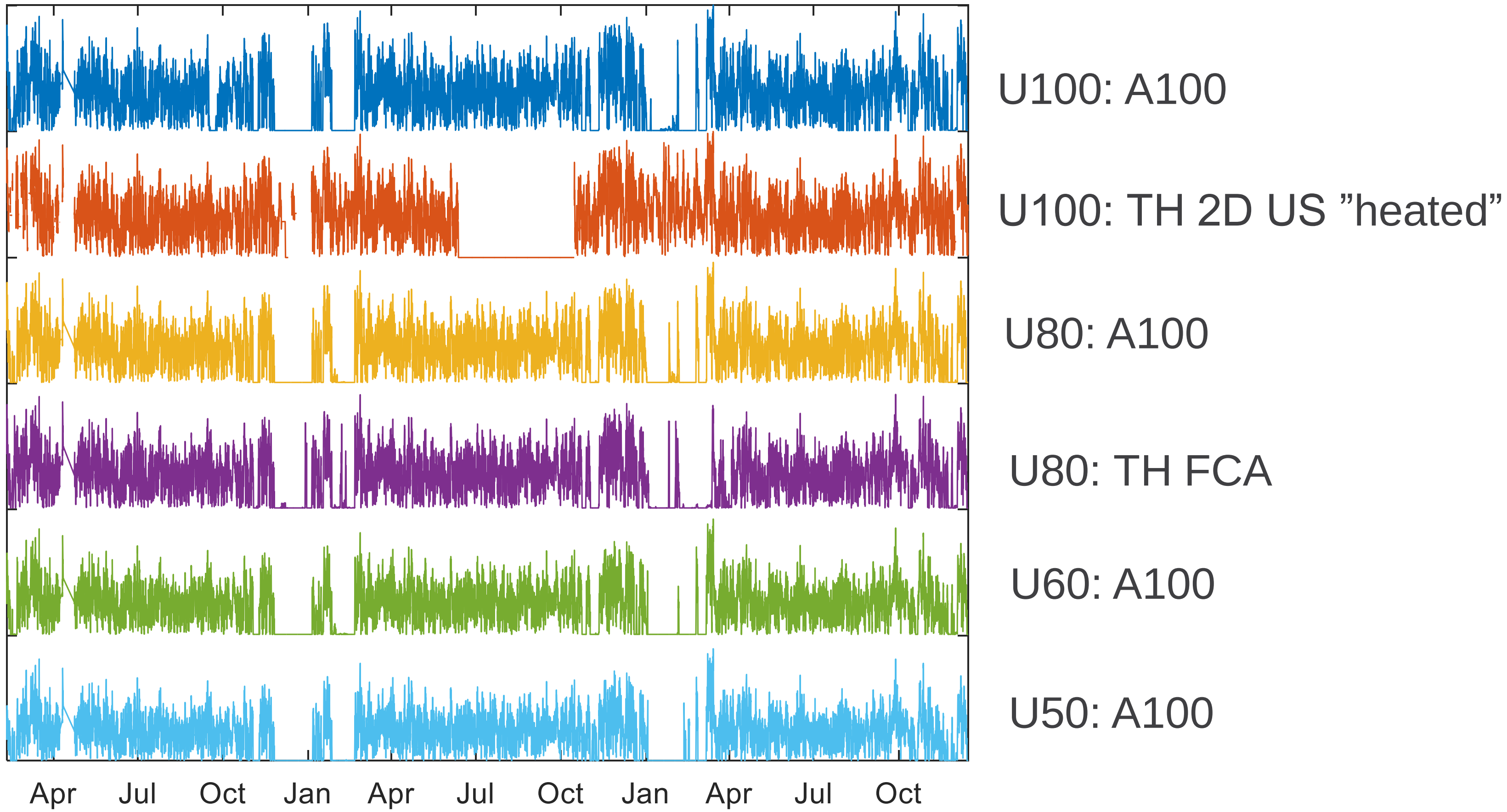
Frozen anemometers and bias in the wind resource

Lasse Johansson, Sweco

Today

- Data loss patterns due to freezing
- Consequences of loss
- Bad data: ignore, remove or substitute?
- Seasonality
- Reduce variance. Then transform. Then reinsert variance. Finally substitute.
- Examples



Apr
Jul
Oct
Jan
Apr
Jul
Oct
Jan
Apr
Jul
Oct
U100: A100
U100: TH 2D US ”heated”
U80: A100
U80: TH FCA
U60: A100
U50: A100

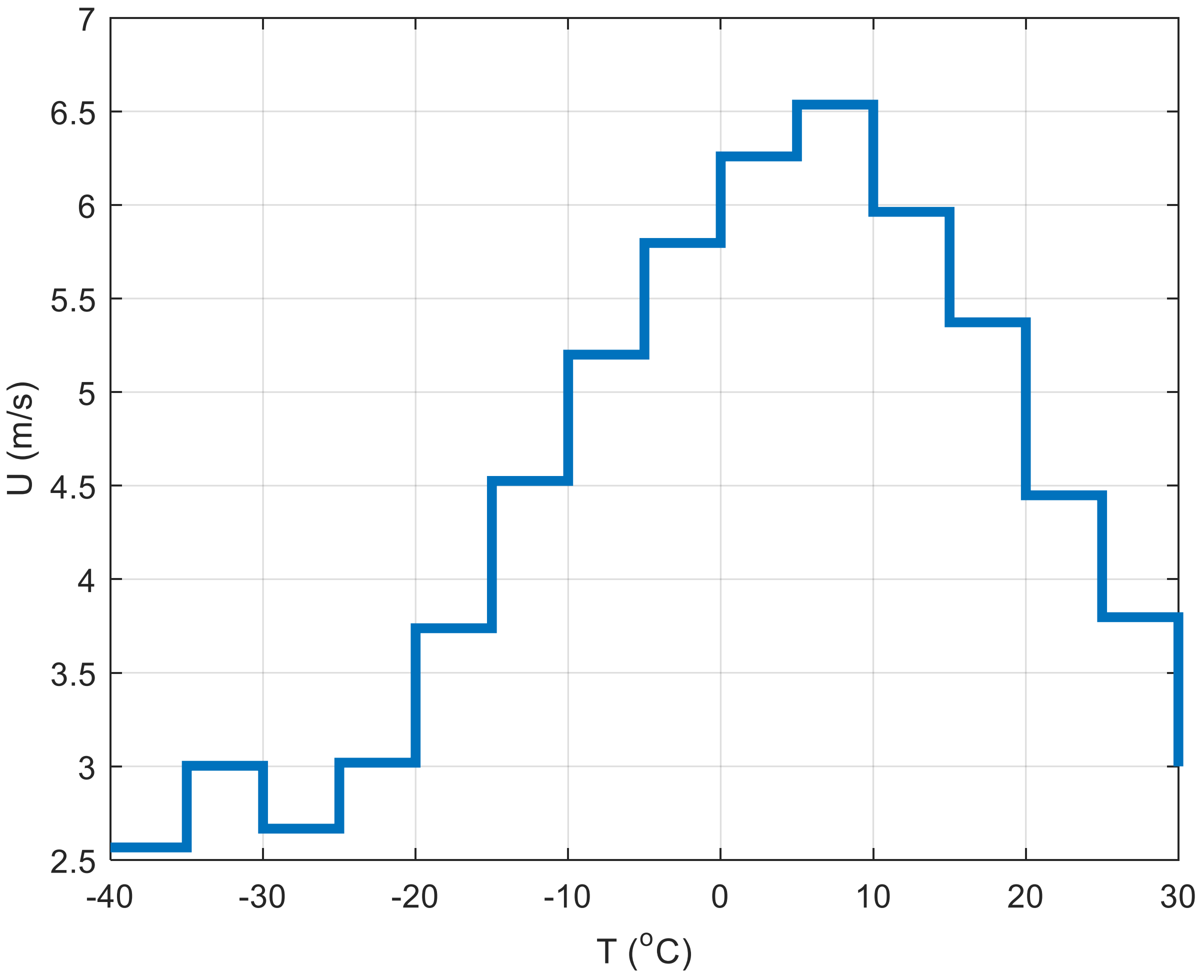
The page before shows a typical mast measurement from northern Sweden. Measurements were done at 100, 80, 60 and 50 m with cup anemometers, unheated. At 100 m there was a heated Thies ultrasonic anemometer.

The tick distance is three months so we can see data interruptions of several months each winter.

The Thies ultrasonic delivers data the second winter. However it was out of order the preceeding summer and winter. The data during the second winter also appear somewhat strange in the interval when only this instrument works. Compare with the observations by the sam instrument immediately before and after the period when all the cup anemometers are frozen.

So, we are left with the unpleasant but not uncommon situation with no or possibly unreliable mast data during the winter.

Is this a problem? Why dont we just remove all frozen and suspicious data, continue and hope for the best?



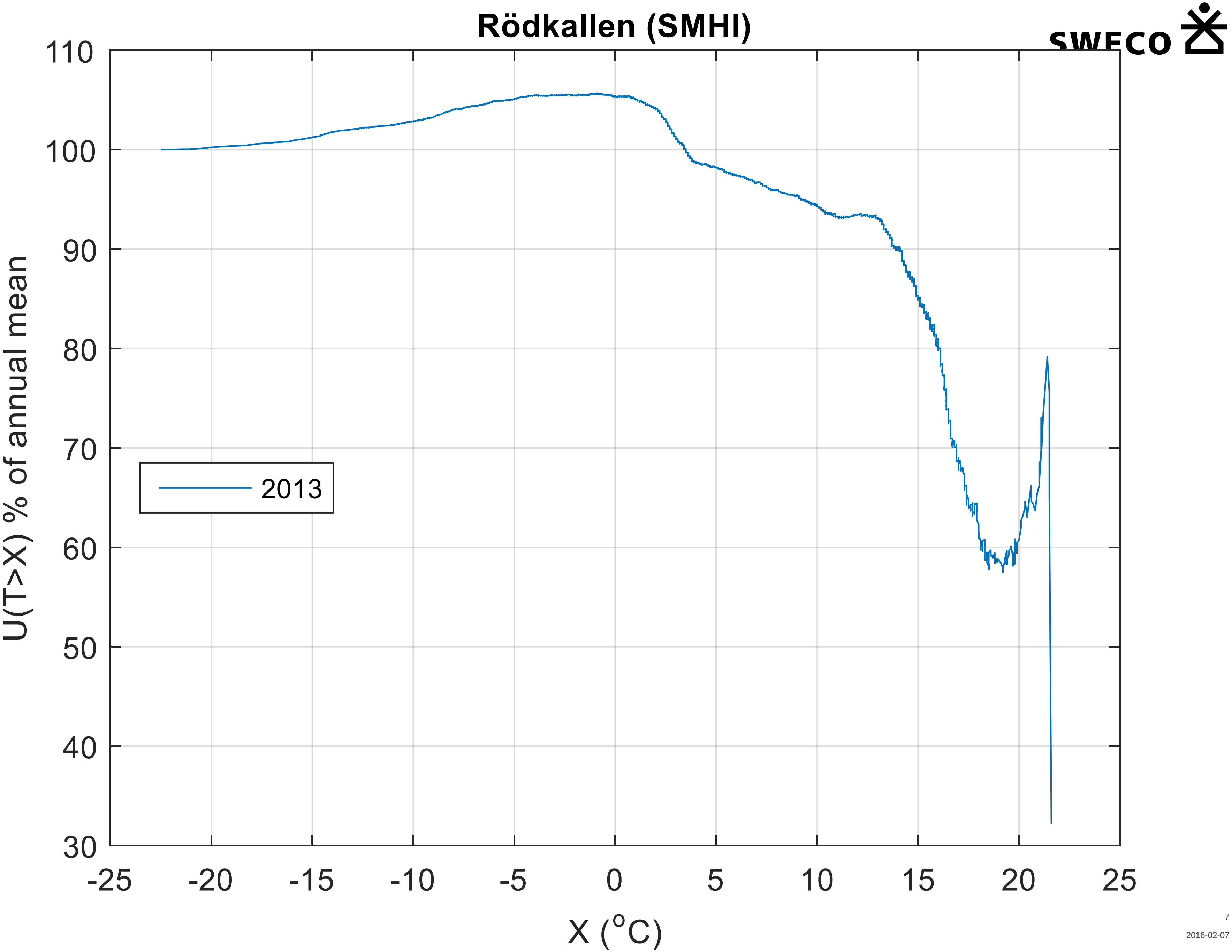
7
6.5
6
5.5
5
4.5
4
3.5
3
2.5
U (m/s)
-40
-30
-20
-10
0
10
20
30
T (^{o}C)

So if we just delete the frozen and suspicious data and ignore the holes – what will happen? What was the true wind during the hole periods?

If the wind is independent of temperature, T, the holes, typically times with $T<0$, will have no or only small effects. As we can see in the previous figure this is far from true. On the contrary, there is a strong connection between wind speed and temperature. The wind at this location is strongest for $0<T<10$ degrees Celsius. Below zero it gets weaker the lower the temperature.

A freezing anemometer will induce a dataloss that significantly will distort the annual wind resource.

Thus – ignoring the data holes will bias the wind resource.

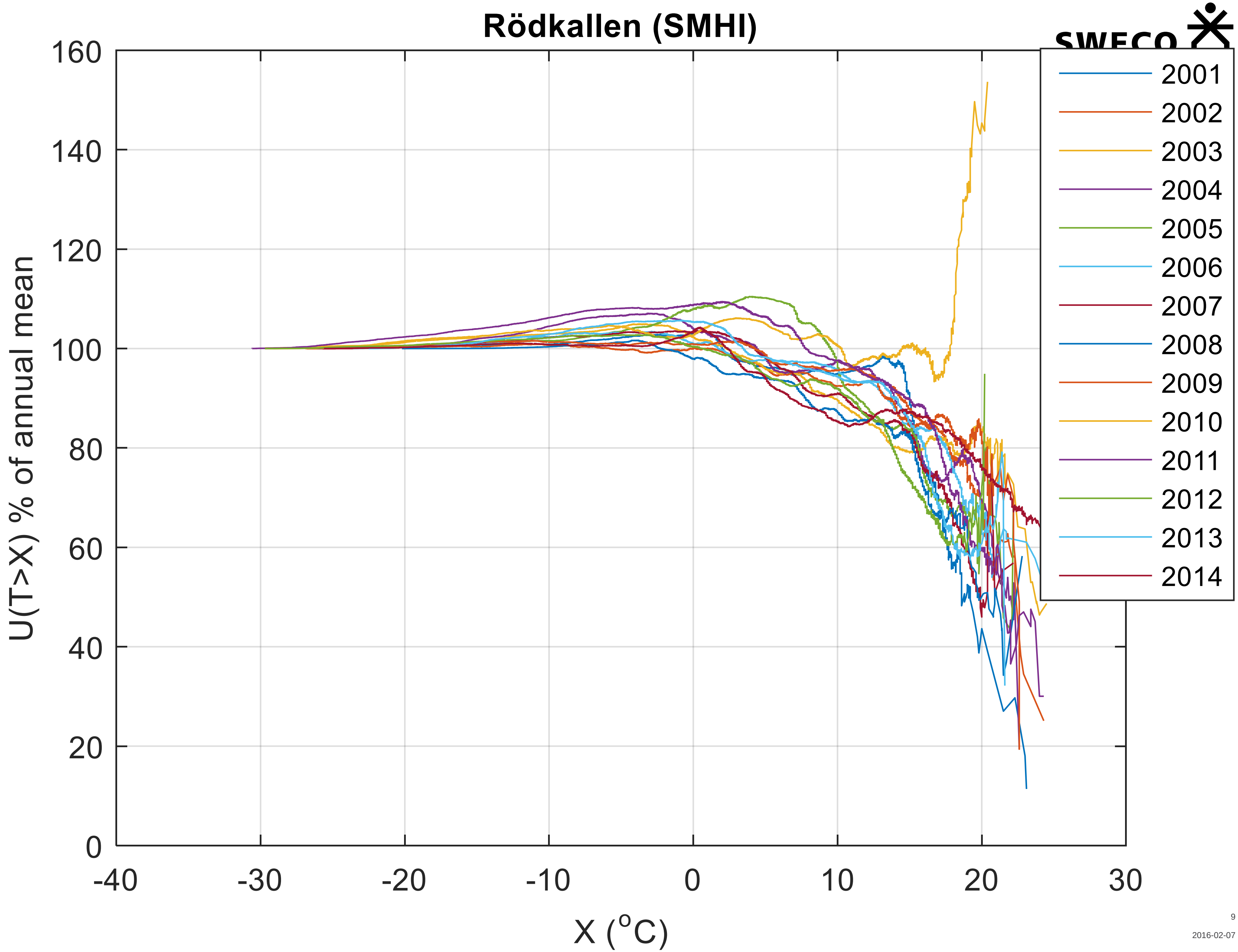

Rödkallen (SMHI)
110
100
90
80
70
60
50
40
30
U(T>X) % of annual mean
-25
-20
-15
-10
-5
0
5
10
15
20
25
X (°C)
2013

To quantify the influence of temperature dependent dataloss, we may draw the cumulative average wind speed as in this graph.

On the x-axis we have the temperature and on the y-axis windspeed in % of the annual mean for 2013 at Rödkallen in Bottenviken off Luleå. The curve starts to the right with the observation on the warmest hour of the day. The temperature was +27oC and the wind was just 30 %, of the annual mean. For example At X=+10oC the average includes all observations with T>10oC. Since wind speed increases with temperature for T>10, the cumulative mean catches up and is 94 % of average and at T= 3oC the cumulative mean gets stronger, >100 %, than the annual. At T=0oC the cumulative is 5 % above the annual mean.

So, assuming that the instrument stops working at T=0, missing these data will make us over-estimate the mean wind speed by 5 %.

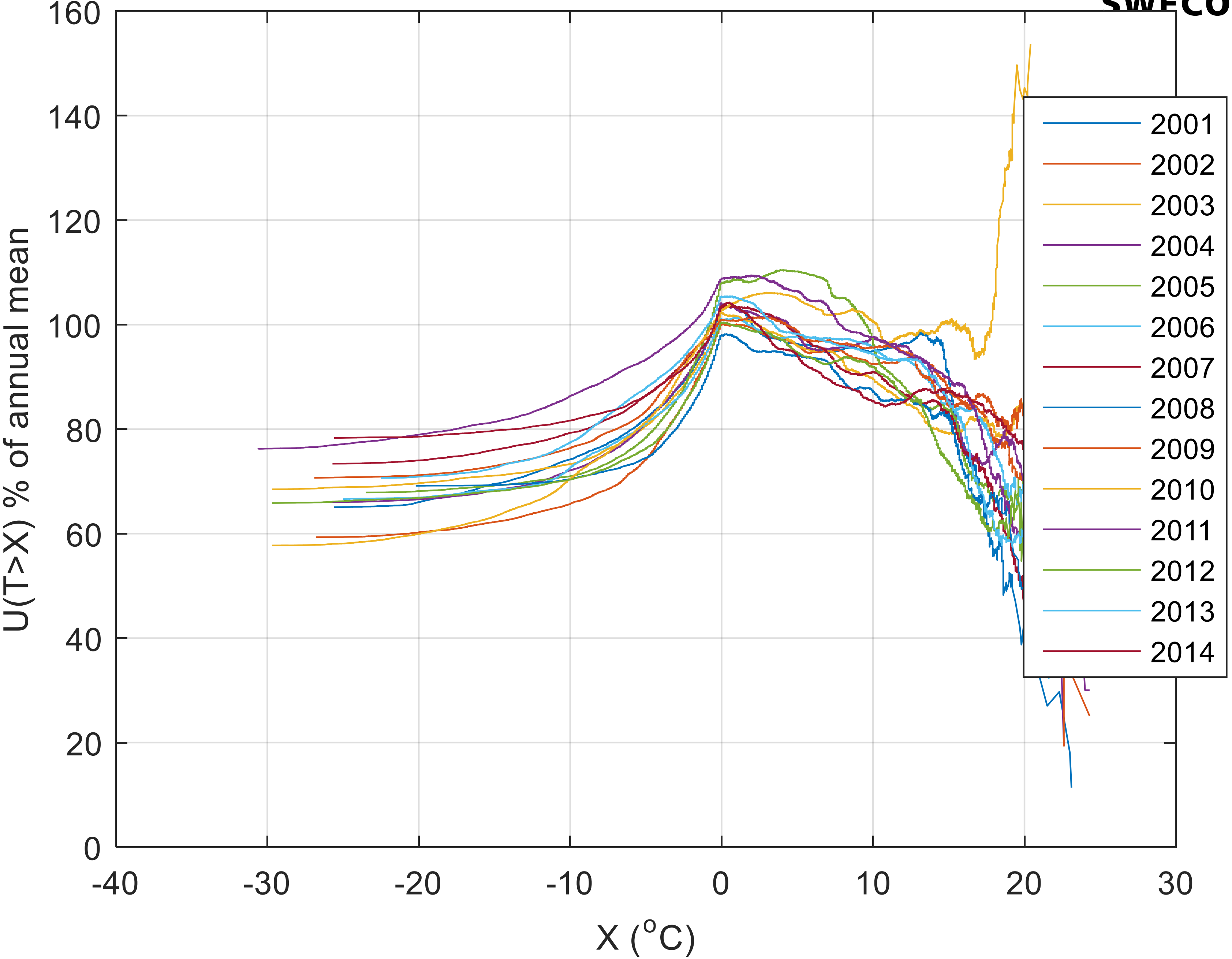
So, if data holes will bias the wind resource – can we fill them with something sensible?


Rödkallen (SMHI)
SWECO
U(T>X) % of annual mean
X (°C)
0
20
40
60
80
100
120
140
160
-40
-30
-20
-10
0
10
20
30
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014

The same graph as earlier but for single years 2001-2014 demonstrates that the cumulative mean develops differently. At T=0 oC for example the cumulative mean wind U(T>0) varies from -2 to + 8 % of the annual mean.

Of course this variability was expected considering the varying weather but it is also influenced by the anemometer performing differently due to icing and malfunction.

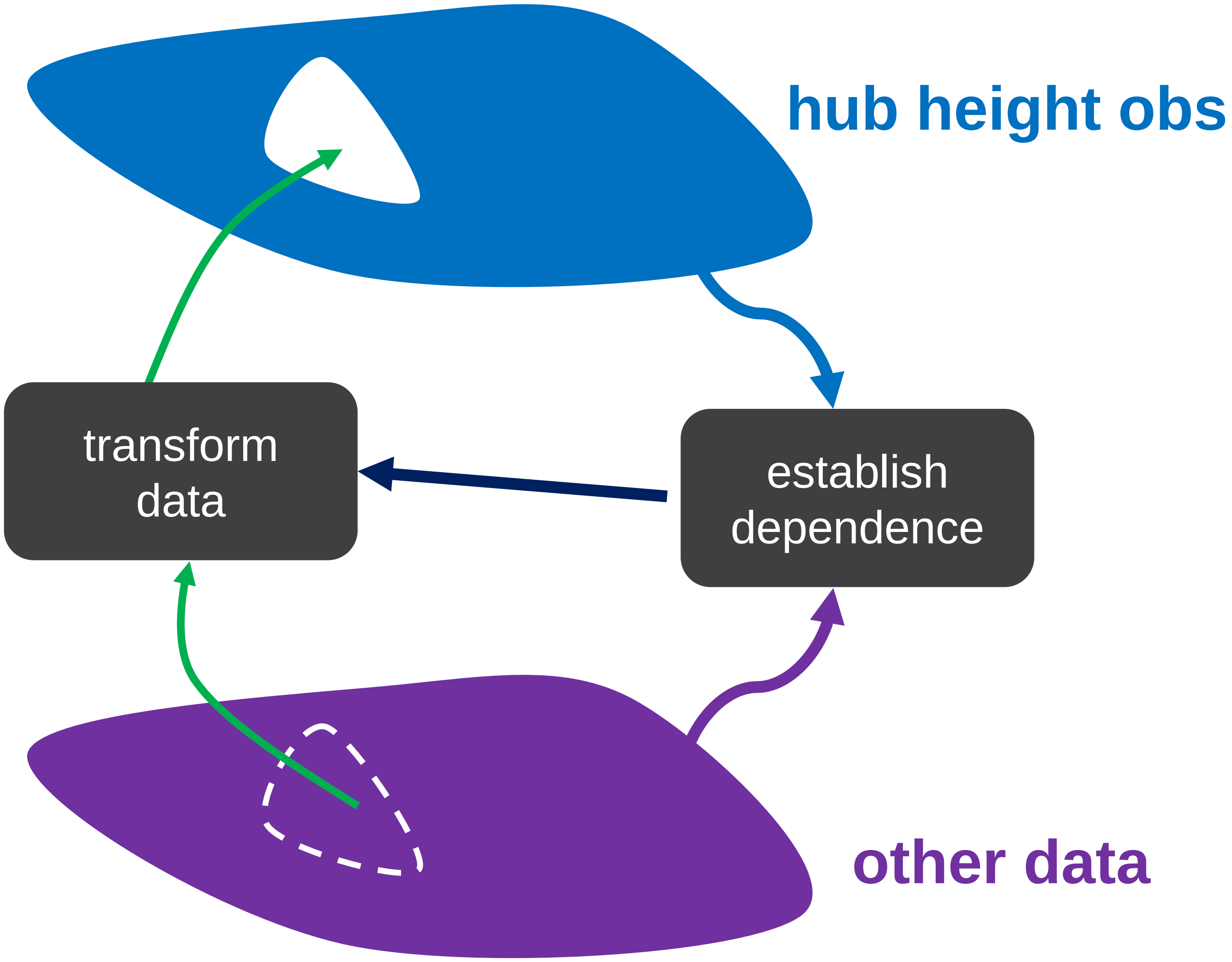
Bad obs = 0 for T<0°C



SWECO

Here i have modelled the hypotetical situation that the anemometer stops working when T<0 oC and only produces zero wind speed. If this would go unnoticed, one would en up with enormous biases of annual wind speed amounting to -20 to -40 % of the actual annual mean.

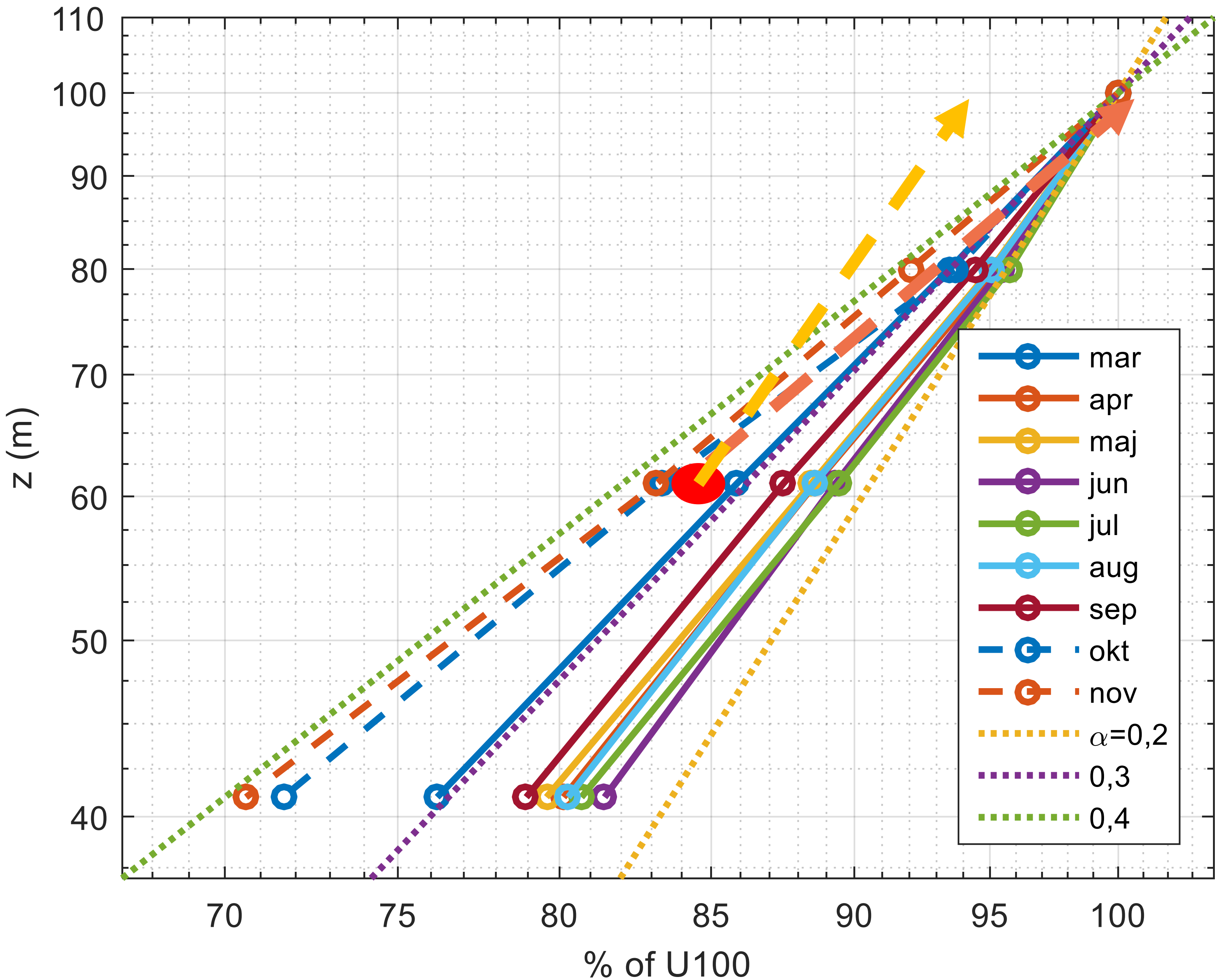
That this would go unnoticed is less likely, but illustrates the strong influence that the presence of even small amounts of erroneous data has. Also, detecting periods with zero wind speed (or constant) is much easier than finding less obvious error periods with, say, much too low but not constant wind speed.


hub height obs
transform data
establish dependence
other data

The idea is to use some other set which has data where the hub height data set has holes. We cant just throw in the other data without doing something to the first to make them represent the hub height. (Of course, if they already represent the hub height we could just throw the whole hub height data set away and replace it by the new data so there wouldnt be an issue in the first place.)

To adapt the new data to hub height we may establish a relation between the hub height data and the new data, Then, we use the relation to transform the other data we want to put in the hole. By inserting the transformed other data into the hole we have a complete hub height data set. But is it correct?

The weakness here lies in that we determine the relation between hub height and the other data from data sets deifferent from the ones that we want to transform. We are for example using summer data to establish a transformation for winter data. Needless to say, its not possible to do it any other way and its is not possible to quantify the error – not based on the data at hand anyway.



110
100
90
80
70
60
50
40
z (m)
70
75
80
85
90
95
100
% of U100
mar
apr
maj
jun
jul
aug
sep
okt
nov
α=0,2
0,3
0,4

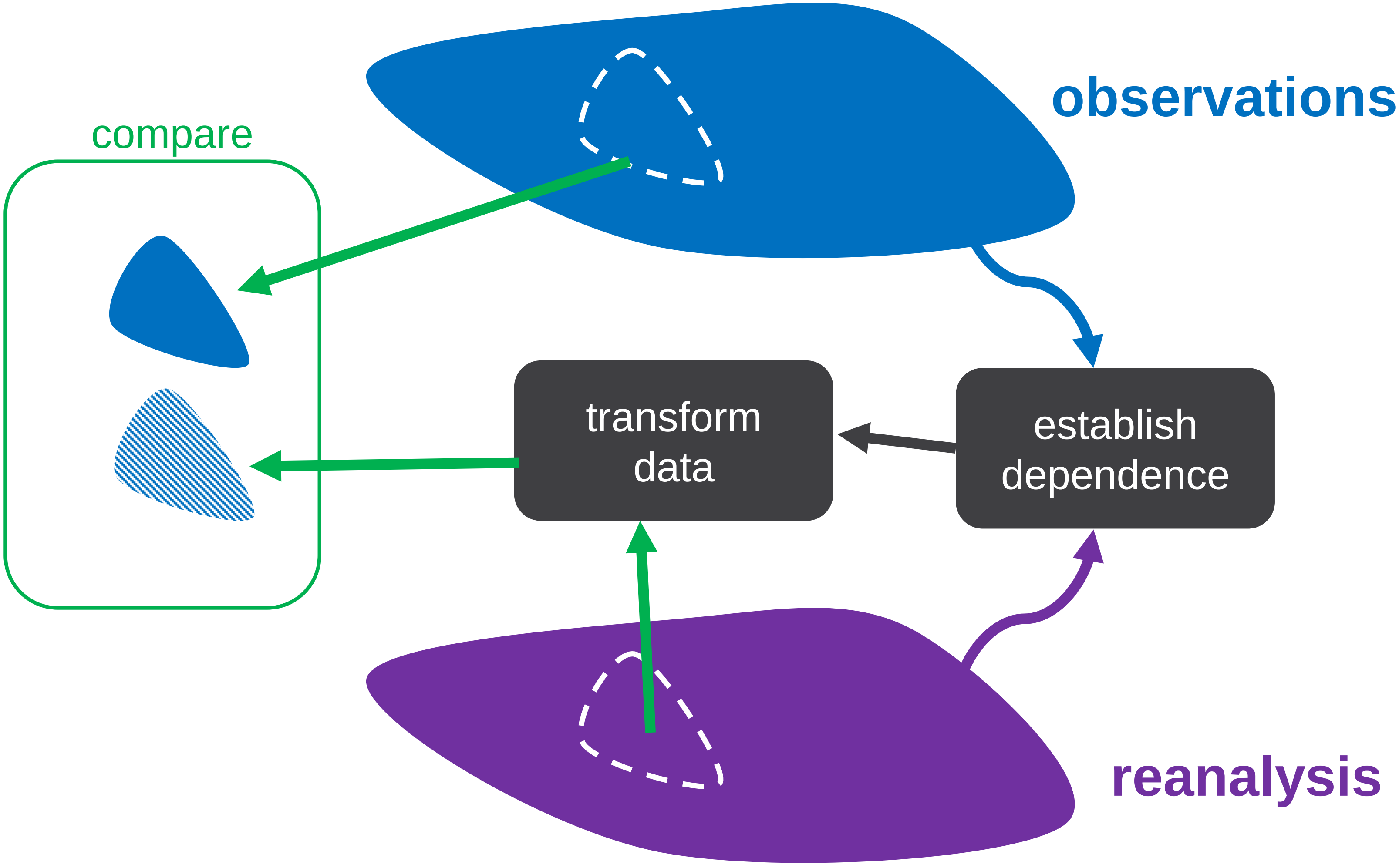
The diagram shows average monthly wind profiles. The wind speed at 100 m has been set to 100 % for easy comparison. We can see a clear pattern in wind shear with lower shear suring summer and higher during fall. (Three profiles corresponding to power model α of 0,2 0,3 and 0,4 are showed for comparison.) Due to icing december-february are missing constituting 25 % of the year. Clearly, we cant just ignore this.

Assuming that the 60 m instrument has been operating correctly during the missing months, we could move this data to 100 m and substitute for the missing observations. However, we dont know the wind shear during this period but we know that is changes a lot depending on the season.

The yellow line indicates what happens if we move the 60 m data (mean for dec-feb, red dot) to 100 m assuming a summer shear, $\alpha \approx 0,2$, compared with an autumn shear, $\alpha \approx 0,4$ respectively. The extrapolated 100 m value will be 95 % and 100 % respectively which is a difference of 6 %.

It is probably wiser to use the autumn value than the summer one, but the point here is to illustrate that since data is missing we are unable to determine an accurate transformation and thus we must guess and end up with an uncertainty.

If we had lots of data we could do some experiments. For this purpose, we need regular continuous SYNOP data and some other data that we could trasform and replace. Reanalysis data are fit for this purpose.



observations
compare
transform data
establish dependence
reanalysis

2016-02-07

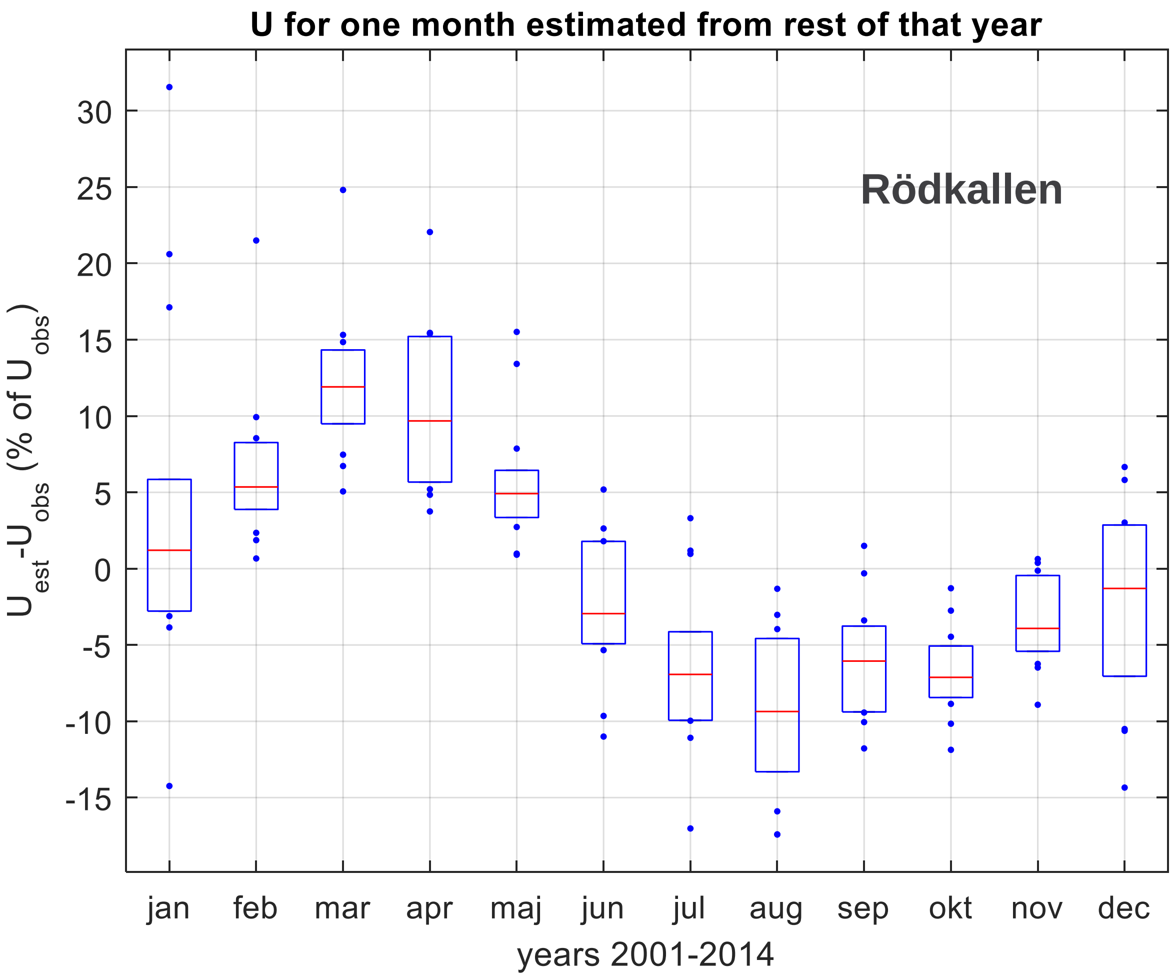
If we have a complete set of observations, i.e. without holes, we could do some experiments.

The idea is illustrated above. First, we take out a subset of the observations, a month for example. Then, we establish a relation between the other data, reanalysis data here, and the corresponding observations for all months except the one we removed.

For this purpose, SYNOP data are suitbale as observations since there are long and reasonably continous data sets. As ”other data” that we could transform and substitute into the hole, reanalysis data are good. (One should always regard reanalysis and other numerical data sceptically since they can be very bad descriptions of reality. However, in wind resourcing, one will likely use them anyway later in the process of long time adjustment. The present use will add a smaller amount of uncertainty tha the long time adjustment process.)

The reanalysis data exist for the whole period (year) and dont suffer from icing. After transforming the reanalysis data, instead of inserting them into the hole in the observation dataset, we compare them with the observations that we took from the hole.

If our method was successful, the subsets should compare well. Lets see…


U for one month estimated from rest of that year
Rödkallen
$U_{est}-U_{obs}$ (% of U_{obs})
30
25
20
15
10
5
0
-5
-10
-15
jan
feb
mar
apr
maj
jun
jul
aug
sep
okt
nov
dec
years 2001-2014

We have14 years of obs and reanalysis data. We took out january from the obs dataset and performed as just described. We then got 14 obs mean values and 14 transformed reanalysis mean values to compare. We do this by calculating the difference, Uest = mean of transformed reanalysis montly average and Uobs = mean of observations. We express Uest-Uobs as a percentage of Uobs.

We then proceed in the same way for february, getting 14 differences. March and so on to december.

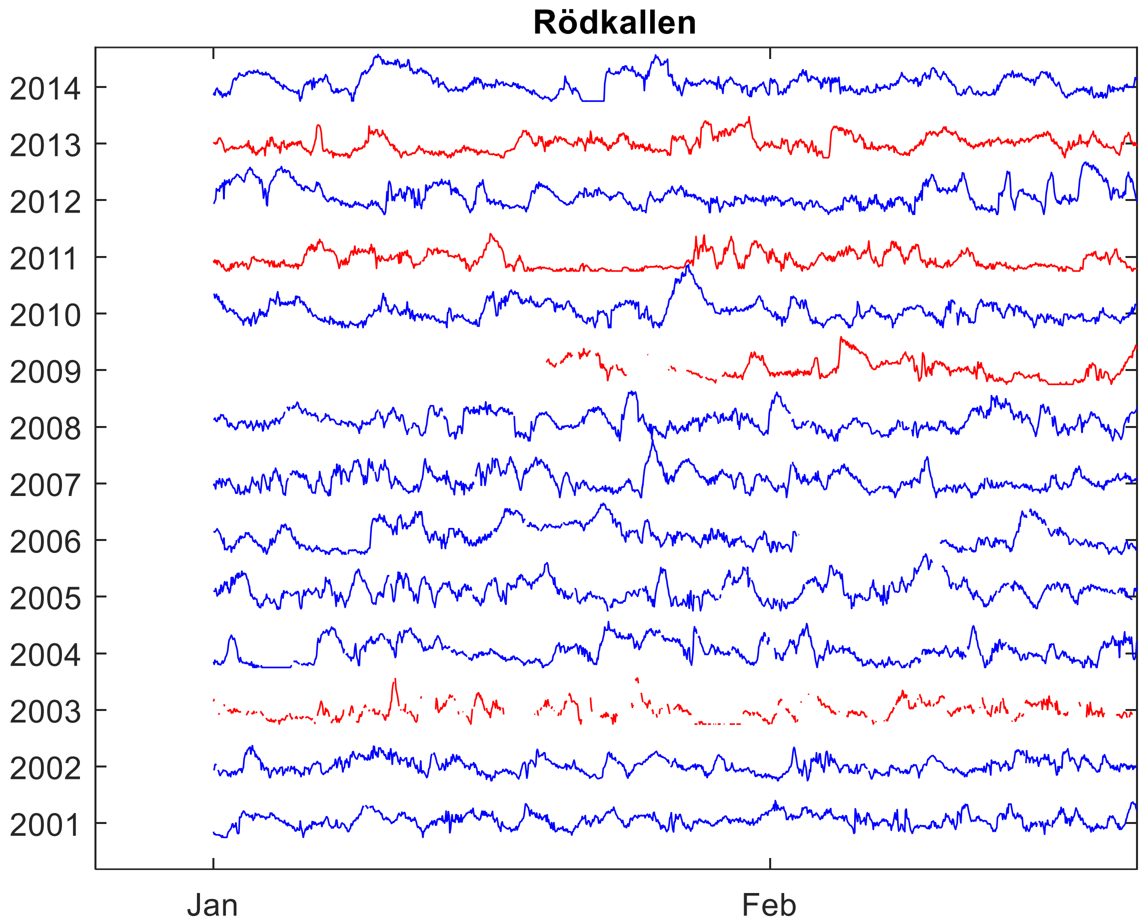
The result is seen in the graph above. The red line for each month is the median. The box contains 7 of the 14 values and the dots indicate the other 7 values.

For january, the median is a percent or so above zero. Thus, the method was good in the median sense. The box is centered reasonably on zero too and ±5 % high which is acceptable – remember that we are operating on just one month worth of data for each subset. However, there are a number of outliers of up to 30 %.

The results for march and august are disappointing because of the strong bias. Generally we see a very clear seasonality. We also note that the outliers are most pronounced in the winter months.

This appearance was to be expected since we transform data from a particular month with data from all other months, some of which are not at all representative for that month,

SWECO

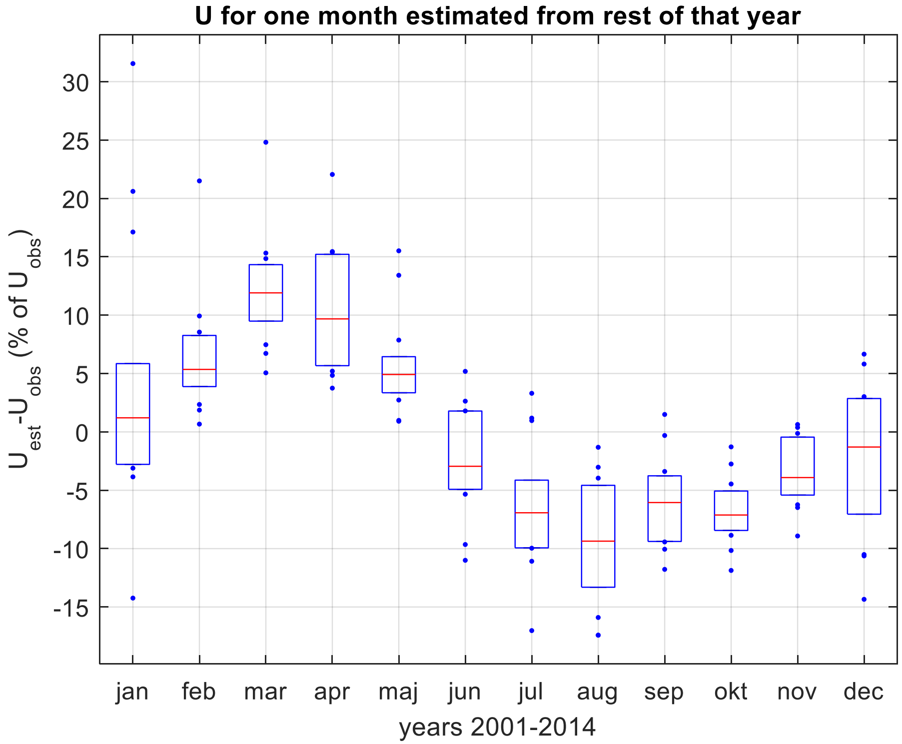

Rödkallen
2014
2013
2012
2011
2010
2009
2008
2007
2006
2005
2004
2003
2002
2001
Jan
Feb

The largest four outliers for january come from the time series marked red here. Each one of them has some kind of problem: absent or frozen data. So, it is plausibel that the large deviations are due to erroneous observational data rather than sudden large malfunctions of the method.

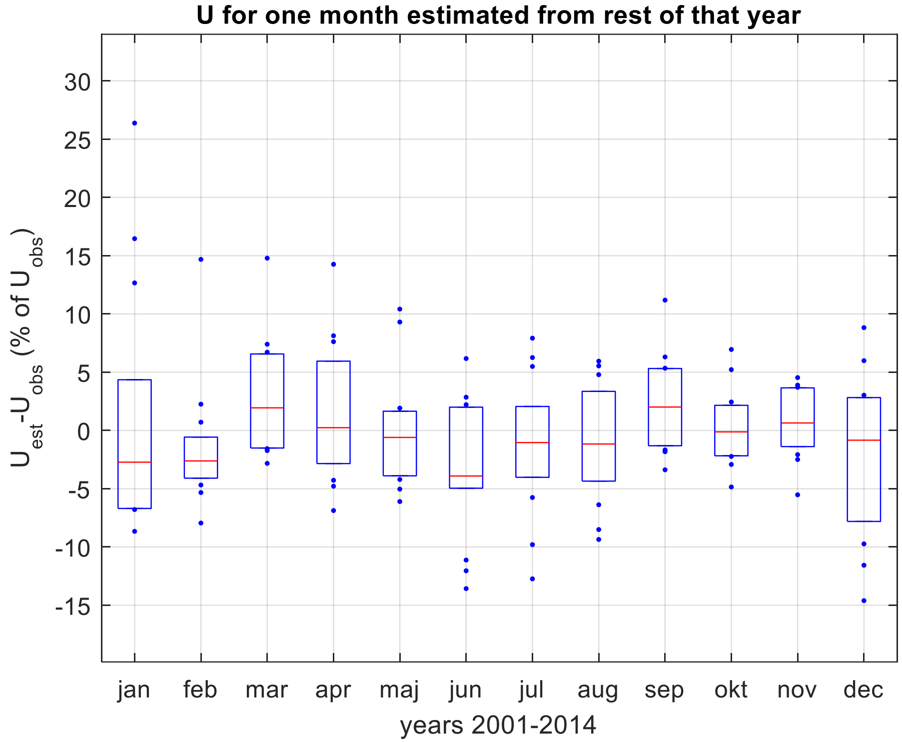
However, something must still be done about the seasonality…

Rödkallen

seasonality ignored



seasonality treated

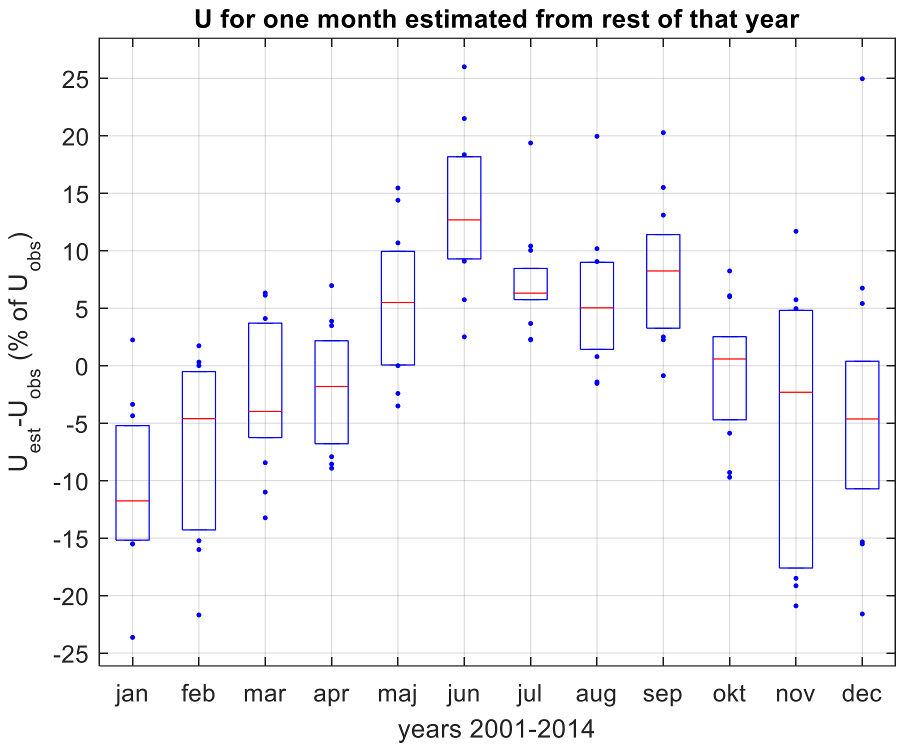


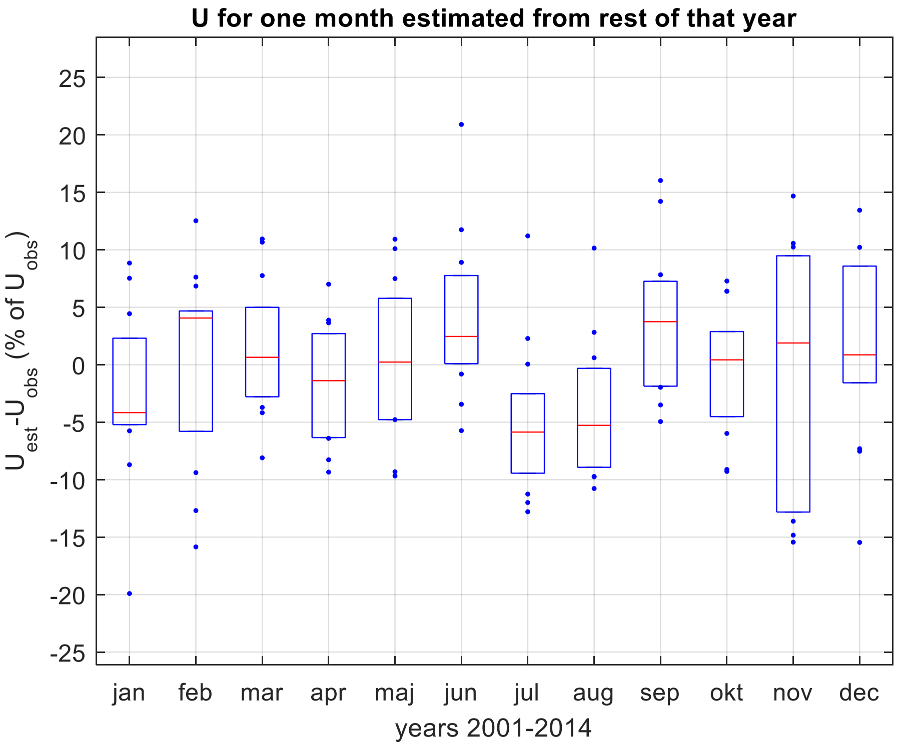
By removing seasonality in the data before establishing the dependence between the observations and the other data (and reinserting it afterwards) we get a much better result. The outliers are still present of course, but the bulk of the spread i.e. the uncertainty is much smaller. (The outliers should have been removed by a more careful filtering of the observations.)

Treatment of seasonality is a standard procedure in statistics. Seasonality should be treated in wind resourcing and it definitely must be adressed when trying to assess effects of data holes and attempts to substitute such holes with other data.

SWECO

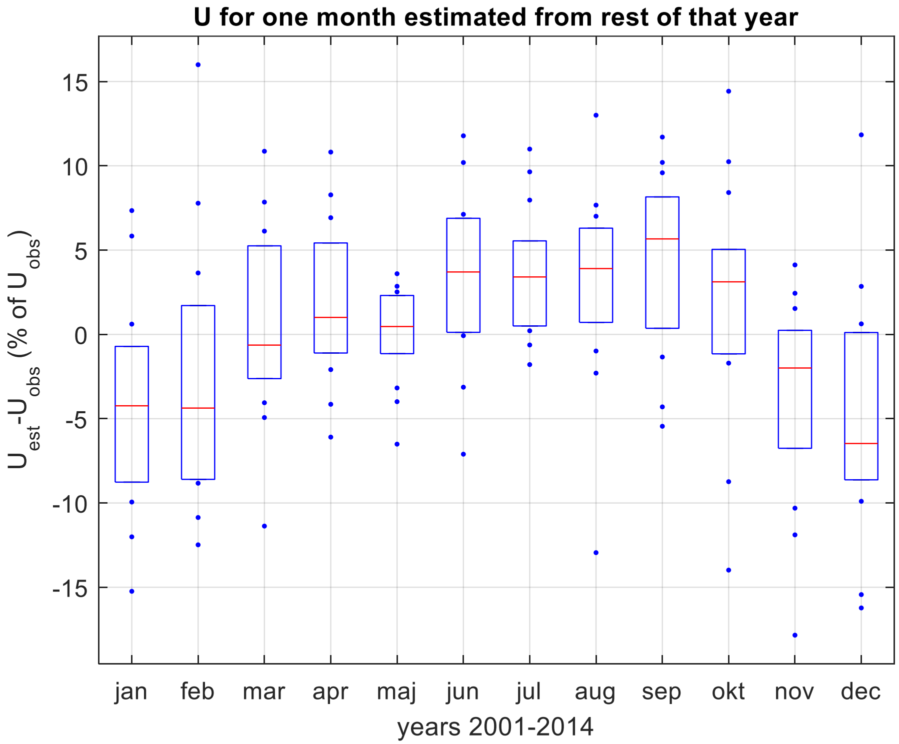
Sylarna



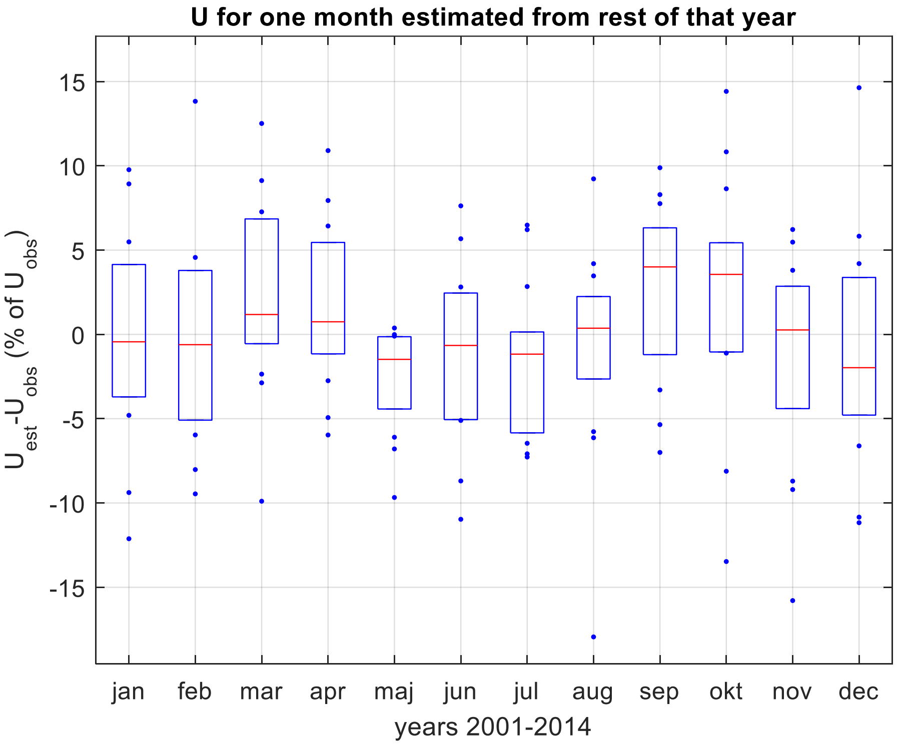


SWECO

Hanö



U for one month estimated from rest of that year
15
10
5
0
-5
-10
-15
U_{est}-U_{obs} (% of U_{obs})
jan feb mar apr maj jun jul aug sep okt nov dec
years 2001-2014



U for one month estimated from rest of that year
15
10
5
0
-5
-10
-15
U_{est}-U_{obs} (% of U_{obs})
jan feb mar apr maj jun jul aug sep okt nov dec
years 2001-2014

complete and correct

data with holes

data with errors

no bias

bias
fill holes
treat seasonality

BIAS
find and remove bad data
treat holes

If you have complete and correct data you have the best foundation for your data wind resource assessment.

If you have correct but incomplete data, there is imminent risk for biasing due to conditioning. If the data holes are randomly distributed, the bias is smaller and possible to estimate. If the data holes are conditioned on something, for example temperature or windiness, there will be a bias that has to be estimated and if necessary, adjusted for. I have demonstrated how this can be done by using other data that covers the periods with absent data, transforming the other data to fill the holes. It is however strongly adviced to assess and treat effects of seasonality in accordance with statistical practice.

If you have data with errors, it is necessary to take action. Even small amounts of erroneous data can have devastating effects on the quality of a wind resource assessment. It is necessary to find and delete all erroneous data! It is much better to remove some correct data and treat the larger gaps than to risk having erroneous data left in the data set. After having purged your data form errors, you are in the ”data with holes” situation as described just above.