

Optimization of Stochastic Epidemiological Models for Disease Control and Prediction

Dr. Khalil Al Handawi

Joint work with: Prof. Michael Kokkolaras

Journées de l'optimisation 2022

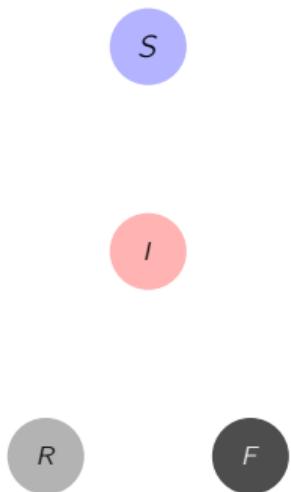
July 24, 2023



Background: epidemiological models

What are compartmental epidemiological models?

S susceptible I infected R recovered F fatality



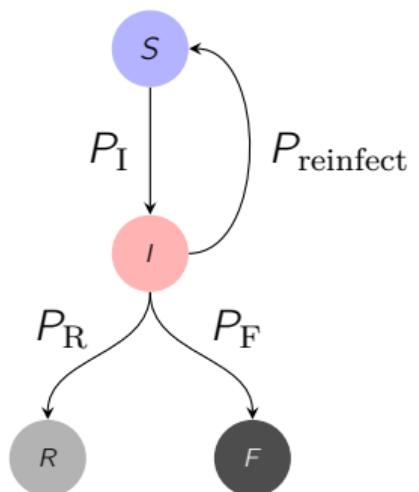


Background: epidemiological models

What are compartmental epidemiological models?

S susceptible I infected R recovered F fatality

- Described by a *stochastic* process

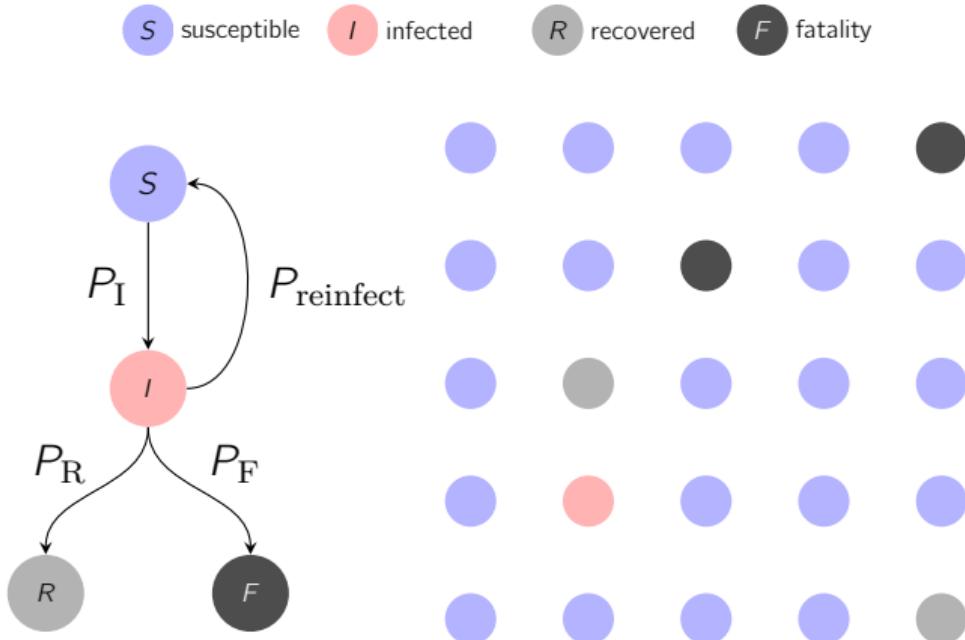




Background: epidemiological models

What are compartmental epidemiological models?

- Described by a *stochastic* process



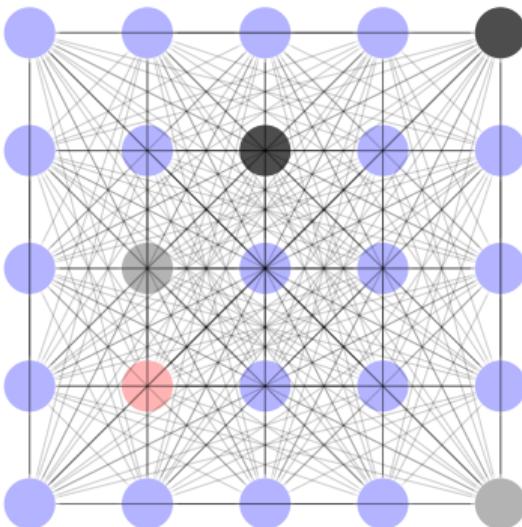
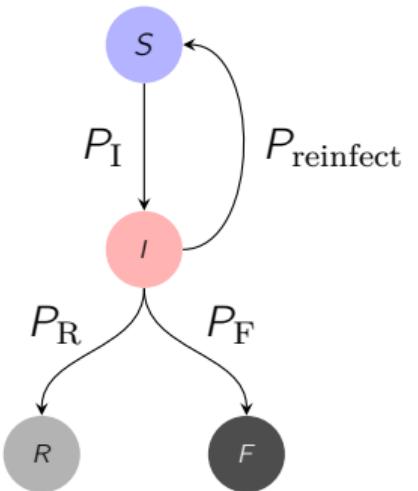


Background: epidemiological models

What are compartmental epidemiological models?

- Described by a *stochastic* process
- Assumes *homogenous* interaction

S susceptible I infected R recovered F fatality

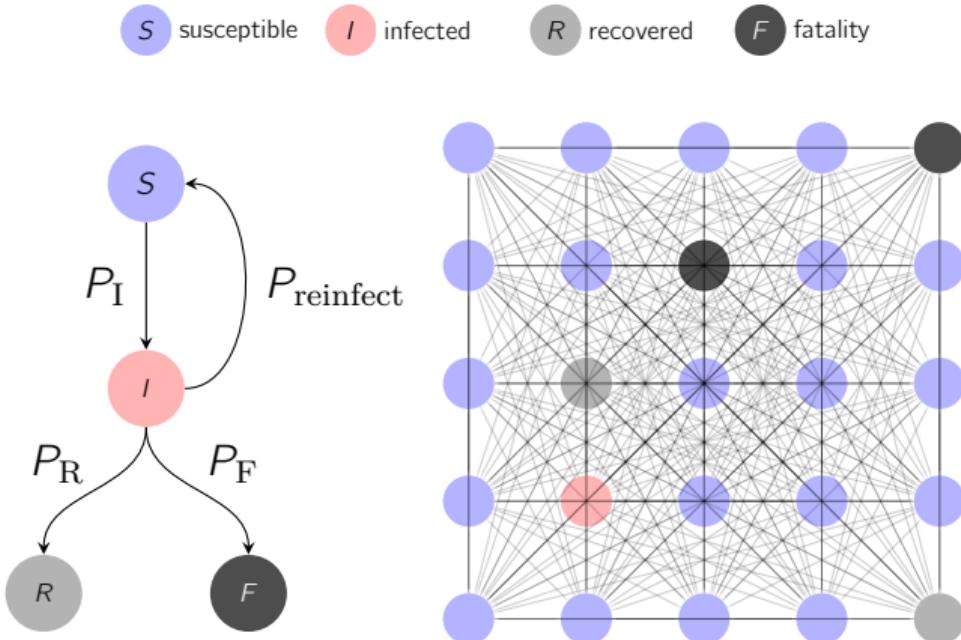




Background: epidemiological models

What are compartmental epidemiological models?

- Described by a *stochastic* process
- Assumes *homogenous* interaction
- Deterministic response for large N



Background: epidemiological models

What are compartmental epidemiological models?

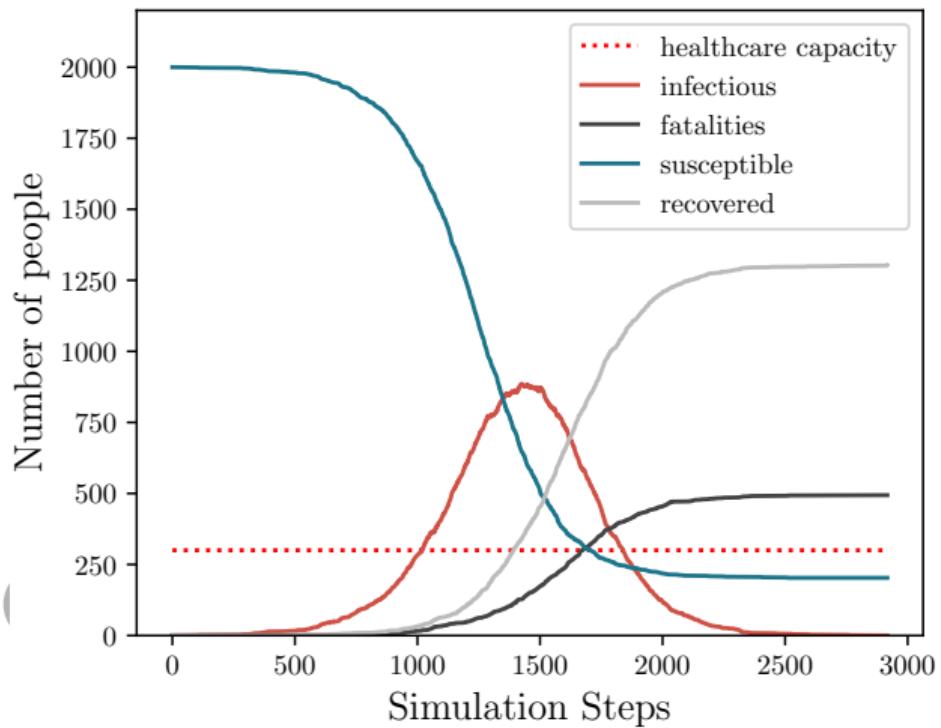
- Described by a *stochastic* process
- Assumes *homogenous* interaction
- Deterministic response for large N

$$\frac{dS}{dt} = -\frac{\beta IS}{N},$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I,$$

$$\frac{dR}{dt} = \gamma I,$$

where $N = S + I + R$, β controls infection spread, and γ controls recovery rate



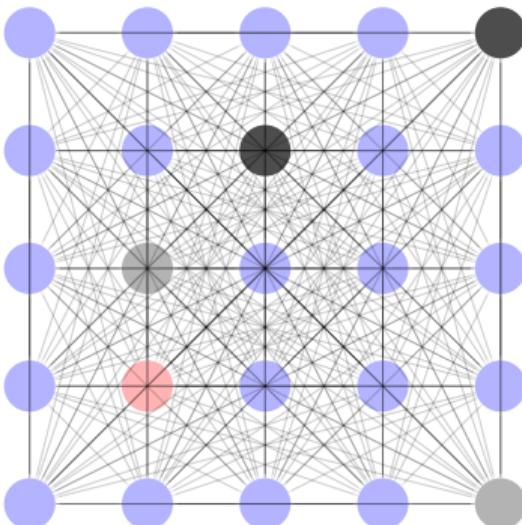
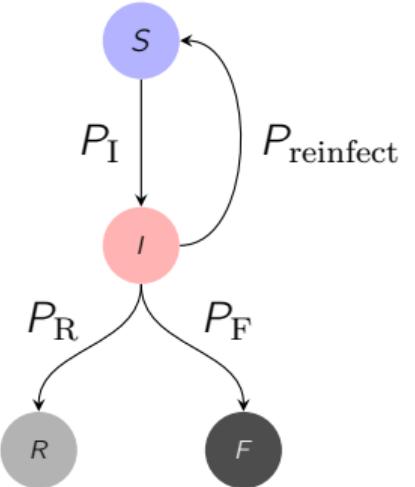


Background: epidemiological models

What are compartmental epidemiological models?

- ✓ Analytical solutions are available

S susceptible I infected R recovered F fatality



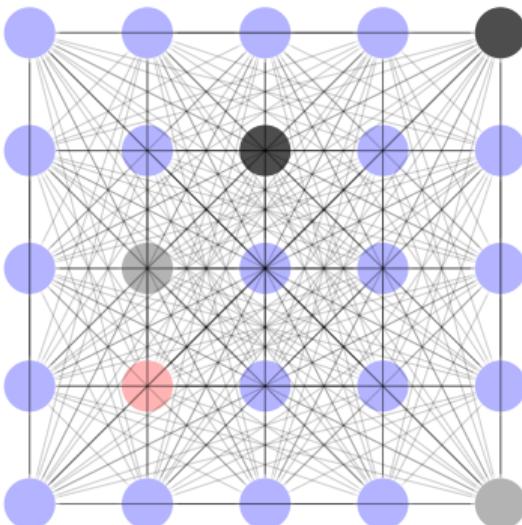
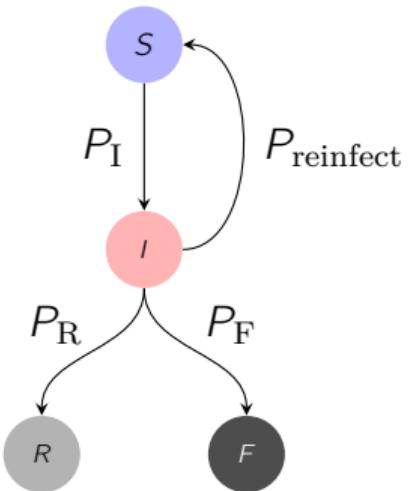


Background: epidemiological models

What are compartmental epidemiological models?

- ✓ Analytical solutions are available
- ✓ Captures large-scale population dynamics

S susceptible I infected R recovered F fatality

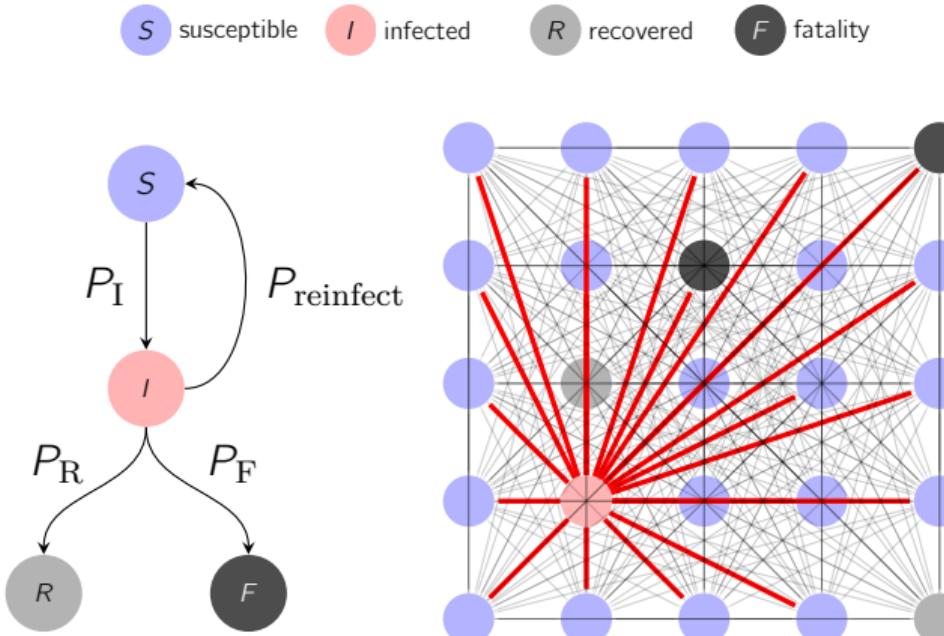




Background: epidemiological models

What are compartmental epidemiological models?

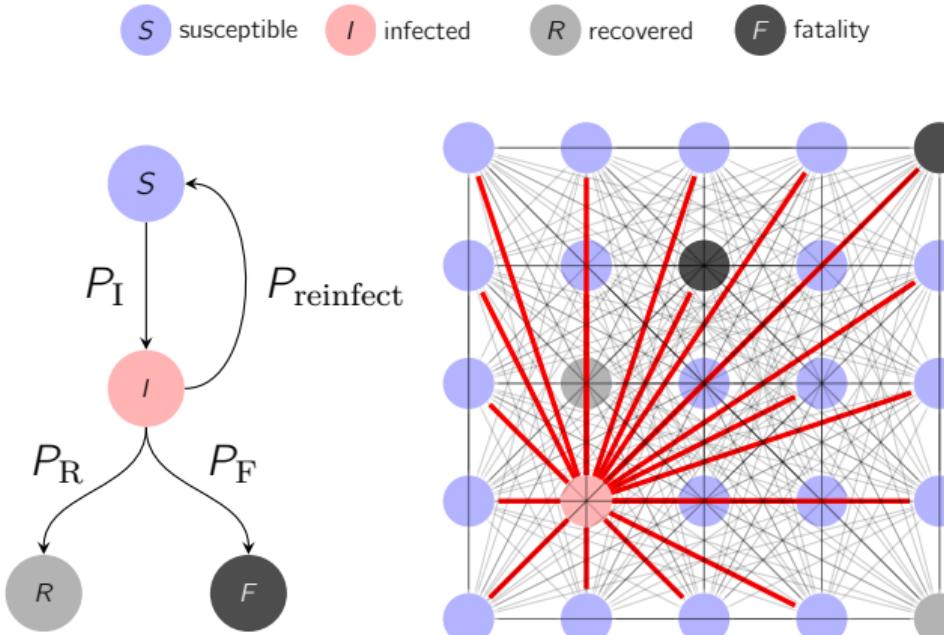
- ✓ Analytical solutions are available
- ✓ Captures large-scale population dynamics
- ✗ Does not account for *geography*



Background: epidemiological models

What are compartmental epidemiological models?

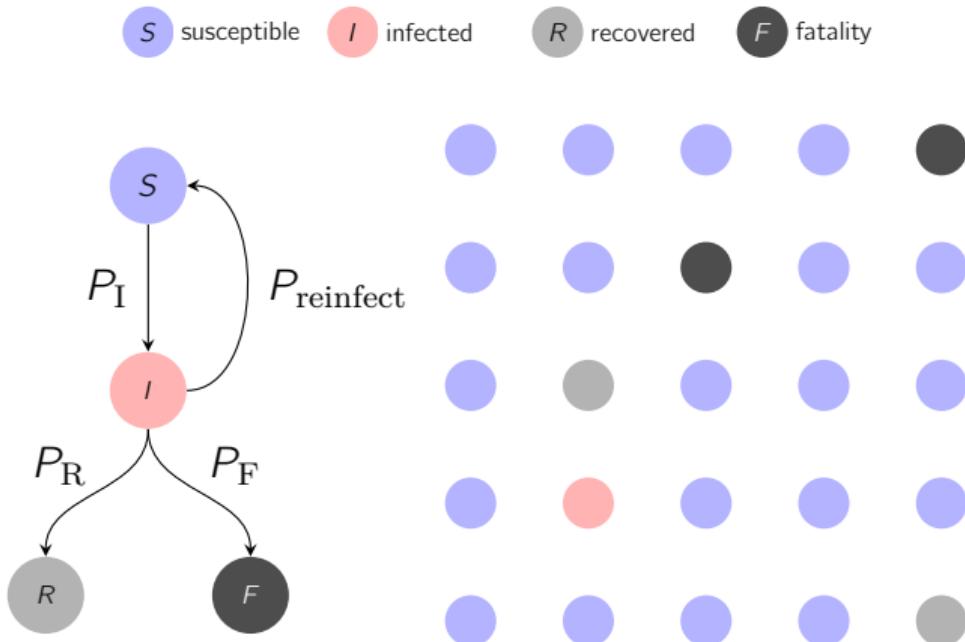
- ✓ Analytical solutions are available
- ✓ Captures large-scale population dynamics
- ✗ Does not account for *geography*
- ✗ Cannot model effect of intervention policies





Background: epidemiological models

What are agent-based epidemiological models?



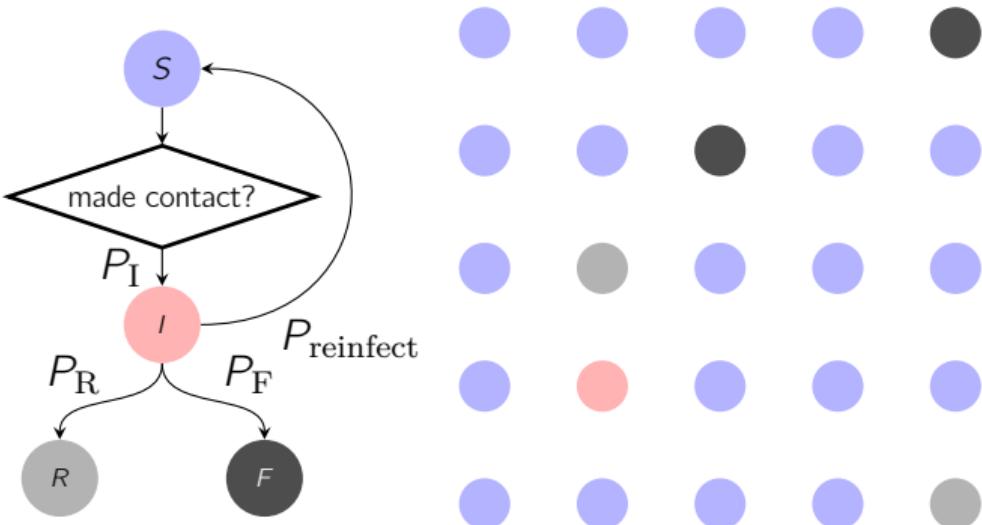


Background: epidemiological models

What are agent-based epidemiological models?

S susceptible I infected R recovered F fatality random event

- *Stochastic process*



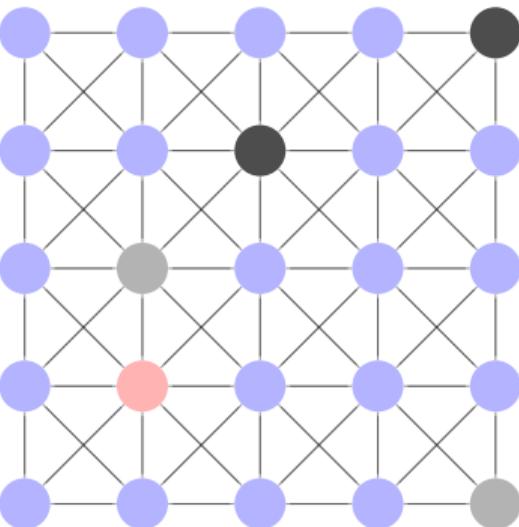
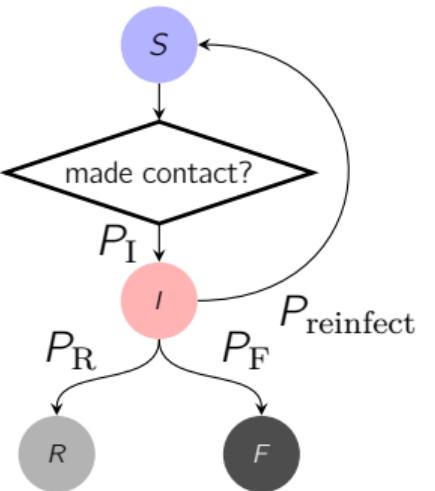


Background: epidemiological models

What are agent-based epidemiological models?

- *Stochastic process*
- Assume *heterogenous* interaction

S susceptible I infected R recovered F fatality \diamond random event

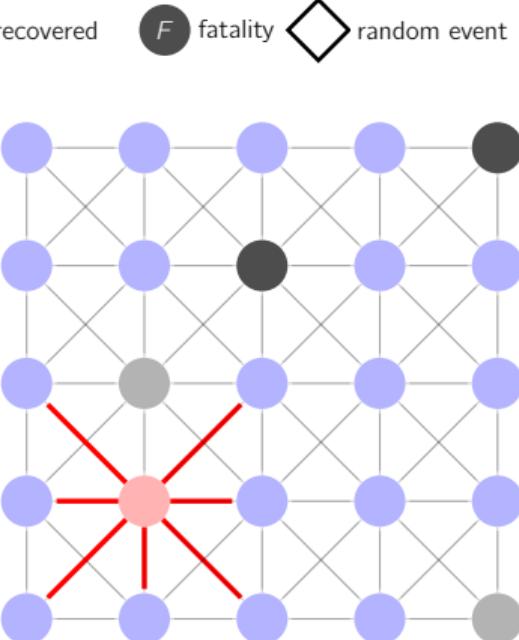
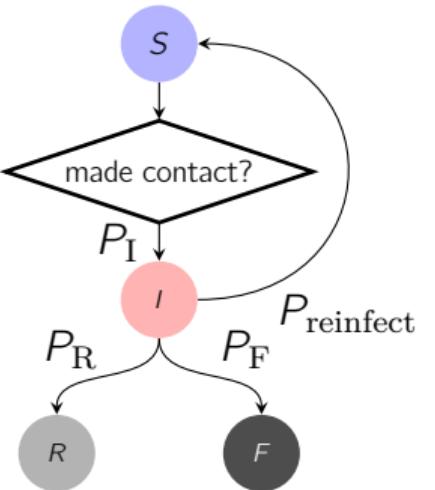


Background: epidemiological models

What are agent-based epidemiological models?

- *Stochastic process*
- Assume *heterogenous* interaction

 susceptible  infected  recovered  fatality  random event





Background: epidemiological models

What are agent-based epidemiological models?

Realization 1

- *Stochastic* process
- Assume *heterogenous* interaction
- Stochastic response



Background: epidemiological models

What are agent-based epidemiological models?

Realization 2

- *Stochastic* process
- Assume *heterogenous* interaction
- Stochastic response

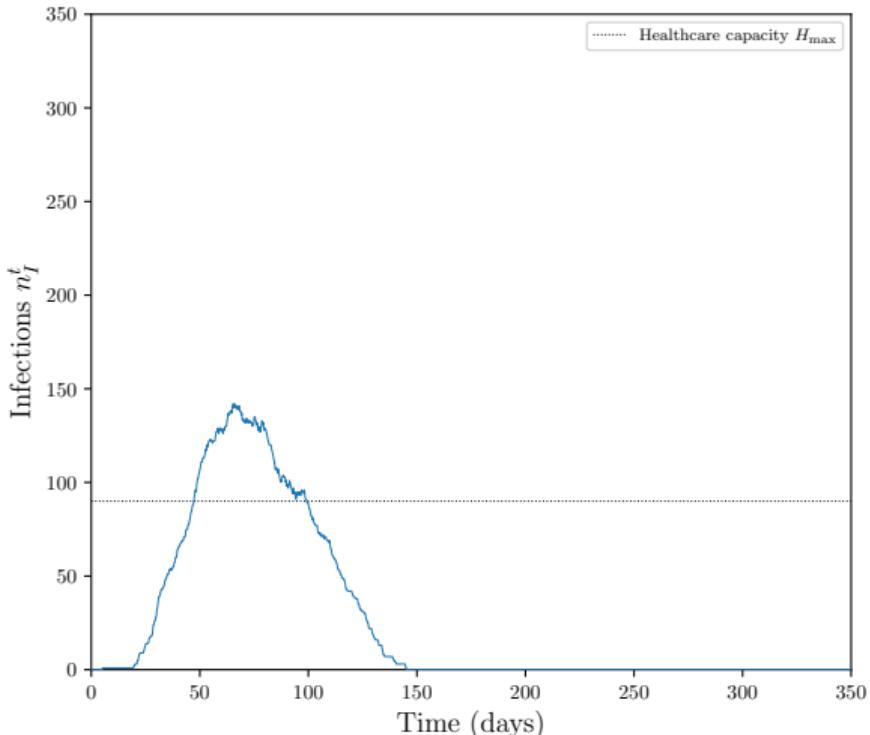


Background: epidemiological models

What are agent-based epidemiological models?

- *Stochastic* process
- Assume *heterogenous* interaction
- Stochastic response

Realization 1



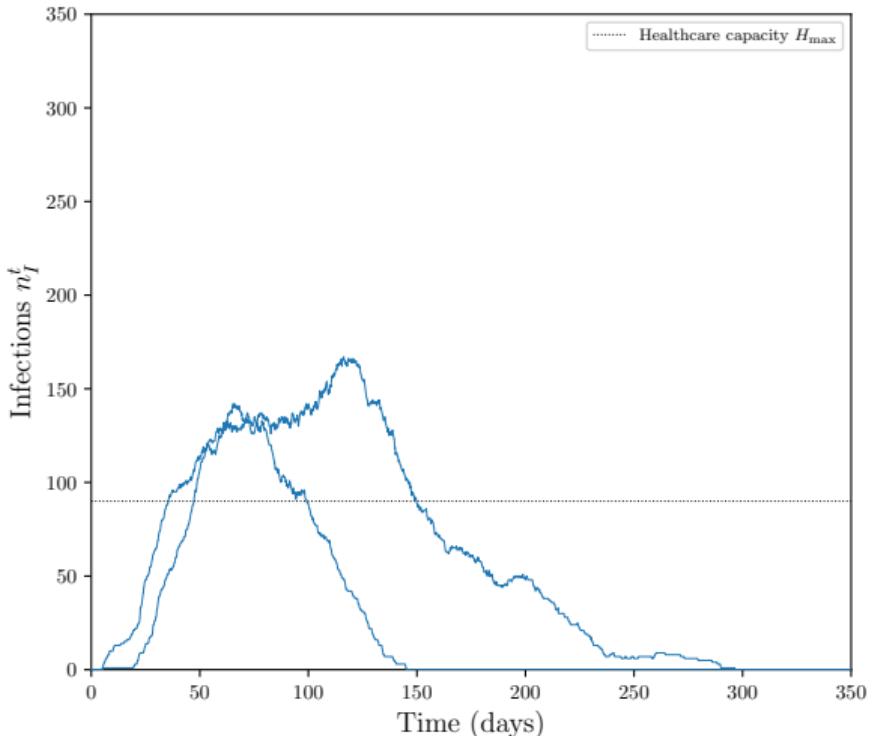


Background: epidemiological models

What are agent-based epidemiological models?

- *Stochastic process*
- Assume *heterogenous* interaction
- Stochastic response

Realization 2

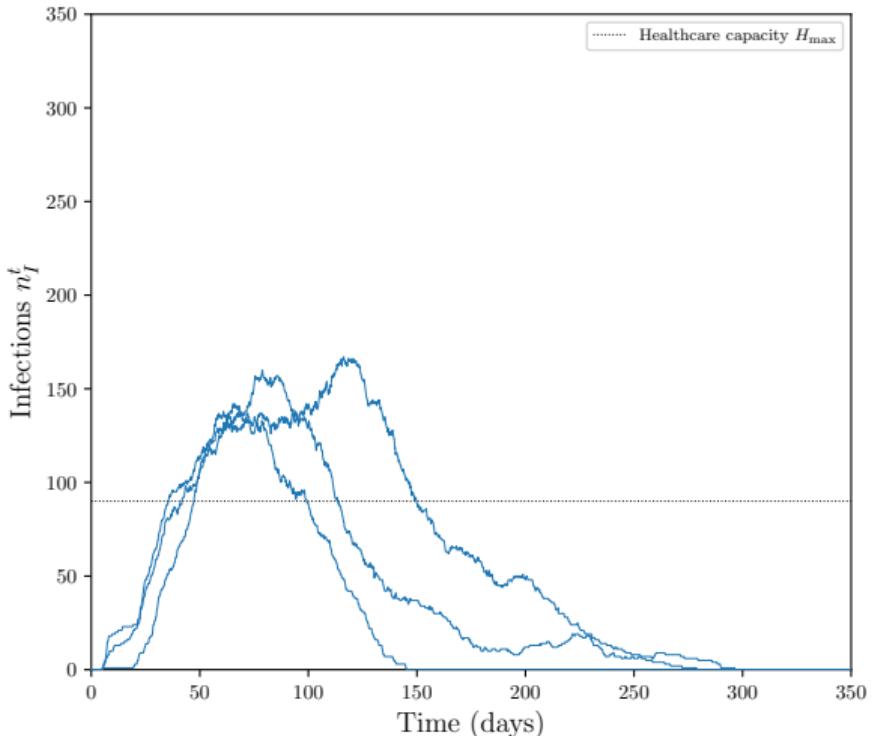




Background: epidemiological models

What are agent-based epidemiological models?

- *Stochastic process*
- Assume *heterogenous* interaction
- Stochastic response

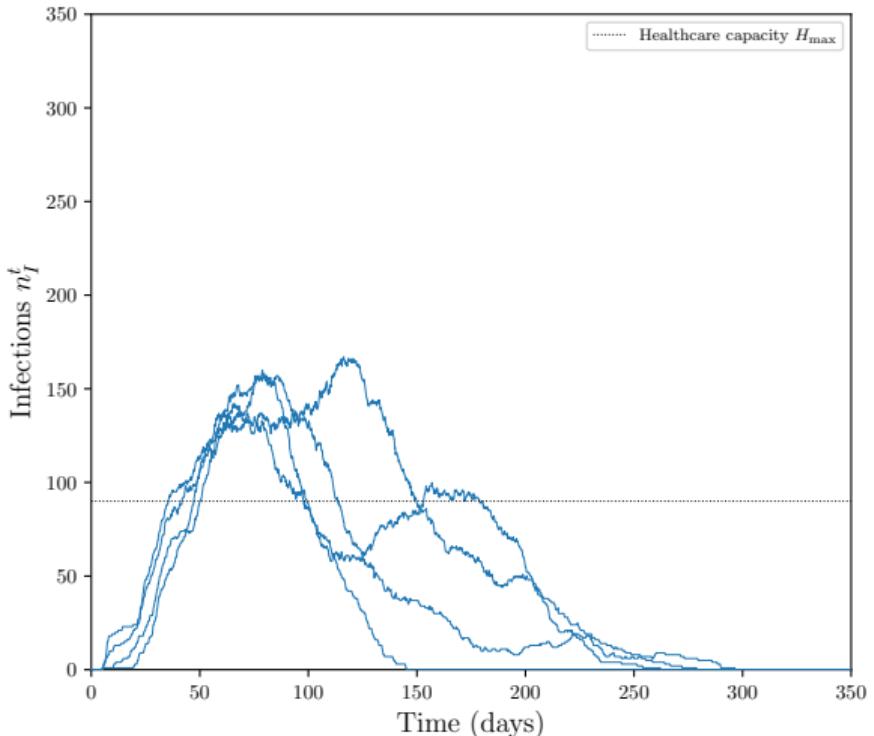




Background: epidemiological models

What are agent-based epidemiological models?

- *Stochastic process*
- Assume *heterogenous* interaction
- Stochastic response

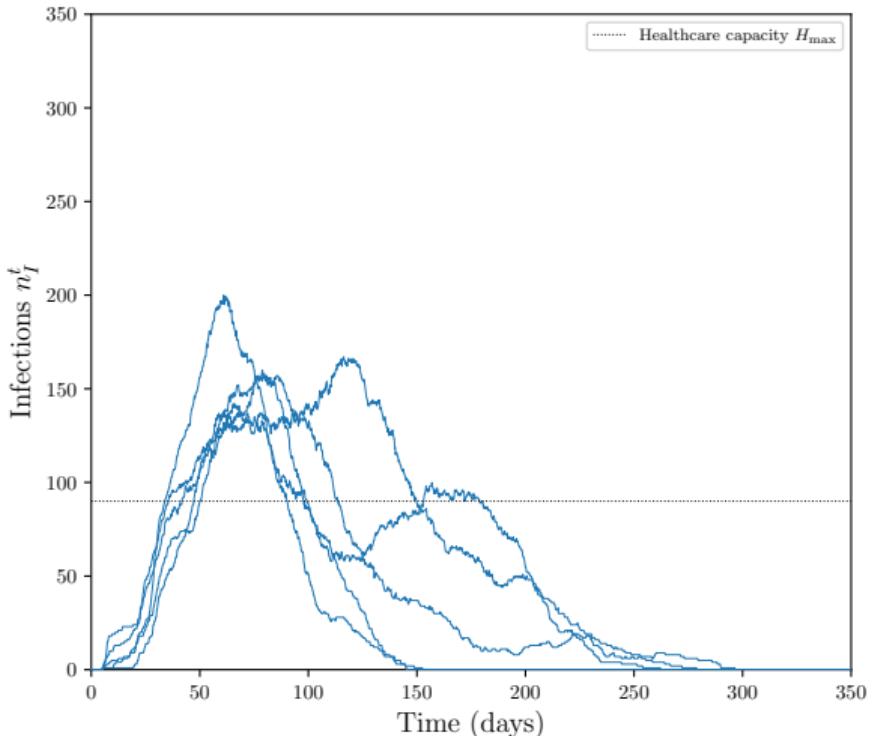




Background: epidemiological models

What are agent-based epidemiological models?

- *Stochastic process*
- Assume *heterogenous* interaction
- Stochastic response

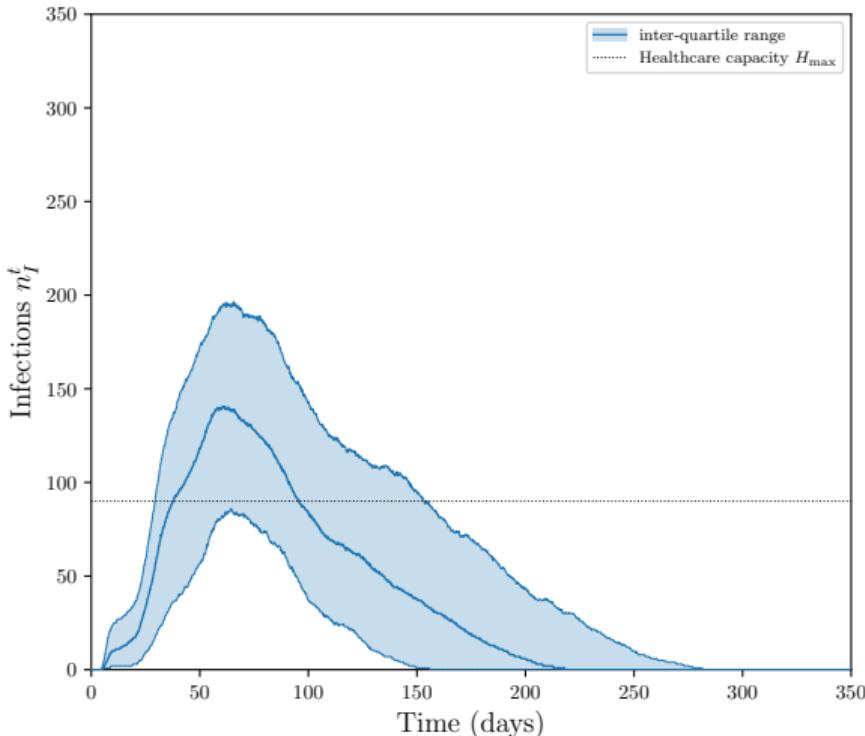




Background: epidemiological models

What are agent-based epidemiological models?

- *Stochastic process*
- Assume *heterogenous* interaction
- Stochastic response



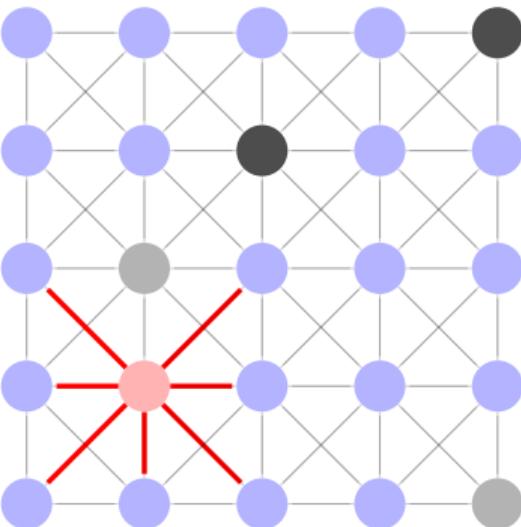
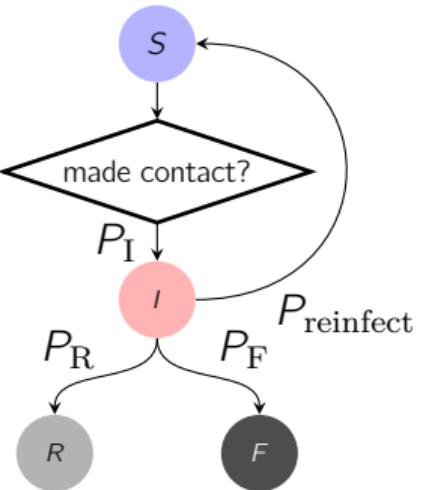


Background: epidemiological models

What are agent-based epidemiological models?

S susceptible I infected R recovered F fatality \diamond random event

- ✓ Account for *geography* and *demographics*

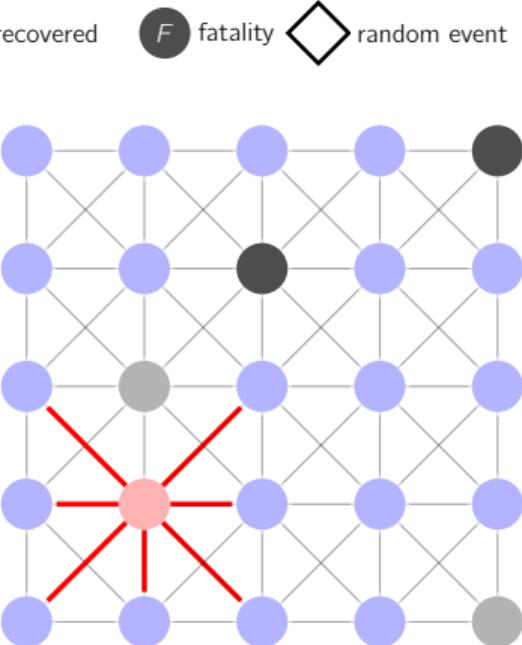
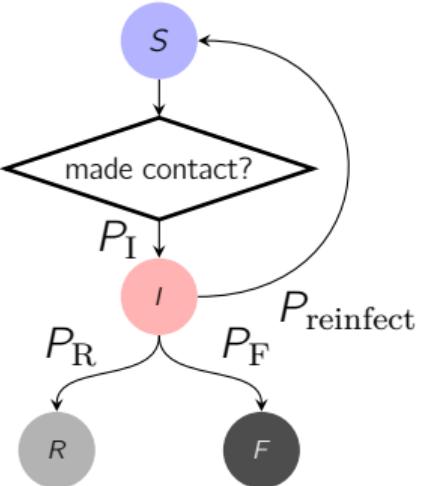


Background: epidemiological models

What are agent-based epidemiological models?



- ✓ Account for *geography* and *demographics*
- ✓ Describe local phenomena



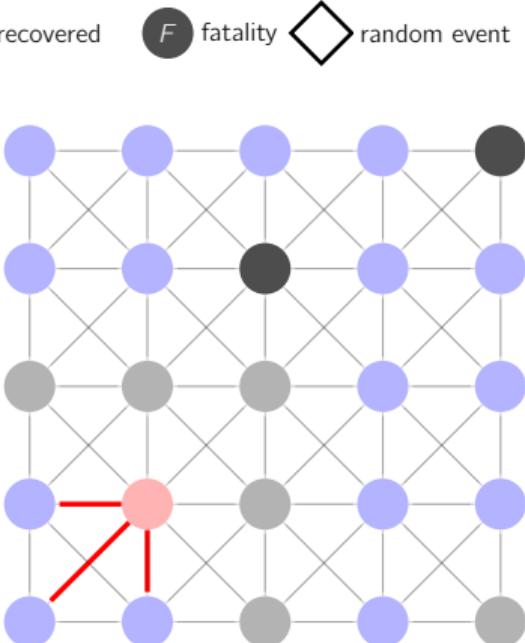
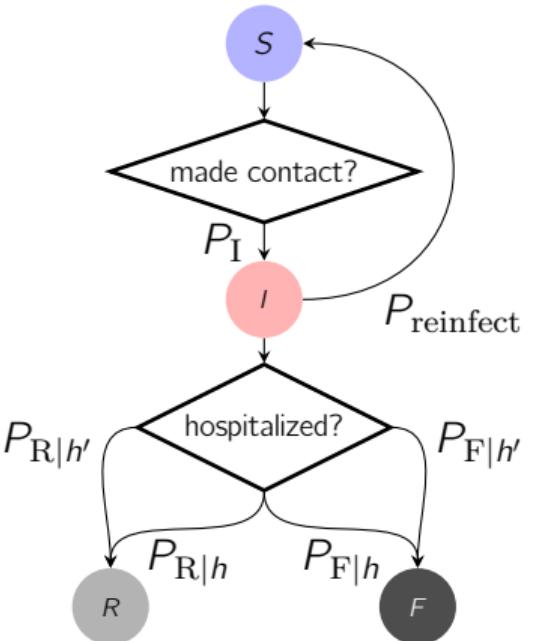


Background: epidemiological models

What are agent-based epidemiological models?

- ✓ Account for *geography* and *demographics*
- ✓ Describe local phenomena
- ✓ Can be used to model intervention policies

susceptible infected recovered fatality random event



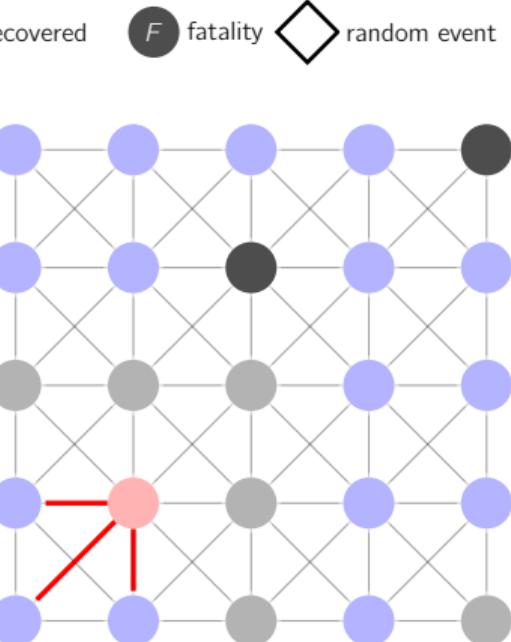
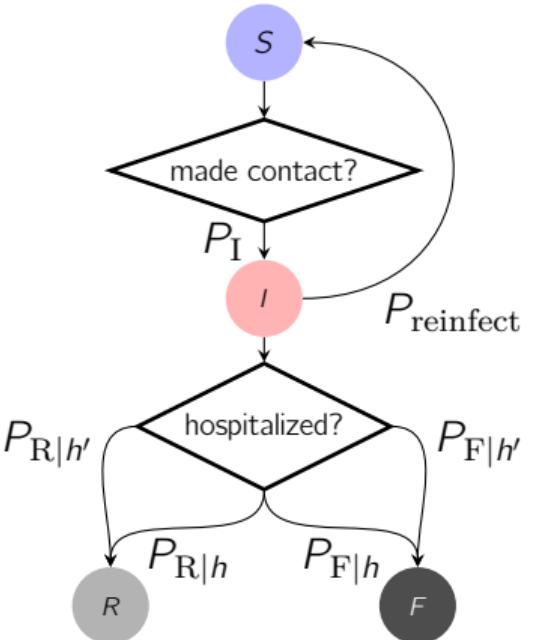


Background: epidemiological models

What are agent-based epidemiological models?

- ✓ Account for *geography* and *demographics*
- ✓ Describe local phenomena
- ✓ Can be used to model intervention policies
- ✗ Stochastic response makes decision making challenging

susceptible infected recovered fatality random event



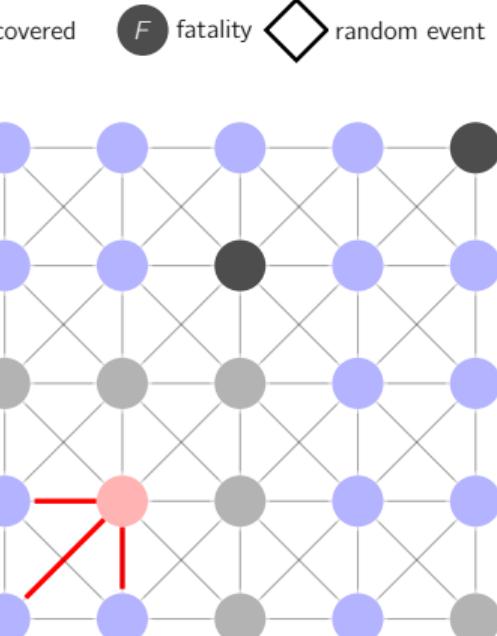
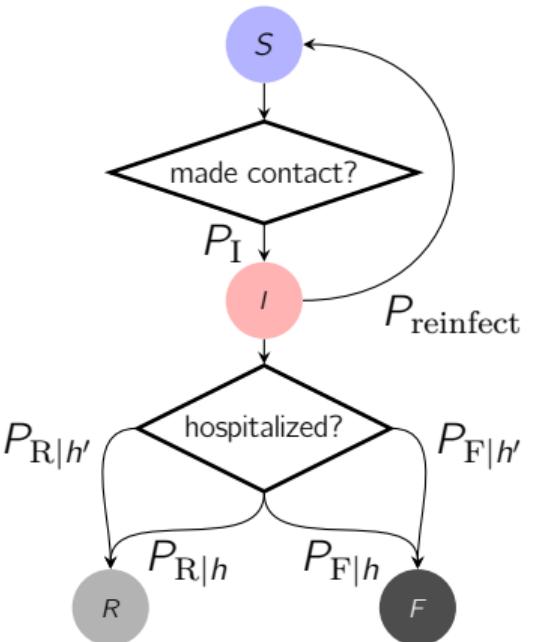


Background: epidemiological models

What are agent-based epidemiological models?

- ✓ Account for *geography* and *demographics*
- ✓ Describe local phenomena
- ✓ Can be used to model intervention policies
- ✗ Stochastic response makes decision making challenging
- ✗ Computationally expensive

susceptible infected recovered fatality random event



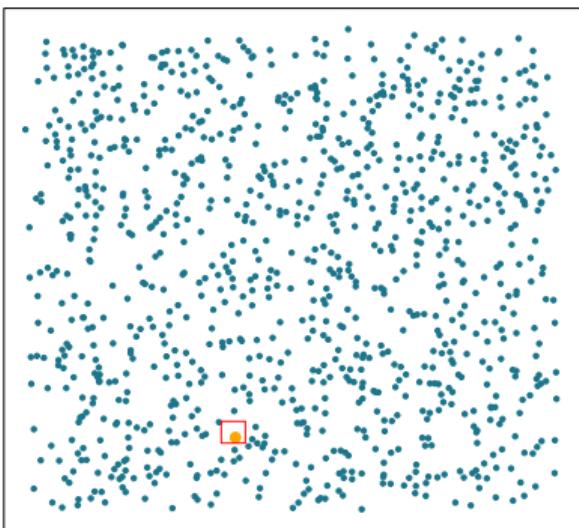


Public health policy-making problem formulation

What is the **cost** of public health interventions?

No interventions applied

- susceptible
- infected
- recovered
- fatalities
- tracked



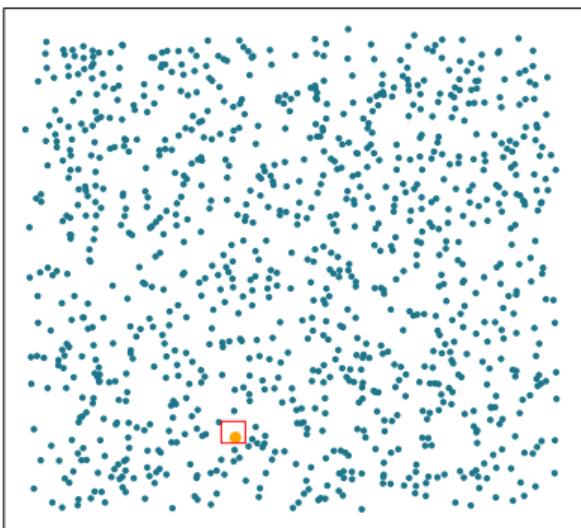
Public health policy-making problem formulation



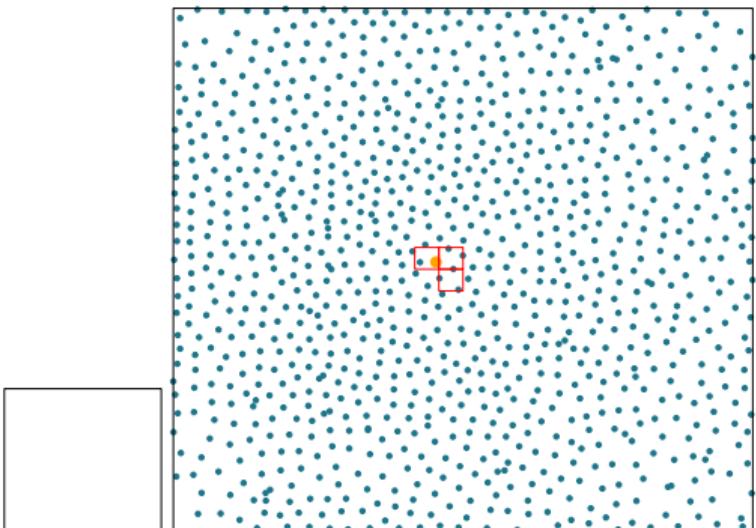
What is the **cost** of public health interventions?

No interventions applied

• susceptible • infected • recovered • fatalities • tracked



• susceptible • infected • recovered • fatalities • tracked



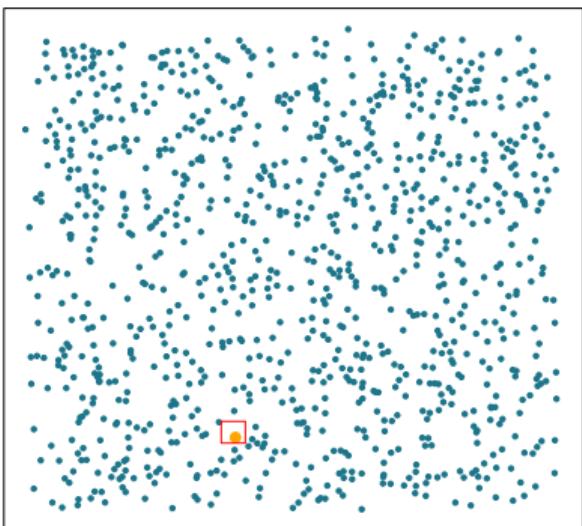
Public health policy-making problem formulation



What is the **cost** of public health interventions?

No interventions applied

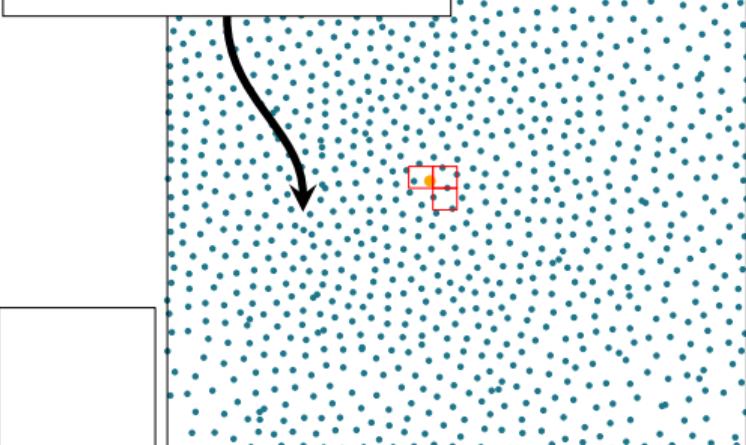
• susceptible • infected • recovered • fatalities • tracked



with intervention

• susceptible • infected • recovered • fatalities • tracked

essential workers



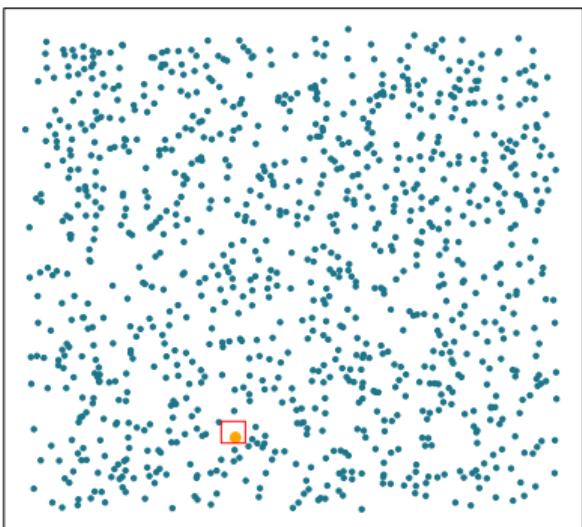
Public health policy-making problem formulation



What is the **cost** of public health interventions?

No interventions applied

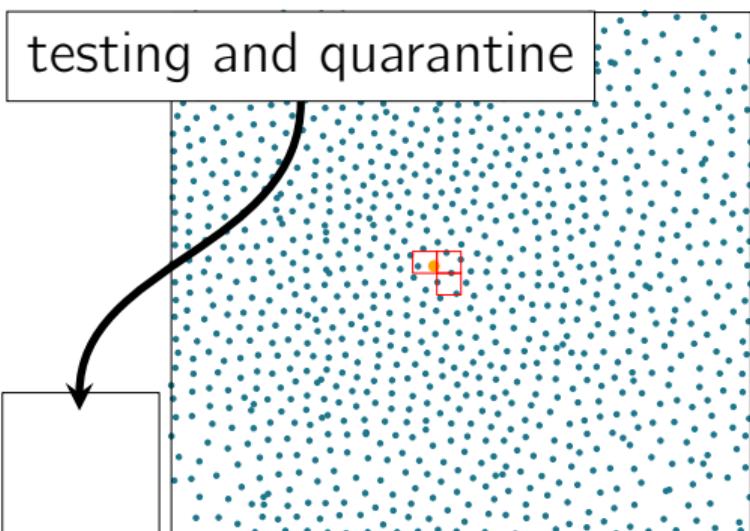
• susceptible • infected • recovered • fatalities • tracked



with intervention

• susceptible • infected • recovered • fatalities • tracked

testing and quarantine



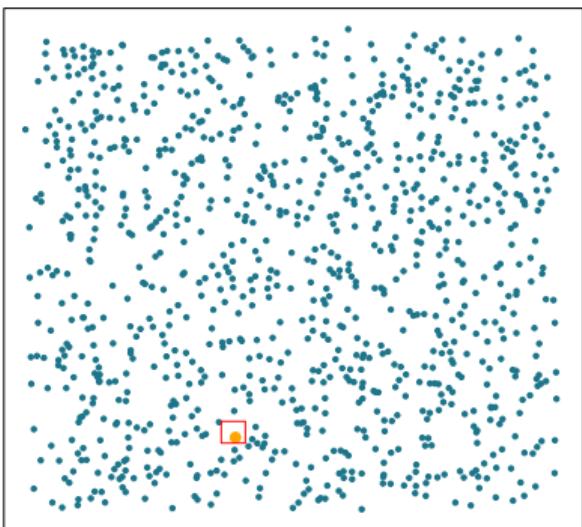
Public health policy-making problem formulation



What is the **cost** of public health interventions?

No interventions applied

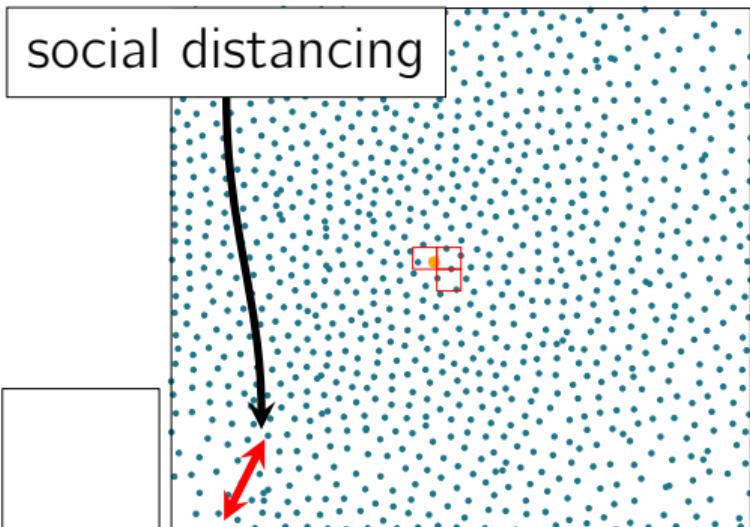
• susceptible • infected • recovered • fatalities • tracked



with intervention

• susceptible • infected • recovered • fatalities • tracked

social distancing





What is the **cost** of public health interventions?

No interventions applied

with intervention

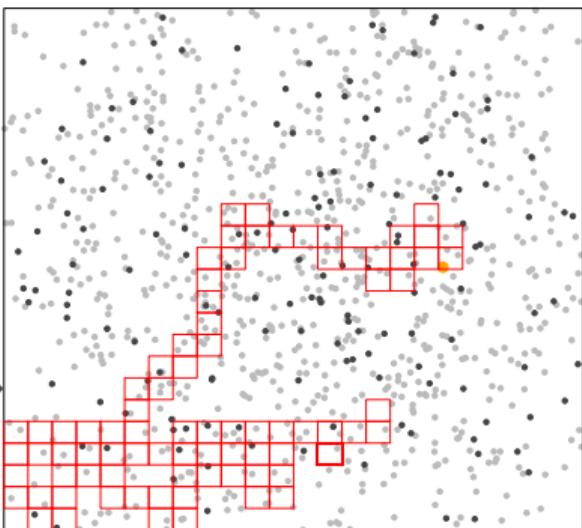
Public health policy-making problem formulation



What is the **cost** of public health interventions?

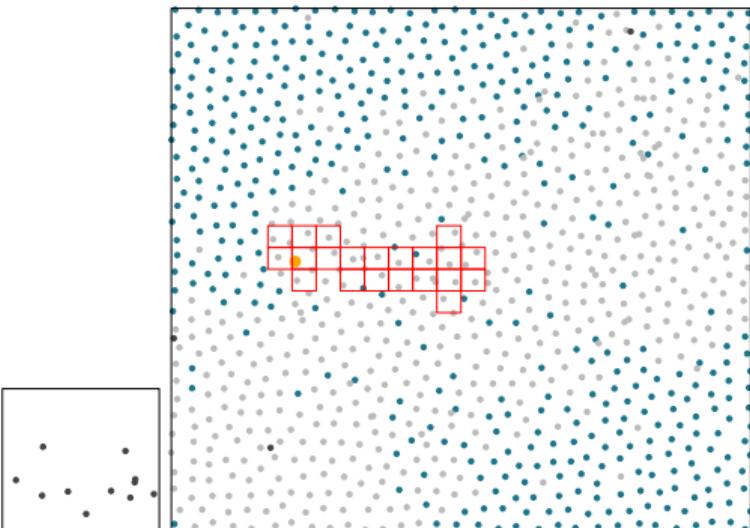
No interventions applied

• susceptible • infected • recovered • fatalities • tracked



with intervention

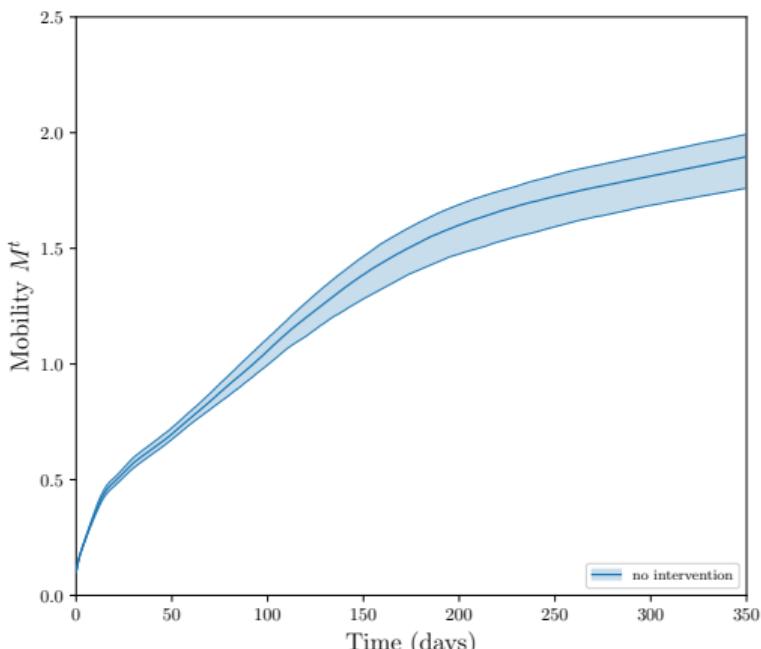
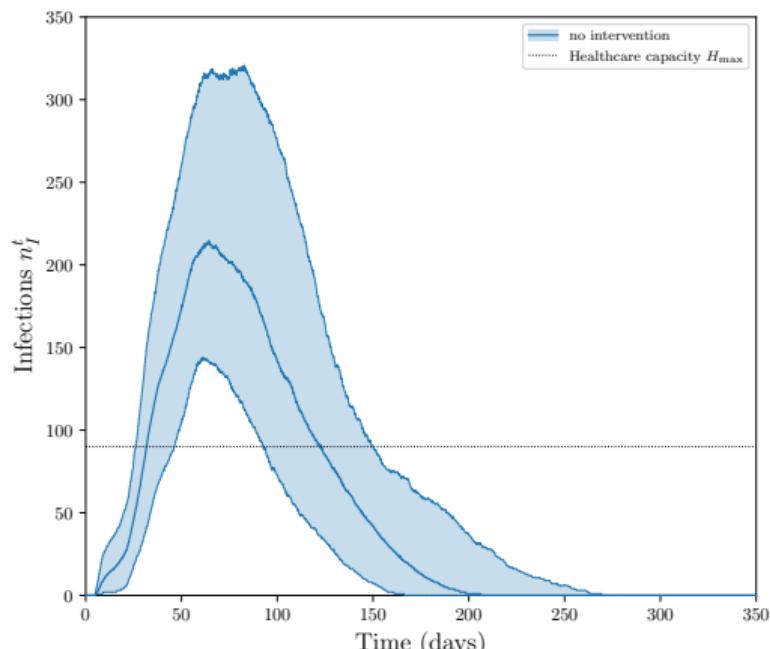
• susceptible • infected • recovered • fatalities • tracked





Public health policy-making problem formulation

What is the **cost** of public health interventions?

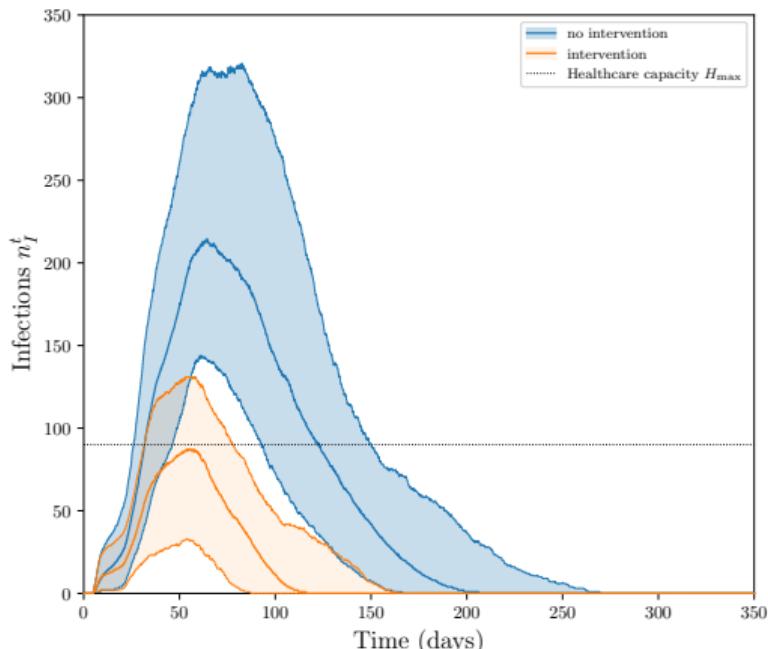




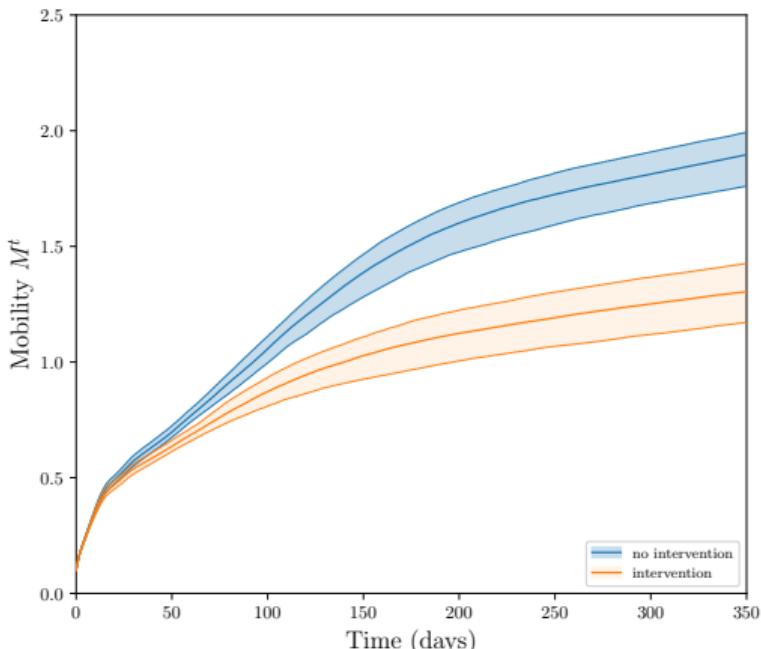
Public health policy-making problem formulation

What is the **cost** of public health interventions?

infections ↓



mobility ↓





Optimization problem

Objective and constraints

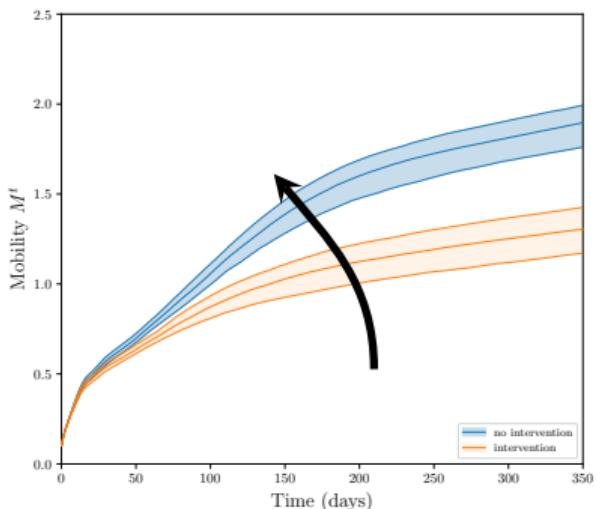
$$\min_x \quad f(x) = -M^T$$

subject to

where $x = [n_E, S_D, n_T]^T$

Design variables

- n_E : Number of essential workers
- S_D : Social distancing factor
- n_T : Number of tests daily



Optimization problem

No gradient information available, blackbox is expensive and noisy

Objective and constraints

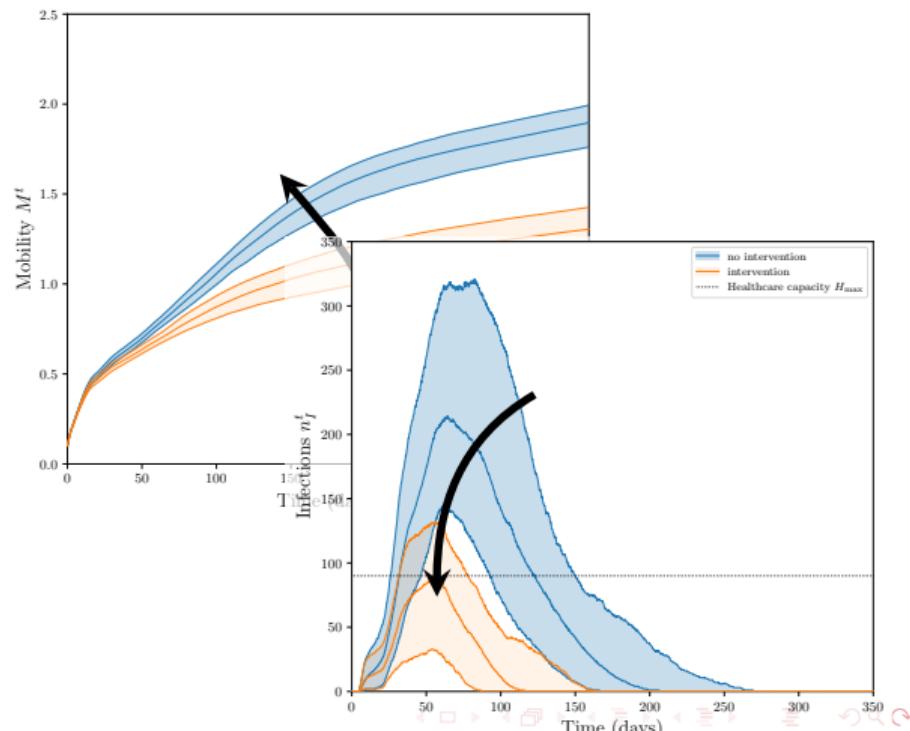
$$\min_x \quad f(x) = -M^T$$

$$\text{subject to} \quad c(x) \equiv n_{I,\max} - H_{\max} \leq 0$$

$$\text{where} \quad x = [n_E, S_D, n_T]^T$$

Design variables

- n_E : Number of essential workers
- S_D : Social distancing factor
- n_T : Number of tests daily



Optimization problem

No gradient information available, blackbox is expensive and **noisy**

Objective and constraints

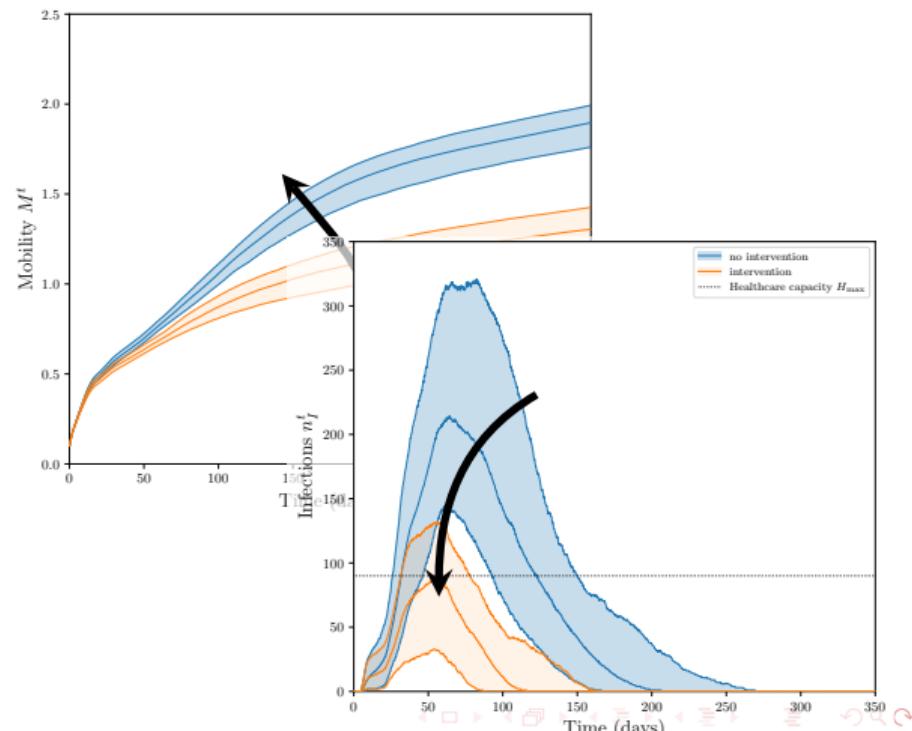
$$\min_x \quad f(x) = -M^T$$

$$\text{subject to} \quad c(x) \equiv n_{I,\max} - H_{\max} \leq 0$$

$$\text{where} \quad x = [n_E, S_D, n_T]^T$$

Design variables

- n_E : Number of essential workers
- S_D : Social distancing factor
- n_T : Number of tests daily



Optimization problem

No gradient information available, blackbox is expensive and **noisy**

Objective and constraints

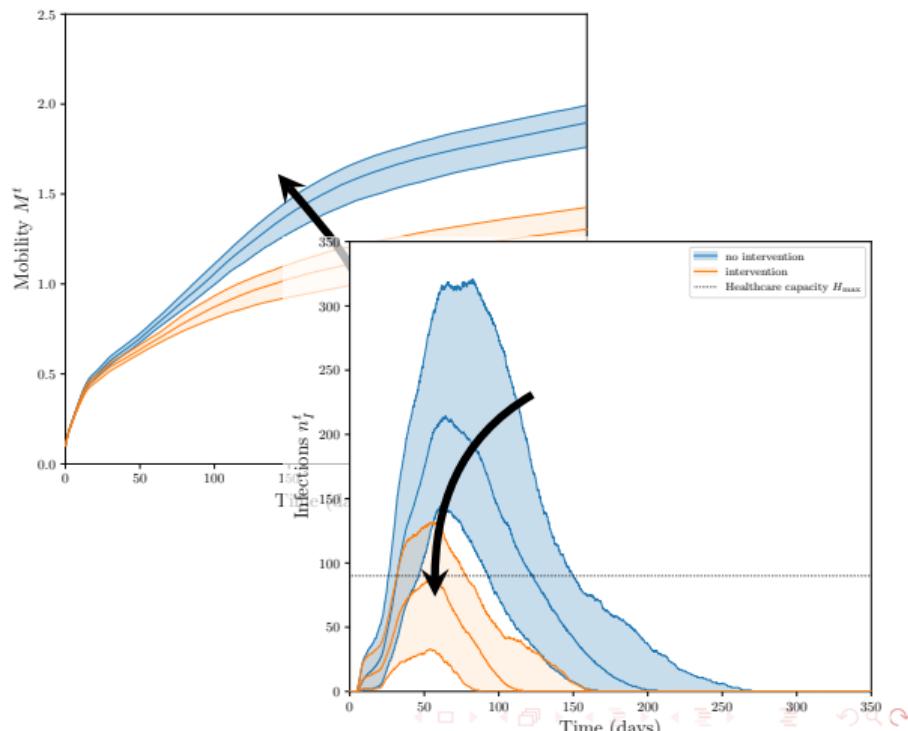
$$\min_x \quad f(x) = \mathbb{E}_{\Theta} [f_{\Theta}(x) = -M^T]$$

$$\text{subject to} \quad c(x) = \mathbb{E}_{\Theta} [c_{\Theta}(x) \equiv n_{I,\max} - H_{\max}] \leq 0$$

where $x = [n_E, S_D, n_T]^T$, Θ : realizations

Design variables

- n_E : Number of essential workers
- S_D : Social distancing factor
- n_T : Number of tests daily



Optimization problem

No gradient information available, blackbox is expensive and **noisy**

Objective and constraints

$$\min_x \quad f(x) = \mathbb{E}_{\Theta} [f_{\Theta}(x) = -M^T]$$

$$\text{subject to} \quad c(x) = \mathbb{E}_{\Theta} [c_{\Theta}(x) \equiv n_{I,\max} - H_{\max}] \leq 0$$

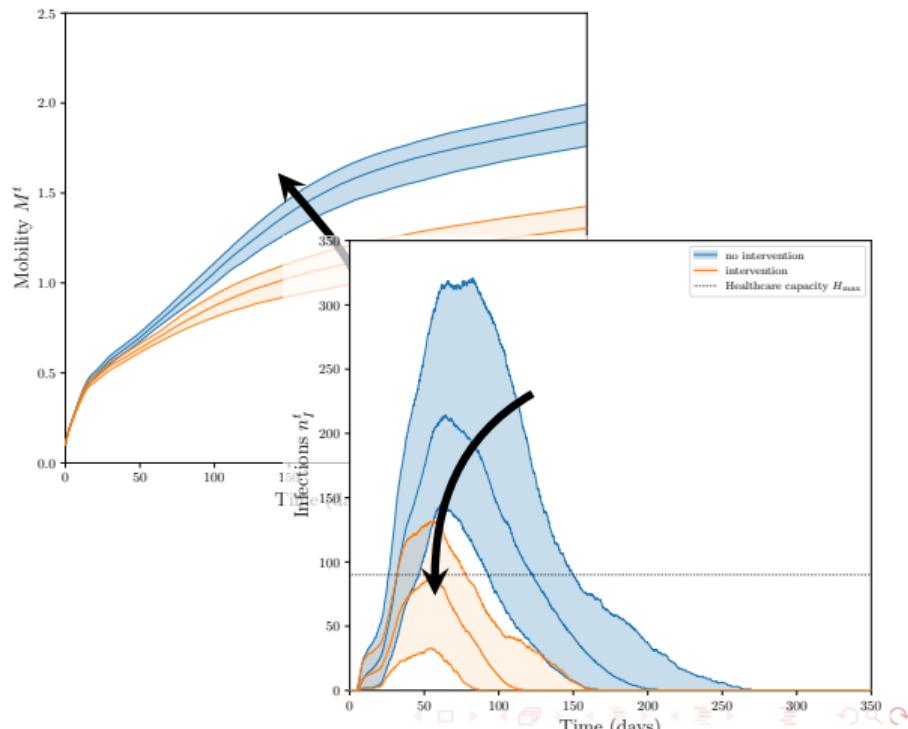
where $x = [n_E, S_D, n_T]^T$, Θ : realizations

Design variables

- n_E : Number of essential workers
- S_D : Social distancing factor
- n_T : Number of tests daily

Randomly seeded parameters

- Initial conditions
- Interactions, demographics





Overview of stochastic mesh adaptive direct search

No gradient information available, blackbox is expensive and **noisy**¹

Objective and constraints

$$\min_x \quad f(x) = \mathbb{E}_{\Theta} [f_{\Theta}(x)]$$

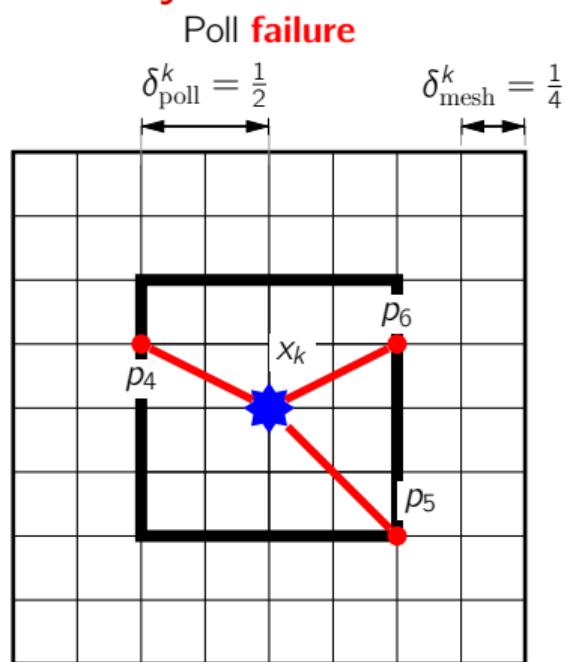
$$\text{subject to} \quad c(x) = \mathbb{E}_{\Theta} [c_{\Theta}(x)] \leq 0$$

where x : variables Θ : realizations

- Constructs estimates of objective:

$$f^k = \frac{1}{n^k} \sum_{i=1}^{n^k} f_{\Theta_0,i}(x_k)$$

- n^k is the sampling rate





Overview of stochastic mesh adaptive direct search

No gradient information available, blackbox is expensive and **noisy**¹

Objective and constraints

$$\min_x \quad f(x) = \mathbb{E}_{\Theta} [f_{\Theta}(x)]$$

$$\text{subject to} \quad c(x) = \mathbb{E}_{\Theta} [c_{\Theta}(x)] \leq 0$$

where x : variables Θ : realizations

- Constructs estimates of objective:

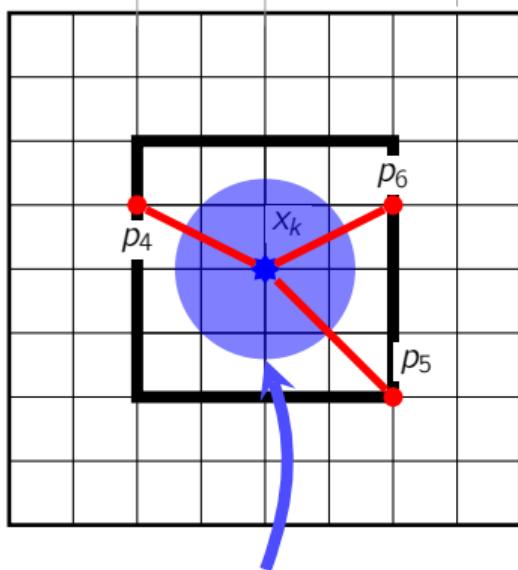
$$f^k = \frac{1}{n^k} \sum_{i=1}^{n^k} f_{\Theta_{0,i}}(x_k)$$

- n^k is the sampling rate
- Tracks uncertainty interval $\mathcal{I}(\delta_{\text{poll}}^k)$ in the estimate

Poll **certain failure**

$$\delta_{\text{poll}}^k = \frac{1}{2}$$

$$\delta_{\text{mesh}}^k = \frac{1}{4}$$



$$\{x : f_s^k - f_0^k \in \mathcal{I}\}$$



Overview of stochastic mesh adaptive direct search

No gradient information available, blackbox is expensive and **noisy**¹

Objective and constraints

$$\min_x \quad f(x) = \mathbb{E}_{\Theta} [f_{\Theta}(x)]$$

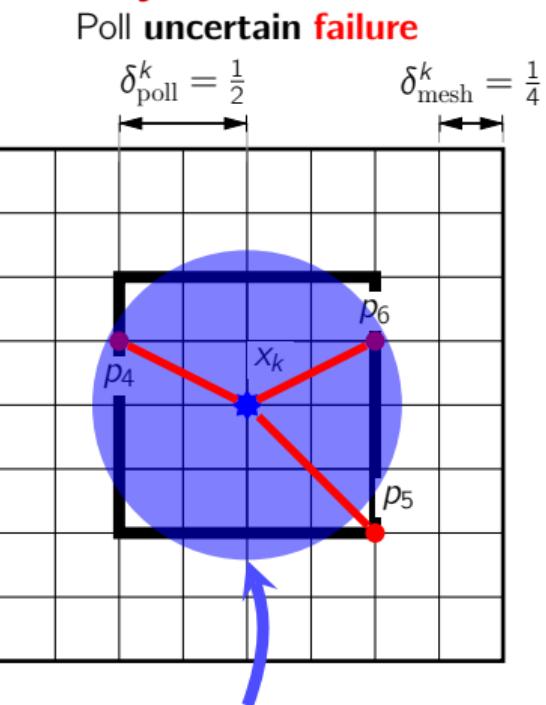
$$\text{subject to} \quad c(x) = \mathbb{E}_{\Theta} [c_{\Theta}(x)] \leq 0$$

where x : variables Θ : realizations

- Constructs estimates of objective:

$$f^k = \frac{1}{n^k} \sum_{i=1}^{n^k} f_{\Theta_{0,i}}(x_k)$$

- n^k is the sampling rate
- Tracks uncertainty interval $\mathcal{I}(\delta_{\text{poll}}^k)$ in the estimate





Overview of stochastic mesh adaptive direct search

No gradient information available, blackbox is expensive and **noisy**¹

Objective and constraints

$$\min_x \quad f(x) = \mathbb{E}_{\Theta} [f_{\Theta}(x)]$$

$$\text{subject to} \quad c(x) = \mathbb{E}_{\Theta} [c_{\Theta}(x)] \leq 0$$

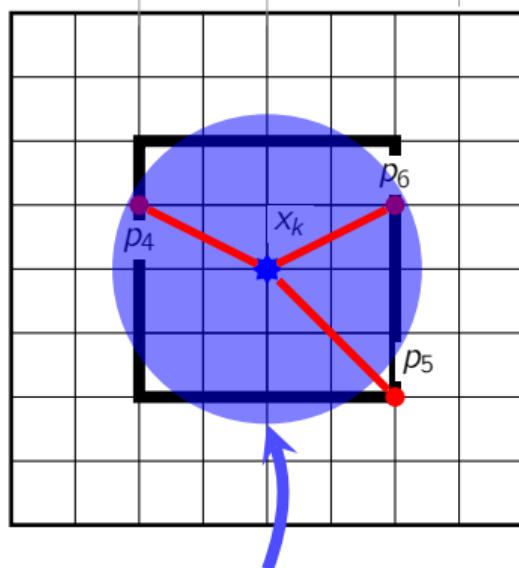
where x : variables Θ : realizations

Constraint handling using the *progressive barrier* approach²

Poll **uncertain failure**

$$\delta_{\text{poll}}^k = \frac{1}{2}$$

$$\delta_{\text{mesh}}^k = \frac{1}{4}$$



$$\{x : f_s^k - f_0^k \in \mathcal{I}\}$$

[1]Audet2020

[2]Dzahini2022



Optimization results

StoMADS, NOMAD¹, and genetic algorithms were used to solve the problem²

Objective and constraints

$$\min_x \quad f(x) = \mathbb{E}_{\Theta} [f_{\Theta}(x) = -M^T]$$

$$\text{subject to} \quad c(x) = \mathbb{E}_{\Theta} [c_{\Theta}(x) \equiv n_{I,\max} - H_{\max}] \leq 0$$

where $x = [n_E, S_D, n_T]^T$, Θ : realizations

Design variables

- n_E : Number of essential workers
- S_D : Social distancing factor
- n_T : Number of tests daily

Randomly seeded parameters

- Initial conditions
- Interactions, demographics

[1]NOMAD

[2]Alhandawi2021b

Optimization results

StoMADS, NOMAD¹, and genetic algorithms were used to solve the problem²

Objective and constraints

$$\min_x \quad f(x) = \mathbb{E}_{\Theta} [f_{\Theta}(x) = -M^T]$$

$$\text{subject to} \quad c(x) = \mathbb{E}_{\Theta} [c_{\Theta}(x) \equiv n_{I,\max} - H_{\max}] \leq 0$$

where $x = [n_E, S_D, n_T]^T$, Θ : realizations

Design variables

- n_E : Number of essential workers
- S_D : Social distancing factor
- n_T : Number of tests daily

Randomly seeded parameters

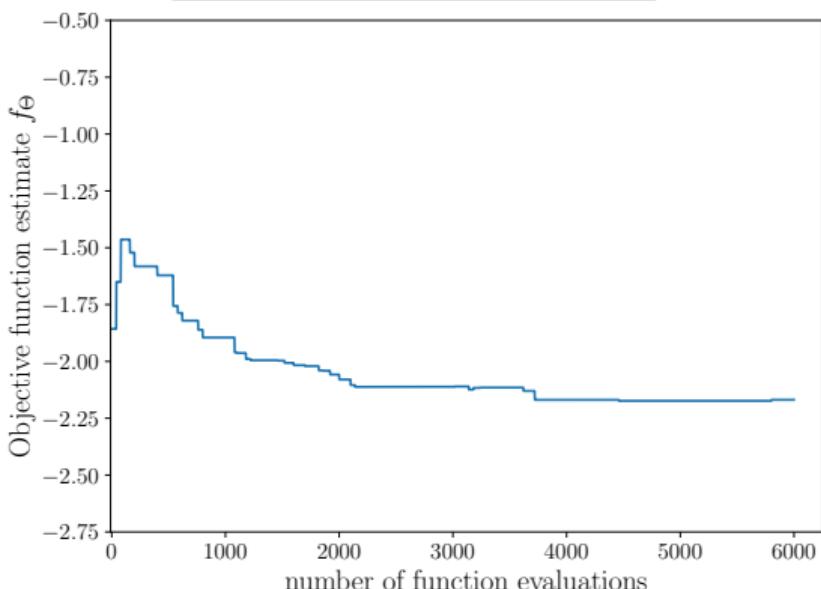
- Initial conditions
- Interactions, demographics

[1]NOMAD

[2]Alhandawi2021b

sampling rate $n^k = 4$

— StoMADS-PB - - - $\bar{c}_{\Theta} = 0$



Optimization results

StoMADS, NOMAD¹, and genetic algorithms were used to solve the problem²

Objective and constraints

$$\min_x \quad f(x) = \mathbb{E}_{\Theta} [f_{\Theta}(x) = -M^T]$$

$$\text{subject to} \quad c(x) = \mathbb{E}_{\Theta} [c_{\Theta}(x) \equiv n_{I,\max} - H_{\max}] \leq 0$$

where $x = [n_E, S_D, n_T]^T$, Θ : realizations

Design variables

- n_E : Number of essential workers
- S_D : Social distancing factor
- n_T : Number of tests daily

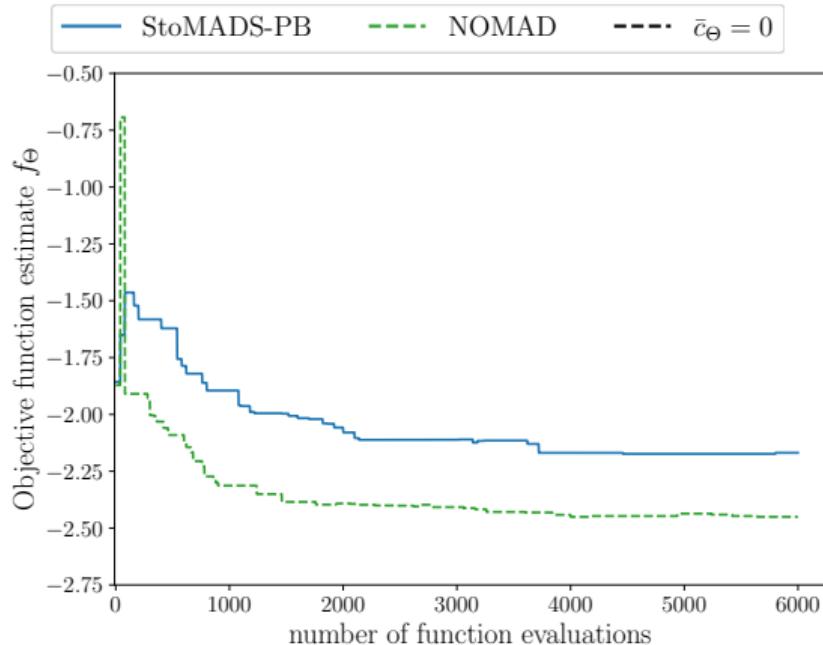
Randomly seeded parameters

- Initial conditions
- Interactions, demographics

[1]NOMAD

[2]Alhandawi2021b

sampling rate $n^k = 4$



Optimization results

StoMADS, NOMAD¹, and genetic algorithms were used to solve the problem²

Objective and constraints

$$\min_x \quad f(x) = \mathbb{E}_{\Theta} [f_{\Theta}(x) = -M^T]$$

$$\text{subject to} \quad c(x) = \mathbb{E}_{\Theta} [c_{\Theta}(x) \equiv n_{I,\max} - H_{\max}] \leq 0$$

where $x = [n_E, S_D, n_T]^T$, Θ : realizations

Design variables

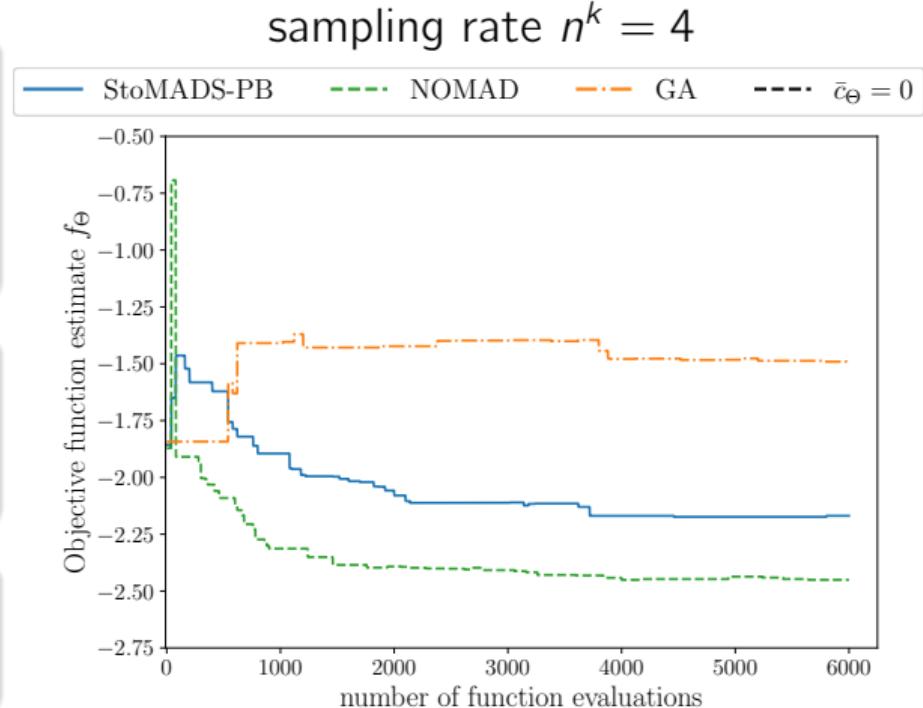
- n_E : Number of essential workers
- S_D : Social distancing factor
- n_T : Number of tests daily

Randomly seeded parameters

- Initial conditions
- Interactions, demographics

[1]NOMAD

[2]Alhandawi2021b



Optimization results

StoMADS, NOMAD¹, and genetic algorithms were used to solve the problem²

Objective and constraints

$$\min_x \quad f(x) = \mathbb{E}_{\Theta} [f_{\Theta}(x) = -M^T]$$

$$\text{subject to} \quad c(x) = \mathbb{E}_{\Theta} [c_{\Theta}(x) \equiv n_{I,\max} - H_{\max}] \leq 0$$

where $x = [n_E, S_D, n_T]^T$, Θ : realizations

Design variables

- n_E : Number of essential workers
- S_D : Social distancing factor
- n_T : Number of tests daily

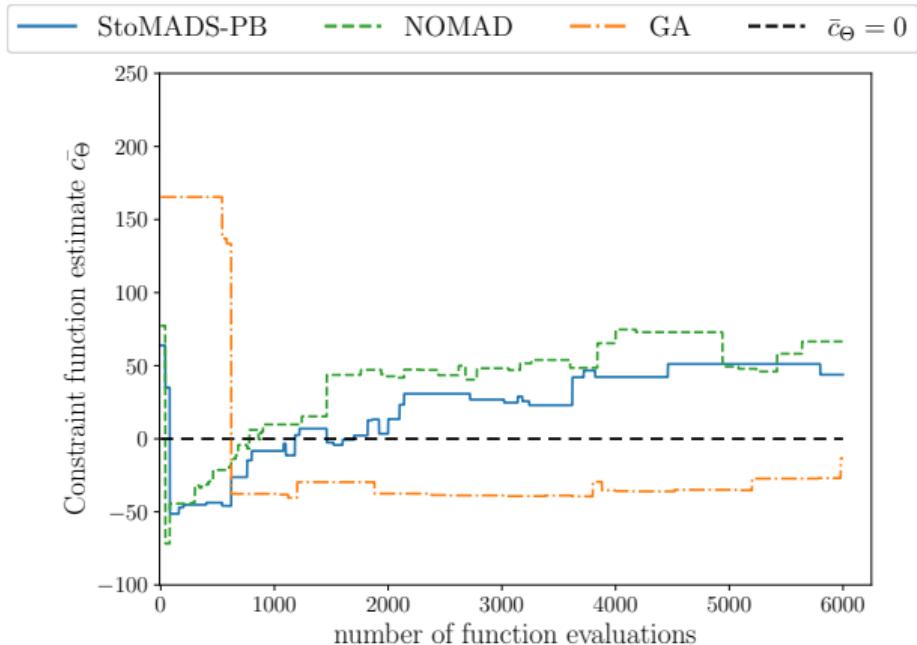
Randomly seeded parameters

- Initial conditions
- Interactions, demographics

[1]NOMAD

[2]Alhandawi2021b

sampling rate $n^k = 4$





Optimization results

StoMADS, NOMAD¹, and genetic algorithms were used to solve the problem²
best known solution

Objective and constraints

$$\min_x \quad f(x) = \mathbb{E}_{\Theta} [f_{\Theta}(x) = -M^T]$$

$$\text{subject to} \quad c(x) = \mathbb{E}_{\Theta} [c_{\Theta}(x) \equiv n_{I,\max} - H_{\max}] \leq 0$$

where $x = [n_E, S_D, n_T]^T, \Theta: \text{realizations}$

Design variables

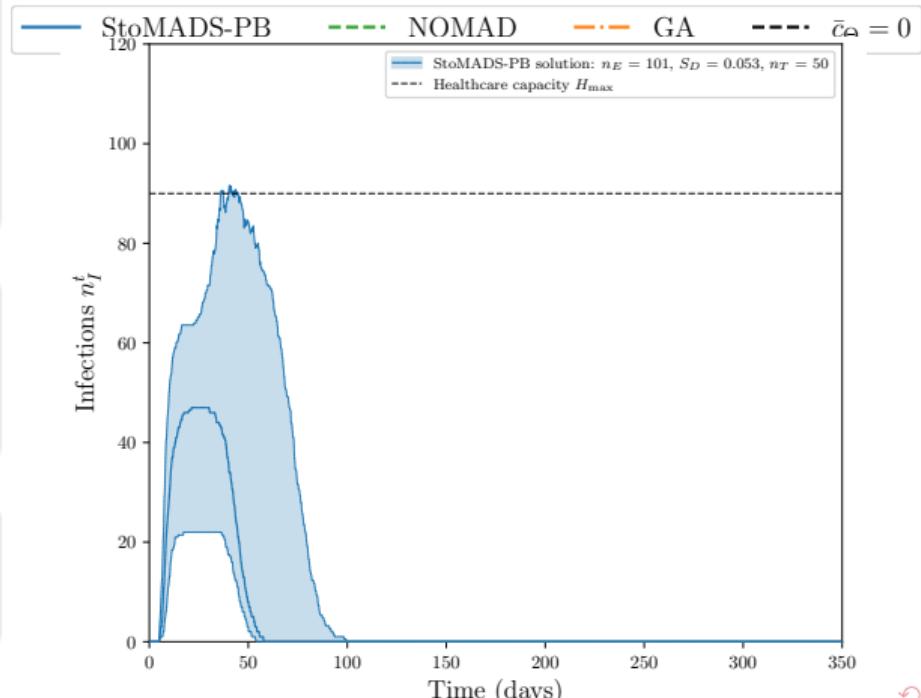
- n_E : Number of essential workers
- S_D : Social distancing factor
- n_T : Number of tests daily

Randomly seeded parameters

- Initial conditions
- Interactions, demographics

[1]NOMAD

[2]Alhandawi2021b





Optimization results

StoMADS, NOMAD¹, and genetic algorithms were used to solve the problem²
best known solution

Objective and constraints

$$\min_x \quad f(x) = \mathbb{E}_{\Theta} [f_{\Theta}(x) = -M^T]$$

$$\text{subject to} \quad c(x) = \mathbb{E}_{\Theta} [c_{\Theta}(x) \equiv n_{I,\max} - H_{\max}] \leq 0$$

where $x = [n_E, S_D, n_T]^T, \Theta: \text{realizations}$

Design variables

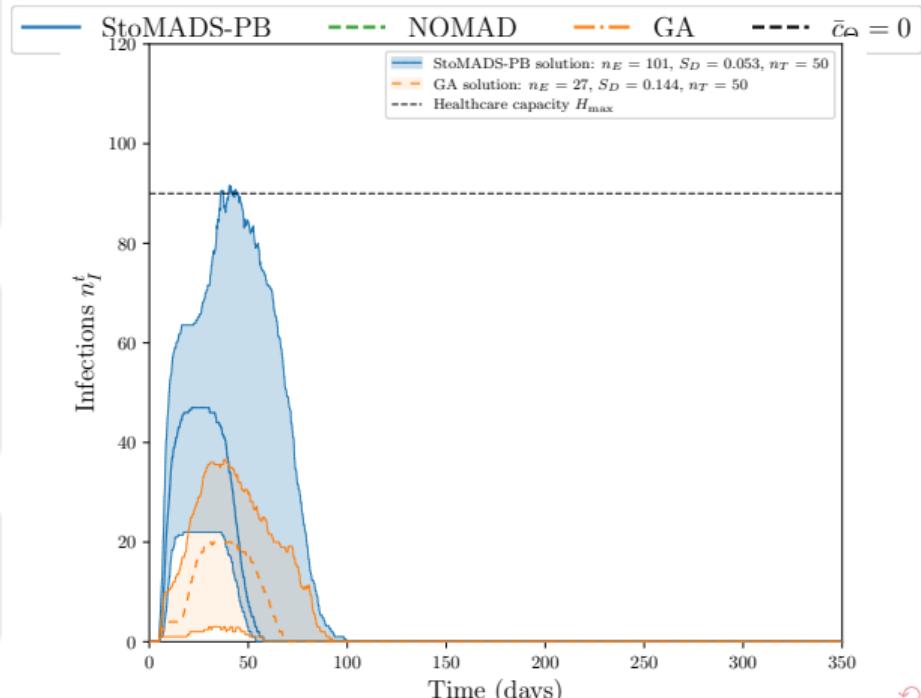
- n_E : Number of essential workers
- S_D : Social distancing factor
- n_T : Number of tests daily

Randomly seeded parameters

- Initial conditions
- Interactions, demographics

[1]NOMAD

[2]Alhandawi2021b



Optimization results

StoMADS, NOMAD¹, and genetic algorithms were used to solve the problem²
best known solution

Objective and constraints

$$\min_x \quad f(x) = \mathbb{E}_{\Theta} [f_{\Theta}(x) = -M^T]$$

$$\text{subject to} \quad c(x) = \mathbb{E}_{\Theta} [c_{\Theta}(x) \equiv n_{I,\max} - H_{\max}] \leq 0$$

where $x = [n_E, S_D, n_T]^T, \Theta: \text{realizations}$

Design variables

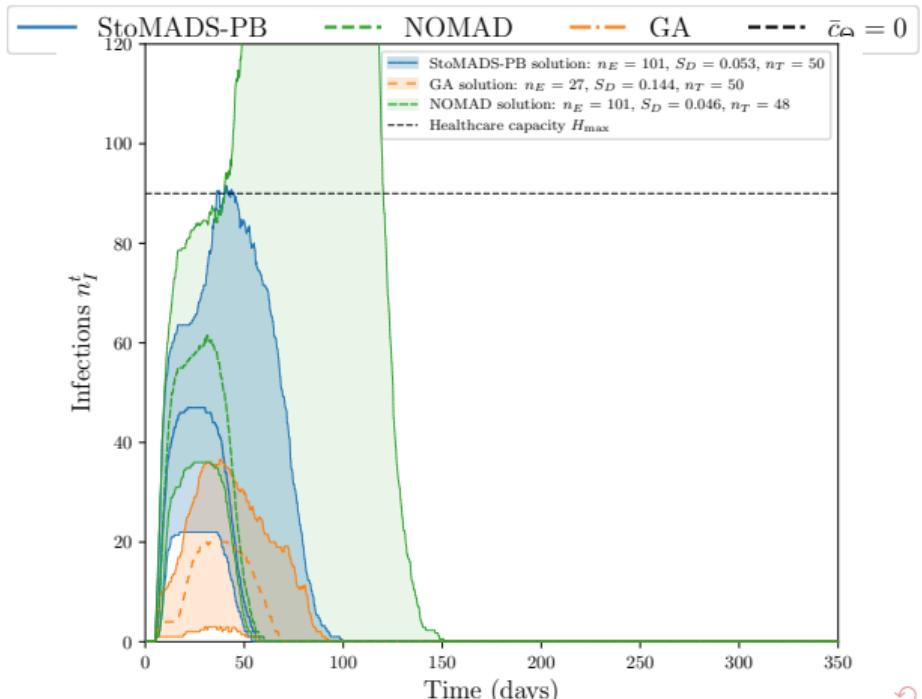
- n_E : Number of essential workers
- S_D : Social distancing factor
- n_T : Number of tests daily

Randomly seeded parameters

- Initial conditions
- Interactions, demographics

[1]NOMAD

[2]Alhandawi2021b



Optimization results

StoMADS, NOMAD¹, and genetic algorithms were used to solve the problem²
best known solution

Objective and constraints

$$\min_x \quad f(x) = \mathbb{E}_{\Theta} [f_{\Theta}(x) = -M^T]$$

$$\text{subject to} \quad c(x) = \mathbb{E}_{\Theta} [c_{\Theta}(x) \equiv n_{I,\max} - H_{\max}] \leq 0$$

where $x = [n_E, S_D, n_T]^T$, Θ : realizations

Design variables

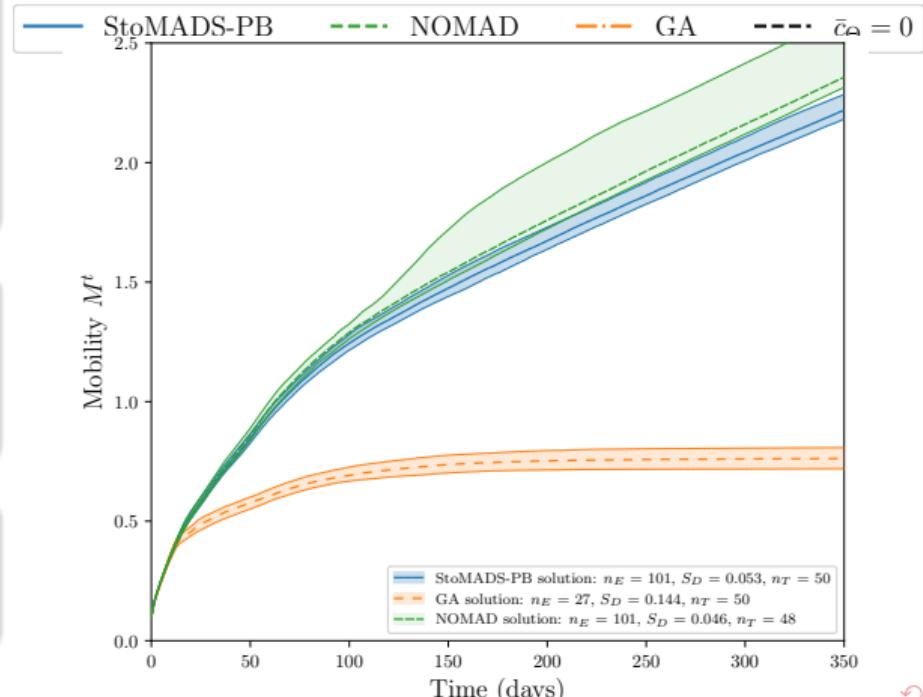
- n_E : Number of essential workers
- S_D : Social distancing factor
- n_T : Number of tests daily

Randomly seeded parameters

- Initial conditions
- Interactions, demographics

[1]NOMAD

[2]Alhandawi2021b



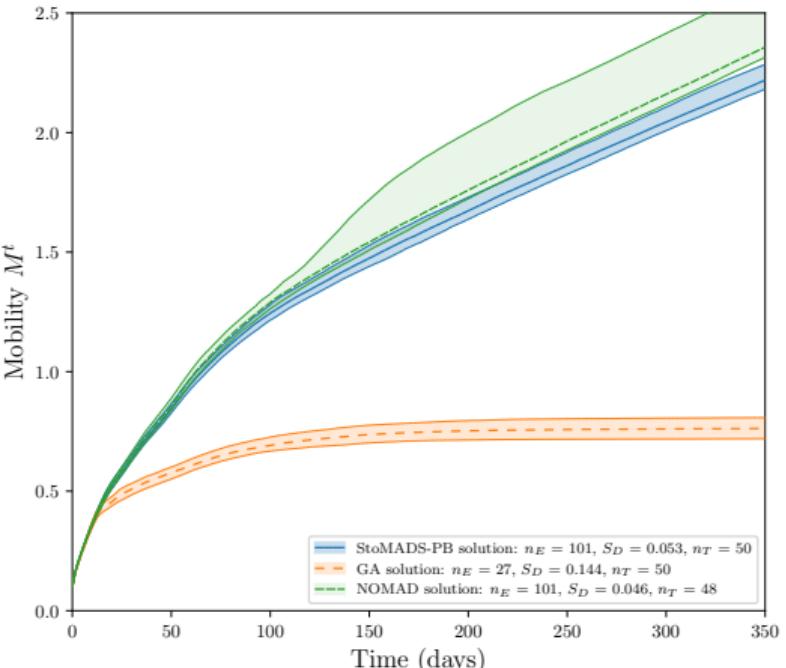


Conclusion

Formulated and solved public health policy-making problems

- Identified a public health policy that favored

best known solution

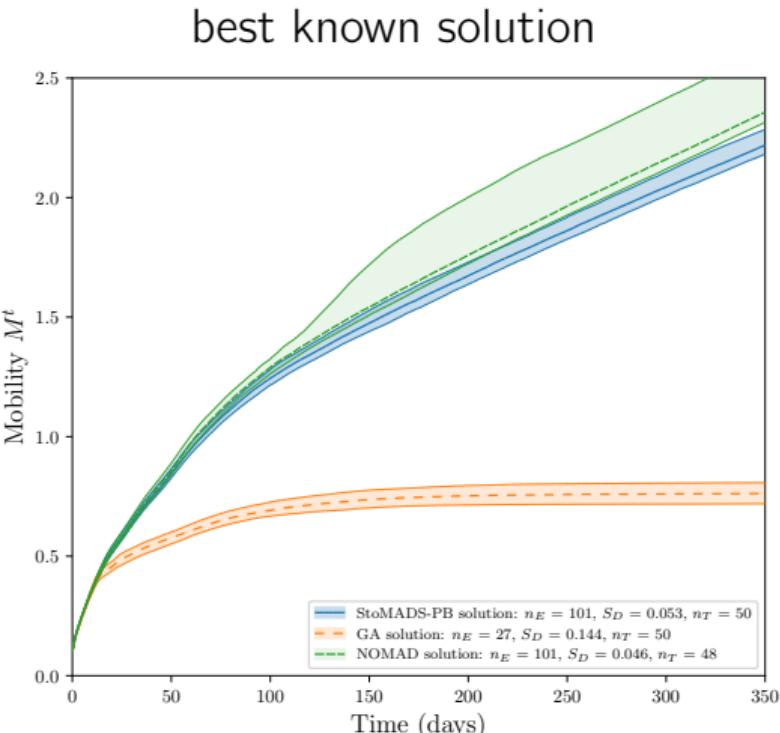




Conclusion

Formulated and solved public health policy-making problems

- Identified a public health policy that favored
 - High testing capacity n_T

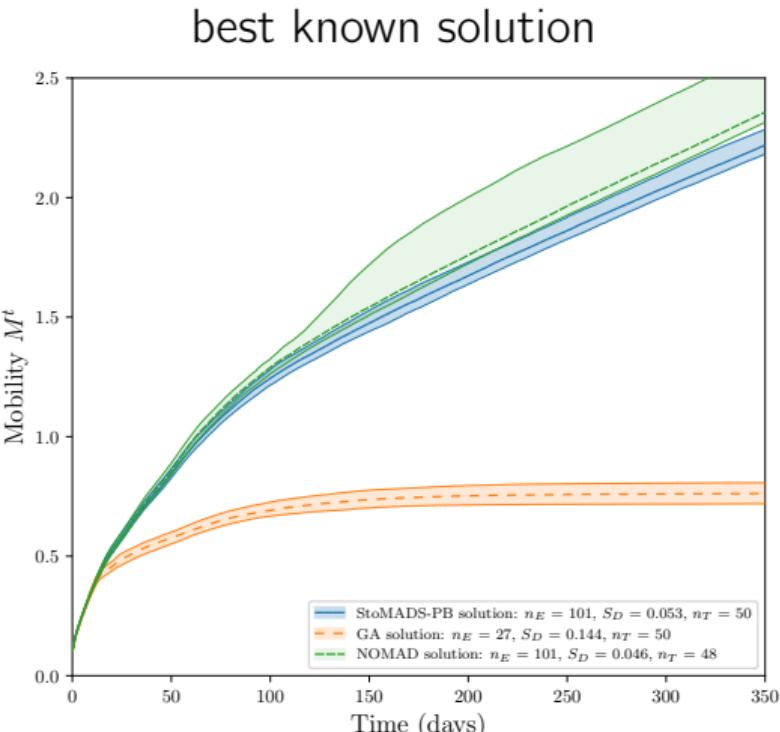




Conclusion

Formulated and solved public health policy-making problems

- Identified a public health policy that favored
 - High testing capacity n_T
 - Large number of essential workers n_E

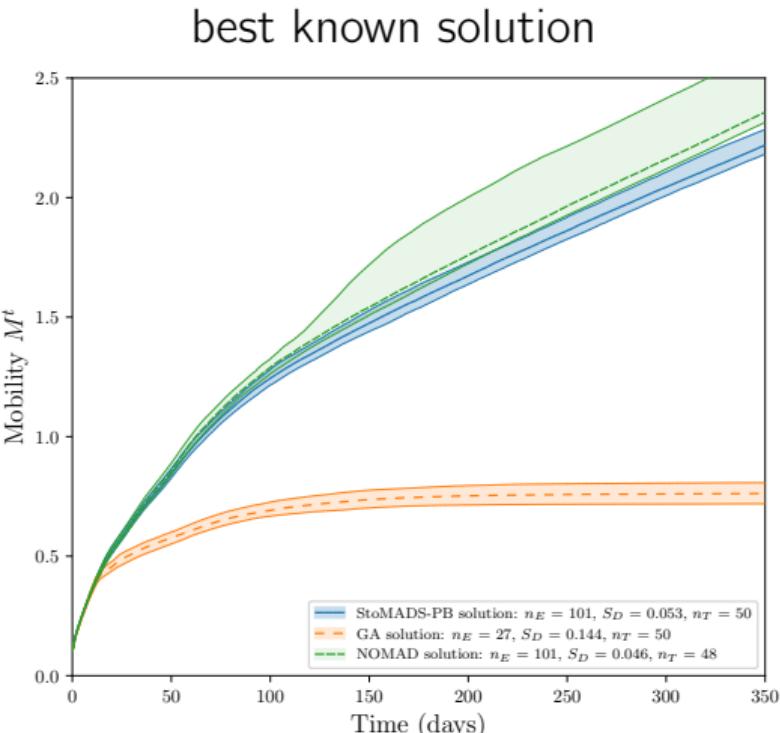




Conclusion

Formulated and solved public health policy-making problems

- Identified a public health policy that favored
 - High testing capacity n_T
 - Large number of essential workers n_E
 - Modest social distancing S_D

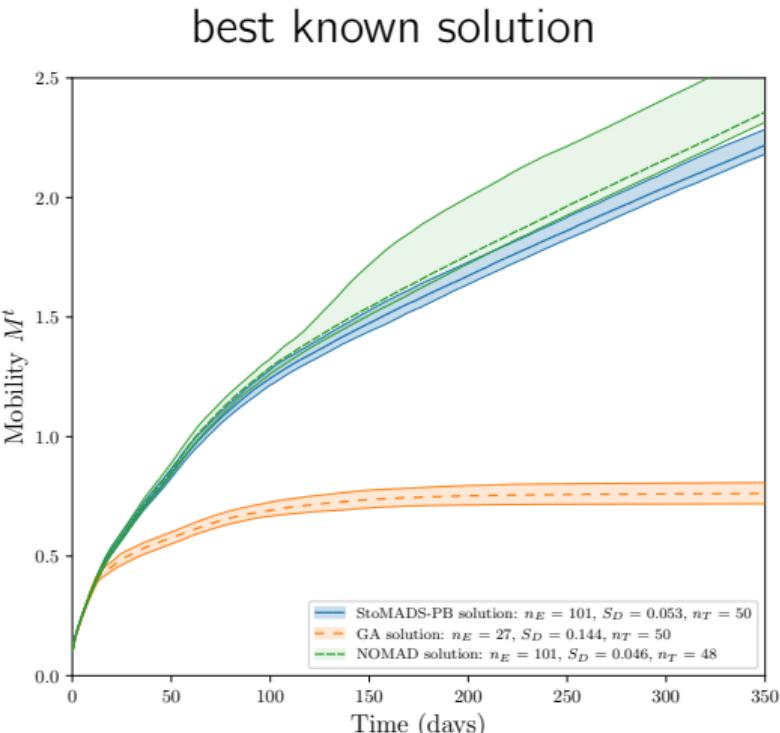




Conclusion

Formulated and solved public health policy-making problems

- Identified a public health policy that favored
 - High testing capacity n_T
 - Large number of essential workers n_E
 - Modest social distancing S_D
- StoMADS outperformed GAs and NOMAD on active constraints





Conclusion

Formulated and solved public health policy-making problems

- Identified a public health policy that favored
 - High testing capacity n_T
 - Large number of essential workers n_E
 - Modest social distancing S_D
- StoMADS outperformed GAs and NOMAD on active constraints
- Applicable to other large-scale agent-based models¹

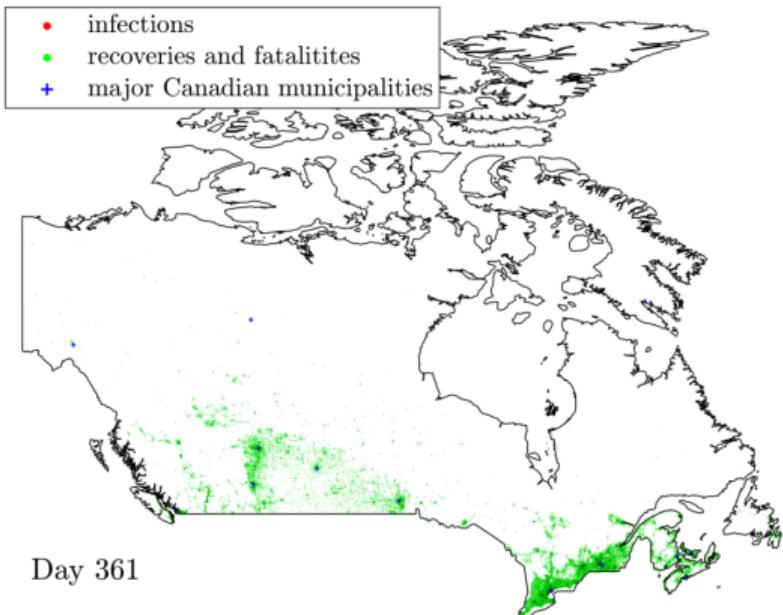
[1] Ferguson2020



Conclusion

Formulated and solved public health policy-making problems

- Identified a public health policy that favored
 - High testing capacity n_T
 - Large number of essential workers n_E
 - Modest social distancing S_D
- StoMADS outperformed GAs and NOMAD on active constraints
- Applicable to other large-scale agent-based models¹

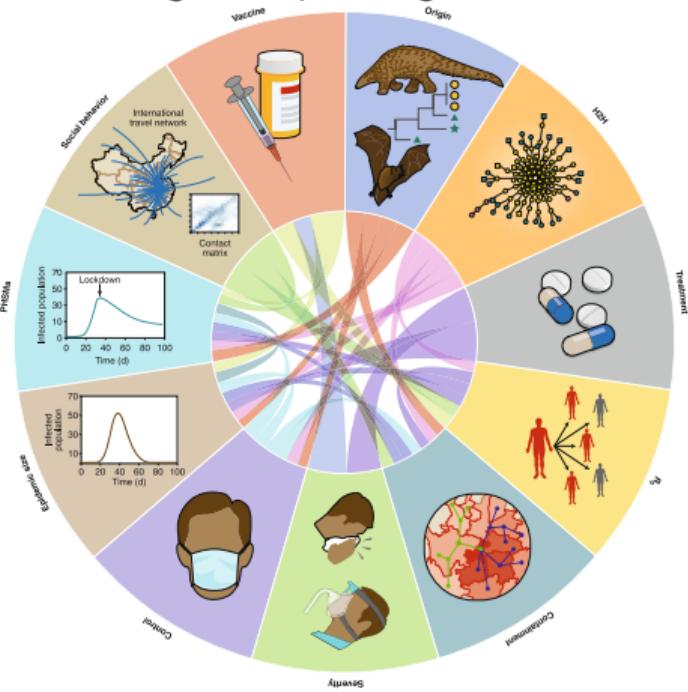


[1] Ferguson2020



Background: COVID-19 forecasting targets

Forecasting novel epidemics is a multidisciplinary field involving multiple *targets*¹
Inputs for forecasting epidemic size:



Nowcasting targets

[1]Wu2021



Background: COVID-19 forecasting targets

Forecasting novel epidemics is a multidisciplinary field involving multiple *targets*¹
Inputs for forecasting epidemic size:

- Growth rate indicators



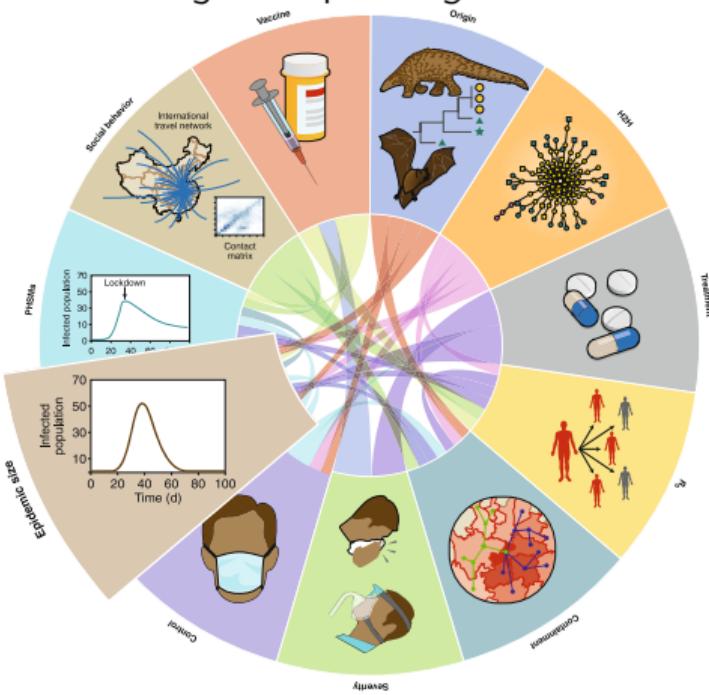
Nowcasting targets

[1]Wu2021

Background: COVID-19 forecasting targets

Forecasting novel epidemics is a multidisciplinary field involving multiple *targets*¹
Inputs for forecasting epidemic size:

- Growth rate indicators
- Historical incidence rate data



Nowcasting targets

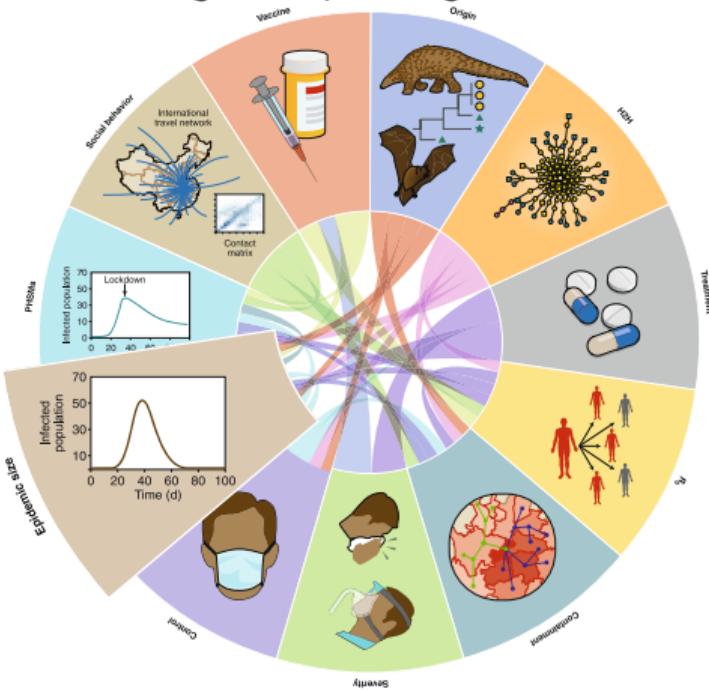
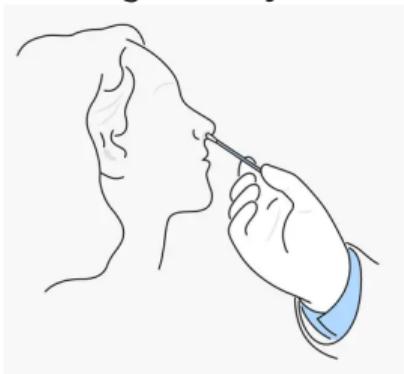
[1]Wu2021



Background: COVID-19 forecasting targets

Forecasting novel epidemics is a multidisciplinary field involving multiple *targets*¹
Inputs for forecasting epidemic size:

- Growth rate indicators
- Historical incidence rate data
- Serologic assays



Nowcasting targets

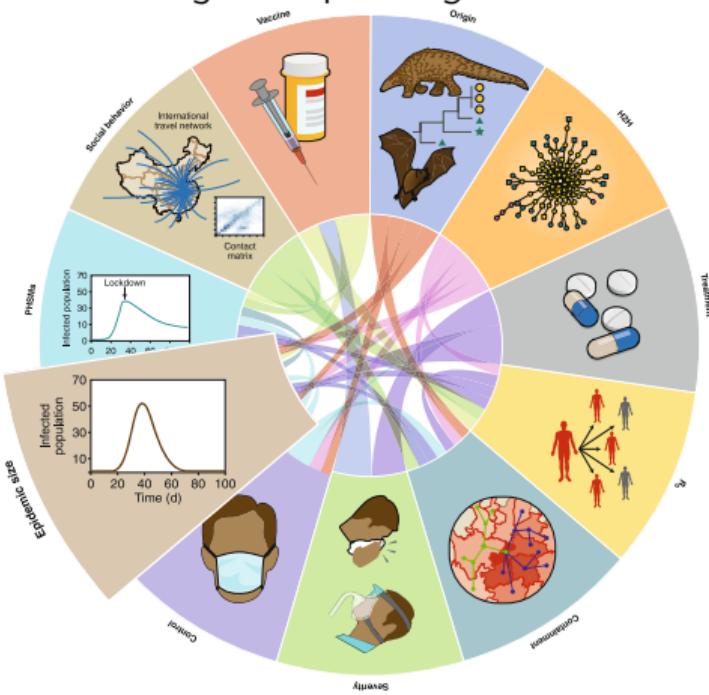
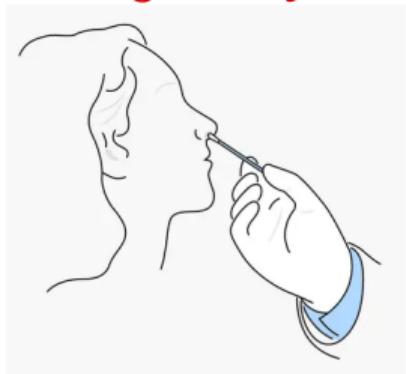
[1]Wu2021



Background: COVID-19 forecasting targets

Forecasting novel epidemics is a multidisciplinary field involving multiple *targets*¹
Inputs for forecasting epidemic size:

- Growth rate indicators
- **Historical incidence rate data**
- **Serologic assays**



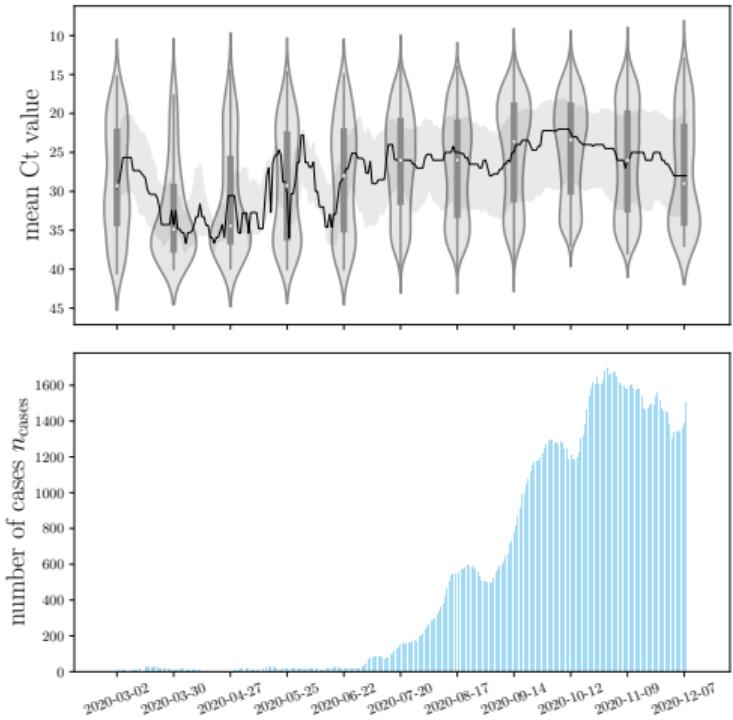
[1]Wu2021

Nowcasting targets



COVID-19 incidence forecasting

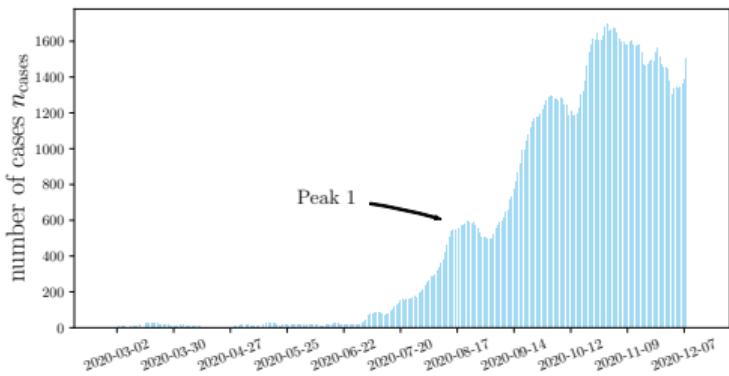
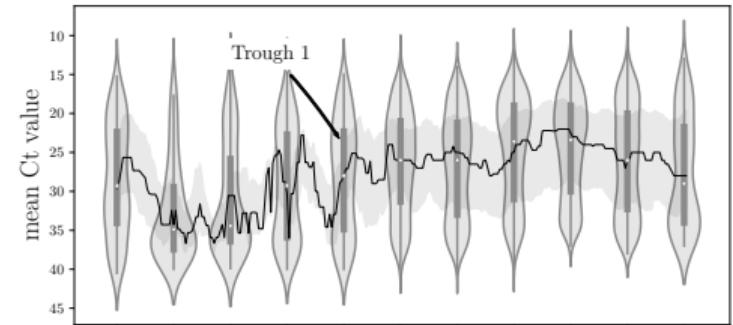
How to *design* a machine learning model for forecasting a time-series?





COVID-19 incidence forecasting

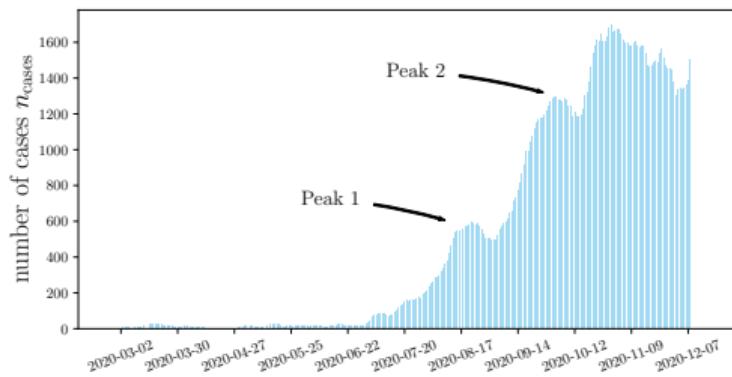
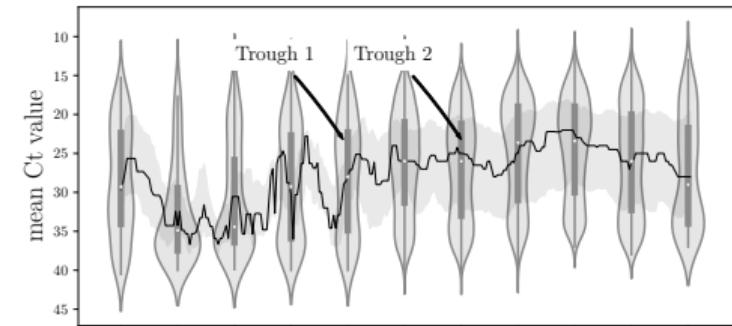
How to *design* a machine learning model for forecasting a time-series?





COVID-19 incidence forecasting

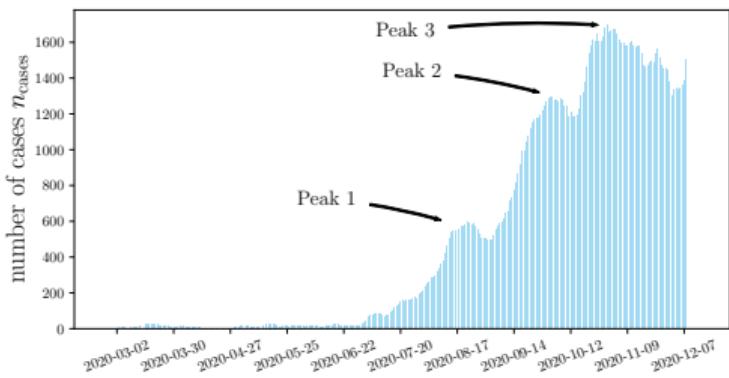
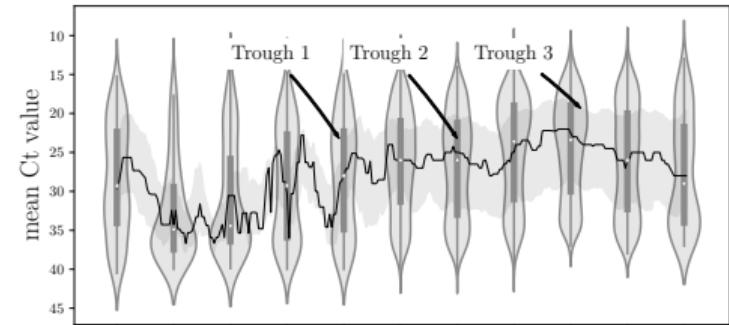
How to *design* a machine learning model for forecasting a time-series?





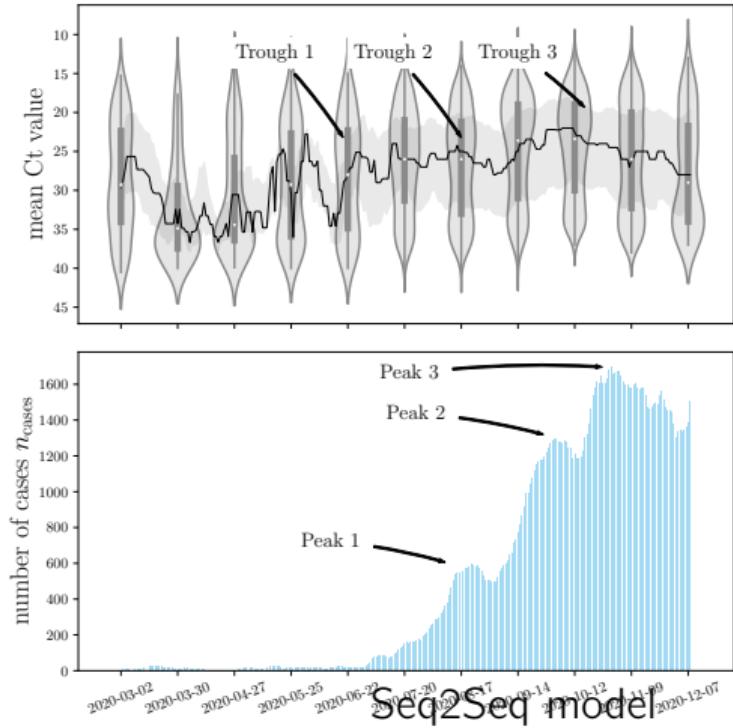
COVID-19 incidence forecasting

How to *design* a machine learning model for forecasting a time-series?



COVID-19 incidence forecasting

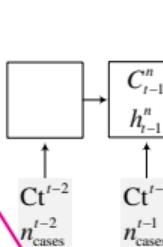
How to *design* a machine learning model for forecasting a time-series?



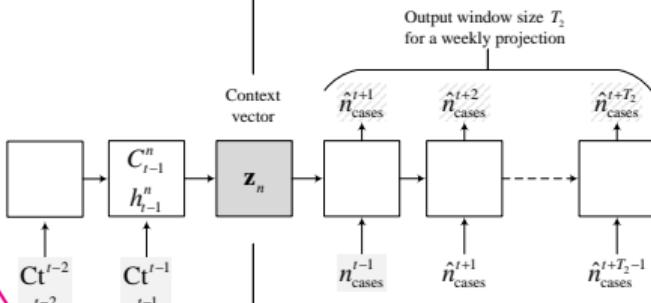
Seq2Seq model



Encoder



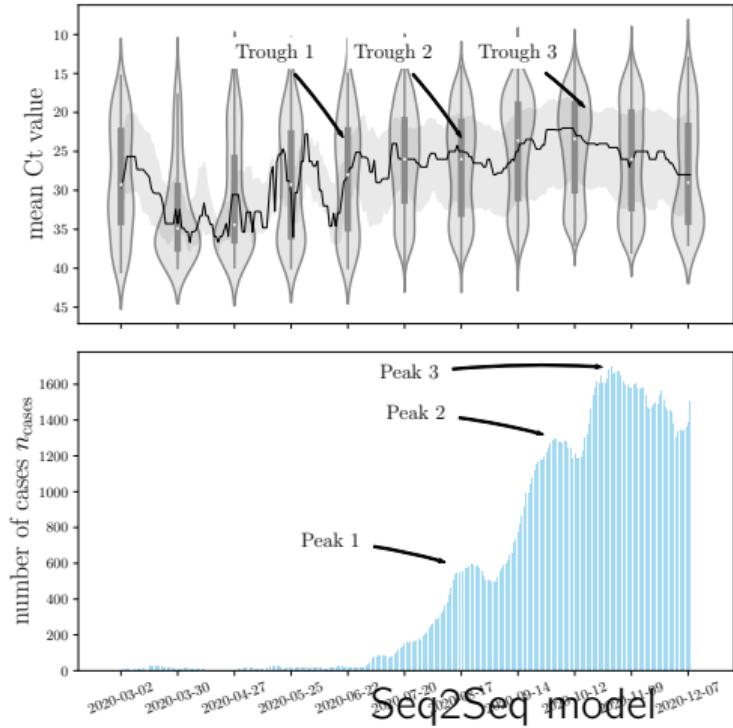
Decoder



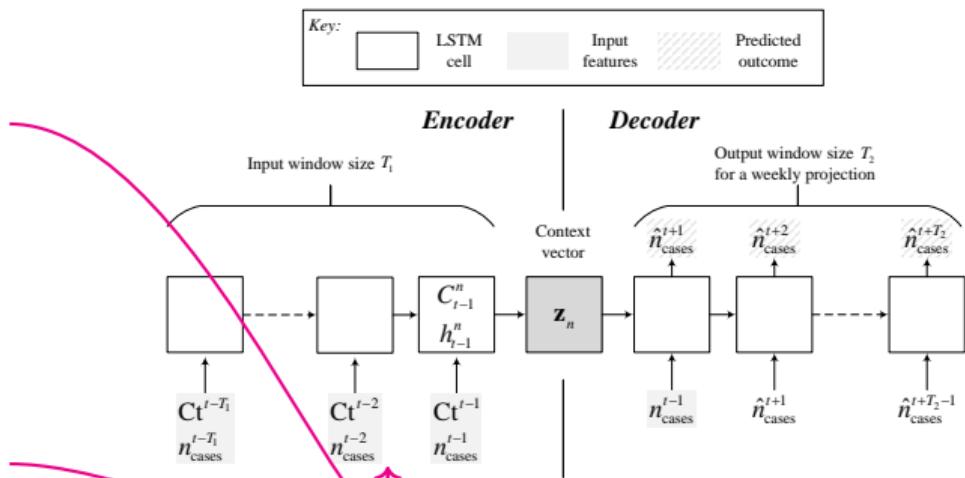


COVID-19 incidence forecasting

How to *design* a machine learning model for forecasting a time-series?



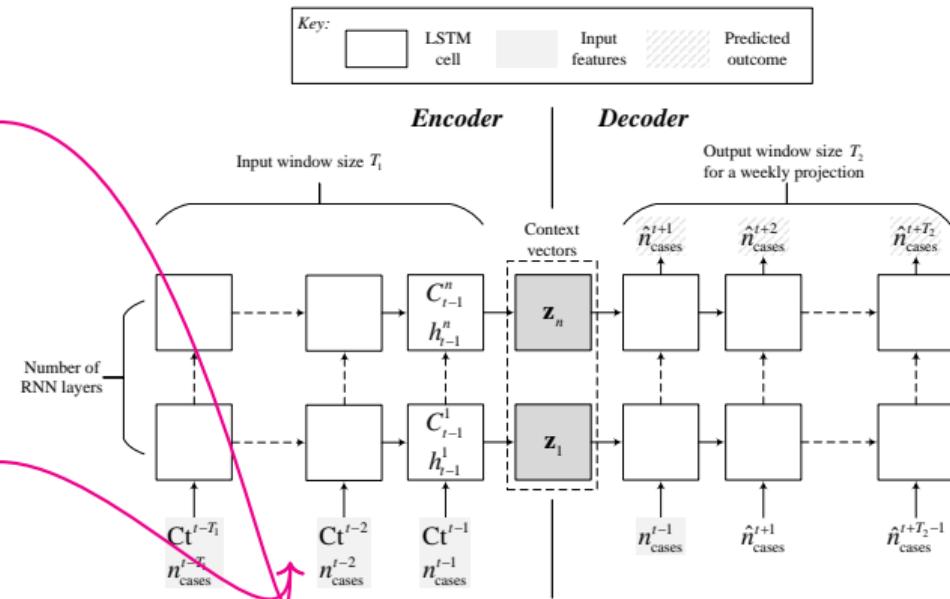
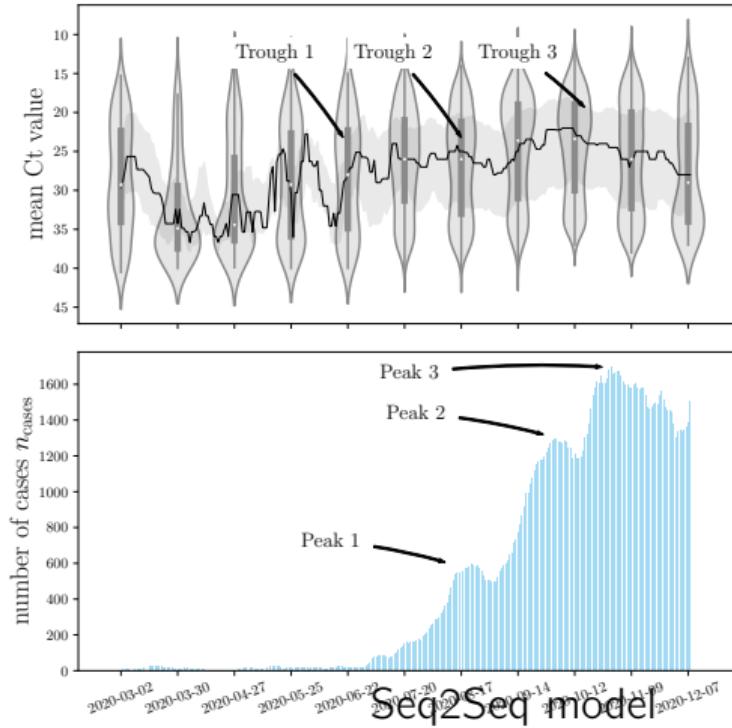
Seq2Seq model





COVID-19 incidence forecasting

How to *design* a machine learning model for forecasting a time-series?

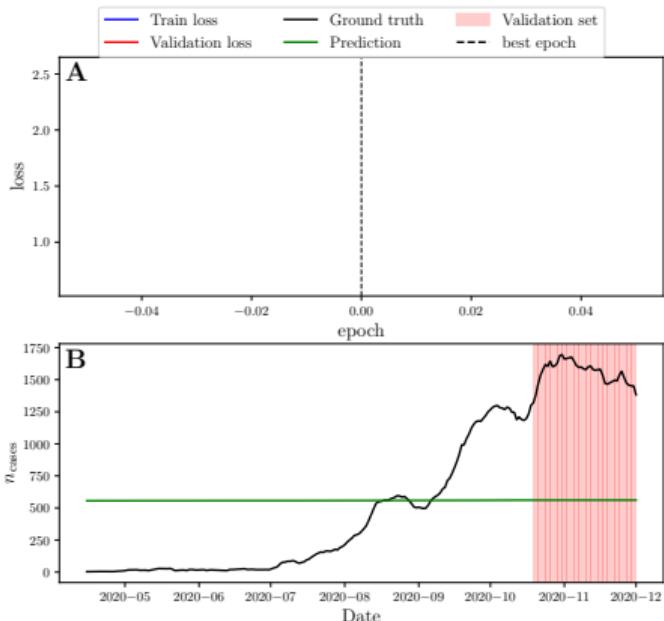




Training the model

There are several challenges associated with hyperparameter optimization

- The number of epochs can be tuned using *early stopping*
- This is a form of *regularization* to reduce over-fitting



Effect of number of epochs on testing loss



Training the model

There are several challenges associated with hyperparameter optimization

- The number of epochs can be tuned using *early stopping*
- This is a form of *regularization* to reduce over-fitting

Effect of number of epochs on testing loss

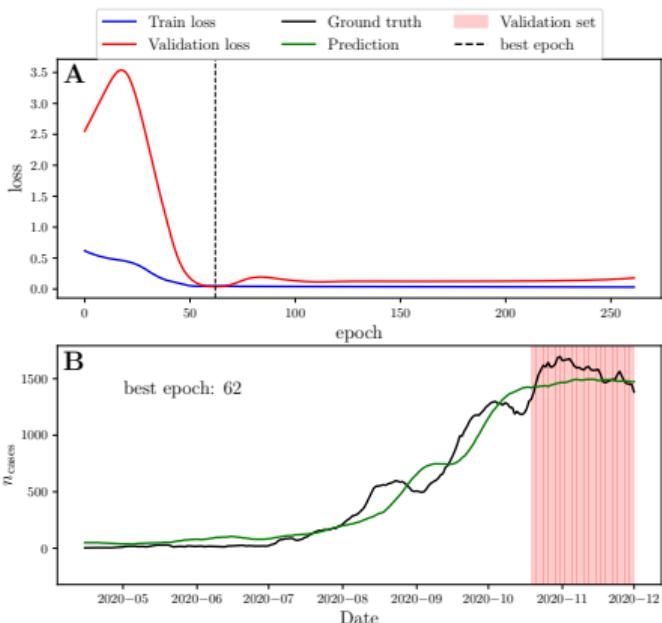


Training the model

There are several challenges associated with hyperparameter optimization

- The number of epochs can be tuned using *early stopping*
- This is a form of *regularization* to reduce over-fitting

However,



Effect of number of epochs on testing loss



Training the model

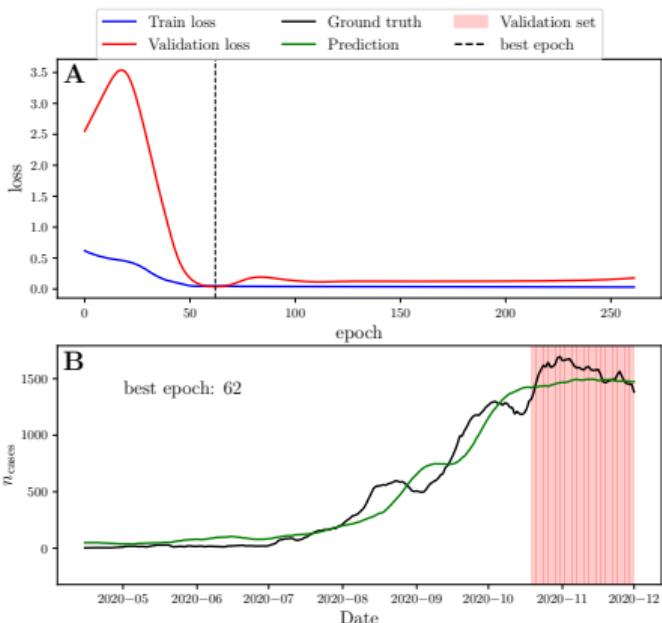
There are several challenges associated with hyperparameter optimization

- The number of epochs can be tuned using *early stopping*
- This is a form of *regularization* to reduce over-fitting

However,

1) Other hyperparameters to tune

- Dropout
- Training batch size
- Activation function(s)



Effect of number of epochs on testing loss



Training the model

There are several challenges associated with hyperparameter optimization

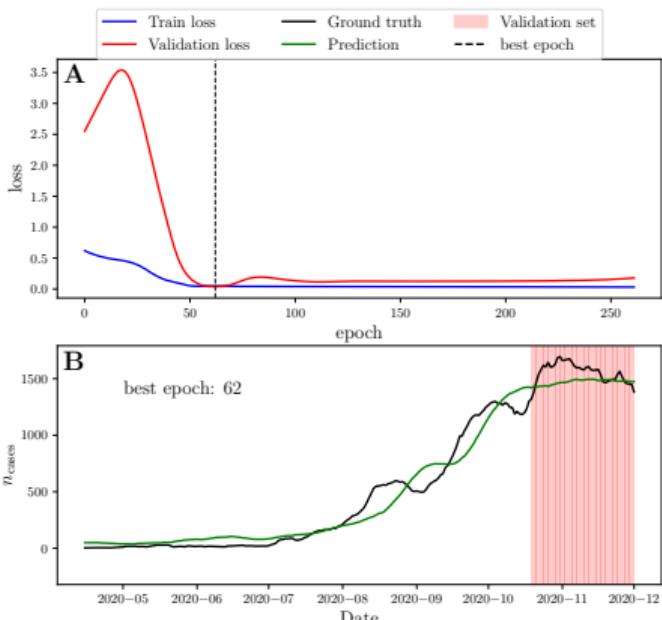
- The number of epochs can be tuned using *early stopping*
- This is a form of *regularization* to reduce over-fitting

However,

- 1) Other hyperparameters to tune

- Dropout
- Training batch size
- Activation function(s)

- 2) Training can be **expensive**



Effect of number of epochs on testing loss



Training the model

There are several challenges associated with hyperparameter optimization

- The number of epochs can be tuned using *early stopping*
- This is a form of *regularization* to reduce over-fitting

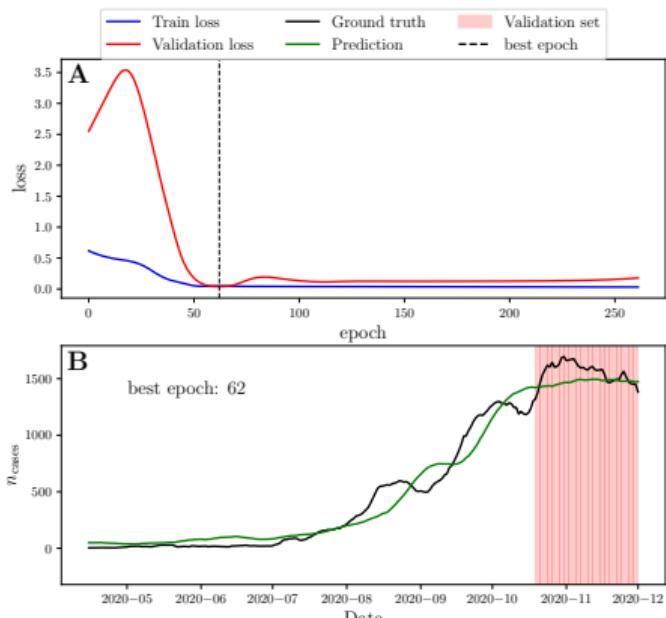
However,

- 1) Other hyperparameters to tune

- Dropout
- Training batch size
- Activation function(s)

- 2) Training can be **expensive**

- 3) Backpropagation is **stochastic**

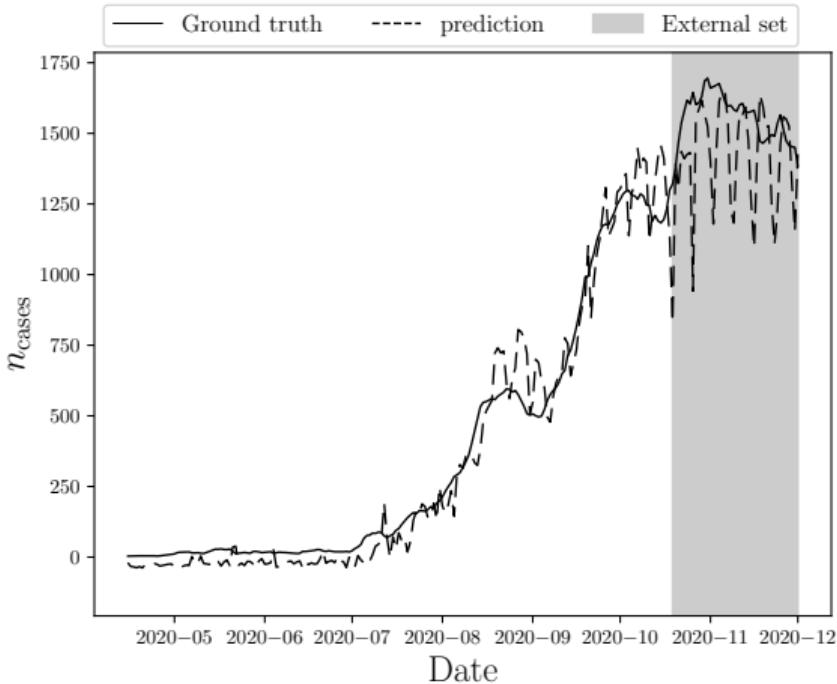
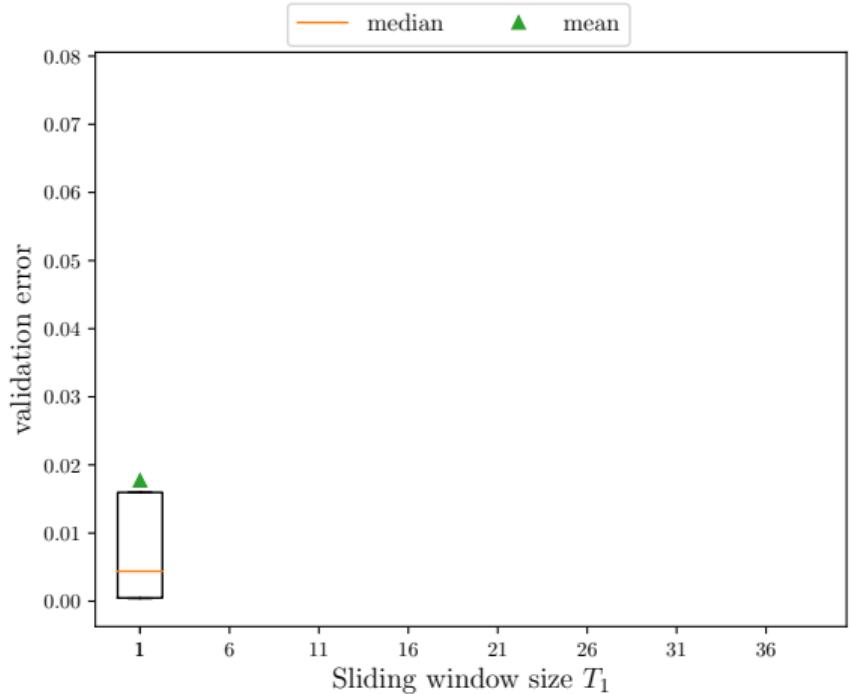


Effect of number of epochs on testing loss



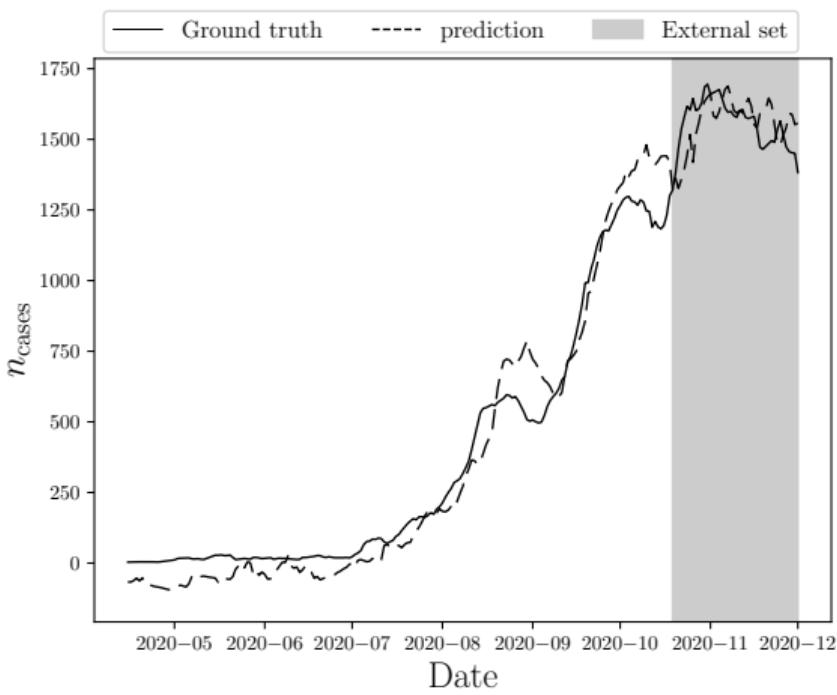
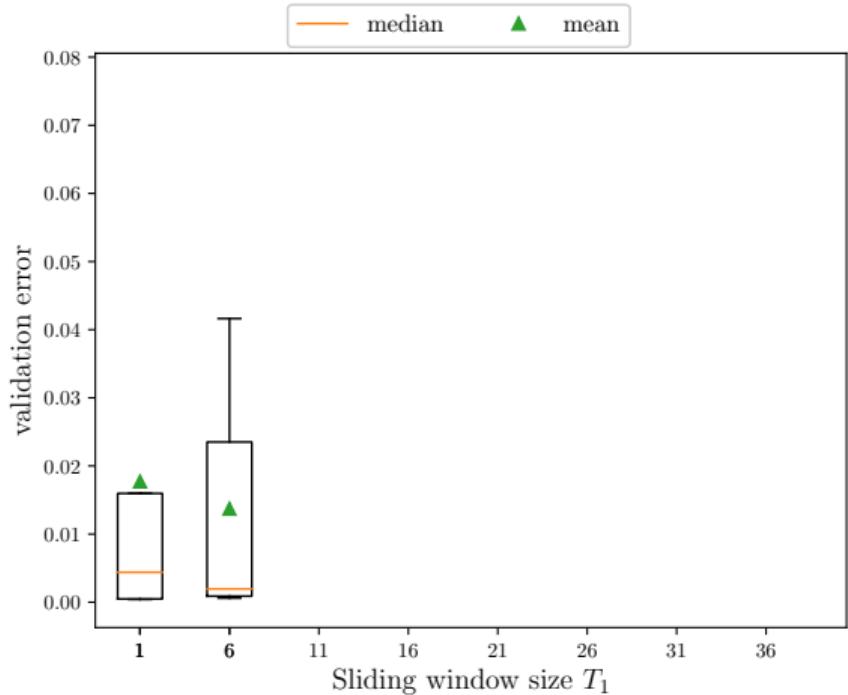
Hyperparameter tuning

We can use StoMADS to solve such hyperparameter optimization problems¹



Hyperparameter tuning

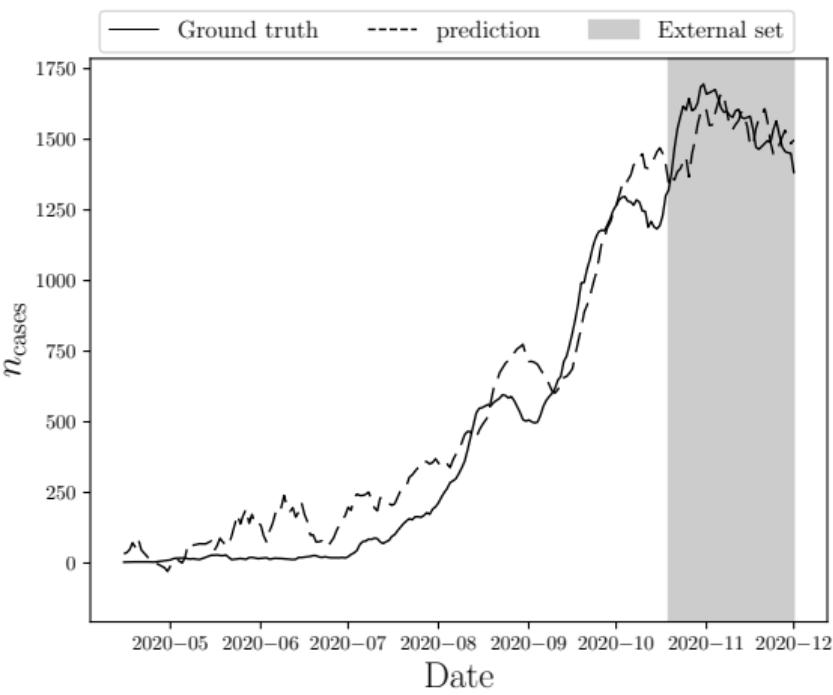
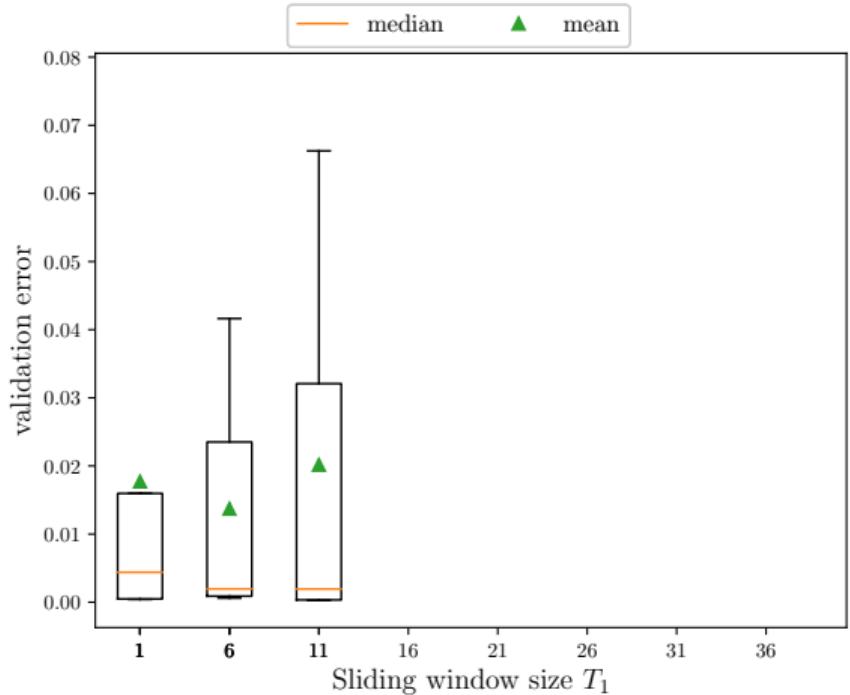
We can use StoMADS to solve such hyperparameter optimization problems¹



[1] Khalil2021

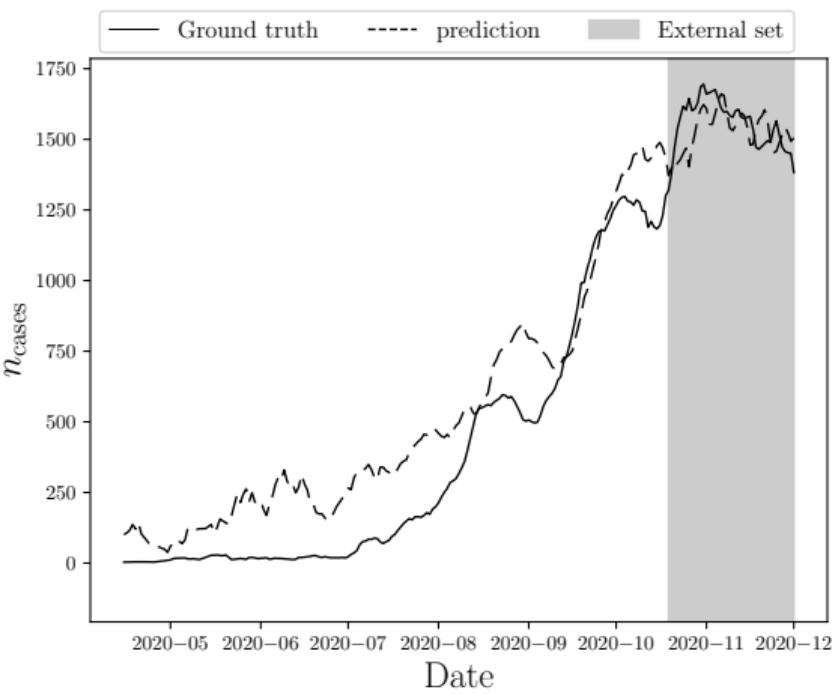
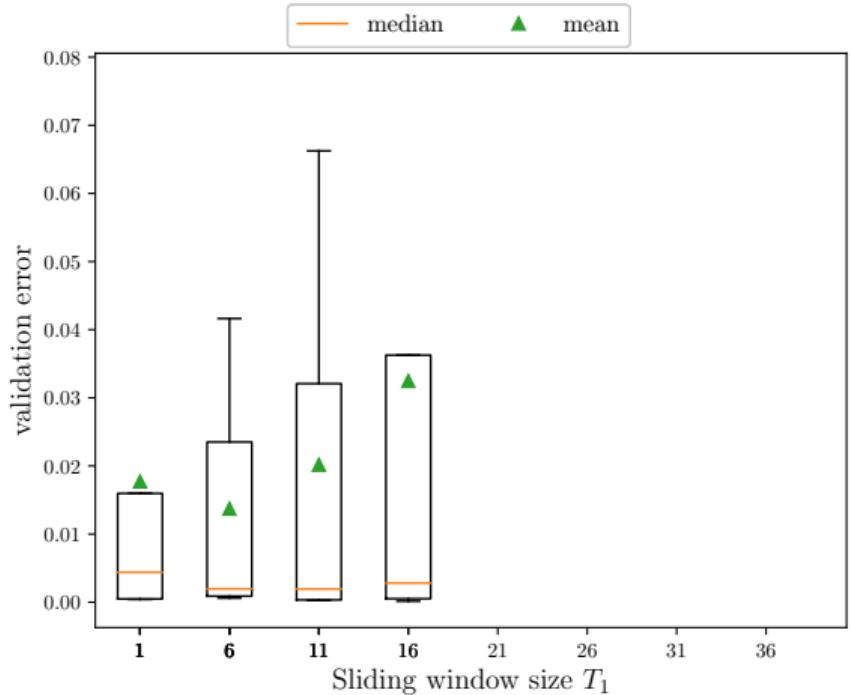
Hyperparameter tuning

We can use StoMADS to solve such hyperparameter optimization problems¹



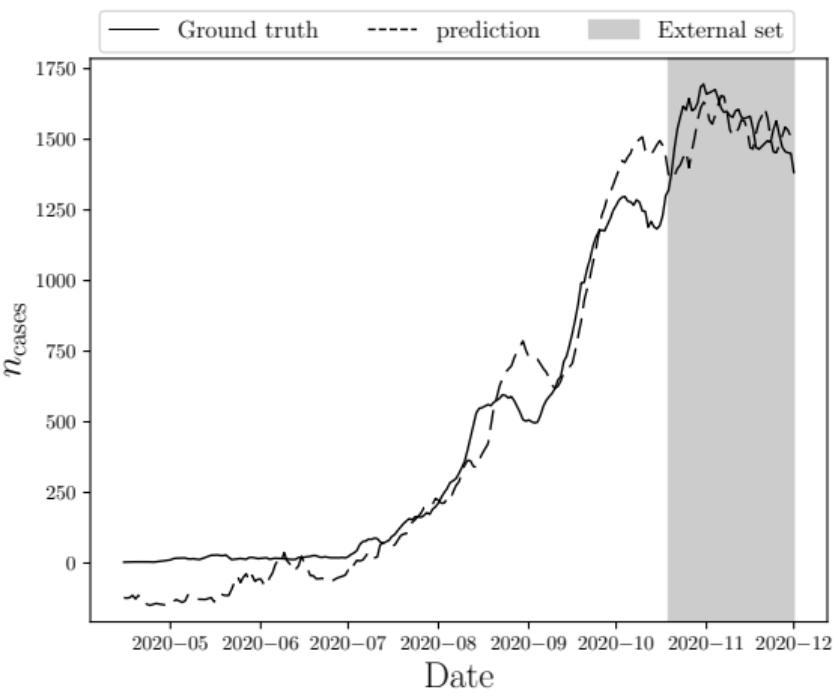
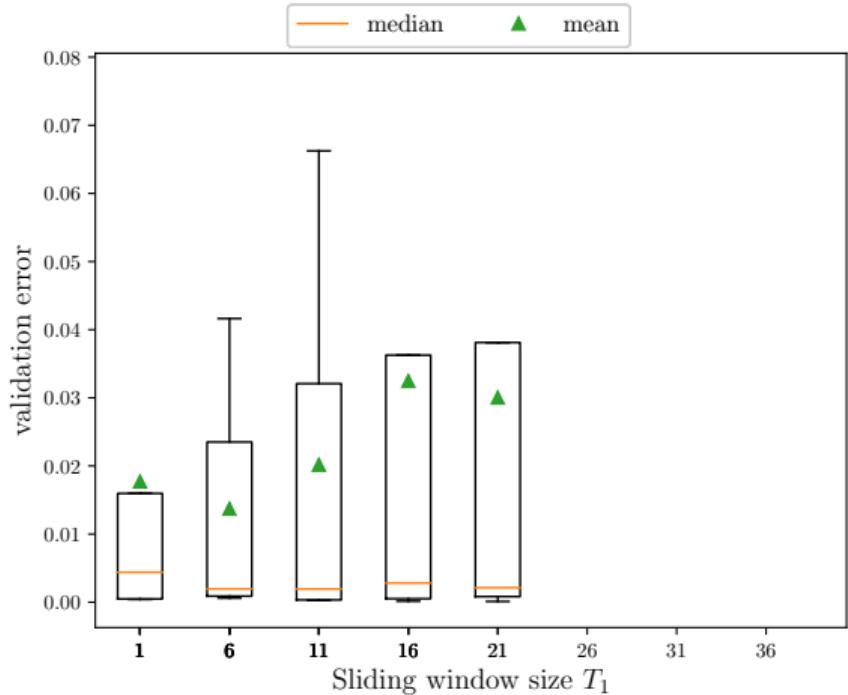
Hyperparameter tuning

We can use StoMADS to solve such hyperparameter optimization problems¹



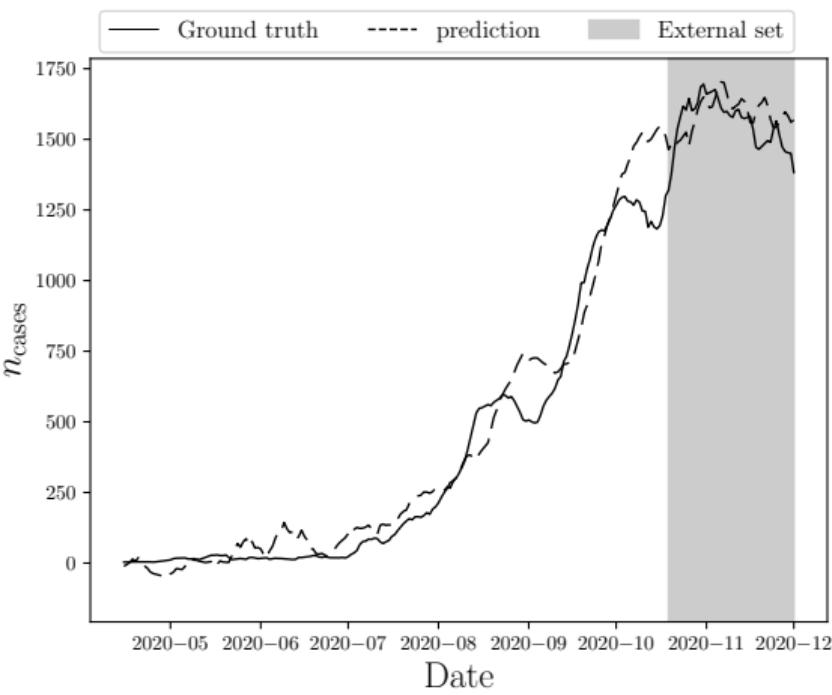
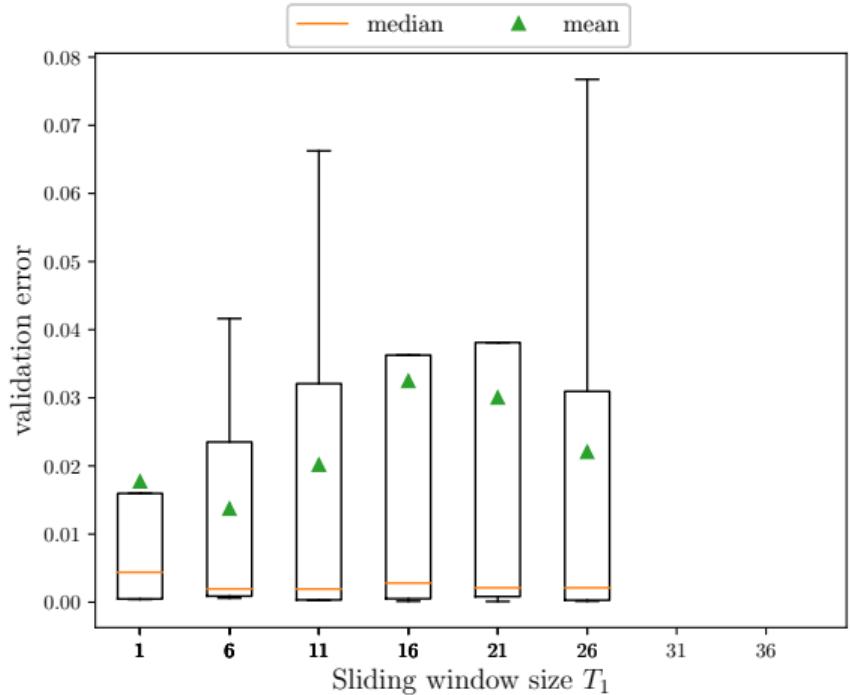
Hyperparameter tuning

We can use StoMADS to solve such hyperparameter optimization problems¹



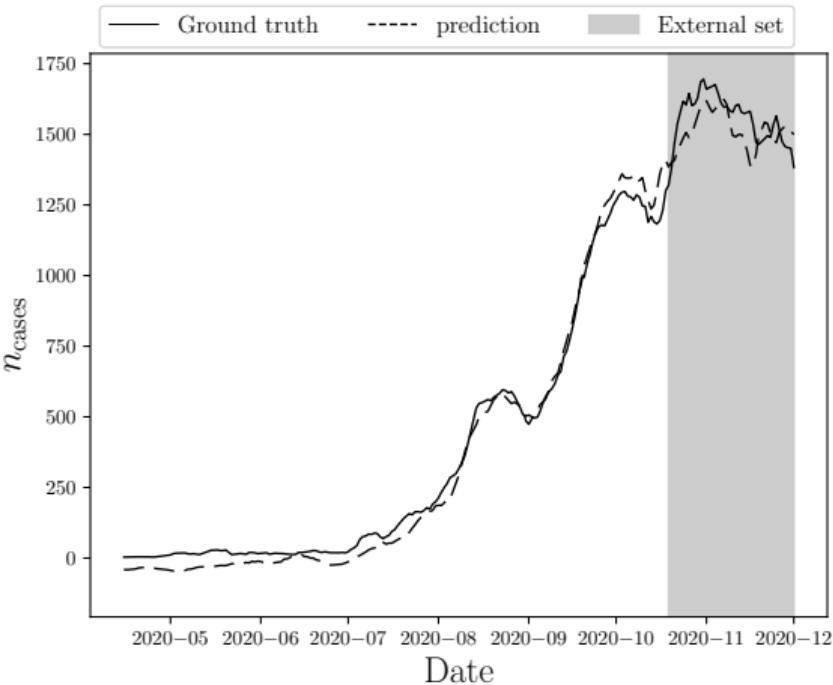
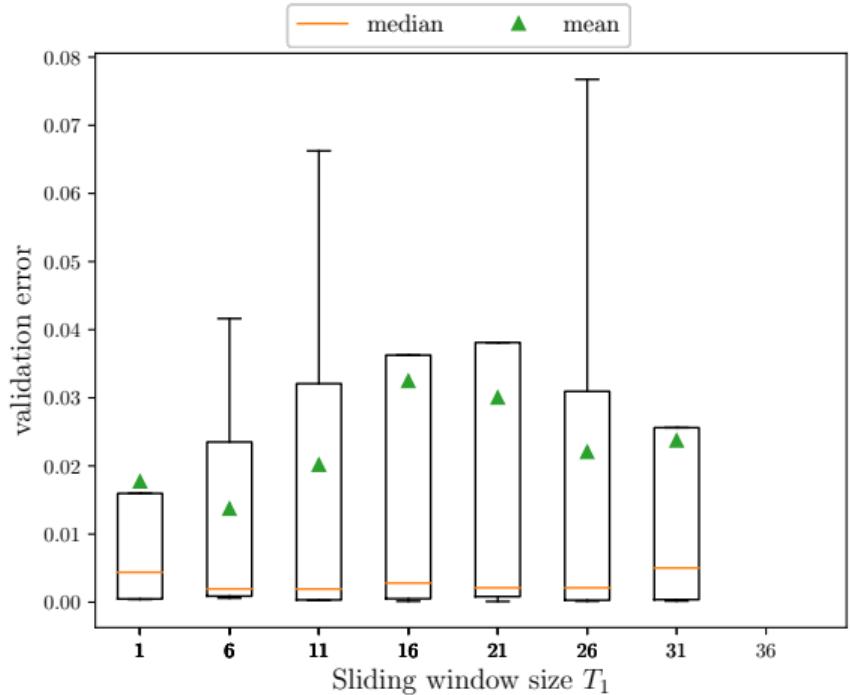
Hyperparameter tuning

We can use StoMADS to solve such hyperparameter optimization problems¹



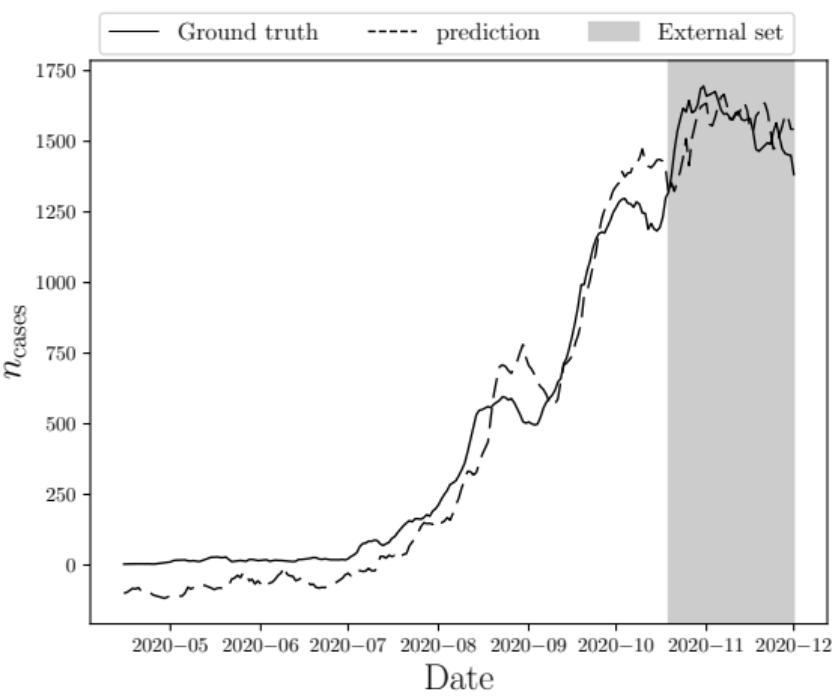
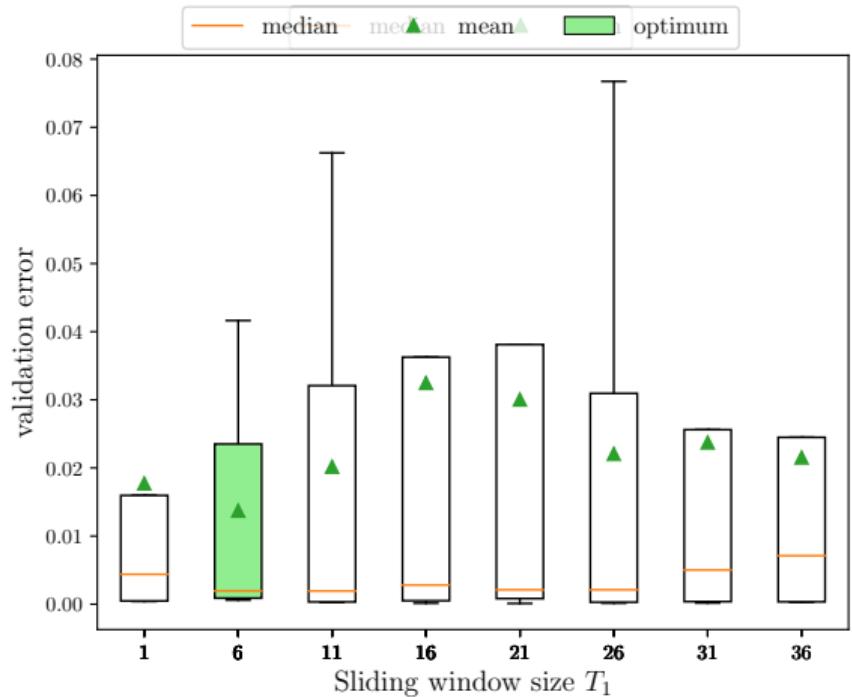
Hyperparameter tuning

We can use StoMADS to solve such hyperparameter optimization problems¹



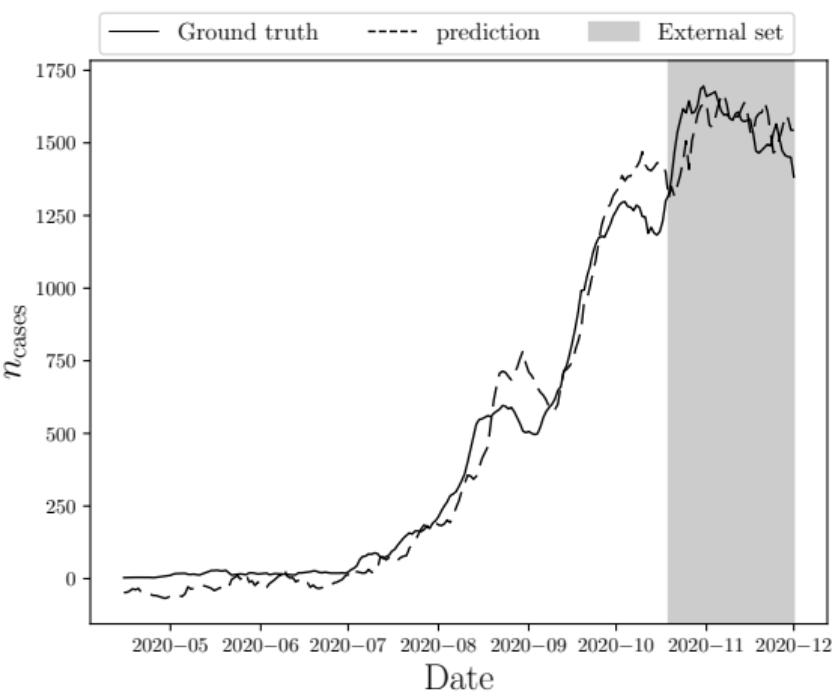
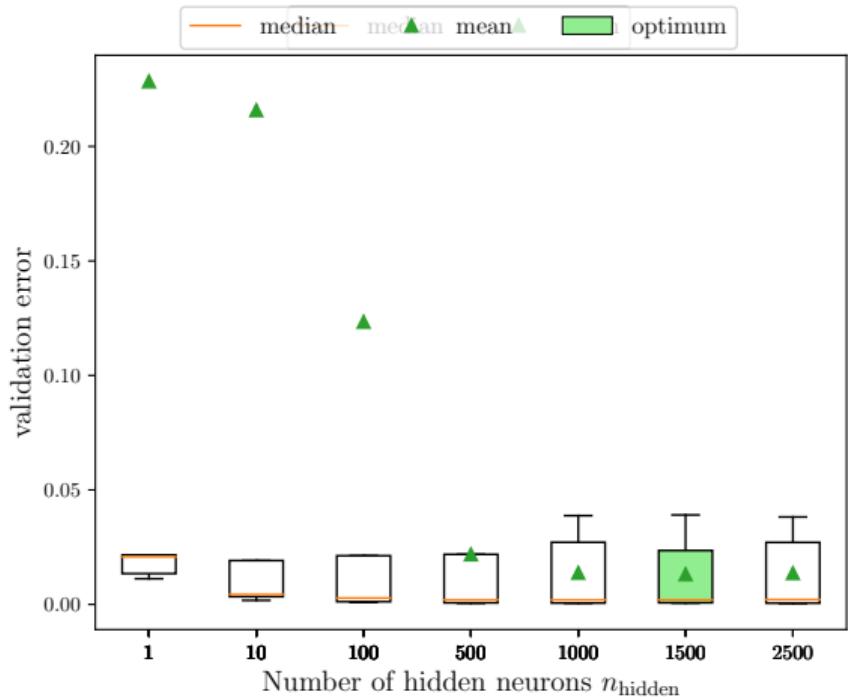
Hyperparameter tuning

We can use StoMADS to solve such hyperparameter optimization problems¹



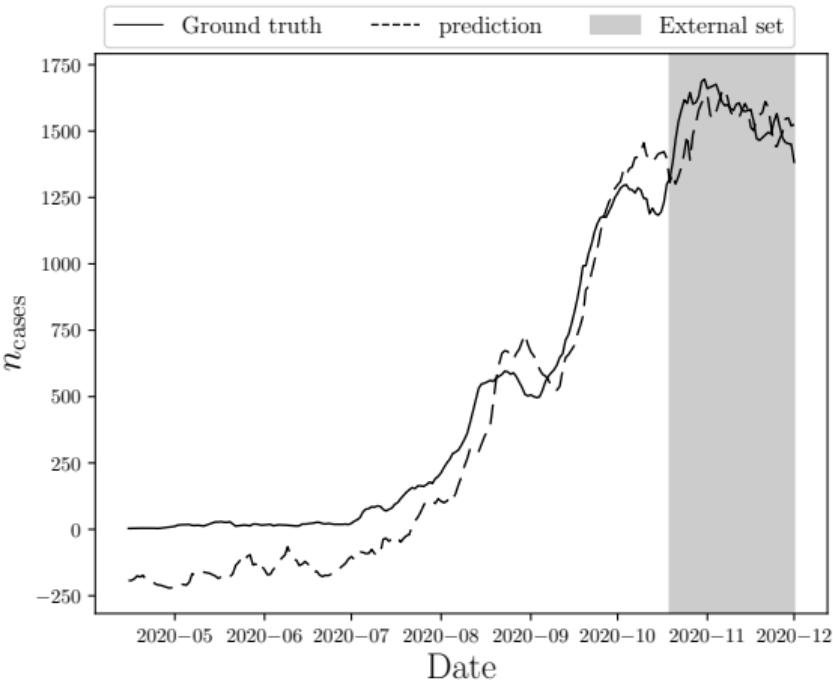
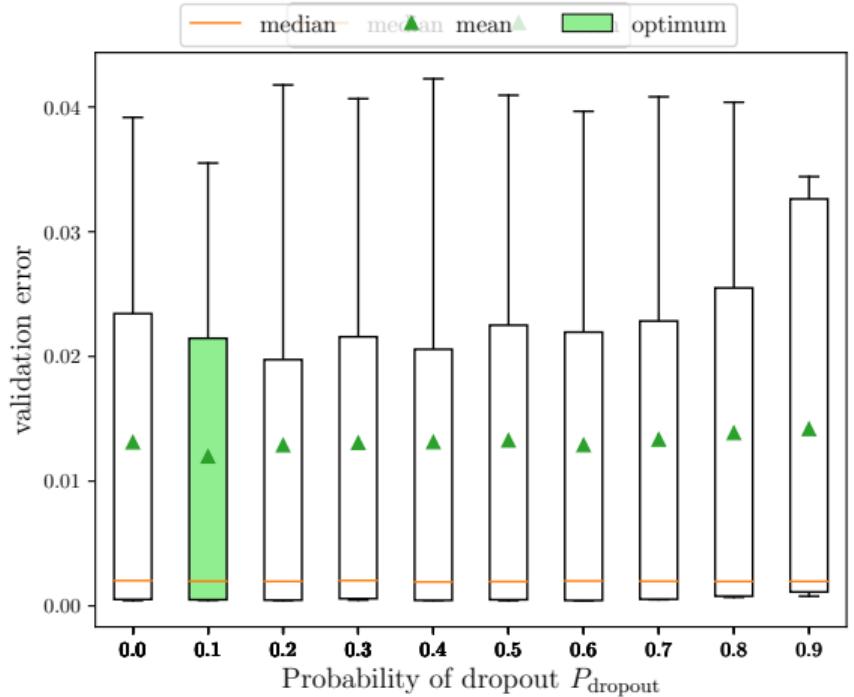
Hyperparameter tuning

We can use StoMADS to solve such hyperparameter optimization problems¹



Hyperparameter tuning

We can use StoMADS to solve such hyperparameter optimization problems¹



Hyperparameter tuning

We can use StoMADS to solve such hyperparameter optimization problems¹

Objective and constraints

$$\min_x \quad f(x) = \mathbb{E}_{\Theta} [f_{\Theta}(x) = \text{error}_{\text{CV}}]$$

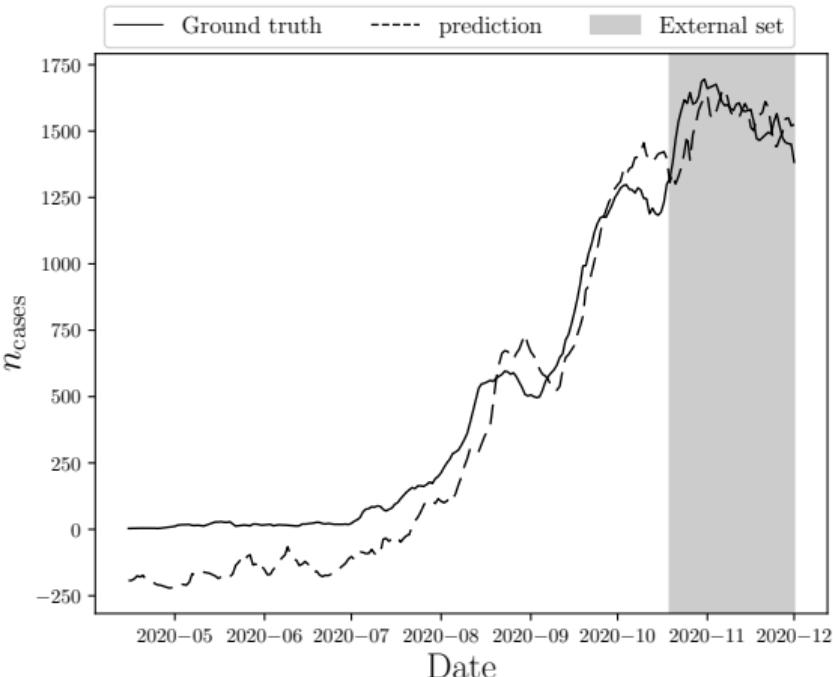
where Θ : realizations

Design variables (\mathbf{x})

- T_1 : Input dimension
- n_{hidden} : Number of hidden neurons
- P_{dropout} : Probability of dropout, etc.

Randomly seeded parameters

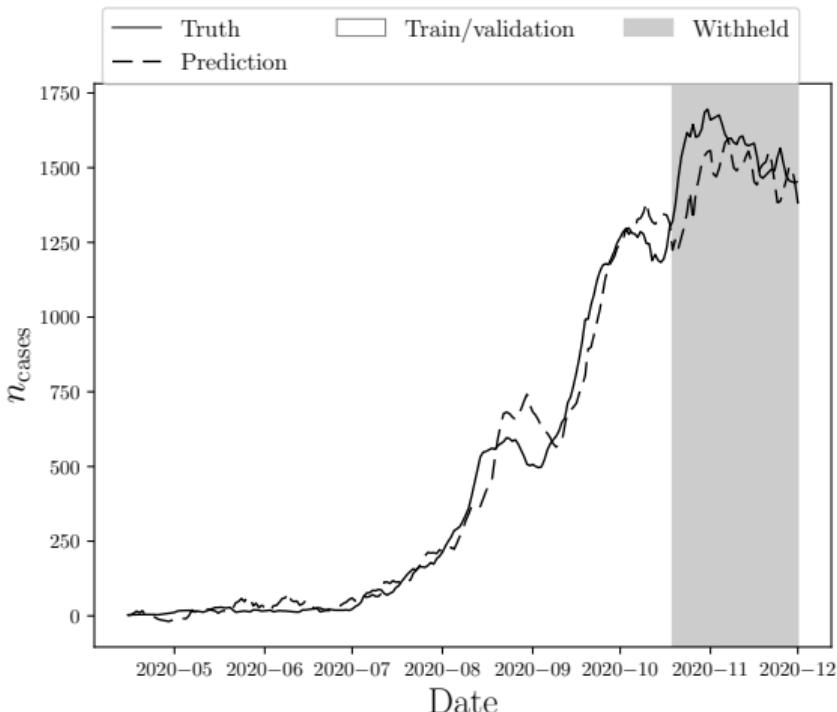
- Initial weights
- Gradient descent steps



Results: Optimal hyperparameters

Optimal hyperparameters for the *Seq2Seq* model:

Hyperparameter		Value
Sliding window size	T_1	6
Number of hidden neurons	n_{hidden}	1500
Probability of dropout	P_{dropout}	0.8
Number of hidden layers	n_{hidden}	2
Teacher forcing probability	P_{teacher}	0.3
Learning rate	l_{rate}	1×10^{-4}
batch size	b_{size}	32

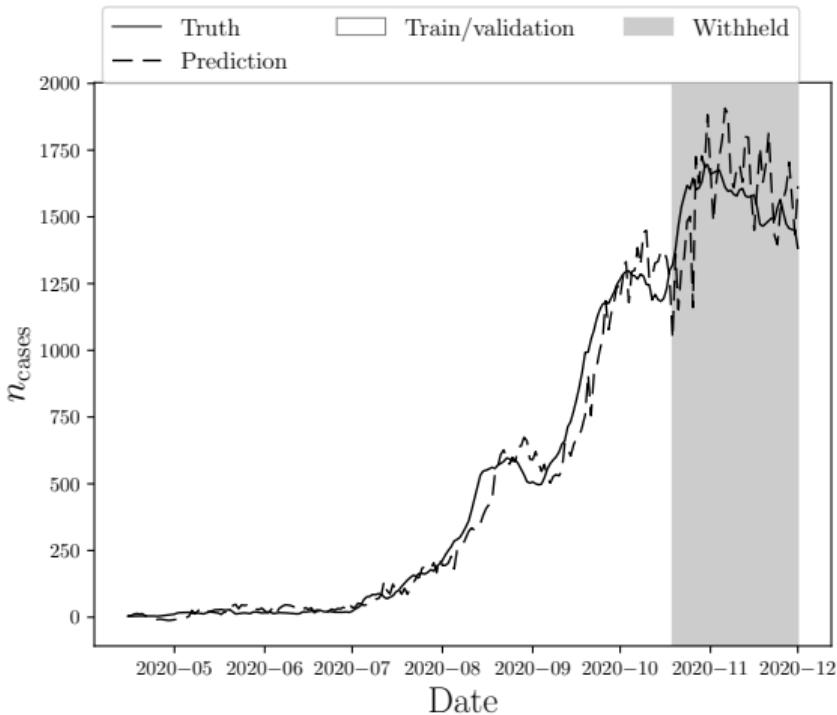




Results: Optimal hyperparameters

Optimal hyperparameters for the *support vector machine regression (SVR)* model:

Hyperparameter	Value
Sliding window size	T_1
Ridge factor	λ
Margin of tolerance	ϵ
Stopping criteria tolerance	ϵ_{tol}
Learning rate	l_{rate}



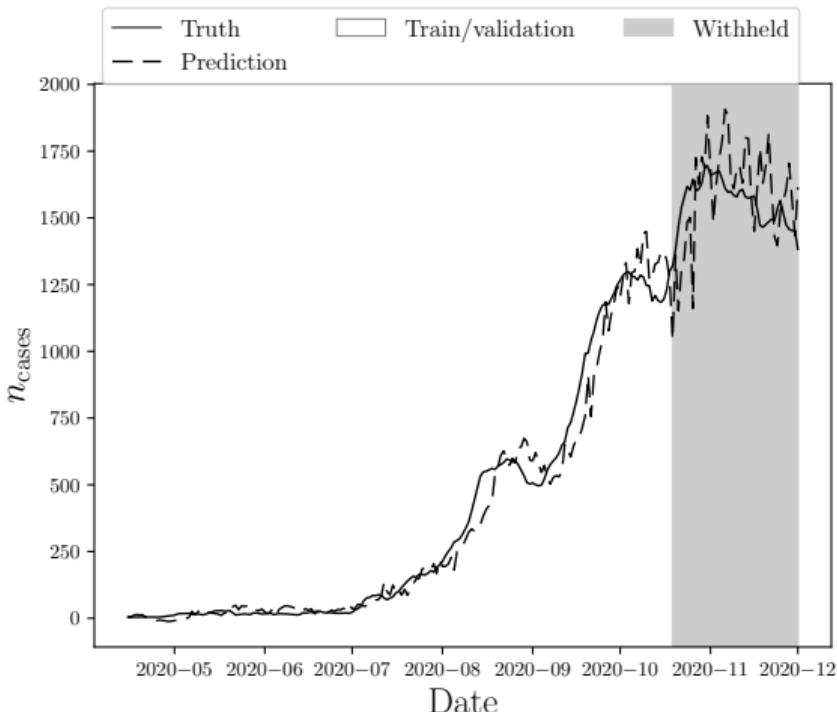


Results: Optimal hyperparameters

Optimal hyperparameters for the *support vector machine regression (SVR)* model:

Hyperparameter	Value
Sliding window size	T_1
Ridge factor	λ
Margin of tolerance	ϵ
Stopping criteria tolerance	ϵ_{tol}
Learning rate	l_{rate}

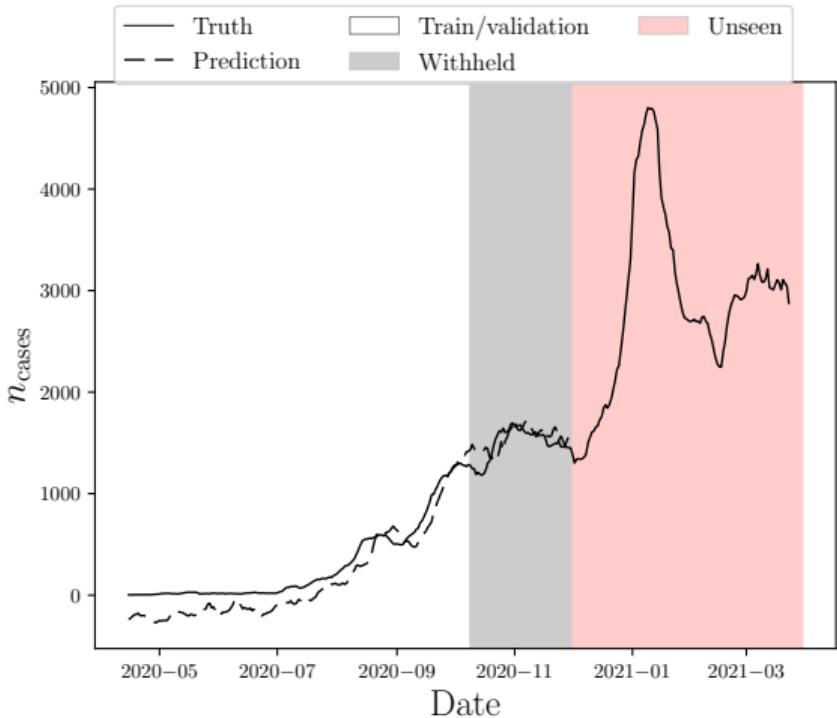
Support vector machine models have
deterministic performance



Results: Prospective validation

Performance of models on **unseen** data (first 4 months of 2021):

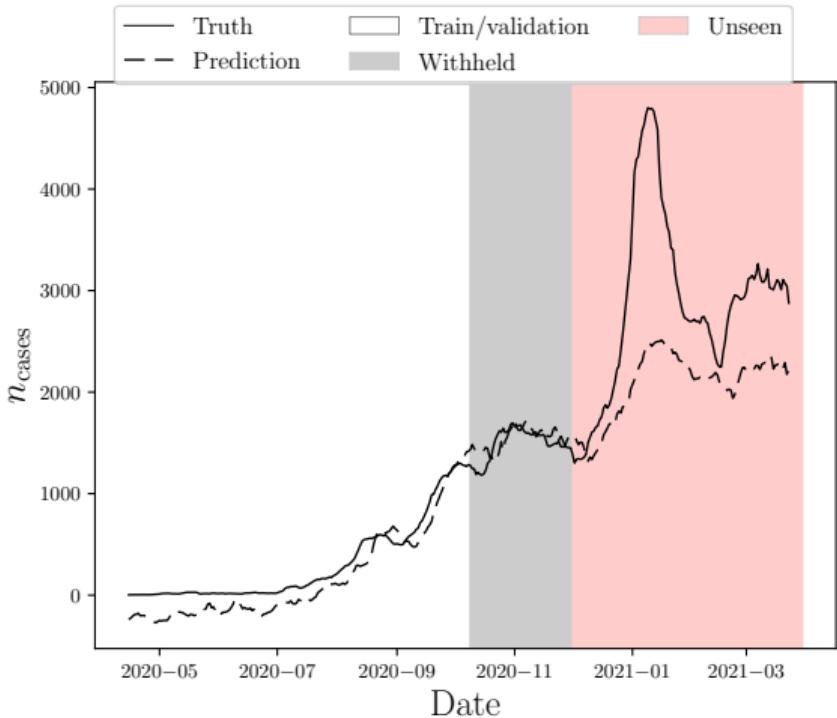
Model	Test error
Seq2Seq	0.571
Long short term memory (LSTM) cell	0.326
feedforward neural network	0.255
Support vector machine	0.168
Gradient boosting	1.444
Linear regression	0.160



Results: Prospective validation

Performance of models on **unseen** data (first 4 months of 2021):

Model	Test error
Seq2Seq	0.571
Long short term memory (LSTM) cell	0.326
feedforward neural network	0.255
Support vector machine	0.168
Gradient boosting	1.444
Linear regression	0.160

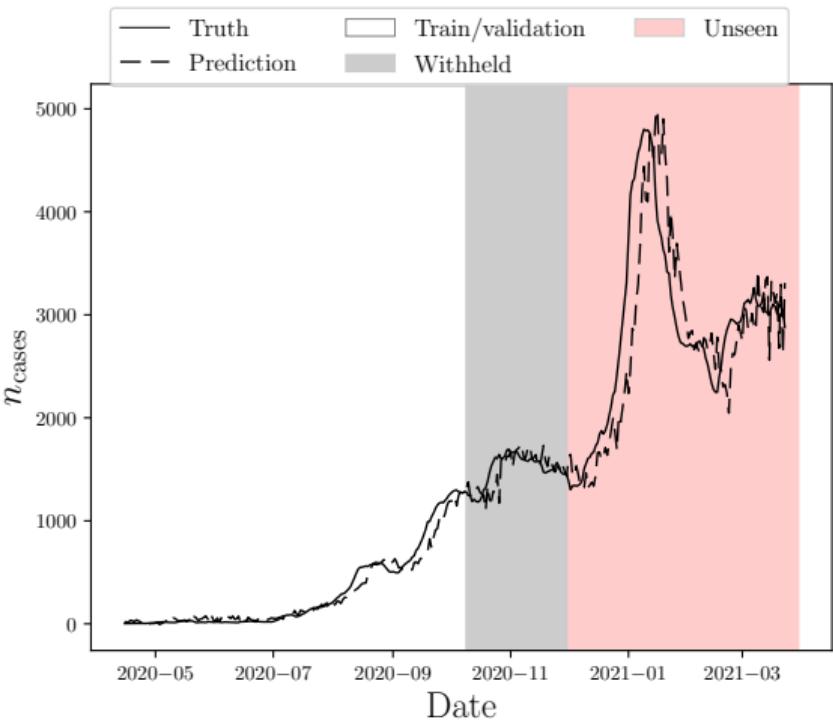




Results: Prospective validation

Performance of models on **unseen** data (first 4 months of 2021):

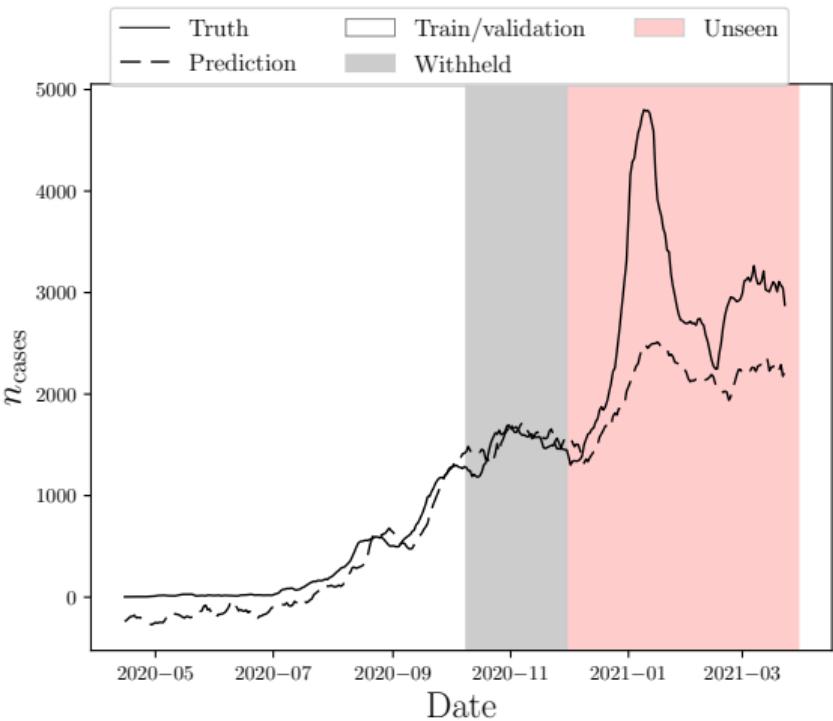
Model	Test error
Seq2Seq	0.571
Long short term memory (LSTM) cell	0.326
feedforward neural network	0.255
Support vector machine	0.168
Gradient boosting	1.444
Linear regression	0.160



Results: Prospective validation

Effect of increasing number of training days (Adding 1 month of data):

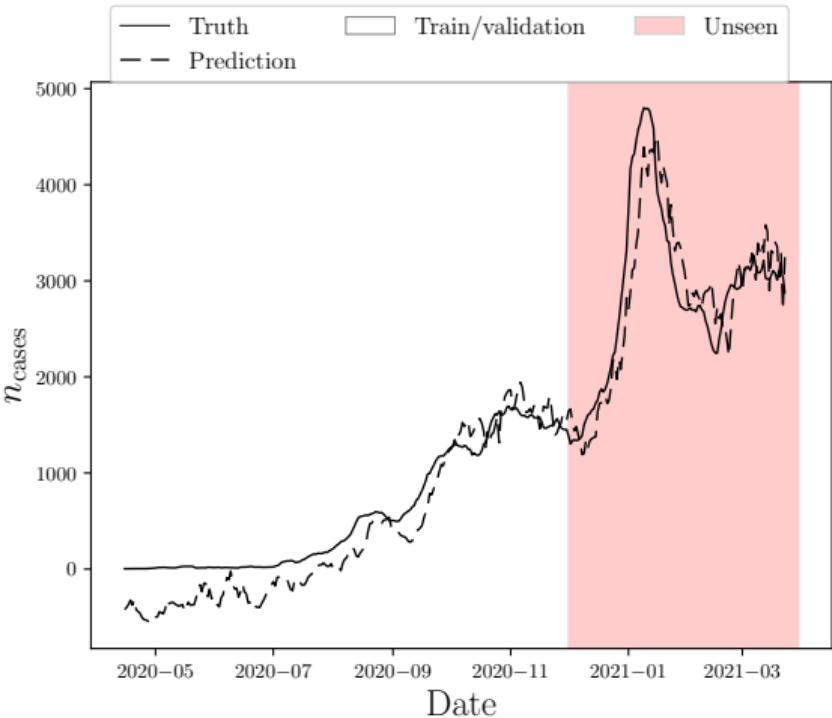
Model	Test error
Seq2Seq	0.571
Long short term memory (LSTM) cell	0.326
feedforward neural network	0.255
Support vector machine	0.168
Gradient boosting	1.444
Linear regression	0.160



Results: Prospective validation

Effect of increasing number of training days (Adding 1 month of data):

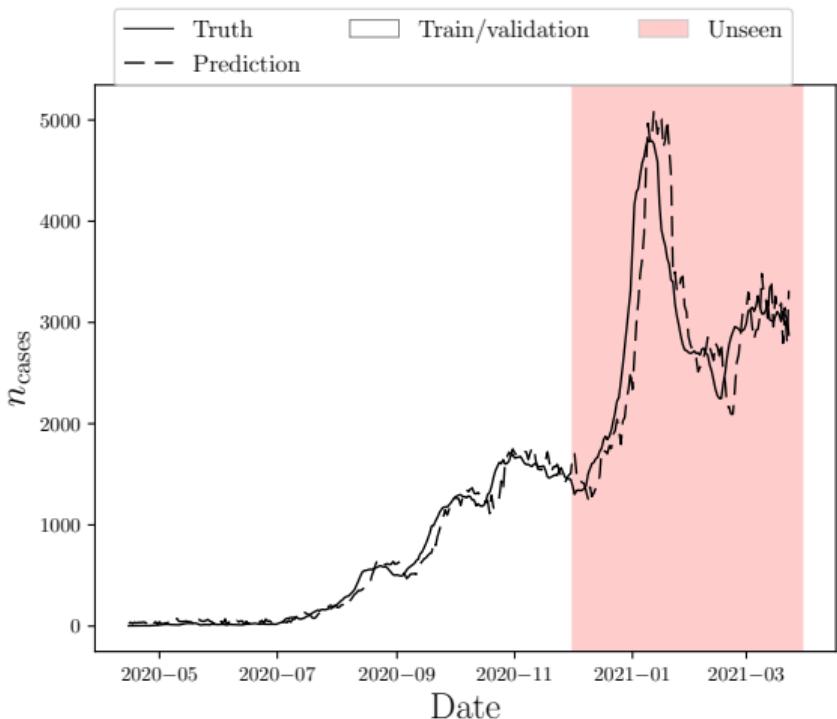
Model	Test error
Seq2Seq	0.106
Long short term memory (LSTM) cell	0.326
feedforward neural network	0.255
Support vector machine	0.168
Gradient boosting	1.444
Linear regression	0.160



Results: Prospective validation

Effect of increasing number of training days (Adding 1 month of data):

Model	Test error
Seq2Seq	0.106
Long short term memory (LSTM) cell	0.326
feedforward neural network	0.255
Support vector machine	0.140
Gradient boosting	1.444
Linear regression	0.160





Conclusion and future directions

Model discovery and development facilitated by hyperparameter optimization



Conclusion and future directions

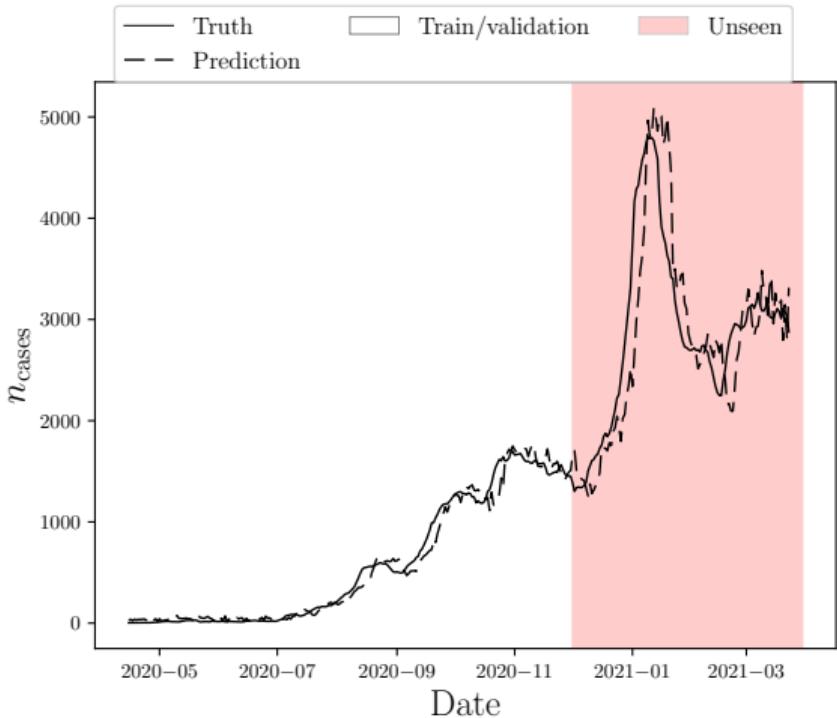
Model discovery and development facilitated by hyperparameter optimization

- Cycle Threshold (C_t) is a useful feature for incidence projection



Model discovery and development facilitated by hyperparameter optimization

- Cycle Threshold (C_t) is a useful feature for incidence projection
- Model that generalizes well on unseen data

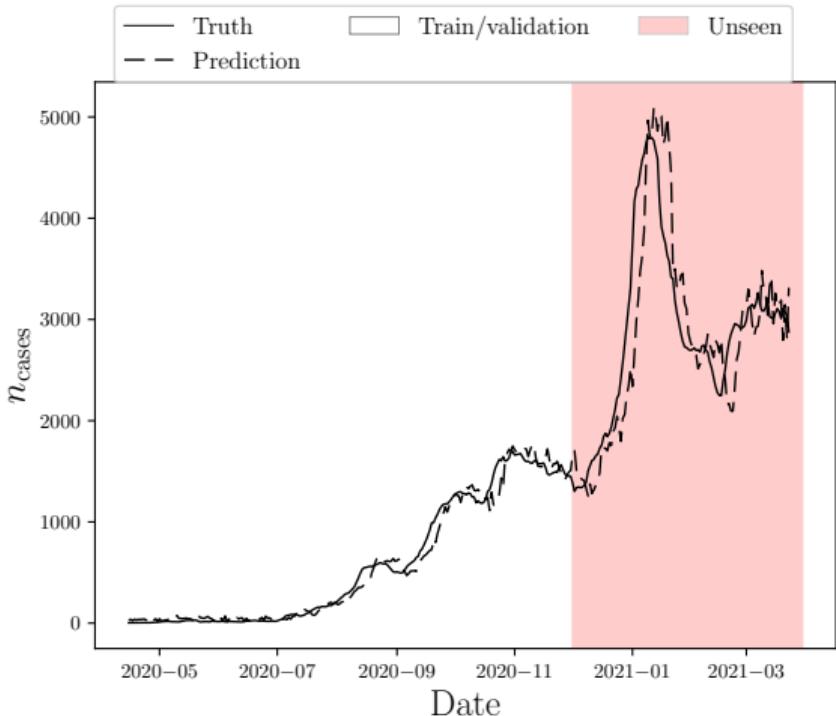


Conclusion and future directions



Model discovery and development facilitated by hyperparameter optimization

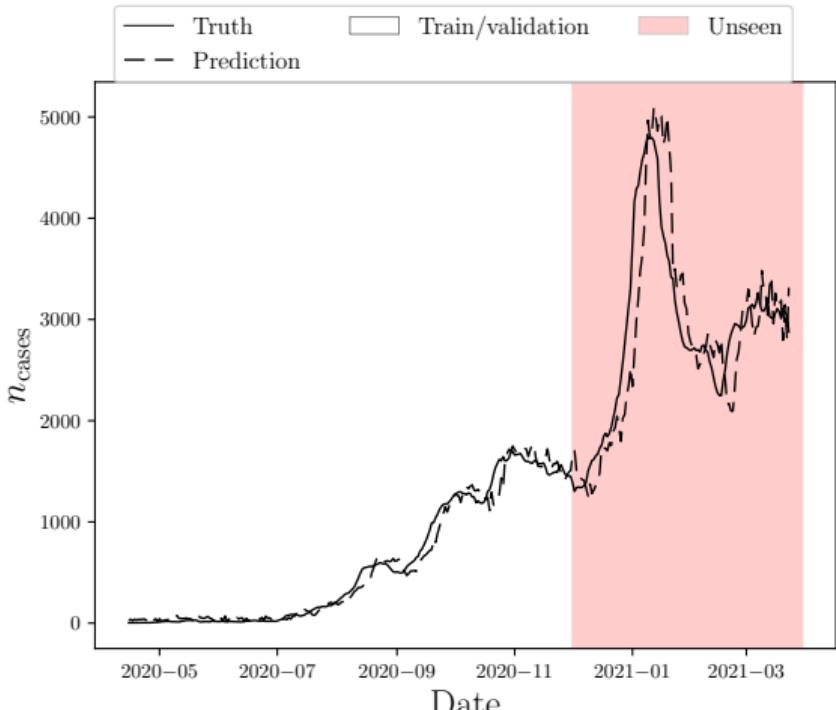
- Cycle Threshold (C_t) is a useful feature for incidence projection
- Model that generalizes well on unseen data
- Simpler models perform well when historical data is limited





Model discovery and development facilitated by hyperparameter optimization

- Cycle Threshold (C_t) is a useful feature for incidence projection
- Model that generalizes well on unseen data
- Simpler models perform well when historical data is limited
- Works well on other datasets¹



[1] Hay2021



StoMADS can be improved to solve a wide variety of HPO problems

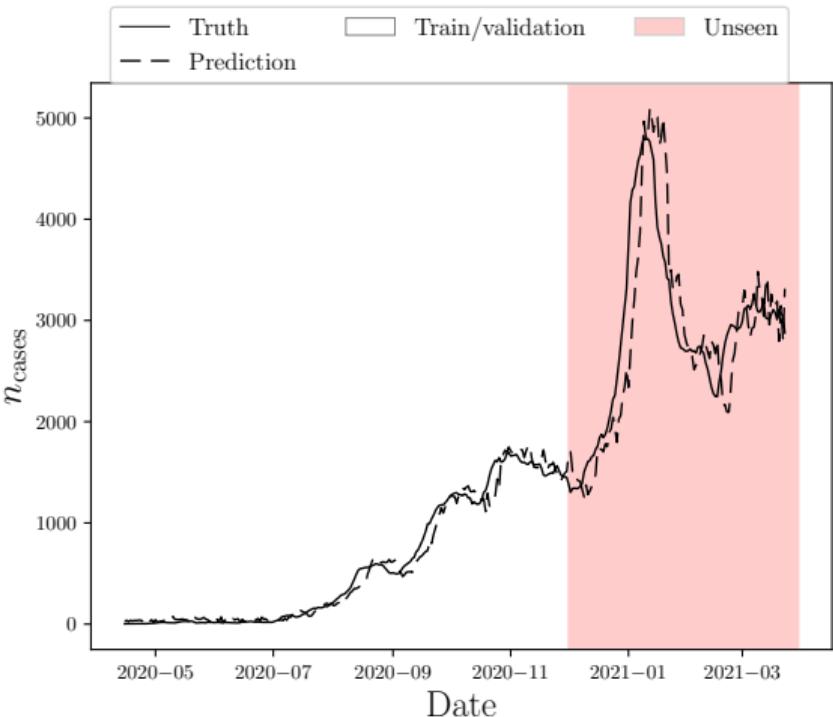
- Can be used to meet deployment targets

Objective and constraints

$$\min_x \quad f(x) = \mathbb{E}_{\Theta} [f_{\Theta}(x) = \text{error}_{\text{CV}}]$$

$$\text{subject to } c(x) = \text{inference time} - \text{threshold} \leq 0$$

where Θ : realizations





StoMADS can be improved to solve a wide variety of HPO problems

- Can be used to meet deployment targets

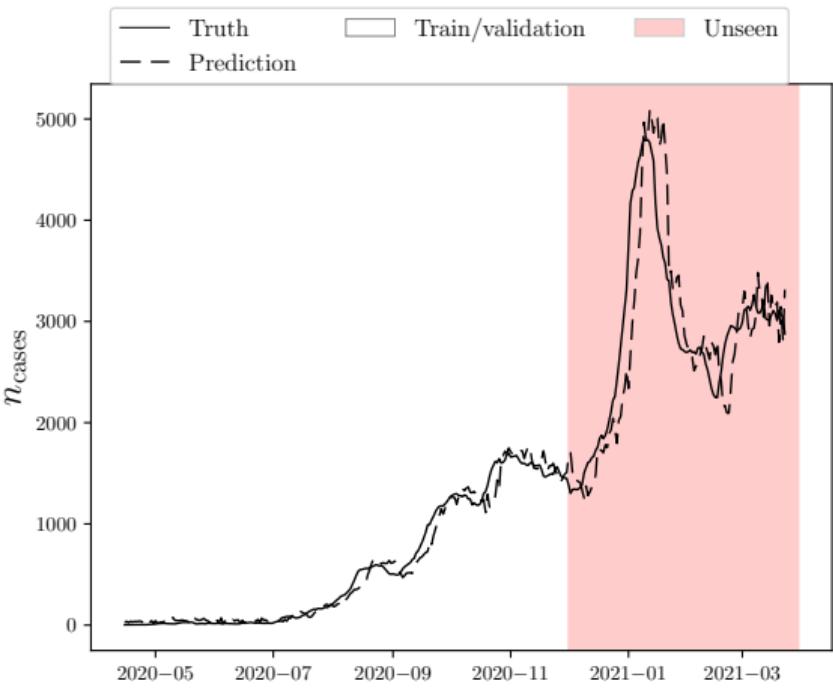
Objective and constraints

$$\min_x \quad f(x) = \mathbb{E}_{\Theta} [f_{\Theta}(x) = \text{error}_{\text{CV}}]$$

$$\text{subject to} \quad c(x) = \text{inference time} - \text{threshold} \leq 0$$

where Θ : realizations

- Should be benchmarked against HyperNOMAD¹, Bayesian optimization²



[1]Lakhmiri2019

[2]Bergstra2011



StoMADS can be improved to solve a wide variety of HPO problems

- Can be used to meet deployment targets

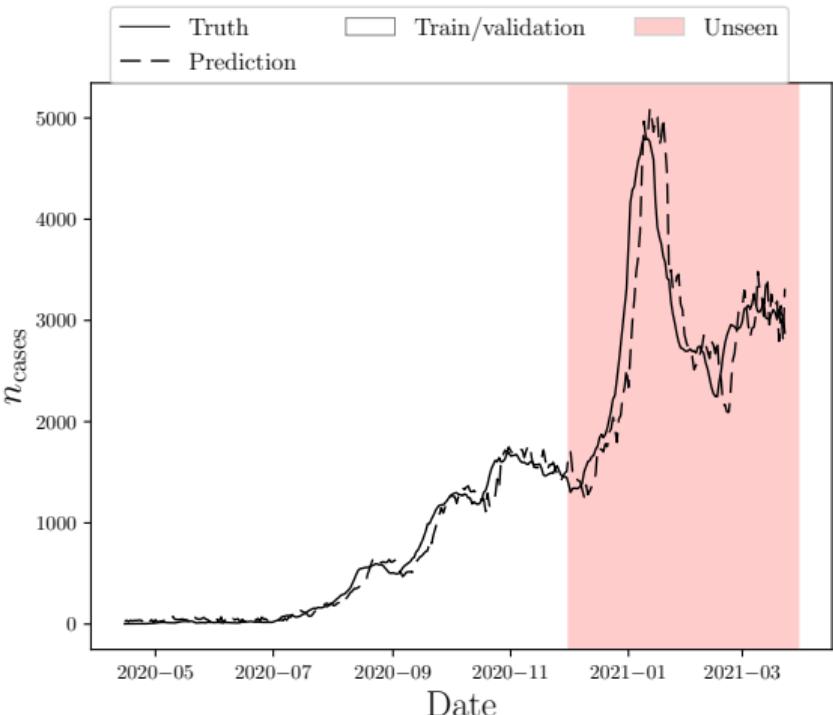
Objective and constraints

$$\min_x \quad f(x) = \mathbb{E}_{\Theta} [f_{\Theta}(x) = \text{error}_{\text{CV}}]$$

$$\text{subject to} \quad c(x) = \text{inference time} - \text{threshold} \leq 0$$

where Θ : realizations

- Should be benchmarked against HyperNOMAD¹, Bayesian optimization²
- Mixed variable version is needed



[1]Lakhmiri2019

[2]Bergstra2011

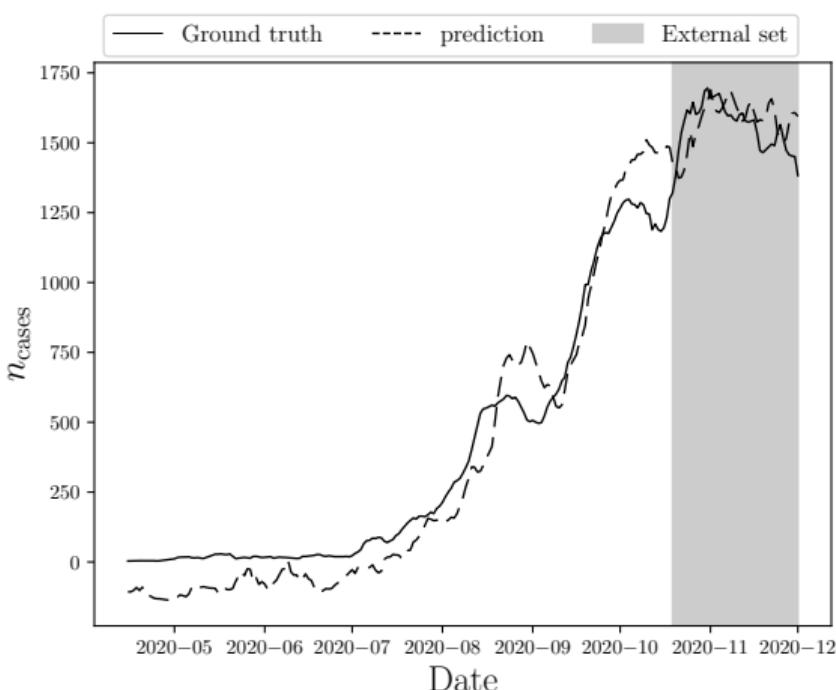
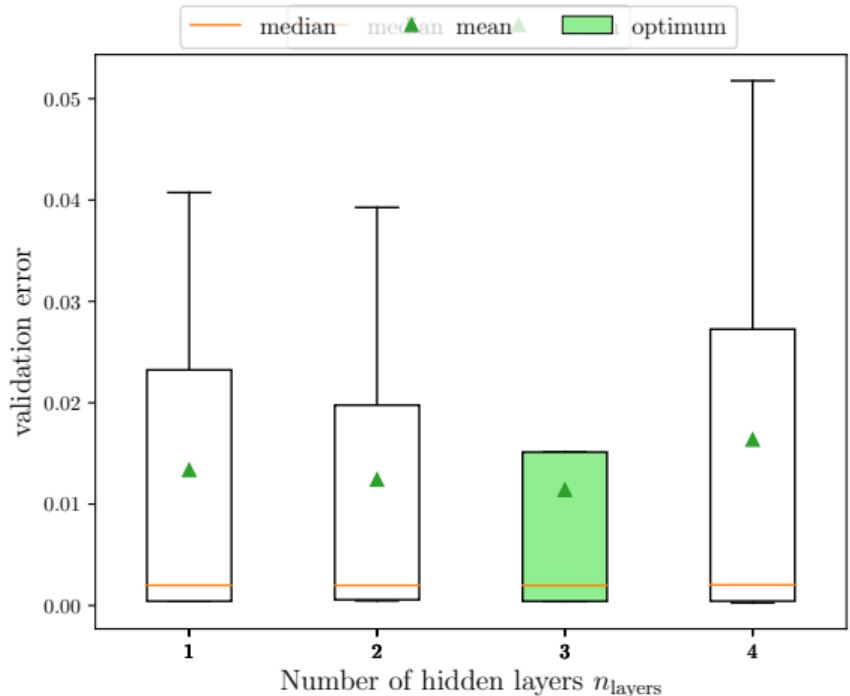
Thank you for your time

Open source GPU implementation of agent-based modeling¹



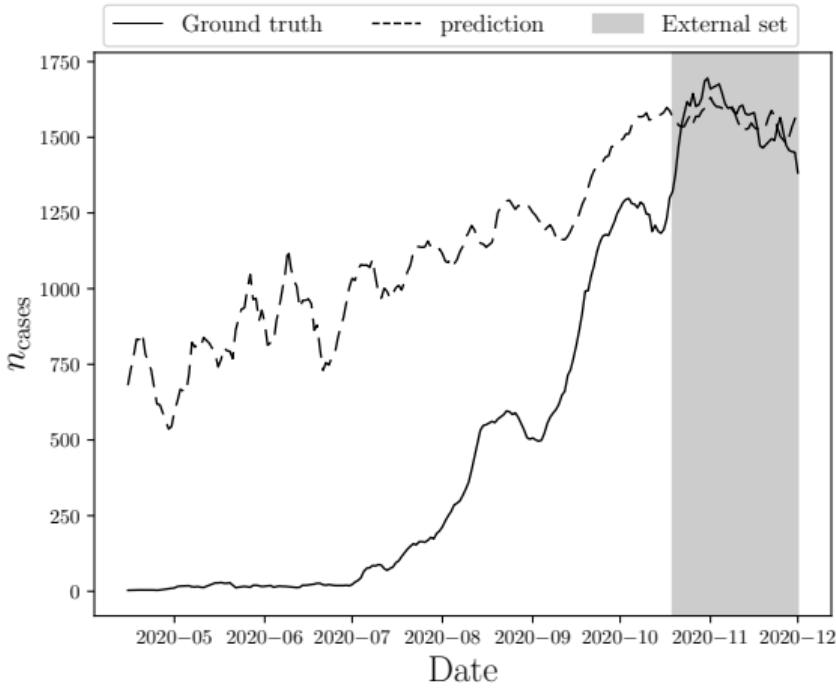
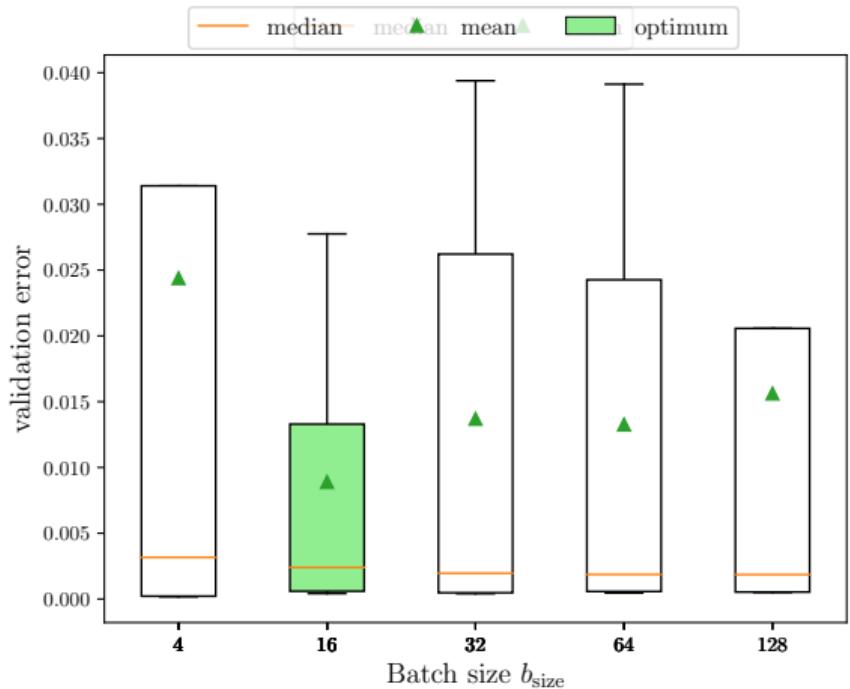
Hyperparameter tuning: other hyperparameters

We can use StoMADS to solve such hyperparameter optimization problems¹



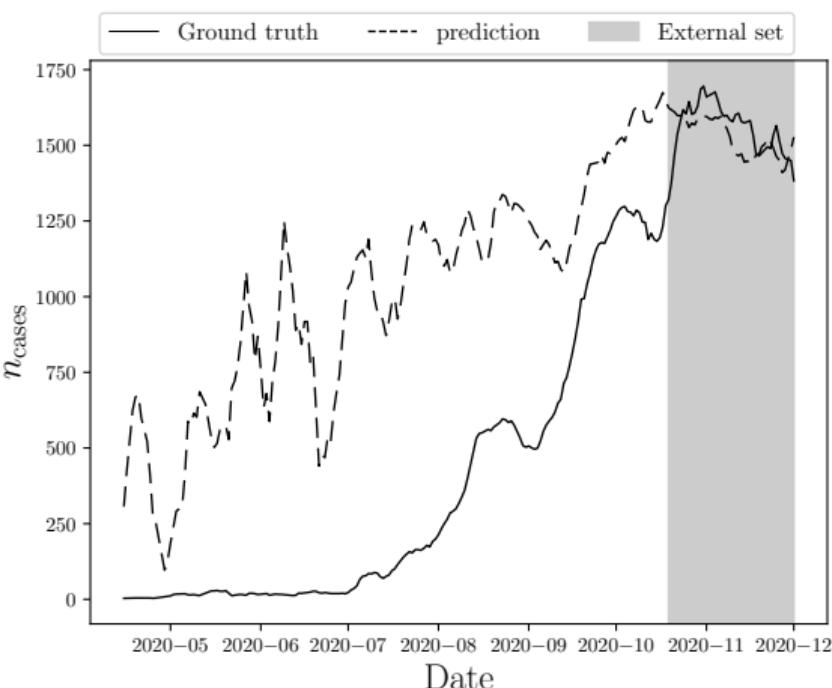
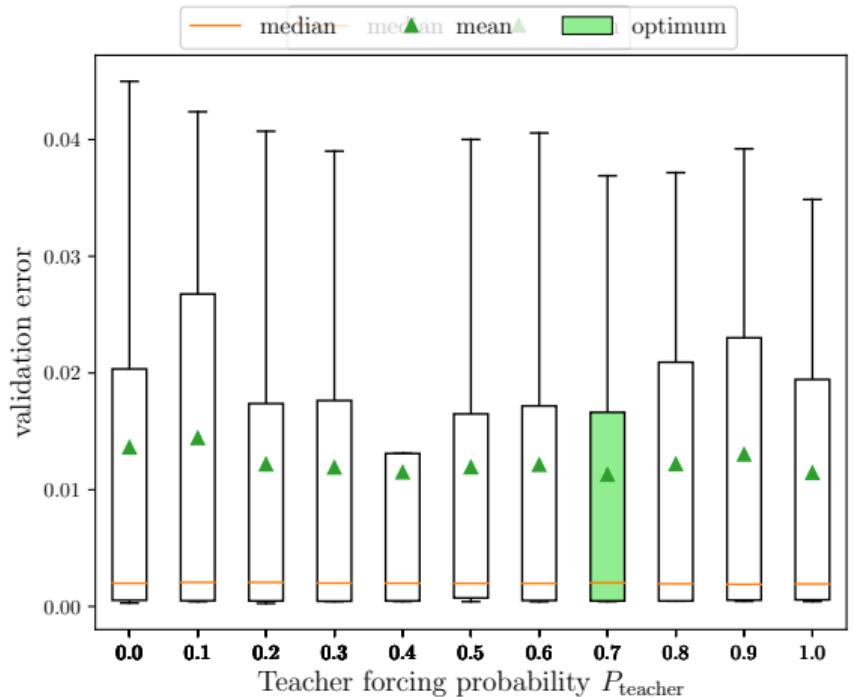
Hyperparameter tuning: other hyperparameters

We can use StoMADS to solve such hyperparameter optimization problems¹



Hyperparameter tuning: other hyperparameters

We can use StoMADS to solve such hyperparameter optimization problems¹



Hyperparameter tuning: other hyperparameters

We can use StoMADS to solve such hyperparameter optimization problems¹

