SENSITIVITY ANALYSIS WITH SOBOL

LARGER COMPLEX MODELS

Only do sensitivity on some inputs/parameters
Use Latin Hypercube to generate samples
Run model for samples
Generate Summary Statistics and graph

MORE INFORMATION ON SENSITIVITY ANALYSIS

[Applying Sensitivity Analysis in Biology](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2570191/)
Nice paper on sensitivity analysis for environmental modeling
Saltelli, Andrea, and Paola Annoni. “How to avoid a perfunctory sensitivity analysis.” Environmental Modelling & Software 25, no. 12 (2010): 1508-1517

STEPS IN SENSITIVITY ANALYSIS

Define the distribution of input parameters (e.g normal with its mean and standard deviation)
Define the outputs to be considered (e.g if your output is time series are you looking at daily, max, min, annual ?)
Sample from pdf of input parameters - we use *randomLHS* for this
Run the model for each sample
Graph the results
Quantify sensitivity * we used *pcc*

SOBOL

Sobol sensitivity analysis is similar to LHS approach as a way to efficiently sample parameter space
Sobol is more general - it can use measured parameter samples (we will approximate sampling here by using known distributions but you could use actual samples!)
Sobol is a *variance-based* method
Sobol quantifies sensitivity by breaking the variance of output into contributions by parameters

SOBOL INDICES - QUANTIFYING SENSITIVITY

Indices are computed using parameter-output variance relationships; they are estimates so the indices themselves have uncertainty bounds
Several Indices
First order sensitivity or Main Effect

- variance associated directly with parameter alone
- fraction associated with each parameter, sum to 1 (although because it is estimated can be slightly more or less)

Total Effect

- variance associated with parameter and interaction with other parameters
- sum can be more than 1 if parameters interact

Second Order Indices

- less used but quantify how parameter pair-wise parameter interactions contribute to output variation

SOBOL IN R

- *Sensitivity* package
- Sobol Indices require estimation - and there are different methods to do that
- The *Sensitivity* package has several of those
- today we will use *sobolSalt* - which uses a method by Saltelli (who has written extensively on Sensitivity analysis)
- R help pages for *Sensitivity* provide many good references
- This is a nice overview paper

Variance Based Methods

SOBOL - HOW TO

- run Sobol to get parameter sets in a sensitivity analysis object
- run model with those parameter sets
- *tell* the senstivity object about results associated with each parameter set
- look at sensitivity analysis indices from Sobol

  Generation of parameter sets slightly different

- generate **two** samples of parameter sets by sampling from a-priori (expected) distributions
- these would be the “hypothetical” distributions based on assumptions about the data
- ideally these would be distributions that you actually sampled

EXAMPLE ATMOSPHERIC CONDUCTANCE

Often when we are estimating vegetation or crop water use we need to know the atmospheric conductance - which is essentially how easily water diffuses into the air and depends largely on windspeed (you get more evaporation in windier conditions)
Atmospheric conductance is also influenced by the vegetation itself and the turbulence it creates
I’ve provided a function to compute atmospheric conductance C_{at} (how easily vapor diffuses from vegetation surfaces)
The function *Catm.R* is provided

ATMOSPHERIC CONDUCTANCE (SIMPLE MODEL)

So that you know what it does - here's some background on the function

$$C_{at} = \frac{v_m}{6.25 * ln(\frac{z_m - z_d}{z_0})^2}$$

$$z_d = k_d * h$$

$$z_0 = k_0 * h$$

z_m is the height at which windspeed is measured - must be higher than the vegetation (cm), it is usually measured 200 cm above the vegetation

h is vegetation height (cm)

v is windspeed (cm/s)

Typical values if k_d and k_o are 0.7 and 0.1 respectively (so use those as defaults)

SENSITIVITY ANALLYSIS GOAL

For a given forest, perform a sensitivity analysis of model predictions of conductance Consider the sensitivity to uncertainty in the following parameters and inputs

- $height$
- k_d
- k_0
- v

Windspeeds v are normally distributed with a mean of 250 cm/s with a standard deviation of 30 cm/s
For vegetation height assume that height is somewhere between 9.5 and 10.5 m (but any value in that range is equally likely)
For the k_d and k_0 parameters you can assume that they are normally distributed with standard deviation of 1% of their default values

STEPS: GENERATE PARAMETERS

First we need to generate parameter sets for the Sobol analysis

- 2 sets of samples of all of the parameters, we will make these up by sampling from expected distributions
- use *sobolSalt* to generate the parameter sets

```
source(here("R/Catm.R"))

# generate two examples of random number f

np <- 1000
k_o <- rnorm(mean = 0.1, sd = 0.1 * 0.1, n
k_d <- rnorm(mean = 0.7, sd = 0.7 * 0.1, n
v <- rnorm(mean = 250, sd = 30, n = np)
height <- runif(min = 9.5, max = 10.5, n =

X1 <- cbind.data.frame(k_o, k_d, v, height

# repeat sampling
k_o <- rnorm(mean = 0.1, sd = 0.1 * 0.1, n
k_d <- rnorm(mean = 0.7, sd = 0.7 * 0.1, n
v <- rnorm(mean = 250, sd = 30, n = np)
height <- runif(min = 9.5, max = 10.5, n =

X2 <- cbind.data.frame(k_o, k_d, v, height
```

STEPS: COMPUTE SOBOL INDICES

now run model for Sobol generated parameter sets and compute indices

- pay attention to values of the indices and confidence intervals
  - if 0 is within the confidence interval, parameter uncertainty is not influencing output
- substantial differences between total effect and first order indices suggest parameter interactions

TIP: a useful plotting strategy is to plot model output against parameter with the highest total effect and then use the parameter with second highest total effect for color

```
# run model for all parameter sets
# make sure you give the parameters names

parms <- as.data.frame(sens_Catm_Sobol$X)
colnames(parms) <- colnames(X1)
res <- pmap_dbl(parms, Catm)


sens_Catm_Sobol <- sensitivity::tell(sens_C

# main effect:  partitions variance (main e
sens_Catm_Sobol$S
```

```
     original         bias std. error  min.
c.i.  max. c.i.
X1 0.23227725  0.002082769 0.02913414
0.1678434 0.28654280
X2 0.56512106 -0.001339263 0.03033162
0.5100473 0.62345926
X3 0.22400354  0.006861115 0.03277670
0.1379324 0.27116373
X4 0.03490354  0.005841568 0.03146518
-0.0414199 0.08890148
```

```
# useful to add names
row.names(sens_Catm_Sobol$S) <- colnames(pa
sens_Catm_Sobol$S
```

```
         original         bias std. error
min. c.i.  max. c.i.
k_o    0.23227725  0.002082769 0.02913414
0.1678434 0.28654280
k_d    0.56512106 -0.001339263 0.03033162
0.5100473 0.62345926
v      0.22400354  0.006861115 0.03277670
0.1379324 0.27116373
height 0.03490354  0.005841568 0.03146518
-0.0414199 0.08890148
```

```
# total effect - accounts for parameter int
row.names(sens_Catm_Sobol$T) <- colnames(pa
sens_Catm_Sobol$T
```

```
          original          bias   std.
error   min. c.i. max. c.i.
k_o    0.224026427  2.278809e-04
```

```
0.0167220483 0.192247038 0.2582154
k_d    0.590569549 -7.497935e-03
0.0370026338 0.525584090 0.6707337
v      0.220518359  3.789481e-03
0.0172329018 0.185413408 0.2469731
height 0.004692158  4.405831e-05
0.0004008086 0.003724768 0.0053911
```

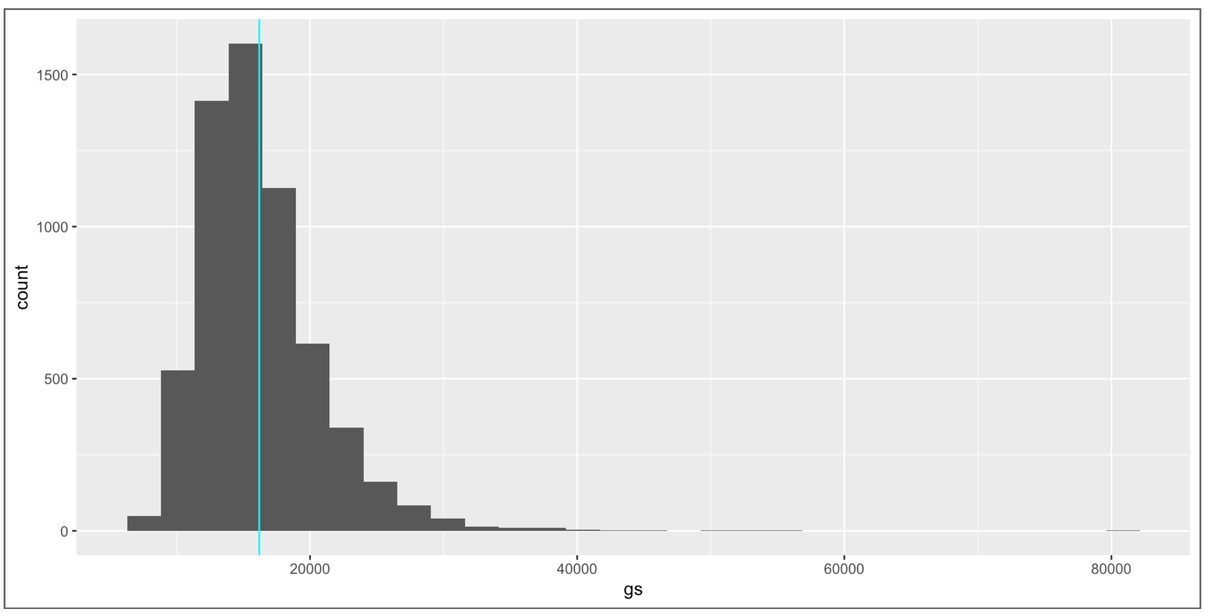
```
# Both the main effect and total effect car


print(sens_Catm_Sobol)
```

```
Call:
sobolSalt(model = NULL, X1 = X1, X2 = X2,
nboot = 100)

Model runs: 6000

Model variance: 20235596

First order indices:
         original         bias std. error
min. c.i.  max. c.i.
k_o    0.23227725  0.002082769 0.02913414
0.1678434 0.28654280
k_d    0.56512106 -0.001339263 0.03033162
0.5100473 0.62345926
```

STEPS: PLOTTING

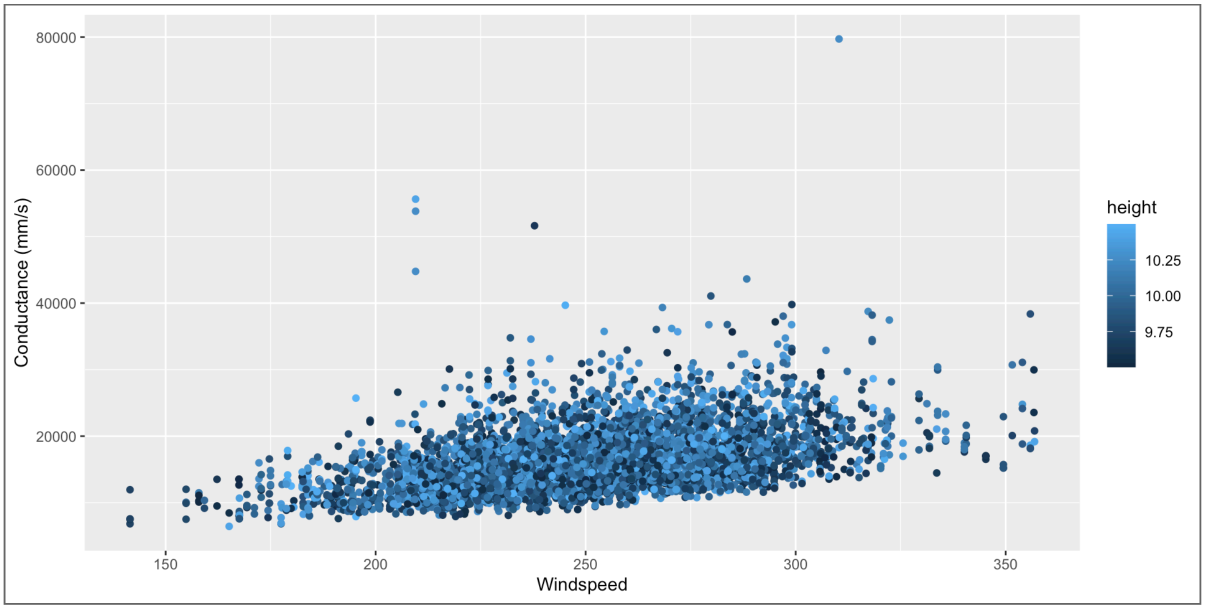
- uncertainty in the output
- relationships you are interested in
- response to most sensitive parameters

```
# graph two most sensitive parameters
both <- cbind.data.frame(parms, gs = sens_C

# look at overall gs sensitvity to uncertai
ggplot(both, aes(x = gs)) +
  geom_histogram() +
  geom_vline(xintercept = mean(both$gs), co
```

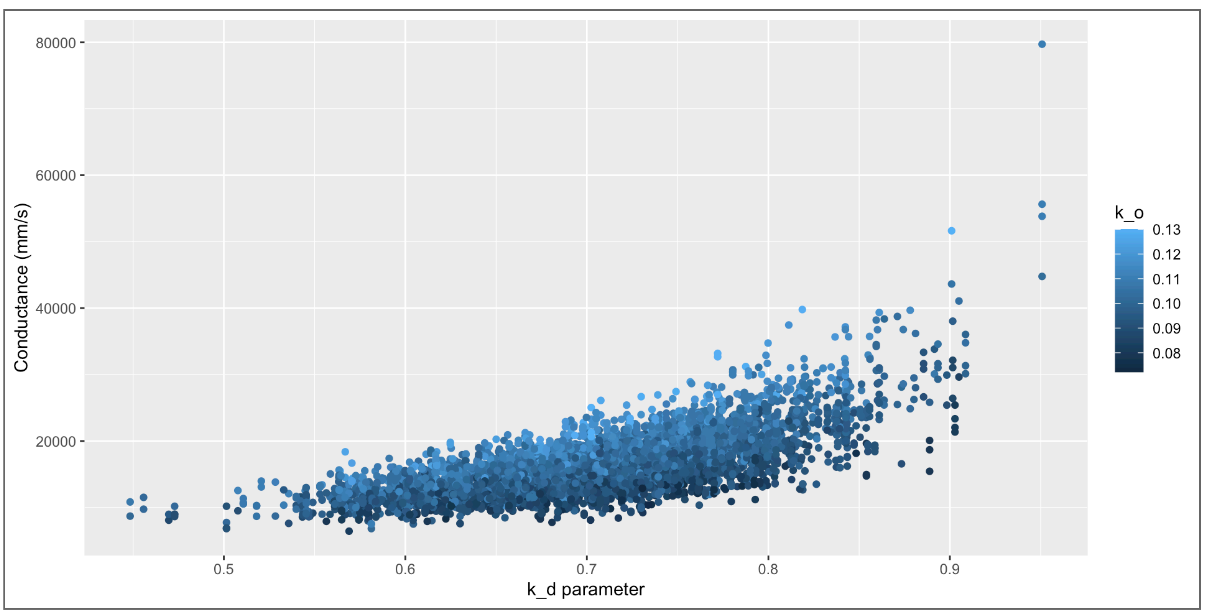



```
# look at response of conductance to the tw
ggplot(both, aes(v, gs, col = height)) +
```

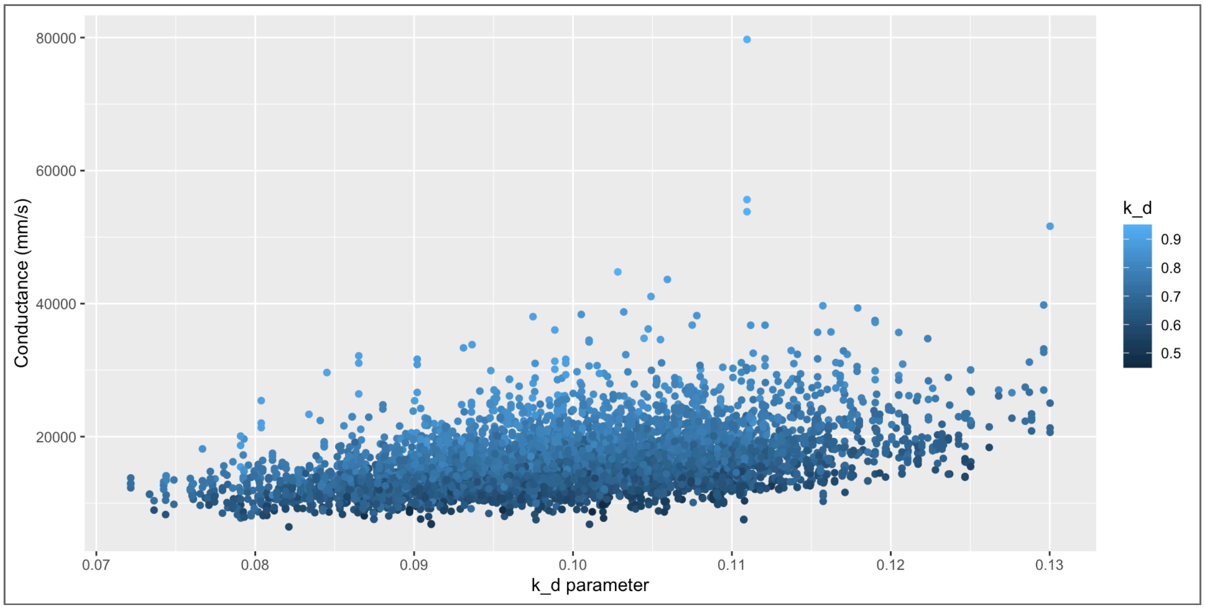
```
3   geom_point() +
4   labs(y = "Conductance (mm/s)", x = "Winds
```




```
1 # look at response of conductance to the tw
2 ggplot(both, aes(k_d, gs, col = k_o)) +
3   geom_point() +
4   labs(y = "Conductance (mm/s)", x = "k_d p
```



```
# use second most sensitive parameter (usir
ggplot(both, aes(k_o, gs, col = k_d)) +
  geom_point() +
  labs(y = "Conductance (mm/s)", x = "k_d p
```



SECOND ORDER INDICES

Optional for this course

If you want to also compute a second order indices you need to use a different variation. (scheme=B)

There are multiple implementation of Sobol in the *senstivity* package, they can be more or less stable I find *sobolSalt* works well

```
sens_Catm_Sobol2 <- sobolSalt(model = NULL,

parms <- as.data.frame(sens_Catm_Sobol2$X)
colnames(parms) <- colnames(X1)
res <- pmap_dbl(parms, Catm)


sens_Catm_Sobol2 <- sensitivity::tell(sens_

# main effect:  partitions variance (main e
row.names(sens_Catm_Sobol2$S) <- colnames(p
sens_Catm_Sobol2$S
```

```
          original          bias std. error
min. c.i. max. c.i.
k_o    0.22127668 -0.0004113736 0.02758751
0.16691220 0.2766504
k_d    0.58015690 -0.0072121449 0.02699959
0.53052314 0.6442312
v      0.22652414  0.0002367047 0.03069473
0.17293051 0.2857273
```

```
height 0.03616299 -0.0020311632 0.03088224
-0.01811522 0.1007975
```

```
# total effect - accounts for parameter int
row.names(sens_Catm_Sobol2$T) <- colnames(p
sens_Catm_Sobol2$T
```

```
          original          bias   std.
error   min. c.i.   max. c.i.
k_o    0.227906190  1.955752e-03
0.0133575920 0.199819181 0.255644518
k_d    0.587074007 -1.313624e-03
0.0304211185 0.524137931 0.645819730
v      0.220252365  2.844003e-03
0.0120806053 0.195198308 0.244027113
height 0.004618199  4.465597e-05
0.0002781498 0.003942629 0.005128816
```

```
# second order parameters interaction in co
# parameters are in order, interactiosn are
sens_Catm_Sobol2$S2
```

```
        original        bias std. error
min. c.i.  max. c.i.
X1X2 -0.01263902 0.004524975 0.03394789
-0.08228495 0.04323694
X1X3 -0.02669965 0.001377749 0.03403418
-0.09645933 0.03920953
X1X4 -0.03593402 0.001797324 0.03096804
-0.09999669 0.02103481
X2X3 -0.03349987 0.002973166 0.03382784
-0.10937990 0.02554487
X2X4 -0.03149249 0.002327382 0.03123475
-0.09518969 0.02120359
```

```
X3X4 -0.03263918 0.002132550 0.03118150
-0.09873748 0.02400883
```

ASSIGNMENT PART I

Choose one of the 3 papers below that provide an example of sensitivity analysis of model parameters. After going through the paper, write a paragraph describing how results of the *sensitivity analysis* reported on in the paper might contribute to understanding (or prediction) within an environmental problem solving or management context.

Snow modeling

Building Cooling Energy Mdoel

Uranium in Groundwater Model

ASSIGNMENT PART 2

Recall our model of atmospheric conductance

$$C_{at} = \frac{v_m}{6.25 * ln(\frac{z_m - z_d}{z_0})^2}$$

$z_d = k_d * h$

$z_0 = k_0 * h$

z_m is the height at which windspeed is measured - must be higher than the vegetation (cm), it is usually measured 200 cm above the vegetation

h is vegetation height (cm)

v is windspeed (cm/s)

Typical values if k_d and k_o are 0.7 and 0.1 respectively (so use those as defaults)

YOUR TASK

Repeat the sensitivity analysis that we have been working on in class BUT lets assume that we are in a different locations - where windpeeds are substantially higher and more variable AND vegetation is shorter - See details below
Consider the sensitivity of your estimate to uncertainty in the following parameters and inputs

- $height$
- k_d
- k_0
- v

Windspeeds v are normally distributed with a mean of 300 cm/s with a standard deviation of 50 cm/s
For vegetation height assume that height is somewhere between 3.5 and 5.5 m (but any value in that range is equally likely)
For the k_d and k_0 parameters you can assume that they are normally distributed with standard deviation of 1% of their default values

a. Use the Sobel approach to generate parameter values for the 4 parameters
b. Run the atmospheric conductance model for these parameters

c. Plot conductance estimates in a way that accounts for parameter uncertainty
d. Plot conductance estimates against windspeed use the parameter that is 2nd in terms of total effect on response
e. Estimate the Sobel Indices for your outut
f. Comment on what this tells you about how atmospheric conductance and its sensitivity to variation in windspped differs in this setting as compared to the setting that we examined in class where windspeed was lower and less variable and vegetation was taller.

Submit the Quarto on Canvas as usual

GRADING RUBRIC

PART I (Paper Examples)

- Discussion of the implication of parameter uncertainty from example paper (20pts)
  *explanation directly relates to a specific parameter (10pts)
  *explanation explores how parameter uncertainty might meaningfully impact results (10pts)

PART II (Atmospheric Conductance)

- Generation of parameter values using Sobol (10pts)
- Running model for the parameters (10pts)
- Graph of uncertainty of the response variable
  - meaningful graph (5pts)
  - graphing style (axis labels, legibility) (5 pts)
- Graph of relationship between output and windspeed
  - choice of color (see instructions) (5pts)
  - graphing style (axis labels, legibility) (5 pts)
- Computing Sobol Indicators (10 pts)
- Discussion (10pts)
  - correctly identifying how sensitivity to windspeed changed with setting (5pts)