Statistical and dynamical downscaling of precipitation over Spain from DEMETER seasonal forecasts

By E. DÍEZ[1], C. PRIMO[2], J. A. GARCÍA-MOYA[1], J. M. GUTIÉRREZ[3] and B. ORFILA[1]*,
[1]*Instituto Nacional de Meteorología (INM), C/Leonardo Prieto Castro 8, 28040 Madrid, Spain;* [2]*Instituto Nacional de Meteorología, CMT/CAS, Santander, Spain;* [3]*Dept. of Applied Mathematics and Computer Science. University of Cantabria, 39005 Santander, Spain*

ABSTRACT

Statistical and dynamical downscaling methods are tested and compared for downscaling seasonal precipitation forecasts over Spain from two DEMETER models: the European Centre for Medium-Range Weather Forecasts (ECMWF) and the UK Meteorological Office (UKMO). The statistical method considered is a particular implementation of the standard analogue technique, based on close neighbours of the predicted atmospheric geopotential and humidity fields. Dynamical downscaling is performed using the Rossby Centre Climate Atmospheric model, which has been nested to the ECMWF model output, and run in climate mode for six months. We first check the performance of the direct output models in the period 1986–1997 and compare it with the results obtained applying the analogue method. We have found that the direct outputs underestimate the precipitation amount and that the statistical downscaling method improves the results as the skill of the direct forecast increases. The highest skills – relative operating characteristic skill areas (RSAs) above 0.6 – are associated with early and late spring, summer and autumn seasons at zero- and one-month lead times. On the other hand, models have poor skill during winter with the exception of the El Niño period (1986–1988), especially in the south of Spain. In this case, high RSAs and economic values have been found. We also compare statistical and dynamical downscaling during four seasons, obtaining no concluding result. Both methods outperform direct output from DEMETER models, but depending on the season and on the region of Spain one method is better than the other. Moreover, we have seen that dynamical and statistical methods can be used in combination, yielding the best skill scores in some cases of the study.

1. Introduction

One of the main objectives of DEMETER (Development of a European Multi-model Ensemble system for seasonal to inTER-annual prediction) is to demonstrate the utility of seasonal climate forecasts for different end-users, such as crop yield and malaria modellers (Palmer et al., 2004). However, existing application models typically require weather input (precipitation, temperature, wind, radiation, etc.) on a spatial scale of a substantially higher resolution than that of the DEMETER climate models (see, for example, Doblas-Reyes et al., 2005; Hagedorn et al., 2005). Moreover, the statistics of soil variables derived from the model can differ substantially from the local statistics of station data (e.g. precipitation in regions of steep orography). Therefore, in order to make DEMETER outputs useful for end-users it is necessary to perform some form of downscaling process

to the climate model output, either by some statistical/empirical scheme or by nesting a high-resolution limited-area model to the coarser resolution global climate model (GCM; Goddard et al., 2001).

Statistical downscaling methods work by mapping one or more large-scale fields from a reanalysis project to the simultaneous records of fine-scale observations required by application models. Standard techniques include global mappings derived by regression methods or artificial neural networks (see Wilby and Wigley, 1997, for a review). These methods are relatively straightforward to apply and are computationally cheap. The main problem in terms of the skill of the statistical methods is the implicit assumption that the future climate variability is the same as the past climate variability and, in particular, the same as in the reanalysis period from which the statistical relationship is derived; this leads to stationary models which cannot take into account trends, or fluctuations. This problem is partially addressed by local downscaling algorithms, such as the method of analogues, where different relationships between model outputs and fine-scale variables are locally established considering

*Corresponding author.
e-mail: orfila@inm.es

neighbourhoods of patterns in the reanalysis data base (see Zorita and von Storch, 1999, for more details).

Dynamical downscaling models work by nesting a high-resolution model to the GCM in areas of interest. These methods do not require local observed data and have the potential to outperform statistical methods, particularly regarding the prediction of extreme events. However, there are outstanding problems, such as propagation of systematic biases from the global to the regional model (Giorgi et al., 2001). In addition, the computational expense of running a high-resolution regional climate model and the requested data storage can be comparable to that of running the global seasonal forecast model. In spite of this, regional climate models have been applied to study local effects of global climate change resulting from increasing concentration of greenhouse gases in the atmosphere (Christensen et al., 2001). By comparison, little has been done for the seasonal scales (Misra et al., 2003).

The main goal of this paper is to explore the skill of the seasonal DEMETER forecast of precipitation over Spain, and to assess the feasibility of statistical and dynamical downscaling techniques applied to global direct model outputs to gain detail in the precipitation forecasts. Both statistical methods and regional dynamical climate models are tested and compared. The statistical method is a particular implementation of the standard analogue technique based on close neighbours of the predicted atmospheric geopotential and humidity patterns. This method is applied to the outputs of the European Centre for Medium-Range Weather Forecast (ECMWF) and the UK Meteorological Office (UKMO) DEMETER models (nine + nine ensemble members) for the period 1986–1997; the resulting precipitation downscaled forecasts are compared with the direct outputs, and the skill for each season is computed. On the other hand, dynamical downscaling is performed using the Rossby Centre Climate Atmospheric (RCA) model, which is nested to the ECMWF model output, and run in climate mode for six months. Due to the computational cost, we could only perform a limited number of regional experiments to test the skill of dynamical downscaling methods. In particular, we ran a 0.5° resolution RCA model with a three-member ensemble during the period 1986–1989, and a three-member ensemble with a 0.2° resolution RCA model only for the six-month period starting in November 1986 (four seasons). These experiments allow for an overall comparison of the two direct model outputs (global and regional) and the statistical method, for the period 1986–1989. However, the skill of the two RCA dynamic models with different resolutions can be compared only for four seasons corresponding to a single integration time (in this last case, no general conclusion can be obtained from the analysis).

To represent the spatial and overall performance of these methods, different scatter diagrams, plots and maps of precipitation over Spain have been used in the paper. Moreover, standard validation techniques, such as the relative operating characteristic (ROC) skill area (RSA) and economic value curves, have been

used to compare different deterministic and probabilistic forecasts. A special verification index based on quintile distances between predictions and observations is used to quantify the spatial performance of the methods. All these graphical and verification tools are referred to or briefly described in the paper.

In Section 2 we describe the climate networks and the interpolated mesh of observations used in this work. In Section 3 we describe the visualization and diagnostic tools used in the analysis. In Section 4 we give a general introduction to analogue methods and describe in detail the two-step technique used in the paper. In Section 5 we describe the experiments undertaken with the regional RCA model. A comparison of dynamical and statistical downscaling results is given in Section 6. Conclusions are found in Section 7.

2. Data and observation networks

The spatial and temporal variability of precipitation in Spain can only be appropriately characterized using a high-resolution network. The main climatological network (see Fig. 1a) has a poor resolution for this task. The pluviometric network began to be set up in 1913, and nowadays the number of operative stations is about 4000, covering the main river basins (see Fig. 1b). However, this network is inhomogeneous and observations contain missing data; hence, it is not suitable for validation purposes. Figure 1c shows a 203-point mesh obtained by interpolating and homogenizing the pluviometric network, yielding a 47-yr data set (1961–1997; Fernández et al., 2001) which is used through the paper for verification purposes. Figures 1d and e show the 0.2° and 0.5° grid resolutions used for the regional RCA model.

In this work we use the output of the DEMETER project, which consists of seven different global coupled atmosphere–ocean models. The objective of this project was to develop a multi-model ensemble forecast system for reliable seasonal to interannual prediction. These models run from sets of initial conditions, each of them slightly different from each other, but consistent with the available observations. For each single model, nine-member ensemble integrations run four times per year (1 February, 1 May, 1 August and 1 November, respectively) over six-month periods. In this work, we only consider the ECMWF and UKMO models. The ECMWF model is a coupled GCM with a resolution of 1.8°. On the other hand, the UKMO model is based on the HadCM3 climate model (Gordon et al., 2000) with a 2.5° latitude by 3.75° longitude (see http://www.ecmwf.int/research/demeter for more details about these models).

3. Display and validation of spatial forecasts

Ensemble prediction systems provide both probabilistic and deterministic seasonal forecasts (for instance, the mean of the ensemble is usually considered a deterministic forecast). When the prediction is spatially extended over a network, or mesh, of
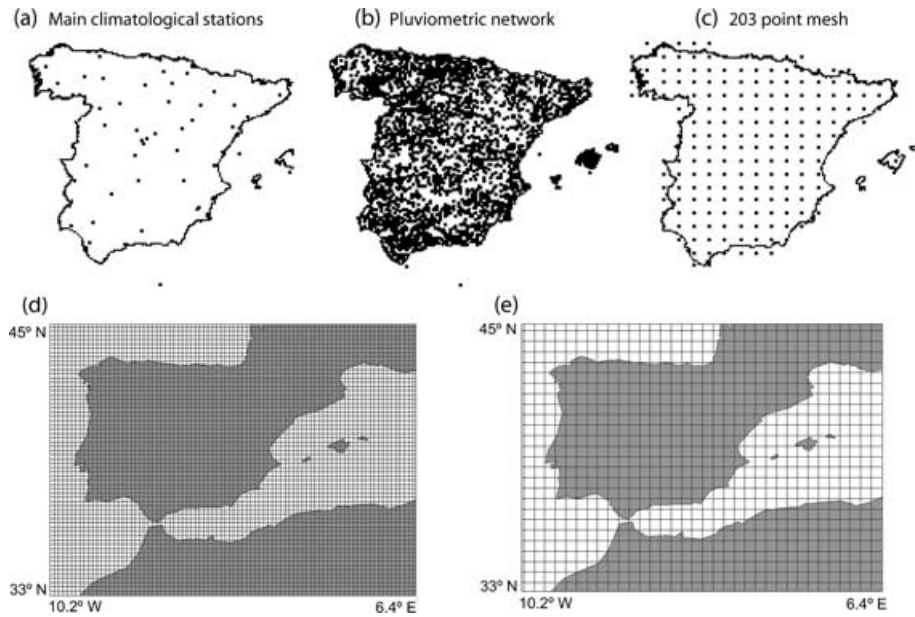
Fig. 1. (a) Spanish main climatological network, (b) pluviometric network and (c) 203-point mesh. (d), (e) 0.2° and 0.5° grid resolutions of the regional RCA model over the Iberian peninsula.

stations, then a deterministic precipitation forecast can be displayed as a map of accumulated precipitation, a map of anomalies, or a map of terciles reflecting the dry/normal/wet character of the season. Validation can be performed by comparing the seasonal forecasts with the accumulated precipitation observed values at each station, obtaining maps of errors, or maps of percentiles by which the precipitation and observation differ, regarding the climatology. The percentage of Spanish area below a certain error, or with a difference of percentiles less than a certain amount, is useful to estimate the overall quality of the forecast over Spain. In particular, we consider a difference of 20 percentiles, i.e. a running quintile, to discriminate between accurate and wrong forecasts (note that this criterion corresponds to the usual division in quintile considered to draw categoric forecasts). When the regular 203-point mesh is used, the ratio of grid points with percentiles smaller than the above threshold gives an estimation of the area with the desired skill; we refer to this verification measure as the 'running quintile verification'.

On the other hand, probabilistic seasonal forecasts can be displayed as maps with the probability of occurrence of an event (e.g. dry/normal/wet events if we use terciles for precipitation). These probabilities can be estimated as the frequency of each of the events within the ensemble forecast. In this case, we can use standard validation techniques derived from 'hit' and 'false alarm' rates for the validation of the forecasts. ROC curves and RSAs provide a global description of the skill of the method (see Katz and Murphy, 2002, for an overview of validation for probabilistic forecasts). Combining the above indices and the climatology of the event for the period considered, economical value curves associated with a simple cost/loss decision model

can be obtained. This validation technique is very convenient in the seasonal forecast framework, because it allows end-users to make a decision driven by economic profit (see Palmer et al., 2000, for a detailed description). The basic idea of this decision method is to quantify the value of taking, or not, precautionary action based on the information provided by a seasonal forecast. Taking action incurs a cost C on each occasion, whilst not taking action incurs a loss L when the event occurs. The economic value of a forecast is defined as the reduction in the user's expected mean expense per unit loss with respect to the action considered when the only information available is the climatology. For a forecast system that is no better than climate, the economic value is zero, whereas a perfect deterministic forecast system has a maximum economic value (one). An ensemble forecast gives hit and false alarm rates as a function of different probability thresholds (from zero to one), providing a set of economic value curves that define an envelope which is considered the economic value of the probabilistic forecast. This verification measure has been used in this paper to objectively compare the skill of the different ensemble methods and to compare them with the deterministic forecasts obtained as an average of the ensemble members.

4. Statistical downscaling using analogue methods

Different global and local statistical methods (regression, canonical correlation, neural networks, analogues, clustering, etc.) have been proposed in the literature (see Zorita and von Storch, 1999; Gutiérrez et al., 2004, for a survey). In this paper we
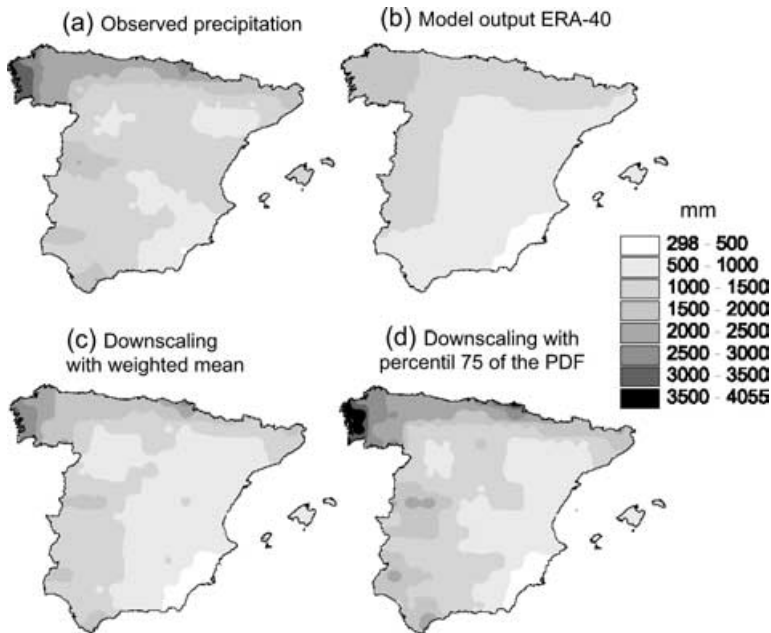
(a) Observed precipitation

(b) Model output ERA-40

mm

	298 - 500
	500 - 1000
	1000 - 1500
	1500 - 2000
	2000 - 2500
	2500 - 3000
	3000 - 3500
	3500 - 4055

(c) Downscaling
with weighted mean

(d) Downscaling with
percentil 75 of the PDF

Fig. 2. (a) Observed accumulated precipitation for 1987–1988; (b) direct ERA-40 analysis; (c) and (d) downscaled values using the weighted mean of the analogue ensemble and the percentile 75 of the PDF analogue ensemble, respectively.

use a two-step standard analogue technique (ANALO-ONE) which is based on the search for analogues of 1000- and 500-hPa geopotential height and 1000-, 925-, 850- and 700-hPa relative humidity fields of the ECMWF and UKMO models. In the first step, 100 analogues are obtained using information from the geopotential; afterwards, the analogue ensemble is lowered down to 30 members using information from the humidity fields. This method is currently used operationally for short-range precipitation forecast (Fernández et al., 2001) in Spain (http://www.inm.es/web/infmet/predi/preci.html). The empirical probability density function (PDF) given by the ensemble of analogues provides a probabilistic forecast for any event of interest. Moreover, this function is also the key for obtaining a numeric forecast, such as the weighted mean, or a given percentile. The performance of the method in perfect-prog mode using ERA40 analysis data is shown in Fig. 2, where two different numerical estimations of the expected precipitation value are taken. The estimation based on the weighted mean is shown to underestimate the observed values (see Fig. 2c). Thus, a skill oriented procedure was conducted to obtain the optimal percentile for each of the stations, based on the whole ERA40 results. The obtained percentile was close to 75 in most pluviometric stations, so we use this value in this work as a consensus estimation for the accumulated precipitation from the ensemble of analogues (see Fig. 2d).

When applied to an ensemble forecast system, the method of analogues can be used in probabilistic mode (considering the joint PDF obtained by combining the analogue sets for each of the ensemble members), or in numeric mode (considering the percentil 75 estimation of the set of analogues for each of the

ensemble members). In this last case, the result is an ensemble of numerical forecasts and can also be validated using RSAs and economic values.

4.1. Overall performance

The overall performance of the statistical downscaling method when applied to the DEMETER data in seasonal mode is shown in Fig. 3. Ensembles of 18 members for 180-d integrations combining the ECMWF and UKMO DEMETER models (ECMO18) during the 1986–1997 period are considered. This figure shows the RSA values obtained from the direct ensemble output versus the RSA values obtained from the corresponding downscaled forecasts (using the percentile 75 as estimator). Results for different lead times are reported on different panels, from a zero-month lead time (seasons NDJ, FMA, MJJ, ASO) to a three-month lead time (in this last case the seasons coincide with those of the zero-month lead time, so the influence of lead time in four different seasons can be compared). In all cases, results for wet, normal and dry seasons are shown separately (according to the observed climatological terciles of the grid points for the period of analysis). From these figures it can be shown that, as the skill of the direct output increases, the statistical downscaling method outperforms the direct output of the models (note that, in those cases where the model skill is poor, the downscaling method is not expected to improve the forecast, because it introduces an additional source of uncertainty). These figures show that, when considering the 1986–1997 period and the whole area of Spain, the highest skills (above 0.6) are associated with early and late spring, summer and autumn seasons at zero- and one-month
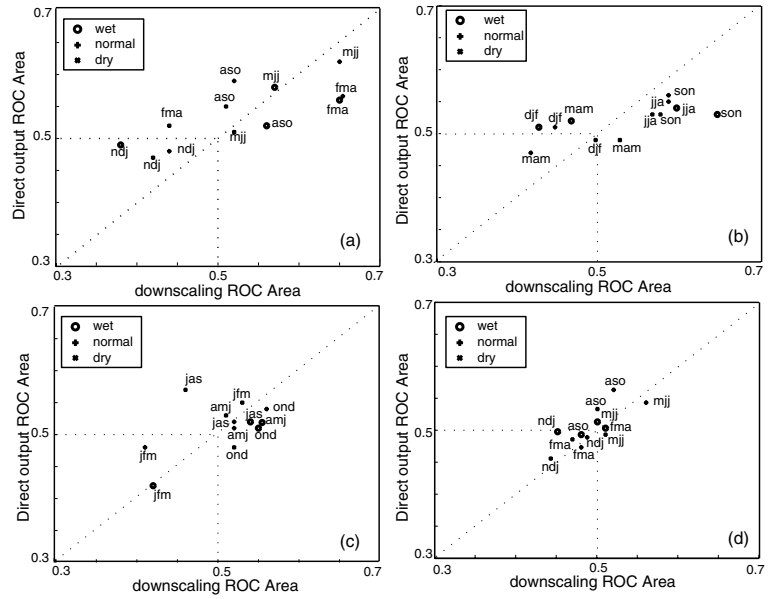
Fig. 3. RSAs of the probabilistic forecasts at the 203 grid points of Fig. 1c stations for each of the 21 seasonal forecasts started in November, February, May and August and all the years overall. (a)–(d) correspond to the lead times 0–3.
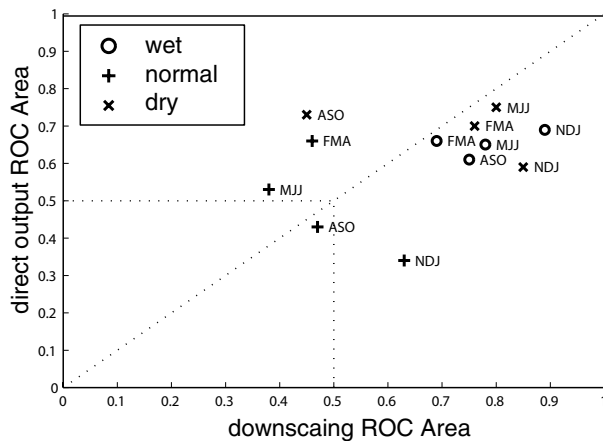

Fig. 4. ROC areas for the four zero-month lead time seasons (NDJ, FMA, MJJ, ASO) for the south of Spain and the period November 1986 to October 1988 (including an active El Ninño period).

lead times. On the other hand, the winter season exhibits a poor performance. Overall, no season and no lead time exceed the RSA 0.65 for this period of 11 yr, as could be expected in this mid-latitude region.

However, the skill of the method can be higher in particular periods, at least in some regions of Spain. For instance, Fig. 4 shows the comparison of the direct output and the downscaled values for a particular El Niño period (1986–1988) in the south of Spain. This figure shows a clear division between normal events (with poor performance) and wet and dry periods, with high RSA values (in the range 0.7–0.85) for all seasons. This is not surprising, because dry precipitation anomalies have been found in Spain in connection with El Niño events (see, for example, Rodo et al., 1997). The high skill found in this period

for the winter season is also remarkable, as opposed to the poor performance found in the previous general case. Moreover, the downscaling method outperforms the direct ensemble output in all cases with significative skill. From the above results, it would be worth making a detailed validation of the skill of direct and downscaled outputs considering different areas in Spain and different El Niño periods; however, this is beyond the scope of this paper.

4.2. Temporal and spatial performance

The above results show that the skill of the models highly fluctuates temporal and spatially. Figure 5 shows the results obtained applying the running quintile criterion to determine the percentage of the peninsular area where the forecasts are accurate, season by season for lead times 0–3. We consider the deterministic forecasts starting in May and November and extended along the 11-yr period 1986–1997 for (a) the analogue method applied to the ECMO18 outputs (ECMO18AN) and (b) the ECMO18 realizations. From this figure, the following results are obtained. For the seasons JJA90, MJJ91, JJA91 and JJA94, the running quintile criterion is simultaneously satisfied for ECMO18 and ECMO18AN for a percentage of peninsular area higher than 60%; for JAS88, MJJ95 and ASO95, the area percentage is 50%. We found that the character of all these seasons was dry for at least 50% of the territory. The ECMO18AN forecasts starting in November are good only for El Niño years 1986 and 1987 and are especially poor in autumn and winter of the very dry years 1991–1994. In contrast, the ECMO18 forecasts are good for these dry years.

In order to investigate the relationship between the wet/dry character of the season and the skill of the above methods, Fig. 6
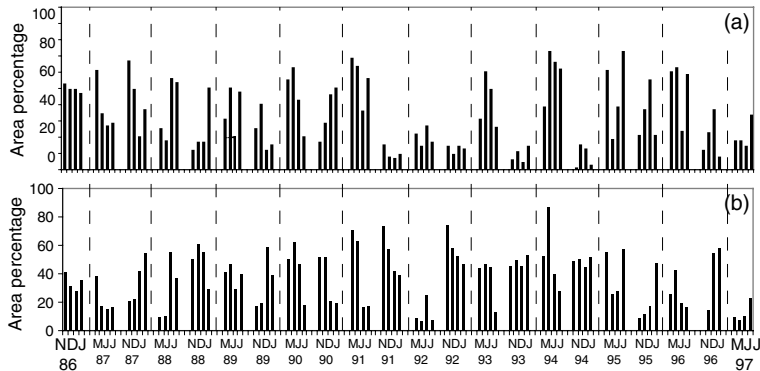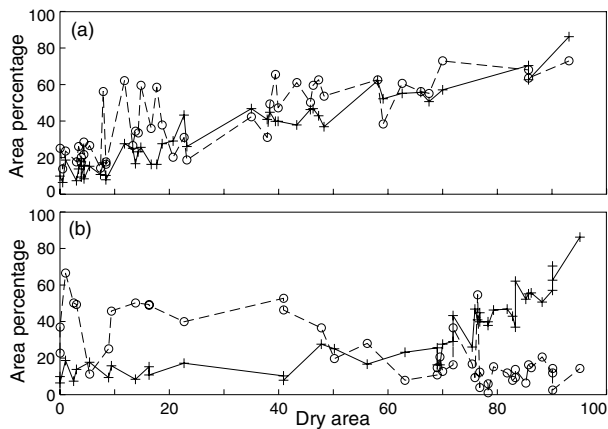
Fig. 6. Percentages of the peninsular areas satisfying the quintile criteria both for ECMO18 (solid line) and ECMO18AN (dashed line) versus the extension of the dry area for the realizations from (a) May and (b) November.

shows the percentages of the peninsular areas satisfying the running quintile criteria, both for ECMO18 and ECMO18AN, versus the extension of the dry area for the realizations from (a) May and (b) November. This figure shows that when the running quintile is applied to the forecasts starting in May, the ECMO18AN forecasts outperform those of ECMO18 in 36 out of 44 cases, not being related to the eight remaining cases with the character of the season. Moreover, a percentage of more than 40% of the peninsular area satisfying the running quintile criteria is attained in approximately half of the cases. This allows us to conclude that both methods produce good forecasts for the May case in relatively large areas of the territory, but with a clear prevalence of the analogue method. However, when the same criterion is applied to the forecasts starting in November, then there is a clear dependence of the results with the character of the season. When the dry character of the season is over 40% of the territory, then the ECMO18 forecasts perform better than those of ECMO18AN in 27 out of 30 cases. In 13 out of the remaining 14 forecasts, ECMO18AN performs better than ECMO18. So, the skill of November forecasts is related to both the character of the season and the method used. For dry seasons, the direct

model prevails, and for the normal and wet seasons the analogue method performs better. This can be explained by the fact that the direct precipitation of the global model has a negative bias as shown, for instance, in Fig. 12, which is corrected by the analogue method. The very dry character of the autumn and winter seasons of the years 1991–1994 in Spain also contributes to these biased results.

Further results about the skill of the analogue method are presented later when describing some case studies.

4.3. Bias-correction

We have also performed some preliminary work to correct the bias of DEMETER models, thus reducing this source of error from the forecast. The bias was derived from the available forecasts and observed seasonal precipitation values from 1961–1997. For the 11 yr, new RSA outputs using the bias-corrected seasonal values were computed. However, no clear conclusions about the improvements attained with the bias reduction of the combined ECMWF and UKMO DEMETER models could be established, and further research is needed to address this point. The bias of the ECMO18AN results could not be corrected because the statistical downscaled values for 1961–1986 were not available.

5. Dynamical downscaling

In this section we explore the ability of dynamical downscaling models to provide high regionalized detail to the seasonal forecast (Chen et al., 1999). Integrations with the dynamical RCA limited area model with 31 levels in the vertical and a horizontal resolution of 0.5° (see Fig. 1e) were undertaken covering a wide European Atlantic domain (15.5°N–65.0°N and 67.5°W–31.0°E); the RCA model is a climate version of the HIRLAM regional weather prediction model (Rummukainen et al., 2001). We also considered a more restricted south-west Europe domain surrounding Iberia (30.2°N–50°N and 23.6°W–15°E) with a horizontal resolution of 0.2° (RCH; see Fig. 1d). In the first phase, a test was conducted under perfect model conditions using ERA40 data from March–June 1998. With this aim, the above versions of the RCA model and two versions of the HIRLAM
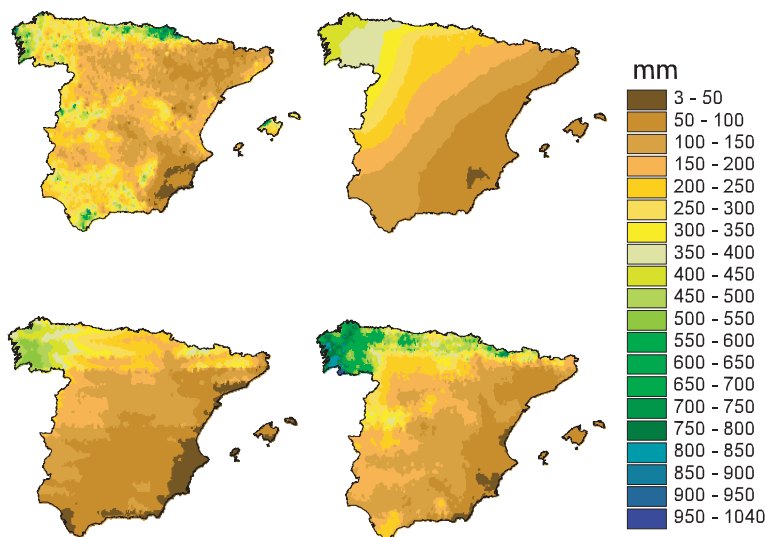
Fig. 7. Observed precipitation for DJF 1986/1987 from the Spanish pluviometric network, three-member ensemble mean predictions from the ECMWF global model and the RCA and RCH limited area models interpolated to the pluviometric network.
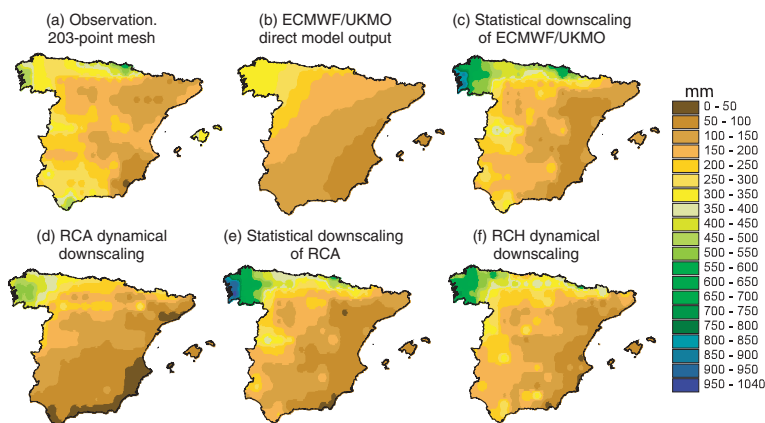

Fig. 8. Accumulated precipitation for DJF 1986/1987. (a) Gridded observations, (b) average of joint ECMWF/UKMO 18-member ensemble prediction, (c) analogue downscaling of ECMWF/UKMO prediction, (d) RCA limited area model prediction (ensemble mean of three-member ensemble), (e) analogue downscaling of RCA prediction, and (f) RCH limited area model prediction (ensemble mean of three-member ensemble).

model (HIRLAM short-range model run in operative mode, and HIRLAM model run in climate mode) were compared. The test showed that RCA and RCH models were appropriate for six-month integrations.

Afterwards, the RCA and RCH models were fed with global model boundaries from the ECMWF DEMETER and RCA models, respectively. Due to the computational cost of the integrations, only some particular periods and combinations of ensembles were performed in this work for the sake of comparison with statistical techniques (in particular, integrations starting in November, from 1986–1989, and in May, from 1987–1989, for three out of nine ensemble members). For instance, Fig. 7 shows the precipitation obtained from the ECMWF global model (with the three ensemble members, which provides the boundaries to the RCA model), and from the RCA and RCH limited area models (with three ensemble members), all interpolated to the locations of the pluviometric network for the season DJF 1986/1987 given by the 1986 November integration. This figure shows that the regional models substantially increase the detail

of the forecast, gaining accuracy in some regions and losing in others.

In the next section we present several comparative results of the above statistical and dynamical downscaling techniques for some seasons within the period 1986–1989, with available dynamical downscaled values.

6. Comparison of dynamical and statistical techniques

6.1. December–February 1986/1987 case

Figure 4 shows that the skill of seasonal forecasts for winter 1986–1988 is significantly high, compared with the overall performance for this season (see Fig. 3). For this reason, we start by comparing the performance of statistical and dynamical methods using the DJF season from the 1986 November integration. Figure 8 shows the maps of the seasonally accumulated observed and predicted precipitation. Ensemble averages for five different

models are compared in this graph: (b) the joint ECMWF/UKMO DEMETER (ECMO18) direct model outputs (18-member ensemble); (d), (f) the RCA3 and RCH3 direct model outputs (both with three-member ensembles); (c), (e) the analogue downscaled precipitation from the joint ECMO18 and from the RCA3 models. The plots are based on values interpolated or computed at the 203-point mesh. They can be compared with those of Fig. 7 interpolated to the pluviometric network. Thus, maps of Figs. 7a, c and d correspond to Figs. 8a, d and f, respectively.

In this case study, the above figures show that the high-resolution RCH3 model outperforms the direct model outputs from the ECMO18 and RCA3 models, which underestimate the precipitation amount. Moreover, the analogue technique (ECMO18AN and RCA3AN) allows us to correct this bias, leading to results comparable and even better than with the RCH3 model. The above statements can be confirmed by measuring, for each of the models, the percentage of the territory where the relative error between the observed and forecasted accumulated seasonal precipitation is less than ±20% (these percentages were estimated from the 203-point mesh using a Geographic Information System): ECMO18AN (50%), RCH3 (40%), ECMO18 (35%), RCA3AN (32%) and RCA3 (21%). For the other 1986/1987 seasons NDJ, JFM and FMA, the RCH3 model performed better than the RCA3 model, but slightly worse than RCA3AN.

More details are given in Fig. 9, which shows a bar diagram of area percentages for the DJF 1986/1987 season period for the above models. In each of the five panels, the first three bars from left to right correspond to the percentage of grid points where the character of the observed precipitation has been dry, normal and wet (given by the terciles of the 1961–1990 series). The horizontal black lines inside the bars indicate the percentage of those grid points where the forecast is less than 20 percentiles away from the observation (i.e. like a hit rate conditioned to the wet/normal/dry observed character). On the other hand, the first three bars from right to left correspond to the percentage of grid points where the forecast precipitation has been dry, normal and wet, respectively; in this case, the horizontal black lines
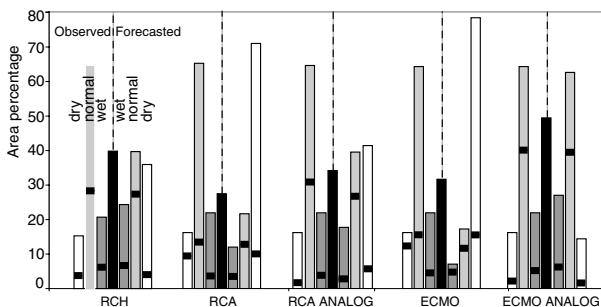
indicate the percentage of grid points where the observation was less than 20 percentiles away from the prediction. A summary measure of the overall quality of the forecast is given by the height of the central bar (black), which shows the percentage of points where the observation and prediction differ less than 20 percentiles (note that this value is the sum of the horizontal black lines either on the left or on the right of the central bar). Therefore, the bars on the left and on the right of the figure allow us to compare observed and forecast dry/normal/wet areas.

Figure 9 shows that both RCA3 and ECMO18 model outputs are biased towards dry events, whereas RCH3 and the analogue methods balance the forecast according to the observed precipitation. In this case, the best forecast is achieved with the analogue version of the combined ECMO18 model, which agrees with the observations in 50% of the Spanish area, and performs better than the RCA3 analogue. However, the second best is RCH3, as already indicated before, and also shown in Fig. 10, which displays the overall measure given by the overall 'hit' areas (the central bars) for each model and each of the 1986/1987 seasons. From this figure it can be seen that RCH3 performs better than RCA3 in all seasons, but it performs better than RCA3AN only in the season DJF 1986/1987. This result, supported also by the fact that RCA3 performs better than ECMO18 in two of the four seasons, indicates that an increase of the resolution of the model benefits the overall skill of the dynamical forecasts. This skill gain can alternatively be achieved by using analogue statistical techniques. In Section 6.4, it can be seen that these conclusions are maintained when a larger sample of seasonal forecasts is considered. It has to be noted however that ECMO18AN is a two-model 18-member ensemble, which does not make completely fair the comparison with the other three-member ensembles used.

6.2. November–January 1986/1987 and 1987/1988 case during an El Niño event

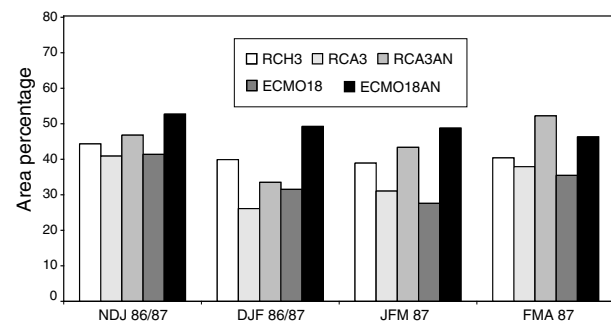Figure 11 shows a bar diagram for the joint period NDJ 1986/1987 and NDJ 1987/1988 and for the outputs

Fig. 9. Bar diagram of area percentages for the period DJF 1986/1987 for the methods ECMO (18 ensemble members) RCA and RCH (three ensemble members) together with the analogue versions of ECMO and RCA.

Fig. 10. Bar diagram of area percentages for the seasons NDJ 1986/1987 (zero-month lead time) to FMA 1987 (three-month lead time) for the integrations started in November 1986 for three different models (and their analogue counterparts).
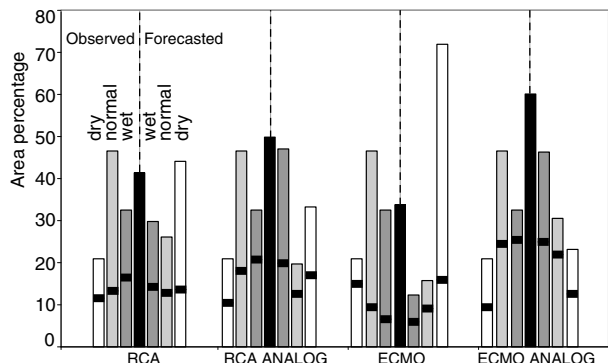
Fig. 11. Bar diagram of area percentages for the period NDJ 1986/1987 and 1987/1988 for the methods ECMO (18 ensemble members) and RCA (three ensemble members) together with the analogue versions.
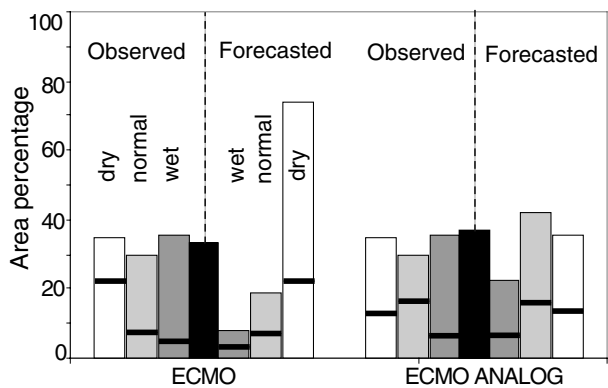

Fig. 12. Bar diagram of area percentages for NDJ 1986–1997 for the methods ECMO18 (18 ensemble members) and the analogue versions.

Fig. 13. Regions of Spain (interpolated from the 203 mesh) where the difference between the observation and forecast differ less than 20 percentiles in two (black), one (dark grey) and zero (light grey) cases out of the two NDJ seasons for the ECMO18AN in the period 1986–1987.
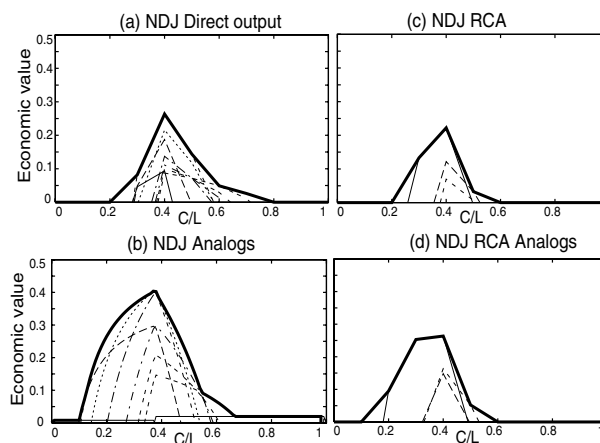

Fig. 14. Economic values and their envelopes for the 'wet' event during NDJ of 1986/1987 and 1987/1988 for ECMO18 (18 ensemble members) and RCA (three ensemble members), and the analogue versions. Economical values are shown as a function of *C/L*. The thin lines show the individual graphs for probability thresholds. The thick lines show the maximum economical value (see explanation of the curve in the last paragraph of Section 3).

corresponding to ECMO18, ECMO18AN, RCA3 and RCA3AN. This bar diagram shows that also for this period the direct model outputs are biased towards dry events, whereas the analogue method balances the forecast according to the observed climatology. In this case, the best forecast is achieved also with ECMO18AN, which agrees with the observations in 59% of the Spanish area. We also analysed both years separately obtaining a 53% area for 1986/1987 and 77% area for 1987/1988; this last year turned out to be the most skilful and was also wet in 62% of the Spanish area.

In order to check the significance of the above forecasts, Fig. 12 displays the same diagram for the period 1986–1997 for the ECMO model. In this case, the performance of the methods decreases and both the direct output and the analogue downscaling exhibit similar performance; however, the balance of the dry/normal/wet events is maintained between the climatology and the forecast for the analogue method.

Figure 13 shows the regions of Spain where the forecast is more accurate. For the two seasons NDJ 1986/1987 and 1987/1988, the percentile criterion was applied to the 203 mesh

points; differences between predictions and observations smaller than 20 percentiles were considered right. The map shows the regions with two right forecasts (black), one (dark grey) and zero (light grey). This figure shows that the southern region of Spain is more skilful during this El Niño period.

In order to check the performance of the different downscaling techniques when considering a probabilistic framework, Fig. 14 displays the economical values of the probabilistic forecasts obtained with the different downscaling models for the period (note that now the ensemble PDFs are used as forecasts, instead of means or percentiles obtained from this distribution; see Section 3 for more details). It shows that the economical value corresponding to the two ensemble forecasts from ECMO18 with 18
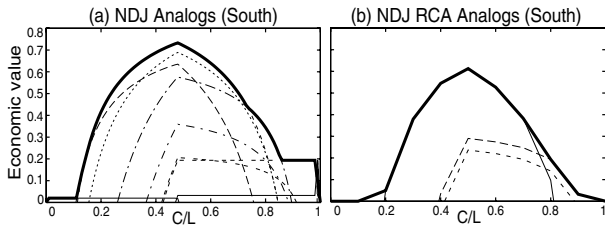
Fig. 15. Economic values and their envelopes for the 'wet' event during NDJ of 1986/1987 and 1987/1988 applying the analogue method to the ECMO (18 ensemble members) and the RCA (three ensemble members) models for the south of Spain.

members presents a higher peak and a wider range of *C/L* than that corresponding to the two ensembles from RCA3 with only three members. The ECMO18AN forecast is the best ranked for deterministic and probabilistic forecasts, but the ECMO18 model is no longer fourth, as it was in the deterministic forecast (see Fig. 11) when the ensemble forecast was considered. In Fig. 15 we also present the economic value for southern Spain. In this case, the economic value of the ensemble seasonal forecasts is remarkably high.

6.3. *February–April 1988 case*

As we have seen in Figs. 3 and 4, the models have a high skill in the period February–April, especially during one El Niño episode. The season FMA88 corresponds to the late winter/early

spring of this El Niño episode, and its character is dry for almost 50% of the territory, in contrast with the DJF 1986/1987 season considered above. Figure 16 shows the observation and the deterministic forecasts given by different models: the 18-member direct model precipitation from the ECMWF and UKMO models (ECMO18) and the corresponding statistically downscaled values (ECMO18AN), the dynamical downscaling realizations from the RCA3 model and their statistically downscaled outputs (RCA3AN) described above. From this figure we can see that the RCA3 model outperforms the ECMO18 model, gaining resolution and improving the amount of precipitation in the north-west. On the other hand, the analogue techniques give very similar patterns for both numerical models in both cases.

From the bar diagram in Fig. 17, it can be seen that the character of the FMA88 season in the peninsula was dry in about 48% of the territory, normal in about 30% and wet in the other 22%. From the map of observed anomalies shown in Fig. 18b it can be seen that a positive anomaly covers most of the high Ebro basin and the north basin, except Galicia, and a negative anomaly covers most of the north-west basin, the south-west part of the peninsula Iberia and part of Catalonia. These anomalies are also reflected in the map of terciles of the observed precipitation (Fig. 18a).

The deterministic forecasts reproduce this pattern only partially. As can be seen from the bar diagram in Fig. 17, each of the four cases gives good forecasts for areas between 37% and 55% of the territory (when a good forecast is defined by the running quintile criterion). ECMO18 and RCA3 perform better
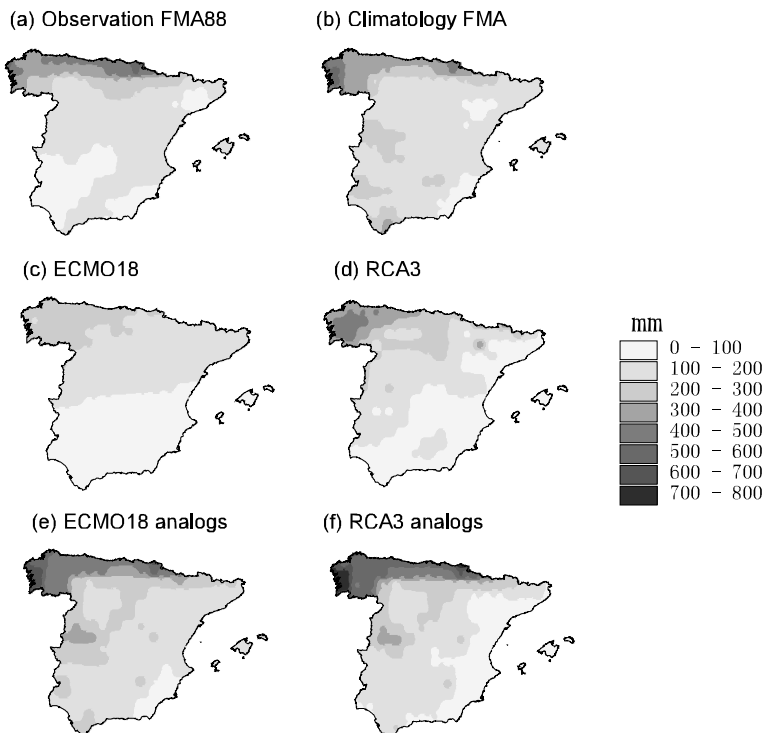

Fig. 16. Maps of (a) amounts of precipitation for the season FMA88; (b) mean map for the period 1961–1990. (c) ECMO18, (d) RCA3 (e) ECMO18 analogues and (f) RCA3 analogues.
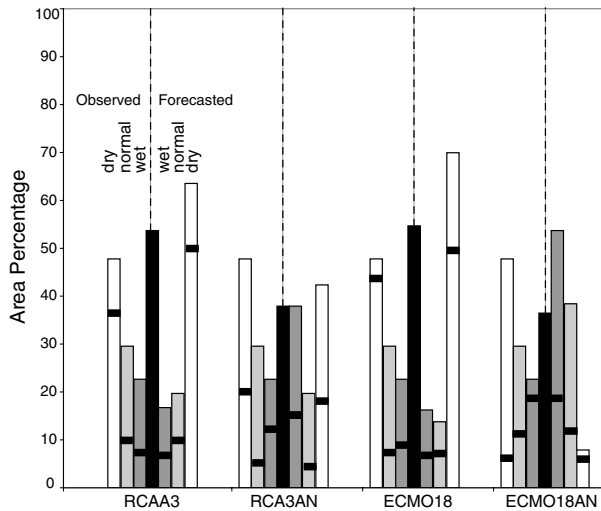
Fig. 17. Bar diagram of area percentages for FMA88 for the methods ECMO (18 ensemble members) and RCA (three ensemble members) together with the corresponding analogue versions.

predicting the dry areas (good forecasts over close to 40% of the territory) where ECMO18AN fails near completely. However, ECMO18AN succeeds in predicting the wet areas and RCA3AN makes a more even distribution of good forecasts for the three events, but reaching overall only 40% of the territory.

The geographical distribution is remarkably different after the forecasts, as shown in Fig. 18. There, observed terciles intersected with those regions where observed and forecast percentiles differ less than 20 show that ECMO18 performs relatively well for most of the country out of the north and north-west, the 'wet Spain', where the wet conditions are better reproduced by ECMO18AN and RCA3AN.

In summary, for the FMA88 season, the deterministic forecast from the direct global models is quite unable to reproduce the higher values of precipitation in the north of Spain. The analogue method does it for both models, but with poor results for the dry areas. The deterministic forecast from RCA3 ensemble members thus gives the best results overall.

This is confirmed by the corresponding economical value curves for the deterministic forecast (bold dashed lines of
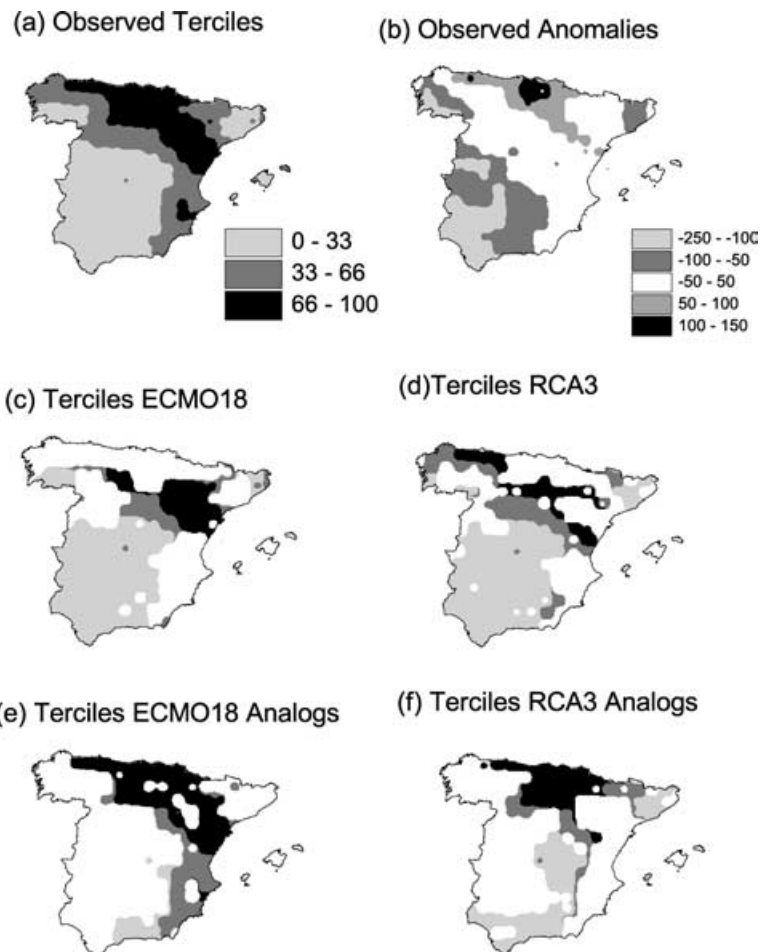

Fig. 18. (a) Map of observed terciles for precipitation with respect to the 1961–1990 precipitation; (b) map of observed anomalies for FMA88; and the intersection of the observed terciles maps and the maps of the areas where the observed and forecast percentiles differ less than 20 percentiles (c) ECMO18, (d) RCA3, (e) ECMO18AN and (f) RCA3AN for FMA88.
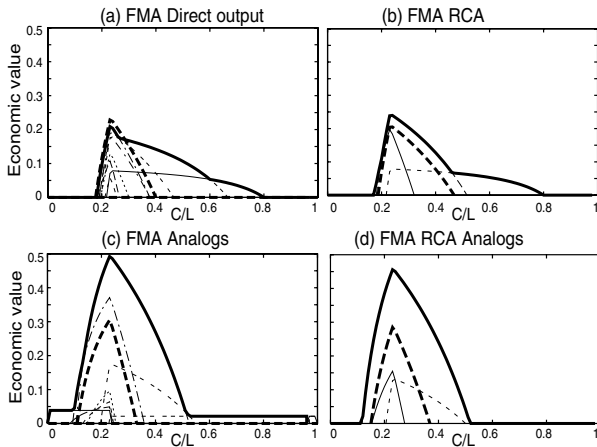
Fig. 19. Economic values for the 'wet' event during FMA of 1988 for ECMO (18 ensemble members) and RCA (three ensemble members), and the analogue versions. The envelopes of the economic values for the probabilistic forecasts (solid bold line) and the economic value for the deterministic forecasts obtained with the ensemble mean (dashed bold line) are displayed for each case (see also the caption for Fig. 14).
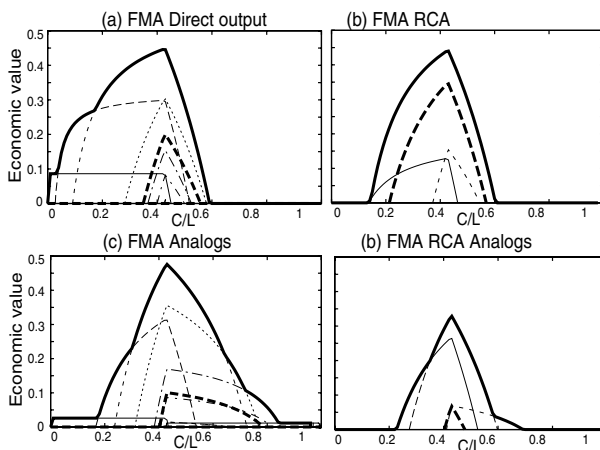

Fig. 20. As in Fig. 19, but for the 'dry' event 1988.

Figs. 19 and 20). The RCA3 curve in Fig. 20, corresponding to the dry event, is the one that presents a remarkable higher peak and area. In the case of the wet event, shown in Fig. 19, the curves corresponding to the analogue methods present the highest peaks, in agreement with what was said previously for the deterministic forecasts. However, when the probabilistic forecast and its economical value curves are considered (the solid bold lines for Figs. 19 and 20) it comes out that the economical values of the probabilistic forecast are higher than the corresponding deterministic forecasts, and that the forecasts performing better are those from ECMO18 and ECMO18AN, both for the wet and dry events. Moreover, the higher number of members of the ensemble (18) is probably the reason why the ECMO18 model

outperforms RCA3 for the dry event, in contrast with what happened for the corresponding deterministic forecast.

6.4. The period 1986–1989

In the previous sections, seasonal forecasts for special periods have been analysed in detail using different validation measures. In this section, in order to compare the overall performance of the methods, we have chosen first the running quintile criterion to rank each of the four models for each of the available deterministic forecasts. The skill computed by the running quintile criterion is the percentage of area in Spain where the observation and prediction differ less than 20 percentiles (they both are in the same running quintile). There is a total of 28 available forecasts, coming from the seven 180-d integrations started in November and May from November 1986 to November 1989 (for each of these integrations, four seasonal forecasts are available, with different lead times). This gives a total of 16 forecasts for the November 1986/1989 integrations and 12 forecasts for the May 1987/1989 integrations. Results are summarized in Table 1. The first figure in each couple of Table 1 indicates the number of times each model ranks in the first and second positions, whereas the second figure indicates the number of times the model ranks in positions third and fourth. This table shows that the analogue methods overall improve the forecast for the period 1986–1989. The best performance for November corresponds to model RCA3AN, whereas for May it corresponds to ECMO18AN. This result is more pronounced during the El Niño episode. For this episode, covering the cases of columns two and five of Table 1 (eight forecasts corresponding to November 1986 and 1987 integrations and four forecasts from May 1987), RCA3AN ranks in the first or second position in 11 out of the 12 cases and ECMO18AN in 10 out of the 12 cases.

When we make a similar comparison considering the ranks for the 28 forecasts only for the dynamical forecasts ECMO18 and RCA3, it turns out that in 11 of the 16 November forecasts and in eight of the 12 May forecasts, RCA3 outperforms ECMO18. This confirms that the seasonal deterministic forecasts benefit from an increase of resolution using a large number of cases (Section 6.1).

Finally, we compare the above reported results for deterministic downscaling with those obtained considering the probabilistic forecast; we also analyse the relevance of the number of members of the ensembles. For this purpose, we compare the direct model outputs of ECMWF9 and UKMO9 (each of the nine members of the ECMO18 ensemble from the ECMO and UKMO models), ECMW3 (the three members of the ECMWF ensemble used as boundary conditions for the RCA model) and RCA3. RSAs for the dry, normal and wet events at the main climatological stations (Fig. 1a) were computed, and are displayed in Fig. 21. It can be seen that RCA3 presents better scores than ECMWF3 for dry and wet events in 14 out of 16 cases; for the normal events, ECMWF3 scores are better in five of the eight cases. However,

Table 1. Summary of model performance for the period 1986–1989 according to their skill rank (see details in the text)

| | November | | | May | | |
	1986–89	1986/1987	1988/1989	1987–89	1987	1988/1989
RCA3	8–8	2–6	6–2	4–8	0–4	4–4
RCA3AN	11–5	7–1	5–3	7–5	4–0	3–5
ECMO18	5–11	1–7	3–5	4–8	0–4	4–4
ECMO18AN	8–8	6–2	2–6	9–3	4–0	5–3

Fig. 21. RSA values for ECMWF9, UKMO9, ECMWF3 and RCA3 probabilistic seasonal forecasts. Left panels correspond to dry, normal and wet events and seasons with lead times zero to three from the three May integrations. Right panels show the same information for the four November integrations.
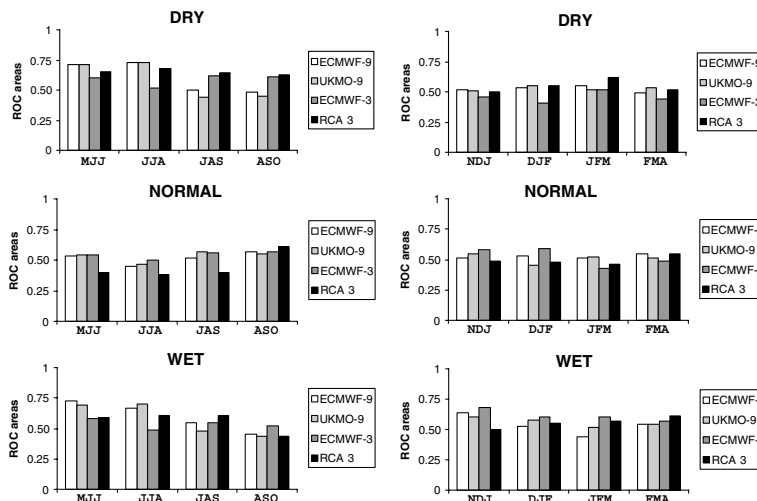
when ECMWF9 and RCA3 are compared, ECMWF9 outperforms RCA3 for normal and wet events in 11 of 16 cases and RCA3 outperforms ECMWF9 for the dry event in five of eight cases. Therefore, for the probabilistic forecasts, the number of ensemble members seems to balance the effects of using a model with higher resolution.

The most comprehensive way to compare deterministic and probabilistic seasonal forecasts with a different number of ensembles is to compute and display the economical value curves in a similar way as done in Figs. 19 and 20 for the FMA88 case. So, the 28 seasonal forecasts for the full 1986–1989 period were split into two groups corresponding to the November and May integrations, and the curves for the different events and methods obtained. From Fig. 22 it turns out that for the November integrations and the events 'dry' and 'wet', probabilistic forecasts outperform the deterministic ones. Moreover, RCA3AN followed by RCA3 outperform the 18-member cases giving an example where successive application of dynamical and statistical downscaling improve each separate method. For the May integrations, the economical value of probabilistic forecasts also outperforms the deterministic forecasts, but the direct model methods prevail over the statistical methods; note that, in this case, the running quintile and economical value methods are not in good agreement.

7. Conclusions

The main goal of this paper was to compare the skill of seasonal DEMETER precipitation forecasts over Spain obtained both from the direct model outputs and by applying statistical and dynamical downscaling techniques. Although several comparative analyses between statistical and dynamical downscaling techniques can be found in the earlier literature, this study has the originality that is applied to seasonal ensemble forecasts.

One of the conclusions reached is that the skill of direct DEMETER forecasts greatly fluctuates from season to season. Overall, the highest skills over Spain are associated with early and late spring, summer and autumn seasons at zero- and one-month lead times. On the other hand, models have a poor skill during winter with the exception of the El Niño period (1986–1988), especially in the south of Spain. In this case, economical values close to 0.8 have been obtained.

In those cases with larger skill, statistical downscaling methods allow us to further improve the results, increasing the skill of the resulting downscaled forecast. Thus, statistical downscaling techniques allow us to improve those predictions with some initial skill (in other cases, statistics do not help, because the predicted anomaly is incorrect). An opposite situation is obtained for the deterministic autumn forecasts, and also with those cases
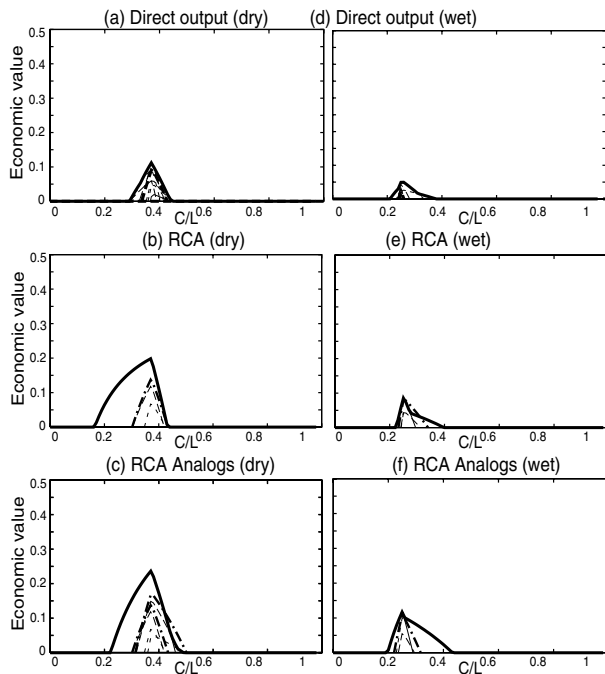
Fig. 22. Economic values for the wet and dry events corresponding to all lead time seasonal forecasts from the November integrations of years 1986–1989, for ECMO18, RCA3 and RCA3AN. The envelopes of the economic values for the probabilistic forecasts (solid bold line) and the economic value for the deterministic forecasts obtained with the ensemble mean (dashed bold line) are displayed for each case (see also caption for Fig. 14).

when the dry character of the overall precipitation in Spain is higher than 40% of the territory. The trend of the global models to underpredict the precipitation and the extreme dry character of several autumn and winter seasons of the period makes the direct model forecasts outperform the statistical downscaling method in those cases.

We have also tested the performance of dynamical downscaling methods. In this case, the skill of the direct DEMETER outputs were compared with the skill of high-resolution models nested to DEMETER models. It turned out that in the four seasons simulated (from the integration of November 1986) the highest-resolution RCH3 model was superior to the lower-resolution direct model outputs ECMO18 and RCA3, which underestimate the precipitation amount. This results shows the feasibility of dynamical downscaling for seasonal forecasts, at least in this case of study. However, no conclusion was obtained when comparing statistical and dynamical downscaling methods, because no method was significantly superior than the others.

Therefore, from the above results, it would be worth making a detailed validation of the skill of direct and downscaled outputs considering different areas in Spain and different El Niño periods. However, this is beyond the scope of this paper.

Another important conclusion is that dynamical and statistical downscaling methods can be used in combination, not as alternatives, obtaining the best skill scores in some of the analysed cases.

Finally, we want to remark that the results of the different dynamical experiments performed in this paper have been stored at the ECMWF MARS system and are available for the scientific community. Those of the statistical experiments are available through the authors. As a result, there is a data set of direct model and downscaled values of daily precipitation over the Spanish networks and mesh referred in the paper that could be useful for testing applications. The inventory of data and downloading instructions are available at http://www.ecmwf.int/research/demeter/.

8. Acknowledgments

References

Chen, S. C., Roads, J. O., Juang, H. M. H. and Kanamitsu, M. 1999. Global to regional simulations of California wintertime precipitation. *J. Geophys. Res. Atmos.* **104**(D4), 31 517–31 532.

Christensen J. H., Raisanen, J., Iversen, T., Bjorge, D., Christensen, O. B. and Rummukainen, M. 2001. A synthesis of regional climate change simulations. A Scandinavian perspective. *Geophys. Res. Lett.* **28**, 1003–1006.

Doblas-Reyes, F. J., Hagedorn, R. and Palmer, T. N. 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting – II. Calibration and combination. *Tellus* **57A**, 234–252.

Fernández, A., Del Hoyo, J., Mestre, A. and Peral, C. 2001. Local Probabilistic Forecasts of Precipitation by the Use of an Analogical Approach. *Fifth European Conference on Applications of Meteorology. ECAM 2001.* http://www.inm.es/ecam/ecam_2001_expanded_abstract.pdf

Giorgi, F., Hewitson, B., Christensen, H. J., Hulme, M., von Storch, H. and co-authors, 2001. Regional climate information. In: *Evaluation and Projections in Climate Change 2001; The Scientific Basis, Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change (IPCC)* (eds J. T. Houghton, Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, and D. Xiaoxu). Chapter 10, Cambridge University Press, Cambridge, 583–638.

Goddard, L., Mason, S. J., Zebiak, S. E., Ropelewski, C. F., Basher, R. and Cane, M. A. 2001. Current approaches to seasonal to interannual climate predictions. *Int. J. Climatology* **21**(9), 1111–1152.

Gordon, C., Cooper, C., Senier, C. A., Banks, H., Gregory, J. M. and co-authors, 2000. The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dyn.* **16**, 147–168.

Gutiérrez, J. M., Cano, R., Cofiño, A. S. and Rodríguez, M. A. 2004. Clustering methods for statistical downscaling in short-range weather forecast. *Mon. Wea. Rev.* **132**(9), 2169–2183.

Hagedorn, R., Doblas-Reyes, F. J. and Palmer, T. N. 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting – I. Basic concept. *Tellus* **57A**, 219–233.

Katz, R. W. and Murphy, A. H. editors. 2002. *Economic Value of Weather and Climate Forecasts*. Cambridge Univ. Press, Cambridge.

Misra, V., Dirmeyer, P. A. and Kirtman, B. P. 2003. Dynamic downscaling of seasonal simulations over South America. *J. Climate* **16**, 103–117.

Palmer, T. N., Brankovic, C. and Richardson, D. S. 2000. A probability and decision-model analysis of PROVOST seasonal multi-model ensemble integrations. *Q. J. R. Meteorol. Soc.* **126**, 2013–2034.

Palmer, T. N., Alessandri, A., Andersen, U., Cantelaube, P., Davey, M. and co-authors 2004. Development of a European multi-model ensemble system for seasonal to interannual prediction (DEMETER). *Bull. Am. Meteorol. Soc.* **85**, 853–872.

Rodo, X., Baert, E. and Comin, F. A. 1997. Variations in seasonal rainfall in southern Europe during the present century: relationships with the North Atlantic Oscillation and the El Niño Southern Oscillation. *Climate Dyn.* **13**, 275–284.

Rummukainen, M., Räisänen, J., Bringfelt, B., Ullerstig, A., Omstedt, A. and co-authors. 2001. A regional climate model for northern Europe: model description and results from the downscaling of two GCM control simulations. *Climate Dyn.* **17**, 339–359.

Wilby, R. L. and Wigley, T. M. L. 1997. Downscaling general circulation model output. a review of methods and limitations. *Prog. Phys. Geog.* **21**, 530–548.

Zorita, E. and von Storch, H. 1999. The analog method as a simple statistical downscaling technique: comparison with more complicated methods. *J. Climate* **12**, 2474–2489.