The Effect of Colored Noise on Automatic Offset Detection in GNSS Time Series

M. S. Bos [1]

R. M. S. Fernandes [1]

M. Karegar [2]

[1] SEGAL (UBI/IDL), Covilhã, Portugal

[2] School of Geosciences, University of South Florida, USA

Part 1:

The limitations of automatic offset detection

Problem statement

- Nowadays there are thousands of GNSS stations and most of the derived coordinate time series contain offsets.
- Gazeaux et al. (2013) have shown that undetected offsets are now the largest contributors to the estimated trend error (±0.2 mm/yr manual detection, ±0.4 mm/yr automated detection).
- We need an **improved** automatic offset detection algorithm.

Still a lot of unknown offsets in the time series

GAZEAUX ET AL.: DETECTING OFFSETS IN GPS TIME SERIE

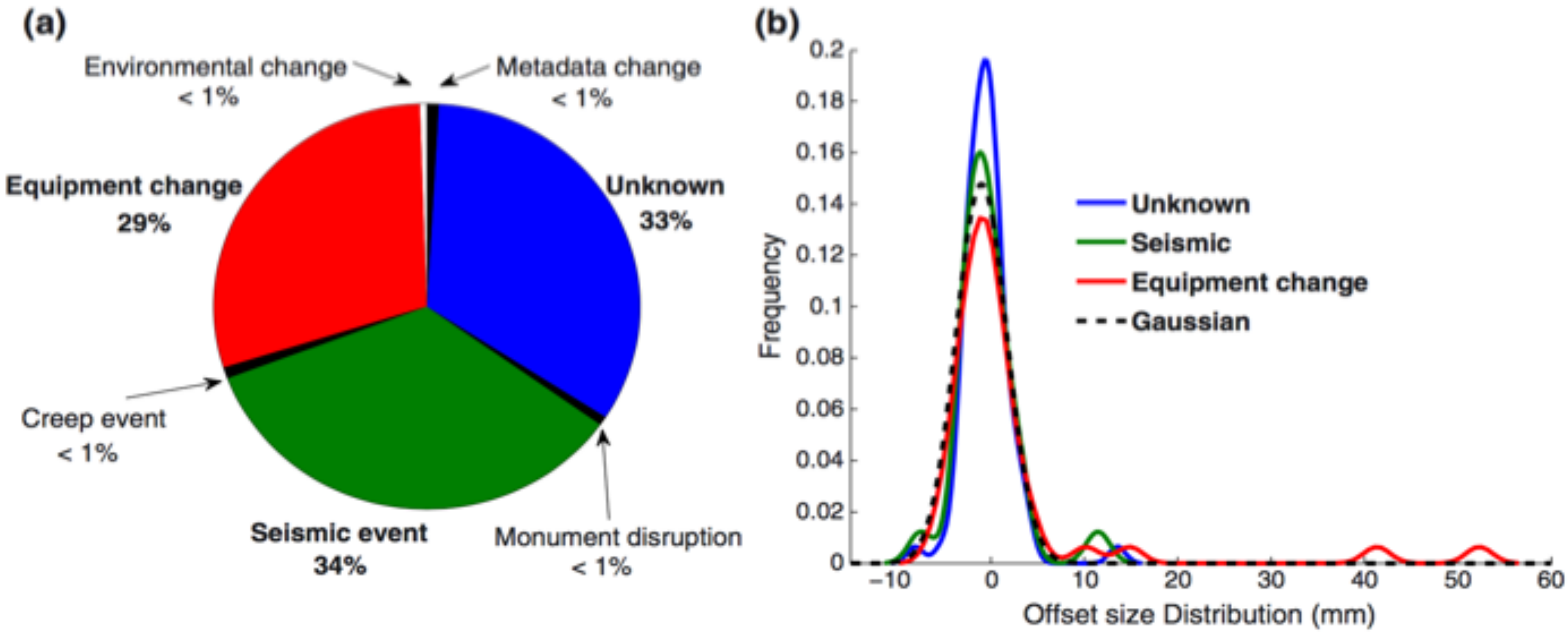


Figure 1. SOPAC offset (a) description and (b) magnitude distribution since 1995 over 340 sites (560 offsets).

Gazeaux et al. (2013)

How can we detect an offset?

- Simply test for each point of the time series if adding an offset improves the fit with the observations.
- To quantify the improvement we use the log-likelihood function L:

$$L = -\frac{1}{2}\left[N\log(2\pi) + \log\det(\mathbf{C}) + \mathbf{r}^T\mathbf{C}^{-1}\mathbf{r}\right]$$

constant

C=covariance matrix

r=residuals

- The higher the value of L (less negative), the more likely the offset has occurred in reality

- **Our strategy is to evaluate which location gives the most likely epoch for an offset.**

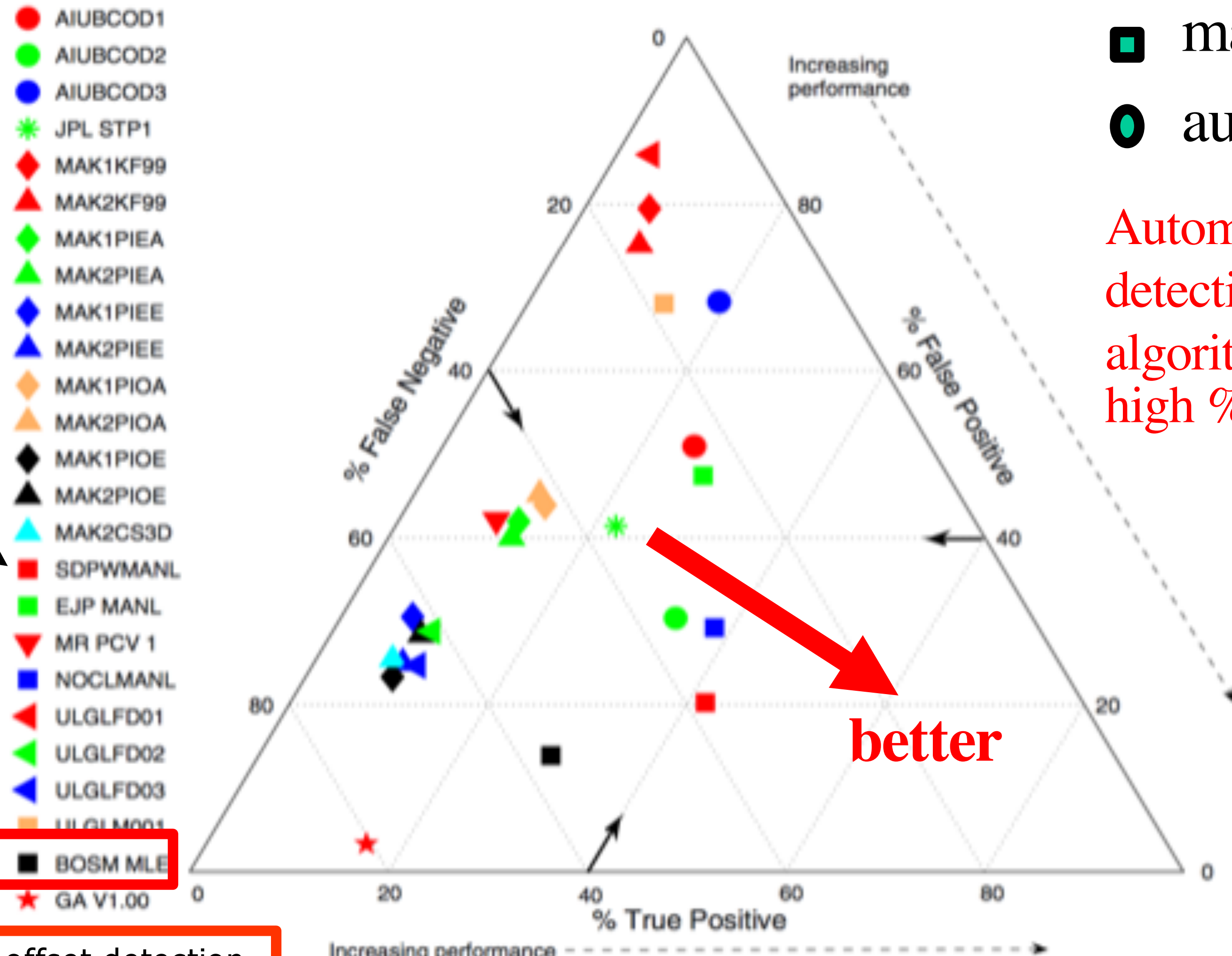
What are the possible outcomes of the offset detection algorithm?

- **True Positive (TP): Algorithm has detected a real offset.**
- **False Positive (FP): Algorithm has detected a false offset.**
- **False Negative (FN): Algorithm has NOT detected a real offset.**

- **“True Negative (TN)” is missing but can be computed from the other three**

Detection of Offsets in GPS Experiment (DOGEx)

GAZEAUX ET AL.: DETECTING OFFSETS IN GPS TIME SERIE



manual

automatic

Automatic offset detection algorithm have high % FP

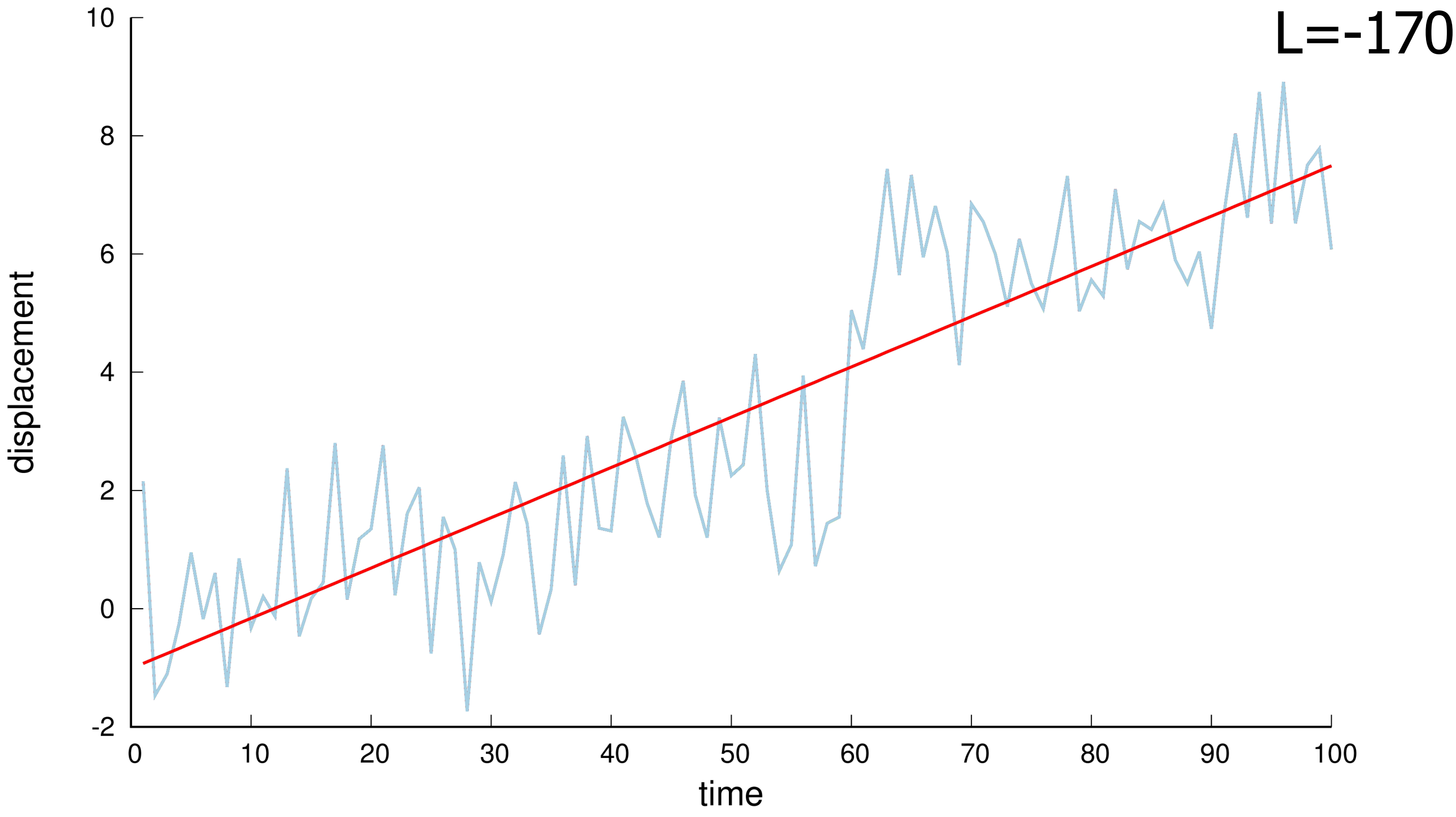
Simon Williams (visual)

Hector + visual offset detection

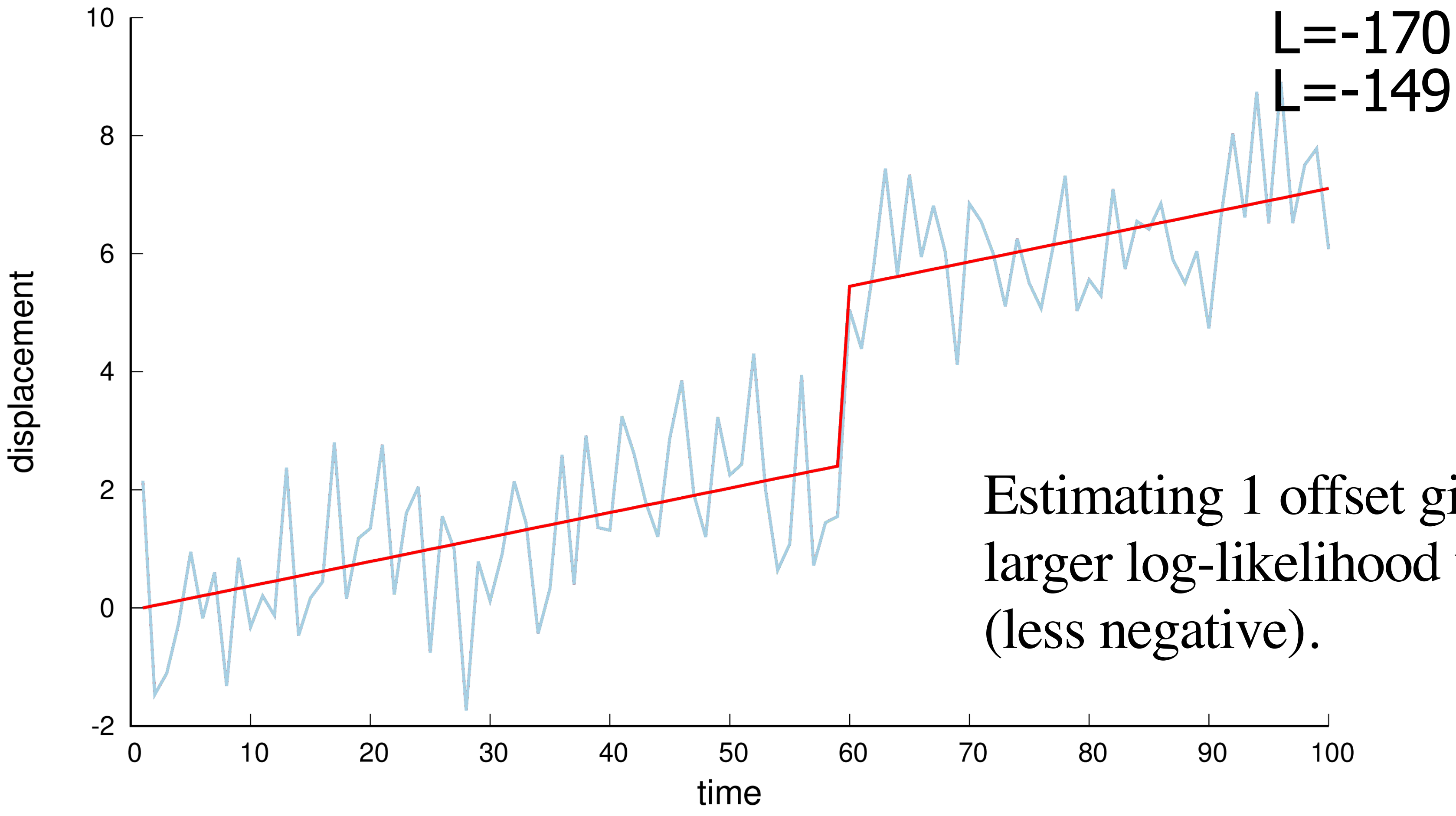
Figure 4. The ratio between the three performance measures of the solutions (TP, FP, and FN). Dotted arrows show the direction of a perfect solution. Manual solutions are represented by square markers.

Gazeaux et al. (2013)

Automatic Detection using Synthetic Data



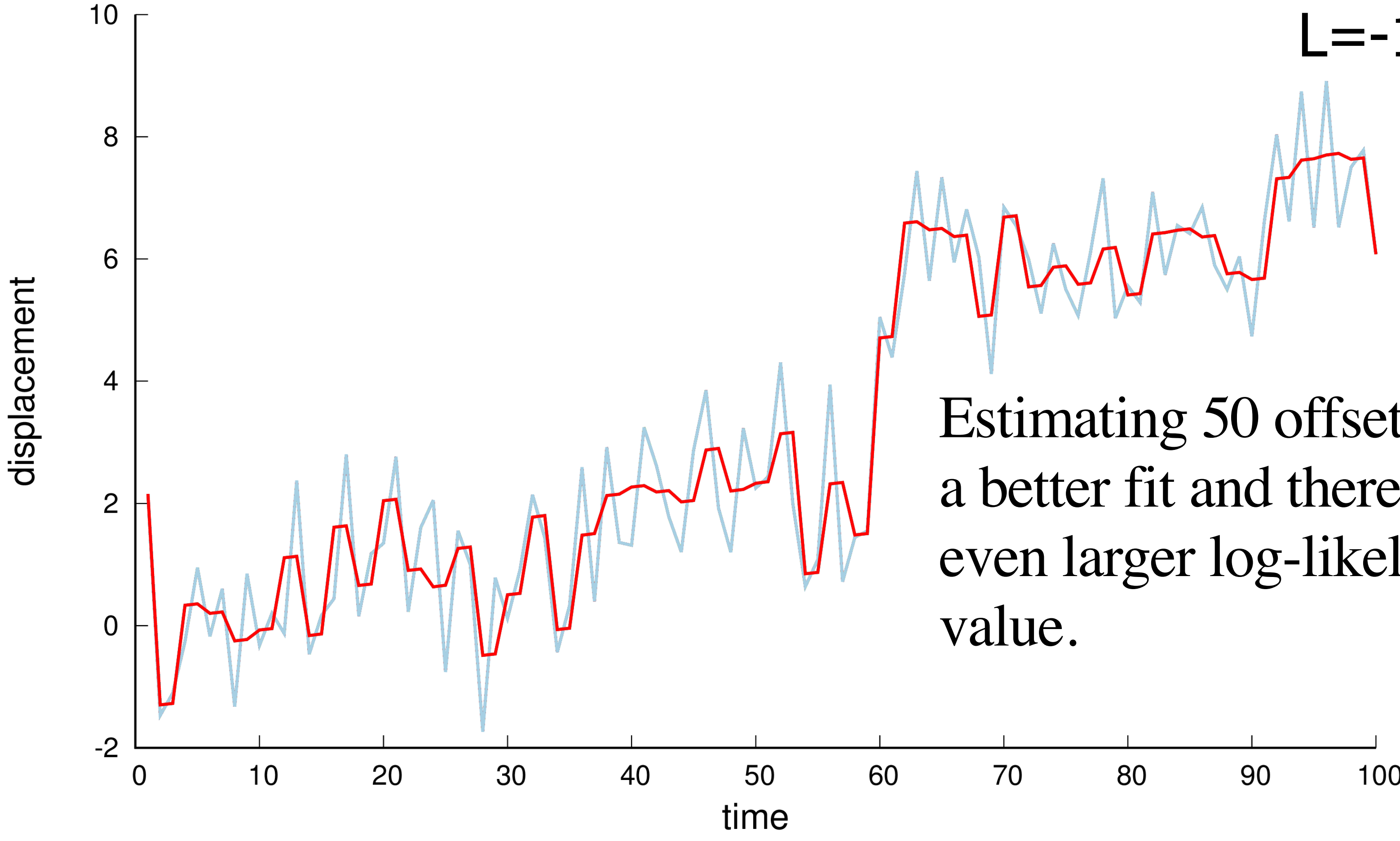
Automatic Detection using Synthetic Data



Estimating 1 offset gives a larger log-likelihood value (less negative).

Automatic Detection using Synthetic Data

L=-170
L=-149
L=-115



Estimating 50 offsets gives a better fit and therefore an even larger log-likelihood value.

When do we stop adding more offsets?

Statement: Offsets does not occur everyday

- We have assumed we have no prior knowledge about the likelihood of having any number of offsets.
- However, we know from experience that the probability of an offset occurs is low (only in exceptional situations) and this needs to be taken into account in the log-likelihood function.
- So, we need to have a criterium to stop the search!

Bayesian Information Criterion

Popular stopping criteria is the Bayesian Information Criterion (BIC).

Assume *L* is the likelihood, *k* the number of parameters in the noise model and *n* the number of observations:

$$AIC = 2k - 2L$$

$$BIC = \ln(n)k - 2L$$

The higher the likelihood *L (better the noise model)*, the lower the AIC/BIC value.

More offsets, higher penalty. **This avoids to have too many offsets.**

Limitations of BIC

- The Bayesian Information Criterion has been derived by taking the limit to infinite long time series.
- BIC is probably widely used because it is easy implemented and it gives a first good rough idea of the amount of offsets.
- Good alternative empirical penalty functions have been developed.
- However, empirical penalty functions depend on the type of data being analyzed and the Bayesian interpretation can be lost.

BICc – Corrected BIC

- For finite time series, which is the case of GNSS daily solutions, we are using a corrected BIC (Bos et al., in preparation):

$$\mathrm{BIC}_c = -2\ln L + m\ln(\frac{n}{2\pi}) + \ln\det(\mathcal{I}) - 2\ln(f(\hat{\theta}_k)) - 2\ln f(\mathbf{s}_k)$$

Depends on noise

A priori info about size of the offset

A priori info about number of offsets

- This is BIC with more a priori information = **BICc**

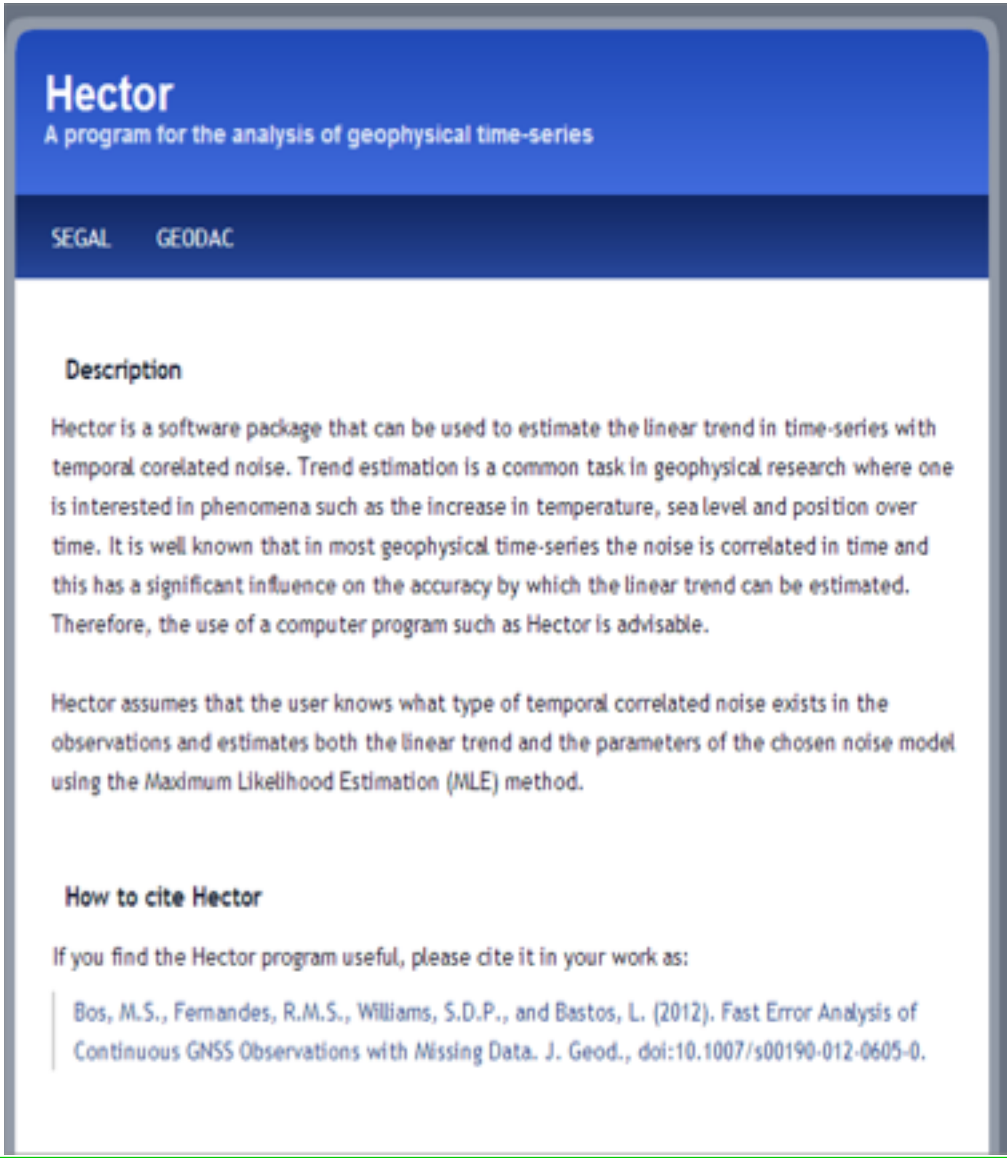
Part 2:

Does this theoretical hogwash actually work?

HECTOR – Time-Series Analysis (http://segal.ubi.pt/hector/)

Simultaneously Computation of:

- **Secular Trend**
- **Seasonal Signals**
- **Automatic Offset Detection**
- **Exponential / Logarithmic Post-relaxation**
- **Power-law errors**
- **Spectrum Index**



Hector

A program for the analysis of geophysical time-series

SEGAL GEODAC

Description

Hector is a software package that can be used to estimate the linear trend in time-series with temporal corelated noise. Trend estimation is a common task in geophysical research where one is interested in phenomena such as the increase in temperature, sea level and position over time. It is well known that in most geophysical time-series the noise is correlated in time and this has a significant influence on the accuracy by which the linear trend can be estimated. Therefore, the use of a computer program such as Hector is advisable.

Hector assumes that the user knows what type of temporal correlated noise exists in the observations and estimates both the linear trend and the parameters of the chosen noise model using the Maximum Likelihood Estimation (MLE) method.

How to cite Hector

If you find the Hector program useful, please cite it in your work as:

> Bos, M.S., Fernandes, R.M.S., Williams, S.D.P., and Bastos, L. (2012). Fast Error Analysis of Continuous GNSS Observations with Missing Data. J. Geod., doi:10.1007/s00190-012-0605-0.

Detection of Offsets in GPS Experiment (DOGEx)

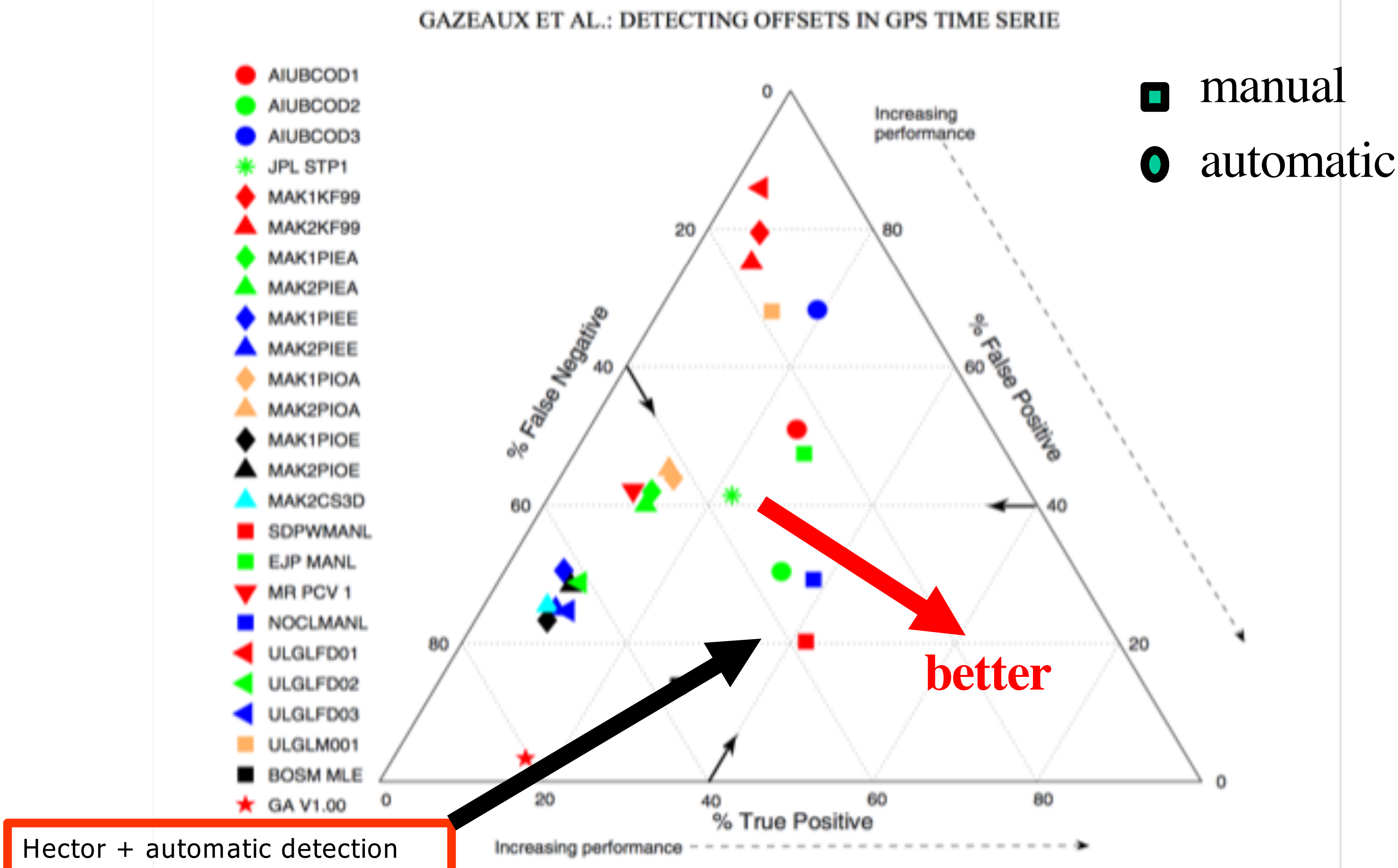


Figure 4. The ratio between the three performance measures of the solutions (TP, FP, and FN). Dotted arrows show the direction of a perfect solution. Manual solutions are represented by square markers.

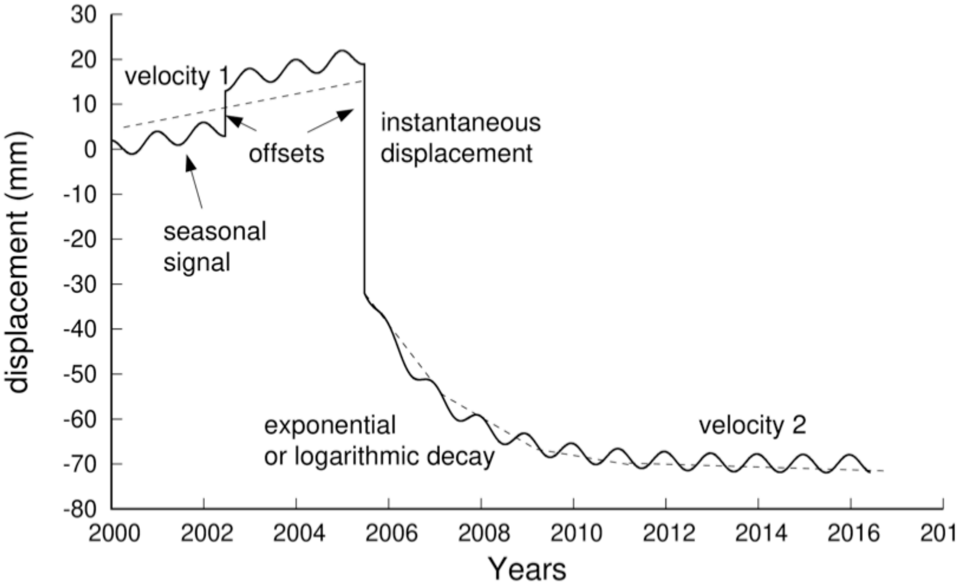
Gazeaux et al. (2013)

Part 3:

Is it Least Squares dead to estimate velocities?

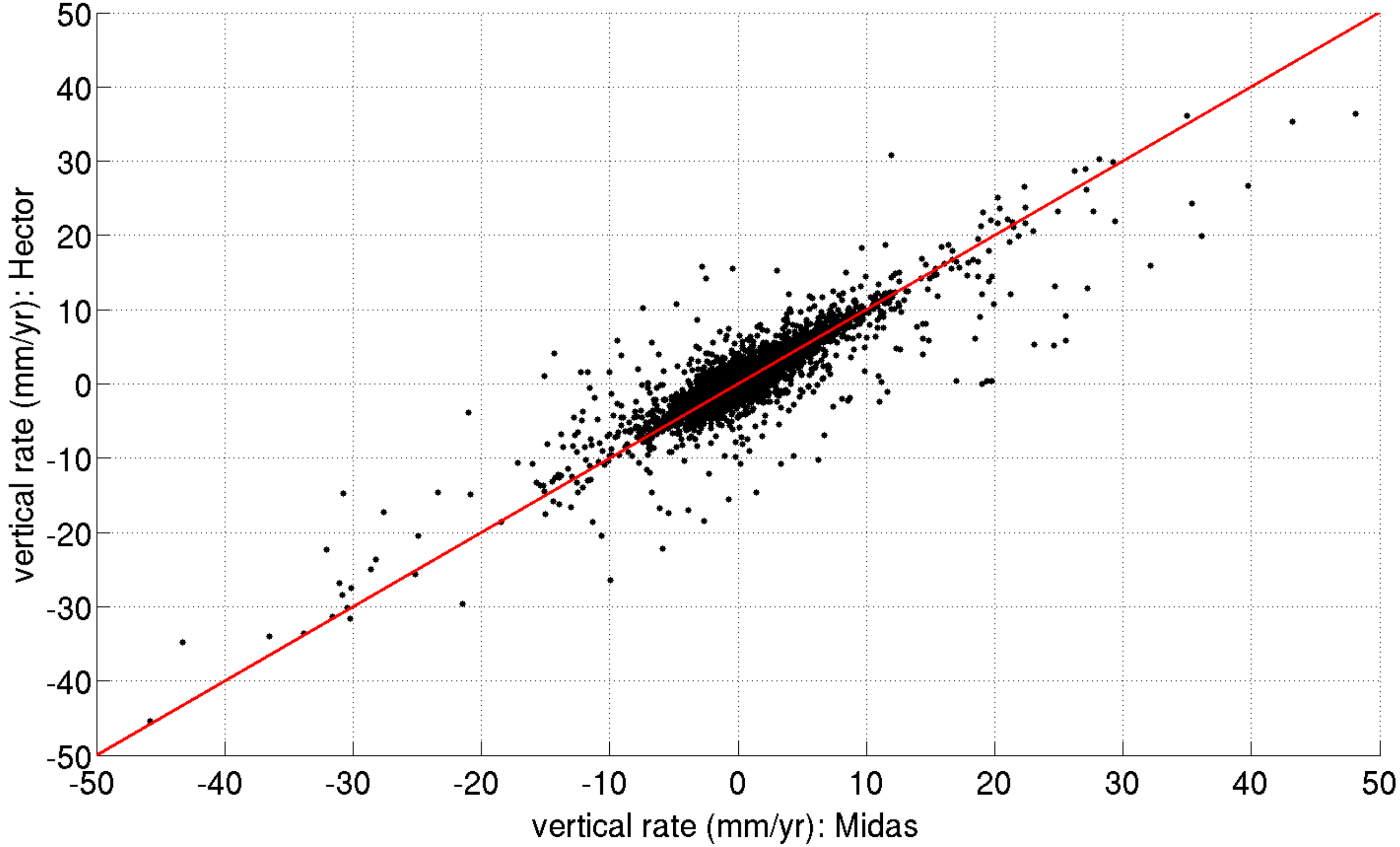
Do we really need to estimate offsets?

- MIDAS (Blewitt et al., 2016), based on the median of velocities computed using a temporal sliding window, is a good solution when you only want to estimate one (and only one) velocity from your time-series.
- However, other signals exist on the time-series that can be of interest.

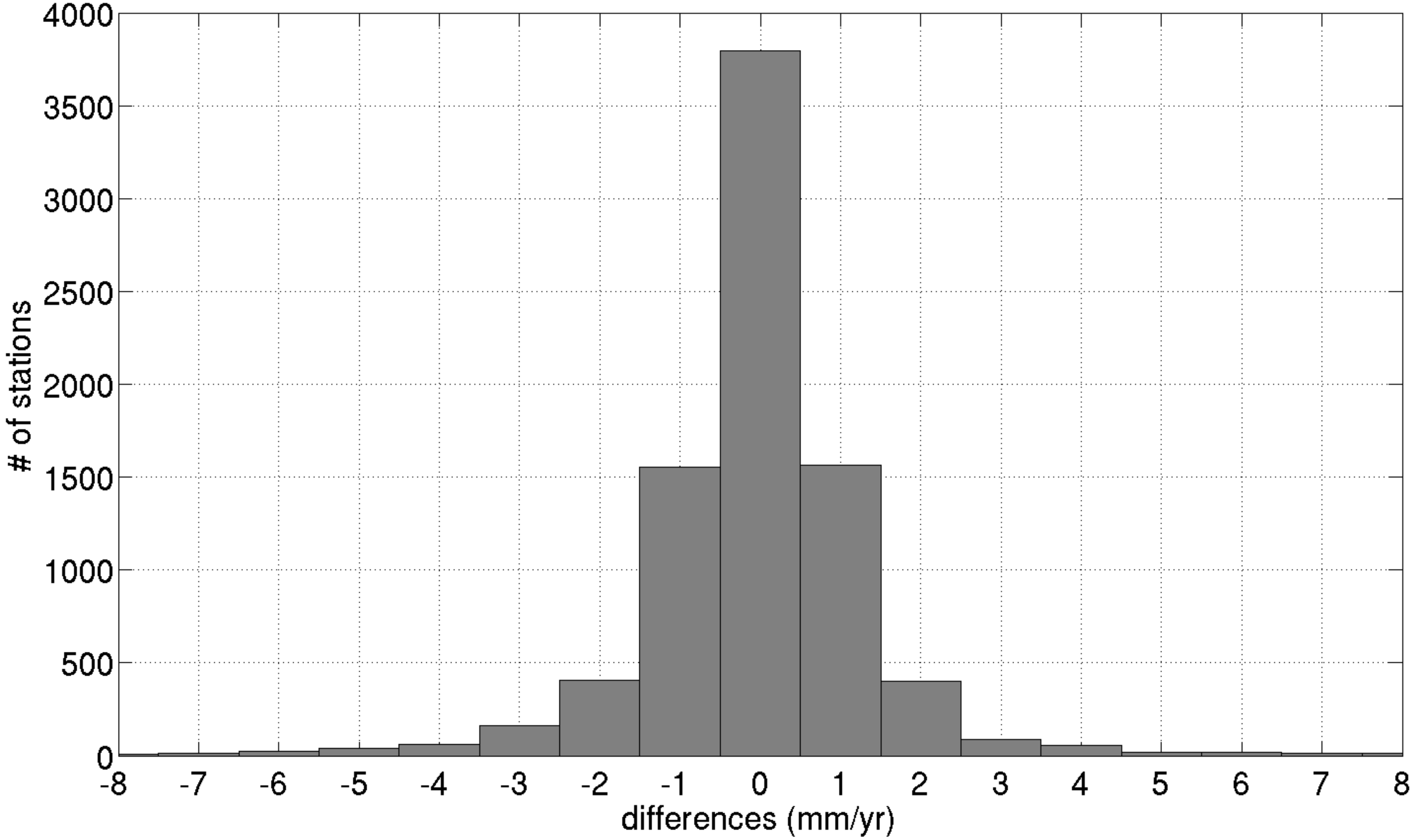


MIDAS vs. Hector + Offset Detection

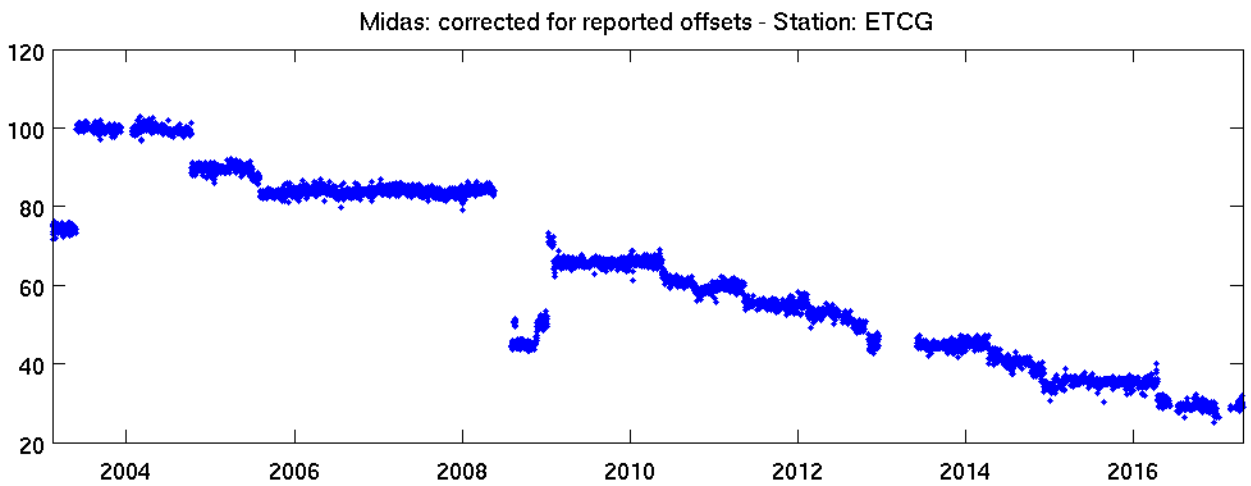
- We used all time series of NGL (~3000 stations = ~9000 time-series) and analyzed them with MIDAS and Hector.

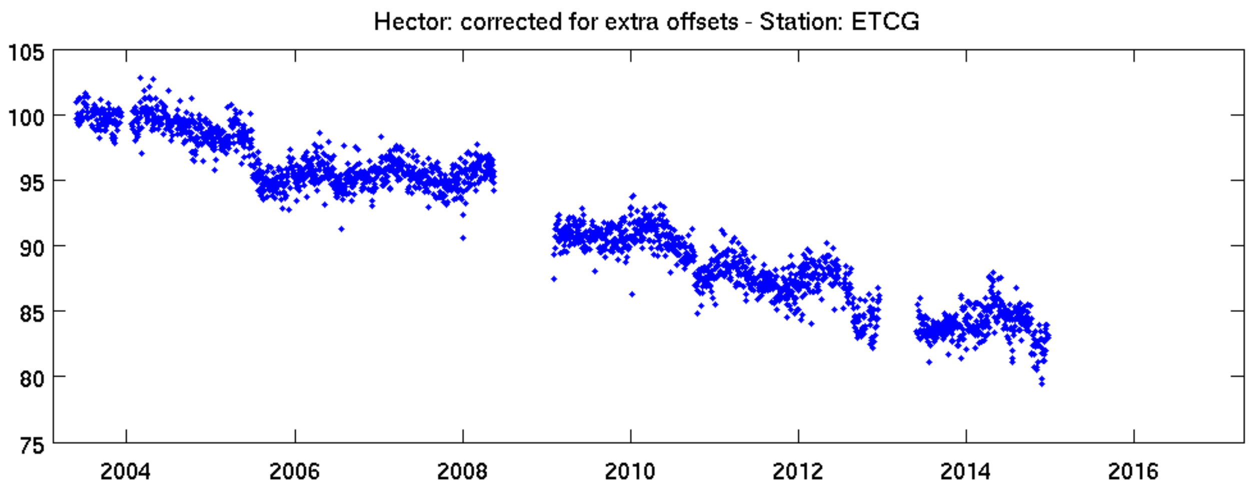


MIDAS vs Hector
9000 NGL time series

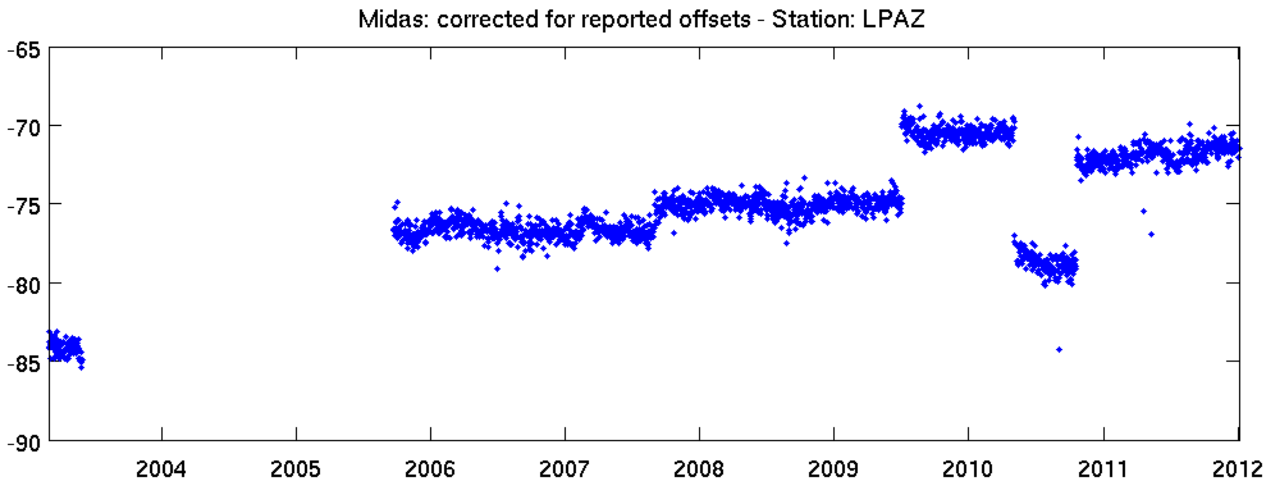


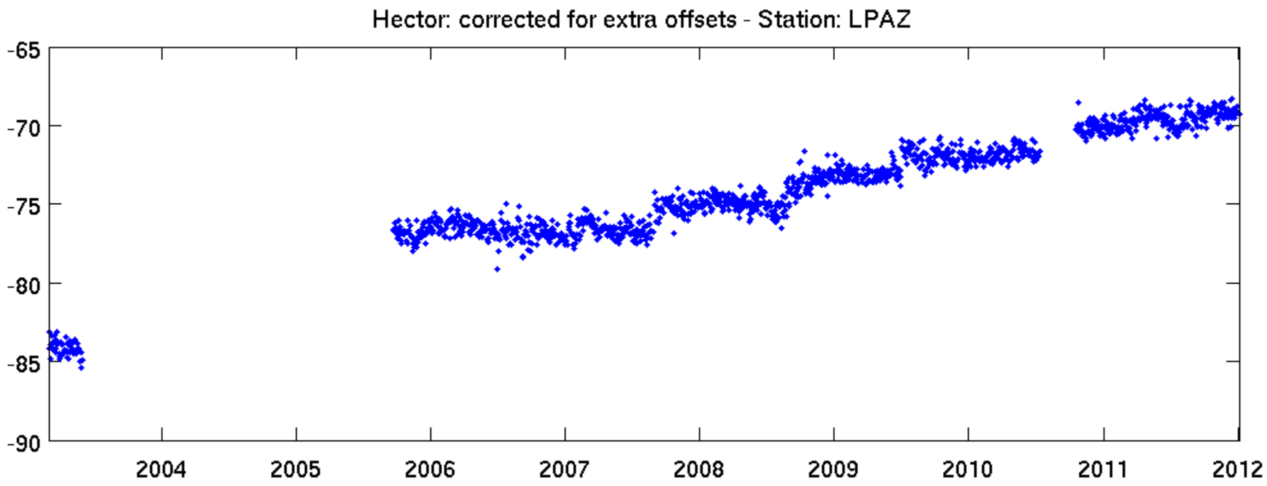
velocity difference > 10 mm/yr





velocity difference > 10 mm/yr



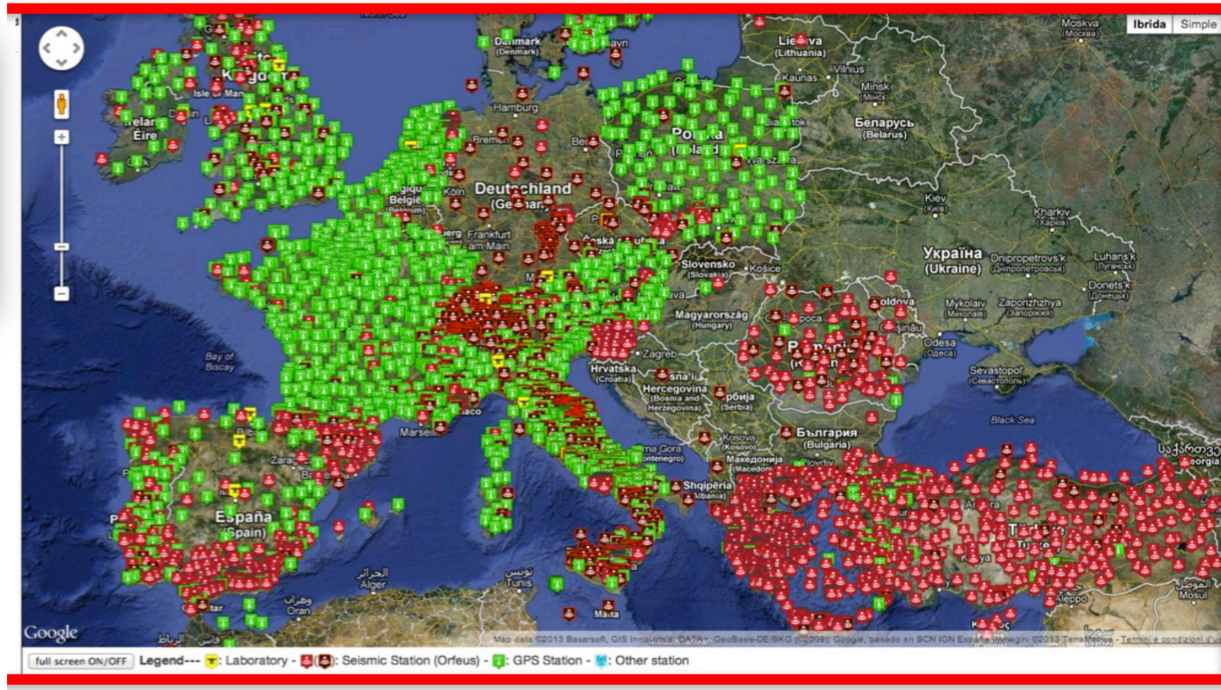


Summary

- Automatic offset detection is becoming a practical necessity.
- To obtain this result, we had to improve the Bayesian Information Criterion to incorporate information about the probability of the size and spacing between offsets.
- Using DOGEx as the benchmark: our algorithm is the best automated one (comparison with the others in 2013)!
- HECTOR (BICc) is slower but we have demonstrated it can handle thousands of stations (the comparison using the 9000 NGL time-series took 2 weeks time).
- Agreement of MIDAS with Hector is good. For problematic stations Hector seems to have the edge but it is difficult to quantify.

EPOS: European Plate Observing System

- etc....



HECTOR (trend estimation and offset detection) will be used in EPOS-GNSS Operational Services

Thanks / Obrigado / 非常感謝

Nice/Productive Questions/Remarks/Suggestions:

msbos@segal.ubi.pt
rui@segal.ubi.pt
makan.karegar@gmail.com

Nasty Ones:

noone@nowhere.noplace