Jostein Lillestøl, NHH/SNF April 2015

Weibull Wind Worth: Wait and Watch?

1. The problem

Suppose W is the yearly worth of a wind mill project. In the case that the worth increase linearly with average wind speed, we may just as well assume W to be wind speed itself. In our illustrations we will do that. We assume W to be a random variable with probability distribution $F(w|\theta)$ depending on a location parameter θ, typically the mean, so that higher the better (except for extreme winds). We will assume that F is known, but θ partly unknown. The project is profitable if θ is at or above a certain threshold t and not profitable otherwise. In this context the issue is: Do we have enough information on θ to launch the project now or should we at a cost set up a wind mill and observe, say for a year, and the decide: launch the full project or scrap it. We have the following actions

a_1 = Accept project now
a_2 = Reject project now
a_{31} = Accept project after observation
a_{32} = Reject project after observation

so that $a_3 = \{a_{31}, a_{32}\}$ corresponds to the action "wait and watch". If θ>t (threshold) acceptance is preferred, otherwise rejection is preferred.

We will assume a linear loss function as follows (negative loss being gain):

Action: a	a_1	a_2	a_{31}	a_{32}
Loss: $L(\theta, a)$	$c_1 \cdot (t - \theta)$	$c_2 \cdot (\theta - t)$	$c_0 + c_1 \cdot (t - \theta)$	$c_0 + c_2 \cdot (\theta - t)$

Here the cost components are given by

c_0 = Cost of setting up a windmill and observe
c_i = Cost per unit difference from threshold t, i=1,2

where c_2 represents an opportunity cost for not launching a profitable project.

The loss function may be made more general by having different constants c', i=1,2 for negative losses. An alternative loss function may simply be the regret of wrong action, i.e. replacing the linear term $(t - \theta)$ by $I_{(-\infty,t]}(\theta)$ etc.

2. A Bayesian solution

We will analyze the two-stage decision problem within the common Bayesian framework, minimizing the expected loss, see Berger (1995). In general we have

Prior distribution of θ: $\pi(\theta)$
Distribution of observation W given θ: $f(w|\theta)$
Posterior distribution of θ given $W = w$: $\pi(\theta|w)$

The decision rule after observing $W = w$ depends on $\theta(w) = E(\theta|w)$, while the decision rule before the observation will depend on the predictive distribution of $\theta(W)$, which may be derived from the predictive (pre-posterior) distribution of *W*. We illustrate this in the case of Normal observations and Normal prior, for which we have a (well known) closed form solution. We may typically expect a solution depending on the prior expectation $\theta_0 = E(\theta)$, so that we accept/reject now when θ_0 is sufficiently large/small, and we go for observations when θ_0 is in a middle region

Example: Normal-Normal

Prior: $\theta \sim Normal(\theta_0, \tau^2)$
Observation: $W|\theta \sim Normal(\theta, \sigma^2)$
Posterior given *W*=*w*: $\theta|w \sim Normal(\theta(w), \tau^2(w))$

where $\theta(w) = \frac{\sigma^2}{\sigma^2+\tau^2} \cdot \theta_0 + \frac{\tau^2}{\sigma^2+\tau^2} \cdot w$ and $\tau^2(w) = \frac{\sigma^2 \cdot \tau^2}{\sigma^2+\tau^2}$.

The pre-posterior (predictive) distributions now are

$$W \sim Normal(\theta_0,\, \tau^2 + \sigma^2\,)$$

$$\theta(W) \sim Normal(\theta_0,\, \tau^2 \cdot \frac{\tau^2}{\sigma^2 + \tau^2})$$

In case of n independent observations we have to replace w by the mean $\bar{w}$ and $\sigma^2$by $\frac{\sigma^2}{n}$.

Returning to the general theory we now consider the expected loss (risk) by taking action *a* when the uncertainty about θ is given by the distribution π:

$$r(\pi, a) = E_\pi L(\theta, a)$$

For a_1=*Accept project now* we use the given prior π and get

$$EL(\theta, a_1) = \int c_1(t-\theta)d\pi(\theta) = c_1(t-\theta_0)$$

where in general θ_0 denotes the prior expectation of θ. Similarly for a_2=*Reject project now* $EL(\theta, a_2) = c_2(\theta_0 - t)$.

If we do not accept or reject the project right away, we set up the observational windmill and after having observed *W*=*w* the relevant distribution is the posterior $\pi(\theta|w)$. We now get

$$\begin{aligned} EL(\theta, a_{31}) &= \int (c_0 + c_1(t-\theta))d\pi(\theta|w) = c_0 + c_1(t - \theta(w)) \\ EL(\theta, a_{32}) &= \int (c_0 + c_2(\theta - t))d\pi(\theta|w) = c_0 + c_2(\theta(w) - t)) \end{aligned}$$

where in general $\theta(w)$ denotes the posterior expectation of θ. Of the two possible actions we choose the one that minimizes the expected loss, i.e.

$$a_{31} \text{ if } \theta(w) \geq t \text{ with expected loss } c_0 - c_1(\theta(w) - t))$$

$$a_{32} \text{ if } \theta(w) < t \text{ with expected loss } c_0 + c_2(\theta(w) - t))$$

The preposterior loss (*W* random) is:

$$c_0 - c_1(\theta(W) - t)) \text{ if } \theta(W) \geq t$$

$$c_0 + c_2(\theta(w) - t)) \text{ if } \theta(W) < t$$

The expected preposterior loss is therefore

$$\begin{aligned} EL_3 &= c_0 + \int_{-\infty}^{t} c_2(x-t)dH_0(x) - \int_{t}^{\infty} c_1(x-t)dH_0(x) \\ &= c_0 - c_1(E\theta(W) - t) + (c_1 + c_2)(H_1(t) - tH_0(t)) \end{aligned}$$

where H_0 is the distribution of $\theta(W)$ and $H_1(t) = \int_{-\infty}^{t} x dH_0(x)$

Initially we compare EL_3 for the action set $\{a_{11}, a_{12}\}$ with $EL_1 = c_1(t - \theta_0)$ for action a_1 and $EL_2 = c_2(\theta_0 - t)$ for action a_2 and decide according to the minimum.

Example: Normal-Normal (cont'd)

In the case of Normal prior and Normal observations we have

$$H_0 \sim Normal(\theta_0, \tau_0^2) \text{ where } \tau_0^2 = \frac{\tau^4}{\sigma^2 + \tau^2}$$

This gives

$$\begin{aligned} H_0(x) &= G(\tfrac{x-\theta_0}{\tau_0}) \\ H_1(x) &= \theta_0 G\left(\tfrac{x-\theta_0}{\tau_0}\right) - \tau_0 g(\tfrac{x-\theta_0}{\tau_0}) \end{aligned}$$

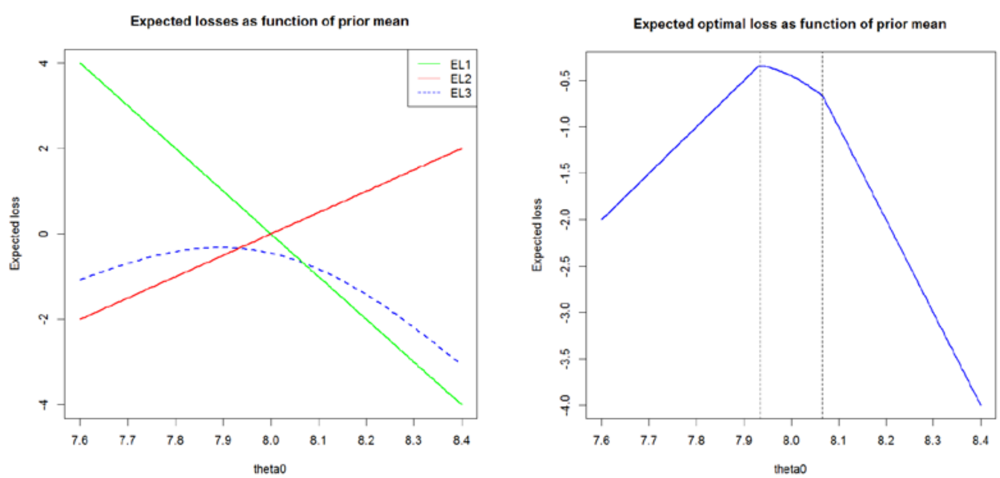
where G and g are the standard Gaussian cumulative distribution and density respectively. We then get the preposterior expected loss for going on to observe (with possible actions a_{11} or a_{12})

$$EL_3 = c_0 - c_1(\theta_0 - t) + (c_1 + c_2)(\,(\theta_0 - t)G\left(\frac{t-\theta_0}{\tau_0}\right) - \tau_0 g(\frac{t-\theta_0}{\tau_0}))$$

This is then compared with $EL_1 = c_1(t - \theta_0)$ for action a_1 and $EL_2 = c_2(\theta_0 - t)$ for action a_2. The initial decision is according to the minimum. In the case that EL_3 is the minimum and we take action $a_3 = \{a_{31}, a_{32}\}$ we have, after having observed *W=w:* If $\theta(w) \geq t$ take action a_{31} = Accept project, if $\theta(w) < t$ take action a_{32} = Reject project.

<u>Illustration:</u> c_0=1, c_1=10, c_2=5, $\sigma = 4.0$ $\tau = 2.0$ and t=8.0

Computation of the expected losses for accepting now, rejecting now and for postponing the decision, for varying prior expectation θ_0, are graphed below. We also give the resulting loss curve taking the optimal choices.



We see that we accept now (a_1) if the prior mean $\theta_0 > 8.065$, reject now (a_2) if $\theta_0 < 7.935$ and postpone the decision and observe (a_3) if $7.935 < \theta_0 < 8.065$.

From the above formulas it is easy to investigate the sensitivity of the conclusion with respect to varying the parameters involved.

The Normal-Normal model provided a complete analytic solution, which is very attractive. Although most models are a compromise between convenience and realism, normality may be quite unrealistic in our context. In general the main opportunities are:

- A complete analytic solution, e.g. using conjugate priors
- Simulations based partly on some analytic features of the model
  e.g. simulate from posterior using Markov chain Monte Carlo or Laplace approximation
- Simulation of the decision process itself

For non-normal prior and observation the exact closed form solution will be somewhat involved, and we typically gave to resort to numeric calculations and simulations.

3. The Weibull model

In the case of *W* being wind speed itself available data have supported the Weibull model, and to illustrate the opportunities above we will examine this model.

Observation distribution: Weibull(k, γ),

Cumulative distribution: $F(w|\gamma) = 1 - e^{-\frac{w^k}{\gamma}},\ w \geq 0$

Probability density: $f(w|\gamma) = \frac{k}{\gamma} w^{k-1} e^{-\frac{w^k}{\gamma}},\ w \geq 0$

Here k is a shape parameter assumed to be known and γ an unknown scale parameter.

Expectation $\theta = E(W|\gamma) = \Gamma\left(1 + \frac{1}{k}\right) \cdot \gamma^{\frac{1}{k}}$

Instead of switching to our base parameter θ it is convenient to stick to γ one step further.

Conjugate prior: $\gamma \sim IG(\alpha, \beta)$ (Inverted Gamma distribution)

$$E(\gamma) = \frac{\beta}{\alpha - 1} \qquad Mode(\gamma) = \frac{\beta}{\alpha + 1} \qquad var(\gamma) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$$

Posterior: $\gamma|w \sim IG(\alpha + 1, \beta + w^k)$

Posterior expectation: $E(\gamma|w) = \frac{\beta + w^k}{\alpha}$

Predictive: $W|\ \alpha, \beta \sim F(w|\alpha, \beta) = 1 - (1 + w^k/\beta)^{-\alpha}$

Translating this to our basic parameter θ, it follows that

$$E\theta = EE(W|\gamma) = \Gamma\left(1 + \frac{1}{k}\right) \cdot E(\gamma^{\frac{1}{k}}) = \Gamma\left(1 + \frac{1}{k}\right) \cdot \frac{\Gamma\left(\alpha - \frac{1}{k}\right)}{\Gamma(\alpha)} \cdot \beta^{1/k}$$

For convenience we denote this expression for the expectation by $E_0 = \mathrm{E}(\mathrm{k}, \alpha, \beta)$. We therefore have with $E_1 = \mathrm{E}(\mathrm{k}, \alpha + 1, \beta)$

$$E(\theta|w) = E_1 \cdot (1 + w^k/\beta)^{1/k}$$

so that the preposterior mean of the decision rule is

$$E(\theta|W) = E_1 \cdot (1 + W^k/\beta)^{1/k} = \mathrm{X}\ (\text{say})$$

Its required preposterior distribution $H_0(x)$ is then obtained from the preposterior (predictive) distribution of $X|\ \alpha, \beta$ above as

$$H_0(x) = P(X \leq x\ ) = 1 - \left(\frac{x}{E_1}\right)^{-k\alpha}, \qquad x > E_1$$

from which we get

$$H_1(x) = E_1\ \cdot \frac{k\alpha}{k\alpha - 1} \cdot (1 - \left(\frac{x}{E_1}\right)^{-k\alpha+1}), \qquad x > E_1$$

As the limit as x tends to infinity we get $(\theta(W)) = E_1\ \cdot \frac{k\alpha}{k\alpha-1}$. We now have the necessary expressions to enter the general decision formula. Before we return to that we look at a few more issues.

From $H_0(x)$ we obtain the density of the preposterior (predictive) distribution of $X|\ \alpha, \beta$ as

$$h_0(x) = \frac{k\alpha}{E_1}\left(\frac{x}{E_1}\right)^{-k\alpha-1}, \qquad x > E_1$$

From $H_0(x)$ we also obtain an easy way to simulate from the (predictive) distribution of $X|\ \alpha, \beta$:

Generate $U\ \sim Uniform[0,1]$ and compute $X = E_1 \cdot U^{-\frac{1}{k\alpha}}$

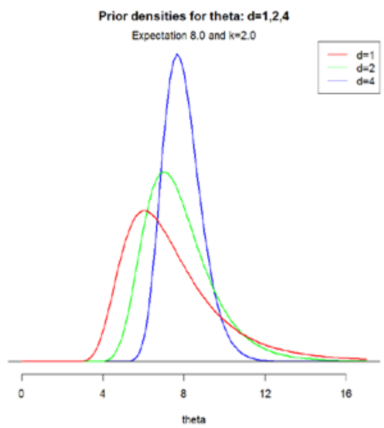
Suppose we want to determine the prior parameter values $(\alpha, \beta) = (\alpha_0, \beta_0)$ so that $\theta = \theta_0$. We may the compute the corresponding γ_0 form the formula above and require $\gamma_0 = \frac{\beta_0}{\alpha_0 - 1}$.

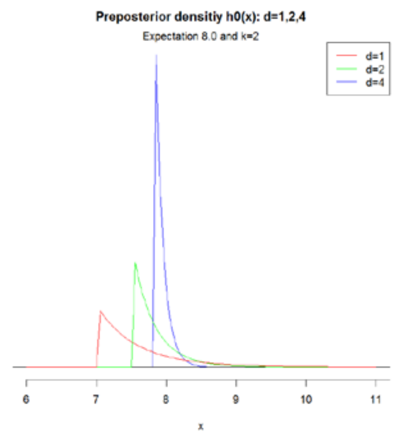
Since α is increased by one for each observation, it is clear that small α corresponds to little information and α>2 is required for the prior distribution of γ to have expectation and variance. If we assume that the standard deviation is a fraction 1/d of γ_0 we get

$$\alpha_0 = 2 + d^2 \quad \text{and} \quad \beta_0 = \gamma_0 \cdot (\alpha_0 - 1)$$

In the case of $\theta_0 = t = 8.0$ (the threshold) we get $\gamma_0 = 52.15$. For d=4 (similar to the previous Normal-Normal example) we get $\alpha_0 = 18$ and $\beta_0 = 938.7$.

In the graph below we show the density $h_0(x)$ for the three cases when $\theta_0 = 8.0$, namely for d=1,2,4 corresponding to $\alpha_0 = 3, 6{,}18$.





The preposterior distribution of $E(\theta|W)$ for W random that was needed in the decision rule and was here derived analytically. Suppose that our theoretical knowledge is limited to the conjugate prior-posterior analysis of the basic parameter γ. The needed distribution can then instead be obtained by simulations as follows: Given (k,d)

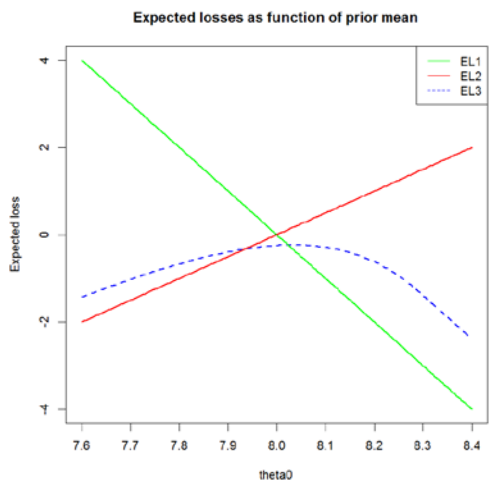
1. Decide on $\theta_0$and compute γ_0 for the given k
2. Compute (α_0, β_0)
3. Simulate W according to the Weibull model
4. Simulate $\gamma|w$ by the posterior Inverted Gamma[1]
5. Compute corresponding θ
6. Repeat 4-5 (for the same w) m times giving $\theta_i, i = 1,2, \dots, m$
   and take $\theta(w) = Mean(\theta_i)$
7. Repeat 3-6 n times to give $\theta(w_j), j = 1,2, \dots, n$

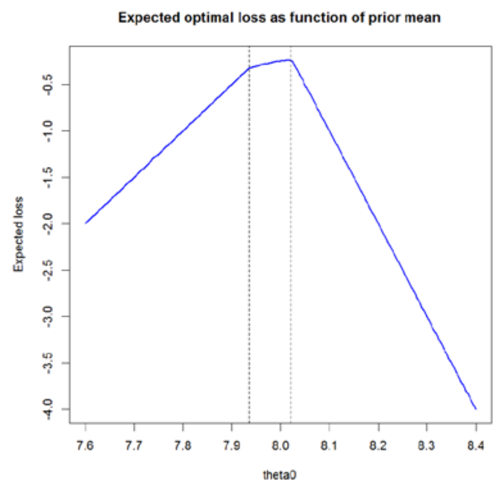
Note that we have avoided the temptation to use the posterior mean of γ and the link between γ and θ, which gets rid of the inner loop, since the means do not transform the same way unless k=1.

The analysis above is heavily dependent on specific knowledge of the posterior of γ. Lacking this for some reason (e.g. non-conjugate prior) there are nevertheless sampling opportunities, e.g. based on Markov chain Monte Carlo principles.

[1] Simulation from the IG-distribution is available in R, among others in the package MCMCpack.

Numerical illustration: From available windmill data we have found that yearly mean wind speed may be modelled by a Weibull model with k=2. At prior expectation $\theta_0 = 8.0$ this will give standard deviation of W about 4.1, close to the specification in our Normal-Normal example. Taking d=4 we have produced the following graphs of expected losses for the cost structure c_0=1, c_1=10, c_2=5:



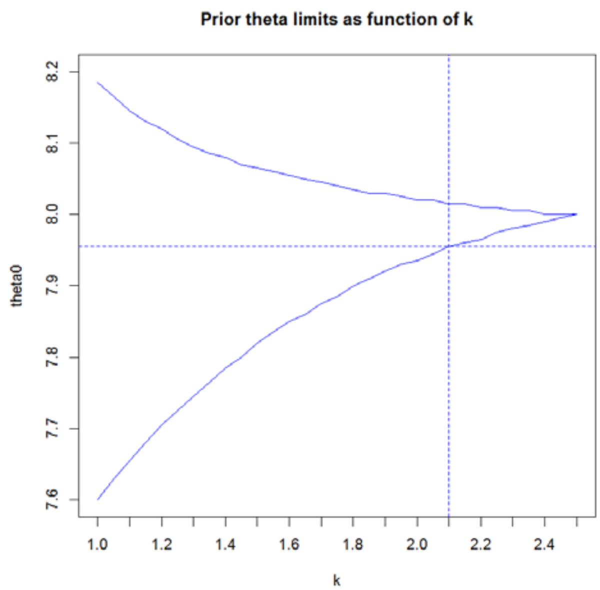


We see that we accept now (a_1) if the prior mean $\theta_0 > 8.020$ reject now (a_2) if $\theta_0 < 7.935$ and postpone the decision and observe (a_3) if $7.935 < \theta_0 < 8.02$. The lower limit is exactly as the Normal example, but the upper limit is slightly lowered, meaning that we less often will postpone the decision. In fact, k=2.1 will give the exact standard deviation 4.0 of the Normal example, for which the region of postponing is narrowed to $7.955 < \theta_0 < 8.015$. Repeated computations for k=2.1 for various combinations of costs are given in the following table, which shows some interesting differences between the Normal and Weibull model that deserves to be studied in more detail:[2]

c_0	c_1	c_2	Normal		Weibull	
1	10	5	7.935	8.065	7.955	8.015
1	10	10	7.890	8.110	7.790	8.055
1	20	5	7.855	8.145	7.845	8.045
1	20	10	7.825	8.110	7.790	8.055
1	20	20	7.790	8.210	7.600	8.100
1	30	5	7.805	8.195	7.780	8.055
1	30	10	7.790	8.210	7.655	8.070
1	30	20	7.760	8.240	7.600	8.095
1	30	30	7.735	8.265	7.600	8.115

[2] The computations are done on a grid size of 0.005

The table shows some interesting differences between the Normal and Weibull model. The Weibull interval for postponement is skewed, while the Normal one is symmetric around 8.0 and typically wider than the Weibull interval. Seemingly the Weibull interval is slightly shifted downwards compared with the Normal one, meaning that we are less likely to reject the project without observation. In the Weibull case it is of interest to see how the continuation interval depends on the shape parameter k. For the cost structure c_0=1, c_1=10 c_2=5 in the first line of the table we have illustrated this in the following graph for the range [1.0, 2.5] :



We see that the continuation interval is very wide at k=1.0 (the exponential distribution), narrows as k increases and collapses to zero at about k=2.5. The case k=2.1 of the first line in the table is illustrated by dotted lines. Beyond the point k=2.5 the wait and watch action is ruled out, and we simply take immediate action according to whether $\theta_0 < t$ or $\theta_0 > t$, i.e. the prior mean less than or greater than the threshold. This behavior is not surprising as increasing k, in a sense, represents more prior information. At some point additional information will be worthless. With a different cost structure like the ones lower down in the table, the continuation interval is wider and breaks down further to the right. As we will see in the next section, the realistic value of k will depend on the time scale of the observation (hourly, daily, weekly, monthly, yearly), which again, in a sense, reflects the amount of information.

4. Empirical support for the Weibull model

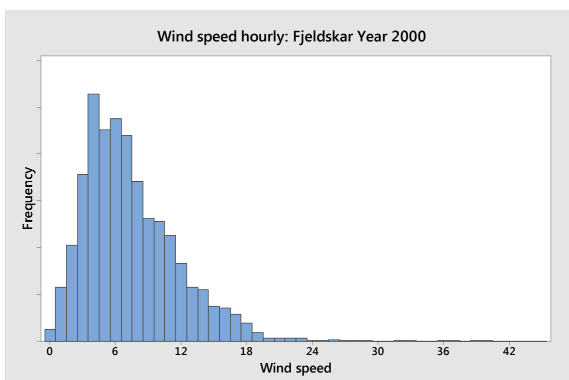
We have wind speed data for 13 years (2000-2012) registered hourly at 68 locations. Wind speed at a location varies, and shows some seasonality over the day and over the year, but the main component of variation is geophysical, which in our context appear as randomness. Hourly wind registration is clearly relevant for a windpower provider running the mills, but for investment decisions it may be sufficient to consider aggregated data, say daily, weekly, monthly or even yearly averages. The kind of model adopted may depend heavily on the chosen time scale. In the case of normal models it does not matter, since aggregates of normal variates turn out normal. For Weibull models we have that the time scale chosen will determine the shape parameter. Being consistent, we cannot switch time scale and stay within the Weibull class, since averages of Weibull variates are not Weibull. However, a model is always a compromise, and a Weibull model may provide good fit and turn out useful on different timescale, without internal consistency. We will therefore look at the Weibull model on different time scales.

First consider the hourly wind speed data for one year (2000) at one location (Fjeldskar). The descriptive statistics are as follows[3]:

Descriptive Statistics: Wind speed hourly at Fjeldskar (2000)

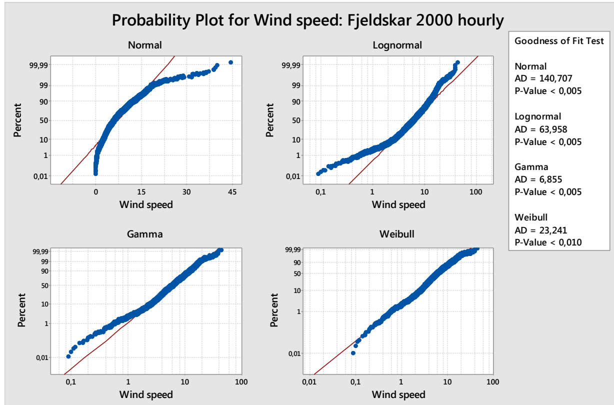
Variable	N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
WindSpeed	8780	7.436	0.047	4.415	0.090	4.300	6.560	9.780	44.610

The observed empirical distribution is displayed in the following histogram:



We see that the distribution is skewed with a fairly log right tail, and definitely different from the normal distribution. Alternative skew distributions to the symmetric normal may be lognormal, Gamma and Weibull, and comparative probability plots show that the Weibull is superior to the others. [4]

[3] The observations started on 01.01.2000 at 04:00:00.
[4] With this many observations the formal test would reject any specified model.



We adopt the Weibull-model, and the maximum-likelihood estimates of its parameters turned out to be:

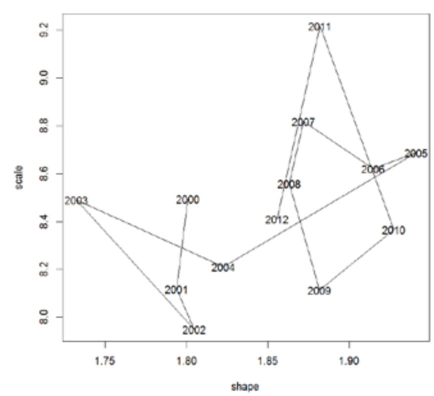
shape=1.78, scale=8.37

This is according to the parameterization commonly used in statistical software (like R and Minitab), which is (x/scale)$^{\text{shape}}$, instead of the one most suitable for Bayesian analysis with conjugate prior, which is x$^{\text{shape}}$ /scale. This means that the estimate of our scale parameter $\gamma = scale^{shape} = 8.37^{1.78}$=43.9.

Repeating the estimation for each of the years 2001-2012 we see that the parameters are fairly stable.

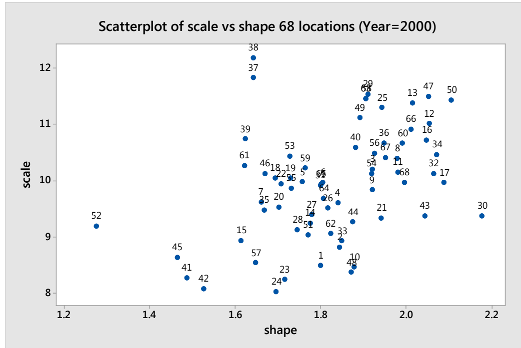
Shape: Mean=1.73, Minimum=1.85, Maximum=1.94
Scale: Mean: 8.47, Minimum=7.95, Maximum=9.22

A plot of these estimates in their time sequence is shown in the following plot, which also shows a noticeable correlation between the shape and scale estimates (of about 0.37).



The estimated Weibull parameters (shape, scale) for the 68 locations for a specific year are of interest to us, since this may indicate the reasonable prior assumptions on these parameters for new locations.[5] The actual values of (shape, scale) for year 2000 are given in the following scatterplot.

[5] This can of course be improved by utilizing some specifics for the new location and selected other locations based on similarities.



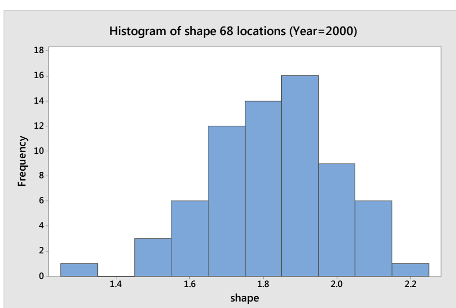
We locate Fjeldskar (1) with (shape=1.78, scale=8.37), in the middle with respect to shape and in low region with respect to scale. The descriptive statistics for the estimated parameters for the 68 locations in year 2000 follows:

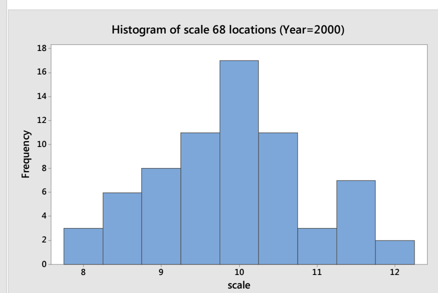
Descriptive Statistics of shape and scale: 68 locations year 2000

Variable	N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
shape	68	1.828	0.021	0.171	1.276	1.709	1.832	1.947	2.178
scale	68	9.933	0.119	0.982	8.033	9.254	9.972	10.575	12.187

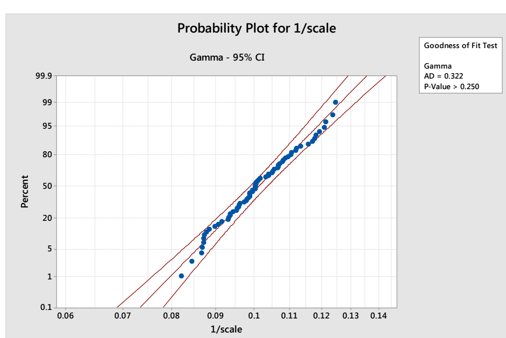
```
Pearson correlation of shape and scale = 0.400
```

The empirical distributions of the estimated shape and scale parameters over the 68 locations are given in histograms as follows:





The latter distribution is well represented by an Inverse Gamma, as shown in the probability plot.

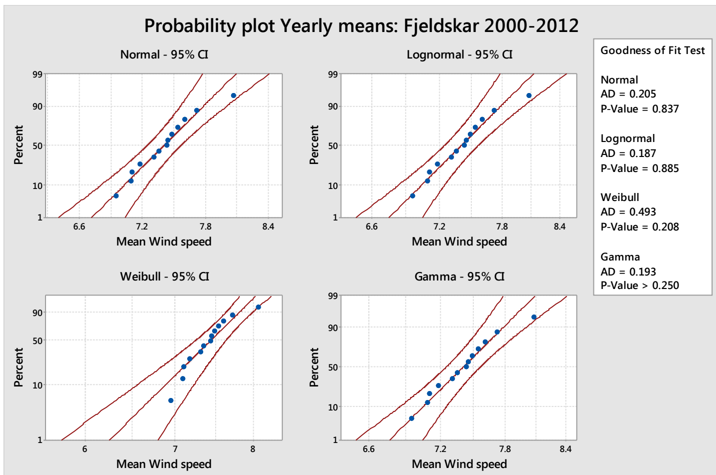


We now look at the Weibull model for the aggregated data, taking the average of the hourly wind speeds over the year at each location for consecutive years 2000-2012. For location 1 (Fjeldskar) we have the following descriptive statistics for wind speed:

Descriptive Statistics: Wind speed Fjeldskar Yearly means

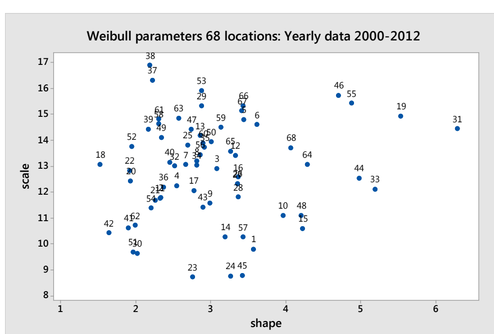
Variable	N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Wind speed	13	7.409	0.082	0.297	6.955	7.146	7.436	7.572	8.071

We get the following probability plot for identification, with focus on the alternatives: Normal, Lognormal, Weibull and Gamma.



We see that the Weibull model is seemingly slightly inferior to the others, but with just 13 observations none of the models can be rejected. The major lack of fit of the Weibull is in the low tail. In our context the upper tail is more important, and there all four models point to the possibility of an even more extreme yearly average wind speed not yet observed in the data.

From estimation of a Weibull model for the 13 yearly averages, we get the following scatterplot for the combinations of estimates (shape, scale) for the 68 locations:



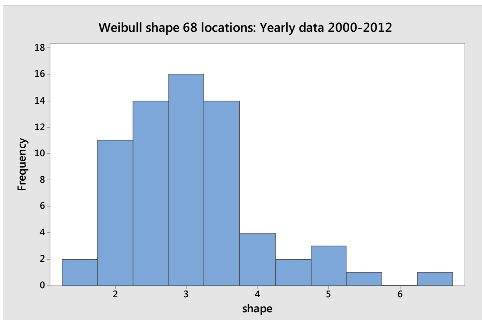
In the scatterplot we locate Fjeldskar (1) with parameters (shape, scale)=(3.57, 9.81).

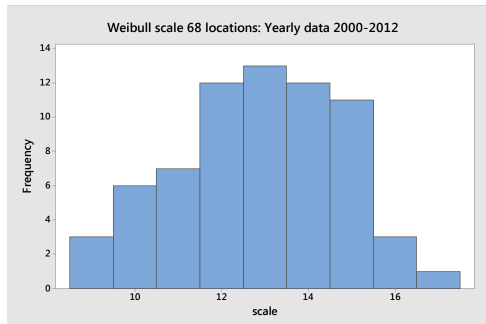
The descriptive statistics for the 68 shape and scale parameter estimates are as follows:

Descriptive Statistics of shape and scale: 68 locations Yearly data 2000-2012

Variable	N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
shape	68	3.042	0.115	0.950	1.526	2.327	2.878	3.420	6.293
scale	68	12.865	0.232	1.916	8.739	11.611	13.054	14.431	16.898

The empirical distributions of the 68 shape and scale parameter estimates are as follows:





The distribution of the scale parameter is not so well fitted to an Inverse Gamma as for the hourly data. Two-parameter alternatives like the lognormal give only marginally better fit. A fairly good fit is obtained by taking 1/scale to have a three-parameter Gamma, i.e. with a lower threshold.

We see that going from hourly to yearly data the shape parameter for location 1 (Fjeldskar) increased from about 1.8 to about 3.0. It turns out that the optimal solution is not very sensitive to this specification, i.e. we get about the same break point for a given expected (average) wind speed.

References

Berger, J.O. (1993) Statistical Decision Theory and Bayesian Analysis, Springer.