Investigating the main drivers of the variation in the outputs of ForestGALES with variance-based sensitivity analysis: The case of Sitka Spruce

Tommaso Locatelli\textsuperscript{1,3,4}, Stefano Tarantola\textsuperscript{2}, Barry Gardiner\textsuperscript{1}, Genevieve Patenaude\textsuperscript{3}

\textsuperscript{1}INRA, UMR 1391 ISPA; \textsuperscript{2}EU-JRC; \textsuperscript{3}The University of Edinburgh; \textsuperscript{4}Forest Research
Research supported by:
1. Why doing a sensitivity analysis?
2. Overview of ForestGALES
3. Overview of Sensitivity Analysis (SA) methods
4. The Global SA method of Sobol’ for correlated variables
5. Results of SA for Critical Wind Speeds
6. Results of SA for Probabilities of damage
7. Conclusions
8. Ideas for future work
1. Why doing a sensitivity analysis?

- We know a lot about what tree and stand variables drive tree failure in real storms
- We build semi-empirical process-based risk and damage models
- We attempt a “validation” of said models
- People apply these models

- Do we really know how our models behave? (investigation more difficult for complex models)

2. Overview of ForestGALES

- Two main modules: GALES and Windiness Module

- Two sets of outputs: 1) Critical Wind Speeds (CWS) for breakage and overturning; 2) Related probabilities of damage

- \( CWS_{(B, O)} \) – Inputs: tree height, dbh, stocking density, rooting depth, soil type, gap size

- Probabilities – Inputs: measure of windiness (DAMS), \( CWS_{(B, O)} \)
2. Overview of ForestGALES

GALES:
• Process-based, semi-mechanistic;
• Calculates the critical moments for stem breakage and uprooting for the mean tree within a stand;
• Calculates the CWS that would match the critical moments

Windiness Module:
• Requires a measure of the windiness of the site to calculate the Annual Exceedance Probability (and the return period)
• Detailed Aspect Method of Scoring (DAMS). Scale: 7 – 22
• DAMS scores converted to Weibull A parameter
• CWS used as inputs
3. Overview of Sensitivity Analysis (SA) methods

Local, One-At-a-Time, and Global methods

Local methods: traditionally used.
Sigma-normalised derivatives recommended by IPCC:

\[ S_{X_i} = \frac{\sigma_{X_i} \delta Y}{\sigma_Y \delta X_i} \]

Advantages: fast computation, scaled to 1.

Disadvantages:
...and when inputs are uncertain (exploring input space)?
...and when model linearity is unknown?
3. Overview of Sensitivity Analysis (SA) methods

Local, **One-At-a-Time**, and Global methods

**One-At-a-Time (OAT) methods**: a first improvement. Inputs are allowed to vary along their full range, OAT. All the others are fixed at a nominal value.

**Advantages**: computationally quite efficient
- Whole range of each input explored

**Disadvantages**: No scope for investigating interactions between variables

E.g.: Morris screening method
3. Overview of Sensitivity Analysis (SA) methods

Local, One-At-a-Time, and **Global** methods

**Global methods**: unambiguous results
All inputs are allowed to vary simultaneously within their range.

*Advantages*: model-free, i.e. make no assumptions about model linearity;
Use whole range of every variable;
Capture interactions.

*Disadvantages*: computationally quite expensive!
(at least $N(k + 2)$ model runs, and $N(2k + 2)$ with correlation)
4. Global SA: Variance-based methods

**Method of Sobol’:** variance decomposition

\[
V = \sum_{i=1}^{k} V_i + \sum_{i \leq j \leq k} V_{ij} + \ldots + V_{12\ldots k} = 1
\]

Allows calculation of 1\textsuperscript{st} order, higher orders, and Total sensitivity indices

Analytical solution: multi-dimensional integrals

Numerical solution: Monte Carlo techniques

*Main problem:* decomposition of variance with correlated variables

*Solution:* method of Kucherenko et al. (2012): using a Copula
4. Summarising the rationale of Global SA:

CWS \((B, O)\)

Probabilities \((B, O)\)

1. Copula
2. Quasi-Monte Carlo random numbers on inputs' joint-PDF

The model
ForestGALES

PDFs of input variables
\(x_1\), \(x_2\), \(x_k\)

Results:
1) Sobol' Sis & STIs
2) Smirnov's \(\alpha\) significance levels

Variance-Based Methods:
Sobol' Indices & Smirnov's two-sample Test

Distribution of \(y\)
\(<y>\)

Output
4. Settings for Global SA:

- **Factor Prioritisation:** Based on $S_j$. Tells the user which variables need to be known accurately to maximally reduce the output variance. Implications for forest managers and fieldwork.

- **Factor Fixing:** Based on $S^T_i$. Tells users and modellers which variables can be fixed to a nominal value to ensure that loss of prediction accuracy is minimal.

- **Factor Mapping:** Based on partitioning of output space in two regions: behavioural ($B$) and non-behavioural ($\overline{B}$).
5. Results – Critical Wind Speeds for Breakage and Overturning

Sobol' Indices for CWS of Breakage

Sobol' Indices for CWS of Overturning

Sobol' Indices

- First-order
- Total
5. Results – Critical Wind Speeds for Breakage and Overturning

Sobol' Indices for CWS of Breakage

- Input Variables: Height, Dbh, Sph, RD, ST, Gap, Size
- Sensitivity

Sobol' Indices for CWS of Overturning

- Input Variables: Height, Dbh, Sph, RD, ST, Gap, Size
- Sensitivity

Sobol' Indices
- First-order
- Total
5. Results – CWS: Factor Prioritisation

Sobol' Indices for CWS of Breakage

Sobol' Indices for CWS of Overturning

Sobol' Indices
- First-order
- Total
5. Results – CWS: Factor Fixing

Sobol' Indices for CWS of Breakage

Sobol' Indices for CWS of Overturning

Sobol' Indices
- First-order
- Total

Input Variables:
- Height
- Dbh
- Sph
- RD
- ST
- Gap
- Size

30/10/2015
5. Results – CWS: Fixing the variables

- **Rooting Depth** Fixed at: Shallow (1), Medium (2), and Deep (3)
5. Results – CWS: Fixing the variables

- **Soil Type** Fixed at: Free-draining mineral (1), Gleyed mineral (2), Peaty mineral (3), and Deep peats (4)
5. Results – CWS: Fixing the variables

- **Gap Size** Fixed at: 0, 1000, 500, 10x Tree height
5. Results – CWS: Fix all three: Rooting depth, Soil type, and Gap size

With Gap size: 10x Tree height
6. Results – Probabilities of Breakage and Overturning

Threshold to differentiate between behavioural and non-behavioural: 0.1
6. Results – Probability of Overturning – Smirnov Tests

Smirnov two-tailed Test - Factor Mapping Setting
Prob of Uprooting

- DAMS
- Tree Height
- Dbh
- Stocking density

- Rooting Depth
- Soil Type
- Gap size

Outcome

- OK
- Fail
6. Results – Probability of Overturning – Density plots

Focus on DAMS

2D - Density plots - Factor Mapping Setting

Prob of Overturning

DAMS vs Dbh: Probability of Overturning

DAMS vs Sph: Probability of Overturning

DAMS vs Height: Probability of Overturning

Outcome:

Fail
OK
6. Results – Probability of Overturning – Density plots

Focus on forest management

2D - Density plots - Factor Mapping Setting

Prob of Overturning

Dbh vs Height: Probability of Overturning

Sph vs Height: Probability of Overturning

Sph vs Dbh: Probability of Overturning

Outcome
- OK
- Fail
7. Conclusions – After this Global SA:

A) We (partially) trust the outputs of ForestGALES? Then:
1. End-users can use novel techniques to obtain accurate measurements of the most significant variables (Dbh, Stocking density, Tree height), such as terrestrial and airborne LiDAR
2. For large scale studies we can fix Rooting Depth, Soil Type, and Gap size, without significant losses in accuracy of model predictions

B) We don’t (entirely) trust the outputs of ForestGALES? Then:
1. Do we seek more experimental data (for empirical module)?
2. Do we take different approaches to model variables that experience shows us to be influential, but whose contribution to the outputs is not?
3. Do we want to develop a research tool, or do we focus on the end-users?
9. Ideas for Future Work (funding permitting?):

1. MORE Global SA?

In ForestGALES there are many parameters that are used deterministically, i.e. their uncertainty is not accounted for in the internal calculations.

For most of these we can calculate uncertainty (e.g. confidence intervals in regression coefficients).

Global SA techniques allow us to group variables, which is useful when treating whole equations.

Issues:
- Algorithms for grouping with the original method of Sobol’ exist (e.g. R package of B. Iooss), but for correlated variables they are not ready yet.
- The number of variables goes up to ~40, needs much more computing time (runs: N(2k +2))

2. Enough Global SA!

We can compare Global SA results with main drivers of overturning moment in tree-pulling experiments.
- Do our models pick up on our fieldwork variables?
- Are our fieldwork techniques adequate?
Thank you for your attention.

Any questions?
Annex to the presentation
Breakage:

\[ CW_{SB} = \frac{1}{kD} \left[ \frac{\pi \times MOR \times dbh^3}{32\rho G(d - 1.3)} \right]^{\frac{1}{2}} \left[ \frac{f_{knot}}{f_{CW}} \right]^{\frac{1}{2}} \ln \left( \frac{h - d}{z_0} \right) \]

Overturning:

\[ CW_{SO} = \frac{1}{kD} \left[ \frac{C_{reg} \times SW}{\rho Gd} \right]^{\frac{1}{2}} \left[ \frac{1}{f_{CW}} \right]^{\frac{1}{2}} \ln \left( \frac{h - d}{z_0} \right) \]
Method of Sobol’

Intuitively:

Clearly different, same mean. Thus we use the variance as a measure of sensitivity
Global SA: Variance-based methods

Method of Sobol’ (orthogonal variables)

Function (a model) decomposition:

\[ f = f_0 + \sum_{i} f_i(X_i) + \sum_{i} \sum_{j > i} f_{i,j} + \cdots + f_{1,2,3,\ldots,k} \]
Global SA: Variance-based methods

Method of Sobol’ (orthogonal variables)
Function (a model) decomposition:

\[ f = f_0 + \sum_i f_i(X_i) + \sum_i \sum_{j>i} f_{i,j} + \cdots + f_{1,2,3,\ldots,k} \]

\[ f_0 = E(Y) \]

\[ f_i(X_i) = E(Y|X_i) - f_0 \]

\[ f_{i,j} = E(Y|X_i, X_j) - f_i(X_i) - f_j(X_j) - f_0 \]
Method of Sobol’ (orthogonal variables)

Variance decomposition:

\[ V = \sum_{i=1}^{k} V_i + \sum_{i\leq j\leq k} V_{ij} + \cdots + V_{12\ldots k} = 1 \]

\[ V_i = V\left(E(Y|X_i = x_i^*)\right) \]

\[ V_{ij} = V\left(E(Y|X_i = x_i^*, X_j = x_j^*)\right) - V_i - V_j \]
Global SA: Variance-based methods

Method of Sobol’ (orthogonal variables)

Empirical calculation of conditional variance:

\[ V(f_i(X_i)) = V[E(Y|X_i)] \]
Method of Sobol’ (orthogonal variables)

Empirical calculation of conditional variance:

\[ V(f_i(X_i)) = V[E(Y|X_i)] \]
Global SA: Variance-based methods

Method of Sobol’ (orthogonal variables)

First-order sensitivity indices:

\[ S_i = \frac{V(E(Y|X_i))}{V_Y} \]

Total sensitivity indices (example with 3 variables \((X_i, X_j, X_l)\)):

\[ S_i^T = S_i + S_{i,j} + S_{i,j,l} \]
\[ S_i^T = 1 - D_{X_{-i}} \]

Global SA: Variance-based methods

Method of Sobol’ (orthogonal variables)
Monte Carlo method: Generate two matrices A and B

\[
A = \begin{bmatrix}
    x_1^{(1)} & x_2^{(1)} & \cdots & x_i^{(1)} & \cdots & x_k^{(1)} \\
    x_1^{(2)} & x_2^{(2)} & \cdots & x_i^{(2)} & \cdots & x_k^{(2)} \\
    \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\
    x_1^{(N-1)} & x_2^{(N-1)} & \cdots & x_i^{(N-1)} & \cdots & x_k^{(N-1)} \\
    x_1^{(N)} & x_2^{(N)} & \cdots & x_i^{(N)} & \cdots & x_k^{(N)} \\
\end{bmatrix}
\]

\[
B = \begin{bmatrix}
    x_{k+1}^{(1)} & x_{k+2}^{(1)} & \cdots & x_{k+i}^{(1)} & \cdots & x_{2k}^{(1)} \\
    x_{k+1}^{(2)} & x_{k+2}^{(2)} & \cdots & x_{k+i}^{(2)} & \cdots & x_{2k}^{(2)} \\
    \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\
    x_{k+1}^{(N-1)} & x_{k+2}^{(N-1)} & \cdots & x_{k+i}^{(N-1)} & \cdots & x_{2k}^{(N-1)} \\
    x_{k+1}^{(N)} & x_{k+2}^{(N)} & \cdots & x_{k+i}^{(N)} & \cdots & x_{2k}^{(N)} \\
\end{bmatrix}
\]

Then define a matrix \( C_i \) formed by all columns of \( B \) except for the \( i \)th column, which is taken from \( A \)

Global SA: Variance-based methods

Method of Sobol’ (orthogonal variables)
Monte Carlo method:

\[ C_i = \begin{bmatrix} 
  x_{k+1}^{(1)} & x_{k+2}^{(1)} & \ldots & x_i^{(1)} & \ldots & x_{2k}^{(1)} \\
  x_{k+1}^{(2)} & x_{k+2}^{(2)} & \ldots & x_i^{(2)} & \ldots & x_{2k}^{(2)} \\
  \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
  x_{k+1}^{(N-1)} & x_{k+2}^{(N-1)} & \ldots & x_i^{(N-1)} & \ldots & x_{2k}^{(N-1)} \\
  x_{k+1}^{(N)} & x_{k+2}^{(N)} & \ldots & x_i^{(N)} & \ldots & x_{2k}^{(N)} 
\end{bmatrix} \]

Then compute the model output for all the inputs in the matrices \( A, B, \) and \( C_i \):

\[ y_A = f(A) \quad y_B = f(B) \quad y_{C_i} = f(C_i) \]

\[ S_i = \frac{V[E(Y|X_i)]}{V(Y)} = \frac{y_A \cdot y_{C_i} - f_0^2}{y_A \cdot y_A - f_0^2} \]

\[ f_0^2 = \left( \frac{1}{N} \sum_{j=1}^{N} y_A^{(j)} \right)^2 \]

Global SA: Variance-based methods

Method of Sobol’ (orthogonal variables)
Monte Carlo method:

\[
S^T_i = 1 - \frac{V[E(Y|X_{\sim i})]}{V(Y)} = 1 - \frac{y_B \cdot y_{C_i}}{y_A \cdot y_A - f_0^2} = 1 - \frac{(1/N) \sum_{j=1}^{N} y^{(j)}_B y^{(j)}_{C_i} - f_0^2}{(1/N) \sum_{j=1}^{N} (y^{(j)}_A)^2 - f_0^2}
\]

Kucherenko et al. (2012) suggest using a Gaussian Copula to map the correlation structure of the inputs (described by arbitrary distributions) to a ”normal” correlation structure:

\[ C(G_1(X_1), \ldots, G_n(X_n); \Sigma_X) = F_n(F^{-1}(G_1(X_1)), \ldots, F^{-1}(G_n(X_n)); \Sigma) \]

Kucherenko et al. (2012) suggest using a Gaussian Copula to map the correlation structure of the inputs (described by arbitrary distributions) to a Normal correlation structure:

\[
C(G_1(X_1), \ldots, G_n(X_n); \Sigma_X) = F_n(F_n^{-1}(G_1(X_1)), \ldots, F_n^{-1}(G_n(X_n)); \Sigma)
\]

Sitka Spruce data:

Source: UK Tree-Pulling database (~1,000 entries)

Fitting PDFs to the inputs:
• Tree height and Dbh: Gaussian
• Stocking density, Rooting Depth, Soil Type, Gap size, DAMS: Uniform

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type of distribution</th>
<th>Distribution parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>mean</td>
</tr>
<tr>
<td><strong>Height (m)</strong></td>
<td>Gaussian</td>
<td>13.64</td>
</tr>
<tr>
<td><strong>Dbh (cm)</strong></td>
<td>Gaussian</td>
<td>19.89</td>
</tr>
<tr>
<td><strong>Sph</strong></td>
<td>Discrete Uniform</td>
<td>300</td>
</tr>
<tr>
<td><strong>Rooting depth</strong></td>
<td>Discrete Uniform</td>
<td>1</td>
</tr>
<tr>
<td><strong>Soil type</strong></td>
<td>Discrete Uniform</td>
<td>1</td>
</tr>
<tr>
<td><strong>Gap size (m)</strong></td>
<td>Uniform</td>
<td>0</td>
</tr>
<tr>
<td><strong>DAMS</strong></td>
<td>Discrete Uniform</td>
<td>7</td>
</tr>
</tbody>
</table>
Correlation structure:

- Tree height
- Dbh
- Sph
- Rooting depth
- Soil type
- Gap size
- DAMS

Sitka Spruce data:
Probability of Breakage – Smirnov Tests

Smirnov two-tailed Test - Factor Mapping Setting

Prob of Breakage

- DAMS
- Tree Height
- Dbh
- Stocking density

- Rooting Depth
- Soil Type
- Gap size

Outcome

OK
Fail
Probability of Breakage – Density plots

Focus on DAMS

2D - Density plots - Factor Mapping Setting
Prob of Breakage

DAMS vs Dbh: Probability of Breakage

DAMS vs Sph: Probability of Breakage

DAMS vs Height: Probability of Breakage

Outcome:
- Fail
- OK
Probability of Breakage – Density plots

Focus on forest management

2D - Density plots - Factor Mapping Setting

Prob of Breakage

Dbh vs Height: Probability of Breakage

Sph vs Height: Probability of Breakage

Sph vs Dbh: Probability of Breakage

Outcome:
- Fail
- OK