

## Workshop 2: calibrating a stochastic model

Danny Scarponi, Andy Iskauskas



# Contents

<b>1 Objectives</b>	<b>5</b>
<b>2 An overview of history matching with emulation and hmer</b>	<b>7</b>
<b>3 Introduction to the model</b>	<b>9</b>
<b>4 ‘wave0’ - parameter ranges, targets and design points</b>	<b>15</b>
<b>5 Emulators</b>	<b>19</b>
5.1 Brief recap on the structure of an emulator . . . . .	19
5.2 Stochastic emulators . . . . .	20
<b>6 Implausibility</b>	<b>25</b>
<b>7 Emulator diagnostics</b>	<b>27</b>
<b>8 Proposing new points</b>	<b>33</b>
<b>9 Second wave</b>	<b>37</b>
<b>10 Visualisation of the non-implausible space by wave</b>	<b>45</b>
<b>11 Dealing with bimodality</b>	<b>51</b>
11.1 ‘bimodal_wave0’ - design points . . . . .	53
11.2 Bimodal Emulators . . . . .	54
11.3 Implausibility . . . . .	58
11.4 Emulator diagnostics . . . . .	59
11.5 Proposing new points . . . . .	61
<b>A Answers</b>	<b>63</b>
<b>B Additional information</b>	<b>75</b>



# Chapter 1

## Objectives

In this workshop, you will learn how to perform history matching with emulation on stochastic models, using the `hmer` package. Much of the calibration process is the same for stochastic and deterministic models, and we therefore recommend that before starting this workshop you complete [Workshop 1](#), which shows how to perform history matching with emulation on deterministic models. In this workshop, we cover the overall process in less detail, with new information that is specific to stochastic models highlighted using boxes:

Boxes like these contain material that is specific to the calibration of stochastic models, and is therefore not covered in the deterministic workshop.

There are many reasons why we might want to use a stochastic model instead of a deterministic model. One reason is that we may be modelling small population sizes with a small initial infected population, where the question of whether a disease spreads through the population ('take-off') or is quickly recovered from without affecting the bulk of the susceptible population ('die-off') can be extremely relevant. A deterministic model cannot give statements on the likelihood of either event occurring, even though both may be possible given a fixed set of initial conditions in a stochastic model.

Another reason to use stochastic models is when some characteristic of the disease that we wish to include requires the use of an **individual-based model** (ABM), which are necessarily stochastic. For example, we may believe that patterns of repeated contact within households are crucial to modelling the disease spread, and we cannot (accurately) represent this in a deterministic model. Here we might not be interested in the individual trajectories for a single run (or *realisation*) of a stochastic model, but only the mean result when aggregated over many such realisations, but in any event we must be aware of the variability between the realisations at a given parameter set.

In this workshop, we will focus on modelling the mean response as described above, using `hmer` to allow us to accurately represent the stochastic variability in realisations. This approach is applicable in a wide variety of situations, and can be useful even if we are ultimately interested in single runs of the model. An implementation of history matching and emulation that focuses on matching to

individual realisations will appear in future versions of the package.

Note that when running the workshop code on your device, you should not expect the hmer visualisation tools to produce the same exact output as the one you can find in the following sections. This is mainly because the maximinLHS function, that you will use to define the initial parameter sets on which emulators are trained, does return different Latin Hypercube designs at each call.

Before starting the tutorial, you will need to run the code contained in the box below: it will load all relevant dependencies and a few helper functions which will be introduced and used later.

Show: Code to load relevant libraries and helper functions on P??

# Chapter 2

## An overview of history matching with emulation and hmer

A video presentation of this section can be found [here](#).

In this section we briefly recap the history matching with emulation workflow and we explain how various steps of the process can be performed using hmer functions. For a more detailed introduction to history matching with emulation in the case of stochastic models you can check [Ian Vernon's presentation](#).

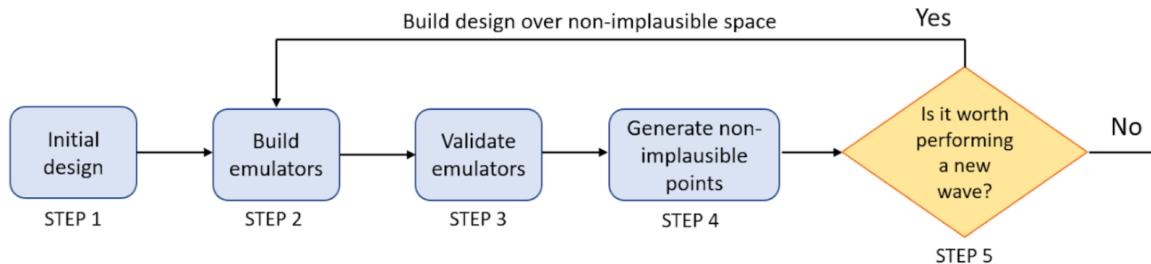


Figure 2.1: History matching with emulation workflow

As shown in the above image, history matching with emulation proceeds in the following way:

1. The model is run on the initial design points, a manageable number of parameter sets that are spread out uniformly across the input space.
2. Emulators are built using the training data (model inputs and outputs) provided by the model runs.
3. Emulators are tested, to check whether they are good approximations of the model outputs.
4. Emulators are evaluated at a large number of parameter sets. The implausibility of each of these is then assessed, and the model run using parameter sets classified as non-implausible.
5. The process is stopped or a new wave of the process is performed, going back to step 2.

The table below associates each step of the process with the corresponding hmmer function. Note that step 1 and 5 are not in the table, since they are specific to each model and calibration task. The last column of the table indicates what section of this workshop deals with each step of the process.

Step of the HME process	Hmer function			Relevant section in this workshop
	Name	Input	Output	
Build stochastic emulators (step 2)	<i>emulator_from_data</i> (setting <i>emulator_type = 'variance'</i> )	<ul style="list-style-type: none"> <li>• The training data</li> <li>• The outputs to emulate</li> <li>• The ranges of parameters</li> </ul>	An emulator of the variance and an emulator of the mean for each output of interest	4 Emulators
Validate emulators (step 3)	<i>validation_diagnostics</i>	<ul style="list-style-type: none"> <li>• The trained emulators</li> <li>• The targets to match to</li> <li>• The validation data</li> </ul>	Three diagnostics for each emulator of interest	6 Emulator diagnostics
Generate non-imausible points (step 4)	<i>generate_new_design</i>	<ul style="list-style-type: none"> <li>• The trained emulators</li> <li>• The number of points to generate</li> <li>• The targets to match to</li> </ul>	A data frame of points deemed non-imausible by all emulators	7 Proposing new points

# Chapter 3

## Introduction to the model

A video presentation of this section can be found [here](#).

In this section we introduce the model that we will work with throughout our workshop.

To facilitate the comparison between the deterministic and the stochastic setting, we will work with the SEIRS model which we used in [Workshop 1](#), but this time we introduce stochasticity. The deterministic SEIRS model used in [Workshop 1](#) was described by the following differential equations:

$$\frac{dS}{dt} = bN - \frac{\beta(t)IS}{N} + \omega R - \mu S \quad (3.1)$$

$$\frac{dE}{dt} = \frac{\beta(t)IS}{N} - \epsilon E - \mu E \quad (3.2)$$

$$\frac{dI}{dt} = \epsilon E - \gamma I - (\mu + \alpha)I \quad (3.3)$$

$$\frac{dR}{dt} = \gamma I - \omega R - \mu R \quad (3.4)$$

where  $N$  is the total population, varying over time, and the parameters are as follows:

- $b$  is the birth rate,
- $\mu$  is the rate of death from other causes,
- $\beta(t)$  is the infection rate between each infectious and susceptible individual,
- $\epsilon$  is the rate of becoming infectious after infection,
- $\alpha$  is the rate of death from the disease,
- $\gamma$  is the recovery rate and
- $\omega$  is the rate at which immunity is lost following recovery.

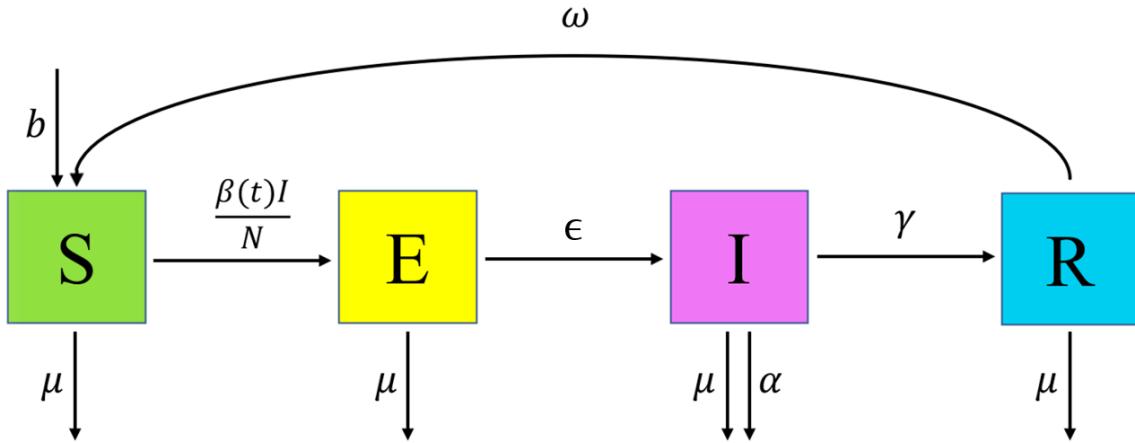


Figure 3.1: SEIRS Diagram

The rate of infection between each infectious and susceptible person  $\beta(t)$  is set to be a simple linear function interpolating between points, where the points in question are  $\beta(0) = \beta_1$ ,  $\beta(100) = \beta(180) = \beta_2$ ,  $\beta(270) = \beta_3$  and where  $\beta_2 < \beta_1 < \beta_3$ . This choice was made to represent an infection rate that initially drops due to external (social) measures and then raises when a more infectious variant appears. Here  $t$  is taken to measure days. Below we show a graph of the infection rate over time when  $\beta_1 = 0.3$ ,  $\beta_2 = 0.1$  and  $\beta_3 = 0.4$ :

In order to obtain the solutions of the stochastic SEIR model for a given set of parameters, we will use a helper function, `get_results` (which is defined in the R-script). This function assumes an initial population of 900 susceptible individuals, 100 exposed individuals, and no infectious or recovered individuals, and uses the Gillespie algorithm to generate trajectories of the model. The minimum specifications for this function are: the parameter set(s) to run the model on, a set of outputs (e.g. `c("S", "R")`), and a set of times (e.g. `c(10, 100, 150)`) that we are interested in. The default behaviour of `get_results` is to run the model 100 times on the parameter set(s) provided, but more or fewer repetitions can be obtained, using the argument `nreps`. The function `get_results` returns a data frame containing a row for each repetition at each parameter set provided, with the first nine columns showing the values of the parameters and the subsequent columns showing the requested outputs at the requested times. If `raw` is set to `TRUE`, all outputs and all times are instead returned: this is useful if we want to plot “continuous” trajectories.

As in [Workshop 1](#), the outputs obtained with the parameter set

```
chosen_params <- c(
  b = 1/(76*365),
  mu = 1/(76*365),
  beta1 = 0.214, beta2 = 0.107, beta3 = 0.428,
  epsilon = 1/7,
```

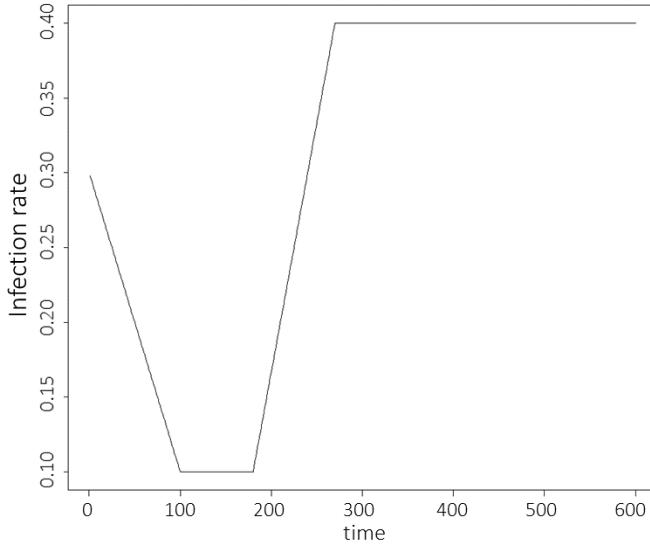


Figure 3.2: Graph of the rate of infection between each infectious and susceptible person

```

alpha = 1/50,
gamma = 1/14,
omega = 1/365
)

```

will be used to define the target bounds for our calibration task.

```

Using get_results on chosen_params
solution <- get_results(chosen_params, outs = c("I", "R"),
                        times = c(25, 40, 100, 200), raw = TRUE)

```

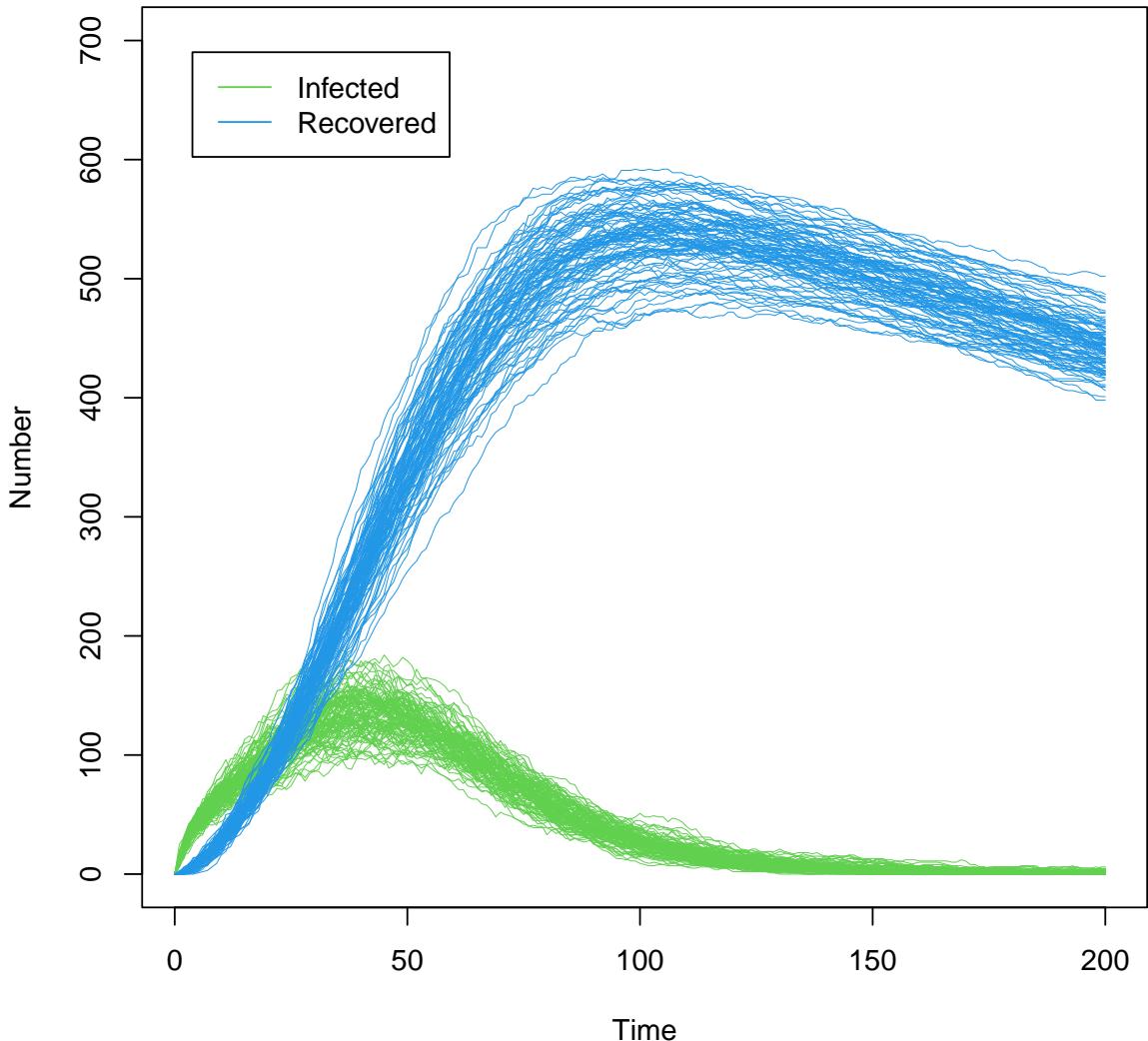
we get a 3-dimensional array `solution` such that `solution[t,j,i]` contains the number of individuals in the  $j$ -th compartment at time  $t$  for the  $i$ -th run of the model at `chosen_params`. In particular,  $t$  can take values 1, 2, ..., 201,  $j$  can take values 1, 2, 3, 4, 5 corresponding to  $S, E, I, R, D$  (where  $D$  stands for the cumulative number of deaths occurred), and  $i$  can be 1, 2, 3, ..., 100.

Plotting the results for “I” and “R”, we have

```

plot(0:200, ylim=c(0,700), ty="n", xlab = "Time", ylab = "Number")
for(j in 3:4) for(i in 1:100) lines(0:200, solution[,j,i], col=(3:4)[j-2], lwd=0.3)
legend('topleft', legend = c('Infected', "Recovered"), lty = 1,
       col = c(3,4), inset = c(0.05, 0.05))

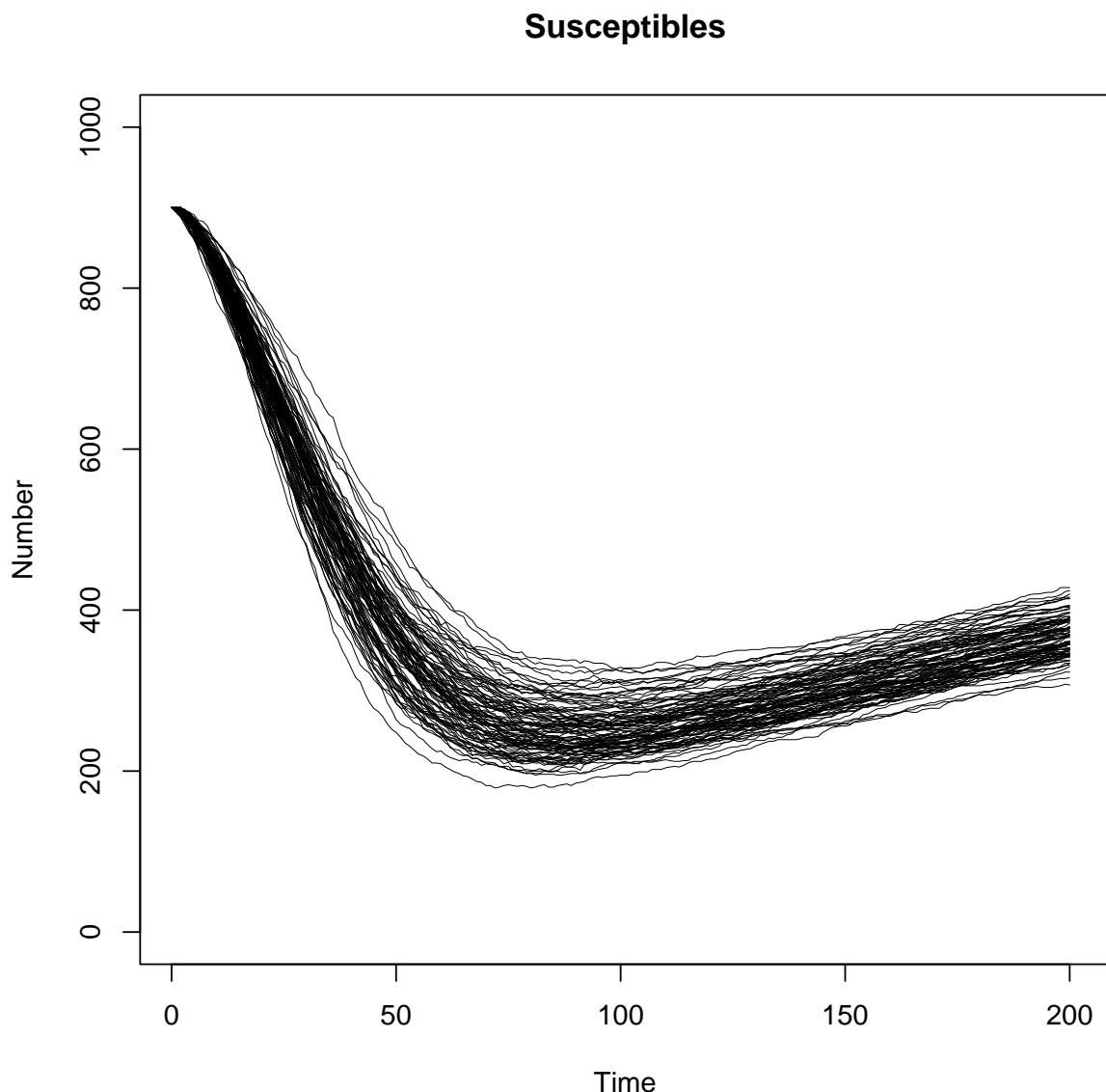
```



The plot above clearly shows the stochasticity of our model. Furthermore we can also appreciate how the spread of the runs varies across the time: for example, the variance in the number of infected individuals is largest around  $t = 40$  and quite small at later times.

Plotting the results for “S” also shows stochasticity:

```
plot(0:200, ylim=c(0,1000), ty="n", xlab = "Time", ylab = "Number", main = "Susceptibles")
for(i in 1:100) lines(0:200, solution[,1,i], col='black', lwd=0.3,
                      xlab = "Time", ylab = "Number", main = "Susceptibles")
```

**Task 1**

If you would like, familiarise yourself with the model. Investigate how the plots change as you change the values of the parameters.

Show: R tip on P77

Show: Solution on P63



# Chapter 4

## ‘wave0’ - parameter ranges, targets and design points

A video presentation of this section can be found [here](#).

In this section we set up the emulation task, defining the input parameter ranges, the calibration targets and all the data necessary to build the first wave of emulators.

Note that, for the sake of clarity, in this workshop we will adopt the word ‘data’ only when referring to the set of runs of the model that are used to train emulators. Empirical observations, that would inform our choice of targets if we were modelling a real-world scenario, will instead be referred to as ‘observations’.

First of all, let us set the parameter ranges:

```
ranges = list(  
    b = c(1e-5, 1e-4), # birth rate  
    mu = c(1e-5, 1e-4), # rate of death from other causes  
    beta1 = c(0.05, 0.3), # infection rate at time t=0  
    beta2 = c(0.1, 0.2), # infection rates at time t=100  
    beta3 = c(0.3, 0.5), # infection rates at time t=270  
    epsilon = c(0.01, 0.21), # rate of becoming infectious after infection  
    alpha = c(0.01, 0.025), # rate of death from the disease  
    gamma = c(0.01, 0.08), # recovery rate  
    omega = c(0.002, 0.004) # rate at which immunity is lost following recovery  
)
```

We then turn to the targets we will match: the number of infectious individuals  $I$  and the number of recovered individuals  $R$  at times  $t = 25, 40, 100, 200$ . The targets can be specified by a pair (val, sigma), where ‘val’ represents the measured value of the output and ‘sigma’ represents its standard deviation, or by a pair (min, max), where min represents the lower bound and max the upper bound for the target. Since in [Workshop 1](#) we used the former formulation, here we will show the latter:

```
targets <- list(
  I25 = c(98.51, 133.25),
  I40 = c(117.17, 158.51),
  I100 = c(22.39, 30.29),
  I200 = c(0.578, 0.782),
  R25 = c(106.34, 143.9),
  R40 = c(218.28, 295.32),
  R100 = c(458.14, 619.84),
  R200 = c(377.6, 510.86)
)
```

The ‘sigmas’ in our `targets` list represent the uncertainty we have about the observations. Note that in general we can also choose to include model uncertainty in the ‘sigmas’ to reflect how accurate we think our model is.

For this workshop, we generate parameter sets using a [Latin Hypercube](#) design. A rule of thumb is to select at least  $10p$  parameter sets to train emulators, where  $p$  is the number of parameters (9 in this workshop). Using the function `maximinLHS` in the package `lhs`, we create a hypercube design with 100 parameter sets for the training set and one with 50 parameter sets for the validation set. Note that we could have chosen 100 parameter sets for the validation set: we chose 50 simply to highlight that it is not necessary to have as many validation points as training points. Especially when working with expensive models, this can be useful and can help us manage the computational burden of the procedure. After creating the two hypercube designs, we bind them together to create `initial_LHS`:

```
initial_LHS_training <- maximinLHS(100, 9)
initial_LHS_validation <- maximinLHS(50, 9)
initial_LHS <- rbind(initial_LHS_training, initial_LHS_validation)
```

Note that in `initial_LHS` each parameter is distributed on  $[0, 1]$ . This is not exactly what we need, since each parameter has a different range. We therefore re-scale each component in `initial_LHS` multiplying it by the difference between the upper and lower bounds of the range of the corresponding parameter and then we add the lower bound for that parameter. In this way we obtain `initial_points`, which contains parameter values in the correct ranges.

```
initial_points <- setNames(data.frame(t(apply(initial_LHS, 1,
                                               function(x) x * purrr::map_dbl(ranges, diff) +
                                                 purrr::map_dbl(ranges, ~ .[[1]]))), names(ranges))
```

We then run the model for the parameter sets in `initial_points` through the `get_results` function, specifying that we are interested in the outputs for  $I$  and  $R$  at times  $t = 25, 40, 100, 200$ . Note that we use the `progressr` package to create a progress bar that updates every time a new parameter set is run: when you run the code on your own, this will indicate at what stage of the process you are.

```
initial_results <- list()
with_progress({
  p <- progressor(nrow(initial_points))
  for (i in 1:nrow(initial_points)) {
```

```

model_out <- get_results(unlist(initial_points[i,]), nreps = 25, outs = c("I", "R"),
                         times = c(25, 40, 100, 200))
initial_results[[i]] <- model_out
p(message = sprintf("Run %g", i))
}
})

```

Note that `initial_results` is a list of length 3750, since it has a row for each of the 25 repetitions of the 150 parameter sets in `initial_points`. Finally, we bind all elements in `initial_results` to obtain a data frame `wave0` and we split it in two parts: the first 2500 elements (corresponding to the first 100 parameter sets) for the training set, `all_training`, and the last 1250 for the validation set (corresponding to the last 50 parameter sets), `all_valid`.

```

wave0 <- data.frame(do.call('rbind', initial_results))
all_training <- wave0[1:2500,]
all_valid <- wave0[2501:3750,]
output_names <- c("I25", "I40", "I100", "I200", "R25", "R40", "R100", "R200")

```



# Chapter 5

## Emulators

A video presentation of this section can be found [here](#).

In this section we will train stochastic emulators, which take into account the fact that each time the model is run using the same parameter set, it will generate different output.

### 5.1 Brief recap on the structure of an emulator

We very briefly recap the general structure of a univariate emulator (see [Workshop 1](#) for more details):

$$f(x) = g(x)^T \xi + u(x),$$

where  $g(x)^T \xi$  is a regression term and  $u(x)$  is a [weakly stationary process](#) with mean zero.

The regression term, which mimics the global behaviour of the model output, is specified by a vector of functions of the parameters  $g(x)$  which determine the shape and complexity of the regression hypersurface we fit to the training data, and a vector of regression coefficients  $\xi$ .

The [weakly stationary process](#)  $u(x)$  (similar to a [Gaussian process](#)), accounts for the local deviations (or residuals) of the output from the regression hypersurface.

In this workshop,  $u(x)$  is assumed to be a Gaussian process, with covariance structure given by

$$\text{Cov}(u(x), u(x')) = \sigma^2 c(x, x')$$

where  $c$  is the square-exponential correlation function

$$c(x, x') := \exp\left(\frac{-\sum_i (x_i - x'_i)^2}{\theta^2}\right)$$

where  $x_i$  is the  $i$ th-component of the parameter set  $x$ . The term  $\sigma^2$  is the **prior emulator variance**, i.e. the variance of  $u(x)$ , and reflects how far we expect the output to be from the regression hypersurface, while  $\theta$  is the **correlation length** of the process, and determines how close two parameter sets must be in order for the corresponding residuals to be non-negligibly correlated.

In the deterministic workshop, we said that in order to train an emulator, the `emulator_from_data` function first estimated the value for  $\sigma$ , using the provided training data. It is important to note that **this estimated value of  $\sigma$  was constant**, i.e., it was the same for all parameter sets. When working with stochastic models, different areas of the parameter space often give rise to quite different levels of stochasticity in the outputs. Taking this phenomenon into account when dealing with stochastic models will allow us to train more accurate and efficient emulators.

## 5.2 Stochastic emulators

To train stochastic emulators we use the function `emulator_from_data` specifying `emulator_type = 'variance'`, which requires the training data, the names of the outputs to emulate, and the ranges of the parameters:

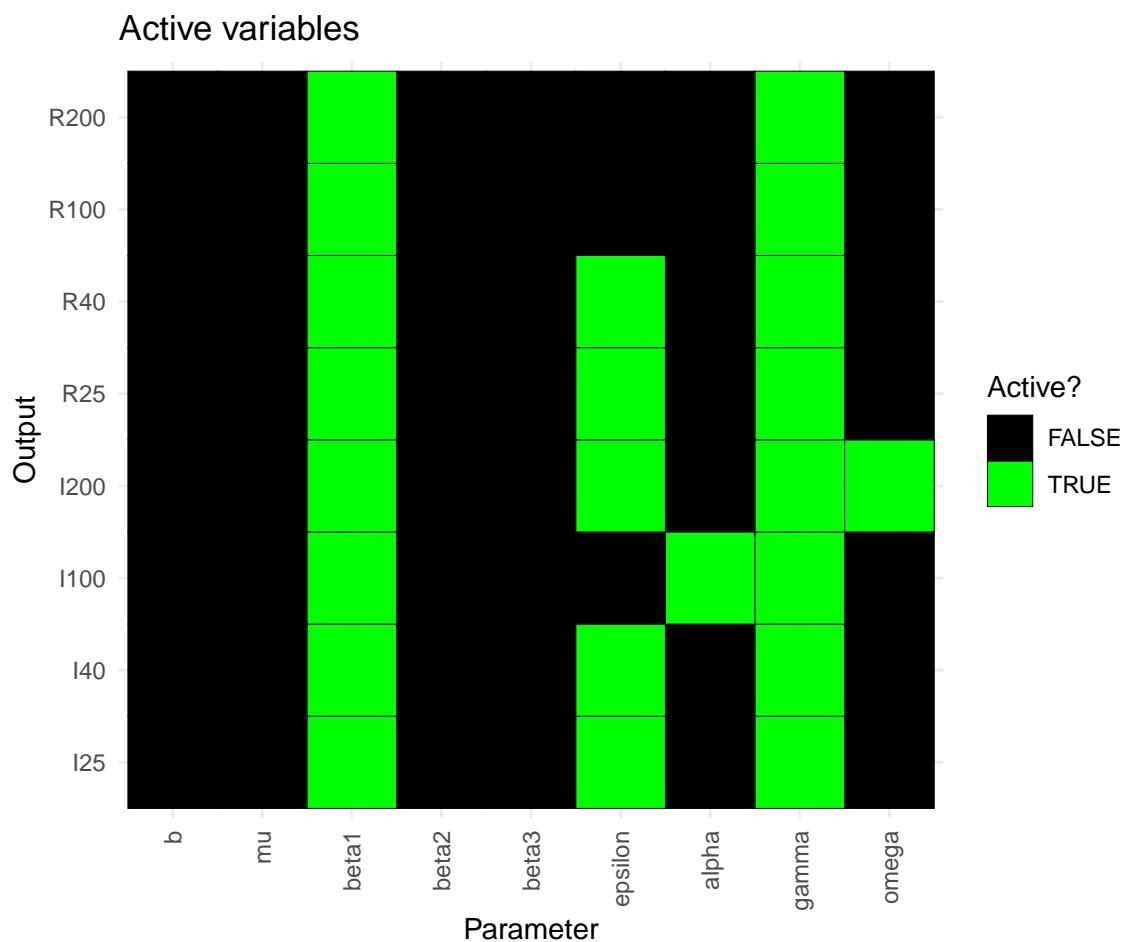
```
stoch_emulators <- emulator_from_data(all_training, output_names, ranges,
                                         emulator_type = 'variance')
```

The function `emulator_from_data` with `emulator_type = 'variance'` returns two sets of emulators: one for the variance and one for the expectation of each model output. Behind the scenes, `emulator_from_data` does the following:

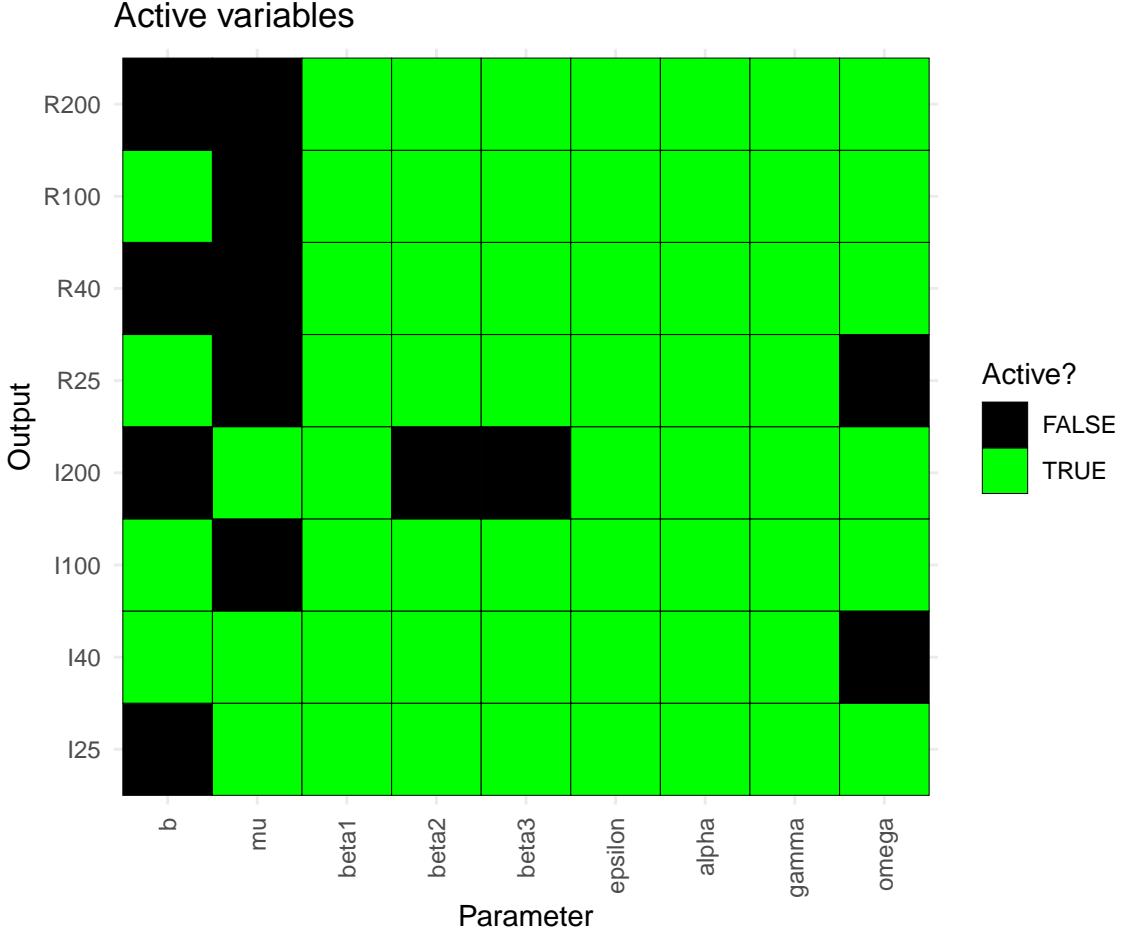
- First, it estimates the variance of each model output of interest at each provided parameter set: this is possible since `all_training` contains several model runs at each parameter set.
- Using the obtained variance data, it then trains an emulator for the variance of each of the outputs of interest. These emulators will be referred to as **variance emulators** and can be accessed by typing `stoch_emulators$variance`.
- Finally, emulators for the mean of each output are built. These emulators will be referred to as mean emulators and can be accessed by typing `stoch_emulators$expectation`. The main difference in the training of emulators between the stochastic and the deterministic case is that here, for each output, we use the variance emulator to inform our choice for a component of  $\sigma$ , that which is due to the stochasticity of the model: for each parameter set  $x$ , the variance emulator gives us an estimate of the variance of the model output at  $x$ . Note that this step is specific to the treatment of stochastic models, where it is important to account for the fact that different areas of the parameter space often give rise to quite different levels of stochasticity in the outputs.

Let's take a look at the two sets of emulators stored in `stoch_emulators`, starting with plots of active variables:

```
plot_actives(stoch_emulators$variance)
```



```
plot_actives(stoch_emulators$expectation)
```



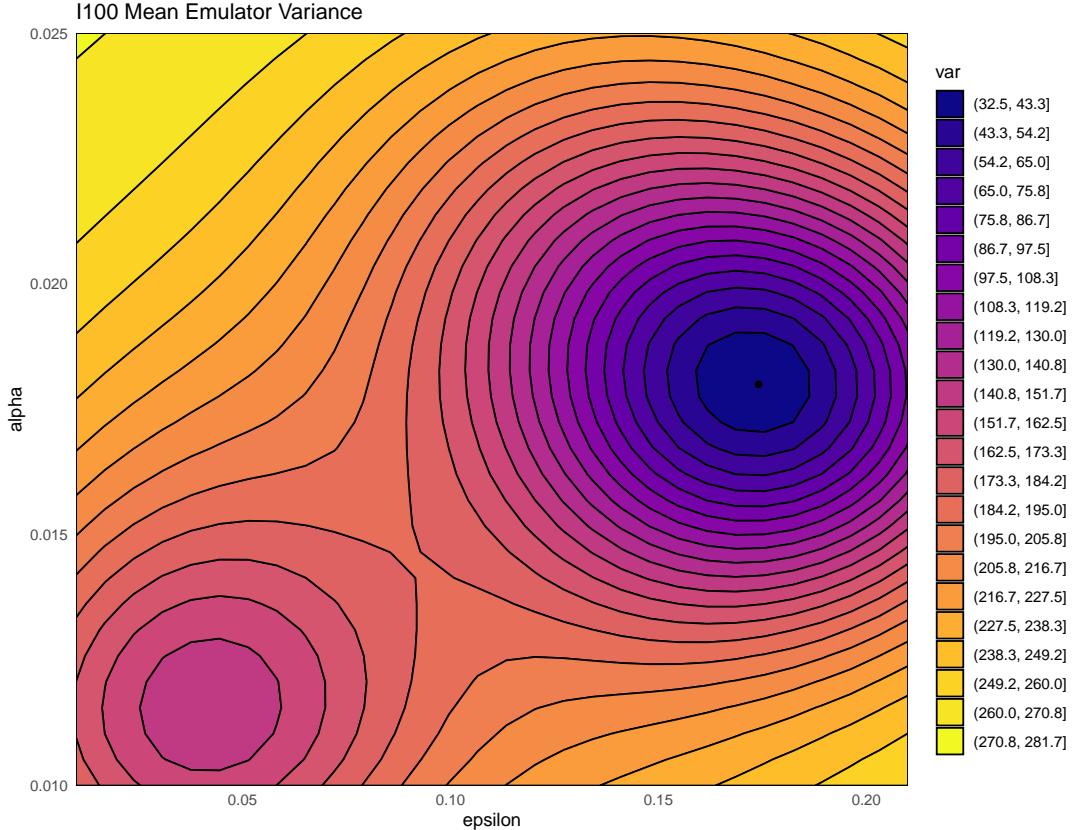
These plots show what variables are active, i.e. have the most explanatory power, for each of the mean and variance emulators. We see that  $b$  and  $\mu$  are inactive for almost all outputs in the variance emulators, even though they are active for some outputs in the mean emulators. Parameters  $\beta_1$ ,  $\epsilon$  and  $\gamma$  are active for most or all outputs in the mean emulators.

As in the deterministic case, the `emulator_plot` can be used to visualise how emulators represent the output space. Since we have two sets of emulators, we have several options in terms of plots. The default behaviour of `emulator_plot` is to plot the expectation of the passed emulator: for example, to plot the expectation of the variance emulator for  $I_{100}$ , we just need to pass `stoch_emulators$variance$I100` to `emulator_plot`.

It is important to stress the following point: unlike in the deterministic case, here mean emulators do not have zero variance at ‘known’ points, i.e., at points in the training set. Let us verify this, by plotting the variance of the mean emulator for  $I_{100}$  in the  $(\epsilon, \alpha)$ -plane, with all unshown parameters to be as in the first row of `all_training`. Note that since the plot produced below depends on the location of the first run, when running the code with

your own `all_training` you will get a different result:

```
emulator_plot(stoch_emulators$expectation$I100, params = c('epsilon', 'alpha'),
  fixed_vals = all_training[1, names(ranges)[-c(6,7)]], plot_type = 'var') +
  geom_point(data = all_training[1,], aes(x = epsilon, y = alpha))
```



We see that the black point in the dark blue area, corresponding to the first point in `all_training`, does not have variance zero. This is because the mean emulator is trying to predict the ‘true’ mean at the point, but it has only been given a sample mean, which it cannot assume is equal to the ‘true’ mean. The emulator internally compensates for this incomplete information, resulting in the variance not being zero, even at points in the training set. This compensation also means that we can work with different numbers of replicates at different parameter sets. In general, for each training/validation parameter set, we should use as many repetitions as is feasible given the model complexity. More repetitions will provide a better estimate of the variance, which in turn allows to train more accurate mean emulators. If a model is relatively fast, then it is worth doing 50-100 repetitions per parameter set; if it is slower, then we can work with fewer repetitions (even just 5-10 per parameter set). This applies to different regions of the parameter space within a single model, too: if one part

of the parameter space is much slower to run (e.g. parameter sets with higher beta values, which tend to produce more infections, slowing down agent-based models), then we could run fewer repetitions in those parts of parameter space.

# Chapter 6

## Implausibility

A video presentation of this section can be found [here](#).

We will first briefly review the definition of an implausibility measure.

For a given stochastic model output and a given target, the implausibility measures the difference between the output and the target, taking into account all sources of uncertainty. For a parameter set  $x$ , a particular choice of implausibility, useful for either matching to the mean of the stochastic model, or its spread of possible values, is:

$$\text{Imp}(x) = \frac{|\text{E}[m(x)] - z|}{\sqrt{V_0 + V_c(x) + V_s(x) + V_m}},$$

where  $\text{E}[m(x)]$  is the expected value of the mean of the stochastic model,  $z$  the target, and the terms in the denominator refer to various forms of uncertainty. In particular

- $V_0$  is the variance associated with the observation uncertainty;
- $V_c(x)$  refers to the uncertainty one introduces when using the emulator output instead of the model output itself;
- $V_s(x)$  is the ‘ensemble variability’ and represents the stochastic nature of the model. This term is used to create an appropriate representation of the uncertainty on either the mean or the set of possible realisations of the stochastic model.
- $V_m$  is the model discrepancy, accounting for possible mismatches between the model and reality.

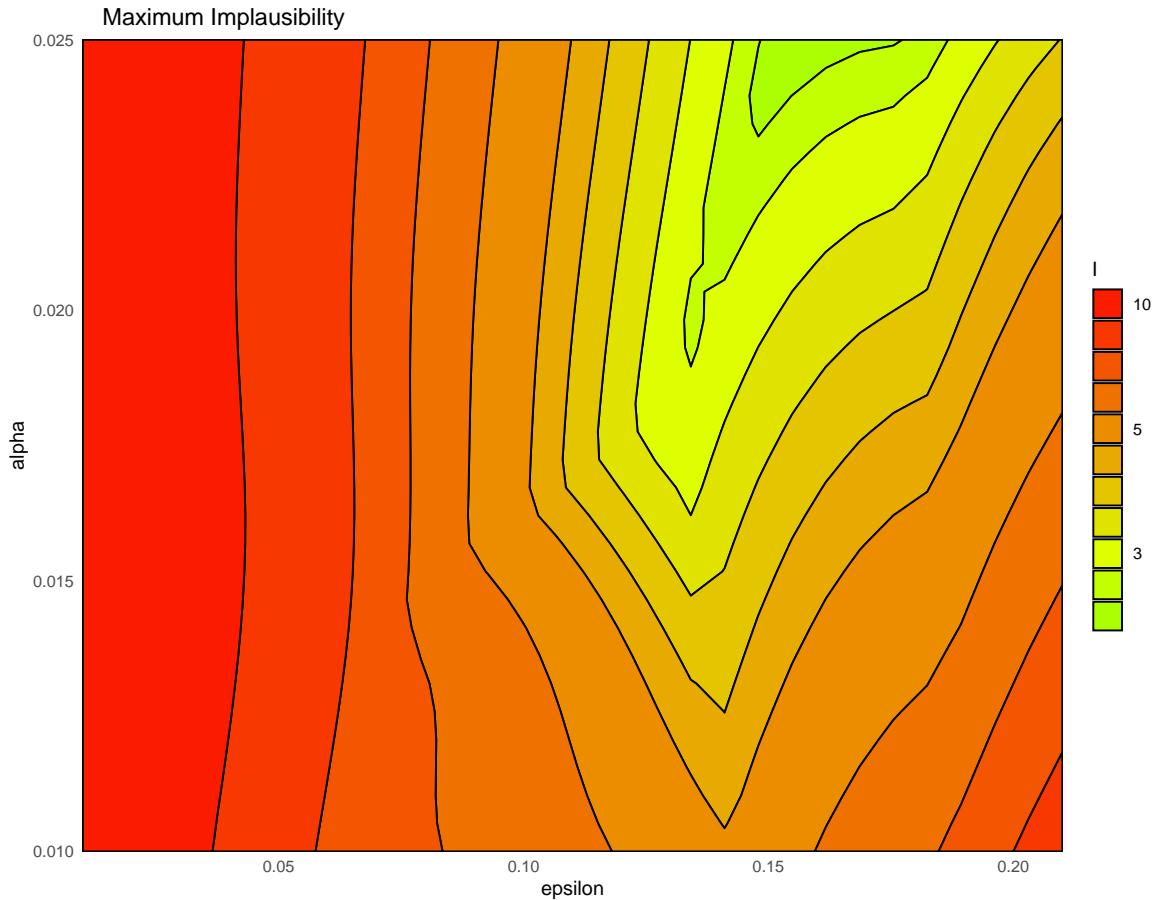
Since in this case study we want to emulate our model, without reference to a real-life analogue, we claim that the model represents the reality perfectly. For this reason we have  $V_m = 0$ . The term  $V_s(x)$ , which was zero when we worked on the deterministic SEIRS model ([Workshop 1](#)), is now non-zero, in order to account for the stochasticity of the model.

A very large value of  $\text{Imp}(x)$  means that we can be confident that the parameter set  $x$  does not provide a good match to the observed data, even factoring in the additional uncertainty that comes with the use of emulators. Conversely, a small value can mean two things: either the predicted mean

is close to the observed value, suggesting that the parameter set may provide a good fit to data; or the collected uncertainties (including the ensemble or stochastic variability) is large, suggesting that exploration of this region of parameter space would be informative.

To plot the maximum implausibility (across all emulators), we set `plot_type` to `nimp` in the call to `emulator_plot`:

```
emulator_plot(stoch_emulators, plot_type = 'nimp',
              targets = targets, params = c('epsilon', 'alpha'))
```



### Task 2

Use the argument `fixed_vals` to set the parameters that are not shown in the plot to be as in `chosen_params`. Verify that the implausibility at `chosen_params` is below 3.

Show: Solution on P68

# Chapter 7

## Emulator diagnostics

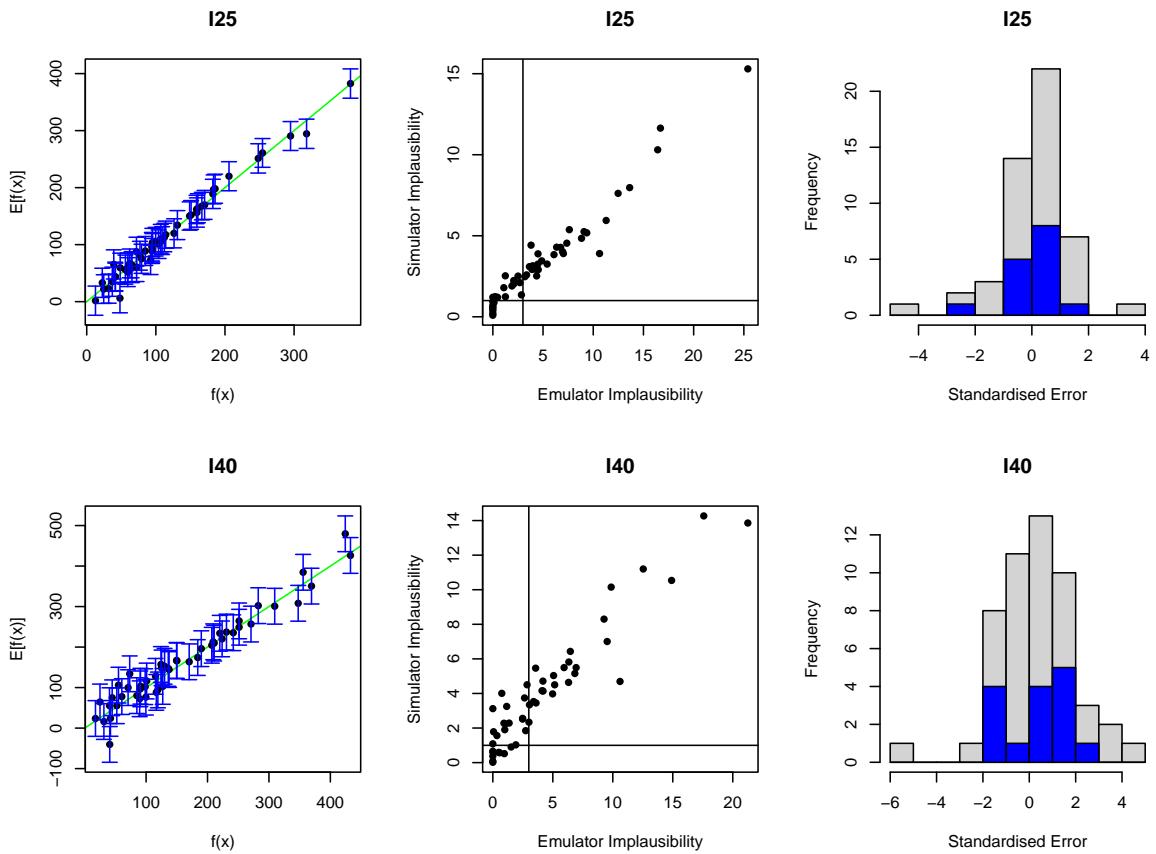
A video presentation of this section can be found [here](#).

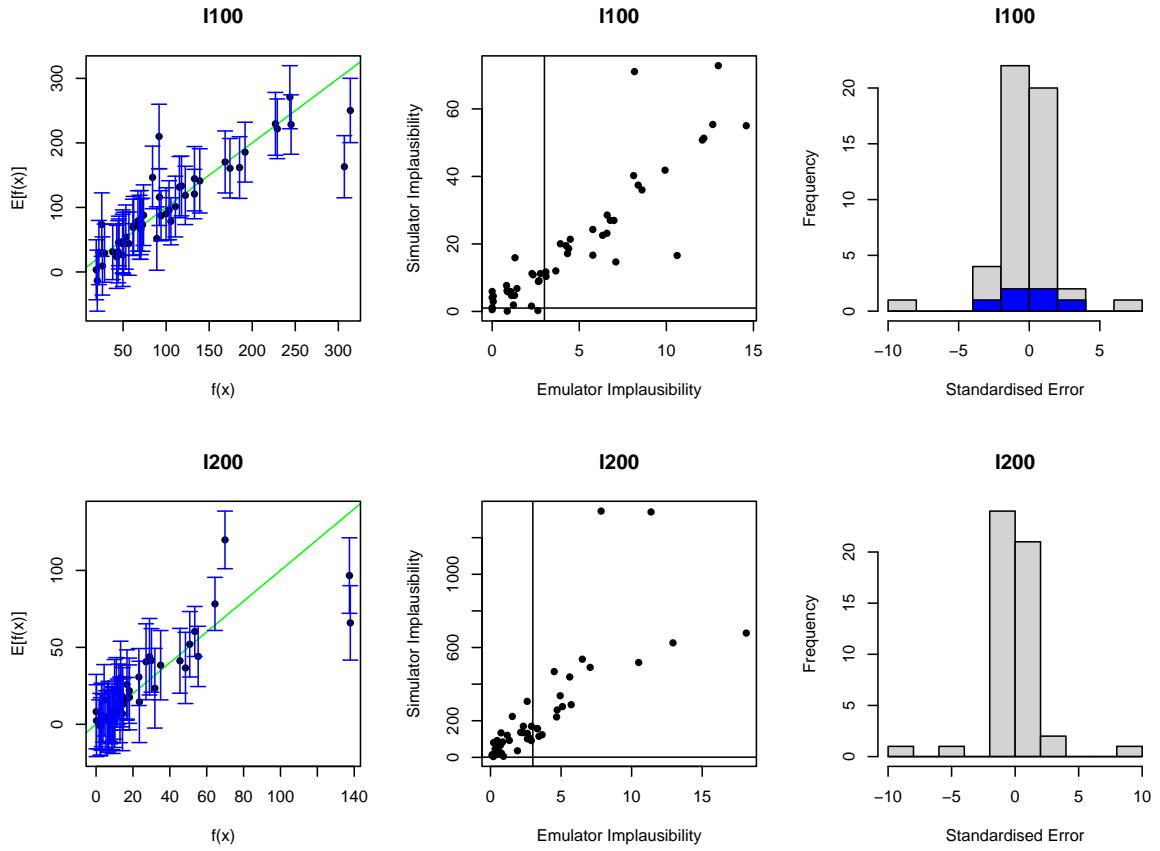
The function `validation_diagnostics` can be used as in the deterministic case, to get three diagnostics for each emulated output.

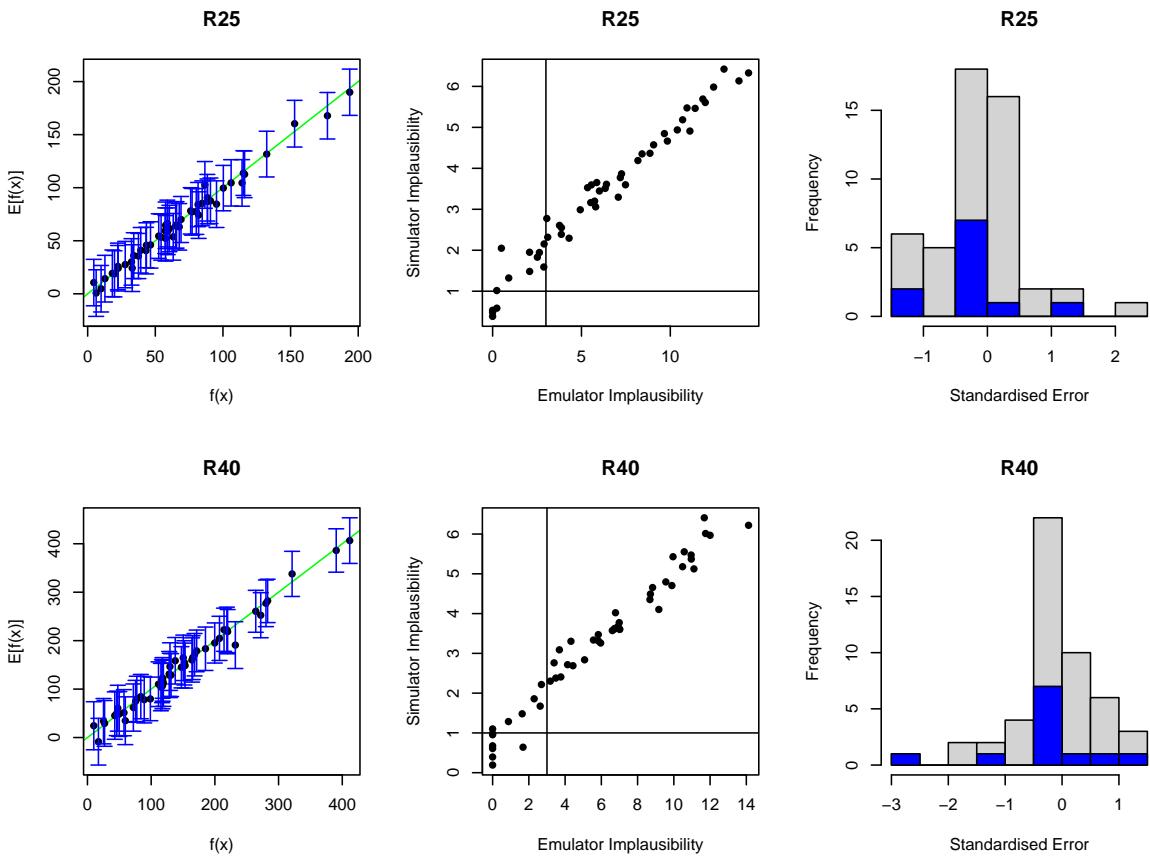
Show: Remind me of what each diagnostic means on P77

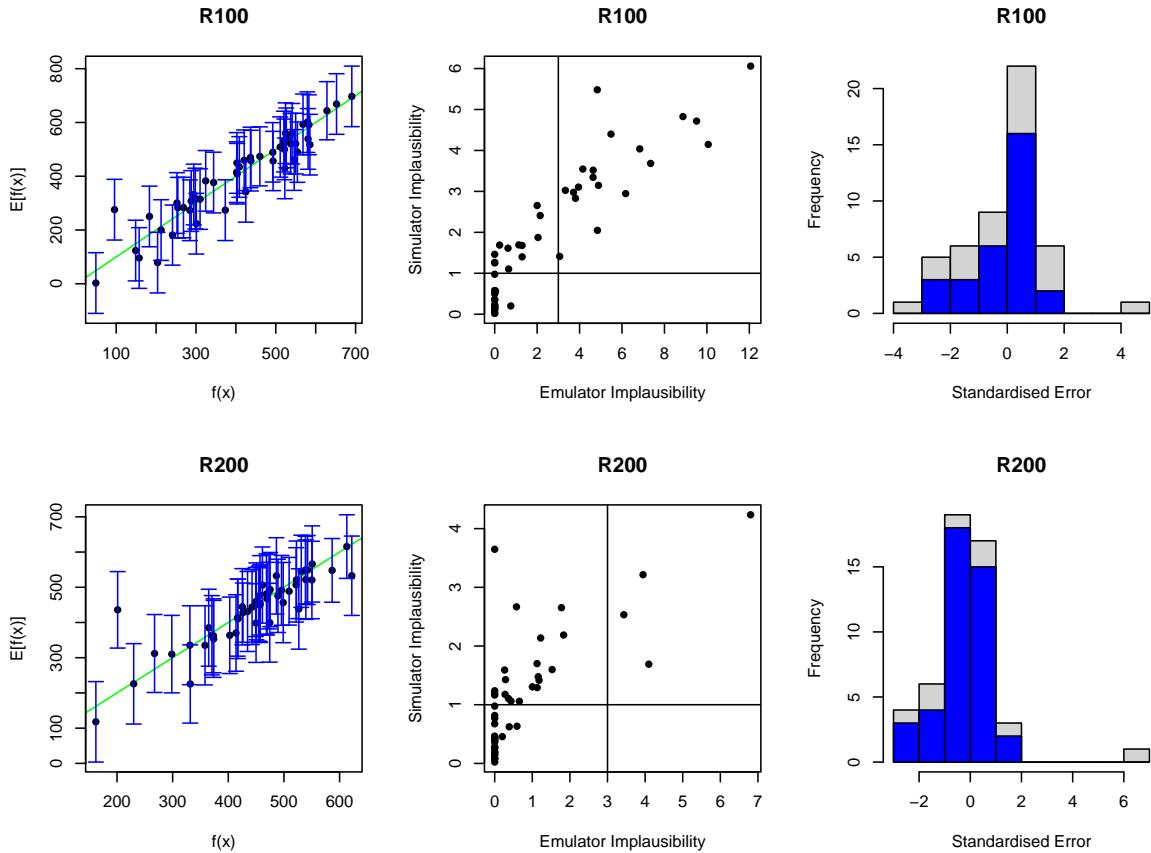
We call `validation_diagnostics` passing it the mean emulators, the list of targets, and the validation set:

```
vd <- validation_diagnostics(stoch_emulators$expectation, targets, all_valid, plt=TRUE, row=2)
```









Note that by default `validation_diagnostics` groups the outputs in sets of three. Since here we have 8 outputs, we set the argument `row` to 2: this is not strictly necessary but improves the format of the grid of plots obtained. As in deterministic case, we can enlarge the  $\sigma$  values of the mean emulators to obtain more conservative emulators, if needed. Based on the diagnostics above, all emulators perform quite well, so we won't modify any  $\sigma$  values here. In order to access a specific mean emulator, you can type `ems$expectation$variable_name`. For example, to double the  $\sigma$  of the mean emulator for `I25`, you would type `stoch_emulators$expectation$I25 <- stoch_emulators$expectation$I25$mult_sigma(2)`.

A more in-depth discussion of diagnostics and how to improve problematic emulators can be found in [Workshop 1](#); note, however, that the `diagnostic_pass()` function is not currently supported for stochastic emulation. We can, of course, automate some aspects of the diagnostic process, which we detail below.

Note that the output `vd` of the `validation_diagnostics` function is a data frame containing all parameter sets in the validation dataset that fail at least one of the three diagnostics. This data frame can be used to automate the validation step of the history matching process. For example, one can consider the diagnostic check to be successful if `vd` contains at most 5% of all points in the validation dataset. In this way, the user is not required to manually inspect each validation diagnostic for each emulator at each wave of the process.

It is possible to automate the validation process. If we have a list of trained emulators `ems`, we can start with an iterative process to ensure that emulators do not produce misclassifications:

1. check if there are misclassifications for each emulator (middle column diagnostic in plot above) with the function `classification_diag`;
2. in case of misclassifications, increase the `sigma` (say by 10%) and go to step 1.

The code below implements this approach:

```
for (j in 1:length(ems)) {
  misclass <- nrow(classification_diag(ems$expectation[[j]], targets, validation, plt = FALSE))
  while(misclass > 0) {
    ems$expectation[[j]] <- ems$expectation[[j]]$mult_sigma(1.1)
    misclass <- nrow(classification_diag(ems$expectation[[j]], targets, validation, plt = FALSE))
  }
}
```

The step above helps us ensure that our emulators are not overconfident and do not rule out parameter sets that give a match with the empirical data.

Once misclassifications have been eliminated, the second step is to ensure that the emulators' predictions agree, within tolerance given by their uncertainty, with the simulator output. We can check how many validation points fail the first of the diagnostics (left column in plot above) with the function `comparison_diag`, and discard an emulator if it produces too many failures. The code below implements this, removing emulators for which more than 10% of validation points do not pass the first diagnostic:

```
bad.ems <- c()
for (j in 1:length(ems)) {
  bad.model <- nrow(comparison_diag(ems$expectation[[j]], targets, validation, plt = FALSE))
  if (bad.model > floor(nrow(validation)/10)) {
    bad.ems <- c(bad.ems, j)
  }
}
ems <- subset_emulators(ems, names(targets)[!seq_along(ems) %in% bad.ems])
```

Note that the code provided above gives just an example of how one can automate the validation process and can be adjusted to produce a more or less conservative approach. Furthermore, it does not check for more subtle aspects, e.g. whether an emulator systematically underestimates/overestimates the corresponding model output. For this reason, even though automation can be a useful tool (e.g. if we are running several calibration processes and/or have a long list of targets), we should not think of it as a full replacement for a careful, in-depth inspection from an expert.

# Chapter 8

## Proposing new points

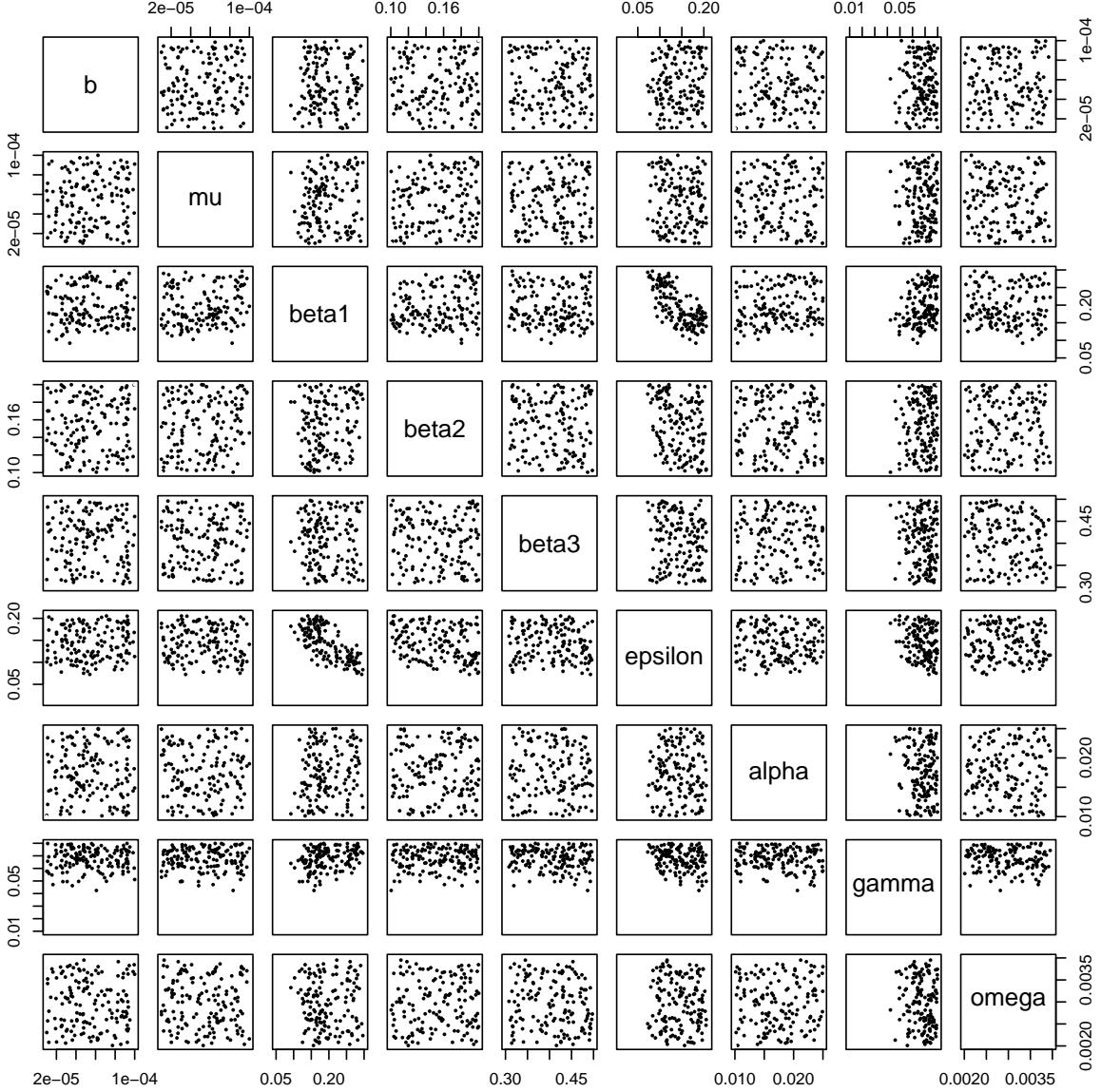
A video presentation of this section can be found [here](#).

To generate a set of non-implausible points, based on the trained emulators, we use the function `generate_new_design`, exactly as in the deterministic case:

```
new_points <- generate_new_design(stoch_emulators, 150, targets)
```

Here we generated 150 points, since we are going to use 100 of them to train new emulators and 50 of them to validate new emulators (as done in the first wave). We can visualise the non-implausible space at the end of this first wave using `plot_wrap`:

```
plot_wrap(new_points, ranges)
```



Here we see which parameters are more or less constrained at the end of the first wave. For example, the penultimate column suggests that low values of  $\gamma$  cannot produce a match. We can also deduce relationships between parameters:  $\beta_1$  and  $\epsilon$  are an example of negatively-correlated parameters. If  $\beta_1$  is large then  $\epsilon$  needs to be small, and vice versa. Other parameters, such as  $\omega$  or  $\mu$ , are instead still spread out across their initial range.

As shown in [Tutorial 2](#), it is also possible to perform a full wave of emulation and history matching using the function `full_wave`, which needs the following information:

- A dataset that will be split by the function into training data and test data;

- A list of ranges for the parameters;
- The targets: for each of the model outputs to emulate, we need a pair (val, sigma) or (min, max) that will be used to evaluate implausibility.



# Chapter 9

## Second wave

A video presentation of this section can be found [here](#).

To perform a second wave of history matching and emulation we follow the same procedure as in the previous sections, with two caveats. First, since the parameter sets in `new_points` tend to lie in a small region inside the original input space, we will train the new emulators only on the non-imausible region found in wave one. This can be done simply setting the argument `check.ranges` to TRUE in the function `emulator_from_data`.

It is often a good strategy to increase the number of repetitions used at each wave: this, paired with the fact later waves emulators are trained on a reduced space, allows to create more and more precise emulators. The rationale behind this is that, as the emulators rule out space, at each wave we obtain a more stable parameter space to explore. This more stable space is likely to ‘behave’ and less likely to result in long model runs (which is common in stochastic models where corners of space are evaluated, often resulting in extremely large numbers of infected people). We can therefore perform more realisations at each parameter combination for the same computational budget as earlier waves.

Using this principle, we run the model 50 times on each parameter set in `new_points` (for the first wave we used 25 repetitions). We then bind all results to create a dataframe `wave1` and we split this into two subsets, `new_all_training` to train the new emulators and `new_all_valid` to validate them. We select the first 5000 rows (corresponding to the first 100 parameter sets in `new_points`) for the training data and the last 2500 (corresponding to the last 50 parameter sets in `new_points`) for the validation data.

```
new_results <- list()
with_progress({
  p <- progressor(nrow(initial_points))
  for (i in 1:nrow(new_points)) {
    model_out <- get_results(unlist(new_points[i,]), nreps = 50, outs = c("I", "R"),
                             times = c(25, 40, 100, 200))
```

```

new_results[[i]] <- model_out
p(message = sprintf("Run %g", i))
})
})
wave1 <- data.frame(do.call('rbind', new_results))
new_all_training <- wave1[1:5000,]
new_all_valid <- wave1[5001:7500,]

```

To train new emulators we use `emulator_from_data`, passing the new training data, the outputs names and the initial ranges of the parameters. We also set `check.ranges` to TRUE to ensure that the new emulators are trained only on the non-imausible region found in wave one, and `emulator_type = 'variance'` to indicate that we want to train stochastic emulators.

```

new_stoch_emulators <- emulator_from_data(new_all_training, output_names, ranges,
                                             check.ranges=TRUE, emulator_type = 'variance')

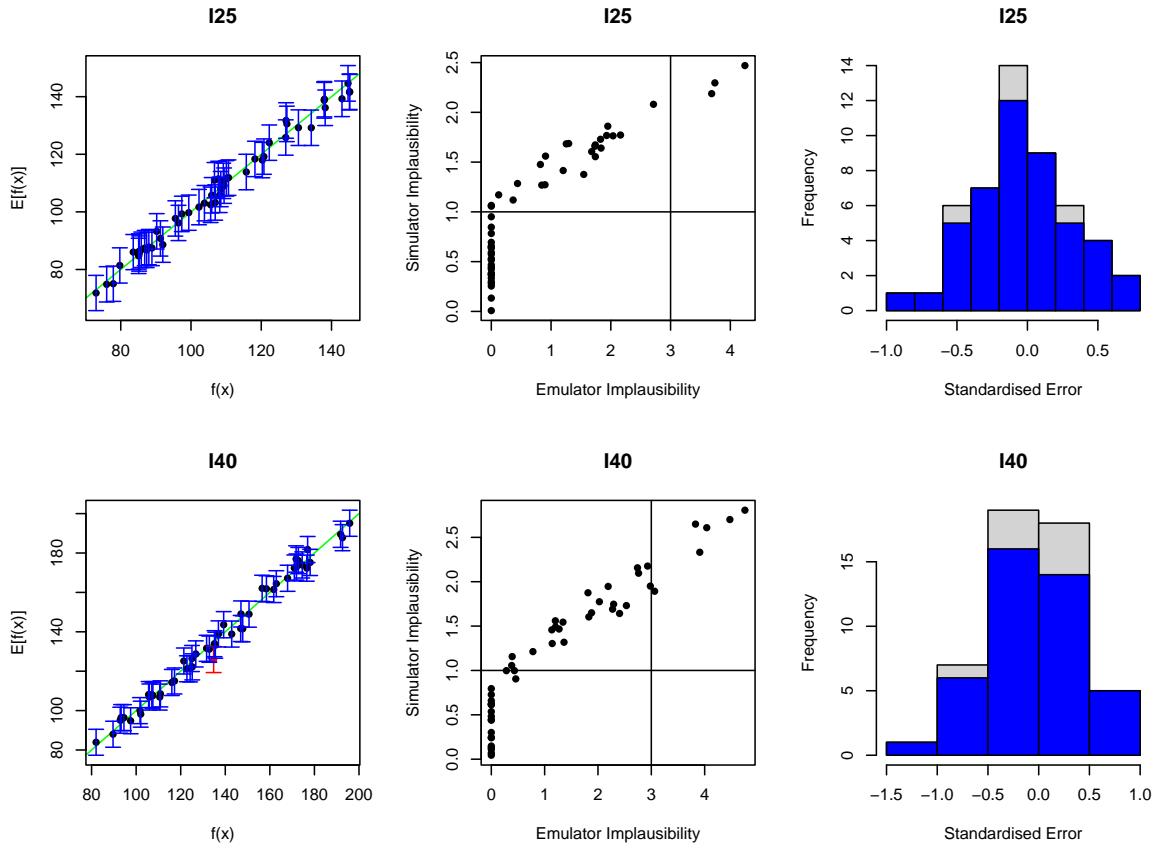
```

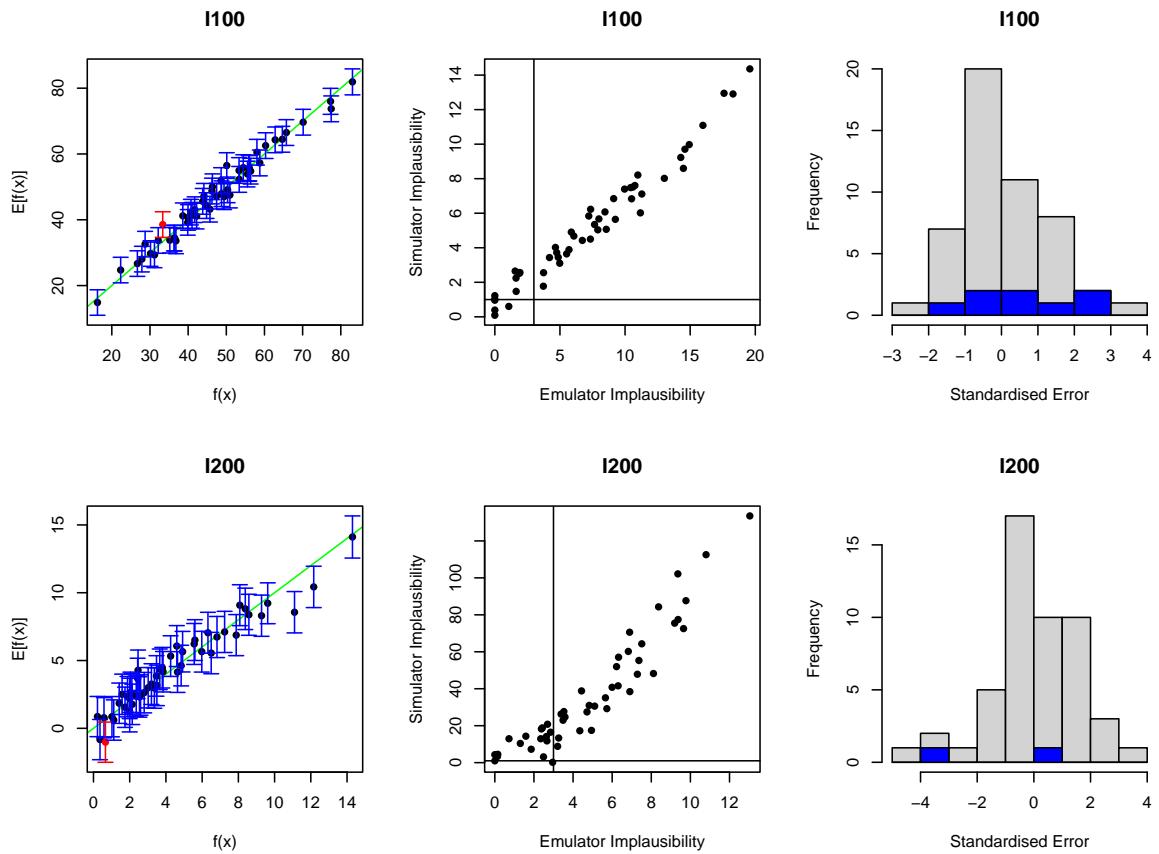
As usual, before using the obtained emulators, we need to check their diagnostics. We use the function `validation_diagnostics` which takes the new emulators, the targets and the new validation data:

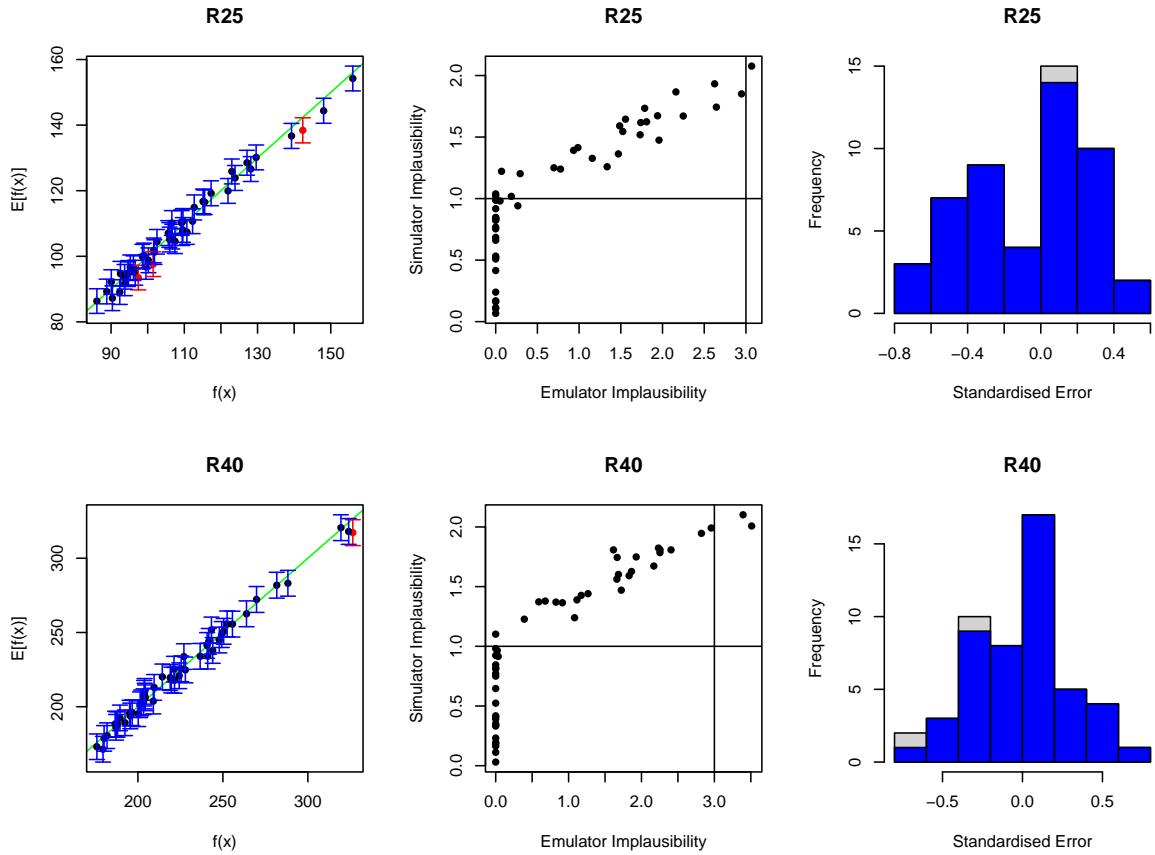
```

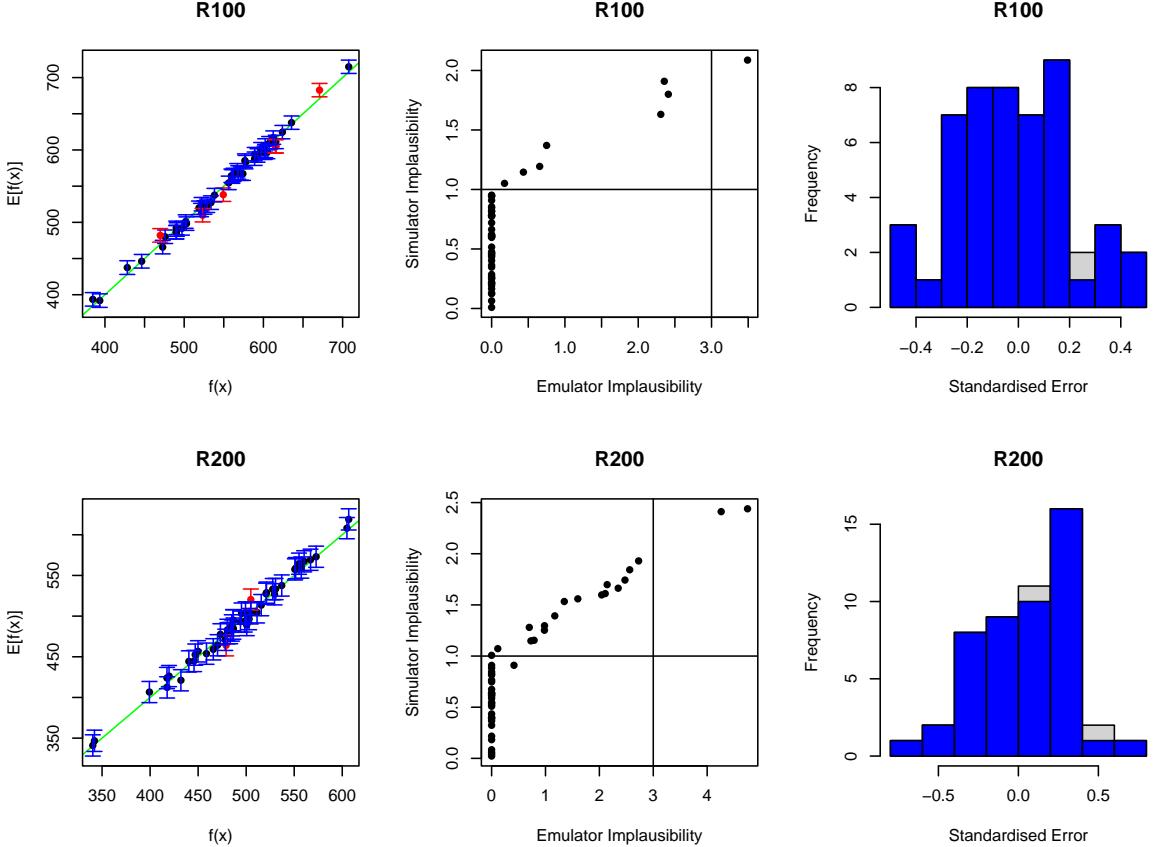
vd <- validation_diagnostics(new_stoch_emulators, targets, new_all_valid, plt=TRUE, row=2)

```









Since these diagnostics look good, we can generate new non-implausible points. Here is the second caveat: we now need to pass both `new_stoch_emulators` and `stoch_emulators` to the `generate_new_design` function, since a point needs to be non-implausible for all emulators trained so far, and not just for emulators trained in the current wave:

```
new_new_points <- generate_new_design(c(new_stoch_emulators, stoch_emulators), 150, targets)
```

Below we run the model on the points provided by the second wave of history matching:

```
new_new_results <- list()
with_progress({
  p <- progressor(nrow(initial_points))
  for (i in 1:nrow(new_new_points)) {
    model_out <- get_results(unlist(new_new_points[i,]), nreps = 100, outs = c("I", "R"),
                             times = c(25, 40, 100, 200))
    new_new_results[[i]] <- model_out
    p(message = sprintf("Run %g", i))
  }
})
wave2 <- data.frame(do.call('rbind', new_new_results))
```

```
new_new_all_training <- wave2[1:10000,]
```



# Chapter 10

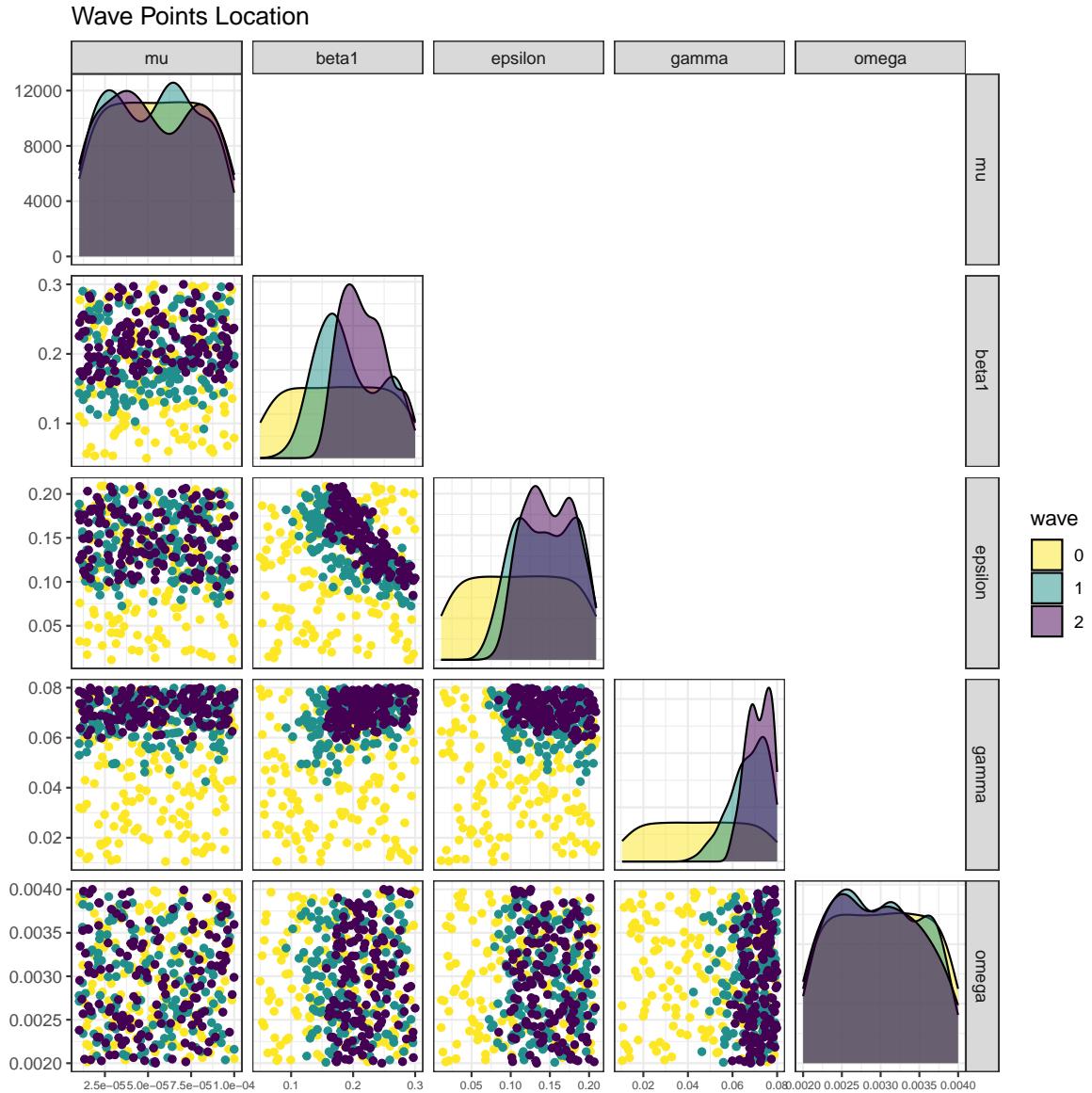
## Visualisation of the non-imausible space by wave

A video presentation of this section can be found [here](#).

In this last section we present three visualisations that can be used to compare the non-imausible space identified at different waves of the process.

The first visualisation, obtained through the function `wave_points`, shows the marginal distributions of the non-imausible space for the waves of interest. For example, let us plot the distribution of five of the parameter sets at the beginning, at the end of wave one and at the end of wave two:

```
wave_points(list(initial_points, new_points, new_new_points),
            input_names = names(ranges)[c(2,3,6,8,9)]) +
            ggplot2::theme(axis.text.x = ggplot2::element_text(size = 6))
```



Here we can easily see that the spread of the second and third wave points are more narrow than those of the first wave.

The second visualisation allows us to assess how much better parameter sets at later waves perform compared to the original `initial_points`.

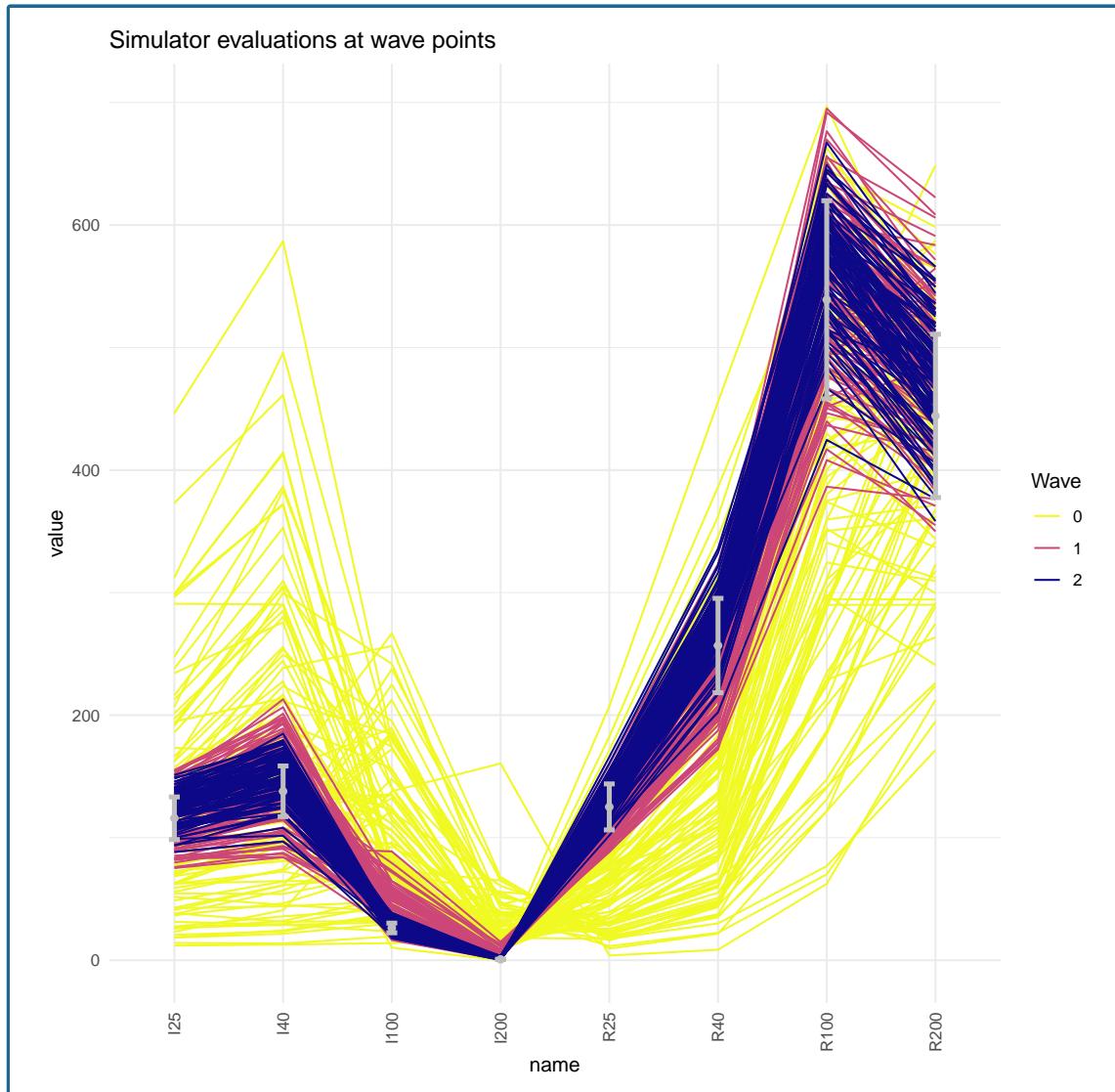
In this workshop we assume that for a given parameter set  $x$ , we are interested in matching the mean of the stochastic model to the observed data, rather than in the output of each realisation at  $x$ , which will be dealt with in future versions. For this reason, we will use

the helper function `aggregate_points`, which calculates the mean of each output across different realisations.

```
all_training_aggregated <- aggregate_points(all_training, names(ranges))
new_all_training_aggregated <- aggregate_points(new_all_training, names(ranges))
new_new_all_training_aggregated <- aggregate_points(new_new_all_training, names(ranges))
all_aggregated <- list(all_training_aggregated, new_all_training_aggregated,
                       new_new_all_training_aggregated)
```

We are now ready to compare the performance of parameter sets at the end of each wave, passing the list `all_aggregated` and the list `targets` to the function `simulator_plot`:

```
simulator_plot(all_aggregated, targets, barcol = "grey")
```

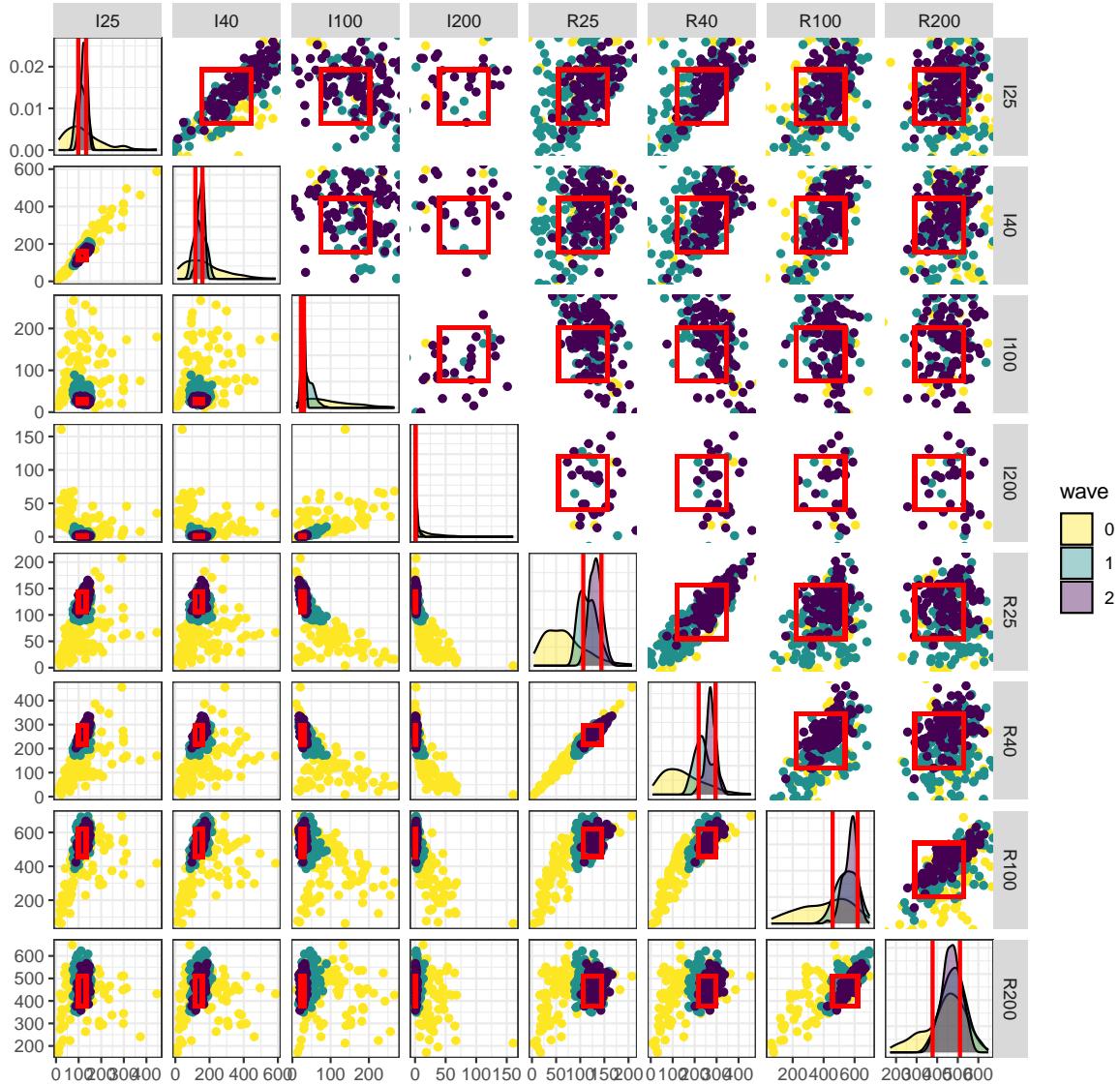


Note that in this call we set `barcol="grey"` in order to have the target intervals in white, instead of black. This plot clearly shows that parameter sets at the end of the first and second wave perform better than the initial set of points.

In the third visualisation, output values for non-imausible parameter sets at each wave are shown for each combination of two outputs:

```
wave_values(all_aggregated, targets, l_wid=1)
```

Output plots with targets



The argument `l_wid` is optional and helps customise the width of the red lines that create the target boxes. The main diagonal shows the distribution of each output at the end of each wave, with the vertical red lines indicating the lower and upper bounds of the target. Above and below the main diagonal are plots for each pair of targets, with rectangles indicating the target area where full fitting points should lie (the ranges are normalised in the figures above the diagonals). These graphs can provide additional information on output distributions, such as correlations between them. For example, here we see positive correlations between  $I25$  and  $R40$ .

In this workshop we have shown how to perform the first three waves of the history matching process on a stochastic model. Of course, more waves are required, in order to complete the calibration task.

Show: Remind me of possible stopping criteria on P78

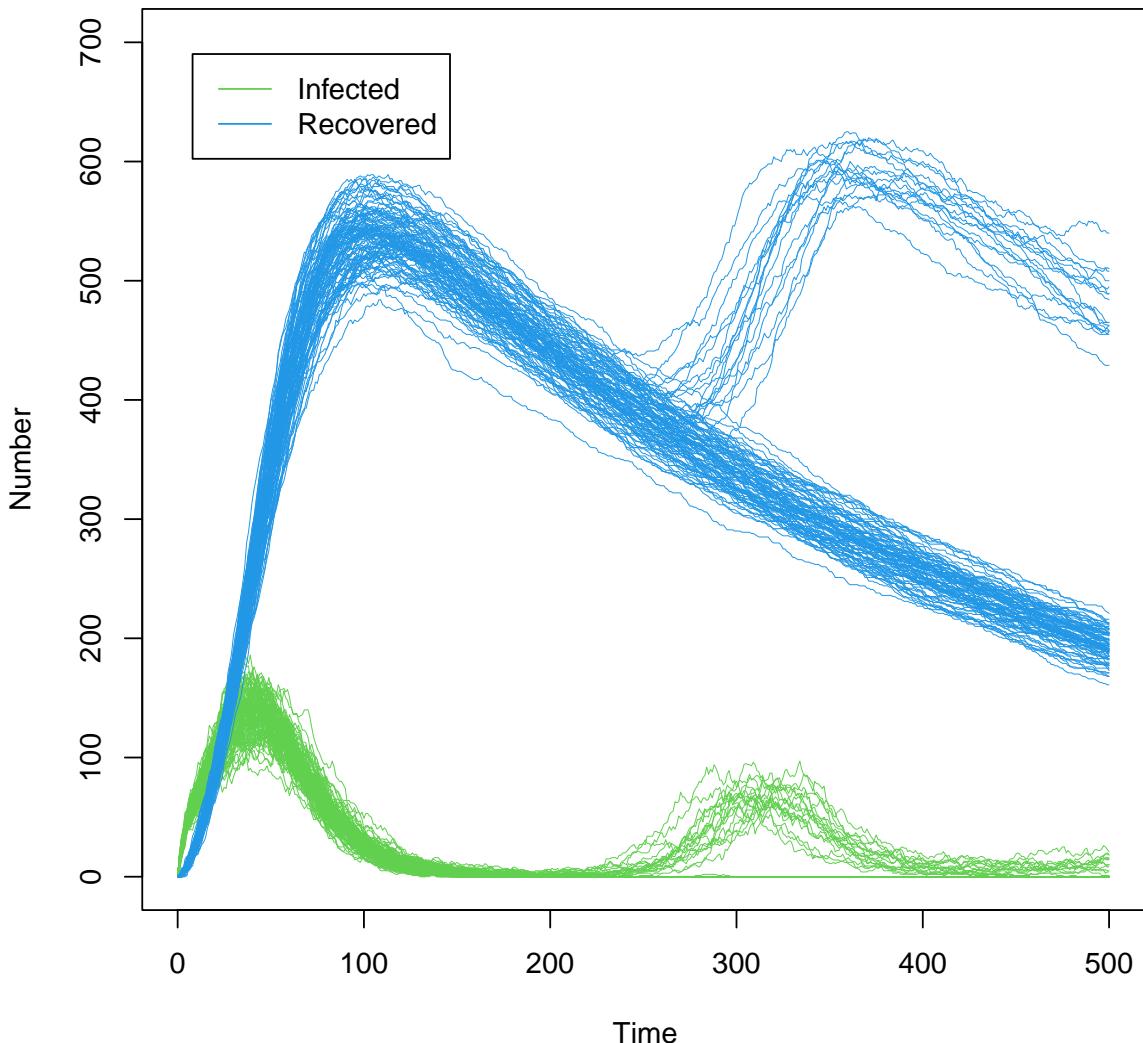
# Chapter 11

## Dealing with bimodality

A video presentation of this section can be found [here](#).

In the previous sections of this workshop we dealt with stochasticity, but not with ‘bimodality’, when one can perform multiple repetitions at a given parameter set and find two distinct classes of behaviour. This is because our SEIRS model did not show any bimodal behaviour in the outputs for the times considered, i.e. for  $t$  up to 200. Let us suppose now that we are interested also in later times. Running the model on `chosen_params` using `get_results` up to time  $t = 500$  and plotting the results for “I” and “R” we obtain:

```
solution <- get_results(chosen_params, outs = c("I", "R"),
                         times = c(25, 40, 100, 200, 300, 400, 500), raw = TRUE)
plot(0:500, ylim=c(0,700), ty="n", xlab = "Time", ylab = "Number")
for(j in 3:4) for(i in 1:100) lines(0:500, solution[,j,i], col=(3:4)[j-2], lwd=0.3)
legend('topleft', legend = c('Infected', "Recovered"), lty = 1,
       col = c(3,4), inset = c(0.05, 0.05))
```



The most common example of bimodality, ‘take off vs die out’, now enters the picture: from time  $t = 250$  on, there are a large number of trajectories in green that have zero infected individuals (note the horizontal green line) and a few trajectories where transmission takes off again, producing a second wave of the epidemic (around  $t = 300$ ).

### Task 3

Investigate how different parameter sets give rise to different levels of bimodality.

Show: R tip on P78

Show: Solution on P69

In the following subsections we will describe how to deal with bimodality when performing history matching with emulation. Since most parts of the process are very similar to what we have seen already for stochastic models, we will only discuss the areas where the process changes.

Similar to the philosophy of the previous sections, we will match the observed data to the mean within each of the two modes. This can be used to then exclude a large portion of the input space without resorting to modelling the specific distributional form of the model output. In later workshops we will also learn to match to the set of possible realisations of a stochastic model, by emulating its covariance structure.

## 11.1 ‘bimodal\_wave0’ - design points

While the parameters’ ranges are the same as before, we need to redefine the targets’ list, adding values for times later than  $t = 200$ :

```
bimodal_targets <- list(
  I25 = list(val = 115.88, sigma = 5.79),
  I40 = list(val = 137.84, sigma = 6.89),
  I100 = list(val = 26.34, sigma = 1.317),
  I200 = list(val = 0.68, sigma = 0.034),
  I250 = list(val = 9.67, sigma = 4.76),
  I400 = list(val = 15.67, sigma = 5.36),
  I500 = list(val = 14.45, sigma = 5.32),
  R25 = list(val = 125.12, sigma = 6.26),
  R40 = list(val = 256.80, sigma = 12.84),
  R100 = list(val = 538.99, sigma = 26.95),
  R200 = list(val = 444.23, sigma = 22.21),
  R250 = list(val = 361.08, sigma = 25.85),
  R400 = list(val = 569.39, sigma = 26.52),
  R500 = list(val = 508.64, sigma = 28.34)
)
```

Note that these are the same exact targets chosen for the deterministic workshop, [Workshop 1](#).

In principle, we now need to run the parameter sets in `initial_points` to include times larger than  $t = 200$ , to obtain our training data. Due to the comparative slowness of the model, in particular since we need to run multiple repetitions at each parameter combination, we have pre-run this to obtain `bimodal_wave0`, a data.frame consisting of 7500 realisations: 50 per point for each of the 150 points in our initial ensemble. To obtain a similar data.frame, one could use the code in Section 4 but modifying the `times` argument to be the numeric vector `c(25, 40, 100, 200, 250, 400, 500)`.

```
bimodal_wave0 <- read.csv(url("https://maths.dur.ac.uk/users/andrew.iskauskas/Resources/bimodal_data.csv"))
bimodal_all_training <- bimodal_wave0[1:5000,]
bimodal_all_valid <- bimodal_wave0[5001:7500,]
bimodal_output_names <- c("I25", "I40", "I100", "I200", "I250", "I400", "I500",
```

```
"R25", "R40", "R100", "R200", "R250", "R400", "R500")
```

## 11.2 Bimodal Emulators

To train bimodal emulators we use the function `emulator_from_data` specifying `emulator_type = 'multistate'`, which requires the training data, the names of the outputs to emulate, and the ranges of the parameters:

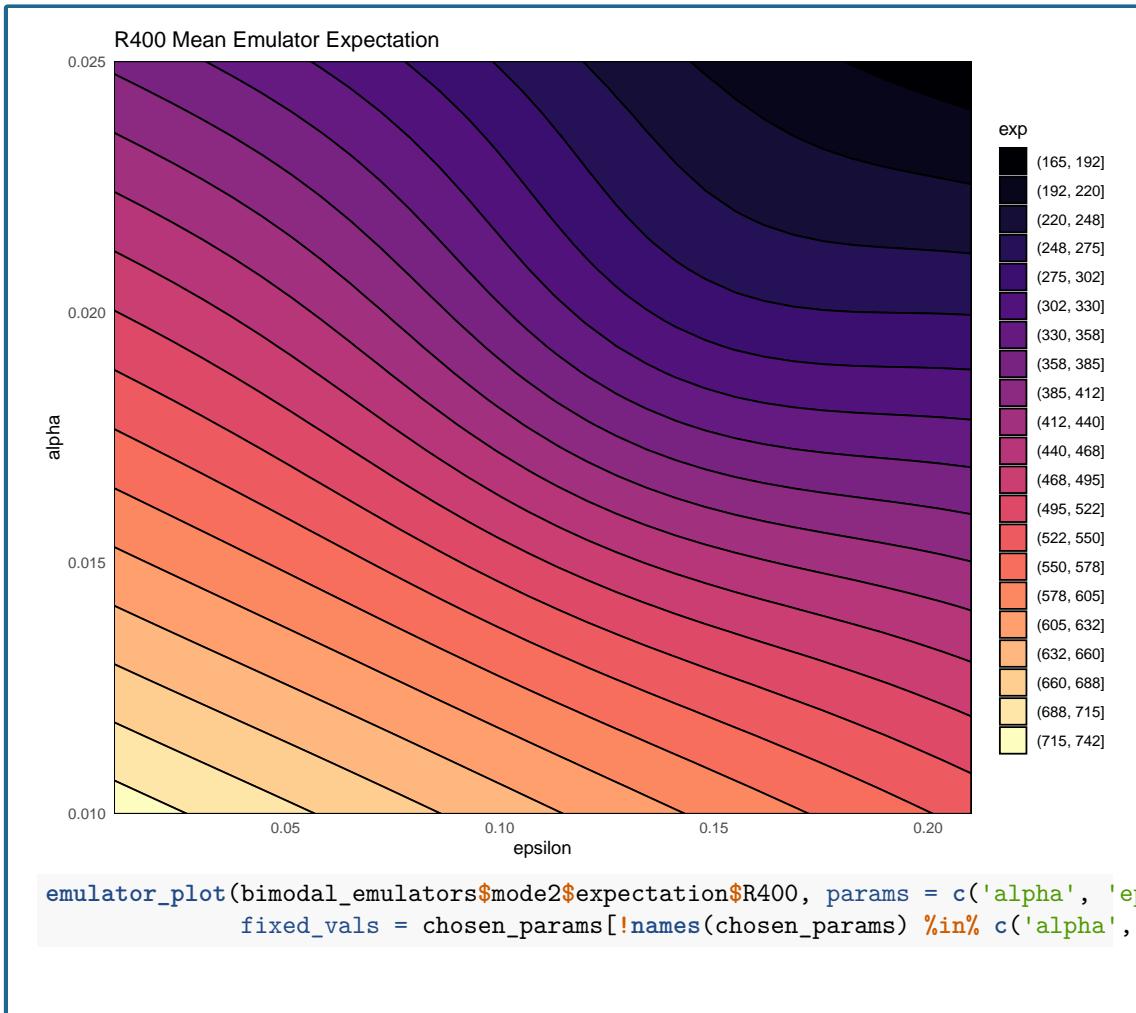
```
bimodal_emulators <- emulator_from_data(bimodal_all_training,
                                         bimodal_output_names, ranges,
                                         emulator_type = 'multistate')
```

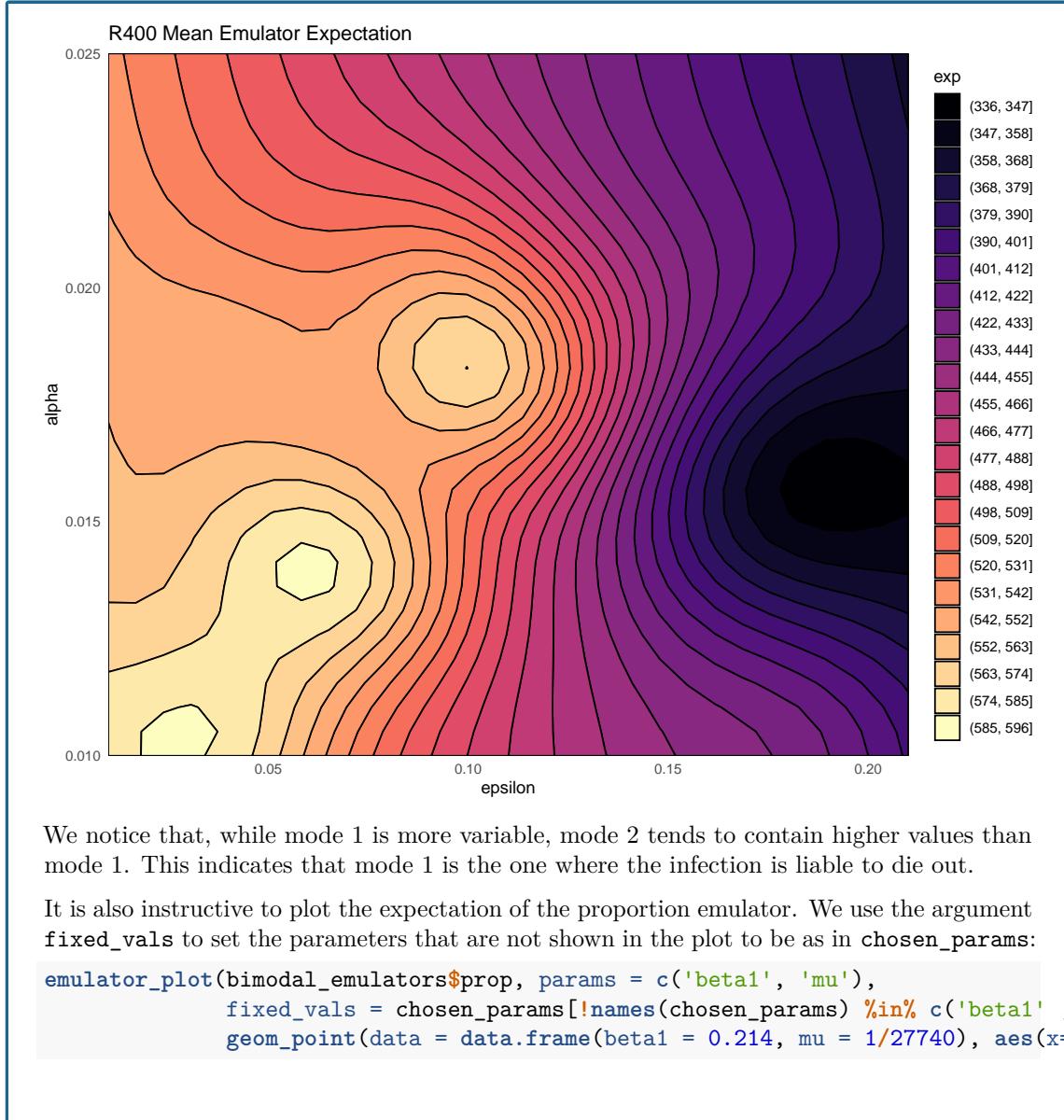
Behind the scenes, this function does the following:

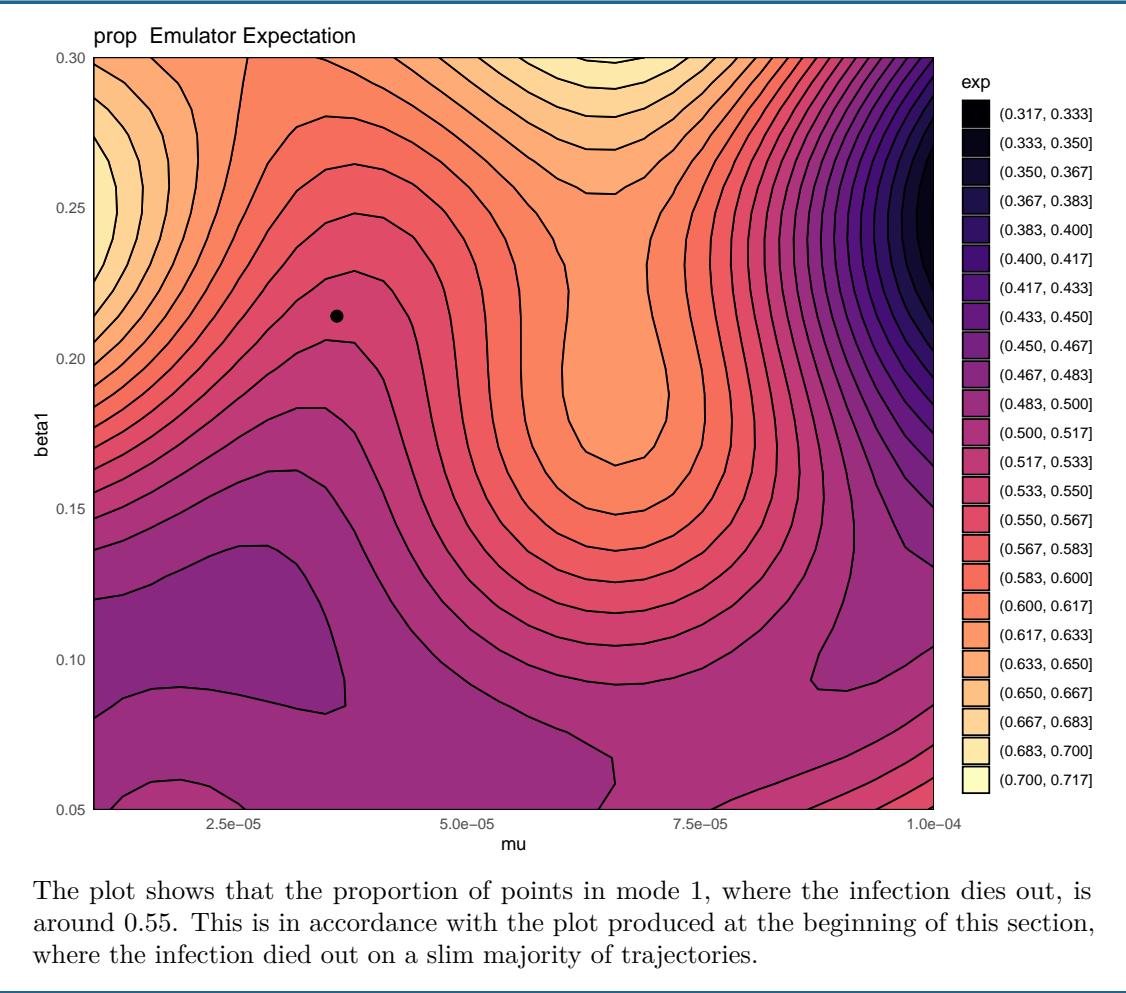
- First it looks at the provided training data and identifies which of the outputs are bimodal (see box below if interested in the method used to identify bimodal outputs).
- For the outputs where bimodality is found, the repetitions at each parameter set are clustered into two subsets, based on the mode they belong to.
- For outputs without bimodality, stochastic emulators are trained as before. For outputs with bimodality, variance and mean emulators are trained separately for each of the two modes. To access the mean emulators for the first mode, we type `bimodal_emulators$mode1$expectation`, and to access the variance emulators for the second mode we type `bimodal_emulators$mode2$variance`.
- Finally, an emulator for the proportion of points in each mode is also trained (this is a single emulator, as in the deterministic case). This emulator can be accessed by typing `bimodal_emulators$prop` and will be referred to as proportion emulator.

Let us now plot the expectation of the mean emulator of *R400* for each mode fixing all non-shown parameters to their values in `chosen_params`:

```
emulator_plot(bimodal_emulators$mode1$expectation$R400, params = c('alpha', 'epsilon'),
               fixed_vals = chosen_params[!names(chosen_params) %in% c('alpha', 'epsilon')])
```







Show: How are bimodal outputs identified? on P78

#### Task 4

Confirm that mode 1 is where the infection dies out by comparing the predictions of the mean emulators for each mode.

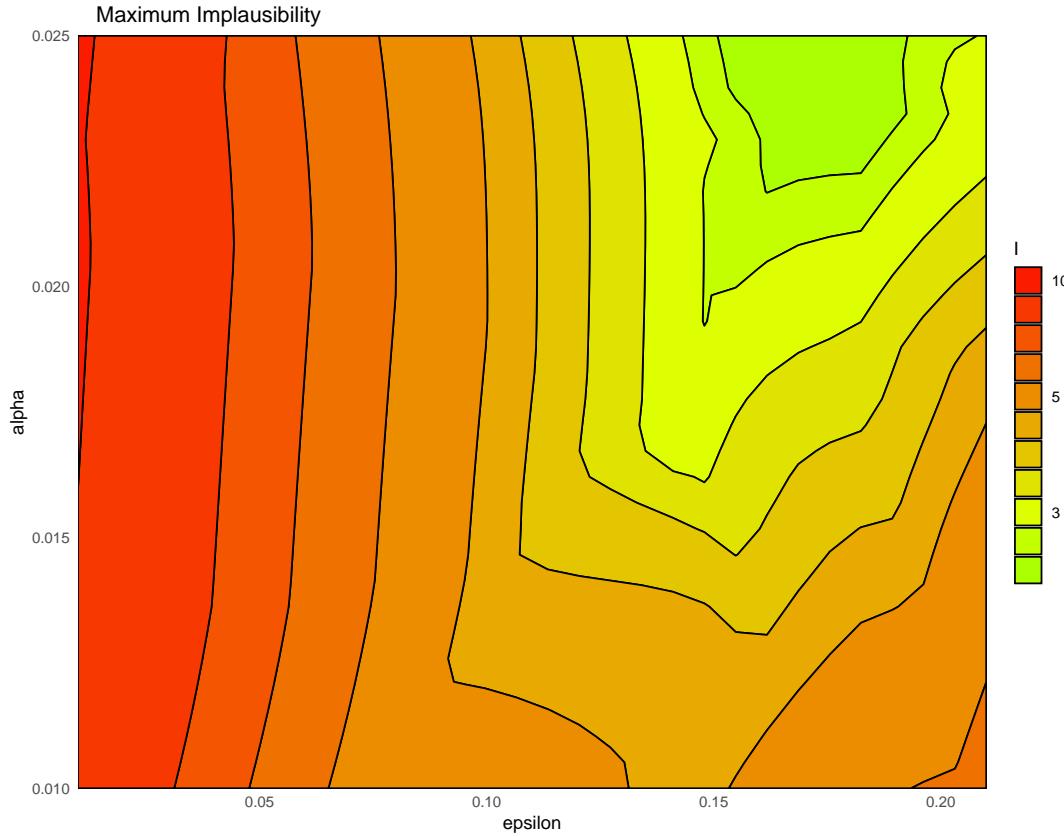
Show: Solution on P71

### 11.3 Implausibility

Since we are now in a bimodal setting, we want to regard a point as non-implausible for a given target if it is valid with respect to either mode. This is the default behaviour of the package, when dealing with bimodal emulators.

For a given output and a given point, the implausibility for mode 1 and the implausibility for mode 2 are calculated, and the minimum of them is taken. The maximum (or second-maximum, third-maximum etc) of this collection of minimised implausibilities is then selected, depending on the user choice. For example, to plot the maximum of these minimised implausibilities, we set `plot_type` to `nimp`:

```
emulator_plot(bimodal_emulators, plot_type = 'nimp', targets = bimodal_targets,
              params = c('alpha', 'epsilon'))
```



## Task 5

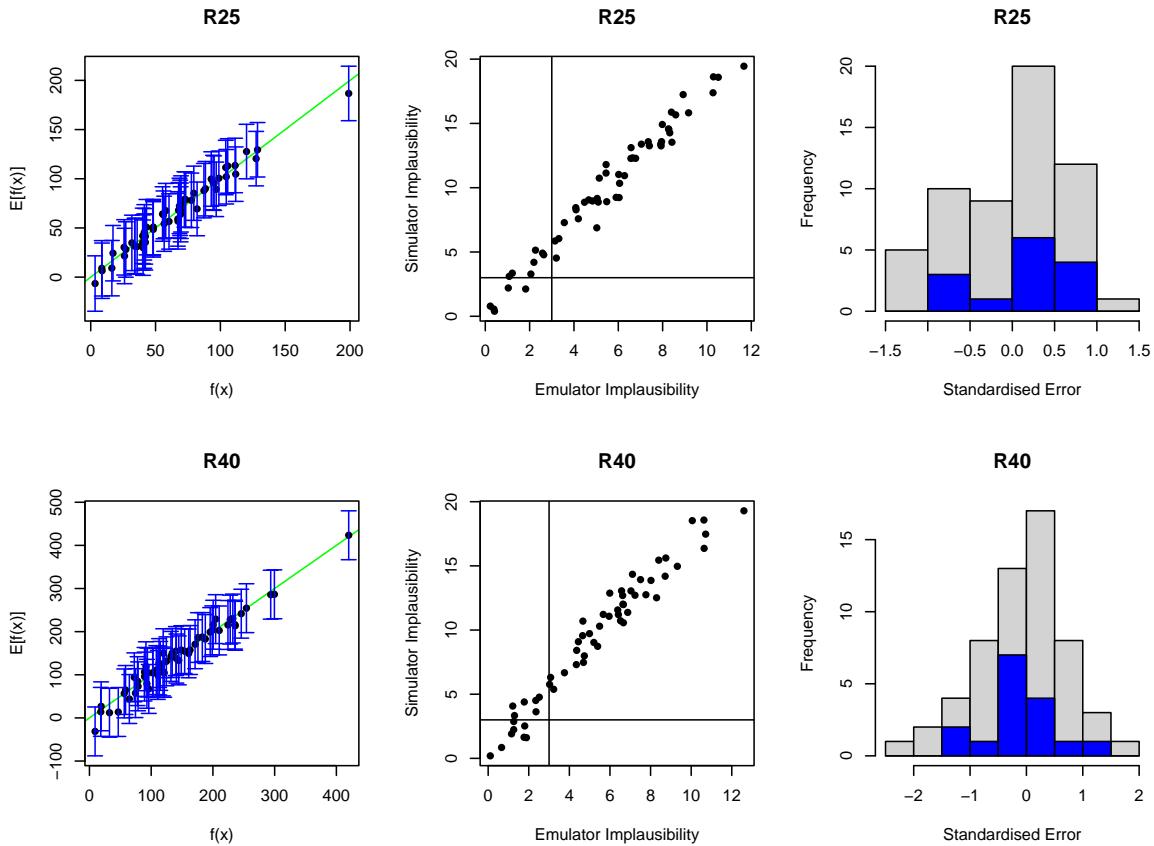
Set the argument `plot_type` to `imp` to produce implausibility plots for each output, for mode 1 and mode 2. Which implausibility plots are the same for each mode, and which are different? Why?

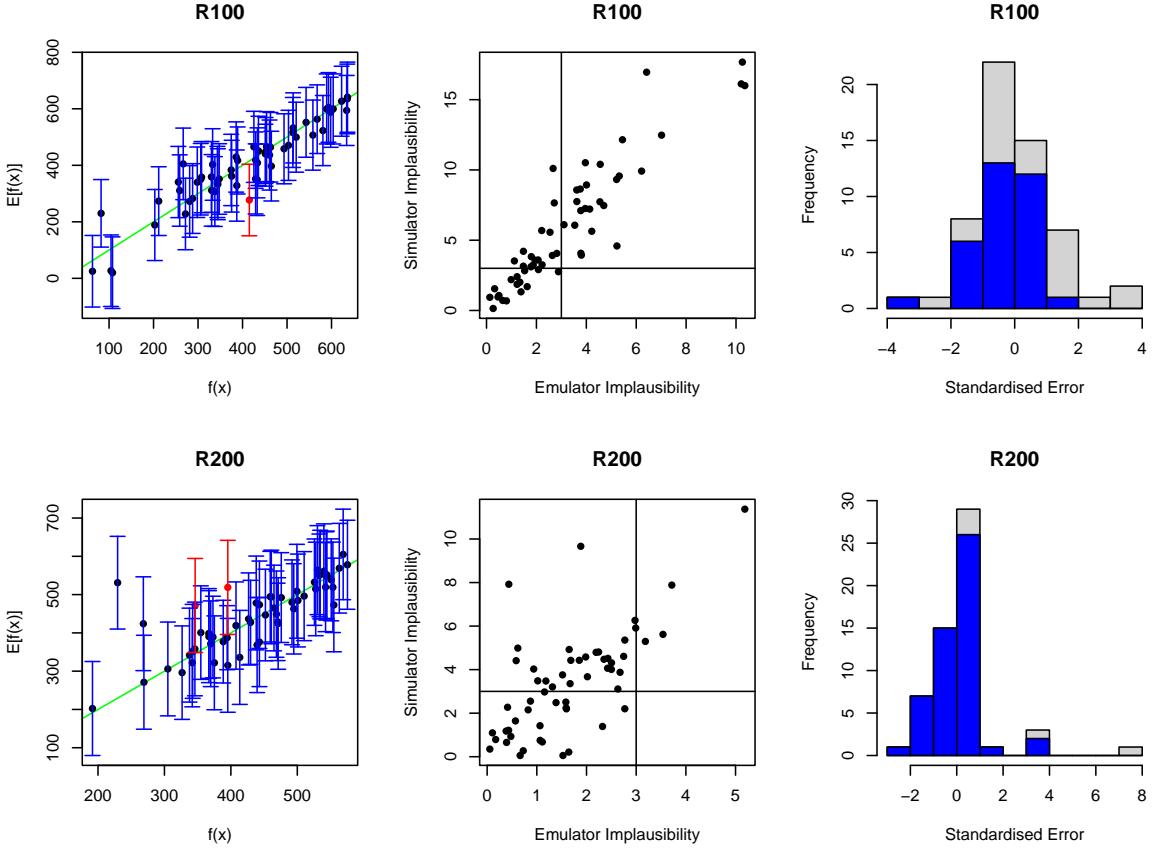
Show: Solution on P73

## 11.4 Emulator diagnostics

As before, the function `validation_diagnostics` can be used to get three diagnostics for each emulated output. For example, to produce diagnostics for the first four emulators of the number of recovered individuals we type:

```
validation_subset <- subset_emulators(bimodal_emulators, c("R25", "R40", "R100", "R200"))
vd <- validation_diagnostics(validation_subset, bimodal_targets, bimodal_all_valid,
                             plt=TRUE, row=2)
```





Note that, in order to restrict to a subset of the emulated outputs, we use the `subset_emulators` function. This ensures that the requisite outputs are selected while maintaining the structure of the emulator lists (i.e. we still have modes 1 and 2, expectation, variance, etc).

You may have noticed that in the first column we have more points plotted than we have validation points. This is because we do diagnostics for each mode for each output, i.e. we compare the predictions of the mean emulators for mode 1 and mode 2 with the model output values, which are also clustered in two subsets, due to bimodality.

As in the deterministic case, one can enlarge the  $\sigma$  values to obtain more conservative emulators, if needed. For example, to increase the  $\sigma$  value slightly for the mean emulators for  $R200$  in both modes we would type:

```
bimodal_emulators$mode1$expectation$R200 <- bimodal_emulators$mode1$expectation$R200$mult_sigma(1)
bimodal_emulators$mode2$expectation$R200 <- bimodal_emulators$mode2$expectation$R200$mult_sigma(1)
```

## 11.5 Proposing new points

Generating a set of non-imausible points, based on the trained emulators, is done in exactly the same way as it was in Section 8, using the function `generate_new_design`.

```
new_points <- generate_new_design(bimodal_emulators, 150, bimodal_targets, nth=1)
```

Similarly, performing the second wave is done in the same as it was in Section 9, replacing `emulator_type = 'variance'` with `emulator_type = 'multistate'` when calling `emulator_from_data`.



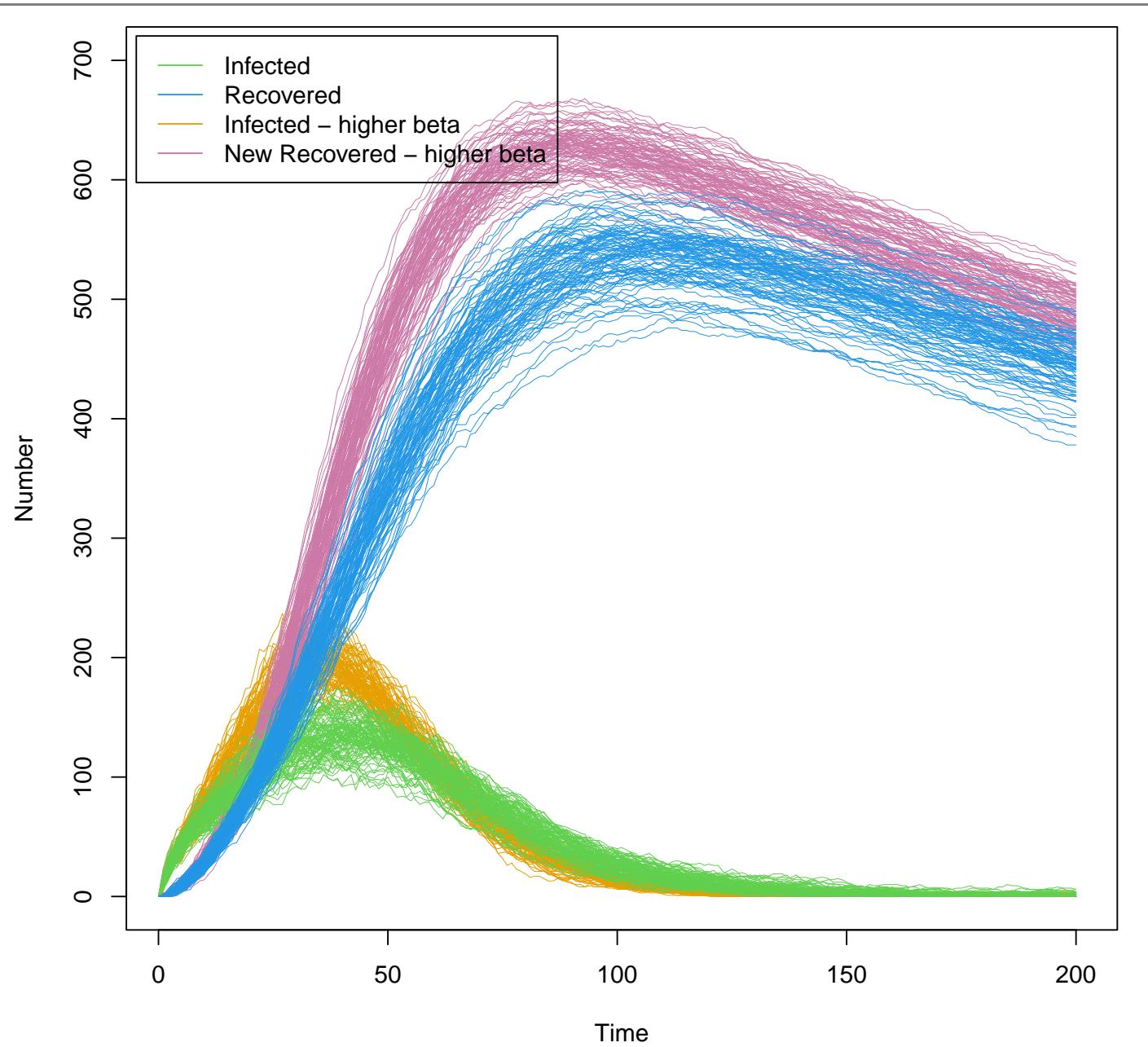
# Appendix A

## Answers

### Solution 1

Let us see what happens when a higher rate of infection between each infectious and susceptible person is considered. We first rerun the model on `example_params` to ensure that `solution` contains the correct trajectories:

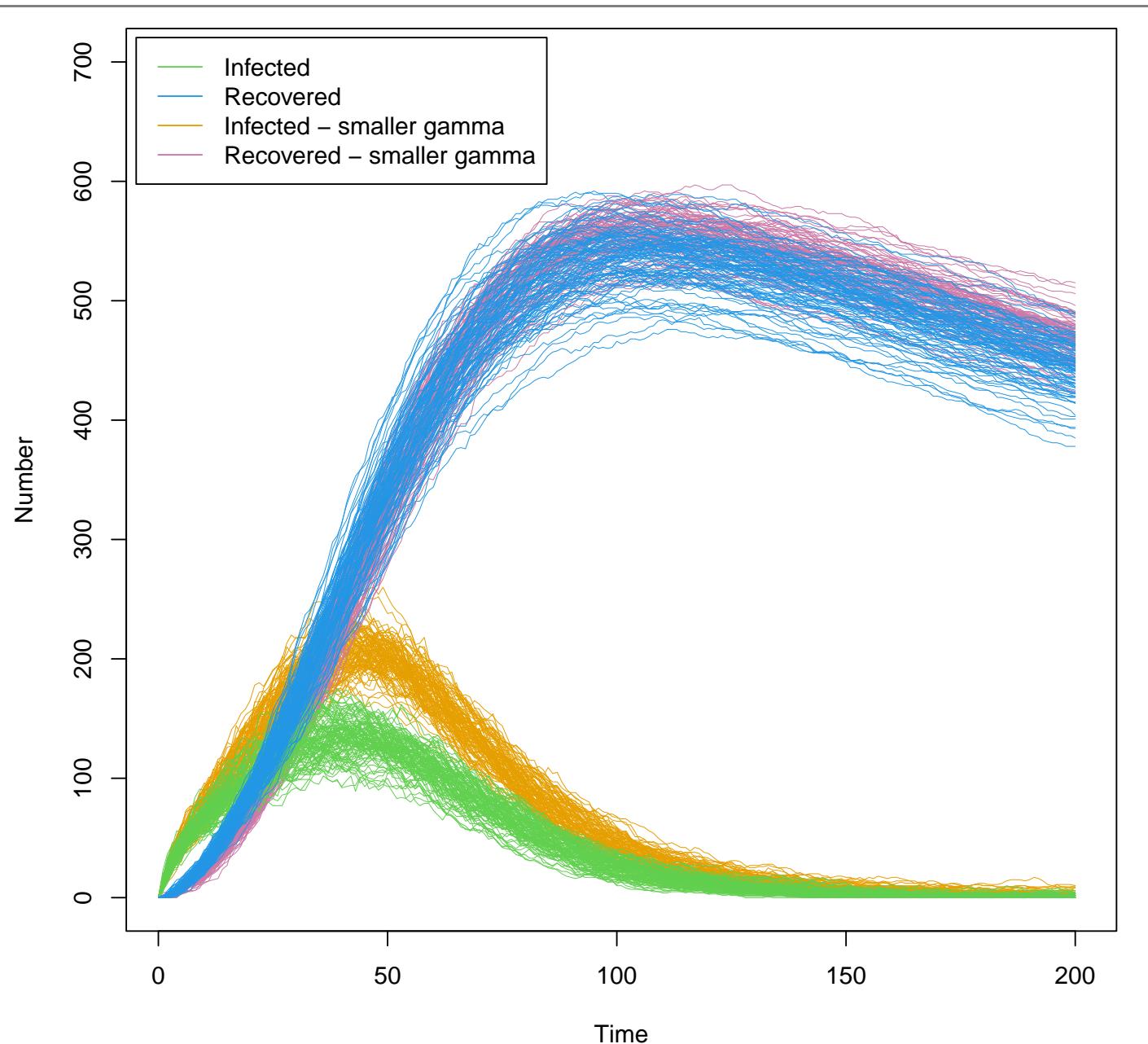
```
example_params <- c(
  b = 1/(76*365), # birth rate
  mu = 1/(76*365), # rate of death from other causes
  beta1 = 0.214, beta2 = 0.107, beta3 = 0.428, # infection rate between each infectious and susceptible individual
  epsilon = 1/7, # rate of becoming infectious after infection
  alpha = 1/50, # rate of death from the disease
  gamma = 1/14, # recovery rate
  omega = 1/365 # rate at which immunity is lost following recovery
)
solution <- get_results(example_params, outs = c("I", "R"),
                        times = c(25, 40, 100, 200), raw = TRUE)
higher_beta_params <- c(
  b = 1/(76*365),
  mu = 1/(76*365),
  beta1 = 0.3, beta2 = 0.1, beta3 = 0.5,
  epsilon = 1/7,
  alpha = 1/50,
  gamma = 1/14,
  omega = 1/365
)
higher_beta_solution <- get_results(higher_beta_params, outs = c("I", "R"),
                                      times = c(25, 40, 100, 200), raw = TRUE)
plot(0:200, ylim=c(0,700), ty="n", xlab = "Time", ylab = "Number")
for(j in 3:4) for(i in 1:100) lines(0:200, higher_beta_solution[,j,i],
                                    col=(c("#E69F00", "#CC79A7"))[j-2], lwd=0.3)
for(j in 3:4) for(i in 1:100) lines(0:200, solution[,j,i], col=(3:4)[j-2], lwd=0.3)
legend('topleft', legend = c('Infected', "Recovered", "Infected - higher beta",
                             "New Recovered - higher beta"), lty = 1,
       col = c(3,4,"#E69F00", "#CC79A7"), inset = c(0.01, 0.01))
```



We see that the peaks of recovered and infectious individuals have now increased, as expected. We might also note that the spread of runs has also changed: our higher  $\beta$  parameter has resulted in a less variable collection of realisations, especially after time  $t = 50$ .

What happens when a lower value of the recovery rate,  $\gamma$ , is used?

```
smaller_gamma_params <- c(
  b = 1/(60*365),
  mu = 1/(76*365),
  beta1 = 0.214, beta2 = 0.107, beta3 = 0.428,
  epsilon = 1/7,
  alpha = 1/50,
  gamma = 1/20,
  omega = 1/365
)
smaller_gamma_solution <- get_results(smaller_gamma_params, outs = c("I", "R"),
                                         times = c(25, 40, 100, 200), raw = TRUE)
plot(0:200, ylim=c(0,700), ty="n", xlab = "Time", ylab = "Number")
for(j in 3:4) for(i in 1:100) lines(0:200, smaller_gamma_solution[,j,i],
                                     col=(c("#E69F00", "#CC79A7"))[j-2], lwd=0.3)
for(j in 3:4) for(i in 1:100) lines(0:200, solution[,j,i], col=(3:4)[j-2], lwd=0.3)
legend('topleft', legend = c('Infected', "Recovered", "Infected - smaller gamma",
                             "Recovered - smaller gamma"), lty = 1,
       col = c(3,4,"#E69F00", "#CC79A7"), inset = c(0.01, 0.01))
```



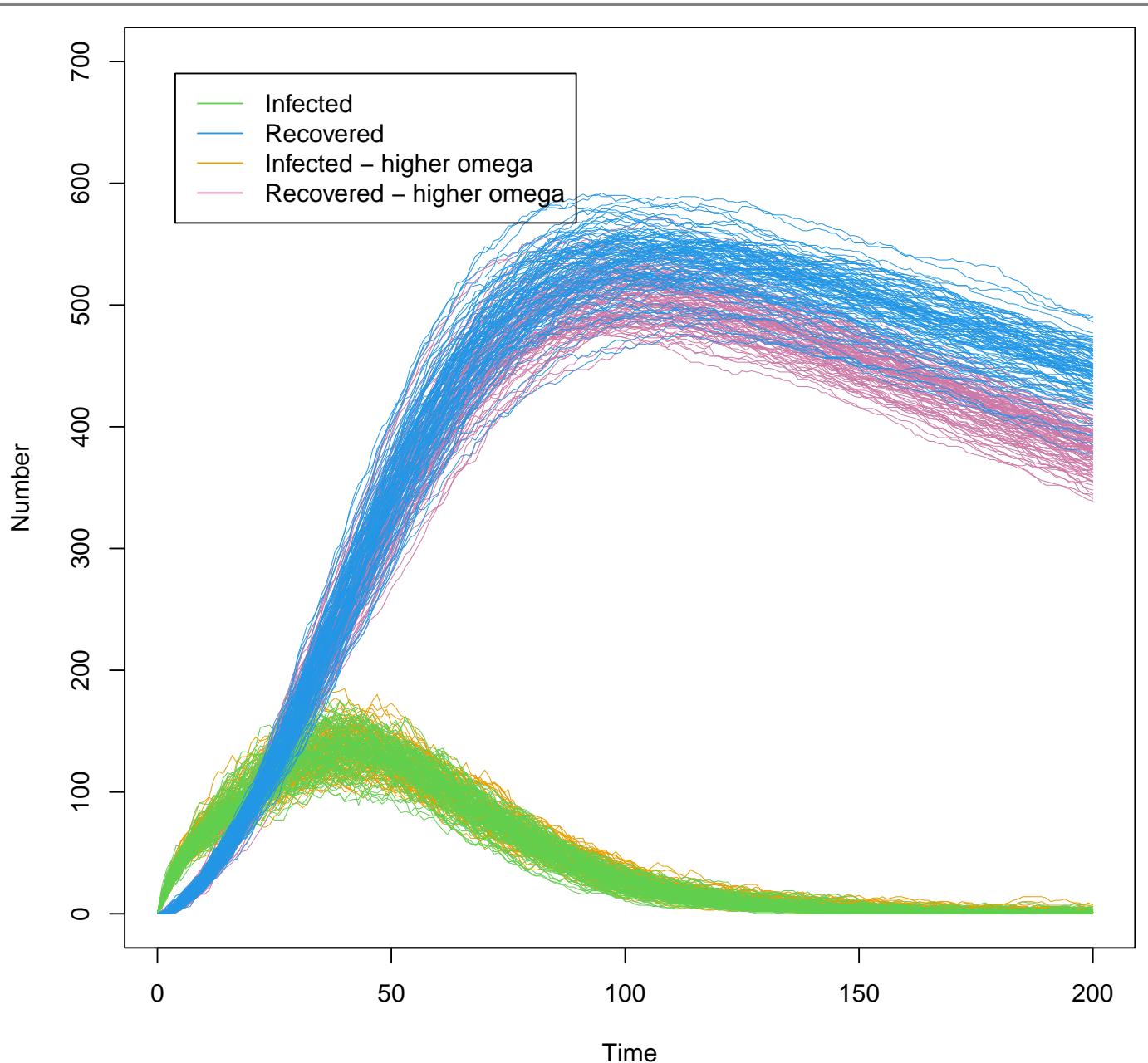
As one expects, this causes the peak of infectious individuals to increase, while the variability of runs does not appreciably change.

Finally, what happens when we increase the rate at which immunity is lost,  $\omega$ ?

```

higher_omega_params <- c(
  b = 1/(60*365),
  mu = 1/(76*365),
  beta1 = 0.214, beta2 = 0.107, beta3 = 0.428,
  epsilon = 1/7,
  alpha = 1/50,
  gamma = 1/14,
  omega = 0.004
)
higher_omega_solution <- get_results(higher_omega_params, outs = c("I", "R"),
                                       times = c(25, 40, 100, 200), raw = TRUE)
plot(0:200, ylim=c(0,700), ty="n", xlab = "Time", ylab = "Number")
for(j in 3:4) for(i in 1:100) lines(0:200, higher_omega_solution[,j,i],
                                    col=(c("#E69F00", "#CC79A7"))[j-2], lwd=0.3)
for(j in 3:4) for(i in 1:100) lines(0:200, solution[,j,i], col=3:4[j-2], lwd=0.3)
legend('topleft', legend = c('Infected', "Recovered", "Infected - higher omega",
                             "Recovered - higher omega"), lty = 1,
       col = c(3,4,"#E69F00","#CC79A7"), inset = c(0.05, 0.05))

```



This causes the peak of recovered individuals to slightly decrease; however, the decrease in peak is not so substantial as to have this change outweigh the effect of stochasticity.

The three examples above give an indication of the importance of considering stochasticity: even where there are appreciable changes in behaviour that align with our physical intuition about the disease, their effects may not be as clear-cut as we would expect. Furthermore, the amount of variability as well as the model output can be different at different parts of parameter space. When matching to such a model, we must therefore be cogniscent of these two aspects.

[Return to task on P13](#)

## Solution 2

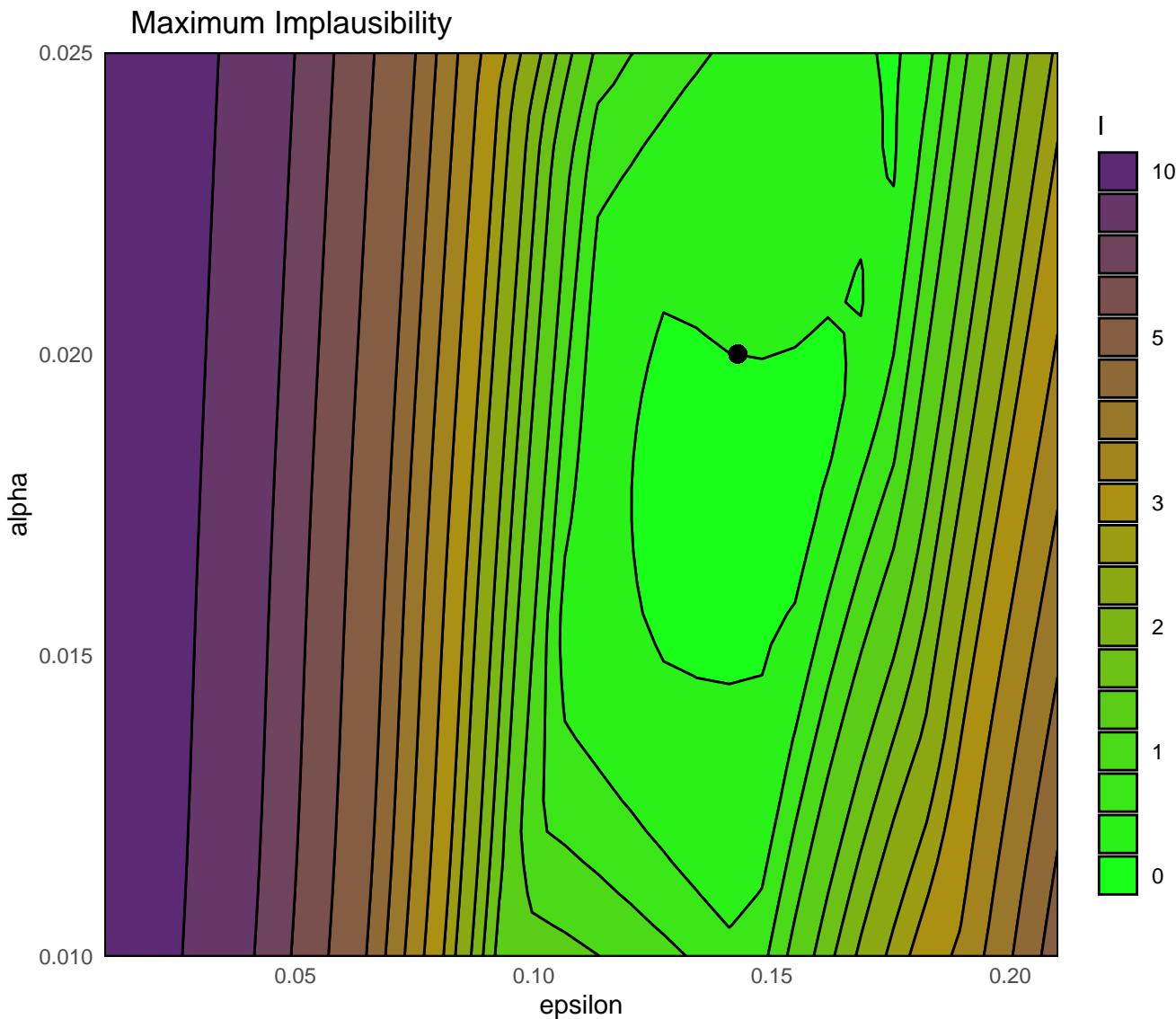
We set `fixed_vals` to `chosen_params[!names(chosen_params) %in% c('epsilon', 'alpha')]`, plot the maximum implausibility and add a point corresponding to the values of  $\sigma$  and  $\alpha$  in `chosen_params`:

```

emulator_plot(stoch_emulators, plot_type = 'nimp', targets = targets,
              params = c('epsilon', 'alpha'),
              fixed_vals = chosen_params[!names(chosen_params) %in% c('epsilon', 'alpha')],
              cb=T) +geom_point(aes(x=1/7, y=1/50), size=3)

## Warning in geom_point(aes(x = 1/7, y = 1/50), size = 1): All aesthetics have length 1, but the data has
## i Please consider using `annotate()` or provide this layer with data containing
##   a single row.

```



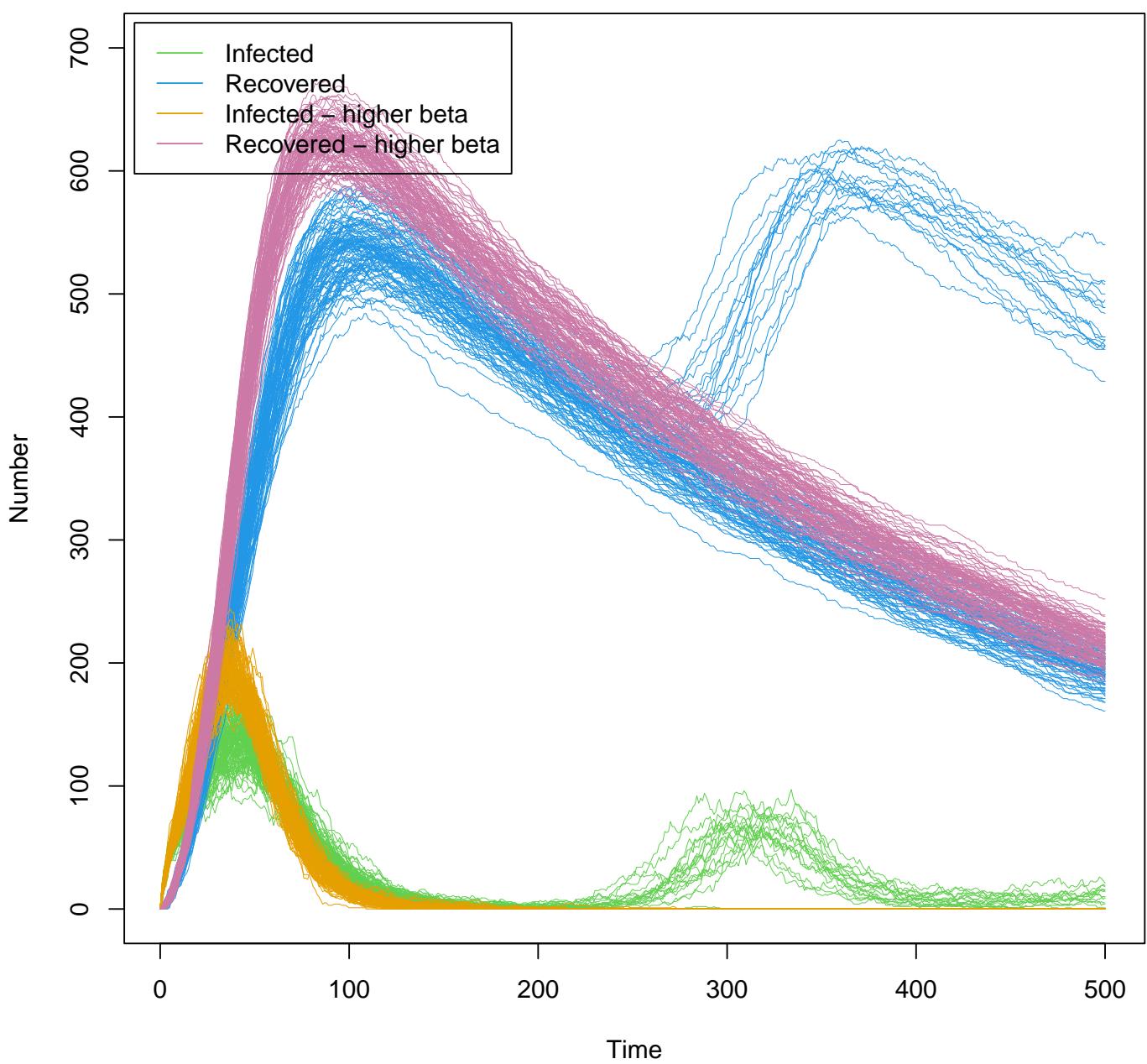
The plot shows what we expected: when  $\epsilon$  and  $\alpha$  are equal to their values in `chosen_params`, the implausibility measure is below the threshold 3 (cf. black point in the box).

[Return to task on P26](#)

Solution 3

Let us see what happens when we simulate a higher rate of infection between each infectious and susceptible:

```
higher_beta_params <- c(
  b = 1/(76*365),
  mu = 1/(76*365),
  beta1 = 0.3, beta2 = 0.1, beta3 = 0.5,
  sigma = 1/7,
  alpha = 1/50,
  gamma = 1/14,
  omega = 1/365
)
higher_beta_solution <- get_results(higher_beta_params, outs = c("I", "R"),
                                      times = c(25, 40, 100, 200, 300, 400, 500), raw = TRUE)
plot(0:500, ylim=c(0,700), ty="n", xlab = "Time", ylab = "Number")
for(j in 3:4) for(i in 1:100) lines(0:500, solution[,j,i], col=(3:4)[j-2], lwd=0.3)
for(j in 3:4) for(i in 1:100) lines(0:500, higher_beta_solution[,j,i],
                                      col=(c("#E69F00","#CC79A7"))[j-2], lwd=0.3)
legend('topleft', legend = c('Infected', "Recovered", "Infected - higher beta",
                             "Recovered - higher beta"), lty = 1,
       col = c(3,4,"#E69F00","#CC79A7"), inset = c(0.01, 0.01))
```



We see that the peaks of recovered and infectious individuals during the first wave have now increased, as expected. Also note how increasing the transmission rate reduced the role of bimodality (very few trajectories in orange take off after the first wave): the fact that more individuals get now infected during the first wave (around  $t = 50$ ), makes the chances of a second “wave” around time  $t = 300$  lower.

[Return to task on P52](#)

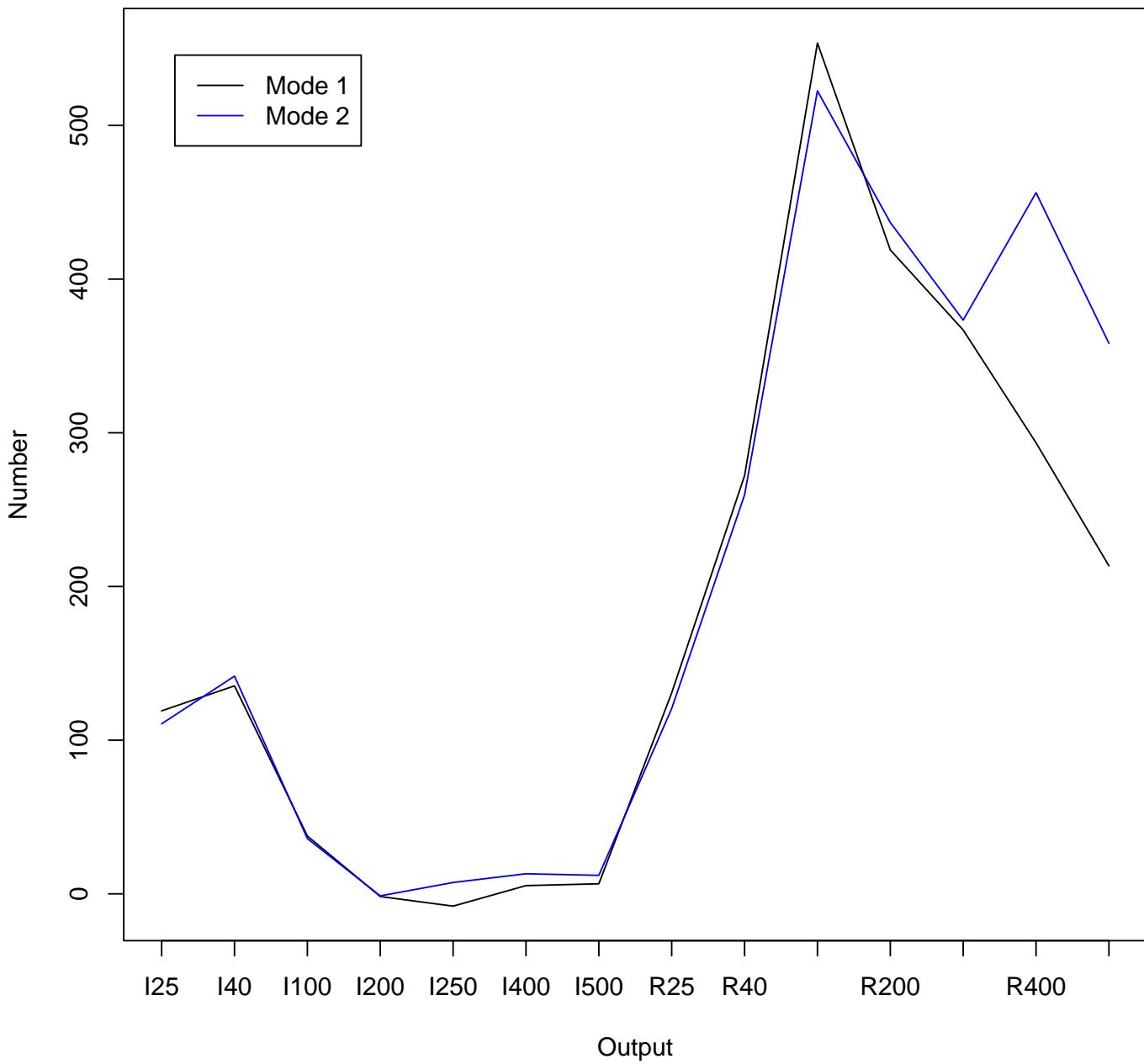
#### Solution 4

We could use `emulator_plot` to perform this, looking at different slices, but it may be more instructive to consider the emulator predictions across all emulated outputs. We use the `chosen_params` as a `data.frame` and predict on both modes, combining the predictions using some base R plotting.

```

chosen_data_frame <- data.frame(t(chosen_params))
mode1_preds <- purrr::map_dbl(bimodal_emulators$mode1$expectation, ~.$get_exp(chosen_data_frame))
mode2_preds <- purrr::map_dbl(bimodal_emulators$mode2$expectation, ~.$get_exp(chosen_data_frame))
plot(mode1_preds, type = 'l', xaxt = "n", xlab = "Output", ylab = "Number")
axis(1, at = seq_len(length(bimodal_output_names)), labels = bimodal_output_names)
lines(mode2_preds, col = "blue")
legend('topleft', inset = 0.05, legend = c("Mode 1", "Mode 2"), lty = 1, col = c('black', 'blue'))

```



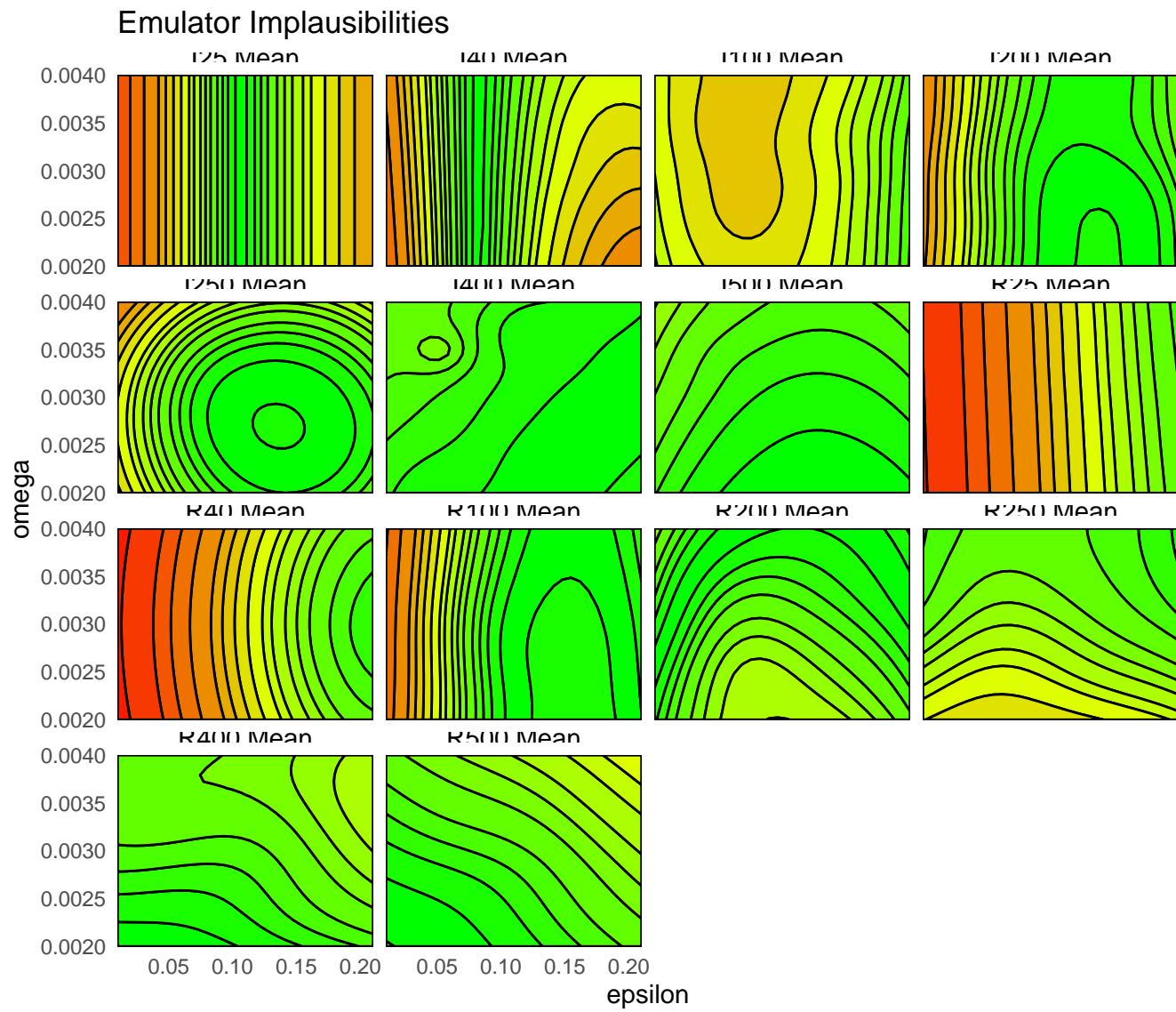
The difference in numbers of infected is small, but appreciable; that of the recovered compartments clearly show the second peak of the disease in the second mode.

[Return to task on P57](#)

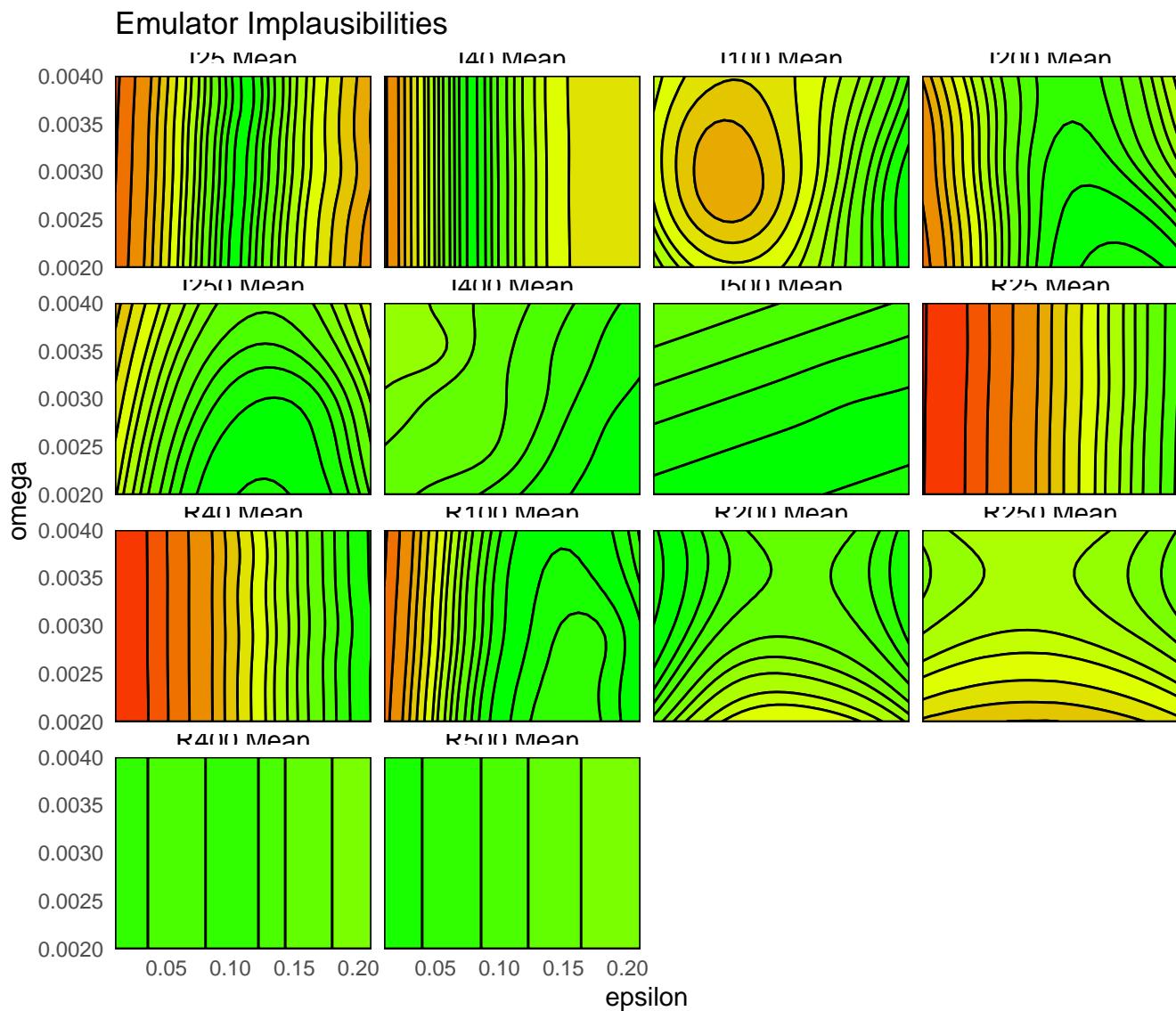
## Solution 5

Setting `plot_type='imp'` in `emulator_plot`, we obtain a collection of plots:

```
emulator_plot(bimodal_emulators$mode1, plot_type = 'imp',
              targets = bimodal_targets, params = c('omega', 'epsilon'))
```



```
emulator_plot(bimodal_emulators$mode2, plot_type = 'imp',
              targets = bimodal_targets, params = c('omega', 'epsilon'))
```



The implausibility plots are broadly the same between modes for early times ( $t = 25, 40, 100$ ) and different between modes for later outputs. This makes sense, since bimodality, which enters the picture after the first wave, should not play a significant role in earlier times.

[Return to task on P59](#)

## **Appendix B**

## **Additional information**

Code to load relevant libraries and helper functions

```

library(hmer)
library(lhs)
library(deSolve)
library(ggplot2)
library(reshape2)
library(purrr)
library(tidyverse)
library(progressr)
handlers("progress")
#####
##### HELPER FUNCTIONS #####
# Function that calculates the mean of the model outputs from several runs
# at the same parameter set
aggregate_points <- function(data, input_names, func = 'mean') {
  unique_points <- unique(data[,input_names])
  uids <- apply(unique_points, 1, rlang::hash)
  data_by_point <- purrr::map(uids, function(h) {
    data[apply(data[,input_names], 1, rlang::hash) == h,]
  })
  aggregate_func <- get(func)
  aggregate_values <- do.call('rbind.data.frame', purrr::map(data_by_point, function(x) {
    output_data <- x[,!(names(x) %in% input_names)]
    apply(output_data, 2, aggregate_func)
  }))
  return(setNames(cbind.data.frame(unique_points, aggregate_values), names(data)))
}
# Function that implements the SEIRS stochastic model using the Gillespie algorithm
gillespied=function (N, T=400, dt=1, ...)
{
  tt=0
  n=T%/%dt
  x=N$M
  S=t(N$Post-N$Pre)
  u=nrow(S)
  v=ncol(S)
  xmat=matrix(0,ncol=u,nrow=n)
  i=1
  target=0
  repeat {
    h=N$h(x, tt, ...)
    h0=sum(h)
    if (h0<1e-10)
      tt=1e99
    else
      tt=tt+rexp(1,h0)
    while (tt>=target) {
      xmat[i,]=x
      i=i+1
      target=target+dt
      if (i>n)
        return(xmat)
    }
    j=sample(v,1,prob=h)
    x=x+S[,j]
  }
  Num <- 1000
  N=list()
  N$M=c(900,100,0,0,0)
  N$Pre = matrix(c(0,0,0,0,0,
                 1,0,0,0,0,
                 1,0,0,0,0,

```

[Return to P6](#)

## R tip

Copy the code below, modify the value of (some) parameters and run it.

```
example_params <- c(
  b = 1/(76*365), # birth rate
  mu = 1/(76*365), # rate of death from other causes
  beta1 = 0.214, beta2 = 0.107, beta3 = 0.428, # infection rate between each infectious and susceptible individuals
  epsilon = 1/7, # rate of becoming infectious after infection
  alpha = 1/50, # rate of death from the disease
  gamma = 1/14, # recovery rate
  omega = 1/365 # rate at which immunity is lost following recovery
)
solution <- get_results(example_params, outs = c("I", "R"),
                         times = c(25, 40, 100, 200), raw = TRUE)
plot(0:200, ylim=c(0,700), ty="n", xlab = "Time", ylab = "Number")
for(j in 3:4) for(i in 1:100) lines(0:200, solution[,j,i], col=(3:4)[j-2], lwd=0.3)
legend('topleft', legend = c('Infected', "Recovered"), lty = 1,
       col = c(3,4), inset = c(0.05, 0.05))
```

[Return to P13](#)

Remind me of what each diagnostic means

**The first column** shows the emulator outputs plotted against the model outputs. In particular, the emulator expectation is plotted against the model output for each validation point, providing the dots in the graph. The emulator uncertainty at each validation point is shown in the form of a vertical interval that goes from  $3\sigma$  below to  $3\sigma$  above the emulator expectation, where  $\sigma$  is the emulator variance at the considered point. An ‘ideal’ emulator would exactly reproduce the model results: this behaviour is represented by the green line  $f(x) = E[f(x)]$  (this is a diagonal line, visible here only in the bottom left and top right corners). Any parameter set whose emulator prediction lies more than  $3\sigma$  away from the model output is highlighted in red. Note that we do not need to have no red points for the test to be passed: since we are plotting  $3\sigma$  bounds, statistically speaking it is ok to have up to 5% of validation points in red (see [Pukelsheim’s  \$3\sigma\$  rule](#)).

**The second column** compares the emulator implausibility to the equivalent model implausibility (i.e. the implausibility calculated replacing the emulator output with the model output). There are three cases to consider:

- The emulator and model both classify a set as implausible or non-implausible (bottom-left and top-right quadrants). This is fine. Both are giving the same classification for the parameter set.
- The emulator classifies a set as non-implausible, while the model rules it out (top-left quadrant): this is also fine. The emulator should not be expected to shrink the parameter space as much as the model does, at least not on a single wave. Parameter sets classified in this way will not be ruled out in this wave, but may be removed on subsequent waves as the emulators grow more accurate on a reduced parameter space.
- The emulator rules out a set, but the model does not (bottom-right quadrant): these are the problem sets, suggesting that the emulator is ruling out parts of the parameter space that it should not be ruling out. As for the first test, we should be alarmed only if we spot a systematic problem, but we might be more circumspect about any of the points in the bottom-right quadrant if we wish to be careful about matching to data: a point that is misclassified in this way may correspond to a neighbourhood of points that would be acceptable and are going to be ruled out.

Finally, **the third column** gives the standardised errors of the emulator outputs in light of the model output: for each validation point, the difference between the emulator output and the model output is calculated, and then divided by the standard deviation  $\sigma$  of the emulator at the point. The general rule is that we want our standardised errors to be somewhat normally distributed around 0, with 95% of the probability mass between  $-3$  and  $3$ . The blue bars indicate

the distribution of the standardised errors when we restrict our attention only to parameter sets that produce outputs close to the targets. When looking at the standard errors plot, we should ask ourselves at least the following questions:

- Is more than 5% of the probability mass outside the interval  $[-3, 3]$ ? If the answer is yes, this means that, even factoring in all the uncertainties in the emulator and in the observed data, the emulator output is too often far from the model output.
- Is 95% of the probability mass concentrated in a considerably smaller interval than  $[-3, 3]$  (say, for example,  $[-0.5, 0.5]$ )? For this to happen, the emulator uncertainty must be quite large. In such case the emulator, being extremely cautious, will cut out a small part of the parameter space and we will end up needing many more waves of history matching than are necessary.
- Is the histogram skewing significantly in one direction or the other? If this is the case, the emulator tends to either overestimate or underestimate the model output.

[Return to P27](#)

Remind me of possible stopping criteria

Since this is an iterative process, at the end of each wave we need to decide whether to perform a new wave or to stop. One possible stopping criterion consists of comparing the emulator uncertainty and the target uncertainty. If the former is larger, another wave can be performed, since new, more confident emulators can potentially help further reduce the non-imausible space. If the uncertainty of emulators is smaller than the uncertainty in the targets, improving the performance of emulators would not make a substantial difference, and additional waves would not be beneficial. We may also choose to stop the iterations when we get emulators that provide us with ‘full-fitting’ points at a sufficiently high rate. In such a case, rather than spending time training new emulators, we can generate new points with the function `generate_new_design` using the current emulators until we find enough full-fitting ones. Finally, we might end up with all the input space deemed implausible at the end of a wave. In this situation, we would deduce that there are no parameter sets that give an acceptable match with the data: in particular, this would raise doubts about the adequacy of the chosen model, or input and/or output ranges.

[Return to P50](#)

R tip

[Return to P52](#)

How are bimodal outputs identified?

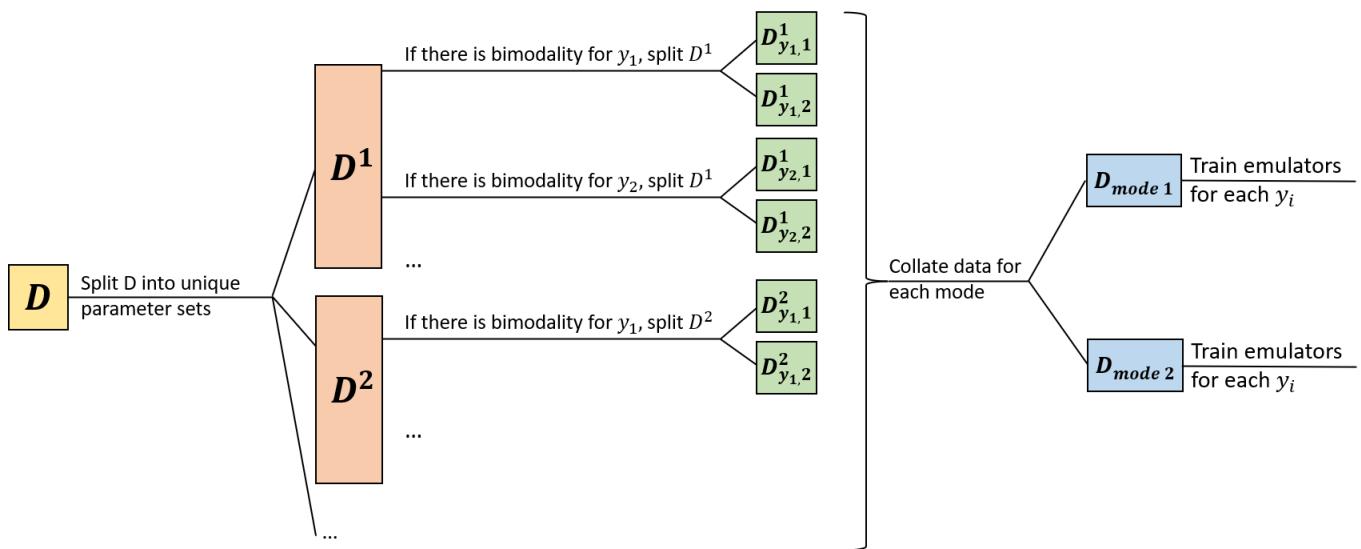
Suppose we are given a dataframe containing parameter sets (with repetitions) and the corresponding outputs. An individual row of that dataframe would be of the form  $(x, y) = (x_1, x_2, x_3, \dots, x_p, y_1, y_2, \dots, y_n)$  where  $x$  are the parameters and  $y$  the outputs. To check and classify bimodality we follow the steps below:

1. We split the dataframe into unique parameter sets, collecting repetitions for each parameter set together. For each unique parameter set  $\bar{x}$  and each  $i = 1, \dots, n$ , we use clustering on  $y_i$  (with finite normal [mixture models](#)), allowing a maximum of two clusters. If the optimal clustering has two clusters, we consider the output  $y_i$  to be bimodal at  $\bar{x}$ ; otherwise we consider it unimodal.
2. If for all parameter sets and all outputs, the optimal clustering is a single cluster, then we have no bimodality: we then perform stochastic emulation using `emulator_from_data` with `emulator_type = 'variance'`. Otherwise, for each output  $y_i$  and each unique parameter set  $\bar{x}$ :
  - 2.1. If there is no bimodality in  $y_i$  for  $\bar{x}$ , then all the data from  $\bar{x}$  for  $y_i$  will be used in each mode.
  - 2.2. If there is bimodality in  $y_i$  for  $\bar{x}$ , then we determine which of the repetitions in the parameter set  $\bar{x}$  correspond to the identified bimodality (again, using clustering). The dataset for that parameter set  $\bar{x}$  gets split into two:

one for each mode.

3. We collate the data for each mode, resulting in (for each output) two dataframes of repetitions. Note that it is unlikely that each mode gets the same number of repetitions to train to.
4. For each mode, we train a list of emulators (one for each output) using the data appropriate to that mode. Note that there is a slight redundancy here, in that if there was no bimodality in an output for any parameter choices, then we have essentially trained the same emulator twice.

The process described above is schematically represented in the figure below:



[Return to P57](#)