Narrative review of infectious disease spread models developed in Poland during COVID-19 pandemic

Andrzej Jarynowski, IBI Wrocław/FU Berlin; Vitaly Belik, FU Berlin

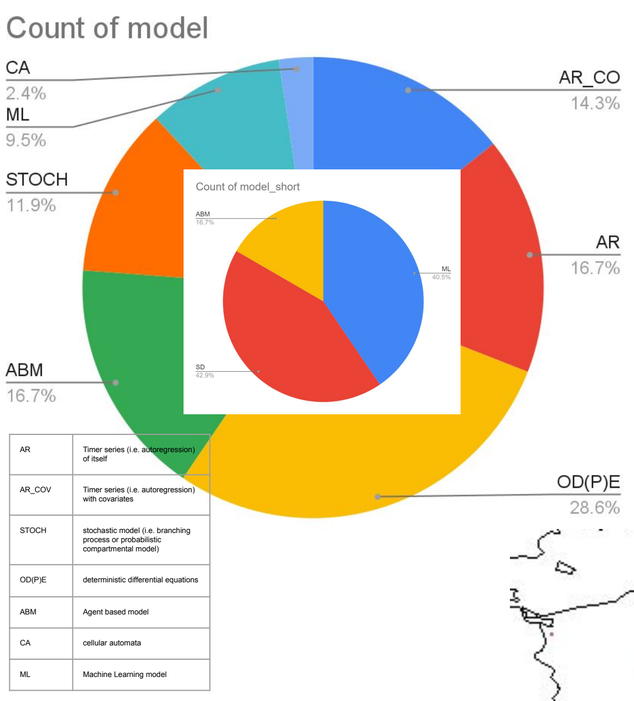
Introduction

Models are elementary tools for timely and targeted responses in the fight against pathogens. The use of Data-based modelling to generate forecast of the epidemic dynamics by harnessing vast, register based and open-source data have potential to be game changing and highly sustainable (i.e. in optimization of the resources [doi:10.15503/emet2020.100.122]). COVID-19 pandemic sped up infection disease spreading model's development and applicability. However, models' predictions for SARS-CoV-2 in Poland and in general in Eastern Europe were often disappointing [doi:10.15503/emet.2021.112.124]. On the other hand, due to the extreme amount of work to gather, analyse and understand data during the time of pandemic with lessons learned in Polish modelling community, for future epidemic, modelling could provide the scientific basis for public health decisions and intervention measures.

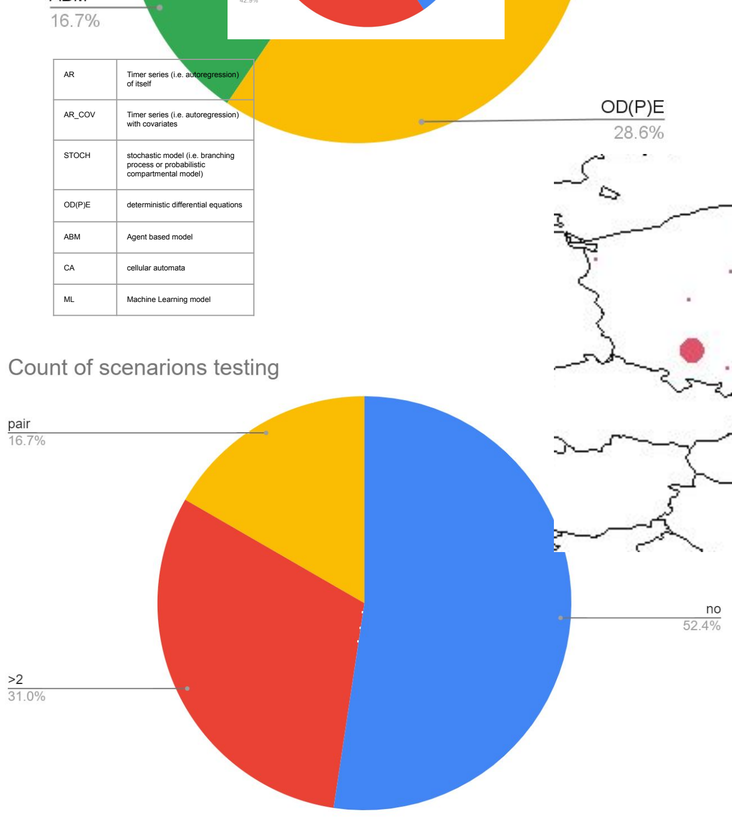
Data and Methods

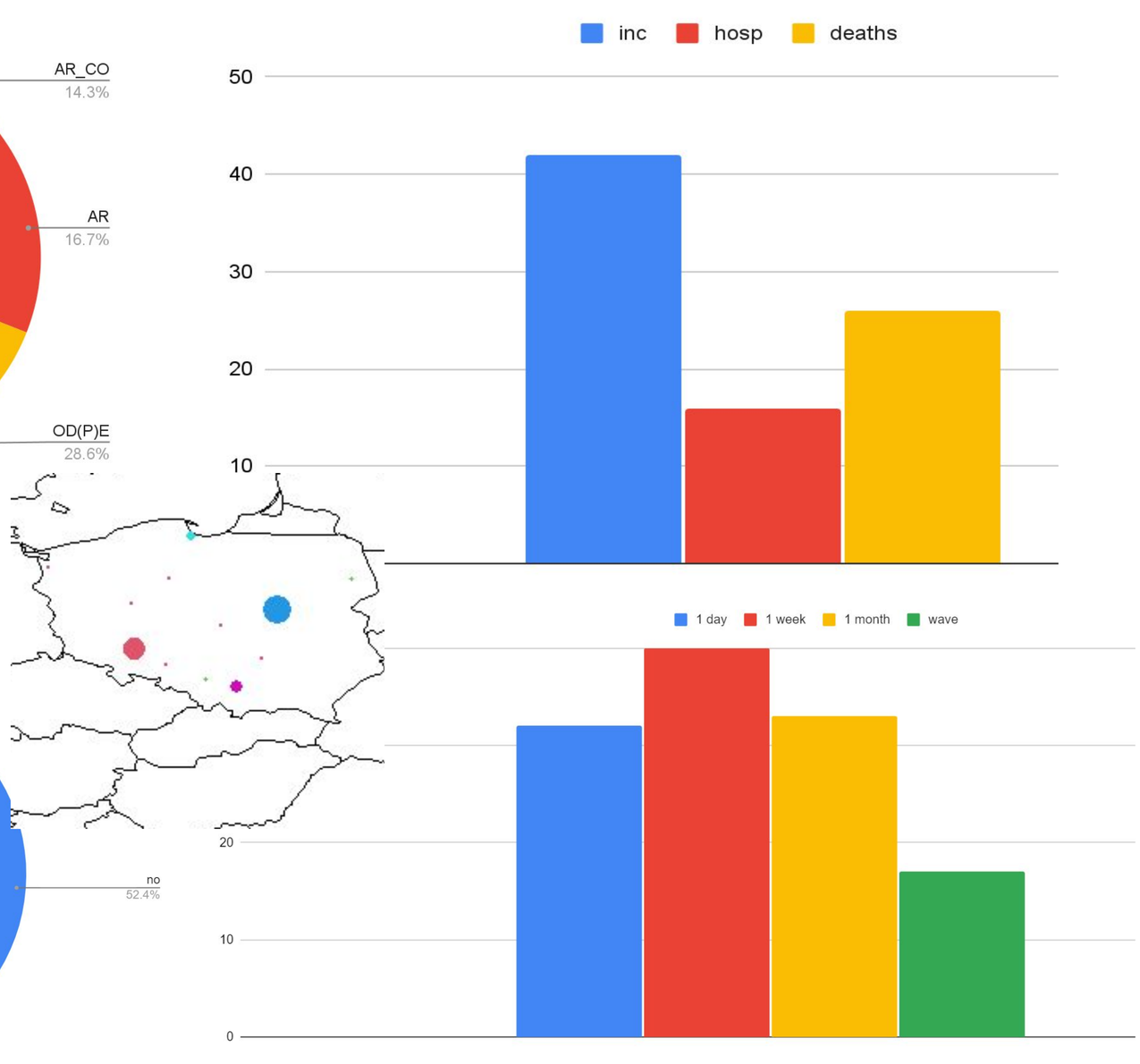
We have performed qualitative and quantitative benchmark of COVID-19 forecasting models for Poland till Omicron wave. We divided models into three categories: 1) system dynamics (i.e. differential equations, stochastic process); 2) agent-based models (i.e. microsimulation); machine learning & time series (i.e. autoregressive models, neural networks). We proposed three categories of forecasting horizons: 1) short term prediction ~1 or 7 days; 2) medium ~ 4 weeks; 3) long ~ full infection season (from a few months up to one year). The selection of the models in qualitative part [doi:10.5604/01.3001.0015.0281] was based on availability principle (both scientific 'white' and unpublished 'grey' sources) and models were included if they provide forecast in any time horizon for at least one outcome for any kind of geographical unit of Poland (i.e. whole country, voivodeship, city): No. cases (any definition), No. hospitalisation, No. COVID-19 related deaths (any definition). For the quantitative part we took data from the European Covid-19 Forecast Hub and compared outcomes for short and medium term forecasts for registered COVID-19 cases.

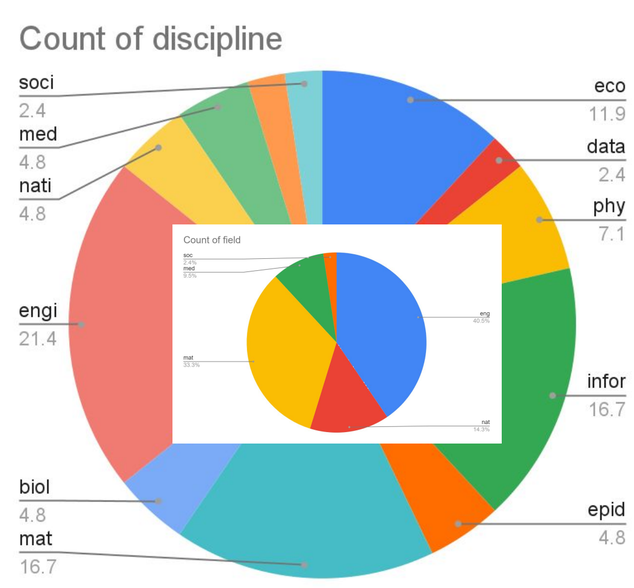
Qualitative analysis

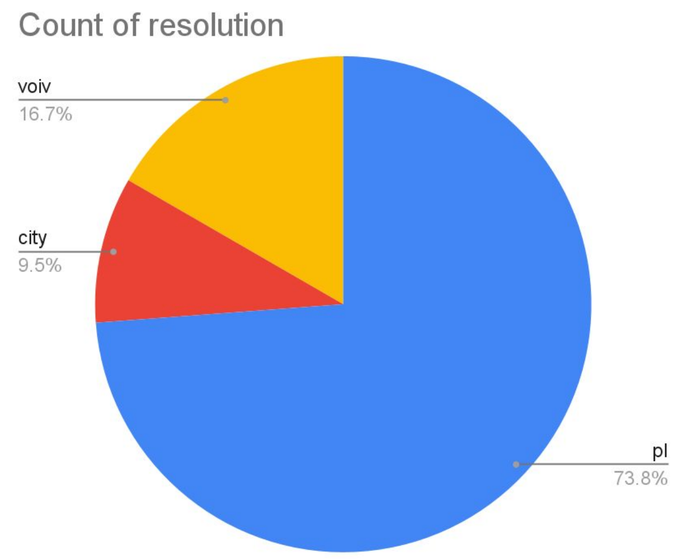


AR	Timer series (i.e. autoregression) of itself
AR_COV	Timer series (i.e. autoregression) with covariates
STOCH	stochastic model (i.e. branching process or probabilistic compartmental model)
OD(P)E	deterministic differential equations
ABM	Agent based model
CA	cellular automata
ML	Machine Learning model



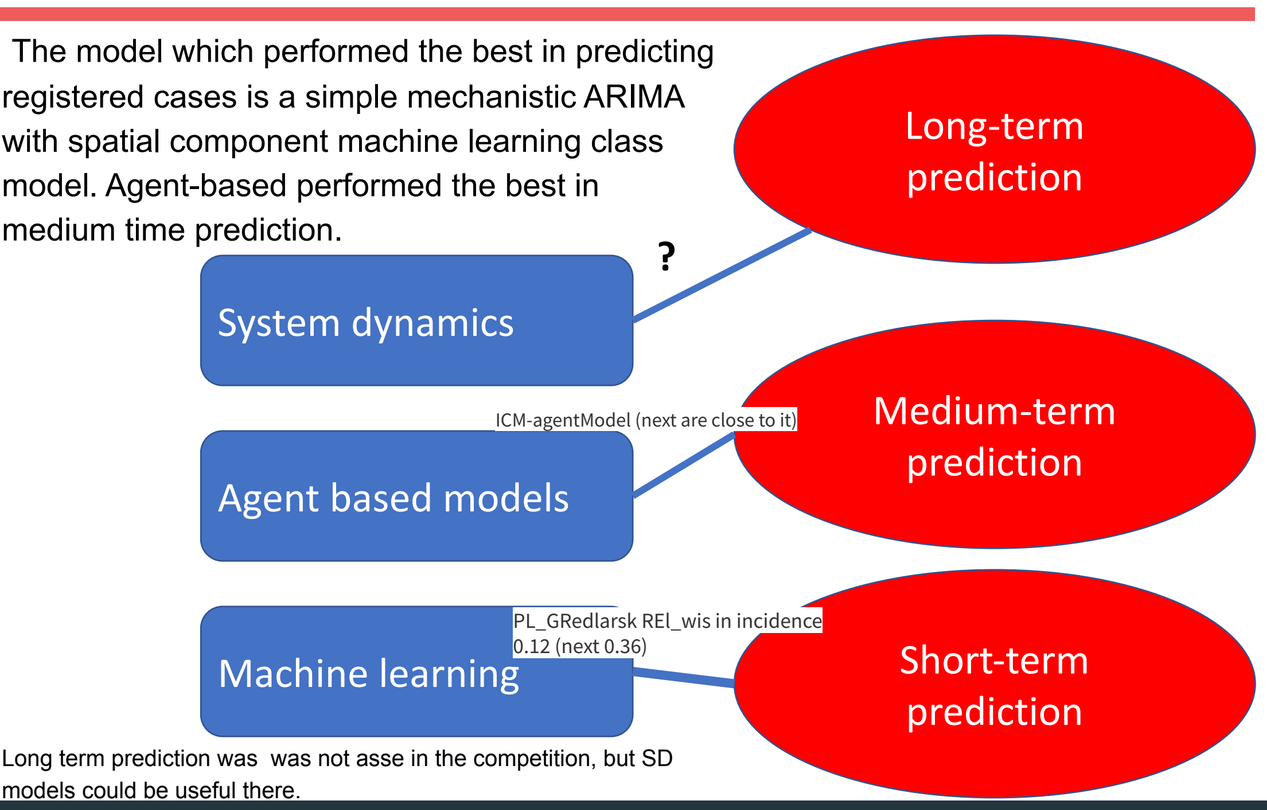






Quantitative comparison

The model which performed the best in predicting registered cases is a simple mechanistic ARIMA with spatial component machine learning class model. Agent-based performed the best in medium time prediction.



Long term prediction was was not asse in the competition, but SD models could be useful there.

Lessons for future?

Due to COVID-19 infectious disease modelling concept has "spread" over the whole Poland (both academic and business entities). As there is no model fit all, both forecasting hub exercise and narrative review of the models allow us to see pros and cons for each approach.

Take home messages:

- Polish System dynamics models are far behind other models (which is not the case i.e. in UK/USA)
- ML models overfit to reporting system, but they are outperforming all other models in fitting to all "official" data (mainly incidence) in short term
- ABM models are as good as ML in medium-term prediction of incidence, but the best in hospitalization and deaths.
- Comparing results of the forecasting performance may be misleading due to dark-figure phenomenon.

https://docs.google.com/spreadsheets/d/1125HWV1Fh38XcQ8epSW0S07rIf6KpN6wiEzFgIu5IA4/edit?usp=sharing