Predicting the future distribution of antibiotic resistance using time series forecasting and geospatial modelling

[version 1; peer review: 2 approved with reservations]

Benjamin Jeffrey¹, David M. Aanensen², Nicholas J. Croucher¹, Samir Bhatt¹

¹MRC Centre for Global Infectious Disease Analysis, Department of Infectious Disease Epidemiology, Imperial College London, London, UK
²Big Data Institute, University of Oxford, Oxford, UK

Abstract

Background: Increasing antibiotic resistance in a location may be mitigated by changes in treatment policy, or interventions to limit transmission of resistant bacteria. Therefore, accurate forecasting of the distribution of antibiotic resistance could be advantageous. Two previously published studies addressed this, but neither study compared alternative forecasting algorithms or considered spatial patterns of resistance spread.

Methods: We analysed data describing the annual prevalence of antibiotic resistance per country in Europe from 2012 – 2016, and the quarterly prevalence of antibiotic resistance per clinical commissioning group in England from 2015 – 2018. We combined these with data on rates of possible covariates of resistance. These data were used to compare the previously published forecasting models, with other commonly used forecasting models, including one geospatial model. Covariates were incorporated into the geospatial model to assess their relationship with antibiotic resistance.

Results: For the European data, which was recorded on a coarse spatiotemporal scale, a naïve forecasting model was consistently the most accurate of any of the forecasting models tested. The geospatial model did not improve on this accuracy. However, it did provide some evidence that antibiotic consumption can partially explain the distribution of resistance.

The English data were aggregated at a finer scale, and expected-trend-seasonal (ETS) forecasts were the most accurate. The geospatial model did not significantly improve upon the median accuracy of the ETS model, but it appeared to be less sensitive to noise in the data, and provided evidence that rates of antibiotic prescription and
bacteraemia are correlated with resistance.

**Conclusion:** Annual, national-level surveillance data appears to be insufficient for fitting accurate antibiotic resistance forecasting models, but there is evidence that data collected at a finer spatiotemporal scale could be used to improve forecast accuracy. Additionally, incorporating antibiotic prescription or consumption data into the model could improve the predictive accuracy.

**Keywords**
Antibiotic Resistance, Antibiotic Consumption, Forecasting, Time Series Modelling, Geospatial Modelling
Introduction

Globally, the prevalence of antibiotic resistance is growing at an alarming rate\(^1-3\). This is a huge threat to public health which is further compounded by the fact that the rate of discovery of new antibiotics has fallen to almost zero\(^4\). It is now understood that the continued use of antibiotics in medicine, agriculture, and aquaculture creates the selection pressure which has driven the increase in the prevalence of resistant bacteria\(^13\). However, rates of resistance are highly heterogeneous between locations, reflecting the impact of multiple covariates on the observed prevalence of antibiotic resistance\(^6\). These include: rates of antibiotic use\(^2\); seasonality\(^4\); prevalence of disease particularly within health-care settings which promotes increased antibiotic consumption; poor quality antibiotics\(^1\); and patient adherence to prescribed antibiotic regimens\(^1\). Socio-economic status has also been observed to be correlated with antibiotic resistance, although it is probable that this is confounded by other factors such as disease prevalence and inappropriate consumption of antibiotics\(^3,10\).

Recommendations for changes to clinical practice and public health policy have been designed to limit the future spread of antibiotic resistance\(^11-14\). Additionally, therapeutic solutions to treating multidrug resistant (MDR) infections are under development\(^15-19\). However, to maximise the efficacy of these clinical interventions, they should be implemented as early as possible when resistance to particular antibiotics is observed to be rising. For this to be achieved, active surveillance of antibiotic resistance is essential\(^20\), and the ability to accurately forecast the future spatial and temporal distribution of antibiotic resistance from surveillance data could be highly informative.

Most published mathematical models of antibiotic resistance epidemiology are mechanistic models of bacterial evolution\(^21\), or transmission of disease\(^22\), from which trends in the prevalence of antibiotic resistance are inferred\(^21-27\). These studies normally use small data sets from a single hospital or clinical trial, typically with the intention of assessing the efficacy of an intervention\(^26\) or improving understanding of the spread of antibiotic resistance and the factors which promote it\(^27\). Although these studies provide useful insight into the biology of resistance they are likely to be unsuitable for making future forecasts of the spatiotemporal distribution of antibiotic resistance\(^9\).

It has been demonstrated that model structure can have a highly significant influence on the predictions of epidemiological studies\(^29\). Therefore, initial testing can employ flexible forecasting algorithms, to identify methods that might be broadly generalisable to different species, antibiotics, and settings. Such models learn their parameters from the data, independent of assumptions regarding the population structure or evolutionary biology of the species in question. To our knowledge there have only been two previous attempts to forecast the future prevalence of antibiotic resistance from retrospective antimicrobial susceptibility testing (AST) data using a model which satisfies these conditions: Lopez-Lozano \textit{et al.}, 2000\(^30\) and Teodoro and Lovis, 2013\(^32\). Lopez-Lozano \textit{et al.} applied an autoregressive integrated moving average (ARIMA) model to a time series of AST data taken from a single hospital and combined it with pharmacy data as a proxy for antibiotic consumption to investigate the correlation between these two factors, whilst also producing future forecasts. A similar approach was taken by Gharbi \textit{et al.}, 2015\(^34\). However, they only included meropenem consumption as an explanatory variable, not retrospective AST data, so their model is only applicable if detailed retrospective data on antibiotic prescription or consumption are available. A more complex approach was taken by Teodoro and Lovis, 2013\(^32\), who applied empirical mode decomposition (EMD) followed by k-nearest neighbour (KNN) analysis to an AST time series from a hospital in Geneva to generate forecasts.

Whilst the authors of both these papers demonstrate their method produces reasonably accurate forecasts, there are important limitations to both studies. Firstly, they only include data from a single geographic location and therefore cannot make inference regarding the future spatial spread of antibiotic resistance.

Secondly, the output of the model devised by Teodoro and Lovis does not include any estimate of uncertainty. This is a limitation of many machine learning models, which do not implicitly provide any measurement of error, unlike ARIMA, which is constructed within a frequentist framework allowing standard error to be inferred from the data\(^3\).

Finally, neither study presents any comparison of their chosen method to other methods of time series analysis (other than a benchmark, random-walk model). Reviews in the econometrics literature have demonstrated that there are large differences in the accuracy of alternative forecasting models, and the most accurate method is often related to the type of data\(^13-19\). Given that neither Lopez-Lozano \textit{et al.}\(^30\) or Teodoro and Lovis\(^32\) present such a comparison, the optimal method of forecasting from AST data remains unclear.

The primary objective of this study was to build on the work of Tedoro and Lovis, 2013\(^32\) and Lopez-Lozano \textit{et al.}, 2000\(^30\) to develop models for accurately forecasting the dissemination of antibiotic resistance. Specifically, we aimed to:

1. Determine the optimal method for predicting future prevalence of antibiotic resistance within a single geographic location from retrospective surveillance data.
2. Extend the model to produce accurate forecasts of the spatiotemporal distribution of antibiotic resistance.
3. Incorporate covariates into the model to investigate their relationship with the outcome and observe how they impact on forecast accuracy.

Methods

Data

Two open access data sets describing the prevalence of antibiotic resistance over time across a given geographic region were used for this study.
The first dataset was the European Antimicrobial Resistance Surveillance Network (EARS-Net) data\(^9\), describing the percentage of recorded isolates that were non-susceptible to carbapenems and non-susceptible to fluoroquinolones. Organisms included in the data set were \textit{Klebsiella pneumoniae}, \textit{Escherichia coli}, \textit{Pseudomonas aeruginosa}, and \textit{Acinetobacter} spp. Data for each species and antibiotic combination was recorded per country per year between the years 2012 and 2016 inclusive. Data representing 21 countries (Belgium, Bulgaria, Croatia, Denmark, Estonia, Finland, France, Greece, Hungary, Ireland, Italy, Latvia, Luxembourg, Malta, The Netherlands, Norway, Portugal, Romania, Slovenia, Sweden, and The United Kingdom) were used for the analyses of resistance to carbapenems; and data representing 29 countries (Austria, Belgium, Bulgaria, Croatia, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, The Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, and The United Kingdom) were used for the analyses of fluoroquinolone resistance. These countries and antibiotics were chosen because data describing the estimated annual consumption of these antibiotics in these countries was available from the European Surveillance of Antimicrobial Consumption Network (ESAC-Net)\(^9\).

Additional covariates included in the analysis of the European data were annual gross domestic product (GDP) per capita, estimated by the International Monetary Fund (IMF)\(^9\), and annual rate of influenza-like-illness (ILI) per 100,000 people, estimated by the World Health Organisation (WHO)\(^9\).

The second dataset was the Fingertips database, including antibiotic resistance surveillance data from England\(^8\). This described the quarterly percentage of \textit{E. coli} isolates in each of 195 clinical commissioning groups (CCGs) that were non-susceptible to treatment with either ciprofloxacin, gentamicin, piperacillin/tazobactam, or 3rd generation cephalosporins (3GC), between the final quarter of 2015 and the final quarter of 2018. We also included the rate of \textit{E. coli} bacteraemia per CCG per quarter from this dataset in our analysis\(^8\). These data were merged with information on the rate of prescribing of each of the above antibiotics per CCG from the Open Prescribing database\(^41\) measured in the number of items prescribed per month, which was divided by the population in each CCG, estimated by the Office of National Statistics\(^42\), and multiplied by 1000 to give the number of items prescribed per 1000 people. We then aggregated the prescription data to quarters to allow comparison with the antibiotic resistance data.

Extended data, Figure 1 and Figure 2 shows the location of each country represented in the European data and the CCG boundaries in England respectively\(^43\).

**Single time series forecasting models**

Four alternative forecasting models were tested on the data from each location individually. These were autoregressive integrated moving average (ARIMA), which is the model used by Lopez-Lozano \textit{et al.}\(^44\); expected-trend-seasonal (ETS) forecasting\(^45\); neural network (NN) autoregression forecasting\(^46\), and the model developed by Teodoro and Lovis\(^38\) (from hereon referred to as Teodoro and Lovis forecasting) which is summarised in the following paragraphs. ARIMA, ETS, and NN autoregression are standard methods of time series prediction, described in 32. Finally, a naïve forecasting model (every point in the future is equal to the last observed value in a time series) was used as a benchmark comparison of forecast accuracy\(^39\).

Teodoro and Lovis forecasting begins by applying empirical mode decomposition (EMD) to the time series. This decomposes the data into a set of intrinsic mode functions (IMFs), defined as a function satisfying two properties\(^44\):

1. The number of extrema must be equal to, or one more or less than, than the number of zero crossings.
2. At any given point in the time series the mean value of the envelopes defined by the local maxima and minima equals zero.

EMD is conducted iteratively by first identifying the local maxima and minima of a time series and connecting them with a cubic spline to define the upper envelope \(e_{\text{max}}(t)\) and the lower envelope \(e_{\text{min}}(t)\). At each point in time the series the mean of the upper and lower envelope is calculated according to Equation 1:

\[
m(t) = \frac{e_{\text{max}}(t) + e_{\text{min}}(t)}{2}
\]

(Equation 1)

The oscillatory component is then calculated as the difference between the data and the local mean at each time point. If this component does not satisfy the IMF conditions, then the steps described in the previous paragraph are repeated using this component as the time series until the resulting component is an IMF. At this stage, the IMF is subtracted from the time series leaving the data minus the highest frequency oscillatory component. Each of the previous steps are then repeated taking the remaining data as the time series. This loop continues until a monotonic function from which no more IMFs can be extracted is left.

A time series decomposed with EMD can be expressed as:

\[
y = \sum_{j=1}^{n} c_j + r
\]

(Equation 2)

where \(c\) is an IMF, \(n\) equals the total number of IMFs, and \(r\) is the residual. The IMFs and the residual are collectively referred to as components of the original time series.

Teodoro and Lovis then remove the highest frequency components as these are assumed to be associated with random noise. They hypothesise that the remaining components are associated with seasonalities of decreasing frequency and finally the overall temporal trend in the rate of antibiotic resistance.

Having selected the relevant, low frequency components of the time series, each component is then embedded into delay vectors. This converts the component into a series of vectors.
each representing one point in the component and containing the previous \( m \) data points. The value of \( m \) is selected using a method derived by Chakrabarti and Faloutsos, 2002\(^{40} \) who use fractal dimensions to specify the optimal value of \( m \). They increment \( m \) between 1 and 10 and calculate the fractal dimension of the resulting matrix using a box-counting algorithm\(^ {41} \). After some value of \( m \) the fractal dimension of the matrix plateaus and this is chosen as the optimal value of \( m \) for the delay embedding vectors.

Finally, k-nearest neighbour (kNN) analysis is applied to the components in the \( m \)-dimensional embedding space. Cross validation is used to select the optimal value of \( k \) which gives the number of points closest in time to the output \((y_{\text{out}})\) to be used to generate a forecast. The mean of these \( k \) variables gives the output of the kNN algorithm. This is done for each low frequency component of the time series and the one-step ahead forecast is calculated as the sum of each of these values. Multi-step ahead forecasts are calculated by iteratively concatenating each predicted value with the time series and repeating the process.

**Spatiotemporal modelling**

Classically, geostatistical data has been modelled as a continuous Gaussian field\(^ {42} \). This is an \( n \)-dimensional process from which any subset of values can be described by a multivariate Gaussian distribution, defined by its mean and a covariance function. Having fitted a model of this kind to spatially indexed data it is possible to use Markov chain Monte Carlo (MCMC) methods to sample from a posterior distribution defined by the Gaussian field and the data. However, this is a computationally expensive method which scales very poorly (\( O(n^3) \)) due to the high cost of performing algebraic operations with large, dense covariance matrices\(^ {43} \). As an alternative to this, we employed an approximate Bayesian computation (ABC) method that uses stochastic partial differential equations (SPDE) to derive an approximate probability distribution from which inference can be made using integrated nested Laplace approximations (INLA)\(^ {44,45} \).

The INLA approach first requires a Gaussian field to be approximated as a discretely indexed Gaussian Markov random field (GMRF)\(^ {46} \). This is defined on a regular lattice with a 2-dimensional mesh at time \( t \), with the same location \( x \) at times \( t + 1 \) and \( t - 1 \).

Given a spatiotemporal model formulated as a GMRF by utilising SPDE, the INLA algorithm can be applied to approximate the posterior distributions of parameters of interest using deterministic simulations\(^ {49} \). This is considerably more efficient than stochastically sampling from the posterior distribution of a continuously indexed field using an MCMC algorithm. For full details see Rue et al.\(^ {46} \), Bakka et al.\(^ {50} \) and Lindgren et al.\(^ {47} \).

**Modelling covariates**

As described above, we also gathered data on covariates that were likely to be associated with the prevalence of antibiotic resistance. An advantage of all the algorithms described above is that they can easily be extended to include covariates by including another set of inputs, each corresponding to the value of a covariate in a given time and location.

**Evaluating forecast accuracy**

The forecasting models described above were applied to the European and English data sets. To allow the accuracy of each model to be determined, for an \( h \)-step ahead forecast, the time series of length \( T \) was split to remove the last \( h \) observations (the testing data). The model was then fitted to the remaining \( T-h \) observations (the training data) so the difference between the forecasted data and the real data could be observed.

The single time series forecasting models were fitted to each location individually and the INLA model was fitted to the entire data set and forecasts produced for each location. Two measures of accuracy were calculated for each forecast to allow comparison of the different models. Firstly, the root mean squared error (RMSE) (Equation 7) was calculated to assess...
the magnitude of the forecast error. RMSE is a proper scoring rule as defined by 51, meaning the objective of the forecasting task will always be to minimise this value.

\[
RMSE = \sqrt{\frac{1}{h} \sum_{t=1}^{h} (e_t)^2} \quad \text{(Equation 7)}
\]

In Equation 3 \( h \) is the forecasting horizon and \( e_t \) is the error associated with each forecasted point in the time series.

Secondly, the percentage of real data points that lie within the 95% prediction interval (PI) of the forecast was calculated to determine how accurately each algorithm assessed its own uncertainty.

All of the above analysis was performed in the R statistical programming language version 3.6.1\(^\text{15}\). The INLA model was implemented within the R-INLA package version 18.07.12\(^\text{52}\), and the single time series forecasting models within the forecast package version 1.26\(^\text{93}\).

**Results**

The following sections describe the results of our analyses of the European and English antibiotic resistance surveillance data sets. For each data set we first provide a comparison of the results of the temporal forecasting algorithms applied to data from individual locations; we then consider the impact of including relevant covariates in the forecasting model; and finally we present the results of the spatiotemporal forecasts produced by the INLA model. In each case the forecasting algorithms were applied to the training data, which were the first \( T-h \) time points describing the proportion of isolates that were resistant in a given location (where \( T \) is the total length of the time series, and \( h \) is the forecasting horizon). Forecast accuracy is evaluated by comparison to the final \( h \) data points in a time series (the testing data).

**European data: single location forecasting**

We began our analysis of these data by splitting the data by country and applying single time series forecasting models with a forecasting horizon of two years to each country’s data. For these data we compared three forecasting models: ARIMA, Expected-Trend-Seasonal (ETS), and Neural Network (NN) autoregression. The Teodoro and Lovis model was not applied to these data because there were only six data points meaning the Intrinsic Mode Function (IMF) algorithm nearly always returned the original time series as the residual, and no additional components. Therefore, the model was reduced to just the k-nearest neighbours (KNN) algorithm. Finally, we also included a naïve forecasting model as a benchmark comparison.

Box and whisker plots in Figure 1 show the root mean squared errors (RMSE) of each forecasting model applied to each location in the data set. The eight individual plots represent each species-antibiotic combination for *Acinetobacter* spp., *E. coli*, *K. pneumoniae* or *P. aeruginosa*, and carbapenems or fluoroquinolones.

The accuracy of the forecasts produced by the alternative models is highly variable between data sets. The relative accuracies of the four forecasting models were comparable between species and phenotypes. The neural network autoregression model consistently produced the least accurate forecasts (indicated by the greatest median RMSE) which is probably a result of the neural network overfitting to a small data set. The worst performing model, based on the highest observed RMSE (53%), was the NN model for the *P. aeruginosa* - fluoroquinolones data. The forecast RMSE produced by the naïve model was the lowest observed RMSE in almost every instance, followed by the ARIMA forecasts. Finally, on average the ETS forecasts had slightly greater RMSE values than the equivalent ARIMA or naïve forecast. Hence only the ARIMA method showed the potential to outperform a naïve model.

The median RMSE for each forecasting model was relatively consistent across data sets: between 0% and 7% for every species - antibiotic combination with every forecasting model. It should be noted however, that this is probably skewed by the fact that the mean value of many of the individual time series was very close to zero. As there are limits applied to the outputs of the forecasting models to prevent them from predicting values greater than 100% and less than 0%, if each observation in a time series is very close to either limit (and there is minimal fluctuation in observed values) then the forecasts will be maintained artificially close to the testing data as any trend is curtailed by the upper or lower forecast limit.

An alternative measure of model accuracy to RMSE is the percentage of observations in the training data that lie within the 95% prediction interval (PI) of the forecasting model (Figure 2). This also demonstrates that the naïve forecasting model produces the most accurate projections with these data for the vast majority of time series, the only exception being the *Acinetobacter* spp.- carbapenems data, for which the mean prediction interval accuracy was slightly higher for the ARIMA forecasts. In every case the mean PI accuracy of the ETS forecasts was lower than the ARIMA and naïve forecasts, but always within one standard error of the ARIMA forecast. Finally, the PI accuracy of the NN forecasting model was as low as a 0% for many of the individual time series, and the greatest mean PI accuracy with this model was 25% for the *E. coli* - carbapenems data. This reflects both the poor accuracy of the point estimates of the NN model (highlighted by Figure 1), and the difficulties of estimating error outside of a defined frequentist or Bayesian framework. Prediction intervals for the NN model are obtained by performing multiple simulations randomly drawing error terms from the training data, and therefore it is a natural tendency of NN models to overestimate their own precision when making predictions outside of the range of the training data. This confirmed that the ARIMA and naïve models best predicted these data.

It is possible that when the coverage of the 95% PI of a forecasting model is high, this is a result of the model predictions having very high variance, meaning the PI often spanned a very wide range of possible values. To test this possibility, we also
calculated and plotted the 75% PI for each forecasting model (Extended data, Figure 3). A perfect forecasting model would have a 95% PI that included 95% of all future data points and a 75% PI that included 75% of all future data points. The reduction in percentage accuracy of the 75% PI compared to the 95% PI was greater than 20 percentage points for every ARIMA and ETS model. However, the difference in coverage of the 95% and 75% PI for the naïve and NN model was not always this great. (For the NN model this is because the 95% PI accuracy was already approaching zero with many of the data sets.) Figure 2 and Extended data, Figure 3 indicate that all the models have a tendency to overestimate their precision when fitted to these data. However, the magnitude of this overestimation was least with the naïve model, followed by ARIMA, ETS, and then the NN autoregression model.

Having tested each of the forecasting models on individual time series we then considered incorporating relevant covariates for which we had data into the forecasting models. These were: rate of consumption of the relevant antibiotic, expressed in defined daily doses (DDD) per 1000 people per day; annual cases of influenza-like-illness (ILI) per 100,000 people; and annual gross domestic product (GDP). Extended data, Figure 4 shows the relationship between each of the covariates and the percentage of *Acinetobacter* spp. isolates that were resistant to carbapenems. It is clear there is no...
correlation between any of the covariates and the outcome. This relationship was observed for every species and resistance phenotype. Therefore, we did not extend any of the single time series forecasting models to include covariates, as there was insufficient evidence that any of the covariates were directly associated with the outcome.

European data: spatiotemporal forecasting

To further investigate the relationship between antibiotic consumption, GDP, and ILI, and the rate of antibiotic resistance, we hypothesised that there could be a more complex association driven by the spatial distribution of the covariates and potentially a lag in the emergence of resistance following a change in the level of a given covariate. An example of this phenomenon is described by Volz and Didelot\(^4\), who observed a lagged spike in the effective population size of methicillin-resistant \textit{Staphylococcus aureus} (MRSA) in the USA following a sharp increase in the rate of consumption of beta-lactam antibiotics.

For this we employed a spatiotemporal model fitted with the integrated nested Laplace approximation (INLA) algorithm. This model also captures the spatial correlation in antibiotic resistance as well as the temporal trend. We first fitted the INLA model to the resistance data without any covariates and a range of values of the autoregressive parameter. Increasing the autoregressive parameter beyond 2 years did not improve the model fit any further and we therefore used this as the optimal value for all analyses of these data with the INLA model. We then fitted the INLA model to the data for each species and phenotype with each covariate individually. Table 1 shows the point estimate and 95\% credible interval of the coefficient for each covariate with each species and resistance phenotype.

The data in Table 1 provides some evidence that the rate of resistance is correlated with consumption when spatial and temporal trends are accounted for by the INLA model. This association was only significant (\(p = 0.05\)) for \textit{Acinetobacter} spp., \textit{K. pneumoniae}, and \textit{P. aeruginosa} resistance to
Table 1. INLA coefficients describing the relationship between each covariate and prevalence of antibiotic resistance in the European data. Coefficients of INLA models fitted to each species and resistance phenotype with each of the three covariates. Asterisk indicates coefficients the 95% credible interval of which did not include zero.

| Organism          | Antibiotic | Covariate | Mean   | 95% CI         |
|-------------------|------------|-----------|--------|----------------|
| Acinetobacter spp | Carbapenems| Ab Consumption | 0.505  | -55.518-56.706 |
| Acinetobacter spp | Fluoroquinolones | Ab Consumption | 13.883* | 5.561-21.797   |
| E. coli           | Carbapenems| Ab Consumption | 0.536  | -4.330-4.815   |
| E. coli           | Fluoroquinolones | Ab Consumption | 0.103  | -1.659-1.937   |
| K. pneumoniae     | Carbapenems| Ab Consumption | 36.630 | -4.238-76.828  |
| K. pneumoniae     | Fluoroquinolones | Ab Consumption | 6.131* | 0.429-12.202   |
| P. aeruginosa     | Carbapenems| Ab Consumption | 26.282 | -23.973-76.053 |
| P. aeruginosa     | Fluoroquinolones | Ab Consumption | 11.653* | 9.359-13.840   |
| Acinetobacter spp | Carbapenems| ILI        | -0.009 | -0.092-0.073   |
| Acinetobacter spp | Fluoroquinolones | ILI        | -0.044 | -0.143-0.053   |
| E. coli           | Carbapenems| ILI        | 0.000  | -0.003-0.004   |
| E. coli           | Fluoroquinolones | ILI        | 0.012  | -0.006-0.030   |
| K. pneumoniae     | Carbapenems| ILI        | -0.718 | -0.818-0.617   |
| K. pneumoniae     | Fluoroquinolones | ILI        | 0.020  | -0.030-0.071   |
| P. aeruginosa     | Carbapenems| ILI        | 0.053  | -0.003-0.108   |
| P. aeruginosa     | Fluoroquinolones | ILI        | -0.005 | -0.058-0.049   |
| Acinetobacter spp | Carbapenems| GDP       | 0.005  | -0.014-0.023   |
| Acinetobacter spp | Fluoroquinolones | GDP       | 0.003  | -0.008-0.014   |
| E. coli           | Carbapenems| GDP       | 0.000  | 0.00-0.00      |
| E. coli           | Fluoroquinolones | GDP       | 0.000  | -0.004-0.004   |
| K. pneumoniae     | Carbapenems| GDP       | 0.000  | 0.00-0.00      |
| K. pneumoniae     | Fluoroquinolones | GDP       | 0.001  | -0.010-0.012   |
| P. aeruginosa     | Carbapenems| GDP       | 0.001  | -0.004-0.006   |
| P. aeruginosa     | Fluoroquinolones | GDP       | 0.002  | -0.007-0.010   |

Fluoroquinolones, and consumption of fluoroquinolones. However, there is also a large coefficient describing the relationship between *K. pneumoniae* resistance to carbapenems and carbapenem consumption (36.63), and *P. aeruginosa* resistance to carbapenems and carbapenem consumption (26.28), but neither of these are statistically significant. Finally, the coefficients of the fitted INLA model provide no evidence that either incidence of ILI or national GDP are significantly associated with antibiotic resistance. Hence, for the purpose of making future forecasts of the prevalence of antibiotic resistance with the INLA model, we included rate of antibiotic consumption as a covariate but excluded incidence of ILI and national GDP. The RMSE of the two year forecasts of antibiotic resistance prevalence with each species are displayed in Figure 3 (the RMSE of the naïve and ARIMA forecasts are also provided for comparison).

The accuracy of the INLA forecasts was broadly comparable to the ARIMA forecasts. However, the median RMSE of the naïve forecasts is marginally lower for every species and resistance phenotype. This indicates that, although there was an observable relationship between antibiotic resistance and consumption when spatial and temporal correlations were accounted for, there was not enough signal in these data to improve upon the accuracy of a naïve forecasting model.
English data: single location forecasting

To test whether finer temporal and spatial resolution would improve the performance of these predictive algorithms, we applied them to the Fingertips database of disease in England. Firstly, we split the data by CCG and applied five forecasting algorithms to the data from each location individually. These were ARIMA, ETS, NN autoregression, Teodoro and Lovis forecasting, and naïve forecasting. For these data we used a range of forecasting horizons between 1 and 4 quarters to observe the difference in accuracy when forecasts are made further into the future. The RMSE of the forecasts produced by each model for each resistance phenotype with forecasting horizons ranging between 1 and 4 quarters are displayed in the box and whisker plots in Figure 4. Plots A-D represent forecasts of *E. coli* resistance to 3rd generation cephalosporins (3gC), ciprofloxacin, gentamicin, and piperacillin/tazobactam respectively, over forecasting horizons of 1–4 quarters.

Box and whisker plots in Figure 4 firstly demonstrate that the median RMSE of each forecasting algorithm increases as forecasts are calculated for a more distant forecasting horizon, increasing by approximately 1.5% between quarters 1 and 4 for each algorithm. This observation is consistent across phenotypes. The median accuracy of the forecasts produced by each algorithm are also broadly comparable across phenotypes. For example, the median RMSE of the ETS forecasts performed with a forecasting horizon of 4 quarters ranged between 2.98% for *E. coli* resistance to gentamicin and 3.84% for resistance to ciprofloxacin. With regards to the relative accuracy of each forecasting model, NN autoregression is consistently the least
accurate algorithm for all phenotypes and forecasting models, followed by ARIMA. The median and maximum RMSE of the forecasts produced by the naïve model are very similar to those produced by the ETS model: for *E. coli* resistance to 3gC, ciprofloxacin, and piperacillin/tazobactam the median RMSE of the naïve forecasts was marginally lower than the median RMSE of the ETS forecasts, whereas for *E. coli* resistance to gentamicin the median RMSE of the ETS forecasts was lower (2.08% and 2.07%). Finally, the Teodoro and Lovis were always equivalent to a naïve forecast in this population (see Discussion).

Plots A–D in Figure 5 refer to *E. coli* resistance to 3gC, ciprofloxacin, gentamicin, and piperacillin/tazobactam respectively, and show the percentage of future observations in the time series that are within the 95% PI of each forecasting algorithm. (Teodoro and Lovis forecasting is excluded because it does not produce any measure of uncertainty). As was observed with the European data, the percentage of future observations that are within the 95% PI of the NN forecasts is far lower than with any other forecasting model (less than 11% for every phenotype with every forecasting horizon). The 95% PI of the naïve forecasting algorithm also contains the most future data points of any of the forecasting models (greater than 90% for every forecasting horizon with every phenotype with the exception of cephalosporins, for which it was between 88% and 90%). This was also observed with the European data. However, unlike the forecasts of the European data, the 95% PI of the ETS model contains a greater percentage of future observations than that of the ARIMA forecasts in every case. The RMSE of the ETS forecasts is also consistently lower than the RMSE of the ARIMA forecasts (Figure 4).

We also plotted the percentage accuracy of the 75% PI for each forecasting model applied to the English data (Extended data, Figure 5). The difference in coverage of the 95% and 75% PI
Figure 5. Percent accuracy of the alternative forecasting algorithms applied to the English data. Plots A–D show the percentage of observations in the training data that were within the 95% prediction interval (PI) of each forecasting algorithm with each forecasting horizon (1–4 quarters) and represent E. coli resistance to 3gC, ciprofloxacin, gentamicin, and piperacillin/tazobactam respectively. The ARIMA forecasts are coloured red, the ETS forecasts are coloured green, the naive forecasts are coloured blue, and the NN autoregression forecasts are coloured purple. Error bars show the standard error of the mean PI accuracy.

was far less with the English data than with European data: the mean percentage accuracy of the 75% PI was between 60% and 75% with both the ETS and naïve model applied to every data set.

Having observed the relative accuracy of each forecasting model applied to the time series of E. coli antibiotic resistance in each CCG, we then considered incorporating covariates into the forecasting models for these data. These were the rate of E. coli bacteraemia and the amount of each antibiotic that was prescribed per CCG. Extended data, Figure 6 shows the relationship between these covariates and resistance of E. coli to 3gC considering each point in space and time individually43. As was observed with the European data, there is no correlation between either covariate and the prevalence of antibiotic resistance (this relationship was consistent for every E. coli phenotype). Therefore, we did not include either covariate in the single time series forecasting models. However, it is important to note that this analysis did not capture the spatial correlation or the temporal correlation of resistance with either covariate.

English data: spatiotemporal forecasting

To further investigate the spatiotemporal distribution of E. coli antibiotic resistance in England and the covariates described above, we fitted the INLA model to the data describing each resistance phenotype and each of the two covariates. The autoregressive parameter was set to 4 so the model could capture any seasonality in the data. Table 2 shows the coefficients of the INLA model fitted to the data for each E. coli resistance phenotype with each covariate. Although the effect
Table 2. INLA coefficients describing the relationship between each covariate and prevalence of antibiotic resistance in the English data. Coefficients of INLA models fitted to each resistance phenotype with each covariate. The asterisk indicates coefficients the 95% credible interval of which did not include zero.

| Organism | Antibiotic       | Covariate     | Mean    | 95% CI     |
|----------|------------------|---------------|---------|------------|
| E. coli  | 3gC              | Bacteraemia   | 0.036*  | 0.030-0.040|
| E. coli  | Ciprofloxacin    | Bacteraemia   | 0.025*  | 0.012-0.037|
| E. coli  | Gentamicin       | Bacteraemia   | 0.035*  | 0.031-0.038|
| E. coli  | Piperacillin/Tazobactam | Bacteraemia   | 0.022*  | 0.015-0.028|
| E. coli  | 3gC              | Ab Prescription| -0.033 | -0.078-0.012|
| E. coli  | Ciprofloxacin    | Ab Prescription| -0.004 | -0.052-0.043|
| E. coli  | Gentamicin       | Ab Prescription| 0.015   | -0.027-0.056|
| E. coli  | Piperacillin/Tazobactam | Ab Prescription| 0.056*  | 0.016-0.095|

The box and whisker plots in Figure 6 demonstrate that the median RMSE of the INLA forecasts is greater than the median RMSE of both the ETS and naïve forecasts. The median RMSE of the INLA forecasts ranged between 3.04% and 3.95%, whilst the median RMSE of the naïve and ETS forecasts ranged between 2.82% and 3.60%, and 3.03% and 3.84% respectively. This implies that on average the INLA model produces less accurate forecasts than either the ETS or naïve forecasting model. However, the maximum forecast RMSE for each resistance phenotype, with the exception of ciprofloxacin, was greater with the naïve (26.51%, 17.82%, 67.74%, 25.00%) or ETS (26.51%, 21.30%, 67.74%, 25.75%) forecasts than the maximum RMSE of the INLA forecasts (18.52%, 20.24%, 19.08%, 18.53%). This indicates that, although the INLA model may not be as accurate on average as the most appropriate single location forecasting model, it is more robust to noise and outliers in the data.

It is important to note that all the forecasting models, including the naïve model, produced more accurate results on average with the English data than with the European data. Therefore, it should not be assumed that the ETS and INLA forecasts only appeared to do better with these data because they are smoothing over noise in the data which is being captured by the naïve model. It is possible that in some cases the ETS model is able to more accurately forecast the English data because it captures seasonality in the data. However, this is unlikely to be the case for the majority of time series because *E. coli* resistance to the drugs of interest here is not highly seasonal, when averaged across English healthcare facilities. These data are shown in Extended data, Figure 7. We also note that the rate of prescription of all four antibiotics in England was not found to be seasonal in these data: Extended data, Figure 8.

Discussion

The purpose of the analysis presented here was to investigate the plausibility of performing accurate and precise data-driven forecasting of the future prevalence of antibiotic resistance. To do this we identified two open access data sets describing the past prevalence of antibiotic resistance at differing spatial and temporal scales (see Methods for a description of the two data sets). This allowed us to question how differences in the data a forecasting model is fitted to affect the accuracy of the model predictions. Significant differences in the accuracy of each forecasting model, including the benchmark naïve forecasting model, were observed between species. These differences are immediately apparent when we consider Figure 1 and Figure 3, showing the RMSE of each forecasting model applied to Europe-wide data describing the prevalence of resistance to...
The RMSE of the forecasts produced by the ETS model, the INLA model, and the naive forecasting model with each *E. coli* resistance phenotype. Plots A–D represent *E. coli* resistance to 3gC, ciprofloxacin, gentamicin, and piperacillin/tazobactam respectively. Bacteraemia was included as a covariate in the INLA model. Boxes show the median, and first and third quartiles. The whiskers denote the value of the lowest and highest value which is no more than 1.5 times the interquartile range from the first or third quartile respectively.

Other factors which could potentially explain species differences in the accuracy of the forecasts produced by the models include total prevalence of the species in the population and the seasonality of disease. For example, the ETS method explicitly models seasonality, whilst other models such as ARIMA, NN autoregression, and INLA will implicitly capture seasonality if the degree of the autoregressive component of the model is at least equal to the period of seasonality in the data. Unfortunately, the European data was aggregated per year, so it was not possible to investigate how seasonality of disease impacts on forecast accuracy with these data. However, the
English data (which was recorded per quarter) did not have noticeable seasonal trends, in contrast with previous studies of antibiotic resistance which have concluded that resistance is highly seasonal.\(^{40,41}\)

Differences in population genomics could be a further contributory factor to species differences in the accuracy of antibiotic forecasting models. Compared to eukaryotes, bacterial species are notoriously difficult to define\(^{32,33}\), and studies have shown that ecological perturbations can result in dramatic changes to the genetic and phenotypic characteristics of bacterial populations\(^{42-45}\). Therefore, ecological differences, and environmental changes that the data we had access to here did not capture, are also likely to be important in driving the dynamics that determine the rate of antibiotic resistance.

There are also some differences in the accuracy of the forecasts produced regarding resistance to different antibiotics. For example, the forecasts of fluoroquinolone resistance were consistently more accurate than those of carbapenem resistance in Europe. However, it is unclear whether these differences reflect biological variation in the epidemiology of resistance to alternative antibiotics, or simply stochastic variation in the amount of noise in alternative datasets. It is reasonable to expect that there may be some natural variation in the accuracy of forecasts produced corresponding to resistance to different antibiotics, related to the way resistance is transmitted within bacterial populations and the alternative ways in which antibiotics are used in healthcare. Secondly, the differences in forecast accuracy between antibiotic resistance phenotypes could be correlated with the alternative ways in which antibiotics are used, such as the mode of delivery or the patients they are offered to (inpatients versus outpatients). One objective of future work in this field should be investigating these species and resistance phenotype differences in forecast accuracy further.

Comparing Figure 5 and Figure 6 (the RMSE and mean PI accuracy of the forecasts with the English data set, which was recorded per quarter per CCG) with Figure 2 and Figure 3 (the corresponding Figures for the European dataset, recorded per year per country), demonstrates that increased spatial and temporal resolution of the data is associated with improved forecast accuracy. An in-depth comparison of the two sets of results could be misleading, as the data describe different species and resistance phenotypes. However, the results do indicate that forecast accuracy is improved with data which is aggregated at smaller units in space and time.

Considering the plots describing the RMSE and mean PI accuracy of the alternative models for both the European and English datasets demonstrates that there are significant differences in the accuracy of forecasts produced with each model. The Teodoro and Lovis\(^{38}\) model is one of only two previously published forecasting models for antibiotic resistance at the time of writing. As indicated previously there are significant flaws with this model: firstly, using the kNN algorithm to produce forecasts from an alternative representation of the data is more susceptible to error caused by noise in the data than fitting coefficients to different components of the time series. Secondly, for any forecasting task it is always important to have an awareness of the uncertainty in the model outputs, but unfortunately this model does not produce any measure of uncertainty. Thirdly, there is a problem applying the method that Teodoro and Lovis adapted from Chakrabarti and Faloutsos.\(^{45}\) When selecting the optimal lag for creating the embedding space in which kNN will be performed, Chakrabarti and Faloutsos state that the maximum lag which should be considered is the point at which increasing the lag any further does not significantly increase the fractal dimensionality of the data any further. This is defined as an increase in the fractal dimensionality of the data of less than \(\max(0.3,0.1 A_m)\) where \(A_m\) is the moving average of the fractal dimension.\(^{45}\) However, with every time series we applied this algorithm to, the increase in the fractal dimension after increasing the lag from 1 to 2 was always less than 0.3, therefore the kNN step of the algorithm was always applied to vectors of length one, containing only the last observation in the time series. Hence, the output was always equivalent to a naïve forecast for the published surveillance data of the type we used here, without changing these parameters.

The NN model performed significantly worse than the simpler statistical forecasting models. This result is consistent with the results of previous comparisons of forecasting models\(^{32-35,66}\) which show that applying complex machine learning models to time series forecasting tasks is rarely justified, and they can often be reduced to simpler models. This second observation was demonstrated by Hyndman and Billah\(^{46}\) who showed that the ‘theta model’\(^{77}\), which won the M3 forecasting competition\(^{31}\), is equivalent to a simple exponential smoothing model with drift\(^{52}\).

It appears that the ARIMA model was more accurate than ETS for forecasting the European data. However, for the finer resolution English data set the ETS model was considerably more accurate. This was the only case in which an alternative forecasting model performed better than the benchmark naïve forecasting model. This indicates that with good quality surveillance data ETS forecasting would be the best method, of those considered in this study, for producing forecasts from a single location.

Finally, we also applied the INLA algorithm\(^{69}\) to extend the forecasts to include spatial correlation, and covariates where justified. Interestingly, this did not improve upon the forecast accuracy achieved with an ETS model. This implies that there may not be sufficient spatial signal in the determinants of antibiotic resistance for any more predictive accuracy to be gained from modelling the spread of resistance as a spatial process. Another objective of future work in this field should be to test this assertion with finer resolution data, which was purposefully collected to minimise sampling bias. It will be useful to extend the work we have begun here by using data aggregated at a range of spatiotemporal scales, so the optimal scale for forecasting can be determined.

Whilst fitting the INLA model to the data did not improve upon the forecasting accuracy of the single time series forecasting models, it did provide some useful insight into the relationship between the covariates, and the outcome. We did not find

...
any evidence of a relationship between GDP and antibiotic resistance, despite the fact that the prevalence of antibiotic resistance is known to be inversely correlated with wealth\textsuperscript{11}. This is probably explained by the very high spatial aggregation of the European data, and the fact that all the countries represented in the data were high income countries\textsuperscript{49} so the gradient of GDP was not great enough for this relationship to be detected by the model. The rate of prescription or consumption of antibiotics was also not found to be significantly correlated with the prevalence of resistance in many cases, which is also an unexpected result. For the European data this is possibly explained by the signal being obscured by the high spatial and temporal aggregation of the data. However, the English data, recorded in the Open Prescribing database\textsuperscript{56}, reports the number of items of antibiotics that were prescribed. Consequently, each prescription was treated as equal, regardless of dosage, length of course of antibiotic, or mode of delivery. This likely introduces some noise into the data. Additionally, the Open Prescribing database describes the amount of antibiotic prescribed in primary care, and therefore does not include information on prescribing in hospitals, which is likely to be highly relevant to HAIs. Finally, fitting the INLA model provided some evidence of an association between the rate of bacteraemia and resistance. However, it is probable that this relationship is due to resistant infections being more likely to result in cases of bacteraemia as the infection is harder to treat, rather than bacteraemia being a determinant of resistance.

Having tested a range of forecasting models, including two previously published antibiotic resistance forecasting models\textsuperscript{26,30}, we conclude that ETS forecasting\textsuperscript{22} is likely to be the most appropriate model for producing forecasts of antibiotic resistance within a given location. Extending the forecasting model to incorporate spatial information using INLA did not improve the average predictive accuracy of the ETS forecasts. However, there were significant caveats with the data that potentially reduced the accuracy of all the models tested compared to a benchmark, naïve forecasting model. To extend this work it will be necessary to gather better quality data from which forecasts can be produced. This data should be collected following a rigorous protocol to limit bias. Ideally, a range of spatiotemporal units would be used so that the optimal scale for forecasting could be determined, but the evidence we provide here indicates that data should be aggregated in smaller spatial units than the countries of Europe and smaller temporal units than years. It may also be useful to incorporate data on the rate of consumption of antibiotics in hospitals and the community, as well their rate of use in agriculture and aquaculture, in order for the model to construct a more complete picture of the ecological forces which drive the spread of antibiotic resistance.

### Data availability

**Source data**

Antibiotic resistance data for Europe and England can be downloaded from https://www.ecdc.europa.eu/en/antimicrobial-resistance/surveillance-and-disease-data/data-ecdc and https://fingertips.phe.org.uk/profile/amr-local-indicators/ respectively.

Data describing the covariates considered in the analysis of the European data can be downloaded from https://www.ecdc.europa.eu/en/antimicrobial-consumption/surveillance-and-disease-data/database (antibiotic consumption), https://www.who.int/influenza/gisrs_laboratory/fluenet/en/ (prevalence of influenza-like illness), and https://www.imf.org/external/datamapper/NGDP_RPCH@WEO/OEMDC/ADVEC/WEOWORLD (gross domestic product).

Data describing the covariates considered in the analysis of the English data can be downloaded from https://fingertips.phe.org.uk/profile/amr-local-indicators/ (bacteraemia), https://openprescribing.net/ (antibiotic prescribing), and https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/clinicalcommissioninggroupmidyearpopulationestimates (population per CCG).

**Extended data**

Figshare: Predicting the future distribution of antibiotic resistance using time series forecasting and geospatial modelling, extended data.pdf. https://doi.org/10.6084/m9.figshare.12750665\textsuperscript{43}

This project contains the following extended data:

- Predicting the future distribution of antibiotic resistance using time series forecasting and geospatial modelling, extended data.pdf (PDF contain additional Figures)

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

**Software availability**

Code to implement the single time series forecasting functions and inspect their outputs can be installed from http://doi.org/10.5281/zenodo.3960236\textsuperscript{40}, and wrapper functions for building and implementing INLA models can be installed from http://doi.org/10.5281/zenodo.3960236\textsuperscript{30}.

License: MIT

### References

1. Ventola CL: The antibiotic resistance crisis: part 1: causes and threats. P T 2015; 40(4): 277-83. PubMed Abstract | Free Full Text

2. Bin ZS, Hussain MA, Nye R, et al.: A Review on Antibiotic Resistance: Alarm Bells are Ringing. Cureus. 2017; 9(6): e1403. PubMed Abstract | Publisher Full Text | Free Full Text
49. Rue H, Martino S, Chopin N: Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J R Stat Soc Ser B Statistical Methodol.* 2009; 71(2): 319–92. Publisher Full Text

50. Bakka H, Rue H, Fuglstad GA, et al.: Spatial modeling with R-INLA: A review. *Wiley Interdisciplinary Reviews: Computational Statistics.* Wiley-Blackwell; 2018; 10. Publisher Full Text

51. Gneiting T, Raftery AE: Strictly Proper Scoring Rules, Prediction, and Estimation. *J Am Stat Assoc.* 2007; 102(477): 359–78. Publisher Full Text

52. R Core Development Team: R: A language and environment for statistical computing. Vienna, Austria; 2018. Reference Source

53. Hyndman RJ, Khandakar Y: Automatic Time Series Forecasting: The forecast package for R. *J Stat Softw.* 2008; 27(3): 1–22. Reference Source

54. Voelz E, Didelet X: Modeling the growth and decline of pathogen effective population size provides insight into epidemic dynamics and drivers of antimicrobial resistance. *Systematic Biology.* 2017; 67(4): 719–728. Publisher Full Text

55. Hentges DJ: Human intestnl microflorain hlth & disease. Elsevier Science; 1983; 587. Reference Source

56. Conway T, Cohen PS: Commensal and Pathogenic *Escherichia coli* Metabolism in the Gut. *Microbial Spectr.* 2015; 3(3). PubMed Abstract | Publisher Full Text | Free Full Text

57. Podschun R, Ullmann U: Klebsiella spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. *Clin Microbial Rev.* 1998; 11(4): 589–603. PubMed Abstract | Publisher Full Text | Free Full Text

58. Centers for Disease Control and Prevention: *Pseudomonas aeruginosa* in Healthcare Settings. 2019. Reference Source

59. Centers for Disease Control and Prevention: *Acinetobacter* in Healthcare Settings. 2019. Reference Source

60. Dagan R, Barkai G, Givon-Lavi N, et al.: Seasonality of Antibiotic-Resistant *Streptococcus pneumoniae*: That Causes Acute Otitis Media: A Clue for an Antibiotic-Restriction Policy? *J Infect Dis.* 2008; 197(8): 1094–102. PubMed Abstract | Publisher Full Text | Free Full Text

61. Elseviers MM, Ferech M, Vander Stichele RH, et al.: Antibiotic use in ambulatory care in Europe (ESAC data 1997–2002): trends, regional differences and seasonal fluctuations. *Pharmacoepidemiol Drug Saf.* 2007; 16(1): 115–23. PubMed Abstract | Publisher Full Text

62. Cohan FM: What are Bacterial Species? *Annu Rev Microbiol.* 2002; 56: 457–87. PubMed Abstract | Publisher Full Text

63. Cohan FM: Bacterial species and speciation. *Syst Biol.* 2001; 50(4): 513–24. PubMed Abstract | Publisher Full Text

64. Shapiro BJ, Friedman J, Cordero-OX, et al.: Population genomics of early events in the ecological differentiation of bacteria. *Science.* 2012; 335(6077): 48–51. PubMed Abstract | Publisher Full Text | Free Full Text

65. Corander J, Fraser C, Gutmann MU, et al.: Frequency-dependent selection in vaccine-associated pneumococcal population dynamics. *Nat Ecol Evol.* 2017; 1(12): 1950–60. PubMed Abstract | Publisher Full Text | Free Full Text

66. Hyndman RJ, Billah B: Unmasking the Theta method. *Int J Forecast.* 2003; 19(2): 287–90. Publisher Full Text

67. Assimakopoulos V, Nikolopoulos K: The theta model: a decomposition approach to forecasting. *Int J Forecast.* 2000; 16(4): 521–30. Publisher Full Text

68. World Bank: World Bank Country and Lending Groups - World Bank Data Help Desk. Reference Source

69. Benjamin J: bjeffrey92/MyForecaster v1.0.0 (Version v1.0.0). Zenodo. 2020. http://www.doi.org/10.5281/zenodo.3960232

70. Benjamin J: bjeffrey92/MyINLA v1.0.0 (Version v1.0.0). Zenodo. 2020. http://www.doi.org/10.5281/zenodo.3960236
Open Peer Review

Current Peer Review Status: ? ?

Version 1

Reviewer Report 12 January 2021

https://doi.org/10.21956/wellcomeopenres.17733.r41604

© 2021 Opatowski L et al. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Eve Rahbe
Institut Pasteur, Paris, France

Lulla Opatowski
UVSQ, Institut Pasteur, Inserm, University of Paris-Saclay, Paris, France

In this paper, the authors aim at assessing different statistical models in their capacity to predict the trends of antibiotic resistance locally or spatially.

They evaluate 5 previously published forecasting algorithms based on the analysis of two distinct public datasets: yearly antibiotic resistance rates per country in Europe over 2012-16 and quarterly antibiotic resistance rates in England over 2015-18. They also attempt to integrate covariates including antibiotic consumption and GDP into the geospatial model to look for associations.

Results highlight that the naïve forecasting model was the most accurate compared with the others when analysing the European data. When analysing the English data, therefore at a finer scale and finer time resolution, expected trend seasonal forecasts were the most accurate. Some association of antibiotic consumption with resistance rates was also obtained from the geospatial analysis.

The article is well written and raises two interesting and important questions: first, the question of which model performs better at analysing and predicting the trends of antibiotic resistance rates; and second the still unqualified link between covariates including antibiotic exposure, and the evolution of resistance at the population level. I strongly appreciate the model comparison approach, in a context where machine learning techniques are more and more promoted but the choice of which should be the appropriate methodology is still unclear and the impact of this choice on the results is unknown.

However, I had several concerns regarding the present version of the article that I would like to highlight and believe that should be addressed. First because I did not find the conclusions regarding model comparison both for the individual dataset and for the different time and spatial scales completely convincing and believe that perhaps other fitting measurements would be more appropriate. Second, I felt that the datasets and method should be made more transparent and
more detailed. Finally, more discussion on the limits of the data used (short time series for the European datasets) and conclusions is clearly missing.

Main comments:

○ Full description on both datasets is lacking: what resistances are described here? Does the data come from hospital infections or community or both? Does it come from carriage? How representative is it from the country? Can the system be assumed to be stable over time? How was resistance level defined (MIC? Which criteria?). It would be essential to clearly indicate the number of available time points for each time series. The authors mention 6 points for European dataset, but between 2012 and 2016, I count 5 points.

○ Power provided in the data for the analysis: our understanding is that the learning period for the European datasets at least include only 1 (or 3) time points? Is this data appropriate and is it expected to provide sufficient power for the different techniques assessed here? A quick simulation study analysing simulated time series of varying length could provide good insights into that question I believe.

○ It would be critical to show the European individual time series as a supplement figures and provide not only links but tables of analysed data as a supplement file. It is hard for the reader to interpret and understand the analysis of times series without actually seeing their trends. Especially as it is mentioned that many of the individual time series have mean value close to zero … did you consider filtering some of the time series to exclude those including too many zeros? How do they impact the results and model comparison?

○ Models description in the methods section: although I understand the different assessed models have been described previously (ARIMA, ETS and NN autoregression) and are considered as standard methods, it would be important to write up the equations for each of them and include them in the manuscript. Indeed, the model comparison is the objective here, it is therefore very important to understand clearly what the differences between investigated approaches are. It would also be good if the authors could provide equations for the models including the covariates (potentially as a supplement if you think this should not be included in the main text).

○ I am not specialized in time series analysis and can't evaluate the section describing the Teodoro and Levis forecasting method.

○ The spatiotemporal modelling part is confusing: it would be better to first introduce the model mathematically, and then go through the description and justification of the estimation algorithm.

○ Quality criteria for the model predictions: Could the RMSE be qualified as a precision measurement and the second evaluation criteria qualified as a sensitivity measurement? Giving equation 7, it is not clear why RMSE is mentioned in % (page 6, sentence starting with “the worst performing model,”). Perhaps that means that the expression of $e$ is normalize (which is not specified anywhere). Could the authors provide the expression of $e$?

Percentage accuracy of the 95% Prediction Intervals: It would be good if the authors could provide an indicator of the width of those prediction intervals for each model. Indeed,
larger intervals are expected to generate higher % of accuracy don't they? How much does that impact the results presented in Figures 2 and 5?

- Comparison of European and UK data + number of time points: If I understand well, the presented results correspond to an assessment for h=2 for the European data. Were different h assessed? When comparing with the English data, it seems that the comparison is made between h=2 for European (2 years) and quarterly=4 for UK (1 year?). The authors should clearly discuss that point.

- To be honest, such a big difference in median differences of RMSE among the different techniques is not clear to me from the figures and results. Would it be possible to find another way to measure the difference, for example by normalizing over absolute fitted values? That may also help understand what is the role of time series for which many of the individual time series have mean value close to zero (as mentioned on page 6). Can we really differentiate between performances of ARIMA, ETS and naïve on Figure 1? Have the authors looked at the estimation direction of the different models. Do we observe trends with models systematically over estimating or under estimating the data? It is frustrating not to see a single forecasting plot! It would be very useful for the reader to be able to see (at least a selection) model's fits and predictions. Could the authors provide them in supplement, showing model and data over the training period and forecasting period for different models and time series?

- Covariate analysis and associations for individual time series:
  Why are bacteremia used as covariate only for the English data? It is not clear to me what figure 4 and figure 6 shows.

- Regarding correlations, it would be important to provide individual results from the correlation analysis for each time series. Would that be also possible to add colouring in Figure S4 to identify countries?

- Another point is that one may consider that the impact of these variables (or at least some of them such as the antibiotic exposure) is potentially related to the variation and not the crude value of resistance. Indeed, the assessed models include autocorrelation. The correlation is assessed based on the measured value at a given time, assuming, if I understand well, independency of time points and countries. Can one really interpret that signal? One could for example, look as well at the deviation between y_{t+1} and y_t. Also my understanding was that for DDD and ILI at least, temporal time series were available, I therefore do not understand why no attempt was done to incorporate directly the related time series into the models.

- Figure 3: is it possible to include the INLA model with no covariate (to compare and see how much is added by including these covariates?)

- Page 11 and Figure 4. Again the differences between methods are not so obvious for me looking at the figure. Any way to provide more details on this? a global measurement over all species?

Several point of the discussion should be improved:

1. My understanding is that the antibiotic data is in DDD for European dataset and in antibiotic
prescriptions for England dataset. How is it expected to change the results? How do antibiotic consumption in England compare with Europe in general?

2. Discussion on seasonality (page 15). It is mentioned that previous studies have concluded that resistance is highly seasonal. My experience is that it strongly depends on the considered species – are cited references related to the species under study?

3. Page 15, paragraph starting with “comparing figure 5 and figure 6...”, also I generally think that increased spatial and temporal resolution is associated with improved forecast accuracy. I am a bit confused here with the comparison and that may come from a misunderstanding of the analysis carried out. Is the forecasting time scale similar for the two analyses? My understanding was that the average yearly accuracy measurement were calculated based on 2 years forecast for European dataset, whereas it was for a shorter time for the England dataset (maximum 4 quarters). Was the number of data-points used for the training analysis comparable? My understanding is that this number is much higher for the UK. The implications should be discussed.

Other points, clarifications and typos:

- Authors state that there are only two papers that predict future resistance in the introduction section. Reviewers found at least another one that predicts resistance over time (but others may exist): Durham et al. 2010 (using a mixed model, with time and antibiotic consumption as covariates to predict fluoroquinolone resistance in E. coli for many countries in Europe and the US). The authors should provide more details on the choices and filters for their literature search or model selection.

- Equations specifications:
  - Add y in “a time series y decomposed...”
  - In the spatiotemporal part: not clear how the dimension of Q is defined
  - Equation 3: should delta be read instead of d in the equation? Otherwise, what is di? It was not clear for me how the neighbouring elements are defined.
  - Equation 4: what is theta_y_t-1?, I suppose it should read as a multiplication instead of a subscript, again need to define y_t and y_t-1 somewhere...
  - The sentence starting with “SPDE describes the method used to define a GMRF...” is unclear to me.
  - Equation 3 is cited instead of equation 7 at the top of page 6. Would be good also to put all the figures on the same scale.

- All figures: would be good to indicate the name of species on each graph instead of only giving them in the figure legend. Also it would be good to put all subfigures on the same scale.

- Table 2: second column presents the resistance => propose to change label of that column and names it “resistance” or “antibiotic resistance”. Suggest also to modify “antibiotic consumption” and use instead “DDD” which is the variable that was used in the model. Also, could you change the notation of intervals on the last column by writing a proper interval or with a ; as separator. Maybe not necessary to have so much precision.

- Figure 4 & 5 - x-axis and colors have changed, specify the color legend (fill color) to show it has changed from forecasting horizon to forecasting method.
Bar plot vs boxplot? Would a box instead of bar plots for mean of 95% PI provide more information?

Figure 5: add x-axis unit (years, or spatial unit, 1 to 4 and not 0.25 to 1) in the label of forecasting horizon?

Page 6, paragraph starting with “an alternative measure of model …” do you actually assess if the training data are in the interval or is it only from the h points (excluding the T points)?

References
1. Durham LK, Ge M, Cuccia AJ, Quinn JP: Modeling antibiotic resistance to project future rates: quinolone resistance in Escherichia coli. *Eur J Clin Microbiol Infect Dis.* 2010; 29 (3): 353-6 PubMed Abstract | Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
No

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Infectious disease epidemiology, amr.

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.

Reviewer Report 17 November 2020

https://doi.org/10.21956/wellcomeopenres.17733.r41205
This paper by Jeffrey et al. represents a solid effort to make predictions of antibiotic resistance based on surveillance data. It is however worth highlighting that it also illustrates how challenging making accurate predictions are, as naïve projections based on recent surveillance data performed better or as well as the tested models in predicting the future ABR distributions in the majority of comparisons.

The authors analyze two prevalence datasets of antibiotic non-susceptibility. The first is from a European surveillance network with the organisms *Klebsiella pneumonia*, *Escherichia coli*, *Pseudomonas aeruginosa*, and Acinetobacter spp. and the drugs carbapenems and fluoroquinolones, spanning the period 2012-2016. The second is from England with 195 clinical commissioning groups England with data on *E. coli* and data on the drugs ciprofloxacin, gentamicin, piperacillin/tazobactam, and 3rd generation cephalosporins, spanning the period 2015-2019.

The paper is well written and employs appropriate methods to predict the percentage of resistant isolates. The authors compare the forecast of individual time series in each location with forecasting that incorporates spatiotemporal modeling. They also investigate the relationship between the observed percentages of resistant isolates with the covariates GDP, antibiotic consumption, and Influenza-like illness. We do however have some suggestions for improvements and clarifications:

Major comments and questions:

1. Figure 2, and Figure 5: I am a little surprised about the poor performance of the Neural Net forecasting. The author states that the poor performance observed is likely due to overfitting. From the code in “forecasting_functions.R”, the neural net seems to have been fitted with the default setting (fit <- forecast::nnetar(training_data)). Could the neural net perform better if tuned appropriately? In the discussion, the authors mention that “applying complex machine learning models to time series forecasting…. they can often be reduced to simpler models”. For the time series, a naïve forecasting model performed very well. Is the neural net unable to approximate a simple model in this case? The paper does not include any figures for the actual forecasts. If you inspect these, what seems to be the problem with the NN-forecast, violent fluctuations?

2. “We hypothesized that there could be a more complex association driven by the spatial distribution of the covariates and potentially a lag in the emergence of resistance following a change in the level of a given covariate. “
   This is a very interesting question. It would be interesting to have a look at the raw pattern
in the data. Figure 4 and Figure 6 in the extended data shows the relationship between the resistance data and the covariates. If you make similar figures for the resistance data shifted forward in time, are there any patterns for the covariates at different lag times? Similarly, are there any relationship between the first-order difference in the resistance data (the change in resistance percentages) and the covariates, or a lagged series of the covariates?

3. In the discussion, the authors state: “Comparing Figure 5 and Figure 6 (the RMSE and MEAN PI) ... with Figure 2 and Figure 3 ... demonstrates that the increased spatial and temporal resolution of the data is associated with increased accuracy.” This might be true, but these are different datasets. Could this be addressed more directly using the English data only? If you construct a new time series from the English dataset, averaging the data over 6mo periods, does this decrease the accuracy of the prediction? In the same way, the CCGs could be combined with their neighboring CCG’s to see if the spatial component contributes significantly.

Minor comments:
The figures would benefit from panel labels indicating the species and/or drug included in each panel, as this is a lot easier for the reader to follow than having to read through the figure legends in each case.

Introduction: “It has been demonstrated that model structure can have a highly significant influence on the predictions of epidemiological studies”. This is quite general, and could perhaps be rephrased in a more specific manner.

First paragraph page 6. “Equation 3” should be “equation 7”.

Is there are reason why bacteraemia was only included as a covariate in the INLA model?

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.
**Reviewer Expertise:** Microbiology, genomics, AMR, mathematical modelling.

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.