TOPICS IN VARIABLE SELECTION AND SPATIAL STATISTICS WITH APPLICATIONS

A Dissertation

by

ZIJUAN CHEN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Suojin Wang
Co-Chair of Committee,	Raymond J. Carroll
Committee Members,	Matthias Katzfuss
	Ignacio Rodriguez-Iturbe
Head of Department,	Daren Cline

August 2020

Major Subject: Statistics

ABSTRACT

Spatial random fields are widely used to model the space-time correlation structure of different objects in various fields, such as the soil moisture field in hydrological, agricultural and climatological studies. Since physically derived representations of soil moisture fail to reproduce the fast decay spatial correlation observed in available data, a jitter process is incorporated in the soil moisture model to deflate its correlation structure. An empirical study shows that the new model has successfully captured the spatial-temporal variability of soil moisture at small scales.

We are then able to study the relationship between the soil moisture field and savannas via simulation after the theory of the space-time model of soil moisture has been well established. The sizes of tree clusters in savannas are found to be following power-law distributions in many studies, while the simulation results also indicate that the size of the soil moisture follows power laws. We infer that the power laws with specific exponents observed in the tree cluster data result from the power law of the soil moisture islands when the impact of fire and herbivores is accounted for.

The distribution properties of clusters or islands defined by specific thresholds and underlying spatial random fields, such as the soil moisture islands, are commonly studied in practice. However, there are no applicable statistical methods to analyze this type of clusters. Therefore, we introduce a well-defined distribution function and provide estimation procedure for the function to study the distribution properties of this type of clusters. Numerical experiments and an application to tree clusters data are carried out to validate the results.

Partially linear single-index models are commonly used in the analysis of different kinds of data. We are interested in the extended partially linear single-index models, which is more flexible than the usual partially linear single-index models. We proposed local smoothing estimators for parameter estimation, and introduced the penalized estimators for variable selection. A chi-squared type of test statistic is also carried out for linear hypotheses. Simulation studies are presented to support our analytic results and a real data analysis is provided for illustration.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professors Suojin Wang, Raymond J. Carroll and Matthias Katzfuss of the Department of Statistics and Professor Ignacio Rodriguez-Iturbe of the Department of Ocean Engineering.

The model derivations in Chapter 2 were conducted in part by Professor Ignacio Rodriguez-Iturbe of the Department of Ocean Engineering and Binayak Mohanty of the Department of Biological and Agricultural Engineering. The analyses in Chapter 3 were conducted in part by Professors Ignacio Rodriguez-Iturbe and contributed in part by Professors Ann Carla Staver of the Department of Ecology and Evolutionary Biology of Yale University and Simon Asher Levin of the Department of Ecology and Evolutionary Biology of Princeton University. The analytical results and discussions of empirical studies in Chapters 4 and 5 were contributed in part by Professor Suojin Wang of the Department of Statistics.

The data sets analyzed for Chapter 2 were provided by Professor Binayak Mohanty. The data set analyzed for Chapter 4 was provided by Professor Ann Carla Staver. The data set analyzed for Chapter 5 was published in [Yeh, 2007] and is publicly available.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported in part by the Texas A&M Engineering Experiment Station and the Simons Foundation Mathematics and Physical Sciences - Collaboration Grants for Mathematicians Program Award #499650.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

Methodologies of spatial statistics, including the studies of stochastic processes, are widely used in various fields. In particular, the space-time variability of soil moisture is of fundamental interest in hydrological, agricultural and climatological studies, where techniques of spatial random fields play an important role. In Chapter 2, a physically derived space-time mathematical representation of the soil moisture field is carried out via the soil moisture balance equation driven by stochastic rainfall forcing. The model incorporates spatial diffusion and in its original version, it is shown to be unable to reproduce the relative fast decay in the spatial correlation functions observed in empirical data. This decay resulting from variations in local topography as well as in local soil and vegetation conditions is well reproduced via a jitter process acting multiplicatively over the space-time soil moisture field. The jitter is a multiplicative noise acting on the soil moisture dynamics with the objective to deflate its correlation structure at small spatial scales which are not embedded in the probabilistic structure of the rainfall process that drives the dynamics. These scales of order of several meters to several hundred meters are of great importance in ecohydrologic dynamics. Properties of space-time correlation functions and spectral densities of the model with jitter are explored analytically, and the influence of the jitter parameters, reflecting variabilities of soil moisture at different spatial and temporal scales, is investigated. A case study fitting the derived model to a soil moisture dataset is presented in detail in Chapter 2.

Tree clusters in savannas are commonly found in sizes that follow power laws with well-established exponents. It is of great interest that where does the power-law property of tree clusters in savannas come from. Thanks to the space-time model of soil moisture described in Chapter 2, we are able to simulate the soil moisture field with fast decay spatial correlation function as shown in real data. In Chapter 3, by simulating soil moisture data, we show that the size distributions of tree clusters could result from the space-time probabilistic structure of soil moisture, estimated over the range of rainfall observed in semi-arid savannas; patterns of soil moisture display islands whose size, for moisture thresholds above the mean, follow power laws. These islands are the

regions where trees are expected to exist and they have a fractal structure whose perimeter-area relationship is the same as observed on field data for the clustering of trees. When the impact of fire and herbivores is accounted for, as acting through the perimeter of the tree clusters, the power law of the soil moisture islands is transformed into a power law with the same exponents observed in the tree cluster data.

In Chapter 3, we simulate the soil moisture field and study the distribution of the sizes of the soil moisture islands. This type of clustering procedure is very different from the traditional statistical clustering analysis, which is based on discrete observations. This method of clustering defines clusters as the connected areas where a well-defined spatial random field is above certain threshold. Note that the defined clusters are correlated, and that the empirical distribution function calculated from the clusters might not converge to a well-defined distribution function induced by a properly defined random variable. Therefore, the problem now is that to define a valid distribution function which can represent the characteristics of the clusters, and propose efficient estimation methodologies for estimating it. However, the available statistical techniques for analyzing clustering models are not applicable to this problem. Thus in Chapter 4, we study the distribution properties of the clusters by defining a distribution function of the clusters rigorously and providing methods to estimate the spatial distribution function. The theoretical results shown in Chapter 4 are illustrated by numerical experiments and an application to a real world problem about tree clusters in savanna.

After studying the modeling of stochastic processes and distributional properties of clusters defined on spatial random fields, our interest lies in other hot topics of statistics: partially linear single-index models and variable selection. In partially linear single-index models, there are two different covariate matrices in the model for the linear part and non-linear part. All covariate information needs to be divided into two parts before the model can be fitted. In contrast, in the extended partially linear single-index models, all the covariate variables are included in one matrix, which is contained in both the linear part and non-linear part of the model. In Chapter 5, we propose local smoothing estimators for the model parameters and unknown function, whose asymptotic properties are demonstrated. The profile estimating procedure for calculating the estimators is also

introduced. We show that the solution to the optimization of the profile objective function is unique and has linear expressions, which leads to fast and accurate computations of the estimators. We also employ the lasso penalty to obtain penalized estimators with consistency and oracle property in order to carry out estimation and variable selection simultaneously. Finally, we develop a linear hypothesis test for the model parameters. Simulation studies are presented to support our analytic results and a real data analysis is provided for illustration.

2. SPACE-TIME MODELING OF SOIL MOISTURE*

2.1 Introduction and general background

The space-time variability of soil moisture is of fundamental interest in hydrological, agricultural and climatological studies. Its probabilistic description at different temporal and spatial scales presents challenges related to the description of the rainfall process, infiltration, and movement of the water on the soil surface (e.g., [Kittredge, 1948] and [Eagleson, 1978]). An important feature of the soil moisture dynamics is that its spatial correlation structure decays much faster than that of the rainfall process and such a decay has not been accomplished with the existing physically based schemes. A number of studies have been carried out to establish a theoretical basis for such a probabilistic description (e.g., [Albertson and Montaldo, 2003] and [Pan et al., 2003]) but none accomplishes a satisfactory result for temporal scales of one day and spatial scales of the order of 1 m^2 to 10,000 m^2 which are of fundamental importance in ecohydrology. Many of these studies (e.g., [Isham et al., 2005] and [Rodriguez-Iturbe et al., 2006]) start from the basic soil moisture balance equation at a point:

$$nZ_r \frac{dS(t)}{dt} = I(t) - E(s) - L(s), \tag{2.1}$$

where in the left-hand side $S(t)$ represents the dimensionless relative soil moisture, n is the soil porosity, and Z_r is the depth of the root zone. On the right had-side, the terms $I(t)$, $E(s)$, and $L(s)$ are the rates of infiltration, evapotranspiration and leakage (the last two function of the current level of soil moisture, S). Equation (2.1) can be rewritten as ([Rodriguez-Iturbe et al., 2006])

$$nZ_r \frac{dS(t)}{dt} = (1-\phi)Y(t) - VS(t), \tag{2.2}$$

*Reprinted with permission from "Space-time modeling of soil moisture" by Chen, Z., Mohanty, B. P. and Rodriguez-Iturbe, I., 2017. *Adv. Water. Resour.*, 109, 343-354, Copyright [2020] by Elsevier.

where $(1-\phi)$ is the net rainfall coefficient which accounts for interception, $Y(t)$ [L] is the rainfall on the point under consideration, and V [L/T] is the soil water loss coefficient accounting for transpiration, leakage and runoff losses. The spatial character of the analysis results from the rainfall Y which is represented via a space-time model from where the characteristics of $Y(t)$ at any point are then derived. This leads to a characterization of $S(t)$ which depends on space and time via the rainfall process and whose space-time correlation structure can be analytically derived.

[Entekhabi and Rodriguez-Iturbe, 1994] used a different soil moisture balance equation which explicitly incorporates the spatial spreading of soil moisture through a diffusion term:

$$nZ_r \frac{\partial S(\boldsymbol{x},t)}{\partial t} = -\eta S(\boldsymbol{x},t) + k\nabla^2 S(\boldsymbol{x},t) + \xi(\boldsymbol{x},t), \tag{2.3}$$

where the diffusion term encompasses not only subsurface diffusion but most importantly the spread of moisture that takes place on the surface of the ground due to the topography and the roughness of the terrain. The term $\eta S(\boldsymbol{x},t)$ is equivalent to the term $VS(t)$ in Equation (2.2) and the rainfall input is explicitly represented in space and time by the term $\xi(\boldsymbol{x},t)$. Similarly to Equation (2.2) if the rainfall input is stationary the soil moisture will evolve to reach a statistical equilibrium state also stationary in space and time.

The representation of the rainfall input in space and time into Equations (2.2) and (2.3) may be carried out through several different probabilistic rainfall models. Thus, [Entekhabi and Rodriguez-Iturbe, 1994] used the model proposed by [Waymire et al., 1984] which models the rain intensity, $\xi(t)$ in Equation (2.3) as rain bands which arrive in space and time according to an homogeneous Poisson process. Each rainband has a random number of cell centers. The rain intensity is a maximum at the center of each cell (random variable exponentially distributed) and decreases from the center of the cell with a Gaussian-like spread. The cells are born at a random distance and time away from the storm origin. More details can be found in [Waymire et al., 1984].

[Isham et al., 2005] used a representation of the spatio-temporal rainfall input $Y(t)$ in Equation (2.2) as that resulting from the [Cox and Isham, 1988] rainfall model. Here, rainfall occurences

are modelled by a sequence of circular rain cells with random area, random intensity and random duration that occur in a Poisson process as described in detail in the next section. In both studies, [Entekhabi and Rodriguez-Iturbe, 1994] and [Isham et al., 2005], the region under consideration is assumed to be uniform in topography, soil, and vegetation chracteristics which make the losses in Equations (2.2) and (2.3) independent of space. [Rodriguez-Iturbe et al., 2006] extended the analysis of [Isham et al., 2005] to the case of an heterogeneous vegetation.

The focus of the analytical framework is well described by [Entekhabi and Rodriguez-Iturbe, 1994]: "The key question is then, how is the input rainfall spectrum modified in its space-time structure when filtered through the land surface processes to yield the evolution of soil moisture?". Regardless of the different rainfall models driving the soil moisture balance equation as well the incorporation in Equation (2.3) of the diffusion term, the results of the analyses carried out by [Entekhabi and Rodriguez-Iturbe, 1994] and [Isham et al., 2005] have a common characteristic which is a key handicap for the use of those schemes towards the development of hydrologic parametrizations and ecohydrological studies. They find that the space-time fluctuations of the rainfall field are basically translated to the soil moisture field. The spatial and temporal variations in soil moisture just reflect an scaled version of those of the rain field (The scale changes from the rain field to the soil moisture field) and thus for scales of say 1 day and less than 1 km^2 lead to spatial correlation structures which decay much slower than what is observed in the empirical data. The modeling schemes with realistic parameters yield large correlations in soil moisture for hundreds of meters and even kilometers as controlled by the rainfall field. The incorporation of the diffusion term in Equation (2.3) with large values of "k" increases the lateral redistribution of soil moisture and tend to decrease its spatial variability on small local scales . For smaller "k's" its impact is negligible. In all cases the resulting spatial correlation structure reflects that of the rainfall field with unrealistic slow decay of the spatial correlation structure.

The next section of this chapter briefly describes the analytics of the [Cox and Isham, 1988] rainfall model to be used in Section 2.3 as input into Equation (2.3) incorporating the diffusion term as suggested by [Entekhabi and Rodriguez-Iturbe, 1994]. This will be followed in Section

2.4 by the development of a new model which incorporates a "jitter" process in the soil moisture dynamics to represent the strong fluctuations in soil moisture arising from local and important small scale variabilities in topography, soil, and vegetation characteristics. This new analytical scheme is shown to lead to realistic space-time soil moisture fields. The new modelling scheme is finally used to represent a set of daily level soil moisture data in an 800 m by 800 m plot.

2.2 The Cox and Isham rainfall model

Rainfall occurrences are modeled by a sequence of circular rain cell that occur in a Poisson process of rate λ in space and time. Each cell has a random radius, W, a random duration, D and random average intensity during its duration, X, so that the total depth of the cell is XD. The random variables are mutually independent and the triples (W, D, X) are independent and identically distributed over the cells. In this model, the cells overlap temporally as well as spatially so that, at any particular spatial location $\boldsymbol{u}$ and time t, the rainfall process is the superposition of all cells overlapping the space-time $(\boldsymbol{u}, t)$. If W and D are all exponentially distributed with parameters ρ and η respectively ($\rho = \mu_W^{-1}$, $\eta = \mu_D^{-1}$), then the mean and covariance function of the total intensity $Y(\boldsymbol{u}, t)$ are (Cox and Isham,1988)

$$E(Y(\boldsymbol{u}, t)) = \pi\lambda E(W^2)\mu_D\mu_X = \frac{2\pi\lambda\mu_X}{\eta\rho^2}, \tag{2.4}$$

$$\Gamma_Y(l, h) \simeq \frac{2\pi\lambda E(X^2)}{\eta\rho^2}\left(1 + \frac{\rho l}{4}\right)e^{-\eta h - \rho l/2}, \tag{2.5}$$

where $\Gamma_Y(l, h)$ denotes the covariance between the rain intensity at two locations at a distance l and with a temporal lag $h \geqslant 0$. For this rainfall model, the temporally aggregated rainfall process is defined as [Rodriguez-Iturbe et al., 2006]

$$Y^T(\boldsymbol{u}, m) = \int_{mT}^{(m+1)T} Y(\boldsymbol{u}, t)dt, \tag{2.6}$$

where $[mT, (m+1)T]$ is the time window of integration with m representing the time lag from a time origin and T the aggregation interval. The mean and variance of the temporally aggregated rainfall are [Rodriguez-Iturbe et al., 2006]

$$\mu_{Y^T} = \frac{2\pi\lambda\mu_X T}{\eta\rho^2}, \quad \sigma^2_{Y^T} = \frac{4\pi\lambda E(X^2)(e^{-\eta T} + \eta T - 1)}{\eta^3\rho^2}, \tag{2.7}$$

and the spatial correlation and temporal correlation of the temporally aggregated rainfall are

$$\gamma_S(l) = \text{corr}(Y^T(\mathbf{0}, 0), Y^T(\boldsymbol{l}, 0)) = \left(1 + \frac{\rho l}{4}\right) e^{-\rho l/2}, \tag{2.8}$$

$$\gamma_T(m) = \text{corr}(Y^T(\mathbf{0}, 0), Y^T(\mathbf{0}, m)) = \frac{e^{-\eta m T}(e^{-\eta T} - 2 + e^{\eta T})}{2(e^{-\eta T} + \eta T - 1)}, \quad m = 1, 2, ... \tag{2.9}$$

By taking the Fourier transform on (2.5), the space-time spectrum of Y is obtained as

$$\Phi_Y(\nu, \omega) = \frac{(2\pi)^2\lambda E(X^2)}{\rho} \frac{2\left(\frac{\rho}{2}\right)^2 + \frac{\nu^2}{2}}{\left[\left(\frac{\rho}{2}\right)^2 + \nu^2\right]^{5/2}(\eta^2 + \omega^2)}. \tag{2.10}$$

where ν and ω are complementary to s and t respectively (see [Whittle, 1962]). We can also obtain the space spectrum and time spectrum of Y by

$$\Phi_Y(\nu) = \frac{1}{2\pi}\int_{\mathbb{R}} \Phi_Y(\nu, \omega) d\omega = \frac{2\pi^2\lambda E(X^2)}{\eta\rho} \frac{2\left(\frac{\rho}{2}\right)^2 + \frac{\nu^2}{2}}{\left[\left(\frac{\rho}{2}\right)^2 + \nu^2\right]^{5/2}}, \tag{2.11}$$

$$\Phi_Y(\omega) = \frac{1}{(2\pi)^2}\int_{\mathbb{R}^2} \Phi_Y(\nu, \omega) d\boldsymbol{\nu} = \frac{4\pi\lambda E(X^2)}{\rho^2} \frac{1}{\eta^2 + \omega^2}. \tag{2.12}$$

In Figure 2.1, the space-time spectrum of the rainfall intensity $\Phi_Y(\nu, \omega)$ is plotted against ν for different temporal scales of variability. It shows that $\Phi_Y(\nu, \omega)$ starts decaying at the 100 km scale for all temporal scales, which indicates that at all temporal scales, the structured spatial variability in the rainfall model is contained within the 0 to 100 km range. At larger scales ($>$ 100 km), there is equal amount of variability at every scale, implying unstructured or white-noise patterns

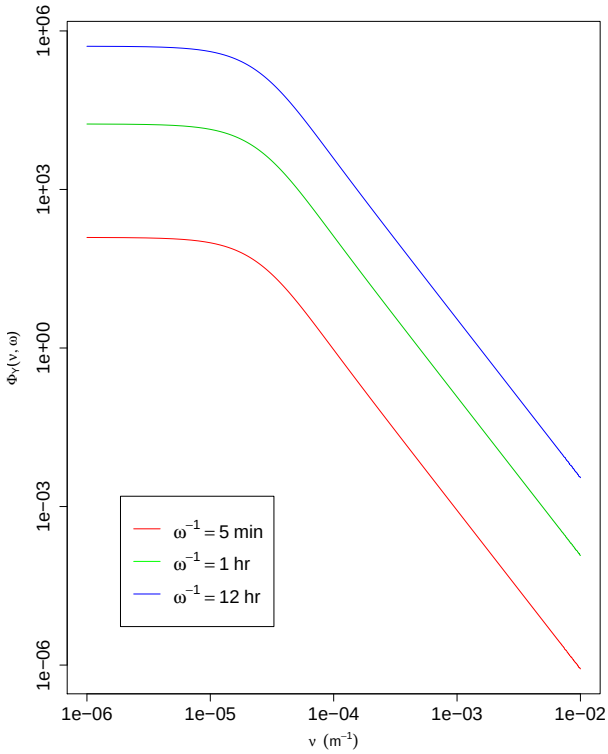
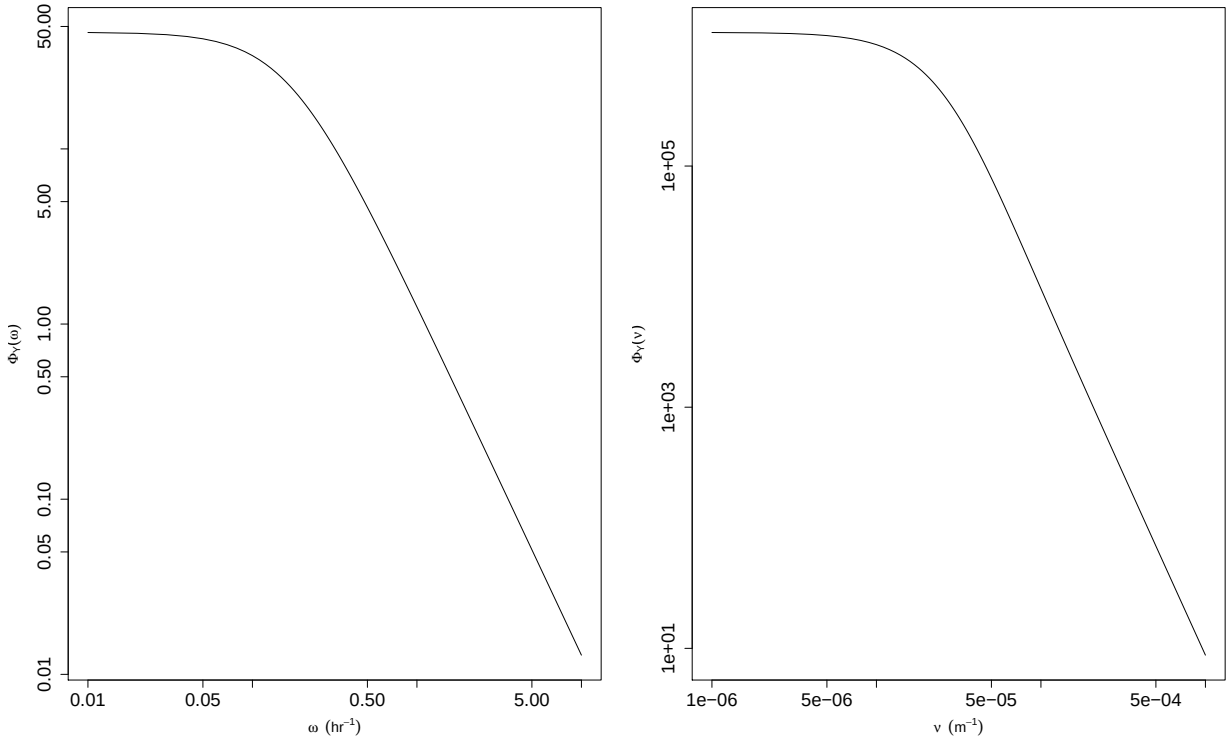


Figure 2.1: The space-time spectrum of rainfall intensity $\Phi_Y(\nu, \omega)$. Most parameters of the model are based on analysis of empirical Italian data (Rodriguez-Iturbe et al. 2006): $\eta^{-1} = 0.25$ days, $\rho^{-1} = 16.6$ km, $E(X^2) = 2\mu_X^2 = 1.29 \times 10^3$ mm^2 day^{-2}, $\lambda = 1.66 \times 10^{-4}$ day^{-1} km $^{-2}$.



(a) Time spectrum

(b) Space spectrum

Figure 2.2: The space spectrum and time spectrum of rainfall intensity $\Phi_Y(\omega)$ for the same parameters of Figure 2.1.

in the rainfall statistics. If we consider the spatial and temporal scales of variability in the rainfall intensity separately from each other, from Figure 2.2a we can see that the structured temporal variability of rainfall is contained in 0 to 20 h scale. Again, the space spectrum in Figure 2.2b shows that the structured spatial variability is contained at $\leqslant$ 100 km scales.

The [Waymire et al., 1984] model used by [Entekhabi and Rodriguez-Iturbe, 1994] to study the space-time variability of soil moisture yields (for typical parameters) the spatial variability of the rainfall process contained in the 1 km to 10 km range, which is smaller than the range of spatial variability of the Cox and Isham rainfall model. In order to consider a wider range in the structured spatial variability of the rainfall input into the soil moisture dynamics we use the Cox and Isham rainfall model in this chapter although the conclusions are the same when using the [Waymire et al., 1984] model.

2.3 Soil moisture dynamics with diffusion

We focus on the space-time soil moisture dynamics represented by Equation (2.3). Thus we will be able to study the impact of the diffusion term on the statistical structure of the soil moisture field. We emphasize that the diffusion term can only account for topographic induced processes. Notice that the coefficient k in Equation (2.3) stands for $k = k^*/nZ_r$, where k^* is the diffusion coefficient controlling the spread of soil moisture in space, considered to be spatially constant. Also, $k = 0$ when there is no rain and thus its value represents an average over all possible values of S throughout the rainfall process occurring continuously in time. For instance, if the diffusion coefficient is 1 m^2 day^{-1}, and the proportion of days with rainfall is 0.25, then $k = 0.25$ m^2 day^{-1}. We will rewrite Equation (2.3) as

$$nZ_r \frac{\partial S(\boldsymbol{u}, t)}{\partial t} = k^* \nabla^2_{\boldsymbol{u}} S(\boldsymbol{u}, t) + I(\boldsymbol{u}, t) - \xi(\boldsymbol{u}, t), \tag{2.13}$$

to emphasize that $I(\boldsymbol{u}, t)$ is the rate of infiltration after interception by vegetation has been accounted for. $\xi(\boldsymbol{u}, t)$ are the evapotranspiration losses. We assume $I(\boldsymbol{u}, t)/nZ_r = bY(\boldsymbol{u}, t)$, where the parameter b represents the proportion of rainfall that become infiltration after being standard-

ized. Also, for the purpose of this chapter and following [Isham et al., 2005] and [Rodriguez-Iturbe et al., 2006], we approximate the standardized loss ξ/nZ_r by the simple linear form $\xi/nZ_r = aS$, where the parameter a will depend on vegetation and soil characteristics. Equation (2.13) is then written as

$$\frac{\partial S(\boldsymbol{u},t)}{\partial t} = k\nabla^2_{\boldsymbol{u}} S(\boldsymbol{u},t) - aS(\boldsymbol{u},t) + bY(\boldsymbol{u},t). \tag{2.14}$$

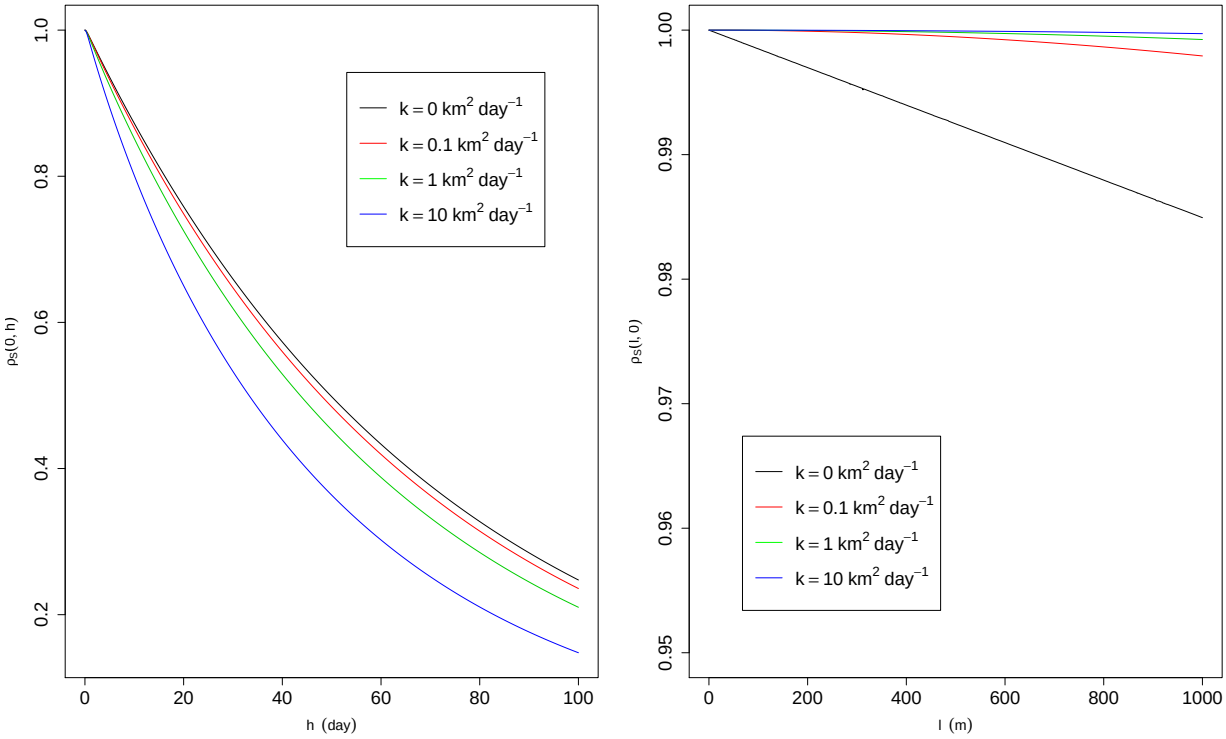
It follows that the expected value of the relative soil moisture is (see Appendix A.1)

$$\mu_S = E(S(\boldsymbol{u},t)) = \frac{2\pi\lambda\mu_X b}{\eta\rho^2 a}, \tag{2.15}$$

which is the same as that for the case when there is no diffusion ($k = 0$). The space-time covariance function of the soil moisture is

$$\begin{aligned}\Gamma_S(l,h) = &\frac{2\pi b^2\lambda E(X^2)}{2\rho}\cdot \\ &\int_0^\infty \frac{\nu\left(2\left(\frac{\rho}{2}\right)^2 + \frac{\nu^2}{2}\right)}{\left[\left(\frac{\rho}{2}\right)^2 + \nu^2\right]^{5/2}\left[\eta^2 - (a+k\nu^2)^2\right]}\left\{\frac{e^{-(a+k\nu^2)h}}{a+k\nu^2} - \frac{e^{-\eta h}}{\eta}\right\} J_0(l\nu)d\nu,\end{aligned} \tag{2.16}$$

where $J_0(\cdot)$ is the Bessel function of the first kind and $l = \|\boldsymbol{l}\|$. Though the analytical form of the covariance function for general (l,h) is not available, numerical calculations for this function are simple. Figure 2.3a demonstrates the behavior of the autocorrelation function of S for different values of k. The range of values of k was studied by [Entekhabi and Rodriguez-Iturbe, 1994], who obtained an upper limit for k around of $O(10^5)$ (m^2 h^{-1}) or $O(1)$ (km^2 day^{-1}). As seen in Figure 2.3a, the auto correlation function for realistic values of k is not very different from that for $k = 0$ (no diffusion). This implies that incorporating the diffusion with $k < 1$ km^2 day^{-1} will make little difference in the temporal correlation. Figure 2.3b shows that the spatial correlation increases with the value of k and remains always unrealistically high for distances at which field data shows a steep decrease in the spatial correlation. This decrease is caused by the variability and fluctuations one observes in local topography and local soil and vegetation conditions in an



(a) Auto correlation functions

(b) Spatial correlation functions

Figure 2.3: Comparison of the auto correlation functions $\rho_S(0,h) = \Gamma_S(0,h)/\Gamma_S(0,0)$ and the spatial correlation functions $\rho_S(l,0) = \Gamma_S(l,0)/\Gamma_S(0,0)$ for different values of k. Note that $k = 0$ is the correlation without diffusion. For comparison purpose, we use $nZ_r = 500$ mm, $V = 7$ mm day $^{-1}$. Other parameters are the same as in Figure 2.1.

otherwise homogeneous terrain. In order to adapt a model like Equation (2.14) for the realistic representation of the soil moisture field at time and space scales frequently used in hydrology and ecological studies it becomes necessary to include an additional source of fluctuations at the scales of interest. This will be described in the next section via a "jitter" type of process.

2.4 Space-time variability of soil moisture with jitter

2.4.1 The space-time covariance function

As discussed before, the correlation structure resulting from the soil moisture balance equation driven only by a stochastic rainfall dynamics is fundamentally controlled by the rainfall process

resulting in a very slow decay of the correlation structure of the soil moisture when compared to that observed in the field. The impact of fluctuations in topography, soil properties and vegetation characteristics is of fundamental importance in leading to a much faster decay of the soil moisture correlation structure. These fluctuations are here modeled via a jitter process acting on the space-time soil moisture resulting from the soil moisture balance equation driven solely by rainfall.

Consider the soil moisture $\{\tilde{S}(\boldsymbol{u},t)\}$ as a process related to the one represented in $\{S(\boldsymbol{u},t)\}$ by

$$\tilde{S}(\boldsymbol{u},t) = Z(\boldsymbol{u},t)S(\boldsymbol{u},t), \tag{2.17}$$

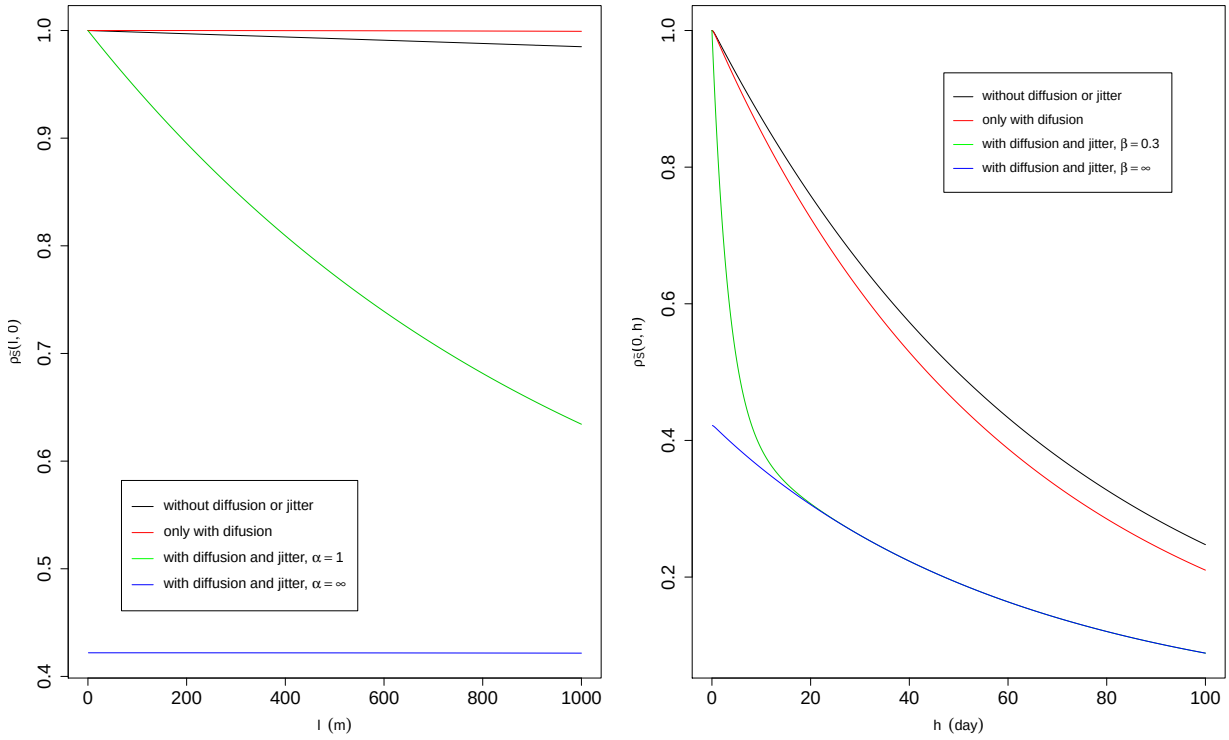
where $\{Z(\boldsymbol{u},t)\}$ is a non-negative stationary jitter process of mean μ_Z and covariance function $\Gamma_Z(s,t)$ independent of $\{S(\boldsymbol{u},t)\}$. The jitter should only increase the variability of soil moisture, but should not change its expected value. Therefore we impose that $\mu_Z = 1$, and a simple calculation shows that the covariance function of $\tilde{S}(\boldsymbol{u},t)$ is

$$\Gamma_{\tilde{S}}(l,h) = \Gamma_S(l,h) + \mu_S^2\Gamma_Z(l,h) + \Gamma_S(l,h)\Gamma_Z(l,h), \tag{2.18}$$

where μ_S and $\Gamma_S(l,h)$ are given by (2.15) and (2.16) respectively. For simplicity we assume $\{Z(\boldsymbol{u},t)\}$ has an exponential correlation structure, $\Gamma_Z(l,h) = \sigma_Z^2\exp(-\alpha l - \beta h)$ with nonnegative parameters α and β. Equation (2.18) then becomes

$$\Gamma_{\tilde{S}}(l,h) = \Gamma_S(l,h) + \mu_S^2\sigma_Z^2 e^{-\alpha l-\beta h} + \sigma_Z^2 e^{-\alpha l-\beta h}\Gamma_S(l,h). \tag{2.19}$$

A special case for this model is when $\beta = +\infty$. In this case, $\Gamma_Z(l,h) = 0$ if $h \neq 0$, which means the jitter $\{Z(\boldsymbol{u},t)\}$ is uncorrelated (white noise) in time. Then $\Gamma_{\tilde{S}}(l,h) = \Gamma_S(l,h)$ for $h \neq 0$, and the effect of jitter deflates the autocorrelation of the soil moisture at a fixed location by a factor $\sigma_S^2/\sigma_{\tilde{S}}^2$. Similarly, if $\alpha = +\infty$, the jitter is uncorrelated in space and deflates the spatial correlation of the soil moisture at a fixed time point by a factor $\sigma_S^2/\sigma_{\tilde{S}}^2$. These points are illustrated by Figures 2.4 and 2.5, which also show that increase of variability of soil moisture caused by the



(a) Spatial correlation functions

(b) Auto correlation functions

Figure 2.4: Comparison of the spatial correlation functions and the auto correlation functions of $\tilde{S}$. Here we use $k = 1$ km^2 day^{-1}, $\sigma_Z^2 = 0.1$. Other parameters are the same as in Figure 2.3.

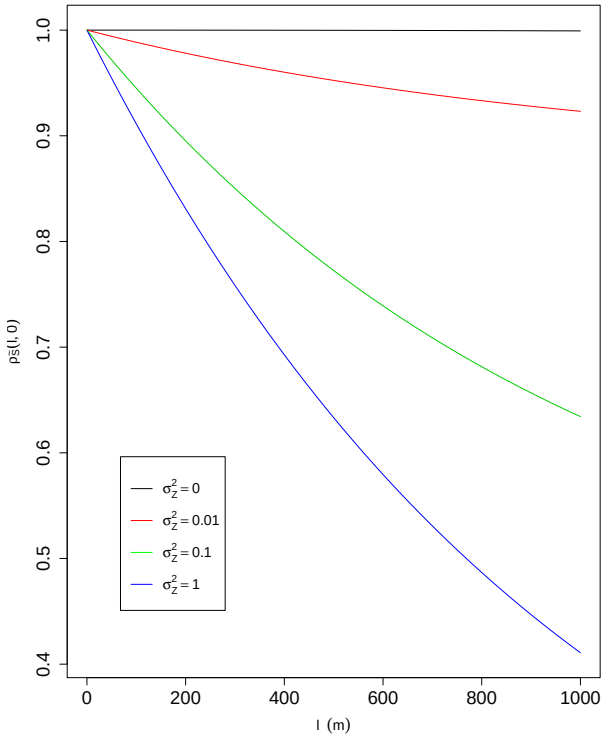


Figure 2.5: Comparison of the spatial correlation functions of $\tilde{S}$ with different values of the variance of jitter Z. Here we use $k = 1$ km^2 day^{-1}, $\alpha = 1$, $\beta = 0.3$. Other parameters are the same as in Figure 2.3.

jitter. Thus the incorporation of the jitter leads to the fast decay of the correlation structure of soil moisture at small and medium scales.

2.4.2 The spectra and hydrologic gain

From (2.19), we are able to obtain the space-time spectrum of the soil moisture with jitter, $\tilde{S}$, whose expressions can be found in Appendix A.2. Figure 2.6 shows that for small time scale ($\leqslant 1$ h), the structured spatial variability in the soil moisture model with jitter process is contained in the 0 to 1 km range. Interestingly, for larger time scales (12 h and above), some structured spatial variability is also contained at 10 km to 100 km scale. Figure 2.7a also demonstrates that all the spatial variability is contained in 0 to 100 km range. As for the temporal variability, Figure 2.7b implies that the structured temporal variability is contained in $\leqslant 10^3$ h (40 day) scales. Again, from

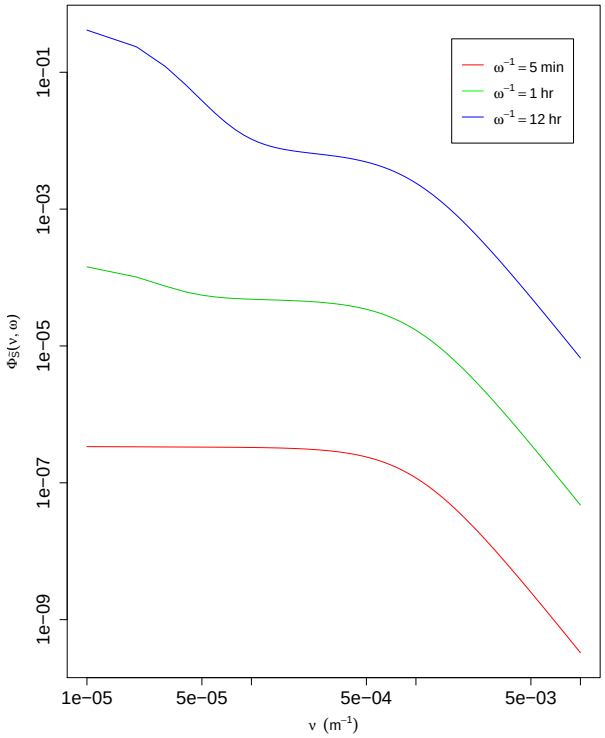


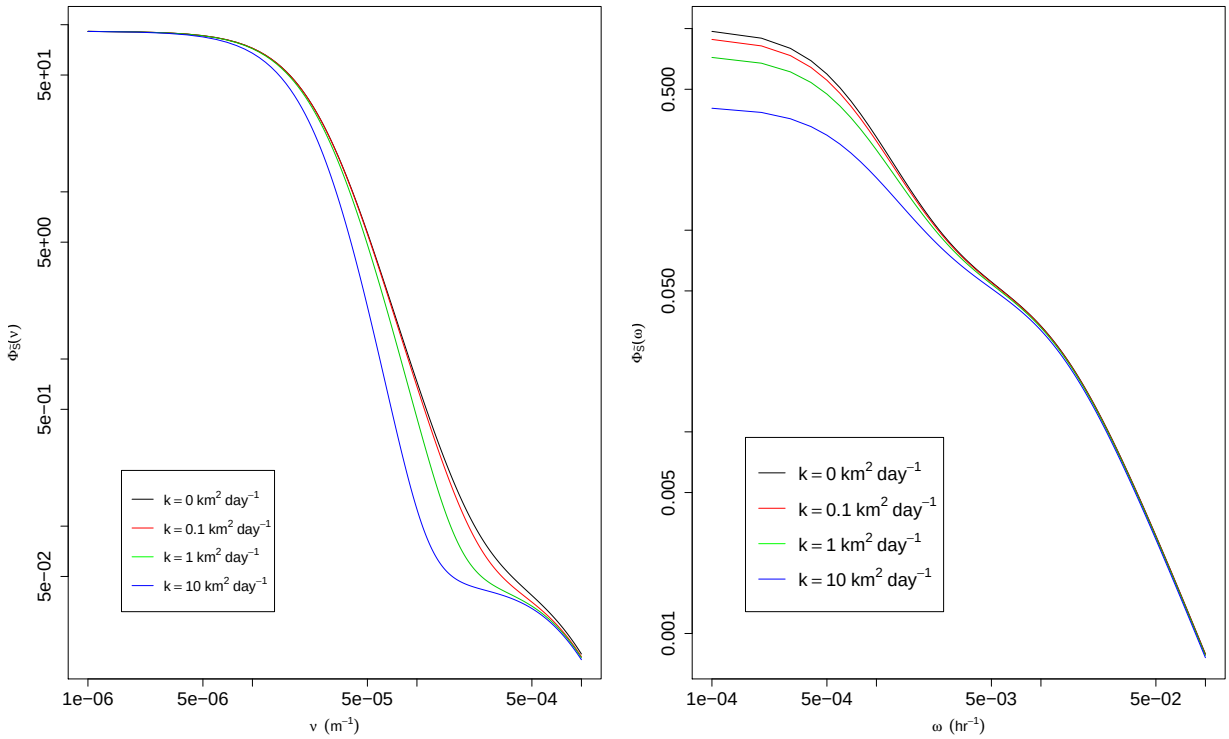
Figure 2.6: The space-time spectrum $\Phi_{\tilde{S}}(\nu, \omega)$ of soil moisture $\tilde{S}$ with different values of ω. Parameters are the same as in Figure 2.5.

Figure 2.7, we can see that only when k is of the order of 10 km^2 day^{-1} or larger that the spectrum starts being impacted by the diffusion coefficient.

[Entekhabi and Rodriguez-Iturbe, 1994] introduced the concept of hydrologic gain function, which can be used to measure the transfer of variability from the rainfall to the soil moisture via hydrologic processes operating at different scales. The hydrologic gain function is defined as

$$G_{\tilde{S}}(\nu, \omega) = \frac{\Phi_{\tilde{S}}(\nu, \omega)}{\Phi_Y(\nu, \omega)}, \tag{2.20}$$

For the soil moisture model without jitter, $G_S(\nu, \omega)$ is a monotone decreasing function, no matter what rainfall model is used ([Entekhabi and Rodriguez-Iturbe, 1994]), which lacks flexibility to describe the relation between the stochastic fluctuations of rainfall and soil moisture.



(a) Space spectrum

(b) Time spectrum

Figure 2.7: The space spectrum $\Phi_{\tilde{S}}(\nu)$ and time spectrum $\Phi_{\tilde{S}}(\omega)$ of soil moisture $\tilde{S}$ with different values of k. Parameters are the same as in Figure 2.5.

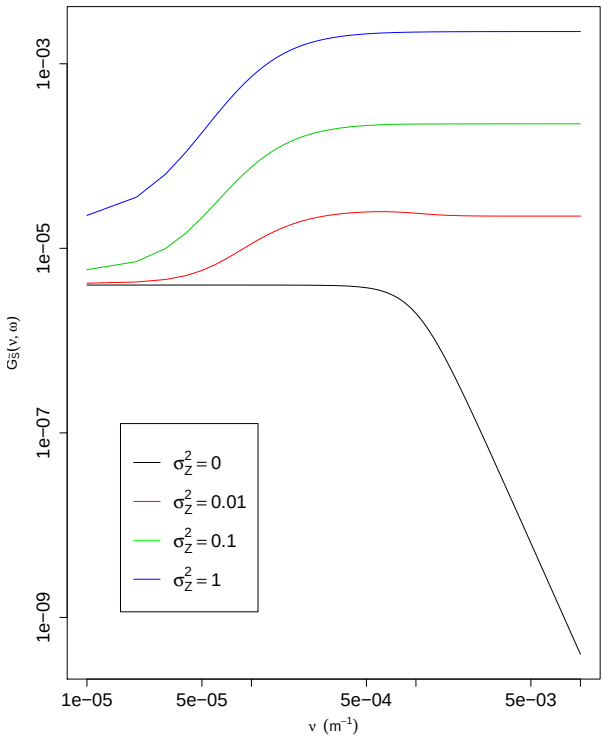
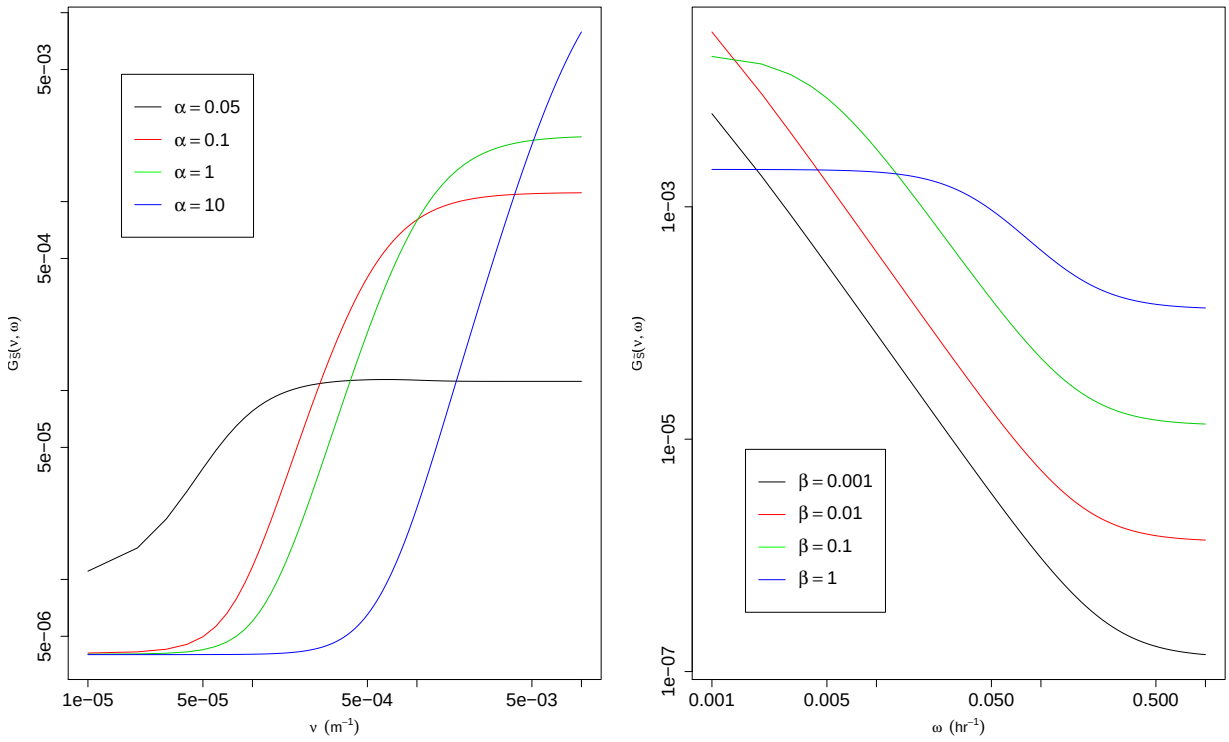


Figure 2.8: The hydrologic gain function $G_{\tilde{S}}(\nu, \omega)$ with $\omega = 1$ day^{-1}, $k = 1$ km^2 day $^{-1}$, $\alpha = 0.1$, $\beta = 0.1$ and different values of σ_Z^2. Other parameters are the same as in Figure 2.3.



(a) Fixed $\omega = 1$ day^{-1}

(b) Fixed $\nu = 10$ km^{-1}

Figure 2.9: The hydrologic gain function $G_{\tilde{S}}(\nu, \omega)$. (a), the hydrologic gain function $G_{\tilde{S}}(\nu, \omega)$ with $\omega = 1$ day^{-1}, $k = 1$ km^2 day $^{-1}$, $\sigma_Z^2 = 0.1$, $\beta = 0.1$ and different values of α. (b), the hydrologic gain function $G_{\tilde{S}}(\nu, \omega)$ with $\nu = 10$ km^{-1}, $k = 1$ km^2 day $^{-1}$, $\sigma_Z^2 = 0.1$, $\alpha = 0.1$ and different values of β. Other parameters are the same as in Figure 2.3.

Figure 2.8 shows the behavior in the hydrologic gain function at different spatial scales with daily temporal scale ($\omega^{-1} = 1$ day) and diffusion coefficient $k = 1$ km^2 day $^{-1}$ for different variances of the jitter. The case $\sigma_Z^2 = 0$ corresponds to the absence of jitter where for spatial scales larger than 1 km at the daily time scale the hydrologic processes of loss and diffusion do not alter the variability of the rainfall field when it is filtered to yield the soil moisture field with a simple scaling between the variances of the two processes. For spatial scales smaller than 1 km, the hydrologic dynamics lead to a transfer in the scales of variability where the spatial fluctuations in the soil moisture field are increasingly smoothed at smaller-local-scales. The behavior of the hydrologic gain function is fundamentally different for the case of the soil moisture with jitter. Thus when $\sigma_Z^2 = 0.1$, which is very close to the estimate we obtain in the empirical study described in Section 2.5, the hydrologic gain function is firstly increasing and then tends to a constant. The hydrologic gain function is constant at both large and small spatial scales, whose ranges depend on the variance of the jitter. Figure 2.9a shows that the spatial range where the hydrologic gain function is increasing and the constant to which it finally tends depends on the jitter parameters. The increasing part of the hydrologic gain function demonstrates that the presence of the jitter leads to an increase in the transfer of variability from the rainfall field to the soil moisture field for an important range of spatial scales controlled by the variance of the jitter. Figure 2.9a also shows that the spatial variability of the rainfall transfers into the spatial variability of the soil moisture with constant rates at both large and small scales, and the rate at smaller spatial scales is greater than the rate at larger spatial scales. This means that at smaller spatial scales, higher proportion of spatial variability of the rainfall is transferred into the spatial variability of the soil moisture through the hydrologic dynamics. Figure 2.9b shows a very different situation for the temporal scales. When $\nu^{-1} = 100$ m is fixed, the hydrologic gain function is also constant at both large and small temporal scales, but it is decreasing in some range. This implies that the rate at smaller temporal scales is less than the rate at larger temporal scales, and a higher proportion of temporal variability of the rainfall is transferred into the temporal variability of the soil moisture through the hydrologic process at larger temporal scales. The asymptotic behavior of the hydrologic gain

function can be analytically described by combining (2.10), (A.15) and (2.20) (see Appendix A.3).

2.5 Empirical study

The empirical study is based on the soil moisture data described in [Mohanty and Skaggs, 2001] and the rainfall data corresponds to the nearest stations from Mesonet. For the rainfall, we choose four Mesonet locations in Oklahoma: ACME, APAC, CHIC and NINN, and we first carry out the estimation of the rainfall parameters based on the daily measurements taken during June and July in 1997, when the soil moisture study was taken place. However, some measurements are missing and the estimations based on only one-year data are highly uncertain. Thus we use daily rainfall data at the four locations mentioned above during June and July from 1996 to 2016. Since measurements for some years are missing, we finally ended with 11 years of data and a total number of 11 (years) $\times$ 61 (days) $\times$ 4 (locations) rainfall data points. We assume that $k = 1$ km^2 day $^{-1}$ as a reasonable value of k ([Entekhabi and Rodriguez-Iturbe, 1994]) although the results are not sensitive to this assumption. The characterization of the rainfall space-time model based on this limited data should be approached with care although it is considered adequate for the purpose of this chapter. The estimation of the rainfall parameters is done as follows:

1. For each year, we estimate the rainfall parameters based on measurements of that year by the methods described in [Rodriguez-Iturbe et al., 2006]. We put all the daily rainfall data obtained during June to July in this year from 4 stations into one dataset. The mean rain cell radius ρ^{-1} is estimated by fitting the spatial correlation function of the temporally aggregated rainfall process to the spatial correlation of the dataset. We have 4 stations and we obtain 6 sample spatial correlations of 6 difference distances between different pairs of stations. Then the estimate of ρ is calculated by least squares through Equation (2.8). The mean storm duration η^{-1} is estimated by fitting the temporal correlation function of the temporally aggregated rainfall process to the temporal correlation of the dataset. For each station, we calculate the sample temporal correlation of time lag $m = 1$ day, and take the median of the 4 obtained temporal correlation as the sample temporal correlation of the whole dataset. Then the estimate of η is calculated through Equation (2.9) with

$m = 1$. To estimate λ and μ_X, we have, from Equation (2.7),

$$\lambda = \frac{2\rho^2(e^{-\eta T} + \eta T - 1)\mu_{Y^T}^2}{\pi T^2 \eta \sigma_{Y^T}^2}, \tag{2.21}$$

$$\mu_X = \frac{T\eta^2\sigma_{Y^T}^2}{4(e^{-\eta T} + \eta T - 1)\mu_{Y^T}}, \tag{2.22}$$

where $T = 1$ day and the mean μ_{Y^T} and variance $\sigma_{Y^T}^2$ of the temporally aggregated rainfall are estimated from the rainfall dataset. We obtain 11 sets of estimated values $(\hat{\rho_k}, \hat{\eta_k}, \hat{\lambda_k}, \mu_{\hat{X}_k})$, $k = 1, 2, ..., 11$.

2. The $(\hat{\rho_k}, \hat{\eta_k}, \hat{\lambda_k}, \mu_{\hat{X}_k})$'s when $\mu_{X_k} \geqslant 100$ mm day^{-1} are excluded since these outcome values of μ_X lead to physically unrealistic values of the other parameters. This is likely to happen because for extreme types of rainfall events the mathematical model is not an adequate representation of the rainfall dynamics.

3. We calculate the means of the $\hat{\rho_k}$'s and $\hat{\eta_k}$'s as the estimates of ρ and η and take the medians of the $\hat{\lambda_k}$'s and $\mu_{\hat{X}_k}$'s as our estimates of λ and μ_X. The reason for that is that the obtained $\hat{\rho_k}$'s and $\hat{\eta_k}$'s are calculated directly from the data and the obtained values are quite close to each other, namely the values are all at a reasonable scale (all ${\hat{\rho_k}}^{-1}$'s are within 10 km to 50 km and all ${\hat{\eta_k}}^{-1}$'s are within 1.5 h to 8 h). The estimates of (λ, μ_X) are calculated using the estimates of (ρ, η) through Equation (2.21) and Equation (2.22), and thus are not directly from the data. They exhibit a large variability in the values obtained for different years. Because of the robustness of medians, we use the medians instead of means as the estimates of λ and μ_X. The estimates finally obtained are: $\hat{\rho} = 0.047$ km^{-1}, $\hat{\eta} = 7.12$ day^{-1}, $\hat{\lambda} = 8.99 \times 10^{-5}$ km^{-2} day^{-1}, and $\hat{\mu}_X = 37.74$ mm day^{-1}.

For the site in consideration the rainfall interception is very low and thus we use $\phi = 0$ in Equation (2.2). Estimated values of $nZ_r = 500$ mm and $V = 7$ mm day^{-1} are used in Equation (2.2). Then, correspondingly, $a = V/nZ_r = 7/500$ day^{-1} and $b = 1/nZ_r = 1/500$ mm^{-1}. From (2.15), we obtain $\mu_{\tilde{S}} = 0.195$, which is very close to the sample mean $\bar{\tilde{S}} = 0.186$ calculated from the soil moisture data.

The soil moisture data is measured by handheld sensors at point locations within three selected

footprints of an air-borne Electronically Scanned Thinned Array Radiometer (ESTAR) during the Southern Great Plains 1997 (SGP97) Hydrology Experiment in Oklahoma ([Mohanty and Skaggs, 2001]). For this study, we use the data corresponding to the LW21 pixel, which is a silt loam, flat, split winter wheat/grass field. The pixel is a regular 7×7 square grid with 100 m spacings. The measurements are taken for each point during 25 days in June to July. However, measurements of 8 days among the 25 days are missing, and we only have $25 - 8 = 17$ days of measurements. Thus our estimation is based on 17×49 soil moisture data points.

We proceed now to the estimation of the jitter parameters $(\sigma_Z^2, \alpha, \beta)$. From (2.19), the spatial correlation function of $\tilde{S}$ is

$$\rho_{\tilde{S}}(l,0) = \frac{\Gamma_S(l,0)/(\mu_S^2 + \Gamma_S(l,0)) + \sigma_Z^2 e^{-\alpha l}}{\Gamma_S(0,0)/(\mu_S^2 + \Gamma_S(0,0)) + \sigma_Z^2}, \tag{2.23}$$

Note from Figure 2.4a that for realistic parameter values, $\Gamma_S(l,0) \simeq \Gamma_S(0,0)$ for $l \leqslant 1000$ m. Thus if we let $C_0 = \Gamma_S(0,0)/(\mu_S^2 + \Gamma_S(0,0))$, then C_0 only depends on the rainfall parameters, and (2.23) becomes

$$\rho_{\tilde{S}}(l,0) \simeq \frac{C_0 + \sigma_Z^2 e^{-\alpha l}}{C_0 + \sigma_Z^2}, \tag{2.24}$$

which allows direct estimations of σ_Z^2 and α. From the data, we can get the estimates of spatial correlations of soil moisture $\hat{\rho}_i = \hat{\rho}(l_i, 0)$ at $l_1 = 100$m, $l_2 = 100\sqrt{2}$ m,..., $l_{26} = 600\sqrt{2}$ m, which are horizontal, vertical and diagonal distances between two points on the grid. Then the estimates of α and σ_Z^2 are calculated via least squares, namely we obtain $\hat{\alpha}$ and $\hat{\sigma}_Z^2$ by minimizing the objective function

$$Q(\alpha, \sigma_Z^2) = \sum_i |\rho_{\tilde{S}}(l_i, 0) - \hat{\rho}_i|^2. \tag{2.25}$$

This yields $\hat{\sigma}_Z^2 = 0.187$, $\hat{\alpha} = 5.84$. We then estimate β through the following equation obtained from (2.19),

$$\beta \simeq -\frac{1}{h_0} \log \left(\rho_{\tilde{S}}(0, h_0) - \frac{C_0(1 - \rho_{\tilde{S}}(0, h_0))}{\sigma_Z^2} \right) \tag{2.26}$$

where we use $h_0 = 1$ day, and $\rho_{\tilde{S}}(0, h_0)$ is estimated from the soil moisture data as the average

lag one autocorrelation computed through all 49 sample points, $\hat{\rho}_{\tilde{S}}(0,1) = 0.635$. This yields $\hat{\beta} = 0.76$.

With these parameters for the jitter process, the model standard deviation is $\sigma_{\tilde{S}} = 0.105$, which is quite close to the sample standard deviation $\hat{\sigma}_{\tilde{S}} = 0.085$.

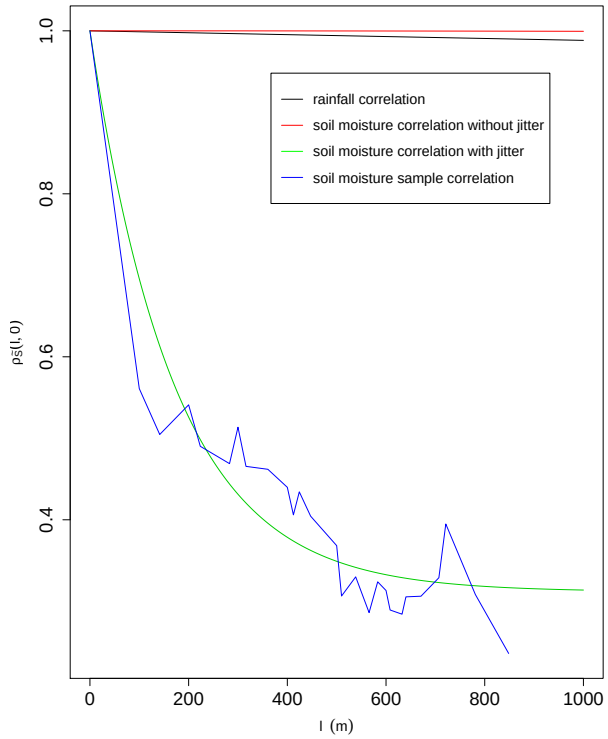


Figure 2.10: The spatial correlation function $\rho_{\tilde{S}}(l,0)$ of soil moisture with $k = 1$ km^2 day $^{-1}$, $\rho = 0.047$ km^{-1}, $\eta = 7.12$ day^{-1}, $\lambda = 8.99 \times 10^{-5}$ km^{-2} day^{-1}, $\mu_X = 37.74$ mm day^{-1}, $a = 7/500$ day^{-1}, $b = 1/500$ mm^{-1}, $\sigma_Z^2 = 0.187$, $\alpha = 5.84$ and $\beta = 0.76$.

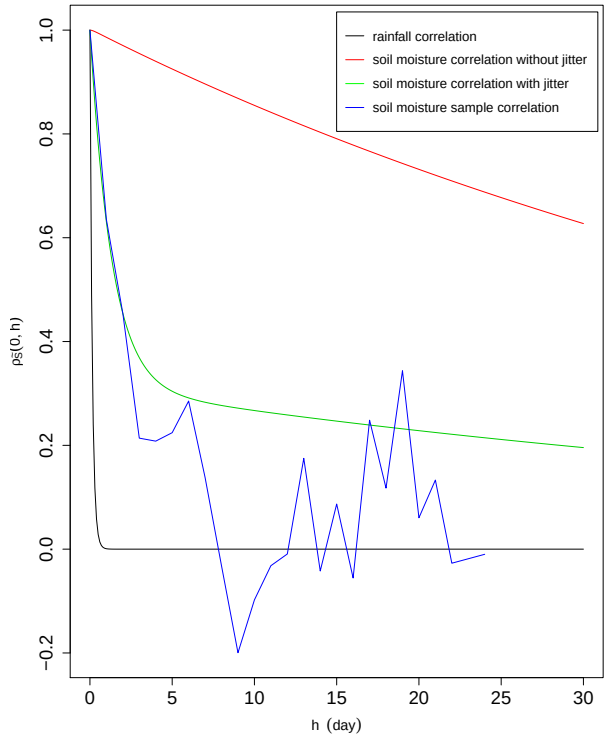
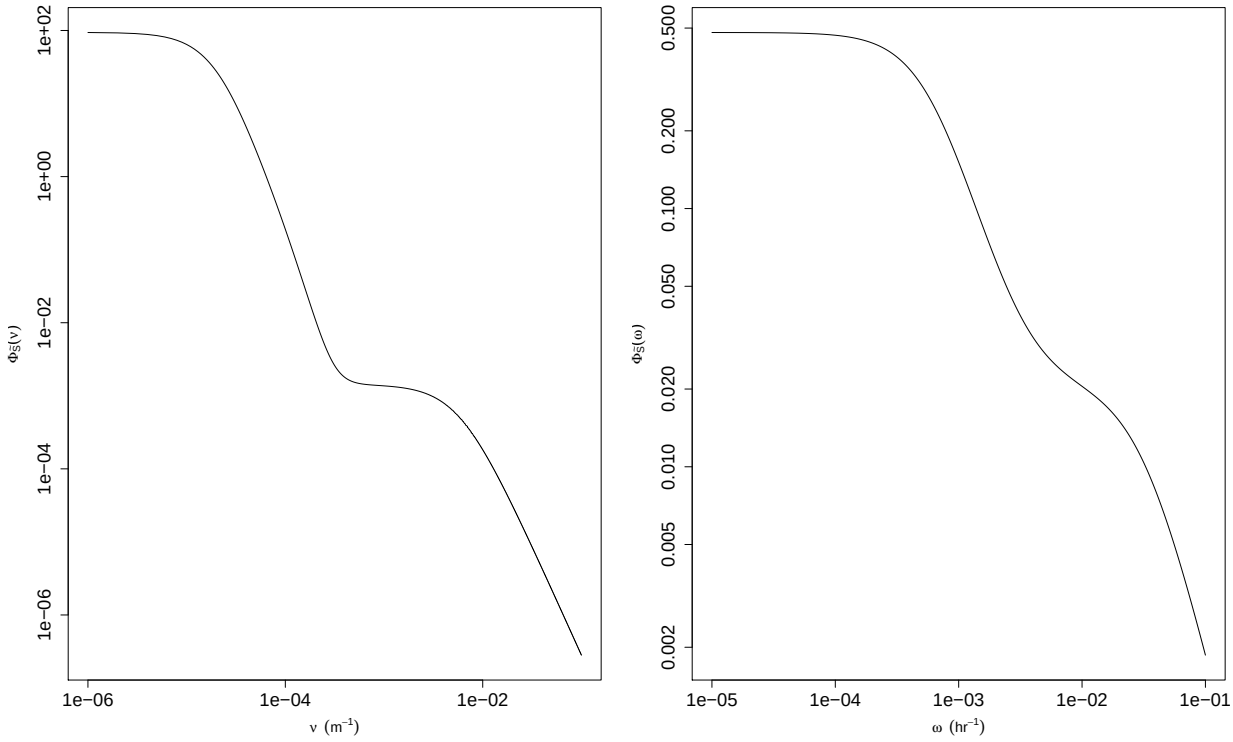


Figure 2.11: The auto correlation function $\rho_{\tilde{S}}(0, h)$ of soil moisture with the same parameters as in Figure 2.10.

Figure 2.10 shows that the spatial correlation functions of rainfall and soil moisture for models without jitter are close to 1 when $l \leqslant 1000$ m, while the spatial correlation function of soil moisture with jitter decays much faster and fits the sample spatial correlation quite well. For the auto correlation, since we only have soil moisture data for 17 days, the sample auto correlations for large time lags obtained from the data are quite unreliable. As shown in Figure 2.11 the sample correlation function oscillates wildly around zero for lag larger than 6 days. The auto correlation function of soil moisture with jitter decays much faster than that without jitter and for lag less than 5 days fits reasonably well the one obtained from the data.

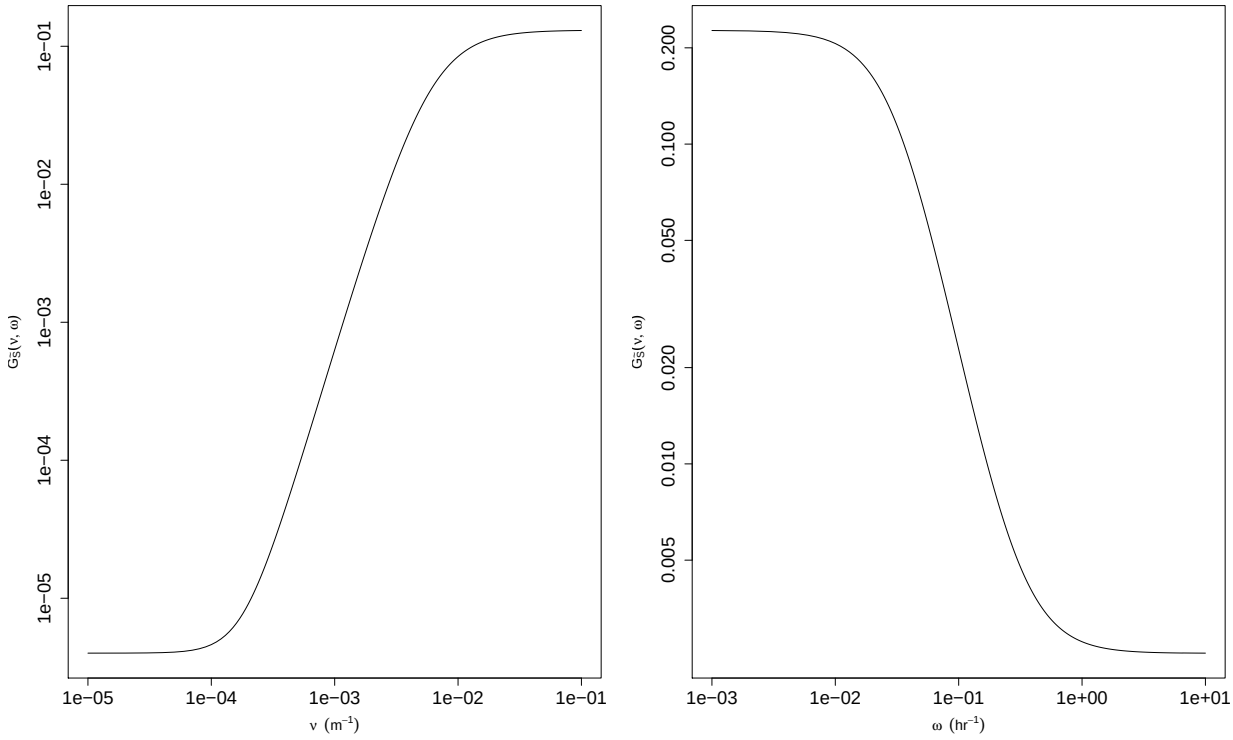
Figure 2.12a shows the behavior in the spatial spectrum of soil moisture with jitter where the structured spatial variability is contained in scales smaller than 100 km. Figure 2.12b shows the behavior of the time spectrum of soil moisture with jitter, from where it is seen that the structured



(a) Space spectrum

(b) Time spectrum

Figure 2.12: The space spectrum $\Phi_{\tilde{S}}(\nu)$ and the time spectrum $\Phi_{\tilde{S}}(\omega)$ of soil moisture with the same parameters as in Figure 2.10.



(a) Fixed $\omega^{-1} = 1$ day

(b) Fixed $\nu = 0.01$ m^{-1}

Figure 2.13: The hydrologic gain function of soil moisture $G_{\tilde{S}}(\nu, \omega)$ for fixed $\omega^{-1} = 1$ day and for fixed $\nu = 0.01$ m^{-1} with the same parameters as in Figure 2.10.

temporal variability is contained in scales less than 10^3 h (40 days).

Figure 2.13a shows the behavior in the hydrologic gain function with the estimated parameters at different spatial scales and daily temporal scale $\omega^{-1} = 1$ day. The behavior of the hydrologic gain function indicates that at the daily time scale, the spatial variability of the rainfall transfers into the spatial variability of the soil moisture with constant rates at scales larger than 10 km and smaller than 10 m and furthermore that the rate at scales smaller than 10 m is much greater than the rate at scales larger than 10 km. This means that at scales smaller than 10 m, a higher proportion of spatial variability of the rainfall is transferred into the spatial variability of the soil moisture through the hydrologic dynamics. It should be emphasized that the interpretation of the soil moisture spectra for large scales, e.g., larger than 1 km, is based on the theoretical results resulting from the model. The available data only cover spatial distances of up to 1 km. Conversely, Figure 2.13b shows a very different behavior for the hydrologic gain function with $\nu^{-1} = 100$ m. The hydrologic gain is now decreasing, and is also constant at scales smaller than 1 hr and scales larger than 100 hr (4 days). This implies that at the 100 m spatial scale, for temporal scales less than 1 hr, the temporal variability of rainfall transfers at a constant rate to the temporal variability of soil moisture. For temporal scales larger than 1 hr the transfer of temporal variability from rainfall to soil moisture increases up to scales of 100 hr where it becomes constant.

2.6 Conclusions

The commonly used space-time soil moisture models yield a transferred variability of the rainfall field to the soil moisture field that leads to spatial correlation structures which decay much slower than what is observed in the field. Furthermore, incorporating a spatial diffusion term into the stochastic soil moisture differential equation leads to an increase of the spatial correlation function at small scales, say, less than 1 km for temporal scales of 1 day. Thus although the incorporation of diffusion is reasonable it does not account for the spatial random fluctuations of soil moisture at small spatial scales (meters to hundreds of meters) which result from important local variation of surface topography, soil characteristics and vegetation conditions. The influence of all the above local factors is accounted for through the incorporation of a jitter process in the

modeling of the soil moisture dynamics.

The jitter process is controlled by three parameters, which make the model capable to represent the spatial-temporal variability of soil moisture at different scales (Figure 2.4 to Figure 2.9). The jitter process also crucially changes the hydrologic gain function of soil moisture resulting in an increase in the transfer of variability from the rainfall output to the soil moisture output via hydrologic processes operating at different scales.

An empirical study confirms the validation of the theory where the new model incorporating the jitter process successfully captures the spatial-temporal variability of soil moisture at small scales.

3. TREE CLUSTERS IN SAVANNAS RESULT FROM ISLANDS OF SOIL MOISTURE*

3.1 Introduction

"A savanna is not an ecologic middle ground between forests and grasslands, but a system with its own characteristics, including a remarkably stable coexistence of trees and grasses" ([Rodriguez-Iturbe and Porporato, 2005]). Savanna ecosystems globally cover near 33 million km^2 ([Ramankutty and Foley, 1999]) and are highly productive and flammable ([Bowman et al., 2009]). In semi-arid savannas, annual rainfall can be highly seasonal and typically between 250 mm and 750 mm of rain during the growing season ([Rodriguez-Iturbe and Porporato, 2005] and [Bowman et al., 2009]) with a pronounced inter-annual rainfall variability, which, combined with the action of fire, has led to interpretations that place savannas as non-stable transitional ecosystems. As the climate becomes drier, the trees become sparser and lower and when moister they grade into woodlands (pages 258-277 of [Scholes, 1997]). The results presented here explain the spatial structure of their vegetation as stable ecosystems with a matrix of grasses and clusters of trees following a well-defined fractal structure in their sizes and perimeters.

Tree clusters in savannas, as many other patterns in nature, may result from endogenous dynamics or exogenous forces. Small-scale patterns are frequently explained via activation-inhibition Turing type of dynamics, where diffusion or other mechanisms of movement play a key role ([Murray, 1993], [Meinhardt, 1993], [Rietkerk and van de Koppel, 2008], [Meron, 2011] and [Staver, 2018]). In addition to the existing endogenous dynamics, large-scale patterns frequently involve a response to exogenous forces, for example in the case of stochastic drivers like precipitation, which are filtered through the dominant dynamics controlling the existence of vegetation. This is also the case of many other large-scale patterns like those existing in Antarctic krill ([Levin et al., 1989]) where large scale oceanic dynamics is a key driver. In semi-arid savannas, soil moisture is the dominant exogenous variable, resulting from stochastic rainfall events filtered through

*Reprinted with permission from "Tree clusters in savannas result from islands of soil moisture" by Rodriguez-Iturbe, I., Chen, Z., Staver, A. C. and Levin, S. A., 2019. *PNAS*, 116, 6679-6683, Copyright [2020] by the National Academy of Sciences of the United States of America.

the space-time soil moisture-balance equation. Tree clusters in savannas and their fractal structure likely respond to these exogenous drivers (without denying the existence of an endogenous dynamics-e.g., competition between trees and grasses) ([February et al., 2013]). This makes the study of the probabilistic structure of soil moisture in savannas of crucial interest to explain their observed vegetation patterns. The soil moisture balance equation can be written in a simplified manner (Chapter 2) as

$$\begin{aligned} \frac{\partial S(\boldsymbol{u},t)}{\partial t} &= k\nabla^2_{\boldsymbol{u}} S(\boldsymbol{u},t) - aS(\boldsymbol{u},t) + bY(\boldsymbol{u},t), \\ \tilde{S}(\boldsymbol{u},t) &= Z(\boldsymbol{u},t)S(\boldsymbol{u},t), \end{aligned} \tag{3.1}$$

where all the terms have been normalized by nZ_r with n being the soil porosity and Z_r the effective root depth. $S(\boldsymbol{u},t)$ is the soil moisture process driven by rainfall at spatial location $\boldsymbol{u}$ and time t, k is the diffusion coefficient, $aS(\boldsymbol{u},t)$ is the moisture loss via evapotranspiration and leakage, b is the normalized infiltration coefficient, $Y(\boldsymbol{u},t)$ is the rainfall process ([Isham et al., 2005]) described in Section 2.2. The losses are treated as linear functions of soil moisture as commonly represented for semiarid regions ([Rodriguez-Iturbe and Porporato, 2005]). The diffusion term is included for completeness but for realistic values of k it has little importance in the overall dynamics (Section 2.3 and [Entekhabi and Rodriguez-Iturbe, 1994]). Within a realistic range of small k values, the correlation structure of the $S(\boldsymbol{u},t)$ is not affected (Section 2.3). $S(\boldsymbol{u},t)$ has a correlation structure with an extremely slow decay fully dominated by the rainfall input which in geographical homogeneous regions may be still of the order of 0.9 at distances of several kilometers (Section 2.3 and [Entekhabi and Rodriguez-Iturbe, 1994]). The slow decay of the rainfall spatial correlation is well represented by the Cox and Isham model ([Cox and Isham, 1988]) described in Section 2.2. When this model is the sole input, $Y(\boldsymbol{u},t)$, in the soil moisture balance equation, it yields $S(\boldsymbol{u},t)$ with an unrealistically slow decay in spatial correlation which is fully dominated by the rainfall input for distances of the order of hundred kilometers in regions with homogeneous topography and climate. It is thus necessary to incorporate the crucial impact of fluctuations in topography, soil properties and vegetation characteristics, which lead to a much faster decay of the soil moisture correlation

Parameter	Average	Wet	Dry
Rainfall model			
ρ, km^{-1}	0.047	0.047	0.047
η, d^{-1}	7.12	7.12	7.12
λ, km^{-2}·d^{-1}	8.99×10^{-5}	1.17×10^{-4}	6.29×10^{-5}
μ_X, mm·d^{-1}	37.74	43.55	26.96
Soil moisture model			
a_1, d^{-1}	0.014	0.014	0.014
b_1, mm^{-1}	0.002	0.002	0.002
a_2, d^{-2}	0.025	0.025	0.025
b_2, mm^{-2}	0.006	0.006	0.006
α	5.84	5.84	5.84
β	0.755	0.755	0.755
σ^2	0.187	0.187	0.187
Mean rainfall during growing season, mm	497.42	746.12	248.71
Mean soil moisture during growing season, case 1	0.195	0.292	0.097
Mean soil moisture during growing season, case 2	0.341	0.511	0.170

Table 3.1: Parameters of the rainfall model and the soil moisture model. The values of parameters in column "average" are estimated from real data (Section 2.5 and [Mohanty and Skaggs, 2001]). The rainfall parameters are changed to what are called the wet and dry cases when the mean growing season rainfall is 50% more or 50% less than the "average" rainfall case. Two sets of parameters are used to represent different infiltration and evapotranspiration conditions. The parameters of the jitter are considered independent of the rainfall and representative of the local fluctuations on soil, topology and vegetation in savannas (Section 2.4).

structure. These fluctuations are modelled via a jitter process ([Rodriguez-Iturbe et al., 1987]), $Z(\boldsymbol{u}, t)$, acting on the space-time soil moisture resulting from the soil moisture balance equation driven solely by rainfall (Section 2.4). The resulting soil moisture field, $\tilde{S}(\boldsymbol{u}, t) = Z(\boldsymbol{u}, t)S(\boldsymbol{u}, t)$, captures the space-time correlation structure observed in the field (see Section 2.3). The jitter process increases the variance and deflates the correlation structure of $S(\boldsymbol{u}, t)$ without changing its mean and without allowing negative values of $\tilde{S}(\boldsymbol{u}, t)$.

Recall that the parameters of the model are estimated from soil moisture data collected via handheld sensors at point locations within three selected footprints of an air-borne Electronically Scanned Thinned Array Radiometer (ESTAR) during the Southern Great Plains 1997 (SGP97) Hydrology Experiment in Oklahoma ([Mohanty and Skaggs, 2001]). The jitter parameters estimated

for the Oklahoma site (Section 2.5) have been used to characterize the jitter in the Kruger site where the unique set of data related to the structure of the special vegetation is available ([Staver et al., 2019]). The Kruger site does not have detailed soil moisture data and moreover, the representation of the impact of very local heterogeneities is likely to be similar in this semiarid/arid region. Rainfall data from the region were used to estimate the parameters of the rainfall model and two different sets of parameters were used for a and b in the soil moisture balance equation ([Rodriguez-Iturbe et al., 2006]). The parameters values are listed in Table 3.1 and their estimation is presented in Section 2.5. The two sets of evapotranspiration parameters, a_i ($i = 1, 2$), and infiltration parameters b_i ($i = 1, 2$) were used to represent different plant and soil conditions. Three different rainfall conditions were analyzed. One denoted by average represents the case of the Oklahoma site with 497 mm of growing season rainfall, the others correspond to the wet case (746 mm, growing season rainfall) and the dry case (249 mm, growing season rainfall).

3.2 Main results

We can now study via simulation the spatial variability of the soil moisture field by focusing on soil moisture islands. At any given time, soil moisture islands are defined as pixels connected through a shared edge (von Neumann neighborhood) where the soil moisture is above the chosen threshold (see Figure B.1 in Appendix B.2 for examples). The simulation scheme for the soil moisture field, $\tilde{S}(\boldsymbol{u}, t)$ is described in Appendix B.1.

Of special interest are the islands of soil moisture where trees have a high likelihood of residence. These "islands" are made up of pixels that have average soil moisture above specific thresholds chosen here as 0.2 and 0.3 for (a_1, b_1), and 0.18 and 0.35 for (a_2, b_2). These thresholds are average values of soil moisture in sandy soils where savanna trees have good conditions for stable permanency ([Rodriguez-Iturbe and Porporato, 2005]). Also, the patterns analyzed refer to $\tilde{S}_T$ which is the average soil moisture, $\tilde{S}$, over each pixel of 10 m $\times$ 10 m during a period of $T = 30$ days. For plants the time-averaged soil moisture is more relevant than the instantaneous or daily values. The resulting patterns were very similar for averaging periods of a few days to over 3 months.

It has been well established that trees in savannas occur in clusters whose areas follow power law distributions, $P(A \geqslant a) \propto a^{-p}$ ([Scanlon et al., 2007]). The p exponent was found to vary in the Kalahari between 0.89 and 1.76 ([Scanlon et al., 2007]) but more recent and precise data in the Kruger National Park region in South Africa found p to be remarkably consistent around a representative value near 1.78 under different rainfall regimes and soil characteristics ([Staver et al., 2019]).

From the simulations of the soil moisture field we find that the islands of connected pixels with average soil moisture above the chosen thresholds closely follow well-defined power laws. Figures 3.1 and 3.2 show $P(A \geqslant a)$ for different values of mean growing season rainfall, different thresholds, and different parameters for losses and infiltration. The power-law distribution fails to describe the size of the islands only when the mean of soil moisture is very small or very large compared to the threshold being used. Thus, when the mean soil moisture is relatively large, as in the wet case with threshold equal to 0.2 (mean soil moisture equal to 0.292), most of the domain is above the threshold and trees will be favored everywhere signaling the ecosystem is more likely to be a forest rather than a savanna. When the mean of soil moisture is very small and the threshold much higher, as in the dry case with threshold equal to 0.3 (mean soil moisture equal to 0.097), islands of soil moisture are very rare signaling that trees will rarely cluster in that situation.

The exponents of the power laws, $P(A \geqslant a) \propto a^{-p}$, are in the range from 0.48 to 0.7 with a typical value around 0.51 for a savanna with the assumed average rainfall regime and threshold of 0.2 for tree stable existence. It is interesting that this range of exponents is very similar to that found for Korcak's law ([Korvin, 1992], [Mandelbrot, 1982] and [Korcak, 1940]) which describes the size distribution of ocean islands as a power law with exponent in the range from 0.5 to 0.75 depending on the region of the world.

These fractal properties of the soil moisture field $\tilde{S}(\boldsymbol{u}, t)$ result from both the correlation structure of the rainfall process, $Y(\boldsymbol{u}, t)$, and the correlation structure of the jitter process, $Z(\boldsymbol{u}, t)$. Firstly, the mean of the soil moisture field is totally determined by the rainfall process and the parameters a and b. For a fixed threshold, the soil moisture islands have fractal properties only

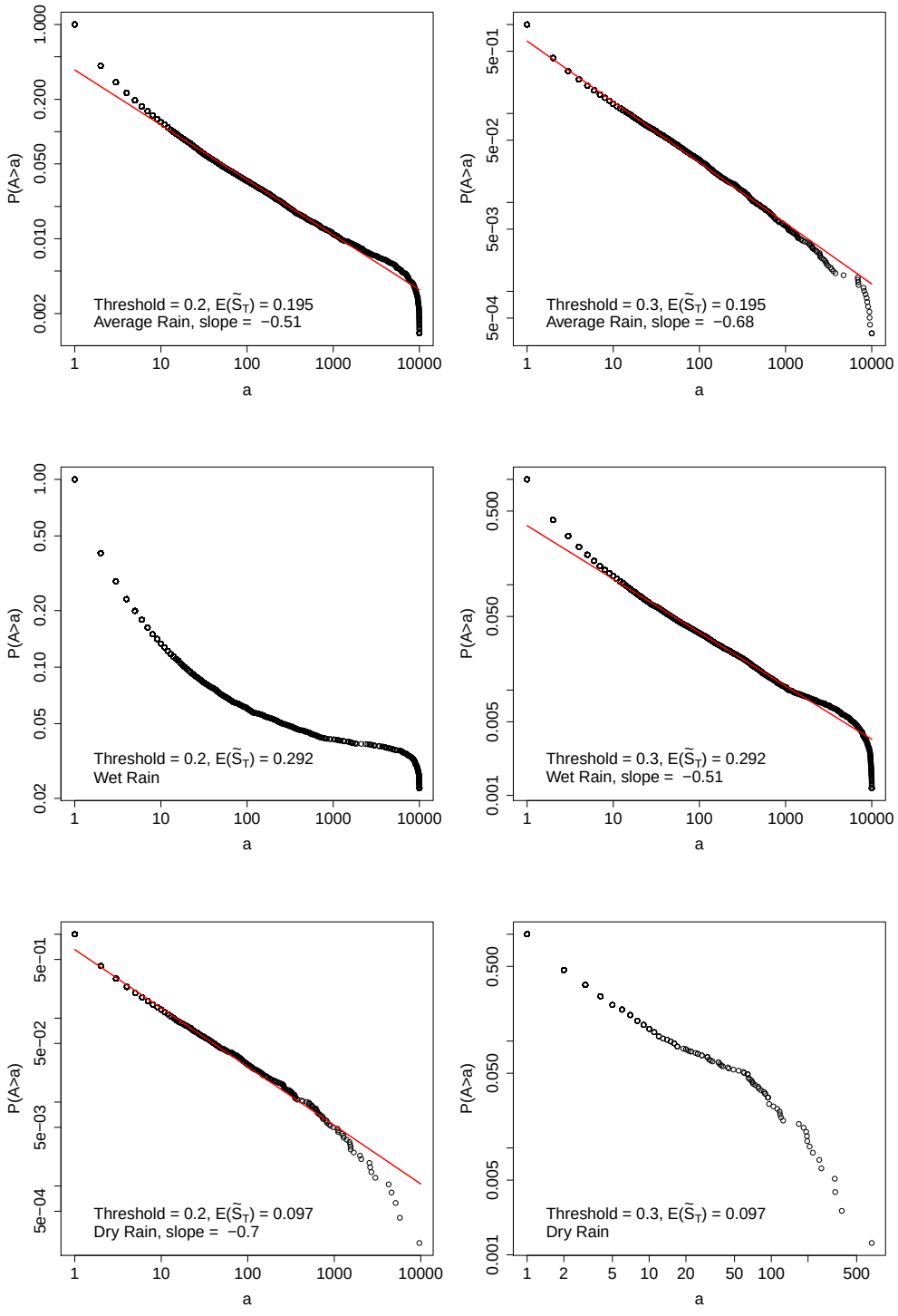


Figure 3.1: Distributions of soil moisture islands (case 1). Log-log plot of the $P(A \geqslant a)$ distribution for loss and infiltration parameters $a_1 = 0.014$ day^{-1} and $b_1 = 0.002$ mm^{-1}. Wet, average and dry season cases are considered with thresholds 0.2 and 0.3. Here the horizontal axis is the number pixels, where each pixel is 10 m $\times$ 10 m on the 1 km $\times$ 1 km field. The number of islands in the last figure is much smaller than those in the other figures.

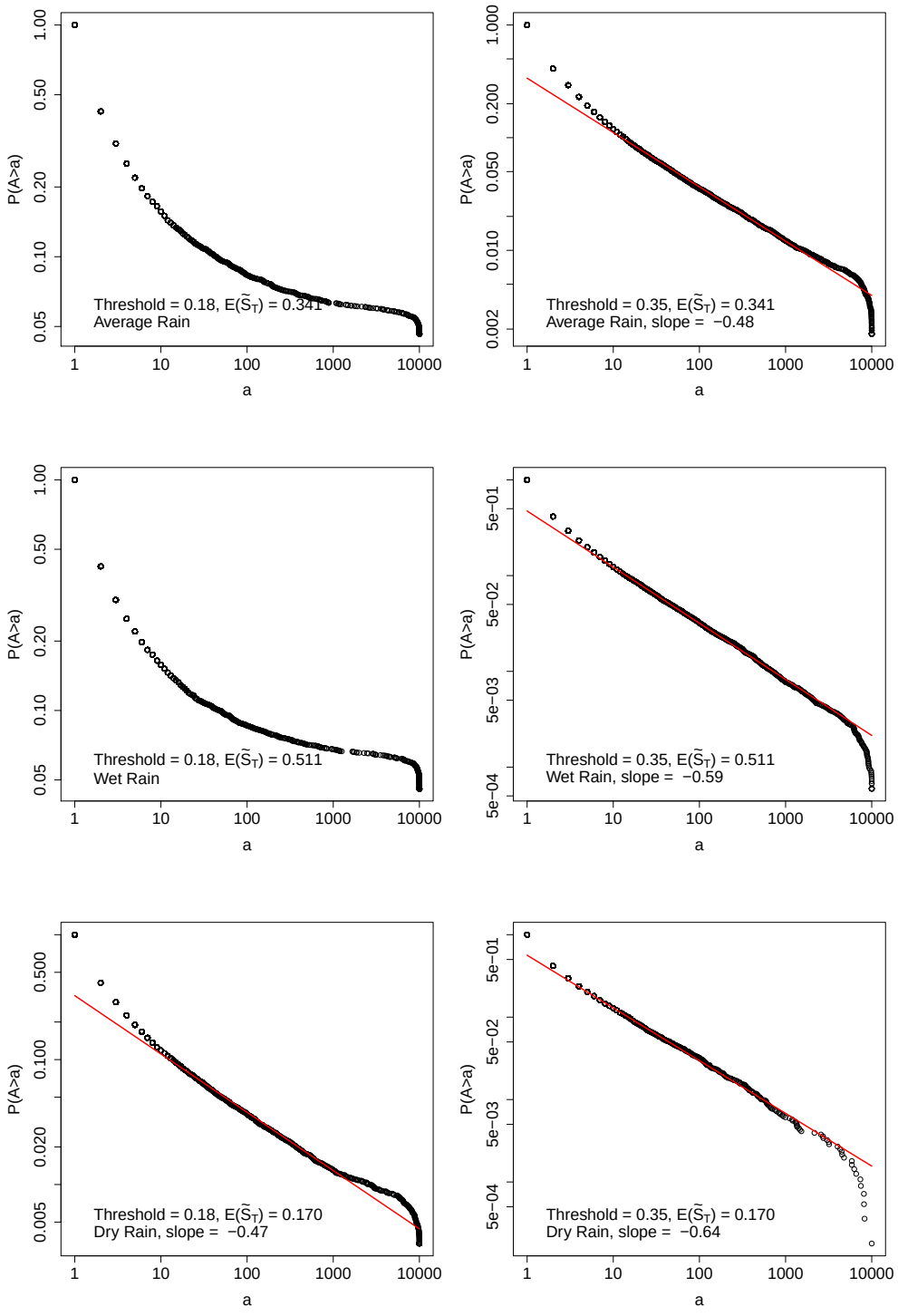


Figure 3.2: Distributions of soil moisture islands (case 2). Log-log plot of the $P(A \geqslant a)$ distribution for loss and infiltration parameters $a_1 = 0.025$ day^{-1} and $b_1 = 0.006$ mm^{-1}. Wet, average and dry season cases are considered with thresholds 0.18 and 0.35.

when the threshold is not too far above or below the mean. As shown in Figure 3.2, when the mean of soil moisture is 0.341, the soil moisture islands do not have fractal properties for a threshold of 0.18. This implies the soil moisture mean, which is independent from the jitter process, is a key factor when analyzing the fractal properties of the islands. Besides, according to Equation (2.18), the spatial correlation function of $\tilde{S}(\boldsymbol{u}, t)$, denoted by $\rho_{\tilde{S}}(l, 0)$, can be expressed as

$$\begin{aligned}\rho_{\tilde{S}}(l,0) &= \frac{(\sigma_S^2\rho_S(l,0)+\mu_S^2)\,\sigma_Z^2\rho_Z(l,0)+\sigma_S^2\rho_S(l,0)}{(\sigma_S^2+\mu_S^2)\,\sigma_Z^2+\sigma_S^2} \\ &\simeq \begin{cases}(1-\gamma)\rho_Z(l,0)+\gamma, & \text{when } \rho_S(l,0)\simeq 1,\\ \gamma\rho_S(l,0), & \text{when } \rho_Z(l,0)\simeq 0,\end{cases}\end{aligned} \tag{3.2}$$

where $\rho_S(l, 0)$ and $\rho_Z(l, 0)$ are the spatial correlation functions of the soil moisture field, $S(\boldsymbol{u}, t)$, driven solely by rainfall and that of the jitter process, $Z(\boldsymbol{u}, t)$, respectively. The terms σ_S^2, σ_Z^2 and μ_S^2 are the variances and mean of $S(\boldsymbol{u}, t)$ and $Z(\boldsymbol{u}, t)$, $\gamma = \sigma_S^2/((\sigma_S^2 + \mu_S^2)\,\sigma_Z^2 + \sigma_S^2)$ is a constant depending on both the rainfall and the jitter. As shown in Figure B.2 in Appendix B.3 and ([Isham et al., 2005]), $\rho_S(l, 0)$ and the spatial correlation of the rainfall process are very close to 1 when l is less than 1 km with the parameters estimated in Section 2.5. Thus on a 1 km by 1 km field, the rainfall process, $Y(\boldsymbol{u}, t)$, and the soil moisture field, $S(\boldsymbol{u}, t)$, are almost constant for any particular realization, and the shape of the correlation function of $\tilde{S}(\boldsymbol{u}, t)$ is totally determined by the jitter $Z(\boldsymbol{u}, t)$, as shown in the above equation and Figure B.2 (Appendix B.3). However, no matter how fast the correlation function of $Z(\boldsymbol{u}, t)$ decays for small scales, the correlation of $\tilde{S}(\boldsymbol{u}, t)$ has a lower bound $\gamma > 0$ controlled by both the jitter and the rainfall, and thus it would not die until l becomes much larger. The size of the islands resulting solely from the jitter process exhibit power-law behavior over a limited range of scales but fail to do so at larger scales due to the fast decay of its correlation function (see Figure B.3 in Appendix B.4). As shown in Figure B.2, the correlation function of $Z(\boldsymbol{u}, t)$ is already very close to 0 at $l = 1$ km. In contrast, the correlation function of the soil moisture has a very heavy tail resulting from the rainfall process, whose spatial correlation function only dies after distances of about 200 km (see [Isham et al.,

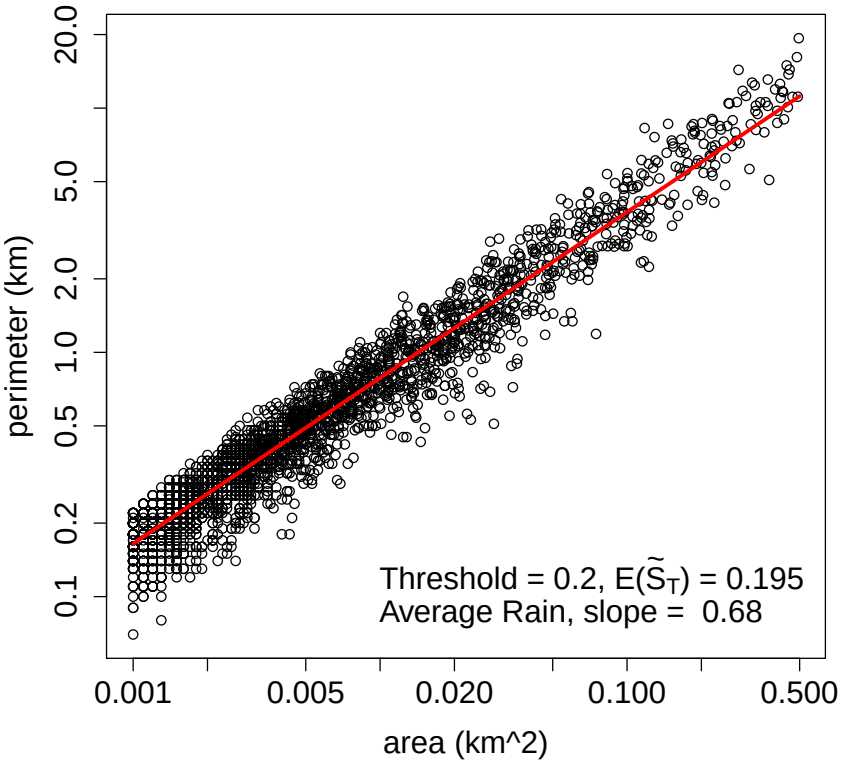


Figure 3.3: Perimeter vs. area of soil moisture islands. Here $a_1 = 0.014$ day^{-1} and $b_1 = 0.002$ mm^{-1} for the case of average rain and threshold equal to 0.2. It shows that $P \propto A^{D/2}$ with $D = 1.36$. Plots of perimeter versus area with different rainfall and thresholds are similar and can be found in Figures B.4 and B.5 (Appendix B.5).

2005] and [Rodriguez-Iturbe et al., 2006] for more details). Therefore, the islands of soil moisture would still have fractal properties at larger scales.

Trees are likely to exist in soil moisture islands whose mean is above an adequate threshold but most certainly they will not occupy the full extent of each island. Fire and herbivores are important factors that will make their occupancy smaller than that of the island. The impact of fire and herbivores will act through the perimeter, P, of the cluster. Let A' denote the area of a tree cluster, and A and P denote the area and perimeter of the soil moisture island on which the tree cluster exists. Then one may write

$$A' \propto A/P. \tag{3.3}$$

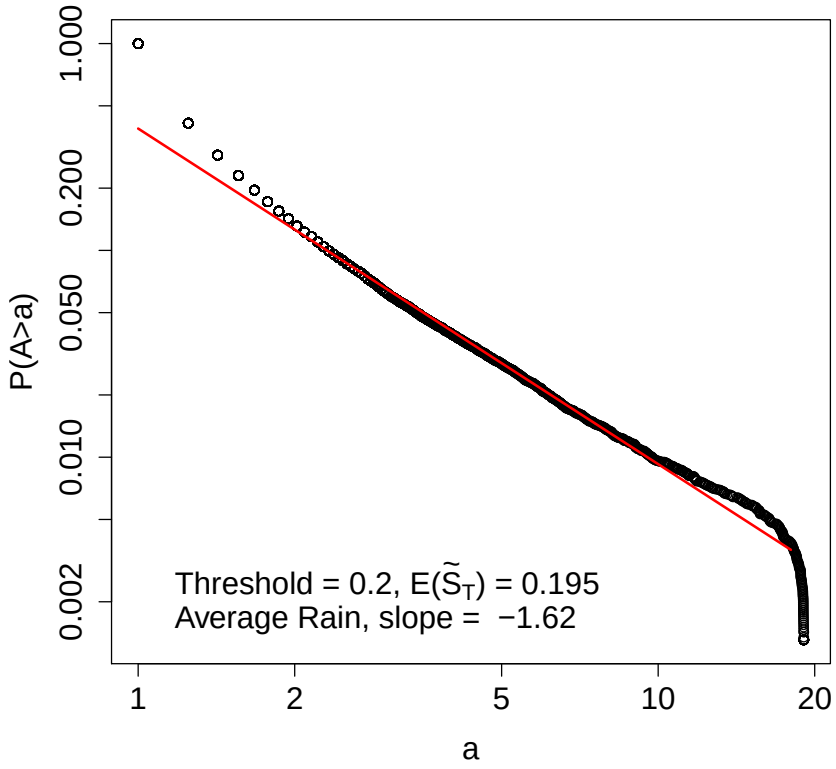


Figure 3.4: The distribution function $P(A' \geqslant a)$ for tree clusters plotted at log-log scale, where $A' \propto A^{1-D/2}$.

More complicated functional relations could be considered between A, A' and P but would be quite difficult to justify. For a fractal island Mandelbrot's Area-Perimeter equation gives $P \propto A^{D/2}$, where D is the fractal dimension of the perimeter ([Mandelbrot, 1982]). Figure 3.3 shows the Area-Perimeter relationship for a typical case of the cases analyzed before. Other cases are shown in Figures B.4 and B.5 (Appendix B.5). An excellent fit is obtained with $D/2 = 0.68$ or fractal dimension 1.36 for the perimeter of the soil moisture islands. This value is very close to the fractal dimension of 1.40 obtained for the perimeters of the tree clusters in the Kruger region with rainfall of 500 mm/year ([Staver et al., 2019]). Figure 3.3 implies that for this typical case $A' \propto A^{1-D/2} = A^{0.32}$. Figure 3.4 shows the distribution of the areas of tree clusters, A', for the typical case described before obtained from the distribution of A corresponding to soil moisture islands derived via simulations. The plot has a slope of -1.62, very close to the exponent of 1.60 observed for the power laws of tree clusters sizes in the Kruger region for the case of rainfall of 500 mm/year ([Staver et al., 2019]).

3.3 Conclusions

We infer that the emergence of large-scale spatial patterns of savanna vegetation characterized by clusters of trees in a matrix of grasses with power-law probability distribution of cluster sizes and fractal perimeters of such clusters, result from spatial patterns of soil moisture in such ecosystems. The islands of soil moisture above thresholds convenient for tree stable existence display perimeters with the same fractal dimension as that of tree clusters and their sizes follow power laws whose exponents match those found for tree clusters when the impact of fires and herbivores is accounted for. When the rainfall regime is outside the range observed in savannas the soil moisture spatial structure does not support the above conclusions reinforcing the understanding that savannas are stable ecosystems and not just transitional unstable states between forests and grasslands. Because of large interannual fluctuations in the rainfall regime the sizes of tree clusters may be affected but the fractal characteristics in their sizes and perimeters will remain relatively stable. The parameter of the cluster-size power law changes in a relatively small range that reflects changes in the rainfall regime as well as the impact of fire and herbivores.

4. DISTRIBUTIONAL PROPERTIES AND ESTIMATION IN SPATIAL IMAGE CLUSTERING*

4.1 Introduction

As shown in Chapter 3, analyses of clusters of soil, water and species have been of great interest in agriculture, ecology and hydrology. See also, for example, [Asnera and Warner, 2003], [Wootton, 2001], [Martin and Goldenfeld, 2006] and [Sole, 2007]. For instance, clusters of trees have been analyzed frequently since their properties are closely related to the environmental conditions. Many studies have been carried out based on spatial modeling (for instance, see Chapter 2) and simulations since data of clusters are often hard to collect. Real data have been used to verify whether models and simulations are capable of reproducing patterns observed in nature (see [Scanlon et al., 2007]).

In ecology, the analyses of clusters often focus on the spatial properties, such as the size of an individual cluster and the locations of the centers of clusters, assuming the object of interest is modeled by a continuous random process $y(s, \omega)$, where s is the parameter of space and ω is some sample point. See Chapter 3 for modeling and simulations, and [Staver et al., 2019] for real data analysis. For instance, $\{y(s, \omega) : s \in D, \omega \in \Omega\}$ represents the soil moisture in some area D, and $\{s \in D : y(s) > 0.2\}$ could be the set of interest where the number of trees might be relatively large. Figure B.1 in Appendix B.2 is an example for illustration. Figure B.1a is a realization of the random field y, and the clusters in red in Figure B.1b are the areas of interest (where $y > 0.19$).

However, traditional statistical cluster analysis mostly studies methods of grouping a set of objects with similar properties based on discrete data points (see [Azzalini and Torelli, 2007], [Cattelan and Varin, 2018], [Li, 2006], [McNicholas, 2016], [Menardi and Azzalini, 2014] and [Steinwart, 2015] for existing clustering methods). In contrast, we are interested in the distributional properties of the clusters based on a stochastic process defined on a continuous domain,

*Reprinted with permission from "Distributional properties and estimation in spatial image clustering" by Chen, Z. and Wang, S., 2019. *Electron. J. Stat.*, 13, 4367-4390, Copyright [2020] by Electronic Journal of Statistics.

assuming that the clusters can be easily identified. Currently there are no available statistical tools to study data about clusters in hydrology and ecology. The most common and fundamental property about this kind of cluster analysis is the size of an individual cluster, which contains much information about the environmental conditions. Data of sizes of clusters of different objects, such as canopies, have been collected and the "distribution" of the size of individual cluster has been studied by many researchers. However, the mathematical and statistical definition of the size of individual cluster has not been well defined and studied, though samples can be easily collected from images obtained from many different ways, such as remote sensors. Note that since samples of clusters are correlated, they can not be regarded as independent samples from an unknown distribution. In fact, given a well-defined spatial statistical model, it is generally difficult to define a random variable as the size of an individual cluster and study its distribution.

Without a well defined distribution function of the clusters, it is difficult for researchers to study the statistical properties of the data of image clustering and perform efficient statistical inferences. Much information in the data is not utilized, which is possible to result in inaccurate conclusions. Therefore, it is important to have a well defined distribution, and derive an efficient method to estimate the defined distribution function. Then we are able to get the distributional properties of the clusters, from which more accurate conclusions can be drawn.

The definition of the distribution function of image clusters in spatial random fields is crucial and cannot be done in a usual way, namely induced by a random variable. Thus in this chapter we define the distribution of the size of an individual cluster in a special way, without defining a random variable to be the size of an individual cluster. The estimation of the defined distribution function will be introduced and the asymptotic properties of the estimators will be investigated, which enable us to make statistical inferences and hypothesis tests.

In the following sections, basic definitions are described in Section 4.2. Then main results are presented in Section 4.3, which include some statistical properties in Section 4.3.1, the definition of the distribution function and its estimation in Section 4.3.2, and applications to Gaussian random fields in Section 4.3.3. A simulation study and a data analysis are carried out in Sections 4.4 and

4.5, respectively. Some concluding remarks are given in Section 4.6.

4.2 Preliminaries

Let D be the spatial domain and $(\Omega, \mathscr{F}, P)$ be the probability space of interest where a random process $y(s,\omega)$, $s \in D$, $\omega \in \Omega$, is defined. Without loss of generality (WLOG), we assume $D = [0,1]^2$ for simplicity. To define the clusters of interest on D, we need the definitions of open sets and connected sets as follows, which are standard definitions from point set topology ([Gemignani, 1990], Chapter 9).

Definition 4.2.1. *(Open in D) A set $S \subset D$ is called an open set in D if there exists an open set $\tilde{S} \subset \mathbb{R}^2$ such that $S = \tilde{S} \cap D$.*

Definition 4.2.2. *(Connected and disconnected in D) A set $S \subset D$ is called a disconnected set in D if it can be divided into two disjoint nonempty open sets in D, i.e., there exist A and B open in D such that $A \neq \emptyset$, $B \neq \emptyset$, $A \cap B = \emptyset$, and $S = A \cup B$. Otherwise, S is called a connected set in D.*

All open sets and connected sets are meant to be open or connected in D if not specified.

Denote the area of interest by $A = A(\omega)$, which is a subset of D depending on y. For instance, A can be the area with soil moisture greater than some value c:

$$A = A(\omega) = \{s \in D : y(s,\omega) > c\}.$$

For simplicity, we assume $c = 0$ and

$$A = A(\omega) = \{s \in D : y(s,\omega) > 0\}$$

in the following sections. However, the conclusions still hold when A is more complicated. We assume further that $\forall \omega \in \Omega$, $y(s,\omega)$ is continuous in D and $\forall s \in D$, $P(y(s) > 0) > 0$. Then for any fixed ω, $A = A(\omega)$ is open in D (an open set with respect to the topology of D). By a basic property of $\mathbb{R}^2$, A can be represented as

$$A = \bigcup_{m=1}^{M} A_m, \tag{4.1}$$

where A_m's are mutually exclusive, open and connected subsets of D, also known as the connected components of A, and M is some positive integer or ∞, depending on ω. We say s' and s'' are connected if there exists m such that $s', s'' \in A_m$. These A_m's can be regarded as "islands" in the spatial domain D. When two islands are "very close", we consider them as a cluster since they would affect each other. This consideration is reasonable in applications: think of two tree canopies which are very close. They would probably be the same species and compete with each other for groundwater. Therefore, they should be considered as one cluster. More clearly, for some fixed positive number δ, if the distance between two islands is less than δ, they should be in the same cluster. We can now define clusters formally by introducing the following relation.

Definition 4.2.3. *Suppose $s', s'' \in A$ and $\delta > 0$. We say that s' and s'' belong to the same cluster, denoted by $s' \sim s''$, if there exist $0 \leqslant n < \infty$ and $s_1, s_2, ..., s_n \in A$, such that for each $i = 1, 2, ..., n+1$, $\|s_{i-1} - s_i\| < \delta$ ($s' = s_0$, $s'' = s_{n+1}$).*

One may ask whether δ could tend to 0 when the resolution k tends to infinity. In general, in practice people would identify clusters of specific objects with $\delta = 0$. Introduce δ here is for the purpose of the theoretical derivations in probability. In applications, a fixed small δ would lead to a negligible difference compared to $\delta = 0$. For instance, when a study is focusing on the islands in oceans one can set δ equal to 1 cm, and when studying clusters of soil moisture one can set δ equal to 1 nm.

The relation $\sim$ groups together the points of A which are close to each other. Since $\sim$ is reflective, symmetric and transitive, it is an equivalence relation in A. Let $S/\sim$ denote the quotient space of a set S by an equivalence relation $\sim$. Now we can define clusters as follows.

Definition 4.2.4. *(Clusters). The equivalence classes partitioned by the equivalence relation $\sim$ in A (elements of $A/\sim$) are called clusters and denoted by $\{C_\beta\}_{\beta\in\Delta}$.*

The definition of clusters seem to be abstract and complicated at a first glance. In fact, each

cluster C_β defined in Definition 4.2.4 is just a union of "islands" (connected components) of A, which is shown in the following theorem. Let

$$d(s,S)=\inf_{s'\in S}\|s'-s\|,\quad d(S',S'')=\inf_{s'\in S',s''\in S''}\|s'-s''\|$$

denote the distance between a point and a set and between two sets in $\mathbb{R}^2$, respectively.

Theorem 4.2.1. *Suppose that $\{C_\beta\}_{\beta\in\Delta}$ are the clusters defined in Definition 4.2.4 and $\{A_m\}_{m=1}^{\infty}$ are the connected components of A. Then for any cluster $C=C_\beta$ of A, we have*

$$C=\bigcup_{i:A_i\subset C}A_i.$$

Proof. See Appendix C.1. □

Let $\lambda(\cdot)$ denote the Lebesgue measure in $\mathbb{R}^2$ and $\overline{\mathbb{R}}_+=[0,\infty)$. Then we have

Definition 4.2.5. *(Number of clusters). For $x\in\overline{\mathbb{R}}_+$, define*

$$N_x=card\left(\{\beta\in\Delta:\lambda(C_\beta)>x\}\right),$$

which is the number of clusters with Lebesgue measure greater than x.

When $x=0$, $N_x=N_0$ is the total number of clusters. Let

$$B(s,r)=\{s':\|s'-s\|<r\}$$

denote the open balls in $\mathbb{R}^2$. For each $\beta\in\Delta$, define

$$\tilde{C}_\beta=\{s\in D:d(s,C_\beta)<\delta/2\}.$$

Then $\tilde{C}_\beta$'s are open and mutually exclusive since by Definition 4.2.3 and 4.2.4, $d(C_\beta,C_{\beta'})\geqslant\delta$ if $\beta\neq\beta'$. Furthermore, each $\tilde{C}_\beta$ contains an open set $B(s_\beta,\delta)\cap D$, where $s_\beta\in C_\beta$. Therefore,

$\lambda(\tilde{C}_\beta) \geqslant \pi\delta^2/16$ (since $\lambda(B(s_\beta, \delta) \cap D) \geqslant \pi(\delta/2)^2/4$ for sufficient small δ) and

$$0 \leqslant N_x \leqslant N_0 \leqslant \frac{\lambda(D)}{\pi}\frac{16}{\delta^2} = \frac{16}{\pi\delta^2} < \infty. \tag{4.2}$$

In applications, collected data are often transferred into images with certain resolutions. Therefore, the information we have is based on pixels or grid points. In this chapter, we assume that

$$G_k = \left\{\frac{1}{2^k}, \frac{2}{2^k}, ..., \frac{2^k - 1}{2^k}, 1\right\}^2 \subset \mathbb{R}^2, \quad k = 1, 2, ...,$$

are the sets of grid points, and

$$G = \bigcup_{k=1}^{\infty} G_k = \lim_{k\to\infty} G_k$$

is the set of all grid points when the resolution goes to infinity.

Now we have similar definitions for the grid points.

Definition 4.2.6. *Suppose $s', s'' \in A \cap G_k$ for some k and $\delta > 0$. We say that s' and s'' belong to the same cluster of G_k, denoted by $s' \overset{G_k}{\sim} s''$, if there exist $0 \leqslant n < \infty$ and $s_1, s_2, ..., s_n \in A \cap G_k$, such that for each $i = 1, 2, ..., n+1$, at least one of the following two conditions is satisfied ($s_0 = s', s_{n+1} = s'', s_i = (y_i, z_i), i = 0, 1, ..., n+1$):*

(1) $|y_{i-1} - y_i| + |z_{i-1} - z_i| = 2^{-k}$ (s_i is in the Von Neumann Neighborhood of s_{i-1});

(2) $\|s_{i-1} - s_i\| < \delta$ (s_{i-1} and s_i are very close).

We say $s' \overset{G}{\sim} s''$ if there exists $K > 0$ such that for all $k > K$, $s' \overset{G_k}{\sim} s''$.

Definition 4.2.6 has one more "neighborhood" condition than Definition 4.2.3. The reason is that when identifying clusters in the continuous domain D, we only need to consider the true distance between points. However, for finite resolution k, the true image, $A(\omega) = \{s \in D : y(s, \omega) > 0\}$ is approximated by the pixels $y(G_k, \omega)$. When $2^{-k} < \delta$, no points of A_k would satisfy condition (2) of Definition 4.2.6, and each cluster can only contain one pixel. In practice, it is reasonable to group these points in $A \cap G_k$ that are neighbors when studying the cluster properties with finite resolution k. Therefore, condition (1) is added to Definition 4.2.6, although it is not necessary for

theoretical derivations for $\delta > 0$ and $k \to \infty$.

Definition 4.2.7. *(Clusters of grids). The equivalence classes partitioned by the equivalence relation $\overset{G_k}{\sim}$ in G_k (elements of $A \cap G_k / \overset{G_k}{\sim}$) are called clusters of G_k and denoted by $\{C_{\beta,k}\}_{\beta \in \Delta_k}$.*

Definition 4.2.8. *(Number of clusters on grids). For $x \in \overline{\mathbb{R}}_+$, define*

$$N_{x,k} = card\left(\{\beta \in \Delta_k : \lambda_k(C_{\beta,k}) > x\}\right),$$

where λ_k is the counting measure defined in G_k (each grid point with mass 4^{-k}).

Similarly to Equation (4.2), define

$$\tilde{C}_{\beta,k} = \{s \in D : d(s, C_{\beta,k}) < \delta/2\},$$

and $\tilde{C}_{\beta,k}$'s are open and mutually exclusive. Thus we have

$$0 \leqslant N_{x,k} \leqslant N_{0,k} \leqslant \frac{\lambda(D)}{\pi(\delta/2)^2/4} = \frac{16}{\pi\delta^2} < \infty. \tag{4.3}$$

4.3 Main results

4.3.1 Some statistical properties

To study the statistical properties of the clusters from real data, we need to ensure that when the resolution gets higher, the plot with pixels obtained from the data becomes closer to the true spatial random field. In other words, the clusters of G_k's should be almost the same as the true clusters. Since the definition of clusters is based on connectivity, we need the following theorem, which shows the relationship between the connectivity of girds and the connectivity in D.

Theorem 4.3.1. *Suppose $s', s'' \in A \cap G$. Then $s' \overset{G}{\sim} s''$ if and only if $s' \sim s''$.*

Proof. See Appendix C.2. □

Before we define the distribution function through $N_{x,k}$, we should make sure that $N_{x,k}$ is well defined, i.e., we should make sure that $N_{x,k}$ is a random variable.

Theorem 4.3.2. *For any $x \in \overline{\mathbb{R}}_+$ and $k \in \mathbb{N}_+$, $N_{x,k} : \Omega \mapsto \mathbb{N}$ is a random variable.*

Proof. See Appendix C.3. □

Let $\emptyset$ be the empty set and

$$\partial S = \{s \in D : \forall \epsilon > 0, B(s, \epsilon) \cap S \neq \emptyset, B(s, \epsilon) \cap S^c \neq \emptyset\}$$

denote the boundary of $S \subset \mathbb{R}^2$. The next lemma and theorem show the behavior of $N_{x,k}$ when the resolution k goes to infinity.

Lemma 4.3.1. *Suppose that the random process y satisfies*

$$\lambda\left(\partial A(\omega)\right) = \lambda\left(\partial\left\{s \in D : y(s, \omega) > 0\right\}\right) = 0, \quad \forall \omega \in \Omega. \tag{4.4}$$

Then for any $x \in \overline{\mathbb{R}}_+$, N_x is a random variable. Define

$$U_x = \{\omega \in \Omega : \lambda(C_i(\omega)) \neq x, \quad i = 1, 2, ..., N_0(\omega)\}, \tag{4.5}$$

where $C_1, C_2, ..., C_{N_0}$ denote the clusters of interest. Then $N_{x,k} \to N_x$ as $k \to \infty$ for all $\omega \in U_x$.

Proof. See Appendix C.4. □

Theorem 4.3.3. *Suppose that the random process y satisfies the condition in Lemma 4.3.1. Then $\forall x \in \overline{\mathbb{R}}_+$, $U_x \subset \Omega$ is measurable with respect to $\mathscr{F}$. Define*

$$V = \{x \in \overline{\mathbb{R}}_+ : P(U_x) = 1\}. \tag{4.6}$$

Then $\overline{\mathbb{R}}_+ \backslash V$ is at most countable. In other words,

$$N_{x,k} \overset{a.s.}{\to} N_x$$

except for an at most countable set in $\overline{\mathbb{R}}_+$.

Proof. According to the proof of Lemma 4.3.1, Equation (C.1) holds for all $\omega \in \Omega$. Thus for any $x \in \overline{\mathbb{R}}_+$, we can rewrite U_x as

$$U_x = \bigcup_{h=1}^{\infty} \bigcup_{j=1}^{\infty} \bigcap_{k=j}^{\infty} \left\{\omega \in \Omega : N_{x-1/h,k}(\omega) = N_{x+1/h,k}(\omega)\right\},$$

which indicates that U_x is measurable.

Now suppose $\overline{\mathbb{R}}_+ \backslash V$ is uncountable. Since

$$\overline{\mathbb{R}}_+ \backslash V = \{x \in \overline{\mathbb{R}}_+ : P(U_x) < 1\} = \bigcup_{m=1}^{\infty} \left\{x \in \overline{\mathbb{R}}_+ : P(U_x) \leqslant 1 - \frac{1}{m}\right\},$$

there exists $m_0 \in \mathbb{N}_+$ such that $\left\{x \in \overline{\mathbb{R}}_+ : P(U_x) \leqslant 1 - \frac{1}{m_0}\right\}$ is uncountable. Then we can choose a sequence

$$\{x_i\}_{i=1}^{\infty} \subset \left\{x \in \overline{\mathbb{R}}_+ : P(U_x) \leqslant 1 - \frac{1}{m_0}\right\}$$

such that $x_i \neq x_j$ if $i \neq j$. Let $U_x^c = \Omega \backslash U_x$. Then U_x^c consists of ω's such that $y(s, \omega)$ has at least one cluster with size x. Since, by Equation (4.2), the number of clusters is finite, $\forall \omega \in \Omega$, it cannot belong to infinite many sets in $\{U_{x_i}^c\}_{i=1}^{\infty}$. In other words, $\limsup_{i \to \infty} U_{x_i}^c = \emptyset$. Thus by Fatou's Lemma,

$$0 = P(\emptyset) = P(\limsup_{i \to \infty} U_{x_i}^c) \geqslant \limsup_{i \to \infty} P(U_{x_i}^c) \geqslant \limsup_{i \to \infty} \frac{1}{m_0} = \frac{1}{m_0},$$

which is a contradiction.

□

By Equation (4.2) we know that N_0 is bounded and $E(N_0) < \infty$. Besides, $P(N_0 > 0) \geqslant P(y((0,0)) > 0) > 0$ and $E(N_0) > 0$. Define

$$\tilde{F}(x) = 1 - \frac{E(N_x)}{E(N_0)}, \quad x \in V. \tag{4.7}$$

Then we have the following corollary:

Corollary 4.3.1. *Suppose that y satisfies the conditions in Lemma 4.3.1 and $x_0 \in V$. Then*

$$\lim_{\substack{x \to x_0 \\ x \in V}} N_x = N_{x_0}, \quad a.s., \tag{4.8}$$

and $\tilde{F}(x)$ is continuous in V.

Proof. Fix $\omega \in U_{x_0}$. Let $C_1, C_2, ..., C_{N_0}$ be the clusters and

$$\gamma = \min_{1 \leqslant i \leqslant N_0} |\lambda(C_i) - x_0|.$$

Then $\gamma > 0$ and when $|x - x_0| < \gamma$ and $x \in V$, $N_x = N_{x_0}$. Since $P(U_{x_0}) = 1$, we have

$$\lim_{\substack{x \to x_0 \\ x \in V}} N_x = N_{x_0}, \quad a.s.$$

Finally, by the Dominated Convergence Theorem, we have

$$\lim_{\substack{x \to x_0 \\ x \in V}} \tilde{F}(x) = 1 - \frac{1}{E(N_0)} \lim_{\substack{x \to x_0 \\ x \in V}} E(N_x) = 1 - \frac{E(N_{x_0})}{E(N_0)} = \tilde{F}(x_0).$$

Hence $\tilde{F}(x)$ is continuous in V. □

Remark 1. Now it is obvious that $\tilde{F}(x)$ has all the properties of distribution functions, but it is only defined in V. Note that V is dense in $\overline{\mathbb{R}}_+$, we can let the right limit of $\tilde{F}(x)$ be the well-defined distribution function. It may look strange that the distribution function of clusters has the form (4.7). One can think of $E(N_0)$ as the expected number of clusters, and $E(N_x)$ as the expected number of clusters with size greater than x. Then $E(N_x)/E(N_0)$ can be regarded as the proportion of clusters with size greater than x, which should be $1 - \tilde{F}(x)$ by definition. This suggests that the right limit of $\tilde{F}(x)$ should be the cumulative distribution function of the size of clusters. This idea comes from practical problems in various fields, such as agriculture, ecology and hydrology

([Pascual et al., 2002], [Scanlon et al., 2007] and [Staver et al., 2019]). More examples will be given after the Empirical Distribution Function (EDF) is defined in the next section.

4.3.2 The distribution function and its estimation

Now we are ready to formally obtain the distribution function of cluster size.

Theorem 4.3.4. *Suppose y satisfies the conditions in Lemma 4.3.1 and $\tilde{F}(x)$ is defined as in Corollary 4.3.1. Define*

$$F(x) = \inf_{z \in V \cap [x, \infty)} \tilde{F}(z), \quad x \in \overline{\mathbb{R}}_+, \tag{4.9}$$

and $F(x) = 0$ when $x < 0$. Then $F(x)$ has the following properties:

(1) $F(x)$ is non-decreasing;

(2) $F(-\infty) = 0$, $F(+\infty) = 1$;

(3) $F(x)$ is right continuous.

Therefore, $F(x)$ is a valid distribution function defined in $\mathbb{R}$.

Proof. (1) follows immediately by the definition of $F(x)$. Note that $F(x) = \tilde{F}(x) = 1$ when $x > \lambda(D) = 1$, we have $F(-\infty) = 0$, $F(+\infty) = 1$ and (2) holds. Besides, $\forall x_0 \in \overline{\mathbb{R}}_+$ and $\forall \epsilon > 0$, by the definition of $F(x_0)$, there exists $x' \in V$ such that $0 \leqslant \tilde{F}(x') - F(x_0) < \epsilon/2$. Note that $\tilde{F}(x)$ is monotone and continuous in V and that $\overline{\mathbb{R}}_+ \backslash V$ is at most countable, there exists $x'' > x'$ such that $0 \leqslant \tilde{F}(x'') - \tilde{F}(x') < \epsilon/2$. Let $\gamma = x'' - x_0$. When $0 < x - x_0 < \gamma$, since $F(x) = \tilde{F}(x)$ when $x \in V$, we have

$$\begin{aligned} 0 \leqslant F(x) - F(x_0) &\leqslant F(x'') - F(x_0) \\ &= \tilde{F}(x'') - F(x_0) \\ &= \tilde{F}(x'') - \tilde{F}(x') + \tilde{F}(x') - F(x_0) \\ &< \epsilon/2 + \epsilon/2 = \epsilon, \end{aligned}$$

which completes the proof. □

For any real function h defined in $\mathbb{R}$, let $\mathscr{C}(h) = \{x \in \mathbb{R} : h \text{ is continuous at } x\}$. Define

$$F_k(x) = 1 - \frac{E(N_{x,k})}{E(N_{0,k})}, \quad x \in \mathbb{R}. \tag{4.10}$$

The following theorem shows explicitly what V in (4.6) is and how we can approximate the true distribution function F at resolution k. Though the distribution functions are not induced by specific random variables, we will show that F_k converges to F in distribution, which means $F_k(x)$ converges to $F(x)$ for all $x \in \mathscr{C}(F)$.

Theorem 4.3.5. *Under the same conditions as in Theorem 4.3.3, we have* $V = \mathscr{C}(F) \cap \overline{\mathbb{R}}_+$. *Moreover,* $F_k(x)$ *is well-defined and*

$$F_k(x) \xrightarrow{\mathscr{D}} F(x) \quad as \quad k \to \infty.$$

Proof. First note that $0 \in V \cap \mathscr{C}(F)$. $\forall x_0 \in V \backslash \{0\}$ and $\epsilon > 0$, since $\tilde{F}$ is continuous at x_0, there exists $x', x'' \in V$, such that $x' < x_0 < x''$ and $\tilde{F}(x'') - \epsilon < \tilde{F}(x_0) < \tilde{F}(x') + \epsilon$. Let $\gamma = \min(x'' - x_0, x_0 - x')$. Then for all x satisfying $|x - x_0| < \gamma$, we have

$$F(x) - \epsilon \leqslant F(x'') - \epsilon = \tilde{F}(x'') - \epsilon < \tilde{F}(x_0) < \tilde{F}(x') + \epsilon = F(x') + \epsilon \leqslant F(x) + \epsilon.$$

Hence $x_0 \in \mathscr{C}(F) \cap \overline{\mathbb{R}}_+$ and $V \subset \mathscr{C}(F) \cap \overline{\mathbb{R}}_+$.

Now suppose $x_0 \in \overline{\mathbb{R}}_+ \backslash V$. Then by definition, $P(U_{x_0}^c) > 0$. $\forall \gamma > 0$, since $\overline{\mathbb{R}}_+ \backslash V$ is at most countable, there exist $x' \in (x_0 - \gamma, x_0) \cap V$ and $x'' \in (x_0, x_0 + \gamma) \cap V$. Note that $\omega \in U_{x_0}^c$ implies

there is at least one cluster with size x_0 and $N_{x'}(\omega) > N_{x''}(\omega)$, we have

$$\begin{aligned} F(x'') - F(x') = \tilde{F}(x'') - \tilde{F}(x') &= \frac{E\left(N_{x'} - N_{x''}\right)}{E(N_0)} \\ &\geqslant \frac{E\left(\left(N_{x'} - N_{x''}\right) 1_{\{N_{x'} - N_{x''} > 0\}}\right)}{E(N_0)} \\ &\geqslant \frac{E\left(1_{\{N_{x'} - N_{x''} > 0\}}\right)}{E(N_0)} = \frac{P(N_{x'} > N_{x''})}{E(N_0)} \\ &\geqslant \frac{P(U_{x_0}^c)}{E(N_0)} > 0. \end{aligned}$$

The last term above is a constant only depending on x_0, which indicates that $\lim_{x \to x_0} F(x) \neq F(x_0)$ and $x_0 \in \overline{\mathbb{R}}_+ \backslash \mathscr{C}(F)$. Therefore, $\overline{\mathbb{R}}_+ \backslash V \subset \overline{\mathbb{R}}_+ \backslash \mathscr{C}(F)$ and $\mathscr{C}(F) \cap \overline{\mathbb{R}}_+ \subset V$. Hence $V = \mathscr{C}(F) \cap \overline{\mathbb{R}}_+$.

Now since $P(N_{0,k} > 0) \geqslant P(y((1,1)) > 0) > 0$ when k is large, by Equation (4.3) we have $0 < E(N_{0,k}) < \infty$ and thus $F_k(x)$ is well defined in $\mathbb{R}$. To show $F_k(x) \xrightarrow{\mathscr{D}} F(x)$, it suffices to show that $\forall x \in V$, $F_k(x) \to F(x)$. Since $N_{0,k}$'s and $N_{x,k}$'s are bounded, we have, by Theorem 4.3.3 and the Dominated Convergence Theorem,

$$\lim_{k \to \infty} F_k(x) = 1 - \frac{\lim_{k \to \infty} E(N_{x,k})}{\lim_{k \to \infty} E(N_{0,k})} = 1 - \frac{E(N_x)}{E(N_0)} = F(x), \quad x \in V,$$

which completes the proof. □

We now address the problem of estimating the distribution function $F_k(x)$ for some specific k. In practice, data are often obtained as images with some fixed resolution. For instance (see Section 2.5), a 800 m by 800 m square field D is divided into 64 pixels, each pixel is 100 m by 100 m. Then a remote sensor detects the soil moisture of the center point of each pixel, and produces a image of soil moisture with resolution $k = 3$. Now let y denote the process of soil moisture and $c = 0.2$ be the threshold. After each measurement, we obtain a realization of y at 64 locations and a series of sizes of clusters $x_1, x_2, ..., x_{N_0}$. Suppose n different fields $D_1, D_2, ..., D_n$ with same size and similar soil properties are chosen, and they are far away from each other so that the soil moisture of each field is considered to be independent of each other. Then we have $y_1, y_2, ..., y_n$

that are independent and identically distributed random processes defined in $D_1, D_2, ..., D_n$, respectively. After measuring the soil moisture of field D_i, $i = 1, 2, ..., n$, a series of sizes of clusters $x_{i,1}, x_{i,2}, ..., x_{i,N_{0,k,i}}$ is obtained, where $N_{x,k,i}$ is the corresponding number of clusters with size greater than x in D_i. Then in practice, a commonly used EDF of clusters $\hat{F}_{k,n}(x)$ is defined as the sample EDF of the whole data set

$$x_{1,1}, ..., x_{1,N_{0,k,1}}, x_{2,1}, ..., x_{2,N_{0,k,2}}, ..., x_{n,1}, ..., x_{n,N_{0,k,n}}.$$

It is obvious that it is equivalent to define $\hat{F}_{k,n}(x)$ as

$$\hat{F}_{k,n}(x) = 1 - \frac{\sum_{i=1}^{n} N_{x,k,i}}{\sum_{i=1}^{n} N_{0,k,i}}. \tag{4.11}$$

We now show an asymptotic property of this EDF in the following theorem.

Theorem 4.3.6. *Suppose that $y_1, y_2, ..., y_n$ are independent and identically distributed random processes that are defined in D and satisfy the condition in Lemma 4.3.1. Define*

$$T_n = \sup_{x \in \mathbb{R}} \left| \hat{F}_{k,n}(x) - F_k(x) \right|.$$

Then we have

$$T_n \xrightarrow{a.s.} 0, \tag{4.12}$$

i.e., $\hat{F}_{k,n}$ converges to F_k almost surely uniformly.

Proof. Let $x_j = j/2^k$, $j = 0, 1, 2, ..., 2^k$, and

$$I_0 = (-\infty, x_1), \quad I_j = [x_j, x_{j+1}), \quad j = 1, 2, ..., 2^k - 1, \quad I_{2^k} = [1, \infty).$$

Then by the definition of $N_{x,k}$ (note that $\lambda_k(\cdot)$ only takes finite values), we have

$$N_{x,k} = N_{x_j,k}, \quad \forall x \in I_j$$

and

$$F_k(x) = F_k(x_j), \quad \hat{F}_{k,n}(x) = \hat{F}_{k,n}(x_j), \quad \forall x \in I_j.$$

Therefore,

$$T_n = \sup_{x\in\mathbb{R}} \left|\hat{F}_{k,n}(x) - F_k(x)\right| = \max_{0\leqslant j\leqslant 2^k} \left|\hat{F}_{k,n}(x_j) - F_k(x_j)\right| \leqslant \sum_{j=0}^{2^k} \left|\hat{F}_{k,n}(x_j) - F_k(x_j)\right|.$$

Note that by the strong law of large numbers (SLLN) we have

$$\left|\hat{F}_{k,n}(x_j) - F_k(x_j)\right| = \left|\frac{\frac{1}{n}\sum_{i=1}^n N_{x_j,k,i}}{\frac{1}{n}\sum_{i=1}^n N_{0,k,i}} - \frac{E(N_{x_j,k})}{E(N_{0,k})}\right| \overset{a.s.}{\to} 0, \quad j = 0, 1, 2, ..., 2^k.$$

Since there are only finite many j's, we conclude that

$$T_n \leqslant \sum_{j=0}^{2^k} \left|\hat{F}_{k,n}(x_j) - F_k(x_j)\right| \overset{a.s.}{\to} 0,$$

as desired. □

4.3.3 Applications to Gaussian random fields

All results of the previous section are based on the condition (4.4) in Lemma 4.3.1. Note that the area of interest $A = \{y > 0\} = \{y \geqslant 0\}\backslash\{y = 0\}$ is the difference between an excursion set and a level set, and many results of properties of level sets and excursion sets have already been obtained; see, for example, [Flores and Leon, 2010], [Worsley, 1995] and [Worsley, 1997]. The following theorem is based on Rice's Formula (see Chapter 11 of [Adler and Taylor, 2007] and [Ulrich, 1984]), which makes the condition (4.4) easy to check when y is a Gaussian random field. Let ∇y denote the almost surely gradient of y: $(\partial y/\partial s_1, \partial y/\partial s_2)$, and $\nabla^2 y$ denote the almost surely Hessian matrix of y with entries $\partial^2 y/\partial s_i \partial s_j$. The joint distribution of $(y, \nabla y, \nabla^2 y)$ is defined as the joint distribution of y, ∇y and the $2(2+1)/2 = 3$ dimensional vector vech$(\nabla^2 y)$. First of all, we introduce the following lemma, which is necessary for the proof of the next theorem.

Lemma 4.3.2. *Suppose that f is a deterministic function defined in D and $f \in \mathscr{C}^1(D)$. Let*

$$B_0 = \partial \{s \in D : f(s) > 0\}, \quad R_0 = \{s \in D : f(s) = \nabla f(s) = 0\}.$$

Then $\lambda(B_0) = 0$ if $\lambda(R_0) = 0$.

Proof. See Appendix C.5. □

Now suppose that y is a centered Gaussian random field (GRF) defined in D. Furthermore, assume that y is twice continuously differentiable almost surely, i.e., $y \in \mathscr{C}^2(D)$ a.s., and the joint distributions of $(y, \nabla y, \nabla^2 y)$ are non-degenerate. Let $C(s,t)$ denote the covariance function of y and $C_{ij}(s,t)$ denote the covariance function of $\partial^2 y / \partial s_i \partial s_j$, namely for $s, t \in D$,

$$C(s,t) = E\left(y(s)y(t)\right), \quad C_{ij}(s,t) = E\left(\frac{\partial^2 y}{\partial s_i \partial s_j}(s)\frac{\partial^2 y}{\partial t_i \partial t_j}(t)\right). \tag{4.13}$$

Then we have the main theorem of this subsection:

Theorem 4.3.7. *Suppose that, for some finite $K > 0$, $\alpha > 0$ and small enough $|t-s|$, C_{ij}'s satisfy*

$$\max_{i,j=1,2} |C_{ij}(t,t) + C_{ij}(s,s) - 2C_{ij}(s,t)| \leqslant K \left|\ln |t-s|\right|^{-(1+\alpha)} \tag{4.14}$$

and

$$|C(t,t) + C(s,s) - 2C(s,t)| \leqslant K \left|\ln |t-s|\right|^{-(1+\alpha)}. \tag{4.15}$$

Then condition (4.4) in Lemma 4.3.1 holds.

Proof. By Lemma 4.3.2, condition (4.4) is satisfied if with probability one,

$$R_0(\omega) = \{s \in D : y(s) = \nabla y(s) = 0\}$$

has zero Lebesgue measure. In Theorem 11.2.1, Corollary 11.2.2 and Lemma 11.2.12 of [Adler and Taylor, 2007], let $T = D$, $B = (0, \infty)$, $f = \nabla y$, $g = y$. Then Lemma 11.2.12 indicates

that when conditions (4.14) and (4.15) are satisfied, $R_0 = \emptyset$ with probability one. This implies $\lambda(R_0) = 0$ a.s., as desired. □

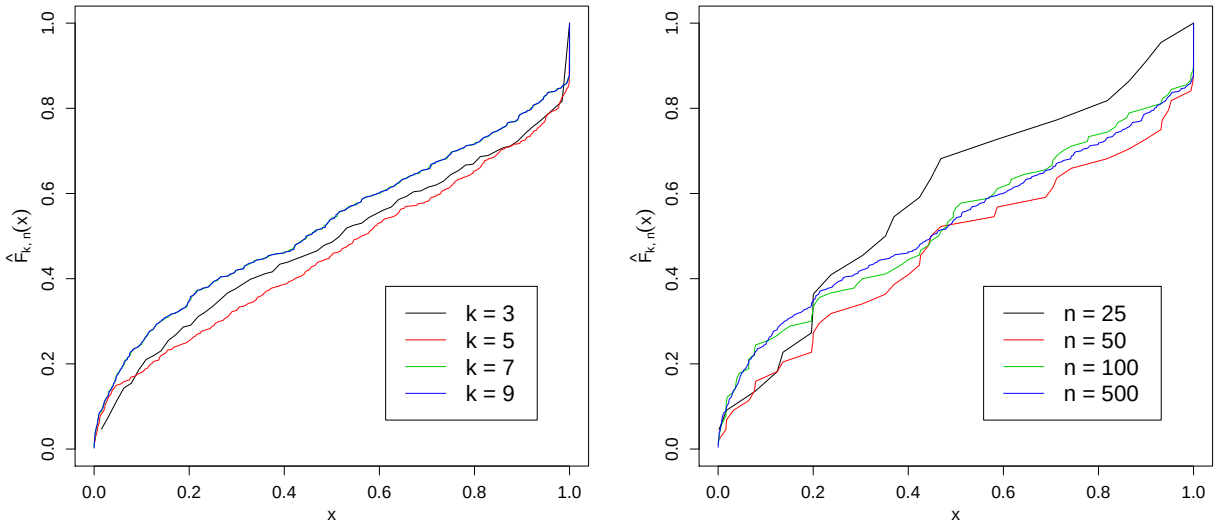
Remark 2. If y is stationary, we can let $C(t) = C(t_1, t_2)$ be the covariance function of y. By the property of GRF, C is fourth differentiable and (4.14) becomes (see Section 5.5 of [Adler and Taylor, 2007])

$$\max_{i,j=1,2} \left| \frac{\partial^4 C}{\partial^2 t_i \partial^2 t_j}(0) - \frac{\partial^4 C}{\partial^2 t_i \partial^2 t_j}(t) \right| \leqslant K \left| \ln |t| \right|^{-(1+\alpha)} \tag{4.16}$$

when t is small enough. (4.15) is not needed anymore since the differentiability of C implies that the left hand side of (4.15) is $O(|t - s|)$. Condition (4.16) is satisfied, for example, when y is isotropic and C is the Matern covariance function with $\nu > 2$.

4.4 Simulation study

Firstly, we simulated y as a Gaussian process in D with mean 0 and isotropic Gaussian covariance function $K_1(r) = e^{-r^2}$. We used $\delta = 0.001$ and y was simulated $n = 500$ times with $k = 3$, $k = 5$, $k = 7$ and $k = 9$ respectively. Then we calculated the EDFs and plotted them against x as in Figure 4.1a. After that, instead of fixing the sample size, we fix $k = 7$ and use $n = 25$, $n = 50$, $n = 100$ and $n = 500$ respectively. The EDFs with different sample sizes are shown in Figure 4.1b.



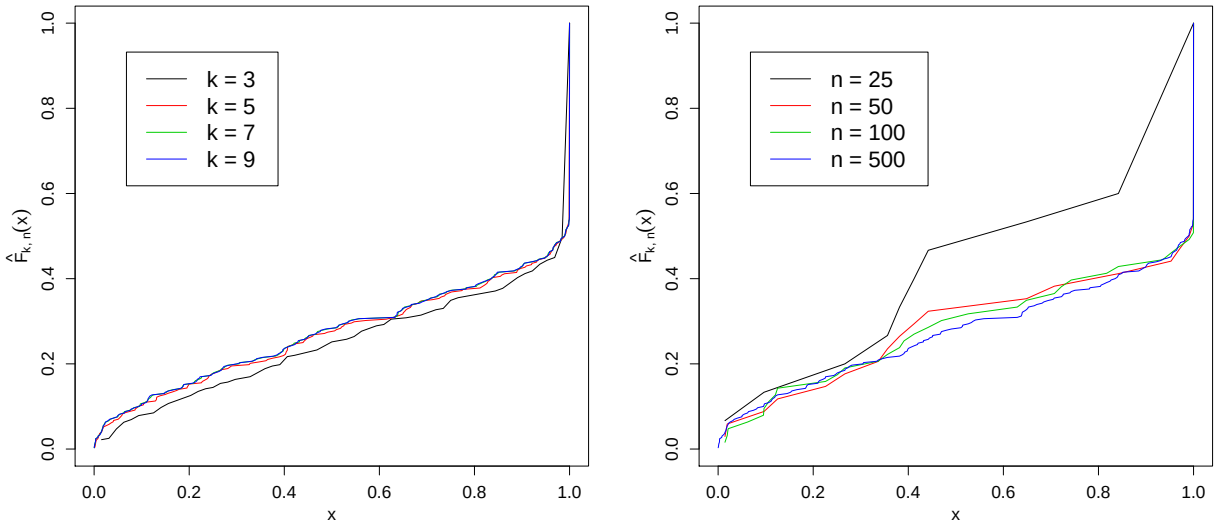
(a) EDFs with different resolutions

(b) EDFs with different sample sizes

Figure 4.1: The EDFs of the areas of clusters with covariance function $K_1(r) = e^{-r^2}$. (a), the EDFs of $n = 500$ samples with different resolutions and $\delta = 0.001$. (b), the EDFs of resolution $k = 7$ with different sample sizes and $\delta = 0.001$.

Figure 4.1a shows the convergence rate of the distribution function. When $n = 500$ is fixed, $\hat{F}_{7,n}(x)$ and $\hat{F}_{9,n}(x)$ are almost identical. This indicates that the convergence of the distribution function in resolution is quite fast. Regarding the sample sizes, Figure 4.1b shows that when $k = 7$ is fixed, $\hat{F}_{7,n}(x)$ is close to $\hat{F}_7(x)$ when $n \geqslant 100$.

We also considered the distribution function with covariance function $K_2(r) = \sin(r)/r$, which is only valid in $\mathbb{R}^d$, $d \leqslant 3$. We used the same value of δ and plotted the EDF curves for different resolutions and sample sizes, as described above. The results are shown in Figure 4.2. Though the shape of the EDF curves of K_2 are different from the curves obtained using K_1, both of them essentially converged at $k = 7$.

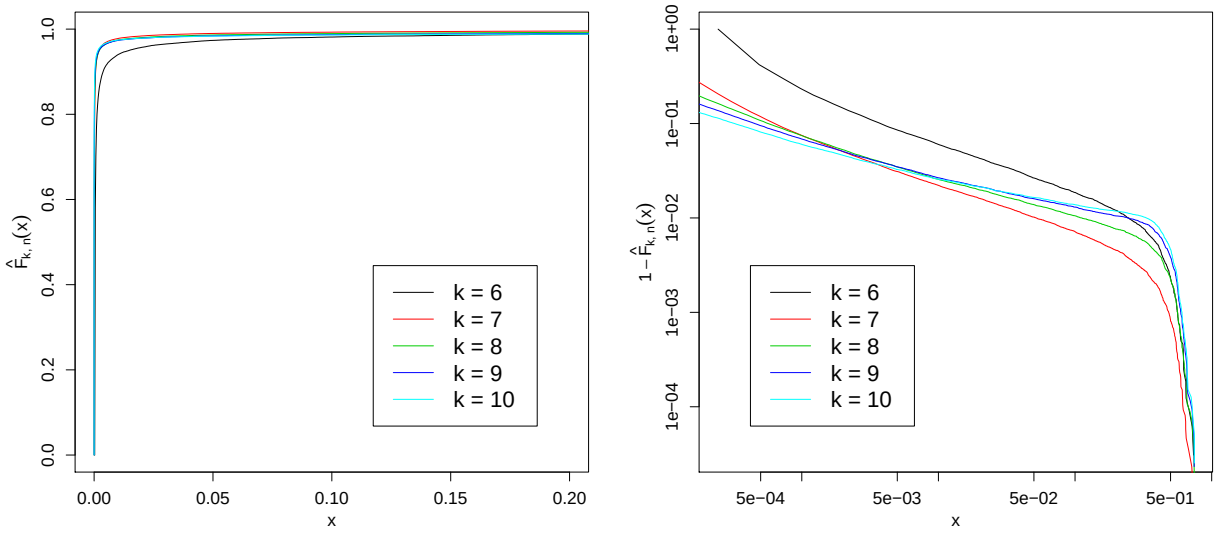


(a) EDFs with different resolutions

(b) EDFs with different sample sizes

Figure 4.2: The EDFs of the areas of clusters with covariance function $K_2(r) = \sin(r)/r$. (a), the EDFs of $n = 500$ samples with different resolutions and $\delta = 0.001$. (b), the EDFs of resolution $k = 7$ with different sample sizes and $\delta = 0.001$.

Finally we changed the covariance function to the exponential covariance function $K_3(r) = e^{-10r}$ and used $\delta = 0.01$. Note that K_3 is not differentiable at $r = 0$ and that a centered Gaussian process with covariance function $K(r) = e^{-10r}$ does not satisfy the conditions of Theorem 4.3.7 ($K_3(r) = e^{-10r}$ is not differentiable at $r = 0$ and y is not differentiable in D). We again simulated y 500 times with k from 6 to 10. The corresponding EDFs $\hat{F}_{k,n}(x)$ against x are shown in Figure 4.3a. We also plot the complimentary EDFs $1 - \hat{F}(x)$ against x in log-log scale in Figure 4.3b.



(a) EDFs in original scale

(b) complimentary EDFs in log-log scale

Figure 4.3: The EDFs of $n = 500$ samples for different resolutions and $\delta = 0.01$ with covariance function $K_3(r) = e^{-10r}$. (a), the EDFs versus x in original scale. (b), the complimentary EDFs versus x in log-log scale.

Figure 4.3 suggests that the EDF still converges as k gets large, though $K_3(r) = e^{-10r}$ does not satisfy the condition of Theorem 4.3.7. However, it converges much slower compared to the EDF with covariance $K_1(r) = e^{-r^2}$ and the EDF with $K_2(r) = \sin(r)/r$.

4.5 Data analysis

In this section we perform an analysis of the tree clusters data introduced in [Staver et al., 2019]. The tree clusters data were collected across $n = 10$ landscapes in April 2012 in Kruger National Park, South Africa, with each pixel $= 56$ cm on a side. Let $y(s)$ denote the height of the tree at location s and $y(s) = 0$ imply that there is no tree presenting at s. According to [Staver et al., 2019], the area of interest is the set of locations where there are trees with height > 3.5 meters presenting, namely

$$A = \{s \in D : y(s) > 3.5\},$$

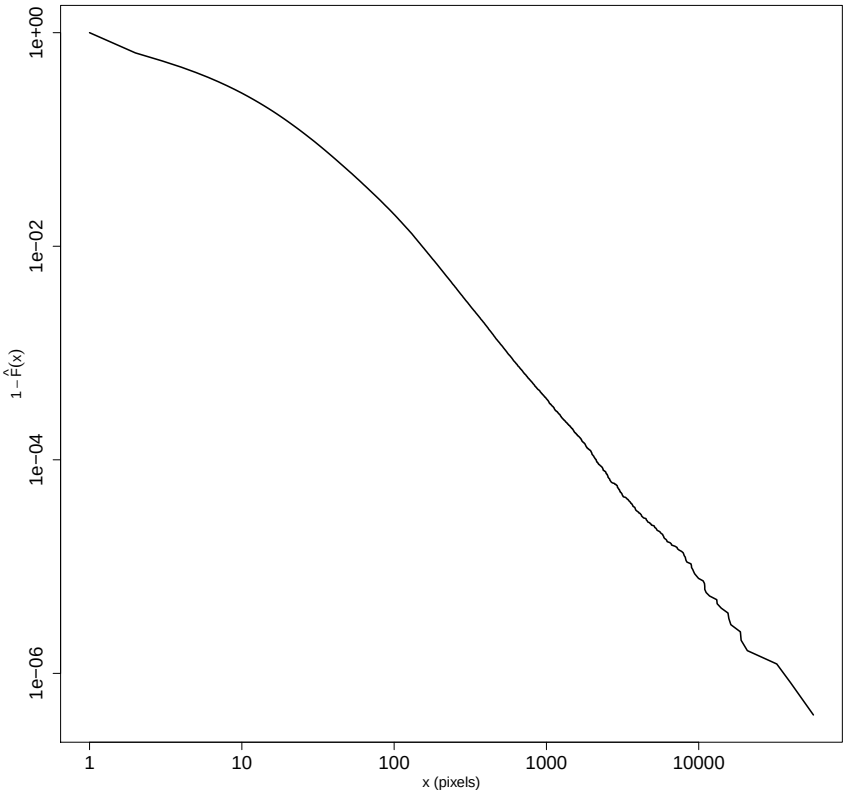


Figure 4.4: The complimentary EDF of the area of the tree clusters in the log-log scale. The horizontal axis, x, is in unit of number of pixels (56 cm $\times$ 56 cm each).

and the tree clusters are the connected components in A. The tree clusters in the data set are identified by Moore Neighborhood, which are equivalent to the tree clusters identified by Von Neumann Neighborhood with δ satisfying $\sqrt{2}l < \delta < 2l$, where $l = 56$ cm is the side length of each pixel.

The data set we analyzed is Dataset S1 in the Supporting Information of [Staver et al., 2019], which contains the areas and the perimeters of $M =$ 2,450,127 identified tree clusters. This data set is available on the PNAS website and more detailed descriptions and analyses of this data set can be found in [Staver et al., 2019].

The complimentary EDF $1 - \hat{F}(x)$ versus x is plotted in the log-log scale in Figure 4.4. We observe in Figure 4.4 that the size of tree clusters, x, may have a log-log linear relationship with

its complimentary distribution function $1 - F(x)$. Let x_i, $i = 1, 2, ..., M$, be the samples of sizes of tree clusters that we obtained, and $y_i = 1 - \hat{F}(x_i)$, $i = 1, 2, ..., M$, be the value of the complimentary EDF at x_i. A simple linear regression of $\log(y_i)$ versus $\log(x_i)$ resulted in a slope of -1.26 and $R^2 = 0.98$. This implies that for the tree clusters,

$$1 - F(x) \propto x^{-\beta},$$

and the size of tree clusters has a power-law distribution. Interestingly, a recently developed soil moisture space-time model has shown that the soil moisture clusters also have a similar power-law distributional property, indicating that the distributional properties of tree clusters result from the space-time probabilistic structure of soil moisture fields (Section 3.2).

This type of power-law clustering distribution is of great interest in various fields. For instance, the power-law cluster size distribution for the nonwetting phase in sandstones reveals the existence of ganglia of all sizes presenting a large surface area for dissolution and reaction in waterflooded oil reservoirs or CO2 storage sites ([Iglauer et al., 2010]). In addition, the power-law distribution of forest fires indicates the relationship between fire probability and population density, which can be used in forest-fire danger rating method and system ([Song et al., 2006]). Moreover, the change in the power-law distribution of vegetation patterns should be regarded as early warning signals of ecological transitions ([Kefi et al., 2014]).

4.6 Concluding remarks

In this chapter, we considered the problem of a particular and practically useful image clustering focusing on the distributional properties of clusters of spatial random fields. It is different from the traditional statistical clustering models. A formal definition of a well-defined distribution function of the clusters $F(x)$ in given in Theorem 4.3.4. The definition of distribution function at specific resolution $F_k(x)$ and the EDF $\hat{F}_{k,n}(x)$ are also defined, respectively. The asymptotic properties of these two functions under general conditions are shown in Theorems 4.3.5 and 4.3.6. This provides an efficient way to estimate $F(x)$ in applications. However, the regularity condition (4.4)

in Lemma 4.3.1 is not easy to verify. Theorem 4.3.7 shows that under the Gaussian assumption, instead of verifying condition (4.4), one can verify the smoothness of the covariance function of the Gaussian random field to ensure the asymptotic results holds.

The simulation studies demonstrated the convergence of $\hat{F}_{k,n}(x)$ with different covariance functions and different values of k and n. The results imply that the convergence rate depends highly on the smoothness of the covariance function $K(\cdot)$. When the isotropic exponential covariance function $K(r) = e^{-10r}$ is used, $\hat{F}_{k,n}(x)$ appears to converge, though $K(r) = e^{-10r}$ does not satisfy the regularity conditions of Theorem 4.3.7 ($K(r) = e^{-10r}$ is not differentiable at $r = 0$ and y is not differentiable in D).

Section 4.5 presented an analysis of a data set of tree clusters that became publicly available recently. Based on this data set, we obtained the empirical distribution function of the size of tree clusters. The result indicates that the tree clusters have a power-law distributional property, which is widely observed in many studies ([Scanlon et al., 2007] [Staver et al., 2019] and Chapter 3). A recently developed space-time model of soil moisture has indicated that the soil moisture clusters also have similar distributional properties. We conjecture that this power-law distributional property may result from some specific covariance functions, such as the isotropic exponential covariance function $K(r) = e^{-cr}$.

As a future research problem, it would be interesting to study the relationship between the model of the random field, y, and F, the distribution of the size of clusters. Assuming that y is a Gaussian Markov random field, F mainly depends on C, the covariance function of y. Studying the relationship between C and F is of great importance in many related fields since one can get the information of the random field through the image data using the relationship. In addition, in Section 4.3.2, we assume that D_1, D_2, ..., D_n are far from each other so that y_1, y_2, ..., y_n are independent. However, in applications, D_1, D_2, ..., D_n might be close to each other and y_1, y_2, ..., y_n would be correlated. In this case, how to obtain the distributional properties of the size of clusters using correlated data still remains a nontrivial question. Furthermore, in this chapter we have obtained the main results when the domain D is bounded. When the area of D tends to

infinity, stronger regularity conditions might be needed, and the asymptotic limiting distribution of cluster size could be very different from the case when D is bounded.

5. INFERENCES FOR EXTENDED PARTIALLY LINEAR SINGLE-INDEX MODELS

5.1 Introduction

Regression models are powerful tools for the analysis of the relationship between a response variable and some covariates. In many simple cases, a linear model $E(y|x) = x^T\beta$ can be used to fit the data, where x is the covariate vector and y is the response. However, the linear assumption may be violated in some cases and the relationship between the response and the covariates can be better explained via a link function g: $E(y|x) = g(x^T\beta)$. This single-index model has been proposed and studied in detail (see [Ichimura, 1993], [Horowitz and Härdle, 1996], and [Liang and Wang, 2005]). However, sometimes neither of these two models is able to sufficiently represent the relationship between the response and covariates. [Carroll et al., 1997] introduced the partially linear single-index model, which is a combination of the linear model and single-index model, and is defined by

$$y = z^T\beta + g(x^T\theta) + e, \tag{5.1}$$

where y is the response, x and z are covariates with dimensions p and q respectively, θ and β are p-dimensional and q-dimensional vectors of parameters, g is the unknown link function, and e is the random error. Different estimation and testing methodologies have been proposed for model selection and estimations of the parameters and the link function; see [Yu and Ruppert, 2002], [Xia et al., 2002], [Xia and Härdle, 2006] and [Liang et al., 2010] for more details. The partially linear single-index models have also been generalized for analysis of data with more complicated correlation structure, such as longitudinal data, which are widely studied in many fields such as epidemiology and biology. See, for example, [Wang et al., 2005], [Li and Hsing, 2010], and [Chen et al., 2015].

In practice, the covariates in the partially linear single-index models (5.1) are manually divided into two parts before the model is fitted to the data: One part becomes the covariates of the linear form z, and the other becomes the covariates of the single-index form x. However, this procedure

may result in a misspecification of the model: some covariate variables of x should actually belong to z, while some covariate variables of z should actually belong to x. The model selection methods for model (5.1) are not able to detect the misspecification after x and z are specified. This problem can be solved by using the extended partially linear single-index model, which was firstly introduced in [Xia et al., 1999] and has the form

$$y = x^T\beta + g(x^T\theta) + e. \tag{5.2}$$

Here x denotes all the covariates in the model, and it appears in both the linear part and the single-index part of the model. The extended partially linear single-index model (5.2) is an extension of the partially linear single-index model (5.1) and is able to prevent the misspecification problem mentioned above.

Since x appears both in the linear part and the single-index part, it is natural to first consider the identifiability of the model parameters before deriving estimation methodologies. [Xia et al., 1999] and [Lin and Kulasekera, 2007] have proposed and investigated the identifiability problem of extended partially linear single-index models, and obtained the regularity conditions that ensure the identifiability. [Xia et al., 1999] also proposed a simple kernel estimation for estimating the model parameters, which is similar to the estimation for the parameters of partially linear single-index models (Equation (5.1)). Since the dimension of the parameters of extended partially linear single-index models is much larger than the dimension of the parameters of partially linear single-index models, similar estimation methods would be less efficient for extended partially linear single-index models. Recently, [Dong et al., 2016] introduced a new estimation method based on orthogonal series expansion. We propose the local linear smoothing estimators for the estimation of extended partially linear single-index models and introduce the profile estimating procedure of the estimation. We show that the solution to the optimization of the profile objective function is unique and can be expressed as linear forms, which leads to fast and accurate computations for the parameter estimation.

Although extended partially linear single-index models (5.2) eliminate the possibility of the model misspecification discussed above, the number of model parameters becomes larger since all covariate variables appear twice. Therefore, it is of importance to conduct variable selection to prevent model overfitting, which may lead to biased or inefficient estimators and predictions. We propose the penalized local linear smoothing estimation for extended partially linear single-index models with the use of penalty functions, such as the least absolute shrinkage and selection operator (lasso). The estimators defined by the approach have several advantages (see Chapters 2 and 3 of [Fan and Gijbels, 1996]), and the variable selection procedure is automatically completed during the estimation procedure. As shown in the empirical study in Section 5.5, when the model parameters contain a substantial number of zeros (sparsity), the penalized estimators have better performance compared with the non-penalized estimators, even if the sample size is large. Therefore, in practice, estimating the parameters using the penalized estimators is always recommended when some covariates are known to be redundant.

After studying the estimation procedure of the model parameters, hypothesis testing for the linear constraints of the parameters is also considered. Based on the difference between the minimum values of the objective function in the null space and in the alternative space, a chi-squared type of test statistic is proposed for this type of linear hypotheses.

Our main contributions are several fold. We propose local linear smoothing estimators for extended partially linear single-index models and provide very efficient ways to compute the estimators. Moreover, we propose penalized local linear smoothing estimators to estimate the parameters with sparsity and to conduct variable selection simultaneously, which is more efficient when the model parameters contain sparsity. In addition, we introduce a chi-squared test statistic for general linear hypothesis testings for the parameters in this setting.

The rest of this chapter consists of the following sections. In Section 5.2, we propose the local linear smoothing estimating methodology for estimating β, θ and the link function $g(\cdot)$. Then we discuss the uniqueness of the solution of the optimization problem resulting from the estimation procedure, and derive the large sample theory for the estimators. In Section 5.3, we introduce the

penalized local smoothing estimators and show the asymptotic properties of the proposed penalized estimators. In Section 5.4, we provide a test statistic for the linear hypothesis testings for the model parameters. In Section 5.5, several numerical studies are conducted to assess the performance of the proposed methods. In Section 5.6, the extended partially linear single-index model is fitted to a publicly available data set about concrete slump test. Section 5.7 gives some additional remarks and concludes the chapter.

5.2 Local smoothing estimators

Formally, an extended partially linear single-index model with independent and identically distributed covariates and errors can be expressed as

$$y_i = x_i^T \beta_0 + g(x_i^T \theta_0) + e_i \quad (i = 1, 2, \ldots, n), \tag{5.3}$$

where β_0 and θ_0 are the true model parameters, (x_i, e_i)'s are independent and identically distributed pairs of covariates and errors, and x_i and e_i are independent. For identifiability, we assume θ_0 has unit L_2 norm, namely $\|\theta_0\| = 1$, the first element of θ_0 is positive and θ_0 is orthogonal to β_0, namely $\beta_0^T \theta_0 = 0$ (see [Lin and Kulasekera, 2007] for more details).

The nonparametric link function $g(u)$ and its derivative $g'(u)$ are estimated by local smoothing estimation as

$$(\hat{g}(u|\beta, \theta), \hat{g}'(u|\beta, \theta)) = \arg\min_{a,b} \sum_{i=1}^{n} \left\{ y_i - x_i^T \beta - a - b \left(x_i^T \theta - u \right) \right\}^2 K_h \left(x_i^T \theta - u \right),$$

where $K_h \left(x_i^T \theta - u \right) = K \left(\left(x_i^T \theta - u \right) / h \right) / h$ with a symmetric kernel function $K(\cdot)$ and bandwidth h. By basic calculations, $\hat{g}(u|\beta, \theta)$ can be expressed as (see [Chen et al., 2015] for similar results for partially linear single-index models)

$$\hat{g}(u|\beta, \theta) = \sum_{i=1}^{n} s_i(u|\theta) \left(y_i - x_i^T \beta \right),$$

which may be abbreviated as $\hat{g}(u)$ when no confusion arises, where $s_i(u|\theta)$ depends on the kernel

function and the observed data but is independent of β (see equations (5.8) and (5.9) below). Then the local smoothing estimators of β_0 and θ_0 can be computed through

$$\left(\hat{\beta}, \hat{\theta}_1\right) = \arg\min_{\beta,\theta} G(\beta, \theta) = \arg\min_{\beta,\theta} \sum_{i=1}^{n} \left\{y_i - x_i^T\beta - \hat{g}(x_i^T\theta|\beta, \theta)\right\}^2, \tag{5.4}$$

with restriction $\hat{\beta}^T\hat{\theta}_1 = 0$. The estimate can be done in two steps with profiling. The first step is to fix θ and calculate

$$\hat{\beta}_\theta = \arg\min_{\beta^T\theta=0} G(\beta, \theta). \tag{5.5}$$

Then the estimator $\hat{\theta}_1$ is

$$\hat{\theta}_1 = \arg\min_{\theta} \sum_{i=1}^{n} \left\{y_i - x_i^T\hat{\beta}_\theta - \hat{g}(x_i^T\theta|\hat{\beta}_\theta, \theta)\right\}^2 = \arg\min_{\theta} G(\hat{\beta}_\theta, \theta). \tag{5.6}$$

Finally, we standardize $\hat{\theta}_1$ and get the estimator of θ by

$$\hat{\theta} = \hat{\theta}_1/||\hat{\theta}_1||. \tag{5.7}$$

The asymptotic properties of the local smoothing estimators are of great interest. We present the following regularity conditions to show the asymptotic normality and consistency of the local smoothing estimators.

Regularity Assumptions:

Assumption 1. The density of $x^T\theta$, $f_\theta(\cdot)$, is positive, bounded away from 0 and second continuously differentiable in $\mathscr{U} = \left\{x^T\theta : \theta \in \Theta, x \in \mathscr{X}\right\}$, where Θ is the compact parameter space of θ and $\mathscr{X}$ is the compact support of x.

Assumption 2. For any $\theta \in \Theta$, the second derivative of the function $\rho_x(u|\theta) = E\left(x|x^T\theta = u\right)$ with respect to u is bounded and continuous.

Assumption 3. The link function $g(\cdot)$ is twice continuously differentiable and $g''(\cdot) \not\equiv 0$ on an open subinterval in $\mathscr{U}$.

Assumption 4. The function $g(x^T\theta)$ and the density of $x^T\theta$, $f_\theta(x)$, are both three times continuously differentiable with respect to x. The third derivatives are uniformly Lipschitz continuous over $\Theta \subset \mathbb{R}^p$ for all $x \in \mathscr{X}$.

Assumption 5. The kernel function $K(\cdot)$ is a symmetric, bounded and continuously differentiable probability density function. Furthermore, $K(\cdot)$ is positive on the whole real line, $\mathbb{R}$, and $\int_{\mathbb{R}} |v|^i (K(v))^j dx < \infty$ $(i, j = 1, 2)$.

Assumption 6. The variance of e, σ^2, is positive, and $E(|e|^\gamma) < \infty$ for some $\gamma \geqslant 3$.

Assumption 7. The bandwidth h satisfies $nh^6 \to 0$ and $nh^{3+3/(\gamma-1)} / \log n \to \infty$ as $n \to \infty$.

One key question is whether Equation (5.5) has a solution and if so, whether the solution is unique. Theorem 5.2.1 below shows that there exists a unique solution to Equation (5.5), and provides an efficient way to calculate the solution. The following definitions are necessary for the introduction of Theorem 5.2.1.

Denote the covariance matrix $X = (x_1, x_2, \ldots, x_n)^T$ and the response vector $Y = (y_1, y_2, \ldots, y_n)^T$, and let $\mathscr{C}(X)$ be the column space of X. Define $T = X\theta$ and $n \times n$ matrix D_θ with entries

$$(D_\theta)_{ij} = \frac{\left(\sum_{s=1}^n K_{is} t_{is}^2\right) K_{ij} - \left(\sum_{s=1}^n K_{is} t_{is}\right) K_{ij} t_{ij}}{\left(\sum_{s=1}^n K_{is}\right)\left(\sum_{s=1}^n K_{is} t_{is}^2\right) - \left(\sum_{s=1}^n K_{is} t_{is}\right)^2} \quad (i, j = 1, 2, \ldots, n), \tag{5.8}$$

where

$$t_{ij} = t_j - t_i = x_j^T\theta - x_i^T\theta, \quad K_{ij} = K_h(t_{ij}) \quad (i, j = 1, 2, \ldots, n).$$

When $T = X\theta \neq c1_n$ for any $c \in \mathbb{R}$, where $1_n = (1, 1, \ldots, 1)^T \in \mathbb{R}^n$, by the Cauchy–Schwarz inequality, we have

$$\left(\sum_{s=1}^n K_{is}\right)\left(\sum_{s=1}^n K_{is} t_{is}^2\right) - \left(\sum_{s=1}^n K_{is} t_{is}\right)^2 > 0.$$

Then for any $Z = (z_1, z_2, \ldots, z_n)^T \in \mathbb{R}^n$ and $i = 1, 2, \ldots, n$, the optimization problem

$$\left(\hat{a}_i, \hat{b}_i\right) = \underset{a_i, b_i \in \mathbb{R}}{\arg\min} \sum_{j=1}^n \left(z_j - a_i - b_i t_{ij}\right)^2 K_{ij},$$

has unique solution $\hat{a}_i$ which can be expressed as $\hat{a}_i = \sum_{j=1}^{n} (D_\theta)_{ij} z_j$. When $Z = Y - X\beta$, we have

$$\hat{g}(x_i^T\theta) = \sum_{j=1}^{n} (D_\theta)_{ij} \left(y_j - x_j^T\beta\right). \tag{5.9}$$

Let $\tilde{X}_\theta = (I_n - D_\theta) X$ and $\tilde{Y}_\theta = (I_n - D_\theta) Y$, where I_n is the $n \times n$ identity matrix. Assume $\theta \neq 0$. Besides, let B^+ be the Moore–Penrose inverse of any matrix B. The following theorem gives an explicit simple expression of the estimators and provides an efficient way to calculate them.

Theorem 5.2.1. *Suppose $n > p \geqslant 2$, $1_n \notin \mathscr{C}(X)$, $\operatorname{rank}(X) = p$ and $K(\cdot) > 0$. Then the optimization problem in (5.5) has a unique solution expressed as*

$$\hat{\beta}_\theta = \left(\tilde{X}_\theta^T \tilde{X}_\theta\right)^+ \tilde{X}_\theta^T \tilde{Y}_\theta = \left(\tilde{X}_\theta^T \tilde{X}_\theta + \theta\theta^T\right)^{-1} \tilde{X}_\theta^T \tilde{Y}_\theta. \tag{5.10}$$

Equation (5.10) provides two different methods for calculating $\hat{\beta}_\theta$ when θ is fixed. The latter equation implies that the solution $\hat{\beta}_\theta$ can be obtained by solving a linear system, which is very efficient and accurate. However, to solve the linear system

$$\left(\tilde{X}_\theta^T \tilde{X}_\theta + \theta\theta^T\right) \beta = \tilde{X}_\theta^T \tilde{Y}_\theta \tag{5.11}$$

in any software, for example in R, the coefficient matrix should be nonsingular. This requires $K(\cdot)$ to be strictly positive on the whole real line $\mathbb{R}$. Many functions with good properties satisfy this condition, such as the standard normal density $\phi(\cdot)$, but they would tend to 0 very quickly since they satisfy the regularity conditions. For instance, $\phi(v)$ would be close to 0 when $|v| > 3$. Consequently, when the sample size is not large enough and h is relative small, many entries of D_θ, defined in (5.8), would be very close to 0. This would cause computational issues since it might make the coefficient matrix in (5.11) close to singular. Therefore, in these cases, though calculating the Moore–Penrose inverse numerically might be inefficient and lead to larger computational errors, we need to use it in order to get rid of singularity issues.

Since we are estimating θ_0 with $||\theta_0|| = 1$, alternatively, we can minimize the profile objective function only on the unit ball:

$$\hat{\theta}_2 = \arg\min_{||\theta||=1} \sum_{i=1}^{n} \left\{ y_i - x_i^T \hat{\beta}_\theta - \hat{g}(x_i^T \theta | \hat{\beta}_\theta, \theta) \right\}^2.$$

With a reasonably chosen bandwidth h, the linear system (5.11) can be solved much more easily with the restriction $||\theta|| = 1$, especially when the dimensions of the parameters are high. Since the solution $\hat{\theta}_2$ is not a universal optimizer, it is expected to be less efficient compared with $\hat{\theta}$ obtained from (5.6) and (5.7). However, $\hat{\theta}_2$ is still useful as it can be easily calculated and used as the initial value in the calculation of $\hat{\theta}$.

An estimating procedure based on kernel smoothing was introduced in [Xia et al., 1999]. This kernel smoothing estimation has an objective function similar to Equation (5.4), with $\hat{g}$ based on kernel smoothing estimation. Therefore, by profiling, one can also obtain the profile estimator $\tilde{\beta}_\theta$ for each fixed θ, as shown in Equation (3.2) of [Xia et al., 1999]. However, this profile estimator is obtained by optimizing the objective function without the constraint $\beta^T \theta = 0$, which implies that the profile kernel smoothing estimator $\tilde{\beta}_\theta$ is not guaranteed to be orthogonal to θ. Therefore, due to identifiability issues, $\tilde{\beta}_\theta$ might not be close to the true value β_0, even if θ is very close to θ_0. This issue is indicated by the simulation results presented in Section 5.5. To resolve this issue, we add the condition $\beta^T \theta = 0$ when optimizing the objective function S_n proposed in [Xia et al., 1999], and implement the method of Lagrange multipliers to calculate the estimators. The performance of this Lagrange kernel smoothing estimators is also assessed as shown in Section 5.5.

The following theorem shows the asymptotic normality of the local smoothing estimators.

Theorem 5.2.2. *Suppose that the regularity assumptions 1-7 are satisfied. Then we have, as* $n \to \infty$*,*

$$n^{\frac{1}{2}} \begin{pmatrix} \hat{\beta} - \beta_0 \\ \hat{\theta} - \theta_0 \end{pmatrix} \xrightarrow{\mathscr{D}} N(0, \sigma^2 \Gamma^+),$$

where

$$\Gamma = E\left(\Lambda\Lambda^T\right), \quad \Lambda = \left(\left\{x - \rho_x(x^T\theta_0)\right\}^T, \left[g'(x^T\theta_0)\left\{x - \rho_x(x^T\theta_0)\right\}\right]^T\right)^T.$$

5.3 Penalized local smoothing estimators

In a real-world problem, the true model is usually unknown and either overfitting or underfitting of the model could happen, especially when the number of parameters is relatively large but not sufficient observations are available. Therefore, in these cases, we would like to estimate the parameters and conduct a variable selection simultaneously. This motivates us to use the penalized local smoothing estimators to perform the data analyses. In this section, we propose the penalized local smoothing estimators with the implementation of the lasso penalty to carry out variable selection as well as parameter estimation. The penalized estimators $\left(\hat{\beta}_{\lambda_1}, \hat{\theta}_{\lambda_2}\right)$ are defined as

$$\begin{aligned}\left(\hat{\beta}_{\lambda_1}, \tilde{\theta}_{\lambda_2}\right) &= \underset{\beta^T\theta=0}{\arg\min}\, G_p(\beta, \theta) \\ &= \underset{\beta^T\theta=0}{\arg\min}\left\{\frac{1}{2}G(\beta, \theta) + n\lambda_1 \left\|\beta\right\|_1 + n\lambda_2 \left\|\theta\right\|_1\right\},\end{aligned}$$

$$\hat{\theta}_{\lambda_2} = \tilde{\theta}_{\lambda_2}/||\tilde{\theta}_{\lambda_2}||,$$

where $G(\beta, \theta)$ is defined in Equation (5.4), λ_1 and λ_2 are the tuning parameters of β and θ respectively, $\left\|\beta\right\|_1 = \sum_{j=1}^{p}|\beta_j|$ and $\left\|\theta\right\|_1 = \sum_{k=1}^{p}|\theta_k|$. Let S and T denote the sets of the subscripts of the nonzero elements of β_0 and θ_0 respectively. For example, $S = \{1, 3\}$ implies $\beta_{01} \neq 0$ and $\beta_{03} \neq 0$. For any $l \in \mathbb{R}^p$ and $A = \{i_1, i_2, \ldots, i_{|A|}\} \subset \{1, 2, \ldots, p\}$, let $l_A = (l_{i_1}, l_{i_2}, \ldots, l_{i_{|A|}})$ be the vector containing elements of l with subscripts in A and $A^c = \{1, 2, \ldots, p\} \setminus A$. Similarly, let X_A be the matrix containing columns of X corresponding to the elements in A.

Theorem 5.3.1. *Suppose that the regularity assumptions 1-7 are satisfied and that* $\lambda_i \to 0$, $n^{1/2}\lambda_i \to \infty$ *for* $i = 1, 2$. *Then we have:*

(a) $\hat{\beta}_{\lambda_1 S^c} = 0$ *and* $\hat{\theta}_{\lambda_2 T^c} = 0$ *for large enough* n*;*

(b)

$$n^{\frac{1}{2}} \begin{pmatrix} \hat{\beta}_{\lambda_1 S} - \beta_{0S} \\ \hat{\theta}_{\lambda_2 T} - \theta_{0T} \end{pmatrix} \xrightarrow{\mathscr{D}} N(0, \sigma^2 \Gamma_r^+),$$

where

$$\Gamma_r = E(\Lambda_r \Lambda_r^T), \quad \Lambda_r = \left(\left\{ x_S - \rho_{x_S}(x^T\theta_0) \right\}^T, \left[g'(x^T\theta_0) \left\{ x_T - \rho_{x_T}(x^T\theta_0) \right\} \right]^T \right)^T.$$

In practice, the tuning parameters, λ_1 and λ_2, can be chosen via cross-validation. Other methods, such as the Akaike information criterion and Bayesian information criterion, can also be applied to determine λ_1 and λ_2 (see [Liang et al., 2010] for more details).

5.4 Hypothesis testing

Consider the general linear hypothesis

$$H_0 : W\xi = 0 \quad \text{versus} \quad H_1 : W\xi \neq 0, \tag{5.12}$$

where $\xi = (\beta^T, \theta^T)^T$ and W is a $m \times 2p$ full rank matrix. Let Ω_0 and Ω_1 be the parameter spaces of H_0 and H_1, respectively. Define

$$G(H_0) = \inf_{\xi \in \Omega_0} G(\xi), \quad G(H_1) = \inf_{\xi \in \Omega_1} G(\xi),$$

and the test statistic

$$V = \frac{n \left\{ G(H_0) - G(H_1) \right\}}{G(H_1)}. \tag{5.13}$$

Then we have the following theorem for testing the hypotheses in (5.12).

Theorem 5.4.1. *Suppose that the regularity assumptions 1-7 are satisfied. We have:*

(a) under H_0 in (5.12), $V \to \chi^2_m$ in distribution;

(b) under H_1 in (5.12), the test is consistent;

(c) under the local alternative of $n^{1/2} W\xi \to d$ for some m dimensional $d \neq 0$, V converges in

distribution to a noncentral chi-squared distribution with m degrees of freedom and noncentrality parameter

$$\psi = \sigma^{-2} d^T (W \Gamma^+ W^T)^{-1} d,$$

where Γ is defined in Theorem 5.2.2.

5.5 Simulation study

In this section, we evaluate our proposed methods empirically, and compare them with the method introduced in [Xia et al., 1999] via simulation. We provide three examples here: the first example is from [Xia et al., 1999], where the number of the parameters is relatively small; in the second example, the number of the parameters is relatively large, where the penalized estimators are expected to have better performance; the third example is about hypothesis testings in extended partially linear single-index models.

Example 5.5.1. We firstly considered the example shown in [Xia et al., 1999], which can be written as

$$y_i = 0.3x_i + 0.4x_{i-1} + \exp\left\{-2\left(0.8x_i - 0.6x_{i-1}\right)^2\right\} + 0.1e_i,$$

where

$$x_i = 0.8x_{i-1} + \epsilon_i + 0.5\epsilon_{i-1}, \quad e_i, \epsilon_i \sim N(0,1),$$

and all e_i, ϵ_i are independent of each other. The model above can also be expressed as

$$y_i = \beta_1 x_i + \beta_2 x_{i-1} + \exp\left\{-2\left(x_i \cos\alpha - x_{i-1}\sin\alpha\right)^2\right\} + 0.1e_i, \tag{5.14}$$

where $\beta_1 = 0.3$, $\beta_2 = 0.4$ and $\alpha = \arcsin(0.6) = 0.6435$.

Five different methods were used in this example: the kernel smoothing estimators introduced in [Xia et al., 1999], the Lagrange kernel smoothing estimators and the local smoothing estimators described in Section 5.2, the penalized kernel smoothing estimators and the penalized local smoothing estimators proposed in Section 5.3. Although penalized estimation was not discussed in [Xia et al., 1999], for comparison purposes, we could simply add the lasso penalty to the objective

n		50		100		200	
Parameters	Methods	Bias	SMSE	Bias	SMSE	Bias	SMSE
	KSE	-0.0040	0.3045	0.0090	0.3094	0.0171	0.3056
	LKSE	0.0004	0.0068	-0.0001	0.0040	0.0000	0.0021
β_1	LSE	-0.0008	0.0054	-0.0005	0.0033	-0.0002	0.0020
	PKSE	-0.0010	0.0154	-0.0009	0.0071	-0.0007	0.0035
	PLSE	-0.0024	0.0087	-0.0014	0.0048	-0.0009	0.0029
	KSE	0.0030	0.2284	-0.0067	0.2322	-0.0130	0.2291
	LKSE	-0.0006	0.0072	0.0001	0.0042	-0.0001	0.0025
β_2	LSE	0.0003	0.0059	0.0003	0.0037	0.0000	0.0023
	PKSE	-0.0010	0.0122	0.0002	0.0063	-0.0006	0.0033
	PLSE	-0.0003	0.0069	0.0005	0.0044	-0.0005	0.0027
	KSE	0.0014	0.0121	-0.0005	0.0066	0.0002	0.0039
	LKSE	0.0013	0.0120	-0.0003	0.0065	0.0001	0.0039
α	LSE	-0.0016	0.0087	-0.0011	0.0054	-0.0003	0.0034
	PKSE	-0.0005	0.0337	-0.0017	0.0161	-0.0005	0.0078
	PLSE	-0.0036	0.0171	-0.0028	0.0099	-0.0009	0.0059

Table 5.1: Simulation results of model (5.14). KSE, kernel smoothing estimator; LKSE, Lagrange kernel smoothing estimator; LSE, local smoothing estimator; PKSE, penalized kernel smoothing estimator; PLSE, penalized local smoothing estimator; SMSE, square root of mean squared error.

function of their kernel smoothing estimators to obtain the penalized kernel smoothing estimators. We simulated 500 independent data sets with sample size $n = 50$, $n = 100$ and $n = 200$. The estimation procedure of the kernel smoothing estimators is the same as shown in [Xia et al., 1999]. For the local smoothing estimators, it is expected that h is determined by cross-validation for each simulated data set. However, due to a substantial computational burden, for each n, we firstly fixed $h \in [0.1, 0.3]$, ran 500 replications and calculated the mean squared error of the parameters. Then h_n was obtained by minimizing the mean squared error. After h was determined, the tuning parameters λ_1 and λ_2 were determined in a similar way. All optimizations were done in R via the *nloptr()* function from the *nloptr* package.

Table 5.1 shows the bias and the square root of mean squared error for the five methods obtained in our simulations. Although the kernel smoothing estimator and the Lagrange kernel smoothing estimator for α have similar performance, the Lagrange kernel smoothing estimator for β is sig-

nificantly better than the kernel smoothing estimator of β in terms of bias and mean squared error. Furthermore, while the biases of the local smoothing estimators and Lagrange kernel smoothing estimators are similar with both of them being nearly negligible relative to the square root of mean squared errors, the square root of mean squared error of the local smoothing estimators is noticeably smaller than the square root of mean squared error of the Lagrange kernel smoothing estimators, especially when the sample size n is relatively small. In addition, both penalized estimators have worse performance compared with the two non-penalized estimators, even for a relatively small sample size. The main reason is that the parameters are not sparse here. Moreover, for sample size $n = 200$, we compared the computation time of calculating the local smoothing estimators using and without using Theorem 5.2.1. About 42% of the computation time had been reduced by applying the results of Theorem 5.2.1, which implies that the calculation methods provided by Theorem 5.2.1 have successfully accelerated the estimating procedure.

Example 5.5.2. We now consider an extended partially linear single-index model with more parameters, which is model (5.3) with link function and parameters

$$g(u) = \left(1 + u^2\right)^{-1}, \quad \beta_0 = (2, -1, 0, 0, 0), \quad \theta_0 = (1, 2, 0, 0, 0)/\sqrt{5}, \tag{5.15}$$

and the covariates and random errors are independent and identically distributed as

$$x_{ij} \sim N(0, 1), \quad e_i \sim N(0, 0.1^2).$$

We simulated 500 independent data sets with sample sizes $n = 50$, $n = 100$ and $n = 200$ from this model. Since the sample size is relatively small and the model parameters are sparse, the penalized estimators are expected to have better performance.

Table 5.2 shows the square root of mean squared error, the average number of the true zero parameters that were correctly set to zero and the average number of the truly nonzero parameters that were incorrectly set to zero for the five methods obtained in our simulations. Again, for β, the mean squared error of the Lagrange kernel smoothing estimator is significantly smaller than

n		50			100			200		
Parameters	Methods	SMSE	C	I	SMSE	C	I	SMSE	C	I
	KSE	0.5492	0.91	0	0.5183	1.34	0	0.4314	1.94	0
	LKSE	0.1158	1.01	0	0.0630	1.67	0	0.0427	2.31	0
β	LSE	0.1171	1.11	0	0.0676	1.77	0	0.0412	2.44	0
	PKSE	0.0436	2.03	0	0.0278	2.53	0	0.0170	2.82	0
	PLSE	0.0280	2.51	0	0.0199	2.75	0	0.0132	2.89	0
	KSE	0.0906	0.66	0	0.0602	0.87	0	0.0402	1.19	0
	LKSE	0.0951	0.60	0	0.0578	0.92	0	0.0391	1.20	0
θ	LSE	0.0984	0.52	0	0.0583	0.85	0	0.0354	1.22	0
	PKSE	0.0196	2.75	0	0.0137	2.77	0	0.0059	2.97	0
	PLSE	0.0057	2.99	0	0.0048	2.98	0	0.0029	3.00	0

Table 5.2: Simulation results of model (5.15). KSE, kernel smoothing estimator; LKSE, Lagrange kernel smoothing estimator; LSE, local smoothing estimator; PKSE, penalized kernel smoothing estimator; PLSE, penalized local smoothing estimator; SMSE, square root of mean squared error; C, the average number of the true zero parameters that were correctly set to zero (less than 0.01); I, the average number of the truly nonzero parameters that were incorrectly set to zero.

the mean squared error of the kernel smoothing estimator. This implies that the implementation of the method of Lagrange multipliers has led to huge improvement in performance of the kernel smoothing estimators. Besides, the results in Table 5.2 also indicate that while the Lagrange kernel smoothing estimators and the local smoothing estimators have similar performance, the penalized estimators have much better performance compared with the estimators without penalty. Although the computation was heavy, we also tried to simulate a small number of replications for $n = 400$, and the results are similar to those for $n = 200$. Therefore, we can conclude that even if the sample size is relatively large, the penalized estimators are more preferable as long as the model contains sparsity. In addition, Table 5.2 also indicates that penalized local smoothing estimators perform significantly better than the penalized kernel smoothing estimators, especially when the sample size is relatively small ($n = 50$ and $n = 100$).

Example 5.5.3. To investigate the performance of the test statistic V described in Section 5.4, we consider model (5.3) with link function and parameters

$$g(u) = 3u^2, \quad \beta_0 = (2, -3, 0, 0), \quad \theta_0 = (3, 2, c, c)/\sqrt{13},$$

where c ranges from 0 to 0.6 with increment 0.05. The covariates and random errors are independent and identically distributed as

$$x_{ij} \sim U(0, 1) \quad e_i \sim N(0, 0.1^2).$$

For each value of c, we simulated 300 independent data sets with sample sizes $n = 50$, $n = 100$ and $n = 200$ from the model, and considered the following null and alternative hypotheses:

$$H_0 : \beta_3 = \beta_4 = 0 \quad \text{versus} \quad H_1 : \beta_3 = \beta_4 = c$$

with the nominal level equal to 0.05. The power function (or type I error when $c = 0$) versus c is plotted in Figure 5.1.

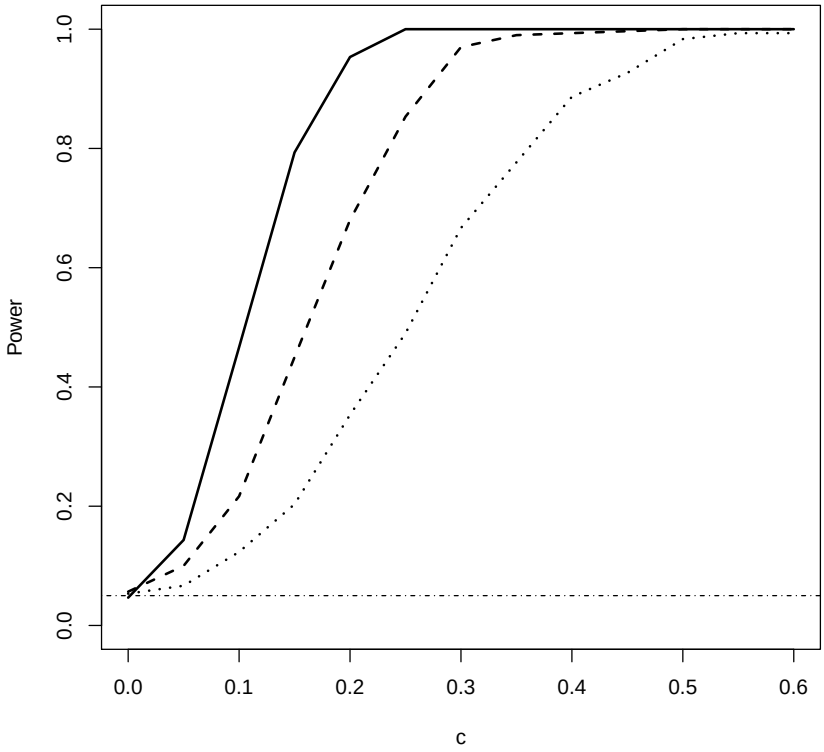


Figure 5.1: The power function (or type I error when $c = 0$) versus c for sample size $n = 50$ (dotted), $n = 100$ (dashed) and $n = 200$ (solid). The nominal level is equal to 0.05 (horizontal dot-dash).

Figure 5.1 shows that when $c = 0$, the type I error of the test is equal to 0.05, 0.06 and 0.05 for $n = 50$, $n = 100$ and $n = 200$, respectively, which is close to the nominal level apart from its standard errors. Also, Figure 5.1 implies that the power function increases quite fast as c increases. Overall, V leads to a powerful test whose size is also well controlled.

5.6 Real data application

We applied the proposed methods to analyze a publicly available data set of concrete slump test data, which was firstly introduced and analyzed in [Yeh, 2007] (see [Yeh, 2006] and [Yeh, 2007] for more related information). The high-performance concrete is highly complex, and thus it is very difficult to model its behavior using available information. In this data set, there are 7

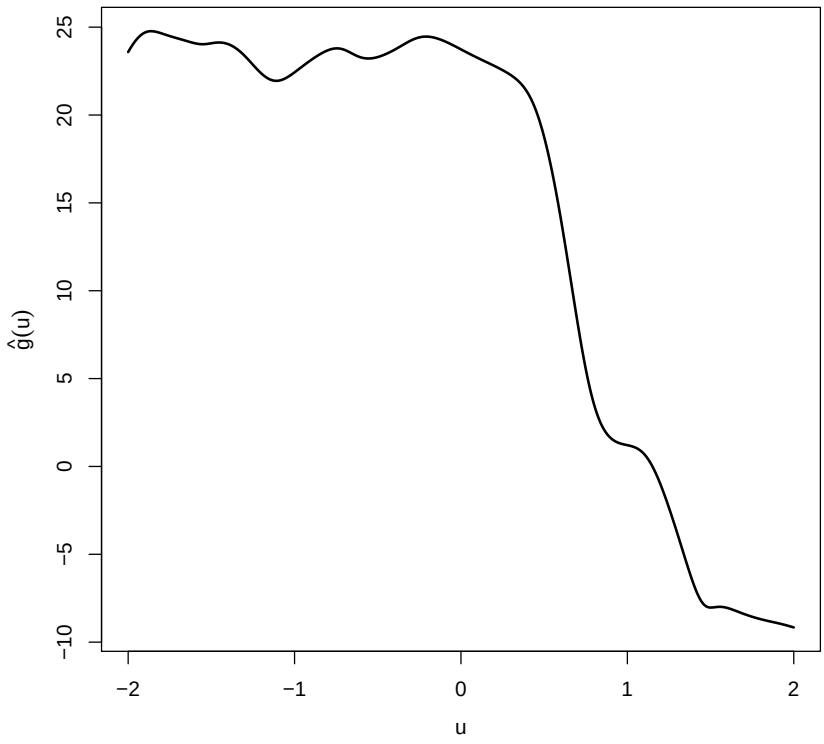


Figure 5.2: The value of $\hat{g}(u)$ versus u obtained from the concrete slump test data set.

input covariates: cement (kg/m^3), blast furnace slag (kg/m^3), fly ash (kg/m^3), water (kg/m^3), superplasticizer (kg/m^3), coarse aggregate (kg/m^3), and fineaggregate (kg/m^3), and 3 output variables: concrete slump (cm), concrete flow (cm) and 28-day compressive strength (mpa). We focused on modeling the concrete slump using all the 7 available input covariates. The data set contains 103 observations and a multiple linear regression model yields an R^2 value of 0.32. Some further exploratory analysis indicates strong nonlinear relationships between the concrete slump and the covariate variables, which leads to the use of nonlinear models for predictions and simulations of concrete slump ([Yeh, 2008] and [Yeh, 2009]).

The local smoothing estimation, Lagrange kernel smoothing estimation and their penalized versions were applied to the analysis of the data set after the covariates were standardized. For the penalized local smoothing estimation, we firstly performed a 10-fold cross-validation to the

data set to select the bandwidth and the tuning parameters. Then we computed the estimates of the parameters with the tuning parameters and obtained

$$\hat{\theta} = (0.152, 0.558, 0.031, -0.722, -0.166, -0.302, -0.158)^T,$$

$$\hat{\beta} = (10.817, 12.927, 12.570, 6.056, 0.032, 11.407, 8.975)^T.$$

The results indicate that the third element of $\hat{\theta}$ and the fifth element of $\hat{\beta}$ are effectively zero. Therefore, the nonparametric part of the model might not depend on the third covariate (fly ash), while the linear part of the model might not depend on the fifth covariate (superplasticizer).

Figure 5.2 shows the estimate of the link function $\hat{g}(u)$ with $-2 \leqslant u \leqslant 2$ obtained from the penalized local smoothing estimation. The function drops rapidly when $u > 0.5$. These estimates of parameters and the link function yields an R^2 value equal to 0.82, while the R^2 values obtained by using the local smoothing estimation, Lagrange kernel smoothing estimation and penalized kernel smoothing estimation are 0.57, 0.44 and 0.47, respectively. The R^2 value of the penalized kernel smoothing estimation is much smaller than the R^2 value of the penalized local smoothing estimation. This, together with the results of the simulations shown in Section 5.5, implies that the penalized local smoothing estimation has better performance and is more robust, especially for real world problems when no prior information of the model parameters is available. Overall, the local smoothing estimation method has the best performance among all the estimation methods, while the other three methods also lead to substantial improvements compared with the simple linear model apporach.

5.7 Discussion

In this chapter, we considered the extended partially linear single-index models (5.3), which are more flexible compared with the partially linear single-index models (5.1). However, extended partially linear single-index models often have more parameters, making it more difficult to estimate the parameters. We proposed the local smoothing estimators in Section 5.2 for parameter estimation, and introduced the chi-squared test statistic in Section 5.4 for testing general linear hy-

potheses. Furthermore, for data sets with too many covariates (which lead to sparse parameters), we proposed the penalized local smoothing estimators in Section 5.3 for conducting parameter estimation and variable selection simultaneously. The uniqueness and linear expressions of the solution to the optimization of the profile objective function are shown in Section 5.2, resulting in fast and accurate computations for the solution. In addition, the performance of the kernel smoothing estimators introduced in [Xia et al., 1999] can be improved by implementing the method of Lagrange multipliers to calculate the profile estimator. Besides, asymptotic properties of the proposed estimators and test statistic were also introduced and discussed in detail.

Simulation studies were presented in Section 5.5 to assess the performance of the proposed estimators and test statistic. We compared the five estimation methods for a model containing a small number of parameters introduced in [Xia et al., 1999], and for a model containing more parameters. The simulation results indicate that the Lagrange kernel smoothing estimators have much better performance compared with the kernel smoothing estimators, especially for β. Besides, the results of the first example implies the local smoothing estimators perform better than the Lagrange kernel smoothing estimators, and the results of the second example implies the penalized local smoothing estimators perform better than the penalized kernel smoothing estimators. The results also indicate that the penalized estimators would generally outperform the non-penalized estimators when the model contains sparsity. For the test statistic V defined in (5.13), the simulation results show that it is powerful with good size control.

An interesting real-world data set of concrete slump test data was analyzed in Section 5.6. We fitted the extended partially linear single-index model to the data and used introduced methods to estimate the parameters and the link function. The estimated link function $\hat{g}(u)$ shown in Figure 5.2 has a special pattern, which might result from some characteristic of the data. The fitted R^2 was more than doubled to 0.82 from 0.32 by fitting the extended partially linear single-index model with penalized local smoothing estimators instead of a multiple linear regression model.

As a future research problem, it would be interesting to study the extended partially linear single-index models with more complicated correlation structure. For instance, the covariates

could be time series with auto correlation, or the measurements are taken from different subjects over time as in longitudinal data.

6. SUMMARY AND CONCLUSIONS

In Chapter 2, a space-time model of soil moisture was developed with a jitter process incorporated to deflate the correlation structure at small spatial scales. The theoretical properties of the model, including the correlation structure, power spectrum and gain function, were carefully studied. Then the new model was fitted to a set of soil moisture data with all the parameters estimated. The results indicate that with the jitter process that accounts for local variation of surface topography, soil characteristics and vegetation conditions, the model has successfully reproduced the fast decay spatial correlation observed in the soil moisture data.

In Chapter 3, the soil moisture field was simulated from the model introduced in Chapter 2 to study the distributional properties of the soil moisture islands. The soil moisture islands, which are defined as the clusters of areas where the soil moisture is above certain threshold, were found to follow a power-law distribution in size. Also, the sizes of tree clusters in savannas are power-law distributed. We infer that the emergence of large-scale spatial patterns of savanna vegetation characterized by clusters of trees in a matrix of grasses with power-law probability distribution of cluster sizes and fractal perimeters of such clusters, result from spatial patterns of soil moisture in such ecosystems.

Motivated by the studies of the practical problems of interest presented in Chapters 2 and 3, in Chapter 4, we considered the statistical distributional properties of image clustering of spatial random fields. A well-defined distribution function was defined and its analytical properties were fully studied. We then showed that the corresponding empirical distribution function, which comes from spatial image data and has already been used in many studies, converges to the defined distribution function as the sample size tends to infinity. Thus the empirical distribution function can be used as an estimator of the distribution function, as shown in numerical experiments and an application to tree clusters data.

In Chapter 5, variable selection and statistical inferences of the extended partially linear single-index models were studied. We proposed the local smoothing estimators to estimate the model

parameters as well as the unknown link function. Due to the relatively large number of parameters contained in these models, we also proposed the penalized local smoothing estimators for the analyses of data sets with sparsity. For linear hypotheses about the parameters, A chi-squared type of test statistic was introduced. The results of simulation studies imply that the local smoothing estimators and the penalized local smoothing estimators are more efficient than the kernel smoothing estimators and the penalized kernel smoothing estimators. Besides, the proposed test statistic were shown to be powerful and control the size of the test well. Finally, a real data analysis was provided to illustrate the proposed methodology.

REFERENCES

[Adler and Taylor, 2007] Adler, R. J. and Taylor, J. E. (2007). Random Fields and Geometry. Springer-Verlag, New York.

[Albertson and Montaldo, 2003] Albertson, J. D. and Montaldo, N. (2003). Temporal dynamics of soil moisture variability: 1. Theoretical basis. Water Resour. Res. *39*, 1274.

[Asnera and Warner, 2003] Asnera, G. P. and Warner, A. S. (2003). Canopy shadow in IKONOS satellite observations of tropical forests and savannas. Remote. Sens. Environ. *87*, 521–533.

[Azzalini and Torelli, 2007] Azzalini, A. and Torelli, N. (2007). Clustering via nonparametric density estimation. Stat. Comput. *17*, 71–80.

[Bowman et al., 2009] Bowman, D. M. J. S. et al. (2009). Fire in the earth system. Science *324*, 481–484.

[Carroll et al., 1997] Carroll, R. J., Fan, J., Gijbels, I. and Wand, M. P. (1997). Generalized partially linear single-index models. J. Am. Stat. Assoc. *92*, 477–489.

[Cattelan and Varin, 2018] Cattelan, M. and Varin, C. (2018). Marginal logistic regression for spatially clusteredbinary data. J. R. Stat. Soc. Ser. C. Appl. Stat. *67*, 939–959.

[Chen et al., 2015] Chen, J., Li, D., Liang, H. and Wang, S. (2015). Semiparametric GEE analysis in partially linear single-index models for longitudinal data. Ann. Statist. *43*, 1682–1715.

[Cox and Isham, 1988] Cox, D. R. and Isham, V. (1988). A simple spatial-temporal model of rainfall. Proc. R. Soc. Lond. A *415*, 317–328.

[Dong et al., 2016] Dong, C., Gao, J. and Tjøstheim, D. (2016). Estimation for single-index and partially linear single-index integrated models. Ann. Statist. *44*, 425–453.

[Eagleson, 1978] Eagleson, P. S. (1978). Climate, soil, and vegetation: 1. Introduction to water balance dynamics. Water Resour. Res. *14*, 705–712.

[Entekhabi and Rodriguez-Iturbe, 1994] Entekhabi, D. and Rodriguez-Iturbe, I. (1994). Analytical framework for the characterization of the space-time variability of soil moisture. Adv. Water Resour. *17*, 35–45.

[Fan and Gijbels, 1996] Fan, J. and Gijbels, I. (1996). Local Polynomial Modelling and Its Applications. Chapman & Hall, London.

[Fan and Huang, 2005] Fan, J. and Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli *11*, 1031–1057.

[February et al., 2013] February, E. C., Higgins, S. I., Bond, W. J. and Swemmer, L. (2013). Influence of competition and rainfall manipulation on the growth responses of savanna trees and grasses. Ecology *94*, 1155–1164.

[Flores and Leon, 2010] Flores, E. and Leon, J. R. (2010). Level sets of random fields and applications: Specular points and wave crests. Int. J. Stoch. Anal. *2010*, 1–22.

[Gemignani, 1990] Gemignani, M. C. (1990). Elementary Topology. 3 edition, Dover Publications.

[Horowitz and Härdle, 1996] Horowitz, J. L. and Härdle, W. (1996). Direct semiparametric estimation of single-index models with discrete covariates. J. Am. Stat. Assoc. *91*, 1632–1640.

[Ichimura, 1993] Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. J. Econom. *58*, 71–120.

[Iglauer et al., 2010] Iglauer, S., Favretto, S., Spinelli, G., Schena, G. and Blunt, M. J. (2010). X-ray tomography measurements of power-law cluster size distributions for the nonwetting phase in sandstones. Phys. Rev. E Stat. Nonlin. Soft. Matter. Phys. *82*, 056315.

[Isham et al., 2005] Isham, V., Cox, D. R., Rodriguez-Iturbe, I., Porporato, A. and Manfreda, S. (2005). Representation of space-time variability of soil moisture. Proc. R. Soc. A. *461*, 4035–4055.

[Kefi et al., 2014] Kefi, S. et al. (2014). Early warning signals of ecological transitions: Methods for spatial patterns. PLoS ONE *9*, e92097.

[Kittredge, 1948] Kittredge, J. (1948). Forest Influences: The Effects of Woody Vegetation on Climate, Water, and Soil, with Applications to the Conservation of Water and the Control of Floods and Erosion. McGraw-Hill Book Co., New York.

[Korcak, 1940] Korcak, J. (1940). Deux types fondamentaux de distribution statistique [Two basic types of statistical distribution]. Bull. Inst. Int. Stat. *30*, 295–299.

[Korvin, 1992] Korvin, G. (1992). Fractal Models in the Earth Sciences. Elsevier, Amsterdam.

[Levin et al., 1989] Levin, S. A., Morin, A. and Powell, T. H. (1989). Patterns and processes in the distribution and dynamics of Antarctic krill. Scientific Committee for the Conservation of Antarctic Marine Living Resources, Selected Scientific Papers (Commission for the Conservation of Antarctic Marine Living Resources, Hobart, Australia), SC-CAMLR-SSP/5 .

[Li, 2006] Li, B. (2006). A new approach to cluster analysis: The clustering-function-based method. J. R. Stat. Soc. Ser. B. Stat. Methodol. *68*, 457–476.

[Li and Hsing, 2010] Li, Y. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. Ann. Statist. *38*, 3321–3351.

[Liang et al., 2010] Liang, H., Liu, X., Li, R. and Tsai, C. (2010). Estimation and testing for partially linear single-index models. Ann. Statist. *38*, 3811–3836.

[Liang and Wang, 2005] Liang, H. and Wang, N. (2005). Partially linear single-index measurement error models. Stat. Sin. *15*, 99–116.

[Lin and Kulasekera, 2007] Lin, W. and Kulasekera, K. B. (2007). Identifiability of single-index models and additive-index models. Biometrika *94*, 496–501.

[Mandelbrot, 1982] Mandelbrot, B. B. (1982). The Fractal Geometry of Nature. Freeman and Co., New York.

[Martin and Goldenfeld, 2006] Martin, H. G. and Goldenfeld, G. (2006). On the origin and robustness of power-law species–area relationships in ecology. PNAS *103*, 10310—10315.

[McNicholas, 2016] McNicholas, P. D. (2016). Model-based clustering. J. Classification *33*, 331–373.

[Meinhardt, 1993] Meinhardt, H. (1993). Models of Biological Pattern Formation. Springer, New York.

[Menardi and Azzalini, 2014] Menardi, G. and Azzalini, A. (2014). An advancement in clustering via nonparametric density estimation. Stat. Comput. *24*, 753–767.

[Meron, 2011] Meron, E. (2011). Modeling dryland landscapes. Math. Model. Nat. Phenom. *6*, 163–187.

[Mohanty and Skaggs, 2001] Mohanty, B. P. and Skaggs, T. H. (2001). Spatio-temporal evolution and time-stable characteristics of soil moisture within remote sensing footprints with varying soil, slope, and vegetation. Adv. Water Resour. *24*, 1051–1067.

[Murray, 1993] Murray, J. D. (1993). Mathematical Biology. Springer, New York.

[Pan et al., 2003] Pan, F., Peters-Lidard, C. D. and Sale, M. J. (2003). An analytical method for predicting surface soil moisture from rainfall observations. Water Resour. Res. *39*, 1314.

[Pascual et al., 2002] Pascual, M., Roy, M., Guichard, F. and Flierl, G. (2002). Cluster size distributions: Signatures of self-organization in spatial ecologies. Phil. Trans. R. Soc. Lond. B *357*, 657–666.

[Ramankutty and Foley, 1999] Ramankutty, N. and Foley, J. A. (1999). Estimating historical changes in global land cover: Croplands from 1700 to 1992. Global Biogeochem. Cycles *13*, 997–1027.

[Rietkerk and van de Koppel, 2008] Rietkerk, M. and van de Koppel, J. (2008). Regular pattern formation in real ecosystems. Trends Ecol. Evol. *23*, 169–175.

[Rodriguez-Iturbe et al., 1987] Rodriguez-Iturbe, I., Cox, D. R. and Isham, V. (1987). Some models for rainfall based on stochastic point processes. Proc. R. Soc. A *410*, 269–288.

[Rodriguez-Iturbe et al., 2006] Rodriguez-Iturbe, I., Isham, V., Cox, D. R., Manfreda, S. and Porporato, A. (2006). Space-time modeling of soil moisture: Stochastic rainfall forcing with heterogeneous vegetation. Water Resour. Res. *42*, W06D05.

[Rodriguez-Iturbe and Porporato, 2005] Rodriguez-Iturbe, I. and Porporato, A. (2005). Ecohydrology of Water-Controlled Ecosystems: Soil Moisture and Plant Dynamics. Cambridge Univ. Press, Cambridge, U.K.

[Scanlon et al., 2007] Scanlon, T. M., Caylor, K. K., Levin, S. A. and Rodriguez-Iturbe, I. (2007). Positive feedbacks promote power-law clustering of Kalahari vegetation. Nature *449*, 209–212.

[Scholes, 1997] Scholes, R. J. (1997). Vegetation of Southern Africa. Cambridge Univ. Press, New York.

[Sole, 2007] Sole, R. (2007). Scaling laws in the drier. Nature *449*, 151–153.

[Song et al., 2006] Song, W., Wang, J., Satoh, K. and Fan, W. (2006). Three types of power-law distribution of forest fires in Japan. Ecol. Model. *196*, 527–532.

[Staver, 2018] Staver, A. C. (2018). Prediction and scale in savanna ecosystems. New Phytol. *219*, 52–57.

[Staver et al., 2019] Staver, A. C., Asner, G. P., Rodriguez-Iturbe, I., Levin, S. A. and Smit, I. (2019). Spatial patterning among savanna trees in high resolution, spatially extensive data. PNAS *116*, 10681–10685.

[Steinwart, 2015] Steinwart, I. (2015). Fully adaptive density-based clustering. Ann. Statist. *43*, 2132–2167.

[Ulrich, 1984] Ulrich, Z. (1984). A general rice formula, palm measures, and horizontal-window conditioning for random fields. Stochastic Process. Appl. *17*, 265–283.

[Wang et al., 2005] Wang, N., Carroll, R. J. and Lin, X. (2005). Efficient semiparametric marginal estimation for longitudinal/clustered data. J. Am. Stat. Assoc. *100*, 147–157.

[Waymire et al., 1984] Waymire, E., Gupta, V. K. and Rodriguez-Iturbe, I. (1984). A spectral theory of rainfall intensity at the meso-β scale. Water Resour. Res. *20*, 1453–1465.

[Whittle, 1962] Whittle, P. (1962). Topographic correlation, power-law covariance functions, and diffusion. Biometrika *49*, 305–314.

[Wootton, 2001] Wootton, J. T. (2001). Local interactions predict large-scale pattern in empirically derived cellular automata. Nature *413*, 841–844.

[Worsley, 1995] Worsley, K. J. (1995). Estimating the number of peaks in a random field using the Hadwiger characteristic of excursion sets, with applications to medical images. Ann. Statist. *23*, 640–669.

[Worsley, 1997] Worsley, K. J. (1997). The geometry of random images. Chance *9*, 27–40.

[Xia and Härdle, 2006] Xia, Y. and Härdle, W. (2006). Semi-parametric estimation of partially linear single-index models. J. Multivar. Anal. *97*, 1162–1184.

[Xia et al., 1999] Xia, Y., Tong, H. and Li, W. K. (1999). On extended partially linear single-index models. Biometrika *86*, 831–842.

[Xia et al., 2002] Xia, Y., Tong, H., Li, W. K. and Zhu, L. (2002). An adaptive estimation of dimension reduction space. J. R. Statist. Soc. B *64*, 363–410.

[Yeh, 2006] Yeh, I. (2006). Exploring concrete slump model using artificial neural networks. J. Comput. Civil. Eng. *20*, 217–221.

[Yeh, 2007] Yeh, I. (2007). Modeling slump flow of concrete using second-order regressions and artificial neural networks. Cement Concr. Compos. *29*, 474–480.

[Yeh, 2008] Yeh, I. (2008). Prediction of workability of concrete using design of experiments for mixtures. Comput. Concr. *5*, 1–20.

[Yeh, 2009] Yeh, I. (2009). Simulation of concrete slump using neural networks. Construct. Mater. *162*, 11–18.

[Yu and Ruppert, 2002] Yu, Y. and Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. J. Am. Stat. Assoc. *97*, 1042–1054.

APPENDIX A

SUPPLEMENTARY INFORMATION FOR CHAPTER 2

A.1 Calculations for equations of soil moisture driven solely by rainfall

The solution to (2.14) is given by

$$S(\boldsymbol{u},t) = b\int_{-\infty}^{\infty} d\boldsymbol{z}\int_{0}^{\infty} d\tau G(\boldsymbol{u},\boldsymbol{z},\tau)Y(\boldsymbol{z},t-\tau), \tag{A.1}$$

where

$$G(\boldsymbol{u},\boldsymbol{z},\tau) = (4\pi k\tau)^{-n/2}\exp\left\{-a\tau - \frac{|\boldsymbol{u}-\boldsymbol{z}|^2}{4k\tau}\right\}. \tag{A.2}$$

From (A.1) and (2.4), it follows that the expected value of the relative soil moisture is

$$\mu_S = E(S(\boldsymbol{u},t)) = b\mu_Y\int_{-\infty}^{\infty} d\boldsymbol{z}\int_{0}^{\infty} d\tau G(\boldsymbol{u},\boldsymbol{z},\tau) = \frac{b\mu_Y}{a} = \frac{2\pi\lambda\mu_X b}{\eta\rho^2 a}, \tag{A.3}$$

which is the same as that for the case when there is no diffusion ($k = 0$). From (2.10) and (2.14), it follows that the spectral density of S is (see [Whittle, 1962])

$$\Phi_S(\nu,\omega) = \frac{(2\pi b)^2\lambda E(X^2)}{\rho}\frac{2\left(\frac{\rho}{2}\right)^2 + \frac{\nu^2}{2}}{\left[\left(\frac{\rho}{2}\right)^2+\nu^2\right]^{5/2}(\eta^2+\omega^2)\left[(a+k\nu^2)^2+\omega^2\right]}. \tag{A.4}$$

Consequently, the space spectrum and time spectrum of S are given by

$$\begin{aligned}\Phi_S(\nu) &= \frac{1}{2\pi}\int_{\mathbb{R}}\Phi_S(\nu,\omega)d\omega \\ &= \frac{(2\pi b)^2\lambda E(X^2)}{2\eta\rho}\frac{2\left(\frac{\rho}{2}\right)^2 + \frac{\nu^2}{2}}{\left[\left(\frac{\rho}{2}\right)^2+\nu^2\right]^{5/2}(a+k\nu^2)(\eta+a+k\nu^2)},\end{aligned} \tag{A.5}$$

$$\begin{aligned}\Phi_S(\omega) &= \frac{1}{(2\pi)^2}\int_{\mathbb{R}^2}\Phi_S(\nu,\omega)d\boldsymbol{\nu} \\ &= \int_0^\infty \frac{2\pi b^2\lambda E(X^2)\left[2\left(\frac{\rho}{2}\right)^2+\frac{\nu^2}{2}\right]\nu}{\rho\left[\left(\frac{\rho}{2}\right)^2+\nu^2\right]^{5/2}(\eta^2+\omega^2)\left[(a+k\nu^2)^2+\omega^2\right]}d\nu.\end{aligned} \tag{A.6}$$

From (A.4), after taking the inverse Fourier transform (see Appendix B for definition), the space-time covariance function of the soil moisture is

$$\begin{aligned}\Gamma_S(l,h) &= \frac{b^2\lambda E(X^2)}{2\rho}\cdot \\ &\int_{\mathbb{R}^2}\frac{2\left(\frac{\rho}{2}\right)^2+\frac{\nu^2}{2}}{\left[\left(\frac{\rho}{2}\right)^2+\nu^2\right]^{5/2}\left[\eta^2-(a+k\nu^2)^2\right]}\left\{\frac{e^{-(a+k\nu^2)h}}{a+k\nu^2}-\frac{e^{-\eta h}}{\eta}\right\}e^{i\boldsymbol{l}^T\boldsymbol{\nu}}d\boldsymbol{\nu} \\ &= \frac{2\pi b^2\lambda E(X^2)}{2\rho}\cdot \\ &\int_0^\infty\frac{\nu\left(2\left(\frac{\rho}{2}\right)^2+\frac{\nu^2}{2}\right)}{\left[\left(\frac{\rho}{2}\right)^2+\nu^2\right]^{5/2}\left[\eta^2-(a+k\nu^2)^2\right]}\left\{\frac{e^{-(a+k\nu^2)h}}{a+k\nu^2}-\frac{e^{-\eta h}}{\eta}\right\}J_0(l\nu)d\nu,\end{aligned} \tag{A.7}$$

where $J_0(\cdot)$ is the Bessel function of the first kind and $l=\|\boldsymbol{l}\|$.

A.2 Space-time spectral density of soil moisture incorporating jitter

We define the Fourier Transform of a d dimensional function f as

$$\mathscr{F}(f)(\boldsymbol{\xi}) = \int_{\mathbb{R}^d} e^{-i\boldsymbol{\xi}^T\boldsymbol{x}}f(\boldsymbol{x})d\boldsymbol{x}, \tag{A.8}$$

then the corresponding inverse transform is

$$\mathscr{F}(\Phi)(\boldsymbol{x}) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} e^{i\boldsymbol{x}^T\boldsymbol{\xi}}\Phi(\boldsymbol{\xi})d\boldsymbol{\xi}. \tag{A.9}$$

According to the property of Fourier Transform,

$$\mathscr{F}(f_1f_2) = (2\pi)^{-d}\mathscr{F}(f_1)*\mathscr{F}(f_2), \tag{A.10}$$

or, equivalently, with $\Phi_1 = \mathscr{F}(f_1)$, $\Phi_2 = \mathscr{F}(f_2)$, we then have

$$(2\pi)^{-d}\Phi_1 * \Phi_2 = \mathscr{F}(\mathscr{F}^{-1}(\Phi_1)\mathscr{F}^{-1}(\Phi_2)), \tag{A.11}$$

where "*" denotes convolution.

Now from (2.18), by taking Fourier Transform, we are able to obtain the space-time spectrum of the soil moisture with jitter $\tilde{S}$ as

$$\Phi_{\tilde{S}}(\nu,\omega) = \Phi_S(\nu,\omega) + \mu_S^2\Phi_Z(\nu,\omega) + \Phi_{S*Z}(\nu,\omega), \tag{A.12}$$

To calculate $\Phi_{S*Z}(\nu,\omega)$ above, note that $\Phi_{S*Z}(\nu,\omega) = \mathscr{F}(\Gamma_S\Gamma_Z)$ and we have

$$\begin{aligned}
\Phi_{S*Z}(\nu,\omega) &\overset{(A.10)}{=} (2\pi)^{-3}\mathscr{F}(\Gamma_S) * \mathscr{F}(\Gamma_Z) \\
&= (2\pi)^{-3}\Phi_S(\nu,\omega) * \Phi_Z(\nu,\omega) \\
&= \frac{2b^2\lambda E(X^2)\alpha\beta\sigma_Z^2}{\rho}\int_{\mathbb{R}^2}\frac{2\left(\frac{\rho}{2}\right)^2 + \frac{x^2}{2}}{\left[\left(\frac{\rho}{2}\right)^2 + x^2\right]^{5/2}(\alpha^2 + \|\boldsymbol{\nu} - \boldsymbol{x}\|^2)^{3/2}}d\boldsymbol{x}\cdot \\
&\int_{\mathbb{R}}\left(\frac{1}{\eta^2 + y^2}\frac{1}{(a + kx^2)^2 + y^2}\right)\left(\frac{1}{\beta^2 + (\omega - y)^2}\right)dy.
\end{aligned} \tag{A.13}$$

Note also that the second integral above is a convolution of two functions, and using (A.11), we obtain

$$\begin{aligned}
\Phi_{S*Z}(\nu,\omega) = \frac{2\pi b^2\lambda E(X^2)\alpha\sigma_Z^2}{\rho}\int_{\mathbb{R}^2}&\frac{2\left(\frac{\rho}{2}\right)^2 + \frac{x^2}{2}}{\left[\left(\frac{\rho}{2}\right)^2 + x^2\right]^{5/2}(\alpha^2 + \|\boldsymbol{\nu} - \boldsymbol{x}\|^2)^{3/2}}\cdot \\
&\frac{1}{\eta^2 - (a + kx^2)^2}\left\{\frac{1}{a + kx^2}\frac{a + kx^2 + \beta}{\omega^2 + (a + kx^2 + \beta)^2} - \frac{1}{\eta}\frac{\eta + \beta}{(\eta + \beta)^2 + \omega^2}\right\}d\boldsymbol{x}.
\end{aligned} \tag{A.14}$$

Thus finally we have

$$\begin{aligned}\Phi_{\tilde{S}}(\nu,\omega) &= \Phi_S(\nu,\omega) + \mu_S^2\Phi_Z(\nu,\omega) + \Phi_{S*Z}(\nu,\omega)\\ &= \frac{(2\pi b)^2\lambda E(X^2)}{\rho}\frac{2\left(\frac{\rho}{2}\right)^2+\frac{\nu^2}{2}}{\left[\left(\frac{\rho}{2}\right)^2+\nu^2\right]^{5/2}(\eta^2+\omega^2)\left[(a+k\nu^2)^2+\omega^2\right]} + \frac{4\pi\alpha\beta\sigma_Z^2\mu_S^2}{(\alpha^2+\nu^2)^{3/2}(\beta^2+\omega^2)}\\ &+ \frac{2\pi b^2\lambda E(X^2)\alpha\sigma_Z^2}{\rho}\int_{\mathbb{R}^2}\frac{2\left(\frac{\rho}{2}\right)^2+\frac{x^2}{2}}{\left[\left(\frac{\rho}{2}\right)^2+x^2\right]^{5/2}(\alpha^2+\|\boldsymbol{\nu}-\boldsymbol{x}\|^2)^{3/2}}\frac{1}{\eta^2-(a+kx^2)^2}\cdot\\ &\left\{\frac{1}{a+kx^2}\frac{a+kx^2+\beta}{\omega^2+(a+kx^2+\beta)^2} - \frac{1}{\eta}\frac{\eta+\beta}{(\eta+\beta)^2+\omega^2}\right\}d\boldsymbol{x}.\end{aligned} \tag{A.15}$$

By using polar coordinates, the last term $\Phi_{S*Z}(\nu,\omega)$ can also be expressed as

$$\begin{aligned}\Phi_{S*Z}(\nu,\omega) = \frac{2\pi b^2\lambda E(X^2)\alpha\sigma_Z^2}{\rho}\int_0^\infty &\left\{\frac{1}{a+kx^2}\frac{a+kx^2+\beta}{\omega^2+(a+kx^2+\beta)^2} - \frac{1}{\eta}\frac{\eta+\beta}{(\eta+\beta)^2+\omega^2}\right\}\cdot\\ &\frac{2\left(\frac{\rho}{2}\right)^2+\frac{x^2}{2}}{\left[\left(\frac{\rho}{2}\right)^2+x^2\right]^{5/2}\left[\eta^2-(a+kx^2)^2\right]}\left\{\int_0^{2\pi}\frac{x}{[\alpha^2+x^2+v^2-2vx\cos\theta]^{3/2}}d\theta\right\}dx,\end{aligned} \tag{A.16}$$

which implies $\Phi_{\tilde{S}}(\nu,\omega)$ only depends on $\nu=\|\boldsymbol{\nu}\|$ and ω. Again, the space spectrum of $\tilde{S}$ can be obtained as

$$\begin{aligned}\Phi_{\tilde{S}}(\nu) &= \frac{1}{2\pi}\int_{\mathbb{R}}\Phi_{\tilde{S}}(\nu,\omega)d\omega\\ &= \frac{(2\pi b)^2\lambda E(X^2)}{2\eta\rho}\frac{2\left(\frac{\rho}{2}\right)^2+\frac{\nu^2}{2}}{\left[\left(\frac{\rho}{2}\right)^2+\nu^2\right]^{5/2}(a+k\nu^2)(\eta+a+k\nu^2)} + \frac{2\pi\alpha\sigma_Z^2\mu_S^2}{(\alpha^2+\nu^2)^{3/2}} + \frac{2\pi b^2\lambda E(X^2)\alpha\sigma_Z^2}{2\eta\rho}\cdot\\ &\int_0^\infty\frac{2\left(\frac{\rho}{2}\right)^2+\frac{x^2}{2}}{\left[\left(\frac{\rho}{2}\right)^2+x^2\right]^{5/2}(a+kx^2)(\eta+a+kx^2)}\left\{\int_0^{2\pi}\frac{x}{[\alpha^2+x^2+v^2-2vx\cos\theta]^{3/2}}d\theta\right\}dx,\end{aligned} \tag{A.17}$$

and the time spectrum of $\tilde{S}$ is given by

$$\begin{aligned}
&\Phi_{\tilde{S}}(\omega) = \frac{1}{(2\pi)^2}\int_{\mathbb{R}^2}\Phi_{\tilde{S}}(\nu,\omega)d\boldsymbol{\nu} \\
&= \frac{2\pi b^2\lambda E(X^2)}{\rho}\int_0^\infty \frac{\left[2\left(\frac{\rho}{2}\right)^2+\frac{\nu^2}{2}\right]\nu}{\left[\left(\frac{\rho}{2}\right)^2+\nu^2\right]^{5/2}(\eta^2+\omega^2)\left[(a+k\nu^2)^2+\omega^2\right]}d\nu \\
&+ \frac{2\beta\sigma_Z^2\mu_S^2}{\beta^2+\omega^2} + \frac{2\pi b^2\lambda E(X^2)\sigma_Z^2}{\rho}\cdot \\
&\int_0^\infty \frac{2\left(\frac{\rho}{2}\right)^2+\frac{x^2}{2}}{\left[\left(\frac{\rho}{2}\right)^2+x^2\right]^{5/2}}\frac{x}{\eta^2-(a+kx^2)^2}\left\{\frac{1}{a+kx^2}\frac{a+kx^2+\beta}{\omega^2+(a+kx^2+\beta)^2}-\frac{1}{\eta}\frac{\eta+\beta}{(\eta+\beta)^2+\omega^2}\right\}dx.
\end{aligned} \tag{A.18}$$

A.3 The asymptotic behavior of the hydrologic gain function

We call $f \sim g$ if $f/g \to cons. \neq 0$. Then when ω is fixed, from (2.10) we know that $\Phi_Y(\nu,\omega) \sim \nu^{-3}$. According to (A.15),

$$\Phi_S(\nu,\omega) \sim \frac{1}{\nu^7}, \quad \Phi_Z(\nu,\omega) \sim \frac{1}{\nu^3}. \tag{A.19}$$

Also, by (A.16),

$$\int_0^{2\pi}\frac{x}{[\alpha^2+x^2+v^2-2vx\cos\theta]^{3/2}}d\theta \sim \frac{1}{\nu^3}, \tag{A.20}$$

and thus $\Phi_{S*Z}(\nu,\omega) \sim \nu^{-3}$. Therefore,

$$\Phi_{\tilde{S}}(\nu,\omega) = \Phi_S(\nu,\omega) + \mu_S^2\Phi_Z(\nu,\omega) + \Phi_{S*Z}(\nu,\omega) \sim \frac{1}{\nu^3} \sim \Phi_Y(\nu,\omega). \tag{A.21}$$

Similar result holds when ν is fixed.

APPENDIX B

SUPPLEMENTARY INFORMATION FOR CHAPTER 3

B.1 Simulation of the soil moisture field

The probabilistic structure of the soil moisture field, $\tilde{S}(\boldsymbol{u}, t)$, is studied via simulation with the parameter values given in Table 3.1 and 500 runs for each set of parameters. For each run the rainfall process $Y(\boldsymbol{u}, t)$ is generated in a 500 km $\times$ 500 km square field. We then divide the 1 km $\times$ 1 km square at the center of the rainfall field into 100 $\times$ 100 square pixels ($\boldsymbol{u}_i, i = 1, 2, \ldots, 10000$) of 10 m $\times$ 10 m each. At the center of each pixel, we calculate the soil moisture solely driven by rainfall, $S(\boldsymbol{u}, t)$, for 30 consecutive days, ($S(\boldsymbol{u}_i, j), j = 1, 2, \ldots, 30$) through the balance equation with $k = 0$. The jitter process is also simulated at the center of each pixel for 30 days and the soil moisture, $\tilde{S}(\boldsymbol{u}, t)$ is obtained by multiplying $S(\boldsymbol{u}, t)$ by $Z(\boldsymbol{u}, t)$. Focusing on the impact of soil moisture on vegetation the average of $\tilde{S}(\boldsymbol{u}, t)$ over a number of days, $\tilde{S}_T(\boldsymbol{u}, t)$, is specially relevant. This average over every pixel was studied at different levels of temporal aggregation with very similar results. Chapter 3 presents the results of $\tilde{S}(\boldsymbol{u}, t)$, for $T = 30$ days. With $k = 0$, we are able to calculate the soil moisture process $S(\boldsymbol{u}, t)$ by ([Isham et al., 2005])

$$S(\boldsymbol{u}, t) = b \int_0^\infty e^{-av} Y(\boldsymbol{u}, t - v) dv = b \int_{-\infty}^t e^{-a(t-v)} Y(\boldsymbol{u}, v) dv. \tag{B.1}$$

By replacing ∞ by some large number T, $S(\boldsymbol{u}, t)$ can be approximated by

$$S(\boldsymbol{u}, t) \simeq b \int_{-T}^t e^{-a(t-v)} Y(\boldsymbol{u}, v) dv. \tag{B.2}$$

Here we use $T = 1$ year, and the rainfall occurrences are simulated in the time period $[-T, 0]$. Then $S(\boldsymbol{u}, t)$ is simulated at $\boldsymbol{u} = \boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{10000}$ and $t = 1, 2, \ldots, 30$ (30 consecutive days). The jitter process, $Z(\boldsymbol{u}, t)$, is simulated as a Chi-squared field on the 1 km $\times$ 1 km area. For $\boldsymbol{u} = \boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{10000}$ and $t = 1, 2, \ldots, 30$, we firstly simulate a Gaussian Random Field $X(\boldsymbol{u}, t)$

with mean and covariance function

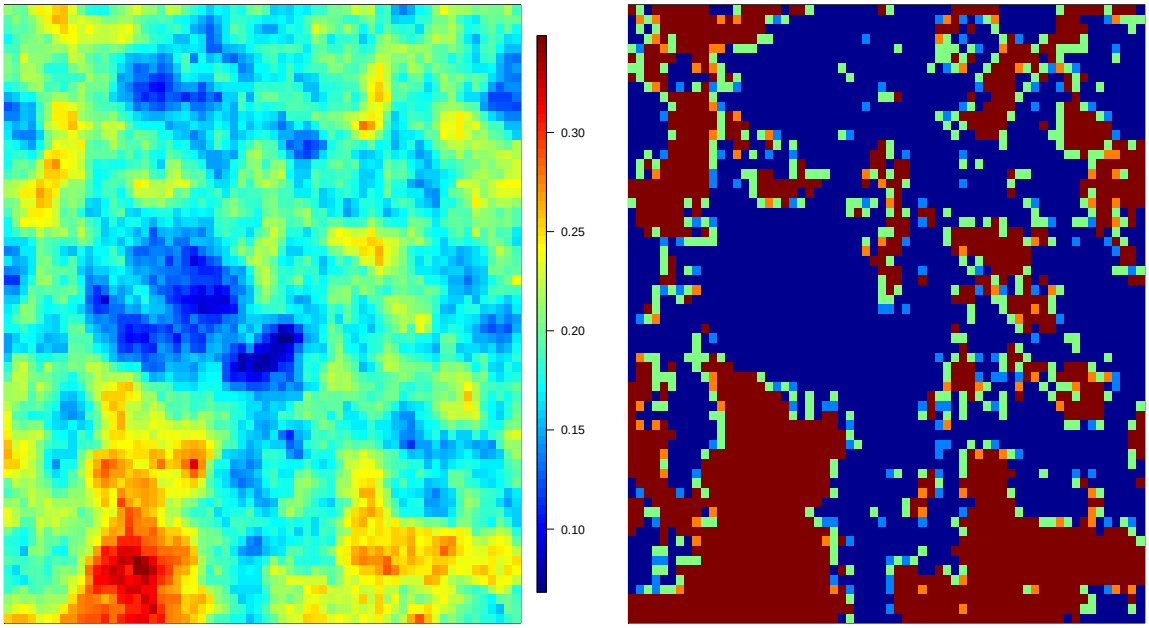
$$\begin{aligned} E\left(X(\boldsymbol{u},t)\right) &= 0, \\ \Gamma_X(l,h) &= e^{-\alpha l/2-\beta h/2}. \end{aligned} \tag{B.3}$$

Then $Z(\boldsymbol{u},t)$ is calculated as

$$Z(\boldsymbol{u},t) = 1 + \frac{\sigma}{2}\left(X^2(\boldsymbol{u},t) - 1\right). \tag{B.4}$$

Simple calculations can show that this transformation results in a shifted Chi-squared distribution of $Z(\boldsymbol{u},t)$ with $E\left(Z(\boldsymbol{u},t)\right) = 1$ and $\Gamma_Z(l,h) = e^{-\alpha l-\beta h}$.

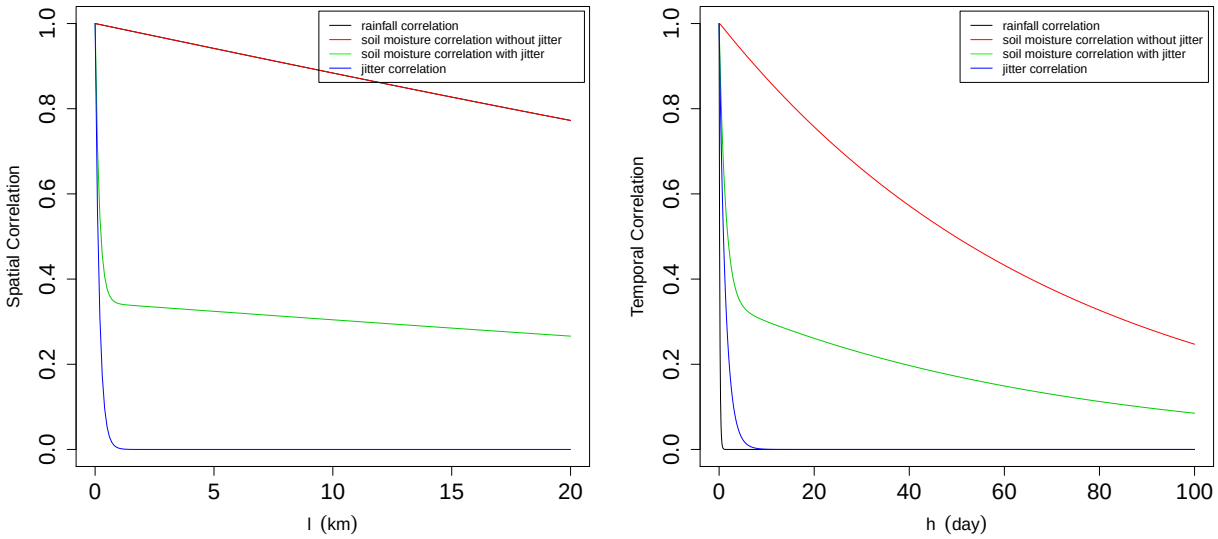
B.2 Examples of the soil moisture field and soil moisture islands



(a) (b)

Figure B.1: Examples of the soil moisture field and soil moisture islands. (a), the soil moisture field at a moment in time. (b), soil moisture islands (in red) above a threshold equal to 0.19.

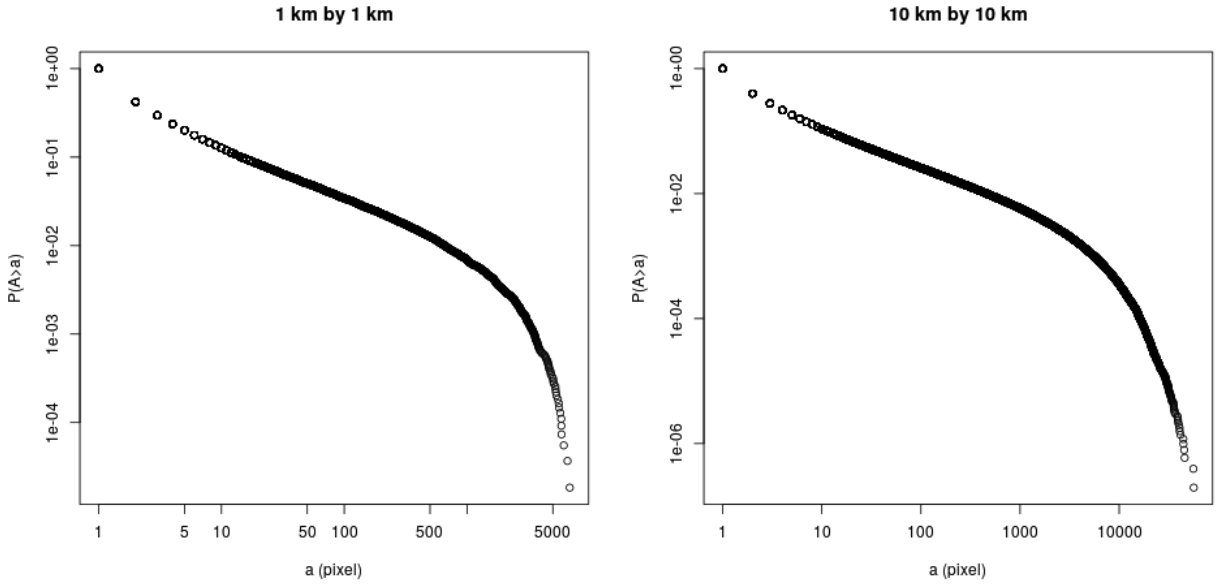
B.3 Correlation functions



(a) (b)

Figure B.2: Correlation functions of the rainfall process $Y(\boldsymbol{u}, t)$, the soil moisture field driven by rainfall $S(\boldsymbol{u}, t)$, the soil moisture field with jitter $\tilde{S}(\boldsymbol{u}, t)$, and the jitter process $Z(\boldsymbol{u}, t)$ with parameters estimated in Section 2.5. (a), the spatial correlation for $0 \leqslant l \leqslant 20$ km. Note that the spatial correlation function of $S(\boldsymbol{u}, t)$ and $Y(\boldsymbol{u}, t)$ are the same ([Rodriguez-Iturbe et al., 1987]). (b), the temporal correlation for $0 \leqslant h \leqslant 100$ days.

B.4 Distributions of jitter islands



(a) (b)

Figure B.3: Distributions of jitter islands plotted in log-log scale. The threshold for both plots is 1.0, and a pixel for both plots is 10 m $\times$ 10 m. (a), the jitter process is simulated on a 1 km $\times$ 1 km field (10^4 pixels in total). The power law exists until the number of pixels is close to the maximum number of pixels. (b), the jitter process is simulated on a 10 km $\times$ 10 km field (10^6 pixels in total). The curve becomes bent before 10^4 pixels, which cannot result from an edge effect since 10^4 pixels only cover 1% of the total area.

B.5 Area vs. perimeter of soil moisture islands

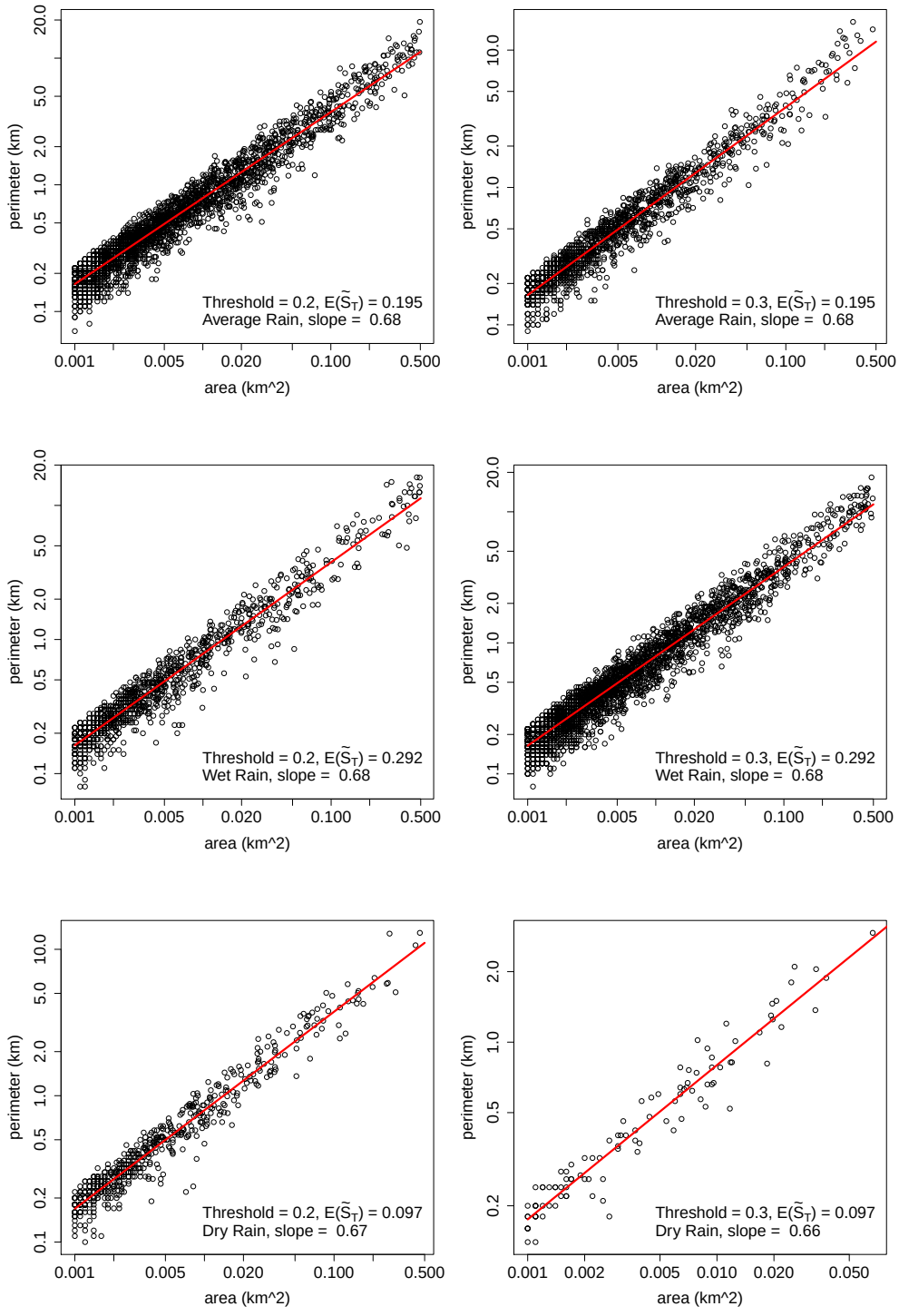


Figure B.4: Perimeter vs. area are plotted at log-log scale with (a_1, b_1) and three different mean growing season rainfall (average, wet and dry). These plots show that the slopes are very close under different cases.

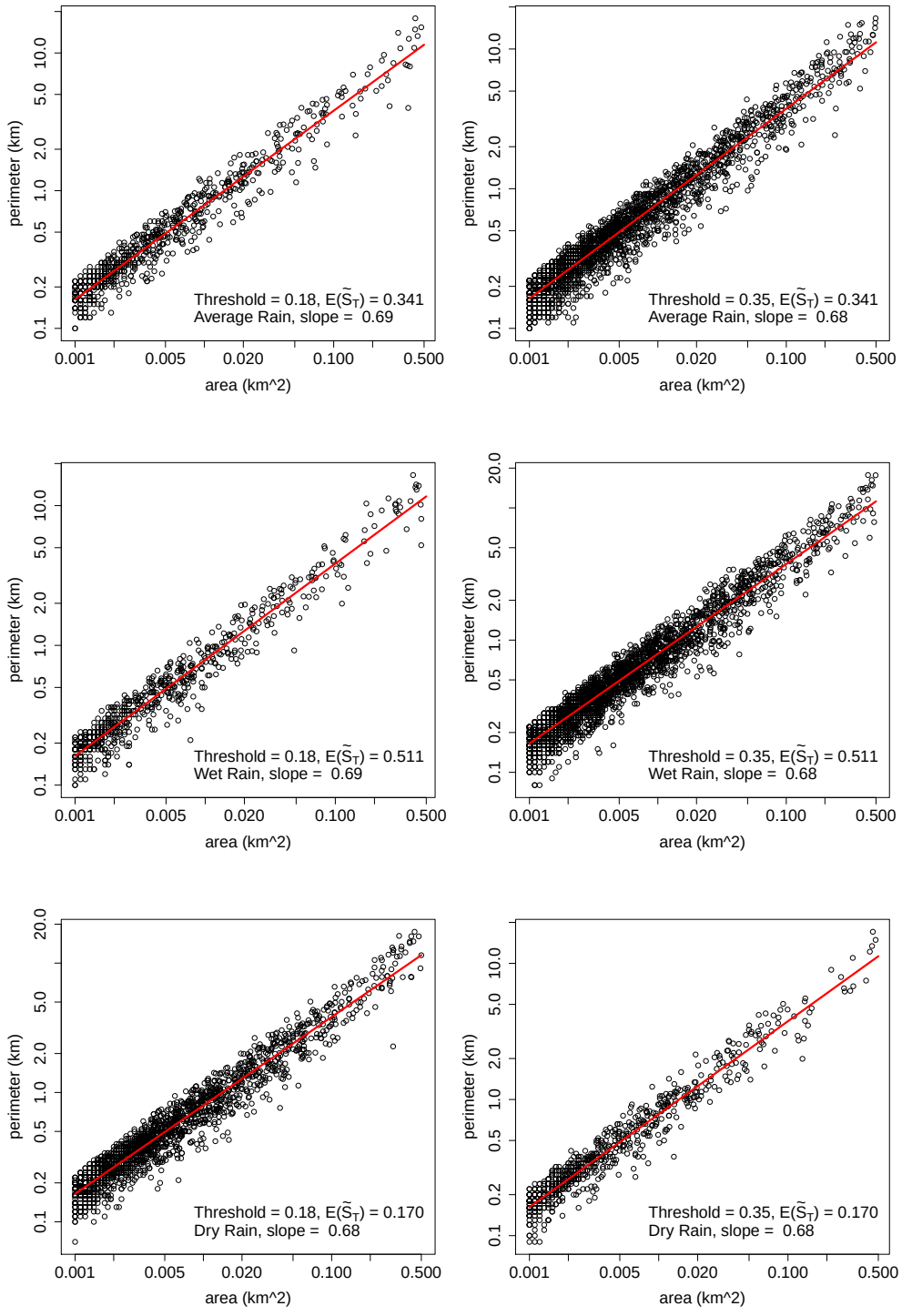
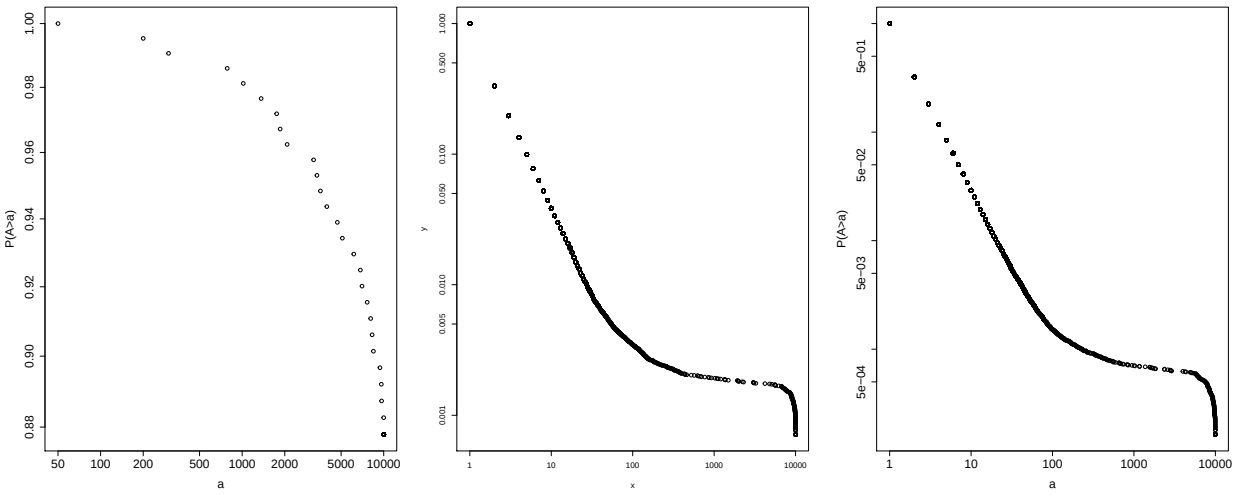


Figure B.5: Perimeter vs. area are plotted at log-log scale with (a_2, b_2) and three different mean growing season rainfall (average, wet and dry). These plots show that the slopes are very close under different cases.

B.6 Distributions of soil moisture islands with and without jitter processes



(a) (b) (c)

Figure B.6: Distributions of soil moisture islands with and without jitter processes. (a), the soil moisture islands on the soil moisture field without jitter. (b), the soil moisture islands on the soil moisture field with independent jitter in both space and time. (c), the soil moisture islands on the soil moisture field with jitter independent in space but correlated in time. Here $a = a_1 = 0.014$ day^{-1} and $b = b_1 = 0.002$ mm^{-1} are used, and the rainfall parameters and jitter parameters are from the 'average' column in Table 3.1. The soil moisture islands are calculated with threshold equal to 0.2.

APPENDIX C

SUPPLEMENTARY INFORMATION FOR CHAPTER 4

C.1 Proof of Theorem 4.2.1

Proof. For any $s \in C$, there exists m such that $s \in A_m$. By Definitions 4.2.2 and 4.2.3, $\forall \tilde{s} \in A_m$, $\tilde{s} \sim s$. Thus we have $\tilde{s} \in C$ by Definition 4.2.4. Hence $A_m \subset C$ and $s \in A_m \subset \bigcup_{i:A_i \subset C} A_i$. Since s is arbitrary, we conclude that $C \subset \bigcup_{i:A_i \subset C} A_i$ and thus $C = \bigcup_{i:A_i \subset C} A_i$.

□

C.2 Proof of Theorem 4.3.1

Proof. The proof consists of two parts.

(Necessity) Assume that $s' \overset{G}{\sim} s''$. Then there exists $K > 0$ such that $s' \overset{G_k}{\sim} s''$ when $k > K$. Fix $k > \max\{K, -\log_2 \delta\}$. Then since $2^{-k} < \delta$, Definition 4.2.6 implies that there exist $s_1, s_2, ..., s_n \in A \cap G_k$ such that $\|s_{i-1} - s_i\| < \delta$, $i = 1, 2, ..., n+1$, where $s_0 = s'$ and $s_{n+1} = s''$. Thus the condition of Definition 4.2.3 is satisfied for all $i = 1, 2, ..., n+1$, which means $s' \sim s''$.

(Sufficiency) Assume that $s' \sim s''$. Then there exist $0 \leqslant n < \infty$ and $s_1, s_2, ..., s_n \in A$, such that for each $i = 1, 2, ..., n+1$, $\|s_{i-1} - s_i\| < \delta$ ($s' = s_0$, $s'' = s_{n+1}$). Note that $s_1, s_2, ..., s_n$ might not belong to G. Let

$$\epsilon = \frac{1}{2}\left(\delta - \max_{i=1,2,...,n+1} \|s_{i-1} - s_i\|\right) > 0.$$

Since G is dense in D and A is open, for each $i = 1, ..., n$, there exists $\tilde{s}_i$ such that $\tilde{s}_i \in G \cap A \cap B(s_i, \epsilon)$. Let $\tilde{s}_0 = s'$, $\tilde{s}_{n+1} = s''$. Then we have

$$\begin{aligned}\|\tilde{s}_{i-1} - \tilde{s}_i\| &\leqslant \|\tilde{s}_{i-1} - s_{i-1}\| + \|s_{i-1} - s_i\| + \|s_i - \tilde{s}_i\| \\ &< \epsilon + \max_{i=1,2,...,n+1} \|s_{i-1} - s_i\| + \epsilon = \delta.\end{aligned}$$

Besides, since for all $i = 0, 1, ..., n+1$, $\tilde{s}_i \in A \cap G$, there exists $K_i > 0$, such that $\tilde{s}_i \in A \cap G_{K_i}$ for all $k > K_i$. Let $K = \max\{K_0, K_1, ..., K_{n+1}\}$, we have, when $k > K$, $\tilde{s}_i \in A \cap G_k$ for all $i = 0, 1, ..., n+1$. This satisfies the condition (2) in Definition 4.2.6, which implies $s' \overset{G_k}{\sim} s''$ for all $k > K$. Hence $s' \overset{G}{\sim} s''$, as desired.

□

C.3 Proof of Theorem 4.3.2

Proof. Let $\sigma(X)$ denote the σ-algebra generated by a random variable X (information contained by X). Fix $x \in \mathbb{R}$ and $k \in \mathbb{N}_+$. It suffices to show that $N_{x,k}$ is a composition of two measurable functions. Let E_k be the set of $2^k \times 2^k$ matrices whose entries are 0 or 1, and the collection of all subsets of E_k is defined as the σ-algebra in E_k. Define $f_k : \Omega \mapsto E_k$ such that

$$f_k(\omega)_{(m,n)} = 1_{\{y\left(m/2^k, n/2^k\right)>0\}}, \quad m = 1, 2, ..., 2^k, \quad n = 1, 2, ..., 2^k,$$

i.e., the $(m,n)^{th}$ entry of $f_k(\omega)$ is equal to 1 if $y\left(m/2^k, n/2^k\right) > 0$ and is equal to 0 otherwise. Then f_k is measurable since $\forall e \in E_k$ we can express $f_k^{-1}(e)$ as

$$f_k^{-1}(e) = \bigcap_{n=1}^{2^k} \bigcap_{m=1}^{2^k} F_{m,n},$$

where $F_{m,n} = \{y\left(m/2^k, n/2^k\right) > 0\}$ or $F_{m,n} = \{y\left(m/2^k, n/2^k\right) \leqslant 0\}$, depending on $e(m,n)$, the $(m,n)^{th}$ entry of e. Now for $e \in E_k$, let

$$Q_e = \left\{(m,n) : e(m,n) = 1, m = 1, 2, ..., 2^k, n = 1, 2, ..., 2^k\right\}.$$

We define an equivalence class $\overset{e}{\sim}$ in Q_e (similar to Definition 4.2.6) as follows: if $q' = (m', n')$ and $q'' = (m'', n'')$, then $q' \overset{e}{\sim} q''$ if there exist $0 \leqslant l < \infty$ and $q_1, q_2, ..., q_l \in Q_e$, such that for each $i = 1, 2, ..., l+1$, at least one of the following two conditions is satisfied ($q_0 = s', q_{l+1} = q'', q_i = (m_i, n_i), i = 0, 1, ..., l+1$):

(1) $|m_{i-1} - m_i| + |n_{i-1} - n_i| = 1$;

(2) $|m_{i-1} - m_i|^2 + |n_{i-1} - n_i|^2 < (\delta \cdot 2^k)^2$.

Now we define $g_{x,k}$ as the number of "clusters" with size larger than x on a given matrix in E_k:

$$g_{x,k} : E_k \mapsto \mathbb{N}, \quad g_{x,k}(e) = card\left(\left\{U \in Q_e/\overset{e}{\sim}: card(U) \cdot 4^{-k} > x\right\}\right).$$

Then automatically $g_{x,k}$ is measurable since any subset of E_k belongs to the σ-algebra defined in E_k. Finally, it is clear that $N_{x,k} = g_{x,k} \circ f_k$, which completes the proof. □

C.4 Proof of Lemma 4.3.1

Proof. We first prove the case when $x = 0$ (note that $U_0 = \Omega$). Fix $\omega \in \Omega$. Note that from Theorem 4.2.1,

$$C_i = \bigcup_{j:A_j \subset C_i} A_j, \quad i = 1, 2, ..., N_0,$$

where C_i's are open, nonempty and mutually exclusive. Since G is dense in D, there exists $K_1 \in \mathbb{N}_+$ such that for any $k \geqslant K_1$, $C_i \cap G_k \neq \emptyset$, $i = 1, 2, ..., N_0$. Besides, from the first part of the proof of Theorem 4.3.1, there exists $K_2 \in \mathbb{N}_+$ such that for any $k \geqslant K_2$, $s' \overset{G_k}{\sim} s''$ implies $s' \sim s''$. Let $K = \max\{K_1, K_2\}$. Then we can choose s_i such that $s_i \in C_i \cap G_K$, $i = 1, 2, ..., N_0$. For $k > K$, if $i \neq j$ then s_i and s_j must belong to different clusters in G_k since $s_i \overset{G_k}{\sim} s_j$ would imply $s_i \sim s_j$. Therefore, there should be at least N_0 different clusters in G_k and thus $N_{0,k} \geqslant N_0$ when $k > K$. Therefore, we have

$$\liminf_{k\to\infty} N_{0,k} \geqslant N_0.$$

Now it suffices to show

$$\limsup_{k\to\infty} N_{0,k} \leqslant N_0.$$

Suppose $\limsup_{k\to\infty} N_{0,k} > N_0$. Then there exists an increasing sequence of integer $\{k_n\}_{n=1}^{\infty}$ such that $k_1 > K$ and $N_{0,k_n} > N_0$ for all n. Let C_{1,k_n}, C_{2,k_n},

$..., C_{N_{0,k_n},k_n}$ be the clusters of G_{k_n} and $s_j \in C_{j,k_n} \subset C_j$, $j = 1, 2, ..., N_0$, as chosen above. Let $\tilde{s}_n$

be a point in C_{N_0+1,k_n}, $n = 1, 2, ...$ Since D is compact in $\mathbb{R}^2$, there exists a subsequence of $\tilde{s}_n$ that converges to a limit $\tilde{s} \in D$. WLOG, we can assume $\tilde{s}_n \to \tilde{s}$ as $n \to \infty$. Then there exists $N \in \mathbb{N}_+$ such that $\|\tilde{s}_n - \tilde{s}_N\| < \delta$ for all $n \geqslant N$. Again, WLOG, assume $\tilde{s}_N \in C_1$. Then $\tilde{s}_N \sim s_1$, which implies $\tilde{s}_N \overset{G}{\sim} s_1$ by Theorem 4.3.1. Hence there exists $n_0 > N$ such that $\tilde{s}_N \overset{G_{k_{n_0}}}{\sim} s_1$. Since $\|\tilde{s}_{n_0} - \tilde{s}_N\| < \delta$, we have

$$s_1 \overset{G_{k_{n_0}}}{\sim} \tilde{s}_N \overset{G_{k_{n_0}}}{\sim} \tilde{s}_{n_0}.$$

However, this is a contradiction since $s_1 \in C_{1,k_{n_0}}$, $\tilde{s}_{n_0} \in C_{N_0+1,k_{n_0}}$, s_1 and $\tilde{s}_{n_0}$ would not belong to a same cluster of $G_{k_{n_0}}$. Hence we conclude that $N_{0,k} \to N_0$ for all $\omega \in \Omega = U_0$, which also implies that N_0 is a random variable.

Now we prove the case when $x > 0$. From the proof above, there exists $K \in \mathbb{N}_+$ such that $N_{0,k} = N_0$ and $s' \overset{G_k}{\sim} s''$ implies $s' \sim s''$ for all $k \geqslant K$. Let $C_{1,k}, C_{2,k}, ..., C_{N_0,k}$ be the clusters of G_k. Then for any i and j, either $C_{j,k} \subset C_i$ or $C_{j,k} \cap C_i = \emptyset$. Assume that $C_{j,k} \subset C_j$, $j = 1, 2, ..., N_0$. Then we have

$$\lambda_k(C_{j,k}) = \left(\frac{1}{2^k}\right)^2 \sum_{s \in G_k} 1_{C_{j,k}}(s) = \left(\frac{1}{2^k}\right)^2 \sum_{s \in G_k} 1_{C_j}(s), \quad j = 1, 2, ..., N_0,$$

which is a Riemann sum. To show that it converges to the Lebesgue integral

$$\int_D 1_{C_j}(s) ds = \lambda(C_j), \quad j = 1, 2, ..., N_0,$$

it suffices to show that the set of discontinuous points of 1_{C_j} has zero Lebesgue measure, i.e., $\lambda(\partial C_j) = 0$ for $j = 1, 2, ..., N_0$. For $j = 1$, suppose $s \in \partial C_1$. Then for any $0 < \epsilon < \delta/2$, there exist s' and s'' such that $s' \in B(s, \epsilon) \cap C_1^c$ and $s'' \in B(s, \epsilon) \cap C_1$, $y(s'') > 0$. Since $\|s' - s''\| < \delta$, if $y(s') > 0$, we would have $s' \sim s''$ and $s' \in C_1$, which contradicts the fact that $s' \in C_1^c$. Thus $y(s') \leqslant 0$. Therefore, for any $0 < \epsilon < \delta/2$, there exist s' and s'' in $B(s, \epsilon)$ such that $y(s') \leqslant 0$,

$y(s'') > 0$. By the definition of $\partial A(\omega)$, we have $s \in \partial A(\omega)$. Therefore, we have shown that

$$\partial C_1 \subset \partial A(\omega) = \partial \{s \in D : y(s, \omega) > 0\},$$

and thus

$$\lambda(\partial C_1) = \lambda(\partial A(\omega)) = 0.$$

Similarly, this equation holds when $j = 2, 3, ..., N_0$. Therefore, $\forall \omega \in \Omega$,

$$\lambda_k(C_{j,k}) \to \lambda(C_j), \quad j = 1, 2, ..., N_0(\omega). \tag{C.1}$$

Now to show N_x is a random variable, it suffices to show that $\forall l \in \mathbb{N}_+$,

$$\{\omega \in \Omega : N_x(\omega) = l\} = \bigcup_{p=1}^{\infty} \bigcap_{q=p}^{\infty} \bigcup_{r=1}^{\infty} \bigcap_{k=r}^{\infty} \left\{\omega \in \Omega : N_{x+1/q,k}(\omega) = l\right\}.$$

WLOG, assume $\lambda(C_1) \geqslant \lambda(C_2) \geqslant \cdots \geqslant \lambda(C_{N_0})$. Suppose $N_x(\omega) = l$. Let $p = 1$ if $l = 0$. Otherwise, choose $p > 0$ large enough such that $x + 1/p < \lambda(C_{N_x})$. Then we have $\lambda(C_i) > x + 1/p$, $i = 1, 2, ..., N_x$ and $\lambda(C_i) < x + 1/p$, $i = N_x + 1, N_x + 2, ..., N_0$. By (C.1), there exists $r > 0$, such that for all $k \geqslant r$, $\lambda_k(C_{j,k}) > x + 1/p$, $j = 1, 2, ..., N_x$ and $\lambda_k(C_{j,k}) < x = 1/p$, $j = N_x + 1, N_x + 2, ..., N_0$. Therefore, for all $q \geqslant p$ and $k \geqslant r$, $N_{x+1/q,k}(\omega) = l$, which implies that ω is in the right hand side (RHS). Thus the left hand side (LHS) $\subset$ RHS.

Now suppose that ω is in RHS. Then there exists $q_0 > 0$ such that

$$\omega \in \bigcup_{r=1}^{\infty} \bigcap_{k=r}^{\infty} \left\{N_{x+1/q_0,k}(\omega) = l\right\},$$

which indicates that $N_x(\omega) \geqslant l$ by (C.1). Suppose $N_x(\omega) > l$. Then for all $q > 1/(\lambda(C_{N_x}) - x)$, we have $x + 1/q < \lambda(C_{N_x})$. Similarly, as the proof above, when k is large enough, we would have $N_{x+1/q,k}(\omega) > l$, which is a contradiction since ω is in RHS. Hence $N_x(\omega) = l$ and ω is in LHS. This implies LHS = RHS.

Finally, assume $\omega \in U_x$. Again, WLOG, we can assume $\lambda(C_i) > x$, $i = 1, 2, ..., N_x$ and $\lambda(C_i) < x$, $i = N_x + 1, N_x + 2, ..., N_0$. Then we have, for sufficiently large k, $\lambda_k(C_{j,k}) > x$, $j = 1, 2, ..., N_x$ and $\lambda_k(C_{j,k}) < x$, $j = N_x + 1, N_x + 2, ..., N_0$. Thus $N_{x,k} = N_x$ for sufficiently large k, which implies $\forall \omega \in U_x$, $N_{x,k} \to N_x$ as $k \to \infty$, as desired. □

C.5 Proof of Lemma 4.3.2

Proof. Firstly, note that for any $x \in B_0$ and $\epsilon > 0$, there exist x' and x'' such that $x', x'' \in B(x, \epsilon)$ and $f(x') > 0$, $f(x'') \leqslant 0$. Since f is continuous, we have $f(x) = 0$ and $B_0 \subset \{f(x) = 0\}$. Since

$$\begin{aligned}
\lambda(B_0) &= \lambda(B_0 \cap \{\nabla f(x) = 0\}) + \lambda(B_0 \cap \{\nabla f(x) \neq 0\}) \\
&\leqslant \lambda(\{f(x) = 0\} \cap \{\nabla f(x) = 0\}) + \lambda(\{f(x) = 0\} \cap \{\nabla f(x) \neq 0\}) \\
&= \lambda(R_0) + \lambda(\{f(x) = 0\} \cap \{\nabla f(x) \neq 0\}),
\end{aligned}$$

it suffices to show $\lambda(\{f(x) = 0\} \cap \{\nabla f(x) \neq 0\}) = 0$. Since $\lambda(\partial D) = 0$, it suffices to show

$$\lambda(\{f(x) = 0\} \cap \{\nabla f(x) \neq 0\} \cap D^{\mathrm{o}}) = 0,$$

where $D^{\mathrm{o}} = (0, 1)^2$ is the interior of D. Let $S_0 = \{f(x) = 0\} \cap \{\nabla f(x) \neq 0\} \cap D^{\mathrm{o}}$. Now suppose $x = (x_1, x_2) \in S_0$. WLOG, we can assume that $\partial f(x)/\partial x_2 \neq 0$. By the Implicit Function Theorem, there exist $\gamma_x^1, \gamma_x^2 > 0$ and a continuously differentiable function $g : I_x \mapsto J_x$ such that

$$\begin{aligned}
\{(y_1, g(y_1)) : y_1 \in I_x\} &= \{(y_1, y_2) \in I_x \times J_x : f(y_1, y_2) = 0\} \\
&= (I_x \times J_x) \cap \{f(y) = 0\},
\end{aligned}$$

where $I_x = (x_1 - \gamma_x^1, x_1 + \gamma_x^1)$ and $J_x = (x_2 - \gamma_x^2, x_2 + \gamma_x^2)$ are the neighborhoods of x_1 and x_2 respectively. Now we show that the above set has zero Lebesgue measure. For any $0 < \gamma < \gamma_x^1$ and $\epsilon > 0$, since g is uniformly continuous in $[x_1 - \gamma, x_1 + \gamma]$, there exists $\xi > 0$ such that when $z', z'' \in [x_1 - \gamma, x_1 + \gamma]$ and $|z_1 - z_2| < \xi$, $|g(z_1) - g(z_2)| < \epsilon$. Choose $K \in \mathbb{N}$ such that $2\gamma/K < \xi$

and let

$$z_k = x_1 - \gamma + \frac{2\gamma k}{K}, \quad k = 0, 1, 2, ..., K.$$

Then we have

$$\{(z, g(z)) : z \in [x_1 - \gamma, x_1 + \gamma]\} \subset$$
$$\bigcup_{k=1}^{K} [z_{k-1}, z_k] \times \left[g\left(\frac{z_{k-1} + z_k}{2}\right) - \epsilon, g\left(\frac{z_{k-1} + z_k}{2}\right) + \epsilon \right],$$

where the Lebesgue measure of RHS is

$$\sum_{k=1}^{K} 2(z_k - z_{k-1})\epsilon = 4\gamma\epsilon.$$

Since ϵ is arbitrary, we conclude that

$$\lambda\left(\{(z, g(z)) : z \in [x_1 - \gamma, x_1 + \gamma]\}\right) = 0.$$

By letting $\gamma \uparrow \gamma_x^1$, we obtain

$$\lambda\left(\{(y_1, g(y_1)) : y_1 \in I_x\}\right) = 0.$$

Finally, note that $I_x \times J_x$ exists for each x in S_0, $\{I_x \times J_x\}_{x \in S_0}$ is an open cover of S_0. Since $\mathbb{R}^2$ is Lindelof, there exists a countable subcover $\{I_n \times J_n\}_{n=1}^{\infty}$ of S_0 ([Gemignani, 1990], Chapter 7). Therefore, we have

$$\begin{aligned}
\lambda(S_0) &= \lambda\left(\bigcup_{n=1}^{\infty} (S_0 \cap (I_n \times J_n))\right) \\
&\leqslant \sum_{n=1}^{\infty} \lambda\left(S_0 \cap (I_n \times J_n)\right) \\
&\leqslant \sum_{n=1}^{\infty} \lambda\left(\{f(y) = 0\} \cap (I_n \times J_n)\right) \\
&= \sum_{n=1}^{\infty} \lambda\left(\{(y_1, g(y_1)) : y_1 \in I_n\}\right) = 0,
\end{aligned}$$

which completes the proof.

□

APPENDIX D

SUPPLEMENTARY INFORMATION FOR CHAPTER 5

D.1 Proof of Theorem 5.2.1

Let $\ker(A)$ denote the kernel (null space) of matrix A. The following lemma is necessary for the proof of Theorem 5.2.1.

Lemma D.1.1. *Suppose* $1_n \notin \mathscr{C}(X)$. *Then*

$$\ker(I_n - D_\theta) = \text{span}\{T, 1_n\}.$$

Proof. Firstly, we have, for $i = 1, 2, ..., n$,

$$(t_i, 0) = \underset{a_i, b_i \in \mathbb{R}}{\arg\min} \sum_{j=1}^{n} (t_j - a_i - b_i t_{ij})^2 K_{ij},$$

$$(1, 0) = \underset{a_i, b_i \in \mathbb{R}}{\arg\min} \sum_{j=1}^{n} (1 - a_i - b_i t_{ij})^2 K_{ij}.$$

Recall that the optimization problem

$$\left(\hat{a}_i, \hat{b}_i\right) = \underset{a_i, b_i \in \mathbb{R}}{\arg\min} \sum_{j=1}^{n} (z_j - a_i - b_i t_{ij})^2 K_{ij}$$

has unique solution

$$\hat{a}_i = \sum_{j=1}^{n} (D_\theta)_{ij} z_j. \tag{D.1}$$

Therefore, we have $D_\theta T = T$ and $D_\theta 1_n = 1_n$. Hence, span$\{T, 1_n\} \subset \ker(I_n - D_\theta)$, and it suffices to show

$$\ker(I_n - D_\theta) \subset \text{span}\{T, 1_n\}.$$

This is equivalent to show that for any $\xi = (\xi_1, \xi_2, ..., \xi_n)^T \in \mathbb{R}^n$ satisfying $(I_n - D_\theta)\,\xi = 0$, there

exist $\alpha_1, \alpha_2 \in \mathbb{R}$ such that $\xi_i = \alpha_1 t_i + \alpha_2$ $(i = 1, 2, ..., n)$. It means that $\{(t_i, \xi_i)\}_{i=1}^n$ are all on a straight line of $\mathbb{R}^2$. Since $(I_n - D_\theta)\,\xi = 0$, we have

$$\left(\xi_i, \hat{b}_i\right) = \underset{a_i, b_i \in \mathbb{R}}{\arg\min} \sum_{j=1}^{n} \left(\xi_j - a_i - b_i t_{ij}\right)^2 K_{ij} \quad (i = 1, 2, ..., n).$$

If $t_{i_1} = t_{i_2}$, we have $t_{i_1 j} = t_{i_2 j}$ and $K_{i_1 j} = K_{i_2 j}$ for $j = 1, 2, ..., n$ by definition. Therefore, the right hand side of the equation above is the same for $i = i_1$ and $i = i_2$, which implies $\xi_{i_1} = \xi_{i_2}$. Let $\xi_{ij} = \xi_j - \xi_i$ and define the slope of the line containing (t_i, ξ_i) and (t_j, ξ_j) by

$$l_{ij} = \frac{\xi_j - \xi_i}{t_j - t_i} I\left(t_j \neq t_i\right) = \frac{\xi_{ij}}{t_{ij}} I\left(t_{ij} \neq 0\right) \quad (i, j = 1, 2, ..., n).$$

Then the observation above indicates that

$$\xi_{ij} = l_{ij} t_{ij}, \quad (i, j = 1, 2, ..., n). \tag{D.2}$$

Again, since $(I_n - D_\theta)\,\xi = 0$, we have that, for all $i = 1, 2, ..., n$,

$$\xi_i = \frac{\left(\sum_{s=1}^n K_{is} t_{is}^2\right)\left(\sum_{j=1}^n K_{ij} \xi_j\right) - \left(\sum_{s=1}^n K_{is} t_{is}\right)\left(\sum_{j=1}^n K_{ij} t_{ij} \xi_j\right)}{\left(\sum_{j=1}^n K_{ij}\right)\left(\sum_{s=1}^n K_{is} t_{is}^2\right) - \left(\sum_{j=1}^n K_{ij} t_{ij}\right)\left(\sum_{s=1}^n K_{is} t_{is}\right)},$$

which, by (D.2), can be rearranged as

$$\sum_{s=1}^{n} K_{is} t_{is}^2 \left(\sum_{j=1}^{n} K_{ij} \left(l_{ij} - l_{is}\right) t_{ij}\right) = 0. \tag{D.3}$$

Without loss of generality, assume $t_1 \leqslant t_2 \leqslant \cdots \leqslant t_n$. Let

$$k = \min\{1 \leqslant s \leqslant n : t_s > t_1\}.$$

Then $(t_1, \xi_1) = (t_2, \xi_2) = \cdots = (t_{k-1}, \xi_{k-1})$. Since $T \neq c1_n$, $k < n$, $K(\cdot) > 0$ and (D.3), we have

$$\sum_{j=1}^{n} K_{1j} \left(l_{1j} - l_{1s}\right) t_{1j} = \sum_{j=k}^{n} K_{1j} t_{1j} \left(l_{1j} - l_{1s}\right) = 0$$

for all $s \geqslant k$. Let $m = \underset{s \geqslant k}{\arg\max}\{l_{1s}\}$ and replace s by m in the equation above. Since $K_{1j} t_{1j} > 0$ for $j \geqslant k$, we have

$$l_{1j} - l_{1m} = 0 \quad (j = k, k+1, ..., n),$$

which implies that $(t_k, \xi_k), (t_{k+1}, \xi_{k+1}), \ldots, (t_n, \xi_n)$ are all on the line containing (t_1, ξ_1) and (t_m, ξ_m). Since $(t_1, \xi_1) = (t_2, \xi_2) = \cdots = (t_{k-1}, \xi_{k-1})$ are also on the same line, the proof is complete.

□

Proof of Theorem 5.2.1. Firstly, we prove that the first equation in Equation (5.10) holds. By the definition of $\hat{g}(x_i^T \theta | \beta, \theta)$ and D_θ, $\hat{\beta}_\theta$ can be rewritten as

$$\hat{\beta}_\theta = \underset{\beta^T \theta = 0}{\arg\min} \sum_{i=1}^{n} \left(y_i - x_i^T \beta - \hat{g}(x_i^T \theta | \beta, \theta)\right)^2 = \underset{\beta^T \theta = 0}{\arg\min} \left\| \tilde{Y}_\theta - \tilde{X}_\theta \beta \right\|^2.$$

By taking the derivative of the right hand side of the equation above, it suffices to show that

$$\beta_\theta^* = \left(\tilde{X}_\theta^T \tilde{X}_\theta\right)^+ \tilde{X}_\theta^T \tilde{Y}_\theta$$

satisfies $\tilde{X}_\theta^T \tilde{X}_\theta \beta_\theta^* = \tilde{X}_\theta^T \tilde{Y}_\theta$ and $\beta_\theta^{*T} \theta = 0$. By basic linear algebra, it is trivial to see that $\tilde{X}_\theta^T \tilde{X}_\theta \beta_\theta^* = \tilde{X}_\theta^T \tilde{Y}_\theta$. Besides, by the properties of the Moore–Penrose inverse and $\left(I_n - D_\theta\right) X\theta =$

0 by Lemma D.1.1, we have

$$\begin{aligned}
\theta^T \beta_\theta^* &= \theta^T \left(\tilde{X}_\theta^T \tilde{X}_\theta\right)^+ \tilde{X}_\theta^T \tilde{Y}_\theta \\
&= \theta^T \left(\tilde{X}_\theta^T \tilde{X}_\theta\right)^+ \left(\tilde{X}_\theta^T \tilde{X}_\theta\right) \left(\tilde{X}_\theta^T \tilde{X}_\theta\right)^+ \tilde{X}_\theta^T \tilde{Y}_\theta \\
&= \theta^T \left[\left(\tilde{X}_\theta^T \tilde{X}_\theta\right)^+ \left(\tilde{X}_\theta^T \tilde{X}_\theta\right)\right]^T \left(\tilde{X}_\theta^T \tilde{X}_\theta\right)^+ \tilde{X}_\theta^T \tilde{Y}_\theta \\
&= \left[\theta^T \tilde{X}_\theta^T\right] \tilde{X}_\theta \left(\tilde{X}_\theta^T \tilde{X}_\theta\right)^+ \left(\tilde{X}_\theta^T \tilde{X}_\theta\right)^+ \tilde{X}_\theta^T \tilde{Y}_\theta \\
&= \left[\left(I_n - D_\theta\right) X\theta\right]^T \tilde{X}_\theta \left(\tilde{X}_\theta^T \tilde{X}_\theta\right)^+ \left(\tilde{X}_\theta^T \tilde{X}_\theta\right)^+ \tilde{X}_\theta^T \tilde{Y}_\theta \\
&= 0.
\end{aligned}$$

Therefore, β_θ^* is a solution to (5.5). To show that it is unique, let $\tilde{\beta}_\theta$ be another solution to (5.5). Then we have

$$\begin{aligned}
\left\|\tilde{Y}_\theta - \tilde{X}_\theta \tilde{\beta}_\theta\right\|^2 &= \left\|\left(\tilde{Y}_\theta - \tilde{X}_\theta \beta_\theta^*\right) + \left(\tilde{X}_\theta \beta_\theta^* - \tilde{X}_\theta \tilde{\beta}_\theta\right)\right\|^2 \\
&= \left\|\tilde{Y}_\theta - \tilde{X}_\theta \beta_\theta^*\right\|^2 + \left\|\tilde{X}_\theta \left(\beta_\theta^* - \tilde{\beta}_\theta\right)\right\|^2,
\end{aligned}$$

which implies $\tilde{X}_\theta \left(\beta_\theta^* - \tilde{\beta}_\theta\right) = \left(I_n - D_\theta\right) X \left(\beta_\theta^* - \tilde{\beta}_\theta\right) = 0$. By Lemma D.1.1, there exist $\alpha_1, \alpha_2 \in \mathbb{R}$ such that $X \left(\beta_\theta^* - \tilde{\beta}_\theta\right) = \alpha_1 X\theta + \alpha_2 1_n$, or

$$X \left(\beta_\theta^* - \tilde{\beta}_\theta - \alpha_1 \theta\right) = \alpha_2 1_n.$$

Since $1_n \notin \mathscr{C}(X)$, $\alpha_2 = 0$. Then $\operatorname{rank}(X) = p$ implies

$$\beta_\theta^* - \tilde{\beta}_\theta = \alpha_1 \theta.$$

Since $\theta^T \beta_\theta^* = \theta^T \tilde{\beta}_\theta = 0$, multiplying both sides of the equation above by θ^T yields $\alpha_1 = 0$. Hence $\beta_\theta^* = \tilde{\beta}_\theta$ and the solution to (5.5) is unique.

Now by the uniqueness of the solution, to show the second equation in Equation (5.10), it

suffices to show

$$\tilde{\beta}_\theta = \left(\tilde{X}_\theta^T \tilde{X}_\theta + \theta\theta^T\right)^{-1} \tilde{X}_\theta^T \tilde{Y}_\theta$$

is also a solution to (5.5). According to Lemma D.1.1, $\mathrm{rank}(\tilde{X}_\theta) = p - 1$ and $\theta \notin \mathscr{C}(\tilde{X}_\theta^T)$. Then we have

$$\mathrm{rank}\left(\tilde{X}_\theta^T \tilde{X}_\theta + \theta\theta^T\right) = \mathrm{rank}\left(\left(\tilde{X}_\theta^T, \theta\right)\right) = p$$

and thus $\tilde{X}_\theta^T \tilde{X}_\theta + \theta\theta^T$ is invertible. Let $\lambda = \tilde{\beta}_\theta^T \theta$. Then

$$\left(\tilde{X}_\theta^T \tilde{X}_\theta + \theta\theta^T\right) \tilde{\beta}_\theta = \tilde{X}_\theta^T \tilde{X}_\theta \tilde{\beta}_\theta + \lambda\theta = \tilde{X}_\theta^T \tilde{Y}_\theta,$$

or

$$\tilde{X}_\theta^T \left(\tilde{X}_\theta \tilde{\beta}_\theta - \tilde{Y}_\theta\right) = \lambda\theta.$$

Multiplying both sides of the equation above by θ^T yields $\lambda = 0$ and $\tilde{X}_\theta^T \tilde{X}_\theta \tilde{\beta}_\theta = \tilde{X}_\theta^T \tilde{Y}_\theta$. Hence, $\tilde{\beta}_\theta$ is also a solution to (5.5), as desired. □

D.2 Proof of Theorem 5.2.2

We first introduce a lemma to be used in the proof of the asymptotic results, whose proof is almost the same as the proofs of Lemmas 1–3 in the Appendix of [Chen et al., 2015] and will be omitted here.

Lemma D.2.1. *Suppose that the regularity assumptions 1-7 are satisfied. Then we have*

$$\sup_{u,\theta} \left| \sum_{j=1}^n s_j(u|\theta) g\left(x_j^T \theta\right) - g(u) \right| = o_p(1), \tag{D.4}$$

$$\sup_{u,\theta} \left| \sum_{j=1}^n s_j(u|\theta) x_j - \rho_x(u) \right| = o_p(1), \tag{D.5}$$

$$\sup_{u,\theta} \left| \sum_{j=1}^n s_j(u|\theta) g'\left(x_j^T \theta\right) x_j - g'(u) \rho_x(u|\theta) \right| = o_p(1), \tag{D.6}$$

$$\sup_{u,\theta}\left|\sum_{j=1}^{n} s_j(u|\theta)e_j\right| = o_p(1). \tag{D.7}$$

We now show some results to simplify the estimating equations.

Lemma D.2.2. *Suppose that the regularity assumptions 1-7 are satisfied. Then we have*

$$\sup_{u,\theta}\left|\frac{\partial}{\partial u}\left(\sum_{j=1}^{n} s_j(u|\theta)g\left(x_j^T\theta\right) - g(u)\right)\right| = o_p(1), \tag{D.8}$$

$$\sup_{u,\theta}\left|\frac{\partial}{\partial \theta}\left(\sum_{j=1}^{n} s_j(u|\theta)g\left(x_j^T\theta\right) - g(u)\right)\right| = o_p(1). \tag{D.9}$$

Proof. Let

$$v_i = \frac{x_i^T\theta - u}{h}, \quad K_i = K\left(v_i\right), \quad K_i' = K'\left(v_i\right) \quad (i = 1, 2, ..., n),$$

$$g_i = g(x_i^T\theta), \quad g_i' = g'(x_i^T\theta), \quad h_i = g_i - g(u) \quad (i = 1, 2, ..., n).$$

We first show (D.8). Noting that $\sum_{j=1}^{n} s_j(u|\theta) = 1$ for all u and θ, we have

$$\begin{aligned}
&\sum_{j=1}^{n} s_j(u|\theta)g\left(x_j^T\theta\right) - g(u) = \sum_{j=1}^{n} s_j(u|\theta)\left(g\left(x_j^T\theta\right) - g(u)\right) \\
&= \frac{\left(\frac{1}{nh}\sum K_i v_i^2\right)\left(\frac{1}{nh}\sum K_j h_j\right) - \left(\frac{1}{nh}\sum K_i v_i\right)\left(\frac{1}{nh}\sum K_j v_j h_j\right)}{\left(\frac{1}{nh}\sum K_i v_i^2\right)\left(\frac{1}{nh}\sum K_j\right) - \left(\frac{1}{nh}\sum K_i v_i\right)^2} \\
&= \frac{\Delta_1 - \Delta_2}{\Delta_3 - \Delta_4}.
\end{aligned}$$

Let $\Delta_j' = \partial\Delta_j/\partial u$. Then

$$\frac{\partial}{\partial u}\left(\sum_{j=1}^{n} s_j(u|\theta)g\left(x_j^T\theta\right) - g(u)\right) = \frac{\left(\Delta_1' - \Delta_2'\right)\left(\Delta_3 - \Delta_4\right) - \left(\Delta_1 - \Delta_2\right)\left(\Delta_3' - \Delta_4'\right)}{\left(\Delta_3 - \Delta_4\right)^2}.$$

Denote

$$\mu_1 = \int K(v)v dv, \quad \mu_2 = \int K(v)v^2 dv.$$

Since

$$E\left(\Delta_3 - \Delta_4\right) \to f_\theta^2(u)(\mu_2 - \mu_1^2) > 0$$

for all u and θ, by the regularity conditions of $f_\theta(u)$, we have

$$\sup_{u,\theta} |\Delta_3 - \Delta_4| = O_p(1), \quad \sup_{u,\theta} |\Delta_3 - \Delta_4|^{-1} = O_p(1).$$

Therefore, to show (D.8), it suffices to show

$$\sup_{u,\theta} \left|\left(\Delta_1' - \Delta_2'\right)\left(\Delta_3 - \Delta_4\right) - \left(\Delta_1 - \Delta_2\right)\left(\Delta_3' - \Delta_4'\right)\right| = o_p(1). \tag{D.10}$$

Since $K(\cdot)$ is symmetric, we have

$$\begin{aligned}
&E\left(\left(\frac{1}{nh}\sum K_i v_i^2\right)'\right) = E\left(\frac{\partial}{\partial u}\left(\left(\frac{1}{nh}\sum K_i v_i^2\right)\right)\right) \\
&= -\frac{1}{h^2}\int K\left(\frac{v-u}{h}\right)\left(\frac{v-u}{h}\right)^2 f_\theta(v)dv - \frac{2}{h^2}\int K\left(\frac{v-u}{h}\right)\left(\frac{v-u}{h}\right) f_\theta(v)dv \\
&= -f_\theta'(u)\tau_3 - 2f_\theta'(u)\mu_2 + o(1) = f_\theta'(u)\mu_2 + O(h) = O(1).
\end{aligned}$$

Similarly, we also have

$$E\left(\frac{1}{nh}\sum K_i v_i^2\right) = f_\theta(u)\mu_2 + O(h) = O(1),$$

$$E\left(\frac{1}{nh}\sum K_j h_j\right) = \left(\frac{f_\theta(u)g''(u)}{2} + f_\theta(u)g'(u)\right)\mu_2 h^2 + O(h^3) = O(h^2),$$

$$E\left(\left(\frac{1}{nh}\sum K_j h_j\right)'\right) = \left(\frac{3f_\theta'(u)g''(u)}{2} + \frac{f_\theta(u)g'''(u)}{2} + f_\theta''(u)g'(u)\right)\mu_2 h^2 + O(h^3).$$

The variances of the summations above are of order $O((nh^3)^{-1}) = o(1)$. Therefore,

$$|\Delta_1'| = \left|\left(\frac{1}{nh}\sum K_i v_i^2\right)\left(\frac{1}{nh}\sum K_j h_j\right)' + \left(\frac{1}{nh}\sum K_j h_j\right)\left(\frac{1}{nh}\sum K_i v_i^2\right)'\right| = O(h^2).$$

Since $E(K_i v_i) = o(h)$ and $E(K_i v_i h_i) = o(h)$, it can be readily seen that $|\Delta_2'| = o_p(1)$. By (D.4),

we have

$$|\Delta_1 - \Delta_2| \leqslant \sup_{u,\theta} \left| \sum_{j=1}^{n} s_j(u|\theta) g\left(x_j^T \theta\right) - g(u) \right| \cdot \sup_{u,\theta} |\Delta_3 - \Delta_4| = o_p(1),$$

which implies that (D.10) holds, and thus (D.8) holds. Similarly, for $\theta = (\theta_1, \theta_2, ..., \theta_p)^T$, we can show that, for $j = 1, 2, ..., p$,

$$\sup_{u,\theta} \left| \frac{\partial}{\partial \theta_j} \left(\sum_{j=1}^{n} s_j(u|\theta) g\left(x_j^T \theta\right) - g(u) \right) \right| = o_p(1),$$

which implies that (D.9) holds, completing the proof.

□

Let

$$\Lambda_i(\theta) = \left(\left(x_i - \rho_x(x_i^T\theta)\right)^T, \left\{ g'(x_i^T\theta) \left(x_i - \rho_x(x_i^T\theta)\right) \right\}^T \right)^T \quad (i = 1, 2, ..., n). \tag{D.11}$$

Lemma D.2.3. *Suppose that the regularity assumptions 1-7 are satisfied. Then we have*

$$\begin{aligned} \sum_{i=1}^{n} \Lambda_i(\hat{\theta}_1) \left(y_i - x_i^T \hat{\beta} - \hat{g}(x_i^T \hat{\theta}_1) \right) = \left\{ \sum_{i=1}^{n} \left(y_i - x_i^T \hat{\beta} - \hat{g}(x_i^T \hat{\theta}_1) \right) \right\} \cdot \\ O_p\left(\left\| \theta - \hat{\theta}_1 \right\| + \left\| \beta - \hat{\beta} \right\| \right). \end{aligned} \tag{D.12}$$

Proof. We take the derivatives of (5.4) with respect to β and θ and obtain

$$\sum_{i=1}^{n} \left(y_i - x_i^T \hat{\beta} - \hat{g}(x_i^T \hat{\theta}_1) \right) \left(\sum_{j=1}^{n} s_j(x_i^T \hat{\theta}_1 | \hat{\theta}_1) x_j - x_i \right) = 0, \tag{D.13}$$

$$\sum_{i=1}^{n} \left(y_i - x_i^T \hat{\beta} - \hat{g}(x_i^T \hat{\theta}_1) \right) \left(\frac{\partial}{\partial \theta} \hat{g}(x_i^T \theta) \Big|_{\theta = \hat{\theta}_1} \right) = 0. \tag{D.14}$$

Firstly, by combining (D.5) and (D.13), we have

$$\sum_{i=1}^{n}\left(y_i - x_i^T\hat{\beta} - \hat{g}(x_i^T\hat{\theta}_1)\right)\left(x_i - \rho_x(x_i^T\hat{\theta}_1)\right) = \sum_{i=1}^{n}\left(y_i - x_i^T\hat{\beta} - \hat{g}(x_i^T\hat{\theta}_1)\right)\cdot o_p(1). \tag{D.15}$$

Since

$$\begin{aligned}\frac{\partial}{\partial\theta}\hat{g}(x_i^T\theta) = &\left\{\sum_{j=1}^{n}\left(\frac{\partial}{\partial u}s_j(u|\theta)\right)\left(y_j - x_j^T\hat{\beta}\right)\right\}x_i\Bigg|_{u=x_i^T\theta} \\ &+\sum_{j=1}^{n}\left(\frac{\partial}{\partial\theta}s_j(u|\theta)\right)\left(y_j - x_j^T\hat{\beta}\right)\Bigg|_{u=x_i^T\theta},\end{aligned}$$

$$y_j - x_j^T\hat{\beta} = x_j^T\left(\beta_0 - \hat{\beta}\right) + \left(g(x_j^T\theta_0) - g(x_j^T\hat{\theta}_1)\right) + e_j + g(x_j^T\hat{\theta}_1),$$

by (D.7) we have

$$\begin{aligned}\frac{\partial}{\partial\theta}\hat{g}(x_i^T\theta)\Bigg|_{\theta=\hat{\theta}_1} = &\left\{\sum_{j=1}^{n}\left(\frac{\partial}{\partial u}s_j(u|\theta)\right)g(x_j^T\theta)\right\}x_i\Bigg|_{u=x_i^T\hat{\theta}_1,\theta=\hat{\theta}_1} \\ &+\sum_{j=1}^{n}\left(\frac{\partial}{\partial\theta}s_j(u|\theta)\right)g(x_j^T\theta)\Bigg|_{u=x_i^T\hat{\theta}_1,\theta=\hat{\theta}_1} + O_p\left(\left\|\theta-\hat{\theta}_1\right\| + \left\|\beta-\hat{\beta}\right\|\right).\end{aligned} \tag{D.16}$$

Next, by (D.8) we have

$$\sum_{j=1}^{n}\left(\frac{\partial}{\partial u}s_j(u|\hat{\theta}_1)\right)g(x_j^T\hat{\theta}_1) = g'(u) + o_p(1). \tag{D.17}$$

Besides, by (D.9) and (D.6),

$$\begin{aligned}\sum_{j=1}^{n}\left(\frac{\partial}{\partial\theta}s_j(u|\hat{\theta}_1)\right)g(x_j^T\hat{\theta}_1) &= -\sum_{j=1}^{n}s_j(u|\hat{\theta}_1)\left(\frac{\partial}{\partial\theta}g(x_j^T\hat{\theta}_1)\right) + o_p(1) \\ &= -\sum_{j=1}^{n}s_j(u|\hat{\theta}_1)g'(x_j^T\hat{\theta}_1)x_j + o_p(1) \\ &= -g'(u)\rho_x(u) + o_p(1).\end{aligned} \tag{D.18}$$

By replacing u by $x_i^T\hat{\theta}_1$ in (D.17) and (D.18), and combining (D.16), we obtain

$$\frac{\partial}{\partial\theta}\hat{g}(x_i^T\theta)\bigg|_{\theta=\hat{\theta}_1} = g'(x_i^T\hat{\theta}_1)\left(x_i - \rho_x(x_i^T\hat{\theta}_1)\right) + O_p\left(\left\|\theta-\hat{\theta}_1\right\| + \left\|\beta-\hat{\beta}\right\|\right).$$

This equation, together with (D.14) and (D.15), implies (D.12), which completes the proof. □

Proof of Theorem 5.2.2. By Lemma D.2.3, the asymptotic normality of the estimators can be shown by following the proof of Theorem 1 of [Chen et al., 2015]. Therefore, the details are omitted here. By Theorem 1 of [Chen et al., 2015] and the Law of Large Numbers, the Moore-Penrose inverse of the asymptotic covariance matrix is

$$\Gamma = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Lambda_i\Lambda_i^T = E\left(\Lambda\Lambda^T\right).$$

□

D.3 Proof of Theorem 5.3.1

Recall that S and T are the sets of the subscripts of the nonzero elements of β_0 and θ_0, respectively.

Proof. This proof is similar to the proof of Theorem 5.2.2 in [Liang et al., 2010] and consists of three steps: Step 1 determines the order of the penalized estimators $(\hat{\beta}_{\lambda_1 S}^T, \hat{\theta}_{\lambda_2 T}^T)^T$; Step 2 demonstrates the sparsity of $(\hat{\beta}_{\lambda_1 S}^T, \hat{\theta}_{\lambda_2 T}^T)^T$; Step 3 shows that the estimators are asymptotically normally distributed.

Step 1. Let κ_n satisfy $\kappa_n \to 0$ and $\kappa_n(\lambda_1+\lambda_2)^{-1}\to\infty$, $u_1, u_2 \in \mathbb{R}^p$ and $||u_1|| = ||u_2|| = C$ for some constant $C > 0$. Denote $\beta = \beta_0 + \kappa_n u_1$ and $\theta = \theta_0 + \kappa_n u_2$ and

$$\begin{aligned} G_p(\beta,\theta) - G_p(\beta_0,\theta_0) =& \frac{1}{2}\left\{G(\beta,\theta) - G(\beta_0,\theta_0)\right\} \\ &+ \left(n\lambda_1\left\|\beta\right\|_1 + n\lambda_2\left\|\theta\right\|_1 - n\lambda_1\left\|\beta_0\right\|_1 + n\lambda_2\left\|\theta_0\right\|_1\right) \\ =& \frac{1}{2}I_{n,1} + I_{n,2}. \end{aligned}$$

With some algebra, we have

$$
\begin{aligned}
I_{n,1} &= \sum_{i=1}^{n} \left\{ y_i - x_i^T\beta - \hat{g}(x_i^T\theta|\beta,\theta) \right\}^2 - \sum_{i=1}^{n} \left\{ y_i - x_i^T\beta_0 - \hat{g}(x_i^T\theta_0|\beta_0,\theta_0) \right\}^2 \\
&= \sum_{i=1}^{n} \left\{ \hat{g}(x_i^T\theta|\beta,\theta) - \hat{g}(x_i^T\theta_0|\beta_0,\theta_0) + x_i^T\beta - x_i^T\beta_0 \right\} \\
&\qquad \times \left\{ x_i^T\beta + \hat{g}(x_i^T\theta|\beta,\theta) + x_i^T\beta_0 + \hat{g}(x_i^T\theta_0|\beta_0,\theta_0) - 2y_i \right\} \qquad \text{(D.19)} \\
&= \kappa_n^2 \sum_{i=1}^{n} \left\{ \left(u_1 + g'(x_i^T\theta_0)u_2 \right)^T \left(x_i - E(x_i|x_i^T\theta_0) \right) \right\}^2 \\
&\quad + \kappa_n \sum_{i=1}^{n} \left\{ \left(u_1 + g'(x_i^T\theta_0)u_2 \right)^T \left(x_i - E(x_i|x_i^T\theta_0) \right) \right\} e_i + o_p(n^{1/2}\kappa_n^3).
\end{aligned}
$$

In addition,

$$
\begin{aligned}
|I_{n,2}| &\leqslant n\lambda_1 \big| \|\beta\|_1 - \|\beta_0\|_1 \big| + n\lambda_2 \big| \|\theta\|_1 - \|\theta_0\|_1 \big| \\
&\leqslant n\lambda_1 \|\beta - \beta_0\|_1 + n\lambda_2 \|\theta - \theta_0\|_1 \\
&= n\kappa_n(\lambda_1 + \lambda_2)C.
\end{aligned}
$$

Since $n^{1/2}(\lambda_1 + \lambda_2) \to \infty$ and $\kappa_n(\lambda_1 + \lambda_2)^{-1} \to \infty$, we conclude that $I_{n,2} = o_p(n\kappa_n^2)$ and the second term on the right-hand side of (D.19) is $O_p(n^{-1/2}\kappa_n) = o_p(n\kappa_n^2)$. Thus, the first term on the right-hand side of (D.19) dominates the second term and $I_{n,2}$. For any $\delta > 0$, there exists C such that

$$
\text{pr}\left\{ \inf_{\mathcal{U}_C} G_p(\beta_0 + \kappa_n u_1, \theta_0 + \kappa_n u_2) > G_p(\beta_0, \theta_0) \right\} \geqslant 1 - \delta,
$$

where $\mathcal{U}_C = \{(u_1^T, u_2^T)^T : ||u_1|| = ||u_2|| = C\} \subset \mathbb{R}^{2p}$. Hence, the convergence rate of $(\hat{\beta}_{\lambda_1 S}^T, \hat{\theta}_{\lambda_2 T}^T)^T$ is $O_p(\kappa_n)$.

Step 2. Let $||\beta_{1S} - \beta_{0S}|| = O_p(n^{-1/2})$ and $||\theta_{1T} - \theta_{0T}|| = O_p(n^{-1/2})$. We then show that

$$
(\beta^{*T}, \theta^{*T})^T = \arg\min_{\mathcal{V}} G_p(\beta, \theta), \qquad \text{(D.20)}
$$

where

$$\mathcal{V} = \{(\beta^T, \theta^T)^T : \beta_S = \beta_{1S}, \theta_T = \theta_{1T}, ||\beta_{S^c}|| \leqslant \tilde{C}n^{-1/2}, ||\theta_{T^c}|| \leqslant \tilde{C}n^{-1/2}\}$$

for some constant $\tilde{C}$ and $(\beta^{*T}, \theta^{*T})^T \in \mathcal{V}$ with $\beta^*_{S^c} = 0$ and $\theta^*_{T^c} = 0$.

For $k \in S^c$, we have

$$\begin{aligned}\frac{1}{2}\frac{\partial G(\beta,\theta)}{\partial \beta_k} &= \frac{1}{2}\frac{\partial}{\partial \beta_k}\sum_{i=1}^{n}\left\{y_i - x_i^T\beta - \hat{g}(x_i^T\theta|\beta,\theta)\right\}^2 \\ &= O_p\left(n||\beta-\beta_0|| + n||\theta-\theta_0||\right).\end{aligned}$$

Since $||\beta - \beta_0|| = O_p(n^{-1/2})$, $||\theta - \theta_0|| = O_p(n^{-1/2})$, when $\beta_k \neq 0$ we obtain

$$\begin{aligned}\frac{\partial G_p(\beta,\theta)}{\partial \beta_k} &= \frac{1}{2}\frac{\partial G(\beta,\theta)}{\partial \beta_k} + \frac{\partial(n\lambda_1 \|\beta\|_1)}{\partial \beta_k} \\ &= O_p(n^{1/2}) + n\lambda_1 \text{sgn}(\beta_k) \\ &= n^{1/2}\left\{n^{1/2}\lambda_1 \text{sgn}(\beta_k) + O_p(1)\right\}.\end{aligned}$$

Since $n^{1/2}\lambda_1 \to \infty$, $\partial G_p(\beta,\theta)/\partial\beta_k$ has different signs for $\beta_k \in (-\tilde{C}n^{-1/2}, \tilde{C}n^{-1/2})$. Therefore, $\beta_k = 0$ minimizes G_p when n is sufficient large. Similarly, we can show that for $k \in T^c$ $\theta_k = 0$ minimizes G_p when n is sufficient large. This completes the proof of (D.20) (sparsity).

Step 3. From Step 2, we have $\hat{\beta}_{\lambda_1 S^c} = 0$ and $\hat{\theta}_{\lambda_2 T^c} = 0$. The asymptotic distribution of $(\hat{\beta}^T_{\lambda_1 S}, \hat{\theta}^T_{\lambda_2 T})^T$ can be obtained by following the proof of Theorem 5.2.2 and thus omitted.

□

D.4 Proof of Theorem 5.4.1

Proof. This proof follows the proof of Theorem 3.2 in [Fan and Huang, 2005] and the proof of Theorem 4 in [Liang et al., 2010]. Denote $\Lambda_i(\theta_0)$ defined in (D.11) by Λ_i. Let

$$M_n = \left(\sum_{i=1}^{n}\Lambda_i\Lambda_i^T\right)^+ W^T\left(W\left(\sum_{i=1}^{n}\Lambda_i\Lambda_i^T\right)^+ W^T\right)^{-1} W.$$

Let $\hat{\xi}_N$ and $\hat{\xi}_A$ denote the values of parameters at which the objective function Denote

$$\hat{\xi}_0 = \arg\min_{\xi \in \Omega_0} G(\xi), \quad \hat{\xi}_1 = \arg\min_{\xi \in \Omega_1} G(\xi). \tag{D.21}$$

With some algebra, the difference $G(H_0) - G(H_1)$ can be written as

$$\begin{aligned} G(H_0) - G(H_1) &= G(\hat{\xi}_0) - G(\hat{\xi}_1) \\ &= (\hat{\xi}_1 - \hat{\xi}_0)^T \left(\sum_{i=1}^{n} \Lambda_i \Lambda_i^T \right) (\hat{\xi}_1 - \hat{\xi}_0) + o_p(1). \end{aligned}$$

Since $\hat{\xi}_0 = \hat{\xi}_1 - M_n \hat{\xi}_1 + o_p(1)$, we have

$$\begin{aligned} G(H_0) - G(H_1) &= \left(M_n \hat{\xi}_1 \right)^T \left(\sum_{i=1}^{n} \Lambda_i \Lambda_i^T \right) \left(M_n \hat{\xi}_1 \right) + o_p(1) \\ &= \left(n^{\frac{1}{2}} W \hat{\xi}_1 \right)^T \left(W \left(n^{-1} \sum_{i=1}^{n} \Lambda_i \Lambda_i^T \right)^+ W^T \right)^{-1} \left(n^{\frac{1}{2}} W \hat{\xi}_1 \right) + o_p(1). \end{aligned} \tag{D.22}$$

(a) Under H_0, $W\xi = 0$, and thus by Theorem 5.2.2 we have

$$n^{\frac{1}{2}} W \hat{\xi}_1 \xrightarrow{d} N\left(0, \sigma^2 W \Gamma^+ W^T\right).$$

This, together with $n^{-1} \sum_{i=1}^{n} \Lambda_i \Lambda_i^T \to \Gamma$ almost surely and Equation (D.22), yields $V \to \chi^2_m$ in distribution.

(b) Under H_1, $W\xi \neq 0$, and thus $n^{\frac{1}{2}} W \hat{\xi}_1 \sim n^{1/2}$. Then from (D.22), $G(H_0) - G(H_1) \sim n$. Since $G(H_1)/n \to \sigma^2$, $V \to \infty$ as $n \to \infty$. Therefore, the power function goes to 1 as $n \to \infty$ and the test is consistent.

(c) Under the local alternative of $n^{1/2} W\xi \to d \neq 0$, by Theorem 5.2.2 we have

$$n^{\frac{1}{2}} W \hat{\xi}_1 \xrightarrow{d} N\left(d, \sigma^2 W \Gamma^+ W^T\right).$$

This, together with $n^{-1} \sum_{i=1}^{n} \Lambda_i \Lambda_i^T \to \Gamma$ almost surely and Equation (D.22), implies $V \to \chi^2_m(\psi)$

in distribution, where the noncentrality parameter is

$$\psi = \sigma^{-2} d^T (W \Gamma^+ W^T)^{-1} d.$$

□