

# ASKEM December 2024 Hackathon Scenarios

The views, opinions, and/or findings contained in this report are those of The MITRE Corporation and should not be construed as an official Government position, policy, or decision, unless designated by other documentation. This technical data deliverable was developed using contract funds under Basic Contract No. W56KGU-18-D-0004.

## USE CASE: EPIDEMIOLOGY

To prepare for the final program evaluation, we have developed hackathon materials that are representative of and exercise similar functionality as our target expectations for the evaluation. These questions are meant to help guide and prioritize critical development for success in the evaluation. The goal is to address as much as possible within the Terarium workbench (including the interactive notebook environment).

## Table of Contents

<b>Decisionmaker Scenario: Mpox.....</b>	<b>2</b>
<b>Background.....</b>	<b>2</b>
<b>Part 1: Finding the Right Starting Point.....</b>	<b>2</b>
<b>Part 2: Developing/Updating the Model.....</b>	<b>3</b>
<b>Part 3: Addressing Decision-maker Priorities .....</b>	<b>4</b>
<b>Independent Questions .....</b>	<b>6</b>
<b>Section 1: TA1 Functionality .....</b>	<b>6</b>
Set 1.1: Model Extraction and Unit Tests, for Covid Domain .....	6
Set 1.2: Model Extraction and Unit Tests, for New Epi Domain .....	7
Set 1.3: Model Search.....	7
Set 1.4: Data Search .....	8
<b>Section 2: TA2 Functionality .....</b>	<b>9</b>
Set 2.1: Multi-model Structural Comparison .....	9
Set 2.2: Model Editing.....	9
Set 2.3: Model Stratification.....	10
Set 2.4: Model Checks.....	10
<b>Section 3: TA3 Functionality .....</b>	<b>11</b>
Set 3.1: Calibration .....	11
Set 3.2: Forecasting or Forward Simulation .....	12
Set 3.3: Interventions.....	12
Set 3.4: Optimizing Interventions.....	13
Set 3.5: Ensembles.....	14
Set 3.6: Sensitivity Analysis .....	14
<b>Section 4: TA4 Functionality .....</b>	<b>15</b>
Set 4.1: High Level Model Summarization, Comparison, and Gap Analysis .....	15
Set 4.2: Annotated Plots/Visualizations and Comparison Against Data.....	15

# Decisionmaker Scenario: Mpox

## Background

Mpox (formerly monkeypox) is a zoonotic disease endemic to parts of Africa. There have been multiple outbreaks in which human-to-human transmission has occurred. In this scenario, you are supporting a European-based decision-maker who needs to decide whether to send mpox vaccine doses to Africa, where an mpox outbreak is currently occurring, and there is concern about it spreading globally. Assume that Africa has a stockpile of 500,000 mpox vaccines and the European Union (EU) has a stockpile of 3 million mpox vaccines. The EU could send vaccines to African countries experiencing mpox outbreaks, as a mitigation measure against importation of diseases from Africa to Europe and elsewhere. For this scenario, assume that a European decision-maker is only concerned with minimizing cases and deaths within the EU.

To support your modeling, use the following papers (which are also provided in the supplementary material):

- A. **Model A:** Al-Shomrani, M.M.; Musa, S.S.; Yusuf, A. Unfolding the Transmission Dynamics of Monkeypox Virus: An Epidemiological Modelling Analysis. *Mathematics* 2023, *11*, 1121. <https://doi.org/10.3390/math11051121>
- B. **Model B:** Peter, O.J., Kumar, S., Kumari, N. *et al.* Transmission dynamics of Monkeypox virus: a mathematical modelling approach. *Model. Earth Syst. Environ.* 8, 3423–3434 (2022). <https://doi.org/10.1007/s40808-021-01313-2>
- C. **Model C:** Rabiou, M., Dansu, E.J., Mogbojuri, O.A. *et al.* Modeling the sexual transmission dynamics of mpox in the United States of America. *Eur. Phys. J. Plus* 139, 250 (2024). <https://doi.org/10.1140/epjp/s13360-024-05020-6>

## Part 1: Finding the Right Starting Point

- 1) **(TA1/TA4 Model extraction)** Ingest the three models referenced above. Implement the following unit tests to ensure that you're able to get them into executable states and can faithfully capture the model structure and parameter values.
  - a. Using Model A, recreate figure 2a from the paper.
  - b. Using Model B, recreate figure 6 from the paper.
  - c. Using Model C, recreate figure 2 from the paper (cases for Men and Women over time graphs only).
- 2) **(TA4 Model Comparison)** Do a model comparison based on key differences in assumptions, strengths, limitations, and distinguishing characteristics. Based on this information, rank each model in terms of its relevance and fit-for-purpose for this scenario context and decision-maker needs.

Model	Distinguishing Characteristics	Assumptions	Strengths	Limitations	Rank fit-for-purpose (1 = most suitable; 3 = least suitable), with reasoning
A					
B					
C					

- 3) **(TA2 Structural Model Comparison)** Now perform structural model comparison between each pair of models. By structural comparison, we seek to understand how compartments and transition pathways overlap or diverge between models. Feel free to create diagrams or use equations in your response.
- 4) **(Model Selection)** Select the model you think is the most appropriate starting point for supporting the decision-maker, given the background context for this scenario. Please explain your reasoning.
- 5) **(Parameter Selection – African Model)** For the chosen model in Q4, determine a range of appropriate parameter values for each parameter in the model, for use in an African context. Fill in the following information about sources and quality. You can pick parameter values from any of the 3 referenced papers or any external sources, regardless of which model you chose to use. Note the size of the time step for the various models when determining parameter ranges.

Parameter	Parameter Definition	Parameter Units	Parameter Value or Range	Uncertainty Characterization	Sources	Modeler Assessment on Source Quality
-----------	----------------------	-----------------	--------------------------	------------------------------	---------	--------------------------------------

- 6) **(Reparameterization – European Model)** The decision-maker is now considering applying this model to locations in Europe instead of Africa. The transmission parameters may or may not be different in Europe and Africa. Determine a plausible range of parameter values for a European context, based on the parameters in the reference papers, or external sources. As in Q5, record information about your chosen parameter values, as well as your justifications for choosing them.

## Part 2: Developing/Updating the Model

- 7) **(TA3 Model Configuration)** Outside of Africa, mpox is not endemic (technically, “enzootic”) in the animal population, and thus the animal parameters would be negligible. Modify the Q6 model (European Model) parameters with the relevant animal parameters set to zero. Ensure the model runs and gives plausible outputs.
- 8) **(TA3 Simulation, TA4 Visualization)** Now simulate the African and European models separately (with all of the modifications in the previous questions), and compare results, using visualizations to support the comparison. For initial conditions, set total susceptible populations with reasonable population numbers from any of the papers referenced in this scenario, or external sources. Assume five humans are infectious at  $t(0)$ , and other compartments have initial conditions of 0 people.
- 9) **(TA2 Model Edits)** Some research suggests that pet animals are susceptible to mpox infection: [Eurosurveillance | The risk of reverse zoonotic transmission to pet animals during the current global monkeypox outbreak, United Kingdom, June to mid-September 2022](#). Starting from the updated European Model in Q7, create a contact pathway from humans to animals to capture this “reverse zoonotic transmission.” The rate of transmission is unknown, so use the animal-to-animal, animal-to-human, and human-to-human transmission parameters as proxies.

Distribution Statement A. Approved for public release: distribution is unlimited

MITRE Public Release Case Number 24-2089

© 2025 The MITRE Corporation.

- a. What is the impact of adding this pathway compared to the model with no reverse zoonotic transmission, in terms of infections per year? Show the impact using a visualization.
  - b. Perform a sensitivity analysis of the reverse zoonotic transmission parameter on the number of infections. For what range of values does this parameter make a material difference in number of human infections? (Material difference means at least 0.1% more infections, or at least 1 infection, whichever is larger.) Provide the range of values and the corresponding difference in number of infections.
- 10) (TA2 Model Edits, TA3 Simulation, TA4 Visualization) The decision-maker now wants to consider a situation where cases can be imported from Africa into Europe. Create and simulate a combined model consisting of an African model with an enzootic animal population and a European model with pet animals. Use realistic values for the starting populations – as before, you may derive these values from any of the papers referenced in this scenario, or external sources. Start with zero European cases at  $t(0)$  and create a connection between the African and European models by which infected people in Africa can expose the susceptible population in Europe through travel. Assume a contact rate of 0.2% annually (i.e., 0.2% of the susceptible population in Europe becomes exposed each year). Create plots showing the infected populations in Africa and Europe over time.
- 11) (TA2 Model Edits) In order to support the decision-maker's questions about vaccination strategy, we need to add a vaccination mechanism in our model. Starting from the combined model in Q10, create a new 'Vaccinated' compartment, in which highly susceptible people move to the 'Vaccinated' compartment at a rate of 100,000 people each day in each continent until that continent's supply is exhausted. Assume the vaccine is 100% effective.

### Part 3: Addressing Decision-maker Priorities

- 12) (TA3 Simulation) Is it more effective for the European Union to send all the vaccine doses to Africa or to vaccinate their own population? Assume the European Union's primary outcome of interest is the number of domestic cases.
- a. (Optional) Find the optimal number of doses for the EU to send to Africa to minimize the number of domestic cases (in the EU).
- 13) (TA3 Scenario Templates) **Horizon Scanning:** A horizon scan is a type of forecast that explores possible outcomes under uncertainty. See Figure 1 below (from [Runge et al, 2024](#)) as an example.

## Horizon Scanning

Example: Characterize potential outcomes if a new viral variant emerges

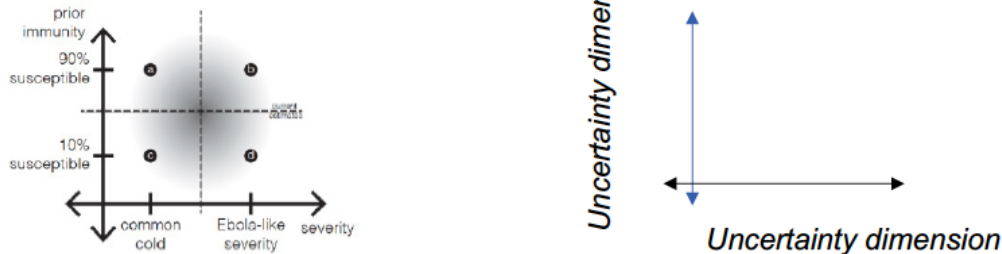


Figure 1. Example of a horizon scanning scenario from Runge et al.

A new strain of mpox, Clade 1b, has emerged, and the modeling parameters for this new subclade are uncertain. Consider the following sources (which are also part of the supplemental materials for this scenario):

- <https://centerforhealthsecurity.org/sites/default/files/2024-06/20240610-mpoxsituationreport.pdf>
- <https://www.cdc.gov/cfa-modeling-and-forecasting/mpox-transmission-technical-brief/index.html>

Perform a horizon scanning analysis to indicate a range of possible outcomes in Europe. Consider the following parameters:

- The case fatality percent (i.e., case fatality rate) could range from 1% to 12%. To account for fatalities in a model without a Deaths compartment, you can simply multiply the population in the R compartment at the end of the simulation by 1% and 12%.
- The human-to-human transmission rate could be up to 100% faster than other clades in a worst-case scenario.

Run this horizon scan for both the scenario in which all of Europe's vaccines are used in Europe, and the scenario in which Europe sends all its vaccines to Africa. In these situations, Europe would also be concerned with domestic deaths rather than just domestic cases.

Given the range of uncertainty, is it better for the European Union to send the vaccine doses to Africa or vaccinate domestically? Develop a recommendation for the decision-maker you are supporting, and provide numerical comparisons and figures from the model to justify your answer.

# Independent Questions

This section describes MITRE's current thinking about the Independent Questions we intend to have evaluation modelers working on, in order to get timing metrics for different pieces of functionality, with minimal interdependencies. This document is provided for Performer review during the Hackathon. Please feel free to address any questions you might have about the described question types to the MITRE team.

These problems are intended to be more like textbook problems than the more detailed and realistic decision-maker scenarios. Each Set will have 3 questions of varying difficulty/complexity (simple, medium-complexity, and high-complexity). All problems are intended to be independent of each other, except for the Stratification questions which will likely build from simple to complex.

## Section 1: TA1 Functionality

### Set 1.1: Model Extraction and Unit Tests, for Covid Domain

*Given a paper, extract the model with human-in-the-loop (HITL) curation, so that it can reproduce the unit test.*

Three papers describing COVID-19 models of increasing complexity will be provided to ingest. The inputs will be a paper to extract a model from, and a model configuration to execute a unit test. The outputs expected will be a compartmental diagram and set of equations describing the extracted model, and some basic simulation plots demonstrating that the extracted model is executable and is correct enough to satisfy the unit test.

An example set of inputs, tasks, and outputs for this set of questions would be:

#### Inputs:

- Paper containing model description and equations
- Model configuration (initial conditions, parameter values, simulation parameters) for unit test: (S, E, I, R, D,  $\alpha$ , dt, sim\_length) = (1000 people, 100 people, 100 people, 100 people, 0 people, 0.1 [parameter units], 1 day, 100 days)
- Expected unit test output (could be a plot or characteristics of the output, e.g. variable  $I$  peaks at time  $t$  days)

#### Tasks:

- Extract model from paper with HITL curation
- Ground model to DKG
- Configure model
- Run unit test

#### Outputs:

- Model in executable state
- Compartmental diagram of the model and set of equations
- Definitions of all variables and parameters, with units
- DKG groundings for all variables and parameters
- Unit test results showing they match expected output

Distribution Statement A. Approved for public release: distribution is unlimited

MITRE Public Release Case Number 24-2089

© 2025 The MITRE Corporation.

## Set 1.2: Model Extraction and Unit Tests, for New Epi Domain

*Given a paper, extract the model and ground concepts with HITL curation, so that it can reproduce the unit test.*

Three papers describing non-COVID-19 epidemiological models of increasing complexity will be provided to ingest. The inputs will be a paper to extract a model from, and a model configuration to execute a unit test. The outputs expected will be a compartmental diagram and set of equations describing the extracted model, and some basic simulation plots demonstrating that the extracted model is executable and is correct enough to satisfy the unit test.

An example set of inputs, tasks, and outputs for this set of questions would be:

### Inputs:

- Paper containing model description and equations
- Model configuration (initial conditions, parameter values, simulation parameters) for unit test: (S, E, X, Y, R,  $\beta$ , dt, sim\_length) = (1000 people, 5 people, 1000 [units of animal population], 5 [units of animal population], 0 [units of animal population], 0.1 [parameter units], 1 week, 4 weeks)
- Expected unit test output (could be a plot or characteristics of the output, e.g. variable  $X$  peaks at time  $t$ )

### Tasks:

- Extract model from paper with HITL curation
- Ground model to DKG (may require extensions to DKG)
- Configure model
- Run unit test

### Outputs:

- Model in executable state
- Compartmental diagram of the model and set of equations
- Definitions of all variables and parameters, with units
- DKG groundings for all variables and parameters
- Unit test results showing they match expected output

## Set 1.3: Model Search

*Find a model that meets some specified criteria, from the public Terarium projects.*

Three sets of model search criteria will be provided. The expected outcome will be the selected model and completion of a checklist/table describing how the model fits the criteria or can be modified to fit the criteria.

An example set of inputs, tasks, and outputs for this set of questions would be:

### Inputs:

- Search criteria for model to meet (e.g., it should output hospitalizations and deaths, and contain parameters related to social distancing and vaccines)

### Tasks:

- Search for relevant models and select the best one (based on user judgement) from search results



#### Outputs:

- List of models, sorted by relevancy and with descriptions for each
- For selected model, a checklist showing which search criteria it meets, and a description of how it could potentially be modified to fit the unmet criteria

**Note:** for *Model Search* we plan to make more concrete decisions about evaluation following the Hackathon and our assessment of what is ready and worth evaluating with respect to these features.

## Set 1.4: Data Search

*Search for already uploaded datasets within public Terarium projects, that meet some specified criteria.*

Three sets of dataset search criteria will be provided. The expected outcome will be the selected dataset and completion of a table describing how the dataset fits the criteria or can be modified/enriched to fit the criteria.

An example set of inputs, tasks, and outputs for this set of questions would be:

#### Inputs:

- Search criteria for dataset to meet, as a list of concepts (e.g., it should contain susceptible population counts, vaccinated population, apply to location X, span time period Y, etc.)

#### Tasks:

- Search for datasets that meet input criteria, and select the best one (based on user judgement) from search results

#### Outputs:

- List of links to datasets, sorted by relevancy and with descriptions of each dataset, including:
  - Title
  - Short summary
  - Any information on spatiotemporal resolution (e.g. daily/weekly/monthly case data, or county/state/national level aggregation)
  - Any information on spatiotemporal extent (e.g. time period and geographic areas covered)
- For selected dataset, a checklist showing which search criteria it meets, and a description of how it could potentially be enriched/transformed/modified or complemented with other sources to fit the unmet criteria

**Note:** for *Data Search* we plan to make more concrete decisions about evaluation following the Hackathon and our assessment of what is ready and worth evaluating with respect to these features.



## Section 2: TA2 Functionality

### Set 2.1: Multi-model Structural Comparison

*Given multiple grounded compartmental models of a similar system or set of processes, do a structural comparison and provide output that clearly illustrates the similarities and differences between them (with respect to parameters, state variables or compartments, and transition pathways).*

Problems will range in complexity from comparison of 2 small, simple models, to multiple larger, more complex, stratified models.

An example set of inputs, tasks, and outputs for this set of questions would be:

Inputs:

- 2-4 (optionally grounded) models to be compared

Tasks:

- Do a structural model comparison with the input models

Outputs:

- A diagram that illustrates common and differing structural components of the models, along with any complementary text summary

### Set 2.2: Model Editing

*Perform a variety of modifications to the basic SIR model.*

Inputs will be a simple base SIR-type model. Outputs will be new models requiring simple, moderate, or complex changes to the model, along with unit test results to confirm that changes were made correctly. Changes will involve adding and or removing compartments and transition pathways, and creating or modifying observables.

An example set of inputs, tasks, and outputs for this set of questions would be:

Inputs:

- Simple base model (e.g. SIR or SEIRHD model)
- Instructions on what edits need to be made, with a diagram of the target model (e.g. add a transition pathway between R and S compartments to signify waning immunity over time and susceptibility to reinfection, add vaccination compartments, etc.)
- Model configuration (initial conditions, parameter values, simulation parameters) for unit test with modified model
- Expected unit test output

Tasks:

- Make model edits of varying complexity according to instructions
- Configure model
- Run unit test

Outputs:

- Modified model in executable state and that meets the structure of the target model diagram
- Unit test results showing they match expected output

Distribution Statement A. Approved for public release: distribution is unlimited

MITRE Public Release Case Number 24-2089

© 2025 The MITRE Corporation.

## Set 2.3: Model Stratification

*Perform a variety of stratifications to a basic epidemiological model.*

Starting with a simple epidemiological model, perform increasingly complex sets of stratification across components that may include disease dynamics, population features (age, sex, other demographic characteristics, etc.), location, treatment types, etc.

An example set of inputs, tasks, and outputs for this set of questions would be

### Inputs:

- Simple base model (e.g. SIR or SEIRHD model)
- Specification of which compartments should be stratified, and number of groups to stratify each by
- Specification of how base model parameters should be modified, and any relevant data needed to inform them (e.g. age contact matrix or mobility matrix, etc.)
- Model configuration (initial conditions, parameter values, simulation parameters) for unit test with stratified model
- Expected unit test output

### Tasks:

- Stratify model according to input specifications
- Configure model
- Run unit test

### Outputs:

- Stratified model in executable state with:
  - New compartments labeled in an understandable way to indicate how they relate to the original unstratified compartments
  - New parameters labeled in an understandable way to indicate how they relate to the original unstratified model parameters
- Unit test results showing they match expected output

## Set 2.4: Model Checks

*Perform model checks of varying complexity against provided models.*

Starting with a fully configured input model, perform increasingly complex types of model checks, which can include monotonicity and population conservation checks, and arbitrary constraints defined as some function of state variables and parameters.

An example set of inputs, tasks, and outputs for this set of questions would be:

### Inputs:

- Fully configured model
- List of constraints or “checks” that a configured model must meet during simulation, which may include:
  - Monotonicity checks (one or more compartment values never increases/decreases over the course of the simulation)

- Population conservation: Some sum of state variables (defined as an observable function), remains constant over the course of the simulation, assuming vital dynamics are not part of the model
- Arbitrary constraints/functions defined in terms of some combination of state variables, parameters, and observables (e.g.  $\delta \geq 2\varepsilon$ )
- Ensure physical realism of prior distributions, e.g. transmission parameter can never be a negative value. This check can be implemented through a combination of the above categories, as needed.

Tasks:

- Perform model checks (can be probabilistic or deterministic)

Outputs:

- Deterministic case - binary output that indicates whether checks passed
- Probabilistic case - diagrams or plots showing combination of parameter ranges that satisfy the model check constraints

## Section 3: TA3 Functionality

### Set 3.1: Calibration

*Calibrate one or more parameters for a provided model based on data.*

Take a given model and dataset and calibrate for requested parameter(s) and compute the error between data and calibrated model output. Problems will vary in complexity from calibrating a single parameter, to simultaneously calibrating multiple parameters.

An example set of inputs, tasks, and outputs for this set of questions would be:

Inputs:

- Input model to be calibrated
- Dataset.csv and indication of which features to use for calibration
- Model configuration, including initial conditions and all parameter values other than those to be calibrated. Known parameters that won't be calibrated, may still have a range of values.
- List of parameters to be calibrated with prior distributions for each

Tasks:

- Perform calibration

Outputs:

- For each calibrated parameter, provide mean parameter value(s) pre- and post-calibration, prior and posterior distributions, and variance pre- and post-calibration
- Some measure of goodness-of-fit for the calibrated parameters, such as mean absolute error (MAE) between projected value for one or more state variables, and observational data from Dataset.csv

## Set 3.2: Forecasting or Forward Simulation

*Create probabilistic or deterministic forecasts/simulations with a provided model.*

Given a model and configuration to apply, simulate and create plots of all state variables for a given number of timesteps. Three models of increasing complexity will be tested.

An example set of inputs, tasks, and outputs for this set of questions would be:

### Inputs:

- Input model to simulate
- Model configuration for forecast/simulation, including initial conditions, parameter values, and simulation parameters (e.g. dt, length of simulation, etc.). For deterministic forecasts/simulations, there will be no uncertainty in initial or parameter values. For probabilistic forecasts/simulation, one or more parameters will be defined as distributions.
- Expected output of simulation, in terms of key inflection points (peak of state variable X is reached at timepoint 10 days, or minimum value of Y occurs at timepoint 2 days, etc.)

### Tasks:

- Configure model according to specification
- Perform forecast/simulation

### Outputs:

- Plots for all state variables, each in a separate plot, and indication that expected output is met (e.g. identification of peaks, etc.). For probabilistic forecasts/simulation, this includes posterior samples of all states.

## Set 3.3: Interventions

*Implement interventions in a provided model, simulate, and assess impact.*

Given a fully configured model, add increasingly complex sets of interventions, simulate, and then create requested plots describing intervention impacts.

An example set of inputs, tasks, and outputs for this set of questions would be:

### Inputs:

- Model with pre-intervention configuration applied (including initial conditions, parameter values, and simulation parameters)
- One or more interventions to implement, specified as changing the value of one or more parameters, and creating and applying a new model configuration. Interventions may also be specified as changing the values of a dataset attached to the model configuration process (e.g. contact matrix). Interventions can be:
  - Static: Setting parameters to values that stay constant over time
  - Dynamic: Setting parameters to values that change over time. This includes implementing interventions at specific timepoints.
- Specific outputs of interest (which could be state variables or observables) we want to measure the impact of the intervention against (e.g. total number of infections, number of hospitalizations, etc.). Problem may request this in terms of Average Treatment Effect

(ATE), Average Treatment Effect on the Treated (ATT) or some other comparison metric.

Tasks:

- Simulate model with pre-intervention configuration to get baseline results with respect to the outputs of interest
- Create new model configuration (intervention configuration) with the specified interventions
- Simulate model with intervention configuration
- Calculate impact of intervention with respect to the outputs of interest

Outputs:

- Simulation output plots for outputs of interest, before and after intervention was implemented
- Impact of intervention with respect to outputs of interest, in numerical numbers (e.g. approximate difference in number of hospitalizations between baseline and after intervention policy applied)

### Set 3.4: Optimizing Interventions

*Find optimal interventions based on provided goals.*

Given a configured model and a goal, optimize interventions and create requested plots to prove that the optimized intervention satisfies the goal. Optimization can be value-based (intervention parameter values), time-based (find the earliest/latest time for an intervention, or minimize the total amount of time an intervention is in effect), and involve a single or multiple objectives.

An example set of inputs, tasks, and outputs for this set of questions would be:

Inputs:

- Model with initial unoptimized intervention configuration applied (initial conditions, parameter values, and simulation parameters)
- Objective function with one or more objectives in terms of variables or observables to optimize. E.g. find minimal mask efficacy required to achieve outcome  $X$  (defined as one or more constraints); find the latest time a social distancing policy can be implemented and still achieve outcome  $Y$ ; minimize cumulative infections AND minimize cost according to some cost function.
- Constraints on the optimization problem, in terms of parameters, state variables, or observable functions. Constraints can apply for the entire simulation, or just for certain time periods. E.g.  $p_1 \geq 2 * p_2$ ,  $R(t) < 1.0$ ;  $I(t) < 5000$  for all  $t$ ;  $I_{cumulative} < 10k$

Tasks:

- Simulate model with unoptimized configuration to get baseline results with respect to the objective
- Perform value- or time-based optimization (or some combination of the two), with single or multiple objectives
- Simulate model with optimal intervention

Outputs:

- A solution to the objective function (e.g., to achieve  $X$ , the minimum required mask efficacy is  $\rho$ ,  $t_{int}$  is the latest time that intervention policy  $Z$  can be implemented, etc.)

- A verification that the provided solution solves the optimization question (e.g. plot showing results with respect to objective and constraints, pre- and post-optimization, etc.)

### Set 3.5: Ensembles

*Create an ensemble of provided models.*

Given models and calibration data, create an ensemble of the models that best fits the calibration data and then perform a forecast.

An example set of inputs, tasks, and outputs for this set of questions would be:

#### Inputs:

- Two or more models that we would like to create ensemble with
- Dataset.csv and indication of which features to use for calibration
- Forecast configuration (e.g. initial conditions, number of timesteps)

#### Tasks:

- Create an ensemble, which can be weighted or non-weighted
- Perform forecast with ensemble according to configuration

#### Outputs:

- Calibration results (weights of component models and measures of error / calibration fit, calibrated parameters for component models)
- Forecast output for all state variables, including uncertainty intervals
- Mean absolute error (MAE) between projected value for one or more state variables from the ensemble model, and observational data from Dataset.csv

### Set 3.6: Sensitivity Analysis

*Determine how one or more parameters influence one or more variables in a model.*

An example set of inputs, tasks, and outputs for this set of questions would be:

#### Inputs:

- Fully configured input model
- List of parameters of interest in sensitivity analysis, with valid ranges or probability distributions
- List of output variable(s) the sensitivity analysis should include

#### Tasks:

- Perform sensitivity analysis - determine how one or more parameters influence one or more output variables

#### Outputs:

- Plots of sensitivity results with units for each parameter and variable, with some indication of how a change in model parameter value(s), results in an amount of change in output variable(s)
- List of model parameters that have greatest impact on output variable(s), in order of impact

## Section 4: TA4 Functionality

### Set 4.1: High Level Model Summarization, Comparison, and Gap Analysis

*Given a decision-maker objective, compare candidate models, provide a high-level summary of each, determine the gap between model capabilities and decisionmaker objectives, and rank fit-for-purpose.*

Perform comparisons of SIR-types of compartmental models for COVID-19. In terms of complexity, we will vary the number of models and the subtlety of the differences between them. The output will be a table describing each model's distinguishing characteristics, assumptions, strengths, limitations, and fit-for-purpose ranking.

An example set of inputs, tasks, and outputs for this set of questions would be:

Inputs:

- 2 or more publications (with model descriptions contained) to be compared
- Decision-maker objective(s)

Tasks:

- Summarize each model, including distinguishing characteristics, assumptions, strengths, and limitations
- With respect to the decision-maker objective(s), perform gap analysis with each model, describing whether the model is sufficient for meeting the objective, and if not, what is missing. If there is functionality missing, determine what could be modified/added to meet the objective(s).
- Rank each model with respect to fit-for-purpose and provide reasoning

Outputs:

- Table describing each model's distinguishing characteristics, assumptions, strengths, and limitations, gap analysis, and fit-for-purpose ranking

Model	Distinguishing Characteristics	Assumptions	Strengths	Limitations	Gap Analysis	Rank fit-for-purpose (1 = most suitable; 3 = least suitable), with reasoning
A						
B						
C						

### Set 4.2: Annotated Plots/Visualizations and Comparison Against Data

*Create decision-maker friendly visualizations.*

Generate requested charts and annotations to compare forecasts against data or thresholds, compare interventions, examine optimization results, etc.

An example set of inputs, tasks, and outputs for this set of questions would be:

Inputs:

- Observational to plot model output against and measure error metrics
- Annotations to include in plots (e.g. peak of an infection curve and time of peak, threshold values, etc.)



Tasks:

- Generate plots according to input specifications

Outputs:

- Output 'ready for report' plots and visualization artifacts

**Note:** for this set of questions we plan to make more concrete decisions about evaluation following the Hackathon and our assessment of what is ready and worth evaluating with respect to these features.