

Combining Fractals and Multifractals to Model Geoscience Records

Carlos E. Puente, PhD

Department of Land, Air & Water Resources, University of California, Davis
(cepuente@ucdavis.edu)

Mahesh L. Maskey, MSc

Department of Land, Air & Water Resources, University of California, Davis
(mmaskey@ucdavis.edu)

Bellie Sivakumar, PhD

School of Civil and Environmental Engineering, The University of New South Wales, Sydney,
Australia

(s.bellie@unsw.edu.au)

and

Department of Land, Air & Water Resources, University of California, Davis
(sbellie@ucdavis.edu)

In Fractals: Concepts and Applications in Geosciences, 2017,
edited by Behzad Ghanbarian and Allen Hunt, CRC Press.

1. Introduction

Understanding hydrological systems is crucial not only for the proper management of water resources but also for elucidating possible climate change impacts. There is not just a single variable playing a vital role in this regard but rather several geoscience signatures, such as precipitation, temperature, streamflow, and others.

With the advancement of computational capabilities, various geophysical-hydrological models have been proposed that try to account for the physical processes behind available data and for the seemingly stochastic character of the records. Irrespective of how this is attempted, the overall objective of these efforts is to try to understand the nonlinear and, hence, complex dynamics involved. Beyond the simplified nature of the assumed notions, these models end up requiring meaningful geophysical parameters, which, as they are often hard to acquire, may lead to misinterpretations. Despite the fact that we are in a computational era, fully understanding and, hence, fully recognizing the intrinsic variability of signals is still challenging. This is the case as records contain intricate, “chaotic,” and, altogether, convoluted details.

Given that natural time series are typically erratic, noisy, and intermittent, it has become natural to model them using stochastic (fractal) theories, that also account for statistical self-similarity (e.g., Mandelbrot 1982, Feder 1988). Specifically, in order to model complex natural signals several efforts have been made implementing stochastic theories (e.g., Rodríguez-Iturbe 1986), stochastic-fractal theories (e.g., Gupta and Waymire 1990, Lovejoy and Schertzer 2013), and chaotic analysis (e.g., Sivakumar 2000, 2004). These kinds of efforts, however, result in representations that are quite difficult to condition. As such, they may miss the specific convoluted details present in a given natural set, such as the exact locations of major peaks.

Given the intrinsic limitation in assuming that what is seen is a realization of a stochastic process, a geometric approach, combining both fractal and multifractal notions, and known as the Fractal-Multifractal (FM) method was introduced (Puente 1996). Such a deterministic method models random-looking natural sets as fractal transformation of multifractal measures and results in a host of patterns that indeed share the same

geometric and statistical features of natural records (Puente 2004). In fact, the FM approach may be conditioned to not only preserve key statistical indicators (e.g. moments, autocorrelation function, power spectrum, multifractal spectrum) but also the inherent textures of data sets.

The present chapter focuses on the application of such a notion in the context of geosciences research, especially in hydrology. From its inception, the FM procedure has advanced from encoding to understanding to predicting the complexity of natural records. In this spirit, the following sub-sections summarize our efforts: (a) encoding of mildly and highly intermittent records, (b) simulating mildly and highly intermittent sets, (c) disaggregating or downscaling such geophysical records, and (d) classifying sets geometrically aiming at the prediction of future scenarios. In what follows, the geophysical sets are limited to rainfall, streamflow, and water temperature, but, as will become apparent, the FM approach may be used to model several other sets.

2. The Fractal-Multifractal Method

Puente (1996) combined fractal functions and multifractal measures in order to model geophysical records in a holistic manner. This section briefly reviews the so-called Fractal-Multifractal (FM) method and also some of its variants, which have been found suitable to encode natural sets.

2.1 Original Approach

A *fractal interpolating function* $f: x \rightarrow y$, passing through $N + 1$ ordered points on the plane $\{(x_n, y_n) | x_0 < x_1 \dots < x_N\}$, is defined as the unique fixed point of N affine maps (Barnsley 1988):

$$w_n \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a_n & 0 \\ c_n & d_n \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e_n \\ f_n \end{pmatrix} \quad n = 1, \dots, N, \quad (1)$$

so that the graph $G = \{(x, f(x) \mid x \in [0,1]\}$ satisfies $G = w_1(G) \cup w_2(G) \cup \dots \dots w_N(G)$. While the parameters d_n are vertical scalings, $|d_n| < 1$, the other parameters, $a_n, c_n, e_n,$ and $f_n,$ are evaluated from contractive initial conditions that guarantee the existence of function f , namely:

$$w_n \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} x_{n-1} \\ y_{n-1} \end{pmatrix}, \quad w_n \begin{pmatrix} x_N \\ y_N \end{pmatrix} = \begin{pmatrix} x_n \\ y_n \end{pmatrix}. \quad (2)$$

These equations give rise to simple linear systems of equations for the “other” parameters in terms of the vertical scalings and the coordinates by which the function passes. At the end, a convoluted “wire” function f is generated whose graph has a fractal dimension D between 1 and 2.

A fractal function may be produced via a point-wise sampling of the “attractor” G iterating successively the affine maps. This process, known as the “chaos game” (Barnsley 1988), starts at a point within G (say, an interpolating point) and progresses guided by arbitrary “coin tosses” that assign distinct usage proportions to the N maps. Allowing enough time for the iterations, the process not only paints a graph G point by point, but also induces a unique invariant measure over G , which typically contains intermittencies and possesses multifractal properties.

Seen over the coordinate x , such invariant measures define *deterministic multinomial multifractals* – with length scales given by the placements of the interpolating points in x and with the proportions for the “coin tosses” defining the intermittencies – (Mandelbrot 1989, Puente 1996). When seen over the coordinate y , the invariant measures over G define deterministic projections, which turn out to encompass some of the irregular shapes encountered in nature, such as rainfall, streamflow, temperature, and other sets.

Fig. 1 illustrates how a binomial multifractal measure dx may be naturally combined with a fractal interpolating function to define an interesting output dy . When the two maps

$$w_1 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0.80 & 0 \\ 2.91 & -0.72 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (3)$$

and

$$w_2 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0.20 & 0 \\ -0.65 & -0.54 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0.80 \\ 2.19 \end{pmatrix}, \quad (4)$$

are iterated according to proportions 68–32%, they: (a) generate a fractal wire f passing through $\{(0,0), (0.80, 2.19), (1,1)\}$ (at the center of circles) and having dimension $D = 1.31$; and (b) induce a simple binomial multifractal dx over x – having length scales 0.8 and 0.2 and redistributions 68–32% – and also a unique and, hence, deterministic derived measure dy over y – exhibiting non-trivial variability.

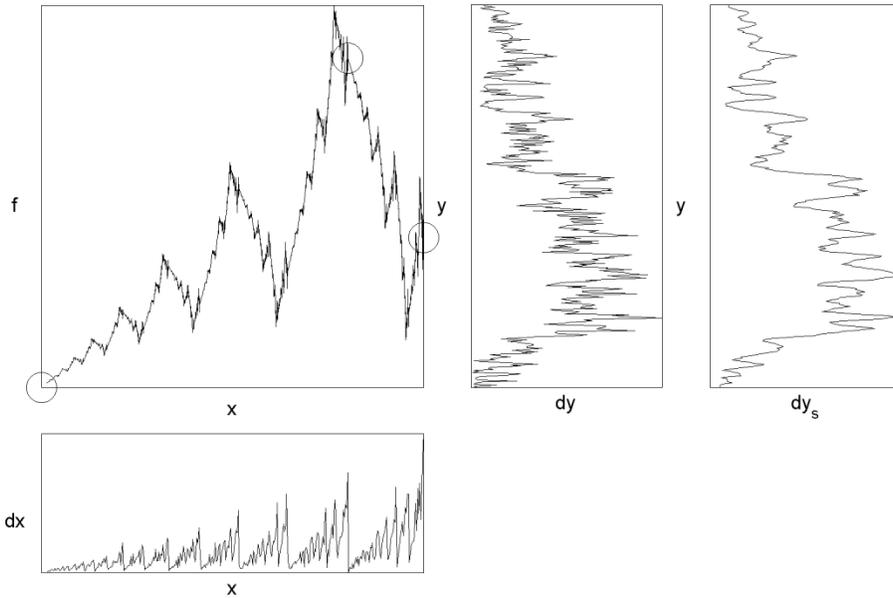


Figure 1: The FM approach: from a multifractal dx to a projection dy , via a fractal interpolating function f , a “wire” from x to y . dy_s is a smoothed version of dy .

In a practical setting, dy is found as a histogram of all chaos game points over y , adding all “events” over x corresponding to the crossings of function f for a given value of y i.e., $dy = f^{-1}(dx)$. By varying the parameters, and as shown on the far right graph in Fig. 1 via local integration, objects dy or dy_s turn out to be “random-looking” sets that do resemble geoscience records (Obregón et al. 2002b, Puente 2004).

The FM procedure, besides transforming multifractals relevant in the study of turbulence, i.e., the case when dx has equal length scales and redistributions 70–30% (Meneveau and Sreenivasan 1987), may be

assigned a more direct physical interpretation (Cortis et al. 2013). This is the case as the measure dy , although entirely deterministic, may be interpreted as a “realization” of a non-trivial but conservative multiplicative cascade, as customarily done via stochastic cascades of tracers (e.g., Lovejoy and Schertzer 2013).

In summary, the modeling of geoscience records via the original FM method requires the solution of an inverse problem for suitable geometric parameters, namely: (a) the interpolating points by which a fractal function passes, (b) the vertical scalings d_n , (c) the frequencies used on chaos game calculations, and (d) a smoothing parameter depending upon the nature of a target set. When using two or three maps, the FM approach requires 6 or 10 parameters, hence leading to sizable compressions.

2.2 Two Extensions

Instead of defining a fractal “wire,” the procedure may be modified to generate more general attractors. Such may be done using N maps as in Equation (1), but using new contractive initial conditions,

$$w_n \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} x_{2n} \\ y_{2n} \end{pmatrix}, \quad w_n \begin{pmatrix} x_{2N-1} \\ y_{2N-1} \end{pmatrix} = \begin{pmatrix} x_{2n+1} \\ y_{2n+1} \end{pmatrix}, \quad n = 1, \dots, N, \quad (5)$$

such that the range of map w_n in x becomes the more general interval $[x_{2n}, x_{2n+1}]$, for $x_0 \leq x_{2n} < x_{2n+1} \leq x_{2N-1}$. When such sub-intervals contain gaps, the resulting attractors are Cantorian in nature (Maskey et al. 2015) and when such ranges overlap or when the corresponding end-points in y do not match as in Equation (2), the resulting attractors are not functions but instead are shaped as interesting “leaves” (Huang et al. 2013).

Fig. 2 presents an example of an FM-based derived measure dy based on a Cantorian case, and Fig. 3 does so for a leafy attractor. While both representations use two maps, the former uses as end-points $\{(0,0), (0.39, -1.54)\}$ and $\{(0.77, -5.0), (1,1)\}$ and the latter $\{(0,0), (0.23, 5.0)\}$ and $\{(0.19, 0.01), (1,1)\}$, as marked by circles. Whereas the scaling parameters and iteration frequencies for the first case are 0.28 and –0.47 and 66–34%, for the second case they are 0.55 and –0.89 and 15–85%, respectively.

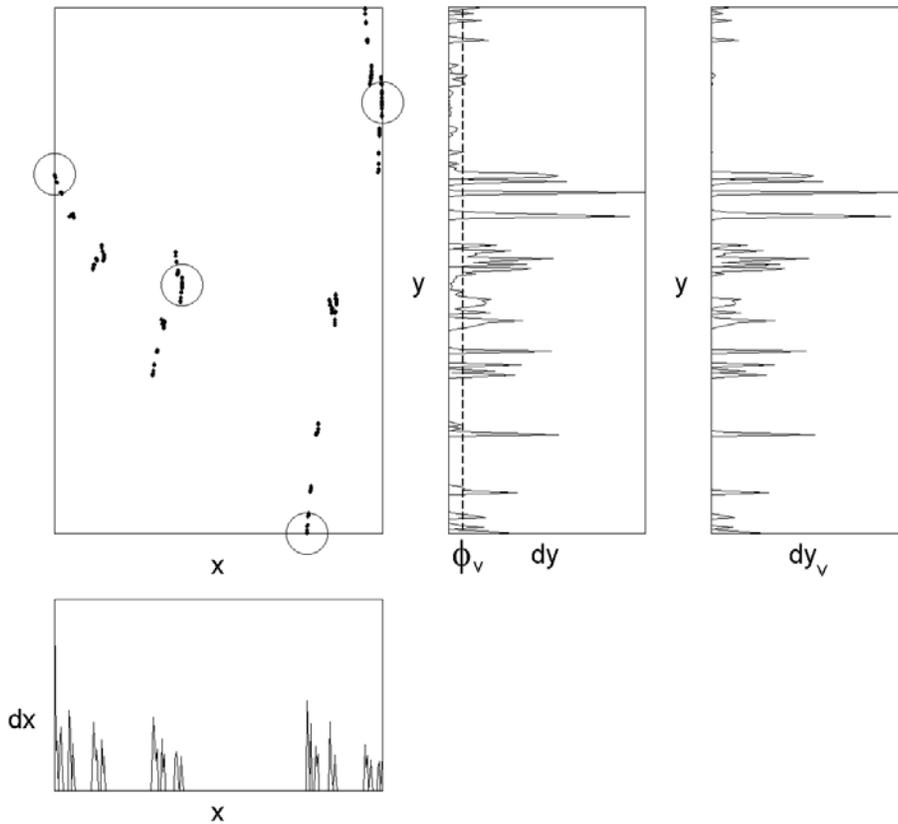


Figure 2: A generalization of the FM approach: from a Cantorian texture dx , to a derived dy , via a disperse attractor. dy_v is found pruning dy below a threshold ϕ_v and renormalizing.

As seen in Fig. 2, the iterations of such maps generate a Cantorian attractor, which, as before, induces projections dx and dy . While dx contains multiple spikes and gaps reflecting the original hole of size (as marked by the second and third open circles), the derived distribution dy also exhibits holes, but not in an obvious repetitive fashion as dx , hence becoming itself useful to model highly intermittent sets. In this spirit, the other entity in Fig. 2 (in the far right), named dy_v , represents yet an additional adaptation found by using a vertical threshold (ϕ_v) in the construction, so that (after proper normalization) new sets having an increased number of zeros may be found.

As seen in Fig. 3, the iteration of the corresponding maps for the second case yields other interesting patterns dx and dy . Now the overlap, also seen by the second and third open circles, results in a leafy

attractor that induces yet another representation dy whose smoothed version (not shown) may resemble, say, the diurnal or yearly variations of air or water temperature.

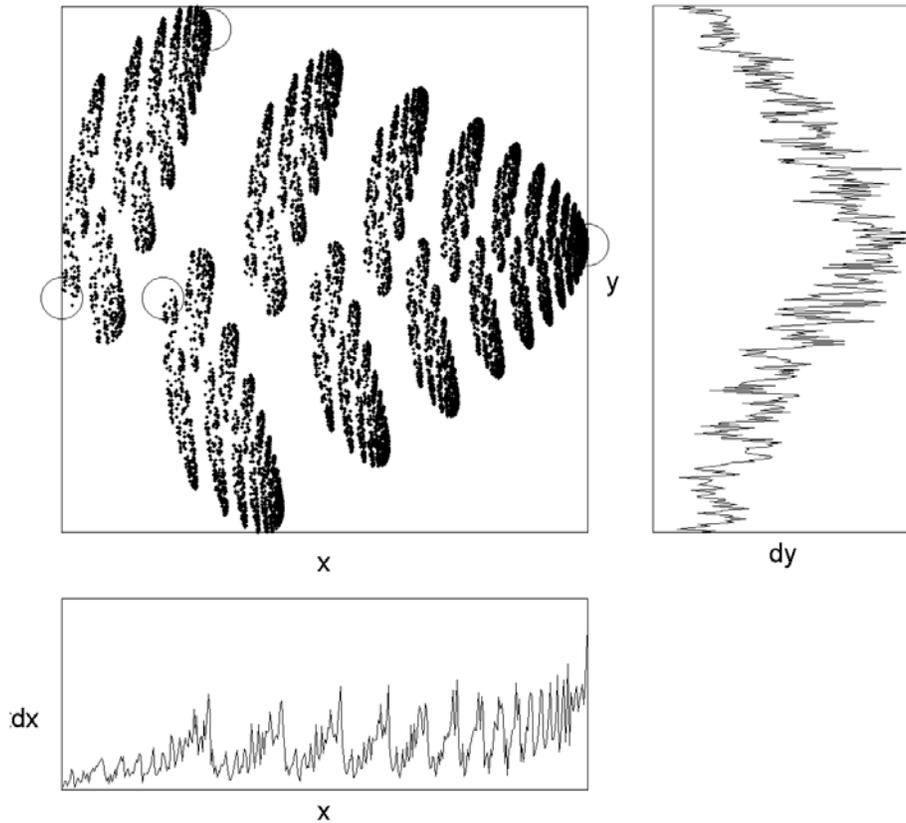


Figure 3: Another generalization of the FM approach: from an input texture dx , to a projection dy , via a “leafy” attractor.

For these extensions of the FM approach, the following are the key parameters needed for a suitable representation: (a) the end-points that define the more general attractor, (b) the vertical scalings d_n , (c) the iteration frequencies, and (d) a smoothing or a threshold parameter, if needed. When using two or three maps, these FM approaches require 8 or 15 geometric parameters, leading still to sizable compressions.

3. General Strategy

Even though the FM methodology is ultimately simple and computationally efficient – once a set of parameters is known – the finding of a suitable representation for a given set is not a trivial task. As there are

neither analytical formulas for attractors nor for the derived measures dy , only a numerical solution is possible. Unfortunately, such is hampered by the dimensionality of the problem, the choice of the objective function, and the optimization algorithm used (Huang et al. 2013), which reflect a highly complex parameter space where local minima may happen based on distinct initial conditions.

The results to be presented herein reflect our own experience throughout the years and are based on a generalized version of the particle swarm optimization (GPSO) algorithm and an objective function that minimizes squared errors of suitable attributes, either the accumulated data sets from beginning to end when encoding sets or statistical qualifiers for simulation purposes.

Inspired by the collective social behavior of animals and the notion that all members of the swarm ought to be leaders, a “cloud” version of the PSO algorithm (Fernández-Martínez et al. 2010) is adopted – with 300 initial swarm populations used for generating FM parameters after 100 successive iterations. As just mentioned, the L^2 norm (i.e., the root mean square error) of statistical attributes of the records is used to define an objective function. As an example, for encoding purposes such a function is

$$\epsilon_{ac} = \sqrt{\frac{1}{N} \sum_{i=0}^N (r_i - \hat{r}_i)^2} \quad (6)$$

where N is the number of data points; and r_i and \hat{r}_i are the i^{th} accumulated values of the original record and FM fit, respectively. In order to ensure that solutions share similar geometrical features with the target set, various penalties are also imposed on the objective function so that: (i) the maximum deviations on accumulated sets, ϵ_{mx} , at each point would not exceed 10%, and (ii) the length of the FM fit and the target set would not differ by more than 10%. At the end, results shown herein do satisfy such constraints.

The evaluation of the resulting FM models is made using various statistics not included in the objective function. For this purpose, various statistical attributes, such as autocorrelation, histogram, and Renyi entropy functions, are evaluated in terms of Nash-Sutcliffe efficiencies (Nash and Sutcliffe 1970).

4. Encoding Geophysical Records

The present section illustrates the applicability of the FM approach for encoding complex geoscience records, which may be categorized into two kinds: (1) mildly intermittent sets, such as rainfall events and typical streamflow and water temperature records gathered daily; and (2) highly intermittent sets, exhibiting considerable swings of activity and inactivity and a large numbers of zeros, such as daily rainfall records gathered over a year. These kinds of sets are now considered step by step.

4.1 Mildly Intermittent Sets

4.1.1 *Rainfall Events*

Faithful FM encodings of high-resolution rainfall events in Boston and Iowa City, lasting for a few hours, have been presented in Puente and Obregón (1996), Obregón et al. (2002a, b), Cortis et al. (2009), and Huang et al. (2012a, b, 2013). Such studies revealed that the FM method, coupled with a suitable search procedure, is capable of preserving not only the overall shapes of the records but also their main statistical and multifractal properties. In regards to the Boston event, measured every 15 seconds, various FM representations based on three to five affine maps yield, for the accumulated objective function, ϵ_{ar} in Equation (6), values smaller than 0.4% and maximum deviations in accumulated sets, ϵ_{max} , that are less than a mere 1.4%. For four events gathered in Iowa City, every five seconds, results are comparable both geometrically and statistically, with ϵ_{max} values that are consistently below 2.5%, for alternative variants of the FM method.

4.1.2 *Daily Streamflow*

Having obtained excellent encodings of rainfall events, the FM approach and its extensions are tested as a modeling method for streamflow records gathered daily over a year. As these mildly-intermittent sets visually share similar “complexity” as the rainfall events, the FM approaches turn out to yield also faithful descriptions. This is illustrated next for 64 years of daily records gathered at the Sacramento River (USGS

station 11447650 near Freeport) and for even smoother, decadal records for a total of 55 years, where a constant base flow is subtracted from the raw data and a water year (labeled by the end year) starts on October 1st and ends on the following September 30th. Prior to encoding, such records are normalized so that the accumulated volume becomes unity, as required for the FM methodology, which is used as in Fig.1, with a constant smoothing of 5 days.

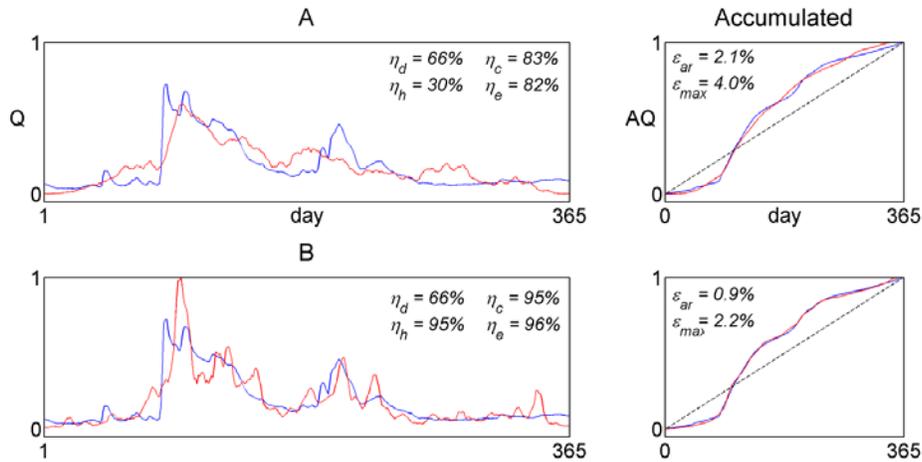


Figure 4: Measured streamflow at the Sacramento River for water year 1965 (blue) and two FM representations (red), corresponding to wires based on two and three maps.

Fig. 4 shows an example of two FM encodings (in red) of the streamflow set corresponding to the water year ending in 1965 (blue). While representation A corresponds to a wire generated via two maps, B emanates from a wire generated via three maps. As seen, the overall geometry of the streamflow set is well captured in both cases, as the corresponding accumulated sets are hard to distinguish from the actual set. Such is corroborated by the small values of the root mean square and maximum errors in accumulated sets (ϵ_{ar} and ϵ_{max}) – below 2.1 and 4.0%, respectively – and the high Nash-Sutcliffe efficiency for the records, η_d , always above 66%, as reported in the figure. As may be expected, given its additional degrees of freedom (10 parameters), the encodings for representations B results in reasonable fittings of the autocorrelation, histogram and entropy functions, as implied by high Nash-Sutcliffe efficiencies (η_c , η_h and η_e) – equal to 95, 95 and 96%, respectively, also as included in the figure.

After encoding all 64 streamflow years, from 1951 to 2014, Fig. 5 shows the implied performance for case B (a fractal wire based on the iteration of three maps), when FM encodings are upgraded by the yearly volumes and then the constant base flow is added. As seen, the overall FM representation of streamflow is excellent visually (in red) and also statistically, as the qualifiers ϵ_{ar} , ϵ_{max} , and η_d have means plus or minus standard deviations of $0.8 \pm 0.3\%$, $1.8 \pm 0.5\%$ and $64 \pm 19\%$, respectively.

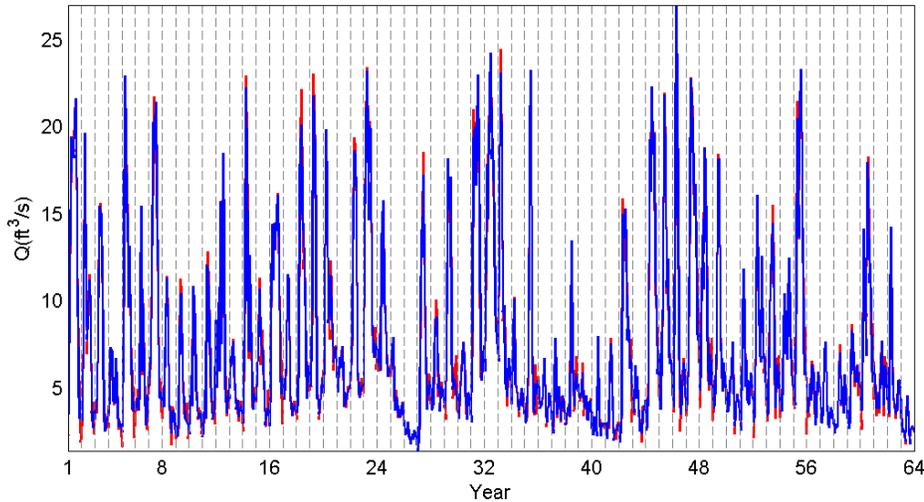


Figure 5: Observed (blue) and FM encoded (red) streamflow for 64 years in the Sacramento River, from water year 1951 till 2014. The FM method uses wires based on three maps. The vertical scale of the graph is in 100,000 ft^3/sec .

As climate change studies often rely on information averaged over a decade, Fig. 6 includes the FM analysis of a typical decadal streamflow set at the Sacramento River (blue) for two variants: a wire model based on two maps (A) and a wire representation that uses three maps (B) (both in red). Notice how the smooth decadal records are also well represented by the FM notions, as expected. The optimization exercise itself turns out to be easier this time than that for yearly sets (in part because parameters for the previous decade may be used as initial conditions for a future search) and such a fact gets reflected in a higher range of statistical qualifiers. As included in the figure, for the decade presented, the ϵ_{ar} and ϵ_{max} values are found to be 0.5 and 1.1% for FM representation A (relying on 6 parameters) and 0.3 and 0.8% for FM variant B (employing 10 parameters), which are rather small numbers, smaller than the typical ones reported in Fig. 4.

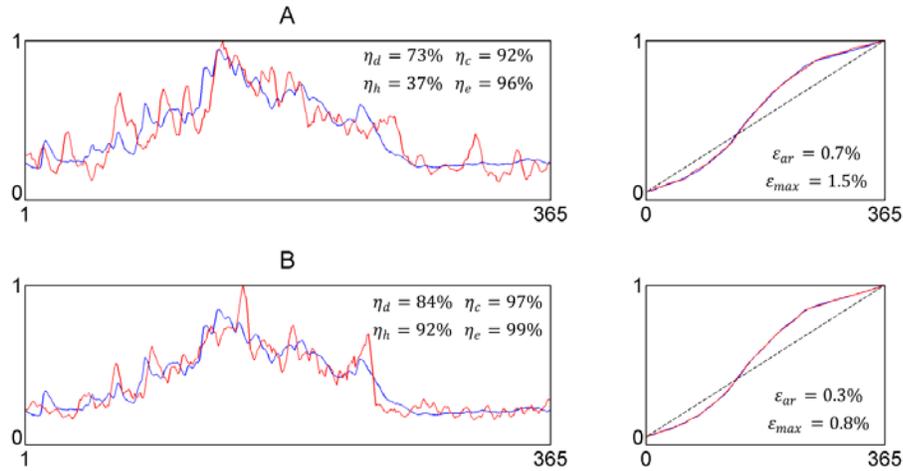


Figure 6: Measured streamflow at the Sacramento River for water decade 1966 (blue) and two FM representations (red), corresponding to wires based on two and three maps.

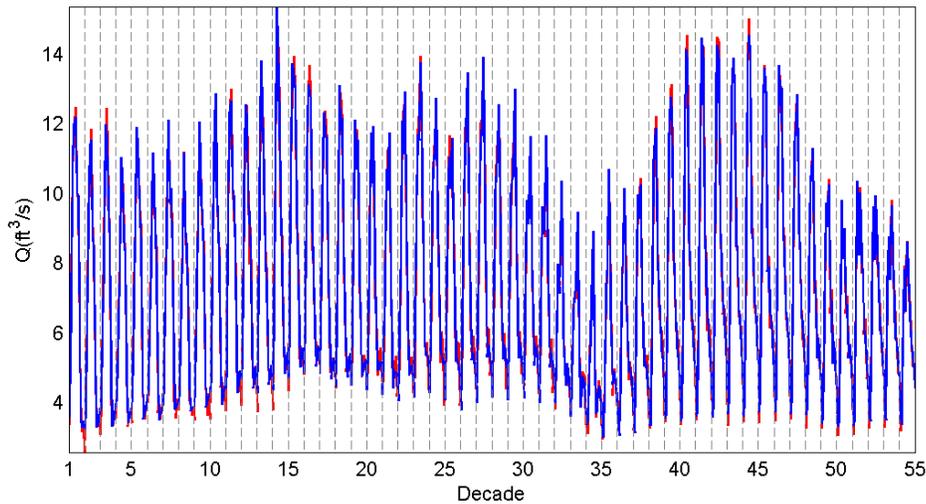


Figure 7: Observed (blue) and FM encoded (red) streamflow for 55 decades in the Sacramento River, from water decade 1959 till 2014. The FM method uses wires based on three maps. The vertical scale of the graph is in $100,000 \text{ ft}^3/\text{sec}$.

As done before, the overall performance over the entire period of 55 streamflow decades is shown in Fig. 7, for the wire model based on three maps. Clearly, the agreement between data (blue) and FM sets (red) is again excellent, and such is reflected by small numbers for the attributes ϵ_{ar} , ϵ_{max} (0.5 ± 0.1 , 1.0 ± 0.3 , respectively), and noteworthy values of η_d (76 ± 12). As seen contrasting Figs. 5 and 7, the fits are better for the smoother sets.

As seen in Fig. 8, the FM representations in Figs. 5 and 7 (red) result in very close preservations of the Spring flows at the Sacramento River (blue), for the months of March, April and May (bottom to top). This fact is particularly useful as the FM encodings may be used to properly represent the main volumes in the river.

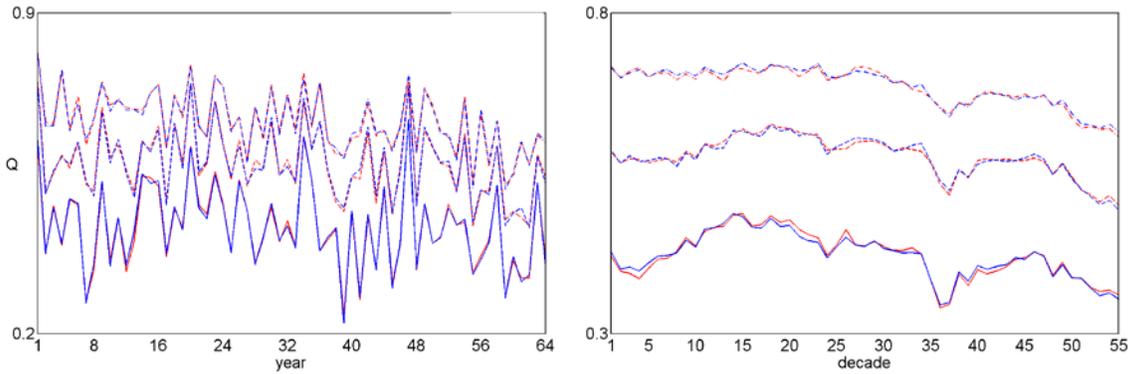


Figure 8: Yearly (left) and decadal (right) evolution of spring flow in the Sacramento River, corresponding to Figures 5 and 7. FM sets are in red and observations in red.

4.1.3 Daily River Water Temperature

Having faithful FM representations of streamflow sets motivates encoding water temperature at the Sacramento River. Since the low temperature over the year happens around December and January, instead of the water year cycle, records from January to December are considered, taking out a minimum temperature and normalizing, as done for streamflow. Once the records are defined, the FM approach with a 7-day internal smoothing is used for 51 years spanning 1962 to 2012.

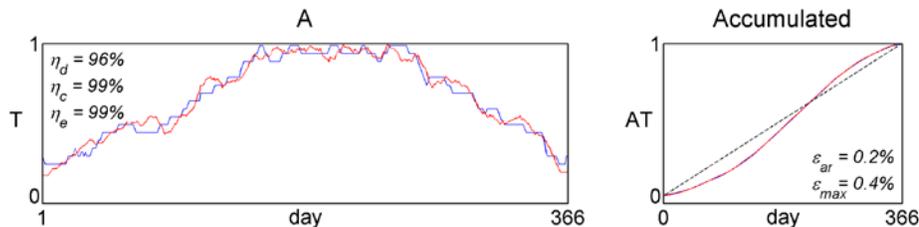


Figure 9: Measured water temperature in the Sacramento River for the year 1968 (blue) and an FM representation (red), corresponding to a wire based on two maps.

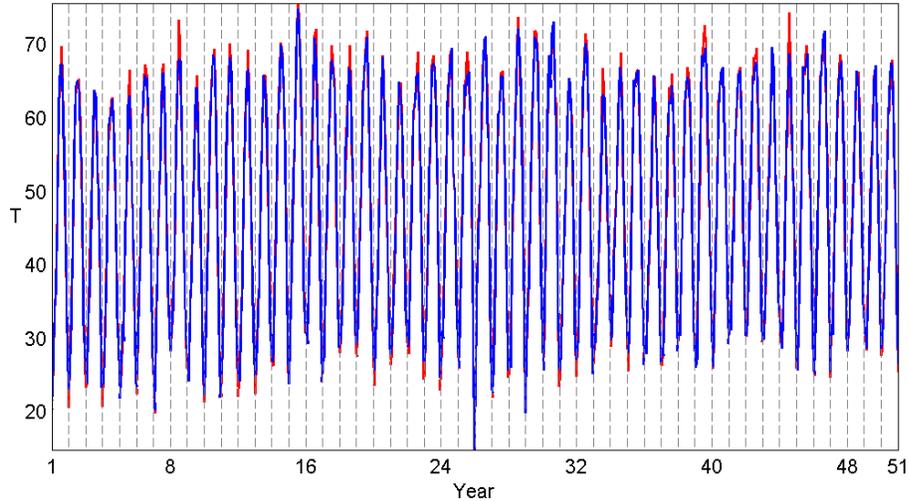


Figure 10: Observed (blue) and FM encoded (red) water temperature records for 51 calendar years in the Sacramento River, from 1962 till 2012. The FM method uses wires based on two maps.

A sample outcome of such an effort is shown in Fig. 9 for the year 1968, which includes close fittings of the accumulated temperature sets based on a wire, defined using two maps (red). As seen, this set has excellent ϵ_{ar} and ϵ_{max} values of 0.2 and 0.4% and close to one Nash-Sutcliffe values η_d , η_c , and η_e . Fig. 10 presents the corresponding overall temperature records as encoded via the wire representation based on a total of 6 parameters. As seen, the FM representations (red) are truly excellent and such is reflected by rather high Nash-Sutcliffe efficiency for the records of $91 \pm 4\%$. Compared to streamflow sets, temperature records are smoother and, at the end, easier to represent using the FM method. As illustrated, this is accomplished with less number of parameters.

4.2 Highly Intermittent Records

The deterministic FM method, in its Cantorian version coupled with a vertical threshold (Fig. 2), may also be used to encode daily rainfall records exhibiting noticeable intermittency. This is illustrated next using 20 years of records gathered at Laikakota, Bolivia.

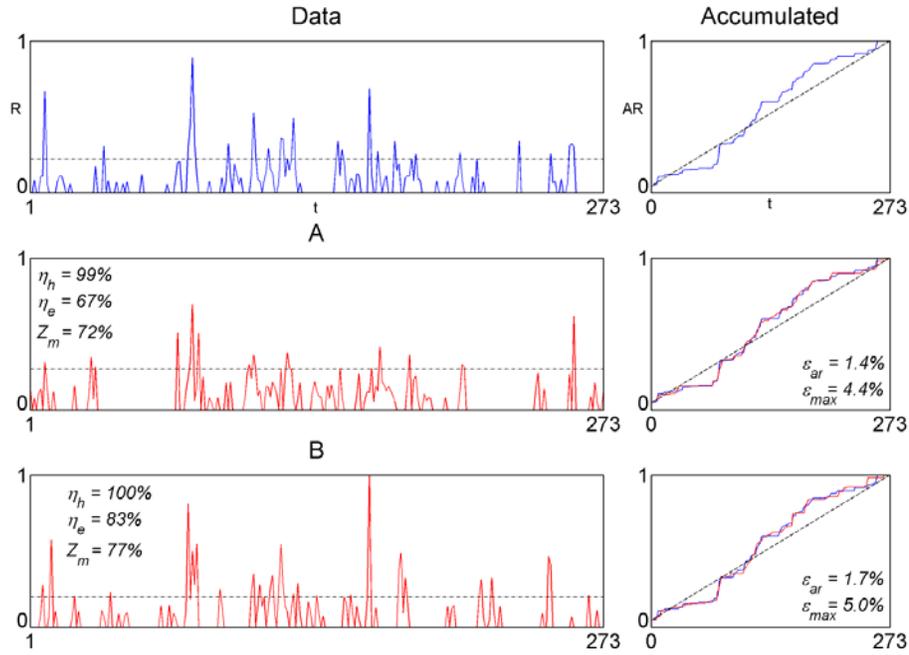


Figure 11: Measured rainfall at Laikakota, Bolivia from September 1965 to May 1966 (top-blue) and two FM representations (bottom-red), corresponding to Cantorian constructions based on three maps and a threshold.

As an example, Fig. 11 in blue shows rainfall records for water year 1966 (lasting nine months from September 1st of the previous year) and the corresponding accumulated mass set (top), and two alternative FM encodings (and accumulated sets) in red, labeled A and B, found via Cantorian inputs and subsequent pruning via a threshold; see Maskey et al. (2015) for further details. These representations rely on the iteration of three maps and approximate the accumulated set closely as they have mean square and maximum errors in ϵ_{ar} and ϵ_{max} that are (as seen in the graph) less than 1.7 and 5%, respectively. Although the precise locations of major peaks on this complex data set are not perfectly captured, the FM encodings do preserve the volume of the most massive second peak, which may be seen in the corresponding steepest section of the accumulated profile. As seen, the FM sets not only share similar textures and overall distribution of rain throughout the season, but also closely capture the extremes in the set, as indicated by the shown horizontal dashed lines at 90% of the mass, especially the one named B. The goodness of the encodings may also be corroborated by the rather high Nash-Sutcliffe efficiencies on the histogram and entropy of the data set: η_h and η_e always above 99 and 67%, and by the percent of zeros matched by the encodings z_m above 72%, as reported in the graph.

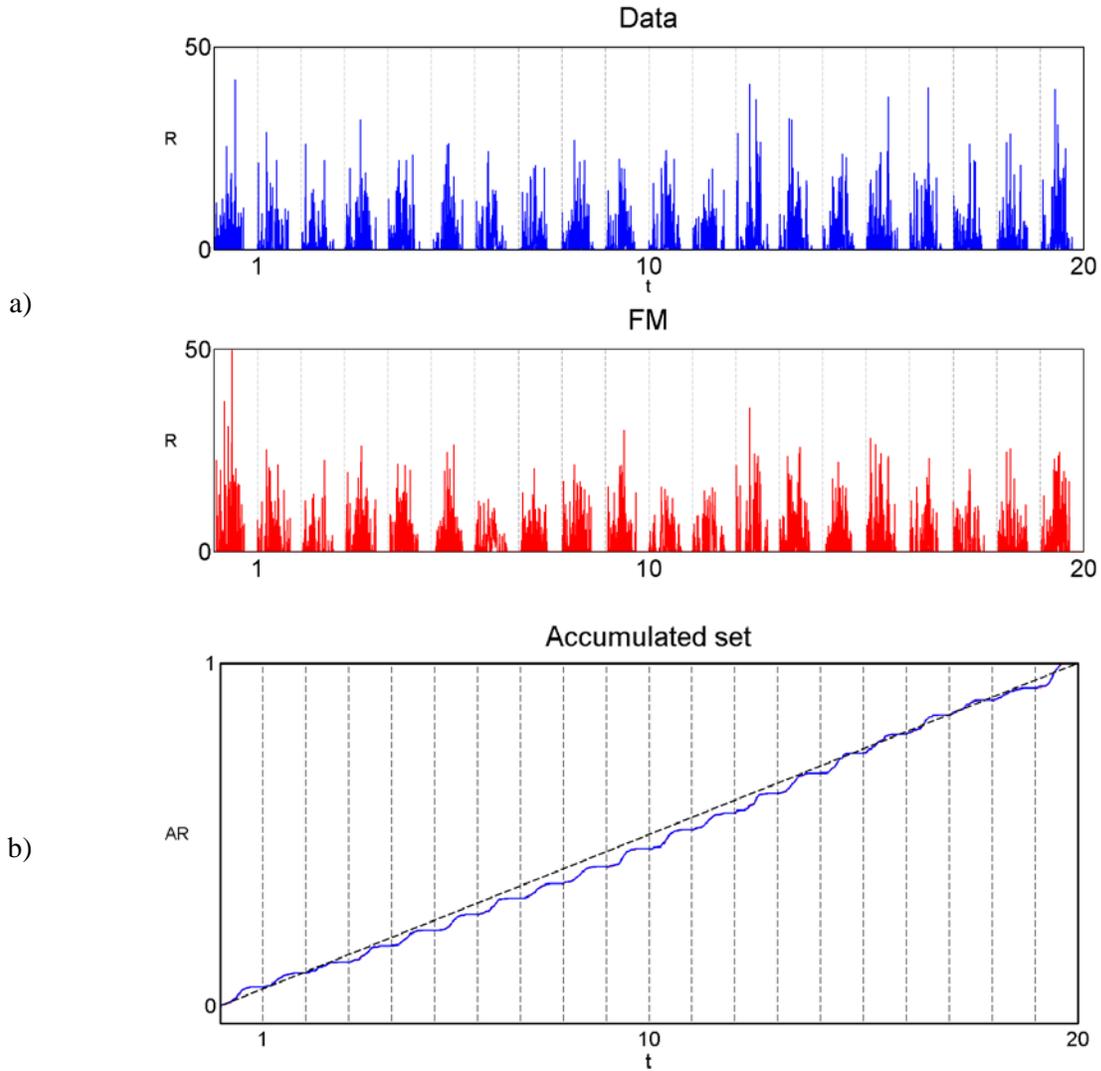


Figure 12: a) Measured rainfall at Laikakota, Bolivia from 1964-1965 to 1983-1984 (top-blue) and a FM representation (bottom-red), corresponding to Cantorian constructions based on two maps and a threshold; and b) corresponding accumulated rainfall sets over the period (After Maskey et al., 2015).

Figure 12 shows the results of applying the FM approach (red) on each of 20 years of data at the Laikakota site (blue, from 1963 to 1984), but using a representation that is best in maximum accumulated error ϵ_{max} , while using only two maps and a threshold. Although the obtained approximation is not as faithful as the one based on three maps, the overall behavior over the years remains good, as optimized accumulated values, ϵ_{ar} , range from 1.3 to 2.2% and maximum accumulated errors, ϵ_{max} , span 3.5 to 8.8%. Notice how the real set and the FM encoding share similar (although non-identical) characteristics which, by having the mass properly accommodated within each individual year, result in almost indistinguishable overall

accumulated rain over the 20 years, as seen in Fig. 12(b). It is quite remarkable that the FM approach, relying on only 9 parameters, may accomplish such a feat.

4.3 Remarks

The above results have clearly illustrated how the FM approach and its variants (depending on 6 to 15 parameters) may be coupled with a heuristic optimization scheme, especially the GPSO, in order to approximate the specific geometry of a host of geophysical patterns. As seen, the more the intermittency on a set the harder it is to encode it, with daily water temperature patterns being the easiest followed by daily streamflow records and daily rainfall sets, which are the most difficult. Certainly, the more complex-looking a given target is, the higher the number of maps required in the iterations and, hence, the higher the number of parameters, leading to more involved searches, which are compounded if a threshold and penalties, involving an inherent discontinuity, are required. Although some calculations may be accomplished on a standard personal computer, calculations for alternative sets require running the optimization procedure from scratch, as distinct sets have variable geometries.

As both rainfall and streamflow records are not measured perfectly (e.g., Lanza and Vuerich 2009, Hammel et al. 2006), the encodings shown, within 3% in accumulated records, represent sensible approximations of the natural phenomena. These results, providing geometries that match the overall shapes and textures present in the sets, clearly substantiate the notion that complex geophysical patterns may be wholly characterized in a deterministic way (Puente and Sivakumar 2007). As the information in the sets is now compressed into the FM parameters of subsequent sets, an assessment of such evolutions shall be presented later in Section 7.

5. Simulation of Geophysical Records

Having shown that the FM method may produce faithful encodings of geophysical records, this section shows that the deterministic parsimonious approach may also be used to obtain simulations having increased levels

of complexity, namely, rainfall events, daily streamflow sets, and daily rainfall sets. This is accomplished by replacing the objective function, so that instead of accumulated sets it now uses suitable statistical information, such as autocorrelations, histograms, entropies, and numbers of zero values.

5.1 Mildly Intermittent Sets

To illustrate the notions with mildly intermittent sets, one rainfall event in Iowa City and one year of daily streamflow records gathered at the Sacramento River, California are analyzed. In both cases, alternative plausible simulations are found by minimizing the root mean square errors for correlations ϵ_c and histograms ϵ_h , employing two maps that generate either a wire or a leaf (as in Figs. 1 and 3). While the simulated streamflow records rely on a smoothing of the obtained FM set, the rainfall set does not require of a local average of the output derived projection. As a consequence, the numbers of FM parameters for the wire and leaf representations of rainfall are 5 and 7, while for streamflow they are 6 and 8.

5.1.1 Rainfall Events

Figure 13 displays a high-resolution storm event gathered in Iowa City in 1990 and lasting 11.4 hours (Georgakakos et al. 1994) together with the records' autocorrelation function and histogram (top-blue), followed by two FM simulations that preserve either the histogram or the autocorrelation function of the records (bottom-red). As seen, while the original storm contains two prominent regions of activity, the simulations yield sensible sets having similar-looking geometries that spread the mass and peaks at different locations.

As appreciated via the statistics in the graph, both representations (A from a wire fitting the histogram only and B via a leaf seeking the autocorrelation only) preserve very well both the autocorrelation and histogram. Nash-Sutcliffe values for the histogram η_h are rather close to 100% and the location of the extreme 90% values μ_{90} match very well, as seen by the close horizontal and vertical dashed lines in the

graphs. Although Nash-Sutcliffe statistics for the autocorrelation for case A is not as high as in case B, 67% vs. 96% in η_c , notice the close agreement on the decay of the function for case A not optimizing such an attribute, as indicated by the number of lags when zero and 1/e correlations are found, τ_0 and $\tau_{1/e}$.

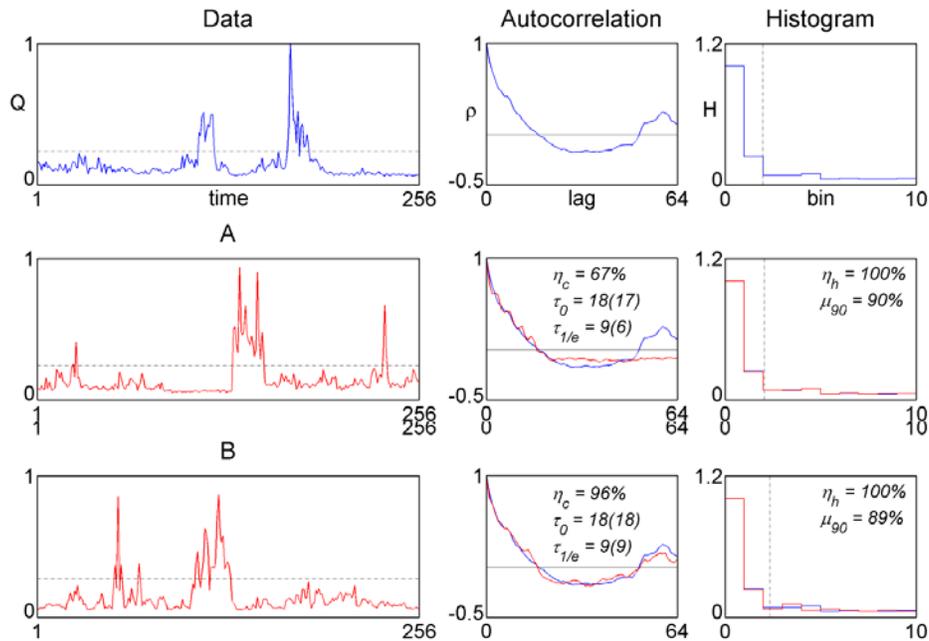


Figure 13: A rainfall event in Iowa City (top) and two suitable FM simulations (bottom), based on a wire and a leaf representation using two maps each. While A is based on fitting the histogram, B optimizes the autocorrelation.

5.1.2 Streamflow

Fig. 14 shows a year of daily streamflow record gathered at the aforementioned station of the Sacramento River, in water year 2003, together with the record's autocorrelation function and histogram (top-blue), followed by two FM simulations that, once again, preserve either the histogram or the autocorrelation function of the record (bottom-red). As seen, while the original record and the simulations do look alike geometrically, the latter spread the mass in different ways, hence yielding peaks at various locations.

As done for the rainfall events just described, while simulation A uses a wire based on two maps in order to preserve the record's histogram, the one labeled B emanates from a leaf based on two maps that seeks to fit the data's autocorrelation function. Unlike what was done for rainfall, though, a smoothing parameter of

7 days is used to generate the simulated sets. As may be seen, the shown simulations preserve almost perfectly the statistic included in the objective function: while simulation A matches the histogram, simulation B very closely preserves the autocorrelation function. Such may be corroborated by the corresponding Nash-Sutcliffe statistics close to 100%, rather close values of μ_{90} close to 90%, and perfect fittings on the decay of the autocorrelation function via the lags τ_0 and $\tau_{e^{-1}}$. Although preserving one attribute does not imply close fittings of the other statistic, notice how the shown simulations do reasonably well on a statistic not included in the objective function, with η_c and η_h values being substantially large at values of 76 and 83%.

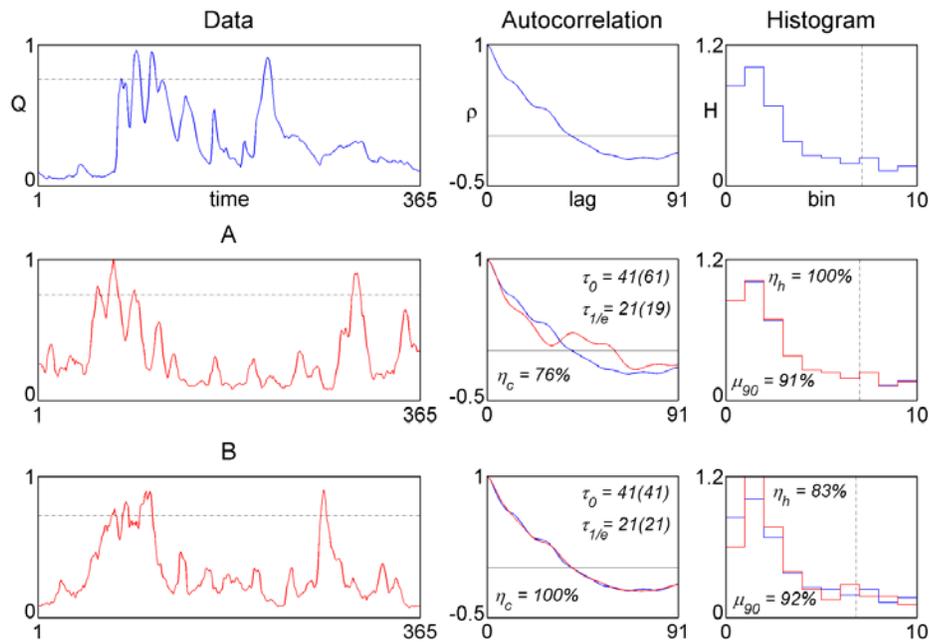


Figure 14: A streamflow set at the Sacramento River from October 2002 to September 2003 (top) and two suitable FM simulations (bottom), based on a wire and a leaf representation using two maps each. While A is based on fitting the histogram, B optimizes the autocorrelation.

5.2 Highly Intermittent Records

To illustrate that the FM approach may also be used to simulate highly intermittent sets, a year of daily rainfall gathered at Tinkham Creek, Washington State, USA is analyzed. As autocorrelations of rain decay towards zero rather quickly, alternative plausible simulations are found by minimizing the root mean square errors for entropy ϵ_e and histograms ϵ_h , employing two maps that generate either a wire or a Cantorian

attractor, both coupled with a threshold (as in Figs. 1 and 2). As such, the numbers of FM parameters for the wire and Cantorian simulations that follow are just 6 and 8.

Figure 15 displays rainfall record at Tinkham Creek for water year 2001 together with the record's histogram, Renyi entropy, and autocorrelation function (top-blue), followed by two FM simulations that preserve either the entropy function or the histogram of the record (bottom-red).

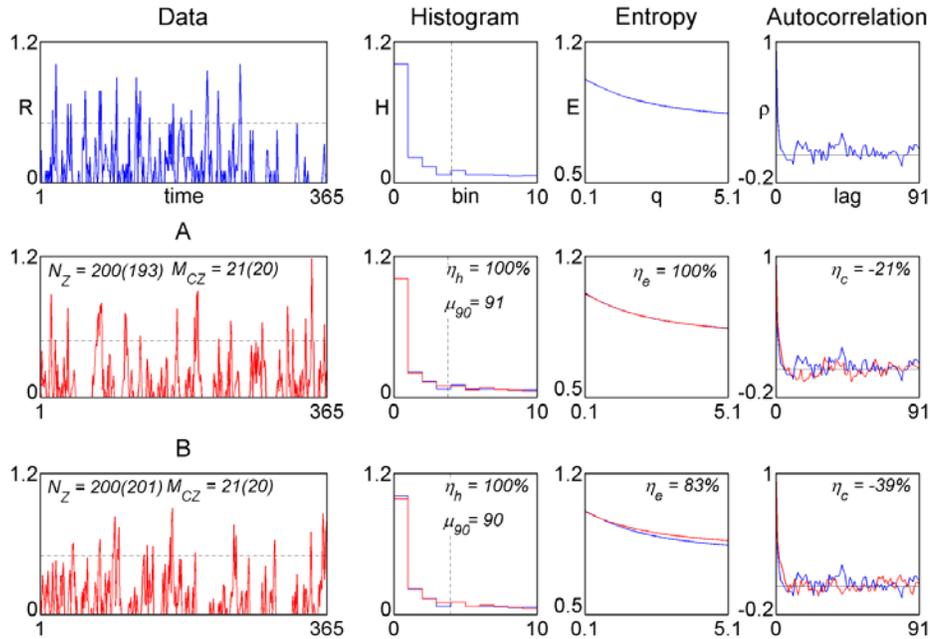


Figure 15: A rainfall set at Tinkham Creek, Washington from October 2000 to September 2001 (top) and two suitable FM simulations (bottom), based on a wire and a Cantorian representation using two maps each. While A is based on fitting the entropy, B optimizes the histogram. The two simulations also account for consecutive zeroes in the records.

As seen, the original set and the two simulations share similar looking intermittent geometries that, as found for mildly intermittent sets, spread the mass throughout the year giving rise to distinct peaks. While simulation A uses a wire based on two maps and uses a threshold in order to preserve the record's entropy, the one labeled B comes from a Cantorian attractor generated via two maps and a subsequent threshold that seeks to fit the histogram of the data. As seen, the simulations preserve almost perfectly the statistic included in the objective function: while simulation A matches the entropy (and also the histogram, but not optimized), simulation B very closely preserves the histogram (and does not do poorly on the entropy as implied by an η_e

value of 83%). This performance, no doubt reflected by the addition of penalties into the objective function so that the number of zeros on the records and FM simulation differ at most by 10%, explains the intuitively correct feeling on the simulations. For as seen on the graphs, the total number of zeros N_Z , the maximum consecutive number of zeros M_{CZ} , and the extreme μ_{90} values are all almost perfectly preserved, even if negative Nash-Sutcliffe values for properly decaying but statistically insignificant autocorrelations η_c appear.

5.3 Remarks

The above results have shown that the FM approach (and its variants) may be coupled with a heuristic optimization scheme in order to simulate a host of geophysical patterns. This is achieved encoding suitable statistics of the records rather than the records themselves, such as autocorrelation, histogram, entropy, and others. The examples herein present just a couple of possibilities, but additional sets may also be obtained by having representations based on more than two maps, by selecting various FM parameter combinations in the vicinity of an optimum, and by combining two or more attributes on an alternative objective function.

It has been our experience that finding suitable simulations may be achieved easily and a fraction of the time required for doing encodings. As the employed statistics are much less complex than the sets, the simulations may be found (for both mildly and highly intermittent cases) iterating just two maps, hence yielding parsimonious FM representations. As it was illustrated, the simulated sets do resemble geometrically the processes under study, including their overall texture and intrinsic complexity. As such, the FM simulations may be useful to analyze the redistribution of patterns throughout the event duration and to provide alternative scenarios from which to study the intrinsic variability of a phenomenon. Certainly, the FM patterns may be used to supplement other plausible simulations found via alternative (stochastic) methods.

6. Downscaling Geophysical Sets

This section explains how the deterministic geometric FM approach may also be used to downscale geophysical sets, in particular streamflow and rainfall sets from coarse scales (e.g., weekly, monthly) to fine scales (e.g., daily). While various stochastic disaggregation techniques exist (e.g., Valencia and Schaake 1973, Koutsoyiannis 1992, 1994, Olsson 1998), they require simplified assumptions that often prevent them from adequately accounting for nonlinearities present in the records. As such, the FM method may provide an alternative approach to complement such stochastic techniques.

Given information at a coarse resolution, say accumulated every τ days, the FM method may be used to find daily values, as follows. First, accumulate the coarse records during the duration of the season in question, say a year, and then seek a suitable FM encoding of such accumulated records (as explained in Section 3). Second, given the FM parameters of such an approximation, compute the disaggregation as the output projection dy obtained at the daily resolution, just by re-computing the output histogram over the appropriate number of (increasing) bins. As the accumulated sets for fine and coarse records surely match every τ days, a close FM approximation at the coarse scale should result in a good representation at the daily scale, one that, by definition, should remain close to the overall accumulated mass. In the present study, these notions are tested for streamflow and rainfall sets using coarse scales for $\tau = 7, 14$ and 30 days.

6.1 Mildly Intermittent Records

Figure 16 shows the streamflow records at the Sacramento River for water year 2005 (minus baseflow), evaluated (from available daily values) at the aforementioned levels of aggregation (blue) and the subsequent FM fits and disaggregations found via wires based on three maps and a local smoothing of 5 days (red).

As seen on the left of the figure and as it may be expected given the previous results herein, all FM representations at the weekly, bi-weekly, and monthly resolutions produce rather faithful fits of sets and

accumulated masses, with relevant search errors ϵ_{ar} and ϵ_{max} that are always below a mere 0.7 and 3.9%, respectively, and with corresponding Nash-Sutcliffe values η_d that are above a healthy 84%. As appreciated on the right hand side of the graph, the FM variants at distinct resolutions yield reasonably-looking disaggregated patterns at the daily scale, which not only closely follow the accumulated records but also preserve, even when using 30-day data, the overall texture and three main peaks of the original set.

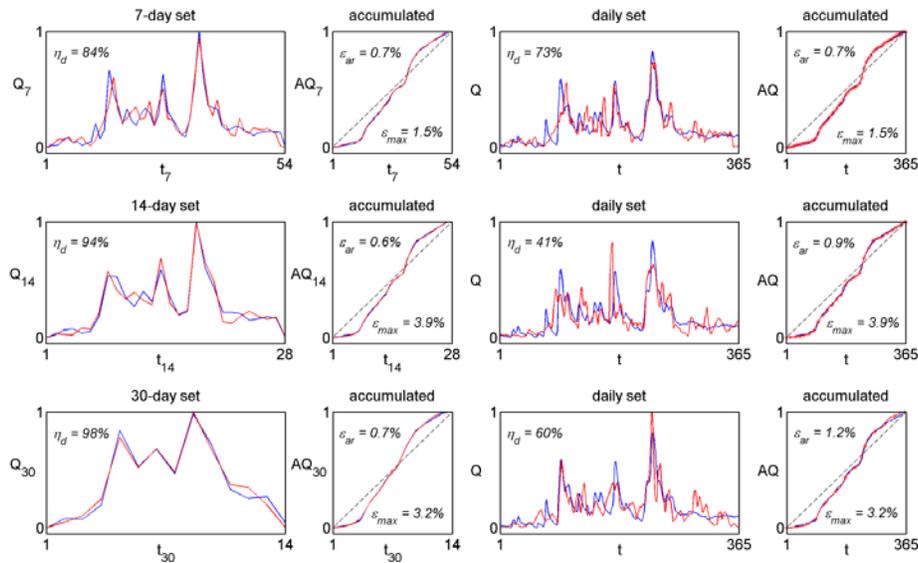


Figure 16: Sets and accumulated records associated with streamflow gathered at Sacramento River, during water year 2005. Observations (blue), FM fits at coarse resolutions (red-left); and corresponding downscalings at a daily scale (red-right). The aggregation scales are 7, 14 and 30 days. Measurements are shown in blue and FM-related information, via a wire based on three maps, are depicted in red.

As expected, the best disaggregated pattern, both statistically and by the naked eye, corresponds to the 7-day case. This is so, as the accumulated records at such a resolution are the closest to reality of the ones considered. Although in terms of Nash-Sutcliffe values η_d the downscaled data from 30 days are better than those from 15 days, notice that both representations are suitable indeed as they have rather small values in ϵ_{ar} and ϵ_{max} that are less than 1.2 and 3.9%, respectively. As seen, all downscales are quite reasonable and they also happen to provide sensible approximations of relevant statistical information, such as autocorrelation and histogram.

6.2 Highly Intermittent Records

Figure 17 depicts rainfall records at Laikakota Bolivia for water year 1966 evaluated (from available daily values) at 7, 14 and 30 days (blue) and the subsequent FM fits and rainfall disaggregations defined via wires based on three maps and a threshold (red).

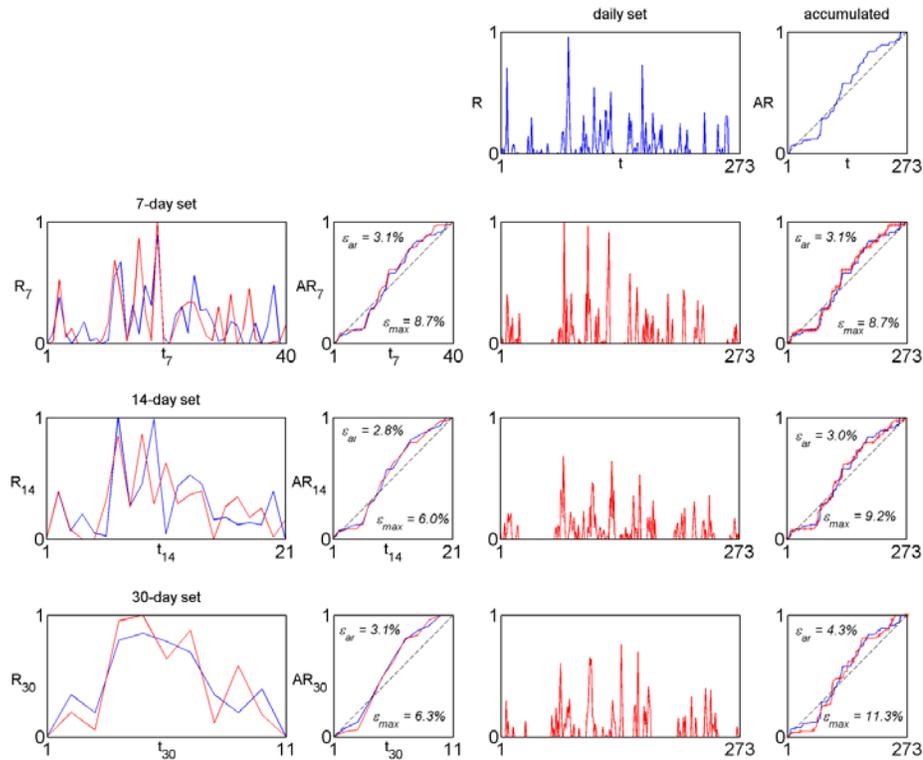


Figure 17: Sets and accumulated records associated with rainfall gathered at Laikakota, Bolivia, during water year 1966. Observations (top), FM fits at coarse resolutions (bottom-left); and corresponding downscalings at a daily scale (bottom-right). The aggregation scales are 7, 14 and 30 days. Measurements are shown in blue and FM-related information, via a wire based on three maps and a threshold, are depicted in red.

As observed on the left hand side of the figure and as it may be expected given the intrinsic complexity of rainfall sets, even though the accumulated records may be encoded with errors ϵ_{ar} and ϵ_{max} that are always below 3.1 and 8.7%, respectively, the corresponding encodings do not show the same degree of faithfulness as just described for streamflow. These lead, however, to the downscals shown on the right hand side, that while corresponding to increased errors ϵ_{ar} and ϵ_{max} below 4.3 and 11.3%, respectively, (about twice as much as what may be obtained while encoding a daily set, as in Fig. 11), do provide suitable shapes that may be useful in a practical setting. Certainly, the geometries of the implied downscals are

sensible and they may also be used as suitably close simulations of the records, that is, representations that follow closely the distribution of rainfall throughout the year.

6.3 Remarks

The above results have clearly shown that the FM approach may be used in order to downscale streamflow and rainfall sets to the daily scale, achieving reasonable results for scales that span up to 30 days. It has been illustrated that the downscaled sets resemble the texture and implicit complexity of the involved process, as they represent suitable geometries that pass closely to the accumulated daily records. The results have established that while runoff may be downscaled with noticeable precision, the disaggregation of more complex rainfall results in only some reasonable approximations that maintain the intermittency and overall distribution of rain during the year.

The application herein shows that only 10 FM parameters may be adequate in disaggregating weekly, bi-weekly and monthly sets, with compression ratios that are as high as 37:1 for a given year. The present downscaling technique may supplement others based on stochastic methods and may clearly be used to disaggregate outputs from global circulation models (GCMs) in order to assess climate change impacts.

7. Geometric Classification and Prediction of Geophysical Records

This section is concerned with the possibility of predicting geophysical records based on the time evolution of successive FM parameters of such annual sets. Specifically, the analysis herein centers on the streamflow sets at the Sacramento River encoded before (Section 4.1.2), which shall be attempted to be forecasted based on FM-based classifications of patterns, both at the yearly and at the decadal scales. Due to space considerations and as streamflow is a relevant indicator of catchment dynamics and of potential climate change effects (e.g., Döll and Schmied 2012), the study centers on such an attribute. A similar approach, however, may also be carried out on other geophysical variables (e.g., rainfall, water temperature, etc.).

7.1 Evolution of FM Parameters for Streamflow

As previously reported, 64 years and 55 decades of successive streamflow records at the Sacramento River have been faithfully encoded via the FM approach, i.e., Figs. 5 and 7. As all such representations emanate from FM wires based on the iteration of three simple maps, the corresponding FM parameters of such sets, together with the evolving volumes of years or decades, allow visualizing (in a compressed geometric fashion) the dynamics of streamflow at the two scales.

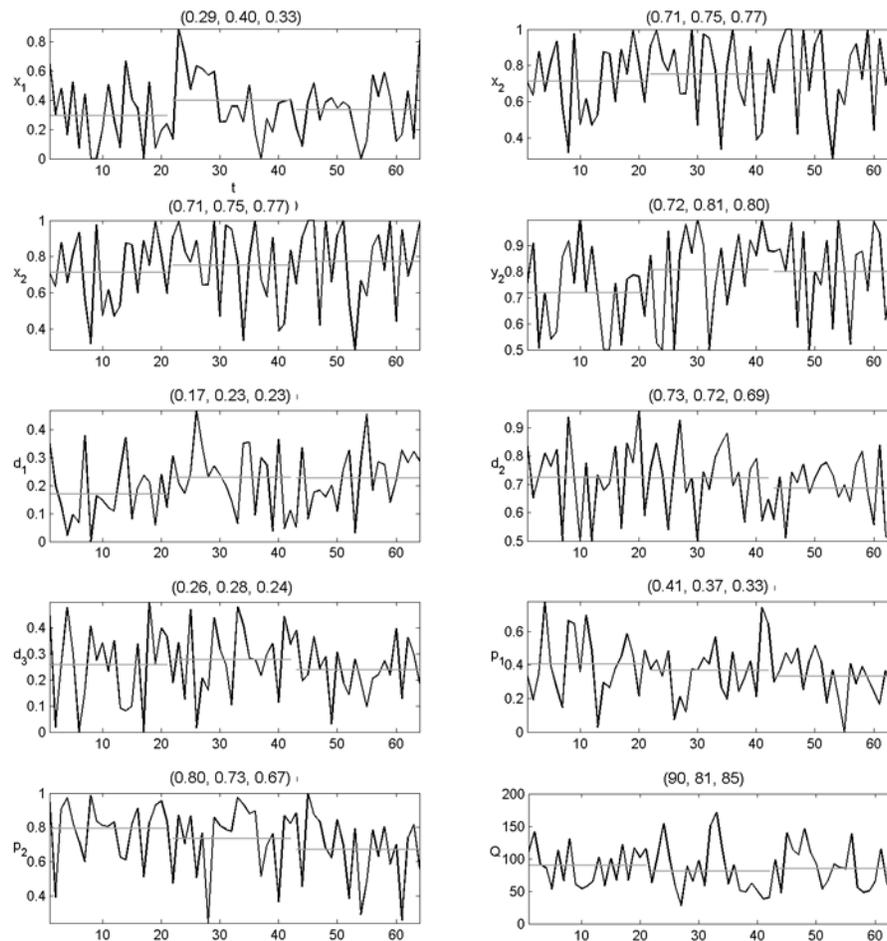


Figure 18: Time evolution of FM parameters for 64 years in the Sacramento River, corresponding to Figure 5. Gray horizontal lines represent mean trends over thirds of the time domain, with values shown above. Q is the total volume over a year in 100,000 cfs.

Figures 18 and 19 present the time evolution of FM parameters for yearly and decadal encodings, respectively, as follows. As all encodings use a local smoothing parameter of 5 days (as in Fig. 1), the two sets of graphs include a total of nine FM parameters, followed by the total volume Q of a year, namely: the coordinates of the second and third interpolating points (x_1, y_1) and (x_2, y_2) (having set, without a loss of generality, the first interpolating point to $(0,0)$), the vertical scaling parameters of the three maps, d_1, d_2 and d_3 , and the weights that determine how the maps are iterated, p_1, p_2 and $(1 - p_1 - p_2)\%$ of the time. As may be appreciated, the graphs also include local means (in gray) of each parameter over a third of the domain (shown on top of each parameter frame) in order to visualize possible trends in the parameters.

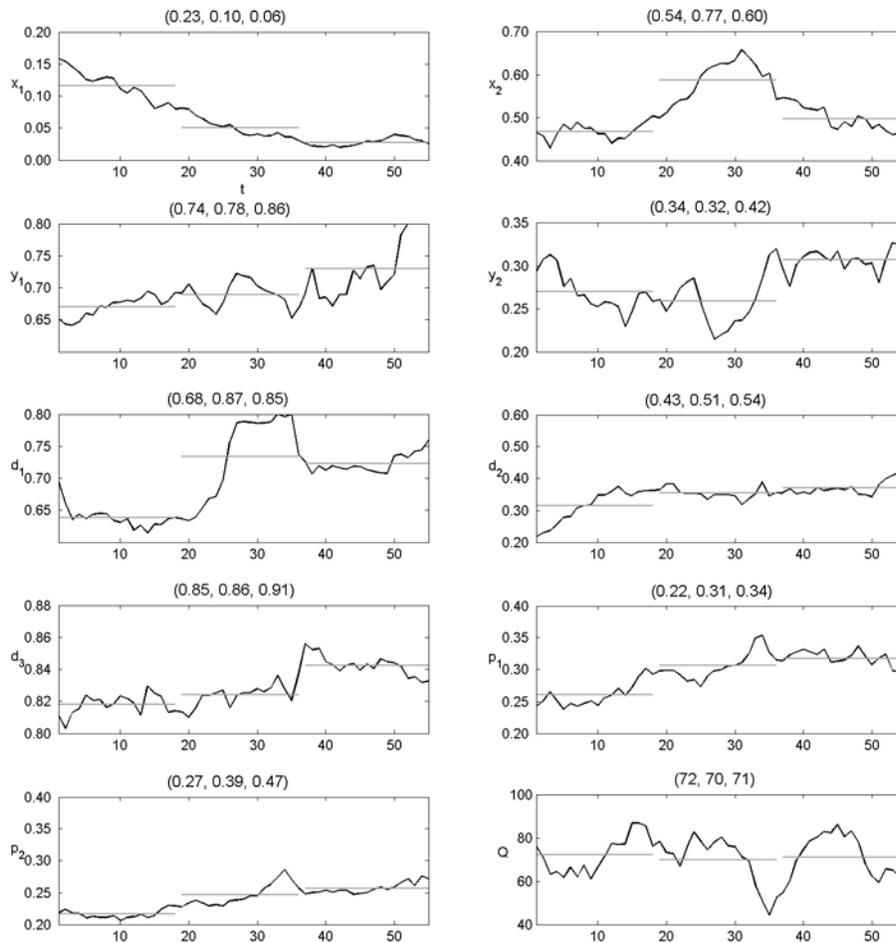


Figure 19: Time evolution of FM parameters for 55 decades in the Sacramento River, corresponding to Figure 7. Gray horizontal lines represent mean trends over thirds of the time domain, with values shown above. Q is the total volume over a decade in 100,000 cfs.

As seen comparing the graphs, it becomes obvious that there is much more variability at the yearly scale (Fig. 18) than at the decadal scale (Fig. 19), in a manner that resembles the different shapes already reported for the spring flows at those scales in Fig. 8. Clearly, yearly FM parameters (and total flow) vary wildly and swing from high to low values and vice-versa in a manner that precludes the possibility of safely extrapolating trends into the future. Although it is clear that these variable signals do play a key role in understanding the complexity of the evolving streamflow, the geometry of the observed patterns change in a non-trivial fashion from year to year and such leads to non-specific changes every 21 years (as shown in gray), which exhibit so much variability within that forecasts of individual parameters are not possible.

The evolution of FM decadal parameters and total volume in Fig. 19, on the other hand, do exhibit the expected smoothing produced by the fact that successive decadal patterns, besides being smoother, do look alike. The decadal evolutions clearly exhibit a degree of smoothness and some well-established trends when seen every 18 years (as shown by the gray horizontal lines) and such suggests that extrapolations of some FM parameters may be made into the future. Even if the flow Q remains quite variable at the decadal scale, the smaller variation in FM parameters could be used to foresee the overall geometry of decades into the future, which in turn may be translated into a yearly (disaggregated) normalized geometric prediction when decades and years share relevant correlations. Regarding plausible geometric implications of the decadal streamflow patterns to global climate change, it is interesting to note that some of the parameters in Fig. 19 do exhibit increasing and/or decreasing trends that may perhaps be related to climatic indicators.

Similar trends as those found in Figs. 18 and 19 are also found for other rivers within the United States (not reported here). While yearly FM parameter values, reflecting variable geometries, vary substantially from year to year, decadal FM parameters exhibit smoothness and some trends that allow extrapolating trends. It is anticipated that the information on such graphs may be used to study the inherent complexity of alternative locations (say by the strength of the swings in parameters) and to evaluate the presence of climatic effects at distinct regions.

7.2 Classification of Streamflow Sets via Clustering of FM Parameters

As the evolutions of FM parameters at yearly and decadal scales show ample variations (especially at the yearly scale), it becomes natural to inquire if a classification of patterns into classes may be made in order to arrive at more stable descriptions. As such, Fig. 20 shows the streamflow centroids for ten classes obtained via an unsupervised classifier, the k -means clustering analysis based on the Euclidean distance of FM parameters (e.g., Arthur and Sergi 2007). The shown graphs, which once again reflect the wider spread of yearly records, summarize the records only up to 1999 (and not all the records), and such are computed in an attempt to study if sensible streamflow predictions for the decade ending in year 2000 and, subsequently, for the water year 2000 may be obtained.

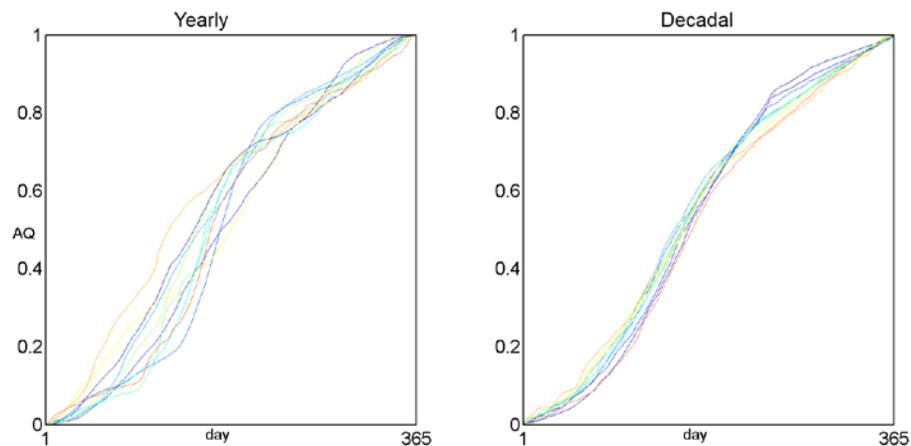


Figure 20: Yearly and decadal accumulated streamflow centroids based on a classification of FM parameters for data up to 1999 at the Sacramento River.

Having defined ten classes on yearly and decadal streamflow at the Sacramento River, Fig. 21 now shows their evolution from 1951, that is, 49 years (blue) and 40 decades (red). As expected, the classes at the yearly scale still exhibit noticeable variability, but the decadal information steadily grows in the selected classes, in a manner that suggests a stable class prediction for the decade ending in year 2000.

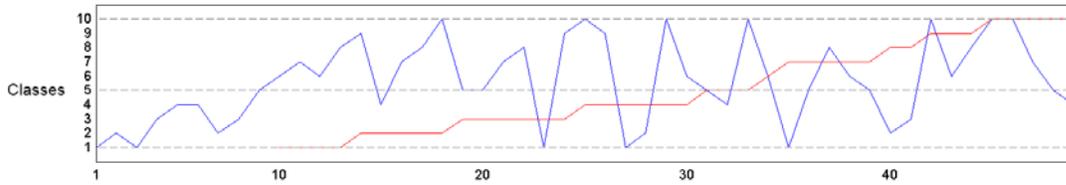


Figure 21: Yearly (blue) and decadal (red) streamflow evolution by classes for the Sacramento River up to 1999, as defined in Figure 20.

With this evolving information in hand, it is explained next how it may be possible to further synthesize the class evolutions in order to try to establish relations between past years and decades that may perhaps be useful in predicting as well. Following a Markovian framework, Fig. 22 summarizes the transition matrices that may be defined based on class information: from yearly to yearly, from decade to decade, and from decade to yearly sets. For instance, while the decade to decade shows simple diagonal patterns that suggest a reasonable predictability at such a scale, the yearly to yearly and decadal to yearly matrices exhibit broader distributions that nonetheless may trim away some states, say, once a prediction of the next decade is known.

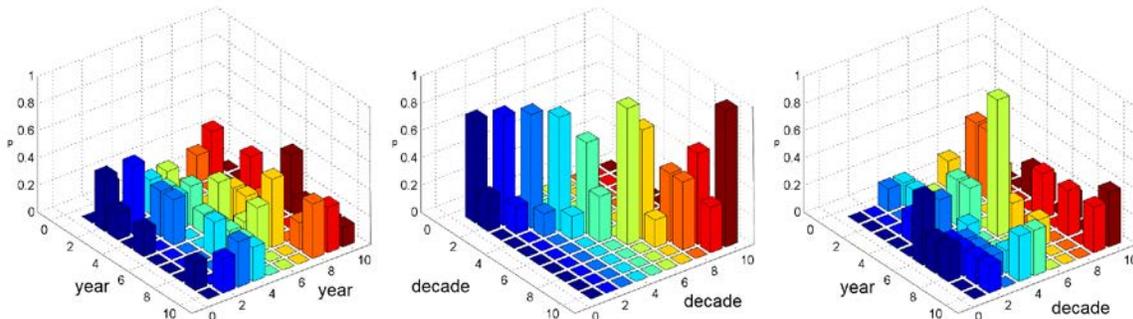


Figure 22: Transition matrices for streamflow dynamics at the Sacramento River up to 1999 corresponding to the classes in Figure 20. Graphs are read from right to left.

7.3 Predictions of Streamflow via FM Parameters and Classes

Given the synthesis of FM parameters into centroids of classes and the calculation of transition matrices up to a given year, various ideas may be used in order to define suitable predictions for the next water decade and then the next water year. As the FM parameters and classes for yearly streamflow records exhibit notorious

variability for the Sacramento River (i.e., Figs. 18 and 21), predicting the next year would have to rely on forecasts of the next decade and then transfers of such via transition matrices. Some variants for predicting streamflow at the decadal and then the yearly scales are given next.

7.3.1 Decadal Predictions

Given that the evolution of decadal FM parameters and classes exhibit smoothness, it becomes natural to employ a statistical time series representation (i.e., an ARMA model) on the individual parameters and to extrapolate the class of the last available decade to the following decade.

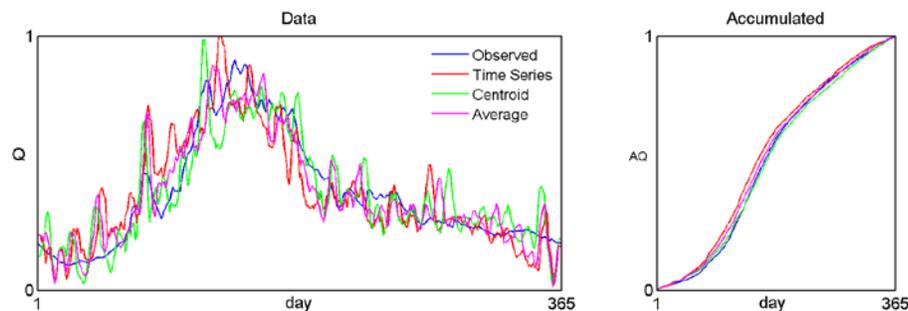


Figure 23: Plausible predictions for water decade 2000 based on FM parameter evolution and FM classes.

In this spirit, Fig. 23 shows the water decade ending in year 2000 (blue) and three plausible predictions stemming from information gathered up to year 1999: (a) a decadal set given by the FM decadal parameters extrapolated via time series models to 2000 (red), (b) the set corresponding to the predicted class for year 2000 defined as the one having the largest transition probability emanating from the class of decade 1999 (green), and (c) the set obtained by averaging the FM parameters used in defining the previous two sets (magenta). As may be seen, although the three “predictions” are less smooth than the records, their accumulated sets – depicting the timing of decadal streamflow over the year – yield rather reasonable patterns for the Sacramento River. In fact, the mean square and maximum errors in accumulated sets (ϵ_{ar} and ϵ_{max}), are, in order, (3.4, 7.2), (1.3, 2.3), and (1.7, 4.0), in percent, whose small ranges are typically encountered while repeating the analysis for other years (decades).

7.3.2 Yearly Predictions

As already mentioned, it is not possible to build meaningful time series models from the parameters in Fig. 18. However, predictions may be defined by: (a) looking at the class evolutions at the yearly scale (considering the future class that maximizes the transition probability from the current state) and (b) considering predictions at the decadal scale and transforming such into the yearly scale using the decadal to yearly transition matrix, both as reported in Fig. 22.

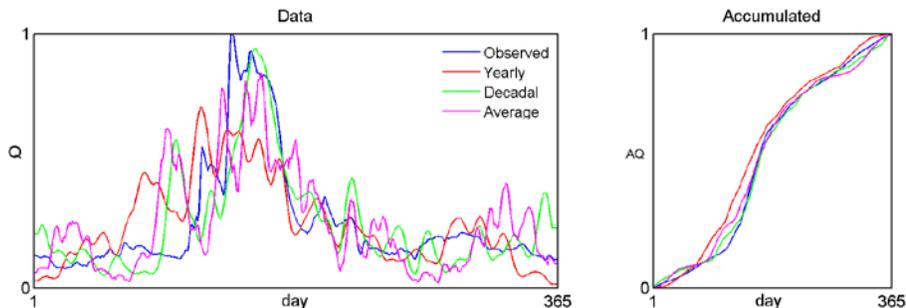


Figure 24: Plausible predictions for water year 2000 based on FM classes and transition matrices.

To continue with the example, Fig. 24 illustrates the notions for water year 2000 (blue) and “successful” predictions based on: (a) year to year class predictions as just explained (red), (b) decadal to yearly extrapolation based on the centroid-decadal prediction for decade ending in 2000 (i.e., from the best decadal prediction in Fig. 23) (green), and (c) the set obtained by averaging the FM parameters used in defining the previous two sets (magenta). As may be seen, although the three “predictions” exhibit intrinsic variability and do not capture in detail the main peak on the yearly records, their accumulated sets – portraying the timing of streamflow over the year – yield indeed reasonable patterns for the Sacramento River in year 2000, especially the prediction obtained by transferring the faithful predictions at the decadal scale. The mean square and maximum errors in accumulated sets for these predictions (ϵ_{ar} and ϵ_{max}), are, in order, (5.5, 15.0), (2.0, 4.9), and (3.1, 8.8), in percent, which, although larger than the ones obtained for decadal information, hold promise that a geometric approach may produce holistic streamflow forecasts.

Even though the quality of the forecasts just presented do vary with the year considered (at the Sacramento River and elsewhere) and there are instances in which the notions provide unfaithful predictions (with maximum errors in accumulated sets greater than 30%, as reflected by the spread in Fig. 20), there is yet another idea that may be used to try to create alternative predictions. Following the same overall notions, the idea is not only to use a predicted centroid but also the years associated with such a class and then generate, in a combinatorial sense, various sets of FM parameters. These may be used to obtain a certain number of predicted “realizations” that may be analyzed to find average holistic forecasts augmented by their spreads. Results from this idea shall be reported elsewhere.

7.4 Remarks

The above results have illustrated how the FM ideas may be employed, in conjunction with further synthesis of available information, in order to define plausible future holistic scenarios of a geophysical data set at the yearly and decadal scales, and beyond. It has been shown that the FM parameters of successive streamflow sets allow us to: (a) visualize the evolution of patterns providing relevant hints about the geometric complexity of the process, (b) establish a geometric classification of sets based on clustering of FM parameters, (c) develop relations from past information into the future via transition matrices from year to year, decade to decade, and decade to year, and (d) propose a methodology for finding holistic forecasts of future years and decades.

Although surmising accurate future scenarios remains a challenge and should be tried broadly before offering widespread generalizations, it is envisioned that the ideas herein may be useful to quantify the geometry of streamflow in different rivers in distinct geographical regions, leading to a better understanding of their implicit complexity. Certainly, the same notions may be tried on sets beyond streamflow and such may naturally include observations of rainfall, water and air temperature, evaporation, and others. In regards to the water temperature and rainfall records shown before (Figs. 10 and 12), it may be said that their FM

parameters exhibit ample variability (qualitatively as much as found in the yearly streamflow records), something that could be expected for the rainfall set but not so for the smoother temperature patterns. What these results stress (not shown due to space restrictions) is the fact that geometric variability from year to year is rather common in nature, which explains why it is not easy to describe and forecast geoscience phenomena.

8. Conclusions and Future Research

This chapter encapsulates the application of the deterministic fractal-multifractal FM approach to the encoding, simulation, disaggregation/downscaling, classification, and prediction of geoscience records. This work has illustrated that the FM method (and its variants), when coupled with a suitable optimization scheme, may be used to parsimoniously describe a host of patterns including rainfall events, daily rainfall sets over a year, daily streamflow sets over a year, and also daily water temperature records. Overall, mildly intermittent sets are easier to encode, as it takes, for a given set, from three to five hours of CPU on a personal computer. Highly intermittent rainfall sets are harder to process, as their optimization process takes, for a given set, about a day of CPU time.

As hinted in Section 3, the FM parameter space corresponding to a wire or a leaf is highly complex, and even more so for Cantorian representations that also include thresholds. Although there is continuity between the FM parameters and the sets they induce (i.e. the FM graphs vary a little when a single FM parameter varies a little), multiple changes of parameters end up generating similarly looking attractors from various combinations of parameters and, hence, there is no unique optimal solution. Although the patterns shown throughout this chapter are close renderings of the target sets, the presence of alternative close solutions ought to be studied in detail. It certainly would be relevant to ascertain the dynamics of, say, the best ten solutions to a given process and to study such representations on a multi-dimensional sense trying to encounter further relationships that may help in producing improved forecasts. It is envisioned that data mining techniques and the general notion of principal components may be useful in such a research.

In regards to the plausibility of improving predictions, few ideas may be tried for further improvements. Such encompasses combining FM parameters at various resolutions beyond years and decades, using say “pentades” (aggregating the records every five years) and employing in the definition of the classes a metric that weighs the distinct parameters according to their intrinsic variability, and not equally as it has been reported herein. Certainly these notions, coupled with the usage of (ten) alternative “solutions,” may yield improvements that may help elucidate climate change effects and trends.

Overall, there is much that needs to be done in order to fully study geoscience records using the FM approach. Besides trying the ideas in a variety of catchments, the following are relevant questions that represent future research. How do the FM parameters vary spatially when streamflow records of sub-catchments (upstream) are compared to those of the catchment (downstream)? How are the FM parameters of rainfall observations related to those of streamflow at the same site? Do the FM parameters vary in a systematic way when performing downscales at various scales? How are the FM parameters of various attributes related on a given site? How are the FM parameters of, say, rainfall, streamflow, and evaporation related to climate indicators? Are trends in climate relatable to discernible changes on the various attributes such that a changing hydrology may be elucidated? How are the FM parameters of distinct processes related to the underlying physics (conservation laws) of the phenomena? Are there discernible physical explanations for each one of the FM parameters?

Clearly, the scope of the FM approach is not limited to the study of geoscience records only, as, in a rather natural way, it may also be used to enhance other disciplines, such as physics, engineering, pattern recognition, general statistics, medical sciences, and finance. It is envisioned that applications of the FM approach, and also for patterns in higher dimensions (Puente 2004), will appear in such fields in the future.

Keywords: Cantor attractor, deterministic encoding, disaggregation, downscaling, fractal-multifractal method, k-means clustering, multiplicative cascade, pattern classification, prediction, projection, rainfall, simulation, streamflow, transition matrices, water temperature.

Symbols and Notation

ϵ_{ar} : root mean square error on accumulated records

ϵ_{max} : maximum error on accumulated records

E_q : entropy function of a data set

η_c : Nash-Sutcliffe efficiency on autocorrelation

η_d : Nash-Sutcliffe efficiency on data

η_h : Nash-Sutcliffe efficiency on histogram

η_e : Nash-Sutcliffe efficiency on entropy

Z_m : percent of zeros matched

τ_0 : lag when autocorrelation becomes zero

$\tau_{1/e}$: lag when autocorrelation becomes $1/e$

μ_{90} : mass of FM fitted histogram equivalent to 90% mass of observed histogram

M_{cz} : maximum consecutive zero values

N_Z : number of zero values

References Cited

- Arthur, D. and V. Sergi. 2007. K-means++: The Advantages of Careful Seeding. In: SODA '07 Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. pp 1027–1035.
- Barnsley, M. F. 1988. *Fractals Everywhere*. Academic Press, San Diego.
- Cortis, A., C. E. Puente and B. Sivakumar. 2009. Nonlinear extensions of a fractal–multifractal approach for environmental modeling. *Stoch. Environ. Res. Risk. Assess.* 23(7):897-906.
- Cortis, A., C. E. Puente, B. H. H. Huang, M. L. Maskey, B. Sivakumar and N. Obregón. 2013. A physical interpretation of the deterministic fractal-multifractal method as a realization of a generalized multiplicative cascade. *Stoch. Environ. Res. Risk Assess.* 28(6):1421–1429.
- Döll, P. and H. M. Schmied. 2012. How is the impact of climate change on river flow regimes related to the impact on mean annual runoff? A global-scale analysis. *Environ. Res. Lett.* 7(1):014037.
- Feder, J. 1988. *Fractals*. Plenum Press, New York.
- Fernández-Martínez, J. L., E. García Gonzalo, J. P. Fernández Álvarez, H. A. Kuzma and C. O. Menéndez Pérez. 2010. PSO: A powerful algorithm to solve geophysical inverse problems Application to a 1D-DC resistivity case. *J. Appl. Geophys.* 71(1):13–25.
- Georgakakos, K. P., A. A. Carsteanu, P. L. Sturdevant and J. A. Cramer. 1994. Observation and analysis of midwestern rain rates. *J. Appl. Meteor.* 33(12):1433-1444.
- Gupta, V. K. and E. Waymire. 1990. Multiscaling properties of spatial rainfall and river flow distributions. *J. Geophys. Res.* 95(D3):1999–2009.
- Hammel, R. D., R. J. Cooper, R. M. Slade, R. L. Haney and J. G. Arnold. 2006. Cumulative uncertainty in measured streamflow and water quality data for small watersheds. *Tran. Am. Soc. Agric. Eng.* 49(3):689.

- Huang, H. H., C. E. Puente and A. Cortis. 2012a. Geometric harnessing of precipitation records: reexamining four storms from Iowa City. *Stoch. Environ. Res. Risk. Assess.*, 27(4):955-968.
- Huang, H. H., C. E. Puente, A. Cortis and B. Sivakumar. 2012b. Closing the loop with fractal Interpolating functions for geophysical encoding. *Fractals*, 20(3-4):261-270.
- Huang, H. H., C. E. Puente, A. Cortis and J. L. Fernández Martínez. 2013. An effective inversion strategy for fractal–multifractal encoding of a storm in Boston. *J. Hydrol.* 496:205-216.
- Koutsoyiannis, D. 1992. A nonlinear disaggregation method with a reduced parameter set for simulation of hydrologic series. *Water Resour. Res.* 28(12):3175-3191.
- Koutsoyiannis, D. 1994. A stochastic disaggregation method for design storm and flood synthesis. *J. Hydrol.* 156(1):193-225.
- Lanza, L. G. and E. Vuerich. 2009. “The WMO field intercomparison of rain intensity gauges.” *Atmos. Res.* 94 (4):534–543.
- Lovejoy, S. and D. Schertzer. 2013. *The Weather and Climate: Emergent Laws and Multifractal Cascades.* Cambridge University Press, Cambridge.
- Mandelbrot, B. B. 1982. *The Fractal Geometry of Nature.* HB Fenn and Company.
- Mandelbrot, B. B. 1989. Multifractal measures especially for the geophysicist. pp. 1-42. *In:* C.H. Scholz and B. B. Mandelbrot (eds.). *Fractals in geophysics.* Birkhanser Verlag, Basel.
- Maskey, M. L., C. E. Puente and B. Sivakumar. 2015. Encoding daily rainfall records via adaptations of the fractal multifractal method. *Stoch. Environ. Res. Risk. Assess.* DOI 10.1007/s00477-015-1201-7.
- Meneveau, C. and K.R. Sreenivasan. .1987. Simple multifractal cascade model for fully developed turbulence. *Phys. Rev. Lett.* 59:1424–1427.
- Nash, J. and J. V. Sutcliffe. 1970. River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* 10(3):282–290.

- Obregón N, C. E. Puente and B. Sivakumar 2002a. Modeling high resolution rain rates via a deterministic fractal–multifractal approach. *Fractals* 10(3):387–394.
- Obregón N, C. E. Puente and B. Sivakumar 2002b. A deterministic geometric representation of temporal rainfall. Sensitivity analysis for a storm in Boston. *J. Hydrol.* 269(3–4):224–235.
- Olsson, J. 1998. Evaluation of a scaling cascade model for temporal rainfall disaggregation. *Hydrol. Earth Syst. Sc.* 2(1):19-30.
- Puente, C. E. 1996. A new approach to hydrologic modelling: derived distribution revisited. *J. Hydrol.* 187:65–80.
- Puente, C. E. 2004. A universe of projections: may Plato be right? *Chaos, Solitons and Fractals* 19(2):241–253.
- Puente, C. E. and N. Obregón. 1996. A deterministic representation of temporal rainfall: result for a storm in Boston. *Water. Resour. Res.* 32(9):2825-2839.
- Puente, C. E. and B. Sivakumar. 2007. Modeling geophysical complexity: a case for geometric determinism. *Hydrol. Earth Syst. Sci.* 11, 721–724.
- Rodriguez-Iturbe, I. 1986. Scale of fluctuation of rainfall models. *Water. Resour. Res.* 22(9):15S–37S.
- Sivakumar, B. 2000. Chaos theory in hydrology: important issues and interpretations. *J. Hydrol.* 227(1-4):1-20.
- Sivakumar, B. 2004. Chaos theory in geophysics: past, present and future. *Chaos, Solitons Fractals* 19(2):441-462.
- Valencia, D. and J. C. Schaake. 1973. Disaggregation processes in stochastic hydrology. *Water Resour. Res.* 9(3):580-585.