HOMOGENIZATION OF SPANISH MEAN WIND SPEED MONTHLY SERIES

José A. Guijarro
State Meteorological Agency (AEMET), Balearic Islands office, Spain (jguijarrop@aemet.es)

Abstract

Monthly mean wind speed data were gathered from all Spanish series with a minimum of 10 years of data in the period 1951-2013, resulting in the selection of 233 series. Monthly wind speed averages were initially drawn from daily wind runs, but since they had too many missing data, mean wind speed recorded at 07, 13 and 18 UTC were obtained as well. These datasets were homogenized by means or the R package Climatol twice: 1) using a ratio normalization of the data; 2) applying a cubic root transformation to the data and standardizing them. Around two thirds of the series were found inhomogeneous through both normalization methods, which gave also similar results in terms of mean RMSE when estimating the series from the neighboring stations and mean SNHT of the final homogenized series. But the overall correlations of the wind series were not good enough, and showed a poor spatial coherence. Wind speed series were then extracted from the NCEP reanalysis to explore their potential value as reference series, but more than 80% of them were found inhomogeneous, probably because of their less noisy nature. Therefore, wind speed seems an element very prone to inhomogeneities, since it is very sensitive to obstacles and surface roughness changes in the surroundings of the observatories, and at the same time difficult to homogenize, because local air circulations as thermal winds may contribute to a significant part of the wind speed values, worsening the correlations between neighboring stations in complex regions. Anyway, wind speed trends were computed from these preliminary homogenization exercises, yielding negative figures mostly ranging between -1 and -2 m/s/century in the colder months of the year.

1. INTRODUCTION

Wind is an important climatic element for many economic areas: agriculture (modulating evapotranspiration), water resources (controlling evaporation from dams and natural surfaces), leisure (outdoor activities, sailing, etc), and renewable energy production. For this reason, many work has been devoted to study its spatial and temporal variability (McVicar *et al.*, 2012, refer 148 papers on wind speed trends).

Wind speed has been traditionally measured in meteorological observatories with cup anemometers, although FUESS type used differential dynamic air pressure, and in recent times sonic anemometers are deployed as well. Changes of instrumentation or calibration drifts (e. g., increase of friction in the rotating axis) are a source of inhomogeneities in the series, as are instrument relocation or changes in the surroundings (new buildings, growing trees, etc), since wind is very sensitive to obstacles, orography, and surface roughness.

Yet many variability studies do not try to homogenize these series, but just to select those having a long period of observation which appear of a reasonable quality according to meta-data, visual inspection or basic comparison with a suitable reference (Dadaser-Celik and Cengiz, 2014).

Wan *et al.* (2010) did a thorough adjustment and homogenization of 117 Canadian wind stations with a minimum of 45 years of observation using the package RHtestV2 (Wang and Feng, 2007), and a recent paper by Azorín-Molina *et al*. (2014) also applied an homogenization package (AnClim, by Stepanek, 2004, using MM5 output as reference series) in their study of 67 wind speed series from Portugal and Spain selected for completeness in the period 1961-2011.

In this work, a more extensive homogenization is applied to most Spanish wind speed series, testing different approaches whose results are discussed, to end with a preliminary evaluation of the trends of the homogenized series.

2. METHODOLOGY

2.1. Data

Monthly mean wind speed data were gathered from all Spanish series with a minimum of 10 years of data in the period 1951-2013, resulting in the selection of 233 series. Monthly wind speed averages were initially drawn from daily wind runs, but since they had too many missing data, mean wind speed recorded at 07, 13 and 18 UTC were obtained as well. These latter values were an 8 % higher in average than those computed from daily wind runs. Figure 1 shows the number of data from both origins, the sharp increase in 1961 being due because data digitization from that year on were prioritized.

To complement observational series, wind speed monthly averages from NCEP reanalysis (Kalnay *et al.*, 1996) were also downloaded from NOAA servers.

2.2. Homogenization method

These series were homogenized with the 'Climatol' R package (Guijarro, 2014), that provides automatic quality control (outlier correction), homogenization (shift correction) and missing data attribution. The package begins by normalizing all data and computing a reference series for each observed series by averaging up to 10 data (if available) at every time step. As reference data are chosen by proximity, nearest data can be used even without any common period of observation with the problem series, taking advantage of short observational series that otherwise would be disregarded.

Series of anomalies are then computed by subtracting the reference series from the original series, allowing a simple detection of outliers (which are rejected if lying beyond a prescribed threshold) and breaks (shifts in the mean). Shift detection is performed by the well known SNHT test (Alexandersson, 1986), applied in stepped windows first to cope with multiple breaks, and then on the whole series to get all the power of the test.

These reference series are not assumed to be homogeneous, but only significantly less inhomogeneous than the original series. Therefore, an iterative application of the detection algorithm from big to small inhomogeneities in successive passes is performed, splitting the series at each noticeable break. Finally, newly computed reference series are straightforwardly used to fill any missing data in the series, including the reconstruction of the split series generated in the break detection process.

This methodology is able to yield results of a quality comparable with other good methods (as shown in http://www.climatol.eu/DARE/testhomog.html), and was applied to monthly wind speed series from wind runs (WRun), wind speed measured three times per day (WSm3) and NCEP reanalysis (WSRe), with two kind of normalizations: ratio to the series means, and full standardization. Ratio normalizations are normally used with variables with a zero lower limit and an L-shape probability distribution, while full standardization (removing the mean and dividing by the standard deviation) is applied to variables with a (near) normal distribution. Therefore, wind speed data were cubic root transformed (when greater than 1.0) in order to normalize their probability distribution.

Finally, trends of the homogenized series were obtained by regression with time, with the help of a post-processing function of the same computer package.

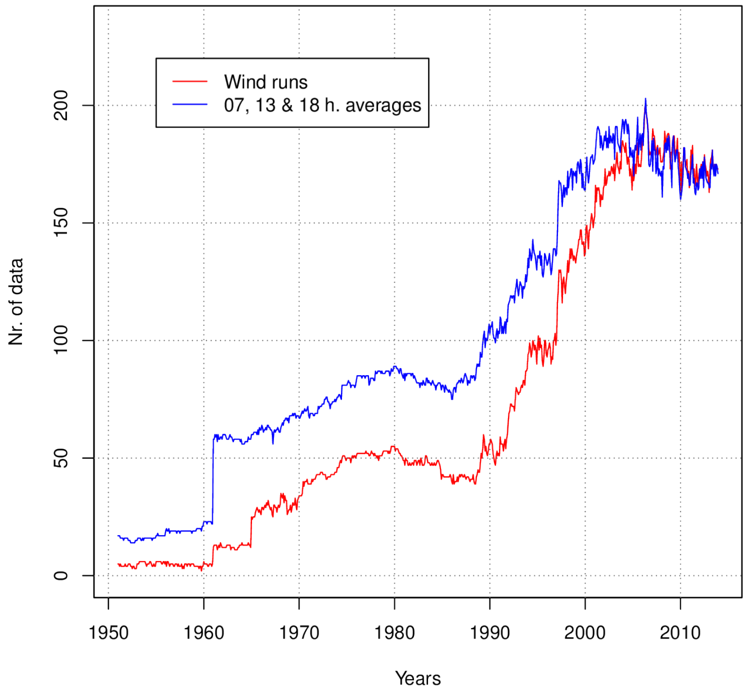

Figure 1. Number of average monthly wind speed data available from daily wind runs and from observations at 07, 13 and 18 hours UTC.

3. RESULTS AND DISCUSSION

3.1. Homogenization results

The first exploratory graphics yielded by Climatol show correlograms quickly decaying with distance, resulting in a spatially incoherent distribution of stations clustered according to their inter-correlations (Figure 2, upper row). This points to a high influence of topography and other features of the surroundings of the observatories on their wind measurements, precluding the use of nearby series as the better references. Wang (2008) already noticed that a reference wind speed series built by averaging neighboring stations gave worse results than another of geostrophic wind calculated from homogenized series of pressure. Yet the use of pressure gradients does not account for local thermal winds (see or valley breezes) that may contribute to a high portion of the average wind speed in regions with complex orography and coastal configuration.

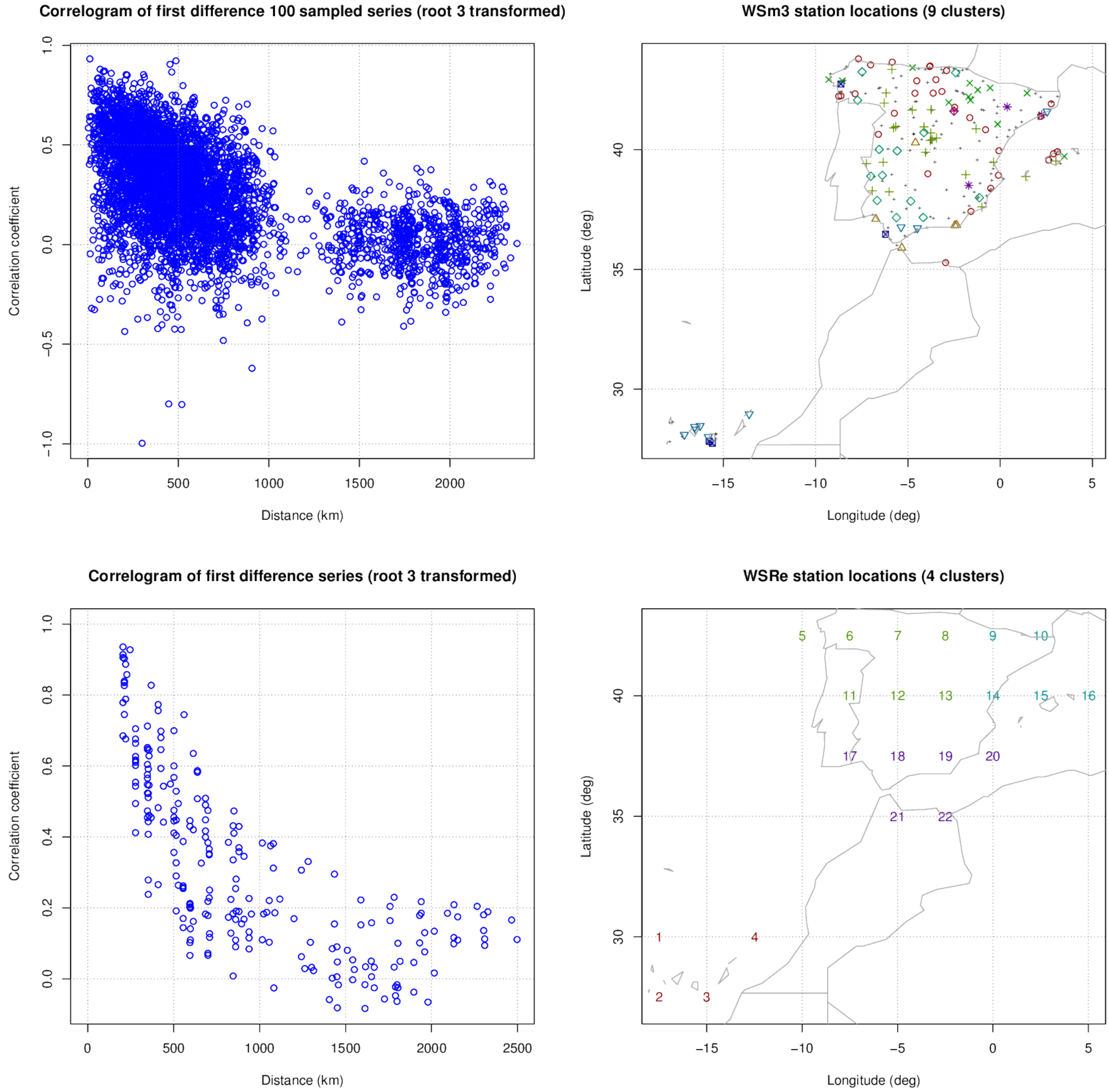

Figure 2. Correlograms (left) and spatial distribution of clustered stations (right) of observed (up) and reanalysis (down) wind speed series. Cluster analysis was limited to a maximum of 100 stations, the other 133 being represented by dots in the upper right map.

To account for this local wind circulations, outputs from mesoscale model simulations would be a better reference, as those from MM5 model used by Azorín-Molina *et al*. (2014), although its 10 km resolution is insufficient to capture most small scale thermal winds. Resolutions of 1 km or less would be needed to achieve a full picture of air circulation near the ground, but these simulations are very costly in computer requirements, hindering their use as references for the homogenization of long wind series. Therefore, reanalysis products are a more affordable source of reference series for wind homogenization studies, and the

NCEP series gathered here display a better spatial consistency than the observational series (Figure 2, lower row). However, their density is much lower than that of the observational series, and then a direct application of Climatol to a joint (observed plus analyzed) data-set would be using as references more nearby measured series than those more distant of the reanalysis. For these reason, homogenization has been applied separately to each dataset as a first approach in this work.

Results of the different homogenizations performed are summarized in Table 1, containing the number of corrected outliers and breaks, the percentage of inhomogeneous series, the mean RMSE of the data when computed from nearby stations and the mean SNHT of the homogenized series. Both ratio and standardization normalization types gave similar results in the wind run series, with slightly better (lower) values of RMSE and SNHT averages with the ratio normalization, more breaks and less outliers, making this the preferred normalization strategy for this variable. But this is not so clear in the series computed from three hourly observations (WSm3): Mean RMSE is also slightly lower with the ratio normalization (R), but the mean SNHT of the homogenized series is lower with the full standardization of cubic root transformed data (S3r). This is probably due to the higher number of breaks corrected, that could be explained by a lower noise in the series of cubic root transformed data. The number of outliers is also noticeable, more than doubling that of the ratio normalization. Around two thirds of the observational series appear as inhomogeneous, with one or more breaks corrected, while only about one third of the Spanish series analyzed by Azorín-Molina *et al*. (2014) were found inhomogeneous during 1961-2011.

No outlier was detected in the reanalysis series with any of the normalization types, but they are not free from shifts in the mean, with 31 and 36 breaks detected and corrected in the two homogenization processes. Moreover, as there are only 22 series coming from reanalysis, the percentage of inhomogeneous series is far higher than expected: 81.8% with the ratio normalization and 90.9% with the full standardization. Most of the breaks are detected in the second stage of the process, when SNHT is applied to the whole series of anomalies, since only 2 and 3 breaks are detected in the first stage respectively, with the stepped windows SNHT. A possible explanation, to be further investigated, is that the presumed lower noise of the reanalysis series allows the test to achieve significant values that would not be reached in more irregular observational series. As to RMSE and SNHT figures, the full standardization of cubic root values strategy yield better results in this case.

Table 1. Outliers and breaks corrected, percentage of inhomogeneous series, mean RMSE of the data when computed from nearby stations and mean SNHT of the resulting homogenized series, for the three data-sets WRun (wind speeds computed from wind daily runs), WSm3 (average wind speed measured three times per day) and WSRe (wind speed from reanalysis). Homogenizations were applied with two different settings: ratio normalization of original data (R) and full standardization of cubic root transformed data (S3r).

	Outliers	Breaks	% Inhom.	Mean RMSE	Mean SNHT
WRun (R)	71	268	64.4	0.38	8.30
WRun (S3r)	75	240	60.1	0.41	9.24
WSm3 (R)	38	360	66.5	0.46	10.64
WSm3 (S3r)	97	409	68.2	0.48	9.50
WSRe (R)	0	31	81.8	0.42	10.2
WSRe (S3r)	0	36	90.9	0.40	8.28

3.2. Trends of the homogenized wind speed series

Annual trends computed from the three homogenized monthly wind speed datasets and both methods of normalization are shown in Figure 3, displaying a majority of decreasing values between -0.02 and -2.50 m/s/century. Wind runs present less negative trends than wind observed three times per day, and the ratio normalization also yield less negative trends than the full standardization, which generates some very negative outliers. This fact makes the ratio normalization to be preferred to the standardization of cubic root transformed data, although both gave similar results in therms of RMSE and SNHT of the homogenized series.

On the other hand, reanalysis series have less negative trends than the observational datasets, backing the hypothesis of the influence of increasing surface roughness on the negative trends of wind speed series observed in many regions (Vautard *et al*., 2010; Wever, 2012).

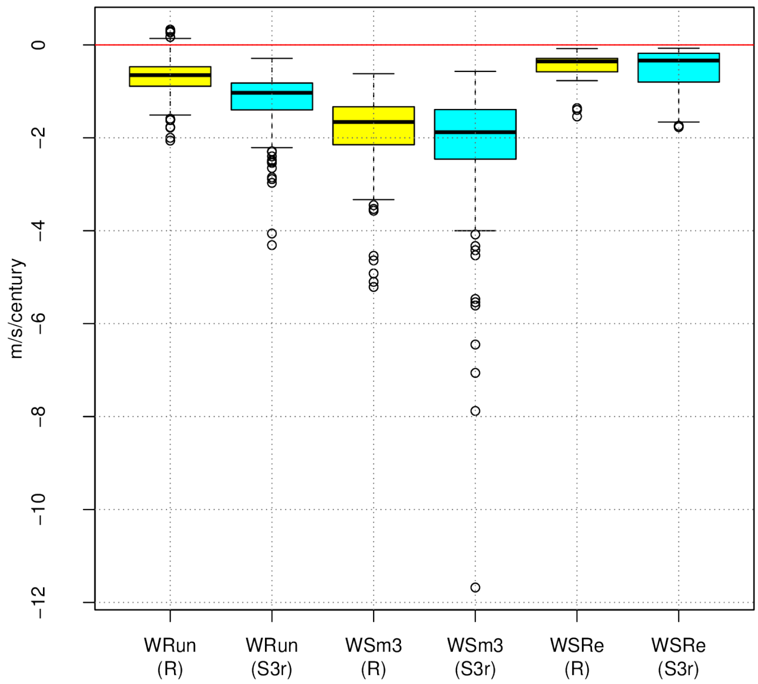

Fig. 3: Annual trends computed from the three homogenized monthly wind speed datasets and both methods of normalization.

Monthly trends of the three observations per day wind speed monthly averages are presented in Figure 4, showing the higher wind decreases of around -2 m/s/century from November-December until May, while in the warmer months, from June to October, trend values are near -1.5 m/s/century.

This seasonal distribution of trends is in accordance with Azorín-Molina *et al*. (2004) results, although their values were weaker and even positive in summer. But they used a lower number of stations (less than 50 from Spain), did not include the Canary islands, and the period of study was shorter (1961-2011).

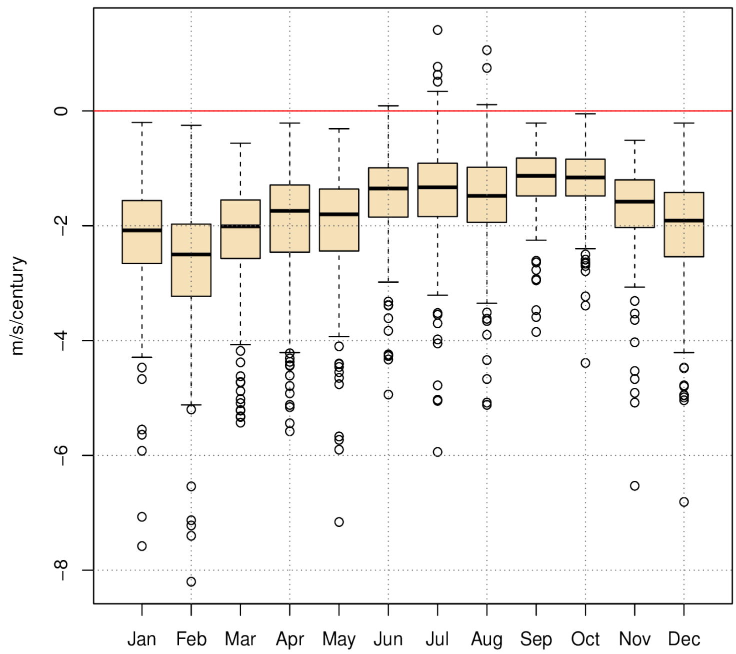

Fig. 4: Monthly trends of the homogenized (with ration normalization method) wind speed averages of three observations per day.

4. CONCLUSIONS AND FUTURE WORK

The Climatol package has allowed an easy homogenization of two datasets of 233 wind speed Spanish series with two different normalization methods.

Wind series appear to be very sensitive to changes and local influences, and are difficult to homogenize, especially in regions with complex orography and coastal configuration, because nearby stations may be poorly correlated.

Most wind speed trends are negative, especially in winter, with typical values between -1 and -2 m/s/century. Trends of reanalysis series are less negative than the observational series, pointing at a possible influence of an increasing roughness in the surroundings of the observatories.

Future work will be devoted to further investigating the benefits of using reanalysis products as a source of reference series to improve the homogenization of the wind speed climatological series, and also to study the geographical distribution of wind speed trends on land and sea, to ascertain the influence of roughness changes on the observed trends.

Acknowledgements

NCEP Reanalysis data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at http://www.esrl.noaa.gov/psd/

References

Alexandersson H (1986): A homogeneity test applied to precipitation data. Jour. of Climatol., 6, 661-675.

Azorín-Molina C, Vicente-Serrano SM, McVicar TR, Jerez S, Sánchez-Lorenzo A, López-Moreno JI, Revuelto J, Trigo RM, López-Bustins JA, Espirito-Santo F (2014): Homogenization and Assessment of Observed Near-Surface Wind Speed Trends over Spain and Portugal, 1961-2011. J. of Climate, 27:3692-3712.

Dadaser-Celik F, Cengiz E (2014): Wind speed trends over Turkey from 1975 to 2006. Int. J. Climatol., 34:1913-1927.

Guijarro JA (2014): User's Guide to Climatol. 40 pp., http://www.climatol.eu/climatol-guide.pdf

Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu W, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Leetmaa A, Reynolds J, Jenne R, Joseph D (1996): The NCEP/NCAR 40-year reanalysis project. Bull. Amer. Meteor. Soc., 77, 437-470.

Li Z, Yan Z, Tu K, Liu W, Wang Y (2011): Changes in wind speed and extremes in Beijing during 1960-2008 based on homogenized observations. Advances in Atmospheric Sciences, 28:408-420.

McVicar TR, Roderick ML, Donohue RJ, Li LT, Niel TGV, Thomas A, Grieser J, Jhajharia D, Himri Y, Mahowald NM, Mescherskaya AV, Kruger AC, Rehman S, Dinpashoh Y (2012): Global review and synthesis of trends in observed terrestrial near-surface wind speeds: Implications for evaporation. *J. Hydrol*., 416-417: 182-205.

Stepanek P (2004): *AnClim: Software for time series analysis and homogenization*. Department of Geography, Faculty of Natural Sciences, Masaryk University.

Vautard R, Cattiaux Yiou P, Thépaut JN, Ciais P (2010): Northern Hemisphere atmospheric stilling partly attributed to an increase in surface roughness. *Nat. Geosci*., 3, 756–761, doi:10.1038/ngeo979.

Wever N (2012): Quantifying trends in surface roughness and the effect on surface wind speed observations. *J. Geophys. Res.*, 117, D11104, doi:10.1029/2011JD017118.

Wan H, Wang XL, Swail VR (2010): Homogenization and Trend Analysis of Canadian Near-Surface Wind Speeds. *J. of Climate*, 23:1209-1225.

Wang XL, Feng Y (2007): RHtestV2 user manual. Climate Research Division, Science and Technology Branch, Environment Canada, 19 pp.

Wang XL (2008): Accounting for Autocorrelation in Detecting Mean Shifts in Climate Data Series Using the Penalized Maximal t or F Test. *Jour. Appl. Meteor. and Climatol.*, 47:2423-2444.