The observed clustering of damaging extra-tropical cyclones in Europe

S. Cusack

Risk Management Solutions, Peninsular House, 30 Monument Street, London, EC3R 8NB, U.K.

Correspondence to: S. Cusack (stephen.cusack@rms.com)

Abstract

The clustering of severe European windstorms on annual timescales has substantial impacts on the re/insurance industry. Our knowledge of the risk is limited by large uncertainties in estimates of clustering from historical storm datasets typically covering the past few decades. Eight storm datasets are gathered for analysis in this study in order to reduce these uncertainties. Six of the datasets contain more than 100 years of severe storm information to reduce sampling errors, and observational errors are reduced by the diversity of information sources and analysis methods between storm datasets. All storm severity measures used in this study reflect damage, to suit re/insurance applications.

The shortest storm dataset of 42 years provides indications of stronger clustering with severity, particularly for regions off the main storm track in central Europe and France. However, clustering estimates have very large sampling and observational errors, exemplified by large changes in estimates in central Europe upon removal of one stormy season, 1989/90. The extended storm records place 1989/90 into a much longer historical context to produce more robust estimates of clustering. All the extended storm datasets show increased clustering between more severe storms from return periods (RP) of 0.5 years to the longest measured RPs of about 20 years. Further, they contain signs of stronger clustering off the main storm track, and weaker clustering for smaller-sized areas, though these signals are more uncertain as they are drawn from smaller data samples. These new ultra-long storm datasets provide new information on clustering to improve our management of this risk.

1 Introduction

European windstorms caused economic losses in excess of 25B USD (indexed to 2008) during the landmark years of 1990 and 1999 (Barredo 2010, using data from the NATHAN database of Munich Re). These huge losses were caused by multiple occurrences of multi-billion dollar loss events, as can be seen in figure 2 of Barredo (2010), and strongly suggested severe European windstorms are temporally clustered. Mailier et al. (2006) analysed clustering in the NCEP reanalysis dataset (Kalnay et al. 1996) and found clustering of winter wind storm occurrences in Europe, with evidence that clustering may be stronger for more severe storms. An analysis of similar data by Vitolo et al. (2009), and of other reanalysis datasets by Pinto et al. (2013), found similar results, and supplied clearer evidence of stronger clustering of the more severe storms.

The most important practical issue caused by significant clustering of severe storms is the threat to the solvency of re/insurance companies. The first step towards a more robust re/insurance industry, one which can better withstand extreme annual losses, is to measure the observed annual clustering of storms for different severities. Meteorological measures of storm severity are common in published work, such as relative vorticity at 850 hPa used by Mailier et al. (2006) and Vitolo et al. (2009), or the depth of the central pressure used by Pinto et al. (2013) and Economou et al. (2015). The damage potential of these storms is a more appropriate measure of storm severity for insurance purposes, taking into account its variability with local wind climate (Klawa and Ulbrich, 2003), and will be used throughout this study to characterise storm strength.

Karremann et al. (2014a) used severity metrics which were validated for re/insurance purposes, and measured storm severity in terms of local return levels. This use of standard insurance industry expressions of severity makes their results more relevant to end-users, but perhaps of more importance is that all storm severity metrics can be easily translated to this common scale of return levels to enable inter-comparison of disparate severity measures. Return levels will be used in this study to allow inter-comparison of a wide variety of storm datasets. Karremann et al. (2014b) extend results from Germany to many other countries impacted by wind storms to provide a fuller picture of clustering as a function of local storm severity in Europe. However, the true clustering climate is obscured by large uncertainties due to sampling errors, as illustrated by the 90% bootstrap confidence interval (CI) in Fig. 6 of Vitolo et al. (2009), based on 50 years of data. Those results imply a very wide range of true, underlying climates of storm clustering could produce the 30-year sample data of severe storms analysed by Karremann et al. 2014a and Karremann et al. 2014b.

Uncertainties from standard datasets are particularly large because clustering depends on the variance of annual storm counts, rather than mean behaviour. These large impacts of sampling and observational errors limit our knowledge of clustering from standard multi-decadal storm datasets. There are two options to reduce this uncertainty: (i) build models of the physical processes which produce clustering to fill in observational gaps; (ii) gain more knowledge of clustering either by new analysis methods or new historical datasets.

Regarding option (i), climate models attempt to simulate climate system processes and their long simulations have the potential to provide much smaller sampling errors. However, previous studies find significant differences between climate models and observed behaviour (e.g. Kvamsto et al., 2008, and the underestimate of clustering for most severe storms in Tables 3a and b of Karremann et al., 2014a). New research by Pinto et al. (2014) looks for the underlying mechanisms generating the cyclone families and persistent climate states that produce severe clusters on seasonal timescales. This information could be used to improve climate models, or as the foundation of simpler statistical models of the underlying processes which produce clustering, both of which could fill gaps in clustering knowledge.

Regarding option (ii), the novel analysis of a standard dataset by Hunter et al. (2015) reveals a link between annual frequency and severity of storms which informs on clustering behaviour. Alternatively, we can gain new knowledge of clustering from new storm datasets. This article presents new extended storm datasets and analyses their clustering character to produce a fuller picture of clustering. To this end, seven extended records of historical storms are described in Section 2, in addition to a more standard dataset of 42 years in length. The seven extended historical records reduce sampling errors by their increased length, and provide insight into impacts of observational errors, since these datasets are based on independent data sources and analysis methods.

Section 3 describes the method of analysing data and has two main parts: first, the measure of clustering for a group of storms is defined, and second, the method of converting the disparate measures of storm severity in the eight different datasets to a common form is described. The observed clustering of European windstorms is presented in Section 4, together with a discussion of estimates and errors. A summary is given in the final section.

2 Data

A total of eight storm datasets are used in this study, all of which contain the date and a measure of damage severity of each storm. Table 1 provides a summary description of all storm datasets described in more detail below. The last column of Table 1 provides the brief name used for each dataset.

Two extended datasets of storms in the U.K. are studied. The first (UK-Lamb-300) is the list of storms and their Storm Severity Index (SSI) values listed in pages 8 to 10 of Lamb and Frydendahl (1991). Their SSI measures are estimated from surface weather reports, meteorological analyses and damage information from a variety of documentary sources and reflect the damage severity of storms. The clustering analysis presented in Sect.4 is restricted to the storms in the period from 1690 to 1989, due to incompleteness of reportage in earlier times, and to those 44 storms with SSI values of 2000 or higher. This high severity threshold ensures both a more homogeneous time-series and more confident estimates of their severity, due to the increased attention and better documentation of the most severe storms in this period.

The second U.K. dataset (UK-RMS-160) is a list of storm fatalities in the UK in the period 1835 to 1994 gathered by Risk Management Solutions (hereafter RMS). This was extracted from archives of The Times newspaper by searching its Index using the terms “storm” and “gale” (Robert Muir-Wood, personal communication). The fatalities are considered to be accurately reported throughout this period, and the dataset is considered complete since bigger national issues would reduce space or prominence attached to more minor storms events but not remove them completely. Two factors were applied to reported fatalities to homogenise this dataset: first, a population factor indexes all fatalities to 1994 national population levels, and second, night-time storm fatalities are scaled by a factor four to produce as-if daytime fatalities. Storm fatalities reflect population densities hence this index is more closely related to actual damage than wind speed intensity, and given the much more densely populated southern half of the U.K., the dataset is viewed as a proxy of storm damage severity in the southern half of the country. Figure 1 shows a time-series of standardised storm fatalities for the full 160 year record.

Extended 105-year records of winds at five stations from the Royal Netherlands Meteorological Institute (KNMI) are used to define storminess in the Netherlands in the period from 1910 to 2014 (NL-KNMI-105). The data and analysis are described in Cusack (2013). In brief, the winds from five weather stations are merged to form an aggregate SSI value for each storm. The data are complete, and the spread of station locations geographically ensures the storm severity represents national values. The largest uncertainties arise from several significant changes in wind measurement practice in the first few decades. Intensive homogenisation methods are applied, based on station metadata made available by KNMI, complemented with statistical methods (see Supplementary Information of Cusack, 2013). The homogenisation serves to reduce but cannot completely remove observational errors, and the final time-series of storm severities will inevitably contain uncertainties. The top 30 or so storms have been compared with documentary sources such as the KNMI list available at http://projects.knmi.nl/hydra/cgi-bin/storm_list.cgi , and other independent sources based on documentary records, and corroborate the significant storms in this KNMI-derived dataset.

The public website of Deutscher Wetterdienst (DWD) provides peak gust data and associated metadata for climate stations covering the past 60 years (DE-DWD-60). Seven stations with minimal changes to the wind observing system over their entire records were chosen, with locations shown in Fig. 2. SSI values for Germany were computed for individual storms over the past 60 years using the method from Cusack (2013) applied to these seven stations. While the stationary observational practices reduce uncertainties in results from inhomogeneities, the small number of selected stations covering such a large area introduces errors in estimated severity. The top storms produced by this analysis were compared to the list of DWD storms provided in Table 1 of Karremann et al. (2014a) – based on much higher station density – and there is high correlation. The larger spatial extent of more severe storms leads to this result.

Brázdil et al. (2004) describe windstorm damage in the Czech Republic from 1500 to 1999 based on research of a wide variety of documentary sources (CZ-Brázdil-500). Their detailed descriptions have been manually analysed into two storm severity classes: class 1 for local-scale damage, or large-scale weak damage, and class 2 for widespread, intense damage. Summer storms forced by convection have been removed. Figure 3 displays the number of storms per century for each severity class. Strong temporal trends can be seen in these data: there is a large increase in frequency of weaker storms in the last 200 years, and increasing occurrence of the stronger storms throughout the period. These temporal trends are most likely due to changes in amount of documentary evidence through time. Figure 3 indicates that the reduction in sampling error achieved by such a long dataset will be offset to some extent by larger uncertainties from reporting inhomogeneities. The impact of these non-stationarities will be explored in the Results section.

Stucki et al. (2014) describe a database of wind storms in Switzerland during the period 1859-2011 (CH-Stucki-153). In brief, they use a wide variety of information, including damage information from buildings and forestry and meteorological information from anemometers and reanalyses, to identify storm events then assign one of three severity ratings to each storm, depending on the severity and spatial scale of damage in Switzerland. Summer wind storms are not infrequent in Switzerland, and all damaging wind events from May to September in the Stucki et al. database are excluded from this analysis of extra-tropical cyclone clustering. A full listing of the wind damage events in their database is given in the Supplementary Information of Stucki et al. (2014).

Emmanuel Garnier (private communication) provided a dataset of storms in France covering the period 1500-1999 based on his research of documentary archives for descriptions of wind damage (FR-Garnier-350). The historical storms are assigned a severity using the Beaufort scale, based on the documented damage severity and spatial extent. The present analysis will focus on the 1650 to 1999 period when documentary evidence is considered more complete and homogeneous for severe storms. Our internal validation indicated gaps in the Garnier record in the 19th century, and this has been alleviated by the inclusion of storm information from Bessemoulin (2002) to produce a more complete dataset of historical storm damage in France, though the completeness of the dataset in the late 18th and early 19th centuries is uncertain, since both sources contain little information in a period when other parts of Europe were stormy, especially in the 1815 to 1840 period. The count of storms by decade is shown in Fig. 4, split by their damage severity measured using the Beaufort Scale.

The storm footprints described in Bonazzi et al. (2012) are the most spatially comprehensive set of storms (EU-RMS-42). In brief, these footprints are derived from datasets of weather station peak gusts from fifteen countries beginning in 1972. The gust datasets comprise freely available data from national meteorological services, together with some RMS purchases from private providers. The locations of weather stations with 15 or more years of peak gust measurements are shown in Figure 5. There are several hundred stations with shorter records, particularly in east Europe, to complement the stations shown in Figure 5. Each storm footprint consists of the maximum observed gusts at each station for the entirety of the storm, which are then spatially interpolated to a more regular grid. The SSI is used to characterise the damage severity of these storms using the method described in Cusack (2013). The 135 footprints used in Bonazzi et al. (2012) are supplemented with seven storms in 2011 to 2013 (Yoda in November 2011; Friedrich, Joachim and Patrick/Dagmar in December 2011; Ulli in January 2012; Christian in October 2013 and Xaver in December 2013) using the same data and methods, to form a set spanning the 42 year period from 1972 to 2013. This set of 142 storms contains the top 20 to 25 of the strongest storms in major countries such as Germany, France, U.K. and Netherlands, and the top 10 to 20 storms in other countries, for the 1972-2013 period. However, due to the large footprints, all fifteen countries are affected by many more storms than these limits. More specific details on this dataset are given in Bonazzi et al. (2012).

3 Analysis methods

The strength of clustering used in most research to date adopts the metric first proposed by Mailier et al. (2006). Given a time-series of annual storm counts, X_i, where i=1, 2, …, N and N is the total number of storm years, Mailier et al. (2006) defined clustering using the dispersion statistic D:

$$D = \frac{Var(X)}{E(X)} - 1 \quad (1)$$

where Var(X) is the variance and E(X) is the expected (or mean) value of observed yearly storm counts. As the variance of a Poisson process is equal to its expected value, Eq. (1) can be re-written as:

$$D = \frac{Var(X) - Var(Pois)}{Var(Pois)} \quad (2)$$

where Var(Pois) is the variance of a Poisson process with expected value E(X). Mailier et al.'s metric of clustering is the relative excess variance of the data above a Poisson process.

Raschke (2015) described how D is proportional to the total rate of storms in the set being analyzed. Therefore, D reflects both the strength of clustering and the size of the storm group studied. Raschke proposed a new metric of clustering called "Beta" which isolates clustering strength from the size of storm group being studied. Raschke's metric simplifies to the dispersion statistic in Eq.(1) normalized by the expectation of observed yearly storm counts (the mean rate):

$$\beta = \frac{D}{E(X)} \quad (3)$$

Since the variation of clustering with storm strength will be explored, and more severe storms are rarer, Eq. (3) will be used for all results in Sect. 4 to ensure no artefact of dependence on storm numbers.

For each dataset, all storms matching or exceeding a specified damage threshold in storm years defined from July to following June were identified, then estimates of variance and mean annual occurrence rates are estimated directly from the data, which are used to specify β in Eq. (3). Various damage thresholds are used in each dataset to explore the variation of clustering strength with storm severity. These severity thresholds are expressed as return levels, following Karremann et al. (2014a; 2014b), and we refer to them as return periods (RP). In brief, the RP is defined to be the inverse of the annual frequency of storms greater than or equal to the particular threshold severity. For example, if a group of storms contain an average annual rate of 0.5 storms per year matching or exceeding the threshold, then the storm severity is defined to be RP = 2 years. This representation unifies dissimilar measures of severity (e.g. SSI, damage classes in Switzerland, U.K. storm fatalities) to enable their inter-comparison.

The uncertainties in the best estimates of β are analysed to provide more information on estimates of storm clustering. The first source of uncertainty is due to the effect of finite sample sizes on estimates of β and is related in concept to the standard error. It is a measure of the spread of β values associated with finite sampling of the true storm population and its estimation is now described. From the historical sample containing N years of historical storms, the parameters of a Negative Binomial model are estimated. Then, an artificial set of N data points are randomly drawn from this model, and repeated to make 50,000 artificial datasets. The β values of each of the 50,000 time-series are computed, from which the 95th confidence interval (CI) is obtained. The 95th CI is used to represent impacts of finite sample sizes on β estimates.

The second source of uncertainty is referred to as observational error and is due to inaccuracies in measured data which are independent of errors due to finite sample sizes. This type of error is unique to the observational datasets being studied. A method of approximating its impact was created for storm datasets, and is described using an illustrative example in which observational errors are to be computed for the subset of storms exceeding RP1 severity in a 40-year dataset. There are 40 storms with RP1 or greater severity in a 40-year time-series. It is assumed that the strongest storms in the top half of this subset – 20 storms – are known and fixed, while the storms of rank 21 to 40 are subject to measurement uncertainty. This uncertainty is simulated by randomly selecting 20 storms from ranks 21 to 60 of the original storm set, to form a new subset of 40 RP1+ storms. The random selection of 20 storms from ranks 21 to 60 is repeated to make 1,000 storm sets, and the 95th CI is formed from the 1,000 β values. This method is intended to produce a plausible guide to impacts of measurement errors on estimates of β values.

4 Results and discussion

Figure 6 displays the variation of clustering with storm severity based on the EU-RMS-42 dataset. The dashed lines in Fig. 6 represent the 95th CI for each β estimate, while the dotted lines represent uncertainty due to observational errors, and they indicate large uncertainty in estimated β values from both sampling and observational errors. Combining these two sources of uncertainty leads to the conclusion that the amount of clustering at any specific severity threshold would not be distinguished from a Poisson process (β=0) at the 5% level.

This assessment of uncertainties at individual points is distinct from the broader question of whether the entire collection of data in Fig. 6 is clustered. This is assessed as follows: a set of storms equal to the largest rate (2.0 in Fig. 6, or RP=0.5) is created, with randomly assigned storm strengths, then a time-series of occurrence following a random Poisson process is generated; the clustering coefficient is computed for each severity threshold, depending on the earlier designated severity assignments; this is repeated with 50,000 random sets of data, to form 50,000 Poisson samples of β vs RP. The empirical probability that the β of the observed storms is greater than the Poisson sample is recorded at each RP, and the probabilities at each RP are multiplied together to form a score corresponding to the likelihood that the observed β values are above that of a Poisson process. The likelihood score is computed for each of the 50,000 Poisson samples, and it is found that the observations exceed 99.6% of all Poisson samples. This finding suggests European storms with severity between RPs of 0.5 and 3 years are significantly different from a Poisson process at the 1% level.

Results in Fig. 6 suggest greater clustering for more severe storms, though the uncertainties are large. The question of whether there is an increase in the clustering for more severe storms is now addressed by analysing β gradients, as follows: compute the best linear fit between observed β and severity expressed as the logarithm of RP; fit Negative Binomial model parameters to observed time-series at RP=0.5 threshold; generate a random Negative Binomial sample and assign storm strength ranks randomly to it, then form subsets for each RP severity threshold (this is essentially the same method as above, except for a Negative Binomial rather than a Poisson); compute β vs RP for this random sample, then find the best fitting gradient of β vs log(RP); finally, repeat this 50,000 times to obtain a set of 50,000 gradients. It was found that the gradient of β versus severity in the observed storm set was more positive than 98.9% of all randomly generated samples. This leads to the conclusion that greater clustering with stronger storms at the Europe-scale is much more likely than not, though the fact that 1.1% of samples with randomly assigned severity relationships have a more positive gradient indicates some uncertainty in this finding.

The relationship between clustering strength and storm severity in previous studies is obscured by the rate dependency of the dispersion parameter described in Raschke (2015). However, some previous studies contain storm rate information which enables β to be derived from dispersion values, and these are now described. Figure 3 of Pinto et al. (2013) indicates higher β for more severe storms in North Atlantic and Europe from three different re-analyses products. Figure 6 of Vitolo et al. (2009) contains storm numbers as well as dispersion and conversion to β suggests a general upward trend of clustering strength with storm severity. Both observational studies are in general agreement with behaviour in the extended storm datasets analysed here, though the different measures of storm severity in the three studies confound their comparison. In contrast, Raschke (2015) finds a constant β is appropriate for RPs from 1 to 5 years, using storm occurrences from a modern coupled climate model simulation. The climate model data are described in Karremann et al. (2014a) and they employ a severity measure similar to that used in analysis of the long historical datasets. This suggests we cannot gain the benefits of smaller sampling errors from long integrations of the ECHAM5 climate model at the present time, due to its inability to simulate observed stronger clustering of more severe storms. Kvamsto et al. (2008) note significant differences in clustering between a different climate model and observations, though β versus storm severity is not analysed. These two studies suggest climate models have different clustering behaviour from observed, however, they represent a small sample, and analysis of more climate models is needed to make firmer, useful conclusions on climate models' quality of clustering simulations. Finally, it is worth noting how constant β with severity is explained by a model assuming independent storm events following an inhomogeneous Poisson process (Raschke, 2015). An alternative model is needed to explain increased β values for more severe storms found in historical storm datasets.

The clustering behaviour at national scales in the EU-RMS-42 dataset is now explored. Figure 7a displays β versus RP curves for some countries in the northern part of the European area shown in Figure 5, while Fig. 7b displays curves for some of the more southern countries. The large uncertainties in β values discussed above apply to national scales too. Thus the differences between northern countries in Fig. 7a lie well within the limits of error, and similarly for southern countries in Fig. 7b. However, comparison of Figs. 7a and b reveals a signal of stronger clustering for more severe storms in the southern part of the domain. The main driver of this north-south difference is the exceptional nature of the storminess in January to March 1990 in the southern countries. Figure 8 contains β versus RP curves for southern countries when the 1989/90 storm season is removed, and it can be seen how clustering strengths at RPs of 1 to 3 years are now much more similar between northern (Fig. 7a) and southern (Fig. 8) parts of the domain. This exemplifies the large sampling errors shown in Fig. 6: if this season had not occurred, the clustering strengths in more southern countries would be very different (Fig. 8 versus 7b). The conclusion is that sampling errors have a major impact when storm datasets are limited to the past few decades. Longer records help to reduce such large sampling errors and place 1989/90 into a fuller historical context. This is the motivation for analysing longer historical datasets.

Figure 9 contains results from an analysis of the longer storm datasets in the U.K. and Netherlands. Figure 7a indicates low values of β in the U.K. at all RPs, and a test of the hypothesis that the group of all data points are significantly different from a sample of Poisson data is rejected at the 0.1 significance level, in common with most northern countries. The results from extended U.K. storm datasets in Fig. 9a show β values of about 1.0 for storms with severities exceeding RPs of 5 years. The lengths of UK-Lamb-300 and UK-RMS-160 datasets, and their independent methods of gathering and assessing storm severities, combine to produce significantly smaller uncertainties than those shown in Fig. 6, raising confidence that more severe U.K. storms are clustered. Figure 9b shows low levels of clustering in the Netherlands from the NL-KNMI-105 storm dataset, which is consistent with analysis of EU-RMS-42 in NL. The raised clustering value at the RP of 6 years in NL-KNMI-105 is very uncertain due to limited sample sizes. However, similar behaviour in the longer and independent datasets in the neighbouring U.K. supports the raised clustering of storms above RP6 severity in NL-KNMI-105.

Figure 10 contains the clustering strengths found in four extended datasets in the southern part of the study area. Results in Fig. 10a indicate lower levels of clustering in DE-DWD-60 compared to the EU-RMS-42 dataset. The DWD clustering is more similar to the EU-RMS-42 dataset with 1989/90 removed. This may be due to greater weighting of far northern Germany in the DWD dataset (3 of the 7 stations), since the 1989/90 season was less extreme in this area, relative to local storm climate. The dotted lines in Fig. 10a represent β versus RP when one station is removed from DE-DWD-60, and show DWD clustering is not especially sensitive to any single weather station.

The results of analysing FR-Garnier-350 dataset are shown alongside those of EU-RMS-42 in Fig. 10b. The much longer storm dataset contains clear signs of clustering of the most severe storms in France. The independence of the information sources, and the increased length of the Garnier-Bessemoulin dataset, raises confidence in the conclusion of stronger clustering of more severe storms in France.

Figure 10c shows the results from analysing CZ-Brázdil-500 dataset in Czech Republic, alongside those from EU-RMS-42. The results from the shorter dataset showed great sensitivity to the inclusion of the 1989/90 storm season and an independent, longer dataset is very useful to help place 1989/90 in historical context. However, the reporting inhomogeneities in this long dataset (Sect. 2) are a source of significant uncertainty in results. Table 2 shows the clustering coefficient for class 1 storms for a range of different time periods in CZ-Brázdil-500, and Table 3 shows results for class 2 storms. Β varies substantially according to the time period studied, though a clear signal emerges of lower values at RP threshold of around 1 year, and significantly stronger clustering of more severe storms (RP threshold of around 10 years). Using the information in Fig. 3, the 1800 to 1999 period is chosen to represent clustering of class 1 storms and stronger, whereas 1700 to 1999 is chose to represent class 2 storms, in Fig. 10c. The main finding from this much longer dataset is weaker clustering around RP1 thresholds and notably stronger clustering of more severe storms. Further investigation of EU-RMS-42 at shorter RP thresholds reveals a 6 year period of elevated gust readings from about 1989 to 1995 suggesting inhomogeneous observation practices. This adds to the acute sensitivity of β to the inclusion of the 1989/90 season in the shorter dataset, as shown in Fig. 10c. The existence of significant observational errors in the most recent records of storms illustrates the benefits of analysing multiple, independent storm datasets.

Figure 10d contains the results from an analysis of Swiss storms. The extended CH-Stucki-153 dataset indicates weak clustering at shorter RPs, and slightly larger values at longer RPs, which supports the findings from EU-RMS-42. Β values are lower than in nearby France, Germany and Czech Republic around

RP1 to 3 thresholds. The most unique feature of Switzerland relative to these nearby countries is its much smaller spatial extent. This suggests a dependence of local β values on size of area studied, which is consistent with the lower dispersion values for narrower latitudinal barriers reported in Vitolo et al. (2009).

Results from all extended storm datasets are presented in Fig. 11. The results contain two main features. First, there is generally stronger clustering in southern countries: at shorter RPs, the Netherlands β values are generally below those of Germany, Czech Republic and Switzerland, while the U.K. values at longer RPs are generally lower than in France and the Czech Republic. This geographical variation is consistent with that found by comparing Figs. 7a and b, however, the signal is smaller in longer datasets. Given the varied nature and independence of these datasets, and their much longer records of storm history, there is some confidence that countries further from the main storm track in Europe experience stronger clustering of storms, though significant uncertainties in our clustering knowledge remain. The second notable aspect of results in Fig. 11 concerns the earlier finding of a strong sensitivity of β values in more southern countries to inclusion of the 1989/90 season (Figs. 7b and 8). The β values around RP1 to 3 year thresholds from the extended datasets are lower than those in Fig. 7b (with 1989/90) and closer to those in Fig. 8 (without 1989/90). This is a practical illustration of large impacts from sampling errors in datasets spanning a few recent decades: too much weight is placed on the big cluster in 1989/90 inflating β values, and longer-term records are needed to place the 1989/90 storm cluster in fuller historical context.

5 Summary

The clustering of extra-tropical cyclones in Europe has been investigated from the perspective of the re/insurance sector since they suffer the most material impacts from this phenomenon. Specifically, storms were gathered into groups according to exceedance of damage severity thresholds expressed as return periods (RP), and clustering on annual timescales was studied.

Perhaps the most notable characteristic of clustering is the unusually large uncertainties of estimates based on typical storm dataset lengths of a few decades, due to its dependence on storm count variance. This was found in previous research and has been explored in more detail in this study. Both the sampling and observational errors are large for estimates of clustering for any single group of storms.

Eight different storm datasets were gathered to reduce these large uncertainties. The mix of different information sources and storm severity measures reduce observational errors, and six of the datasets were more than 100 years in length and help reduce sampling errors. Quality control was applied to each dataset: the biggest issue with such long datasets is temporal inhomogeneity and the period of analysis was shortened for some datasets to improve this aspect. Finally, the inter-comparison of data with different units of storm severity (e.g. SSI, damage severity classes, fatalities) was made possible by expressing each dataset’s storm severities in units of local RP.

The evidence from all datasets strongly suggests that clustering increases with storm severity, for the range of severities analysed, from RP 0.5 up to about 20 years. The 42-year RMS storm database shows a distinction between northern areas with weaker clustering, to regions off the main storm track in central Europe and France with stronger clustering of severe storms. However, the removal of one very stormy season (1989/90) eliminates differences between the two regions. This epitomises the large sampling errors of clustering estimates based on a few decades of data. The longer datasets also contain signs of stronger clustering in countries off the main storm track, with notable years in history of multiple severe storms. Conversely, countries closer to the storm track show little signs of clustering of storms at RPs around one year, though three longer datasets in the U.K. and Netherlands indicate some clustering of storms at RPs longer than 5 years. While the differences between individual countries are less significant due to large uncertainties, there is evidence from multiple, diverse historical datasets for the difference between regions on and off the storm track. Finally, the comparison of clustering in Switzerland with larger neighbours indicates weaker clustering with smaller spatial scales of analysis, which is consistent with earlier published findings.

While the multiple datasets used in this study reduce uncertainties in estimates of severe storm clustering, there is plenty of scope for further reductions. Europe is relatively rich in historical documentation and expanded research into these archives would be very beneficial. Climate models have the capability to provide much smaller sampling errors via millennial-scale simulations, and it is hoped models with validated relations between clustering strength and storm severity will be available in the future.

Acknowledgements

The author greatly appreciates Prof. Brázdil, Peter Stucki and all their collaborators for making their storm data publicly available, and more generally to KNMI and DWD as well. The suggestions in the reviews by Mathias Raschke, Joaquim Pinto and an anonymous reviewer significantly improved this manuscript and were much appreciated. The study also benefitted from many discussions with RMS colleagues, particularly Robert Muir-Wood, Pete Holland, Laurent Marescot and Christos Mitas.

References

Barredo, J. I.: No upward trend in normalised windstorm losses in Europe: 1970-2008. Nat Hazards Earth Syst Sci 10: 97-104, 2010

Bessemoulin, P.: Les tempêtes en France, Annales des Mines, Réalités Industrielles, p. 9-14, 2002.

Bonazzi, A., Cusack, S., Mitas, C., and Jewson, S.: The spatial structure of European wind storms as characterized by bi-variate extreme-value copulas, Nat. Hazards Earth Syst. Sci., 12, 1769–1782, 2012.

Brázdil, R., Dobrovolny, P., Stekl, J., Kotyza, O., Valasek, H., Jaroslav, J., History of weather and climate in the Czech lands VI: Strong winds, Masaryk University, Brno, Czech Republic, 2004.

Cusack, S.: A 101 year record of windstorms in the Netherlands, Climatic Change, 116, 693-704, 2013.

Economou, T., Stephenson, D. B., Pinto, J. G., Shaffrey, L. C., and Zappa, G.: Serial clustering of extratropical cyclones in a multi-model ensemble of historical and future simulations, Q. J. Roy. Meteorol. Soc., doi: 10.1002/qj.2591, 2015.

Hunter, A., Stephenson, D. B., Economou, T., Holland, M., and Cook, I.: New perspectives on the collective risk of extratropical cyclones, Q. J. Roy. Meteorol. Soc., doi:10.1002/qj.2649, 2015

Kalnay, E., Kanamitsu, M., Kirtler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha ,S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetma, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, B. Am. Meteorol. Soc., 77, 437-471, 1996.

Karremann, M. K., Pinto, J. G., von Bomhard, P. J., and Klawa, M.: On the clustering of winter storm loss events over Germany, Nat. Hazards Earth Syst. Sci., 14, 2041–2052, 2014a.

Karremann, M. K., Pinto, J. G., Reyers, M., and Klawa, M.: Return periods of losses associated with European windstorm series in a changing climate, Environ. Res. Lett., 9, doi:10.1088/1748-9326/9/12/124016124016, 2014b.

Klawa, M., and Ulbrich U.: A model for the estimation of storm losses and the identification of severe winter storms in Germany, Nat. Hazards Earth Syst. Sci., 3, 725-732, 2003.

Kvamsto, N. G., Song, Y., Seierstad, I. A., Sorterberg, A., and Stephenson, D. B.: Clustering of cyclones in the ARPEGE general circulation model, Tellus A, 60, 547-556, 2008

Lamb, H., and Frydendahl, K., Historic storms of the North Sea, British Isles, and Northwest Europe, Cambridge University Press, Cambridge, 1991.

Mailier, P. J., Stephenson, D. B., Ferro, C. A. T., and Hodges, K. I.: Serial clustering of extratropical cyclones, Mon. Weather Rev., 134, 2224–2240, 2006.

Pinto, J. G., Bellenbaum, N., Karremann, M. K., and Della-Marta, P. M., Serial clustering of extratropical cyclones over the North Atlantic and Europe under recent and future climate conditions, J. Geophys. Res.-Atmos., 118, 12476-12485, 2013.

Pinto, J. G., Gomara, I., Masato, G., Dacre, H. F., Woollings, T. and Cabellero, R., 2014, Large-scale dynamics associated with clustering of extratropical cyclones affecting Western Europe, J. Geophys. Res.-Atmos., 119, 13704–13719, 2014

Raschke, M.: Statistical detection and modeling of the over-dispersion of winter storm occurrence, Nat. Hazards Earth Syst. Sci., 15, 1757-1761, 2015.

Stucki, P., Bronnimann, S., Martius, O., Welker, C., Imhof, M., von Wattenwyl, N., and Philipp, N.: A catalog of high-impact windstorms in Switzerland since 1859, Nat. Hazards Earth Syst. Sci., 14, p. 2867-2882, 2014.

Vitolo, R., D. B. Stephenson, D. B., Cook, I. M., and Mitchell-Wallace, K.: Serial clustering of intense European storms, Meteorol. Z., 18, 411–424, 2009.

Table 1: summary of storm datasets

Country	Source	Time period	Storm data type	Processing	Brief name*
Europe	RMS storm database	1972 to 2014	Measured wind speeds from 1000's of weather stations across EU	Compute SSI values for 142 major storms	EU-RMS-42
United Kingdom	Lamb and Frydendahl (1991)	1690 to 1989	Estimated storm severity index based on surface weather reports, meteorological analysis and documentary damage information	Restricted analysis to 1690-1989 due to incompleteness of reportage in earlier times.	UK-Lamb-300
United Kingdom	RMS (internal) fatality list	1835 to 1994	List of fatalities compiled from newspaper archives	Applied population factor to index fatalities to 1994 population, and night factor to scale night-time fatalities to daytime	UK-RMS-160
Netherlands	Cusack (2013)	1910 to 2014	Measured wind speeds based on Royal Netherlands Meteorological Institute (KNMI) data	Homogenized using KNMI metadata, and computed national SSI values for each storm	NL-KNMI-105
Germany	Deutsche Wetterdienst (DWD)	past 60 years	Publicly available peak gust data for climate stations, available on the DWD web site	Selected seven stations with minimal changes to wind observing system over time, then compute national SSI values for each storm	DE-DWD-60
Czech Republic	Brázdil et al. (2004)	1500 to 1999	Detailed damage descriptions	Assigned storms into two severity classes: (1) local or large-scale weak damage, (2) widespread intense damage	CZ-Brázdil-500
France	Emannuel Garnier (private communication) and Bessemoulin (2002)	1650 to 1999	List of storms with documentary descriptions of wind damage	For Bessemoulin dataset, we assigned a Beaufort scale severity based on documentary damage severity and spatial extent	FR-Garnier-350
Switzerland	Stucki et al. (2014)	1859 to 2011	List of storms, some with documentary descriptions	Damage severity taken directly from Stucki et al. dataset. Summer storms (May to September) are excluded.	CH-Stucki-153

* This brief name will be used in text to refer to each dataset

Table 2: β for class 1 storms in the Brazdil dataset, for various time periods.

	1500–1599	1600–1699	1700–1799	1800–1899	1900–1999	1800–1999	1500–1999
RP (years)	1.80	2.38	1.54	0.90	0.58	0.71	1.12
CC	1.19	0.27	-0.32	0.52	0.11	0.28	0.59

Table 3: as Table 2, for class 2 storms.

	1700–1849	1850–1999	1700–1999	1500–1999
RP (years)	7.89	10.00	8.82	11.09
CC	4.07	1.74	3.19	2.97

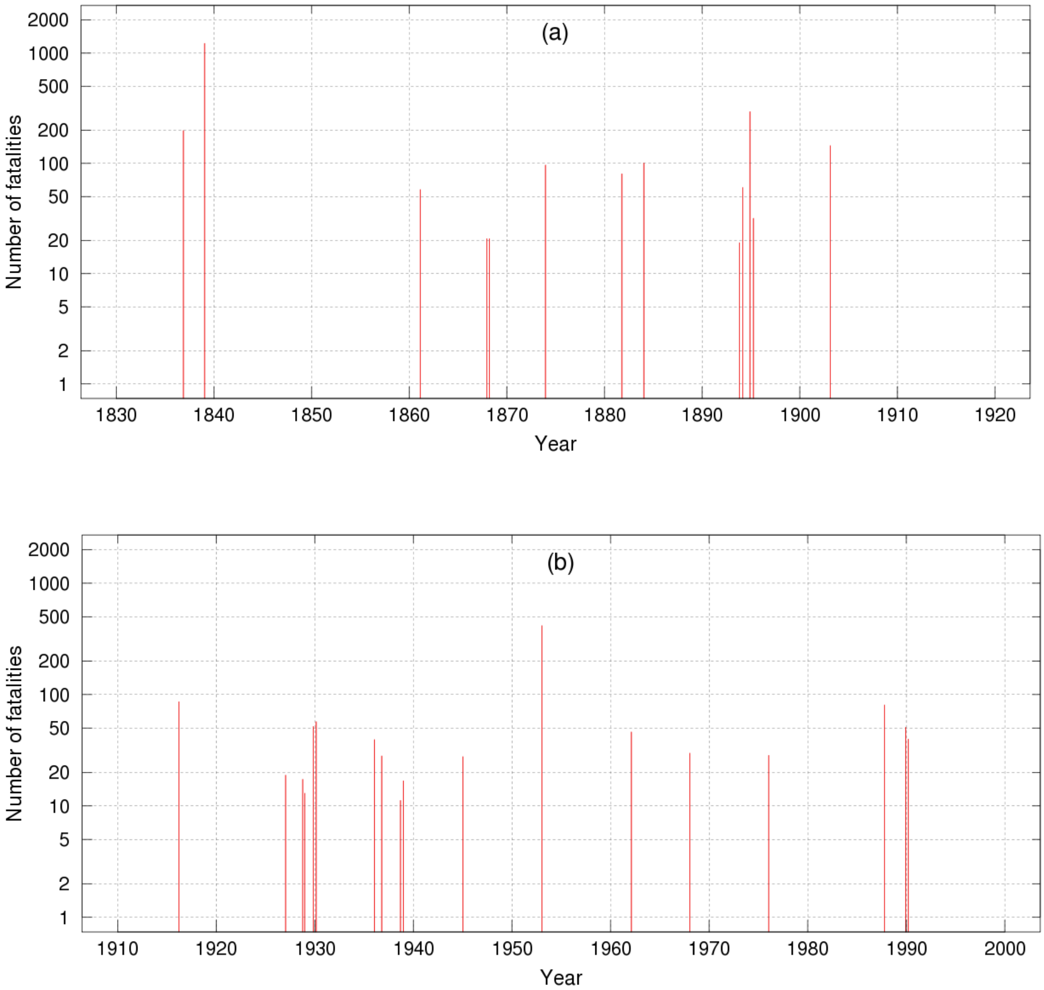

Figure 1: time-series of storm fatalities in the U.K. from the UK-RMS-160 dataset. All data are adjusted as-if storms occurred during daytime, and trended to 1994 population levels.

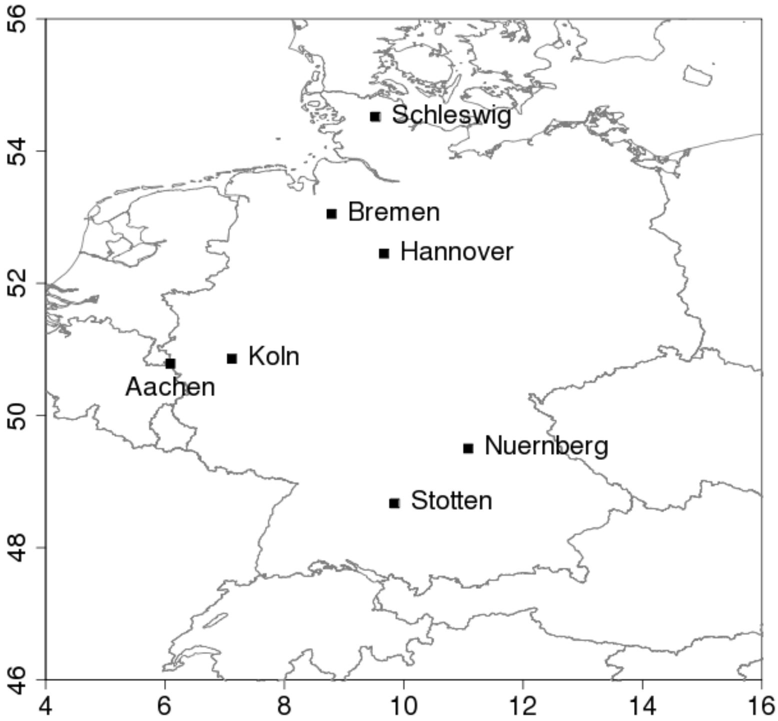

Figure 2: the location of the seven DWD weather stations in the DE-DWD-60 dataset.

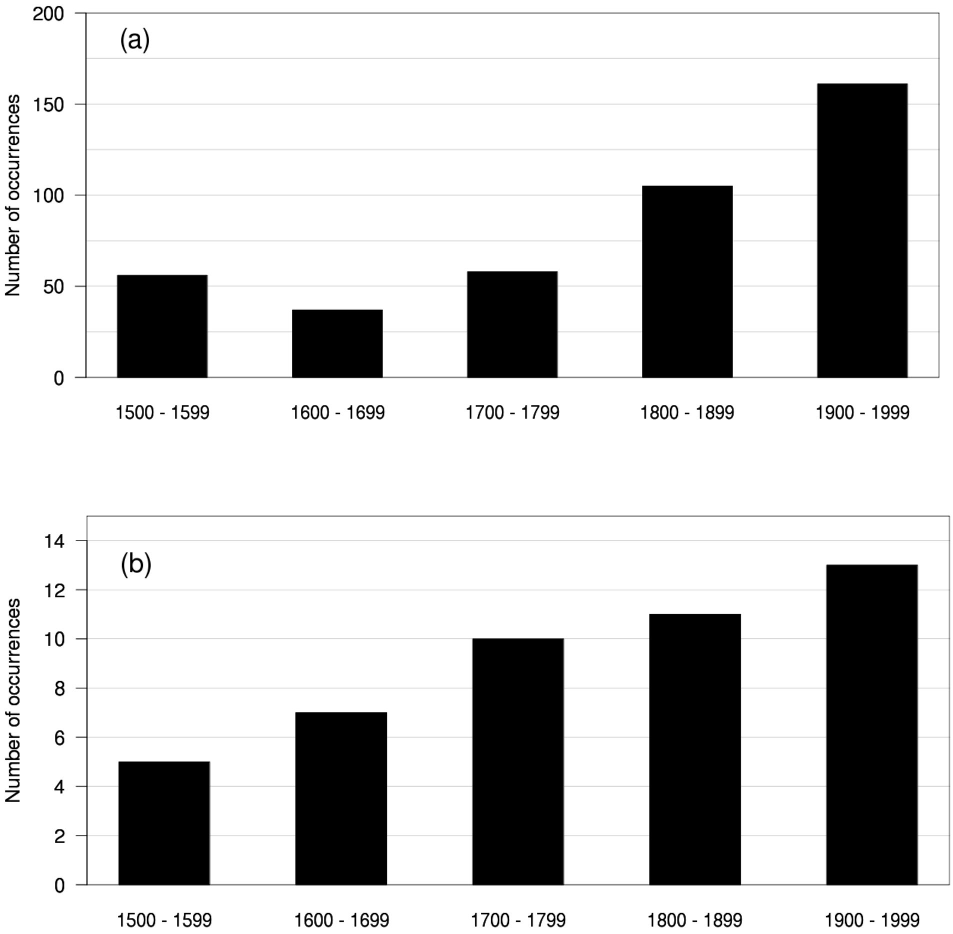

Figure 3: Histogram of storm occurrences per century in Czech Republic from the CZ-Brázdil-500 dataset for (a) weaker class 1, and (b) stronger class 2 storms.

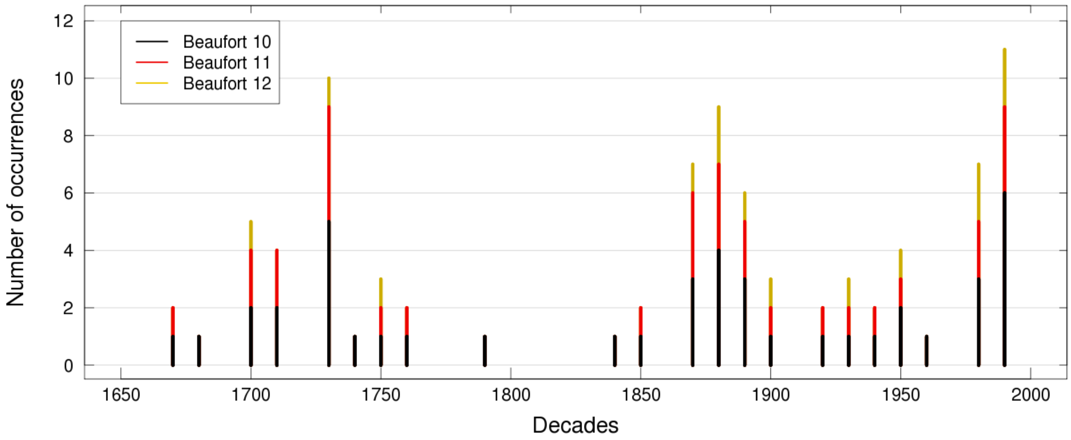

Figure 4: Count of storm occurrences per decade in France from the FR-Garnier-350 dataset, split into three damage severity categories.

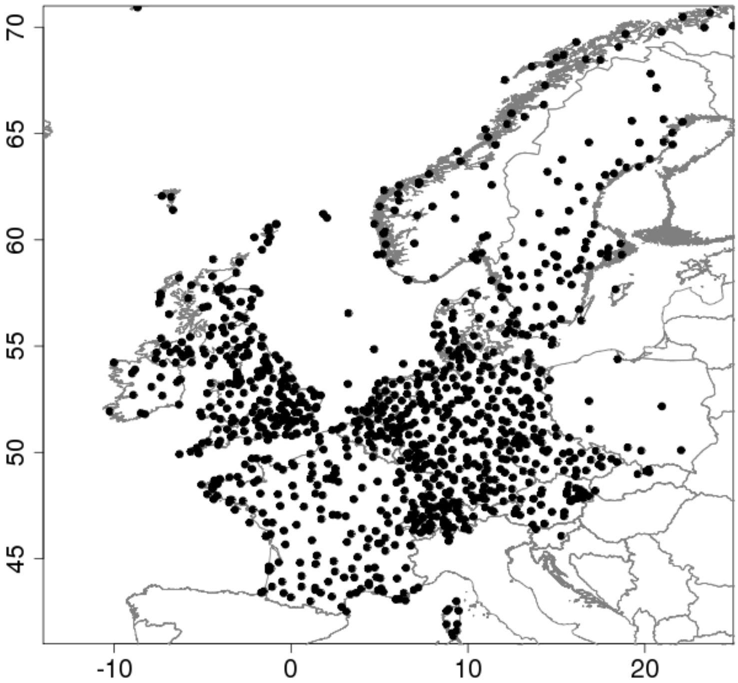

Figure 5: the location of weather stations with 15 or more years of peak gust measurements in the EU-RMS-42 dataset.

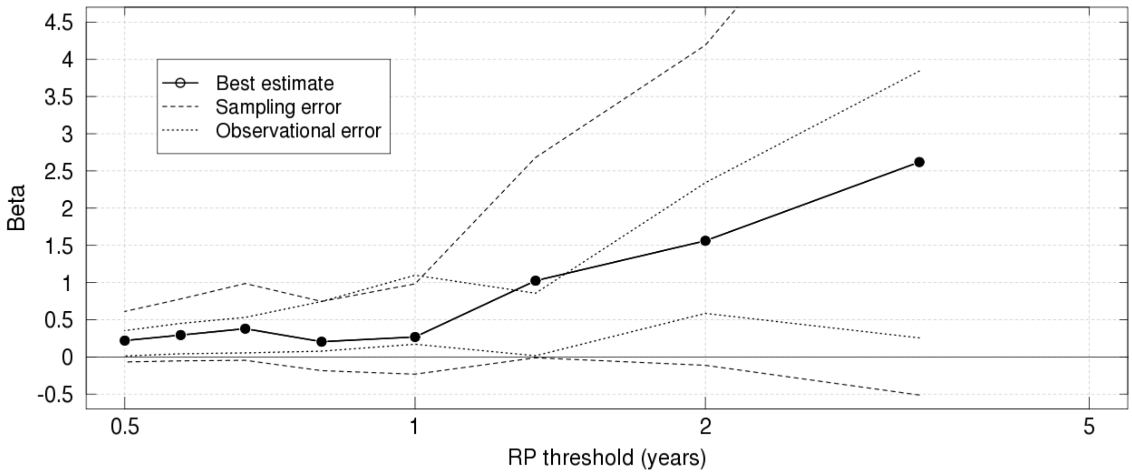

Figure 6: clustering strength (β) as a function of the storm severity groupings for historical storms in the EU-RMS-42 dataset. The dashed lines show the 95th confidence interval based on sampling error, and the dotted lines represent the 95th confidence interval of observational errors.

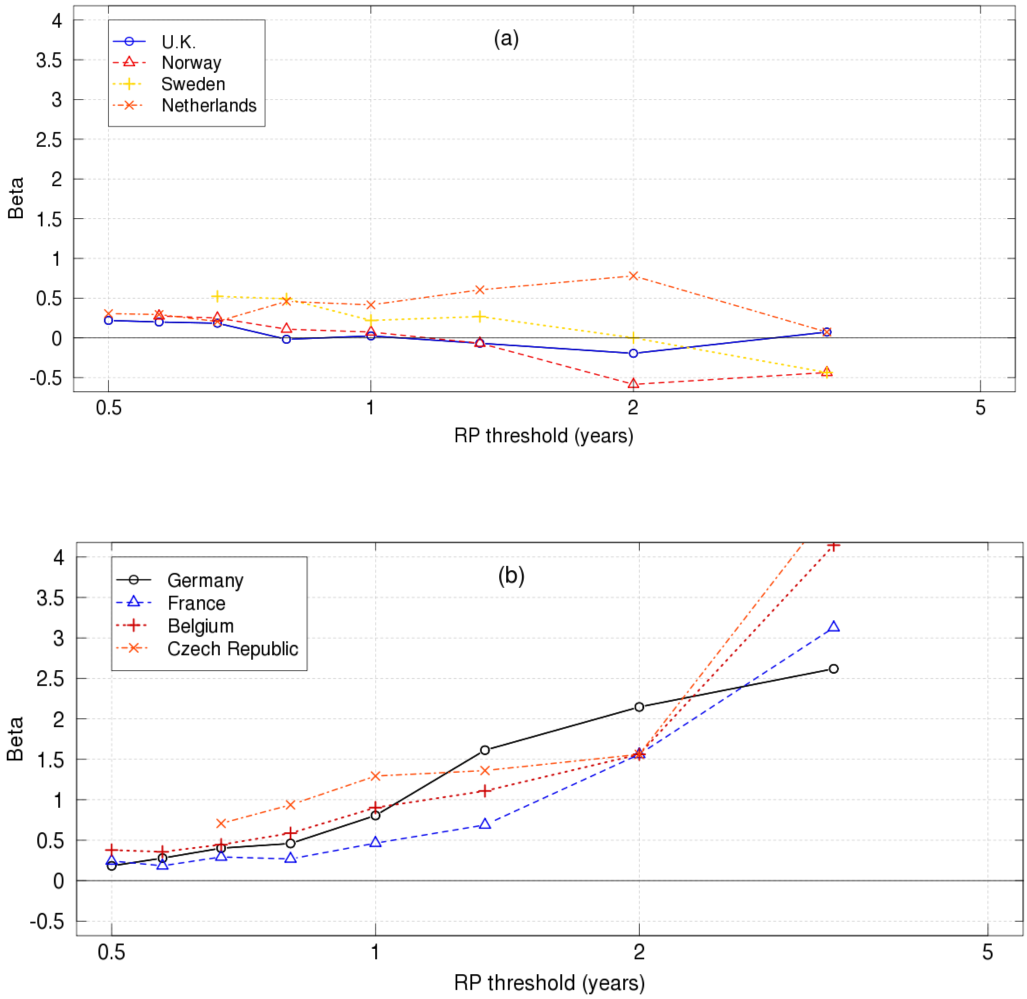

Figure 7: As Fig. 6, for various countries in (a) northern part and (b) southern part of the study area.

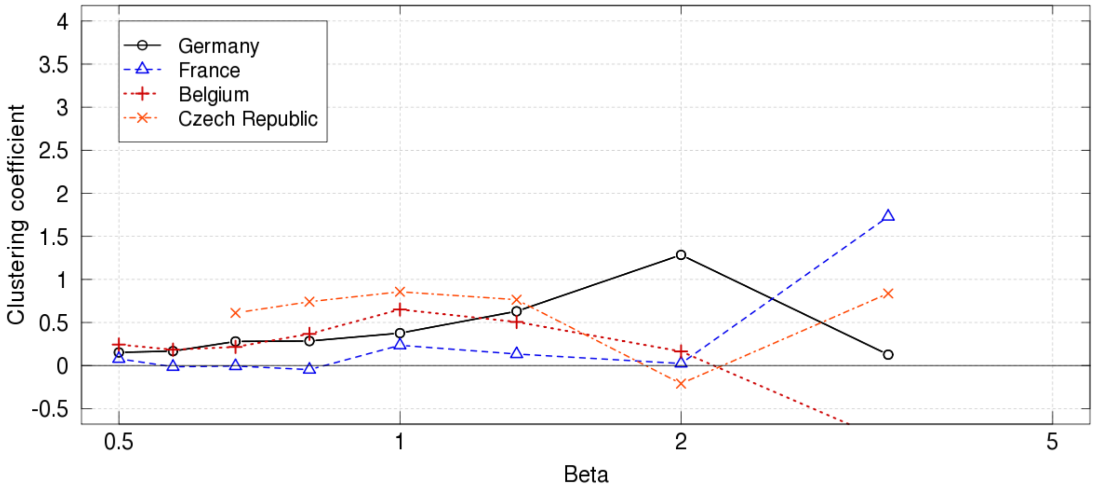

Figure 8: as Fig. 7b, with 1989/90 storm season excluded.

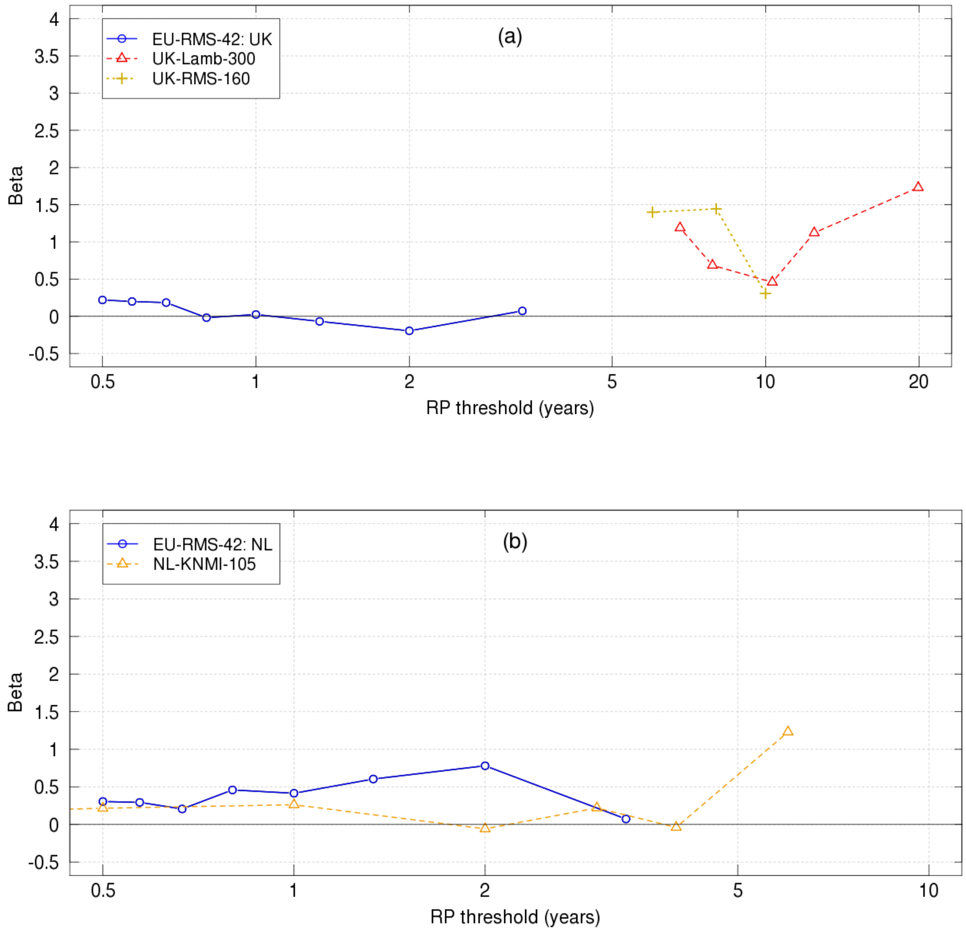

Figure 9: clustering strength (β) as a function of the storm severity groupings for (a) U.K. and (b) Netherlands.

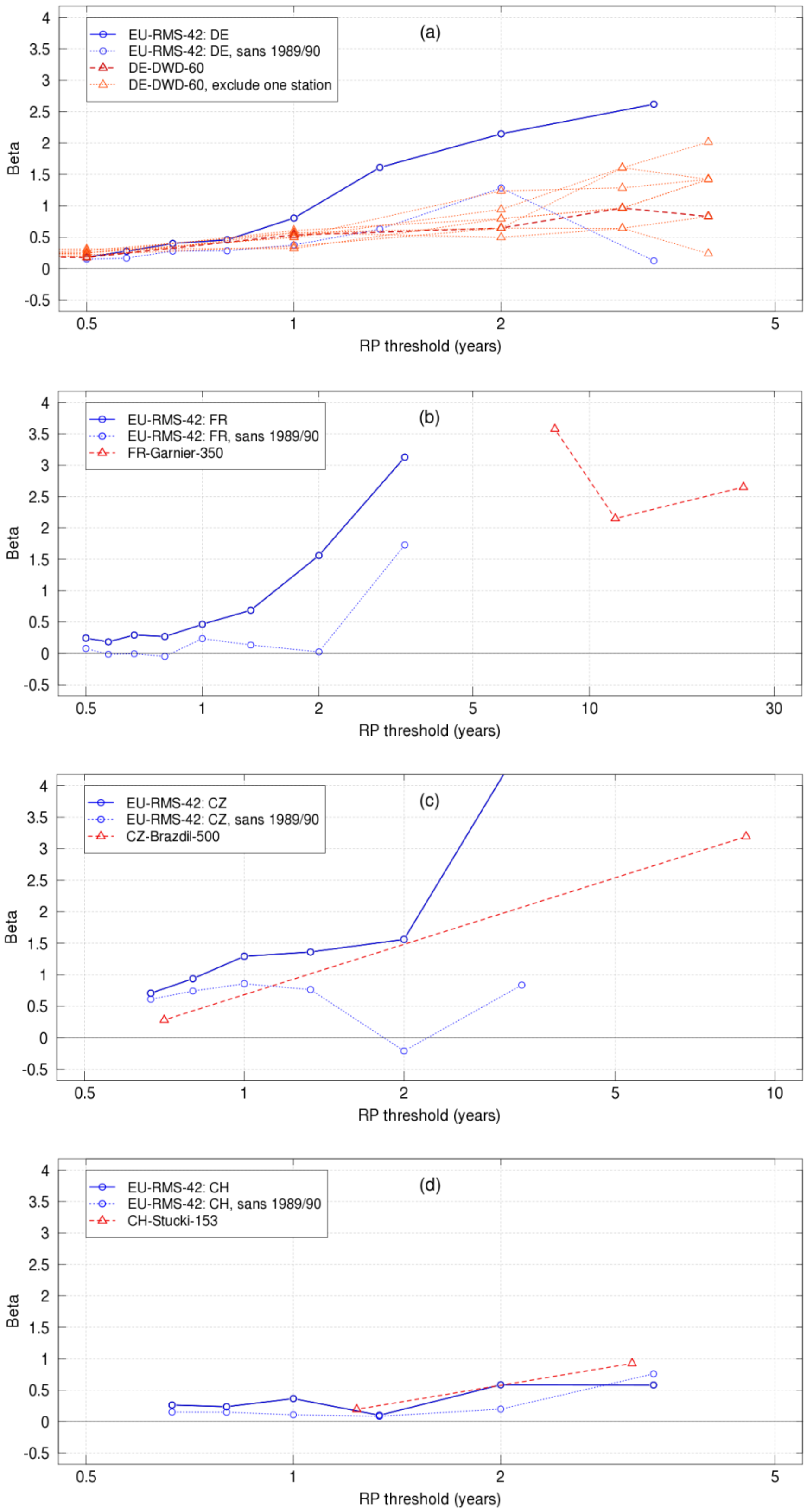

Figure 10: clustering strength (β) as a function of storm severity for (a) Germany, (b) France, (c) Czech Republic and (d) Switzerland.

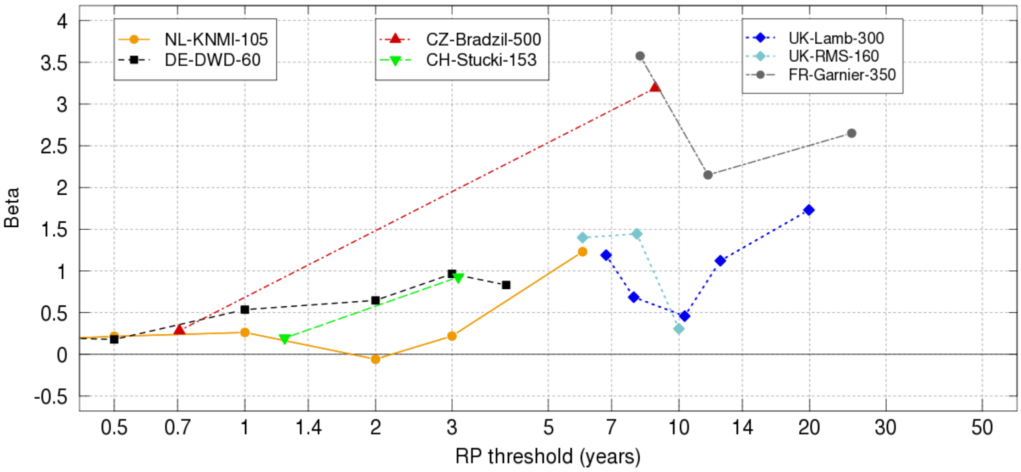

Figure 11: clustering strength (β) versus storm severity from extended historical storm datasets.