ACCURACY AND PRECISION OF GPS CARRIER-PHASE CLOCK ESTIMATES

K. Senior and J. Ray
U.S. Naval Observatory
3450 Massachusetts Avenue NW, Washington, DC 20392, USA
Tel: (202)762-1456, (202)762-1444
E-mail: *ksenior@usno.navy.mil, jimr@maia.usno.navy.mil*

Abstract

The accuracy of GPS-based clock estimates is determined by the pseudorange data. For 24-hour arcs of global data sampled at 5-minute intervals, the formal errors for the clock estimates are typically about 115 ps. An internal test of the actual time transfer measurement accuracy has been made by comparing clock estimates at the boundaries between successive analysis arcs for receivers equipped with very stable oscillators, using the combined daily clock estimates provided by the International GPS Service (IGS). During the period 29 October 2000 to 28 July 2001, the observed day-boundary discontinuities for individual IGS stations have distributions well described as zero-mean and Gaussian. However, the variances among the 23 stations span a wide range, from rms values of about 170 ps to 1200 ps, implying time transfer accuracies ranging from about equal to the formal errors to nearly an order of magnitude greater. For a few stations, the performance changes dramatically with time. Since the same receiver and antenna models are common for many stations, it is likely that the dominant site-dependent effects are related more to local factors affecting data quality than to specific hardware choices. The ALGO and NRC1 stations display notable temporal variations that might be seasonal. We find that a portion of the variability is caused by sensitivities to long-term temperature changes, with coefficients of -101 ps/°C and 156 ps/°C. Smaller, less significant temperature-dependent effects are seen at some of the other stations. After allowing for temperature-dependent effects, much larger day-boundary jumps remain for ALGO and NRC1 during winter 2000/2001, possibly due to signal reflection off snow-covered surfaces near the antennas. Temporal changes in time transfer accuracy at HOB2 are probably related to damage to the antenna cable. The causes for similar changes at MATE and briefly at NYAL/NYA1 have not been identified. Using our thermal sensitivity results for USNO, where the cable and receiver systems are well isolated from environmental changes, we have extended our previous study of diurnal temperature effects on AOA Dorne Margolin choke-ring antennas to put an upper limit of 10.1 ps/°C on any possible pseudorange-induced long-term variations due to this type of antenna.

The precision of clock estimates within a given analysis arc is usually assumed to be better than indicated by the formal errors or the accuracy measures because the relative clock estimates are determined mostly by the carrier-phase observations. We confirm this for intervals shorter than about 1 day, the analysis arc length. Average Allan deviations are well correlated with the day-boundary accuracy measures and imply a stability floor for carrier-phase time transfer up to 1 day of $2 \times 10^{-13}\tau^{-0.44}$, *consistent within measurement errors with a random-walk process. For longer intervals we expect the stability to improve as* τ^{-1}, *but this cannot be tested until more stable station clocks are available.*

INTRODUCTION

GPS-based techniques have been the basis for high-accuracy time transfer for about two decades. The "common view" method [1], employed by the Bureau International des Poids et Mesures (BIPM) to form the international UTC atomic time scale, uses single-frequency C/A pseudorange observations and relies heavily on cancellation of common-mode errors to achieve intercontinental comparisons between timing laboratories with uncertainties of a few ns at 5-day intervals. By contrast, high-accuracy geodetic methods using dual-frequency GPS carrier-phase observations achieve positioning repeatabilities at the sub-cm level (30 ps) for 1-day integrations. Assuming such positioning results can be realized also as equivalent light-travel times, the potential of GPS carrier-phase and geodetic techniques for global time and frequency comparisons at the sub-ns level is evident. This was the rationale in establishing a joint pilot project between the BIPM and the International GPS Service (IGS) to develop and demonstrate the operational capabilities [2].

GPS phase data are essential for modern geodesy. With the "double differencing" technique, for example, which incidentally removes all clock-like effects, pseudorange data are not normally even included in the geodetic analysis, as these observables are about 100 times less precise than the phase measurements. However, to analyze undifferenced data and extract clock estimates, it is necessary to add the pseudorange data in order to permit separation of the otherwise indistinguishable clock offset and phase cycle ambiguity parameters. See Larson and Levine [3] for a complete description of the GPS phase and pseudorange observation equations. For each tracking station-satellite pair, the quality of the clock estimates is maximized by ensuring the longest possible spans of continuous phase data free of cycle slips, thus minimizing the number of ambiguity parameters. High-performance geodetic receivers typically track individual satellite passes with arcs of up to 3 to 4 hours. Apart from viewing obstructions, the most problematic tracking is usually at the lowest elevation angles, where the signal strength is most attenuated and the atmospheric path delay is greatest and most variable.

Thus, the "absolute" accuracy of geodetic GPS clock estimates (neglecting the instrumental calibration biases here) is determined solely by the pseudorange data. More properly, it is the average over the chosen analysis span of all the pseudorange data involving a particular clock that determines the accuracy of the estimates for that clock. For an analysis arc of 24 hours using a global tracking network sampled at 5-minute intervals, the formal error of each clock estimate is typically about 100 to 125 ps, assuming each pseudorange observation has an uncertainty of 1 m and using standard analysis strategies and *a priori* variances. The thermal noise figure for pseudorange data is only a few cm, but the effects of multipath are much larger and account for the widely adopted 1-m value. On the other hand, the far more precise carrier phase data determine the detailed time variation, and hence precision, of the clock estimates. Several experiments seem to bear out the expectation of timing precisions (or frequency stabilities) better than the formal errors, down to the level of 10^{-15} or better at 1-day intervals [4,5]. But these tests have only been conducted for zero-length or very short baselines where an independent ground truth can be established. Due to differences in common-mode error cancellation, such results cannot be safely extrapolated to longer baselines where it is more difficult to determine the true time transfer precision, especially for averaging intervals much less than a day.

In this paper we perform an internal test of actual clock measurement accuracy by comparing independent clock estimates at the boundaries between consecutive 1-day analysis arcs. This is analogous to the classic geodetic repeatability test using a time series of position determinations. To ensure that the extrapolation error due to instabilities in the station clocks is minor, only those receivers equipped with H-maser external frequency standards are included. Alternatively, one could consider the clock differences during overlapping intervals, which would not require high-stability oscillators. Larson *et al.* [6] have done this for a single long baseline to remove the day-boundary discontinuities. As an estimate of measurement accuracy, however, that approach suffers from the unknown correlation between the overlapping arcs. For this study we prefer to avoid this difficulty by using statistically independent clock estimates from adjacent, non-overlapping analyses. Similar, preliminary results were reported by Ray *et al.* [7] based on clock solutions posted by the U.S. Naval Observatory (USNO). Those were for pairs of stations, however. Here individual stations are analyzed from the combined clock results of the IGS for a 39-week period. Time series of day-boundary clock discontinuities are obtained for 23 tracking stations, together with computed root-mean-squared (rms) deviations. From these we can infer lower bounds on the true accuracy of the clock estimates.

Before proceeding it is worth noting several distinctions between the observed estimates for clock parameters from a geodetic analysis and direct laboratory measurements of clock standards. In geodetic analyses, any effect which is bias-like will be absorbed into the clock estimates. The observability of clock parameters depends strongly on being able to model the geodetic problem adequately, with all other parameters of importance having partials that are much different from unity (or, more generally, from any constant value). In addition, to distinguish carrier-phase cycle ambiguities, pseudorange observables are required (as described above). These conditions are generally well satisfied for GPS data where the dynamics favor a very clear separation of clock effects from other parameters. Moreover, proper interpretation of the clock estimates in terms of the behavior of a particular frequency standard requires that additional effects be understood to relate the internal receiver clock, determined in the geodetic analysis, to an external reference. These effects include the various contributing timing biases within the GPS tracking antenna, receiver, and associated electronics. This latter area is broadly considered as "instrumental calibration" and is regarded here as a separable problem that must be addressed independently through suitable comparison measurements involving absolute or differential standards (e.g., [8,9]. The present study attempts to characterize the intrinsic performance of the geodetic time transfer method for remote measurements of clock estimates, independent of the instrumental calibration aspects. To the extent that the instrumental biases may not be constant, the analysis here will be affected and this is considered below where appropriate. The level of geodetic modelling is assumed to correspond to the standards realized by the IGS for its products.

OBSERVATIONS OF DAY-BOUNDARY CLOCK DISCONTINUITIES

Our raw data set consists of the combined clock estimates of the IGS for a subnetwork of the global tracking network [10]. The IGS "Final" clock series is used here, starting 29 October 2000 for the 9-month period through 28 July 2001. These data in their original form are not suitable for this test because of large instabilities in the underlying time scale, which would introduce spurious day-boundary discontinuities. The IGS time scale is obtained by linear alignment of the full constellation of satellite clocks to broadcast GPS time for each day independently. This alignment strategy keeps the IGS clock products close to GPS time, but causes large breaks in time and frequency between the individual days. These instabilities have no impact on the geodetic utility of the clock information, since the inter-clock and clock-orbit consistency is fully maintained, but do limit the usefulness of IGS clocks for timing studies. To address this latter issue, Senior *et al.* [11] have developed a procedure which relies on the numerous high-stability frequency standards within the IGS tracking network to synthesize a much better internal frequency ensemble that can be used to re-reference the IGS clocks. The time scale of the realigned IGS clocks has an estimated stability of about 10^{-15} at 1 day, better than that of any individual H-maser. These realigned clock products, which fully preserve the internal consistency of the original IGS clocks, are available at the Web site *http://clockdev.usno.navy.mil/igst/final.*

The daily clock estimates of Senior *et al.* have been analyzed for 23 tracking stations, all equipped with H-maser frequency standards. Figure 1 illustrates the day-boundary clock jumps for five stations during a representative 8-day period. The algorithm for measuring day-boundary clock jumps must recognize and ignore a variety of data anomalies. First, each 2-day span is detrended using a linear fit. The interval between clock observations of adjacent days closest to midnight must not exceed 30 minutes to avoid excessive interpolation error due to oscillator instability. The formal error of the clock difference at the midpoint between the 2 days must be less than 500 ps to ensure adequate observability. The rms variation of the clock values for each individual day must be less than 150 ps, compared with typical H-maser variations of 45 ps over 24 hours, to avoid confusion by other data problems unrelated to time transfer. Any clock jump greater than 5 ns is rejected as an outlier. The following results are not sensitive to these specific selection criteria. Statistics for the resulting day-boundary clock jumps are given in Table 1, together with the receiver and antenna model used at each site. Antennas equipped with a radome cover are so indicated.

The results in Table 1 are listed in order of increasing rms clock jumps. Not all stations are equally well sampled and results for some are clearly not robust, particularly GODE. Four stations show discrete changes in clock-jump behavior with time: ALGO (Algonquin, Ontario, Canada), HOB2 (Hobart, Tasmania, Australia), MATE (Matera, Italy), and NRC1 (Ottawa, Ontario, Canada). For HOB2 and MATE, the clock performances appear to sharply degrade following brief data gaps. At ALGO and NRC1, the clock jumps are much greater during winter than during other times of the year; Figure 2a shows the day-boundary jumps for NRC1. For each of these cases, statistics are given in Table 1 for the distinct periods (dates indicated in the table). The two receivers at Ny Alesund, Norway (NYA1 and NYAL), both connected to the same H-maser frequency standard but using different antennas, also seem to display a period of larger clock jumps during May-June 2001. Because the number of samples during the affected period is relatively small compared to the total data set, these have not been subdivided. In all other cases, the day-boundary clock jumps do not show distinct variations with time, although some are not well sampled.

ANALYSIS OF TIME TRANSFER ACCURACY RESULTS

The mean values for the clock jumps tabulated in Table 1 are always comparatively small, except for the sparsely sampled case of FORT. This is shown graphically in Figure 3. The FORT station is plagued with poor data quality due to its older generation receiver (AOA TurboRogue SNR-8000), which cannot track L2 under conditions of high ionospheric activity, causing the data coverage to be fragmentary most of the time during this period near solar maximum. The overall smallness of the mean clock jumps confirms that the geodetic time transfer technique does not induce spurious long-term drifts into the clock estimates, at least not at a significant level. To examine the distribution of clock jumps, histograms for the well-sampled USNO and WSRT stations are shown in Figure 4 compared with Gaussian distributions of like mean and standard deviations. All points outside the plot range are placed in the endpoint bins. The WSRT distribution matches a Gaussian quite closely, whereas the USNO data are overly peaked near the mean. The means in both cases are insignificantly different from zero.

The most striking feature of the rms clock jumps listed in Table 1 is the very wide range of values observed, from about 170 to 1200 ps. Since each day-boundary difference involves two measurements, the implied error in any single clock measurement will be smaller by $\sqrt{2}$ if the correlations between successive days are negligible. Interpreting the results in terms of lower limits on the time transfer accuracy, the performance of the technique is evidently highly station-specific, ranging from about equal to the formal error expectation to nearly an order of magnitude greater. Generally, the choice of receiver and antenna model does not seem to be the determining factor, judging from Table 1. However, the best performance of any station equipped with an older TurboRogue (IRKT) is poorer than either the newer AOA (ACT models) or Ashtech receivers, though possibly for other reasons. Use of antenna radomes also does not affect time transfer performance.

To investigate possible sources of station-specific time transfer error, it is instructive to consider those cases of clear temporal changes in performance. The HOB2 clock discontinuities are shown in Figure 5a. Following an interruption in the time series during April-May 2001, the HOB2 day-boundary clock jumps became much noisier. The gap in the time series is caused by frequent receiver clock resets beginning on 16 April (MJD 52015), which make it impossible to determine reliable day-boundary clock jumps. Inspection of IGS Mail messages and the site log does not indicate any particular event that might be responsible for the change in performance. Correspondence with station personnel (P. Digney, private communications) revealed that the clock reset period corresponds with a large increase in phase cycle slips. This can be seen in Figure 5b, which plots the daily number of cycle slips as determined by the quality-checking utility 'TEQC' [12]. The increase in cycle slips was traced to a sharp drop in the signal-to-noise ratio (SNR) due to corrosion and moisture in the connection of the Heliax 50 ohm underground cable to the antenna. Judging from Figure 5b, this problem seems to have begun around 7 March 2001. The connection was repaired, but not until 17 August (see IGS Mail #3477), after the end our clock study period. Apparently, some temporary corrective action was taken earlier, on about 14 May (MJD 52043), after which day-boundary clock jumps were again measurable, but with rms values increased by more than a factor of two. Figure 5b shows that the frequency of cycle slips improved in mid-May, but not to the level seen earlier in the year before the problem began. Another diagnostic output from TEQC has been consulted to check for correlations with the post-May time transfer performance. MP1 and MP2 report the rms variations of the pseudorange multipath at the L1 and L2 frequencies, allowing for unknown biases for each satellite and assuming the effect of phase multipath is negligible [12]. Aggregate MPi statistics are computed by TEQC, as well as values for individual satellites and for 5-degree elevation angle bins. Inspection of MPi measures for HOB2 fails to reveal any patterns clearly correlated with either the day-boundary clock jumps or the cycle-slip trends except for MP2 at high elevation angles. Figure 5c plots MP2 for the 85° to 90° bin as a function of time. In mid-May (around MJD 52041) the MP2(85-90) measure begins a sharp increase closely matching the greater clock jump rms in Figure 5a and coincident with a dramatic drop in the number of cycle slips (Figure 5b).

We hesitate to assert that high-elevation pseudorange multipath at the L2 frequency is directly or causally linked to clock measurement accuracy. For one thing, the correlation seen for HOB2 is not seen for other sites with time-varying day-boundary clock jumps. More likely, the high-elevation MP2 index reflects some other underlying error source, which also affects the number of cycle slips and leads to poorer time transfer accuracy. Without further information it is not possible to isolate the ultimate source. However, the sequence of events at the station suggests that perhaps some attempt in mid-May to ameliorate the cable corrosion problem, prior to its ultimate repair in August, may have improved the SNR, but inadvertently introduced some other signal problem, such as an impedance mismatch, giving rise to internal multipath at the L2 frequency.

The case with MATE is less informative because no other quantitative measure seems to correlate with the degradation in time transfer accuracy beginning in late April 2001. There is no public record of any specific events that could have affected the data quality, but correspondence with station personnel (D. Del Rosso, private communication) reveals that the frequency synthesizer of the H-maser was adjusted on 26 April 2001. This inadvertently caused a tracking problem in the GPS receiver that was corrected on 2 May. These events caused a gap in our time series of day-boundary clock jumps, but it is unclear how the degraded time transfer accuracy afterwards could have resulted. The receiver was swapped on 25 September 2001, to a new Trimble 4000SSI unit (see IGS Mail #3538), but it was not stated whether this action was taken to correct any specific problem. Our clock results do not yet extend to that epoch to check for any possible relationship with the receiver change.

Interpretation of the NRC1 and ALGO clock measurement results is complex. The indication of a possible seasonal trend in the day-boundary jumps (see Figure 2a), even though our data set spans less than 1 year, suggests checking for temperature dependencies. Of the stations in our study set, meteorological data are available for 11. We have extracted daily mean temperatures during the period 29 October 2000 to 28 July 2001, and tested for correlations between the day-boundary clock jumps and differences in mean daily temperatures. The results are given in Table 2. Only ALGO and NRC1 exhibit temperature dependencies that are significant at a level greater than 5 sigmas, while WTZR, DRAO, and ALBH show effects at about 4 sigmas. None of the remaining six stations shows any correlation of day-boundary clock variations with temperature.

Before examining the ALGO and NRC1 results more carefully, let us first consider the implications of those stations with no evidence of temperature sensitivities. Various studies [4,5,13] have demonstrated the effects of temperature variations on individual components of the tracking stations. Typical sensitivities for the receivers alone are of order ±100 ps/°C with large variations among individual units, including for the same model [14]. Common RF antenna cables have thermal sensitivities around 1 ps/°C/m and cable runs frequently exceed 50 m. Much better cable types are available, having temperature coefficients of about 0.03 ps/°C/m [15], which are coming into use at timing labs, though not so commonly at other geodetic installations. Trenching and other means of environmental isolation are sometimes used to minimize the effects of temperature cycling on antenna cables. We have previously shown [16] that the short-term (diurnal) temperature stability of AOA Dorne Margolin choke-ring antennas is better than 2 ps/°C. However, that study could not exclude the possibility of longer-term temperature-induced effects due to sensitivity in the daily average of the pseudorange observations.

The results in Table 2 for USNO, which show a long-term null temperature sensitivity of 3.7 ± 6.4 ps/°C, can now be used to extend our earlier antenna results. As documented by Ray and Senior [16], the USNO cable and receiver systems are very well isolated from environmental variations. The only significant uncontrolled component of the station hardware is the outside antenna unit, an AOA Dorne Margolin T choke-ring (see Table 1). The uncertainty in our estimate of the long-term USNO station thermal coefficient therefore places an upper limit of 10.1 ps/°C (1 sigma) on any possible pseudorange-induced thermal variation due to this type of antenna.

Returning to ALGO and NRC1, their inferred temperature-sensitivities are very significant, about 10 sigmas. J. Kouba (2001, unpublished report) has documented diurnal temperature dependencies in the raw pseudorange data from ALGO and DRAO, with the effects being larger at ALGO by a factor of two to three. The antenna cables in both cases are underground and should not be especially sensitive to diurnal changes. The cable type at ALGO is Andrew LDF4 (M. Caissy, private communications), which has a temperature coefficient between 0.027 and 0.061 ps/°C/m. For this reason, Kouba concluded that the thermal variations are caused by either the antennas or the receivers. Our results above for USNO exclude any significant effect due to the Dorne Margolin antennas, assuming that individual units behave similarly. Thus, the temperature sensitivities at ALGO, NRC1, and to a lesser extent at DRAO, point to inadequate isolation of the receivers from external environmental changes. Station personnel for ALGO and NRC1 (M. Caissy, private communications) indicate that this possibility cannot be excluded. Efforts are underway to begin monitoring temperature variations near those receivers. However, other components of these systems must also be considered, such as the connectors and the antenna power splitter used at NRC1.

Regardless of the underlying source of the temperature sensitivity, this effect alone cannot fully account for the magnitude of the clock jumps at ALGO and NRC1. That can be seen in the rms day-boundary clock residuals after removing temperature-dependent trends (Table 2), which are reduced by only 20% to 30% compared with the original rms values (Table 1). Even more important is the persistence of apparent seasonal trends after removal of the temperature dependencies. Table 3 gives the day-boundary statistics for ALGO and NRC1 after accounting for the temperature trends in Table 2. Results are given for the same winter-summer periods as in Table 1. The rms clock jumps remain two to three times greater in winter compared with summer. Of the diagnostic metrics we have checked for data quality and station performance (such as those discussed above for HOB2), only the multipath MPi measures from TEQC show any correlations with the ALGO or NRC1 day-boundary clock jumps. However, the daily multipath variability is much smaller (roughly 10% to 50% between seasons compared with factors of two to three for the clock jumps), and the multipath indices are smaller in winter than in summer. Despite these inconclusive results, we speculate that greater wintertime multipath might occur due to signal reflection off snow-covered surfaces near the antennas, leading to poorer time transfer accuracy during those periods. It should be noted that MPi measures multipath variations with periods from 1 minute (IGS data sampling is 30 s) to about 3 hours (the typical satellite pass) and is not sensitive to longer wavelength or quasi-static biases. The ALGO antenna is mounted over a concrete pillar with the L1 and L2 phase centers being 21 and 22.8 cm above the top of the pillar, and with a space of 13.5 cm between the bottom of the antenna choke rings and the pillar. This configuration might be prone to standing-wave reflections, especially during winter when the top of the pillar is covered by snow or ice. Elosegui *et al.* [**17**] described serious back-scattering problems for a similar antenna configuration, although not specifically involving snow cover.

The NYAL and NYA1 stations are separated by about 20 m and use independent antennas and receivers, although they are driven from the same external H-maser. NYA1 is intended as an eventual replacement for the older station due to concerns for the stability of the NYAL antenna mount. The parallel operation provides an interesting test case, although the tracking hardware differs only in the choice of antennas. It is perhaps not surprising then that their time transfer accuracies are very similar. During May-June 2001 the day-boundary clock jumps for both stations experienced greater than usual variations (included in the statistics in Table 1). This coincidence implies an error source external to the tracking systems themselves. However, the absence of any correlation with diagnostic metrics or any published events at Ny Alesund prevents us from identifying a cause for the large day-boundary jumps.

Without further information, the causes for the general dispersion of time transfer accuracies among the various stations in Table 1 cannot be further isolated. We should expect essentially any factor that can degrade the pseudorange and/or carrier phase observables to be considered a candidate. Because multipath is usually the dominant observational error, it deserves special attention. Doing so will require well constructed models tailored for the individual stations, a task beyond this scope of this paper. Other local factors to be considered include temperature sensitivities (examined above for some stations), RFI and electromagnetic environment, internal impedance mismatches, receiver firmware, and so forth.

ANALYSIS OF TIME TRANSFER

The precision of clock estimates within a given analysis arc (and, hence, the frequency stability) is generally expected to be much better than either the formal errors or the absolute accuracy measures, because the relative clock values are determined mostly by the carrier-phase data (under normal circumstances). However, the actual performance has not been well characterized experimentally. For the few long baselines that have been well studied, the observed frequency stabilities are rarely better than about 2×10^{-15} at 1-day intervals (e.g., [**6**]). Also, the question remains whether the noise of the carrier-phase time transfer process behaves as white noise (Allan deviation as τ^{-1}) or random walk ($\tau^{-0.5}$) in the time domain. If the time transfer errors were perfectly white, the formal clock measurement uncertainties would imply 1-day stabilities at the level of 1.3×10^{-15} assuming no other effects are significant (such as intrinsic clock instability), but a very poor 3.8×10^{-13} at 5-minute intervals.

Given the very large dispersion in accuracy performance among IGS stations, it is natural to consider whether short-term stability varies similarly. Figure 6 shows the average Allan deviations at 300-s intervals compared with the rms day-boundary clock jumps for the 23 IGS stations in Table 1. For each station and day in the study period, Allan deviations were computed from the realigned IGS clock products for intervals of 300 s, 10^3 s, 10^4 s, and 3 hours. Then average values were computed for each station and each interval, rejecting as outliers any values greater than $(8 \times 10^{-13}) \sqrt{300\mathrm{s}/\tau}$. For the four stations with variable day-boundary behaviors, separate points are plotted for each of the two periods listed in Table 1 and those points are joined by lines. Five stations (ONSA, TIDB, CRO1, IRKT, and METS) have much poorer short-term stabilities than the rest. Inspection of the time domain plots for ONSA and METS shows clear diurnal variations, likely to be related to environmentally induced effects, while the CRO1 results appear "noisy" at high-frequencies, for unknown reasons. In each of these five cases, aspects of the local station configurations are more likely to be responsible for the poor short-term stability than is the time transfer method. For this reason, these stations have been excluded from a fit of the average Allan deviations versus rms day-boundary discontinuities. Our objective is to determine an approximate lower envelope of the short-term stability regime. The existence of a correlation between these distinct measures of long-term accuracy and short-term stability is reinforced by the observation that the two separate periods for ALGO, HOB2, and NRC1 parallel the general trend of the other included stations; only MATE behaves somewhat differently.

By extrapolating the trend in Figure 6 to an intercept of zero rms discontinuities (that is, presumably perfectly accurate time transfers), we can infer a limiting value for the short-term stablity due to the time transfer method. Repeating for 10^3 s, 10^4 s, and 3 hours gives the values plotted in Figure 7 and an overall stability limit power law of $2.01 \times 10^{-13} \tau^{-0.44}$. The formal error for the determination of the power law exponent is 0.07, so the behavior is not significantly different from $\tau^{-0.5}$, as expected for a random-walk error process. At an interval of 1 day, the inferred stability is 1.4×10^{-15}. Also shown in Figure 7 are the average Allan deviations for WSRT, the station with the best overall short-term stability and one of the best in terms of long-term accuracy. Our estimate for the limiting time transfer stability is about half that of WSRT.

We conclude that the short-term time transfer stability is indeed much better than implied by the instantaneous formal errors (as good as 2×10^{-14} versus 3.8×10^{-13} at 300 s), at least for intervals shorter than 1 d (see Figure 7). The benefit of implicit carrier-phase "smoothing" to improve the frequency stability, which must introduce important temporal correlations, probably explains the observed random walk rather than white noise behavior. In this regard, it is noteworthy to recall that GPS carrier-phase observations are also recognized as being equivalent to an integrated Doppler data type. At 1-d intervals the inferred stability is approximately equal to the level expected from the formal time transfer measurement errors and a white noise process (1.4×10^{-15} versus 1.3×10^{-15}). We expect, but cannot demonstrate, that the time transfer stability for intervals longer than 1 d will follow a white noise trend for the appropriate station accuracy. However, internal investigations near 1 d and longer are limited by the stability of the underlying time scale, its steering, and the frequency standards themselves [11]. To explore the actual stability performance over those intervals will require comparisons between stations equipped with more stable frequency standards, such as the new cold atom standards now under development. Our observational results here accord amazingly well with the expected performance predicted by Petit and Thomas [18] based on a purely theoretical error budget analysis.

DISCUSSION AND CONCLUDING REMARKS

Our central finding is that the accuracy of carrier-phase clock estimates is, in the best cases, about equal to the formal errors of approximately 115 ps for 1-day arcs. However, the performance can be much poorer, by nearly an order of magnitude, and seems to be highly site-dependent. Since several different antenna and receiver models in common use appear to give similar results, other site-dependent factors are probably responsible for the variable performance. Our study has identified only a few of the local error sources, notably temperature-induced variations and antenna cable problems. Generally, however, any effect which degrades the quality of the pseudorange or carrier-phase data must be considered. Multipath errors, which have been well studied for their effects on phase data, deserve particularly close scrutiny for pseudorange and clock effects.

Some authors who have compared carrier phase results with other time transfer techniques have chosen to remove the day-boundary discontinuities analyzed here. Larson *et al.* [**6**], for example, use half-day overlaps between successive 4.5-day analysis arcs to estimate and remove the discontinuities. They found a discontinuity rms of 222 ps over a 236-day period for the baseline between USNO and a station in Colorado Springs, CO (AMC2). This is reasonably consistent with our own rms estimate for 1-day USNO arcs of 236 ps, especially if one considers that the longer analysis arcs used by Larson *et al.* might be expected to yield somewhat more accurate results. However, the approach of successively removing discontinuities presents some drawbacks that should be considered. If, for example, the statistical relationship between the time transfer errors of successive analysis arcs can be described by a Gauss-Markov process, then the accumulated error in the adjusted clock time series will have a variance that grows with time as

$$\sigma^2(t) = 2T\sigma_0^2 \left[t - T + Te^{-t/T}\right]$$

where T is a characteristic correlation time constant for the clock errors between successive analysis arcs and σ_0 is the standard error for clock estimates of an individual arc (see [**19**] for background). Such an accumulation error could degrade the frequency stability if the correlation time is not very short (i.e., less than 0.5 d) and should be checked whenever this procedure is applied; our data set is not suited to analyze the autocorrelation properties. Dach *et al.* [**20**] have used simulations to illustrate other problems that can be introduced by long-term accumulation of systematic errors.

Another approach to remove discontinuities between processing arcs, by simply continuing the analysis forward without allowing any interruption in the time series ([**21,22**], raises other questions. While this will give the appearance of "error-free" clock estimates, no physical measurement process can actually be free of uncertainty. Potentially, only the information needed to objectively quantify the measurement errors is removed. It could well be that longer analysis arcs give more accurate clock estimates, as shown for the formal errors [**22**], but an error "floor" always intervenes in any real physical process to eventually deny $1/\sqrt{N}$ improvements. Whether the floor is significant at 1-day intervals or for longer periods remains to be established. It should also be noted that the time transfer stability could suffer for intervals longer than about 1 d due to the continuation of carrier phase-based correlations that would not otherwise occur with 1-d analysis arcs. That is, the expected transition from random walk to white noise stability at about 1 d (and at a stability of roughly 1.3×10^{-15}) for 1-d analysis arcs could very well lead to random-walk behavior extending to n days for an n-day analysis arc.

We have also shown that the short-term stability of carrier-phase time transfer results varies linearly with the inferred 1-day accuracy. This conclusion is not necessarily expected, since the pseudorange data largely determine the accuracy, whereas the carrier-phase data dominate the stability. This result implies that the error sources that affect pseudorange data and accuracy are highly correlated with the quality of phase data. Certainly that should be true for multipath errors, but other effects should be studied as well.

ACKNOWLEDGMENTS

We are indebted to all components of the IGS without whose accomplished efforts this work would not have been possible. We also thank Mark Caissy, François Lahaye, Paul Digney, and Domenico Del Rosso for their assistance in diagnosing some of the station-related effects, as well as Kristine Larson and Jan Kouba for helpful discussions.

REFERENCES

[1] D. Allan and M. Weiss, 1980, *"Accurate time and frequency transfer during common-view of a GPS satellite,"* in Proceedings of the 34th Annual IEEE Frequency Control Symposium, 28-30 May 1980, Philadelphia, Pennsylvania, USA (IEEE Publication AD-A213670), pp. 334-356.

[2] J. Ray, 1999, *"IGS/BIPM time transfer project,"* GPS Solutions, 2(3), 37-40.

[3] K. Larson and J. Levine, 1999, *"Carrier-phase time transfer,"* IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 46(4), 1001-1012.

[4] C. Bruyninx, P. Defragne, and J. Sleewaegen, 1999, *"Time and frequency transfer using GPS codes and carrier phases: Onsite experiments,"* GPS Solutions, 3(2), 1-10.

[5] G. Petit, C. Thomas, Z. Jiang, P. Uhrich, and F. Taris, 1998, *"Use of GPS Ashtech Z12T receivers for accurate time and frequency comparisons,"* in Proceedings of the 1998 IEEE International Frequency Control Symposium, 27-29 May 1998, Pasadena, California, USA (IEEE Publication 98CH36165), pp. 306-314.

[6] K. Larson, J. Levine, L. Nelson, and T. Parker, 2000, *"Assessment of GPS carrier-phase stability for time transfer applications,"* Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 47(2), 484-494.

[7] J. Ray, F. Arias, G. Petit, T. Springer, T. Schildknecht, J. Clarke, and J. Johansson, 2001, *"Progress in carrier phase time transfer,"* GPS Solutions, 4(4), 47-54.

[8] G. Petit, Z. Jiang, J. White, R. Beard, and E. Powers, 2001, *"Absolute calibration of an Ashtech Z12-T GPS receiver,"* GPS Solutions, 4(4), 41-46.

[9] G. Petit, Z. Jiang, P. Moussay, J. White, E. Powers, G. Dudle, and P. Uhrich, 2001, *"Progresses in the calibration of "geodetic like" GPS receivers for accurate time comparisons,"* in Proceedings of the 15th European Frequency and Time Forum (EFTF), 6-8 March 2001, Neuchtel, Switzerland (FSRM, Neuchâtel), pp. 164-166.

[10] J. Kouba and T. Springer, 2001, *"New IGS station and satellite clock combination,"* **GPS Solutions, 4(4),** 31-36.

[11] K. Senior, P. Koppang, D. Matsakis, and J. Ray, 2001, *"Developing an IGS time scale,"* in Proceedings of the 2001 IEEE Frequency Control Symposium, 6-8 June 2001, Seattle, Washington, USA (IEEE Publication 01CH37218), pp. 211-218.

[12] L. Estey and C. Meertens, 1999, *"TEQC: The multi-purpose toolkit for GPS/ GLONASS data,"* **GPS Solutions, 3(1),** 42-49.

[13] F. Overney, T. Schildknecht, G. Beutler, L. Prost, and U. Feller, 1997, *"GPS time transfer using geodetic receivers: Middle-term stability and temperature dependence of the signal delays,"* in Proceedings of the 11th European Frequency and Time Forum (EFTF), 4-6 March 1997, Neuchâtel, Switzerland, pp. 504-508.

[14] T. Schildknecht and G. Dudle, 2000, *"Time and frequency transfer: High precision using GPS phase measurements,"* **GPS World, 11(2),** 48-52.

[15] E. Powers, P. Wheeler, D. Judge, and D. Matsakis, 1999, *"Hardware delay measurements and sensitivities in carrier phase time transfer,"* in Proceedings of the 30th Annual Precise Time and Time Interval (PTTI) Systems and Applications Meeting, 1-3 December 1998, Reston, Virginia, USA (U.S. Naval Observatory, Washington, DC), pp. 293-305.

[16] J. Ray and K. Senior, 2001, *"Temperature sensitivity of timing measurements using Dorne Margolin antennas,"* **GPS Solutions, 5(1),** 24-30.

[17] P. Elosegui, J. Davis, R. Jaldehag, J. Johansson, A. Niell, and I. Shapiro, 1995, *"Geodesy using the Global Positioning System: The effects of signal scattering on estimates of site positions,"* **Journal of Geophysical Research, 100,** 9921-9934.

[18] G. Petit and C. Thomas, 1996, *"GPS frequency transfer using carrier phase measurements,"* in Proceedings of the 1996 IEEE International Frequency Control Symposium, 5-7 June 1996, Honolulu, Hawaii, USA (IEEE Publication 96CH35935), pp. 1151-1158.

[19] R. Brown and P. Hwang, 1992, **Introduction to Random Signals and Applied Kalman Filtering** (John Wiley and Sons, New York).

[20] R. Dach, G. Beutler, U. Hugentobler, S. Schaer, T. Schildknecht, T. Springer, G. Dudle, and L. Prost, 2001, *"Time transfer using GPS carrier phase: Error propagation and results,"* submitted to **Journal of Geodesy.**

[21] G. Petit, Z. Jiang, F. Taris, P. Uhrich, R. Barillet, and F. Hamouda, 1999, *"Processing strategies for accurate frequency comparison using GPS carrier phase,"* in Proceedings of the 1999 Joint European Frequency and Time Forum (EFTF) and IEEE International Frequency Control Symposium, 13-16 April 1999, Besançon, France (IEEE Publication 99CH36313), pp. 235-238.

[22] K. Senior, E. Powers, and D. Matsakis, 2000, *"Attenuating day-boundary discontinuities in GPS carrier-phase time transfer,"* in Proceedings of the 31st Precise Time and Time Interval (PTTI) Systems and Applications Meeting, 7-9 December 1999, Dana Point, California, USA (U.S. Naval Observatory, Washington, DC), pp. 481-489.

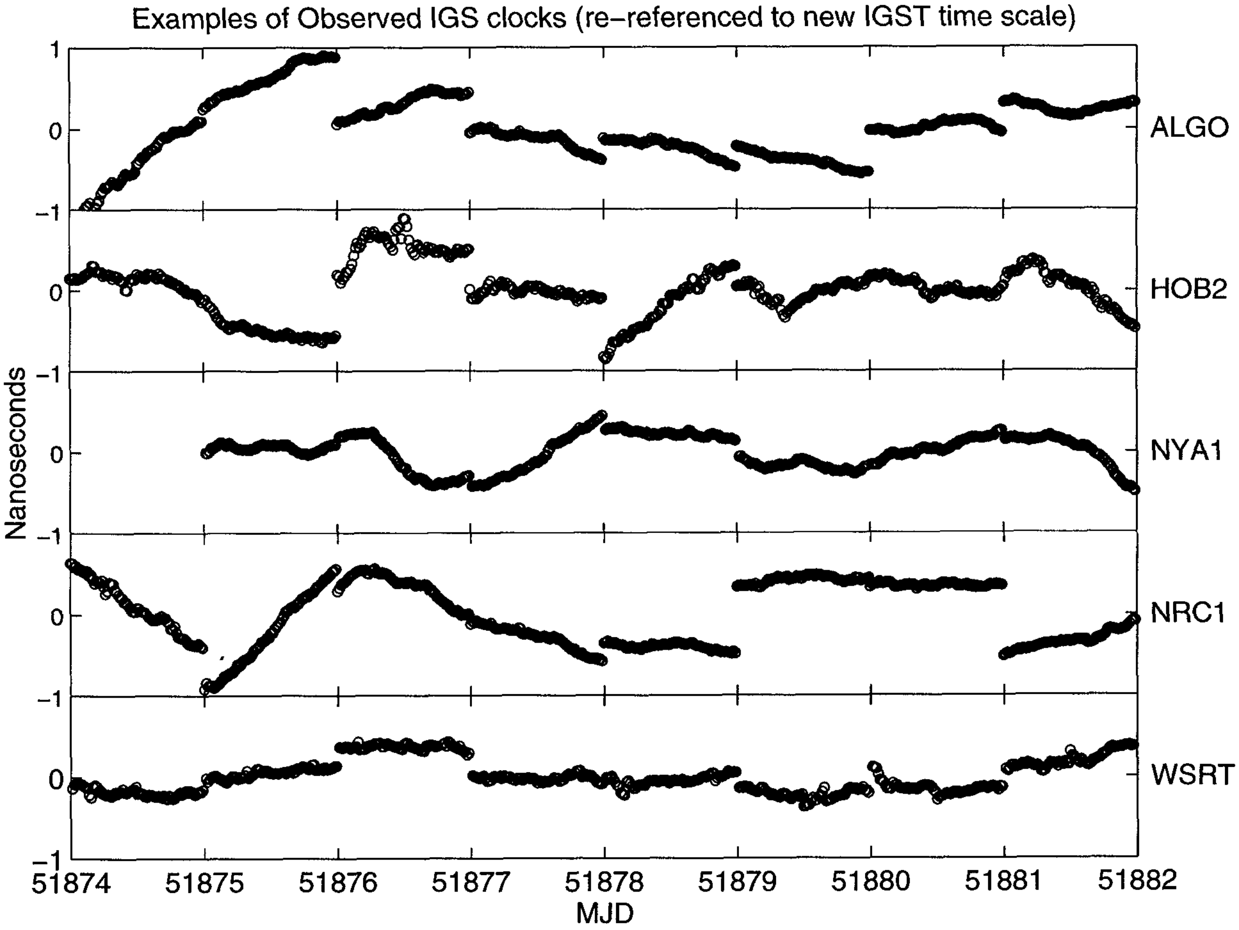

Figure 1: Time series of IGS clock estimates (after internal time scale realignment) for a sample of five stations equipped with H-maser frequency standards to illustrate day-boundary clock discontinuities. The period shown is 26 November to 4 December 2000. A separate quadratic has been removed over the entire interval from each series for the purpose of plotting.

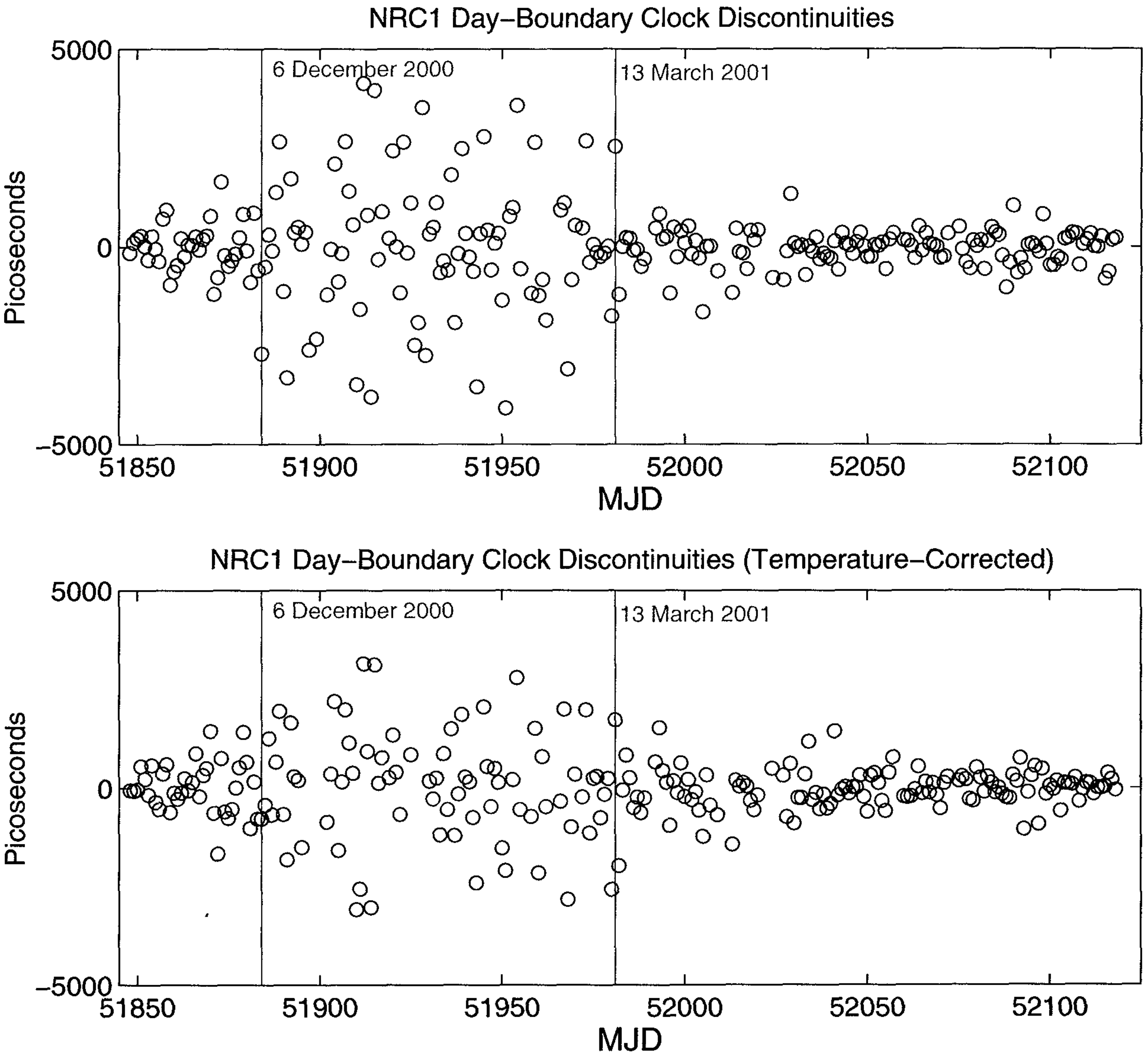

Figure 2: Time series of day-boundary clock discontinuities for NRC1 (Ottawa, Canada). Upper panel (a) shows the raw results, with much larger variations during winter than at other times of the year. Lower panel (b) shows the same results after removing a linear temperature-dependent trend (see Table 2). While the level of discontinuities during winter has been reduced, they remain significantly larger than at other times.

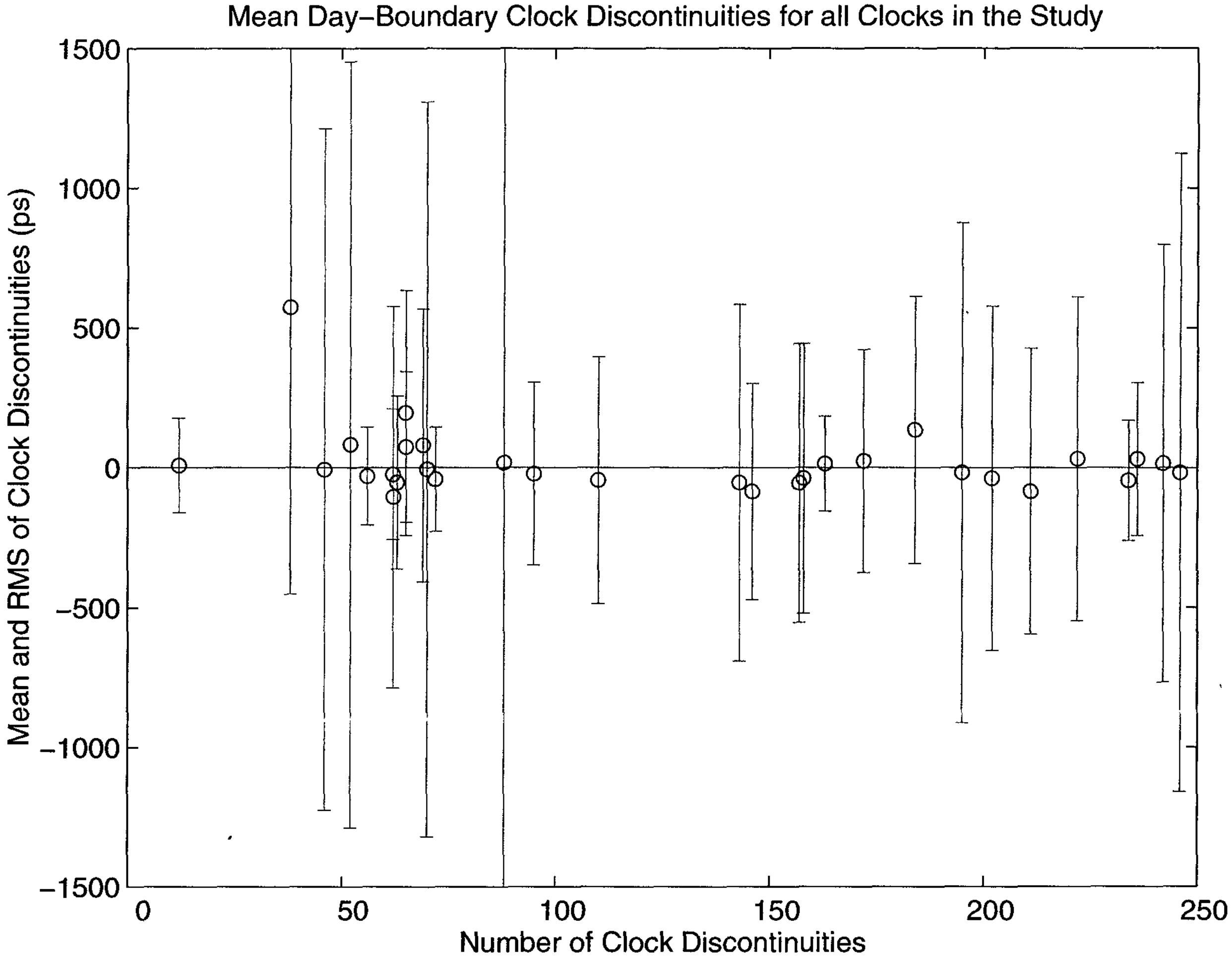

Figure 3: The mean and RMS values for the day-boundary clock discontinuities of all stations in the study as a function of the number of discontinuities available.

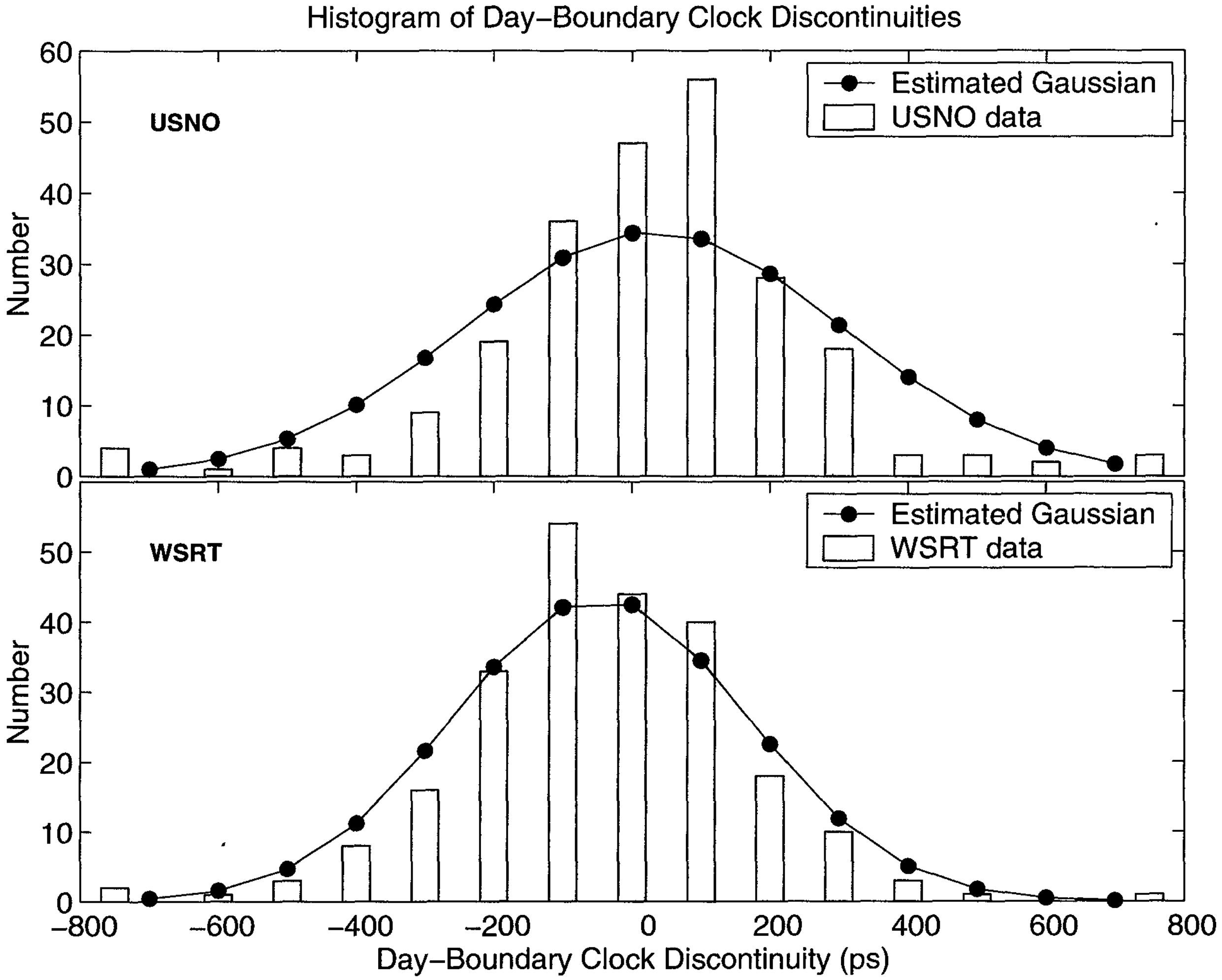

Figure 4: Histograms of the day-boundary clock discontinuities for USNO and WSRT compared with Gaussian distributions for each.

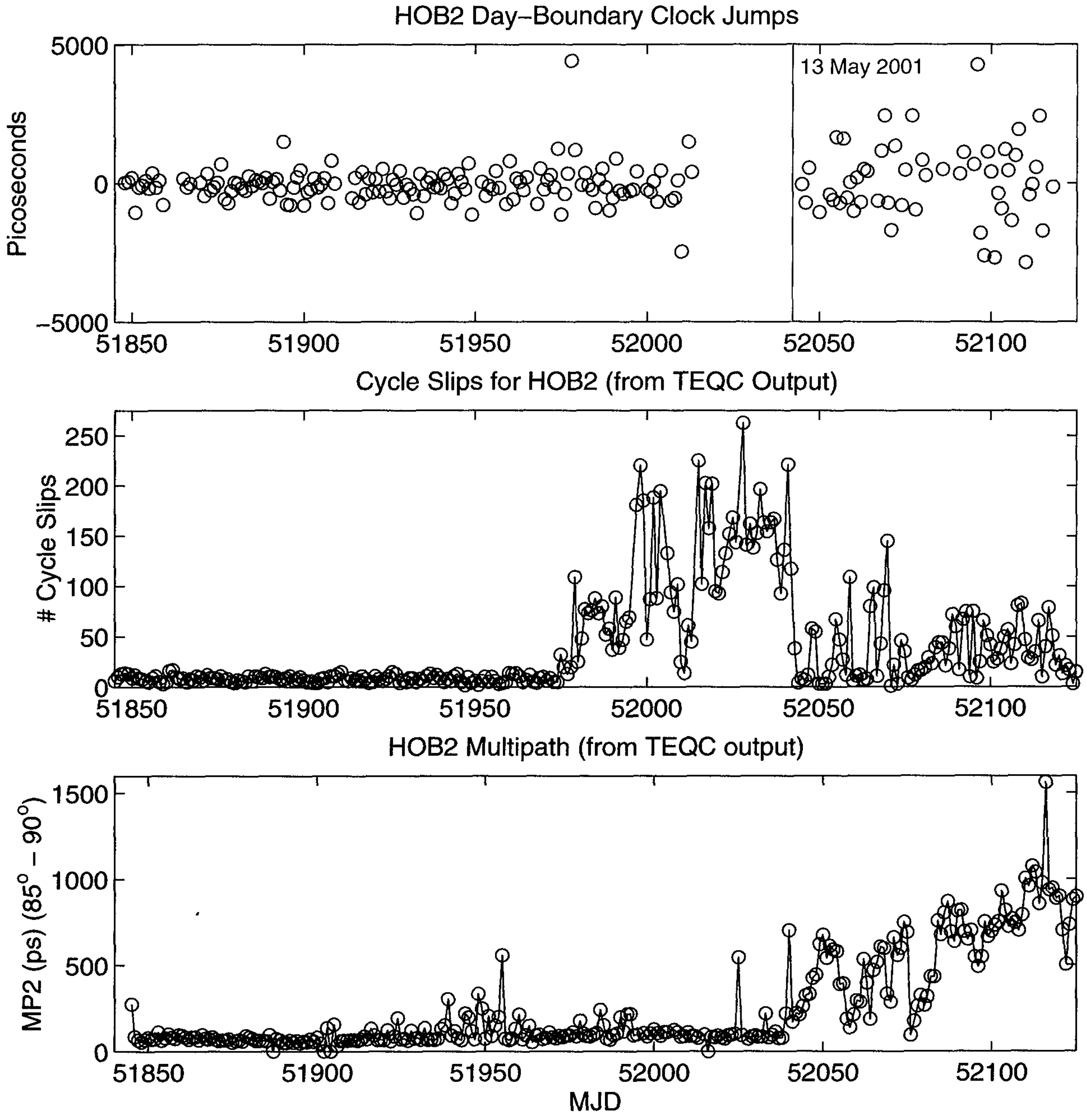

Figure 5: Time series of results for HOB2 (Hobert, Australia). Upper panel (a) shows the raw day-boundary clock discontinuities illustrating the large increase in variations after 13 May 2001. Middle panel (b) shows the daily number of carrier phase cycle slips as reported by the quality-checking utility TEQC. Lower panel (c) shows the TEQC multipath index MP2 for observations in the elevation angle bin from 85° to 90°.

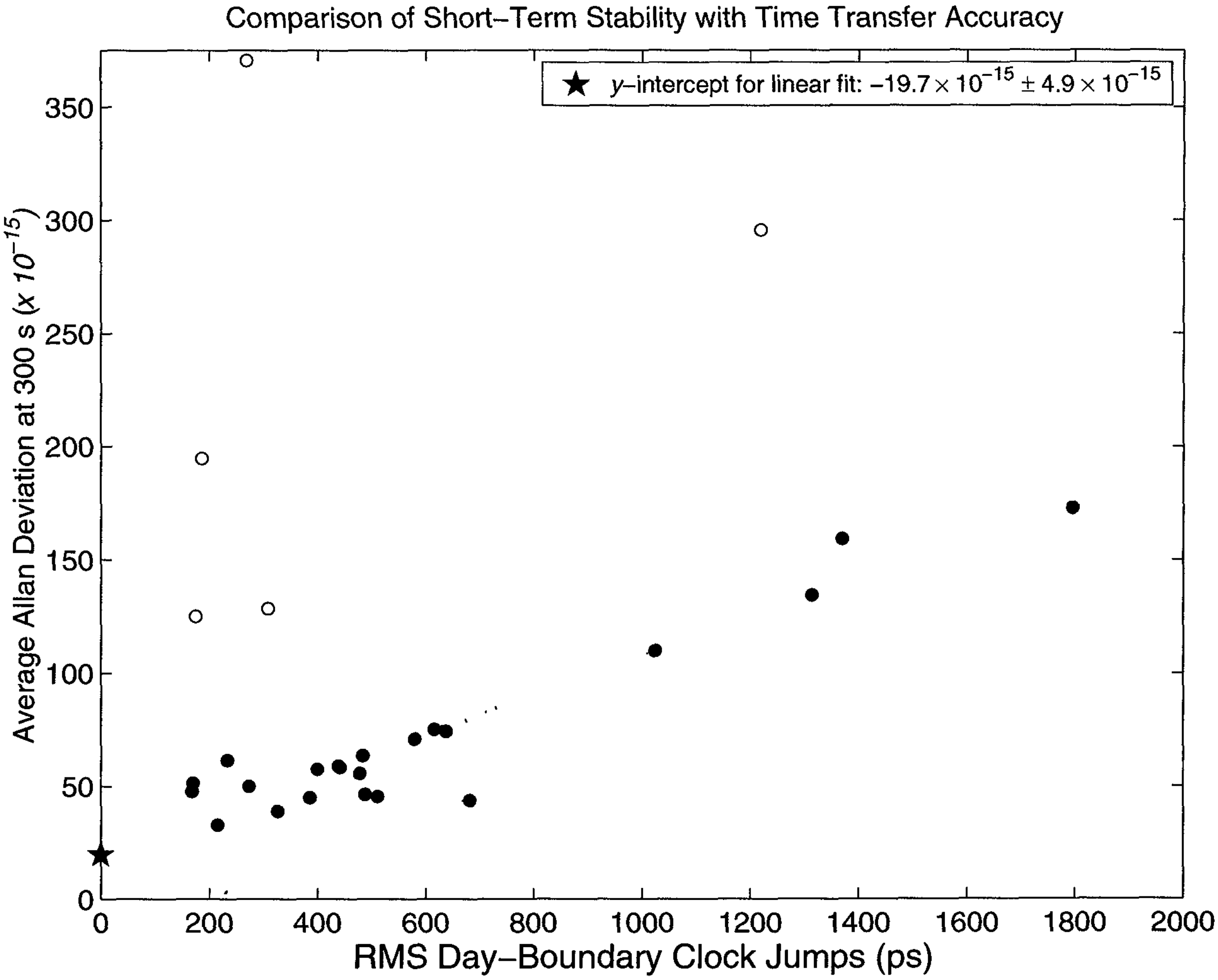

Figure 6: The average Allan deviations computed at 300-s intervals versus the RMS day-boundary clock discontinuities for all stations. Two connected points are shown for each of the stations ALGO, HOB2, MATE, and NRC1, for the intervals given in Table 1. The stations with local timing instabilities (ONSA, TIDB, CRO1, IRKT, and METS, indicated by open circles) are not included in the linear fit for the general trend.

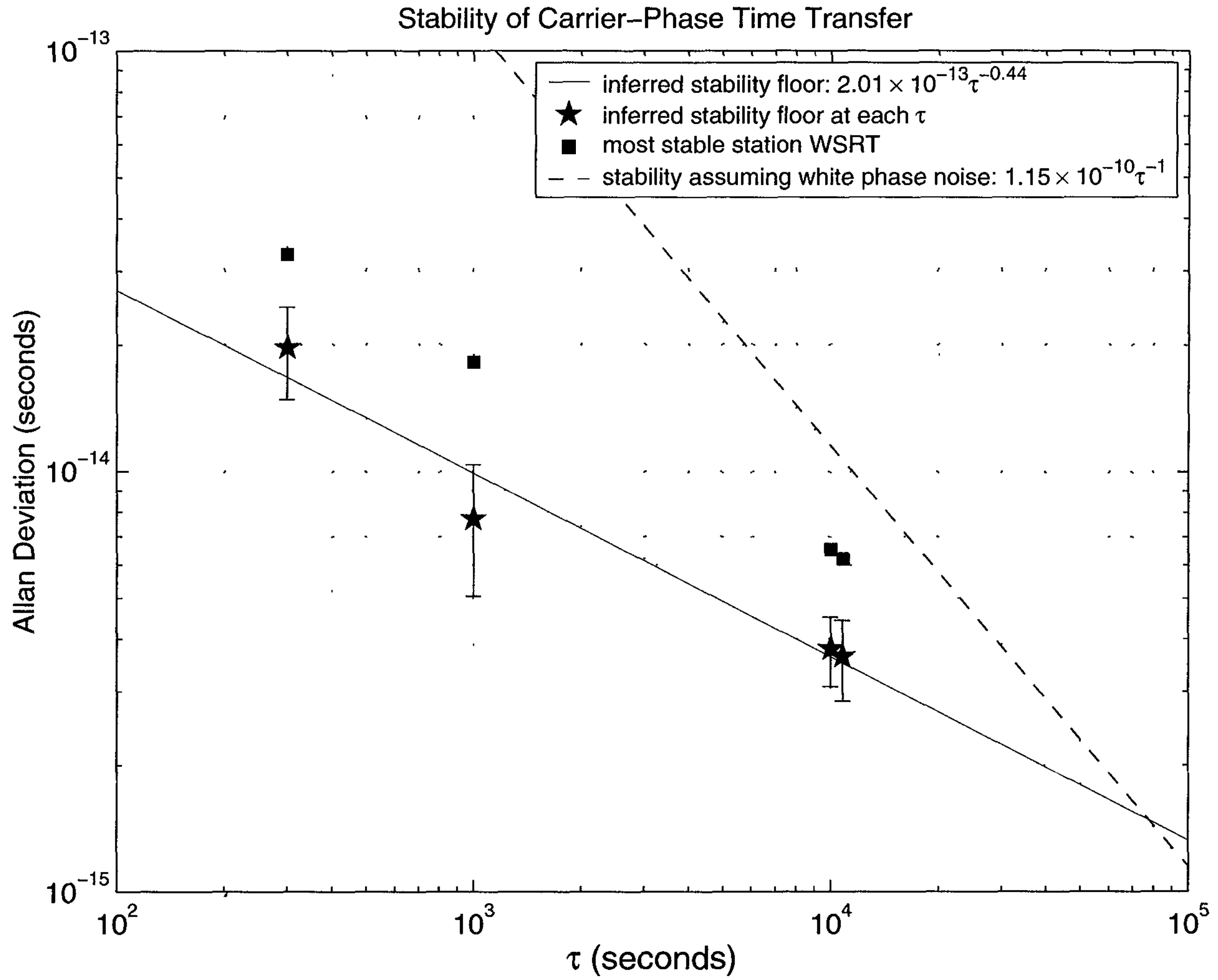

Figure 7: The inferred floor for Allan deviation stability of the carrier phase time transfer method compared with the average stability for WSRT, the station with the observed best overall stablility. Also shown (dash line) is the expected behavior if the time transfer stability behaved as a white noise process with an instantaneous uncertainty equal to the typical measurement formal errors (115 ps), which is consistent with the observed time transfer accuracy for the best performing stations.

QUESTIONS AND ANSWERS

VICTOR REINHARDT (Boeing Satellite Systems): Exactly what do you mean by "day boundary?"

KEN SENIOR: Looking at the carrier-phase estimates, which are independently reduced over 1-day arcs, I get the pseudo-range and carrier phase data, I put them through this big filter, and I determine what the clock estimates are for a given site. This is repeated independently each day. If you look, say, 30 minutes before the day boundary and 30 minutes after the day boundary – again, assuming you have a good stable clock – you determine the jump discontinuity across the processing boundary. So here are some sample estimates – each little tick mark represents a day. And these are estimates of Algonquin, Hobart, Ny-Alesund, NRC-1, and Westerbork, all referenced to the IGS Time Scale.

REINHARDT: So you say "discontinuity" just because you take the data of your stop and then you start again?

SENIOR: Essentially what is happening is that you are averaging pseudorange data, if you will, over each individual day, and this determines your timing accuracy. Whereas, the precision of the estimates is being determined by the carrier observables. So the variations within each processing arc are dominated by the carrier. But the real timing information is sort of an average of the pseudorange data. And that is where the accuracy is coming from.

REINHARDT: One other question and then I will let you go. Are you using single carrier here or are you using multiple carriers?

SENIOR: Dual.

REINHARDT: Okay, are you trying to resolve ambiguities here?

SENIOR: Oh, definitely.

REINHARDT: Okay. So, this is with the ambiguity ... ?

SENIOR: The ambiguities are estimated. And I know that some of the analysis centers in the IGS that contribute do ambiguity resolution. I cannot tell you which ones. They generally do double-difference ambiguity resolution, which tweaks the network, although generally that does not impact the clock estimates all that much. Usually it is the east component of positioning that is chiefly impacted.

TOM CLARK (NASA/Goddard and Syntonics): I was going to make two comments. First of all, you mentioned the pillar assembly; it is a somewhat bizarre site. For those of you who do not know the site, it is at 79 degrees north in Pittsburgh, and has very bizarre geometry, since its latitude is so much higher than the inclination of the GPS satellites, so all intuitive thoughts about what happens there tend to break down.

The other comment I had to make has to do with your high elevation multipath. This was a problem that I had looked at sometime in the past in terms of what does it come from. And you, I think, hinted that you thought that part of it had to do with the height of the pier.

SENIOR: Not in this case.

CLARK: But still, I was going to explain what high-elevation multipath came from – at least, I believe. All of these antennas have some radiation into their backfield, which is impingent on the top of the cylindrical pier that tends to hold these things up. Usually it is made of concrete, sometimes it will have some metal in the top of it. No one of them is the same; they are all different things. And the measurements I made and the theory I developed basically said that the multipath beam pattern that is affecting this is essentially due to the array disk, the defraction disk from something like the diameter of the pier. We simulated this on the antenna range and, yes, in fact, it happens – which means that the multipath is confined to an area on the sky is slightly larger in angle than λ/D radians. It looks like classical J-1 Bessel function for the antenna pattern of this stuff. So a high-elevation multipath is not surprising.

One thing that would be interesting to see, some people have gone to mounting antennas on much smaller diameter piers or tripods. And it would be interesting to do the comparison where you believe that you are having some of these multipath effects between antennas that are mounted on things that are smaller than the antenna versus ones that are mounted on these big concrete cylinders that go under the ground.

SENIOR: Thank you. Yes, I think it was NGS Web site that has an experiment with this pillar assembly which is in Algonquin, and I was only speculating that perhaps the multipath – and I don't think for that case, necessarily, high-elevation multipath – was a cause for the remaining 60 percent of those temporal variations in the rms that occurred seasonally. And that was pure speculation. In the NGS Web site, though, they described that particular situation setup that they have at Algonquin as being sort of the classic textbook case of what one would expect with pseudorange multipath. So –

MARC WEISS (National Institute of Standards and Technology): I have a question on IGST. If it is based on the IGS clocks, then it would somehow have time steps built in each day?

SENIOR: Yes, and we believe that they are much smaller, almost negligible compared to any of the sites. First of all, the banding and the stability of all the clocks are pretty close together. So we believe that we, with respect to stability, are getting square-root-N-type improvement in the stability, as one would expect with ensembling. But from our determinations, the clock that jumped discontinuities at the day boundaries of the scale are very, very small – much smaller than any other stations.