Methodology for Computation of UK Windstorm Gust Return Levels on a High Spatial Resolution (100 m) Grid

Professor Mark Saunders and Dr Adam Lea
Department of Space and Climate Physics, University College London
Version 1.4 20th April 2017

Summary: This document outlines the methodology employed by the Climate Extremes Group in the Department of Space and Climate Physics at University College London (UCL) to compute windstorm gust return levels (from 1 year to 1000 years) at high spatial resolution (100 m grid) for the whole UK. The methodology is summarised under eight headings. A ninth section addresses the precision of the gust return levels. A product format section, references and an appendix provide additional information.

List of Contents:

1. UK Windstorm Definition. A UK windstorm is defined here as a synoptic-scale or mesoscale storm which has a wind strong enough to cause at least light damage to trees and buildings, and which may or may not be accompanied by precipitation. Extratropical cyclones are responsible for virtually all (98.9%) of the maximum 3-sec gusts defined here to be a UK windstorm event [*Saunders and Lea*, 2016]. The high winds associated with tornadoes and micro-scale storm phenomena are not included within the UK windstorm definition. In terms of scale, the definition includes all high windspeed storms with a size greater than ~10 km. Synoptic-scale storms (extratropical cyclones) have a size which ranges from 100 km to over 1000 km. Mesoscale storms (large thunderstorms and squall lines) have a scale which ranges from ~10 km to 100s km. Three-second wind gust is the wind intensity parameter that underpins this product because it has a robust link to windstorm damage [e.g. *Klawa and Ulbrich*, 2003] and is the standard parameter used in catastrophe models to assess UK windstorm risk and loss [e.g. *Waisman*, 2015].

2. Input Data. Three main input data sets are used. These are: **2.1** Hourly peak 3-sec gust recordings from 282 UK weather stations within the Met Office Integrated Data Archive System (MIDAS) [*Met Office, 2012*]. These data span the 45 year period 1969-2013. The 282 stations are spread uniformly across the whole UK (see Appendix figure). **2.2** UK land cover data at 100 m spatial resolution from the European Environment Agency's CORINE land cover/land use database 2006, version 17 [*European Environment Agency,* 2013]. **2.3** UK terrain model height data on a 50 m grid from the Ordnance Survey [*OS*, 2015] and Ordnance Survey of Northern Ireland [*OSNI,* 2016]. Our use of the MIDAS data for commercial value-added products is permitted via a perpetual licence with the Met Office. The re-use of CORINE and digital terrain model height data for commercial purposes is permitted free of charge. The other, more minor, input data sets employed are also available free of charge.

3. Data Cleaning. Standard and advanced quality control tests are used to clean the MIDAS station gust data; namely to identify, remove and correct bad data. Standard tests used include the identification of periods of data constancy and data repetition, the identification of sharp data jumps, and the identification of instances where data have clearly been digitised wrongly. The more advanced tests used include the examination of gust factors and the identification of incorrect outliers through the use of station pressure data and the use of wind and pressure data from nearby stations. The latter test has led to the removal of several high gust values due to them being localised in nature and not associated with significant wind damage. It is concluded that these high gusts are either associated with a localised micro-scale phenomenon or are spurious. The data cleaning has resulted in over 4000 MIDAS wind and gust records from UK anemometer stations 1969-2013 being either removed or corrected.

4. <u>Data Homogeneity Corrections</u>. Three corrections are made to homogenise the MIDAS UK station gust records. These are:

4.1 The location of each of the 282 MIDAS stations has been determined to 10 m precision. High and uniform precision in the location of stations is desirable to reduce errors in the modelling of gust perturbations due to local orography and to upstream surface roughness. This homogenisation has been achieved through study of station metadata stored in boxes in the Met Office Meteorological Archive combined with examination of satellite imagery available on Google Earth, Bing Maps and Google Maps. The 10 m precision is provided for the current location of open stations and for the final location of closed stations. Many of the station locations documented by MIDAS were found to be in error by between 100 m and 1 km.

4.2 Gust data imputation has been performed to give complete 45 year hourly gust coverage at all 282 stations. This homogenisation is desirable as no MIDAS anemometer station has complete hourly gust data 1969-2013; the completeness level of the 45 year hourly gust data for the 282 stations varies between 99.7% and 5%. Only 10 stations have less than 20% complete hourly gust coverage 1969-2013, of which seven are included to boost otherwise data-sparse parts of the UK. Missing hourly gust data imputation is necessary to avoid biases in computed return levels caused by gaps in the hourly gust record and by differences in the number of active stations. Hourly gust data imputation has been performed for every MIDAS station using a development of the method described in *Perry and Hollis* [2005]. Our imputation method is available on request. It has provided ≥99.5% complete hourly gust coverage over 45 years at 267 stations and ≥98% complete hourly gust coverage over 45 years at 281 stations.

4.3 The effective heights of the MIDAS 3-sec gust recordings 1969-2013 have been standardised to 10 m where possible, as 10 m is the international standard height to which wind and gust measurements are referenced. This homogenisation is desirable as the raw MIDAS gust recordings are not corrected for non-standard effective heights. Examination of station metadata stored in the Met Office Meteorological Archive showed that 73 station anemometers had non-standard effective heights (ranging between 5 m and 160 m) during at least part of the 1969-2013 period. The gust recordings for the relevant periods at these stations have been corrected to an effective height of 10 m by using a standard power law formula. The exponent in this formula varies between 0.085, and 0.14 depending upon local surface roughness [*Sachs*, 1978; *Met Office*, 1997]. The effective heights for 34 anemometer stations have not been determined and for these an effective height of 10 m has been assumed.

5. <u>Extreme Value Analysis</u>. Extreme gust return levels are computed from the cleaned and homogenised hourly station data by using the method of independent storms combined with a robust peak-over-threshold analysis. The method of independent storms is used to create the data base of independent extreme gusts at each station. This method increases the number of extremes within the analysis whilst ensuring their independence (eg *Palutikof et al*. [1999], *Brabson and Palutikof* [2000], *Della-Marta et al*. [2009]). We employ a dead-time of 24 hours between events at each station to ensure that the selected storm gust extremes are independent. A 24 hours dead-time agrees with the findings of *Brabson and Palutikof* [2000]. An 'event' is defined as a peak 3-second gust with a magnitude of at least 40 mph. Thus our method uses all the hourly gust data over 45 years for every station but our database of extreme gusts employed for the extreme value analysis comprises the highest gust value per independent storm at each station. The derivation of gust return levels and their associated confidence intervals at each station is made using a robust peak-over-threshold extreme value analysis [*Coles*, 2001, Chapter 4]. The distribution of gust exceedances above different extreme gust thresholds is modelled using the Generalised Pareto Distribution (GPD). We employ the *Grimshaw* [1993] method to determine maximum likelihood estimates for the GPD parameters as a function of extreme gust threshold. We employ the *Coles* [2001] diagnostic tests to select the optimum extreme gust threshold and thus the GPD parameters to use at each station. These tests are that the GPD shape parameter is approximately stable within limits and that the mean of the gust exceedances depends linearly on threshold. Asymmetric 95% confidence intervals on each gust return level at each station are determined by using profile log-likelihood intervals [*Coles*, 2001].

6. <u>Influence of Orography</u>. The influence of the surrounding three dimensional orography on local gust speed is computed at 100 m spatial resolution (10^{-3} of a degree) by using the model of *Troen and de Bass* [1986] applied to UK terrain height data on a 50m grid. The *Troen and de Bass* [1986] model employs the theoretical work of *Jackson and Hunt* [1975] who solved a simplified linear and steady-state version of the Navier-Stokes equations to give the boundary layer flow in the presence of orography for neutrally-stratified flow. The Jackson-Hunt theory was extended to three dimensions by *Mason and Sykes* [1979]. Here the velocity perturbations due to three-dimensional orography are computed using the Jackson-Hunt theory combined with a spectral Fourier transform solution. The final flow field is obtained by adding the sum of different velocity perturbations to the background flow. Speed-ups and speed-downs in gust speed occur respectively on upslopes and downslopes with the change in gust speed being proportional to the orographic slope. The wind direction employed to compute the orographic gust speed change for each 100 m grid cell is determined by spatially interpolating a single weighted windstorm gust direction at each station. A 'windstorm gust' is defined to be a gust with a return level of at least 1-year, and 'windstorm gust direction' is the gust direction at the time/hour of the windstorm maximum gust. A single weighted windstorm gust direction is obtained by applying the following weightings to the windstorm gust directions: 1 (for gusts with a return level of 1-1.99 years), 2 (for gusts with a return level of 2-4.99 years), and 3 (for gusts with a return level of at least 5 years). A global third order polynomial is used for the spatial interpolation [*Luo et al*., 2008].

7. <u>Influence of Upstream Surface Roughness</u>. The influence of upstream surface roughness on local high gust speed is computed at 100 m spatial resolution by using a new model developed in-house. This model utilises the European Environment Agency's CORINE land cover database 2006 (version 17) on a 100 m grid [*European Environment Agency,* 2013]. Upstream surface roughness lengths are assigned to each of the 44 separate CORINE land cover codes. Effective upstream surface roughnesses, z_{eff}, are computed for centered 30 degree sectors, each offset by 5 degrees, for each UK 100 m grid square). This is achieved by using the CORINE data with their assigned surface roughnesses together with the upstream wind directions employed to compute the gust speed change due to orography and an optimised upstream inverse-distance weighting model. The latter is determined empirically by selecting pairs of nearby stations where orography has little influence on the gust speed and where a sizeable difference occurs in the upstream roughness for the same 30 degree sector between the stations. The error in the high gust speed difference between the station pairs due to the deduced z_{eff} values is then minimised for different inverse-distance weightings. The z_{eff} values are converted to a corresponding percentage change in gust speed by using the *Wieringa* [1980, equation 3] exposure correction model. The deduced optimised inverse distance model has upstream weightings of 0.45 (0-1 km), 0.25 (1-2 km), 0.15 (2-3 km), 0.1 (3-4 km) and 0.05 (4-5 km).

8. <u>Spatial Interpolation</u>. Gust return levels are created on a 100 m grid for the whole UK (34 million grid cells) from an input of 282 irregularly spaced station values by using multiple regression and spatial interpolation (e.g. *Perry and Hollis* [2005]). Multiple regression analysis is used to model the spatial variation of each gust return level at high (100 m) granularity in terms of five geographic variables: orography, upstream surface roughness, altitude, latitude and longitude. The orography and upstream surface roughness variables are computed as described in §6 and §7 and are input as percentage perturbations. Altitude is the elevation above mean sea level (m). Latitude and longitude are included to capture spatial trends. The multiple regression is performed for each gust return level using the 282 station values (§6) as the dependent variable and the five geographic variables at each station as the independent variables. The regression estimate for the gust return level at each station is subtracted from the empirical gust return level to leave a set of 282 regression residuals which are largely free of geographic effects. These residuals are then interpolated onto a regular 100 m grid by using local polynomial interpolation [eg *Luo et al*., 2008]. Local polynomial interpolation captures short-range variability better than global polynomial interpolation. In applying local polynomial interpolation we used a polynomial of order 2, a 180 km search radius (which ensured the inclusion of at least 6 stations at all UK extremities) and Gaussian kernel weighting. Finally the regression model is applied at each

100 m grid cell and the output value is added to the interpolated residual to give a complete UK mapping of the gust return level at 100 m spatial resolution. For a few locations where the regression model performed poorly (parts of the Outer Hebrides and Scilly Isles) synthetic station data were introduced – a process called bogussing – to enable a sensible fit.

9. Precision of Gust Return Levels. Our computation of UK windstorm gust return levels at high (100m) resolution is underpinned by careful and rigorous data cleaning, data homogeneity corrections, physical modelling of the influence of orography and upstream surface roughness on wind gust, extreme value analysis and spatial interpolation. The product is offered at a spatial resolution which is over 40 times better than the current state-of-the-art for UK windstorm gust return level [*Met Office*, 2016]. Despite the methodological and analytical rigour, uncertainties remain. The two main sources of uncertainty come from the length of the observational input data and from the spatial interpolation (§8). The extreme value analysis (§5) outputs the deterministic (or the most likely) gust return level to high precision. However, because the input data are only 45 years long, the confidence intervals in the output gust levels become sizeable at the longer period gust return levels due to the uncertainty in the extreme tail fitting. For example, at Heathrow the 95% confidence intervals for the 10 yr, 50 yr, 200 yr and 1000 yr gust return levels are 68-75 mph, 73-86 mph, 76-95 mph and 79-106 mph respectively (computed using the *Coles* [2001] GPD likelihood-based method). The uncertainty arising from the spatial interpolation (§8) is computed by excluding each of the 282 stations in turn from the multiple regression and spatial analysis, and then computing the percentage error between the modelled gust return level value and the extreme value analysis gust return level value at each removed station location. The mean absolute percentage errors across 282 station locations are 6.1%, 6.4% and 6.8% for the 10 yr, 50 yr and 200 yr gust return levels respectively. The mean percentage error in each case is 0%. Figure 1 displays the distribution of the 282 absolute percentage errors in terms of probability of exceedance for the 50 yr gust return level. Similar error distributions occur for other gust return levels. This distribution of errors is the uncertainty in gust return level due to spatial interpolation. The main source of uncertainty for gust return levels longer than 50 years comes from the length of the observational input data.

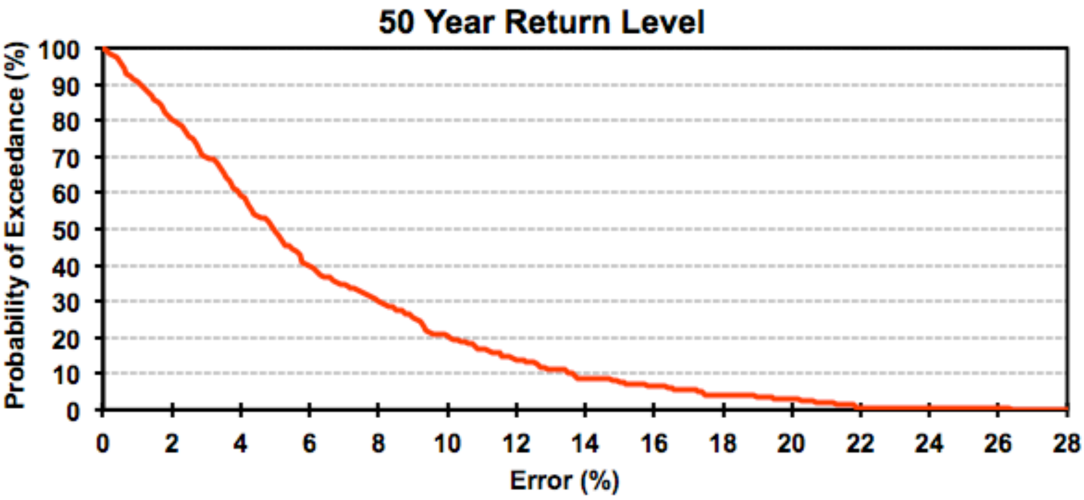

Figure 1. Distribution of absolute percentage errors in the 50 yr gust return level due to spatial interpolation. The distribution is nearly independent of gust return level.

10. Product Format. The UCL gust return levels are provided for display in ArcGIS. The data are gridded at 100 m resolution for the whole UK on the Ordnance Survey National Grid, which is also called the British National Grid [*OS*, 2016]. This grid uses a 1936 datum which is different to the datum used in the World Geodetic System 1984 coordinate system. The product is also available in high resolution static PNG image form.

Acknowledgement: We thank Dr Paul Northrop (Department of Statistical Science, UCL) for assistance concerning the extreme value analysis part of the methodology. This includes recommending the *Grimshaw* [1993] method as the most thorough and reliable for computing maximum likelihood estimates for the Generalised Pareto Distribution parameters, the kind provision of computer code to implement the Grimshaw method, the provision of computer code to compute

asymmetric confidence intervals for return levels, and the double-checking of our output gust return levels and confidence intervals.

11. References

Brabson, B. B. and J. P. Palutikof (2000): Tests of the Generalised Pareto Distribution for predicting extreme wind speeds, *Journal of Applied Meteorology and Climatology*, **39**, 1627-1640. DOI: http://dx.doi.org/10.1175/1520-0450(2000)039<1627:TOTGPD>2.0.CO;2

Coles, S. (2001): An Introduction to Statistical Modelling of Extreme Values, Springer-Verlag, London, 209pp. DOI: 10.1007/978-1-4471-3675-0.

Della-Marta, P. M., H. Mathis, C. Frei, M. A. Liniger, J. Kleinn and C. Appenzeller (2009): The return period of wind storms over Europe, *Int. J. Climatol.*, **29**, 437-459. DOI: 10.1002/joc.1794.

European Environment Agency (2013): CORINE land cover 2006 seamless vector data (version 17), accessed on multiple dates during 2013 and 2014. http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster-3

Grimshaw, S. D. (1993): Computing maximum likelihood estimates for the Generalised Pareto Distribution, *Technometrics*, **35**, 185-191. DOI: 10.2307/1269663. Stable URL: http://www.jstor.org/stable/1269663

Jackson, P. S. and J. C. R. Hunt (1975): Turbulent wind flow over a low hill, *Quart. J. R. Met. Soc.*, **101**, 929-955.

Klawa, M. and U. Ulbrich (2003): A model for the estimation of storm losses and the indentification of severe winter storms in Germany, *Nat. Hazards Earth Sci.*, **3**, 727-732. DOI:10.5194/nhess-3-725-2003.

Luo, W., M. C. Taylor and S. R. Parker (2008): A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from England and Wales, *Int J. Climatol.*, **28**, 947-959. DOI: 10.1002/joc.1583.

Mason, P. J. and R. I. Sykes (1979): Flow over an isolated hill of moderate slope. *Quart. J. R. Met. Soc.*, **105**, 383-395.

Met Office (1997): Source Book to the Forecasters' Reference Book, Met.O 1024, 279pp http://www.metoffice.gov.uk/media/pdf/9/3/Source_Book_to_Forecasters_Reference_Book_Complete.pdf

Met Office (2012): Met Office Integrated Data Archive System (MIDAS) land and marine surface stations data (1853-current). NCAS British Atmospheric Data Centre, accessed on multiple dates between 2012 and 2016. http://catalogue.ceda.ac.uk/uuid/220a65615218d5c9cc9e4785a3234bd0

Met Office (2016): Euro Windstorm Hazard Maps, accessed August 2016. http://www.metoffice.gov.uk/media/pdf/c/n/European_Hazard_maps.pdf

OS (2015): Ordnance Survey OS Terrain 50 opendata, accessed October 2015. https://www.ordnancesurvey.co.uk/opendatadownload/products.html

OS (2016): A guide to coordinate systems in Great Britain, v3.0, August 2016. https://www.ordnancesurvey.co.uk/docs/support/guide-coordinate-systems-great-britain.pdf

OSNI (2016): Ordnance Survey of Northern Ireland open data 50m digital terrain model CSV, accessed April 2016. https://www.arcgis.com/sharing/rest/content/items/a3863aa3c79e4372bcd23a629e7d6f81/data

Palutikof, J. P., B. B. Brabson, D. H. Lister and S. T. Adcock (1999): A review of methods to calculate extreme wind speeds, *Meteorol. Appl.*, **6**, 119-132. DOI: 10.1017/S1350482799001103.

Perry, M. and D. Hollis (2005): The development of a new set of long-term climate averages for the UK, *Int. J. Climatol.*, **25**, 1023-1039. DOI: 10.1002/joc.1160.

Sachs, P. (1978): *Wind Forces in Engineering*, 2nd Edition, Pergamon Press, Oxford, 410pp, ISBN: 9781483148359.

Saunders, M. A. and A. S. Lea (2016): Computation of windstorm catalogues and windstorm severity indices 1969-2013 for the UK and 11 UK regions, University College London internal document, 7pp, December 2016.

Troen, I. and A. F. de Baas (1986): A spectral diagnostic model for wind flow simulation in complex terrain. Proceedings of the European Wind Energy Association Conference and Exhibition, Rome, October 7-9, 1986, 243-249.

Waisman, F. (2015): European windstorm vendor model comparison. IUA Catastrophe Risk Management Seminar presentation, London, October 2015, accessed on 13th October 2016. https://www.iua.co.uk/IUA_Member/Events/Catastrophe_Risk_Management_Presentations/European_Windstorm_Vendor_Model_Comparison.aspx

Wieringa, J. (1980): Representativeness of wind observations at airports. *Bull. Am. Meteorol. Soc.*, **61**, 962-971.

12. Appendix

The locations of the 282 Met Office historical anemometer stations that contribute hourly 3-sec gust data 1969-2013 to the UCL UK windstorm gust return level product are displayed below.