June 28 2022

Agro-ecological modeling for sustainable crop production at field scale

David Makowski

What is a model?

Mathematical model = Tool to compute outputs Y from inputs X

$$Y \longleftarrow \text{Model} \longleftarrow X$$

What is a model?

Mathematical model = Tool to compute outputs Y from inputs X

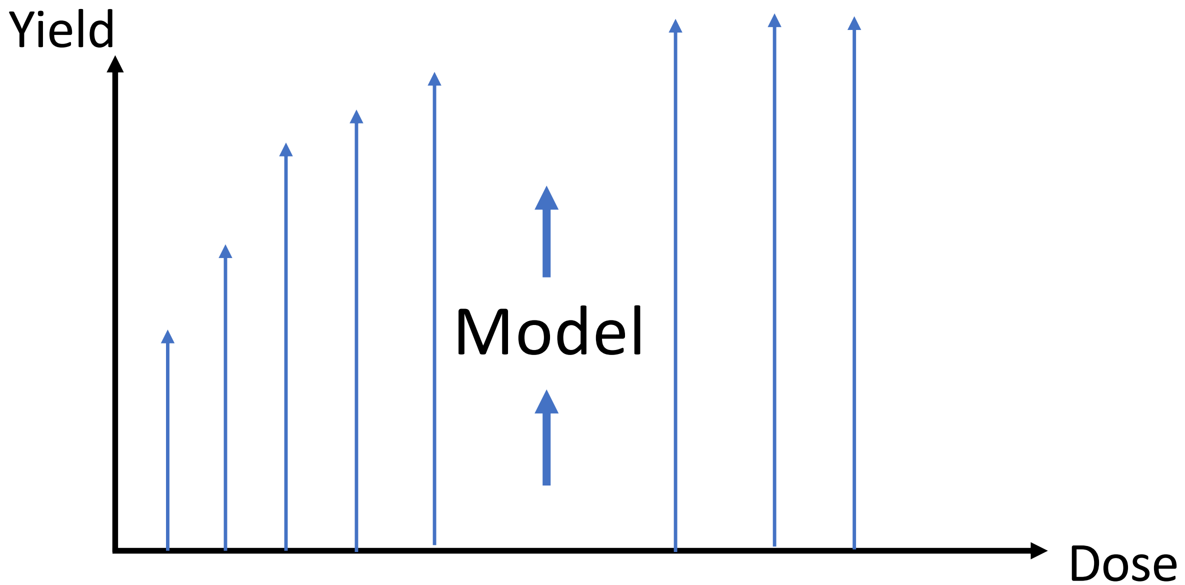
Wheat yield ← Model ← N fertilizer dose

What is a model?

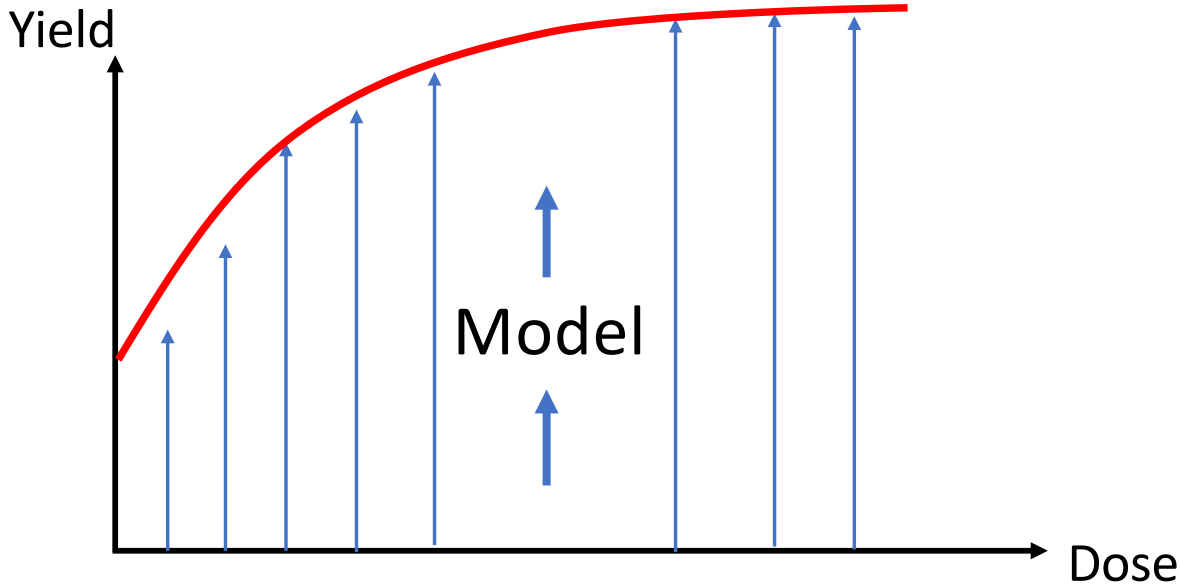
Mathematical model = Tool to compute outputs Y from inputs X

8 t ha^{-1} ⟵ Model ⟵ 200 kg ha^{-1}

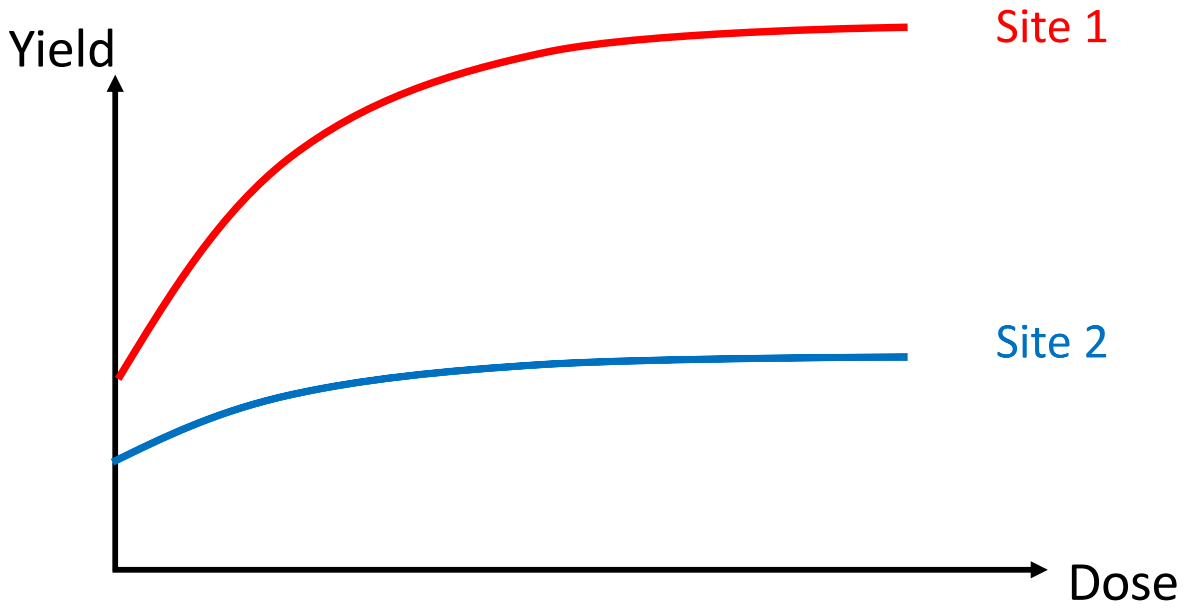
What is a model?



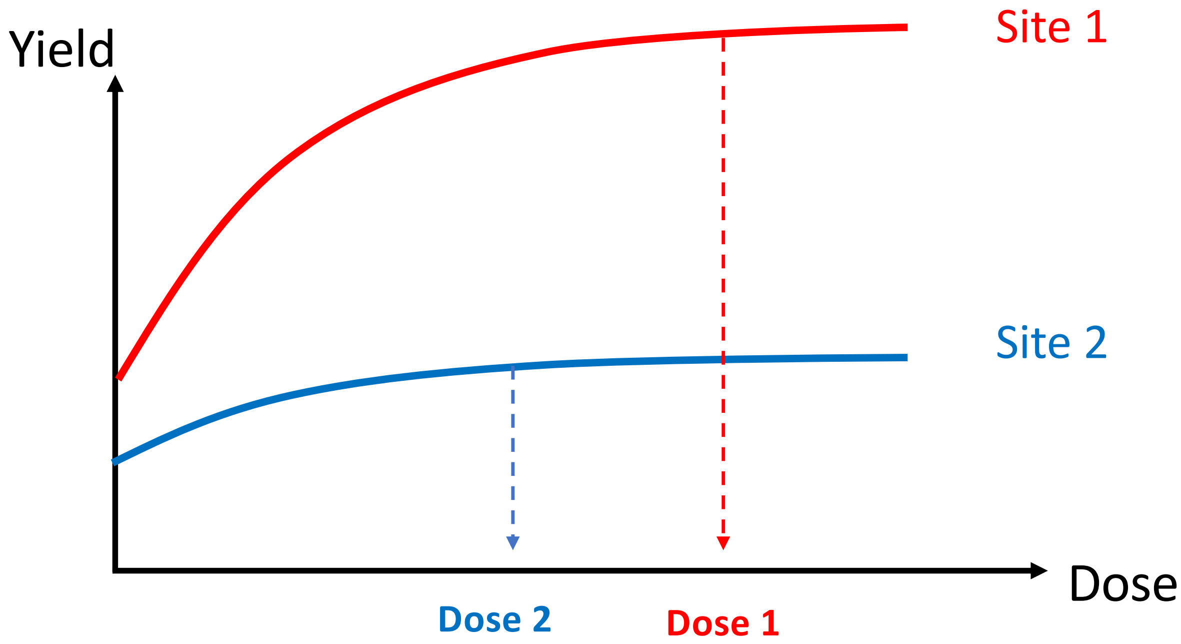
What is a model?



What is a model?

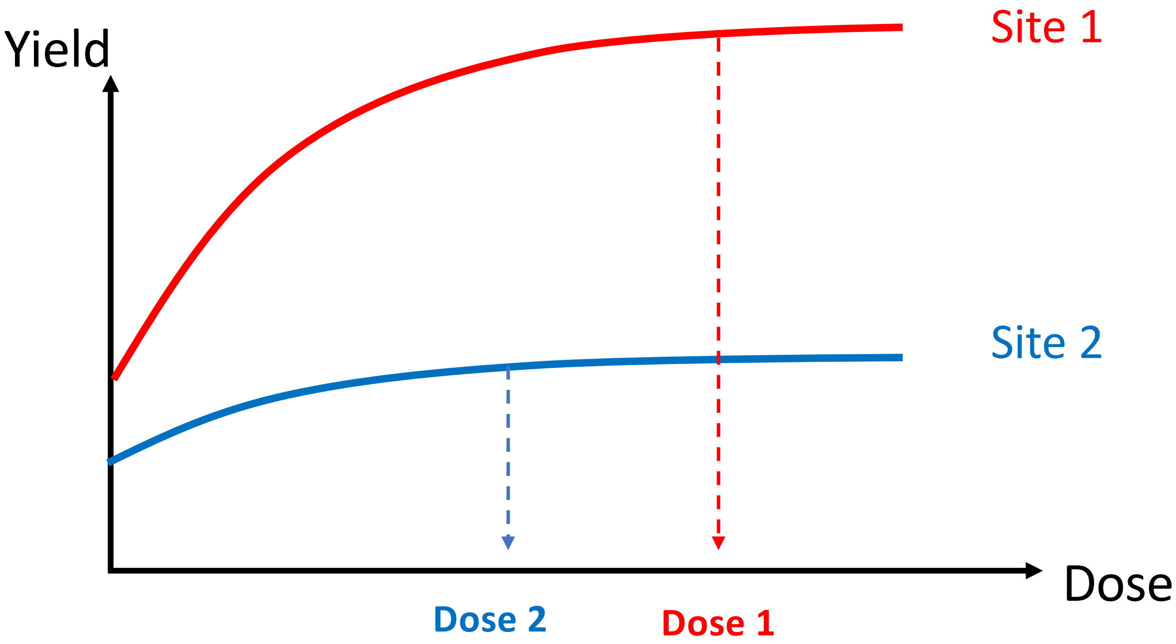


What is a model?



What is a model?

Model-based recommandations can contribute to chemical input use reduction



What is a model?

Mathematical model = Tool to compute outputs Y from inputs X

Yield

Greenhouse gas emission

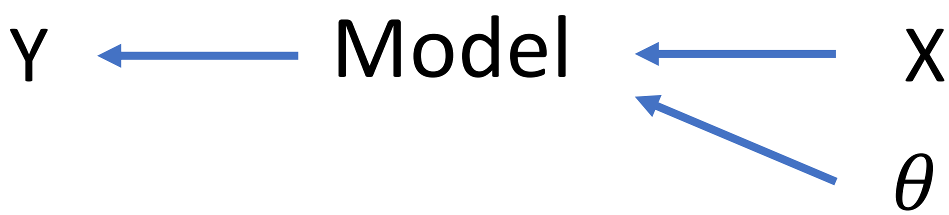
Soil carbon

Model

Fertilizer dose
Irrigation
Temperatures
Precipitations
Radiations
Soil characteristics
etc.

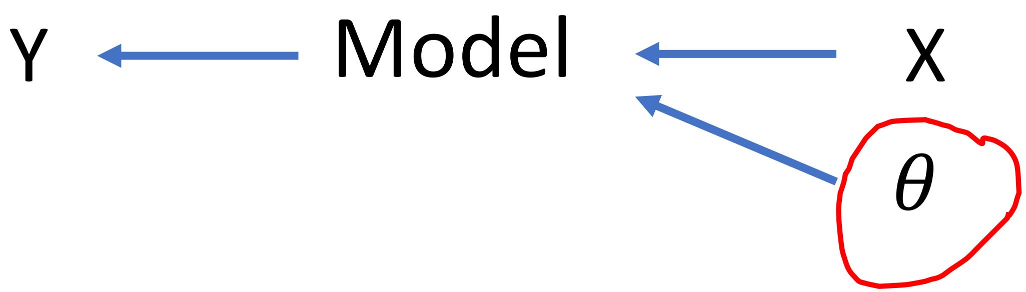
What is a model?

Mathematical model = Tool to compute outputs Y from inputs X **and parameters** θ



What is a model?

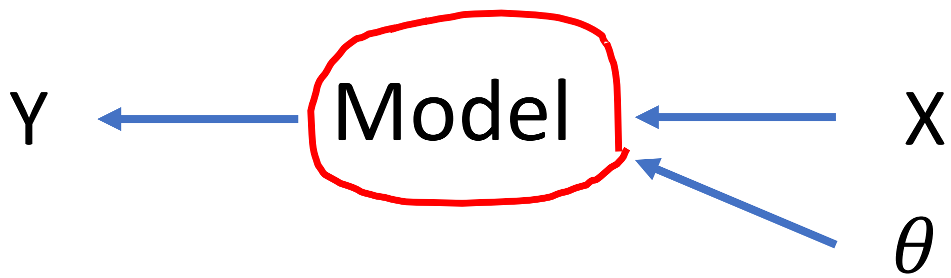
Mathematical model = Tool to compute outputs Y from inputs X **and parameters** θ



Parameter values are unknown and need to be estimated before running the model.

What is a model?

Mathematical model = Tool to compute outputs Y from inputs X **and parameters** θ



A wide variety of models can be used to address a specific objective

Models can be very simple or… very complex

Model type	**Example**	**Complexity**
Linear model	$Y = \theta_1 + \theta_2 X$	1 input, 2 parameters

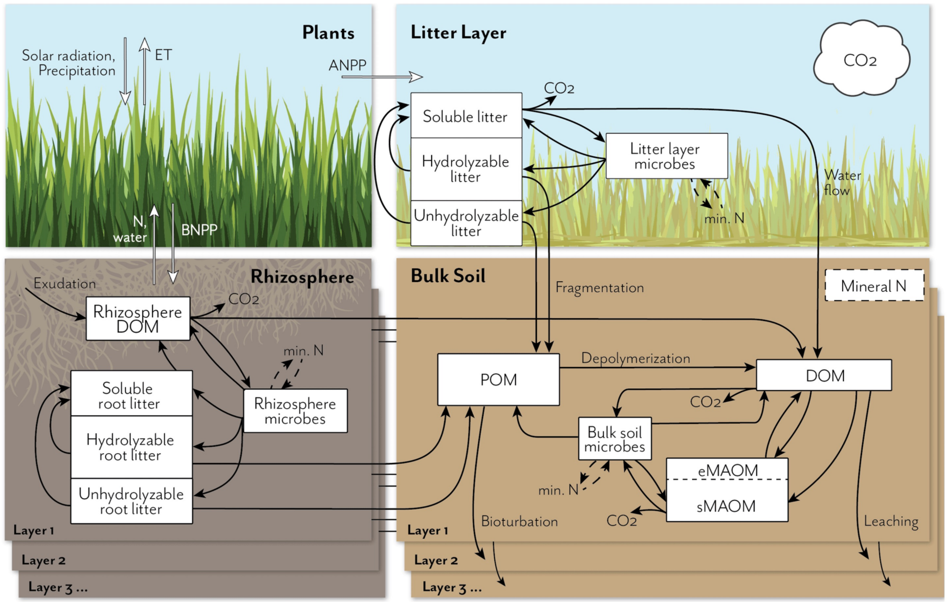
Models can be very simple or… very complex

Model type	**Example**	**Complexity**
Linear model	$Y = \theta_1 + \theta_2 X$	1 input, 2 parameters
Linear model	$Y = \theta_1 + \theta_2 X_1 + \theta_3 X_2$	2 inputs, 3 parameters

Models can be very simple or… very complex

Model type	Example	Complexity
Linear model	$Y = \theta_1 + \theta_2 X$	1 input, 2 parameters
Linear model	$Y = \theta_1 + \theta_2 X_1 + \theta_3 X_2$	2 inputs, 3 parameters
Nonlinear model	$Y = \theta_1[1 - \theta_2 exp(-\theta_3 X)]$	1 input, 3 parameters

Models can be very simple or… very complex

Model type	Example	Complexity
Mechanistic	**model simulating carbon and nitrogen dynamics**  https://doi.org/10.5194/bg-18-3147-2021	~ 100

Models can be very simple or... very complex

Model type	**Example**	**Complexity**
Mechanistic	**model simulating carbon and nitrogen dynamics** (equations below)	~100

Equations

Surface litter

$$\frac{dC_{Ssoluble}}{dt} = -C_{Ssoluble} * k_{soluble} * T_{eff} * W_{eff} * LCI_{eff} * MicCN_{eff} - C_{Ssoluble} * k_{solubleLeach} * W_{leach} + C_{Shydro} * k_{hydro} * T_{eff} * W_{eff} * LCI_{eff} * MicCN_{eff} + C_{Sunhydro} * k_{unhydro} * T_{eff} * W_{eff} * MicCN_{eff} + C_{SmicLitter} * k_{micDeath} * frac_{toSoluble}$$

$$\frac{dC_{Shydro}}{dt} = -C_{Shydro} * k_{hydro} * T_{eff} * W_{eff} * LCI_{eff} * MicCN_{eff} - C_{Shydro} * k_{fragment} * T_{eff} * W_{eff} + C_{SmicLitter} * k_{micDeath} * frac_{toHydro}$$

$$\frac{dC_{Sunhydro}}{dt} = -C_{Sunhydro} * k_{unhydro} * T_{eff} * W_{eff} * MicCN_{eff} - C_{Sunhydro} * k_{fragment} * T_{eff} * W_{eff} + C_{SmicLitter} * k_{micDeath} * frac_{toUnhydro}$$

Note: unlike the soluble and hydrolysable pools, no LCI_{eff} on unhydrolysable pool decay.

$$\frac{dC_{SmicLitter}}{dt} = -C_{SmicLitter} * k_{micDeath} + C_{Ssoluble} * k_{soluble} * T_{eff} * W_{eff} * LCI_{eff} * MicCN_{eff} * CUE_{Ssoluble}$$

$$\frac{dC_{SCO_2}}{dt} = C_{Ssoluble} * k_{soluble} * T_{eff} * W_{eff} * LCI_{eff} * MicCN_{eff} * (1 - CUE_{Ssoluble})$$

Rhizosphere litter

$$\frac{dC_{Rsoluble}}{dt} = -C_{Rsoluble} * k_{solubleLeach} * LCI_{eff} + C_{Rhydro} * k_{hydro} * T_{eff} * W_{eff} * LCI_{eff} * MicCN_{eff} + C_{Runhydro} * k_{unhydro} * T_{eff} * W_{eff} * MicCN_{eff} + C_{RmicLitter} * k_{micDeath} * frac_{toSoluble}$$

$$\frac{dC_{Rhydro}}{dt} = -C_{Rhydro} * k_{hydro} * T_{eff} * W_{eff} * LCI_{eff} * MicCN_{eff} - C_{Rhydro} * k_{fragment} * T_{eff} * W_{eff} + C_{RmicLitter} * k_{micDeath} * frac_{toHydro}$$

$$\frac{dC_{Runhydro}}{dt} = -C_{Runhydro} * k_{unhydro} * T_{eff} * W_{eff} * MicCN_{eff} - C_{Runhydro} * k_{fragment} * T_{eff} * W_{eff} + C_{RmicLitter} * k_{micDeath} * frac_{toUnhydro}$$

$$\frac{dC_{RDOM}}{dt} = -C_{RDOM} * k_{soluble} * T_{eff} * W_{eff} * MicCN_{eff} - C_{RDOM} * k_{RDOMLeach} * WFPS^3 + C_{Rsoluble} * k_{solubleLeach} * LCI_{eff} + C_{exudate} * k_{exudate}$$

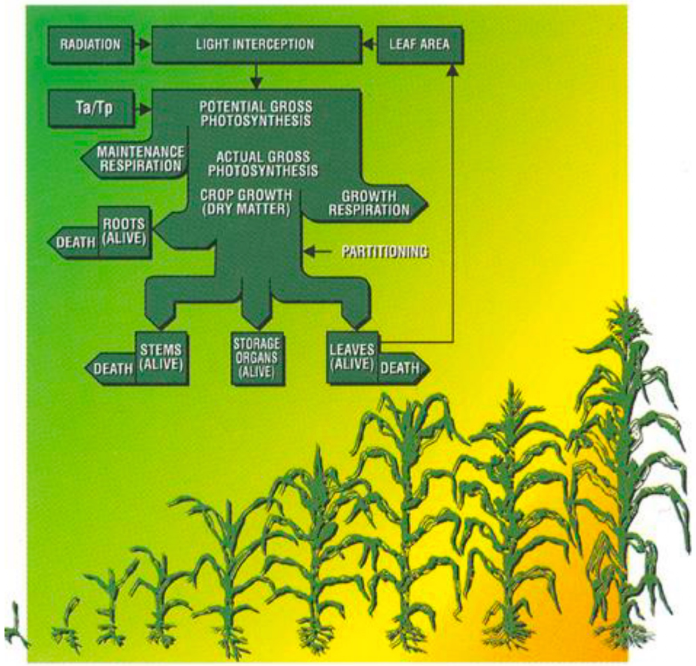
Note: the decay rate of surface soluble litter $k_{soluble}$ is also used for RDOM.

$$\frac{dC_{RmicLitter}}{dt} = -C_{RmicLitter} * k_{micDeath} + C_{RDOM} * k_{soluble} * T_{eff} * W_{eff} * MicCN_{eff} * CUE_{RDOM}$$

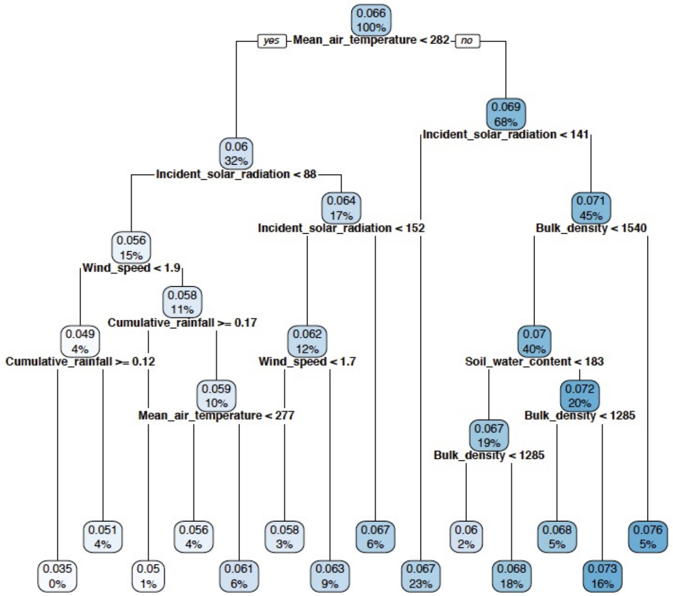
$$\frac{dC_{RCO_2}}{dt} = C_{RDOM} * k_{soluble} * T_{eff} * W_{eff} * MicCN_{eff} * (1 - CUE_{RDOM})$$

https://doi.org/10.5194/bg-18-3147-2021

Models can be very simple or… very complex

Model type	**Example**	**Complexity**
Mechanistic	Crop model WOFOST 	~100

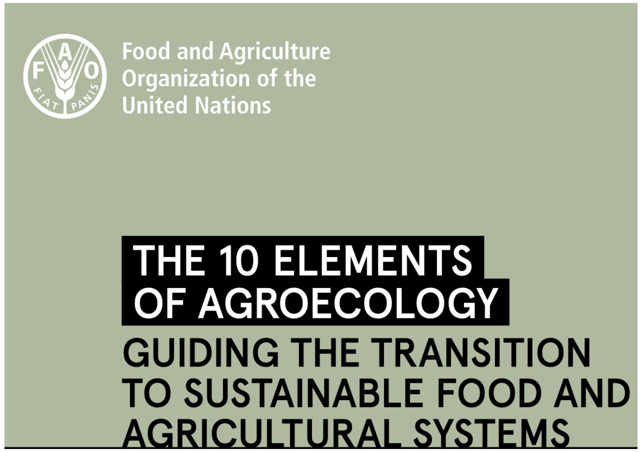
Models can be very simple or… very complex

Model type	**Example**	**Complexity**
Machine learning		100, 1000 or more

Why modelling is useful for agro-ecology?

Why modelling is useful for agro-ecology?

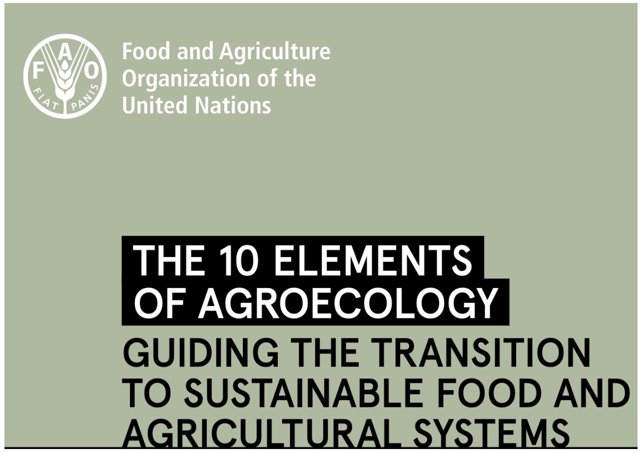
« Agroecology seeks to optimize the interactions between plants, animals, humans and the environment while taking into consideration the social aspects that need to be addressed for a sustainable and fair food system »


Food and Agriculture
Organization of the
United Nations

THE 10 ELEMENTS
OF AGROECOLOGY
GUIDING THE TRANSITION
TO SUSTAINABLE FOOD AND
AGRICULTURAL SYSTEMS

Why modelling is useful for agro-ecology?

- Agroecology has multiple objectives
  - ➢Improve soil health
  - ➢Increase biodiversty
  - ➢Input reduction (ex: fertilizer)
  - ➢Improve resilience to climate change
  - ➢Efficiency
  - ➢Promote healthy diets
  - ➢Etc.



https://doi.org/10.1007/s13593-020-00646-z
https://www.fao.org/documents/card/fr/c/I9037EN/

Why modelling is useful for agro-ecology?

- Agroecology promotes a great diversity of practices
  - Diversification of rotation
  - Agroforestry
  - Intercropping
  - Use of decision support tools to reduce inputs
  - No tillage
  - Etc.

Cocoa agroforestry system (Photo E. Malezieux)

DOI: 10.1051/agro:2007057

Why modelling is useful for agro-ecology?

- Agroecology has multiple objectives
- Agroecology promotes a great diversity of practices

➡ Models = powerful tools to compare many practices for many objectives.

Why modelling **at field scale**?

- Many decisions are taken at the field scale
  - ➢ choice of crop species,
  - ➢ choice of cropping systems,
  - ➢ fertilization,
  - ➢ pest & disease control,
  - ➢ irrigation etc.

Why modelling **at field scale**?

- Many decisions are taken at the field scale
  - choice of crop species,
  - choice of cropping systems,
  - fertilization,
  - pest & disease control,
  - irrigation etc.
- A lot of data is available at the field scale to assist in model development
  - trials in experimental stations,
  - on-farm experiments,
  - field observations,
  - expert knowledge (from farmers, advisors etc.)

Main steps in a modelling project

Step 1: Definition of the objective

Step 2: Data collection

Step 3: Definition of candidate models

Step 4: Model training with data (parameter estimation)

Step 5: Model testing with data (model evaluation)

Step 6: Model application

Example of modelling project:
N, P, K fertilization models for potato crops in Eastern Canada

https://doi.org/10.1371/journal.pone.0230888

PLOS ONE

RESEARCH ARTICLE

Site-specific machine learning predictive fertilization models for potato crops in Eastern Canada

Zonlehoua Coulibali[1], Athyna Nancy Cambouris[2], Serge-Étienne Parent[1]*

1 Department of Soils and Agrifood Engineering, Université Laval, Québec City, Quebec, Canada, **2** Quebec Research and Development Centre, Agriculture and Agri-Food Canada, Québec City, Quebec, Canada

Example of modelling project:
N, P, K fertilization models for potato crops in Eastern Canada

https://doi.org/10.1371/journal.pone.0230888

Data and code available!

https://github.com/rgoals/2019_Site-specific-potato-npk-model

Random forest (RF) algorithm

The default parameters

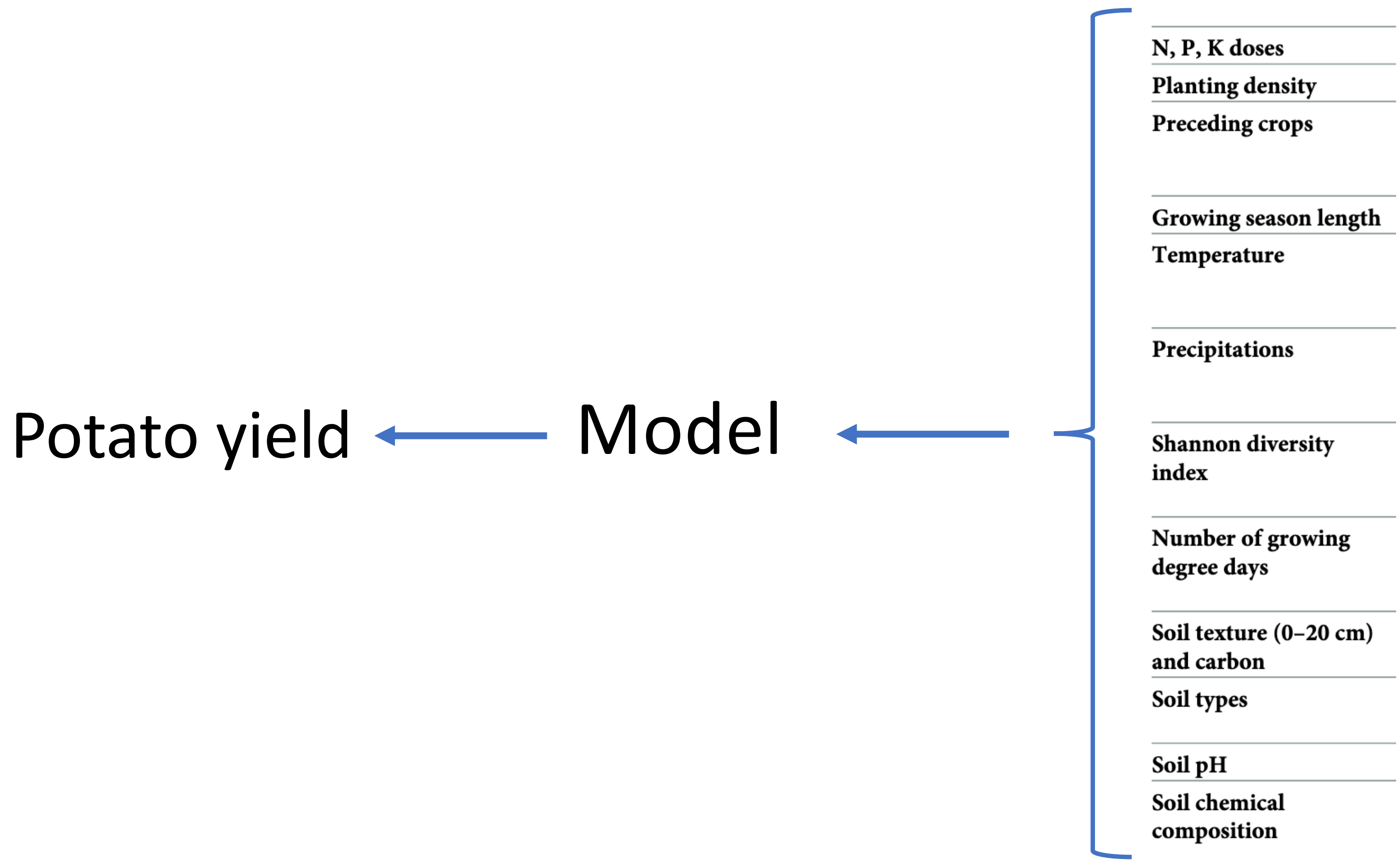
In [43]:

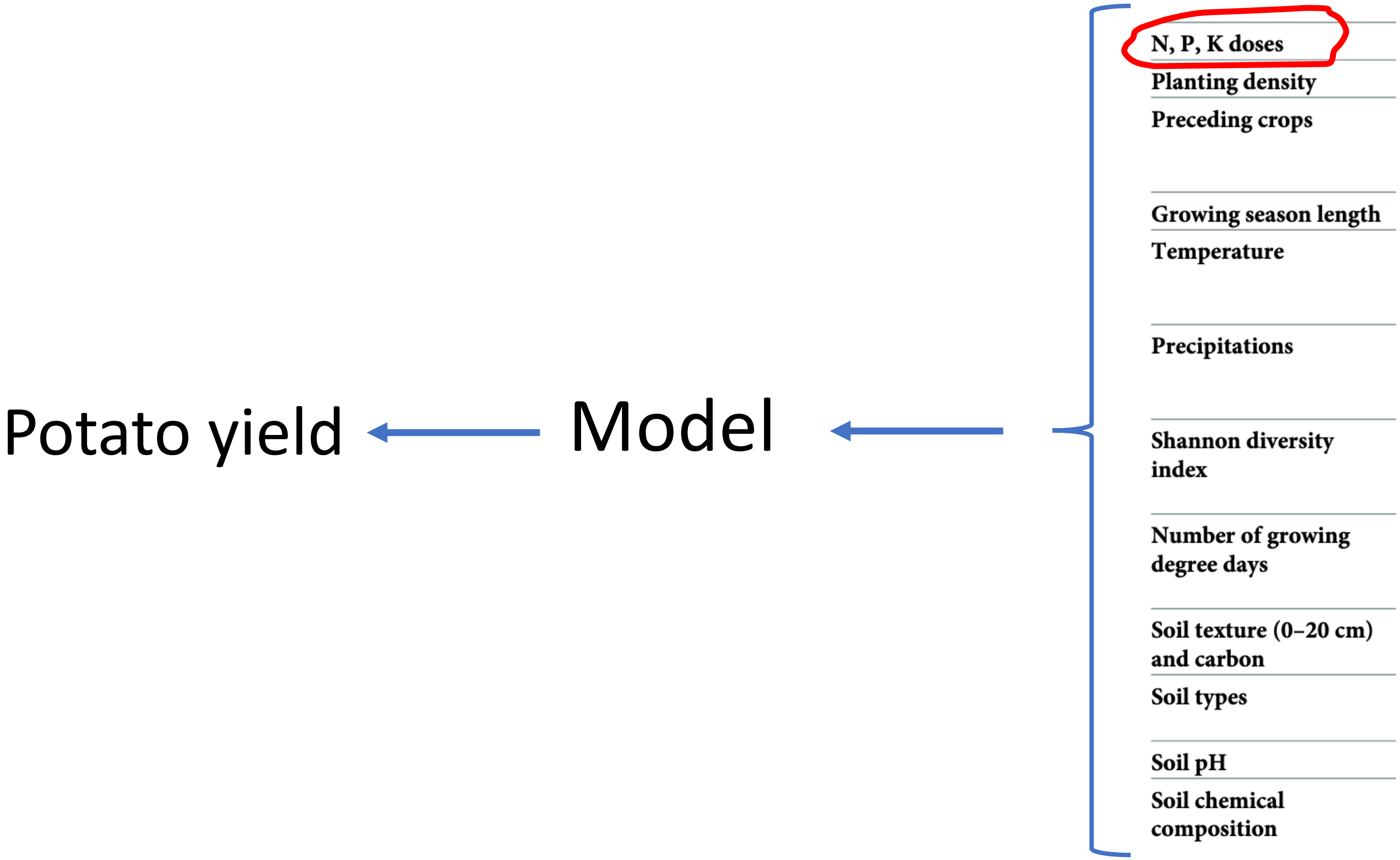
```
from sklearn.ensemble import RandomForestRegressor
```

In [44]:

```
rf_reg = RandomForestRegressor(random_state = 1)
print('Parameters currently in use:\n')
pprint(rf_reg.get_params())
```

```
Parameters currently in use:
```


Potato yield
Model
N, P, K doses
Planting density
Preceding crops
Growing season length
Temperature
Precipitations
Shannon diversity index
Number of growing degree days
Soil texture (0–20 cm) and carbon
Soil types
Soil pH
Soil chemical composition


Potato yield
Model
N, P, K doses
Planting density
Preceding crops
Growing season length
Temperature
Precipitations
Shannon diversity index
Number of growing degree days
Soil texture (0–20 cm) and carbon
Soil types
Soil pH
Soil chemical composition

Main steps in a modelling project

Step 1: Definition of the objective

Step 2: Data collection

Step 3: Definition of candidate models

Step 4: Model training with data (parameter estimation)

Step 5: Model testing with data (model evaluation)

Step 6: Model application

Main steps in a modelling project

Step 1: Definition of the objective

Develop models to predict yields and calculate optimal N, P, K fertilizer doses for potato crops in Eastern Canada

Step 2: Data collection

Step 3: Definition of candidate models

Step 4: Model training with data (parameter estimation)

Step 5: Model testing with data (model evaluation)

Step 6: Model application

Main steps in a modelling project

Step 1: Definition of the objective

Step 2: Data collection

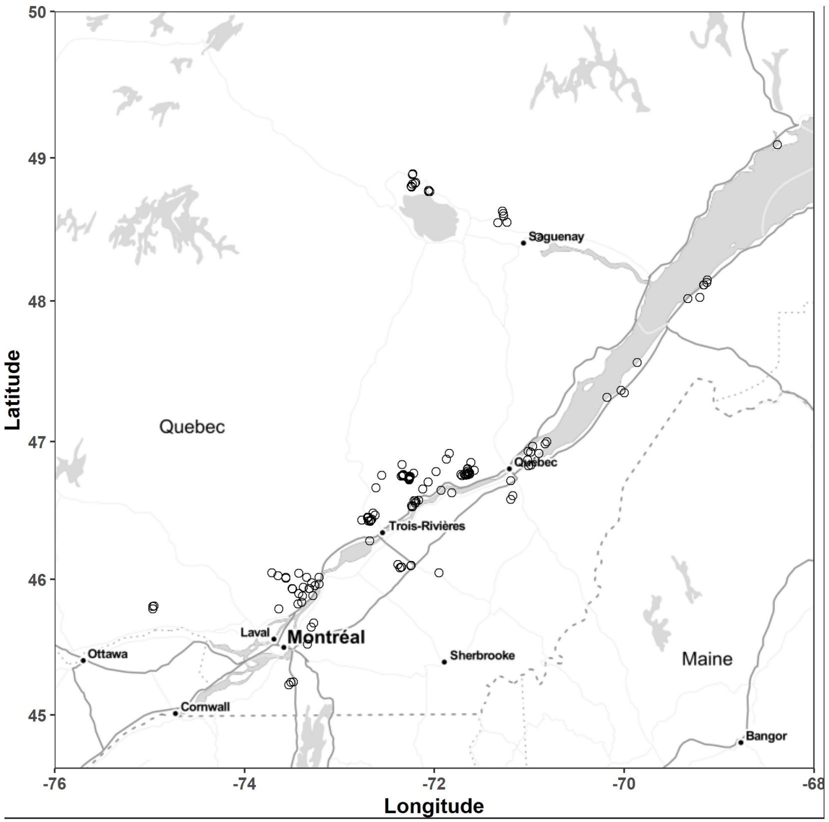
Step 3: Definition of candidate models

Step 4: Model training with data (parameter estimation)

Step 5: Model testing with data (model evaluation)

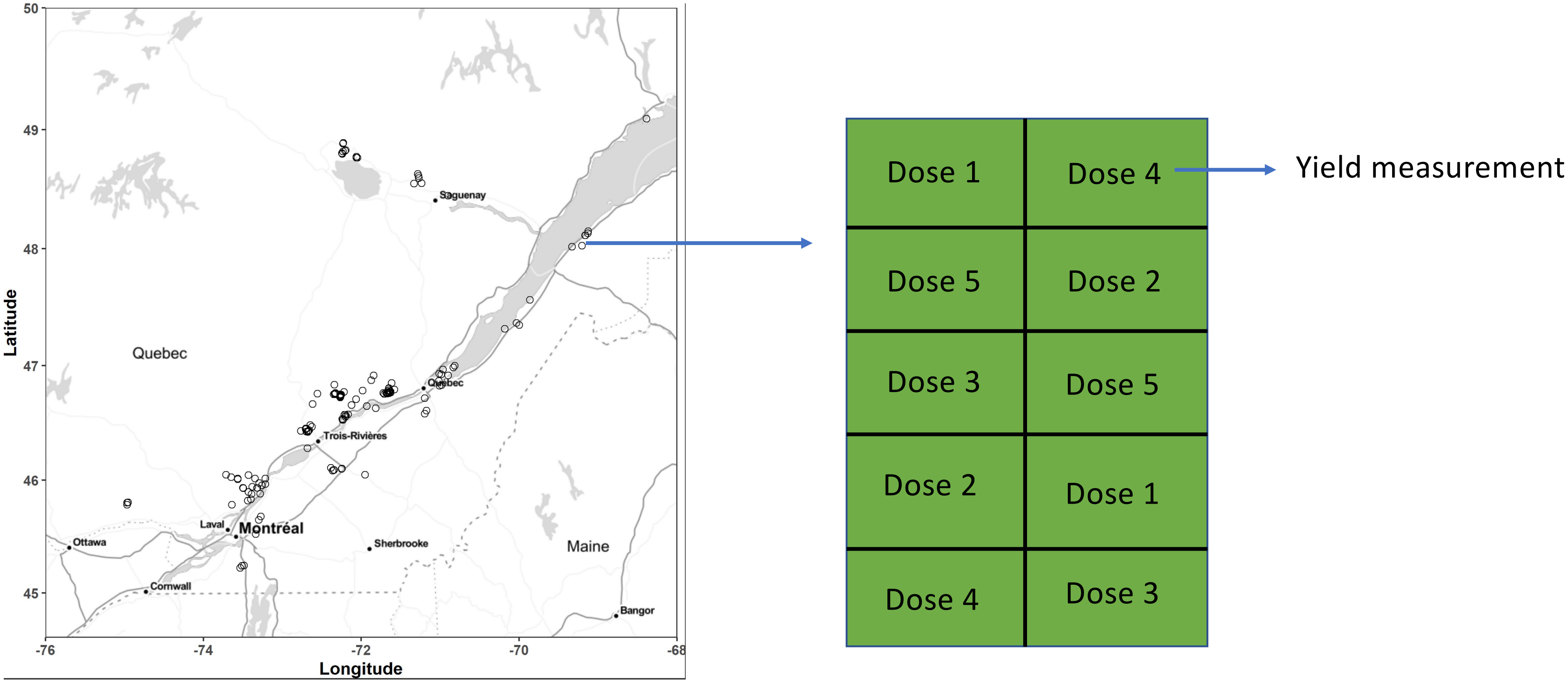
Step 6: Model application

237 field trials



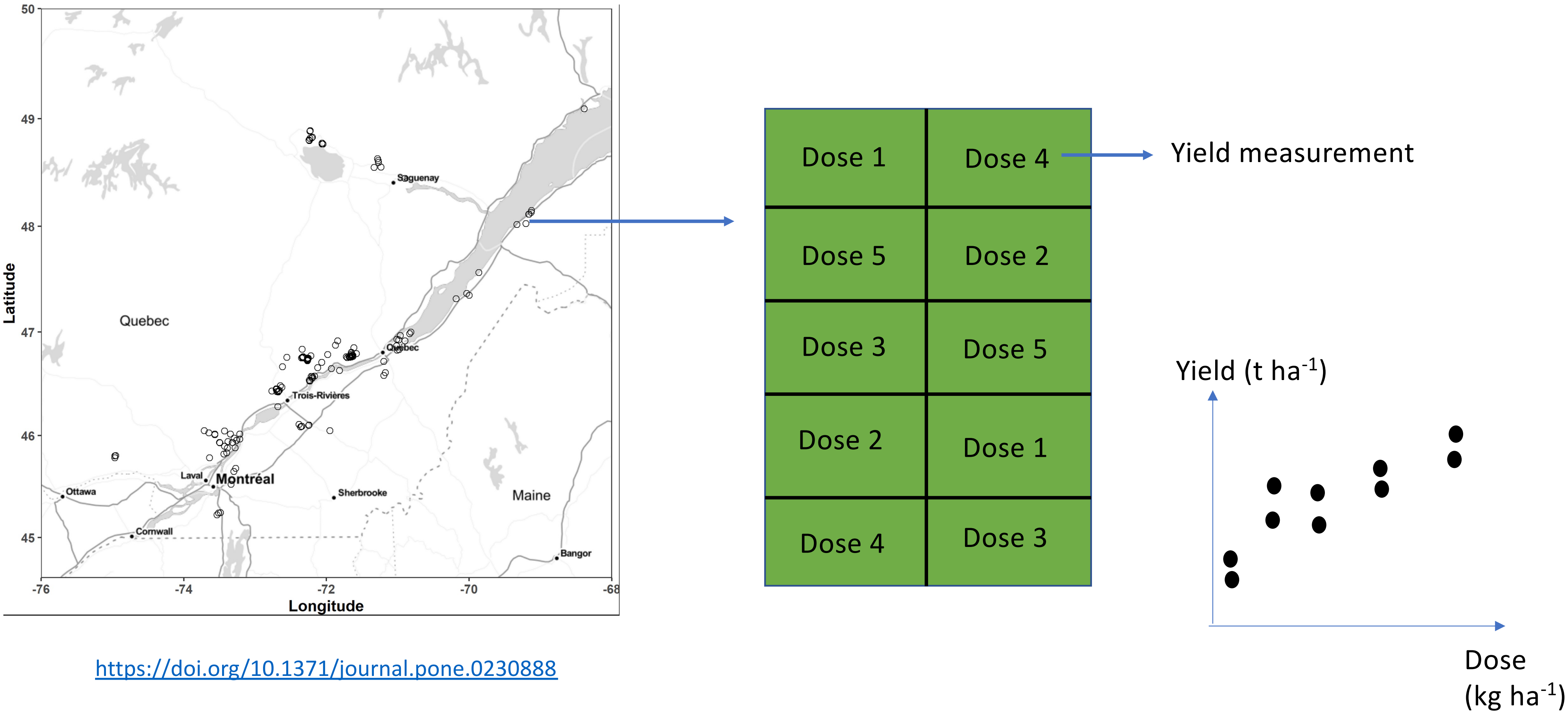
https://doi.org/10.1371/journal.pone.0230888

237 field trials



https://doi.org/10.1371/journal.pone.0230888

237 field trials



https://doi.org/10.1371/journal.pone.0230888

Main steps in a modelling project

Step 1: Definition of the objective

Step 2: Data collection

Step 3: Definition of candidate models

Step 4: Model training with data (parameter

Step 5: Model testing with data (model evalu

Step 6: Model application

Five models

1. Mitscherlich
2. KNN
3. Random forest
4. Neural network
5. Gaussian process

Main steps in a modelling project

Step 1: Definition of the objective

Step 2: Data collection

Step 3: Definition of candidate models

Step 4: Model training with data (parameter

Step 5: Model testing with data (model evalu

Step 6: Model application

Five models

1. Mitscherlich
2. KNN
3. Random forest
4. Neural network
5. Gaussian process

$$Y = A\,x(1 - e^{-R_N \mathrm{x}(E_N + dose_N)})\mathrm{x}(1 - e^{-R_P \mathrm{x}(E_P + dose_P)})\mathrm{x}(1 - e^{-R_K \mathrm{x}(E_K + dose_K)})$$

Main steps in a modelling project

Step 1: Definition of the objective

Step 2: Data collection

Step 3: Definition of candidate models

Step 4: Model training with data (parameter (

Step 5: Model testing with data (model evalu

Step 6: Model application

Five models

1. Mitscherlich
2. KNN
3. Random forest
4. Neural network
5. Gaussian process

Standard machine learning models

Main steps in a modelling project

Step 1: Definition of the objective

Step 2: Data collection

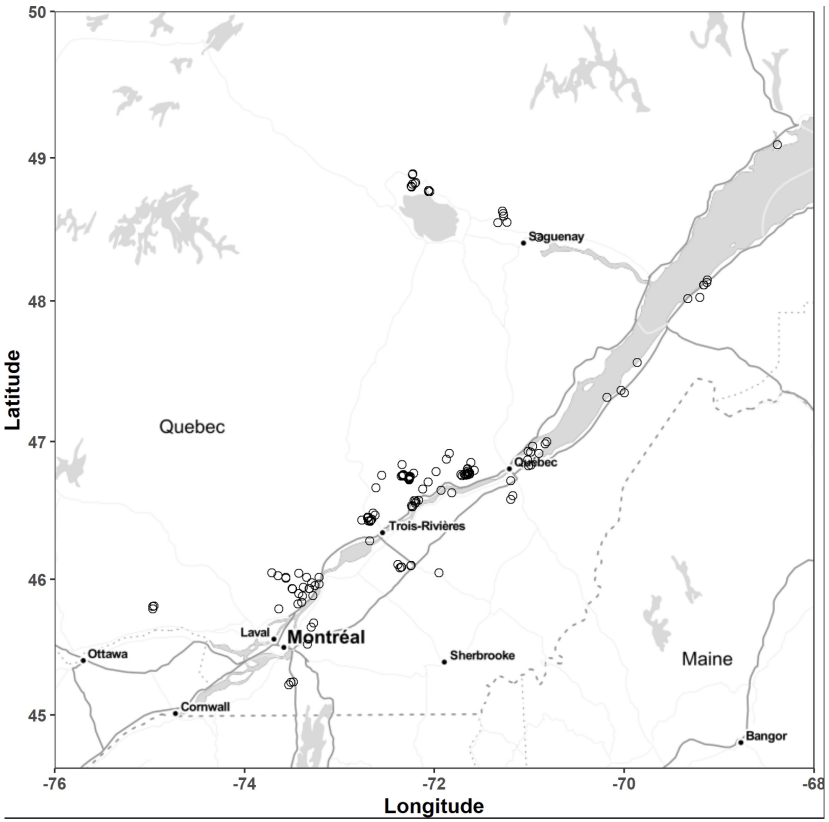
Step 3: Definition of candidate models

Step 4: Model training with data (parameter estimation)

Step 5: Model testing with data (model evaluation)

Step 6: Model application

237 field trials



https://doi.org/10.1371/journal.pone.0230888

Training dataset
60% of the trials

Parameter estimation
for the five models

Main steps in a modelling project

Step 1: Definition of the objective

Step 2: Data collection

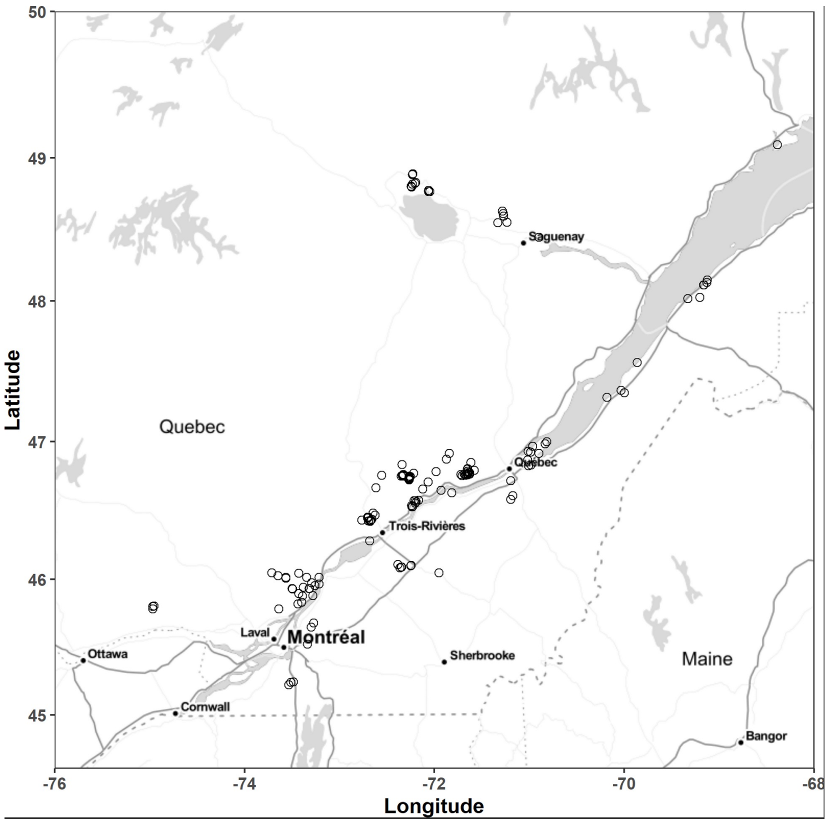
Step 3: Definition of candidate models

Step 4: Model training with data (parameter estimation)

Step 5: Model testing with data (model evaluation)

Step 6: Model application

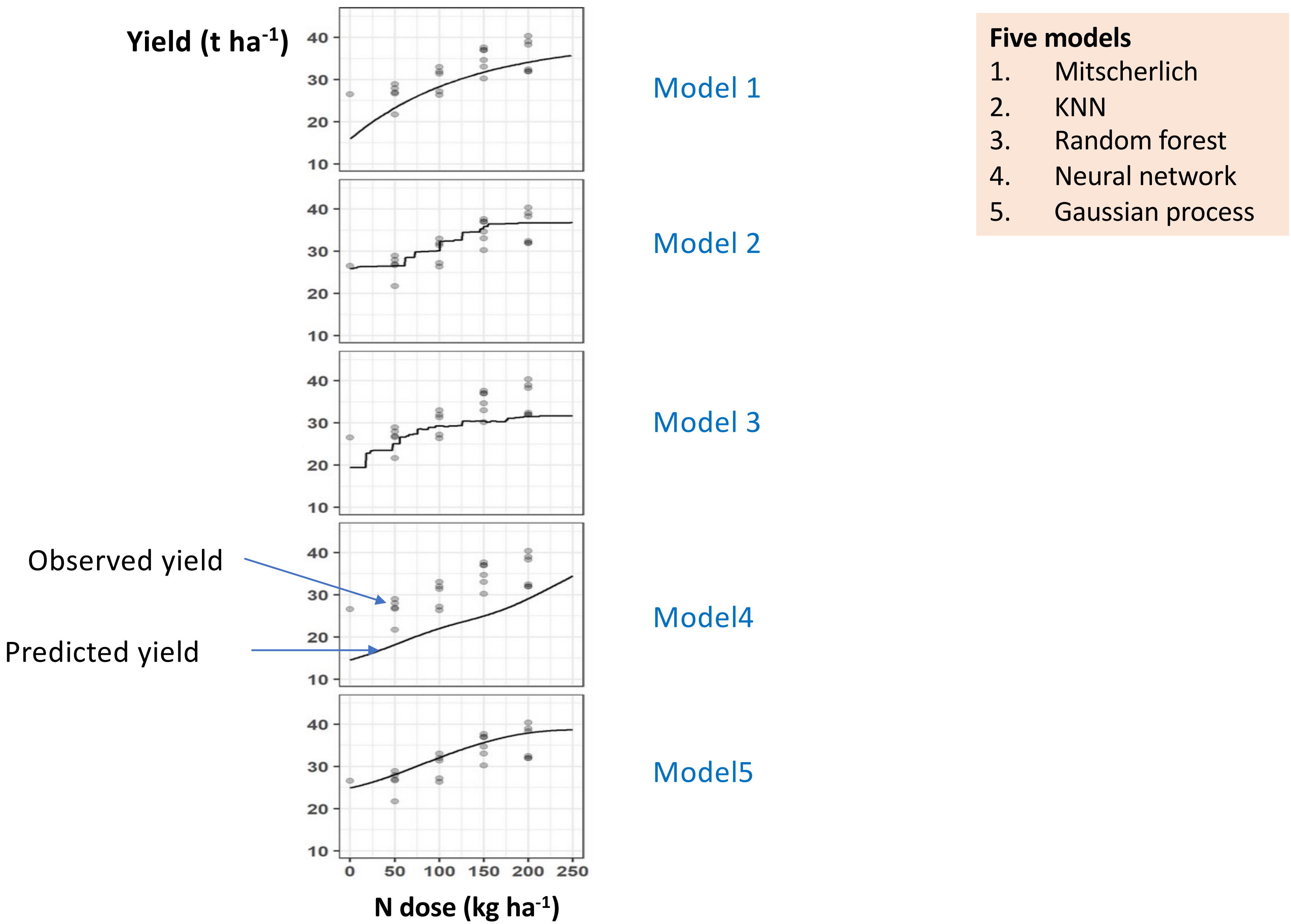
237 field trials



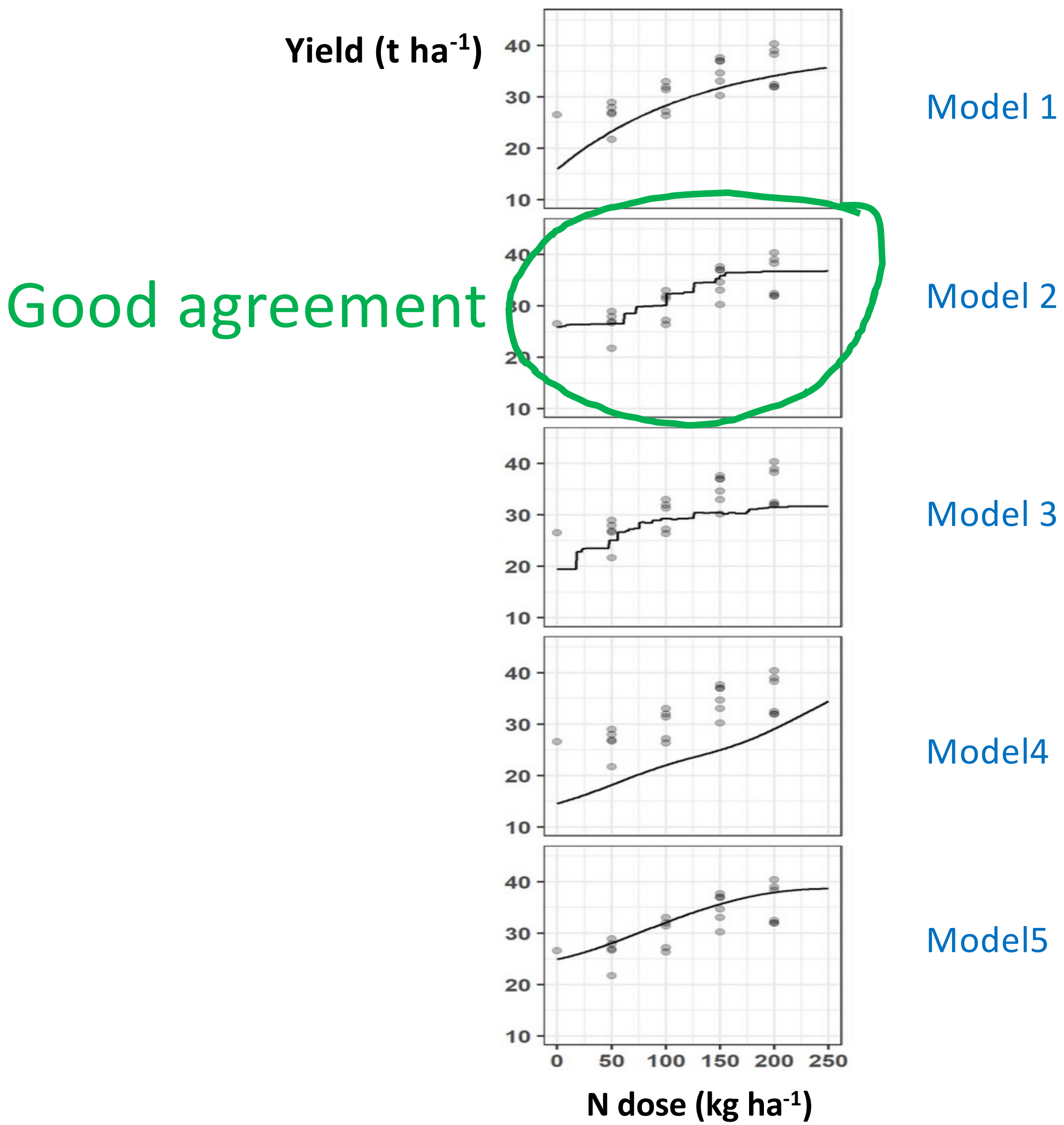
https://doi.org/10.1371/journal.pone.0230888

Testing dataset
40% of the trials

Evaluation of the
model performances

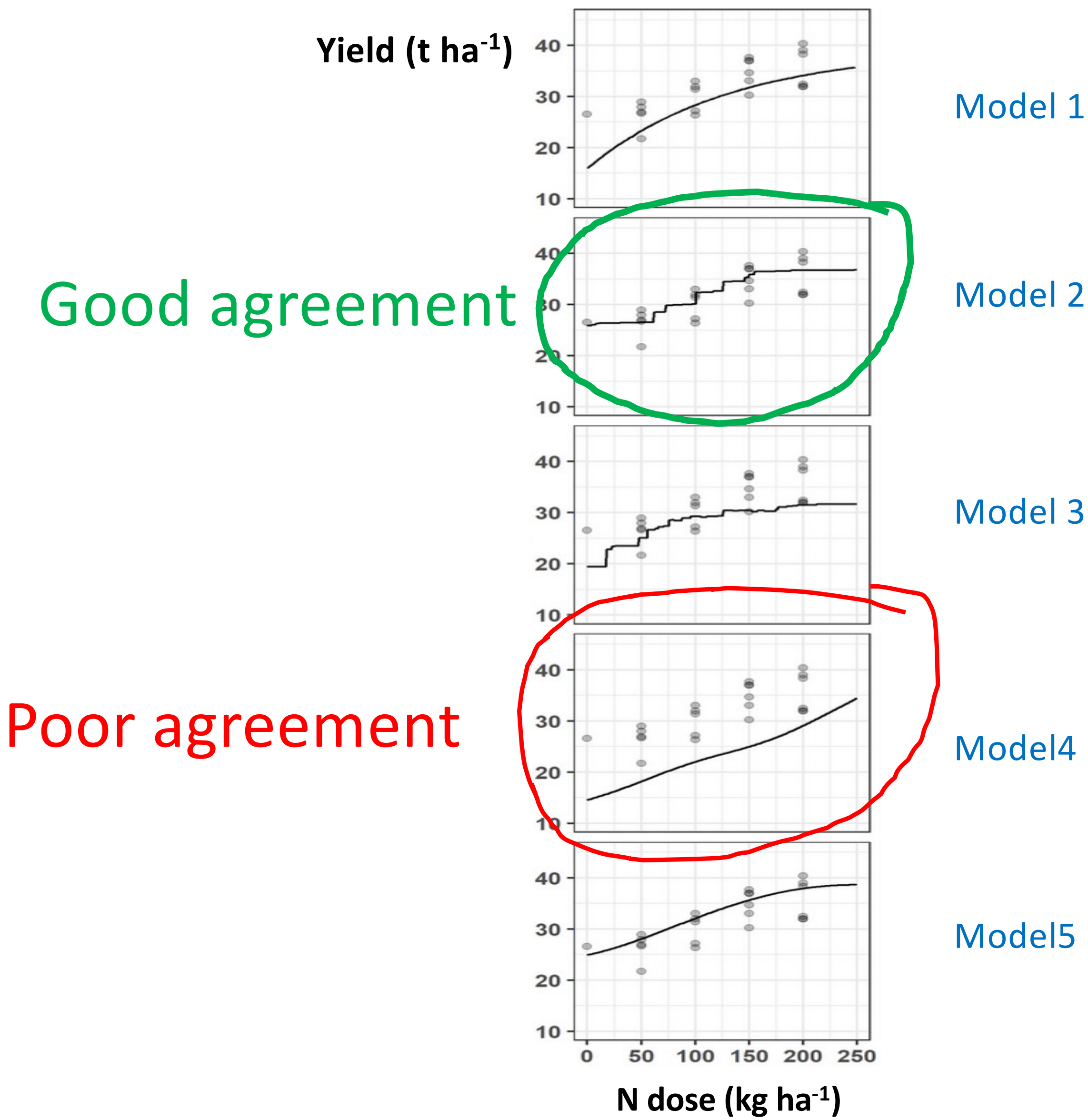


Yield (t ha-1)
40
30
20
10
Model 1
Model 2
Model 3
Model4
Model5
Observed yield
Predicted yield
0
50
100
150
200
250
N dose (kg ha-1)
Five models
1. Mitscherlich
2. KNN
3. Random forest
4. Neural network
5. Gaussian process



Five models

1. Mitscherlich
2. KNN
3. Random forest
4. Neural network
5. Gaussian process

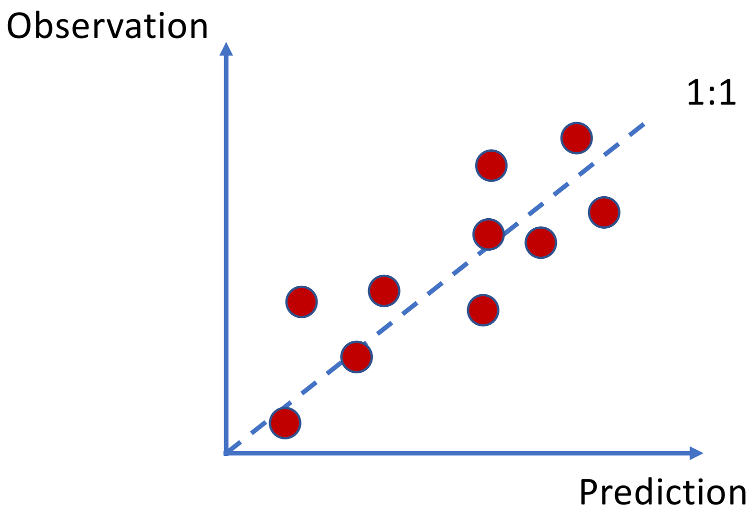


Five models

1. Mitscherlich
2. KNN
3. Random forest
4. Neural network
5. Gaussian process

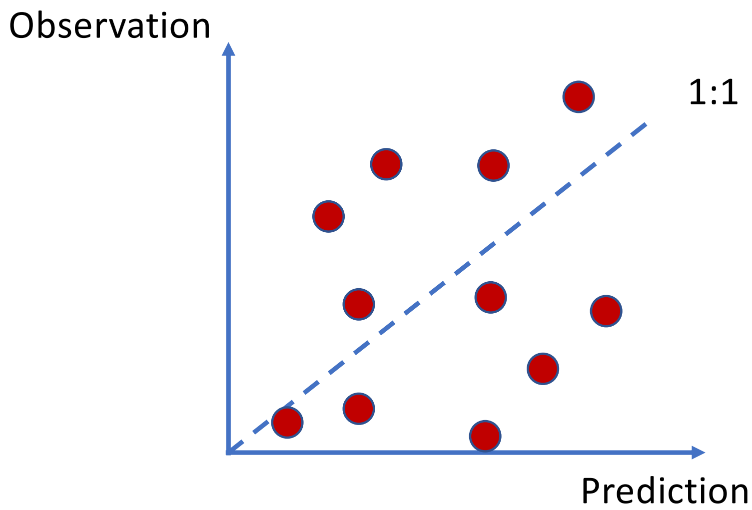
R^2 is a popular evaluation criterion

Good agreement



R^2 close to 1

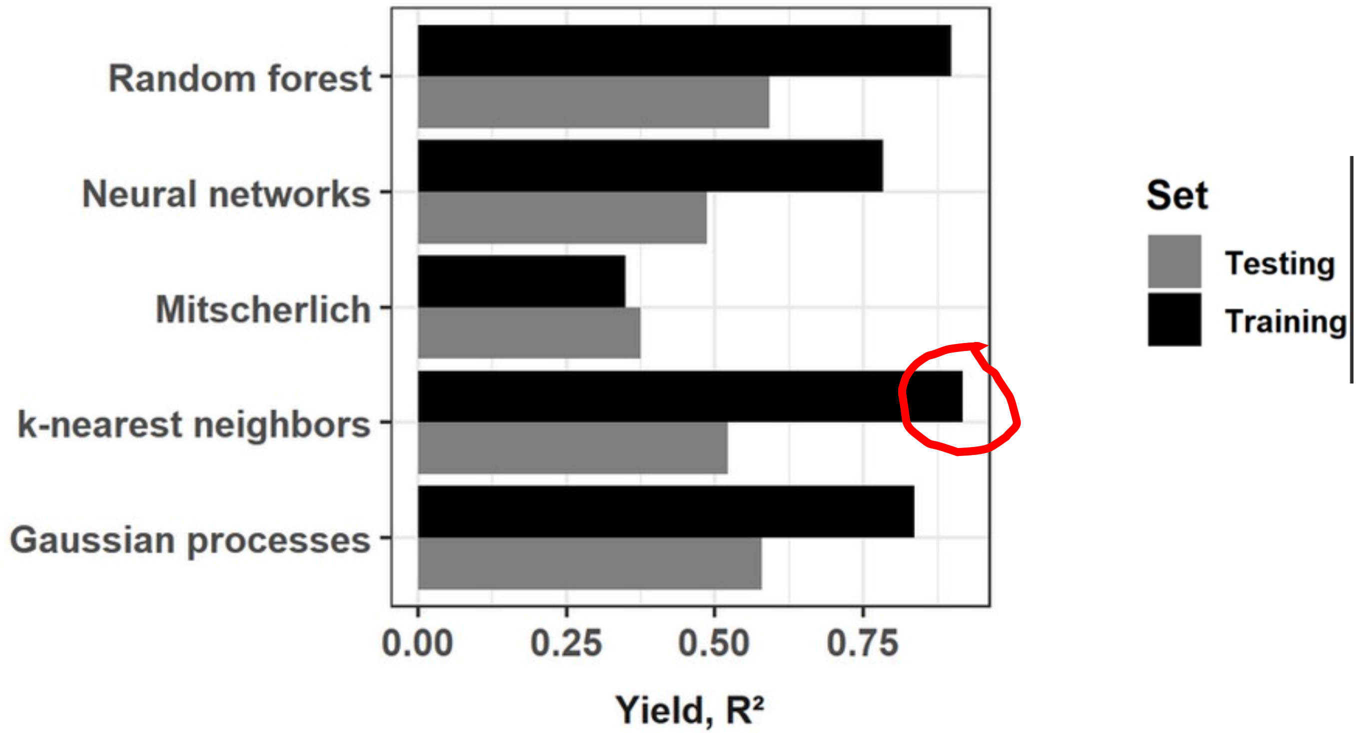
Poor agreement



R^2 close to 0

R^2

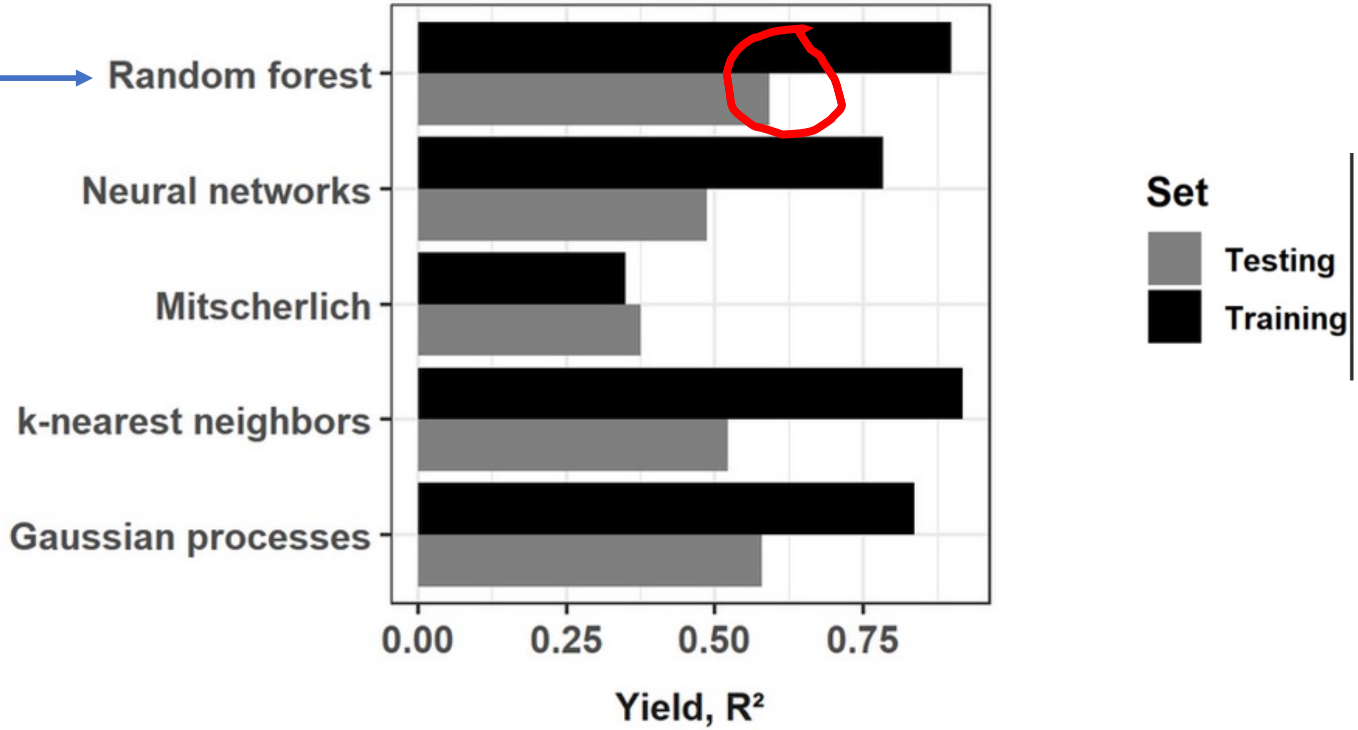
Best model according to the training dataset



https://doi.org/10.1371/journal.pone.0230888

R^2

Best model according to the test dataset

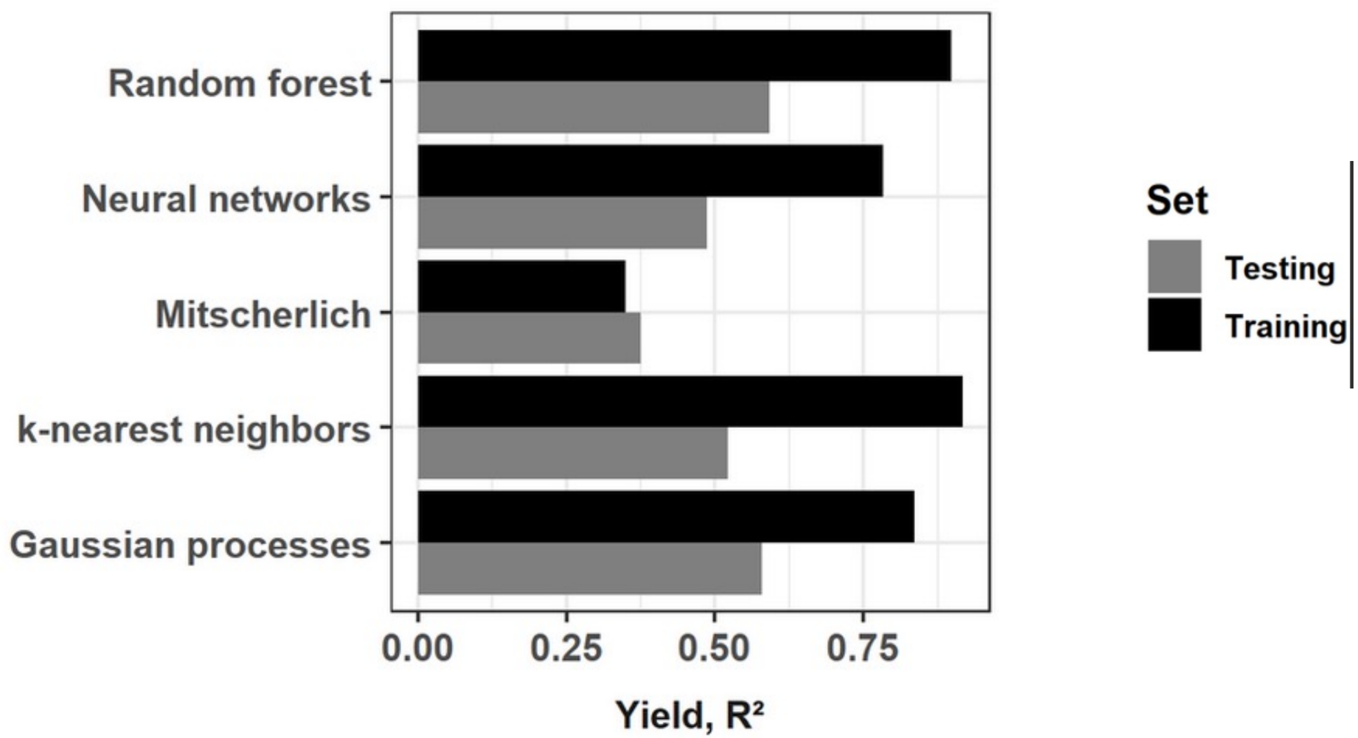


https://doi.org/10.1371/journal.pone.0230888

R^2

Model performances are too optimistic according to the training dataset.

Important to use an independent test dataset !



https://doi.org/10.1371/journal.pone.0230888

Step 1: Definition of the objective

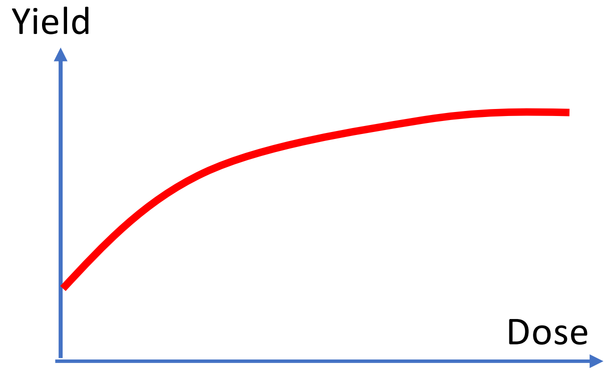
Step 2: Data collection

Step 3: Definition of candidate models

Step 4: Model training with data (parameter estimation)

Step 5: Model testing with data (model evaluation)

Step 6: Model application



Step 1: Definition of the objective

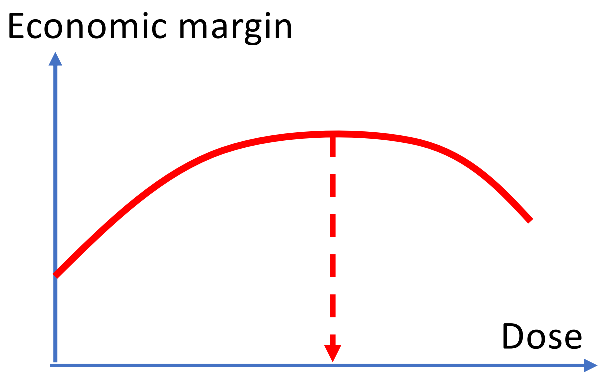
Step 2: Data collection

Step 3: Definition of candidate models

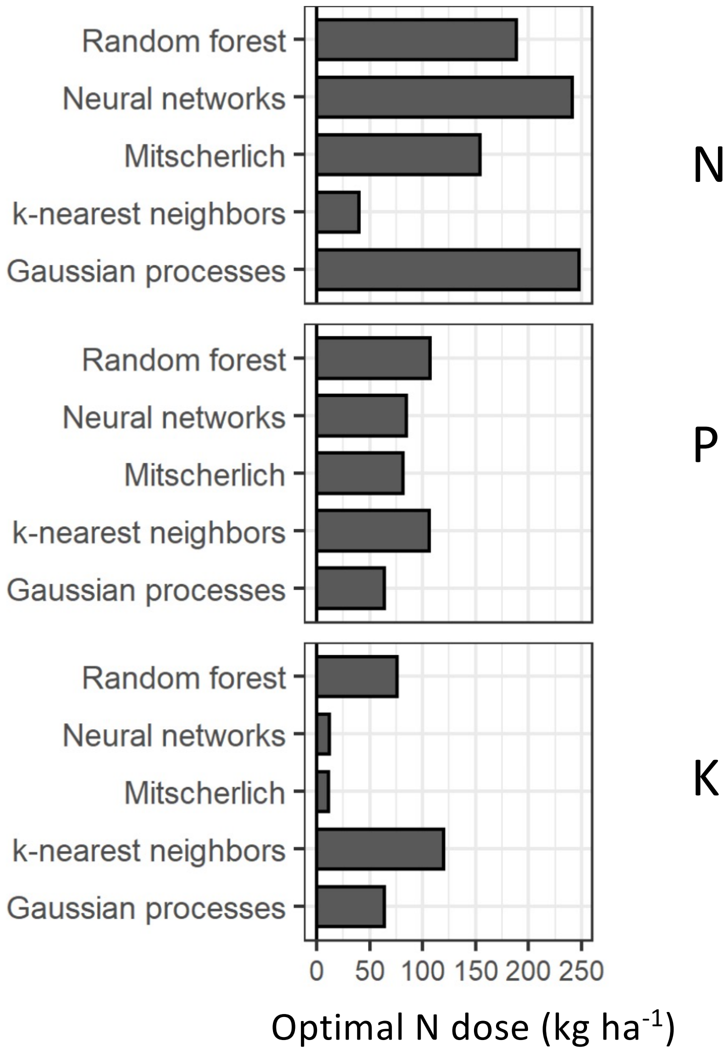
Step 4: Model training with data (parameter estimation)

Step 5: Model testing with data (model evaluation)

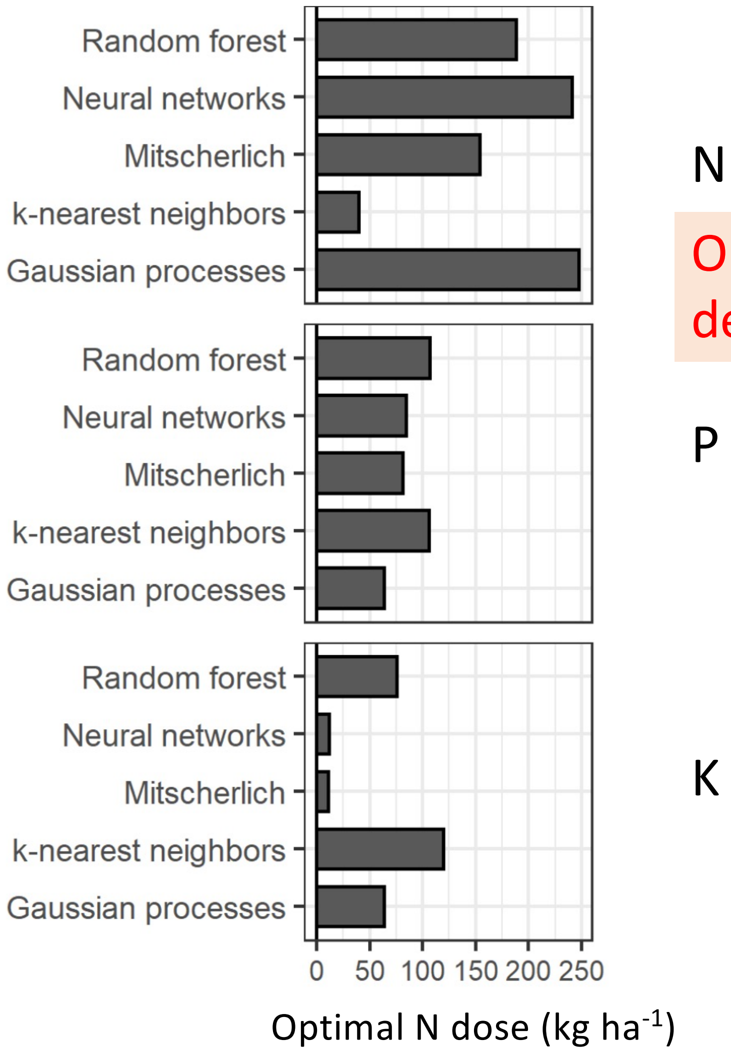
Step 6: Model application



Examples of optimal economic fertilizer doses at one site in Canada



Examples of optimal economic fertilizer doses at one site in Canada



Optimal doses are highly dependent on the selected model!

Why modelling is powerful?

- Models aggretate different sources of knowledge
- Useful for assessing & optimising farming practices
- Useful in foresight studies (ex: climate change)
- Support decisions of farmers, companies, policy makers

Why modelling can be dangerous?

- All models are wrong!
- Some models are not correctly evaluated
- Some models can be lead to poor predictions and decisions

Why modelling can be dangerous?

- All models are wrong!
- Some models are not correctly evaluated
- Some models can be lead to poor predictions and decisions
- Complex models can be fascinating… but can be highly uncertain.

➡ **Important to assess models carefully!**

Limit of modelling at field scale

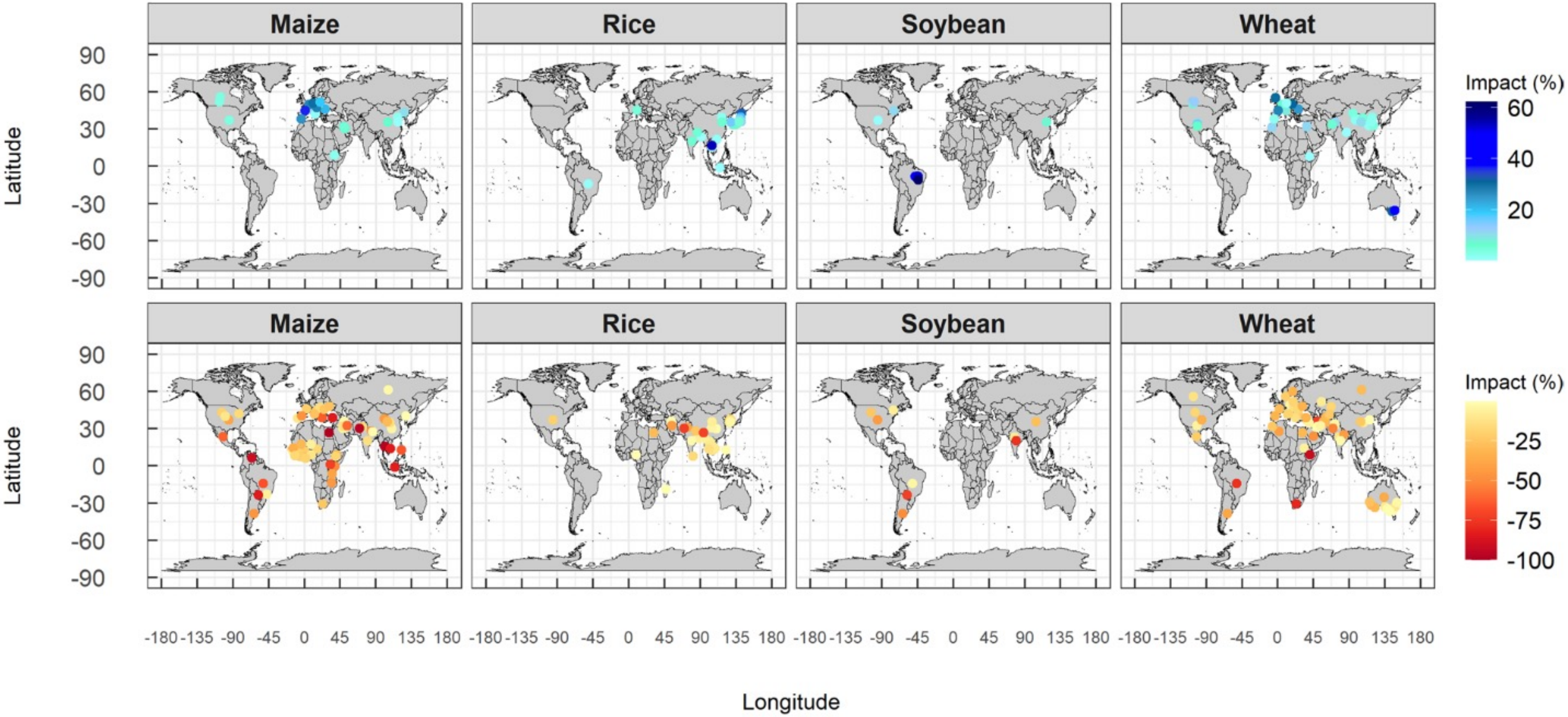
- Not the best scale to evaluate all economic and social impacts
- Not always the best scale to address global issues

BUT

Field-scale models can be upscaled at a larger scale.

Simulated yield variations due to climate change

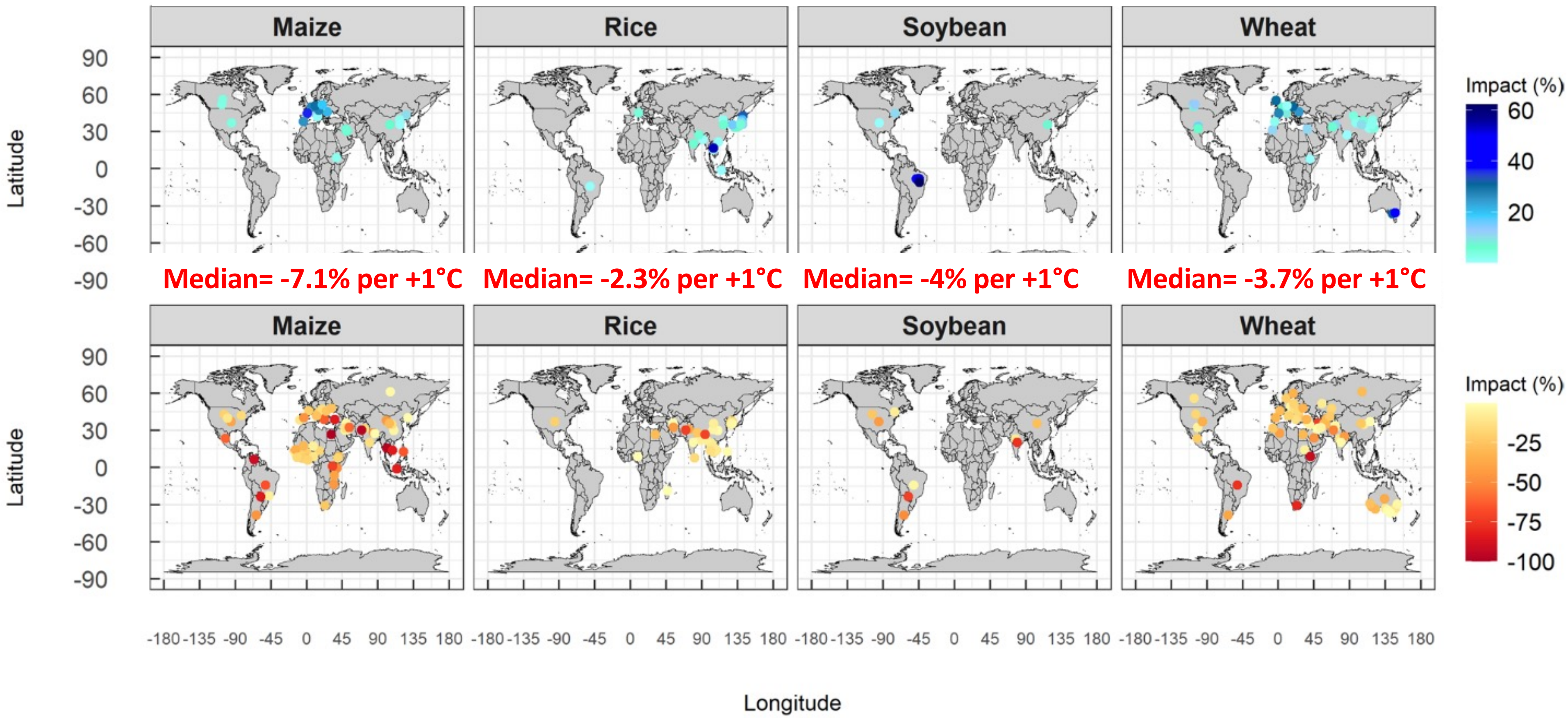
for scenario « RCP8.5 », time horizon 2050, baseline 2001-2010



https://doi.org/10.1038/s41597-022-01150-7

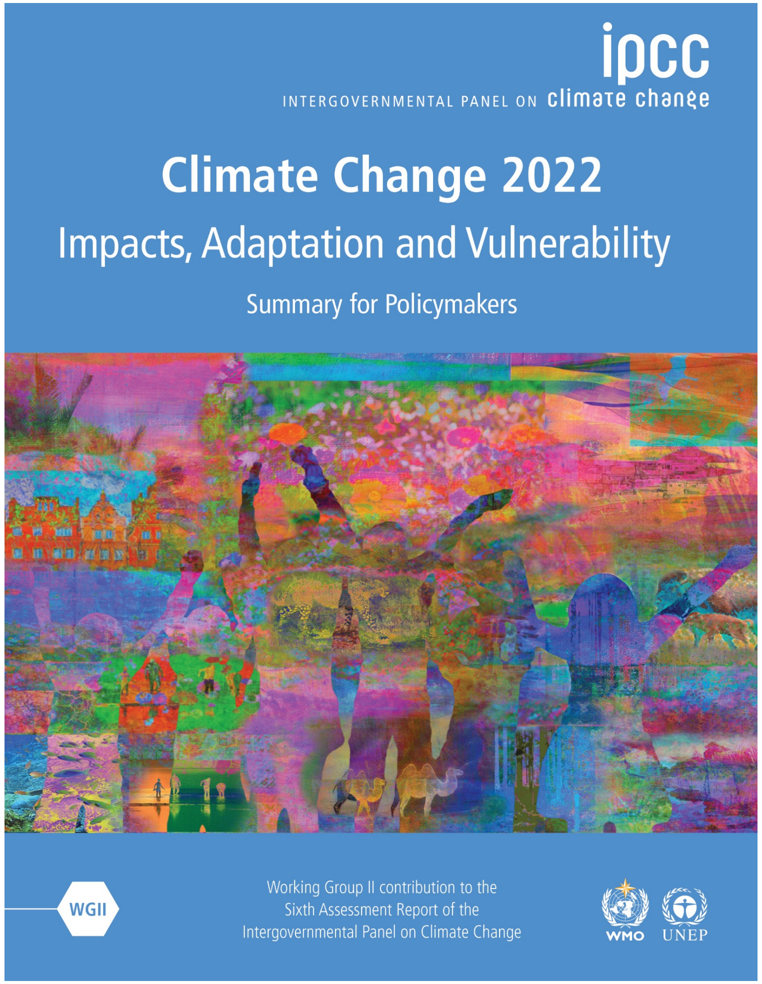
Simulated yield variations due to climate change

for scenario « RCP8.5 », time horizon 2050, baseline 2001-2010



https://doi.org/10.1038/s41597-022-01150-7

Crop model simulations are used in the IPCC reports on climate change



What you need to learn to be a good modeller

- Principles of agrosystem functioning

What you need to learn to be a good modeller

- Principles of agrosystem functioning
- Potential sources of data (data on weather, soil, management, yields etc.)

What you need to learn to be a good modeller

- Principles of agrosystem functioning
- Potential sources of data (data on weather, soil, management, yields etc.)
- Key modelling techniques

What you need to learn to be a good modeller

- Principles of agrosystem functioning
- Potential sources of data (data on weather, soil, management, yields etc.)
- Key modelling techniques
    - ➢Statistics
    - ➢Machine learning
    - ➢Sensitivity/Uncertainty analysis
    - ➢Data assimilation

What you need to learn to be a good modeller

- Principles of agrosystem functioning
- Potential sources of data (data on weather, soil, management, yields etc.)
- Key modelling techniques
- Coding

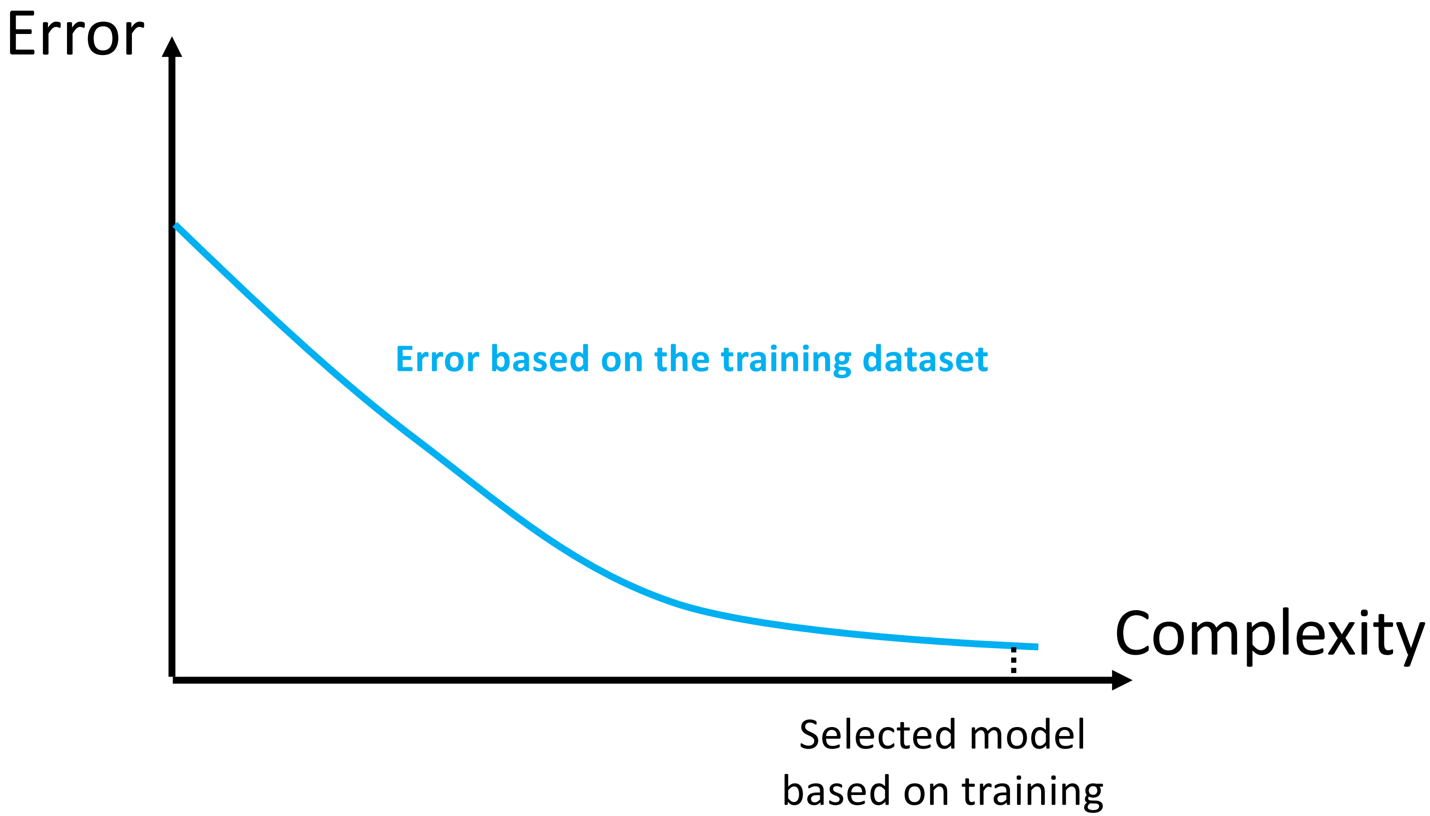
What you need to learn to be a good modeller

- Principles of agrosystem functioning
- Potential sources of data (data on weather, soil, management, yields etc.)
- Key modelling techniques
- Coding
- Team work

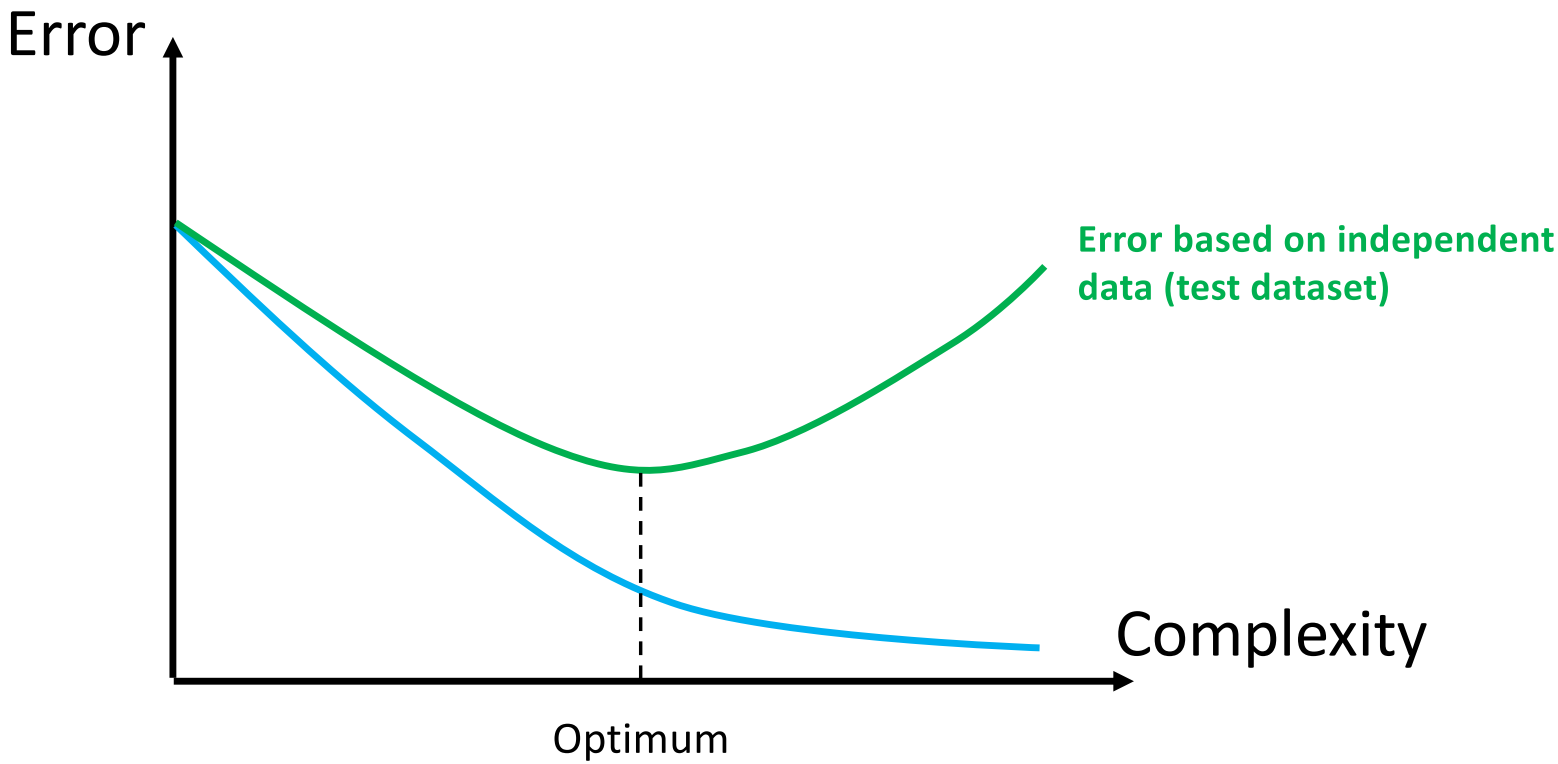
Check my github site for more lectures, data, and codes!

https://github.com/davemakowski?tab=repositories

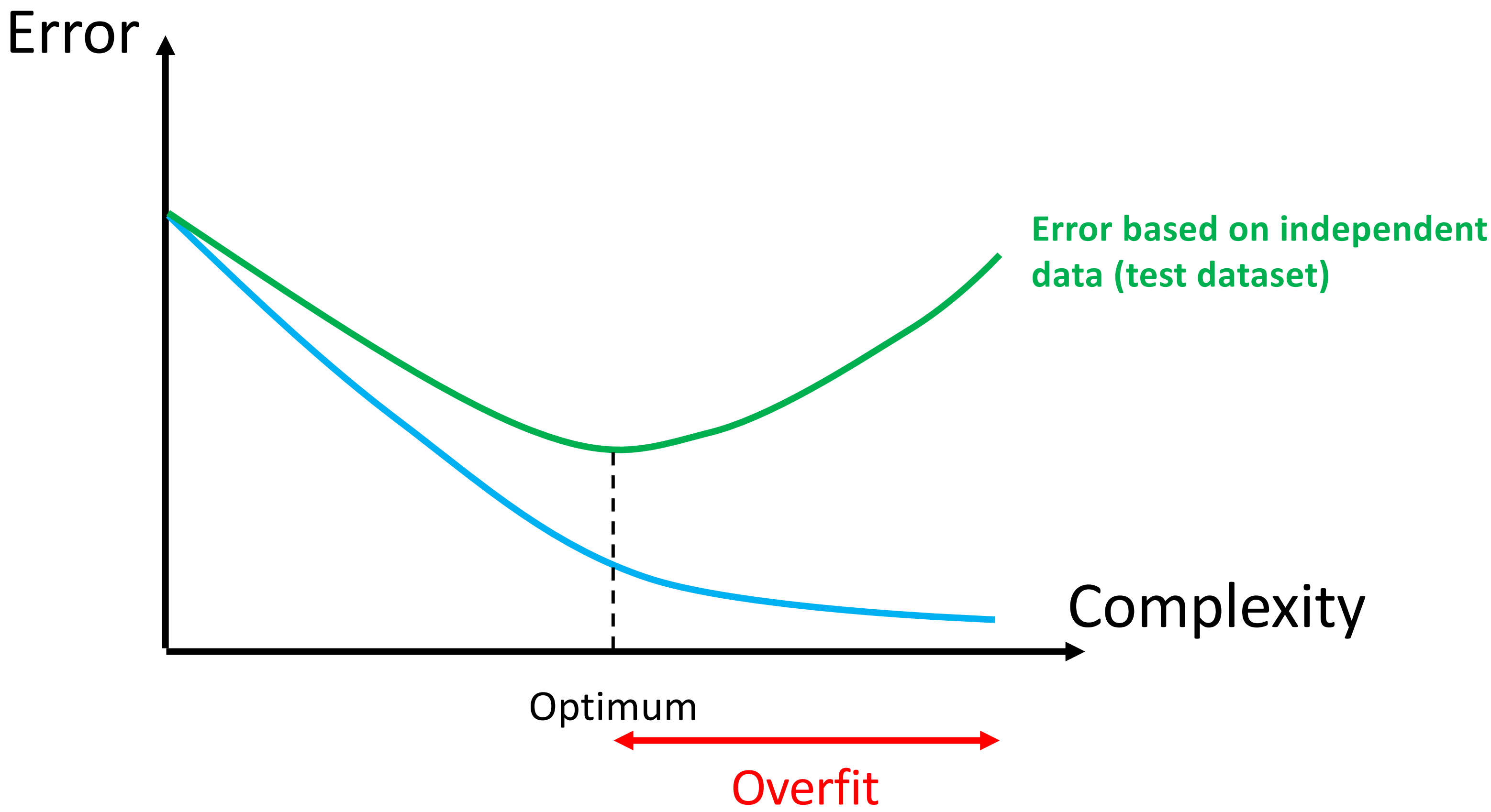
Model testing should be taken seriously to avoid risk of overfitting



Model testing should be taken seriously to avoid risk of overfitting



Model testing should be taken seriously to avoid risk of overfitting



Model testing should be taken seriously to avoid risk of overfitting