MODEL CARD: OCHA-Bucky COVID-19 Model

The Centre for Humanitarian Data's Peer Review Framework for Predictive Analytics in Humanitarian Response aims to create standards and processes for the use of models in our sector. The Framework consists of five steps: model submission, technical review, implementation plan submission, ethical review, and final report. The output of the model submission is a completed Model Card that includes information about the intended use, model development, model evaluation, and operational readiness. For further information, please contact Leonardo Milano: leonardo.milano@un.org.

OCHA-Bucky COVID-19 Model - October 2020

Model Summary

- Organization developing the model: OCHA Centre for Humanitarian Data in collaboration with the Johns Hopkins University Applied Physics Laboratory (APL).
- Model date: Official release on 28 October 2020
- Model version: V1 - October 2020
- Model type: Spatially distributed SEIR model
- License: MIT License
- Main contacts: Leonardo Milano, Predictive Analytics team lead, UN OCHA Centre for Humanitarian Data leonardo.milano.un.org
- Link): GitHub Repository and Methodology paper

1. Intended Use

A. In-scope use cases: what is the actual and potential scope of the model? Describe the situations in which the model output is expected to be reliable.

OCHA-Bucky model is a COVID-19 metapopulation compartment model initially designed to estimate medium-term (on the order of weeks) case incidence and healthcare usage at the second administrative (admin-2, ADM2) level. These ADM2 regions are all coupled using mobility information to approximate the inter- and intra-regional contacts between the members of the populations. The model is therefore expected to be reliable for short to medium term projections in all situations in which data is available to tune all the model's parameters.

B. Out-of-scope use cases: what are the model's limits and constraints?

The OCHA-Bucky methodology is not expected to be applicable for estimating:

- Spread and patterns at a very granular level (city, neighborhood, refugee camp, etc.). Other models such as Agent-Based Models are expected to be more reliable in these situations.
- Non-human-to-human communicable diseases. The main transmission route simulated by OCHA-Bucky is based on contacts between people.

MODEL CARD: OCHA-Bucky COVID-19 Model

C. Describe the situations in which the model output may not be reliable?

The main know limitations of the model are linked to:

- The lack of data to tune the model parameters. In absence of such datasets, the model could be tuned based on assumptions but the projections will depend on them and should be interpreted accordingly. The main inputs to OCHA-Bucky are:
    - Subnational historical COVID-19 cases and deaths data to tune the reproduction rate
    - Subnational age-disaggregated population data
    - Data on non-pharmaceutical interventions currently in place or planned, as it impacts the projected reproduction rate
    - Contact matrices estimating the social mixing of different age groups
    - Location-specific vulnerability factors which may impact the severity of symptoms or the mortality.
- The time horizon of the projections. Given the very dynamic nature of the situation, the model assumptions change over time. The model has been used so far to produce four week projections at the national level and two week projections at the subnational level. OCHA-Bucky could be used to explore longer term trends but the results should be interpreted as a scenario exploration/comparison tool and not as a quantitative model.

D. Model interpretation: what does the output represent?

Model output consists of future projections of geographically distributed COVID-19 cases and deaths, as well as severe cases (defined as a proportion of total cases). Moreover, the model projects the impact of implemented Non-Pharmaceutical Interventions (NPIs).

2. Model Development

A. Details of the datasets used to build the model.
- Describe the sources of data, size, and scope of the datasets.

For a detailed description of the input datasets and their use in OCHA-Bucky please refer to section 2.3 in the OCHA-Bucky methodology paper.
These are the datasets needed to run OCHA-Bucky's projections:

- Population Data: Administrative boundaries shapefiles at the ADM2 level from the Humanitarian Data Exchange (HDX) and the United Nations-adjusted WorldPop age and sex-disaggregated population raster data.
- Vulnerability Parameters: the fraction of the population in an ADM2 region that is particularly vulnerable to COVID-19 is estimated by calculating the proportion of the population affected

by the following factors:

- Food insecurity: Food insecurity data is obtained from the Integrated Food Security Phase Classification (IPC) Global Platform
- Indoor air pollution - indoor cooking fuels. Data on the use of solid fuels is obtained from the World Health Organization.

- Mobility Data: Road density data from Humanitarian OpenStreetMap (HOTOSM) is used to estimate the strength of the connection between all ADM2 regions considered.
- COVID-19 Cases and Deaths: In most countries of interest, the source for COVID case counts is the nation's Ministry of Public Health. Additional federal (ADM0) historical data is also obtained from the WHO.
- Contact Matrix: Contact matrices quantify how much people from different age groups interact, and are extracted from Prem et al.
- Non-pharmaceutical Interventions (NPIs): For each country, an initial list of NPIs was obtained from the ACAPS COVID-19 Government Measures Dataset.

- Is it representative of the population being sampled?

For the population figures and vulnerability estimates, we largely rely on data collected by partner organizations directly in the countries that we are modeling. The population data obtained from World Pop is compared and adjusted to match the population figures used for Humanitarian Response Planning. Vulnerability parameters may not be up-to-date. For each indicator, we look for the most recent and reliable data.

- How accurate or reliable is the training data?

One of the main inputs to the model is the historical cumulative number of COVID-19 cases and deaths. This dataset is used to tune the parameters of the model related to disease transmission. OCHA-Bucky partially corrects for the known gaps and reporting delays in the COVID-19 data reported by the authorities in the following way:

- OCHA-Bucky uses recent data to estimate the fraction of cases expected to be reported and corrects the projections to account for large underreporting observed in these countries.
- Rather than explicitly including a parameter controlling the transmission rate, Bucky uses the recent historical data to estimate transmission at the local level.

- How is missing data treated? (e.g., exclusion, single imputation, multiple imputation)

Missing data is generally related to the number of cases and deaths at the subnational level. We deal with it in the following ways:

- Parameters extracted from national level data (mainly the Doubling Time and the Case Reporting Rate) are applied at the ADM2 level when the information for calculating them at the local level is missing or insufficient.
- We apply a rolling average, typically with a 14-day rolling window, to smooth out fluctuations in the input datasets.

B. What are the model assumptions and approximations?

Like most of the SEIR models, OCHA-Bucky assumes that:

- Disease transmission is happening exclusively via human-to-human contacts
- People that recover from the disease will be immune and therefore cannot be reinfected
- As we are focusing on short to medium term projections we assume birth rates to be negligible
- The severity of severe symptoms and covid related deaths are assumed to be dependent on people's age and vulnerability factors. The impact of COVID-19 on access to healthcare facilities is not simulated in OCHA Bucky.
- Death rates from causes other than COVID-19 are also assumed to be negligible within the timeframe considered in the simulations.

C. Methodology - provide a description of the different analysis steps and how the input datasets are used to train the model.

The OCHA-Bucky model consists of a series of adjustments to an existing COVID-19 model (JHUAPL-Bucky) that was developed as part of the Centers for Disease Control and Prevention (CDC) COVID-19 Mathematical Modeling Forecasting Ensemble.
OCHA-Bucky is a collection of coupled and stratified SEIR models. Since COVID-19 exhibits heavily age-dependent properties, wherein a majority of severe cases are in older individuals, SEIR models are stratified via the age demographic structure of a geographic region to obtain accurate estimates of case severity and deaths. Additionally, to model the spatial dynamics of COVID spread, we consider a set of SEIR sub-models at the smallest geographic level for which we have appropriate data. For a full description of the methodology please refer to the OCHA-Bucky methodology paper.

3. Model Evaluation

A. Are there other similar existing models and how does this model compare?

OCHA-Bucky is a subnational implementation of a broad class of SEIR models that have been used to project the impact of COVID-19. In the United States context, the parent model (JHUAPL-Bucky) is included in the Centers for Disease Control and Prevention (CDC) COVID-19 Mathematical Modeling

Forecasting Ensemble.

The figure below highlights some of the features of OCHA-Bucky which makes it more adaptable to humanitarian contexts.

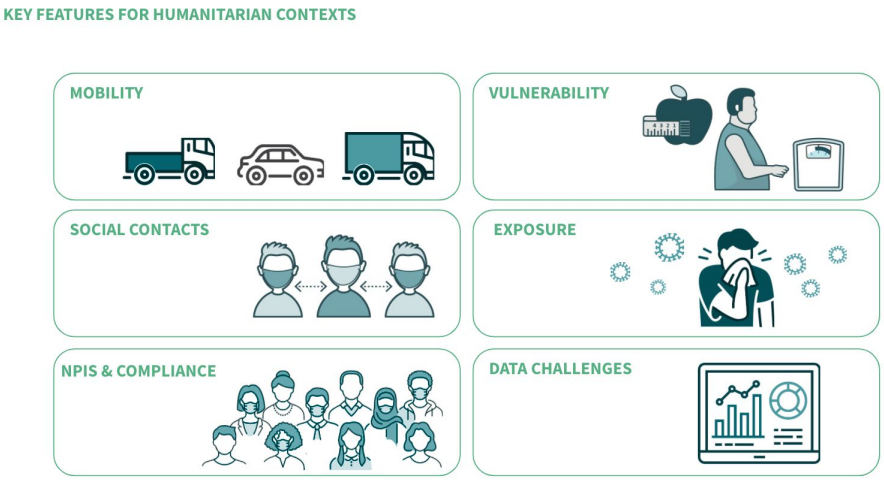

Figure 1: features of OCHA-Bucky making it adaptable to humanitarian contexts.

B. Have there been any past peer reviews, such as published papers or other review exercises?

The model hasn't been published in a peer-reviewed journal yet.

C. Is there any reference or benchmark used to evaluate the performance?

The Bucky model has been validated using U.S. case data. The validation has shown good agreement on daily figures with no significant model bias. Some noise on daily case & death counts, especially at small numbers is observed.
The OCHA team is currently working on the validation of the OCHA-Bucky in Afghanistan and potentially other humanitarian contexts.

D. What are the metrics used for model evaluation? Why have these metrics been selected? Describe current model performance.

The model has been evaluated by looking at residuals between projected and actual cases on historical model runs. The figure below shows the validation of the JHUAPL-Bucky model in the U.S. Most of the data points lie within the estimated model uncertainty band and no significant bias is observed.

MODEL CARD: OCHA-Bucky COVID-19 Model

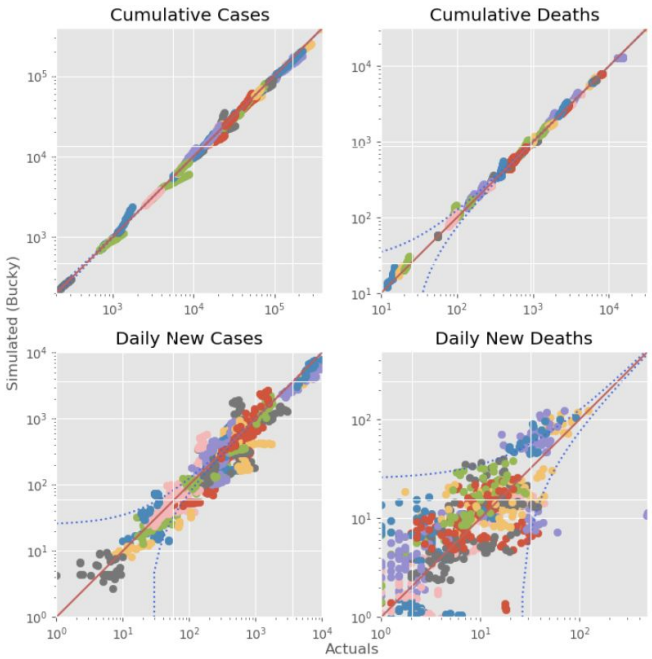

Figure 2: validation of the JHUAPL-Bucky model in the U.S. Most of the data points lie within the estimated model uncertainty band and no significant bias is observed.

E. How does performance depend on forecast lead time?

The reliability of the projections decreases with lead time. The main factor having an impact on this is not a limitation of the methodology used but rather the dynamic nature of the disease. At the beginning of the outbreak, COVID-19 was spreading in an uncontrolled way in many countries resulting in a high reproduction rate (a measure of how fast cases are growing). As governments implemented mitigation measures to slow the spread of the disease, the reproduction rate dropped. Long term projections couldn't capture this evolution. We suggest using OCHA-Bucky for projecting cases and deaths up to four weeks out. The model can however be used as a scenario exploration tool (i.e. not as a quantitative model) for longer lead times.

F. What is the risk tolerance of the model? What happens if the model produces false positives, false negatives?

Inaccuracies of the model are limited in the following ways:

- Projections are limited to four weeks at the national level and to two weeks at the subnational level.
- Every time a new model run is performed, the disease transmission parameters are recalibrated from recent data.

- Uncertainties are calculated by varying most of the model parameters within uncertainties (Monte Carlo simulation).
- Projections are always presented as a range which covers the projection uncertainties.

4. Operational Readiness

A. Is the model ready to be used to inform humanitarian response?
-Is the model kept up-to-date with the latest datasets?

Yes, the Centre for Humanitarian Data is updating the input datasets and the projections on a regular basis. Based on the projections reports are produced and shared with the OCHA country offices and other stakeholders. All the reports can be accessed here.

- Who is responsible for model updating and/or recalibration?

The Predictive Analytics team of the OCHA Centre for Humanitarian Data.

- Is the model ready to be deployed? If not, what are the additional steps needed (e.g. further research, validation, updates)?

The model has been already deployed and used at different levels to better inform COVID-19 response strategies.

B. Has the model been developed in collaboration with operational partners? Has the model previously been used in humanitarian situations?

OCHA-Bucky is the first model fully developed by the Centre. The Centre acted as an intermediary between OCHA field offices and the JHU team of technical experts. The model projections are available for six countries: Afghanistan, the Democratic Republic of Congo (DRC), Iraq, Somalia, South Sudan and Sudan. Since July 2020, we have worked with OCHA colleagues and partners in these locations to finetune the model with country-specific data. Currently, the reports are shared with the field offices in these six countries.

C. What happens if the model is inaccurate, produces false positives or false negatives? Describe expected impact according to the in-scope use cases highlighted in Section 1A above.

The main risks for OCHA-Bucky are related to the underestimation and the overestimation of COVID-19 cases. Underestimation of cases could potentially lead to a response that is not adequate

to the actual impact of the outbreak. Overestimation of COVID-19 impact may lead to ill-informed allocation of resources and prioritisation. Some of these biases are due to the poor quality of the input data. These risks are mitigated in the following ways: i) by reducing the lead time of the projections and by updating them as frequently as possible; ii) by presenting the results together with a clear description of the assumptions and the corresponding uncertainties and iii) by inviting partners not to rely solely on the model projections but use it as additional evidence to support their decisions.

D. List additional considerations for the use of the model in humanitarian response.

OCHA-Bucky has been initially developed to inform COVID-19 response but it can be potentially adapted to other human-to-human infectious diseases (e.g. influenza, measles, polio, ebola). The team is currently working on extending the model to include vaccination scenarios in the projections.
OCHA-Bucky's code and documentation are publicly available and interested partners can use it to run their own simulations. For more information and support please contact the Centre for Humanitarian Data centrehumdata@un.org.