A Stochastic Space-time Model for Intermittent Precipitation Occurrences

Ying Sun

Environmental Statistics Group

es.kaust.edu.sa

King Abdullah University of Science and Technology

EnvStat @KAUST

KAUST King Abdullah University of Science and Technology

Joint work with Michael Stein from the University of Chicago

Stochastic Weather Generators

- Stochastic weather generators:
  - Provide fast and realistic simulations of atmospheric variables.
  - Facilitate understanding the probabilistic structure.
  - Required by impact studies.
- Note:
  - Statistical models, not forecast models.
  - Reproduce reality accurately in a distributional sense.
  - Different from climate numerical models.
- Precipitation has been a key variable of interest.
  - Serves as input into hydrologic and agricultural models.

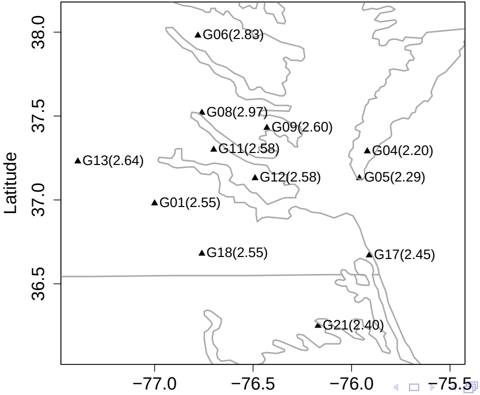
Features of Precipitation Field

- Intermittent nature and high variability.
- Spatio-temporal dependence.
- Precipitation statistics are strongly scale dependent.
  - Data are measured as averages over space-time scales.
- Rainfall zeros: precipitation occurrences.
  - Especially for small time scales.
  - Contribute significantly to the observed dependence.
  - An important component in stochastic weather simulations.
- Intensity: non-Gaussian.

Our Contributions

- Consistent model for precipitation occurrences:
  - Shorter time scale: 15-minute rain gauge data.
  - Generate realistic spatio-temporal dependence.
- Model framework:
  - Truncated spatio-temporal non-Gaussian random field model.
- Model diagnostics:
  - Define meaningful precipitation statistics.
  - Develop visualization methods for the assessment of spatio-temporal dependence.
- Model assessment is important.
  - Reproducing means of the precipitation.
  - Reproducing "higher moments": dependence.

Gauge Locations



Rain Gauge Data

- High-frequency gauge-based space-time data:
  - Time records of each bucket tip (0.01 inches) at each location.
  - 12 Gauges from 05/19/2004 to 05/17/2007.
  - Aggregate to different time scales.
  - Unit: mm/hr.
- Percentage of rain for different time scales:

Time	10-min	15-min	30-min	1-hr	3-hr
Occurrence	1.77	2.55	4.91	6.47	10.42

Time	6-hr	1-day	1-week	1-mon	3-mon
Occurrence	14.77	32.72	88.57	99.76	100

Truncated Model

- Bell (1987): the truncated Gaussian random field model

$$W(\mathbf{x}) = \begin{cases} f(Z(\mathbf{x})), & Z(\mathbf{x}) > c; \\ 0, & Z(\mathbf{x}) \leq c. \end{cases}$$

- Stein (1992) considered Monte Carlo methods for prediction and inference for truncated spatial data based on this model.
- For 15-minute rain rates:
  - The value of c is high.
  - GRF has thin tail distributions.
  - The probability of exceeding c at many locations is low.
  - It may not be sufficient to describe the dependence in the occurrence.

More Flexible tRF

- t random field model (Røislien and Omre, 2006) is specified by
  - Mean function: $\mu(\mathbf{x})$.
  - Positive definite scale function: $\kappa(\mathbf{x}, \mathbf{x}')$.
  - The degrees of freedom: ν.
- The tRF tends towards a GRF as $\nu \to \infty$.

Representation of Spatial tRF Model

- Univariate t random variable has the representation

$$T = \frac{Z}{\sqrt{V/\nu}}.$$

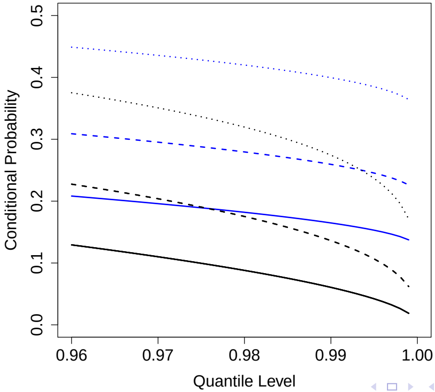
  - T has a heavier tail due to the random scaling $\sqrt{V/\nu}$.
- Similarly, the multivariate t random vector for n locations

$$\mathbf{Y} = \boldsymbol{\mu} + \frac{\mathbf{Z}}{U}, \quad \nu U^2 \sim \chi^2(\nu).$$

  - Given $U = u$, $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Omega}/u)$.
  - The variability across realizations of $\mathbf{Y}$ is larger.
  - A higher probability that $\mathbf{Y}$ exceeds c at many locations.
  - tRF tends towards GRF as $\nu \to \infty$.

Bivariate Illustration

$P(Y_2 > c | Y_1 > c)$: $\nu = 3, \infty$ and $\rho = 0.3, 0.5, 0.7$.



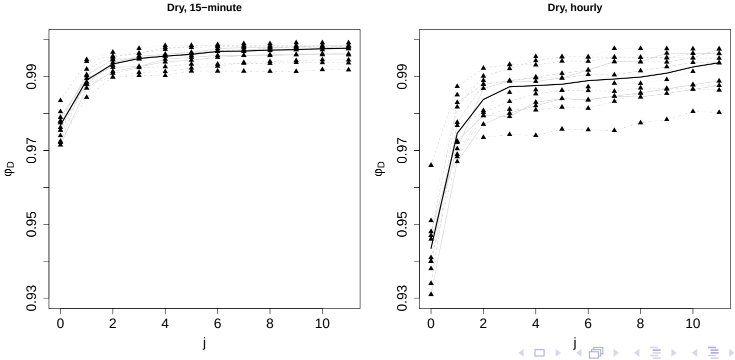
Precipitation Statistics

- Define the dry and the rain event:

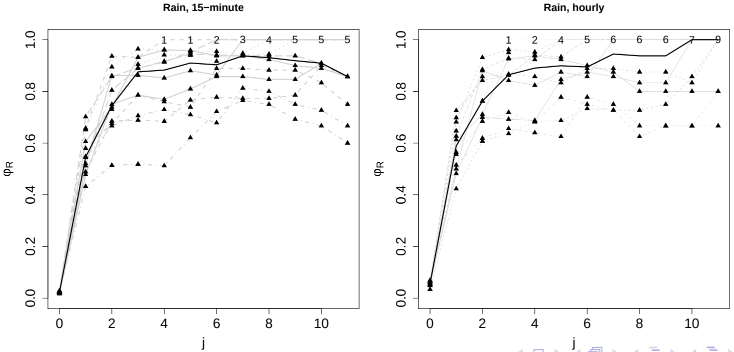
$$D(\mathbf{x}) = \{Y(\mathbf{x}) \leq c\}, \quad R(\mathbf{x}) = \{Y(\mathbf{x}) > c\}.$$

- Precipitation statistics:
  - Marginal dry: $p_D = Pr\big(D(\mathbf{x})\big)$.
  - Marginal rain: $p_R = Pr\big(R(\mathbf{x})\big) = 1 - p_D$.
  - Conditional dry: $p_{D|D} = Pr\big(D(\mathbf{x})|D(\mathbf{x}')\big)$.
  - Conditional rain: $p_{R|R} = Pr\big(R(\mathbf{x})|R(\mathbf{x}')\big)$.

Conditional Dry Probability Plot

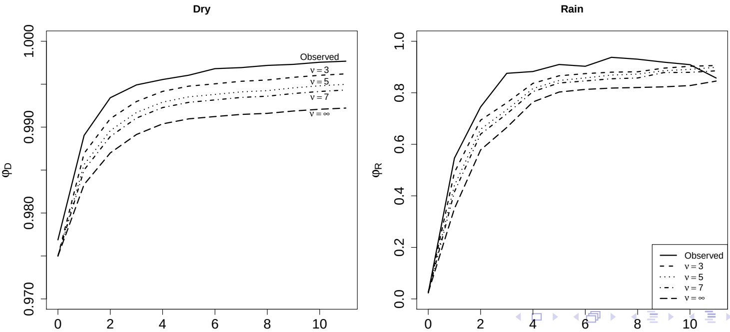


Conditional Rain Probability Plot



Comparing Model Simulations

- Purely spatial truncated tRF models with zero-mean.
- Degrees of freedom: $\nu = 3, 5, 7, \infty$.
- The marginal dry probability: 97.5%.



Space-time tRF Model

EnvStat

KAUST

- We propose a more flexible space-time tRF model

$$Y(\mathbf{x}, t) = \frac{Z(\mathbf{x}, t)}{U(t)},$$

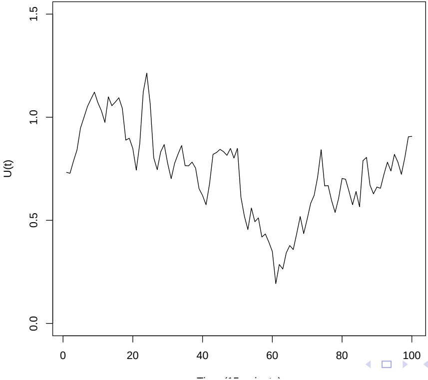
  - $Z(\mathbf{x}, t)$: a zero-mean stationary GP with ACF $K(\mathbf{x}, t)$.
  - $\nu U^2(t)$: a stationary Gamma process, where

$$U^2(t) = \frac{1}{\nu} \sum_{j=1}^{\nu} X_j^2(t),$$

    - $X_j(t)$'s are i.i.d. zero-mean stationary Gaussian processes.
- For any given time $t = t^*$,
  - $\nu U^2(t^*)$ is χ^2_ν distributed.
  - It follows that $Y(\mathbf{x}, t^*)$ is a spatial tRF.

Simulated $U(t)$ Process

Temporal range parameter $\alpha_u = 0.5$, $\nu = 3$.



Spatio-temporal Covariance Function

- Stein (2005): spectral-in-time representation of K for data taken regularly in time at a modest number of sites,

$$K(\mathbf{x}, t) = \int_{\mathbb{R}} S(\omega) C\Big(|\mathbf{x}|\gamma(\omega)\Big) e^{i\mathbf{u}'\mathbf{x}\theta(\omega)+i\omega t} d\omega.$$

- All the functions have natural interpretations.
- Parameterization for even positive functions:

$$\log\Big\{\gamma(\omega)\Big\} = \sum_{k=0}^{L} a_k \cos(k\omega),$$

$$\log\Big\{S(\omega)\Big\} = -\beta \log\left(\sin\left|\frac{1}{2}\omega\right|\right) + \sum_{k=0}^{L} c_k \cos(k\omega).$$

Parameters

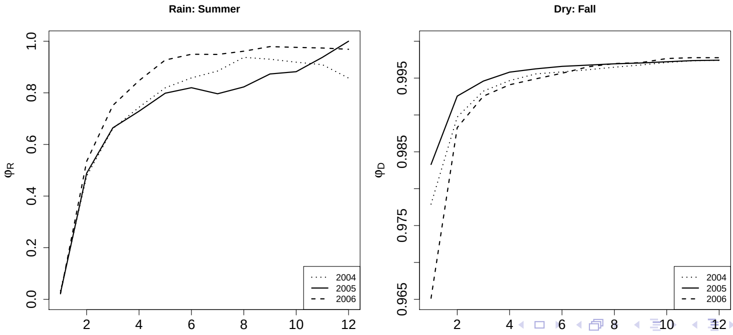
- Parameters of interest are:
  - α: the spatial range in C (Matérn).
  - β: measures the long-range time dependence in S.
  - α_u: dependence range in the $U(t)$ process.
  - ν: degrees of freedom.

Application to Rain Gauge Data

- Conditional probability plots for four seasons from Summer 2004 to Spring 2007.
- Look for patterns of the precipitation occurrences:
  - Precipitation over a larger region and/or last longer time.
  - Local storms.

El Niño Years?

- The years 2004 and 2006: weak El Niño years.
- The study region is not typically affected by the El Niño effect in terms of total precipitation; How about spatial patterns?



Model Fitting

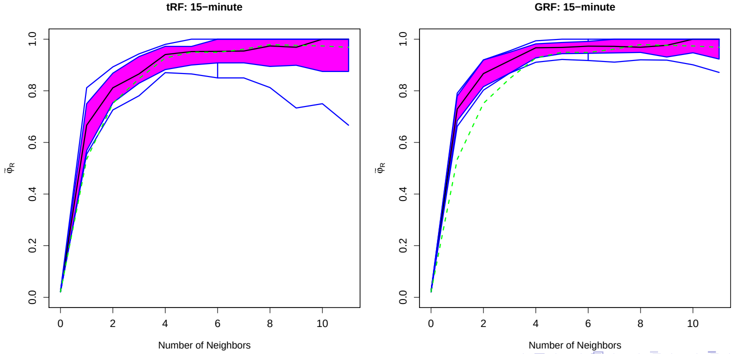
- Empirical approach: based on Monte Carlo simulations.
  - Minimizing the difference of the conditional probabilities obtained from the model and from the observations.
- The estimates of $(\alpha, \beta, \alpha_u, \nu)$ for Summer 2004-2006.

Year	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}_u$	$\hat{\nu}$
2004	0.485	0.486	0.199	4
2005	0.495	0.558	0.232	5
2006	0.500	0.652	0.175	3

- Summer 2006:
  - $(\hat{\alpha}, \hat{\beta}) = (0.811, 0.123)$ for the truncated GRF model.
  - The estimate of the spatial range parameter is much larger.

Model Diagnostics: Summer 2006

- Compare the conditional rain probabilities by the functional boxplot (Sun and Genton, 2011).

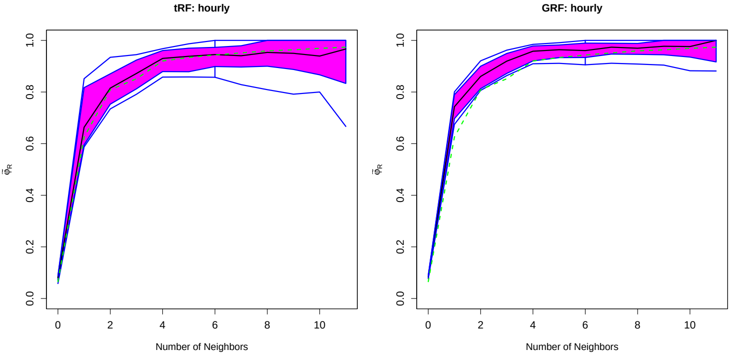


Consistent Spatio-temporal Model

- Another important aspect: wet or dry spell.
- Temporal dependence is necessary to produce them correctly.
- From a statistical modeling point of view
  - It is desirable to have a consistent space-time model to produce precipitation features at different scales.
  - Rather than to have a separate model for each scale.
- In order to produce the desired rain or dry spells at larger scales through aggregation
  - Need to characterize the complex dependence structure precisely at small space-time scales.
- We can use aggregation to assess model performance.

Model Diagnostics: Summer 2006

- Aggregate to hourly data



Computation

- Estimation:
  - Generate data at 12 locations and 8736 time points.
  - Circulant embedding techniques.
  - FFT for fast and exact simulations (Wood and Chan, 1994; Helgason et al., 2011).
- Simulation on a grid.

Discussion

- A stochastic space-time model for precipitation occurrence:
  - Truncated tRF model with random scaling $U(t)$.
  - The increased variability is useful for shorter time scale.
- Model assessment:
  - Conditional probability plot, and aggregation.
  - Functional boxplot: the fast algorithm (Sun et al., 2012).
- Precipitation intensity:
  - More complete analysis: includes the positive rainfall amounts.
  - Investigate Bayesian inference methods.