Stoch Environ Res Risk Assess (2015) 29:411–416
DOI 10.1007/s00477-014-0923-2

ORIGINAL PAPER

A Matérn model of the spatial covariance structure of point rain rates

Ying Sun · Kenneth P. Bowman · Marc G. Genton · Ali Tokay

Published online: 15 July 2014

Abstract It is challenging to model a precipitation field due to its intermittent and highly scale-dependent nature. Many models of point rain rates or areal rainfall observations have been proposed and studied for different time scales. Among them, the spectral model based on a stochastic dynamical equation for the instantaneous point rain rate field is attractive, since it naturally leads to a consistent space–time model. In this paper, we note that the spatial covariance structure of the spectral model is equivalent to the well-known Matérn covariance model. Using high-quality rain gauge data, we estimate the parameters of the Matérn model for different time scales and demonstrate that the Matérn model is superior to an exponential model, particularly at short time scales.

Y. Sun
Department of Statistics, Ohio State University, Columbus, OH 43210, USA
e-mail: sunwards@stat.osu.edu

K. P. Bowman
Department of Atmospheric Sciences, Texas A&M University, College Station, TX 77843, USA

M. G. Genton (✉)
CEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia
e-mail: marc.genton@kaust.edu.sa

A. Tokay
Joint Center for Earth Systems Technology, University of Maryland Baltimore County, Baltimore, MD 21228, USA

A. Tokay
NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA

Keywords Covariance model · Exponential covariance · Matérn covariance · Point rain rates · Spectral model · Time scales

1 Introduction

Because of its intermittent nature, high variability, and small length and time scales, precipitation poses significant challenges for both observations and modeling. In addition, rainfall statistics are strongly scale dependent. For example, spatial correlation lengths for monthly rain rates are much larger than for hourly or daily rain rates. Similarly, autocorrelation time scales for area-averaged rain rates are larger for larger areas. For many precipitation-related problems, such as estimating area-averaged precipitation from a set of rain gauges or validating satellite precipitation observations with surface observations, it is useful to have a statistical model with the characteristics of the precipitation field. Much progress has been made for precipitation modeling; Onof et al. (2000) reviewed the development of Poisson-cluster processes, Bruno et al. (2014) investigated how to calibrate radar measurements via rain gauge data, Oliveira (2004) constructed separate models for rainfall occurrences and the positive rainfall amounts, and Chakraborty et al. (2014) developed an adaptive spatial model for precipitation data from multiple satellites.

Many types of inference and estimation problems can be addressed if the covariance structure of the rain field is known (Bell and Kundu 2003). In this study we analyze the spatial covariance of high-quality data from a rain gauge network for different averaging times and compare it with a semi-analytical stochastic model of rain developed by Laughlin (1981), North and Nakamoto (1989), and Bell and Kundu (1996). Related problems have been addressed by Li et al. (2009), who investigated the space–time covariance structure of propagating precipitation features, and Kundu and Siddani (2011), who developed an empirical model of space and time scaling properties of the occurrence of rain. North et al. (2011) recently discussed a similar problem for the surface temperature field. The distribution of non-zero rain rates is non-Gaussian with a heavy tail at high rain rates. A number of different distributions have been used to represent the conditional rain rate (the rain rate when raining), such as the log-normal or gamma distributions (Essenwanger 1985) and log-skew-elliptical distributions (Marchenko and Genton 2010). The rain model and the resulting form of the space–time covariance function are described in more detail in Sect. 3.

Developing and calibrating such a statistical model requires consistent, accurate data, particularly when estimating the second moments, e.g., covariances, or even higher moments. In this study we take advantage of high-quality precipitation data from a network of research rain gauges in Virginia, Maryland, and North Carolina that was deployed as part of the NASA Tropical Rainfall Measuring Mission (TRMM) ground validation effort (Tokay et al. 2010). We use the spatial covariance structure of the gauge data to estimate the parameters of the statistical model of Bell and Kundu (1996) for the purely spatial case, and show that it falls into the family of covariance models described by Matérn (1986).

2 Data

The deployment of the rain gauge network is described in detail in Tokay et al. (2010). For quality control and reliability, each site in the network has two or three research-quality 8-inch tipping-bucket rain gauges manufactured by Met One Inc. These gauges are colocated with at least one rain gauge from an operational rainfall monitoring network. From the 20 sites in the network, we select 12 that have essentially complete data for the 3-year period from 2004-05-19 to 2007-05-17. The map in Fig. 1 shows the locations of the 12 gauges used here. The gauges record the time of each bucket tip; one tip is equal to 0.254 mm (0.01 inches) of rain. Bucket tips are converted to rain rates by counting the number of tips within specified intervals. The longitudes and latitudes of the gauge sites and the mean annual rain rate at each gauge for the period of study are given in Table 1, which shows that the long-term mean rate is relatively constant across the network.

3 Rain rate covariance model

3.1 Spectral covariance model

Bell and Kundu (1996) developed a spectral stochastic model for rain rates that is a simplification of the rainfall model in Bell (1987). The analytical properties of a similar model for temperature are discussed in more detail in North et al. (2011). The surface rain rate $R(\mathbf{x}, t)$ at location $\mathbf{x}$ and time t is written as a sum of spatial Fourier modes with amplitudes $a(\mathbf{k}, t)$ of the rain field fluctuation defined as the deviation from the mean, where $\mathbf{k}$ is the spatial frequency. If we assume isotropy, then $k = |\mathbf{k}|$, and the amplitude of each mode is given by a first-order differential equation of the form

$$\frac{\mathrm{d}a(k,t)}{\mathrm{d}t} = -\frac{1}{\tau_k} a(k,t) + f_k(t), \tag{1}$$

where $f_k(t)$ is a white-noise forcing. The damping time scale for each mode τ_k is taken to have the form

$$\tau_k = \frac{\tau_0}{(1 + k^2 \beta_0^2)^{1+\eta}}, \tag{2}$$

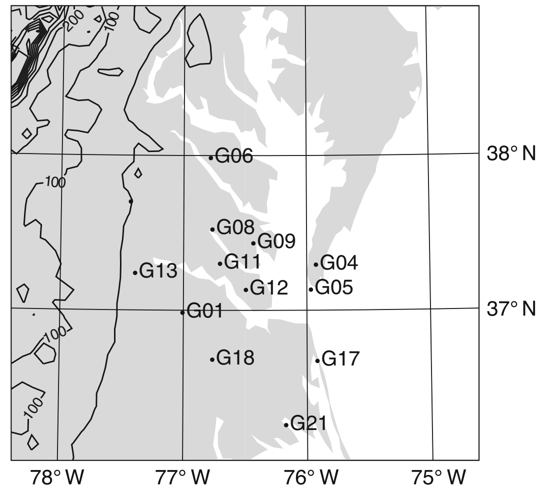

Fig. 1 Locations of the 12 rain gauges used in this study. The *solid black lines* are contours of surface elevation in meters. The contour interval is 50 m

Table 1 Gauge locations and annual mean rain rate at each gauge for the 3-year period of study

Gauge	Longitude	Latitude	Mean (mm)
G01	−77.00	36.98	1,309
G04	−75.92	37.29	966
G05	−75.96	37.13	1,195
G06	−76.78	37.98	1,334
G08	−76.76	37.52	1,141
G09	−76.43	37.43	1,318
G11	−76.70	37.30	1,258
G12	−76.49	37.13	1,287
G13	−77.39	37.23	1,267
G17	−75.91	36.67	1,358
G18	−76.76	36.68	1,308
G21	−76.17	36.25	1,305

where τ_0 is the correlation time scale of the area-averaged field (k = 0), and β_0 is a characteristic length scale. The exponent η affects the scale dependence of the damping time scale τ_k, which provides additional control over the shape of the covariance function.

From (1), Bell and Kundu (1996) found that the power spectrum of the stochastic rain field R can be shown to be

$$\tilde{c}(k,\omega) = \frac{F_0\tau_0^2}{\tau_0^2\omega^2 + (1 + k^2\beta_0^2)^{2+2\eta}}, \tag{3}$$

where ω is the temporal frequency and F_0 is the magnitude of the white-noise forcing, which determines the total variance of the rain field. The space–time covariance of the point rain rate at distance $r = |\mathbf{r}|$ and time lag s implied by this model is given by the inverse transform of the power spectrum

$$c(r,s) = (2\pi)^{-3/2}\int \mathrm{d}\omega \int \mathrm{d}k\; e^{(ikr-\omega s)}\; \tilde{c}\;(k,\omega). \tag{4}$$

This is a stationary spatio-temporal covariance model for instantaneous point rain rates with four parameters: F_0, τ_0, β_0 and η. The covariance of the time-averaged or the area-averaged rain rates can be expressed as suitable integrals over the spectrum for a given set of parameters that are fixed across different time or spatial scales.

For point rain rates, integration of (4) over ω gives

$$c(r,s) = (2\pi)^{-1}\int \mathrm{d}k\; e^{ikr}c(k,s), \tag{5}$$

where

$$c(k,s) = \sqrt{\frac{\pi}{2}}F_0\tau_k e^{-|s|/\tau_k}. \tag{6}$$

Bell and Kundu (1996) noticed that Abramowitz and Stegun (1964) gives the result

$$c(r,0) = \gamma_0\; C_\eta\left(\frac{r}{\beta_0}\right), \tag{7}$$

where $C_\eta(z) = (\frac{z}{2})^\eta K_\eta(z)$, with $K_\eta(z)$ denoting the modified Bessel function of order η, and γ_0 is related to F_0 by

$$F_0 = \sqrt{\frac{2}{\pi}}\Gamma(1+\eta)\left(\frac{\beta_0^2}{\tau_0}\right)\gamma_0.$$

3.2 Matérn covariance model

Among many available covariance models, the Matérn family (Matérn 1986) has gained widespread interest in recent years. Handcock and Stein (1993) introduced the Matérn form of spatial correlations into statistics as a flexible parametric class with one parameter determining the smoothness of the underlying spatial random field. The varied history of this family of models can be found in Guttorp and Gneiting (2006). The Matérn form also naturally arises as the correlation for temperature fields described by simple energy balance climate models (North et al. 2011). The Matérn class of covariance functions is defined as

$$c(r) = \frac{2\sigma^2}{\Gamma(\nu)}\left(\frac{r}{2L_0}\right)^\nu K_\nu\left(\frac{r}{L_0}\right), \tag{8}$$

where $\nu > 0$ depends upon the smoothness of the random field, with larger values of ν corresponding to smoother fields; and $L_0 > 0$ is a spatial range parameter that measures how quickly the correlation of the random field decays with distance, with larger L_0 corresponding to a slower decay (keeping ν fixed). When $\nu = 1/2$, the Matérn covariance function reduces to the exponential covariance model and describes a rough field. The value $\nu = \infty$ corresponds to a Gaussian covariance model which describes a very smooth field, in fact a field which is infinitely differentiable.

The spatial covariance function of the spectral model in (7) is a Matérn form, with $\eta = \nu$, $\gamma_0 = 2\sigma^2/\Gamma(\nu)$ and $\beta_0 = L_0$. Thus, (1) in the spatial domain is equivalent to the Matérn class of covariance models defined in (7).

4 Methods

We analyze the spatial covariance structure of rainfall over the gauge network by first computing average rainfall rates at each gauge at time resolutions varying from 5 min to 3 weeks. The covariances between each pair of gauges are then computed for each time resolution. Specifically, for each time resolution, we use the observed rain rates within the 3-year period to compute the sample correlation between any two sites from the 12 selected gauges. The resulting $\binom{12}{2} = 66$ correlation estimates can be plotted as a function of distance between the two corresponding gauges. The minimum and maximum distances between the gauges in the network are 18 and 200 km, respectively. To describe and model the spatial correlation patterns, we fit the general Matérn and the exponential (special case of $\nu = 1/2$) spatial correlation models to the correlation estimates. Parameters are estimated by ordinary least squares (OLS). The general Matérn model fits two parameters, L_0 and ν. For the exponential model the only adjustable parameter is L_0. In principle the correlation should go to 1 as the separation between instruments goes to zero, but instrumental error, for example, would cause correlations to be less than 1 even for instruments at the same location. Both models are fit without accounting for these so-called nugget effects by forcing the correlation to go to 1 at $r = 0$ since the high-quality colocated gauges do not suggest measurement errors. We aim to show that the flexible Matérn covariance model is more appropriate than choosing the exponential model with an unrealistic nugget effect.

5 Results

Figure 2 shows the correlations of time-averaged rain rates between each pair of rain gauges as a function of station separation for four different time-averaging windows ranging from 10 min to 1 day. Results for averaging windows longer than 1 day are very similar to those for 1 day (panel d). The solid lines are the least-squares fits to the data using the Matérn model, while the dashed lines are for the exponential model. Krajewski et al. (2000) plotted similar figures for long-term seasonally-averaged precipitation values, while Habib et al. (2001) examined correlations at station separations less than 10 km. As expected, the correlations between gauges are larger for gauges that are closer to one another and for longer averaging windows. The model fits also show that as the aggregation time increases, the spatial correlation becomes stronger for a certain distance. In other words, the spatial range L_0 is larger for longer averaging times.

The parameter estimates and the resulting mean squared errors (MSEs) from the Matérn and exponential model fittings are given in Table 2. For each averaging time window, the smaller MSE of the Matérn model indicates a better fit compared to the exponential model. Although the exponential model always has a bigger MSE than the Matérn model does, the difference generally decreases as the averaging time window increases. The quality of the fits can be seen graphically in Fig. 2, which shows that the Matérn model better captures the shape of the correlation function than the exponential model, particularly for shorter averaging times.

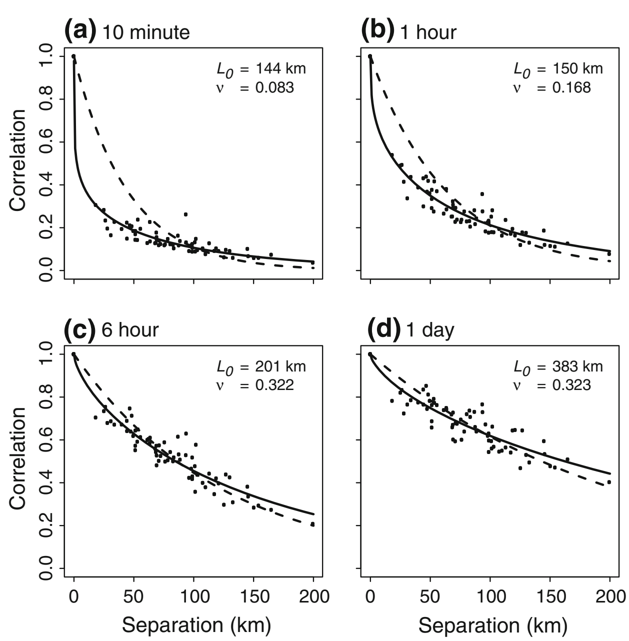

Fig. 2 Correlation as a function of site separation distance for different gauge averaging times. Parameters shown are for the Matérn fits. *Solid lines* least-squares fit to Matérn function; *dashed lines* least-squares fit to exponential

Table 2 Parameter estimates and mean squared errors (MSEs)

Averaging window	Matérn model			Exponential model	
	L_0 (km)	ν	MSE ($\times 10^3$)	L_0 (km)	MSE ($\times 10^3$)
5 min	125	0.080	0.85	42	14.04
10 min	144	0.083	0.91	45	13.49
15 min	159	0.087	0.93	48	12.96
30 min	173	0.105	1.14	54	11.06
1 h	150	0.168	1.48	64	7.15
3 h	165	0.268	2.05	91	3.87
6 h	201	0.322	2.22	124	3.04
1 day	383	0.323	2.50	206	3.15
1 week	396	0.320	2.50	207	3.16
3 weeks	267	0.390	5.06	189	5.80

For the Matérn model the MSE becomes bigger for longer time-averaging windows. This is not surprising because for a fixed record length, the number of rain rate samples within the record becomes smaller as the averaging time window gets longer. Moreover, the parameter estimates of the Matérn model in Table 2 show that when the time-averaging window increases beyond 6 h, the length scale becomes larger than the maximum separation between the gauges (202 km). This can also be seen in the plots in Fig. 2. For longer time scales, therefore, the gauge network is probably too small to accurately estimate the natural length scale of variability of the rain field.

Figure 3 shows graphically the parameter estimates of L_0 and ν from the Matérn model shown in Table 2 for the given averaging time windows (unit: hour). For time scales out to a day, the spatial range parameter L_0 increases approximately linearly. The value of L_0 does not have an increase from 30 min to 1 h or 3 h because the spatial correlation depends on both L_0 and ν in the Matérn model, unlike the exponential model where L_0 itself controls the spatial range. Indeed, up to 1 week, L_0 in the exponential model always increases as the averaging time window gets longer. However, the value of L_0 becomes smaller for the 3 weeks time window, possibly due to the fact that fewer available observations make the estimation less accurate. Therefore, we need a period of time longer than 3 years to estimate the parameters for the averaging time window greater than 1 week. Moreover, the limiting value for very short averaging times appears to be near 100 km, but it is

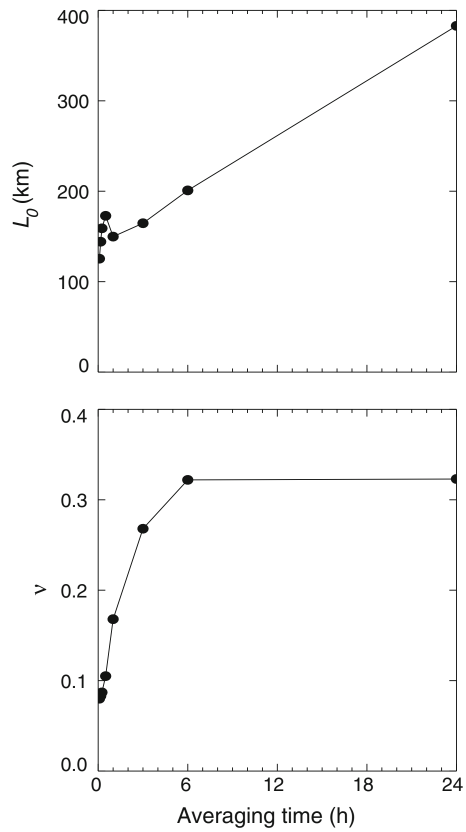

Fig. 3 *Top panel* the plot of parameter estimates of L_0 from the Matérn model (shown in Table 2) as a function of averaging time (unit: hour). *Bottom panel* the plot of parameter estimates of ν from the Matérn model (shown in Table 2) as a function of averaging time (unit: hour)

important to keep in mind that the estimates of the parameter ν indicate that the precipitation field is very non-smooth. The smoothness parameter itself increases from a value less than 0.1 to around 0.3 for longer averaging times. In all cases the estimates of the exponent ν from fitting the Matérn model are less than the value for the exponential model (0.5). This indicates that rain fields are rougher than a random field that has an exponential spatial covariance. At short time scales the rain field appears to be much rougher than exponential, but this conclusion is limited by the fact that the minimum separation of the rain gauges is 18 km, so it is not possible to directly observe the variability at length scales shorter than that. The smoothness parameter ν increases as the averaging time increases, indicating a smoother random field. This is not surprising, as time averaging would be expected to reduce the variability of the rain field.

The results in Table 2 can be compared with Table 1 in Bell and Kundu (2003). They fit a single consistent space–time covariance model of Bell and Kundu (1996), and the parameters of the covariance model are estimated using radar rainfall maps from the several different tropical field experiments. The model in Bell and Kundu (1996) describes the covariance structure of the instantaneous point rain rates, so that the covariance of the time-averaged rain gauge measurements and the area-averaged radar observations can be expressed as suitable integrals. Therefore, its parameters are scale independent. In our analysis, we fit independent spatial covariance models for different time scales and obtain different parameter estimates, which are thus dependent on time scales.

In Bell and Kundu (2003), the spatial and temporal resolution of the radar data from those experiments is typically ~ 4 km and ~ 15 min, respectively. They found values of β_0 from 61 to 104 km for instantaneous rain rates. In our analysis, when averaging the rain gauge data with windows of 15 min or less, the parameter L_0 in the Matérn model ranges from 125 to 159 km, which corresponds reasonably well to the radar data with a slightly larger spatial range due to the short-time aggregation. The greatest differences between this study and Bell and Kundu (2003) are found for the values of the smoothness parameter. Bell and Kundu (2003) estimated η to be in the range $-1/2 \leq \eta < 0$, arguing that negative values of η fit the data best. Although for the space–time spectrum in (2), the part, $(1 + k^2\beta_0^2)^{2+2\eta}$, in the denominator still makes the spectrum integrable with respect to k, the purely spatial model, which is a Matérn form with the spectrum having $(1 + k^2\beta_0^2)^{1+\eta}$ in the denominator, has infinite variance at the origin. In our fitting procedure, however, we only fit a purely spatial covariance model, therefore, we constrain the values of ν to be positive, which results in a finite variance at the origin and a valid covariance model.

6 Discussion

In geostatistics, it is common to treat observations in time as replicates and fit a spatial exponential covariance model to precipitation data. In this paper, we aimed to show that the flexible Matérn covariance model is more appropriate than the exponential model for characterizing the spatial dependence of the precipitation especially on short time scales. Although maximum likelihood estimates have better properties, it is not trivial to apply the maximum likelihood method due to the fact that the distribution of precipitation is usually non-Gaussian with rainfall zeros. Therefore, the least squares method is a reasonable alternative, widely used in the precipitation literature, in the sense that we do not have to assume distributions and temporal replicates are available in our setting. In the least squares method, the MSE plays the same role as the negative likelihood function, where a smaller MSE indicates a better model. We have shown that the spatial component of the spectral covariance model for precipitation developed by Bell and Kundu (1996, 2003) is equivalent to the family of covariance models introduced by Matérn (1986). We used high-quality, high-frequency rain rate data from a network of research rain gauges (Tokay et al. 2010) to estimate the parameters of the Matérn covariance model for a range of different rain-rate averaging times (accumulation times). The results indicated that for averaging times less than a few weeks, the rain field is rougher than is the case for a random field with an exponential spatial covariance structure. The roughness increases as the averaging time decreases. In contrast to Bell and Kundu (2003), our model is time scale dependent. We estimate the parameters of the Matérn model using the rain gauge data with different aggregation to characterize the precipitation field at different time scales in terms of the smoothness and the spatial correlation range.

The datasets described in this paper have many interesting features to explore. We have focused on estimating the spatial correlation of the precipitation using the Matérn covariance model to compare the roughness of the spatial fields on different time scales. Considering the relatively small geographic region in this study, the spatial stationarity of the precipitation process was assumed. If the spatio-temporal dependence is of interest, one may further develop more complicated spatio-temporal models for each given time scales. However, the stationarity assumption in time is unlikely to be realistic as precipitation usually shows seasonal trend. Consequently, a longer period of data will be needed to study the long-term characteristics.

It is also worth noting that in addition to the Matérn class, there are other available covariance families that one can choose from, depending on the properties of the precipitation field. For example, the generalized Cauchy family proposed by Gneiting (2004) is suitable to model the long-range dependence in a process. Whether such a covariance family may arise from a stochastic differential equation similar to (1) is an open problem. Further extensions to multiple variables and multivariate Matérn cross-covariance functions (Gneiting et al. 2010; Apanasovich et al. 2012; Genton and Kleiber 2014) would be of interest too.

Acknowledgments The research in this article was partially supported by Award No. KUSC1-016-04 made by King Abdullah University of Science and Technology (KAUST) and by the Spanish Ministry of Science and Innovation (Project MTM2011-22664) which is co-funded by FEDER.

References

Abramowitz M, Stegun IA (1964) Handbook of mathematical functions with formulas, graphs, and mathematical tables. No. 55 in National Bureau of Standards Applied Mathematics Series. Superintendent of Documents, U.S. Government Printing Office, Washington, DC

Apanasovich TV, Genton MG, Sun Y (2012) A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components. J Am Stat Assoc 107:180–193

Bell TL (1987) A space–time stochastic model of rainfall for satellite remote-sensing studies. J Geophys Res 92:9631–9643. doi:10.1029/JD092iD08p09631

Bell TL, Kundu PK (1996) A study of the sampling error in satellite rainfall estimates using optimal averaging of data and a stochastic model. J Clim 9:1251–1268

Bell TL, Kundu PK (2003) Comparing satellite rainfall estimates with rain gauge data: optimal strategies suggested by a spectral model. J Geophys Res 108:4121. doi:10.1029/2002JD002641

Bruno F, Cocchi D, Greco F, Scardovi E (2014) Spatial reconstruction of rainfall fields from rain gauge and radar data. Stoch Environ Res Risk Assess 28:1235–1245

Chakraborty A, De S, Bowman K, Sang H, Genton MG, Mallick B (2014) An adaptive spatial model for precipitation data from multiple satellites over large regions. Stat Comput (in press)

Essenwanger OM (1985) General climatology: elements of statistical analysis, vol 1B., World survey of climatology, Elsevier, Amsterdam

Genton MG, Kleiber W (2014) Cross-covariance functions for multivariate geostatistics. Stat Sci (in press)

Gneiting T (2004) Stochastic models that separate fractal dimension and the hurst effect. SIAM Rev 46:269–282

Gneiting T, Kleiber W, Schlather M (2010) Matérn cross-covariance functions for multivariate random fields. J Am Stat Assoc 105:1167–1177

Guttorp P, Gneiting T (2006) Studies in the history of probability and statistics XLIX: on the Matérn correlation family. Biometrika 93:989–995. doi:10.1093/biomet/93.4.989

Habib E, Krajewski WF, Ciach GJ (2001) Estimation of rainfall interstation correlation. J Hydrometeorol 2:621–629

Handcock MS, Stein ML (1993) A Bayesian analysis of kriging. Technometrics 35:403–410

Krajewski WF, Ciach GJ, McCollum JR, Bacotiu C (2000) Initial validation of the Global Precipitation Climatology Project monthly rainfall over the United States. J Appl Meteorol 39:1071–1086

Kundu PK, Siddani RK (2011) Scale dependence of spatiotemporal intermittence of rain. Water Resour Res 47:318–340. doi:10.1029/2010WR010070

Laughlin CR (1981) On the effect of temporal sampling on the observation of mean rainfall. In: Atlas D, Thiele OW (eds) Precipitation measurement from space. National Aeronautics and Space Administration, Greenbelt, pp D59–66

Li B, Murthi A, Bowman K, North G, Genton MG, Sherman M (2009) Statistical tests of Taylor's hypothesis: an application to precipitation fields. J Hydrometeorol 10:254–265

Marchenko YV, Genton MG (2010) Multivariate log-skew-elliptical distributions with applications to precipitation data. Environmetrics 21:318–340

Matérn B (1986) Spatial variation. In: Lecture notes in statistics, vol. 36, 2nd edn. Springer, Berlin

North GR, Nakamoto S (1989) Formalism for comparing rain estimation designs. J Atmos Oceanic Technol 6:985–992

North GR, Wang J, Genton MG (2011) Correlation models for temperature fields. J Clim 24:5850–5862. doi:10.1175/2011JCLI4199.1

Oliveira VD (2004) A simple model for spatial rainfall fields. Stoch Environ Res Risk Assess 18:131–140

Onof C, Chandler RE, Kakou A, Northrop P, Wheater HS, Isham V (2000) Rainfall modelling using Poisson-cluster processes: a review of developments. Stoch Environ Res Risk Assess 14:384–411

Tokay A, Bashor PG, McDowell VL (2010) Comparison of rain gauge measurements in the mid-Atlantic region. J Hydrometeorol 11:553–565. doi:10.1175/2009JHM1137.1