

A clustering framework for conditional extremes models: supplementary materials

Patrick O’Toole, Christian Rohrbeck, Jordan Richards

May 13, 2025

1 Simulation study

1.1 Gaussian copula

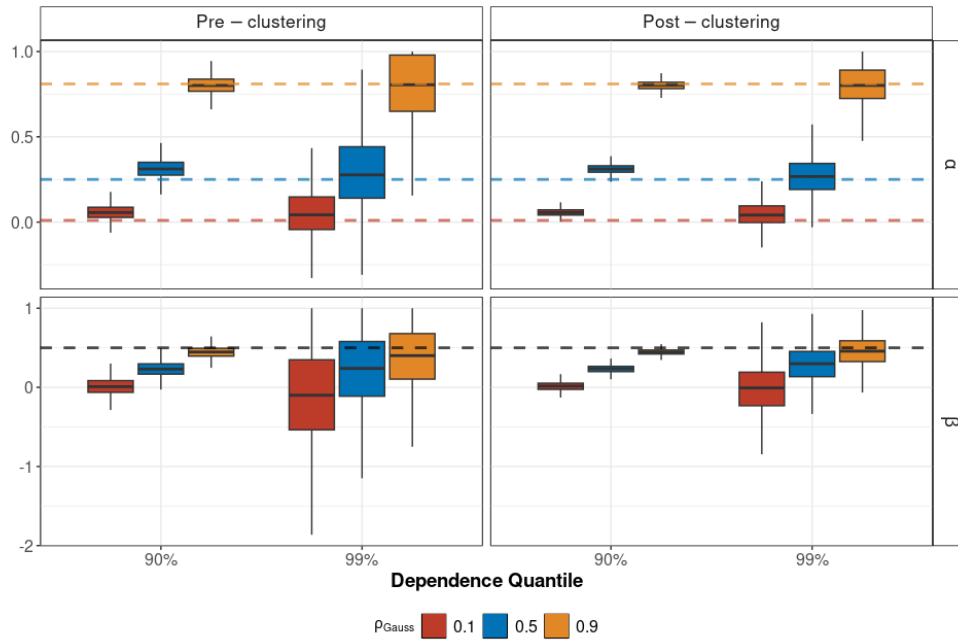


Figure 1: Boxplots of α and β parameter estimates for the conditional extremes model pre- and post-clustering for a bivariate Gaussian copula simulation of 12 datasets, or “locations”, each with 1000 observations of each variable, belonging equally to three known clusters. Simulations were repeated 500 times. The x-axis represents the conditional quantile used to fit the conditional extremes model, and the boxes are coloured by their Gaussian copula correlation parameter, ρ_{Gauss} . Dotted horizontal lines indicate the theoretical asymptotic values of α and β for a bivariate Gaussian copula, coloured by the same ρ_{Gauss} values for α and black for β to indicate the same value across all clusters, at $1/2$.

1.2 Mixture models

1.2.1 Comparison to competing methods

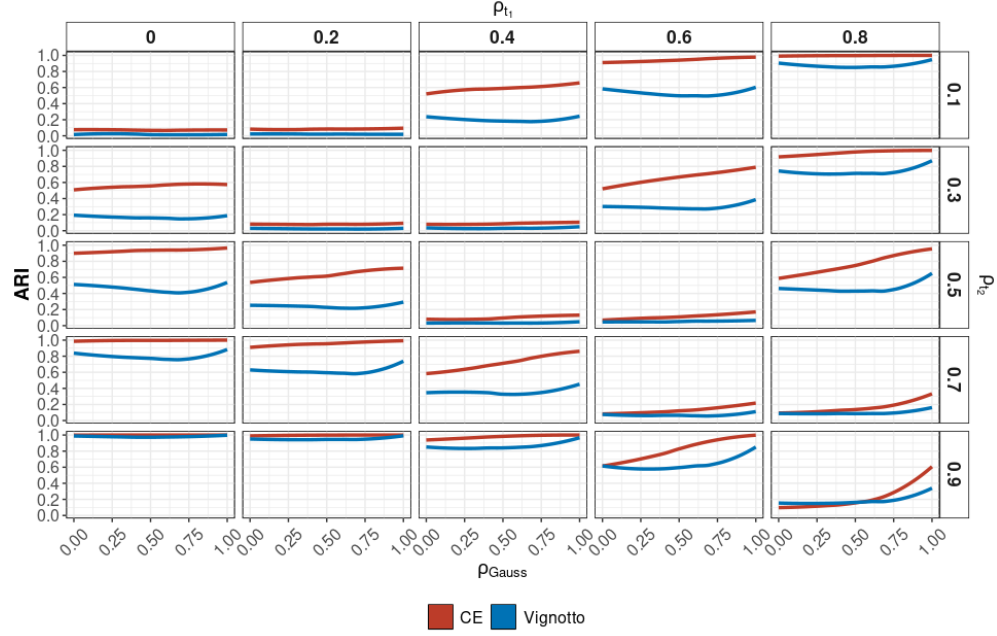


Figure 2: Comparison of clustering methods for two variables and two equally sized clusters for simulations of 12 “locations”, each with 1000 observations of each variable, from a mixture of Normal and Gaussian copulas. A grid search was performed, with the x-axis representing the Gaussian correlation parameter used for both clusters, and the facet labels showing the t -copula correlation parameters for each “known” cluster. This grid search was repeated 500 times. The lines show the smoothed Adjusted Rand Index for both clustering methods across all simulations, with associated uncertainty.

1.2.2 Extension to > 2 dimensions

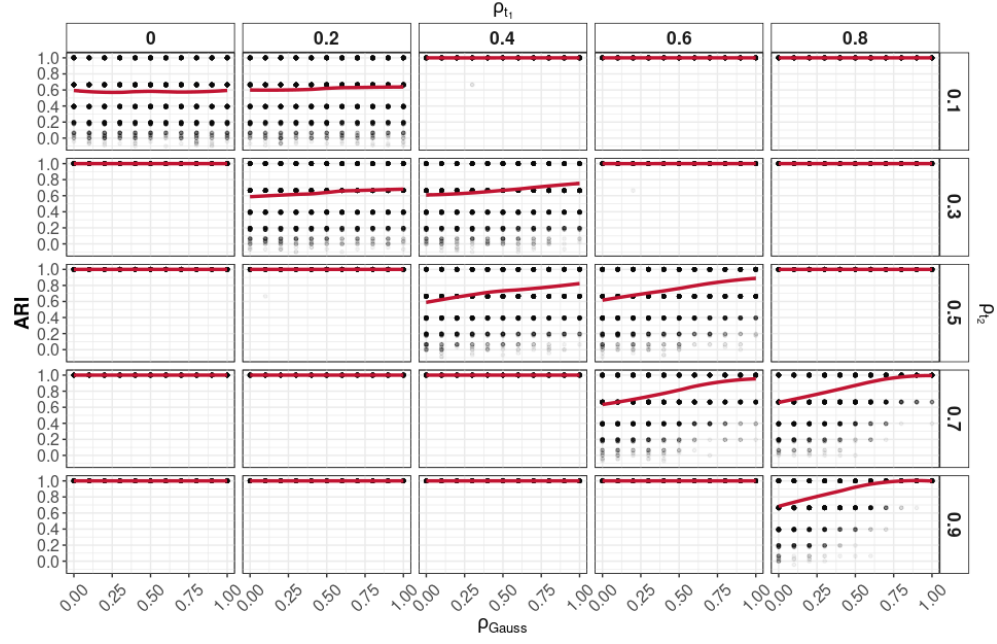


Figure 3: *Evaluation of clustering performance for three variables and two equally sized clusters for simulations of 12 “locations”, each with 1000 observations for each variable, from a mixture of Normal and Gaussian copulas. A grid search was performed, with the x-axis representing the Gaussian correlation parameter used for both clusters, and the facet labels showing the t-copula correlation parameters for each “known” cluster. This grid search was repeated 500 times. The lines show the smoothed Adjusted Rand Index for both clustering methods across all simulations, with associated uncertainty. Points show individual values of the ARI for a given simulation.*

1.2.3 More realistic example

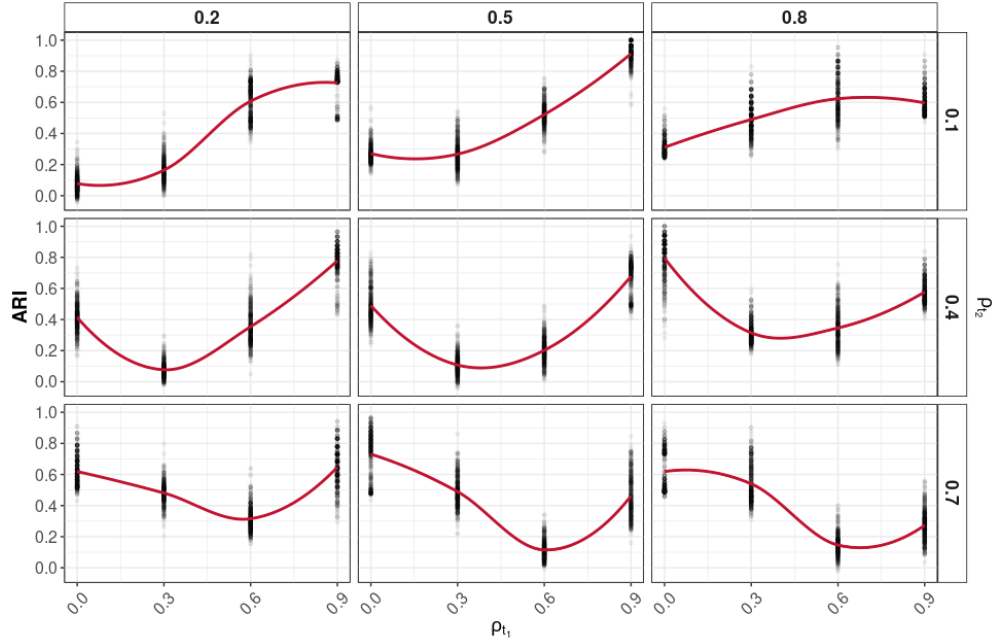


Figure 4: *Evaluation of clustering performance for two variables and three clusters for simulations from a mixture of Normal and Gaussian copulas in a more realistic clustering setting. 60 locations with 1000 observations for each variable were simulated, with 10, 20 and 30 belonging to each respective cluster. A grid search was performed, with the x-axis and facet labels showing the t-copula correlation parameters for each of the three “known” clusters, for a Gaussian copula correlation of 0.5. Perturbations were added to these parameters. This grid search was repeated 500 times. The lines show the smoothed Adjusted Rand Index for both clustering methods across all simulations, with associated uncertainty. Points show individual values of the ARI for a given simulation.*

1.2.4 Parameter estimation pre- and post-clustering

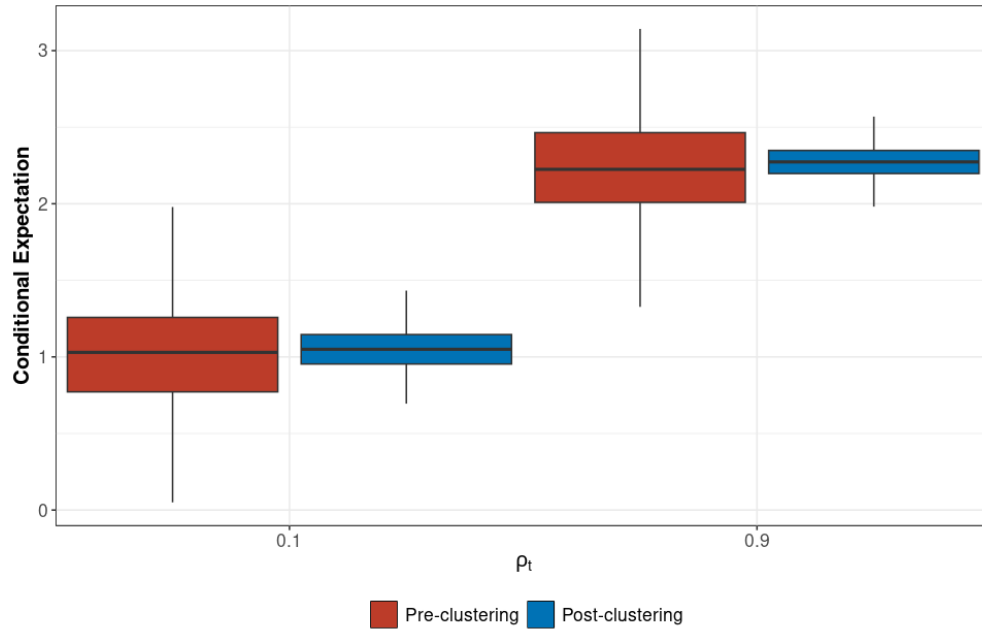


Figure 5: *Boxplots of 500 bootstrapped conditional expectations of one variable given the other is at the 98% marginal quantile u , pre- and post-clustering for a simulation of 12 “locations”, each with 1000 observations for each variable, from a mixture of Normal and Gaussian copulas, with the Gaussian correlation set to 0.5. The t -copula correlation parameters for the two clusters were set to 0.1 and 0.9, as shown on the x -axis.*

2 Application to Irish meteorological data

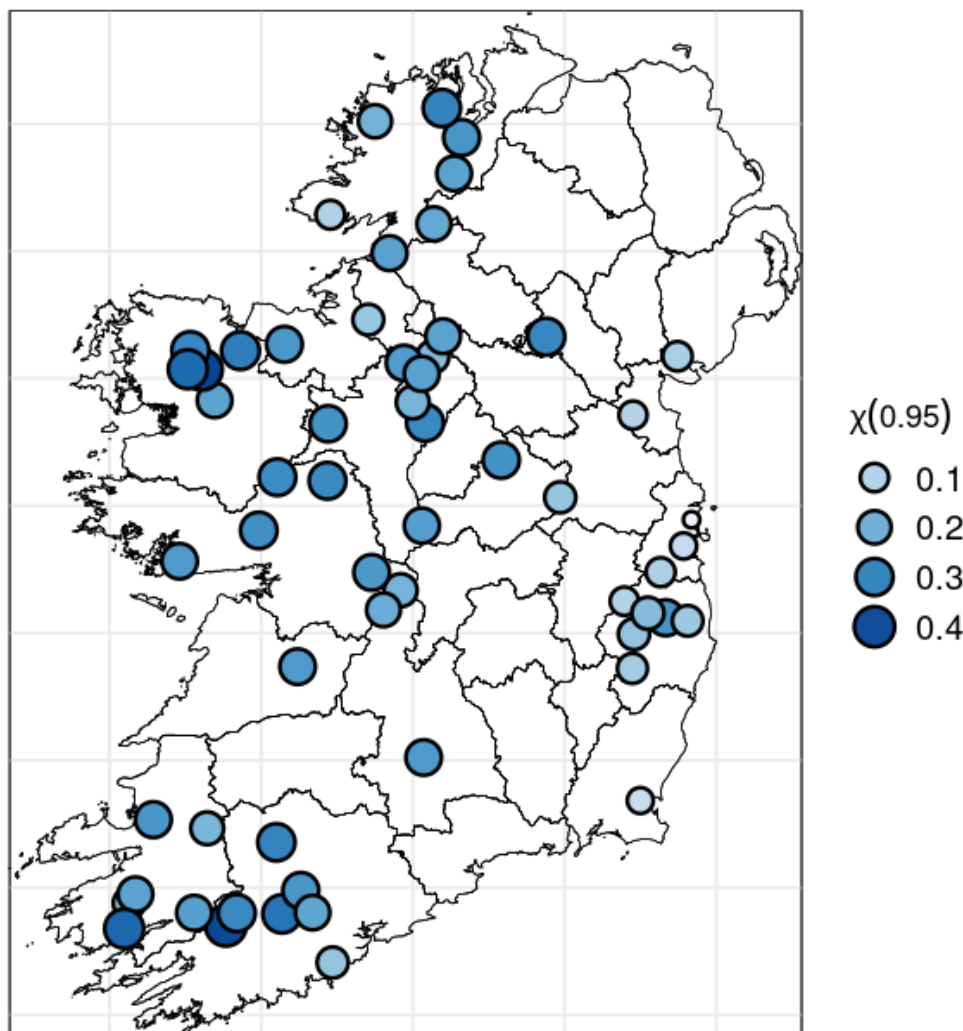


Figure 6: χ estimates for rain and wind speed at the 95th quantile across all 59 Met Éireann weather sites in Ireland.

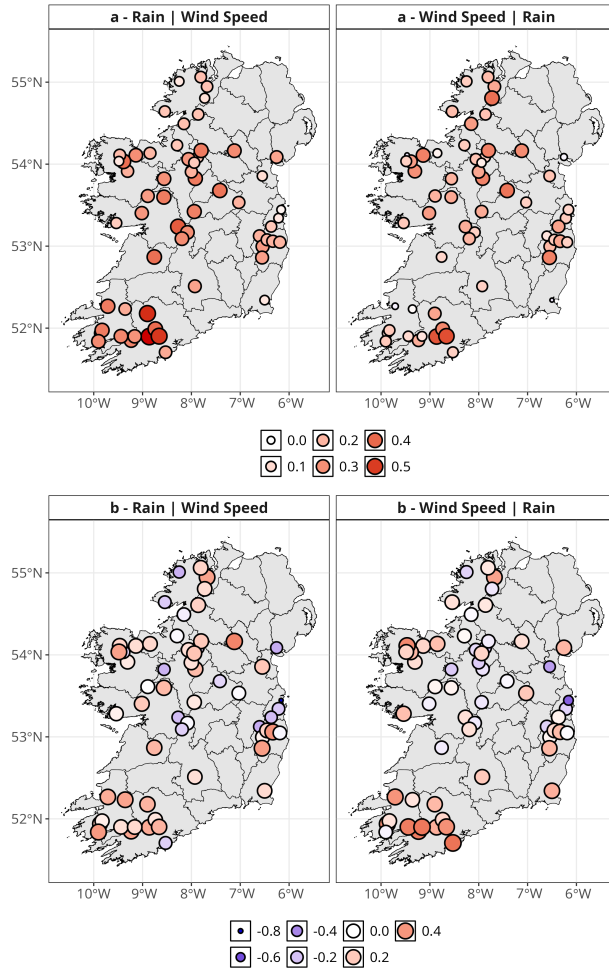


Figure 7: Estimates for the conditional extremes model α and β parameters for the 59 Met Éireann weather sites in Ireland, for the model of rain given extreme wind speed, and vice versa.

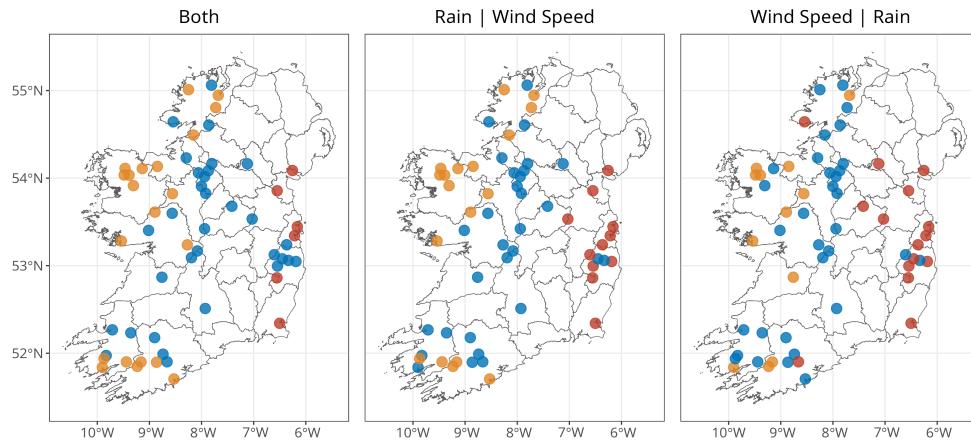


Figure 8: *Clustering solution based on the conditional extremes model fits at each of the 59 Met Éireann weather sites in Ireland, for the model of rain given extreme wind speed, vice versa and combined.*

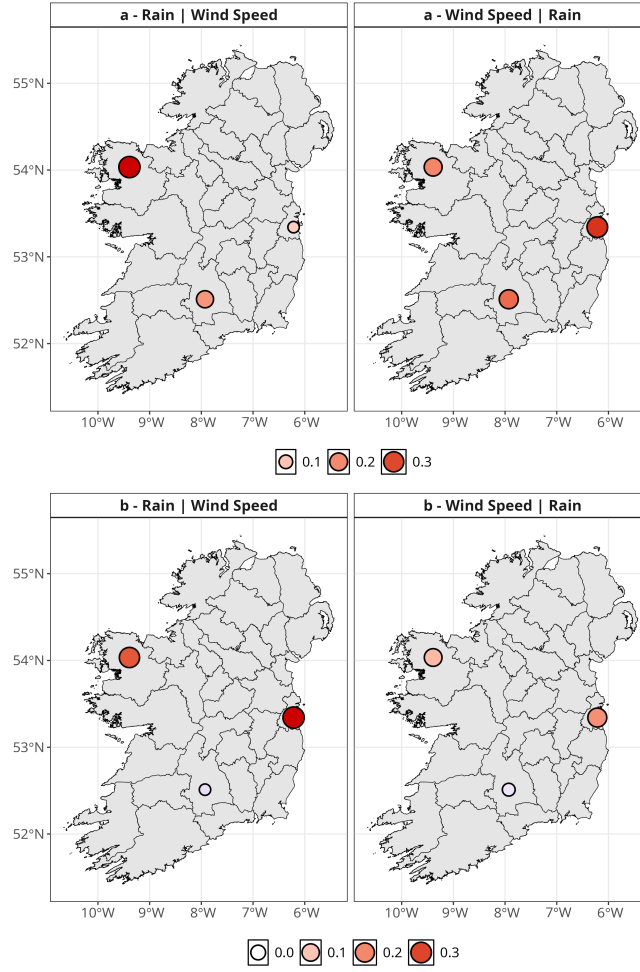


Figure 9: *Post-clustering estimates for the conditional extremes model α and β parameters for the 59 Met Éireann weather sites in Ireland, for the model of rain given extreme wind speed, and vice versa. Data was re-assigned to their respective cluster medoids, aggregating rainfall and taking the maximum wind speed for a given date across all members of each cluster.*