AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1002/2015JD023793

Key Points:
- A useful tool for operational management of snowplows has been developed
- Model uncertainty and probabilistic information are provided by a microphysical and PBL ensemble
- In the study area, systematic errors of the parameterizations can be minimized by a postprocessing

Correspondence to:
S. Fernández-González,
sefern04@ucm.es

Citation:
Fernández-González, S., F. Valero, J. L. Sánchez, E. Gascón, L. López, E. García-Ortega, and A. Merino (2015), Numerical simulations of snowfall events: Sensitivity analysis of physical parameterizations, *J. Geophys. Res. Atmos.*, *120*, 10130–10148, doi:10.1002/2015JD023793.

Received 12 JUN 2015
Accepted 14 SEP 2015
Accepted article online 19 SEP 2015
Published online 10 OCT 2015

Numerical simulations of snowfall events: Sensitivity analysis of physical parameterizations

S. Fernández-González[1], F. Valero[1], J. L. Sánchez[2], E. Gascón[2], L. López[2], E. García-Ortega[2], and A. Merino[2]

[1]Department of Astrophysics and Atmospheric Sciences, Faculty of Physical Sciences, Complutense University of Madrid, Madrid, Spain, [2]Atmospheric Physics Group, IMA, University of León, León, Spain

Abstract Accurate estimation of snowfall episodes several hours or even days in advance is essential to minimize risks to transport and other human activities. Every year, these episodes cause severe traffic problems on the northwestern Iberian Peninsula. In order to analyze the influence of different parameterization schemes, 15 snowfall days were analyzed with the Weather Research and Forecasting (WRF) model, defining three nested domains with resolutions of 27, 9, and 3 km. We implemented four microphysical parameterizations (WRF Single-Moment 6-class scheme, Goddard, Thompson, and Morrison) and two planetary boundary layer schemes (Yonsei University and Mellor-Yamada-Janjic), yielding eight distinct combinations. To validate model estimates, a network of 97 precipitation gauges was used, together with dichotomous data of snowfall presence/absence from snowplow requests to the emergency service of Spain and observatories of the Spanish Meteorological Agency. The results indicate that the most accurate setting of WRF for the study area was that using the Thompson microphysical parameterization and Mellor-Yamada-Janjic scheme, although the Thompson and Yonsei University combination had greater accuracy in determining the temporal distribution of precipitation over 1 day. Combining the eight deterministic members in an ensemble average improved results considerably. Further, the root mean square difference decreased markedly using a multiple linear regression as postprocessing. In addition, our method was able to provide mean ensemble precipitation and maximum expected precipitation, which can be very useful in the management of water resources. Finally, we developed an application that allows determination of the risk of snowfall above a certain threshold.

1. Introduction

Snow is one of the most hazardous meteorological phenomena during the cold season on the Iberian Peninsula. Heavy snow can impede transport by road and rail, cause airport closures, and even cut electricity supplies, as what occurred during the heavy snowfall episode in February 2015 in the northwest of the Iberian Peninsula. Moreover, the accumulation of snow on transportation infrastructure can increase the number of injuries and fatalities caused by traffic accidents [*Datla and Sharma*, 2008]. However, the adverse effects of heavy snowfall can be minimized if people are warned in advance. One of the great challenges to modern meteorology is the improvement of quantitative precipitation estimation by numerical weather prediction models during the cold season, because of the important socioeconomic impacts involved [*Pielke and Downton*, 2000]. This challenge is caused mainly by great difficulty in representing the complicated habits and different growth mechanisms of solid hydrometeors [*Iguchi et al.*, 2012]. To address this problem, numerical models classify solid-phase hydrometeors into distinct categories, i.e., cloud ice, snow, and graupel. Regarding liquid phase, two different classes are considered, cloud water and rainwater. The water vapor mixing ratio in the troposphere is also estimated. Numerical models also simulate conversions from one category to another [*Hong and Lim*, 2006; *Thompson et al.*, 2008; *Tao et al.*, 2009; *Morrison et al.*, 2009].

Because of the difficulties in improving deterministic precipitation outcomes, recent interest has been focused on probabilistic techniques [*Jankov et al.*, 2005]. The accuracy and spread of deterministic simulations need to be improved, particularly for mesoscale weather phenomena, which are the most relevant to human activities [*Eckel and Mass*, 2005]. However, advances in computing during recent decades allow for the combination of single-moment and multimoment bulk water microphysics schemes within operational forecast models

[*Molthan and Colle*, 2012]. Probabilistic weather simulations are often based on ensemble techniques, in which numerical weather prediction models are run several times with different initial conditions or model physics, resulting in a finite set of deterministic simulations. Probabilistic methods have the advantage of estimating outcome uncertainty, owing to the spread between distinct deterministic outcomes. Further, they provide probabilistic information that is more useful to users than a single deterministic outcome [*Tracton and Kalnay*, 1993]. These methods are very useful in forecasting specific weather risks [*Homar et al.*, 2006].

In the beginning, ensembles were designed based on perturbed initial conditions, and ensemble mean values were used to estimate the verifying state better than those based on a single ensemble member [*Molteni et al.*, 1996]. This was the strategy of the Global Forecast System and European Centre for Medium-Range Weather Forecasts ensemble prediction systems for medium range, but these produced inadequate spread in the short range because initialization perturbations require time to grow and provide consistent dispersion [*Buizza et al.*, 1999]. Moreover, ensembles built using only perturbations of initial conditions generally have insufficient dispersion [*Hamill and Colucci*, 1998], which may be a consequence of the original assumption that errors only result from uncertainties in the initial conditions. This concurs with *Stensrud et al.* [2000], who demonstrated that both initial conditions and physics perturbations are important in generating ensemble spread, and spread increases more rapidly when model physics variations are included among ensemble members.

During recent years, ensemble techniques have been increasingly investigated for short-term precipitation forecasting [*Alhamed et al.*, 2002]. Introducing several alternatives in model physics is an interesting method to develop an ensemble, especially since there are many uncertainties in parameterizations used in any numerical model [*Harrison et al.*, 1999]. An alternative way to introduce physical uncertainty is the use of different models, leading to a multimodel ensemble [*Krishnamurti et al.*, 2013; *Yuan and Wood*, 2012]. Ensembles that include model uncertainties are more skillful than those that do not, owing to imperfect representation of atmospheric processes in the model [*Evans et al.*, 2000]. In this study we analyzed the influence that different microphysical and planetary boundary layer (PBL) parameterizations of the Weather Research and Forecasting (WRF) model have on estimated precipitation in the northwestern Iberian Peninsula. Moreover, we selected cumulus, radiation, and land surface parameterizations that were the most accurate in similar research [*Yuan et al.*, 2012; *Pennelly et al.*, 2014].

To improve reliability and accuracy of the model, a postprocessing technique can be applied. Multiple options have been used to develop statistical postprocessing, namely, linear regression, logistic regression [*Wilks*, 2009], quantile regression [*Bremnes*, 2004], and neural networks [*Lauret et al.*, 2014]. According to *Yuan et al.* [2008], multiple linear regression with error correction via a cross-validation procedure reduces error as effectively or better than nonlinear methods. This type of correction was used in similar research involving the WRF model [*Erickson et al.*, 2012].

Generating physics ensembles on seasonal time scales is computationally very demanding, although it is possible using low horizontal resolution [*Yuan et al.*, 2012]. In this study we preferred to analyze several days at high resolution instead of long periods of time. *Johnson and Wang* [2012] showed that precipitation estimation can be effectively improved after calibration with 10 days of training. Only marginal improvements were found upon increasing the training period to 25 days. These findings agree with *Gallus and Bresch* [2006], who claimed that ensemble techniques can be evaluated using 15 days. Because the objective of the present research is to improve WRF accuracy for snowfall, we selected 15 days between January 2013 and March 2014 with significant snowfall in the Duero Basin, as determined by requests for snowplows in at least 20 municipalities there. During these 15 days, maximum snowfall was between 15 and 56 mm of snow liquid water equivalent.

In this paper we analyze the effects of four microphysical parameterizations (WRF Single-Moment 6-class scheme (WSM) [*Hong and Lim*, 2006], Goddard [*Tao et al.*, 2009], Thompson [*Thompson et al.*, 2008], and Morrison [*Morrison et al.*, 2009]) and two PBL schemes (Yonsei University (YSU) [*Hong et al.*, 2006] and Mellor-Yamada-Janjic (MYJ) [*Janjic*, 2001]), as well as their interactions, on precipitation amounts and snowfall detection on the northwestern Iberian Peninsula. The main goal is to optimize WRF-simulated precipitation amount expected across the northwestern Iberian Peninsula and to precisely distinguish rain from snow. Comparison between observational data and values produced by numerical simulation is important for evaluating simulation reproducibility and analysis of factors that cause differences in results of distinct microphysical and PBL settings.

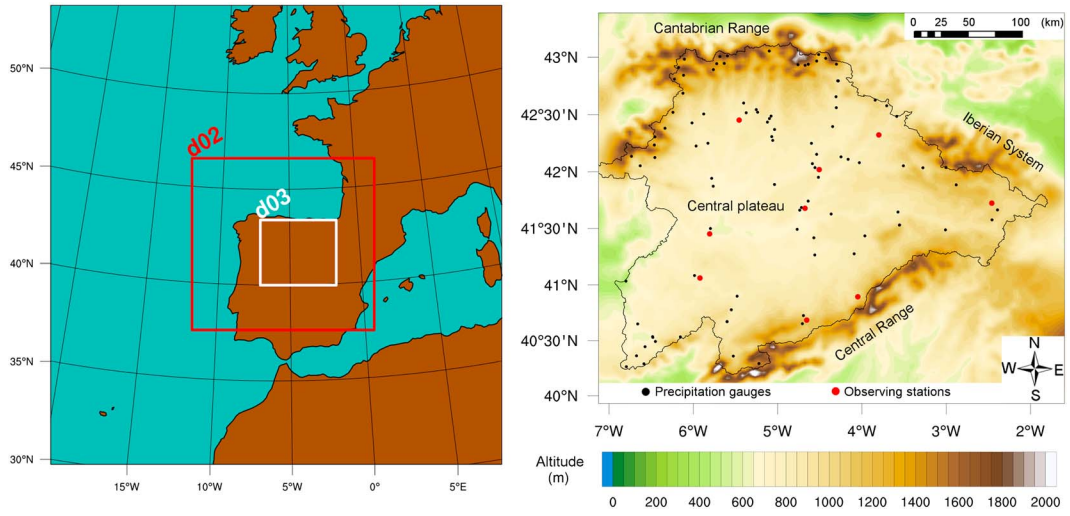

Figure 1. (left) Domains of WRF model. (right) Map of study area (D03), in which black dots represent locations of precipitation gauges and red dots represent locations of Spanish Meteorological Agency observatories; Spanish portion of Duero Basin is outlined in black.

2. Study Area

The study area is the Spanish portion of the Duero basin on the northwestern Iberian Peninsula (Figure 1). This area is mostly in Castilla and León, one of the largest and most varied areas of Europe, which makes it suitable for the purposes of this study. This basin is composed of a central plateau with elevations 600–900 m above sea level (asl), which is surrounded by mountain ranges with altitudes up to 2500 masl (the Cantabrian Range on the north, Iberian System on the east, and Central Range on the south). Altitudes are lower in the southwest and northeast of the area, where oceanic air masses can penetrate the central plateau more easily. The differences in orography cause distinct climatic conditions. Mountainous areas have average annual precipitation >1000 mm, whereas in lower areas of the plateau (rarely exposed to moist air masses from the Atlantic Ocean), average annual precipitation is 400 mm [*Morán-Tejeda et al.*, 2011]. The different climatic conditions existing in the study area derived from the oceanic, continental, and orographic influences [*Caramelo and Manso-Orgaz*, 2007] enable the application of this methodology in similar climatic regions at midlatitudes.

Winter precipitation in this area is largely associated with undulations of the polar jet stream. The most common situations conducive to snow in the area were described by *Merino et al.* [2014]. Snowfall is frequent in mountainous areas. In spite of less frequent snowfall on the central plateau, it causes major problems because most of the population is found there.

During winter, the snowline may undergo drastic oscillations. Upon the entry of cold air masses, snowfall can occur at all elevations in the study area. However, when warm air masses approach, the snow level may rise above 2500 masl, producing rain even at the highest altitudes of mountain ranges.

Morán-Tejeda et al. [2010] stated that in the highest altitudes of the study area, snow can accumulate from late autumn to spring, with a maximum rate of accumulation during winter. In the present study, we considered the cold season to be between November and April, inclusive.

The frequency of snowfall during the cold season strongly depends on altitude. For instance, at altitudes below 700 masl, <10% of precipitation episodes are as snow. This percentage increases to 20% at altitudes 800 masl and >30% above 900 masl. At altitudes near 1100 masl, the percentage of days with snow reaches 50% of all days with precipitation. This percentage increases exponentially to in excess of 90% above 1800 masl. Snow thickness behaves similarly. On the plateau, accumulations rarely exceed 10–20 cm, whereas in mountain areas they can exceed 2 m or even 3 m. There is also great interannual variability of snow accumulation in the mountains of the Iberian peninsula [*López-Moreno and García-Ruiz*, 2004]. Freezing rain events are very infrequent in the study area and are virtually restricted to mountainous areas [*Fernández-González et al.*, 2014].

3. Database

First, we wish to clarify the difference between the databases used. Sections 5.1 and 5.2 quantitatively analyze the daily amount of total precipitation (rain + snow) registered by the precipitation gauges. Therefore, it was not our aim to differentiate between rain and snow amounts in this section. These data have been verified with total precipitation estimated by the model (stratiform and convective precipitation), which includes snow + rain amounts. For this purpose, weighing gauges accurately measure total precipitation (rain + snow). In contrast, tipping-bucket rain gauges can underestimate snow precipitation when the temperature is subfreezing, as well as when they are in areas exposed to strong winds. For this reason, we performed a thorough quality control, removing rain gauges with suspicious or erroneous data.

In sections (5.3 and 5.4), a qualitative (presence/absence) validation of the snow forecast is developed, using a different database for the validation, namely, snow plow requests for snowfall presence and the nine principal observing stations of the Spanish Meteorological Agency for absence. Light snowfall cases not requiring snowplows were not considered. We only considered as presence locations those where snowplow requests ensured that considerable amounts of snow had fallen. For absence of snow, we only considered locations where trained personnel verified that the total precipitation amount was 0 mm or that the precipitation was entirely rain, because the minimum temperature was >5°C. In this case, the WRF variable used was total snow.

The amount of precipitation accumulation on the surface is very difficult to measure accurately. To make these measurements, precipitation gauges are the most common instrument. The magnitude of error from these devices varies as a function of precipitation rate and wind speed [*Huang et al.*, 2015]. Underestimation of ~20%–30% of total precipitation is typical for tipping-bucket precipitation gauges, with wind-induced undercatch the major source of error [*Villarini et al.*, 2008]. For solid precipitation, *Wolff et al.* [2015] found a nonlinear relationship between wind speed and undercatch, with strong temperature dependence. There is also undercatch caused by evaporation and sublimation [*Rasmussen et al.*, 2012]. Weighing gauges probably are the most accurate automatic technique for onsite measurements, because their sensitivity is greater than those of traditional tipping-bucket rain gauges [*Colli et al.*, 2014], especially for snowfall. *Ikeda et al.* [2010] recommended the use of weighing gauges at high altitudes for snowfall measurement.

With the aim of quantitative validation of daily precipitation amount estimated by the WRF model, we used a daily precipitation database of the precipitation gauge network of the Duero Water Management Agency. This network has weighing precipitation gauges at altitudes >1200 masl, which consist of a 200 cm^2 collection tank partially filled with antifreeze, which rests on an electronic weighing system. The surface of the bucket is impregnated with an oil film, preventing evaporation or loss of hydrometeors by blowing out of the bucket before they can be weighed. Below 1200 masl, there are tipping-bucket precipitation gauges with resolution of 0.2 mm.

The network has 165 precipitation gauges, and we analyzed data for all 15 days. It was necessary to filter the original data for quality control. This was because of measurement difficulties (heavy snowfall and strong winds) causing insufficient accuracies at some gauges. Average daily accumulated precipitation from the various WRF settings was compared with the database. A Student's *t* test was applied, removing precipitation gauges with Pearson's correlation coefficient not reaching significance level $\alpha = 0.1$. We selected this significance level to ensure elimination of all stations with data suspected of being erroneous. In this way, we eliminated rain gauges within areas exposed to strong winds, where precipitation is underestimated. Analogously, data from gauges incapable of measuring snow at temperatures <0°C were eliminated. Therefore, we only selected precipitation gauges in locations sheltered from strong winds and those capable of accurately measuring snow.

After this process, a final database of 97 precipitation gauges was used for verification of WRF simulations. Most deleted gauges were at altitudes 1000–1200 masl, where tipping-bucket gauges were unable to accurately measure liquid water equivalent during snowfall with subfreezing temperatures.

Figure 1 shows that the spatial distribution of precipitation gauges was not homogeneous, with greater density in mountainous areas and less density on the central plateau. The main reason for this is that the gauges are concentrated near reservoirs to optimize management of water resources. This discouraged the interpolation of data from precipitation gauges to a grid for comparison to WRF precipitation outputs. This interpolation could have led to erroneous values where gauges density was low, especially considering the abrupt orography of the study area. Therefore, it was decided to validate the gauge network observations with

WRF-simulated precipitation at the closest grid point to each gauge. We have considered the option to apply a bias correction using lapse rate developed by *Gao et al.* [2015]. However, owing to the complicated nature of precipitation processes in mountainous areas and accurate representation of the orography in our study area by the high resolution of the model (mean bias between altitude of the closest grid point and actual altitude is 63 m for D02 and only 54 m for D03), we decided to use the nearest grid point for validation purposes.

The precipitation gauges of the Duero Water Management Agency cannot identify precipitation type. Therefore, to validate the presence/absence of snow simulated by the model, it was necessary to use a different database. Snowfall was validated by locations of snowplow requests. These data were obtained through the emergency service of Spain. The absence of snow was determined based on surface synoptic observations of the nine principal observing stations of the Spanish Meteorological Agency in Castilla and León at which there are trained personnel taking measurements. Therefore, these data are considered reliable. These data were compared with the WRF simulation at their locations. The database to validate snowfall presence or absence was smaller than that of precipitation, owing to the difficulty of obtaining reliable data. There were 488 reports of the snowfall presence from the snowplow requests. There were 65 reports of snowfall absence from the principal observing stations of the Spanish Meteorological Agency.

4. Methodology

4.1. WRF

The WRF model is a numerical weather prediction and atmospheric simulation system designed for operational forecasting and atmospheric research. It is a nonhydrostatic, three-dimensional model, whose details are described in *Skamarock and Klemp* [2008]. In the present study, version 3.5.1. of the Advanced WRF was used, and sensitivities to microphysics and PBL parameterizations were evaluated.

The model was initialized with initial and boundary conditions from the National Centers for Environmental Prediction (NCEP) Global Forecast System analyses with 1° horizontal grid spacing. NOAA/NCEP global analyses provide time-varying lateral boundary conditions at 6 h intervals, with data of the surface and 26 pressure levels, from 1000 to 10 hPa. Simulations of the 15 days analyzed were run individually.

We used the scheme defined by *Dudhia* [1989] for shortwave radiation, and the Rapid Radiative Transfer model [*Mlawer et al.*, 1997] was selected for longwave radiation. Furthermore, the Noah Land Surface Model [*Chen and Dudhia*, 2001] was used, which is a four-layer soil temperature and moisture model with canopy moisture and snow cover estimation, which provides data of sensible and latent heat fluxes to the PBL scheme. With the aim of minimizing computational costs, we decided not to evaluate other radiation and land surface parameterizations. This configuration was selected after a thorough literature review [e.g., *Evans et al.*, 2012; *Yuan et al.*, 2012; *Liang et al.*, 2012]. We selected parameterizations with superior results in similar studies for winter conditions.

Most cumulus parameterizations have been developed for low-resolution models, in which they are useful. However, it is more accurate to simulate convective processes explicitly at mesoscale resolution, making unnecessary the use of cumulus parameterizations for resolutions higher than 10–15 km [*Wang and Seaman*, 1997; *Arakawa*, 2004]. *Pennelly et al.* [2014] compared different cumulus parameterizations, obtaining greater accuracy in the forecasting of precipitation using the Kain-Fritsch cumulus scheme [*Kain*, 2004] in domains with lower resolution and explicit cumulus for higher resolutions. In this research, we performed a test using the new Kain-Fritsch cumulus parameterization for D02 and D03 instead of explicitly resolved convection, but the results were unsatisfactory, so it was not used. Considering the above, the new Kain-Fritsch cumulus parameterization was chosen for convection simulation in domain D01. In the domains with higher resolution, cumulus was resolved explicitly.

Finally, we used the Eta surface layer scheme described by *Janjic* [1996]. These schemes have been used by *Gascón et al.* [2015] and *Fernández-González et al.* [2015] for the mainland Iberian Peninsula, with satisfactory results.

A total of 54 sigma levels were defined to study meteorological conditions in the troposphere. We selected a large number of vertical levels to more clearly ascertain differences between the distinct PBL and microphysical schemes used. The spacing of sigma levels was progressive, with greater density near the surface for better representation of the PBL.

Three nested domains were defined to achieve high spatial resolution. The domains D01, D02, and D03 had respective spatial resolutions 27, 9, and 3 km. Intervals between outputs were 3, 1, and 1 h, respectively, for D01, D02, and D03. D01 had 100×100 grid points in both south-north and east-west directions, covering the southwest of Europe and nearby Atlantic Ocean. D02 had 121 east-west grid points and 112 north-south grid points. D02 nearly coincided with the extent of the Iberian Peninsula. To cover the entire Spanish part of the Duero Basin, D03 had 151 east-west grid points and 127 north-south grid points. The selected number of days simulated and resolutions were limited by available computational resources.

An approach to generating ensembles is the use of different physical parameterization schemes to construct various versions of a model and produce an ensemble of simulations that start from the same initial conditions. All model configurations are assumed equally skillful, and the ensemble average should ideally provide a more robust simulation than any individual ensemble member [*Stensrud et al.*, 2000].

The WRF offers multiple options for most physics schemes, enabling the user to optimize the model for a range of spatial and temporal resolutions and climatologically different regions. Here we investigate the performance of various options commonly used for real data applications. We tested four microphysics and two PBL schemes, which are described in the following paragraphs. In this way, we obtained eight deterministic simulations that were combined into an ensemble that permitted development of a probabilistic simulation.

While single-moment bulk microphysics schemes estimate only the mixing ratios of hydrometeors, double-moment methods include an additional prognostic variable that is related to the size distribution, such as number concentration. The single-moment schemes use an assumed distribution function for hydrometeor size [*Walko et al.*, 1995], whereas the double-moment approach explicitly calculates hydrometeor number concentrations, thereby allowing flexibility in the hydrometeor size distribution [*Morrison et al.*, 2005; *Thompson et al.*, 2008]. Double and multimoment approaches require cloud condensation nuclei (CCN) information. The main disadvantage of those approaches is that they require greater computational time than the single-moment schemes. In this research, we investigated a combination of the most commonly used single-moment and double-moment schemes. Following are the microphysical schemes treated herein.

The first microphysical scheme was the WSM [*Hong and Lim*, 2006]. In this parameterization, properties of cloud ice and related cold microphysical processes are calculated as defined by *Hong et al.* [2004]. Further, various saturation techniques and a process for the conversion of cloud ice to snow were added by *Tao et al.* [1989] and *Braun and Tao* [2000].

The next selected parameterization was the Goddard, which is a single-moment 6-class microphysics scheme [*Tao et al.*, 2009]. This includes cold rain processes defined by *Rutledge and Hobbs* [1984] and *McCumber et al.* [1991]. *Tao et al.* [1989] added several saturation techniques, a new process for the conversion of cloud ice to snow, and corrections regarding graupel-related variables following the specifications of *Braun and Tao* [2000] and *Lang et al.* [2007].

The third scheme is the Thompson 6-class microphysics [*Thompson et al.*, 2008]. It is a single-moment scheme, although for cloud ice and rain it also acts as a double-moment scheme. The Thompson scheme determines the hydrometeor mixing ratio and, for rain and cloud ice hydrometeors, also the number concentration. Snow shape is considered nonspherical, with bulk density varying inversely with diameter as in observations [*Thompson et al.*, 2004], and its size distribution is represented as a sum of exponential and gamma distributions.

Finally, the Morrison 6-class double-moment scheme [*Morrison et al.*, 2009] was used. It includes more complicated microphysics, estimating number concentrations and mixing ratios of four hydrometeor species (cloud water, cloud ice, rainwater, and snow), rain size distribution, and different rates of rain evaporation in stratiform and convective regions. Cloud number concentration is diagnosed. The Morrison microphysical parameterization has options to optimize simulations by accommodating the selection of ice nucleation method and CCN spectra [*Morrison et al.*, 2005].

All four schemes have the same number of water substances, water vapor, cloud water, rainwater, cloud ice, snow, and graupel.

The surface layer determines friction velocities and exchange coefficients for surface heat and moisture fluxes. In the WRF model, these processes are mainly defined by the PBL scheme, which determines vertical subgrid

fluxes owing to eddy transport in the rest of the atmospheric column. PBL schemes deal with the problem of parameterizing the turbulent layer, which develops above the surface because of surface heating, wind shear and friction. Vertical transport of heat, moisture and momentum, and low-level cloud development are influenced by PBL processes. For these reasons, accurate parameterization of this layer is essential to achieve realistic simulations, especially for surface variables such as precipitation type and amount. In this research we used the following PBL schemes.

The first PBL scheme was that of YSU [*Hong et al.*, 2006]. It is a first-order scheme that computes turbulent fluxes using nonlocal eddy diffusivity coefficients. The YSU scheme is based on the Medium Range Forecast model PBL scheme but improves it with explicit treatment of entrainment. PBL height is diagnosed using a critical Richardson number.

The second scheme is MYJ [*Janjic*, 2001], which is a local closure scheme of order 1.5 and level 2.5. Equations for heat and moisture fluxes include a term that allows them to go against the local gradient, which permits simulation of countergradient fluxes caused by large eddies. This scheme also calculates local vertical mixing. These characteristics, together with estimation of turbulent kinetic energy diffusion outside the mixed layer, improves the representation of entrainment. PBL height is diagnosed using a turbulent kinetic energy threshold. The top of the layer depends on turbulent kinetic energy as well as buoyancy, stability, and shear of the driving flow.

4.2. Validation Methodology

The relationship between quantitative precipitation estimated by the model and observed data is often characterized by their correlation and standard deviation, whereas accuracy is often estimated via the root-mean-square difference (RMSD). In this study, we tested the reliability of each WRF physics scheme and their ensembles in simulating accumulated precipitation with the RMSD, which is especially useful for evaluating discontinuous variables such as precipitation [*Stensrud and Wandishin*, 2000]. Furthermore, spatial similarity has been measured with Pearsons correlation coefficient, which has been used in similar analyses [*Evans et al.*, 2012]. Moreover, standard deviation has been used to estimate similarity between deterministic simulations and observed values. As claimed by *Schwartz et al.* [2010], standard deviation can also be used to determine the spread of the ensemble members. Standard deviation, correlation coefficient, and RMSD results were summarized in a Taylor diagram, following the methodology in *Taylor* [2001].

Verification of dichotomous outcomes was performed by constructing a contingency table composed of elements representing all possible scenarios, including probability of detection (POD) (the model detects an observed event), frequency of misses (FOM) (an event occurs that was not detected by the model), false alarm ratio (FAR) (the model detects an event that did not occur), and probability of correct rejection (PCR) (the model correctly determines that an event does not occur). A score of POD or PCR of 1 represents perfect estimation. A score <0.5 has no skill, and 0.7 represents the lower limit of a useful simulation. In contrast, FOM and FAR scores equal to 0 indicate perfect estimation, whereas values close to 1 indicate large errors [*Buizza et al.*, 1999]. The contingency table and equations of these parameters are found in *López et al.* [2007].

Finally, the Brier score (BS) has been used in verification of snowfall outcomes. The BS, which was defined by *Brier* [1950], is one of the most common methods for verification of probabilistic outcomes of dichotomous events. This skill score measures the total probability error, averaging squared differences between pairs of estimated probabilities and corresponding observations. The observation is set to 1 if the event occurs and 0 if it does not. The BS can take on values between 0 and 1, 0 being perfect estimation. Values near 1 indicate less accurate estimation. The BS has been used to validate quantitative precipitation simulations from the WRF [*Yuan et al.*, 2008]. For deterministic outcomes, the probability is 1 when the model estimates snow and 0 when it does not indicate snow risk.

4.3. Postprocessing

Postprocessing can improve deterministic and probabilistic outcomes. Model Output Statistics (MOS) is an objective technique that determines a statistical relationship between a predictand and variables output from a numerical model. The approach taken herein was to produce these MOS probabilities through multiple linear regression. The error reduction in *Lu et al.* [2007] recommends application of a linear regression method over an ensemble outcome.

Multiple linear regression relates one variable (Y, known as the predictand or dependent variable) to several other variables (X_i, called predictors or independent variables). The result is an equation in which the predictand can be estimated as a linear combination of the predictors:

$$Y = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + a_5X_5 + a_6X_6 + a_7X_7 + a_8X_8$$

The fundamental concept of this postprocessing is that when one model configuration is significantly less skillful than the others, then the ensemble members should be weighted unequally to obtain better results. The parameters of the linear regression equation were obtained by introducing observed data and deterministic outcomes during 10 of the 15 days analyzed. Data from the remaining 5 days were used for validation of the postprocessing equation. Days used for constructing the regression equation were randomly chosen.

5. Results

This section will present the results obtained with various microphysical parameterizations and PBL schemes (as well as their combinations) both in the case of total precipitation accumulated and in the form of snow. Besides deterministic simulations, a mean estimation has been calculated, in order to achieve a more accurate verification. What is more, the existing spread between the eight deterministic simulations gives an idea of the reliability of the model.

5.1. Precipitation Estimation

We ran the WRF model, combining the four microphysical parameterizations and two PBL schemes, to obtain eight deterministic outcomes capable of simulating different scenarios. In principle, all simulations were equally valid. The spread of ensemble members (measured by the standard deviation) is typically associated with perceived model uncertainty. Otherwise, probabilistic outcomes can be readily obtained by considering the total number of members detecting an event at a given grid point. Alternatively, information from all the members can be averaged into an ensemble mean precipitation [*Schwartz et al.*, 2010].

We analyze in depth the case of 13 March 2013, which is representative of the influence of the various microphysical and PBL parameterizations used. In this case, light snowfall was expected in northern, central, and eastern Castilla and León, with less likelihood in the southwest of the region. Moderate and even heavy snowfall was predicted on the north side of the Cantabrian Range and locally in the highest mountains of the Central Range and Iberian System. The snowfall episode was caused by Arctic air advection associated with a surface high in the Atlantic, coupled with a trough in the northeastern Iberian Peninsula. This pattern usually produces orographic precipitation on the windward side of the mountain ranges, mainly in the Cantabrian region. In the eastern Duero Basin, the lower altitudes of the Cantabrian Mountains there allowed the entrance of moisture [*Merino et al.*, 2014]. The synoptic situation was characterized by the entrance of northerly flow. This was produced by an anticyclone over the Atlantic and low pressure centered over the northeastern Iberian Peninsula, which caused instability in northern Spain. The episode was characterized by heavy precipitation on the north (windward) side of mountain ranges, especially the Cantabrian Range. Precipitation was weaker on the plateau and nearly nonexistent over the southwestern Duero Basin. This situation indicates that orographic forcing was the main cause of the precipitation enhancement on the windward side of the mountain ranges.

First, we analyze and show quantitative precipitation estimation from the method developed in the research. In Figure 2, the eight deterministic outcomes in D03 are shown for the case of 13 March 2013. The ensemble mean and spread (given by standard deviation between the eight individual outcomes) have been included in this figure, in addition to the maximum precipitation expected from any of the eight deterministic simulations. Images for the other 14 days are not shown for space reasons, although the results indicate similar characteristics as the day illustrated.

The individual deterministic simulations produced varying spatial patterns of precipitation, and it appears that these differences are attributable to the different PBL and microphysics schemes. Similar to *Jankov et al.* [2005], these findings indicate that spread in precipitation can be realized by varying the physical parameterizations within an ensemble method. Observed precipitation during this day strongly depended on the orographic forcing, so WRF-simulated precipitation is extremely sensitive to variations in the microphysics and PBL configurations used. This is consistent with *Stensrud et al.* [2000], who indicated that a model physics ensemble was more skillful than an initial-condition ensemble when large-scale forcing of upward motion is weak.

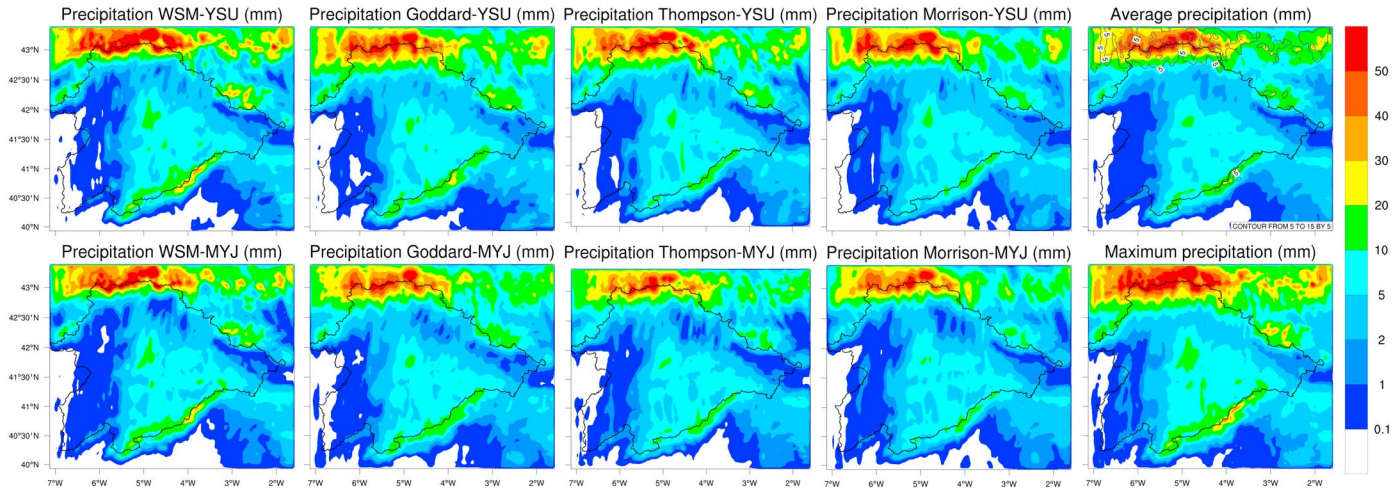

Figure 2. Daily accumulated precipitation estimated by WRF model for 13 March 2013. Eight deterministic outcomes are shown. (top rightmost) Ensemble mean precipitation and spread. (bottom rightmost) Maximum precipitation.

Analyzing the results from different microphysical parameterizations, it was seen that the WSM estimated stronger orographic forcing. In consequence, the estimated precipitation amount on the windward side of the mountains was greater than those from the other three parameterizations. However, the WSM produced the least precipitation on the plateau, confirming that it has a larger contribution of orographic effects.

Another difference of the WSM relative to the other microphysical parameterizations is that abundant precipitation was much more restricted on the north side of the Cantabrian Range. In the other three simulations, there was precipitation >20 mm on the lee side of that range. The explanation for this may be that orographic precipitation was able to cross the range and reach its leeward side when the pressure gradient was strong and there was a cold air mass at upper levels, which causes instability and increases the likelihood of convective precipitation [*Merino et al.*, 2014]. Therefore, the WSM appears more accurate when the pressure gradient is weak. The Goddard, Thompson, and Morrison schemes are expected to be superior around the lee side of the Cantabrian Range when that gradient is strong.

Differences between the YSU and MYJ schemes were more significant. In general terms, precipitation estimated by the MYJ was less than that from the YSU. Although there was a slight decrease of observed precipitation on the plateau, the most significant decline was recorded in mountainous areas. The reduced precipitation from the MYJ scheme may be connected with the fact that PBLs simulated by that scheme are typically cooler, with stronger thermal inversion layers. This can hamper development of orographic forcing and resultant clouds and precipitation.

The ensemble mean in Figure 2 (top rightmost) shows the average precipitation of the eight deterministic outcomes. Also depicted is the standard deviation between the deterministic simulations, revealing a greater spread in mountainous areas, particularly the Cantabrian Range. The main cause of this significant spread is the different formulations for orographic forcing in the microphysical parameterizations and PBL.

This method can be used operationally and would be very helpful to water resource management by the Spanish hydrographic confederations. Managers of these resources in the country generally use deterministic simulations, which typically have large errors, especially for extreme events. The present tool provides more information, because the spread gives some idea of confidence in the model outcome. Moreover, the method gives maximum expected precipitation (shown in Figure 2 (bottom rightmost)) and risk (in probabilistic terms) that it will be reached. Although this maximum is not the most likely value, managers are advised to take appropriate action in case it occurs.

In addition, we analyzed the temporal sensitivity of each of the various combinations of parameterizations to see which had greater reliability in determining the temporal evolution of rainfall. To do this, we analyzed precipitation accumulated every 3 h throughout the day of 13 March 2013.

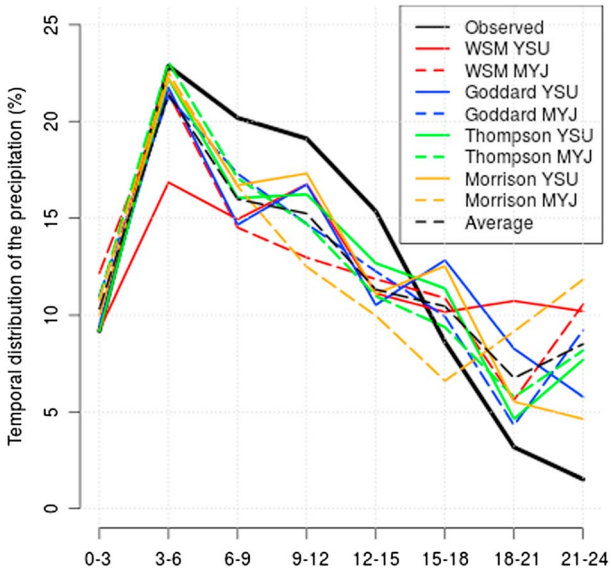

Figure 3. Precipitation accumulated every 3 h for 13 March 2013, both observed and estimated by WRF model.

There were a total of 23 precipitation gauges with valid data every 15 min, all of them weighing gauges and therefore reliable for analysis of temporal sensitivity to various WRF settings. From these data, we calculated precipitation accumulated every 3 h and then compared it to WRF-estimated precipitation in D03.

Figure 3 shows the evolution of precipitation every 3 h during 13 March 2013 (both observed and model estimated). We have represented the percentage of precipitation every 3 h to more clearly visualize temporal similarities and differences between various parameterizations and ground truth.

As seen in Figure 3, all combinations of parameterizations were very accurate for the first 6 h (except WSM-YSU, which underestimated considerably the percentage of recorded precipitation over 3–6 h). Later, between 6 and 15 h, all settings underestimated the percentage of precipitation recorded during this period. However, between 18 and 24 h the opposite occurred; i.e., all WRF model configurations estimated precipitation percentages higher than actually recorded.

Regarding microphysics parameterization, Thompson had the most accurate temporal sensitivity during the day, with correlation coefficient $r = 0.92$. Goddard followed closely with $r = 0.90$, whereas WSM and Morrison had a significantly lower sensitivity ($r = 0.83$ and 0.82, respectively). Regarding PBL schemes, YSU ($r = 0.90$) was considerably more accurate than MYJ ($r = 0.83$). Individually, the best combination would be the Thompson microphysics parameterization and YSU PBL scheme. By calculating the average precipitation of the eight combinations, very satisfactory results were obtained, with $r = 0.92$. These results are comparable to those of *Evans et al.* [2012].

Table 1 shows mean values in D03 during 13 March 2013 for the following variables: temperature, integrated water vapor (IWV), liquid water path (LWP), ice water path (IWP), and precipitation. For the last four, the values are averages over D03 of totals accumulated throughout the day. Mean values of the two deterministic outcomes from each microphysical parameterization and averages of the four simulations using the same PBL scheme are listed, with the goal of more clearly differentiating the influences of each parameterization.

Table 1. D03 Mean Temperature, IWV, LWP, and IWP

	WSM	Goddard	Thompson	Morrison	YSU	MYJ
Temperature (°C)	0.06	−0.18	−0.11	−0.18	0.18	−0.39
IWV (kgm^{-2})	12.85	13.15	13.42	13.64	13.58	12.94
LWP (gm^{-2})	31.67	81.21	126.81	212.02	117.55	108.31
IWP (gm^{-2})	349.10	572.13	435.49	346.77	464.96	386.84
Precipitation (mm)	7.11	7.15	6.31	6.17	7.19	6.19

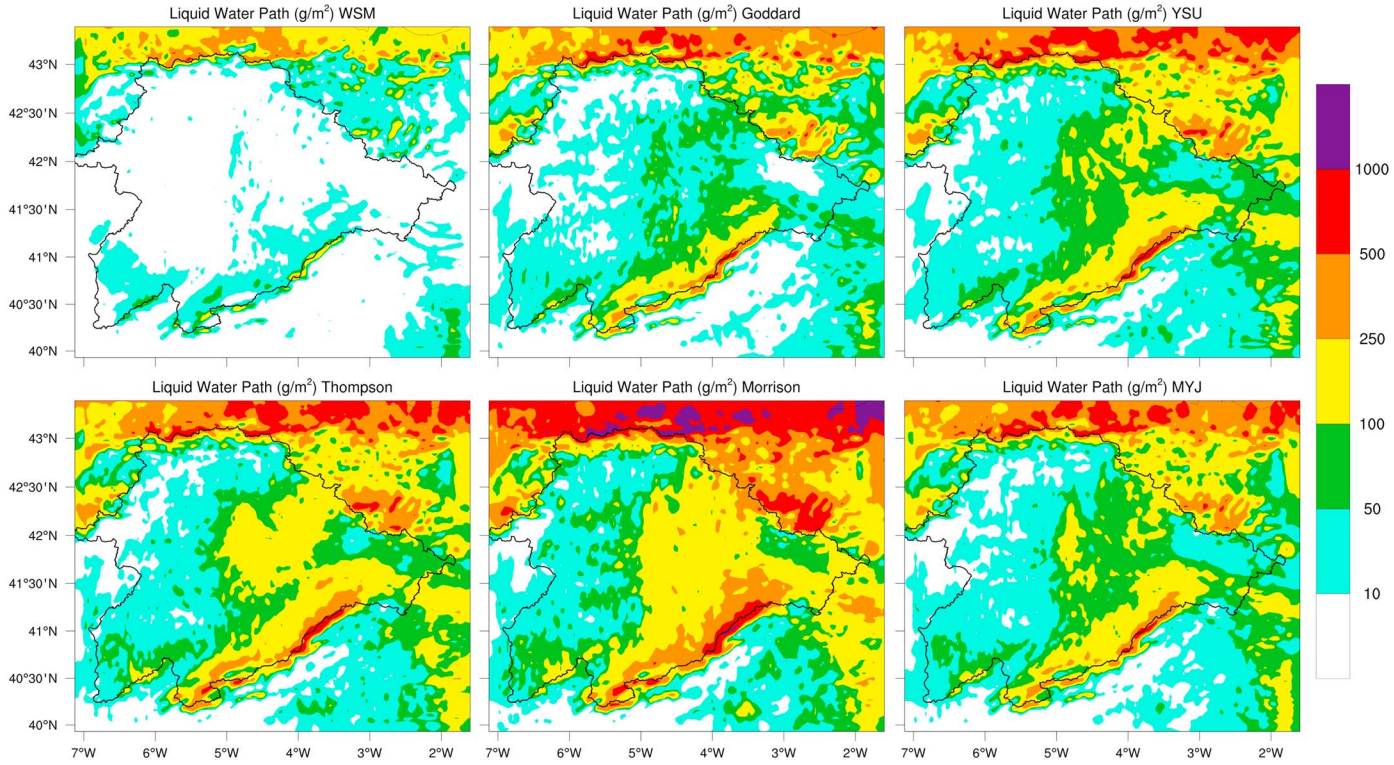

Figure 4. LWP (gm^{-2}) accumulated during 13 March 2013 in D03 for various microphysical parameterizations and PBL schemes.

Surface temperatures simulated by the various microphysical schemes were very similar. There were only slight differences between them (0.25°C at most), with the warmest being the WSM. However, the difference was greater among the PBL schemes. On average for D03, simulations with the YSU scheme had temperatures almost 0.6°C warmer than those with the MYJ scheme. The cooler temperature of MYJ may have contributed to the lesser precipitation simulated with that PBL scheme, because a lower temperature can reduce precipitation generated by convective processes. The aforesaid temperature difference might appear small, but in snowfall episodes a slight temperature increase can alter the precipitation type at the surface, from snow to rain and vice versa.

The YSU scheme produced tropospheric IWV 5% larger than the MYJ scheme. The water vapor mixing ratio is very important for the development of precipitation. It seems logical that the greater the water vapor available in the troposphere, the greater the expected precipitation. Therefore, the less water vapor from the MYJ could be a major cause of the lesser precipitation output by this PBL scheme compared to the YSU. The various IWV values estimated by the distinct microphysical parameterizations do not appear to have significant correlation with estimated precipitation. Mean IWVs determined by the different WRF settings were similar to those obtained by *Morland and Mätzler* [2007].

LWP and IWP accumulated during a particular day of the database are analyzed in Figures 4 and 5 and Table 1. One sees that the single-moment microphysical schemes (WSM and Goddard) estimated a larger proportion of solid phase and lower concentration of liquid hydrometeors. Simulations with such schemes have difficulty determining accurate size distributions. Furthermore, because of their similarity in modeling warm rain processes, the outputs of these schemes present similar concentrations of both liquid water and ice phase [*Kim et al.*, 2013]. Nevertheless, Morrison (a double-moment scheme) produced a lower amount of solid phase and greater liquid phase, likely because its formulation for heterogeneous nucleation processes is more complex and accurate. Another reason is that modified warm rain processes in double-moment schemes yield smaller amounts of cloud water [*Lim and Hong*, 2010]. This means that the Morrison scheme has less efficient glaciation processes. The Thompson microphysical scheme is in an intermediate position, likely because it is a partial double-moment scheme.

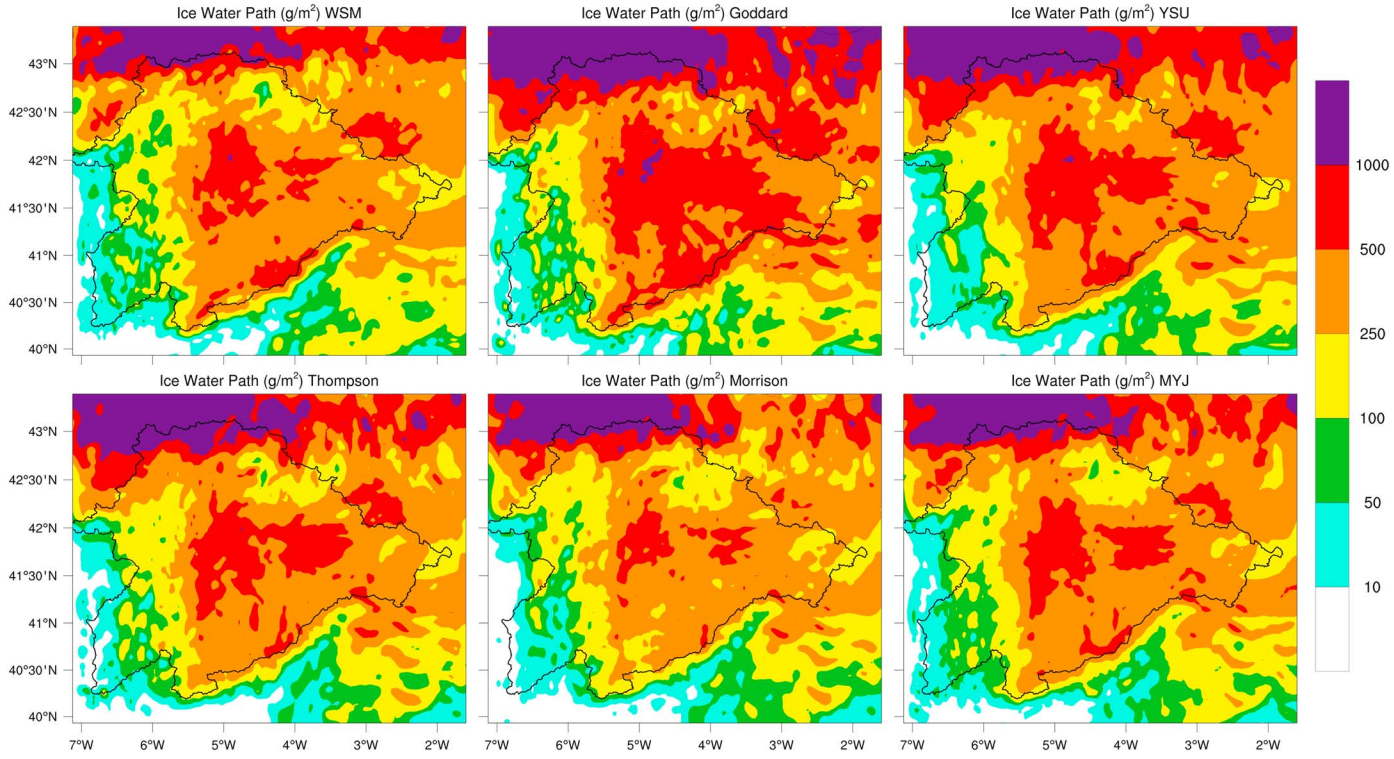

Figure 5. IWP (gm^{-2}) accumulated during 13 March 2013 in D03 for various microphysical parameterizations and PBL schemes.

The shorter IWP simulated using the Morrison scheme relative to Goddard and Thompson may have been caused by its absence of enhanced melting processes of snow and graupel [*Morrison et al.*, 2009]. Our results show that the LWP is very important in determining total amounts of precipitation because, according to *Li et al.* [2009], it modulates related microphysics processes, especially evaporation rate. Differences between PBL schemes were minor for LWP and IWP.

The small LWP from the WSM (Figure 4) is particularly significant, suggesting a clear underestimation of this parameter. In association, a gradient of LWP is evident between single-moment and double-moment schemes. The more complex formulation in double-moment schemes may simulate more accurate LWP and IWP, recommending their use for estimation of hydrometeor number concentration when accurate estimation is required.

For average precipitation in D03, we detected notable differences between the various settings. First, precipitation estimates from single-moment schemes were ~10%–15% greater than those from double-moment schemes (considering Thompson as a double-moment scheme). Similarly, average precipitation output from the YSU scheme was 16% greater than that from the MYJ. As will be shown in section 5.2, this appears to represent a systematic deviation between the various parameterizations, making postprocessing mandatory. In summary, the results confirm that the PBL scheme and microphysical parameterizations used have a strong influence on the surface precipitation pattern, consistent with *Iguchi et al.* [2012].

5.2. Precipitation Validation

The performance of each physics scheme setting and ensembles of multiple physics parameterizations were evaluated separately. In theory, all ensemble members are assumed equally likely to represent atmospheric conditions at initialization and therefore have the same probability of producing the most accurate outcome [*Schwartz et al.*, 2010].

Numerical results are shown in a Taylor diagram (Figure 6). Analyzing each scheme, it is seen that correlation coefficients are similar, with greater differences in standard deviation and RMSD. Overall, no particular configuration stands out as much better than another, concurring with *Gallus and Bresch* [2006]. However, a large RMSD was not uniform across the model configurations. The best microphysical parameterization was the Thompson scheme, which had the smallest RMSD. Although the WSM had a correlation coefficient

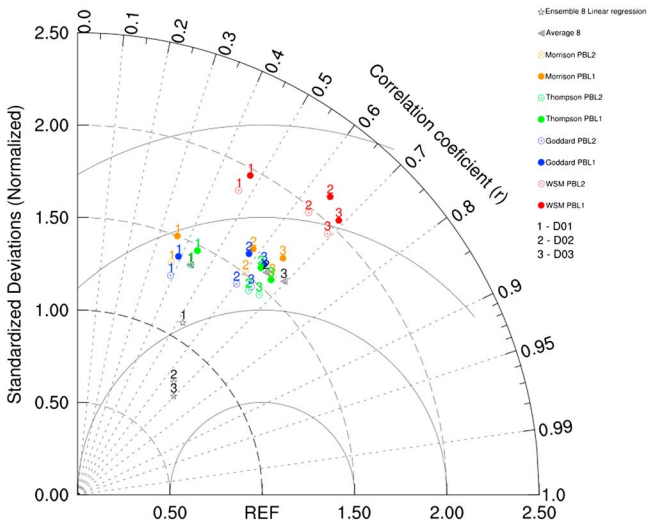

Figure 6. Taylor diagram showing correlation coefficient, RMSD, and standard deviation resulting from different settings of WRF model, ensemble averaging, and postprocessing. Results for D01, D02, and D03 are represented.

comparable to the other schemes, it had the greatest standard deviation and consequent larger RMSD. The Goddard and Morrison schemes were similar to the Thompson scheme but with slightly larger RMSD. On average, all deterministic simulations overestimated precipitation.

Regarding PBL parameterizations, we detected a systematic positive bias in precipitation estimated by the YSU scheme, whereas simulations using MYJ PBL physics gave an overall smaller daily RMSD in the Duero Basin. Analyzing the results of the four microphysical parameterizations, in all cases there was a reduction of RMSD and standard deviation with use of the MYJ versus YSU. *Weisman et al.* [2008] stated that this may be because the YSU typically forms deeper PBL, tending to eliminate capping inversions. However, the MYJ used to simulate PBL conditions is characteristically cooler with stronger thermal inversions, which reduces the convective component of precipitation and thereby its total amount. These results agree with those of *Evans et al.* [2012].

As a first approach to improving the skill of precipitation estimation, it was decided to average the deterministic outcomes to obtain an ensemble average. The objective was more robust and accurate precipitation outputs. Simple equal-weight ensemble averaging of precipitation was compared with observed precipitation. Upon averaging the eight deterministic outcomes, the correlation coefficient increased moderately, indicating a closer approximation to the precipitation spatial distribution. Nevertheless, there was no significant improvement of RMSD. The results show that the ensemble average generally produces more realistic precipitation estimates than the individual deterministic simulations. Improvements over individual ensemble members were seen in the accuracy of spatial patterns (measured by the correlation coefficient), consistent with *Ji et al.* [2014]. However, the RMSD remained large, necessitating postprocessing.

Therefore, we decided to apply a linear regression method with the aim of improving the accuracy of quantitative precipitation estimation over the study area. This was done after discovering a systematic error in the YSU parameterization and WSM. The large standard deviation and RMSD produced with these settings make postprocessing convenient. In this way, we reduced uncertainties of the deterministic simulations and mitigated individual errors. Furthermore, a postprocessing allows us to give a greater weight to the most accurate deterministic members and vice versa. The greatest improvement after using the multiple linear regression was a significant reduction in standard deviation and RMSD, with slight improvements to the spatial distribution of precipitation (the correlation coefficient only increased slightly). In conclusion, by using the eight deterministic outcomes as sole predictors and without considering spatial variations, linear regression reduced RMSD, in agreement with *Wilson* [2007] and *Yuan et al.* [2008].

Upon analyzing model response to increased spatial resolution, accuracy improvement was dramatic for the step from 27 to 9 km, but there was only slight improvement for 9 to 3 km. Examining the plateau surrounded by mountain ranges, this improvement was especially notable in areas of more rugged terrain.

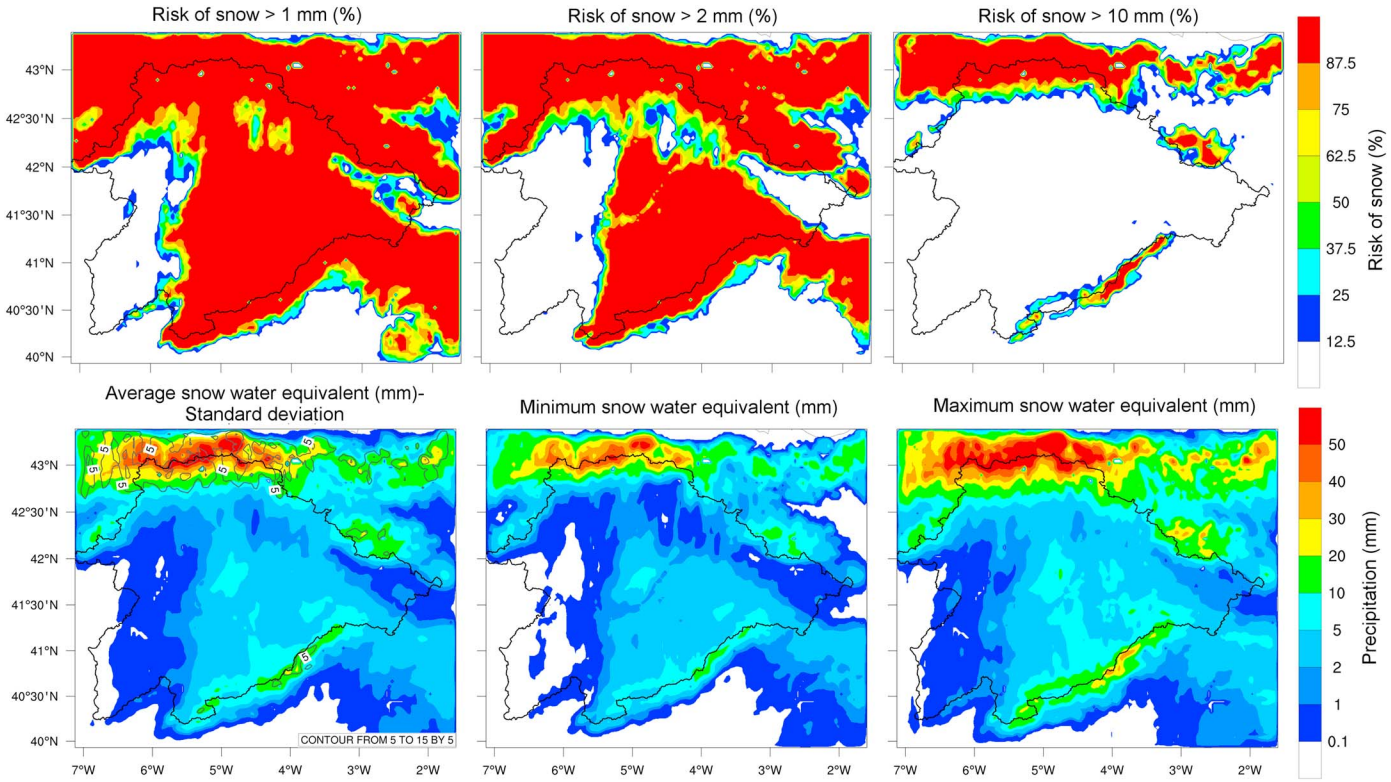

Figure 7. (top row) Risk of reaching snow water equivalent values >1, 2, and 10 mm. (bottom left) The ensemble average together with the spread between eight deterministic simulations. (bottom middle) Minimum and (bottom right) maximum snow water equivalents.

The major improvement from increased model spatial resolution was better representation of the orography and resulting greater accuracy in the simulation of mesoscale processes [*Fargey et al.*, 2014]. However, the RMSD decrease was more noticeable upon combining several parameterizations and applying postprocessing. These results agree with those of *Du et al.* [2004], who found greater improvement in probabilistic simulation than in a single run with twice the horizontal grid resolution. Therefore, for operational use of the WRF model, it would be interesting to add different physical schemes. This is because the outcomes would likely be more accurate and computational costs would be saved. Our RMSD and correlation coefficient values have the same order of magnitude as those of *Kim et al.* [2013]. As demonstrated by *Ikeda et al.* [2010], a horizontal resolution of at least 6 km is necessary to accurately simulate mesoscale processes. Nevertheless, it is unclear if it is worth increasing resolution beyond 3 km, given the associated great increase in computational cost, as found by *Weisman et al.* [2008].

In short, deterministic outcomes are acceptable in terms of precipitation spatial distribution (correlation coefficients >0.65 for D03), but they overestimate precipitation (large standard deviation and RMSD). Averaging the eight deterministic outcomes produced a more accurate spatial distribution (correlation coefficient = 0.70) but only slightly reduced RMSD. The large RMSD necessitates postprocessing to reduce systematic overestimation of precipitation. A reduction of ~25% was achieved through multiple linear regression. This postprocessing facilitates more accurate precipitation values.

5.3. Snowfall Estimation

This section explores improvement of snowfall estimation by WRF. Firstly, we ran eight deterministic simulations by combining the aforementioned four microphysical parameterizations and two PBL schemes. Then, we combined all these in a probabilistic simulation, which could furnish additional information on the occurrence probability of an event and model reliability. Furthermore, we attained information on minimum and maximum snowfall associated with the various deterministic outcomes.

Probabilistic outcomes of snowfall were realized by summing the number of deterministic simulations that output snowfall exceeding a given amount. Ensemble relative frequencies for a selected variable

Table 2. Summary of Skill Scores From Various WRF Model Settings for Snowfall Detection

	WSM		Goddard		Thompson		Morrison		Ensemble
	YSU	MYJ	YSU	MYJ	YSU	MYJ	YSU	MYJ	Average
Snow Water Equivalent, 1 mm									
POD	0.92	0.92	0.92	0.92	0.92	0.93	0.86	0.89	0.94
FOM	0.08	0.08	0.08	0.08	0.08	0.07	0.14	0.11	0.06
PCR	0.61	0.51	0.51	0.52	0.52	0.49	0.57	0.58	0.54
FAR	0.39	0.49	0.49	0.48	0.48	0.51	0.43	0.42	0.46
BS	0.08	0.09	0.09	0.09	0.10	0.08	0.17	0.12	0.07
Snow Water Equivalent, 2 mm									
POD	0.85	0.87	0.88	0.88	0.86	0.88	0.83	0.84	0.90
FOM	0.15	0.13	0.12	0.12	0.14	0.12	0.17	0.16	0.10
PCR	0.72	0.63	0.67	0.66	0.69	0.64	0.72	0.74	0.69
FAR	0.28	0.37	0.33	0.34	0.31	0.36	0.28	0.26	0.31
BS	0.15	0.13	0.11	0.11	0.14	0.12	0.19	0.16	0.09
Snow Depth, 1 cm									
POD	0.89	0.90	0.92	0.90	0.89	0.91	0.86	0.86	0.91
FOM	0.11	0.10	0.08	0.10	0.11	0.09	0.14	0.14	0.09
PCR	0.66	0.55	0.60	0.60	0.64	0.60	0.67	0.63	0.63
FAR	0.34	0.45	0.40	0.40	0.36	0.40	0.33	0.37	0.37
BS	0.11	0.11	0.08	0.10	0.11	0.09	0.14	0.15	0.08
Snow Depth, 2 cm									
POD	0.76	0.83	0.84	0.83	0.80	0.85	0.74	0.77	0.82
FOM	0.24	0.17	0.16	0.17	0.20	0.15	0.26	0.23	0.18
PCR	0.88	0.72	0.75	0.70	0.77	0.75	0.78	0.85	0.75
FAR	0.12	0.28	0.25	0.30	0.23	0.25	0.22	0.15	0.25
BS	0.29	0.18	0.16	0.19	0.24	0.15	0.36	0.26	0.19

corresponded closely with the estimated probability. Figure 7 shows the probability of snow water equivalent amounts greater than 1, 2, and 10 mm for 13 March 2013.

Figure 7 portrays average snow water equivalent estimated by the eight deterministic simulations, together with the standard deviation. The spread between the eight deterministic outcomes gives some idea of model reliability, i.e., the less spread, the greater the reliability. As in the case of total precipitation, there is greater spread between the eight deterministic outcomes in the Cantabrian Range, with less spread in other mountainous regions and very consistent results over the plateau.

Finally, minimum and maximum snow water equivalent generated by the eight deterministic simulations are depicted in Figure 7. Although they are not the most likely, it is very useful to know maximum values of expected snowfall, to be ready and take measures in case those values are reached.

5.4. Snowfall Validation

Finally, in the snowfall episode assessment, we analyzed accuracy of the various deterministic simulations and ensemble average. Skill scores used were the POD, FOM, PCR, FAR, and BS. We also examined the simulated WRF variables snow water equivalent and snow depth, to determine which was more reliable in establishing the snowfall threshold causing traffic problems (evaluated using snowplow requests). We focused here on finer-resolution D03, because it is known that low spatial resolutions underestimate snowfall by 15%–30% because of terrain smoothing and consequent poor representation of mesoscale processes [*Leung et al.*, 2003; *Ikeda et al.*, 2010].

As seen in Table 2, we obtained very satisfactory values of POD and FOM, but PCR and FAR show room for improvement. One reason for those low scores may be the enormous difficulty in obtaining reliable data about absence of snowfall and resultant small database for the study locations. The database of snowfall locations

is more abundant, providing the model with sufficient samples for evaluation. The overall database produced satisfactory results using the BS.

The snow water equivalent appeared to be a more reliable variable than snow depth, as demonstrated by higher skill scores. The largest BS was attained by setting the threshold to 1 mm. However, the PCR improved markedly upon using a 2 mm threshold, with FOM and FAR values more balanced. The threshold could be established based on the purpose for which the model is used.

Among deterministic simulations, the Morrison scheme had a higher PCR than the Goddard and Thompson parameterizations but a larger error in terms of FOM. This may be attributable to the higher percentage of liquid water relative to the solid phase in the Morrison parameterization compared to the Thompson and Goddard. A combination of WSM and YSU schemes produced very satisfactory values of PCR and FAR, possibly owing to the higher temperature predicted by these parameterizations. There were also very satisfactory results for POD, with the second best BS. The best deterministic simulations used the Thompson microphysical and MYJ schemes (as in the case of total precipitation amount), followed by the setting with the WSM parameterization and YSU scheme. The major difference between snow and rain is the amount of cooling of rain droplets necessary to generate ice particles, which is a function of the various microphysical parameterizations used. These results agree with those of *McMillen and James Steenburgh* [2015], who claimed that the Thompson microphysical scheme estimates areal coverage and amount of precipitation as snow more accurately than the Goddard and Morrison schemes, likely owing to more accurate calculation of graupel and snow production rates.

Regarding PBL schemes, there was no appreciable influence of the Goddard simulations. Nevertheless, with the Thompson and Morrison microphysical schemes, there was clear improvement in FOM reduction using the MYJ scheme. This may be due to the stronger thermal inversion layers simulated by this scheme relative to the YSU. This could more accurately simulate snowfall, in that a strong thermal inversion layer at low levels facilitates snow reaching the surface. However, the YSU scheme had a better PCR (especially for the WSM and Thompson schemes), which might be due to the higher temperature simulated by that PBL scheme. This suggests that in this case, the colder temperatures simulated by the MYJ scheme could have overestimated the presence of snow instead of rain.

The BS shows that the skill of the probabilistic outcome was markedly better than those of the deterministic simulations. The combination of the eight deterministic outcomes in an ensemble yielded higher skill scores. Adding physics variability improved model ability to provide model uncertainty, as described in *Du et al.* [2004]. In summary, it is worth making an ensemble simulation, because it notably improves the results of each deterministic outcome, removing part of the individual errors. The developed ensemble permits a total error <10%. This is considered very satisfactory, especially given the enormous difficulty of snowfall estimation.

Despite it being desirable to reduce the FAR, in this study it is more important to minimize the FOM because the purpose is to avoid risks. If the model estimates snow but it does not snow, it could increase the cost of road maintenance. However, if the model determines no snow and it does snow, it could have dramatic consequences, for example, by increasing the number of traffic accidents because drivers are not warned in advance of the danger of slippery roads.

We wanted to do postprocessing of the snowfall estimates, but the scarce database of snowfall episodes hampered the development of accurate multiple linear regression.

6. Conclusions

This research addressed the enormous difficulty of simulating snowfall episodes with numerical weather prediction models. We compared the results of four different microphysical parameterizations and two PBL schemes, finding that there is no unique solution for every synoptic situation. For the 15 snowfall days analyzed, a combination of the Thompson microphysical parameterization and MYJ scheme was the most accurate in terms of precipitation spatial distribution and smaller RMSD of precipitation amount, relative to the precipitation gauge network database. The same occurred when we analyzed the ability of the various settings to distinguish snow and rain. However, the YSU PBL scheme is more accurate than MYJ regarding temporal sensitivity.

The accuracy of the precipitation and snowfall outcomes increased significantly by combining the eight deterministic simulations in an ensemble average, especially for spatial distributions. Moreover, the spread expressed by standard deviation between the eight individual outcomes provides useful information about model reliability. Furthermore, the method gives probabilistic information about risk of a given amount of precipitation or snowfall predicted.

Multiple linear regression was used as postprocessing, achieving a substantial reduction of RMSD. In our case, this postprocessing was especially useful for reducing systematic overestimation of precipitation amount in the study area, exercising the YSU and WSM parameterization options of the WRF model.

Improvement of model accuracy by increasing its spatial resolution was substantial in the step from 27 to 9 km but weak from 9 to 3 km. The next targets should be focused on determining the optimal horizontal resolution of the WRF model, with regard to both forecasting accuracy and computational cost. The estimation of these parameters will require the availability of a network with higher density of observations, as well as instrumentation more accurate in the measure of snow water equivalent. For the current study we used the most reliable instrumentation existing in the study area, although we are aware of their limitations in measuring snowfall under extreme weather conditions. In this regard, we are studying the possibility of introducing in the study area additional instrumentation (such as snow pillows) for future researches.

Operational use of this tool would be helpful in making decisions for various human activities, mainly in the management of snowplows. Current information offered in Spain is based on melting level and is generally incomplete and inaccurate. The tool can give detailed information about the exact location of snowfall expected, at high spatial (3 km) and temporal (1 h) resolution. The accuracy of this information would reduce management costs and optimize the use of salt, equipment, and personnel.

Differences in precipitation distribution and amount simulated by various PBL and microphysics schemes reveal uncertainty present in these formulations. This necessitates improvement of these parameterizations, in particular by enhancing the modeling of various hydrometeors and microphysical processes that interact.

Future work may be directed at increasing the spread of the ensemble simulation by perturbing initial conditions. It is also possible that other microphysical and PBL schemes will be evaluated. Furthermore, we believe that the results of this study might be extrapolated to other similar climatic regions at midlatitudes. Therefore, we intend to expand the study area in future studies. Finally, we anticipate applying this method in the warm season to evaluate its accuracy during convective episodes. We expect that the accuracy of our method will increase in the analysis of weather phenomena other than snowfall. This is because of the enormous difficulties in the measurement of snowfall and their simulation by numerical weather models.

Acknowledgments

Data support was from the Atmospheric Physics Group (University of León, Spain), Duero Hydrographic Confederation, Spanish Meteorological Agency, and the emergency service of Spain. To request the data, please contact S. Fernández-González (sefern04@ucm.es). This paper was supported by the following grants: TEcoAgua, METEORISK PROJECT(RTC-2014-1872-5), Granimetro(CGL2010-15930) and MINECO(CGL2011-25327, RTC-2014-1872-5 and ESP2013-47816-C4-4P), and LE220A11-2 and LE003B009 awarded by the Junta de Castilla and León. Special thanks to Roberto Weigand, Ángel Guerrero, Steven Hunter, and Analisa Weston. S. Fernández-González acknowledges the grant support from the FPU program (AP 2010-2093).

References

Alhamed, A., S. Lakshmivarahan, and D. Stensrud (2002), Cluster analysis of multimodel ensemble data from SAMEX, *Mon. Weather Rev.*, *130*(2), 226–256.

Arakawa, A. (2004), The cumulus parameterization problem: Past, present, and future, *J. Clim.*, *17*(13), 2493–2525, doi:10.1175/1520-0442(2004)017<2493:RATCPP>2.0.CO;2.

Braun, S., and W.-K. Tao (2000), Sensitivity of high-resolution simulations of Hurricane Bob (1991) to planetary boundary layer parameterizations, *Mon. Weather Rev.*, *128*(12), 3941–3961.

Bremnes, J. (2004), Probabilistic forecasts of precipitation in terms of quantiles using NWP model output, *Mon. Weather Rev.*, *132*(1), 338–347.

Brier, G. (1950), Verification of forecasts expressed in terms of probability, *Mon. Weather Rev.*, *78*, 1–3.

Buizza, R., M. Miller, and T. Palmer (1999), Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System, *Q. J. R. Meteorol. Soc.*, *125*(560), 2887–2908.

Caramelo, L., and M. D. Manso-Orgaz (2007), A study of precipitation variability in the Duero Basin (Iberian Peninsula), *Int. J. Climatol.*, *27*, 327–339, doi:10.1002/joc.1403.

Chen, F., and J. Dudhia (2001), Coupling and advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: Model implementation and sensitivity, *Mon. Weather Rev.*, *129*(4), 569–585.

Colli, M., L. Lanza, P. La Barbera, and P. Chan (2014), Measurement accuracy of weighing and tipping-bucket rainfall intensity gauges under dynamic laboratory testing, *Atmos. Res.*, *144*, 186–194, doi:10.1016/j.atmosres.2013.08.007.

Datla, S., and S. Sharma (2008), Impact of cold and snow on temporal and spatial variations of highway traffic volumes, *J. Transport Geogr.*, *16*(5), 358–372, doi:10.1016/j.jtrangeo.2007.12.003.

Du, J., et al. (2004), The NOAA/NWS/NCEP Short Range Ensemble Forecast (SREF) system: Evaluation of an initial condition versus multiple model physics ensemble approach, in *16th Conference on Numerical Prediction*, p. 21.3, Am. Meteorol. Soc., Seattle, Wash.

Dudhia, J. (1989), Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model, *J. Atmos. Sci.*, *46*(20), 3077–3107.

Eckel, F., and C. Mass (2005), Aspects of effective mesoscale, short-range ensemble forecasting, *Weather Forecasting*, *20*(3), 328–350, doi:10.1175/WAF843.1.

Erickson, M., B. Colle, and J. Charney (2012), Impact of bias-correction type and conditional training on Bayesian model averaging over the Northeast United States, *Weather Forecasting*, *27*(6), 1449–1469, doi:10.1175/WAF-D-11-00149.1.
Evans, J., M. Ekström, and F. Ji (2012), Evaluating the performance of a WRF physics ensemble over South-East Australia, *Clim. Dyn.*, *39*(6), 1241–1258, doi:10.1007/s00382-011-1244-5.
Evans, R., M. Harrison, R. Graham, and K. Mylne (2000), Joint medium-range ensembles from the Met. Office and ECMWF systems, *Mon. Weather Rev.*, *128*(9), 3104–3127.
Fargey, S., W. Henson, J. Hanesiak, and R. Goodson (2014), Characterization of an unexpected snowfall event in Iqaluit, Nunavut, and surrounding area during the Storm Studies in the Arctic field project, *J. Geophys. Res. Atmos.*, *119*(9), 5492–5511, doi:10.1002/2013JD021176.
Fernández-González, S., F. Valero, J. L. Sánchez, E. Gascón, L. López, E. García-Ortega, and A. Merino (2014), Observation of a freezing drizzle episode: A case study, *Atmos. Res.*, *149*, 244–254, doi:10.1016/j.atmosres.2014.06.014.
Fernández-González, S., F. Valero, J. L. Sánchez, E. Gascón, L. López, E. García-Ortega, and A. Merino (2015), Analysis of a seeder-feeder and freezing drizzle event, *J. Geophys. Res. Atmos.*, *120*, 3,984–3,999, doi:10.1002/2014JD022916.
Gallus, W., Jr., and J. Bresch (2006), Comparison of impacts of WRF dynamic core, physics package, and initial conditions on warm season rainfall forecasts, *Mon. Weather Rev.*, *134*(9), 2632–2641, doi:10.1175/MWR3198.1.
Gao, Y., J. Xu, and D. Chen (2015), Evaluation of WRF Mesoscale Climate Simulations over the Tibetan Plateau during 1979–2011, *J. Clim.*, *28*, 2823–2841, doi:10.1175/JCLI-D-14-00300.1.
Gascón, E., J. Sánchez, S. Fernández-González, L. Hermida, L. López, E. García-Ortega, and A. Merino (2015), Monitoring a convective winter episode of the Iberian Peninsula using a multichannel microwave radiometer, *J. Geophys. Res. Atmos.*, *120*(4), 1565–1581, doi:10.1002/2014JD022510.
Hamill, T. M., and S. J. Colucci (1998), Evaluation of Eta-RSM ensemble probabilistic rainfall forecasts, *Mon. Weather Rev.*, *126*, 711–724.
Harrison, M., T. Palmer, D. Richardson, and R. Buizza (1999), Analysis and model dependencies in medium-range ensembles: Two transplant case-studies, *Q. J. R. Meteorol. Soc.*, *125*(559), 2487–2515.
Homar, V., D. J. Stensrud, J. J. Levit, and D. R. Bright (2006), Value of human-generated perturbations in short-range ensemble forecasts of severe weather, *Weather Forecasting*, *21*(3), 347–363.
Hong, S.-Y., and J.-O. J. Lim (2006), The WRF Single-Moment 6-Class Microphysics Scheme (WSM6), *J. Korean Meteorol. Soc.*, *42*, 129–151.
Hong, S.-Y., J. Dudhia, and S.-H. Chen (2004), A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation, *Mon. Weather Rev.*, *132*(1), 103–120.
Hong, S.-Y., Y. Noh, and J. Dudhia (2006), A new vertical diffusion package with an explicit treatment of entrainment processes, *Mon. Weather Rev.*, *134*(9), 2318–2341, doi:10.1175/MWR3199.1.
Huang, G.-J., V. Bringi, D. Moisseev, W. Petersen, L. Bliven, and D. Hudak (2015), Use of 2D-video disdrometer to derive mean density-size and Ze-SR relations: Four snow cases from the light precipitation validation experiment, *Atmos. Res.*, *153*, 34–48, doi:10.1016/j.atmosres.2014.07.013.
Iguchi, T., T. Matsui, J. Shi, W.-K. Tao, A. Khain, A. Hou, R. Cifelli, A. Heymsfield, and A. Tokay (2012), Numerical analysis using WRF-SBM for the cloud microphysical structures in the C3VP field campaign: Impacts of supercooled droplets and resultant riming on snow microphysics, *J. Geophys. Res.*, *117*, D23206, doi:10.1029/2012JD018101.
Ikeda, K., et al. (2010), Simulation of seasonal snowfall over Colorado, *Atmos. Res.*, *97*(4), 462–477, doi:10.1016/j.atmosres.2010.04.010.
Janjic, Z. (1996), *The Surface Layer Parameterization in the NCEP Eta Model*, 444 pp., World Meteorol. Organ., Geneva, Switzerland.
Janjic, Z. (2001), Nonsingular implementation of the Mellor-Yamada level 2.5 scheme in the NCEP Meso model, in *Note 437*, p. 61, NCEP Office. [Available at http://www.emc.ncep.noaa.gov/officenotes/FullTOC.html.]
Jankov, I., W. Gallus Jr., M. Segal, B. Shaw, and S. Koch (2005), The impact of different WRF model physical parameterizations and their interactions on warm season MCS rainfall, *Weather Forecasting*, *20*(6), 1048–1060, doi:10.1175/WAF888.1.
Ji, F., M. Ekström, J. Evans, and J. Teng (2014), Evaluating rainfall patterns using physics scheme ensembles from a regional atmospheric model, *Theor. Appl. Climatol.*, *115*(1–2), 297–304, doi:10.1007/s00704-013-0904-2.
Johnson, A., and X. Wang (2012), Verification and calibration of neighborhood and object-based probabilistic precipitation forecasts from a multimodel convection-allowing ensemble, *Mon. Weather Rev.*, *140*(9), 3054–3077, doi:10.1175/MWR-D-11-00356.1.
Kain, J. (2004), The Kain-Fritsch convective parameterization: An update, *J. Appl. Meteorol.*, *43*(1), 170–181.
Kim, J.-H., D.-B. Shin, and C. Kummerow (2013), Impacts of a priori databases using six WRF microphysics schemes on passive microwave rainfall retrievals, *J. Atmos. Oceanic Technol.*, *30*(10), 2367–2381, doi:10.1175/JTECH-D-12-00261.1.
Krishnamurti, T. N., C. M. Kishtawal, T. E. LaRow, D. R. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran (2013), Improved weather and seasonal climate forecasts from multimodel superensemble, *Science*, *285*, 1548–1550, doi:10.1126/science.285.5433.1548.
Lang, S., W.-K. Tao, R. Cifelli, W. Olson, J. Halverson, S. Rutledge, and J. Simpson (2007), Improving simulations of convective systems from TRMM LBA: Easterly and westerly regimes, *J. Atmos. Sci.*, *64*(4), 1141–1164, doi:10.1175/JAS3879.1.
Lauret, P., M. Diagne, and M. David (2014), A neural network post-processing approach to improving NWP solar radiation forecasts, *Energy Procedia*, *57*, 1044–1052, doi:10.1016/j.egypro.2014.10.089.
Leung, L., Y. Qian, J. Han, and J. Roads (2003), Intercomparison of global reanalyses and regional simulations of cold season water budgets in the western United States, *J. Hydrometeorol.*, *4*(6), 1067–1087, doi:10.1175/1525-7541(2003)004<1067:IOGRAR>2.0.CO;2.
Li, X., W.-K. Tao, A. Khain, J. Simpson, and D. Johnson (2009), Sensitivity of a cloud-resolving model to bulk and explicit bin microphysical schemes. Part I: Comparisons, *J. Atmos. Sci.*, *66*(1), 3–21, doi:10.1175/2008JAS2646.1.
Liang, X.-Z., et al. (2012), Regional Climate-Weather Research and Forecasting Model (CWRF), *Bull. Am. Meteorol. Soc.*, *93*, 1363–1387, doi:10.1175/2008JAS2646.1.
Lim, K., and S.-Y. Hong (2010), Development of an effective double-moment cloud microphysics scheme with prognostic cloud condensation nuclei (CCN) for weather and climate models, *Mon. Weather Rev.*, *138*(5), 1587–1612, doi:10.1175/2009MWR2968.1.
López, L., E. García-Ortega, and J. Sánchez (2007), A short-term forecast model for hail, *Atmos. Res.*, *83*(2–4), 176–184, doi:10.1016/j.atmosres.2005.10.014.
López-Moreno, J. I., and J. M. García-Ruiz (2004), Influence of snow accumulation and snowmelt on streamflow in the central Spanish Pyrenees, *Hydrol. Sci.*, *49*(5), 787–802, doi:10.1623/hysj.49.5.787.55135.
Lu, C., H. Yuan, B. Schwartz, and S. Benjamin (2007), Short-range numerical weather prediction using time-lagged ensembles, *Weather Forecasting*, *22*(3), 580–595, doi:10.1175/WAF999.1.
McCumber, M., W.-K. Tao, J. Simpson, R. Penc, and S.-T. Soong (1991), Comparison of ice-phase microphysical parameterization schemes using numerical simulations of tropical convection, *J. Appl. Meteorol.*, *30*(7), 985–1004.
McMillen, J., and W. James Steenburgh (2015), Impact of microphysics parameterizations on simulations of the 27 October 2010 great Salt Lake-effect snowstorm, *Weather Forecasting*, *30*(1), 136–152, doi:10.1175/WAF-D-14-00060.1.

Merino, A., S. Fernández, L. Hermida, L. López, J. Sánchez, E. García-Ortega, and E. Gascón (2014), Snowfall in the northwest Iberian Peninsula: Synoptic circulation patterns and their influence on snow day trends, *Sci. World J.*, *2014*, 480,275, doi:10.1155/2014/480275.

Mlawer, E., S. Taubman, P. Brown, M. Iacono, and S. Clough (1997), Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-*k* model for the longwave, *J. Geophys. Res.*, *102*(14), 16663–16682.

Molteni, F., R. Buizza, T. Palmer, and T. Petroliagis (1996), The ECMWF ensemble prediction system: Methodology and validation, *Q. J. R. Meteorol. Soc.*, *122*(529), 73–119.

Molthan, A., and B. Colle (2012), Comparisons of single- and double-moment microphysics schemes in the simulation of a synoptic-scale snowfall event, *Mon. Weather Rev.*, *140*(9), 2982–3002, doi:10.1175/MWR-D-11-00292.1.

Morán-Tejeda, E., A. Ceballos-Barbancho, and J. M. Llorente-Pinto (2010), Hydrological response of Mediterranean headwaters to climate oscillations and land-cover changes: The mountains of Duero River basin (Central Spain), *Global Planet. Change*, *72*, 39–49, doi:10.1016/j.gloplacha.2010.03.003.

Morán-Tejeda, E., J. López-Moreno, A. Ceballos-Barbancho, and S. Vicente-Serrano (2011), River regimes and recent hydrological changes in the Duero basin (Spain), *J. Hydrol.*, *404*(3–4), 241–258, doi:10.1016/j.jhydrol.2011.04.034.

Morland, J., and C. Mätzler (2007), Spatial interpolation of GPS integrated water vapour measurements made in the Swiss Alps, *Meteorol. Appl.*, *14*(1), 15–26, doi:10.1002/met.2.

Morrison, H., J. Curry, and V. Khvorostyanov (2005), A new double-moment microphysics parameterization for application in cloud and climate models. Part I: Description, *J. Atmos. Sci.*, *62*(6), 1665–1677, doi:10.1175/JAS3446.1.

Morrison, H., G. Thompson, and V. Tatarskii (2009), Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes, *Mon. Weather Rev.*, *137*(3), 991–1007, doi:10.1175/2008MWR2556.1.

Pennelly, C., G. Reuter, and T. Flesch (2014), Verification of the WRF model for simulating heavy precipitation in Alberta, *Atmos. Res.*, *135-136*, 172–192, doi:10.1016/j.atmosres.2013.09.004.

Pielke, R. A., Jr., and M. Downton (2000), Precipitation and damaging floods: Trends in the United States, 1932–97, *J. Clim.*, *13*(20), 3625–3637.

Rasmussen, R., et al. (2012), How well are we measuring snow: The NOAA/FAA/NCAR winter precipitation test bed, *Bull. Am. Meteorol. Soc.*, *93*(6), 811–829, doi:10.1175/BAMS-D-11-00052.1.

Rutledge, S., and P. Hobbs (1984), Mesoscale and microscale structure and organization of clouds and precipitation in midlatitude cyclones. XII: A diagnostic modeling study of precipitation development in narrow cold-frontal rainbands, *J. Atmos. Sci.*, *41*(20), 2949–2972.

Schwartz, C., J. Kain, S. Weiss, M. Xue, D. Bright, F. Kong, K. Thomas, J. Levit, M. Coniglio, and M. Wandishin (2010), Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership, *Weather Forecasting*, *25*(1), 263–280, doi:10.1175/2009WAF2222267.1.

Skamarock, W., and J. Klemp (2008), A time-split nonhydrostatic atmospheric model for weather research and forecasting applications, *J. Comput. Phys.*, *227*(7), 3465–3485, doi:10.1016/j.jcp.2007.01.037.

Stensrud, D., and M. Wandishin (2000), The correspondence ratio in forecast evaluation, *Weather Forecasting*, *15*(5), 593–602.

Stensrud, D., J.-W. Bao, and T. Warner (2000), Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems, *Mon. Weather Rev.*, *128*, 2077–2107.

Tao, W.-K., J. Simpson, and M. McCumber (1989), An ice-water saturation adjustment, *Mon. Weather Rev.*, *117*, 231–235.

Tao, W.-K., et al. (2009), The Goddard multi-scale modeling system with unified physics, *Ann. Geophys.*, *27*(8), 3055–3064, doi:10.5194/angeo-27-3055-2009.

Taylor, K. (2001), Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.*, *106*(D7), 7183–7192.

Thompson, G., R. Rasmussen, and K. Manning (2004), Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part I: Description and sensitivity analysis, *Mon. Weather Rev.*, *132*(2), 519–542.

Thompson, G., P. Field, R. Rasmussen, and W. Hall (2008), Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization, *Mon. Weather Rev.*, *136*(12), 5095–5115, doi:10.1175/2008MWR2387.1.

Tracton, M., and E. Kalnay (1993), Operational ensemble prediction at the National Meteorological Center: Practical aspects, *Weather Forecasting*, *8*(3), 379–398.

Villarini, G., P. Mandapaka, W. Krajewski, and R. Moore (2008), Rainfall and sampling uncertainties: A rain gauge perspective, *J. Geophys. Res.*, *113*, D11102, doi:10.1029/2007JD009214.

Walko, R., W. Cotton, M. Meyers, and J. Harrington (1995), New RAMS cloud microphysics parameterization. Part I: The single-moment scheme, *Atmos. Res.*, *38*(1–4), 29–62, doi:10.1016/0169-8095(94)00087-T.

Wang, W., and N. L. Seaman (1997), A comparison study of convective parameterization schemes in a mesoscale model, *Mon. Weather Rev.*, *125*, 252–278, doi:10.1175/1520-0493(1997)125<0252:ACSOCP>2.0.CO;2.

Weisman, M., C. Davis, W. Wang, K. Manning, and J. Klemp (2008), Experiences with 0–36-h explicit convective forecasts with the WRF-ARW model, *Weather Forecasting*, *23*(3), 407–437, doi:10.1175/2007WAF2007005.1.

Wilks, D. (2009), Extending logistic regression to provide full-probability-distribution MOS forecasts, *Meteorol. Appl.*, *16*(3), 361–368, doi:10.1002/met.134.

Wilson, T. (2007), The forecast accuracy of Australian Bureau of Statistics national population projections, *J. Population Res.*, *24*(1), 91–117.

Wolff, M., K. Isaksen, A. Petersen-Øverleir, K. Ødemark, T. Reitan, and R. Brækkan (2015), Derivation of a new continuous adjustment function for correcting wind-induced loss of solid precipitation: Results of a Norwegian field study, *Hydrol. Earth Syst. Sci.*, *19*(2), 951–967, doi:10.5194/hess-19-951-2015.

Yuan, H., J. McGinley, P. Schultz, C. Anderson, and C. Lu (2008), Short-range precipitation forecasts from time-lagged multimodel ensembles during the HMT-west-2006 campaign, *J. Hydrometeorol.*, *9*(3), 477–491, doi:10.1175/2007JHM879.1.

Yuan, X., and E. F. Wood (2012), On the clustering of climate models in ensemble seasonal forecasting, *Geophys. Res. Lett.*, *39*, L18701, doi:10.1029/2012GL052735.

Yuan, X., X.-Z. Liang, and E. F. Wood (2012), WRF ensemble downscaling seasonal forecasts of China winter precipitation during 1982–2008, *Clim. Dyn.*, *39*, 2041–2058, doi:10.1007/s00382-011-1241-8.