Supplement of Hydrol. Earth Syst. Sci., 29, 1–25, 2025
https://doi.org/10.5194/hess-29-1-2025-supplement

Hydrology and Earth System Sciences
Open Access

Supplement of

Modelling convective cell life cycles with a copula-based approach

Chien-Yu Tseng et al.

Correspondence to: Li-Pen Wang (lpwang@ntu.edu.tw)

Supplement

S1: Selection and fitting of the optimal Vine Copula model

This section details the procedure for selecting and fitting the optimal vine copula model used to represent the interdependencies between convective cell properties. Note that the selection process utilised here is based on the deterministic vine-copula structures, in which the 4D copula is D-Vine (2-3-1-4), while the 3D copulas are C-Vine (2-3-1). Here, we begin by outlining the three distinct fitting strategies under consideration and then elaborate on the rationale behind choosing the TLL (Transformation Local Likelihood) copula model as the final model based on both quantitative metrics and visual assessment.

We evaluated the following three strategies for constructing vine copula models:

- TLL (Non-parametric): This strategy leverages the flexibility of the non-parametric TLL family, as implemented in the `kdecopula` R package (Nagler, 2018), to estimate bivariate dependencies at each edge of the vine structure. This method is particularly suitable for capturing complex dependencies that may not be well-represented by standard parametric families.
- Parametric: In this strategy, we fit a range of parametric copula families to the data, including Gaussian, Student's t, Clayton, Gumbel, Frank, Joe, BB1, BB6, BB7, and BB8. The best-fitting parametric copula for each edge is selected based on goodness-of-fit criteria, primarily the Akaike Information Criterion (AIC).
- Combined TLL and Parametric: This strategy aims to harness the strengths of both TLL and parametric approaches by combining them within the vine structure. This allows for flexibility in modelling certain dependencies non-parametrically while potentially benefiting from the parsimony and interpretability of parametric copulas for other edges.

Here, we employed AIC (Akaike information criterion) to be the primary model selection criterion since AIC balances the goodness-of-fit with model complexity.

It is important to note that for non-parametric copulas like TLL, the complexity term in the AIC calculation is represented by effective degrees of freedom. These effective degrees of freedom are determined by the smoothing parameters (e.g., bandwidth) used in the kernel density estimation process (Nagler, 2018). The fitting process for TLL copulas involves transforming the input data for each variable to uniform margins on the interval

[0, 1] and then applying kernel density estimation to smooth the copula density in the transformed space. The smoothing parameters, particularly the bandwidth, are crucial for determining the accuracy of the estimation. In this work, we utilised the `pyvinecopulib` Python package for fitting TLL copulas, leveraging its cross-validation capabilities to optimise the bandwidth parameter and ensure a balance between bias and variance in the density estimation.

Table S1 summarises the fitting results, with bold text indicating the optimal strategy eventually chosen in this study for each copula model. Figures S1-S4 illustrate the pairwise correlation structures of distinct vine copula models (C_{peak}, C_{Imax}, C_{Smaj} and C_{Smin}) for each of the three fitting strategies. In each sub-figure, $\Delta\tau$ represents the difference in Kendall's τ between simulated and observed samples (i.e. $\Delta\tau = \tau_{sim} - \tau_{obs}$).

As can be seen, while the optimal copula family selected through visual inspection (Figures S1-S4) generally aligns with the AIC and log-likelihood values in Table S1, the C_{Smaj} model presents an exception. Specifically, although the combined TLL and Parametric strategy initially appeared to be the best strategy for C_{Smaj} model according to AIC and log-likelihood values, Figure S3 reveals inconsistent result. The TLL strategy provides a better visual match between observed and simulated data compared to the other two strategies. This is particularly evident in capturing tail dependencies, while the combined and purely parametric strategies exhibit a poorer fit, especially in the tails. After further investigation, we found that this is mainly caused by numerical instability during the process of fitting parametric models. Therefore, despite the initial AIC results, the TLL model was chosen for C_{Smaj} to prioritise the clear visual agreement between observed and simulated dependencies and ensure a more reliable representation of the underlying data structure.

This analysis suggests that selecting the optimal bivariate copula family should involve both numerical evaluation (AIC and log-likelihood) and visual inspection of the correlation distribution.

Table S1. Comparative evaluation of different copula models based on Akaike Information Criterion (AIC) and log-likelihood metrics.

Vine-copula model	Control of Bivariate Family	AIC	log-likelihood
C_{peak}	TLL	-37314.823	18912.031
	Parametric	-32815.206	16416.603
	Parametric and TLL	**-37362.077**	**18893.488**
C_{Imax}	**TLL**	**-40263.157**	**20269.259**
	Parametric	-29212.414	14611.207
	Parametric and TLL	-40262.747	20269.104
C_{Smaj}	**TLL**	**-46064.656**	**23172.394**
	Parametric	-92409.25	46210.625
	Parametric and TLL	-95783.66	47942.412
C_{Smin}	**TLL**	**-42428.612**	**21353.235**
	Parametric	-39033.233	19521.616
	Parametric and TLL	-42427.892	21352.897

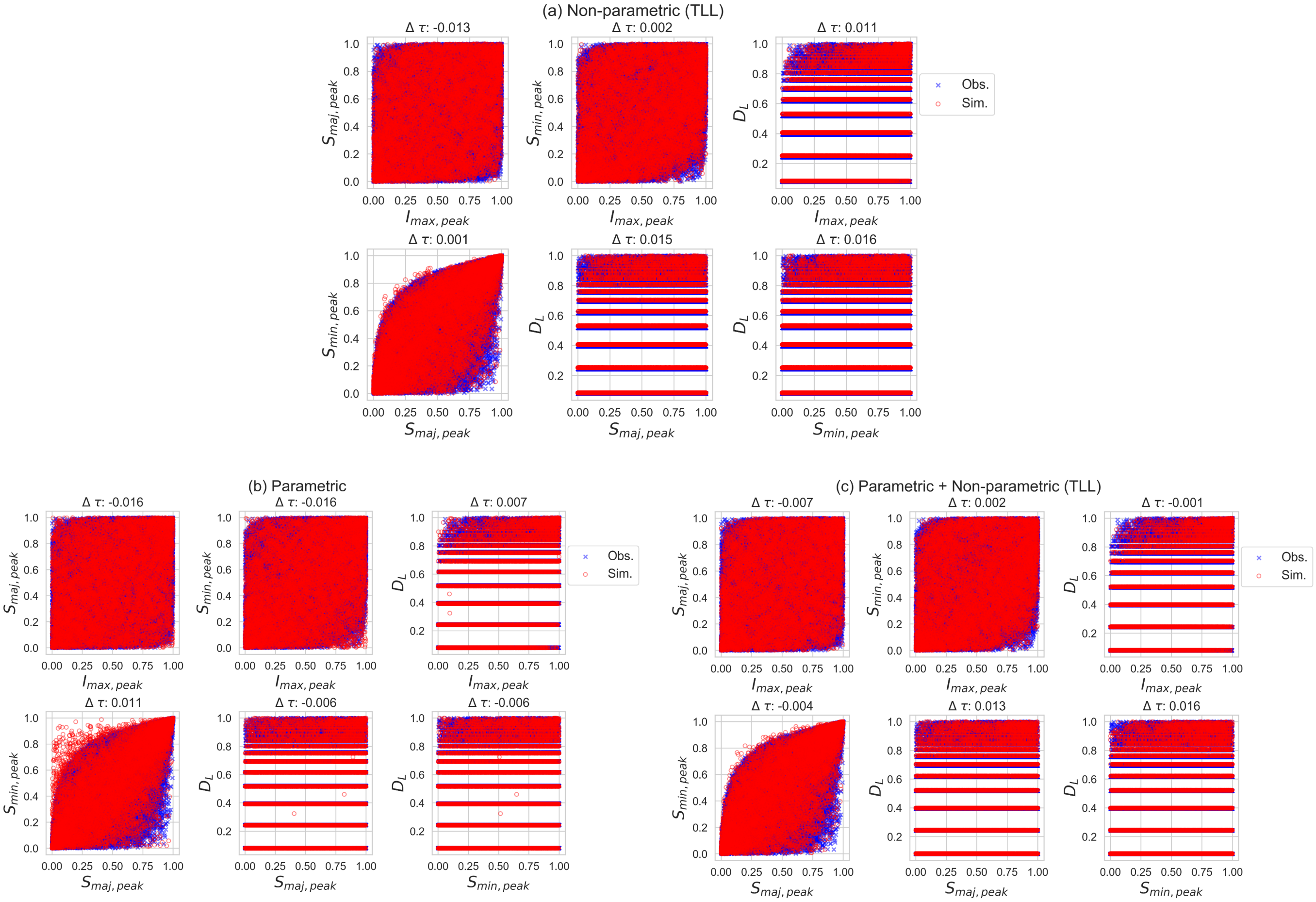


Figure S1. Visual inspection of the fitting results of parametric, non-parametric, and mixed copula models for C$_{peak}$.

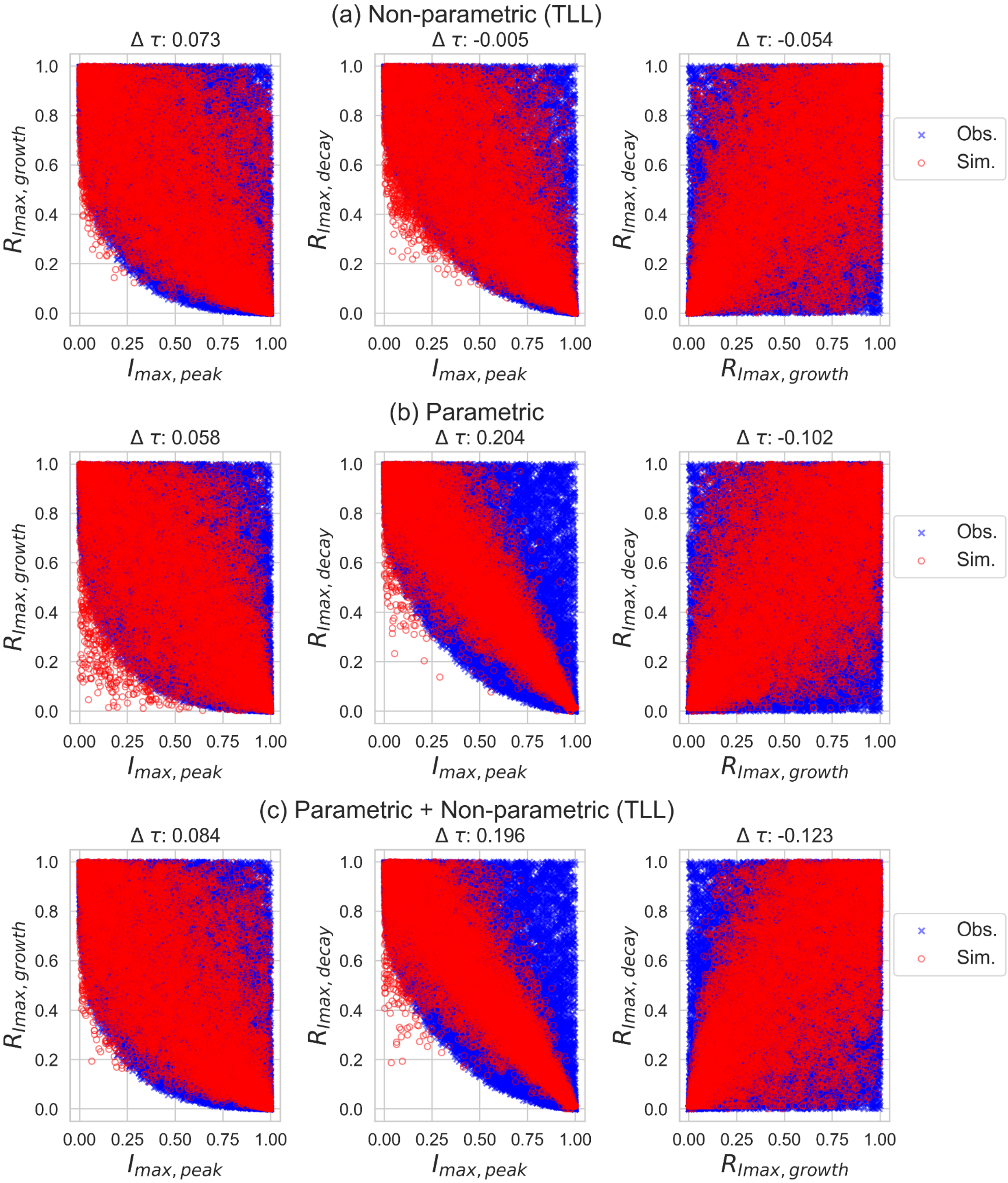
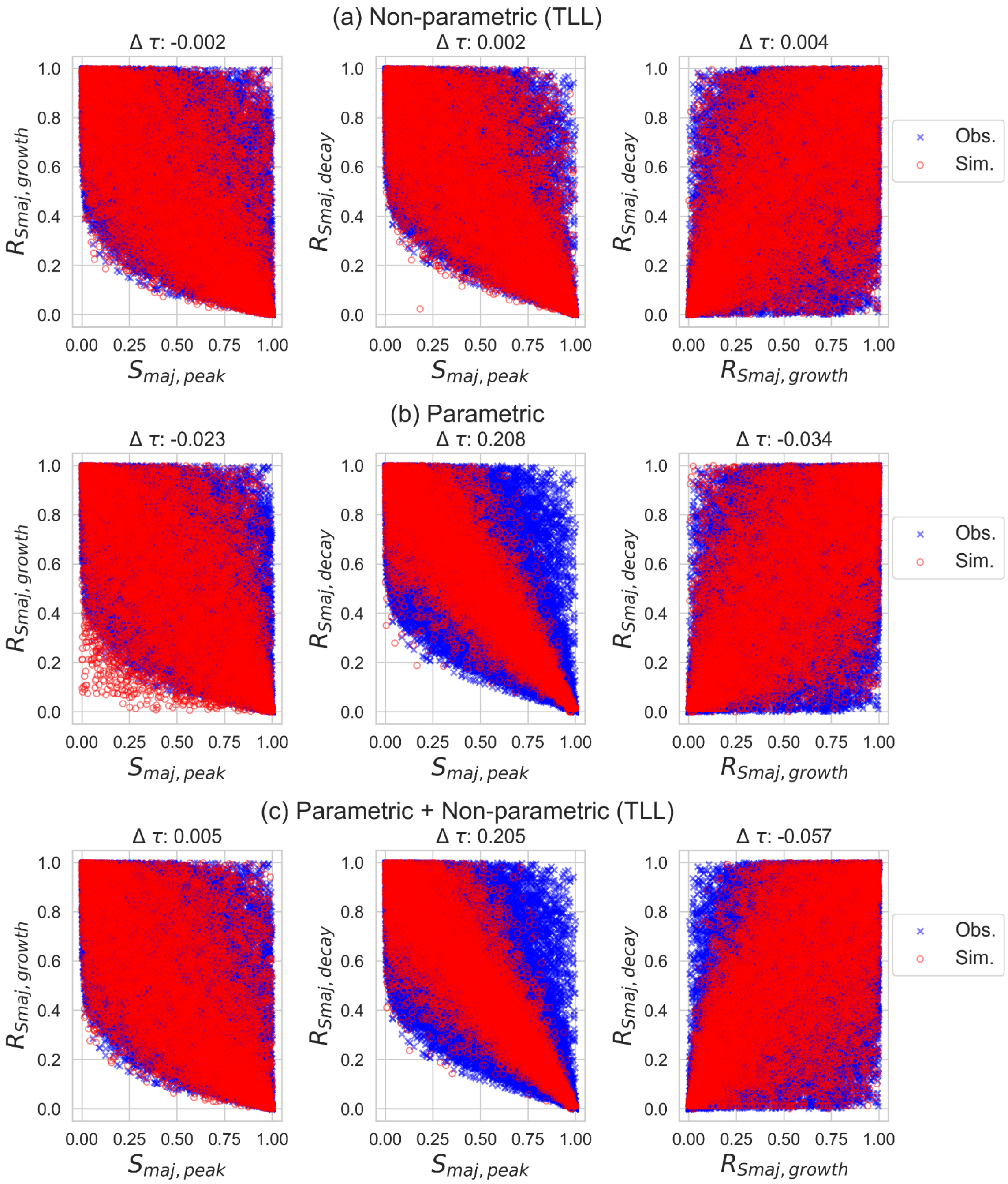


Figure S2. Visual inspection of the fitting results of parametric, non-parametric, and mixed copula models for C_{Imax}.



(a) Non-parametric (TLL)

(b) Parametric

(c) Parametric + Non-parametric (TLL)

Figure S3. Visual inspection of the fitting results of parametric, non-parametric, and mixed copula models for C_{Smaj}.

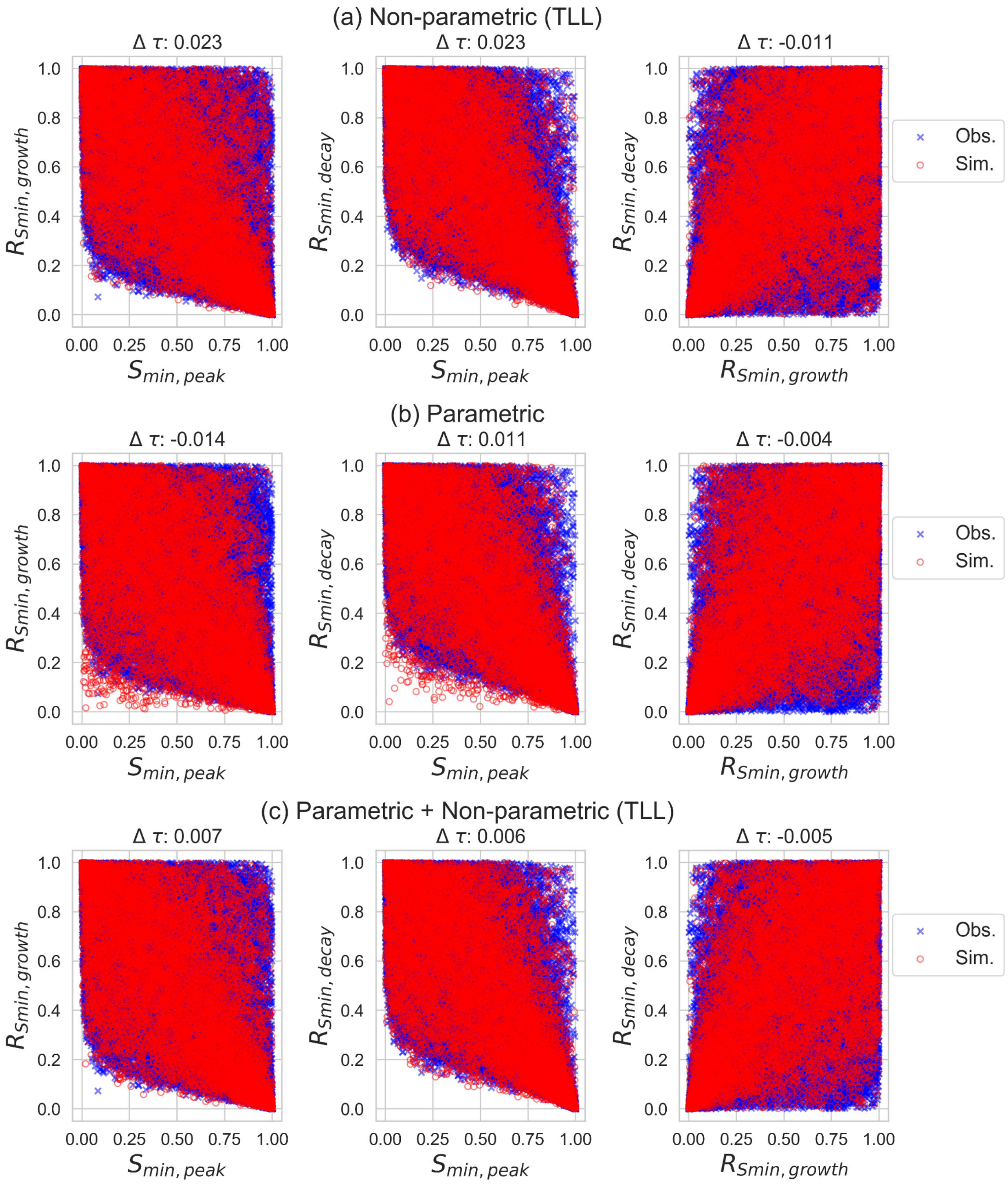


Figure S4. Visual inspection of the fitting results of parametric, non-parametric, and mixed copula models for C_{Smin}.

References.

Bedford, T. and Cooke, R. M.: Probability density decomposition for conditionally dependent random variables modeled by vines, Annals of Mathematics and Artificial Intelligence, 32(1-4), 245-268, https://doi.org/10.1023/A:1016725902970, 2001.

Nagler, T. (2018). kdecopula: An R Package for the Kernel Estimation of Bivariate Copula Densities. Journal of Statistical Software, 84(7), 1–22. https://doi.org/10.18637/jss.v084.i07

Nagler, T., & Vatter, T. (2023). pyvinecopulib (v0.6.4). Zenodo. https://doi.org/10.5281/zenodo.10435751