EUREF 2017 Symposium

WROCŁAW May 17-19, 2017

Accuracy estimation of regional TEC maps by cross-validation and maximum likelihood

W. Jarmołowski, A. Krypiak-Gregorczyk, P. Wielgosz

University of Warmia and Mazury in Olsztyn, Poland

UWM
UNIWERSYTET
WARMIŃSKO-MAZURSKI
W OLSZTYNIE

When do we require **precise parameters** in spatial or temporal modeling

- **High accuracy** requirements
- **Sparse data** or data including gaps
- **Combination of data** that have different accuracy, reference, include outliers
- **Significant noise in relation to the signal** (crustal movements, plate velocities)
- **Significant noise** in general

Motivation and objectives of precise parameters estimation for TEC maps

- **Noisy data** (separation of noise and signal is difficult but is crucial)
- **A very limited information about noise**
  (even no info about measurement error)

- **Best accuracy maps for positioning** purposes
  (therefore least-squares method)
- **High accuracy grids for further processing**
  (spectral methods, spherical harmonics, etc.)

Motivation of **Fisher Scoring choice** instead of Empirical Covariance Function and Cross-validation

- **Empirical covariance function does not give info about noise** variance (as it is calculated from signal+noise)
- **ECF usually needs manual steps**
- **Cross-validation needs selection of some range of parameters** and is time consuming
- **Cross-validation (e.g. LOO) can be difficult for noisy sets,** as there is no good data for comparison at points
- **All parameters vary in time (near real time estimation)** (empirical covariance functions/variograms would be extremely inefficient)

Observational data

- **L1&L2 carrier phase** data from:
- **50** GNSS stations of Polish **ASG-EUPOS** network.
- **>200** GNSS stations of **EPN (EUREF Permanent Network)**.
- dual-frequency carrier phase and pseudorange **GPS + GLONASS** data.
- sampling interval: **60 seconds**
- **elevation cut-off: 30°**

Data, Fundamental observation equations

$$L1_i^k = q_i^k + c(\Delta t_i - \Delta t^k) + \Delta T_i^k - \boldsymbol{\Delta I_i^k} - \lambda_1 N1_i^k + c(b_{L1}^k + b_{L1.i}) + \varepsilon$$

$$P1_i^k = q_i^k + c(\Delta t_i - \Delta t^k) + \Delta T_i^k + \boldsymbol{\Delta I_i^k} + c(b_1^k + b_{1.i}) + \varepsilon$$

where:

$L1_i^k$	– the carrier phase observations on L1 frequency.
$P1_i^k$	– the P-code observations on L1 frequency.
q_i^k	– the geometric distance between receiver i and satellite k.
c	– the speed of light.
$\Delta t_i.\ \Delta t^k$	– offsets of the receiver (i) and satellite (k) clocks.
ΔT_i^k	– delay of the signal due to the troposphere.
ΔI_i^k	**– delay of the signal due to the ionosphere.**
$b_1^k.\ b_{L1}^k$	– the satellite hardware delay.
$b_{1.i}.\ b_{L1.i}$	– the receiver hardware delay.
$N1_i^k$	– the initial carrier phase ambiguity.
λ_1	– the wavelength.
ε	– indicates a random error.

Geometry-Free Linear Combination (L4) of carrier phase data

$$L1_i^k = q_i^k + c(\Delta t_i - \Delta t^k) + \Delta T_i^k - \boldsymbol{\Delta I_i^k} - \lambda_1 N1_i^k + c(b_{L1}^k + b_{L1.i}) + \varepsilon$$

$$L2_i^k = q_i^k + c(\Delta t_i - \Delta t^k) + \Delta T_i^k - \xi\boldsymbol{\Delta I_i^k} - \lambda_2 N2_i^k + c(b_{L2}^k + b_{L2.i}) + \varepsilon$$

$$\xi = \frac{f_1^2}{f_2^2} \approx 1.647; \qquad \xi_4 = 1 - \xi.$$

$$\boldsymbol{L4_i^k} = L1_i^k - L2_i^k = -\,\xi_4 \Delta \mathrm{I}_\mathrm{i}^\mathrm{k} + \mathrm{B}_{\mathrm{i}.4}^\mathrm{k}.$$

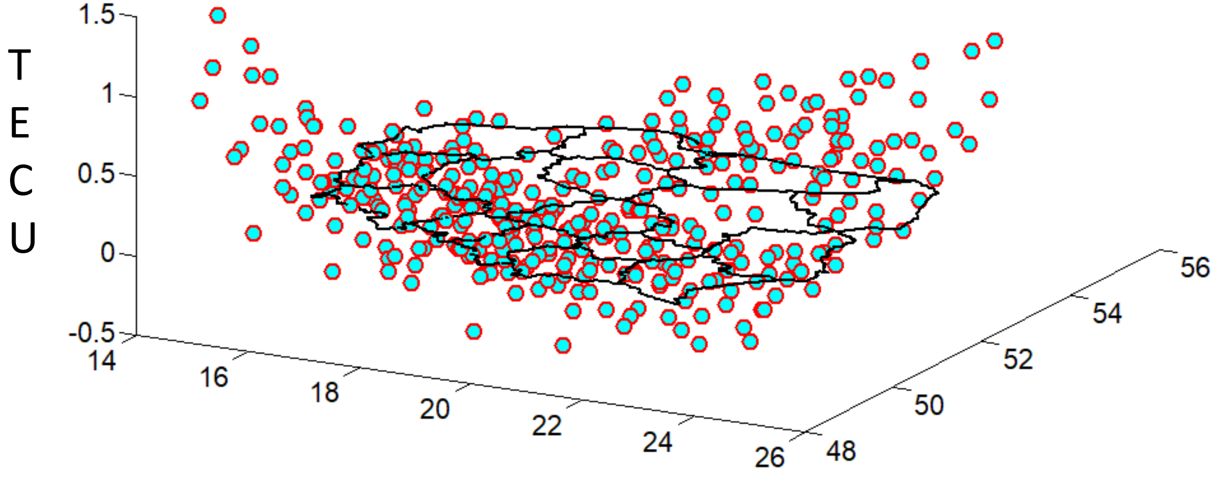
$$where: \quad B_{i.4}^k = \lambda_1 N_{i.1}^k - \lambda_2 N_{i.2}^k - (b_{L1}^k - b_{L2}^k) - (b_{L1.i} - b_{L2.i})$$

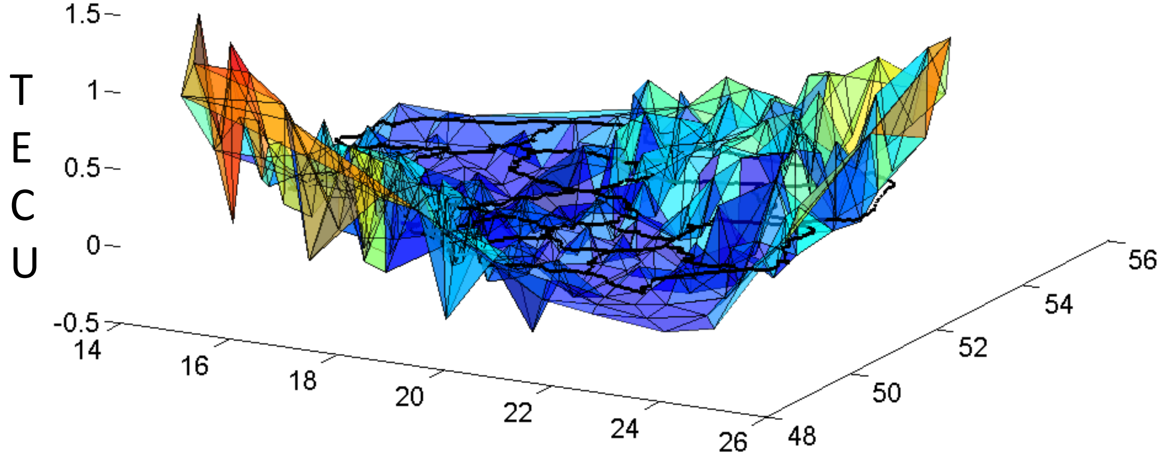
$$\Delta I_i^k = \frac{L4_i^k - B_{i.4}^k}{-\,\xi_4}$$

Carrier phase bias:
constant for continuous data arc

Data after polynomial trend romoval

- The signal power of TEC decreases quickly (spatially) at higher frequencies.
- To keep a measurable part of signal, first order polynomial as a trend is useful for very local investigations.
- The variance of remaining signal is larger than noise variance only a little





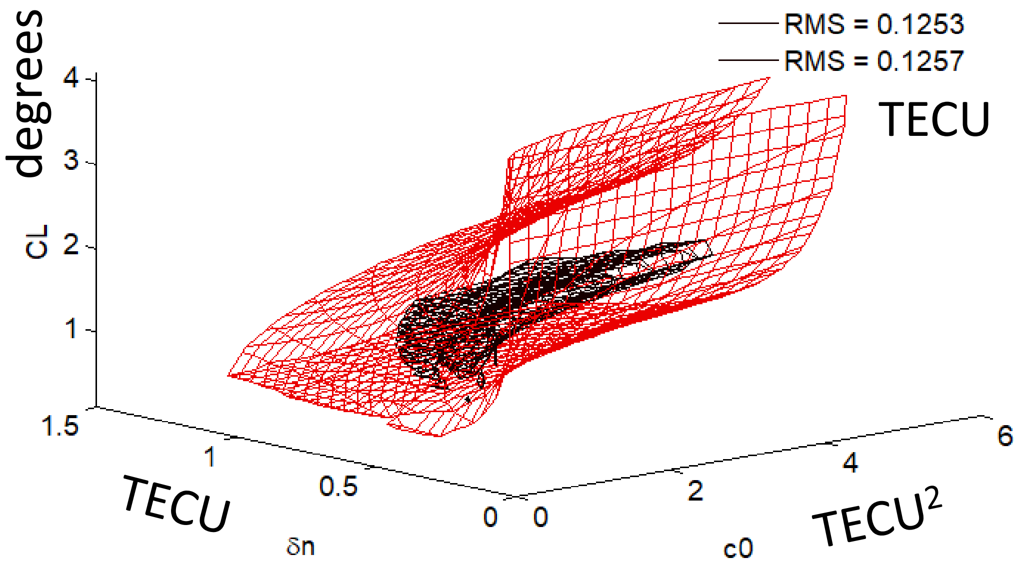
Covariance matrices (signal + noise)

$$\begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{bmatrix} + \begin{bmatrix} \delta n_{11} & 0 & \cdots & 0 \\ 0 & \delta n_{22} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \delta n_{nn} \end{bmatrix}$$

- Covariance matrices of signal and noise added.
- Therefore C_0 and δn are correlated, covariance function can be rescaled, but their ratio should be kept

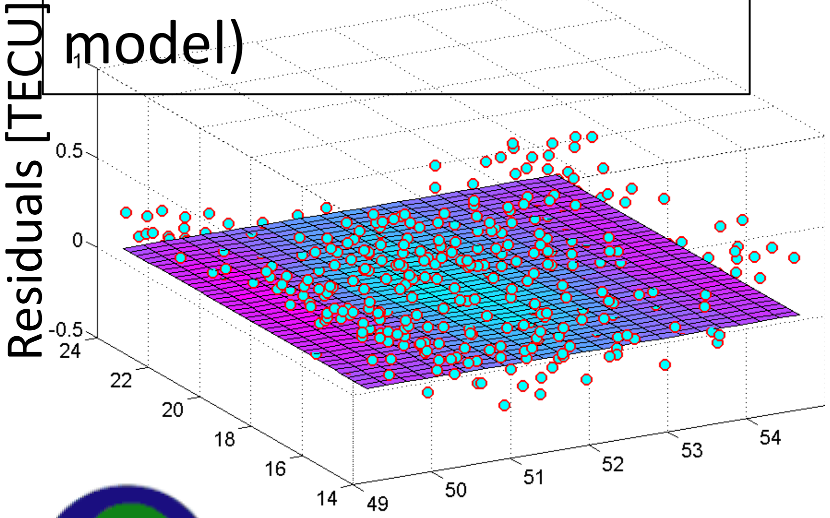
$$GM3(C_0, CL, s) = C_0\left(1 + \frac{s}{CL} + \frac{s^2}{3 \cdot CL^2}\right) \cdot \exp\left(\frac{-s}{CL}\right)$$

Leave-one-out validation of 3 parameters in 3D. Minimum is elongated along C_0 (signal variance)

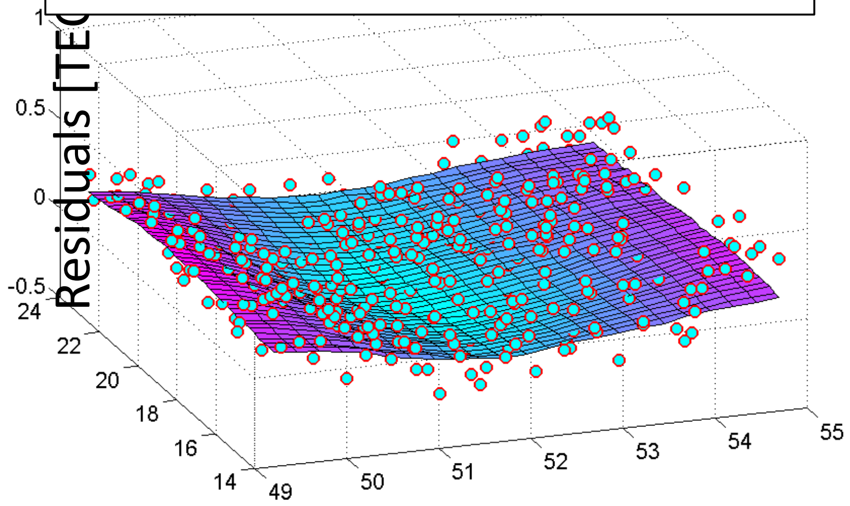


Noise issue. LSC is a spatial technique, but we manipulate spherical harmonic degrees (frequencies) in some sense

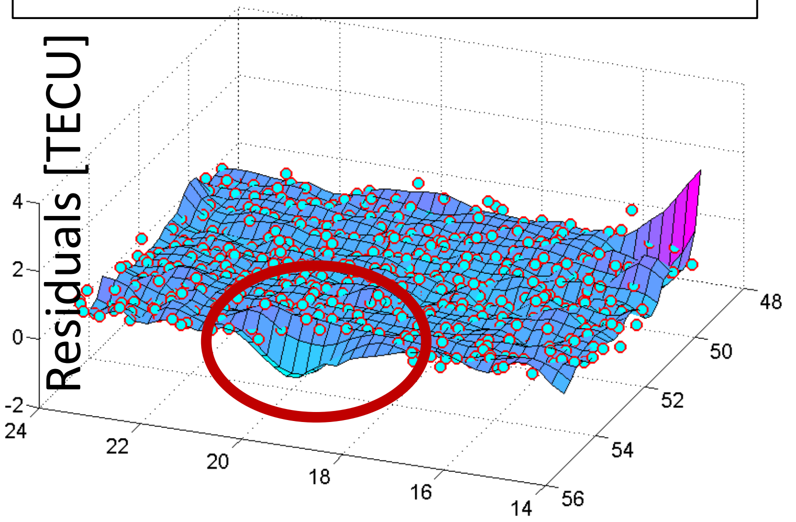
In case of overestimated a priori noise (here 5 TEC to emphesize the effect) LSC looses upper frequencies (smooths the model)



In case of correct a priori noise (0.2 TEC) LSC interpolates the signal and ignore the noise to the level that we permit



In case of underestimated a priori noise (here 0.001 TEC to emphesize the effect) LSC interprets noise as a signal



Fisher Scoring with Levenberg-Marquardt optimization

- In Fisher Scoring we use so-called Fisher Information Matrix of the form:
- Which has a size dependent on the numer of covariance parameters estimated and its elements are computed as:
- Where R is based on projection matrix and Ci are covariance parameters derivatives:
- These parameters in the current case are:
- And the optimized Fisher Scoring reads:

$$\boldsymbol{S}(\boldsymbol{\theta}) = \begin{bmatrix} S_{1,1} & S_{1,2} & \cdots & S_{1,jmax} \\ S_{i,1} & S_{2,2} & \cdots & S_{2,jmax} \\ \vdots & \vdots & \ddots & S_{i,jmax} \\ S_{imax,1} & S_{imax,2} & S_{imax,j} & S_{imax,jmax} \end{bmatrix}$$

$$S_{ij} = tr\left(\boldsymbol{R}\boldsymbol{C}_i\boldsymbol{R}\boldsymbol{C}_j\right) \qquad i,j = \theta_i$$

$$\boldsymbol{C}_i = \frac{\partial \boldsymbol{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i}$$

$$\boldsymbol{\theta} = [\delta n;\ CL]$$

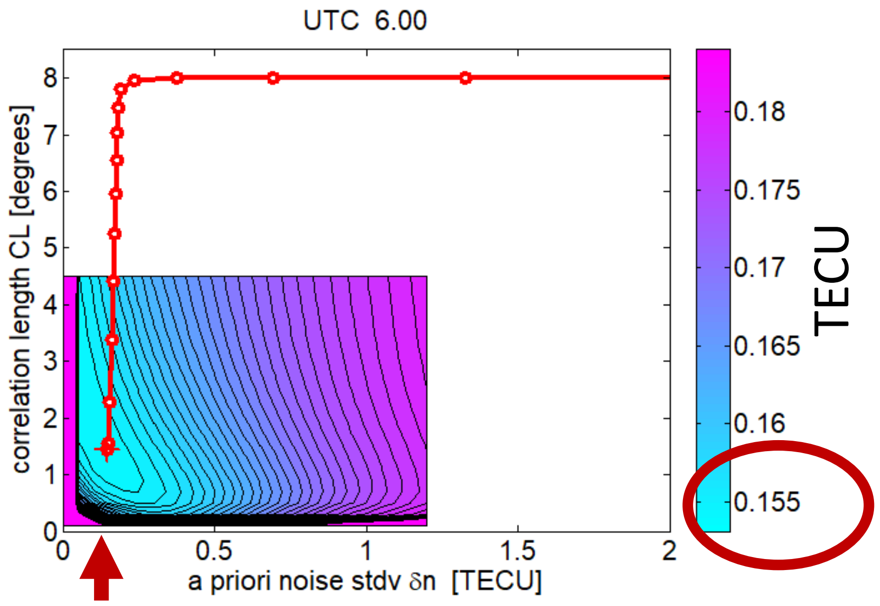
$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - (\boldsymbol{S}(\boldsymbol{\theta}_k) + \mu \cdot \boldsymbol{I}_{imax})^{-1} \cdot \boldsymbol{d}_k(\boldsymbol{\theta}_k), \qquad k \in \{1,2..z\}$$

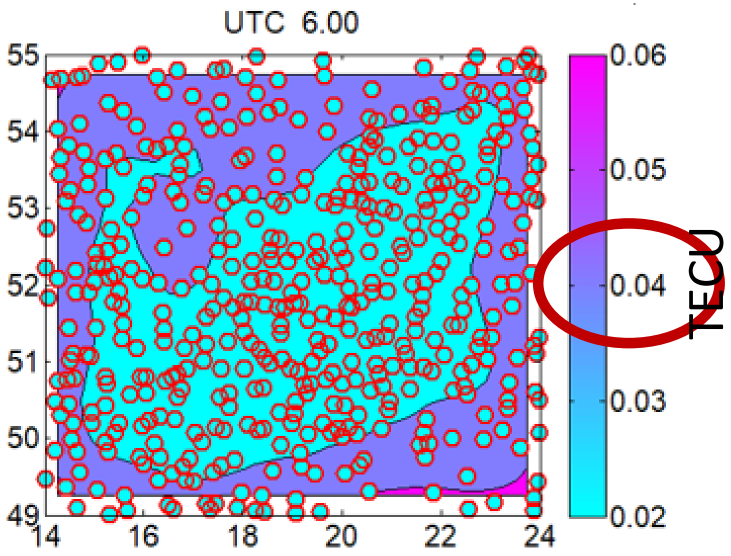
Jarmołowski W., 2017, Fast estimation of covariance parameters in least squares collocation by Fisher scoring with Levenberg-Marquardt optimization. Surveys in Geophysics: DOI : 10.1007/s10712-017-9412-8

Comparison of 3 accuracy markers

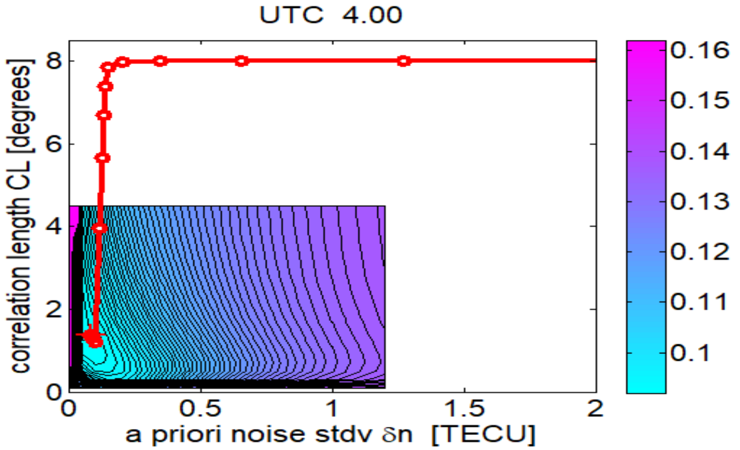
There are many factors related to data distribution causing differences between these factors, however it is clear that 3 indicators of accuracy must be comparable (must be always compared and cannot diverge significantly).

- A priori noise **δn**
- **RMS** of LOO validation
- **a posteriori error** calculated on the basis of a priori noise (depends at least on δn and data density in place)

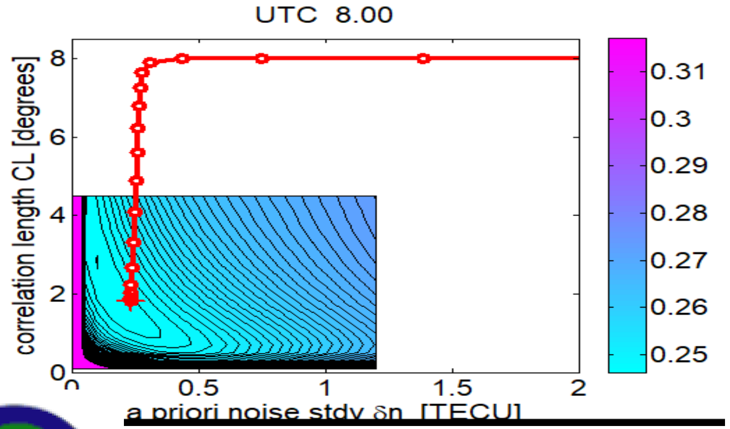


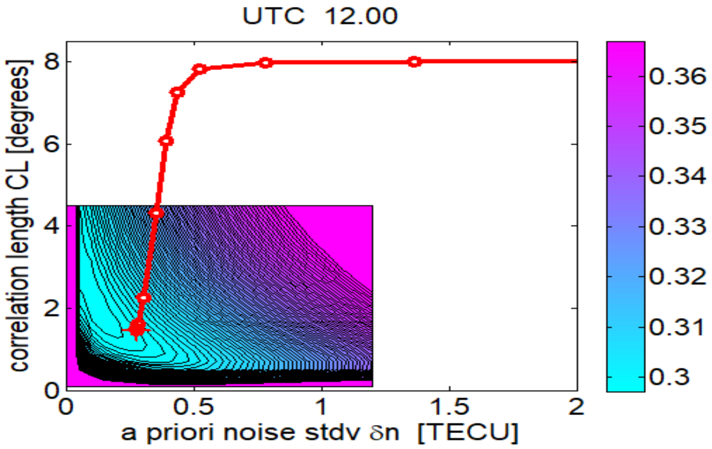


Fisher scoring of two parameters (CL and δn) day 78 (2015) (leave-one-out (LOO) validation in the background

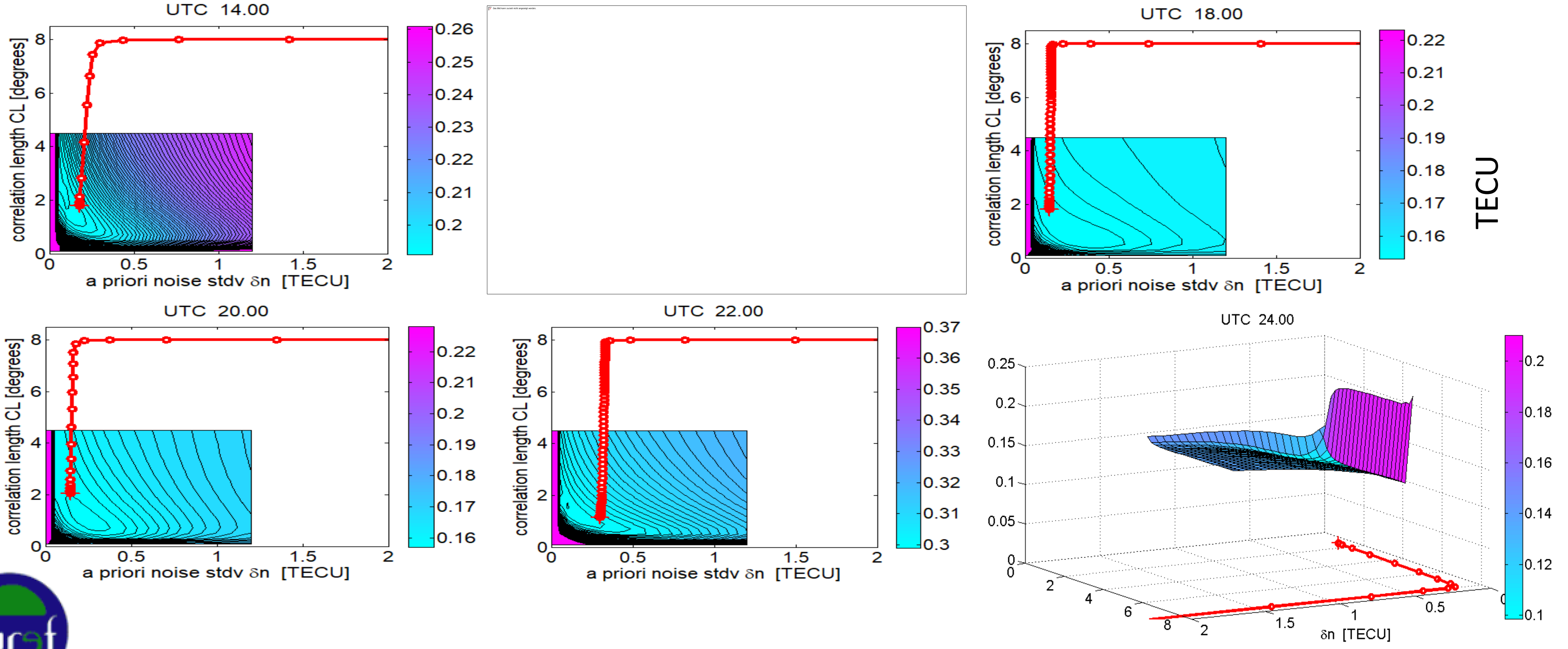


TECU

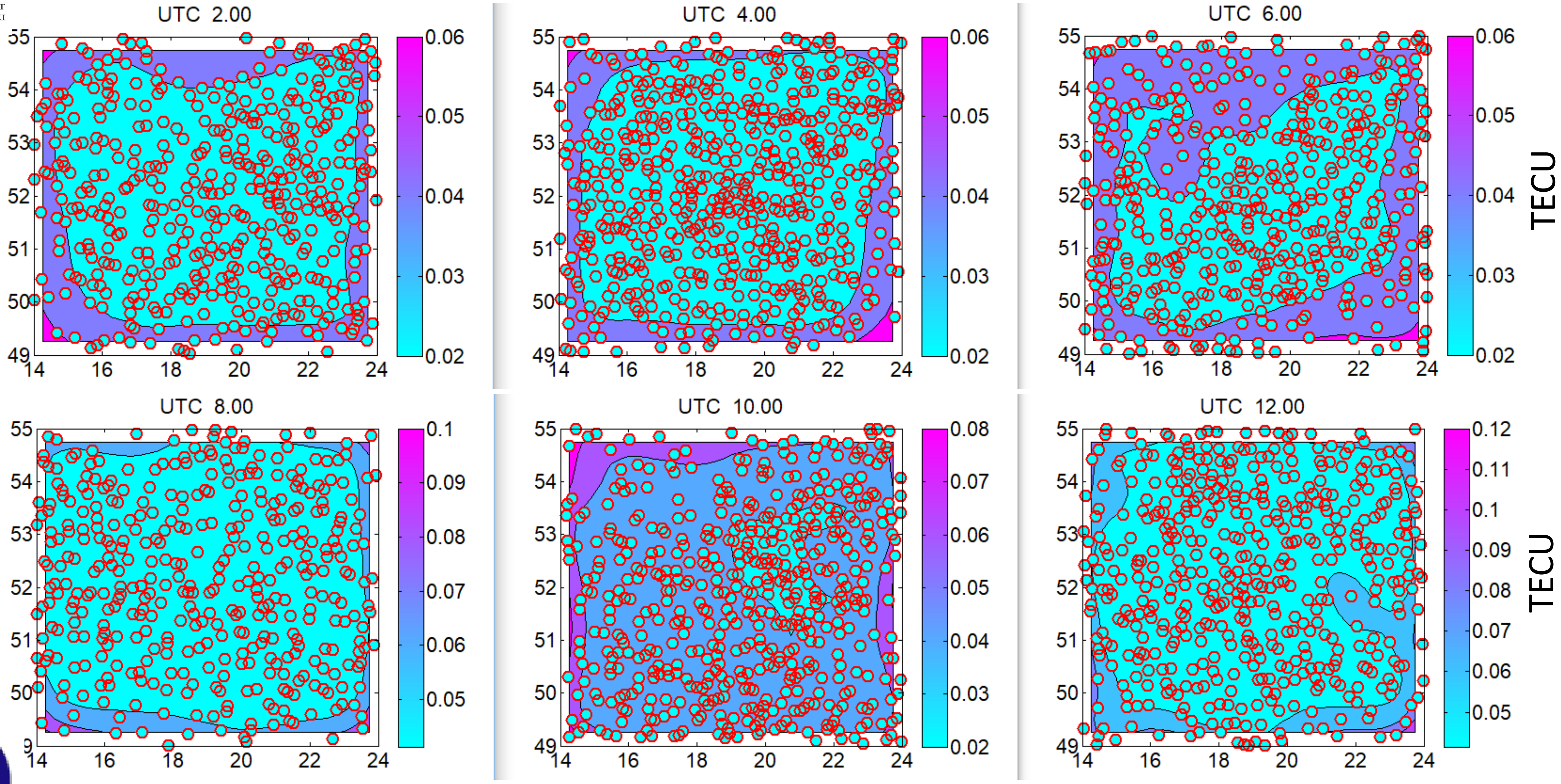




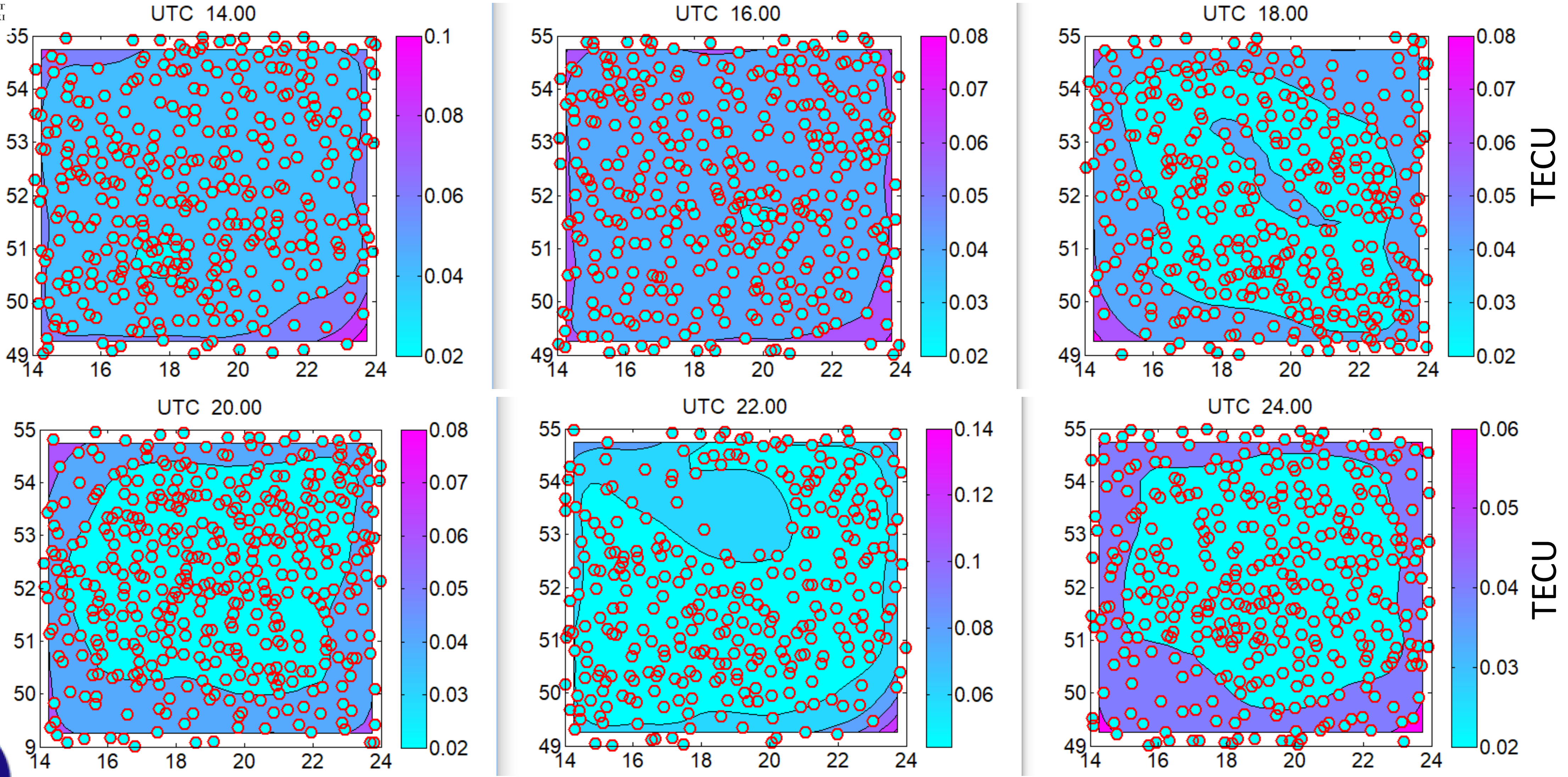
Fisher scoring of two parameters (CL and δn) day 78 (2015) (leave-one-out (LOO) validation in the background



A posteriori error (accuracy of TEC) day 78 (2015), **δn = 0.2**



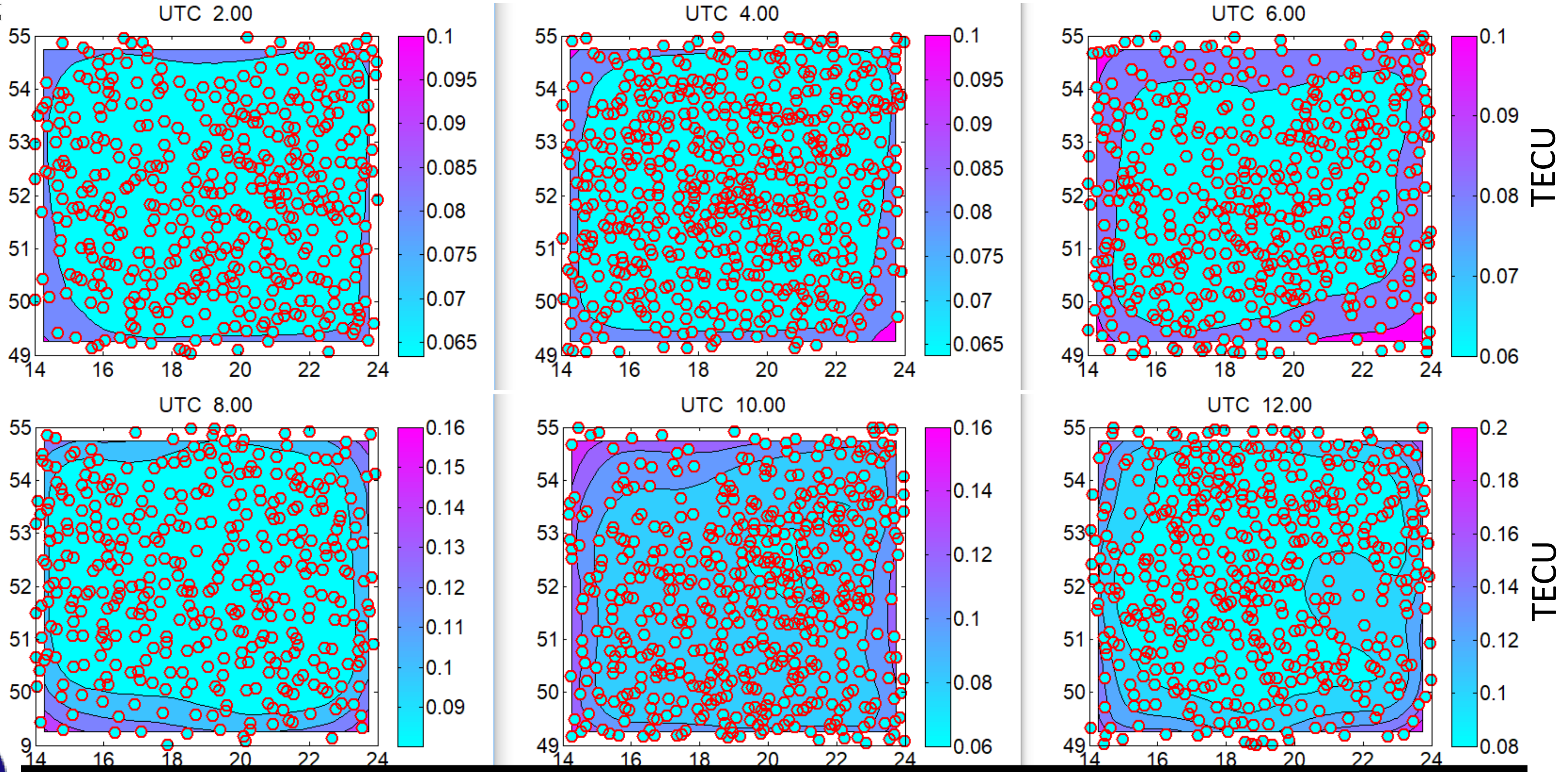
A posteriori error (accuracy of TEC) day 78 (2015), **δn = 0.2**



EUREF 2017 Symposium

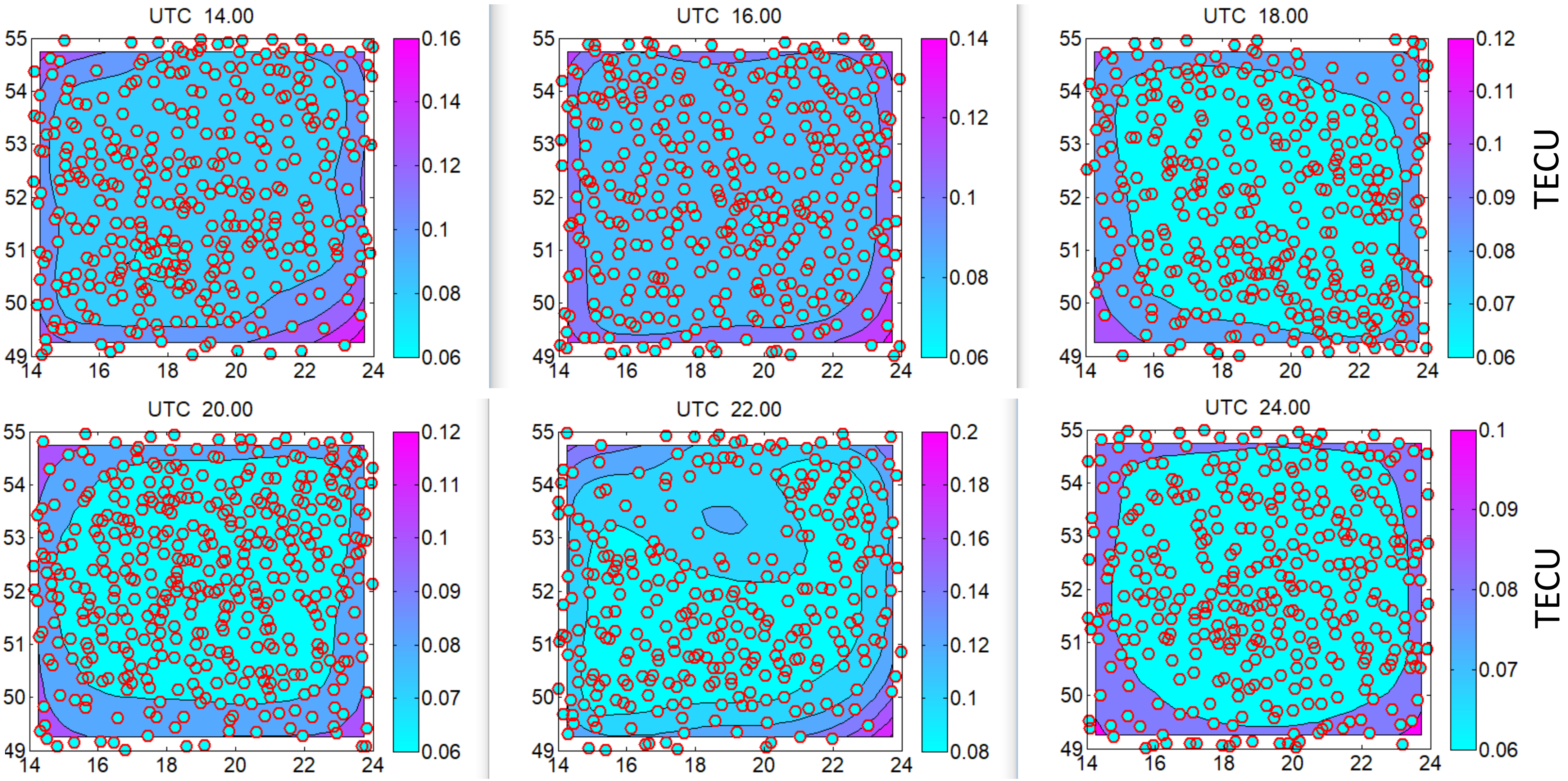
WROCŁAW May 17-19, 2017

A posteriori error (accuracy of TEC) day 78 (2015) , **δn = 0.5**



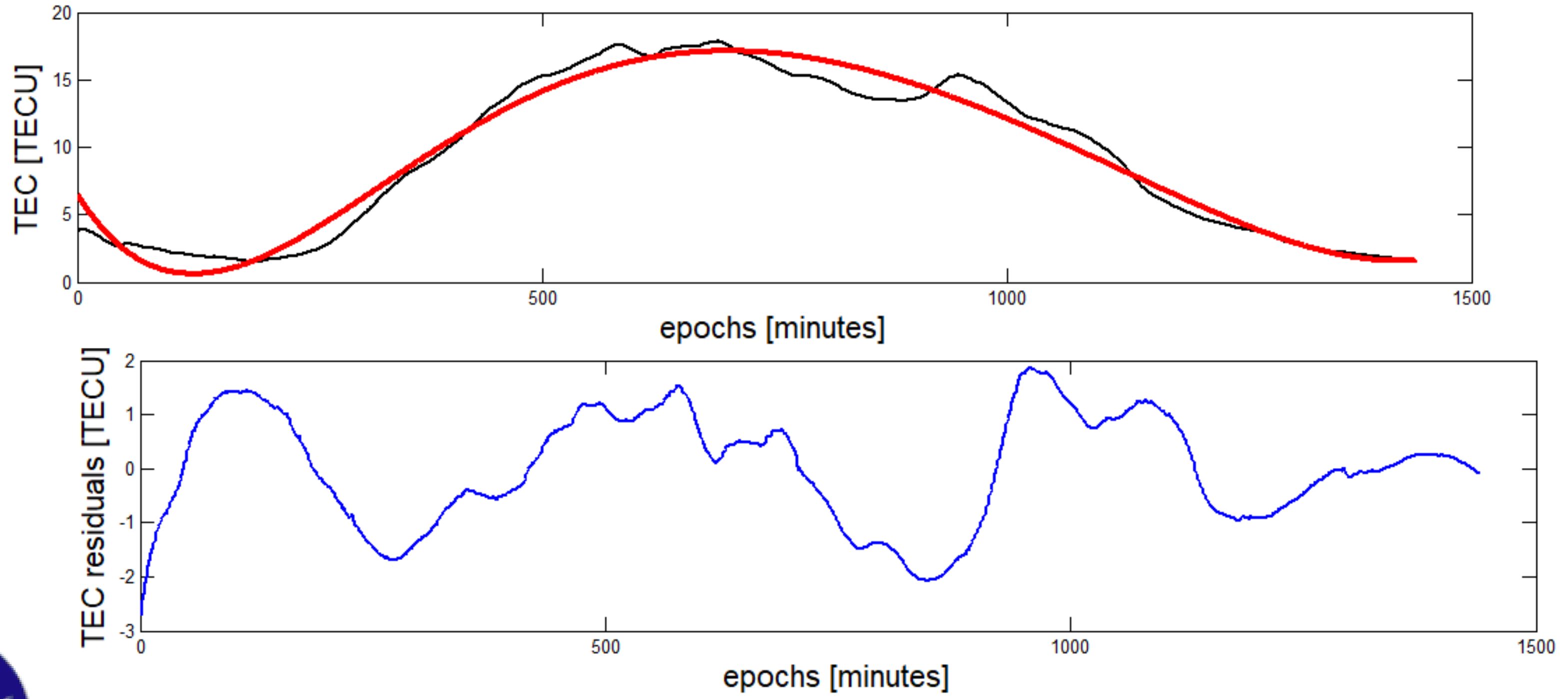
UNIWERSYTET WARMIŃSKO-MAZURSKI W OLSZTYNIE

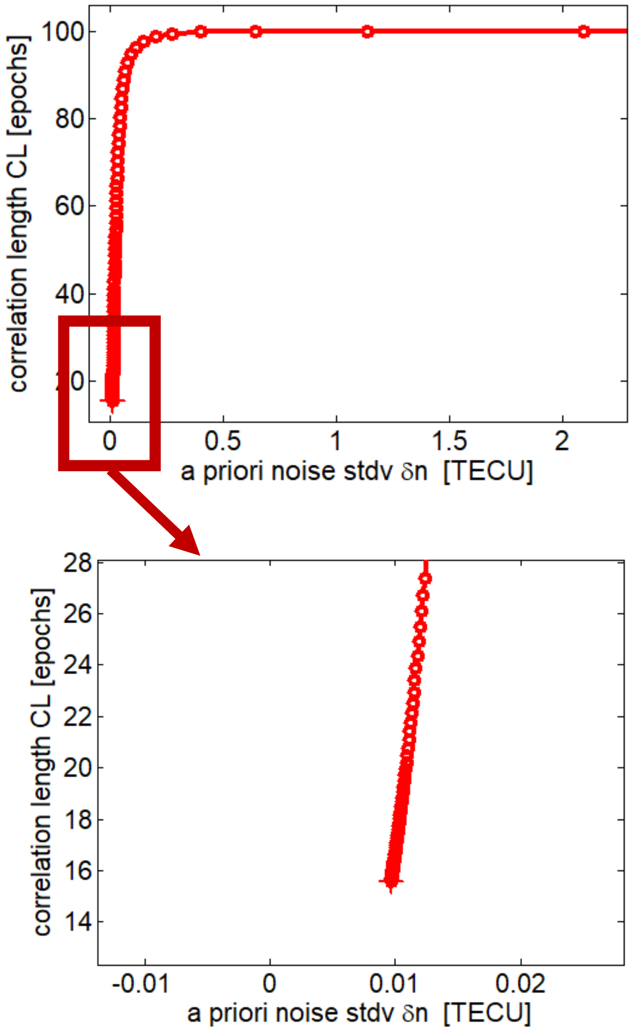
A posteriori error (accuracy of TEC) day 78 (2015) , **δn = 0.5**



Fisher scoring for TEC noise between time samples (minutes) (one selected node of the model, day 77, 2015)

The 6th order polynomial has been removed, however the time change stochastic proces in more non-stationary. Nevertheless, the noise is apparently smaller, FS may be true...





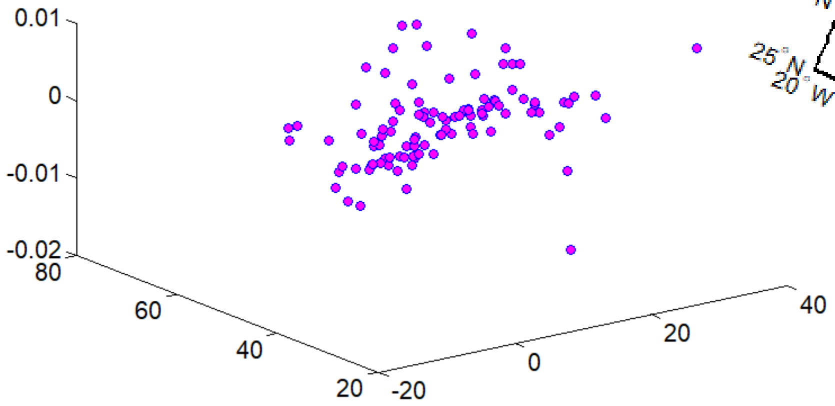
UNIWERSYTET WARMIŃSKO-MAZURSKI W OLSZTYNIE

Plate velocities example (vertical)

```
           CLASS_A EPN STATION POSITIONS AND VELOCITIES
           REFERENCE FRAME: ETRF2000 AT EPOCH OF 2005.0
           CUMULATIVE SOLUTION OF GPSWEEKS [   834 - 1934 ]
           RELEASE NAME: EPN_A_ETRF2000_C1934
           RELEASED ON 25/03/2017 BY EPN REFERENCE FRAME COORDINATOR(AMBRUS KENYERES, FOMI, HUNGARY)
DOMES NB. SITE NAME          TECH. ID.        X/Vx          Y/Vy          Z/Vz          Sigmas          SOLN  DATA_START
          CLASS                        ---------------------------m/m/Y------------------------
-------------------------------------------------------------------------------------------------------------------------
13434M001 ACOR                GPS ACOR  4594489.895  -678368.036  4357065.898  0.001  0.000  0.001  1 99:237:00000
13434M001       A                          -0.0016       0.0035      -0.0022 0.0001 0.0000 0.0000
13434M001 ACOR                GPS ACOR  4594489.890  -678368.032  4357065.900  0.000  0.000  0.000  2 00:044:00000
13434M001       A                          -0.0016       0.0035      -0.0022 0.0001 0.0000 0.0000
```

Data cover the period of observations
2001-2007
(from 6 to 21 years of obs.)

I assume no bias (no detrending)



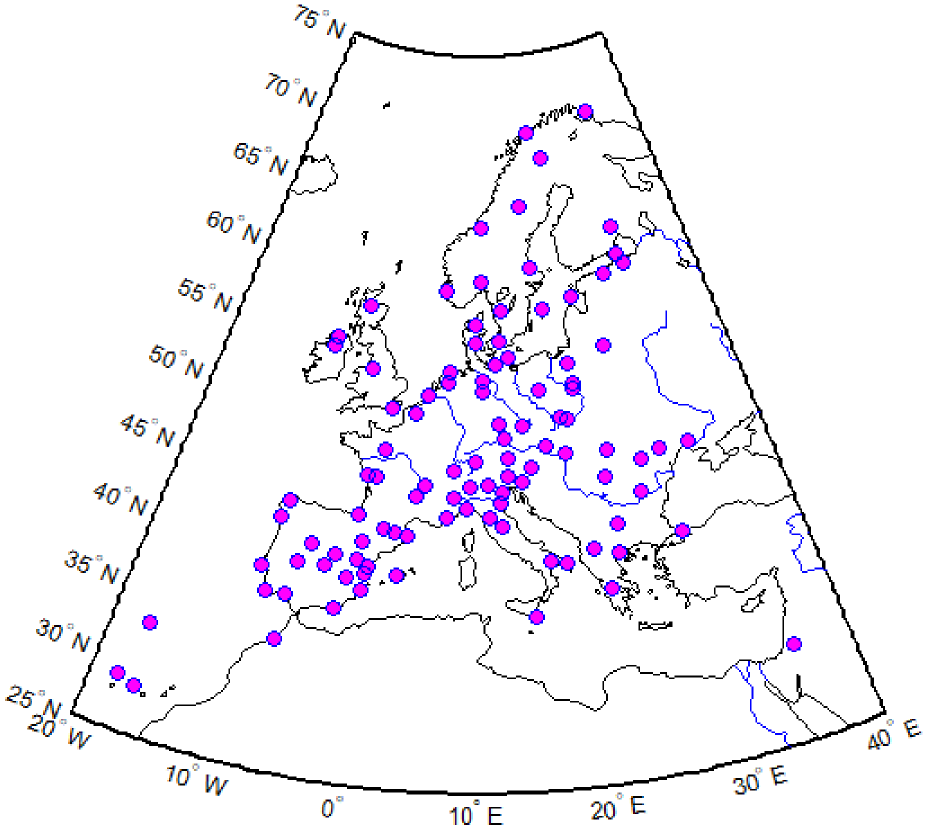
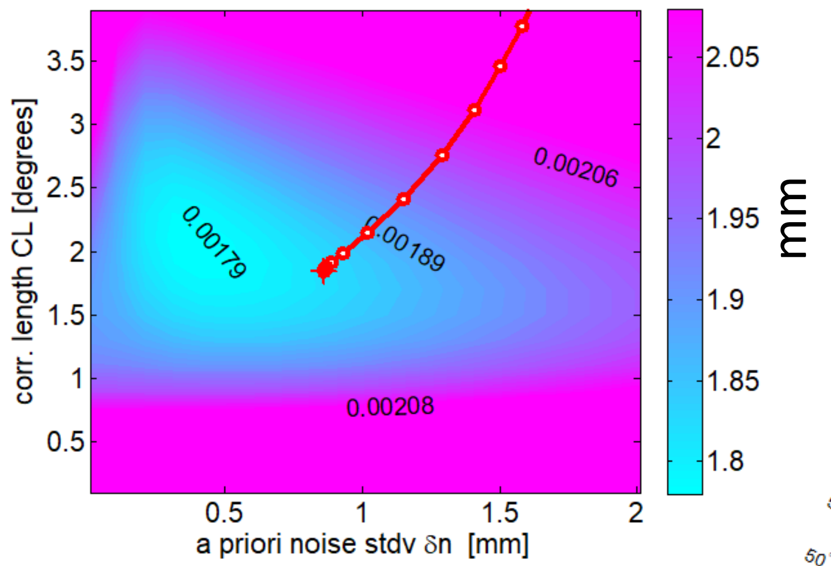
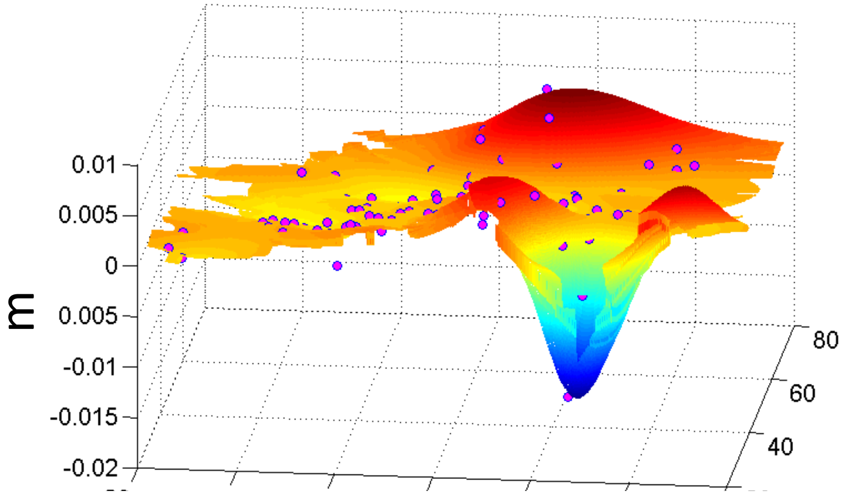


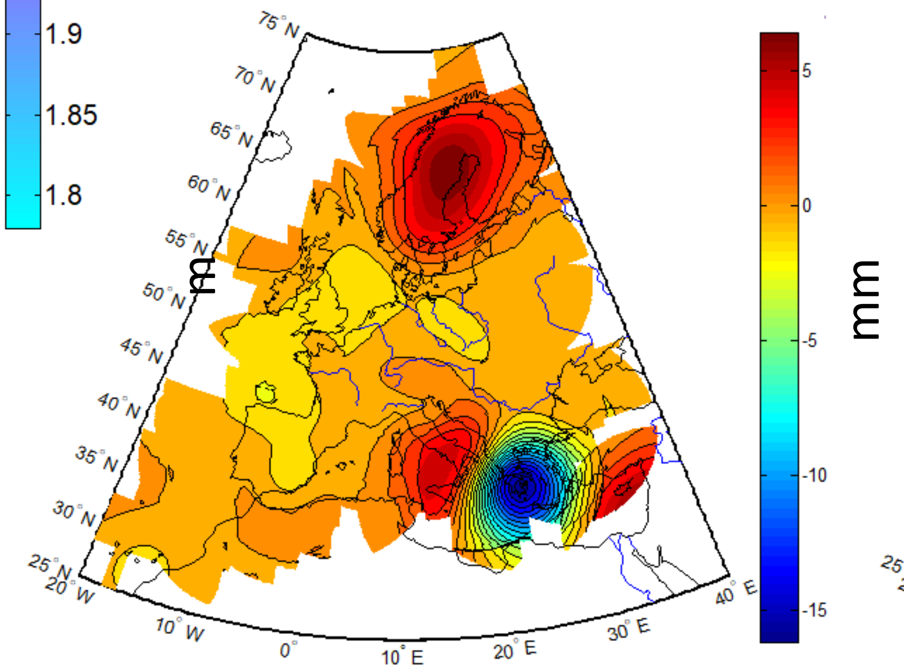
Plate velocities (vertical)

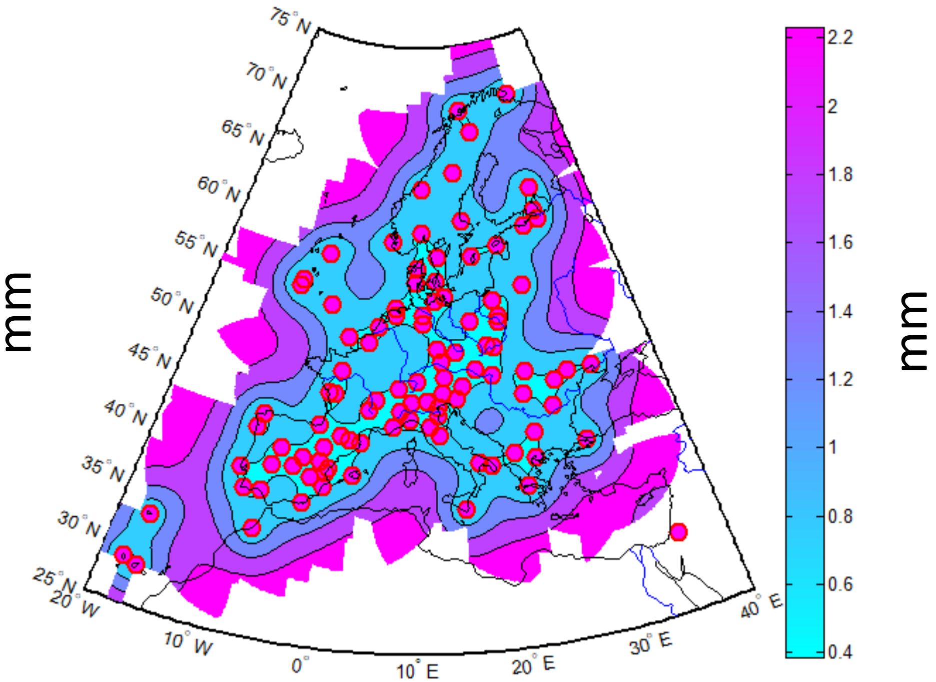


Fisher Scoring
+ LOO validation

Results and accuracy estimates







UNIWERSYTET WARMIŃSKO-MAZURSKI W OLSZTYNIE

Conclusions

- **Fisher Scoring works** evidently, however **data distribution issue** must be taken into account
- Non parametrized (or incorrectly parametrized) techniques can apllied only **if you do not expect accuracy** (eg. graphics or maps)
- **A larger a priori noise is always better** than underestimated (which occurs more frequently due to the link with survey error)
- Always compare and discuss **a priori noise, a posteriori error** and **RMS** from cross-validation