

Homogeneity of monthly wind speed time series in the Northeast of the Iberian Peninsula

Nuria **Casabella**, J. F. **Gonzalez-Rouco**, J. **Navarro**, A. **Hidalgo**, E. E. **Lucio-Eceiza**, J. L. **Conte**, E. **Aguilar**

Renewable Energy Dept., CIEMAT, Madrid, Spain

Faculty of Physical Sciences, Univ. Complutense of Madrid, Spain

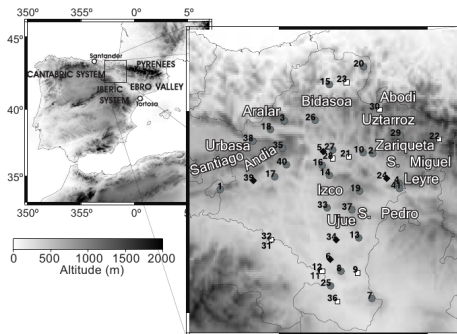
GLOBAL Forecasters, S.L., Madrid, Spain

Center for Climate Change, Univ. Rovira i Virgili, Tarragona, Spain

May, 14th 2014

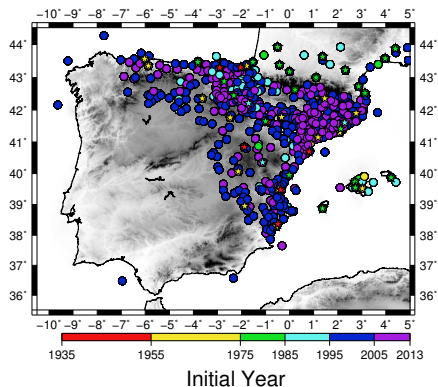


From 41 to 752 stations



Location of Navarra within the Iberian Peninsular (IP) and meteorological stations (Jimenez et al., *J. of Atm. and Oc. Technology*, 2010).

- 41 stations
- 13 years of data
- 10/30 min resolution

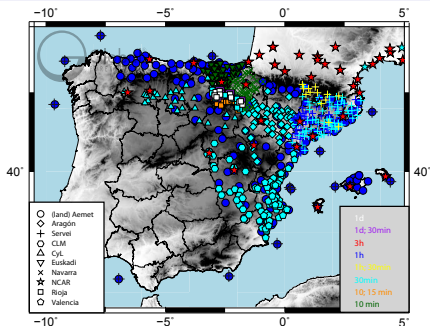


Location of the extended database and meteorological stations.

- 752 stations
- 77 years of data in some stations
- Different resolution times

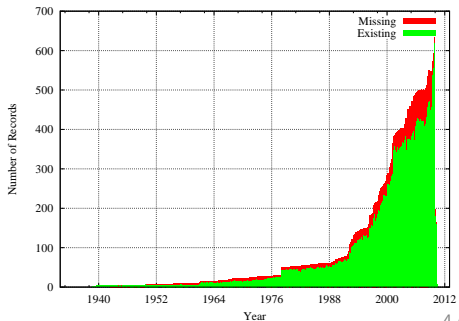
- Database overview
- Developed Quality Control methodology and examples
- Homogenisation process
- Future Work

Characteristics



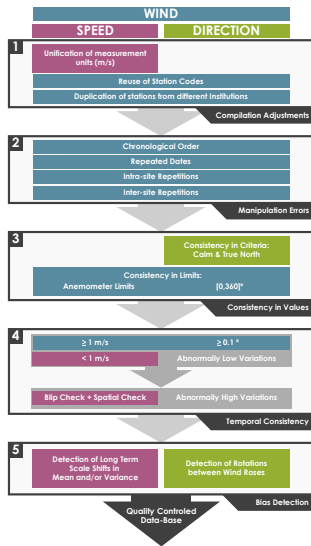
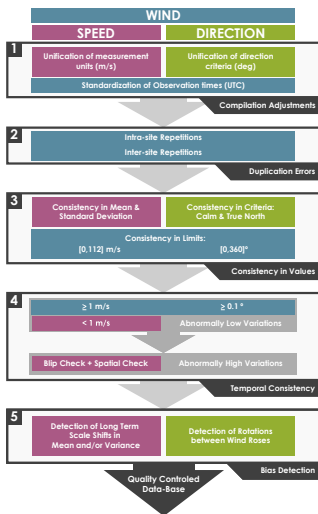
- 14 buoys distributed over the IP coast provided by Puertos del Estado (1990-2010)
- 738 additional land sites provided by 8 different Spanish institutions (1933-2010) and the National Center for Atmospheric Research (NCAR; 1978-2010)

Time evolution of the monthly number of existing and **above it** missing records in the whole dataset.



QC Diagram

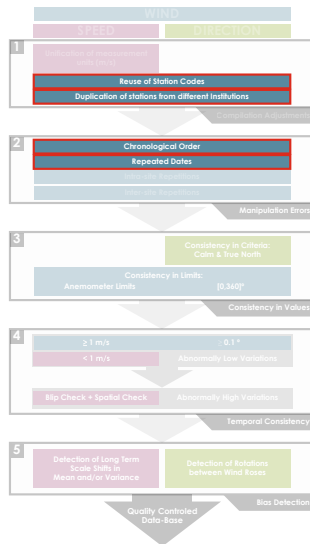
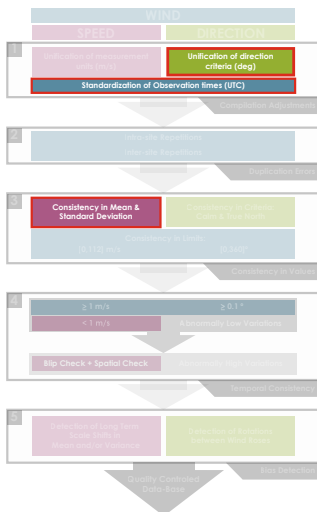
North Eastern North America QC:



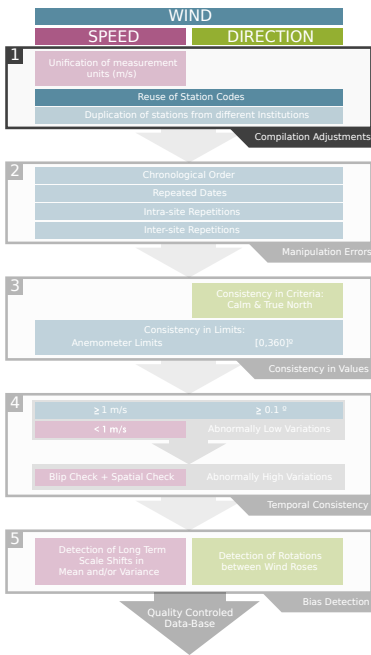
Quality Control of a Surface Wind Observations Database for North Eastern North America, E. E. Lucio-Eceiza

QC Diagram

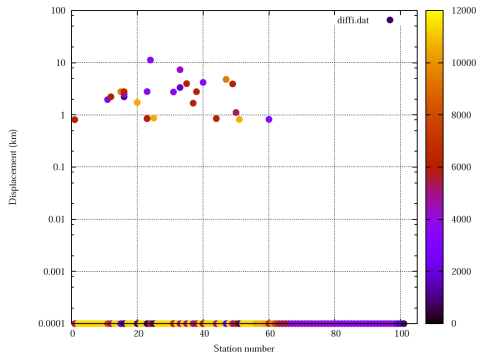
North Eastern North America QC:



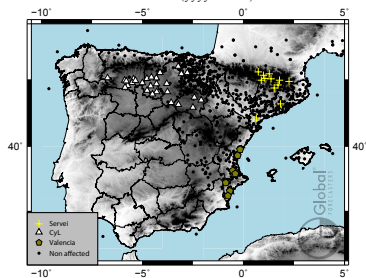
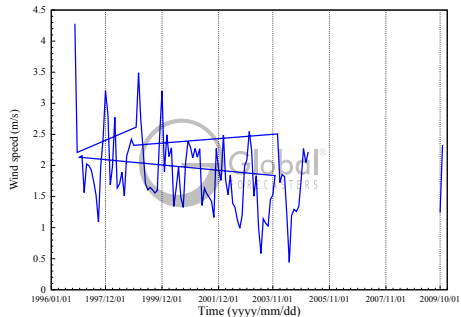
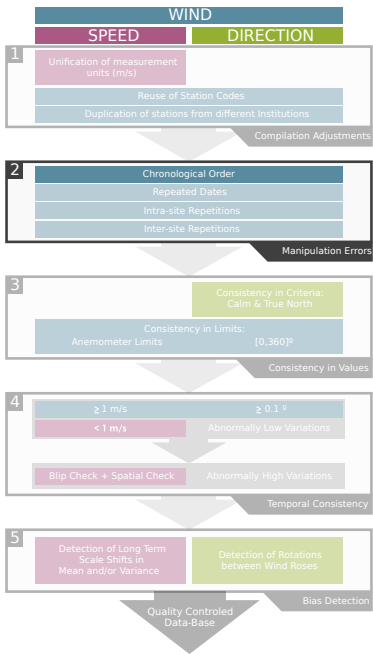
Quality Control of a Surface Wind Observations Database for North Eastern North America, E. E. Lucio-Eceiza



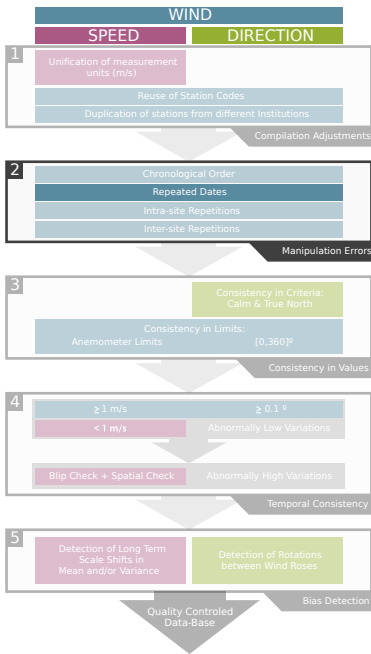
Displacements of the stations:



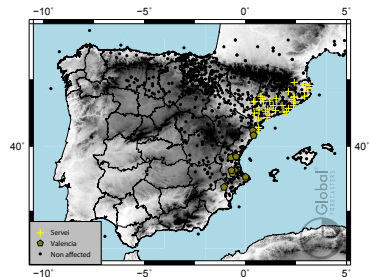
Possible displacements (km) of the stations, versus the station number and the number of days corresponding to certain values of latitude and longitude.



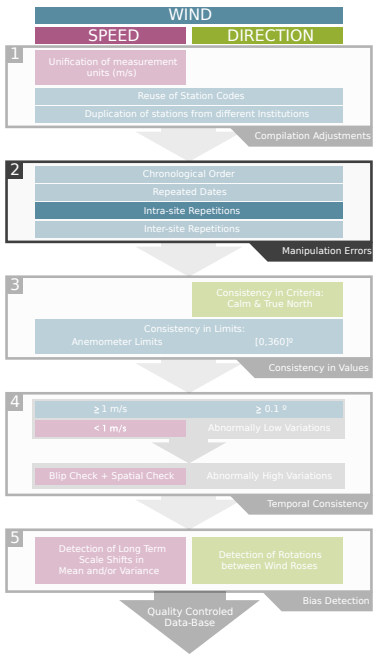
48 sites were affected by this manipulation error



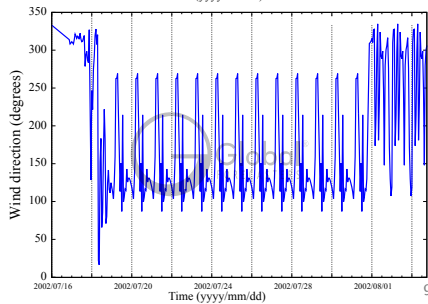
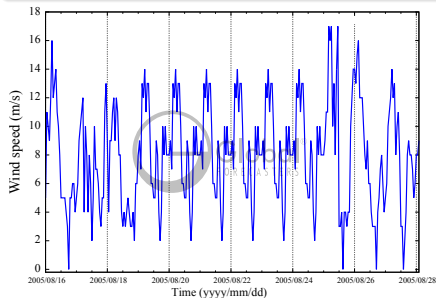
- Same wind variable values
- Different wind variable values

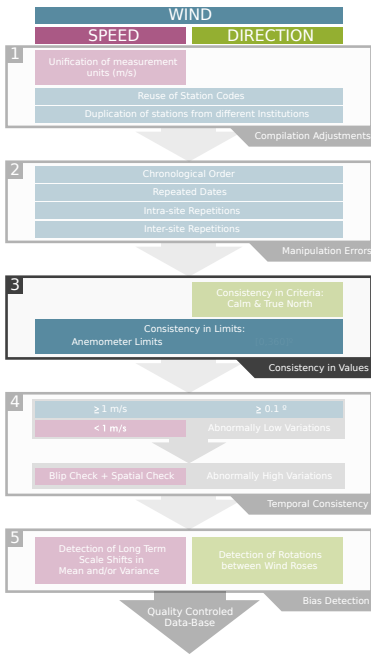


45 sites presented repeated dates with the same wind variables values

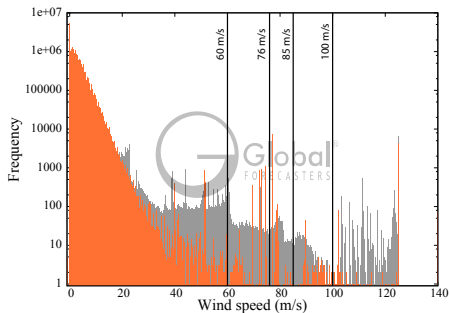


Repeated daily cycle on successive days



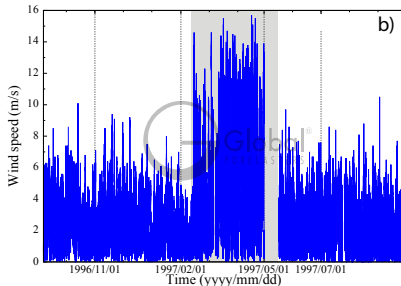
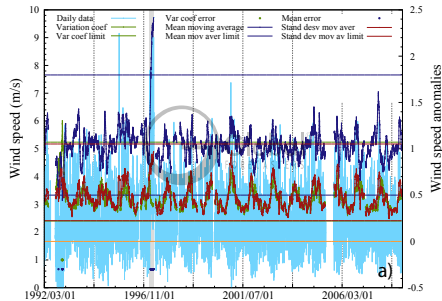
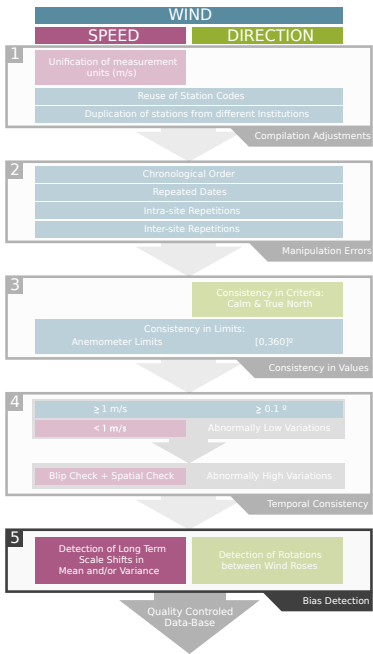


Wind speed histogram:

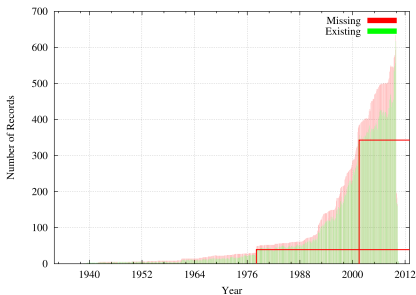


Whole database except 17 sites presenting an anomalous behavior.

The vertical lines correspond to the maximum recording values for the anemometers used by Navarra, Puertos del Estado, AEMET and the rest of institutions.



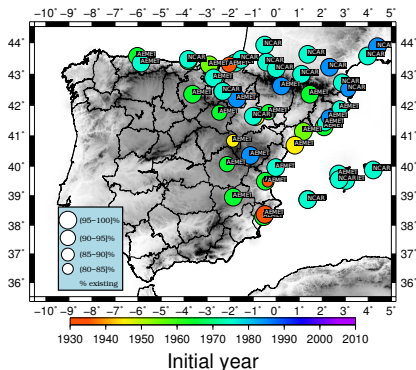
Homogeneity study



A subgroup of **monthly** wind speed time series spanning between 1978 and 2009 are subjected to the homogeneity study for fulfilling these criteria:

- several years of data
- lack of missing data
- being closest to the present

Homogeneity study



A subgroup of **monthly** wind speed time series spanning between 1978 and 2009 are subjected to the homogeneity study for fulfilling these criteria:

- several years of data
- lack of missing data
- being closest to the present

Software recommended by COST HOME (Mestre et al., 2013)

Software	References
HOMER	Mestre et al., 2013
Craddock	Craddock et al., 1979
MASH	Szentimrey, 2007
USHCN	Menne et al., 2009
ACMANT	Domonkos et al., 2011

SNHT (Alexandersson, 1986):

- Clear, methodologically simple
- Robust
- Good performance, thoroughly tested

Wind speed homogenisation work

Method/Software	Comparison	References
PMTred	Geostrophic Wind	Wan et al. 2010
t-test	Metadata	Keevallik et al., 2012
MASH	Multiple references	Lakatos et al., 2013
AnClim & ProClimDB	Regional reference series	Stepanek et al. 2013
AnClim	MM5	Azorin-Molina et al., 2014

Standard Normal Homogeneity Test

P: test series

Q: reference series

$$F_i = \frac{P_i}{\bar{P}}, \quad G_i = \frac{Q_i}{\bar{Q}}, \quad i = 1, \dots, n$$

$$q_i = \frac{F_i}{G_i}, \quad z_i = \frac{(q_i - \bar{q})}{S_q}; \quad N(0, 1)$$

Null hypothesis, (H0)

The test series is homogeneous. Any subset of z_i is distributed as $N(0, 1)$.

Alternative hypothesis, (H1)

The test series is inhomogeneous.

$$T_m = m\bar{z}_1^2 + (n - m)\bar{z}_2^2, \quad m = 1, \dots, n - 1$$

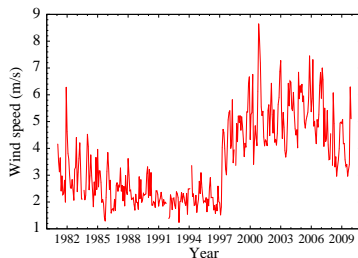
A high T value in year m implies that \bar{z}_1 and \bar{z}_2 depart significantly from zero.

Selecting the best reference series

- Short: 20 years of monthly data
- Potentially we expect a high number of inhomogeneities
- Different subregional wind variability dynamics

- **Normalized series through deseasonalization**
- **Detrend: Trend inhomogeneities won't be detected**
- **To select series for being similar in low frequency: Filter high frequency variability**

Pre-selection of the best reference series

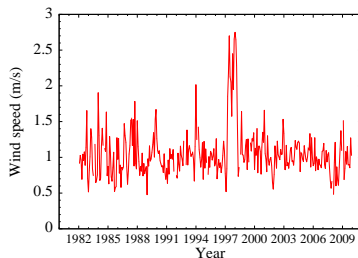


- 1 Make series of ratios:

$$d_i = \frac{y_i}{y_{i-12}}; \quad i = 1, n$$

- 2 Those values d_i exceeding the median ± 2 times the interquartile range are removed
 - 3 Construct d_i series of one year moving average
 - 4 Calculate the correlation between test series and each of the 49 reference series
- The best correlated stations will be pre-selected to assess the classification

Pre-selection of the best reference series

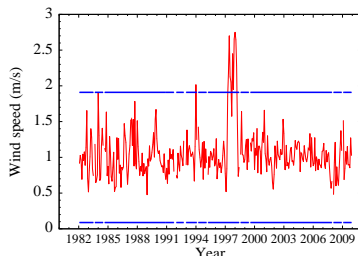


- 1 Make series of ratios:

$$d_i = \frac{y_i}{y_{i-12}}; i = 1, n$$

- 2 Those values d_i exceeding the median ± 2 times the interquartile range are removed
- 3 Construct d_i series of one year moving average
- 4 Calculate the correlation between test series and each of the 49 reference series
The best correlated stations will be pre-selected to assess the classification

Pre-selection of the best reference series

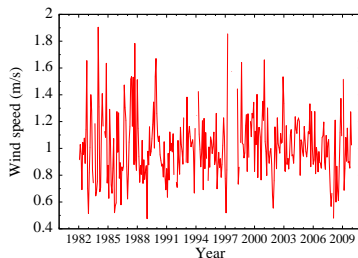


- 1 Make series of ratios:

$$d_i = \frac{y_i}{y_{i-12}}; \quad i = 1, n$$

- 2 Those values d_i exceeding the median ± 2 times the interquartile range are removed
 - 3 Construct d_i series of one year moving average
 - 4 Calculate the correlation between test series and each of the 49 reference series
- The best correlated stations will be pre-selected to assess the classification

Pre-selection of the best reference series

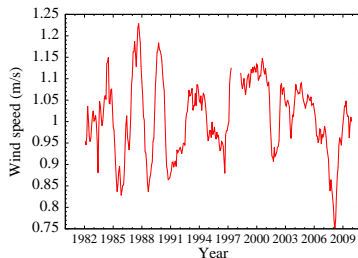


- 1 Make series of ratios:

$$d_i = \frac{y_i}{y_{i-12}}; i = 1, n$$

- 2 Those values d_i exceeding the median ± 2 times the interquartile range are removed
 - 3 Construct d_i series of one year moving average
 - 4 Calculate the correlation between test series and each of the 49 reference series
- The best correlated stations will be pre-selected to assess the classification

Pre-selection of the best reference series

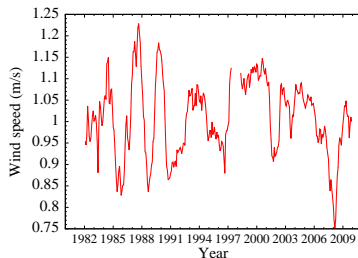


- 1 Make series of ratios:

$$d_i = \frac{y_i}{y_{i-12}}; \quad i = 1, n$$

- 2 Those values d_i exceeding the median ± 2 times the interquartile range are removed
- 3 Construct d_i series of one year moving average
- 4 Calculate the correlation between test series and each of the 49 reference series
The best correlated stations will be pre-selected to assess the classification

Pre-selection of the best reference series



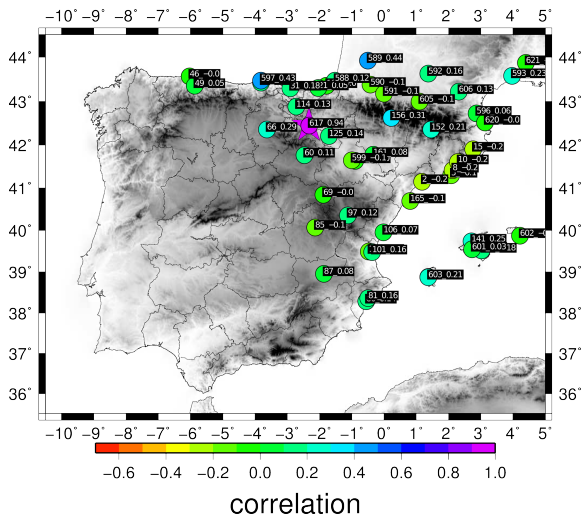
- 1 Make series of ratios:

$$d_i = \frac{y_i}{y_{i-12}}; \quad i = 1, n$$

- 2 Those values d_i exceeding the median ± 2 times the interquartile range are removed
 - 3 Construct d_i series of one year moving average
 - 4 Calculate the correlation between test series and each of the 49 reference series
- The best correlated stations will be pre-selected to assess the classification

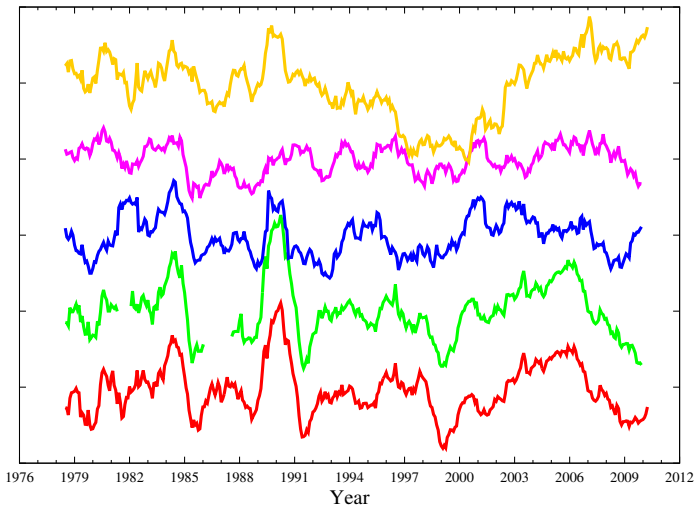
Example: Homogeneous Series

Correlation among ratio, truncated and one year moving average time series:

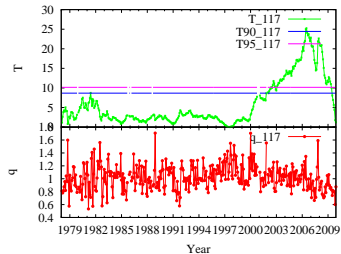
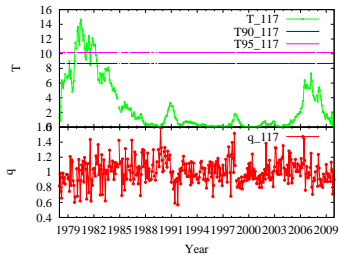
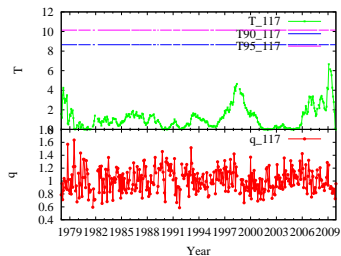
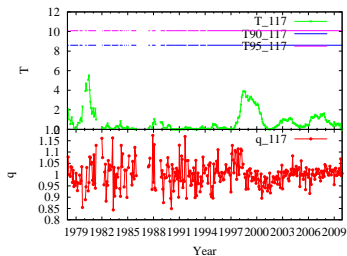


Example: Homogeneous Series

Selected original series once they are deseasonalized and detrended, with a one year moving average:

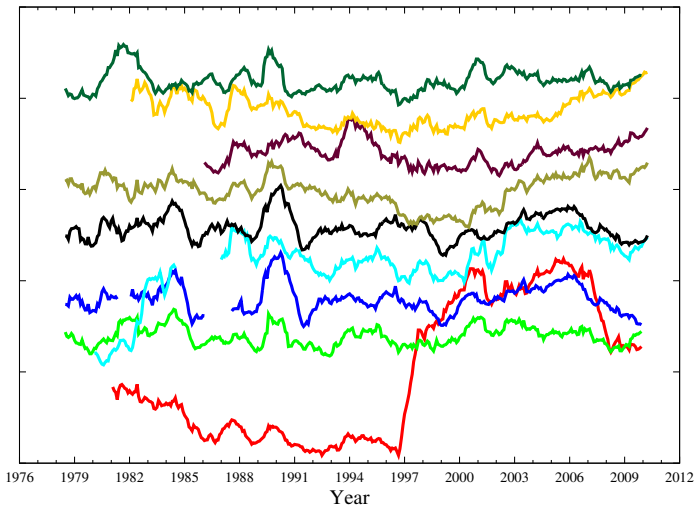


Example: Homogeneous Series

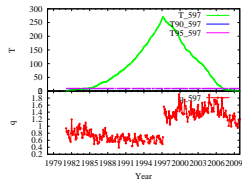
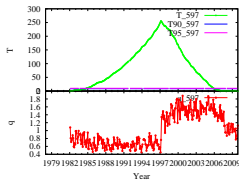
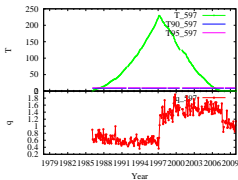
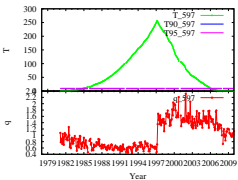
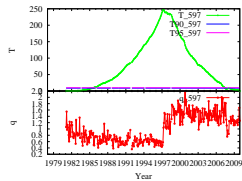
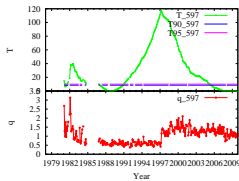
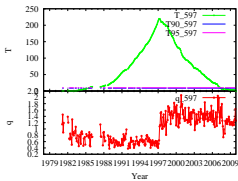
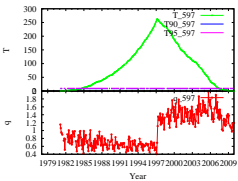


Example: Inhomogeneous Series

Selected original series once they are deseasonalized and detrended, with a one year moving average:

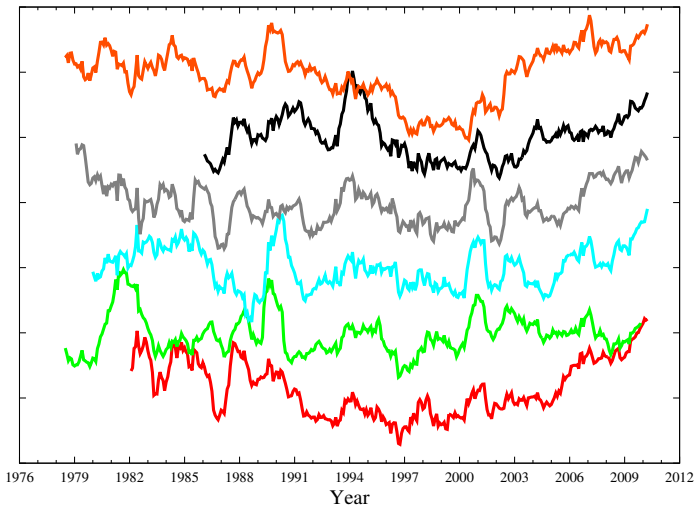


Example: Inhomogeneous Series

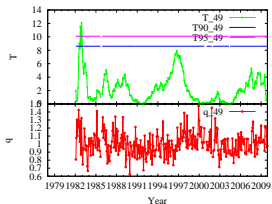
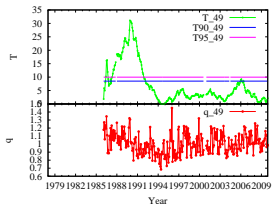
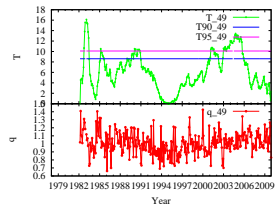
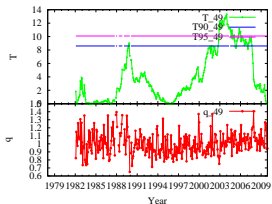
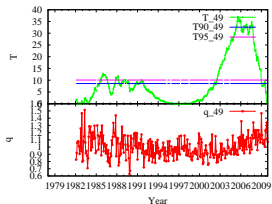


Example: Unclassified Series

Selected original series once they are deseasonalized and detrended, with a one year moving average:



Example: Unclassified Series

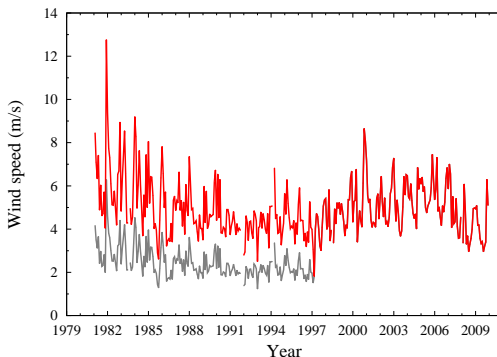


Correction method (Gonzalez-Rouco et al., *J. of Climate*, 2001)

The data **before** the inhomogeneity date are corrected through multiplying them by:

$$f = \frac{q_a}{q_b}$$

q_a and q_b : mean values of q_i after and before the inhomogeneity



Next Steps

- 1 Make an algorithm that reproduces well these conditions for classification and follow an iterative procedure.
- 2 Extend the homogeneity procedure to other stations in the database.
- 3 Analyze the spatial and temporal occurrence of inhomogeneities.
- 4 Analyze the impacts of the inhomogeneities in the long-term wind speed trends.
- 5 Homogenize the wind direction.

Thank you! Questions/Suggestions...?

Homogeneity of monthly wind speed time series in the Northeast of the Iberian Peninsula

Nuria **Casabella**, J. F. **Gonzalez-Rouco**, J. **Navarro**, A. **Hidalgo**, E. E. **Lucio-Eceiza**, J. L. **Conte**, E. **Aguilar**

Renewable Energy Dept., CIEMAT, Madrid, Spain

Faculty of Physical Sciences, Univ. Complutense of Madrid, Spain

GLOBAL Forecasters, S.L., Madrid, Spain

Center for Climate Change, Univ. Rovira i Virgili, Tarragona, Spain

May, 14th 2014

