

COS-D407. Scientific Modeling and Model Validation

Lecturer: Christina Bohk-Ewald

Week 3

University of Helsinki, Finland
01.11.2021–15.12.2021

Third week's class:

Scientific modeling and model validation in practice

- Q & A: recap of material of previous session
- Present your findings of previous lab session
- Brief introduction to assessing methods in general
- Guest lecturer Ricarda Duerst:
 - ▶ Introduction to forecasting human mortality & evaluating a method's suitability to do so
 - ▶ Introduction to basic concept of method validation in the context of demographic forecasting

Third week's class in the lab:
Ricarda will tell you later about this :-)

→ Present and discuss your findings in class at the beginning of the next session on Monday.

Brief Q&A: recap material of previous session:

- What are the main steps of the scientific method?
- Why can the scientific method or process be at least as important as the scientific finding?
- Why is it so important that an explanation for a phenomenon is testable or falsifiable?
- Why is it important to re-test a model with high predictive power for a phenomenon?
- What are common pitfalls during scientific work?

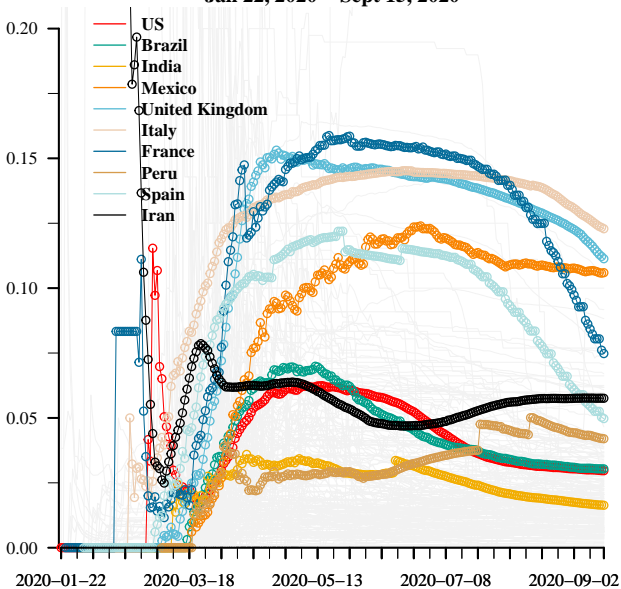
→ Open questions?

Present your findings of previous lab session:

- What are recent trends in the case fatality rate (CFR) of COVID-19 in the ten countries with most COVID-19 deaths as of last week?
- What are possible explanations for cross-country differences in CFR?

→ Open issues?

Case fatality rate for ten countries with most COVID-19 deaths Jan 22, 2020 – Sept 15, 2020



Data source: JHU CSSE: <https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases>

Cross-country variation in CFR may be due to

- Real differences in mortality attributable to COVID-19
- Age structure of population and, consequently, of deaths and cases attributable to COVID-19
- Stage of progress of COVID-19 outbreak in each country
- Classification of COVID-19 deaths
- Testing practices wrt to test coverage and test specificity
- Capacity and occupancy rate of health-care systems for intensive care
- Effectiveness of control measures to curb severe COVID-19 infections
- ...

⇒ Confirmed cases and reported deaths may be biased

Confirmed cases may underestimate the number of infections

- Cases with mild symptoms or asymptomatic cases might go undetected
- Test coverage
 - ▶ Test kits may not be available in large numbers
 - ▶ Focus on sub-populations, e.g., cases with proven contact to other COVID-19 cases or hospitalized cases
- False negatives
 - ▶ People get tested after the first week of infection, when it is likely that SARS-CoV-2 cannot be detected in pharynx anymore (PCR)
 - ▶ Test for antibodies could be carried out before a body has had a chance to produce them
- ...

Reported deaths may be biased

- Reporting delays may amount to several days.
- Inconsistent practices for classifying COVID-19 deaths within and between countries. For example, only deceased individuals who (1) were hospitalized or (2) died from COVID-19 as primary and / or secondary cause of death may be counted.
- Test coverage and test specificity may be insufficient. For example, not all deaths are tested for COVID-19. Persons dying at home or in other institutions may not be counted.

Model validation

In the next three weeks, you will be introduced to two basic ways for evaluating a model and the validity of its output.

- 1 Comparing the model output to its *true realizations* using, e.g., ex post errors and cross validation. Week 3.
- 2 Analyzing the *process* that a model uses to generate its output and, if possible, analyzing the *sensitivity of the results* with respect to, e.g., using input data from different sources. This can also be done if true realizations are unknown. Weeks 4 & 5.

Input parameters \Rightarrow Model \Rightarrow Outcome variables

Model validation in the context of demographic forecasting

Guest lecturer Ricarda Duerst:

- Introduction to forecasting human mortality & evaluating a method's suitability to do so
- Introduction to basic concept of method validation in the context of demographic forecasting

Course learning materials

Course learning materials on GitHub:

<https://github.com/christina-bohk-ewald/2021-COS-D407-scientific-modeling-and-model-validation>

Recommended learning material for today's class

- **Bohk-Ewald et al. (2018)**

Forecast accuracy hardly improves with method complexity when completing cohort fertility. PNAS 115(37), 9187–9192.

DOI: <https://doi.org/10.1073/pnas.1722364115>

Thank you for your attention!

`christina.bohk-ewald@helsinki.fi`