GPS Meteorology: Sliding-Window Analysis*

JAMES FOSTER
SOEST, University of Hawaii at Manoa, Honolulu, Hawaii

MICHAEL BEVIS
The Ohio State University, Columbus, Ohio

STEVEN BUSINGER
SOEST, University of Hawaii at Manoa, Honolulu, Hawaii

(Manuscript received 3 April 2004, in final form 13 August 2004)

ABSTRACT

The sliding-window technique uses a moving time window to select GPS data for processing. This makes it possible to routinely incorporate the most recently collected data and generate estimates for atmospheric delay or precipitable water in (near) real time. As a consequence of the technique several estimates may be generated for each time epoch, and these multiple estimates can be used to explore and analyze the characteristics of the atmospheric estimates and the effect of the processing model and parameters. Examples of some of the analyses that can be undertaken are presented. Insights into the phenomenology of the atmospheric estimates provided by sliding-window analysis permit the fine-tuning of the GPS processing as well as the possibility of both improving the accuracy of the near-real-time estimates themselves and constraining the errors associated with them. The overlapping data windows and the multiple estimates that characterize the sliding-window method can lead to ambiguity in the meaning of many terms and expressions commonly used in GPS meteorology. In order to prevent confusion in discussions of sliding-window processing, a nomenclature is proposed that formalizes the meaning of the primary terms and defines the geometric and physical relationships between them.

1. Introduction

Networks of global positioning system (GPS) receivers are now routinely used to provide near-real-time estimates of precipitable water vapor (PWV) for use in weather models (e.g., Wolfe and Gutman 2000; Dick et al. 2001). In order for GPS-derived PWV estimates to be usefully included in a weather model the estimates need to be more accurate than the model is capable of analyzing without the GPS data, and the estimates must be available for ingest into the model during the current assimilation cycle. These criteria vary considerably, depending on the specific applications and models being implemented; however, a reasonable rule of thumb is that estimates should have accuracies better than 2 mm of PWV and be available within 1 h of the data collection (Gutman and Benjamin 2001). Whereas GPS processing has traditionally been performed in daily 24-h batches, the new need for rapid estimates has led to more flexible processing approaches, of which the most commonly used is the “sliding window” (Fang and Bock 1998; Dick et al. 2001; Ge et al. 2002). In this approach a time window (of constant width) is used to select data for geodetic processing and is stepped forward in regular increments in order to include the most recently available data from the network. In order to optimize the processing speed, one possible sliding-window approach is to include only the data collected since the previous solution. This has the advantage of providing the most current estimates more quickly than an approach that includes all the data within a broader
window. While using a wider time window to select data for processing requires more CPU time, in practice, with the relative affordability of fast computers, this does not preclude the estimates being available in time for application in numerical models (Gutman and Benjamin 2001), and it has the advantage that multiple estimates are made for every time epoch. Each of these multiple estimates is generated from a different position within the data window, providing the operator with an opportunity to examine a wide range of phenomena associated with the processing and parameter estimation. The focus of this paper is on the second sliding-window approach, in which all the data in a broader time window are processed. Abandoning the traditional daily batch processing and instead generating multiple, overlapping solutions leads to ambiguity in some commonly used GPS meteorology terminologies. Therefore, a nomenclature is proposed here to formalize the meaning of the primary terms and to define the geometric and physical relationships between them.

2. Sliding-window terminology

The sliding-window technique foregoes the traditional rigid 24-h-batch approach and instead applies a finite time window to select the dataset for processing. The window is then moved forward in time incrementally, generating a full solution each time, resulting in multiple estimates for atmospheric parameters at each epoch in time. The analysis of any given time window is similar to standard batch processing in that a single (often strongly constrained) solution is obtained for the spatial coordinates of each GPS station, whereas multiple atmospheric delay parameters are estimated for each station in order to determine how the delay varies with time within the period of consideration (e.g., Bevis et al. 1992; Duan et al. 1996). Although the general technique is conceptually simple, this approach, with its multiple solutions overlapping in time, requires a more careful definition of many terms often used loosely in GPS meteorology in order to discuss meaningfully its results and implications.

In this paper the term “solution” is reserved to refer to the full set of parameters estimated in a single processing run (the parameters obtained in a single geometric analysis of all data associated with a given window position or time interval). In general this will include the site position and satellite orbit parameters, in addition to the atmospheric delay estimates; however, as this paper is focused on GPS meteorology, “solution” will be used to denote the full set of atmospheric delay estimates from a processing run rather than employing the more cumbersome, albeit rigorous, term “atmospheric solution.” For any window of data, the number of atmospheric delays estimates, or “knots,” is determined by the time width of the window and the knot interval (i.e., the time step between successive knots, or estimates, in the solution). Typically the atmosphere is modeled by a stochastic process similar to a random walk and is fitted by either a piecewise linear or a piecewise constant function (e.g., King and Bock 2000). The former function provides one more estimate per solution for a given knot interval than the latter, as its knots begin and end at the edges of the data window rather than at the midpoints of each constant section. Rather than address each of these cases individually, in the following formalism the form for the piecewise linear approach is given and the reader is invited to make the conversion to one-fewer knots for the piecewise constant case.

When describing the sliding-window process it is possible to take either of two perspectives: time fixed or window fixed. In the former perspective one views the window moving forward in time, generating an estimate for the most current epoch and successively increasing the number of estimates for older epochs until the window passes that epoch and moves on (Fig. 1). This is the most natural perspective for an intuitive understanding of the sliding-window approach. In the window-fixed perspective the process appears as a growing sequence of solutions stretching back in time from the most current window (Fig. 2). This perspective is the most useful for programming and handling of solutions. The approach from both perspectives is presented, and the transformations required to switch between them are described.

3. Fundamental parameters

Three parameters are needed to define a window and the sliding-window process: (i) the position of the window, (ii) the width of the window, and (iii) the increment with which the window moves forward in time. As this technique is designed for near-real-time (NRT) GPS processing, it is most natural to define the window position as \( t \), the end time of the window. In an NRT processing environment this will be the most current (or nearly real time) epoch. The “width” of the window can be specified by any combination of two of three parameters: the knot interval (or knot step) \( \delta \), number of knots \( n \), and the time width \( W \) of the window. For these three parameters the following equation holds (for a piecewise-linear solution):

\[
(n - 1)\delta = W. \tag{1}
\]
The ratio of the window step ($\Delta$), the time step with which the window progresses, and the knot step is a value that appears repeatedly in this sliding-window definition. Due to its prominence this ratio is referred to as the step ratio ($S$), where

$$S = \frac{\Delta}{\delta}. \quad (2)$$

For all practical applications $S$ is expected to be a positive integer, as it makes little sense to set the window step interval smaller than the knot interval or have the knot interval not a factor of the window step interval.

By defining a reference epoch $\tau_0$ (most typically, but not necessarily, the most current epoch) a time frame is established within which epochs can be identified by their “lead” time or “lag” time relative to $\tau_0$. Epochs at an earlier time-of-day than the reference time are said to “lead” that time, while those at a later time “lag.” By defining an integer scale for lead and lag times where the integer is the number of knot intervals they are from the reference epoch, epochs can be identified with a compact and intuitive notation:

$$\tau_i = \tau_0 - i\delta, \quad (3)$$

where the $\tau_i$ is said to lead if $i$ is positive and lag if $i$ is

---

**Fig. 1.** Schematic diagram of a sliding-window process viewed from a time perspective. The window is nine knots wide and the step ratio is 1. The reference time for solution N is shown as $t$.  

**Fig. 2.** Schematic diagram of a sliding-window process viewed from a solution perspective. The window is nine knots wide and the step ratio is 2.
negative. A similar equation can be written to relate the window times to a reference time:

$$T_i = T_0 - j\Delta,$$  \hspace{1cm} (4)

Here $j$ simply represents the sequential solution number relative to the reference time. Using the epoch and window lead/lag numbers and setting $T_0 = T_p$, the epochs can now be related to knots and windows using

$$i = jS + n - k,$$  \hspace{1cm} (5)

where $k$ is the knot number, that is, the consecutive atmospheric estimate within a solution window.

The “order” of any epoch is the number of estimates ($N$) available for that epoch. For any epoch still within the active window (i.e., any epoch whose lead is less than the number of knots in the window), $N$ can be determined from the following equation, setting the most current epoch as the reference epoch:

$$N(\tau) = \text{floor} \left( \frac{i}{S} \right) + 1, \text{ for } i = 0 \rightarrow n - 1,$$  \hspace{1cm} (6)

where “floor” indicates that the resultant is rounded down to the nearest integer.

The final number of estimates, or maximum order, $N_{\text{max}}$ available for any epoch is dependent on $S$ and whether the epoch position ($\tau$) is an integer multiple of $\Delta$ from the window position ($T$):

- case 1: \( \text{mod} \left( \frac{T - \tau}{\Delta} \right) = 0; \quad N_{\text{max}} = \frac{n + 1}{S}, \)  \hspace{1cm} (7)
- case 2: \( \text{mod} \left( \frac{T - \tau}{\Delta} \right) \neq 0; \quad N_{\text{max}} = \frac{n - 1}{S}, \)

where “mod” indicates the modulo or remainder after division by 2. One possible sliding-window configuration is illustrated in Fig. 1. This could represent an 8-h piecewise-linear window with a 1-h knot interval incremented in hourly steps. As an example, the solution and knot numbers for all estimates of the highlighted epoch ($t + 1$) would be given by solving the equality given by (5): $-1 = j + 9 - k, k = [1, 2, \ldots, 9]$, resulting in a set of solution lag numbers $j = [-1, -2, \ldots, -9]$.

The view of the process for Fig. 1 is from a time perspective, which is the easiest way to visualize the sequential solutions and their temporal relationships. For the implementation of a process, however, the most useful practical perspective is from a solution viewpoint, where each successive solution is a new row on the bottom of a stack. This is illustrated by Fig. 2, which once again shows a process with a nine-knot window. In this case, however, the step ratio is set to 2, representing, for example, a knot step of half an hour and a window step of an hour or, alternatively, a knot step of 1 h and a window step of 2 h. The problem of locating all the estimates for a given epoch requires consideration of where the epoch lies relative to the window, as in (6). As an example, we can calculate the knot-window locations of all other estimates for the epoch estimated at knot 6 in the most recent solution ($N + 4$).

If we define the most current estimate as our reference time, we require the locations for all estimates of epoch number $i = 3$:

$$3 = 2j + 9 - k, k = \{1, 2, \ldots, 9\},$$  \hspace{1cm} \Rightarrow 2j = -6 + k, k = \{1, 2, \ldots, 9\}.$$

Integer solutions are only available for knots $k = \{2, 4, 6, 8\}$ with matching solution numbers $j = \{-2, -1, 0, 1\}$. Only solution numbers 0 and 1 currently exist in the stack; the final two estimates will be generated by the next two solutions.

### 4. Applications

The multiple solutions that are generated with a sliding-window process simulate Gibb’s sampling, resulting in a family of model parameters sampled from the solution space close to the optimum solution. The multiple estimates for precipitable water (or atmospheric delay) for each epoch provide the opportunity to examine a wide variety of problems related to the reproducibility of PWV estimates and the characteristics of the errors. In this section we will introduce a few examples of the types of analysis that can be undertaken with the sliding-window approach.

The sliding-window technique was used to process GPS data collected as part of the 1997 Water Vapor Intensive Operations Period (WVIOP97). Seven continuous GPS stations in Oklahoma and Kansas were running for 3 weeks during the fall of 1997 and four of these sites had collocated water vapor radiometer (WVR) instruments operating continuously and radiosonde launches (Lesht and Liljegren 1997; Lesht 1999) every 3 h (Fig. 3). With this well-constrained dataset the GPS PWV estimates can be examined in detail to investigate the effect of the processing window on the atmospheric solutions. In an operational system both the global and regional solutions would be processed using the sliding-window technique, with the global solution using data from sites around the world and a broader time window to generate precise orbits, clock models, etc. For this study, however, the standard precise orbit solutions from Scripps Orbit and Permanent Array Center (SOPAC) have been used (see http://sopac.ucsd.edu/processing/orbits.html) and the sliding-window technique has been applied for the regional

...
solution only. This approach allows the focus to remain on the phenomenology of the sliding window in a simple, specific context, without any potential confusion as to which of the two sliding-window processes might be responsible for any observed behavior in the solutions. For example, it is common to note that the delay estimates for an epoch have a trend that is a function of the knot number of the estimate. If the orbits are also being generated by a sliding-window process, it is then unclear whether the source of such a trend is purely due to the changing orbital parameters used for each solution or arises from the regional sliding-window process and is due to the interplay between the changing data window and the regional processing parameters, or is even some combination of both these potential sources. The trade-off for this is the possibility of artifacts in the solutions introduced from the potentially discontinuous orbit solutions. These effects, however, are visible in the solutions as sliding-window process steps from 1 day's orbital solutions to the next.

The GPS data were processed using Gamit (King and Bock 2000) with an 8-h window, half-hour knot step, and 1-h window step. The atmospheric delay was estimated with a piecewise-linear function, using the Niell (1996) mapping function. The hydrostatic component was calculated using Saastamoinen’s (1972) formula, and the parameter II for mapping the wet delay to PWV was from Bevis et al. (1994). Although Ross and Rosenfeld (1997, 1999) provide a more location- and season-specific set of functions for II, the differences were so small as to be insignificant for this study (less then 0.05 mm of PWV). A fiducial site from each coast of the United States, plus one from Canada and one from South America, were included in the processing to provide the long baselines needed to provide absolute PWV estimates (Duan et al. 1996). Accurate International Terrestrial Reference Frame 1997 (ITRF97) locations and velocities for all sites were used to constrain the site coordinates to their prior estimates as tightly as their confidences permitted, and the orbits were also tightly constrained. Site coordinates and orbital parameters were estimated in addition to the atmospheric parameters. The 8-h width of the sliding window was chosen because experience suggests that this is an effective compromise between processing time and solution accuracy, as there is generally little measurable improvement in the accuracy of atmospheric estimates for wider windows (see, e.g., Baker et al. 2001).

In order to investigate the precipitable water estimates in detail it is necessary to define a reference measurement against which they can be compared. Although comparisons against an absolute measurement would be ideal, there is no absolute reference for PWV data. Radiosondes and WVRs provide independent estimates of PWV that are useful for corroboration and identification of gross biases and trends, but each system has weaknesses that make them unsuitable for use as a standard for this type of study. Thus, rather than adopting an independent platform for a reference, a GPS-derived reference PWV time series was chosen for determining the performance of the new technique. A PWV time series was generated using the traditional 24-h-batch method, but with an extra 1 h of data added before and after each 24-h data file in order to minimize the window effect at the day boundaries. Figure 4 shows the scatterplot of the results plotted against the radiosonde data. Also shown are plots for the WVR results, the “nearly real time” GPS PWV estimates (i.e., the final estimates from each sliding-window solution), and the time series of medians, formed from all the estimates for each epoch from the sliding-window analysis. These plots confirm that each of the GPS time series compares well with the radiosonde data, with the time series of median estimates providing the best match (orthogonal standard deviation = 1.74 mm), performing slightly better than the WVR (2.14 mm). The results match those found by previous investigations (e.g., Emardson et al. 1998; Tregoning et al. 1998) and confirm that even though the NRT is the weakest of the GPS estimates (1.89 mm) it is sufficiently accurate for weather prediction applications. Note that the somewhat large standard deviations quoted are the orthogonal standard deviations and so include the contribution from the radiosonde measurements, suggesting that the actual scatter in the GPS estimates can be expected to
be rather smaller than the quoted numbers. Investigations into the accuracy of PWV estimates derived using predicted orbits (e.g., Dodson and Baker 1998; Kruse et al. 1999) suggest that as long as the prediction lead is only on the order of hours, the accuracy of the PWV estimates is not badly degraded. Because the median estimate compared best with the radiosondes and the median is a robust operator, the time series of median estimates is taken as the reference time series with which to examine in detail the results of the sliding-window process.

Removing the median estimate for each epoch from all the estimates gives a set of residuals whose behavior can be analyzed as a function of their window position. Plotting the residuals for all seven WVIOP97 GPS sites as a function of their knot number within the window reveals the position-dependent scatter of the residuals (Fig. 5). As might be expected, the rms of the residuals rises near the edges of the window, with the rms scatter for estimates from the middle 3 knots approximately 5 times lower than for estimates from the first and last knots. This identifies one of the weaknesses with GPS real-time estimates: the estimate that is of most interest to weather forecasting is the most poorly constrained. For epochs leading the most current epoch, however, the opportunity exists to update the “best” estimate with each successive window solution. Since the median preserves steps and is robust in the presence of outliers, it is a good choice for the best estimate. However, other operators might reasonably be used. One of the simplest alternatives to the median is a weighted mean combination (see Fig. 5). The epochs estimated by the first 2 knots in the window have order 1, so all estimates are identical by definition. As the order (and therefore the number of estimates available) for each epoch increases the best estimate for that epoch can be cumulatively updated. The rms for the cumulative best estimate drops continuously until it (by definition) matches the reference estimate at $N_{\text{max}}$. The weighted mean performs slightly better than the median, but what is most notable is that neither it, nor the median, is able to provide a better estimate than simply using the most recent estimate for each epoch until the lead is more than half the window width. This indicates that the estimates are not independent, and that improving convergence to the final estimate would require that the correlation between estimates be taken into account by a more complicated operator, such as a Kalman filter.

One way to investigate the dependency of successive estimates is shown in Fig. 6. Here we plot the means of the residuals for each position within the window. This shows whether there is a general tendency for the estimates to have a particular bias based on their window position, giving us some insight into trends in the estimates that might be being introduced by the processing itself. Figure 6 shows that although the overall trend of all the sites combined is close to zero, several of the sites, in particular LMNO, have small individual trends that might be of concern. The source of these trends is...
not clear; possibly some processing parameter is slightly overconstrained for some of the conditions experienced during WVIOP97. Further, more detailed sliding-window analysis of this dataset might provide more insight into the cause(s). The figure indicates that the correlation between the residuals is small over the time period of the WVIOP97 experiment. Mean residual plots for other networks (not shown) tend toward zero as the length of time considered increases; however, more significant window trends may be present over time periods on the order of days. Another visualization of window trends is provided by a 2D image of the PWV residuals from the median estimates (Fig. 7). This indicates a quasi-periodic component in the residuals with a time period on the order of a few hours. In the most severe case, at the beginning of day 13 of WVIOP97 the residuals range from −3 mm too moist at the beginning of the window to ∼3 mm too dry at the end of the window. Two similar, but reversed, examples appear toward the end of the same day. The pairing of these residuals, with opposite signs appearing at opposite edges of the window, suggests that they might be due to the zenith delay change constraint parameter being too tight. If there is an event with a temporal delay gradient that is greater than the delay change parameter can accommodate, there will be under- and overshoot in the delay estimates as the window slides past the event (reversed if the gradient is negative). Although many of these zones of extreme residuals at the edges of the windows are paired, there are also examples that are not: the epochs near the end of day 12 show only small residuals near the end of the windows despite high residuals near the beginning of the window; the source of these anomalies is unknown.

The potential for sliding-window analysis as a tool for fine-tuning processing parameters is illustrated by Fig. 8. The data here come from a network of GPS sites on the island of Hawaii during a large storm event (Foster et al. 2003). The dataset was processed using the sliding-window technique with an 8-h window and an hourly window step. Two piecewise-constant atmospheric gradients were estimated per window. By comparing the differences between the two gradients estimated within each processing window with the differences between gradient estimates for corresponding epochs but from windows 4 h apart, we can examine whether the gradient variation constraint used for the processing is allowing the model sufficient range. The results clearly show that the gradient differences are significantly smaller when calculated from within the windows than when calculated between windows (Fig. 8a). This suggests that the gradient variation parameter was overconstrained. As the magnitudes of the gradients suggest, this was an extreme storm event with peak

Fig. 6. Plot of the mean residuals from the median epoch estimate. The heavy solid black line is the mean for all the WVIOP97 sites. The black dashed line is HBRK, the black dotted line is LMNO, the medium gray solid line is HVLK, the medium gray dashed line is PRCO, the medium gray dotted line is VCIO, the light gray dashed line is HKLO, and the dotted light gray line is NDSK.

Fig. 7. Plot of residuals from the median estimates for PRCO for days 11–15 of WVIOP97.
rainfall rates of over 4 in. h\(^{-1}\), and the default choices of processing parameters were unable to accommodate it. The processing was repeated, this time with the gradient variation constraint loosened from the original 0.02 to 0.04 m h\(^{-1/2}\). The new results (Fig. 8b) fall almost exactly on the 1:1 line, indicating that we are now getting estimates that are consistent. (The effect of the relaxed constraint is also visible in the overall scatter: the loosely constrained estimates are \(\sim 50\%\) more scattered.)
Window width $W$  
Knot interval $\delta$  
Window increment/step $\Delta$  
(Window) solution  
Epoch estimate  
Nearly real-time estimate  
Best estimate  
Epoch order

cessing, the most relevant time is the most recent—last—epoch.

The time width ($h$) of the window.

Time between each atmospheric estimate in a solution.

Time step between successive solutions.

The complete set of (atmospheric/ZND) parameters estimated by a processing run.

Atmospheric delay (ZND) estimate for a fixed epoch.

The first estimate for an epoch. (Strictly only appropriate for the last estimate of the most recent window when the processing is in a near-real-time mode; otherwise the term “most recent estimate” is more correct.)

The median (or other statistically derived) estimate for an epoch determined from all the estimates available for that epoch.

The number of estimates available for the epoch.

REFERENCES


