

# 1. Evaluation

- focus on CRPS
- look at IQR or other ranges → see whether 50 percent of the observations fall inside the IQR
- compare to 1 day-ahead-forecasts (i.e. multiples of 1 day ahead forecasts. Mail to Fabian Krüger to ask whether ratios or differences of crps are in any way meaningful)
- Diebold-Mariano-Tests (basically t-Tests for scores to see whether they are significantly different, see e.g. Gneiting and Katzfuss, Annual Review of Statistics and Its Application, 2014).
- would be good to also have something more interpretable like MSE or absolute error for intuition.
- I would like to see a plot where we forecast like 200 days into the future, just to get a feeling what the forecast looks like asymptotically.

# 2. Models and Estimation

- AutoAR
  - how many data points do we actually want to include? All past data points?
  - maybe have a mechanism that resets the number of past data points once predictions deteriorate? ⇒ probably should have a look how quickly the model is able to correct for changes in trends.
  - what about any kind of trend? AR assumes basically a constant mean.
  - should we maybe to classical time series analysis things like (p)AC plots, differencing, stationarity checks etc.?
  - do we have to pay attention to stationarity?
- external regressors
  - code certain shocks / interventions as 0/1?
  - use google searches?  
⇒ general idea: can we make use of the delay between infection/symptom/report and use lagged predictors?
- think about multivariate time series? lags ⇒ probably makes less sense now with lock downs.
- Post-hoc correction of predictions
  - some models seem to be consistently underestimating  $R_t$ -values. Can we simply multiply these with a factor to make sure bias is equal to 0.5? One way would be to use model stacking, if we e.g. took the 90% quantile of the forecast in and treat it as a different model, that could draw all estimates upwards where needed.
  - Can we correct the predictive distributions so that they are well calibrated?  
⇒ need to know about correlation between Bias and bias distribution over time.
- Model Stacking
  - use Regression where all parameters are constrained to sum up to one.
  - literature research: find other ways to do that.
- Forecast package

- offers state space exponential smoothing models and ARIMA models.
- Also a theta function (don't know what it does, it creates a straight line) and things like linear forecast from cubic spline. We don't want that.
- Also offers a tsdisplay function for showing ACF and PACF. That might be useful.
- has a checkresiduals function for diagnostics. Not sure we care too much about it if the predictions are fine.
- automatically chooses a model from the chosen class ETS or ARIMA.
- offers either calculated or simulated prediction intervals
- offers running a fit model on new data for validation. Do we want that?
- there is also an interesting thing about grouped timeseries here: <https://otexts.com/fpp3/gts.html>
- fable package
  - Like forecast, but better. Has auto ETS and auto ARIMA functions and also Auto Plots if we want them.
  - current issues to resolve
- opera package
  - does model stacking. Handling is a bit clunky, but it should work. Problem: what are sensible predictive intervals?
  - also: what are the weights. Should we just use the latest weights and keep them constant? Should we use the average weight over the last x days? Should we follow a trend in these weights
  - ask whether there is any support like that in fable implemented.

### 3. Time Series Considerations

- Forecast package offers ARIMA models and exponential smoothing state space models
- ARIMA models are stationary, exponential smoothing models are non-stationary and therefore maybe more applicable exponential smoothing state space models
- ARIMA models are stationary, exponential smoothing models are non-stationary and therefore maybe more applicable here. Especially, ARIMA models assume homoskedasticity. This might not be true as  $R_t$ -values wiggle around much more in the beginning. Not sure how important that is to the prediction.
- ARIMA forecast from forecast package has very wide predictive intervals and forecasts basically a straight line
- state space smoothing model is asymptotically a flat line, but seems more useful in short /midterm.

## 4. Literature

### 4.1 A simple approach to measure transmissibility and forecast incidence

<https://www.sciencedirect.com/science/article/pii/S1755436517300245>

#### 4.1.1 Aim

Forecasting of Ebola using  $R_t$

#### 4.1.2 Summary

The study does exactly the same thing we do, albeit probably less sophisticated. They take (simulated?) cases and a linelist, use the renewal equation to compute  $R_t$ -values using an averaging window of 2-4 weeks, then project these cases into the future using a branching process. I am not sure whether they just assumed the  $R_t$  to be constant in the future.

For evaluation, they calculated the number of true values outside the IQR (should be 50%) and used MSE.

#### 4.1.3 Take away

“Our approach worked best when near-future patterns of incidence were well described by an exponential trend.” Might be due to constant  $R_t$ ?

I have the feeling the paper is interesting for us as a proof of concept, but ultimately pretty useless...

### 4.2 Nonmechanistic forecasts of seasonal influenza with iterative one-week-ahead distributions

<http://www.stat.cmu.edu/~ryantibs/papers/deltadens.pdf>

#### 4.2.1 Aim

Improve Flu forecasting by using delta-densities and ensemble models.

#### 4.2.2 Summary

They use “delta densities”, which assumes an autoregressive dependency structure, but uses a kernel density estimation approach to model these dependencies rather than the common choice of linear relationships plus Gaussian noise → similar to method of analogues (haven’t researched it yet), also similar to Kernel Conditional Density Estimation forecasting. Also includes a lot of references / info on model stacking / averaging.

For evaluation, they use binned log scores (not sure why binned) and absolute errors.

#### 4.2.3 Take Away

Very interesting, sophisticated and complicated. Might be useful to look the ensemble things more closely. Also the delta density seems very interesting. One problem with the paper for us is that for flu, past data from previous years is available. So not sure how good we can use their methods, as we don’t have that.