Environmental concentrations as ratios of random variables

Saverio Perri\(^1\)∗ and Amilcare Porporato\(^{1,2}\)

\(^1\) High Meadows Environmental Institute, Princeton University, Princeton, NJ 08544, United States of America
\(^2\) Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ 08540, United States of America

∗ Author to whom any correspondence should be addressed.
E-mail: sperri@princeton.edu

Keywords: environmental concentrations, ratio distributions, risk assessment, nitrates, salinity, environmental change, heavy-tailed distributions

Abstract

Human-induced environmental change increasingly threatens the stability of socio-ecological systems. Careful statistical characterization of environmental concentrations is critical to quantify and predict the consequences of such changes on human and ecosystems conditions. However, while concentrations are naturally defined as the ratio between solute mass and solvent volume, they have rarely been treated as such, typically limiting the analysis to familiar distributions generically used for any other environmental variable. To address this gap, we propose a more general framework that leverages their definition explicitly as ratios of random variables. We show that the resulting models accurately describe the behavior of nitrate plus nitrite in US rivers and salt concentration in estuaries in the Everglades by accounting for heavy tails potentially emerging when the water volume fluctuates around low values. Models that preclude the presence of heavy tails and the related high probability of extreme concentrations could significantly undermine the accuracy of diagnostic frameworks and the effectiveness of mitigation interventions, especially for soil contamination characterized by a water volume (i.e. soil moisture) frequently approaching zero.

1. Introduction

Anthropogenic activities and climate change are responsible for deteriorating water, soil, and air quality, calling for improved diagnostic frameworks to assess primary contamination threats and devise appropriate remediation interventions [1–4]. As increased pollutant concentrations pose substantial threats to human health and, more generally, to the Earth’s system resilience [5–7], environmental concentrations are key variables to quantify the consequences of environmental changes [8, 9].

The impact of pollutants on humans, plants, and animals can be characterized by either a threshold or non-threshold response [10]. The majority of carcinogenic contaminants are generally described by a non-threshold behavior [11, 12], meaning that their adverse effects occur at all levels of exposure with linearly increasing risk associated with increasing concentration. On the contrary, most environmental contaminants result in threshold response [13, 14], with non-linear exposure-response curves, where vulnerability suddenly escalates as critical thresholds are exceeded [15, 16], growing exponentially for concentrations larger than a threshold value [17–20].

Threshold response behaviors arise, for example, from the response of ecological communities to contaminants [14, 21]. Multi-species assemblies of plants and animals are known to be resilient to biotic and abiotic stress, with species diversity being positively correlated with ecosystem stability [13]. Yet, abrupt changes and regime shifts take place when critical concentrations are crossed [14]. These events tend to be related to the presence of ecological tipping points [22–24]. Sea-level rise, for instance, has caused significant salinization of coastal wetlands due to seawater intrusion [25, 26]. Vegetation can cope with increasing salinity up to a species-specific salt-tolerance threshold, after which mortality rates soar and regime shifts to salt-tolerant vegetation occur [27–31]. Such
A major challenge with analyzing environmental quantities is that their concentration exhibits high spatio-temporal variability [32, 33]. This variability is the result of stochastic geophysical and hydroclimatic processes occurring at various temporal and spatial scales [34–36]. For example, the concentration of nitrate (NO$_3^-$) and nitrite (NO$_2^-$) in rivers depends on rainfall-induced leaching events, water temperature, crop development stage, and other specific factors (e.g. frequency, timing, and magnitude of agricultural fertilization) [37–42]. Similarly, rivers salinity is a function of rainfall, evapotranspiration, rock weathering, atmospheric salt deposition, and seawater intrusion into estuarine ecosystems [43–46].

Because of this highly fluctuating and unpredictable nature, environmental concentrations have been studied through the lens of probabilistic tools [47, 48]. However, unlike any generic random variable, they possess very specific properties due to their mathematical structure [49]. By definition, the concentration of an environmental quantity is the ratio between solute mass and solvent volume. This aspect has been largely overlooked in the literature, potentially causing inaccurate estimates of the probability of harmful concentrations. In fact, as solvent volumes decrease for various reasons, it is not uncommon that these ratios tend to very large values, resulting in heavy-tailed probability distributions [49, 50], which are characterized by sub-exponential tails linked to elevated probabilities of extreme values [51, 52].

One of the most intuitive examples of such behavior is represented by soil contaminants. Although the mass of a contaminant in the soil varies at times scales order of magnitude slower than soil water content, frequent variations in soil moisture generate extreme fluctuations in the concentration [29, 53] (see figure 1). During a drydown, soil moisture can approach extremely low values, and, as a result, the concentration can quickly spike to critical levels [29]. Similar patterns could be observed for atmospheric aerosols that are known to absorb pollutants and facilitate their transport [55, 56]. They also transport pathogens like viruses, which concentration (and associated hazard) increases dramatically as the droplet volume and diameter are reduced by evaporation [57, 58]. The increasing pathogen concentration and reduced droplet size increase the hazard and biological uptake efficiency until the droplet fully evaporates and the pathogen becomes inactive [59, 60]. Hence, an accurate prediction of extreme concentrations necessitates accounting for heavy tails related to ratio distributions, and specific probability density functions (pdf) should be employed for this purpose [50].

In this work, we analyze general probabilistic models that can be obtained by treating environmental concentrations as the ratio of two random variables, thus accounting for their specific mathematical properties. While, a variety of distributions has been already recommended in the literature (figures 2(A) and (B)) [61–63], these models are not tailored to ratios of random variables. Among the most widely used, we can recall Exponential [64–66], Log-normal [67–69], Gamma [70–73], Beta [74–76], Pareto [75–77], and Extreme Value Distributions (e.g. the Weibull distribution) [76, 78–81]. Most of these distributions can be successfully adopted in scenarios where the volume of solvent never approaches zero or when mass fluxes are strongly correlated with discharge (see section 2.2) [82–85]. This is often the case for atmospheric plumes characterized by turbulent diffusion, and rapid mixing [64, 84, 86, 87]. However, they fall short in the more critical cases in which the volume fluctuates around small values. Extreme Value and Pareto distributions can account for the asymptotic behavior of extreme conditions and, therefore, tend to perform better than other classic distributions in predicting critical values in water, soil, and biota. However, their applicability remains limited, as they do not consider the possibility that the concentration distribution may be strongly skewed to the right (or even diverge) when the volume of solvent fluctuates near zero [8], and they do not typically result from ratios of random variables.

Based on these premises, we propose a more general framework to interpret concentration data and account for right tails heavier than the exponential decay. We apply our framework to the concentration pdf obtained as the ratio of two beta distributions. While beta distributions can reproduce the behavior of most environmental variables constrained by an upper limit [88, 89], our approach applies to the ratio of different distribution families [49]. The theoretical pdf’s are tested for nitrate plus nitrite and salt concentration in US rivers and coastal streams. To further emphasize the importance of accounting for heavy tails, we use a stochastic soil contamination model that reproduces the dynamics of a generic contaminant in a semi-arid environment. Our ratio distribution shows good agreement with data, especially in the right tails, of interest in many applications. The role of statistical dependence is also discussed, as the mass of contaminants and volume of solvent may be correlated. Finally, the broad implications of the properties of ratio random variables on data analysis and signal processing are briefly discussed.

2. Probabilistic description of concentration fluctuations

The solute concentration $C$ of an environmental quantity is defined as the ratio between solute mass $M$ and solvent volume $V$:
Figure 1. (A)–(C) Numerical simulations of soil moisture (A), salt mass per unit of ground area (B), and salinity (C) as described by equations (9) and (10). The dynamics are obtained by forcing the system with random inputs of precipitation modeled as a compound Poisson process. The hydroclimatic conditions are representative of semi-arid regions (average rainfall about 300 mm per year). The semi-transparent lines represent an ensemble of possible realizations. Panels (A.1)–(C.1) display the histogram of the simulated variables. The parameters used in the simulations are typical of dry regions with $n = 0.45$, $Z_r = 30$ cm, effective soil moisture at field capacity $s_1 = 0.8$, and maximum evapotranspiration rate $ET_{max} = 4.5$ mm d$^{-1}$ (see [33, 54] for details).

(D) Right tail of the empirical concentration distribution. The histogram is compared to the best fit of exponential decay (dashed-dotted red line) and power-law decay (green line).

Figure 2. (A), (B) Examples of probability density function (pdf, (A)) and cumulative distribution function (CDF; (B)) of classic pdf’s used to model environmental concentration. (C), (D) Pdf (C) and CDF (D) of the ratio between two Beta distributions for $C_{max} = 8$ mg L$^{-1}$ and various parameters combinations.
The distribution of \( C \) is obtained as the ratio distribution of two continuous non-negative random variables \([49, 89]\):

\[
p(C) = \int_0^\infty V \varphi(CV, V) dC dV,
\]

where \( \varphi(CV, V) \) is the joint probability density function of \( M \) and \( V \). As discussed in the following sections, the shape of the \( p(C) \) depends on the properties of the \( M \) and \( V \) distributions as well as the strength of their statistical dependence.

### 2.1. Ratio distribution for statistically independent mass and volume

In the case of independent random variables, the joint distribution of \( M \) and \( V \) reduces to the product of their marginal pdfs: \( \varphi(M, V) = p(M) \cdot p(V) \). This hypothesis allows for an exact analytical derivation of \( p(C) \) for several marginal \( M \) and \( V \) distributions \([49]\) and is remarkably realistic when different stochastic processes drive the dynamics of mass and volume. Anthropogenic point-input of a contaminant in a natural water body or soil salinization—where soil moisture evolves at time scales orders of magnitude faster than salt mass \([91]\)—are exemplary cases of largely independent solute and solvent dynamics.

As a typical and important case, we solved equation (2) for marginal pdf’s derived from a Beta distribution. Beta distributions have been used to characterize a number of environmental variables, as they can accurately describe most quantities that lead to substantial percolation of contaminants from the soil to rivers can feature simultaneous peaks in solute mass and water volume (see mass and volume dynamics in figure 3). This raises the question of how statistical dependence affects concentration fluctuations and its pdf.

\[
p(M) = \frac{M_{\max}^\alpha (1 - M_{\max})^{\beta - 1}}{M_B[\alpha, \beta]} \cdot \frac{(1 - M_{\max})^{\gamma - 1}}{M_B[\gamma, \delta]},
\]

\[
p(V) = \frac{V_{\max}^\gamma (1 - V_{\max})^{\delta - 1}}{V_B[\gamma, \delta]},
\]

where \( B[\cdot, \cdot] \) is the Beta function, and \( \alpha, \beta, \gamma, \delta \) shape parameters.

Finally, solving equation (2) for (5) and (6), the concentration distribution reads:

\[
p(C) = N C^\alpha \sum_2 \beta^\gamma [1 - \beta, \epsilon - \delta, \alpha; \epsilon, \alpha + 1; C/C_{\max}],
\]

where \( N \) is a constant,

\[
N = \sum_2 \beta^\gamma [\alpha, 1 - \beta, \epsilon - \delta; \alpha + 1, \epsilon; 1],
\]

where \( \sum_2 \beta^\gamma [\cdots; \cdots; \cdots] \) is the regularized generalized hypergeometric function, and \( \epsilon = \alpha + \gamma + \delta \). Figures 2(C) and (D) shows equation (7) for various combinations of the shape parameters. The ratio distribution tends to display heavy tails or power-law growth, especially when the \( p(V) \) is skewed to the left (black curves in figures 2(C) and (D)). This implies high probabilities of crossing critical values. Even in the case of symmetrical \( p(M) \) and \( p(V) \) (green curves in figures 2(C) and (D)), the \( p(C) \) shows elevated probabilities of concentration approaching the maximum \( (C_{\max} = 8 \text{ mg L}^{-1} \text{ in this example}) \). Classic probability distributions may be suitable in the case of \( p(V) \) skewed to the right (orange and red curves in figures 2(C) and (D)), as heavy tails do not emerge in this scenario.

### 2.2. Impact of statistical dependence of mass and volume

Environmental concentrations may result from the ratio of statistically dependent variables. For instance, the assumption of statistical independence may not hold for diffuse contamination processes. Events that lead to substantial percolation of contaminants from the soil to rivers can feature simultaneous peaks in solute mass and water volume (see mass and volume dynamics in figure 3). This raises the question of how statistical dependence affects concentration fluctuations and its pdf.

We investigate the impact of statistical dependence on the ratio distribution by extracting two random \( M \) and \( V \) vectors from a Beta copula \([95–98]\). The ratio \( C = M/V \) was then obtained for a given set of shape parameters of the marginal Beta distributions and different correlation matrices (figure 4).

This analysis suggests that strong positive statistical dependence between mass and volume reduces the probability of encountering elevated concentrations (figure 4(C)), as it is unlikely that high solute mass co-occurs with low solvent volume. Hence, in the case of a strong positive correlation, the absence of heavy tails
Figure 3. (A) Map of the major rivers in the contiguous USA with the location of the USGS sites selected in this study. Nitrate plus nitrite data were collected along major rivers (sites 1–3), while salinity was measured in coastal wetlands in the Florida Everglades (sites 4–6). The inset in (A) shows a Google Earth zoom-in of the area of interest in the Everglades. (B)–(G) Time-series of solute mass (nitrate plus nitrite in (B) and salt in (E)), discharge volume ((C) and (F)), and concentration ((D) and (G)) at site 2 (B)–(D) and site 4 (E)–(G) for the hydrological year June 2019–May 2020. Nitrate plus nitrite concentrations were obtained by USGS using the technique of colorimetric determination in water by enzymatic reduction with precision \( \pm 1\% \) \[93\], while salinity was estimated from specific conductance, typically accurate to \( \pm 5\% \) \[94\]. The map of the major rivers in the contiguous USA was reproduced from Esri (2010). Sources: Esri; Rand McNally; Bartholemew and Times Books; Digital Chart of the World (DCW); U.S. National Geospatial-Intelligence Agency; i-cubed.

allows treating environmental concentrations as regular random variables, and classic pdf’s may be used to estimate exceedance probabilities.

On the contrary, when mass and volume are negatively correlated, \( p(C) \) tends to have heavier tails (figure S1I available online at stacks.iop.org/ERL/17/024011/mmedia). Although negative statistical dependence seems unlikely to occur in the environment, it could cause underestimates of high concentration return periods. Therefore, when strong negative correlations are observed, the joint pdf should be used to solve equation (2).

3. Case studies

Nitrogen and salt concentration in streams, rivers, and estuaries are excellent examples of environmental concentrations forced by stochastic hydroclimatic processes \[46, 99, 100\], raising growing concern. Over the past decades, excessive nitrogen fertilization has resulted in increasing nitrogen losses to leaching \[101\]. Rising concentration of \( \text{NO}_3^- \) and \( \text{NO}_2^- \) in rivers in turn has caused eutrophication and contamination of surface- and ground-water resources \[102–104\]. Analogously, aridification and increasing
seawater intrusion into estuarine ecosystems due to sea-level rise have exacerbated rivers, and soil salinity [46, 105, 106]. As a result, salinization reduces the productivity of a natural and managed ecosystem, threatening to become a global environmental issue [46, 106–108]. For these reasons, and given the extensive network of measurements available, the concentrations of salts and nitrogen compounds in rivers and estuaries are chosen here as case studies to test our probabilistic framework.

3.1. Nitrate concentration in rivers

Historical hourly optical measurements [109] of NO$_3^-$ plus NO$_2^-$ over the period 2014-2021 were obtained from the U.S. Geological Survey (USGS) at three station along the Mississippi (Site 1; USGS 07289000), Missouri (Site 2; USGS 06934500), and Ohio river (Site 3; USGS 03612600, figure 3). For each site, concentration time series were combined with discharge data to derive water volumes (figures 3(C) and (F)) and solute mass (figures 3(B) and (E)) via integration under the assumption of well-mixed conditions. Finally, all the data were aggregated at the daily timescale and the empirical concentration distributions were used to fit the theoretical pdf (equation (7)).

A good agreement with the empirical pdf was observed across all the sites (figures 5(A)–(C)). While nitrate mass and discharge volume are correlated ($\rho \geq 0.8$), our distribution is particularly effective in describing the right-tail and the slow decay of $p(C)$. Because of discontinuous fertilizer applications and seasonality in plants nitrogen uptake, nitrate concentration peaks occur at the beginning of the growing season (April–June, figure 3(D)). Such high NO$_3^-$ plus NO$_2^-$ values do not necessarily depend on large
Figure 5. Comparison between observed environmental concentrations in six USGS stations (histograms) and the ratio distribution (solid lines) obtained in equation (7) for nitrate plus nitrite (A–(C)) and salinity (D–(F)). Concentrations are obtained as daily averages under the assumption of perfect mixing.

discharge volumes (see figure 3(C)), but rather on the amount of nitrate available in the basin soil. Under these conditions, the Beta distribution of the solute mass tends to collapse to an Exponential distribution with frequent values near zero. Rivers discharge, on the contrary, could be well described by a Log-normal distribution \[110, 111\], of which the Beta distribution represents a more general and viable alternative \[112\].

3.2. Salinity in coastal wetlands

Salinity data at 15 minutes temporal resolution for the 2007–2021 period were retrieved from three sites in the Florida Everglades: Taylor River at Mouth (Site 4; USGS 251127080382100), Trout Creek (Site 5; USGS 251253080320100), and East side creek (Site 6; USGS 250802081035500). The data were processed similarly to the nitrogen data as described in section 3.1.

Also for salinity, the model shows excellent agreement with observations and captures the high probability of values greater than seawater salinity \((\approx 35 \text{ g L}^{-1}; \text{figures } 5(\text{D})–(\text{F}))\). While salinity is usually lower than the average seawater concentration, high evapotranspiration and scarce freshwater inputs from rainfall and rivers can occasionally reduce the volume of water and result in high salinity. In the Everglades, the increase in salinity is particularly evident during the dry season, when lowering water levels enhance seawater intrusion and inundations \[113, 114\]. The frequency of these peaks in salinity may increase in the near future due to sea-level rise and consequent coastal ecosystems salinization \[27, 115\].

Compared to nitrate concentration, salt mass and water volume at Sites 4 and 5 show weaker positive statistical dependence \((\rho = 0.4 \text{ for Site 4 and } \rho = 0.5 \text{ for Site 5})\). Part of the reason is that salt mass encroaches into coastal streams as a result of periodic inundations and seawater intrusion \[116\], and it is not transported by upstream river discharge. On the contrary, strong positive correlation is observed at Site 6 \((\rho = 0.98)\), where water fluxes are dominated by tides, which carry most of the salt mass. Although the salt mass distribution is still skewed to the left, the discharge volume tends to display a more symmetrical distribution due to tidal inundations. Similarly to the nitrate case study, the Beta distribution thus represents a good general fit for both \(M\) and \(V\).

3.3. Soil contamination

While nitrate and salt concentration represent an important example of environmental concentrations that should be generally treated as a ratio of random variables, large and interconnected water bodies tend to have relatively stable properties \[117, 118\], and their water volume rarely approaches zero. A more compelling case of ratio concentration is the one of conservative soil contaminants, such as heavy metals and salt (NaCl). Unlike water volume in rivers (figures 3(C) and (F)), the volume of solvent in the soil, namely soil moisture, can quickly and frequently approach extremely low values during drydowns \[119\]. As a result, the concentration occasionally reaches critical values, and heavy tails in the probability distribution emerge. If concentration rises to the solubility limit, solute precipitation could take place and in this case the approach needs to be modified accordingly. However, this limit is unlikely reached for NaCl \((C_{\text{max}} \approx 360 \text{ g L}^{-1} \text{ at } 20 \degree \text{C})\) because maximum concentration is normally limited by the ratio between salt mass input and output \[120\] or the salinity at which vegetation water uptake is completely impaired \[29, 121\].
Unfortunately, long-term high-frequency soil quality data are much less available than nitrate and salt concentration in rivers. Here, this limitation is overcome by using a parsimonious stochastic model of coupled soil moisture and concentration dynamics based on previously proposed soil moisture dynamics [54] and salinization models [53], to which we refer for a complete description of the dynamical system. In this context, balance equations for the vertically averaged relative soil moisture \( s \equiv V \) and salt mass \( M \) dynamics can be written as:

\[
\frac{dM}{dt} = Y - C(t) L [s(t), t],
\]

\[
\frac{dZ_r}{dt} = P(t) - ET[s(t)] - L [s(t), t],
\]

where \( n \) is the soil porosity, \( Z_r \) the root depth, \( P(t) \) the stochastic precipitation input, \( ET[s(t)] \) the evapotranspiration rate depending on soil moisture, \( L [s(t), t] \) the deep percolation rate, \( Y \) is the average salt input from precipitation and dry deposition [53, 122], and \( C(t) L [s(t), t] \) represents leaching salt removal driven by percolation. The concentration of salt \( C \) is calculated as \( C = M / (nZ_r s) \). Stochastic precipitation is modeled as a compound Poisson process [123], and all the rain that cannot be allocated in the soil water storage \((nZ_r s_1)\) with \( s_1 \) being the effective soil moisture at field capacity) is assumed to be lost to deep percolation [54] at a frequency equal to the frequency of soil moisture crossing the threshold \( s_1 \) [124]. The model also assumes an instantaneous mixing of salt within the active soil depth [29, 53, 91].

Figure 1 shows relative soil moisture (figure 1(A)), solute mass (figure 1(B)) and solute concentration (figure 1(C)) resulting from equations (9) and (10), solved for a semi-arid region. It can be noticed that the dynamics of \( s \) and \( M \) are largely decoupled, and \( C \) trajectories closely follow those of \( s \). Drydowns result in exponentially increasing salinity, even when the mass of salt remains almost constant. As a consequence, the salinity probability distribution (figures 1(C.1) and (D)) exhibits a heavy tail with sub-exponential decay (figure 1(D)).

As shown by Suweis et al [53], the pdf of \( C \) can be analytically obtained by assuming complete statistical independence between \( s \) and \( M \). The resulting pdf, obtained as the ratio between the truncated Gamma distribution of \( s \) [54] and the Gamma distribution of \( M \) [120], displays a power-law decay. However, the hypothesis of statistical independence between \( s \) and \( M \) may not hold in highly vegetated and salt-affected ecosystems. There, salinity reduces \( ET \) depending on vegetation salt tolerance [29, 46]. Yet, even accounting for the salinity-induced reduction of \( ET \), \( s \) variability time scales are much shorter than those of \( M \), and heavy tails in the concentration distribution still emerge (see [29] for details).

4. Discussion

The analysis of environmental concentrations concerns applications spanning from ecotoxicology to ecology, biogeochemistry, and ecohydrology [125–130]. In particular, the prediction of the response of socio-ecological systems to environmental change and the design of proper interventions and mitigation strategies requires an improved probabilistic characterization of extreme concentrations [131]. Classic probability distributions widely used in the literature (figures 2(A) and (B)) do not capture the extreme nature of environmental concentrations fluctuations related to their definition as ratios between random variables.

Our alternative formulation of the concentration pdf is specifically derived from the ratio distribution of solute mass \( M \) and solvent volume \( V \) treated as random variables with Beta distributions. Similar formulations could be obtained as the ratio of variables extracted from different distribution families [49] and can accommodate statistical dependence when important. The ratio distribution proposed here accurately described observed concentrations of nitrate plus nitrite and salinity patterns. In particular, it accounts for the possibility that the pdf is particularly skewed to the right with fat tails (see figures 2(C), (D) and 5). Equation (7) can be used to estimate the probability of exceeding extreme concentration values and, therefore, can provide a direct estimate of the average return period of a contamination event.

Although it was obtained initially under the assumption that solute mass and solvent volume are independent, this framework can conservatively reproduce also the behavior of weakly positively correlated variables (figure 4). A positive correlation between mass and volume leads to less frequent extreme events, with faster decay of the distribution’s right tail. When positive correlations are particularly strong, however, the heavy tail tends to disappear (figures 4(C) and (F)), and neglecting statistical dependence could lead to significant overestimates of the probability of crossing critical values. Under these circumstances, simple Gamma or Log-normal distributions may capture the probabilistic behavior of concentration dynamics. Classic pdf’s could be reasonably employed also to analyze concentration data in situations in which the solvent volume remains conspicuously above null values at all times. This is the case with many atmospheric pollutants and water quality indicators. Yet, ratio distributions should be used in the circumstances like soil contaminants, characterized by highly variable solvent volume (i.e. soil moisture) and concentrations growing exponentially during a dry period (figure 1).

The emergence of heavy tails may also have implications for the analysis of tracers to understand chemical and ecohydrological processes at the
catchment scale. For example, the transport of contaminants in the whole catchment can be characterized by comparing the power spectra of concentration at the input (e.g. in the rainfall or at a point source) to the spectrum of concentration in the catchment discharge [132–134]. Because environmental concentrations are defined as the ratio of two variables, their variance can be extremely high (when the volume fluctuates near zero), calling for extreme care in evaluating power spectral scaling and any other time series analysis based on variance decomposition of concentrations.

Predicting harmful concentrations statistics will also imply dealing with nonstationarity of extreme-event probabilities due to climate change and anthropogenic activities [84, 135–138]; alternative methods should be implemented to relax the ergodicity assumption [139–141], while still taking advantage of the special mathematical properties of the concentration of random variables.

Data availability statement

The data that support the findings of this study are openly available at the Geological Survey’s (USGS) National Water Information System (NWIS) website.

Acknowledgments

Funding in support of this research was provided by Princeton University’s Dean for Research, High Meadows Environmental Institute, Andlinger Center for Energy and the Environment, and the Office of the Provost International Fund. We also acknowledge support from the US National Science Foundation (NSF) Grant Nos. EAR-1331846 and EAR-1338694, the BP through the Carbon Mitigation Initiative (CMI) at Princeton University, and the Moore Foundation.

ORCID iDs

Saverio Perri http://orcid.org/0000-0001-6382-1381

Amilcare Porporato http://orcid.org/0000-0001-9378-207X

References

[1] Meybeck M 2003 Phil. Trans. R. Soc. B 358 1935–55
[2] Melnick D, McNeely J and Navarro Y K 2005 Environment and Human Well-Being: A Practical Strategy (London: Earthscan)
[3] Vörösmarty C J et al 2010 Nature 467 555–61
[4] Mateo-Sagasta J, Zadeh S M and Tural H 2018 More People, More Food, Worse Water? A Global Review of Water Pollution from Agriculture (Rome: Food and Agriculture Organization of the United Nations (FAO) and International Water Management Institute (IWMI))
[5] Silva R A et al 2013 Environ. Res. Lett. 8 034005
[6] Diamond M L et al 2015 Environ. Int. 78 8–15
[7] Whitmee S et al 2015 Lancet 386 1973–2028
[8] Achley W R, Gaskins C T and Anderson D 1976 Syst. Biol. 25 137–48
[9] de Barros F P J and Fiori A 2014 Water Resour. Res. 50 4018–37
[10] Vesilind P A, Peirce J J and Weiner R F 2013 Environmental Pollution and Control (Amsterdam: Elsevier)
[11] Tchounwou P 1999 Rev. Environ. Health 14 211–30
[12] Orsini N, L R, Wolk A, Khudyakov P and Spiegelman D 2012 Am. J. Epidemiol. 175 66–73
[13] May R M 1977 Nature 269 471–7
[14] Clements W H and Rohr J R 2009 Environ. Toxicol. Chem. 28 1789–800
[15] Alloway B and Ayres D C 1997 Chemical Principles of Environmental Pollution (London: Blackie Academic & Professional)
[16] Kroeze R, Kleiner J and Renwick A 2005 Toxicol. Sci. 86 226–30
[17] Pope III C A 2000 Am. J. Epidemiol. 152 407–12
[18] Bell M L, Peng R D and Dominici F 2006 Environ. Health Perspect. 114 532–6
[19] Bell M L and Dominici F 2006 Epidemiology 17 5223
[20] Cox L A T 2018 Environ. Res. 164 635–46
[21] Connell J H and Sousa W P 1983 Ann. Nat. 121 789–824
[22] Scheffer M and Carpenter S R 2003 Trends Ecol. Evol. 18 648–56
[23] Horan R D, Fenichel E P, Drury K L S and Lodge D M 2011 Proc. Natl Acad. Sci. 108 7333–8
[24] Crépin A S, Biggs R, Polasky S, Troell M and de Zeeuw A 2012 Ecol. Econ. 84 15–22
[25] Michener W K, Blood E R, Bildstein K L, Brinson M M and Gardner L R 1997 Ecol. Appl. 7 770–801
[26] Yu M, Rivera-Oceasso E, Heartill-Scalley T, Davila-Casanova D, Rios-López N and Gao Q 2019 Wetlands 39 1343–55
[27] Saha A K, Saha S, Sadle J, Jiang J, Ross M S, Price R M, Sternberg L S O and Wendellberger K S 2011 Clim. Change 107 81–108
[28] Herbert E R et al 2015 Ecosphere 6 1–43
[29] Perri S, Suweis S, Entekhabi D and Molini A 2018 Geophys. Res. Lett. 45 669–82
[30] Pagherazzi S, Anisfeld S C, Blum L K, Long E V, Feagin R A, Fernandes A, Kearney W S and Williams K 2019 Front. Environ. Sci. 7 25
[31] Smart L S, Taille G P, Joulter B, Vukomanovic J, Singh K K, Swenson J J, Mitsasova H, Smith J W and Memetmeyri R K 2020 Environ. Res. Lett. 15 104028
[32] Roff D A 2002 Life History Evolution (Sunderland, MA: Sinauer Associates)
[33] Hipel K W 2012 Stochastic and Statistical Methods in Hydrology and Environmental Engineering: Stochastic and Statistical Modelling with Groundwater and Surface Water Applications vol 2 (Berlin: Springer)
[34] Yu W H, Harvey C M and Harvey C F 2003 Water Resour. Res. 39 1146–63
[35] Jerrett M, Arain A, Kanaroglou P, Beckerman B, Potoglou D, Sahuvarough T, Morrison J and Giovs C 2005 J. Expo. Sci. Environ. Epidemiol. 15 185–204
[36] Luque-Espinjar J A, Pardo-Igúzquiza E, Grima-Olmedo J and Grima-Olmedo C 2018 J. Hydrol. 561 348–71
[37] Mitchell J, McIsaac G, Walker S and Hirschi M 2000 Trans. ASAE 43 337–42
[38] Malvi A, Nouri J, Babaei A and Nabizadeh R 2005 Int. J. Environ. Sci. Technol. 2 41–47
[39] Schlesinger W H 2009 Proc. Natl Acad. Sci. 106 203–8
[40] Gascuel-Odoux C, Aurousseau P, Durand P and Ruiz I. and Molenaar J 2010 Sci. Total Environ. 408 5657–66
[41] Billen G, Ramarson A, Thieu V, Théry S, Silvestre M, Pasquier C, Hénault C and Garnier J 2018 Biogeochemistry 139 31–51
[42] Thomas Z and Abbott B W 2018 J. Contam. Hydrol. 215 51–61
[43] Pillsbury A F 1981 Sci. Am. 245 54–65
[118] Hansen C H, Burian S J, Dennison P E and Williams G P 2017 Remote Sens. 9 409
[119] D’Odorico P, Porporato A and Runyan C W 2006 Dryland Ecohydrology vol 9 (Dordrecht: Springer)
[120] Mau Y, Feng X and Porporato A 2014 Phys. Rev. E 90 21–28
[121] Perri S, Viola F, Noto L V and Molini A 2017 Hydrol. Process. 31 1271–82
[122] Hillel D 2000 Salinity Management for Sustainable Irrigation: Integrating Science, Environment and Economics (Washington, DC: The World Bank)
[123] Rodríguez-Iturbe I, Porporato A, Ridolfi L, Isham V and Coxi D R 1999 Proc. R. Soc. A 455 3789–805
[124] Botter G, Porporato A, Daly E, Rodriguez-Iturbe I and Rinaldo A 2007 Water Resour. Res. 43 W06404
[125] Moriarty F 1988 Hum. Toxicol. 7 457–457
[126] Carpenter S R, Caraco N F, Correll D L, Howarth R W, Sharples A N and Smith V H 1998 Ecol. Appl. 8 559–68
[127] Erisman J, Bleeker A, Galloway J and Sutton M 2007 Environ. Pollut. 150 140–9
[128] Botter G, Daly E, Porporato A, Rodriguez-Iturbe I and Rinaldo A 2008 Water Resour. Res. 44 W03416
[129] Perri S, Entekhabi D and Molini A 2018 Water Resour. Res. 54 2781–98
[130] Ashraf M A and Faheem M 2020 Environ. Sci. Pollut. Res. 27 37173–5
[131] Amiard-Triquet C, Amiard J C and Mouneyrac C 2015 Aquatic Ecotoxicology: Advancing Tools for Dealing with Emerging Risks (Waltham, MA: Academic)
[132] Feng X, Kirchner J W and Neal C 2004 J. Hydrol. 292 296–307
[133] Kirchner J W, Feng X and Neal C 2000 Nature 403 524–7
[134] Li L et al 2021 WIREs Water 8 e1495
[135] Patz J A and Olson S H 2006 Ann. Trop. Med. Parasitol. 100 535–49
[136] Milly P C, Betancourt J, Falkenmark M, Hirsch R M, Kundzewicz Z W, Lettenmaier D P and Stouffer R J 2008 Science 319 573–4
[137] Sivapalan M and Samuel J M 2009 Hydrol. Process. 23 1671–5
[138] Ren H, Hou Z J, Wigmosta M, Liu Y and Leung L R 2019 Water 11 1296
[139] Koutsoyiannis D and Montanari A 2015 Hydrol. Sci. J. 60 1174–83
[140] Serinaldi F and Kilsby C G 2015 Adv. Water Resour. 77 17–36
[141] Kwan M P 2021 Health Place 70 102609