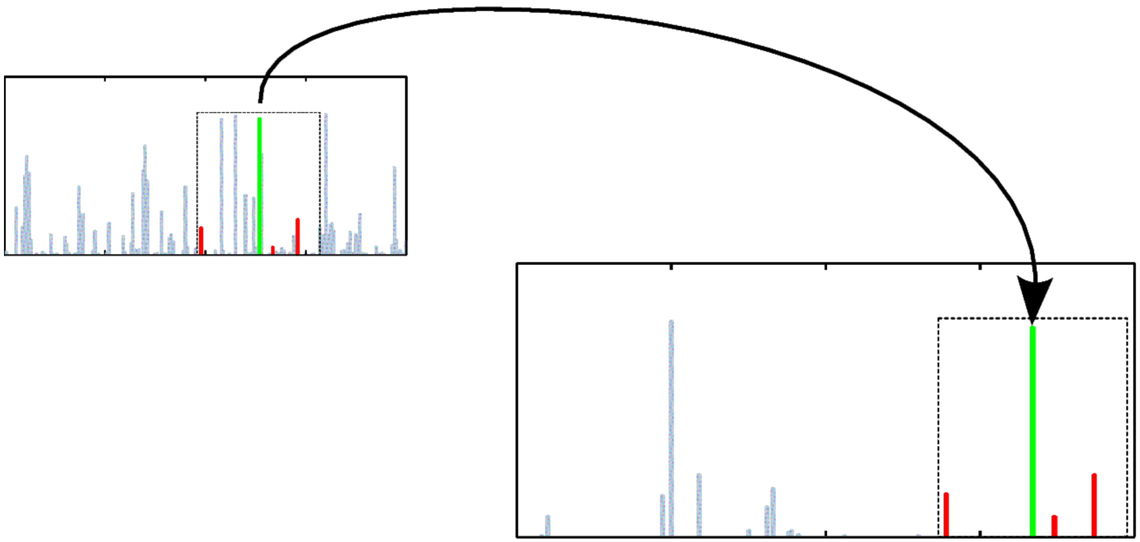
Daily rainfall simulation: reproducing high-order statistics with the Direct Sampling technique



F.Oriani[1], R.Mehrotra[2], G.Mariethoz[2], J.Straubhaar[1], A.Sharma[2], P.Renard[1]

1. Centre for Hydrogeology and Geothermics, University of Neuchâtel, Neuchâtel, Switzerland.
2. School of Civil and Environmental Engineering, University of New South Wales, Sydney, New South Wales, Australia.

FNSNF Swiss National Science Foundation

SWG2014, Avignon, September 2014

NATIONAL CENTRE FOR GROUNDWATER RESEARCH AND TRAINING

unine UNIVERSITÉ DE NEUCHÂTEL

Centre for Hydrogeology and Geothermics (CHYN)

UNSW THE UNIVERSITY OF NEW SOUTH WALES

Overview

Simulating rainfall

The Direct Sampling technique

Simulating the Australian rainfall

A comparison with the Markov-chain approach

Conclusions

Simulating rainfall

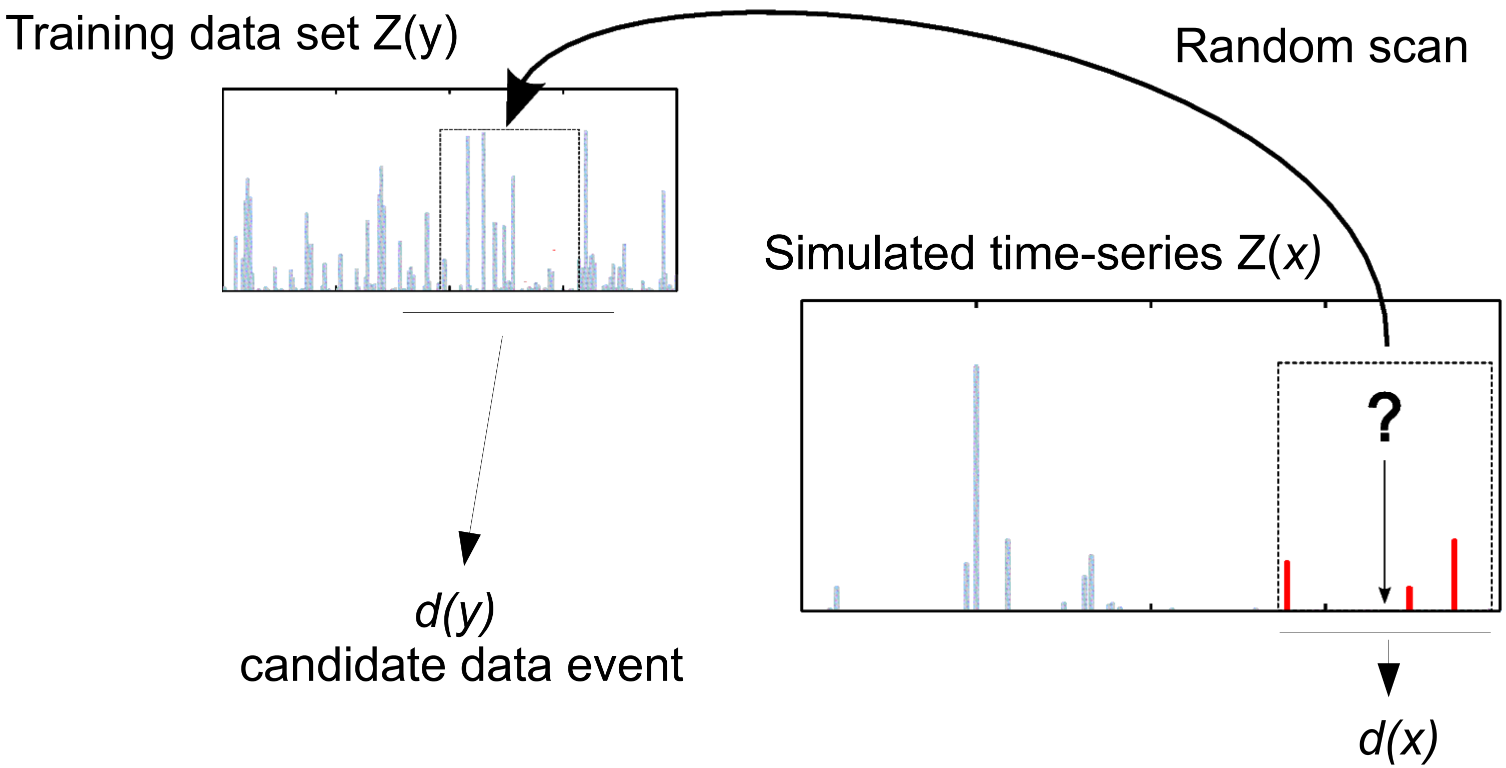
- A regular annual seasonality, inter-annual fluctuations but also a chaotic behavior at the daily scale [see e.g. Sivakumar1998].

- **The challenge**: to simulate synthetic time-series honoring the reference statistics and persistence from the daily to the higher temporal scale.

- **The problem of over-dispersion**: if the model is focused on the daily scale, extremes are poorly reproduced at higher scales (= reference is more dispersed than the simulation).

Direct Sampling (DS)

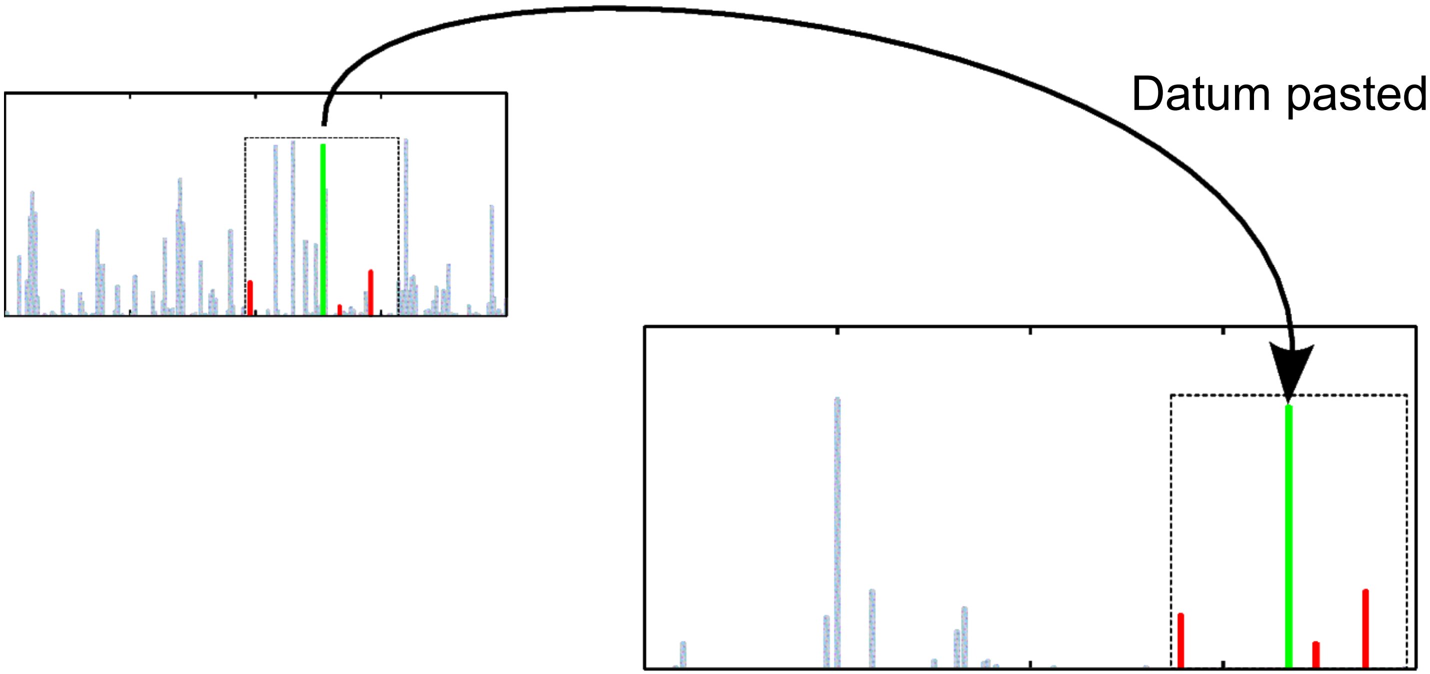
[Mariethoz 2010,...]



d(x) = *n* informed time steps closest to *x* inside the search window.

Direct Sampling (DS)

[Mariethoz 2010,...]



A sampling rule based on a distance measure (dissimilarity between patterns).

$$D(d(x),d(y))$$

If $D(d(x),d(y)) < T$, Z(y) is assigned to Z(x).

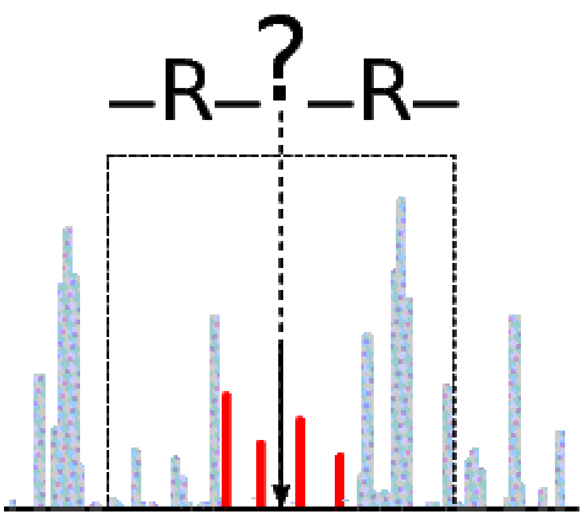
Distance categorical variables:

$$D(\vec{d}(x_t), \vec{d}(y_i)) = \frac{1}{n}\sum_{j=1}^{n} a_j, \qquad a_j = \begin{cases} 1 & if \quad Z(x_j) \neq Z(y_j) \\ 0 & if \quad Z(x_j) = Z(y_j) \end{cases}$$

For continuous variables:

$$n \leq N$$

$$D(\vec{d}(x_t), \vec{d}(y_i)) = \frac{1}{n}\sum_{j=1}^{n} |Z(x_j) - Z(y_j)|$$

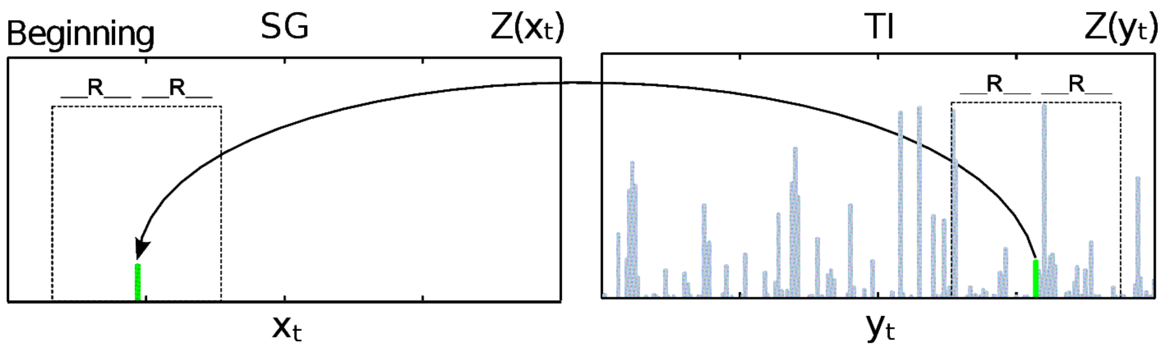


Main parameters:

N = max number of considered neighbors;
R = search neighborhood radius;
T = distance threshold;
F= max scanned TI fraction;

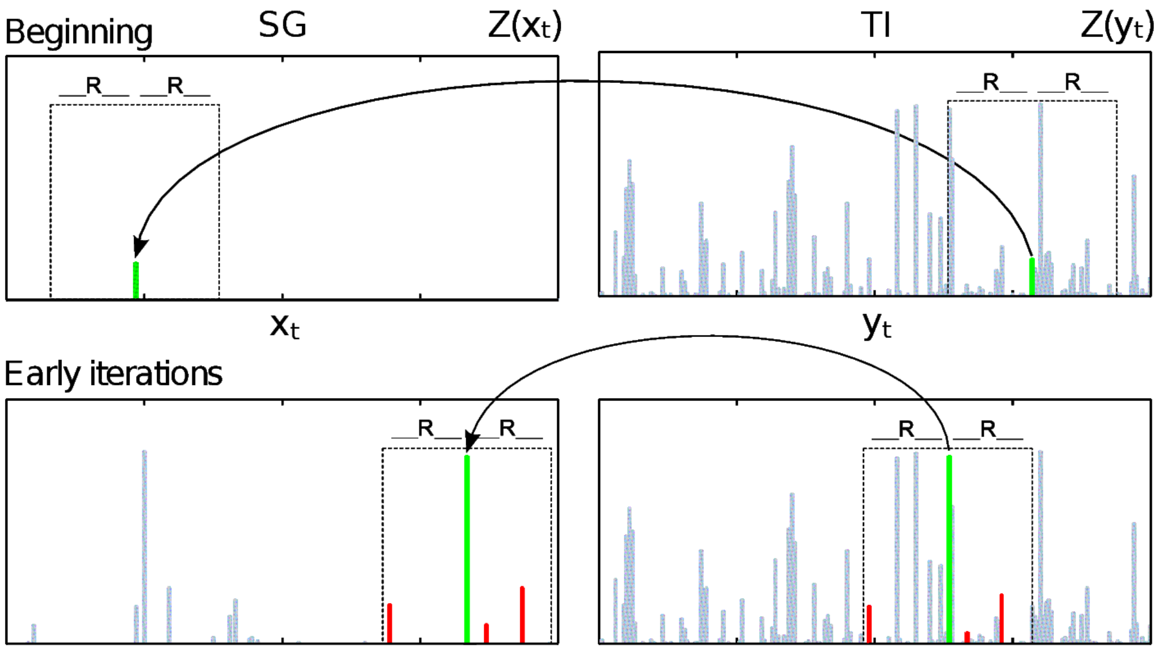
Direct Sampling (DS)

[Mariethoz 2010]



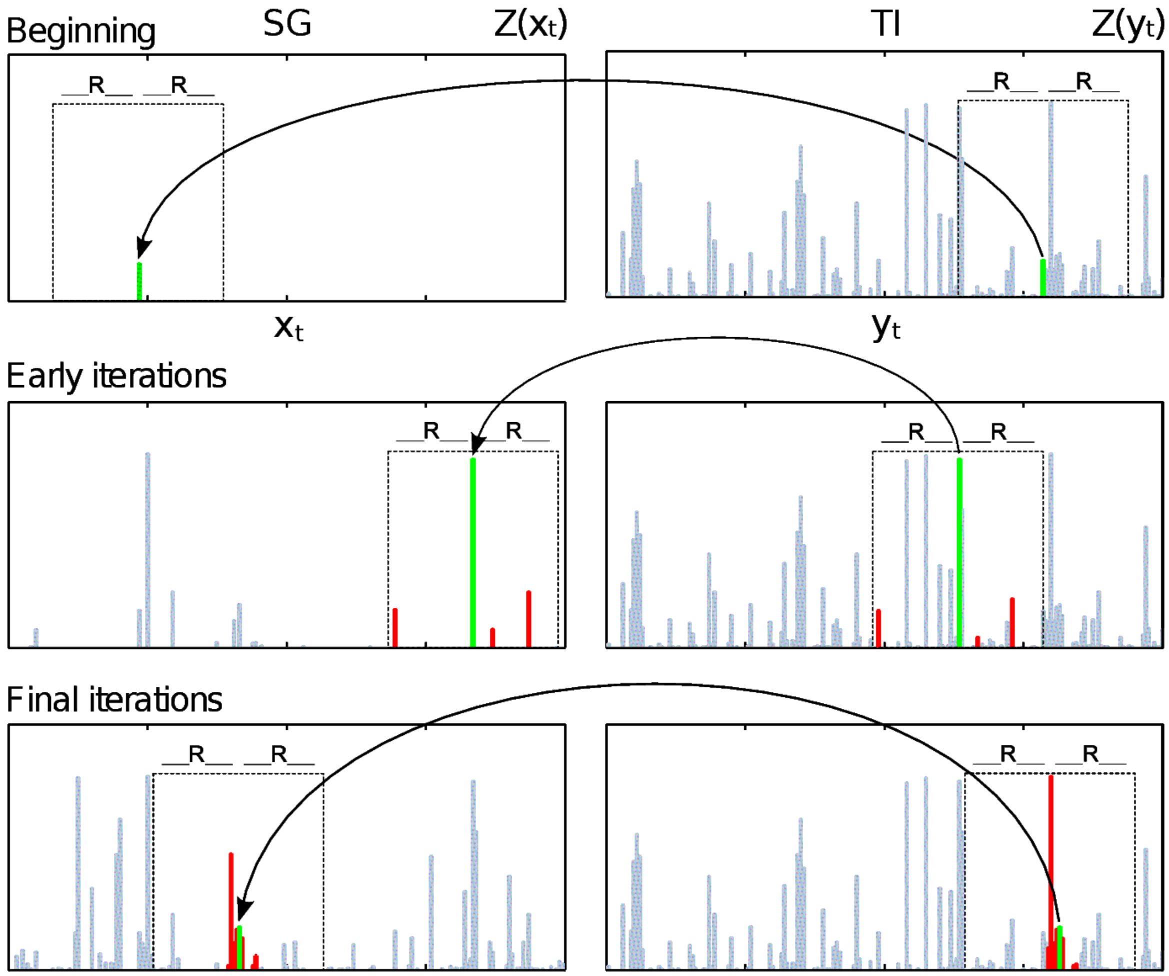
Direct Sampling (DS)

[Mariethoz 2010]



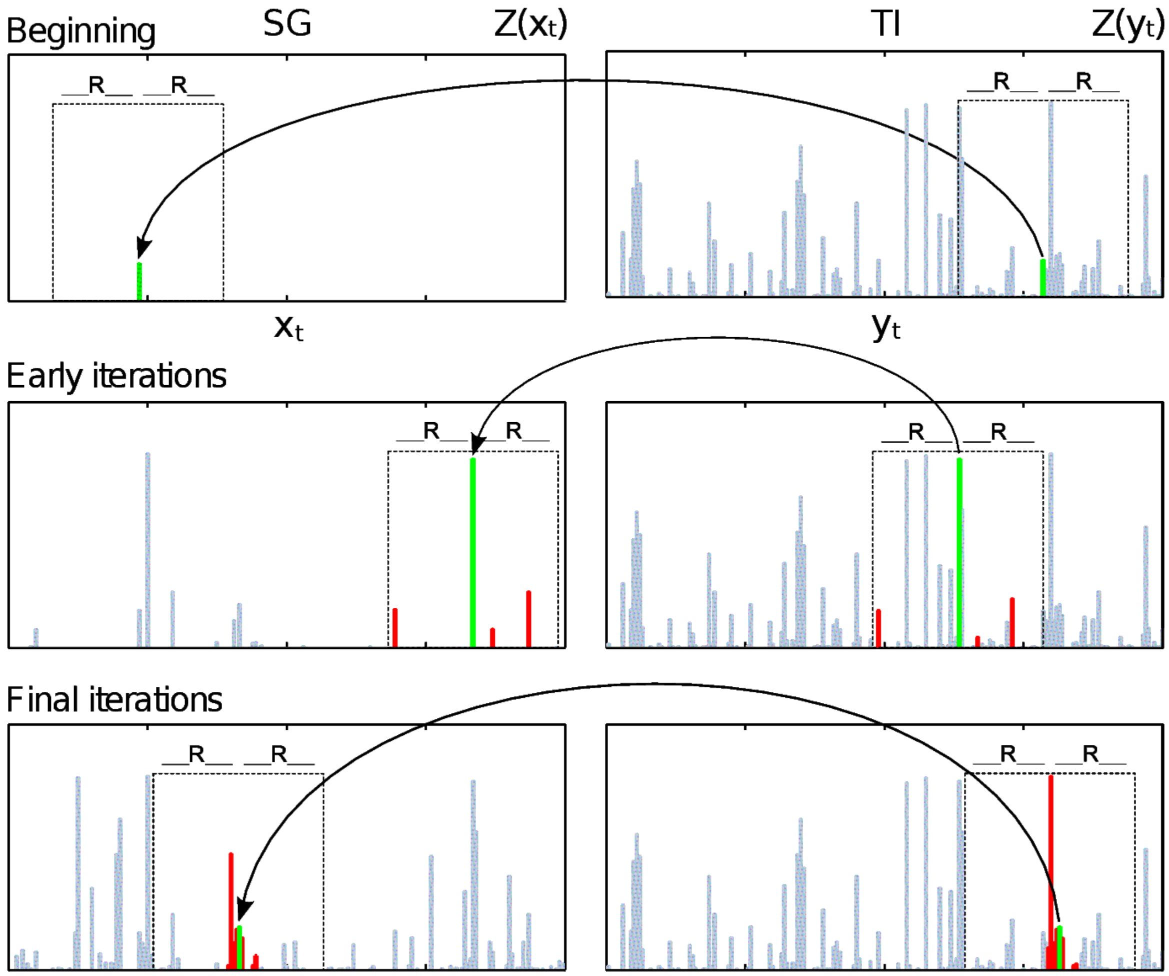
Direct Sampling (DS)

[Mariethoz 2010]



Direct Sampling (DS)

[Mariethoz 2010]



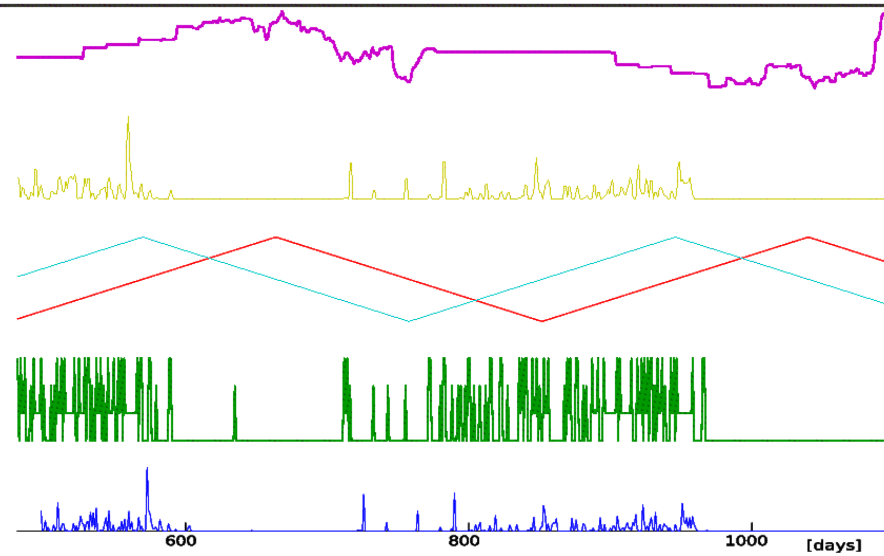
Multiple scale pattern reproduction

High-order statistics reproduction.

No need of a high-order prior!

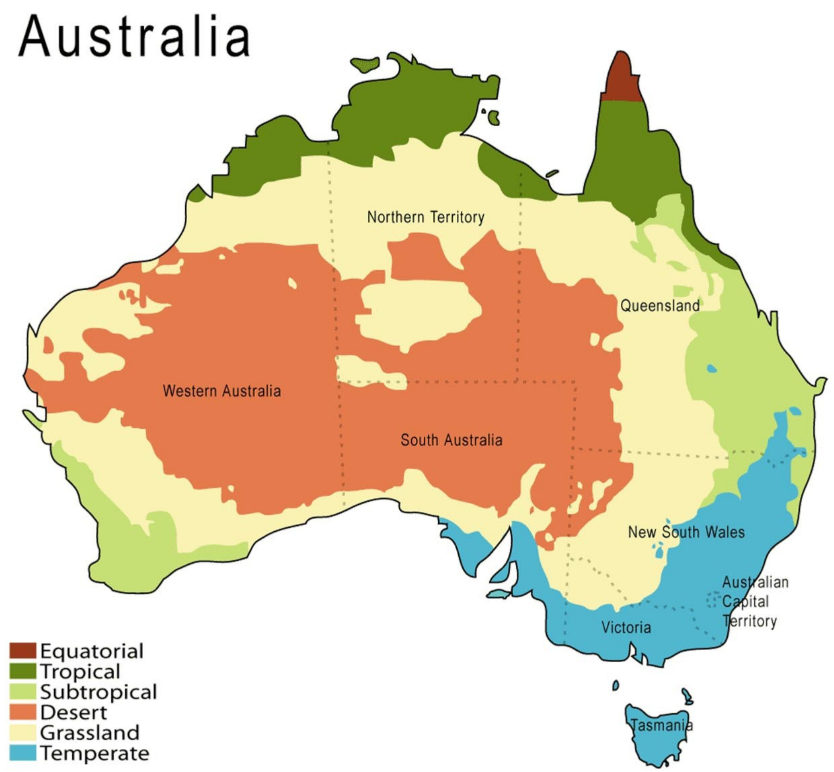
Standard DS setup for daily rainfall simulation

Variable	R	N	T
1) 365MA	5000	21	0.05
2) 2MS	1	1	0.05
3) tr1	1	1	0.05
4) tr2	1	1	0.05
5) dw	10	5	0.05
*) rainfall	5000	21	0.05



(Multivariate simulation)

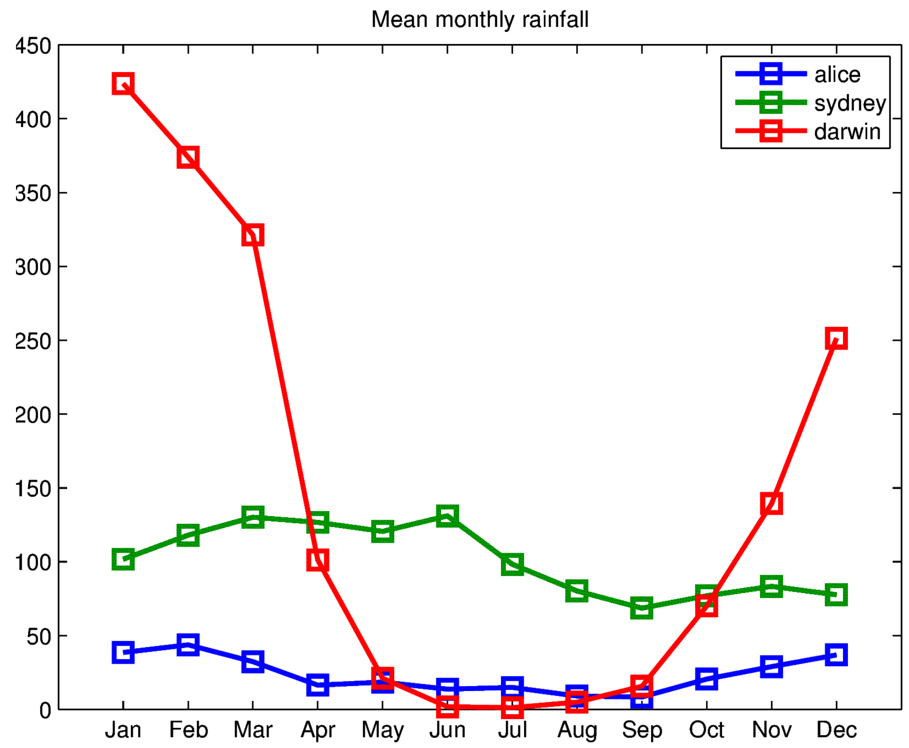
Some results: the Australian rainfall [Oriani et al.2014]



Sydney (1858-2013, temperate)

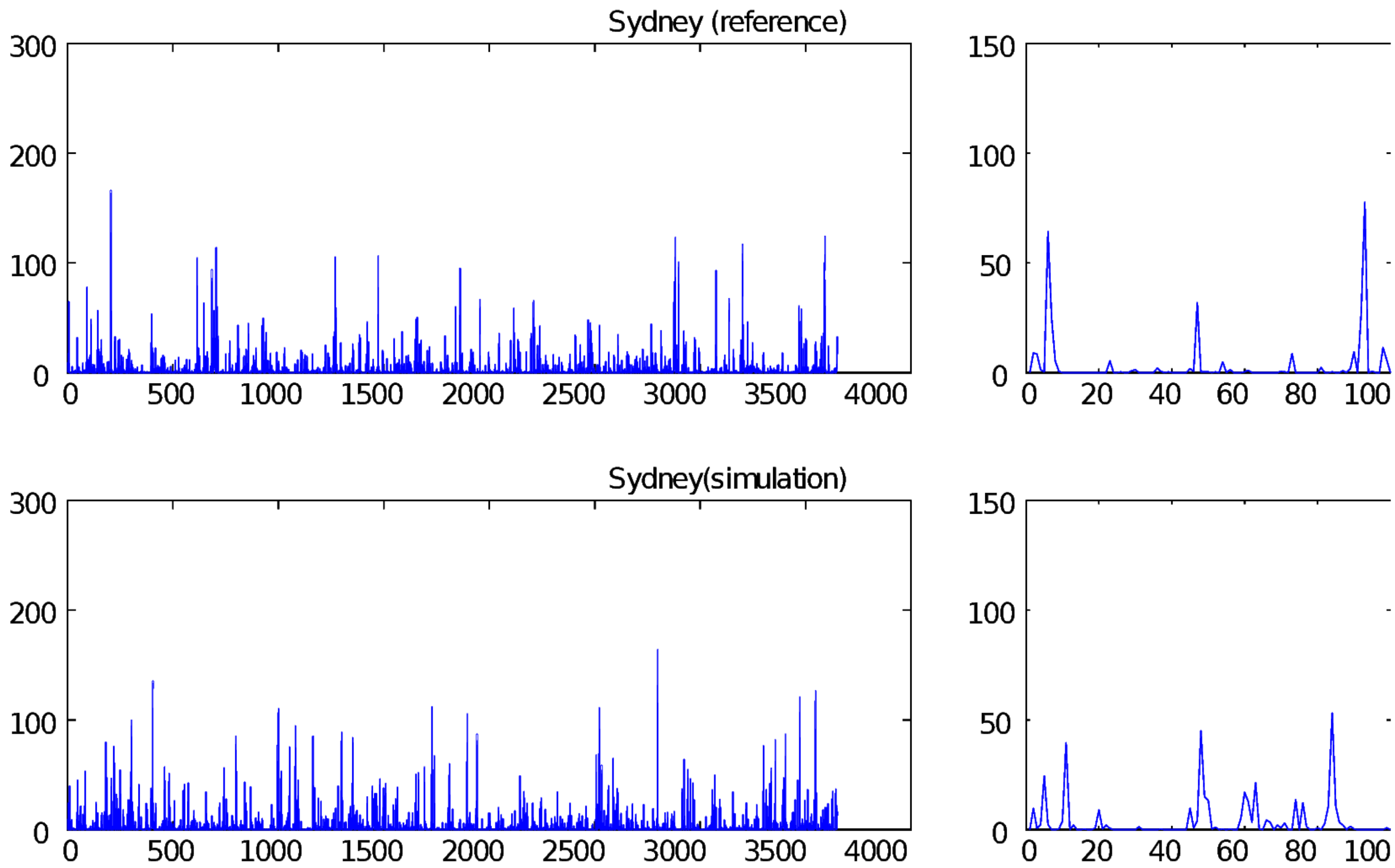
Alice Springs (1941-2013, hot desert)

Darwin (1941-2013, tropical savannah)

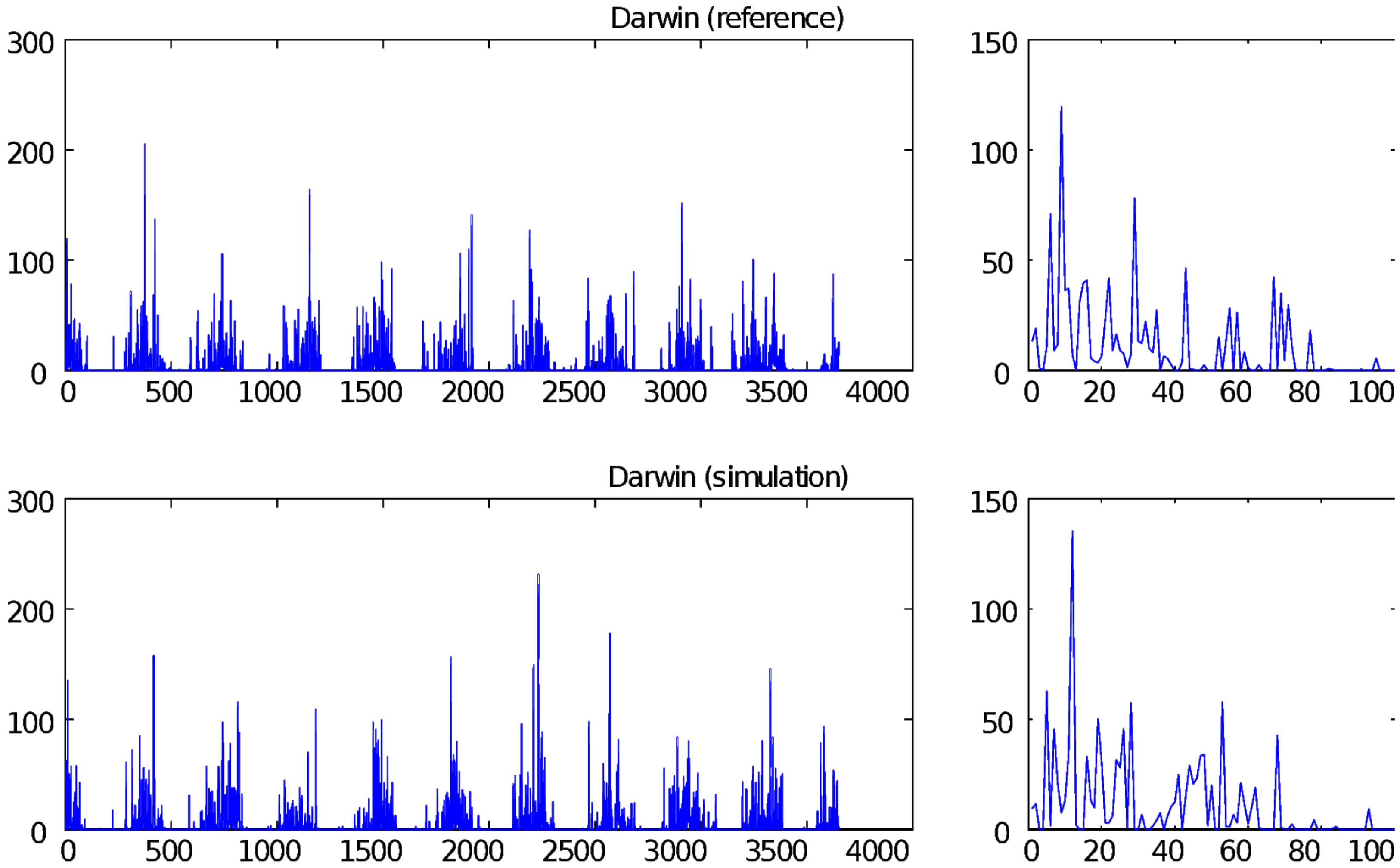


100 realizations of the same size of the TI.

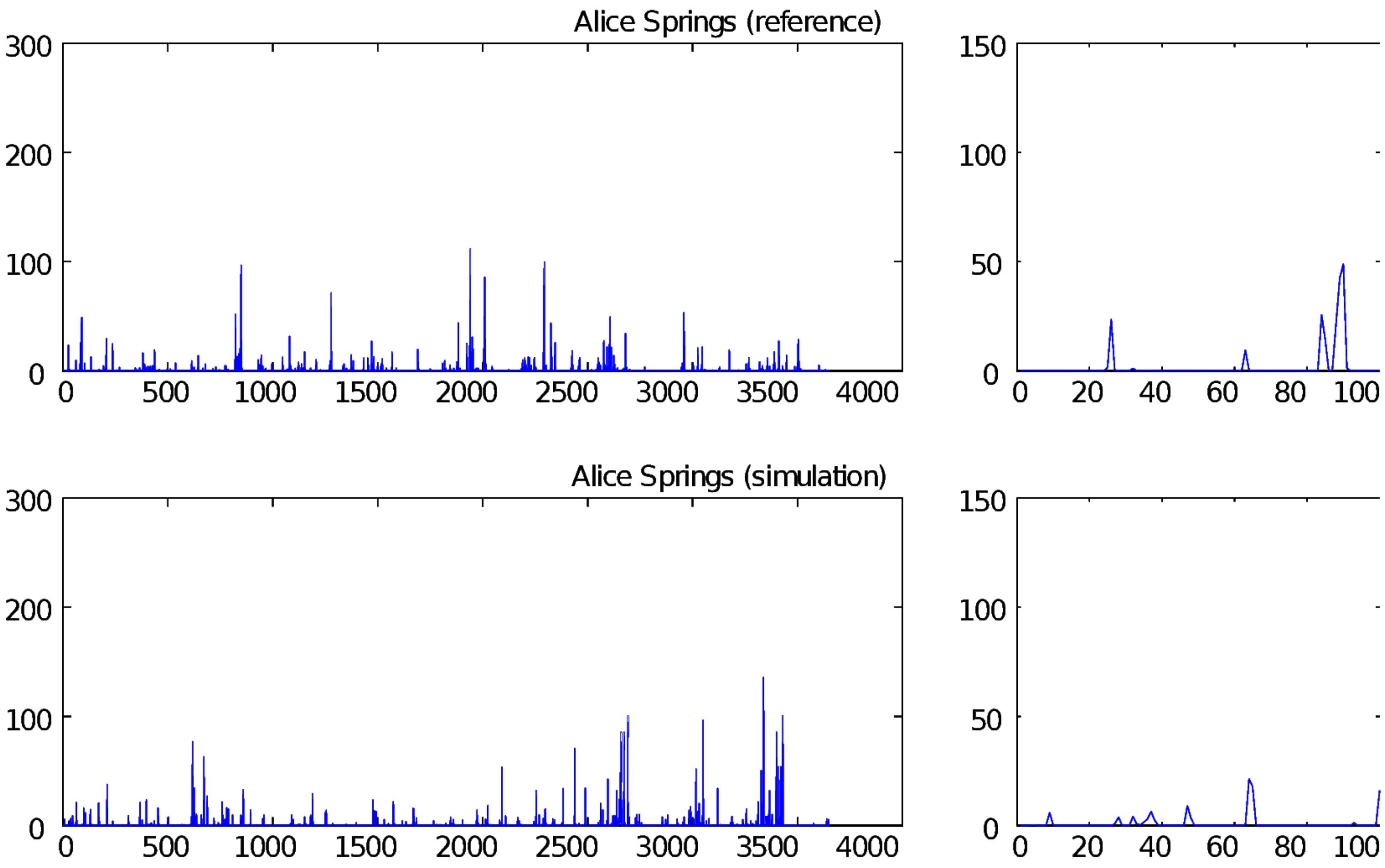
Visual comparison



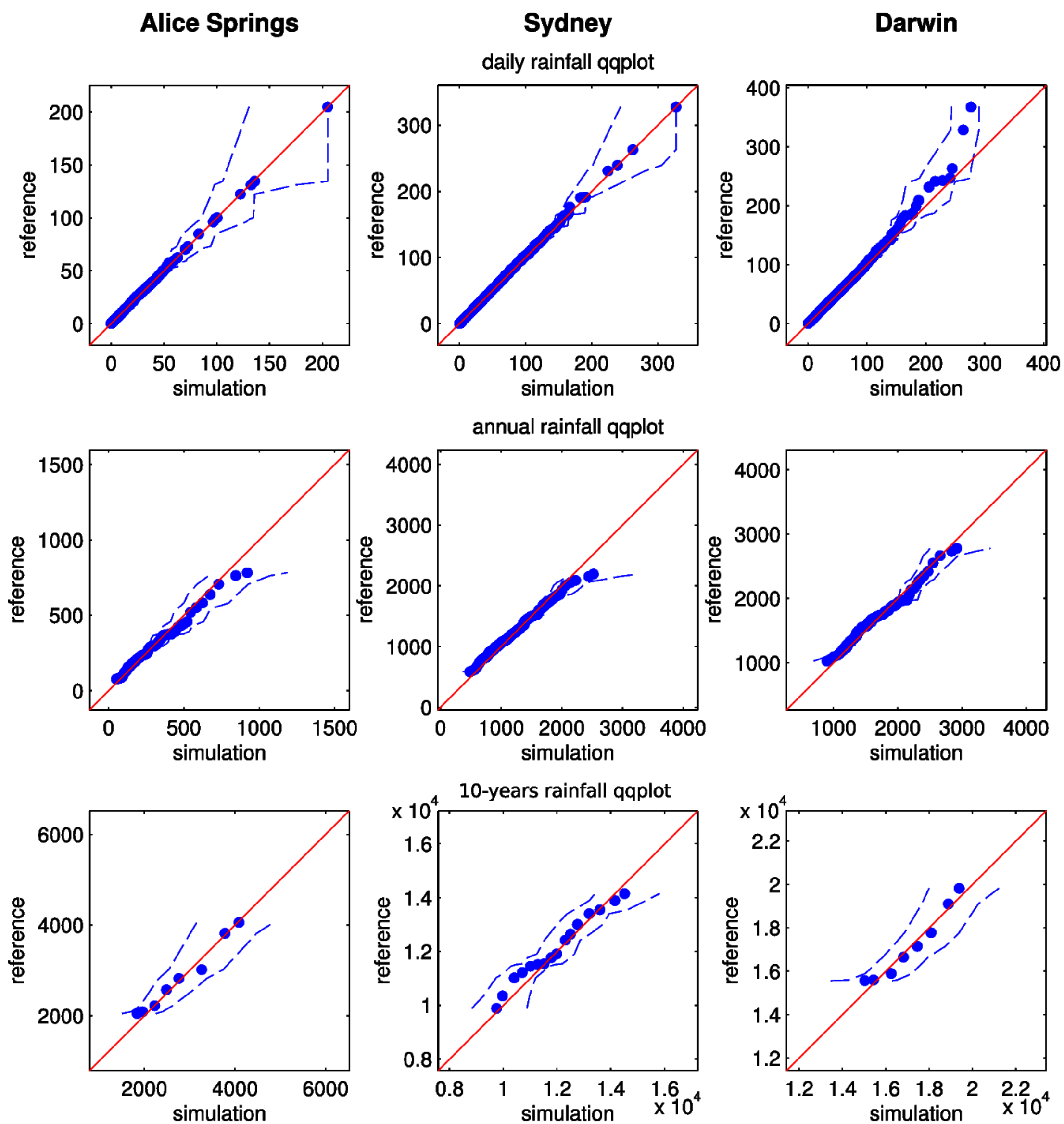
Visual comparison



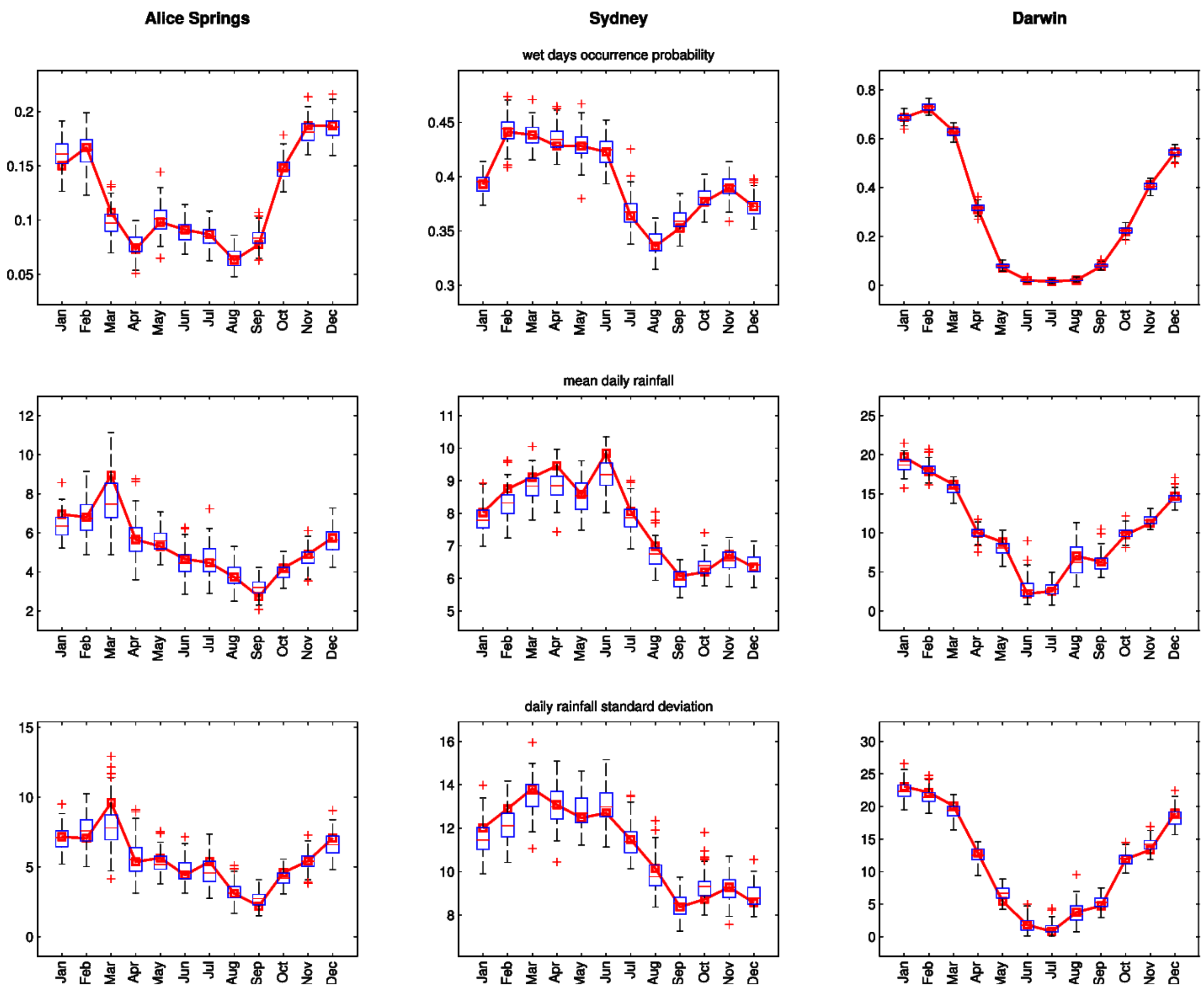
Visual comparison



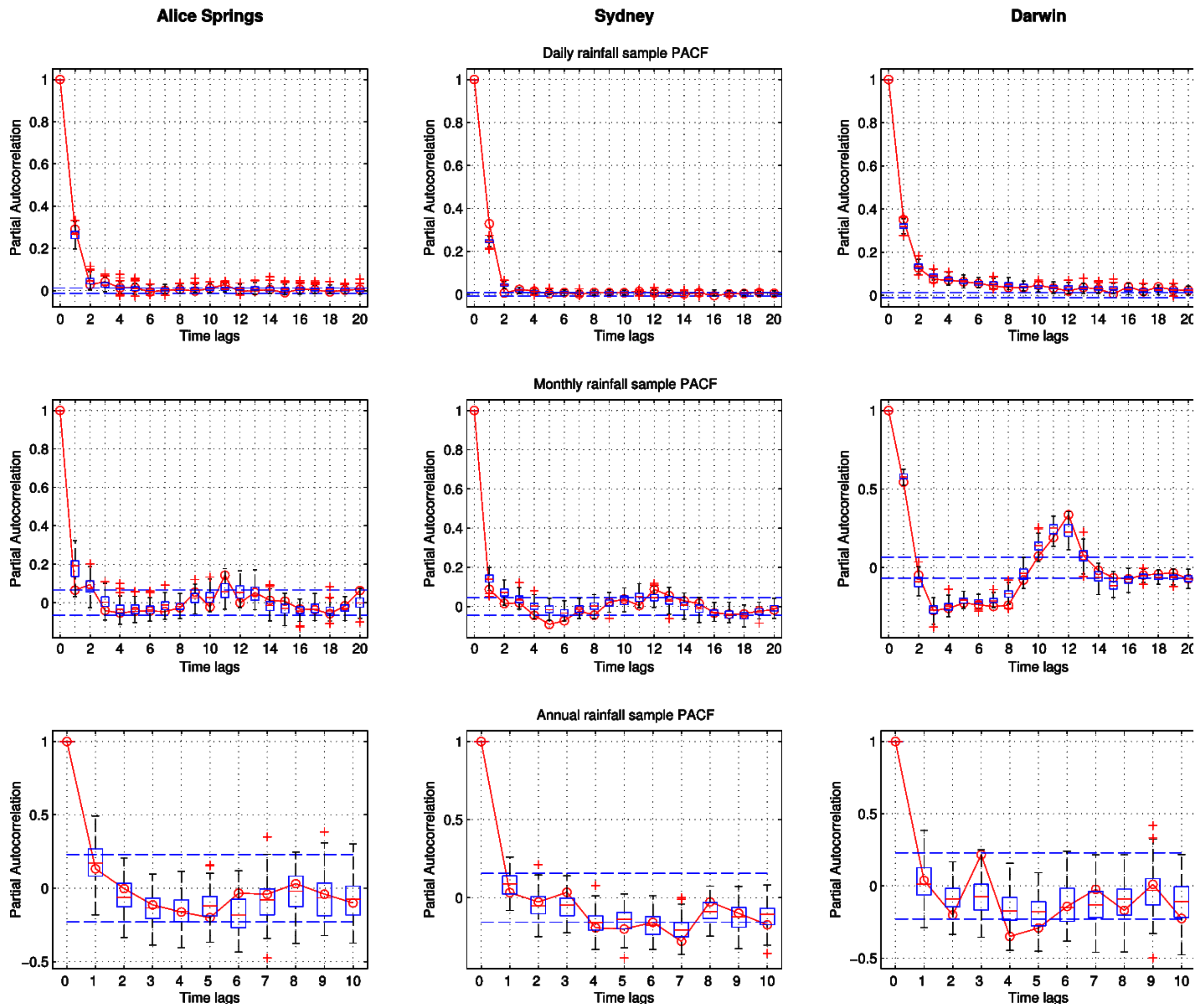
Marginal probability distribution at multiple scales.



Annual seasonality



Sample partial autocorrelation function (PACF) at multiple scales.



A comparison with a state-of-the-art MC model

[Oriani, Merothra et al., 2014, preparation]

Some recent Markov-Chain (**MC**) based algorithms [Harrold 2003, Mehrotra 2007] introduce non-linearity in the time dependence, i.e. the low-order conditional probability varies as a function of some low frequency covariates.

↓

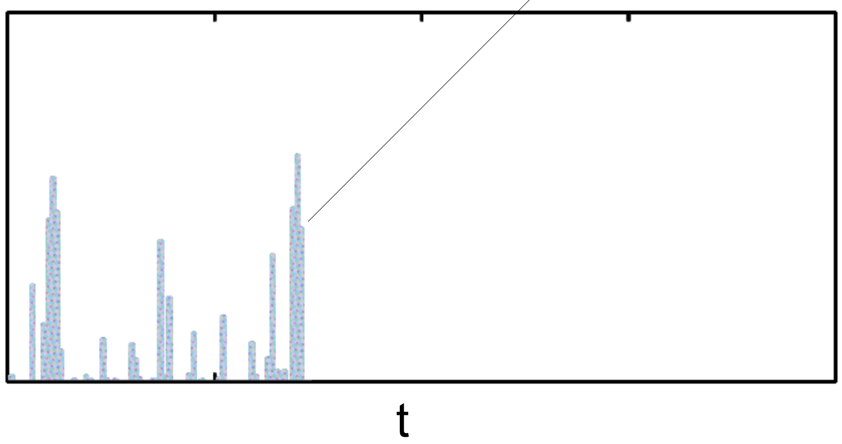
Low frequency fluctuations can be reproduced in the daily rainfall simulation.

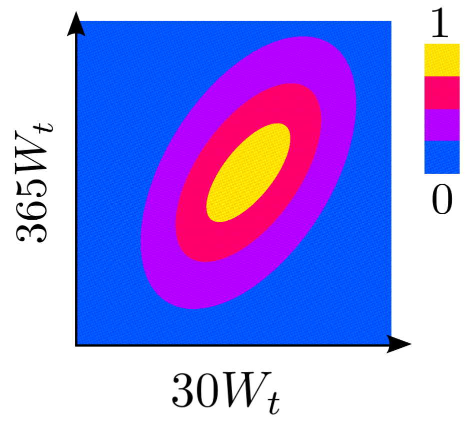
Modified Markov Model (MMM)

[Mehrotra 2007]

OCCURRENCE: $P(R_t = 1 | R_{t-1} = i, 30W_t, 365W_t)$

$$(30W_t, 365W_t | R_t = i, R_{t-1} = j) \sim N(\mu_{ij}, V_{ij})$$



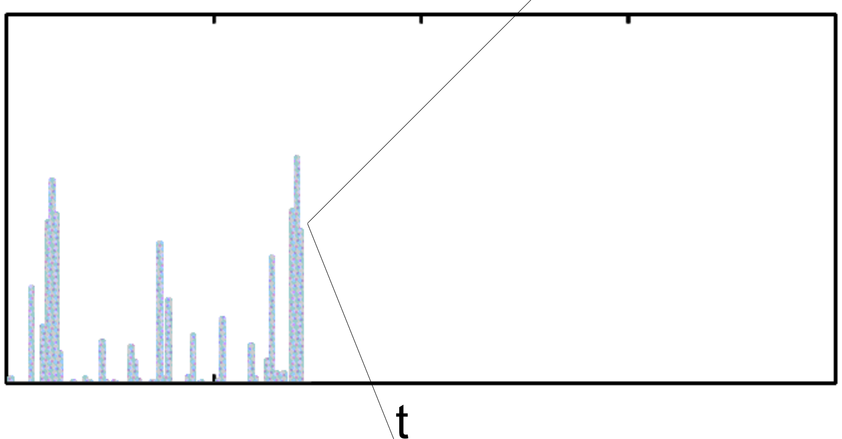


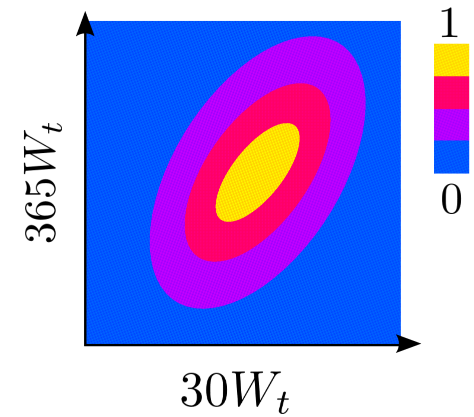
Modified Markov Model (MMM)

[Mehrotra 2007]

OCCURRENCE: $P(R_t = 1 | R_{t-1} = i, 30W_t, 365W_t)$

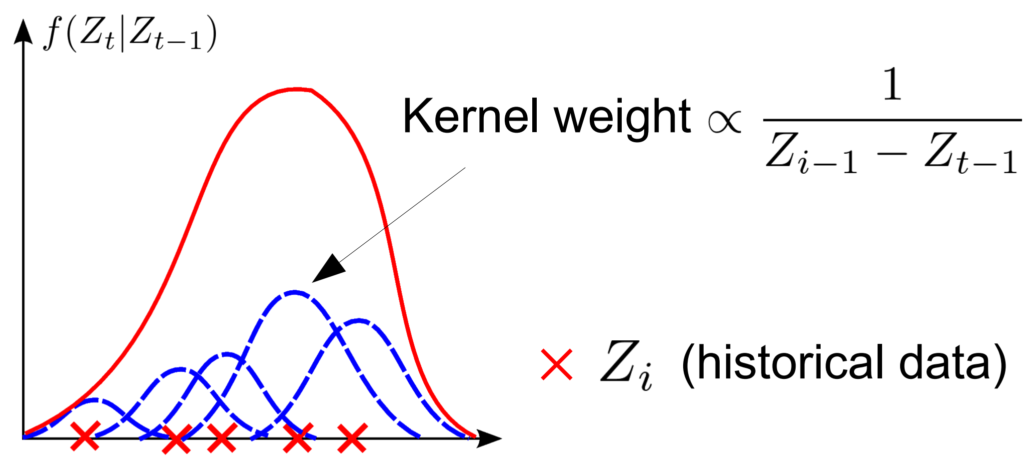
$$(30W_t, 365W_t | R_t = i, R_{t-1} = j) \sim N(\mu_{ij}, V_{ij})$$



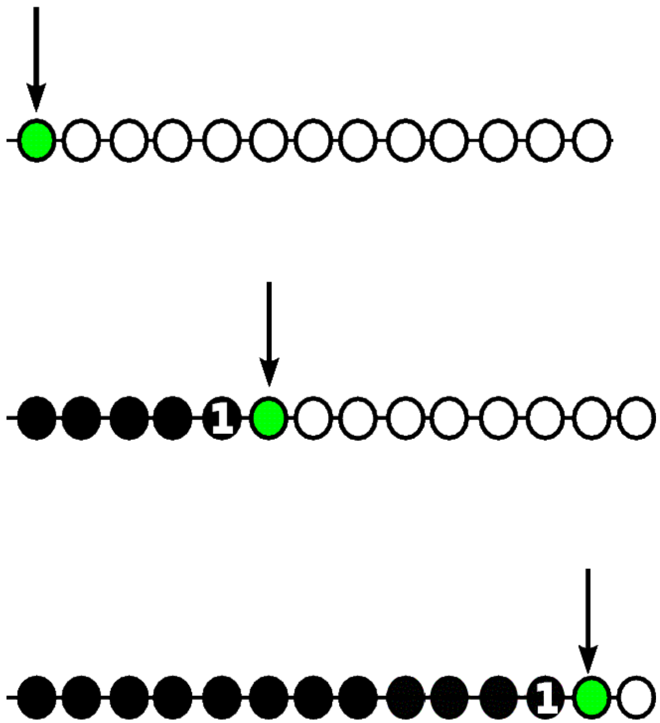


AMOUNT: $f(Z_t | Z_{t-1})$

using a conditional KDE [Sharma 1997]

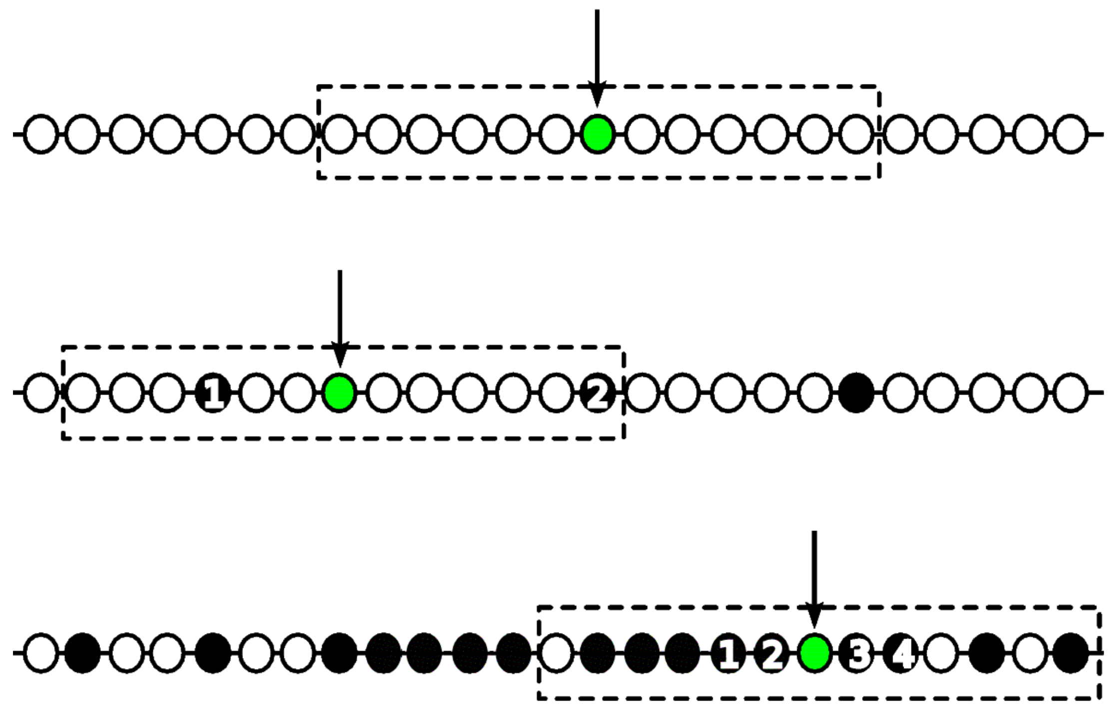


MMM



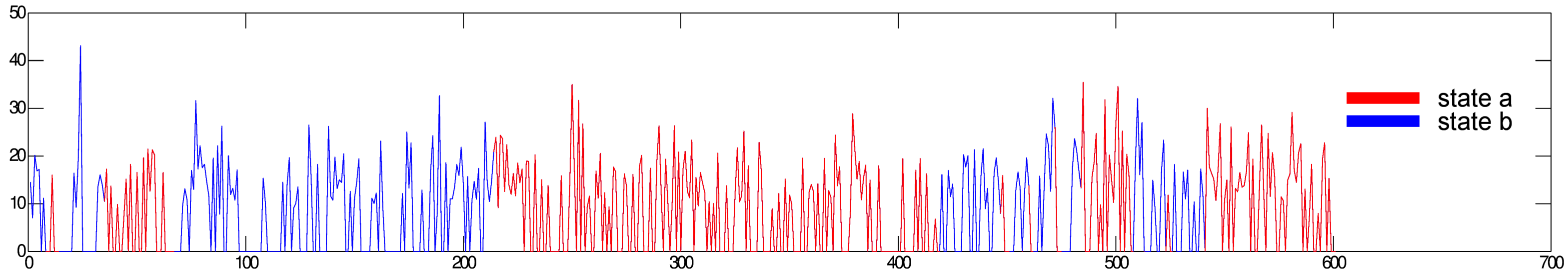
Linear simulation path
+
Fixed non-linear dependence

DS



Random simulation path
+
Variable non-parametric dependence

The reference: A synthetic signal with a chaotic seasonality and variable time-dependence



OCCURRENCE MODEL: 2 states Markov model

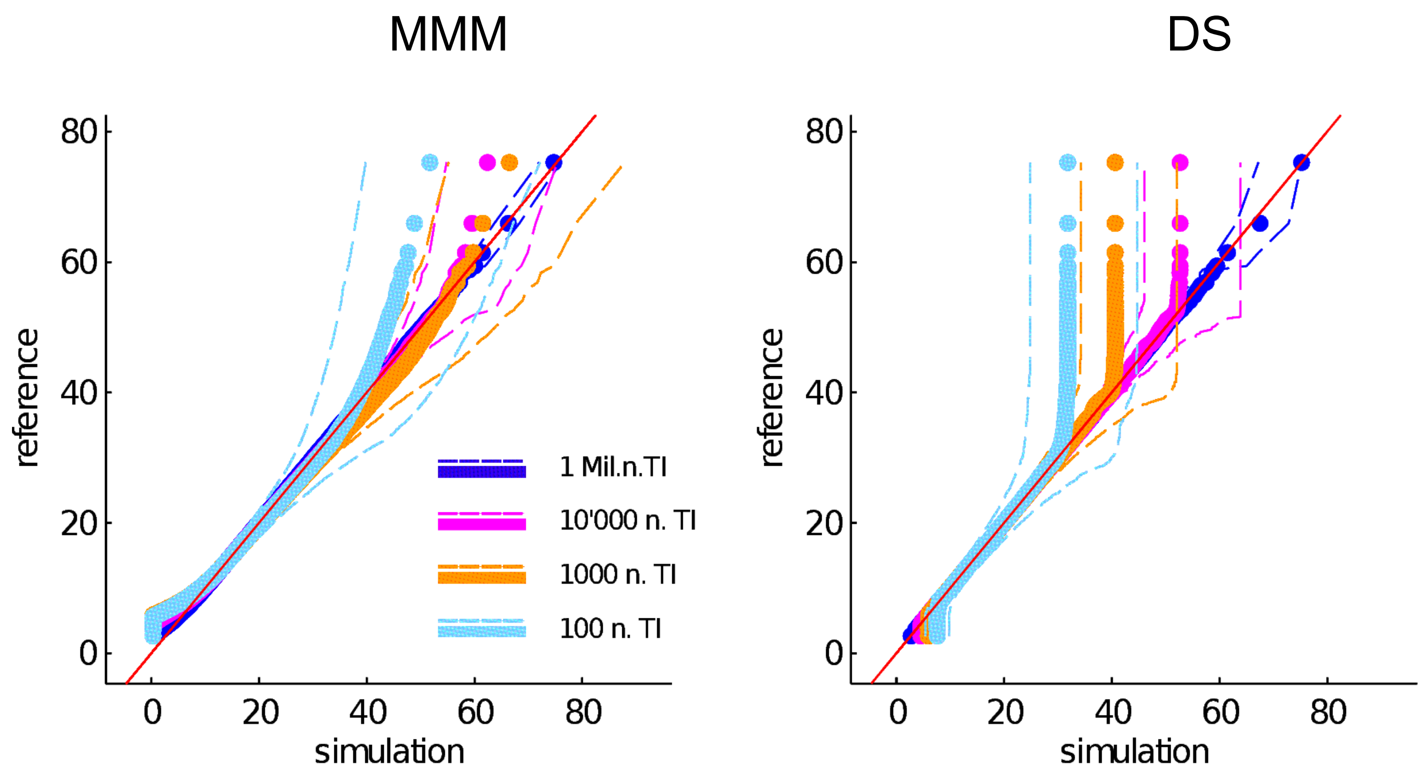
$$R_t \in \{0, 1\}$$

$$P(R_t = 1) = \begin{cases} \text{state a}: \quad P(R_t = 1 \mid R_{t-6} = i, R_{t-12} = j) & if \sum_{z=1}^{200} R_{t-z} > 95 \\ \text{state b}: \quad P(R_t = 1 \mid R_{t-1} = k) & otherwise \end{cases}$$

AMOUNT MODEL: $IID \sim lnN(\mu, \sigma^2)$

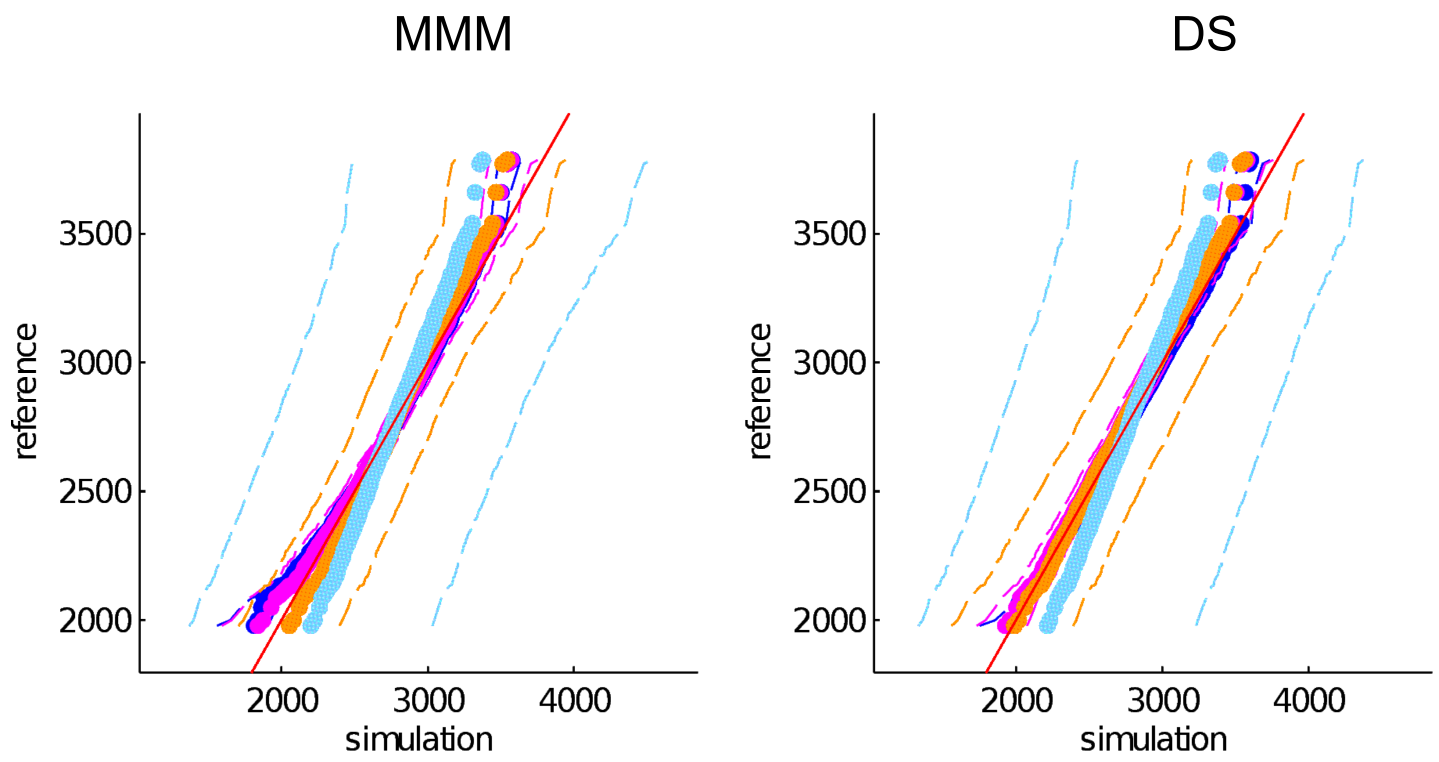
Simulation of the 1Mil nodes reference using progressively smaller training data sets (TI).

Daily rainfall distribution (mm)



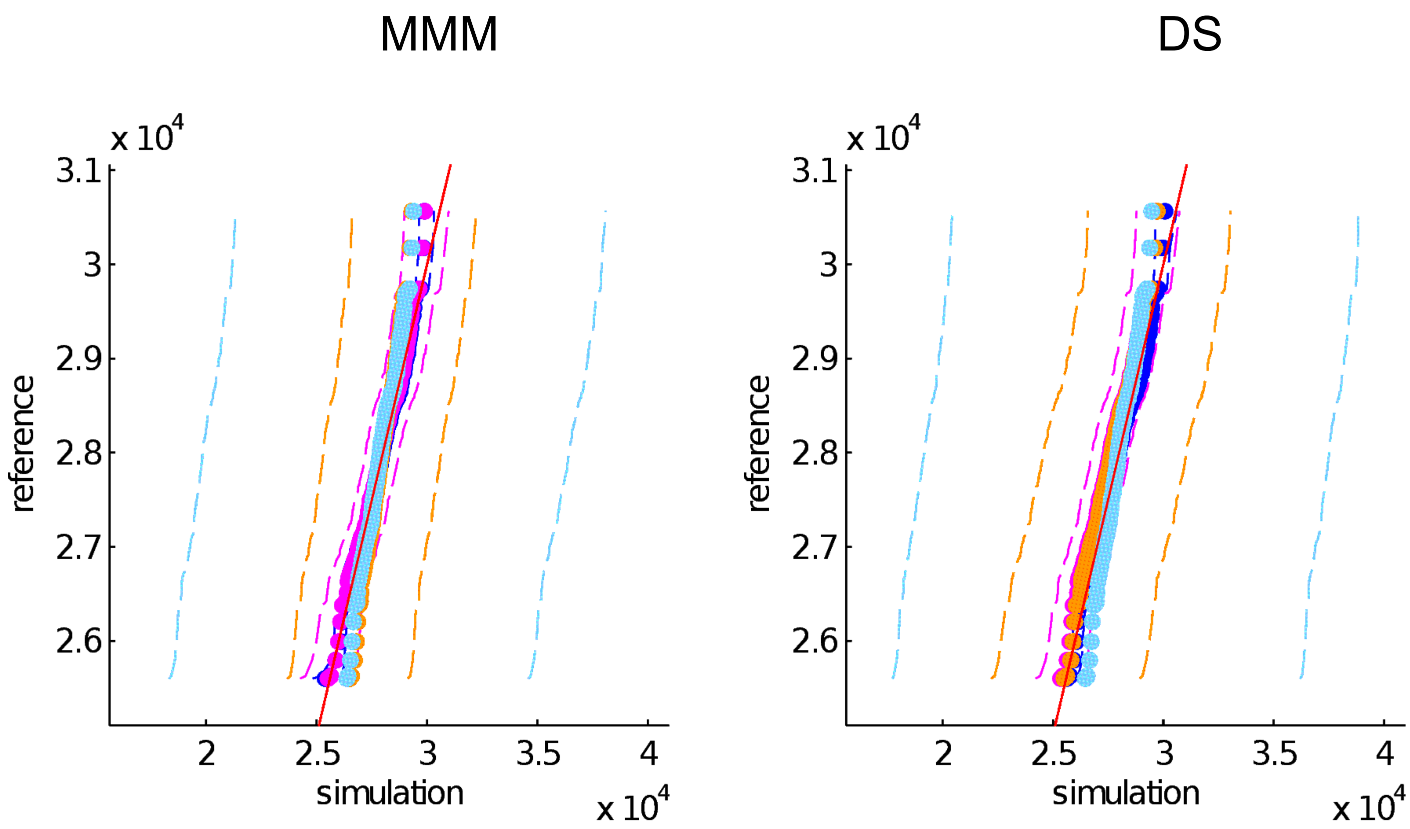
Simulation of the 1Mil nodes reference using progressively smaller training data sets (TI).

Annual rainfall distribution (mm)



Simulation of the 1Mil nodes reference using progressively smaller training data sets (TI).

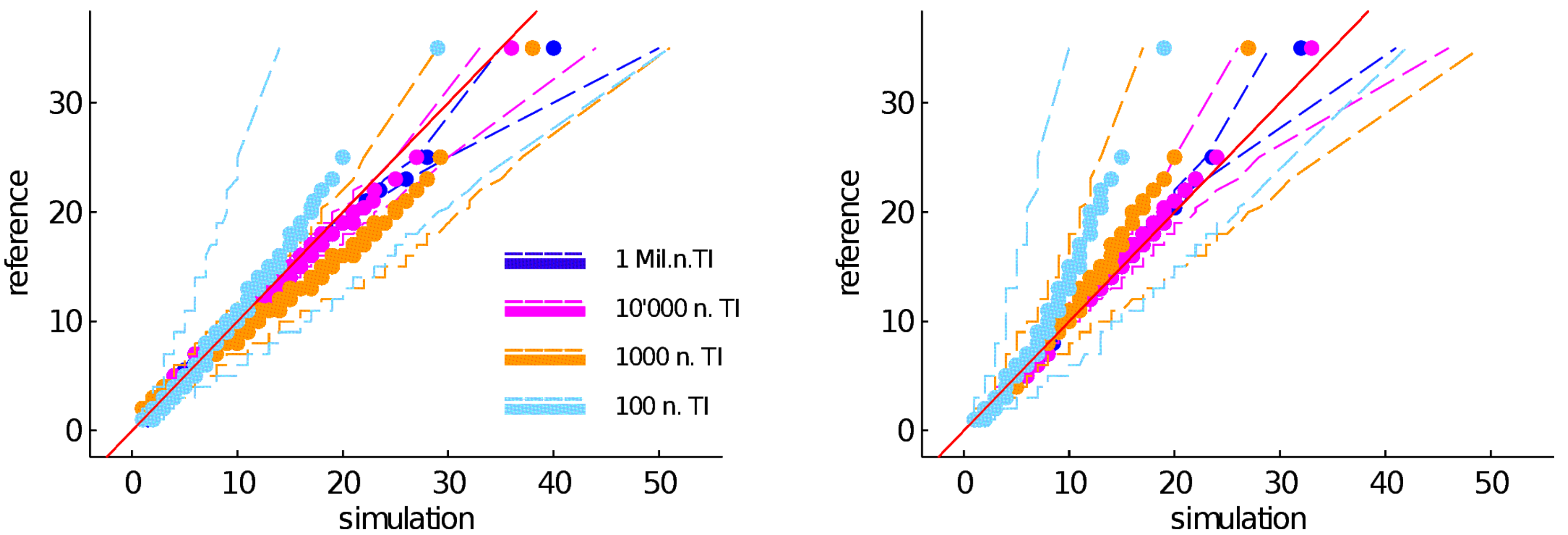
10-year rainfall distribution (mm)



Simulation of the 1Mil nodes reference using progressively smaller training data sets (TI).

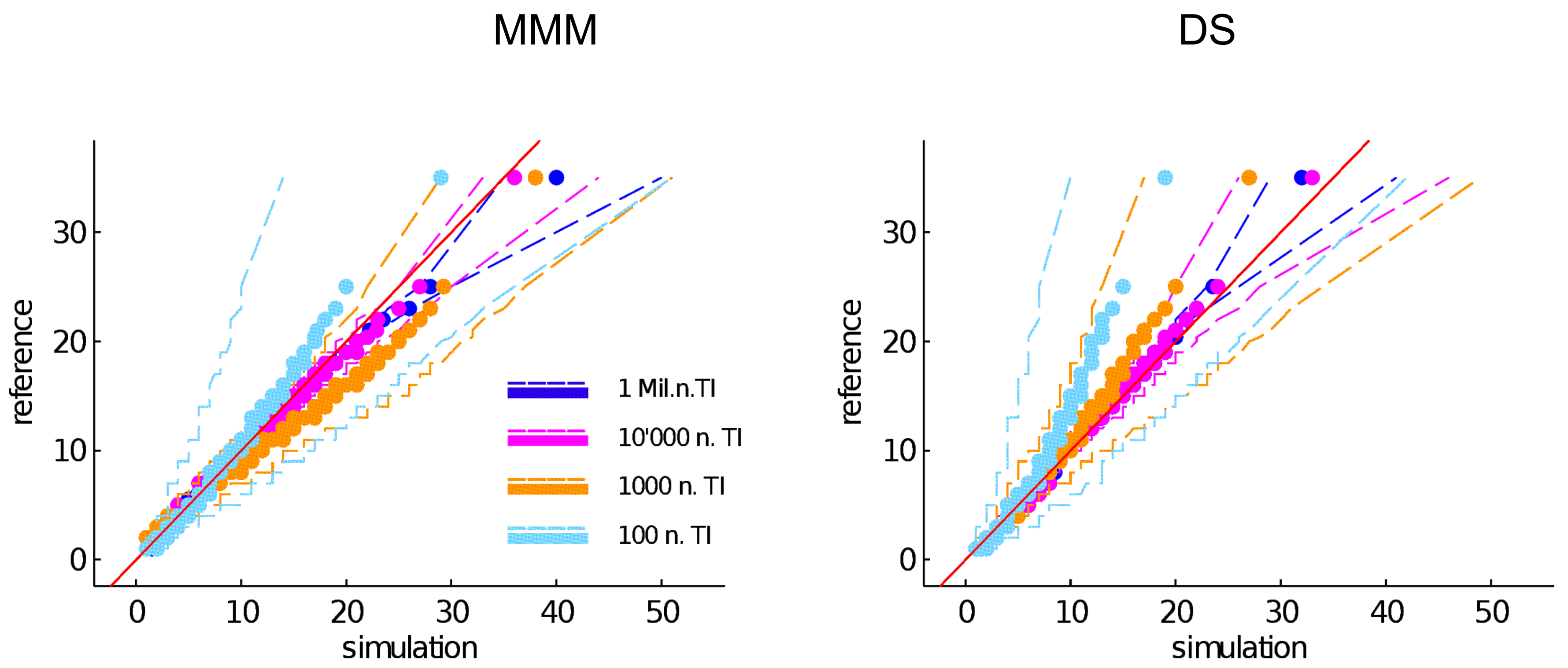
Dry spell distribution (days)

MMM DS



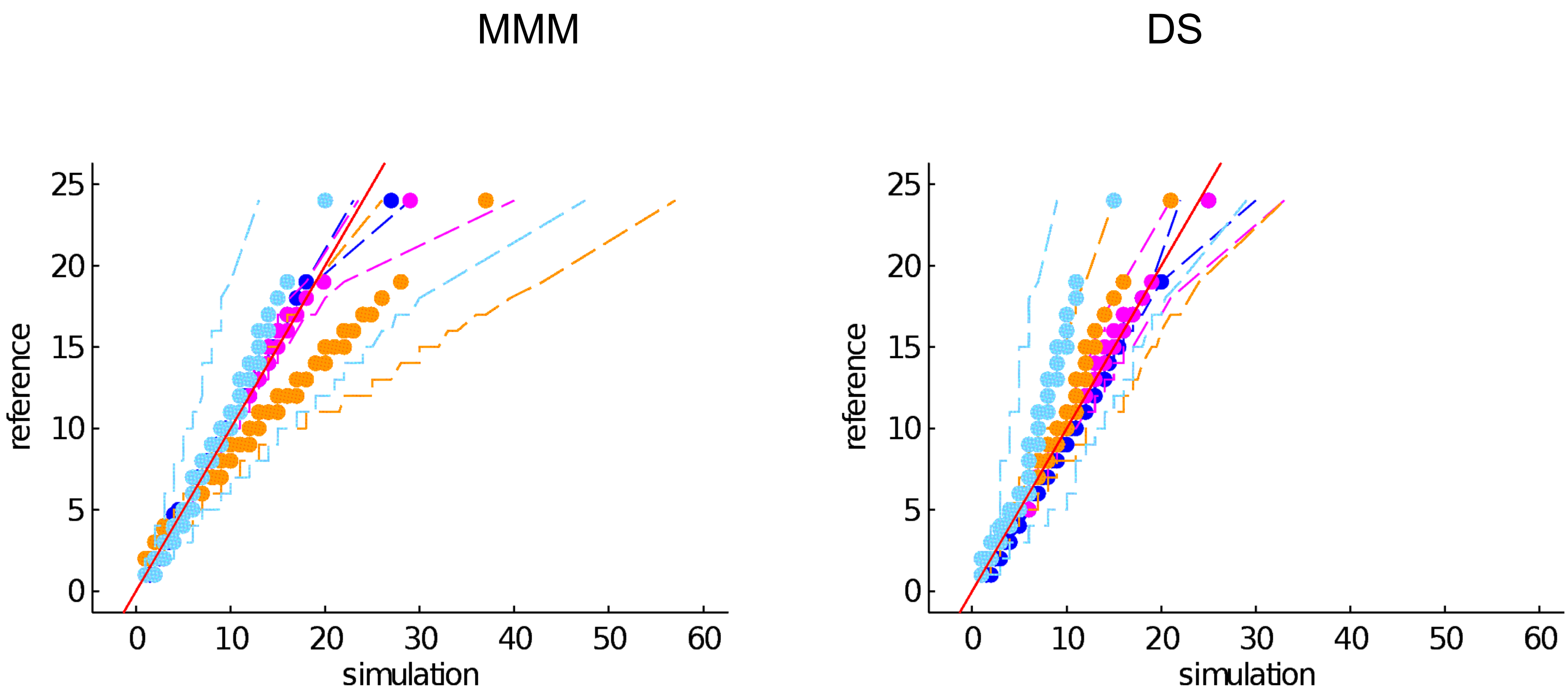
Simulation of the 1Mil nodes reference using progressively smaller training data sets (TI).

Dry spell distribution (days)



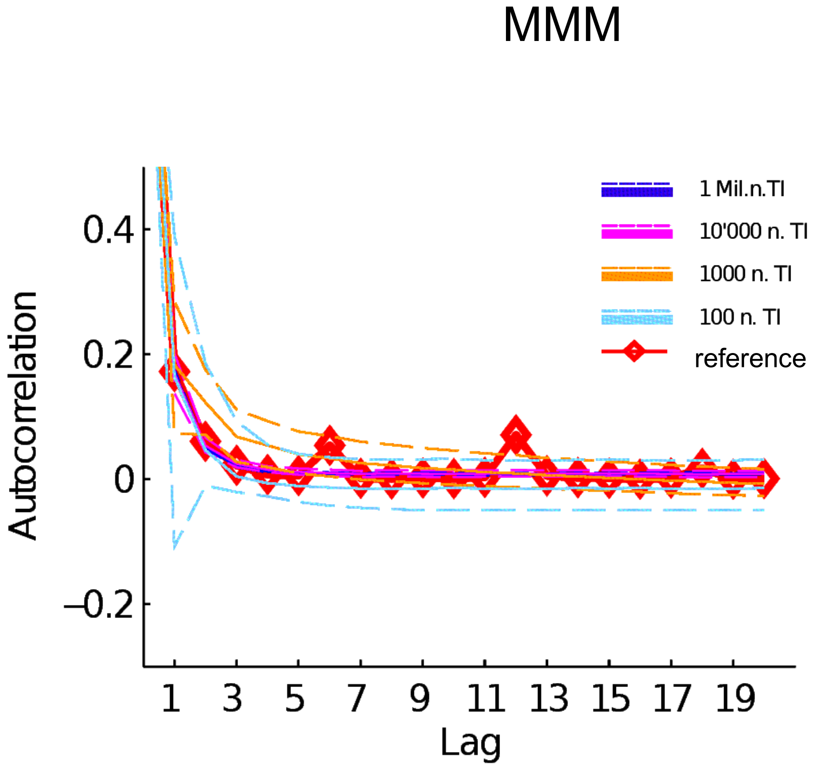
Simulation of the 1Mil nodes reference using progressively smaller training data sets (TI).

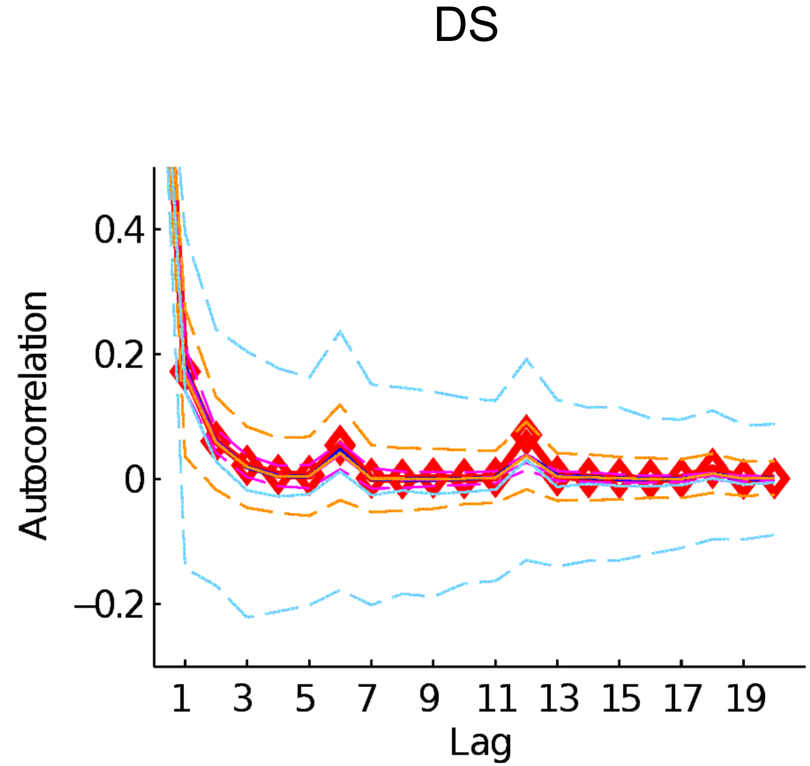
Wet spell distribution (days)



Simulation of the 1Mil nodes reference using progressively smaller training data sets (TI).

Sample autocorrelation function.
TOTAL SIGNAL



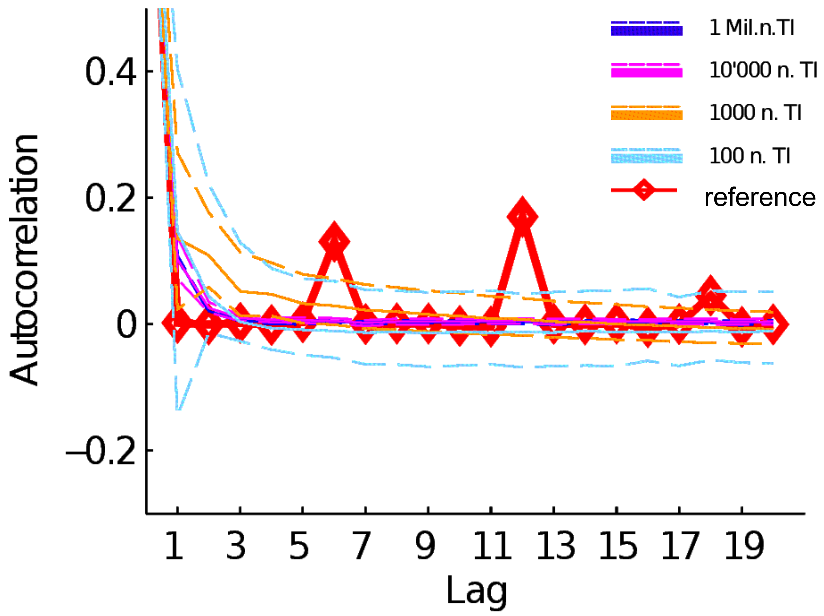


Simulation of the 1Mil nodes reference using progressively smaller training data sets (TI).

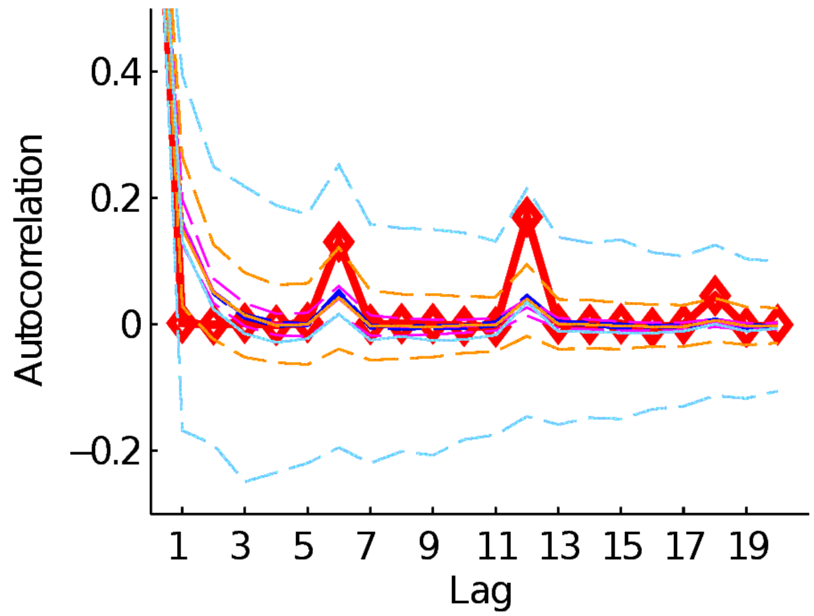
Sample autocorrelation function.

State a

MMM

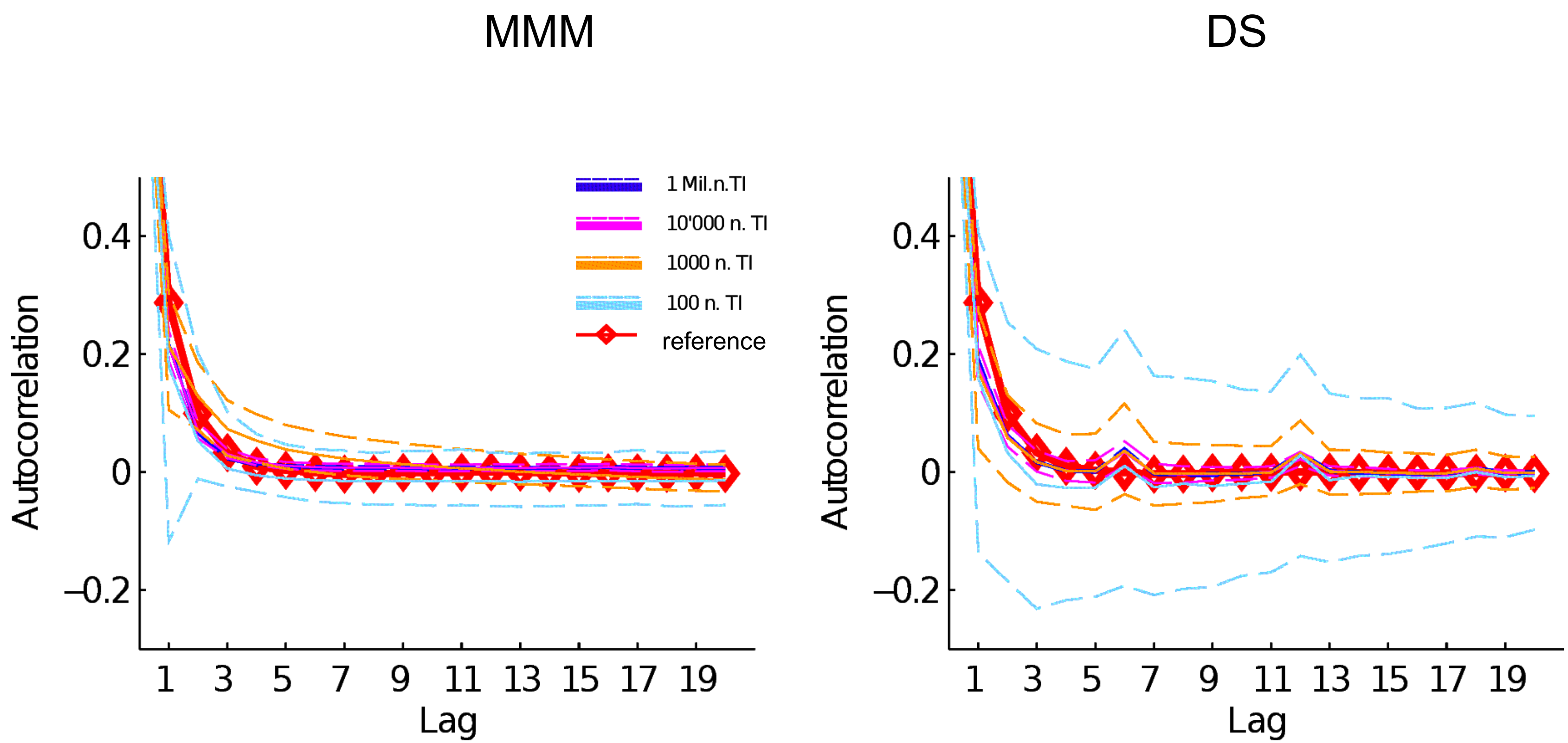


DS



Simulation of the 1Mil nodes reference using progressively smaller training data sets (TI).

Sample autocorrelation function.
State b



Conclusions

	High-order statistics	Extremes extrapolation	Multiple scale features	Non-stationarity detection
MMM (low order, nonlinear Markov model + conditional kernel smoothing)	🔴	🟢	🟠	🔴
DS (variable, high-order, multivariate time dependence + non parametric framework)	🟢	🔴	🟢	🔴

Which one of the two approaches is more efficient?

DS

If

- Large dataset
- Reproducing complex data structure is critical
- Avoid a prior model structure

MMM

If

- Limited amount of data
- The long-term structure is known and not overly complex
- A low order MC model can represent the short-term time-dependence adequately

Thank you!

Fabio Oriani

Bibliography

Harrold, T. I., Sharma, A., and Sheather, S. J.: A nonparametric model for stochastic generation of daily rainfall amounts, Water Resour. Res., 39, 1343, doi:10.1029/2003WR002570, 2003b.

Mariethoz, G., Renard, P., and Straubhaar, J.: The direct sampling method to perform multiple-point geostatistical simulations, Water Resour. Res., 46, W11536,doi:10.1029/2008WR007621, 2010.

Oriani, F., Straubhaar, J., Renard, P., Mariethoz, G., Simulation of rainfall time-series from different climatic regions using the Direct Sampling technique, Hydrology and Earth System Sciences. dx.doi.org/10.5194/hess-18-3015-2014

Mehrotra, R. and Sharma, A.: A semi-parametric model for stochastic generation of multi-site daily rainfall exhibiting low-frequency variability, J. Hydrol., 335, 180–193, doi:10.1016/j.jhydrol.2006.11.011, 2007a.

Mehrotra, R. and Sharma, A.: Preserving low-frequency variability in generated daily rainfall sequences, J. Hydrol., 345, 102–120, doi:10.1016/j.jhydrol.2007.08.003, 2007b.