Supplementary Information

Contextual gating modifies climate-conditioned dengue transmission potential across India

Peri Subrahmanya Hari Prasad[1,*] and **A. N. V. Satyanarayana**[1]

[1]Centre for Oceans, Rivers, Atmosphere and Land Sciences (CORAL), Indian Institute of Technology Kharagpur, Kharagpur 721302, West Bengal, India

[*]Correspondence: pshari19@gmail.com

Supplementary methods, results, and supporting tables

S1. Climate forcing and construction of annual transmission fields

Daily meteorological fields (near-surface air temperature, precipitation and near-surface specific humidity) from MPI-ESM1-2-LR, MRI-ESM2-0 and ACCESS-ESM1-5 were harmonised spatially and calendrically to the common analysis grid, and the grid-cell median was used to construct the daily forcing supplied to CliMed (daily three-GCM median forcing). The historical experiment (the reference period in the expressions below) covered 2005–2014 and the near-term experiments covered 2026–2035. SSP2-4.5 forcing was used for the analyses supporting Figs. 1–4, whereas SSP1-2.6, SSP2-4.5 and SSP5-8.5 were compared for Fig. 5. Optional contextual gates were disabled for the climate-only comparisons, and all spatial comparisons used the common analysis domain of 8,292 valid grid cells.

Daily entomological inoculation rate (EIR) and force of infection (FOI) outputs were aggregated to annual mean fields for the historical (reference) and near-term periods. All comparisons were performed on the same common grid and using the same geographical mask. Invalid cells were retained as missing values and were not replaced with zeros.

For a model variable X, the near-term change at the grid cell $i$was calculated as

$$\Delta X_i = \bar{X}_{i,\mathrm{near-term}} - \bar{X}_{i,\mathrm{reference}},$$

where the overbar denotes the temporal mean over the corresponding analysis period. Positive values, therefore, indicate an increase under near-term conditions, whereas negative values indicate a decrease.

S2. Statistical characterisation of climate–EIR relationships

The statistical characterisation supporting Fig. 2 used two components. Grid-cell EIR values were first summarised across 20 equal-frequency classes of near-surface air temperature, relative humidity and precipitation, with the median EIR within each class providing an empirical description of the marginal climate-response structure. Nonlinear conditional relationships were then represented using additive spline–ridge regression: each meteorological predictor was transformed with a cubic spline basis of degree three with six knots and no bias term, and the combined basis was fitted using ridge regression with a regularisation parameter of 10.

Predictive performance for ΔEIR was evaluated using five-fold grouped spatial cross-validation, in which grid cells were assigned to non-overlapping 3° × 3° blocks retained whole within either the training or validation subset. The contribution of each predictor was examined with a drop-one procedure ($\Delta R^2 = R^2_{full} - R^2_{reduced}$, so negative values indicate that removing the predictor improved out-of-block prediction). For descriptive spatial attribution, the full additive model was fitted to all valid grid cells and each grid cell was assigned the meteorological component with the largest absolute centred fitted contribution (Fig. 2e); spatial residuals were calculated as simulated minus cross-validated predicted ΔEIR (Fig. 2f). The principal settings and results are summarised in Supplementary Table S1.

Supplementary Table S1. Spline–ridge specification and spatial-block cross-validation results for the ΔEIR analysis.

Analysis / setting	Result
ΔEIR spatial-block cross-validated R^2	−0.071
Drop temperature ΔR^2	−0.0047
Drop relative humidity ΔR^2	−0.0098
Drop precipitation ΔR^2	−0.0318
Spatial blocks	3° × 3°
Cross-validation folds	5
Spline basis	Cubic (degree 3), six knots, no bias term
Ridge regularisation α	10

S3. Seasonal climatology, peak timing and L_{80}

Monthly transmission and climate climatologies were calculated from daily spatial summaries over the common modeling domain. For each day, EIR and FOI were averaged across all finite model grid cells,

$$\bar{EIR}_d = \frac{1}{N_d}\sum_{i=1}^{N_d} EIR_{i,d},$$

and

$$\bar{FOI}_d = \frac{1}{N_d}\sum_{i=1}^{N_d} FOI_{i,d},$$

where N_d is the number of valid model cells on day d. These are unweighted common-grid-cell means and were not population weighted. Missing or invalid grid cells were excluded from the daily spatial mean rather than assigned zero values.

Daily climatic summaries were constructed over the same geographical domain. Near-surface air temperature was converted from Kelvin to degrees Celsius, and precipitation flux was converted from kg m $^{-2}$ s $^{-1}$ to mm day $^{-1}$ using

$$P_{\mathrm{mm\,day^{-1}}} = P_{\mathrm{kg\,m^{-2}\,s^{-1}}} \times 86400.$$

Relative humidity was derived from near-surface air temperature, specific humidity and pressure. Vapour pressure was calculated as

$$e = \frac{qp}{0.622 + 0.378q},$$

where q is specific humidity and p is pressure. Saturation vapour pressure was calculated from air temperature as

$$e_s = 611.2 \exp\left(\frac{17.67 T_C}{T_C + 243.5}\right),$$

where $T_C$is air temperature in degrees Celsius. Relative humidity was then obtained as

$$RH = 100 \frac{e}{e_s},$$

with values constrained to the physical range of 0–100%.

Monthly climatologies were constructed in two stages. Daily spatial means were first averaged within each year and calendar month to produce one value for every available year–month combination,

$$X_{y,m} = \frac{1}{n_{y,m}} \sum_{d \in (y,m)} X_d \,,$$

where Xrepresents EIR, FOI, temperature, relative humidity or precipitation. Monthly climatological means were then calculated across the year-specific monthly values,

$$\bar{X}_m = \frac{1}{Y_m} \sum_{y} X_{y,m} \,.$$

This procedure prevents years containing different numbers of daily observations from receiving disproportionate weight in the climatological summaries. Transmission and climatic variables were treated identically in the monthly aggregation.

Interannual variability for each calendar month was represented by the 10th and 90th percentiles of the corresponding year-specific monthly means,

$$Q_{10,m} = P_{10}\left(X_{y,m}\right),\ Q_{90,m} = P_{90}\left(X_{y,m}\right).$$

The shaded envelopes in Fig. 3 represent the 10th–90th percentile range across year-specific monthly means and are not confidence intervals. The twelve monthly climatological values were plotted without temporal smoothing.

Reference and near-term climatologies were calculated independently using the corresponding daily model outputs. Differences between corresponding monthly means were calculated from the unrounded climatological values. Where relative changes are reported in the main text, the percentage change for calendar month $m$was calculated as

$$\%\Delta X_m = 100 \left(\frac{\bar{X}_{m,\text{near-term}} - \bar{X}_{m,\text{reference}}}{\bar{X}_{m,\text{reference}}}\right).$$

Spatial changes in transmission seasonality were quantified from cell-wise monthly climatologies of EIR. Daily EIR was first averaged separately within each year and calendar month for every valid model grid cell. The resulting year-specific monthly means were subsequently averaged across years for the corresponding calendar month. Each available year therefore contributed equally to the 12-month climatology rather than individual days being pooled across the complete analysis period.

For grid cell i, the monthly climatological EIR for calendar month $m$was calculated as

$$E\bar{I}R_{i,m} = \frac{1}{Y_{i,m}} \sum_{y} \left(\frac{1}{n_{i,y,m}} \sum_{d \in (y,m)} E\,IR_{i,d} \right),$$

where $Y_{i,m}$is the number of years contributing to calendar month m, and $n_{i,y,m}$is the number of valid daily observations for grid cell $i$during month $m$of year y.

Peak transmission month

The climatological peak EIR month at each grid cell was defined as

$$M_i = \underset{m\in\{1,\ldots,12\}}{\arg\max}\ \bar{EIR}_{i,m}\,.$$

Cells for which the annual climatology contained no positive EIR were excluded from the timing analysis.

Because calendar month is circular, the difference between near-term and reference peak timing was calculated using the shortest signed displacement around the annual cycle,

$$\Delta M_i = \left[\left(M_{i,\mathrm{near}} - M_{i,\mathrm{ref}} + 6\right)\bmod(12)\right] - 6.$$

Positive values indicate that the near-term peak occurs later in the annual cycle, whereas negative values indicate an earlier peak. This formulation prevents transitions across the December–January boundary from being interpreted as shifts approaching 12 months. Under the implemented convention, the displacement ranges from -6 to $+5$months.

Changes in the temporal concentration of annual EIR were quantified using L_{80}, defined as the minimum number of contiguous calendar months required to contain at least 80% of the annual climatological EIR at an individual grid cell.

For monthly climatological EIR values

$$E_{i,1}, E_{i,2}, \ldots, E_{i,12},$$

the annual climatological total was

$$E_{i,\mathrm{annual}} = \sum_{m=1}^{12} E_{i,m}\,,$$

and the corresponding 80% target was

$$T_i = 0.8E_{i,\mathrm{annual}}.$$

For every possible starting calendar month s, contiguous monthly EIR was accumulated forward through the annual cycle. Calendar months were treated circularly, so a sequence beginning in December could continue through January and subsequent months. For each starting month, the smallest number of months $k$satisfying

$$\sum_{j=0}^{k-1} E_{i,\ [(s+j-1)\ 12]+1} \geq 0.8E_{i,\mathrm{annual}}$$

was identified. $L_{80}$was then defined as the minimum such value across all twelve possible starting months,

$$L_{80,i} = \min_{s} k_{i,s}.$$

The implemented calculation therefore identifies the shortest contiguous portion of the annual cycle containing at least 80% of climatological EIR, rather than simply counting the highest individual months irrespective of their temporal continuity.

Cells without positive annual climatological EIR were excluded. $L_{80}$is bounded between 1 and 12 months for valid cells. A smaller value indicates that annual EIR is concentrated within a shorter portion of the year, whereas a larger value indicates that the same annual climatological transmission is distributed across a longer contiguous period.

The change in seasonal concentration was calculated as

$$\Delta L_{80,i} = L_{80,i,\mathrm{near-term}} - L_{80,i,\mathrm{reference}}.$$

Thus,

$$\Delta L_{80} < 0$$

indicates greater seasonal concentration under near-term conditions, whereas

$$\Delta L_{80} > 0$$

indicates that annual EIR is distributed over a larger number of contiguous months.

The seasonal-timing and concentration statistics for the current historical (2005–2014) versus near-term (2026–2035) SSP2-4.5 comparison are summarised in Supplementary Table S2.

Supplementary Table S2. Peak-timing and seasonal-concentration statistics (SSP2-4.5, 7,634 paired grid cells).

Metric	Result
Paired valid grid cells	7,634
Peak month unchanged / later / earlier	83.0% / 11.1% / 5.9%
Most frequent peak month	August (81.4% → 79.5% of cells)
Mean L80 (historical → near-term)	2.960 → 2.968 months
L80 unchanged / decreased / increased	82.8% / 8.0% / 9.2%

S4. Cross-scenario hotspot consistency

Hotspot definitions and agreement metrics for the three-scenario comparison are described in the main Methods. Supplementary Table S3 summarises the principal scenario-level statistics over the 8,292 common valid grid cells.

Supplementary Table S3. Scenario-level transmission and hotspot-agreement statistics (2026–2035 versus 2005–2014).

Metric	SSP1-2.6	SSP2-4.5	SSP5-8.5
Mean EIR (infectious bites person^{-1} day^{-1})	0.0678	0.0697	0.0673
Change from historical	+14.6%	+17.8%	+13.8%
Cells with positive ΔEIR	72.3%	84.0%	78.7%
Historical–future hotspot Jaccard	0.885	0.890	0.897

Of the 765 hotspot cells identified in each near-term scenario, 714 were shared across all three SSPs.

S5. Contextual-gating configurations and spatial reprioritisation

Four configurations were compared under identical daily three-GCM median SSP2-4.5 forcing for 2026–2035: ungated, vector-occurrence gating, VIIRS gating and full contextual gating. Supplementary Table S4 lists the status of each contextual component in each configuration, and Supplementary Table S5 summarises the corresponding attenuation and spatial-reprioritisation statistics.

Supplementary Table S4. Contextual-component status in the four model configurations.

Component	Ungated	Occurrence	VIIRS	Full
Core climate–LULC–population–vector mechanism	Active	Active	Active	Active
Vector-occurrence gate (Ae. aegypti and Ae. albopictus)	Off	On	Off	On
VIIRS settlement/contact gate	Off	Off	On	On
Population-pressure modifier	Off	Off	Off	On
Population-context classification	Off	Off	Off	On
Exposed-population adjustment	Off	Off	Off	On

Component	Ungated	Occurrence	VIIRS	Full
Susceptible-population adjustment	Off	Off	Off	On

Supplementary Table S5. Attenuation and spatial reprioritisation of mean EIR under the contextual-gating configurations (SSP2-4.5, 2026–2035, 8,292 grid cells).

Configuration	Mean EIR (bites person^{-1} day^{-1})	Attenuation	Rank ρ vs ungated	Hotspot Jaccard
Ungated	0.0697	—	1.000	1.000
Vector occurrence	0.0240	65.6%	0.859	0.399
VIIRS	0.0132	81.1%	0.817	0.360
Full contextual gating	0.0116	83.4%	0.760	0.301

Ungated EIR represents climate-conditioned mechanistic transmission potential, and the contextual gates constrain that potential according to ecological and human-environmental structure; the attenuation magnitude is therefore a model response to the specified contextual configuration rather than an estimate of the fraction of nationally reported dengue cases.

State ranking by mean EIR | SSP2-4.5

Rank	Ungated	Vector occurrence	VIIRS	Full gating
1	Andaman & Nicobar	Andaman & Nicobar	Andaman & Nicobar	Andaman & Nicobar
2	Arunachal Pradesh	Kerala	Kerala	Kerala
3	Mizoram	Mizoram	Goa	Goa
4	Manipur	Assam	Assam	Assam
5	Nagaland	Nagaland	Mizoram	Mizoram
6	Meghalaya	Meghalaya	Bihar	Bihar
7	Assam	Manipur	Tamil Nadu	Tamil Nadu
8	Chhattisgarh	Tamil Nadu	Nagaland	Nagaland
9	Kerala	West Bengal	West Bengal	West Bengal
10	Odisha	Bihar	Meghalaya	Puducherry
11	Karnataka	Goa	Puducherry	Manipur
12	Telangana	Tripura	Manipur	Meghalaya
13	Andhra Pradesh	Karnataka	Karnataka	Karnataka
14	Tripura	Odisha	Chhattisgarh	Odisha
15	Tamil Nadu	Jharkhand	Jharkhand	Jharkhand
16	Madhya Pradesh	Andhra Pradesh	Odisha	Andhra Pradesh
17	Jharkhand	Chhattisgarh	Uttar Pradesh	Uttar Pradesh
18	Goa	Uttar Pradesh	Arunachal Pradesh	Tripura
19	West Bengal	Arunachal Pradesh	Telangana	Arunachal Pradesh
20	Maharashtra	Telangana	Andhra Pradesh	Chhattisgarh
21	Bihar	Puducherry	Tripura	Telangana
22	Uttar Pradesh	Maharashtra	Maharashtra	Haryana
23	Uttarakhand	Madhya Pradesh	Haryana	Maharashtra
24	Haryana	Haryana	Punjab	Punjab
25	Gujarat	Punjab	Madhya Pradesh	Delhi
26	Punjab	Dadra and Nagar Haveli and Daman and Diu	Uttarakhand	Uttarakhand
27	Puducherry	Uttarakhand	Delhi	Madhya Pradesh
28	Jammu & Kashmir	Gujarat	Dadra and Nagar Haveli and Daman and Diu	Dadra and Nagar Haveli and Daman and Diu
29	Dadra and Nagar Haveli and Daman and Diu	Delhi	Jammu & Kashmir	Gujarat
30	Himachal Pradesh	Himachal Pradesh	Gujarat	Jammu & Kashmir
31	Rajasthan	Jammu & Kashmir	Himachal Pradesh	Himachal Pradesh
32	Delhi	Rajasthan	Rajasthan	Rajasthan
33	Sikkim	Sikkim	Sikkim	Sikkim
34	Ladakh	Ladakh	Ladakh	Ladakh

Supplementary Fig. S1 | State and union-territory ranking by mean EIR under alternative contextual gating configurations. Administrative units were ranked by unweighted mean EIR across contributing model grid cells for 2026–2035 under SSP2-4.5. Rankings are shown for ungated, vector-occurrence, VIIRS and full contextual-gating configurations. Full gating substantially reorganised relative geographical prioritisation: only five of the ten highest-ranked units under ungated conditions remained within the top ten after full gating. Rankings describe relative modelled transmission potential and do not represent observed dengue incidence or burden.

S6. Persistent-hotspot robustness

Persistent hotspots were defined in the main Methods as grid cells classified within the annual upper-decile EIR distribution in at least eight of the ten years (2026–2035, SSP2-4.5). Supplementary Table S6 summarises persistent-hotspot reorganisation relative to the ungated configuration, and Supplementary Table S7 reports the sensitivity of the full-gating comparison to alternative percentile thresholds and persistence requirements.

Supplementary Table S6. Persistent-hotspot reorganisation relative to the ungated configuration (90th percentile, at least eight of ten years).

Configuration	Persistent hotspots	Retained	Lost	Newly prioritised	Jaccard
Ungated	645	—	—	—	—
Vector occurrence	616	346	299	270	0.378
VIIRS	594	301	344	293	0.321
Full contextual gating	605	267	378	338	0.272

Supplementary Table S7. Sensitivity of the full-gating persistent-hotspot comparison: Jaccard similarity relative to the ungated configuration across hotspot-percentile thresholds and minimum persistence requirements.

Hotspot threshold	≥7 years	≥8 years	≥9 years
85th percentile	0.338	0.323	0.314
90th percentile	0.288	0.272	0.251
95th percentile	0.215	0.207	0.197

S7. Mechanistic configuration and fixed parameters

The reference and near-term simulations were conducted using the same CliMed mechanistic configuration. The common model grid, study-area mask, population input, LULC raster, species parameter profiles, species-combination setting and contextual-gate configuration were retained between the two experiments. For the climate-comparison experiment, climatic forcing was the quantity allowed to vary between periods; static ancillary layers were otherwise held constant. This ensures that differences between the reference and near-term outputs are not inadvertently produced by changes in spatial resolution, population input, land-cover classification or gate configuration. The CliMed implementation explicitly requires historical/reference and near-term runs to use the same common grid, AOI, population, LULC, species configuration, gates and static rasters for a climate-only comparison. The fixed 2018 population and ancillary layers do not represent projected demographic or settlement change. The 20-km analysis grid does not create independent climate information at that resolution.

All variables were represented on the common latitude–longitude grid before transmission calculations. Continuous environmental variables were interpolated to the target grid, categorical LULC was transferred using nearest-neighbour sampling, and population-count inputs were handled as extensive quantities so that population mass was preserved. The common-grid implementation applies fixed categorical and extensive-data rules, including nearest-neighbour treatment of LULC and conservative treatment of population counts.

The LULC-dependent habitat and rainfall-response, humidity-suitability, thermal-response, vector-to-host, contextual-gate and coupled infection-dynamics formulations, together with their parameter values, are given in full in the main Methods and are not repeated here. Supplementary Table S9 summarises the principal fixed configuration values.

Population entered the model in two different forms. Population density contributed to the upstream vector-to-host calculation, whereas population count determined the number of people over which transmission probability was subsequently expressed.

Population-density context was normalized relative to the 90th percentile of population density,

$$D_{\mathrm{norm}}(x) = \mathrm{clip}\left[\frac{\ln\big(1 + H(x)\big)}{\ln(1 + H_{90})}, 0, 1\right].$$

The contextual index combined the bounded settlement/contact field and normalized population-density information as

$$U(x) = 0.55\, G_V(x) + 0.45\, D_{\mathrm{norm}}(x).$$

Grid cells were classified as

$$C(x) = \begin{cases} \text{rural}, & U < 0.33, \\ \text{peri} - \text{urban}, & 0.33 \le U < 0.66, \\ \text{urban}, & U \ge 0.66. \end{cases}$$

The population base entering the human transmission model was

$$N_{\text{risk},0}(x) = N_{\text{cell}}(x) f_{\text{exp}}(x) f_{\text{sus}}(x),$$

where $N_{\text{cell}}$is the population count represented within the model cell, $f_{\text{exp}}$is the context-dependent exposed fraction and $f_{\text{sus}}$is the susceptible fraction.

Supplementary Table S8. Context-dependent exposed and susceptible population fractions.

Context	Exposed fraction range	Susceptible fraction
Urban	0.03–0.18	0.12
Peri-urban	0.05–0.28	0.16
Rural	0.08–0.40	0.22

These are fixed model-configuration parameters rather than empirically estimated epidemiological fractions.

Aedes aegypti and *Aedes albopictus* were simulated separately before combination. Although the configuration retained nominal baseline species-weight fields of 0.5/0.5 for ancillary diagnostics, these weights did not multiply the dynamically simulated human forces of infection or realised EIR in the executed full-SEI configuration. Species-specific dynamic forces of infection were combined additively, consistent with the main Methods.

Supplementary Table S9. Principal fixed configuration values used in the CliMed analysis

Model component	Parameter	Value
Rainfall memory	Antecedent window	7 days
Flushing	P_{thr}	20 mm day $^{-1}$
Flushing	k_f	0.1 mm $^{-1}$
Relative humidity	RH_{min}	30%
Relative humidity	RH_{opt}	75%
Adult mortality	μ_{min}	0.05 day $^{-1}$
EIP	Lower bound	3 days
EIP	Upper bound	30 days
Context index	Settlement/contact contribution	0.55
Context index	Population-density contribution	0.45
Rural/peri-urban threshold	T_{peri}	0.33
Peri-urban/urban threshold	T_{urban}	0.66
Urban exposure	Range	0.03–0.18
Peri-urban exposure	Range	0.05–0.28
Rural exposure	Range	0.08–0.40
Urban susceptibility	f_{sus}	0.12
Peri-urban susceptibility	f_{sus}	0.16
Rural susceptibility	f_{sus}	0.22

Model component	Parameter	Value
Ae. aegypti	Diagnostic baseline weight	0.50
Ae. albopictus	Diagnostic baseline weight	0.50

The complete species-specific thermal coefficients, LULC-dependent $E_0$and K lookup values, occurrence-gate settings and other implementation constants are retained in the archived CliMed parameter files.