN-DET (WRF in legend). Observations are represented with grey bars. Boxes on the upper left of each panel show the total amounts of forecast and observed 24-hour precipitation as well as their relative value taking observations as reference.

A performance diagram (Roebber 2009) for 24-h precipitation forecast for different thresholds (colour) and lead times (triangles for 30 h and circles for 54 h) is shown in Figure 20, with SAP.SMN-DET on the left panel, GFS in the centre, and CTRL (ex-HRES) on the right. The SAP.SMN-DET shows a rather low frequency bias at all thresholds while both global models show a concave shape indicating an underestimation of strong precipitation events and an overestimation of weak precipitation. This behaviour is typical of coarse resolution models, which tend to under-forecast intense precipitation and over-forecast the trace (due to the representativeness mismatch and the intrinsic limits inherited by the grid resolution). For CTRL (ex-HRES) the least biased value corresponds to 25 mm accumulated in 24 hours, while for the GFS it seems to be between 5 and 10 mm. As expected, higher performance is associated with weaker precipitation accumulations, as measured by the Critical Success Index (shaded colour). Longer lead times, naturally, display lower performance for the three models.

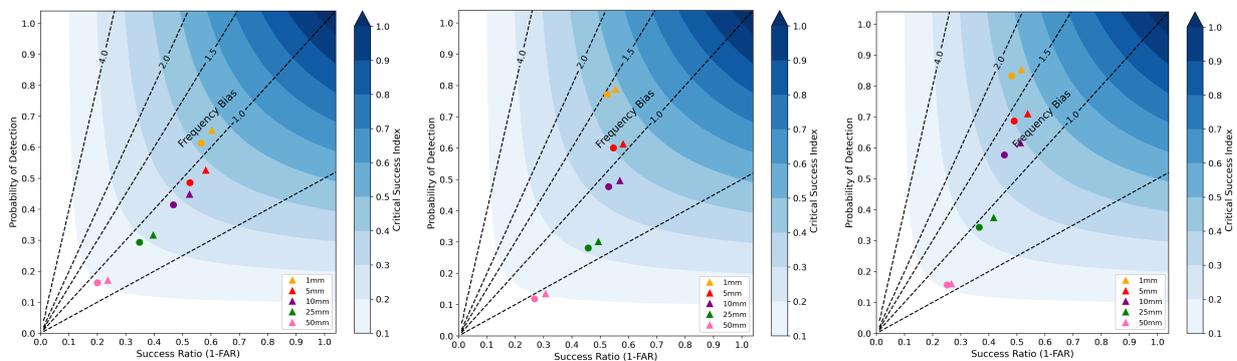


Figure 20: Performance diagram for 24-h precipitation for different thresholds (colour) and forecast lead times (triangles for 30 h, and circles for 54 h). The left panel presents the SAP.SMN-DET, the center panel the GFS and the right panel the CTRL (ex-HRES). The abscissa indicates the Success Ratio, the ordinate the Probability of detection and the shaded colour the Critical Success Index (see Roebber 2009).

3.3.3 10-metre wind speed

Figure 21 shows the ME (bias) and RMSE of the 10-metre wind speed forecast with respect to Argentine SYNOP observations for CTRL (ex-HRES), GFS and SAP.SMN-DET. The left panel uses raw forecast data while the right panel includes calibrated data for GFS and for the SAP.SMN-DET (not for CTRL, ex-HRES).

The raw data show that the three models have significant diurnal cycle in the bias, with the CTRL (ex-HRES) having the stronger fluctuation. Unlike the case for the 2-metre temperature, all models have the maximum of the bias cycle during the night hours and the minimum during the day. The SAP.SMN-DET and the GFS seem to have a net positive bias while the CTRL (ex-HRES) seems to have a negligible net bias. The CTRL (ex-HRES) shows an advantage in RMSE over both models. Unlike for 2-metre temperature and precipitation, the SAP.SMN-DET in this case exhibits the largest bias and error.

4.1 Windstorm of 16-17 December 2023

During the period 16-17 December 2023 a series of convective-related events took place in the centre of Argentina, leaving 14 fatalities and severe damage in several locations. In particular, the cities of Bahia Blanca and Buenos Aires were the most affected. Due to the severity of the case the SMN of Argentina conducted a study to capture the perspectives of the different components involved, from observations to forecasting to impact (see Saucedo et al. 2024). As a complement, a questionnaire of the warning chain developed by HIWeather was filled, which included additional information (see Saucedo et al. 2025). Here we will briefly discuss the event from the perspective of the capabilities and limitations of the IFS to capture and predict it.

The meteorological situation associated with the event was characterized by the presence of a warm, very humid and unstable air mass located over the centre and north of Argentina. A frontal zone was located over the north of the Patagonian region which began to move towards the centre of the country. To the north of this system during the afternoon hours of Saturday, December 16, temperatures were between 29 and 35°C with dew point temperatures between 19 and 25°C.

At middle and high levels of the troposphere, a deep trough was located over the Pacific Ocean, which advanced towards the east, giving rise to the formation of rain and storms over the provinces of Mendoza, Río Negro, La Pampa and San Luis. Radiosonde information from the morning of Saturday 16th showed the instability of the air mass over the centre and north of the country with MUCAPE values above 2000 J/kg, with peaks during the afternoon between 3000 and 4500 J/kg, together with precipitable water values of the order of 40 to 50 mm. Likewise, an evening sounding showed a northeasterly flow of 20 knots at 500 meters high, turning north at 30 knots at 1000 meters, which denotes a strong vertical wind shear at the lower levels. Consequently, there was a clear risk of severe storms and favourable conditions for the occurrence of very violent instability lines and even tornadoes. Orange warnings were issued.

The events and the associated impacts corresponded with the warnings issued. Figure 22 shows the hazards registered between the 16th and the 18th December in the SAMHI database (see Salio et al. 2024), giving a good idea of the area impacted and the focus in the cities of Bahia Blanca and Buenos Aires. The AWOS located at Bahía Blanca's airport, recorded wind speed peaks of 154 km/h, within a period of around 30 minutes during which gusts remained above 120 km/h. In the AMBA region, at Buenos Aires airport there were maximum wind speeds of 142 km/h.

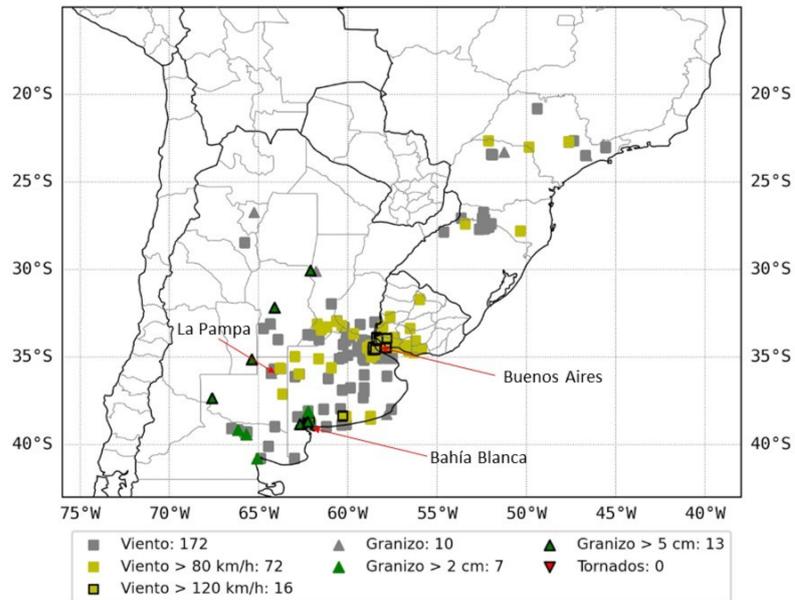


Figure 22: Map of meteorological phenomena extracted from the SAMHI database for December 16, 17 and 18, 2023. Most of this data comes from social media posts and, in the case of wind, from measurements by unofficial meteorological stations. Note that viento=wind and granizo=hail. Source: SAMHI database (see Salio et al. 2024).

Figure 23 shows the evolution of EFI and SOT for CAPE-shear, from the operational IFS forecast, valid for 16 December 2023, the day the Bahía Blanca event took place. Both indices are tools to identify in the “model world” potentially extreme weather compared to model climatology for a given location and time of year. Near Bahía Blanca, EFI values are greater than 0.8 with SOT Index taking positive values at least 5 days in advance (December 11) representing a very good guidance for forecasters.

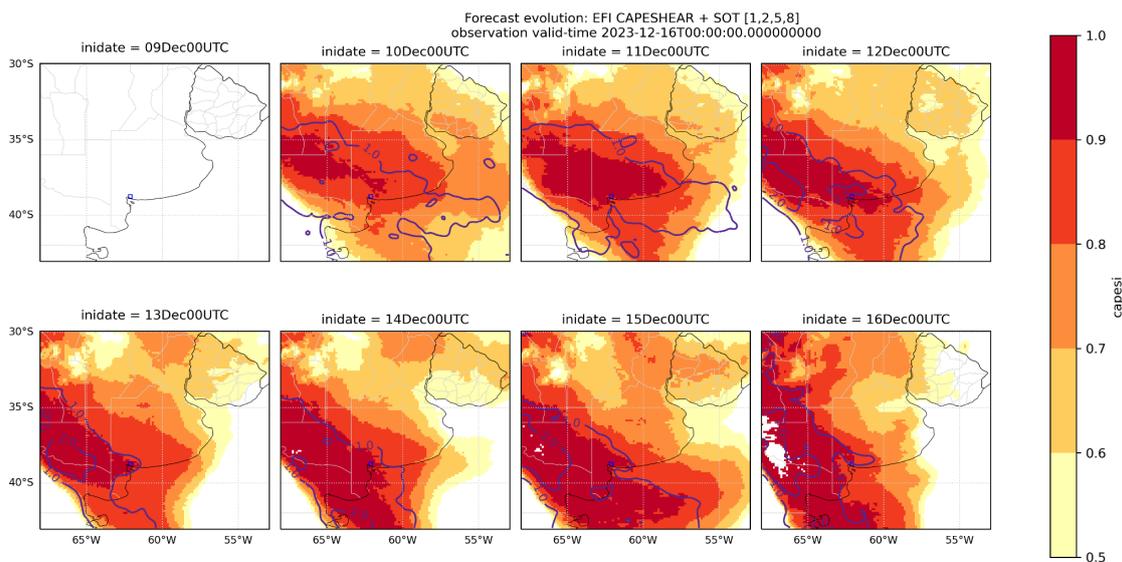


Figure 23: Evolution of the EFI (colours) and SOT (contours) valid for December 16 for CAPE-shear, based on the operational IFS forecast. Each field corresponds to a different initialization date. No data is available for the initialization on December 9.

Figure 24 displays forecast evolution of EFI Index values for the two large cities affected, Bahia Blanca (left panel), and Buenos Aires (right panel). For this calculation, a 0.25° box around each location was considered to obtain the mean EFI values. EFIs for five variables are presented: CAPE, CAPE-shear, precipitation, wind speed and wind gusts. For both cities, EFIs for CAPE, CAPE-shear and precipitation are considerably large several days in advance, particularly in Bahia Blanca (left panel). Both parameters related with wind speed do not present values associated with danger. For the case of Buenos Aires (right panel), the EFI for wind speed is stronger but inspection shows that these values are mostly associated with the synoptic scale northerly wind bringing humidity prior to the convective developments.

Forecasters were not relying only on global models to warn about the appropriate hazards. They were also using conceptual models that relate a given weather situation with a given hazard. In addition, they had at their disposal the higher-resolution limited area operational forecast produced by the SAP.SMN-DET (described in section 2.2.2). Figure 25 shows the wind speed and wind gusts forecasts by the time hazards hit both cities, Bahia Blanca (left panel) and Buenos Aires (right panel). Although these values are below the observed ones, they are unusual enough to confirm the suspicions and subsequent warning issued by the forecasters. This signal was absent in the CTRL (ex HRES) forecast (lower panels). It is important to mention that when the regional ensemble SAP.SMN-ENS was analysed, the strong wind signals mentioned above were only present in some of the 20 members, which indicates a high level of uncertainty associated with the wind hazard (not shown).

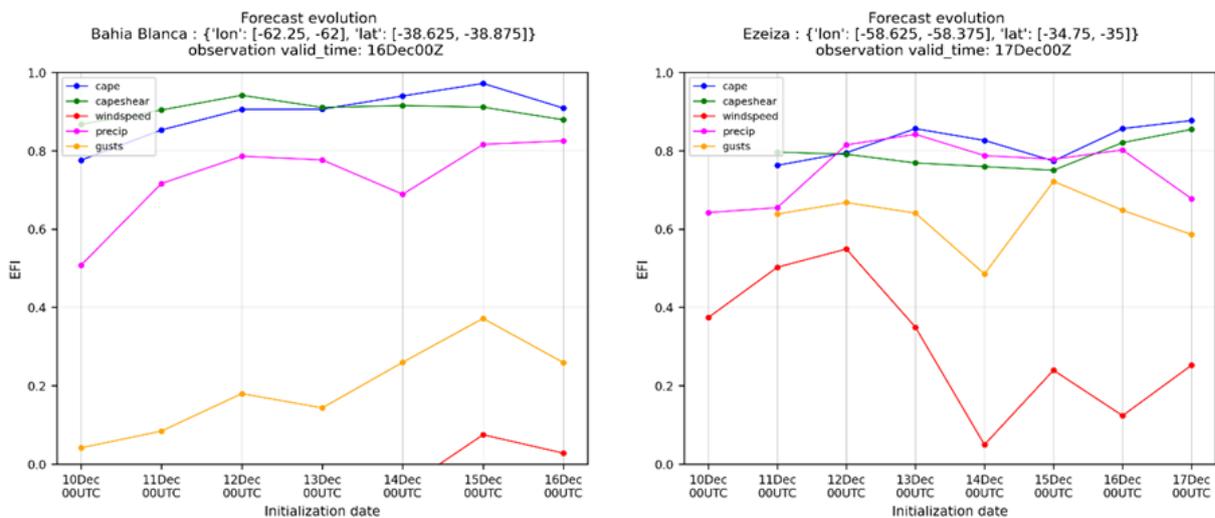


Figure 24: Evolution of the EFI forecast for different variables for Bahía Blanca (left panel) and Buenos Aires (right panel). The abscissa indicates the start day of the forecast, all of which are valid for the day of the event in the respective cities.

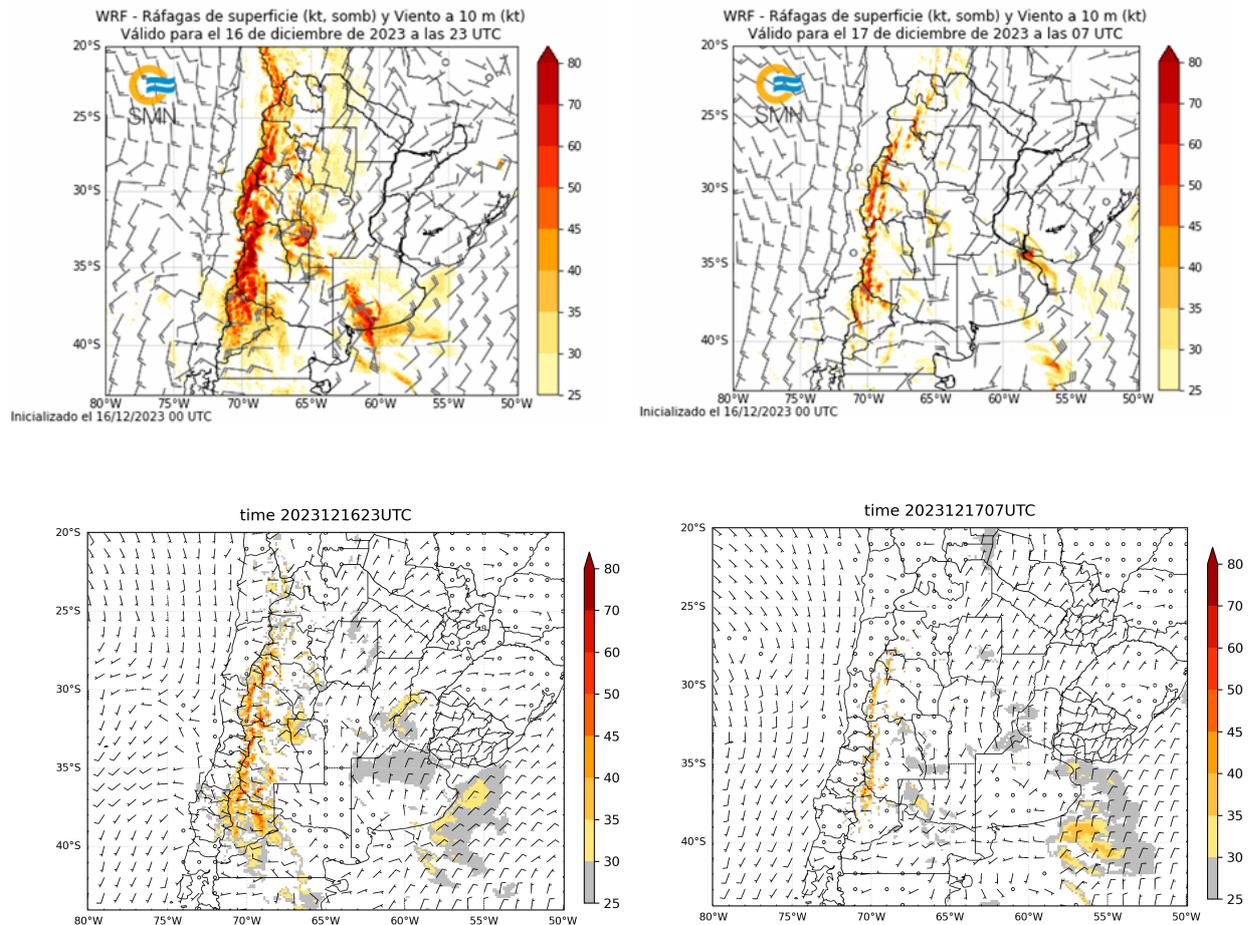


Figure 25: Forecasts for 10 m wind and surface gusts initialized on 16 December 2023 at 00 UTC. The upper row shows the operational SAP.SMN-DET forecasts valid for 16 December 2023 at 23 UTC (left panel, related to event in Bahia Blanca), and for 17 December 2023 at 07 UTC (right panel, related to event in Buenos Aires). The lower row is similar to the upper row but for CTRL forecasts (ex HRES).

The convective case analysed here involves two distinct challenges for the forecasts: on the one hand they are expected to predict well in advance the general weather conditions (or “ingredients”) that may create an environment favourable for severe storms, while on the other, they are scrutinized for signs of small-scale hazards related to wind or precipitation.

With respect to the former perspective, EFIs related to the general environment (CAPE and CAPE-shear) show a very useful level of predictability for the weather system –with clues about the severity of the event already present 6 days in advance–, and a clear potential for direct use as an early warning tool. These results are comparable to those obtained for Europe and North America (e.g., Tsonevsky et al 2018) and go along the same line as those of section 3.

For the latter perspective, however, CTRL (ex HRES) forecasts do not show direct signs of the presence of remarkable wind gusts, while SAP.SMN-DET does a good job in capturing the event. Given the scale of the phenomena and the respective resolution of the forecasts this is not unexpected, as shown for example in Magnusson et al (2018), although it is a good indication that the operational limited-area forecast from the SMN is adding value to the information provided by global models.

4.2 Zonda wind of 21 July 2023

Zonda wind (Argentinian foehn) is a strong, warm, dry wind associated with adiabatic compression upon descending the eastern slopes of the Andes in central Argentina (for a detailed review of the Zonda wind see Norte 2015). Zonda has a remarkable spatial and temporal variability, with hazards associated primarily with unseasonably high temperatures, exceedingly low humidity, and dangerous winds. In populated regions, it produces a range of damage according to the intensity of wind gusts (e.g., blowing off roofs, felling trees and high-voltage electricity lines, toppling vehicles and shutting down power supply and communication lines). In addition, it favours the ignition and propagation of fires and causes damage to crops due to the strong gusts, sudden dryness and high temperatures, and it may also be responsible for premature fruit flowering. At high altitudes in the mountains, its occurrence accelerates snow melting and evaporation, modifying the snowpack depth, contributing to avalanches, and influencing the hydrological cycle.

On 21 July 2023, a severe Zonda wind event occurred in the city of Mendoza. This event started at 18 UTC and lasted for 9 hours and caused significant damage, including 2 deaths, with wind gusts of over 100 km/h, a temperature increase of more than 10°C, and a decrease in dewpoint temperature of 20°C in one hour.

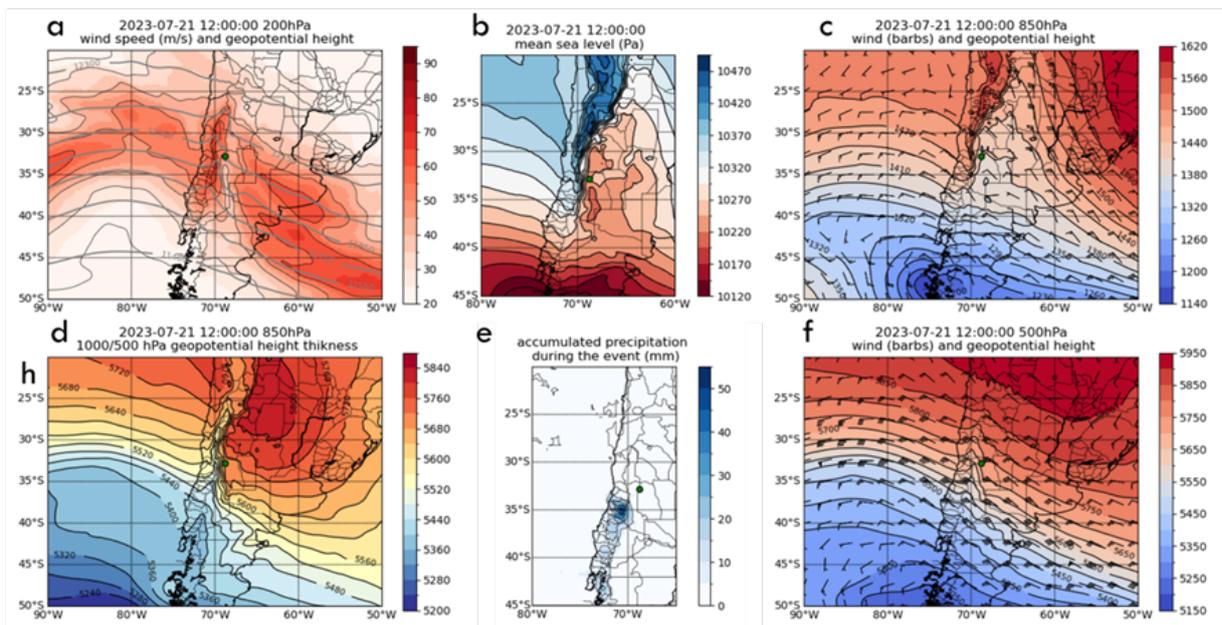


Figure 26: Synoptic situation previous to the development of the Zonda wind event on 21 July 2023 at 12 UTC. a) 200 hPa wind speed (shaded, m/s) and geopotential height (grey lines, gpm), b) mean sea level pressure (Pa), c) and f) geopotential height (shaded, gpm) and horizontal winds (barbs) for 850 hPa and 500 hPa, d) geopotential height thickness and e) hourly accumulated total precipitation during the event. All data taken from ERA5 reanalysis at 0.25° resolution.

As Figure 26 shows, the meteorological situation was characterized by an intense mid-level trough located over the Pacific, advancing toward the continent. At lower levels, an intense low-pressure system, ahead of the upper-level trough, was also moving eastward. Over the central-east and northern parts of Argentina, there was a very strong advection of warm

northerly winds (possibly of the low-level jet type). These conditions generated an intense pressure gradient across the Andes Mountains (a factor that favors the development of Foehn winds). On the Chilean side, the westerly flow was blocked from crossing the mountain range up to approximately 2500 meters. South of 35°S precipitation was observed, possibly associated with forced lift over the Andes. Above this humid air mass, there was a much drier air mass, which ultimately crossed the barrier, creating the leeward Zonda conditions. On the leeward side, the 12 UTC Mendoza vertical profile taken a few hours (12 UT, figure not shown) before the onset showed a dewpoint depression of more than 45°C at 850 hPa, while at the surface, it was only a few degrees.

Figure 27 shows meteograms for the Zonda event at Mendoza Aero station, comparing observations with different forecast hours from the CTRL (ex HRES) (from 5 to 53 hours) and a WRF (IANIGLA) hindcast initialized 11 hours before the event. Observations (in black) show typical winter temperature and humidity until around 12 UTC where temperature rises over 6 hours to reach just below 30 degrees to sustain this value for 9 hours. In parallel to the fast increase in temperature, the dewpoint falls first to -20°C and then to below -40°C, values which remain until the rapid cessation of the event. Wind speed reaches a maximum of 12.5 m/s at 18 UTC, and a maximum wind gust of 16.5 m/s (not shown).

For the CTRL (ex HRES, left panel) at 53-h forecast (light violet line), the event is present, but no sharp increase in temperature or drop in dewpoint is observed. The forecast underestimates extreme values, particularly for wind speed, and it also underestimates the duration of the event (starting later and ending earlier). The 41-h forecast (red line) shows results similar to the 53-hour forecast, but extreme values are closer to observations. At 29 hours, the CTRL (ex HRES) forecast shows much better agreement in wind and temperature, as well as good agreement in the event onset, but it finishes earlier. Dewpoint temperature, however, remains significantly higher than observed. At 17 hours, the CTRL (ex HRES) forecast performance improves significantly, particularly for 2m temperature, relative humidity, and wind speed and direction. However, at 5 forecast hours, the performance deteriorates, predicting a shorter event duration.

The right panel of Figure 27 shows the same analysis but for different IANIGLA (WRF) model configurations. Among them, the Mellor-Yamada-Janjic (MYJ) TKE scheme was tested with the following settings: smooth topography on (orange line) and off (red line), adaptive time stepping (blue), increased vertical resolution (80 levels, lilac line), and 3-hour input boundary conditions (brown line). Additionally, the MYNN 2.5-level TKE scheme was evaluated with smooth topography off (green line).

Before the event onset, all forecasts show very good agreement in 2m temperature, 2m dewpoint, relative humidity, and 10m wind speed, but not in wind direction. For the WRF model, the event starts at least 5 hours earlier and ends 3 hours later. A sharp increase in temperature and a corresponding drop in dewpoint are observed. However, forecasts underestimate extreme values (temperature, dewpoint, and wind speed) and overestimate the duration of the event.

A key difference between WRF and CTRL (ex HRES) is the abrupt increase in temperature simulated by WRF, in contrast to the more gradual warming trend in CTRL (ex HRES). In addition, WRF systematically anticipates the event onset (from 6 hours to more than 3 hours

earlier). The maximum temperature during the event (28.3°C) was underestimated in all configurations. A similar behaviour is observed in the dewpoint temperature drop, although none of the WRF simulations reach the extreme observed values.

For relative humidity, WRF does not drop below 10%, whereas CTRL (ex HRES) does, suggesting that the ECMWF model provides a more accurate representation of the extremely dry conditions during the event. Wind speeds are initially overestimated in some WRF configurations but tend to be underestimated in the later stages. Some configurations reflect the early onset of strong winds between 12 and 15 UTC, while others fail to clearly capture Zonda activity in wind speed values. Wind direction is poorly represented by WRF, showing significant deviations from observations, whereas ECMWF provides a more stable, though still imperfect, forecast.

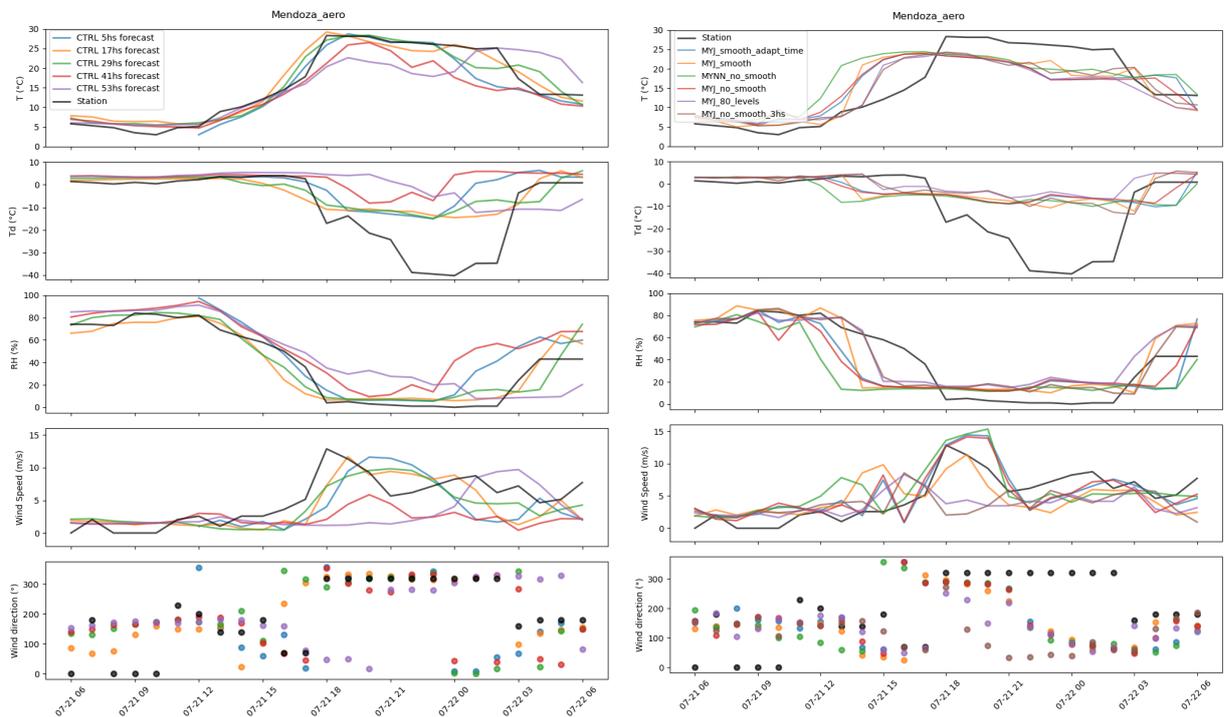


Figure 27. Time series for the closest model point to Mendoza Aero station of 2-metre temperature and dewpoint temperature (°C), relative humidity (%), wind speed (m/s), and wind direction (°). The left panel shows different forecast lead times for the CTRL (ex HRES) model, initialized at 2023-07-19 12 UTC (lilac line), 2023-07-20 00 UTC (red line), 2023-07-20 12 UTC (red line), 2023-07-21 00 UTC (orange line), and 2023-07-21 12 UTC (blue line). The right panel displays hindcasts from the IANIGLA (WRF) model, initialized at 2023-07-21 00 UTC, for different configurations: the Mellor-Yamada-Janjic (MYJ) TKE scheme with smooth topography enabled using an adaptive time step (blue), smooth topography enabled (yellow), smooth topography disabled (red), smooth topography disabled with 80 vertical levels (lilac), and smooth topography disabled with 3-hourly ERA5 boundary input (brown). Additionally, the MYNN 2.5-level TKE scheme was tested with smooth topography disabled (green). Black lines and dots correspond to observations.

Figure 28 shows the evolution of EFI and SOT for 10-metre wind gusts and 2-metre maximum temperature forecast for 24 and 48 forecast hours, valid for 21 July 2023, the day of the Zonda wind event at Mendoza Aero station (green dot). High EFI values are present at several locations over the lee slopes corresponding to some valleys of the Andes Mountain. EFI's greater than 0.8 are observed in the wind gusts for the 48 forecast hours and higher than 0.9 for

the 24 forecast hours. EFIs values with SOT Index above zero are noted in very small areas south of the station in the 48 forecast hours and just over the station (higher than 2) in the 24 forecast hours. In the same way, the maximum temperature reveals high EFI's and SOT values both, in the 24 and 48 forecast hours indicating the severity of this event. EFIs values with SOT Index above zero are noted in very small areas near (north and south) the surface station at least 5 days in advance (July 17, figures not shown) which denote the great spatial variability of this kind of event. For what we know of the event, the 24h forecast seems to constitute a very good numerical guidance for forecasters.

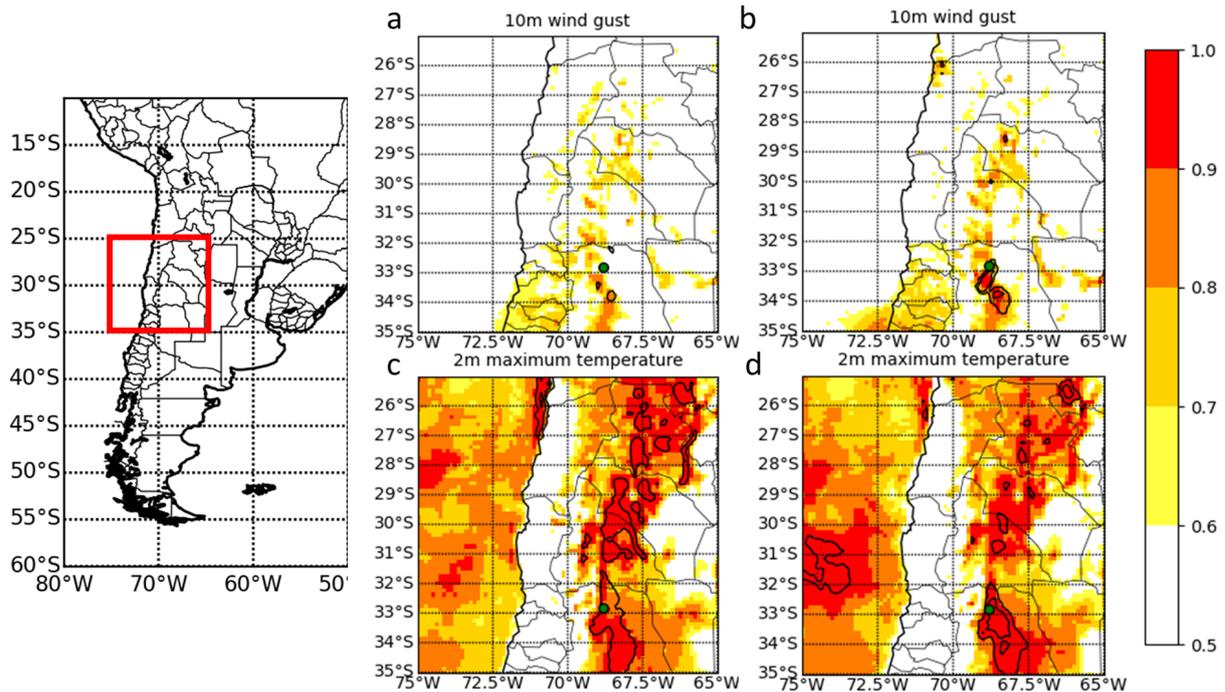


Figure 28: Evolution of EFI (shaded) and SOT (contours) valid for 21 July 2023 (00:00 to 23:59 UTC) in the IFS-ENS forecast for a) 48 hours forecast wind gust, b) 24 hours forecast wind gust, c) 48 hours forecast maximum temperature and d) 24 hours forecast maximum temperature. The green dot indicates the Mendoza Aero station.

Figure 29 presents the time series of 2-metre temperature (t_{2m} , solid lines) and 2-metre dew point temperature (t_{2d} , dashed lines) at Mendoza Aero station location from 06 UTC on 21 July to 06 UTC on 22 July. Various ECMWF IFS-based high-resolution forecasts are compared against observations (in black), in a similar way to the comparison shown in Figure 27. Among the model simulations, the CTRL run at 9 km resolution (blue) shows a temporal evolution of 2m temperature that closely follows the observations, though with a slightly earlier onset and a somewhat extended duration of the event. The reduced cloud-base mass flux (RCBMF) of the convective scheme experiment at 4.4 km resolution (purple) closely mirrors the performance of the DestinE forecast (IFS forecast runs at 4.4 km with the same initial conditions as the operational 9 km IFS), with both simulations capturing the event onset approximately two hours earlier than observed. The similarity between these two experiments suggests that, in this particular case horizontal resolution may play a more significant role than differences in model

physics. The IFS simulation at 2.8 km resolution (green) anticipates the event even earlier, by approximately 3 to 4 hours, and also simulates the temperature reduction sooner than the other models. It also considerably underestimates the dew point drop, as do all other configurations.

Overall, all IFS experiments considerably underestimate the magnitude of the 2m dew point temperature depression observed during the event. While the timing of the temperature increase is generally well captured, the persistence and intensity of the dew point anomaly are not fully reproduced and exhibit a time shift of 3 to 4 hours earlier than observed. This shift becomes more pronounced with increasing model resolution, as seen with the evolution of 2m temperature as well.

It is also worth noting that the 10m wind speed evolution (not shown) is similar across all IFS experiments, though all significantly underestimate surface 10m wind speeds and 10m wind gusts throughout the event. However, the 2.8 km simulation produces slightly higher wind gust values than the other configurations.

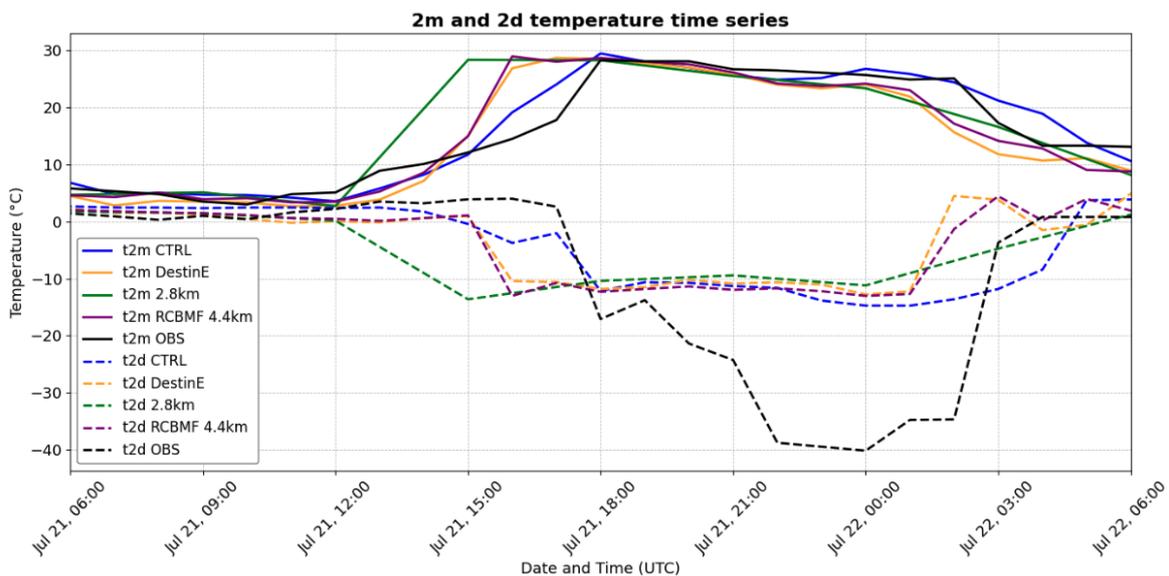


Figure 29: Time series for the closest model point to Mendoza Aero station of 2-metre temperature (t2m) and 2-metre dew point temperature (t2d) at Mendoza Aero station. Coloured lines represent forecasts from various ECMWF IFS model configurations: blue corresponds to the operational CTRL forecast at 9 km resolution; yellow to the DestinE forecast at 4.4 km; green to an IFS experiment at 2.8 km resolution; and purple to an IFS experiment at 4.4 km with reduced cloud base mass flux (RCBMF) applied in the deep convective scheme. All the forecast models are initialised on 21 July 2023 at 00 UTC. Observations are shown in black lines.

4.3 Heat wave of January 2024

In the second half of January 2024 an intense heat wave developed over central Argentina. The establishment of a high-pressure centre over the South Atlantic, with very slow movement and its associated circulation, led to the development of a heat wave between 21 January and 12 February 2024. It began in the south of the country and gradually spread northward and north-eastward, causing persistently high maximum and minimum temperatures. Several locations surpassed their previous record for highest daily maximum temperature during this prolonged

heat wave period. The evolution of the extreme heat index during the first week of the event is shown in Figure 30a, from January 23 to 29, 2024. Red areas indicate locations meeting extreme heat index criteria (Tmax and Tmin above heat wave thresholds), orange areas indicate locations where at least one of the criteria was met, and white areas indicate that neither criterion was satisfied. In locations where the extreme heat index persists for more than three days, it is considered a heat wave. ECMWF model forecasts for both the medium-range (days 1-15) and the sub-seasonal timescale up to three weeks in advance are used.

Analysing the evolution of daily EFI forecasts (Figure 30 b and c) from two initial dates, it can be observed that, with a lead time of two weeks, positive values are concentrated in central and southern Argentina during the first days of the target week. These values slowly shift northward while decreasing in magnitude towards the end of the target week. Similarly, forecasts made one week ahead (Figure 30b) show EFI temperature values above 0.7 in southern Argentina on the first day of the week, which then moved to the north of Argentina by the last day of the week. This pattern of high EFI temperature values matches the observed stations reporting high extreme heat index (Figure 30a) during the target week. This result suggests that the ensemble forecast was able to predict the air mass movement even with a 2 week lead time, although excluding Patagonia.

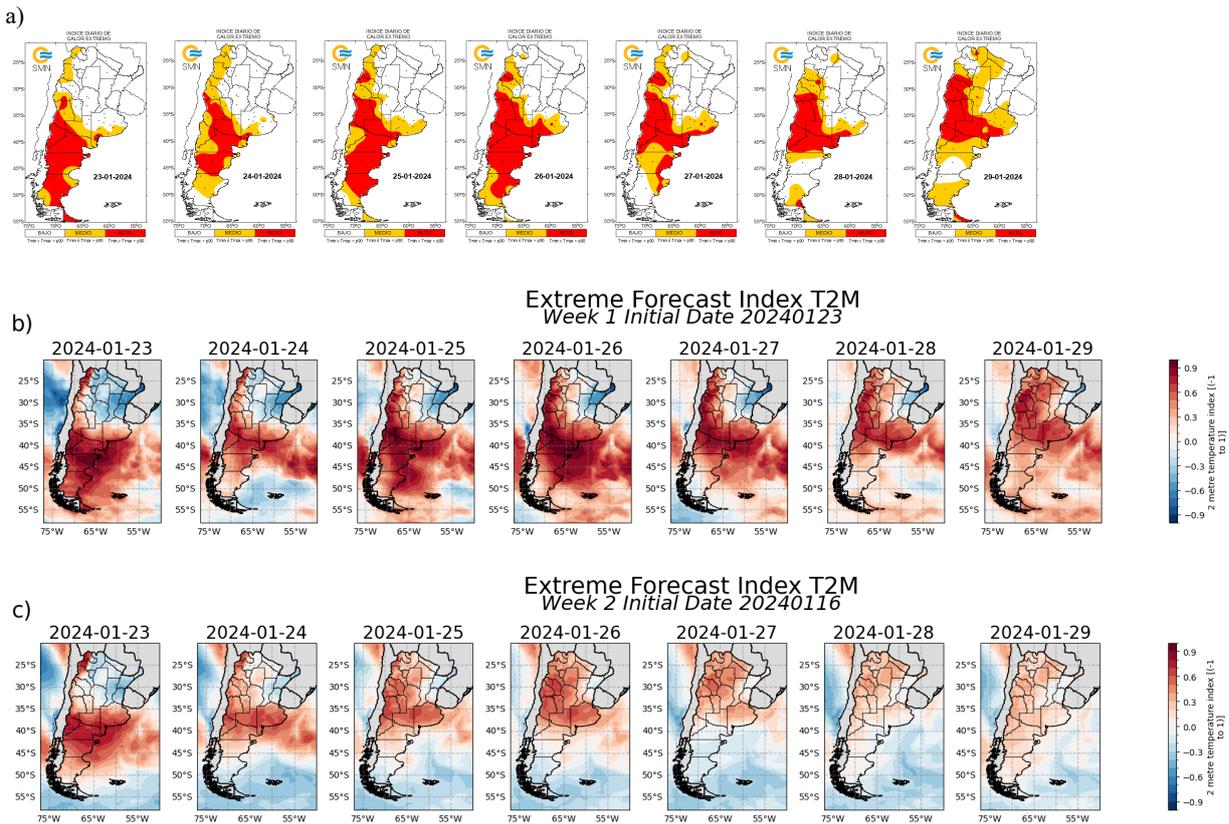


Figure 30: (a) Evolution of interpolated observed daily Extreme Heat Index from 23-29 January 2024 in Argentina. Red areas indicate regions where both Tmax and Tmin are larger than the 90th percentiles, orange areas where either Tmax or Tmin is larger than the 90th percentile, and white indicates both variables are smaller than the 90th percentiles. Below are shown evolutions of daily Extreme Forecast Index (EFI) of 2-metre temperature forecasts from medium range ECMWF model for the target week from 23-29 January 2024, initialized one week (b) and two weeks (c) prior the event. Note that these EFIs are obtained from the operational ENS ensemble.

Figure 31a (right panel) shows the 2-metre temperature weekly mean EFI for the first week of the heat wave event, from January 23 to 29, 2024. One week prior to the event (Figure 30a, right panel) the predicted EFI values were greater than 0.5 in the centre and northwest of Argentina while being negative in the northeastern area. Most stations that reached heatwave conditions (Figure 31b) during the targeted week are within the area with 2-metre temperature EFI weekly mean greater than 0.5. Two weeks prior to the event (Figure 30a, left panel), EFI values were also positive, above 0.3 in the centre and northwest of Argentina. Areas of high EFI also include the stations with heatwave, with positive but less intense EFI values in Patagonian stations.

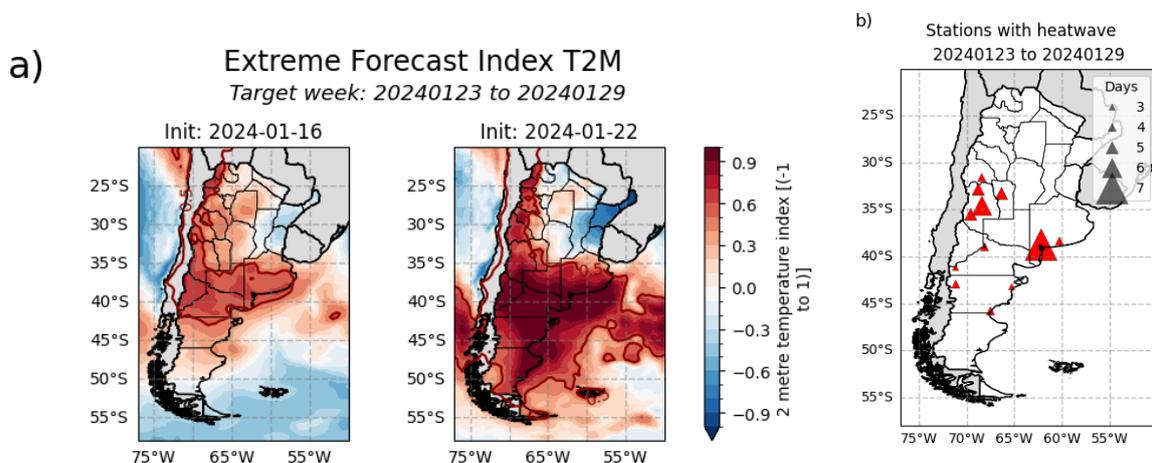


Figure 31: a) Weekly mean 2-metre temperature Extreme Forecast Index (EFI) for the target week initialized one day (right panel) or eight days (left panel) prior the beginning of the event. The red contour line indicates areas with values greater than 0.5. Note that these EFIs are obtained from the sub-seasonal ensemble. b) Observed stations that meet the criteria of heatwave event during that week. Only stations with 3 or more days in heatwave condition are considered.

From sub-seasonal forecasts (Figure 32a) it can be seen that the Cuyo region (Region 1, see Figure 32b) had EFI temperature values of 0.2 to 0.4 up to 2 weeks before the beginning of the target week. Similarly, the north of Patagonia (Region 3) presented EFI temperature values between 0.1 and 0.4 in that week, with a slight drop during 10 and 11 of January. At the same time, forecasts initialized from 9 till 14 of January predicted smaller EFI values for the southern half of Buenos Aires (Region 2), ranging from less than 0 to 0.4. A week prior to the beginning of the event, the forecasts of Region 1 increased to almost 0.6 on 15th of January. By 18 January all areas had predicted EFI values greater than 0.5 and continued to increase with shorter lead times. In brief, initializations showed increasing EFI values 21 to 15 days before the target week, from 0 to above 0.4 in all three strategic areas. This indicates that the sub-seasonal ensemble could have provided an early signal to monitor this heat wave event.

Both medium-range and sub-seasonal ensemble forecasts were able to predict positive 2-metre temperature EFI for the first week of the heatwave event with at least 2 weeks of anticipation.

Moreover, even with forecasts initialized ten days prior to the event, the area of the highest predicted EFI values matches with the location of high daily extreme heat index. Notably, the sub-seasonal ensemble forecast indicates higher EFI 2-metre temperature values in northern Patagonia than the medium-range. Thus, both models should be used to complement each other, not only for different time leads, but for different performance per region as well.

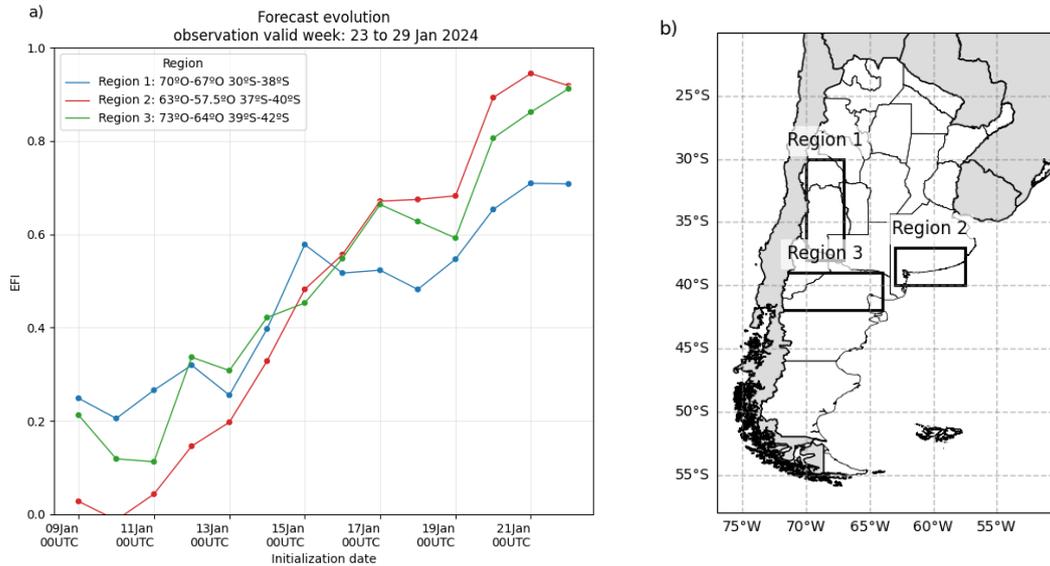


Figure 32: Panel (a) shows the evolution of weekly mean 2-metre temperature Extreme Forecast Index (EFI) for three regions in central Argentina obtained from the sub-seasonal ECMWF forecast. The abscissa indicates the start day of the forecast, all of which are valid for the week of the event in the respective areas. Panel (b) presents the three regions where EFIs latitude weighted average is computed.

Within South America CPTEC/INPE produces regularly (every week) numerical sub-seasonal forecasts using the BAM-1.2 model (Guimarães et al., 2019; Guimarães et al., 2021; Coelho et al., 2021; Coelho et al., 2022). Forecasts are calibrated using retrospective data from 1999–2018. The NOAA’s CPC Unified Temperature dataset is used as the observational reference. Figure 33 shows forecasts issued on January 17 and 24 January for the 24–30 January 2024 week, offering one to two weeks of lead time. The January 17 forecast already suggested upper tercile temperatures as the most likely outcome in central Argentina. The January 24 forecast increased the probability and spatial extent for occurrence of above upper tercile temperatures. The model demonstrated the ability to anticipate this extreme temperature event.

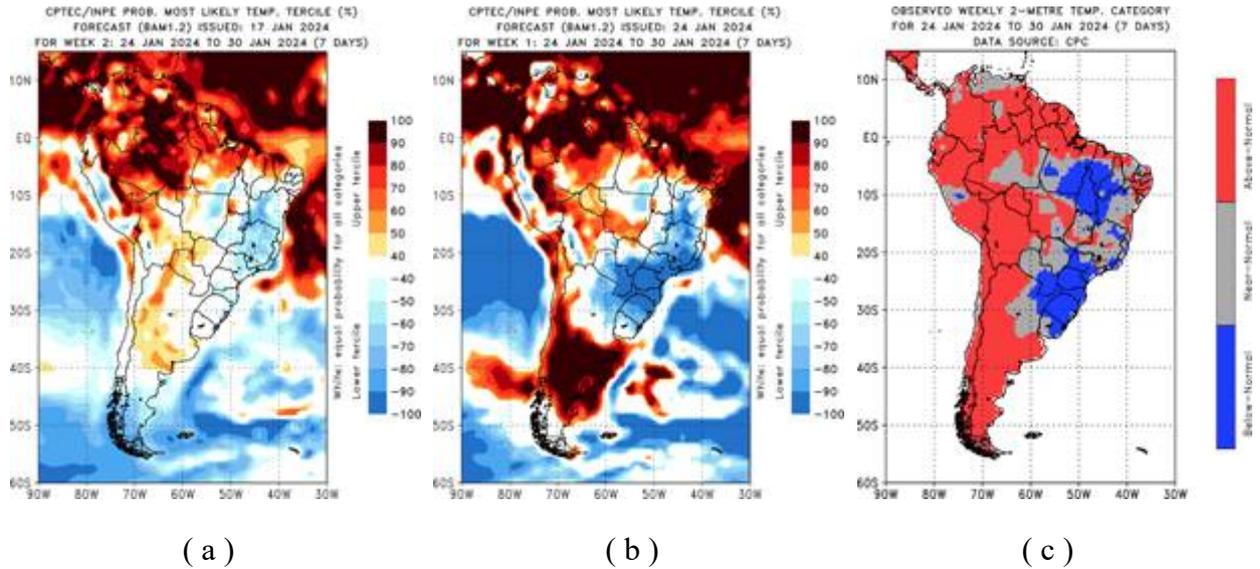


Figure 33: Calibrated numerical subseasonal most likely tercile probability forecasts for the target week 24-30 January 2024 produced with CPTEC/INPE BAM-1.2 model a) issued on 17 January 2024, b) issued on 24 January 2024, and c) observed tercile category based on Climate Prediction Center (CPC) Global Unified Temperature dataset of the National Oceanic and Atmospheric Administration (NOAA).

5 Conclusions

The results presented in this work are possibly the first documented attempt to analyse the performance of operational IFS forecasts over South America. This is important for two reasons: first as a source of objective information on ECMWF forecast capabilities over South America for local forecasters, and second for the developers of the model as a testing ground on which to evaluate the IFS potential and limitations.

First, an overarching analysis evaluating summary verification statistics was carried out for upper air variables by comparing forecasts with their own analysis, and at the surface by comparing surface forecast fields with SYNOP data.

Subsequently, in order to get a sense of the model performance in specific cases, three HIW events with very different characteristics were analysed: a convective windstorm causing severe damage and loss of life, a strong Zonda wind case, and a prolonged heat wave.

The general verification on upper air variables showed that the model performance over South America was comparable to that over Europe. This result was somewhat surprising for the participants in the pilot project, given the landmass/ocean ratio in the southern hemisphere and the sparseness of soundings over South America and the Southern Pacific upstream of the continent (especially its southern tip, see WIGOS Data Quality Monitoring System at https://wdqms.wmo.int/gbon/land_upper-air/). However, this low expectation was probably misplaced since forecast performance in both hemispheres at 500hPa has been comparable since the first decade of this century (see lead time of the anomaly correlation coefficient (ACC) at https://charts.ecmwf.int/products/plwww_m_hr_ccaf_adrian_ts?single_product=latest).

Surface verification generally shows low bias, although there are regions where the results are less favourable, such as the Amazon and the Andes. Outside these regions, the performance is

also comparable to that obtained for Europe. The three HIW case studies presented were designed to test the capabilities of the model in representing complex weather phenomena.

The convective case was a double test in the sense that not only the predictability of the general weather system was checked, but also the forecast's ability to reproduce the hazards (in particular wind gusts). The results based on the evolution of the CAPE and CAPE-shear EFIs show a very useful level of predictability for the weather system and a clear potential for direct use as an early warning tool. Clues about the severity of the event were already present 6 days in advance. However, the forecast failed to suggest the presence of dangerous wind gusts. Given the scale of the phenomena compared with the model resolution, this was not surprising. The higher resolution operational limited-area forecast from the SMN was able to capture the wind gust hazards and was the key for forecasters to issue the orange warning with more conviction (see Saucedo et al. 2024).

The Zonda wind case study was an attempt to test the model performance for a phenomenon induced by a very high, narrow topographic barrier. Overall, the representation of the case was very realistic while the predictability was good enough to allow for an effective early warning. For this task, the CTRL (ex HRES) was more effective than a higher resolution WRF and comparable to different versions of IFS at higher resolution. Despite the overall remarkable success, all models failed in reaching the very low humidity level observed at the station --a difference of at least 25 degrees of dewpoint temperature. It is important to keep in mind that this difference is less impressive when expressed in terms of relative humidity (just a few percentage points).

The study of the heat wave was centred on the predictability of the phenomenon. Results show that both medium-range and sub-seasonal forecasts of the IFS captured the entire evolution quite well and with a reasonable lead time. In addition, the area of the highest predicted EFI values matched well with the location of high daily extreme heat index, which suggests it could be a valuable tool for direct use in early warning for heat waves. In order to further this analysis, multiple heat wave events should be studied to evaluate the model's performance on both timescales.

This work is a first but important step towards better understanding of the performance of global models over South America. In this study we have analysed in particular the performance of the ECMWF's CTRL (ex HRES), future studies could consider other global forecasting systems. For a more complete evaluation, ensemble forecast and machine-learning forecasts (e.g. AIFS) could be assessed, and a verification of extremes as well as conditional verification (e.g. stratified by flow regimes) could be performed. In addition, verification of the ECMWF's sub-seasonal forecast will complement the work presented here on heat waves. These are potential topics for further collaboration.

In the near future (from 2026 onwards), this continent, as well as Africa, will be part of the WIPPS standardized verification score exchange (WMO 2023) between global numerical weather prediction centres. It is hoped that this WMO coordinated activity will intensify the interaction between the South American and the global numerical weather prediction communities, thus accelerating the achievements of the objectives of Early Warning for All in the region.

References

- Bouallègue, Z. B., F. Cooper, M. Chantry, P. Düben, P. Bechtold, and I. Sandu, 2023: Statistical Modeling of 2-m Temperature and 10-m Wind Speed Forecast Errors. *Mon. Wea. Rev.*, 151, 897–911, <https://doi.org/10.1175/MWR-D-22-0107.1>.
- Chesini F, Herrera N, Skansi MLM, Morinigo CG, Fontán S, Savoy F, Titto E. Mortality risk during heat waves in the summer 2013-2014 in 18 provinces of Argentina: Ecological study. *Cien Saude Colet.* 2022 May;27(5):2071-2086. doi: 10.1590/1413-81232022275.07502021.
- Coelho, C. A. S., de Souza, D. C., Kubota, P. Y., Costa, S. M. S., Menezes, L., Guimarães, B. S., Figueroa, S. N., Bonatti, J. P., Cavalcanti, I. F. A., Sampaio, G, Klingaman, N. P., Baker, J. C. A., 2021: Evaluation of climate simulations produced with the Brazilian global atmospheric model version 1.2. *Climate Dynamics.* 56, 873- 898. <https://doi.org/10.1007/s00382-020-05508-8>
- Coelho, C. A. S., Baker, J. C. A., Spracklen, D. V., Kubota, P. Y. Souza, D. C., Guimarães, B. S., Figueroa, S. N., Bonatti, J. P., Sampaio, G., Klingaman, N. P., Chevuturi, A., Woolnough, S. J., Hart, N., Zilli, M., Jones, C. D., 2022: A perspective for advancing climate prediction services in Brazil. *Climate Resilience and Sustainability.* <https://doi.org/10.1002/cli2.29>
- Cutraró, F., S. Righetti, Y. García Skabar y M. Sacco, 2020: Implementación del sistema de pronóstico numérico en el HPC: Calibración de temperaturas pronosticadas. *Nota Técnica SMN 2020-81.*
- ECMWF, 2022: ECMWF forecast user guide. ECMWF Tech. Rep., <https://confluence.ecmwf.int/display/FUG/Forecast+User+Guide>
- Gascón Salvador E, Maier-Gerber M, Vannière B, Zaplotnik Z, Becker T, Magnusson L, Chevallier M, Sandu I, 2024: Evaluating km-scale simulations in Destination Earth. *Newsletter Number 181 - Autumn 2024*, 14-15.
- Guimarães, B.S., Coelho, C.A.S., Woolnough, S.J., Kubota, P.Y., Bastarz, C.F., Figueroa, S.N., Bonatti, J.P., Souza, D.C., 2019: Configuration and hindcast quality assessment of a Brazilian global sub-seasonal prediction system. *Quarterly Journal of the Royal Meteorological Society.* Vol 146 (728). Pages 1067-1084. <https://doi.org/10.1002/qj.3725>
- Guimarães, B. S., Coelho, C. A. S., Woolnough, S. J., Kubota, P. Y., Bastarz, C. F., Figueroa, S. N., Bonatti, J. P., de Souza, D. C., 2021: An inter-comparison performance assessment of a Brazilian global sub-seasonal prediction model against four sub-seasonal to seasonal (S2S) prediction project models. *Climate Dynamics.* 56. Pages 2359-2375. <https://doi.org/10.1007/s00382-020-05589-5>
- Haiden, T., M. Janousek, F. Vitart, Z. Ben Bouallegue, L. Ferranti, F. Prates, and D. S. Richardson, 2021: Evaluation of ECMWF forecasts, including the 2021 upgrade. ECMWF Technical Memorandum No. 880, 54pp.
- Hewson T., 2021: Use and Verification of ECMWF products in Member and Co-operating States. Technical Memorandum No. 885, 33pp.

Magnusson, L., I. Tsonevsky, T. Hewson, 2017: Two storm forecasts with very different skill, ECMWF Newsletter, 154, 4–5.

Matsudo C., Y. García Skabar, 2023: Verificación de los pronósticos del Sistema de Asimilación y Pronóstico numérico del SMN para el período 2020-2022. Nota Técnica SMN 2023- 145.

McTaggart-Cowan, R., and Coauthors, 2019: Modernization of atmospheric physics parameterization in Canadian NWP. *J. Adv. Model. Earth Syst.*, 11, 3593–3635, <https://doi.org/10.1029/2019MS001781>.

NCEP, 2015: NCEP GFS 0.25 Degree Global Forecast Grids Historical Archive. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. <https://doi.org/10.5065/D65D8PWK>. Accessed 14 02 2025. updated daily.

Otero, F. and D. Araneo, 2020: Zonda wind classification using machine learning Algorithms. *International journal of climatology*, 41, E342–E353.

<https://rmets.onlinelibrary.wiley.com/doi/epdf/10.1002/joc.6688>

Righetti, S., F. Cutraro, Y. García Skabar y M. Sacco, 2024: Calibración de los pronósticos horarios de magnitud del viento. Nota Técnica SMN 2024-162.

Rodwell, M. J., D. S. Richardson, T. D. Hewson, and T. Haiden, 2010: A new equitable score suitable for verifying precipitation in numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, 136, 1344–1363

Roebber, P.J., 2009: Visualizing Multiple Measures of Forecast Quality. *Wea. Forecast.* 24, 601–608.

Salio, P., and Coauthors, 2024: Toward a South American High-Impact Weather Reports Database. *Bull. Amer. Meteor. Soc.*, 105, E1204–E1217,

<https://doi.org/10.1175/BAMS-D-23-0063.1>.

Saucedo, M., R. de Elía, L. Vidal, C. Matsudo, S. Righetti, A. Cejas, D. Diaz, N. Pérez, P. Hobouchian, R. Vasques Ferro, C. Ribero, M. Patanella, P. Irurzun, H. Bechis, D. D'Amén, P. Salio, Y. García Skabar, S. Pérez, C. Sánchez, C. Fiol, R. Vidal, 2024: Destructive wind events over central Argentina during December 16 and 17, 2023. Nota Técnica SMN 2024-191. <https://repositorio.smn.gob.ar/handle/20.500.12160/2890>

Saucedo, M., R. de Elía, L. Vidal, C. Matsudo, S. Righetti, A. Cejas, D. Diaz, N. Pérez, P. Hobouchian, R. Vasques Ferro, C. Ribero, M. Patanella, P. Irurzun, H. Bechis, D. D'Amén, P. Salio, Y. García Skabar, S. Pérez, C. Sánchez, C. Fiol, R. Vidal, F. Bertoli, 2025: Warning Value Chain Questionnaire for Argentina Diciembre 17, 2023. <https://docs.google.com/document/d/1Q4Wpld4BMGaacPadBHPhHwanSRukoibTc/edit?tab=t.0>

Suli, S., Barriopedro, D., García–Herrera, R. & Rusticucci, M. 2023 Regionalisation of heat waves in southern South America. *Weather Clim. Extrem.* 40, 100569.

Tsonevsky I. 2015: New EFI parameters for forecasting severe convection. ECMWF Newsletter No. 144.

<https://www.ecmwf.int/sites/default/files/elibrary/2015/17324-new-efi-parameters-forecasting-severe-convection.pdf>

Tsonevsky, I., C. A. Doswell, and H. E. Brooks, 2018: Early Warnings of Severe Convection Using the ECMWF Extreme Forecast Index. *Wea. Forecasting*, 33, 857–871.

Walters, D., and Coauthors, 2019: The Met Office unified model global atmosphere 7.0/7.1 and JULES global land 7.0 configurations. *Geosci. Model Dev.*, 12, 1909–1963, <https://doi.org/10.5194/gmd-12-1909-2019>.

Wedi et al., 2025. Implementing Digital Twin technology of the Earth System in Destination Earth Journal: Journal of the European Meteorological Society (under review).

WMO, 2022: EARLY WARNINGS FOR ALL. The UN Global Early Warning Initiative for the Implementation of Climate Adaptation Executive Action Plan 2023-2027. Geneva, 56 pp.

WMO, 2023: Manual on the WMO Integrated Processing and Prediction System. WMO-No. 485, 166 pp.

Zipser, E. J., D. J. Cecil, C. Liu, S. W. Nesbitt, and D. P. Yorty, 2006: Where are the most intense thunderstorms on earth? *Bull. Amer. Meteor. Soc.*, 87, 1057–1072, <https://doi.org/10.1175/BAMS-87-8-1057>.