From formal errors towards realistic uncertainties

Leonid Petrov *(NASA GSFC)*, Nlingi Habana *(SSAI, Inc./NASA GSFC)*

Problem statement:

It is very common to report uncertainties that are unrealistic, usually underestimated. Consequences of reporting wrong uncertainties:

- inability to solve a scientific problem. In general, solving a scientific problem is reduced to testing a statistical hypothesis which requires a precise and reliable error distribution;
- inability to guide the measurement technique development. If we do not know what is accuracy of the current technology, we will likely make a wrong prediction of the impact of a proposed technology modification;
- inability to predict how much resources we need to allocate to reach the desired accuracy.

What is the origin of wrong uncertainties?

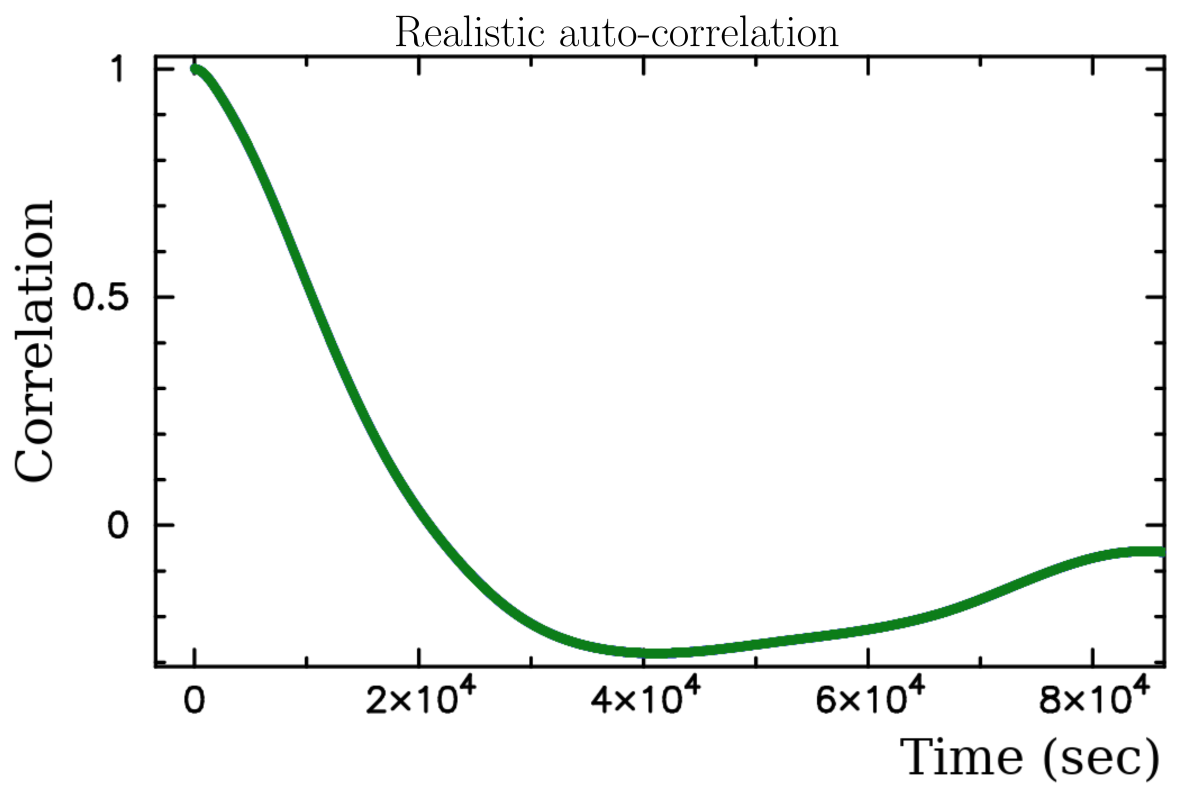
The most common problem is neglecting correlations and a belief to *"the square root of N law"*. Probably, it comes from a syllogism: if correlations are small (to infer statistical dependence), their impact (on error estimation) is also small, which is wrong.

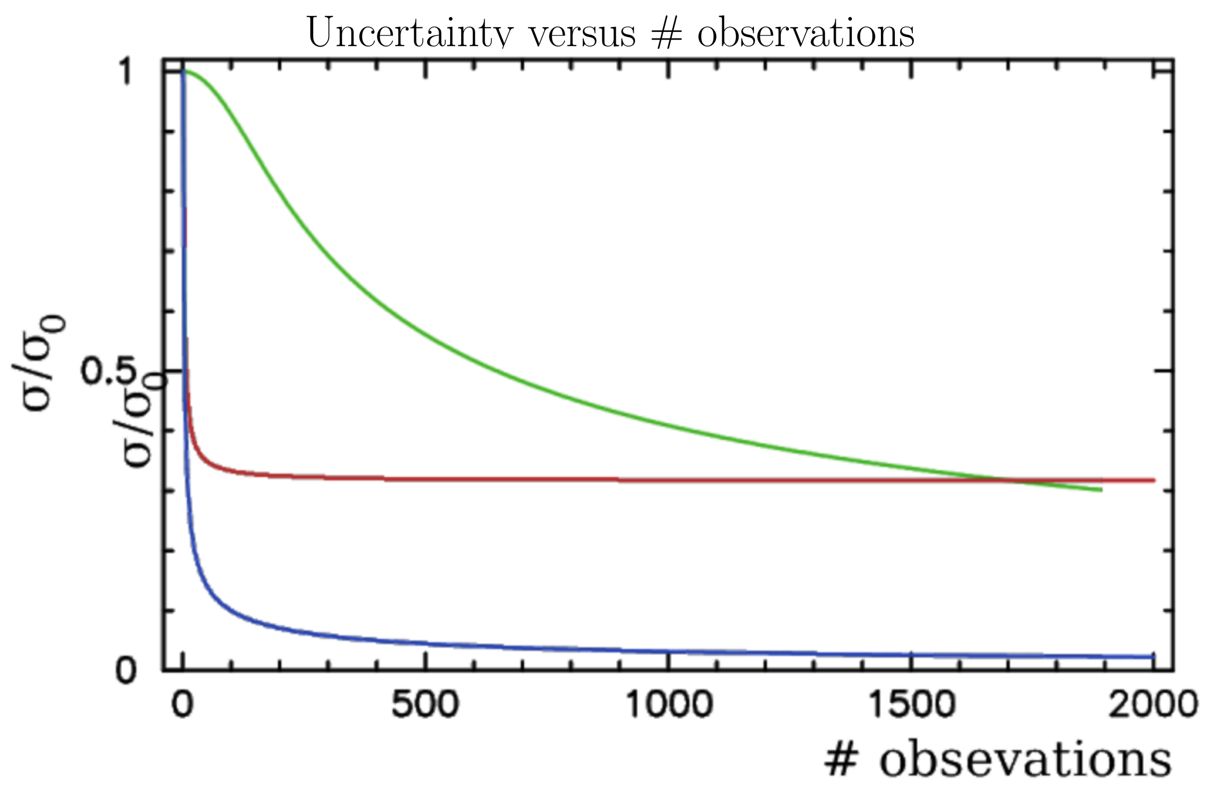
Will more observations improve geodetic results?

1 **Correlations $\rho = 0$; $\sigma_p \to 0$ rapidly** — **Covariance matrix is diagonal;**

2 **Correlations $\rho = \rho_0 =$ const; $\sigma_p \to \lim > 0$** — **Covariance matrix is constant, except main diagonal;**

3 **Correlations are stationary; $\sigma_p \to 0$ extremely slow** — **Covariance matrix is Töplitz**





What is the limit factor of space geodesy accuracy?

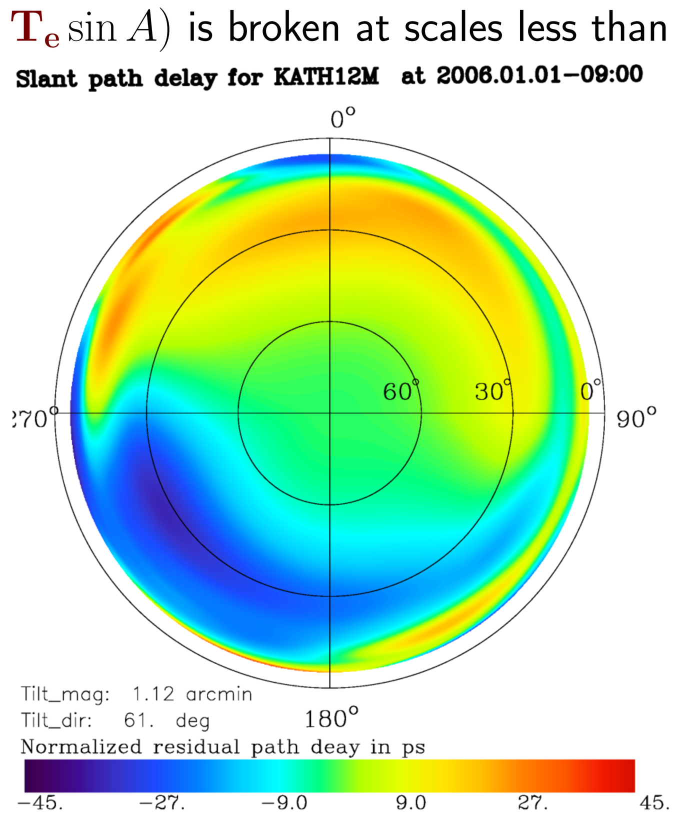
Path delay in the neutral atmosphere.

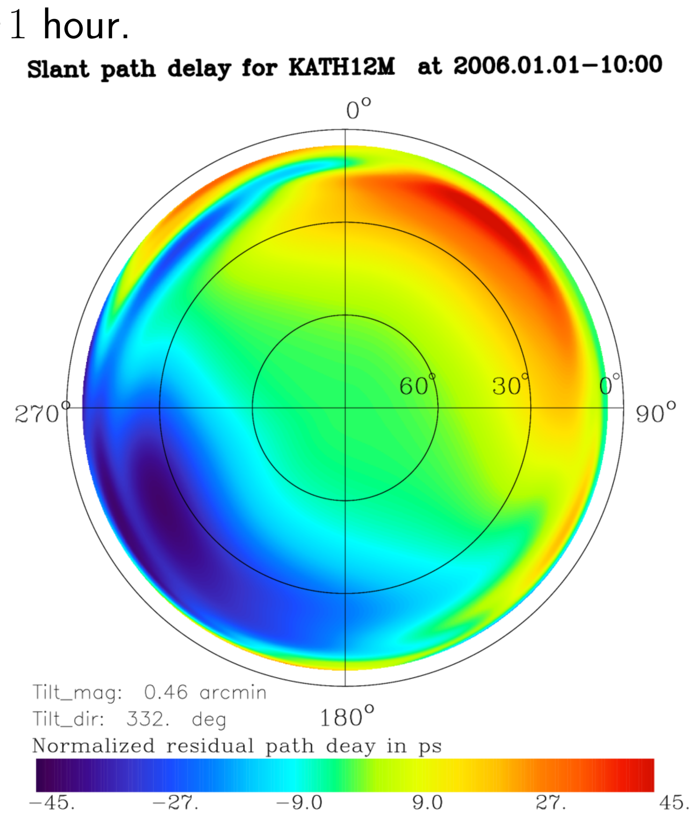
Why estimation of path delay in zenith is not enough?

Because $\tau(A,e) = \tau_z \cdot M(e + \mathbf{T_n} \cos A + \mathbf{T_e} \sin A)$ is broken at scales less than ~ 1 hour.

Residuals from fitting path delay in zenith direction and tilts of the refractivity symmetry axis to slant path delay computed from the high resolution 7 km numerical weather model.

These plots illustrate the time variability of the residual path delay.





Current approach:

The a priori covariance matrix is diagonal

$$
\mathrm{Cov}(x, x^{\top}) = \begin{pmatrix}
\sigma_1^2 + a_{ij}^2 & 0 & 0 & 0 & 0 \\
0 & \sigma_2^2 + a_{ij}^2 & 0 & 0 & 0 \\
0 & 0 & \sigma_3^2 + a_{ij}^2 & 0 & 0 \\
0 & 0 & 0 & \sigma_4^2 + a_{ij}^2 & 0 \\
0 & 0 & 0 & 0 & \sigma_5^2 + a_{ij}^2
\end{pmatrix}
$$

Parameter a_{ij} are found by iterations.

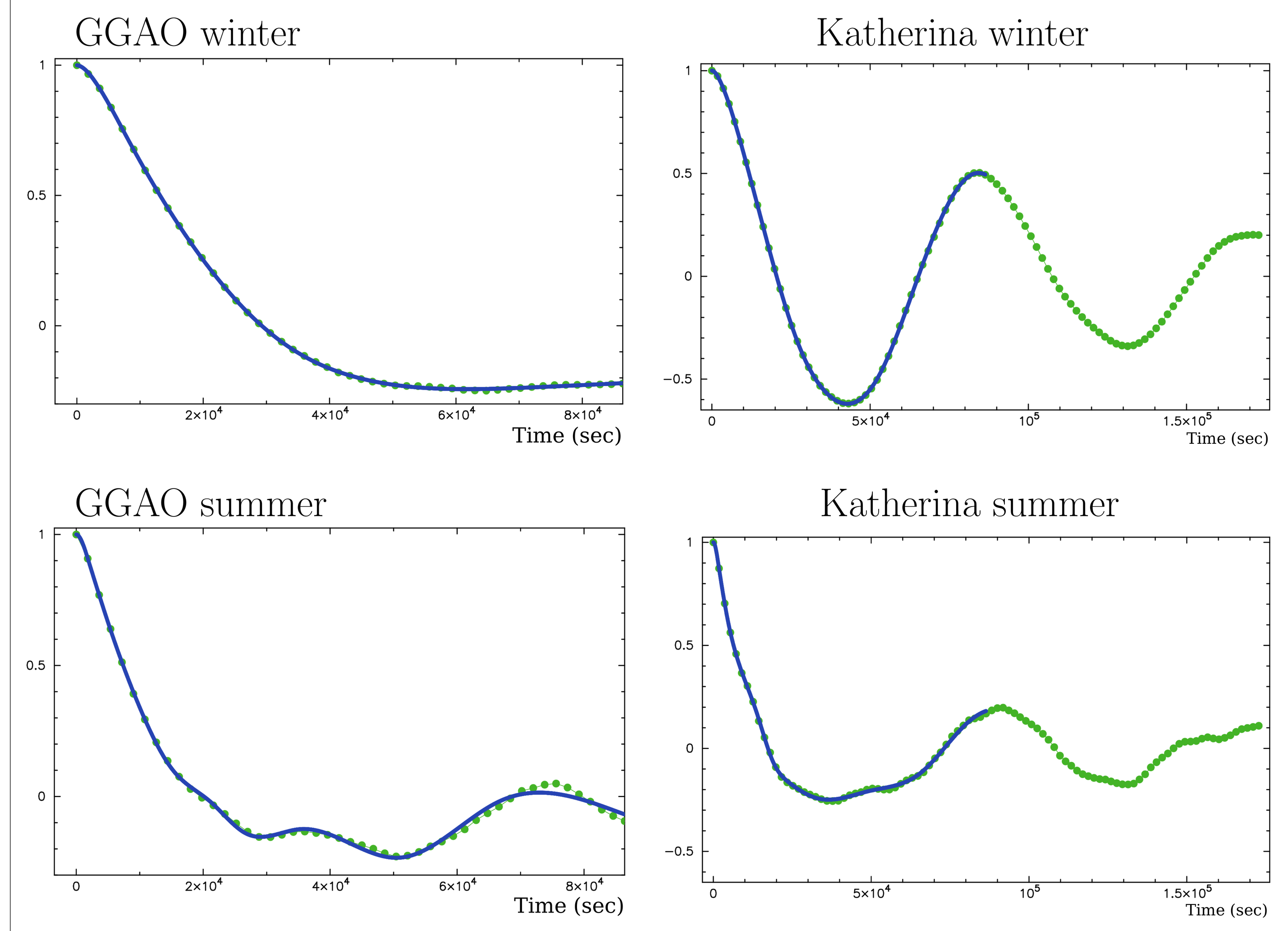
Next generation approach:

To use a full, Töplitz covariance matrix with off-diagonal terms.

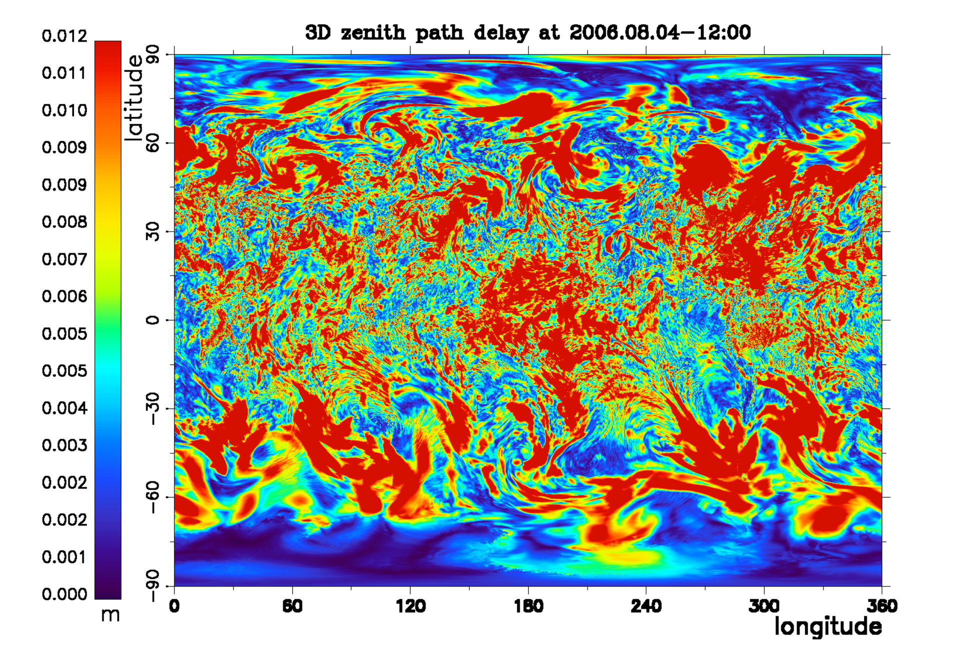
$$
\mathrm{Cov}(x, x^{\top}) = \sigma_0^2 \begin{pmatrix}
1 & \rho_a & \rho_b & \rho_c & \rho_d \\
\rho_a & 1 & \rho_a & \rho_b & \rho_c \\
\rho_b & \rho_a & 1 & \rho_a & \rho_b \\
\rho_c & \rho_b & \rho_a & 1 & \rho_a \\
\rho_d & \rho_c & \rho_b & \rho_a & 1
\end{pmatrix}
$$

Autocorrelation functions:

Auto-correlations that are computed from the NWM after fitting to zenith path delay and tilts (**the green circles**) and then approximation with a B-spline (**the blue line**):



The rms of zenith path delay over a 7 day period computed from the output of NASA GMAO high resolution NWM:



Difficulties:

- The approximated autocorrelation should correspond to a positively defined covariance matrix. Solution: to run the eigenvalue decomposition of the covariance matrix, update the minimum eigenvalue and repeat fitting the empirical autocorrelation.
- Autocorrelation is not stationary. Solution: to compute a representative autocorrelation for different seasons.
- An NWM is blind do autocorrelation at small lags. Solution: to run a dedicated VLBI program to assess the autocorrelation at lags of 1–60 minutes.

Road map:

- Process a high resolution NWM and derive the approximated auto-correlations for all the station for several seasons (done).
- Modify the LSQ solver to ingest auto-correlations, compute the a priori covariance matrices with all off-diagonal terms, and use them for computing normal equations (done).
- To propagate this code to the operational environment (prototype).
- To process a large dataset and investigate the realism of reported uncertainties (in the to do list).