

macpan-workshop

Syllabus

License CC BY-NC-SA 4.0

This is a (currently draft) syllabus for a workshop on using [macpan2](#) for applied public health modelling.

(Note: I highlight areas that cannot be finalized yet, with notes like this)

Table of Contents

- [Instructor](#)
- [Audience](#)
- [Background](#)
- [Objectives](#)
- [Materials](#)
- [Outcomes](#)
- [Schedule](#)
- [Preparation](#)
- [Footnotes](#)

Instructor

Steve Walker – swalk@mcmaster.ca – github.com/stevencarlislewalker

Audience

We assume that workshop participants are epidemiological/disease modellers who use, or want to use, compartmental models to address public health issues. The material is specifically tailored to such modellers at the [Public Health Agency of Canada](#).

Participants should be comfortable with the following.

- Basic terminology and ideas about [compartmental models in epidemiology](#).
- Basic coding to solve scientific problems.¹
- Familiarity with common terminology in statistical modelling.²

Details on how to prepare for the workshop are covered [here](#).

Background

The [macpan2](#) package is a flexible compartmental modelling tool that has been optimized for fast calibration to data. This package grew out of lessons learned developing the [McMasterPandemic](#) COVID-19 model, which was used to support public health during the pandemic.

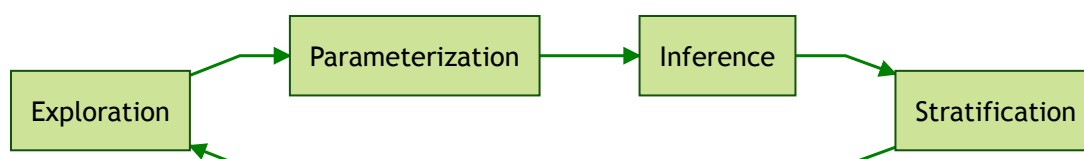
Applied compartmental modelling is a big topic. The following activities describe the aspects of this topic for which `macpan2` is designed.

- **Exploration:** Compare simulations freely under alternative parameterizations, and with real and/or hypothetical data.
- **Parameterization:** Refine default parameters for a specific population and question, possibly by formally calibrating parameters with data.
- **Inference:** Use realistically parameterized models to make inferences (e.g. confidence intervals for parameters, forecasts, counter-factuals).
- **Stratification:** Separate individuals in one compartment into many (e.g. by age, space, vaccination status) and/or find stratified data.

We stress the importance of data and other factual inputs, including expert opinions, into the modelling process. We are concerned with modelling specific real-world problems, and so none of these activities should be too abstract.

Example Workflow

Starting with exploration, and then iterating within the following model refinement cycle, is a good default workflow for applied public health compartmental modelling projects.



In this cycle, exploration often clarifies if our existing parameter values need refinement. In turn, parameterization refines a model so that it can be used to make defensible inferences in a specific context. Such inferences can raise questions about whether a model should be stratified to make it more realistic. Finally, stratification can create new model behaviours that may require further exploration, starting the cycle over again.

We have learned from our experience in applied public health modelling that it is better to complete each iteration of this cycle as quickly as possible, rather than trying to produce the perfect model outright. Keeping the scope of each step as limited as possible, or even skipping steps if they are not a priority, is useful for avoiding [analysis paralysis](#). Every time we get back to the **inference** step we have another chance of providing valuable inputs to a public health debate, and so we want to do this as often as possible.

The final stratification step will increase model complexity, which has both advantages and disadvantages. Therefore, complexity will tend to increase as a modelling project iterates through the cycle. Starting with a simple model will help ensure that the project converges on an appropriate level of complexity. Sometimes it is worth resetting the cycle by starting again with a simpler model.

We will use this model refinement cycle to provide context when concepts are introduced in the workshop. The instructor will take a modelling example through one cycle during the workshop (Note: The example that we will use needs to be discussed as the contract proceeds). Participants will be given time to try some of these techniques on [their own models that they have come prepared with](#).

Objectives

The overall goal of the workshop is to train epidemiological modellers to use `macpan2` software in their efforts to use compartmental modelling to support public health. Each [session](#) will cover techniques associated with one step in the [model refinement cycle](#).

There will not be enough time in the workshop to cover all of `macpan2`, so there will be a [companion set of online materials](#) that will cover all of the options available to modellers. In the workshop, we will follow one example through one full iteration of the [model refinement cycle](#). Variations on each skill that is covered will be described in the [companion material](#). Where applicable, the lists of workshop objectives link to lists of these variations.

Session 1: Exploration

Participants will learn how to do the following types of tasks required for exploring model simulations.

- Find candidate simulation models in the `macpan2` [library](#) of starter models.
- Prepare certain [types of data](#) so that they can be compared with `macpan2` simulation output, both visually and numerically.
- Make [modifications to models](#) in the library.
- Compute [epidemiological summaries](#) (e.g., basic reproduction number, \mathcal{R}_0).
- Cast a model as a particular [dynamical model type](#) (e.g. discrete-time recursion, ordinary differential equation).

Session 2: Parameterization

Participants will learn how to do the following types of tasks required for parameterizing models for making inferences about a particular population and public health problem.

- Modify the default values of parameters in a model.
- Express uncertainty in model parameters (e.g., transmission rate) or in [epidemiological summaries](#) (e.g., \mathcal{R}_0) with prior distributions.
- Use [optimization](#) to calibrate parameters (e.g., transmission rate) so that the discrepancy between observed and simulated data is minimized.
- Parameterize the [initial values of the state variables](#) (e.g. S , I) so that they can be optimized.
- Calibrate the functional form of time-variation of parameters using machine learning sub-models embedded within epidemiological models. This is a useful technique when the reasons for parameter time-variation are not well-understood.

Session 3: Inference

Participants will learn how to do the following types of tasks that are often necessary when making inferences using realistically parameterized models.

- Visualize goodness-of-fit.
- Generate confidence intervals for estimated parameters.
- Forecast model variables beyond the last data point.
- Calculate prediction intervals measuring uncertainty about these forecasts.
- Compare alternative scenarios for counter-factual causal analysis (e.g., how many deaths were saved due to vaccination?).
- Produce uncertainty estimates for [epidemiological model summaries](#) like \mathcal{R}_0 .

Session 4: Stratification

Participants will learn the following stratification strategies.

- [Stratify every compartment](#) in the same way (e.g. by age, location).
- [Stratify infectious compartments](#) (e.g. by symptom status).
- Expand a single compartment into a sequential chain, to model different distributions of time spent in that compartment.
- Combine several single-strain models into a single multi-strain model.
- Stratify compartments based on immunity status. This is particularly useful in cases where immunity is partial, can wane, and is caused by a mixture of past infection, vaccination, and cross-immunity.

Materials

(Note: This section will not be in the final syllabus, but I've put it in the draft to give an idea of what the companion materials will contain once they are produced in a subsequent milestone.)

Library models

The [companion material](#) will describe each of the following starter models. Although every emerging pathogen has a unique set of characteristics, there typically exist disease models that can serve as starting points in a crisis situation. The `macpan2` [library of candidate models](#) includes a variety of pathogens that cover the range of likely characteristics for emerging threats (Note: Currently this is a placeholder list, which will be finalized using PHAC input in a subsequent milestone. Still, suggestions as soon as possible would be great).

- Seasonal respiratory illnesses (e.g., influenza, COVID-19, RSV).
- Childhood infections (e.g., measles).
- Sexually transmitted diseases (e.g., HIV and MPox).
- Vector-mediated pathogens (e.g., dengue, Zika, and bubonic plague).

Types of data

The [companion material](#) will describe how to prepare the following types of data for comparisons with model simulations, either for [exploration](#) or [parameterization](#).

- Case, symptom, and outcome reports.
- Hospital admissions and occupancy.
- Seroprevalence.
- Published contact matrices

- Cell phone mobility data.
- Genomic data on the relative abundance of pathogen strains.
- Wastewater pathogen concentrations.
- Social attitude data (e.g., vaccine hesitancy).
- Operational and intervention schedules (e.g., vaccination supply, school closures).

Model modification tools

The [companion material](#) will describe how to make the following modifications, or to quickly find and use instructions for making them.

- Add/remove compartments or flows amongst compartments (e.g., a compartment for pathogen concentration in wastewater).
- Add behavioural responses to outbreaks.
- Filter numbers of cases to account for reporting delays, under-reporting, and weekend effects.
- Add schedules of parameter variation in response to exogeneous factors (e.g., transmission rate in response to vaccine supply, school term schedules).
- Add stochastic exogeneous factors (e.g., importation).
- Add/remove/modify vital dynamics (i.e., birth and death).

Epidemiological model summaries

The [companion material](#) will describe how to compute the following epidemiological model summaries.

- Basic reproduction number, \mathcal{R}_0 .
- Effective reproduction number, \mathcal{R}_t .
- Intrinsic per-capita rate of increase, r .
- Moments of the generation interval distribution.

Dynamical model types

The [companion material](#) will describe how to cast an existing model as one of the following types of dynamical models.

- Discrete-time recursion (i.e., Euler ODE solver).
- Runge-Kutta 4 (a more accurate ODE solver).
- Discrete-time recursion with process error (using the [Euler-multinomial distribution](#)).

Optimization

The [companion material](#) will describe how to use the following techniques to calibrate observed data using trajectory matching.

- Recover parameter values by calibrating to data simulated from the model being calibrated. This is a useful sanity check.
- Relate simulations of specific model variables to observed versions of those variables.
- Model over-dispersion using the negative-binomial distribution.
- Manipulate optimizer settings when troubleshooting lack of convergence.
- Determine whether the data provide information about a parameter of interest.
- Transform parameters to be calibrated so that the optimizer doesn't converge on impossible values (e.g. calibrating the log of the transmission rate will ensure that the transmission rate is positive).

Calibrate initial state variables

The [companion material](#) will describe how to calibrate the initial values of the state variables using the following approaches.

- Simple expressions involving the total population size, N (e.g., initializing S as N minus the initial sum of the infectious and exposed classes).
- Optimizing the initial values of selected state variables based on fit to the data (e.g., estimate numbers of individuals that were already recovered, R , by the time data collection started).
- Estimating the distribution of exposed and infectious individuals among compartments in stratified models at the beginning of an epidemic. This is useful for addressing convergence issues caused by dynamical instability.

Cartesian product models

The [companion material](#) will describe the following types of stratification with Cartesian product models.

- Stratified by age group.
- Spatially stratified such that locations are completely isolated.
- Neighborhood spatial models where compartments that are geographically close can interact.
- Stratified by vaccine status.

Stratify infectious compartments

The [companion material](#) will describe the following types of stratification of infectious compartments.

- Stratified by symptom status (e.g. mild, severe).
- Stratified by isolation status (e.g. hospitalized, ICU, self-isolation).
- Stratify to incorporate different assumptions (or information) about diagnostic test reliability.

Outcomes

After participating in the workshop, modellers will be able to do the following.

- Determine if compartmental modelling is an appropriate tool for a particular applied public health problem.
- Use `macpan2` to create a compartmental model for a real public health problem by following the [model refinement cycle](#).
- Navigate the [documentation](#) and the [workshop companion material](#) to learn how to solve compartmental modelling problems that `macpan2` is able to solve.
- Suggest improvements to `macpan2`.

Schedule

(Note: This is just a suggested schedule for now. We will likely break this up into several days. We will also explicitly target two types of users in the schedule: (1) those who want to translate a model diagram into simulations and (2) those who want to dig into the details of optimization, ODE solvers, etc.)

Time	Activity
8:00-9:00	Registration and Breakfast
9:00-9:15	Welcome and Introduction
9:15-10:30	Session 1: Exploration
10:30-10:45	Break
10:45-12:00	Session 2: Parameterization
12:00-1:00	Lunch

Time	Activity
1:00-2:15	Session 3: Inference
2:15-2:30	Break
2:30-3:45	Session 4: Stratification
3:45-4:00	Break
4:00-5:15	Session 5: Open Discussion
5:15-5:30	Wrap-up and Closing Remarks

Preparation

Required Preparation

Participants **must** take the following steps to prepare for the workshop.

- Successfully install `macpan2` using these [instructions](#).
- (Note: The final syllabus will include installation instructions for supporting packages as well – includes at least `dplyr`, `ggplot2`, `tidyr`, `broom.mixed`).
- Successfully run the `macpan2` [hello world](#) programme.

If you have any issues with these steps, please contact the [instructor](#) as soon as possible.

Suggested Preparation

Participants **should** take the following steps to prepare for the workshop.



- Sketch a state-transition diagram for a [compartmental model](#) that is of interest to you; preferably one that you use or would like to use in your own work.
- Be prepared to load in some data that are relevant for this model.
- Try using the [companion material](#) to take their model through the [model refinement cycle](#).
- Take a look at the [tidyverse](#), which is a data analysis framework and set of packages in R with which `macpan2` is designed to work.
- Install [RStudio](#) if they have not already done so, as I will be using RStudio during the workshop. But facility with another program for creating R scripts is fine.

Optional Preparation

Participants **could** take the following steps to prepare for the workshop.

- (Note: This will need to be done later.)

Footnotes

1. Although participants familiar with [R](#) will be more comfortable with the material, familiarity will not be assumed. 
2. Deep technical knowledge is not required. We will briefly review some statistical concepts, including prediction intervals, parameter estimates, likelihood functions, and prior distributions, so that everyone is on the same page. 

This site is open source. [Improve this page](#).