June 28 2022

Agro-ecological modeling for sustainable crop production at field scale

David Makowski

What is a (mathematical) model?

Mathematical model = Tool to compute outputs Y from inputs X

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

What is a (mathematical) model?

Mathematical model = Tool to compute outputs Y from inputs X

Wheat yield ⟵ Model ⟵ N fertilizer dose

What is a (mathematical) model?

Mathematical model = Tool to compute outputs Y from inputs X

8 t ha-1 ⟵ Model ⟵ 200 kg ha-1

What is a (mathematical) model?

Mathematical model = Tool to compute outputs Y from inputs X

6 t ha-1 ⟵ Model ⟵ 100 kg ha-1

What is a (mathematical) model?

Mathematical model = Tool to compute outputs Y from inputs X

6.1 t ha-1 ← Model ← { 100 kg ha-1, Tillage }

What is a (mathematical) model?

Mathematical model = Tool to compute outputs Y from inputs X

5.8 t ha-1 ← Model ← 100 kg ha-1, No tillage

What is a (mathematical) model?

Mathematical model = Tool to compute outputs Y from inputs X

5.8 t ha^{-1} grain
10.5 kg ha^{-1} $N_2 0$ ← Model ← 100 kg ha^{-1}
No tillage

What is a (mathematical) model?

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

$$\begin{tikzcd}
Y & \text{Model} \arrow[l] & X \arrow[l] \\
 & & \theta \arrow[ul]
\end{tikzcd}$$

What is a (mathematical) model?

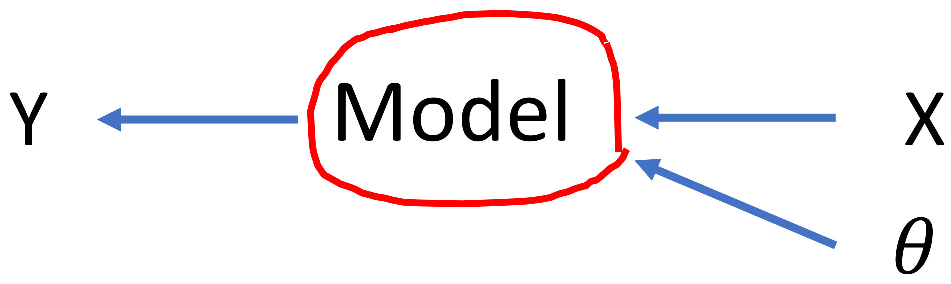
Mathematical model = Tool to compute outputs Y from inputs X
and parameters θ

$$
\begin{tikzcd}
Y & \text{Model} \arrow[l] & X \arrow[l] \\
 & & \theta \arrow[ul]
\end{tikzcd}
$$

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

What is a (mathematical) 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

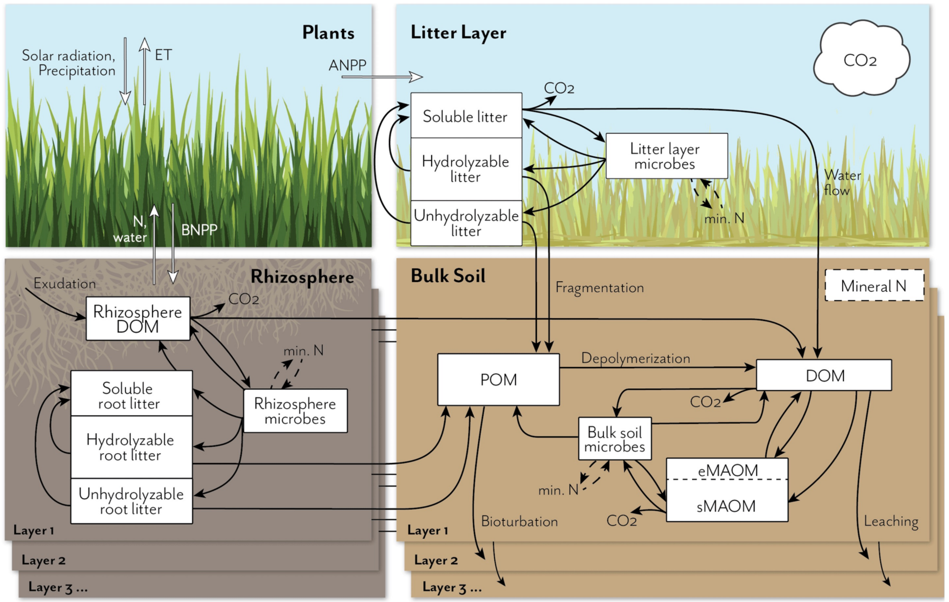
Models can be very simple or… very complex

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** 	~ 100

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

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

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

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

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

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

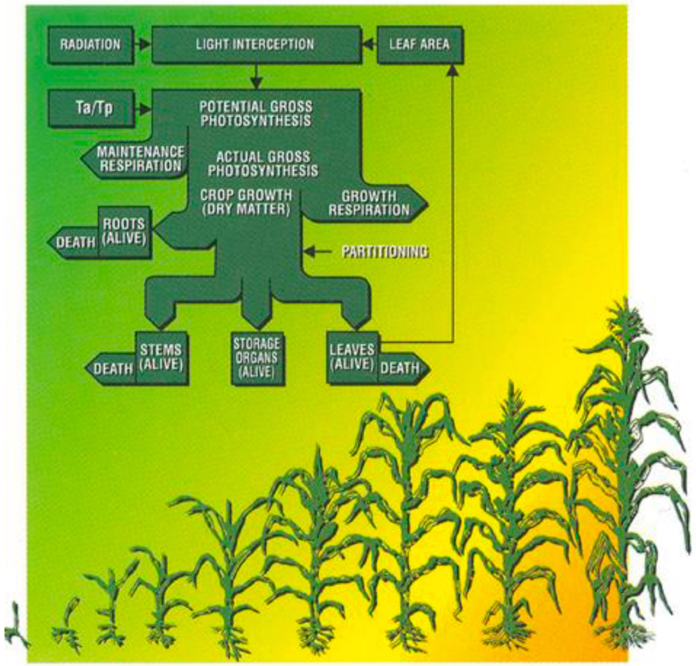
Note: the decay rate of surface soluble litter $k_{soluble}$ is also used for RDOM.

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

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

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

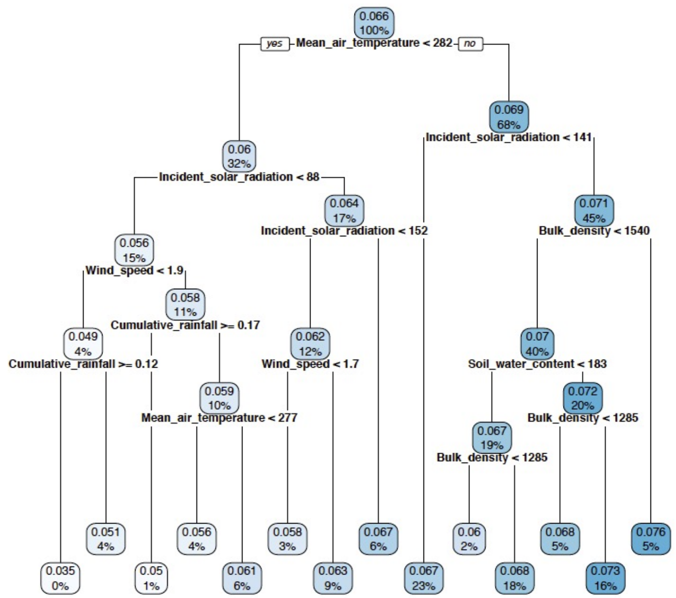
Models can be very simple or… very complex

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

Models can be very simple or... very complex

Model type | **Example** | **Complexity**

Machine learning

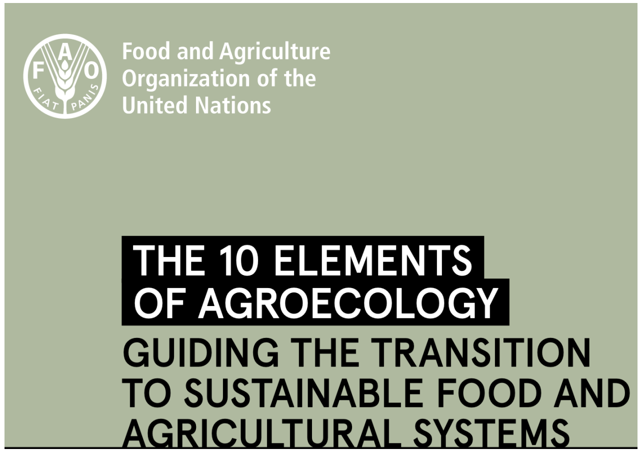


100, 1000 or more

Why modelling is useful for agro-ecology?

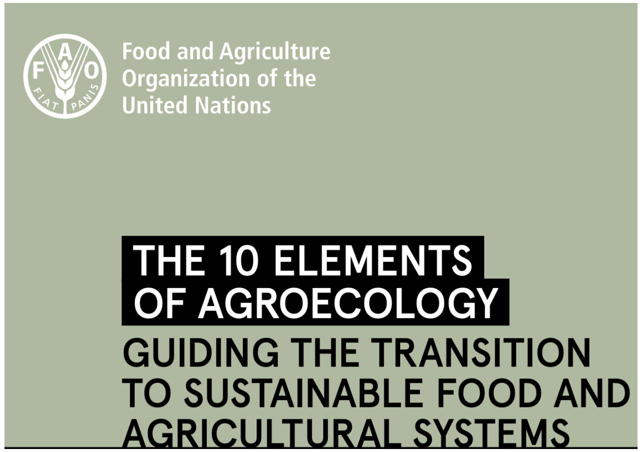
Why modelling is useful for agro-ecology?

« Agroecology is an integrated approach that simultaneously applies ecological and social concepts and principles to the design and management of 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

➡ Need to evaluate many practices for environmental, agronomic, economic and societal objectives.

Why modelling is useful for agro-ecology **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 is useful for agro-ecology **at field scale**?

- Many decisions are taken at the field scale
  - choice of crop species,
  - 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.)

Example 1: N, P, K, fertilization models for potato crops in Eastern Canada

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

Example 1: N, P, K, fertilization models for potato crops in Eastern Canada

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

Step 1: Definition of the objective

Develop models to 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

Example 1: N, P, K, fertilization models for potato crops in Eastern Canada

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)

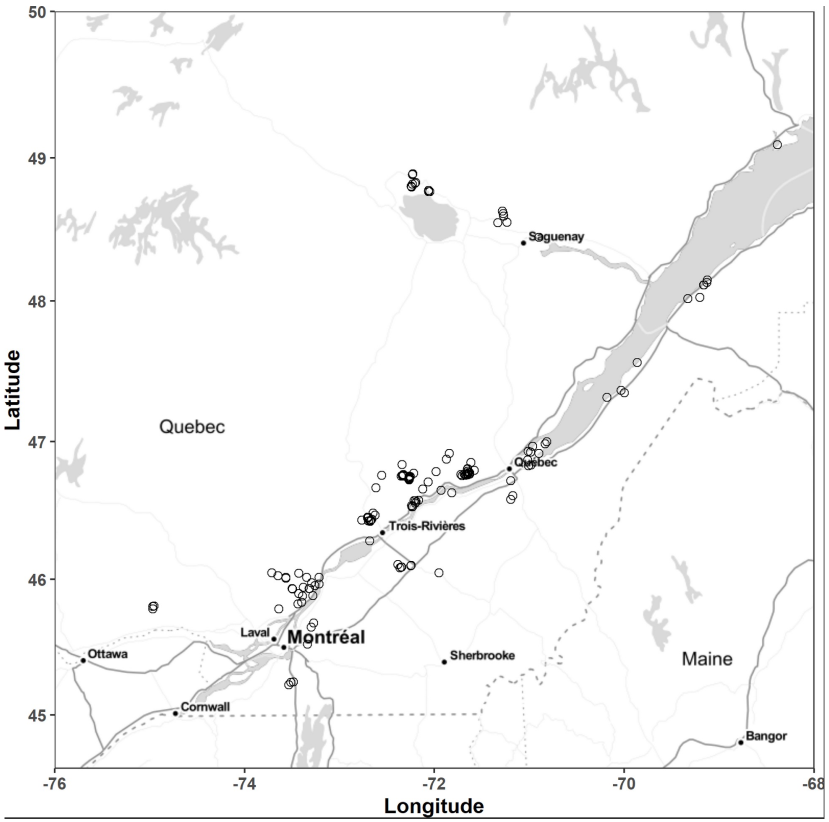
Step 5: Model testing with data (model evaluation)

Step 6: Model application

Example 1: N, P, K, fertilization models for potato crops in Eastern Canada

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

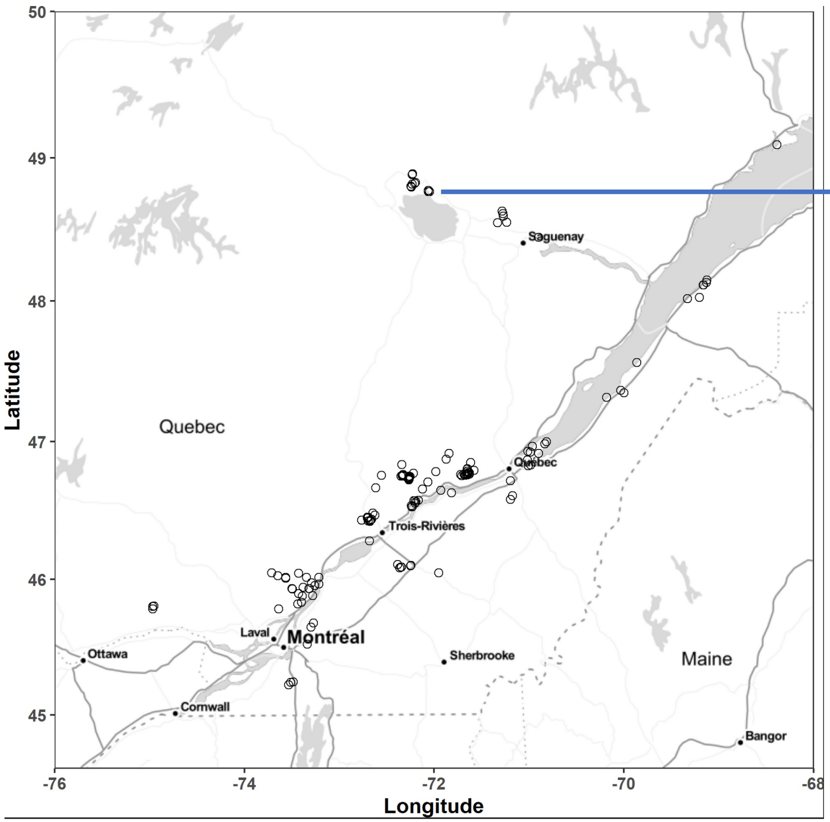
237 field trials

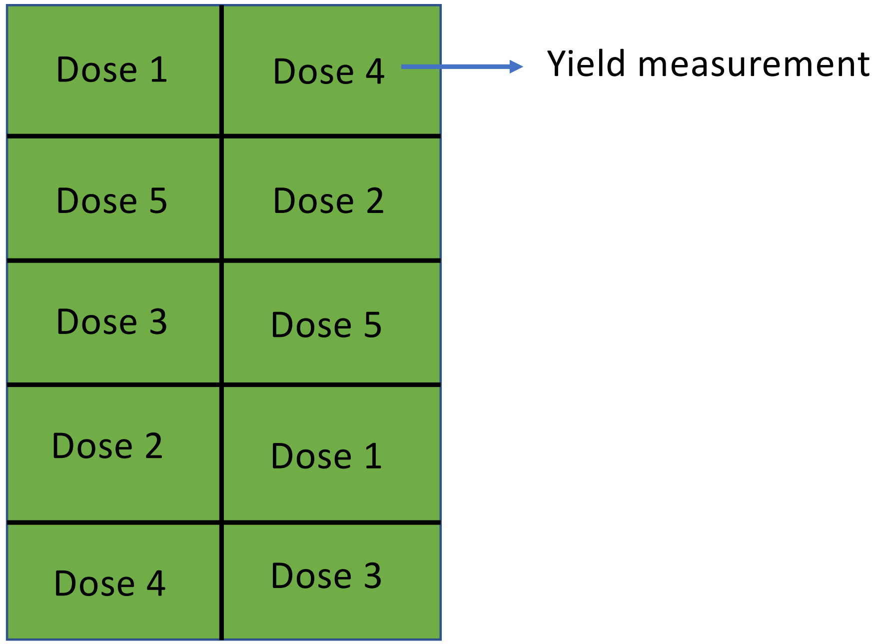


Example 1: N, P, K, fertilization models for potato crops in Eastern Canada

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

237 field trials

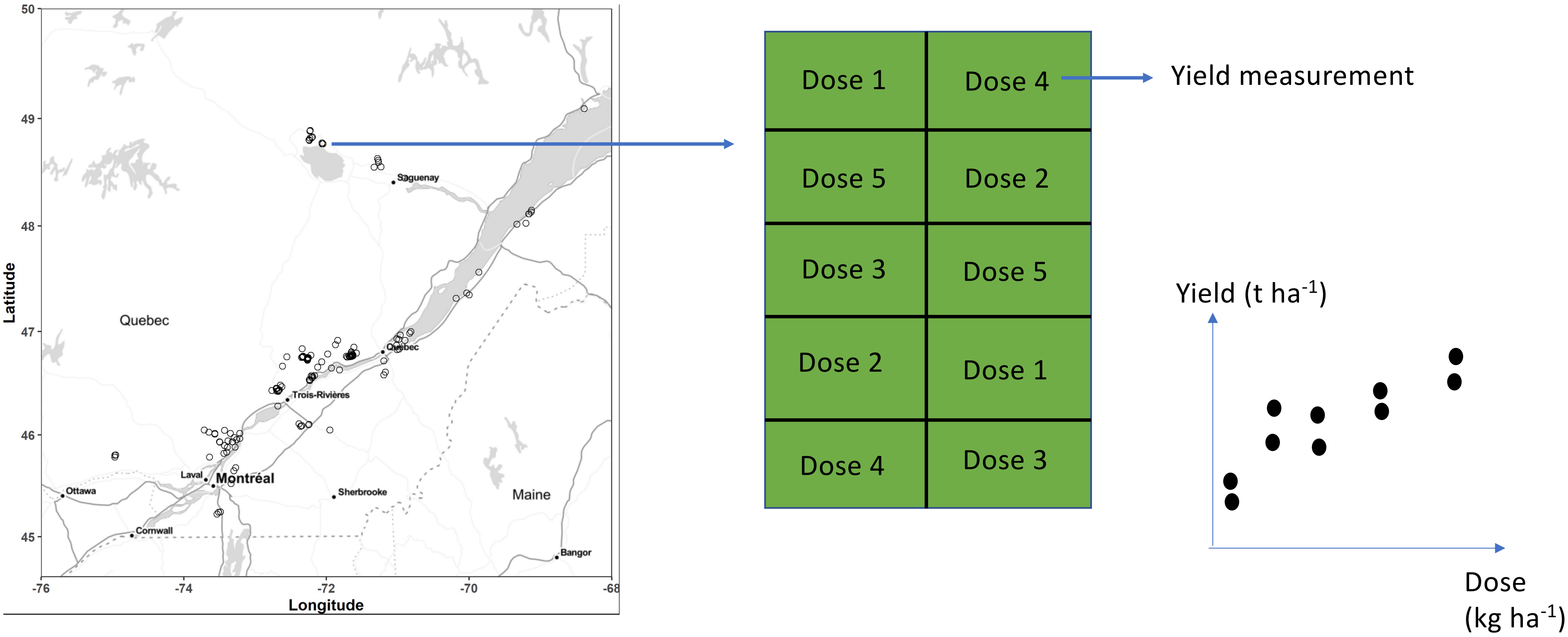




Example 1: N, P, K, fertilization models for potato crops in Eastern Canada

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

237 field trials



Example 1: N, P, K, fertilization models for potato crops in Eastern Canada

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

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

Example 1: N, P, K, fertilization models for potato crops in Eastern Canada

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

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

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

Example 1: N, P, K, fertilization models for potato crops in Eastern Canada

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 (

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

Example 1: N, P, K, fertilization models for potato crops in Eastern Canada

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)

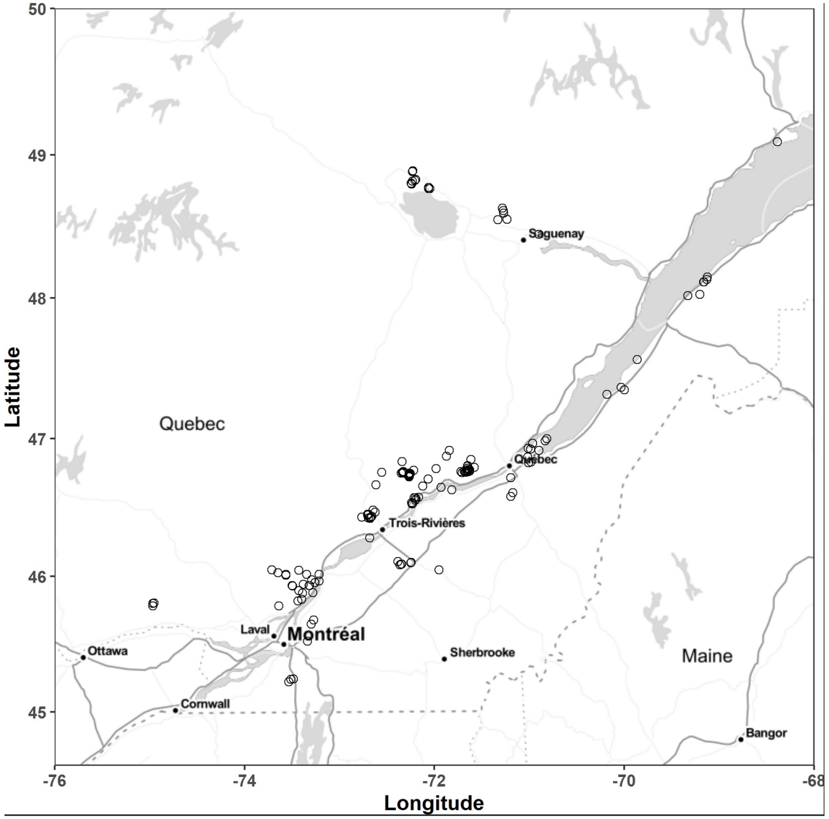
Step 5: Model testing with data (model evaluation)

Step 6: Model application

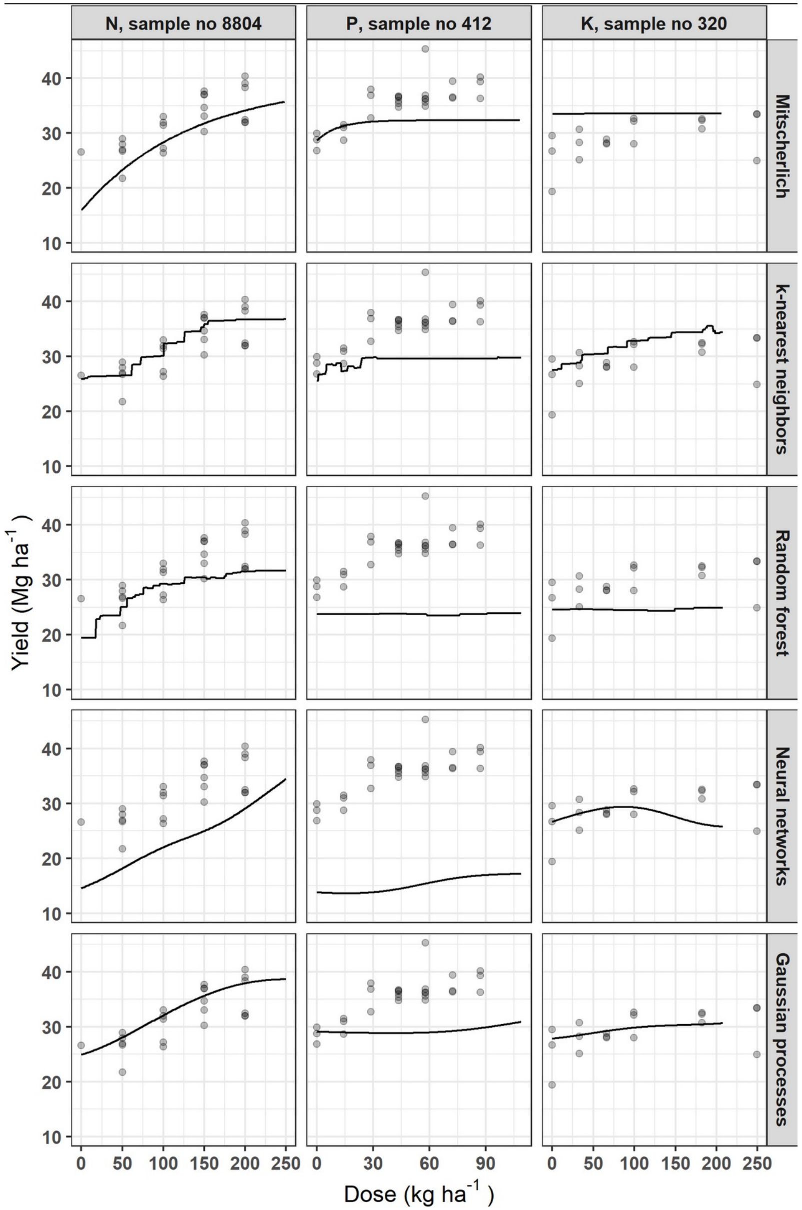
Example 1: N, P, K, fertilization models for potato crops in Eastern Canada

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

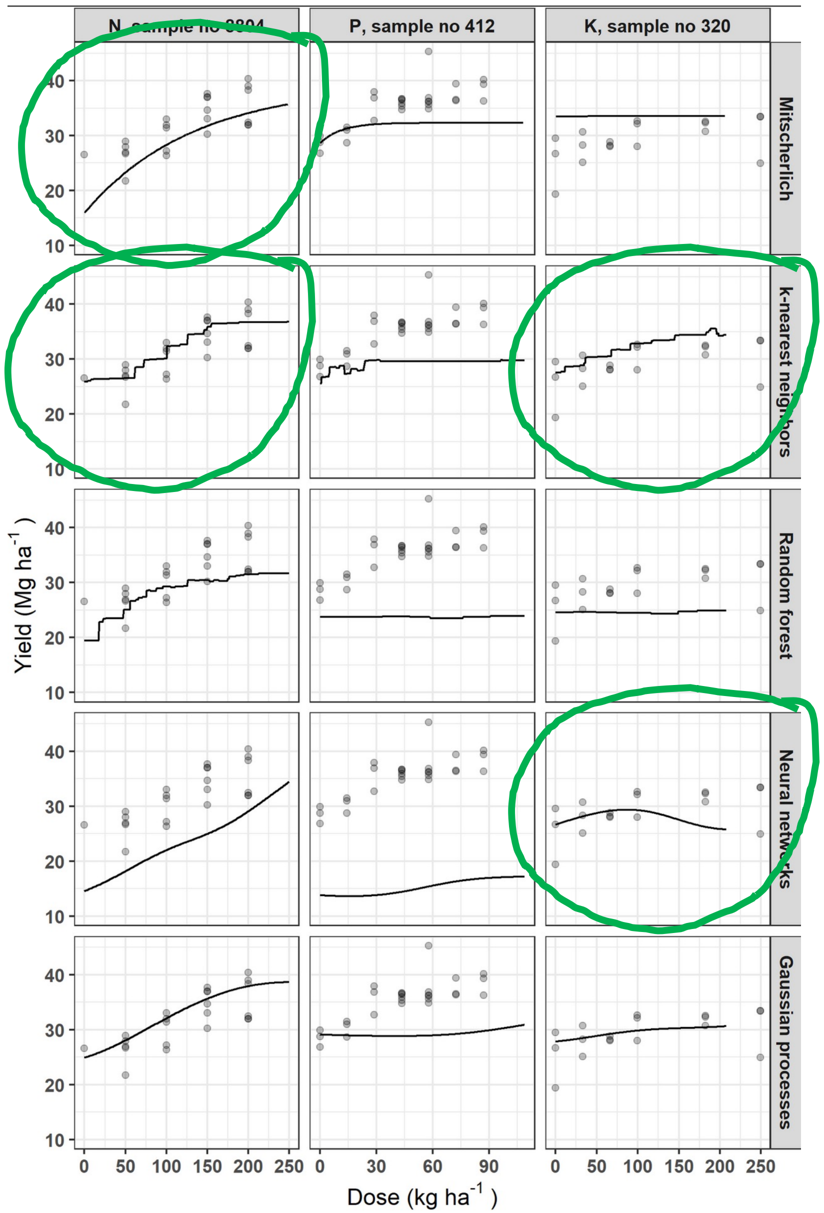
237 field trials



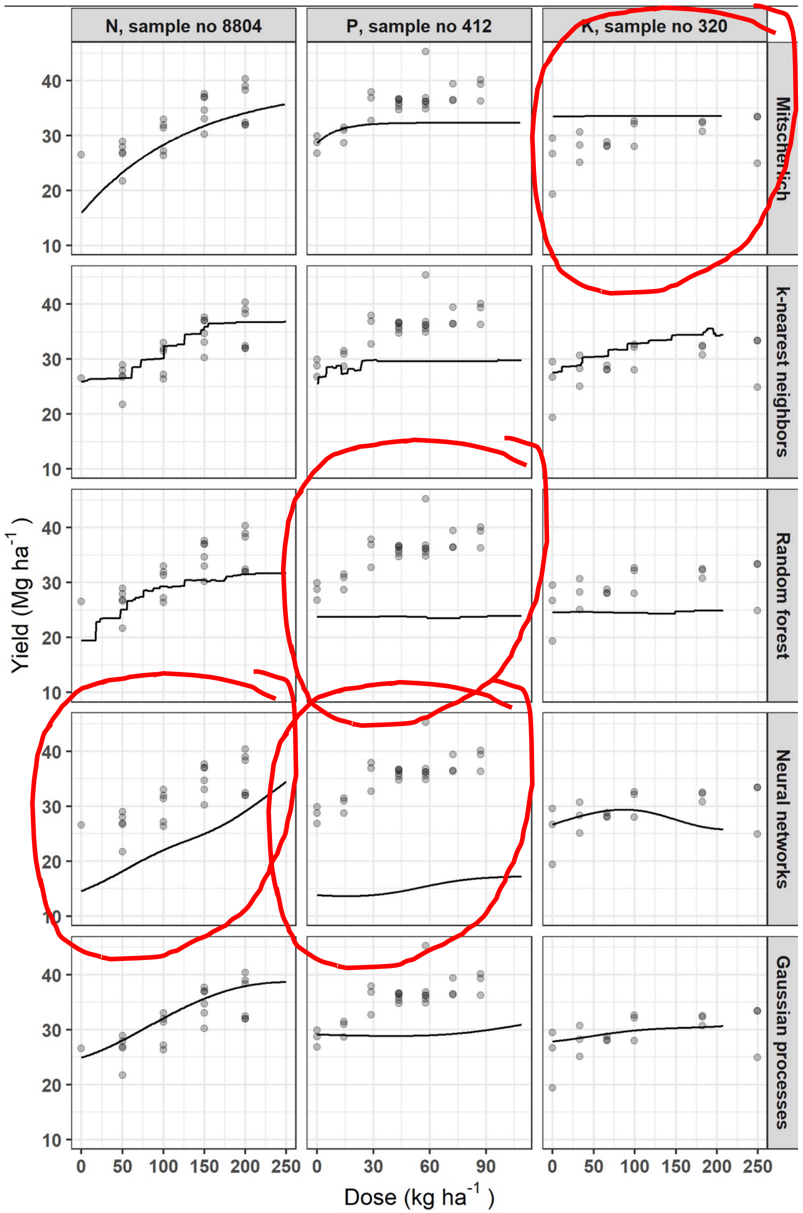
Training dataset
60% of the trials



N, sample no 8804
P, sample no 412
K, sample no 320
Mitscherlich
k-nearest neighbors
Random forest
Neural networks
Gaussian processes
Yield (Mg ha^{-1})
Dose (kg ha^{-1})



Good quality of fit



Poor quality of fit

Example 1: N, P, K, fertilization models for potato crops in Eastern Canada

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)

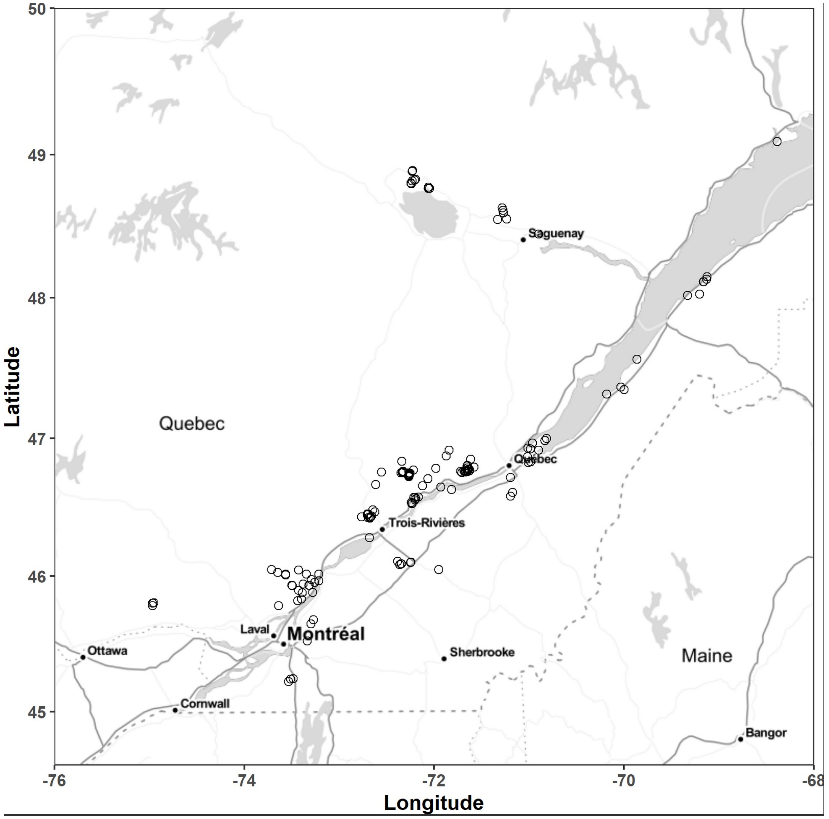
Step 5: Model testing with data (model evaluation)

Step 6: Model application

Example 1: N, P, K, fertilization models for potato crops in Eastern Canada

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

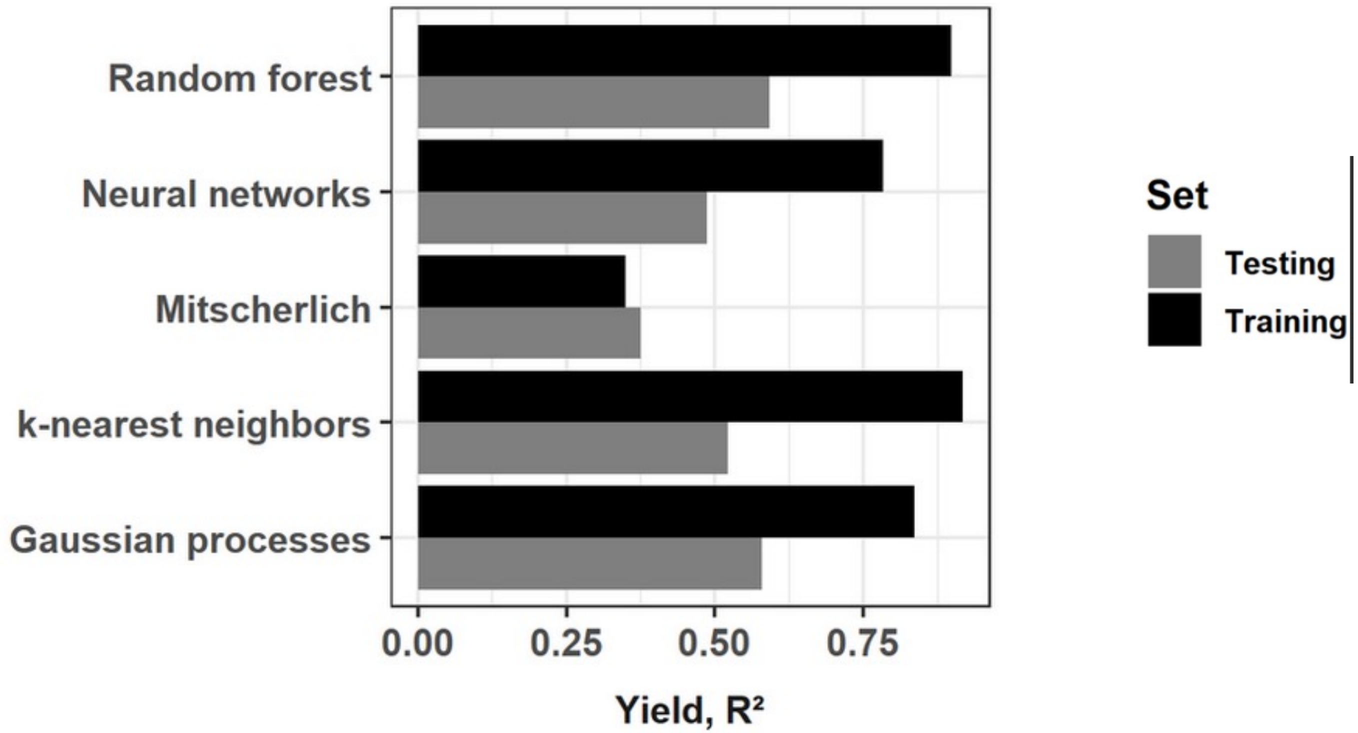
237 field trials



Testing dataset
40% of the trials

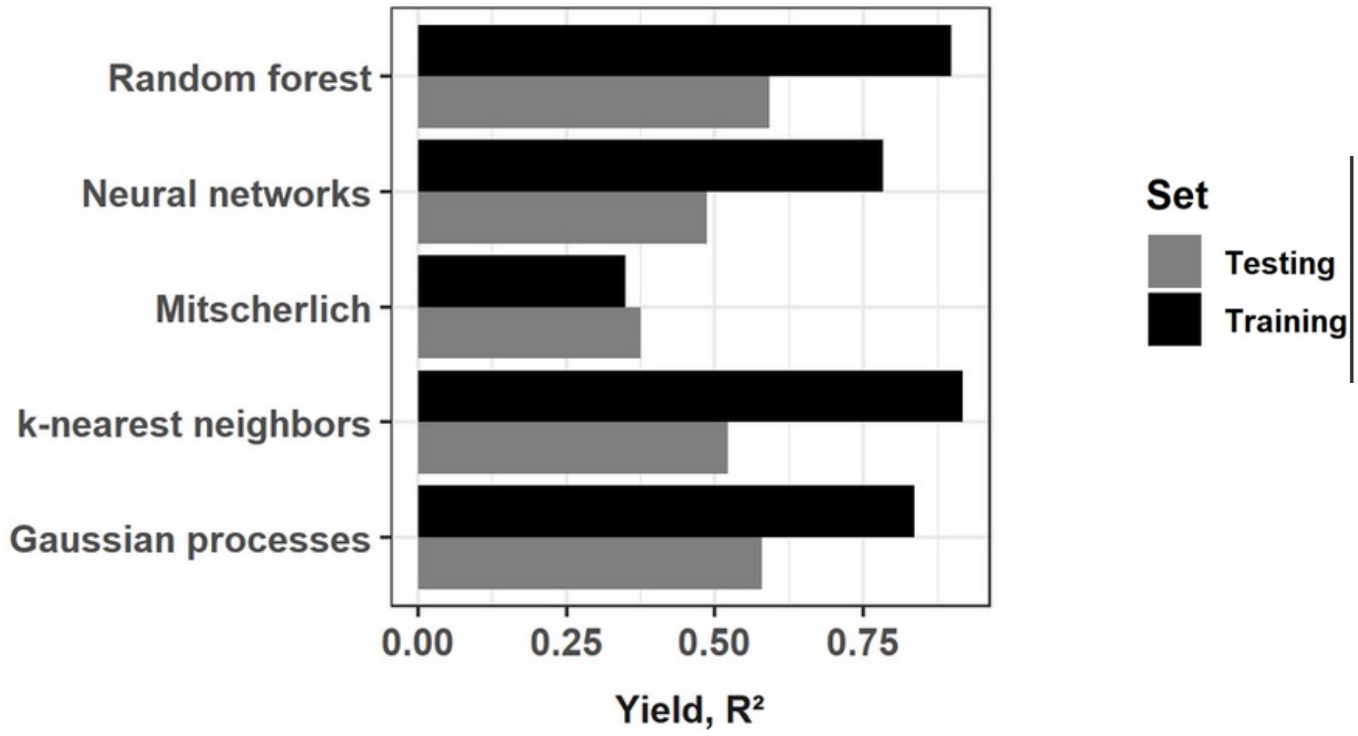
R^2, a popular evaluation criterion

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$$



R², a popular evaluation criterion (the higher, the better)

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$$

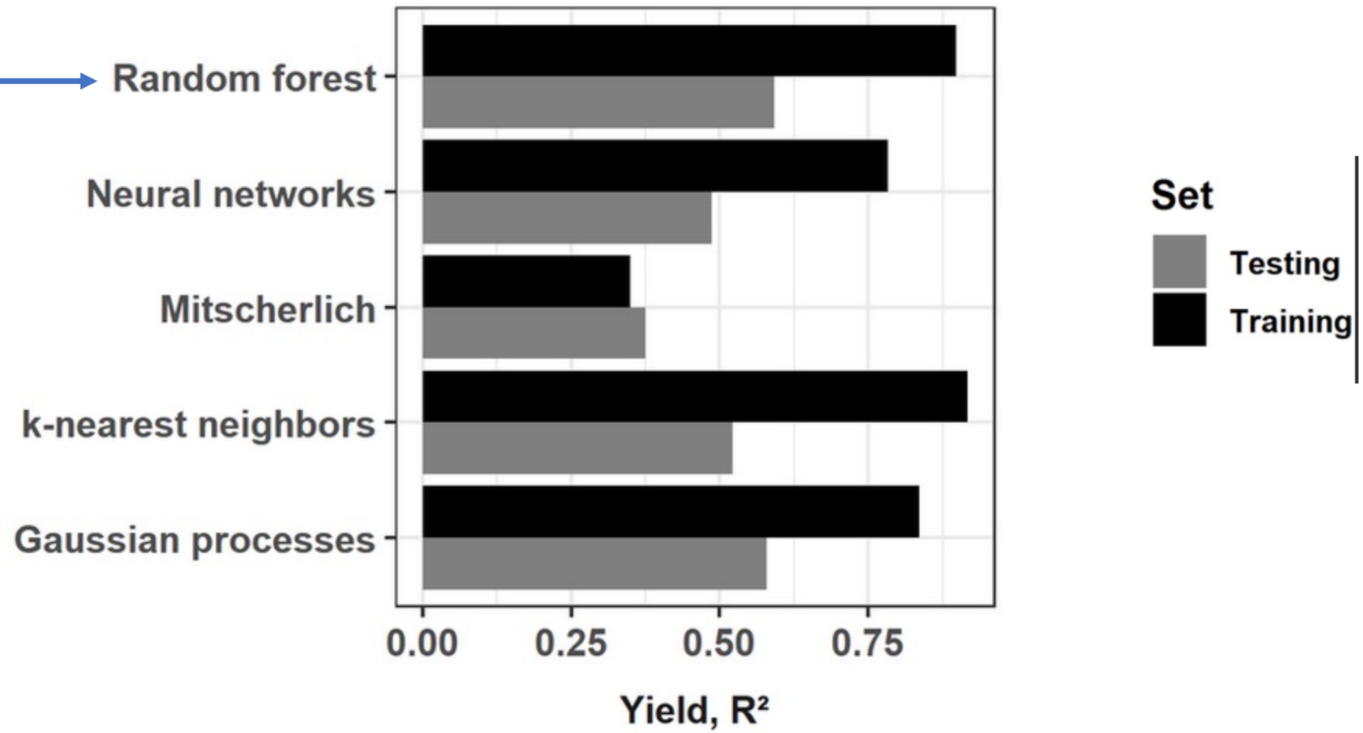


Best model according to the training dataset

R^2, a popular evaluation criterion (the higher, the better)

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2}$$

Best model according to the test dataset

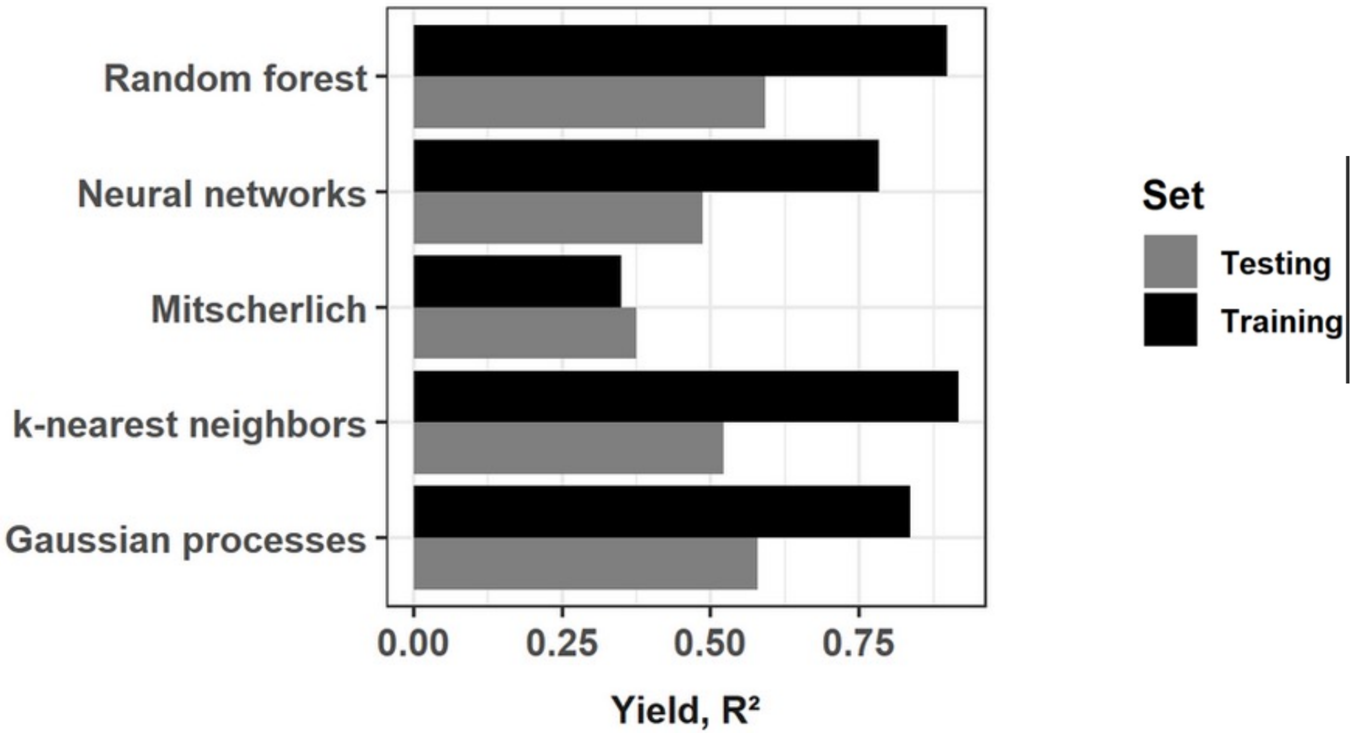


R^2, a popular evaluation criterion

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2}$$

Model performances are too optimistic according to the training dataset.

Important to use an independent test dataset !



Example 1: N, P, K, fertilization models for potato crops in Eastern Canada

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

Step 1: Definition of the objective

Step 2: Data collection

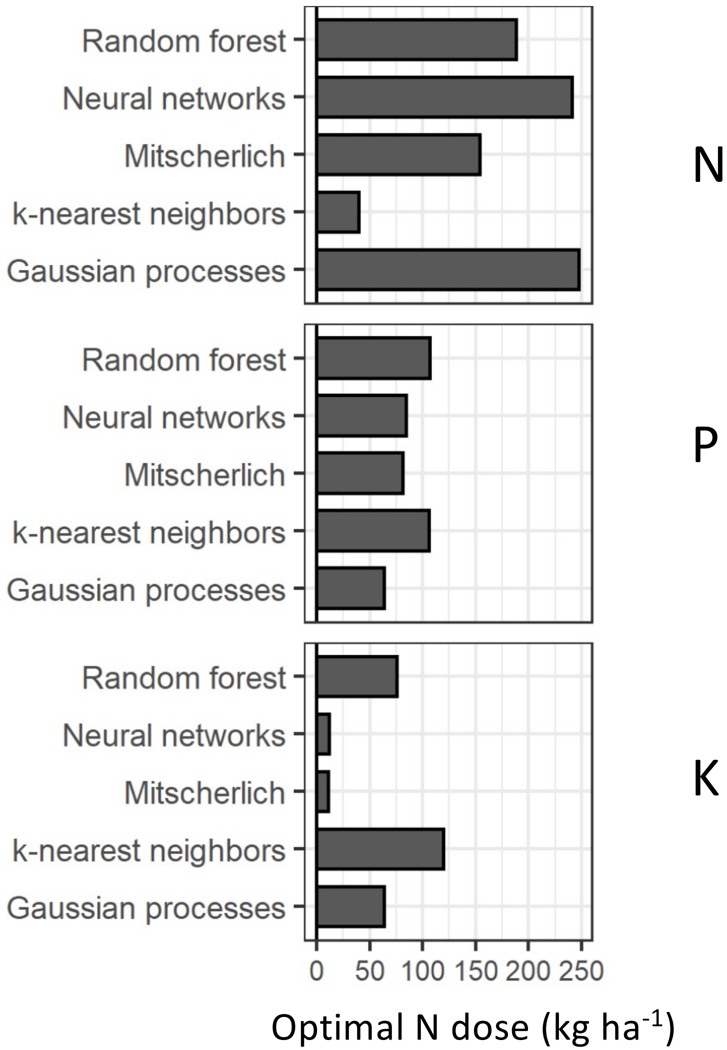
Step 3: Definition of candidate models

Step 4: Model training with data (parameter estimation)

Step 5: Model testing with data (model evaluation)

Step 6: Model application

Examples of optimal economic N doses at one site in Canada



Why modelling is useful for agro-ecology **at field scale**?

- Many decisions are taken at the field scale
  - choice of crop species,
  - 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.)
- **They can be used to address issues at larger scales as well !**

Example 2: Model-based projected impacts of climate change on crop yields

Climate change scenarios

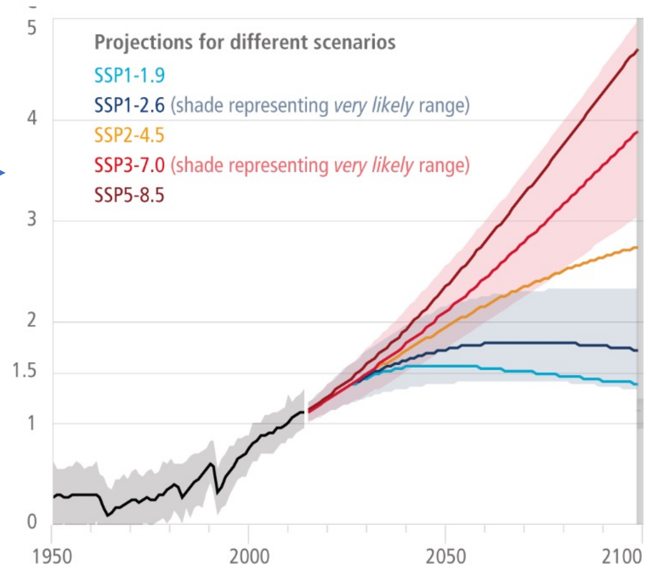
Climate model

Example 2: Model-based projected impacts of climate change on crop yields

Climate change scenarios

Climate model

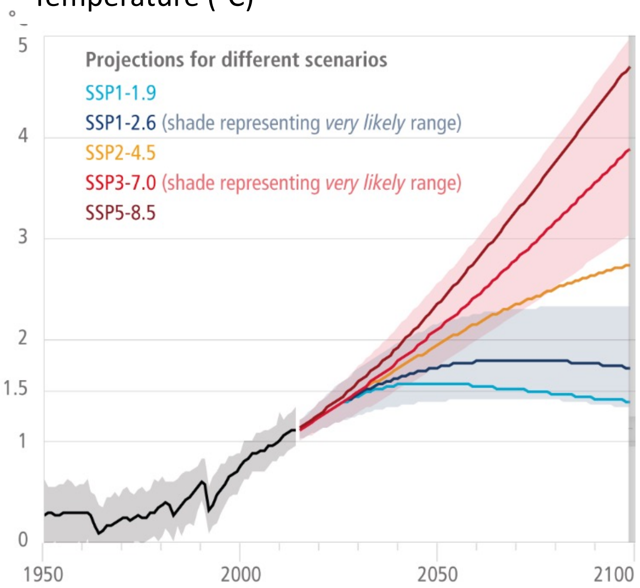
Temperature (°C)



Example 2: Model-based projected impacts of climate change on crop yields

Climate change scenarios

Climate model



Downscaling method

Local weather data **with** climate change

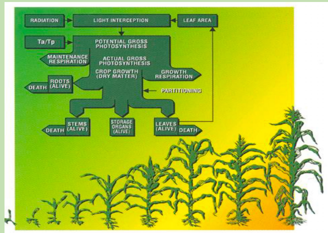
Local weather data **without** climate change (baseline)

Example 2: Model-based projected impacts of climate change on crop yields

Local weather data **with** climate change

32° 32° 33° 34° 34°

Crop model



23° 23° 22° 22° 20°

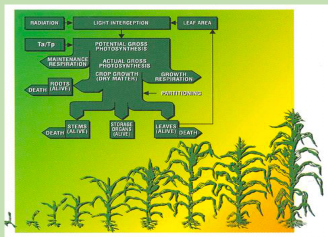
Local weather data **without** climate change (baseline)

Example 2: Model-based projected impacts of climate change on crop yields

Soil characteristics

Local weather data **with** climate change

Crop model



Local weather data **without** climate change (baseline)

Adaptation strategies
- Irrigation
- Drought resistant cultivars
- Sowing/harvest dates

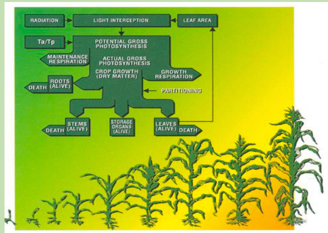
Example 2: Model-based projected impacts of climate change on crop yields

Soil characteristics

Local weather data **with** climate change

32°	32°	33°	34°	34°

Crop model



Yc=Yield with climate change

Y_0=Yield without climate change

23°	23°	22°	22°	20°

Local weather data **without** climate change (baseline)

Adaptation strategies

- Irrigation
- Drought resistant cultivars
- Sowing/harvest dates

Example 2: Model-based projected impacts of climate change on crop yields

Soil characteristics

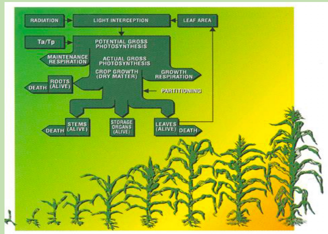
Local weather data **with** climate change

32° 32° 33° 34° 34°

23° 23° 22° 22° 20°

Local weather data **without** climate change (baseline)

Crop model



Yc=Yield with climate change

Y_0=Yield without climate change

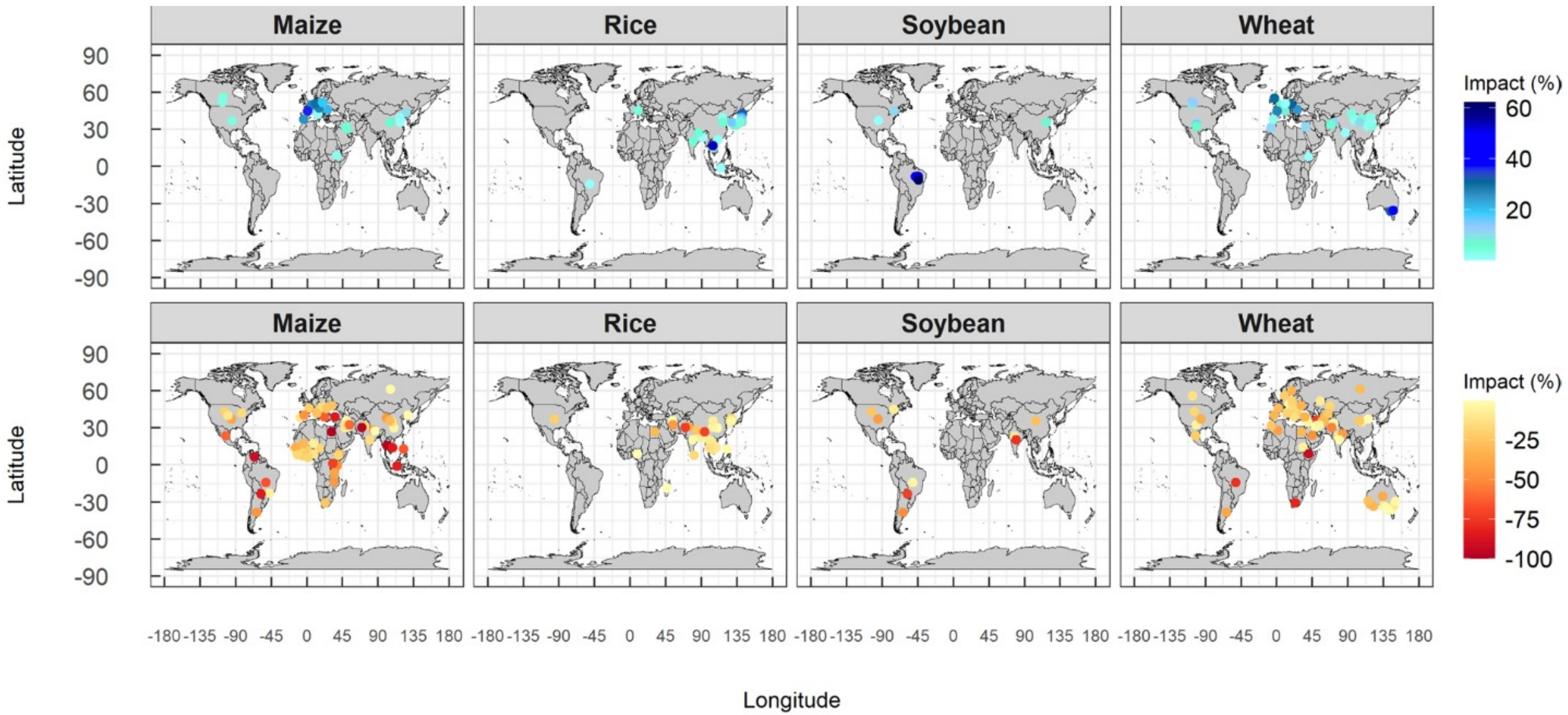
Yield variation= $(Yc-Y_0)/_{Y0}$

Adaptation strategies

- Irrigation
- Drought resistant cultivars
- Sowing/harvest dates

Simulated yield variations due to climate change

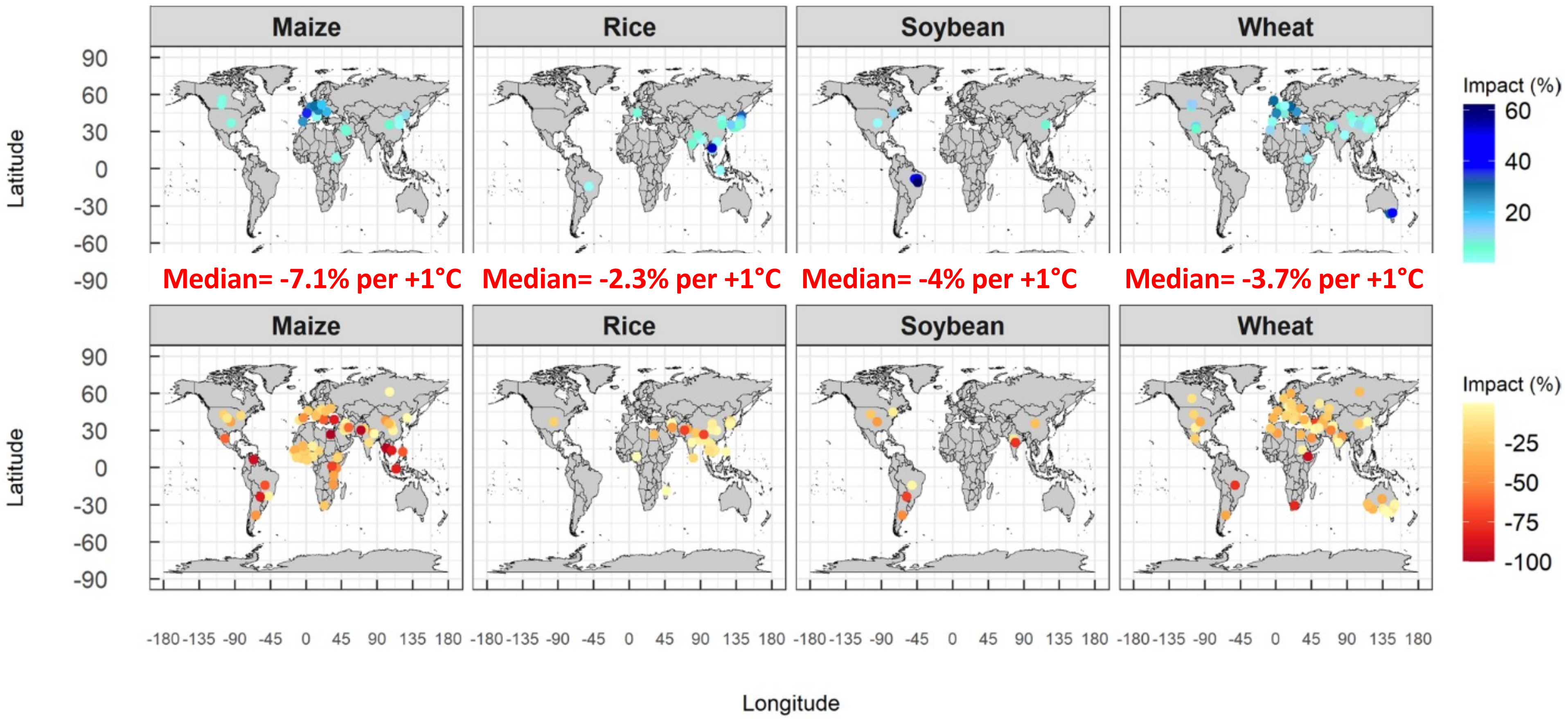
for scenario « RCP8.5 », time horizon 2050, baseline 2001-2010



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

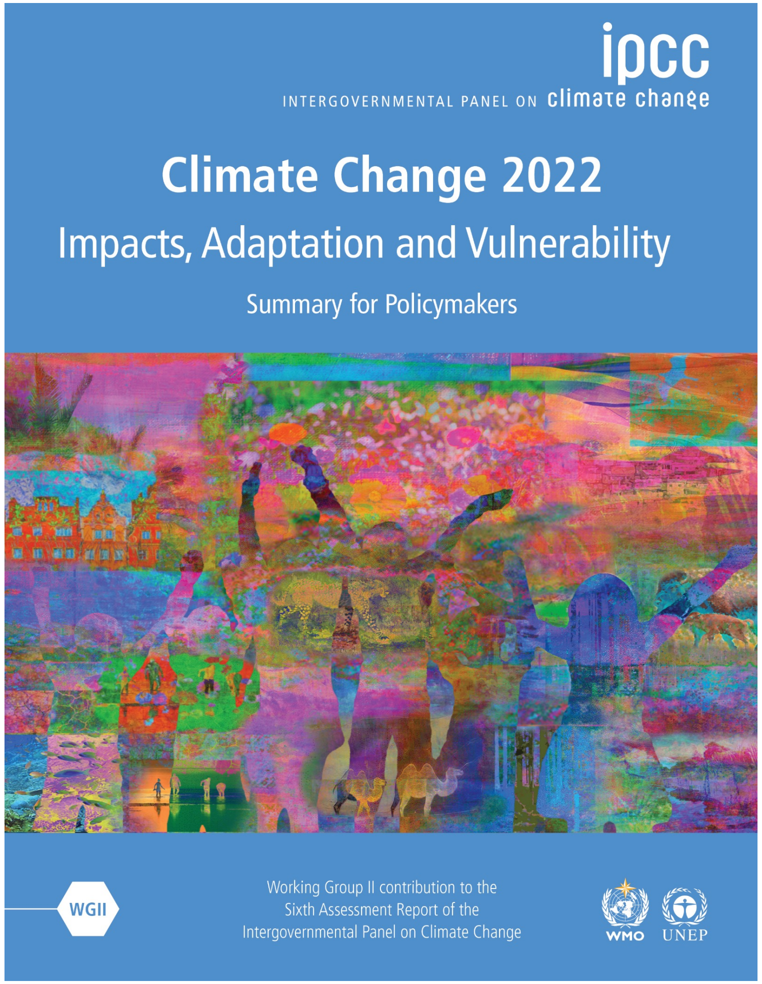
Simulated yield variations due to climate change

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



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

These model simulations were included in the last IPCC report on climate change



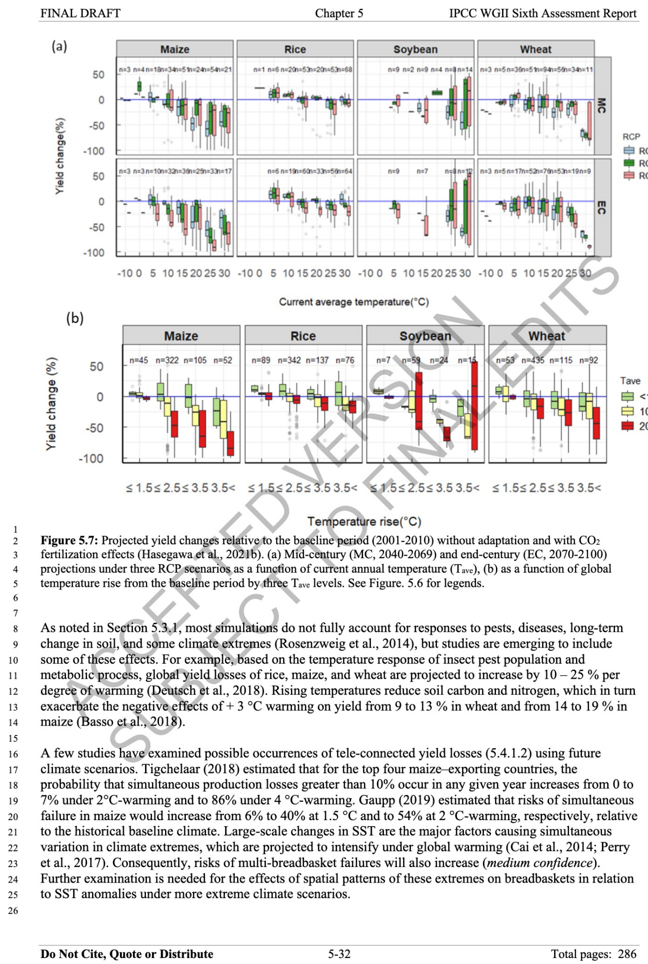


Figure 5.7: Projected yield changes relative to the baseline period (2001-2010) without adaptation and with CO_2 fertilization effects (Hasegawa et al., 2021b). (a) Mid-century (MC, 2040-2069) and end-century (EC, 2070-2100) projections under three RCP scenarios as a function of current annual temperature (T_{ave}), (b) as a function of global temperature rise from the baseline period by three T_{ave} levels. See Figure. 5.6 for legends.

As noted in Section 5.3.1, most simulations do not fully account for responses to pests, diseases, long-term change in soil, and some climate extremes (Rosenzweig et al., 2014), but studies are emerging to include some of these effects. For example, based on the temperature response of insect pest population and metabolic process, global yield losses of rice, maize, and wheat are projected to increase by 10 – 25 % per degree of warming (Deutsch et al., 2018). Rising temperatures reduce soil carbon and nitrogen, which in turn exacerbate the negative effects of + 3 °C warming on yield from 9 to 13 % in wheat and from 14 to 19 % in maize (Basso et al., 2018).

A few studies have examined possible occurrences of tele-connected yield losses (5.4.1.2) using future climate scenarios. Tigchelaar (2018) estimated that for the top four maize–exporting countries, the probability that simultaneous production losses greater than 10% occur in any given year increases from 0 to 7% under 2°C-warming and to 86% under 4 °C-warming. Gaupp (2019) estimated that risks of simultaneous failure in maize would increase from 6% to 40% at 1.5 °C and to 54% at 2 °C-warming, respectively, relative to the historical baseline climate. Large-scale changes in SST are the major factors causing simultaneous variation in climate extremes, which are projected to intensify under global warming (Cai et al., 2014; Perry et al., 2017). Consequently, risks of multi-breadbasket failures will also increase (*medium confidence*). Further examination is needed for the effects of spatial patterns of these extremes on breadbaskets in relation to SST anomalies under more extreme climate scenarios.

Why modelling is fascinating?

- Multi-criteria assessment of a diversity of farming practices
- Explore the future (ex: climate change)
- Quickly generate a lot of data at low cost
- Sometimes useful to support decisions of farmers, companies, policy makers

Why modelling can be dangerous sometimes?

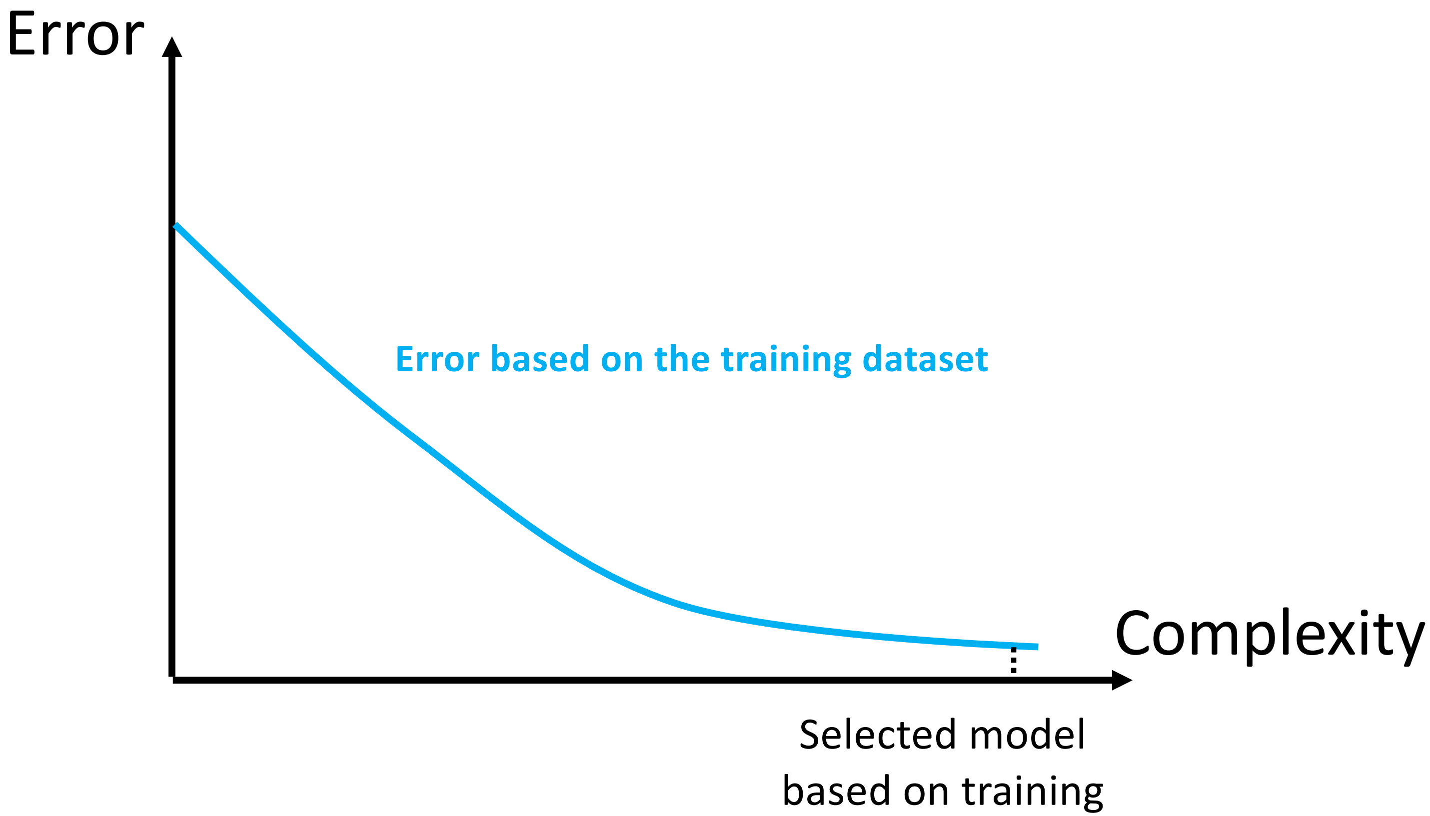
- All models are wrong
- Some models can be lead to poor predictions and decisions
- Some models are not correctly evaluated using reliable data

Why modelling can be dangerous sometimes?

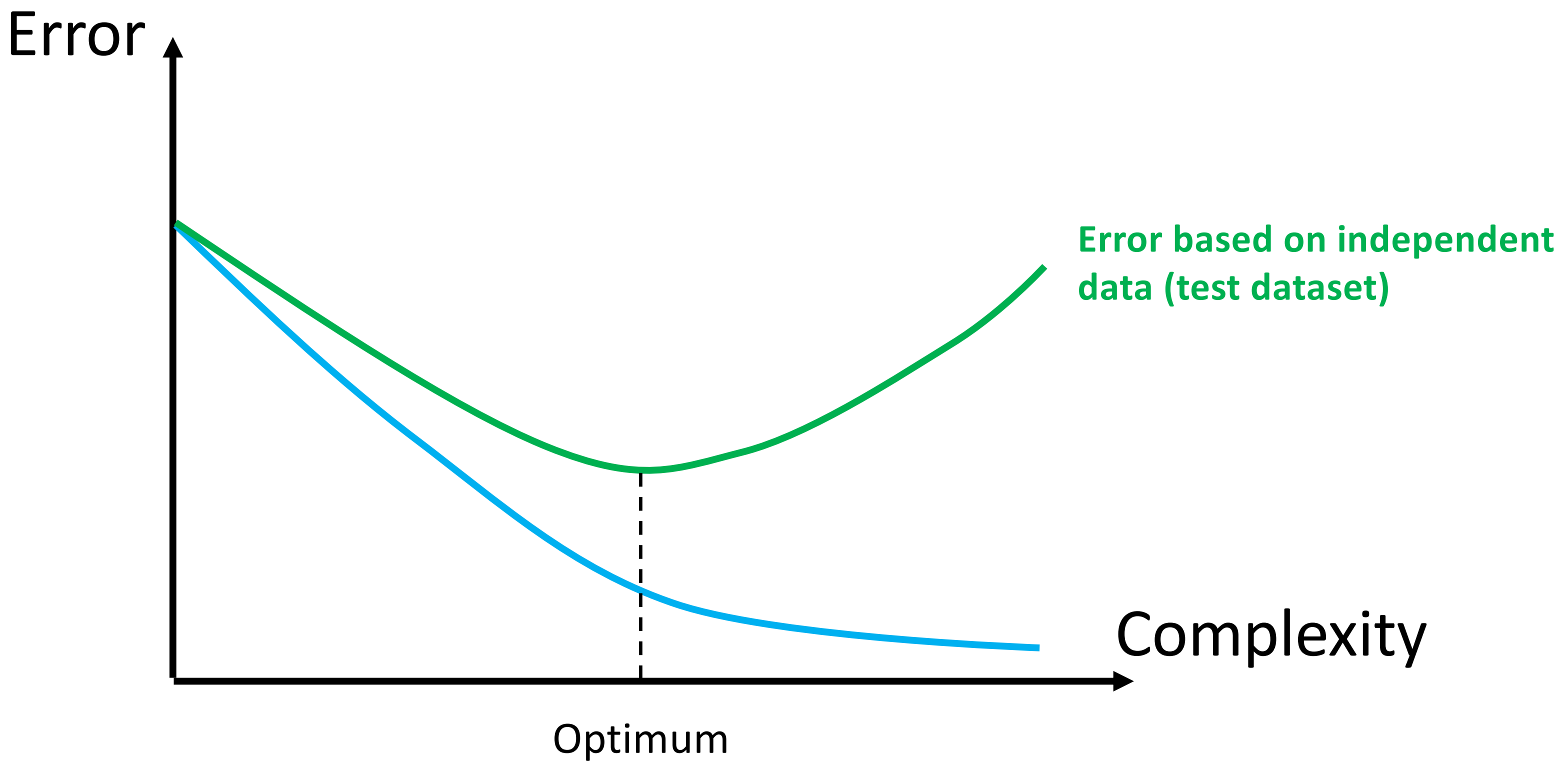
- All models are wrong
- Some models can be lead to poor predictions and decisions
- Some models are not correctly evaluated using reliable data
- Complex models are not easy to understand and are not necessary more accurate than simple ones.

➡ **Important to develop models using rigorous methods**

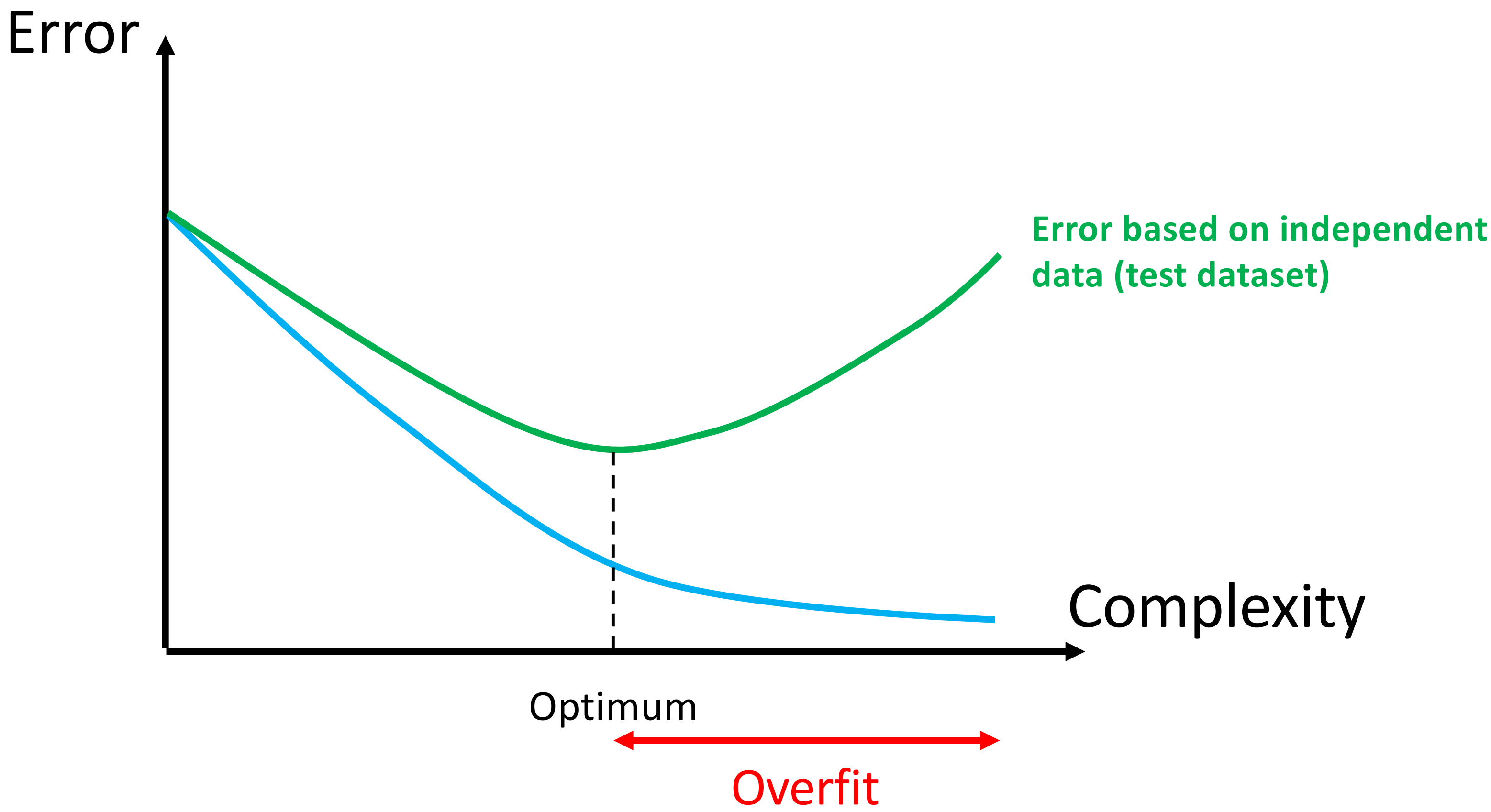
Model testing should be taken seriously to avoid risk of overfitting



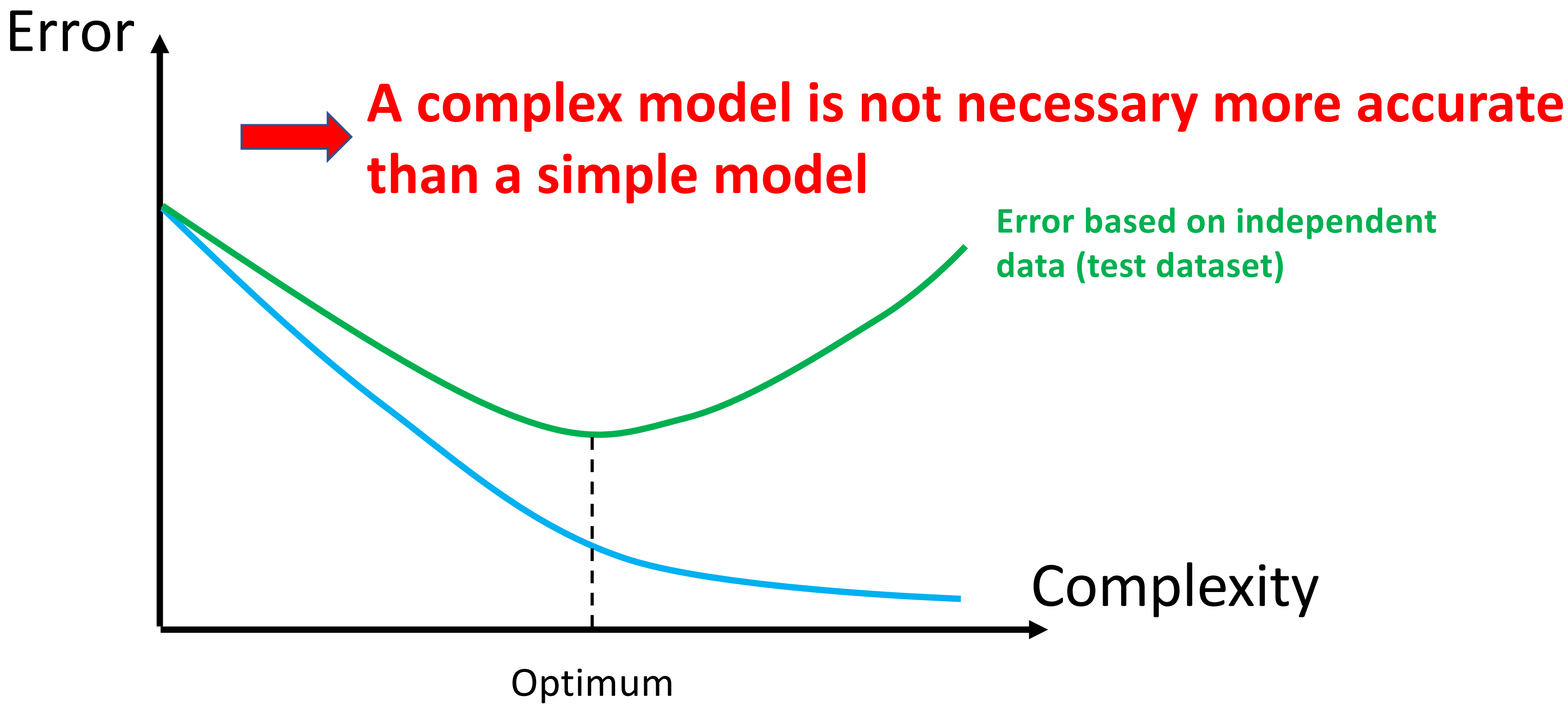
Model testing should be taken seriously to avoid risk of overfitting



Model testing should be taken seriously to avoid risk of overfitting



Model testing should be taken seriously to avoid risk of overfitting



Limit of modelling at field scale

- Not the best scale to evaluate all economic and social impacts of agroecology
- Not always the best scale to evaluate the environmental impact of agroecology

BUT

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

Being a responsible model developer and user; what you should **not do**

1. Don’t believe that your models represent the truth

    « All models are wrong, some are useful »

2. Don’t use a single model
3. Don’t overfit your models to data
4. Don’t protect your models from external evaluation

Being a responsible model developer and user; what you should **not do**

1. Always try to challenge your model
2. Use ensemble of models and compare conclusions
3. Be aware of the risk of overfitting with complex models
4. Be parsimonious (always consider the option of using a simple model)
5. Be transparent, share your data, equations, and codes

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
    - ➢Parameter estimation
    - ➢Sensitivity analysis
    - ➢Uncertainty analysis
    - ➢Forecasting
    - ➢Data assimilation
    - ➢Model performance evaluation
    - ➢Optimization & decision support

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
  - Parameter estimation
  - Sensitivity analysis
  - Uncertainty analysis
  - Forecasting
  - Data assimilation
  - Model performance evaluation
  - Optimization & decision support
- Coding
  - R
  - Python
  - etc.

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

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