

Demographically-Informed Heat-Mortality Risk Curves via Risk Graph Neural Networks

Alex O. Davies¹ (✉), Eunice Lo¹, and Rui Zhu¹

School of Geographical Sciences, University of Bristol, UK
alexander.davies@bristol.ac.uk

Abstract. Estimating heat-related mortality risk is a core task in environmental epidemiology, typically addressed with Distributed Lag Non-linear Models (DLNMs); interpretable exposure-response surfaces fitted to temperature-mortality time series. DLNMs are effective but ignore demographic and geographic context, despite well-established relevance to heat vulnerability. We propose Risk Graph Neural Networks (RGNNs), a hierarchical GNN encoder that uses granular census features to optimise DLNM coefficient vectors, preserving interpretable risk curve outputs while substantially improving predictive calibration. Evaluated across 10 regions of England and Wales on two unprecedented heat years, RGNN variants maintain both lower point-errors and near-nominal uncertainty coverage during the 2022 heatwave where baselines collapse.

Keywords: temperature-mortality modelling · graph neural networks · distributed lag non-linear models · spatial epidemiology

1 Introduction

A central task in environmental epidemiology is estimating mortality risk from temperature exposure. The dominant tool is the Distributed Lag Non-linear Model (DLNM) [4], which produces an interpretable exposure-response surface over a lag dimension. This captures, for instance, that cold mortality peaks 7–14 days after exposure while heat mortality concentrates in the immediate 3-day window. DLNMs operate purely on temporal signals, however, making no use of demographic or geographic features despite their established importance in mediating heat vulnerability. Recent ML approaches [1] sometimes improve point estimates but rarely match the interpretability of the DLNM risk curve, and Boudreault et al. [2] note that the DLNM is rarely used as a proper benchmark.

A core reason for the wide adoption of DLNMs over other predictive models is the interpretability of their outputs. A risk curve is immediately transparent to technical and non-technical users, making it a powerful tool in informing policy decisions. We introduce Risk Graph Neural Networks (RGNNs), which augment DLNMs with a hierarchical GNN encoder over census features. The result preserves the interpretable risk curve while substantially improving calibration under distributional shift — precisely the regime that matters most for public

health decision-making during extreme heat events. Early counterfactual experiments produce spatially-resolved, policy-relevant mortality predictions across urban areas.

2 Background

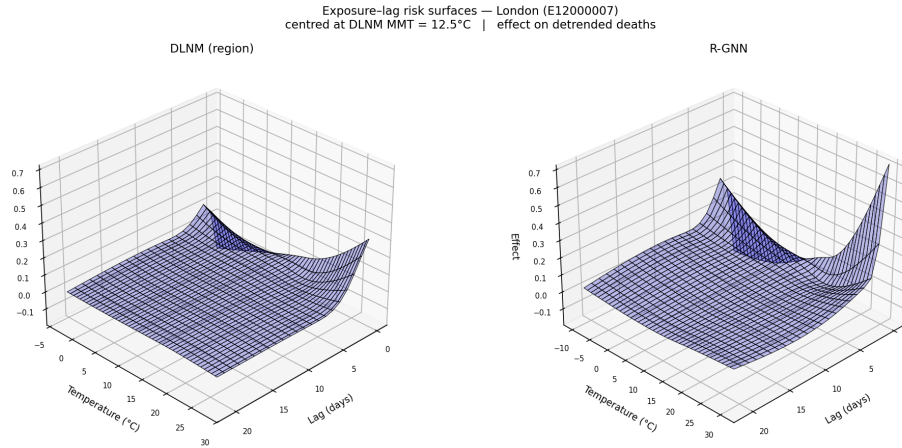


Fig. 1. Risk surfaces, in units of deaths per 100k, for London. Left: Vanilla DLNM. Right: RGNN-Calibrated.

Here we briefly summarise the current state of modelling the heat-mortality relationship. DLNMs are proposed and formalised in-detail in Gasparrini et al. [4], and Boudreault et al. [2] review ML approaches.

2.1 Distributed Lag Non-linear Models (DLNMs)

A DLNM [4] models a scalar health outcome at location i and time t as a linear combination of a *cross-basis* matrix that encodes the joint non-linear and lagged relationship between an exposure and the outcome:

$$\hat{y}_{it} = \mathbf{w}_i \cdot \mathbf{CB}_{it}, \quad (1)$$

where $\mathbf{w}_i \in \mathbb{R}^{n_{cb}}$ is a coefficient vector estimated by ordinary least squares and $\mathbf{CB}_{it} \in \mathbb{R}^{n_{cb}}$ is the cross-basis row for location i at time t . The primary output is a exposure-outcome effect surface along lag and the exposure variable, which in our case is lag-temperature-risk, as visualised in Figure 1. Fitting one DLNM per location independently produces unstable estimates for small areas with sparse death counts, and provides no mechanism for borrowing strength from neighbouring or demographically similar locations.

2.2 Graph Neural Networks (GNNs)

Graph Neural Networks [9] generalise neural networks to graph-structured data by iteratively aggregating information from local neighbourhoods. GATv2 [13] extends graph attention networks with a more expressive attention mechanism in which both source and target node features jointly determine attention weights, making it well-suited to heterogeneous hierarchies where child nodes vary substantially in informativeness. GraphSAGE [6] performs mean aggregation over sampled neighbourhoods and is widely used for spatial smoothing on geographic graphs. GNNs have been applied to urban analytics tasks including traffic forecasting [8] and air quality prediction [3], but their application to epidemiological exposure-response estimation is, to our knowledge, novel.

3 Risk Graph Neural Networks (RGNNs)

The RGNN replaces per-location independent OLS with a learned mapping from census features to DLNM coefficient vectors, structured as a two-level hierarchy: a frozen per-region OLS baseline is modulated by GNN-derived corrections at both the region and LAD levels. The output preserves both the interpretability and strong prior of the OLS-initialised DLNM, while improving performance with spatially-shared demographic information, as well as being comparably cheap to run compared to full ML models. An overview of the architecture is given in Figure 2.

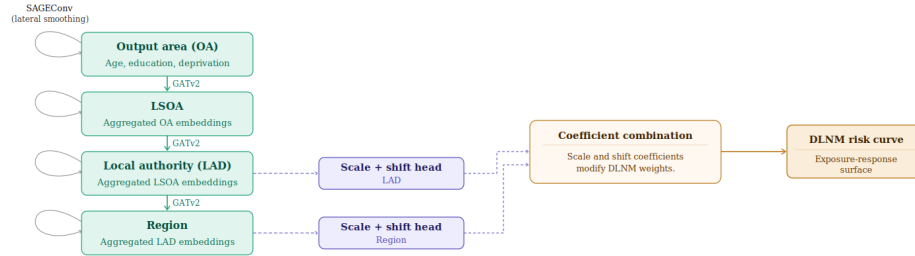


Fig. 2. RGNN architecture. Census features at the lowest administrative level are projected to a hidden embedding and propagated upward and laterally in separate stages, with heads predict modifying weights (Eqn. 2) over an OLS-initialised DLNM.

Initialisation. Before GNN training, per-region OLS coefficients $\mathbf{b}_r$ are estimated on region-level Pearson mortality residuals and frozen. All GNN heads are zero-initialised, so at the start of training every LAD’s coefficient vector equals its region OLS baseline.

Encoder. A bottom-up heterogeneous GNN propagates static demographic features from the finest available administrative unit upward through the hierarchy. At each level, child embeddings are aggregated into parent nodes using multi-head *GATv2Conv* attention; geographic neighbours at the same level are smoothed using *SAGEConv* mean aggregation. Higher-level nodes receive no direct demographic features; their representations are built entirely from upward message passing, forcing the model to learn meaningful aggregations.

Coefficient prediction. Zero-initialised linear heads at each predictive level map GNN embeddings to corrections in $\mathbb{R}^{n_{cb}}$:

$$\mathbf{w}_i = \mathbf{b}_{r(i)} + \text{clamp}(\mathbf{b}_{r(i)} \odot f_{ls}(\mathbf{h}_i) + f_{lsh}(\mathbf{h}_i), \pm\delta), \quad (2)$$

where $\odot$ is element-wise multiplication, δ is a hyperparameter controlling the maximum correction magnitude, and the scale and shift heads modulate the shape and location of the baseline curve respectively. Importantly, weights at each level anchor independently to the same frozen baseline rather than being chained, keeping granularities independently interpretable and preventing cascading errors. With dropout, and repeat predictions of coefficients, we can form useful uncertainties that can be propagated into the vanilla DLNMs uncertainty estimations.

Training objective. The primary loss is Lin’s CCC [10], which jointly penalises errors in correlation, variance ratio, and mean bias, summed at LAD and region levels with population weighting. Three auxiliary terms are added: (i) a monotonicity penalty encouraging physiologically plausible cold and heat tails; (ii) a consistency penalty penalising divergence between population-weighted LAD aggregates and region-level predictions; and (iii) a population-inverse-weighted L2 delta penalty pulling data-sparse LADs toward their regional baseline. MC-dropout over the coefficient heads propagates uncertainty into the final risk curves.

4 Experiments

We collect daily temperature data from HadUK [7] and LAD-level daily mortality through a data-sharing agreement with the UK ONS (2000–2024). For demographic features we use OA-level census information from Goodwin and Singleton [5] for 2021, with the implicit assumption that demographic drift is minimal across the study period. [Adjacent_To] edges connect areas sharing a boundary, e.g. (Wales) $\leftrightarrow$ (South West); [Contains] edges encode the administrative hierarchy, e.g. (Wales) $\rightarrow$ (Cardiff).

We demonstrate three RGNN variants, which share the same codebase and architecture, differing only in the depth of the administrative hierarchy: **rgnn** (LAD $\rightarrow$ Region, 2 levels) **rgnn_ Isoa** (LSOA $\rightarrow$ LAD $\rightarrow$ Region, 3 levels) **rgnn_oa** (OA $\rightarrow$ LSOA $\rightarrow$ LAD $\rightarrow$ Region, 4 levels) As baselines we use the DLNM, Random

Table 1. Region-level holdout evaluation across 10 regions of England and Wales. Values are mean $\pm$ standard deviation across regions. Bold indicates the best result.

Year	Model	RMSE $\downarrow$	MAE $\downarrow$	Pearson $\uparrow$	CRPS $\downarrow$	Cov@90% $\uparrow$
2018	DLNM	0.33 $\pm$ 0.07	0.25 $\pm$ 0.05	0.32 $\pm$ 0.05	0.19 $\pm$ 0.04	0.73 $\pm$ 0.03
	RF	0.29 $\pm$ 0.06	0.22 $\pm$ 0.04	0.45 $\pm$ 0.04	0.19 $\pm$ 0.04	0.36 $\pm$ 0.02
	GBM	0.28 $\pm$ 0.06	0.21 $\pm$ 0.04	0.37 $\pm$ 0.06	0.16 $\pm$ 0.03	0.78 $\pm$ 0.03
	LSTM [†]	0.27 $\pm$ 0.04	0.21 $\pm$ 0.03	0.55 $\pm$ 0.10	—	—
	MLP [†]	0.29 $\pm$ 0.05	0.23 $\pm$ 0.04	0.48 $\pm$ 0.07	—	—
	RGNN	0.24 $\pm$ 0.04	0.19 $\pm$ 0.03	0.52 $\pm$ 0.08	0.14 $\pm$ 0.02	0.97 $\pm$ 0.02
	RGNN-LSOA	0.24 $\pm$ 0.05	0.19 $\pm$ 0.04	0.56 $\pm$ 0.05	0.14 $\pm$ 0.02	0.99 $\pm$ 0.01
	RGNN-OA	0.23 $\pm$ 0.04	0.19 $\pm$ 0.03	0.56 $\pm$ 0.05	0.14 $\pm$ 0.02	0.98 $\pm$ 0.01
2022	DLNM	0.54 $\pm$ 0.09	0.44 $\pm$ 0.07	0.49 $\pm$ 0.11	0.36 $\pm$ 0.05	0.43 $\pm$ 0.06
	RF	0.52 $\pm$ 0.08	0.43 $\pm$ 0.06	0.46 $\pm$ 0.07	0.40 $\pm$ 0.06	0.11 $\pm$ 0.03
	GBM	0.52 $\pm$ 0.08	0.43 $\pm$ 0.06	0.40 $\pm$ 0.08	0.35 $\pm$ 0.05	0.39 $\pm$ 0.07
	LSTM [†]	0.53 $\pm$ 0.07	0.46 $\pm$ 0.06	0.40 $\pm$ 0.13	—	—
	MLP [†]	0.53 $\pm$ 0.08	0.46 $\pm$ 0.07	0.48 $\pm$ 0.11	—	—
	RGNN	0.39 $\pm$ 0.04	0.31 $\pm$ 0.03	0.47 $\pm$ 0.07	0.22 $\pm$ 0.02	0.87 $\pm$ 0.03
	RGNN-LSOA	0.39 $\pm$ 0.03	0.32 $\pm$ 0.02	0.40 $\pm$ 0.07	0.22 $\pm$ 0.02	0.89 $\pm$ 0.03
	RGNN-OA	0.43 $\pm$ 0.04	0.36 $\pm$ 0.03	0.40 $\pm$ 0.07	0.25 $\pm$ 0.02	0.83 $\pm$ 0.04

Forests (RF), Gradient Boosted Models (GBM), Long Short-Term Memory networks (LSTMs) and Multi-Layer Perceptrons (MLP). RFs report their own uncertainty estimates through their ensemble, and for GBMs we use heteroscedastic noise on training residuals ($\text{std}(y - \hat{y})$).

We take two holdout sets in 2018 and 2022, both unprecedented years for heat exposure in the UK. We detrend for long-term trend, seasonality, day-of-week effects, and public holidays using linear and Fourier terms. The outcome y_{it} is a Pearson mortality residual; these are considered our ‘excess deaths’ not covered by these simple detrending terms, including heat-related mortality. We then normalise units to excess deaths per 100k population. For each holdout year and model we report RMSE, MAE and Pearson-correlation, as well as CRPS and Coverage@90% where uncertainty is available. Hyper-parameters for each model are optimised through a Bayesian search over RMSE with 2017 as validation year, though as is standard practise in environmental epidemiology we use the DLNM parameters from Lo et al. [11], a government-cited study on heat risk in the UK.

All three RGNN variants substantially outperform DLNM on RMSE in both holdout years. In the Supplementary we provide per-region results and example risk curves from the vanilla DLNM alongside the optimised curves from our RGNNs. Performance does not monotonically increase with hierarchy depth (RGNN > RGNN-LSOA < RGNN-OA). This may be an artifact of hyper-parameter selection, and is an area of ongoing research. The most striking result is the behaviour of predictive coverage under distributional shift. In the moderate

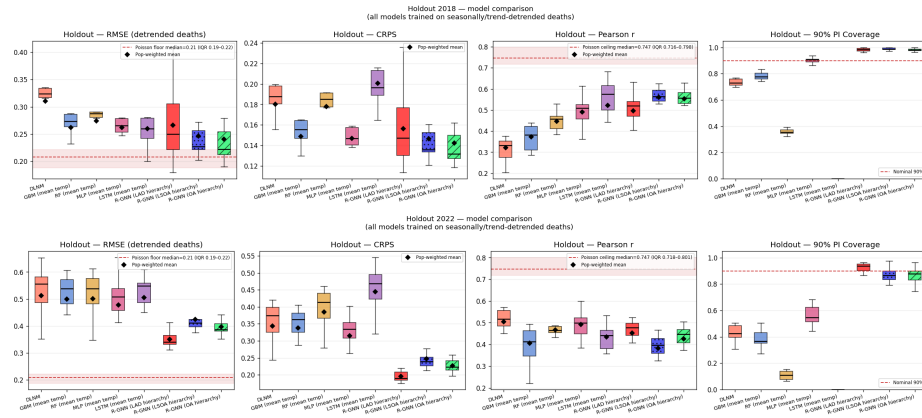


Fig. 3. RMSE ($\downarrow$), CRPS ($\downarrow$), Correlation ($\uparrow$) and Coverage@90% for the holdout years 2018 (top) and 2022 (bottom).

2018 holdout, all models with uncertainty estimates achieve reasonable coverage. In 2022, however, the DLNM’s coverage collapses to well below the nominal 90% level—while the RGNN variants maintain coverage between 0.83 and 0.89. For a public health tool intended to support decision-making during extreme heat events, calibration under tail conditions is arguably more consequential than average-year point accuracy.

As an early set of results, we vary census features at OA and examine the changes in predictions made using the RGNN-OA. We perform two experiments here; increasing the proportion of OA population with low educational qualifications (level 1/2), as well as the proportion of young adults (25-29), by one standard deviation upwards. We perform this analysis for the London region using data for the 21st July 2022, the hottest day on-record in the UK. While this does not establish a causal link, with future research it may prove a useful interrogative tool.

Chloropleth visualisation in Figure 4 shows heterogeneity across spatial areas, and a sweep over delta magnitudes in Figure 5 (Appendix) shows the expected relationships. Increased proportions of population with low educational levels increases mortality, and increased proportions of young people decrease it, in-line with epidemiological literature on deprivation and clinical risks [12]. The North East in-general shows greater sensitivity to demographic changes, particularly in education levels, possibly due to the broadly lower socio-economic status of the region in-comparison to reason. This spatially-resolved attribution, mapping demographic vulnerability onto urban geography, exemplifies the kind of actionable, area-level intelligence that can directly inform local heat-health interventions. Further analysis will follow in future iterations of this work.

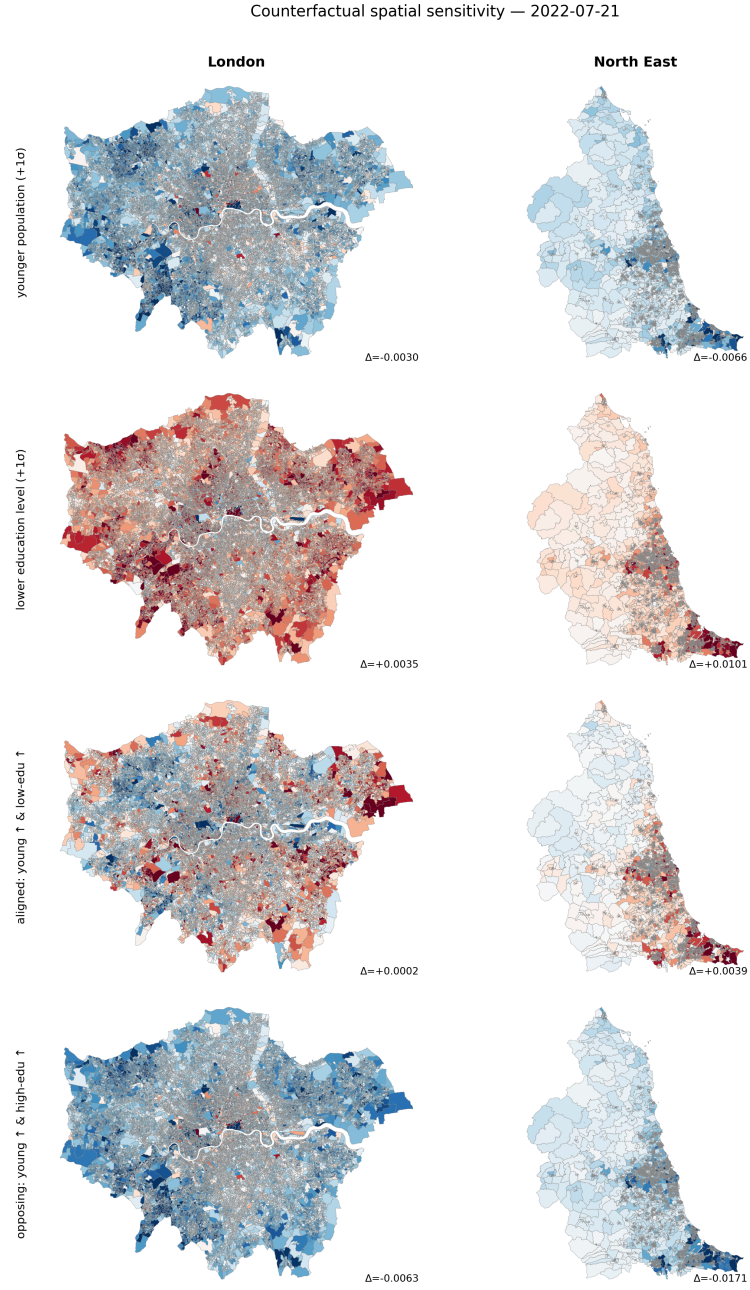


Fig. 4. Δ mortality with varying census features for the RGNN-OA model on the 21st July 2022 in the London and North East regions.

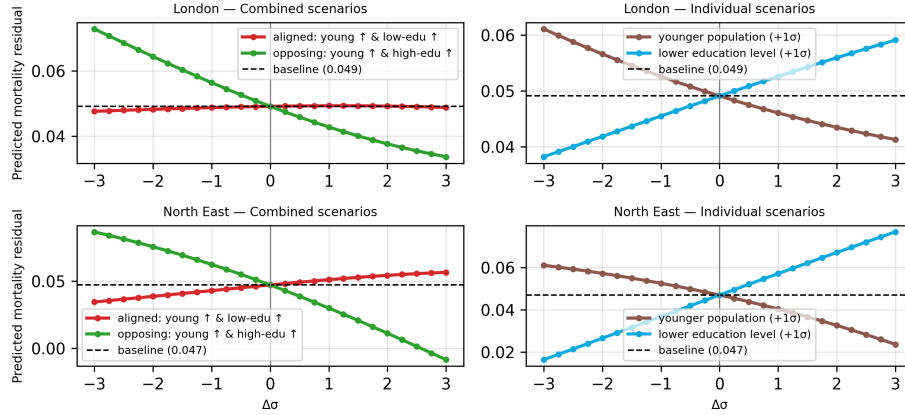


Fig. 5. Δ predictions with varying proportions of young and lower-educated people in OAs, as predicted by the RGNN-OA in London.

4.1 Discussion

The most consequential result is calibration: where DLNM and RF coverage collapse during the 2022 heatwave, all RGNN variants maintain near-nominal coverage. For a public health tool intended to support decision-making during the events where it is needed most, this robustness under distributional shift is a meaningful contribution.

Early counterfactual experiments provide evidence that the model is learning epidemiologically coherent structure rather than spurious correlations, with deprivation and age effects in the expected directions. More systematic feature attribution analysis will follow in future iterations. A natural extension is to assess whether the learned demographic embeddings transfer across geographic contexts, enabling risk curve estimation for areas with limited mortality data.

5 Conclusion

We introduced Risk Graph Neural Networks, which augment classical DLMs with a hierarchical GNN encoder over census features, preserving interpretable risk curve outputs while substantially improving predictive calibration under extreme heat. Evaluated across 10 regions of England and Wales, RGNN variants maintain near-nominal uncertainty coverage where standard baselines collapse. Future work will broaden the geographic scope, investigate adaptive hierarchy depth, and provide fuller feature attribution analysis. We also intend to explore whether representations learned over England and Wales generalise to other health outcome data.

Acknowledgments. A. O. Davies, E. Lo and R. Zhu thank the University of Bristol and respective grants for their funding.

Bibliography

- [1] J. Boudreault, A. Ruf, C. Campagna, and F. Chebana. Multi-region models built with machine and deep learning for predicting several heat-related health outcomes. *Sustainable Cities and Society*, 115:105785, Nov. 2024. ISSN 2210-6707. <https://doi.org/10.1016/j.scs.2024.105785>. URL <https://www.sciencedirect.com/science/article/pii/S2210670724006097>.
- [2] J. Boudreault, F. Lamothe, C. Campagna, and F. Chebana. Machine learning for modelling the health impacts of extreme heat: A comprehensive literature review. *Environment International*, 206:109965, Dec. 2025. ISSN 0160-4120. <https://doi.org/10.1016/j.envint.2025.109965>. URL <https://www.sciencedirect.com/science/article/pii/S0160412025007160>.
- [3] E. Fillola, R. Santos-Rodriguez, R. Tunnicliffe, J. N. Clark, N. Keshtmand, A. Ganesan, and M. Rigby. Enabling fast greenhouse gas emissions inference from satellites with GATES: a Graph-Neural-Network Atmospheric Transport Emulation System. *Geoscientific Model Development*, 19(5):1893–1915, Mar. 2026. ISSN 1991-959X. <https://doi.org/10.5194/gmd-19-1893-2026>. URL <https://gmd.copernicus.org/articles/19/1893/2026/>.
- [4] A. Gasparrini, B. Armstrong, and M. G. Kenward. Distributed lag non-linear models. *Statistics in Medicine*, 29(21):2224–2234, 2010. ISSN 1097-0258. <https://doi.org/10.1002/sim.3940>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.3940>. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.3940>.
- [5] O. Goodwin and A. Singleton. A unified dataset of UK census variables for small areas: Harmonised data tables from the 2021 England, Wales, and Northern Ireland censuses and the 2022 Scotland census. *Environment and Planning B: Urban Analytics and City Science*, page 23998083261429563, Feb. 2026. ISSN 2399-8083. <https://doi.org/10.1177/23998083261429563>. URL <https://doi.org/10.1177/23998083261429563>.
- [6] W. L. Hamilton, R. Ying, and J. Leskovec. Representation learning on graphs: Methods and applications. *arXiv*, 1(1):1–24, Sept. 2017. URL <http://arxiv.org/abs/1709.05584>.
- [7] D. Hollis, M. McCarthy, M. Kendon, T. Legg, and I. Simpson. HadUK-Grid—A new UK dataset of gridded climate observations. *Geoscience Data Journal*, 6(2):151–159, 2019. ISSN 2049-6060. <https://doi.org/10.1002/gdj3.78>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/gdj3.78>. _eprint: <https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/gdj3.78>.
- [8] W. Jiang and J. Luo. Graph neural network for traffic forecasting: A survey, Nov. 2022.
- [9] T. N. Kipf and M. Welling. Semi-Supervised Classification with Graph Convolutional Networks. In *In Proceedings of the 5th International Conference on Learning Representations (ICLR)*, Apr. 2017.

- [10] L. I.-K. Lin. A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics*, 45(1):255–268, 1989. ISSN 0006-341X. <https://doi.org/10.2307/2532051>. URL <https://www.jstor.org/stable/2532051>.
- [11] Y. T. E. Lo, D. M. Mitchell, R. Thompson, E. O’Connell, and A. Gasparri. Estimating heat-related mortality in near real time for national heatwave plans. *Environmental Research Letters*, 17(2):024017, Feb. 2022. ISSN 1748-9326. <https://doi.org/10.1088/1748-9326/ac4cf4>. URL <https://doi.org/10.1088/1748-9326/ac4cf4>.
- [12] Y. T. E. Lo, D. M. Mitchell, and A. Gasparri. Compound mortality impacts from extreme temperatures and the COVID-19 pandemic. *Nature Communications*, 15(1):4289, May 2024. ISSN 2041-1723. <https://doi.org/10.1038/s41467-024-48207-2>. URL <https://www.nature.com/articles/s41467-024-48207-2>.
- [13] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio. Graph Attention Networks. In *In Proceedings of the 6th International Conference on Learning Representations (ICLR)*, Feb. 2018. <https://doi.org/10.48550/arXiv.1710.10903>.

A Appendix

Table 2: Per-region holdout results for all models across both holdout years. RMSE↓, MAE↓, Pearson r ↑, CRPS↓, and Coverage-90 are shown for each region under the 2018 (moderate) and 2022 (severe heatwave) holdout years. **Bold** indicates the best result per metric in each region-year block. For Coverage-90, best is defined as closest to the nominal 0.90 level. The Poisson noise floor (fl) and Pearson ceiling (cl) are shown for reference alongside each region header. †CRPS and Coverage-90 are not reported for LSTM and MLP, as these models do not produce well-defined predictive distributions.

Region	Model	RMSE↓	MAE↓	Pearson↑	CRPS↓	Cov-90
North East — 2018 $fl = 0.327$, $cl = 0.630$						
	DLNM	0.455	0.346	0.204	0.265	0.715
	RF	0.384	0.286	0.405	0.251	0.342
	GBM	0.360	0.272	0.286	0.203	0.808
	LSTM†	0.372	0.296	0.540	—	—
	MLP†	0.413	0.332	0.429	—	—
	RGNN	0.409	0.341	0.266	0.236	0.962
	RGNN _L	0.378	0.306	0.432	0.220	0.984

(continued on next page)

(continued from previous page)

Region	Model	RMSE↓	MAE↓	Pearson↑	CRPS↓	Cov-90
	RGNN _{OA}	0.366	0.298	0.426	0.212	0.970
North East — 2022 $fl = 0.326$, $cl = 0.631$						
	DLNM	0.653	0.521	0.161	0.420	0.504
	RF	0.594	0.482	0.287	0.445	0.148
	GBM	0.594	0.482	0.221	0.382	0.506
	LSTM [†]	0.527	0.449	0.240	—	—
	MLP [†]	0.654	0.534	0.179	—	—
	RGNN	0.362	0.260	0.255	0.202	0.964
	RGNN _L	0.393	0.306	0.213	0.223	0.975
	RGNN _{OA}	0.379	0.285	0.247	0.211	0.964
North West — 2018 $fl = 0.188$, $cl = 0.798$						
	DLNM	0.321	0.245	0.258	0.187	0.737
	RF	0.287	0.211	0.471	0.185	0.351
	GBM	0.271	0.205	0.303	0.155	0.758
	LSTM [†]	0.249	0.194	0.612	—	—
	MLP [†]	0.275	0.212	0.362	—	—
	RGNN	0.291	0.242	0.459	0.171	0.997
	RGNN _L	0.228	0.180	0.566	0.140	0.997
	RGNN _{OA}	0.230	0.183	0.558	0.139	0.989
North West — 2022 $fl = 0.187$, $cl = 0.801$						
	DLNM	0.587	0.491	0.484	0.410	0.307
	RF	0.582	0.488	0.486	0.461	0.071
	GBM	0.576	0.485	0.494	0.406	0.272
	LSTM [†]	0.581	0.523	0.534	—	—
	MLP [†]	0.506	0.435	0.554	—	—
	RGNN	0.336	0.248	0.485	0.184	0.959
	RGNN _L	0.425	0.356	0.397	0.244	0.882
	RGNN _{OA}	0.397	0.320	0.446	0.223	0.893
Yorkshire — 2018 $fl = 0.215$, $cl = 0.743$						
	DLNM	0.330	0.242	0.262	0.189	0.721
	RF	0.289	0.212	0.459	0.185	0.351
	GBM	0.275	0.206	0.295	0.156	0.783
	LSTM [†]	0.239	0.182	0.618	—	—
	MLP [†]	0.270	0.209	0.508	—	—
	RGNN	0.319	0.270	0.403	0.185	0.981
	RGNN _L	0.265	0.213	0.539	0.157	0.986
	RGNN _{OA}	0.262	0.210	0.521	0.153	0.981
Yorkshire — 2022 $fl = 0.215$, $cl = 0.743$						
	DLNM	0.558	0.452	0.485	0.372	0.414

(continued on next page)

(continued from previous page)

Region	Model	RMSE↓	MAE↓	Pearson↑	CRPS↓	Cov-90
	RF	0.536	0.436	0.433	0.408	0.115
	GBM	0.532	0.435	0.447	0.358	0.369
	LSTM [†]	0.545	0.467	0.435	—	—
	MLP [†]	0.535	0.456	0.437	—	—
	RGNN	0.312	0.235	0.442	0.175	0.953
	RGNN _L	0.375	0.309	0.347	0.213	0.934
	RGNN _{OA}	0.351	0.278	0.404	0.196	0.923
East Midlands — 2018 $fl = 0.221$, $cl = 0.716$						
	DLNM	0.335	0.262	0.346	0.199	0.710
	RF	0.290	0.218	0.425	0.190	0.392
	GBM	0.283	0.212	0.335	0.162	0.742
	LSTM [†]	0.240	0.188	0.618	—	—
	MLP [†]	0.279	0.221	0.526	—	—
	RGNN	0.256	0.210	0.503	0.150	0.986
	RGNN _L	0.226	0.184	0.551	0.136	0.989
	RGNN _{OA}	0.225	0.182	0.545	0.133	0.986
East Midlands — 2022 $fl = 0.222$, $cl = 0.718$						
	DLNM	0.548	0.439	0.561	0.359	0.449
	RF	0.535	0.432	0.454	0.402	0.101
	GBM	0.537	0.433	0.340	0.353	0.386
	LSTM [†]	0.553	0.469	0.358	—	—
	MLP [†]	0.539	0.457	0.482	—	—
	RGNN	0.339	0.254	0.495	0.184	0.945
	RGNN _L	0.415	0.342	0.396	0.237	0.866
	RGNN _{OA}	0.389	0.310	0.449	0.217	0.868
West Midlands — 2018 $fl = 0.202$, $cl = 0.751$						
	DLNM	0.327	0.251	0.353	0.194	0.699
	RF	0.291	0.218	0.529	0.192	0.345
	GBM	0.288	0.215	0.426	0.165	0.764
	LSTM [†]	0.253	0.205	0.683	—	—
	MLP [†]	0.247	0.193	0.555	—	—
	RGNN	0.238	0.193	0.519	0.144	1.000
	RGNN _L	0.214	0.173	0.562	0.133	1.000
	RGNN _{OA}	0.210	0.169	0.557	0.130	1.000
West Midlands — 2022 $fl = 0.202$, $cl = 0.751$						
	DLNM	0.554	0.457	0.533	0.376	0.375
	RF	0.541	0.450	0.530	0.420	0.068
	GBM	0.539	0.447	0.469	0.369	0.325
	LSTM [†]	0.559	0.497	0.460	—	—

(continued on next page)

(continued from previous page)

Region	Model	RMSE↓	MAE↓	Pearson↑	CRPS↓	Cov-90
	MLP [†]	0.481	0.405	0.527	—	—
	RGNN	0.330	0.239	0.508	0.180	0.948
	RGNN _L	0.397	0.325	0.433	0.224	0.904
	RGNN _{OA}	0.383	0.304	0.471	0.213	0.901
East of England — 2018 $f_l = 0.191$, $cl = 0.753$						
	DLNM	0.319	0.238	0.348	0.182	0.742
	RF	0.285	0.205	0.407	0.179	0.378
	GBM	0.265	0.194	0.395	0.147	0.769
	LSTM [†]	0.281	0.216	0.494	—	—
	MLP [†]	0.261	0.199	0.508	—	—
	RGNN	0.215	0.164	0.579	0.126	0.973
	RGNN _L	0.220	0.172	0.597	0.130	0.973
	RGNN _{OA}	0.220	0.168	0.590	0.127	0.964
East of England — 2022 $f_l = 0.191$, $cl = 0.756$						
	DLNM	0.445	0.359	0.571	0.290	0.496
	RF	0.438	0.354	0.476	0.326	0.153
	GBM	0.442	0.355	0.327	0.288	0.447
	LSTM [†]	0.450	0.388	0.365	—	—
	MLP [†]	0.412	0.358	0.383	—	—
	RGNN	0.327	0.269	0.407	0.186	0.901
	RGNN _L	0.398	0.342	0.324	0.234	0.827
	RGNN _{OA}	0.383	0.325	0.373	0.224	0.816
London — 2018 $f_l = 0.135$, $cl = 0.801$						
	DLNM	0.201	0.159	0.375	0.119	0.764
	RF	0.182	0.142	0.454	0.123	0.367
	GBM	0.173	0.136	0.427	0.101	0.797
	LSTM [†]	0.200	0.165	0.442	—	—
	MLP [†]	0.183	0.145	0.525	—	—
	RGNN	0.180	0.150	0.633	0.114	1.000
	RGNN _L	0.225	0.192	0.625	0.136	1.000
	RGNN _{OA}	0.190	0.160	0.629	0.118	1.000
London — 2022 $f_l = 0.134$, $cl = 0.819$						
	DLNM	0.352	0.299	0.450	0.243	0.392
	RF	0.347	0.301	0.459	0.279	0.066
	GBM	0.343	0.294	0.363	0.239	0.350
	LSTM [†]	0.351	0.320	0.466	—	—
	MLP [†]	0.334	0.299	0.599	—	—
	RGNN	0.366	0.328	0.471	0.218	0.901
	RGNN _L	0.420	0.381	0.448	0.254	0.847

(continued on next page)

(continued from previous page)

Region	Model	RMSE↓	MAE↓	Pearson↑	CRPS↓	Cov-90
	RGNN _{OA}	0.379	0.341	0.468	0.227	0.874
South East — 2018 $fl = 0.157$, $cl = 0.807$						
	DLNM	0.272	0.205	0.315	0.156	0.770
	RF	0.246	0.181	0.383	0.156	0.359
	GBM	0.232	0.172	0.370	0.130	0.806
	LSTM [†]	0.275	0.199	0.300	—	—
	MLP [†]	0.252	0.200	0.442	—	—
	RGNN	0.198	0.154	0.519	0.119	0.986
	RGNN _L	0.202	0.161	0.530	0.121	0.984
	RGNN _{OA}	0.203	0.159	0.526	0.119	0.975
South East — 2022 $fl = 0.157$, $cl = 0.808$						
	DLNM	0.466	0.382	0.569	0.314	0.419
	RF	0.464	0.380	0.462	0.354	0.126
	GBM	0.465	0.381	0.430	0.316	0.356
	LSTM [†]	0.474	0.414	0.426	—	—
	MLP [†]	0.449	0.393	0.510	—	—
	RGNN	0.340	0.282	0.442	0.194	0.888
	RGNN _L	0.407	0.351	0.394	0.242	0.792
	RGNN _{OA}	0.412	0.354	0.418	0.247	0.759
South West — 2018 $fl = 0.213$, $cl = 0.740$						
	DLNM	0.316	0.239	0.319	0.180	0.770
	RF	0.279	0.205	0.469	0.177	0.375
	GBM	0.263	0.195	0.439	0.147	0.833
	LSTM [†]	0.266	0.193	0.518	—	—
	MLP [†]	0.258	0.191	0.510	—	—
	RGNN	0.244	0.196	0.543	0.141	0.970
	RGNN _L	0.231	0.178	0.580	0.134	0.984
	RGNN _{OA}	0.220	0.170	0.565	0.127	0.981
South West — 2022 $fl = 0.214$, $cl = 0.741$						
	DLNM	0.570	0.462	0.544	0.378	0.430
	RF	0.557	0.456	0.468	0.426	0.090
	GBM	0.560	0.460	0.394	0.377	0.361
	LSTM [†]	0.558	0.487	0.448	—	—
	MLP [†]	0.509	0.439	0.511	—	—
	RGNN	0.379	0.294	0.524	0.209	0.866
	RGNN _L	0.521	0.447	0.467	0.323	0.630
	RGNN _{OA}	0.441	0.362	0.505	0.259	0.745
Wales — 2018 $fl = 0.303$, $cl = 0.651$						
	DLNM	0.451	0.345	0.376	0.265	0.701

(continued on next page)

(continued from previous page)

Region	Model	RMSE↓	MAE↓	Pearson↑	CRPS↓	Cov-90
	RF	0.386	0.291	0.474	0.257	0.321
	GBM	0.380	0.288	0.411	0.219	0.756
	LSTM [†]	0.340	0.271	0.611	—	—
	MLP [†]	0.361	0.295	0.615	—	—
	RGNN	0.310	0.246	0.534	0.179	0.973
	RGNN _L	0.272	0.216	0.616	0.160	0.989
	RGNN _{OA}	0.279	0.219	0.610	0.162	0.981
Wales	— 2022 $fl = 0.302$, $cl = 0.651$					
	DLNM	0.637	0.504	0.502	0.405	0.496
	RF	0.612	0.483	0.543	0.447	0.142
	GBM	0.607	0.480	0.490	0.385	0.461
	LSTM [†]	0.610	0.546	0.475	—	—
	MLP [†]	0.619	0.545	0.488	—	—
	RGNN	0.412	0.302	0.499	0.219	0.926
	RGNN _L	0.486	0.399	0.413	0.277	0.868
	RGNN _{OA}	0.456	0.358	0.468	0.252	0.882

fl = Poisson RMSE noise floor; cl = Pearson ceiling. RGNN_L = RGNN-LSOA;
 RGNN_{OA} = RGNN-OA.

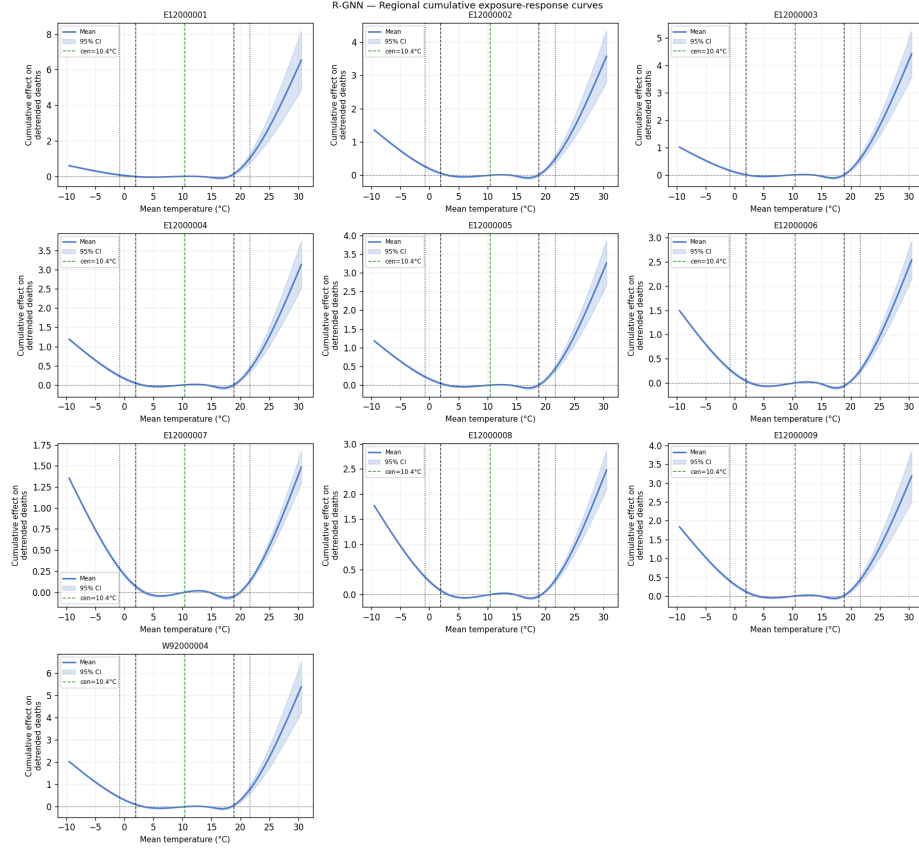


Fig. 6. Vanilla DLNM curves for each UK region in our experiments. Important to note in contrast with the optimised curves below is the flat portion at moderate temperatures, with a ‘bump’ due to the combination of spline specification and heat risk falling below detectable SNR levels.

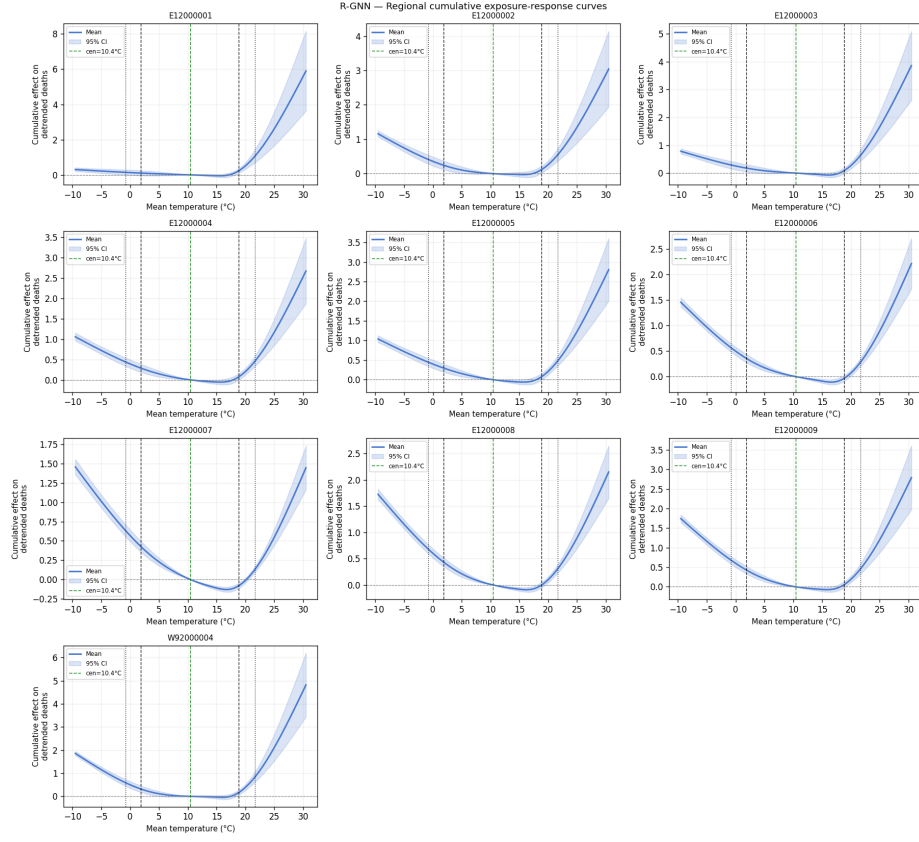


Fig. 7. DLNM curves for each UK region in our experiments, optimised by RGNN-LAD.

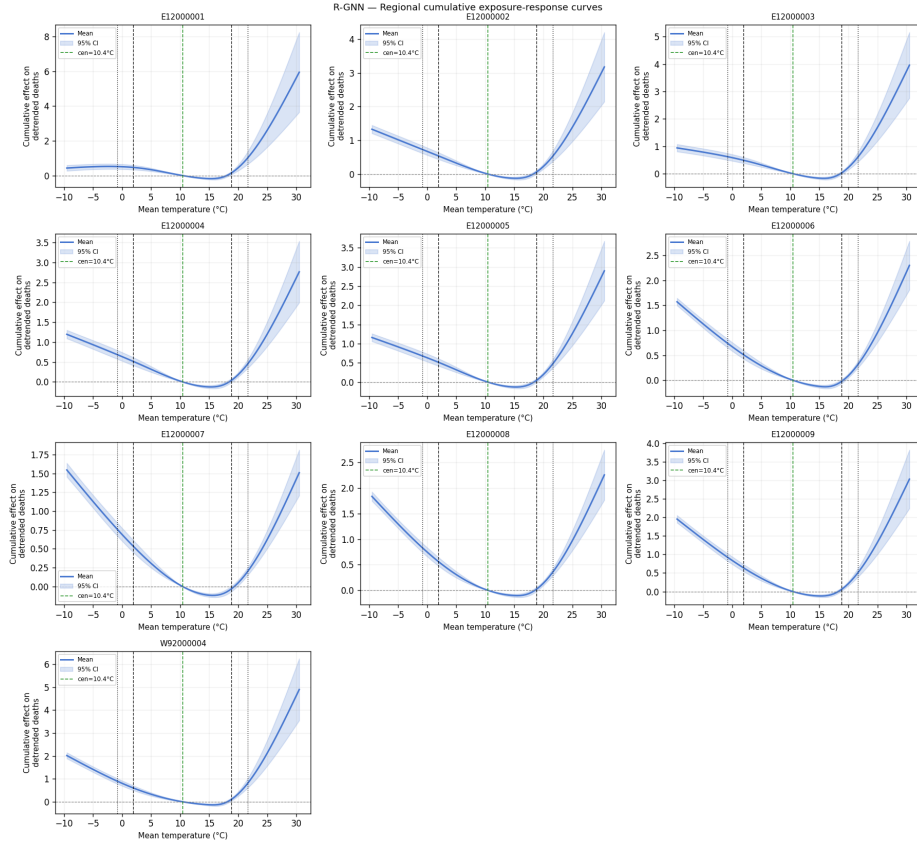


Fig. 8. DLNM curves for each UK region in our experiments, optimised by RGNN-LSOA.

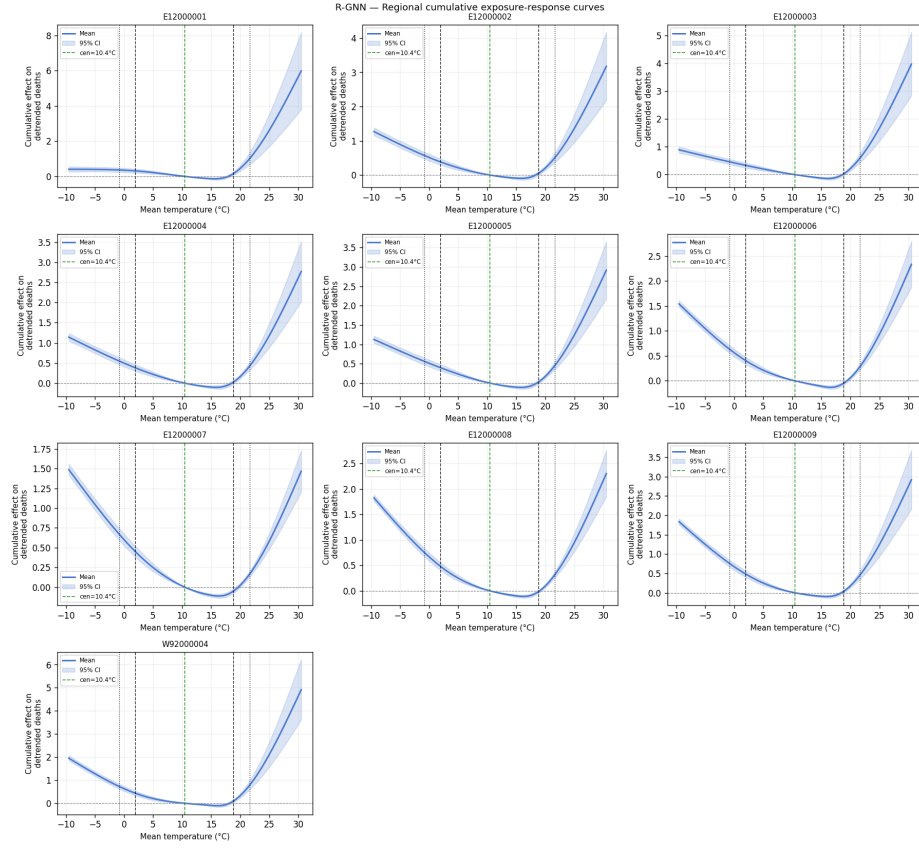


Fig. 9. DLNM curves for each UK region in our experiments, optimised by RGNN-OA.