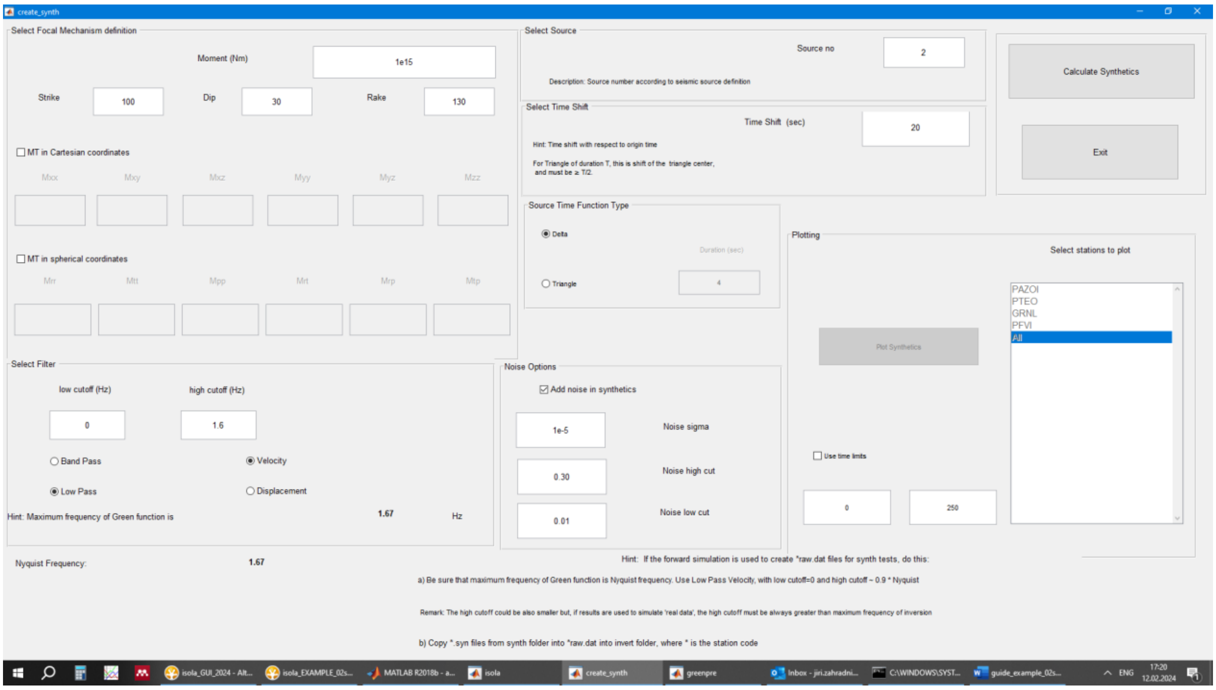
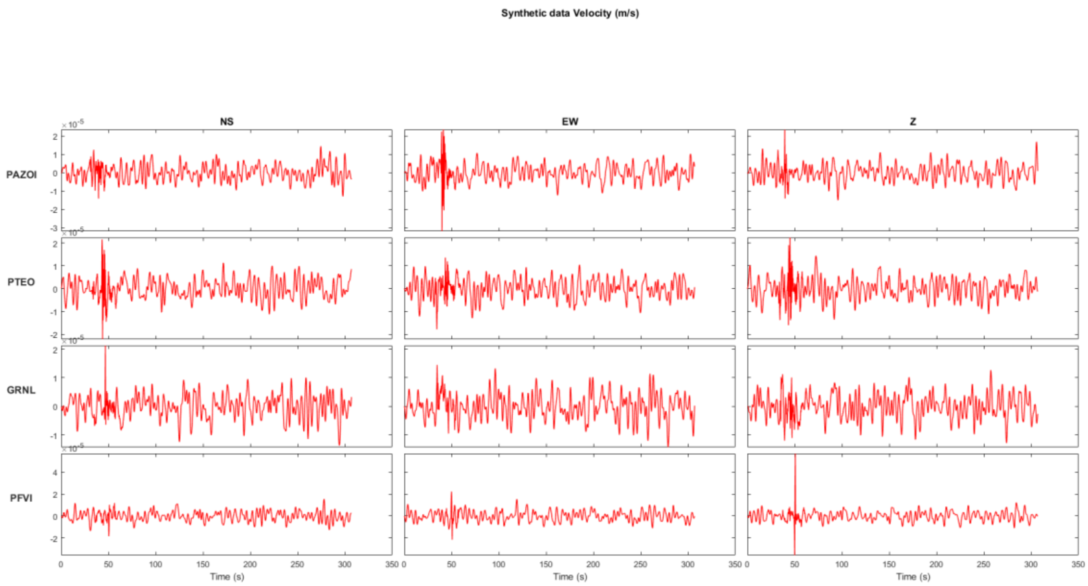
Example 02
Forward simulation for synthetic test with noisy data. Inversion of synthetic data resembling a Portugal shallow Mw 4 earthquake.

Before this, see the introductory hints in isola_EXAMPLE_01.

Preparation of synthetic data
Event Info. Open the tool and see the data. We mimic a real Mw 4 event of 20230107, so we set Lat 37.891, Lon -9.534. We choose the time window length,TWL= 307.2 s, implying waveform sampling with dt= TWL/1024= 0.3 s. Save.
Station Selection. Taken from iberia_4stations.stn from the root folder. Select all, and exit.
Define Crustal Model. Load the velocity model from file crustal.cru from the root. Plot, save and exit.
Seismic Source Definition. Trial source positions, starting at depth 15 km, depth step 5 km, 3 sources (= 15, 20, 25 km). Calculate and exit.
Green function calculation. GF is calculated up to the Nyquist frequency 1.67 Hz, assuming delta-function moment-rate (default values); 3 trial sources, 4 stations. Run. Results are in folder green (gr*hea, gr*hes) and in the invert folder (elemse*) for all trial positions *.

Forward simulation. The tool is accessible from the main Isola menu. See the options in the screenshot copied below (and the 'hints' therein). Source no. 2 means that synthetic waveforms are created for source position 2 at the depth of 20 km, delta-function acting at t=20 s. Code interacts with the user, asking in the system command window how long is the noise window; we request noise to be generated in the whole TWL range (307.2 s) for all stations [in future Isola version this option will be automated]. Further, check how the options were saved in the folder synth, in the files mechan.dat and 99.syn. Choose Plot and you should see the figure saved in synth subfolder as wave_velocity_synth.png, shown below. We created ground-motion velocity data 0 - 1.6 Hz where a strong random Gaussian noise, with a flat spectrum in ground-motion velocity, is present between 0.01 and 0.3 Hz; these values are corners of a Butterworth filter, which means that still some noise is present out of this interval. The velocity waveforms appear as *raw.dat in synth folder, for all four stations *. I copied manually these data also into a spare subfolder data_Isola_raw_synth. This is a good practice with which we avoid the situation that some files are possibly deleted, by mistake. Later, *raw.dat files will be used in the synthetic inversion test exactly in the same way as we normally use real instrumentally corrected records (velocities, *raw.dat). Caution: Code has no repeatability (random seed number). Thus, your results may slightly differ numerically from the example.


create_synth
Select Focal Mechanism definition
Moment (Nm)
1e15
Strike
100
Dip
30
Rake
130
MT in Cartesian coordinates
MT in spherical coordinates
Select Source
Source no
2
Description: Source number according to seismic source definition
Select Time Shift
Time Shift (sec)
20
Calculate Synthetics
Exit
Source Time Function Type
Delta
Triangle
Plotting
Select stations to plot
PAZOI
PTEO
GRNL
PFVI
All
Select Filter
low cutoff (Hz)
high cutoff (Hz)
0
1.6
Band Pass
Low Pass
Velocity
Displacement
Hint: Maximum frequency of Green function is
1.67
Hz
Noise Options
Add noise in synthetics
1e-5
Noise sigma
0.30
Noise high cut
0.01
Noise low cut
Use time limits
0
250
Nyquist Frequency:
1.67
Hint: If the forward simulation is used to create *raw.dat files for synth tests, do this:
a) Be sure that maximum frequency of Green function is Nyquist frequency. Use Low Pass Velocity, with low cutoff=0 and high cutoff ~ 0.9 * Nyquist
b) Copy *.syn files from synth folder into *raw.dat into invert folder, where * is the station code


Synthetic data Velocity (m/s)
NS
EW
Z
PAZOI
PTEO
GRNL
PFVI
Time (s)

Inversion in the almost noise-free band

Copy *raw.dat files from data_Isola_raw_synth into the invert folder.

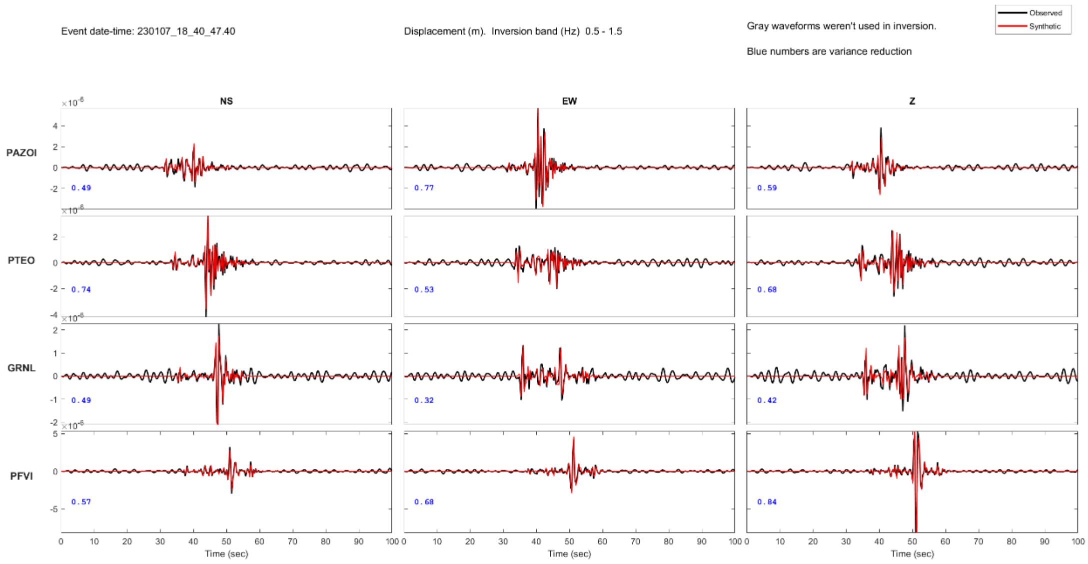
Waveform Inversion tool. Set frequency range 0.5-1.5 Hz, which should be almost noise-free, and invert for full MT; time search 15-25 s with step 0.3 s. Use standard Isola (no covariance, “cova” matrix). In the invert folder, the results appear in inv1.dat and inv2.dat. Your results may have a small difference due to the random seed, mentioned above.

Retrieved is Source position 2 (the one which was prescribed),
and time 20.1 (which is nearest to the prescribed value within the time search step)

moment (Nm): 9.5029084E+14
moment magnitude: 4.0
VOL % : 0.5
DC % : 99.2
CLVD % : -0.3
strike,dip,rake: 102 29 133
strike,dip,rake: 235 68 68
varred= 0.6969212 (variance reduction)

As already briefly commented above, the small noise remaining above 0.3 Hz caused a small deviation from the prescribed moment 1.0e15, VOL=CLVD=0 and from s/d/r = 100/30/130.

Plot Results tool provides the waveform plot which I manually saved in subfolder EXAMPLE_02syn\invert\0.5-1.5Hz_without_cova



Inversion in a broad frequency band (including noise) without cova
Go again back to Waveform Inversion, and now invert the same data in the high-noise range 0.01-1.00 Hz. See saved results in folder invert, files inv1.dat and inv2.dat. The result is totally wrong:

moment (Nm): 1.7642834E+15 … (instead of 1.0e15)
moment magnitude: 4.1
VOL % : 39.6 … instead of 0
DC % : 8.0 …..instead of 100
CLVD % : -52.4
strike,dip,rake: 87 66 69 … instead of 100/30/130
varred= 3.1453967E-03 … (variance reduction is extremely low, waveforms are not fitted at all)

Moreover, the inversion is so unstable that this result can significantly change due to random seed. Again, you can compare your results with those saved in subfolders of invert.

Inversion in a broad frequency band (including noise) with cova, iteration 1
After the previous inversion, we obtained residual wavefield res = obs-syn, i.e., observed minus synthetic waveforms. Now we repeatedly go to the Waveform Inversion tool and run with covariance matrix, the Toeplitz type. which is calculated from the residual. Frequency range 0.01-1. Hz, inversion for full MT. See your results in folder invert, e.g. inv1.dat, inv2.dat, copied into subfolder EXAMPLE_02syn\invert\0.01-1.00Hz_with_covaFULL\iteration1

Source position 1 instead of 2
moment (Nm): 7.8230440E+14
moment magnitude: 3.9
VOL % : 31.0
DC % : 45.7
CLVD % : 23.3
strike,dip,rake: 103 54 89
varred= 1.3955832E-03

The result is still far from the true mechanism. In file inv1.dat we can see

isour,ishift,corr,moment,DC%,str,dip,rak,str,dip,rak

1	64	0.030459	0.7823E+15	45.742	103	54	89	284	35	91
2	67	0.029138	0.9061E+15	80.636	100	30	131	235	67	68
3	69	0.016811	0.2811E+15	54.071	276	29	67	122	62	102

Isour is trial source position and Ishift is time shift expressed as a number of steps dt; the other parameters are selfexplanatory.

Although the preferred solution (row 1, the largest correlation 0.030 in column 3) is wrong, a solution much closer to the correct one is in row 2, featuring just a small difference in the correlation value 0.029. It means that if in practice we were sure with depth (position 2), the wrong solution (in position 1) would be avoided, but that would be rarely the case. If the depth is not known precisely, which is the practical case, we should prefer position 1 so we have failed with MT.

Despite this negative result, we can obtain a better solution in the next iteration.

Inversion in a broad frequency band (including noise) with cova, iteration 2

The previous calculation created new synthetic data (syn(1)). We take a new residual res = obs-syn(1), and, still in Waveform Inversion tool, 0.01-1. Hz, we again choose Run, with the same cova type (Toeplitz). [Now we ignore the warning that the preceding calculation must be made in standard isola, because it was relevant only for iteration 1]. Check your results in invert folder and compare them with the saved files EXAMPLE_02syn\invert\0.01-1.00Hz_with_covaFULL\iteration2.
We obtain inv1.dat like this:

Source position 2 … (correct, at the input model value)
moment (Nm): 8.9889337E+14 …. close to 1e15
moment magnitude: 3.9
VOL % : 2.8 … close to 0
DC % : 86.0 … close to 90
CLVD % : 11.2
strike,dip,rake: 98 29 130 close to 100/30/130
varred= 7.9447031E-04

Although variance reduction 7.9e-4 is still extremely low, we have obtained an almost exact solution.

Warning: Inspection of file inv1.dat shows

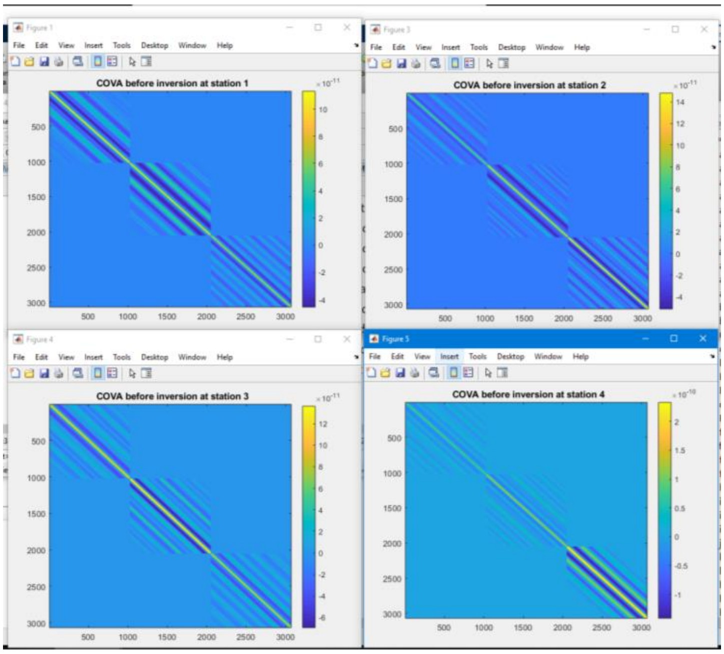
1	64	0.024181	0.6850E+15	17.514	108	52	91	285	37	88
2	67	0.028058	0.8989E+15	86.015	98	29	130	233	68	70
3	69	0.016725	0.2876E+15	55.417	279	31	69	122	61	101

Again, we find almost the same correlation value (column 3) in positions 1 and 2, so it was rather a matter of ‘good luck’ that this time the code formally preferred position 2 with the correct solution. However, if you make further iterations, you confirm the solution near the correct one. I am not sure how much this is still ‘good luck’ or a true improvement due to iterations and this particular example.

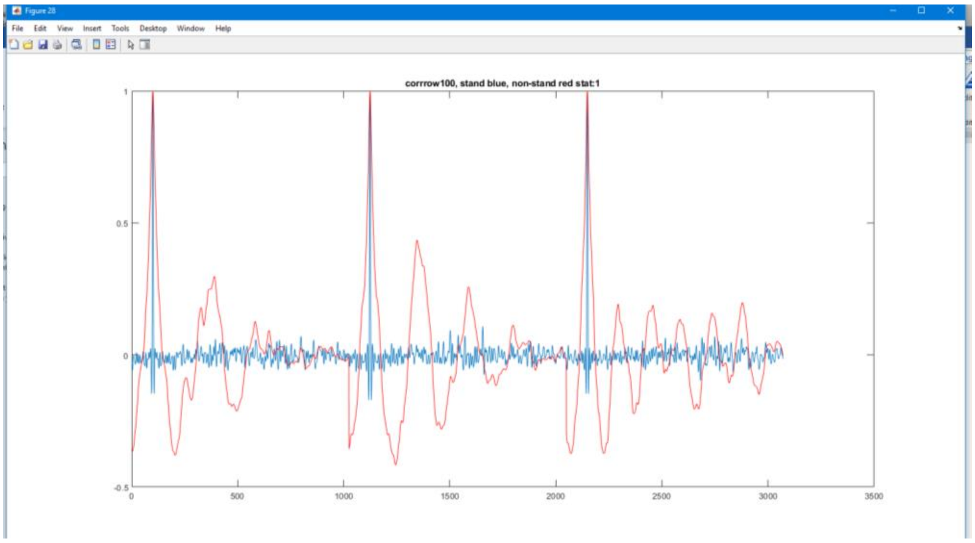
Why did we obtain the correct solution? How is it supported if not known in advance?

We can be satisfied that the inversion with cova provided the correct solution under extreme noise conditions in this synthetic test. In a real data application, it would be difficult to accept the solution because variance reduction is very low, no fit can be seen when comparing obs and syn waveforms. Here we try to answer the questions from the title. To this goal, we discuss the results of iteration 2 in detail as follows.

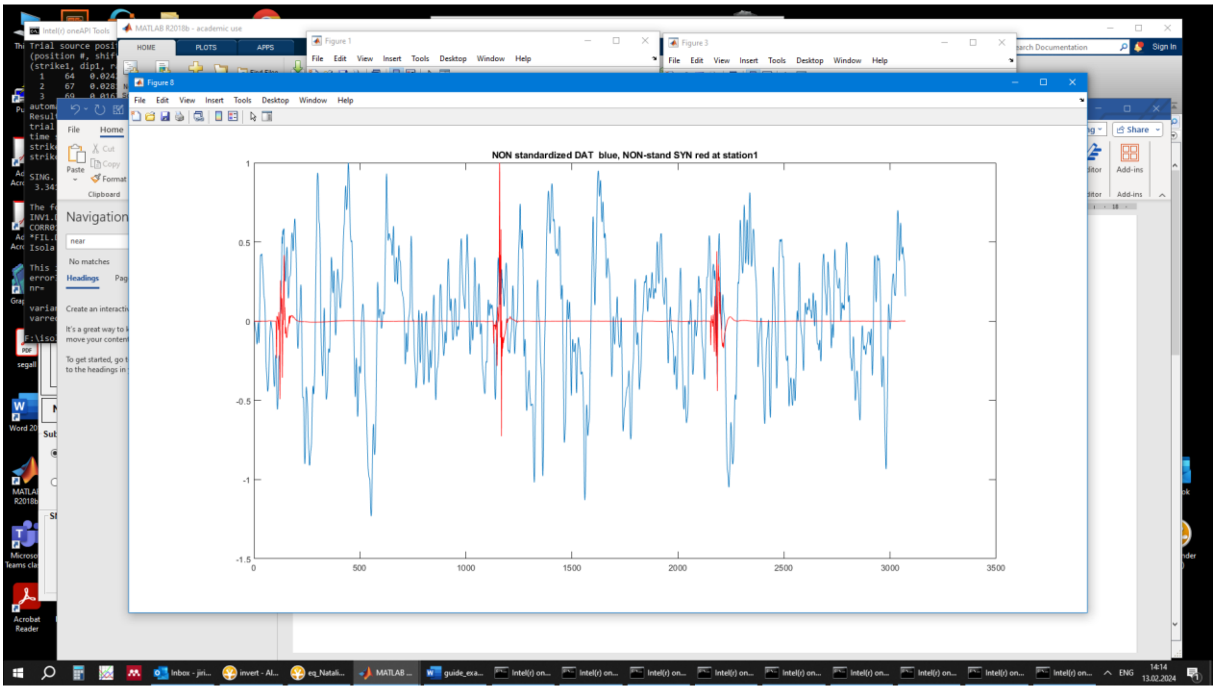
Many plots appear on the output. Here we copy plots of the covariance matrix (of res) for the four stations.



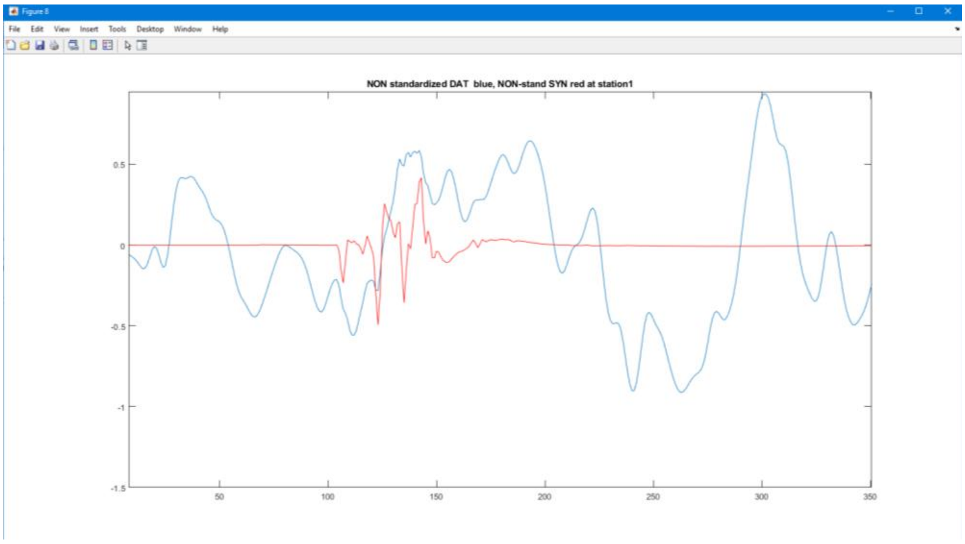
From the top left to the bottom right in each panel, are the N, E, and Z components. The plots show strong non-diagonal elements, i.e. a highly correlated data residual error (res). Inversion with cova suppresses the correlated part of data error. It is demonstrated below by the comparison between one row of the cova matrix Cd based on res (red line), and another matrix of so-called standardized residuals, Sres (blue line). Sres=res * L^T, where L^T is Choleski factor of Cd^{-1} (eqs. 13-15 of Vackar et al., 2017). From the left to right are N, E, and Z autocorrelations, respectively, for station 1 (PAZOI). The blue curves are almost delta-like autororrelations (concentrated near the zero lag), indicating strong suppression of the error correlation in the standardized residuals.



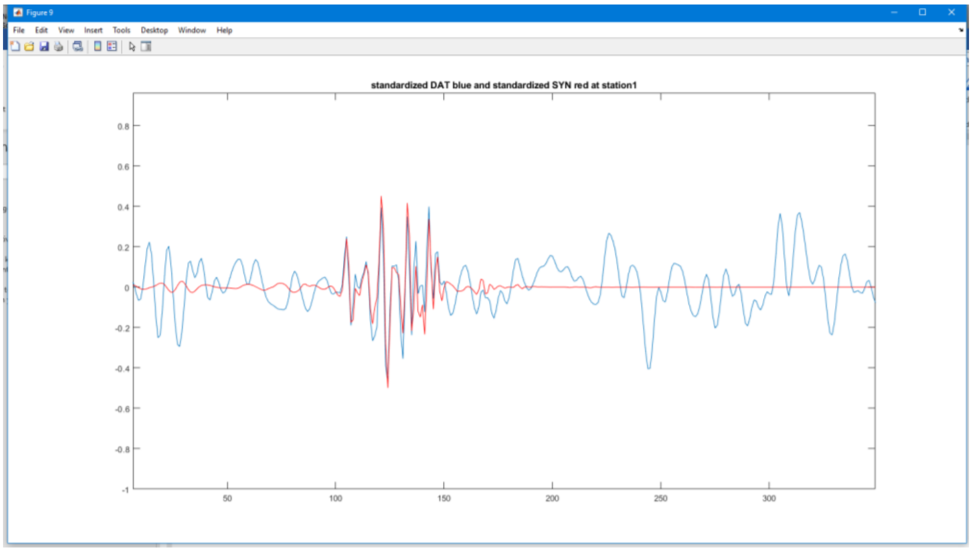
In the time domain, again for station 1, we further compare obs (blue) and syn (red) waveforms before standardization, i.e. corresponding to the previous res with high noise with long-large error correlation; again - from left to right is N, E, and Z.



Here we zoom in on station 1, the N component, and see a very poor misfit between data obs (blue) and syn (red).

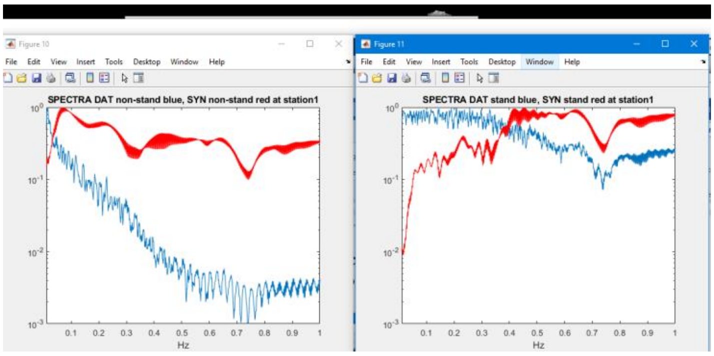


The same but with standardization shows a significantly better fit.



Numeric output in Matlab command window confirms the improvement by a great (100-times !) increase of VR from varred_classical = 0.00079385 to varred_standardized = 0.1063.

In the spectral domain, again for station 1 (spectrum of displacement calculated from all 3 components), before and after standardization (left and right), both normalized to its maximum we observe this:



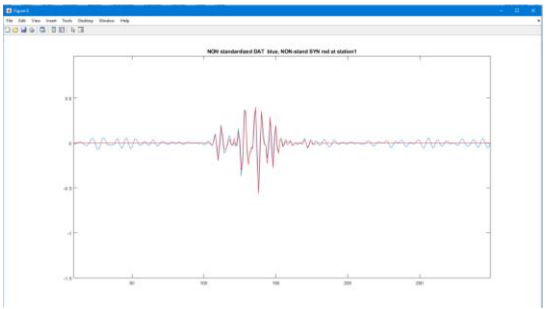
These plots show that when inverting without cova (left), data power of displacement (blue) grows toward zero frequency due to the (correlated) noise at f< 0.5 Hz [we remind the reader that synthetic noise was created as flat in velocity for 0.01 and 0.3 Hz]. With cova (right), the observed and standardized data spectrum (blue) is ‘whitened’, becoming closer to a constant, which is a spectrum of a delta-like autocorrelation function (uncorrelated noise). The whitening acts as a relative suppression of standardized data in the noisy range f<0.5 Hz. The spectrum of synthetics (red), with cova (right) is also strongly suppressed at f< 0.5 Hz. Yet there is not a good fit between the blue and red curves in the right panel. This will be improved more below in this text.

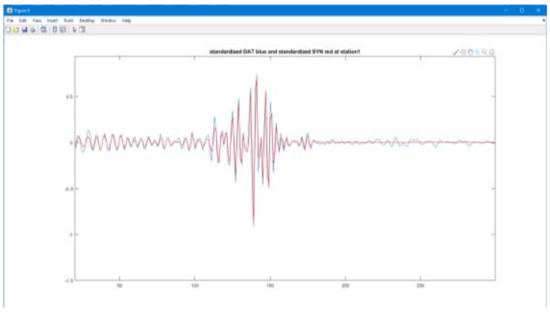
Lesson learned for practical inversion.
Imagine that in real application we obtained a similar plot as in the last spectral comparison. The standardized synthetic spectrum (red curve in the right-hand plot above) indicated that the main noise was present in the frequency band where the red spectrum was decreased by standardization, i.e. f< 0.5 Hz. It suggests that even a standard inversion (without cova) may be useful at f > 0.5 Hz, where the noise is small. This is exactly what we did in the previous test [Inversion in the almost noise-free band, 0.5-1.5 Hz], and is also what a skilled user would do after quickly testing a few frequency ranges and not having codes with cova.

Or, we can make a two-step inversion restricted to the range 0.5-1.0 Hz, first without and then with cova. Here is the result after iteration 1 with cova (Toeplitz type). As expected, we obtain almost exact solution in this noise-free range:

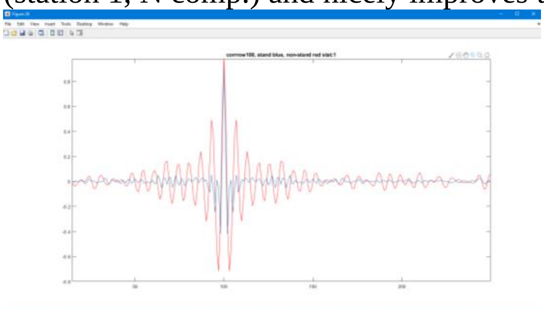
```
moment (Nm):  9.7001836E+14
  moment magnitude:   4.0
         VOL % :   3.0
          DC % :  93.0
        CLVD % :  -4.0
 strike,dip,rake:       102        28        132
 strike,dip,rake:       235        69         69
 varred=  0.7762061
```

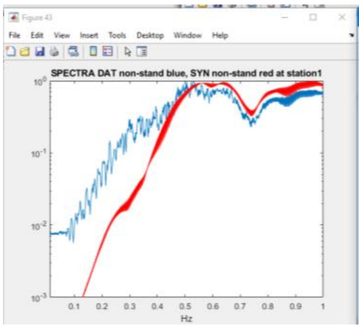
Standardization slightly improves VR: from varred_classical = 0.7762, to varred_standardized = 0.8376. This is demonstrated below on station 1, N component.

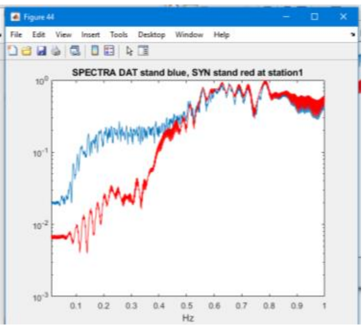




The effect of cova is small but not negligible, it clearly narrows the autocorrelation function (station 1, N comp.) and nicely improves the spectral fit at f> 0.5 Hz.

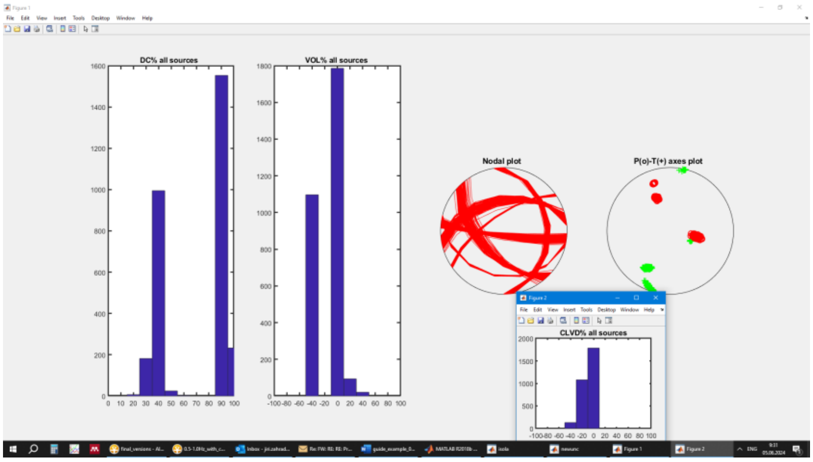


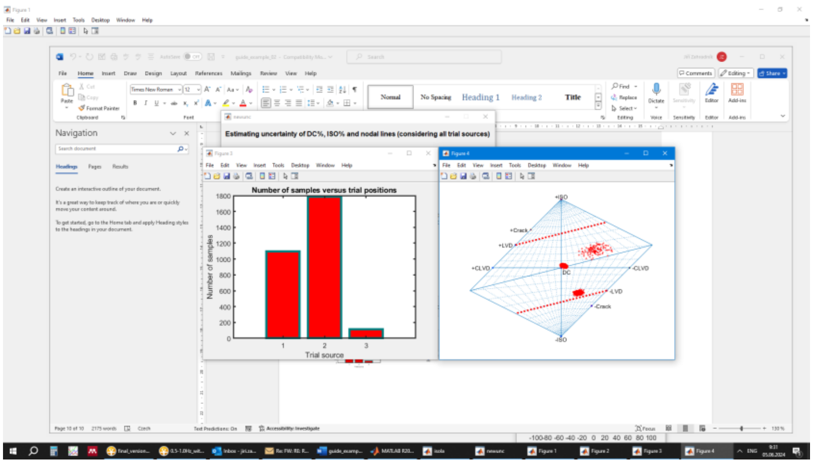




Uncertainty Estimate

The next step is to study MT uncertainty which is jointly determined by waveform fit and the posterior covariance of MT $Cm= (G^T Cd^{-1} G)^{-1}$ at all trial source positions.
Using the tool Uncertainty Estimate we find a possibly surprising result, see below.





We do not obtain a single ‘cloud’ of MTs near the best-fit solution. Each trial depth produces its own “cloud” whose size (number of the random samples) is determined by the red histogram. In this case, the non-uniqueness of the solution is strongly caused by the uncertain depth, because each depth prefers very different non-DC components. Besides the best-fit position 2 (depth 20 km), with its almost 100% DC (center of the diamond source-type plot), we obtain an important contribution (many MT samples) of a strongly non-DC mechanism at position 1 and position 3 (the fewest samples). The number of samples (red histogram) is caused by a complex interplay of misfit and the Cm matrix. This effect can be seen in the file newunc_log.dat, in folder newunc; note that misfit varies from 0.3 to 1.1, while the determinant of Cm varies by two orders of magnitude.
These results are saved in subfolder EXAMPLE_02syn\invert\0.5-1.0Hz_with_covaFULL