Studying the Complexity of Rainfall within California via a Fractal Geometric Method

Carlos E. Puente[1], Mahesh L. Maskey[1] and Bellie Sivakumar[1,2]

Abstract A deterministic geometric approach, the fractal–multifractal (FM) method, useful in modeling storm events and recently adapted in order to encode highly intermittent daily rainfall records, is employed to study the complexity of rainfall sets within California. Specifically, sets –from south to north– at Cherry Valley, Merced, Sacramento and Shasta Dam and containing respectively 59, 116, 115 and 72 years, all ending at water year 2015, are studied. The analysis reveals that: (a) the FM approach provides faithful encodings of all records, by years, with mean square and maximum errors in accumulated rain that are less than a mere 2% and 10%, respectively; (b) the evolution of the corresponding "best" FM parameters, allowing visualization of the inter-annual rainfall dynamics from a reduced vantage point, exhibit a highly entropic variation that prevents discriminating between sites and extrapolating to the future; and (c) the rain signals at all sites may be termed "equally complex," as usage of k-means clustering and conventional phase space analysis of FM parameters yields comparable results for all sites.

1. Introduction

As rainfall is a fundamental input to the hydrologic system, quantifying its temporal and spatial complexity is paramount for the proper planning, design and implementation of water resource infrastructure. As the process often exhibits complex non-linear behavior and high-intermittency, it is desirable to develop improved techniques that may allow further understanding of its structure.

With the development of stochastic and fractal notions and the advent of modern computation, a substantial effort has been made in the past few decades at conceptualizing numerous reainfall models. Attempting to capture the erratic, intermittent, random, and, in short, complex nature of rainfall, various frameworks have been proposed. Such include attempts to quantify complexity based on: (a) chaotic features of the records (e.g., Rodríguez-Iturbe et al. 1989; Sharifi et al. 1990; Ghilardi and Rosso 1990; Rodríguez-Iturbe 1991; Jayawardena and Lai 1994; Kout-

[1]Department of Land, Air and Water Resources, University of California, Davis

Davis, CA 95616, USA

Email: cepuente@ucdavis,edu

[2] School of Civil and Environmental Engineering, The University of New South Wales,

Sydney, NSW 2052, Australia

In **Nonlinear Advances in Geosciences**, 2017,
edited by A. A. Tsonis, Springer Nature.

soyiannis and Pachakis 1996; Sivakumar et al. 1999, 2001a; Peters et al. 2001; Men et al. 2004; Dhanya and Kumar 2010; Jothiprakash and Fathima 2013), (b) nonlinear time series models (e.g., French et al. 1992; Luk et al. 2000; Jin et al. 2005; Ramirez et al. 2005; Nasseri et al. 2008; Kim et al. 2009; Sivakumar 2009; Sivakumar and Singh 2012; Sivakumar et al. 2014), and (c) representations aiming at preserving statistical and fractal and multifractal rainfall properties (e.g., Rodríguez -Iturbe 1986; Gupta and Waymire 1990; Tessier et al. 1993; Lovejoy and Schertzer 2013; Puente and Obregón 1996; Sivakumar 2000, 2004; Sivakumar et al. 2001b; Maskey et al. 2015).

Relevant to this research, Puente (1996) developed a simple geometric procedure, the so-called fractal-multifractal (FM) method, which generates "seemingly-random" sets as fractal transformation of multifractal measures without requiring any statistical assumptions. This method, which fits within the modern notion of a fourth paradigm in data-intensive scientific discovery (Hey et al. 2009), produces a vast class of patterns defined over one and higher dimensions that not only preserves key statistical indicators, viz. moments, autocorrelation function, power spectrum, multifractal spectrum, but also captures intricate details and the textures present in the data sets, something which is quite difficult to accomplish using (physical) stochastic models (Puente 2004; Cortis et al. 2009).

While the FM approach has already been used to model: (a) rainfall events (e.g., Puente 1996; Obregón et al. 2002a,b; Huang et al. 2012a,b), (b) daily rainfall sets gathered over a year (Maskey et al. 2015, 2016b; Puente et al. 2016), (c) daily streamflow records over a year (Puente et al. 2016; Maskey et al. 2016a), (d) daily temperature measurements (Puente et al. 2016), and (e) even spatial contaminant plumes (Puente et al. 2001a,b), this article represents the first effort in using the FM method as a tool to quantify rainfall complexity, using for the purpose data sets collected within California.

The organization of the paper is as follows. Given first is an introduction to the FM notions and, in particular, the specific adaptation used to model intermittent rainfall records. This is followed by the methodology employed in fitting specific data sets via a numerical optimization exercise and an explanation of how FM geometric parameters will be used to quantify complexity. Then, the analysis of the records' complexity at four stations, from south to north, Cherry Valley, Merced, Sacramento and Shasta Dam, is advanced, including a study of the inter-annual dynamics and data-mining classifications at each site and a comparison of complexity features in space among the sites. The article concludes with its conclusions and recommendations.

2. The Fractal-Multifractal method

The transformation of multifractal measures via fractal interpolating functions, leading to the fractal-multifractal (FM) method (Puente 1996), is reviewed here.

A *fractal interpolating function* $f: x \to y$, passing through $N+1$ ordered points in plane $\{(x_n, y_n) |\, x_0 < x_1 < \cdots < x_N\}$ and having a graph $G = \{(x, f(x) | x \in$

$[x_0, x_N] = [0,1]\}$, is defined as the unique fixed point of N affine maps (Barnsley 1988):

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

such that, $G = w_1(G) \cup w_2(G) \cup \dots \cup w_N(G)$. While the vertical scalings d_n are free parameters satisfying $|d_n| < 1$, the other coefficients in Eq. (1), a_n, c_n, e_n and f_n, are evaluated via the contracting initial conditions:

$$w_n \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} x_{n-1} \\ y_{n-1} \end{pmatrix}, \qquad w_n \begin{pmatrix} x_N \\ y_N \end{pmatrix} = \begin{pmatrix} x_n \\ y_n \end{pmatrix}, \; n = 1, \dots, N, \tag{2}$$

which guarantee the existence of a stable attractor and yield N systems of linear equations that may be easily solved in terms of the interpolating points and the vertical scalings. Upon successive iterations of the maps, a convoluted "wire" function f, whose graph has a fractal dimension $1 \leq D < 2$, is found.

The notions may be generalized so that a more general attractor, other than a function, is obtained. Such is easily accomplished replacing the contractile initial conditions (Eq. 2) by more general contractions:

$$w_n \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} x_{2n} \\ y_{2n} \end{pmatrix}, \qquad w_n \begin{pmatrix} x_{2N-1} \\ y_{2N-1} \end{pmatrix} = \begin{pmatrix} x_{2n+1} \\ y_{2n+1} \end{pmatrix}, \; n = 1, \dots, N, \tag{3}$$

such that the range in x of map w_n becomes the interval $[x_{2n}, x_{2n+1}]$. Notwithstanding the need of additional end-point parameters y_{2n}, y_{2n+1}, a disperse attractor, over a Cantor set (Mandelbrot, 1982), is defined whenever the domain of the attractor contains gaps (Huang et al. 2013; Maskey et al. 2015).

Figure 1 illustrates how a disperse attractor is constructed iterating two affine maps whose end-points are $\{(0,0), (0.41,1.08)\}$ and $\{(0.80,4.02), (1,-0.35)\}$, namely:

$$w_1 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0.41 & 0 \\ 0.97 & -0.32 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{4}$$

$$w_2 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0.19 & 0 \\ -4.53 & -0.43 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0.80 \\ 4.02 \end{pmatrix}, \tag{5}$$

and when such are iterated following independent outcomes of a 33-67% biased coin.

As seen, the Monte Carlo procedure, known as the "chaos game" (Barnsley, 1988), defines (say after 2^{14} iterations) a Cantorian function from x to y, and also ultimately induces stable projections (histograms) dx and dy, which are hence functionally related and deterministic. While the former is clearly defined over a Cantor set (as there is a gap of 0.39 in end-points over x) and exhibits a *multifractal* structure containing noticeable repetition, the latter, which exhibits ample intermittency, is the derived measure over y found transforming the input measure dx via the fractal func-

tion from x to y, hence explaining the notation *fractal-multifractal* approach. For the sake of rainfall modeling, Fig. 1 also includes an adaptation of the notions via set dy_v so that additional zero values may be defined. Such a pattern is simply found trimming dy below a threshold ϕ, in a manner that evokes removing "traces" of rain.

By varying the parameters associated with the construction, the ideas herein yield indeed a host of rainfall-looking sets that shall be used later on to encode rainfall measurements via an inverse problem that uses recorded information (duly normalized) as the target of an FM optimization exercise that depend on the following parameters: (a) the end-points that define where the attractor would pass, (b) the ings d_n, (c) the frequencies used to carry the iterations p, and (d) the threshold ϕ used to trim the original output. Without a lack of generality, the first member of the first end-point and the second member of the second end-point have been set, respectively, to (0,0) and (1, y_{2N+1}). This implies a total of nine geometric FM parameters when iterating two maps.

Although, as mentioned earlier, the FM approach do fit within the modern notion of data-intensive scientific discovery (Hey et al. 2009), it is worth remarking that the notions have been assigned a physical interpretation, as the FM sets produced are deterministic realizations of non-trivial stochastic conservative multiplicative cascades, ultimately belonging to the class of "universal multifractals" (Cortis et al. 2013). Although no physical meaning may be assigned to the specific FM parameters nor such may be measured in any obvious way given a pattern (hence requiring an optimization exercise), the collective representations provided by the FM notions are part of a family of sensible physical entities and members of a collection of pertinent geometries with which to attempt to represent rainfall patterns and others (Puente et al. 2016). Certainly, it is not easy to capture the overall geometries of rainfall sets parsimoniously, but the FM approach does hold promise for such a desirable task that also aims at finding a suitable language for describing complexity (Maskey et al. 2015).

3. Methodology

This section explains how the FM approach is used to study the inter-annual and spatial variability of rainfall sets. First, the methodology used to find suitable encodings, for a year of data at a time, is introduced. Second, the validation statistics employed to quantify performance are given. Then, the analysis carried out to classify and quantify rainfall complexity is advanced.

3.1 FM encodings

Even though the FM methodology is ultimately rather simple and computationally efficient – once a set of parameters is known –, the finding of an appropriate representation for a given set is challenging. As there are neither analytical formulas for the attractors nor for the derived measures dy, only a numerical solution is possible. Also, alternative parameter sets may exist, i.e., equifinality (Beven 2006; Huang et al. 2013).

Following previous efforts (e.g., Maskey et al. 2015; Puente et al. 2016), a generalized particle swarm optimization (GPSO) algorithm, with swarm members having dynamic capabilities, is used in the study. Specifically, the GPSO procedure is run 200 times to find that many plausible solutions using swarms made of 500 randomly defined elements – FM parameter values defined uniformly between bounds – and allowing them to evolve following 100 successive iterations. The best parameter values for the 200 runs, even if local optima, are recorded for further study.

In trying to account for the inherent complexity in daily rainfall (Obregón et al. 2002a,b; Huang et al. 2013; Maskey et al. 2015), the objective function to minimize is defined adding three L^2 norms, i.e., root mean square errors, of accumulated rainfall vs. accumulated FM fitted values, at the daily, ε_1, three-day, ε_3, and seven-day, ε_7, scales, over the period of consideration (i.e., a year) plus a few penalties aimed at discarding unacceptable renderings. This gives for the objective function:

$$\varepsilon = \varepsilon_1 + \varepsilon_3 + \varepsilon_7 + \varepsilon_p, \tag{6}$$

where

$$\varepsilon_j = \sqrt{\frac{1}{M_j}\sum_{i=1}^{M_j}\left(c_{i,j} - \hat{c}_{i,j}\right)^2}, \qquad j = 1,3,7, \tag{7}$$

M_j is the number of data points at scale j, i.e., $M_1 = 365(366)$, $M_3 = 123$ and $M_7 = 53(54)$, $c_{i,j}$ is the accumulated measured rainfall up to period i for scale j, $\hat{c}_{i,j}$ is the corresponding value obtained via an FM representation, and ε_p are penalties dealing with the maximum allowable deviations between $c_{i,j}$ and $\hat{c}_{i,j}$ and information regarding their distribution of zero values. Although GPSO calculations may not always keep the penalty restrictions all the time, the results reported herein do fulfill such constraints.

3.2 Model performance

To assess the quality of individual FM approximations, various qualifiers are computed at the aforementioned scales. Such include:

(a) Nash-Sutcliffe efficiencies for, accumulated, rainfall vs. FM sets:

$$\eta_j = 1 - \frac{\sum_{i=1}^{M_j}\left(c_{i,j} - \hat{c}_{i,j}\right)^2}{\sum_{i=1}^{M_j}\left(c_{i,j} - \bar{c}_j\right)^2}, \qquad j = 1,3,7, \tag{8}$$

where the notation is as in Eq. (7) and $\bar{c}_j$ is the accumulated rainfall mean for scale j;

(b) Nash-Sutcliffe efficiencies of rainfall vs. FM sets (all duly normalized):

$$\nu_j = 1 - \frac{\sum_{i=1}^{M_j}\left(r_{i,j} - \hat{r}_{i,j}\right)^2}{\sum_{i=1}^{M_j}\left(r_{i,j} - \bar{r}_j\right)^2}, \qquad j = 1,3,7, \tag{9}$$

where $r_{i,j}$ is the measured rainfall at period i for scale j, $\hat{r}_{i,j}$ is the corresponding FM value and $\bar{r}_j$ is the rainfall mean for scale j;

(c) the number of zero values at the three scales, ζ_j in real vs. FM representations; and

(d) the percent of zero values matched by an FM encoding, π_j.

3.3 Complexity analysis

The encoding of a host of rainfall sets, by year and at various sites, allows visualizing, beyond the annual depth, the dynamics of the process from the vantage point of geometry, i.e., from the perspective of the parameters of the FM representations employed. As shall be illustrated, such FM parameters permit performing more complete inter-annual and spatial comparisons, which enable quantifying the complexity of the records, as follows: (a) yearly patterns may be classified by parameters via data mining techniques, e.g., k-means clustering (Arthur and Sergi 2007), such that rainfall dynamics may be studied by classes, and (b) parameter values may be studied following a "classical" complexity study, computing correlations of them all and building phase-space diagrams to identify the presence or lack of geometric trends.

4. Rainfall encodings in California

The FM notions are tested next for daily rainfall sets gathered over water years (October 1st to September 30th) in four sites in California, from south to north: Cheery Valley, Merced, Sacramento and Shasta Dam, as summarized in Table A1 in the Online Appendix http://puente.lawr.ucdavis.edu/pdf/nag_puente_appendix.pdf. As seen, all sites contain at least 59 years of contiguous sets gathered by NOAA's National Climate Data Center (NCDC) and the average annual rains (from south to north) are 46.7, 12.9, 17.6 and 63.7 inches. Prior to FM encoding, all data sets are normalized so that the accumulated depth (over a given year) becomes unity.

Given the geometric intricacies of the records that contain substantial number of zero values, the FM representations used are associated with the Cantorian construction based on the iteration of two maps as highlighted in Fig. 1. As they rely on nine FM geometric parameters, the computed sets have associated compression ratios of about 40:1 (365/9). Since encoding over 360 years of records (see Table A1) takes a substantial amount of time for solving the associated inverse problems (over 12 hours per year on a personal computer) and in order to study distinct possible solutions for the optimization exercises, the results reported here not only correspond to the "best" objective functions (over all 200 cases, as explained in Section 3) but also include, for sensitivity reasons, up to the best twenty solutions. In what follows and for clarity purposes, the quantity ε_1 is renamed as ε_{ac} (see Eq. (7)), the root mean

square error in daily accumulated sets over a year, and the results also report on the maximum daily deviation in accumulated rain over a year, ε_{mac}.

4.1 Examples of FM encodings

To demonstrate the capability of the FM notions, encodings of four distinct rainfall sets, with varied geometries and for each location, are reported here together with an extensive statistical evaluation of their performance. As an example, Fig. 2 contains rainfall sets (black), best FM encodings (gray) and comparisons of accumulated sets at Cherry Valley for water years ending at 1970, 1980, 1990 and 2000. As readily seen, all the FM sets do capture very well the overall "rough" distribution of rain and, although the rainfall intensities themselves are not perfectly captured –for the locations of all peaks do not necessarily match– the accumulated profiles of FM sets are indeed close to their targets.

The goodness of the best FM representations is further illustrated in Table 1 that contains, by blocks, a host of statistical information at various scales, as in the objective function (Eq. (6)). As seen in the top block, average and maximum errors in accumulated sets at the daily scale, ε_{ac} and ε_{mac}, are rather small, with magnitudes that do not exceed 2.2 and 9.8%, respectively. As noticed on the second and third blocks, while the Nash-Sutcliffe indices on the accumulated sets, η_j, for a day, three days and seven days are all close to perfection, the Nash-Sutcliffe indices on the records themselves, ν_j, exhibit low values at the daily and three day scales (due to non-matching rainfall intensities) but reasonable values –above 60%– at the weekly scale. As reported on the last two blocks, the numbers of zeros, ζ_j, in data and FM sets (in parenthesis) are close for all scales and the actual percent of zeros matched by the FM representations, π_j, is consistently higher than 75%, for all scales.

Similar analyses for the other three sites reveal comparably good performance, which, due to space limitations, have been included in the aforementioned Online Appendix. There, the interested reader shall find, for the Merced, Sacramento and Shasta Dam locations, four examples each of FM fits similar to those in Fig. 2 (Figs. A1, A2 and A3) followed by corresponding statistical information as in Table 1 (Tables A2, A3 and A4). As may be verified, such information supports the usage of the FM method to encode highly intermittent rainfall sets, as follows: (a) all FM representations, for the twelve examples, do resemble the real sets both in texture and overall locations of peaks and cannot be taken apart from data sets by the naked eye, as they do all look reasonable and "real," (b) all FM fits are excellent in accumulated sets as the encoding errors ε_{ac} and ε_{mac} are rather low, always less than merely 2.1% and 9.0%, respectively, (c) all Nash-Sutcliffe indices for accumulated sets, at the scales of one, three and seven days, η's, remain close to 100%, (d) Nash-Sutcliffe values for the records, ν's, increase with aggregation scale and the numbers at the seven day scale are typically larger than 65%, as with Cherry Valley, (e) the number of zeros in the records and those in the FM representations, ζ's, are close to each other, and (f) the FM fits do preserve well the location of zeroes in data, with π's that always exceed 54% but that could be, sometimes, as high as 96%.

4.2 Overall performance

Having studied in detail a few examples of rainfall patterns, this section includes the best FM representations over the whole records available: 59 years for Cherry Valley, 116 for Merced, 115 for Sacramento and 72 for Shasta Dam. In such spirit, Fig. 3 for Cherry Valley, and Figs. A4, A5 and A6 on the Online Appendix for the other sites, include the observed rainfall set (top), the corresponding FM fit (middle) obtained by upgrading annual volumes (depths), and their implied accumulated sets over the whole period (bottom).

As readily seen, the textures of FM sets associated with the best solutions (year by year) are indistinguishable from measured sets, not only by the naked eye, but also statistically as illustrated in the previous section. As such, the FM fitted accumulated sets turn out to be excellent renderings of the "real" sets, now over much longer periods of time. The goodness of such overall representations is further reflected in Table 2, which contains, for all sites, encoding errors and some of the statistical information used earlier at three aggregation scales, but averaged over all years and together with their plus and minus standard deviations. As seen, encoding errors are consistently low, for all sites, with ε_{ac} values that are on the average less than 1.8% and with standard deviations less than 0.3%, and with ε_{mac} values lower than 8% and standard deviations less than 2.1%. Such behavior translates into almost perfect Nash Sutcliffe values for accumulated sets at all sites, reasonable Nash-Sutcliffe indices for the records at the 7 day aggregation scale, ν_7, with averages above 65% and standard deviations less than 22%, and a large percent of zeroes matched by the FM representations for all scales, as mean values of π's are greater than 65% on the average and as the values for one day, π_1, are all greater than 83% with a standard deviation less than 8%.

These results illustrate that the Cantorian-based FM notions are useful to model highly intermittent rainfall sets containing a notorious amount of no-rain activity, with small errors that are within the accuracy of rainfall measurements (Lanza and Vuerich 2009). These results, even if only at four sites, do suggest that the FM geometric parameters of successive sets, year by year, may be used to study the evolution and complexity of rainfall patterns.

4.3 Rainfall dynamics

Having sensible approximations of rainfall sets at all sites leads us to hypothesize that the time evolution of the best FM parameters, as defined in Sect. 3, may help elucidate the inter-annual dynamics of rainfall. As such, Fig. 4 and Figs. A7, A8 and A9 (the latter on the Online Appendix) include the time evolution of the best FM parameters for Cherry Valley and the three other sites: the coordinates of the right end-point of the first map, (x_1, y_1), the coordinates of the left end-point of the second map, (x_2, y_2), the y-value of the right end-point of the second map, y_3, the vertical scalings of the two maps, d_1 and d_2, the frequency used to carry the iterations, p, and a rain-trace threshold ϕ. While the x_i values are bounded from the first and last end-points, i.e., from 0 to 1, the y_i values ranged from -5 to 5, and the d_i's between -1 and 1. All of

such values are shown normalized in the figures between 0 and 1, superimposing on them local parameter averages computed over 5 years.

As seen for all sites, the FM geometric parameters vary wildly and often swing from high to low values and vice-versa. Such variability is also seen in the averages every 5 years as reported in the graphs. Consistently with the early figures year by year (e.g., Fig. 2), the geometries of successive years do vary significantly, a feature that has already been reported not only for yearly rainfall sets but also for streamflow records (Maskey et al., 2015, 2016). This is corroborated in Table A5 on the Online Appendix, which includes entropy calculations for all parameters at all sites based on histograms containing eight bins resulting in maximum entropies of three (based on log_2 calculations for a purely uniform case). As seen and as hinted from the evolutions, all entropy values (except for the iteration frequency p and the threshold ϕ that a bit more orderly) reflect near uniformity, as all are between 2.75 and 2.98.

At the end, there are no noticeable trends in the best FM parameters (not even when averaged every five years) and such a fact clearly precludes the possibility of readily finding rainfall forecasts from such geometric information or discerning effects due to climate change. As seen, all sites, irrespective of their variable average annual volumes, exhibit a high degree of complexity, a trait that shall be further elaborated later on.

4.4 Sensitivity

In order to further understand the overall effectiveness of the FM method at all sites and in an attempt to find "sub-optimal" solutions that may exhibit parameter trends, Fig. 5 shows the evolution of the encoding errors ε_{ac} and ε_{mac}, but not only for the best solution every year, but for the best three. As seen, while the three root mean square errors (left) remain close to each other and at values that do not exceed small quantities of 2.4, 2.5, 2.8 and 2.6% from top to bottom (for sites from south to north), the three maximum errors - not explicitly optimized - (right) exhibit an increased variability that is bounded, from top to bottom, by 9.9, 10.0, 10.0 and 14.0%. Noticeably, few years at Shasta Dam exhibit large maximum errors and such may be interpreted as saying that from the point of view of the FM approach such a site is a bit more complex than the others. Although not shown here, a similar analysis of the best twenty solutions confirms the nature of the results. In terms of the employed objective function, the four sites within California may be considered comparable, but in terms of the maximum error in accumulated records (not explicitly accounted for), rainfall at Shasta Dam remains a bit more difficult to encode than in the other sites.

Figure 5 suggests that the FM encodings do not possess a single best solution but that there are other close solutions. As such, it becomes natural to inspect how the parameter evolution of such alternative FM representations may look. Figure 6, for Cherry Valley, and Figs. A10, A11 and A12 on the Online Appendix, for the other sites, present such information for the three best FM parameter values. As seen, those solutions that have close encoding errors as reported in Fig. 5, evolve wildly in time, implying solutions in various regions of FM parameter space that reveal the presence of equifinality (Beven 2006). There are indeed distinct parameter configurations yielding close solutions (Hill and Tiedeman 2007), a common feature regarding the

solution of inverse problems dealing with complex processes. This fact, and their implied evolutions of parameters, further elaborates the intrinsic complexity of rainfall, as comparisons of this block of figures exhibit rather similar degrees of variability. Certainly, a similar comparison of various solutions in the inherently complex and nonlinear rainfall process, may allow further quantifying climate complexity around the globe.

As there is no perceptible trend in the best parameters and as there are several close solutions, a set of alternative parameter values was defined for each site based on fixed lower bounds over time, as included in Table A6 (for un-normalized parameters) on the Online Appendix. Such "filtered" representations resulted in slightly worse performance as reflected in Table 3 when compared to Table 2. Still, however, such new FM parameters led to ample variation in parameters (albeit a bit less), as reflected in Fig. 7 for Cherry Valley and Figs. A13, A14 and A15 (on the Online Appendix) for Merced, Sacramento and Shasta Dam and as corroborated in the entropy analysis reported in Table A7 on the Online Appendix. Such filtered representations shall be used next together with the best parameters in order to classify rainfall patterns and further assess the inherent complexity of the rainfall records.

5. Classification and complexity analysis

5.1 Geometric and data classification

As there is ample variation in FM dynamics at all sites, both for best and filtered parameters, this section presents a discretized analysis of the rainfall records via FM parameter classifications (for both best and filtered solutions) yielding eight distinct classes using k-means clustering, as described in Sect. 3.3.

Figure 8 shows the time evolution of the aforesaid eight classes for all sites (all starting at "class 1" for the initial year), based on best solutions (on the left) and filtered solutions (on the right), together with their implied Markovian matrices summarizing transitions from time to time (to be read from right to left). As seen and as expected, the classification based on filtered parameters differs from that obtained from best parameters and all evolutions, at all sites, exhibit notable swings from low to high classes and vice-versa and broad Markovian matrices.

In order to further qualify the results just discussed, Fig. 9 includes a similar classification into eight classes obtained by using the deciles of the yearly records, that is, the same number of parameters as the FM representations in Fig. 8. As seen, the rainfall class evolutions based on deciles exhibit yet similar swings as those based on FM parameters and the corresponding Markov matrices remain fairly broad, except for two noticeable transitions for Cherry Valley and Shasta Dam. As reported in Table A8 on the Online Appendix, the class evolutions for all sites turn out to be, at the end, rather similar in terms of their class entropies, H_c, (ranging from 2.71 to 2.96 and hence close to uniformity) and average class orbit lengths, $\overline{OL_c}$, measured, in an absolute sense, over the evolving classes (spanning from 1.98 to 2.77 classes). Although the records for Sacramento may be termed a bit more complex based on these two attributes, there is not enough separation to conclude that sets are not similarly complex nor that any effects due to climate change may be identified.

Figure 9 also includes an inter-comparison between the classes implied by the decile classification and those obtained via best and filtered FM parameters, in the form of scatter plots having a larger circle depending on multiple occurrences of a given combination of classes. As seen, both for best and filtered FM classifications, there is no salient visible patterns that emerge from the analysis but rather very broad diagrams, which give credence to the notion that the three distinct classifications represent different ("orthogonal") views of the records (as other "equifinal" FM solutions would likely produce), which altogether exhibit similar degrees of complexity.

5.2 Comparative analysis in space

Having explained that rainfall may be considered equally complex from a geometric point of view at the four sites under study, it becomes sensible to compare the class evolutions in space, one site against another. For this purpose, Fig. 10 includes pairwise comparisons associated with the best FM solutions, and Figs. A16 and A17 (on the Online Appendix) do so for the filtered FM and decile classifications, respectively. As exemplified in Fig. 10, such graphs include a visual comparison of the class evolutions for concurrent years and the corresponding scatter plots.

As seen, the rather erratic class evolutions result in fairly uncorrelated scatter plots, which, although having instances where class combinations do not exist (as high classes for Cherry Valley vs. low classes for Sacramento in Fig. 10), do not exhibit noticeable trends. There are no clear correlations among the classes for all pairs of stations for any of the three classifications employed, hence emphasizing the complexity of the rainfall records, not only in time, but also in space, at least for a distance of about 550 miles within California.

5.3 Additional complexity analysis

Having obtained "best" FM solutions for every site allows computing autocorrelation functions and phase diagrams for such individual parameters, as usually performed when trying to identify chaotic properties of records, e.g., Sivakumar and Berndtsson (2010). In that spirit, Fig. 11 includes such an analysis for the nine FM parameters at Merced together with the total rainfall depth over the years (in inches) and Figs. A18, A19 and A20 on the Online Appendix do so, in order, for Cherry Valley, Sacramento and Shasta Dam.

As seen, Fig. 11 is divided into two parts, with the first one including the endpoint FM parameters x_1, x_2, y_1, y_2, y_3, and the second comprising the other FM parameters d_1, d_2, p and ϕ and the aforementioned total rainfall depth. As observed, for each attribute there are shown five features: (a) the time series itself, (b) the autocorrelation function, (c) the two-dimensional phase diagram with a lag equal to one year, (d) the two-dimensional phase diagram with a lag equal to half the total number of years in the records (58 for Merced), and (e) the three-dimensional phase diagram obtained using lags 0, 1, and 2.

As observed in Fig. 11, the annual rainfall depths for Merced, as well as the best FM parameters there, exhibit noticeable swings from high to low and vice-versa and, as expected, their corresponding autocorrelations decay rather quickly and re-

main almost always, within the shown $\pm 1.96/\sqrt{n}$ (where n is the length of records) bands for up to 58 lags. As seen, all phase diagrams for all FM parameters and total rainfall depth exhibit noticeable scattering at the scales considered. The diagrams shown, and others for additional scales not included here, confirm the complexity of the rainfall records, as there is no obvious attractor that may be discerned for any of the geometric attributes of the daily rainfall patterns. As the analysis for the other sites, shown in the corresponding figures on the Online Appendix, reveals rather similar results, it may indeed be surmised that the studied sets lack a low-dimensional structure that may define different degrees of (geometric) complexity for rainfall in space.

6. Conclusions and further research

This research illustrates how a Cantorian variant of the fractal-multifractal, FM, approach, which relies on the iteration of two simple maps and uses eight geometric parameters, may be adapted adding a rain trace threshold parameter to closely encode highly intermittent daily rainfall sets in California. By studying over 360 combined years at (from south to north) Cherry Valley, Merced, Sacramento and Shasta Dam, it is shown that it is possible to nicely approximate the geometry of individual years at the daily scale (i.e., optimizing their mass functions) with root mean square errors that are less than a mere 1.8 ± 0.3% and maximum errors in accumulated sets that are less than 7.9 ± 2.1%, both well within measurement errors reported for the rainfall process, Lanza and Vuerich (2009). As the FM encodings also reasonably preserve information pertaining to the distribution of zero rainfall values and Nash-Sutcliffe attributes for rainfall at the weekly scale, the results support the notion that hidden determinism may lie at the root of natural complexity (Puente 1996; Puente and Sivakumar 2007).

Once FM representations are established, the dynamics of the aforementioned nine parameters are used in an attempt to study the inter-annual dynamics of rainfall at the four sites, aiming also at a spatial comparison. The analysis revealed; however, that the evolutions of “best” FM parameters, obtained via an optimization exercise, fail to exhibit any noticeable trends but rather ample variations, for all sites, in a manner that does not reflect any variations due to climate change. As the optimization process revealed the presence of other FM parameter combinations having close objective functions, i.e., equifinality, “filtered” FM sets, narrowing the range of parameter variations, were also defined, but such evolutions also ultimately resulted in non-trivial and highly-entropic behaviors, for all sites.

As the evolutions of FM parameters were not useful in discerning differences in the complexity of rainfall among the chosen locations, a more complete analysis of such geometric parameters was carried via classifications (using k-means clustering) and conventional phase space diagrams. This investigation resulted in further understanding of the rather erratic signals and confirmed that the geometries of the rainfall sets at the four sites, and irrespective of distinct annual rainfall averages, cannot be distinguished from one another, as they all may be termed as “equally complex.”

The results of this work emphasize the "deterministic complexity" of the rainfall process, i.e., deterministic, as the individual sets may be represented by the FM method, but complex, as there are no obvious trends in FM parameters over time. The fact that there are equifinal FM solutions certainly suggests further investigating within the space of parameters, using explicit bounds in the numerical search aiming at defining trends. Although the results herein suggest that such may be unlikely, there may still be solutions that could allow discriminating rainfall complexity between sites. Certainly, the analysis should be extended further north to include rainfall sets with less numbers of zero values.

The quantification of rainfall complexity in time and space, as attempted in this work, is certainly a scientifically relevant problem, especially in relation to climate change studies. If trends in dynamics may be discerned, such would have obvious benefits, and if such trends do not exist, that would also be relevant information. Although several questions remain regarding the FM approach, for instance, finding a physical explanation for each of its parameters, it is envisioned that similar analysis may also be carried to assess the complexity of other (less complex) hydrometeorological attributes such as streamflow and temperature records. Such an avenue of research, with less variable geometric patterns, is also being investigated and will be reported in the future.

Acknowledgements

This article is dedicated to Panayiotis Tsonis, whom the first author hugged in Rhodes, as if he was his brother.

Notation

ε_{ac}: Root mean square error in accumulated set

ε_{mac}: Maximum error in accumulated set

ν_j: Nash-Sutcliffe efficiency on records at j-day scale

η_j: Nash-Sutcliffe efficiency on accumulated sets at j-day scale

ζ_j: Number of zeros present in the sets at j-day scale

π_j: Percent of zeros matched in the FM set at j-day scale

H_c: Entropy of class distribution

$\overline{OL_c}$: Average orbit length by classes

References

Barnsley MF, (1988) Fractals everywhere. Academic Press, San Diego

Beven K (2006) A manifesto for the equifinality thesis. J Hydrol 320(1):18-36

Cortis A, Puente CE, Sivakumar B (2009) Nonlinear extensions of a fractal-multifractal approach for environmental modeling. Stoch Environ Res Risk Assess 23(7):897–906

Cortis A, Puente CE, Huang HH, Maskey ML, Sivakumar B, Obregón N (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

Dhanya CT, Kumar DN (2010) Nonlinear ensemble prediction of chaotic daily rainfall. Adv Water Resour 33(3):327-347

French MN, Krajewski WF, Cuykendall RR (1992) Rainfall forecasting in space and time using a neural network. J Hydrol 137(1-4):1-31

Ghilardi P, Rosso R (1990) Comment on "Chaos in rainfall" by I. Rodríguez -Iturbe et al. Water Resour Resour 26(8):1837-1839

Gupta VK, Waymire E (1990) Multiscaling properties of spatial rainfall and river flow distributions. J Geophys Res 95(D3):1999–2009

Hey T, Stewart T, and Kristin MT (2009) The fourth paradigm: data-intensive scientific discovery. Vol 1, Microsoft research, Redmond, WA

Hill MC, Tiedeman CR (2007) Effective groundwater model calibration: With analysis of data, sensitivities, predictions and uncertainty. John Wiley & Sons, Inc., Hoboken, New Jersey

Huang HH, Cortis A, Puente CE (2012a) Geometric harnessing of precipitation records: reexamining four storms from Iowa City. Stoch Environ Res Risk Assess 27(4):955-968

Huang HH, Puente CE, Cortis A, Sivakumar B (2012b) Closing the loop with fractal interpolating functions for geophysical encoding. Fractals 20(3–4):261–270

Huang HH, Puente CE, Cortis A, Fernández Martínez JL (2013) An effective inversion strategy for fractal–multifractal encoding of a storm in Boston. J Hydrol 496:205–216

Jayawardena AW, Lai F (1994) Analysis and prediction of chaos in rainfall and stream flow time series. J Hydrol 153:23–52

Jothiprakash V, Fathima TA (2013) Chaotic analysis of daily rainfall series in Koyna reservoir catchment area, India. Stoch Environ Res Risk Assess 27(6):1371-81

Jin YH, Kawamura A, Jinno K, Berndtsson R (2005) Nonlinear multivariate analysis of SOI and local precipitation and temperature. Nonlinear Processes Geophys 12:67–74

Kim HS, Lee KH, Kyoung MS, Sivakumar B, Lee ET (2009) Measuring nonlinear dependence in hydrologic time series. Stoch Environ Res Risk Assess 23:907–916

Koutsoyiannis D, Pachakis D (1996) Deterministic chaos versus stochasticity in analysis and modeling of point rainfall series. Geophys Res Atmos, 101(D21):26441-26451

Lanza LG, Vuerich E (2009) The WMO field intercomparison of rain intensity gauges. Atmos Res 94(4):534–543

Lovejoy S, Schertzer D (2013) The Weather and Climate: Emergent Laws and Multifractal Cascades. Cambridge University Press, Cambridge

Luk KC, Ball JE, Sharma A (2000) A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting. J Hydrol 227(1):56-65

Maskey ML, Puente CE, Sivakumar, B, Cortis A (2015) Encoding daily rainfall records via adaptations of the fractal multifractal method. Stoch Environ Res Risk Assess 30:1917 doi 10.1007/s00477-015-1201-7

Maskey ML, Puente CE, Sivakumar, B (2016a) A comparison of fractal-multifractal techniques for encoding streamflow records. J Hydrol 542: 564-580

Maskey ML, Puente CE, Sivakumar, B, Cortis A (2016b) Deterministic Simulation of Highly Intermittent Hydrologic Time Series. Stoch Environ Res Risk Assess doi 10.1007/s00477-016-1343-2

Men B, Xiejing Z, Liang C (2004) Chaotic analysis on monthly precipitation on Hills Region in Middle Sichuan of China. Nature and Science, 2(2):45-51

Nasseri M, Asghari K, Abedini MJ (2008) Optimized scenario for rainfall forecasting using genetic algorithm coupled with artificial neural network. Expert Syst Appl 35(3):1415-1421

Obregón N, Puente CE, Sivakumar B (2002a) Modeling high resolution rain rates via a deterministic fractal–multifractal approach. Fractals 10(3):387–394

Obregón N, Puente CE, Sivakumar B (2002b) A deterministic geometric representation of temporal rainfall. Sensitivity analysis for a storm in Boston. J Hydrol 269(3–4):224–235

Peters O, Hertlein C, Christensen K (2001) A complexity view of rainfall. Phys Rev Lett 88(1):018701

Puente CE (1996) A new approach to hydrologic modelling: derived distribution revisited. J Hydrol 187:65–80

Puente CE (2004) A universe of projections: may Plato be right? Chaos, Solitons and Fractals 19(2):241–253

Puente CE, Obregón N (1996) A deterministic representation of temporal rainfall: result for a storm in Boston. Water Resour Res 32(9):2825–2839

Puente CE, Sivakumar B (2007) Modeling hydrologic complexity: a case for geometric determinism. Hydrol Earth Syst Sci 11:721–724

Puente CE, Robayo O, Díaz MC, Sivakumar B (2001a) A fractal–multifractal approach to groundwater contamination. 1. Modeling conservative tracers at the Borden site. Stoch Environ Res Risk Assess 15(5):357–371

Puente CE, Robayo O, Sivakumar B (2001b) A fractal–multifractal approach to groundwater contamination. 2. Predicting conservative tracers at the Borden site. Stoch Environ Res Risk Assess 5(5):372–383

Puente CE, Maskey ML, Sivakumar B (2016) Combining Fractals and Multifractals to Model Geoscience Records. In: Ghanbarian B, Hunt A (eds) Fractals: Concepts and Applications in Geosciences, CRC Press, in Press

Ramirez MCV, de Campos Velho, HF, Ferreira, NJ (2005) Artificial neural network technique for rainfall forecasting applied to the Sao Paulo region. J Hydrol 301(1):146-162

Rodríguez -Iturbe I (1986) Scale of fluctuation of rainfall models. Water Resour Res 22(9):15S–37S

Rodríguez-Iturbe I (1991) Exploring complexity in the structure of rainfall. Adv Water Resour 14(4):162-167

Rodríguez-Iturbe I, De Power FB, Sharifi MB, Georgakakos KP (1989) Chaos in rainfall. Water Resour Res 25(7):1667–1675

Sharifi MB, Georgakakos KP, Rodríguez-Iturbe I (1990) Evidence of deterministic chaos in the pulse of storm rainfall. J Atmos Sci 47(7):888-893

Sivakumar B (2000) Fractal analysis of rainfall observed in two different climatic regions. Hydrolog Sci J 45(5):727-738

Sivakumar B (2004) Chaos theory in geophysics: past, present and future. Chaos, Solitons and Fractals 19(2):441–462

Sivakumar B (2009) Nonlinear dynamics and chaos in hydrologic systems: latest developments and a look forward. Stoch Environ Res Risk Assess 23:1027–1036

Sivakumar B, Liong SY, Liaw CY, Phoon KK (1999) Singapore rainfall behavior: chaotic? J Hydrol Eng 4(1):38-48

Sivakumar B, Berndtsson R, Persson M (2001a) Monthly runoff prediction using phase space reconstruction. Hydrolog Sci J, 46(3), 377-387

Sivakumar B, Sorooshian S, Gupta HV, Gao X (2001b) A chaotic approach to rainfall disaggregation. Water Resour Res 37(1):61–72

Sivakumar B, Berndtsson R (2010) Advances in data-based approaches for hydrologic modeling and forecasting. World Scientific, Singapore

Sivakumar B, Singh VP (2012) Hydrologic system complexity and nonlinear dynamic concepts for a catchment classification framework. Hydrol Earth Syst Sci 16:4119–4131

Sivakumar B, Woldemeskel FM, Puente CE (2014) Nonlinear analysis of rainfall variability in Australia. Stoch Environ Res Risk Assess 28(1):17–27

Tessier Y, Lovejoy S, Schertzer D (1993) Universal multifractals: Theory and observations for rain and clouds. J Appl Meteor 32:223–250

List of Tables

List of Figures

Table 1: Performance of the best FM models for Cherry Valley in Fig. 2 ($\varepsilon_{ac}, \varepsilon_{mac}, \eta$'s, ν's and π's are in percent, ζ's are for data followed by FM fit in parenthesis)

Statistics	Period			
	1970	**1980**	**1990**	**2000**
$\boldsymbol{\varepsilon_{ac}}$	1.8	1.4	2.2	1.8
$\boldsymbol{\varepsilon_{mac}}$	8.0	7.0	8.4	9.8
$\boldsymbol{\eta_1}$	99.8	99.9	99.5	99.8
$\boldsymbol{\eta_3}$	99.8	99.9	99.5	99.8
$\boldsymbol{\eta_7}$	99.8	99.9	99.6	99.8
$\boldsymbol{\nu_1}$	1.0	17.9	-40.0	-40.0
$\boldsymbol{\nu_3}$	36.4	67.9	14.5	29.9
$\boldsymbol{\nu_7}$	65.2	87.1	61.9	71.4
$\boldsymbol{\zeta_1}$	295(300)	276(288)	300(319)	291(320)
$\boldsymbol{\zeta_3}$	85(85)	73(80)	84(91)	81(89)
$\boldsymbol{\zeta_7}$	27(30)	18(26)	26(32)	23(29)
$\boldsymbol{\pi_1}$	86.1	85.2	90.7	91.8
$\boldsymbol{\pi_3}$	78.8	78.1	80.6	84.0
$\boldsymbol{\pi_7}$	77.8	88.9	76.9	78.3

Table 2: Overall performance of best FM encoding in all locations ($\varepsilon_{ac}, \varepsilon_{mac}, \eta$'s, ν's and π's are in percent)

Statistics	Site			
	Cherry Valley	**Merced**	**Sacramento**	**Shasta Dam**
$\boldsymbol{\varepsilon_{ac}}$	1.6 ± 0.2	1.6 ± 0.2	1.6 ± 0.2	1.8 ± 0.3
$\boldsymbol{\varepsilon_{mac}}$	7.1 ± 1.5	7.0 ± 1.3	7.1 ± 1.2	7.9 ± 2.1
$\boldsymbol{\eta_1}$	99.8 ± 0.1	99.8 ± 0.1	99.8 ± 0.1	99.7 ± 0.1
$\boldsymbol{\eta_3}$	99.8 ± 0.1	99.8 ± 0.1	99.8 ± 0.1	99.7 ± 0.1
$\boldsymbol{\eta_7}$	99.8 ± 0.1	99.8 ± 0.1	99.8 ± 0.1	99.7 ± 0.1
$\boldsymbol{\nu_1}$	-11 ± 38	-7 ± 32	-9 ± 28	-18 ± 29
$\boldsymbol{\nu_3}$	34 ± 31	34 ± 27	43 ± 23	32 ± 20
$\boldsymbol{\nu_7}$	65 ± 22	68 ± 16	74 ± 13	66 ± 17
$\boldsymbol{\pi_1}$	85 ± 8	88 ± 7	91 ± 5	83 ± 8
$\boldsymbol{\pi_3}$	76 ± 11	81 ± 10	85 ± 8	75 ± 12
$\boldsymbol{\pi_7}$	68 ± 16	76 ± 14	80 ± 10	65 ± 16

Table 3: Overall performance of filtered FM encoding in all locations ($\varepsilon_{ac}, \varepsilon_{mac}, \eta$'s, ν's and π's are in percent)

Statistics	Site			
	Cherry Valley	**Merced**	**Sacramento**	**Shasta Dam**
$\boldsymbol{\varepsilon_{ac}}$	1.9 ± 0.3	1.8 ± 0.3	1.8 ± 0.3	2.0 ± 0.3
$\boldsymbol{\varepsilon_{mac}}$	7.4 ± 1.4	7.4 ± 1.4	7.5 ± 1.2	8.5 ± 2.6
$\boldsymbol{\eta_1}$	99.7 ± 0.1	99.7 ± 0.1	99.7 ± 0.1	99.7 ± 0.2
$\boldsymbol{\eta_3}$	99.7 ± 0.1	99.8 ± 0.1	99.8 ± 0.1	99.7 ± 0.2
$\boldsymbol{\eta_7}$	99.7 ± 0.1	99.8 ± 0.1	99.8 ± 0.1	99.7 ± 0.2
$\boldsymbol{\nu_1}$	-23 ± 44	-23 ± 40	-13 ± 32	-27 ± 35
$\boldsymbol{\nu_3}$	24 ± 36	24 ± 33	38 ± 25	27 ± 23
$\boldsymbol{\nu_7}$	59 ± 26	60 ± 22	69± 16	62 ± 20
$\boldsymbol{\pi_1}$	85 ± 7	89 ± 6	90 ± 5	85 ± 7
$\boldsymbol{\pi_3}$	76 ± 10	82 ± 9	83 ± 7	76 ± 10
$\boldsymbol{\pi_7}$	68 ± 13	77 ± 13	78 ± 9	68 ± 14

.

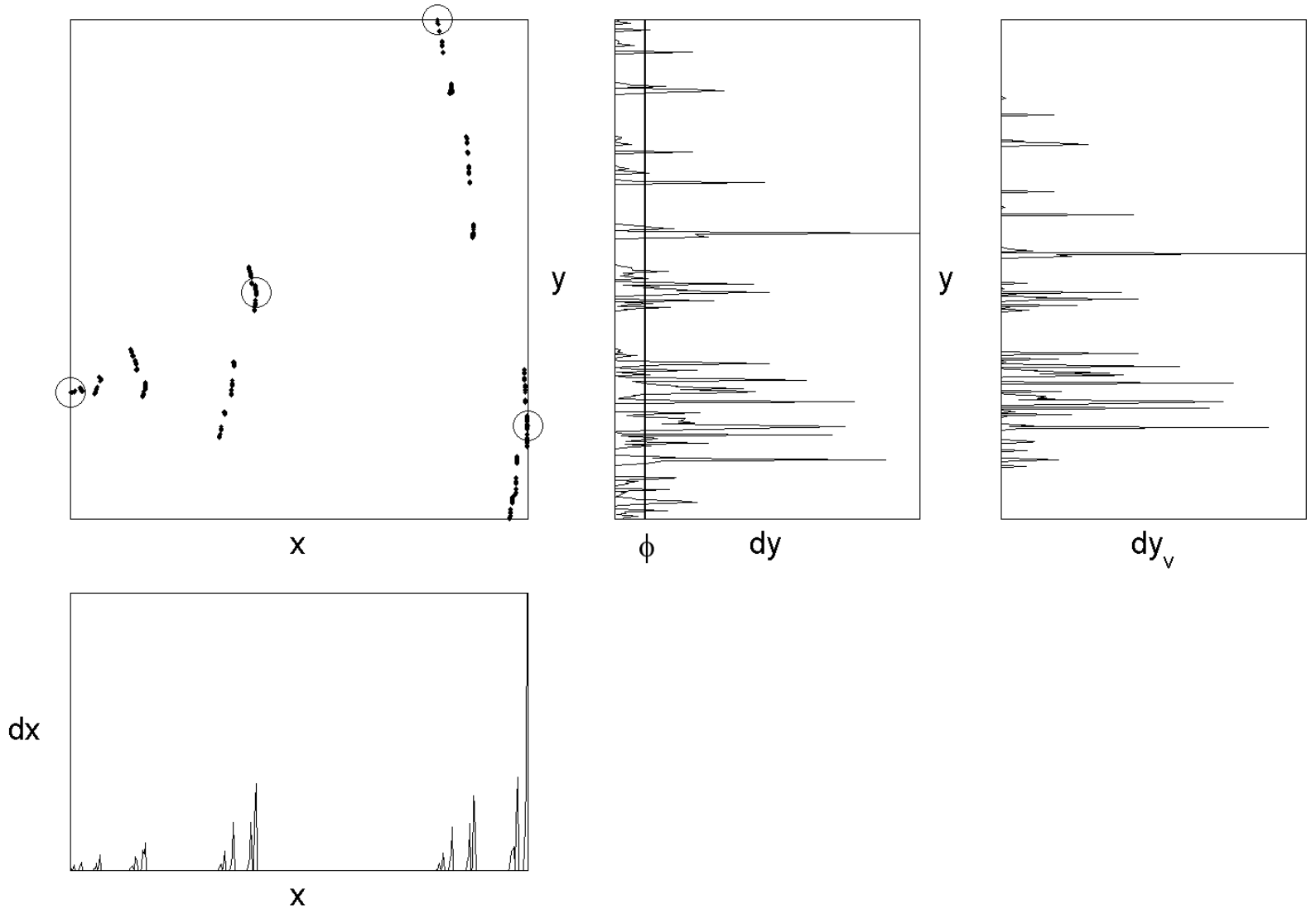

Fig. 1:

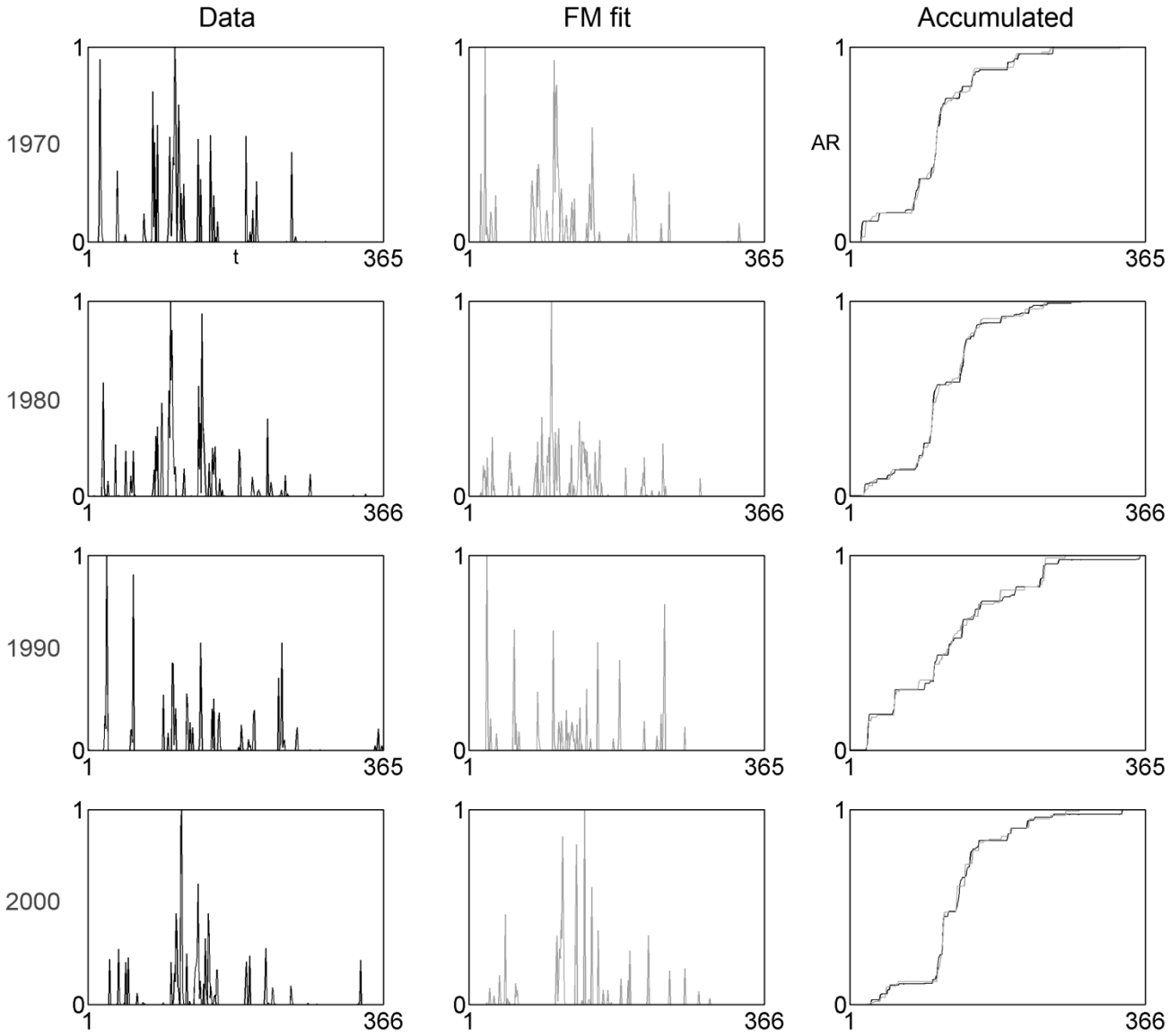

Fig. 2:

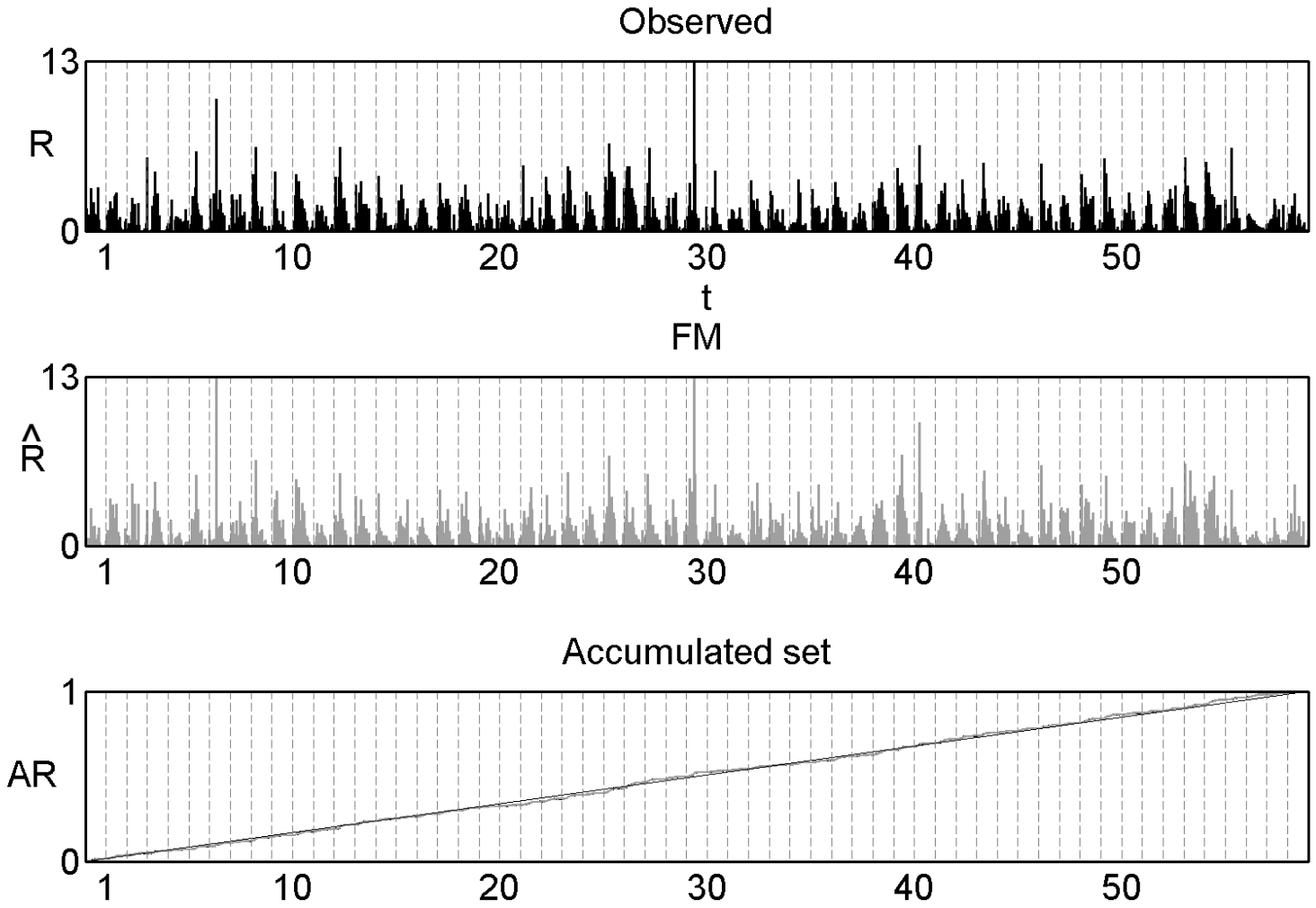
Observed
13
R
0
1
10
20
30
40
50
t
FM
13
R̂
0
1
10
20
30
40
50
Accumulated set
1
AR
0
1
10
20
30
40
50

Fig. 3:

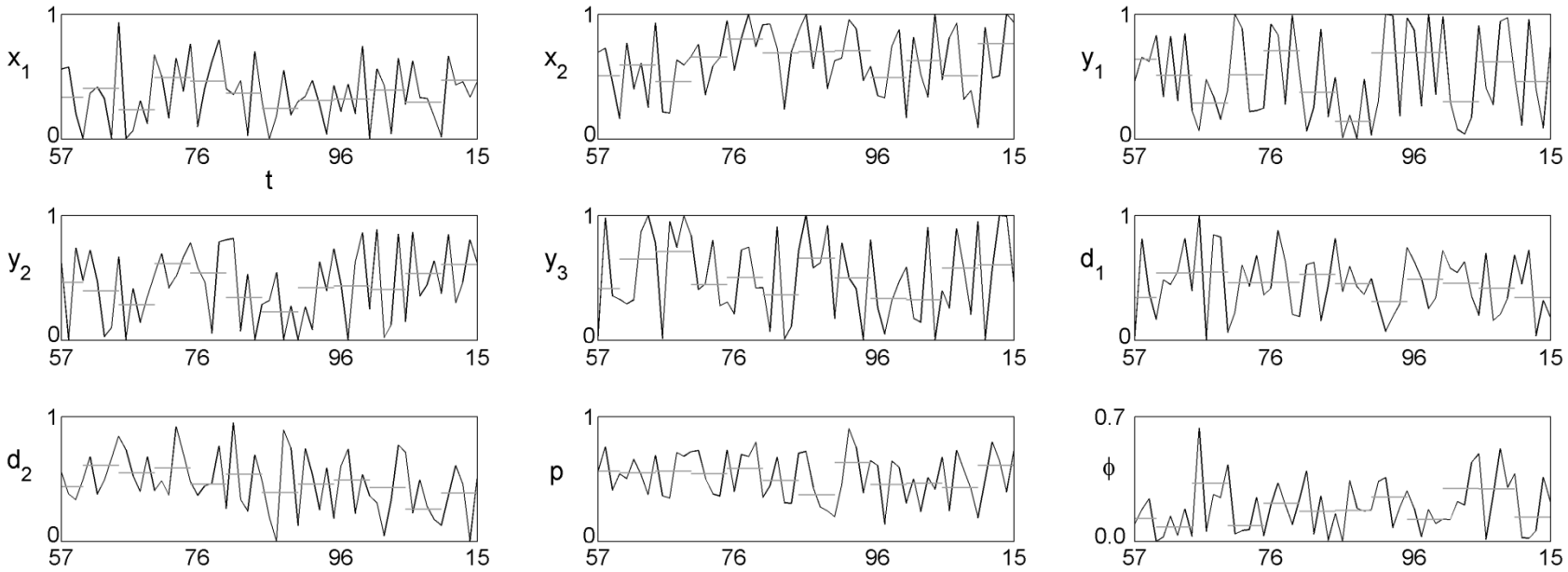

Fig. 4:

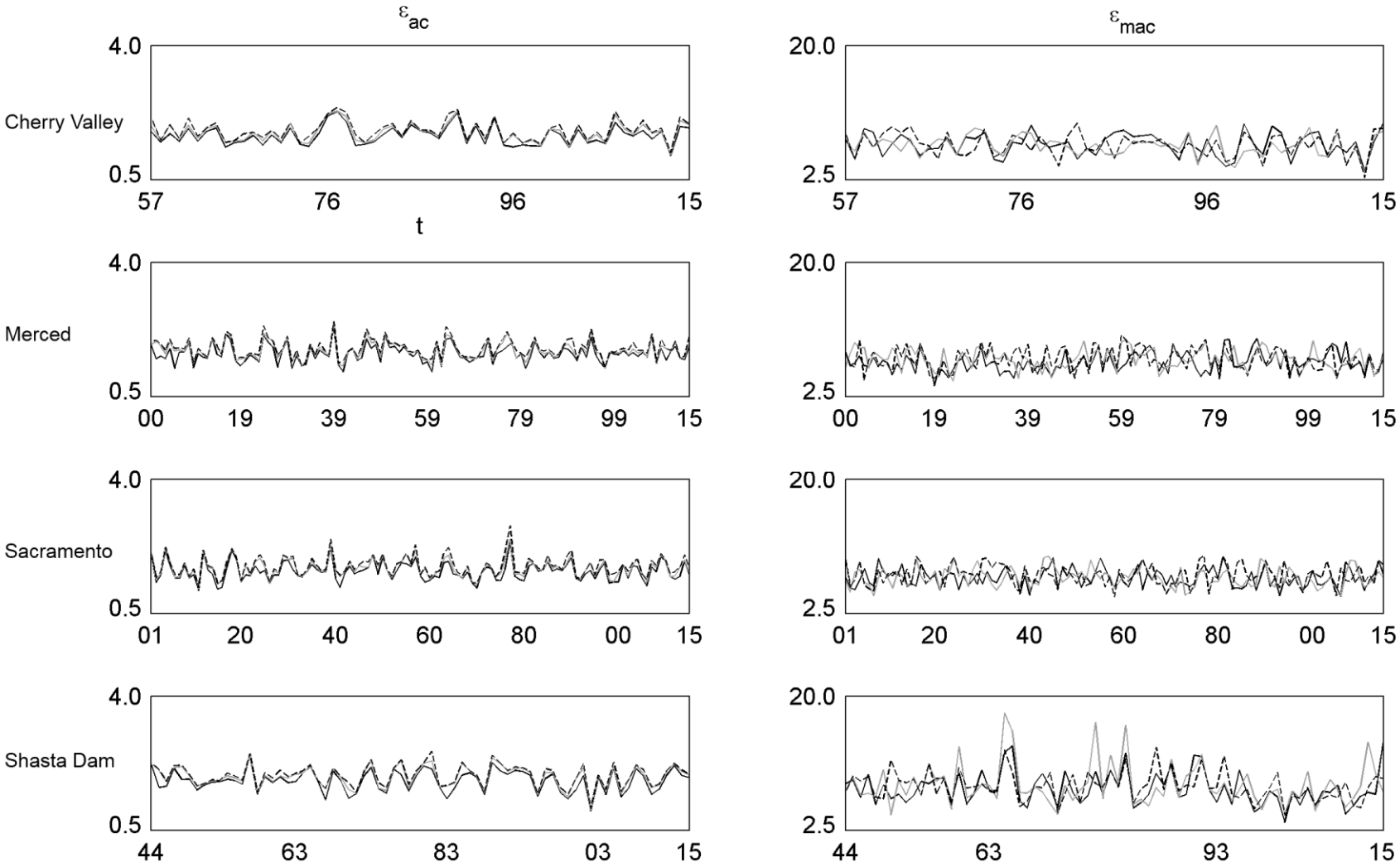
ε_{ac}
ε_{mac}
Cherry Valley
Merced
Sacramento
Shasta Dam
4.0
0.5
20.0
2.5
57
76
96
15
t
00
19
39
59
79
99
01
20
40
60
80
44
63
83
03
93

Fig. 5:

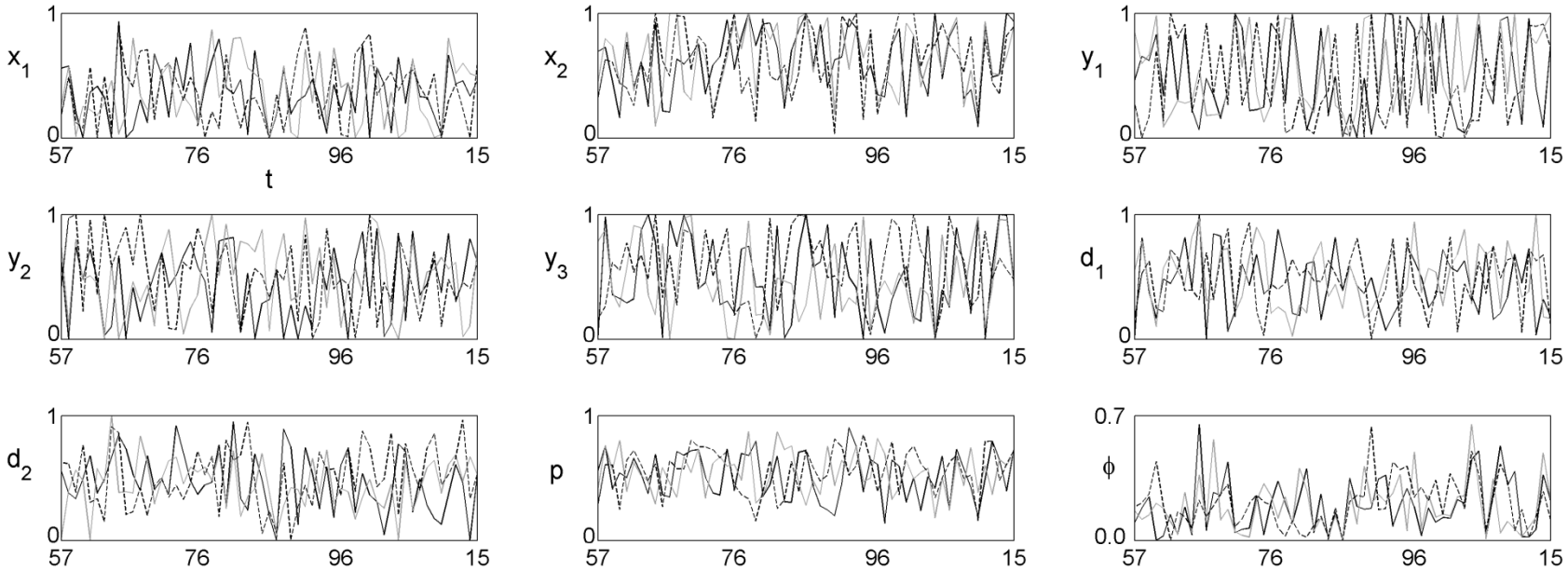

Fig. 6:

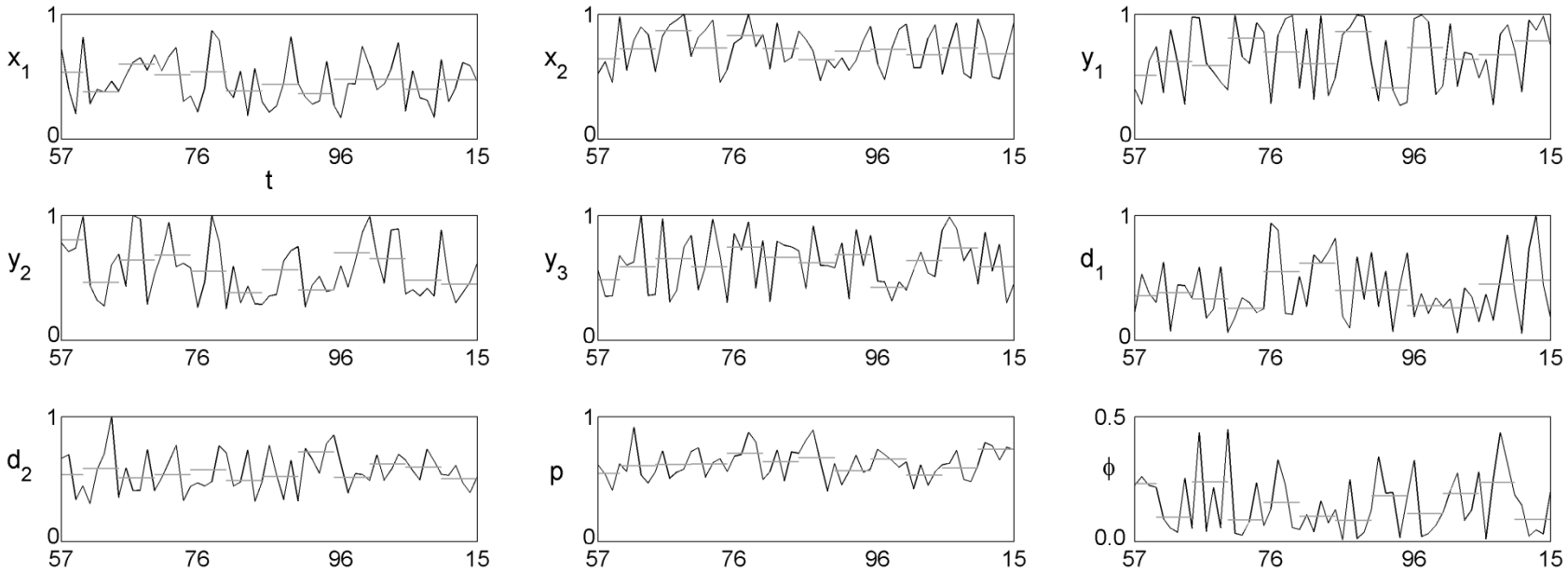

Fig. 7:

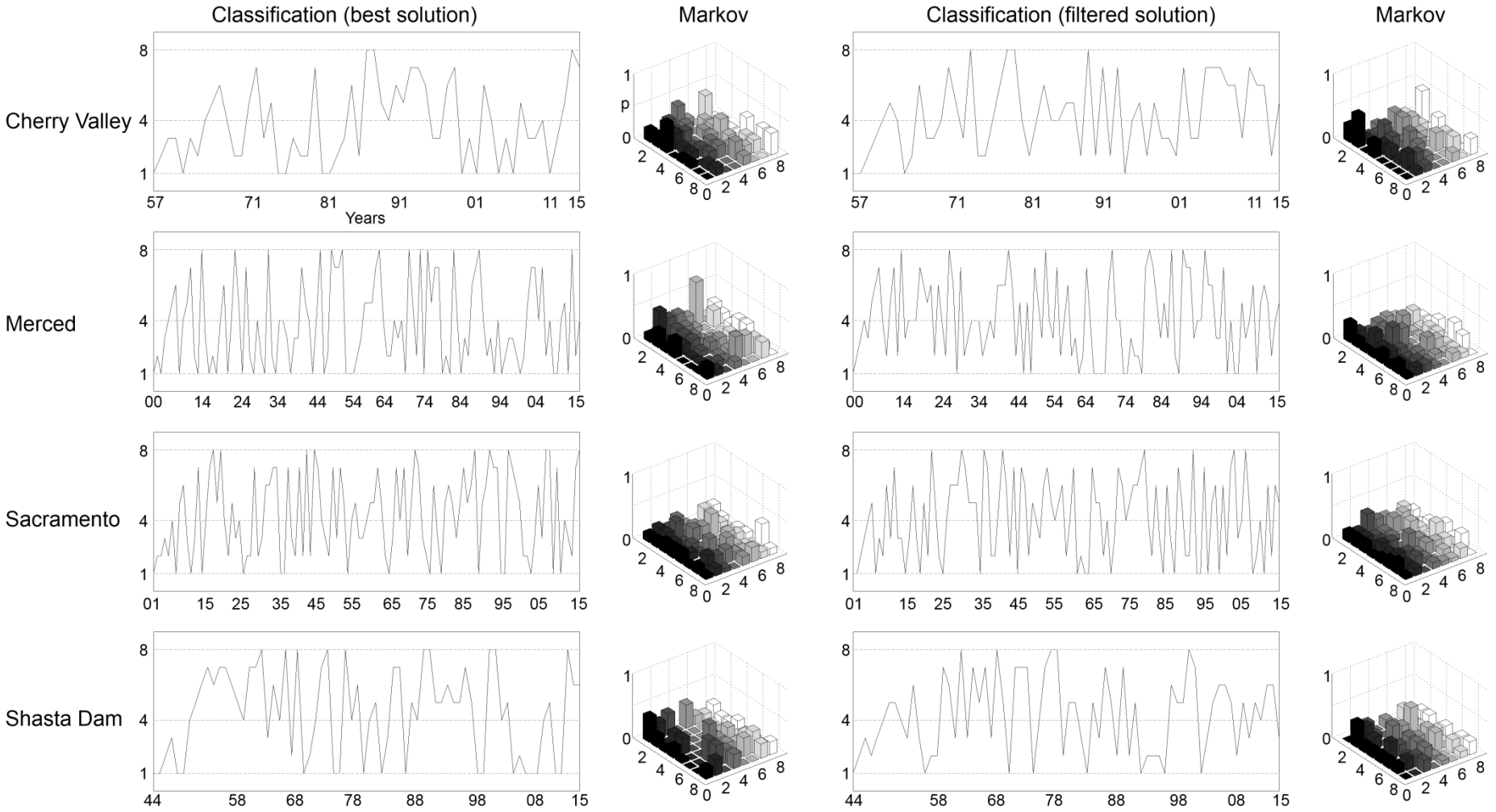
Classification (best solution)
Markov
Classification (filtered solution)
Markov
Cherry Valley
Merced
Sacramento
Shasta Dam
Years
p

Fig. 8:

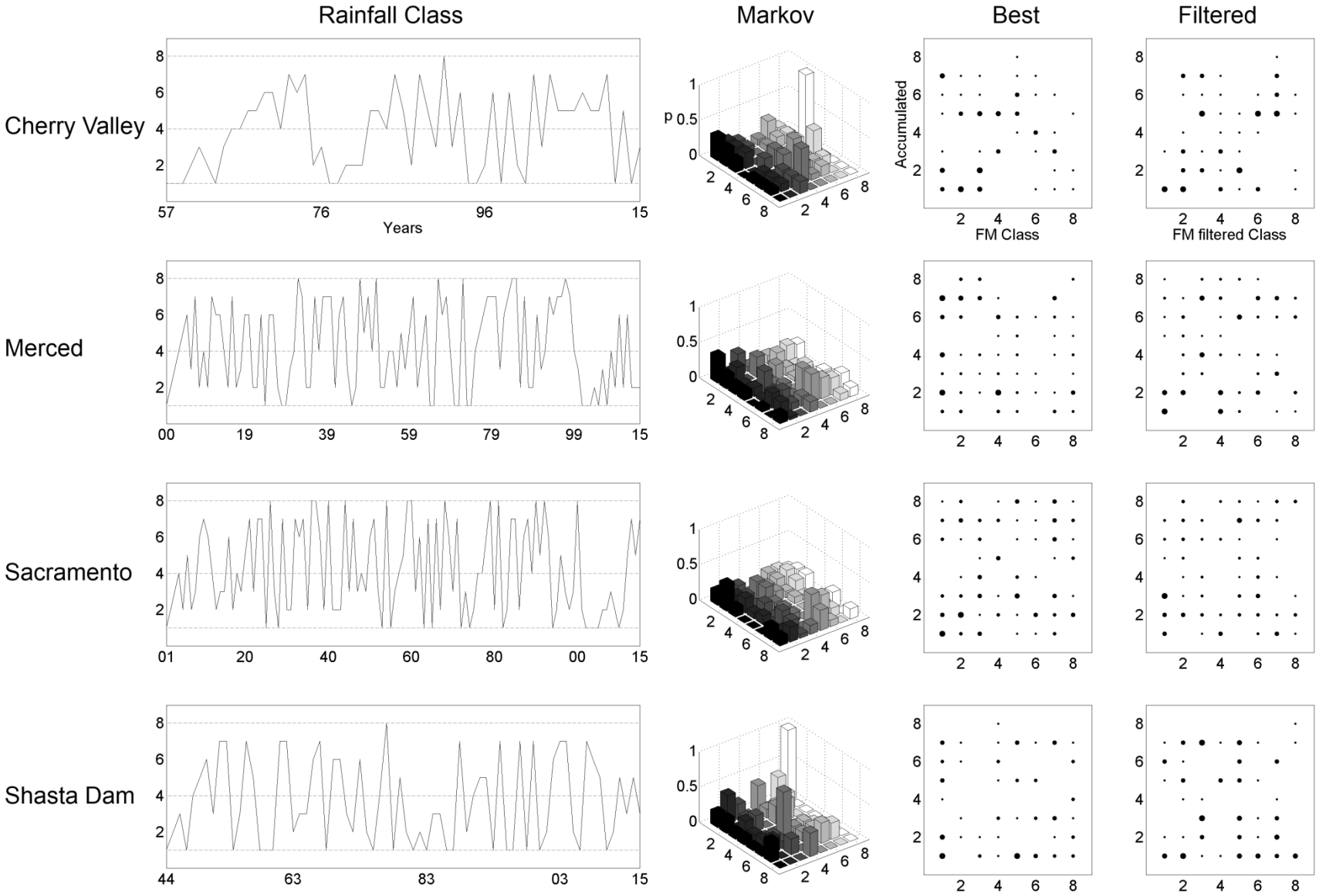
Rainfall Class
Markov
Best
Filtered
Cherry Valley
Merced
Sacramento
Shasta Dam
Years
p
Accumulated
FM Class
FM filtered Class

Fig. 9:

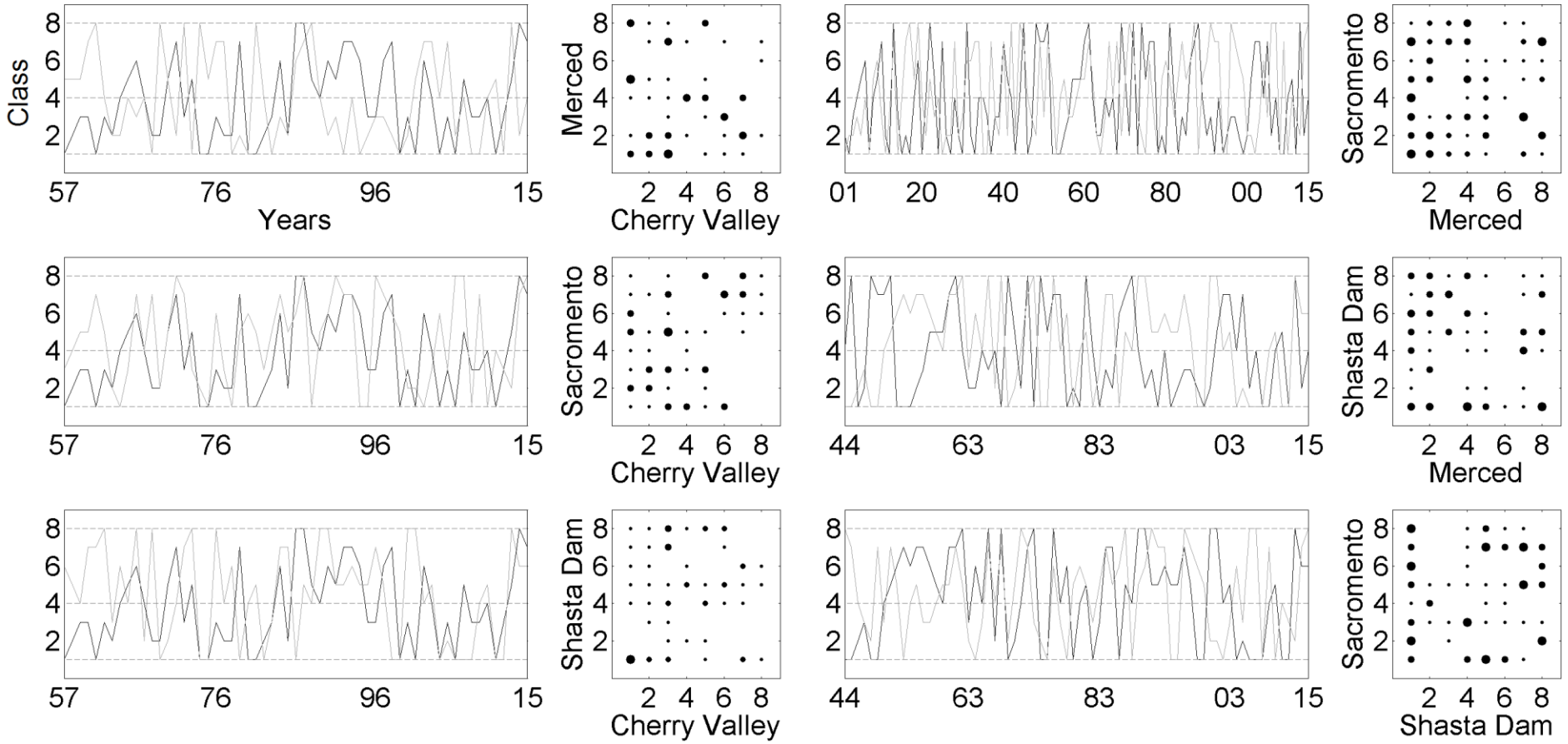
Class
Years
Merced
Cherry Valley
Sacromento
Merced
Sacromento
Cherry Valley
Shasta Dam
Merced
Shasta Dam
Cherry Valley
Sacromento
Shasta Dam

Fig. 10:

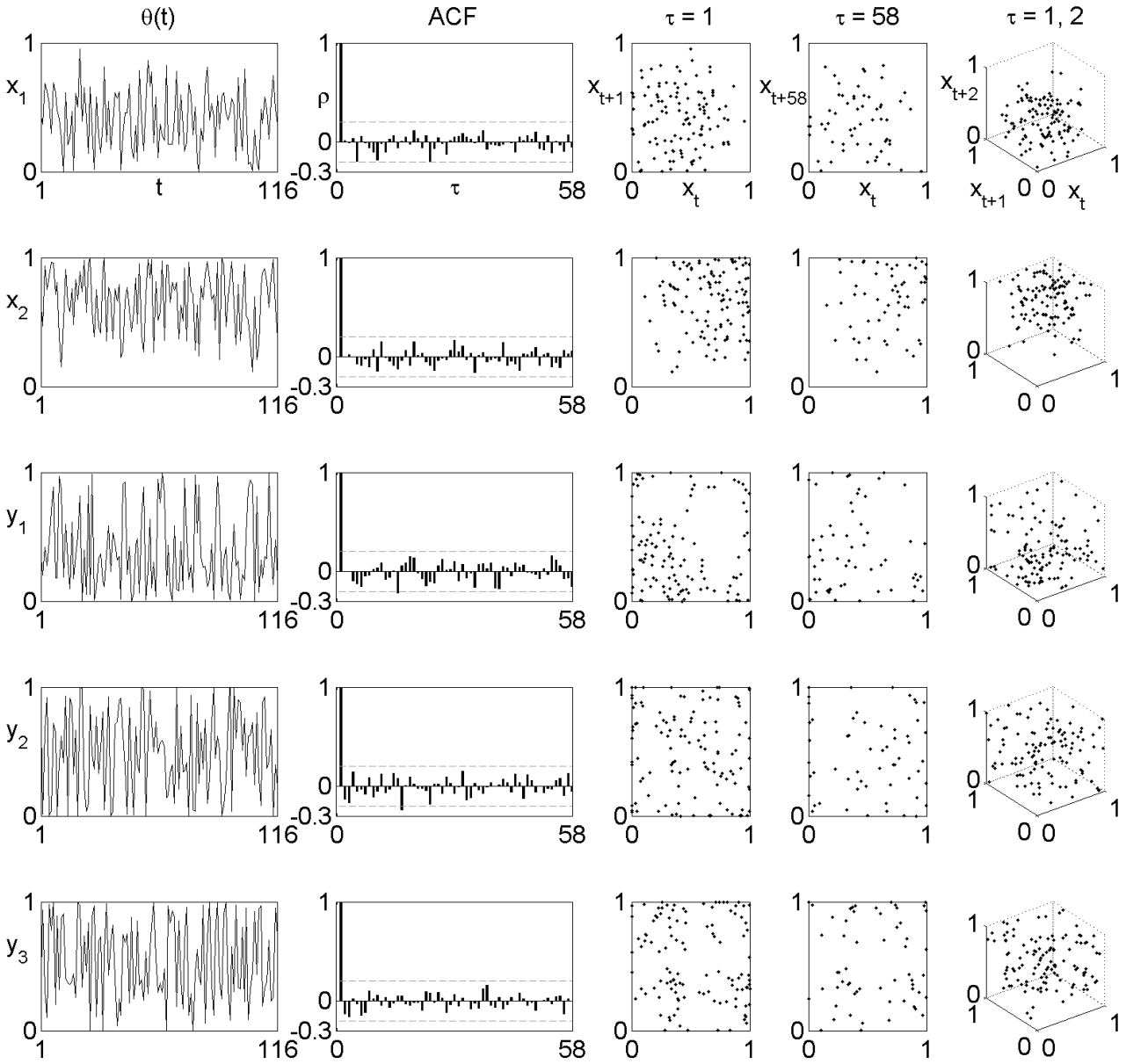

Fig. 11a:

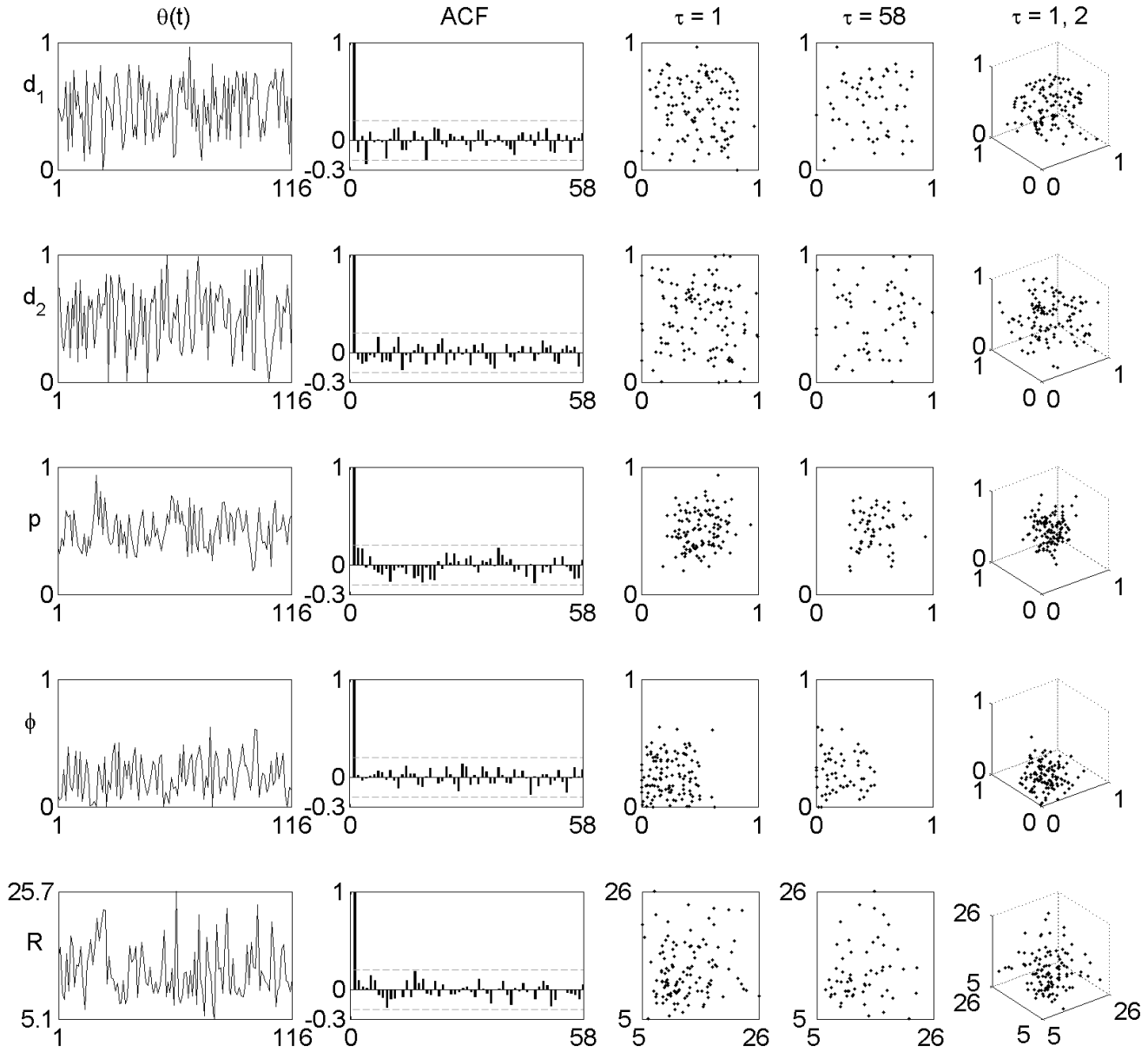

Fig. 11b: