Newcomb-Benford Law as a generic flag for changes in the derivation of long-term solar terrestrial physics timeseries

A.M. Benedito Nunes[1⋆], J. Gamper[1⋆], S.C. Chapman[1,2,3†], M. Friel[4], J. Gjerloev[4,5]
[1]*Centre for Fusion, Space And Astrophysics, Physics Department, University of Warwick, UK*
[2]*Department of Mathematics and Statistics, University of Tromso, Norway*
[3]*International Space Science Institute, Bern, Switzerland*
[4]*Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA*
[5]*Department of Physics and Technology, University of Bergen, Bergen, Norway*

Accepted 23/08/2023. Received 05/08/2023; in original form 22/06/2023

ABSTRACT

The Newcomb-Benford Law (NBL) prescribes the probability distribution of the first digit of variables which explore a broad range under conditions including aggregation. Long-term space weather relevant observations and indices necessarily incorporate changes in the contributing number and types of observing instrumentation over time and we find that this can be detected solely by comparison with the NBL. It detects when upstream solar wind magnetic field OMNI High Resolution (HRO) Interplanetary Magnetic Field incorporated new data from the WIND and Advanced Composition Explorer (ACE) spacecraft after 1995. NBL comparison can detect underlying changes in the geomagnetic Auroral Electrojet (AE) index (activity dependent background subtraction) and the SuperMAG Electrojet (SME) index (different station types) that select individual stations showing the largest deflection, but not where station data are averaged, as in the SuperMAG Ring Current (SMR) index. As composite indices become more widespread across the geosciences, the NBL may provide a generic, data processing independent flag indicating changes in the constituent raw data, calibration or sampling method.

Key words: Newcomb-Benford Law – Solar Wind – Geomagnetic Indices – Data Methods – Calibration–Observational uncertainty

1 INTRODUCTION

Benford's Law, also known as the Newcomb-Benford Law (NBL) (Newcomb 1881; Benford 1938), prescribes the probability distribution of the first digit of numbers from large sequences under conditions (see Berger & Hill (2021) and refs. therein) that can include scale and base invariance (Pietronero et al, 2001), aggregation, and the absence of a cut-off (Nigrini 2000). Products of random samples from continuous distributions converge to the NBL (Hill 1995). The NBL gives the probability of digit d being the first digit of a standard form number in the sequence as $P(d) = \log_{10}(\frac{d+1}{d})$, so that digits $d = 1$ and 2 occur at around 30.1% and 17.61% of the time, respectively, whereas $d = 9$ occurs only 4.58% of the time. Benford (1938) demonstrated it in a wide range of domains including physical constants and physical and societal data. It has been found to apply in a broad range of observations of physical systems (Sambridge et al. 2010) and in the social (Mir 2012; Pietronero et al, 2001), and biological (Pröger et al. 2021) sciences. In particular, it has been proposed as a means to detect 'anomalies', that is, changes in time sequences of data, for example providing a means to detect earthquakes (Diaz et al. 2014; Sambridge et al. 2010).

Space weather can have significant impact over a wide range of technological systems including power grids, aviation, satellites and communications (Hapgood 2019; Oughton et al. 2016, 2017; Knipp et al. 2016; Knipp et al 2021) . In common with studies across the geophysical sciences, the study of space plasma physics and the climatology of space weather (Pulkkinen 2007) requires long term space and ground-based parameters and indices that necessarily aggregate multiple observations, the details of which can change with time. In this first application to space weather parameters and indices, we show that the NBL can detect changes in the instrumentation and calibration underlying long-term geophysical records, solely from the processed data records.

In space weather, as in other fields such as climate change, it is critical to be able to verify that any observed secular change is not a result of changes in how the data record is constructed. As composite indices are becoming more widespread across the geosciences, the NBL may provide a generic data flag indicating changes in the constituent raw data, calibration or sampling method. Magnetic field observations, both from satellites in-situ and from ground based magnetometers, are an essential component of the modelling and prediction of space weather. Geophysical data is often multipoint in character, with several hundred station observations sampling time-varying fields across the earth's surface. It is common practice across the geosciences to construct indices that capture relevant aspects of a multipoint-sampled spatial field, that is, indices based for example on the average, the variance, a threshold crossing, or an extremum across multiple station data.

An observation of a plasma parameter such as the magnetic field,

⋆ Joint first authors
† E-mail: S.C.Chapman@warwick.ac.uk

either in-situ in space, or on the ground, includes various stages of processing of the raw data, involving calibration, removing offsets or background fields, coordinate rotation, and interpolation onto a common, uniform time-base. Geomagnetic indices are derived by combining data from multiple ground based magnetometer stations. The physical processes underlying these observations are also often aggregating, or multiplicative processes such as mixing and turbulence. Given sufficient dynamic range, and in the absence of a cut-off, the NBL might be expected to be followed by both solar wind parameters and geomagnetic indices, at least to some precision.

The station locations, instrumentation, calibration and processing required to derive observed parameters naturally change with time. In this paper we will show that quite subtle changes in the derivation of a parameter or index can be reflected in a statistically significant change in how closely the final data product or index follows the NBL. Our results suggest that the NBL can provide a flag that indicates if changes have occurred in the details of how long-term observations and indices are derived. This data flag is generic, acting solely on the fully processed data record without any information on the details of how the data product or index is derived. In space weather, as in other fields such as climate change, it is critical to be able to verify that any observed secular change is not a result of changes in how the data record is constructed.

The solar wind magnetic field upstream of earth has been observed in-situ around L1 since the 1960s (Papitashvili et al. 2020) by a succession of satellites. Ground-based magnetometers comprise the Auroral Electrojet AE (Davis & Sugiura 1966), SuperMAG Electrojet SME (Newell and Gjerloev 2011) and SuperMAG Ring current SMR (Newell & Gjerloev 2012) geomagnetic indices, at 1 minute resolution from 1981. Unlike AE, the number of stations that comprise SME and SMR has increased by over an order of magnitude, and some more recent stations have different instrumentation. AE and SME have different baseline subtraction procedures but are both extremal in the sense that they are both comprised of data from the stations with the largest deflections, whereas SMR is based on a multi-station average. We will see that whilst all these quantities follow the NBL to high precision, quite subtle changes in the underlying data and processing can be detected simply by changes in how closely some parameters and indices follows the NBL. It should be emphasised that the closeness to which a quantity follows the NBL is not an indicator of relative quality or precision per-se. Rather, it offers an indicator that there has been a change in the underlying raw observations and the process by which the final parameter or index is derived. To be detectable, any change in the underlying data methodology would need to generate a statistically significant change in how well the NBL is followed, hence the lack of change in the NBL in a data sequence does not guarantee that the underlying methodology is preserved.

This paper is organised as follows. In section 2, we describe the datasets and detail the method to estimate the fit-parameter which quantifies the precision to which the NBL is followed by a finite length sequence of data. In section 3, we estimate how the fit parameter changes over the full records of the OMNI High Resolution Interplanetary Magnetic Field (IMF), and the AE, SME and SMR geomagnetic indices. We conclude and discuss the prospect of a generic data flag in section 4.

2 METHODS

2.1 The datasets

We focus on parameters that monitor the state of the solar wind upstream of earth, and indices that monitor geomagnetic activity. As well as being central to space weather, these multi decadal time-series provide a test-bed to see under what circumstances changes in their construction can be detected solely by applying the NBL to the processed data record.

A series of solar wind monitors located at the L1 point upstream of the Earth have provided solar wind parameters including the Interplanetary Magnetic Field (IMF). We will consider the IMF for the time interval [1981 - 2022] inclusive at 1 minute resolution as extracted from NASA/GSFC's Modified (Level-3) High Resolution OMNI (HRO) data set through OMNIWeb (Papitashvili et al. 2020). The parameters are interpolated onto a uniform timebase and mapped to the Earth's bow shock nose. The HRO 1 minute resolution IMF is derived from observations from a series of satellites, and from 1st January 1995 there was a transition from IMP8 only to IMP8, WIND and later, other satellites such as ACE. The data processing method was also modified in 1995.

Auroral indices are designed to monitor the high latitude ionospheric electrojets. The Auroral Electrojet (AE) is the difference between the Auroral Upper (AU) and the Auroral Lower (AL) indices (Davis & Sugiura 1966). AU and AL are derived from the 1 minute resolution geocentric solar magnetospheric coordinates (GSM) $\hat{e}$ local magnetic east field component from one of 12 high latitude ground based magnetometer stations in the northern hemisphere. The index takes the value of the data from the stations which at that instant have the largest positive (AU) and largest negative (AL) deflection. Recently, a SuperMAG (Gjerloev 2012) analog of AE, SME, has been derived from the full set of available stations between +40 and +80 degrees in northern hemispheric latitude (Newell and Gjerloev 2011). We will consider AE for the interval [1981-2018] inclusive and SME for the interval [1981-2022] inclusive.

Ring-current indices are based on averages over multiple low-latitude station observations. Our study relies on a statistical analysis, therefore rather than focus on the 1 hour time resolution DST index (Sugiura 1964), we will consider the 1 minute resolution SuperMAG (Gjerloev 2012) ring-current index SMR (Newell & Gjerloev 2012). SMR is derived from all available magnetometer stations within ±50 degrees of latitude. Following a latitudinal correction, the GSM $\hat{n}$ local magnetic north displacement is first averaged over stations within four 6 hour wide local time windows to give the SMR-00, SMR-06, SMR-12 and SMR-18 local indices. These four local indices are then averaged to give SMR. We consider 1 minute SMR for the interval [1981-2022] inclusive.

Studies of the variations caused by electric currents flowing in the ionosphere and magnetosphere require a subtraction of the dominant and slowly varying Earth main field from the constituent magnetometer observations. The AE index baseline is determined from identified quietest days (Davis & Sugiura 1966), whereas the SuperMAG indices employ an automated procedure that removes the yearly trend as well as daily variation (Gjerloev 2012). The number of stations comprising the AE index does not change over the interval that we will consider here. The SME and SMR indices draw upon a set of SuperMAG collated stations where there is an increase in the number, and changes to the type, of stations over time.

Our analysis will also utilize yearly mean total sunspot number (SSN) version 2.0 and dates of solar maxima and minima provided by the World Data Centre-Sunspot Index and Long-term Solar Observations (WDC-SILSO). Increased dynamic range of the data record

may improve the NBL fit precision, which might be expected to come into play during active intervals of the solar cycle.

2.2 Testing data records with the Newcomb-Benford Law

We quantify the statistical goodness of fit to the NBL of the distribution of first digits in samples taken successively in time from the data record to be tested. The sample length is optimised to test if any secular change revealed by the data record is of physical origin, or is due to a change in the underlying instrumentation, calibration or processing. For space weather parameters and indices an optimal sample over which to estimate first digit distributions is 1 year, since it is (i) long enough to provide a statistically significant sample (at 1 minute resolution, 525600 data points); (ii) is a sufficiently long time interval for the system to explore its full dynamics (quiet times, substorms and storms, coronal mass ejections) and (iii) is a timescale which is short compared to the 11 year cycle of solar activity and changes across solar cycles. Any given year-long sample may contain data gaps, and we also exclude records that read zero, this is reflected in bootstrap estimated confidence intervals described below.

A variety of statistical estimators have been used to test for the NBL (e.g. Durtschi et al. (2004); Druicá et al. (2018); Sambridge et al. (2010); Mir (2012); Pröger et al. (2021)). We performed a systematic comparison of three commonly used estimators in order to select an estimator that best supports sample length optimisation, specifically, it has fast convergence with increasing sample size. Using notation that leading digit $d = 1..9$ has a theoretical occurrence frequency $T_d = \log_{10}(\frac{d+1}{d})$ from the NBL and observed occurrence frequency O_d in the data sequence, the three methods are as follows:
the Chi-squared test (Chi) per degree of freedom: (e.g. Peren (2022)):

$$\theta_{Chi^2} = \sum_{d=1}^{9} \frac{(T_d - O_d)^2}{T_d} \tag{1}$$

the Mean Absolute Deviation (MAD) as defined in Druicá et al. (2018); Pröger et al. (2021) (see also Bacon (2012)):

$$\theta_{MAD} = \frac{1}{9} \sum_{d=1}^{9} \frac{|O_d - T_d|}{N} \tag{2}$$

and the Root Mean Square Error (RMSE): (e.g. Cichosz (2015))

$$\theta_{RMSE} = \sqrt{\frac{1}{9} \sum_{d=1}^{9} (T_d - O_d)^2} \tag{3}$$

The Fibonacci sequence asymptotically obeys the NBL in the large N limit (Washington 1981) and we use it to estimate a lower bound for θ as a function of the length of the data sample for each of the three estimators. The left panel of Figure 1 overlays the 1st digit distribution for the first $N = 525600$ Fibonacci numbers, a sample length equivalent to one 'year' of 1 minute data records, on the NBL distribution $T_d = \log_{10}(\frac{d+1}{d})$. The right panel plots the fit parameter θ obtained using the different estimators as a function of N. We can see that the MAD and Chi-squared estimators both perform well, they have a large dynamic range and are converging faster with increasing N. We will use the MAD estimator here. The MAD lower bound on the fit parameter (estimated from the Fibonacci sequence) for a sequence that is the length to be tested here, that is, 1 year of minute observations or N=525600, is $\theta \approx 10^{-9}$.

For the data analysis to follow we estimate 95% confidence intervals for the fit parameter θ using the stationary bootstrap (Politis & Romano 1994). The bootstrap method estimates uncertainties by randomly re- sampling with replication from the data sample multiple times. It provides a reliable uncertainty estimate under conditions of weak stationarity, and where the sample means form a stable distribution. The optimal length of the bootstrapping block was obtained using the method outlined in (Politis & White 2004). The stationary bootstrap and block length selection algorithm were implemented using the python library *arch* (Sheppard 2021). The Python function *arch.bootstrap.StationaryBootstrap.conf_int*, used to calculate the confidence interval, required the following inputs: seed, number of bootstrap replications, method, size, and sampling which we set to the following values, respectively: 66, 1000, "basic" (also known as empirical bootstrap), 0.95, nonparametric. We checked the validity of the bootstrap estimates by examining the distribution of the fit parameter obtained from the bootstrap re-samples. We discarded estimates of the confidence interval where the distribution of the fit parameter for the bootstrap re-samples was not single-peaked, as well as where the confidence interval did not converge.

3 RESULTS

3.1 Solar wind Interplantary Magnetic Field at L1

The HRO IMF dataset provides a test case to see if the NBL can detect changes in instrumentation and processing in a single-point observational time series. In 1995 there was a change in the contributing satellites to OMNI HRO and to the processing procedure (https://omniweb.gsfc.nasa.gov/html/HROdocum.html see also (Alterman 2022)). Prior to 1995, the underlying observations were from IMP8 only, post 1995, they also included recently launched WIND and later ACE.

Figure 2 plots the MAD estimated fit parameters θ for non-overlapping year-long samples of 1 minute resolution HRO IMF geocentric solar ecliptic (GSE) x,y, and z components, along with 95% confidence intervals. The NBL is followed reasonably closely, the fit parameter $\theta < 10^{-4}$ across the entire record. In 1995 there is a step-change improvement in the fit parameter θ which drops by about a factor of three, a step-change which significantly exceeds the 95% confidence intervals. The fit parameter is constant thereafter. This step-change in θ in 1995 is coincident with the inclusion of WIND and later, ACE observations in the OMNI HRO record.

The dynamic range explored by the IMF will increase in more disturbed solar wind which over year-long samples will correlate with the overall level of solar activity. We can check if this is a factor in how closely the IMF record follows the NBL. The yearly averaged SSN is overplotted on Figure 2; it shows that the precision to which the NBL is followed is not sensitive to the overall level of activity.

As well as a change in instrumentation, the coverage also improves with time across the OMNI HRO record. Figure 2 plots estimates of θ and its bootstrap uncertainties for annual samples, but as the record contains data gaps, each of these samples may not contain a complete year-long interval of data. Smaller sample sizes will result in larger bootstrap estimated confidence intervals. Before 1986 the average annual coverage is 16.55%, θ is more variable and the confidence intervals are relatively large compared to later times. Between 1986-1995 the annual coverage is 28.94%, and post 1995 it is 91.27%. The availability of WIND and ACE resulted in significantly fewer data gaps per year on average, and the bootstrap confidence intervals can be seen to be significantly smaller post 1995. However, the step-change in θ in 1995 cannot be solely attributed to an improvement in coverage as in that case, the confidence limits pre- and post- 1995

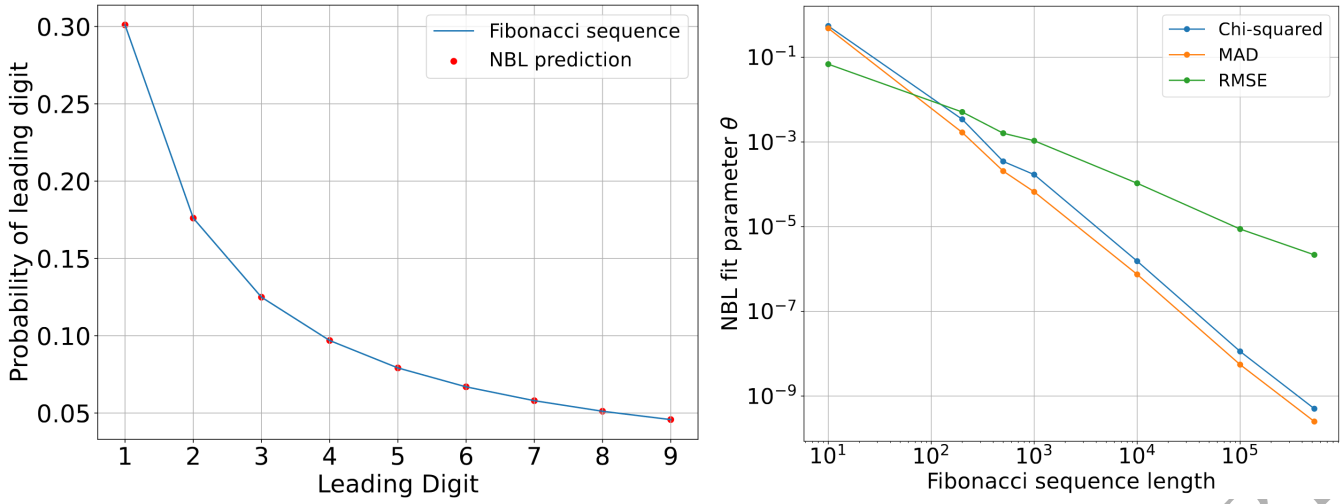

Figure 1. Left: the first digit distribution of the first 525600 numbers in the Fibonacci sequence (blue line) overplotted on the NBL distribution (red circles). Right: fit parameter of finite Fibonacci sequences plotted as a function of length of the sequence $N = [10, 200, 500, 1000, 10000, 100000, 525600]$ for three estimators: Chi-squared(blue), MAD (orange) and RMSE (green).

would overlap. This suggests that the step change in the NBL fit parameter θ does discriminate the inclusion of new instrumentation from WIND and later ACE in the OMNI HRO record post 1995.

3.2 Geomagnetic indices

Geomagnetic indices are derived from observations from multiple individual magnetometer stations. Before considering geomagnetic indices, we first investigated how closely the data from individual magnetometer stations follow the NBL. Some sample values for the NBL fit parameter of year-long samples of GSM magnetometer data with SuperMAG baseline subtraction are: Pebek [2014], $\hat{n}$ component: $\theta = 1.22 \times 10^{-6}$; Yellowknife [2001], $\hat{e}$ component: $\theta = 1.24 \times 10^{-7}$; Abisko [1990], $\hat{z}$ component $\theta = 2.06 \times 10^{-7}$.

The AE and SME auroral indices are essentially comprised of data taken from the pair of ground stations that at any time observe the maximum (positive and negative) magnetic field deflections. SMR on the other hand is a multi-station average. Estimates of the NBL fit parameter θ from non-overlapping year-long samples are plotted in Figure 3 for the SME, AE and SMR indices, with 95% confidence limits. The figure examines the effect on SME and SMR of changing station number and coverage and changes in class of magnetometer. The figure also examines the effect of different baseline removal in construction of the index by comparing SME and AE.

Panel (a) of Figure 3 plots the overall coverage provided by the ground based magnetometers collated by SuperMAG. For each year we sum over the fraction of the year that each station is taking data to obtain the total operating station-years, so that if m stations were taking data for the entire year, this would give m operating station-years. Colours discriminate a subset of stations which were introduced after 2003 which use different classes of magnetometer, these are mostly THEMIS project, and later, Magstar, CARISMA, and McMac operated stations coded as T, R, C, and M in the SuperMAG catalog (Gjerloev 2009, 2012). The coverage from all other stations are indicated by grey in panel (a) of Figure 3. On panels (b) and (d) the red line plots the annual mean of the number of stations contributing to each minute observation of SME, and of SMR, the error bar indicates the standard deviation for this mean.

The SuperMAG collated stations then provide a test dataset to see if the NBL is sensitive to (i) an increase in station number but no change in class of magnetometer, as occurs before 2003 and (ii) the inclusion of different classes of magnetometer as occurs after 2003. Figure 3 panel (b) then plots the NBL parameter θ for the SME index derived from all available SuperMAG stations (green) overplotted on the NBL parameter for SME constructed excluding the T stations (blue). The Figure then shows that increasing the number of stations, that is, the spatial coverage, for stations of the same magnetometer class, does not change the NBL fit parameter: there is no change in the NBL fit parameter between the early record, and 1996-2002, over which period the number of magnetometers has increased by an order of magnitude. However, after 2005 there is a statistically significant divergence between the NBL for SME obtained using the full set of stations (which now include the R, C, M and T stations), and with the T stations excluded.

Panel (c) of Figure 3 plots the NBL fit parameter for the AE index which is comprised of a fixed number of stations during this interval. In the first half of the AE data record there is a statistically significant correlation between the NBL fit parameter and the variation in the SSN over stronger solar cycles 22 (maximum in 1989) and 23 (maximum in 2001), it is less evident over weaker cycle 24. For SME, there is no statistically significant solar cycle variation in the NBL fit parameter over cycles 22 and 23. The AE index baseline is determined from identified quietest days (Davis & Sugiura 1966), whereas the SuperMAG indices do not use the concept of quietest days, instead, an automated procedure that removes the yearly trend as well as daily variation is employed (Gjerloev 2012). The AE baseline will therefore track the overall level of geomagnetic activity in a different manner to SME. If the quietest days around strong solar maxima are more active than the quietest days around solar minima, then a baseline determined from those most quietest days will in turn track the yearly averaged SSN. During active years, a raised baseline would then act as a low-end cut-off which would increase the value of the NBL fit parameter. It should be emphasised that both the AE and SME records follow the NBL to high precision; changes in the NBL fit parameter are nevertheless sensitive to quite small changes in the underlying magnetometers and in the baselines used.

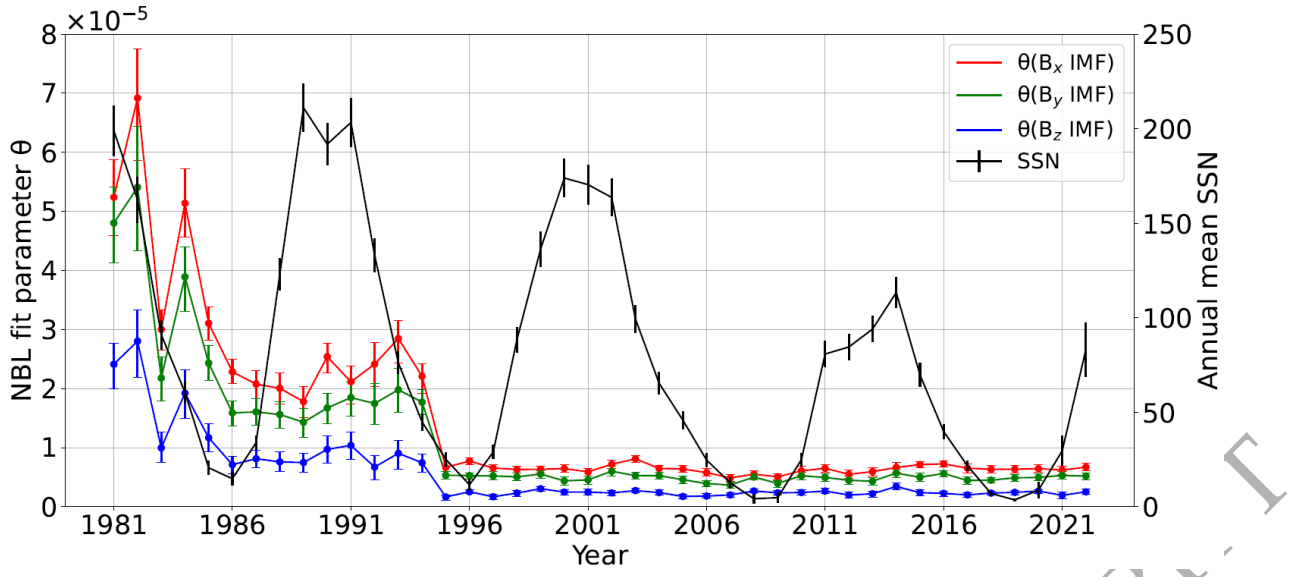

Figure 2. The MAD-estimated fit parameter θ (left ordinate) for solar wind IMF GSE $\hat{x}$ (red), $\hat{y}$ (green) and $\hat{z}$ (blue) components, estimated for 1 year non-overlapping samples, with bootstrap 95% confidence limits, are plotted versus time. Smaller fit parameter values indicate closer correspondence to the NBL first digit distribution. Yearly averages of daily sunspot number (right ordinate) is plotted (black), error bars denote the standard deviation for that year.

The NBL fit parameter for SMR is plotted in panel (d) of Figure 3, alongside the yearly averaged SSN and the total number of SuperMAG constituent stations. The SMR fit parameter is essentially constant within the bootstrap 95% confidence intervals. This suggests that the NBL fit parameter of an average over many stations is much less sensitive to changes in its constituent data, in this case, the inclusion of different instrumentation post 2006.

4 CONCLUSIONS

The Newcomb-Benford Law (NBL) prescribes the probability distribution of the first digit of standard form number sequences under conditions which include aggregation (the values arise from multiple operations) scale and base invariance, and the absence of strong truncation. Long-term parameters and indices are in widespread use across the geosciences and their constituent instrumentation and construction methodology will necessarily change over time. We have demonstrated that changes in the instrumentation and calibration underlying long-term records can be detected directly just from the processed records by how closely they follow the NBL.

We determined the precision to which the NBL is followed over time by long-term parameters and indices that are central to the monitoring of space weather. We considered non-overlapping yearly samples of the solar wind interplanetary magnetic field (IMF) monitored at L1, and the AE, SME and SMR geomagnetic indices available at minute resolution over multiple solar cycles. Our results are as follows:

(i) The OMNI (HRO) IMF, and indices AE, SME and SMR all follow the NBL to high precision.

(ii) A change in the NBL fit parameter θ for the OMNI high resolution IMF parameter occurs when the data source changes from IMP8 to include data from other spacecraft such as WIND and ACE following their launch, and the processing method was modified.

(iii) The SMR index which averages over multiple ground-based magnetometer time-series, follows the NBL to a consistent precision across changing solar activity, a ten-fold increase in the number of stations comprising the index, and the introduction of different classes of constituent magnetometer.

(iv) A change in the NBL fit parameter for the SME auroral index, which is comprised of data from stations recording the largest deflections, occurs when there is a change in the class of constituent magnetometer but not when the number of the same class of stations is increased.

(v) The AE and SME indices are both based on stations recording the largest deflections but use different baseline determination. Unlike the SME index, the AE index follows the NBL to a precision that tracks the relatively strong SSN variation of solar cycles 22 and 23, consistent with the latter using a baseline determined fom geomagnetically quietest days.

The tendency towards the NBL includes conditions (see Berger & Hill (2021) and refs. therein) of scale and base invariance (Pietronero et al, 2001), aggregation, and the absence of a cut-off (Nigrini 2000). Since products of random samples from continuous distributions converge to the NBL (Hill 1995), departure from the NBL may not be expected to be a sensitive indicator of a change simply in the amplitude of uncorrelated noise in the input to a derived parameter or index. A change in range or resolution, that is, the number of digits available to represent each data record, or a change in baseline or threshold, or saturation, can all affect how closely the NBL is followed.

These results have practical implications for the design and use of long-term parameters and indices. In space weather, as in other fields such as climate change, there is requirement to verify that any observed secular change is not a result of changes in how the data record is constructed. We have examined geophysical parameters and indices which in all cases follow the NBL to high precision. Quite subtle changes in the underlying instrumentation and differences in the subtracted baseline can be detected by the NBL in long-term records of parameters (here, the IMF) and in indices that select single time-series from the set of stations (here, auroral indices). The latter may also be expected to apply to indices that select on a high threshold, again being comprised of a few timeseries selected from

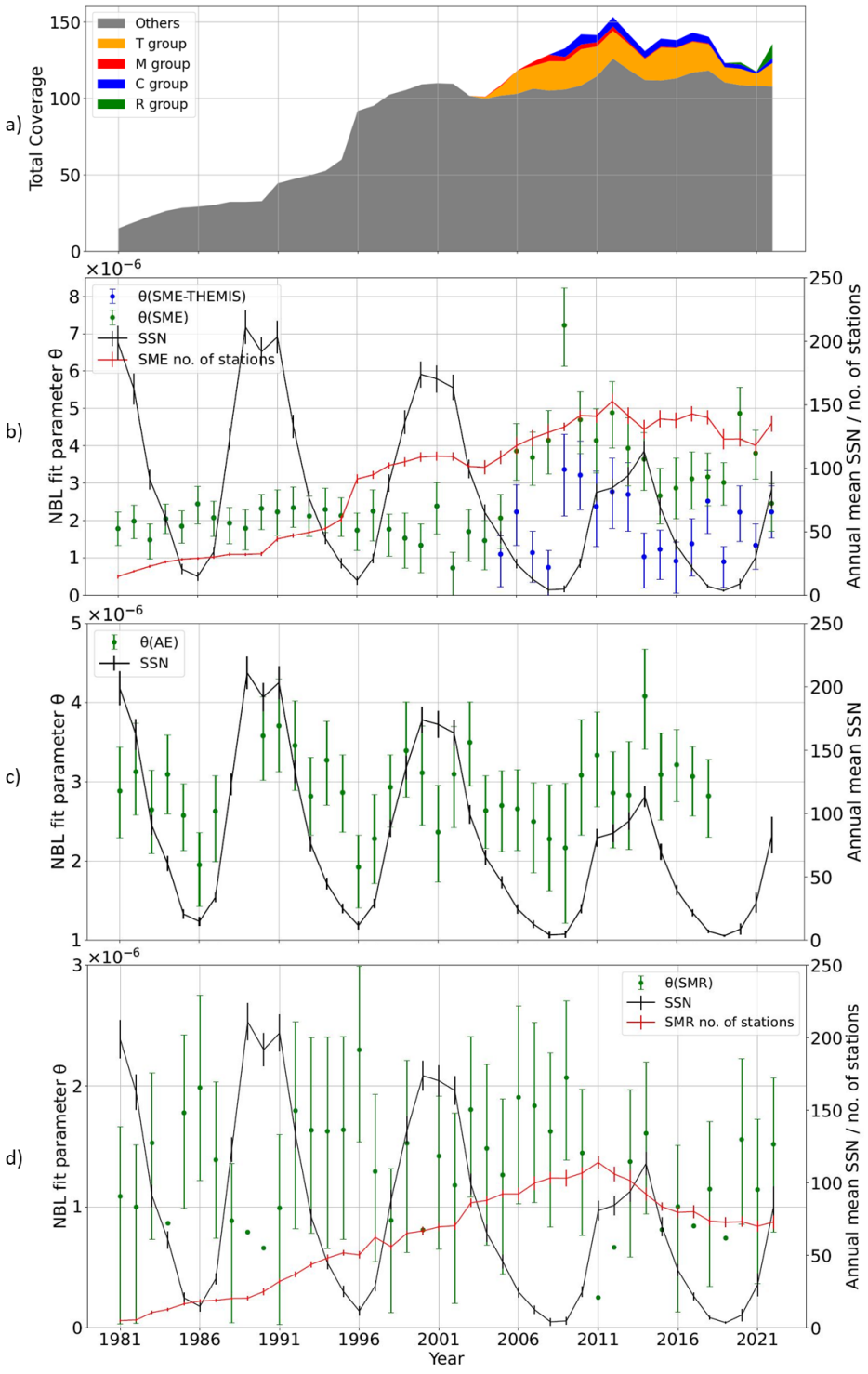

Figure 3. Panel (a): Stack plot of the coverage (total operating station-years) for different classes of SuperMAG stations. A different class of instrumentation is introduced after 2003, colours indicate specific SuperMAG station classifications. Panel (b): Left ordinate refers to the NBL fit parameter θ for non-overlapping yearly samples of the SME index. Green circles plot θ for SME derived from all stations overplotted on blue circles which plot θ for SME derived excluding T group stations. Panel (c) Left ordinate refers to the NBL fit parameter θ for non-overlapping yearly samples of the AE index (green circles). The fit parameter is not plotted for years 1988 and 1989 where there are significant data gaps in AE. Panel (d) Left ordinate refers to the NBL fit parameter θ for non-overlapping yearly samples of the SMR index (green circles). On panels (b-d), error bars plot bootstrap estimated 95% confidence interval uncertainties on the NBL fit parameter. The right ordinate refers to the yearly averaged SSN (black line), and in panels (b) and (d), to the annual mean number of all SuperMAG stations that contribute within each year (red line).

the set of observing stations. In all these cases, the NBL could provide a data flag that would indicate to the user that further investigation is needed in how a long-term parameter or index is utilised. Such a data flag would be informative without any detailed knowledge of how the parameter or index is constructed, important since parameters and indices are designed for widespread application as benchmarks of activity.

Importantly, a lack of change in how closely the NBL is followed does not in all cases guarantee that there has been no change in how the data is derived. We have found that the NBL is not sensitive to changes in the construction of indices that average or aggregate over many stations (here, ring current indices), consistent with the aggregating process driving the data records towards closer correspondence to the NBL. Furthermore, to be detectable, any change in the underlying data methodology would need to generate a statistically significant change in how well the NBL is followed, which will depend on sample size and details of the specific parameter or index.

We have found that how closely the NBL first digit distribution is followed is sensitive to changes in how parameters and indices are constructed. This is distinct from tracking physical changes in the system that they are designed to parameterize. The NBL fit parameter does not track the variation in activity (smoothed SSN), of the last four solar cycles in the IMF at L1, in SME or SMR. The distribution of solar wind parameters do show solar cycle variation (Tindale & Chapman 2016) and the top few percent of the data records of both AE and SMR also track the solar cycle (Bergin et al. 2022). Auroral indices such as AE and SME sample the ground magnetic perturbations from high-latitude current systems, the largest of which are the auroral electrojets. Auroral electrojet intensity tracks the solar cycle (Smith et al. 2017) and will have a maximum possible intensity, this is seen in auroral indices (Nakamura et al 2015). The electrojets are geographically localized, so that as the number of SME stations is increased, it is more likely that a station will be located in the vicinity of the maximum ground magnetic deflection. It has indeed been shown that the AE record systematically undersamples when compared to SME for later solar cycles (Bergin et al. 2020) as the number of stations comprising SME has increased. This change is not seen in the NBL fit parameter; for SME it does not change as the number of constituent stations is increased over an order of magnitude.

ACKNOWLEDGEMENTS

Artur Miguel Benedito Nunes and Jekaterina Gamper are co-first authors.

We thank the University of Warwick Institute for Advanced Teaching and Learning. SCC acknowledges support from ISSI via the J. Geiss fellowship and AFOSR grant FA8655-22-1-7056 and STFC grant ST/T000252/1. We acknowledge the SuperMAG collaborators: https://supermag.jhuapl.edu/info/?page=acknowledgement
We acknowledge use of NASA/GSFC's Space Physics Data Facility's OMNIWeb service, and OMNI data and the experiment teams that acquired, processed and provided the OMNI-included data, and J.H. King and N.E. Papitashvili of NASA/GSFC for creating the OMNI data set. We acknowledge the WDC for Geomagnetism, Kyoto for the provision of the AE index data. We thank the World Data Center SILSO, Royal Observatory of Belgium, Brussels for provision of sunspot data.

DATA AVAILABILITY

All data used in this study is freely available from the following sources (accessed on 30th July 2023).

SuperMAG (Gjerloev 2012) indices: https://supermag.jhuapl.edu/
The AE index from the WDC for Geomagnetism, Kyoto (Nose et al 2015) http://wdc.kugi.kyoto-u.ac.jp/wdc/Sec3.html
OMNI (Papitashvili et al. 2020) Solar wind parameters: https://omniweb.gsfc.nasa.gov/form/omni_min_def.html OMNI HRO Documentation: https://omniweb.gsfc.nasa.gov/html/HROdocum.html
SILSO Royal Observatory of Belgium, Brussels daily total sunspot number version 2.0 from 1818: http://www.sidc.be/silso/home
The dates of solar cycle maxima and minima are as determined from the smoothed sunspot number record by SILSO: http://www.sidc.be/silso/cyclesmm
Stationary bootstrap and block length selection algorithms were implemented using the Python library of Sheppard (2021): bashtage/arch. Retrieved from https://doi.org/10.5281/zenodo.7822947
The Python function *arch.bootstrap.StationaryBootstrap.conf_int* is used to calculate the confidence interval.

REFERENCES

Alterman, B. L., (2022) , Plasma Data Sources in the OMNI Database, Res. Notes AAS 6 135 doi:10.3847/2515-5172/ac7a2f
Bacon, C. R. (2012) Practical risk-adjusted performance measurement. Wiley, doi: 10.1002/9781118673621
Benford, F. (1938) The law of anomalous numbers. Proc. Am. Philos. Soc. 78 (4): 551–572
Berger, A., Hill, T.P. (2021) The mathematics of Benford's law: a primer, Stat. Methods Appl. 30, 779–795 doi:10.1007/s10260-020-00532-8
Bergin, A., S. C. Chapman, J. Gjerloev, (2020) AE, DST and their SuperMAG Counterparts: The Effect of Improved Spatial Resolution in Geomagnetic Indices, J. Geophys. Res., doi:10.1029/2020JA027828
Bergin, A., S. C. Chapman, N. Moloney, N. W. Watkins, (2022) Variation of geomagnetic index empirical distribution and burst statistics across successive solar cycles, J. Geophys. Res, doi:10.1029/2021JA029986
Cichosz (2015). Data Mining Algorithms: Explained Using R. John Wiley & Sons. doi:10.1002/9781118950951
Davis, T. N., Sugiura, M. (1966). Auroral electrojet activity index AE and its universal time variations. J. Geophys. Res., 71 , 785–801 doi:10.1029/JZ071i003p00785
Diaz, J., Gallart, J., Ruiz, M. (2014) On the ability of the Benford's law to detect earthquakes and discriminate seismic signals. Seismological Res. Lett. 86 192-201. doi:10.1785/0220140131.
Druicá, E., Oancea, B., Vâlsan, C. (2018) Benford's law and the limits of digit analysis. Intl. J. Accounting Information Systems, 31 , 75–82 doi: 10.1016/j.accinf.2018.09.004
Durtschi, C., Hillison, W., Pacini, C. (2004). The effective use of Benford's law to assist in detecting fraud in accounting data. Journal of forensic accounting research, 5 , 17–34.407
Gjerloev, J. W. (2009), A Global Ground-Based Magnetometer Initiative, EOS, 90, 230-231, doi:10.1029/2009EO270002.
Gjerloev, J. W. (2012). The SuperMAG data processing technique. J. Geophys. Res., 117, 411 doi:10.1029/2012JA017683
Hapgood, M. (2019). The great storm of May 1921: An exemplar of a dangerous space weather event. Space Weather, 17, 950–975.
Hill, T.P. (1995) A Statistical Derivation of the Significant-Digit Law. Statistical Science, 10, 354-363. doi: 10.1214/ss/1177009869
Knipp, D. J., A. C. Ramsay, E. D. Beard, A. L. Boright, W. B. Cade, I. M. Hewins, R. H. McFadden,W. F. Denig, L. M. Kilcommons, M. A. Shea, and D. F. Smart (2016), The May 1967 great storm and radio disruption event: Extreme space weather and extraordinary responses, Space Weather ,14, 614–633, doi:10.1002/2016SW001423.

Knipp, D. J., Bernstein, V., Wahl, K., Hayakawa, H., (2021) Timelines as a tool for learning about space weather storms J. Space Weather Space Clim. 11, 29 doi:10.1051/swsc/2021011

Mir, T. A. (2012). The law of the leading digits and the world religions. Physica A 416, doi: 10.1016/j.physa.2011.09.001

Nakamura, M., Yoneda, A., Oda, M., Tsubouchi, K. (2015). Statistical analysis of extreme auroral electrojet indices. Earth Planets Space, 67, 153 doi:10.1186/s40623-015-0321-0

Newcomb, S, (1881) Note on the frequency of use of the different digits in natural numbers. American Journal of Mathematics. 4 (1/4): 39–40. doi:10.2307/2369148.

Newell, P. T., Gjerloev, J. W. (2011). Evaluation of SuperMAG auroral electrojet indices as indicators of substorms and auroral power. J. Geophys. Res. 116 (A12) 422 doi: 10.1029/2011JA016779

Newell, P. T., Gjerloev, J. W. (2012). SuperMAG-based partial ring current indices J. Geophys. Res. 117 424 doi:10.1029/2012JA017586

Nigrini, M.J.: Digital Analysis Using Benford's Law. Global Audit Publications, Vancouver, B.C., Canada (2000)

Oughton, E., Copic, J., Skelton, A., Kesaite, V., Yeo, Z. Y., Ruffle, S. J., Tuveson, M., Coburn, A. W., Ralph, D. Helios (2016) Solar Storm Scenario, Cambridge Risk Framework series, Centre for Risk Studies, University of Cambridge

Oughton et al (2017), Quantifying the daily economic impact of extreme space weather due to failure in electricity transmission infrastructure, Space Weather,15,65–83

Papitashvili, N. E., King, J. H. (2020). Omni 1-min data – [GSE solar wind, AE index, 1981-2021]. NASA Space Physics Data Facility. Retrieved from doi:10.48322/45bb-8792 (Accessed on 23-09-2022)

Peren, Franz W. (2022). Statistics for Business and Economics. Springer Berlin Heidelberg, doi: 10.1007/978-3-662-65846-8, isbn:978-3-662-65845-1

Pietronero, L., E. Tosatti, V. Tosatti, A. Vespignani (2001) Explaining the uneven distribution of numbers in nature: the laws of Benford and Zipf, Physica A Stat. Mech. 293, 297 doi: 10.1016/S0378-4371(00)00633-6

Politis, D. N., Romano, J. P. (1994). The stationary bootstrap. J. Am. Stat. Assoc., 89, 1303–1313 doi:10.2307/2290993

Politis, D. N., White, H. (2004). Automatic Block-Length selection for the dependent bootstrap. Econometric Reviews, 23, 53–70 doi:10.1081/ETC-120028836

Pröger, L., Griesberger, P., Hackländer, K., Brunner, N., Kühleitner, M. (2021) Benford's law for telemetry data of wildlife. Stats, 4, 943–949, doi:10.3390/stats4040055

Pulkkinen, T. Space Weather: Terrestrial Perspective. Living Rev. Sol. Phys. 4, 1 (2007) doi:10.12942/lrsp-2007-1

Sambridge, M., Tkalci¸, H., Jackson, A. (2010). Benford's law in the natural sciences. Geophys. Res. Lett., 37, 437 doi: 10.1029/2010GL044830

Sheppard, K. (2021). bashtage/arch: Release 5.3.1 (version 5.3.1). Retrieved from doi: 10.5281/zenodo.593254439

Smith, A. R. A., Beggan, C. D., Macmillan, S., Whaler, K. A. (2017). Climatology of the auroral electrojets derived from the along-track gradient of magnetic field intensity measured by POGO, Magsat, CHAMP, and swarm. Space Weather, 15, 1257–1269 doi: 10.1002/2017SW001675

Sugiura, M. (1964). Hourly values of equatorial Dst for the IGY. Ann. Int. Geophys., 35 (9).445

Tindale, E., Chapman, S. C. (2016). Solar cycle variation of the statistical distribution of the solar wind ϵ parameter and its constituent variables. Geophys. Res. Lett., 43, 5563–557 doi:10.1002/2016GL068920

Washington, L. C. (1981). Benford's Law for Fibonacci and Lucas Numbers. The Fibonacci Quarterly. 19 (2): 175–177.

World Data Center for Geomagnetism, Kyoto, M. Nose, T. Iyemori, M. Sugiura, T. Kamei (2015), Geomagnetic AE index, doi:10.17593/15031-54800