CSE 8803 EPI: Data Science for Epidemiology, Fall 2024

Lecturer: B. Aditya Prakash October 10;17 2024
Scribe: Grace Driskill, Lauren Fogel, Anoch Mohan Lecture 15 : FORECASTING I

---

1 Summary

So far we have looked at methods of modeling diseases, detecting outbreaks, and surveilling their spread. The next step here is to use our modeling of the network and surveillance techniques to try to predict the future of the disease spread. Predicting disease spread is essential for making informed public policy decisions and taking action to minimize the impact of the disease on the population. While the influx of large data collection sources has accelerated research into real-time epidemic forecasting, it is imperative to incorporate the different causal factors while forecasting to ensure accurate predictions.

Epidemic Forecasting can be split into different categories based on the tasks in hand, indicators of interest, and the spatial/temporal scale. These categories also dictate how we evaluate our model's forecasts. We further look into hybrid models which combine the previously discussed mechanistic models with machine learning methods. These models are able to efficiently utilize the variety of datasets available while incorporating the domain-based prior knowledge of mechanistic models.

2 Why Forecasting?

Forecasting diseases is very similar to weather forecasting in the sense that it gives people a sense of what to expect. Additionally, it allows governments to prepare medical resources and make informed public policy decisions, such as allocating budget for ventilators or placing mandates. Not just the government, at some level an individual is a stakeholder too. We make many decisions based on weather forecasts and similarly having insights about public health we would be able to take that into consideration when making a decision. The importance of forecasting makes it crucial that it is accurate and incorporates the different causal factors.

The causal factors mainly are -

- Current Number of Infections
- Interventions in place
- Contact patterns
- Exposure to disease

The advent of social media and communication technologies has brought along vast amounts of data which at first sight might seem unrelated to epidemics but, in fact, are very useful. Google Search Trends, Facebook's Covid Impact Survey, and Safe Graph's mobility data are just some of the examples of different types of data available and are an *indirect* measure of the causal factors. Advancements in machine learning have also made models capable of ingesting such data readily available. Example citation [1]. There are non-linearities in data that traditional methods like ODE model and agent-based methods can't handle.

3 Epidemic Forecasting

The epidemic forecasting pipeline typically consists of data processing, model training & validation, and utilization & decision-making components, as shown in Fig. 3. This section focuses on the objectives of model training & validation. Specifically, we will discuss the following four aspects sequentially: Forecasting Tasks, Targets of Interest, Spatial and Temporal Scale, and Model Evaluation.

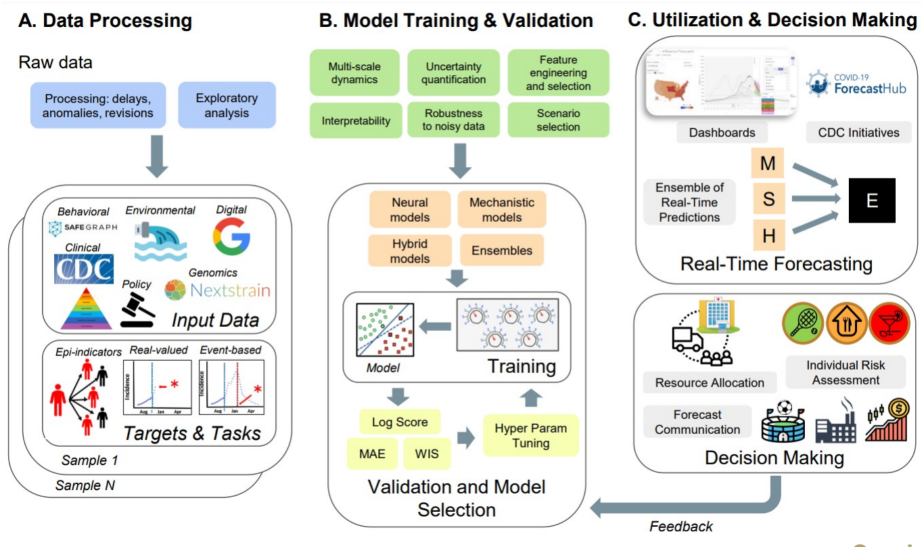

Figure 1: The illustration of epidemic forecasting pipeline

3.1 Forecasting Tasks

Forecasting tasks can be broadly separated into three categories:

- **Real Valued Predictions** - look at epidemic indicators like mortality rate and cases to try and understand the trend of the disease. These can be predicting the future incidence or the peak intensity of the epidemic. For example, during the COVID-19 pandemic, real-valued predictions were used to estimate the number of infections and deaths in different countries. This information was crucial for governments and health organizations to allocate resources and implement containment measures.

  Since much of the dataset is updated 1–3 weeks in the past, such delays in data make nowcasting, predicting the current values, useful. Short-term forecasting of 4 weeks or less ahead helps increase preparedness. Long-term forecasting of greater than 4 weeks ahead, although challenging, can help us understand how the epidemic is going to pan out. However, long-term predictions may not be useful, since there may be some policies implemented or interventions that might take place that would change the curve completely. They have some silent assumptions that kind of make them a projection itself.

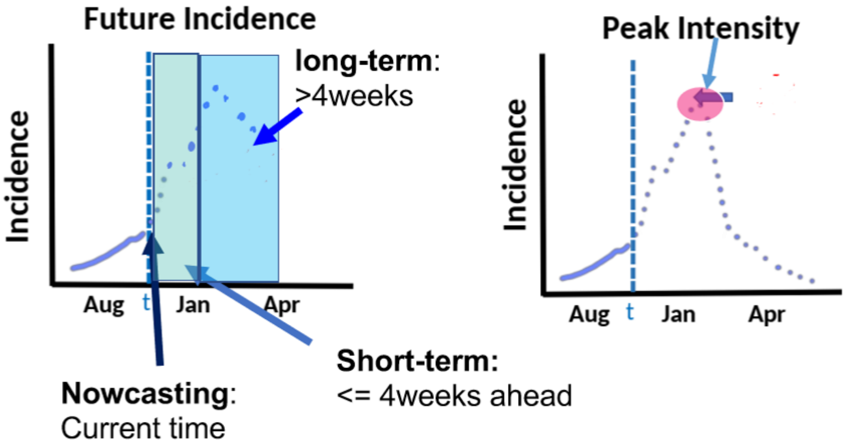

Figure 2: The illustration of real-valued predictions.

- **Event-based Predictions** - look into identifying times of particular events related to the epidemic. Events such as peak-time and onset time, which are indicators of the intensity of the epidemic [5]. Peak-time refers to the time when the peak value is observed. Onset refers to the time when an indicator first surpasses a baseline decided by forecast organizers. For example, the CDC sets a baseline, for each US region, as the average incidence value during non-flu season from the past three years. These events are used as signals for interventions like shutting off schools and vaccination drives. For instance, during the H1N1 influenza pandemic in 2009, event-based predictions were employed to determine the best time to close schools and initiate vaccination campaigns.

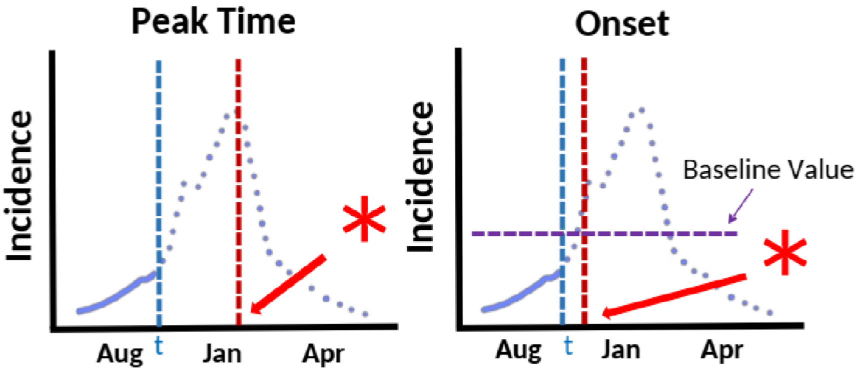

Figure 3: The illustration of event-based predictions.

- **Epidemiological indicator Predictions** - look into the composite indicators that characterize the behavior of the epidemic, like the reproduction number and the final infected size [2]. Reproduction number is the expected number of secondary infections caused by one infected individual. Final size is the total fraction of the population that will be infected over the course of an epidemic. Characterizing the final size can be challenging for diseases that allow for reinfection, like COVID-19, which is why this metric is usually only considered when the disease dynamics are well understood.

These indicators help in understanding the potential spread and severity of an epidemic. For example, during the Ebola outbreak in 2014, epidemiological indicator predictions were used to estimate the basic reproduction number (R0) and the final infected size. This information was valuable in guiding public health response and intervention strategies.

3.2 Targets of Interest

The target of interest for a forecasting model will depend on what the target disease is, what data is available, and what surveillance is being done. The typical important indicators analyzed are number of cases, mortality, and hospitalizations. Regarding influenza, the percentage of influenza-like-illnesses (%ILI) and lab-tested hospitalizations are common targets of interest. %ILI is an example of syndromic surveillance, as it keeps track of outpatients experiencing flu-like symptoms that are reported by health care providers to the CDC Fig. 4. Regarding COVID-19, reported deaths, hospitalization and cases are common indicators that may be used for targets of interest. It is important to note that these indicators might not always be accurate because during the COVID season, Influenza %ILI might get mixed with symptomatic COVID outpatients, so additional indicators like lab-tested hospitalizations would be needed.

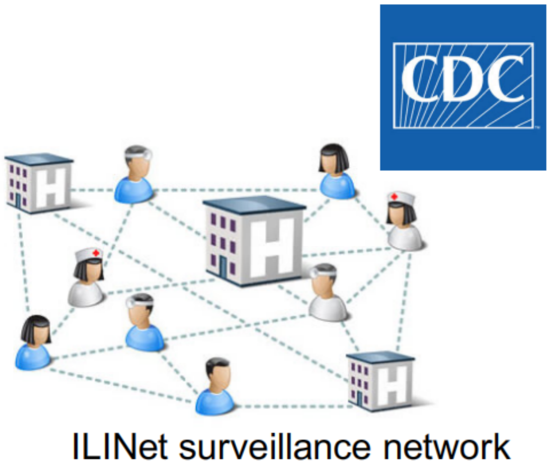

Figure 4: The illustration of an influenza-like illness surveillance network.

3.3 Spatial and Temporal Scales

Forecasting of epidemics is done on different spatial granularity [4]. Regions are grouped together to make monitoring and reporting easier. For example, during the COVID-19 pandemic, forecasts were made at various spatial levels, such as global, national, and even regional or city levels. This allowed authorities to identify hotspots, allocate resources effectively, and implement targeted interventions.

Additionally, forecasts are made for different time scales like weekly and daily. For instance, the Centers for Disease Control and Prevention (CDC) in the United States provides weekly forecasts of COVID-19 cases, hospitalizations, and deaths. On a daily basis, the World Health Organization (WHO) updates its COVID-19 dashboard with the latest data on cases and deaths worldwide. These varying time scales help policymakers and public health officials assess the current state of the epidemic, monitor trends, and evaluate the effectiveness of interventions. Generally, the more you aggregate data either spatially or temporally, the more your model becomes robust.

3.4 Model Evaluation

The next most important step after training models is to evaluate them. We want to measure the success of our predictions, and the success might be different in various forecasting scenarios. Forecasting outcomes can be split into two categories:

- Point Forecasts
- Probabilistic Forecasts

Point Forecasts have single-valued results like the number of infections or deaths. Metrics used for point forecasting include RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), and MAPE (Mean Absolute Percentage Error). RMSE and MAE measure the error in L2 and L1 norms, respectively, whereas MAPE allows us to get the ballpark of our prediction using the prediction error. For example, if a model is predicting the number of daily new cases for COVID-19, RMSE would penalize larger errors more severely, while MAE would give equal weight to all errors. MAPE, on the other hand, would provide a relative measure of the error, making it easier to compare the model's performance across different magnitudes of cases.

Since these forecasts are of great social and economic significance, it becomes essential that we are certain about our predictions. Point forecasts, however, are not able to capture the confidence of the model, and this is why probabilistic forecasts are preferred. Probabilistic forecasts [8] capture the uncertainty of model predictions by using confidence intervals while also considering the accuracy of the model, as shown in Fig. 5. For example, a probabilistic forecast might predict a range of possible cases (e.g., 1000-2000) rather than a single number (e.g., 1500). This approach provides a better understanding of the uncertainty in the model's predictions.

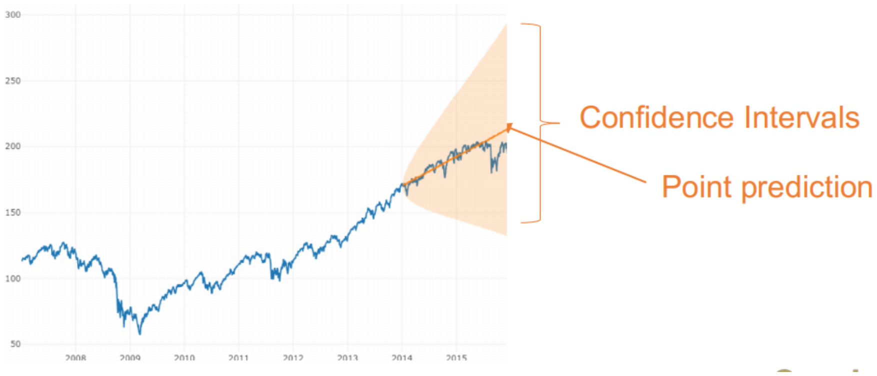

Figure 5: The illustration of point and probabilistic forecasts

The Log Score calculates the binned log probability of the ground truth, while Interval scores are used to penalize how far the model's predictions are from the ground truth. Additionally, we want to penalize flat distributions because that means the model is not confident in predicting and assigns equal probability to all outcomes. We also want to penalize if the ground truth is lower than the lower bound and greater than the upper bound. Coverage score is another metric that measures the fraction of times the ground truth actually lies in the confidence interval or, in simple words, penalizes an overconfident model. For example, a model with a high coverage score would have a higher percentage of its predictions containing the actual number of cases within its confidence interval, indicating its ability to account for uncertainty more accurately.

More recently, researchers have adopted the Weighted Interval Score (WIS), which aggregates interval scores for multiple intervals and aggregates them. WIS also builds upon log score in the sense that it is unbounded, and especially for forecasting diseases like COVID-19, it is a more suitable metric. WIS combines the advantages of both interval scores and log scores, making it a comprehensive metric for evaluating probabilistic forecasts. It helps assess a model's ability to accurately predict the range of possible outcomes while considering the uncertainty inherent in the data.

Researchers are still trying to find better ways of measuring success and incorporating qualitative metrics. Just giving the number isn't enough. It still needs to be useful for epidemiologists to make decisions.

4 Modeling Paradigms

[7] talks about different types of paradigms, datasets used by them, the type of task, and features used.

4.1 Mechanistic Models

These models have been thoroughly discussed in the previous lectures. Mechanistic models are the workhorses in epidemiology. The population is divided into different compartments based on the disease state, and they move between compartments based on the disease progression. ODEs like the SIR model, Metapopulation models, and Agent-based models (ABMs) are the primary examples. These models require domain knowledge and their parameters require intensive testing for sensitivity. The future of these models include mechanistic forecasting using mobility networks to expand model heterogeneity, improving speed and efficiency using ABMs, and real-time forecasting using ML and Bayesian frameworks.

4.1.1 Forecasting

The methods and models described here are most useful for long-term forecasting, and not very accurate when it comes to immediate forecasting.

There is a lack of real-world features in models tracking mobility networks for disease spread. This paper [3] uses multi-scale mobility networks to better model the spread of infectious disease. With limited empirical data, the GLEaM (Global Epidemic and Mobility) model is a computational framework that integrates population data, mobility patterns, and epidemic dynamics to simulate the spread of diseases globally. Worldwide populations are split up into subpopulations and fed into the three-layer model, made up of a population layer, which splits up the population, mobility layer, and epidemic layer. A gravity model is used to describe commuting patterns, where the flow of people between two subpopulations is a function of their populations and the distance between them. The model uses real-world data to capture the complex nature of human movement, combining it with stochastic simulations of disease transmission across the mobility networks. This paper finds that smaller-scale commuting flows are much larger data wise than something worldwide like air travel, but air travel largely shapes epidemic patterns across the globe. Also, short-range mobility, while it does not significantly alter global epidemic patterns, enhances synchronization among nearby populations and affects the spread in regions weakly connected by air travel.

4.2 Statistical/ML Models

Advancements in machine learning methods have led to a large-scale collection and maintenance of publicly available datasets. Newer architectures have been successful in finding patterns even in complex data forms. Due to their success, numerous optimization algorithms are also available. Some approaches are regression-based, language and/or vision models, neural networks, and density estimation models. ML models [6] are useful as they can handle different types of datasets like Languages, Images, time series, etc.

The idea here is to find a function that can forecast the target based on the input data, however, do note that this would generate an approximate forecast. This is achieved by minimizing the comparison between the prediction from the function and the ground truth. The comparison is called Loss, and the function that generates the comparison is known as a loss function. The goal is to minimize the loss generated by the loss function, generally best found using past training data.

4.2.1 Specific Model Types

1. Regression Models
   Regression models assume a linear relationship between input features and the forecast of the future features:
   $$\tilde{y} = w_0 + w^T x$$
   It is most useful to use more than just linear regression and couple it with another method, such as Nearest Neighbor Based Regression.
   AutoRegressive models are also an interesting way to use past values of an epidemic curve to model the future. An example of this can be found in the ARGO paper, which uses AutoRegression on Google search data [9].

2. Language and Vision Models
   Language and vision models take in large sources of data, like online text data from social media (language) or satellite data with images of places sensitive to outbreak (vision) to use in their models. These often incorporate the use of other techniques types, such as regression models, to use for modeling. This is an up and coming area of research, so it has a lot of unknowns.

3. Neural Models (Deep Learning)
   Deep learning models are very useful, as they capture non-linear patterns in high-dimensional data without requiring lots of assumptions. They are very flexible wen it

comes to learning complex domains, as well as leveraging a multitude of data sources. In fact, it is said that a single layer neural model, if it is large/wide enough, can model almost any function. There are many different types of neural models that are specialized to different data types and desired outcomes, making them varied and widespread when it comes to areas of usefulness.

4.3 Hybrid Models

Given the widespread use of ML models, an ML engineer doesn't need to have the domain knowledge to apply a model on an epidemic data set - this might work in certain scenarios, but in most scenarios, domain knowledge combined with an ML model would produce better and more reliable results. Hybrid models offer a powerful approach in modeling by blending mechanistic and statistical frameworks, leveraging the strengths of both paradigms. These models are particularly useful in complex, real-world applications like disease forecasting, where both biological mechanisms and data-driven patterns play key roles. The flexibility of hybrid models comes from their ability to combine the structure and interpretability of mechanistic models with the adaptability of statistical models, which can incorporate large, diverse datasets that mechanistic models alone struggle to handle. For instance, hybrid models can integrate environmental data (e.g., temperature, humidity) or genomic data (e.g., viral mutations) in ways that traditional mechanistic models cannot.

By utilizing both model types, hybrid models improve calibration, especially in real-time forecasting. While mechanistic models may face challenges with parameter estimation due to sparse or noisy data, statistical models can step in to provide more accurate estimates or adjust outputs based on observed data. This leads to more accurate overall predictions, particularly in short-term forecasts where real-time data is crucial. By incorporating mechanistic priors into statistical models, hybrid approaches ensure predictions remain grounded in biological reality, while benefiting from the scalability and flexibility of machine learning techniques.

However, hybrid models also present challenges. Their complexity requires expertise in both mechanistic and statistical modeling, and combining these approaches is not always straightforward. Care must be taken to ensure the two models complement each other without introducing conflicting assumptions. Additionally, calibrating hybrid models can be computationally intensive, especially when dealing with large datasets, often requiring specialized software or high-performance computing resources. The accuracy of hybrid models heavily depends on the quality of input data, which in many epidemiological settings can be sparse, noisy, or incomplete. Addressing these data challenges and ensuring that hybrid models can handle missing or inconsistent information is critical for their successful implementation.

References

[1] B. Adhikari, X. Xu, N. Ramakrishnan, and B. A. Prakash. Epideep: Exploiting embeddings for epidemic forecasting. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 577–586, 2019.

[2] E. Andersson, S. Kühlmann-Berenzon, A. Linde, L. Schiöler, S. Rubinova, and M. Frisén. Predictions by early indicators of the time and height of the peaks of yearly influenza outbreaks in sweden. *Scandinavian Journal of Public Health*, 36(5):475–482, 2008.

[3] D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J. J. Ramasco, and A. Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the national academy of sciences*, 106(51):21484–21489, 2009.

[4] A. N. Desai, M. U. Kraemer, S. Bhatia, A. Cori, P. Nouvellet, M. Herringer, E. L. Cohn, M. Carrion, J. S. Brownstein, L. C. Madoff, et al. Real-time epidemic forecasting: challenges and opportunities. *Health security*, 17(4):268–275, 2019.

[5] A. Kattan, S. Fatima, and M. Arif. Time-series event-based prediction: An unsupervised learning framework based on genetic programming. *Information Sciences*, 301:99–123, 2015.

[6] B. Lim and S. Zohren. Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A*, 379(2194):20200209, 2021.

[7] A. Rodríguez, H. Kamarthi, P. Agarwal, J. Ho, M. Patel, S. Sapre, and B. A. Prakash. Data-centric epidemic forecasting: A survey, 2022.

[8] D. Salinas, V. Flunkert, J. Gasthaus, and T. Januschowski. Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3):1181–1191, 2020.

[9] S. Yang, M. Santillana, and S. C. Kou. Accurate estimation of influenza epidemics using google search data via argo. *Proceedings of the National Academy of Sciences*, 112(47):14473–14478, 2015.