Can losses caused by wind storms be predicted from meteorological observations?

by
Holger Rootzén† and Nader Tajvidi‡
†Chalmers University of Technology
‡Linköping University and Australian National University

September 15, 1999

Abstract

This paper contains a study of the extent to which aggregate losses due to severe wind storms can be explained by wind measurements. The analysis is based on 12 years of data for a region, Skåne, in southern Sweden. A previous investigation indicated that wind measurements from six recording stations in Skåne was insufficient to obtain accurate prediction. The present study instead uses geostrophic winds calculated from pressure readings, at a regular grid of size 50 kilometers over Skåne. However, also this meteorological data set is seen to be insufficient for accurate prediction of insurance risk. The results indicate that currently popular methods of evaluating wind storm risks from meteorological data should not be used uncritically by insurers or reinsurers. Nevertheless, wind data does contain some information on insurance risks. There is a need for further research on how to use this information to improve risk assessment.

[0]

Keywords and phrases: Wind storm claims, meteorological prediction, geostrophic winds.

Research supported by Stiftelsen Länsförsäkringsbolagens forskningsfond.

1 Introduction

Wind storm insurance poses difficult problems for insurance companies and reinsurers. Significant losses caused by strong winds occur irregularly, with long periods in between, but when such losses do happen they can be very large. The magnitude of the problem can be seen e.g. from the Catastrophe Reinsurance (1993) list of the biggest aggregated claims from catastrophic events. The five largest losses were results of wind storms and the biggest one, the hurricane Andrew in August 1992, is listed as having caused an insured damage of about \$ 16×10^9. Also in Scandinavia the single largest aggregated claims so far have been caused by wind storms. Insurers are further concerned about the growth of insured value in wind exposed areas, and possible increases in wind storm frequencies and intensities which could be caused by global warming or by inherent instabilities in the wind climate, see [3].

Maximal wind speeds depend strongly on geographical location, and building codes, and the state of buildings may vary very much from one region to another, even within a single country. The problem is compounded by the facts that a given average (geostrophic) wind, depending on the terrain, may lead to very different local wind speeds, and that a change of a few degrees in wind direction may be the difference between very considerable damage and a totally unharmed building. Further, factors like temperature, precipitation and season may also influence the amount of damage.

The large insurance and reinsurance companies spend substantial effort on evaluating risks connected with wind storm insurance. Available loss experience often is insufficient. The obvious way to supplement it is with data from the meteorological offices. A currently popular approach is to use this data to build computer simulation models connecting loss ratios to wind velocities, with the connecting function estimated from a few historical events. Alternatively the wind data is just used for qualitative considerations.

The aim of the present paper is to investigate if risk connected with wind storm insurance can be accurately inferred from meteorological information. We use parts of a data base containing meteorological measurements and insured loss over a 12-year period for a province, Skåne, in southern Sweden. Our approach is to try to construct the best method we can for prediction of loss from wind measurements. The idea is that if the resulting predictor performs well it would show that it indeed can be possible to use meteorological information as a reliable basis for risk assessment.

On the other hand, if the prediction error contains substantial unexplained random variation, this would indicate that risk assessment based on wind data alone was not possible for Skåne. By extension, risk predictions based on meteorological data, but without support from extensive loss experience, then couldn't be trusted for other areas either, unless further information shows that the area, or the meteorological data is substantially different from the ones in the present study. A possible bias in this conclusion could be that a good predictor might exist although we were not able to find it. We have done our best to ensure that this hasn't happened.

Our background is a previous study, [4], of wind storm insurance which used the same loss data. As a part of the study we correlated losses with actual wind measurements from 6 recording stations in Skåne. The conclusions were that (i) the best predictor of the losses left substantial random variation unexplained, e.g one standard deviation of the prediction error corresponded to to a factor of about 5 up and a factor of .2 down, (ii) there were storms which had higher wind speeds in all recording stations but which caused substantially less damage than a corresponding storm with lower wind speeds in all recording stations, and (iii) the 12-year time period studied showed no discernible trend in the sizes of the aggregated losses after correction for inflation. There was nevertheless some indication of a minor increase in the average size of small claims.

Further, the main conclusion of [4] was that wind storm insurance always includes an element of gambling. The best that can be hoped for is that the odds in this gamble can be better understood: the randomness in storm occurrence is inherent in nature. Thus, in this perspective, the present investigation is aimed at studying what meteorological measurements can and cannot teach us about these odds.

These results from [4] indicated that it may be difficult to predict losses from wind data. However, wind speeds recorded at a specific measuring station are strongly influenced by the local topography. In addition, six recording stations are rather few for covering Skåne effectively (the original data contained more recording stations, but we kept only those which didn't have any missing observations). Furthermore, it seemed worthwhile to investigate if including more covariates could improve prediction.

In this paper we instead study geostrophic wind speeds computed from air pressure readings. These wind speeds, computed at a regular grid of size 50 kilometers over Skåne are also included in the data base. They have the advantages of not being influenced by a local topography, may actually be more representative of the weather than wind recordings from local stations, and the finer grid used makes it less likely for storms to pass in between the grid points. We further investigated if prediction could be improved by taking wind direction, the length of storm events and the time of the year into account.

The database is described more in detail in Section 2, and the method used to analyze the data is introduced and motivated in Section 3. Section 4 summarizes the results and Section 5 contains a discussion of the results and our conclusion.

2 The data

The loss database was put at our disposal by the Swedish insurance group Länsförsäkringar. It covers the period 1982 to 1993 and contains the individual amounts of wind storm claims, the place and time of the claims, and the type of the claim. Approximately 65% of the total amount claimed stems from farm insurance.

To obtain as homogeneous data as possible, we only consider claims from farm insurance. In addition we restrict attention to Skåne. Skåne is an important farming area, it contains much open terrain, and 43% of the total claims from farm insurance in the windstorm loss data come from it. All claims were corrected for inflation, but since the portfolio was relatively stable, we made no adjustments for portfolio changes.

The wind storm data base contains 78 storms and was provided by the Swedish Meteorological and Hydrological Institute. The criteria for inclusion are described in Rootzén and Tajvidi (1997, p. 87, 88). However, 6 of these storms were considered parts of other storms in the loss database, and were then merged or deleted. Thus the resulting basic data consists of 72 storm events.

The part of the database used in this paper is the geostrophic winds and wind directions at equally spaced grid points 50 kilometers apart (Fig. 1), calculated from air pressure measurements, and normalized to a height of 10 meters over flat terrain with a roughness parameter of 5 cm.

The geostrophic winds were not available for 14 of the original storm events. For 12 of these the reason was that they had not been selected by the original objective storm criteria, but because the events had caused losses in excess of .9 MSEK for all of Sweden. The remaining 2 events consisted of one storm which occurred after March 1993, which was the endpoint for the geostrophic wind data, and one storm where the geostrophic winds were absent for reasons unknown to us.

The statistical analysis reported on below was performed on the remaining 58 storm events. For each of them and each of the grid points we computed the maximum wind speed and a main wind direction. Much of the analysis

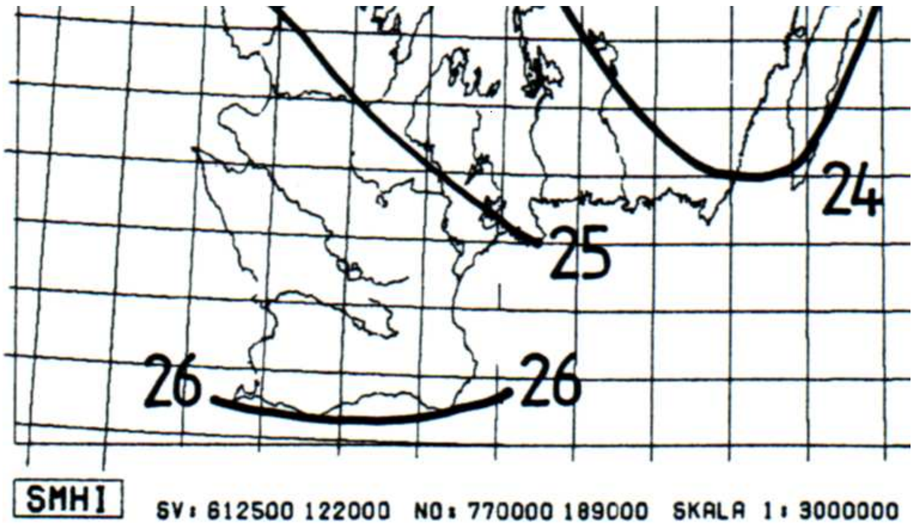

Figure 1: Grid points for geostrophic wind calculation

used wind pressure, which was taken to be proportional to the square of the wind speed. Deviations from this proportionality caused by variations in air density were not taken into account. Further, we often used the logarithms of the aggregated claims, since previous experience indicated that these may be simpler to analyze.

3 Statistical methods

We first made scatter plots of the maximal wind speeds at the different grid points, as a preliminary check on whether the grid was dense enough to catch all storms. We also performed a number of preliminary checks on the distribution of the claims and their relation to the wind data, and a number of alternative analyses of the entire data set. Most of these are not reported below.

Rootzén and Tajvidi (1997, p. 87, 88) used the simple model that the logarithm of the aggregate claim caused by a wind storm was a linear function of the wind at the six recording stations, plus random noise. As a first model in the present paper we used the corresponding approach, with the logarithm of insured loss a linear function of the wind pressures at the recording stations. We tried both the assumption that the losses had a lognormal distribution and that they followed a Generalized Pareto (GP) distribution, and used a downwards stepwise regression analysis to reduce the number of grid points in the final prediction equation. The same analyses were also performed with the storm events divided up according to season of the year, according to main wind direction, and with storm duration taken into account. Further, since stronger winds ought to lead to more damage, we also made the analyses under the restriction that the regression coefficients are non-negative.

In the introduction (cf. also [4]) we discussed the currently popular approach of using computer simulation models for risk prediction. This approach corresponds to a second model which was derived from the following reasoning:

If a building is exposed to a specific wind pressure (from a certain direction), the "average cost" of the damage caused is a (highly nonlinear) function of the wind pressure. In principle, it is possible to compute the "average maximum wind pressure" for each insured building using knowledge of the local topography and the geostrophic winds. One possibility would be to use the geostrophic wind at the nearest grid point as a basis. This "average" maximum wind pressure for a building would be a (nonlinear) function of the maximal wind at the grid point, and the "average" claim for the building would be obtained as a composition of two functions. The aggregate claim for all buildings nearest to a grid point is the sum of the individual damages and hence also is a, possibly very complicated, nonlinear ("connecting") function of the wind at the grid point. Finally, the total damage in Skåne would then be obtained as a sum of these nonlinear connecting functions corresponding to the different grid points.

In summary, in the second model, damage is determined as a sum of nonlinear functions of the maximal wind pressures at the grid points. The statistical analysis of this model was performed in two steps. In the first, scatter plots of aggregated claim amount against wind pressure were made separately for each grid point, and used to determine a suitable form of the connecting nonlinear functions. In the second step, the forms of the connecting functions were taken as known up to a few unknown parameters which were fitted by maximum likelihood.

4 Results

The sum of the losses due to the 12 storm events which were not selected by the objective storm selection criteria was 13.9 MSEK. This was, e.g., only 12% of the largest single aggregate claim (119.3 MSEK) However, still the 7-th, 8-th, 10-th and 11-th largest of the 58 losses were not included in the wind storms chosen by the criterion.

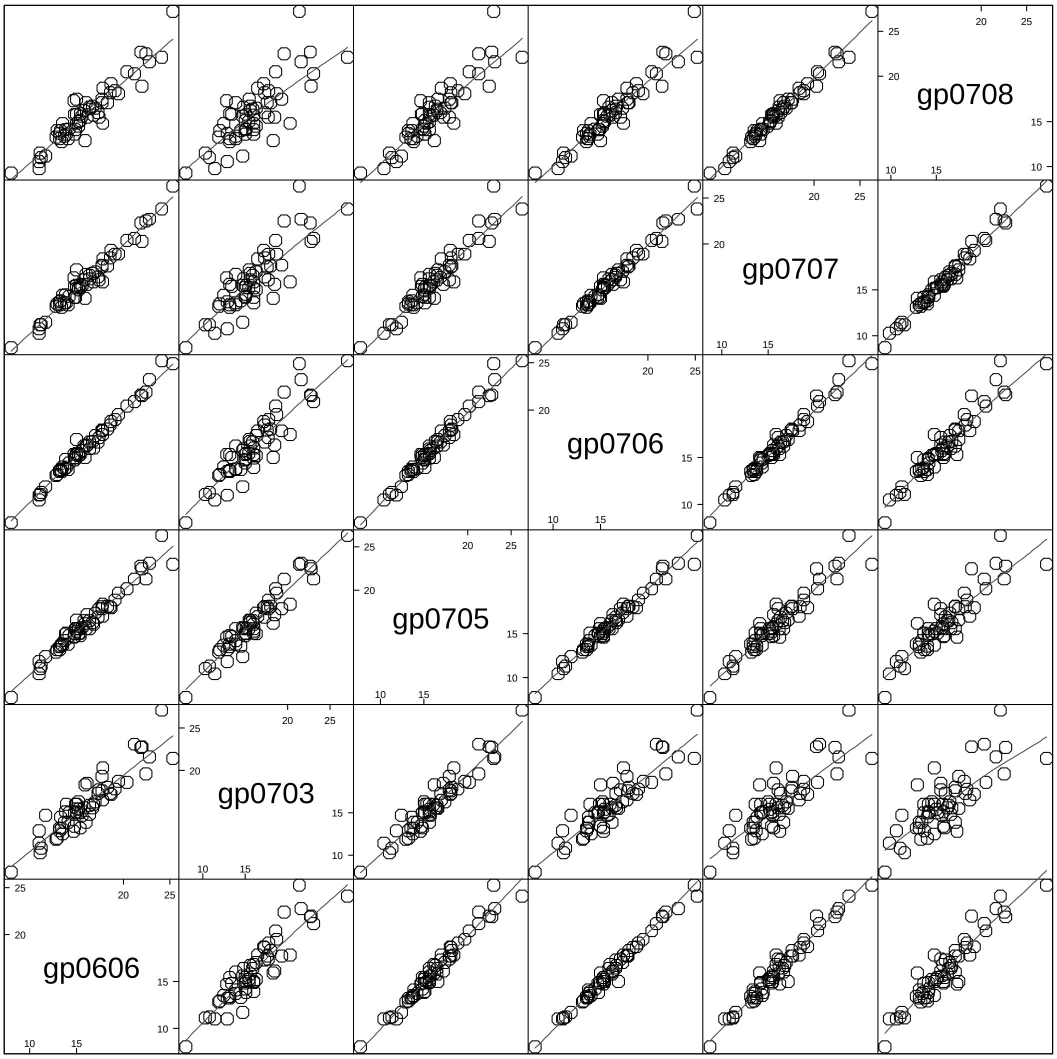

Figure 2: Scatter-plots of wind-speeds at some of the grid points. Nonparametric regression lines calculated using Loess

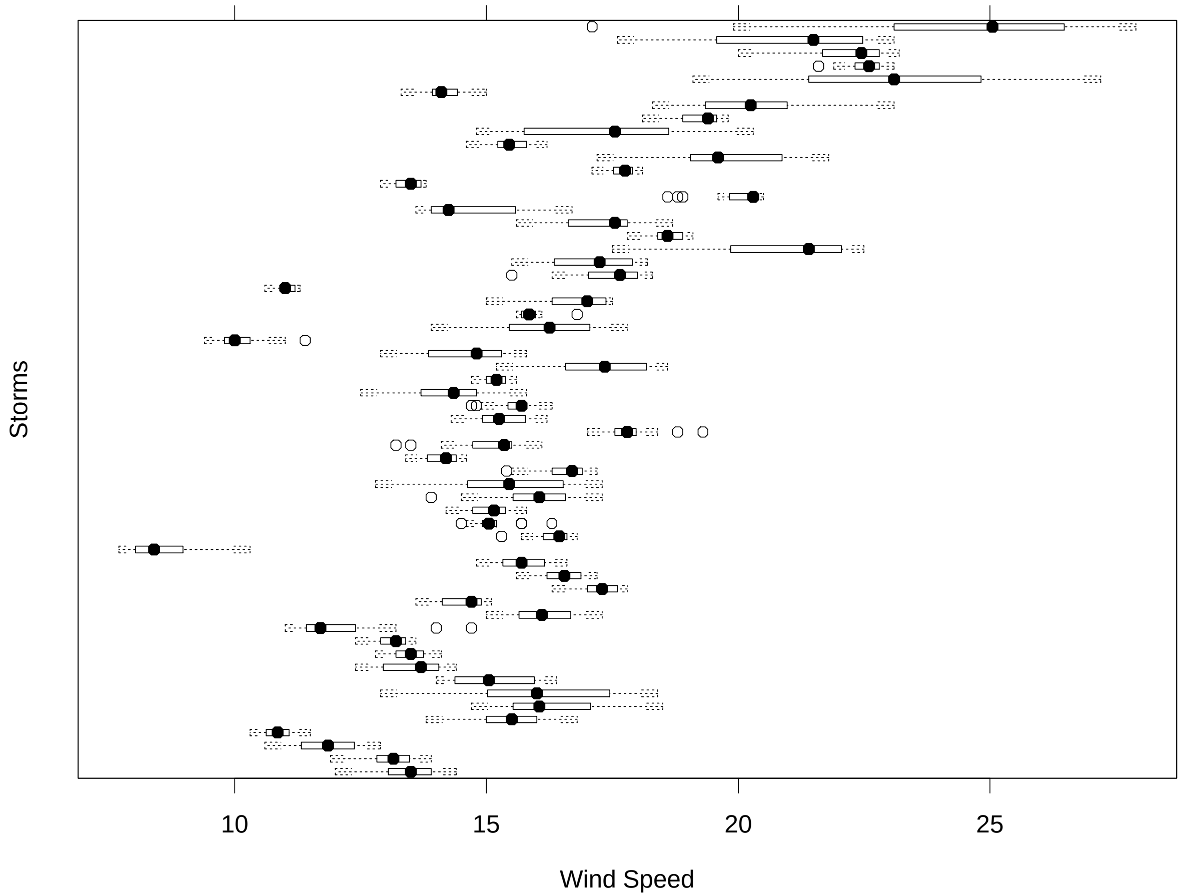

Figure 3: Maximal windspeeds at the 18 gridpoints. One boxplot for each storm event. The plots are ordered after size of the insured loss, with the largest claims on top. (The appendix, Table 1, lists storms and insured losses.)

Figure 2 contains scatter plots of the maximal wind speeds at some of the grid points. The plots for the remaining points were similar. For each grid point there are other points with very similar wind speeds, but still the correlation isn't perfect between any two grid points. This indicates that probably not much would be gained by having a more dense grid, but also that a substantial coarsening of the grid might lead to some loss of information. Figure 3 shows boxplots of the wind speeds from the 18 grid points for each of the 58 storm events. The storms are ordered after the size of the insured loss, with the largest on top. It can e.g. be seen that all the maximal wind speeds at the individual grid points for a storm with a loss of MSEK 6.2 are lower than the corresponding maximal wind speeds for another storm which had an insured loss of .4 MSEK (storm events 8905 and 9102 in the appendix). Hence substantially weaker winds caused about 15 times more damage than the stronger winds. There are also may other such "reversals".

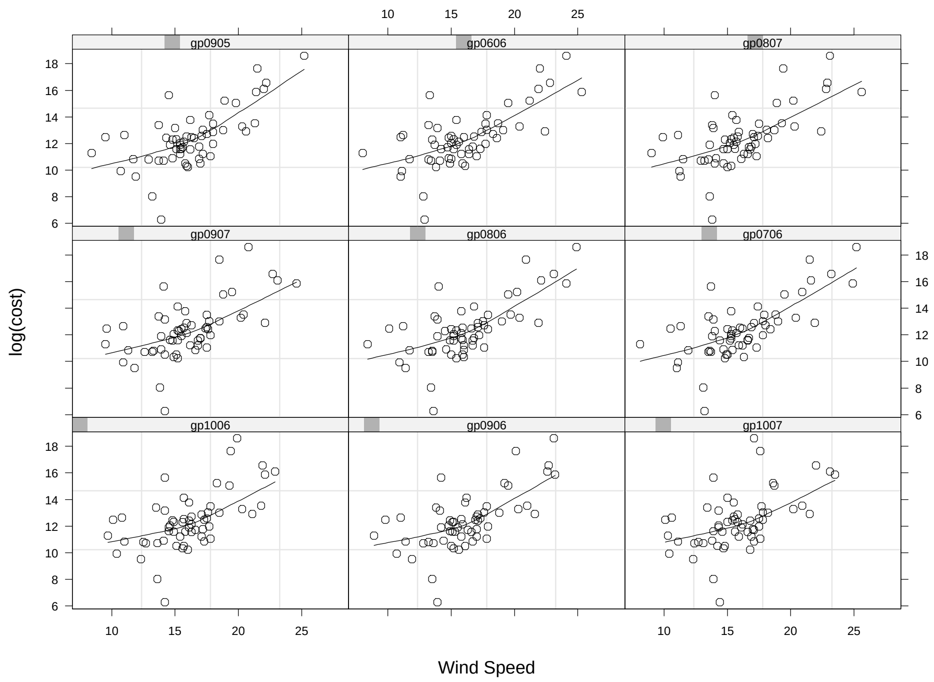

Figure 4: Logarithm of insured loss plotted against maximal wind pressure for grid points 10-18, one for each storm event. Nonparametric regression lines fitted by Loess.

Figures 4, 5 contain scatter plots of the logarithm of loss against wind pressure, for each of the grid points, together with nonparametric regression lines obtained by loess. In the plots, the gridpoints are ordered after the size of the maximal wind speed, starting with the smallest at bottom left in Figure 4 and the largest at top right in Figure 5. The plots suggest that the relation between logarithm of losses and wind pressure is linear, and also show that the data contains substantial random deviations from the regression lines.

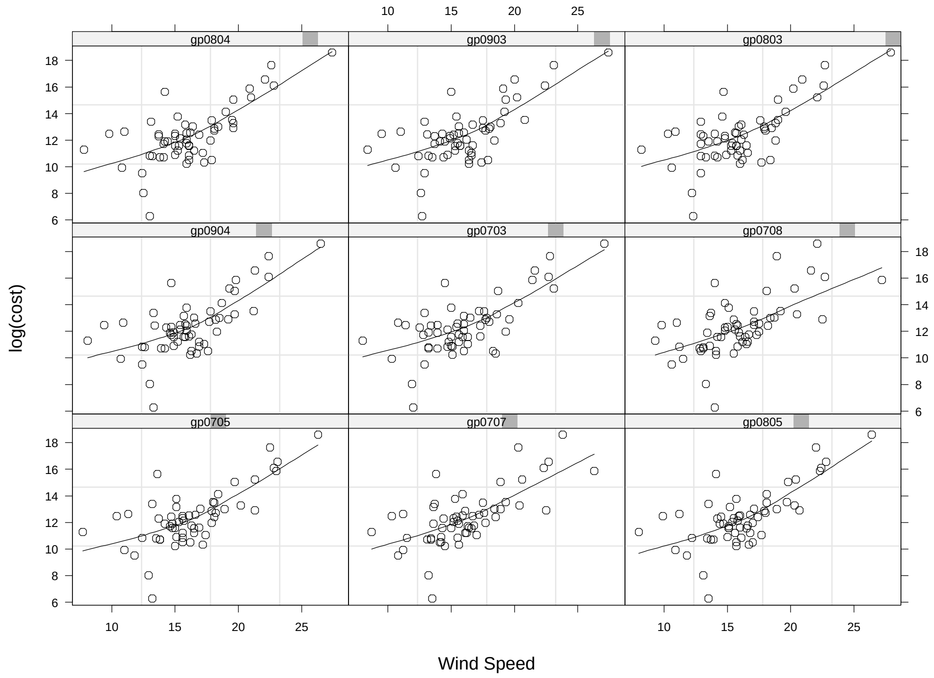

Figure 5: Logarithm of insured loss plotted against maximal wind pressure for grid points 10-18, one for each storm event. Nonparametric regression lines fitted by Loess.

The result of fitting the first model from Section 3, i.e. a stepwise multiple linear regression of the logarithm of loss, assumed to be normal, on the maximal wind pressure is illustrated in Figure 6. The fit is rather good. Nevertheless, the residual standard deviation, 1.4, corresponds to a factor 4 up or .25 down in the actual amounts. There is some indication of a better fit to the regression line in the important extreme right region of the plot. However, also there the spread is rather large.

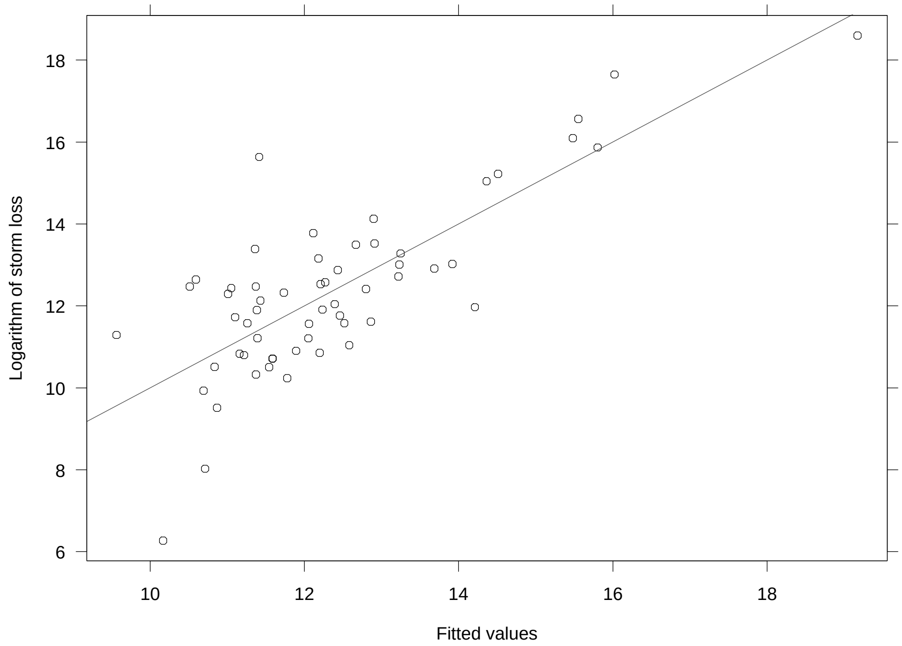

Figure 6: Stepwise multiple regression of logarithm of insured loss on maximal wind pressure in the grid points.

Three of the seven regression coefficients were negative. When the coefficients in the final model were forced to be non-negative the residual standard deviation increased to 1.5. It was interesting to note that then only two of the seven coefficients were non-zero, and that the original value for one of the two had been very close to zero.

Figures 7, 8 show the result of applying the same computations separately to storms occurring during the winter months and storms occurring during the remaining months. For ease of comparison we used the same grid points as in Figure 6. The fit for the winter months was somewhat better, with a residual standard deviation of 1.2, while the residual standard deviation for the non-winter months was 1.4. In particular for the winter months, negative regression coefficients were influential. If the winter regression coefficients were restricted to be non-negative, the residual standard deviation increased to 1.3.

Corresponding computations using a GP distribution lead to quite similar results. QQ-plots indicated satisfactory fit for both lognormal and GP distributions. Analyses taking wind direction and/or length of storm into account didn't lead to any interesting results.

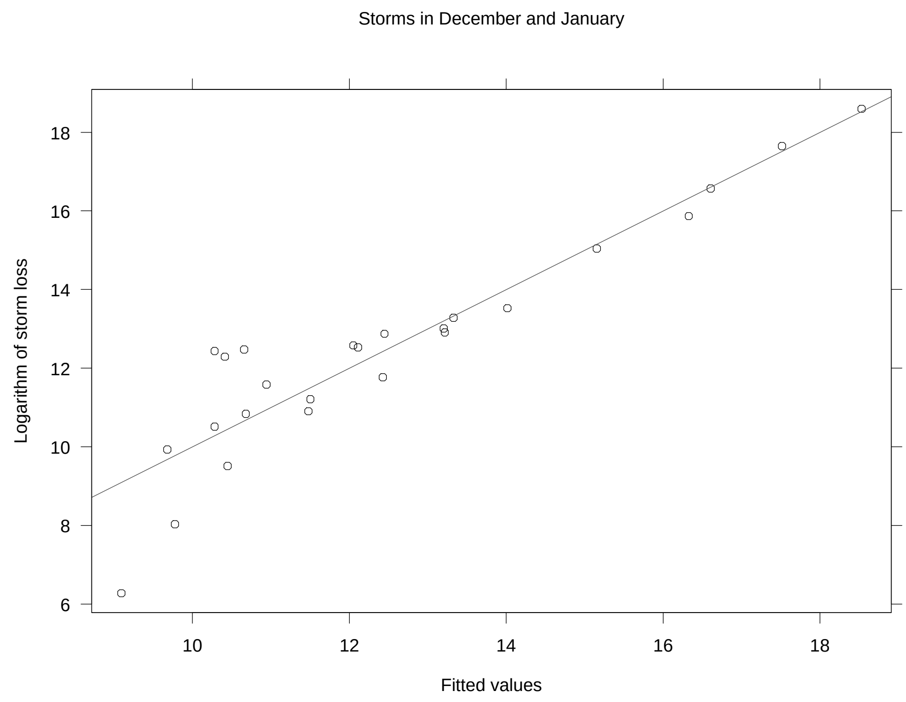

Figure 7: Stepwise multiple regression of logarithm of insured loss on maximal wind pressure in the grid points for storms occurring during December and January.

As motived by Figures 4, 5 (cf. the discussion above), for the second model we assumed that the connecting function was the exponential of a linear function. We thus used a GP model with the same shape parameter γ for all observations , and with scale parameter $\sigma = \sum_i a_i e^{b_i p_i}$, where p_i is the wind pressure at grid point i and a_i, b_i are unknown parameters. Since σ is a scale parameter, $\log \sigma$ is a location parameter, and the model is additive on the "natural" log scale, with "errors" following a log GP distribution. The Maximum Likelihood procedure only converged if six or fewer grid points were included. The parameter estimates of course were highly correlated.

Figure 9 shows the result for a suitably choosen subset of six gridpoints. The "regression line" corresponds to the predicted means. The observed residual standard deviation in Figure 9 (on log scale) was 1.46. The same quantity

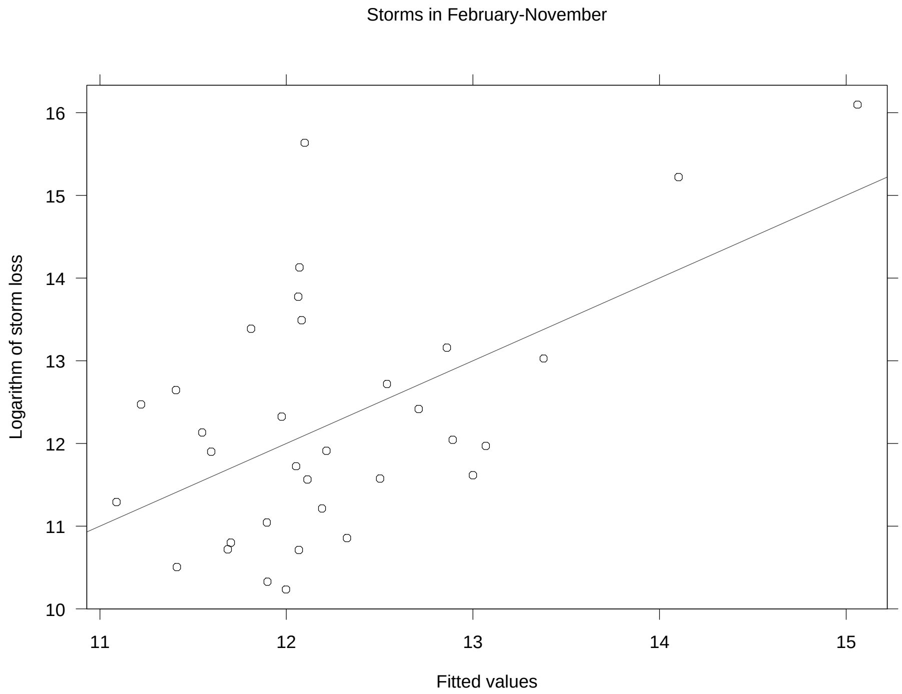

Figure 8: Stepwise multiple regression of logarithm of insured loss on maximal wind pressure in the grid points for storms occurring during February - November.

can be estimated by computing the standard deviation of $\log X$, where X has a standard GP-distribution (distribution function $1 - (1 + \gamma x)_+^{-1/\gamma}$) and inserting the estimated value of γ. This lead to the value 1.49. Again QQ-plots indicated good fit, also if the error distribution instead was assumed to be lognormal.

5 Discussion and conclusions

12 of the storm events in the loss data base were not picked up by the selection criteria based on wind speeds only. The total loss caused by these storms was small compared to the largest aggregate claim, and compared to the total amount claimed. However, nevertheless e.g. the 7-th and 8-th largest aggregate claims belonged to those which were not selected, and it is possible that this has affected the fit of the models considerably. We believe that

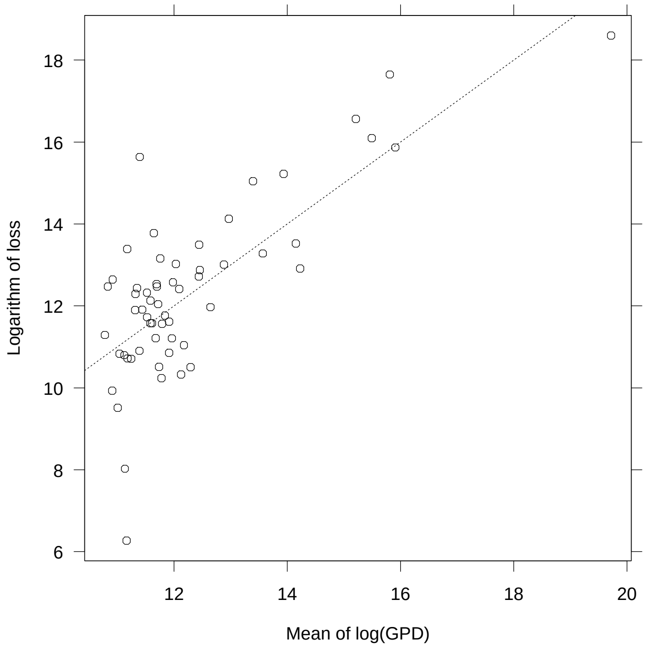

Figure 9: Nonlinear regression of loss using windpressure from 6 selected grid-points and the GP-distribution (log scale).

similar problems are likely to occur for other data sets too — it is difficult to find meteorological selection criteria which correspond well to the size of the economic damage caused by a wind storm.

The best predictor we could find still left random variation corresponding to a factor of 4 up or .25 down in the loss amounts unexplained. A similar result was obtained in Rootzen and Tajvidi (1997) for predictions based on measured, instead of computed, wind pressure. No improvement of the fit was obtained if the length of the storm or wind direction was taken into consideration or if wind speed was used instead of wind pressure. Some improvement may be possible by treating summer and winter storms separately. Conceivable explanations for this might be different conditions of the ground and/or the presence of leaves on the trees. However, we are convinced that it is not possible to obtain a substantially better predictor using the present data base. One indication of this has already been dicussed in Section 4: there were storm events where substantially weaker winds caused about 15 times more damage than stronger winds, cf. Figure 3.

The non-linear model didn't lead to better predictions. This may point to a need for careful statistical evaluation of the often used computer simulation models.

Some improvement of prediction might result if the data had included detailed information about the amount of precipitation during the storm. However, we still do not believe this would change the general conclusion, i.e. that for the situation studied in this paper it is not possible to make accurate predictions of insured loss from available meteorological data.

A further limitation in the present data is that the measurements were spaced 3 hours apart. This might be enough for the peak of a storm to pass without being registered. The future, with automatic computerized recording opens the possibility of better temporal resolution. Concievably this could lead to better predictions of loss. From the present data it isn't possible to evaluate if this actually would happen. Similarly, better spatial resolution of wind measurements might also improve predictions. However, in view of the costs involved, this is less likely to happen.

As discussed in the introduction, this leads to the conclusion that also for other areas and other data sets it is unwise to rely on a risk predictions based on meteorological data, unless they have been extensively validated against loss data. In fact, one might suspect that it very seldom is possible to predict insured losses caused by wind storms from available meteorological data very accurately. However, whether this in fact is correct, of course only can be established by further studies.

Finally, the idea behind using meteorological data for windstorm risk assesment, as discussed above, is the hope that there is a close ("deterministic") relation between measured winds and the size of the losses. If this is true, one could catch all of the randomness in wind storm losses by using a few storm events to determine this relation, and then confidently use long meteorological records of high quality to "find the odds in the wind storm insurance gamble". For our data base this hope wasn't substantiated. Still meteorological data contains some information about the risks. At present, however, methods to use this information efficiently have not been developed. To do this seems an interesting and potentially useful area for further research.

Acknowledgment: We are grateful to Håkan Pramsten for initiating the present study, for providing us with the wind storm loss data base and helping us to use it, and for many stimulating and helpful discussions and comments, which has lead to numerous improvements. We also want to thank Roger Taesler and Roland Kriek from the Swedish Meteorological and Hydrological Institute for providing us with the meteorological data, and for helpful discussions.

References

[1] Catastrophe Reinsurance Newsletter (1993). No. 8, 1-2.

[2] Geographic Analysis Project, Greig Fester (1995). A windstorm study for Länsförsäkringsbolagens AB.

[3] *Munich Reinsurance company.* Windstorms – new loss dimensions of a natural hazard.

[4] Rootzén, H. and Tajvidi, N. (1997) Extreme value statistics and wind storm losses: a case study. *Scand. Actuarial J.* No. 1, 70 - 94.

Appendix

Table 1: Storm events, numbered consecutively within years, ordered after size of insured loss (MSEK).

storm no	loss	storm no	loss	storm no	loss	storm no	loss	storm no	loss
8902	.001	8908	.054	9106	.158	8607	.390	8905	6.173
8808	.003	8703	.063	8801	.170	9102	.404	8402	7.781
9202	.013	8901	.074	9004	.186	8205	.445	9003	9.791
8807	.020	8603	.074	9302	.218	8904	.454	9002	15.659
8406	.028	8805	.080	8903	.225	8802	.519	8301	46.188
8803	.031	9204	.105	8702	.246	9303	.586	9301	119.299
9005	.037	9205	.106	8606	.251	8405	.651		
9101	.037	8701	.107	8907	.261	9203	.723		
8202	.045	9107	.111	9001	.262	8302	.748		
8705	.045	8305	.123	8206	.277	9105	.959		
8704	.049	9201	.128	8306	.290	8503	1.367		
8507	.051	8604	.147	8403	.310	8605	3.413		
8303	.052	8804	.149	8504	.334	8806	4.081		