FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
FACULTY OF MEDICINE

Forecasting based on surveillance data

Sebastian Meyer
Institute of Medical Informatics, Biometry, and Epidemiology
Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

GEOMED 2019, Glasgow, 27 August 2019

Based on joint work with Leonhard Held (University of Zurich)
and Junyi Lu (FAU Erlangen-Nürnberg)

Epidemics are hard to predict

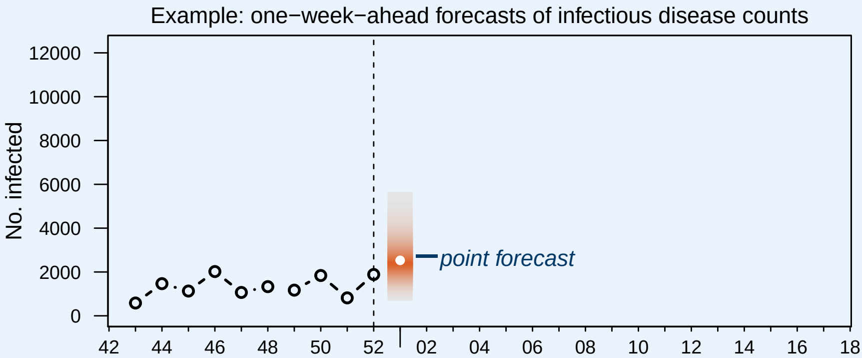
World Health Organization (2014)

> *Forecasting disease outbreaks is still in its infancy, however, unlike weather forecasting, where substantial progress has been made in recent years.*

Meanwhile ...

- **Epidemic Prediction Initiative** in the USA (https://predict.cdc.gov/): online platform to collect real-time forecasts from multiple research groups
- Integration of **social contact patterns** (Meyer & Held, 2017), **human mobility data** (Pei, Kandula, Yang, & Shaman, 2018), and **internet data** (Osthus, Daughton, & Priedhorsky, 2019)
- Adoption of **forecast assessment techniques** from weather forecasting

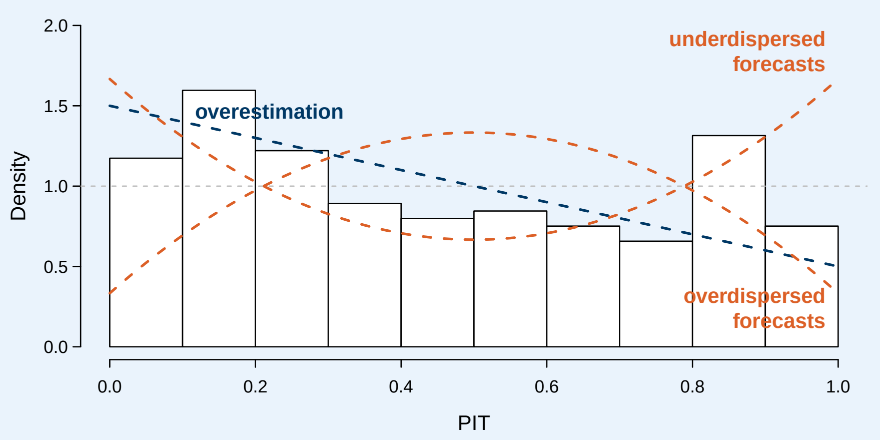
“Forecasts should be probabilistic” (Gneiting & Katzfuss, 2014)



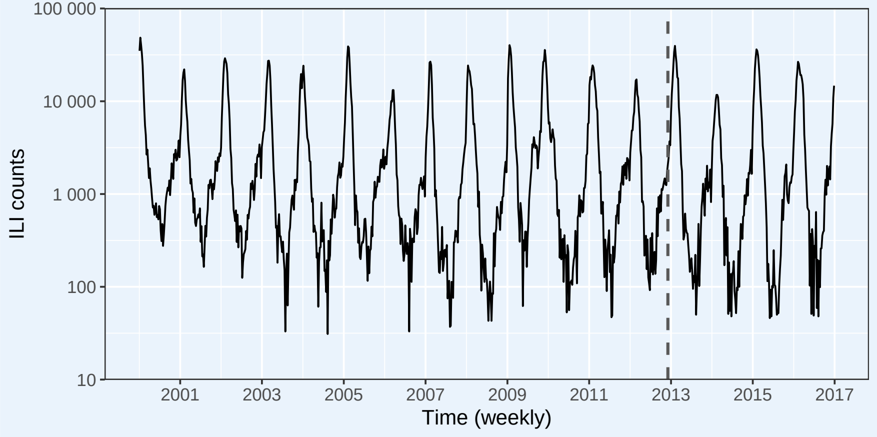
Proper scoring rules $S(F, y)$

- Quantify discrepancy between forecast F and observation y
- "Proper": forecasting with true distribution is optimal
- Most scoring rules are easy to compute:
  - **Squared error score**: $SES(F, y) = (y - \mu_F)^2$
  - **Logarithmic score**: $LS(F, y) = -\log f(y)$
  - **Dawid-Sebastiani score**: $DSS(F, y) = \log(\sigma_F^2) + \frac{(y-\mu_F)^2}{\sigma_F^2}$
- Scoring rules summarize two complementary measures of forecast quality:
  - **Sharpness**: width of prediction intervals (property of F)
  - **Calibration**: statistical consistency of forecast F and observation y

Histogram of $F(y) = $ PIT (probability integral transform) values



Case study I: Weekly ILI counts in Switzerland, 2000–2016



- Compute one-week-ahead forecasts in the test period (from December 2012)
- Compare average scores between different models

Useful *statistical models* to forecast epidemic spread

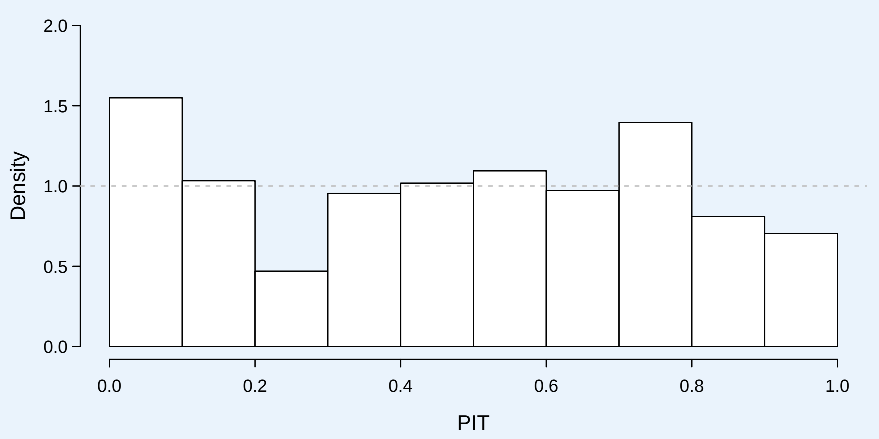
- Scope: well-documented open-source R implementations
- We compare five different models:
  - `forecast::auto.arima()` for log-counts $\longrightarrow$ ARMA(2,2)
  - `glarma::glarma()` $\longrightarrow$ NegBin-ARMA(4,4)
  - `surveillance::hhh4()`: "endemic-epidemic" NegBin model (lag 1)
  - Kernel conditional density estimation (`kcde`) by Ray et al. (2017)
  - `prophet::prophet()` for log-counts: harmonic regression with changepoints
- Naive historical reference forecast: log-normals by calendar week

Performance of 213 one-week-ahead forecasts

Method	RMSE	$\overline{\text{DSS}}$	$\overline{\text{LS}}$	runtime [s]
`arima`	2287	13.78	7.73	0.51
`glarma`	2450	13.59	7.71	1.49
`hhh4`	**1769**	**13.58**	**7.71**	**0.02**
`kcde`	1963	13.79	7.80	1128
`prophet`	5614	15.00	8.03	3.01
naive	5010	14.90	8.06	0.00

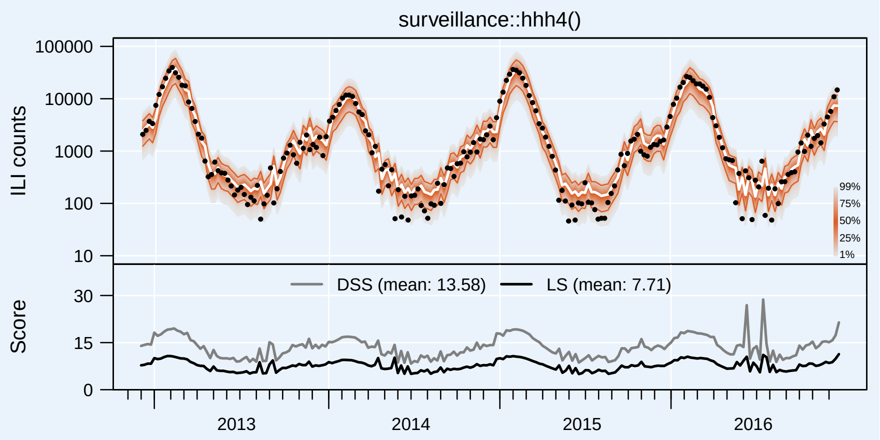
- Runtimes vary considerably (time for *single* refit and forecast)
- The two *autoregressive NegBin* models score best
- Non-dynamic methods: `prophet` does not outperform naive forecasts

PIT histogram for hhh4-based one-week-ahead forecasts



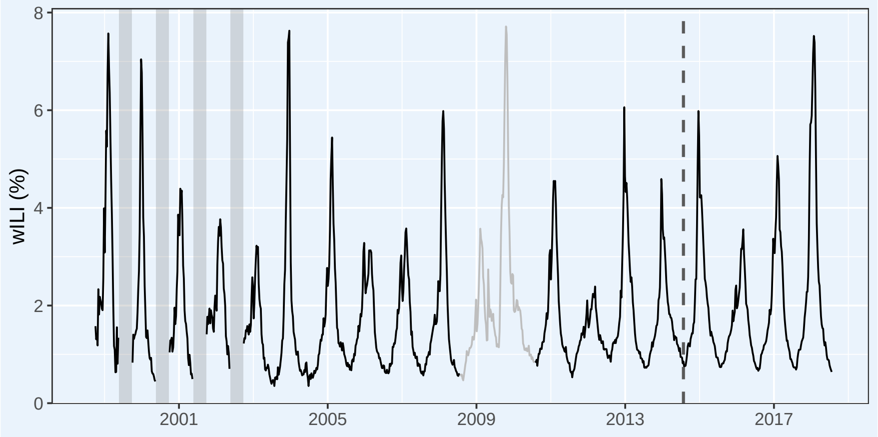
- Some evidence of miscalibration

hhh4-based one-week-ahead forecasts



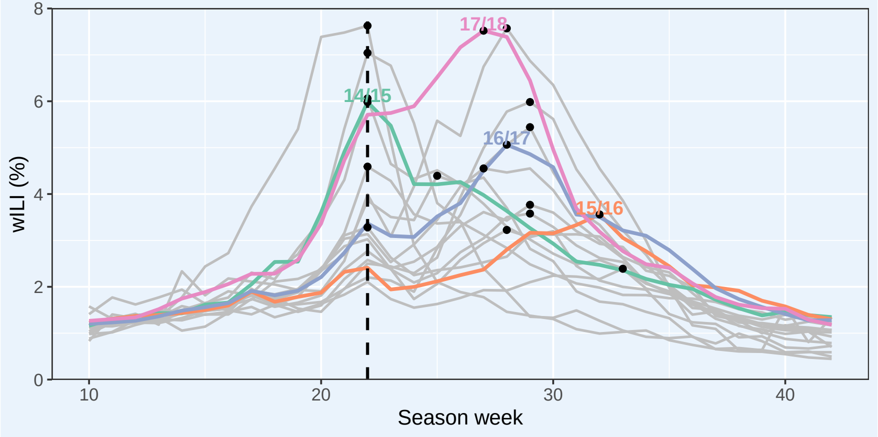
- Relatively sharp forecasts → penalty in wiggly off-season 2016
- Off-season counts tend to be lower than predicted

Case study II: Weekly ILI activity in the USA, 1998–2018



- Inspired by CDC's *FluSight* competition (https://predict.cdc.gov/)
- Forecast ILI proportion 1 to 4 weeks ahead, plus peak week & proportion

Seasonal epidemic curves



- (Intermediate) peak at the end of the year (dashed line)
- Test seasons with late peak (15/16) and high intensity (17/18)

Forecasting machinery

- Gaussian models of logit-transformed proportions:

  [S]ARIMA, **Prophet**, **naive historical**

- Kernel conditional density estimation (**KCDE**)

- hhh4 not applicable for proportions

  $\rightarrow$ Idea: “Endemic-epidemic” beta regression (**Beta(*p*)**), via `betareg`:

$$X_t | \mathscr{F}_{t-1} \sim \mathrm{Beta}(\mu_t, \phi_t)$$

$$\mathrm{logit}(\mu_t) = \nu_t + \sum_{k=1}^{p} \beta_k \, \mathrm{logit}(X_{t-k})$$

$$\nu_t = \alpha^{(\nu)} + \beta^{(\nu)^T} \mathbf{z}_t^{(\nu)}$$

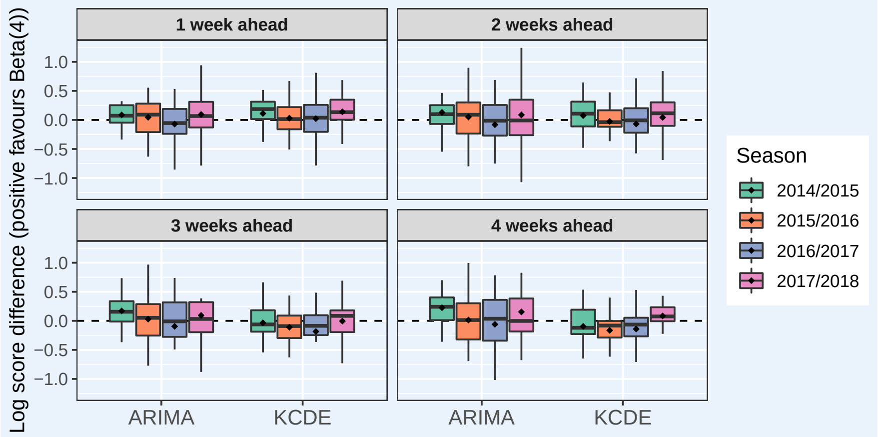
$$\log(\phi_t) = \alpha^{(\phi)} + \beta^{(\phi)^T} \mathbf{z}_t^{(\phi)}$$

Overall performance of short-term forecasts (all horizons)

Method	$\overline{\text{DSS}}$	$\overline{\text{LS}}$	max(LS)	runtime [min]	#par
ARIMA(5,1,0)	–1.81	–0.02	5.24	6.2	16
SARIMA(1,0,0)(1,1,0)[52]	–1.69	0.04	4.92	110.4	**3**
Beta(1)	–2.02	–0.11	5.59	2.9	19
Beta(4)	–2.07	**–0.12**	4.34	**2.6**	20
KCDE	**–2.29**	–0.12	**4.08**	266.6	28
Prophet	–0.75	0.48	5.04	11.8	50
Naive	–1.13	0.42	5.29	0.1	106

- Runtimes vary considerably (total time for [re]fitting and forecasting)
- Higher order lags improve Beta forecasts
- Worst case prediction is less worse with KCDE than with Beta(4)
- Non-dynamic methods: `prophet` does not outperform naive forecasts

Relative performance wrt Beta(4), by season and horizon



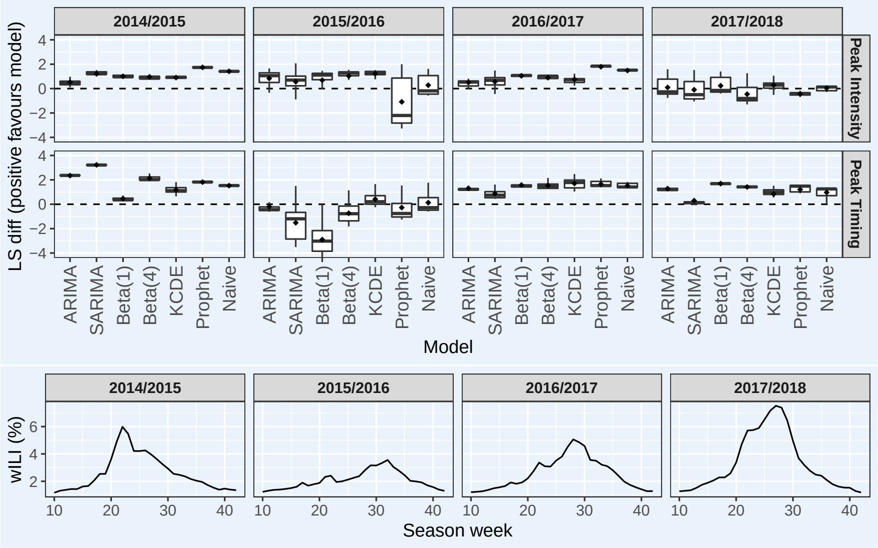
- No model consistently outperforms another, and rankings vary by season
- KCDE tends to produce better 3- and 4-week-ahead forecasts

Overall performance of peak forecasts

Method	Timing ($\overline{\text{LS}}$)	Intensity ($\overline{\text{LS}}$)
ARIMA(5,1,0)	1.44	1.59
SARIMA(1,0,0)(1,1,0)[52]	1.78	1.57
Beta(1)	1.99	1.46
Beta(4)	1.47	1.51
KCDE	**1.43**	**1.41**
Prophet	1.44	1.68
Naive	1.46	1.46
Equal bin (uniform)	3.50	3.30

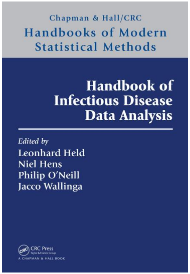
- KCDE has best peak forecasts overall
- Naive historical forecasts are not that bad either

Relative performance wrt equal-bin forecast, by season



Discussion

- Endemic-epidemic approach useful for short-term forecasts:
  fast, performant, and easy to implement
- Peak prediction is hard: no model outperformed naive historical forecasts in all seasons (KCDE did the best job)
- Any missing competitive forecasting method with a well-documented implementation in open-source software?
- Ensemble forecasts (Reich et al., 2019)
- Underreporting and reporting delays
- Multivariate forecasting by region or age group



References

Gneiting, T., & Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application*, *1*(1), 125–151. https://doi.org/10.1146/annurev-statistics-062713-085831

Held, L., & Meyer, S. (2019). Forecasting based on surveillance data. In L. Held, N. Hens, P. D. O'Neill, & J. Wallinga (Eds.), *Handbook of infectious disease data analysis*. Chapman & Hall/CRC.

Meyer, S., & Held, L. (2017). Incorporating social contact data in spatio-temporal models for infectious disease spread. *Biostatistics*, *18*(2), 338–351. https://doi.org/10.1093/biostatistics/kxw051

Osthus, D., Daughton, A. R., & Priedhorsky, R. (2019). Even a good influenza forecasting model can benefit from internet-based nowcasts, but those benefits are limited. *PLOS Computational Biology*, *15*(2), 1–19. https://doi.org/10.1371/journal.pcbi.1006599

Pei, S., Kandula, S., Yang, W., & Shaman, J. (2018). Forecasting the spatial transmission of influenza in the United States. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(11), 2752–2757. https://doi.org/10.1073/pnas.1708856115

Ray, E. L., Sakrejda, K., Lauer, S. A., Johansson, M. A., & Reich, N. G. (2017). Infectious disease prediction with kernel conditional density estimation. *Statistics in Medicine*, *36*(30), 4908–4929. https://doi.org/10.1002/sim.7488

Reich, N. G., Brooks, L. C., Fox, S. J., Kandula, S., McGowan, C. J., Moore, E., . . . Shaman, J. (2019). A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the united states. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1812594116

Data and reproduction code for case study I:
https://HIDDA.github.io/forecasting/