

Chapter 3

Mathematical Challenges in Measuring Variability Patterns for Precipitation Analysis



Maria Emelianenko and Viviana Maggioni

Abstract This chapter addresses some of the mathematical challenges associated with current experimental and computational methods to analyze spatiotemporal precipitation patterns. After a brief overview of the various methods to measure precipitation from *in situ* observations, satellite platforms, and via model simulations, the chapter focuses on the statistical assumptions underlying the most common spatiotemporal and pattern-recognition techniques: stationarity, isotropy, and ergodicity. As the variability of Earth's climate increases and the volume of observational data keeps growing, these assumptions may no longer be satisfied, and new mathematical methodologies may be required. The chapter discusses spatiotemporal decorrelation measures, a nonstationary intensity-duration-function, and 2-dimension reduction methodologies to address these challenges.

Keywords Centroidal Voronoi tessellation · Data reduction · Decorrelation · Empirical orthogonality functions · Ergodicity · Isotropy · Precipitation patterns · Stationarity · Statistical assumptions

3.1 Introduction

Precipitation occurs when a portion of the atmosphere becomes saturated with water vapor, so that the water condenses and precipitates by gravity. Precipitation is a critical component of the water and energy cycles, providing moisture for processes such as runoff, biogeochemical cycling, evapotranspiration, groundwater recharge, carbon exchange, and heat fluxes. The main forms of precipitation include rain,

M. Emelianenko (✉)

Department of Mathematical Sciences, George Mason University, Fairfax, VA, USA
e-mail: memelian@gmu.edu

V. Maggioni

Sid and Reva Dewberry Department of Civil, Environmental, and Infrastructure Engineering,
George Mason University, Fairfax, VA, USA
e-mail: vmaggion@gmu.edu

sleet, snow, and hail, but this chapter discusses liquid precipitation only, and the term “precipitation” is used here as a synonym for “rain.”

Precipitation is highly variable, both in space and time. This variability affects the dynamics of many hydrological processes at and near ground level. Information on precipitation characteristics and precipitation patterns is therefore critical for understanding these complex hydrological processes, as well as for monitoring and predicting extreme events such as floods and droughts [63]. Access to high-resolution high-quality rainfall data and information about spatiotemporal precipitation patterns can benefit applications at all levels; examples are hazard mitigation, agricultural planning, and water resources management at the regional level [33, 37, 46]; controlling stormwater runoff, managing reservoirs and detention ponds, cleaning streams and channels, and closing roads or parking lots during extreme precipitation events at the local level.

However, estimating precipitation is challenging because it involves many factors, including the natural temporal and spatial variability of precipitation, measurement errors, and sampling uncertainties, especially at fine temporal and spatial scales. The spatiotemporal variability of precipitation patterns is changing heterogeneously due to climate change, and those changes have an impact on the tools used to make decisions and optimize water management. This chapter focuses on some of the mathematical and statistical issues related to variability of precipitation patterns.

Outline of the Chapter In Sect. 3.2, we briefly discuss various methods to measure precipitation, whether *in situ*, remotely, or by using model simulations. In Sect. 3.3, we review the strengths and limitations of current methods to analyze spatiotemporal precipitation patterns. We discuss decorrelation measures in Sect. 3.4 and dimension reduction strategies in Sect. 3.5. In Sect. 3.6, we present some concluding remarks.

3.2 Estimating Precipitation

Precipitation can be estimated through three main approaches: (1) *in situ* measurements, (2) remote sensing (including weather radars and satellite sensors), and (3) model simulations [52].

3.2.1 In Situ Measurements

The only direct method to measure precipitation is through rain gauges (also known as *pluviometers*) which collect and measure the amount of rain over a period of time. There are several types of rain gauges; the most common one is the tipping bucket. Precipitation is collected in a funnel and channeled into a small container. After a set amount of precipitation is collected, the device tips, dumping the water,

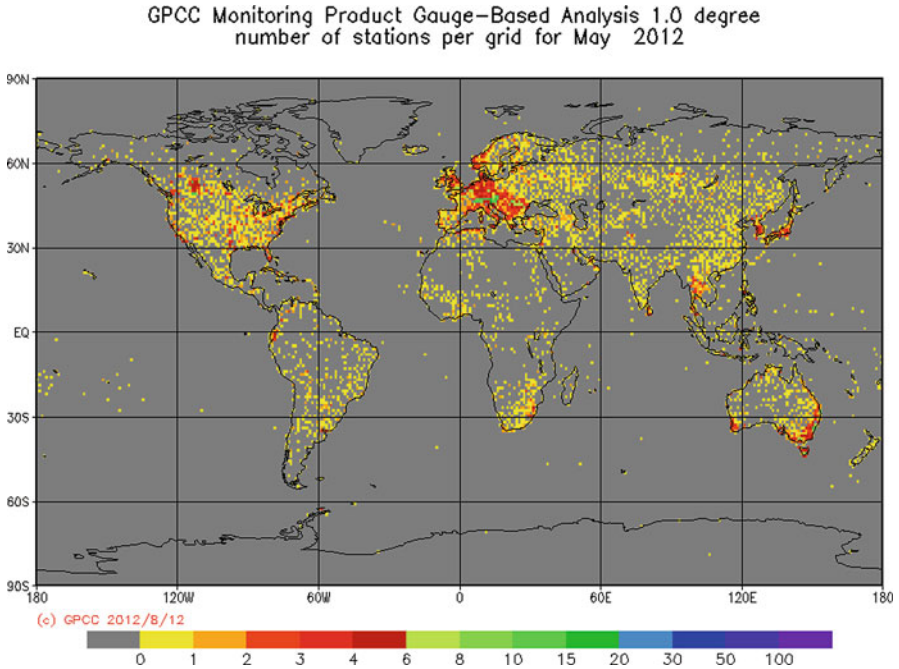


Fig. 3.1 Number of stations used by the global precipitation climatology center (GPCC) for May 2012. Figure produced with “GPCC Visualizer” [61], courtesy of National Center for Atmospheric Research Staff (Eds), last modified 29 Jun 2018. Retrieved from <https://climatedataguide.ucar.edu/climate-data/gpcc-global-precipitation-climatology-centre>

and sending a signal that is automatically recorded by a data logger. Rain gauges may underestimate rainfall because of wind effects and evaporation.

Rain-gauge networks can provide measurements with high temporal resolution, but obtaining a spatially representative measurement requires a sufficiently large number of samples to account for variability of terrain, microclimate, and vegetation. Moreover, *in situ* measurements are localized and limited in spatial and temporal coverage [43]. One of the main applications of ground-monitoring networks is for assessing flood risk through early warning systems [3]. However, their usefulness is limited by the spatial representativeness of local measurements and the network density, especially over important climatic regions like the tropical rain forests and mountainous areas (Fig. 3.1).

A ground-based alternative to monitor precipitation is weather radar which provides spatially distributed information on rainfall (Fig. 3.2). Weather radars send directional pulses of microwave radiation connected to a parabolic antenna. Wavelengths are of the order of a few centimeters, which is about ten times larger than the average diameter of water droplets and ice particles. These particles bounce part of the energy in each pulse back to the radar (reflectivity). As they move farther from the source, the pulses spread out, crossing a larger volume

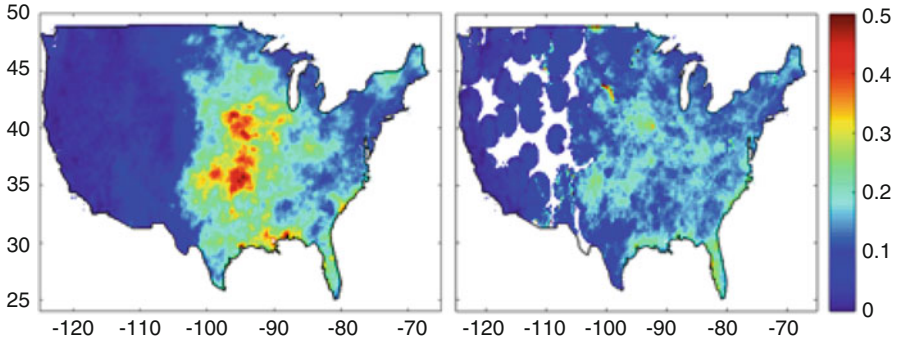


Fig. 3.2 Average precipitation maps for Summer 2015 from a satellite precipitation product that combines infrared and microwave observations (left) and ground-based weather radars (right) across the continental USA

of air, and therefore their resolution decreases with distance. Doppler radars are common and observe not only rainfall rates, but also the motion of rain droplets. However, weather radar estimates are affected by uncertainties associated with rain-path attenuation, the lack of uniqueness in the reflectivity-to-rain-rate relationship, radar calibration and contamination by ground return problems, sub-resolution precipitation variability, and complex terrain effects [10, 46, 51]. Moreover, ground-based monitoring systems, like rain gauges and weather radars, require substantial financial and technological investments to support their operation and maintenance on a continuous basis over a long period.

3.2.2 Remote Sensing

A way to overcome these issues is the use of satellite precipitation products, which are nowadays available on a global scale at increasing spatial and temporal resolution. Precipitation estimates can be derived from a range of observations from many different on-board satellite sensors. Specifically, rainfall can be inferred from visible imagery, since thick clouds, which are more likely to be associated with rainfall, tend to be brighter than the surface of the Earth. Infrared (IR) images are more suitable because they are available night and day, and heavier convective rainfall tends to be associated with larger taller clouds with colder cloud tops. Another method uses passive microwave (PMW) sensors, since emissions from rain droplets lead to an increase in PMW radiation. And scattering caused by precipitating ice particles leads to a decrease in PMW radiation.

Several techniques have been developed to exploit the synergy between IR radiances and PMW observations (Fig. 3.2). Examples include the TRMM multi-satellite precipitation analysis (TMPA) [41], the climate prediction center morphing (CMORPH) technique [42], and, most recently, the integrated multi-satellite retrievals for GPM (global precipitation measurement) (IMERG) [40], which

merges precipitation estimates from PMW and IR sensors and monthly surface rain-gauge data to provide half-hourly precipitation estimates on a 0.1° grid over the 60°N-S domain. In other cases, artificial neural networks (ANNs) are used to derive precipitation estimates by combining information from multichannel and multisensor observations, like the precipitation estimation from remotely sensed information using ANNs (PERSIANN) [39]. The availability of these products has opened new venues to support water management and hydrologic applications globally. Especially in poorly gauged regions, satellite precipitation products may be the only input data to allow flow predictions downstream with enough lead time to implement management and response actions [64].

Satellite observations can be affected by detection errors, as well as systematic and random errors. Detection errors include missed events (when satellite observes no rain, but there is rain at the ground) and false alarms (when the satellite sees rain, but it does not rain). In the case of successful detection, the estimated rain rate may still be affected by systematic and/or random errors, which depend on the accuracy of the remote sensor (retrieval error) and the lack of continuity in the coverage by low earth-orbiting satellites (sampling error, [7]). Typical sources of retrieval error are due to sub-pixel inhomogeneity in the rainfall field [48], whereas sampling errors are related to the satellite orbit, swath width, and space-time characteristics of rainfall [14]. The performance of satellite precipitation products is also influenced by factors such as seasonal precipitation patterns, storm type, and background surface [31, 33, 57, 66]. Detection, systematic, and random errors all play a pivotal role in hydrological applications (e.g., flood forecasting) and water resource management.

High-mountain regions are among the most challenging environments for precipitation measurements (whether from the ground or from satellites) due to extreme topography and large weather and climate variability. These regions are typically characterized by a lack of *in situ* data, but are also prone to flash floods whose consequences can be devastating.

3.2.3 Model Simulation

Numerical weather prediction (NWP) models provide a third option for estimating precipitation at global and regional scales. NWP models estimate the state of the atmosphere (including air density, pressure, temperature, and velocity) at a given time and location using fluid dynamics and thermodynamics equations. These models are rather accurate for large-scale organized systems. However, their performance deteriorates in the case of more localized events that are not governed by large-scale flows and whose spatial and temporal variability cannot be explicitly captured by the model resolution. NWP model forecasts can be improved by more accurate parameterizations and by constraining model analyses with moisture-, cloud-, and precipitation-related observations through data assimilation systems, such as 4D-Var and ensemble Kalman filter methods [6, 50].

3.3 Assessment of Spatial and Temporal Patterns

Changes in spatiotemporal precipitation patterns have a direct impact on the spatial and temporal distribution of water resources and the occurrence of natural hazards [69]. The hydrological community has adopted a set of geostatistical tools for measuring spatiotemporal correlations in rainfall [5, 65]. As mentioned in multiple sources [65], some of the notions come with tacit assumptions that often lead to their misuse in practice. While the complete list is beyond the scope of this chapter, we will review some of the key elements of such analyses and point out some of their strengths and limitations.

3.3.1 Definitions

Assume that rainfall corresponds to a stochastic process $\eta(\mathbf{u}, t)$, where $\mathbf{u} \equiv (x, y)$ is a vector representing the spatial coordinates in a given area, t stands for time, and $\eta(\cdot)$ is a measure of the intensity of the rainfall. In a practical setting, one typically considers an observation map in the form of a *snapshot matrix* $A = A_{i,j} \in \mathbb{R}^{N \times n}$, where $A_{i,j} = \eta(\mathbf{u}_i, t_j)$ is the rainfall observed at location i at time t_j ($i = 1, \dots, N; j = 1, \dots, n$). Typically, for hydrological applications, $N \gg n$. Different statistical characterizations of the process are used, depending on the purpose of the study.

Spatial Variability If the focus is on spatial correlations, time series may be integrated over time at each location. Following [5], we define the *depth*, Z , of the rainfall over a time interval of length T at the location \mathbf{u} , by the integral

$$Z(\mathbf{u}) = \int_t^{t+T} \eta(\mathbf{u}, \tau) d\tau, \quad (3.3.1)$$

and its *intensity*, X , by the integral

$$X(\mathbf{u}) = \frac{1}{T} \int_t^{t+T} \eta(\mathbf{u}, \tau) d\tau. \quad (3.3.2)$$

The *mean*, m , of the rainfall at \mathbf{u} is

$$m(\mathbf{u}) = E[Z(\mathbf{u})], \quad (3.3.3)$$

where $E[\cdot]$ denotes the expected value over all realizations of the process—that is, over all different measurements at a certain location. After subtracting the mean, we obtain the *detrended* or *centered process*, Y ,

$$Y(\mathbf{u}) = Z(\mathbf{u}) - m(\mathbf{u}). \quad (3.3.4)$$

The *covariance function* is defined in terms of the detrended process.

$$\text{Cov}(\mathbf{u}_1, \mathbf{u}_2) = E[Y(\mathbf{u}_1)Y(\mathbf{u}_2)] = E[Z(\mathbf{u}_1)Z(\mathbf{u}_2)] - m(\mathbf{u}_1)m(\mathbf{u}_2). \quad (3.3.5)$$

Similarly, the *covariance matrix* is $\Sigma = E[Y^T Y] = E[Z^T Z] - m^T m$, where the (i, j) th entry represents the covariance between the depth of rainfall at the i th and j th spatial location. The *correlation function* is a normalized version of the covariance function,

$$R(\mathbf{u}_1, \mathbf{u}_2) = \frac{\text{Cov}(\mathbf{u}_1, \mathbf{u}_2)}{\sigma(\mathbf{u}_1)\sigma(\mathbf{u}_2)}, \quad (3.3.6)$$

where σ is the *standard deviation*,

$$\sigma(\mathbf{u}_i) = \left\{ E[Z(\mathbf{u}_i) - m(\mathbf{u}_i)]^2 \right\}^{1/2} = E[Y^2(\mathbf{u}_i)]^{1/2}. \quad (3.3.7)$$

A concept that is commonly used in hydrology is that of a *semivariogram function*,

$$\Gamma(\mathbf{u}_1, \mathbf{u}_2) = \frac{1}{2} E\{[Y(\mathbf{u}_1) - Y(\mathbf{u}_2)]^2\}. \quad (3.3.8)$$

The covariance and semivariogram functions are symmetric,

$$\text{Cov}(\mathbf{u}_1, \mathbf{u}_2) = \text{Cov}(\mathbf{u}_2, \mathbf{u}_1), \quad \Gamma(\mathbf{u}_1, \mathbf{u}_2) = \Gamma(\mathbf{u}_2, \mathbf{u}_1).$$

Note that the covariance is a measure of the association between the two variables $Z(\mathbf{u}_1)$ and $Z(\mathbf{u}_2)$, while the semivariogram function is a measure of their dissociation.

The above definitions of the various statistical quantities work for any time interval $[t, t + T]$. For instance, one may decide to study daily, monthly, or yearly averages, as appropriate. The longer the period over which the data are integrated, the more one may expect temporal variations to be suppressed.

Temporal Variability If temporal variability is of interest, it is important to keep as much of the original temporal information as possible when computing variograms and correlations. So, while integrated data are attractive from the processing point of view, in climate research one always defines statistical characteristics using the original map $\eta(\mathbf{u}, t)$. Thus, the *mean* is defined as a time average,

$$m(\mathbf{u}) = \langle \eta(\mathbf{u}, t) \rangle, \quad (3.3.9)$$

and the *centered data* (also called *anomalies*) are given by

$$Y = Y(\mathbf{u}, t) = \eta(\mathbf{u}, t) - m(\mathbf{u}). \quad (3.3.10)$$

The correlation function, standard deviation, and semivariogram are defined in terms of anomalies as in (3.3.6)–(3.3.8).

The statistical quantities defined above all have discrete analogs. For example, $m_i = \frac{1}{n} \sum_{j=1}^n A_{i,j}$ is the time average of a certain realization of the rainfall field at location i ; the anomalies $y_{i,j} = a_{i,j} - m_i$ are the entries of the *anomaly matrix* $Y = y_{i,j}$, and the corresponding *covariance matrix* is $\Sigma = Y^T Y \in \mathbb{R}^{n \times n}$. The eigenvectors of this covariance matrix Σ are the *empirical orthogonal functions*, which we will discuss in Sect. 3.5. Note that, while the size of the matrix Σ is normally much smaller than that of the original detrended matrix Y , the condition of the covariance matrix is given by $\text{cond}(\Sigma) = \text{cond}(Y)^2$, so it is not surprising that ill-conditioning in the original data presents an issue for many geospatial applications [1].

The correlation function, standard deviation, and semivariogram function are collectively referred to as *variograms* of the process; they represent the structure of the spatial dependence of the process and variability in the reference area A .

3.3.2 Statistical Assumptions in Hydrological Analyses

The effective use of the statistical quantities defined in Sect. 3.3.1 depends critically on a number of regularity assumptions for the underlying stochastic process. In hydrological analyses, the rainfall process is commonly assumed to be second-order stationary, isotropic, and ergodic. We briefly recall the relevant definitions.

Stationarity The field $Z(\mathbf{u})$ is *first-order stationary* if

$$E[Z(\mathbf{u})] = m = \text{constant}, \quad \forall \mathbf{u} \in A, \quad (3.3.11)$$

and *second-order stationary* or *weakly stationary* if it is first-order stationary and, in addition,

$$\text{Var}[Z(\mathbf{u})] = \sigma^2 = \text{constant}, \quad \forall \mathbf{u} \in A, \quad (3.3.12)$$

$$\text{Cov}(\mathbf{u}_1, \mathbf{u}_2) = \text{Cov}(\mathbf{u}_1 - \mathbf{u}_2), \quad \forall \mathbf{u}_2, \mathbf{u}_2 \in A. \quad (3.3.13)$$

For a second-order stationary process, $\Gamma(\mathbf{u}_1, \mathbf{u}_2) = \Gamma(\mathbf{u}_1 - \mathbf{u}_2) = \Gamma(\mathbf{h})$, where $\mathbf{h} = \mathbf{u}_1 - \mathbf{u}_2$, and $\text{Cov}(\mathbf{u}_1, \mathbf{u}_2) = \text{Cov}(\mathbf{h}) = E[Z(\mathbf{u} + \mathbf{h})Z(\mathbf{u})] - m^2$ for all $\mathbf{u}_2, \mathbf{u}_2 \in A$. Furthermore,

$$\begin{aligned} \Gamma(\mathbf{h}) &= \frac{1}{2} E[Z(\mathbf{u} - \mathbf{h}) - Z(\mathbf{u})]^2 \\ &= \frac{1}{2} E[Z(\mathbf{u} + \mathbf{h})^2] - E[Z(\mathbf{u})Z(\mathbf{u} + \mathbf{h})] + \frac{1}{2} E[Z(\mathbf{u})]^2 \\ &= \text{Cov}(\mathbf{0}) - \text{Cov}(\mathbf{h}). \end{aligned}$$

Isotropy The field $Z(\mathbf{u})$ is *isotropic* if spatial variability, measured by the covariance or semivariogram function, does not depend on the direction of the vector $\mathbf{h} = \mathbf{u}_1 - \mathbf{u}_2$,

$$\text{Cov}(\mathbf{h}) = \text{Cov}(|\mathbf{h}|) = \text{Cov}(h), \quad (3.3.14)$$

$$\Gamma(\mathbf{h}) = \Gamma(|\mathbf{h}|) = \Gamma(h), \quad (3.3.15)$$

where $h = |\mathbf{h}|$ is the distance between two locations \mathbf{u}_1 and \mathbf{u}_2 .

Ergodicity A dynamic process is said to be *ergodic* if time averages coincide with sample averages,

$$E(\eta(\mathbf{u}, t)) = \langle \eta(\mathbf{u}, t) \rangle. \quad (3.3.16)$$

In the case of an ergodic process, the estimates of the moments obtained on the basis of the available realizations converge in probability to the theoretical moments as the sample size increases. The process will tend to a limiting distribution regardless of the initial state [44]. In practice, this enables one to obtain estimates even from a single realization of the process.

3.3.3 What If the Assumptions Are Not Satisfied?

Figure 3.3 shows a realization of the precipitation process. The data (blue dots) represent the annual maximum precipitation (in inches) recorded at Beardstown in the State of Illinois (USA) during the period 1903–2000. Connecting the dots, we see that the maximum moves up and down without much regularity, but a linear regression analysis shows an overall upward trend (solid blue line). The mean (solid purple line) is approximately 2.3 in. over the first 55 years (1903–1958) and approximately 2.8 in. over the next 42 years (1958–2000), an increase of more than 20%. The variance (dotted red lines) also increases over time, albeit more slowly. The example shows that the rainfall process is clearly not stationary, so at least one of the hypotheses discussed in Sect. 3.3.2 is violated. Then the question is, what to do?

Nonstationarity The paper by Milly et al. [53] entitled *Stationarity Is Dead: Whither Water Management?*, which appeared in *Science* in 2008, served as a wake-up call for scientists in the field of hydrology and water resources engineering. Water management systems have been designed and operated for decades under the assumption of stationarity. However, this assumption has long been compromised by human disturbances in river basins such as dams, diversions, irrigation, land-use change, channel modifications, and drainage work. In addition, the timing and characteristics of precipitation—the most critical hydrological input—are also being modified by a changing climate, as demonstrated in Fig. 3.3. The hydrological

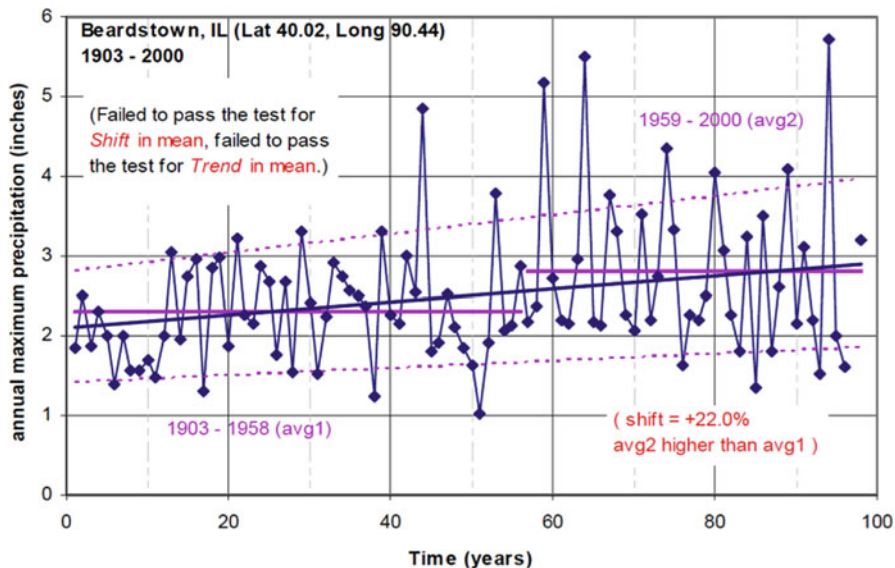


Fig. 3.3 Example of nonstationary changes in both the precipitation mean and variance [9]

literature on the analysis of long-term precipitation and runoff data is very thin [35]. Hodgkins and Dudley [36] show that most North American streams are experiencing earlier spring runoff, and DeGaetano et al. [24] show that nearly two-thirds of the trends in the 2-, 5-, and 10-years return-period rainfall amounts are positive. At the same time, the expected recurrence intervals have decreased by about 20%; for example, the 50 year storm based on 1950–1979 data is expected to occur once every 40 years based on 1950–2007 data.

Nonstationarity introduces multiple challenges for hydrological analysis, as recognized by several authors [35, 36, 49, 53]. Bonnin et al. [9] show trends in the intensity-duration-frequency (IDF) rainfall curves for the Ohio river basin. A particularly active area of research is the development of nonstationary rainfall IDFs, where theoretical advances in extreme value theory (EVT) turn out to be especially useful (see [16] and references therein). In particular, Cheng et al. [16] describes a new framework for estimating stationary and nonstationary return levels, return periods, and extreme points, which relies on Bayesian inference; the framework is implemented in NEVA software [15]. Ref. [16] offers a case study based on a global temperature dataset, comparing predictions based on stationary and nonstationary extreme value analysis. The study combines local processes (urbanization, local temperature change) and global processes (ENSO cycle, IOD, global temperature change) as time covariates for rainfall IDF, based on Hyderabad data [2]. The comparison shows that the IDF curves derived from the stationary models are underestimating the extreme events of all duration and for all return periods.

Nonisotropy Hydrological processes (soil moisture, streamflow, evapotranspiration) are extremely sensitive to small-scale temporal and spatial rainfall variability. Although ground-based weather radars have been particularly popular for forcing hydrological models that simulate a basin hydrological response, several authors have indicated that the interaction between the variability of precipitation (including spatial and temporal variations) and the resolution of a hydrological model is still poorly understood, especially when radar data are used in an urban environment [12, 22, 54]. If we assume a perfect hydrological model, and we force it with perfect rainfall input, we should expect that the accuracy of a streamflow simulation increases as the resolution of the model and the input increase. However, the finest available radar rainfall temporal resolution does not necessarily provide the best estimation of peak streamflow in a distributed hydrological model. This is the result of uncertainty and errors related to both the precipitation measurement techniques, as discussed in Sect. 3.2, and the model physics [4, 58].

The spatial resolution of precipitation data must be functionally coupled with the temporal resolution to fully reproduce the hydrological response of an urban catchment. For instance, Berne et al. [8] proposed the relation $\Delta s = \frac{3}{2}\sqrt{\Delta t}$ to couple the spatial scale (Δs , in km) with the temporal scale (Δt , in minutes) for rainfall processes in urban catchments.

More recently, Ochoa-Rodriguez et al. [55] fitted the variogram of the spatial structure of rainfall over a peak storm period with an exponential model. They concluded that the minimum required spatial resolution was one-half the characteristic length scale r_c of the storm, which they defined in terms of the variogram range $r[L]$, $r_c = (2\pi/3)^{1/2} r[L]$. A unique relationship linking the temporal and spatial resolutions of precipitation adequate for the reproduction of the hydrological response of a catchment basin is yet to be found.

Nonergodicity Most of the literature simply assumes without evidence that precipitation and hydrological processes in general are ergodic; for example, see [27, 45, 56]. However, a recent study [67] indicates that the assumption may not be fully justified. The author proposed an approach to assess the mean ergodicity of hydrological processes based on the autocorrelation function of a dataset. The approach was tested on monthly rainfall time series at three locations, two in China and one in the State of Michigan (USA). The results showed that, at all three locations, the ergodicity assumption was met only during a few months of the year. Therefore, statistical metrics computed on the basis of data collected during those months do not meet the ergodicity assumption (sample statistics) and cannot be used as proper approximations for the population statistics. Moreover, the ergodicity assumption was met in different months at different locations, so ergodicity cannot be transferred to a different region and/or period. More work is clearly needed to establish the limits of validity of the ergodicity assumption.

Scenarios where the ergodicity assumption is not met have been studied even less frequently than scenarios where the stationarity and isotropy assumptions are not met, partially because of the difficulty of testing it in the absence of large quantities of high-quality data spanning a reasonable period of time.

In statistical mechanics, one often uses nonergodic Monte Carlo simulations to create multiple realizations for estimating statistical information on the dynamic processes over the region in question [47]. In the geospatial sciences, this approach is often infeasible.

An attempt has been made to formulate nonergodic versions of covariograms for the case of preferentially sampled data. However, as argued in [23], these measures do not offer a clear advantage over standard ergodic statistics for studying spatial dependence or making spatial predictions. Developing appropriate data transformations is considered a more promising direction.

In the mathematical literature, much attention is currently being paid to fractional diffusion processes, which typically generate nonergodic behavior. Some recent work aims to develop a metric quantifying nonergodicity [62]. This direction may also be useful for hydrological applications.

3.4 Decorrelation Measures

Correlation functions are standard tools for measuring spatial and temporal dependencies in the rainfall fields [11, 17]. Figure 3.4 shows both the temporal and spatial correlation functions for a precipitation dataset for the State of Oklahoma (USA) during the period March–October, 2011.

In the case of spatial correlations, one computes the correlation of the two time series associated with any two measurement points (for example, two rain gauges or two pixels) as a function of their distance. A common approximation is the exponential model with the so-called *nugget effect* [19, 20],

$$\rho_g(d) = c_0 \exp\left[-(d/d_0)^{s_0}\right]. \quad (3.4.1)$$

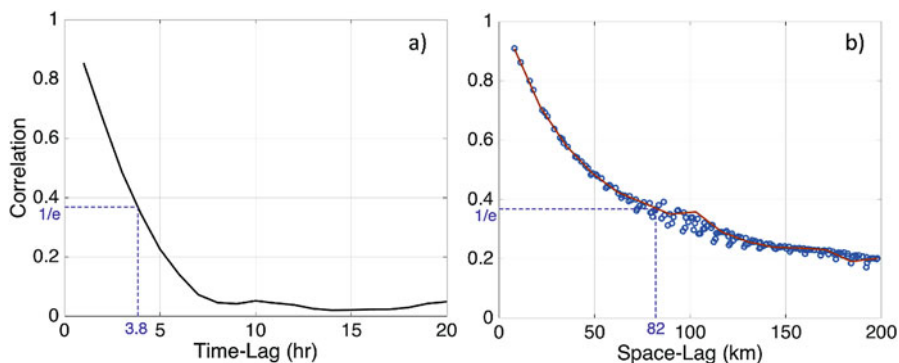


Fig. 3.4 (a) Temporal and (b) spatial correlations of CMORPH precipitation data for the State of Oklahoma (USA) at 8 km/1 h resolution during the period March–October, 2011

Here, c_0 is the nugget parameter, which corresponds to the correlation value for near-zero distances [21]; d is the separation distance, the distance between the two measurement points; d_0 is the scale parameter, which corresponds to the spatial decorrelation distance; and s_0 is the correlogram shape parameter, which controls the behavior of the model near the origin for small separation distances. The quantity $(1 - c_0)$ is the instant decorrelation due to random errors in the rainfall observations [18]. The separating distance at which the correlation is $1/e$ defines the correlation length for the (assumed) exponential variogram model.

In the case of temporal dependencies, autocorrelations are plotted as a function of the time lag. The lag-1 correlation is commonly adopted as a viable index of rainfall decorrelation in time [38, 60].

The exponential model (3.4.1) with the corresponding “ $1/e$ rule” is only one of several models for fitting semivariograms; linear, spherical, and Gaussian models are possible alternatives [5]. The choice of model has to be made based on the analysis of statistical data, and one should not adopt the decorrelation definition provided above as the default option. In fact, one may suspect that for regions with slowly decaying correlations (for example, flat regions with low spatial variability), the “ $1/e$ rule” might only work after a sufficient increase in the domain size. In other cases, the data might not support the exponential modeling assumption at all, and corresponding adjustments of the methodology would have to be performed. These modeling subtleties and tacit assumptions are sometimes a source of ambiguity in the literature, which may lead to erroneous conclusions.

3.5 Dimension Reduction Techniques

One of the many challenges of modeling and understanding spatiotemporal precipitation patterns is the large amount of data that needs to be processed. For example, in the relatively small-scale NASA Merra dataset, precipitation is given by monthly averages on a 50×91 grid representing a map of the contiguous USA at 50 km resolution over a period of 35 years, amounting to a total of 1,911,000 entries. However, much more detailed information at higher spatial (on the order of 100 m regionally and 1 km globally) and temporal (hourly) resolutions is required to assess the storage, movement, and quality of water at and near the land surface [68]. Higher-resolution data bring higher data volumes: for the previous example, there would be more than $3 \cdot 10^{12}$ entries for a map of the contiguous US at 1 km resolution and hourly intervals. Some form of data and dimension reduction is called for.

In a general sense, one may attempt to find a decomposition of the data (*signal*) of the form

$$\eta(\mathbf{u}, t) = \sum_{k=1}^N \alpha_k(t) \mathbf{p}_k(\mathbf{u}) + \text{noise}, \quad (3.5.1)$$

where the \mathbf{p}_k are *characteristic patterns* used to approximate the data (also called *guess patterns* or *predictors*), and the α_k are the *amplitudes* or *principal components*

of the corresponding patterns. The patterns \mathbf{p}_k are spatial structures that account for temporal variations of the rainfall data $\eta(\mathbf{u}, t)$. When plotted as functions of time, the amplitudes α_k convey information on how the patterns evolve in time.

Mathematically, finding the “best” patterns and principal components for a given dataset is a projection problem, “Find a subspace approximating a given set of data in an optimal (for example, least-squares) sense.” To solve this problem, various treatments have been proposed within the geophysical community by different groups and authors [65]. Here, we attempt to place these methods in perspective against methodologies developed independently in the mathematics community. While some techniques exist in both literatures (sometimes under different names), other methods have not yet penetrated the language barrier between the two disciplines.

EOF Method The method of empirical orthogonal functions (EOF) is one of the staple tools in geostatistics, which has received much attention in the hydrological literature. As mentioned in Sect. 3.3.1, EOFs are the eigenvectors of the covariance matrix $Y^T Y$. In the mathematical and statistical literature, the EOF method is referred to as *singular value decomposition* (SVD) or *principal component analysis* (PCA) and belongs to the class of *proper orthogonal decomposition* (POD) methods. In geospatial theory, it goes by the name *Karhunen–Loève analysis*.

Let Y denote the $N \times n$ matrix of detrended observations (also called “snapshot matrix” if $n < N$), whose columns are modified snapshots of rainfall data at a given time. If $C = \frac{1}{n} Y^T Y$ is the normalized correlation matrix, then a POD basis is comprised of the vectors

$$\phi_i = \frac{1}{\sqrt{n\lambda_i}} Y \chi_i, \quad i = 1, \dots, n,$$

where χ_i is the normalized eigenvector ($|\chi_i| = 1$) corresponding to the i th largest eigenvalue λ_i of C . The POD basis vectors are the first n left singular vectors of the snapshot matrix Y obtained by using the SVD decomposition of Y , $Y = U \Sigma V^T$, so $\phi_i = \mathbf{u}_i$ for $i = 1, \dots, n$.

Let $\{\psi_i\}_{i=1}^n$ be an arbitrary orthonormal basis for the span on the modified snapshot set $\{x_j\}_{j=1}^n$. Then the projection onto the d -dimensional subspace spanned by $\{\psi_i\}_{i=1}^n$ is

$$P_{\psi,d} x_j = \sum_{i=1}^d (\psi_i, x_j) \phi_i. \quad (3.5.2)$$

The POD basis is optimal in the sense that the approximation error

$$\varepsilon = \sum_{j=1}^n |x_j - P_{\psi,d} x_j|^2 \quad (3.5.3)$$

is minimized for $\psi_i = \phi_i$, $i = 1, \dots, d$.

While EOFs present an attractive tool for studying spatiotemporal variability patterns in precipitation data, care should be taken when interpreting the results of such analysis, as pointed out in [26]. In short, while it is tempting to find physical relevance for each of the EOF “modes,” the orthogonality condition built into this methodology often renders such interpretation useless. Rotated EOF technique is often used as a better alternative; however, a deeper analysis is normally needed to decipher the meaning of the EOF-based patterns.

CVT-Based Techniques In the mathematical community, an alternative dimension reduction technique based on *centroidal Voronoi tessellations* (CVTs) has recently gained popularity. While the list of applications is growing quickly, the method remains relatively under-explored in hydrological applications. The presentation below is based on [13].

The idea of the CVT technique is to find a fixed number of representative points (“generators”) to decompose the original high-dimensional space into a finite number of subspaces with relatively small loss of accuracy. The main ingredient of this method is the “density function,” usually denoted $\rho(\mathbf{x})$, which can be constant or a function of \mathbf{x} , depending on the application. For instance, ρ can be used to represent a variety of physical characteristics such as the local characteristic length scale [59], signal intensity [32], or the desired grid resolution [28]. In [25], ρ is used to represent spatial rainfall variability.

More precisely, given a snapshot matrix $X = \{\mathbf{x}_j\}_{j=1}^n \in W \subset \mathbb{R}^N$, the goal is to find a set of points $\{\mathbf{z}_i\}_{i=1}^k \in \mathbb{R}^N$, such that W can be decomposed in Voronoi regions, $W = \cup_{i=1}^k V_i$, with a minimum tessellation error, $\mathcal{E}[\{\mathbf{z}_i, V_i\}_{i=1}^k]$. A Voronoi region V_i is defined as

$$V_i = \{\mathbf{x} \in W : |\mathbf{x} - \mathbf{z}_i| \leq |\mathbf{x} - \mathbf{z}_j|, j = 1, \dots, k, j \neq i\}, \quad (3.5.4)$$

and the tessellation error is given by

$$\mathcal{E}[\{\mathbf{z}_i, V_i\}_{i=1}^k] = \sum_{i=1}^k \sum_{\mathbf{x} \in V_i} \rho(\mathbf{x}) |\mathbf{x} - \mathbf{z}_i|^2. \quad (3.5.5)$$

It can be shown that the tessellation error is minimal if and only if $\mathbf{z}_i = \mathbf{z}_i^*$ for $i = 1, \dots, k$, where \mathbf{z}_i^* is the mass centroid of the Voronoi region V_i [30]. At the minimum,

$$\sum_{\mathbf{x} \in V(\mathbf{z}^*)} \rho(\mathbf{x}) |\mathbf{x} - \mathbf{z}^*|^2 = \inf_{\mathbf{z} \in \mathbb{R}^N} \sum_{\mathbf{x} \in V(\mathbf{z})} \rho(\mathbf{x}) |\mathbf{x} - \mathbf{z}|^2. \quad (3.5.6)$$

Figure 3.5 gives two examples of CVTs for different types of densities.

In the discussion of the EOF method, we saw that the optimal basis was comprised of the set of vectors $\{\phi_i\}_{i=1}^d$. In the CVT method, the situation seems similar: the optimal basis is the set of generators $\{\mathbf{z}_i\}_{i=1}^k$. However, there are many

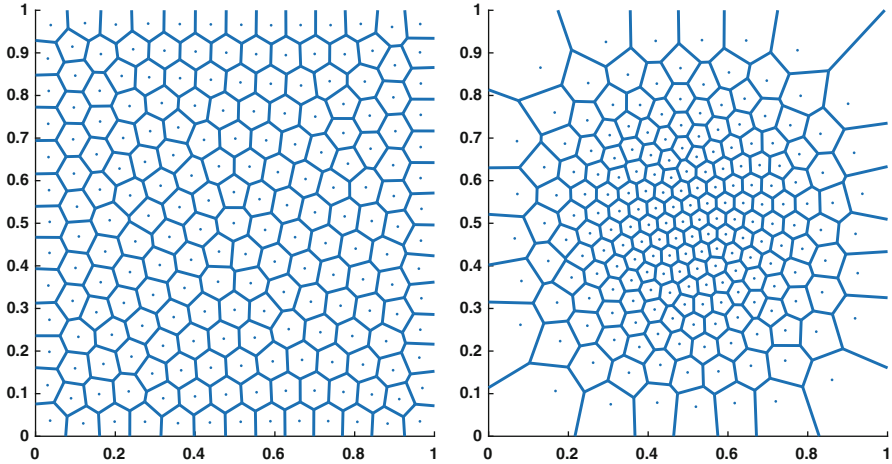


Fig. 3.5 Two CVT tessellations of the unit square, $W = [0, 1]^2$; (left) $\rho(x, y) = 1$, (right) $\rho(x, y) = \exp^{-20(x^2+y^2)}$

differences between the two approaches. POD minimizes the functional ε as in (3.5.3), while CVT minimizes the error \mathcal{E} given by (3.5.5). POD requires one to solve an $n \times n$ eigenvalue problem, where n is the number of snapshots, which is not very amenable to adaptive computations. While the CVT methodology is in general cheaper than POD, there are often numerical caveats associated with CVT computations. For an overview of CVT-related numerical techniques, we refer the reader to [29, 34]. Several case studies based on rainfall data highlighting the features of the CVT and POD approaches are presented in [25].

3.6 Concluding Remarks

In this chapter, we have presented an overview of experimental and computational methodologies and reviewed some of the mathematical challenges associated with the field of precipitation analysis. In particular, we focused our attention on the statistical assumptions underlying some of the commonly used pattern-recognition techniques. Because of the instability of the current climate, the validity of these assumptions should naturally fall under scrutiny. As abundant satellite and *in situ* observation data continue to pour in, one must reconsider the long-standing notions of stationarity, homogeneity, and ergodicity and be prepared to adopt new mathematical methodologies. In this chapter, we reviewed decorrelation measures, nonstationary extensions of intensity-duration-functions, and two types of dimension reduction methodologies with associated challenges. While some of these efforts are well under way, others are still in their infancy, and rigorous mathematical analysis is needed to address these challenges.

Acknowledgements This work was instigated at the Mason Modeling Days workshop held at George Mason University, generously supported by the National Science Foundation grant DMS-1056821. The authors are grateful to Paul Houser for stimulating discussions at the initial stages of this collaboration. ME also wishes to thank Hans Engler and Hans Kaper for their encouragement over the years, and for introducing this research group to the MPE community.

References

1. Ababou, R., Bagtzoglou, A.C., Wood, E.F.: On the condition number of covariance matrices in kriging, estimation, and simulation of random fields. *Math. Geol.* **26**(1), 99–133 (1994). <https://doi.org/10.1007/BF02065878>
2. Agilan, V., Umamahesh, N.V.: What are the best covariates for developing non-stationary rainfall intensity-duration-frequency relationship? *Adv. Water Resources* **101**, 11–22 (2017)
3. Artan, G., Gadain, H., Smith, J.L., et al.: Adequacy of satellite derived rainfall data for streamflow modeling. *Nat. Hazards* **43**, 167–185 (2007)
4. Atencia, A., Mediero, L., Llasat, M.C., et al.: Effect of radar rainfall time resolution on predictive capability of a distributed hydrological model. *Hydrol. Earth Syst. Sci.* **15**, 3809–3827 (2011)
5. Bacchi, B., Kottegoda, N.: Identification and calibration of spatial correlation patterns of rainfall. *J. Hydrol.* **165**, 311–348 (1995)
6. Bauer, P., Lopez, P., Benedetti, A., et al.: Implementation of 1D + 4D-Var assimilation of precipitation-affected microwave radiances at ECMWF. I: 1D-Var. *Q. J. Roy. Meteorol. Soc.* **132**(620), 2277–2306 (2006)
7. Bell, T.L., Kundu, P.K.: Dependence of satellite sampling error on monthly averaged rain rates: comparison of simple models and recent studies. *J. Climate* **13**(2), 449–462 (2000)
8. Berne, A., Delrieu, G., Creutin, J.D., et al.: Temporal and spatial resolution of rainfall measurements required for urban hydrology. *J. Hydrol.* **299**, 166–179 (2004)
9. Bonnin, G.M., Maitaria, K., Yekta, M.: Trends in rainfall exceedances in the observed record in selected areas of the United States 1. *J. Am. Water Resour. Assoc.* **47**(6), 1173–1182 (2011)
10. Borga, M., Anagnostou, E.N., Frank, E.: On the use of real-time radar rainfall estimates for flood prediction in mountainous basins. *J. Geophys. Res.* **105**(D2), 2269–2280 (2000)
11. Bras, R.L., Rodriguez-Iturbe, I.: *Random Functions and Hydrology*. Courier Corporation, Chelmsford (1985)
12. Brown, P.E., Diggle, P.J., Lord, M.E., et al.: Space-time calibration of radar rainfall data. *J. Royal Statistical Society: Series C (Applied Statistics)* **50**(2), 221–241 (2001)
13. Burkardt, J., Gunzburger, M., Lee, H.C.: Centroidal Voronoi tessellation-based reduced order modeling of complex systems. *SIAM J. Sci. Comput.* **28**(2), 459–484 (2006)
14. Chang, A.T., Chiu, L.S.: Nonsystematic errors of monthly oceanic rainfall derived from SSM/I. *Mon. Weather Rev.* **127**(7), 1630–1638 (1999)
15. Cheng, L.: Nonstationary Extreme Value Analysis (NEVA) software package, version 2.0. <http://amir.eng.uci.edu/neva.php> (2014)
16. Cheng, L., AghaKouchak, A., Gilleland, E., et al.: Non-stationary extreme value analysis in a changing climate. *Clim. Chang.* **127**(2), 353–369 (2014). <https://doi.org/10.1007/s10584-014-1254-5>
17. Chumchean, S., Sharma, A., Seed, A.: Radar rainfall error variance and its impact on radar rainfall calibration. *Phys. Chem. Earth, Parts A/B/C* **28**(1–3), 27–39 (2003)
18. Ciach, G.: Local random errors in tipping-bucket rain gauge measurements. *J. Atmos. Ocean. Technol.* **20**(5), 752–759 (2003)
19. Ciach, G.J., Krajewski, W.F.: On the estimation of radar rainfall error variance. *Adv. Water Resour.* **22**(6), 585–595 (1999)

20. Ciach, G.J., Krajewski, W.F.: Analysis and modeling of spatial correlation structure in small-scale rainfall in Central Oklahoma. *Adv. Water Resour.* **29**(10), 1450–1463 (2006)
21. Cressie, N.A.C.: *Statistics for Spatial Data*. John Wiley and Sons, Hoboken (1993)
22. Cristiano, E., Ten Veldhuis, M.C., van de Giesen, N.: Spatial and temporal variability of rainfall and their effects on hydrological response in urban areas – a review. *Hydrol. Earth Syst. Sci.* **21**, 3859–3878 (2017)
23. Curriero, F.C., Hohn, M.E., Liebhold, A.M.: A statistical evaluation of non-ergodic variogram estimators. *Environ. Ecol. Stat.* **9**, 89–110 (2002)
24. DeGaetano, A.T.: Time-dependent changes in extreme-precipitation return-period amounts in the continental united states. *J. Appl. Meteor. Climatol.* **48**, 2086–2099 (2009)
25. Di, Z., Maggioni, V., Mei Y., Vazquez M., Houser P., Emelianenko M., 2019, arXiv, arXiv:1908.10403
26. Dommenges, D., Latif, M.: A cautionary note on the interpretation of EOFs. *J. Climate* **15**, 216–225 (2001)
27. Duan, J., Goldys, B.: Ergodicity of stochastically forced large scale geophysical flows. *J. Math. Math. Sci.* **28**, 313–320 (2001)
28. Du, Q., Gunzburger, M.: Grid generation and optimization based on centroidal Voronoi tessellations. *Appl. Math. Comput.* **133**, 591–607 (2002)
29. Du, Q., Faber, V., Gunzburger, M.: Centroidal Voronoi tessellations: applications and algorithms. *SIAM Review* **41**, 637–676 (1999)
30. Du, Q., Emelianenko, M., Ju, L.: Convergence of the Lloyd algorithm for computing centroidal Voronoi tessellations. *SIAM J. Num. Anal.* **44**, 102–119 (2006)
31. Ebert, E.E., Janowiak, J.E., Kidd, C.: Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bull. Amer. Meteor. Soc.* **88**, 47–64 (2007)
32. Emelianenko, M.: Fast multilevel CVT-based adaptive data visualization algorithm. *Numer. Math. Theor. Meth. Appl.* **3**(2), 195–211 (2010)
33. Gottschalck, J., Meng, J., Rodell, M., et al.: Analysis of multiple precipitation products and preliminary assessment of their impact on global land data assimilation system land surface states. *J. Hydrometeorol.* **6**, 573–598 (2005)
34. Hateley, J.C., Wei, H., Chen, L.: Fast methods for computing centroidal Voronoi tessellations. *J. Sci. Comput.* **63**(1), 185–212 (2015)
35. Hirsch, R.M.: A perspective on nonstationarity and water management. *J. Amer. Water Resources Assoc. (JAWRA)* **47**(3), 436–446 (2011)
36. Hodgkins, G.A., Dudley, R.W.: Changes in the timing of winter–spring streamflows in eastern North America. *Geophys. Res. Lett.* **33**, 1913–2002 (2006)
37. Hossain, F., Anagnostou, E.N.: Assessment of current passive-microwave- and infrared-based satellite rainfall remote sensing for flood prediction. *J. Geophys. Res.* **109** (2004)
38. Hossain, F., Anagnostou, E.N.: A two-dimensional satellite rainfall error model. *IEEE Trans. Geosci. Remote Sens.* **44**(6), 1511–1522 (2006)
39. Hsu, K., Gao, X., Sorooshian, S., et al.: Precipitation estimation from remotely sensed information using artificial neural networks. *J. Appl. Meteor.* **36**, 1176–1190 (1997)
40. Huffman, G.J., Bolvin, D.T., Nelkin, E.J., et al.: The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.* **8**(1), 38–55 (2007)
41. Huffman, G.J., Bolvin, D., Braithwaite, D., et al.: Integrated Multi-satellite Retrievals for GPM (IMERG), version 4.4. NASA’s Precipitation Processing Center. Accessed 31 March 2015. <ftp://arthurhou.pps.eosdis.nasa.gov/gpmpdata/>
42. Joyce, R.J., Janowiak, J.E., Arkin, P.A., et al.: Cmorph: a method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeorol.* **5**, 487–503 (2004)
43. Kidd, C., Bauer, P., Turk, J., et al.: Intercomparison of high-resolution precipitation products over northwest Europe. *J. Hydrometeorol.* **13**, 67–83 (2012)
44. Kottegoda, N.T.: *Stochastic Water Resources Technology*. Palgrave, Macmillan (1980). <https://books.google.com/books?id=3SiuCwAAQBAJ>

45. Koutsoyiannis, D.: Stochastic simulation of hydrosystems. *Water Encyclopedia* **3**, 421–430 (2005)
46. Krajewski, W.F., Anderson, M.C., Eichinger, W.E., et al.: A remote sensing observatory for hydrologic sciences: a genesis for scaling to continental hydrology. *Water Resour. Res.* **42**(7), W07,301 (2006)
47. Krauth, W.: *Statistical Mechanics: Algorithms and Computations*. Oxford Master Series in Physics. Oxford University Press, UK (2006). <https://books.google.com/books?id=B3koVucDyKUC>
48. Kummerow, C.: Beamfilling errors in passive microwave rainfall retrievals. *J. Appl. Meteorol.* **37**(4), 356–370 (1998)
49. Lins, H.F.: A note on stationarity and non-stationarity. 14th Session of the Commission for Hydrology (2012)
50. Lorenc, A.C.: The potential of the ensemble Kalman filter for NWP—a comparison with 4D-Var. *Q. J. R. Meteorol. Soc.* **129**(595), 3183–3203 (2003)
51. Marzano, F.S., Picciotti, E., Vulpiani, G.: Rain field and reflectivity vertical profile reconstruction from c-band radar volumetric data. *IEEE Trans. Geosci. Remote Sens.* **42**(4), 1033–1046 (2004)
52. Michaelides, S., Levizzani, V., Anagnostou, E.N., et al.: Precipitation science: measurement, remote sensing, climatology and modeling. *Atmos. Res.* **94**, 512–533 (2009)
53. Milly, P.C.D., Betancourt, J., Falkenmark, M., et al.: Stationarity is dead: whither water management? *Science* **319**, 573–574 (2008)
54. Nikolopoulos, E., Borga, M., Zoccatelli, D., et al.: Catchment scale storm velocity: quantification, scale dependence and effect on flood response. *Hydrol. Sci. J.* **59**, 1363–1376 (2014)
55. Ochoa-Rodriguez, S., Wang, L., Gires, A., et al.: Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: a multi-catchment investigation. *J. Hydrol.* **531**, 389–407 (2015)
56. Oliveira, T.F., Cunha, F.R., Bobenrieth, R.F.M.: A stochastic analysis of a nonlinear flow response. *Probab. Eng. Mech.* **21**, 377–383 (2006)
57. Oliveira, R., Maggioni, V., Vila, D., et al.: Characteristics and diurnal cycle of GPM rainfall estimates over the Central Amazon Region. *Remote Sens.* **8**(7), 544 (2016)
58. Rafieenasab, A., Norouzi, A., Kim, S., et al.: Toward high-resolution flash flood prediction in large urban areas: analysis of sensitivity to spatiotemporal resolution of rainfall input and hydrologic modeling. *J. Hydrol.* **531**, 370–388 (2015)
59. Ringler, T., Ju, L., Gunzburger, M.: A multiresolution method for climate system modeling: application of spherical centroidal Voronoi tessellations. *Ocean Dyn.* **58**, 475–498 (2008)
60. Rodríguez-Iturbe, I., Isham, V.: Some models for rainfall based on stochastic point processes. *Proc. R. Soc. Lond. A* **410**(1839), 269–288 (1987)
61. Schneider, U., Fuchs, T., Meyer-Christoffer, A., et al.: Global precipitation analysis products of the GPCC. Global Precipitation Climatology Centre (GPCC), DWD, Internet Publication **112** (2008)
62. Schwarzl, M., Godec, A., Metzler, R.: Quantifying non-ergodicity of anomalous diffusion with higher order moments. *Sci. Rep.* **7**, 3878 (2017)
63. Scofield, R.A., Kuligowski, R.J.: Status and outlook of operational satellite precipitation algorithms for extreme-precipitation events. *Weather Forecast.* **18**, 1037–1051 (2003)
64. Serrat-Capdevila, A., Valdes, J.B., Stakhiv, E.: Water management applications for satellite precipitation products: synthesis and recommendations. *J. Am. Water Resour. Assoc.* **50**, 509–525 (2014)
65. von Storch, H., Navarra, A.: *Analysis of Climate Variability Applications of Statistical Techniques*. Springer, Berlin (1999)
66. Tian, Y., Peters-Lidard, C.D., Choudhury, B.J., et al.: Multitemporal analysis of TRMM-based satellite precipitation products for land data assimilation applications. *J. Hydrometeorol.* **8**, 1165–1183 (2007)
67. Wang, H., Wang, C., Zhao, Y., et al.: Toward a practical approach for ergodicity analysis. *Nonlin. Processes Geophys. Discuss.* **2**, 1425–1446 (2015)

68. Wood, E., Roundy, J.K., Troy, T.J., et al.: Hyper-resolution global land surface modeling: meeting a grand challenge for monitoring Earth's terrestrial water. *Water Resour. Res.* **47**, W05,301 (2011)
69. Zhang, Q., Sun, P., Singh, V.P., et al.: Spatial-temporal precipitation changes (1956–2000) and their implications for agriculture in China. *Global Planet. Change* **82**, 86–95 (2012)