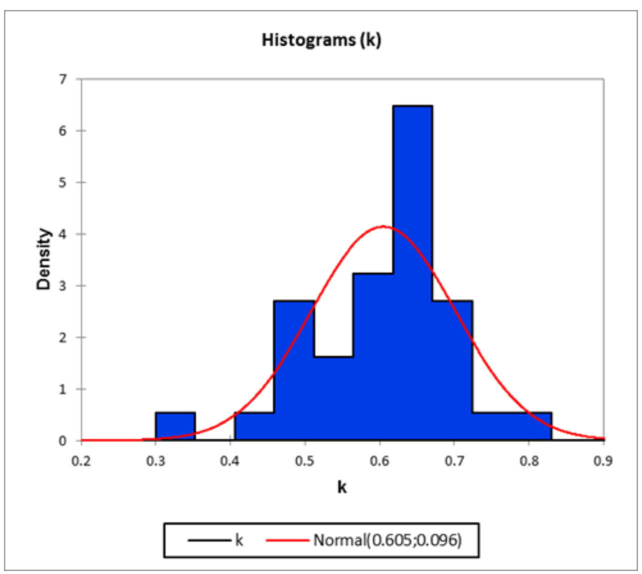
# In silico ideotyping

Development of crop ideotypes

1. Define **ranges/statistical distributions** for **trait** values



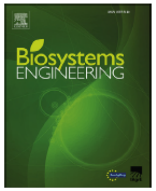
BIOSYSTEMS ENGINEERING 164 (2017) 1–12

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110

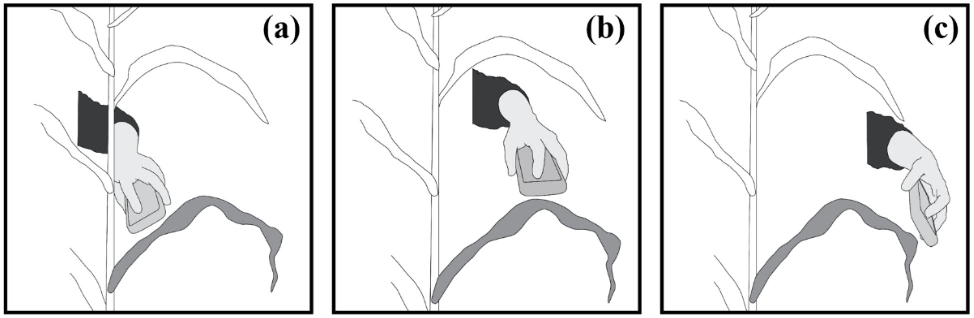




Research Paper

## PocketPlant3D: Analysing canopy structure using a smartphone

CrossMark


Roberto Confalonieri [a,*], Livia Paleari [a], Marco Foi [b], Ermes Movedi [c], Fosco M. Vesely [a], William Thoelke [a,d], Cristina Agape [d], Giulia Borlini [d], Irene Ferri [d], Federico Massara [d], Roberto Motta [d], Riccardo A. Ravasi [d], Sofia Tartarini [d], Camilla Zoppolato [d], Luca M. Baia [d], Andrea Brumana [d], Davide Colombo [d], Antonio Curatolo [d], Valerio Fauda [d], Denise Gaia [d], Andrea Gerosa [d], Antonio Ghilardi [d], Enrico Grassi [d], Andrea Magarini [d], Francesco Novelli [d], Fatima B. Perez Garcia [d], Andrea Rota Graziosi [d], Michele Salvan [d], Tommaso Tadiello [d], Laura Rossini [e]

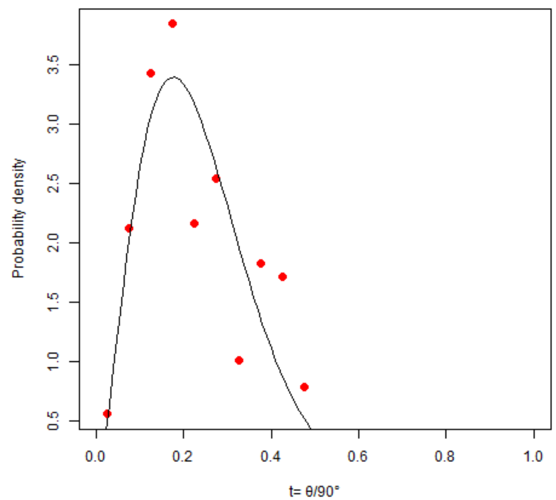



# In silico ideotyping



1. Define **ranges/statistical distributions** for **trait** values

- Phenotyping **barley** lines (e.g., *k*)
  - ✓ line7165
    - $\chi = 0.527$
    - $k = 0.38$

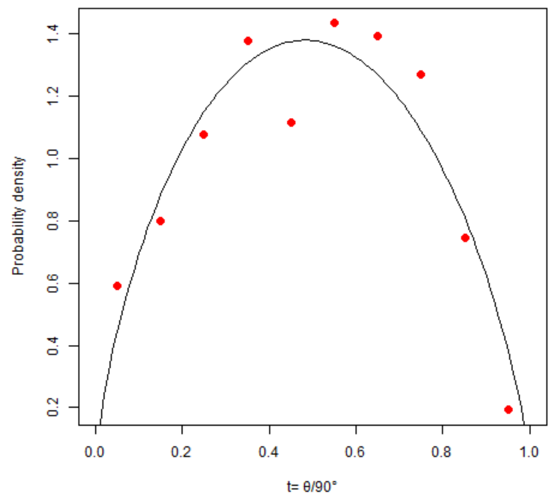



# In silico ideotyping

Development of crop ideotypes

1. Define **ranges/statistical distributions** for **trait** values

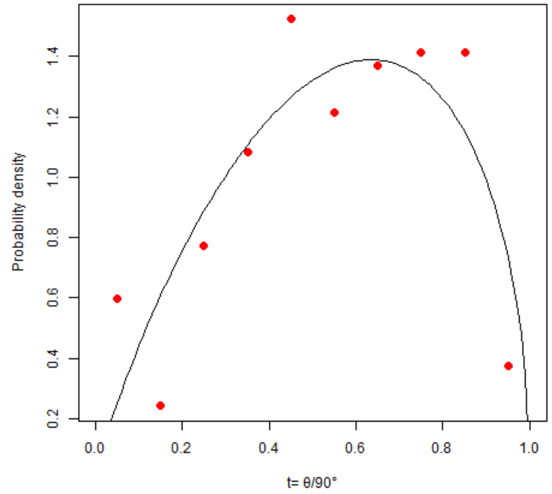
- Phenotyping **barley** lines (e.g., *k*)
  - ✓ Calanque
    - o $\chi$ = 1.537
    - o k = 0.76




# In silico ideotyping

1. Define **ranges/statistical distributions** for **trait** values

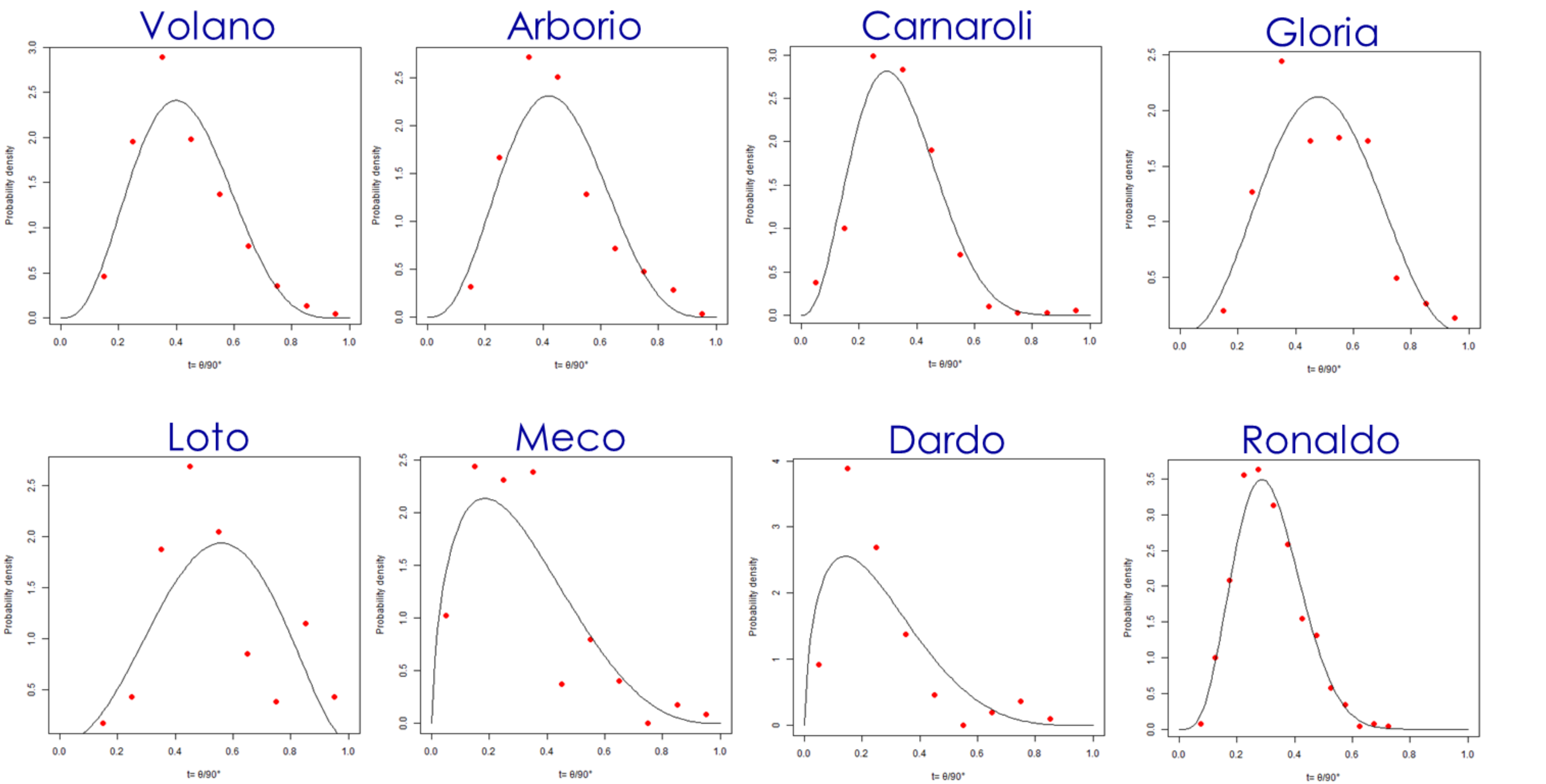
- Phenotyping **barley** lines (e.g., *k*)
  - ✓ Cometa
    - $\chi = 1.835$
    - $k = 0.84$




# In silico ideotyping

1. Define **ranges/statistical distributions** for **trait** values

- Phenotyping **rice** varieties (e.g., *k*)

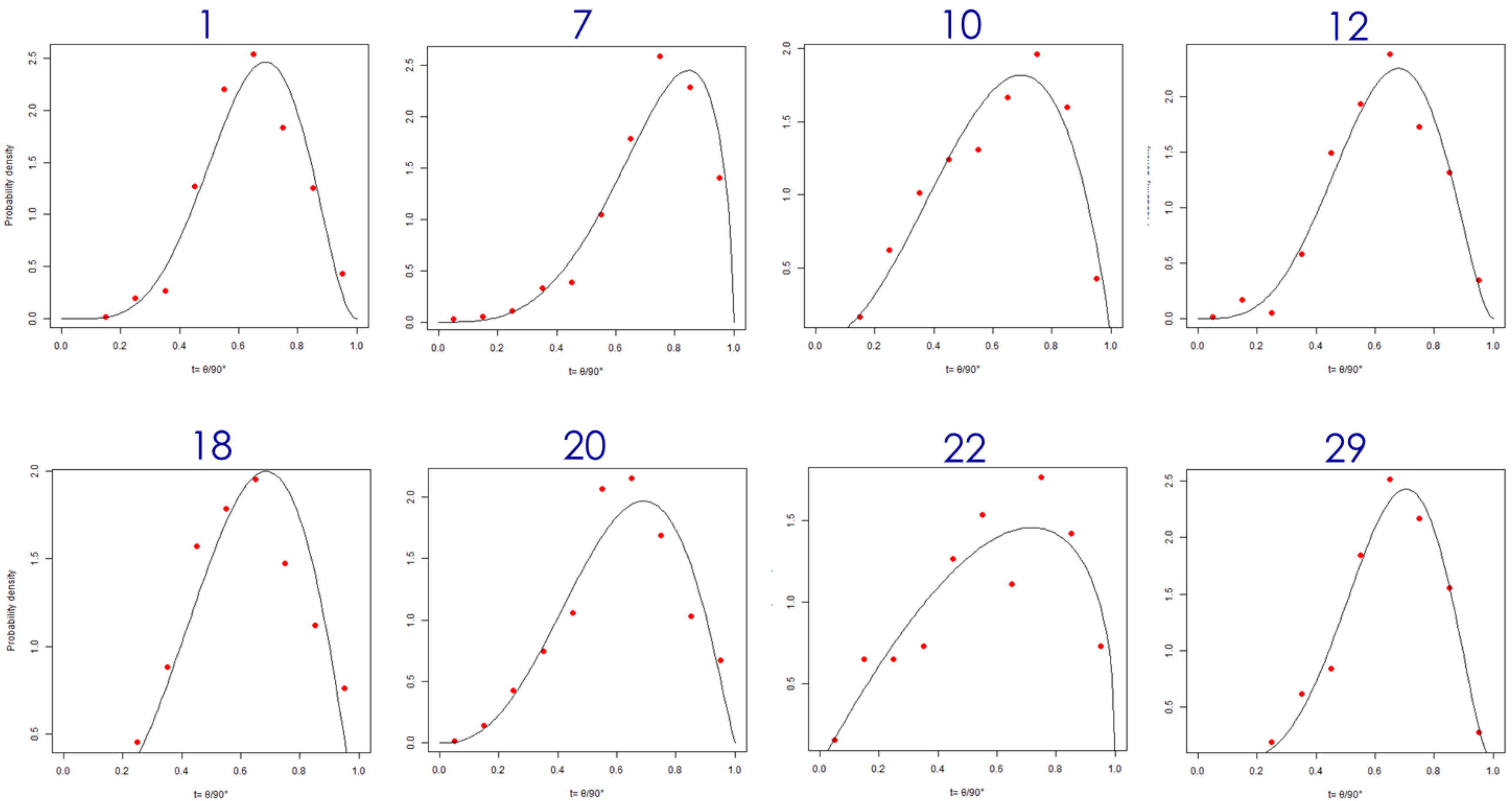


# In silico ideotyping



1. Define **ranges/statistical distributions** for **trait** values

- Phenotyping **bean** lines (e.g., *k*)



# In silico ideotyping

1. Define **ranges/statistical distributions** for **trait** values

| Parameter | Relevance for breeding (e.g.) | Distribution | Source |
|---|---|---|---|
| Radiation use efficiency (RUE; g $MJ^{-1}$) | Peng et al. 2008; Dingkhun et al. 2015 | Normal (m 2.7; s 0.1) | Kiniry et al. 2001; Boschetti et al. 2006 |
| Extinction coefficient (k; -) | Peng et al. 2008; Sheehy et al. 2013 | Normal (m 0.47; s 0.04) | Casanova et al. 1998; Dingkhun et al. 1999; Kiniry et al. 2001; Boschetti et al. 2006 |
| SLA at emergence (SLAini; $m^2$ $kg^{-1}$) | Peng et al. 2008; Kush et al. 2012 | Normal (m 41.6; s 5.9) | Kropff et al. 1994; Ash et al. 1998; Confalonieri and Bocchi 2005 |
| SLA at tillering (SLAtill; $m^2$ $kg^{-1}$) | Ashikari et al. 2005; Peng et al. 2008; | Normal (m 28.7; s 3.9) | Laza et al. 2015; Boschetti et al. 2006 |
| Threshold T for cold sterility (T-ColdSter; °C) | Suh et al. 2010; Sanchez et al. 2014 | Normal (temp. m 13.5; s 1.4) (trop. M 16.6; s 1.2) | Satake 1969; Da Cruz et al. 2006; Farrel et al. 2006; Thakur et al. 2010; Deng et al. 2011; Dreni et al. 2012; National Rice Authority |
| Threshold T for heat sterility (T-HeatSter; °C) | Matsui 2009; Jagadish et al. 2010 | Normal (m 34.4; s 1.5) | Yoshida 1981; Satake 1995; Nakagawa et al. 2002; Matsui 2009; Ishimaru et al. 2010; Jagadish et al. 2010; Shah et al. 2011; Maruyama et al. 2013 |
| Blast resistance (BlastRes; -; 1 to 3) | Fisher et al. 2005; Fukoka et al. 2009 | Discrete (1, 2, 3) | National Rice Authority |
| Threshold T for chalkiness (T-Chalkiness; °C) | Yamakawa et al. 2007; Usui et al. 2014 | Normal (m 26.4; s 0.9) | Wakamatsu et al. 2007; Yamakawa et al. 2007; Morita et al. 2008; Madan et al. 2012; Usui et al. 2014; Matsutomi et al. 2015 |
| Threshold T for grain breakage (T-HeadRIce; °C) | Siebenmorgen et al. 2013; Sreenivasulu et al. 2015 | Normal (m 23.9; s 2.1) | Ambardekar et al. 2011; Okada et al. 2011; Siebenmorgen et al. 2013 |

# In silico ideotyping



1. Define **ranges/statistical distributions** for **trait** values
2. Identify **most relevant traits**

✓ Global **sensitivity analysis**

Received: 29 October 2016 | Revised: 24 February 2017 | Accepted: 27 February 2017

DOI: 10.1111/gcb.13682

PRIMARY RESEARCH ARTICLE

WILEY Global Change Biology

## Surfing parameter hyperspaces under climate change scenarios to design future rice ideotypes

Livia Paleari[1] | Ermes Movedi[1] | Giovanni Cappelli[2*] | Lloyd T. Wilson[3] | Roberto Confalonieri[4]

# Sensitivity analysis



- **Objective**: quantifying the role of uncertain input factors in explaining the variability of the outputs of mathematical models.
- It is often used to **identify** the model **parameters** that – under **specific conditions** – have the **largest effect** on model outputs.

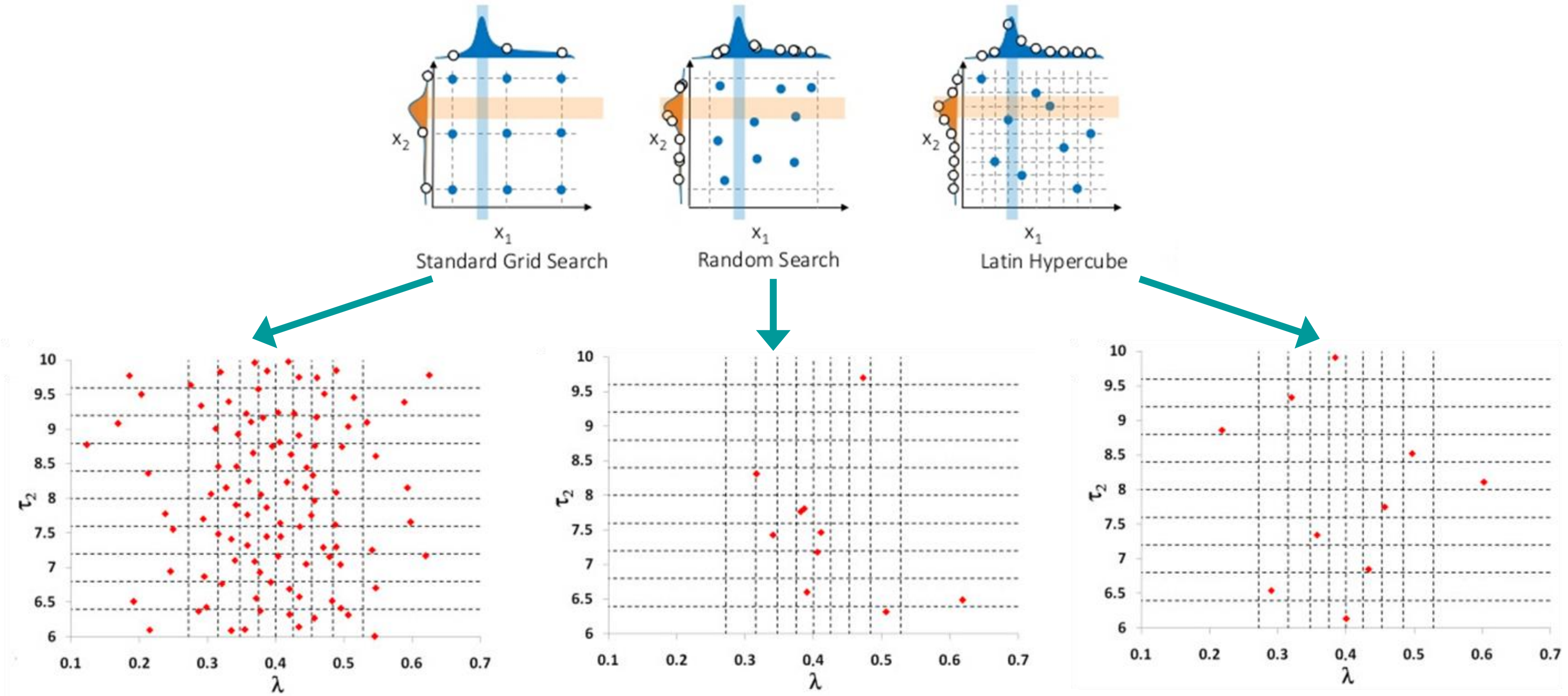
# Sensitivity analysis



- The **rationale** is to quantify changes in model outputs occurring because of changes in model inputs.
- The first idea could be:
  - ✓ **Dividing** the biophysical **range** of each input in a certain number of **regular intervals**.
  - ✓ For each input, running simulations assigning to the input the value of each interval.
    - One input at a time
    - *N*-dimensional grids (*N* being the number of inputs)

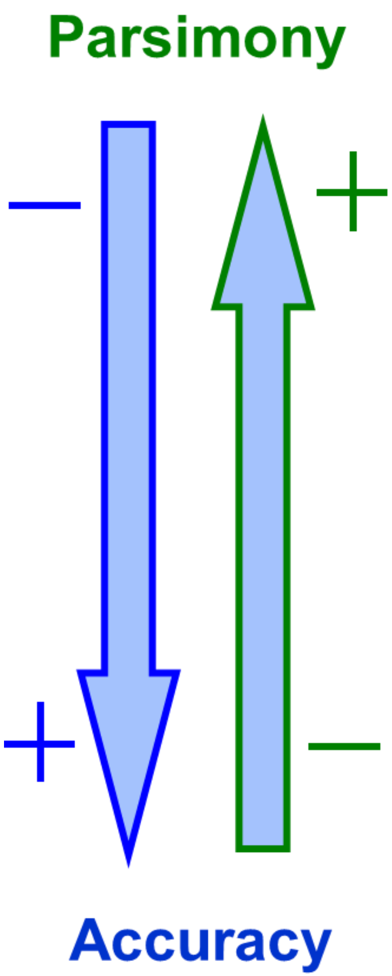
# Sensitivity analysis

- **Problems**:
  - ✓ **How many intervals**? (response functions often discontinuous)
  - ✓ **If many** intervals and ***N*-dimensional grids** → the number of simulations can be **huge**



# Sensitivity analysis

- **Sensitivity analysis methods** were proposed to efficiently **explore** the parameter **hyperspace**




- **Screening** methods (mean and standard deviations of incremental ratios)
  - ✓ Morris
- **Regression-based** methods
  - ✓ Latin Hypercube Sampling (LHS)
  - ✓ Random
  - ✓ Quasi-Random LpTau (LpTau)
- **Variance-based** methods
  - ✓ Sobol’
  - ✓ Fourier Amplitude Sensitivity test (FAST)
  - ✓ Extended FAST

## Morris method

- It is **very fast** (few model executions are needed).
- **Useful** in case of models with **many inputs** and/or very **demanding** in terms of computational time.
- **Sometimes** used to identify parameters with a low impact on output variability
  - ✓ The others are then analyzed using methods requiring many executions (**2-step approach**).
- In **comparative studies**, it demonstrated **effectiveness in ranking** parameters according to their relevance
  - ✓ Used alone (1 step) when the aim is obtaining a **qualitative ranking** of inputs.

# Sensitivity analysis

- Assuming
  - ✓ $X = (x_1, \dots, x_i, \dots, x_N)$ as the vector of the $\boldsymbol{N}$ **inputs** on which the sensitivity analysis is being performed
  - ✓ $y(X)$ as the model **output**.
- **Rescale** each parameter $x_i$ in the interval $\mathbf{[0, 1]}$.
- **Force** each $x_i$ **to assume values** in the set $\{0, 1/(p-1), 2/(p-1), \dots, 1\}$, $\boldsymbol{p}$ being the number of **levels**.
- The **parameter space** $\Omega$ is then defined as a $N$-dimensional, $p$-level unit hypercube.

# Sensitivity analysis



- Assuming $\boldsymbol{\Delta}$ as $1/[2(p-1)]$, a number of **incremental ratios** (**elementary effects**) $(R_i)$ is calculated as:

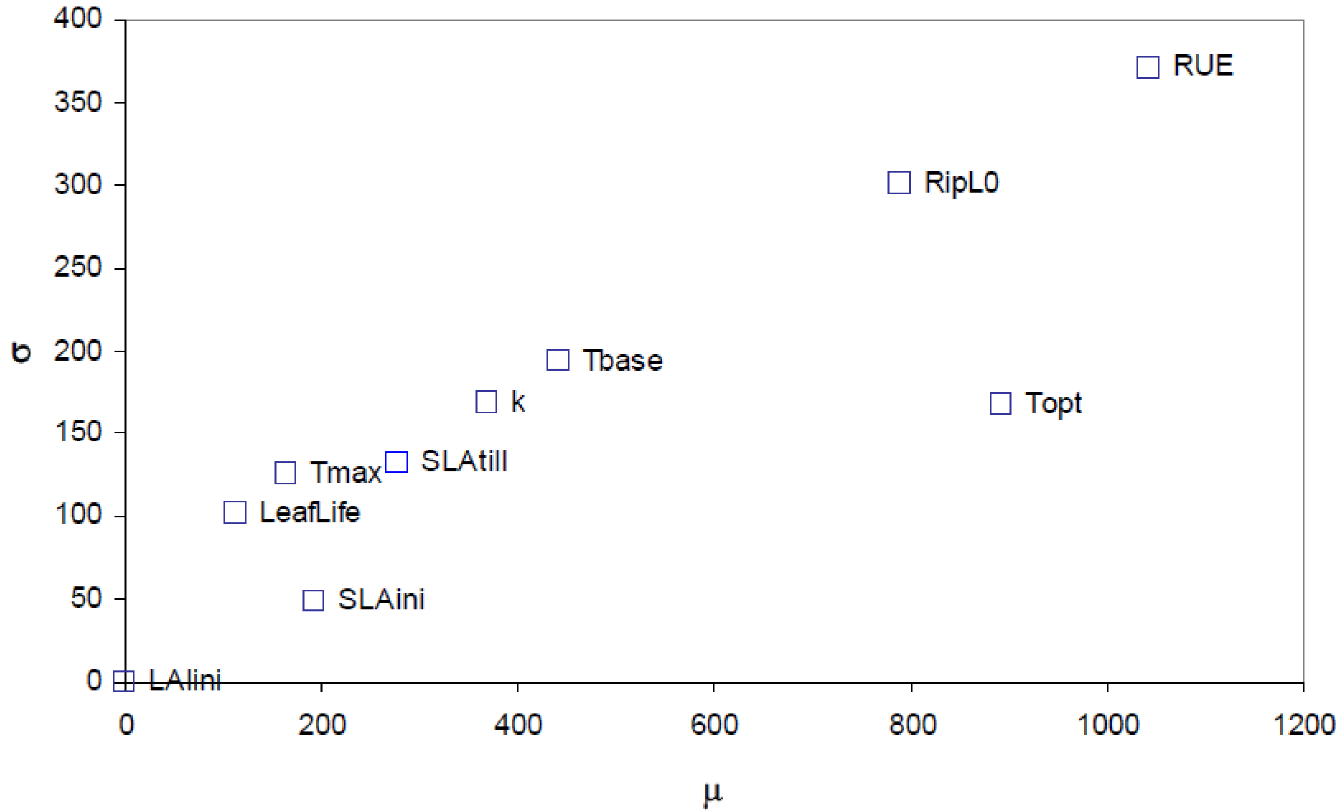
$$R_i(x_1, \dots, x_N, \Delta) = \frac{y(x_1, \dots, x_{i-1}, x_i + \Delta, \dots, x_{i+1}, \dots, x_N) - y(x_1, \dots, x_N)}{\Delta}$$

- $\boldsymbol{\Omega}$ is randomly **sampled** over $r$ **different trajectories**.
- **After the sampling**, parameters are **scaled back** to their biophysical values.
- The total number of model executions is $\boldsymbol{r(k+1)}$.
- **Mean** $(\mu_i)$ and **standard deviation** $(\sigma_i)$ of each distribution of $R_i$ are then calculated
  - ✓ $\mu_i$ represents the **overall influence** (total effect) of the parameter $x_i$
  - ✓ $\sigma_i$ identifies (for high values) **non-linearities or interactions** with other parameters.

# Sensitivity analysis

Development of crop ideotypes

- **Sample results**
  - ✓ WARM rice model, northern Italy
  - ✓ Output: aboveground biomass at maturity



# Sensitivity analysis

## Regression-based methods (LHS, Random, LpTau)

- The rationale is **approximating the relationship** between inputs ($x_i$) and output ($y$) through the following equation:

$$y = b_0 + \sum_{i=1}^{N} b_i \cdot x_i + \varepsilon$$

  where $b_i$ is the coefficient to be estimated for $x_i$ and $\varepsilon$ is the random error.

- **When the inputs $x_i$ are independent**, it is possible to use the **standardized regression coefficient** (SRC) to get an **estimate of the sensitivity index**:

$$SRC(x_i) = b_i \cdot \frac{\hat{s}_i}{\hat{s}}$$

  where $\hat{s}_i$ and $\hat{s}$ are the **standard deviations** of the input and of the output

# Sensitivity analysis



- Each **SRC** gives information about the effect of changing the value of an input from its standard value by a fixed fraction of its standard deviation, while **maintaining the other factors at their default** values.
- This family of methods also provides the **coefficient of determination** (**$R^2$**), indicating the portion of total variance explained by the regression model.
- **If the regression model is actually able** to explain the relationship between inputs and output, the larger the value of a **SRC**, the more **sensitive** the model to that input.
- Among the **sampling techniques** more popular for generating the combination of model inputs, a key role is played by **Latin Hypercube**, **Random** and **Quasi-Random LpTau**.

## Variance-based methods (Sobol', FAST, E-FAST)

- Variance-based methods use a **variance ratio** to estimate the **importance of inputs**.
- The rationale is the **partitioning of the total variance** of model **output $\boldsymbol{V(Y)}$** (analogous to ANOVA) using the following equation:

$$V(Y) = \sum_{i=1}^{N} D_i + \sum_{i \le j \le N}^{N} D_{ij} + \dots + \sum_{i \le \dots N}^{N} D_{i \dots N}$$

where:

✓ $D_i = V[E(Y/x_i)]$ is the **first order effect** for each input $x_i$

✓ values from $D_{ij} = V\left[E\left(Y/x_i, x_j\right)\right] - D_i - D_j$ to $D_{1 \dots N}$ are the **second to $\boldsymbol{N}$° order** effects (**interactions**).

# Sensitivity analysis

- The sensitivity index for the **first order effect** (effect of the single input, no interactions) of the input $x_i$ is:

$$S_i = \frac{V[E(Y/x_i)]}{V(Y)}$$

- According to **most implementations**, the **number of iterations** (decidedly larger than in other methods) is kept to a "reasonable" value by calculating only **first and total order** (St) effects (no explicit estimates of second, third, … orders).
- All the effects for **orders higher than first**

$$St_i = \sum S_i + \sum_{j \neq i} S_{ij} + \dots + S_{1 \dots N}$$

are **estimated together** (total – first order).

- **Sobol'** is based on the **Monte Carlo** integration method, **FAST and E-FAST** use the **Fourier series** (approximation but more efficient).

# Sensitivity analysis



- **Sample results**
  - ✓ WOFOST model
  - ✓ Rice in Northern Italy
  - ✓ Output: yield