Supplement to: A temperature-conditioned spatial-temporal storm generator with evolving convective cell life-cycles

Chien-Yu Tseng[1], Li-Pen Wang[1,2], and Christian Onof[2]

[1]Department of Civil Engineering, National Taiwan University, Taipei, 106319, Taiwan
[2]Department of Civil and Environmental Engineering, Imperial College London, London, SW7 2AZ, United Kingdom

Correspondence: Chien-Yu Tseng (cy12tseng@caece.net) and Li-Pen Wang (lpwang@ntu.edu.tw)

This supplement provides additional implementation and verification details to the proposed generator described in the main manuscript. It is organised according to modelling roles: the GAMLSS intensity module, event-scale inputs, temporal clustering, validation evidence, lifecycle dependence and rainfall-field rendering.

S1 Marginal distributions and temperature-conditioned intensity specification

S1.1 Marginal distribution fitting

Figure S1 shows representative goodness-of-fit results for the marginal families selected for the storm-scale quantities and for the cell-property intensities and axes by cell-path type. During storm generation, the base reference-state intensity marginal is replaced by the storm-specific temperature-conditioned distribution described below.

S1.2 Temperature-conditioned intensity fitting protocol

The storm-specific intensity models are fitted using GAMLSS, with a fixed scale random-effect penalty $\lambda_\sigma = 1$ in the `gamlss` call (Stasinopoulos and Rigby, 2007),

```
sigma.formula = ~ pb(T) + random(event_id, lambda = 1).
```

The scale random-effect penalty was fixed at $\lambda_\sigma = 1$ to stabilise the estimation of storm-to-storm variation in within-event spread. Exploratory fits using the package-default penalty produced stronger shrinkage of this variation, especially for the short-lived path types, and more tightly clustered simulated event maxima.

The production fitting protocol is shared across path types:

- maximum-likelihood estimation under a truncated likelihood,
- penalised B-spline smooths of temperature,
- a bivariate event-level random effect on the location and scale-related parameters,
- fixed optimisation control settings (`n.cyc = 500`, `mu.step = 0.3`, `sigma.step = 0.3`, `c.crit = 1e-4`).

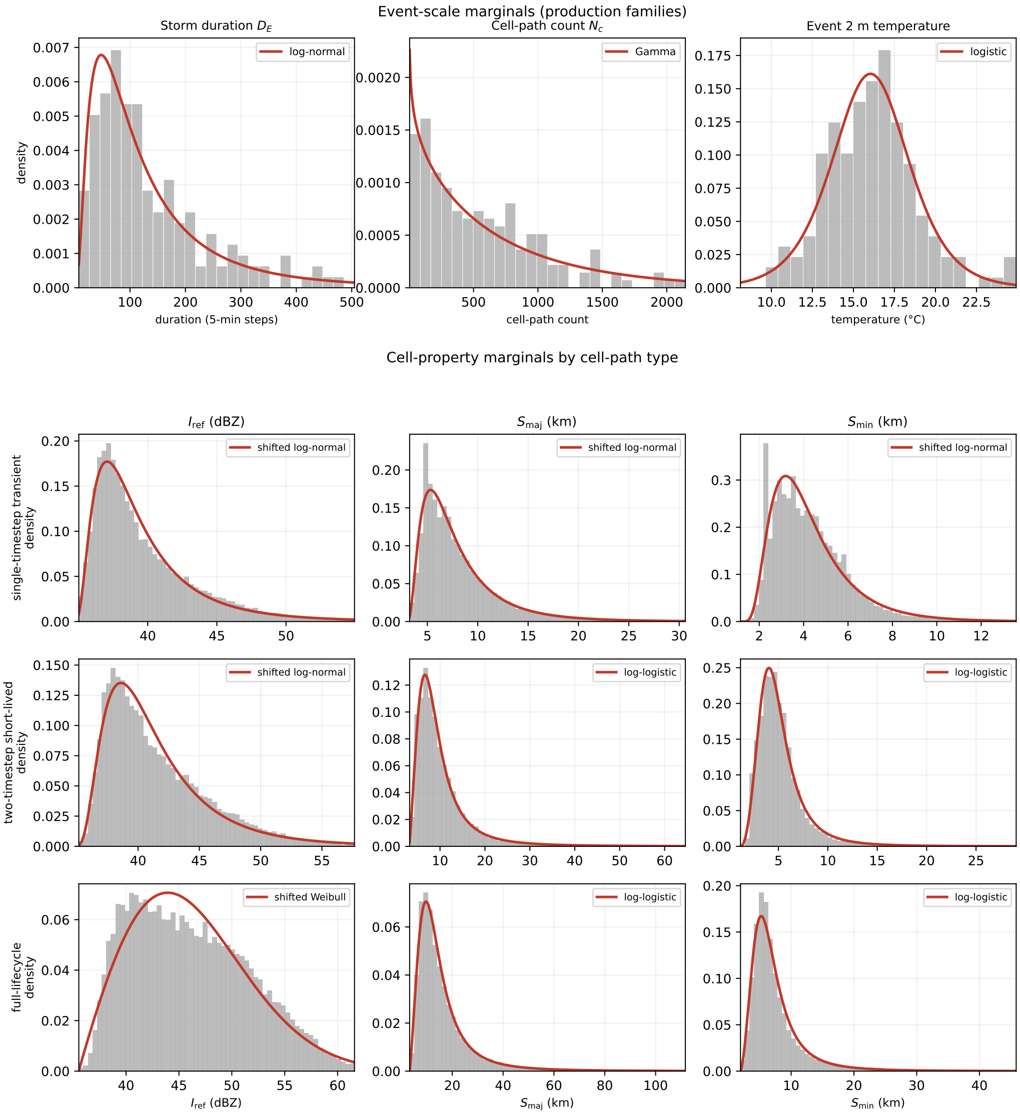

Figure S1. Selected marginal fits used by the generator. The top row shows storm-scale quantities and the lower grid shows the reference-state intensity, S_{maj} and S_{min} by cell-path type. Histograms show the observed distributions, and solid curves show the selected fitted families.

At simulation time, the generator evaluates the fitted penalised-spline temperature smooth directly through a tabulated lookup over the observed temperature range. The intercept-and-slope coefficients reported below provide a compact linear summary of this smooth and are retained as auxiliary information. Reproducing the published simulator output therefore requires the tabulated smooth used by the model implementation rather than the linear summary alone.

S1.3 Distribution selection and production parameters

The full-lifecycle path type is fitted with a left-truncated Weibull type 3 family (`WEI3tr`) at a lower threshold of 35.528 dBZ (Stasinopoulos et al., 2017). The model estimates the location and shape parameters, while the threshold is fixed by the radar cell-extraction floor. The truncation-aware likelihood is required because fitting an untruncated family to observations that have already been truncated biases the fitted distribution parameters and can lead to under-dispersed simulated event maxima when threshold-based sampling is subsequently applied.

The fitted temperature smooth is summarised by the following linearised coefficients (see Table S1); the simulator itself evaluates the full smooth described in Sect. S1.2.

Table S1. Production parameters of the temperature-conditioned intensity model for full-lifecycle paths used by the simulator.

Quantity	Value
μ intercept	34.5735
μ slope	0.62755
$\log(\text{shape})$ intercept	1.9428
$\log(\text{shape})$ slope	0.01265
Threshold (dBZ)	35.528

The residual covariance of the event-level random effects, relative to the fitted temperature smooth, is

$$\mathbf{\Sigma}_{\text{full}} = \begin{pmatrix} 11.304484 & 0.395428 \\ 0.395428 & 0.044995 \end{pmatrix}, \qquad r = +0.554.$$

In this parametrisation, a larger Weibull shape corresponds to a narrower within-event spread. The positive correlation therefore indicates that storms with larger event-level location deviations also tend to have more concentrated within-event intensity distributions.

The single-time-step transient and two-time-step short-lived path types are modelled using left-truncated shifted log-normal families (`LOGNOtr`). Cells at or above the radar encoding ceiling are removed before fitting so that the estimated parameters represent the cell population below the encoding limit rather than the saturation artefact. The fitted production parameters are summarised in Table S2.

Table S2. Production parameters of the temperature-conditioned intensity model for single-time-step transient and two-time-step short-lived paths used by the simulator.

Path type	μ intercept	μ slope	$\log(\sigma)$ intercept	$\log(\sigma)$ slope
single-time-step transient	3.5350	0.00809	-2.8758	0.03077
two-time-step short-lived	3.5743	0.00725	-2.9480	0.03491

The corresponding event-level residual covariance matrices, relative to the fitted temperature smooth, are

$$\boldsymbol{\Sigma}_{\text{single}} = \begin{pmatrix} 0.002577 & -0.001073 \\ -0.001073 & 0.062261 \end{pmatrix}, \qquad r = -0.085,$$

$$\boldsymbol{\Sigma}_{\text{two}} = \begin{pmatrix} 0.001431 & 0.001260 \\ 0.001260 & 0.063609 \end{pmatrix}, \qquad r = +0.132.$$

Multiple alternative families were explored for both short-lived path types. The shifted log-normal family was retained because it provided stable convergence under truncation and avoided boundary parameter values encountered by several alternatives under the same event-level random-effect specification (Stasinopoulos et al., 2017).

The linearised coefficients above summarise the fitted temperature smooth over the observed range. At simulation time, the generator evaluates the full smooth, while the coefficients are retained only as an auxiliary linear summary. The fitted smooth curves are retained in the fitting scripts and saved model objects.

S1.4 Integration with the cell-property copulas

All three cell-path types use a common GAMLSS–copula integration pattern:

1. sample event temperature;
2. evaluate the fitted GAMLSS parameters for the realised event;
3. sample reference-state intensities from the type-specific truncated family;
4. rank-match the sampled intensities to the copula-generated order;
5. condition the remaining lifecycle properties through the corresponding copula component.

This design replaces only the reference-state intensity marginal while continuing to use the fitted copula model to couple the reference-state intensity with the remaining lifecycle properties. The rank-matching step is therefore not an additional model layer; it is the coupling point that lets the temperature-conditional intensity population enter the existing generator (Tseng et al., 2025).

For the full-lifecycle type, lifecycle placement is resolved before reconstruction: once the temperature-conditioned peak intensity has been drawn, the two axes and pathLength are obtained jointly from the same conditional inverse-Rosenblatt draw, and the resulting pathLength is used to place the cell on the time grid. Thus, the peak–axis and peak–pathLength associations arise from the joint conditional generation rather than from a post hoc rank match. For the short-lived types, the rank-matched substitution is used to update the reference-state intensity values before the path-specific properties are reconstructed.

S1.5 Fitting diagnostics

The diagnostics below were calculated directly from the fitted objects used to generate the production ensemble, rather than reconstructed from rounded manuscript constants. All three fits converged under the common RS (Rigby–Stasinopoulos) algorithm with the optimisation controls listed above. The reported AIC, Schwarz Bayesian criterion (SBC) and global deviance values support comparisons within a path type on identical data; they are not comparable across path types because the response samples, truncation thresholds and selected families differ. Family selection therefore used within-type convergence, parameter stability and residual diagnostics. SMEV is reported separately as a downstream evaluation of return-level behaviour and was not used in fitting or family selection.

Table S3. Production GAMLSS fitting diagnostics. Total effective degrees of freedom (edf) are dominated by the event-level random effects and should not be interpreted as the complexity of the temperature smooth alone.

Path type	Converged	Iter.	edf_μ	edf_σ	G-deviance	AIC	SBC
single-time-step transient	yes	89	151.4	151.6	265828	266434	269134
two-time-step short-lived	yes	25	109.8	112.6	91915	92360	94093
full-lifecycle	yes	110	129.6	130.6	158549	159069	161201

Total edf values are approximately the number of event-level effects on each predictor. The temperature $\mathrm{pb}(T)$ contribution is small, so edf does not indicate a highly complex temperature response.

The production simulation uses the covariance relative to the fitted penalised-spline smooth: $r = -0.085$, $+0.132$ and $+0.554$ for the single-time-step transient, two-time-step short-lived and full-lifecycle types, respectively. The covariance relative to a linearised temperature trend is retained as a secondary validation quantity and is not used when the penalised-spline lookup is enabled.

Table S4. Randomised normalised quantile-residual diagnostics for the production GAMLSS fits.

Path type	Mean	Variance	Skewness	Kurtosis	Filliben r
single-time-step transient	0.033	0.859	0.957	4.305	0.9749
two-time-step short-lived	0.023	0.899	0.856	3.717	0.9772
full-lifecycle	0.031	0.858	0.528	3.250	0.9916

The residual diagnostics are close to the reference normal pattern in their means and Filliben correlations, with mild positive skewness in all three types and slight excess kurtosis for the single-time-step transient type. The residual variances are mildly below unity. These diagnostics support the adequacy of the fitted conditional distributions for the intended simulations, but do not establish perfect tail calibration; held-out cross-validation below and the multi-return-period SMEV assessment in the main text remain the relevant downstream checks.

The per-event dispersion values underlying the main-text GAMLSS assessment are listed below. They are provided as fitting-output detail because the main text reports the corresponding pattern in Fig. 11 rather than repeating the values for each path type.

Table S5. Within-event standard deviation σ_{ev} of reference-state intensity (dBZ): observed versus GAMLSS-fitted, averaged over the calibration events.

Cell-path type	Family	Obs σ_{ev} (dBZ)	GAMLSS σ_{ev} (dBZ)
single-time-step transient	shifted log-normal, left-truncated	3.44	3.30
two-time-step short-lived	shifted log-normal, left-truncated	3.64	3.54
full-lifecycle	truncated Weibull type 3 (WEI3)	4.74	4.94

S2 Storm-scale inputs and cell-path allocation

S2.1 Lifespan-based cell-path representations

The supplementary analyses use the three lifespan-based path representations defined in Sect. 2.4 of the main manuscript:

- **Full-lifecycle paths**: at least three 5-min observations, with a resolved growth–peak–decay sequence.
- **Two-time-step short-lived paths**: exactly one observable transition.
- **Single-time-step transient paths**: no resolved temporal transition.

The observed counts are 27,730 full-lifecycle paths, 19,252 two-time-step short-lived paths and 55,314 single-time-step transient paths. These classes discretise a continuous longevity–vigour gradient inferred from radar tracking and describe the lifecycle information resolvable at the 5-min sampling interval; they are not interpreted as distinct convective regimes.

S2.2 Storm-level temperature covariate

The event-level temperature covariate is derived from ERA5-Land 2 m air temperature. For each radar event, the hourly temperature field is averaged over the approximate 250 km $\times$ 250 km extraction domain centred on the pilot site. The event window is rounded to whole hours by extending the start downward and the end upward to the nearest hour, and the maximum of the resulting hourly spatial means is retained as the event-level covariate T_e. Temperatures are retrieved in kelvin and converted to degrees Celsius for fitting and reporting.

This event-centred definition supplies the single storm-scale covariate used by the GAMLSS module to represent systematic between-event variation in the cell-intensity population.

For simulation, the event-level temperature covariate is sampled from a logistic distribution fitted to the calibration-event values. The fitted location is $16.041\,^{\circ}$C and the fitted scale is $1.551\,^{\circ}$C. These parameters define the Birmingham temperature marginal used when no externally supplied scenario covariate is provided.

S2.3 Logistic-normal cell-path allocation

The event-level allocation of cell paths among the three lifespan-based representations is treated as compositional data. For compact notation, PL1 denotes single-time-step transient paths, PL2 denotes two-time-step short-lived paths and PL3 denotes full-lifecycle paths. Taking PL2 as the reference component, the additive log-ratio coordinates are modelled as

$$\mathbf{y}_{e'} = \left(\log\frac{p_{e',\mathrm{PL3}}}{p_{e',\mathrm{PL2}}}, \log\frac{p_{e',\mathrm{PL1}}}{p_{e',\mathrm{PL2}}}\right)^{\mathsf{T}} \sim \mathcal{N}\left(\begin{pmatrix}0.418151\\1.042250\end{pmatrix}, \begin{pmatrix}0.084129 & -0.010667\\-0.010667 & 0.060483\end{pmatrix}\right).$$

The proportions are recovered by the inverse ALR (softmax) transformation, and integer counts are then drawn from a multinomial distribution conditional on the event total. The covariance was estimated after subtracting the expected multinomial sampling covariance from the observed event-level composition covariance and mapping the result to ALR coordinates. This prevents finite-cell-count noise from being counted a second time during simulation.

In observation-sized ensemble checks, the simulated standard deviations of the full-lifecycle, two-time-step short-lived and single-time-step transient proportions were 0.080, 0.039 and 0.086, compared with observed values of 0.081, 0.039 and 0.088, respectively. For the pairs (full-lifecycle, short-lived), (full-lifecycle, transient) and (short-lived, transient), the simulated correlations were -0.38, -0.89 and -0.08, compared with observed values of approximately -0.40, -0.90 and -0.05, respectively. The logistic-normal therefore reproduces both the unequal event-to-event spread and the observed composition dependence that could not be represented by the single-concentration Dirichlet specification.

S3 Temporal clustering of cell-path arrivals

Cell-path arrivals are generated through independent per-type latent Gaussian AR(1) processes mapped to negative-binomial marginals (Onof and Wang, 2020). The latent process is

$$z_t = \phi z_{t-1} + \sqrt{1-\phi^2}\,\varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0,1),$$

with `norm.cdf` used to map the latent sequence to uniform variates before inverse-transform sampling from the negative binomial distribution. The calibrated value is $\phi = 0.55$.

This construction preserves the marginal count distribution while recovering the observed temporal clustering of peak arrivals. In the validation run, the observed lag-1 autocorrelation of the active-count series is 0.396 and the simulated value with AR(1) enabled is 0.393. The active-count Perkins Skill Score increases from 0.84 to 0.86.

An alternative common-factor AR(1) variant was tested but was discarded because it synchronised activity patterns across path types and reduced the active-count skill score. The independent-per-type structure is therefore retained in the manuscript.

As an independent consistency check, the lag-1 autocorrelation was recomputed from the final $M = 100$ ensemble used downstream. The ensemble median and mean were both 0.391, with a 5–95% range of 0.376–0.407, containing the observed value of 0.396. The statistic was calculated as the mean of event-level lag-1 autocorrelations, with events treated separately.

S4 Additional validation of storm-to-storm intensity variability

S4.1 Held-out event cross-validation

The event-level cross-validation uses held-out events, not held-out cells. Events are partitioned into five folds; the GAMLSS component is refitted on four folds and then evaluated on the held-out fold using the observed temperature and the fitted event-level random-effect distribution from the training folds. Upstream generator components are kept fixed. This checks whether the conditional-intensity substitution generalises beyond the calibration events; it does not refit the full storm generator fold by fold.

Table S6. Held-out cross-validation summary for the GAMLSS substitution. Values are absolute differences between the held-out and in-sample SMEV residuals at the N=100 evaluation setting.

Path type	full-lifecycle	single-time-step transient	two-time-step short-lived
$\lvert\Delta\rvert$ (pp)	1.1	1.3	1.3

These small differences provide evidence that the conditional-intensity response is retained when evaluated on held-out events.

S4.2 Observed storm-to-storm intensity variability

For each cell-path type, observed cells were grouped by storm and the raw response-scale variance was decomposed as

$$SS_{\text{between}} = \sum_{e} n_e(\bar{I}_e - \bar{I})^2, \qquad SS_{\text{within}} = \sum_{e}\sum_{c}(I_{e,c} - \bar{I}_e)^2, \tag{1}$$

with $f_{\text{between}} = SS_{\text{between}}/(SS_{\text{between}} + SS_{\text{within}})$. We also report the matching one-way random-effects ICC, using method-of-moments variance components. Storm means were regressed on the fitted production temperature smooth by weighted least squares with weights n_e; the temperature fraction is the resulting reduction in cell-weighted between-storm variance. These are raw-response-scale quantities and must not be identified with the GAMLSS random-effect covariance on the family-specific predictor scale.

Table S7. Observed between- and within-storm variance of reference-state intensity on the raw response scale. The temperature column is the fraction of between-storm variance removed by the fitted temperature smooth, and the final column is the residual between-storm fraction after that adjustment.

Path type	Events	Cells	σ_B^2	σ_W^2	f_B	temp. frac. / f_B^T
single-time-step transient	150	54 776	2.83	12.55	18.5%	19.0% / 15.5%
two-time-step short-lived	111	17 942	2.72	13.88	16.7%	20.9% / 13.7%
full-lifecycle	129	26 737	7.77	24.42	24.3%	17.9% / 20.9%

Across the production calibration set, the between-storm fraction is 16.7–24.3% and the ICC agrees with it within 0.5 percentage points. The fitted temperature smooth accounts for approximately one-fifth of the between-storm component under the production map; sensitivity to the functional form gives approximately 13–22%. The residual component is therefore substantial and motivates the event-level random effects in addition to temperature conditioning.

S4.3 Simulated storm-to-storm intensity variability

The observed decomposition was repeated on each member of the 100-member, 13-year ensemble, using the same event grouping as the observational decomposition. The copula-only values were taken before the GAMLSS substitution, whereas the full-pipeline values were taken after it. The copula-only between-storm fraction was 0.2–0.6%, close to the random-grouping expectation $(E-1)/(N-1)$, and its event-mean standard deviation collapsed to 0.28–0.47 dBZ. The full GAMLSS medians were 21.4%, 17.4% and 25.7% for the between-storm fraction, with 5–95% intervals of 17.7–26.2%, 13.9–21.5% and 20.4–30.9% for the three path types. The observed fractions (18.5%, 16.7% and 24.3%) lie inside all three intervals.

The same inclusion result holds for the event-mean standard deviation and ICC. This comparison is a round-trip integration test: the random-effect variance was estimated from the observed calibration population and then recovered after simulation. It indicates that the GAMLSS module reinstates the missing storm-level spread on the raw response scale; it is not an out-of-sample validation.

S4.4 Temperature-intensity relationship

Temperature-gradient validation was performed per ensemble member rather than on a pooled collection of simulated cells. Each member contains approximately the observed number of calibration events, so the ensemble 5–95% interval represents the sampling variability relevant to an observation-sized record. For event-mean cell intensity, the observed slopes are 0.29, 0.31 and 0.53 dBZ $^{\circ}\mathrm{C}^{-1}$ and the ensemble medians are 0.31, 0.31 and 0.50 for the single-time-step transient, two-time-step short-lived and full-lifecycle types, respectively; all observed values lie inside the corresponding intervals. For within-event standard deviation, the observed and simulated slopes are both $+0.16$ for the two short-lived types, while the full-lifecycle values are $+0.13$ and $+0.09$.

Event-maximum intensity is an emergent quantity obtained by taking the maximum of the cell intensities generated within each event. Its distribution combines the effects of the per-cell intensity marginal, event cell counts and between-storm variation, so it is treated as a descriptive diagnostic and not as a separately fitted response. The predictor-scale GAMLSS random-effect covariance and the raw-scale ICC reported in this section are distinct quantities.

S4.5 Event-maximum distributional comparison

The event-maximum comparison complements the SMEV analysis by showing how the cell-intensity distribution appears after taking the maximum within each event. The main text presents the three path-type comparisons and a pooled aggregate view as empirical cumulative distribution functions. Figure S2 isolates the pooled comparison and gives its Wasserstein summary;

the type-specific values are reported in Table S8. The event maximum is an emergent quantity combining the cell-intensity marginal, event cell count and between-storm variation; return-level behaviour is quantified separately with SMEV below.

To provide a bin-free numerical summary, the Wasserstein-1 distance was calculated for each member against the fixed observed event-maximum sample. The full pipeline has a smaller mean distance for all three path types and for the pooled sample, with the paired reductions reported in Table S8. These values quantify distributional agreement over the observed event-maximum range; they are not a separate tail-calibration criterion.

The corresponding pooled cell-level marginal comparison is not uniformly improved by the conditional substitution: the event-level temperature and random-effect mixture changes the unconditional cell-level distribution, with the largest change occurring for the single-time-step transient type. The mean per-member PSS changes from 0.950 to 0.841 for the single-time-step transient type, from 0.947 to 0.875 for the two-time-step short-lived type, and from 0.931 to 0.924 for the full-lifecycle type. This is a scale-dependent trade-off between cell-level bulk fidelity and event-scale behaviour.

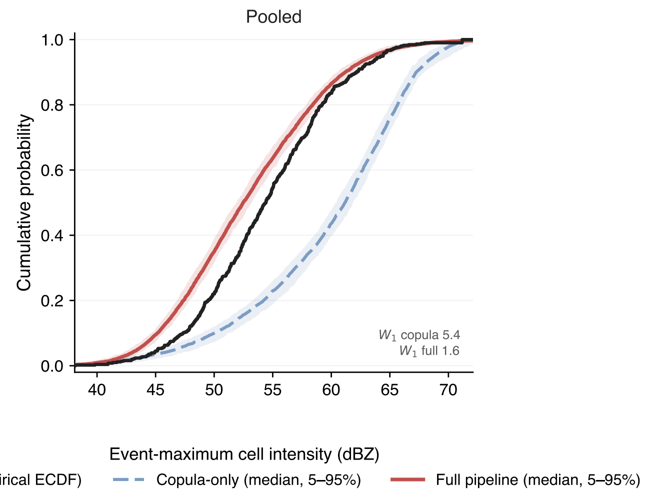

Figure S2. Pooled event-maximum cell-intensity distribution across the three path types, with the event maxima calculated from the path-specific reference-state intensities. The observed empirical distribution is shown as a charcoal step function; coloured curves and bands show the median and 5–95% range across 100 observation-sized production members for the copula-only baseline and full pipeline. The annotations report the mean Wasserstein-1 distance to observation in dBZ for each configuration.

Table S8. Wasserstein-1 distance between observed and simulated event-maximum distributions. Values are means over 100 observation-sized members; positive paired reductions indicate a smaller distance for the full pipeline. The intervals are percentile bootstrap intervals over paired members.

Path type	Copula-only	Full pipeline	Reduction	95% interval
Single-time-step transient	7.24	2.39	4.86	[4.70, 5.01]
Two-time-step short-lived	4.98	2.01	2.98	[2.82, 3.14]
Full-lifecycle	3.88	1.11	2.77	[2.68, 2.86]
All types pooled	5.36	1.64	3.73	[3.61, 3.84]

S4.6 SMEV return-level details

The main text presents the SMEV result through the absolute return-level curves and multiplicative validation ratios. The same event definition, top-13 fitting rule, annual event rate, plotting positions and return-level calculation are used for the observations, the copula-only baseline and the full pipeline. The corresponding numerical biases are provided here for reproducibility across all evaluated return periods. They are not used to single out one return period; the main-text interpretation concerns the overall RP2–100 behaviour.

Table S9. Percentage SMEV return-level bias $B_k(T) = 100[Q_k(T) - 1]$ with respect to the observed event-maximum reference-state-intensity return level over RP2–100. Values are ensemble means over 100 members of 13-year simulations; the corresponding ensemble bands are shown in the main-text SMEV figure.

Cell-path type	Configuration	RP2	RP5	RP10	RP20	RP50	RP100
PL1 transient	copula-only	$+10.2$	$+3.9$	$+0.7$	-1.8	-4.6	-6.4
	+ GAMLSS	-0.3	-0.9	-1.2	-1.5	-1.7	-1.9
PL2 short-lived	copula-only	$+12.9$	$+8.4$	$+6.0$	$+4.2$	$+2.1$	$+0.8$
	+ GAMLSS	$+2.0$	$+2.4$	$+2.6$	$+2.8$	$+3.0$	$+3.2$
PL3 full-lifecycle	copula-only	$+1.0$	-1.1	-2.2	-3.1	-4.0	-4.7
	+ GAMLSS	-0.2	$+0.2$	$+0.4$	$+0.5$	$+0.7$	$+0.8$

The fitted Weibull shape parameter is much larger for the copula-only event maxima than for the observations and the full pipeline in the transient and full-lifecycle cases. The corresponding approximate values are 62, 4.9 and 6 for the copula-only, observed and full-pipeline transient maxima, and 96, 13.2 and 12.5 for the corresponding full-lifecycle maxima. This reflects the tighter clustering of event maxima across storms and explains the under-estimation of the long-return-period tail. A sensitivity check using 8, 10, 13, 16 and 20 upper-order event maxima gives the same qualitative RP100 comparison: the full-pipeline biases remain close to zero, while the copula-only transient and full-lifecycle cases remain negative. SMEV is consequently used as a common downstream diagnostic with a documented fitting-window sensitivity, not as a family-selection criterion.

S5 Cell-property dependence and lifecycle reconstruction

The pairwise rank-dependence diagnostics in this section evaluate the cell-level Kendall's τ values for the production copula models. These copulas were fitted using cells pooled across the observed storm cohort, so the diagnostics describe the rank relationships in the combined cell sample. They include the intensity–geometry triples, intensity–rate associations, the six full-lifecycle peak-state copula links including pathLength, and all fifteen pairwise links of the two-time-step short-lived six-dimensional copula. The observed values are generally close to the production ensemble, with small systematic departures concentrated in some intensity-related pairs. Across the supplementary links, the maximum absolute difference between the observed Kendall's τ and the production-ensemble median is 0.033. Together with the endpoint-reconstruction checks below, these diagnostics provide a pairwise assessment of the dependence relationships represented in the production cell-property models. The event-level duration–cell-count copula is also checked below, while other event-scale conditional dependencies and cross-type temporal dependence remain outside the present validation scope.

The Kendall-τ and endpoint-PSS diagnostics are summarised at the member level. Each member spans the same 13-year record length as the observations, while its event and cell counts are sampled by the production model. The rank-dependence spread therefore reflects realisation-to-realisation variation at the record length used for comparison. The endpoint distributions use the same observation-length members because histogram-based Perkins Skill Scores are sensitive to small-sample binning noise. These choices affect the spread displayed by the validation ensembles, not the fitted copulas or the underlying sampling distributions.

The full-lifecycle check adds the three links between pathLength and the peak-state properties to the compact intensity–geometry diagnostic in the main text. The observed values remain close to the corresponding production-ensemble distributions, with the clearest small departure in the intensity–major-axis link. The full two-time-step short-lived diagnostic covers all 15 pairwise links in the production vine, including the six initial-state–cross-rate links and the three rate–rate links that are not represented in the compact intensity–rate figure below. These supplementary figures provide the pairwise diagnostics underlying the compact main-text assessment.

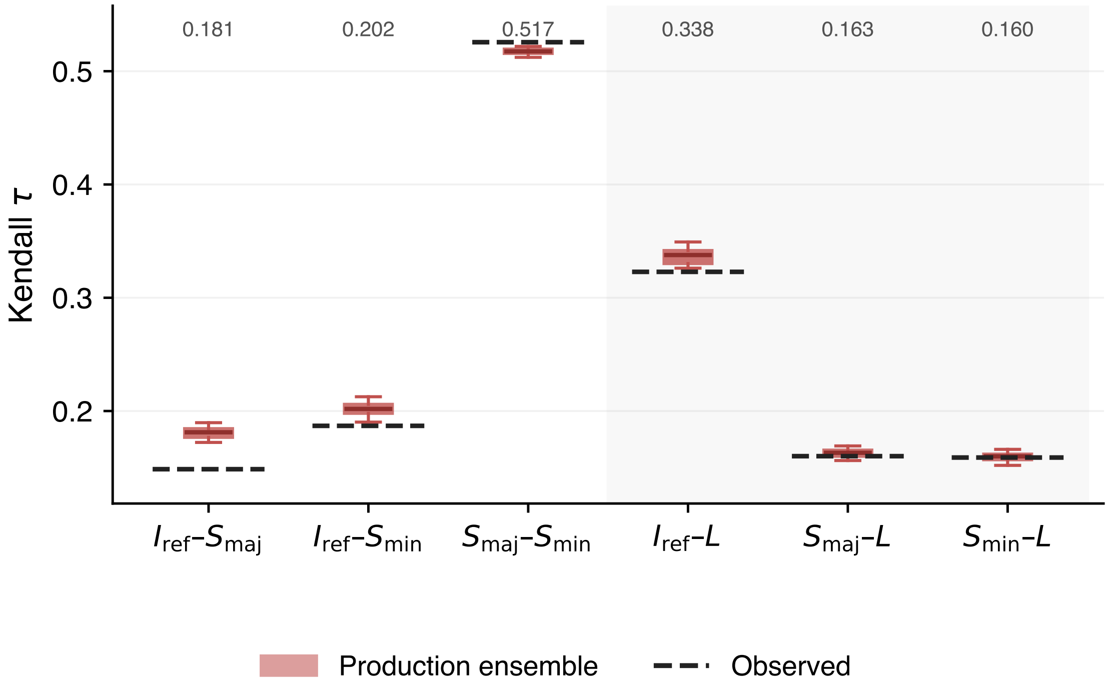

Figure S3. Cell-level pairwise Kendall-τ diagnostics for the four-dimensional full-lifecycle peak copula, including pathLength. The three pathLength links extend the compact peak-property diagnostic in the main text. For PL3, I_{ref} corresponds to the peak-state intensity. Dashed lines indicate the observed values.

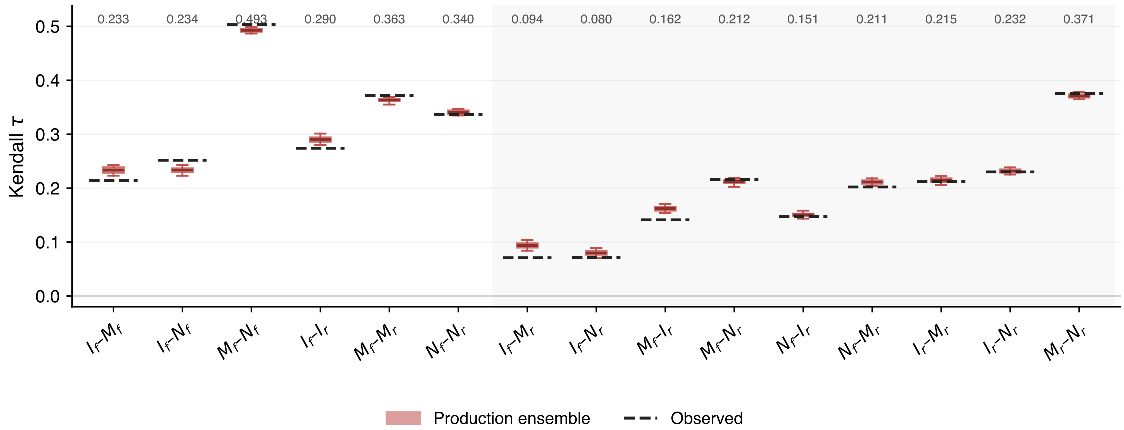

Figure S4. Cell-level pairwise Kendall-τ diagnostics for the six-dimensional two-time-step short-lived copula. The panels cover all 15 links among the three initial-state properties and their three resolved rates. In the labels, I, M and N denote intensity, major-axis length and minor-axis length, respectively; subscripts f and r denote the initial state and the corresponding resolved one-step rate, respectively. Dashed lines indicate the observed values.

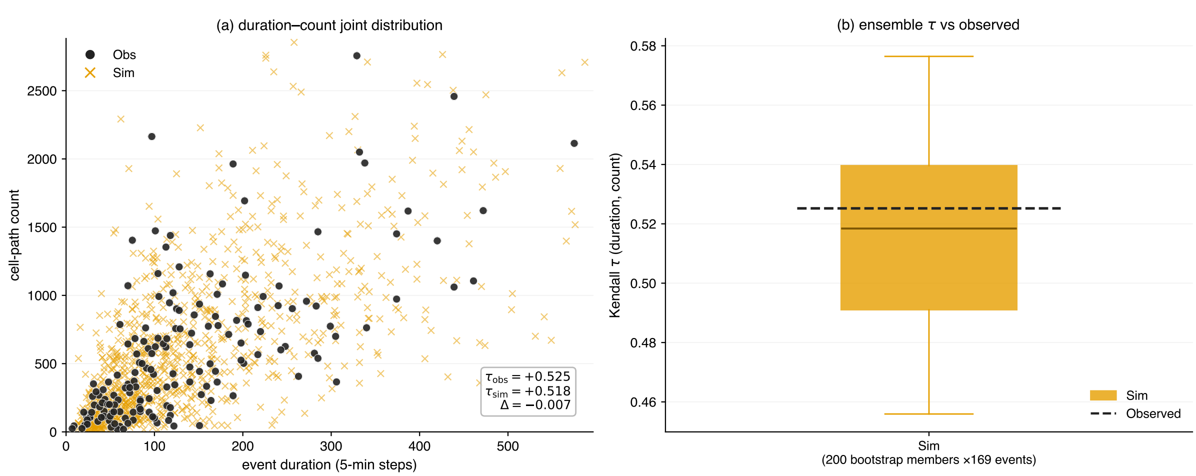

Figure S5. Event-level dependence between storm duration and total cell-path count. The observed joint structure is compared with the Gaussian-copula event sampler using observation-sized ensemble members.

The duration–count copula reproduces the positive association and joint support of the two sampled quantities. This is a direct check of the event-scale copula, separate from the marginal validation of duration and cell count.

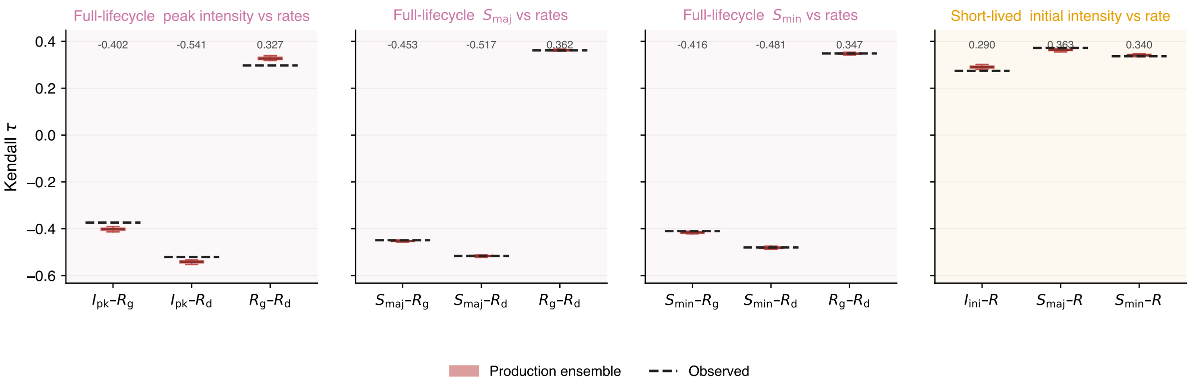

Figure S6. Cell-level intensity–growth/decay-rate dependence for the lifecycle properties used by the generator. The panels cover the three full-lifecycle peak–rate groups and the two-time-step short-lived initial-intensity–rate group; dashed lines indicate the observed values.

For the peak–rate checks, the observed Kendall τ values are generally close to the production ensemble, with small departures for some intensity–rate links. The GAMLSS substitution changes the reference-state intensity marginal, while rank matching and subsequent copula conditioning are used to keep the reference-state intensity coupled to the lifecycle rates during generation.

The endpoint check reconstructs the first and last full-lifecycle cells from the sampled peak properties and the corresponding growth and decay rates. For two-time-step short-lived paths, the second cell is reconstructed from the first cell and the single resolved rate. The resulting endpoint distributions are compared with the observed distributions below.

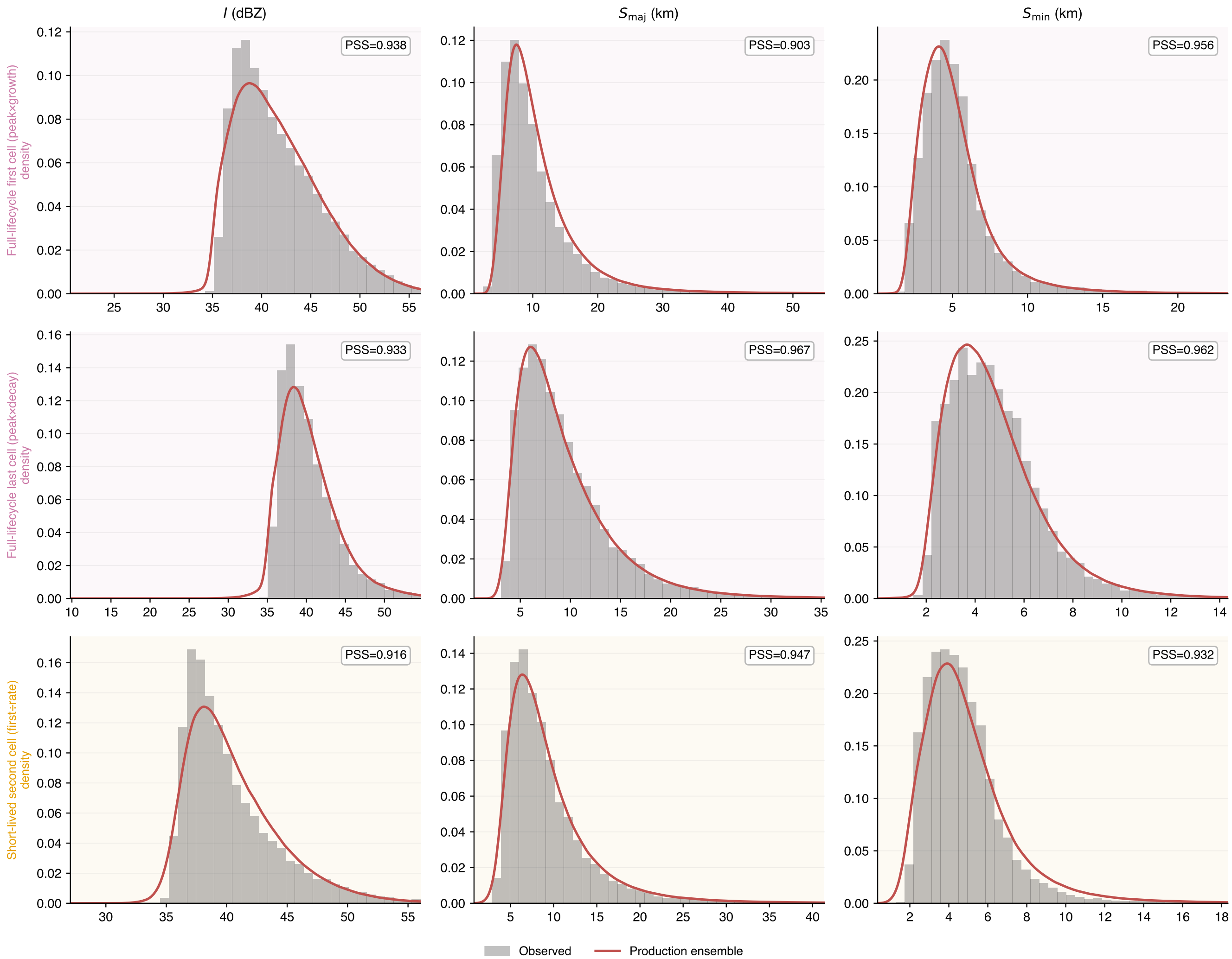

Figure S7. Distributions of inferred lifecycle endpoint properties compared with observations. The upper two rows show the reconstructed first and last cells for full-lifecycle paths; the lower row shows the reconstructed second cell for two-time-step short-lived paths. The panels report the Perkins Skill Score for each comparison. Here, I denotes the reconstructed cell intensity at the corresponding lifecycle endpoint, which is not the reference-state intensity I_{ref}.

The endpoint distributions have Perkins Skill Scores between 0.90 and 0.98. The full-lifecycle first-cell intensity is the least
245 closely matched endpoint, with the small difference arising from the broader full-GAMLSS peak-intensity distribution; the remaining endpoint properties show close overlap with the observations.

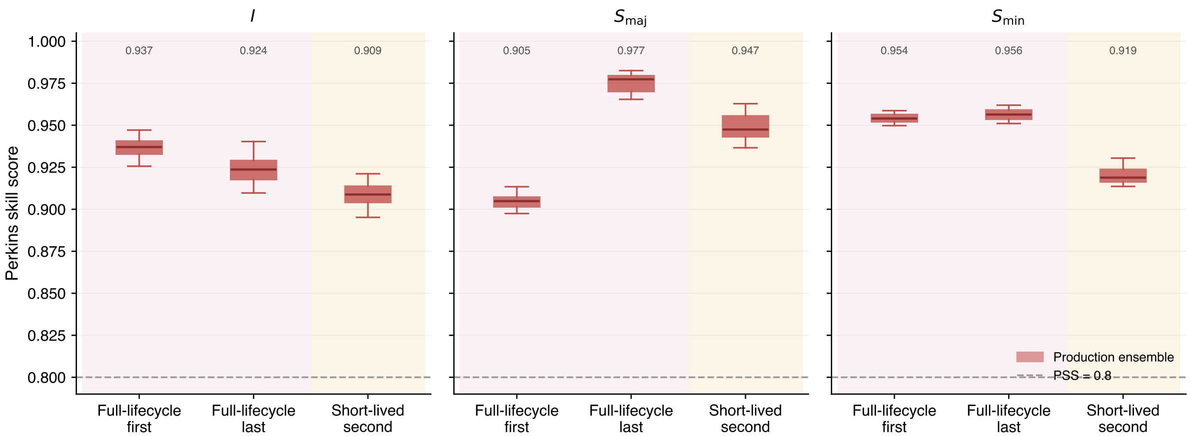

Figure S8. Ensemble Perkins Skill Scores for reconstructed lifecycle endpoints. The ensemble distributions summarise 100 independent observation-sized members. All endpoint comparisons exceed the working PSS threshold of 0.8 used in this study.

Across the 100 observation-sized members, the endpoint PSS medians range from 0.90 to 0.98 and all 5–95% intervals remain above 0.85. These results provide an end-to-end check that the peak–rate dependence and the reconstruction equations jointly reproduce the observed temporal endpoints beyond matching their component marginals.

The upper-tail coefficient at $q = 0.95$ was retained as an exploratory summary in the analysis archive rather than as a separate manuscript figure. It is not used as a fitting criterion or as a general dependence diagnostic.

S6 Scale-consistent rainfall-field rendering

S6.1 Cell representations and finite-resolution observation

The theoretical renderer distinguishes three representations of cell geometry. The tracking output provides second-moment semi-axes $(S_{\mathrm{maj}}^{\mathrm{mom}}, S_{\mathrm{min}}^{\mathrm{mom}})$, which describe the spatial dispersion of the tracked cell support. The continuous EXCELL field uses decay scales (a, b), whereas the rainfall footprint is the set of grid cells exceeding the rendering threshold I_{thr}. These objects have different roles and are not treated as interchangeable.

For an underlying point-scale peak $I_{\max,p}$, the continuous field is

$$I(x,y) = I_{\max,p}\exp(-r), \qquad r = \sqrt{\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2}.$$

The observed peak $I_{\max}^{\mathrm{obs}}$ is not the point value $I_{\max,p}$. It is the average intensity over the 1 km radar grid cell $D_{\max}$ containing the peak,

$$I_{\max}^{\mathrm{obs}} = \frac{1}{|D_{\max}|}\iint_{D_{\max}} I(x,y)\,\mathrm{d}A.$$

The continuous coordinate system is centred on the point peak, whereas $D_{\max}$ is the radar cell that contains it. The point peak can therefore be offset from the cell centre; the production renderer evaluates the integral at the corresponding sub-grid position. This is the finite-resolution observation principle used for peak inversion by Tseng et al. (2025). The present construction additionally couples the inversion to a sampled threshold-footprint geometry before solving for the point peak. Here the integral is evaluated numerically on the same grid support used by the renderer.

The exponential radial form is supported by a cell-scale check using the tracked Birmingham radar cells. Figure S9 shows the volume-normalised mean profile after alignment by moment-normalised radius. Over the fitted radial interval, the observed mean profile is closely described by an exponential; the small departure near the centre is retained in the figure. This check supports the chosen cell-profile form but is not a validation of rainfall-field spatial skill.

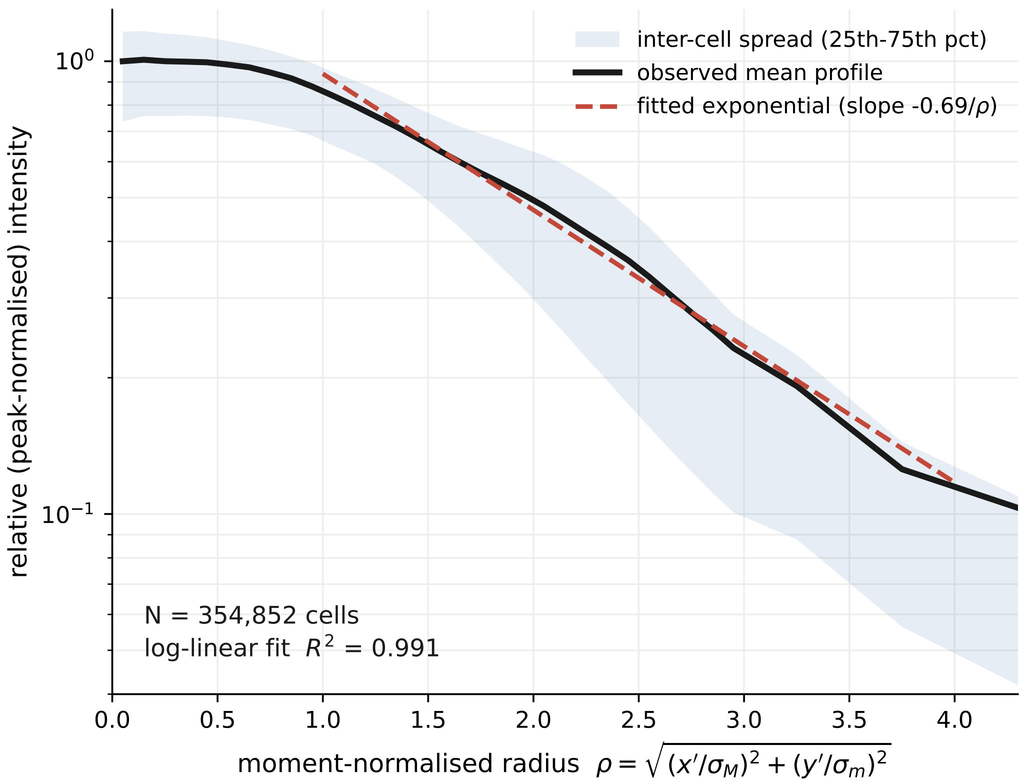

Figure S9. Volume-normalised mean radial intensity profile of the 354,852 tracked Birmingham radar cells, aligned by moment-normalised radius. The red dashed line is an exponential fitted over $\rho \in [1, 4]$ (log-linear $R^2 = 0.991$); the shaded band shows the inter-cell 25th–75th percentile range rather than a confidence interval. This cell-scale form check supports the exponential EXCELL cell model and is not a rainfall-field spatial validation.

S6.2 Moment-to-threshold-area mapping

The moment ellipse defines the reference area

$$A_{\mathrm{ref}} = \pi S_{\mathrm{maj}}^{\mathrm{mom}} S_{\mathrm{min}}^{\mathrm{mom}}.$$

It is related statistically, rather than deterministically, to the observed threshold-footprint area A_{obs} through

$$R = \frac{A_{\mathrm{ref}}}{A_{\mathrm{obs}}}.$$

For each cell-path type, the production renderer samples $Y = R - 1$ from a log-normal distribution calibrated by quantile matching to the observed median and upper quartile in the 169-event calibration cohort, and caps the sampled value at the observed 99th percentile. The resulting area-equivalent threshold semi-axes are

$$\widetilde{S}_{\mathrm{maj}} = \frac{S_{\mathrm{maj}}^{\mathrm{mom}}}{\sqrt{R}}, \qquad \widetilde{S}_{\mathrm{min}} = \frac{S_{\mathrm{min}}^{\mathrm{mom}}}{\sqrt{R}}.$$

This isotropic mapping retains the moment ellipse's orientation and aspect ratio while changing its area. It is a parsimonious representation of threshold-footprint area, not a reconstruction of the irregular observed footprint. The empirical medians used as calibration targets are 1.58, 1.61 and 1.71 for the single-time-step transient, two-time-step short-lived and full-lifecycle types, respectively; their corresponding empirical interquartile ranges are [1.23, 2.75], [1.28, 2.58] and [1.38, 2.34]. A 2×10^6-draw check of the calibrated distributions gives sampled medians of 1.580, 1.609 and 1.713, respectively.

S6.3 Joint inversion and numerical solution

Let

$$L = \log\left(\frac{I_{\mathrm{max},p}}{I_{\mathrm{thr}}}\right), \qquad \rho(x,y) = \sqrt{\left(\frac{x}{\widetilde{S}_{\mathrm{maj}}}\right)^2 + \left(\frac{y}{\widetilde{S}_{\mathrm{min}}}\right)^2}.$$

With $a = \widetilde{S}_{\mathrm{maj}}/L$ and $b = \widetilde{S}_{\mathrm{min}}/L$, the field becomes

$$I(x,y) = I_{\mathrm{thr}} \exp[L\{1 - \rho(x,y)\}].$$

The unknown L is obtained by solving

$$\frac{1}{|D_{\mathrm{max}}|} \iint_{D_{\mathrm{max}}} \exp[L\{1 - \rho(x,y)\}] \,, \mathrm{d}A = \frac{I_{\mathrm{max}}^{\mathrm{obs}}}{I_{\mathrm{thr}}}.$$

Conditional on the sampled R, the solution determines the point peak and decay scales while reproducing the target grid-scale peak and the area-equivalent threshold geometry. A sequential procedure that solves for the point peak before applying the area mapping would generally no longer satisfy the observation operator after the geometry changes. In a closure diagnostic, this sequential construction produced a median grid-scale peak residual of approximately 3.5%, with values reaching approximately 16% in the upper tail of the residual distribution.

For fixed area-equivalent geometry, the left-hand side of the inversion equation is denoted by $F(L)$. The positive root of $F(L) = I_{\max}^{\mathrm{obs}}/I_{\mathrm{thr}}$ exists uniquely because $\log F(L)$ is a convex log-moment-generating function of $1-\rho$ over $D_{\max}$, with $F(0) = 1 < I_{\max}^{\mathrm{obs}}/I_{\mathrm{thr}}$ and $F(L) \to \infty$ as $L \to \infty$. This argument does not require $F(L)$ to be monotonic over its full domain. The production implementation brackets the positive root and obtains it numerically for each rendered cell state.

The isotropic mapping is supported as a parsimonious approximation by the observed alignment of the threshold-footprint orientation and aspect ratio with their moment-ellipse counterparts. It does not imply that the idealised ellipse reproduces the detailed boundary or internal texture of an observed radar cell.

S6.4 Observation-conditioned cell analogues (exploratory comparison)

An observation-conditioned cell-analogue renderer was explored as a visual reference. It constructs a field from matched observed cell patterns rather than from the self-contained theoretical field used by the production generator. The resulting intra-cell texture is informative, but the output remains conditioned on a finite library of observed patterns. It is therefore not a generative, externally applicable rendering path and is not used in the production simulations.

Figure S10 juxtaposes the theoretical and analogue renderings of the same synthetic storm over the same spatial window and colour scale. This comparison is included to distinguish the production renderer from an observation-conditioned reference, not to establish spatial skill or to select a preferred renderer. Quantitative field-scale validation remains outside the scope of the present manuscript.

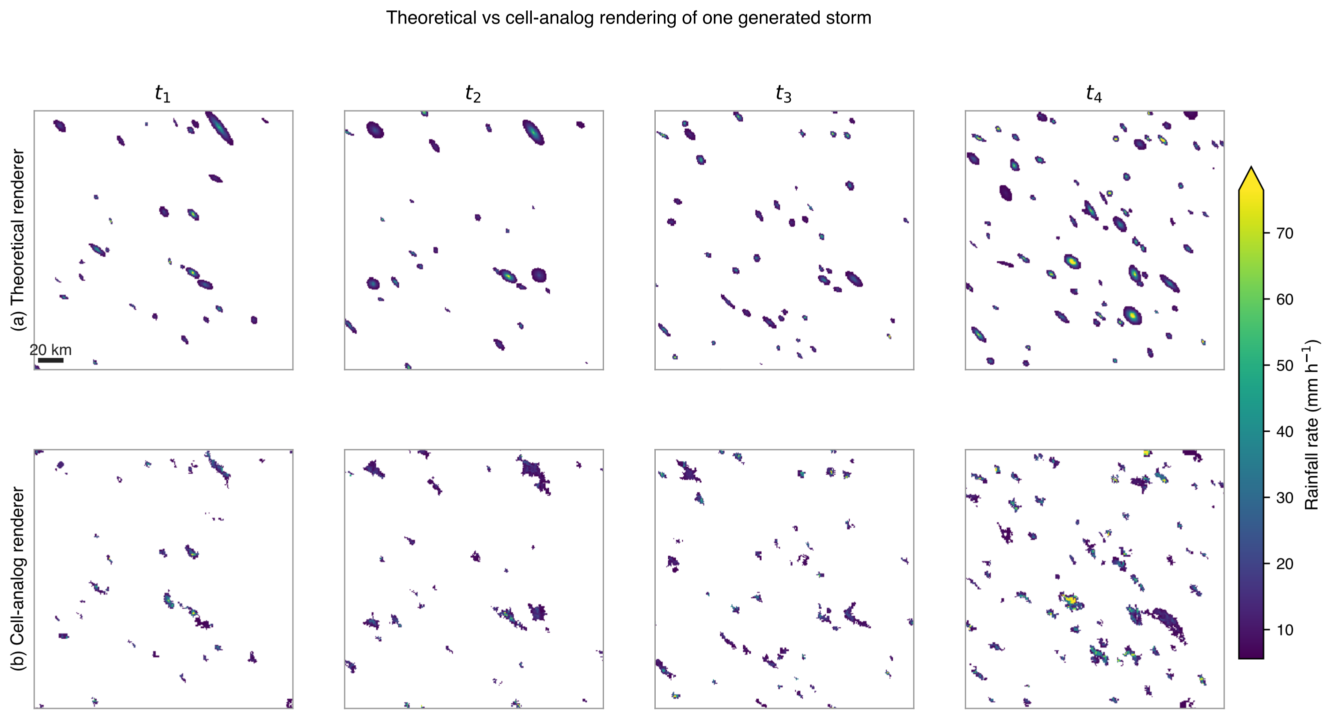

Figure S10. Exploratory rendering comparison for one synthetic storm. The upper row uses the production scale-consistent theoretical renderer; the lower row uses an observation-conditioned cell-analogue reference. Each column is a 5-min snapshot, and all panels use the same spatial window and rainfall-rate colour scale. The analogue reference is not part of the production generator and is not interpreted as a spatial-skill benchmark.

References

Onof, C. and Wang, L.-P.: Modelling rainfall with a Bartlett–Lewis process: new developments, Hydrology and Earth System Sciences, 24, 2791–2815, https://doi.org/10.5194/hess-24-2791-2020, 2020.

Stasinopoulos, D. M. and Rigby, R. A.: Generalized additive models for location scale and shape (GAMLSS) in R, Journal of Statistical Software, 23, 1–46, https://doi.org/10.18637/jss.v023.i07, 2007.

Stasinopoulos, D. M., Rigby, R. A., Heller, G. Z., Voudouris, V., and De Bastiani, F.: Flexible Regression and Smoothing: Using GAMLSS in R, Chapman & Hall/CRC, https://doi.org/10.1201/b21973, 2017.

Tseng, C.-Y., Wang, L.-P., and Onof, C.: Modelling Convective Cell Life Cycles with a Copula-Based Approach, Hydrology and Earth System Sciences, 29, 1–25, https://doi.org/10.5194/hess-29-1-2025, 2025.