

Macromodel

Poster Title: An Open-Source Probabilistic Modeling Framework for Personal Pandemic Risk Assessment

By Quantum Risk Analytics, Inc.: <https://pandemonium.dev/>

Our macromodel is the main “workhorse” within our modeling framework. It is built on the SEIRD compartmental model, but with extensions to permit heterogeneity of people within a geographic region and to take advantage of probabilistic programming. Our model is designed primarily to simulate risk to an individual user of our app. However, to do so skillfully, it must also model groups of people with whom the individual may interact and, in turn, other groups who are connected to those groups, etc. Nevertheless, our model is flexible enough to also model specific individuals simply by treating them as populations of size one, which is sufficient to evaluate individual risk and the possible results of different interventions for a specific group of identified individuals.

In our model, geography and demographic strata are treated as essentially the same for most cases: as categories that group or separate otherwise nondescript people *vis-a-vis* belonging or not belonging. Therefore, each populated “node” in the model is characterized distinctly by those primary factors: geographic location, age, sex, and race/ethnicity. When geographical COVID-19 data is lacking stratification according to a demographic category, we do not subdivide the node by that category.

Geographic-demographic nodes are connected or “coupled” through different kinds of “coupling factors”. Each coupling factor between two nodes ranges from 0 (no direct coupling) to 1 (fully coupled). Demographic strata in widely-separated geographical locations are nevertheless assumed to have similar disease-affecting behaviors. As such, we optionally partially couple their basic reproductive rate numbers and mortality rates. Regular patterns of movement between geographical regions (e.g., commuting for work) are parameterized through static coupling factors, which can be set by the user or from provided mobility data. Although the coupling factors associated with the aforementioned couplings are fixed or “static” through time, we also have “dynamic” coupling factors to represent the time-varying, special movements of a person or a group of people.

A noteworthy feature of our macromodel is its capacity to model individuals and their interaction with others. When an individual app user provides specifics about their congregations with others indoors (indoor events), we calculate their event risk using an indoor airborne transmission model. This risk, in turn, is mixed in with a risk derived from our macromodel, based on what fraction of time the app user is in the inputted event(s). Apart from these user-inputted events, an individual is modeled similarly to a collection of people, essentially by cloning the individual many times in order to handle fractional probabilities that he/she is infected, and by making necessary adjustments for that duplication when modeling the individual’s interactions.¹

Our probabilistic programming approach allows the SEIRD model parameters, their drift through time, and their respective probability distributions to be inferred from data (deaths, cases or wastewater-based modeled cases). In particular, for a given posited Susceptible-to-Exposed parameter value, there is transmission uncertainty due to the stochastic boolean nature of individual-to-individual potential transmission events and their repetition, which we capture using a binomial distribution. We analogously specify and handle unvarying SEIRD hyperparameters for

¹ This is done for computation efficiency and is not adequate to maintain the full stochastic nature of individuals when considering transmission to others. We will handle multiple individuals with other techniques (*i.e.*, importance sampling and potentially subsampling).

incubation, recovery and mortality periods.

Finally, within a given node, vaccination and behavioral and disease condition risk factor prevalence data can also be incorporated and applied if available.

Below is a comparison of the capacities of our macromodel versus those of other popular stochastic modeling software:

Feature	Pandemonium	Starsim	EMOD
Simulation type	Population/Group/Individual modeling as desired, including simulating at multiple scales within one session.	Agent-based	Agent-based ²
Availability	Python library (soon), web app (in user testing)	Python or R library, web app	C++ library, python library, standalone app
Language	Python	Python with Numba JIT	C++
Handling of stochastic variance	Sampling is handled automatically by Pyro framework	One sample per simulation run with no built-in variance estimation	One sample per simulation run with no built-in variance estimation
Parameter inference	Yes, uses Pyro library for Bayesian inference from data	No, must be tuned manually or with external software to fit data	No, must be tuned manually or with external software to fit data
Compute scaling	A typical 28-day run in New York takes 10-20 minutes on a 16-core server. ³	Cited as 5 to 10 million agent-days per second on typical personal hardware	No performance figures given, but explicitly intended to be run on HPC clusters
Memory scaling	Scales quadratically with the number of distinct subpopulations	Appx. 1 GB / 1M agents (linear scaling)	High (exact statistics unavailable, 3GB for a single-region simulation is not unexpected)
Heterogeneity of demographic groups [within regions]	Yes, through subdivision of geographical nodes into homogeneous subpopulations	Yes	Yes
Regional granularity	Arbitrary; derived from administrative regions (cities, states/provinces, countries) by default.	One region only ⁴	Square grid only
Region-level statistics support	Yes	No (cannot model regions directly)	Yes
Heterogeneity of infection between regions	Yes	Officially "possible, but not easy" ³	Yes
Individual risk assessment support	Yes ⁵	Partially (individual events available, no subsampling)	Partially (individual events available, no subsampling)
Spatial movement	Explicitly supported ⁶	Not explicitly supported	Explicitly supported (migration files)
Data-driven mobility	Yes	Planned	Yes
Incorporates peer-reviewed studies for vaccine efficacy	Yes	No	No
Data scraping tools included	Yes	Yes	No
Applicability to new diseases	Relatively easy (directly extensible with Python)	Relatively easy (directly extensible with Python)	Difficult (must build C++ app from source to extend).

² Can use an assumption of node-homogeneity to reduce the required computation, becoming more of a standard regional model.

³ Pandemonium is designed to produce distributions of outcomes to reflect variance/uncertainty, and it can infer latent variables including parameters. As such, one Pandemonium run is equivalent to many Starsim or EMOD simulations for the purposes of statistical analysis, and will fully self-tune to input data.

⁴ Starsim directly supports only one region per model run. While modeling interactions between regions or expressing heterogeneity is possible through individuals' contact graph, these must be constructed manually or with external automation, making it challenging to implement.

⁵ We currently provide scaled individual modeling when evaluating risk for a single individual by modeling them as a population without self-interaction and proportionally reduced impact on other populations.

⁶ We consider there to be 3 types of spatial movement: 1. Continuous/Cyclic/Commuting 2. Migration/Permanent Movement 3. Dynamic/Arbitrary. We support 1 & 3.

Feature	Pandemonium	Starsim	EMOD
Epidemiological model	Supports arbitrary compartmental models as a framework. SEIRD model included in webapp.	Model configurability seems poor	Supports SEIR, SEIRS, SIR, SIRS, SI, and SIS models; can be extended with new diseases and presumably new models in C++.
Granularity of compartmental model state	Arbitrary (integer) division within subpopulation	Each agent can be in one state, even if representing multiple people	Each agent can be in one state, even if representing multiple people
Asymptomatic carriers	Not currently explicitly modeled (but will be)	Yes	Possible (Typhoid model includes asymptomatic carriers)
Contact networks	Yes ⁷ (via coupling factors)	Yes	Yes
Health system load modeling	No (but could be added to compartmental model)	Yes	No
Viral load model	Yes (only in MicroModel ⁸)	Yes	In disease-specific models, not Generic, but has HINT
Biophysical airborne transmission model	Yes (via [Bazant & Bush] MicroModel integration)	No	No
Importance sampling?	Not Yet (via Pyro)	No	No
Cluster support	No (hopefully unnecessary)	Yes, natively	Yes, with MPIexec
Common Random Numbers technique for intervention impact noise reduction	Planned but very difficult (would likely require modifying Pyro and/or implementing automatic graph surgery)	Yes	No
Models adaptive behavior of people	Planned (psychological-behavioral model)	Yes	Yes

⁷ caveat: multiple individuals have not yet been tested

⁸ Viral load modeling occurs within the integrated MicroModel, specifically during MicroEvents, and is not generalized or directly exposed within the broader MacroModel.