

1 Introduction

In the final tutorial on sensitivity analysis, we will look at the factorial ANOVA (ANalysis Of VAriance) method. This tutorial builds on both the Morris and Sobol method tutorials. If you have not had a look at these yet, it might be useful to do so before continuing.

The factorial ANOVA is very closely related to the Sobol method. Both approaches rely on the idea that a mathematical function of m inputs can be decomposed into 2^m functions, each corresponding to one combination of inputs. The factorial ANOVA uses a much simpler (if more computationally intensive) sampling approach.

A factorial ANOVA (just like the Sobol method) identifies inputs with a large contribution to output variability. The method can again quantify (as a proportion of total output variance), the contribution of each input. In the same way, the contribution of each interaction term can be quantified.

The advantage of the factorial ANOVA is that it can easily cope with categorical and composite inputs. Remember that a composite input refers to a collection of values that are typically not changed independently (e.g. a whole soil node, or `met` file). The method implemented here makes use of the factorial functionality in APSIM, and is likely to be intuitive to the experienced APSIM user.

Allowing the currently implemented Sobol method to do this, would require a number of significant changes. The factorial ANOVA offers a much simpler, more intuitive solution.

A disadvantage of the factorial ANOVA is that the required number of simulations increases exponentially with the number of inputs. The factorial ANOVA is likely to be used as an alternative to the Sobol method, following a preliminary sensitivity analysis (using the Morris method for example).

This tutorial will demonstrate the use of the factorial ANOVA analysis, including setup, and interpretation of results. We will also compare the results to Morris and Sobol method findings. As in previous tutorials, the modelling of nitrogen leaching from a cow urine patch will be used for context.

2 Factorial Anova Setup

3 Experiment Node

The structure of the factorial ANOVA is quite different, when compared to the Morris and Sobol methods. At the centre of the analysis is the experiment node (simply named `Experiment` here). The experiment node is used to run all simulations. It also contains the base simulation (`UrinePatchSimulation`). Hopefully you are reasonably familiar with the base simulation by now, but feel free to remind yourself of its structure. The factorial ANOVA can easily cope with multiple values of the same output. Hence, as in the Morris method tutorial, the simulation will run from 1972 to 1983, and report five values of nitrogen leached (`LeachN`).

The experiment node also contains a factors node (`Factors`). The factors node contains individual factor nodes, each corresponding to one sensitivity analysis input. Expand `Factors` to confirm the presence of an `SOC`, `Load`, and `DUL` factor. Note that we do not need to include the year as one of the factors, although we will analyse its effect. Year is already a reporting variable in the base simulation, meaning the required data will be recorded.

The experiment node is set up in exactly the same way as in any other APSIM simulation. Select each factor in turn to get an idea of how they are defined. If you need more help with setting up the experiment, explore the `Factorial.apsimx` simulation, from the examples folder. You may also wish to look at the [factorial simulation tutorial](#) (note this is somewhat outdated now).

The most important part of the sensitivity analysis (as usual) is selecting the values for each factor (i.e. input). Remember that this may significantly affect the results of the analysis. For categorical and composite inputs, you simply need to include each possible value. Numeric input values should adequately cover the complete range of interest. The problem is similar to defining the sample grid in the Morris method. Revisit the Morris method tutorial if you require additional guidance.

4 Factorial Anova Node

The factorial ANOVA node is a model of type `FactorialAnova`. Here it is named `FactorialAnovaExample`. The factorial ANOVA node is primarily a post-simulation tool. It works by extracting the experiment data from the `DataStore`, analysing it using the R language, and recording the analysis results back in the `DataStore`. The `FactorialAnova` node should be added at the same level as the `DataStore`. Note that you should only add one experiment node and one `FactorialAnova` node per APSIMX file.

The setup of the ANOVA node is simple. Select `FactorialAnovaExample` to have a look. In the top panel of the user interface, you need to enter each input on a separate line. The input names must be the same as the factor names in the factors node. If you wish to analyse temporal variability, you also need to include the name of the time variable. Here we are using `[Clock].Today`, renamed to `Year` (since reporting happens only once every two years). In the bottom panel of the interface, you enter the output names. Each output name should match an output name defined in the `Report` node, or a reporting script. The R script will attempt to match these input and output names with column names in the experiment results dataset. At this stage there is no checking mechanism, so you need to ensure the names are entered correctly!

5 Factorial ANOVA Algorithm

The factor values entered in the factors node are used to construct a sampling grid, similar to the one used in the Morris method:

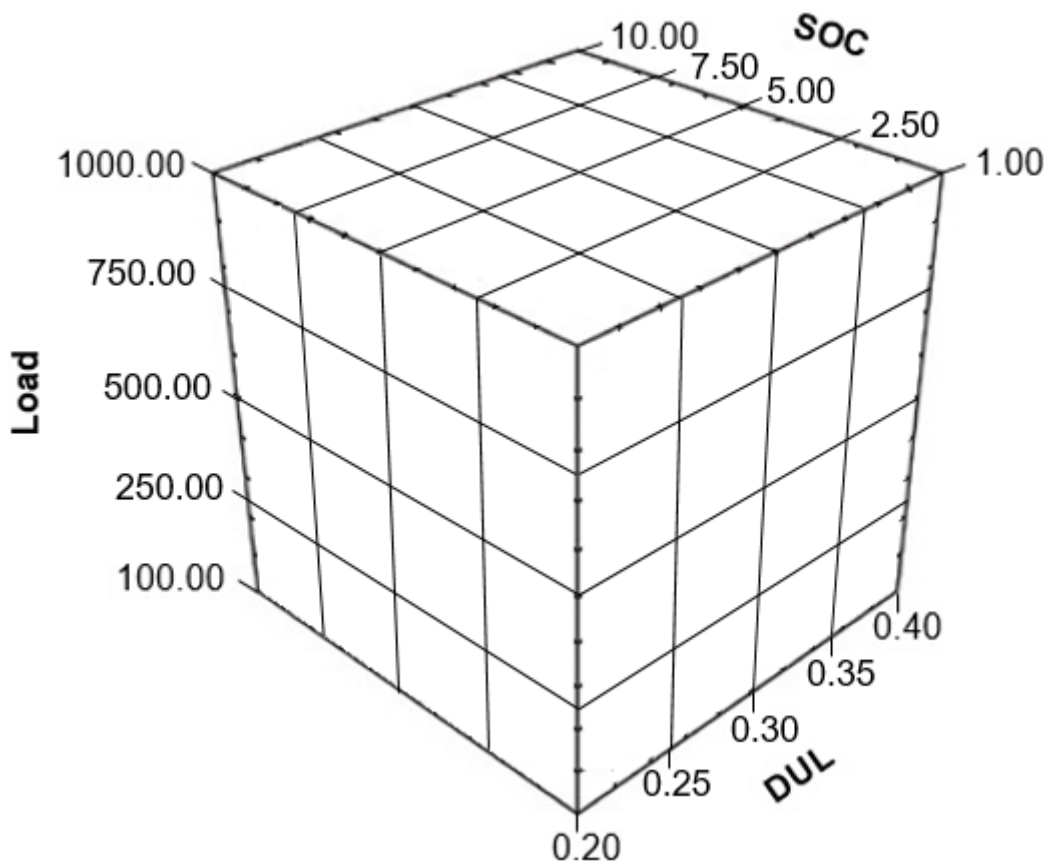


Figure 1: SampleGrid

In this case, *all* points in the grid (i.e. combinations of input values) are used for running simulations. Run the analysis from the `Simulations` node - it may again take a few minutes to finish. The analysis will produce the same sensitivity measures as the Sobol method - a first order and total order index for each input. If you are unsure what these measures are, revisit the Sobol method tutorial.

6 FactorialAnovaExample

Encapsulates a factorial ANOVA parameter sensitivity analysis.

7 Analysis Results

The factorial ANOVA (just like the Sobol method) produces two data sources: `Report` and `Statistics`. We will again focus on `Statistics`, which contains the sensitivity indices. Note that the factorial ANOVA involves no random sampling. Hence as long as the results of the base simulation are reproducible, the results of the sensitivity analysis will also be reproducible.

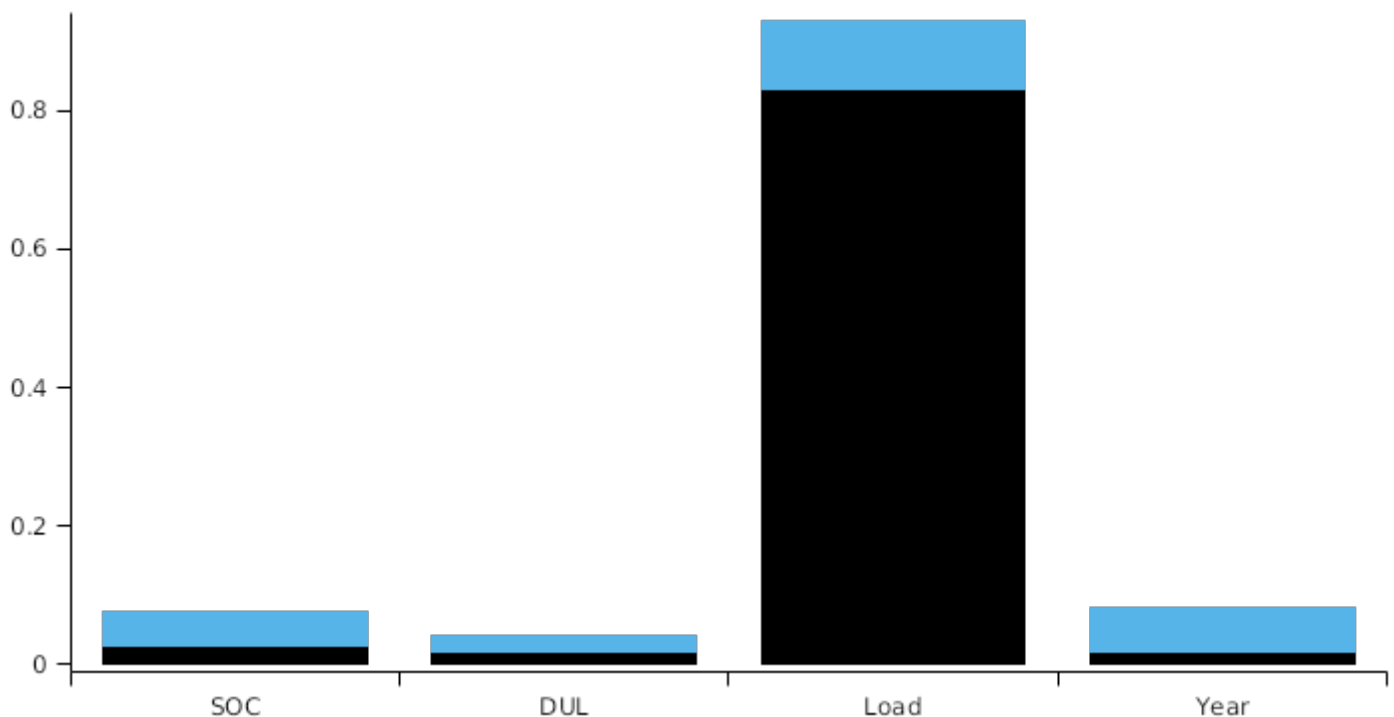
Select the graph node named `Indices`. Figure 1 shows the overlaid first and total order indices for each input. We can see that `Load` is the most important input, regardless of whether interaction effects are considered. Without considering interaction effects, `Load` accounts for almost 25% of variation in `LeachN`. If we add on all interaction effects involving `Load`, it accounts for over 70% of variation in `LeachN`.

`DUL`, `SOC`, and `Year` are not particularly important when interactions are ignored (i.e. their first order indices are low). All three inputs are important when interaction effect are considered however (the total order indices are high). This is consistent with the Morris method findings, where `DUL` and `SOC` were shown to significantly contribute to `LeachN` variation in some, but not all, years.

Note that neither the first or total order indices add up to one (i.e. 100% of output variation). The sum of all first order indices is typically less than one, as output variation introduced by interaction effects is ignored. The sum of all total order indices is typically greater than one, as each interaction effect contributes to more than one total order index.

The factorial ANOVA method has no convenient tool for assessing the reliability of results right now. If the associated computational cost is not prohibitive, it may be sensible to change the factor values, and observe how this affects the sensitivity. Try changing the factor values in this example, to get some practice at setting up the analysis. You may also wish to try recreating the `Indices` graph. Hint: the order of the two series nodes is important!

Indices



8 Conclusion

This tutorial concludes the series on sensitivity analysis. Like the Sobol method, the factorial ANOVA was used to obtain quantitative sensitivity measures (first and total order indices). `Load` was again shown to be the most influential input, but `year`, `DUL`, and `SOC` were also important. If you look back at the conclusions drawn from the Sobol and Morris method, you will see the three methods are in general agreement. This suggests the sensitivity analyses we conducted are sensible, and should give us some confidence about the robustness of the results.