CoSMoS-MATLAB (v0.9): Complete Stochastic Modelling Solution

Summary

Hydroclimatic processes come in all "shapes and sizes". They are characterized by different spatiotemporal correlation structures and probability distributions that can be continuous, mixed-type, discrete or even binary. Simulating such processes by reproducing precisely their marginal distribution and linear correlation structure, including features like intermittency, can greatly improve hydrological analysis and design. Traditionally, modelling schemes are case specific and typically attempt to preserve few statistical moments providing inadequate and potentially risky distribution approximations.

Here, a single framework is proposed that unifies, extends, and improves a general-purpose modelling strategy, based on the assumption that any process can emerge by transforming a specific "parent" Gaussian process. A novel mathematical representation of this scheme, introducing parametric correlation transformation functions, enables straightforward estimation of the parent-Gaussian process yielding the target process after the marginal back transformation, while it provides a general description that supersedes previous specific parameterizations, offering a simple, fast and efficient simulation procedure for every stationary process at any spatiotemporal scale.

This framework, also applicable for cyclostationary and multivariate modelling, is augmented with flexible parametric correlation structures that parsimoniously describe observed correlations. Real-world simulations of various hydroclimatic processes with different correlation structures and marginals, such as precipitation, river discharge, wind speed, humidity, extreme events per year, etc., as well as a multivariate example, highlight the flexibility, advantages, and complete generality of the method. From (Papalexiou, 2018).

1. Description

CoSMoS-MATLAB is developed to help scientists and researchers generate time series with desired properties. In this beta version (v0.9), you can generate time series from six marginal distributions, i.e., Normal ($\mathcal{N}$), Weibull ($\mathcal{W}$), Pareto type II ($\mathcal{P}$II), Burr type III ($\mathcal{Br}$III), Burr type XII ($\mathcal{Br}$XII), and Generalized Gamma ($\mathcal{GG}$); and four autocorrelation structures (ACS), i.e., Pareto type II (PII), Weibull (W), Burr type XII (BrXII), Generalized Logarithmic (GL) , Markovian (M), and fractional Gaussian noise (fGn). Details on these distributions and autocorrelation structures are given in Appendix (see also Section 3.1 and 3.2 in Papalexiou, 2018).

CoSMoS-MATLAB offers a Graphical User Interface (GUI) that makes easy to select the properties of the time series you wish to generate (Figure 1). Details on how to use the GUI are given in the next section.

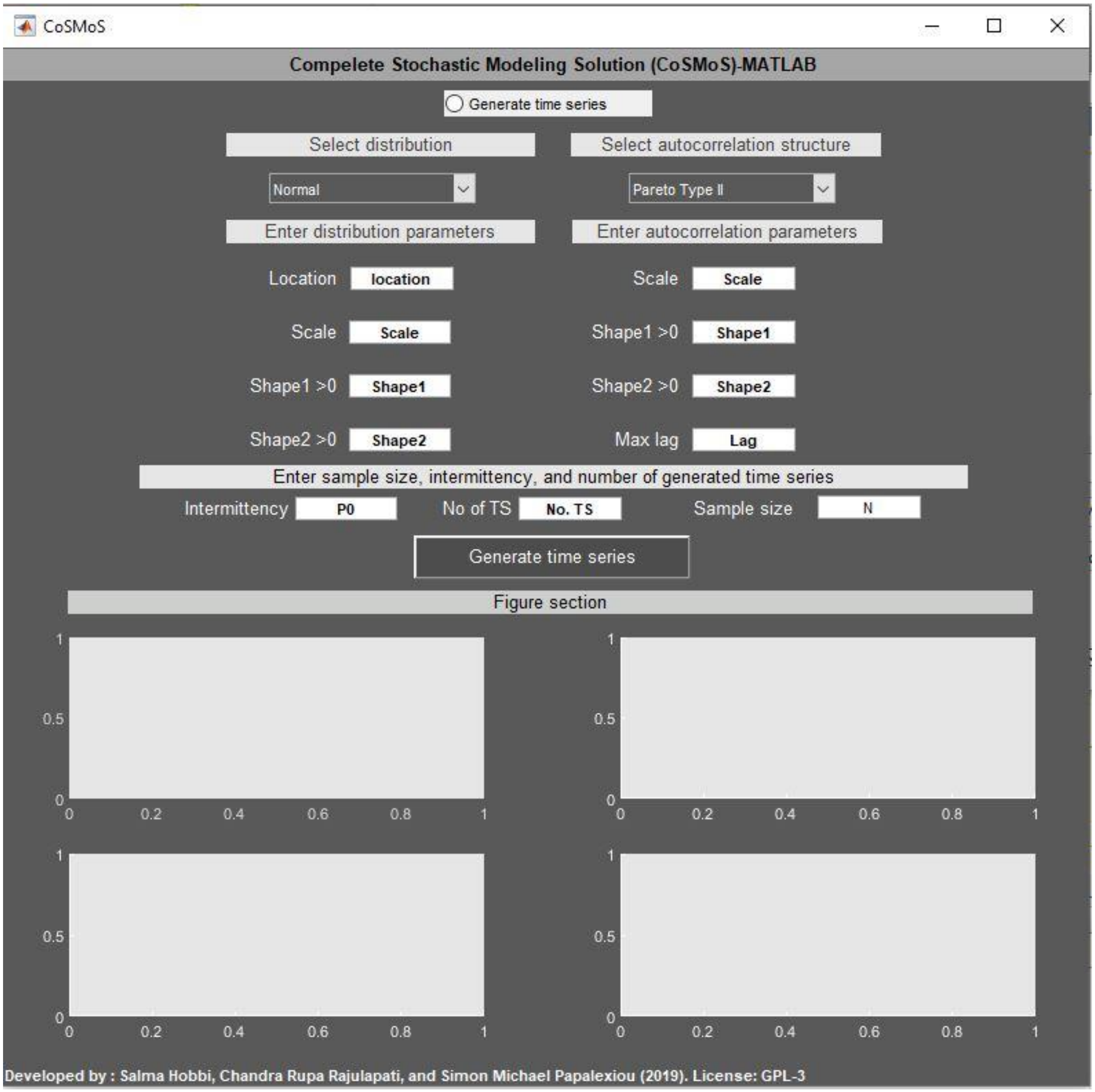

Figure 1. Graphical User Interface of CoSMoS-MATLAB

2. Installation

1. Visit https://github.com/SMPLab/CoSMoS-MATLAB and download the files: CoSMoS.m, call_CoSMoS.m, CoSMoS.fig, and Visualize.m.
2. Create a directory and save these files.
3. Open MATLAB and change the current MATLAB directory to the directory created in step 2.

3. Using the GUI (see also QuickGuide.pdf)

1. Double click on Call_CoSMoS.m and run it to open the GUI.

2. In the GUI select the probability distribution and autocorrelation structure from the drop-down lists.
3. Enter the parameters of selected distribution and autocorrelation structure.
4. Enter the intermittency value (as probability zero), sample size (time series length), and number of time series you wish to generate. Default values are: probability zero $p_0 = 0$; time series length = 10,000; number of time series = 1.
5. Click the “Generate time series” button.

4. Results (QuickGuide.pdf)

1. The GUI offers four plots: (1) the generated time series, (2) the target distribution compared to the empirical, (3) the target autocorrelation structure compared to the empirical, (4) the correlation transformation function. (If more than one time series are generated the GUI shows results for the first one).
2. The ‘Results’ folder is stored in the directory created in Installation Step 2. The generated .mat files in the folder store data to create the four plots for any of the time series. Particularly, five files are created:
    a. SimulationInfo.mat – summary of the simulation parameters.
    b. GeneratedTS.mat – values of the generated time series.
    c. Distribution.mat – data to create the exceedance probability plot.
    d. ACS.mat – data to create autocorrelation plot.
    e. ACTF.mat – data to visualize the autocorrelation transformation function.
3. Use Visualize.m to create the plots shown in GUI for any of the time series and perform any additional analysis in MATLAB.

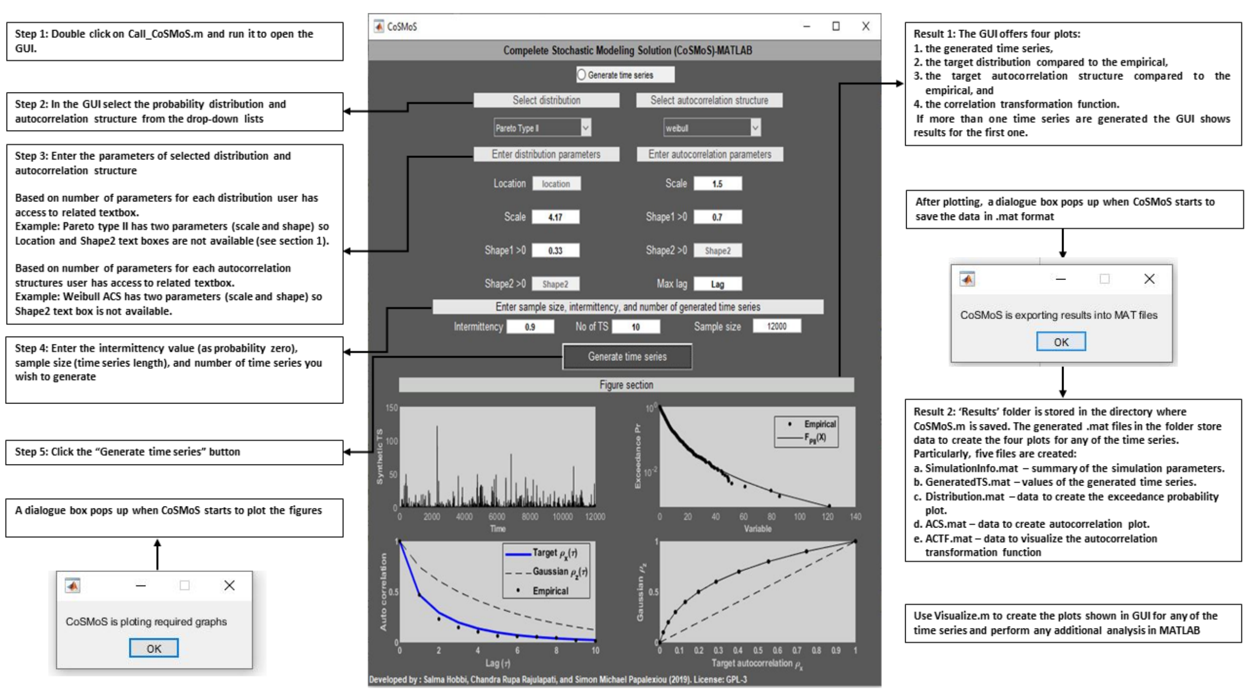

Figure 2. Guide to use GUI of CoSMoS

5. Summary of the method used

The generated time series preserve specified probability distribution and autocorrelation structure. Users can generate as many and as long time series from processes such as precipitation, wind, and temperature. The Toolbox is developed based on a framework that unified, extended, and improved a modelling strategy that generates time series by transforming "parent" Gaussian time series having specific characteristics (Papalexiou, 2018). **Table 1** summarizes the steps followed to generate synthetic time series along with functions used in MATLAB. Equation numbers referred in the table correspond to equation numbers in the paper Papalexiou (2018).

Table 1. Steps to generate synthetic time series along with functions called in CoSMoS-MATLAB.

Step 1.	For a chosen distribution $F_X(x)$ use the Auto Correlation Transformation Integral $\mathcal{R}\big(\theta_x, \rho_z(\tau)\big)$ (ACTI; Eq.9†) to estimate (ρ_X, ρ_Z) points. Where $\rho_Z = (0, 0.1, \dots, 0.95, 1)$ are autocorrelations of the Gaussian process and ρ_X the corresponding ones for the process with the selected $F_X(x)$. *MATLAB function*: actpnts_t
Step 2.	Estimate the parameters b and c of the Auto-Correlation Transformation Function (ACTF; Eq.27) using the points estimated in Step 1. *MATLAB function*: fitactf
Step 3.	Use the fitted ACTF to transform a parametric target ACS $\rho_X(\tau)$ into the parent Gaussian ACS (pGACS) $\rho_Z(\tau)$. *MATLAB function*: ACS_Z_lag
Step 4.	Generate Gaussian time series using an AR(p) (Eq.31) the reproduces the estimated parent-Gaussian ACS. Transform the Gaussian time series using the marginal-back transformation $X(t) = Q_X\Big(\Phi_Z\big(z(t)\big)\Big)$. The resulting time series preserve the marginal $F_X(x)$ and the target ACS $\rho_X(\tau)$. *MATLAB function*: AR_p_LTS

†Equation number refer to the supporting paper (Papalexiou, 2018).

6. Funding

The package was partly funded by the Global institute for Water Security (GIWS; https://www.usask.ca/water/) and the Global Water Futures (GWF; https://gwf.usask.ca/) program.

7. License and Disclaimer Information

Package: 'CoSMoS-MATLAB'
Title: Complete Stochastic Modelling Solution (CoSMoS)
Version: v0.9 (beta)
Description: A single framework, unifying, extending, and improving a general-purpose modelling strategy, based on the assumption that any process can emerge by transforming a specific "parent" Gaussian process (Papalexiou, 2018).
License: GPL-3
Depends: MATLAB (tested on MATLAB R2018)

Repository: SMPLab
Date/Publication: 2019-10-17 16:30:00 CST
Redistribution and use the source code, with or –without modification, are permitted if the following conditions are met:

- Redistribution of the source code must retain the developers' and affiliated institution names.
- Redistribution of source code must retain all the information in this section along with the following disclaimer

This software is designed for educational and research purposes only, and the commercial use of this package is strictly prohibited. It is provided by developers 'as is' and any express or implied warranties of merchantability for a particular purpose are disclaimed. Authors, developers and the affiliate institutions accept neither responsibility nor liability for direct/indirect errors, mistakes, and consequential damages; loss of use, data, or profits; or business interruption. User is responsible for any such errors or damages as the package works based on his/her discretion.
Neither the University of Saskatchewan nor the authors and contributors may be used to endorse or promote products derived from this software without specific prior written permission.

8. Author(s)

Coded by: Salma Hobbi
Conceptual design by: Salma Hobbi, Chandra Rupa Rajulapati, Simon Michael Papalexiou,
Maintainer: Salma Hobbi and Chandra Rupa Rajulapati
Contact: Salma Hobbi (salma.hobbi@usask.ca)

Reference

Papalexiou, S. M. (2018). Unified theory for stochastic modelling of hydroclimatic processes: Preserving marginal distributions, correlation structures, and intermittency. *Advances in Water Resources*, *115*, 234–252.

Appendix

Distributions

Burr type III distribution

$$f_{\mathcal{B}r\mathrm{III}}(x) = \frac{1}{\beta}\left(\frac{x}{\beta}\right)^{-\frac{1}{\gamma_2}-1}\left(\frac{1}{\gamma_1}\left(\frac{x}{\beta}\right)^{-\frac{1}{\gamma_2}}+1\right)\left(\frac{1}{\gamma_1}\left(\frac{x}{\beta}\right)^{-\frac{1}{\gamma_2}}+1\right)^{-\gamma_1\gamma_2-1} \tag{A.1}$$

$$F_{\mathcal{B}r\mathrm{III}}(x) = \left(\left(\frac{1}{\gamma_1}\left(\frac{x}{\beta}\right)^{-\frac{1}{\gamma_2}}\right)+1\right)^{-\gamma_1\gamma_2} \tag{A.2}$$

$$Q_{\mathcal{B}r\mathrm{III}}(x) = \mathrm{\beta}\left(\gamma_1\left(u^{-\frac{1}{\gamma_1\gamma_2}}-1\right)\right)^{-\gamma_2} \tag{A.3}$$

$$m_{\mathcal{B}r\mathrm{III}}(x) = \frac{\beta^q {\gamma_1}^{\gamma_2(-q)}\Gamma\big((q+\gamma_1)\gamma_2\big)\Gamma(1-q\gamma_2)}{\Gamma(\gamma_1\gamma_2)} \tag{A.4}$$

where $\beta > 0$ is scale, $\gamma_1 > 0, 0 < \gamma_2 < 0.5$ are shape parameters and Γ is the gamma function.

Burr type XII distribution

$$f_{\mathcal{B}r\mathrm{XII}}(x) = \frac{1}{\beta}\left(\frac{x}{\beta}\right)^{\gamma_1-1}\left(\gamma_2\left(\frac{x}{\beta}\right)^{\gamma_1}+1\right)^{-\frac{1}{\gamma_1\gamma_2}-1} \tag{A.1}$$

$$F_{\mathcal{B}r\mathrm{XII}}(x) = 1-\left(\left(\frac{x}{\beta}\right)^{\gamma_1}+1\right)^{-\frac{1}{\gamma_1\gamma_2}} \tag{A.2}$$

$$Q_{\mathcal{B}r\mathrm{XII}}(x) = \beta\left(-\frac{1-(1-u)^{-\gamma_1\gamma_2}}{\gamma_2}\right)^{-\frac{1}{\gamma_1}} \tag{A.3}$$

$$m_{\mathcal{B}r\mathrm{XII}}(x) = \frac{1}{\gamma_1}\left(\beta^q {\gamma_2}^{-\frac{q}{\gamma_1}-1}\mathrm{B}\left(\frac{q+\gamma_1}{\gamma_1},\frac{1-q\gamma_2}{\gamma_1\gamma_2}\right)\right) \tag{A.4}$$

where $\beta > 0$ is scale, $\gamma_1 > 0, 0 < \gamma_2 < 0.5$ are shape parameters, and Γ and B are gamma and beta functions respectively.

Generalized Gamma distribution

$$f_{\mathcal{GG}}(x) = \frac{1}{\beta\Gamma\left(\frac{\gamma_1}{\gamma_2}\right)}\gamma_2 e^{-\left(\frac{x}{\beta}\right)^{\gamma_2}}\left(\frac{x}{\beta}\right)^{\gamma_1 - 1} \tag{A.9}$$

$$F_{\mathcal{GG}}(x) = Q\left(\frac{\gamma_1}{\gamma_2}, 0, \left(\frac{x}{\beta}\right)^{\gamma_2}\right) \tag{A.10}$$

$$Q_{\mathcal{GG}}(u) = Q^{-1}\left(\frac{\gamma_1}{\gamma_2}, 0, u\right)^{\frac{1}{\gamma_2}} \tag{A.11}$$

$$m_{\mathcal{GG}}(x) = \frac{1}{\Gamma\left(\frac{\gamma_1}{\gamma_2}\right)}\beta^q\Gamma\left(\frac{q}{\gamma_2} + \frac{\gamma_1}{\gamma_2}\right) \tag{A.12}$$

where $\beta > 0$ is scale, $\gamma_1 > 0, 0 < \gamma_2 < 0.5$ are shape parameters, and Γ and Q are gamma and incomplete gamma functions respectively.

Pareto type II distribution

$$f_{\mathcal{P}\mathrm{II}}(x) = \frac{1}{\beta}\left(\frac{\gamma x}{\beta} + 1\right)^{\frac{1}{\gamma} - 1} \tag{A.13}$$

$$F_{\mathcal{P}\mathrm{II}}(x) = 1 - \left(\frac{\gamma x}{\beta} + 1\right)^{\frac{1}{\gamma}} \tag{A.14}$$

$$Q_{\mathcal{P}\mathrm{II}}(x) = \frac{\beta((1 - u)^{-\gamma} - 1)}{\gamma} \tag{A.15}$$

$$m_{\mathcal{P}\mathrm{II}}(x) = \frac{1}{\Gamma\left(\frac{1}{\gamma}\right)}\Gamma(q + 1)\left(\frac{x}{\beta}\right)^q \Gamma\left(\frac{1}{\gamma} - q\right) \tag{A.16}$$

where $\beta > 0$ is scale, $0 < \gamma < 0.5$ is shape parameter, and Γ is gamma function.

Autocorrelation Structures

The following are parametric autocorrelation structures and not to be confused with distribution functions. They express how the autocorrelation decreases with time τ.

Weibull

$$\rho_{\mathrm{w}}(\tau; b, c) = \exp\left(-\left(\frac{\tau}{b}\right)^{c}\right) \tag{A.17}$$

with $b > 0$ and $c > 0$

Pareto type II distribution

$$\rho_{\mathrm{PII}}(\tau; b, c) = \left(1 + c\frac{\tau}{b}\right)^{-1/c} \tag{A.18}$$

with $b > 0$ and $c > 0$

Burr Type XII

$$\rho_{\mathrm{BrXII}}(\tau; b, c_1, c_2) = \left(1 + c_2\left(\frac{\tau}{b}\right)^{c_1}\right)^{\frac{-1}{c_1 c_2}} \tag{A.19}$$

with $b > 0$, $c_1 > 0$ and $c_2 > 0$

Generalized Logarithmic

$$\rho_{\mathrm{GL}}(\tau; b, c) = \left(1 + \ln\left(1 + c\frac{\tau}{b}\right)\right)^{-1/c} \tag{A.20}$$

with $b > 0$ and $c > 0$

Fractional Gaussian Noise

$$\rho_{\mathrm{fGn}}(\tau; \mathrm{H}) = \frac{1}{2}(|\tau - 1|^{2\mathrm{H}} - 2|\tau|^{2\mathrm{H}} + |\tau + 1|^{2\mathrm{H}}) \sim \tau^{-2(1-\mathrm{H})} \tag{A.21}$$

with $0 < H < 1$, H is the Hurst coefficient

Markovian

$$\rho_{\mathrm{M}}(\tau) = \rho_1^{\tau} \tag{A.22}$$

where $\rho_1 > 0$ is lag-1 correlation.