IVS Memorandum 2006-011v01

16 December 2005

"A Monte Carlo Simulator for Geodetic VLBI"

Bill Petrachenko

Subject: A Monte Carlo Simulator for Geodetic VLBI
From: Bill Petrachenko
Date: Dec. 8, 2005
Last Revised: Dec. 16, 2005

1. Why do we need a better simulator?

WG3 (VLBI2010) was an exciting time for many of us because it gave our imaginations free rein to consider cost-effective scenarios of how geodetic VLBI should evolve into the future. However, it was also a frustrating time. We simply did not have the tools to be able to come up with hard numbers for many basic parameters, e.g.:

- How many stations are required?
- What are some realistic options for effective networks?
- Is there a benefit to having multiple antennas at each site?
- How much CRF is enough CRF?
- How effective are new observing strategies?
- How effective are new schedules?
- How effective are new analysis strategies?
- How many scans are required per day?
- What slew rates are required of the antennas?
- What is the minimum flux that must be detected?
- What is the maximum allowable integration time per scan?
- What accuracy is required of the delay observable?
- How stable do the h-maser oscillators need to be?

…and the list goes on. Most frustrating and basic to me was our inability to understand in any detail the error budget and exactly how it effects our final results - most notably, the interaction of observing scenario, observable accuracy, atmosphere estimation, and reference frequency and timing characteristics. [Note: To be fair, in some cases, we did come up with numbers, e.g. accuracy of the delay observable, and performance of the h-masers, but at the same time, we realized that these numbers included many caveats and could change greatly depending on details of the observing scenario.]

I think it is important to look at the tools we had at our disposal at the time and to ask ourselves why these tools were not good enough to answer these pressing questions.

- **Analytic Studies**. Although analytic studies can be useful to gain insight into simple questions (e.g. determining how a pcal error will bias the delay observable), the interactions are far too complex for them to deal with the follow on question of how the bias will affect the accuracy of our final products.

- **R&D Experiments**. R&D experiments are an interesting option, since they allow us to test hypotheses. However they have limitations:
    - They use up a lot of time and resources. In a majority of cases, the most reliable measure of comparative performance is parameter repeatability. This implies that each hypothesis that needs to be tested requires several sessions to build up the

required statistics. Years of R&D experiments may be necessary to test competing hypotheses.
  - Only network and hardware configurations that exist today can be tested. This makes extrapolation into unobserved parameter space unreliable. If, for example, the impact of an antenna with much higher slew rates needs to be tested, this would require the actual construction of prototype antennas, and is unlikely to be funded unless firm evidence (probably from a reliable simulator) were available.
  - It is difficult to perform truly controlled experiments where only isolated parameters are varied. In real data, there are always uncontrolled factors that confuse conclusions. For example, when testing to see if the way we handle the atmosphere can cause an annual baseline term, other processes that have incompletely understood annual terms can confuse our conclusions.
  - Baseline length repeatability is our best figure of merit (since it is independent of orientation error). This means that performance with respect to other parameters needs to be determined indirectly.
  - The true answer is never known. Repeatability is our best indicator of performance and it bears no information about bias.

- **Re-analysis or interpretation of previous campaigns or global data sets**. This suffers from most of the problems of the R&D experiments except that it is the by-product of past observations. Hence, most of the time and resources required have already been expended. However, it suffers more strongly from the limitation that only networks, scenarios, etc that have already been observed can be evaluated.

- **Inter-technique co-location studies.** This has most of the limitations of the previous two options except that it has the advantage that it allows us to look at bias (although we can't tell which technique (or possibly both) is biased). However, it requires co-located sites with strong ties, and is limited to the length of time that both techniques have been observing, and once again, to networks, scenarios, etc that are already in existence.

- **Simulations**. Simulations seem like a very cost/time/resource-effective way to over-come the limitations of the previous options. However, the simulators that currently exist or are planned in the near term are all of the covariance type. Although these simulators can be useful for some applications, it is clear that, for a large fraction of the more interesting of today's questions, covariance simulators are inadequate and, in fact, may lead to misleading conclusions. In the simplest mode, covariance simulators generate the equivalent of formal parameter uncertainties based purely on estimates of the "thermal" errors generated in the fringing process. Unfortunately, this totally ignores clock, atmosphere and modelling errors, along with potential correlation of the observables. The best we can do at the present time to handle these deficiencies, is to use the crude expedient of a quadratic additive noise term. The size of the term is based on experience gained from real observations. Unfortunately, in practice, its size varies greatly from session to session, and is known to be dependent (in only partially understood ways) on network size, observation precision, etc. In general, covariance simulators produce reasonably reliable results when the scenarios (network, stations, schedule, etc) are fairly close to something already in use. This is because a realistic value for the additive noise term is available. However, when pushing the envelope to larger networks, faster antennas, higher sensitivity, better clocks, new atmosphere models, creative observing scenarios, etc there is little basis for trusting their results.

This is not to say that the above options don't have their place. But that is just the point. They are already in place and can be used whenever needed. However, that is not true for a more realistic simulator. I do believe we are strongly limited by its absence. In fact, it is my belief that correcting that omission is the single action that will have most impact on improving the future performance of geodetic VLBI. If a well-tested realistic simulator had been in existence three years ago when we started the VLBI2010 process, the final report would have been very different in nature. Instead of making some vague recommendations and suggestions for future studies, the report could have laid out a detailed road map, including specs, schedules and costs, for how to proceed into the future. If we had realized the omission three years ago and, at that time, placed a priority on building such a simulator, one would be in existence today and we could start immediately into answering many of the questions. As it is, Oleg Titoc is asking us for antenna specs for Australia and we don't have the answers or any reliable way to get those answers.

Clearly, if we are to be taken seriously when proposing a major upgrade (or perhaps a new instrument) to our funding agencies, we will have to have **believable** numbers backing up our proposal. A more realistic simulator will be a major step in that direction.

2. What is a Monte Carlo Simulator?

Conceptually, a Monte Carlo Simulator is simple. In our case, it would involve the following steps:

- Schedule an observation.
- Generate fake data.
- Analyse the fake data using our existing analysis software.
- Repeat steps 2 and 3 (i.e. generate fake data with different random number seeds and analyse it) several times to generate a sample of results for statistical analysis.
- Analyse statistically the output parameters of the sample (e.g. to get the mean and variance of the estimated parameters, the break-down of post-fit residuals, etc) and compare the results to the input.
- Display the results in a meaningful way.

3. Benefits of a Monte Carlo Simulator

A Monte Carlo Simulator has many benefits:

- It automatically handles difficult to analyse situations like the interaction of the clocks, geometry and atmosphere.
- If designed properly, it can handle situations very different from existing scenarios, e.g. multiple antennas at each station, many more stations, antennas with much faster slew rates, etc.
- Controlled experiments can be run in which only one parameter is varied.
- The inputs are known so biases can be evaluated in addition to the usual repeatability.

- Inputs are known, so repeatability and bias of **all** parameters (station coordinates, EOPS, clocks, atmosphere, etc) can be reliably and meaningfully determined and not just baseline length repeatability.
- The breakdown of the input data is known so the contributions of each component (atmosphere, clocks, random) to the post-fit residuals can be determined.

4. Limitations of a Monte Carlo Simulator

A Monte Carlo Simulator is limited primarily by the quality of the models used to generate the fake data. Also, when evaluating more unusual scenarios, it will be limited by its built-in flexibility.

5. A Plan for Implementing a Monte Carlo Simulator

The main new element for a Monte Carlo Simulator is the fake data generator – although, scheduling and analysis software for more unique scenarios would also need to be altered too.

I would propose that the fake data generator be constructed in the most general, modular and open manner possible. For example, it could be separated into the following modules:

- Raw data generator (possibly including instrumental biases along with thermal uncertainty)
- Source effects (e.g. effects of structure on delay, effects of frequency dependent structure, instability of effective brightness centre)
- Geometry (using Calc, Occam, etc)
- Atmosphere (using ray tracing, anisotropy, physical models, etc)
- Clocks
- Probably more…..

The important thing is to start by defining the modules and interfaces. At the outset, simple models for each module could be installed. If the interfaces are well defined, then experts can provide more refined models as time progresses. An ability to which model is used (simple or sophisticated, Calc or Occam) would also be beneficial.