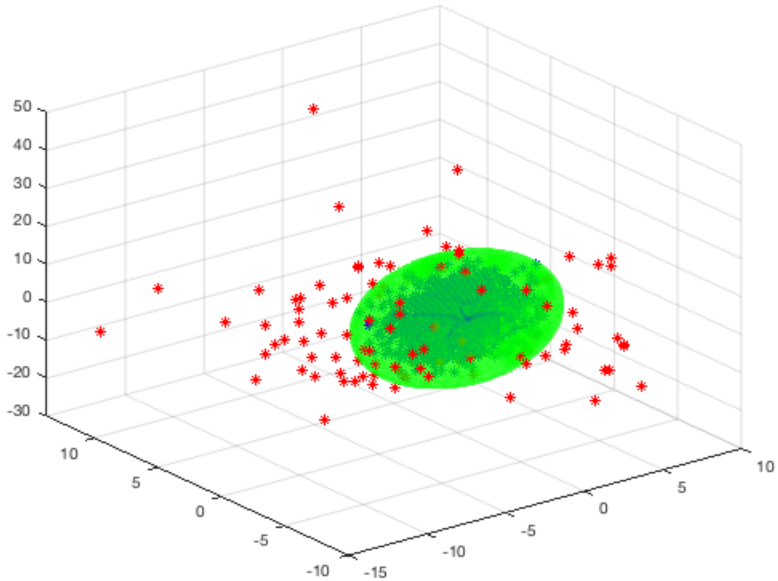
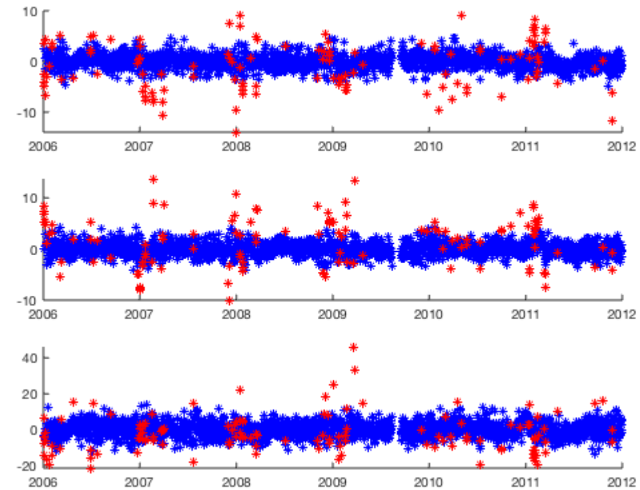
A new 3D approach to automated outlier rejection in GNSS time series



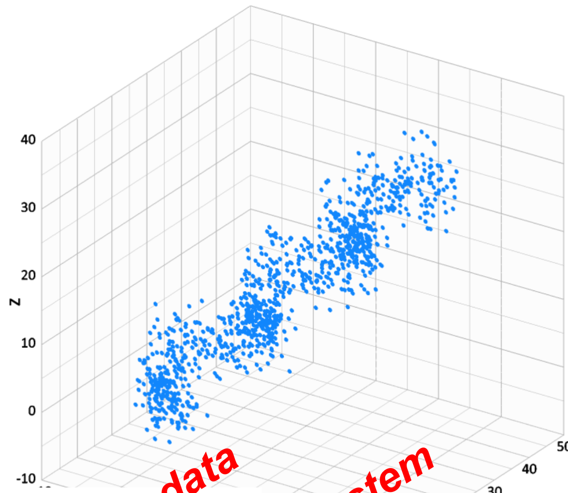


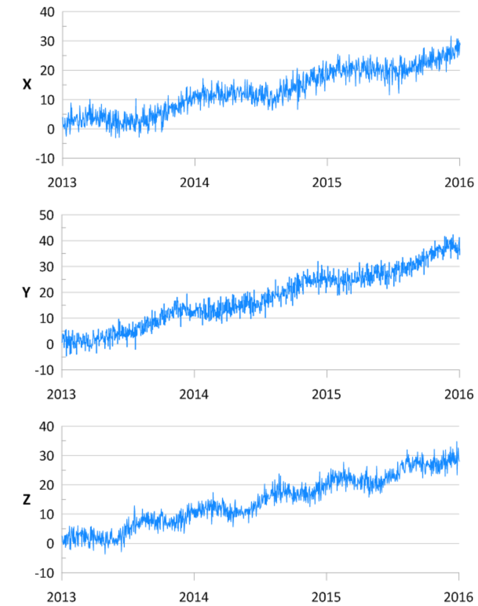
L. Tavasci[1], M. Di Donato[2], M. Errico[2], S. Gandolfi[1], S. Zerbini[2]

[1] DICAM, University of Bologna, Italy
[2] DIFA University of Bologna, Italy
Ref.: luca.tavasci2@unibo.it

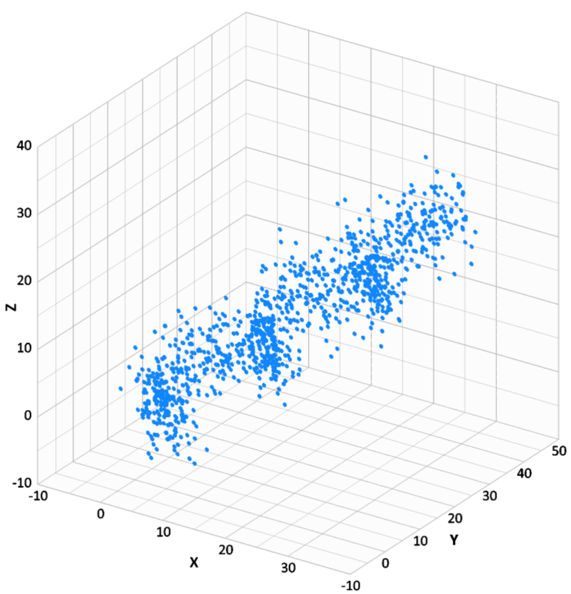
Time-series analysis results usually depends on the chosen RS…

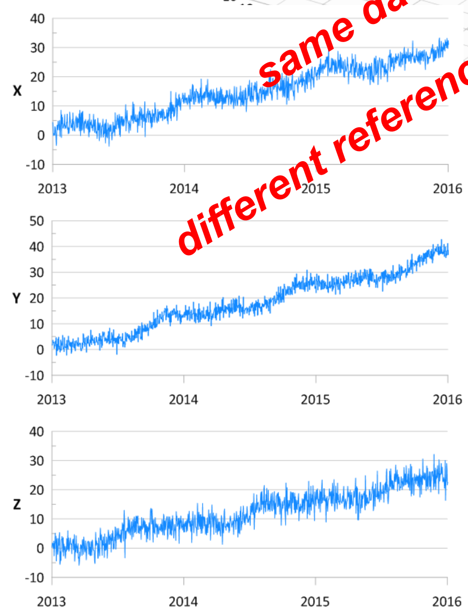
A **GNSS time series** is made of **3D** coordinates. Nevertheless, we usually analyse the three mono-dimensional projections of these coordinates independently.
The **projections depend upon the reference system** used to define the coordinates, typically an ECEF (XYZ) reference system or a topocentric reference system (ENU).





same data
different reference system



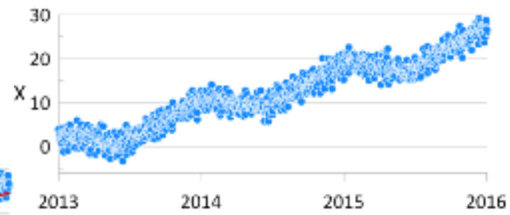


Depending on the used reference system (in particular its orientation), the mono-dimensional time series that we analyse are different and would lead to different results.

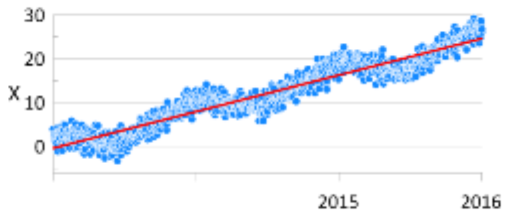
In this work we focus on the detection of outlier solutions. We propose a method which is invariant with respect to the choice of the RS.

…outlier detection phase has to be performed on the residuals with respect to the station positions

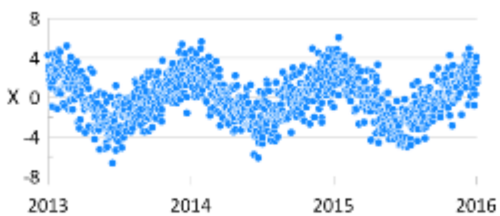
A typical approach for mono-dimensional time series analysis could be:



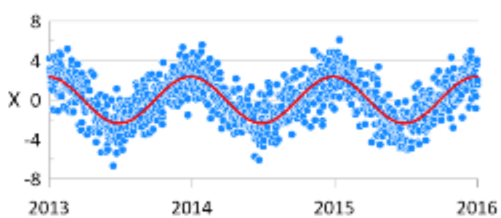
Original data



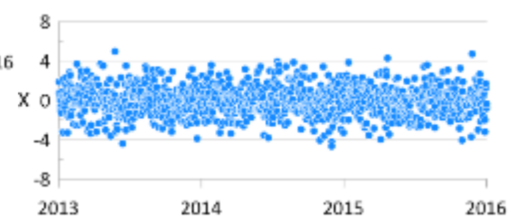
Regression line



Linear residuals

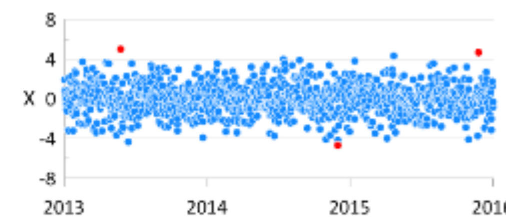


Periodical model(s)



Model residuals

Outliers analysis



Whatever approach is used to outlier-detection, this phase of the analysis should be performed **iteratively** after having removed both linear and non-linear trends

Estimation of the residuals to be analyzed for the outlier rejection

We define:

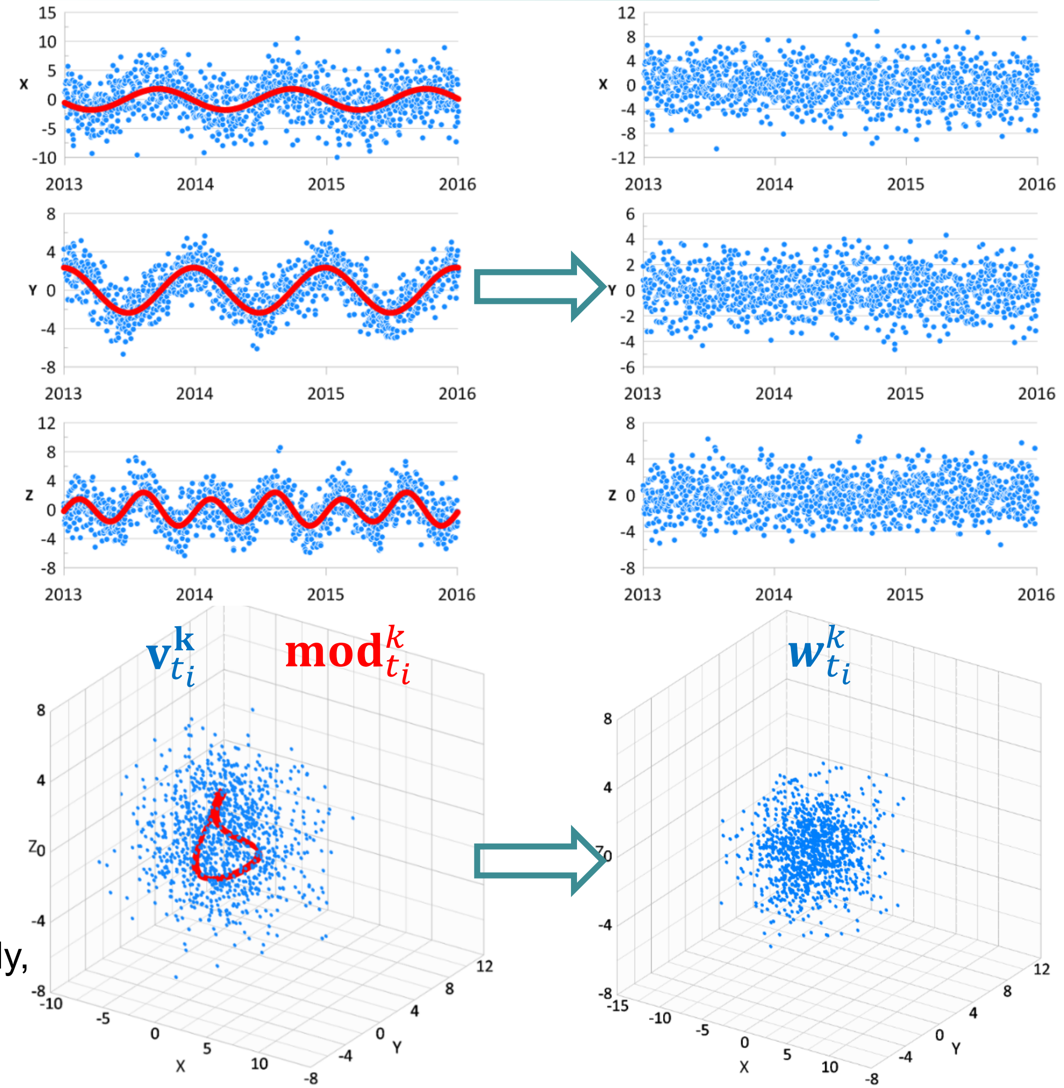
- $i = 1, .., n$ with n number of position solution in of the GNSS time series.
- k = generic reference system.
- $\boldsymbol{v}_{t_i}^{k} = \left[v_{x_{t_i}}^{k}, v_{y_{t_i}}^{k}, v_{z_{t_i}}^{k} \right]$ residuals with respect to the regression straight line.

$$\mathbf{mod}_t^k = \begin{vmatrix} \sum_m \left[A_{xm}^k \sin\left(2\pi f_{xm}^k * t\right) + B_{xm}^k \cos\left(2\pi f_{xm}^k * t\right) \right] \\ \sum_m \left[A_{ym}^k \sin\left(2\pi f_{ym}^k * t\right) + B_{ym}^k \cos\left(2\pi f_{ym}^k * t\right) \right] \\ \sum_m \left[A_{zm}^k \sin\left(2\pi f_{zm}^k * t\right) + B_{zm}^k \cos\left(2\pi f_{zm}^k * t\right) \right] \end{vmatrix}$$

- $\mathbf{mod}_t^k$ example of models of displacement for the monitored position.

$$\boldsymbol{w}_{t_i}^{k} = \boldsymbol{v}_{t_i}^{k} - \mathbf{mod}_{t_i}^{k} = \left[w_{x_{t_i}}^{k}, w_{y_{t_i}}^{k}, w_{z_{t_i}}^{k} \right]^{\mathbf{T}}$$

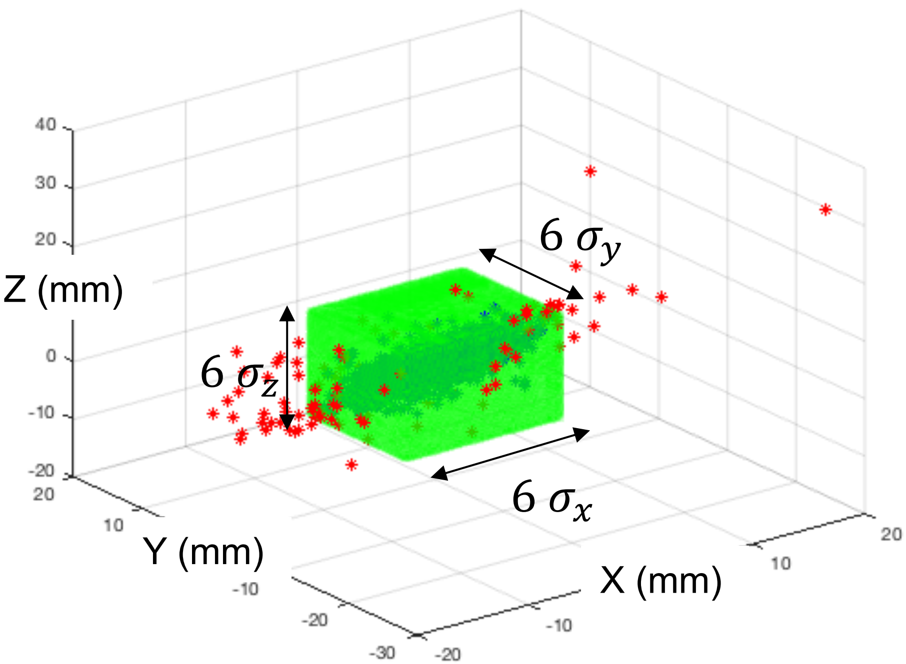
- $\boldsymbol{w}_{t_i}^{k}$ residuals with respect to the models. These should represent measuring errors only, therefore $\boldsymbol{w}_{t_i}^{k}$ can be used for the automated rejection of the outliers →

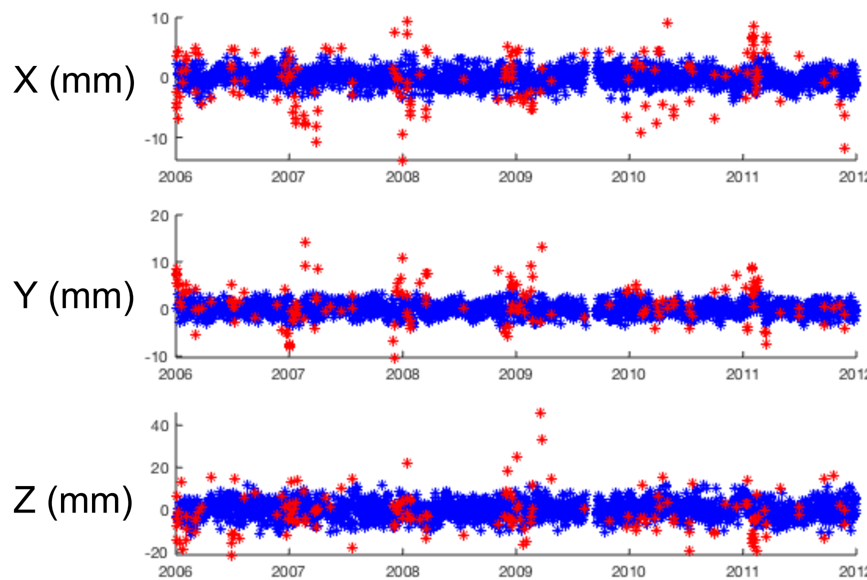


Classical 1-D approach

If we analyse the three mono-dimensional time series of residuals $\boldsymbol{w}_x^k, \boldsymbol{w}_y^k, \boldsymbol{w}_z^k$ separately, the usual approach is to define the STD of each series $\sigma_x^k, \sigma_y^k, \sigma_z^k$ and then fix a confidence interval (usually 3σ) that we use to reject the points laying outside it.

This leads to define a volume of rejection for the residuals which is a parallelepiped aligned to the RS axes.

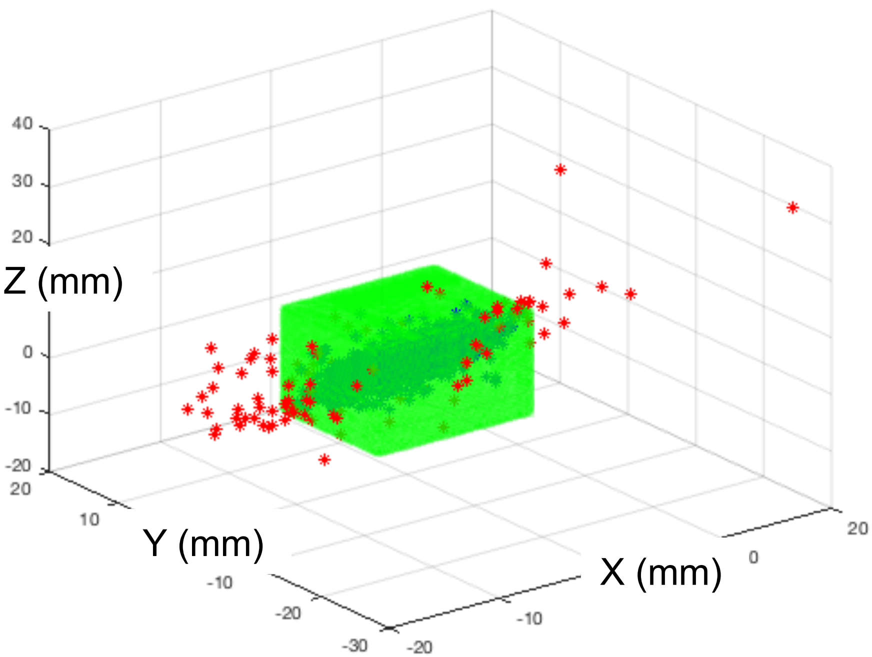


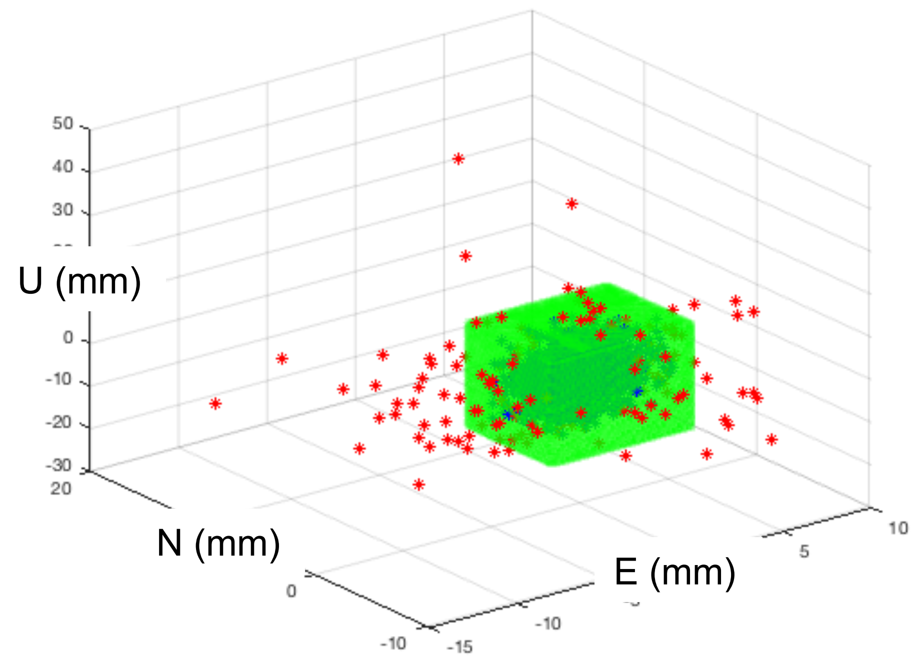


Classical 1-D approach

If we do the same analysis using the ENU geodetic axes instead of the XYZ ones, the result is not exactly the same since the dimension of each volume is different.

The volumes are different because they are defined without taking into account the correlations between the three time series, that change when using different reference systems.





The new 3-D approach

We calculate the covariance matrix of the residuals $\boldsymbol{w}_{t_i}^k$:

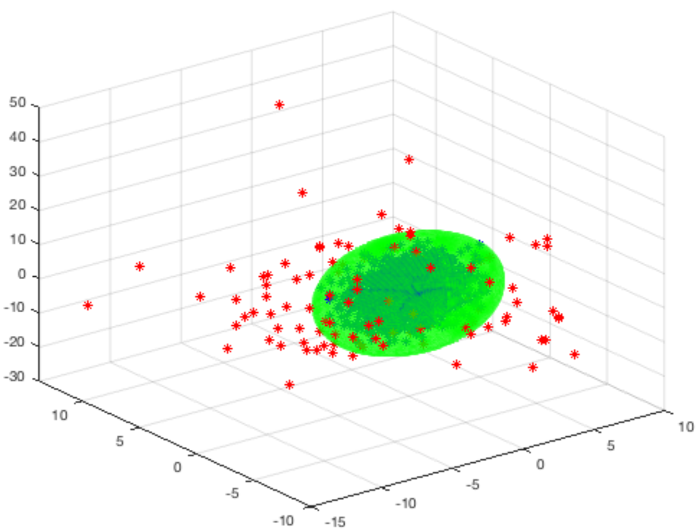
$$\Sigma_{ww}^k = \frac{1}{n}\begin{bmatrix} \sum_i \left(w_{x_{t_i}}^k\right)^2 & \sum_i w_{x_{t_i}}^k w_{y_{t_i}}^k & \sum_i w_{x_{t_i}}^k w_{z_{t_i}}^k \\ \sum_i w_{y_{t_i}}^k w_{x_{t_i}}^k & \sum_i \left(w_{y_{t_i}}^k\right)^2 & \sum_i w_{y_{t_i}}^k w_{z_{t_i}}^k \\ \sum_i w_{z_{t_i}}^k w_{x_{t_i}}^k & \sum_i w_{z_{t_i}}^k w_{y_{t_i}}^k & \sum_i \left(w_{z_{t_i}}^k\right)^2 \end{bmatrix}$$

Then, it is possible to estimate a parameter $\xi_{t_i}^k$ representing the square of the normalized residuals $\boldsymbol{w}_{t_i}^k$:

$$\xi_{t_i}^k = \left[\boldsymbol{w}_{t_i}^k\right]^{\boldsymbol{T}} {\Sigma_{ww}^k}^{-1} \boldsymbol{w}_{t_i}^k$$

Under the hypothesis of normal distribution of the three coordinate components of the residuals w^k the normalized residuals ξ follow a chi-square (χ^2) distribution with three degrees of freedom.

Note that the set of points $\boldsymbol{x}^k$ that verify the equation $[\boldsymbol{x}^k]^{\boldsymbol{T}} {\Sigma_{ww}^k}^{-1} [\boldsymbol{x}^k] = 1$ represent the standard error ellipsoid characterizing the residuals w^k.



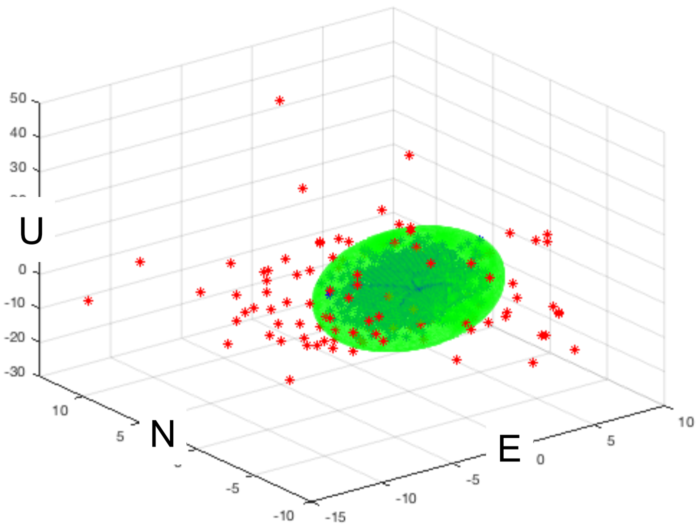
Having chosen a confidence level α a point can be rejected as an outlier if its residual is:

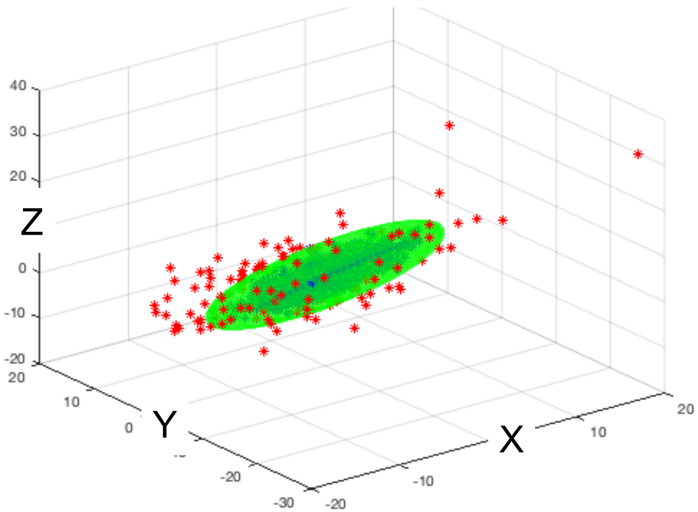
Rejection criteria

$$\max(\xi^k) \leq \chi_\alpha^2$$

$$\left(\xi_{t_i}^k > \chi_\alpha^2 \mid \xi_{t_i}^k = \max(\xi^k)\right)$$

The new 3-D approach





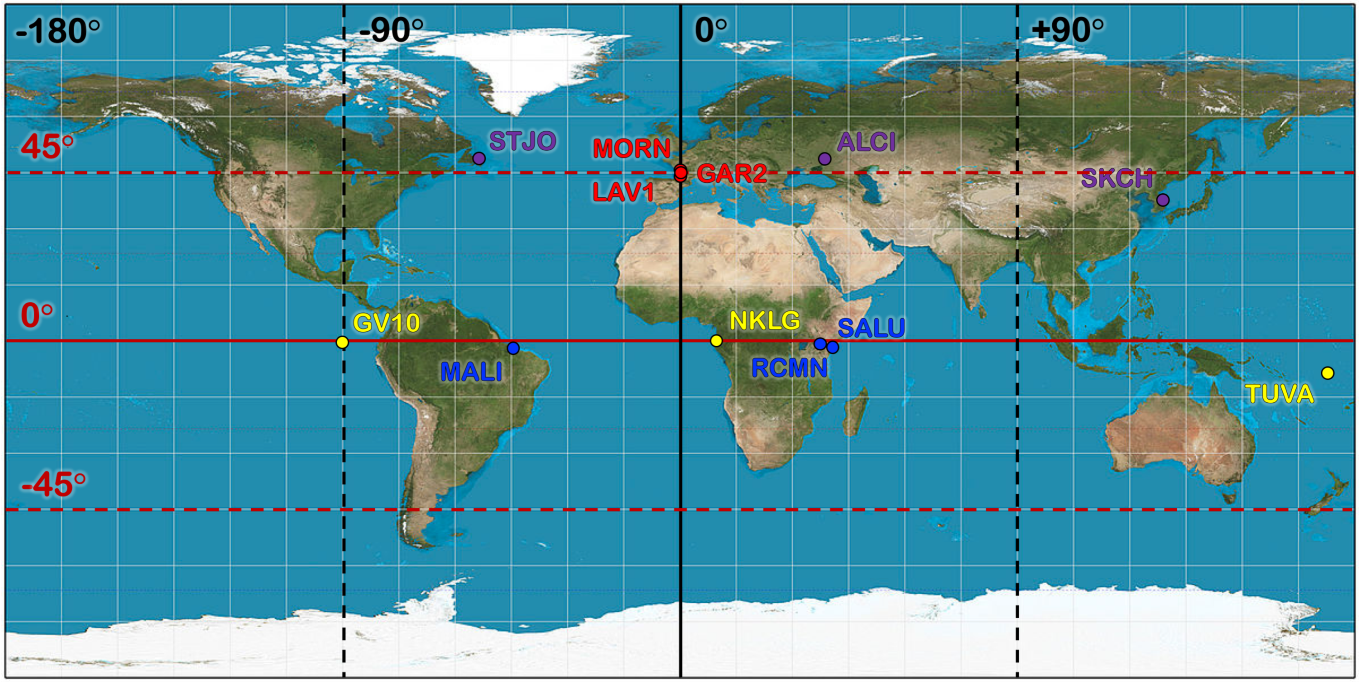
Note that since the ellipsoid is defined using the **full covariance matrix** of the residuals, it is also **oriented in the space** and **invariant in dimensions**.

This means that the 3-D approach to outlier rejection we are proposing is invariant depending on the reference system that is used to represent the coordinates.

IS IT ALWAYS TRUE? …

Test on real GNSS data

DATASET: 6 years of daily position solutions from 12 permanent stations freely available on the Nevada Geodetic Laboratory web site (http://geodesy.unr.edu/). Time series are expressed in both the geocentric reference system (.txyz2 files) and topocentric reference systems (.tenv3 file). For more details see Blewitt et al. (2018)*.



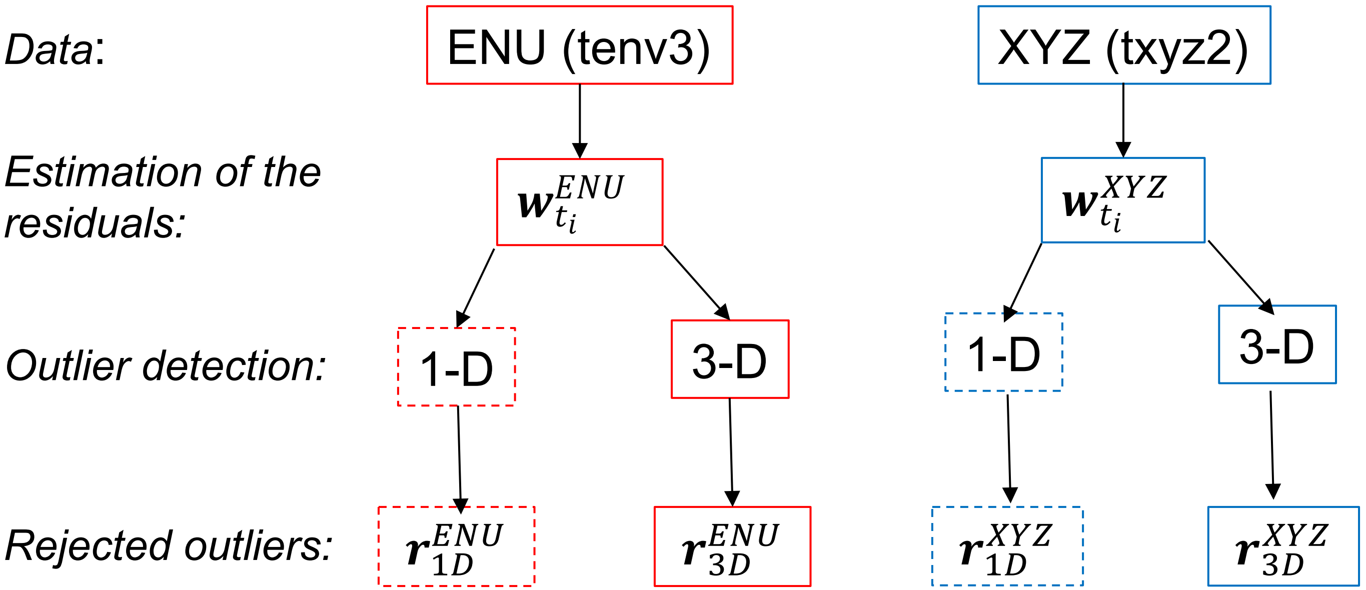
All the time series were analysed following exactly the same workflow, with the exception of the outlier rejection phase:

This has been done once using the classical 1-D approach and once using the proposed 3-D approach…doing this for both the ENU and the XYZ time series…

* Blewitt, G., Hammond, W. C., & Kreemer, C. (2018). Harnessing the GPS data explosion for interdisciplinary science. *Eos*, *99*, 1-2.

Test on real GNSS data

The following workflow was performed for each of the 12 considered GNSS sites:



Note that in the Outlier detection phase the same confidence level (99.73%) has been considered in both the 1-D and 3-D cases

$$[\boldsymbol{x}^k]^T \, {\Sigma_{ww}^k}^{-1} \, [\boldsymbol{x}^k] = c = 3{,}76248$$

(3.76248 is the value of a χ^2 having 3 degrees of freedom for the 99,73% confidence level)

Test on real GNSS data

- Columns 2 and 4: rejected solutions which are present both in $\boldsymbol{r}_{1D-3\text{D}}^{ENU}$ and $\boldsymbol{r}_{1D-3\text{D}}^{XYZ}$
- Columns 3 and 5: number of rejected solutions present in $\boldsymbol{r}_{1D-3D}^{ENU}$ but not in $\boldsymbol{r}_{1D-3D}^{XYZ}$ and vice versa

	n. rejected points 1-D		n. rejected points 3-D	
SITE	Different elements $\boldsymbol{r}_{1D}^{ENU}, \boldsymbol{r}_{1D}^{XYZ}$	Different elements $\boldsymbol{r}_{1D}^{ENU}, \boldsymbol{r}_{1D}^{XYZ}$	Common elements $\boldsymbol{r}_{1D}^{ENU}, \boldsymbol{r}_{1D}^{XYZ}$	Different elements $\boldsymbol{r}_{1D}^{ENU}, \boldsymbol{r}_{1D}^{XYZ}$
ALCI	16	31	19	0
GAR2	31	11	17	0
GV10	18	3	23	0
LAV1	40	19	30	0
MALI	46	43	55	0
MORN	31	24	25	0
NKLG	44	37	38	0
RCMN	29	15	22	0
SALU	27	17	21	0
SKCH	123	100	166	0
STJO	93	77	123	0
TUVA	44	19	37	0

Conclusions

We propose a new approach to outlier detection in time series, which is particularly suitable for GNSS data.

With respect to the classical approaches based on the analysis of each coordinate component separately, the 3-D approach has the following advantages:

- It is coherent with the probability distribution of a three-dimensional stochastic variable.
- It is invariant with respect to the reference system used to represent the coordinates.