

Rainfall Simulation from an Inverse Problems Perspective

J. Piantadosi ^a and R. S. Anderssen ^b

^a*School of Information Technology and Mathematical Sciences, University of South Australia, SA 5095*
^b*CSIRO Digital Productivity & Services, GPO Box 664, Canberra, ACT 2601*
Email: Bob.Anderssen@csiro.au

Abstract: The recovery of information from indirect measurements is the paramount goal that underlies the solution of inverse problems. The indirect measurements, both practical and theoretical, come in a wide variety of disparate forms from histology, tomography, pathology, preventive medicine, geophysical exploration, rheology, financial management, etc. Often the information recovered is used to answer quite specific practical questions such as "Are cancer cells present?"; "Is the bone density normal?"; "Are minerals present in the surveyed area?"; "What is the risk associated with investing in this type of option?"; "What is the molecular structure of this type of polymer?"; "How does the DNA chromatin folded in response to a drought event?".

In many situations involving experimentally determined indirect measurements, the recovered information is the answer, because the experimental protocol being utilized has been designed to simply test for the presence or absence of some specific feature. Such situations arise in pathology where body fluids are used to test for the presence of infections or cancer.

In other situations, such as in stereology and in rheology, the answer is obtained directly from the indirect measurements without the need to formulate and solve a model relating the required information to the indirect measurement. Such "*direct use of the indirect measurements*" is based on the use of mathematical results that have established an equivalencing of the information to be recovered with some specific formula defined on the indirect measurements. A popular strategy for performing the equivalencing is the "*linear functional strategy*".

The underlying "*modus operandi*" is modelling, in that some model which relates the indirect measurements to the information to be recovered has been formulated and solved using some appropriate form of regularization.

Equally important, yet not so common, is the recovery of information for the construction of a simulator, which will be used to test various scenarios. A prototypical example considered here is the construction of a synthetic rainfall simulator to test urban drainage design or to assess flood mitigation measures.

Here, the indirect measurements are the historic records of rainfall at various locations $\{\ell_k\}$ ($k = 1, 2, \dots, K$) in the region Ω of interest, such as a new suburb being designed or the catchment of a river upstream from a dam. The information to be recovered is the structure of the simulator. A popular design for such a simulator is a multivariate (cumulative) probability distribution $J(u_1, u_2, \dots, u_d)$ which is a function of the rainfall intensities $\{r_i\}$ ($i = 1, 2, \dots, d$) for the d (successive) months of interest. The corresponding indirect measurements are the d monthly rainfall intensities $\{r_i\}$ modelled as the marginal distribution functions $u_i = J_i(r_i)$ of $J(u_1, u_2, \dots, u_d)$.

Without additional assumptions, there is a continuum of possibilities for $J(u_1, u_2, \dots, u_d)$ with respect to the given marginals $\{u_i\}$. When it is appropriate to conclude that the marginal are independent, it follows automatically that

$$J(u_1, u_2, \dots, u_d) = \prod_{i=1}^d u_i.$$

However, this often represents a gross simplification of the situation since it is known that, in the rainfall situation, successive monthly rainfalls tend to be correlated. In order to take the nature of such correlation into account in the recovery of a suitable approximation for $J(u_1, u_2, \dots, u_d)$, a popular strategy is to use an appropriate *copula* to perform a regularized recovery.

Other questions such as spatial correlation and modelling rainfall intensities at shorter timescales is not included in this paper. In this paper, the focus is on the various *inverse subproblems* that must be solved in implementing a copula recovery strategy.

Keywords: Rainfall simulation, copulas, inverse problems

1 INTRODUCTION

For each of the given K locations $\{\ell_k\}$ ($k = 1, 2, \dots, K$) in the region of interest Ω , in order to perform a simulation to test the consequence of representative rainfall events with respect to some scenario under investigation, the first step is the determination of how realizations of the rainfall events are to be modelled and simulated. This specifically involves the choice of a stochastic model, with parameterization θ , that will be sampled to generate the realizations. A popular choice for a stochastic model is an appropriate joint cumulative probability distribution $J(u_1, u_2, \dots, u_d; \theta)$ with the available rainfall intensity data $\{r_i\}$ at the $\{\ell_k\}$ modelled as its uniform marginal monthly rainfall intensity distributions $u_i = J_i(r_i)$.

The next step then becomes the determination of the parameterization α that allows the joint distribution model to be matched to the marginals $\{u_i\}$.

For the solution of this inverse problem, one starts with the recorded monthly rainfall intensities $\{r_i\}$ at the locations $\{\ell_k\}$ and solves a sequence of inverse subproblems which includes the choice of a copula model for the joint distribution and the determination of an estimate for the parameterization θ for that model. Copulas are functions that join or couple multivariate distribution functions to their one-dimensional marginal distributions.

The specific details are discussed in Section 3.

Other approaches to the stochastic modelling of rainfall dynamics include that of Cox and Isham (1988) and Rodriguez-Iturbe et al. (1988) who proposed and analysed Poisson point process models. Finally we cite two papers which provide a more comprehensive review of weather generation models, Srikanthan and McMahon (2001) and Wilks and Wilby (1999).

2 MODELLING THE JOINT DISTRIBUTION AS A COPULA

The formulation of the concept of a *copula* dates from the seminal paper of Sklar (1959), where the concept is initially defined and a uniqueness theorem is proved that, for the recovery of a joint (cumulative) distribution from its (continuous) uniform marginal distribution functions, there is a unique copula that does this.

However, it is a purely theoretical non-constructive result which gives no insight about this copula other than its existence and uniqueness. Nevertheless, if the distribution $J(u_1, u_2, \dots, u_d; \theta)$ and uniform marginals $u_i = J_i(r_i)$ are known, then the corresponding unique copula is given by

$$C(u_1, u_2, \dots, u_d; \theta) = J(J_1^{-1}(u_1), J_2^{-1}(u_2), \dots, J_d^{-1}(u_d); \theta),$$

where the uniform distribution property of the individual marginals has been explicitly exploited.

When only the marginals are given, the strategy reduces to choosing, on the basis of the problem context, an appropriate structure for the copula that satisfies Sklar's definition, or some equivalent forms of it, and to then use this resulting choice. Depending on the nature of the choice, it may be necessary to invoke some appropriate regularity to guarantee uniqueness. On the basis of Sklar's theorem, this then becomes the unique copula with respect to the given uniform marginals and the joint distribution that this choice of copula generates.

A detailed discussion of the choice of appropriate copulas for rainfall modelling and simulation can be found in Piantadosi et al. (2007a,b, 2009, 2011, 2012) and Borwein et al. (2013, 2014). In particular, their principal focus is maximum entropy checkerboard copulas. By seeking to maximise entropy, they add as little information to the problem as possible. They find a multi-dimensional checkerboard copula of maximum entropy that matches an observed set of grade correlations coefficients. This problem is formulated as a maximization of a concave function on a convex polytope. Under mild constraint qualifications they show that a unique solution exists in the core of the feasible region. By reformulating the problem as an unconstrained concave optimization problem using the theory of Fenchel duality they show that solution of the dual problem and subsequent recovery of the primal solution is a much more tractable procedure. The underlying entropy model assures one that the dual problem has many attractive features both theoretically and numerically.

Useful background information about copulas can be found in Frees and Valdez (1998), Kulpa (1999), Nelsen (1999) and Mikosch (2006).

3 THE METHODOLOGY

As already mentioned, the actual construction of a synthetic rainfall simulator, starting with the available monthly rainfall intensity data $\{r_i\}$ at the locations $\{\ell_k\}$ can be viewed as solving the following sequence of

interconnected inverse subproblems.

Step 1. With respect to the chosen region Ω , identify the locations $\{\ell_k\}$ where rainfall data is required for the construction of the simulation required to study the consequential effect of various rainfall patterns consistent with historical records. That such a step can be considered as an *solving an inverse problem* activity relates to the fact that the essential character associated with *solving of an inverse problem* is the recovery of information in response to answering a specific question.

Here, the question is “*What are the key locations $\{\ell_k\}$ for which rainfall data is required?*”. For example, in a large river catchment, the choice of the locations should, where possible, relate to where different types of rainfall events occur.

Step 2. Collect together the available historical records for Ω .

Here, the question is “*What and where are data available for the chosen locations?*”.

Step 3. For a particular month M_1 , using the available data, form the histogram H_1 of intensities for that month.

Here, the question is “*What is the appropriate way to organize the available data with respect to the construction of the simulator?*”.

Step 4. The gamma distribution has been widely used to model strictly positive intensities at a single site Srikanthan and McMahan (2001). Fit a Gamma density distribution, with shape parameters α (shape) and β (rate),

$$\gamma(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x), \quad (1)$$

to the H_1 histogram to obtain $\gamma(x; \alpha_1, \beta_1)$.

This clearly fits the pattern of a more traditional *solving an inverse problem* activity.

- (a) Though of an elementary nature, it corresponds to *parameter identification*.
- (b) It can be solved by traditional maximum likelihood methods which, computationally, reduces to solving a system of non-linear equations.
- (c) From an inverse problems perspective, it represents one of those special situations where its underlying mathematical structure can be exploited to yield functionals (here, the mean and variance) that can be directly applied to the observational data. Because the functionals are the mean and variance of the observational data, it is referred to as “*moment matching*”.
- (d) Moment matching is based on the fact that the mean and variance of a Gamma distribution, in terms of the shape parameters α and β , are given by

$$\mu = E(\gamma) = \frac{\alpha}{\beta} \quad \sigma^2 = \text{Var}(\gamma - E(\gamma)) = \frac{\alpha}{\beta^2}. \quad (2)$$

Consequently, once estimates $\hat{\mu}$ and $\hat{\sigma}^2$ have been derived for μ and σ^2 , it is a simple matter to solve the equations (2) to derive the following estimates for α and β

$$\hat{\alpha} = \frac{\hat{\mu}^2}{\hat{\sigma}^2}, \quad \hat{\beta} = \frac{\hat{\mu}}{\hat{\sigma}^2}. \quad (3)$$

- (e) The cumulative distribution $\Gamma(x; \alpha, \beta)$ corresponding to the density distribution (1)

$$\Gamma(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)} \gamma(\alpha, \beta x), \quad \gamma(s, x) = \int_0^x t^{s-1} \exp(-t) dt. \quad (4)$$

Step 5. Repeat Step 4 for the months M_i , $i = 2, 3 \dots, I$, to obtain the corresponding sets of $\{H_i\}$ and $\{\gamma(\alpha_i, \beta_i)\}$.

Step 6. Estimate the grade correlations between the $\{H_i\}$ (Piantadosi et al. (2007b)).

Here, the question is “*What is the nature of the correlations between the $\{H_i\}$ as estimated as the correlations between the $\{\gamma(\alpha_i, \beta_i)\}$?*”.

Step 7. For a selection of the $\{M_i\}$, determine the joint distribution J which has the corresponding marginals $\{\gamma(\alpha_i, \beta_i)\}$ such that the grade correlations determined at Step 6 are taken into account.

This is where the key inverse problem related to the construction of the simulator, to which the other steps relate, arises.

Here, the question is “*How is the joint distribution J to be modelled and determined?*”.

(a) It is at this stage that the role and importance of *copulas*, in terms of the discussion of Section 2, arise, since, in general, it cannot be assumed that the marginals $\{\gamma(\alpha_i, \beta_i)\}$ are independent.

(b) Consequently, solving this inverse problem, reduces to performing the following steps:

i. Formulate a parametric model $\hat{J}(x, y; \theta)$ for the joint distribution $J(x, y)$.

ii. A popular strategy for constructing a parametric model is to use even grided partitions of unity to construct a checkerboard model. Any continuous copula can be uniformly approximated by a checkerboard copula.

iii. The corresponding parametric model is referred to as a checkerboard copula (Kulpa (1999)).

(c) Define how the parameters θ are to be identified. This represents a key inverse subproblem that must be solved as part of the overall process.

i. As for the solution of any inverse problem, there is a need to guarantee that the parameters θ are chosen such that the marginals of the copula closely approximate the observed marginals and that subsidiary constraints are satisfied.

ii. In Piantadosi et al. (2007b), this has been achieved by using a checker board copula.

4 COPULAS AND TOMOGRAPHY

Interestingly, from an inverse problem’s perspective, the recovery of a two-dimensional joint distribution from its two one-dimensional marginals corresponds, tomographically, to the recovery of information about a two-dimension density structure when only two orthogonal projections of that structure have been recorded.

In the tomography situation, it is necessary to invoke strong assumptions about the density structure before a full recovery of that structure is possible. For example, if it is assumed that the density structure is radially symmetric, then only one projection is required, since both projections will be the same, and the recovery of that structure reduces to performing a half differentiation of one of the projections, since the inverse problem corresponds to solving the Abel integral equation Anderssen (1976); Jakeman and Anderssen (1975).

As is well known theoretically, for the recovery of a general density structure, infinitely many projections are required. Consequently, a comparison of the recovery of a two-dimensional joint distribution from its marginals with the tomography situation yields a clear intuitive understanding about the special nature of the joint distribution inverse problem. In fact, through the use of an appropriate copula, Sklar’s theorem guarantees the existence of a unique solution. Its underlying improperly posedness relating to the number of parameters needed to be estimated to fully define the copula.

A complementary discussion about copulas and tomography, with an insightful graphical illustration, can be found in Pougaza et al. (2010).

REFERENCES

Anderssen, R. (1976). Stable procedures for the inversion of abel s equation. *IMA Journal of Applied Mathematics* 17, 329–342.

J. Piantadosi and R. S. Anderssen, Rainfall modelling

- Borwein, J., P. Howlett, and J. Piantadosi (2013). Modelling and simulation of seasonal rainfall. In Gaertner, P., Bowden, F., Piantadosi, J., Anderssen, R. S. and Mobbs, K. (Ed.), *22nd National Conference of the Australian Society for Operations Research (ASOR 2013). The Australian Society for Operations Research, Adelaide, December 2013*, pp. 300–306.
- Borwein, J., P. Howlett, and J. Piantadosi (2014). Modelling and Simulation of Seasonal Rainfall Using the Principle of Maximum Entropy. *Entropy* 16, 747–769.
- Cox, D. R. and V. Isham (1988). A simple spatial-temporal model of rainfall. *Proc. Roy. Soc. A.* 415, 317–328.
- Frees, E. W. and E. A. Valdez (1998). Understanding relationships using copulas. *North American actuarial journal* 2, 1–25.
- Jakeman, A. and R. Anderssen (1975). Abel type integral equations in stereology: I. general discussion. *Journal of Microscopy* 105, 121–133.
- Kulpa, T. (1999). On approximation of copulas. *International Journal of Mathematics and Mathematical Sciences* 22, 259–269.
- Mikosch, T. (2006). Copulas: Tales and facts. *Extremes* 9, 3–20.
- Nelsen, R. (1999). *An Introduction to Copulas, Lecture Notes in Statistics*. New York: Springer-Verlag.
- Piantadosi, J., J. Boland, and P. Howlett (2009). Generating Synthetic Rainfall on Various Timescales-Daily, Monthly and Yearly. *Environment Modelling & Assessment* 14, 431–438. 17th Biennial Conference on Modelling and Simulation, Christchurch, NEW ZEALAND, DEC 10-13, 2007.
- Piantadosi, J., I. Howlett, J. Borwein, and J. Henstridge (2011). Generation of simulated rainfall data at different time-scales. In Chan, F. and Marinova, D. and Anderssen, R. S. (Ed.), *19TH International Congress on Modelling and Simulation (MODSIM2011)*, pp. 2654–2660. 19th International Congress on Modelling and Simulation (MODSIM), Perth, AUSTRALIA, DEC 12-16, 2011.
- Piantadosi, J., P. Howlett, and J. Boland (2007a). Matching the grade correlation coefficient using a copula with maximum disorder. *J. Industrial Management Optimization* 3, 305–312. 5th Ballarat Workshop on Global Optimization and Non-Smooth Optimization, Univ Ballarat, Ballarat, AUSTRALIA, NOV 28-30, 2006.
- Piantadosi, J., P. Howlett, and J. Boland (2007b). Matching the grade correlation coefficient using a copula with maximum disorder. *Journal of Industrial and Management Optimization* 3, 305–312.
- Piantadosi, J., P. Howlett, and J. Borwein (2012). Copulas with maximum entropy. *Optimization Letters* 6, 99–125.
- Pougaza, D.-B., A. Mohammad-Djafari, and J.-F. Bercher (2010). Link between copula and tomography. *Pattern Recognition Letters* 31, 2258–2264.
- Rodriguez-Iturbe, I., D. R. Cox, and V. Isham (1988). A point process model for rainfall: further developments. *Proc. Roy. Soc. A.* 417(1853), 283–298.
- Sklar, A. (1959). Fonctions de repartition n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Universit de Paris* 8, 229–231.
- Srikanthan, R. and T. McMahon (2001). Stochastic generation of annual, monthly and daily climate data: A review. *Hydrology and Earth System Sciences* 5(4), 653–670.
- Wilks, D. and R. Wilby (1999). The weather generation game: A review of stochastic weather models. *Progress in Physical Geography* 23(4), 329–357.