Quality Control of a surface wind observations database for North Eastern North America

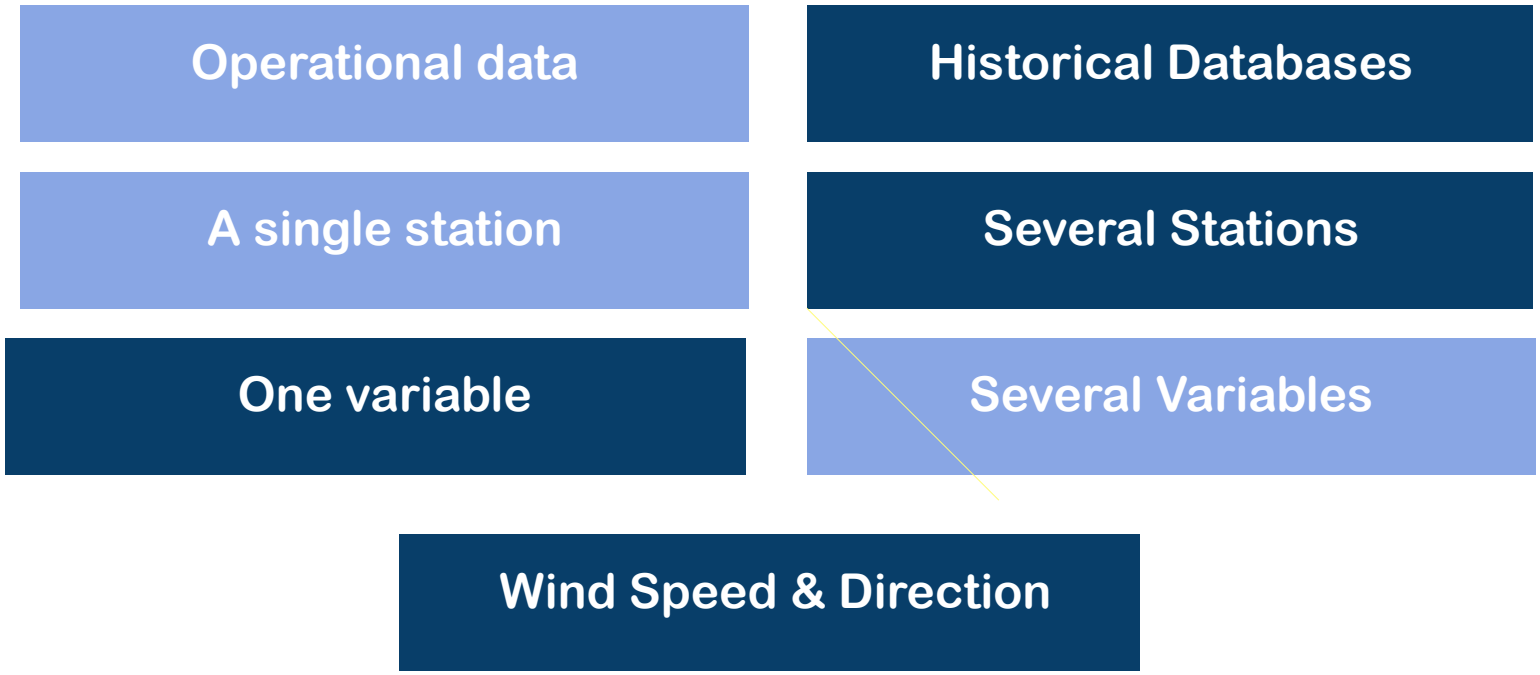
E.E. Lucio-Eceiza	**UCM, IGEO (eelucio@fis.ucm.es)**
J.F. González-Rouco	**UCM, IGEO**
J. Navarro	**CIEMAT**
A. Hidalgo	**Global Forecasters S.L.**
P.A. Jiménez	**CIEMAT**
E. García-Bustamante	**U. Murcia**
N. Casabella	**CIEMAT**
J.L. Conte	**Global Forecasters S.L.**
H. Beltrami	**CASI (St. Francis Xavier U.)**

Paleoclimate Modeling & Analysis (PalMA)
Dpto. Física de la Tierra, Astronomía y Astrofísica II
Universidad Complutense de Madrid
Instituto de Geociencias CSIC-UCM

8th Seminar for Homogenization and
3rd Conference on Spatial Interpolation

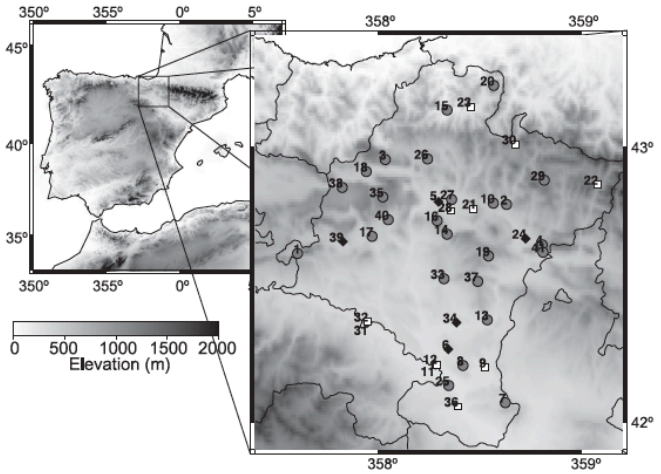
12 - 16 May 2014
Budapest, Hungary

Quality Control Procedures, suited for different purposes:



DeGaetano, A. T., 1997: *A quality-control routine for hourly wind observations. Journal of Atmospheric and Oceanic Technology*, **14**, 308–317.

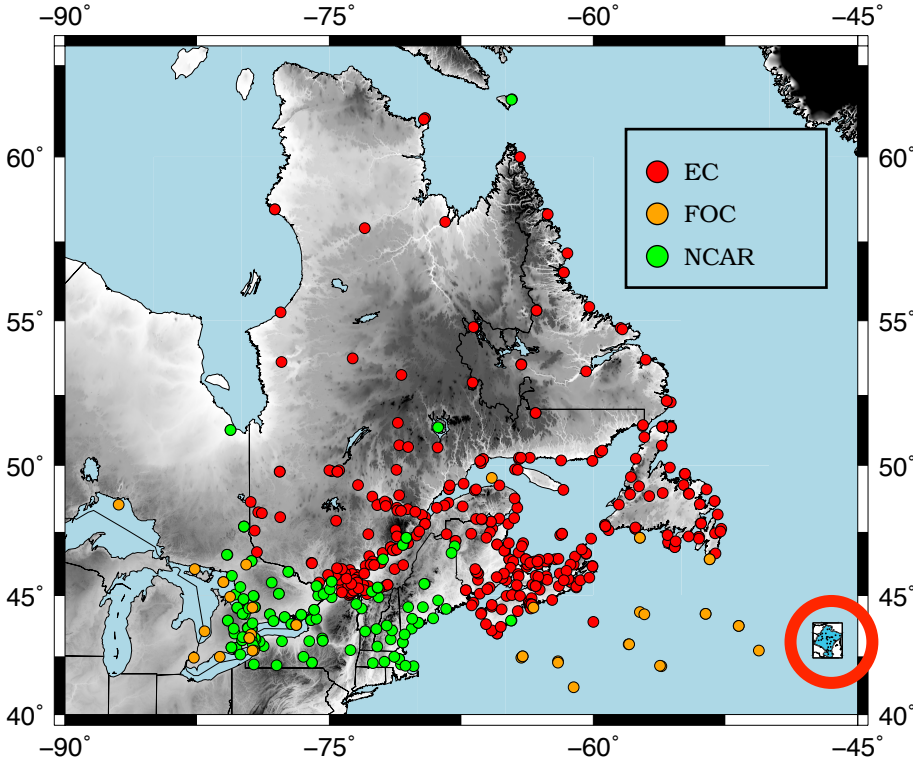
Jimenez, P., J. Gonzalez-Rouco, J. Navarro, J. Montavez, and E. García-Bustamante, 2010: Quality assurance of surface wind observations from automated weather stations. *Journal of Atmospheric and Oceanic Technology*, **27**, 1101-1122.



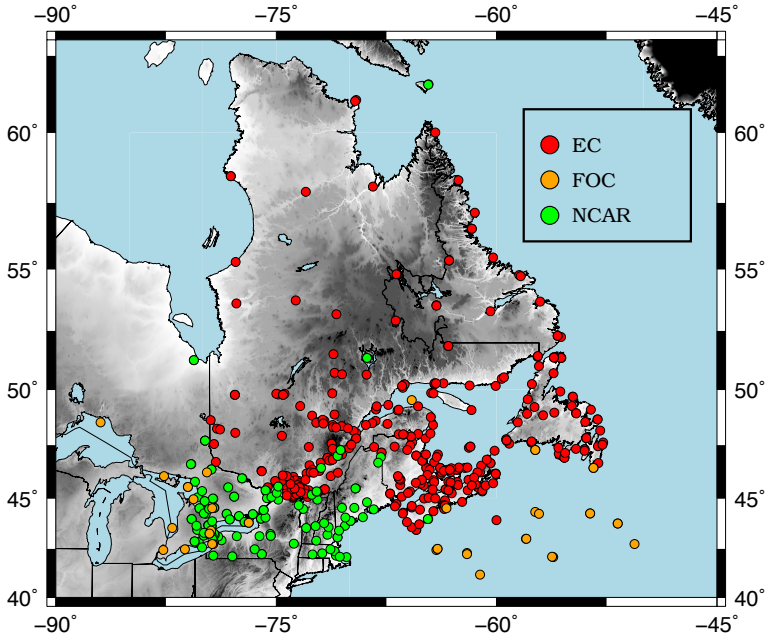
- 41 automatic stations
- 13 years of data
- 10min/30min resolution

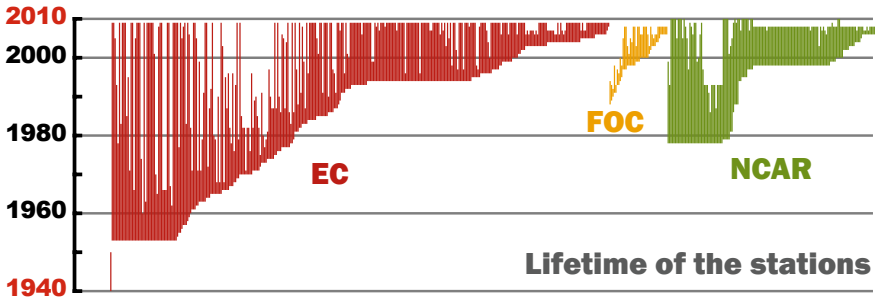
Jiménez et al., JOAT, 27, 2010

Location of Navarra and spatial distribution of the 41 stations (Jimenez et al. 2010).



Size comparison between Navarra (Jimenez et al. 2010, circled) and the area of North Eastern North America.



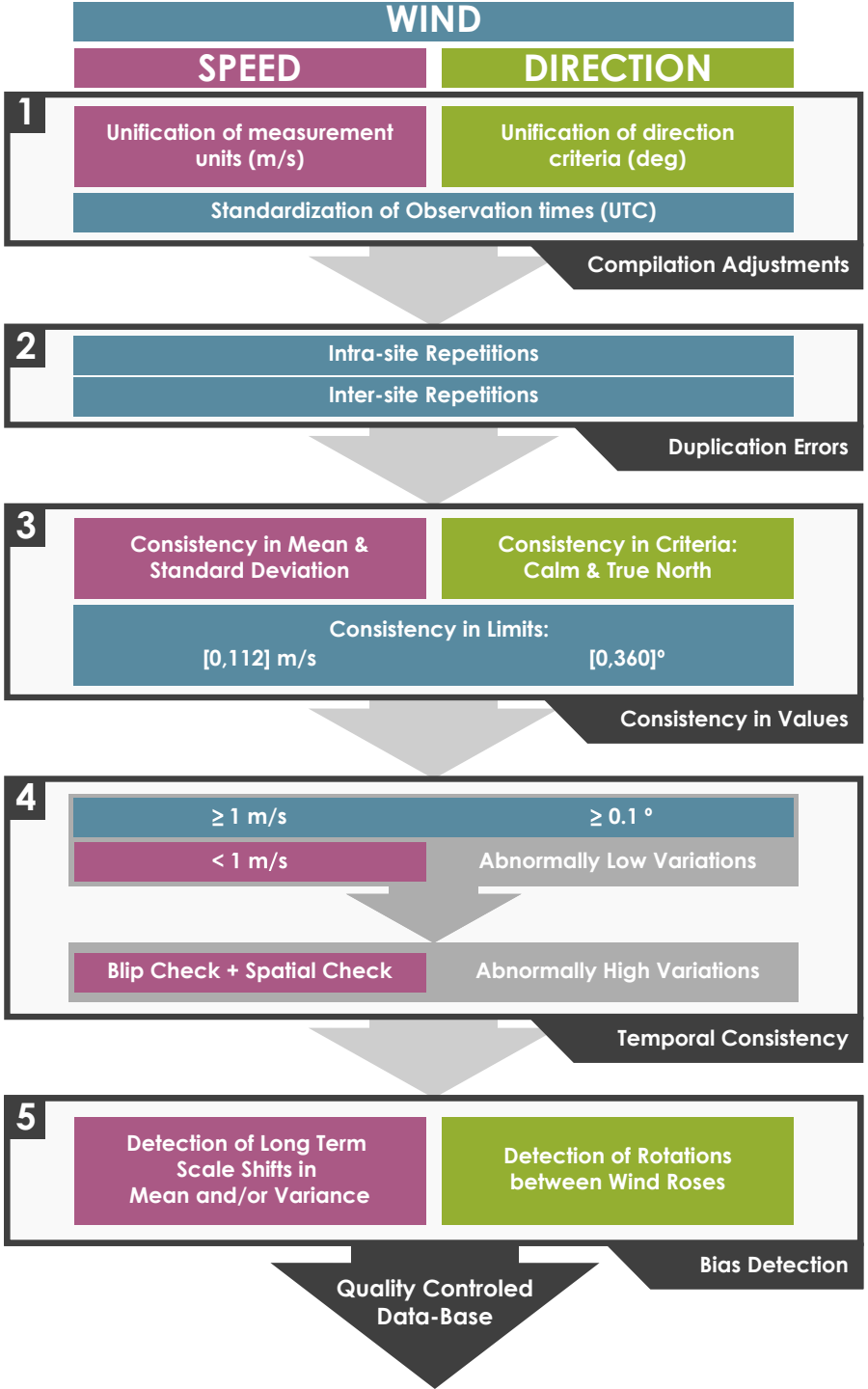


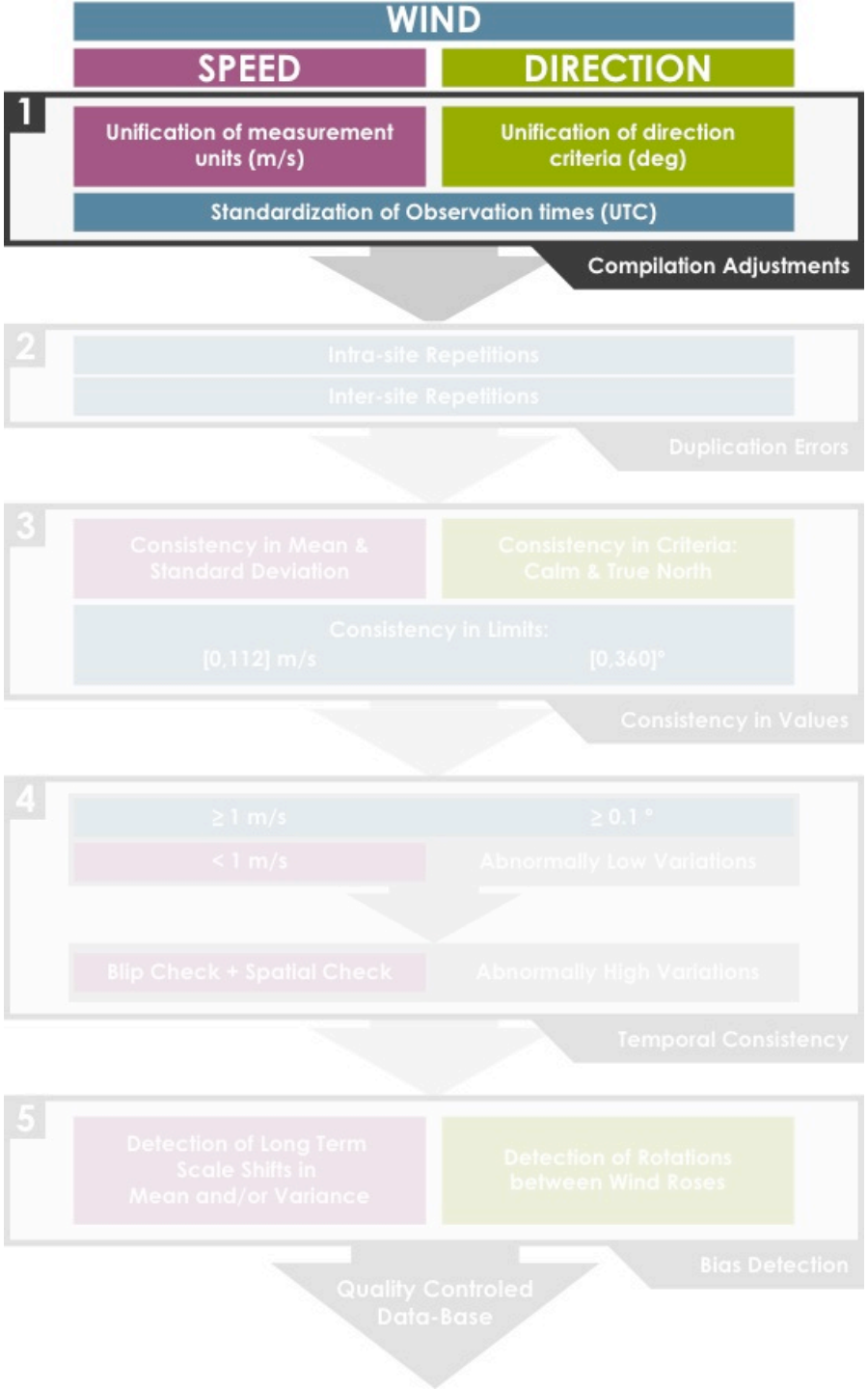
The vertical lines correspond to the lifetime of each station.

The database:

- 527 stations: 344 from Environment Canada (red), 40 Fisheries and Oceans (orange) and 143 from NCAR (green).
- Time resolution: hourly, 3hourly and synoptic, sometimes within the same station.
- Time span: 1940 - 2010.
- Over 54×10^6 data pair values.

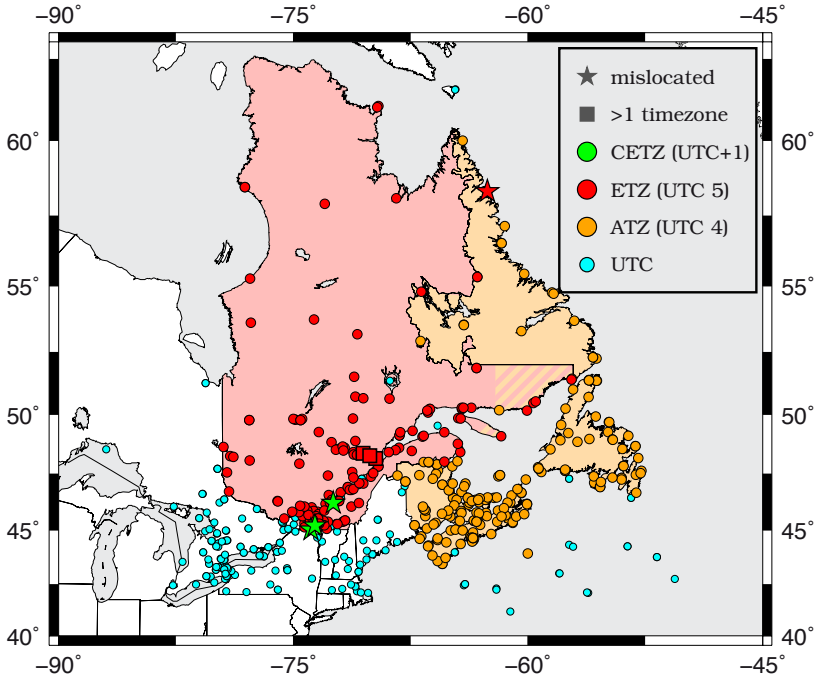
Motivation | Database Compilation | Quality Assurance



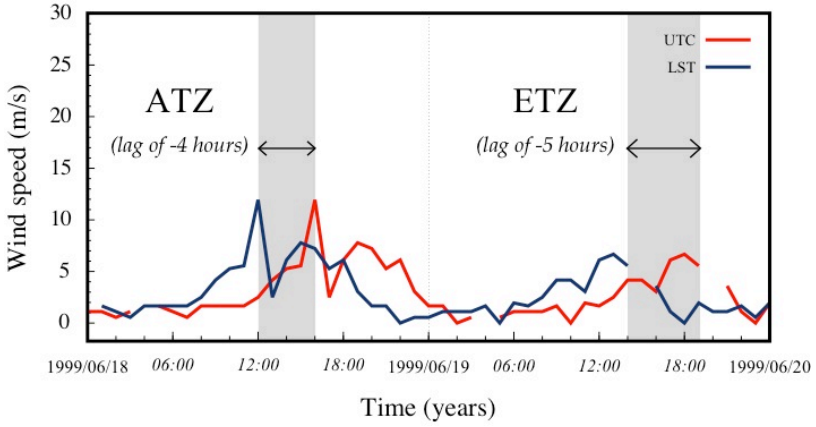

WIND
SPEED
DIRECTION
1
Unification of measurement units (m/s)
Unification of direction criteria (deg)
Standardization of Observation times (UTC)
Compilation Adjustments
2
Intra-site Repetitions
Inter-site Repetitions
Duplication Errors
3
Consistency in Mean & Standard Deviation
Consistency in Criteria: Calm & True North
Consistency in Limits:
[0,112] m/s
[0,360]°
Consistency in Values
4
≥ 1 m/s
≥ 0.1 °
< 1 m/s
Abnormally Low Variations
Blip Check + Spatial Check
Abnormally High Variations
Temporal Consistency
5
Detection of Long Term Scale Shifts in Mean and/or Variance
Detection of Rotations between Wind Roses
Bias Detection
Quality Controled Data-Base

Standardization of Observation Times (UTC)

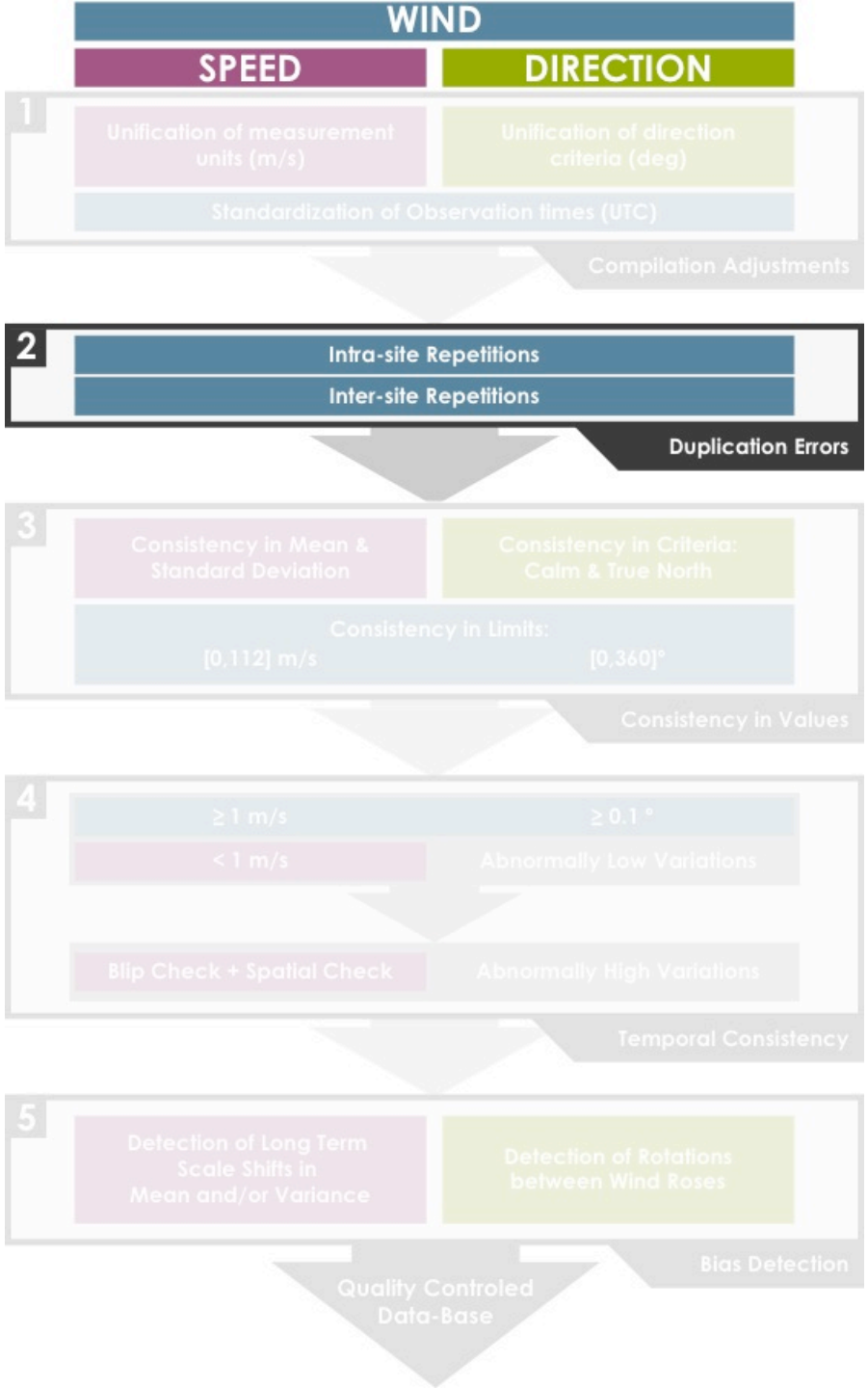
- FOC & NCAR record in UTC.
- EC does it in Local Standard Time (LST).
- All stations are transformed to UTC.



- Spatial distribution of the time zones that the stations belong to.
- Most follow their geographical timezones.
- Three are located in Central European Time Zone (CETZ) despite belonging to Quebec.

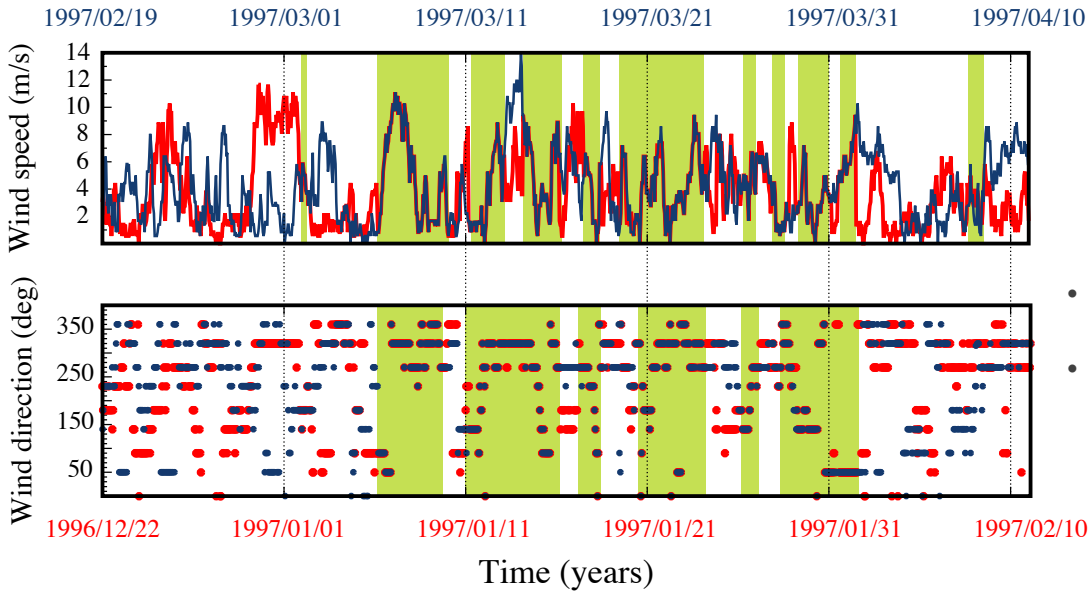


Example of a station that moves from being recorded in ATZ to ETZ.


WIND
SPEED
DIRECTION
1
Unification of measurement units (m/s)
Unification of direction criteria (deg)
Standardization of Observation times (UTC)
Compilation Adjustments
2
Intra-site Repetitions
Inter-site Repetitions
Duplication Errors
3
Consistency in Mean & Standard Deviation
Consistency in Criteria: Calm & True North
Consistency in Limits:
[0,112] m/s
[0,360]°
Consistency in Values
4
≥ 1 m/s
≥ 0.1 °
< 1 m/s
Abnormally Low Variations
Blip Check + Spatial Check
Abnormally High Variations
Temporal Consistency
5
Detection of Long Term Scale Shifts in Mean and/or Variance
Detection of Rotations between Wind Roses
Bias Detection
Quality Controled Data-Base

Intra-site repetitions:

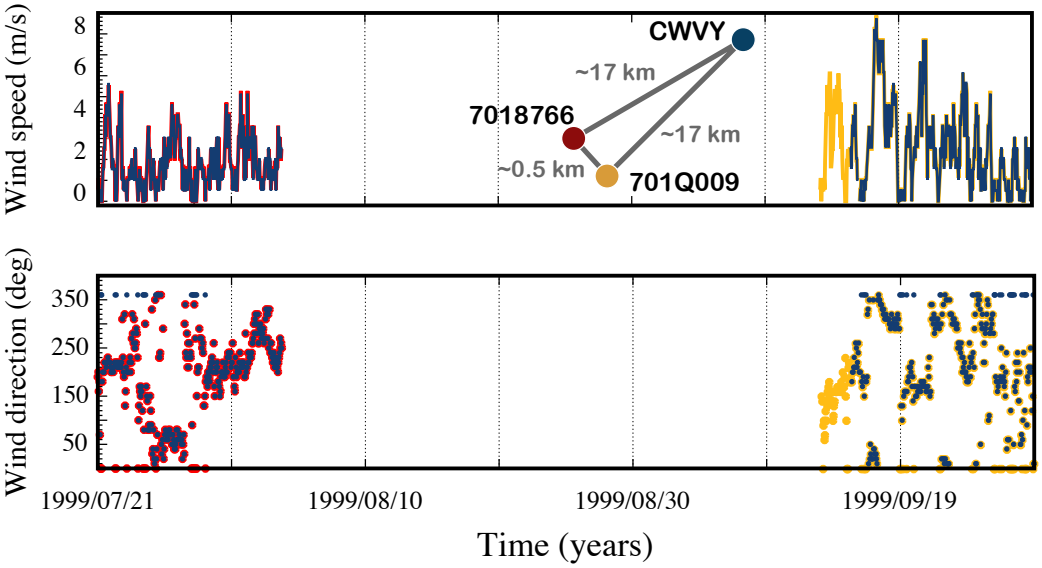
Duplications of data periods within the same series



- Example of a station with ~1 month of duplicated data.
- The month of March (blue) is overwritten over January (red) for both wind speed & direction.

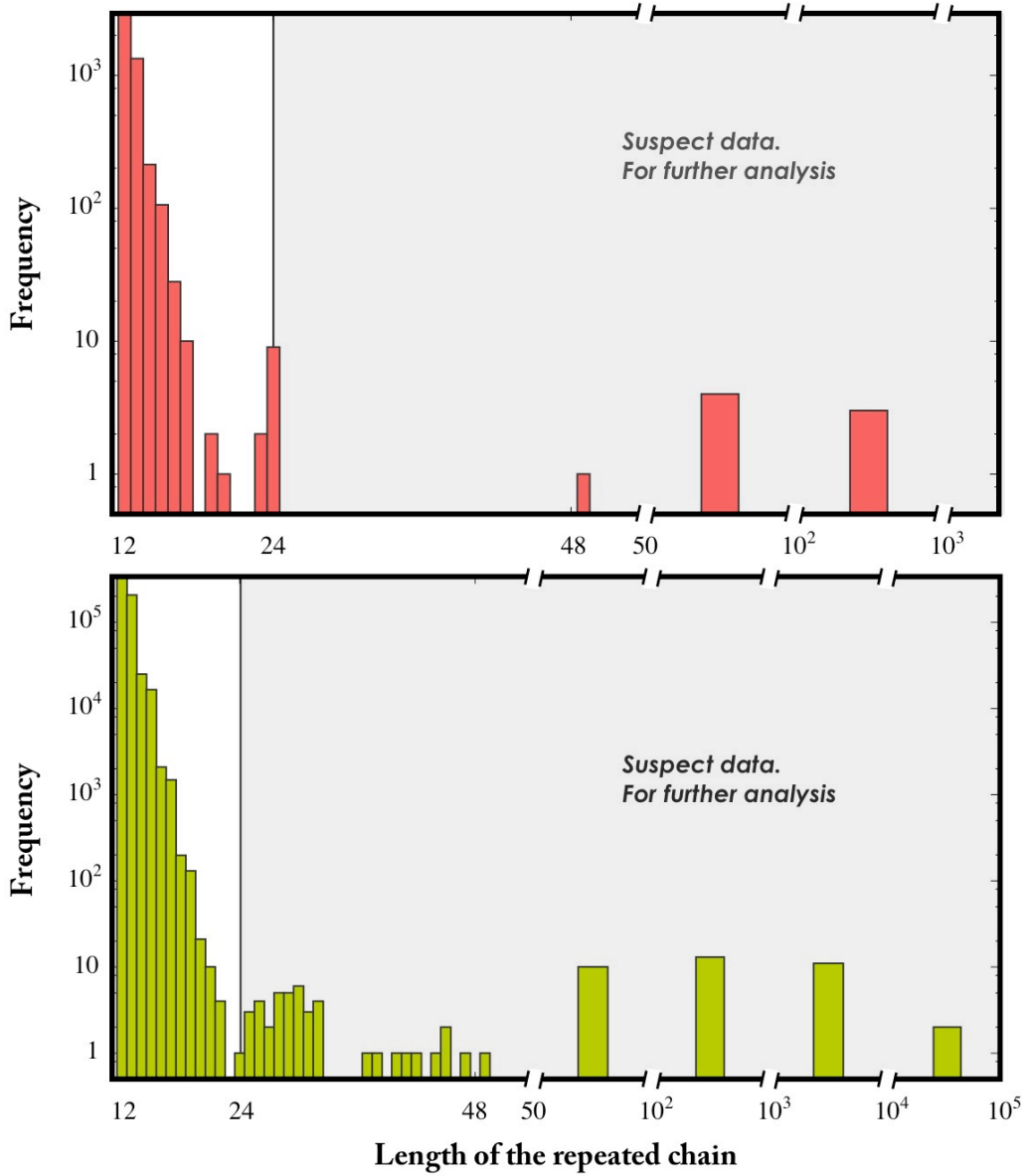
Inter-site repetitions:

Data transfer from one serie to another.



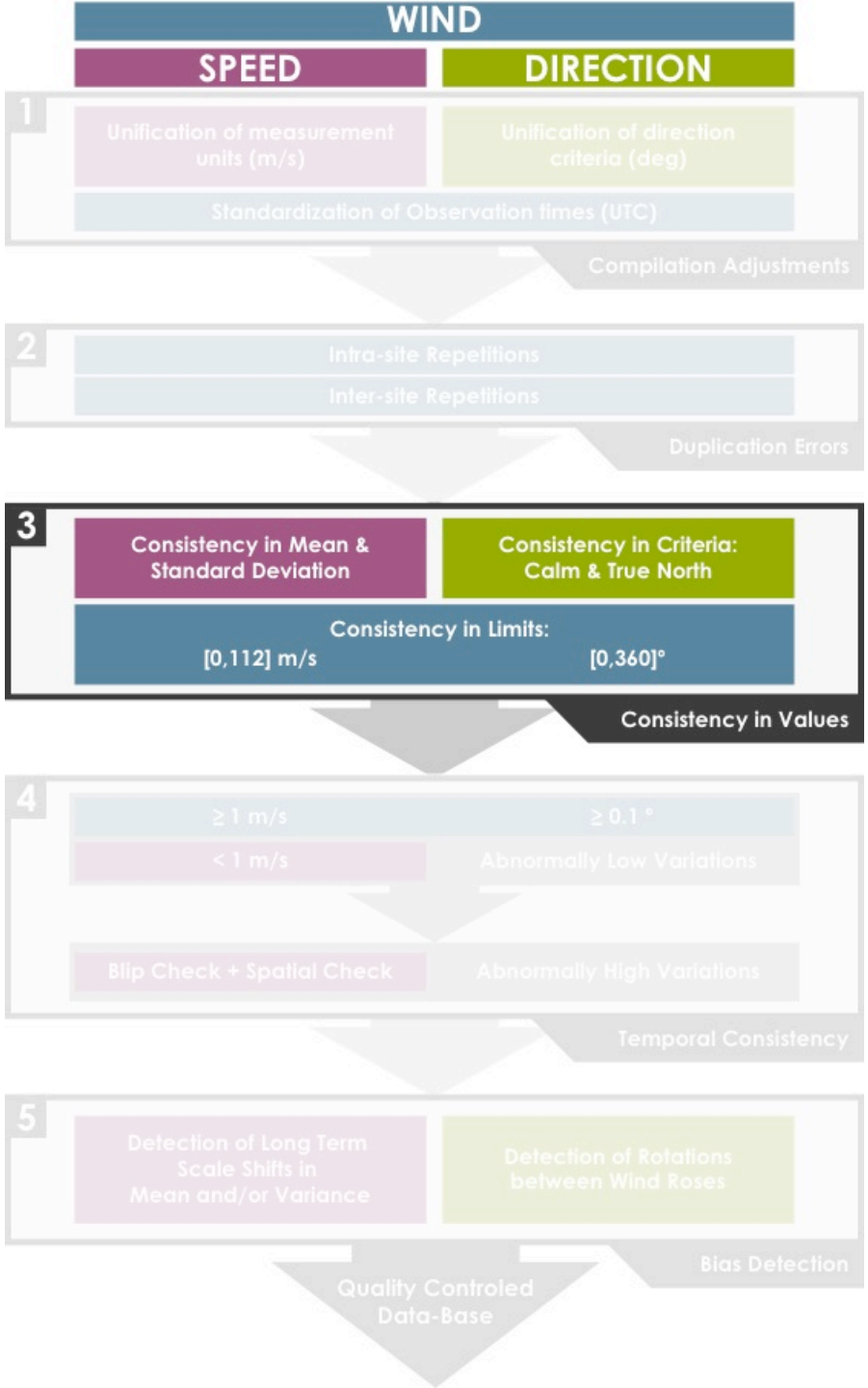
- Example of a station (in blue) that shares data with two others (red and orange) for both wind speed & direction.
- A total of ~7 years of duplicated data were found.

- Both duplication cases are analyzed separately but in a similar manner.
- Once the repeated chains are identified, we distribute them according to their length.
- The absolute frequency of the chains diminishes with their length.
- A threshold is set when the distribution tends to zero.



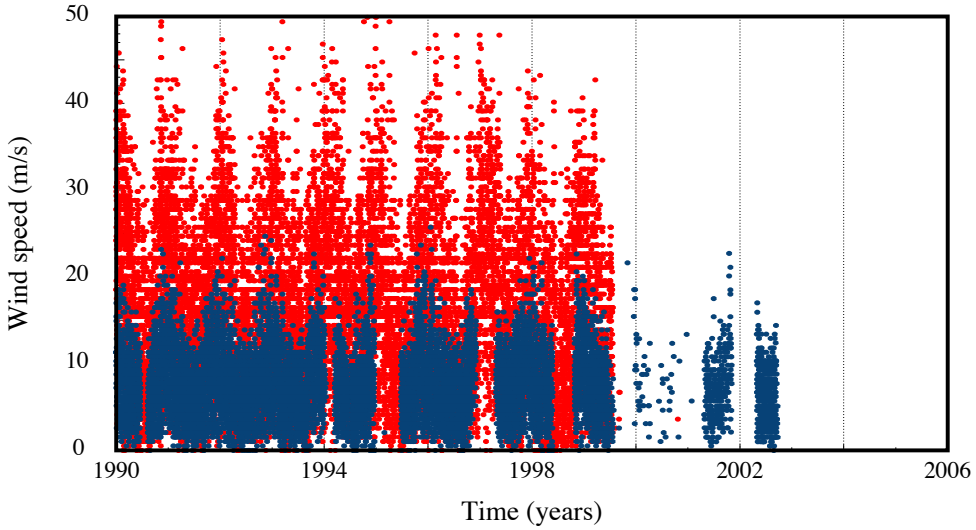
Intra-site repetitions.
Case of wind speed.

Inter-site repetitions.
Case of wind direction.


WIND
SPEED
DIRECTION
1
Unification of measurement units (m/s)
Unification of direction criteria (deg)
Standardization of Observation times (UTC)
Compilation Adjustments
2
Intra-site Repetitions
Inter-site Repetitions
Duplication Errors
3
Consistency in Mean & Standard Deviation
Consistency in Criteria: Calm & True North
Consistency in Limits:
[0,112] m/s
[0,360]°
Consistency in Values
4
≥ 1 m/s
≥ 0.1 °
< 1 m/s
Abnormally Low Variations
Blip Check + Spatial Check
Abnormally High Variations
Temporal Consistency
5
Detection of Long Term Scale Shifts in Mean and/or Variance
Detection of Rotations between Wind Roses
Bias Detection
Quality Controled Data-Base

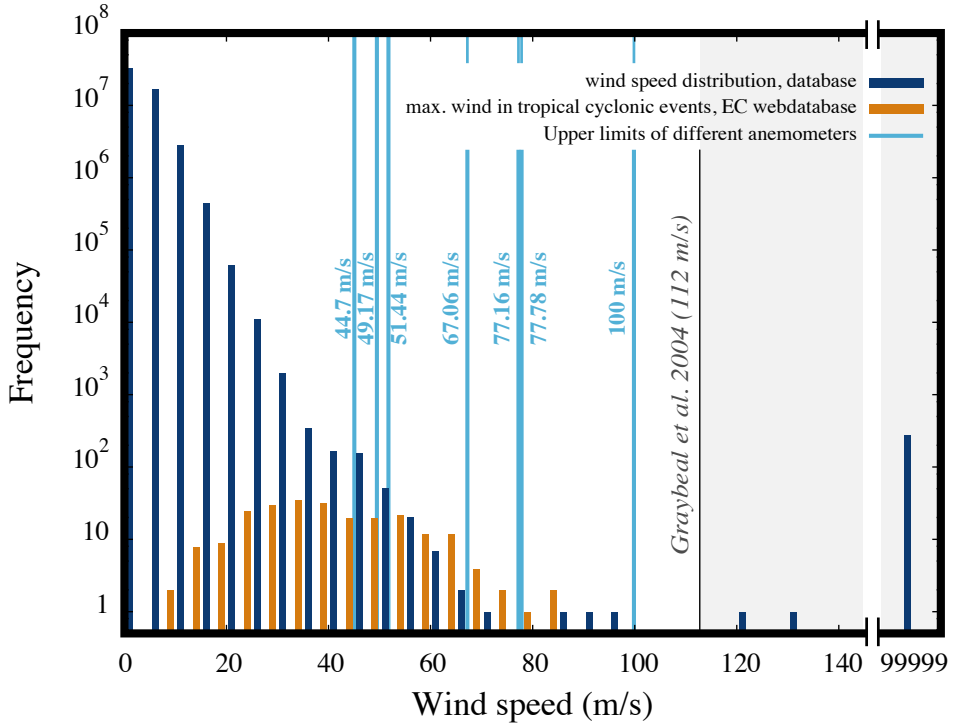
Consistency in Mean & Std. Dev.

It identifies whole stations with values that are clearly unrealistic.



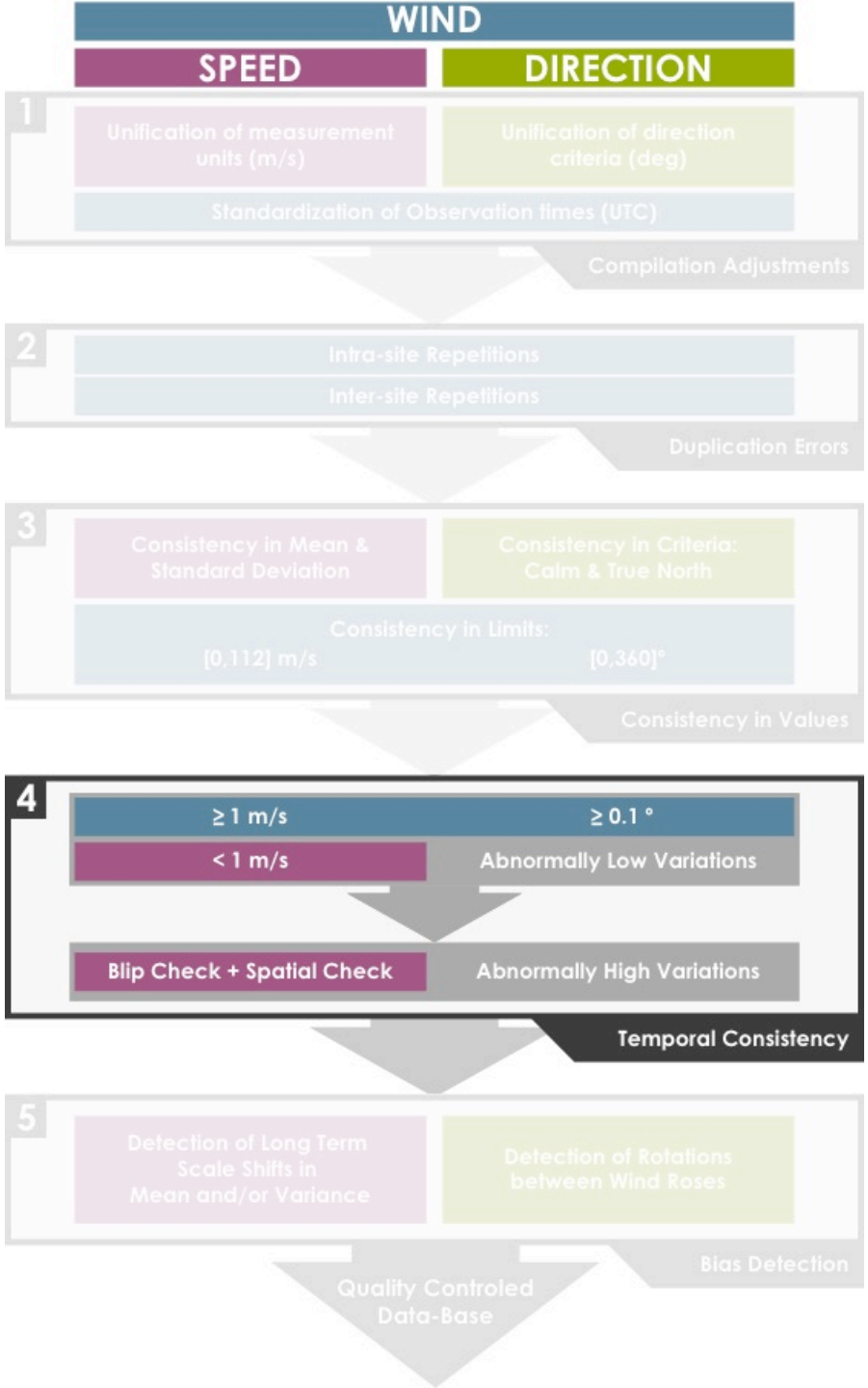
Comparison between an erroneous station and good one:

- Station in red: mean of 16 m/s
- Station in bue: mean of 7.6 m/s



Consistency in limits. Wind speed: [0,112] m/s

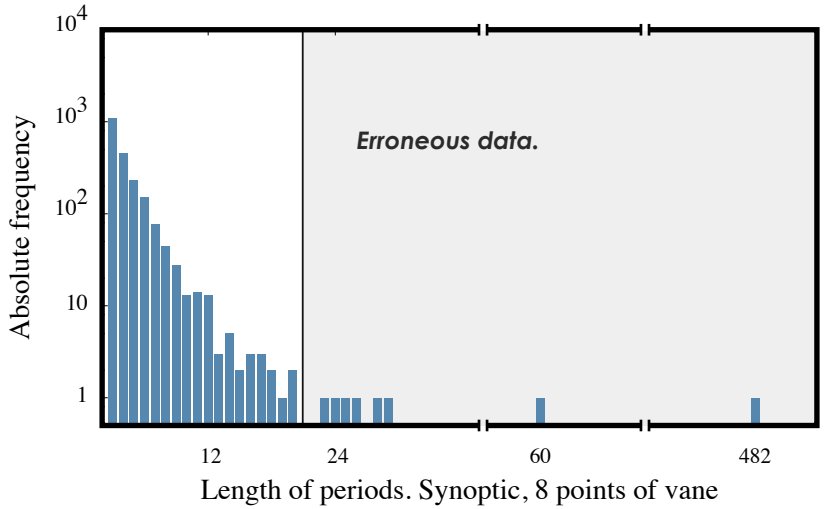
- Wind speed distribution for the whole database (blue bars).
- Anemometer limits (vertical lines).
- Hurricane speed distribution (orange bars).
- Shaded area: unrealistic values.
- The limit was imposed based on Graybeal et al. 2004.


WIND
SPEED
DIRECTION
1
Unification of measurement units (m/s)
Unification of direction criteria (deg)
Standardization of Observation times (UTC)
Compilation Adjustments
2
Intra-site Repetitions
Inter-site Repetitions
Duplication Errors
3
Consistency in Mean & Standard Deviation
Consistency in Criteria: Calm & True North
Consistency in Limits:
[0,112] m/s
[0,360]°
Consistency in Values
4
≥ 1 m/s
≥ 0.1 °
< 1 m/s
Abnormally Low Variations
Blip Check + Spatial Check
Abnormally High Variations
Temporal Consistency
5
Detection of Long Term Scale Shifts in Mean and/or Variance
Detection of Rotations between Wind Roses
Bias Detection
Quality Controled Data-Base

Abnormally Low variations

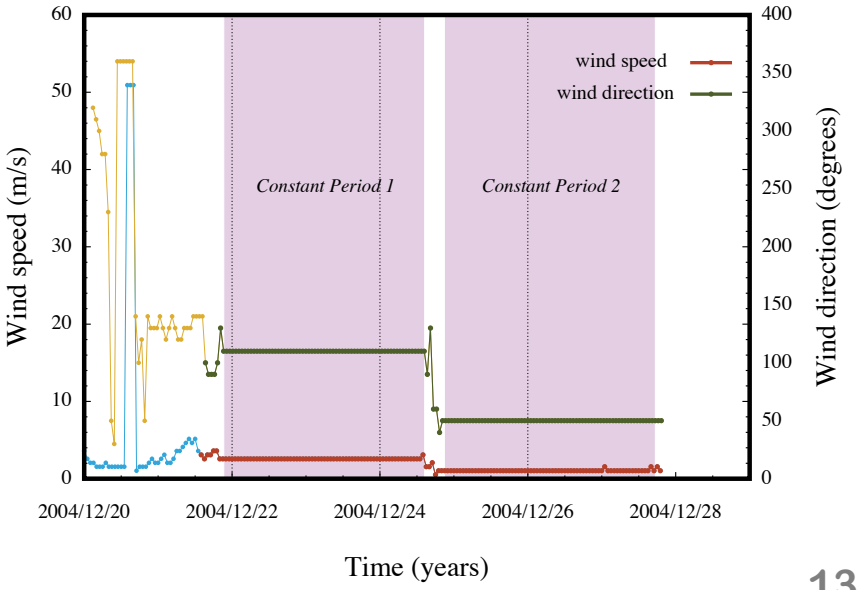
Wind speeds ≥ 1m/s
Wind directions ≥ 0.1°

- The procedure is similar to that of the duplication errors.
- The method is only able to identify erroneous periods longer than certain threshold.
- The threshold is specific for each time resolution and instrument precission.



Distribution example for synoptic constant periods with 8 points of vane:

- Wind direction constant periods (blue).
- The periods of the shaded area are considered erroneous.

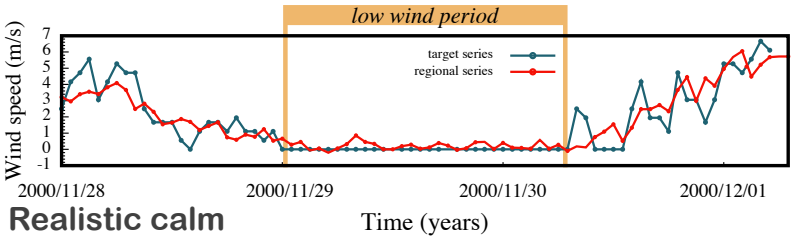


Example of a detected error with 2 consecutive constant periods for both wind speed (red) and direction (green).

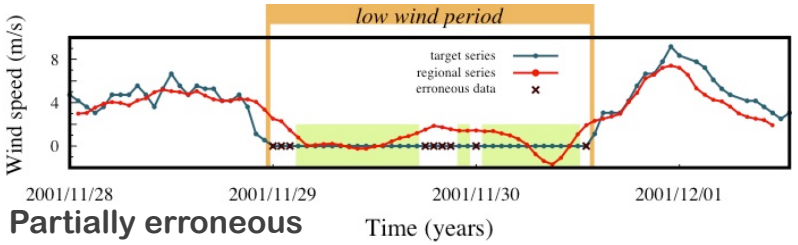
Abnormally Low variations

Wind speeds < 1m/s (calm/low wind periods)

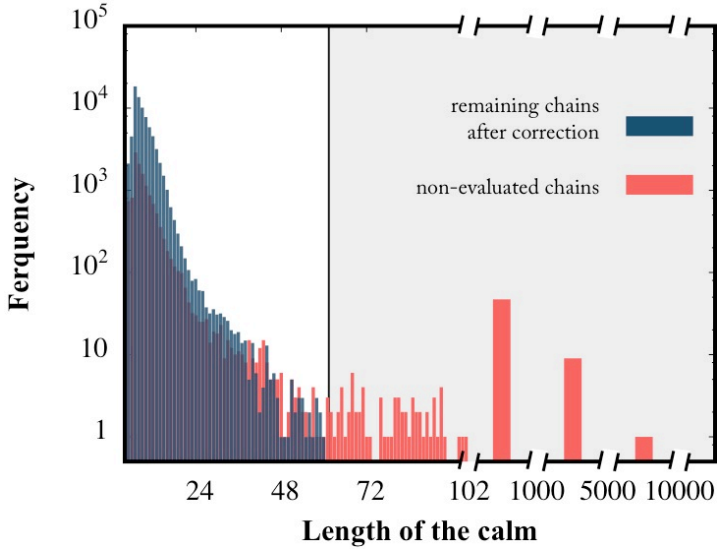
- The procedure is based on the comparison with neighboring stations.
- The method is able to identify erroneous periods of any length.



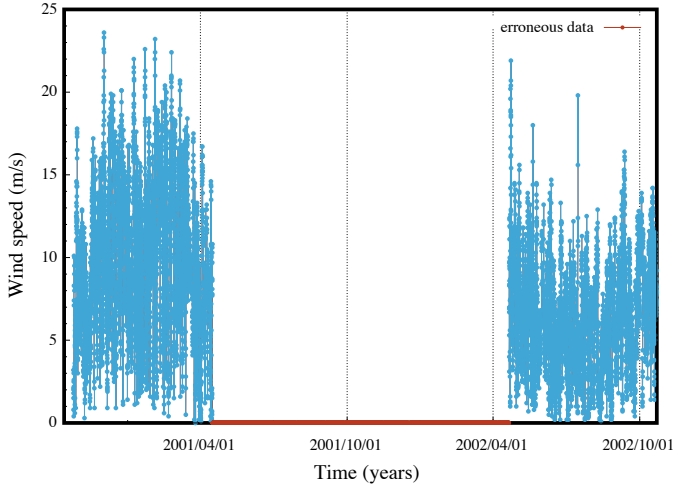
Realistic calm



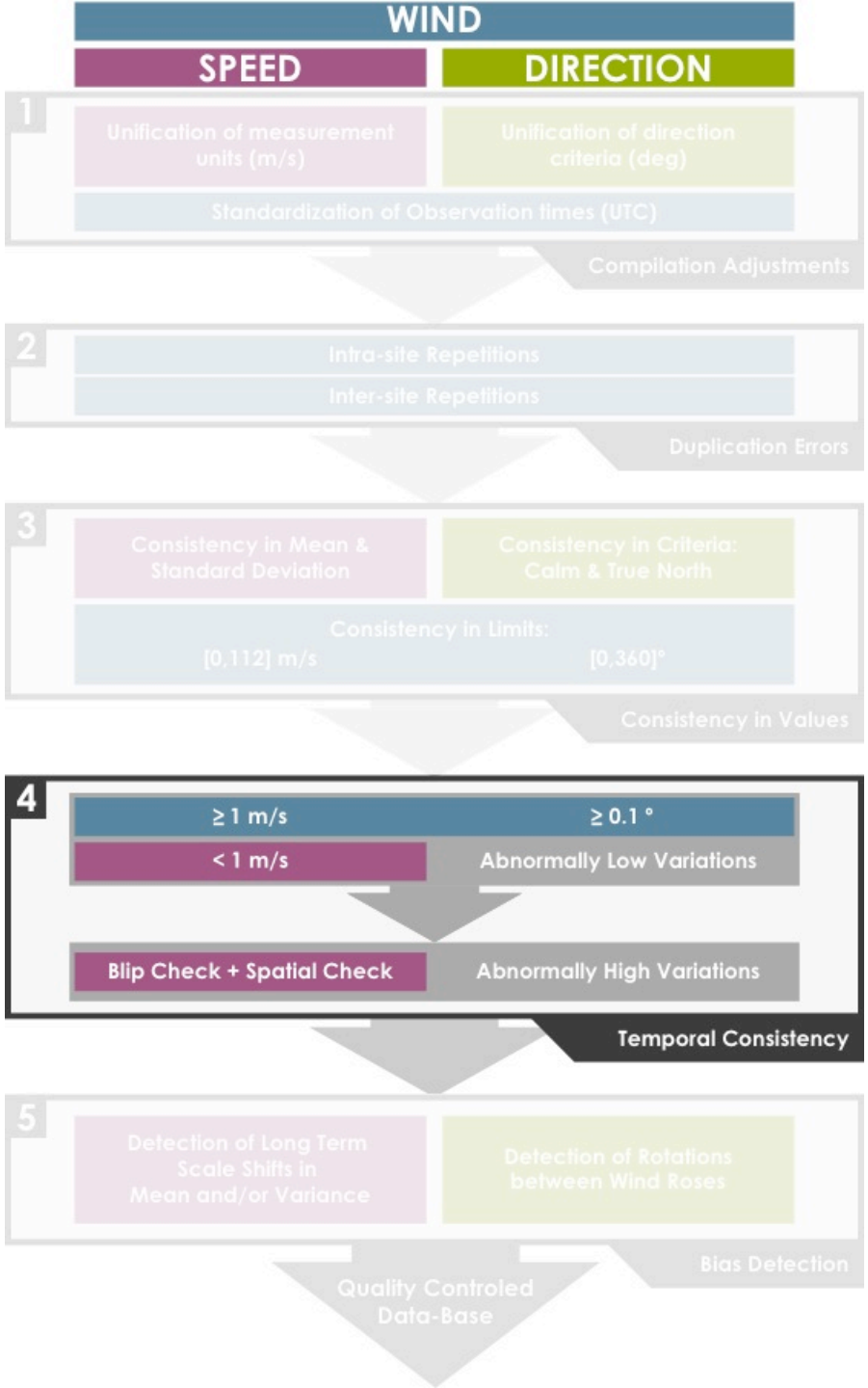
Partially erroneous



- Absolute freq. distribution of low wind constant periods regarded as realistic (blue).
- Distribution of non-analyzed periods, those with no neighboring regional stations, (red).
- All the periods in the shaded area are regarded as suspect.

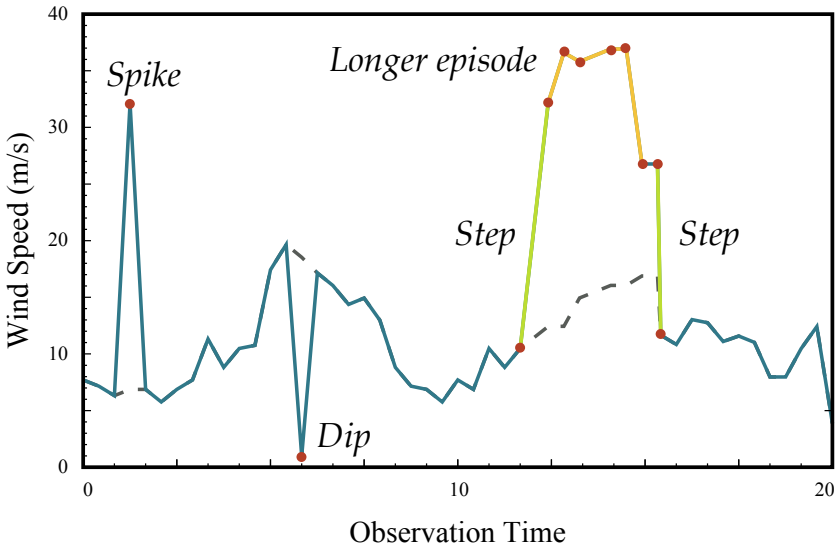


Example of an extremely long calm period of almost 2 years.


WIND
SPEED
DIRECTION
1
Unification of measurement units (m/s)
Unification of direction criteria (deg)
Standardization of Observation times (UTC)
Compilation Adjustments
2
Intra-site Repetitions
Inter-site Repetitions
Duplication Errors
3
Consistency in Mean & Standard Deviation
Consistency in Criteria: Calm & True North
Consistency in Limits:
[0,112] m/s
[0,360]°
Consistency in Values
4
≥ 1 m/s
≥ 0.1 °
< 1 m/s
Abnormally Low Variations
Blip Check + Spatial Check
Abnormally High Variations
Temporal Consistency
5
Detection of Long Term Scale Shifts in Mean and/or Variance
Detection of Rotations between Wind Roses
Bias Detection
Quality Controled Data-Base

Abnormally High Variations (blip check + spatial check)

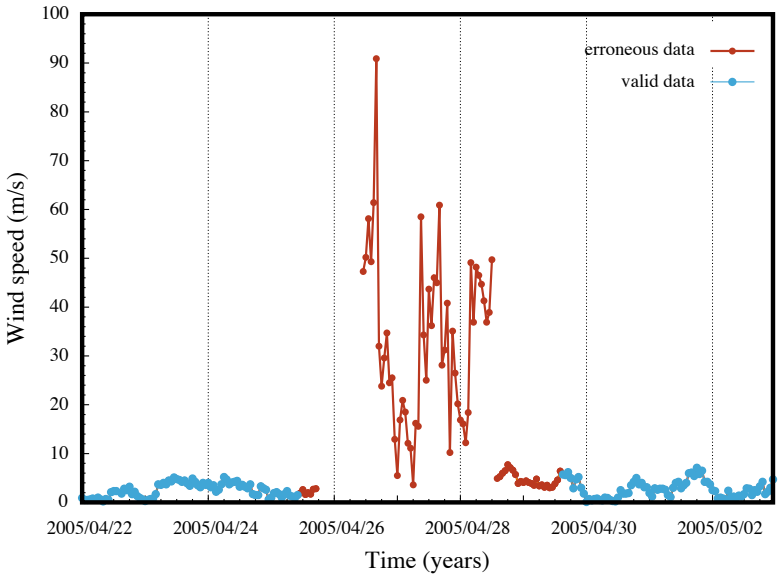
- The procedure evaluates the jump differences between 2 values in the context of the whole time serie (blip check).
- Is supplemented with an spatial evaluation of each data value (spatial check).



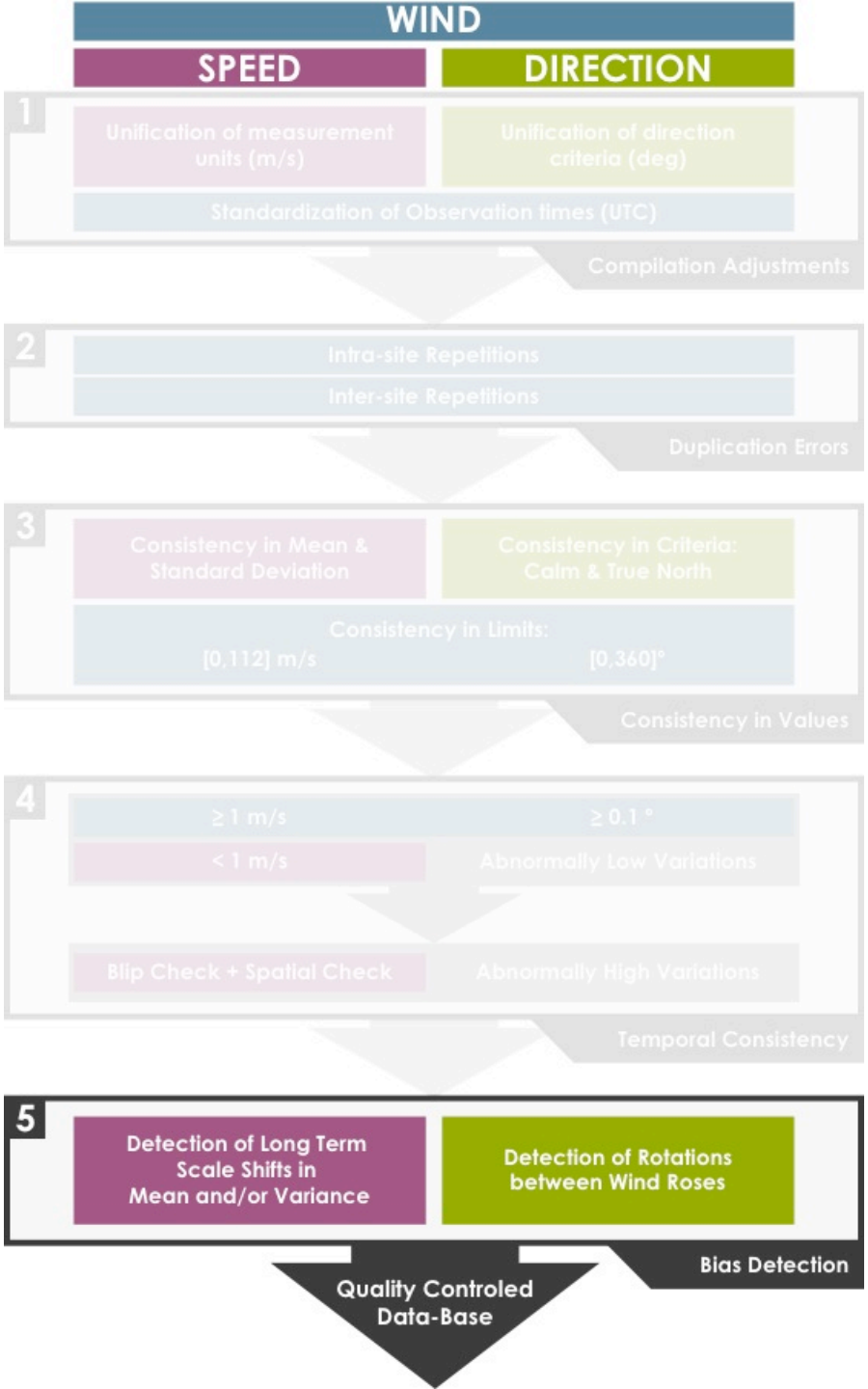
Different typologies in abnormally high variations:

- **Spikes** (detected with blip check)
- **Dips** (blip check)
- **Steps** (blip check)
- **Longer episodes** (blip check + spatial check)

(figure based on Fiebrich et al. 2010).

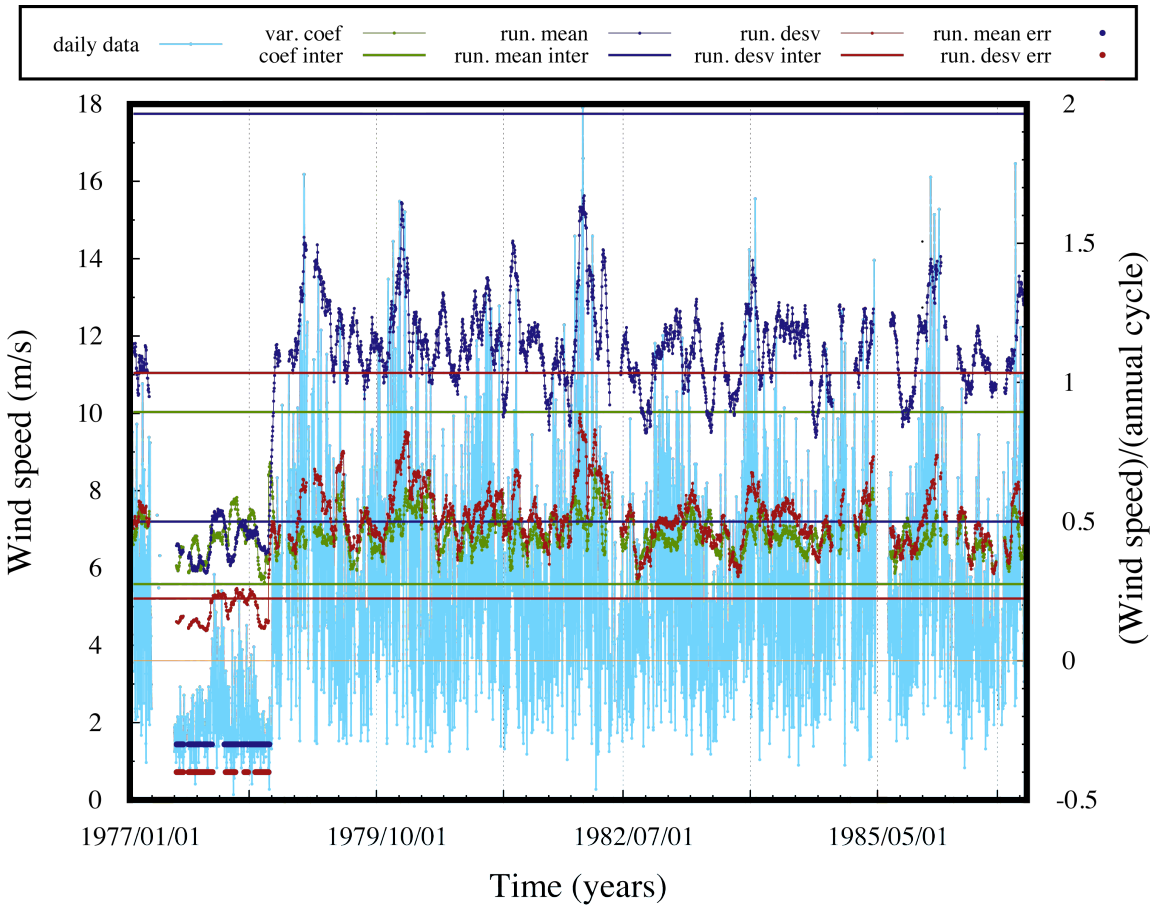


Example of a detected erroneous long episode (in red).


WIND
SPEED
DIRECTION
1
Unification of measurement units (m/s)
Unification of direction criteria (deg)
Standardization of Observation times (UTC)
Compilation Adjustments
2
Intra-site Repetitions
Inter-site Repetitions
Duplication Errors
3
Consistency in Mean & Standard Deviation
Consistency in Criteria: Calm & True North
Consistency in Limits:
[0,112] m/s
[0,360]°
Consistency in Values
4
≥ 1 m/s
≥ 0.1 °
< 1 m/s
Abnormally Low Variations
Blip Check + Spatial Check
Abnormally High Variations
Temporal Consistency
5
Detection of Long Term Scale Shifts in Mean and/or Variance
Detection of Rotations between Wind Roses
Bias Detection
Quality Controled Data-Base

Detection of Long Term Scale Shifts in Mean and/or Variance

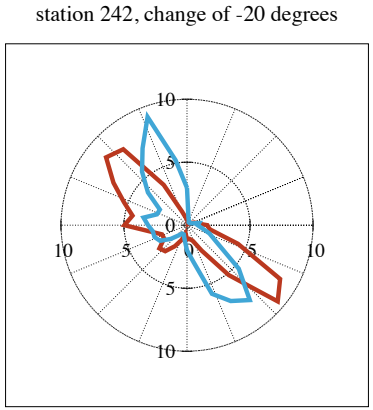
- The procedure works with deseasonalized and 1 month running averaged 3 variables: mean wind speed, standard deviation and coefficient of variation.
- For each of these variables a lower and upper interval are stablished.
- The outliers are thoroughtfully analyzed and if erroneous, the corresponding wind speed data is erased (in the original time resolution).

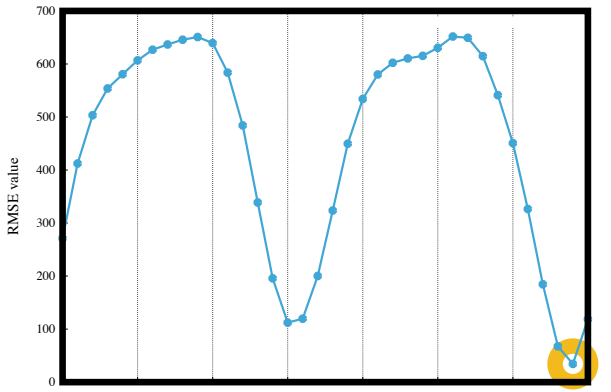


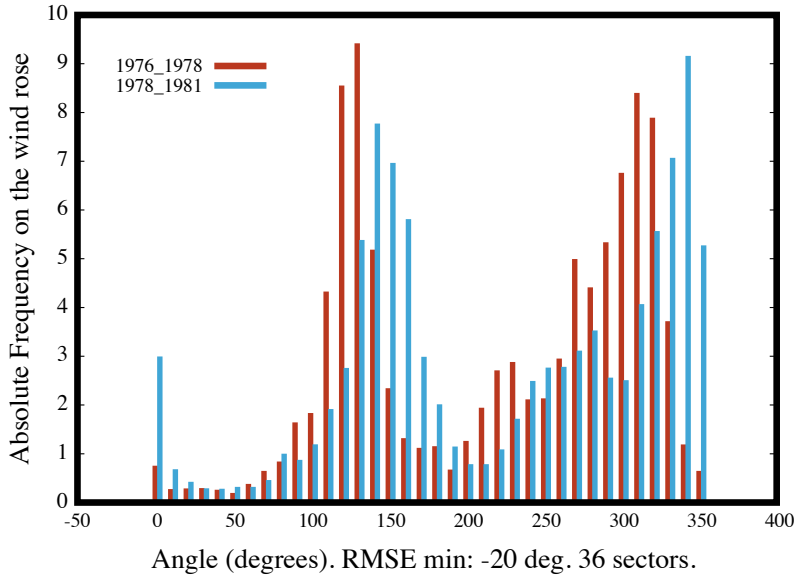
Example of an analyzed station.
~1 year of data was erased.

Detection of Rotations between Wind Roses

- Assumption: annual wind roses remain constant throughout time or vary slightly year to year.
- Limitation: rotations only identified with annual resolution.
- Consecutive annual wind roses (or periods) are compared in search for rotations.
- The roses are veered relative to each other, and RMSE values are calculated between them for each angle spin.
- The minimum RMSE gives the angle between the roses.
- Only rotations corresponding to angles of at least 20º are considered erroneous.

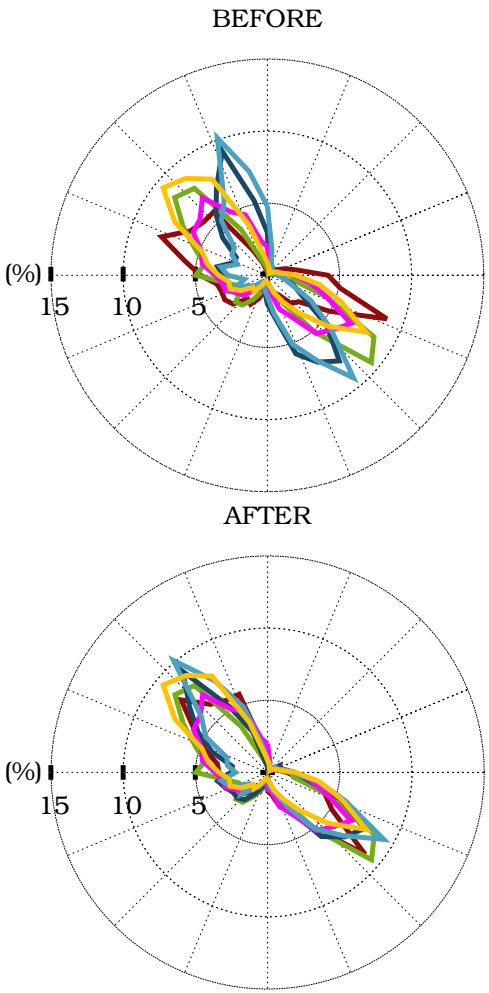




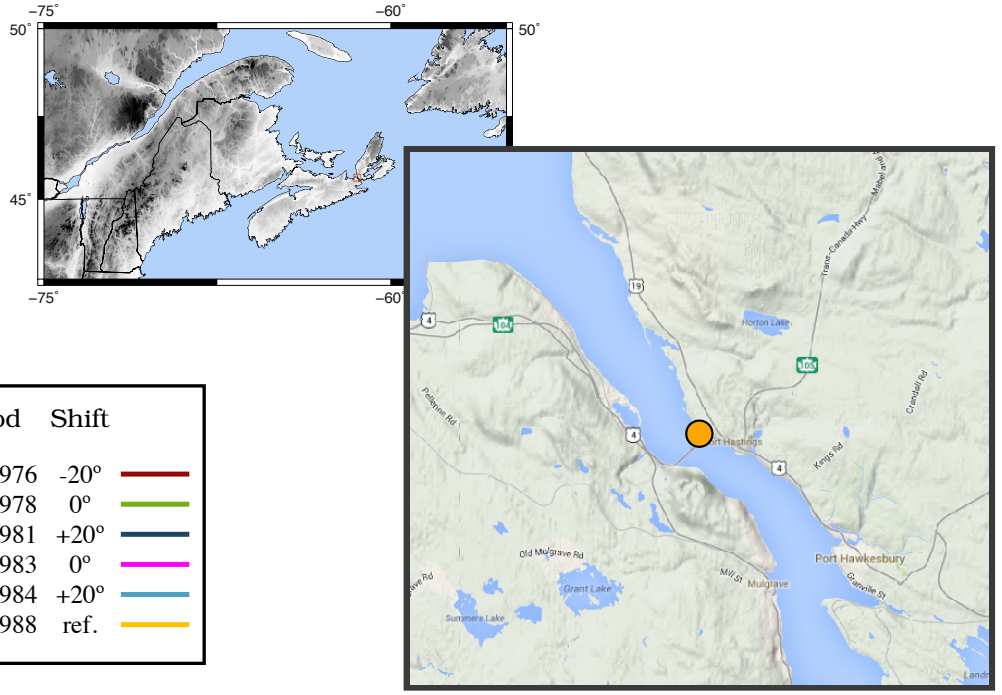


Detection of Rotations between Wind Roses

- All the rotated periods are corrected (rotated) to match the most recent one.
- One possible, the topography is also checked in search of accordance.

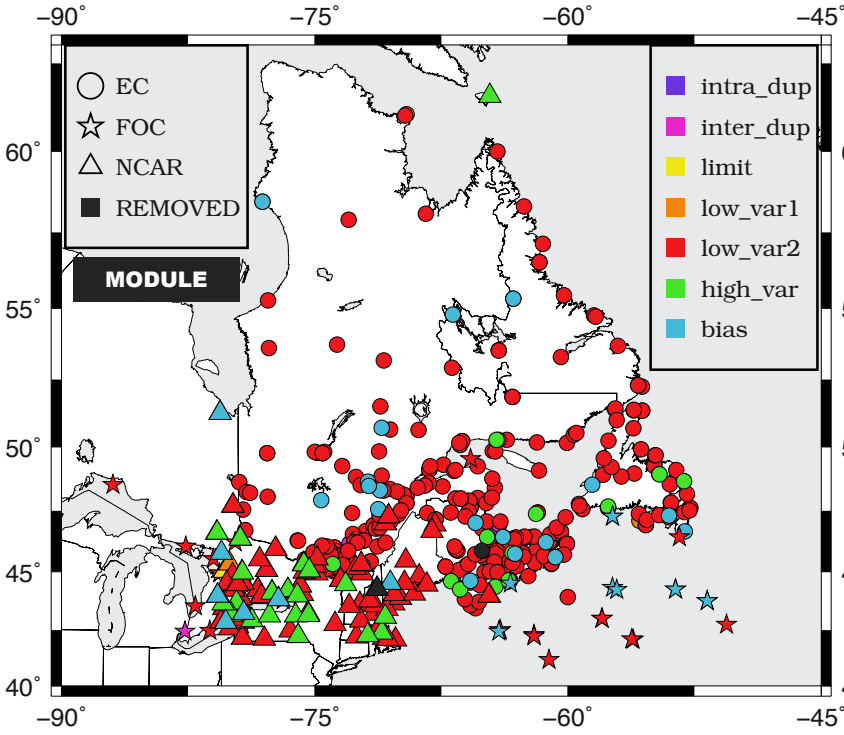


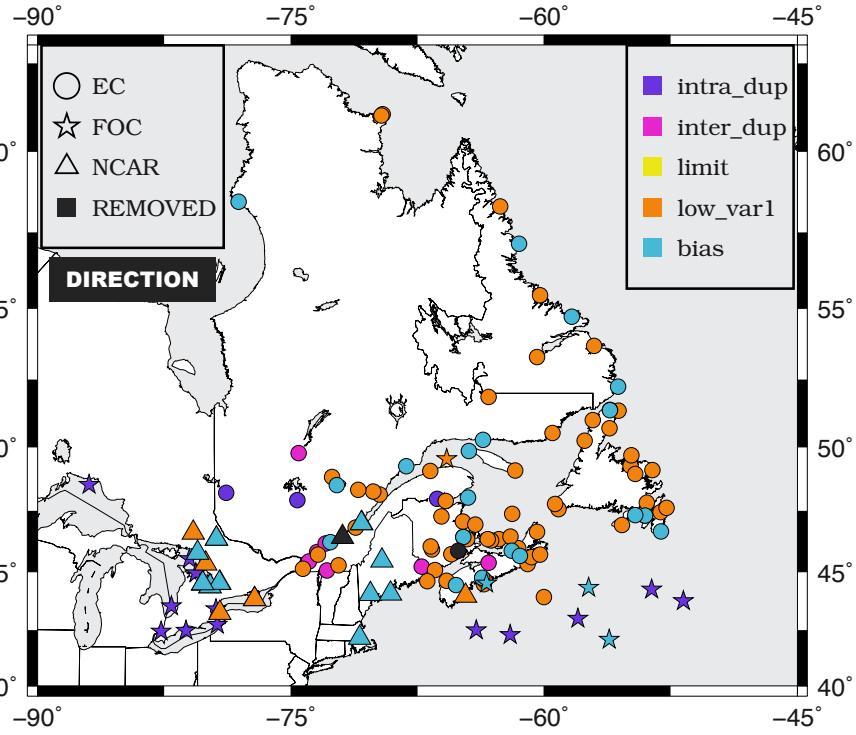
Period	Shift
1974-1976	-20°
1976-1978	0°
1978-1981	+20°
1981-1983	0°
1983-1984	+20°
1984-1988	ref.



Example of a station with 5 rotations.
The topographic channeling makes the various roses unlikely to happen.

Overview of the most important errors per station:

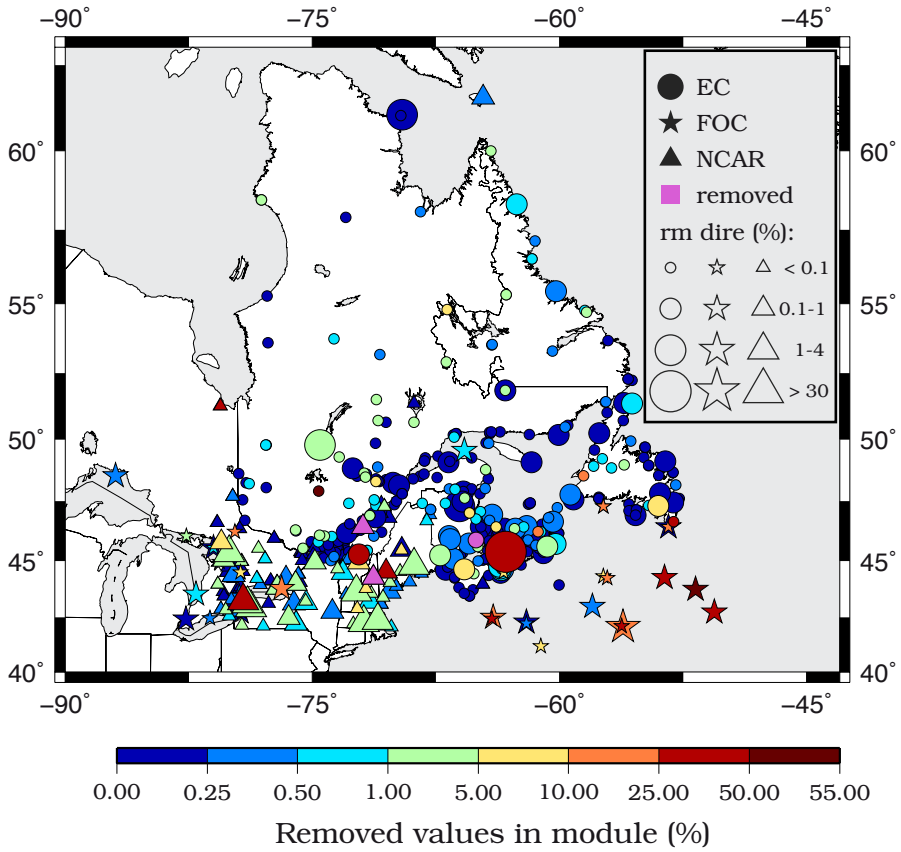




- Commonest error (by far): false calms.
- In buoys: false calms and bias.
- Two stations erased due to unrealistic mean.
- Stations with high var errors (NCAR, mostly) had few errors in total.

- Commonest errors: constant periods, then bias.
- In buoys: intra duplications (systematic failure)
- One station was erased due to inter duplication.

Impact on the Quality Controlled Database: Removed Data



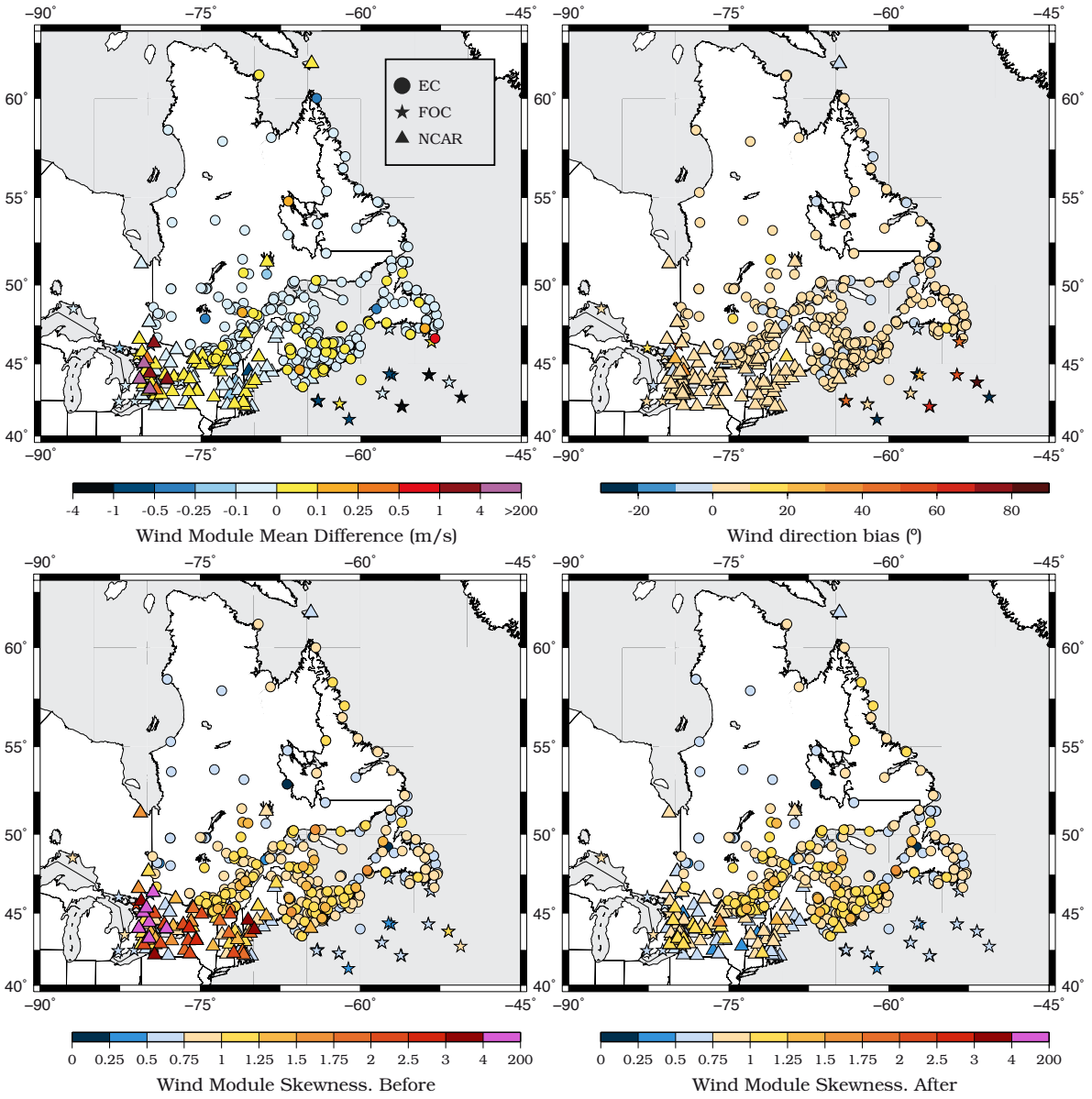
527 stations at the beginning:

- **Erased module data: 506 st. (900,000 records or 1.7%)**
- **Erased direction data: 318 st. (180,000 records or 0,3%)**
- **In TOTAL: 1,000,000 records (1.8%)**
- **4 stations removed**

Not included:

- **Modified due to compilation adjusments: ~98% of the database**
- **Direction wind rose correction: 1,300,000 records (2.4%)**

Impact on the Quality Controlled Database: Changes in Mean and Skewness



Thank you for your attention!

For further questions we can meet after the talk or you can email me at eelucio@fis.ucm.es