LJMU Research Online

Wilby, RLW, Clifford, NJ, De Luca, P, Harrigan, S, Hillier, J, Hodgkins, R, Matthews, TR, Johnson, MF, Murphy, C, Noone, S, Parry, S, Prudhomme, C, Rice, S, Slater, L, Smith, KA and Wood, P

 The "dirty dozen" of freshwater science: Detecting then reconciling hydrological data biases and errors

http://researchonline.ljmu.ac.uk/5641/

Article

The "dirty dozen" of freshwater science:

Detecting then reconciling hydrological data biases and errors

OVERVIEW

Robert L. Wilby (*)
Department of Geography, Loughborough University, LE11 3TU (r.l.wilby@lboro.ac.uk)
Nicholas J. Clifford
School of Social, Political and Geographical Sciences, Loughborough University, LE11 3TU
Paolo De Luca
Department of Geography, Loughborough University, LE11 3TU
Shaun Harrigan
Centre for Ecology & Hydrology, Wallingford, OX10 8BB
John K. Hillier
Department of Geography, Loughborough University, LE11 3TU
Richard Hodgkins
Department of Geography, Loughborough University, LE11 3TU
Matthew F. Johnson
School of Geography, University of Nottingham, NG7 2RD
Tom K.R. Matthews
Natural Sciences and Psychology, Liverpool John Moores University, L3 2AJ
Conor Murphy
Department of Geography, Maynooth University, County Kildare, Ireland
Simon J. Noone
Department of Geography, Maynooth University, County Kildare, Ireland
Simon Parry
Centre for Ecology & Hydrology, Wallingford, OX10 8BB
Christel Prudhomme
Centre for Ecology & Hydrology, Wallingford, OX10 8BB
Steve P. Rice
Department of Geography, Loughborough University, LE11 3TU
Louise J. Slater
Department of Geography, Loughborough University, LE11 3TU
Katie A. Smith
Centre for Ecology & Hydrology, Wallingford, OX10 8BB
Paul J. Wood
Department of Geography, Loughborough University, LE11 3TU

ABSTRACT

Sound water policy and management rests on sound hydrometeorological and ecological data. Conversely, unrepresentative, poorly collected or erroneously archived data introduces uncertainty regarding the magnitude, rate and direction of environmental change, in addition to undermining confidence in decision-making processes. Unfortunately, data biases and errors can enter the information flow at various stages, starting with site selection, instrumentation, sampling/ measurement procedures, post-processing and ending with archiving systems. Techniques such as visual inspection of raw data, graphical representation and comparison between sites, outlier and trend detection, and referral to metadata can all help uncover spurious data. Tell-tale signs of ambiguous and/or anomalous data are highlighted using 12 carefully chosen cases drawn mainly from hydrology ('the dirty dozen'). These include evidence of changes in site or local conditions (due to land management, river regulation or urbanisation); modifications to instrumentation or inconsistent observer behaviour; mismatched or misrepresentative sampling in space and time; treatment of missing values, post-processing and data storage errors. As well as raising awareness of pitfalls, recommendations are provided for uncovering lapses in data quality *after* the information has been gathered. It is noted that error detection and attribution are more problematic for very large data sets, where observation networks are automated, or when various information sources have been combined. In these cases, more holistic indicators of data integrity are needed that reflect the overall information life-cycle and application(s) of the hydrological data.

Key words

INTRODUCTION

High quality hydrometeorological measurement contributes to high quality policies and management of natural resources. Examples of data sensitive (hydro-) decisions include: compliance monitoring for environmental regulation; water resource allocation between riparian states; planning, design and investment in long-lived water infrastructure; post-project evaluation; safety and performance reviews of critical infrastructure. All such activities rely on high-integrity data collection and archiving processes. Conversely, poor measurement and information management practices can seriously undermine confidence in data[1].

International bodies such as the World Meteorological Organisation (WMO) provide detailed guidelines on best measurement practices, beginning with how to choose a site for a meteorological station, followed by protocols for site maintenance and instrument use[2]. Likewise, seminal texts such as *Streamflow Measurement*[3] and *Hydrology in Practice*[4] explain the strengths and weaknesses of different types of equipment for measuring water balance terms. These points of reference are intended to avoid erroneous practices *before they occur*; there is surprisingly little advice on how to discern lapses in sound practice *after* the information has been gathered. Of course, there are quality assurance systems to protect the veracity of data holdings in major collections such as the UK National River Flow Archive (NRFA)[5]. But even these systems are fallible – erroneous entries can still slip through automated checking procedures when data values lie within plausible ranges.

This overview exposes some common data recording and handling errors, to explain how they might arise and be detected. We refer to our collection of 'rogue' data as *The Dirty Dozen*. This is in homage to the classic 1967 film by the same name in which a band of U.S. Army convicts are brought together to achieve an honourable but near impossible military objective. Similarly, by bringing together a portfolio of suspect data we are aiming for a positive outcome of raised awareness amongst researchers and practitioners. Although we draw our exhibits largely from observed data and personal experience, some of the same pitfalls might apply to modelled information. Likewise, while our case studies are mainly based on hydrological data the issues raised are relevant to related disciplines of ecology, meteorology and water quality.

The order of our dirty dozen follows a typical information flow. We begin with examples of artificial influences on monitoring sites (#1 to #4), then cover equipment changes (#5 and #6), quirks of sampling and observer bias (#7 to #9), interpretation of outliers (#10 and #11), and techniques for infilling missing data (#12). We then add examples of errors that can occur at post-processing and archiving stages, along with recommendations for detecting these kinds of erroneous values. Some

supporting data are provided as *Supplementary Information* so readers can examine the same data for themselves. It is our intention that the dirty dozen(s) assembled in this paper will provide a basis for practical exercises and expose some of the tell-tale signs when things go wrong with hydrometric data.

## EXHIBIT #1: CHANGING SITE LOCATION AND THE VALUE OF METADATA

Lengthy hydrometeorological records are essential for understanding climate variability and change, detecting emergent trends and contextualising extreme weather events. To be fit for purpose, these data need to be homogeneous (i.e. collected in consistent ways and places) so that variability is only caused by changes in climate rather than by artificial influences such as station moves. Homogeneity may be tested by a) identifying break-points in single series (absolute homogeneity)[6]; or b) comparing records from neighbouring stations (relative homogeneity)[2]. In both cases, metadata are invaluable for confirming detected breaks and for highlighting questionable parts of data that might elude statistical tests. The value of metadata increases with the age of the record because the earlier the data, the smaller the number of stations for implementing relative homogenisation tests.

For example, absolute and relative homogenisation methods were applied alongside metadata to build a quality assured, long-term rainfall network for the Island of Ireland[7]. One part of that record for Malin Head illustrates how station moves (and other factors) can influence trends identified in data and the importance of metadata in building confidence in adjusted series. This station was used in an earlier analysis of trends that claimed a large increase in annual rainfall totals[8]. However, metadata (Supplementary Information page #1) indicate: changes in the time and frequency of readings (throughout the record, but particularly after 1950 with the onset of hourly measurements); a move of the station from a cliff top at 230ft (70 m) to a location at 20ft (6 m) above sea level (in 1921); opening of a new station (same elevation) in 1955. Detected breaks in the annual rainfall series were consistent with the station relocation in 1921 and changes in the time of observations in the 1950s. This evidence was used to guide data homogenisation – that is correction for gauge under-catch during decades with less frequent measurement and more exposed site conditions[7].

A significant increasing trend is evident in the pre-homogenised annual rainfall series (**Figure 1a**). However, post-homogenisation, the gradient for the entire series (1890-2010) is only a quarter of that for the un-corrected record. **Figure 1b** shows a double mass plot which compares the cumulative sums of annual rainfall for the corrected Malin Head (MH) annual series with Derry (the nearest long-

running neighbouring station). The break-points and cumulative departure of the MH homogenised record from Derry (the 1:1 line) are smaller than those for the MH original record.

Recommendation: Use metadata to check the continuity of site location and environs; use techniques such as linear regression or Pettitt's test for break-points to expose trends and abrupt changes respectively that may be due to undocumented changes in site properties.
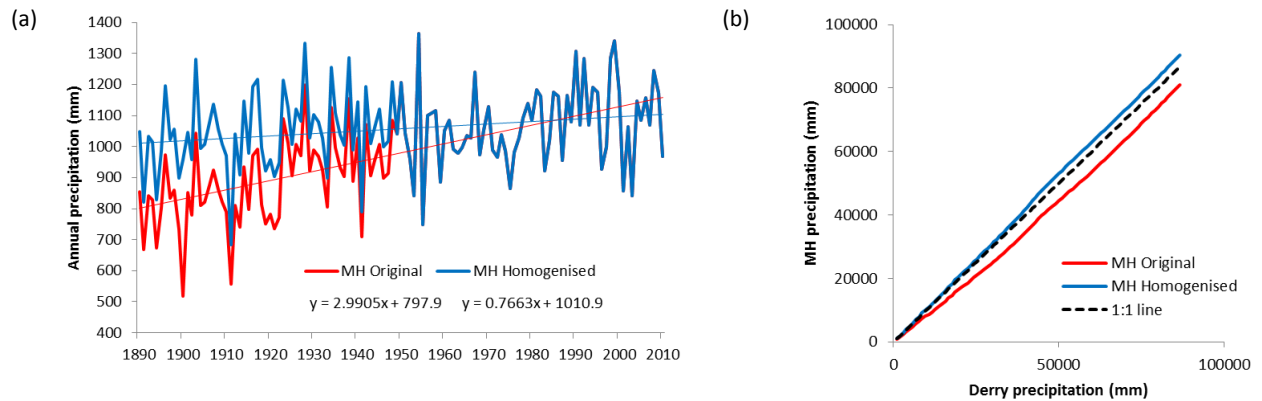

FIGURE 1 (a) Comparison of the original (red line) and post-homogenized (blue line) annual rainfall record for Malin Head, Ireland. Also shown are the regression equations for the linear trend in each series. (b) Double mass plots for original and homogenised annual precipitation series at Malin Head compared with a nearby station in Derry.

EXHIBIT #2: ARTIFICIAL INFLUENCE ON RECORDS (ARTERIAL DRAINAGE)

Agricultural productivity is greatly reduced where there is persistent waterlogging and flooding. In an effort to combat this problem, arterial drainage schemes involving channel deepening and widening may be undertaken to improve flow conveyance. Field drains might also be installed to drain the land. Newly dredged river channels have a greater capacity to receive additional water from previously waterlogged soils. While arterial drainage has economic advantages it can introduce hydrological discontinuities to river flow records.

For example, a break-point in the measured flows of the River Boyne in east Ireland was detected around the 1970s (**Figure 2**) (Supplementary Information page #2). Early studies linked this abrupt change in regime to increased precipitation caused by a shift in the North Atlantic Oscillation to a predominantly positive phase[9]. Subsequent research[10] attributes the change to an extensive arterial drainage scheme that took place over the period 1969-1986. Between this pre- and post-drainage period observed flow volumes increased by approximately 30%. Hydrological modelling was used to simulate flows in the Boyne catchment as if in a natural state (i.e. with no arterial drainage but with

observed climate variability). The results showed that modelled and observed flows did not match after the change point so increased precipitation does not fully account for the regime change. It was deduced that change in the Boyne must, therefore, be driven by a change within the catchment – most likely arterial drainage.
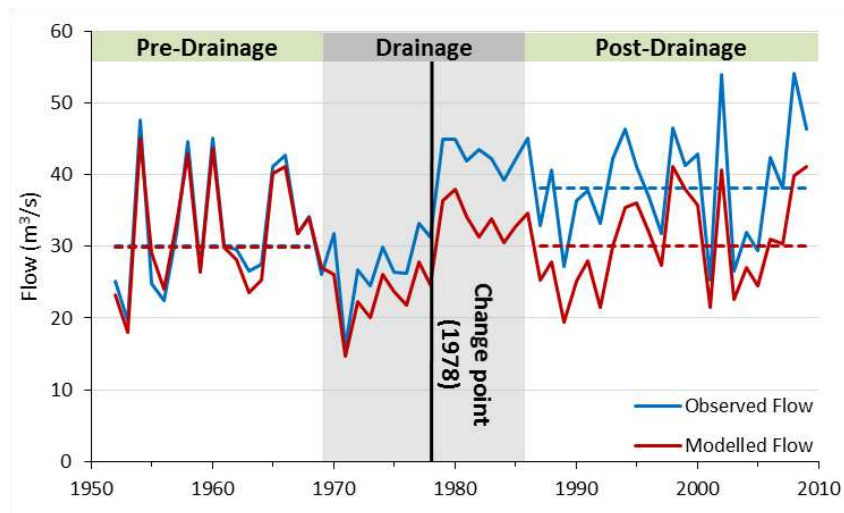
FIGURE 2 Observed (blue) and modelled (red) annual mean flow for the River Boyne catchment 1952-2009 (Office of Public Works station number 07012). The grey shaded area represents the years in which arterial drainage took place (1969-1986). Dashed horizontal lines are median observed and modelled flows in pre- and post-drainage periods. The black vertical line is the change point in observed flow in 1978, detected by Pettitt's test. Adapted from ref 10.

This case demonstrates how human modification to river channels and drainage properties can have a substantial impact on river flow. Such artificial changes can be misinterpreted as a natural consequence of, for example, an intensification of the hydrological cycle due to climate change. The process of setting up multiple hypotheses and systematically testing[11] the most likely cause(s) of a detected change (i.e. attribution) is fundamental to developing appropriate management responses and long-term adaptation strategies. Detection only studies can be of limited use for planning and could even lead to mal-adaptation.

Recommendation: Keep an open-mind about the cause(s) of hydrological change and set up analytical frameworks that can test multiple working hypotheses, including the effects of both anthropogenic and natural drivers of change.
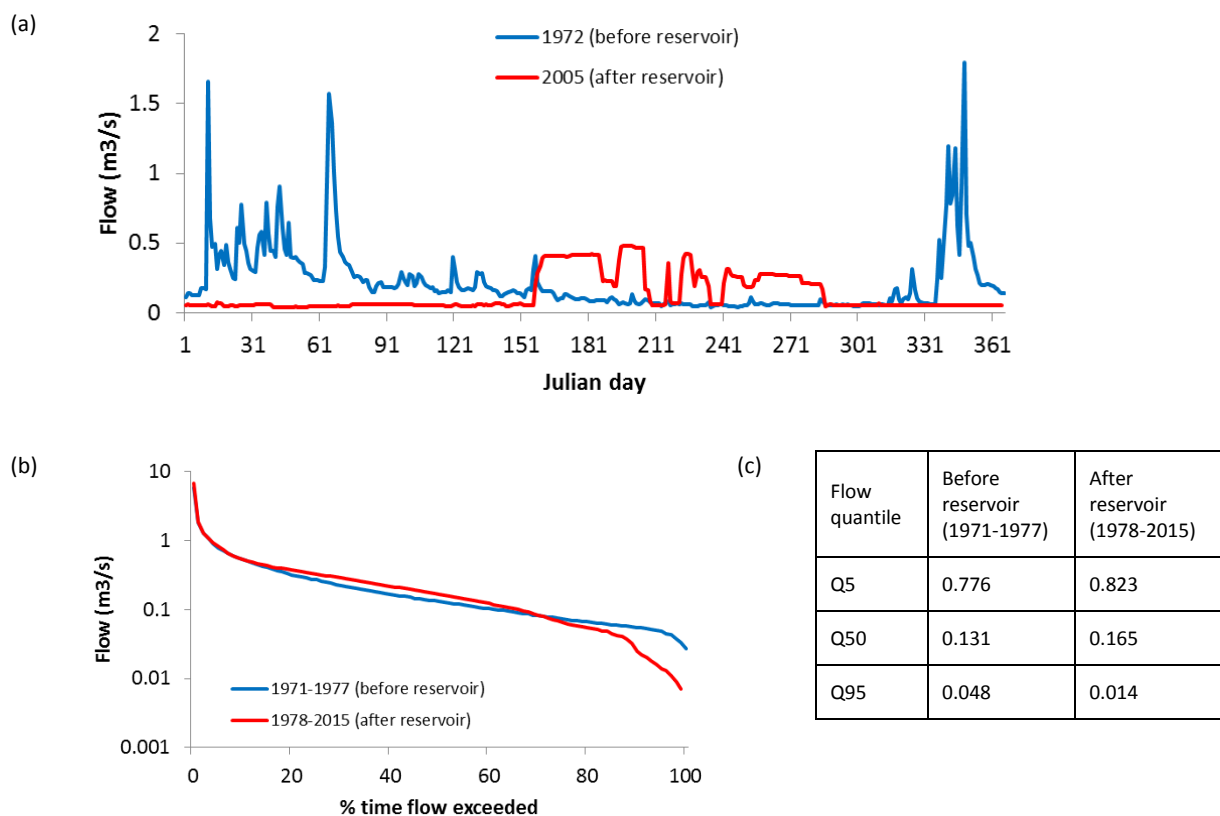
EXHIBIT #3: ARTIFICIAL INFLUENCE ON RECORDS (REGULATED RIVERS)

The construction and operation of reservoirs can substantially impact gauged river flows[12] (and other quantities such as water temperature[13]), predominantly through the introduction of compensation

flows[14], the suppression of flood maxima and/or the timing and magnitude of releases. One such time series in the UK NRFA is the Shell Brook in southern England. This gauging station began recording river flows in 1971 and these early data reflect the natural flow regime. In 1978, Ardingly reservoir was constructed immediately upstream of the gauging station.

The post-1978 river flow record is clearly influenced by the reservoir, with sustained periods of similar flows and abrupt step changes (as in 2005; **Figure 3a**). These anomalous patterns predominantly impact the drier half of each water year (April to September), although there are some years in which all flow data are affected. The sustained periods of both low flows and moderate flows are particularly apparent when comparing the pre- and post-reservoir flow duration curves (**Figure 3b**) and flow quantiles (**Figure 3c**). Adjustments to reservoir operations have also introduced substantial inter-annual variability.

(a)

(b)

(c)

Flow quantile	Before reservoir (1971-1977)	After reservoir (1978-2015)
Q5	0.776	0.823
Q50	0.131	0.165
Q95	0.048	0.014

FIGURE 3 Pre- and post-reservoir data for Shell Brook, UK (NRFA 41024): (a) daily river flow hydrographs in 1972 and 2005; (b) flow duration curves for 1971-1977 and 1978-2015; (c) 5th, 50th and 95th flow quantiles for the same periods as (b).

In this example, simply plotting the data should highlight the impact on flows. However, reservoir influence can be more subtle, for example the truncation of low flows in summer. Such effects are more difficult to detect, although a flow duration curve (e.g. **Figure 3b**) can help to highlight deviations from the expected distribution of river flows in a natural series. Where modelling approaches can

generate naturalised river flow data, a range of ecologically relevant indicators can be calculated to summarise changes in seasonal flow regime, hydrological extremes and variability (e.g. ref 15).

Recommendation: Plot hydrographs and search metadata to identify more obvious erroneous river flow data; plot flow duration curves and calculate flow quantiles to quantify the influence or to highlight more subtle impacts.

EXHIBIT #4: ARTIFICIAL INFLUENCE ON RECORDS (URBAN HEAT ISLAND)

Near surface air temperatures are influenced by regional- and local-scale energy balances. In mid-latitudes, for example, summer anticyclones generally elevate air temperatures by synoptic-scale subsidence and by diabatic warming through amplified surface heat fluxes. The latter can be highly sensitive to spatial variations in the physical properties of the underlying land cover which modulate surface energy fluxes[16]. Non-homogeneity can emerge in temperature records at fixed sampling locations if these site-specific properties change in time. This can be problematic for the interpretation of trends in long-term temperature records. For instance, without detailed interrogation, it can be difficult to separate the impact of global-scale anthropogenic warming on temperature records, from local processes driven by land cover modification (e.g. ref 17). This attribution uncertainty extends to associated water balance terms such as evapotranspiration[18].

Urbanisation is known to affect air temperature records, as the surface properties of cities modify energy fluxes in ways that strongly favour nocturnal warming[19]. Sampling locations experiencing urbanisation over time may, therefore, contribute to a warm bias in the study of larger-scale temperature trends[20]. An assessment of data collected by the U.S. Historical Climatology Network, found much greater 20th Century warming for urban stations relative to their rural counterparts, particularly for minimum air temperatures[21]. **Figure 4a** demonstrates this tendency for two stations separated by only a few hundred kilometres, with the urban site experiencing more than twice the rate of rural warming.

Where such localised heating effects are detected it may be desirable to exclude the sample location from the study. However, removal of the artificial warming signal is also possible, for example by homogenization techniques[22] (see Exhibit #1), or via methods that explicitly identify and adjust urban records to yield trends consistent with rural neighbours[23]. Satellite observations of night lights (**Figure 4b**) can be used to independently discriminate between rural and urban sites[21].

Recommendation: Use independent indicators of the extent of urban development (such as maps of nocturnal light) to identify surface air temperature records that may be affected by urbanisation; apply urban-rural pairing procedures to correct localised warming trends at urban sites.
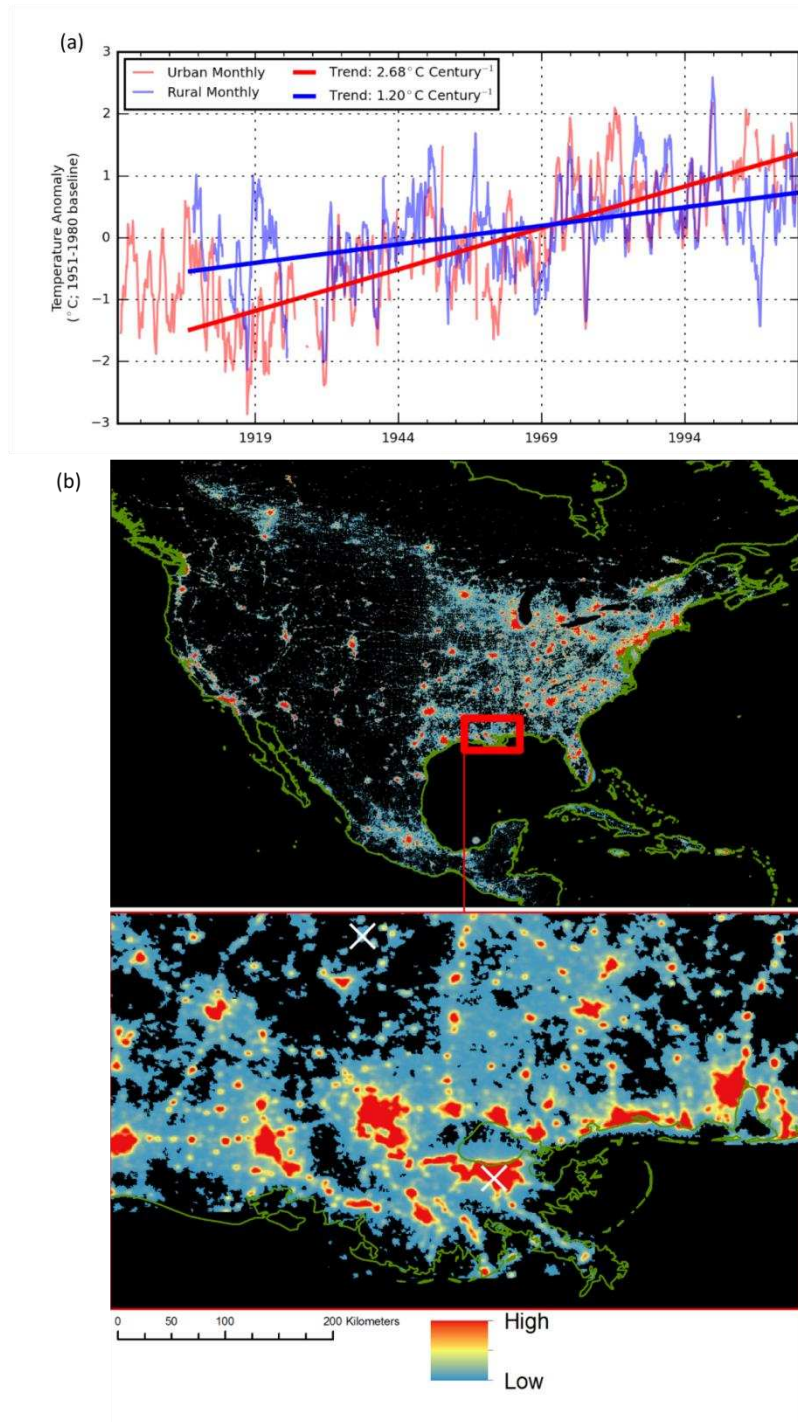
FIGURE 4 (a) Unadjusted monthly mean minimum temperatures smoothed with 12-month running mean at urban (USHCN ID: 166664) and rural (USHCN ID: 168163) weather stations. (b) The locations of stations in A are shown on a map of night time lights generated from the Defence Meteorological Satellite Program's Operational Linescan System (http://ngdc.noaa.gov/eog/). The lower panel shows the locations in detail, with example stations marked with a white cross. The urban station is situated in the bright area of New Orleans.

EXHIBIT #5: CHANGING INSTRUMENTS

There are many ways of collecting river flow data. Fixed gauging stations such as weirs and flumes aim to stabilise the relationship (rating) between flow depth and volume to enable more accurate measurement of discharge. Ultrasonic gauges and electromagnetic gauges measure velocity using acoustic pulses and magnetic fields respectively. Structures and equipment at a gauging site may be installed, then changed or upgraded in time. For example, a velocity-area station may be superseded by a weir, which may in turn have ultrasonic equipment installed if a stable relationship between water level and flow cannot be achieved. Weirs can alter the level of the river significantly, and may affect only certain aspects of the flow regime. For example, in the case of the Harper's Brook (**Figure 5**), only the annual maximum flows appear to be affected. Like all field equipment, electromagnetic gauges suffer from deterioration over time which introduces errors to the flow data (e.g. due to degrading insulation of detecting electrodes, or siltation of the weir cross-section).
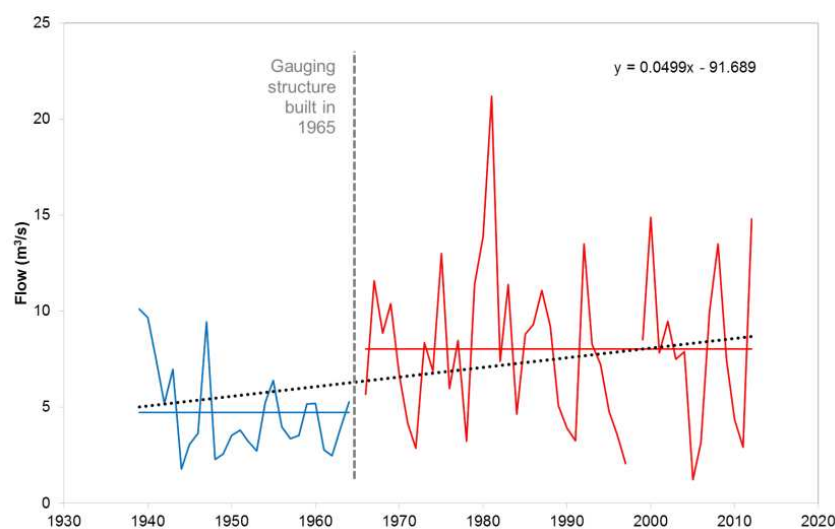
FIGURE 5 AMAX series for the Harper's Brook at Old Mill Bridge, UK (NRFA 32003). A compound crump profile weir was installed in 1965. The black dots show the linear trend for the whole record, with the equation given in the top right corner. Horizontal blue and red lines show the AMAX mean of the records pre- and post-1965 respectively.

When a gauging station is being installed or modified, data are generally not recorded leading to gaps in the time series (see Exhibit #12). Weirs and electromagnetic gauges require substantial building works and the disruption to the flow in this period is significant. However, where ultrasonic gauges are fitted, an overlap period may be used to calibrate the instruments. In the majority of cases, installation of a weir or alterations to it are accounted for by taking spot gaugings of river velocity and cross sectional area, and altering the rating curve (which defines the relationship between the stage and the discharge). Despite this, testing can reveal step changes and/or false trends as a result of gauging alterations and/or gaps in records[24].

For example, metadata for Harper's Brook at Old Mill Bridge, central England shows that the record began with a velocity area station measuring the natural channel, until a compound crump weir was built in 1965 (**Figure 5**). A simple linear regression fit to the annual maximum (AMAX) flow series reveals a substantial trend, whereby the AMAX values appear to increase by 0.5 m^3/s per decade. Plotting mean values for the data before and after the installation of the compound crump weir highlights the effect of the structure on the high flows in this river. The Pettitt statistical change point test also detects the year 1965. Even so, the increase in AMAX could still be partly explained by multi-decadal climate variability leading to a flood-rich period in the later portion of the record[25].

Recommendation: Use metadata to check the continuity of instrumentation at a site; use the Pettitt test to expose abrupt changes that may be due to undocumented changes in equipment at the site.

EXHIBIT #6: CHANGES IN GAUGING STATION DATUM

Measurements of stage, or *water surface elevation* (alternatively referred to as the *gauge height*), are generally made above an established datum[26]. The zero elevation point is often located in the ground beneath the riverbed. Ideally, the datum should be fixed over time, such that there is a consistent reference point for the entire record. However, sometimes the datum is changed, for example, following degradation of the riverbed. Unfortunately, it is estimated that between a third and half of all U.S. Geological Survey (USGS) stream gauges have had a change of datum or major change of location during their period of record (Kolva, *pers. comm.*).

Changing the datum alters the gauge height that is referenced for a given water surface elevation. For example, at the Comite River near Comite (**Figure 6**), the datum was lowered by two feet (0.6 m) on 1 October 1996 [note the imperial units that are routinely used in the US]. Hence, a stage of two feet in September 1996 is equivalent to a stage of four feet (1.2 m) in October 1996 (for the same water surface elevation). Such changes can be detected relatively easily in historical time series when a large datum shift is applied (**Figure 6**), but not necessarily when the change is small or gradual, for example due to ground subsidence. Note also that switches of units such as between imperial (in **Figures 6** and **7**) and metric can be problematic too. Shifts in the stage-discharge relation may further be indicative of natural geomorphic processes at the site (e.g. changes in riverbed elevation or channel width due to accretion or erosion)[27]. Information about changes in datum can usually be found in the USGS gauging station water-year summary report (see: http://waterdata.usgs.gov/LA/nwis/wys_rpt/?site_no=07378000&agency_cd=USGS).
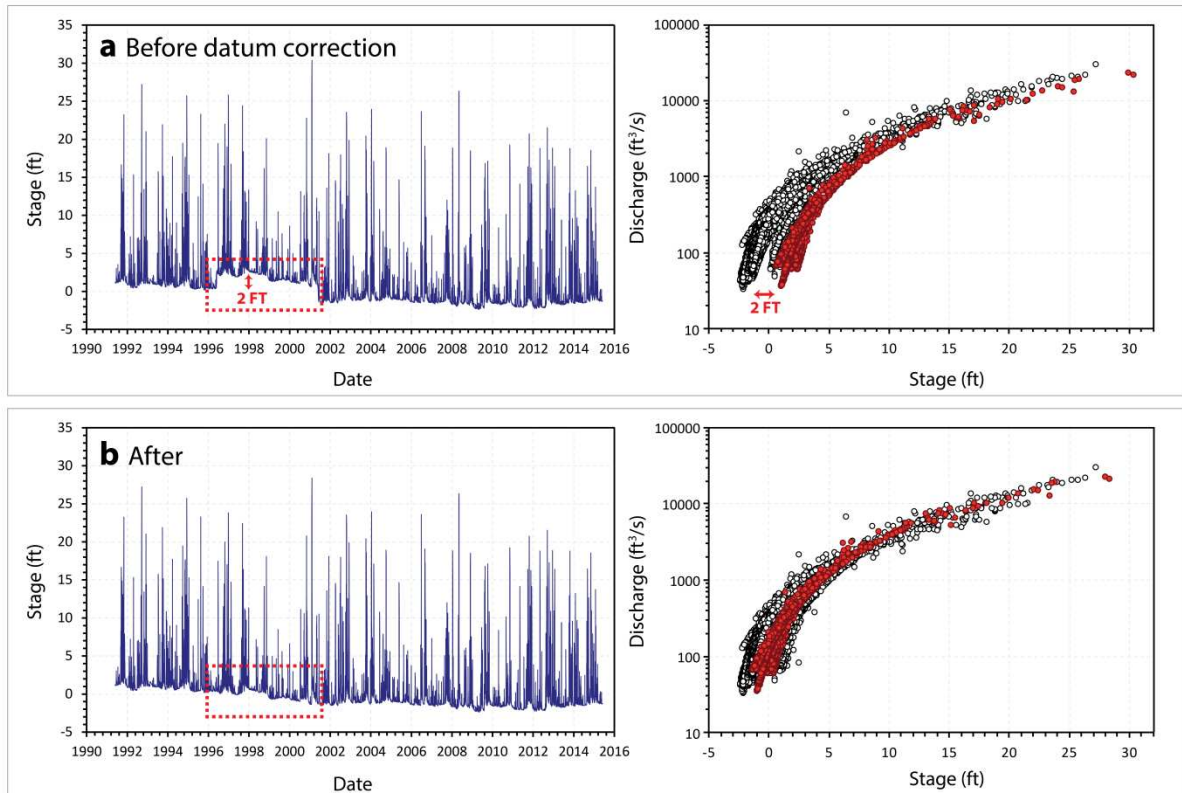
FIGURE 6 Before and after correcting for change in datum. Stage records for the Comite river near Comite, Louisiana (USGS site number 07378000) are publicly available on the USGS National Water Information Service website[28]. The online Water Year Report states "*From Oct. 1, 1978 to Sept. 30, 1996, at current datum. From Oct. 1, 1996 to Sept. 30, 2001, at datum 2.00 ft lower*". Therefore, the stage time series were adjusted to the same datum by subtracting two feet from the measured stage between 1st October 1996 and 30th September 2001 (i.e. water years 1997 to 2001). The measurements made during this period are shown as red circles, before (a) and after (b) datum correction.

The issue of datum correction is particularly important for time series analyses of stage records[29] or of river channel geometry[30]. When computing changes in the frequency of flood events above flood stage, for example, if the measurements are not referenced to a fixed datum, a spurious trend in flood frequency could be inferred. Progressive changes in the datum may also contribute to instability in rating relationships used to estimate discharge from river stage[31].

Recommendation: Plot and visually inspect the stage-discharge relationship and stage time series before conducting any statistical analyses; note any abrupt shifts in stage that may reveal undocumented changes in datum.

EXHIBIT 7: OBSERVER MEASUREMENT BIASES

Benford's Law (BL), also known as the first digit law, recognizes that in many collections of numbers, the leading digit is most often 1 (~30% of the time) and least often 9 (~5% of the time). Such differences in frequency are greater than would be expected to occur by chance. BL holds for a wide variety of socio-economic and natural science data sets[32]. Knowledge of this law can be used as a diagnostic tool. For instance, departures from expected high frequencies of small leading digits are routinely used to pick up rounding errors or fabricated data (e.g. in tax returns).

BL can also be used to detect observer bias or suspect values in hydrometeorological data[33]. Some biases may be unintentional. For example, weather observers tend to favour daily precipitation totals that are divisible by 5 or 10. One evaluation of the U.S. Cooperative Observer Program network found that 97% of stations with complete or near complete records exhibit this 5/10 bias[34]. Observers also tend to under-report the frequency of days with light precipitation, that is, daily totals at the lower limit of measurement – which in the U.S. is often close to 2.54 mm (or 0.1 inches). Both biases were linked to the precision and consistency of use of precipitation measuring sticks which have large, labelled tick marks every 0.10 inches, large, unlabelled tick marks every 0.05 inches, and small, unlabelled tick marks every 0.01 inches[34]. Both number bias and trace wet-day under-reporting skew the overall frequency distribution of precipitation amounts in ways that can affect estimation of extreme values.

Another bias occurs when manual weather observations are not made on a weekend or over a holiday period. Instead, any precipitation falling during the unobserved days is assigned to the first day of return to business, which is typically a Monday or Tuesday. Average precipitation totals on these days tend to be higher than those estimated for days on the weekend. Such under-reporting of rainfall on Sunday has been shown for meteorological stations in Australia[35], the UK[36] and U.S.[34].

To illustrate these points, observer number preference and weekend under-reporting biases are assessed using daily precipitation data for Dushanbe, Tajikistan (**Figure 7**) (Supplementary Information page #7). At this site, observer(s) have a preference for 3.0 and 6.0 inch daily rainfall totals as evidenced by unexpectedly high frequencies of these amounts during the period 1958-1967. In fact, the value 3.0 occurs 14% more frequently than expected by BL. More striking is the lack of any values either side of the 3.0 and 6.0 inch amounts which further raises doubt about the credibility of these entries. Mean intensities are notably higher on Mondays/Tuesdays than on Sundays suggesting that some weekend rainfall has been carried over into weekday totals too.

Recommendation: Use histograms of daily precipitation amounts to reveal under-reporting of light rainfall and/or observer number bias; mean daily amounts plotted by day of week can expose unrecorded aggregation of multi-day precipitation.
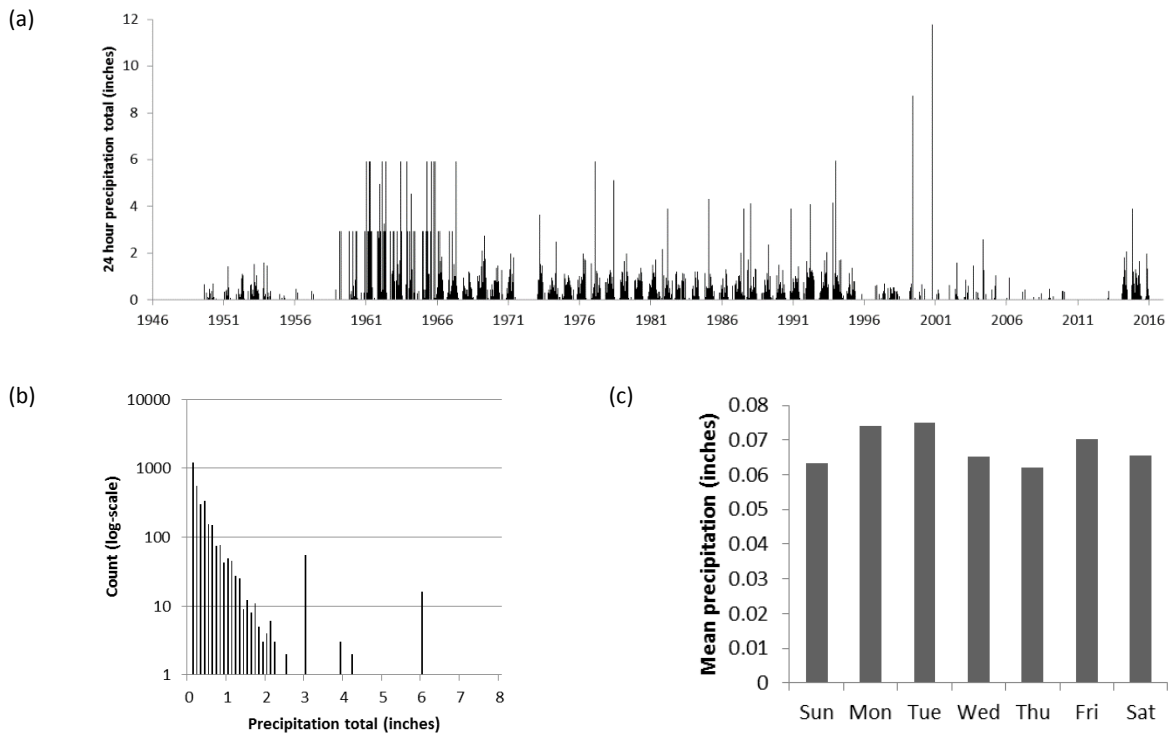
FIGURE 7 Evidence of observer (a, b) value and (c) day of week biases in daily precipitation amounts recorded for Dushanbe, Tajikistan. Data source: NOAA Global Summary of the Day.

EXHIBIT #8: SAMPLING BIAS IN TIME

Spot sampling is widely used to monitor environmental variables in a non-continuous way, perhaps to save time and/or resources. Sampling may be fixed (systematic) or random (without any temporal or spatial structure) according to the purpose of the data collection. Ideally, the sampling frequency, time and location are appropriate to the behaviour of the variable(s) under surveillance. Slowly varying phenomena such as groundwater levels may be adequately sampled once per month at a handful of sites to represent behaviour across an aquifer. Conversely, rapidly varying variables like suspended sediment concentrations (see Exhibits #10 and #11) have to be sampled at hourly or sub-hourly intervals to accurately estimate the amount of material transported. If the sampling frequency is not appropriate, biased estimates may arise.

For example, it has been shown that the 98th percentile water temperature (used for compliance monitoring in the EU Water Framework Directive, WFD) can be 1°C cooler if based on monthly values

rather than the "true" values from hourly sampling[37]. As well as the frequency, the *time* of sampling is also critical for variables like water and air temperatures which have strong diurnal and seasonal cycles[38]. Provided that samples are collected at fixed points in these cycles, repeat measurements are comparable with each other. **Figure 8** provides an example where systematic spot sampling was not applied to water temperature monitoring at a site on the River Dove, UK (Supplementary Information page #8).
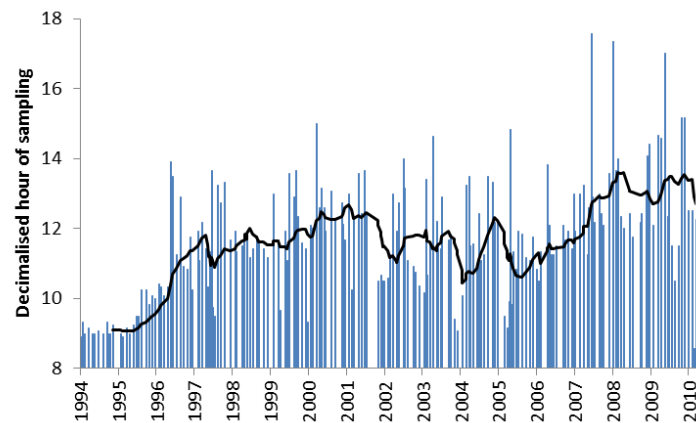
FIGURE 8 Time of day when spot samples of river water temperatures were taken at Glutton, River Dove, Derbyshire, UK. Adapted from ref 39.

Although the water temperature measurements in **Figure 8** were made by trained field staff, following standard procedures, with well-maintained equipment and at a fixed location, the time of day of taking the monthly samples was not consistent. Spot samples in the mid-1990s were taken at around 09:00 hrs, but this drifted to about 13:00 hrs by the 2010s. Given that afternoon water temperatures are typically higher than those in the morning, the change in sampling time alone has introduced a warming bias of ~1.1°C over the course of the record. Even small discrepancies in water temperature are significant because they can lead to a misclassification of a river's health under the terms of the WFD, or exaggerate the pace of warming seen in UK freshwaters[40].

Recommendation: Plot the time of spot sampling to check for hidden biases in the collection of data, particularly for series with strong cyclical variations.

EXHIBIT #9: MISMATCHED SAMPLING IN SPACE AND TIME

Continuous river discharge records are often used to derive 'flow statistics' to match with other environmental indicators such as benthic invertebrate data[41,42]. High resolution flow series may yield point discharge at a predetermined time and date through to daily, seasonal or annual averages and

long-term flow duration statistics (e.g. Q95 – the flow that is exceeded 95% of the time). In contrast, most ecological series represent discrete sampling events, typically collected on a quasi-annual or seasonal basis (**Figure 9**). Hence, timing of eco-sampling may vary from one year to another with, for example, collection of an 'autumn' sample anywhere between 1 September and 30 November. When assessing potential influences of antecedent flow conditions on instream communities it is clearly essential that discharge and ecological series overlap to ensure that the hydrological conditions experienced by instream communities are properly reflected. Two primary sources of error may still arise after quality assurance processes have been undertaken: (i) sites where discharge and ecological series were derived may not be co-located; and (ii) the sampling time-frame of discrete ecological series may miss potentially important hydrological events driving community structure and change. A third potential source of error may occur if discharge statistics drawn from the UK hydrological year (1st October – 30th September) are matched with ecological samples that are collated on a seasonal basis (such as autumn, which spans 1 September to 30 November).
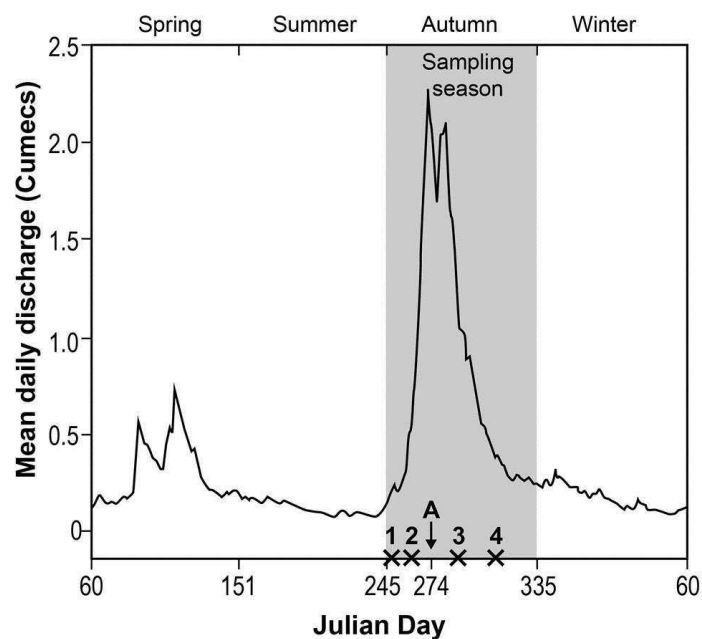
FIGURE 9 Schematic of continuous river discharge measurement with a schedule of discrete biological surveys (numbered 1 to 4) within an autumn sampling season. Eco-sample 1 is collected under steady/low flow conditions; 2 during a period of catchment re-wetting; 3 near to and 4 following the peak discharge. A denotes the start of the hydrological year in the UK.

One study examined 291 long-term (>20-years) paired river flow and autumn season macroinvertebrate community records (>10-years) for sites across England and Wales[43]. Screening of the series resulted in 208 (71%) of the sites being removed due to missing values or because sampling points were not coincident. Removal of some sites was necessary because of flow addition or loss associated with impoundment, abstraction or confluences occurring between the gauge and bio-

monitoring points. A common source of error was due to missing hydrological events because of the mismatch between the hydrological year (October to September) and seasons used to analyse discrete macroinvertebrate samples (such as autumn being September to November).

Errors can arise when (i) an invertebrate sample is collected towards the end of a season with marked variability in river discharge that is not reflected in the seasonal average of the chosen flow metric (points #3 and #4 in **Figure 9**); or (ii) discharge data from the period after the ecological survey is included in the seasonal average flow metric if the 'hydrological year' is not corrected to coincide with ecological sampling window. Most ecohydrological statistics potentially omit some hydrological events due to the mismatch between the continuous hydrological and discrete ecological series. It is, therefore, probably not surprising that the most statistically significant models of river flow–ecology relationships have been developed for less hydrologically variable groundwater dominated systems as opposed to flashier surface runoff dominated systems.

Recommendation: *Plot hydrological time-series alongside dates for discrete ecological samples to confirm that sampling periods are coincident; examine series for the presence of potentially significant discharge events prior to collection of ecological samples (even those falling in another season).*

EXHIBIT #10: SPURIOUS OR CURIOUS SPIKES

Modern instruments deployed in rivers can provide high frequency (≤ 1 minute resolution) data, creating new opportunities for research but also requiring careful quality control. For example, Acoustic Doppler Velocimetry can record flow speeds at >100 Hz but it is widely acknowledged that time series require filtering to remove spurious values, that are an inherent and unavoidable product of the technology. Standard protocols exist for identifying spikes and outliers, which usually involve removing data that fall outside upper and lower thresholds defined relative to the record mean[44, 45].

Similarly, high frequency turbidity records can be subject to considerable noise and other limitations, not least when calibrating turbidity and suspended sediment concentration (SSC) records[46,47]. Noise can be caused by electronic signal errors, but these tend to be small relative to mean values and normally within the error range of the device (**Figure 10**). Larger spikes in data are common and can be caused by dirty optics, particularly biofouling that can be detected by sudden step changes or more gradual, but systematic shifts in turbidity. Wipers on sensors can remove small contaminants but larger debris must be manually removed.

Large spikes can also be caused by biological activity[48,49]. For instance, **Figure 10**, shows spikes in turbidity due to the activity of Signal Crayfish (*Pacifastacus leniusculus*) in both laboratory and field settings, compared with controls where crayfish were excluded. In still-water with no crayfish, spikes are small, so most likely associated with electronic signal errors. In contrast, records with flowing water and crayfish are subject to much larger spikes, which reflect the impact of sediment disturbance by crayfish. Diurnal variations in spikiness are indicative of biological activity. One study reported that spikes are three-times more likely and 20% higher when crayfish are active at night, than during daylight[50].
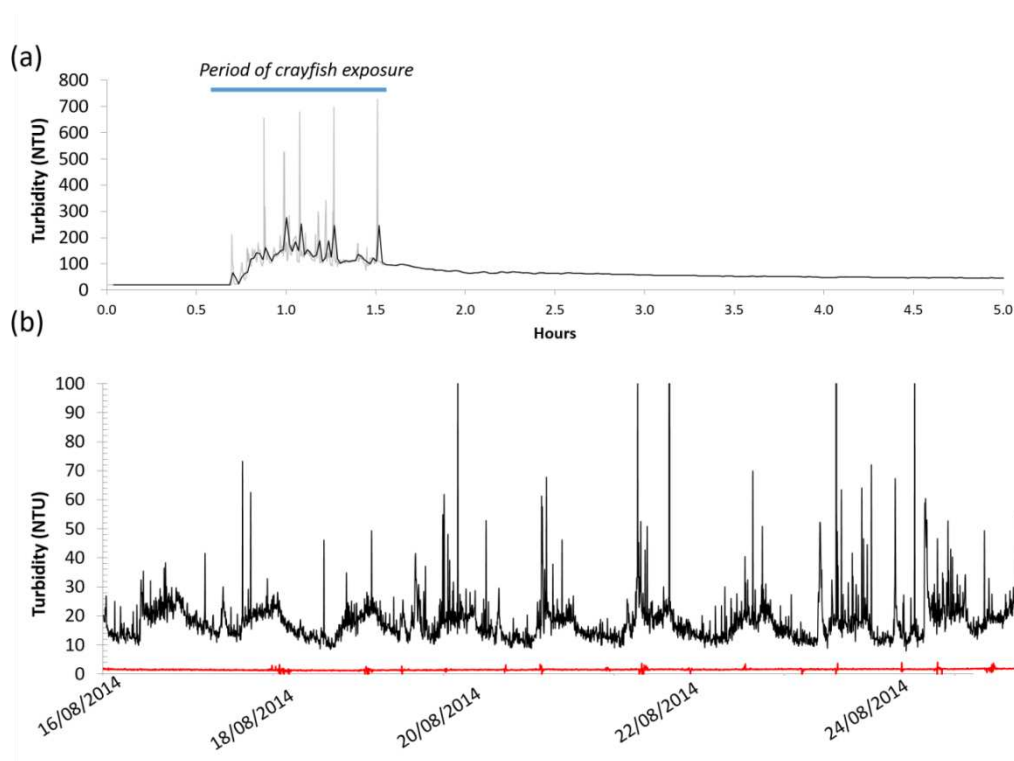
FIGURE 10 (a) 10-second resolution turbidity record (grey) with a 1-minute moving average (black) over 10-hours in a still-water laboratory aquaria with silty substrate and one Signal Crayfish left for 1-hour near the beginning of the experiment, after which time it was removed. Note that spikes occur only when the crayfish is present, with gradually decreasing turbidity after crayfish removal. (b) 5-minute resolution turbidity record for a tributary of the River Nene, UK, colonised by crayfish (black) which records a signal with more frequent spikes during night hours (labels are at midnight) and a strong diurnal structure in the mean turbidity. During this period other instruments confirmed that there were no changes in hydraulics capable of driving these turbidity fluctuations. It was concluded that individual spikes reflect fine sediment entrainment caused by foraging, burrowing or fighting events, which increase at night because crayfish are nocturnal. The diurnal pattern reflects the net effect of this enhanced night time activity on mean turbidity. A second turbidity sensor (red line), identical to that in the river, was deployed in an open-top aquarium filled with clean water and situated on the river bed adjacent to the first. The flat trace confirms that the signal from the river is not an instrument artefact, driven by diurnal variations in light or temperature that can affect the optical measurement of turbidity in some sensors. The small spikes that do occur, fall within the manufacturers stated error, are randomly distributed around the mean and do not show any temporal structure, which suggests that they reflect instrument noise.

Hence, high resolution turbidity data from field deployment needs careful assessment. Systematic changes, such as cumulative increases in turbidity, or step-changes should be removed and are likely the result of sensor fouling. However, remaining spikes exceeding sensor error terms are likely to be associated with biological activity or turbulent events, representing real phenomena.

Recommendation: Understand potential sources of data spikiness that are inherent in some measurement techniques but do not remove spikes and outliers uncritically; cross-check unexpectedly high data values against independent evidence and consider all potential causes of data excursions - they may reveal something unexpected and important.


EXHIBIT #11: MEANINGLESS MEANS

Simple measures of central tendency – such as average annual river flow or mean winter monthly rainfall – are routinely used to characterise hydrometeorological data. Such metrics are meaningful when the properties in question exhibit relatively consistent variability (i.e. when there are slow variations, with few extreme departures from typical levels). This applies not only to mean values over any particular time interval, but also to the nature and extent of any variability over diurnal, monthly, seasonal, or annual scales.

Sometimes, however, time series do not exhibit gradual or at least consistent change; instead, there may be extreme and apparently unpredictable variability at multiple timescales. For example, SSC may exhibit abrupt spikes above background levels, rather than gradual shifts. These effects can be caused by episodic sediment supply from biological activity (e.g. Exhibit #10), or due to bank collapse, flushing associated with rainstorms or meltwater release, or variable entrainment patterns on floodplains. This makes determining a representative value of SSC difficult, because actual values tend to be either very low background readings, or volatile quantities associated with transient events. In other words, the data are multi-modally distributed.

Figure 11 shows discharge and SSC time series for the pro-glacial river of the Finsterwalder Glacier in Svalbard, Arctic Norway[51]. Meltwater-fed systems such as these are useful exemplars of hydrological processes because they exhibit rapid change over relatively short timescales. In this case, it is evident that SSC values are dominated by two brief episodes (corresponding to flushing events) and a mean value that is not representative of the bimodal distribution of concentrations. Furthermore, attempts to quantify variability around the putative mean SSC are unreliable without explicit reference to a

specific timescale – the characteristic diurnal range in this particular example is much smaller than the seasonal range.
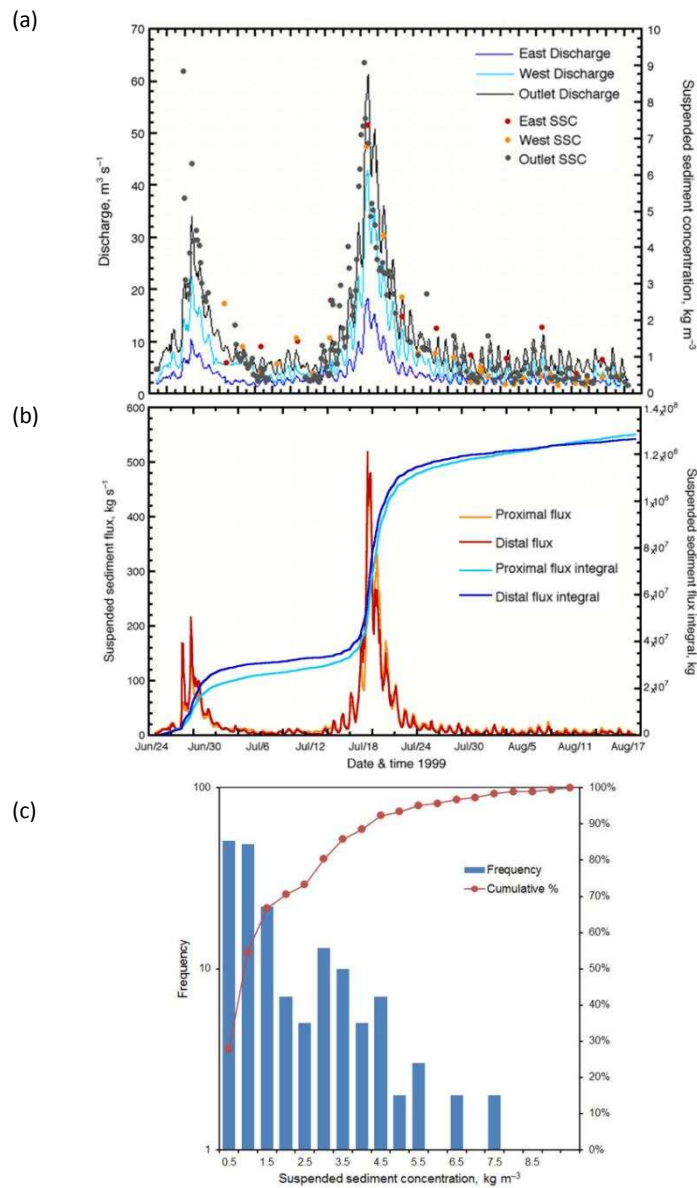
FIGURE 11 (a) Discharge and SSC time series for the proglacial river of the Finsterwalder Glacier, Norway, showing two periods of very high values compared to background levels (East and West are rivers draining the glacier margins, which coalesce downstream to form the Outlet river); (b) The same data converted to SSL and integrated over time – a process that yields improved characterisation of the sediment transport regime compared with simple measures of central tendency and dispersion (the proximal flux is the sum of the East and West fluxes; the distal flux is the outlet flux); (c) Histogram of the SSC showing the multi-modal nature of the data.

However, because of the need to statistically characterise the system, the question remains: what is a representative suspended sediment transport value for – in this case – the Finsterwalder proglacial river? This question is best addressed through temporal and spatial aggregation. Here, the time

integral of Suspended Sediment Load (SSL, the product of SSC and discharge; in units of mass) better quantifies the total transport for the duration of the time series, and forms a reliable basis for calculating sediment flux (in units of mass per unit area per unit time).

Recommendation: Aggregate time- and space-scales as much as possible when describing the 'average' condition of rapidly-changing variables with transient, extreme values when it is useful/ important to define mean conditions.

EXHIBIT #12: INFILLED DATA GAPS

Data may be missing for various reasons, including equipment malfunctions or loss during transmission and storage. Sometimes data are coded as missing because they are of insufficient accuracy, precision or reliability to be retained. Records may begin at different times, be discontinuous, or end before the present day. Individual variables may differ in their completeness even at the same site. Plotting data availability with time (**Figure 12a**), shows the extent of overlap between neighbouring records that might be bridged to create a composite series (as in ref 52). Information in the metadata may contain errors too. For example, header information held in the NRFA 41009 record for gauge **C** incorrectly reported no data between 1977 and 1998.
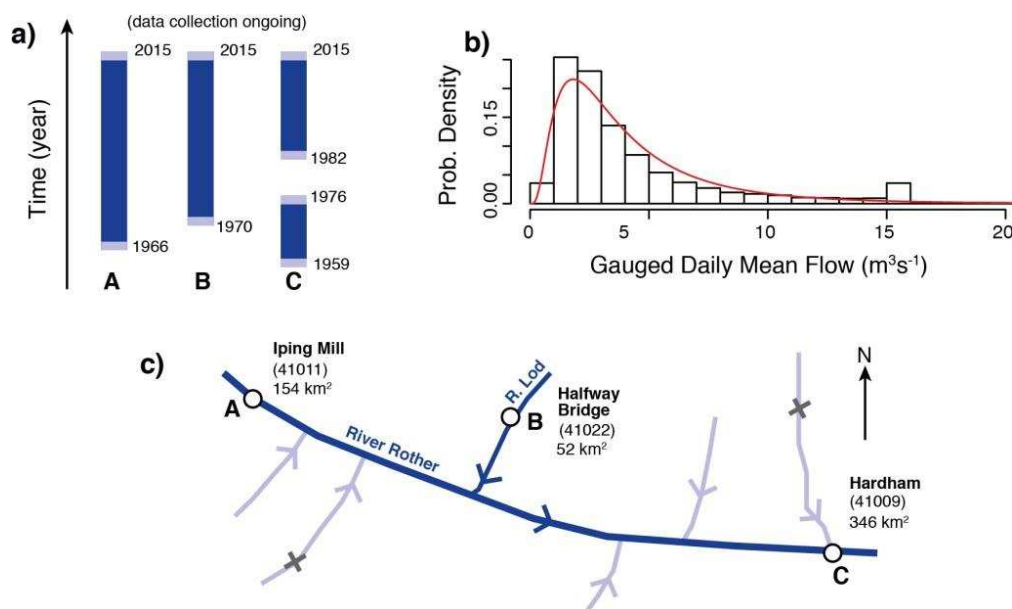
FIGURE 12 Gauged river flow records for the River Rother, UK. (a) Data completeness for the three gauges, with light blue illustrating partial data; (b) Histogram of river flows for gauge **C**, fitted with a log-normal distribution (red); (c) Schematic of gauge locations and metadata. Dark blue river reaches are measured by gauges **A**-**B**-**C**, light blue ones are not; numbers in brackets () are the NNRFA station codes; areas represent the upstream catchment size, and **x** symbols indicate other gauges. Data source: UK National Flood River Archive (http://nrfa.ceh.ac.uk/)

Infilling data gaps may be necessary to create homogeneous hydrological series for assessment of long-term variability (see Exhibits #1 and #2), extreme events or continuous series for running models. However, any infilling by interpolation or extrapolation relies upon assumptions that can introduce artefacts and give an impression of false certainty. For instance, the parameters of a statistical distribution can be estimated from a sub-set of the observations as in **Figure 12b**, but the observations do not exactly conform to the log-normal curve selected. Hence, using the log-normal for infilling would impose some of this assumed shape on the distribution. Critically, if gap filling is needed, beware of using the mean (of the rest of the record or neighbouring stations) as this will suppress variability and underestimate extremes. There are three valid alternatives:

1) *Time substitution* involves taking information from other dates assuming stationarity of the observations. For example, with flow data from gauge **C** fitted to a log-normal distribution, the missing data for 1976-1982 can be resampled from the same distribution. Alternatively, using the relationship between the overlapping records of gauges **B** and **C** for the period 1982-2015, the missing block in gauge **C** for 1976-1982 could be estimated from gauge **B**. Other sources of data, such as newspaper archives or proxy records, can help to corroborate infilled extreme events[53, 54].

2) *Space substitution* involves taking information from equivalent sites. For flood frequency estimation, 'pooled' analysis is common practice. For example, this technique was used to create 1405 annual maxima flow values for the River Trent, UK using ~50-year records[54, 55].

3) *Physical principles* can be used to predict missing data. For instance, **A** and **B** flow towards **C**; this means that **A** and **B** are each hydrologically linked to **C**, and all three are likely meteorologically inter-related given their proximity (<10 km) (**Figure 12c**). Using rainfall-runoff models it would be possible to estimate missing values at gauges **A**, **B** or **C** and any inter-relationships between them. Missing records can then be infilled with synthetic river flow records or even reconstructed for times without river records using historical weather data[56, 57].

Recommendation: Filled data gaps contain assumptions <u>not</u> observations, so beware the techniques used to create apparently complete records to avoid (re)interpreting those assumptions.

DIRTY DOZEN II AND III: POST-PROCESSING AND ARCHIVING ERRORS

Space limitations mean that we have only scratched the surface of the full range of biases and errors that can occur in a hydrological information flow, between site selection and eventual dissemination of data (**Figure 13**). Related disciplines, such as ecology and water quality, would be subject to many

of the same uncertainties such as concerns about instrument drift, fouling, or truncation settings, as well as about equipment maintenance, calibration, and routine updating of instrument logs/meta data to help interpret outliers in data.
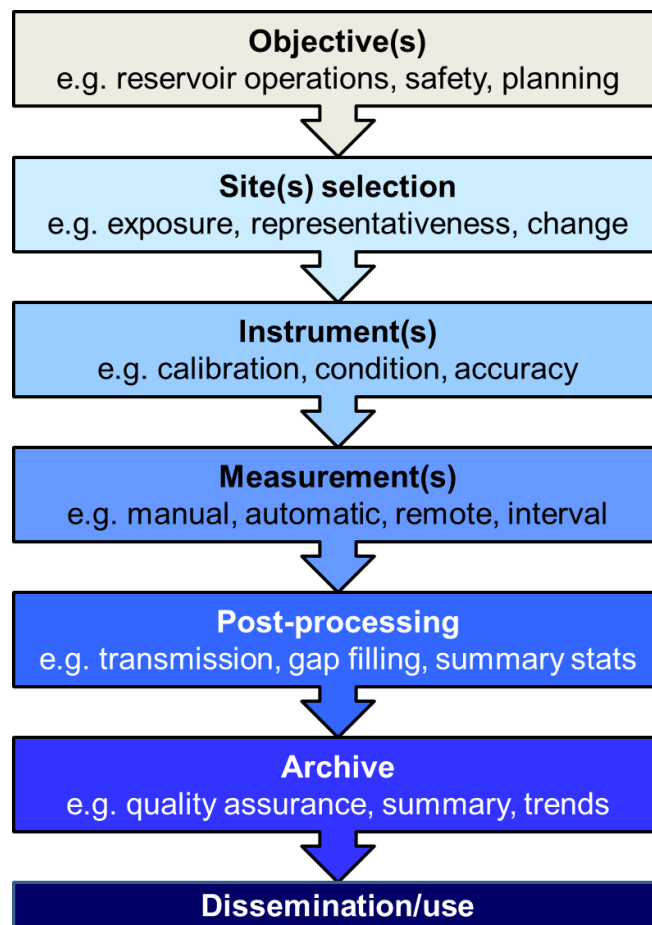
FIGURE 13 An information-flow that begins by setting project objectives and ends with data archiving, dissemination and use. Data biases and errors can enter the information-flow at any point in between.

Table 1 lists other sources of uncertainty that may be encountered by field hydrometry. Here, a distinction is made between errors (E) that relate to problems with instrumentation or measurement practices and biases (B) that are due to changing catchment conditions (outside the control of the field technician). **Table 2** gives examples of errors that can arise at the other end of the process at the point of archiving, with indications of how they might be detected. Ideally, instrument logs would be maintained and made available for open inspection. Such checks might be feasible for individual sites, instruments or records but impracticable for very large data sets – it is simply too labour intensive to visually inspect all entries. Hence, these types of error can present hidden dangers to users of 'global' sets compiled from multiple networks, with varying standards of data collection, types of instrumentation and quality assurance protocols.

TABLE 1 Dirty dozen II: Other causes of errors (E) and biases (B) in hydrometric data

#	Cause	Effect
1	Signal errors in electromagnetic gauging (E)	Unexpected peaks in river flow
2	Aquatic weed growth or weed cutting (E)	Higher or lower than expected water levels
3	Over-flowing rain-gauges (E)	Underestimated heavy rainfall events
4	Over-topping and by-passing gauging stations (E)	Truncated high flows
5	Ice jams and frozen equipment (E)	Constant, elevated or zero values returned
6	Silt enters the stilling well of a float-based water-level gauge (E)	Truncated low flows
7	Road building (B)	Artificial trend in high flows and siltation of instruments
8	Reservoir construction and filling (B)	Temporary and/or unexpected reduction in flows at downstream sites
9	Urban expansion (B)	Shorter time to peak and higher maximum discharges
10	Effluent returns and abstractions (B)	Higher or lower than expected minimum flows
11	Sluice operations (B)	Unexpected spikes in flow
12	Subsidence of equipment due to groundwater abstraction, mining, urban development (B)	Changes to gauge datum with ramifications for stage, rating relationships and flood frequency estimation

TABLE 2 Dirty dozen III: Errors at the point of data transmission and archiving. See also **FIGURE 14**.

#	Causes	Detection
1	Missing days in leap years	Fewer values than expected when sorted by date
2	Double entries of data	More values than expected when sorted by date
3	Decimalization or not (e.g. tenths to whole millimetres of precipitation)	Step change(s) in time series plot
4	Fabricated data	Comparison with neighbouring stations; statistical tests such as BL
5	Miscoded or changed units (e.g. inches to mm; Fahrenheit to Celsius)	Step change(s) in time series plot
6	Truncation and rounding errors	Statistical tests such as BL
7	Inconsistent use of missing data codes	Blocks of data with the same value (e.g. -999 then 198)
8	Suspect or erroneous data coded as zero	Unrealistic occurrences of zero conditions (e.g. river flow)
9	Data entry or key stroke error(s) for manually digitized records	Values outside expected range (four sigma test); comparison with neighbouring sites
10	Miscoded or classified variable (e.g. relative humidity stored as temperature or *vice versa*)	Values outside expected range (four sigma test); comparison with neighbouring sites
11	Values outside calibration/rating curve range	Truncated peak values
12	Regional variations in date (e.g. dd/mm/yy or mm/dd/yy) and decimal (e.g. "," or ".") formats	End of file error messages or data mis-feeds when importing records

Although there is now a tendency towards increased automation of environmental monitoring and quality assurance, there is still high dependency on manual techniques, not least for instrument calibration or evaluation of unexpected results. As we have shown, suspiciously high or low values are not always wrong (see: Exhibits #10 and #11). Moreover, the power to detect outliers and change-

points depends on the choice of the statistical techniques deployed[58]. Such considerations underline the importance of metadata and other circumstantial evidence (not least local knowledge) for ratifying hydrometerological data[59]. Once an issue is detected, the question then arises as to how to handle the error? Ideally, the archivist would set up processes to enable capture of user-community feedback. On the other hand, perhaps one of the conditions attached to the freedom of data access should be a responsibility on users to report errors.
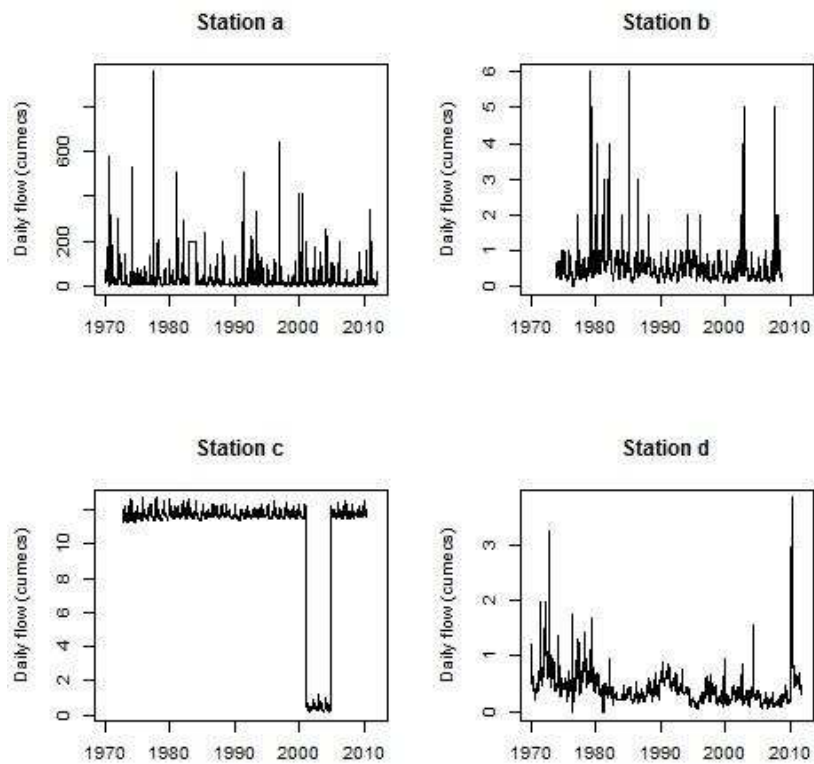
Figure 14 Archiving errors in river flow records: Station a – missing data code (198) interpreted as actual flow data; Station b – rounding of flows greater than 1 cubic metre per second (cumec) to whole integers; Station c – decimalisation change; and Station d – suspect low values set to zero. Data sources: (a): FRIEND European Water Archive (EWA), Germany; (b): German Federal Institute of Hydrology (BfG), Germany (including data acquired from the water authorities of the German Federal States); (c): Office of Public Works (OPW), Ireland; (d): Centre for Hydrographic Studies (CEDEX), Spain.

We have focussed on individual records but the representativeness of the observing network of stations as a whole matters just as much (if not more). Benchmark networks such as the US Geological Survey Hydrologic Benchmark Network[60] and the UK Reference Hydrological Network[61] are comprised of catchments with near-natural conditions and good quality gauge records. Unfortunately, such networks of reference stations are under constant pressure to rationalize and demonstrate cost-benefits. Facilities such as the FRIEND European Water Archive, the UK Acid Waters Monitoring Network and the UK Environmental Change Network all provide a basis for tracking long-term environmental trends (e.g. ref 24).

However, benchmark networks are also critical points of reference for cross-validating data. Measures such as the Representative Catchment Index (RCI) and Catchment Utility Index (CUI) show the extent to which individual gauging station records are amenable to regionalization or comparable with other sites[62]. Indicators of hydrometric data quality, completeness and provision also provide a basis for stabilising 'fluctuating' networks and setting levels of service provision[63]. Increasingly, the case is being made for more holistic measures of data quality that reflect the overall information life-cycle and utility of the data to users (**Figure 13**), rather than a few conventional quality indictors (e.g. record completeness).

CONCLUSIONS

Hydrological data biases and errors are a fact of life but early detection and attribution can help to minimise the risk of costly/poor/dangerous decisions later on. Indeed, future work might catalogue instances where data errors and/or biases have directly changed a management decision or led to a different outcome. One notorious example from space-engineering is the burn up of the Mars Climate Orbiter because different units were used by the constructors (imperial) and modellers (metric) of the satellite's thrusters. Just as systems engineers have examined the causes of famous failures[64] similar appraisals might be undertaken of the robustness of, for example, local flood protection schemes and national water policies to data biases and errors.

We have illustrated a range of techniques but the most dependable are: a) visual inspection of raw data; b) simple line, bar and scatter charts to display changes over time or to compare data from neighbouring sites; c) basic outlier and trend diagnostics; and d) reference to high quality metadata to aid interpretation of unexpected values or abrupt changes in data. Above all, it is necessary to have a critical mind-set when interrogating any field data. Such precautions are not only valid for hydrologists, ecologists and water quality specialists – they are just as essential for other environmental and social science disciplines.

During periods of austerity, conventional observing networks tend to be rationalized. With scarcer resources there is likely to be growing reliance on data gathered by automated systems, non-experts ('citizen scientists') or via the amalgamation of disparate information sources ('big data'). As data sets grow in size and complexity, users may become even more distanced from the processes that produced them – the real danger is that such data are deployed uncritically or in good faith. Hence, the case for building data literate communities has never been stronger.

ACKNOWLEDGEMENTS

REFERENCES

1. Davey CA, Pielke RA. 2005. Microclimate exposures of surface-based weather stations: implications for the assessment of long-term temperature trends. *Bulletin of the American Meteorological Society*, **86**:497-504.

2. World Meteorological Organisation (WMO). 2011. *Guide to Climatological Practices (WMO/No 100)*. World Meteorological Organisation: Geneva, Switzerland.

3. Herschy RW. 2008. *Streamflow Measurement*, Third Edition. CRC Press, Abingdon.

4. Shaw EM, Beven KJ, Chappell NA, Lamb R. 2011. *Hydrology Practice*. Spon Press, Abingdon, UK.

5. Fry MJ, Swain O. 2010. Hydrological data management systems within a national river flow archive. In: Kirby, Celia, (ed.) *Role of Hydrology in Managing Consequences of a Changing Global Environment*. Proceedings of the BHS Third International Symposium. *British Hydrological Society*, 808-815.

6. Bates BC, Chandler RE, Bowman AW. 2012. Trend estimation and change point detection in individual climatic series using flexible regression methods. *Journal of Geophysical Research*, **117**:D16106.

7. Noone S, Murphy C, Coll J, Matthews T, Mullan D, Wilby RL, Walsh S. 2016. Homogenization and analysis of an expanded long-term monthly rainfall network for the Island of Ireland (1850–2010). *International Journal of Climatology*, **36**:2837-2853.

8. McElwain L, Sweeney J. 2007. *Key Meteorological Indicators of Climate Change in Ireland*. Environmental Protection Agency, Johnstown Castle Estate: Wexford, Ireland, 31pp.

9. Kiely G. 1999. Climate change in Ireland from precipitation and streamflow observations. *Advances in Water Resources*, **23**:141-151.

10. Harrigan S, Murphy C, Hall J, Wilby RL, Sweeney J. 2014. Attribution of detected changes in streamflow using multiple working hypotheses. *Hydrology and Earth System Sciences*, **18**:1935–1952.

11. Merz B, Vorogushyn S, Uhlemann S, Delgado J, Hundecha, Y. 2012. HESS Opinions "More efforts and scientific rigour are needed to attribute trends in flood time series". *Hydrology and Earth System Science*, **16**:1379–1387.

12. Piqué G, Batalla RJ, Sabater S. 2016. Hydrological characterization of dammed rivers in the NW Mediterranean region. *Hydrological Processes*, **30**:1691-1707.

13. Webb BW. 1995. Regulation and thermal regime in a Devon river system. In: Foster, I.D.L., Gurnell, A.M. and Webb, B.W. (Eds) *Sediment and Water Quality in River Catchments*. John Wiley & Sons, Chichester.

14. Gustard A. 1989. Compensation flows in the UK: A hydrological review. *Regulated Rivers: Research and Management*, **3**:49-59.

15. Döll P, Fiedler K, Zhang J. 2009. Global-scale analysis of river flow alterations due to water withdrawals and reservoirs. *Hydrology and Earth System Sciences*, **13**:2413-2432.

16. Wilby RL, Jones PD, Lister DH. 2011. Decadal variations in the nocturnal heat island of London. *Weather*, **66**:59–64.

17. Parker DE. 2006. A demonstration that large-scale warming is not urban. *Journal of Climate*, **19**:2882-2895.

18. Hao L, Sun G, Liu Y, Wan J, Qin M, Qian H, Liu C, Zheng J, John R, Fan P, Chen J. 2015. Urbanization dramatically altered the water balances of a paddy field-dominated basin in southern China. *Hydrology and Earth System Sciences*, **19**:3319–3331.

19. Oke TR. 1982. The energetic basis of the urban heat island. *Quarterly Journal of the Royal Meteorological Society*, **108**:1-24.

20. Jones PD, Lister DH, Li Q. 2008. Urbanization effects in large-scale temperature records, with an emphasis on China. *Journal of Geophysical Research: Atmospheres*, **113**:1-12.

21. Hausfather Z, Menne MJ, Williams CN, Masters T, Broberg R, Jones, D. 2013. Quantifying the effect of urbanization on U.S. Historical Climatology Network temperature records. *Journal of Geophysical Research*: *Atmospheres*, **118**:481–494.

22. Menne MJ, Williams CN. 2009. Homogenization of temperature series via pairwise comparisons. *Journal of Climate*, **22**:1700–1717.

23. Hansen J, Ruedy R, Sato M, Lo K. 2010. Global surface temperature change. *Reviews of Geophysics*, **48**:RG4004.

24. Slater L, Villarini G. 2016. On the impact of gaps on trend detection in extreme streamflow time series. *International Journal of Climatology*, doi:10.1002/joc.4954.

25. Hannaford J. 2015. Climate-driven changes in UK river flows: A review of the evidence. *Progress in Physical Geography*, **39**:29-48.

26. Carter RW, Davidian J. 1968. Chapter A6 - General procedure for gaging streams, Book 3 - Applications of hydraulics, in *Techniques of Water Resources Investigations of the USGS*, pp1–13.

27. Smelser MG, Schmidt JC. 1998. *An assessment methodology for determining historical changes in mountain streams*. United States Department of Agriculture, Forest Service General Technical Report RMRS-GTR-6, 1-32.

28. U.S. Geological Survey 2016. *Water Data for the Nation– National Water Information System: Web Interface* http://waterdata.usgs.gov/nwis [Accessed 04/10/16]

29. Slater LJ, Villarini G. 2016. An impact perspective on recent trends in US flood risk, submitted to *Geophysical Research Letters*, under review.

30. Slater LJ, Singer MB, Kirchner JW. 2015, Hydrologic versus geomorphic drivers of trends in flood hazard, *Geophysical Research Letters*, **42**:1-7.

31. Morlot T, Perret C, Favre A-C, Jalbert J. 2014. Dynamic rating curve assessment for hydrometric stations and computation of the associated uncertainties: Quality and station management indicators. *Journal of Hydrology*, **517**:173-186.

32. Sambridge M, Tkalčić H, Jackson A. 2010. Benford's law in the natural sciences. *Geophysical Research Letters*, **37**:L22301.

33. Nigrini MJ, Miller SJ. 2007. Benford's Law applied to hydrology data – results and relevance to other geophysical data. *Mathematical Geology*, **39**:469-490.

34. Daly C, Gibson WP, Taylor GH, Doggett MK, Smith JI. 2007. Observer bias in daily precipitation measurements at United States Cooperative Network Stations. *Bulletin of the American Meteorological Society*, **88**:899-912.

35. Viney NR, Bates BC. 2004. It never rains on Sunday: The prevalence and implications of untagged multi-day rainfall accumulations in the Australian high quality data set. *International Journal of Climatology*, **24**:1171-1192.

36. Wilby RL, Tomlinson OJ. 2000. The "Sunday Effect" and weekly cycles of winter weather in the UK. *Weather*, **55**:214-222.

37. Skeffington RA, Halliday SJ, Wade AJ, Bowes MJ, Loewenthal M. 2015. Using high-frequency water quality data to assess sampling strategies for the EU Water Framework Directive. *Hydrology and Earth System Sciences*, **19**:2491-2504.

38. Wilby RL, Johnson MF, Toone JA. 2014. Nocturnal river water temperatures: Spatial and temporal variations. *The Science of the Total Environment*, **482-483**: 157-173.

39. Toone JA, Wilby RL, Rice SP. 2011. Surface-water temperature variations and river corridor properties. *Water Quality: Current Trends and Expected Climate Change Impacts*. IAHS Publ. **348**:129-134.

40. Orr HG, Simpson GL, des Clers S, Watts G, Hughes M, Hannaford J, Dunbar MJ, Laize CLR, Wilby RL, Battarbee RW, Evans R. 2015. Detecting changing river temperatures in England and Wales. *Hydrological Processes*, **29**:752-766.

41. Worrall TP, Dunbar MJ, Extence CA, Laize CLR, Monk WA, Wood PJ. 2014. The identification of hydrological indices for the characterization of macroinvertebrate community response to flow regime variability. *Hydrological Science Journal* **59**:645-658.

42. Durance I, Ormerod SJ. 2009. Trends in water quality and discharge confound long-term warming effects on river macroinvertebrates. *Freshwater Biology*, **54**:388-405.

43. Monk WA, Wood PJ, Hannah DM, Wilson DA. 2008. Macroinvertebrate community response to inter-annual and regional river flow regime dynamics. *River Research and Applications*, **24**:988-1001.

44. McLelland SJ, Nicholas AP. 2000. A new method for evaluating errors in high-frequency ADV measurements. *Hydrological Processes*, **14**: 351-366.

45. Goring DG, Nikora VI. 2002. Despiking acoustic Doppler velocimeter data. *Journal of Hydraulic Engineering*, **128**:117-126.

46. Billota GS, Brazier RE. 2008. Understanding the influence of suspended solids on water quality and aquatic biota. *Water Research* **42**:2849-2861.

47. Clifford NJ, Richards KS, Brown RA, Lane SN. 1995. Laboratory and field assessment of an infra-red turbidity probe and its response to particle size and variation in suspended sediment concentration. *Hydrological Sciences Journal* **40**:771-791.

48. Loperfido JV, Just CL, Papanicolaou AN, Schnoor JKL. 2010. In situ sensing to understand diel turbidity cycles, suspended solids, and nutrient transport in Clear Creek, Iowa. *Water Resources Research* **46**:W06525.

49. Rice SP, Johnson MF, Mathers K, Reeds J, Extence C. 2016. The importance of biotic entrainment for base flow fluvial sediment transport. *Journal of Geophysical Research: Earth Surface* **121**:890-906.

50. Rice SP, Johnson MF, Reeds J, Extence C, Longstaff H. 2014. Diel patterns of suspended sediment flux and the zoogeomorphic agency of invasive crayfish. *Cuadernos de Investigación Geográfica*, **40**:7-27.

51. Hodgkins R, Cooper R, Wadham J, Tranter M. 2003. Suspended sediment fluxes in a High-Arctic glacierised catchment: implications for fluvial sediment storage. *Sedimentary Geology*, **165**:105-117.

52. Barker PA, Wilby RL, Borrows J. 2004. A 200-year precipitation index for the central English Lake District. *Hydrological Sciences Journal*, **49**:769-785.

53. Bayliss A, Reed D. 2001. The use of historical data in flood frequency estimation. *Report to Ministry of Agriculture, Fisheries and Food,* Centre of Ecology and Hydrology *http://nora.nerc.ac.uk/8060/1/BaylissRepN008060CR.pdf* (Accessed 3/12/2015).

54. Macdonald N. 2013. Reassessing flood frequency for the River Trent, Central England, since AD 1320. *Hydrology Research,* **44**:215-33.

55. Kjeldsen TR, Jones DA. 2009. A formal statistical model for pooled analysis of extreme floods. *Hydrology Research,* **40**:465-480.

56. Cameron D, Beven K, Tawn J, Blazkova S, Naden P. 1999. Flood frequency estimation by continuous simulation for a gauged upland catchment (with uncertainty), *Journal of Hydrology*, **219**:169-187.

57. Jones PD, Lister DH, Wilby RL, Kostopoulou E. 2006. Extended river flow reconstructions for England and Wales, 1865-2002. *International Journal of Climatology*, **26**:219-231.

58. Reeves J, Chen J, Wang XL, Lund R, Lu Q. 2007. A comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology*, **46**:900-915.

59. Aguilar E, Auer I, Brunet M, Peterson TC, Wieringa J. 2003. *Guidance on metadata and homogenization*. *WMO Td*, *1186*, pp53.

60. Murdoch PS., McHale MR, Mast MA, Clow DW. 2005. *The US geological survey hydrologic benchmark network.* Fact Sheet No. 2005-3135, US Geological Survey, pp6.

61. Bradford RB, Marsh TJ. 2003. Defining a network of benchmark catchments for the UK. *Proceedings of the Institution of Civil Engineers – Water and Maritime Engineering*, **156**:109-116.

62. Dixon H, Hannaford J, Fry MJ. 2013. The effective management of national hydrometric data: experiences from the United Kingdom. *Hydrological Sciences Journal*, **58**:1383-1399.

63. Muchan K, Dixon, H. 2014. Ensuring hydrometric data are fit-for-purpose through a national Service Level Agreement. *Hydrology in a Changing World: Environmental and Human Dimensions*. IAHS Publ. **363**:323-329.

64. Bahill AT, Henderson SJ. 2005. Requirements development, verification, and validation exhibited in famous failures. *Systems Engineering*, **8**:1-14.