

Global high-resolution models vs. limited-area models: verification of HRES-IFS and WRF-ARW forecasts over Catalonia

G. Montané¹ and B. Codina¹

¹Universitat de Barcelona

Received: 22-I-2020 – Accepted: 06-XI-2020 – **Original version**

Correspondence to: gimonpin@gmail.com

Abstract

Over recent decades, numerical weather prediction (NWP) has undergone big improvements thanks, in part, to the increase in the computational power available to run forecast models. This has led global weather models to currently run with grid-cell sizes similar to those used in most of the traditionally used limited-area models (LAMs). One example of this is the ECMWF high-resolution deterministic model, the HRES-IFS, which currently runs operationally with a horizontal grid resolution of 9 km. Under these premises, some questions can be raised: does it make sense to continue running LAMs for operational weather forecasting? Is it enough to use the output of a global model at 9 km for regional or even local forecasts? In this preliminary study, the outputs of some simulations performed with the WRF model over Catalonia for January and July 2018 are compared with the HRES-IFS operational forecast for the same dates.

Key words: Numerical Weather Prediction, high-resolution IFS, WRF, forecast verification

1 Introduction

Numerical weather prediction (NWP) has had to overcome several challenges in different areas to achieve increasingly reliable forecasts. We can mainly differentiate between scientific and technological challenges: the modelling of the physical and dynamic processes that describe the behavior of the atmosphere, the definition of a consistent atmospheric initial state from a set of observations, the parameterization of certain physical processes, the availability of computational capacity to resolve the primitive equations using numerical methods, etc. (Bauer et al., 2015).

One key factor that limits the quality of the forecasts is the grid-cell size used by weather models when discretizing the space. Due to the high computational cost of running high-resolution simulations, global models used to run with coarse grids. This was traditionally solved by using limited-area models (LAMs), which allow the run of mesoscale simulations in a higher resolution by reducing the forecast area. Thus, weather phenomena on a smaller scale than in global models can be represented and more detailed forecasts can be obtained at a not so high computational cost.

In Catalonia, the use of mesoscale models dates back to the '90s. The first LAM used was the MASS (Mesoscale Atmospheric Simulation System) model which was able to run with grid-cell sizes between 10 and 100 km allowing nested domains (Codina et al., 1997). Since then, this kind of model has been used in operational forecasting. Specifically, the Meteorological Service of Catalonia (Servei Meteorològic de Catalunya, SMC) has been using the WRF (Weather Research and Forecasting) model operationally for several years now to provide weather forecasts for the region. The WRF model is one of the most used LAMs all over the world and it was developed by the National Center for Atmospheric Research (NCAR) (Skamarock et al., 2008). Nevertheless, mostly thanks to the great technological advances and to the increase of the computational power available in recent decades, global models are currently able to run with very fine grids, which involves a bigger number of calculations but also more detailed forecasts. One example of a high-resolution global model is the HRES (High-RESolution) version of the IFS (Integrated Forecast System) model from the European Centre for Medium-Range Weather Forecasts (ECMWF), which has been running operationally since 2016



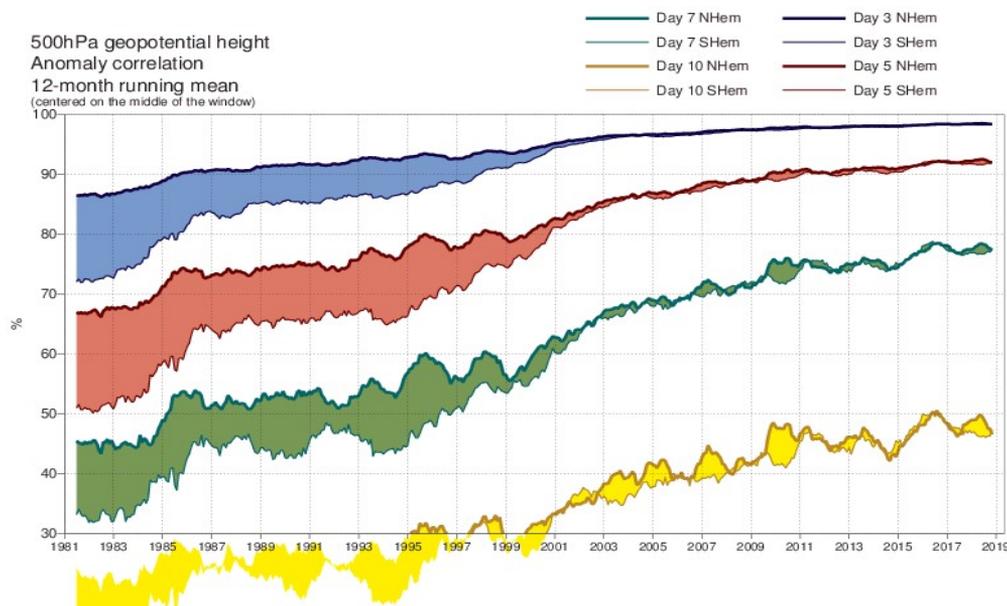


Figure 1: Evolution of the forecasting capability of the ECMWF’s IFS model shown through the Anomaly Correlation (AC) coefficient. It is considered that forecasts are useful if they have an AC score of above 60%, while a score of above 80% indicates a high degree of accuracy. Source: ECMWF.

using a grid-point space of 9 km (Malardel et al., 2016), a resolution commonly used in LAMs.

The IFS is the operational model used by the ECMWF since 1994, and its different components (resolution, data assimilation process, etc.) have been continuously updated and improved since then. These successive improvements, as well as the use of more and more powerful supercomputers, have led to a huge improvement in the quality of its forecasts. Fig. 1 shows the evolution in time of the Anomaly Correlation (AC) coefficient for the geopotential height at 500 hPa. Currently, the forecasts for 3, 5 and 7 days have scores of above 95, 90 and 75% respectively, which show the great capability of the model to forecast anomalies with respect to the climate average.

Since the update in March 2016 (IFS cycle 41r2), the HRES-IFS model has been using an octahedral Gaussian grid with 1,280 latitudinal lines between the equator and the poles, resulting in an equivalent grid-point space of about 9 km at mid latitudes (Malardel et al., 2016; Owens et al., 2018). Besides its high resolution, it is run operationally as a hydrostatic model using parameterizations to represent small-scale physical processes. It also uses the 4D-Var data assimilation method and gets its input from multiple observation sources (ECMWF, 2018).

Having said all this, and considering the current trend of reducing the grid size in global weather models more and more, we may pose questions such as whether it makes sense to run LAMs operationally while having high-resolution global models that are currently run with the same grid-cell size?

The aim of this piece of work is to give an idea of the current status of global high-resolution weather models and

their operational use. Applied to the specific case of Catalonia, we try to find out if it currently adds any value to use LAMs like WRF for local or regional forecasting or if it is enough to use high-resolution global models like HRES-IFS.

In the following sections, the methodology used is explained, the results of the experiments performed are discussed, and the conclusions of the entire work are shown.

2 Methodology

This work mainly consists of the verification of the WRF-ARW (Advanced Research WRF) and HRES-IFS operational forecasts for January and July 2018 in the area of Catalonia. In both cases, the daily 36-hour forecasts initialized at 00 UTC were verified.

In the case of the HRES-IFS model, the data from the operational forecasts for both months were obtained from the Spanish Meteorological Agency (Agencia Estatal de Meteorología, AEMet), which is the representative partner of the ECMWF organization in Spain. Specifically, the forecasts for the following meteorological variables were obtained:

- Temperature at 2 meters
- Relative Humidity at 2 meters
- Wind Speed at 10 meters
- Accumulated precipitation in 24 hours
- Temperature at 850 hPa

The interpolated lat/lon grid obtained from the IFS outputs (with a grid-point space of 0.1°) is shown in Fig 2.

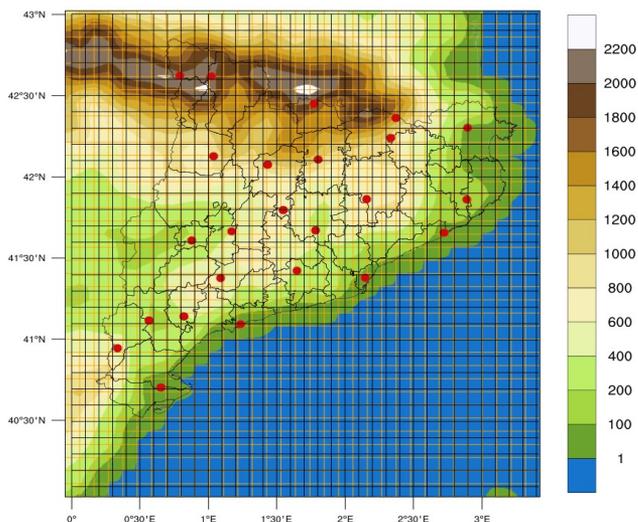


Figure 2: Domain grids used in WRF (orange) and IFS (black) with the selected stations (red dots) and the topography from the WRF model.

Below, the methodology used in the WRF-ARW simulations and in the verification is described.

2.1 Simulations

The simulations with the WRF-ARW model v3.9.1.1 (Wang et al., 2017) were done using a domain that covers Catalonia with a grid-cell size of 9 km. The boundary and initial conditions were taken from ERA5, the ECMWF reanalysis (Hersbach et al., 2018) that has provided data since 1979 with a grid-point space of about 30 km.

Simulations were run for every day during January and July 2018, and they were initialized at 00 UTC to obtain 36-hour forecasts. The model was executed in non-hydrostatic mode, using the parameterization and configuration used by the SMC in its operational forecasts. Specifically, the Kain-Fritsch convective scheme and the WSM5 microphysics scheme were used (Mercader et al., 2010; Cáceres León et al., 2018).

In Table 1, the domain configuration used to run WRF is shown.

2.2 Verification

The output of both models was compared to observational data from the two simulated months. To verify surface variables, this data was obtained from 24 stations of the SMC’s Automatic Weather Stations Network (Xarxa d’Estacions Meteorològiques Automàtiques, XEMA). These stations were selected manually trying to obtain a good representation of the whole territory and to choose stations that are representative of their geographical environment, in keeping with the topography of the model. The selected stations are

Table 1: Domain configuration and parameterizations used in the WRF-ARW simulations.

Parameter	Selected option
DX, DY	9 km
Vertical levels	31
Domain dimensions (WE, SN)	38, 40
Time step	54 s
Initial and boundary conditions	ERA5
Pressure top	50 hPa
Convective scheme	Kain-Fritsch
Microphysics scheme	WSM5
SW radiation scheme	Dudhia
LW radiation scheme	RRTM
Surface layer	MM5
Land surface	Noah Land Surface Model
PBL scheme	YSU
Urban surface	Off

shown in Fig. 2, where the model’s grid and topography are also represented. Regarding temperature at 850 hPa, forecasts were compared to data from the Barcelona radiosonde, which is launched at 00 and 12 UTC every day.

In every case, the results of the simulations were compared with the observations at 12 UTC of day d+1 (36-hour forecast). For the precipitation, the 24-hour accumulation was used (from 12 UTC of day d+0 to 12 UTC of day d+1). The first 12 forecast hours were discarded to avoid spin-up problems, as a model typically takes a few hours to adjust initial conditions to its grid, stabilize them and develop precipitation (Sun et al., 2014).

The verification of surface temperature, relative humidity and wind speed was done by calculating two indices: the Mean Error (ME), which allows us to calculate the bias (positive or negative) of forecasts regarding real observations, and the Root Mean Square Error (RMSE), which shows the typical magnitude of the forecast errors (Collaboration for Australian Weather and Climate Research, 2017). These indices were calculated for every selected station, taking the forecasted values from the closest grid point to every station and treating each month separately to see the differences in the forecast capabilities between two different seasons: winter and summer.

The ME and RMSE were also calculated to verify temperature at 850 hPa in the two months in question, but in this case taking the values from the 4 grid points closest to the point where the radiosonde is launched (the Faculty of Physics at the University of Barcelona). Once both indices had been calculated for each of the points, the best of them was considered as the result of the verification for each of the models and months.

Precipitation was verified using the fuzzy verification method (Ebert, 2008). Obtaining successful forecasts in certain convective precipitation episodes is quite difficult, so this method considers a window of grid points around each station to avoid double penalization in case of spatial errors. Thus, a certain tolerance to forecast errors is given, trying to

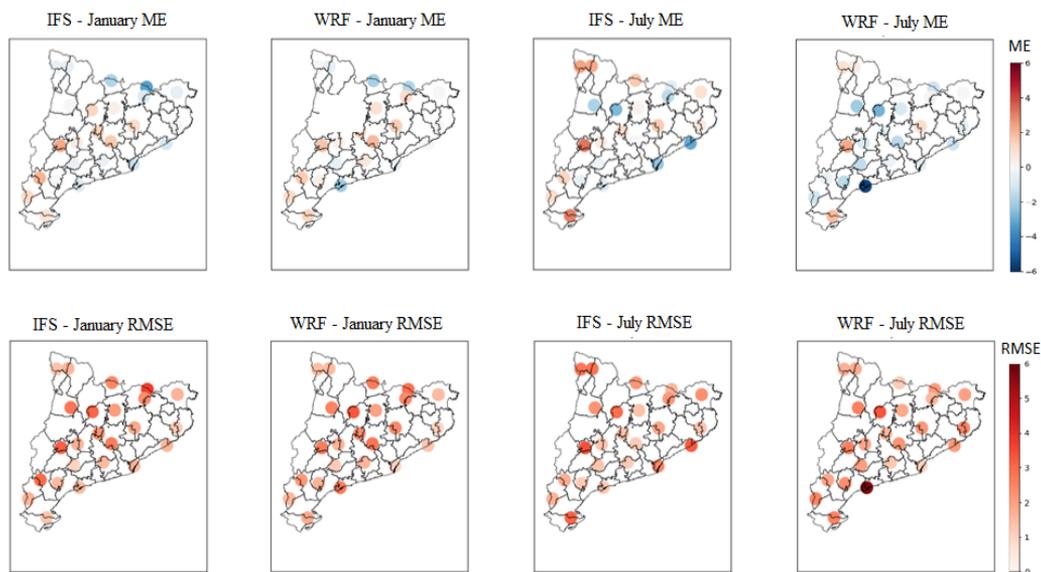


Figure 3: Representation of ME and RMSE for surface temperature (in °C) obtained for both models in each of the stations in question in January and July 2018.

find an approximate agreement between forecasts and observations instead of an exact match in space.

In this work, a 4 grid-point window around each station was used. For each window, the forecast of the grid point with the smallest error compared with the observed values was considered. Four accumulated precipitation ranges were considered in the verification process (in all of them, the first value is not included while the second one is): 0-1, 1-5, 5-20 and more than 20 mm in 24 hours.

A contingency table was created for each threshold which allowed us to determine the concordance between accumulated precipitation forecasts and observations. Using this contingency table, the following metrics were calculated (Collaboration for Australian Weather and Climate Research, 2017):

- **POD (Probability of Detection):** also known as hit rate, indicates the percentage of forecast hits regarding the number of total observed events (ideal value: 1).
- **FAR (False Alarm Ratio):** indicates the percentage of forecasted events that have not been actually observed (ideal value: 0).
- **FBI (Frequency BIAS):** indicates the total bias, calculating the relationship between the total number of forecasted events and the total number of observed events (ideal value: 1).
- **CSI (Critical Success Index):** gives us the relationship between the correctly forecasted events and the total number of forecasted and observed events (ideal value: 1).

Additionally, an ROC (Receiver Operating Characteristic) graph was generated to help determine which model

Table 2: ME and RMSE (in °C) of the 36-hour forecasts for temperature at 850 hPa obtained with the WRF-ARW and HRES-IFS models for January and July 2018 in Barcelona. Best results are in bold.

	ME		RMSE	
	WRF	IFS	WRF	IFS
January	-0.30	-0.67	1.62	1.36
July	0.03	-0.07	1.66	0.96

better forecasts precipitation. This graph consists of representing POD vs. POFD (Probability of False Detection or False alarm rate: the relationship between false alarms and non-observed events) and shows how well a weather model detects an event (i.e., its usefulness) (Fawcett, 2006). The higher the POD and the lower the POFD (i.e., the closer to the upper left-hand corner of the graph a point is), the better the result will be.

3 Results and discussion

In this section the results obtained in the verification of both HRES-IFS and WRF forecasts are shown. First, temperature at 850 hPa is analyzed, then all surface variables and finally the results for precipitation are shown.

3.1 Temperature at 850 hPa

As explained in the previous section, temperature at 850 hPa was verified considering the values from the 4 closest grid points to the Barcelona radiosonde launch point - i.e., the Faculty of Physics at the University of Barcelona. Table 2 shows the results obtained for the ME and RMSE indices.

We can see that WRF obtained the best results in both months for the ME, but it also obtained the worst results for the RMSE. This is probably due to a cancellation effect

Table 3: ME and RMSE (in °C) for surface temperature obtained with WRF-ARW and HRES-IFS in January and July 2018 for each of the stations in question. Best results are in bold. The last two rows show, respectively, the total number of stations in which each model obtained the best result for each index and month and the total mean value of each index and month considering all stations.

Station	ME				RMSE			
	January		July		January		July	
Amposta	1.32	0.81	1.86	3.01	1.82	1.27	2.55	3.1
Ascó	1.52	2.02	-0.42	1.35	2.15	2.99	1.87	1.75
Barcelona - el Raval	-0.44	-1.28	-0.57	-2.26	1.02	1.55	0.97	2.43
Bonaigua (2.266 m)	-0.5	-0.45	0.47	2.25	1.48	1.61	1.75	2.83
Cabanes	0.03	-0.51	0.02	0.91	1.52	1.63	1.98	2.31
Cassà de la Selva	-0.11	-0.13	-0.53	0.61	1.04	1.18	2.47	1.32
Falset	0.36	-0.08	-1.55	-0.47	1.66	1.5	2.33	1.16
Font-rubí	0.62	-0.42	-0.35	-0.15	1.6	1.51	1.22	1.17
Horta de Sant Joan	1.16	1.14	-1.14	1.1	1.54	1.63	2.47	1.66
l'Espuga de Francolí	-0.49	-0.44	-1.52	-0.36	1.15	0.99	2.03	1.31
Lac Redon (2.247 m)	0.14	-0.38	0.86	2.36	1.45	1.35	1.7	2.78
Lladurs	1.32	1.29	-2.82	-2.63	3.3	3.08	3.31	2.92
Malgrat de Mar	-0.09	-0.94	-1.19	-3.07	1.27	1.58	2.1	3.24
Malniu (2.230 m)	-1.97	-1.8	-0.04	1.63	2.64	2.43	1.19	2.14
Mollerussa	1.65	2.27	2.29	3.2	2.63	3.15	2.6	3.41
Molló – Fabert	-1.74	-3.28	-1.21	-0.59	2.9	3.72	2.08	1.59
Muntanyola	1.54	1.12	1.04	1.5	2.42	1.83	2.13	1.87
Pinós	1.3	1.35	-0.54	0.99	2.35	2.16	1.5	1.32
Sant Pau de Segúries	1.08	-0.84	0.18	-1.58	2.43	2.33	1.59	2.15
Sant Romà d'Abella	1.31	0.03	-2.13	-1.91	2.51	2.78	2.57	2.22
Sant Salvador de Guardiola	1.93	1.55	-1.62	-0.09	2.77	2.62	2.2	1.17
Santuari de Queralt	0.87	0.33	-0.92	0.24	1.96	2.05	1.85	1.34
Tarragona - Complex Educatiu	-2.18	-0.78	-5.57	-0.77	2.67	1.51	5.68	1.21
Tàrraga	0.35	0.5	-0.9	0.54	1.66	1.67	1.86	1.29
TOTAL	10	14	13	11	12	12	9	15
TOTAL MEAN	0.37	0.05	-0.68	0.24	2	2.01	2.17	1.99

between positive and negative errors and would show that, while the average of the errors obtained with WRF is smaller, these are bigger in magnitude than IFS errors. So, IFS seems to provide better forecasts than WRF for 850 hPa temperature.

It should also be noted that having obtained a slightly negative ME in both months shows that IFS tends to underestimate temperature at 850 hPa. This is also observed for WRF in January but not in July, when ME is near 0.

3.2 Temperature at 2 meters

Regarding temperature at 2 meters, similar results were obtained for both models when comparing the two verification indices both in January and July (Fig. 3 and Table 3).

If we take a look at ME, we see that in January both models tend to overestimate the temperature in the inland areas of Catalonia, while they slightly underestimate it in the coastal areas and in the Pyrenees. In July, however, WRF in general underestimates the temperature, while IFS is more balanced. Both models present similar results for RMSE in January, but in July, IFS generally obtains slightly better results.

3.3 Relative humidity at 2 meters

Fig. 4 and Table 4 show the results for the relative humidity at 2 meters. If we look at ME, we see that both models tend to underestimate relative humidity except in some stations in the Pyrenees and some coastal zones.

If we also consider the RMSE, we can see that WRF provides better results than IFS in July, while in January the opposite occurs regarding the number of stations where each model gets a better score. Despite this, the total mean of the RMSE values is slightly better for the IFS model.

A quite high RMSE value in WRF stands out in July for the Tarragona station, which could indicate that this station is not so representative of its environment considering the topography of the model, at least in the case of relative humidity. The Pyrenees and Pre-Pyrenees are the areas where both models obtain the worst results.

3.4 Wind speed at 10 meters

Regarding surface wind speed (Fig. 5 and Table 5), we can see that IFS obtains better results than WRF in almost all stations for both ME and RMSE, especially in January. Despite that, it seems to underestimate the real values in some cases in both months, while WRF tends to overestimate them. IFS also seems to have difficulties in forecasting wind speed in some mountain stations like in the Vall d'Aran.

Table 4: ME and RMSE (in %) for surface relative humidity obtained with WRF-ARW and HRES-IFS in January and July 2018 for each of the stations in question. Best results are in bold. The last two rows show, respectively, the total number of stations in which each model obtained the best result for each index and month and the total mean value of each index and month considering all stations.

Station	ME				RMSE			
	January		July		January		July	
Amposta	-3.97	-2.67	-7.36	-13.01	8.54	7.82	10.97	13.97
Ascó	-8.09	-10.64	1.53	-3.35	15.49	18.99	8.69	7.2
Barcelona - el Raval	-1.13	11.54	-1.83	12.73	8.8	13.38	7.85	14.82
Bonaigua (2.266 m)	18.79	5.69	14.96	2.82	26.52	21.69	21.82	15.84
Cabanes	-5.92	-0.6	-4.71	-10.35	10.22	7.18	10.88	14.37
Cassà de la Selva	-8.51	-4.28	0.89	-1.14	13.6	10.58	14.09	7.95
Falset	-3.05	-3.35	-0.42	-5.39	10.22	9.05	7.95	8.85
Font-rubí	-6.78	1.06	-3.38	-4.63	11.45	7.58	9.34	10.5
Horta de Sant Joan	-7.17	-6.56	-0.46	-9.14	9.57	10.34	11.32	12.88
l'Espuga de Francolí	-0.48	-0.5	-2.52	-6.5	6.95	7.68	7.96	9.89
Lac Redon (2.247 m)	16.81	5.4	8.77	-8.12	24.52	16.29	15.27	13.97
Lladurs	-15.33	-14.85	1.36	2.36	23.02	22.55	8.67	6.94
Malgrat de Mar	-6.09	2.09	-2.06	8.37	10.89	7.79	9.54	10.7
Malniu (2.230 m)	24.89	9.5	10.08	-0.67	28.76	17.84	15.4	11.43
Mollerussa	-14.12	-15.37	-15.46	-17.39	18.63	18.88	16.33	18.59
Molló-Fabert	8.2	5.8	3	-3.78	12.26	12.26	9.87	12.7
Muntanyola	-12.08	-7.33	-9.7	-9.78	17.17	11.83	14.7	12.24
Pinós	-17.14	-12.88	-6.35	-10.03	21.83	17.97	10.11	11.54
Sant Pau de Segúries	-5.21	-4.45	-0.52	-1.27	14.16	14.53	10.61	10.57
Sant Romà d'Abella	-11.45	-3.11	1.79	0.99	19.2	16.76	8.25	6.91
Sant Salvador de Guardiola	-15.94	-10.03	-7.51	-9.65	20.5	13.38	11.16	12.36
Santuari de Queralt	-2.75	1.47	0.16	-7.16	14.87	16.18	10.84	12.06
Tarragona - Complex Educatiu	15.89	4.2	33.72	-1.8	17.91	8	34.27	6.75
Tàrraga	-7	-6.01	-4.54	-8.3	11.66	9.89	8.87	10.67
TOTAL	5	19	19	5	7	17	14	10
TOTAL MEAN	-2.82	-2.33	0.39	-4.34	15.70	13.27	12.28	11.40

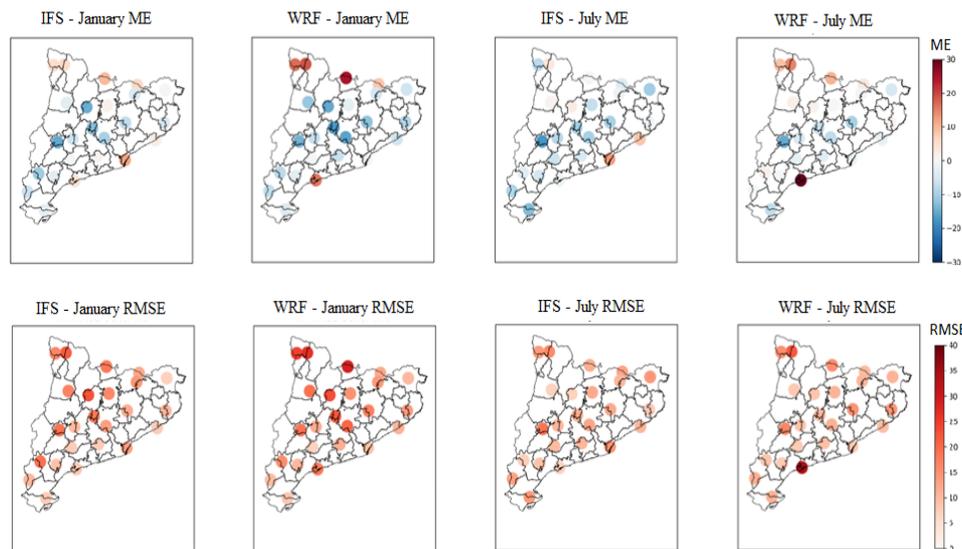


Figure 4: Representation of ME and RMSE for surface relative humidity (in %) obtained for both models in each of the stations in question in January and July 2018.

These results seem to be in keeping with some recent studies, such as Dutour Sikirić et al. (2015) and Hewson (2017), which suggest that the wind speed results obtained with some LAMs are not as good as the ones obtained with IFS.

3.5 Precipitation

Results for the different indices used in the precipitation verification are shown in Table 6. In January, IFS shows greater POD in all thresholds, which indicates that it is the model with better ability to detect precipitation. It stands out

Table 5: ME and RMSE (in m/s) for surface wind speed obtained with WRF-ARW and HRES-IFS in January and July 2018 for each of the stations in question. Best results are in bold. The last two rows show, respectively, the total number of stations in which each model obtained the best result for each index and month and the total mean value of each index and month considering all stations.

Station	ME				RMSE			
	January		July		January		July	
Amposta	1.4	-0.6	0.52	-0.05	2.55	2.06	1.04	0.94
Ascó	4.14	2.64	2.47	1.16	4.7	3.1	3.02	1.95
Barcelona - el Raval	0.97	0.41	1.24	1.39	1.86	1.61	1.46	1.65
Bonaigua (2.266 m)	2.54	-2.32	-0.26	-0.48	3.54	3.24	1.77	1.23
Cabanes	2.06	1.01	2.27	2.09	3.2	2.36	3.08	2.66
Cassà de la Selva	2.49	1.28	1.25	1.28	3.15	1.85	1.84	1.69
Falset	2.95	1.36	1.32	0.73	3.53	1.88	1.77	1.25
Font-rubí	1.36	-0.68	0.65	-0.13	3.03	2.24	1.43	1.13
Horta de Sant Joan	4.05	2.34	2.33	0.73	4.68	2.74	2.74	1.51
l'Espluga de Francolí	3.01	0.43	0.71	-0.24	3.49	1.5	1.45	1.15
Lac Redon (2.247 m)	1.79	-2.16	-3.92	-4.39	3.08	2.9	5.59	4.97
Lladurs	3.56	0.65	2.24	0.83	4.16	0.94	2.36	1.01
Malgrat de Mar	1.68	1.21	0.75	0.99	3.23	2.37	1.29	1.71
Malniu (2.230 m)	7.26	0.11	1.35	1.1	8.65	1.08	2.04	1.49
Mollerussa	1.64	1.49	1.3	0.59	2.46	2.08	1.75	1.32
Molló – Fabert	2.33	-1.29	0.38	-0.12	4.17	1.62	1.94	0.89
Muntanyola	2.24	0.05	0.39	0.17	3.23	1.39	1.26	0.88
Pinós	2.04	0.12	0.85	0.79	2.67	1.21	1.52	1.22
Sant Pau de Segúries	1.13	-0.18	0.6	0.75	2.31	1.44	0.9	1.22
Sant Romà d'Abella	1.98	-0.1	-0.55	-1.12	2.55	1.55	1.6	1.71
Sant Salvador de Guardiola	4.07	1.35	1.64	1.16	5.04	1.8	2.2	1.44
Santuari de Queralt	0.89	-1.2	-1.1	-1.61	2.62	3.33	1.39	1.89
Tarragona - Complex Educatiu	0.77	-0.3	0.76	1.21	2.14	2.02	1.3	1.53
Tàrraga	1.72	0.9	0.47	0.2	2.12	1.36	0.85	0.93
TOTAL	2	22	9	15	1	23	7	17
TOTAL MEAN	2.42	0.27	0.74	0.29	3.42	1.99	1.9	1.56

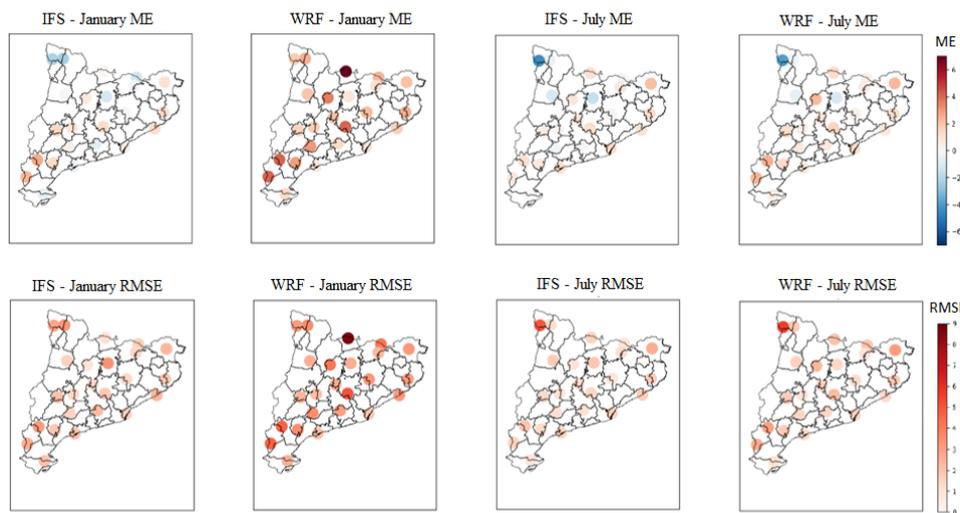


Figure 5: Representation of ME and RMSE for surface wind speed (in m/s) obtained for both models in each of the stations in question in January and July 2018.

in the detection of accumulated precipitation of more than 20 mm in 24 hours with a POD value of 0.93. Nevertheless, it is also the model which gives more false alarms in the same month, especially in the lowest thresholds. In July, WRF is the model that gives more false positives but it is also the one that better detects precipitation over 5 mm. This could be related to its capability to develop convection, and thus better

detect convective precipitation which is the most common one in summer in this region.

Both models obtain similar results regarding FBI and CSI. IFS obtains greater FBI in the 0-1 mm threshold, which means that it tends to overforecast precipitation between 0 and 1 mm. We can also see that CSI values are near or below 0.5 in almost all cases, meaning that both models correctly

Table 6: POD, FAR, FBI and CSI for the accumulated precipitation obtained by the WRF-ARW and HRES-IFS models for January (top) and July (bottom) 2018 for all the stations in question in the 0-1, 1-5, 5-20 and more than 20 mm thresholds. Best results are in bold.

January									
POD		POFD		FAR		FBI		CSI	
WRF	IFS	WRF	IFS	WRF	IFS	WRF	IFS	WRF	IFS
0.49	0.69	0.03	0.13	0.32	0.57	0.73	1.61	0.40	0.36
0.64	0.67	0.02	0.04	0.34	0.50	0.98	1.33	0.48	0.40
0.77	0.83	0.02	0.02	0.26	0.27	1.04	1.15	0.61	0.63
0.54	0.93	0	0	0	0.07	0.54	1.04	0.54	0.90

July									
POD		POFD		FAR		FBI		CSI	
WRF	IFS								
0.55	0.78	0.08	0.24	0.60	0.76	1.40	3.34	0.30	0.22
0.73	0.85	0.07	0.06	0.51	0.41	1.48	1.45	0.42	0.53
0.83	0.65	0.04	0.01	0.41	0.19	1.39	0.80	0.53	0.57
0.29	0.14	0	0	0	0	0.29	0.14	0.29	0.14

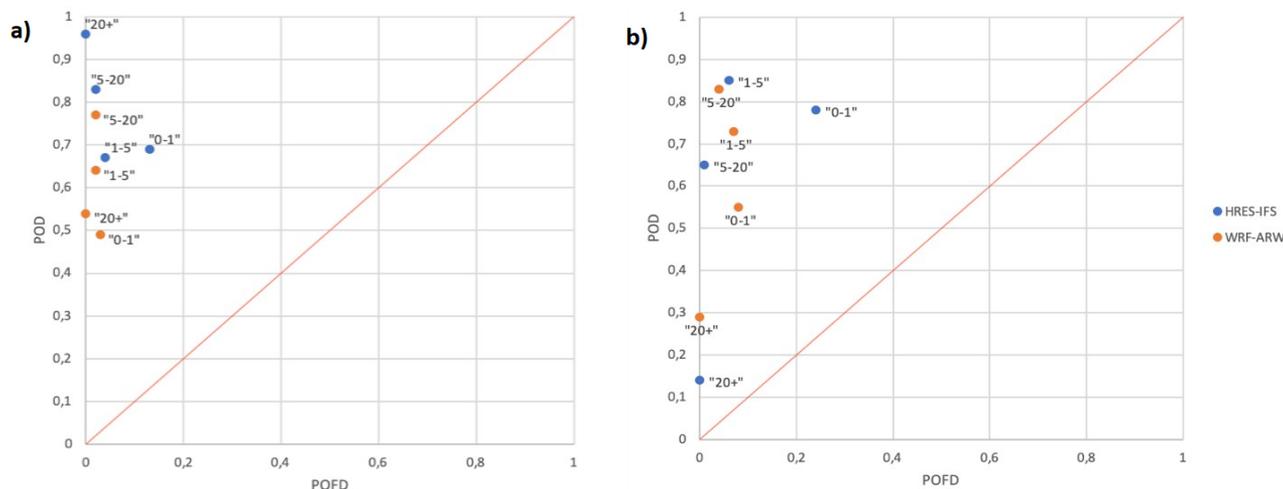


Figure 6: ROC graphs for the WRF and HRES-IFS forecasts of the different precipitation thresholds for (a) January and (b) July 2018. The diagonal line represents randomness: points in the upper-left half of this line are the ones that are useful and provide knowledge.

forecast a maximum of half of the events. Despite that, in January IFS obtains a CSI value of 0.9 for precipitation accumulations of more than 20 mm in 24 hours, which coincides with a high POD. This seems to confirm the IFS ability to detect this kind of precipitation, at least in January.

To see the behavior of both models regarding the precipitation forecast more clearly, two ROC graphs were drawn for each of the months in question (Fig. 6). In these graphs, the relationship between POD and POFD is shown for each threshold and model. We can see that all points are in the upper left-hand part of the graph regarding the so-called random guessing line (in red), a line along which it is considered that the results obtained are random. This means that both models provide some knowledge when forecasting precipitation in all of the ranges considered. If we look at the January graph, we can see that all IFS points are above the WRF ones, which means that IFS forecasts rain better in this month in all ranges. This is also true in July for the lowest precipitation ranges (0-1 and 1-5 mm), but in some cases, it obtains more false alarms than WRF. IFS also seems to

obtain worse results than WRF in July in the 5-20 and more than 20 mm thresholds, despite both models obtaining quite poor results in the latter threshold as they do not give false positives but they hit the forecasts on few occasions.

4 Conclusions

In this work, the 36-hour forecasts obtained by the HRES-IFS global model and the WRF-ARW mesoscale limited-area model for January and July 2018 were compared in the area of Catalonia. The objective was to investigate whether, using current knowledge and technology, the forecasts obtained with a high-resolution global model are better or equal than the ones obtained with a limited-area model. To do so, a set of selected variables were verified comparing them to observations: temperature at 850 hPa, temperature at 2 meters, relative humidity at 2 meters, wind speed at 10 meters and accumulated precipitation in 24 hours. The output of the IFS model was compared to the simulations

done with WRF at a resolution of 9 km.

These verifications showed that IFS results are of a similar quality to the ones obtained with WRF, even better in some cases. We found an example of these better results in the surface wind speed, in which IFS obtained better results in the ME and RMSE indices in almost all the stations in question. Another case in which IFS has obtained better results than WRF is the surface relative humidity in January 2018. Regarding temperature at 850 hPa and at 2 meters, both models obtained quite good, similar results.

Regarding the verification of the precipitation forecasts, the so-called fuzzy verification technique was used. We found that IFS obtains better results than WRF in some of the accumulation thresholds considered in both months. The highest ranges are the ones in which WRF obtains better results in July. That being said, it seems that precipitation is possibly the variable in which IFS could further improve its forecasting capabilities regarding WRF.

Finally, even though the results obtained show that the HRES-IFS model obtains equal or better results than WRF, we cannot take this as an absolute truth. In this work, only two months of 2018 were analyzed for a particular region. A more exhaustive analysis would be necessary to reach a more definitive conclusion, considering whole seasons and years and doing more detailed studies in more areas all over the world. Also, some other parameterizations and data assimilation techniques could be used in the WRF model to optimize the results for the particular area of study and do a more realistic comparison obtaining a more accurate analysis. Moreover, radar data could be used aside from surface observation to improve the verification.

Despite that, with this preliminary study we can get an idea about the forecasting abilities of the IFS model, which will probably improve even more in a near future with the optimizations and improvements that the ECMWF will develop in the next versions of the model.

Acknowledgements

We would like to thank the Spanish Meteorological Agency (AEMet) for providing the output data from the HRES-IFS model, and also the Meteorological Service of Catalonia (SMC) for providing the observations from the weather stations used to validate the forecasts.

References

- Bauer, P., Thorpe, A., and Brunet, G., 2015: *The quiet revolution of numerical weather prediction*, *Nature*, **525**, 47–55.
- Cáceres León, R. H. et al.: Impacto de la asimilación radar en el pronóstico de precipitación a muy corto plazo usando el modelo WRF, Ph.D. thesis, Universitat de Barcelona, 2018.
- Codina, B., Aran, M., Young, S., and Redaño, A., 1997: *Prediction of a mesoscale convective system over Catalonia (Northeastern Spain) with a nested numerical model*, *Meteorol. Atmos. Phys.*, **62**, 9–22.
- Collaboration for Australian Weather and Climate Research: *Forecast verification*, available at: <https://www.cawcr.gov.au/projects/verification>, 2017.
- Dutour Sikirić, M., Janeković, I., Tomažić, I., Kuzmić, M., and Roland, A., 2015: *Comparison of ALADIN and IFS model wind speeds over the Adriatic*, *Acta Adriatica*, **56**, 67–82.
- Ebert, E. E., 2008: *Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework*, *Meteorol. Appl.*, **15**, 51–64.
- ECMWF: *IFS Documentation - CY45R1*, Tech. rep., ECMWF, available at: <https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model/ifs-documentation>, 2018.
- Fawcett, T., 2006: *An introduction to ROC analysis*, *Pattern Recognit. Lett.*, **27**, 861–874.
- Hersbach, H., de Rosnay, P., and Bell, B., 2018: Operational global reanalysis: progress, future directions and synergies with NWP, ECMWF.
- Hewson, T., 2017: Use and verification of ECMWF products in member and co-operating states (2016), ECMWF.
- Malardel, S., Wedi, N., Deconinck, W., Diamantakis, M., Kühnlein, C., Mozdzyński, G., Hamrud, M., and Smolarkiewicz, P., 2016: *A new grid for the IFS*, *ECMWF Newsletter*, **146**, 23–28.
- Mercader, J., Codina, B., Sairouni, A., and Cunillera, J., 2010: *Resultats del model meteorològic WRF-ARW sobre Catalunya, utilitzant diferents parametritzacions de la convecció i la microfísica de núvols*, *Tethys*, **7**, 77–89.
- Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W., and Powers, J. G., 2008: *A description of the Advanced Research WRF version 3*. NCAR Technical note-475+STR.
- Sun, J., Xue, M., Wilson, J. W., Zawadzki, I., Ballard, S. P., Onvlee-Hooimeyer, J., Joe, P., Barker, D. M., Li, P.-W., Golding, B., et al., 2014: *Use of NWP for nowcasting convective precipitation: Recent progress and challenges*, *Bull. Am. Meteorol. Soc.*, **95**, 409–426.
- Wang, W., Berner, J., Bruyere, C., Chen, M., Duda, M., Dudhia, J., Fossell, K., Gill, D., Ha, S.-o., Kavulich, M., Keene, K., Lin, H.-C., Michalakes, J., Rizvi, S., and Zhang, X.: *WRF-ARW V3.9: User's Guide*, Tech. rep., UCAR, available at: https://www2.mmm.ucar.edu/wrf/users/docs/user_guide.V3.9/ARWUsersGuideV3.9.pdf, 2017.