

Snow Trends in Northern Spain. Analysis and Simulation with Statistical Downscaling Methods

M.R. Pons^(a), D. San-Martín^(b), S. Herrera^(c) and J.M. Gutiérrez^(c)

^(a)Agencia Estatal de Meteorología (AEMET), Santander, Spain.

^(b)Dept. of Applied Mathematics. University of Cantabria. Santander, Spain.

^(c)Instituto de Física de Cantabria, CSIC-University of Cantabria, Santander, Spain.

Abstract

In this study we analyze and simulate (with statistical downscaling techniques) the snow trends observed in the Northern Iberian Peninsula using daily snow occurrence data from a network of 33 stations ranging from 60 to 1350 metres. We first analyze the annual snow frequency measured as the annual Number of Snow Days (NSD), obtaining a significant decreasing trend since the mid seventies with a NSD reduction of about 50%; moreover, this magnitude is similar for low and high stations and for winter and spring separately. Then, we analyze the existing correlations with mean temperature and precipitation occurrence obtaining different relationships depending on the season and elevation. Finally, we simulate the observed trends using the connection of Daily Snow Occurrence (DSO) with large-scale fields simulated by a General Circulation Model (GCM); for this purpose we apply an analog-based statistical downscaling method to obtain an estimation of DSO, working in perfect prognosis conditions using reanalysis data. On the one hand, the downscaling method is able to estimate/predict the DSO with typical values of Hit and False Alarm Rates around 60% and 2%, respectively. On the other hand, the annual frequency obtained by averaging the DSO estimations reproduces very well both the observed trends and the high interannual variability. These promising results open the possibility to future research in seasonal or climate change projections of snow frequency.

1. Introduction

The analysis of trends in the observed climate has become an important research topic during the last two decades, mainly due to the interest in global warming; moreover, different sources of information have been used for this purpose: instrumental observations, proxy data, satellite information, etc. As a result, many global and regional trend studies are nowadays available, but mostly for temperature and precipitation (see, e.g. Trenberth et al. 2007). Other variables of interest such as snow have received little attention, basically due to the lack of appropriate data. However, snow frequency and extent are key parameters for the water budget and, thus, present and future trends of these variables are an important source of information for many socio-economic activities.

At a global scale, visible satellite imagery produced by the National Oceanic and Atmospheric Administration (NOAA) since the early 1970s (Dewey and Heim 1982; Matson et al. 1986; Robinson 1993) has been the main data source for studying the variability of snow extent (Foster et al. 1983; Gutzler and Rosen 1992; Frei and Robinson 1999). These satellite measurements have confirmed that the average snow covered area in the Northern Hemisphere has decreased by about 10% in the last 30 years, especially in spring and summer (Serreze et al. 2000). This trend agrees with direct measurement of snow depth and extent on the ground in Western North America (Mote et al. 2005) and in general terms with observations extending from North America to Eurasia (Brown 1999). At a European scale, most of the studies have been performed in the Alps, reporting statistically significant decreases since the early 1980s in the mean snow depth, the duration of snow cover and the number of snowfall days; moreover, more pronounced trends have been found in the medium and lower altitudes (Laternser and Schneebeli 2003). This decrease has been mainly attributed to an increase in mean temperature (Scherrer et al. 2004; Hantel and Hirtl-Wielke 2007). There have also been some attempts to simulate the observed snow behaviour and trends using statistical models. For instance, Martin et al. (1997) simulated the seasonal snow cover of the French Alps using an analog method, and McGinnis (1997) applied multiple regression and neural networks daily snowfall in Colorado, explaining 70% of the daily snowfall variance.

In the case of the Iberian Peninsula, the studies about this topic focus on snow pack and snowfall on the Pyrenees region. For instance, using a dataset from 1985 to 1999, López-Moreno (2005) found that snow depth measures at the end of April were highly correlated with precipitation records during the previous months and also with April temperature; using these relationships they inferred a snow pack series for the period 1950-1999, in which they detected a significant decreasing trend. In the present study, we analyze the regional observed snow trends in a different and broader area: Northern Spain. To this aim, we consider daily snow occurrence data from a network of 33 stations during the period 1957-2002 (observations of snow cover or snow depth are not available in this region). Using this data we analyze annual snow frequency (measured as the number of snow days) uncovering a linear decreasing trend since the mid seventies. Moreover, we simulate daily snow occurrence using an analog-based statistical downscaling model, inferring the local snow observations from large-scale circulation fields provided by a historical reanalysis database (ERA40). The simulated annual frequencies (obtained by aggregating the downscaled values of daily occurrence) reproduce very well the observed annual variability and trends.

The paper is organized as follows. In Section 2 we describe the region of interest and the observed and reanalysis data used in the work. Then, we analyze annual snow frequency and perform a linear analysis to uncover significant trends in the data (Section 3). In Section 4 we simulate daily snow occurrence using statistical downscaling models and we study the annual trend

and interannual variability of the simulated data (Section 5). Finally, some concluding remarks and future work are described in Section 6.

2. Area of Study and Available Data

2a. Local Observations from Meteorological Stations

The study is performed with daily snow occurrence data from 33 stations belonging to the Spanish Meteorological State Agency (AEMET, former National Meteorological Institute, INM), selected from a homogeneous dataset for precipitation for the period 1957-2002 (see Figure 1a). Only 7% of missing data is allowed for the analysis period, defined by the availability of ERA40 data (from September 1957 to August 2002), excluding summer months June, July and August, when practically no snow occurs. A further 5% restriction on the maximum missing values per year is imposed when analyzing annual frequencies, giving a maximum of 14 missing days per year. These percentages were chosen as a compromise between the quality of the resulting time series and the number of stations considered for the study. The region has a complex orography, with station heights ranging from 60 to 1350 metres (see Table 1 for more details).

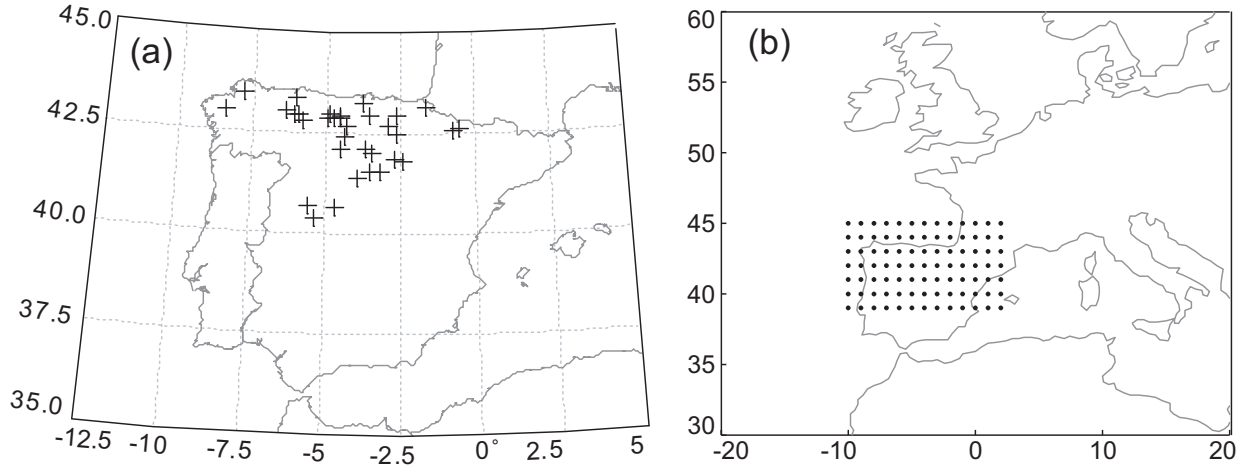


Figure 1: (a) Spatial distribution of the stations used in the study (see Table 1 for geographical details). (b) The reanalysis $1.0^\circ \times 1.0^\circ$ lat and lon grid used to characterize the atmospheric pattern.

The observations account for a 24-hour period, from 06 UTC to 06 UTC of the following day, and refer to snow occurrence. Our database only indicates whether snowfall was reported, it doesn't give information on the amount of snow that fell; moreover, measurable snow on the ground is not even required to issue a snow occurrence report. This data allows studying both the Daily Snow Occurrence (DSO) and the snow frequency, measured as the annual Number of Snow Days (NSD); note that both variables are discrete, the former being a binary variable, $DSO \in \{0, 1\}$, and the latter a positive variable, $0 \leq NSD \leq 365$. Moreover, by convention, the years considered in this study don't correspond to natural years: they cover the period from September to August, preventing the winters to be artificially split into two separate years (this criterion is maintained throughout the whole study).

In this paper we analyze trends, so special care must be taken with the missing data and the non-overlapping periods covered by each of the stations. Figure 2 shows the number of stations

with data for each year and the resulting mean height. Both features remain approximately constant from 1970 onwards, which includes the period in which a significant trend is detected (from 1975 onwards, see Section 3a). No trend analysis has been performed for the period 1957-1970 since the number of stations available each year varied significantly during this period and, thus, the trend results may be biased (see, e.g. Weatherhead et al. 1998, for more information about trend detection influencing factors).

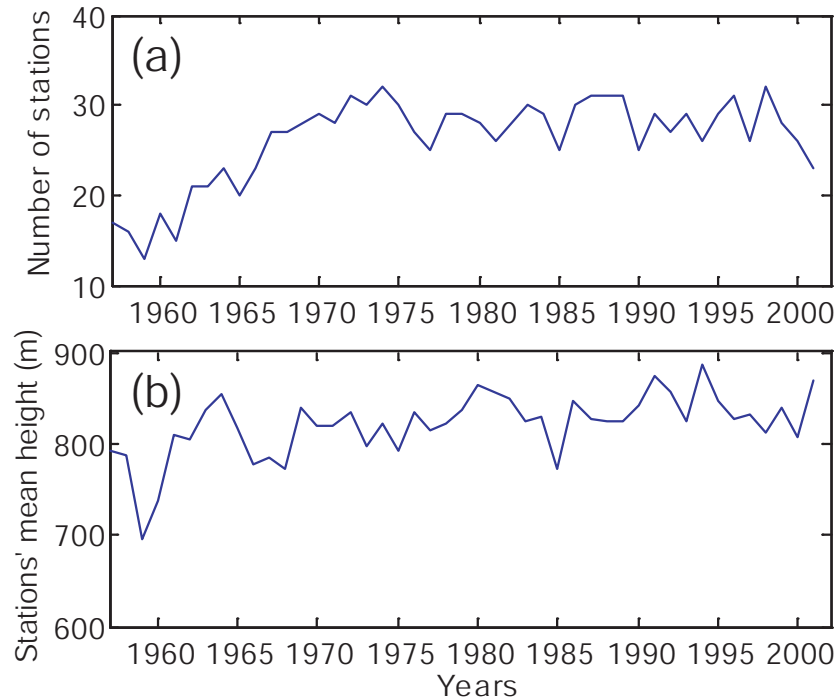


Figure 2: a) Number of stations used each year to calculate the mean annual frequency of snow days. b) Mean height (in meters) of the previously mentioned stations. The characteristics remain approximately constant from 1970 onwards.

Other variables such as precipitation occurrence and maximum, minimum and mean temperature were also analyzed in order to explore possible correlations with snow observations. In the case of temperatures, only a subset of 16 stations satisfying the same quality criteria as for snow was available. Therefore, although there are only slight variations when restricting the snow analysis to the above 16 stations, we shall only use the reduced data set to perform comparative studies involving snow/precipitation and temperature.

2b. Reanalysis Data from Atmospheric GCMs

In this study we also consider large-scale circulation fields provided by a historical reanalysis database. In particular we use the ERA40 reanalysis data from the European Centre of Medium-Range Weather Forecasts (ECMWF), containing gridded atmospheric fields obtained from a T159L60 model running from September 1957 to August 2002 at a horizontal resolution of around 100 km (Uppala et al. 2005). The reanalysis provides us with a set of 6-hourly fields (at 00, 06, 12 and 18 UTC) at different pressure levels. The atmospheric configuration of a certain day over a certain area is approximated through a combination of these fields; the resulting vector is called the *atmospheric pattern*.

In particular, since the geographical area of interest in this work is the northern part of the Iberian Peninsula, we defined an atmospheric pattern on the $1.0^\circ \times 1.0^\circ$ lat and lon grid shown in Figure 1b considering a set of large-scale variables which are physically related to snow precipitation:

$$\mathbf{x}_{12-24} = (Z_{12}^{500}, Z_{12}^{1000}, T_{12}^{500}, T_{12}^{850}, RH_{12}^{850}, Z_{24}^{500}, Z_{24}^{1000}, T_{24}^{500}, T_{24}^{850}, RH_{24}^{850}) \quad (1)$$

where X_i^j denotes the j -pressure level field of variable X at i UTC. Z stands for geopotential height, T for temperature and RH for relative humidity. This pattern defines the daily atmospheric configuration as seen by a GCM and will be used in Section 4 as input for the statistical downscaling method, working in perfect prognosis conditions (i.e., assuming that the only model error is given by the analysis procedure). More details about the choice of the pattern and the downscaling method are given in Section 4.

In order to reduce the dimensionality (91 grid points \times 5 fields \times 2 analysis times) and the spatial correlation of the atmospheric patterns, a Principal Component Analysis (PCA) was performed using the ERA40 period (see Preisendorfer and Mobley 1988, for an overview of PCA in meteorology). Since the variables have different variances, we standardized the data by removing the mean and dividing by the standard deviation, independently for each variable X_i^j and grid point. The first 50 Empirical Orthogonal Functions (EOFs), accounting for more than 98% of the variance, were kept to define the atmospheric patterns, obtaining the corresponding PCs (similar results were found when the standardization process was applied independently for each variable, but considering the spatial mean and standard deviation of the grid points). Note that, in this study, PCA is just applied to compress the data, removing part of the high-frequency spatial variability; this procedure has been shown helpful for downscaling applications (Gutiérrez et al. 2004). The PCA analysis was performed considering the covariance matrix, obtaining the EOFs as their eigenvectors; more efficient methods such as Singular Value Decomposition (SVD) were not necessary in this case.

3. Analysis of Annual Snow Frequency

In the first part of the study the main objective is to analyze the historical records of snow occurrence, looking for trends in the observed annual frequencies. To this aim we use the annual NSD, ranging from 1.5 to 50 in the chosen stations (see Table 1 for more details) and exhibiting an increasing value as a function of the stations' height: snow occurrence increases linearly and smoothly for low-altitude stations and then a crossover appears at approximately 1000 meters with higher increments for the higher stations (see Figure 3). This is the reason why in some sections a separate analysis is performed for stations with high or low altitudes (this distinction is also made due to different performance when applying statistical downscaling methods, shown in Section 4). Note that considering stations with higher snow frequencies is basically equivalent to considering stations located at higher elevations.

3a. Observed Trends

The annual frequency of snow (NSD) averaged over all 33 stations is represented in Figure 4a. This figure exhibits two different trend periods: up to around the mid-seventies the number of snow days per year increased over time, decreasing afterwards until present. As mentioned in Section 2a, only the second period will be further analyzed since the number of stations available each year as well as the resulting mean height remain approximately constant from 1970 onwards.

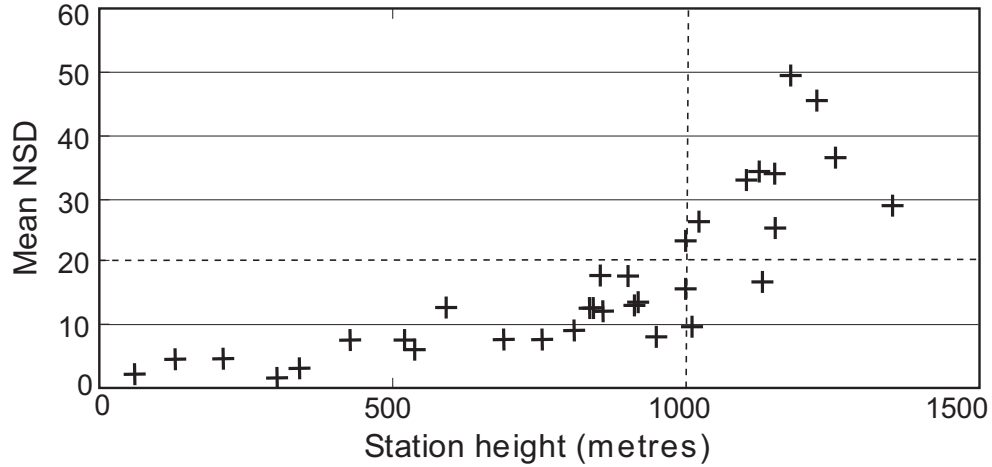


Figure 3: Relationship between the mean Number of Snow Days (NSD) per year and the stations' height. Considering stations with more than 20 days of snow per year is practically equivalent to considering stations above 1000 metres.

A linear fit from 1975 onwards to the annual data gives a decreasing significant trend (at a confidence level of 95%) of around half a day of snow per year (a 2% yearly relative decrease), which means around 13 less days of snow per year in 2002 than in 1975 (an overall relative decrease of 54%). When analyzing separately each season (Figure 4b), the trend is found to be significant only in winter and spring; in absolute values the winters have suffered the largest decrease, with around 8 fewer days of snow per year in 1975-2002 (a 60% relative decrease in this season), followed by spring, with an overall loss of 3.5 annual days of snow (a 50% relative decrease); note that the remaining difference is due to autumn and summer snow.

Similar annual trends for snow occurrence can be observed if we consider separately the stations with high altitude (see Figure 4c), or high snow occurrence (over 20 NSD per year on average for the whole period), and those with low occurrence/altitude (see Figure 4d). High stations show a reduction of around 23 annual days of snow between 1975 and 2002, while low stations have suffered a reduction of around 7 annual days for the same period. This represents a 2% relative yearly decrease in both cases. Despite the overall negative trend since the mid-seventies, a high interannual variability can be observed in both cases.

A further analysis of these figures shows that the last decade has a different behaviour: a stabilization of the decreasing trend seems to occur and the interannual variability decreases as well; thus, removing the last decade would give us an even more pronounced decrease in snow frequency (-0.55 days of snow per year instead of -0.48 ; during the last decade the variation is only of 0.06 days). Nonetheless, the trend analysis was performed for the whole period (from 1975 onwards) corresponding to the observed mean temperature increase.

3b. Correlation with Temperature and Precipitation

To look for possible relationships between snow and other variables, we consider the annual rain frequency and the mean temperature. In this case, in order to keep the consistency of the results, we restrict the study to the subset of 16 stations with temperature records (those marked with “*” in Table 1). The snow trends and interannual variability obtained for this reduced set are very similar to those obtained with the 33 stations (the results are not shown).

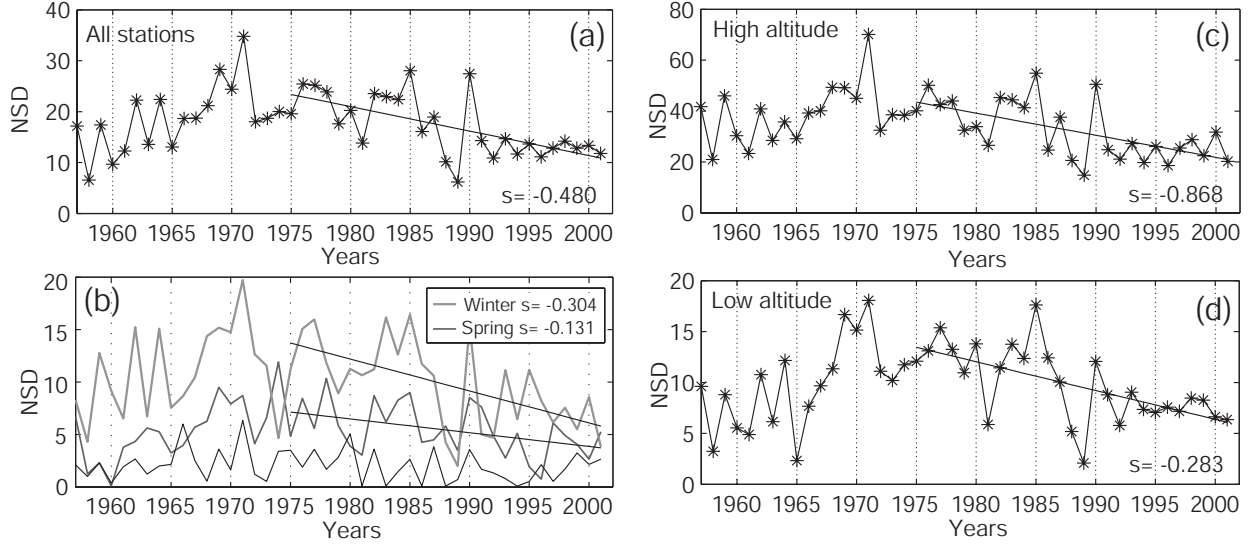


Figure 4: Mean annual frequency of NSD and the regression line for the period 1975-2002 for (a) all stations, (b) all stations analyzing each season separately, (c) stations with high occurrence (more than 20 NSD on average over the ERA40 period), and (d) stations with low occurrence (less than 20 NSD). The value s corresponds to the regression's slope, measured in days/year and it is only displayed, together with the regression line, when it is significant (with a 95% confidence level).

First, we consider the binary series of observed precipitation, that is, the number of days in which some type of precipitation occurred (either in the form of snow or rain). In this case, the annual frequency shows no significant trend during the same period 1975-2002 (see Figure 5a). Similar results can be found in Rodrigo and Trigo (2007) and Gallego et al. (2006), although the data and the period selected for these studies are quite different. Therefore, the reduction in the annual frequency of snow days doesn't seem to be related to an overall decrease of days with precipitation.

We also studied the annual mean temperature for the same period (see Figure 5b). This figure shows that the mean temperature has increased with a positive trend of around 0.056°C per year—a relative increase of 0.5%—(a significant trend at a confidence level of 95%), which means an increase of around 1.5°C during the period 1975-2002. Note that the mean annual temperature curve with the characteristic minimum in the mid seventies is in agreement with previous results (see, e.g., Brunet et al. 2007).

A simple visual inspection shows a clear inverse correlation between the average annual snow frequency and the annual mean temperature, decreasing and increasing respectively since the mid-seventies. With the aim of further analyzing this correlation, both variables are standardized and plotted together in Figure 6a. Annual anomalies (with respect to the whole period 1957-2002) of snow days and mean temperature have a Pearson's correlation coefficient of -0.72 . Both yearly series have been smoothed with a 5-year running mean (solid lines) and, in this case, the correlation coefficient goes up to -0.80 . Three more or less distinct periods can be observed: two periods in which the temperature was above normal and snow frequency under the mean (until the late 1960s and since the mid 1980s) and a cooler period with a greater frequency of snow events between these two periods.

When analyzing separately the correlation of maximum and minimum temperature with snow

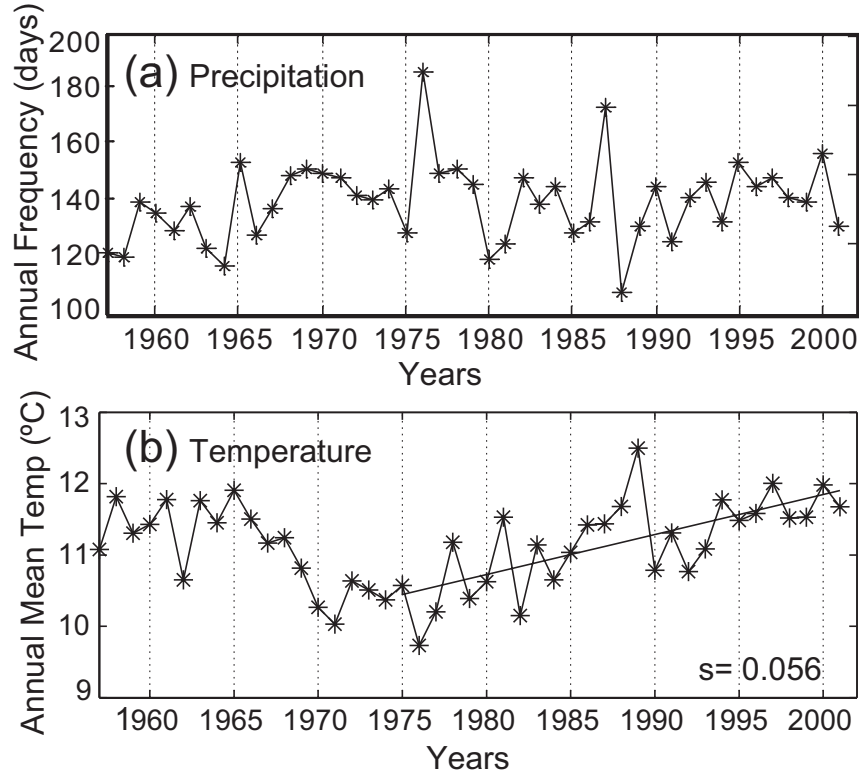


Figure 5: (a) Annual frequency of precipitation days and (b) annual mean temperature. Both figures correspond to averages over the subset of 16 stations. The regression's slope for temperature, measured in $^{\circ}\text{C}/\text{year}$, is shown, whereas precipitation has no significant trend.

days one finds out that the contribution of maximum temperature is much more important. Mean annual maximum temperature and annual snow frequency have a correlation coefficient of -0.72 , while the coefficient goes down to -0.58 for minimum temperature. This is not a surprising result since snow events are usually related to the arrival of a very cold and unstable air mass, which affects both maximum and minimum temperatures while very low minimum temperatures occur frequently during clear winter nights —under the influence of a strong anticyclone— when no precipitation takes place.

The annual correlation coefficient with precipitation occurrence is 0.27 and reaches 0.38 for the 5-year moving average data (Figure 6b). Snow frequency seems to follow the precipitation curve, except for the last period since 1990 when the stabilization takes place. In general terms, positive/negative anomalies of snow are associated with negative/positive anomalies of temperature and positive/negative anomalies of precipitation, except for this last period. Mean annual temperatures continue to increase while the annual NSD seems to remain more or less constant, perhaps due to the simultaneous positive precipitation anomalies.

A more detailed analysis considering high and low stations and seasons separately reveals complex relationships among these three variables (see Table 2); temperature alone does not explain the observed snow variability in most of the cases, although it seems to be the driving factor for the 1975-2002 decreasing trend; similar results have been found for the Swiss Alps by Scherrer et al. (2004). For low stations the correlation with temperature is clearly greater than with precipitation but both values have similar magnitudes when analyzing high stations. This becomes even more clear in winter, when the annual correlation of snow is higher with precipitation than with temper-

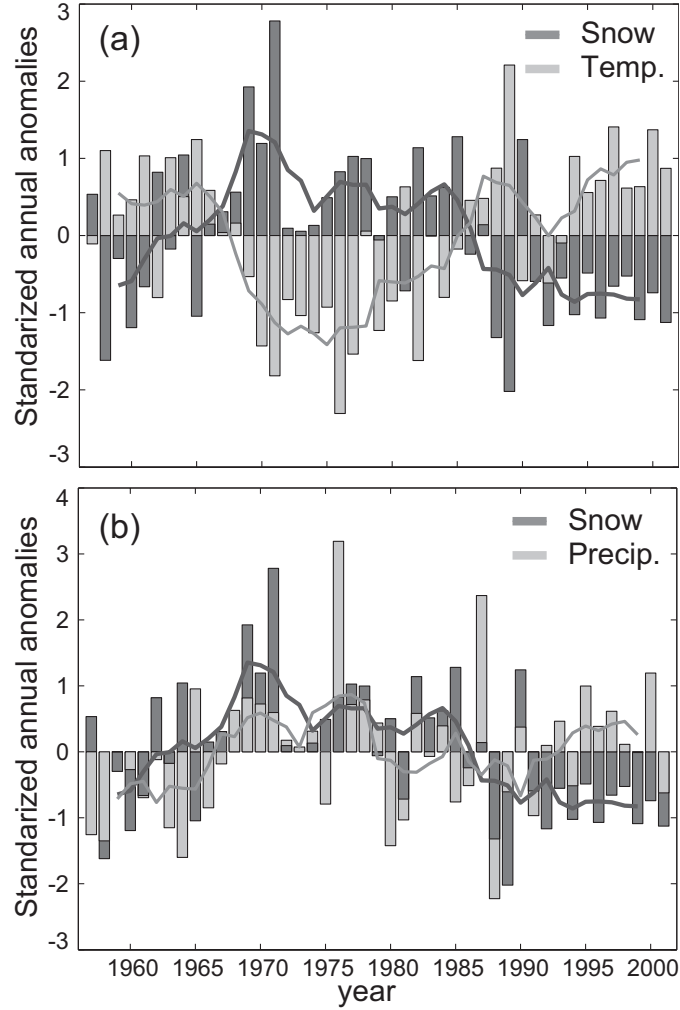


Figure 6: (a) Annual snow frequency and mean temperature standardized anomalies with respect to the period 1957-2002. Continuous lines represent smoothed anomalies using a 5-year moving average for temperature and snow respectively. (b) As in (a) but for precipitation.

ature (0.62 and -0.59, respectively). Winter anomalies of snow, temperature and precipitation are very similar to the annual anomalies represented in Figure 6 (not shown in this paper).

4. Statistical Downscaling. Daily Snow Occurrence

In the second part of this study we test the capability of statistical downscaling techniques to simulate the variability and trends observed in the snow frequency when locally projecting (downscaling) the snow from a large-scale atmospheric pattern given by a GCM (see Wilby and Wigley 1997; Zorita and von Storch 1999, for a general introduction to statistical downscaling methods). To this aim, we focus on the binary variable Daily Snow Occurrence (DSO) and analyze the performance of a statistical downscaling method based on analogs (or nearest neighbours) following the work by Gutiérrez et al. (2004).

We considered the ERA40 reanalysis data described in Section 2b and worked with large-scale variables which are physically related to snow precipitation (upper and low-level circulation, cold

fronts, humidity conditions), bearing in mind that some variables are poorly represented by Global Circulation Models (GCMs). Several tests for different combinations of variables and horizontal and temporal resolutions were performed (see Section 4c for more details), obtaining the best performance to reproduce DSO with the pattern given by Equation (1). Moreover, as we shall see in Section 5, the above pattern also produced the best results to reproduce annual snow frequency. Thus, although most of the snow studies focus only on the relationship between snow observations and temperature, in this study we found that simple patterns obtained considering only temperature fields performed worse than Equation (1), which includes more variables and, hence, can also take into account the correlations with precipitation found in Section 3.

Furthermore, although most of the studies which link local snow observations to atmospheric patterns use large spatial patterns in order to account for large-scale flow, these studies deal with seasonal or monthly-averaged observations (see for example Scherrer and Appenzeller 2006). In our case, we work with daily observations and for such a purpose, we found that a smaller grid (including a temporal component by considering the fields both at 12 and 24 UTC) performed better than a larger pattern, in agreement with Gutiérrez et al. (2004).

4a. Analog Downscaling Scheme

The method of analogues introduced by Lorenz (1969) is a simple and powerful downscaling technique which assumes that similar circulation flows, or atmospheric patterns, over a region lead to similar local outcomes. Following this approach, the local DSO values for a particular day can be estimated from the occurrences recorded in a set of similar historical days, obtained according to the similarity of the corresponding atmospheric patterns. In general, it has been shown that the analogue method performs as well as other complicated downscaling techniques (see e.g. Zorita and von Storch 1999), indicating that this poor-man method is an efficient alternative for many downscaling problems.

The different applications of the analogue methodology in the literature differ basically in the atmospheric pattern, the similarity metric, the number of analogs considered, and the method applied to infer a local value from the resulting ensemble of analogues. In this paper, in order to downscale the local DSO values for the 33 stations for a particular day, represented by the atmospheric pattern given by Equation (1), we first compute an ensemble of 30 analog days (those historical days with closest patterns to the given one) using the standard Euclidean metric. Then the occurrence of snow for each station is estimated using the mean of the 30 corresponding historical observations. Since snow observations are binary, the mean gives the relative frequency within the analog set and, thus, the estimation of DSO is given in a probabilistic form, ranging from 0 (no occurrence) to 1 (occurrence). An advantage of the analog-based methods is that missing values can be handled very easily, since days with missing observations can be discarded from the set of candidate analogues.

In order to test the performance of the analog method to downscale DSO, we consider a leave-one-out cross-validation approach for the period 1957-2002 (Stone 1974). This approach involves using a single observation/day from this period as the validation data, and the remaining observations as the training data. Thus, the ERA40 patterns (Equation 1) for this period, excluding a one-month window around the validation date, are used as the historical database to look for analogs, and the analog method described above is applied to obtain a downscaled probabilistic estimation for DSO, to be compared with the observed one; the one-month exclusion window prevents the validation from persistence and serial correlation overfitting problems (longer windows

were also tried with similar results). This process is repeated for each of the days, such that each observation in the sample is used once as the validation data.

We use the Relative Operating Characteristic (ROC) curves and ROC Skill Areas (RSA) as validation indices to compare the observed and downscaled occurrences for the period 1957-2002. The labels HIR (Hit Rate) and FAR (False Alarm Rate) correspond to the probabilities that DSO is estimated/downscaled to occur given that it has been observed, $P(p = 1|o = 1)$, and it hasn't been observed, $P(p = 1|o = 0)$, respectively (see Jolliffe and Stephenson 2003, for an overview of validation for probabilistic forecasts). Figure 7a shows the ROC curve and the corresponding reliability diagrams (as described in Wilks 2006) for the low and high stations. Since snowfall is a rare or low-probability event, we include in the diagram a vertical dashed line indicating the climatological probability (0.028 and 0.087 for low and high stations, respectively); we also include a borderline dot over the ROC curve corresponding to the HIR and FAR values for the threshold value $prob > 0$; note that this is an extreme situation where snow events are estimated to occur when the downscaled probability is different from zero. Note also that although high HIR values can be attained for appropriate probability thresholds, the rate of false alarms should be kept below the climatological probability. On the other hand, the reliability curves shown in the inset of this figure indicate that the method tends to underforecast in both cases; this deficiency was taken into account with a calibration post-process to render annual snow frequency estimations, as we shall show later.

Figure 7b shows the RSA values of individual stations by displaying the box-and-whiskers plot of the 33 stations (label "all") and of the low and high stations separately. From this figure we can conclude that the performance of the method is better for high stations, with larger frequency of snow occurrence.

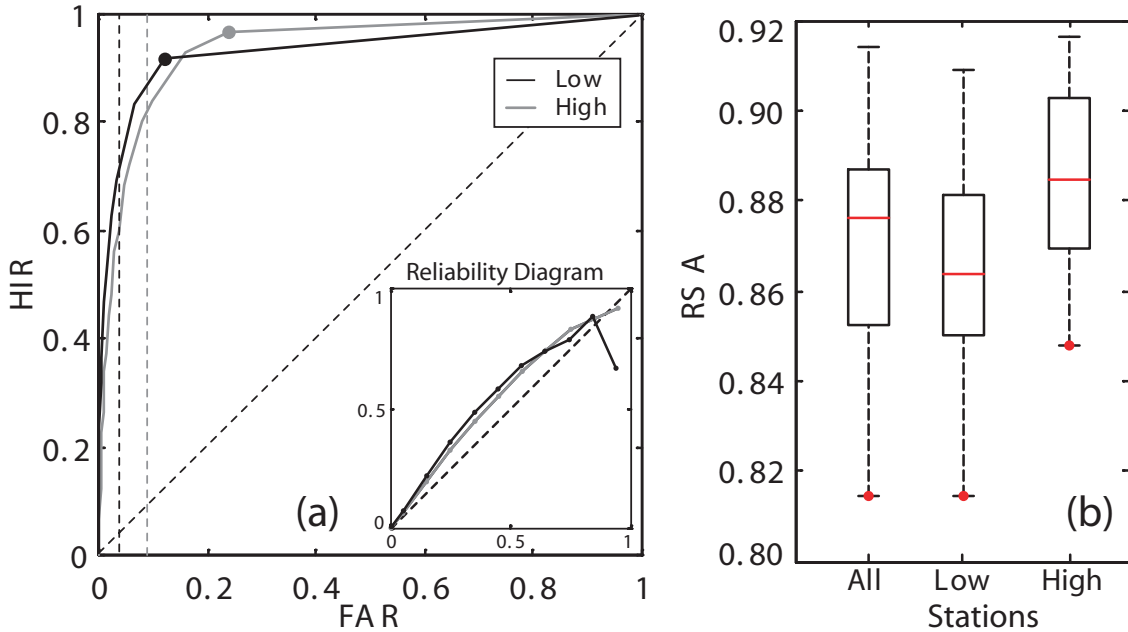


Figure 7: (a) ROC curve and reliability diagram (the probability downscaled for snow occurrence \hat{p} on the x-axis and $P(o = 1|\hat{p})$ on the y-axis) for the subsets of low and high stations. (b) Box-and-whiskers plots for the individual RSA for the 33 stations (label "all") and for the low and high stations separately.

The above cross-validation approach is commonly used in the field of machine learning to determine how accurately a method is able to estimate data that it was not trained on (Schaffer 1993). However, in a climatological context, this approach only provides a descriptive (or interpolation) skill, since it uses the data available from all time steps. Therefore, in order to obtain true predictive skill, we also perform a validation experiment partitioning the data (16434 days) into separate training (1957-1980) and test/validation (1981-2002) sets. Thus, the downscaled values are obtained using only past information of the observed occurrences, providing a forecast of the present ones. The mean of the 33 cross-validation RSA values obtained with this approach is 0.875 (the quartiles are 0.852, 0.877, and 0.907, respectively). These results are very similar to those corresponding to the cross-validation case shown in Fig. 7b (mean 0.870 and quartiles 0.852, 0.876, and 0.887). We also used different training/test periods obtaining similar results, with mean differences lower than 2% in all the cases.

4b. Calibration using Annual Snow Frequency

In this section we analyze the problem of calibrating the probabilistic outputs of the statistical downscaling procedure to produce a deterministic forecast representing correctly the climatological 1957-2002 frequency. To this aim, the daily probabilities downscaled for snow occurrence are converted to 0s (no snow) and 1s (snow day) using a probability threshold which depends on the climatological mean annual frequency of each station. This threshold is empirically chosen to obtain a downscaled mean annual frequency equal to the observed one over the whole period 1957-2002. Obviously, the threshold will be higher for those stations in which snow is more frequent (Table 1 shows the obtained thresholds).

Figure 8 shows an illustrative example of the above calibration process for 150 consecutive days in the station with the highest snow occurrence (Genestoso, see Table 1). The continuous lines represent the probabilistic forecasts obtained with the analog method and the black dots represent the binary snow observations. The optimal calibration threshold for this station for the period 1957-2002 is 0.31. Therefore, all those days in which the probability forecast was higher than 0.31 (red curve above the dotted line) were considered as snow days and those under 0.31 as no snow events (red curve under the dotted line). The grey dots are the projection of the observed snow days onto the probabilistic forecast and, therefore, all the grey dots which lie above the threshold dotted line are a correct detection of a snow day.

The resulting average HIR and FAR values of the binary calibrated deterministic forecast are around 55% and 2%, respectively (individual validation indices for each station are included in Table 1). Note that FARs are very low due to the fact that snow is a rare event (an annual frequency of 5% on average). Thus, we also considered the False Alarm Ratio (FAR2), $P(o = 0|p = 1)$; in this case, the average value of the stations is around 45%. Figure 9 shows the HIR and FAR2 values for the individual stations as a function of the annual frequency of snow days. There is a clear separation between those stations with less than 20 days of snow per year on average (low stations) and those with more than 20 days (high stations). The HIR is only clearly better than the false alarm ratio for the high stations; for these stations the HIR ranges from 58 to 76% whereas the FAR2 ranges from 30-40%. Therefore, the DSO forecast using the statistical downscaling method described in this section is only useful in practice for this second group of stations.

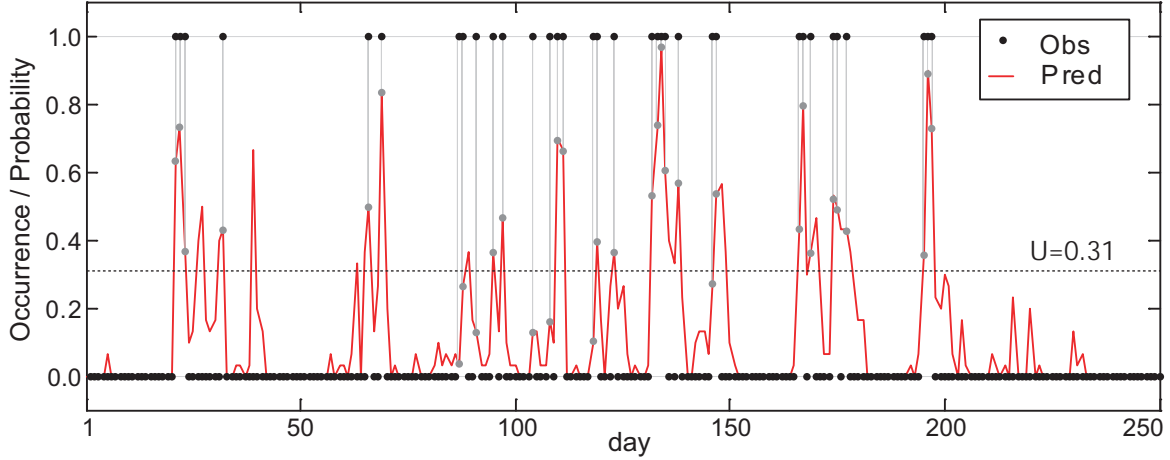


Figure 8: Observations (black dots) and probabilistic forecasts (continuous line) of snow at the station Genestoso for a period of 150 days. Those days in which the probabilistic forecast is above 0.31 –the optimum threshold which equals the annual climatology of observations and deterministic forecasts during the period 1957-2002– are converted to a deterministic forecast of 1. The grey dots are the projection of the observations onto the probabilistic curve: if the grey dot is above the threshold dotted line, the snow event is correctly detected for that specific day.

4c. Sensitivity Studies

As mentioned at the beginning of this section, different variables were tested to find the optimum predictor patterns to be used in the downscaling method for snow events. Apart from trying different fields, several sensitivity analysis were carried out to study the influence of the resolution, the number of analogs, the downscaling method, etc., in the performance of the statistical downscaling method. Figure 10 shows the results of some of these sensitivity studies by plotting the HIRs of two different methods, one versus the other. This figure shows that similar results are obtained considering a dynamic pattern or a static pattern (panel a); in this case, searching for analog fields of temperature, geopotential and relative humidity both at 12 UTC and at 24 UTC gives similar results than those obtained by only looking for analog fields at 24 UTC. Moreover, similar results are also obtained when decreasing the resolution of the predictor fields from 1.0° to 2.0° , considering in both cases the dynamical predictors defined at 12 and 24 UTC (panel b). This is an important result if one wants to apply this method to climate models which have lower temporal and spatial resolutions.

Finally, we also tested an alternative statistical downscaling method recently introduced as an extension of the standard analog methodology based on weather types (see Gutiérrez et al. 2004, for more details). In this case, the method works by first applying a clustering algorithm to the reanalysis patterns in order to obtain homogeneous groups, or weather types. In our case, we applied the k -means algorithm to the patterns given in Equation (1) considering $k = 100$ weather types, as described in Gutiérrez et al. (2004). The obtained groups are used as *analog ensembles* and each pattern to be downscaled is assigned to the closest group, obtaining the downscaled value for the local stations from the occurrences within the pre-defined ensembles. This downscaling procedure is much simpler than the standard analog approach but it gives lower HIR values (see Figure 10c)

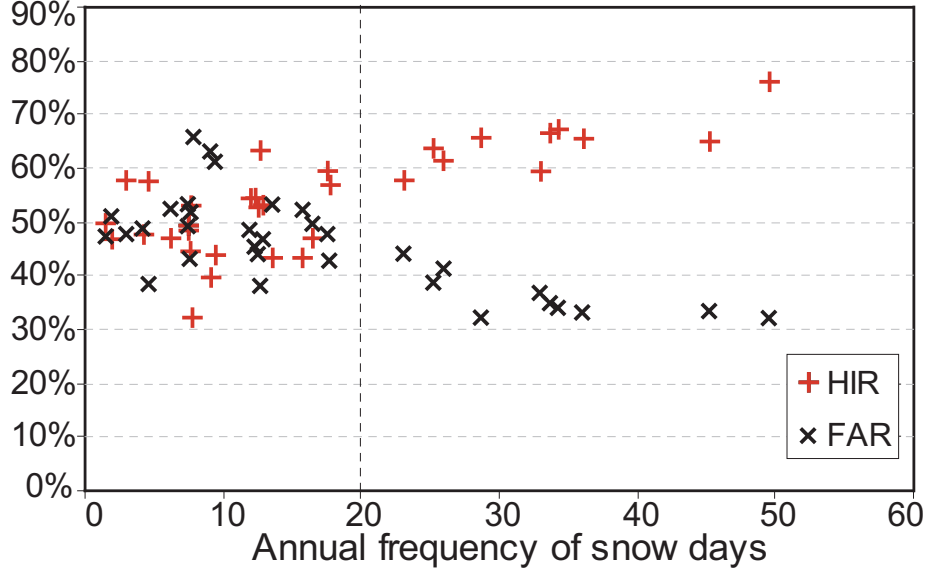


Figure 9: Hit Rate $P(p = 1|o = 1)$ and False Alarm Ratio $P(o = 0|p = 1)$ as a function of the mean annual number of snow days, each dot represents a different station. Daily deterministic forecasts of snow events are only useful for those stations with more than 20 days of snow per year on average, when the hits are clearly higher than the false alarms.

5. Simulation of Annual Snow Frequency Trends

In this section we study whether the downscaling method described in the previous section is able to simulate the annual snow frequency behaviour described in Section 3a and represented in Figure 4. Note that now we are no longer interested in being precise at downscaling snow on a daily basis, but rather in simulating correctly annual frequencies of snow days, capturing the inter-annual variability and the observed trends.

To this aim, the deterministic daily downscaled forecasts of snow occurrence are accumulated for each year, thus obtaining a time series of predicted annual frequency of NSD for each station. These annual series are then averaged over all stations, obtaining a simulation of the mean annual frequency which is then compared with the observed mean annual frequency (see Figure 11a). As it can be observed from this figure, the method simulates very well the observed annual frequency (the Root Mean Square Error, RMSE, is 2.56), not only the decreasing trend observed since 1975, but also the high inter-annual variability (as for example, the abrupt increase observed in 1990, when 20 more days of snow than the previous year were registered). Note that the calibration process only has an effect on the climatological average values of the NSD for the whole period 1957-2002 (i.e., it acts as a translation of the whole predicted curve), but has no effect in the inter-annual variability nor the trend. Note also that we are working under the perfect prognosis approach and, thus, the daily atmospheric patterns for a particular year are expected to appropriately represent the atmospheric meteorological conditions that occurred. Therefore, estimating the annual frequency of a certain year is easier than estimating accurately daily snow occurrence.

In Figures 11b and c the predicted annual frequencies for high and low stations are analyzed separately. The RMSE values for high and low stations are 3.96 and 2.36 respectively, which corresponds to standardized errors of 0.34 and 0.61. The method simulates better the annual frequency of the high stations, especially when reproducing the inter-annual variability. This is an expected

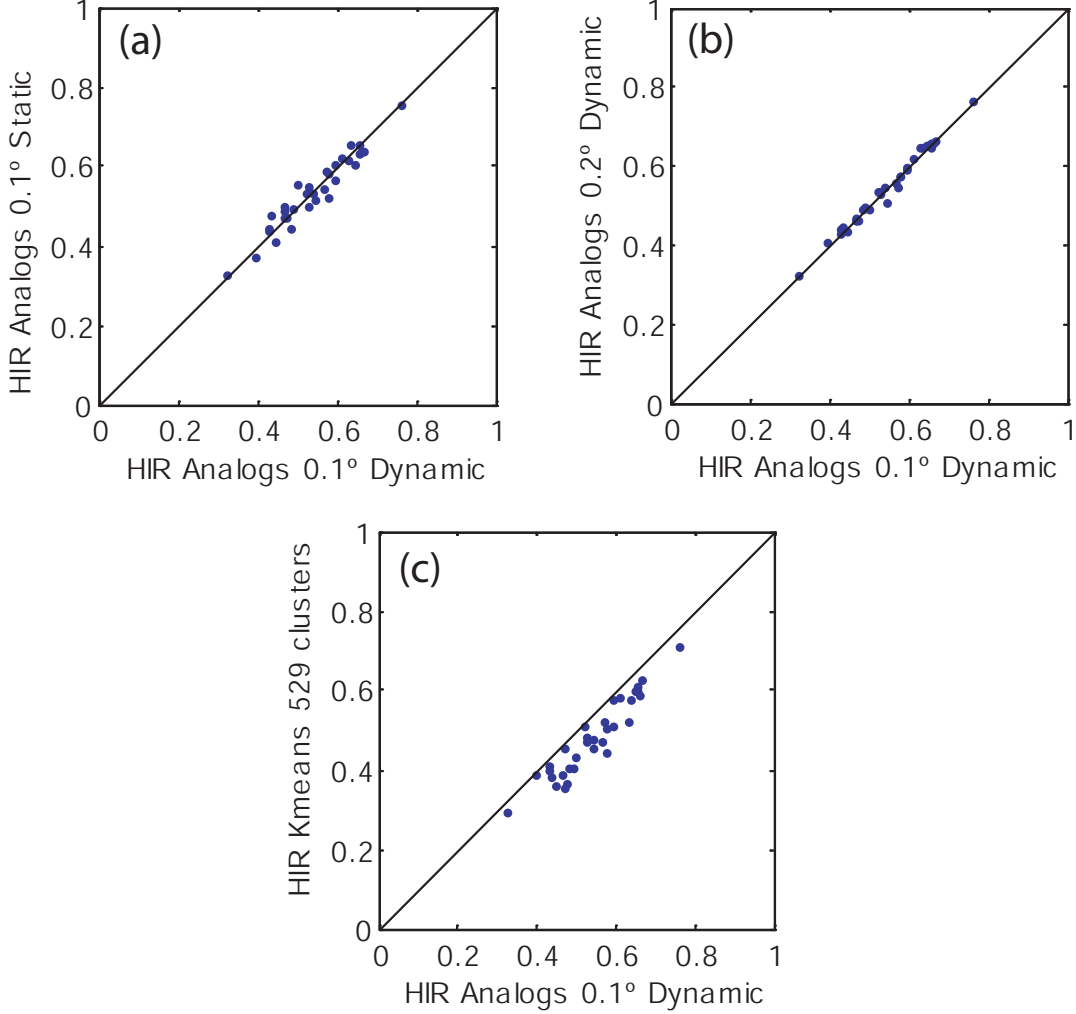


Figure 10: A sensitivity study of certain aspects of the downscaling method measuring performance using only hit rates. (a) A comparison between dynamical and static patterns. (b) A comparison between different resolutions for the predictor fields. (c) A comparison between the standard analog method and a weather-typing approach.

result since in the previous section the downscaling method reproduced better the daily snow occurrence of the high stations. Nevertheless, the observed decreasing trend is well reproduced in both cases.

One could argue that similar results could be obtained using a much simpler pattern, considering only large scale temperature variations, in order to infer snow trends. The great influence of temperature has been studied in previous studies, such as Hantel and Hirtl-Wielke (2007) for Alpine snow cover. We tried to simulate the annual NSD using only the temperature field at 850 mb as our predictor. Although the general features of snow variability were captured, the method was clearly worse at simulating the decreasing trend: the predicted slopes were -0.37 , -0.59 and -0.24 for all, high and low stations respectively. For high stations, the temperature-based pattern predicts a decreasing trend more than 30% lower than the observed one. This is in agreement with the results found in section 3, where the correlation with temperature for high stations was less important than for low stations, and even had similar magnitudes to precipitation correlations. Thus,

simple statistical methods based on temperature do not provide an appropriate methodology for simulating annual snow frequency in this region and, thus, general statistical downscaling methods are required to simulate the observed behaviour.

Finally, we performed the same experiment partitioning the data into separate training (1957-1980) and test/validation (1981-2002) sets, obtaining very similar results for the simulated annual snow frequencies and for the trend; in particular, the RMSE of the validation period for all the stations was 2.84.

6. Conclusions and Future Work

In this paper we have undertaken the study of snow occurrence and frequency in a broad area of Northern Spain characterized by a high inter-annual variability, with low to medium altitudes (the 33 stations' height used in the study range from 60 to 1350 metres). This characteristic makes snow occurrence very sensitive to both temperature and precipitation fluctuations in this region. Therefore, this work complements previous snow studies performed in regions with higher mean elevations (e.g., Pyrenees, such as López-Moreno 2005; López-Moreno and Vicente-Serrano 2006).

We have analyzed snow frequency in terms of the annual number of snow days, obtaining a 50% relative decrease since the mid seventies; moreover this magnitude is similar for low and high elevations and is basically due to winter and spring decreases. The study only takes into account the binary variable snow occurrence; however, the magnitude of the trend is such that one would expect similar behaviours for variables such as the mean snow depth or the snow cover duration; for instance, Laternser and Schneebeli (2003) found significant decreasing trends for all three variables in the Alps since the 1980s. Moreover, López-Moreno (2005) found a significant decreasing trend in snow depth measures in the Spanish Pyrenees during the second half of the 20th century.

A high correlation has been found between snow occurrence and mean temperature (-0.80, -0.69 and -0.76 for all, high and low stations respectively), whereas medium to low correlations (0.38, 0.47 and 0.42, respectively) have been obtained for the case of precipitation occurrence (in both cases we refer to 5-year moving average correlations). Therefore, although temperature appears to be the main driving factor, especially for low stations, the role of precipitation becomes important in some cases, such as in high stations, especially in winter; this is in agreement with the results obtained in Clark et al. (1999) for snow extent. Furthermore, in a later study in the Pyrenees, López-Moreno and Vicente-Serrano (2006) related the observed decreasing snow trend in April's snow extent to a reduction in precipitation from January to March (the snow stakes used in this study were all located above 1700 metres, where temperature is usually below zero during the winter months). Note that this is also in agreement with the important correlations found with precipitation for high stations in our study.

On such a basis, a complete pattern including temperature, circulations and humidity variables was used to simulate these observed trends through an analog-based statistical downscaling method. We found that the method is only suitable for predicting daily snow occurrence in high stations; however, in the case of annual frequencies, the downscaling method is capable of simulating the inter-annual variability and the observed trends for all cases. These results seem to be in disagreement with the belief that statistical downscaling techniques do not provide adequate information at the scales required for snow investigations (Beniston et al. 2003).

The promising results presented in this paper open the possibility to future research in seasonal

prediction or climate change projection of snow frequency. However, several issues have to be cautiously addressed: the study was performed in a perfect prognosis approach, using reanalysis data. Thus, operational applications involving model predictions will include an additional source of error (model error) and further analysis will be necessary in order to test whether the snow frequency can be simulated or not with some skill. Furthermore, when addressing future climate projections one has to assume the hypothesis that the relation between local variables and the large-scale flow still explains a great part of the observed variability, and that this relation can be assumed to be valid also for a changed climate. The analog method will surely perform worse if future weather patterns differ considerably from past weather patterns.

Acknowledgments

The authors are grateful to the Spanish Meteorological State Agency (AEMET) for providing us with partial support and the necessary data for this work. The authors are also grateful to the University of Cantabria, CSIC and the Comisión Interministerial de Ciencia y Tecnología (CICYT, CGL-2007-64387/CLI and CGL2005-06966-C07-02/CLI) for partial support of this work.

References

- Beniston, M., F. Keller, and S. Goyette, 2003: Snow pack in the swiss alps under changing climatic conditions: an empirical approach for climate impact studies. *Theoretical and Applied Climatology*, **74**, 19–31, doi:10.1007/s00704-002-0709-1.
- Brown, R. D., 1999: Northern hemisphere snow cover variability and change, 1915-97. *Journal of Climate*, **13**, 2339–2355.
- Brunet, M., P. D. Jones, J. Sigró, O. Saladié, E. Aguilar, A. Moberg, P. M. Della-Marta, D. Lister, A. Walther, and D. López, 2007: Temporal and spatial temperature variability and change over spain during 1850-2005. *J. Geophys. Res.*, **112**, D12117, doi:10.1029/2006JD008249.
- Clark, M. P., M. C. Serreze, and D. Robinson, 1999: Atmospheric controls on eurasian snow extent. *International Journal of Climatology*, **19**, 27–40.
- Dewey, K. F. and R. J. Heim, 1982: A digital archive of northern hemisphere snow cover, november 1966 through december 1980. *Bulletin of the American Meteorological Society*, **63**, 1132–1141.
- Foster, J., M. Owe, and A. Rango, 1983: Snow cover and temperature relationships in north america and eurasia. *Journal of Climate and Applied Meteorology*, **22**, 460–469.
- Frei, A. and D. A. Robinson, 1999: Northern hemisphere snow extent: regional variability 1972-1994. *International Journal of Climatology*, **19**, 1535–1560.
- Gallego, M. C., J. A. García, J. M. Vaquero, and V. L. Mateos, 2006: Changes in frequency and intensity of daily precipitation over the iberian peninsula. *Journal of Geophysical Research*, **111**, D24105, doi:10.1029/2006JD007280.

- Gutiérrez, J. M., A. S. Cofiño, R. Cano, and M. A. Rodríguez, 2004: Clustering methods for statistical downscaling in short-range weather forecast. *Monthly Weather Review*, **132**, 2169–2183.
- Gutzler, D. S. and R. D. Rosen, 1992: Interannual variability of wintertime snow cover across the northern hemisphere. *Journal of Climate*, **5**, 1441–1447.
- Hantel, M. and L. M. Hirtl-Wielke, 2007: Sensitivity of alpine snow cover to european temperature. *International Journal of Climatology*, **27**, 1265–1275.
- Jolliffe, I. T. and D. B. Stephenson, eds., 2003: *Forecast Verification: : A Practitioner's Guide in Atmospheric Science*. John Wiley and Sons.
- Latnser, M. and M. Schneebeli, 2003: Long-term snow climate trends of the swiss alps (1931–99). *International Journal of Climatology*, **23**, 733–750.
- Lorenz, E. N., 1969: Atmospheric predictability as revealed by naturally occurring analogues. *Journal of the Atmospheric Sciences*, **26**, 636–646.
- López-Moreno, J. I., 2005: Recent variations of snow pack depth in the central spanish pyrenees. *Artic, Antartic and Alpine Research*, **37**.
- López-Moreno, J. I. and S. M. Vicente-Serrano, 2006: Atmospheric circulation influence on the interannual variability of snow pack in the spanish pyrenees during the second half of the 20th century. *Nordic Hydrology*, **38**.
- Martin, E., B. Timbal, and E. Brun, 1997: Downscaling of general circulation model outputs: Simulation of the snow climatology of the french alps and sensitivity to climate change. *Climate Dynamics*, **13**, 45–56.
- Matson, M., C. F. Ropelewski, and M. S. Varnadore, 1986: *An Atlas of Satellite-derived Northern Hemisphere Snow Cover Frequency*. NOAA Atlas, National Oceanic and Atmospheric Administration.
- McGinnis, D. L., 1997: Estimating climate-change impacts on colorado plateau snowpack using downscaling methods. *The Professional Geographer*, **49(1)**, 117–125.
- Mote, P., A. Hamlet, M. Clark, and D. Lettenmaier, 2005: Declining mountain snowpack in western north america. *Bulletin of the American Meteorological Society*, **66**, 39–49.
- Preisendorfer, R. W. and C. D. Mobley, 1988: *Principal Component Analysis in Meteorology and Oceanography*. Elsevier.
- Robinson, D. A., 1993: Global snow cover monitoring: an update. *Bulletin of the American Meteorological Society*, **74(9)**, 1689–1696.
- Rodrigo, F. S. and R. M. Trigo, 2007: Trends in daily rainfall in the iberian peninsula from 1951 to 2002. *International Journal of Climatology*, **27**, 513–529.
- Schaffer, C., 1993: Selecting a classification method by cross-validation. *Machine Learning*, **13**, 135–143.

- Scherrer, S. and C. Appenzeller, 2006: Swiss alpine snow pack variability: major patterns and links to local climate and large-scale flow. *Climate Research*, **32**, 182–199.
- Scherrer, S., C. Appenzeller, and M. Latternser, 2004: Trends in swiss alpine snow days: The role of local- and large-scale climate variability. *Geophysical Research Letters*, **31**, L13215, doi:10.1029/2004GL020255.
- Serreze, M., J. Walsh, F. Chapin, T. Osterkamp, M. Dyurgerov, V. Romanovsky, W. Oechel, J. Morison, T. Zhang, and R. Barry, 2000: Observational evidence of recent change in the northern high-latitude environment. *Climatic Change*, **46**, 159–207.
- Stone, M., 1974: Cross-validation choice and assessment of statistical procedures. *Journal Royal of Statistical Society*, **36**, 111–147.
- Trenberth, K., 1984: Some effects of finite sample size and persistence on meteorological statistics. part i: Autocorrelations. *Monthly Weather Review*, **112**, 2359–2368.
- Trenberth, K., P. Jones, P. Ambenje, R. Bojariu, D. Easterling, A. Klein, D. Parker, F. Rahimzadeh, J. Renwick, M. Rusticucci, B. Soden, and P. Zhai, 2007: *Observations: Surface and Atmospheric Climate Change*. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, chapter 3. 235–336.
- Uppala, S. M., P. W. Kållberg, A. J. Simmons, U. Andrae, V. da Costa Bechtold, M. Fiorino, J. K. Gibson, J. Haseler, A. Hernandez, G. A. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R. P. Allan, E. Andersson, K. Arpe, M. A. Balmaseda, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, S. Caires, F. Chevallier, A. Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Hólm, B. J. Hoskins, L. Isaksen, P. A. E. M. Janssen, R. Jenne, A. P. McNally, J. F. Mahfouf, J. J. Morcrette, N. A. Rayner, R. W. Saunders, P. Simon, A. Sterl, K. E. Trenberth, A. Untch, D. Vasiljevic, P. Viterbo, and J. Woollen, 2005: The ERA-40 re-analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012.
- Weatherhead, E., G. Reinsel, G. Tiao, X. Meng, D. Choi, W. Cheang, T. Keller, J. DeLuisi, D. Wuebbles, J. Kerr, A. Miller, S. Oltmans, and J. Frederick, 1998: Factors affecting the detection of trends: Statistical considerations and application to environmental data. *Journal of Geophysical Research*, **103(D14)**, 149–161.
- Wilby, R. and T. Wigley, 1997: Downscaling general circulation model output. a review of methods and limitations. *Progress in Physical Geography*, **21**, 530–548.
- Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. Academic Press.
- Zorita, E. and H. von Storch, 1999: The analog method as a simple statistical downscaling technique: comparison with more complicated methods. *Journal of Climate*, **12**, 2474–2489.

Station	Height	Lat.	Lon.	NSD	θ	HIR	FAR	FAR2
GRADO *	60	43.38	-6.06	2.00	0.15	0.47	0.00	0.51
SANTESTEBAN *	131	43.13	-1.66	4.36	0.23	0.48	0.01	0.48
VILLACARRIEDO *	212	43.23	-3.80	4.67	0.26	0.57	0.00	0.39
MONTAOS-ORDES	306	43.04	-8.42	1.59	0.19	0.50	0.00	0.47
AS PONTES DE GARCIA *	343	43.45	-7.86	3.14	0.18	0.58	0.00	0.47
CENICERO INDUSTRIAL *	430	42.48	-2.64	7.67	0.23	0.53	0.01	0.43
MIRANDA DE EBRO *	520	42.68	-2.96	7.51	0.23	0.49	0.01	0.49
URRUNAGA 'PRESA'	540	42.96	-2.65	6.30	0.23	0.47	0.01	0.52
VILLARCAYO	595	42.94	-3.57	12.72	0.26	0.63	0.01	0.38
JAVIERREGAY	690	42.59	-0.74	7.69	0.26	0.45	0.01	0.52
MONZON DE CAMPOS *	754	42.12	-4.49	7.61	0.18	0.49	0.01	0.53
OSORNO	809	42.41	-4.36	9.24	0.19	0.40	0.02	0.63
SAN MIGUEL DE BERNUY	839	41.40	-3.95	12.65	0.26	0.52	0.02	0.44
PANTANO DE STA. TERESA *	840	40.67	-5.60	12.46	0.26	0.54	0.02	0.45
ALAR DEL REY *	851	42.66	-4.31	17.60	0.23	0.59	0.03	0.47
SAN ESTEBAN DE GORMAZ *	860	41.57	-3.20	12.11	0.23	0.55	0.02	0.48
RETUERTA	900	42.03	-3.51	17.77	0.28	0.57	0.02	0.43
LINARES DEL ARROYO *	911	41.53	-3.56	12.91	0.26	0.53	0.02	0.47
BESCOS DE GARCIPOLLERA	920	42.63	-0.50	13.58	0.26	0.43	0.02	0.53
TORRECILLA DEL MONTE	949	42.09	-3.69	7.82	0.23	0.32	0.01	0.65
LA MAGDALENA *	998	42.78	-5.80	15.84	0.30	0.43	0.02	0.52
PANTANO DE CERVERA	1000	42.87	-4.53	23.23	0.30	0.58	0.03	0.44
GARRAY	1010	41.82	-2.45	9.59	0.19	0.44	0.02	0.61
PANTANO DE REQUEJADA *	1024	42.91	-4.53	26.13	0.30	0.61	0.03	0.41
BOCA DE HUERGANO	1104	42.97	-4.93	32.98	0.36	0.60	0.04	0.36
PRIORO *	1123	42.89	-4.96	34.24	0.31	0.66	0.04	0.34
AVILA 'OBSERVATORIO'	1130	40.65	-4.68	16.49	0.30	0.47	0.02	0.49
EMBALSE CUERDA DEL POZO	1150	41.88	-2.70	25.32	0.30	0.64	0.03	0.38
RABANAL DE LUNA *	1150	42.93	-5.97	33.80	0.31	0.67	0.04	0.35
GENESTOSO	1180	43.06	-6.40	49.54	0.31	0.76	0.06	0.32
HUERGAS DE BABIA	1222	42.96	-6.09	45.27	0.38	0.65	0.05	0.33
PANTANO DE CAMPORREDONDO *	1253	42.90	-4.74	36.18	0.34	0.66	0.04	0.33
ZAPARDIEL DE LA RIBERA	1353	40.36	-5.33	28.82	0.34	0.66	0.03	0.32
Mean Values	823	42.42	-4.26	16.93	0.26	0.54	0.02	0.45

Table 1: Name, height (m), location (latitude and longitude in degrees), mean annual number of snow days (NSD) and the probability calibration threshold θ considered for the different stations. The last three columns show the results of different verification indices: Hit Rate $HIR \equiv P(p = 1|o = 1)$, False Alarm Rate $FAR \equiv P(p = 1|o = 0)$ and false alarm ratio or $FAR2 \equiv P(o = 0|p = 1)$. The stations have been sorted by increasing height (those marked with an asterisk also have temperature records).

	Annual	Winter	Spring	Summer	Autumn	5-years
Temp All	-0.72 *	-0.63 *	-0.65 *	-0.28 *	-0.19	-0.80 *
Temp High	-0.52 *	-0.59 *	-0.61 *	-0.28 *	-0.30 *	-0.69 *
Temp Low	-0.68 *	-0.60 *	-0.49 *	-0.14	-0.11	-0.76 *
Precip All	0.27 *	0.53 *	0.47 *	0.03	0.05	0.38 *
Precip High	0.39 *	0.62 *	0.52 *	0.15	0.22	0.47 *
Precip Low	0.24	0.40 *	0.33 *	-0.09	-0.13	0.42

Table 2: Annual and seasonal correlations of snow frequency with temperature and with precipitation frequency calculated separately for all, high and low stations. Significant correlations with $p\text{-value} \leq 0.05$ are indicated with an asterisk. The last column is the correlation calculated for a 5-year moving average; in this case, the significance values have been adjusted for autocorrelation (Trenberth 1984).

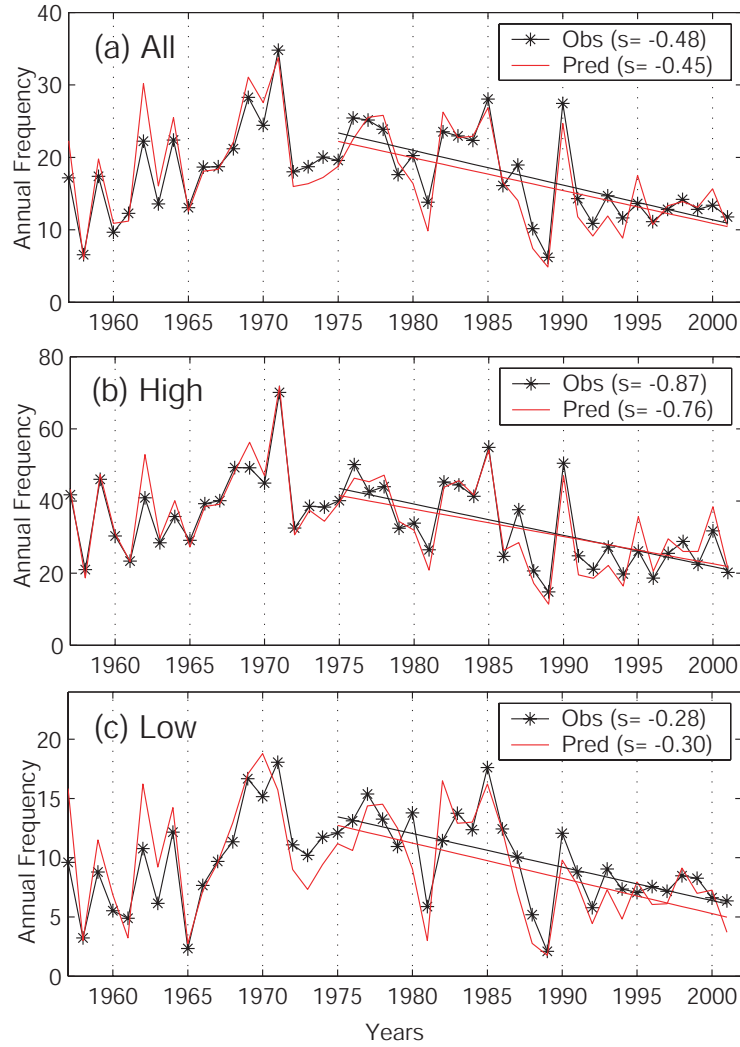


Figure 11: Observed annual frequency of snow days (black line with asterisks), predicted annual frequency (grey curve) and the corresponding regression lines for the period 1975-2002 for (a) all stations, (b) stations with high occurrence (more than 20 days of snow per year on average), (c) stations with low occurrence (less than 20 days of snow per year on average). The values of the observed and predicted slopes, in days/year, are shown in each panel.