

552 Supplemental materials

553 Testing predictability of disease outbreaks with a simple model of pathogen 554 biogeography

555

556 Tad A. Dallas, Colin J. Carlson, Timothée Poisot

557 Further detail on the *plug-and-play* model

558 We present the *plug-and-play* model in the main text as a means to estimate
559 the suitability of a pathogen being found in a given country as a function of 1)
560 year, 2) pathogen community dissimilarity between countries (C_{ij}), 3) pathogen
561 distributional dissimilarity (P_{ij}), and 4) the product of these two values ($C_{ij} \times P_{ij}$).
562 Here, we provide more detail on the modeling approach. Further methodological
563 detail is provided elsewhere (36), as is the application of the approach to estimating
564 potentially missing links in host-parasite networks given host and pathogen trait
565 data (37).

566 The goal of the modeling approach is to estimate the conditional suitability of
567 a given country-pathogen pair given the set of observed ($y = 1$) and unobserved
568 ($y = 0$) country-pathogen pairs. That is, the suitability of a given pathogen and
569 country pair is given by

$$P(y = 1|x) = \frac{f_1(x)P(y = 1)}{f(x)}$$

570 where f_1 is the density of feature set $x_{i,j}$ given that $y_{i,j} = 1$, $P(y = 1)$, and f is
571 the density of all possible combined feature sets. The feature set $x_{i,j}$ here consists
572 of the four variables discussed above, relating to the dissimilarity of pathogen
573 communities in countries and pathogen distributions, in addition to the year for
574 which the dissimilarity values were estimated. We estimate suitability of a country-

575 pathogen pair as the ratio of these two densities

$$q = \frac{f_1}{f}$$

576 The *plug-and-play* model estimates f_1 and f separately to estimate q , using
577 the kernel density estimator `npudens` in the `np` package (38) and the “normal-
578 reference” bandwidth.

579 **Effect of rolling window size**

580 The size of the rolling window we used for model training prior to prediction
581 could influence model performance. To examine this possibility, we used a rolling
582 window of 7 years (compared to the 4 year window used in the main text), finding
583 qualitatively similar results when flu was included (Figure S1) or excluded (Figure
584 S2). We explored this further by examining rolling windows of 2, 4, and 6 years
585 (Figure S3), with qualitatively similar findings. For this analysis, we excluded
586 influenza, as we did in the main text.

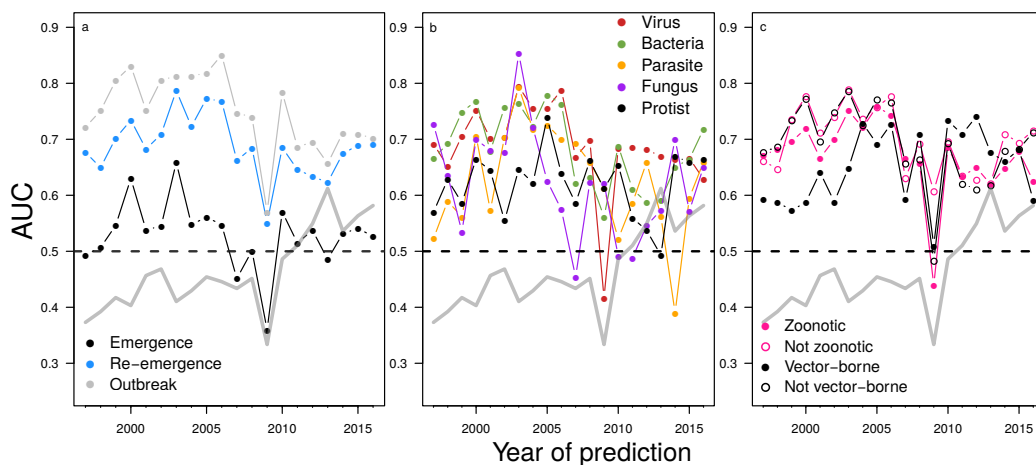


Figure S1: Rolling window size did not strongly influence model performance when considering next year prediction, as a window of 7 years produced qualitatively similar results to the window of 4 years we examine in the main text. The null expectation (grey line) performed consistently worse than our approach.

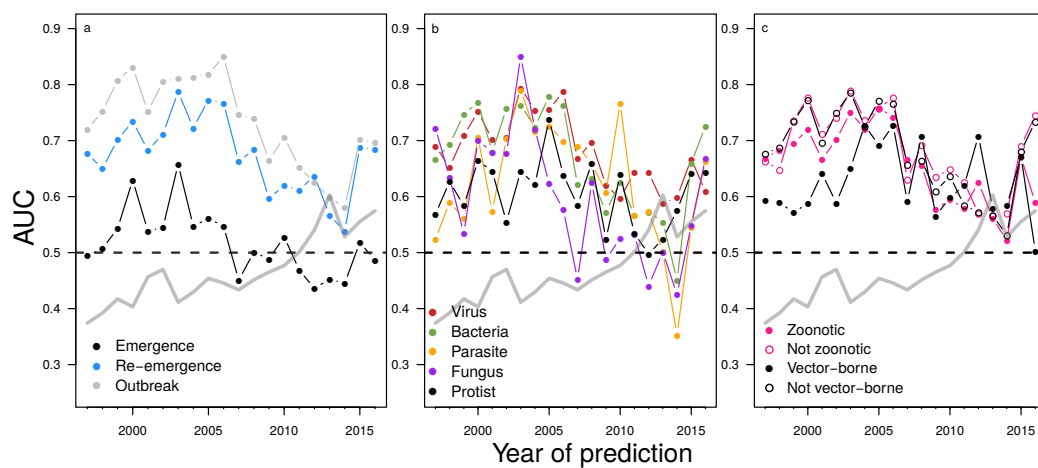


Figure S2: Rolling window size did not strongly influence model performance when considering next year prediction, as a window of 7 years produced qualitatively similar results to the window of 4 years we examine in the main text. The null expectation (grey line) performed consistently worse than our approach.

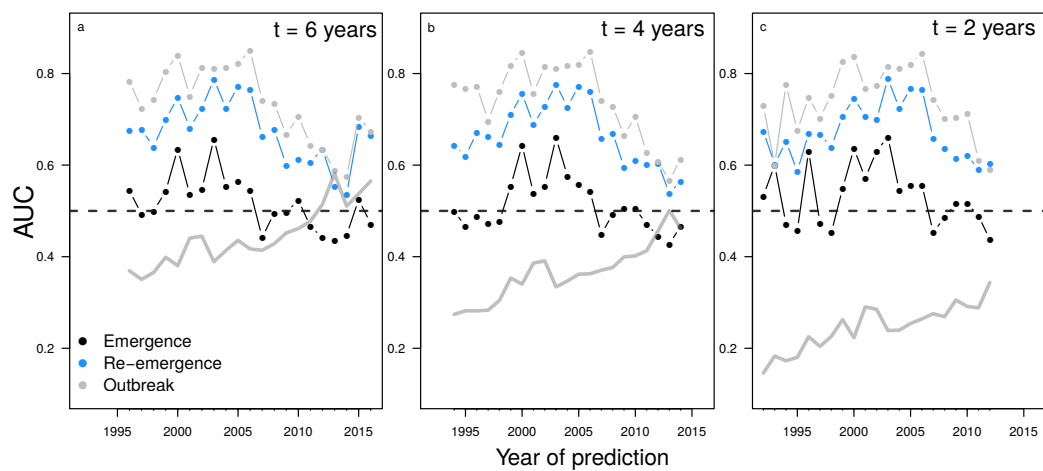


Figure S3: Examining rolling windows of 2, 4, and 6 years provides evidence that rolling window size did not strongly influence model performance. The null expectation (grey line) performed consistently worse than our approach.

587 **Effect of pathogen traits on model hindcasting/forecasting**
588 **ability**

589 Here, we further explore the effect of pathogen type on model performance when
590 hindcasting or forecasting pathogen outbreak or emergence event suitability. There
591 was no predictable variation in model performance as a function of pathogen type
592 (Figure S4) or whether the pathogen is classified as zoonotic or vector-borne (Fig-
593 ure S5).

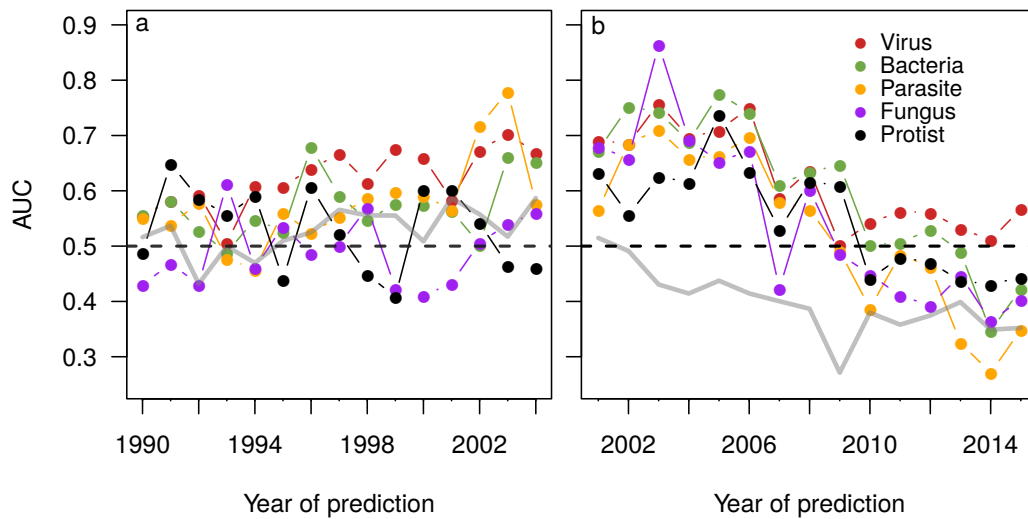


Figure S4: Predictive accuracy decreased when attempting to forecast far into the past or future, independent of pathogen type (grey line corresponds to null model performance). Models were trained on either the period between 2005-2015 (for prediction into the past) or 1990-2000 (for prediction into the future).

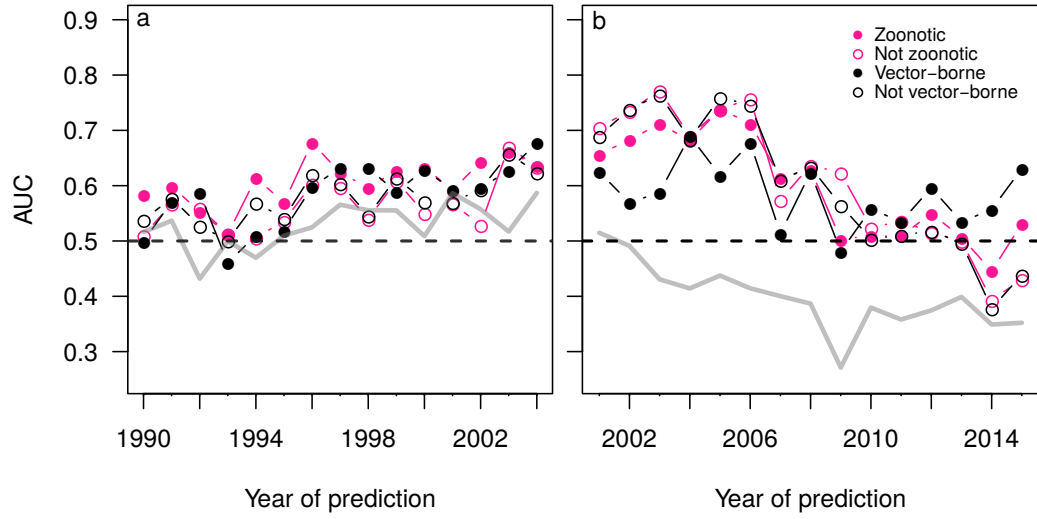


Figure S5: Predictive accuracy decreased when attempting to forecast far into the past or future. This was insensitive to whether the pathogen is considered zoonotic or vector-borne. Models were trained on either the period between 2005-2015 (for prediction into the past) or 1990-2000 (for prediction into the future). The null expectation (grey line) performed consistently worse than our approach.

594 **The effect of including influenza**

595 The 2009 influenza A pandemic fundamentally changed the network of countries
596 and pathogens through the addition of many links to one pathogen (Figure S6).
597 This may be an issue for approaches such as ours, which relies on extracting infor-
598 mation from the similarity between pathogens in their distributions among coun-
599 tries, and similarity between countries in their pathogen composition. When the
600 model wasn't expected to predict a pandemic event, the inclusion of influenza did
601 not substantially influence model predictions when trained on differing numbers
602 of years (Figure S7) or when forecasting or hindcasting to different time periods
603 (Figure S8). However, the effect of the 2009 influenza pandemic can be seen in the
604 substantial declines in model performance when attempting to forecast one year
605 ahead after training on a rolling window of 4 years (Figure S9). Interestingly, the
606 exclusion of influenza results in lower mean performance when the model doesn't
607 have data on many years, likely because influenza is widespread and can influence
608 the pathogen and country dissimilarity values used to train the model. However,
609 once sufficient data is provided, model performance with and without influenza is
610 nearly identical.

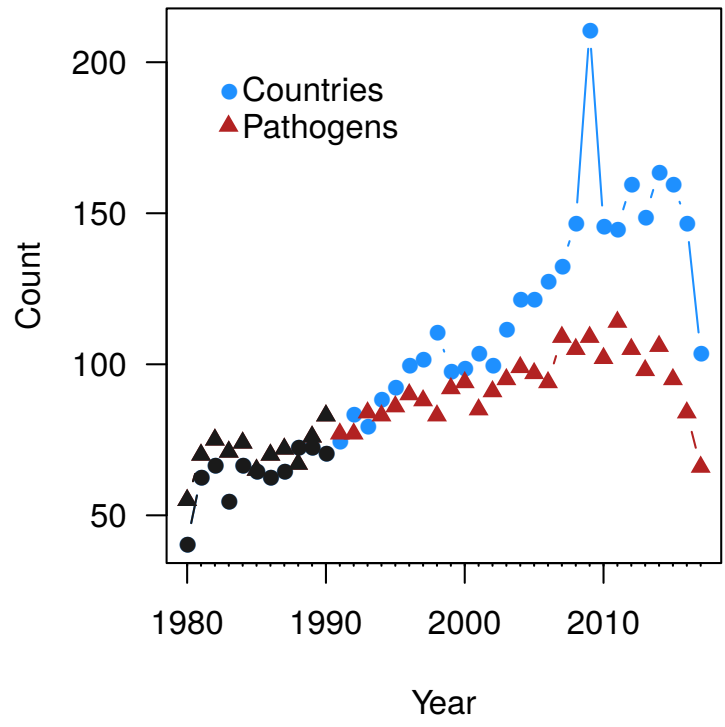


Figure S6: The number of countries with at least one outbreak event and the number of pathogens has increased over time, likely due to more vigilant sampling and description of emerging pathogens in a larger number of countries.

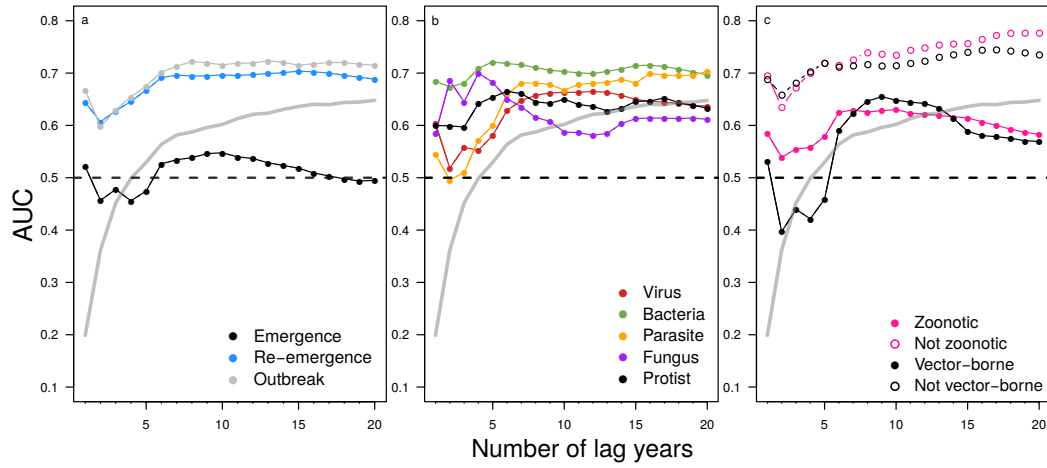


Figure S7: Pathogen events from previous years increased model predictive accuracy after an initial small decrease, suggesting that five years or more of data improves predictions, but accuracy could actually decrease in some data sparse situations where only two or three years of data were available. Performance of the null expectation (grey line) was less than our approach, except when the null was given more than 15 years of previous data.

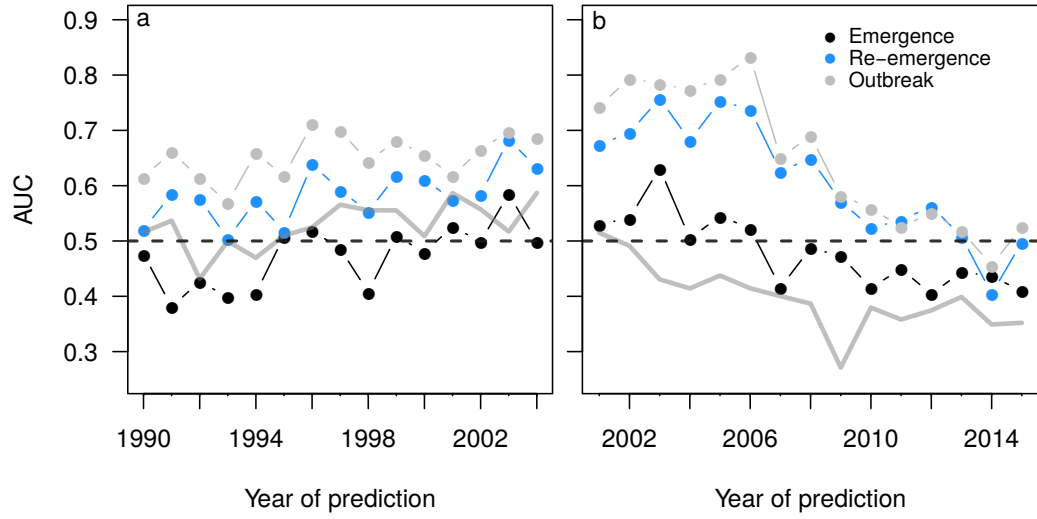


Figure S8: Predictive accuracy decreased when attempting to forecast far into the past or future. Models were trained on either the period between 2005-2015 (for prediction into the past) or 1990-2000 (for prediction into the future). The null expectation (grey line) performed consistently worse than our approach.

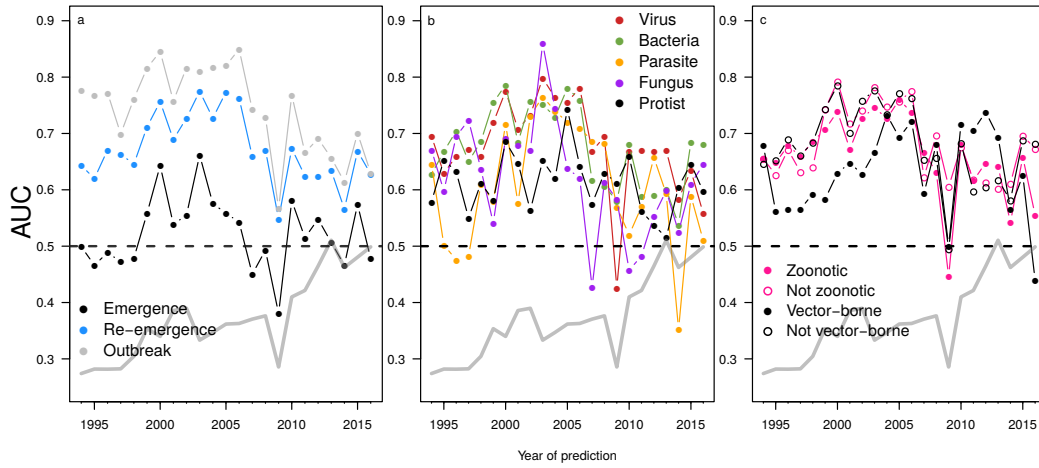


Figure S9: Using a rolling window ($t = 4$ years), we found that predictive accuracy did not increase as a result of enhanced surveillance and data collection of more recent years. The null expectation (grey line) performed consistently worse than our approach.