UNCERTAINTY AND SENSITIVITY ANALYSIS IN QUANTITATIVE PEST RISK ASSESSMENTS

APPLICATION TO THE MODEL OF MAGAREY ET AL.

David Makowski
INRA

Paris, 2015

Sources of uncertainty in a model

$$G(x,\theta)$$

Equations — Input variables — Parameters

Types of uncertainty

- ***Lack of knowledge***
- ***Measurement error***
- ***Variability of the system characteristics***

Uncertainty and sensitivity analysis; two techniques with two different objectives

- Objective of an **uncertainty analysis**:

  to study the consequence of uncertainty by computing a probability distribution on model output from the set of probability distributions on model inputs.

- Objective of a **sensitivity analysis**:

  to rank of uncertain inputs according to their influence on the output

Practical interest

of uncertainty analysis

- Give information about the uncertainty associated with model prediction

- Optimize decision variables

of sensitivity analysis

- Identify the parameters and input variables which strongly influence the model outputs

→ ***Important to know them accurately***

- Identify the parameters and input variables which do not strongly influence the model outputs

→ ***Less important to know them accurately***

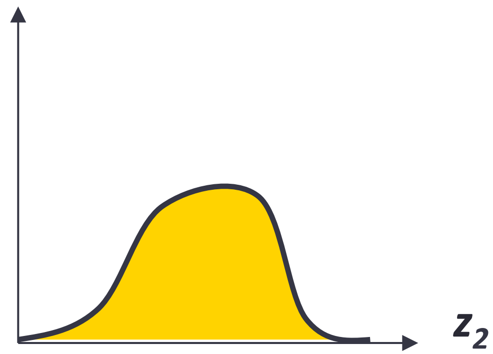
Uncertainty analysis

Its purpose is to answer the following question:

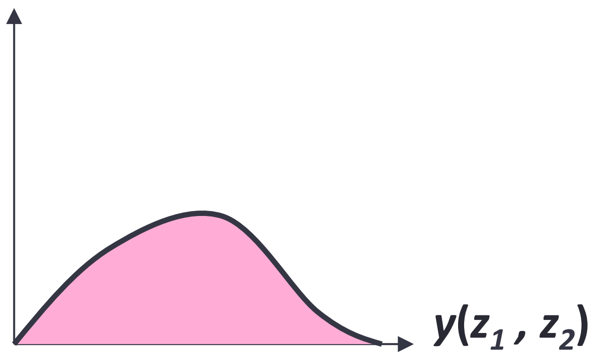
« What is the uncertainty about the *output* resulting from the uncertainty about the *inputs*? »

We have

and



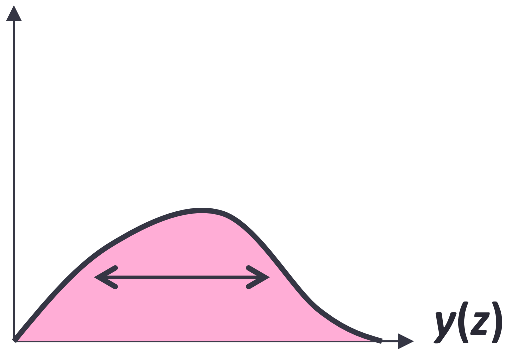
We want to compute



Sensitivity analysis

Its purpose is to answer the question:

« What are the most important uncertain inputs? »



Variance of $y(z)$ = effect of z_1 + effect of z_2 + …

Ten Most Important Accomplishments in Risk Analysis, 1980–2010

Michael Greenberg, Charles Haas, Anthony Cox, Jr., Karen Lowrie, Katherine McComas, and Warner North

As part of the celebration of the 30th anniversary of the Society for Risk Analysis and *Risk Analysis, An International Journal*, a group of your editors engaged in a process to select the 10 most important accomplishments in risk analysis. The article that follows is the product of this process.

Some preliminary decisions were that we would reach out to the full membership for nominations, focus on the period 1980 to 2010, and accept nominations for contributions to theory, methods, and applications. Also, we focused on accomplishments that address health, safety, and the environment, which has been our tradition.(1) All the accomplishments have contributed to answering at least one of the six following risk analysis questions:(2–5)

1. What can go wrong?
2. What are the chances that something with serious consequences will go wrong?
3. What are the consequences if something does go wrong?

TEN MOST IMPORTANT ACCOMPLISHMENTS IN RISK ANALYSIS, 1980–2010

Theory

1. Understanding how affect and trust influence risk perception and behavior
2. Recognizing that personal decisions reflect different processes for valuing and combining anticipated and actual losses, gains, delays, and surprises.
3. Developing an environmental justice ethic and frameworks

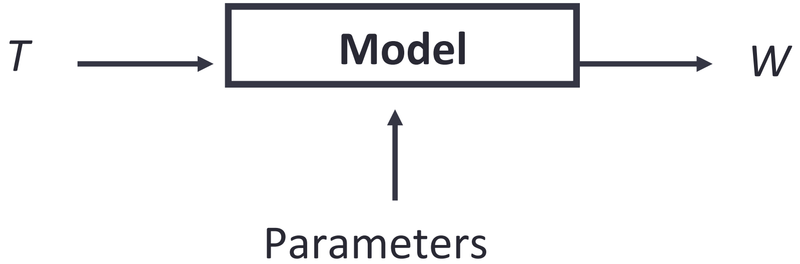
Methods

4. Using formal uncertainty analysis in risk assessment

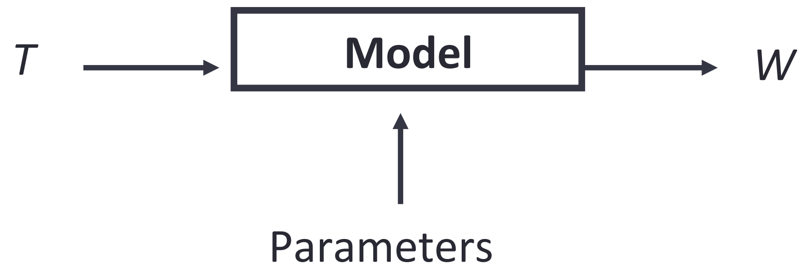
Warning

- Several factors may influence the results of uncertainty and sensitivity analysis
- Validity of conclusion may be limited
- Important to follow some practical rules

Example: Wetness model (Magarey et al., 2005)



W = leaf wetness duration requirement for successful fungal infection (h
T = average temperature (°C)



$$W = \frac{W_{\min}}{f(T)}, \text{ and } W \le W_{\max}$$

$$f(T) = \left(\frac{T_{\max} - T}{T_{\max} - T_{opt}}\right)\left(\frac{T - T_{\min}}{T_{opt} - T_{\min}}\right)^{(T_{opt} - T_{\min})/(T_{\max} - T_{opt})}$$

Five parameters: T_{min}, T_{opt}, T_{max}, W_{min}, W_{max}

TABLE 2. Infection model parameters and statistical comparison between model predictions and observations based on published studies relating fungal infection to temperature and wetness duration

Pathogen	Host[a]	Ref.[b]	Ref. T_{min}[c]	T_{min}[d]	T_{max}[e]	T_{opt}[f]	W_{min}[g]	W_{max}[h]	Obs[i]	r[j]	RMS[k]	SRMS[l]
Albugo occidentalis	Spinach	81	…	6	28	16	3	12	12	0.87	2.8	0.9
Alternaria brassicae	Oilseed rape	38	6	2.6	35	18	6	22	9	0.96	4.0	0.7
Alternaria cucumerina	Muskmelon	31	…	12	25	19	8	24	6	0.98	1.6	0.2
Alternaria mali	Apple	32	…	1	35	23	5	40	16	0.88	5.2	1.0
Alternaria porri	Onion	80	…	1	35	23	8	24	5	1.00	0.7	0.1
Alternaria sp.	Mineola tangelo	18	…	9.4	35	25	8	16	5	0.90	1.3	0.2
Ascochyta rabiei	Chick pea	84	…	1	35	25	12	48	6	0.10	19.2	1.6
Bipolaris oryzae	Rice	59	25	8	35	27.5	10	24	6	0.78	5.0	0.5
Botryosphaeria dothidea	Apple fruit	58	…	8	35	28	8	19	6	0.95	1.6	0.2
Botryosphaeria obtuse	Apple fruit	7	…	1	35	26	5	40	7	0.97	3.2	0.6
Botrytis cinerea	Grape	56	57	10	35	20	4	10	11	0.94	0.8	0.2
Botrytis cinerea	Strawberry flower	15	…	5	35	25	8	18	7	0.13	5.0	0.6
Botrytis cinerea	Grape flower	56	57	1	34	25	1	12	6	0.99	0.6	0.6
Botrytis squamosa	Onion	82	…	1	28	18	15	24	8	0.50	4.7	0.3
Bremia lactucae	Lettuce	67	…	1	25	15	4	10	6	0.98	0.8	0.2
Cercospora arachidicola	Peanut	93	6	13.3	35	24	24	48	5	0.72	8.9	0.4
Cercospora carotae	Carrot	20	…	11	32	24	28	96	5	0.98	16.5	0.6
Cercosporidium personatum	Peanut	17	6	8	35	20	16	33	6	0.33	6.0	0.4
Coccomyces hiemalis	*Prunus* sp.	28	29	4	30	18	5	30	11	0.96	7.8	1.6
Colletotrichum acutatum	Strawberry fruit	92	…	7	35	27.5	6	36	6	0.93	4.4	0.7
Colletotrichum orbiculare	Watermelon	53	…	7	30	24	2	16	7	0.69	5.6	2.8

Magarey et al. (2005)

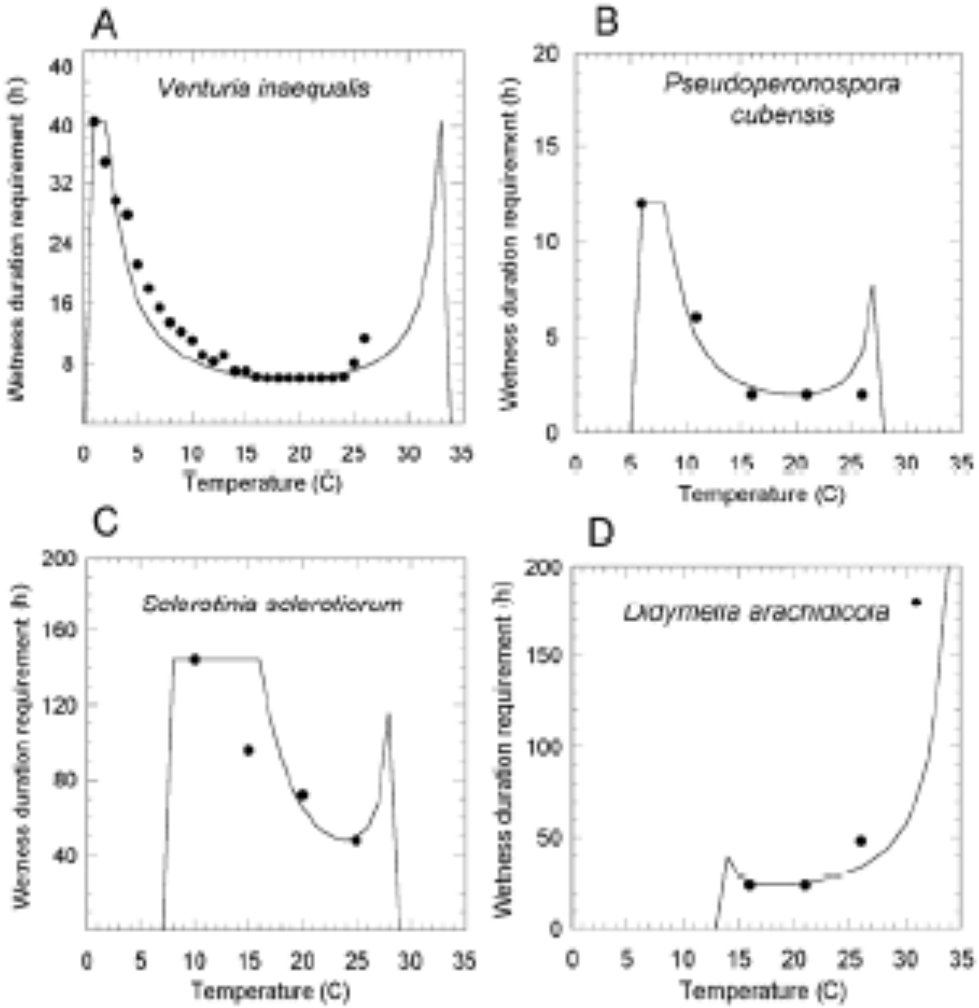
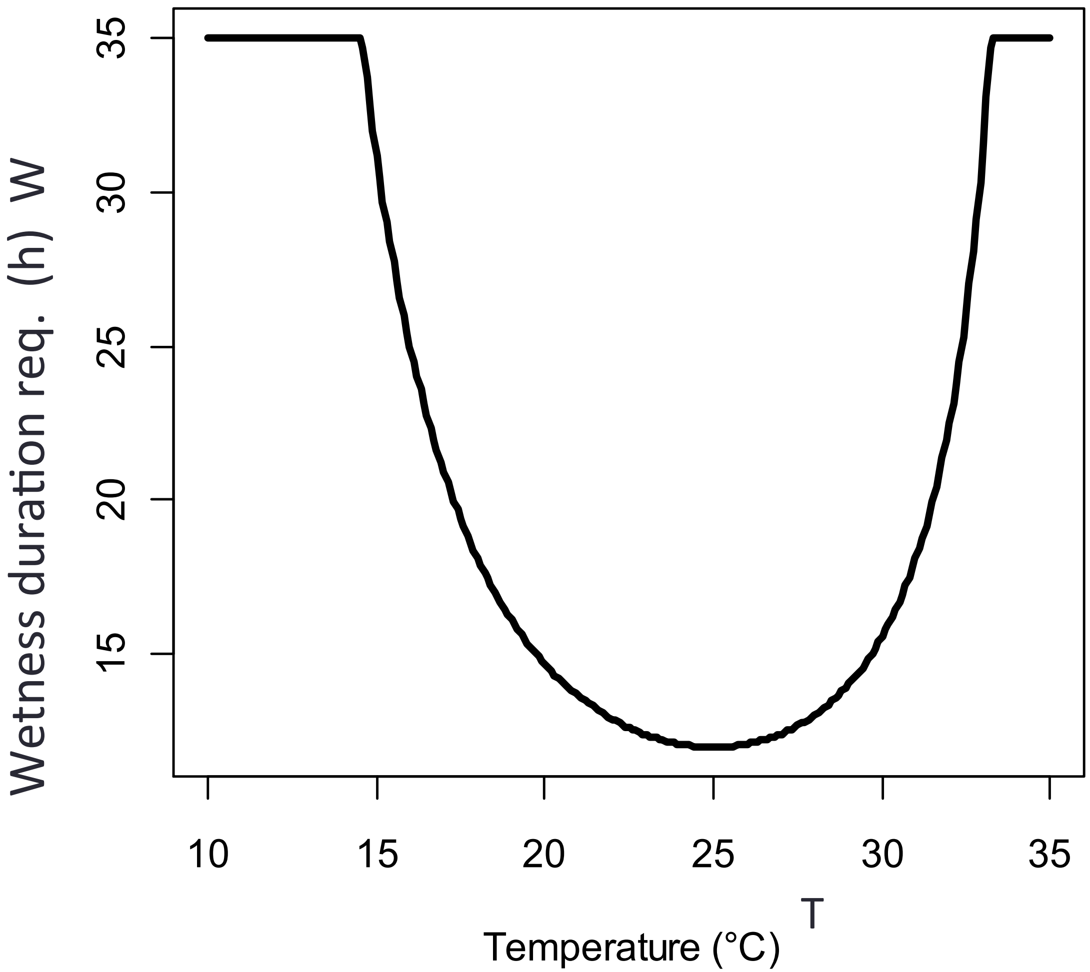


Fig. 1. Examples of goodness of fit of model predictions (solid lines) of wetness requirements at different temperatures compared with experimental observations (solid circles) for **A,** *Venturia inaequalis* (78), **B,** *Pseudoperonospora cubensis* (23), **C,** *Sclerotinia sclerotiorum* (91), and **D,** *Didymella arachidicola* (79). In each case, model predictions were produced with equations 1 and 2 (described in text) and the parameters in Table 2.

Magarey et al. (2005)

Parameter values estimated for pycnidiospores of *Guignardia citricarpa* Kiely , and associated response of *W* to temperature (from EFSA, 2008)

T_{min}= 10 °C, T_{opt}= 25 °C, T_{max}=35 °C, W_{min}=12 h, W_{max}= 35 h



Scientific name:
Guignardia citricarpa
Order: Dothideales, Family: Botryosphaeriaceae
Common Name: Citrus black spot *Guignardia citricarpa* Kiely

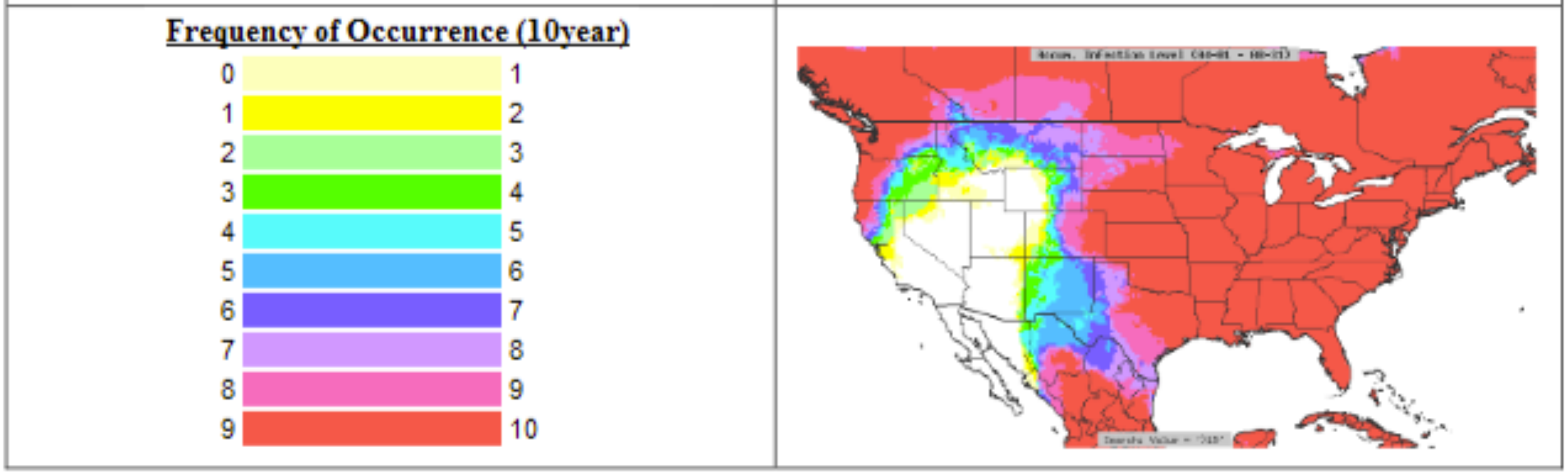


Figure 3B. The probability of more than 15 days suitable for *Guignardia citricapia* pycnidiopsoric infection by continent. (Legend lower left).

Uncertainty about the model parameters *Guignardia citricarpa* Kiely

		Min	Max
T_{min}	(°C)	10	15
T_{max}	(°C)	32	35
T_{opt}	(°C)	25	30
W_{min}	(h)	12	14
W_{max}	(h)	35	48

Panel on Plant Health, from EFSA (2008)

Uncertainty analysis

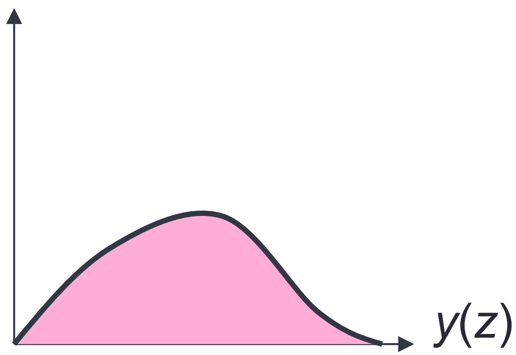
Its purpose is to answer the following question:

« What is the uncertainty about *y*(*z*) resulting from the uncertainty about *z* ? »

We have

and

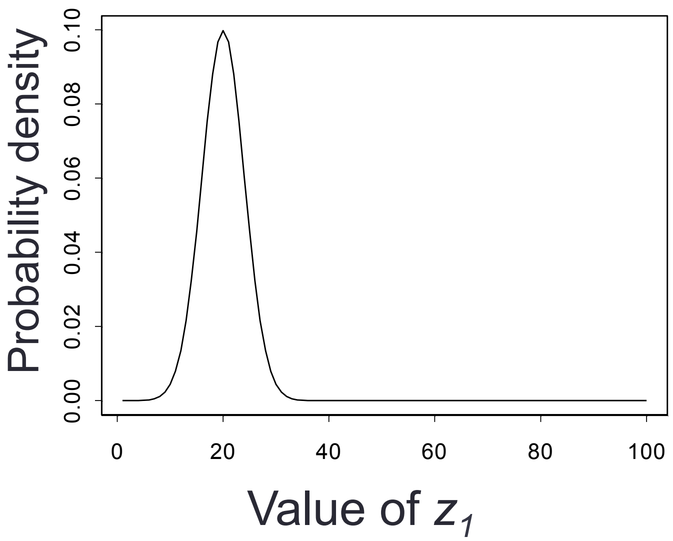
We want to compute

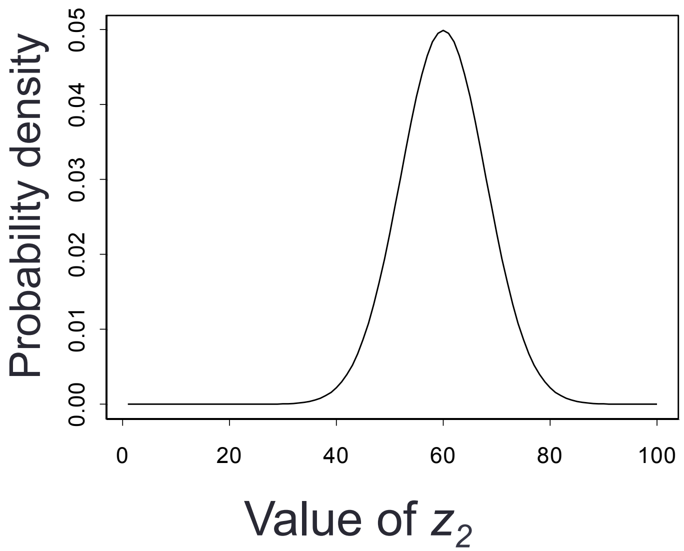


Application to a very simple model

Equation: $y(z_1, z_2) = z_1 + 2\, z_2$

Uncertainty about z_1 and z_2: $z_1 \sim N(20, 16)$ and $z_2 \sim N(60, 64)$





Question: Perform an uncertainty analysis

Application to a very simple model

« You need to determine the probability distribution of $y(z_1, z_2)$ from the distributions of z_1 and z_2 » .

Properties:

If z_1 and z_2 are two independant variables with Gaussian distributions then

$A z_1 + B z_2$ follows a Gaussian distribution

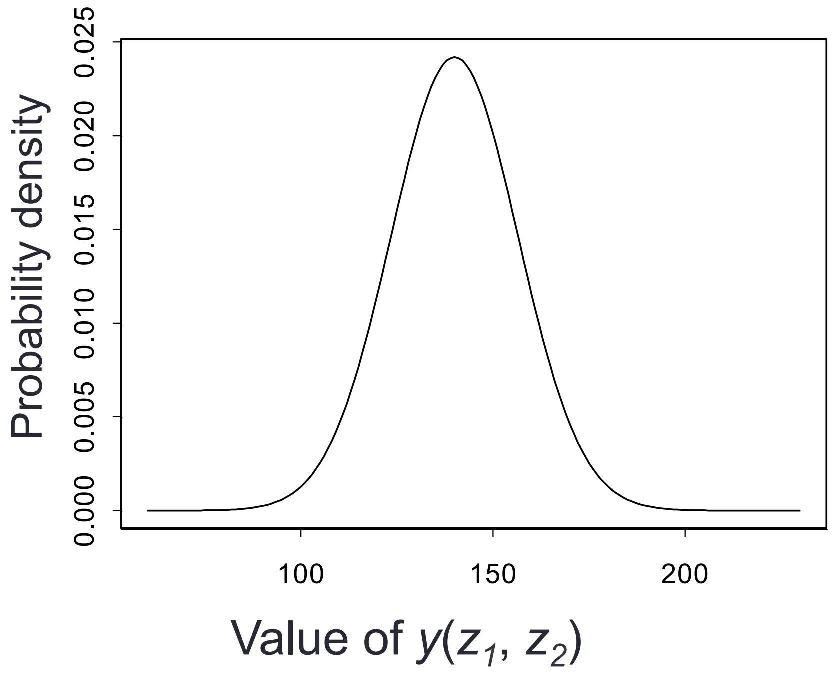
$E(A z_1+B z_2)=A E(z_1)+B E(z_2)$

$var(A z_1+B z_2)=A^2 var(z_1)+B^2 var(z_2)$

Application to a very simple model

For this simple model, it is possible to determine the exact expression of the distribution of $y(z_1,z_2)$:

$$y(z_1,z_2) \sim N(140, 272)$$



In general, it is more difficult

- **More complex equations, non linear relationship between *y*(*z*) and *z***

→The analytical expression of the distribution of *y*(*z*) cannot be determined

- **The distribution of *z* is not always well known**

→ Subjective choice

- **Computation times can be long with some models**

→ The number of simulations is limited

A four step approach based on Monte Carlo simulation

1. Define the distributions of $z_1, \ldots, z_p$.

2. Generate samples from the distributions defined in step 1

3. Compute $y(z)$ for each generated set $z_1, \ldots, z_p$

4. Describe/Approximate the distribution of $y(z)$

A four step approach based on Monte Carlo simulation

1. **Define the probability distributions of T_{min} , T_{opt} , T_{max} , W_{min} , W_{max}**

2. **Generate N values from the distributions defined in step 1**

3. **Compute W for each generated set T_{min} , T_{opt} , T_{max}, W_{min} , W_{max}**

4. **Describe/Approximate the distribution of W**

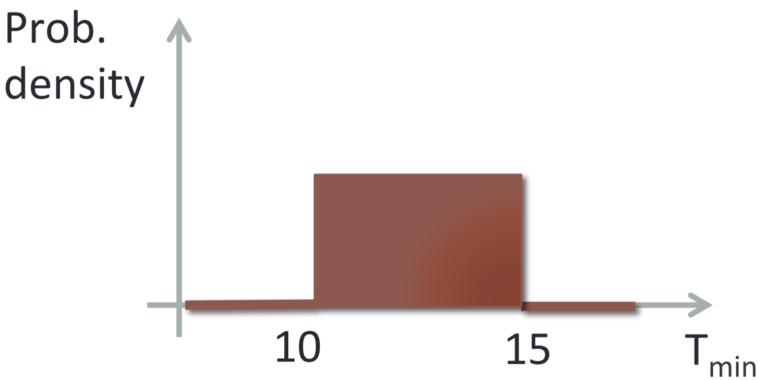
Step 1 – Definition of distributions for the uncertain inputs

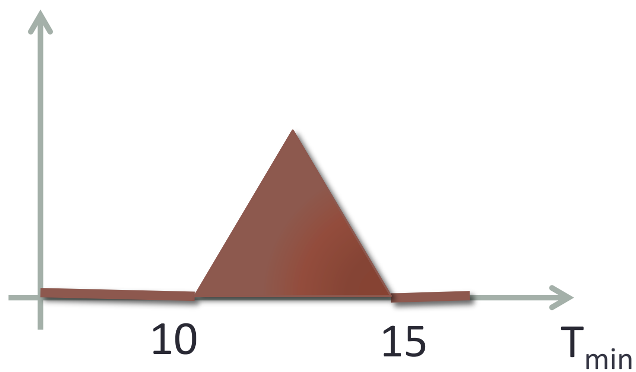
- This step is not straightforward
- Several choices possible for T_{min} , T_{opt} , T_{max} , W_{min} , W_{max}

Step 1 – Definition of distributions for the uncertain inputs

Ex: Three types of distribution are considered here

- ✓ Independent Uniform distributions
- ✓ Independent Triangle distributions
- ✓ Non-independent Triangle distributions (T_{min} and T_{opt} positively correlated)





Step 1 – Definition of distributions for the uncertain inputs

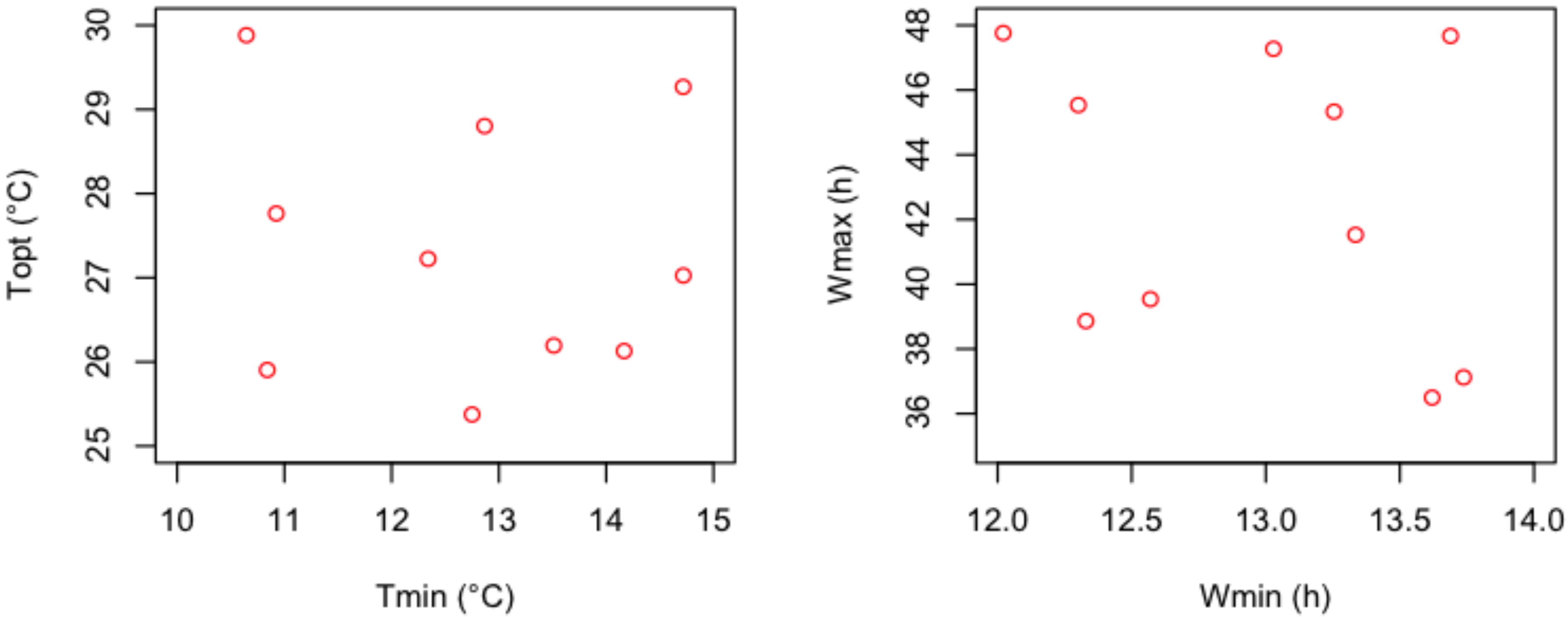
Rule 1: « Be transparent about assumptions »

Step 2 – Random generation of *N* values from the chosen distributions

- The choice of *N* is critical
- Reliability of the conclusion depends on *N*
- Computation time can be a limiting factor

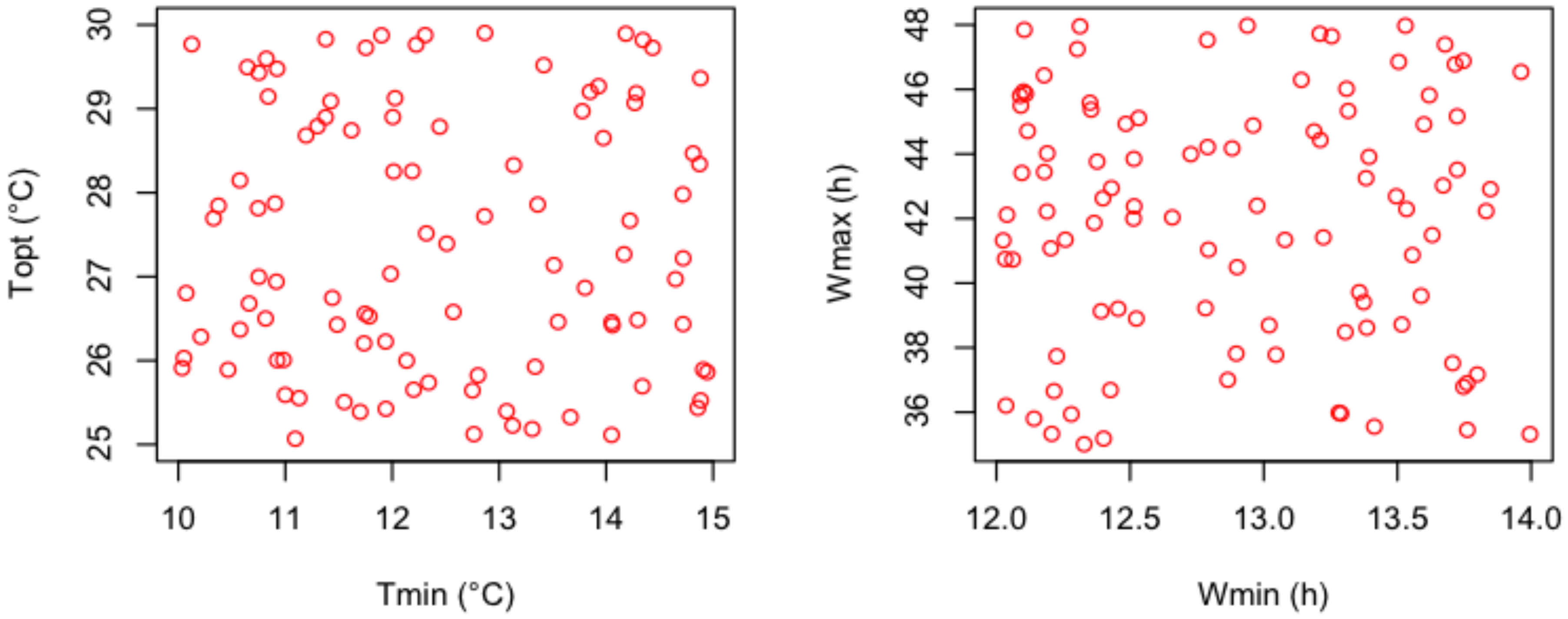
Step 2

N=10 (independent uniform)



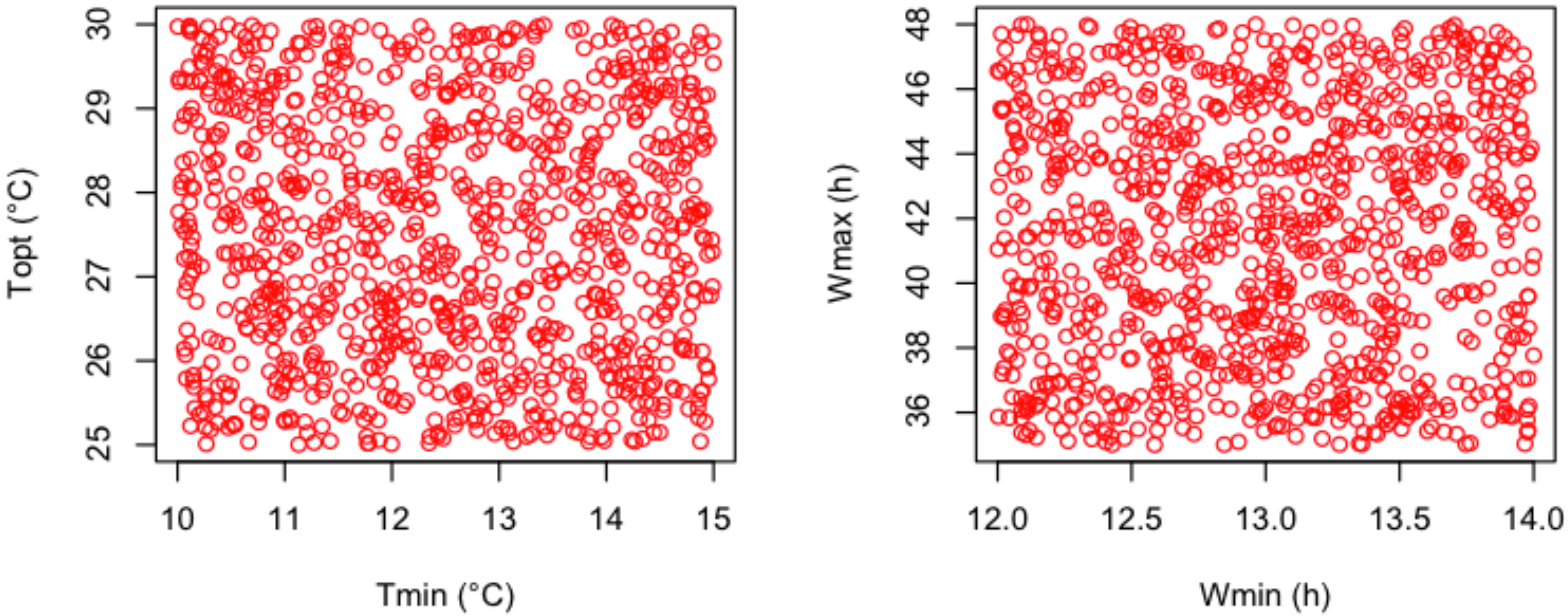
Step 2

N=100 (independent uniform)

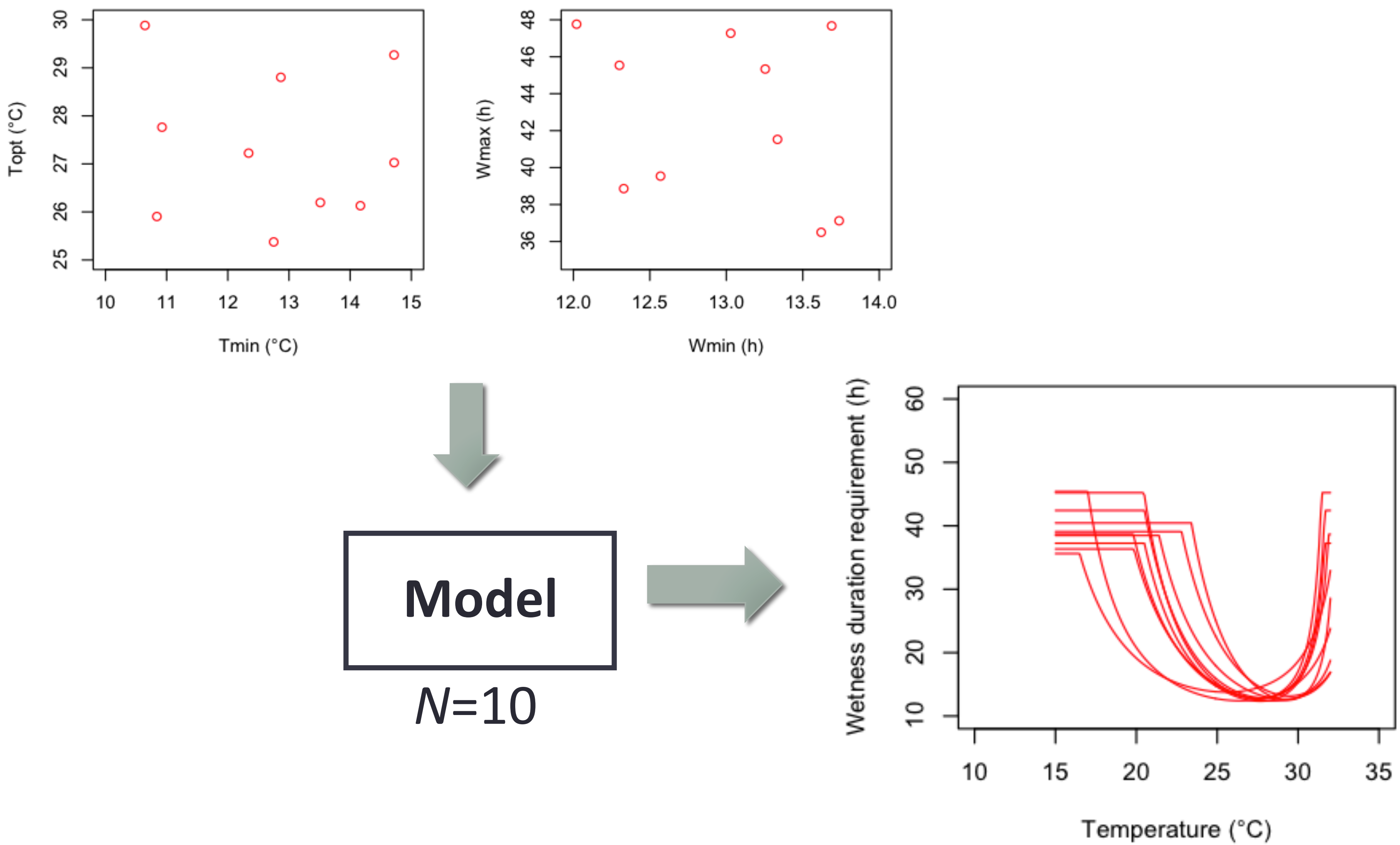


Step 2

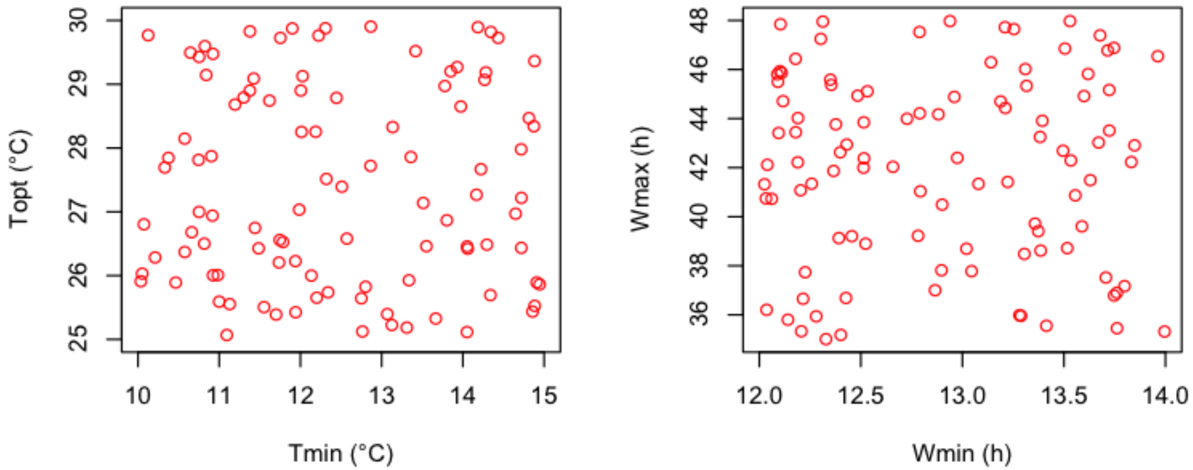
N=1,000 (independent uniform)



Step 3 – Computation of the model output *W* for each generated parameter set

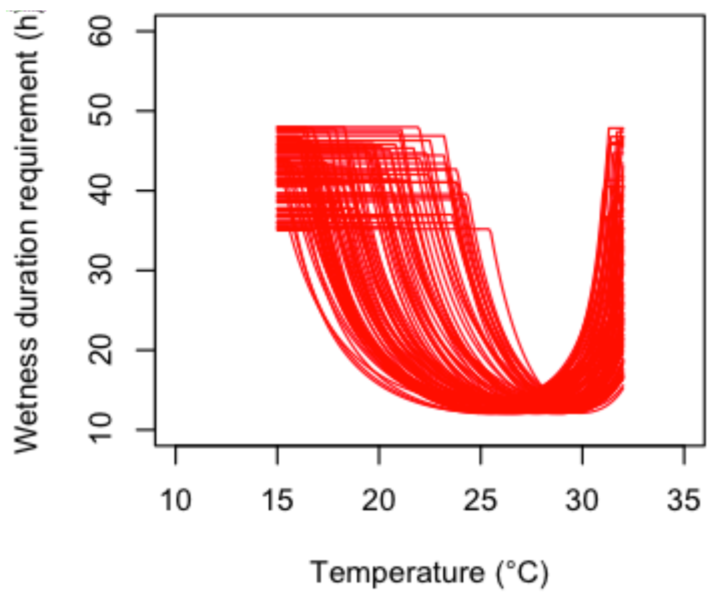


Step 3

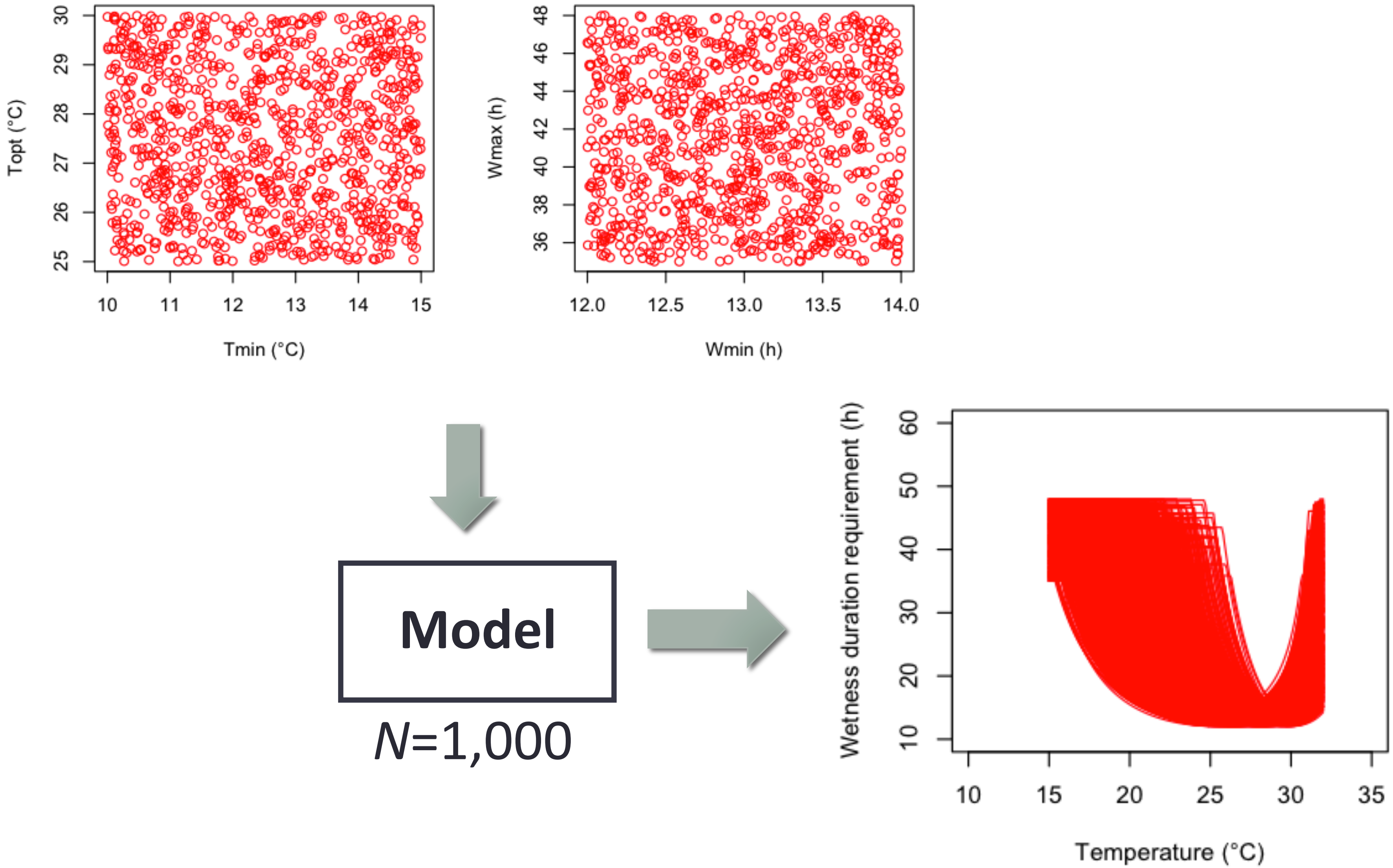


Model

N=100



Step 3



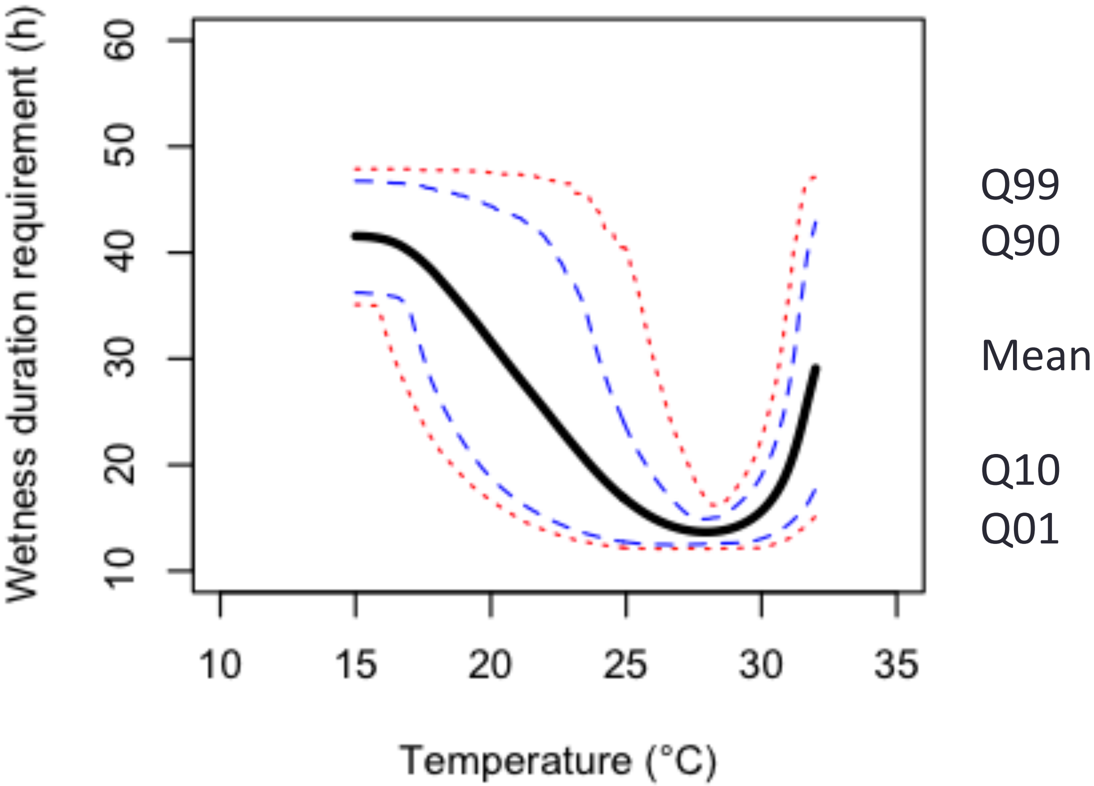
Topt (°C)
30
29
28
27
26
25
10
11
12
13
14
15
Tmin (°C)
Wmax (h)
48
46
44
42
40
38
36
12.0
12.5
13.0
13.5
14.0
Wmin (h)
Model
N=1,000
Wetness duration requirement (h)
60
50
40
30
20
10
10
15
20
25
30
35
Temperature (°C)

Rule 2: « Study the effect of increasing the number of simulations »

Step 4 - Describe/Approximate the distribution of the model output

- Usually, several model outputs can be analysed
- Summary statistics can be calculated for each output of interest (e.g., mean, median, percentiles, standard deviation, correlation)
- Graphics are usually useful (e.g., scatterplot, boxplot, contour plot)

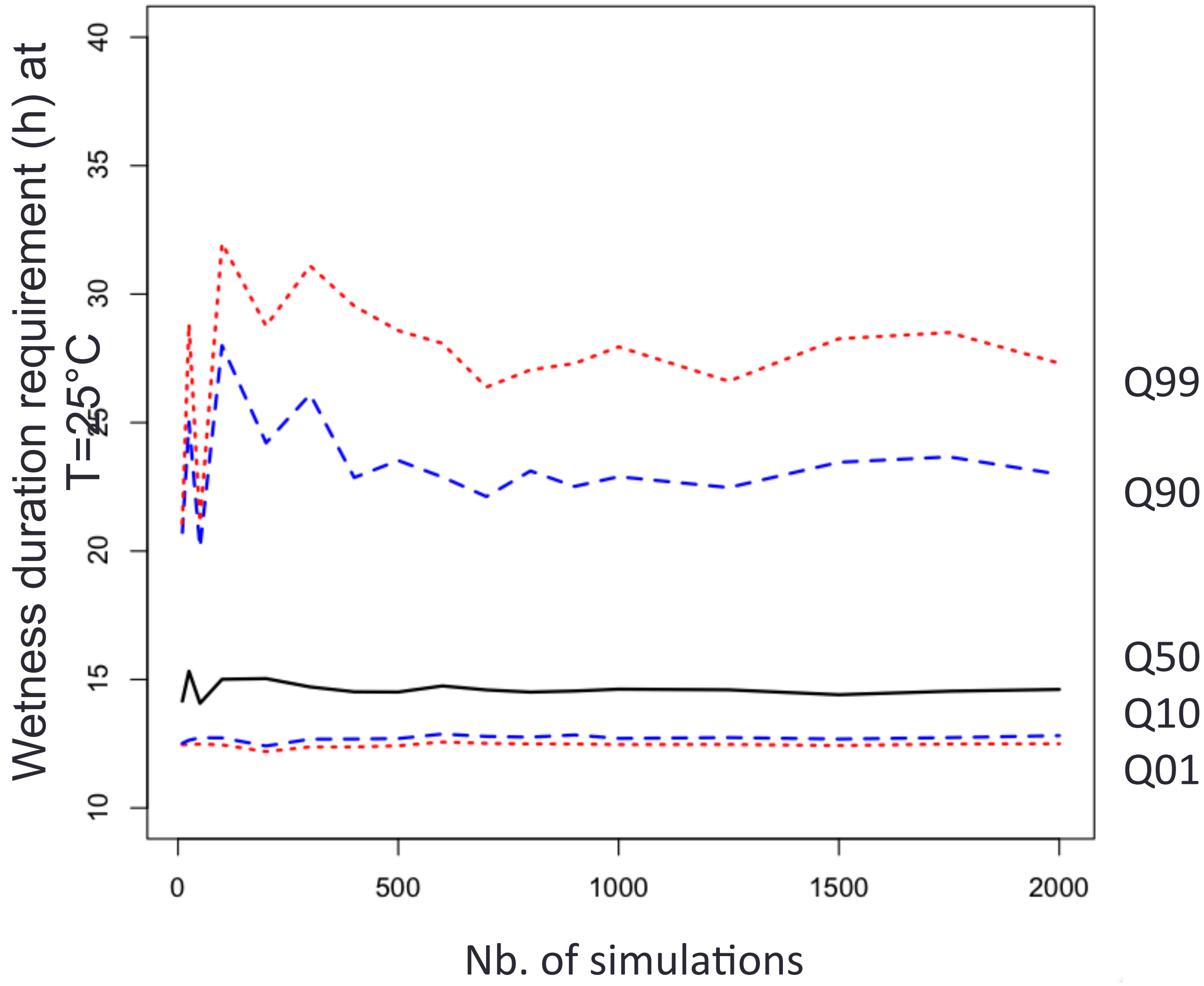
Step 4 - Describe/Approximate the distribution of the model output



N=1,000

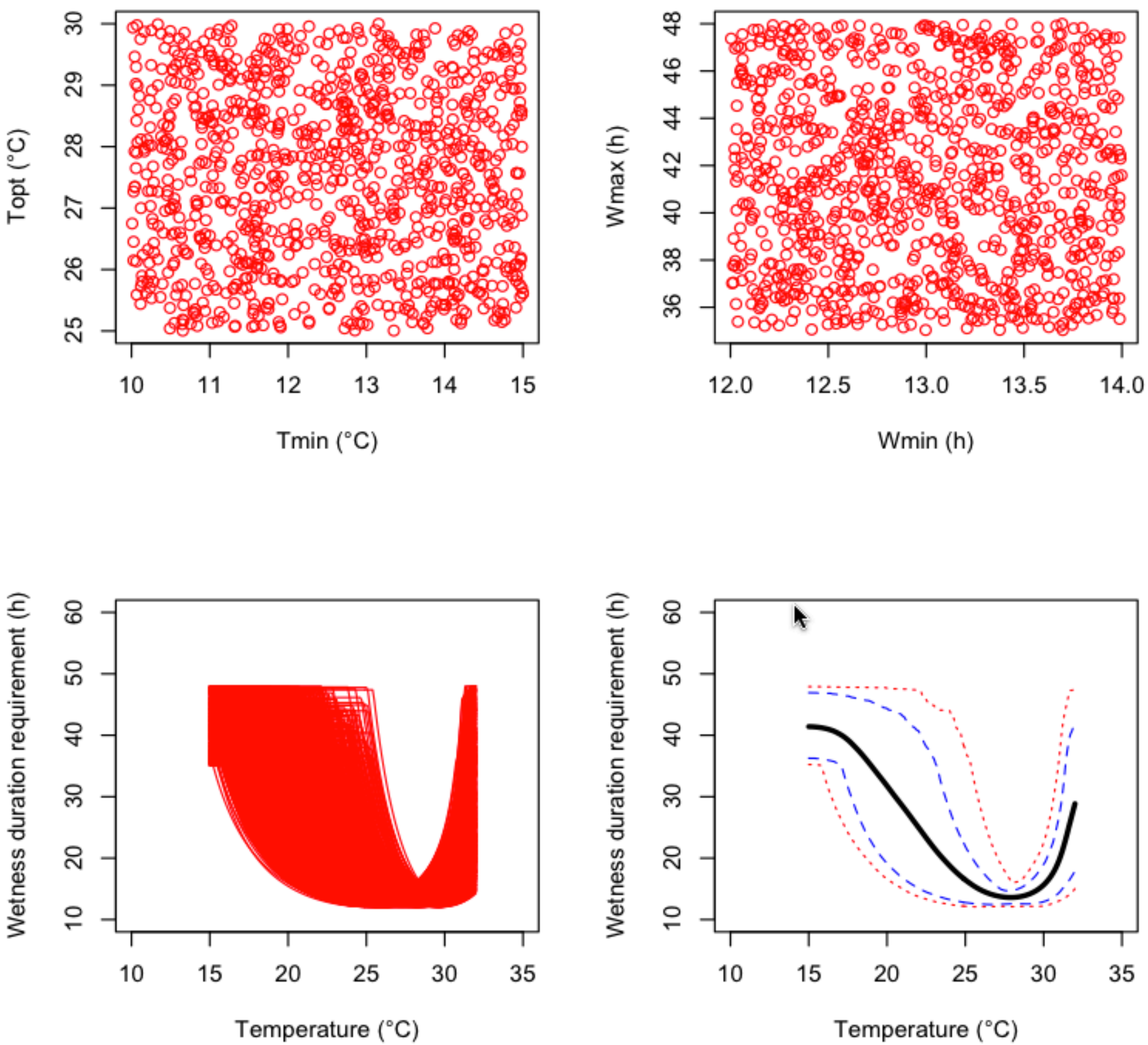
Rule 3: « Define precisely the model outputs of interest, and explore different aspects of the distribution of these outputs »

Estimated extreme quantiles may be unstable!



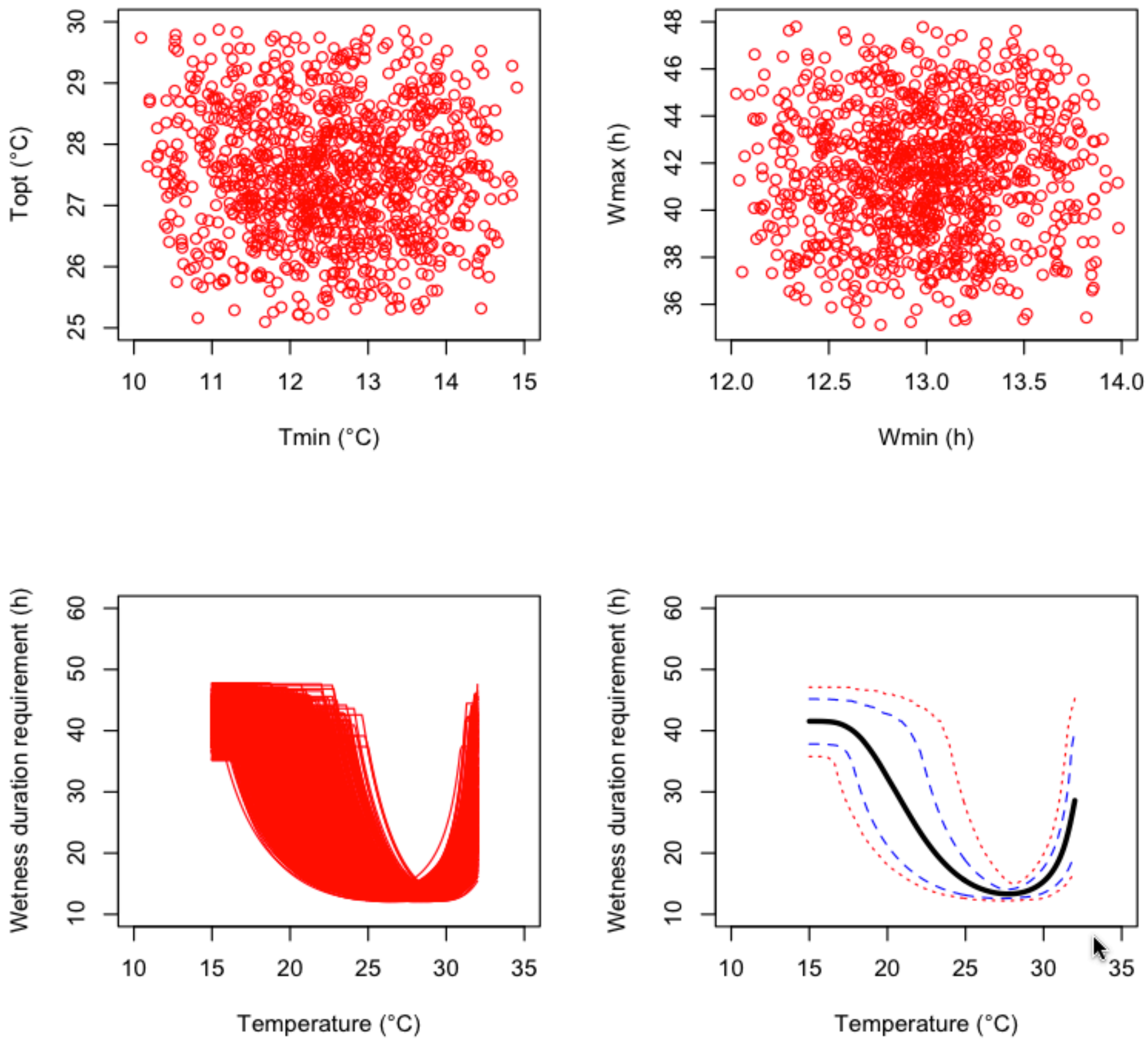
Results may depend on distribution assumptions!

Uniform and independent



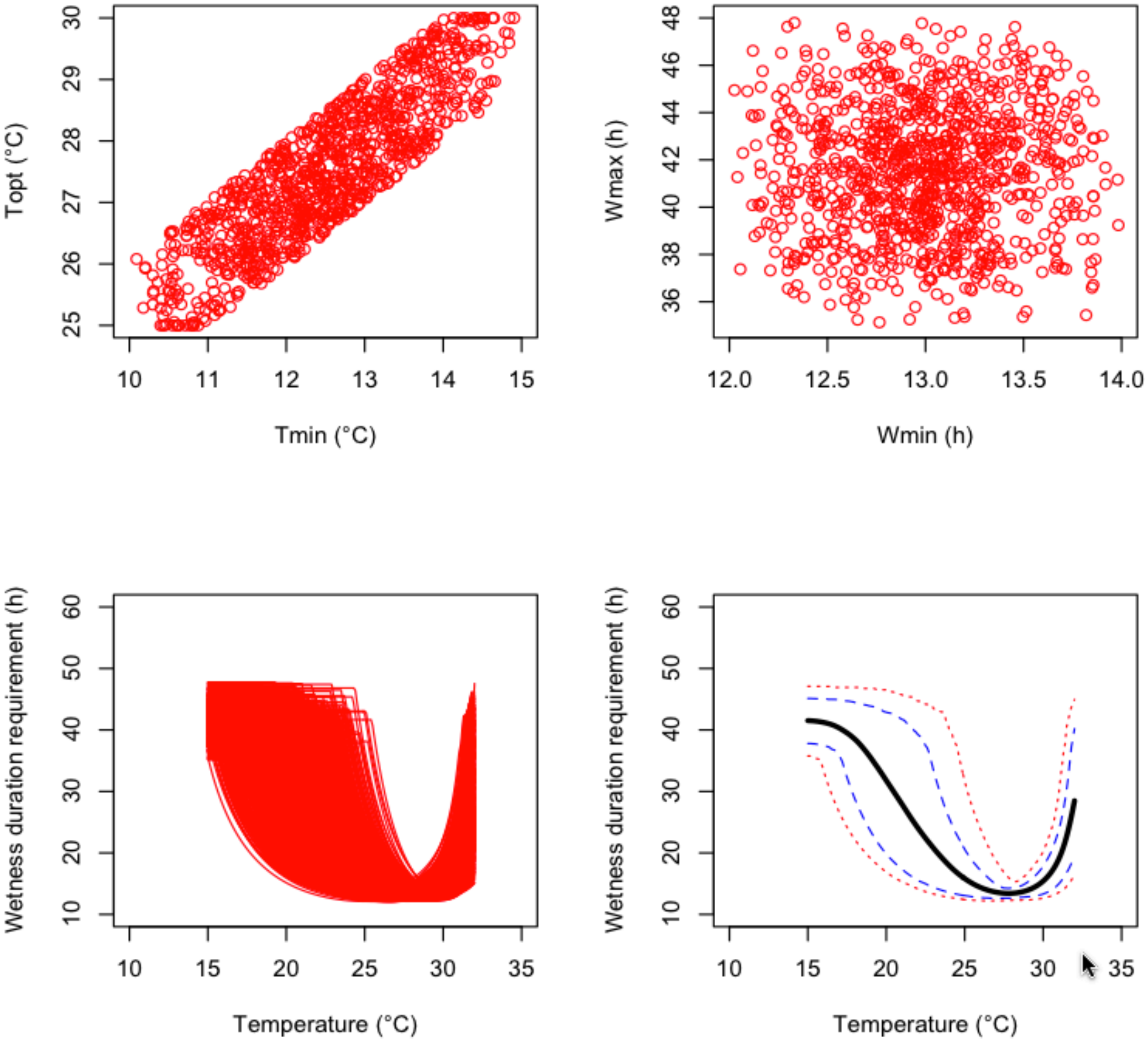
Results may depend on distribution assumptions!

Triangle and independent



Results may depend on distribution assumptions!

Triangle with correlation between T_{opt} and T_{min}



Percentiles of *W* obtained for T=25°C with different input distributions (*N*=10,000)

with independent uniform distributions

50%	95%	99%
14.52	27.78	39.61 h

with independent triangle distributions

50%	95%	99%
14.51	20.82	26.20 h

with non-independent triangle distributions

50%	95%	99%
14.44	23.35	32.38 h

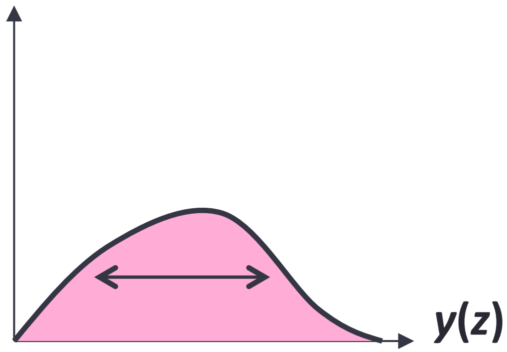
Rule 4: « Assess the robustness of the results to

- the number of simulations
- distribution assumptions »

Sensitivity analysis

Its purpose is to answer the question:

« What are the most important uncertain inputs? »

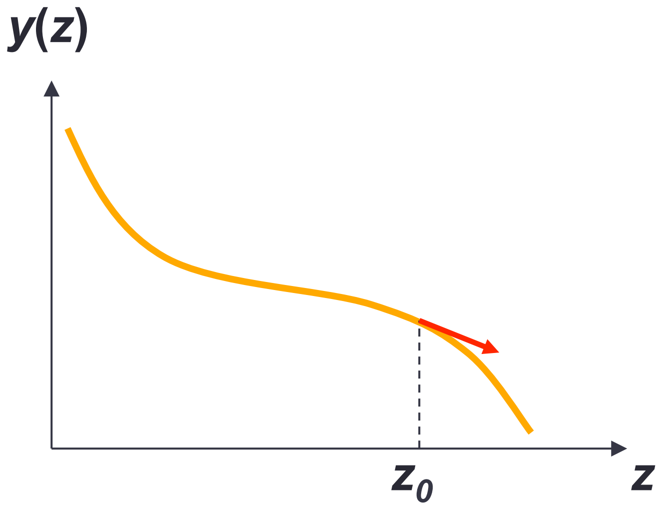


Variance of $y(z)$ = effect of z_1 + effect of z_2 + …

Local sensitivity analysis or Global sensitivity analysis ?

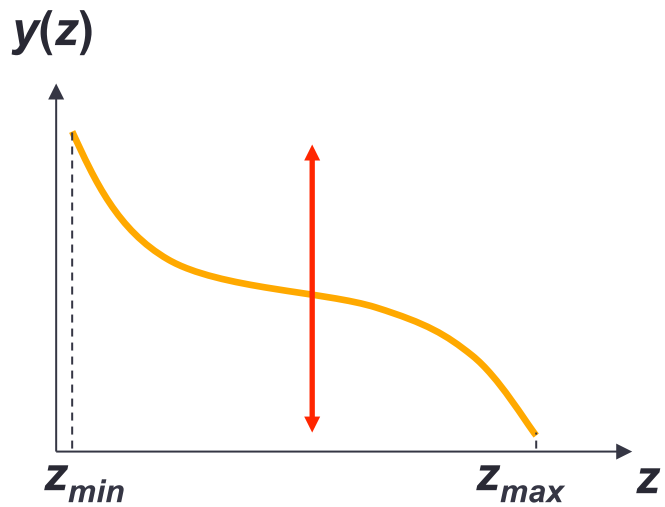
Local SA

variation of $y(z)$ « around » z_0



Global SA

Global variation of $y(z)$ when z varies within its uncertainty range



Practical interest of sensitivity analysis

i) Identify the parameters and input variables which strongly influence the model outputs

→ ***Important to know them accurately***

ii) Identify the parameters and input variables which do not strongly influence the model outputs

→***Less important to know them accurately***

iii) Analyze the behaviour of the model at some points

Local sensitivity analysis

Based on the computation of derivatives

Global sensitivity analysis

It consists in

- Defining sensitivity indices
- Compute the indices by varying the uncertain factors $z_1, \ldots, z_p$ over their ranges

Sensitivity indices

Many methods are available

- ✓ One-at-a-time
- ✓ Morris method
- ✓ Correlations between inputs and outputs
- ✓ Variance-based methods
  - Factorial design + anova
  - Sobol method
  - FAST (Fourrier)

Software a

- ✓ R (packages sensitivity & mtk)
- ✓ @risk, Crystal Ball etc.

Rule 5: « Be aware of the capabilities of different sensitivity analysis techniques and, when possible, compare results »

Method of Morris

- Define a design by combining *k* values of the *p* uncertain factors

- Choose an element of this design
- Add a « jump » Δ_{ij} to the i^th uncertain factor
- Compute the elementary effect

$$d_{ij} = \frac{\left[y\left(z_1, \ldots, z_{i-1}, z_i + \Delta_{ij}, \ldots, z_p\right) - y\left(z\right) \right]}{\Delta_{ij}}$$

- Repeat this procedure for all uncertain factor (i=1, …, p)

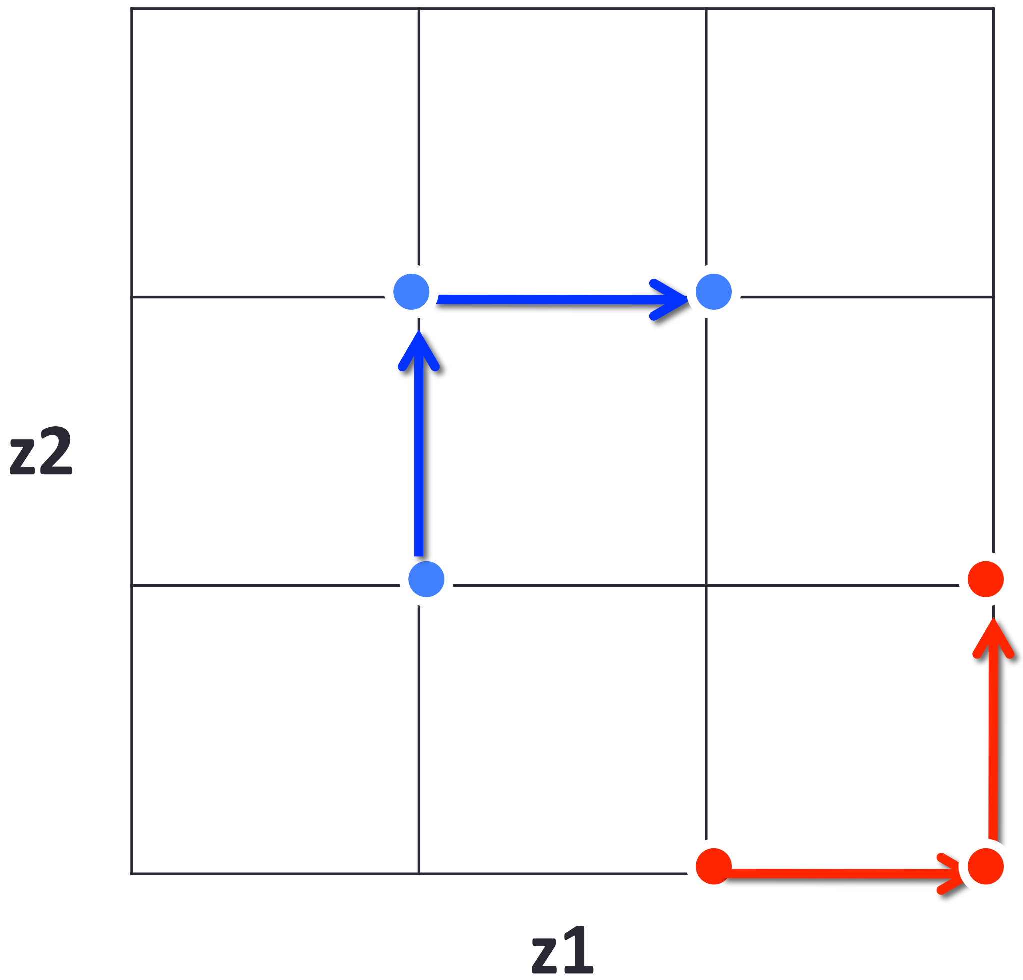
- Repeat *r* times all steps (j =1, …, r)

- Compute mean and variance of elementary effects from *r* replicates

$$\mu_i = \frac{\sum_{j=1}^{r} d_{ij}}{r} \qquad \sigma_i = \sqrt{\sum_{j=1}^{r} \left(d_{ij} - \mu_i\right)^2 / r}$$

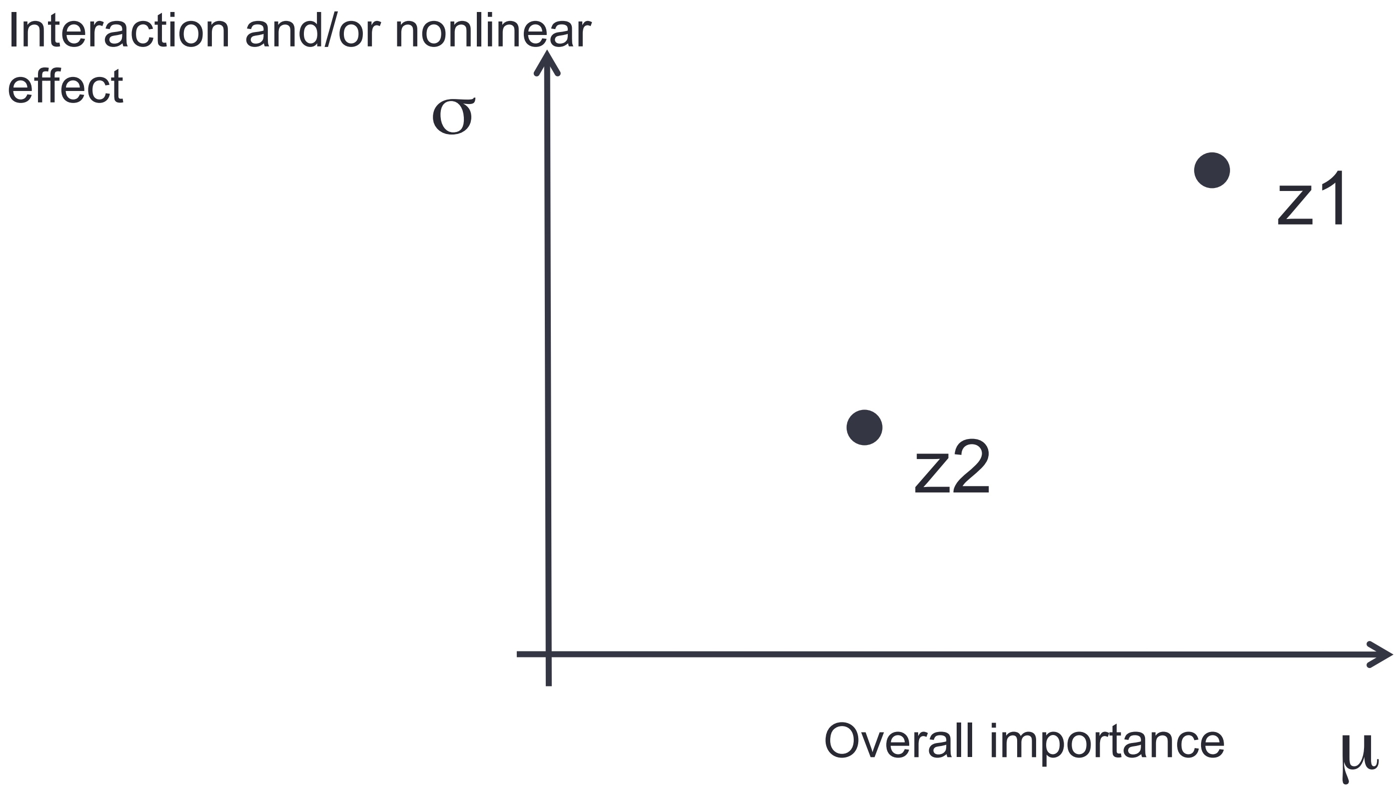
Method of Morris

Two examples of trajectories (p=2, k=3, r=2)

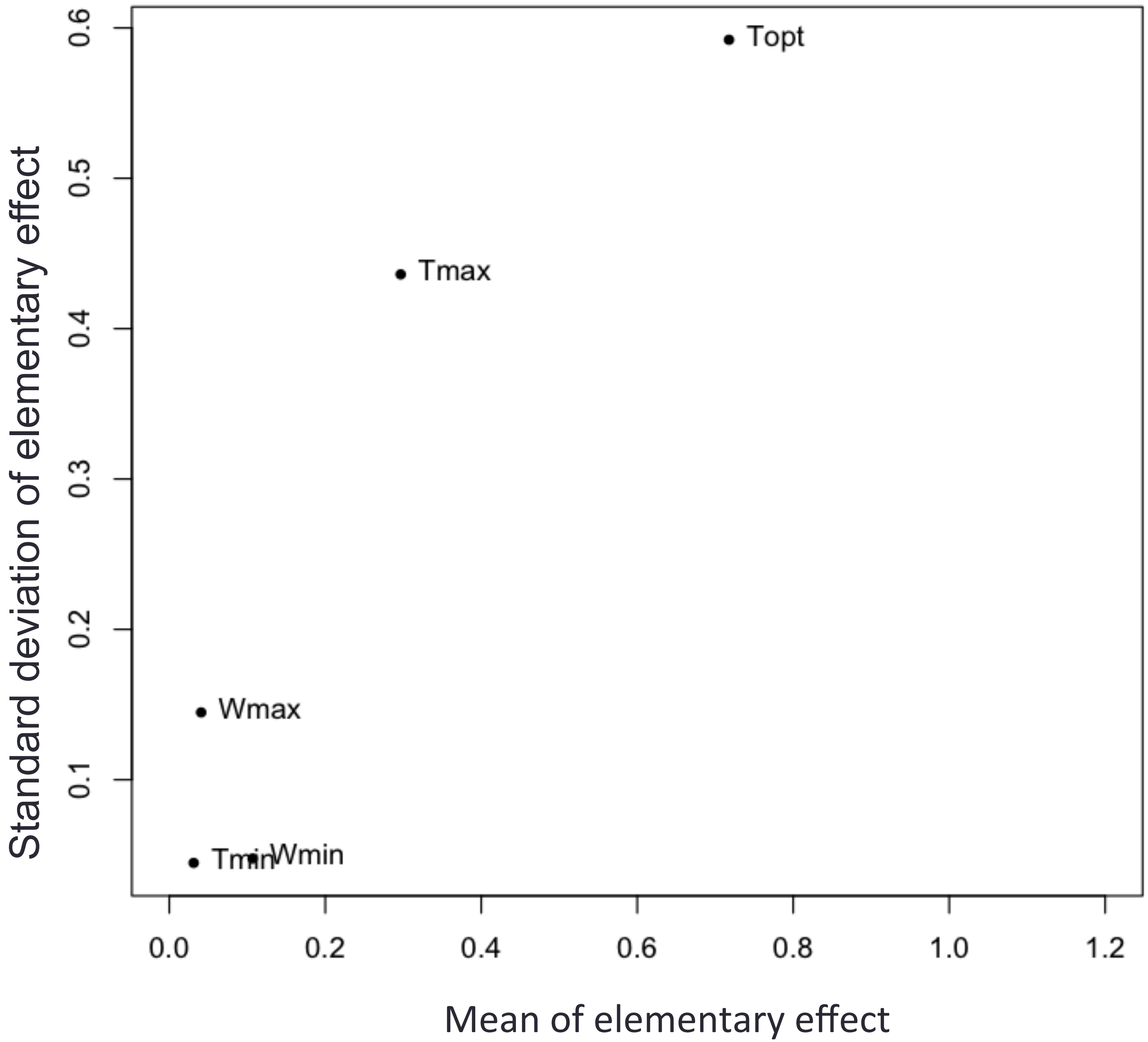


Method of Morris

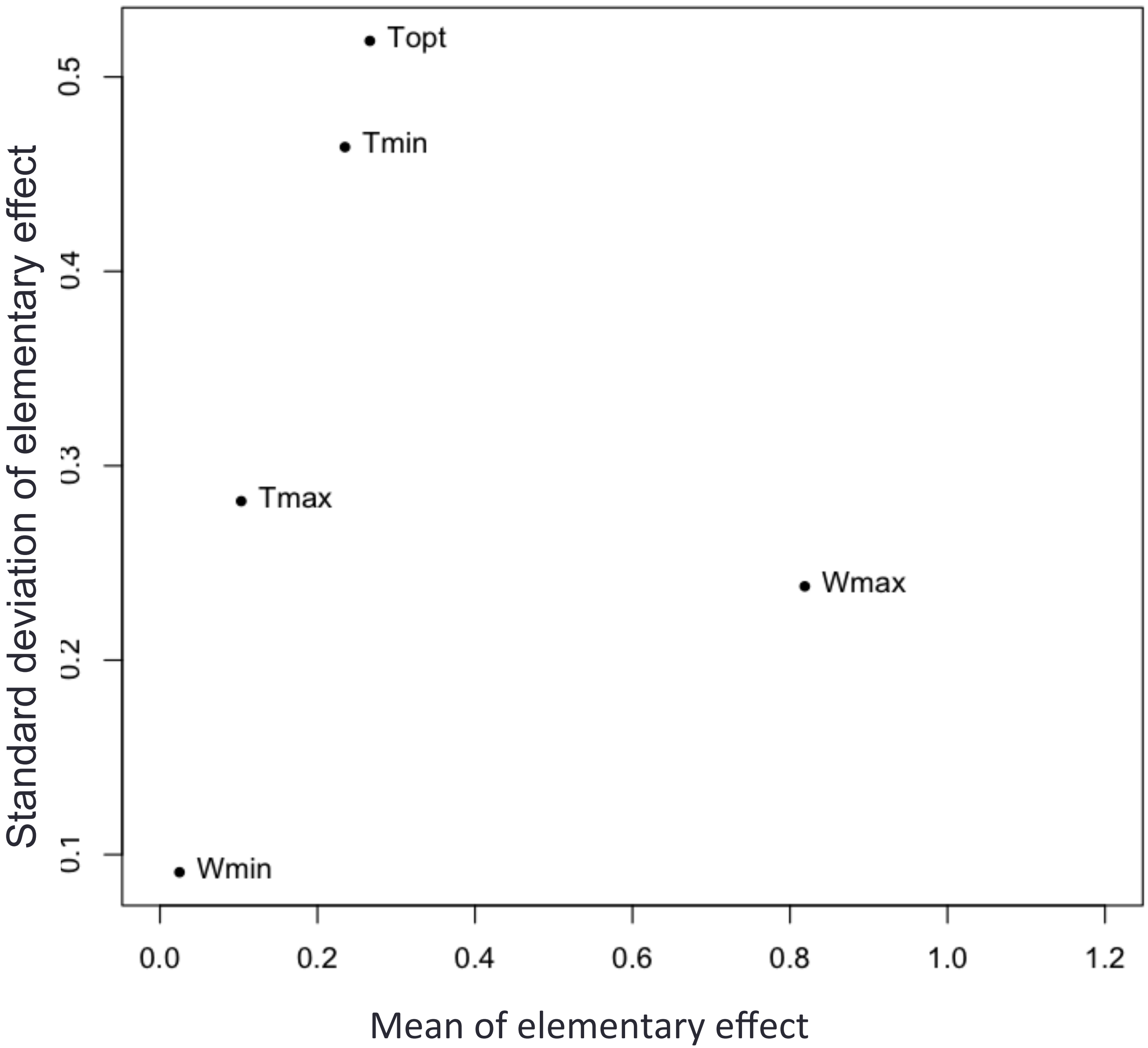
Presentation of the results



Method of Morris (T=25°C)



Method of Morris (T=16°C)



Conclusion (1)

- Several factors may influence the results of uncertainty and sensitivity analysis
- Validity of conclusion may be limited
- Important to follow some **rules**
  - ✓ Be transparent about assumptions and methods
  - ✓ Study the effect of increasing the number of simulations
  - ✓ Define precisely the model output of interest, and explore different aspects of the distribution of this model output
  - ✓ Assess robustness of results to distribution assumptions
  - ✓ Be aware of the capabilities of different sensitivity analysis techniques and, when possible, compare results

Conclusion (2)

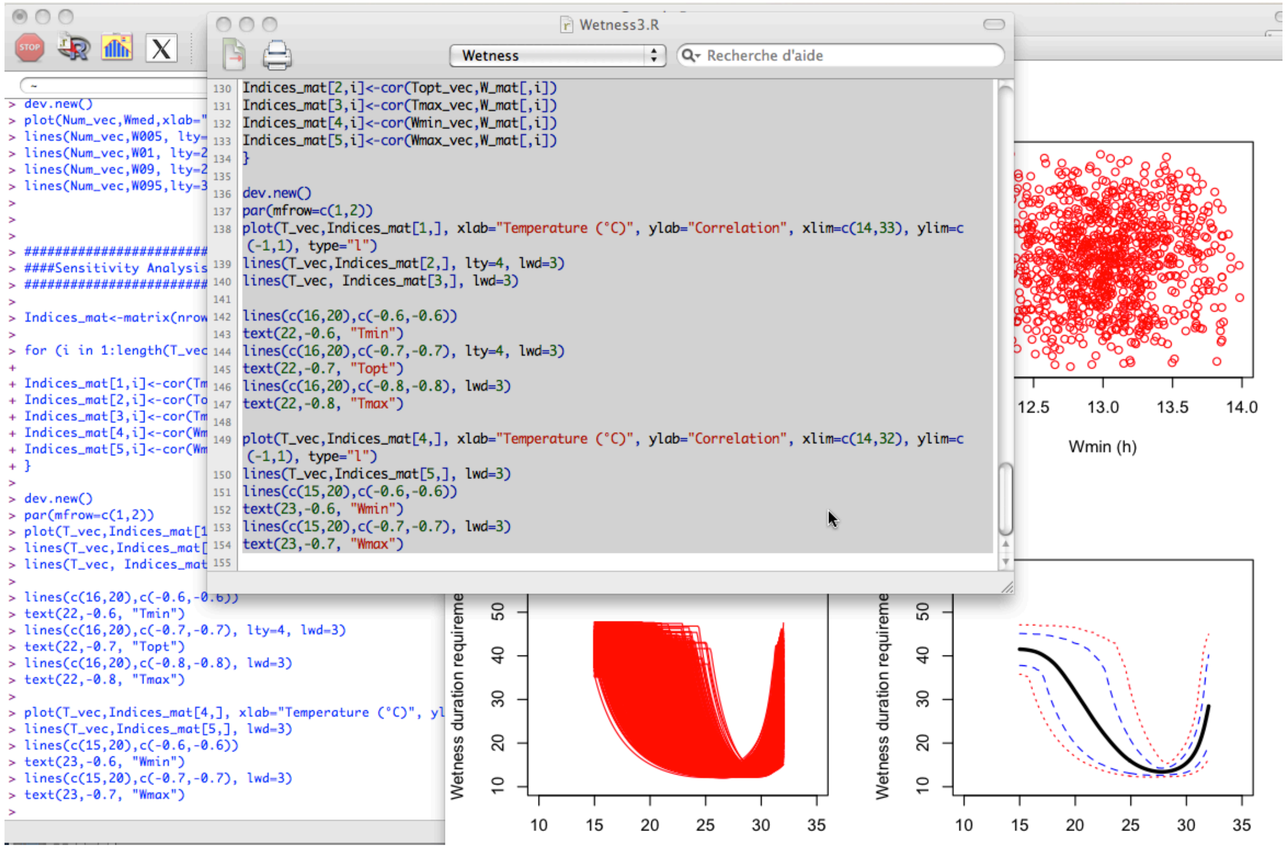
- Software are available for uncertainty and sensitivity analysis, but some training is necessary

  → Organize training sessions focused on models for PRA?
    - Part 1: Training on the use of some models (e.g., simple spread models)
    - Part 2: Uncertainty and sensitivity analysis

- Results may depend on the chosen parameter probability distributions

  → Important to derive reliable model parameter probability distributions

R


Wetness3.R
Wetness
Recherche d'aide
Indices_mat[2,i]<-cor(Topt_vec,W_mat[,i])
Indices_mat[3,i]<-cor(Tmax_vec,W_mat[,i])
Indices_mat[4,i]<-cor(Wmin_vec,W_mat[,i])
Indices_mat[5,i]<-cor(Wmax_vec,W_mat[,i])
}
dev.new()
par(mfrow=c(1,2))
plot(T_vec,Indices_mat[1,], xlab="Temperature (°C)", ylab="Correlation", xlim=c(14,33), ylim=c(-1,1), type="l")
lines(T_vec,Indices_mat[2,], lty=4, lwd=3)
lines(T_vec, Indices_mat[3,], lwd=3)
lines(c(16,20),c(-0.6,-0.6))
text(22,-0.6, "Tmin")
lines(c(16,20),c(-0.7,-0.7), lty=4, lwd=3)
text(22,-0.7, "Topt")
lines(c(16,20),c(-0.8,-0.8), lwd=3)
text(22,-0.8, "Tmax")
plot(T_vec,Indices_mat[4,], xlab="Temperature (°C)", ylab="Correlation", xlim=c(14,32), ylim=c(-1,1), type="l")
lines(T_vec,Indices_mat[5,], lwd=3)
lines(c(15,20),c(-0.6,-0.6))
text(23,-0.6, "Wmin")
lines(c(15,20),c(-0.7,-0.7), lwd=3)
text(23,-0.7, "Wmax")
Wmin (h)
Wetness duration requireme
Wetness duration requireme

References

Anderson R.P., Lew D., Peterson A.T. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling* 162, 211-232.

Dupin M., Reynaud P., […], Makowski D. 2011. Effects of training dataset characteristics on the performance of models for predicting the distribution of Diabrotica virgifera virgifera. *Plos One* 6, 1-11.

Makowski D., Mittinty M. 2010. Comparison of scoring systems for invasive pests using ROC analysis and Monte Carlo simulations. *Risk Analysis* 30, 906-915.

Makowski D., Chauvel B., Munier-Jolain N. 2010. Improving weed population model using a sequential Monte Carlo method. *Weed Research* 50, 373-382

Makowski D., Bancal R., Vincent A. 2011. Estimation of wetness duration requirements of foliar fungal pathogens with uncertain data. Application to *Mycosphaerella nawae*. *Phytopathology* 101, 1346-1354.

Magarey R.D., Sutton T.B., Thayer C.L. 2005. A simple generic infection model for foliar fungal plant pathogens. Phytopathology 95, 92-100

Monod, H., C. Naud, D. Makowski. 2006. Uncertainty and sensitivity analysis for crop models. *In: Working with dynamic crop models.* D. Wallach, D. Makowski, J. Jones Eds, Elsevier. p. 55-100

Philibert A., Desprez-Loustau M-L., […], Makowski D. 2011. Predicting invasion success in forest pathogenic fungi from species traits. *Journal of Applied Ecology* 48, 1381–1390

Scientific Opinion of the Panel on Plant Heath on a request from the European Commission on *Guignardia citricarpa* Kiely. *The EFSA Journal* (2008) 925, 1-108.

Wallach D., Makowski D., Jones J.W., Brun F. 2013. Working with dynamic crop models. Elsevier/ Academic press.