Influenza forecasting framework based on Gaussian processes

Christoph Zimmer(1), Reza Yaesoubi(2)

(1) Bosch Center for Artificial Intelligence,
(2) Yale School of Public Health

International Conference on Machine Learning 2020

Short Summary

Why forecasting seasonal epidemics?

Seasonal epidemics

- pose high burden on public health
- vary from year to year

$\Rightarrow$ Forecast to allocate health ressources

Short Summary

Why forecasting seasonal epidemics?	Algorithms for forecasting
Seasonal epidemics ■ pose high burden on public health ■ vary from year to year $\Rightarrow$ Forecast to allocate health ressources	■ Physical models (e.g. ODE with SIR) ■ Time series models, e.g. SARIMA ■ NOVELTY HERE: Gaussian processes

BOSCH

Short Summary

Why forecasting seasonal epidemics?

Seasonal epidemics

- pose high burden on public health
- vary from year to year

$\Rightarrow$ Forecast to allocate health ressources

Algorithms for forecasting

- Physical models (e.g. ODE with SIR)
- Time series models, e.g. SARIMA
- NOVELTY HERE: Gaussian processes

Our contribution

- Precise point forecasts
- Reliable uncertainty quantification
- Competetive results to state of the art benchmarks

Short Summary

Why forecasting seasonal epidemics?

Seasonal epidemics

- pose high burden on public health
- vary from year to year

$\Rightarrow$ Forecast to allocate health ressources

Algorithms for forecasting

- Physical models (e.g. ODE with SIR)
- Time series models, e.g. SARIMA
- NOVELTY HERE: Gaussian processes

Our contribution

- Precise point forecasts
- Reliable uncertainty quantification
- Competetive results to state of the art benchmarks

Use cases

- Retrospective forecasts on Center for Disease Control and Prevention (CDC) influenza-like illness (ILI) data
- CDC hosts a yearly challenge on ILI forecasting

Introduction

Seasonal Influenza

Seasonal influenza causes a tremendous burden on public health each year.

In the US alone

- 9.2 - 35.6 million cases
- 140 - 710.000 hospitalizations
- 12000 - 56.000 deaths

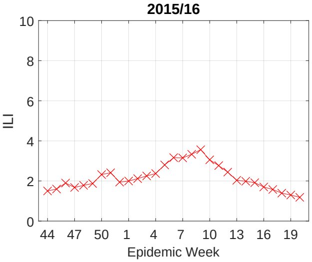
Why do we need forecasting?

The Center for Disease Control and Prevention (CDC) in the US tracks Influenza-like illness.

Why do we need forecasting?

The Center for Disease Control and Prevention (CDC) in the US tracks Influenza-like illness.

Size and timing of influenza-like illness epidemics very different from year to year.

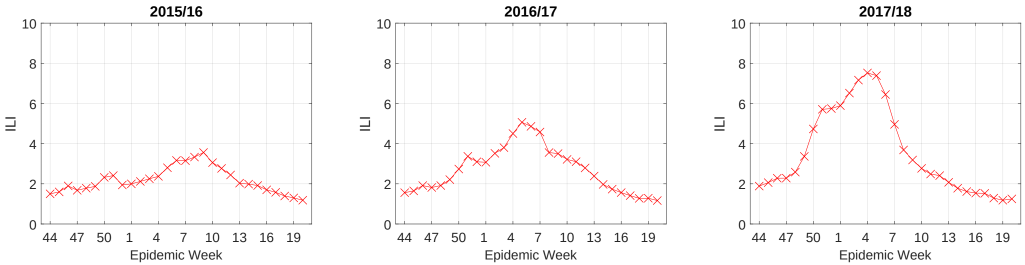


| Christoph Zimmer | 2020-06-15

Why do we need forecasting?

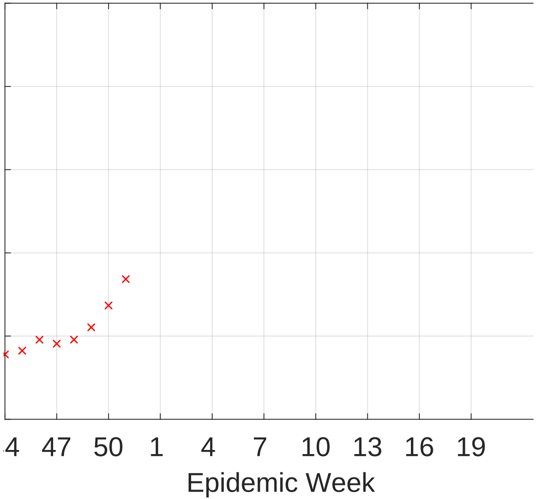
The Center for Disease Control and Prevention (CDC) in the US tracks Influenza-like illness.

Size and timing of influenza-like illness epidemics very different from year to year.

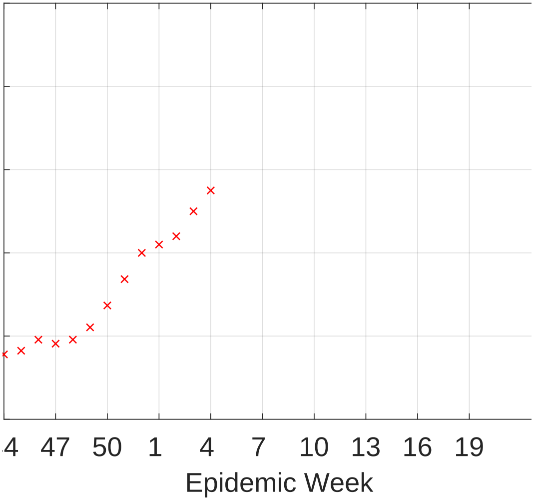


Need for forecasting to allocate public health resources.

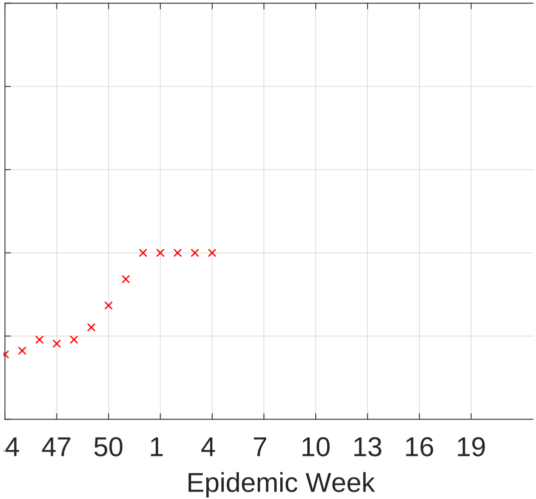
Motivation -- data seen so far, how is it going to continue?



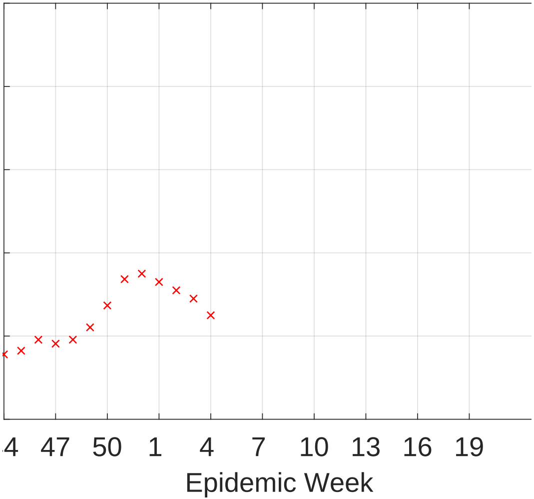
Motivation -- like this?



Motivation -- or this?



Motivation -- or like this?



Algorithm -- Intuition

We have seen previous years, so we can use machine learning.

Algorithm -- Intuition

- Let us assume that we are in week 5 of year 2015
- Let us assume we want to predict week 8 of year 2015.
- We look how past seasons weeks 1-5 have impacted week 8
- Therefore, input training data is weeks 1-5 of 2010, 1-5 of 2011, 1-5 of 2012,...
- Output training data is week 8 of those years
- We want to evaluate the model at week 1-5 of 2015 and predict week 8 of 2015.

BOSCH

Algorithm -- more formally

Seasonal epidemics forecasting framework

Input: current week j^*, current year i^*, forecasting horizon T, one or more feature set of past weeks $J_1, \ldots, J_N$ and seasons I, data recorded so far d_j^i for $j \leq j^*$ and $i \leq i^*$.

FOR $t = 1$ to T % 1 week to T week forecasts

FOR $l = 1$ to N % ensembles

1. Assemble target T specific training data inputs: $\mathcal{X}_{j^*}^{i^*} = (d_j^i | j \in J_l \; , \; i \in I)$
2. and training data outputs $\mathcal{Y}_{j^*}^{i^*} = (y_{t,j^*}^i | i \in I)$
3. Train a GP based on $\{\mathcal{X}_{j^*}^{i^*}, \mathcal{Y}_{j^*}^{i^*}\}$
4. Forecast target according $p\left(y_{T,j^*}^{i^*} | x^{i^*}, \mathcal{X}_{j^*}^{i^*}, \mathcal{Y}_{j^*}^{i^*}\right)$, resulting in μ_l and σ_l

ENDFOR

Build ensemble forecast over N members

ENDFOR

How can we test whether our framework produces accurate and reliable forecasts?

Retrospective Testing

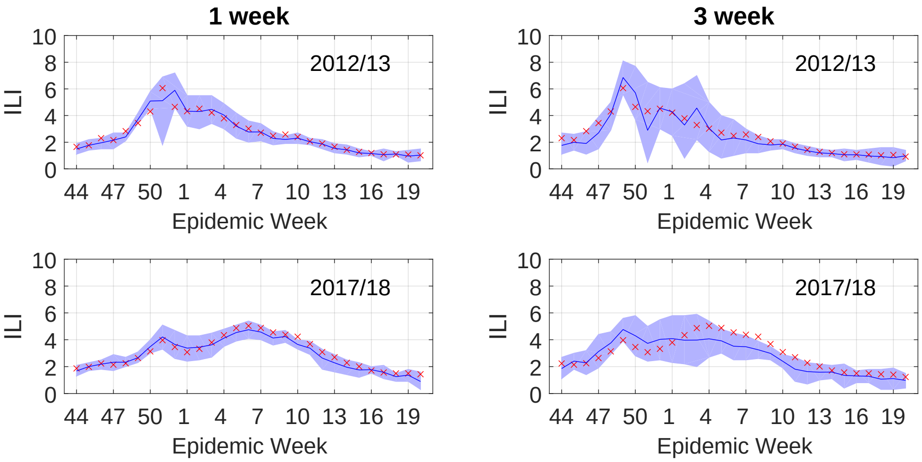
- We use the seasons 2003/04 - 2007/08 as training data
- We use the seasons 2010/11 and 2011/12 as validation (for feature selection)
- We use the seasons 2012/13 - 2018/19 as test data

How can we test whether our framework produces accurate and reliable forecasts?

Retrospective Testing

- We use the seasons 2003/04 - 2007/08 as training data
- We use the seasons 2010/11 and 2011/12 as validation (for feature selection)
- We use the seasons 2012/13 - 2018/19 as test data
- We do retrospective forecasting for each week and target of the test seasons
- Retrospective forecasting means that we do only use data that has been available until the timepoint of forecast
- Targets are 1-4 week forecasts

Retrospective forecasting -- Prediction intervals



Red -- observed value; blue line -- mean prediction,
blue shaded area -- 95% prediction intervals

Our framework for influenza forecasting in action

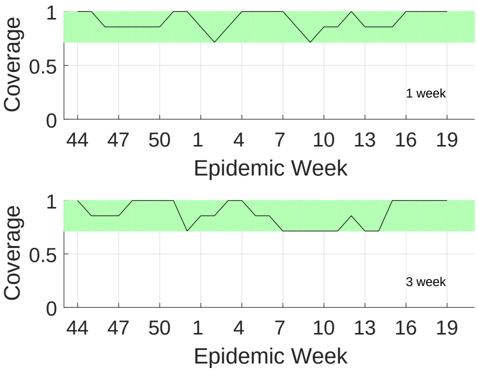
Retrospective Forecasting -- Reliability of uncertainty quantification

Fraction of true values within the 95% prediction intervals (black line).

This is a binomially distributed random number,

we can add its 95% confidence intervals (green shaded area)

⇒ Our framework yields reliable uncertainty estimation.



BOSCH

Benchmarking our framework against state of the art

How to compare probabilistic forecasts?

We use a log-score: logarithm of probability in certain interval around true value.

Benchmarking our framework against state of the art

How to compare probabilistic forecasts?

We use a log-score: logarithm of probability in certain interval around true value.

State of the Art Benchmarks

- (A) Historical averages.

Benchmarking our framework against state of the art

How to compare probabilistic forecasts?

We use a log-score: logarithm of probability in certain interval around true value.

State of the Art Benchmarks

- (A) Historical averages.
- (B) MSS is a recently published framework based on a humidity based SIRS model and a linear noise approximation (Zimmer et al., 2018, SMMR).

Benchmarking our framework against state of the art

How to compare probabilistic forecasts?

We use a log-score: logarithm of probability in certain interval around true value.

State of the Art Benchmarks

- (A) Historical averages.
- (B) MSS is a recently published framework based on a humidity based SIRS model and a linear noise approximation (Zimmer et al., 2018, SMMR).
- (C) Linear regression uses linear models with different sets of past weeks as features. LinEns is a average ensemble over the three best linear models.

Benchmarking our framework against state of the art

How to compare probabilistic forecasts?

We use a log-score: logarithm of probability in certain interval around true value.

State of the Art Benchmarks

- (A) Historical averages.
- (B) MSS is a recently published framework based on a humidity based SIRS model and a linear noise approximation (Zimmer et al., 2018, SMMR).
- (C) Linear regression uses linear models with different sets of past weeks as features. LinEns is a average ensemble over the three best linear models.
- (D) Sarima uses Seasonal auto regressive integrated moving average models as are also used in (Ray et al., 2017, Stat. Med.).

Benchmarking our framework against state of the art

How to compare probabilistic forecasts?

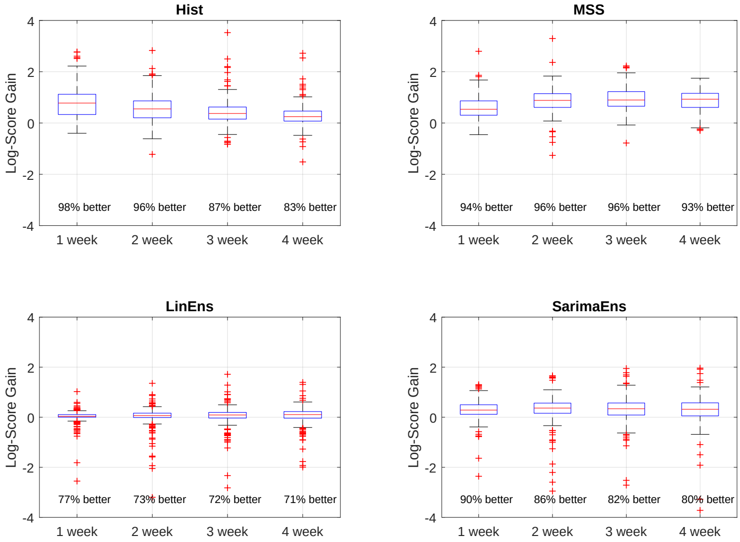
We use a log-score: logarithm of probability in certain interval around true value.

State of the Art Benchmarks

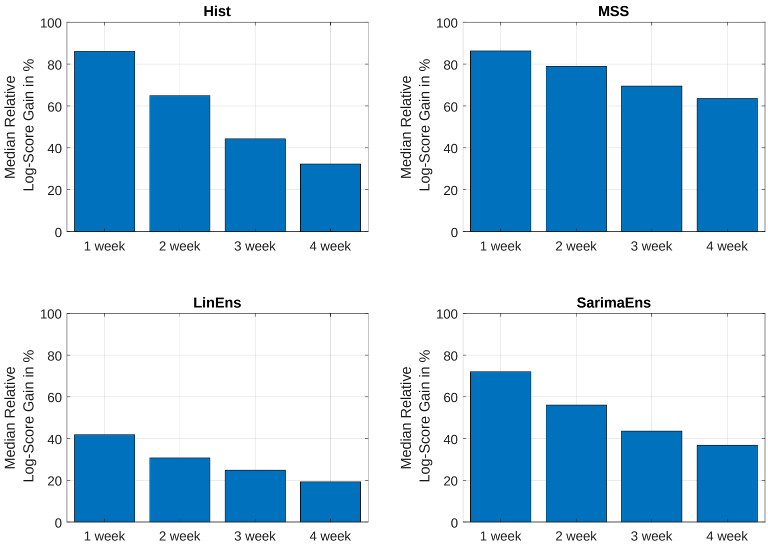
- (A) Historical averages.
- (B) MSS is a recently published framework based on a humidity based SIRS model and a linear noise approximation (Zimmer et al., 2018, SMMR).
- (C) Linear regression uses linear models with different sets of past weeks as features. LinEns is a average ensemble over the three best linear models.
- (D) Sarima uses Seasonal auto regressive integrated moving average models as are also used in (Ray et al., 2017, Stat. Med.).
- (E) Epideep is a recently developed deep learning based influenza forecasting framework (Adhikari et al., 2019, KDD).

BOSCH

State of the art benchmarking: results

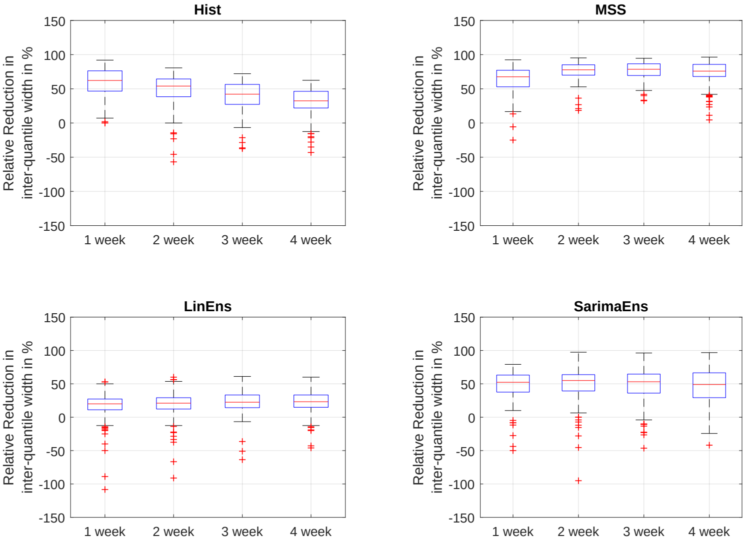


State of the art benchmarking: results



BOSCH

State of the art benchmarking: results



 | Christoph Zimmer | 2020-06-15

BOSCH

Applications / use cases

Center for Disease Control and Prevention (CDC)

- CDC hosts a yearly influenza-like illness forecasting challenge
- Retrospective forecasts of this paper were using same data and forecasting targets

Applications / use cases

Center for Disease Control and Prevention (CDC)

- CDC hosts a yearly influenza-like illness forecasting challenge
- Retrospective forecasts of this paper were using same data and forecasting targets

Join influenza-like illness forecasting

You also have a good algorithm? Come, join the challenge, get a benchmark for your algorithm and help CDC's efforts responding to seasonal epidemics!

- predict.cdc.gov/
- www.cdc.gov/flu/weekly/flusight/index.html
- www.cdc.gov/coronavirus/2019-ncov/covid-data/mathematical-modeling.html

Applications / use cases

Center for Disease Control and Prevention (CDC)

- CDC hosts a yearly influenza-like illness forecasting challenge
- Retrospective forecasts of this paper were using same data and forecasting targets

Join influenza-like illness forecasting

You also have a good algorithm? Come, join the challenge, get a benchmark for your algorithm and help CDC’s efforts responding to seasonal epidemics!

- predict.cdc.gov/
- www.cdc.gov/flu/weekly/flusight/index.html
- www.cdc.gov/coronavirus/2019-ncov/covid-data/mathematical-modeling.html

Question/Comments? → christoph.zimmer@de.bosch.com

Thank you for your attention!