Minimum-Entropy Velocity Estimation from GPS position time series

Jarir Saleh (Jarir.Saleh@noaa.gov) and Richard A Bennett, NOAA/NGS
Simon DP Williams, NOC

Abstract

We propose a nonparametric minimum entropy method for estimating an optimal velocity from position time series which may contain unknown noise, data gaps, loading effects, transients, outliers and step discontinuities. Although nonparametric, the proposed method is based on elementary statistical concepts familiar to least-squares and maximum-likelihood users. It seeks a constant velocity with a best possible (realistic) variance rather than a best variable velocity fit to the closest position data. We show based on information theory, synthetic and real data that minimum-entropy velocity estimation: (1) Accounts for colored noise without assumptions about its distribution or the extent of its temporal correlations; (2) Is unaffected by the series deterministic content such as an initial position and the heights of step discontinuities and insensitive to small-amplitude periodic variations and transients; (3) Is robust against outliers and, for long time series, against step discontinuities and even slight non-stationarity of the noise; (4) Does not involve covariance matrices or eigen/singular value analysis, thus can be implemented by a short and efficient software; (5) Under no circumstances results in a velocity variance that decays as $1/N$, where N is the number of observations. The proposed method is verified based on synthetic data and then applied to a few hundred NGL (Nevada Geodetic Lab) position time series of different characteristics, and the results are compared to those of the MIDAS (Median Interannual Difference Adjusted for Skewness) algorithm. The compared time series include continuous and linear ones used to test the agreement between the two methods in the presence of unknown noise, data gaps and loading effects, discontinuous but linear series selected to include the effect of a few (1 to 4) discontinuities, and nonlinear but continuous time series selected for including the effects of transients. Both the minimum-entropy and MIDAS methods are nonparametric in the sense that they only extract the velocity from a position time series with hardly any explicit assumptions about its noise distribution or correlation structure. Otherwise, the two methods differ in every single possible technical sense. Other than pointing to a close agreement between the derived velocities, the comparisons consistently revealed that minimum-entropy velocity uncertainties suggest a smaller degree of temporal correlations in the NGL time series than the MIDAS does.

Key words: Entropy, Mutual Information, GPS position time series, colored stationary noise, velocity uncertainty, outliers, step discontinuities, loading effects, MIDAS.

1. Introduction

Consider a position time series of the form:

$$P_i = P_0 + V \cdot (t_i - t_0) + \epsilon_i \equiv P_0 + V \cdot \delta t_i + \epsilon_i \quad (1)$$

where $i = 1, 2, \cdots, N$, t_0 is a reference time, $\delta t_i = t_i - t_0$ is a timing variable defined over the range [0, T] years at intervals of Δt, where $(N-1) \cdot \Delta t = T$, P_i and P_0 are positions at times t_i and t_0 and V is a constant velocity. If the noise, ϵ_i, is an independent and identically distributed (iid) random variable (RV) of zero mean and a variance of σ^2 mm^2, linear regression leads to a velocity variance given by (e.g., Rice 1988):

$$\sigma_V^2 = \sigma^2 / \{\sum_{i=1}^{N} (\delta t_i - \bar{t})^2\} \equiv \{\sigma^2/(N-1)\} / \{\sum_{i=1}^{N} (\delta t_i - \bar{t})^2/(N-1)\} \quad \text{(mm/year)}^2 \qquad (2)$$

where $\bar{t}$ is the mean of δt_i and the denominator on the right-hand side of (2) describes its variations in years^2. In the absence of data gaps, the timing variable could be viewed as a uniformly distributed RV (i.e., an RV of independent and equally probable values) on the range [0, T], which by definition has a variance of $T^2/12$ years^2, and (2) reduces to (see also Zhang et al. (1997)):

$$\sigma_V^2 = \{\sigma^2/(N-1)\} / (T^2/12) = 12\sigma^2 / \{(N-1)T^2\} \qquad (3)$$

In position time series, however, the noise is not iid due to temporal correlations and the timing variable is not uniform due to randomly occurring data gaps. Thus, (3) leads to biased, optimistic values (Agnew 1992; Langbein and Johnson 1997; Zhang et al. 1997; Mao et al. 1999; Williams 2003; Williams et al. 2004; Santamaría-Gómez et al. 2011). The power of the temporally correlated noise is concentrated in its low frequencies, biasing those of the position and inevitably biasing the estimated velocity relative to what (3) suggests. The only way for avoiding this bias is by understanding the colored noise, all its possible sources and origins and its interaction with other contents of the position time series (Agnew 1992; van Dam et al. 2001; Dong et al. 2002; Blewitt and Lavallée 2002; Dong et al. 2006; Ray et al. 2008; Ji and Herring 2011; Klos et al. 2018; Santamaría-Gómez and Ray 2021; Gobron et al. 2023; Souza and Monico 2004; Chen et al. 2013; Wu et al. 2015; Gualandi et al. 2016; Alevizakou et al. 2018; Bevis et al. 2020; Wang et al. 2021; Gao et al. 2022) and attempting to eliminate it. This task, however, is not very easily achievable, and we concentrated instead on optimal velocity estimation under these complex circumstances, which get even more complex in the presence of loading effects, transients, outliers and discontinuities.

Two different approaches to velocity estimation from geodetic position time series have been prevalent. The first seeks a secular trend with emphasis on achieving a best possible realistic variance (Langbein and Johnson 1997; Zhang et al. 1997; Mao et al. 1999; Williams 2003; Williams et al. 2004; Amiri-Simkooei et al. 2007; Williams 2008; Bos et al. 2008; Amiri-Simkooei 2009; Hackl et al. 2010; Blewitt et al. 2016; Floyd and Herring 2020, Cucci et al 2023). This is suitable for studying slowly changing phenomena such as tectonic motion away from plate boundaries and GIA (Glacial Isostatic Adjustment), and is especially suited for reference frame research where longstanding professional wisdom excludes reference stations in volatile regions (unless a small minority of them is absolutely necessary). The second approach puts the emphasis on estimating a best variable velocity fit to the closest position data (e.g., Dmitrieva et al. 2015; Didova et al. 2016; Olivares-Pulido et al. 2020; Engels 2020) rather than on achieving a best velocity variance. Such methods, usually based on "short memory", “Markov-type” assumptions, are more suited for studying rapidly changing phenomena such as deformations of plate boundaries and other volatile regions. Most methods within both approaches involve computationally intensive and time-

consuming investigations for guessing the noise types, and assume certain noise distributions (e.g., Gaussian) and correlation characteristics (e.g., power-law, Markov chains).

As mapping infrastructure practitioners, our proposed method classifies within the first approach. Although nonparametric, it is based on the same elementary statistical concepts familiar to least-squares and maximum-likelihood analysts (Montillet and Bos 2020), but differs from most existing methods in several respects: (1) It minimizes the entropy (Shannon 1948; Cover and Thomas 2006) of residual positions, which is based on their probabilities rather than values, and therefore reflects their "true uncertainty" (see section 2). Thus, a minimum-entropy solution is optimal in the sense of minimum true uncertainty; (2) Being nonparametric, the proposed method avoids assumptions about the noise distribution and its temporal correlations, sparing the user from lengthy investigations for guessing the noise types; (3) Because entropy (and hence our method) is based on the probabilities rather than values of the residual positions, it is unaffected by deterministic quantities contained in the position time series, such as a bias (e.g., a reference position), a constant (e.g., annual) sinusoid, a transient or the heights of step discontinuities. Thus, the proposed method ignores their presence rather than investigate and estimate them; (4) Entropy is based on the (empirical) probability density function (pdf) of the position time series, thus contains a complete description of its stochastic properties including any correlations between its elements; (5) It is robust against outliers because their probability is low; (6) The (not necessarily accurate) locations of discontinuities are assumed known based on a pre-analysis, but the method is oblivious to their heights; (7) It involves no heuristic thoughts or algorithms and neither matrices nor singular/eigen value decompositions (see section S4), thus can be implemented by a short and efficient software.

Briefly, the proposed method minimizes the entropy of residual positions for estimating an optimal velocity, then uses the minimum entropy value for decorrelating the time series before computing velocity variance. First, we verify the method based on synthetic data. Then we apply it to a few hundreds of NGL daily position time series (Blewitt et al. 2018) of different characteristics, and the resulting velocities and their uncertainties are compared to those derived by the MIDAS method (Blewitt et al. 2016).

Entropy-based methods, collectively called "information theory", have been useful for several decades (Shannon 1948; Cover and Thomas 2006). Section 2 introduces entropy (also called "Shannon's entropy", "information entropy" or "Shannon's measure of information (SMI)" to distinguish it from thermodynamic entropy) and lists some of its basic properties which we have exploited for formulating the proposed method. The mechanics of minimum-entropy velocity estimation and noise decorrelation are described in section 3. Section 4 describes tests of the method on synthetic data, done for understanding: (1) The performance of the method in the presence of an unknown but stationary mix of white and colored noises, simulated to mimic (but exceed in magnitude) the noise of vertical positions of a typical IGS station; (2) Theoretically, entropy (and hence the proposed method) is unaffected by constant sinusoids, but this insensitivity has limitations in reality. We try to understand these limitations; (3) The effect of step discontinuities on minimum-entropy velocity; (4) The effect of a smooth transient; (5) The effect of non-stationary noise. Section 5 describes tests of the proposed method on real data, using 276 NGL medium- to long-duration

GPS position time series of different characteristics, collected at stations distributed in and around the US. Section 6 presents a brief summary.

More technical aspects are left to the Supplementary Material. Section S1 compares two nonparametric entropy computation methods: (1) Replacing pdfs by histograms (Wallis 2006; Stone 2015) created using GMT (Wessel and Smith 1998); (2) M-Spacings based on the variability within the time series (Vasicek 1976, Beirlant et al. 1997). Section S2 lists more details of entropy properties, some of which are based on "Mutual Information", the second most important information-theoretic quantity. Section S3 lists equations used for building covariance matrices of known mixtures of noises (Williams 2003). Section S4 presents a justification for approximating the entropy rate by the Vasicek (1976) nonparametric M-Spacings method. Section S5 lists the IDs of the 276 NGL stations used to test the proposed method and a map of their geographic distribution. A short software for minimum-entropy velocity estimation which calls some *Numerical Recipes* routines (sort, medfit, locate, rofunc and select) (Press et al. 2007), is also included.

2. Entropy

The entropy of a random experiment is a measure of its irreducible true uncertainty. The term "uncertainty" is used in this paper in two different contexts. The first and more common measures the spread of values of an RV around their mean and is expressed by the variance. The second is based on the probabilities rather than values of the RV, and is therefore called "true uncertainty" hereafter. It expresses the "surprise" or "new information" brought about by the occurrence of a random event. If the probability of occurrence of an event in a random experiment is p, the surprise experienced when this event occurs or equivalently, the true uncertainty about its occurrence just before it occurs, can be quantified as $1/p$. For example, if a biased coin lands heads up 90% of the time, the surprise/true-uncertainty of observing a head when this coin is tossed can be quantified as 1/0.9 = 1.1 while that of observing a tail is 1/0.1 = 10.0, much larger as it should be. Shannon (1948), for reasons explained below, suggested using $log\ (1/p)$ instead (called "the Shannon information"), and defined entropy, H, as the average surprise/true-uncertainty:

$$H \doteq E\{surprise\ or\ true\ uncertainty\} = E\{\log(1/p)\} = \sum_{i=1}^{n} p_i \cdot log(1/p_i) \qquad (4)$$

where E is the expectation and n is the number of all possible independent outcomes of the experiment. The logarithm, base 2 in our calculations, was introduced by Shannon (1948) to make entropy additive ($E\{\log(1/p \cdot 1/q)\} = E\{\log(1/p)\} + E\{\log(1/q)\}$) and to zero it out if there is no surprise/uncertainty, i.e., when the probability of one outcome is 1 and of all others is 0 (because according to L'Hôpital's rule, $lim\{p \cdot log(1/p)\} \to 0$ as $p \to 0$). When the logarithm is base 2, the unit of entropy is the "bit" and when the natural logarithm is used, the unit is the "nat". The entropy (4) is always positive, equals zero only if there is no uncertainty, and increases gently with increasing uncertainty. The entropy of the biased coin example, where $n = 2$ (heads and tails), is 0.469 bits ($0.1 \cdot log_2(10.0) + 0.9 \cdot log_2(1.1)$). Had the coin been fair, its entropy would have been 1.0 bit ($0.5 \cdot log_2(2.0) + 0.5 \cdot log_2(2.0)$), larger because the average uncertainty would have been larger. When extended to a continuous RV x with pdf $f(x)$, (4) is replaced by (Cover and Thomas 2006, chapter 8):

$$H = \int_{-\infty}^{\infty} f(x) \cdot \log(1/f(x)) \cdot dx. \quad (5)$$

Equations (4) and (5) are useful for explaining the meaning of entropy, but they can only be used in the simplest of cases, when the RV is iid, i.e., uncorrelated. In the more realistic cases where the values are correlated, e.g., the elements of a residual GPS position time series, they must be "compressed" into a sequence of independent values before (4) or (5) can be used. This "compression" is done by introducing the "entropy rate". When each new element of the time series depends to some extent on the previous ones, the added uncertainty when we observe a new element decreases as more elements are observed. The "entropy rate" is an average entropy which accounts for this decreasing rate of uncertainty.

A time series, $X_i, i = 1,2,\cdots,N$ of residual positions can be viewed as the outcome of a continuous stochastic process which generates the random vector $X = [X_1, X_2, \cdots, X_N]$ with a joint pdf $f(X) = f(X_1, X_2, \cdots, X_N)$ and a joint entropy, $H(X_1, X_2, \cdots, X_N)$, given by (Cover and Thomas 2006, eq. (8.31)):

$$H(X_1, X_2, \cdots, X_N) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(X_1, \cdots, X_N) \cdot \log(1/f(X_1, \cdots, X_N)) \cdot dX_1 \cdots dX_N. \quad (6)$$

The entropy rate of the process is formally defined as (Cover and Thomas 2006, eq. (4.10)):

$$H(\mathcal{X}) = \lim_{N\to\infty} \frac{1}{N} \cdot H(X_1, X_2, \cdots, X_N) \quad (7)$$

where $\mathcal{X}$ is the "compressed" (uncorrelated) series. Equation (6) states that the residual time series is not considered iid. Rather, its correlated elements are assumed drawn from different marginal distributions and the joint pdf, $f(X)$, describes the statistical relations, including correlations, between all these non-iid elements, which gives rise to the joint entropy $H(X_1, X_2, \cdots, X_N)$. Equation (7) states that the "entropy rate" is the irreducible average entropy of the correlated series as it grows very long, and to approximate it, one must know the joint entropy first. Equations (6) and (7) are formal definitions, difficult to use in practice without further assumptions. Common practice follows by making assumptions about the joint distribution (e.g., multivariate Gaussian, see section S4) and about the depth of the correlations within the series (e.g., Markov chains). We chose a different approach with less explicit assumptions and much smaller computational burdens, which we deduced empirically based on synthetic and real residual GPS position time series (see also section S4). Notice first that - thanks to the additivity property (i.e., the log in (6)) - had the elements of the series been iid, their joint entropy would have been the sum of their identical marginal entropies, i.e., $N \cdot H$, where H is the marginal entropy of an element. Assuming that the series is sufficiently long to satisfy the limit in (7), e.g., $N \gg 30$, and substituting this joint entropy in (7) results in an "entropy rate" equal to H. Thus, for an iid time series, entropy and entropy rate are equal. The residuals, however, are not iid but temporally correlated and there is no such thing as the identical marginal entropy, H. We empirically found that H can be accurately and efficiently replaced by an average entropy, $\mathbb{H}$, computed from the variations within the time series, ignoring the possibility that it might not be iid. In other words, we approximate the joint entropy, (6), by:

$$H(X_1, X_2, \cdots, X_N) \cong N \cdot \mathbb{H} \quad (8)$$

where $\mathbb{H}$ is an average entropy, computed from the variations within the residual positions, by a nonparametric entropy computation method (Beirlant et al. 1997). We considered two of several existing

such methods (see section S1): (1) Replacing the pdf by a histogram; (2) M-Spacings (Vasicek 1976). We chose to proceed using the M-Spacings method of Vasicek (1976) for its simplicity, computational efficiency and better accuracy. Vasicek showed, relying on the empirical cumulative distribution function (see derivations in section S1.2), that the nonparametric entropy of an iid sample (approximated by a residual position time series in our case) can be computed from the variations within its elements by:

$$\mathbb{H} = \frac{1}{N}\sum_{i=1}^{N}\log\left\{\frac{N}{2M}\left(x_{(i+M)} - x_{(i-M)}\right)\right\} \quad (9)$$

where $x_1, x_2, \cdots, x_N$ are the sample's elements, N is its size, $x_{(1)}, x_{(2)}, \cdots, x_{(N)}$ are its order statistics, i.e., its elements sorted in an ascending order, $x_{(1)} < x_{(2)} \cdots < x_{(N)}$, and M is a positive integer < $N/2$. Vasicek (1976) obtained optimal results using small values of M, e.g., 2 and 4. We used *M*=1, mainly to effectively remove the effects of smooth, long periodic content such as annual loading effects and transients. Notice that, because entropy is a function of probabilities rather than values, $\mathbb{H}$ is a function of differences, i.e., variations, rather than the values of the elements of the time series. This is a key to much of what follows, and is the reason why entropy is unaffected by additive constants in the series, such as a reference position, P_0, or an annual sinusoid, $A \cdot \sin(\theta + \Phi)$, of constant parameters.

Let δt be a timing RV with entropy $H(\delta t)$, δV a "velocity" with absolute value $|\delta V|$, and ϵ a stationary noise uncorrelated with other phenomena, then: (Cover and Thomas 2006, Theorems 8.6.3 and 8.6.4):

$$H(P_0 + \delta V \cdot \delta t + \epsilon) = H(\delta V \cdot \delta t + \epsilon) = H(\delta V \cdot \delta t) + H(\epsilon) = H(\delta t) + log_2|\delta V| + H(\epsilon) \quad (10)$$

Notice that the minimal value of the entropy (10) for general RVs δt and ϵ is obtained as $|\delta V| \to 0$.

The sinusoids discussed in geodetic time series analysis are modeled as: $g(\delta t) = A \cdot \sin(\omega \cdot \delta t + \Phi)$ where the parameters A, ω and Φ are usually assumed constants. The presence of the timing RV, δt, in the sinusoid, $g(\delta t)$, makes it stochastic which gives rise to a certain nonzero entropy. However, it can be shown that δt and $g(\delta t)$ are stochastically identical (see proof in section S2), i.e., no additional uncertainty beyond that of δt (already accounted for in (10)) is introduced by the stochastic sinusoid. Thus:

$$H(P_0 + \delta V \cdot \delta t + A \cdot \sin(\omega \cdot \delta t + \Phi) + \epsilon) = H(\delta t) + log_2|\delta V| + H(\epsilon). \quad (11)$$

This theoretical insensitivity of the entropy of a time series to its periodic content (assuming of course that the time series knows the periodic content well, i.e., is long enough to contain several cycles of it (Blewitt and Lavallée (2002)) has two limitations. First, theory is based on continuous and infinite RVs while geodetic time series are discrete and finite. Second, the parameters of "periodic" loading effects are not exactly constants. Nonetheless, these limitations depend on the series length and amplitudes of "periodic" effects, and we demonstrate in section 4.2 that, for small amplitudes and long series, (11) is a good approximation.

Entropy is robust against outliers because their probabilities are very small (see (5)). If it is still desired to protect against huge outliers, the order statistics in (9) could be slightly trimmed on both sides before computing their differences. When *d* step discontinuities are present in the position time series, their locations are assumed known based on a pre-analysis. The time series can then be broken into *d+1* partial series, one between every two consecutive discontinuities. Each partial series is assigned a probability, p_i, equal to the ratio of its length to the total length of the entire series. The entropy of each partial series, $\mathbb{H}_i$, $i = 1,2,\cdots, d+1$, is computed by (9) and the total entropy of the entire series by:

$$\mathbb{H} = \sum_{i=1}^{d+1} p_i \cdot \mathbb{H}_i. \quad (12)$$

Because differences rather than values of the elements are used in (9), equation (12) is oblivious to the heights of step discontinuities. A slight misidentification of locations of discontinuities in the pre-analysis has no effect on the results because entropy is robust to outliers. Unfortunately, however, discontinuities still degrade minimum-entropy velocity because as their number increases, the length of partial series decreases violating convergence requirements underlying the nonparametric entropy (9) (see section S1.2) and weakening its convergence (see also Santamaría-Gómez and Ray 2021 and Gobron et al. 2022).

Finally, a residual time series can be decorrelated by relating its entropy (9) (or (12) if discontinuities exist) to the variance, $\sigma_{\delta P}^2$, of an "effective", uncorrelated, Gaussian RV which has that same entropy (and hence the same true uncertainty). This is achieved by substituting the expression for a univariate Gaussian pdf in (5) (see full derivation in section S2) and leads to:

$$\mathbb{H} = \frac{1}{2} log_2(2\pi e \sigma_{\delta P}^2) = 2.0471 + log_2(\sigma_{\delta P}) \quad \text{bits} \quad (13)$$

$$\Longrightarrow \quad \sigma_{\delta P} = 2^{\mathbb{H}}/\sqrt{2\pi e} \equiv 2^{(\mathbb{H} - 2.0471)} \quad \text{mm}. \quad (14)$$

The notation $\sigma_{\delta p}$ is used because it is the SD of a position change relative to P_0 and not an absolute position.

3. The mechanics of nonparametric minimum-entropy velocity estimation

A velocity search range of width dV mm/year is chosen such that it contains the true velocity. It can be designed by fitting a robust line to the longest partial series and bounding its slope, V_0, in a confidence interval $[V_0 - dV/2\,, V_0 + dV/2]$. This range is then divided into small equal increments. After experimenting with increment sizes from 0.001 to 0.1 mm/year and verifying their insignificant effect on the results, we used increments of 0.01 mm/year. Thus, the search range contains $100 \cdot dV + 1$ equally spaced values, one of which is the true velocity. For each value V_r in this range, a residual time series is defined by:

$$\begin{aligned} e(t_i) &= P(t_i) - P_r(t_i) \equiv [P_0 + V \cdot (t_i - t_0) + \epsilon(t)] - [P_{0r} + V_r \cdot (t_i - t_0)] \\ &= (P_0 - P_{0r}) + (V - V_r) \cdot (t_i - t_0) + \epsilon(t_i) \equiv \delta P_0 + \delta V \cdot \delta t_i + \epsilon(t_i) \end{aligned} \quad (15)$$

where $i = 1,2,\cdots,N$, V is the true, constant (i.e., mean) velocity implicit in the position time series, P_{0r} is an a priori value of the reference position, P_0, and $\epsilon(t)$ contains the rest of the position time series including unknown noise, "periodic" and transient signals and outliers. There is no need for a search range for the initial position, P_{0r}, because entropy is unaffected by constants. If step discontinuities exist, the series is divided into partial series as explained in section 2. The entropy of the residual series is computed using (9) (then (12) if necessary) resulting in a sequence of $100 \cdot dV + 1$ entropies, one per searched velocity value. According to (11), the smallest of these entropies, H_{min}, corresponds to the optimal velocity. Equation (14) is used to convert this minimal entropy to the SD, $\sigma_{\delta P}$, of an "effective" Gaussian iid position which has that same entropy, $\mathbb{H} = H_{min}$, hence reflects the true uncertainty within the position time series. The corresponding velocity SD, σ_V, is finally computed based on the laws of white noise propagation, by:

$$\sigma_V = \sigma_{\delta P}/T \quad \text{mm/year} \quad (16)$$

where T is the total length of the time series in years, during which the position change δP occurred.

4. Tests of minimum-entropy velocity estimation on synthetic data

4.1 The performance of minimum-entropy velocity estimation with colored and white noise

The first test on synthetic data considered colored noise. We simulated 100 independent stationary noise series of $N = 10{,}000$ days (or 27.4 years) (see, e.g., Fig S1 (top)), consisting of a combination of 2.7 mm Gaussian white noise (WN) and 5.0 mm of power-law (PL) noise of spectral index $\kappa = -0.9$ (Agnew 1992). Synthetic position time series were then created by adding each of the noise series to the positions $P_i = P_0 + V \cdot (t_i - t_0)$ at time (day) t_i, $i = 1,2,\cdots,N$, where $P_0 = 10.0$ mm is the true initial position, $V = -18.0$ mm/year is the true velocity and t_0 is the first epoch of t_i. A velocity search range, $[-30.0, -8.0]$ and increment of 0.01 mm/year were used, thus $\mathrm{d}V = 22.0$ mm/year and the search range contained $100 \cdot dV + 1 = 2201$ velocity values. Figure 1 shows a typical behavior, a gradual smooth descent, of the corresponding 2201 entropies as a function of searched velocity values, converging to a distinct minimum ($H_{min} = 4.9786$ bits, at $V = -17.80$ mm/year in this example), very close to the true velocity. This minimum entropy is a measure of the true uncertainty of the position time series at that optimal velocity. Thus, it contains a complete description of its stochastic properties including any correlations that may exist within it. Equations (14) and (16) are then used to estimate the velocity SD (0.278 mm/year in this example).

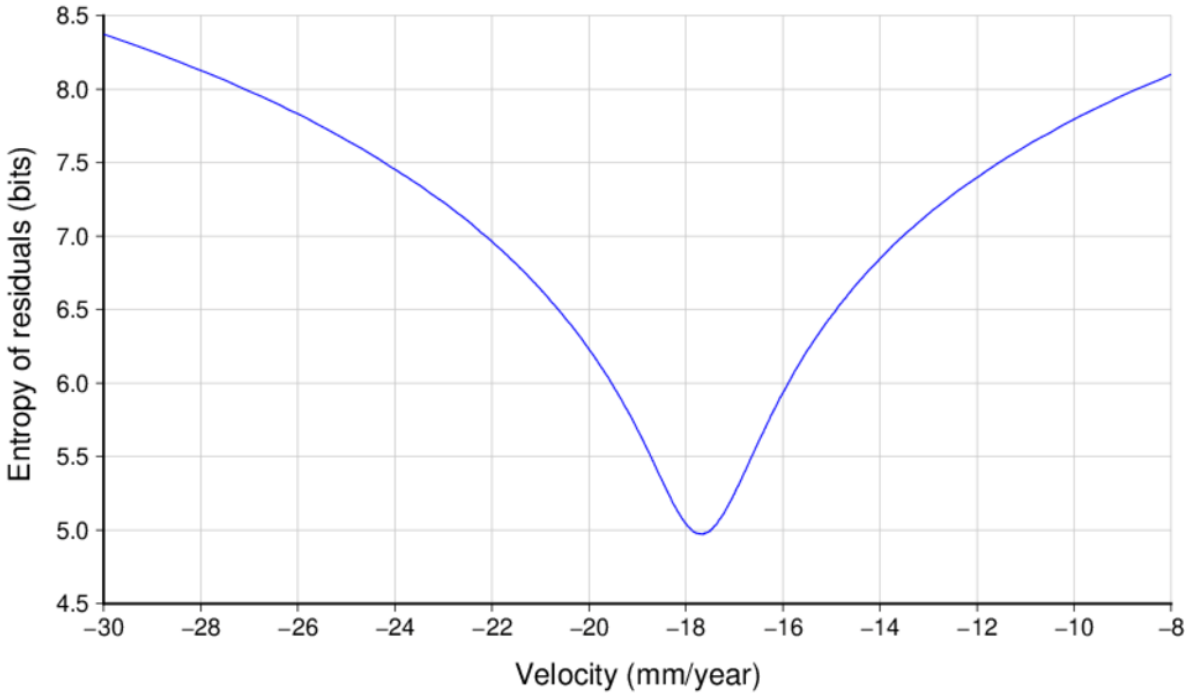

Fig 1: All 2201 values of the entropy in one of 100 minimum-entropy velocity solutions. The minimum entropy value is reached at the minimum-entropy velocity, very close to the true velocity and its value is then used to decorrelated the position time series using eq (14).

For each of the 100 synthetic position time series, we estimated the minimum-entropy-derived velocity, its bias (i.e., estimated minus true velocity) and the “effective” velocity SD as obtained from (14) and (16).

To sum up the results of the 100 solutions, we computed the RMS of the 100 velocity biases and that of the 100 velocity SDs. These computations were first done for a time series of length of 10,000 days (or 27.4 years) and then repeated for time series of 23, 20, 17, 13.7, 10, 7, 5 and 3 years of synthetic data. Figure 2 presents the RMS values of the 100 minimum-entropy velocity biases and SDs as a function of time series length. It also presents corresponding quantities obtained from linear regressions with properly computed covariance matrices of the known noise, based on the equations of section S3 (Williams 2003). Notice that the SDs in Fig 2 are realistic, i.e., the velocity SDs of both methods are practically equal to the velocity biases. Notice also that, in the presence of a stationary mixture of white and colored (close to flicker) noise, the efficient minimum-entropy velocity estimation, while entirely blind to the structure and characteristics of the noise, results in velocity estimates (value and uncertainty) which are practically equal to those obtained by the (computationally expensive) linear regression with a priori knowledge of the noise.

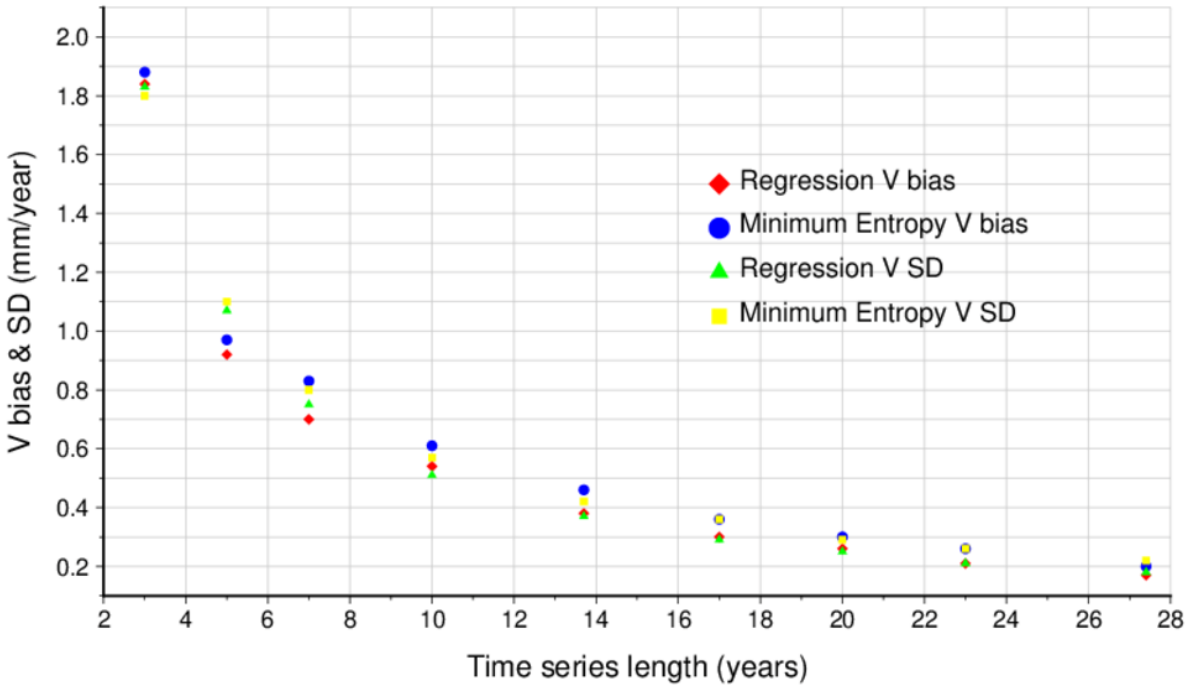

Fig 2: The velocity bias (estimated minus true) and SD from two different methods: (1) Minimum-entropy, which is blind to the noise structure and (2) Regression with a priori knowledge of the noise structure.

When this experiment is repeated using purely white noise of comparable variance ($(2.7\,mm)^2 + (5\,mm)^2 = 32.29\,mm^2$), a slightly different patterns emerge. While the velocity bias (i.e., estimated minus true velocity) became very small (0.05 mm/year) for N=10,000 days, the estimated velocity SD stays larger than the velocity bias, about twice its size (0.1 mm/year), or ~2σ, where σ is the velocity SD derived by linear regression. This is due to the fact that variances of minimum-entropy velocity estimation under no circumstances decrease proportional to $1/N$. Rather, they are evaluated based on the true uncertainty (variability) within the time series.

4.2 On the insensitivity of entropy to small-amplitude periodic signal

Because the entropy of residual positions is based on probabilities (i.e., variations within its elements rather than their absolute values are used in (9)), any smoothly and slowly (compared to Δt) varying signal contained in the time series, such as annual and a semiannual periodic effects and transients, nearly cancel out by the differencing (see also theoretical arguments to this effect in sections 2 and S2). To numerically examine the remaining effects (after the differencing) on minimum-entropy velocity estimation, we used a 13.7-years (or 5000 days) portion of one of the simulated noise series. To this noise, an annual sinusoid with amplitude of 7 mm is added. The upper row of Fig 3 shows the pdfs (or smoothed histograms, i.e., probability versus bin value in mm) of the noise (right) and the combination of noise plus sinusoid (left). The addition of the sinusoid clearly flattens the pdf decreasing its peak by ~25% and increasing its variance by ~36%, from 53.828 mm^2 for the noise alone to 73.616 mm^2 after adding the sinusoid. In contrast, the lower row in Fig 3 presents smoothed histograms of the corresponding weighted "Shannon information", i.e., $p \cdot log_2(1/p)$, of the noise (right) and combination of noise and sinusoid (left), clearly showing that it (and hence the entropy) dilutes the effect of the sinusoid. While the entropy of the noise alone based on (9)

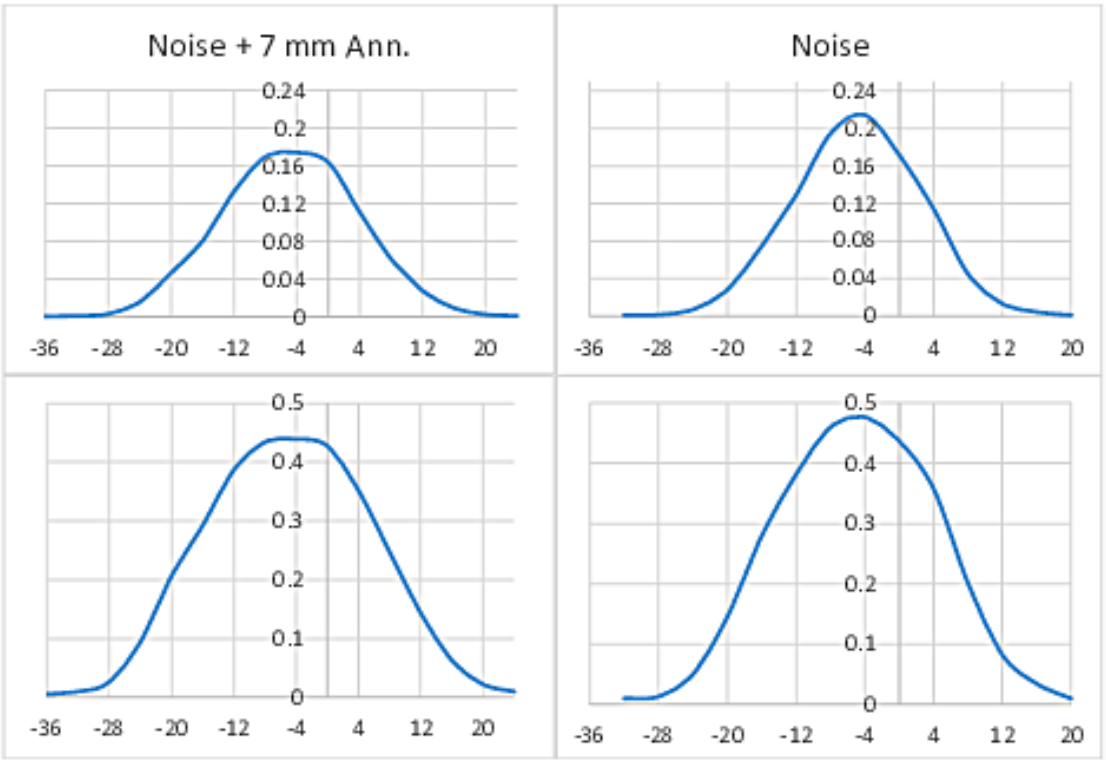

Fig 3: (Top right) Smoothed histogram (i.e., probability versus binned value in mm) of 10 years of simulated position noise (2.7 mm WN plus 5.0 mm power-law noise with spectral index k=-0.9); (Top left) The smoothed histogram of the above noise plus an annual sinusoid of amplitude 7 mm; (Bottom right) The weighted Shannon information ($p \cdot log_2(1/p)$) of the same noise; (Bottom left) The weighted Shannon information of the same noise and annual sinusoid

is 4.50 bits, that of the combination of noise and sinusoid is 4.72 bits, only 5% larger, increasing the velocity uncertainty (see (14) and (16)) by 0.06 mm/year and practically causing no change in the estimated velocity value. In addition, the effect of the added sinusoid on velocity uncertainty decreases with increasing series length, reaching 0.04 mm/year for a series 27.4 years long. However, the effect of an added sinusoid increases with increasing amplitude. For a 27.4-years long daily time series, although the addition of a sinusoid of amplitude 20 mm quadruples the variance of the time series, it increases its entropy by only 20%. This has a negligible effect on minimum-entropy velocity, but doubles the minimum-entropy velocity SD, from 0.2 mm/year to 0.4 mm/year. For a series 13.7-years long, this effect is similar except that the velocity SD goes from 0.4 to 0.8 mm/year.

4.3 The effects of transients, discontinuities and non-stationary noise

To understand the effect of smoothly varying nonlinearities on minimum-entropy velocity estimation, we created 100 synthetic position time series of the form $P_i = P_0 + V \cdot (t_i - t_0) + \frac{1}{2} \cdot a \cdot \left(t_i - t_{N/2}\right)^2 + \epsilon(t_i)$, where a=0.3 mm/year^2, $t_{N/2}$ is the timing variable at the middle of the series and $\epsilon(t)$ is the simulated mixture of white and power-law noise described in section 4.1. The nonlinear part of the synthetic positions (i.e., $\frac{1}{2} \cdot a \cdot \left(t_i - t_{N/2}\right)^2$) is shown in the top of Fig S2. The green curve in the middle of Fig S2 presents this part immersed in the colored noise of one of the 100 previously mentioned simulated noise series and the red curve is its smoothed value. The latter, added only to bring out the colored character of the noise, is computed using an edge detecting filter known as “the weak elastic string” (Saleh 1996). Figure 4 presents the RMS values of velocity biases and SDs resulting from the 100 nonlinear, synthetic position time series. It also shows corresponding results when the underlying motion model is linear (i.e., $P_i = P_0 + V \cdot (t_i - t_0) + \epsilon(t_i)$), which was previously verified and presented in Fig 2. As in the case of small-amplitude “long-periodic” signals, the fact that entropy is based on probabilities (i.e., differences in (9)) dilutes the effect of the a nonlinear transient. In addition, the smaller the value of *M* parameter is (we used the minimal possible value of *M*=1), the more effective is the removal of transient effects. Therefore, the difference between the estimated velocity values in both cases is small, reaching 0.15 mm/year for 3-year long time series and decreasing with series length (Fig 4). As expected, however, the effect of the nonlinearity inflates minimum-entropy velocity uncertainty for very long time series. The velocity uncertainty reaches 0.35 mm/year for a 27.4 years long series when the transient is present but reaches only 0.2 mm/year when the series is linear. In summary, transient motion could slightly inflate the estimated velocity uncertainty for very long time series and slightly bias estimated velocities for short time series.

Discontinuous time series are not stationary but increment-stationary (i.e., their differences are stationary). Thus, the question arises whether (9), which is based on differences, can still be used to give a good approximation of the entropy rate of such series. Unfortunately, unless the heights of the step discontinuities are much larger than the combined effect of the noise and slope, the answer is negative. This is because when the series is sorted in preparation for applying (9), some noise, step functions and change due to the slope get scrambled, inflating the resulting entropy value (see also Santamaría Gómez

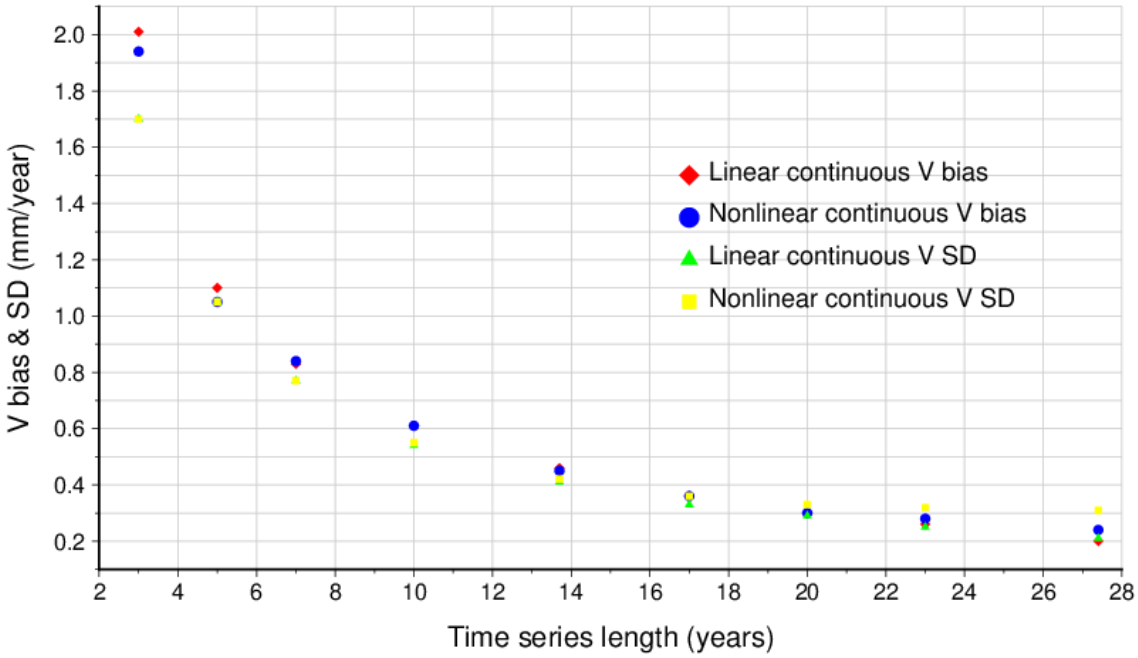

Fig 4: RMS of velocity bias and velocity standard deviation (SD) based on 100 independent simulated noise sequences, both derived by minimal-entropy velocity estimation, for linear and nonlinear position time series. The 100 linear position time series are constructed by $P_t = P_0 + V \cdot (t - t_0) + \varepsilon(t)$ using 100 independent noise series, $\varepsilon(t)$, consisting of combinations of white and power-law noise. The nonlinear series are similar except that $P_t = P_0 + V \cdot (t - t_0) + \frac{1}{2} \cdot a \cdot \left(t - t_{N/2}\right)^2 + \varepsilon(t)$, where a=0.3 mm/year^2

and Ray (2021) and Gobron et al. (2022)). A simple way for preventing this scrambling is by dividing the series into partial series, one between each pair of consecutive discontinuities. The entropies of the partial series are computed by (9), the entropy of the entire time series is computed by (12), and velocity estimation continues as done for a continuous time series. Figure 5 presents the RMS values of velocity biases and SDs based on the 100 synthetic linear position series used in section 4.1. It presents solutions for 2 cases, both computed by the minimum-entropy method: (1) Continuous position time series, which were previously verified and presented in Fig 2; (2) Same position time series except with one discontinuity (of irrelevant height) at the middle of the series. Fig 6 is very similar except that it considers the effect of 2 discontinuities at the middle and three quarters of the series. These tests reveal that the presence of discontinuities hardly affects minimum-entropy velocity uncertainties because, even when discontinuities exist, it is still computed based on the variability within the entire time series. Unfortunately, the effect of discontinuities on the estimated velocity value is more complex. Discontinuities divide the series to shorter pieces than the complete series, which violates the convergence requirements of (9) and weakens its convergence to the true entropy, slightly biasing the location of the minimum value of (11), and slightly increasing the velocity bias with every additional discontinuity. As expected, these additional biases decrease with increasing series length.

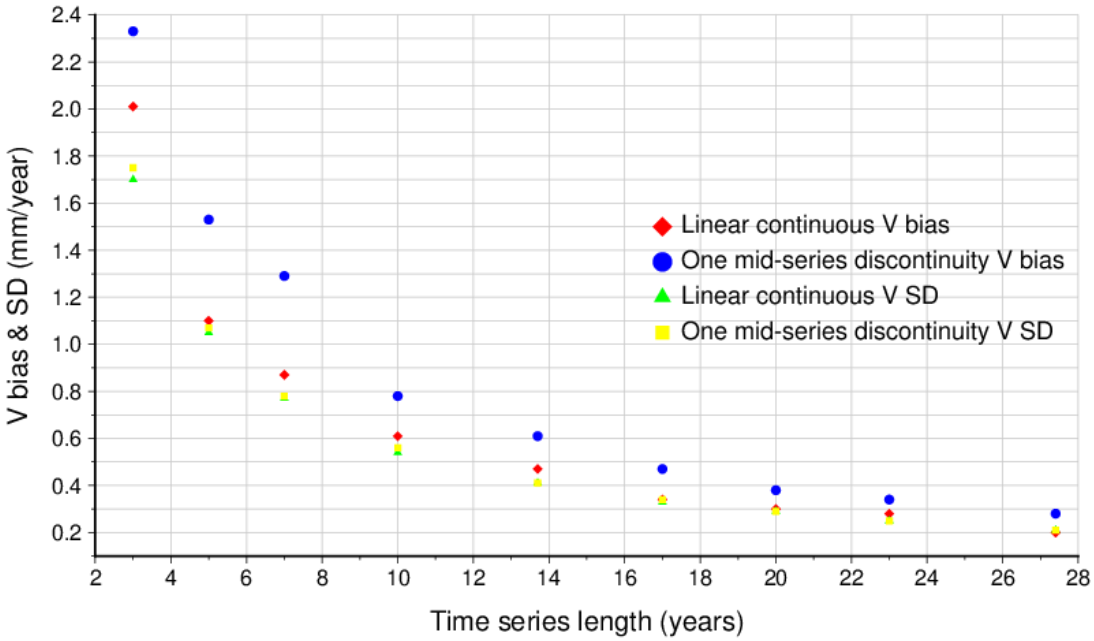

Fig 5: RMS values of velocity bias (i.e., minimum-entropy-derived minus true velocity) and standard deviations (SDs) based on 100 simulated noise series, for 2 cases, both computed using the proposed minimum-entropy method: (1) Continuous position time series and (2) Position time series with one discontinuity (of irrelevant height) in the middle of the time series

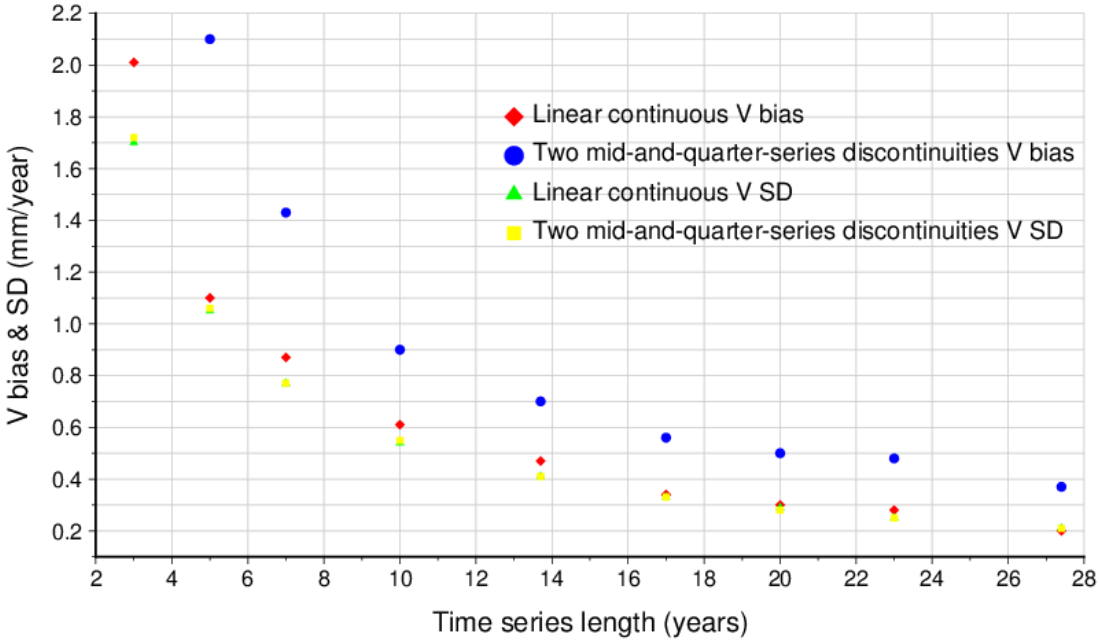

Fig 6: RMS values of velocity bias (i.e., minimum-entropy-derived minus true velocity) and standard deviations (SDs) based on 100 simulated noise series, for 2 cases, both computed using the proposed minimum-entropy method: (1) Continuous position time series and (2) Position time series with two discontinuities (of irrelevant heights), at the middle and at three quarters of the time series

We finally examine the performance of minimum-entropy velocity estimation in the presence of non-stationary noise. To this end, we simulated three noise types: (see also Fig. S1 and section S3): (1) A slightly non-stationary noise with 2.7 mm WN and 5 mm PL noise of spectral index $k = -1.1$; (2) A slightly non-stationary noise similar to that described in section 4.1, namely consists of 2.7 mm WN and 5 mm PL noise of $k = -0.9$, but with the addition of random walk (RW) (i.e., $k = -2.0$) of step size $s = 0.1\ mm$; (3) A significantly non-stationary noise similar to that of (2) except that the RW step size is $s = 0.5\ mm$. For each noise type and for each of the lifespans of 3, 5, 7, 10, 13.7, 17, 20, 23 and 27.4 years, 100 daily noise time series were simulated independently and used to create synthetic position time series based on known true reference position and velocity. The velocity estimation from each synthetic series was first computed for noise type 1 (Fig 7 (top)), then repeated for noise type 2 (Fig 7 (middle)) and 3 (Fig. 7 (bottom)). For each lifespan, an RMS of the 100 velocity biases (estimated minus true velocity) and that of the 100 velocity SDs were computed, first using minimum-entropy, which is unaware of the noise characteristics, and then using regression with proper covariance matrices computed as described in section S3 (Williams 2003).

We know that information theory guarantees the existence of the limit in (7) only for a stationary stochastic processes (Cover and Thomas 2006, Theorem 4.2.1). So, we already know based on theory that minimum-entropy velocity estimation is likely in error when the noise is non-stationary. What we did not know before these simulations is the size and characteristics of that error. The results, presented in Fig 7, provide some estimates and insights. Note first that as usual, the regression solution in Fig 7 is included to provide a best-known estimate, a standard for measuring the performance of the minimum-entropy solutions. The results of these simulations suggest that as the noise of the position time series gradually becomes less stationary, from type 1 to 2 to 3, the minimum-entropy derived velocity slowly and slightly diverges away from the truth. For noise types 1 and 2, which are only slightly non-stationary, the minimum-entropy velocity estimation bias reaches 0.2 mm/year for short (< 8 years) time series and decreases slowly as the series length increases until it almost disappears for very long (> 20 years) series. For significantly non-stationary noise such as that of type 3, although the minimum-entropy velocity bias also decreases with time series length, it still reaches 0.4 mm/year even for very long series. Unfortunately, the minimum-entropy derived velocity uncertainties are under estimated by a factor of ~2 for noise types 1 and 2, and a factor of 3 to 5 for a noise of type 3. This is because minimum-entropy estimation is based on the assumption of stationarity. Although the underestimation of velocity uncertainty does improve slowly with increasing time series length, the improvement is slower if random walk is present in the noise.

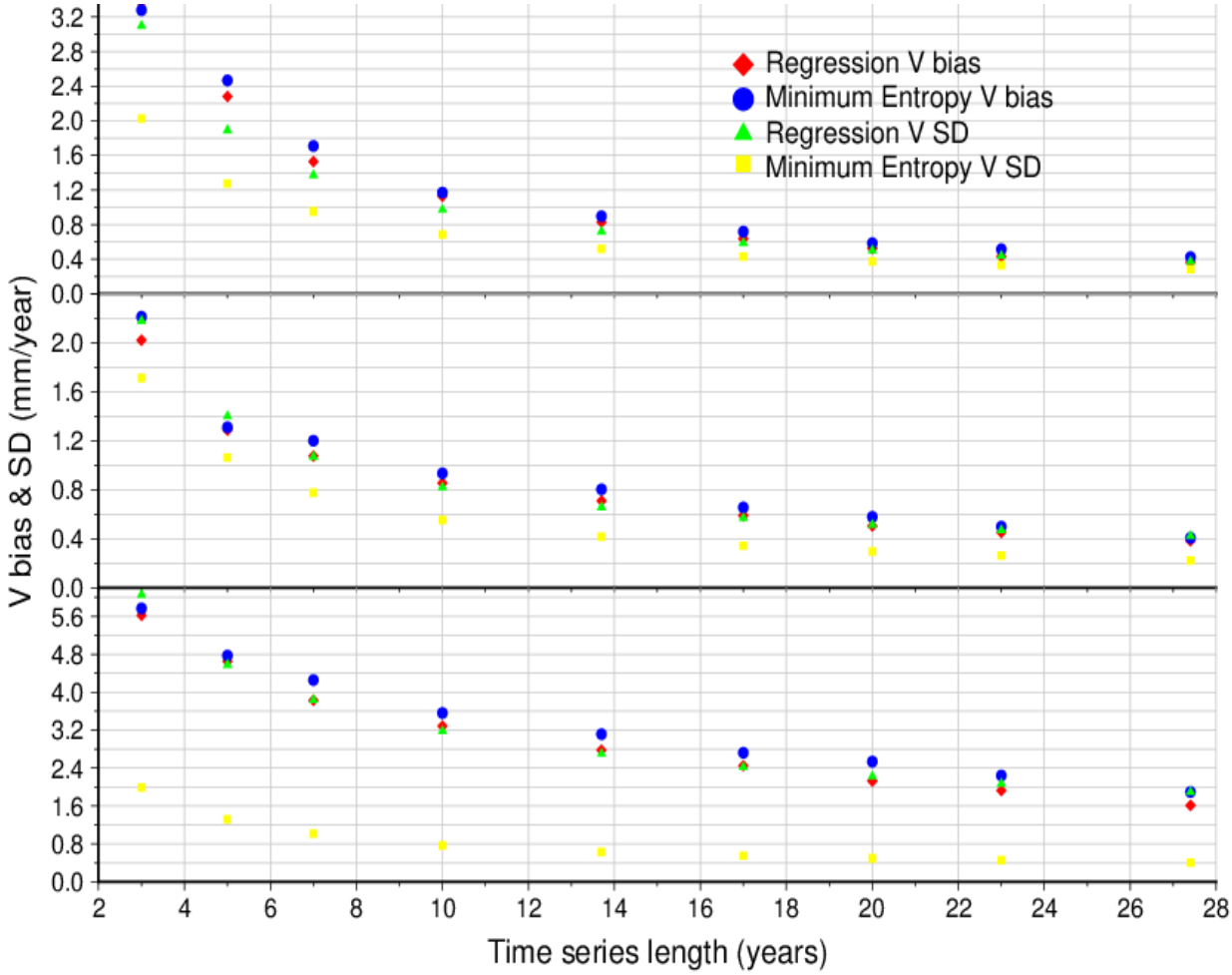

Fig 7: Minimum-entropy and regression estimated velocity bias and its standard deviation in mm/year in the presence of 3 different nonstationary noise types: (top) Noise type 1; (middle) Noise type 2; (bottom) Noise type 3 (see descriptions of noise types in section 4.3)

5. Application of minimum-entropy estimation to real data and comparison to the MIDAS algorithm

We selected 276 daily NGL position time series (Blewitt et al. 2018) from stations distributed mainly over the US (see section S5 for a station list). We initially chose these stations because of their geographic locations and because they had only one GPS antenna throughout their history, which was chosen to be longer than 3 years. However, we found that the time series can be divided into 3 groups. The first includes 171 stations with continuous and linear time series (Fig S3 and S4), the second includes 55 stations with nonlinear transients but no discontinuities and the third includes 50 stations which had experience one to four discontinuities in their history (although they had only one antenna). These time series were processed by the MIDAS algorithm (Blewitt et al. 2016) and the minimum-entropy method and velocities and their uncertainties were compared. It is worth mentioning that both methods, the minimum-entropy and MIDAS, are nonparametric velocity estimators in the sense that they extract only the velocity from a position time

series with hardly any explicit assumptions about its noise distribution or correlation structure, thus both are efficient. Otherwise, however, the two methods differ in every single possible technical sense.

Figure 8 presents the differences between the estimated velocities (minimum-entropy minus MIDAS) versus time series lifespan in years showing that, in the great majority of cases, they are smaller than the minimum-entropy velocity uncertainty. Figure 9 presents histograms of the velocity differences (mean: -0.002, 0.009 and 0.002; RMS: 0.118, 0.136 and 0.322 mm/year for the East, North and Up components, respectively). Figure 10 presents velocity uncertainties versus time series lifespan for both methods on a log-log plot (log base 2) showing that, as expected, the uncertainties cluster linearly in time. In addition, the uncertainties derived by both methods are similar for short time series (< ~5 years). However, the minimum-entropy linear clusters, in all components, exhibit larger slopes than the MIDAS ones. Assuming that the noise is dominated by power-law characteristics (Agnew 1992), these larger slopes imply that the minimum-entropy method senses smaller/shorter temporal correlations (i.e., more “whiteness”) within the noise than the MIDAS does (see, e.g., Fig 1 in Williams (2003)). In other words, while both methods lead to similar velocity uncertainty for short time series, their uncertainties for long time series differ significantly, almost by a factor of 2. It is also worth noting that, based on the simulations with non-stationary noise of section 4.3, the fact that both methods result in similar uncertainties for short time series most likely suggests that random walk is absent in the examined NGL time series. Had even the smallest random walk behavior been present, minimum-entropy velocity uncertainties would have been significantly underestimated.

The second group of analyzed NGL time series includes 55 stations with visible nonlinearities but no discontinuities. Although simulations in section 4.3 showed that transients inflate minimum-entropy velocity uncertainty for long time series, comparisons to MIDAS uncertainties (Fig S5) still revealed similar patterns to those of Fig 10, namely, the minimum-entropy method senses whiter noise than the MIDAS. Since most of the nonlinear behavior occurred in the vertical component, the histogram of its velocity differences is multimodal (Fig S6), with a mean (-0.164) and RMS (0.353 mm/year), slightly larger than those of the linear time series of Fig 9. This might be because both methods define velocity differently. While minimum-entropy estimates a constant (i.e., mean) velocity, the MIDAS estimates a median velocity.

The third compared group consists of 50 stations with clear discontinuities but no visible nonlinearities, chosen in order to examine the effect of discontinuities while avoiding ambiguities related to the definition of the velocity of nonlinear time series. Of the 50 stations, 26 had experienced one to four discontinuities in the East, 37 in the North and 44 in the Up components. Before applying the minimum-entropy method, we detected the locations (dates in decimal years) of their discontinuities, using an edge detection filter called “the weak elastic string” (Saleh 1996) followed by a manual removal of false negatives. As mentioned in sections 2, 4.3 and S2, discontinuous time series are not strictly stationary but increment-stationary. To convert them to stationary, we divide them to partial series. The entropy of each partial series is computed using (9) and that of the entire series using (12), and velocity estimation proceeded as usual (as described in sections 3 and 4.1). This method is oblivious to the heights of step discontinuities because the entropies of the partial series are oblivious to their biases.

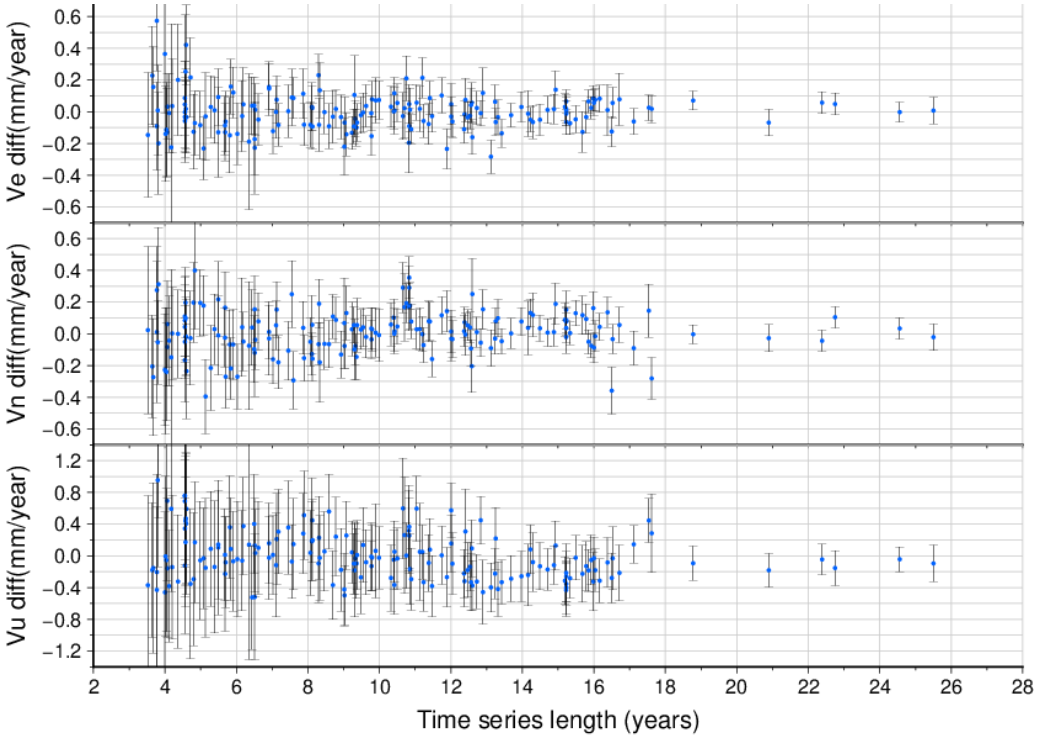

Fig 8: Velocity differences (minimum-entropy minus MIDAS) derived from 171 linear and continuous NGL position time series

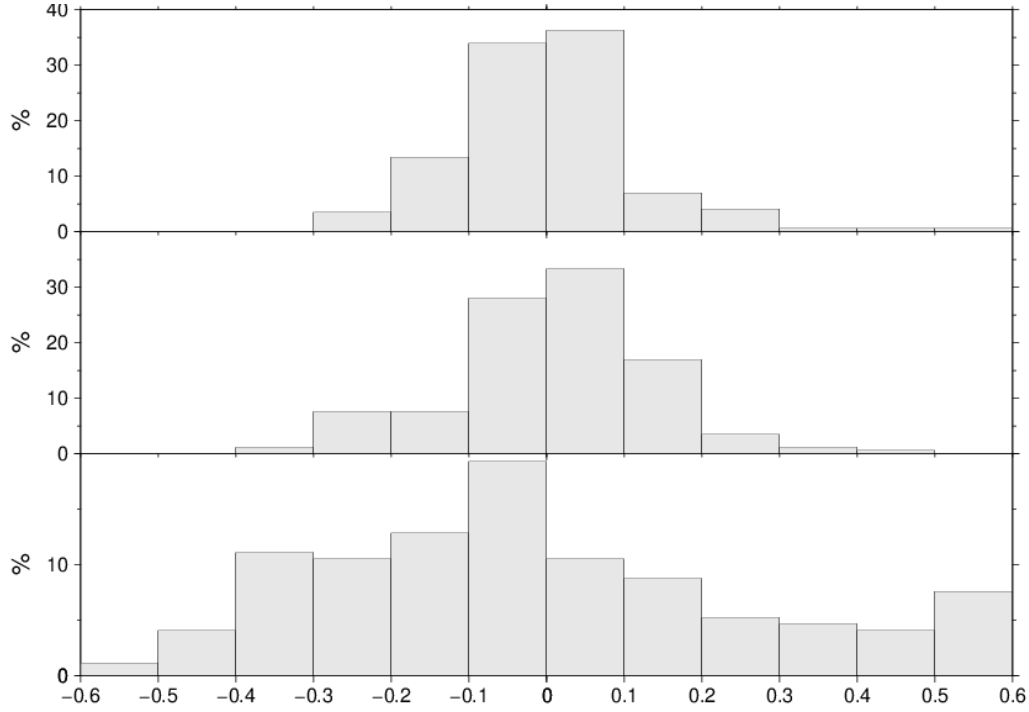

Fig 9: Histograms of velocity differences (Top: East, Middle: North, Bottom: Up) in mm/year for the 171 continuous and linear NGL series. The means of the 171 differences are -0.002, 0.009 and 0.002 mm/year and their RMS are 0.118, 0.136 and 0.322 mm/year in the East, North and Up components, respectively

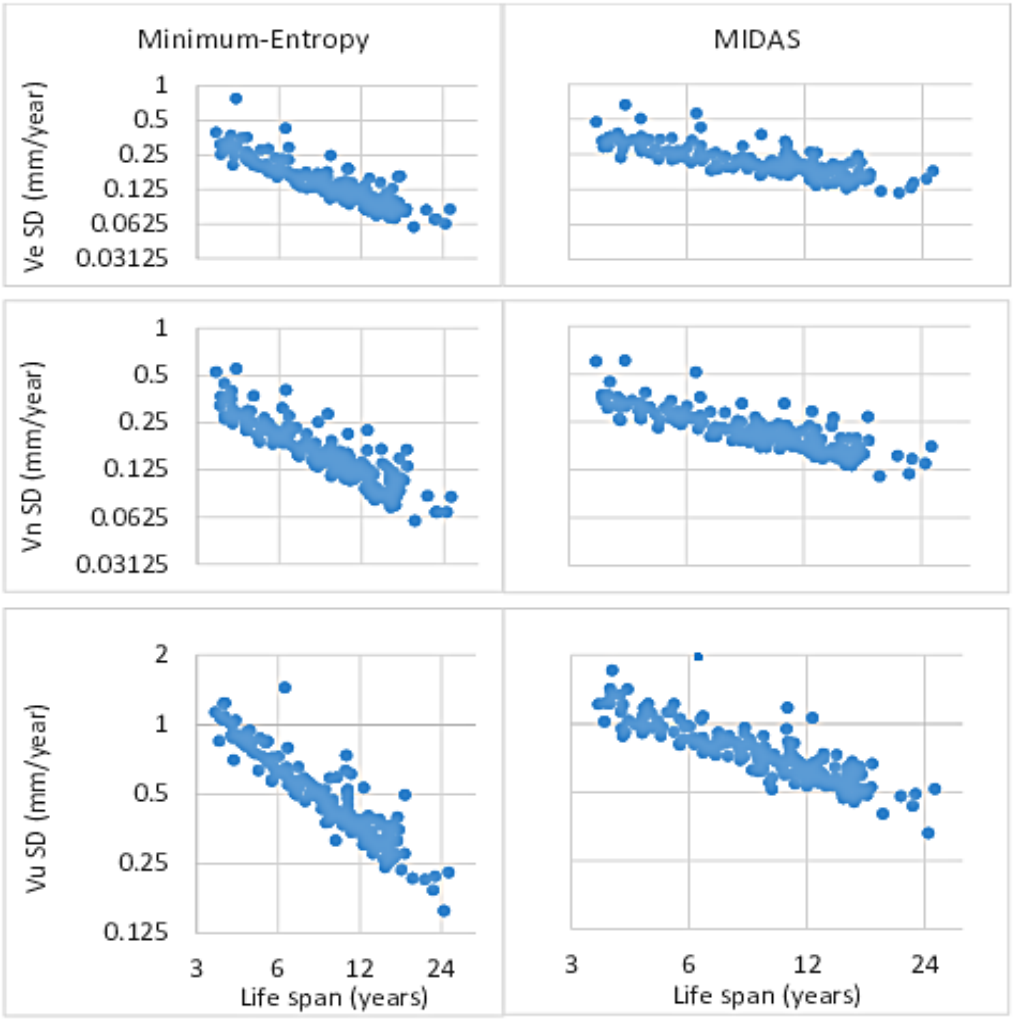

Fig 10: (Left) Minimum-entropy-derived velocity standard deviations (SDs); (Right) MIDAS-derived velocity SDs, for 171 continuous and linear NGL time series

Inaccuracy in identifying the locations of discontinuities in the pre-analysis does not affect the results because entropy is robust to outliers. However, as the number of discontinuities increases, the length of the partial series decreases and the convergence of (9) to the true entropy rates become weaker. As a consequence, the velocity bias gets worse with every added discontinuity. On the other hand, this has a negligible effect on the velocity uncertainty for sufficiently long series, as mentioned in section 4.3. Figure S7 presents histograms of velocity differences (minimum-entropy minus MIDAS) for the 50 discontinuous time series, showing a slightly wider spread (RMS of 0.187, 0.196 and 0.315 mm/year for the E, N and U components, respectively) than in the linear continuous group of Fig 9. Figure S8 presents the velocity uncertainties as a function of lifespan confirming that, also in discontinuous time series, minimum-entropy reveals a smaller degree of temporal correlations within the NGL time series than the MIDAS method does. Figures S7 and S8 present only the components of the time series which were affected by the disturbance that caused the discontinuities (26 in the East, 37 in the North and 44 in the Up series).

6. Summary

Information theory has been instrumental in the development of digital communications, data compression, genome mapping and econometrics. In this work, we apply it for simplifying one aspect of geodetic time series analysis, namely, optimal velocity estimation. For an iid time series, entropy, the most important information-theoretic quantity, is a unique measure of the series irreducible true uncertainty, beyond which there is no simplification or "compression". For correlated time series, entropy is replaced by the "entropy rate", the average entropy of the series when "compressed" to become iid. The distinction between entropy and "entropy rate" is often dropped for conciseness, including in our paper. For sufficiently long (e.g., $N \gg 30$), near-stationary time series, entropy can be computed efficiently and nonparametrically based on the probabilities, i.e., the true variability, within the series rather than on norms of its elements. Therefore, while the entropy of a position time series senses all its stochastic properties, it is theoretically unaffected by its additive deterministic content, such as its reference position, the heights of its step discontinuities, constant periodic and transient content. In reality however, these insensitivities to "constant" contents of the time series have limitations. Nevertheless, because entropy is based on the probabilities (or differences) rather than on the value of the residual positions, it significantly dilutes the effects of smooth, long periodic and slowly varying (relative to Δt) content of the series. These properties make entropy a good candidate tool for velocity estimation from position time series, especially in the presence of colored near-stationary noise.

We proposed a nonparametric minimum-entropy velocity estimation method from geodetic position time series, verified its viability and efficiency based on synthetic data, applied it to hundreds of NGL position time series and compared the results to those of the MIDAS algorithm. We demonstrated based on synthetic and real data that: (1) Minimum-entropy leads to an accurate velocity and its best realistic variance in the presence of an unknown noise; (2) Is insensitive to additive "constant" content of the time series such as a reference position, constant sinusoids and transients. While the addition of a sinusoid of a typical amplitude (~7 mm) to a geodetic position time series increases its variance by as much as 35%, it increases its entropy by only 5%, has no effect on the estimated velocity and only an insignificant effect on velocity uncertainty; (3) For sufficiently long time series, a transient does not affect the estimated velocity except by slightly inflating the velocity uncertainty, but this increase gets worse with increasing series length; (4) Step discontinuities do not affect minimum-entropy velocity uncertainty, but do degrade its estimated velocity value, and the degradation decreases with increasing length; (5) Minimum-entropy leads to accurate velocity when the noise is stationary. However, the presence of slightly non-stationary noise, although hardly affects the derived velocity, leads to significant underestimation of velocity uncertainty; (6) Under no circumstances does the proposed method lead to variances that decay as 1/N; (7) Minimum-entropy is efficient (a non-optimized, non-parallelized short and straightforward software takes ~1.5 CPU to process 16 years of daily data) because it avoids, on the one hand, covariance matrices, determinants and eigen value analysis and on the other hand the modeling and estimation of deterministic time series contents; (8) Comparisons to the MIDAS algorithm revealed that minimum-entropy and MIDAS derived velocity uncertainties are similar for short time series (< ~5 years) but differ by a factor of 2 for long time series.

Acknowledgements

We thank NGS for supporting this work and John Galetzka for allowing its completion. We thank professor Duncan Agnew and two anonymous reviewers for comments that focused this paper and improved its presentation. Thanks to NGS colleagues who reviewed this paper in-house and to colleagues who continued to answer our questions over the years.

Author contribution statement

SW provided simulated noise series and RB facilitated the extraction of the NGL data and the MIDAS software and edited initial versions of the manuscript. As long-time experts on GNSS time series analysis, SW and RB oversaw the soundness of ideas, logic and results. JS did the rest.

Data availability

We downloaded the NGL time series from http://geodesy.unr.edu/gps_timeseries/tenv/IGS14. Section S5 in the supplementary information lists the 4-character ID of all downloaded stations. The MIDAS software was downloaded from the NGL website http://geodesy.unr.edu/MIDAS_release and compiled on NGS computers. Our little software and its short documentation are included with the supplementary material.

References

Agnew D (1992) The time-domain behavior of power-law noises. Geophys Res Lett 19:333–336. DOI: 10.1029/91GL02832

Alevizakou EG, Siolas G, Pantazis G (2018) Short-term and long-term forecasting for the 3D point position changing by using artificial neural networks. ISPRS Int J Geo Inf 7(3):86. DOI: 10. 3390/ijgi7030086

Amiri-Simkooei AR, Tiberius CCJM, Teunissen PJG (2007) Assessment of noise in GPS coordinate time series: Methodology and results, J. Geophys. Res., 112, B07413, DOI: 10.1029/2006JB004913.

Amiri-Simkooei AR (2009) Noise in multivariate GPS position time-series. J Geod 83, 175–187. DOI: 10.1007/s00190-008-0251-8.

Beirlant J, Dudewicz E, Gyor L, van der Meulen EC (1997) Nonparametric entropy estimation, An overview. Inter Jour Math Statist Sci, 6 (1):17-39.

Bevis M, Bedford J, Caccamise D (2020) The art and science of trajectory modelling. In: Montillet JP and Bos M (eds) Geodetic Time Series Analysis in Earth Sciences. Springer Geophysics. DOI: 10.1007/978-3-030-21718-1.

Blewitt G, Lavallée D (2002) Effect of annual signals on geodetic velocity. J Geophys Res 107(B7):ETG9-1–ETG 9-11. DOI: 10.1029/2001JB000570

Blewitt G, Kreemer C, Hammond WC, Gazeaux J (2016) MIDAS robust trend estimator for accurate GPS station velocities without step detection, J. Geophys. Res. Solid Earth, 121, 2054–2068, DOI:10.1002/2015JB012552.

Blewitt G, Hammond WC, Kreemer C (2018) Harnessing the GPS data explosion for interdisciplinary science. Eos, 99. DOI:10.1029/2018EO104623.

Bos MS, Fernandes RMS, Williams SDP, Bastos L (2008) Fast error analysis of continuous GPS observations, J. Geod., 82(3), 157–166. DOI:10.1007/s00190-007-0165-x

Chen Q, van Dam T, Sneeuw N, Collilieux X, Weigelt M, Rebischung P (2013) Singular spectrum analysis for modeling seasonal signals from GPS time series, J. Geodyn., 72, 25-35. DOI: 10.1016/j.jog.2013.05.05

Cover TM, Thomas JA (2006) Elements of information theory, 2nd edition, John Wiley and Sons, New York.

Cucci DA, Voirol L, Kermarrec G, Montillet JP, Guerrier S (2023) The Generalized Method of Wavelet Moments with eXogenous inputs: a fast approach for the analysis of GNSS position time series. J Geod 97, 14. DOI: 10.1007/s00190-023-01702-8

Didova O, Gunter B, Riva R, Klees R, Roese-Koerner L (2016) An approach for estimating time-variable rates from geodetic time series. J Geod 90, 1207–1221 (2016). DOI: 10.1007/s00190-016-0918-5

Dmitrieva K, Segall P, DeMets C (2015) Network-based estimation of time-dependent noise in GPS position time series. J. Geod, 89, 591–606. DOI: 10.1007/s00190-015-0801-9

Dong D, Fang P, Bock Y, Cheng MK, Miyazaki S (2002) Anatomy of apparent seasonal variations from GPS-derived site position time series. J Geophys Res. DOI: 10.1029/2001JB000573

Dong D, Fang P, Bock Y, Webb F, Prawirodirdjo L, Kedar S, Jamason P (2006) Spatiotemporal filtering using principal component analysis and Karhunen-Loeve expansion approaches for regional GPS network analysis. J. Geophys. Res, 111, B03405. DOI: 10.1029/2005JB003806

Engels O (2020) Stochastic modelling of geophysical signal constituents within a Kalman filter framework. In: Montillet JP, Bos M (eds) Geodetic time series analysis in earth sciences. Springer Geophysics. DOI: 10.1007/978-3-030-21718-1.

Floyd MA and Herring TA (2020) Fast statistical approaches to geodetic time series analysis. In: Montillet JP, Bos M (eds) Geodetic time series analysis in earth sciences. Springer Geophysics. DOI: 10.1007/978-3-030-21718-1.

Gao W, Li Z, Chen Q, (2022) Modelling and prediction of GNSS time series using GBDT, LSTM and SVM machine learning approaches. J Geod 96, 71. DOI:10.1007/s00190-022-01662-5

Gobron k, Rebischung P, de Viron O, Demoulin A, Van Camp M (2022) Impact of offsets on assessing the low-frequency stochastic properties of geodetic time series. J Geod 96 (7), 46. DOI: 10.1007/s00190-022-01634-9

Gobron K, Rebischung P, Van Camp M., Demoulin A, de Viron O (2023) Influence of aperiodic non-tidal atmospheric and oceanic loading deformations on the stochastic properties of global GNSS vertical land motion time series. J Geophys Res, 126. DOI: 10.1029/2021JB022370

Gualandi A, Serpelloni E, Belardinelli ME (2016) Blind source separation problem in GPS time series. J Geod 90, 323–341 (2016). DOI: 10.1007/s00190-015-0875-4

Hackl M, Malservisi R, Hugentobler U, Wonnacott R (2010) Estimation of velocity uncertainties from GPS time series: Examples from the analysis of the South African TrigNet network. J. Geophys. Res., 116, B11404, DOI:10.1029/2010JB008142

Ji KH, Herring T (2011) Transient signal detection using GPS measurements: transient inflation at acutan volcano, Alaska during early 2008, Geophys Res Lett, 38:L06307. DOI: 10.1029/2011GL046904

Klos A, Olivares G, Teferle FN, Hunegnaw A, Bogusz J (2018) On the combined effect of periodic signals and colored noise on velocity uncertainties. GPS Solut 22, 1. DOI: 10.1007/s10291-017-0674-x

Langbein J, Johnson H (1997) Correlated errors in geodetic time series: Implications for time-dependent deformation, J. Geophys. Res., 102, 591-604. DOI: 10.1029/96JB02945

Mao A, Harrison GA, Dixon TH (1999) Noise in GPS coordinate time series. J Geophys Res 104(B2):2797–2816. DOI: 10.1029/1998JB900033.

Montillet JP and Bos MS (2020) Geodetic time series analysis in earth science. Springer Geophysics. DOI: 10.1007.978-3-030-21718-1.

Olivares-Pulido G, Teferle FN, Hunegnaw A (2020) Markov chain monte carlo and the application to geodetic time series analysis. In: Montillet JP, Bos M (eds) Geodetic time series analysis in earth sciences. Springer Geophysics. DOI: 10.1007.978-3-030-21718-1.

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) Numerical recipes 3rd edition: The art of scientific computing. Cambridge University Press, New York.

Ray J, Altamimi Z, Collilieux X, van Dam T (2008) Anomalous harmonics in the spectra of GPS position estimates. GPS Solut 12(1):55–64. DOI:10.1007/s10291-007-0067-7.

Rice JA (1988) Mathematical statistics and data analysis, Wadsworth & Brooks/Cole advanced books & software, Pacific Grove, California.

Saleh J (1996) The weak elastic string and some applications in geodesy. J Geod 70:203–213. DOI:10.1007/BF00873701.

Santamaria-Gómez A, Bouin MN, Collilieux X, Wöppelmann G (2011) Correlated errors in GPS position time series: Implications for velocity estimates. J Geophys Res 116:B01405. DOI:10.1029/2010JB007701.

Santamaría-Gómez A, Ray J (2021). Chameleonic noise in GPS position time series. J Geophys Res: Solid Earth, 126, e2020JB019541. DOI: 10.1029/2020JB019541

Shannon C (1948) A mathematical theory of communication, Bell Sys Tech Journal, 27:379-423.

Stone JV (2015) Information Theory, a tutorial introduction, Sebtel Press, Sheffield, UK

Souza E, Monico J (2004) Wavelet shrinkage: high frequency multipath reduction from GPS relative positioning. GPS Solut 8(3):152–159. DOI: 10.1007/s10291-004-0100-z

van Dam T, Wahr J, Milly PCD, Shmakin AB, Blewitt G, Lavallée D, Larson KM (2001) Crustal displacements due to continental water loading. Geophys Res Lett 28:651–654. DOI:10.1029/2000GL012120.

Vasicek O (1976) A Test for Normality Based on Sample Entropy, J Royal Statist Soc, B38, 54.

Wallis KF (2006) A note on the calculation of entropy from histograms. Technical report, Dept. of Economics, University of Warwick, UK.

Wang J, Jiang W, Li Z, Lu Y (2021) A new multi-scale sliding window LSTM framework (MSSW-LSTM): a case study for GNSS time-series prediction. Remote Sens 13(16):3328. DOI: 10.3390/rs13163328

Wessel P, Smith WHF (1998) New improved version of generic mapping tools released, Eos Trans. AGU, 79(47):579. DOI:10.1029/98EO00426

Williams SDP (2003) The effect of coloured noise on the uncertainties of rates estimated from geodetic time series. J Geod 76:483–494. DOI:10.1007/s00190-002-0283-4

Williams SDP, Bock Y, Fang P, Jameson P, Nikolaidis R, Prawirodirdjo L, Miller M, Johnson D (2004) Error analysis of continuous GPS position time series. J Geophys Res, 109: B03412. DOI: 10.1029/2003JB002741

Williams SD (2008) CATS: GPS coordinate time series analysis software. GPS Solut 12(2):147–153

Wu H, Li K, Shi W, Clarke K, Zhang J, Li H (2015) A wavelet-based hybrid approach to remove the flicker noise and the white noise from GPS coordinate time series. GPS Solut 19, 511–523. DOI: 10.1007/s10291-014-0412-6

Zhang J, Bock Y, Johnson H, Fang P, Williams S, Genrich J, Wdowinski S, Behr J (1997) Southern California permanent GPS geodetic array: Error analysis of daily position estimates and site velocities. J Geophys Res 102(B8):18035–18055. DOI: 10.1029/97JB01380