

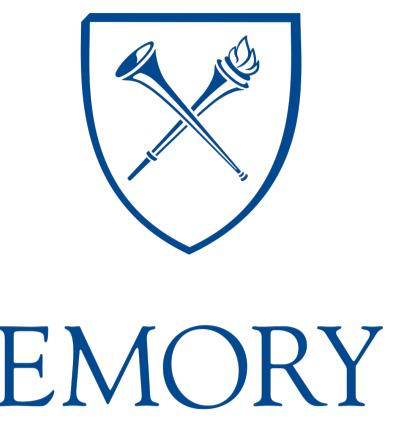
# Neural Network Reveals Gravitational Coupling of Endemic Measles Dynamics



Wyatt G. Madden<sup>†</sup>, Jessica E. Metcalf<sup>‡</sup>, Bryan T. Grenfell<sup>‡</sup>, Max S.Y. Lau<sup>†</sup>

<sup>†</sup>Department of Biostatistics & Bioinformatics, Emory University,

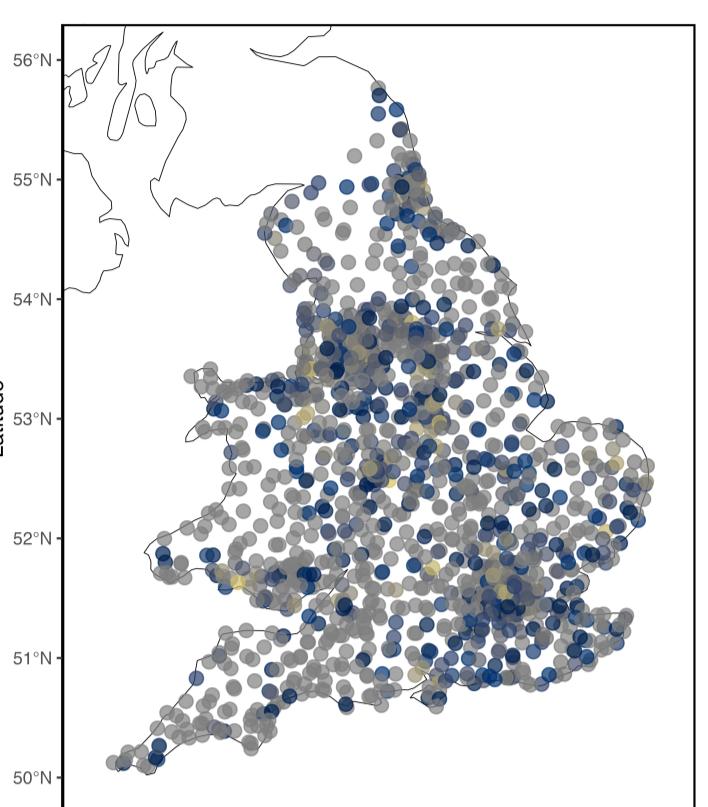
<sup>‡</sup>Princeton University, NJ, USA



## Introduction

Measles is an important infectious disease system, both for its immediate relevance to public health and for the study of non-linear spatio-temporal disease dynamics. Traditional mechanistic models are often unable to fully capture the complicated nonlinear spatio-temporal dynamics inherent in measles transmission. Here we develop a high-dimensional neural-network-based model to forecast endemic measles outbreaks, with the aim of providing both prediction accuracy and interpretability and inference of mechanistic drivers.

Fig. 1: Map of England and Wales, with cities/towns colored by log measles cases on first biweek of 1961.



We demonstrate forecast accuracy for high and low population cities/towns for time steps ahead ranging from 1 to 52 biweeks, and employ feature importance methods to demonstrate the learning of coupled gravity dynamics.

## Data & Methods

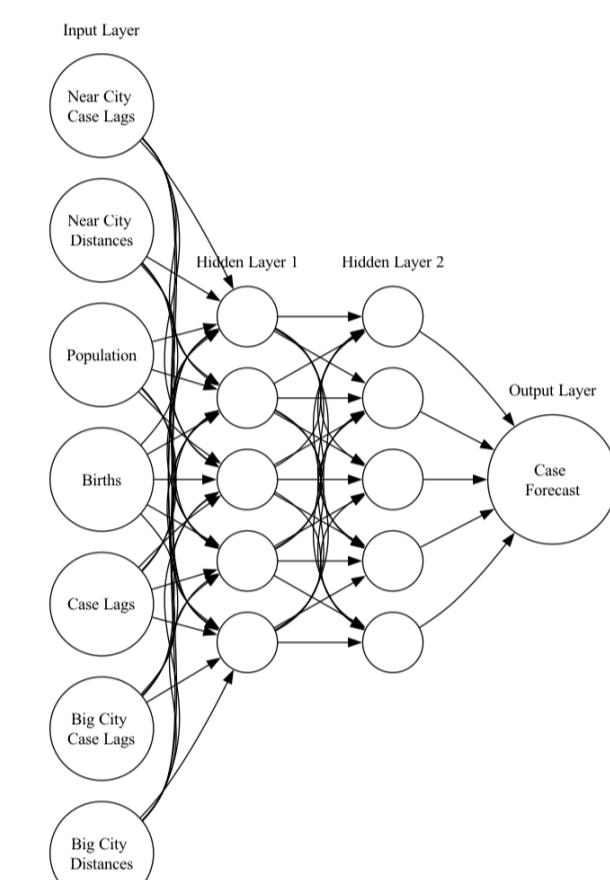


Fig. 2: Feed-forward neural network architecture.

Our data consists of biweekly measles case counts across 1,452 cities/towns in the United Kingdom during the pre-vaccination period from 1951 to 1964. The forecast goal was to predict  $k$ -step ahead case counts for all cities/towns, for  $k \in \{1, 2, \dots, 52\}$  using a range of features. For each  $k$ -step ahead we fit a separate feed-forward neural network with 2 hidden layers of dimension 962, linear input/output layers, and ReLu activation functions (Figure 2).

Features include lagged case counts, high population city lagged case counts and distances (defined as a population greater than 300k, of which there are seven), nearest ten city lagged case counts and distances, births, population. Lagged features range from  $t - k$  to  $t - 130$ , where  $t$  is the target time step. Birth and population features are from the nearest time step less than or equal to  $t - k$  while still sharing the same biweek of the year. Neural networks are fit using the Adam optimizer with MSE loss, and are trained on 70% of the data (cases ranging from 1951 to 1960), with 30% of the data (cases ranging from 1960 to 1964) held out for testing.

We compare the neural networks to TSIR (time-series susceptible-infected-recovered) models, a popular semi-mechanistic technique that has been shown to handle the dynamics of measles outbreaks well [1]. TSIR provides a computationally inexpensive and highly tractable alternative to the classic SIR compartmental model, and is described by the following equations:

$$E[I_{t+1}] = \beta_{t+1} I_t^\alpha S_t \quad S_{t+1} = B_{t+1} + S_t - I_{t+1}$$

## Performance

The neural network generally performed better than TSIR for all  $k$ -steps ahead, across different population sizes. While the neural network performed well for long forecasts in high-population cities (Figure 3), the level of improvement is most pronounced for small  $k$ -steps ahead in low population cities/towns that have sporadic and brief outbreaks (Figures 4,5).

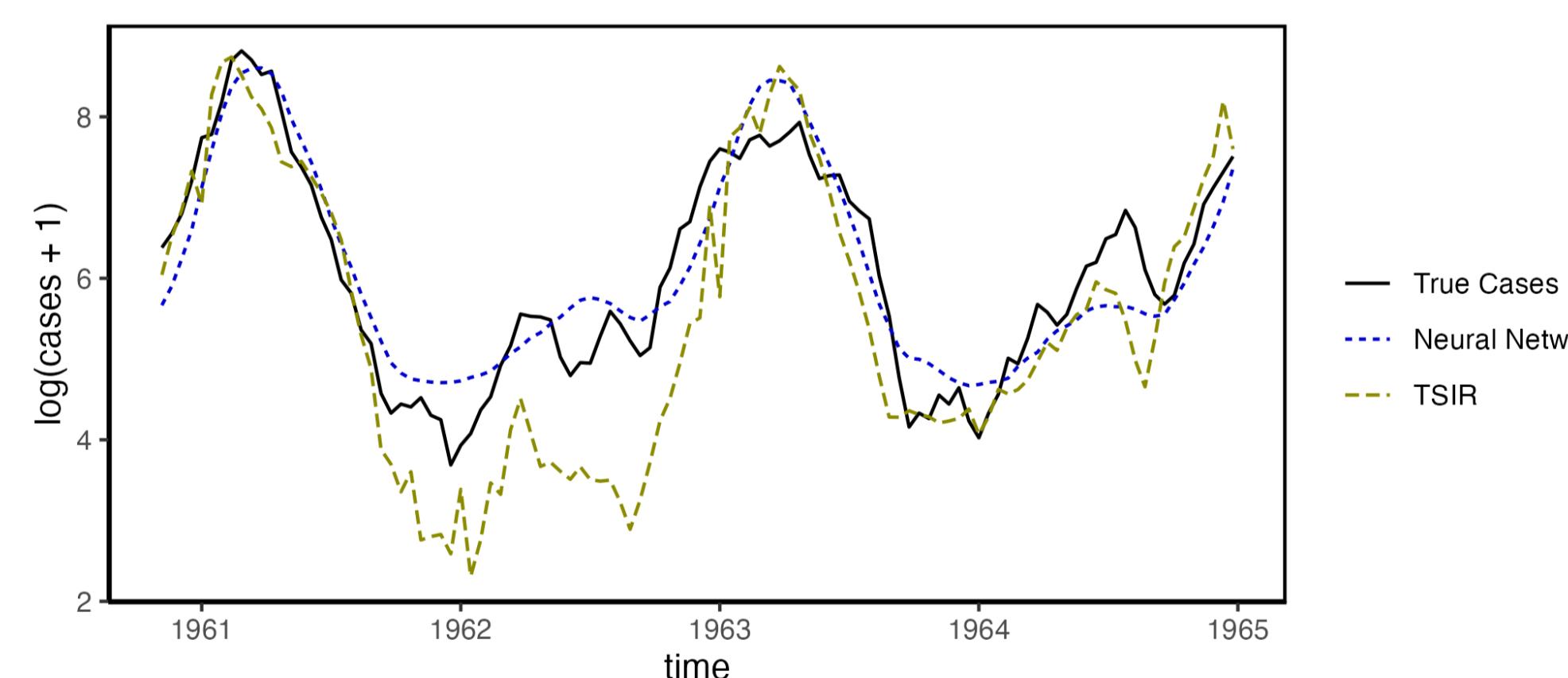


Fig. 3: 52-step ahead test set neural network forecasts, TSIR forecasts, and true case values for London and Manchester

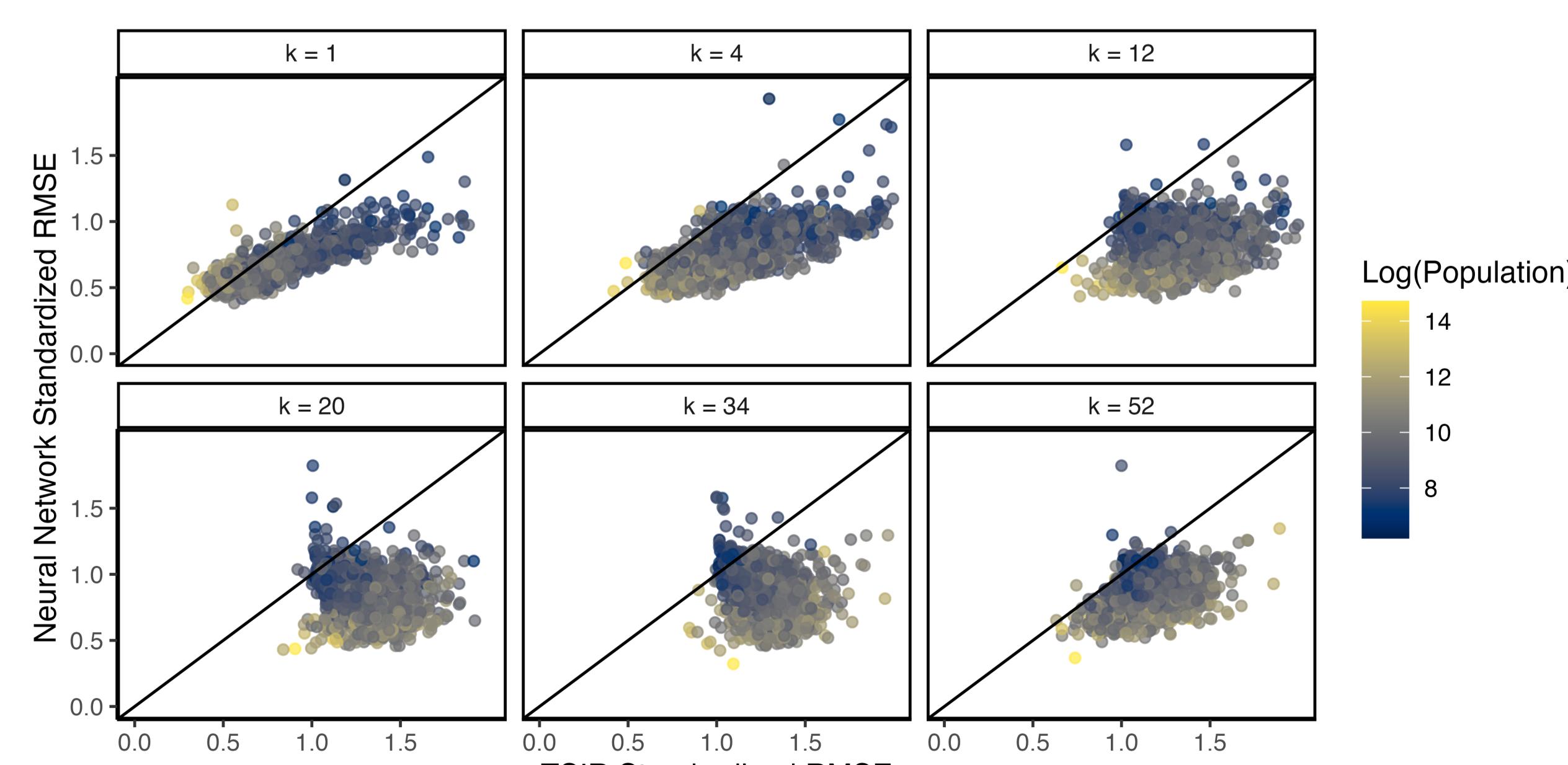


Fig. 4: Within-city neural network RMSE over TSIR RMSE, colored by log(population), faceted by  $k$ -step ahead forecast.

## Performance

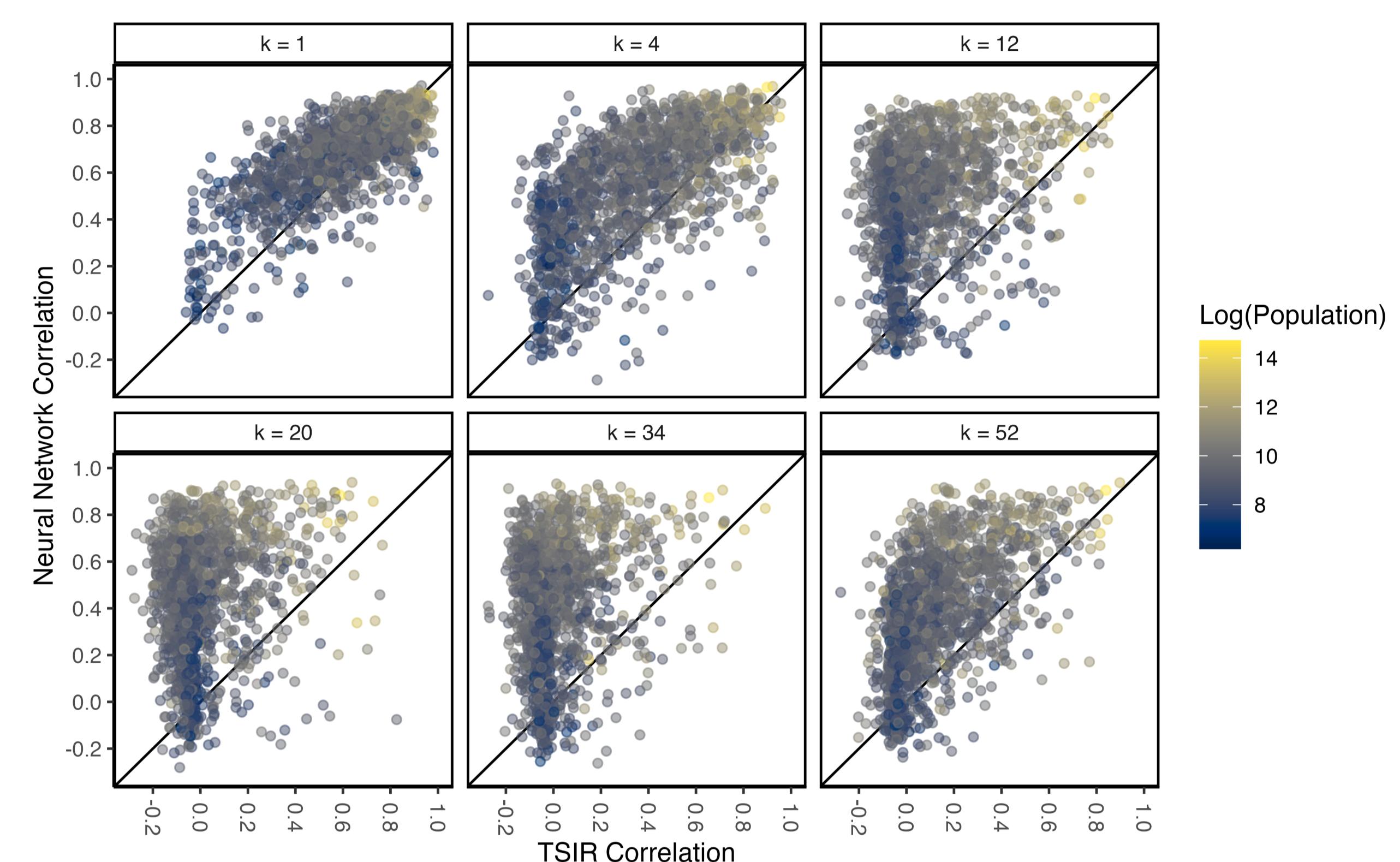


Fig. 5: Within-city neural network prediction correlation with true cases over TSIR prediction correlation with true cases, colored by log(population), faceted by  $k$ -step ahead forecast.

## Feature Importance

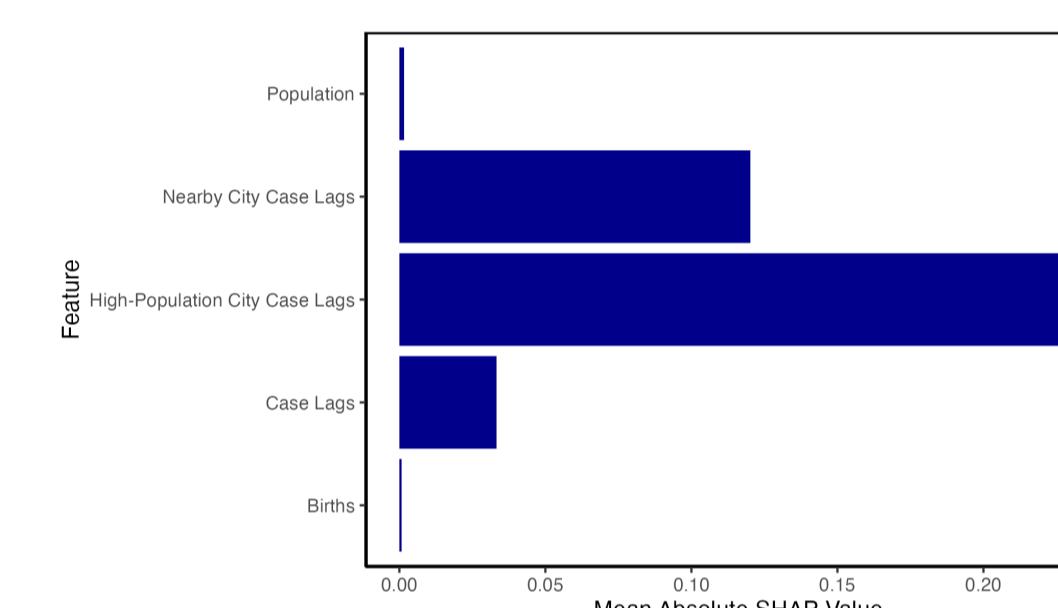


Fig. 6: Mean Absolute SHAP values by feature group.

We use the SHAP (SHapley Additive exPlanations) method [2] to assess feature importance. SHAP values are calculated by selected random permutations of feature groups, calculating the change in model output as each feature group is added back to a baseline value, and finally averaging across all permutations.

We choose to group features by lag type, e.g. all lagged nearest-city cases are grouped and permuted together. The mean absolute SHAP value for each feature group suggests that the most important features are the lagged case counts of high-population cities, followed by the lagged case counts of nearby cities/towns, and then the lagged case counts of the target city (Figure 6).

## Gravity Dynamics

Previous work has demonstrated the presence of gravity dynamics in measles outbreaks, where outbreaks in low population cities/towns are driven by outbreaks in nearby high population cities [3]. To assess if the neural network is learning high/low population epidemic coupling, we first calculate the relative absolute SHAP value for each feature group at each observation. We then calculate the within-city mean value of the relative absolute SHAP values. This provides a measure of the relative importance of each feature group for each city. Plotting these against the log of the population of the target city, we see that the lagged case counts of high population cities are relatively more important for low population cities/towns, which suggests that our neural-network is able to reveal gravity-like core-satellite dynamics [4] present in the endemic measles spatio-temporal dynamics (Figure 7).

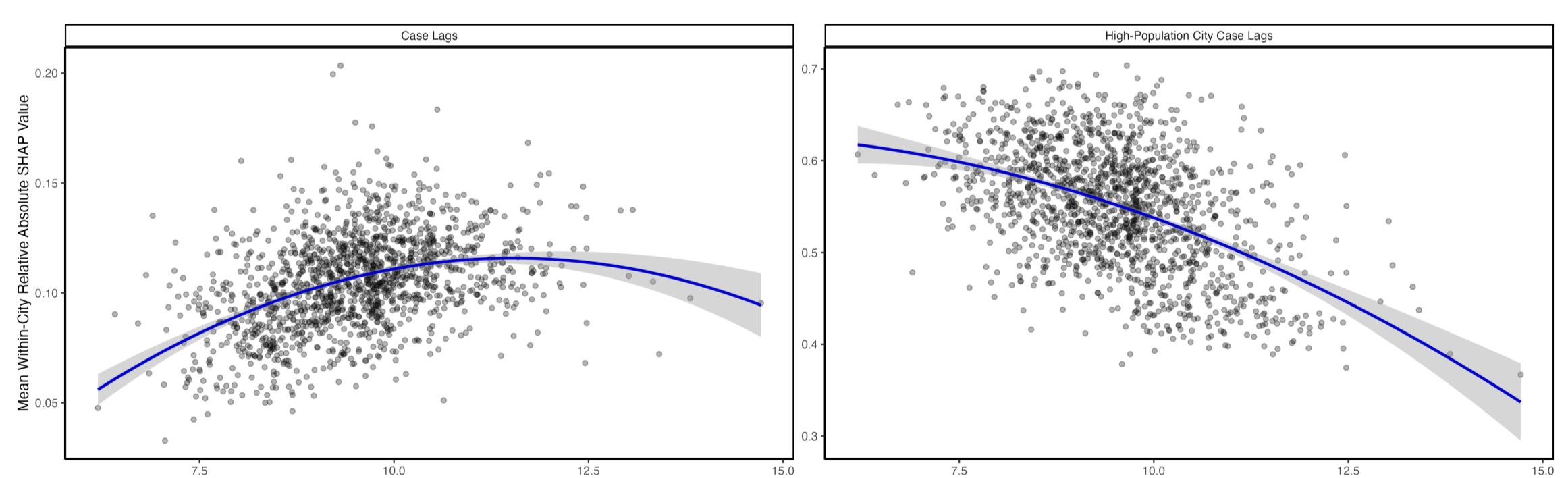


Fig. 7: With-city mean relative absolute SHAP values over log(population) for each feature group, with LOESS fits.

## Future Directions

- Extend analyses to more chaotic post-vaccination period.
- Explore joint approaches combining mechanistic models with neural networks, such as Physics-Informed Neural Networks (PINNs) and Neural-ODEs.

## Acknowledgements & References

This work is supported by the Dean's Pilot and Innovation Awards provided by the Dean's office at Rollins School of Public Health at Emory University.

- [1] Barbel F. Finkenstadt and Bryan T. Grenfell. "Time Series Modelling of Childhood Diseases: A Dynamical Systems Approach". In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 49.2 (2000), pp. 187–205.
- [2] Scott M. Lundberg and Su-Lee Lee. "A Unified Approach to Interpreting Model Predictions". In: *Proceedings of the 31st International Conference on Neural Information Processing Systems. NIPS'17*. Long Beach, California, USA: Curran Associates Inc., 2017, pp. 4768–4777.
- [3] Yingcun Xia et al. "Measles Metapopulation Dynamics: A Gravity Model for Epidemiological Coupling and Dynamics". In: *The American Naturalist* 164.2 (2004), pp. 267–281.
- [4] Max Lau et al. "A competing-risks model explains hierarchical spatial coupling of measles epidemics en route to national elimination". In: *Nature Ecology & Evolution* 4 (July 2020), pp. 1–6.