IOP Publishing | Bureau International des Poids et Mesures Metrologia

Metrologia **59** (2022) 035007 (9pp) https://doi.org/10.1088/1681-7575/ac6a95

A three-cornered hat analysis of instabilities in two-way and GPS carrier phase time transfer systems

T E Parker[1,*], V Zhang[2], G Petit[3], J Yao[4], R C Brown[1] and J L Hanssen[5]

[1] National Institute of Standards and Technology, Boulder, CO, United States of America
[2] National Institute of Standards and Technology, Boulder, CO, United States of America, retired
[3] Bureau International des Poids et Mesures, Sèvres, France
[4] Now at UCAR/NCAR, Boulder, CO, United States of America
[5] United States Naval Observatory, Washington DC, United States of America

E-mail: tom.parker@nist.gov

Received 3 December 2021, revised 14 April 2022
Accepted for publication 26 April 2022
Published 18 May 2022

Abstract

A three-cornered hat analysis has been performed on three independent time transfer systems between the United States Naval Observatory (USNO) and the National Institute of Standards and Technology (NIST). These include (1) a direct Two-Way Satellite Time and Frequency Transfer (TWSTFT) link provided by the USNO, (2) an indirect TWSTFT link, and (3) GPS carrier phase (GPSCP) links. Time transfer instabilities for individual links, as quantified with the time deviation and TOTAL time deviation, have been measured for the first time for averaging periods between 2 h and 85 days. Time transfer instabilities beyond one day are roughly flicker phase in nature and generally range between 40 ps and 120 ps. Frequency transfer instabilities, as quantified by the modified Allan deviation, are observed to be as low as 1×10^{-16} for a 20 day comparison and into the mid 10^{-17} range at longer comparison times.

Keywords: three-cornered hat, time and frequency transfer instabilities, two-way satellite time and frequency transfer, GPS carrier phase time transfer, time deviation and total time deviation

Supplementary material for this article is available online

(Some figures may appear in colour only in the online journal)

1. Introduction

In November 2019 a direct Two-Way Satellite Time and Frequency Transfer (TWSTFT) link, provided by the United States Naval Observatory (USNO), became fully operational between the National Institute of Standards and Technology (NIST) and USNO. This direct link, along with the well-established indirect TWSTFT link between NIST and USNO through the intermediate station at Physikalisch-Technische Bundesanstalt (PTB) in Germany [1], and existing GPS carrier phase (GPSCP) links, provide the unique situation of having three independent time transfer links between NIST and USNO. This allows the calculation of three double difference pairs. A double difference is obtained by subtracting the clock differences between two stations obtained by two different transfer techniques. Therefore, a three-cornered hat (TCH) analysis [2], commonly used to calculate individual clock instabilities, can also be performed using double difference data to obtain previously unavailable information on the long-term instabilities of individual time transfer links.

The instabilities in time and frequency transfer systems are important because they influence a number of significant

* Author to whom any correspondence should be addressed.

performance characteristics. These include: (1) the accuracy of time transfer between laboratories, (2) the calibration interval of transfer links, (3) the stability of clock data supplied to the Bureau International des Poids et Mesures (BIPM) for the calculation of International Atomic Time, TAI, (4) the frequency uncertainties of reports from primary and secondary frequency standards to the BIPM, which impact the frequency accuracy of TAI, and (5), the uncertainty of frequency standard comparisons between widely separated metrology laboratories. The long-term instabilities (months to years) are particularly important in determining how rapidly the uncertainties of time transfer calibrations increase over time, and therefore how often calibrations need to be performed. Time transfer instabilities may also impose multi-month averaging times on remote optical frequency comparisons and yet still be limited to $\sim 10^{-16}$ uncertainty (approximately the uncertainty of cesium primary frequency standards) as the long-distance comparison of reference [3] has shown. Though time and frequency transfer over optical fiber is by far the most stable technique [4], its implementation is not straightforward and not yet realized across oceans. Therefore, most laboratories still rely on TWSTFT [5] and GPSCP [6] for high performance time and/or frequency transfer.

For time intervals less than about 1 day the clock noise of hydrogen masers is typically less than the noise of either TWSTFT or GPSCP [6]. Thus, the short-term instabilities of these two transfer techniques can generally be observed in links between stations with masers or maser ensembles. The instabilities of TWSTFT and GPSCP beyond 1 day are more difficult to evaluate since clock noise usually dominates in this range. Cesium or rubidium fountains and optical frequency standards have better long-term stability than hydrogen masers and can therefore allow better observation of medium-term time and frequency transfer instabilities. With continuous operation of fountains at two laboratories, the time transfer instabilities of TWSTFT or GPSCP may be observable out to intervals of about 20 days. In the future, continuously operated optical frequency standards will increase this interval, but such data is not currently available. Two-way carrier phase has shown very good performance [7], as well as frequency transfer with very long baseline interferometry [8], but these techniques are still in development and not widely utilized.

Another approach to gain information about long-term time transfer instabilities that is not limited by clock noise is to evaluate the double difference between TWSTFT and GPSCP under the assumption that they are independent. This eliminates the clock instabilities. The double difference does not give the instabilities of the individual systems but does show the combined instabilities. TWSTFT, using geostationary communication satellites, and GPSCP are very different systems, but may not be totally independent. Some of the equipment is outside and thus both may see the same variable environment. However, it is worthwhile to note that the delay diurnals seen quite often in TWSTFT are not observed in GPSCP. Thus, it is generally considered a safe assumption that TWSTFT and GPSCP are substantially independent. This is discussed in more detail in section 3.2. For some laboratories a parallel, highly stable, optical fiber link may be available [9] that allows the direct observation of instabilities in TWSTFT or GPSCP, but this is not the case for NIST and many other laboratories, particularly for long baselines and for long time intervals.

In section 2 of this report, we discuss the results of a TCH analysis of the instabilities in two independent TWSTFT links and GPSCP links between NIST and USNO. Three different GPSCP analysis techniques are investigated. In section 3 we compare the link instabilities for both time and frequency and consider the impact of correlations between links. Results are summarized in section 4.

2. Three-cornered hat analysis

For the measurement of UTC(NIST)-UTC(USNO), a baseline of ∼2600 km (∼1600 miles), we now have three independent, high-performance time transfer techniques available. These are: (1) direct TWSTFT (DTW) between NIST and USNO, (2) indirect TWSTFT (ITW) through PTB, and (3) GPSCP. For the GPSCP we have used three different data analysis techniques. These are revised RINEX shift (RRS) [10], TAIPPP [11] and IPPP [12], all using the Novatel[6] GPS receivers identified as 'nist' and 'usn6'. We are not treating the three GPSCP techniques as being independent since they all use the same pair of GPS receivers. One analysis technique used is RRS-rapid (RRS-r), which is an operational system at NIST with a latency of only 2.75 days. A sacrifice of stability of RRS-r compared to RRS is expected due to the low latency. Second, we use TAIPPP data from the BIPM. TAIPPP is available to anyone once a month and is used by many national laboratories. Finally, we use integer ambiguity resolution precise point positioning (IPPP) data following the computation method in [12]. This technique offers improved ambiguity resolution for better long-term stability but is not yet operational because of a present latency of about two weeks. The direct and indirect two-way systems are independent, with different earth stations, modems, phase code chip rates, up and down converters, satellites, etc. The only common elements are UTC(NIST) and UTC(USNO) as the time references which are derived from maser ensembles. The direct two-way link is configured like a typical TWSTFT link using a geostationary satellite. The indirect link uses Telstar-11N[6] at 37.5 degree west. Both TWSTFT links operate at Ku band, and use TimeTech SATRE[6] modems. ITW uses a chip rate of 1 MChip/s and DTW uses 2.5 MChip/s.

We have collected 568 days of data with the new DTW system from early November 2019 to late May 2021 which can be compared to ITW and GPSCP data. An intentional change in the transmission characteristics of the ITW link in May of 2021 provided a convenient stopping point for this phase of the investigation. Time differences for UTC(NIST)—UTC(USNO) using the DTW, ITW, and GPSCP(IPPP) are shown in figure 1 and illustrate a clear difference in short-term stability. Note that only the ITW link is

[6] Commercial products or services are identified for technical completeness. NIST does not endorse any commercial products or services.

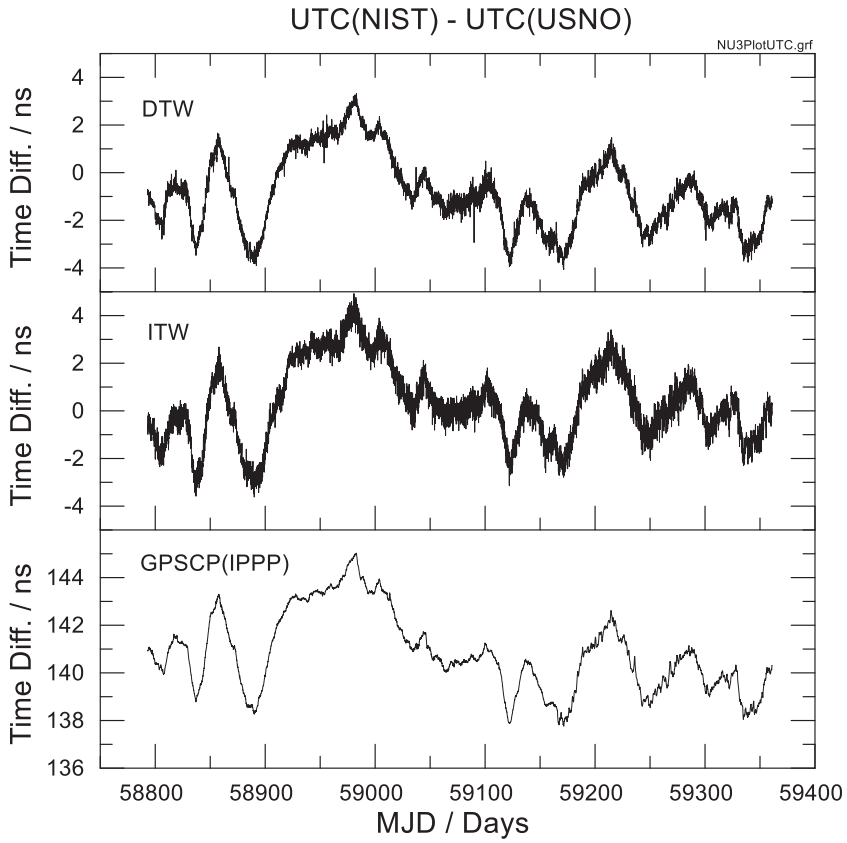

Figure 1. Time differences of UTC(NIST)—UTC(USNO) as a function of modified Julian date (MJD) for three different time transfer links. These include direct two way (DTW), indirect two way (ITW) and GPS carrier phase (GPSCP(IPPP)).

calibrated and therefore the time offsets are not all the same. Only GPSCP(IPPP) data are shown since the curves for the other GPSCP techniques are all very similar on this scale.

2.1. Data processing procedures and TCH confidence limits

In order to perform an accurate TCH analysis, the data for all double differences needs to have identical time tags. Thus, missing data must be filled in and the time tags must be aligned. Details of this process are provided in the supplemental material (https://stacks.iop.org/MET/59/035007/mmedia). These adjustments have been made to all the data in figure 1, as well as the GPSCP RRS-r and TAIPPP data.

Figure 2 shows the time deviations (TDEV) calculated from the data in figure 1 as a function of averaging time τ. TDEV is used because it clearly distinguishes between white, flicker and random walk phase noise and is calculated using the Stable32[6] software [13]. For $\tau < 1 \times 10^5$ s transfer noise dominates and the link noise varies significantly for the different transfer techniques. The GPSCP(IPPP) noise is close to the scale noise and exhibits the same white FM (random walk phase) noise characteristic of a typical clock. Stability curves for RRS-r and TAIPPP are not shown because they are similar to the IPPP curve. Confidence limit error bars are 1σ. The process of filling in missing data and aligning time tags changed the TDEV values for UTC(NIST)—UTC(USNO) by less than an average of 2% from the original unaltered data sets (except for outlier removal). Thus, cleaning up the data for the three-corned hat analysis had a minimal impact on the instability statistics for the various links.

We now consider TCH confidence limits. Accurate values of the TCH confidence limit error bars are not available at this time because they have not been theoretically calculated for TDEV or TOTAL TDEV (which is used below). Confidence limits in a TCH calculation are complicated because they depend on the relative noise levels of the three entities being examined [14]. If the noise of one of the transfer techniques dominates, the error bars for this technique will be nearly the same in a fractional sense as in figure 2. However, if the noise of one of the techniques is much smaller than the other two, as is the case for GPSCP(IPPP) at small τ, the error bars can be much larger because the effective number of degrees of freedom (EDOF) is reduced significantly [14]. Though confidence limits have not been calculated for TDEV or TOTAL TDEV in a TCH analysis, they have been evaluated for the Allan deviation (ADEV) and the TOTAL ADEV [14, 15]. Since TDEV and ADEV are derived from related two sample variances, the ADEV analysis can provide some guidance on how the TDEV and TOTAL TDEV confidence limit error bars are influenced by relative noise levels. The analysis in [15] shows that confidence limit error bars can get quite large for the entity with the lowest noise level. When the degrees of freedom (DOF) are low, the chi-square statistical distribution of a TDEV (or ADEV) calculation is asymmetric and skewed in the direction of making it more probable that a calculated TDEV

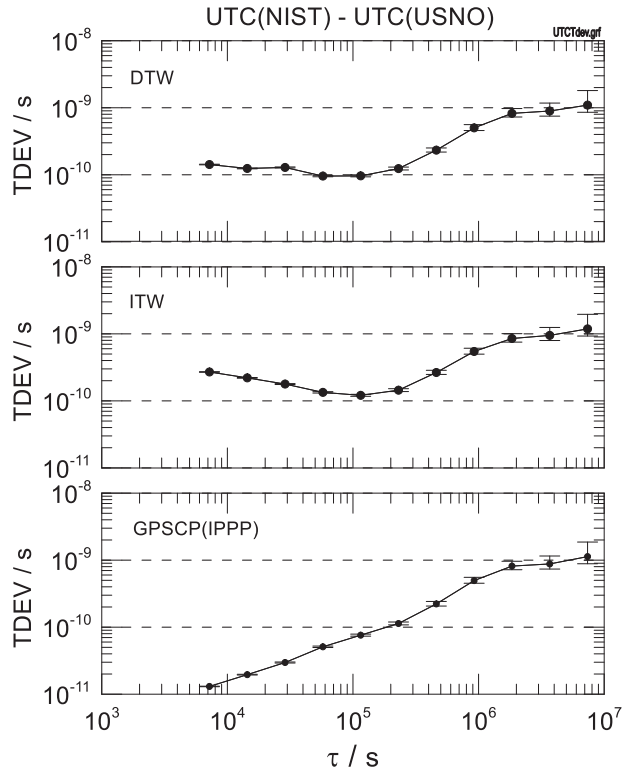

Figure 2. Time deviation (TDEV) of the data in figure 1 for direct two-way in the top curve, indirect two-way in the middle curve, and GPS carrier phase (IPPP) in the bottom curve.

will be lower than the true value. In a TCH analysis this is particularly true for the link with the lowest level of instabilities. The improved confidence limits obtained with TOTAL TDEV help mitigate to a small extent the impact of the TCH analysis.

The TCH analysis is sensitive to correlations between the clocks or links being analyzed, particularly for the lower noise entity. This is discussed in more detail in section 3.2.

The TCH analysis for DTW, ITW and GPSCP(IPPP) is discussed in section 2.2 below and details of the similar analyses for GPSCP(TAIPPP) and GPSCP(RRS-r) are presented in the supplementary material.

2.2. DTW, ITW and GPSCP(IPPP)

The time differences, $\delta\Delta$, for the three double difference combinations of ITW-DTW, ITW-GPSCP(IPPP) and DTW-GPSCP(IPPP) are shown in figure 3 where it is clear that the clock instabilities have been removed. The fact that there are three double difference pairs can reveal the sources of some of the structure simply by inspection of the data. For example, figure 3 shows what appears to be a slow upward step in ITW around MJD 58920. There is also long-term drift in the double differences that ranges from $+1.5$ ps day^{-1} to $+3.9$ ps day^{-1}.

Figure 4 shows TOTAL TDEV calculated with Stable32 for each of the three double difference pairs. The confidence limit error bars are 1σ. TOTAL TDEV is used in order to get improved confidence limits and is equivalent to TDEV at all but the largest τ [16]. It also provides data at larger τ. As can be seen, the TOTAL TDEV curves are fairly flat (roughly flicker phase in nature over most of the range) with values ranging from 50 ps to 300 ps. The combination of DTW and GPSCP(IPPP) shows the lowest level of instabilities. The double difference serves as an upper bound on individual link instabilities. Because there is some missing data from all three transfer techniques the cancellation of the clock noise is not perfect. However, because the intervals of missing data are all equal to, or less than, 1 day, it is only the short-term clock instabilities that are not completely cancelled by the double difference technique. Fortunately, it is in this region that we have knowledge independent of the double differences regarding the instabilities of the transfer techniques as illustrated in figure 2.

A TCH analysis has been performed to extract TOTAL TDEV for the individual time transfer techniques and the results are shown in figure 5. Both octave and decade intervals are used. Confidence limits for the TCH results will be discussed and added in section 3. As noted earlier regarding figure 2, the time scales at NIST and USNO are quiet enough at $\tau < 1$ day for the observed instabilities to be mostly from the time transfer processes, particularly for the TWSTFT links. TDEVs for the four lowest τ values from figure 2 are shown in figure 5 as $+$ and x symbols and identified as 'scale-to- scale'. Beyond one day the clock noise obscures the transfer noise in figure 2 and these TDEV values are not shown. The TOTAL TDEV values from the TCH analysis are shown as dots, triangles and open circles and are the only measure of the transfer instabilities for $\tau \geqslant 72\,000$ s.

In figure 5 we see that the 'scale-to-scale' values agree fairly well with the TCH results for ITW and DTW at the smallest τ. However, there is some disagreement between these at the third and fourth smallest τ values. This is due to a correlated diurnal and is discussed in more detail in section 3.2. The large scatter and missing point (negative variance) in the TCH GPSCP(IPPP) results at small τ reflect the poor confidence limits for this data due to the very low relative noise level. Here the DOF are reduced as much as a factor of 20 000. Fortunately, the scale-to-scale values are available and are a better measure.

The higher ITW noise level at small τ compared to DTW is largely due to the lower chip rate of the ITW link. However, there are other factors that can affect the TOTAL TDEVs at small τ. For example, ITW has the longest effective baseline that involves an intermediate station (NIST to PTB to USNO), has lower dish antenna elevations causing increased ionospheric propagation systematics, and a smaller data interval than DTW (see supplemental material). DTW and ITW are decreasing as white PM at small τ, while GPSCP(IPPP) shows a very different characteristic in that the noise increases with τ.

For $\tau > 1$ day ITW has the highest instabilities and is relatively flat, with an average value near 90 ps. GPSCP(IPPP) is also relatively flat at a level of about 50 ps out to $\tau = 4 \times 10^6$ s. The drop off beyond this reflects the reduced DOF from the lower instability level. DTW has a missing point due to a negative variance at $\tau = 2 \times 10^6$ s and shows a very low level of instability around $\tau = 1 \times 10^6$ s that very likely is not real.

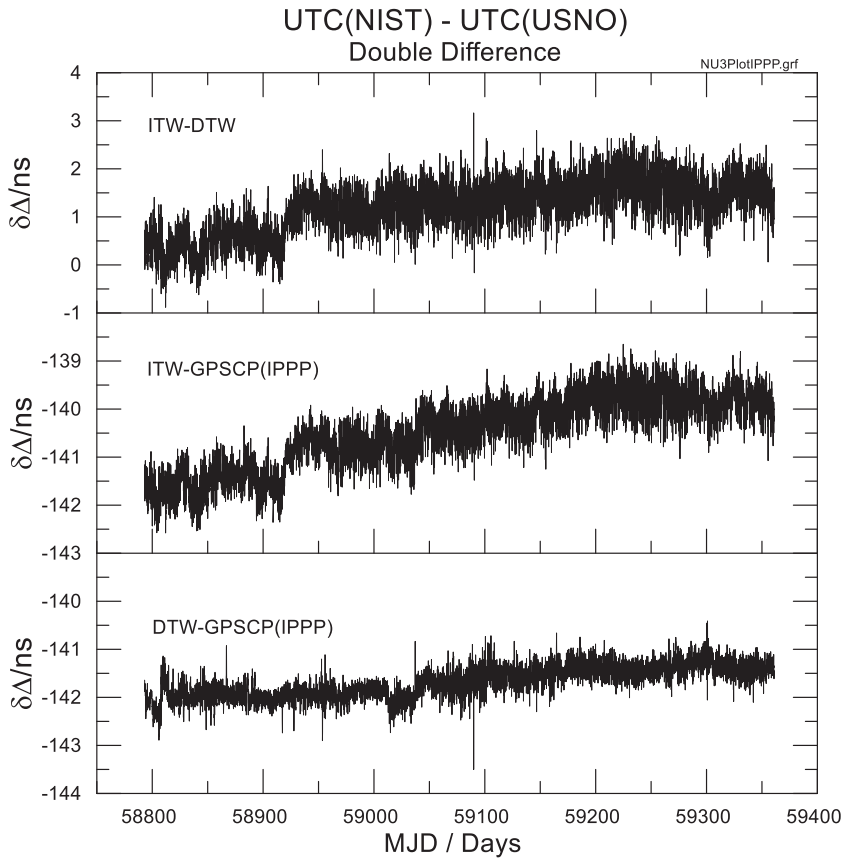

Figure 3. Time difference, $\delta\Delta$, as a function of MJD for the three different double difference combinations used for the TCH analysis with GPSCP(IPPP).

This low level is not consistent with the results shown below using TAIPPP and RRS-r.

2.3. Results for GPSCP(TAIPPP) and GPSCP(RRS-r)

A similar TCH analysis has also been performed for GPSCP(TAIPPP) and GPSCPRRS-r) and the details are presented in the supplemental material. Here we show the TCH results in figures 6 and 7. In addition to the TAIPPP and RRS-r curves, we also see more results for ITW and DTW. The ITW curves are similar to that in figure 5, but DTW shows more variation.

3. Comparison of links and link correlations

To better compare the instabilities in the various links figure 8 shows selected curves from figures 5–7. The ITW plot is a straight unweighted average of the three ITW curves. The DTW plot is also a similar average and does include the unusually low values in figures 5 and 6 under the assumption that they are random fluctuations due to the large confidence limits. The three GPSCP plots come from the corresponding TCH figures. For clarity the two-way curves are shown in figure 8(a) and the GPS curves in figure 8(b). With all five curves the scale-to-scale TDEVs are used for the four smallest τ values rather than the TCH results because they are a better measure. Estimated confidence limits are shown and are based on the EDOF calculated by the method of [14] in combination with TOTAL TDEV confidence limits as shown in figure 4. Only data with an EDOF greater than 3 are shown which limits the DTW and GPSCP(IPPP) curves to $\tau < 2 \times 10^6$ and the ITW, GPSCP(RRS-r) and GPSCP(TAIPPP) curves to $\tau < 1 \times 10^7$. In calculating the EDOF of ITW, TAIPPP and RRS-r the DOF were reduced an average of a factor of two and an average of a factor of ten for DTW and IPPP. Even though the TWSTFT curves are obtained from an average of three TCH analyses, the confidence limits have not been reduced. These results are not totally independent since the same ITW and DTW data were used in all three.

3.1. Comparison of links

All of the GPSCP links are much quieter than the TWSTFT links at small τ, though RRS-r and TAIPPP have higher noise levels than IPPP. The TWSTFT link noise averages down roughly as white phase, while the GPSCP link noise tends to increase as white FM (random walk phase). Some of this rise may be caused by clock noise, but there are at least two other sources of random walk PM in GPSCP. One is troposphere residuals for τ in the range of a few hours, the other is the effect of code errors on ambiguity resolution (1 day and more). However, the latter is specifically addressed by IPPP, which explains the better long-term performance of this technique. For $\tau \leqslant 5.76 \times 10^4$ s (16 h) the instabilities are a combination of clock and transfer noise while for $\tau \geqslant 7.2 \times 10^4$ s (20 h) the

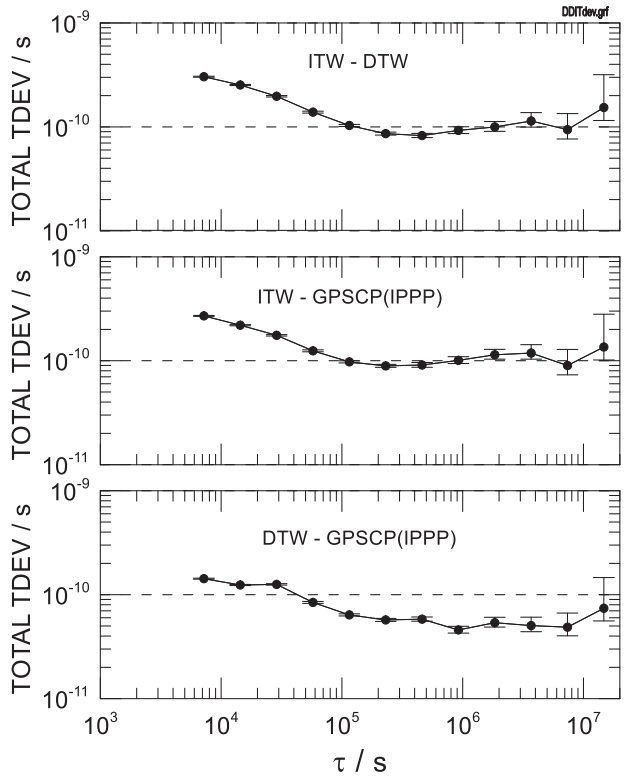

Figure 4. TOTAL TDEV plots of the double difference data in figure 3. These curves provide an upper bound on individual link instabilities.

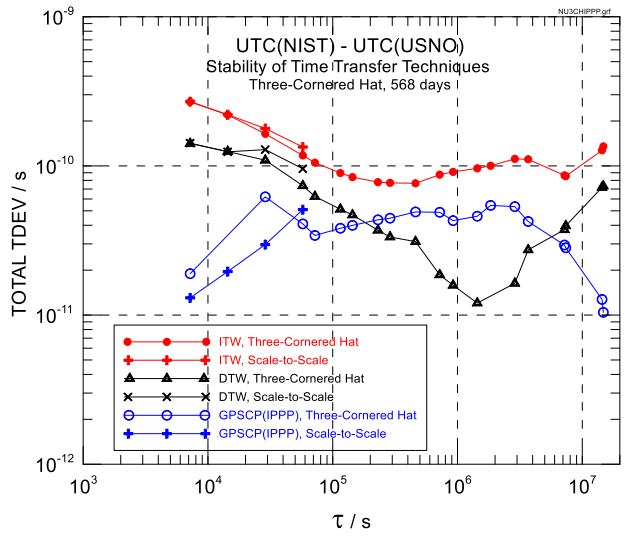

Figure 5. Three-cornered hat results for the indirect two-way (ITW), direct two way (DTW) and GPS carrier phase using IPPP (GPSCP(IPPP)) data in figure 3.

instabilities are only from transfer noise. This information, in combination with the fact that both the clock noise and GPSCP transfer noise are white FM in nature, allows us to estimate that the scale-to-scale clock noise is in the range of 20 to 40 ps at 16 h. However, the discrepancies between TCH and scale-to-scale noise levels observed in figures 5–7 for DTW and ITW at 8 and 16 h are too large to be consistent with this level of clock noise. These discrepancies are due to the presence of correlated diurnals in DTW and ITW.

By $\tau = 1$ day, the five curves have converged considerably and remain within about a factor of three at a level that is largely at or below 100 ps all the way out to τ of about 20 days (1.7×10^6 s). The RRS-r and TAIPPP GPSCP links tend to slowly increase in instability as τ increases, while the TWSTFT and IPPP links show less of a rise in instability beyond one day, though all are roughly flicker phase in nature. Beyond 4×10^6 s some of the links show evidence of increasing instabilities. This TCH analysis has provided a good measure of the long-term instabilities in ITW, GPSCP(TAIPPP) and GPSCP(RRS-r) out to 85 days. Beyond about two days the DTW and GPSCP(IPPP) links have the lowest level of instabilities and consequently have the largest error bars. Both DTW and GPSCP(IPPP) exhibit instabilities near 50 ps at $\tau = 1 \times 10^6$ s. The very low DTW TDEV values in figures 5 and 6 are inconsistent with the confidence limits in figure 8 and may be due to a correlation between DTW and ITW as discussed in section 3.2. With the exception of DTW, the scatter in the link stability data is consistent with the estimated confidence limits. Note that the double difference curve for DTW—GPSCP(IPPP) in figure 4 provides an upper limit for both DTW and IPPP with a high degree of confidence that is close to 60 ps all the way out to 10^7 s. The double difference curve is also consistent with the DTW and GPSCP(IPPP) values from the TCH analyses.

At this time, it is not clear where the sources of instabilities in any of the transfer techniques at $\tau > 1$ day originate. There are certainly environmental factors that have long-term fluctuations and there are annual variations. These could be contributing to the observed delay fluctuations. For GPSCP TAIPPP and RRS-r there is also the issue of ambiguity resolution. There is evidence of long-term monotonic time difference drift in the double differences that might be caused by aging in the electronics, which could also be contributing to the long-term fluctuations. The observed drift rates of up to almost 4 ps day^{-1} in the double difference data, combined with the evidence of increasing noise levels beyond 100 days, suggests that link calibrations should occur on the order of once a year in order to maintain an accuracy near 1 ns.

To relate the TOTAL TDEV data to frequency transfer uncertainty (FTU) [17], a simple relation exists to calculate the TOTAL modified Allan deviation (TOTAL MDEV) from TOTAL TDEV.

$$\text{TOTAL MDEV} = \sqrt{3}(\text{TOTAL TDEV})/\tau. \qquad (1)$$

This relation also applies to MDEV and TDEV. Figure 9 shows TOTAL MDEV as calculated from the data in figure 8. Confidence limits are not shown for clarity, but they are the same as in figure 8. To use MDEV as an estimate of FTU it is necessary to take into account that for a given comparison length T, the relevant τ value is $T/2$ due to the phase averaging of the MDEV process. For example, for a 30 day comparison of frequency standards, the relevant τ value is 15 days. It has been confirmed with simulated flicker phase noise for the range of τ values in this paper that MDEV at $T/2$ is a good estimate for the FTU. Thus, TOTAL MDEV can be viewed as

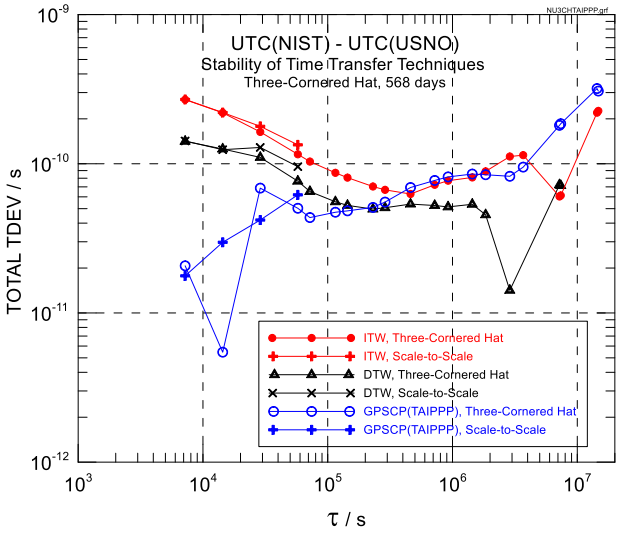

Figure 6. Three-cornered hat results for the indirect two-way (ITW), direct two way (DTW) and GPS carrier phase using TAIPPP (GPSCP(TAIPPP)).

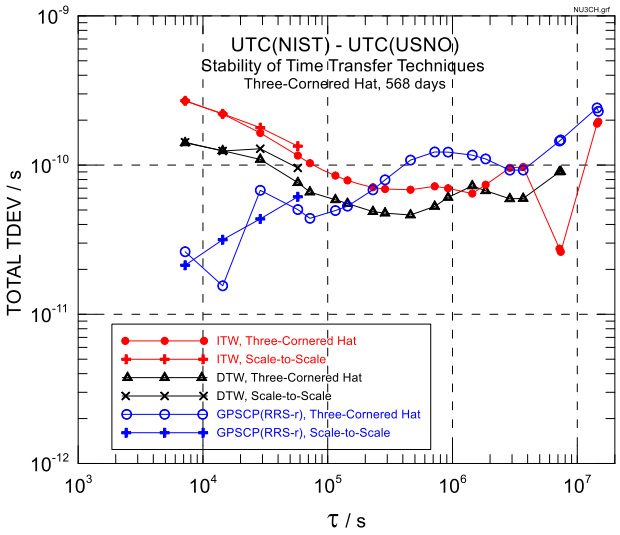

Figure 7. Three-cornered hat results for the indirect two-way (ITW), direct two way (DTW) and GPS carrier phase using revised RINEX shift rapid, (GPSCP(RRS-r)).

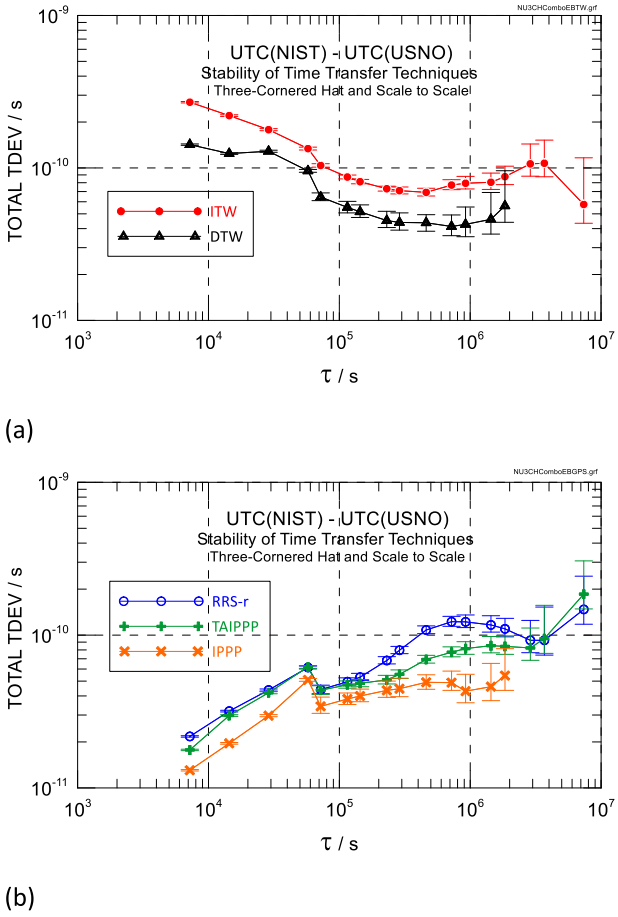

Figure 8. Selected curves from figures 5–7 with estimated confidence limits. The ITW and DTW curves in (a) are the unweighted averages of the three curves in figures 5–7. Scale-to-scale data are used for the four smallest τ values. TCH data are shown only for τ values where the estimated degrees of freedom are greater than 3.

an estimate of the FTU if used properly. The data in figure 9 shows that MDEV at or below 1×10^{-16} at $\tau = 1 \times 10^6$ s (∼24 day comparison interval) can be achieved with DTW and GPSCP(IPPP) and reaches the mid 10^{-17} range at 85 days for the other links. MDEV in figure 8 can be approximated for DTW and GPSCP(IPPP) to +/− 20% over the range of 1 to 20 days by the relation $\mathrm{MDEV} = 7.7 \times 10^{-17}(\tau/\tau_R)^{-1.0}$ for $\tau_R = 1 \times 10^6$ seconds. The upper limit for DTW and GPSCP(IPPP) is in the low 10^{-17} range at 100 days as determined from double difference data. Just how well MDEV works as an estimate of FTU depends on the noise type and the τ/τ_o ratio, where τ_o is the minimum τ [18]. Though MDEV is used here because it is easily derived from TDEV, ADEV (with or without a small amount of pre-averaging) is a better estimate of the true FTU, though it is generally biased high by about 15% [17].

The GPSCP(IPPP) data here shows an instability level that is comparable to, though likely a little higher, than that observed in [9] for GPSCP(IPPP) compared to optical fiber links in the range of 10^5 to 10^6 s. This higher level may be due to the longer baseline or the particular receivers that are used.

It must also be noted that the long-term linear systematic time drifts observed in figure 3 constitute a frequency offset ranging from 0.6×10^{-17} to 4.5×10^{-17}. These long-term drift rates are for double differences and not individual links. We cannot determine the drift of individual links from this data or the clock-to-clock data in figure 1. Thus, to be conservative, these values would have to be considered as a frequency bias, or a type B systematic frequency uncertainty, in a comparison of frequency standards. A regular schedule of link calibrations could help determine the drift in individual links.

3.2. Correlations

Correlations between the instabilities in the links (or clocks) can cause errors in a TCH analysis, including negative variances [19, 20]. They will of course also affect any double difference data. In the DTW and ITW links there is evidence of

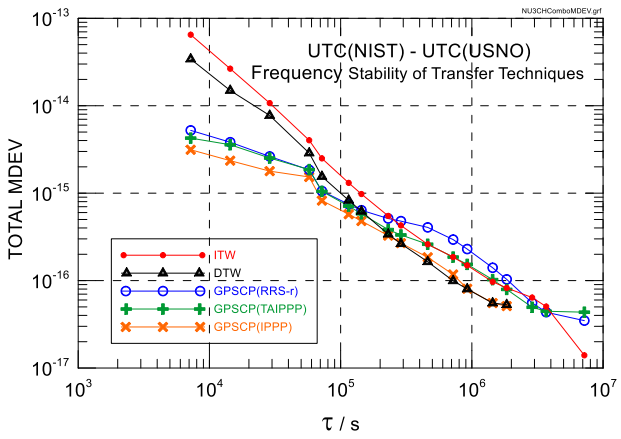

Figure 9. TOTAL modified Allan deviation, TOTAL MDEV, calculated from the data in figure 8.

a diurnal variation, probably caused by environmental instabilities. The diurnals are barely visible with TDEV, but are quite pronounced in a power spectral density analysis. Careful examination of the daily data shows that there is a positive correlation in the time differences with the DTW link leading the ITW link by about 3 h. This correlation results in the small discrepancy in TDEV values between the scale-to-scale and TCH data at 8 and 16 h seen in figures 5–7. The magnitude of the diurnal is reduced in the TCH analysis. Whatever influence that causes the diurnal may also exist at longer averaging times and this may be responsible for the abnormally low DTW TOTAL TDEV values at large τ. Any evidence of a diurnal in GPSCP is at least 20 dB lower based on a power spectral density analysis. We have seen no evidence of a significant correlation between TWSTFT and GPSCP. An analysis of covariance terms will be the subject of a future investigation. Two areas of concern are the impact of long-term environmental variations and atmospheric instabilities on the various links. Including an analysis of covariance terms can improve the confidence limits and eliminate negative variances [19, 20].

An interesting question is whether GPSCP links using different receiver pairs and possibly different data processing techniques are sufficiently independent that they could be used as different legs in a TCH analysis. This will require a detailed look at covariance terms to assess the level of correlation. Redundant GPSCP links are considerably easier to implement than redundant two-way links.

4. Summary

The unique situation of having three independent time transfer links between NIST and USNO has enabled the use of the three-cornered-hat technique to determine the level of instabilities of the individual links. This investigation has provided the best long-term stability data on individual ITW, GPSCP(TAIPPP) and GPSCP(RRS-r) time transfer links that has ever been achieved, with stability data out to 85 days. Though the NIST/USNO TWSTFT and GPSCP links have very different levels of instabilities at small τ, ranging from 13 ps with GPSCP(IPPP) to nearly 270 ps for ITW, we have observed that by $\tau = 1$ day they have converged significantly. Beyond one day GPSCP is similar to TWSTFT with TOTAL TDEVs ranging from 40 ps to 120 ps for τ up to about 20 days and beyond and both are roughly flicker phase in nature. The GPSCP method of analysis makes a difference, with IPPP showing the lowest instabilities at all averaging periods, though it currently has the longest latency as long as rapid integer products are not available. Overall, DTW and GPSCP(IPPP) have the lowest instabilities at averaging intervals longer than 1 day, with values in the 40 ps to 60 ps range, though they have larger confidence limit error bars and the data is limited to 21 days in order to provide confidence limits. For frequency comparisons, GPSCP(IPPP) and DTW can provide frequency transfer uncertainties near 1×10^{-16} at 20 days and ITW, GPSCP(TAIPPP) and GPSCP(RRS-r) in the mid parts in 10^{-17} at 85 days.

The fact that the instabilities in TWSTFT and GPSCP are similar for $\tau > 1$ day makes the interpretation of double difference data easier if that is all that is available. One can simply divide the double difference by the square root of two and get a reasonable estimate of the individual link noise. It would be very useful to obtain similar stability information for other station pairs, in particular links with longer baselines.

Acknowledgments

The authors thank Jeff Sherman, Mike Lombardi, Biju Patla, Judah Levine and Dave Howe for valuable discussions.

ORCID iDs

T E Parker https://orcid.org/0000-0003-3264-9245
V Zhang https://orcid.org/0000-0002-8061-5157
G Petit https://orcid.org/0000-0002-6837-3052
J Yao https://orcid.org/0000-0003-1702-4757
R C Brown https://orcid.org/0000-0002-8228-4283

References

[1] Jiang Z *et al* 2019 Improving two-way satellite time and frequency transfer with redundant links for UTC generation *Metrologia* **56** 025005
[2] Gray J E and Allan D W 1974 A method for estimating the frequency stability of an individual oscillator *Proc. IEEE 28th Annual Symp. Freq. Control* pp 243–6
[3] McGrew W F *et al* 2019 Towards the optical second: verifying optical clocks at the SI limit *Optica* **6** 448–54
[4] Predehl K *et al* 2012 A 920 km optical fiber link for frequency metrology at the 19th decimal place *Science* **336** 441–4
[5] Kirchner D 1999 Two-way satellite time and frequency transfer (TWSTFT): principle, implementation, and current performance *Review of Radio Science 1996-1999* (Oxford: Oxford University Press) pp 27–44
[6] Larson K, Levine J, Nelson L and Parker T 2000 Assessment of GPS carrier-phase stability for time-transfer applications *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **47** 484–94

[7] Fujieda M *et al* 2018 Advanced satellite-based frequency transfer at the 10^{-16} level *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **65** 973–8
[8] Pizzocaro M *et al* 2021 Intercontinental comparison of optical atomic clocks through very long baseline interferometry *Nat. Phys.* **17** 223–7
[9] Petit G 2021 Sub-10^{-16} accuracy GNSS frequency transfer with IPPP *GPS Solutions* **25** 22
[10] Yao J *et al* 2014 An improvement of RINEX-shift algorithm for continuous GPS carrier-phase time transfer *Proc. 27th Institute of Navigation GNSS+ Conf.* pp 1253–60
[11] Petit G and Jiang Z 2008 Precise point positioning for TAI computation *IJNO* **2008** 1–8
[12] Petit G, Kanj A, Loyer S, Delporte J, Mercier F and Perosanz F 2015 1×10^{-16} frequency transfer by GPS PPP with integer ambiguity resolution *Metrologia* **52** 301–9
[13] Riley W J 2020 www.stable32.com Stable32 Software
[14] Ekstrom C and Koppang P 2006 Error bars for the three-cornered hats *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **53** 876–9
[15] Lantz E, Calosso C, Rubiola E, Giordano V, Fluhr C, Dubois B and Vernotte F 2019 KLTS: a rigorous method to compute the confidence intervals for the three-cornered hat and for Groslambert covariance *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **66** 1942–9
[16] Weiss M and Howe D TOTAL TDEV *Proc. 1998 IEEE Int. Frequency Control Symp.* pp 192–8
[17] Panfilo G and Parker T E 2010 A theoretical and experimental analysis of frequency transfer uncertainty, including frequency transfer into TAI *Metrologia* **47** 552–60
[18] Lesage P and Ayi T 1984 Characterization of frequency stability: analysis of the modified Allan variance and properties of its estimate *IEEE Trans. Instrum. Meas.* **IM-33** 332–6
[19] Torcaso F, Ekstrom C R, Burt E A and Matsakis D N 2000 Estimating the stability of *N* clocks with correlations *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **47** 1183–9
[20] Tavella P and Premoli A 1994 Estimating the instabilities of *N* clocks by measuring differences of their readings *Metrologia* **30** 479–86

SUPPLEMENTAL MATERIAL

1 – Data Processing Details

The TWSTFT data are taken at two-hour intervals, but DTW and ITW sessions are 3.5 minutes apart. The ITW and DTW sessions are respectively two-minute and four-minute averages. The GPSCP data is recorded at 5-minute intervals and is processed as two-hour averages (straight average of 24 consecutive points). There are some missing data in all five techniques. In order to perform an accurate three-cornered hat analysis, the data for all double differences should have identical time tags. Thus, missing data must be filled in and the time tags must be aligned. The ITW time tags are used as reference time tags and any missing ITW data was filled in by taking data from the previous or following days at the same time of day and randomizing it by ~0.05 ns. This reduces the impact of missing data on the statistics of the instabilities. 2.0 % of the ITW data is missing. The DTW data are interpolated to match the ITW time tags. 2.3 % of the DTW data is missing and is filled in the same manner as with the ITW data. The 2-hour average GPSCP data is also interpolated to match the ITW time tags. Missing GPSCP data ranged from 0.48% to 0.58% and all gaps were filled in by interpolation. No unexplained time steps were removed from any of the data. Outliers for individual data points larger than 2.5 ns have been removed. One outlier was removed from the DTW data and two were removed from the ITW data. About one days' worth of large outliers had to be removed from the GPSCP(RRS-r) data and this contributes to the RRS-r missing data. One outlier was removed from the TAIPPP data and no outliers were removed from the IPPP data. All intervals of missing data for all five links were less than or equal to one day.

2 - Details of the TCH analysis of GPSCP(RRS-r) and GPSCP(TAIPPP).

2.1 - DTW, ITW and GPSCP(RRS-r)

The time differences for the three double difference combinations of ITW-DTW, ITW-GPSCP(RRS-r) and DTW-GPSCP(RRS-r) are shown in Fig. 1. The fact that there are three double difference pairs reveals the sources of some of the structure simply by inspection of the data. For example, between MJD 58810 and 58850 there are some periodic fluctuations in the top and middle plots that very likely come from the ITW link. There is a spike at MJD 58910 and structure from MJD 59090 to 59170 that originates in the GPSCP(RRS-r) data. There is also long-term drift in the data, with ITW-DTW averaging +2.4 ps/day, ITW-GPSCP(RRS-r) showing +3.1 ps/day and DTW-GPSCP(RRS) averaging +0.7 ps/day.

Figure 2 shows the TOTAL TDEV values calculated with Stable32 for each of the three double difference pairs. TOTAL TDEV is used in order to get the best confidence intervals at large τ, and is equivalent to TDEV at all but the largest τ. The confidence interval error bars are 1 σ. As can be seen, the TOTAL TDEV curves are fairly flat (roughly flicker phase in nature over most of the range) with values ranging from 80 ps to 300 ps. The error bars at the largest τ value of 1.47×10^7 s are +107% and -25%. Because there is some missing data from all three transfer techniques the cancellation of the clock noise is not perfect. However, because the intervals of missing data are all equal to, or less than, 1 day, it is only the short-term clock instabilities that are not completely cancelled by the double difference technique. Fortunately, it is in this region ($\tau < 1$ day) that we have knowledge independent of the double differences regarding the instabilities of the transfer techniques.

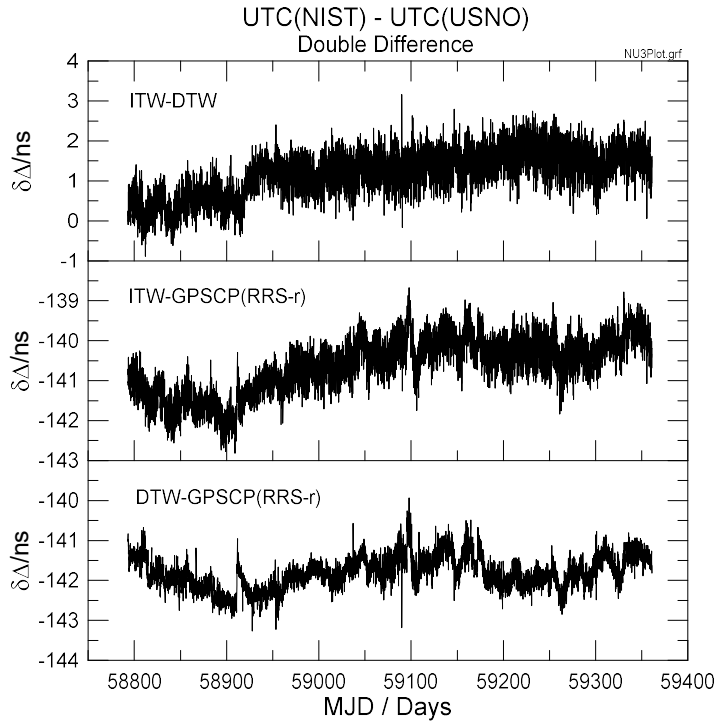

Figure 1. Time difference, $\delta\Delta$, as a function of MJD for the three different double difference combinations used for the three-cornered hat analysis using GPSCP(RRS-r).

A three-cornered hat analysis has been performed to extract TOTAL TDEV for the individual time transfer techniques. Figure 3 shows these TOTAL TDEV values for the three-time transfer techniques used in the double differences in Fig. 1. For this analysis, both octave and decade intervals are used. As noted in the main paper, the time scales at NIST and USNO are quiet enough at $\tau < 1$ day for the observed instabilities to be mostly from the time transfer processes, particularly for the TWSTFT links. The TDEV values for the two TWSTFT links for the four lowest τ values are shown in Fig. 3 as + and x symbols and identified as "scale-to-scale". TDEV is also calculated for GPSCP(RRS-r) and shown as "scale-to-scale" in Fig.3. Beyond one day the clock noise obscures the transfer noise and these TDEV values are not shown. The TOTAL TDEV values from the TCH analysis are shown as dots, triangles and open circles and are the only measure of the transfer instabilities for $\tau > 1$ day.

In Fig. 3 we see that the "scale-to-scale" values agree fairly well with the TCH results for ITW and DTW at the smallest τ. However, there is some disagreement between these at the third and fourth smallest τ values. The large scatter in the TCH GPSCP(RRS-r) results at small τ reflect the poor confidence limits for this data. Fortunately, the scale-to-scale values are a better estimate at small τ.

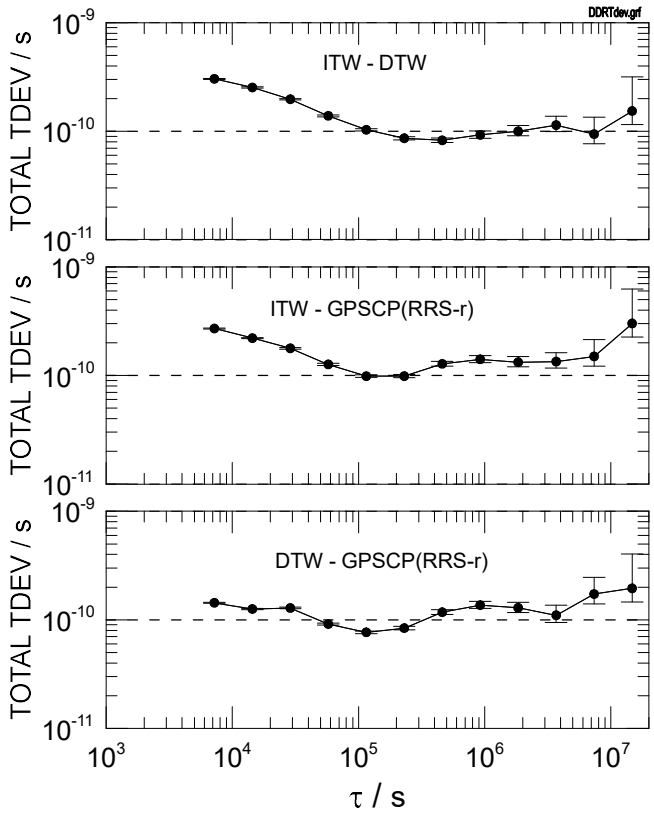

Figure 2. TOTAL TDEV plots for the double difference data in Fig. 1.

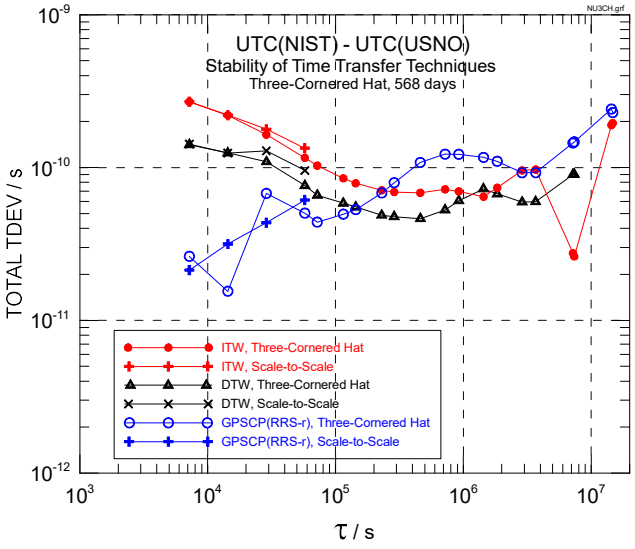

Figure 3. Three-cornered hat results for the Indirect Two-Way (ITW), Direct Two Way (DTW) and GPS Carrier Phase using Revised RINEX Shift rapid, (GPSCP(RRS-r)).

The curves in Fig. 3 show the noise characteristics of the individual transfer techniques. At $\tau < 1$ day ITW has the highest level of instabilities with TOTAL TDEV = 270 ps at τ = 7200 s. The instability of DTW at this

τ is 140 ps and for GPSCP(RRS-r) is 21 ps. The higher noise level with ITW as compared to DTW is largely due to the lower chip rate of the ITW link. The ITW and DTW instabilities average down roughly as white phase noise at small τ. Also, both ITW and DTW show some evidence of a weak diurnal, possibly due to temperature (this is discussed in Section 3.2 of the main paper). Any evidence of a diurnal in GPSCP is at least 20 dB lower based on a power spectral density analysis. However, GPSCP does show a very different characteristic at small τ in that the noise increases with τ like white FM (equivalent to random walk phase noise).

When τ has increased to $1x10^5$ s (~ 1.2 days) the TOTAL TDEV values for all three techniques have converged to a range of 50 ps to 85 ps and remain relatively flat to beyond $1x10^7$ s. At $3x10^6$ s (~ 35 days) the TOTAL TDEV values are still in the range of 60 ps to 97 ps. Generally, DTW has the lowest level of instabilities beyond $\tau = 1x10^5$ at ~ 60 ps, while ITW is closer to 80 ps. The missing DTW data near $\tau = 1.5x10^7$ s are due to negative variances and the unrealistically small values in the ITW data near $\tau = 8x10^6$ s are due to poor confidence limits. TOTAL TDEV values for GPSCP(RRS-r) are closer to 100 ps. All three transfer techniques exhibit roughly flicker phase noise characteristics beyond $\tau = 1x10^5$ s, though GPSCP(RRS-r) shows some increase. The data also suggests that TOTAL TDEV for all three links is increasing at $\tau > 1x10^7$ s.

2.2 - DTW, ITW and GPSCP(TAIPPP)

The time differences for the three double difference combinations of ITW-DTW, ITW-GPSCP(TAIPPP) and DTW-GPSCP(TAIPPP) are shown in Fig. 4. The plot for ITW-DTW is the same as in Fig. 1 and is shown for ease in comparison. Some of the structures observed in Fig. 1 are also observed in Fig. 4. The periodic fluctuations between MJD 58810 and 58850 from the ITW link are present in both figures. What is not seen at the same level in Fig. 4 is the spike at MJD 58910 and the structure from MJD 59090 to 59170 that is present in the bottom two curves of Fig. 1. The GPSCP(TAIPPP) data is more stable. At MJD 58989 there is a small, instantaneous time step in the GPSCP(TAIPPP) curve due to one day of missing data that is not present in the RRS-r curve. As in Fig. 1 there is also some drift in the data in Fig. 4, with ITW-DTW averaging +2.4 ps/day (the same as in Fig.1), ITW-GPSCP(TAIPPP) showing +2.9 ps/day and DTW-GPSCP(TAIPPP) averaging +0.5 ps/day.

Figure 5 shows the TOTAL TDEV curves for each of the three double difference pairs in Fig. 4 and they are similar to those in Fig. 2. The top curve is exactly the same, and is shown for ease in comparison to the lower two curves. At the largest τ value the TOTAL TDEVs are little larger in the middle and bottom curves than in Fig. 2, but are lower in the range of $\tau = 1x10^5$ s to $\tau = 1x10^6$ s. The differences at the largest τ are not statistically significant.

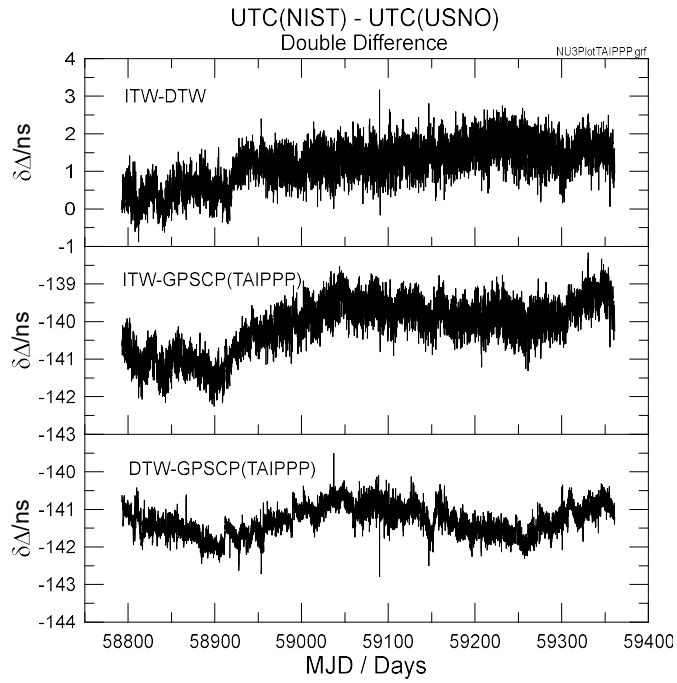

Figure 4. Time difference, $\delta\Delta$, as a function of MJD for the three different double difference combinations used for the three-cornered hat analysis using GPSCP(TAIPPP).

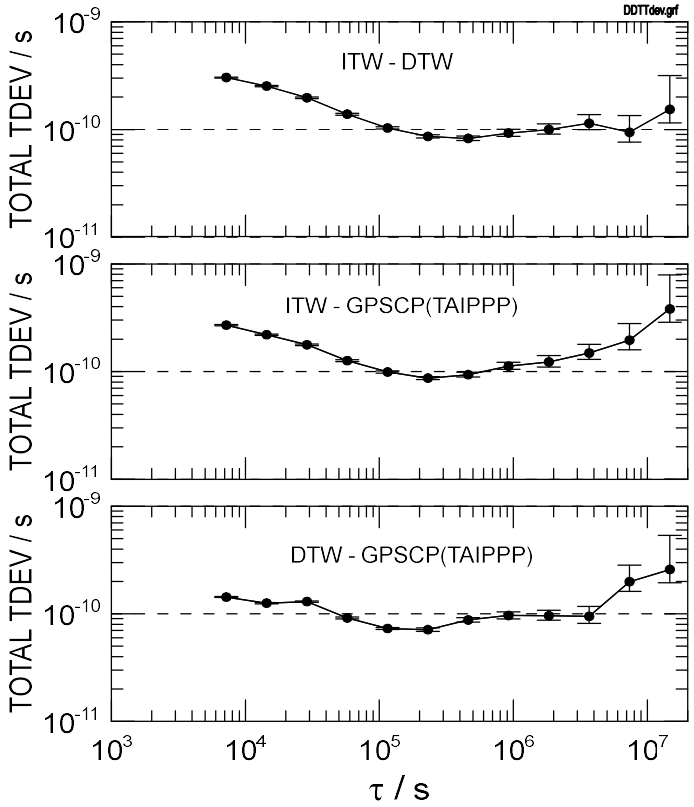

Figure 5. TOTAL TDEV plots of the double difference data in Fig. 4.

Figure 6 shows the three-cornered hat results for the DTW, ITW and GPDCP(TAIPPP) links. For $\tau < 1.5 \times 10^5$ s the results are nearly the same as those in Fig. 3. For $\tau > 2 \times 10^5$ s the ITW curves in Fig. 3 and Fig. 6 are also very similar, with values ranging from 63 ps at $\tau = 4.6 \times 10^5$ s to 114 ps at $\tau = 3.7 \times 10^6$ s in Fig. 6. The GPSCP(TAIPPP) curve in Fig. 6 shows a generally lower TOTAL TDEV than the GPSCP(RRS-r) curve in Fig. 3 in the range of 2×10^5 s $< \tau < 3.7 \times 10^6$ s. Beyond $\tau = 3.7 \times 10^6$ s the GPSCP(TAIPPP) data is a little higher than the GPSCP(RRS-r) values, though this may be within the range of the confidence limits. For $\tau > 2 \times 10^5$ s the TOTAL TDEV values for both GPSCP versions in Figs. 3 and 6 range from 51 ps to 320 ps. The TOTAL TDEV curves for DTW are very similar out to $\tau = 1 \times 10^6$ s in Figs. 3 and 6. However, beyond this point the values are lower or missing (due to negative variances) in Fig. 6. The discrepancies in TOTAL TDEV values in this region are related to poor confidence limits in the DTW data because the confidence limit error bars are always larger at the largest τ and that DTW has the lowest noise level. The higher values for DTW in Fig. 3 are probably a better estimate of the truth since there is no obvious physical explanation for the sudden drop in TOTAL TDEV at $\tau = 2.9 \times 10^6$.

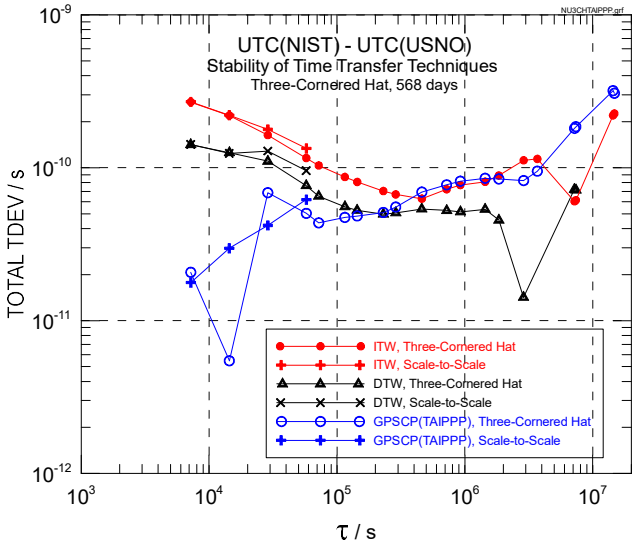

Figure 6. Three-cornered hat results for the Indirect Two-Way (ITW), Direct Two Way (DTW) and GPS Carrier Phase using TAIPPP (GPSCP(TAIPPP)) data in Fig. 4.