Filtering of raw data

Achim Stahl, RWTH Aachen University, stahl@physik.rwth-aachen.de
August, 18th, 2026

ABSTRACT

This report summarizes the impact of the filters applied to the raw data of the seismic sensors prior to their analysis.

Introduction

Data from the seismic sensors are usually filtered before they are analysed. These filters might introduce biases. Here I present a brief study of these biases.

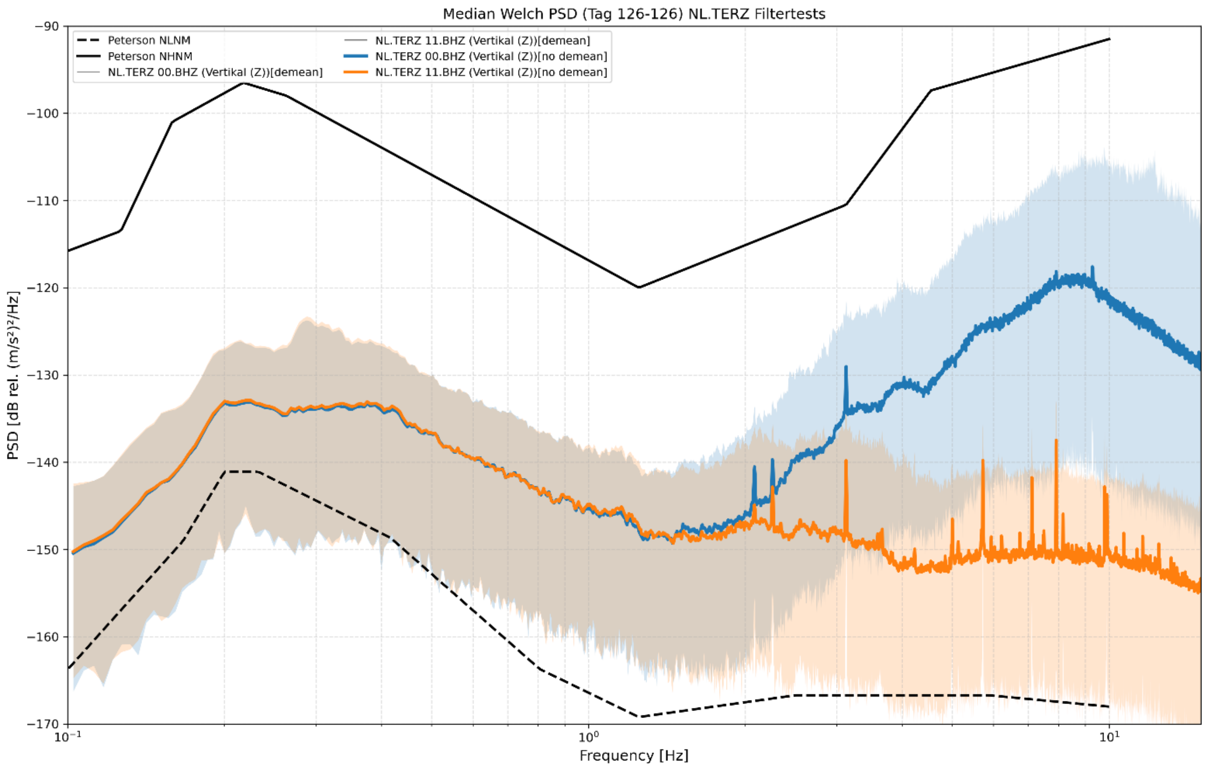
I use data from the Terziet sensor. This is the cleanest data available at this time. A single day of data is sufficient for the analysis. For no particular reason, I selected Wednesday, May 6th, 2026, the 126th day of the year. I analysed the z-component only. No additional information is expected from the horizontal components. I use Welch's method for the estimate of the spectral density with a segment length of 102.4 seconds, i.e. 4096 samples per segment for this data sampled with 40 Hz, and an overlap fraction of 0.5.

I use the obspy framework (www.obspy.org) with scipy in a jupyter lab. I study four different filters. They are applied to the traces of the seismometers before converting them into the Fourier domain. Technically, the filters replace the raw data with the filtered traces.

a. The demean filter (tr.detrend("demean"))
   This filter calculates the mean value of the whole trace and then subtracts it from the trace, i.e. it removes a constant offset from the trace.
b. The linear detrend filter (tr.detrend("linear"))
   This filter goes one step further. It fits a linear function (slope) to the trace and subtracts the resulting line from the trace, i.e. it corrects for an offset that is changing linearly in time.
c. The taper (tr.taper(TAPER) with typical value TAPER = 0.05)
   The taper smoothes the start and end of a trace to avoid artifacts in the Fourier spectrum, for example a sharp end of a trace would generate high-frequency components in the Fourier spectrum. A window is applied to the trace so that the amplitude starts at zero and increases uniformly to its full value over the window and a corresponding window is applied at the end of the trace. There are different window functions available. The default is the Hann function, which is based on a $\cos^2 t$. The parameter TAPER gives the length of each of the two windows as a percentage of the total length of the trace.

d. The bandpass filter (tr.filter("bandpass", ...))
The trace is Fourier transformed. Then a bandpass filter is applied and the trace is transformed back into ordinary space. Various parameters need to be specified, for example the lower and upper end of the accepted band (parameters freqmin and freqmax) or the steepness of the filter.

a. Demean



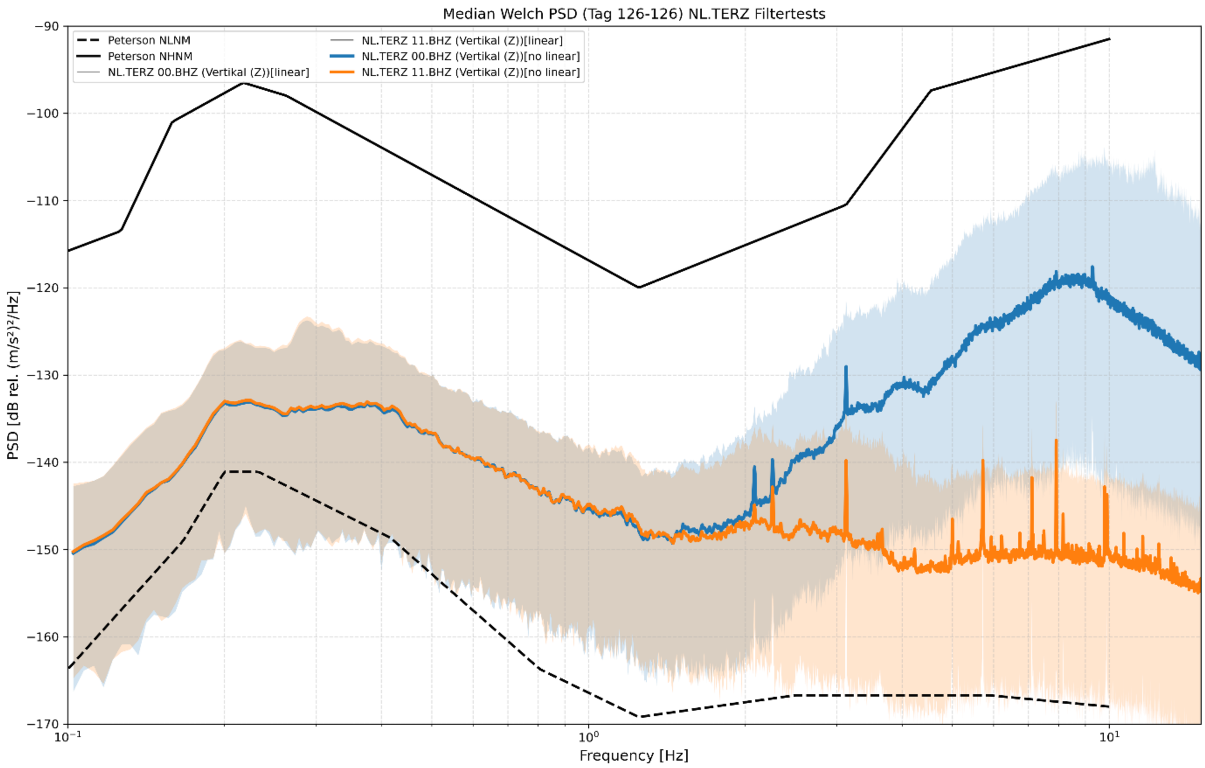
The figure shows the noise spectra of that day at Terziet measured with the surface sensor (blue) and the downhole sensor at 250 m depth (orange line). The solid line gives the median, the shaded areas represent the 90% percentiles (5% to 95%). The black lines show the Peterson limits. Noise should mostly fall between those two lines. The data has been processed with typical filter values:

a. Demean was applied.
b. The linear detrend filter was applied.
c. The trace was tapered with TAPER=0.05
d. A bandpass filter with a window from $f_{min} = 0.05$ Hz to $f_{max} = 15$ Hz with four corners, describing the order of the Butterworth filter, was applied.

The data was reloaded and reanalysed without subtracting the mean (no demean). The result is plotted in gray lines behind the blue and orange lines. It is barely visible. There is no significant impact on the data.

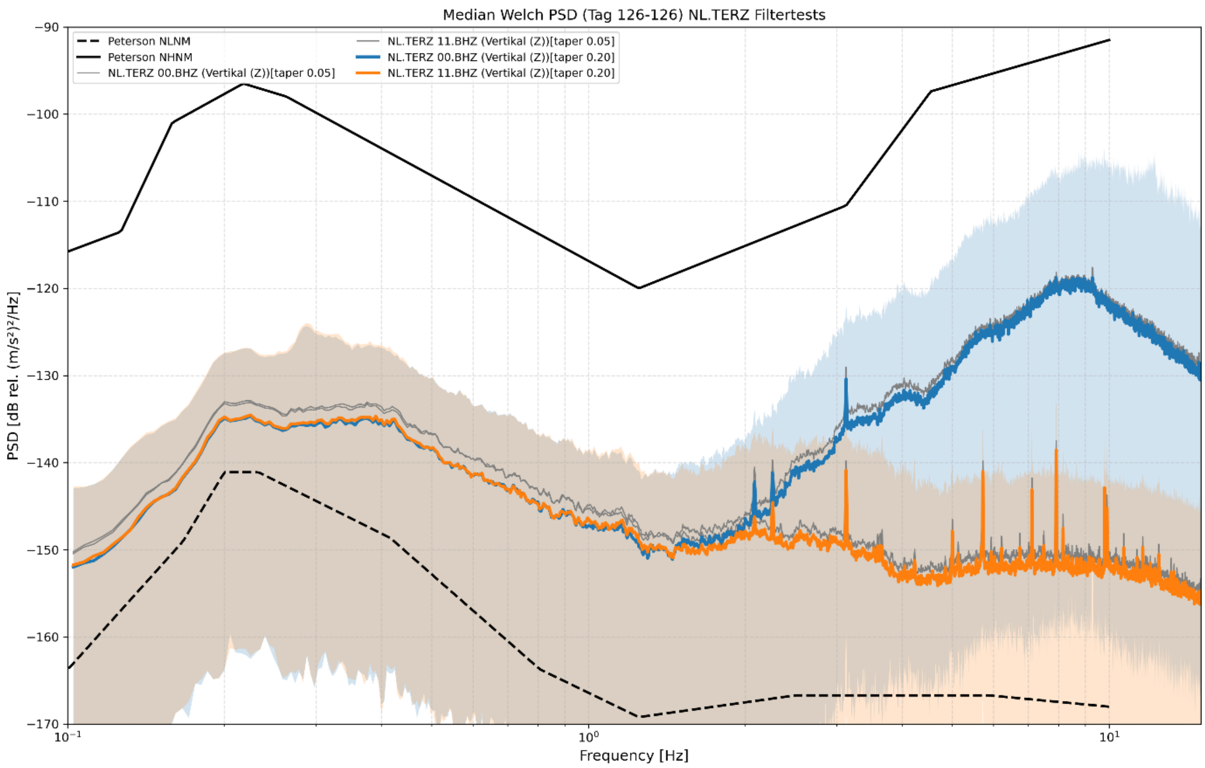
Recommendation: The computational effort of the filter is low. I recommend to keep it, even if it has limited impact. It might give some robustness against more problematic data.

b. Linear detrend



The linear filter has no visible impact on the data either.

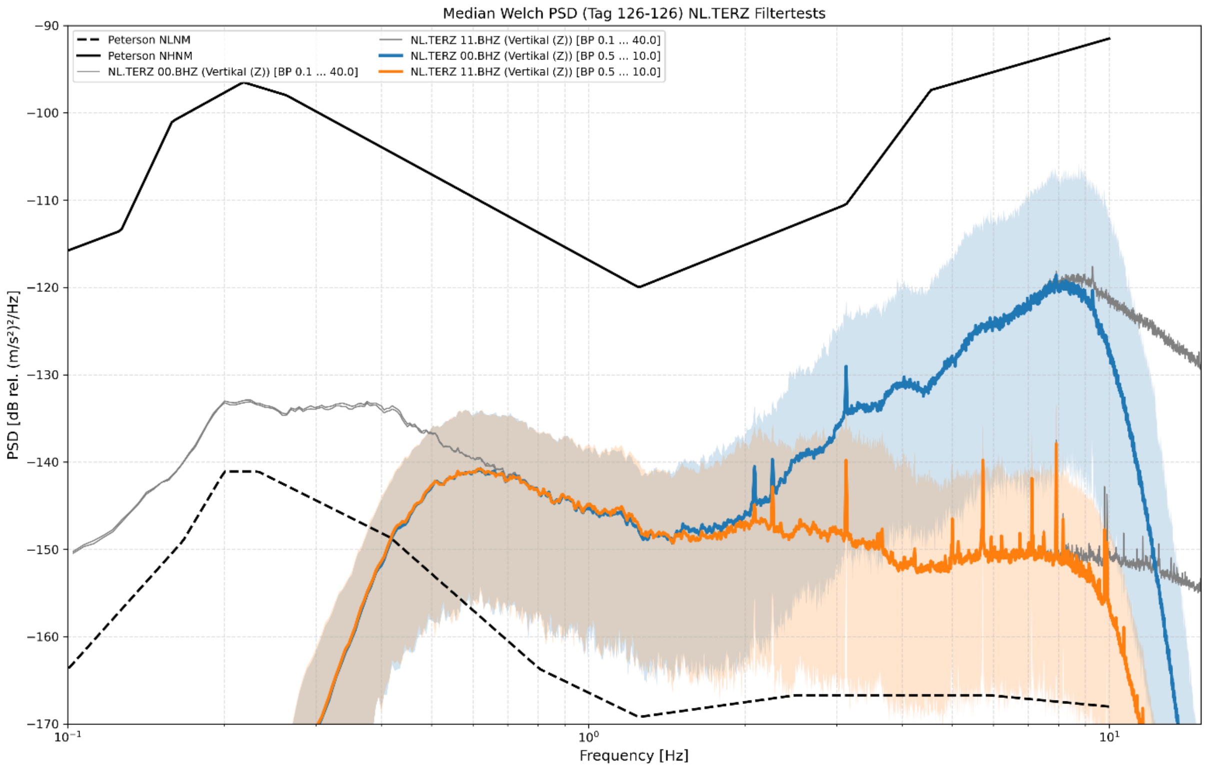
c. Taper



To study the impact of the taper, I increased the tapered windows from 5% of the trace to 20%. This is quite a significant increase. The impact on the spectrum is visible. There is a slight reduction of the overall noise level due to the strong taper. It goes back to an inconsistent use of the taper. The noise in the first and last hour of the day is reduced by the smoothing of the taper, but this data is still included in the average of the noise of the data. A proper procedure would add data of a certain length T_w before and after the time to be analysed and taper those. One should not include the tapered part of the trace in the averages. If I reduce the tapered fraction to 10%, there is hardly any difference of the median to no taper visible, but a strong impact on the lower 5% percentile remains, which is clearly visible in the comparison to the plots in section a. and b.

Recommendation: Adding time before and after the relevant data is the correct procedure, but it might be cumbersome. As long as the taper is below 10%, there is no significant bias on the data.

d. Band pass



Unlike the other filters, the bandpass filter has a clear effect on the spectra, even within the window defined by the bandpass filter's cutoff frequencies. The default range of frequencies accepted by the filter is 0.05 Hz to 40 Hz (gray lines). I compare to a filter tightened to the range of 0.5 Hz to 10 Hz. The impact at the edges – outside of the bandpass window – is dramatic, but the impact clearly extends into the range of accepted frequencies. With an upper cutoff at 10 Hz the impact is visible down to 8 Hz.

The modified frequency range is not chosen arbitrarily. It is the recommendation of our colleagues in Lusatia (see https://www.gpi.kit.edu/329_ET.php). This recommendation creates an unacceptable bias (reduction) of the noise at the upper end of the analysed frequency range of 2 Hz to 10 Hz.

Recommendation: The cutoff frequencies should be chosen carefully. I recommend maintaining a factor-of-2 margin relative to the frequency range under consideration.

Conclusion

Except for the bandpass configured too tightly, the filtering does not introduce any significant biases.