

# Optimal scheduling of geodetic VLBI observations

Submitted in response to Research Opportunities in Space and Earth Sciences (ROSES) — 2018 Solicitation: NNH18ZDA001N-ESI, program A. 24. Earth Surface and Interior, section 2.2

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Previous work</b>	<b>1</b>
<b>3</b>	<b>Known deficiencies of the current approach</b>	<b>2</b>
<b>4</b>	<b>Goals of the project</b>	<b>4</b>
<b>5</b>	<b>Proposed methodologies</b>	<b>4</b>
5.1	Computation of the refractivity field from the GMAO 7km nature run . . . . .	4
5.2	Optimal estimation of atmospheric parameters from VLBI observations . . . . .	6
5.3	Development of a non-sequential VLBI scheduling algorithms . . . . .	10
5.4	Development of a VLBI observation simulator . . . . .	11
<b>6</b>	<b>Relevance to the program elements</b>	<b>12</b>
<b>7</b>	<b>Previous work and risk assessment</b>	<b>12</b>
<b>8</b>	<b>Anticipated impact</b>	<b>13</b>
<b>9</b>	<b>Deliverables</b>	<b>13</b>
<b>10</b>	<b>Management plan and Milestones</b>	<b>14</b>
10.1	Data sharing plan . . . . .	15

## 1 Introduction

Geodetic VLBI is an integral part of the space geodesy technique. NASA was actively involved in development of VLBI technique since mid 70s and is internationally recognized as a leader in this field. NASA made significant investments in development of VLBI hardware in the frame of development of the VLBI Global Observing System in last years and funded construction of several VLBI stations of next generation. **The focus of our proposal is to study how the VLBI observing strategy should be modified in order to maximize return from this long-term investment.**

When we plan VLBI observations, typically conducted either as 1-hour long or 24-hour long observing sessions, we have a freedom to select which extragalactic radio sources to observe, how long to integrate, and in which sequence to observe them. A VLBI schedule is a table that specifies for each station the source name, recording start time, and recording stop time. The freedom to select the sequence of target sources can be exploited for optimization, depending on the goal of the VLBI observing session. Considering that a list of 300 sources can be observed at a given observing session and a VLBI experiment may have up to 1000 observations, the total number of possible schedules can reach  $300^{1000}$ . Although not all listed 300 sources are above the physical horizon at a given time, and taking into account other constraints, the total number of admissible schedules is considerably less, but still so huge that selection of the best schedule is serious challenge. In our view, this problem did not receive a sufficient attention in the past. Several factors prompted us to revisit the problem now. First, development of VGOS antennas and hardware comes to the end, and the system will become operational in mid-term range, within 1–2 years. VGOS antennas are fast slewing what makes possible to explore scheduling strategies that were not feasible with slowly slewing antennas. Third, the output of high resolution numerical weather models became available. Path delays through the neutral moist atmosphere computed from the output of these models can be used for modeling realistic noise for right hand-sides of VLBI simulators that is a spatially and time correlated non-stationary process. Fourth, astrometric surveys reached completeness at the 150 mJy level, i.e., now we know all the sources that can be potentially used for geodetic VLBI observations. Fifth, availability of multi-processor high performance computers makes it feasible to run large scales simulations.

Here we formulate the problem. A time range (typically 1 to 24 hours) is allotted at a given network of VLBI stations. Stations execute a sequence of command: slew to a given source, record the signal, and then slew to a next source. We aim to develop an algorithm that generates such a sequence of commands, hereafter called VLBI schedule, that provides the extremum to the selected quality functional and satisfies specified constraints imposed on antenna motion. The search of extremum will be performed with the use of the measurement noise model that we aim to make as realistic as possible. The measurement noise model will consider the contribution of thermal noise of group delay determination; and atmospheric path delay fluctuations that are temporarily and spatially correlated.

## 2 Previous work

At the moment, there are five software packages that are used for scheduling VLBI observations:

- **sked**. It is used since 80s and is supported by the Goddard VLBI geodetic group.

- **sched.** Developed and supported by NRAO. Widely used for scheduling VLBI imaging experiments.
- **views-sched.** Developed by Vienna VLBI group. Used approaches similar to sked, but written using modern software development approaches.
- **mvievs.** Developed by M. Schratner.
- **sur\_sked.** Developed by L. Petrov and optimized for scheduling astrometry surveys but also usable for geodetic applications.

### 3 Known deficiencies of the current approach

We both developed independent VLBI scheduling software from scratches. They were used for scheduling a number of VLBI experiments. Results of data analysis of the experiments scheduled that way were considered good. But not very good and not excellent. All software packages, including our MVIEVS and SUR\_SKED use different variants of the same scheduling algorithm: let us assume  $N$  scans have been scheduled. The algorithm checks the pool of sources and generates the list of admissible sources, i.e. those that are currently above the physical horizon and were not recently observed. The algorithm checks each admissible source and assigns a score depending on some criteria. The source with the highest score will be added to the schedule and the procedure will be repeated. The above mentioned software packages differ mainly in algorithms for score computations and they support more than such algorithm. Some algorithms perform a simplified simulation on the fly and assign higher weights to that next source that minimized diagonal element(s) of the covariance matrix of simulated observations assuming the right hand-side of the equations of conditions is a zero-mean Gaussian noise with *a diagonal covariance matrix*.

We call this approach sequential. The advantage of this approach is that it appeals to intuition and is easy to implement. Originally, VLBI scheduling software SKED was designed for manual interactive work and a user selected the next source manually. The scoring scheme provided a guidance which next source is better. The ability to generate the schedule without human intervention was added later. The disadvantage of this approach is that it does not generate an optimal schedule. Even if at each step of the sequential algorithm we select a source that provides a maximum reduction of the designated diagonal element of the covariance matrix (for instance, UT1 uncertainty), in general, there exists another schedule that will provide a diagonal element(s) the covariance matrix of simulated observations that will be less. The reason is that the sequential algorithm looks only one step ahead. It can be proven that an algorithm that looks ahead of  $k$  steps will provide a schedule with the diagonal element(s) the covariance matrix that is less than the algorithm that looks ahead of  $k - 1$  steps. **We aim to overcome this deficiency and develop a non-sequential scheduling algorithm.**

The second deficiency of the current scheduling algorithms is that during schedule optimization the on-the-fly simulation considers the noise as an uncorrelated Gaussian process. We know that the dominating source of errors in geodetic observations, the path delay in the neutral atmosphere, has time variable temporal and spatial correlation. In the absence of correlation, the standard deviation of the mean value of  $n$  elements of a sample of Gaussian noise with the second moment  $\sigma_0^2$ , is  $\sqrt{1/n}\sigma_0$ , i.e. it is approaching to 0 when  $n \rightarrow \infty$ . If the noise has correlation

matrix with all non-diagonal elements equal to  $\rho \in (0, 1)$ , than the standard deviation of the mean value of  $n$  elements is  $\sqrt{\rho + 1/n} \sigma_0$ , i.e. it converges to  $\sqrt{\rho} \sigma_0 > 0$  when  $n \rightarrow \infty$ . Both cases are extreme and are not encountered in nature and are shown here for illustrative purpose only, but any realistic correlations causes a deviation of scaling the uncertainties of LSQ estimation of  $\sqrt{1/n}$  rule. The schedule optimized for uncorrelated noise is not optimized for correlated noise. **We aim to overcome this deficiency by a rigorous modeling temporal and spatial correlations in the noise contributed by the neutral atmosphere.**

The third deficiency of the current scheduling algorithms is that they use an over-simplified approach for estimation parameters related to the atmosphere path delay. This problem extends beyond just scheduling. The current paradigm of data analysis of GPS and VLBI observations is to estimate zenith path delay and (optionally) the tilt of the axis symmetry of the zenith path delay, also known as atmosphere gradients, at each station. These parameters are usually treated as a function of time, f.e. a B-spline of the 1st degree or as a stochastic variables. But this approach is based on the assumption that path delay  $\tau_a$  at a given moment of time as a function of elevation  $E$  and azimuth  $A$  is symmetric and can be characterized either by one parameter (path delay in zenith direction) or by three (path delay at direction of the axis of symmetry and two small tilt angles). Thus, making several measurements at a given stations and given  $E$  and  $A$  over interval of time  $\Delta t$ , we can represent the path delay at that station as a function  $E$  and  $A$ . The turbulence in the atmosphere breaks this assumption. Assuming frozen flow turbulence  $\tau$ , temporal and spatial variations are connected via average wind velocity  $v$  that is typically around 10 m/s. That means that at short time intervals  $\tau_a(E, A)$  cannot be represented by three parameters. An area of the atmosphere at elevations  $E > 10^\circ$  above the horizon with a characteristic height  $H = 3000$  m can be considered as uniform at time scales  $2H/(v \sin E) \approx 3500$  s, i.e around 1 hour. Modern VGOS antennas with slewing rate  $6^\circ/s$  can sample the atmosphere for time intervals significantly shorter than 1 hour, and a model of axially symmetric local atmosphere with 3 parameters becomes not adequate. **We aim to overcome this deficiency and develop methods for data analysis that consider path delay  $\tau_a(E, A)$  at a given station axially asymmetric and incorporate these methods in the schedule generating algorithm.**

All existing VLBI software packages have many options that control schedule generation. Changing these options we can generate a number of trial schedules. Surprisingly, we do not have a convenient tool that would run simulation and compare a large number (say, over 10,000) schedules. The International VLBI Service for Geodesy and Astrometry (IVS) ran in the past a number of campaigns to assess new scheduling strategies. This approach is very expensive. Nowadays, it is customary in the aero-space industry to develop sophisticated flight and mission simulators that are used for observation planning. Some existing VLBI software packages have rudimentary capabilities for simulation. In particular, SKED can generate a simulated database, process it with Calc/Solve software and compute uncertainties of estimated parameters. However, first, it cannot be used for bulk processing of the large number of trial schedules, second, it does not feed the simulator with realistic correlated noise, and third, reports only formal uncertainties. In scientific papers we usually do not claim that a new method provided improvement because formal uncertainties have been reduced. For instance, a frequently used metric of quality VLBI results is baseline length repeatability, not formal uncertainties. We call the uncertainties “formal” because we know they are not realistic and show rather a lower level of true errors. Simulating software that reports metrics that are not regarded as realistic by the geodetic community has

a limited value. **We aim to overcome these limitations and develop a realistic simulator of VLBI observations.**

Putting all these steps together, **we aim to develop new VLBI scheduling algorithms, implement them in software, validate them using the simulator, and assess quantitatively the improvement with respect to existing VLBI scheduling packages.**

## 4 Goals of the project

Considering the deficiencies of the current approaches for planning observing strategies of geodetic VLBI observations, we formulate our goals:

1. To develop advanced methods for parameterization of asymmetric dependence of path delay on azimuth and elevation exploiting the capability of VGOS antenna slew fast from source to another.
2. To develop an advance simulator that uses realistic atmosphere-driven noise computed with the use of the output NASA GMAO Nature Run high resolution numerical weather model. The simulator should be able to process bulk simulations (over 10,000 schedules) and compute repeatability quantities.
3. To develop a new non-sequential VLBI scheduling algorithm that is optimized to provide an extremum to a selected quality functional. In particular, we will consider problems of minimizing the UT1 uncertainty, nutation daily offset uncertainty, vertical position uncertainty for selected stations, and uncertainties in coordinates of selected sources.
4. To generate a large number of trial schedules run them through simulator and assess quantitatively the improvement of readabilities of in time series of estimates of selected parameters with respect existing observations.

## 5 Proposed methodologies

We propose develop methodologies in four areas: 1) computation of the global 4D refractivity field with very high spatial temporary and spatial resolution using the output of GMAO 7km nature run that will be used for modeling path delay; 2) development optimal estimation of the parameters that characterize atmospheric path delay from real or simulated VLBI observations; 3) development of a non-sequential VLBI scheduling algorithm that provides extremum to the specified functional; 4) development of a simulator that for a given experiment runs an ensemble of trial schedules and computes statistics.

### 5.1 Computation of the refractivity field from the GMAO 7km nature run

Numerous models for spatial and temporal covariance of path delays through inhomogeneous, turbulent neutral atmosphere were proposed in literature. Any models are based on a number of assumptions and on *average* characteristics of the atmosphere. We are trying to reduce the number of assumption to the minimum and consider realistic atmosphere. According to ?, we consider the atmosphere as “locally homogeneous random medium with smoothly varying characteristics”. We formulate the following assumptions:

**Assumption 1.** *We assume that although the output of GMAO numerical model may not represent the state of the atmosphere precisely enough to compute path delay with accuracy required for data analysis of space geodesy observations, it adequately represent variability of the atmosphere, i.e. such properties as covariance, structure function, etc.*

**Assumption 2.** *We assume that the hypotheses of frozen turbulence is valid at time scales of 2 to 300 minutes.*

**Assumption 3.** *We assume that the atmosphere has a 3D turbulence at heights 0–4000 m above the ground.*

We will compute the refractivity field using the output of 3D GMAO 7km Nature Run numerical weather model. The GMAO 7km Nature Run was computed in 2014 for years 2005–2007. The model spatial resolution is  $1/16^\circ \times 1/16^\circ$ , 72 layers and time resolution 30 minutes. We will use the following six instantaneous variables provided at a native 3D grid: pressure thickness, air thickness, specific humidity, and three components of wind. Using the first three variables we compute at each grid element the density of dry and wet air,  $\rho_d$  and  $\rho_w$  respectively, using the equation of state. Then, according to ?, refractivity, defined as  $\frac{c-v}{v}$ , where  $v$  is the group speed of light in moist air and  $c$  is the speed of light in vacuum, depends on  $\rho_d$  and  $\rho_w$ , and absolute air temperature  $T$  the following way:

$$r = r_o + \frac{r_o^2}{6} \quad r_o = \left( a + \frac{b}{T} \right) \rho_d + \left( c + \frac{d}{T} \right) \rho_w, \quad (1)$$

where  $a, b, c, d$  are constants. By solving numerically the hypsometric equations for each profile of atmospheric parameters, we compute the geometric height above the geoid of each layer at a given longitude and latitude, and as a result get refractivity as a function of longitude, latitude, and geometric height. The refractivity field will be expanded in a tensor product of 4D B-spline basis. We will process the 7km Nature Run output for entire year 2006, in total  $48 \times 365.25 = 17532$  time epochs and store the coefficients of this 4D global refractivity field.

However, the spatial and time resolution of the the 7km Nature Run is not sufficient for our purposes. Although the GMAO model has integration time step 5 minutes, the output data rate is restricted to 30 minutes due to logistical reasons. We will compute the refractivity fields for epochs intermediate epochs with 600 sec step invoking the frozen turbulence hypothesis:  $\Delta r(x, t + \tau) = \Delta r(x - v_x \tau, t)$ , where  $v_x$  is the component of wind along  $x$  axis using wind from the GMAO model output. In the framework of frozen turbulence hypothesis the refractivity at the node  $x, y, z$  is translated to node  $x + v_x \tau, y + v_y \tau, z + v_z \tau$  due to wind. The refractivity increment at node will be computed as  $r(x, y, z, t + \tau) = r(x, y, z, t) - r(x - v_x \tau, y - v_y \tau, z - v_z \tau, t)$  using the expansion coefficients. Considering typical wind velocity 10 m/s, spatial grid with resolution 7 km will approximately correspond to time grid with resolution 10 minutes.

The refractivity fields with 10 minute time step will be derived from the original refractivity field with 30 minute steps. Using these refractivity fields, we will compute slant path delay for 71 operating and planned VLBI sites. We will compute slant path delay on azimuth-elevation grid with a step of 16 over azimuth and 24 over elevation in a range of  $0^\circ$  to  $360^\circ$  in azimuth and  $3^\circ$  to  $90^\circ$  over elevation. The grid will have equal resolution over azimuth, but the step over elevation

will be unequal. The grid step is selected in such a way that mapping function  $M_{\text{isa}}(E)$  defined as slant path delay at a given elevation to path delay in zenith computed according to ISA Standard atmosphere  $M_{\text{isa}}(E_i)$  forms a sequence with equal step. Using this grid, the line of site pierces atmospheric layers with approximately equal steps over longitude and latitude.

For computing slant path we will solve differential equations of wave propagation in the heterogeneous continuum media that are a solution of the variational problem governed by Fermat principle. Our method is similar to (?). The trajectory  $\eta(\xi)$  is sought as a perturbation of the straight line connecting the receiver and emitter denoted as vector  $\xi$ . Then path delay in neutral atmosphere is found by integration along the trajectory

$$\tau_{na} = \frac{1}{c} \int_0^{\xi_{\text{ta}}} \left( (1 + r(\xi, \eta)) \sqrt{1 + \left( \frac{d\eta}{d\xi} \right)^2} - 1 \right) d\xi \quad (2)$$

from the receiver to the top of the atmosphere  $\xi_{\text{ta}}$  that corresponds to a layer at height 80 km. We call the time series of these slant path delays as 600s time series.

Time resolution of 600 s and spatial resolution of 7 km are relatively high, but we would like to explore the contribution of turbulence at time scales 90s and spatial scale 1km. The GMAO 7km Nature Run does not provide that level of horizontal resolution, however at lower atmosphere, below 4000 m, it provides vertical resolution of 100–500 m. Invoking assumption 3, we can reconstruct statistical properties of fluctuations in horizontal direction by analyzing fluctuations in vertical directions. In particular, in the area of 50 km in the vicinity of each station we will compute structure function  $S(d) = \langle [\Delta r(x) - \Delta r(x + d)]^2 \rangle$ , where axis  $x$  is along the vertical direction,  $\langle \dots \rangle$  denotes the ensemble average, and  $d$  is around 1000 m.  $\Delta r(x)$  represents the deviation of refractivity wrt the average value. Using the Kolmogorov power law  $S(d) = C_n^2 d^{5/3}$ , we will scale the derived structure function  $S(d)$  to  $d=1000, 2000$ , and  $3000$  m. Then assuming that at scales of 3000 m the deviation of refractivity fluctuations from stationarity is small, we will convert the structure function to the covariance function:  $C(d) = 2(S(0) - S(d))$ . Using random number generator, we will compute refractivity at new grid nodes at 1000, 2000, and 3000 m over latitude and  $1000 \cos \phi$ ,  $2000 \cos \phi$ , and  $3000 \cos \phi$  m over longitude for layers below 4000 m above the surface. Then using the frozen turbulence hypothesis, we compute the refractivity field with time step of 90 s within 50 km of each VLBI stations. Then we will compute slant path delay using the new refractivity field with spatial resolution of 1 km and time resolution of 90 s. Noticing, that the contribution of the ray trajectory to path delay is the second order effect, we will use the ray trajectory from 600s time series of slant path delay and compute slant path delay by integrating expression 2. The procedure will provide us time series of simulated slant delay for each VLBI site with time resolution 90 s. Hereafter, we will call them 90s time series.

Both, 600 s and 90 s series will be used for simulation of VLBI observations.

## 5.2 Optimal estimation of atmospheric parameters from VLBI observations

Figure 1 demonstrates that a three-parameter estimation for the path delay in the direction of axial symmetry and the tilts of the symmetry axis at a given station is insufficient. The atmosphere turbulence in spatial domain requires more parameters to describe azimuth-elevation dependence



of path delay and turbulence in time domain makes these parameter time-dependent. To make things worse, the frozen turbulence assumption implies that finer details of spatial path delay dependence are the most time variable.

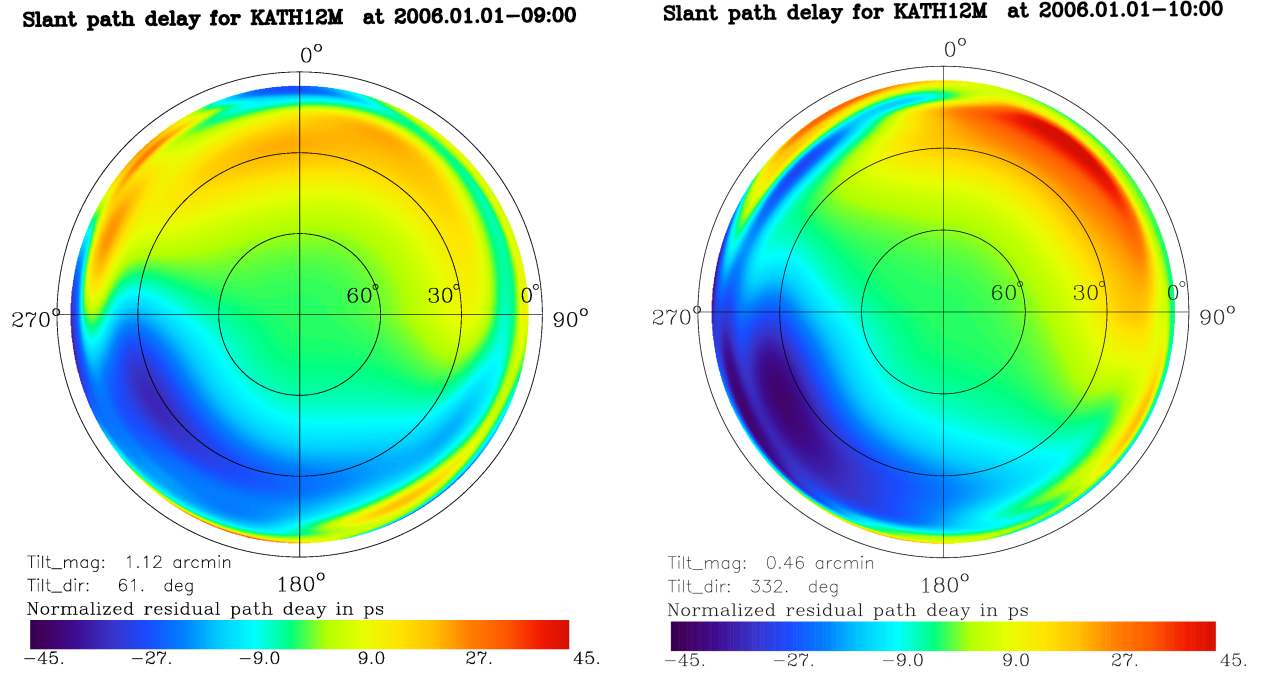


Figure 1: Azimuth-elevation diagram of normalized slant path delay residuals at the IVS station KATH12M (North Australia) after fitting a three parameter model that includes estimation of the path delay in the direction of axial symmetry and the tilt of the symmetry axis. The residuals are divided by the mapping function for normalization. The path delays were computed using the output of the GMAO 7km Nature Run.

The current state-of-art model of VLBI data analysis includes computation of the a priori slant path delays using the refractivity field of moist air derived from an output of numerical weather models. The residual path delay is estimated with a sum of B-spline of the 1st degree that models zenith path delay and B-spline that models two angles of the axis symmetry tilt. The root mean squares (rms) of post-fit residual zenith path delay range from 20 to 80 ps.

Fast slewing VLBI antennas are able to sample the atmosphere at almost any azimuths and elevation. This freedom can be used for optimization of the parametric model of residual path delay.

We propose what we call a slab parametric model for residual path delays. We assume that residual delay occurs in a thin slab that is located at height  $H$  above the station. The slab is split into  $N * N$  tiles of size  $S$  and it covers the range of elevation down to the specified minimum elevation. When the ray at elevation  $e$  pierces a slab with index  $ij$ , it acquires delay  $\tau_{ij}^s(t) M(E)$ , where  $M(E)$  is the mapping function. Path delays  $\tau_{ij}^s(t)$  are modeled with a B-spline of the 1st degree with time step  $t_s$ .

Obviously, parameters  $\tau_{ij}^s(t)$  cannot be estimated without imposing some sort of constraints, since at sufficiently short time interval  $t_s$  we merely do not have enough observations to estimate



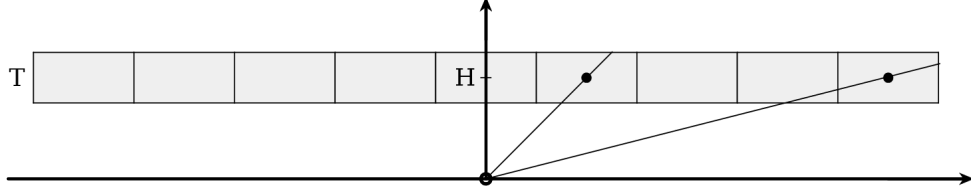


Figure 2: The diagram of slab parametric model for estimation of residual atmosphere. A projection of a 2D slab on a given direction is shown. The slab has thickness  $T$  and height  $H$ .

parameters. We will estimate parameters using least squares (LSQ) with full a priori variance-covariance matrix that describes both spatial and temporal correlations. We will compute the estimate of the variance-covariance matrix from 600s time series of the refractivity field. We will map the grid of tiles of the slab at height  $H$  to the closest elements of the 600s grid and compute zenith path delay for these grid elements. Then we will shift the entire grid at  $\pm 1, \pm 2, \dots, n$  steps and form an ensemble of path delays. The variance-covariance matrix will be computed from this ensemble. We should note that computation of zenith path delay from the expansion of refractivity into the tensor product of B-splines is performed very fast. Then the estimate of vector  $\hat{x}$  is sought using LSQ as

$$\hat{x} = \left( A^\top C_a^{-1} A + B^\top C_b^{-1} B \right)^{-1} A^\top C_a^{-1} y, \quad (3)$$

where  $A$  is the observation matrix,  $C_a$  — a priori estimate of the covariance of observation vector  $y$ ,  $B$  is the matrix of constraints and  $C_b$  — the covariance matrix of constraints. It is essential that these two covariance matrices are full, i.e. have non-zero off-diagonal elements.

As Figure 3 shows, path delay may show quick variations at time scales less than 1 hour. The time step of the spline for  $\tau_{ij}^s(t)$  should be set short enough to catch these variations.

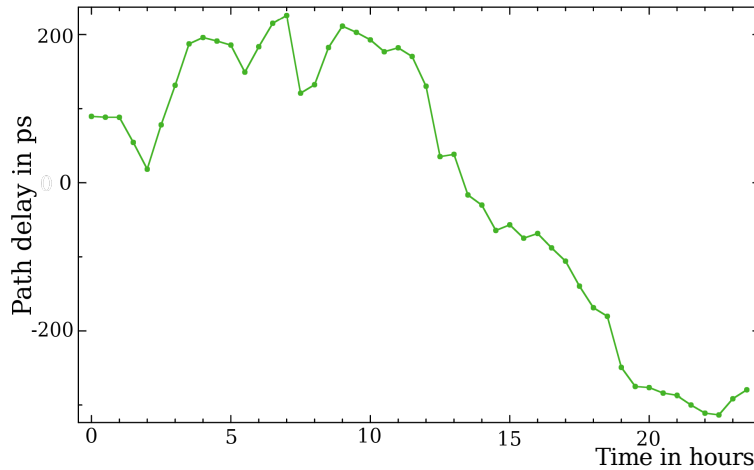


Figure 3: Variation of moist (dry and wet) path delay in zenith direction for station GGAO12M (suburb of Washington DC) with respect to the average value 8157 ps on 2006.01.01 computed using the output of GMAO 7km Nature Run.

We are going to run a number of computational experiments to investigate optimal estimation of the residual path delay in the framework of slab model. We will generate a synthetic dataset of direct pseudo-measurements of the atmosphere at a given station using 90s time series at points with random elevation and azimuths sampled with a step 1 to 2 minutes — a typical sampling rate of upcoming VGOS observations. The right hand-side of the simulated dataset will be a sum of path delay from 90s time series and the measurement noise with the rms of 10–20 ps. For processing this simulated dataset we will compute path delay using the output of MERRA2 (available from <http://pathdelay.org> — results of our previous work) that has spatial resolutions  $0.67^\circ \times 0.5^\circ$  and time resolution 6 hours. We will form residual path delays and will perform LSQ solution estimating parameters of the slab model. We will compute two statistics that are considered as metric of parameter estimation: the rms of residual path delay divided by mapping function for normalization and the rms of unnormalized residuals. For simulations we restrict elevations and azimuth to a specified area that we call “azel admissible area”. That is the area is determined by the elevation limits of a given antenna, the horizon mask and the zone of mutual visibilities with other stations of the network. Figure 4 shows the azel admissible area for IVS stations Gs, Wn, and Is. In the framework of this simulation we assume that a) a natural radio source can be found at any direction; b) antenna slewing time and integration time (i.e. scan duration) is constant.

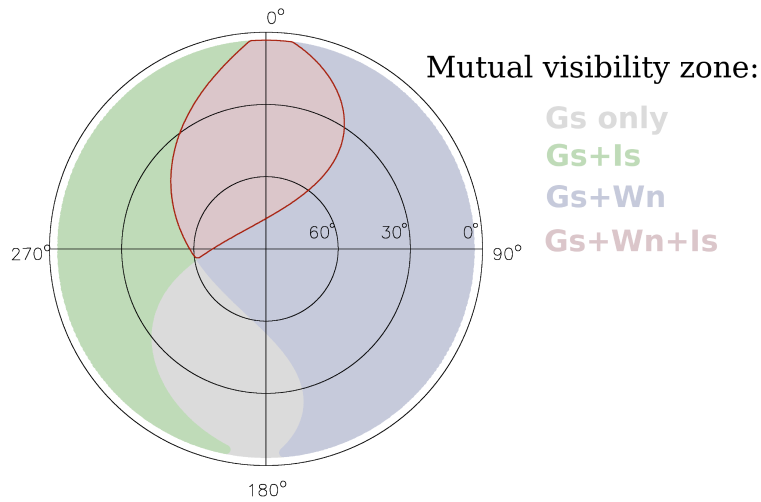


Figure 4: The azimuth-elevation diagram of mutual visibility zone at the IVS station GGAO12M, aka Gs (suburb of Washington DC) with stations Wn (Germany), and Is (Japan). Different colors show mutual visibility at different sub-networks.

In these computational experiments we will explore the impact of different variants of the parameter estimation model on the metrics:

- slab size  $S$  (initial value: 7km)
- slab height  $H$  (initial value: 3000 m)
- slab thickness  $T$  (initial value: 500 m)
- time interval of spine  $\tau_{ij}^s(t)$  (initial value: 15 minutes)

- elevation dependent weight  $\alpha M(E)$  (initial value: 0.0 )
- distribution of pseudo-observations over elevations: equally distributed, distributed as  $M(E)$  over elevations, i.e. more observations at low elevation, distributed as  $1/M(E)$  over elevations, i.e. less observations at low elevation.

The simulations will be run for 1-hour and 24-hour periods of time. We will use the brute force approach for 1 hour schedules: cast a large number of trial schedules and select the best. Our preliminary estimates show that it is feasible to run a large scale simulation of 30–50 observations acquired for 1 hour at a workstation, but it is not feasible to simulate that way a 24 hour run since computational complexity grows as the 3rd degree of the number of observations. Instead, we first generate a number of 1-hour blocks (50–200) optimized using the above mentioned approach, concatenate them into a 24-hour schedule, and test.

The goal of these simulations is to find an optimal setup for the parametric mode for estimation of residual path delay and optimal distribution of observations over elevations that reduce the metrics.

In addition to simulations, we propose to validate this scheme of data analysis using real VLBI data. In particular, we are going to use CONT14 and CONT17 VLBI campaign. we will need reduce slab size and time interval of spline for this analysis since the slewing speed of participating antennas is not as high as VGOS antennas. The goal of the validation runs will be 1) to confirm results of simulations; 2) to compare results of the slab parametric model with respect to the traditional data analysis technique (estimation of zenith path delay and tilt angles with B-splines of the 1st degree).

### 5.3 Development of a non-sequential VLBI scheduling algorithms

We propose to develop a multi-stage process for schedule generation. Our approach is to generate a schedule that is good for estimation of the residual atmosphere path delay and then modify it in such a way that it would provide an extremum to the specified functional. We favor this scheme arguing that the dominating error source is the atmosphere. Therefore, if we can minimize the contribution of the mismodeled atmospheric delay, we will get a good initial schedule that can later be perturbed to minimize the specific functional.

At the first step, we will generate pseudo-schedules for each station separately. These pseudo-schedules will be optimized using the approached discussed in the previous subsection. For a short network (say,  $d < 0.3R_{\oplus}$ ), we can compute azimuth-elevation diagrams for each station similar to that presented in Figure 4 and restrict all trial schedules for the zone of mutual visibility. For a large network ( $R_{\oplus} < d < 2R_{\oplus}$ ), the zone of mutual visibility may appear too small. To overcome this problem, the network is split into several sub-networks of smaller size not exceeding the specified parameter  $d_{\max}$ . Pseudo-schedules are generated for each sub-networks separately.

These pseudo-schedules will be gradually transformed to observable schedule with a number of steps. These steps are briefly outlined here.

- **Tying together pseudo-schedules.** The pseudo schedules were generated independently at the previous step, and therefore did not “observe” the same pseudo-sources. As the

first approximation, we find the average position of this pseudo-source,  $\alpha_{\text{avr}}, \delta_{\text{avr}}$  that is common at all stations of all sub-networks. Then we find optimal position of that source by minimizing the following functional:

$$J_w = \sum_i (A_i x_i - y_i)^\top C_{ia}^{-1} (A_i x_i - y_i), \quad (4)$$

i.e. the sum of squares of weighted rms for all the stations of the (sub)network.

- **Perturbing the schedule to optimize the target functional.** The goal of VLBI experiments can be set in terms of minimizing some functional. For instance, if the goal of the experiment is to get the best estimate of UT1, then the target functional is a diagonal element of the variance-covariance matrix of estimates that correspond to UT1. We can set a goal to get the best estimate of vertical positions of some (or all) stations. The target functional will be a sum of diagonal elements of the variance-covariance matrix of estimates of these parameters. We will run in a cycle over all sources the optimization problem: perturbation of a position of the  $k$ th source that minimizes the target functional. A sub-cycle over sub-networks will run and all sub-networks will be checked. The source that provides the highest decrease of target functional among all checked sub-networks will be selected, and the schedule will be updated. The cycle will be repeated till reduction of the functional at a given iteration will be less than a given threshold.
- **Mapping pseudo-sources to real sources.** We will replace a pseudo-sources with the closest real sources.
- **Adjusting start-stop time stamps.** At this step for each observation we will compute slewing time and adjust observation start time that in general different for different stations. We will check observation stop time that is supposed to be common for all stations. Using source images we will compute the expected correlated flux densities at each baseline for a given observation and then compute the signal to noise ratios (SNR). We will adjust integration time of a VLBI observation in order to reach the target SNR and have enough time to slew to the next source. A collection of images of all sources suitable for geodetic observations is already prepared by the PI and available at [http://astrogeo.org/vlbi\\_images](http://astrogeo.org/vlbi_images). Although observation start time of a given source may be different at different stations, the stop time will be the same. In the framework of our approach integration time will be variable.

The first part of the VLBI scheduling procedure, generating pseudo-schedules for sub-networks depends on the seed of the random number generator. Runs with different seeds will produce different schedules. Running a set of schedules for the same date and the same network, we will get a family of schedules. Examining the mean value of the target functional and its spread, will allow us to assess robustness of the scheduling algorithm for a given network.

#### 5.4 Development of a VLBI observation simulator

Although the scheduling block is a 1-hour or 24-hour session, a planning unit of VLBI observations is a campaign that consists of many sessions. An example of a campaign is 14-day long

CONT17 or one year long UT1 intensive campaign. Before running an observing campaign, we need run a simulation and investigate expected results. Typical questions that we need solve are what is the best network, what is an impact of including or excluding a given station, what is uncertainty of expected results either in terms of uncertainties or from repeatability. We propose to incorporate scheduling and analysis in one program that will call a routine for schedule generation, compute simulated right hand-sides using 90s time series, perform LSQ analysis, store results, and compute the specified metric for the entire campaign. The maximum duration of a simulated campaign is set to one year. This is sufficient for catching annual variations in precision of estimated parameters. It is known for a long time that in local winter wrms of residuals is smaller and accuracy of VLBI results is better because the contribution of atmospheric path delay is smaller. We will be able to model these seasonal effects with the proposed simulator.

We are going to use the simulator and generate schedules for past campaigns, such as CONT14 and CONT17, IVS Intensives and compare the metrics of the schedules that were used for these campaigns with the schedules generated by using our proposed approach. This will allows us to quantify the improvement.

## 6 Relevance to the program elements

This proposal address the statement of the A.24 program call, paragraph 2.2 for “*modeling, and analysis efforts that explore the tradeoffs between different data collection strategies*”, in particular, the strategy for optimization of the sequence of observed sources and integration time at geodetic VLBI ground network with the emphasis on advance modeling the contribution of the atmosphere to path delay. We also address a call “*to conduct Observing System Simulation Experiments (OSSE) that consider real and simulated observations*” by using the output of GMAO past OSSE run, transforming it in a such a way that if can be used for developing the VLBI simulator, and performing a number of OSSE runs for reducing “*errors associated with solid-Earth science questions*”.

## 7 Previous work and risk assessment

This activity is built on our past experience in running existing scheduling software SKED, SCHED, and developing from scratches our own scheduling software packages. We learned in detail subtleties of VLBI schedules, for instance constraints set by cable wrap, modeling correlated flux densities from source brightness distribution, and we know how to format the output schedule in order to make a schedule executable by VLBI stations. Our software packages were used for scheduling VLBI campaigns and results of these campaigns resulted in publications (f.e. (??)). Therefore, we consider risk of failure to generate valid schedules as very low.

We have invested heavily in development of the infrastructure for ingestion of the output of GMAO numerical weather models for computation of slant path delay in the atmosphere, computation of atmospheric mass loading loading (ESI award NNX12AQ29G, PI: L. Petrov), computation of the atmospheric angular momentum and Earth orientation prediction (ESI award NNX15AC10G, PI: L. Petrov), and processing GPS radio occultation data using GMAO model GEOS-FPIT (ESI award, NNX16AD88G, PI: L. Petrov). The first three projects resulted into services that are operating non-stop 24/7 on hourly basis: <http://massloading.net>, <http://pathdelay.net>, <http://earthrotation.net>. It took about half a day to

adapt the existing infrastructure for processing the GMAO 7km Nature Run. The results of computation of slant path delay from the GMAO 7km Nature Run are shown in Figures 1 and 3. Therefore, we consider the risk of generating 600s and 90s time series is low. We estimate required CPU time at a 64-core cluster 50–60 days, which is manageable.

Our vast previous experience in VLBI data analysis (the PI has processed all VLBI geodetic experiments since 1980 through April 2018) backed by publication records makes development of the VLBI simulator and that includes coding the slab parametric model an easy task.

The most risky part of the project is our non-sequential procedure of schedule optimization. Since the problem has complexity well over  $10^{1000}$ , it cannot be examined rigorously and we should honestly say we do not see a way to prove that a given schedule is indeed the best possible among  $10^{1000}$  others. In order to get a practical solution for a finite time, we have to resort to some sort of heuristics. We employ the heuristic that a good schedule should optimize retrieval of the atmospheric parameters that are considered to have spatial and temporal correlations. This heuristic is based on a consensus that the troposphere is the dominating sources of errors in VLBI and GNSS. Therefore, we first optimize a pseudo-schedule for optimal estimation of atmospheric parameters and then gradually apply the secondary criteria for optimization, such as uncertainties of UT1, nutation angles, vertical site positions that accounts for temporarily and spatially correlated right hand-sides.

Although we cannot quantify whether our schedule is the best we can and we will quantify that our schedule is better or worse than a reference schedule.

## 8 Anticipated impact

We expect that proposed activity will result to development of a better scheduling VLBI strategy that fully exploits new VLBI hardware that was being developed during last 5–10 years. We expect improvement in determining the Earth orientation parameters such as UT1 and nutation angles. We expect improvement in accuracy of site position determination that will contribute to the NASA goal to reach 1 mm level of position accuracy and 0.1 mm/yr for rate of changes in the terrestrial reference frame.

We expect that implementation of the slab parametric model into VLBI data analysis for estimation for the residual atmosphere path delay will be useful for improving results of analysis and making their uncertainties more realistic. ? has demonstrated that using full variance-covariance model based on a simplified model for parameter estimation made uncertainties much more realistic and reduced wrms by 9 ps in quadrature. Since our variance-covariance model is based on processing the output of numerical weather model and we propose a more refined estimation model, we expect a more sizable improvement.

The 600s and 90s time series of path delays that we will develop may be useful for other applications, for instance for simulation of GNSS observations.

## 9 Deliverables

The following will be delivered in the course of the project:

- 600s and 90s time series of slant path delay over 2006 for 71 VLBI sites.

- Detailed algorithm of the slab parametric model for estimation or residual atmospheric parameters, results of simulations with different parameters of the model, and results of validation using CONT14 and CONT17 IVS campaigns.
- Detailed non-sequential VLBI scheduling algorithm and its implementation into software.
- Simulator of VLBI campaigns based on the GMAO 7km Nature Run.
- Results of trial schedules for a one year campaigns focus on a) determination of UT1, b) determination of nutation angles; c) determination of vertical station positions.

## 10 Management plan and Milestones

The chart below shows the schedule for implementing the tasks. The schedule is arranged to give an approximately uniform deployment of effort for the team.

Table 1: Schedule chart

Activity name	PY1 H1	PY1 H2	PY2 H1	PY2 H2
Computation of 600s and 90s time series	•			
Development of the algorithm for slab parametric model and its implementation		•		
Validation of the algorithm for slab parametric model			•	
Development of the non-sequential scheduling algorithm and its implementation			•	
Development of the VLBI simulator				•
Running simulations of one-year long VLBI campaign				•
Writing papers and reports				•

The Principal Investigator, Leonid Petrov, who works for ADNET Systems Inc. will manage the project. Leonid Petrov will develop all computational procedures for processing the GMAO 7km Nature Run and for computation of 90s and 600s time series of slant path delay. He will develop the algorithm for slab parametric model and implement it in software. He will also validate this model using VLBI observations. Leonid Petrov will develop the scheduling algorithms and software, as well as the simulator, and run it.

Matthias Schartner (Co-I) who works in Vienna Technical University will be using the 90s and 600s time series computed by the ADNET. That results will be shared between the PI and Co-I. Other development will be done in parallel and to a significant extent independent. We selected this scheme of relatively loose cooperation recognizing that the Co-I is funded independently, and our ability to visit each other will be limited. We will communicate via email and telephone. The Co-I will be working on its own implementation of the slab parametric model and he will incorporate it to VLBI analysis software VieVS. Similarly, M. Schartner will be developing his



own program of VLBI scheduling software using the methodology described in the proposal. The results will be compared and checked. Those elements in this development chain where the co-I got better results will be replicated by the PI and vice versus.

The advantage of this scheme is that work will be done independently, errors can be easily caught, and a personal bias will be alleviated. The disadvantage of this approach is duplication of efforts, but for this project duplication of efforts does not result in duplication of cost. The contribution of the Co-I is essential, but not critical for the project.

No funds are requested for foreign co-investigator Matthias Schartner.

## **10.1 Data sharing plan**

The major results of this project will be accessible as electronic attachments to peer-reviewed publications. The 600s and 90s time series of slant path delays from the GMAO 7km Nature Run will be made available from the project web site. The source code of the VLBI scheduling software and VLBI simulator will be made freely available from the project web site. Examples of our past data-sharing practices can be found at <http://massloading.net>, <http://pathdelay.org>, and <http://earthrotation.net>.