Probabilistic modelling of extreme wind speeds from ensemble forecasts for the generation of pre-warnings at the RMI

Pascal Mailier

Royal Meteorological Institute of Belgium
Avenue circulaire 3
B-1180 Brussels
T: +32 (0)2 7903966
E: pascal.mailier@meteo.be

Background

- The RMI issues
  - warnings of high wind speeds in the short range:
    - Own web site
    - Meteoalarm
    - Shipping
    - Media
  - pre-warnings in the early medium range:
    - ‘Bruxelles Environnement’
      - Authorities must close roads through wooded areas ahead of storms (e.g. Bois de la Cambre, Forêt de Soignes);
      - Pre-alerts 3 to 4 days beforehand helps with the planning so as to make sure that the necessary manpower and equipment are available;
      - Pre-warnings are triggered automatically when a given proportion or more of EPS members exceed a prescribed threshold.
      - Pre-warnings are disseminated via emails and SMS.
- This talk focuses on pre-warnings in the medium-range generated through the EPS

Limitations of the current method

- Current method essentially empirical

$$P(WSP \geq thr) = \frac{n}{N}.$$

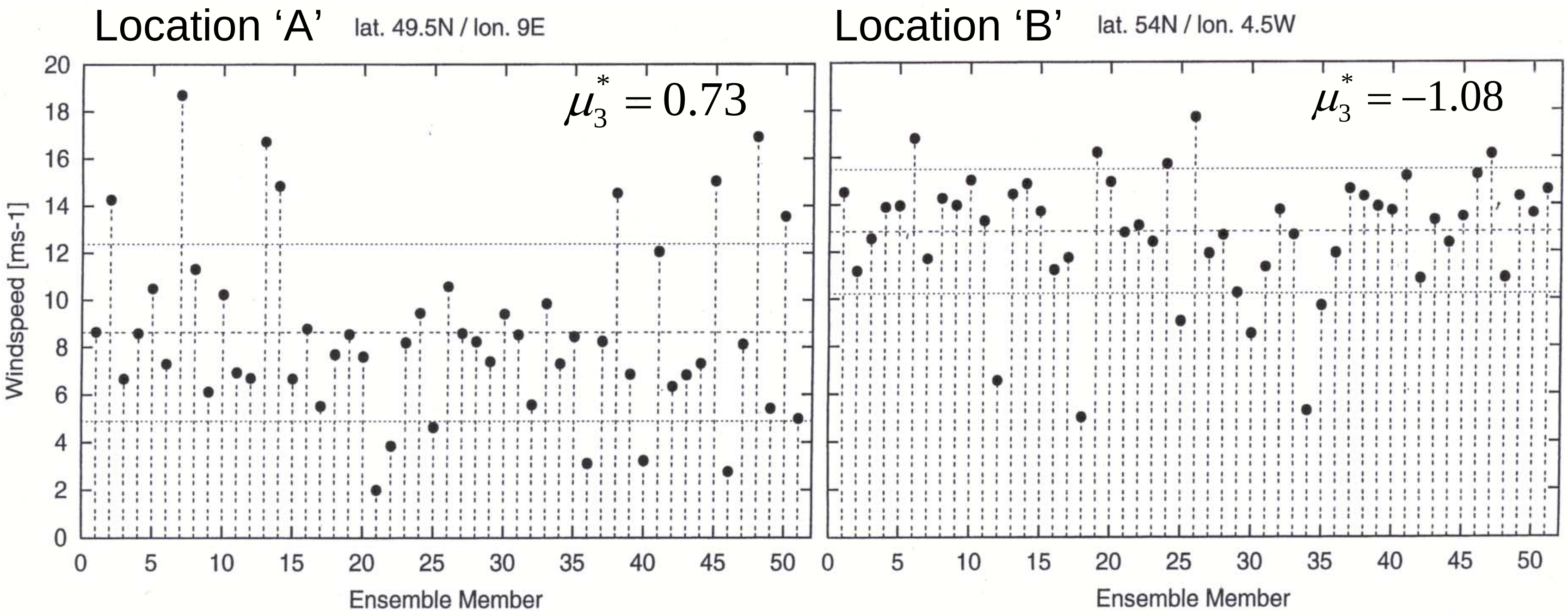
n: number of EPS members above some threshold *thr*

N: total number of EPS members

- Fairly small ensemble size poses a problem to estimate upper-tail probabilities when the distribution is skewed to the right.
- More in particular, probabilities collapse to zero at thresholds above the maximum of the EPS distribution:

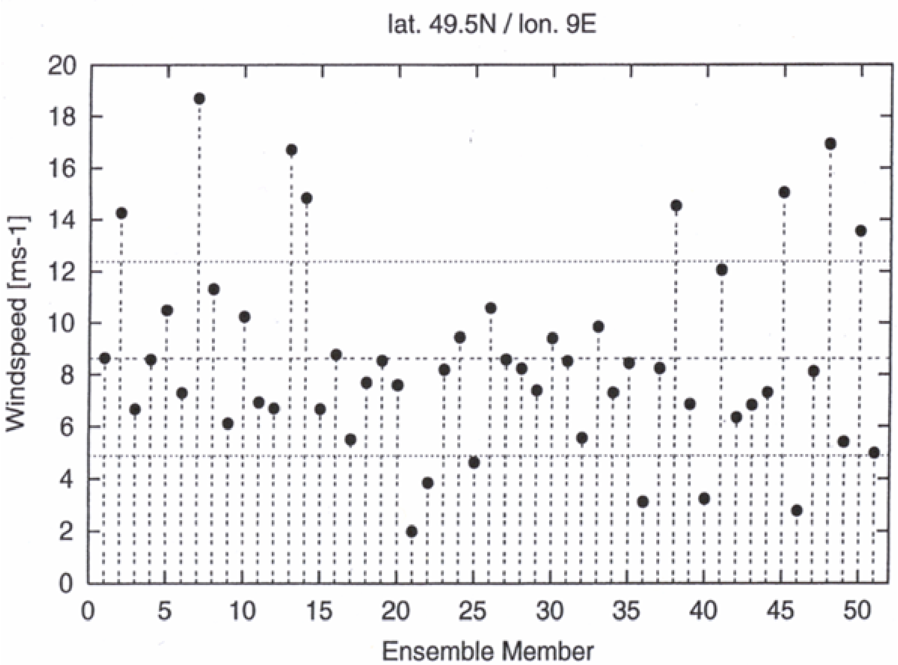
$$P(WSP > \max(wsp_1, wsp_2, \ldots, wsp_N)) = 0.$$

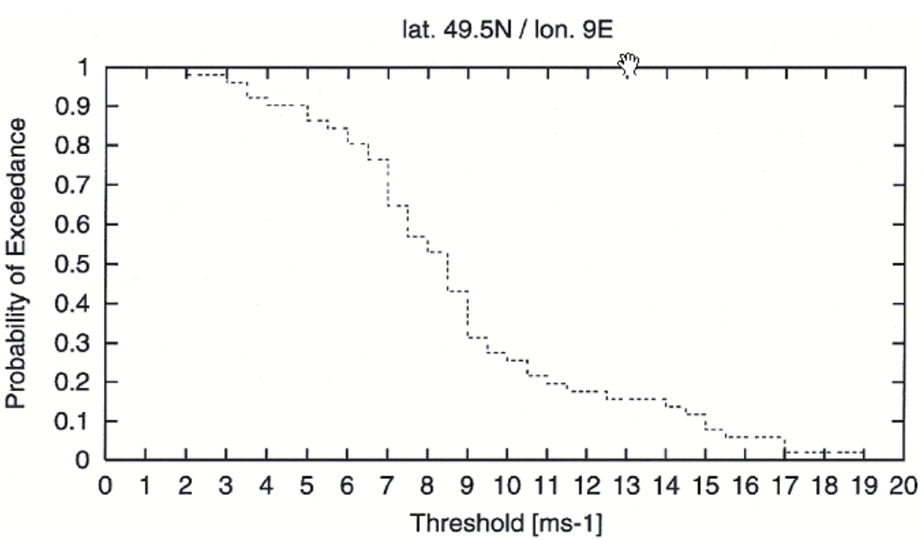
Typical EPS distributions with high wind speeds



- EPS distributions significantly asymmetric: skewed to the right at 'A' and to the left at 'B'
- EPS mean larger at 'B', but EPS maximum larger at 'A'
- Larger spread at 'A' due to longer upper tail
- $P(WSP > 19\ m/s) = 0$

Probability of exceeding a moving threshold





- Below *2 m/s*, all points are above threshold.
- More than ½ of all EPS members fall below threshold between *6 and 10 m/s*.
- Above *19 m/s*, all points are below threshold.
- Q: How can we estimate the probability of exceeding thresholds beyond *19 m/s*?
- A: Parameterise … but fit a GEV distribution on one sample maximum only doesn't work.
- The generalised Pareto distribution makes use of more data in the upper tail than just the maximum.

The Generalised Pareto Distribution (GPD)

- Model for the distribution of excesses above high thresholds (Pickands, 1975)
- The probability to exceed any value above a sufficiently high threshold u is:

$$P(X > u + x \mid X > u) \approx \left[1 + \xi \frac{x}{\sigma}\right]^{-1/\xi}.$$

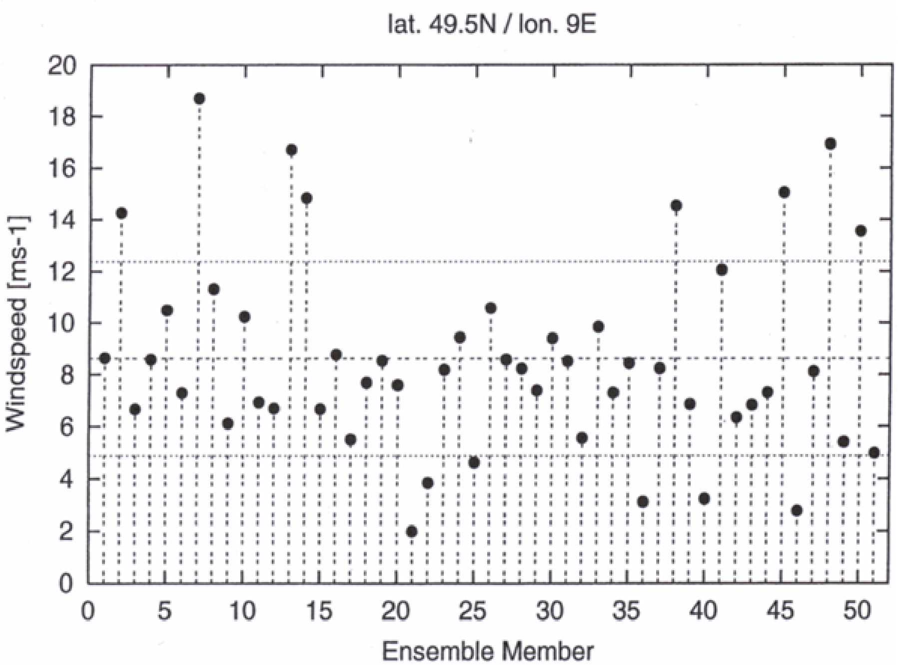
- The *maximum exceedance* x_e is :

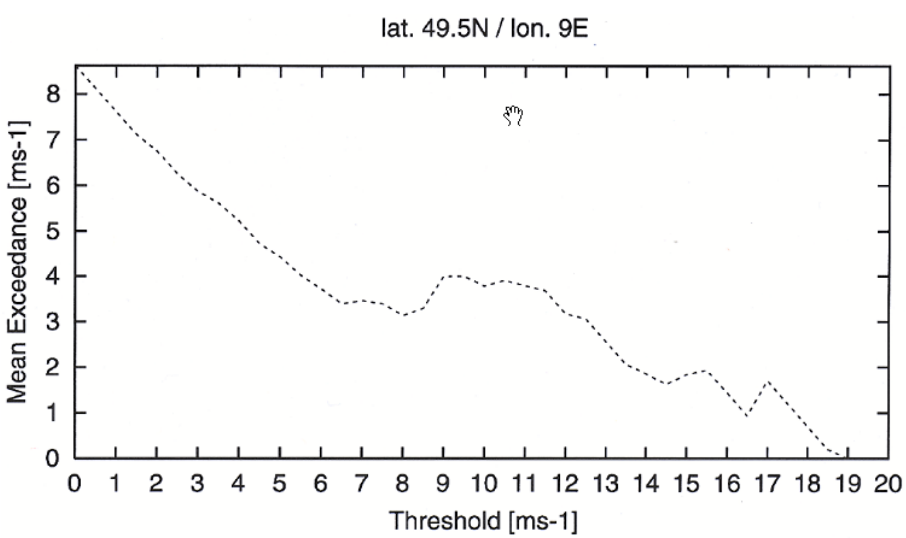
$$x_e = -\frac{\sigma}{\xi}.$$

- The *mean exceedance* over u is a linear function of u, provided $\xi < 1$:

$$E(X - u \mid X > u) = \frac{\sigma + \xi u}{1 - \xi}.$$

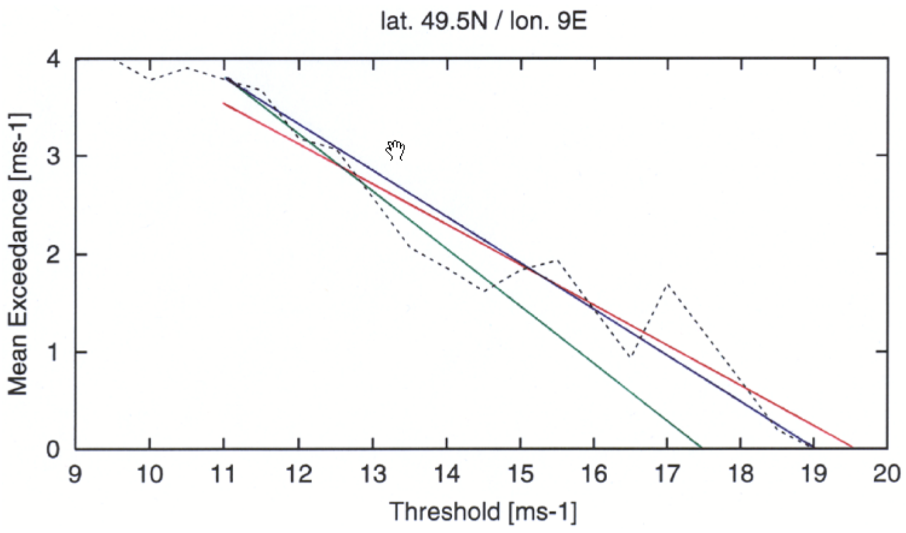
Mean residual life plot





- The *mean exceedance* (*ME*) is the mean distance between the threshold u and the points above.
- For $u = 0$ *m/s*, *ME* corresponds to the sample mean *8.6 m/s.*
- For $u \geq 19$ *m/s*, *ME* $= 0$ *m/s*.
- The decrease towards zero is not monotonic:
  - Linear below *6 m/s*
  - Nonlinear between *6* and *11 m/s* with local max at $u \approx 9$ *m/s*
  - Linear behaviour returns above *11 m/s* … but variability increases because of the small number of EPS members remaining above the threshold.

Estimation of the GDP parameters



- GPD assumed valid at thresholds above *11 m/s*.
- 3 straight lines:
  - Red: simple linear fit;
  - Green: linear fit discounts points ‘too far’ in the tail;
  - Blue: mean slope between *11* and *19 m/s*.

Linear fit

$$E(X-u \mid X>u) \approx a\,u+b$$

$$\hat{\xi} = \frac{a}{1+a}$$

$$\hat{\sigma} = \frac{b}{1+a}$$

$$\hat{x}_e = -\frac{\hat{\sigma}}{\hat{\xi}}$$

$$\hat{u}_e = u + \hat{x}_e = u - \frac{\hat{\sigma}}{\hat{\xi}}$$

GPD parameter and probability estimates

Fit	Red	Green	Blue
$\hat{\xi}$	*-0.7*	*-1.4*	*-0.9*
$\hat{\sigma}$	*13.8 m/s*	*24.8 m/s*	*17.1 m/s*
$\hat{x}_e$	*19.6 m/s*	*17.5 m/s*	*19.0 m/s*
$\hat{u}_e$	*30.6 m/s*	*28.5 m/s*	*30.0 m/s*
$P(X > 25 m/s)$	*0.03*	*0.06*	*0.04*

Observed wind speed was *25 m/s*

Conclusions and suggestions for future work

- An exploratory threshold method based on the GPD has been used to make inferences on extreme wind speeds from EPS forecast data at one grid point location.
- The approach taken is manual and subjective
  - Linearity does not necessarily guarantee that the GPD model is valid. Was *11 m/s* a sufficiently high threshold?
  - Uncertainty increases at higher thresholds as exceedances gets scarce.
- An objective procedure that could be automated is difficult to put in place.
  - Method of moments and maximum-likelihood estimation only work for restricted ranges of ξ.
  - Look at alternative estimation techniques, e.g. based on the principle of maximum entropy or other Bayesian methods.

Conclusions and suggestions for future work

- The small EPS size is an issue
  - High uncertainty on parameter values
  - Look at how successive EPS runs can be pooled together.
  - Look at how grid points over a region can be pooled together
    - Loss of spatial resolution
    - Non-independence of events at neighbouring grid points needs to be taken care of.
- GPD value: is it worth bothering?
  - Results from the GPD should be compared with tail probabilities obtained from standard Weibull fits.