Including (Station Dependent) (Correlated) Noise in VLBI Improves Your Estimates

John Gipson

NVI,Inc/NASA GSFC

Talk Given at

Technische Universität (TU) Wien

11-April-2011

Outline (1/2)

- Prior Work
- Review of Least Squares
- How correlation modifies the Normal Equations
- Problems in Paradise(=VLBI)
- Station Dependent Noise Improves Things
  - Diagonal or Lazy ~~Man’s~~ Geodesist’s Approximation
- What did we forget? (Hint: Correlation)

Outline (2/2)

- Evidence of Correlation
- Simplifying assumptions:
  - Only station and scan dependent correlation.
- Even better results.
- Tricks for speeding up the calculation
- Review & Conclusions
- Questions

Prerequisites

- Basic linear algebra
- Knowledge of least squares
- Basic calculus

Goals

Understand:

- Effect of correlated observations.
- Importance of incorporating in VLBI analysis.
- How to incorporate in VieVS

Prior Work

- Qian Zhi-han, (1985). “The Correlation on VLBI Observables and Its Effects for the Determination of ERP”. Shanghai Observatory.
- Schuh, H. and Wilkin A. (1989) “Determination of Correlation Coefficients between VLBI-Observables”.,Proc. Of the 7th Working Group Meeting on European VLBI.
- Schuh, H. and Tesmer, V: “Considering A Priori Correlations in VLBI Data Analysis” 2000 IVS General Meeting.
- Tesmer, V. (2003a) Refinement of the Stochastic VLBI Model: First Results. 16th Working Meeting on European VLBI.
- Tesmer, V., „Das stochastische Modell bei der VLBI-Anwertung”, PhD thesis, Munich, Germany, 2003.
- Tesmer, V., and Kutterer, H. (2004). “An advanced Stochastic Model for VLBI Observations and its Applications to VLBI Data Analysis”. 2004 IVS General Meeting Proceedings.
- Gipson: 2005 East Coast VLBI Meeting, 2006 IVS GM, 2007 EVGA, 2008 IVS GM

Review of Least Squares

1. Assume you have observations which are modeled as follows:

When considering FA summation is assumed.

Observations

$$O_j = F_j^a A_a + \varepsilon_j$$

Noise

Design Matrix or "Partials"

Parameters

Review of Least Squares

2. Furthermore, assume that the noise is uncorrelated between observations and you know the expected size of the noise:

$$\langle \varepsilon_j \varepsilon_k \rangle = \sigma_j^2 \delta_{jk}$$

Expectation value

Review of Least Squares

The least squares solution are the values of the parameters ***A*** which minimizes:

$$\chi^2 = \sum_j (O_j - F_j^a A_a) \frac{1}{\sigma_j^2} (O_j - F_j^b A_b)$$

Weighting by 1/(error squared)—observations with large error shouldn't effect the solution much.

Review of Least Squares

The minimum is found by differentiating with respect to ***A*:**

$$\frac{\partial \chi^2}{\partial A_a} = \sum_j F_j^a \frac{1}{\sigma_j^2}(O_j - F_j^b A_b) = 0$$

Or in matrix notation:

$$F^T \frac{1}{\sigma^2} F A - F^T \frac{1}{\sigma^2} O = 0$$

Review of Least Squares

Normal Equations:

$$\left[F^T \frac{1}{\sigma^2} F\right] A = F^T \frac{1}{\sigma^2} O$$

Normal Matrix

B-Vector

This has the formal solution:

$$\hat{A} = \left[F^T \frac{1}{\sigma^2} F\right]^{-1} F^T \frac{1}{\sigma^2} O$$

Hat indicates the solution to the least square equations—
NOT the real value of the parameters.

Review of Least Squares

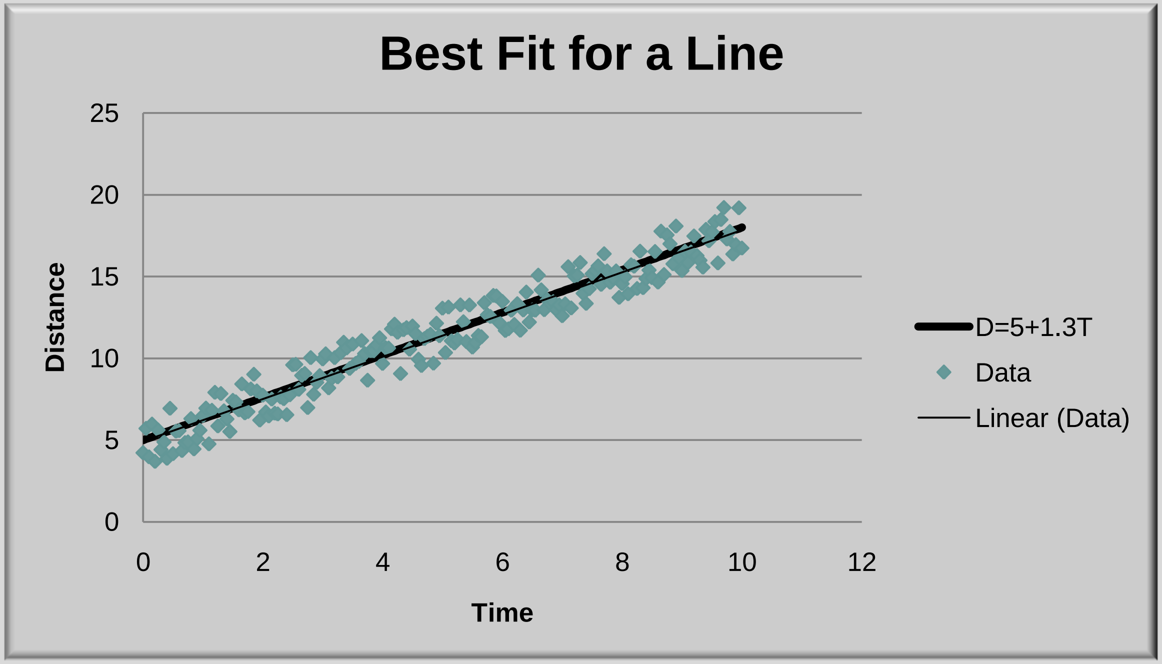
The uncertainty in the estimated parameter A_j is:

$$\sigma_j = \sqrt{\left[F^T \frac{1}{\sigma^2} F\right]^{-1}_{jj}}$$

No summation this time.

This assumes normally distributed noise.

Review of Least Squares



Least Squares Values are:
Offset = 5.126 ± 0.054
Rate = 1.286 ± 0.024

Detour: Expected Value of χ^2

One can show with modest assumptions that:

$$\left\langle \chi^2 \right\rangle = NumObs - NumPar$$

Usually deal with "reduced Chi-squared":

$$\chi^2_{red} \equiv \frac{\chi^2}{NumObs - NumPar}$$

This has the advantage that the expectation value is 1.

Detour: Expected Value of χ^2

$$\left\langle \chi^2_{red} \right\rangle = 1$$

If this differs significantly from 1, then one of our assumptions is wrong.

Detour: Expected Value of χ^2

$$\left\langle \chi^2_{red} \right\rangle = 1$$

If this differs significantly from 1, then one of our assumptions is wrong.

…. and we only made two:

$$O = FA + \varepsilon$$

$$\left\langle \varepsilon_j \varepsilon_k \right\rangle = \sigma_j^2 \delta_{jk}$$

Correlated Observations

Suppose that the first assumption is correct:

$$O = FA + \varepsilon$$

But our observations are correlated (and we know the correlation):

$$\left\langle \varepsilon_j \varepsilon_k \right\rangle = Cov_{jk}$$

What changes?

Correlated Observations

Can always diagonalize the covariance matrix:

$$SCovS^{-1} = \sigma'^2$$

Here S is a square orthogonal matrix: $S^{-1} = S^T$ of dimension NumObs.

Correlated Observations

Can always diagonalize the covariance matrix:

$$S Cov S^{-1} = \sigma'^2$$

Apply this transformation to assumption 1:

$$O = FA + \varepsilon \Rightarrow SO = SFA + S\varepsilon$$

Correlated Observations

Can always diagonalize the covariance matrix:

$$S Cov S^{-1} = \sigma'^2$$

Apply this transformation to assumption 1:

$$O = FA + \varepsilon \Rightarrow SO = SFA + S\varepsilon$$

Defining: $O' = SO \quad F' = SF \quad \varepsilon' = S\varepsilon$

Correlated Observations

Can always diagonalize the covariance matrix:

$$S Cov S^{-1} = \sigma'^2$$

Apply this transformation to assumption 1:

$$O = FA + \varepsilon \Rightarrow SO = SFA + S\varepsilon$$

Defining: $O' = SO \quad F' = SF \quad \varepsilon' = S\varepsilon$

…we have: $O' = F'A + \varepsilon'$

and by construction: $\langle \varepsilon'_j \, \varepsilon'_k \rangle = \sigma'^2_j \, \delta_{jk}$

Correlated Observations

Can always diagonalize the covariance matrix:

$$S Cov S^{-1} = \sigma'^2$$

Apply this transformation to assumption 1:

$$O = FA + \varepsilon \Rightarrow SO = SFA + S\varepsilon$$

Defining: $$O' = SO \quad F' = SF \quad \varepsilon' = S\varepsilon$$

…we have: $$O' = F' A + \varepsilon'$$

and by construction: $$\left\langle \varepsilon'_j \, \varepsilon'_k \right\rangle = \sigma'^2_j \, \delta_{jk}$$

Look familiar?

Correlated Observations

By changing basis we have diagonalized the covariance matrix. Everything we learned before goes through in terms of the new variables O'. We start with--

$$\chi^2 = \left(O'^T - A^T F'^T\right)\frac{1}{\sigma'^2}\left(O' - F' A\right)$$

And differentiate with respect to A to obtain the normal equations.

The main problem is finding the transformation *S*. It may be easier to work in the original basis.

Correlated Observations

Let's go back to the original basis:

Start with:

$$\chi^2 = \left(O'^T - A^T F'^T\right)\frac{1}{\sigma'^2}\left(O' - F'A\right)$$

$$\chi^2 = \left(O^T - A^T F^T\right)S^{-1}\frac{1}{\sigma'^2}S\left(O - FA\right)$$

By definition of O', F'

Moving S downstairs

$$\chi^2 = \left(O^T - A^T F^T\right)\frac{1}{S\sigma'^2 S^{-1}}\left(O - FA\right)$$

$$\chi^2 = \left(O^T - A^T F^T\right)\frac{1}{Cov}\left(O - FA\right)$$

Definition of Cov

Correlated Observations

Chi-square in the original basis is:

$$\chi^2 = \left(O^T - A^T F^T\right)\frac{1}{Cov}\left(O - FA\right)$$

Differentiating with respect to A and setting the result to 0 we obtain the normal equations:

$$\left[F^T \frac{1}{Cov} F\right] A = F^T \frac{1}{Cov} O$$

Problems in Paradise

Hints of problems in VLBI Session Analysis

χ_{red} for a single session is usually >2 or larger.

- This implies that we are not correctly modeling things on a session-by-session basis.
- The standard fix is to reweight the error to get $\chi_{red}^2 = 1$

$$\sigma_i^2 = \sigma_{i,meas}^2 + \sigma_{i,rewt}^2$$

- This increases the formal errors of estimates.

Problems in Paradise

Several Ways Reweighting Observations

1. Multiplicative reweighting. Doesn't change value of estimates.
2. Additive reweighting. Will change estimates.
   1. Add same constant for all observations.
   2. Add constants that depend on stations in observation.
   3. Add constants that depend on baseline. This is Goddard default.

Problems in Paradise

Several Ways Reweighting Observations

1. Multiplicative reweighting. Doesn’t change value of estimates.
2. Additive reweighting. Will change estimates.
    1. Add same constant for all observations.
    2. Add constants that depend on stations in observation.
    3. Add constants that depend on baseline. This is Goddard default.

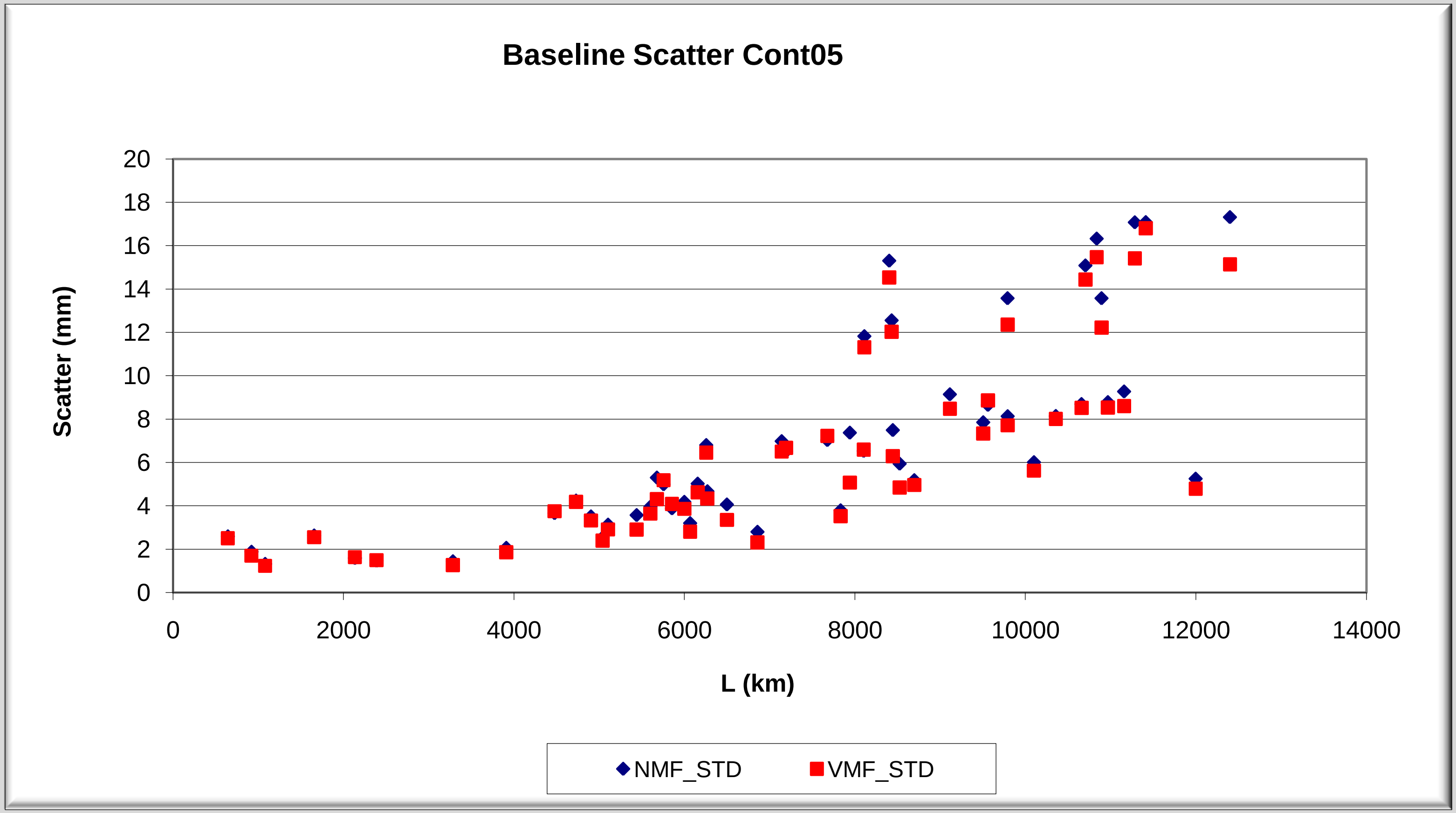
Reweighting tries to sweep problems under the rug.

More Problems in Paradise

Even after reweighting…

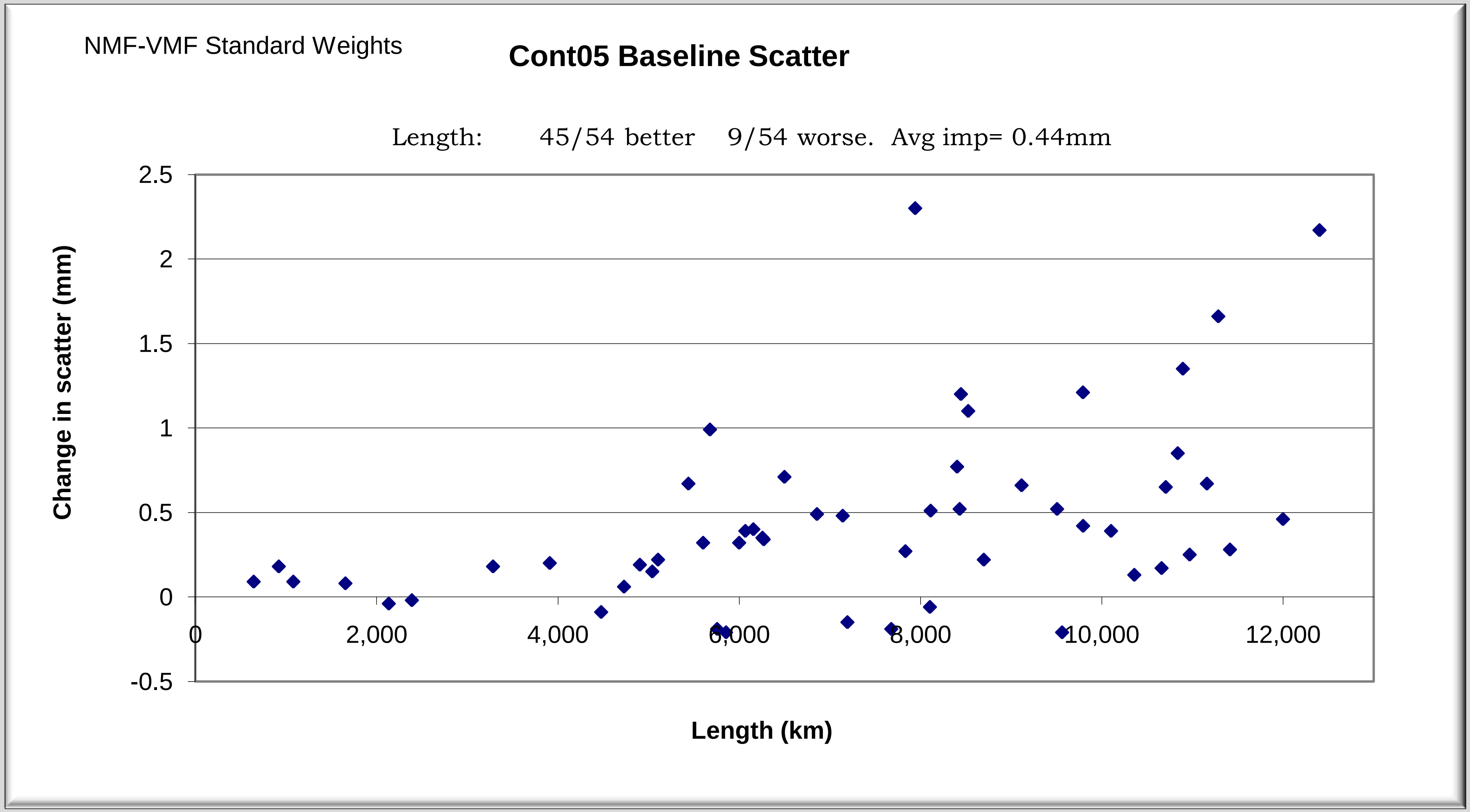
- Baseline scatter plots show too much variation based on formal errors of estimates. χ_{red} >2

Detour: Baseline Plots



This plots the baseline scatter as a function of baseline length. These are sometimes used to see if an alternative model is better.

Detour: Difference Plots



In comparing two options, it is usually better to plot the difference in scatter. For example, the above illustrates that VMF is better than NMF during CONT05.

More Problems in Paradise

Even after reweighting…

- Baseline scatter plots show too much variation based on formal errors of estimates.
- EOP estimates from simultaneous independent VLBI networks differ more than they should.
- Differences between VLBI and GPS derived Polar Motion are too large based on formal errors.

Review Our Assumptions

If we know the physics

$$O_j = \underbrace{F_j^a A_a}_{} + \varepsilon_j$$

Review Our Assumptions

$$\underbrace{O_j = F_j^a A_a}_{\text{If we know the physics}} + \varepsilon_j$$

… we must be wrong in our assumptions about the noise.

$$\langle \varepsilon_j \varepsilon_k \rangle = \sigma_j^2 \delta_{jk}$$

Station Dependent Noise

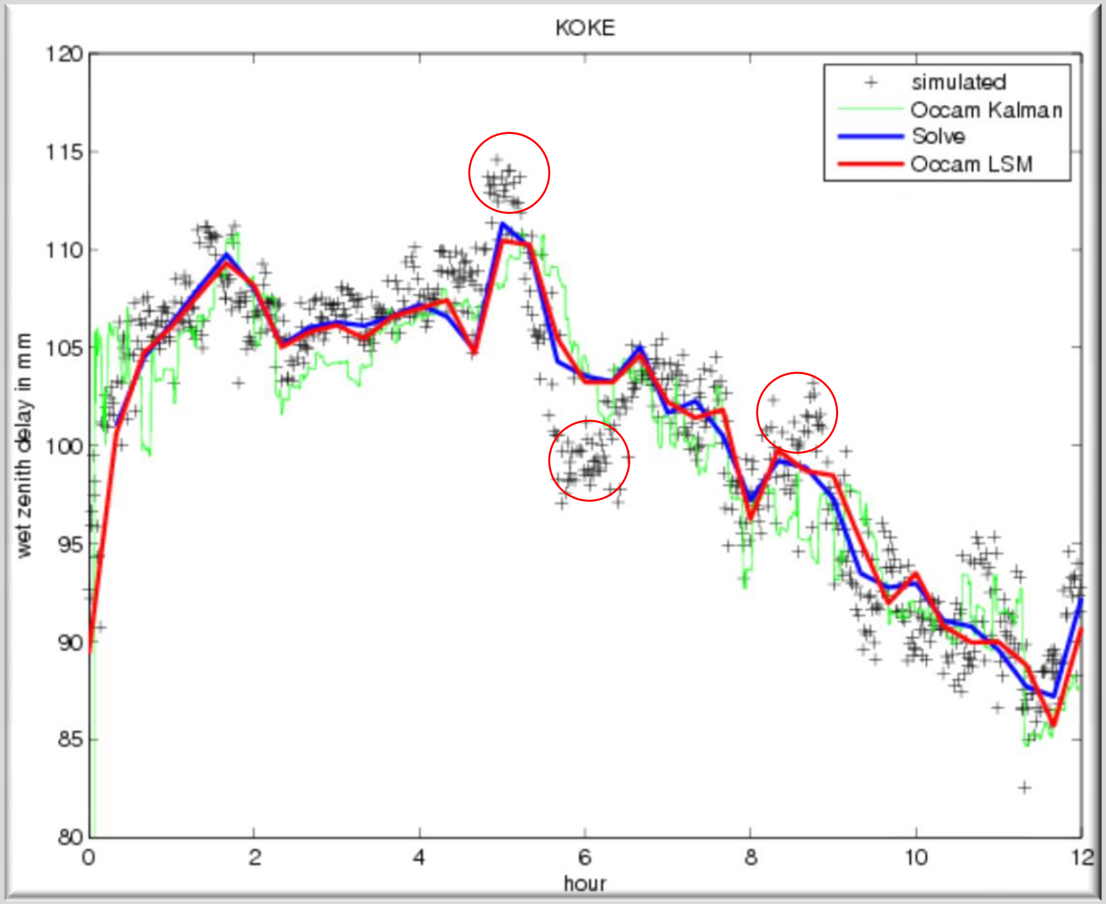
Several sources of station dependent noise:

1. Atmosphere modeling error.
2. Cable cal error.
3. Phase cal error.
4. Clock modeling error.
5. Loading corrections, hydrology, etc.

Many slowly varying noise terms will be soaked up in the clocks, or result in a shift of station position.
This will not effect Chi-square.

Evidence from Simulations

Solution departs from input data for extended periods.

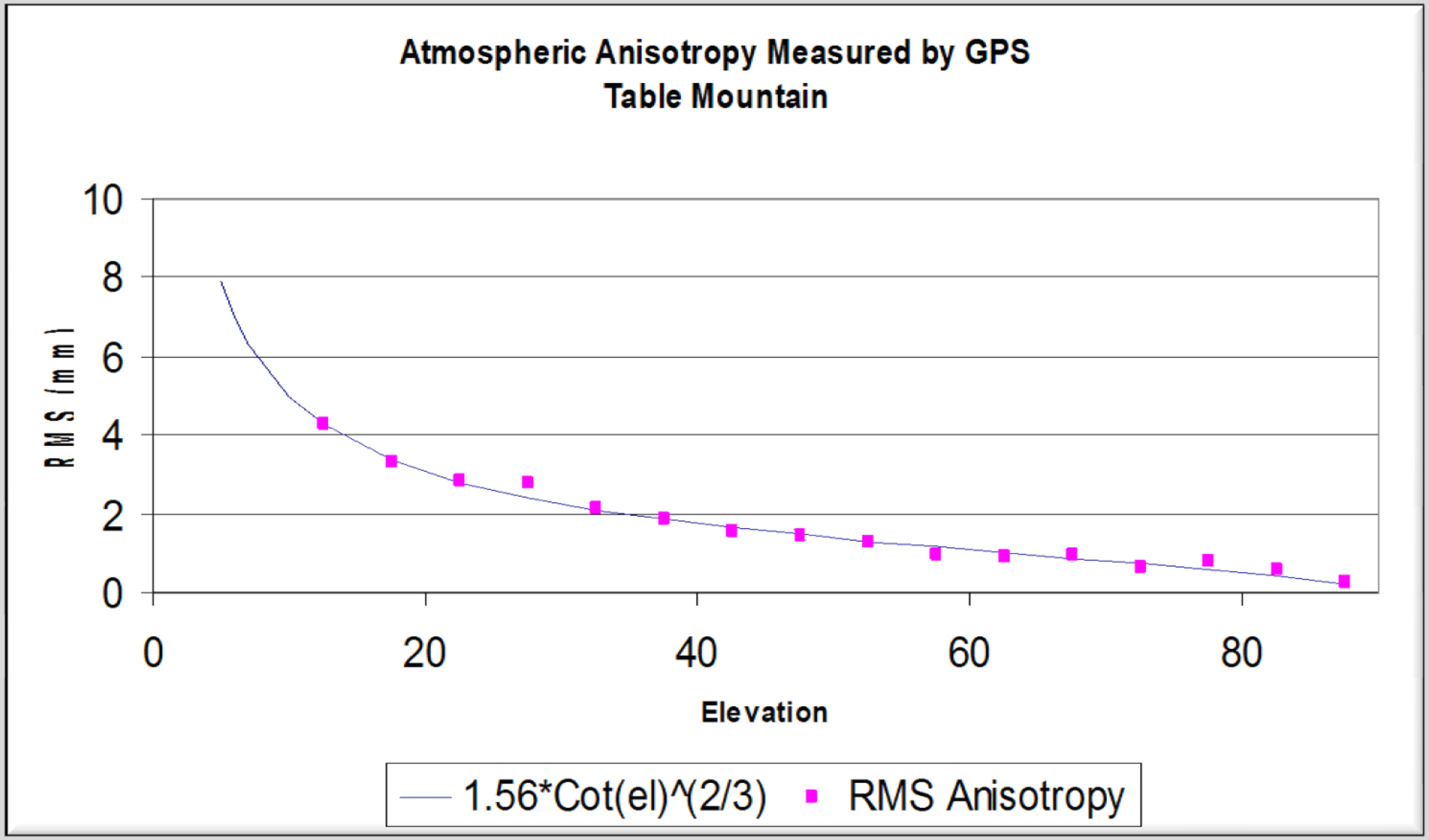


+'s indicate simulated atmosphere mapped to zenith based on realistic turbulence model from T. Nillson

Lines indicate recovered zenith delay from different analysis techniques.

Chart courtesy of J. Boehm. Data courtesy of T. Nillson,D. MacMillan, J. Boehm

Evidence from GPS



Residual scatter as a function of elevation determined by GPS. Functional form of curve is based on turbulence model.

Station Dependent Noise

ij, label the baseline

$$\sigma_{ij}^2 = \sigma_{ij,meas}^2 + \sigma_i^2 + \sigma_j^2$$

From correlator

Station dependent noise terms for stations *i* and *j*

Station Dependent Noise

ij, label the baseline

$$\sigma_{ij}^2 = \sigma_{ij,meas}^2 + \sigma_i^2 + \sigma_j^2$$

From correlator

Station dependent noise terms for stations i and j

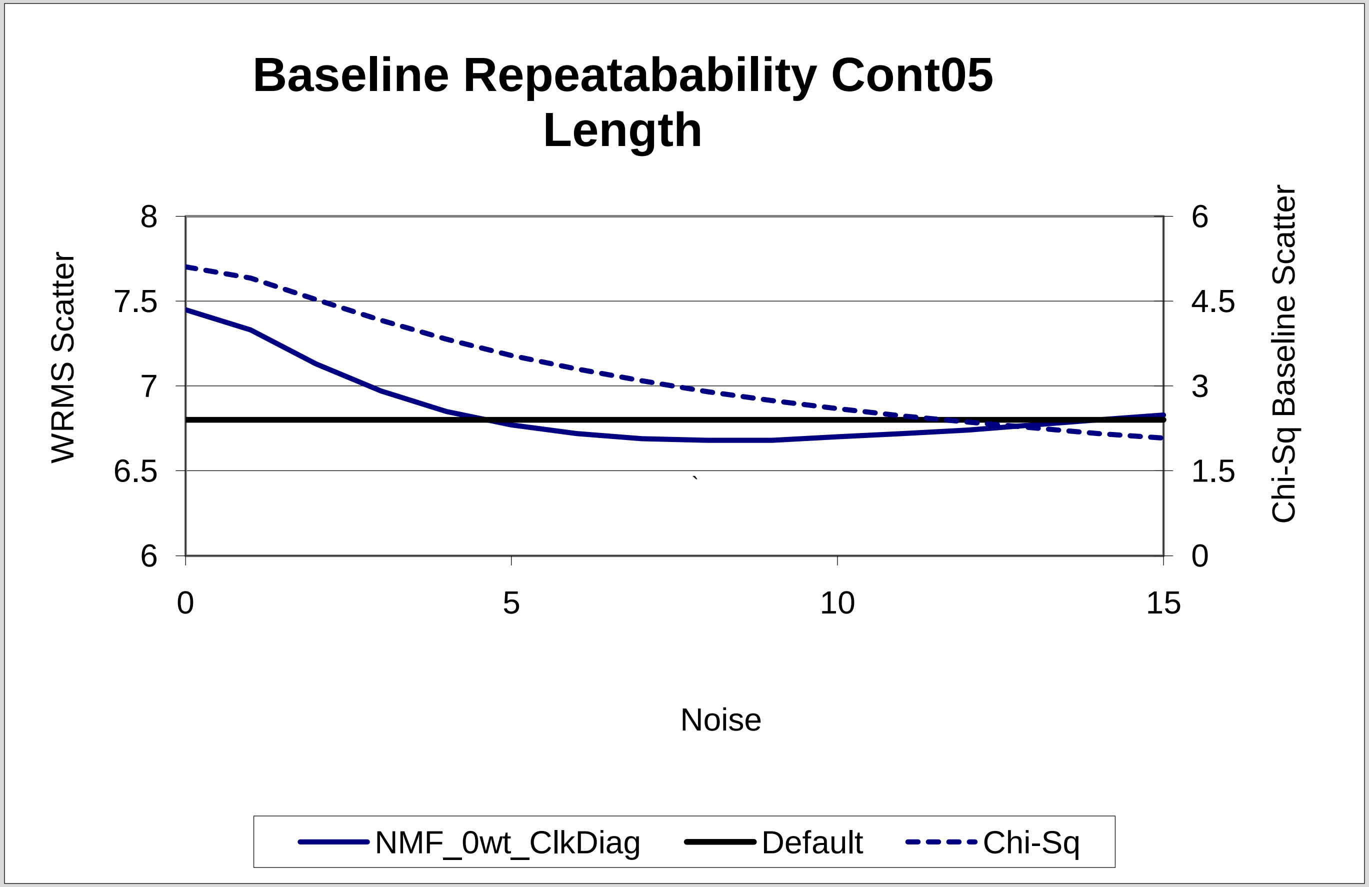
Asymmetry error

$$\sigma_i = c_i + a_i \times \cot^{2/3}(el_i) + m_i \times map(el_i)$$

Clock-like error

Mapping error

Effect of Adding Clock Error

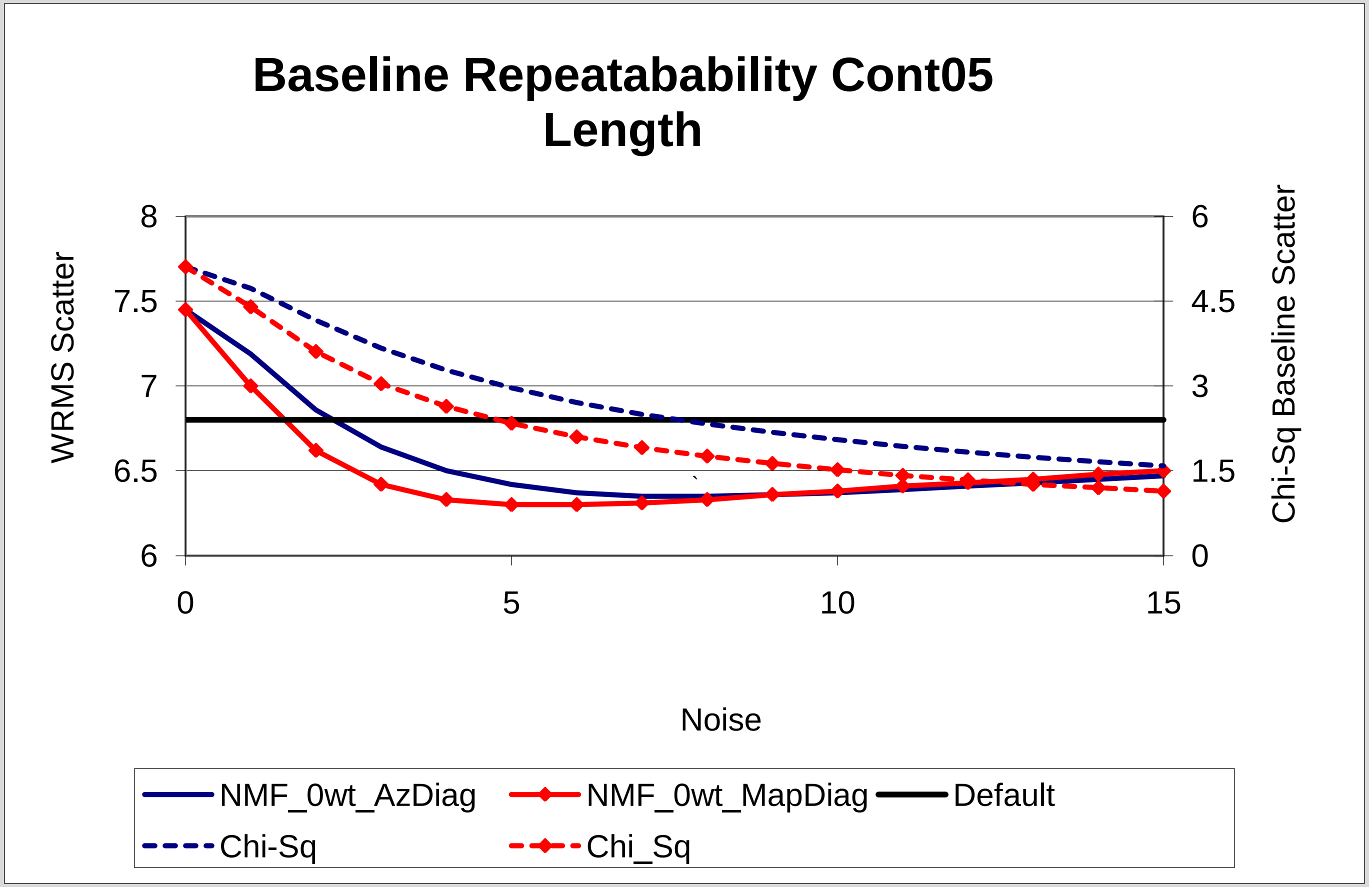


Adding clock-like noise results in:

1. Modest improvement compared to standard solution.

2. More realistic Chi-square.

Effect of Adding Atm Error



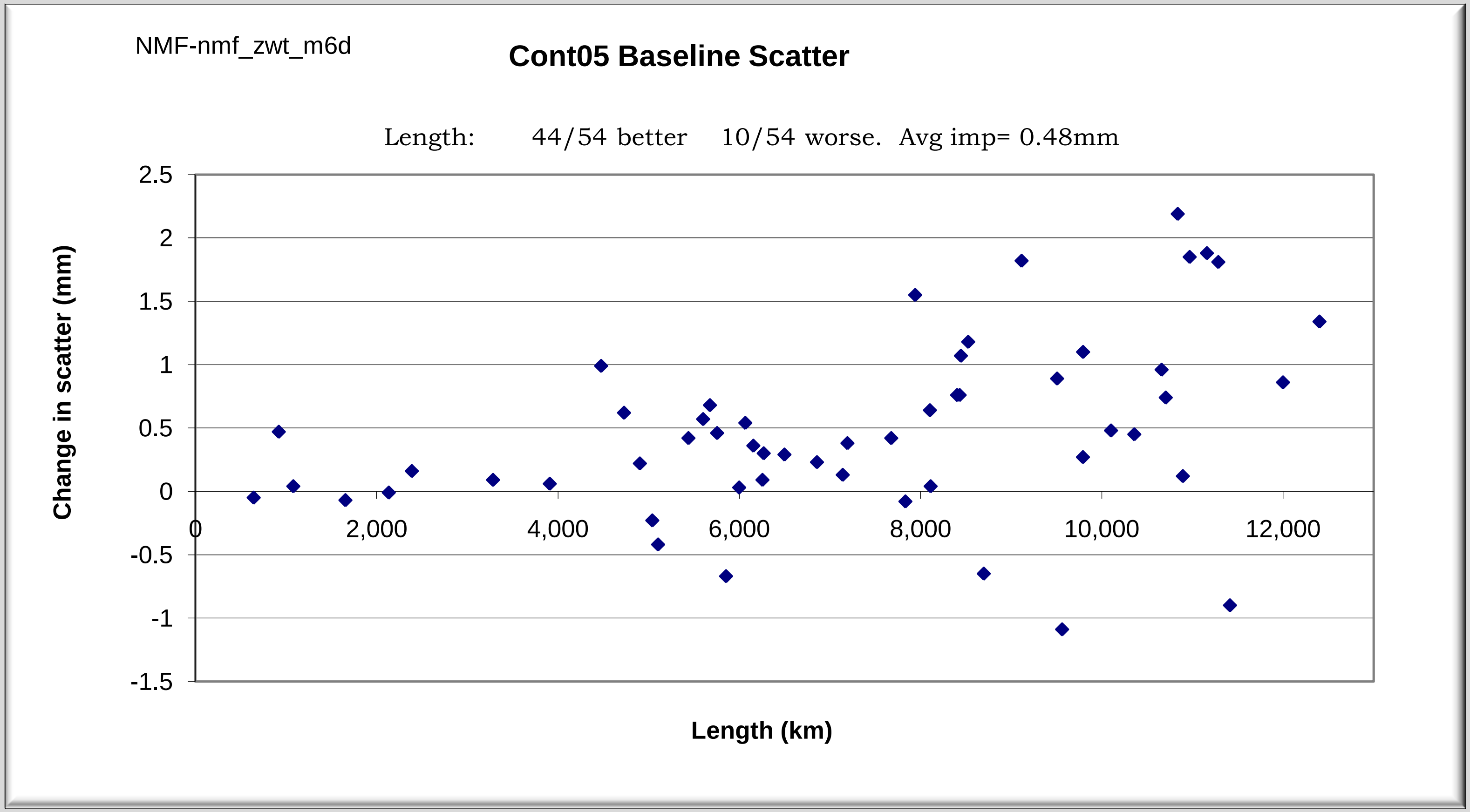
Adding atm-like noise results in:

1. Substantial improvement compared to standard solution.

2. Much more realistic Chi-square.

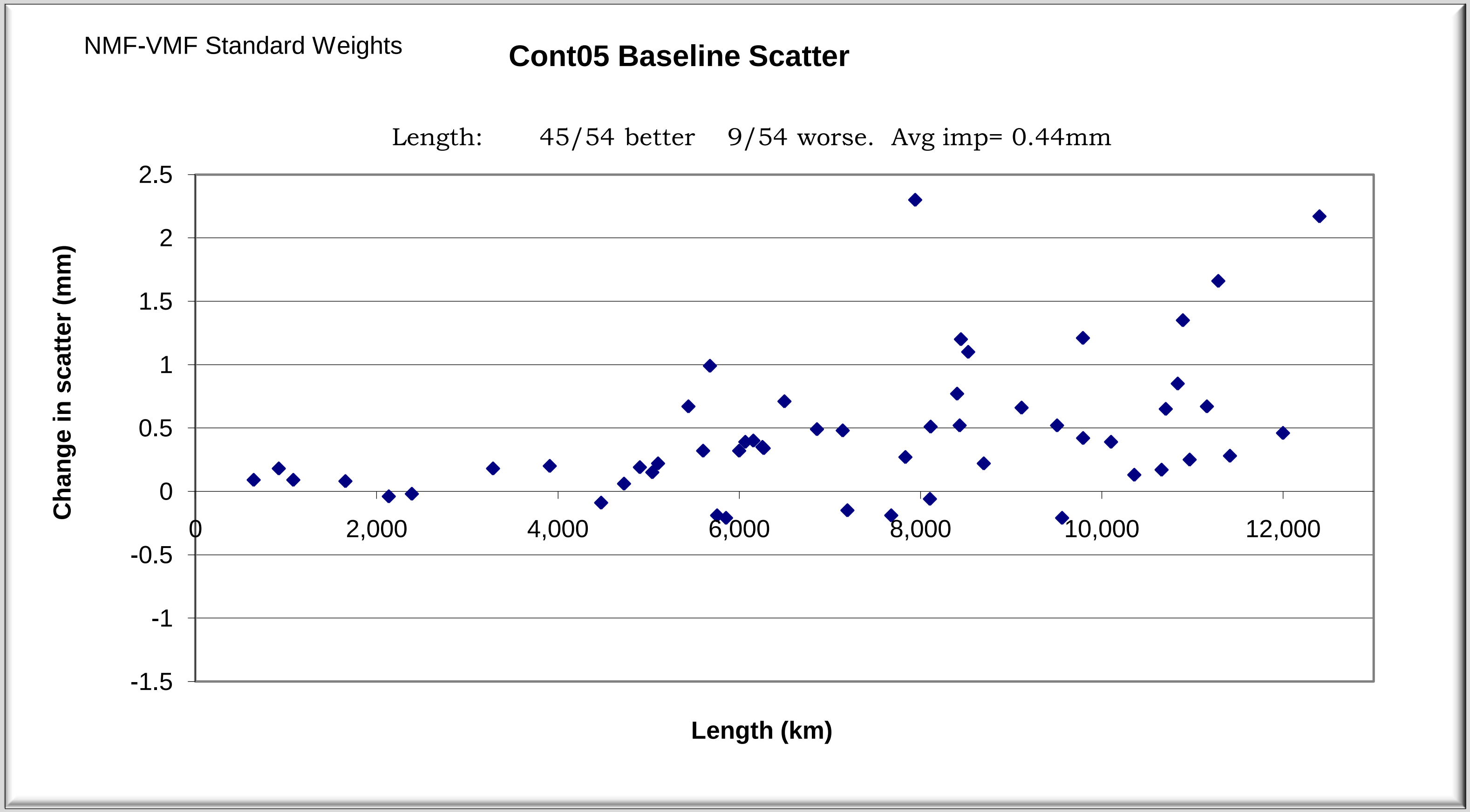
Effect of Mapping Noise is similar too, but better than, Az Noise.

Atm Baseline Difference Plot



The improvement in adding station dependent noise is 0.48 mm.
Slightly better than the improvement form using VMF which was 0.44 mm.

Reprise of VMF Effect



A repeat of the VMF chart for quick comparison.

We Forgot Something…

Station dependent noise introduces correlation between observations.

We Forgot Something…

Station dependent noise introduces correlation between observations.

$$O_j = F_j^a A_a + \varepsilon_j$$

Not only does this term change…

We Forgot Something…

Station dependent noise introduces correlation between observations.

$$O_j = F_j^a A_a + \varepsilon_j$$

Not only does this term change…

..but so does this…

$$\langle \varepsilon_j \varepsilon_k \rangle = Cov_{jk}$$

Correlation Between Observations

The VLBI observable is the differential delay:

Measurement noise

$$\tau(t)_{ij,meas} = \tau(t)_{ij,true} + \varepsilon(t)_{ij,meas}$$

Correlation Between Observations

The VLBI observable is the differential delay:

Measurement noise

$$\tau(t)_{ij,meas} = \tau(t)_{ij,true} + \underbrace{\varepsilon(t)_{ij,meas} + \overbrace{\varepsilon(t)_i - \varepsilon(t)_j}^{\text{Station dependent noise.}}}_{\text{Total noise}}$$

Correlation Between Observations

The VLBI observable is the differential delay:

$$\tau(t)_{ij,meas} = \tau(t)_{ij,true} + \varepsilon(t)_{ij,meas} + \varepsilon(t)_i - \varepsilon(t)_j$$

Assuming this form, what is the covariance?

Correlation Between Observations

The VLBI observable is the differential delay:

$$\tau(t)_{ij,meas} = \tau(t)_{ij,true} + \varepsilon(t)_{ij,meas} + \varepsilon(t)_i - \varepsilon(t)_j$$

Assuming this form, what is the covariance?

$$\left\langle \varepsilon_{ij}(t), \varepsilon_{lm}(t') \right\rangle = \left\langle \varepsilon_{ij,meas}(t) + \varepsilon_i(t) - \varepsilon_j(t), \varepsilon_{lm,meas}(t') + \varepsilon_l(t') - \varepsilon_m(t') \right\rangle$$

Correlation Between Observations

Simplifying Assumptions

$$\left\langle \varepsilon_{ij}(t), \varepsilon_{ij}(t') \right\rangle = \delta_{t,t'} \times \{something\}$$

Observations at different times are uncorrelated.

$$\left\langle \varepsilon_{ij,meas}, \varepsilon_{i} \right\rangle = 0$$

Station noise is uncorrelated with measurement noise.

$$\left\langle \varepsilon_{i}, \varepsilon_{j} \right\rangle = \delta_{ij} \left\langle \varepsilon_{i}^{2} \right\rangle$$

Station noise at different stations is uncorrelated.

Correlation Between Observations

Based on our assumptions we only need consider observations at a common time with one or two stations in common.

Correlation Between Observations

Based on our assumptions we only need consider observations at a common time with one or two stations in common.

$$\left\langle \varepsilon_{ij}, \varepsilon_{ij} \right\rangle = \left\langle \varepsilon_{ij,meas}^2 \right\rangle + \left\langle \varepsilon_i^2 \right\rangle + \left\langle \varepsilon_j^2 \right\rangle$$

Diagonal term.
This is what we just did.

Correlation Between Observations

Based on our assumptions we only need consider observations at a common time with one or more stations in common.

$$\left\langle \varepsilon_{ij}, \varepsilon_{ij} \right\rangle = \left\langle \varepsilon_{ij,meas}^2 \right\rangle + \left\langle \varepsilon_i^2 \right\rangle + \left\langle \varepsilon_j^2 \right\rangle$$

Diagonal term.
This is what we just did.

$$\left\langle \varepsilon_{ij}, \varepsilon_{il} \right\rangle = \left\langle \varepsilon_{ji}, \varepsilon_{li} \right\rangle = \left\langle \varepsilon_i^2 \right\rangle$$

The first or second station on two observations is the same.

Correlation Between Observations

Based on our assumptions we only need consider observations at a common time with one or more stations in common.

$$\langle \varepsilon_{ij}, \varepsilon_{ij} \rangle = \langle \varepsilon^2_{ij,meas} \rangle + \langle \varepsilon^2_i \rangle + \langle \varepsilon^2_j \rangle$$

Diagonal term.
This is what we just did.

$$\langle \varepsilon_{ij}, \varepsilon_{il} \rangle = \langle \varepsilon_{ji}, \varepsilon_{li} \rangle = \langle \varepsilon^2_i \rangle$$

The first or second station on two observations is the same.

$$\langle \varepsilon_{ij}, \varepsilon_{li} \rangle = \langle \varepsilon_{ji}, \varepsilon_{il} \rangle = -\langle \varepsilon^2_i \rangle$$

The first station in one observation is the second station in the other observation.

Correlation Between Observations

Based on our assumptions we only need consider observations at a common time with one or more stations in common.

$$\left\langle \varepsilon_{ij}, \varepsilon_{ij} \right\rangle = \left\langle \varepsilon_{ij,meas}^2 \right\rangle + \left\langle \varepsilon_i^2 \right\rangle + \left\langle \varepsilon_j^2 \right\rangle$$

Diagonal term.
This is what we just did.

$$\left\langle \varepsilon_{ij}, \varepsilon_{il} \right\rangle = \left\langle \varepsilon_{ji}, \varepsilon_{li} \right\rangle = \left\langle \varepsilon_i^2 \right\rangle$$

The first or second station on two observations is the same.

$$\left\langle \varepsilon_{ij}, \varepsilon_{li} \right\rangle = \left\langle \varepsilon_{ji}, \varepsilon_{il} \right\rangle = -\left\langle \varepsilon_i^2 \right\rangle$$

The first station in one observation is the second station in the other observation.

All other elements are 0.

Evidence of Correlation in VLBI

RDV48

Delay Residuals for Source 1418+546 @ 2004-12-01-18:06:59

		BR	FD	HN	KP	LA	NL	OV	PI	SC	WS	ME	ON	TS	#	AVG	SIG	Avg/Sig
1	**BR-VLBA**	-	-5	31	-20	-18	1	12	-21	8	36	-19	-11	-8	12	-6.1	2.4	-2.6
2	**FD-VLBA**	5	-	31	31	14	22	15	7	16	2	4	-39	-3	12	14.1	2.1	**6.8**
3	**HN-VLBA**	-31	-31	-	-6	-29	-6	-5	-22	-12	2	-12	-63	-31	12	-17.2	3.0	**-5.7**
4	**KP-VLBA**	20	-31	6	-	10	1	-45	-3	20	-11	-	-	46	10	-1.2	3.6	-0.3
5	**LA-VLBA**	18	-14	29	-10	-	2	-14	-3	-11	-22	-24	12	-24	12	-2.8	2.1	-1.5
6	**NL-VLBA**	-1	-22	6	-1	-2	-	-11	-14	20	-8	-17	-39	-17	12	-7.8	1.9	**-4.1**
7	**OV-VLBA**	-12	-15	5	45	14	11	-	14	-24	0	8	16	-36	12	-1.34	2.3	-0.6
8	**PIETOWN**	21	-7	22	3	3	14	-14	-	20	11	-21	16	-3	12	6.2	1.8	**3.6**
9	**SC-VLBA**	-8	-16	12	-20	11	-20	24	-20	-	6	4	64	23	12	-3.1	2.9	-1.1
10	**WESTFRD**	-36	-2	-2	11	22	8	0	-11	-6	-	34	-	32	11	0.24	4.3	0.1
11	**MEDICINA**	19	-4	12	-	24	17	-8	21	-4	-34	-	55	-28	11	2.8	4.3	0.7
12	**ONSALA60**	11	39	63	-	-12	39	-16	-16	-64	-	-55	-	-62	10	-13.6	12.0	-1.1
13	**TSUKUB32**	8	3	31	-46	24	17	36	3	-23	-32	28	62	-	12	13.9	3.4	**4.0**

Correlation Between Observations

By assumption, there is no correlation between different observations on different scans.

$$Cov = \begin{bmatrix} Cov_{Scan1} & 0 & 0 & 0 \\ 0 & Cov_{Scan2} & 0 & 0 \\ 0 & 0 & Cov_{Scan3} & 0 \\ 0 & 0 & 0 & \ldots \end{bmatrix}$$

The covariance matrix is block diagonal.

Correlation Between Observations

By assumption, there is no correlation between different observations on different scans.

$$Cov = \begin{bmatrix} Cov_{Scan1} & 0 & 0 & 0 \\ 0 & Cov_{Scan2} & 0 & 0 \\ 0 & 0 & Cov_{Scan3} & 0 \\ 0 & 0 & 0 & \ldots \end{bmatrix}$$

The covariance matrix is block diagonal.

This makes it easier to invert:

$$Cov^{-1} = \begin{bmatrix} Cov_{Scan1}^{-1} & 0 & 0 & 0 \\ 0 & Cov_{Scan2}^{-1} & 0 & 0 \\ 0 & 0 & Cov_{Scan3}^{-1} & 0 \\ 0 & 0 & 0 & \ldots \end{bmatrix}$$

Normal Equations

The normal equations are:

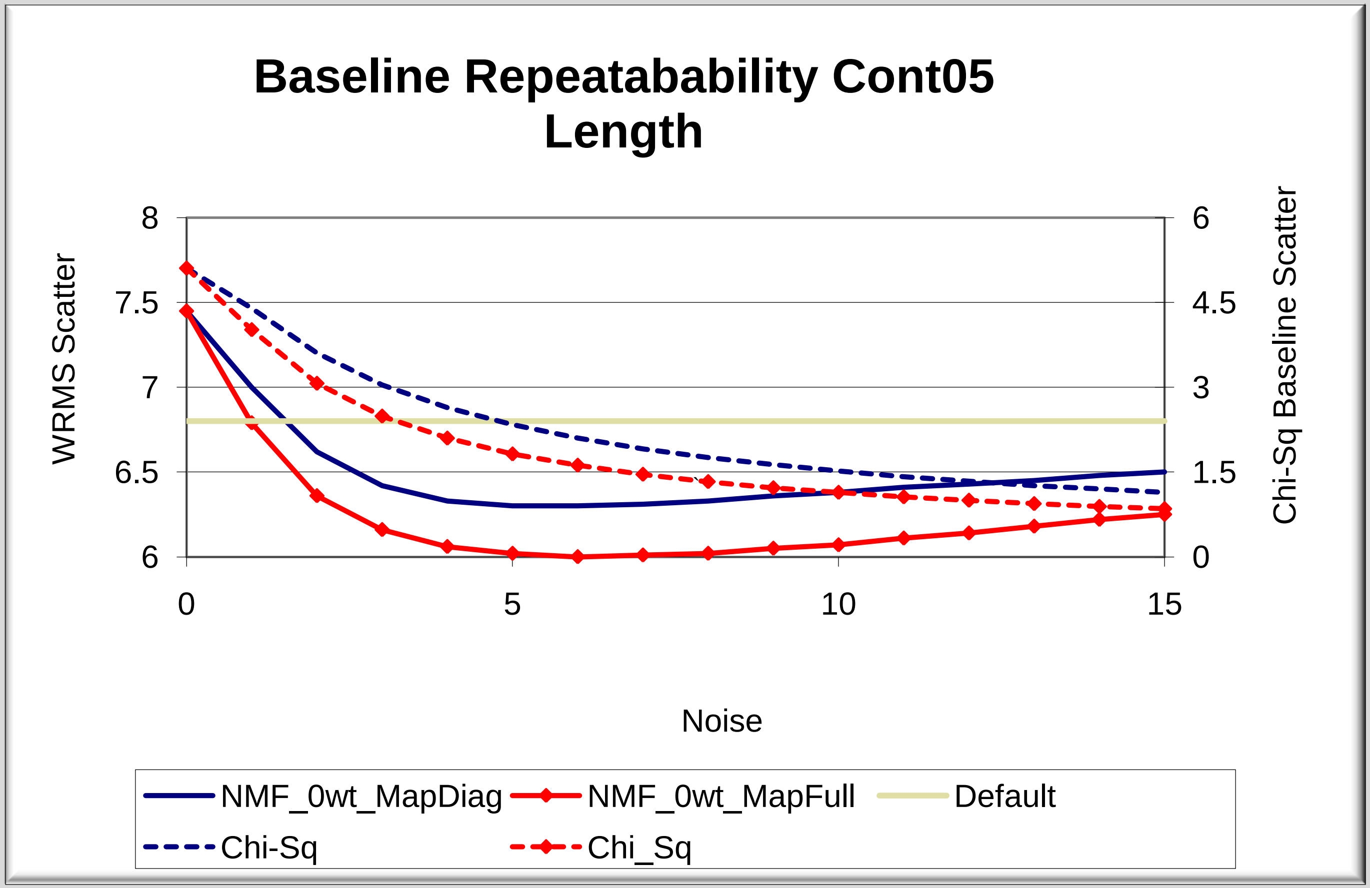
$$\left[F^T \frac{1}{Cov} F\right] A = F^T \frac{1}{Cov} O$$

Because of block diagonal structure, this becomes:

$$\left\{\sum_{scans}\left[F^T \frac{1}{Cov} F\right]\right\} A = \sum_{Scans}\left[F^T \frac{1}{Cov} O\right]$$

Instead of building up the normal equations observation-by-observation, we build them up scan- by- scan.

Effect of Adding Corr. Map Error

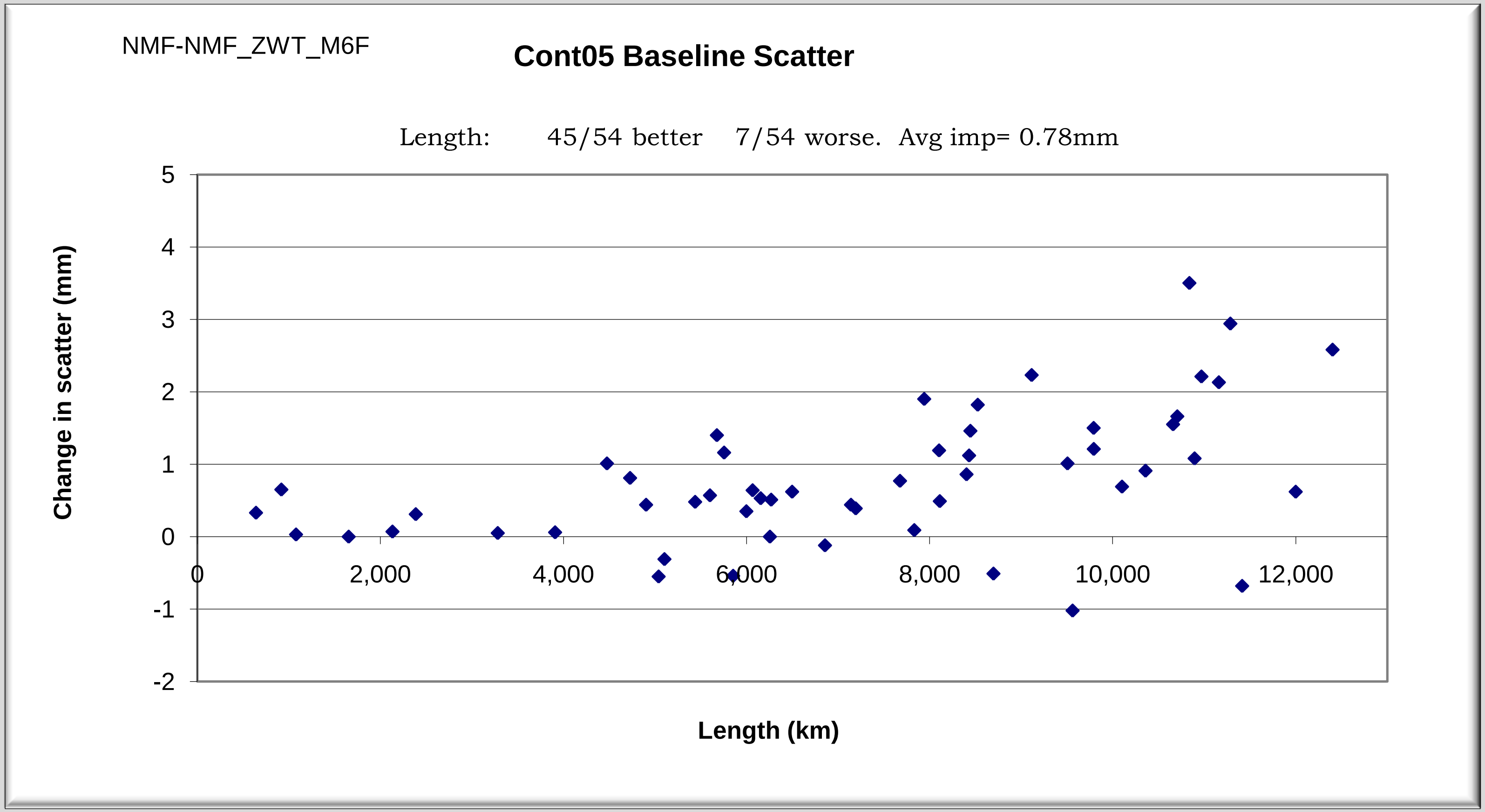


Including correlated atm-like noise results in:

1. Substantial improvement compared to standard solution.
2. Much more realistic Chi-square.
3. Added benefit: No reweighting!

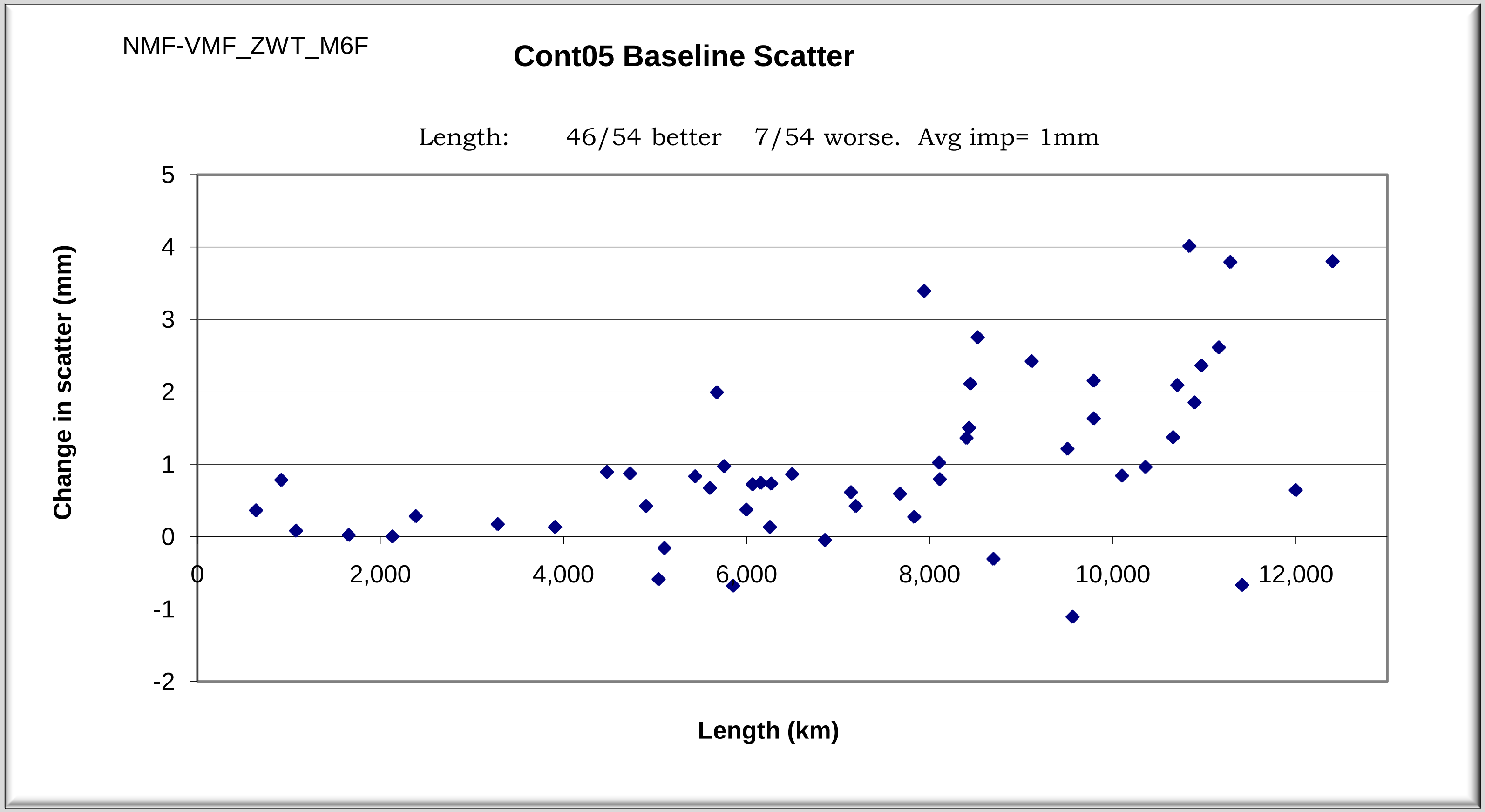
Including correlation improves the results.

Difference Plots



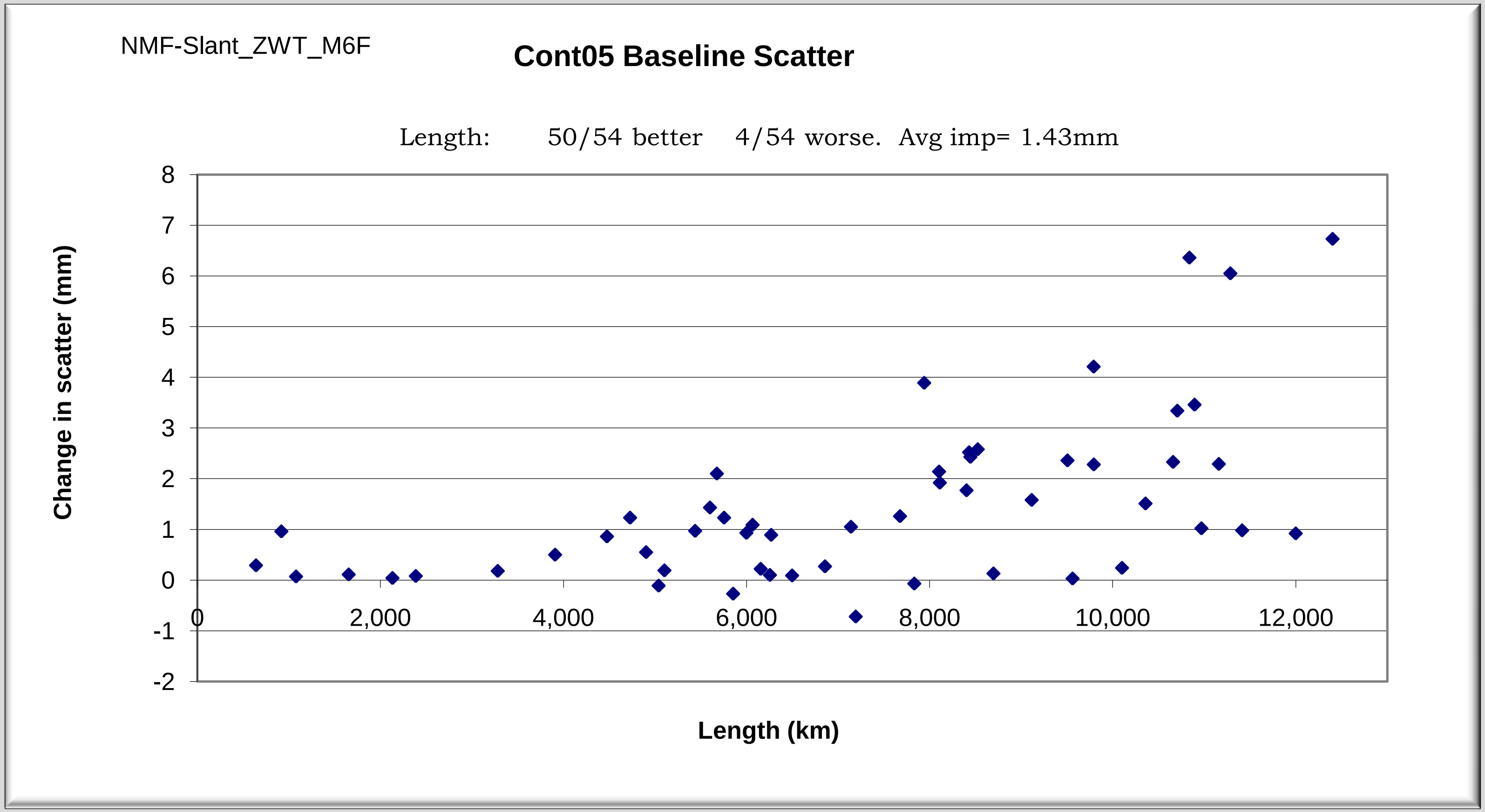
On average the scatter is reduced by 0.78 mm; 45/54 baselines are improved.

Difference Plots



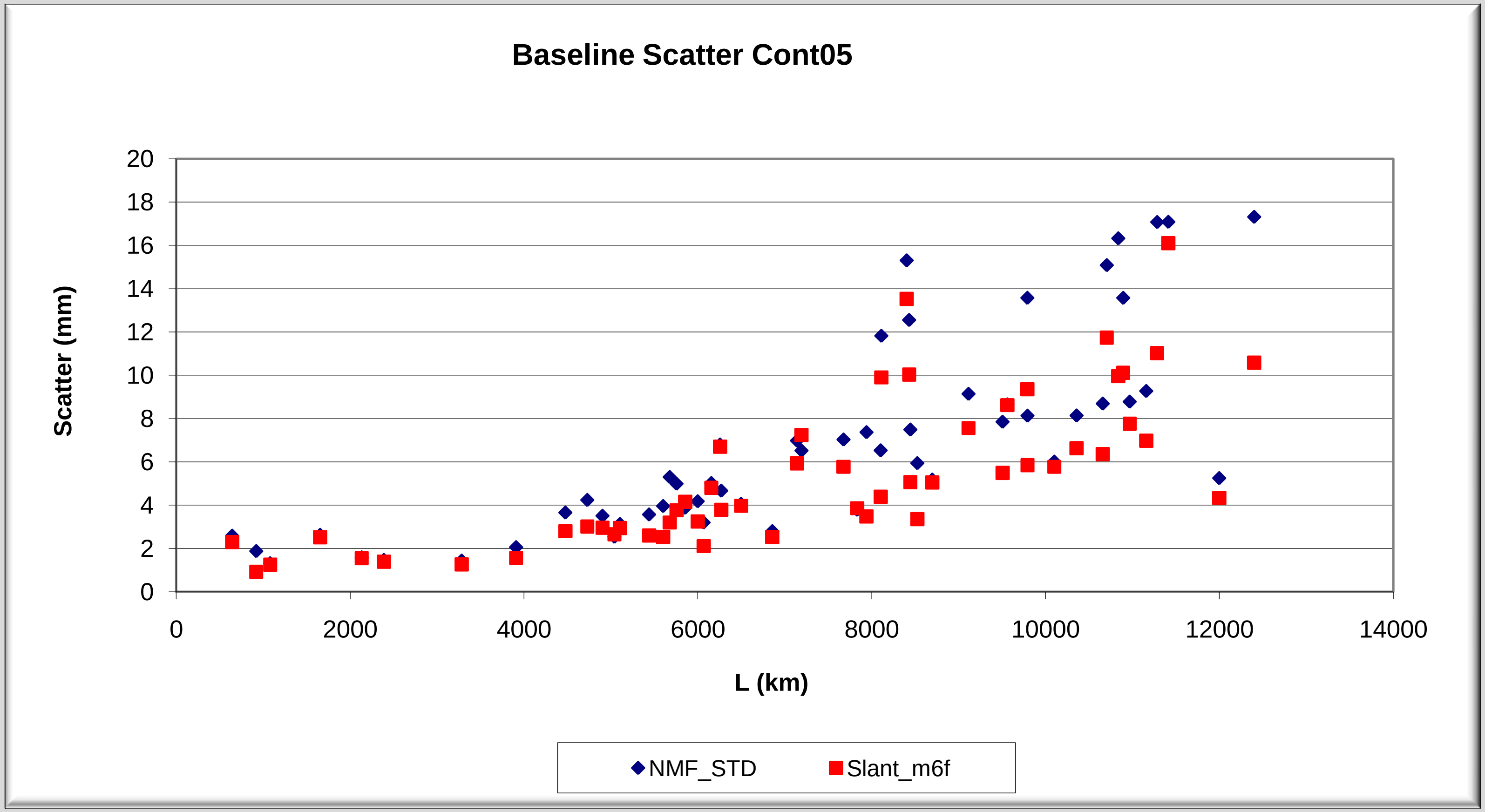
Using VMF mapping function results in further improvement.
Reduction of 1 mm versus 0.78 mm with NMF.

Difference Plots



Using ray-traced delays as a priori results in an even larger improvement. Reduction of 1.43 mm versus 1.0 mm with VMF and 0.78 with NMF.

Baseline Plot



The improvement using correlated noise is very clear in the baseline plot.

Calculation of Cov^-1

Previously we derived….

$$\langle \varepsilon_{ij}, \varepsilon_{ij} \rangle = \langle \varepsilon_{ij,meas}^2 \rangle + \langle \varepsilon_i^2 \rangle + \langle \varepsilon_j^2 \rangle$$

$$\langle \varepsilon_{ij}, \varepsilon_{li} \rangle = \langle \varepsilon_{ji}, \varepsilon_{il} \rangle = -\langle \varepsilon_i^2 \rangle$$

$$\langle \varepsilon_{ij}, \varepsilon_{il} \rangle = \langle \varepsilon_{ji}, \varepsilon_{li} \rangle = \langle \varepsilon_i^2 \rangle$$

All other elements are 0.

Using these you can show that

$$Cov = \sigma^2 + \sum_j V_j V_j^T$$

There is one V-vector for each station, and the sum is over stations. The dimension of V is the number of observations in the scan.

Covariance in absence of noise.

The elements of V are labeled by the stations in the observation. They are non-zero if and only if the observation contains station j:

$$V_{kl}^j = \varepsilon_j \times \{\delta_{jk} - \delta_{kl}\}$$

Calculation of Cov^-1

$$V_{kl}^{j} = \varepsilon_j \times \left\{ \delta_{jk} - \delta_{kl} \right\}$$

From previous page.

Example: 4 station scan which has 6 scans.
Then there are 4 V vectors and they take the form:

$$\begin{array}{lcccccc}
 & 12 & 13 & 14 & 23 & 24 & 34 \\
V_1 = & (\varepsilon_1 & \varepsilon_1 & \varepsilon_1 & 0 & 0 & 0) \\
V_2 = & (-\varepsilon_2 & 0 & 0 & \varepsilon_2 & \varepsilon_2 & 0) \\
V_3 = & (0 & -\varepsilon_3 & 0 & -\varepsilon_3 & 0 & \varepsilon_3) \\
V_4 = & (0 & 0 & -\varepsilon_4 & 0 & -\varepsilon_4 & -\varepsilon_4)
\end{array}$$

An Aside: Cov is positive definite.

Previously we showed…

$$Cov = \sigma^2 + \sum_j V_j V_j^T$$

Claim this is positive definite. $\|Cov\| > 0 \Leftrightarrow X^T Cov X > 0$

Proof: $$X^T Cov X = X^T \sigma^2 X + \sum_j X^T V_j V_j^T X = X^T \sigma^2 X + \sum_j \left[V_j^T X\right]^2$$

This means it is invertible.

Calculation of Cov^-1

It is straightforward to calculate Cov, and Cov^-1.
But it is slow….

Matrix inversion goes like N^3, where N is the dimension of the matrix.

$$N = N_{BL} = N_{Stat} \times \left(N_{Stat} - 1\right)/2$$

Hence calculating Cov^-1 goes like $N_{Stat}^{6}/8$

Fortunately can speed it up using a trick.

Calculation of Cov^-1: Trick

Consider

$$Cov^{-1} = \frac{1}{\sigma^2 + \sum_j V_j V_j^T} = \frac{1}{\sigma}\left[\frac{1}{I + \sum_j \frac{V_j}{\sigma}\frac{V_j^T}{\sigma}}\right]\frac{1}{\sigma}$$

This is just a symbolic notation.

Calculation of Cov^-1: Trick

Consider

$$Cov^{-1} = \frac{1}{\sigma^2 + \sum_j V_j V_j^T} = \frac{1}{\sigma}\left[\frac{1}{I + \sum_j \frac{V_j}{\sigma}\frac{V_j^T}{\sigma}}\right]\frac{1}{\sigma}$$

We can expand the term in […] in a power series:

$$\left[\frac{1}{I + \sum_j \frac{V_j}{\sigma}\frac{V_j^T}{\sigma}}\right] = I - \sum_j \frac{V_j}{\sigma}\frac{V_j^T}{\sigma} + \sum_j \frac{V_j}{\sigma}\frac{V_j^T}{\sigma}\sum_k \frac{V_k}{\sigma}\frac{V_k^T}{\sigma} - \sum_j \frac{V_j}{\sigma}\frac{V_j^T}{\sigma}\sum_k \frac{V_k}{\sigma}\frac{V_k^T}{\sigma}\sum_l \frac{V_l}{\sigma}\frac{V_l^T}{\sigma}\dots$$

Strictly speaking, need to worry about convergence of the series, etc.
But I am a physicist, not a mathematician. When I am done I will make sure everything works.

Calculation of Cov^-1: Trick

Consider

$$Cov^{-1} = \frac{1}{\sigma^2 + \sum_j V_j V_j^T} = \frac{1}{\sigma}\left[\frac{1}{I + \sum_j \frac{V_j}{\sigma}\frac{V_j^T}{\sigma}}\right]\frac{1}{\sigma}$$

We can expand the term in […] in a power series:

$$\left[\frac{1}{I + \sum_j \frac{V_j}{\sigma}\frac{V_j^T}{\sigma}}\right] = I - \sum_j \frac{V_j}{\sigma}\frac{V_j^T}{\sigma} + \sum_j \frac{V_j}{\sigma}\underbrace{\frac{V_j^T}{\sigma}\sum_k \frac{V_k}{\sigma}}\frac{V_k^T}{\sigma} - \sum_j \frac{V_j}{\sigma}\frac{V_j^T}{\sigma}\sum_k \frac{V_k}{\sigma}\frac{V_k^T}{\sigma}\sum_l \frac{V_l}{\sigma}\frac{V_l^T}{\sigma}\ldots$$

Contract these vectors to obtain

$$\frac{V_j^T}{\sigma} \bullet \frac{V_k}{\sigma} = W_{jk}$$

Calculation of Cov^-1: Trick

$$\left[\frac{1}{I+\sum_j \frac{V_j}{\sigma}\frac{V_j^T}{\sigma}}\right] = I - \underbrace{\sum_j \frac{V_j}{\sigma}\frac{V_j^T}{\sigma} + \sum_{jk}\frac{V_j}{\sigma}W_{jk}\frac{V_k^T}{\sigma} - \sum_{jkl}\frac{V_j}{\sigma}W_{jk}W_{kl}\frac{V_l^T}{\sigma} + \ldots}$$

These terms can be gathered together to give:

$$\left[\frac{1}{I+\sum_j \frac{V_j}{\sigma}\frac{V_j^T}{\sigma}}\right] = I - \sum_{jk}\frac{V_j}{\sigma}\left[I+W\right]^{-1}_{jk}\frac{V_k^T}{\sigma}$$

Calculation of Cov^-1

Putting it all together

$$Cov^{-1} = \frac{1}{\sigma^2} - \sum_{jk} \frac{V_j}{\sigma^2} [I + W]^{-1}_{jk} \frac{V_k^T}{\sigma^2}$$

Where W is a matrix with elements $\frac{V_j^T}{\sigma} \bullet \frac{V_k}{\sigma} = W_{jk}$

Note that $I+W$ has dimension $Nstat$.

Naive time: $N_{Stat}^6 / 8$

Clever time: N_{Stat}^3

Speeded up by a factor of: N_{Stat}^3

One Last Trick….

$$Cov^{-1} = \frac{1}{\sigma^2} - \sum_{jk} \frac{V_j}{\sigma^2} \left[I + W\right]^{-1}_{jk} \frac{V_k^T}{\sigma^2}$$

In calculating the normal equations…

$$\left\{ \sum_{scans} \left[F^T \frac{1}{Cov} F \right] \right\} A = \sum_{Scans} \left[F^T \frac{1}{Cov} O \right]$$

You encounter terms like:

$$F^T \sum_{jk} \frac{V_j}{\sigma^2} \left[I + W\right]^{-1}_{jk} \frac{V_k^T}{\sigma^2} F$$

Resist the temptation to expand the sum first! It is much better to do the calculation like this…

$$\sum_{jk} \left(\frac{V_j^T F}{\sigma^2} \right)^T \left[I + W\right]^{-1}_{jk} \left(\frac{V_k^T F}{\sigma^2} \right)$$

V has many 0's elements, which makes the VF multiplication fast.

Review & Conclusions

1. Correlation modifies the normal equations:

$$\chi^2 = \left(O^T - A^T F^T\right)\frac{1}{Cov}\left(O - FA\right)$$

$$\left[F^T \frac{1}{Cov} F\right] A = F^T \frac{1}{Cov} O$$

2. Evidence of problems with VLBI Analysis.

- Chi-square >1 for individual sessions.
- Baseline scatter larger than it should be.
- Disagreement between independent measurements too large.

3. Baseline scatter plots as a tool.

- VMF better than NMF because it reduces scatter by 0.45mm on average.

Review & Conclusions

4. Evidence for Station Dependent Noise and some sources of it.

5. Adding Station Dependent Noise helps.
   - Clock-like helps a little.
   - Mapping function helps a lot.
   - Baseline scatter for CONT05 reduced by 0.48mm on average.
   - Some baselines reduced as much 2.2mm.

6. Evidence for correlation between observations.

Review & Conclusions

7. Simplifying assumptions:
   - Station and Scan Dependent.

8. Incorporating correlation in analysis helps (a lot!)
   - RMS scatter reduced by 0.78mm. Some baselines 3.6mm
   - With VMF RMS reduced by 1.0mm avg. Max 4.0mm
   - With raytrace RMS reduced by 1.43mm. Max 6.8mm

9. Some steps to speed construction of normal equations.

Questions & Comments

?