Sensitivity using 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 Click to Download

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

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 samplng from apriori (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 as a function of windspeed, vegetation height and parameters

First lets get the parameters

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

# generate two examples of random number from parameter distributions

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

X1 = cbind.data.frame(k_o, k_d, v, height=height)

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

X2 = cbind.data.frame(k_o, k_d, v, height=height)

# there are different versions of sobol functions that have different approaches for estimating
        parameters and indices, we use an approach implemented by jansen

sens_Catm_Sobol = sobolSalt(model = NULL, X1, X2, nboot = 100)


# Take a look at the Sobol generated sensitivity object
# your parameters sets for sensitivity analysis are in X
```

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_Catm_Sobol,res, res.names="ga")

# main effect:  partitions variance (main effect without co-variance) - sums approximately to
        one
sens_Catm_Sobol$S
```

```
##      original          bias std. error   min. c.i.  max. c.i.
## X1 0.16770090 -0.0008131935 0.04128852  0.08696936 0.24463221
## X2 0.54305928 -0.0006958346 0.03325617  0.48010119 0.61600283
## X3 0.20404861 -0.0080903259 0.03431641  0.14048079 0.27659075
## X4 0.01008973 -0.0018972568 0.03111242 -0.05368225 0.07569566
```

```
# useful to add names
row.names(sens_Catm_Sobol$S) = colnames(parms)
sens_Catm_Sobol$S
```

```
##          original          bias std. error   min. c.i.  max. c.i.
## k_o    0.16770090 -0.0008131935 0.04128852  0.08696936 0.24463221
## k_d    0.54305928 -0.0006958346 0.03325617  0.48010119 0.61600283
## v      0.20404861 -0.0080903259 0.03431641  0.14048079 0.27659075
## height 0.01008973 -0.0018972568 0.03111242 -0.05368225 0.07569566
```

```
# total effect - accounts for parameter interactions
row.names(sens_Catm_Sobol$T) = colnames(parms)
sens_Catm_Sobol$T
```

```
##           original          bias   std. error   min. c.i.   max. c.i.
## k_o    0.245600252 -1.460095e-03 0.0125893615 0.218038258 0.272810571
## k_d    0.608941203  8.766847e-03 0.0379342633 0.522615446 0.659447329
## v      0.194229641 -3.038151e-03 0.0205420086 0.152559692 0.236797938
## height 0.004586504  3.264492e-06 0.0005510147 0.003490607 0.005766445
```

```
# Both the main effect and total effect can tell us something about how the parameter
       influences results


print(sens_Catm_Sobol)
```

```
##
## Call:
## sobolSalt(model = NULL, X1 = X1, X2 = X2, nboot = 100)
##
## Model runs: 6000
##
## Model variance: 19666435
##
## First order indices:
##          original          bias std. error   min. c.i.  max. c.i.
## k_o    0.16770090 -0.0008131935 0.04128852  0.08696936 0.24463221
## k_d    0.54305928 -0.0006958346 0.03325617  0.48010119 0.61600283
## v      0.20404861 -0.0080903259 0.03431641  0.14048079 0.27659075
## height 0.01008973 -0.0018972568 0.03111242 -0.05368225 0.07569566
##
## Total indices:
##           original          bias   std. error   min. c.i.   max. c.i.
## k_o    0.245600252 -1.460095e-03 0.0125893615 0.218038258 0.272810571
## k_d    0.608941203  8.766847e-03 0.0379342633 0.522615446 0.659447329
## v      0.194229641 -3.038151e-03 0.0205420086 0.152559692 0.236797938
## height 0.004586504  3.264492e-06 0.0005510147 0.003490607 0.005766445
```

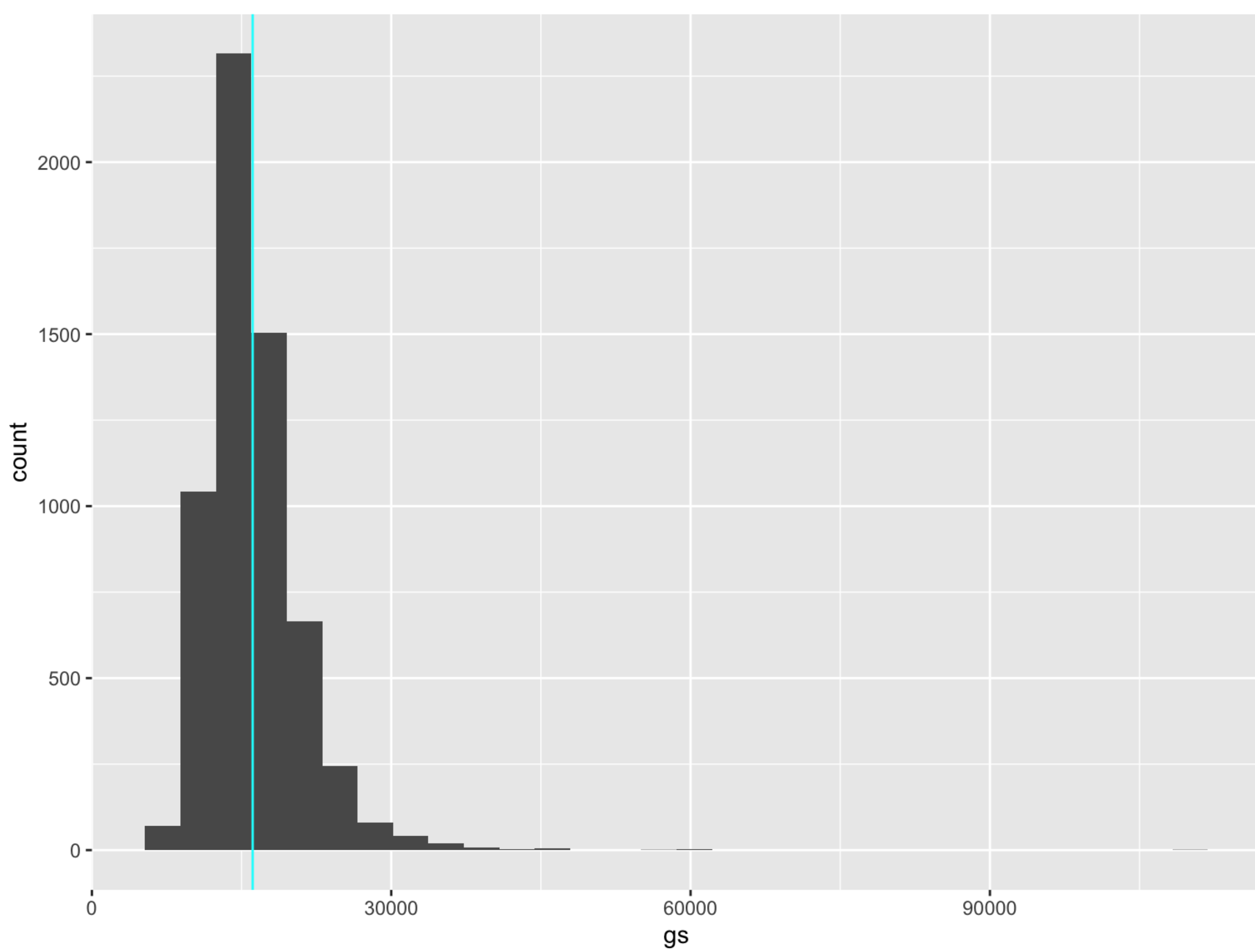
Plotting

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

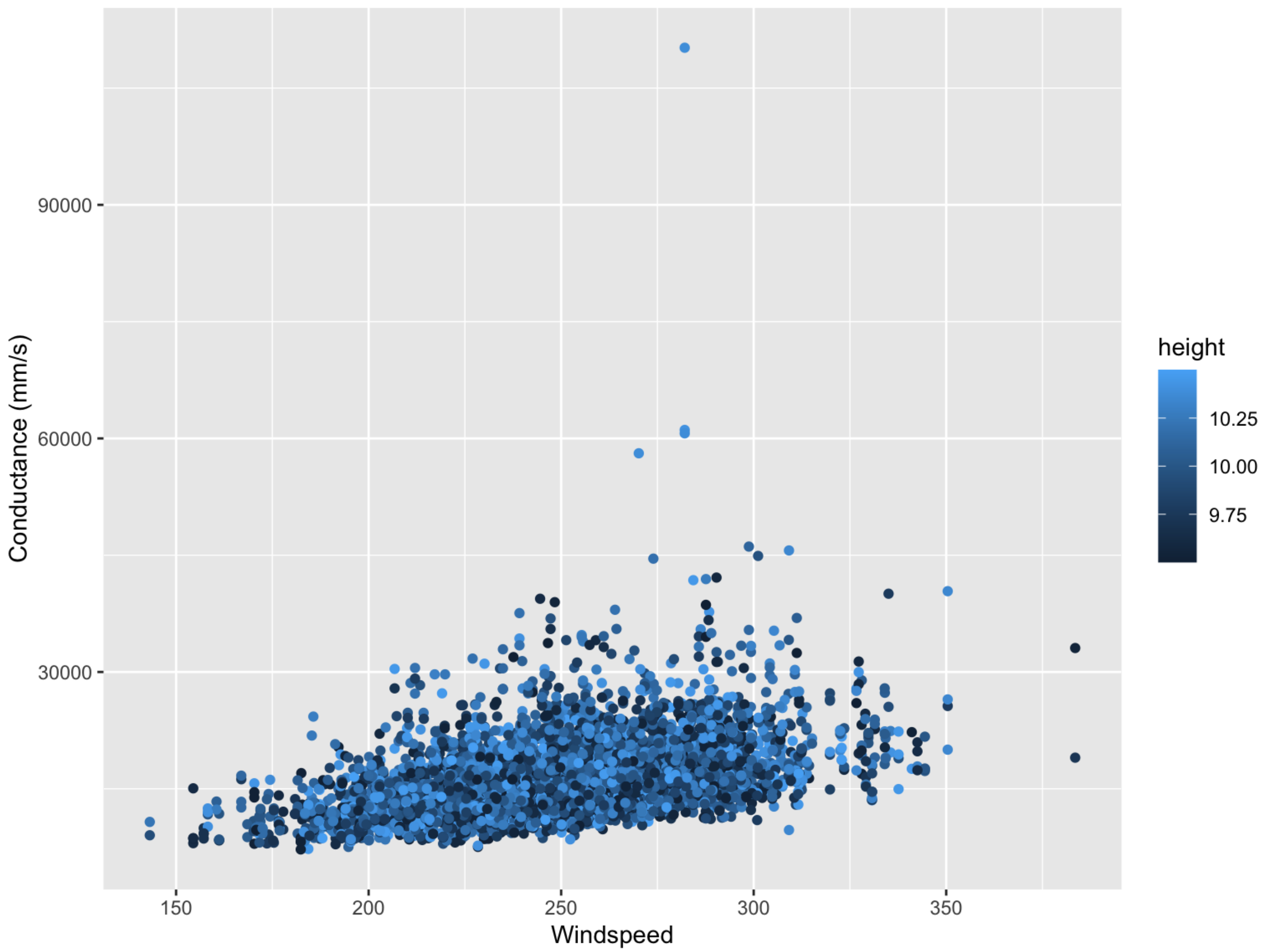
```
# graph two most sensitive parameters
both = cbind.data.frame(parms, gs=sens_Catm_Sobol$y)

# look at overall gs sensitvity to uncertainty
ggplot(both, aes(x=gs))+geom_histogram()+geom_vline(xintercept=mean(both$gs), col="cyan")
```

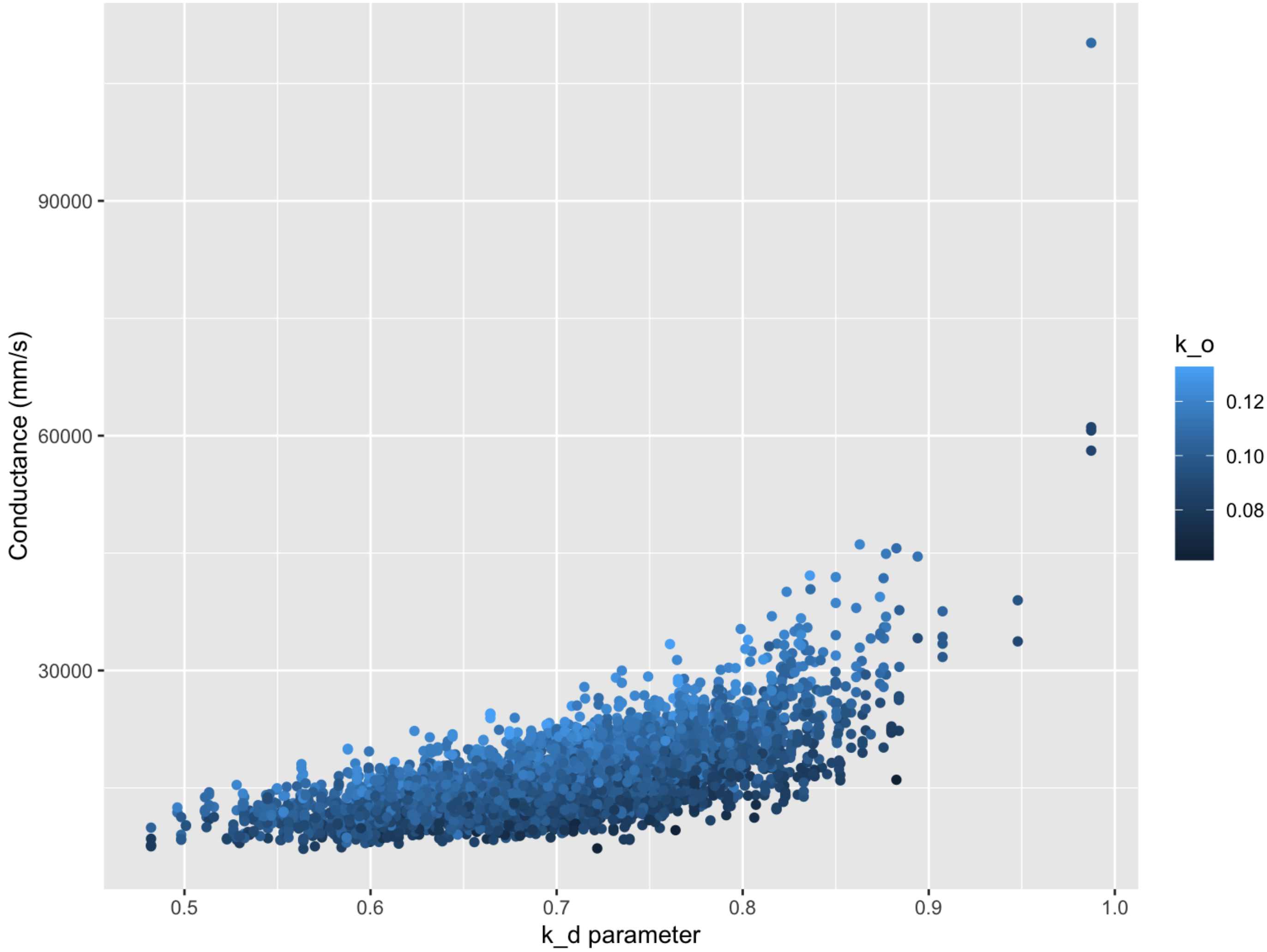
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



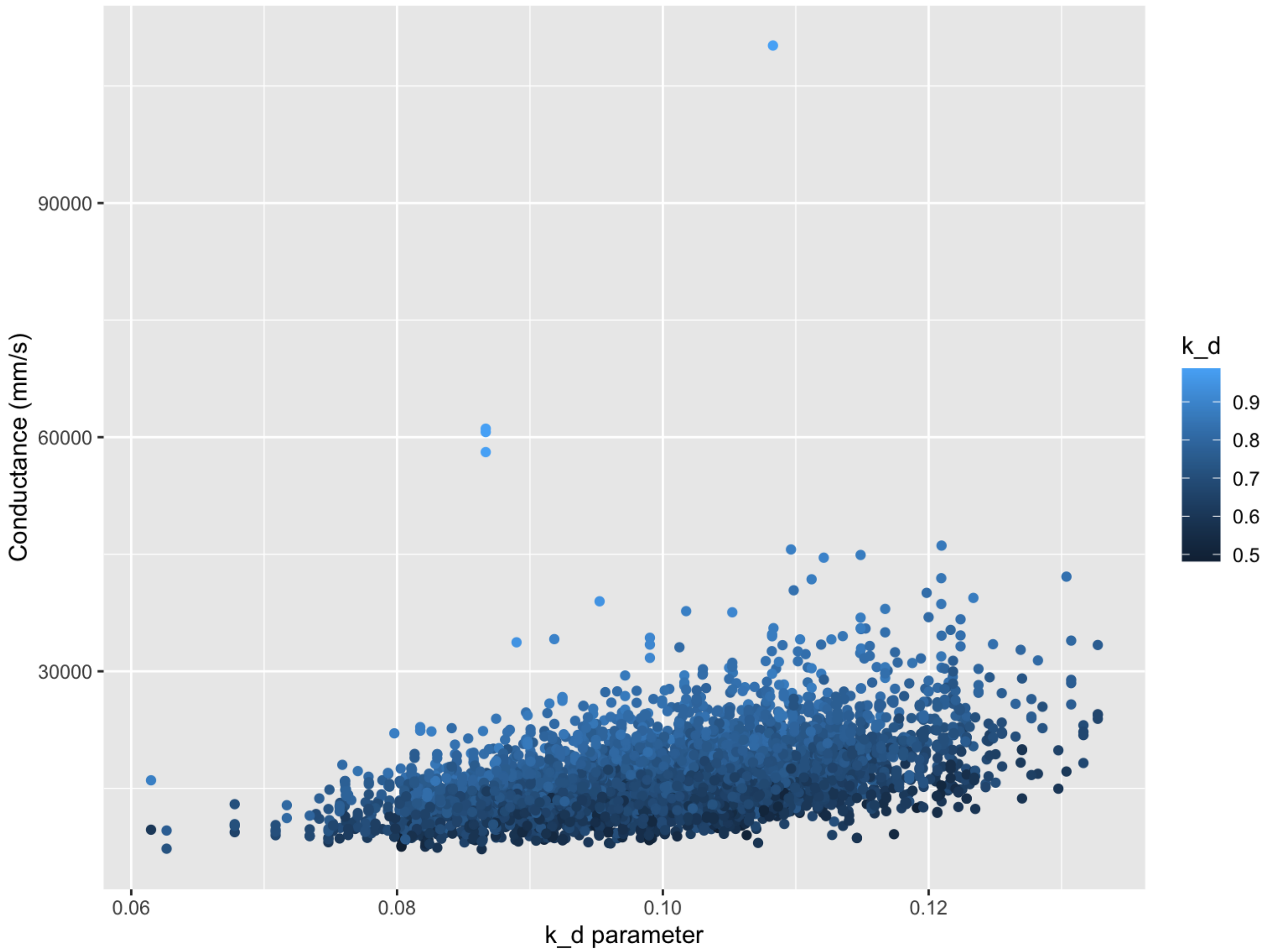
```r
# look at response of conductance to the two interesting variables
ggplot(both, aes(v,gs, col=height))+geom_point()+labs(y="Conductance (mm/s)", x="Windspeed")
```




```r
# look at response of conductance to the two most important variables
ggplot(both, aes(k_d,gs, col=k_o))+geom_point()+labs(y="Conductance (mm/s)", x="k_d parameter")
```




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


k_d
0.9
0.8
0.7
0.6
0.5
90000
60000
30000
Conductance (mm/s)
0.06
0.08
0.10
0.12
k_d parameter

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, X1, X2, nboot = 100, scheme="B" )

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


sens_Catm_Sobol2 = sensitivity::tell(sens_Catm_Sobol2,res, res.names="ga")

# main effect:  partitions variance (main effect without co-variance) - sums approximately to
      one
row.names(sens_Catm_Sobol2$S) = colnames(parms)
sens_Catm_Sobol2$S
```

```
##           original          bias std. error   min. c.i.  max. c.i.
## k_o    0.18414797  0.0026372728 0.03321070  0.1108856 0.24617377
## k_d    0.57658342 -0.0067165915 0.03112930  0.5264093 0.64878037
## v      0.20002299  0.0010380482 0.03136168  0.1438808 0.25913701
## height 0.01205946 -0.0006326814 0.03280418 -0.0522181 0.08455264
```

```
# total effect - accounts for parameter interactions
row.names(sens_Catm_Sobol2$T) = colnames(parms)
sens_Catm_Sobol2$T
```

```
##            original          bias   std. error   min. c.i.   max. c.i.
## k_o    0.251598063  1.521125e-03 0.0165170050 0.220242757 0.284639073
## k_d    0.629533694 -7.400534e-03 0.0477194939 0.530748588 0.720573410
## v      0.203111396  1.689281e-03 0.0137193075 0.174935224 0.235232140
## height 0.004606306  3.352621e-05 0.0003689713 0.003861689 0.005425464
```

```
# second order parameters interaction in controlling sensitivity
# parameters are in order, interactiosn are small here
sens_Catm_Sobol2$S2
```

```
##          original          bias std. error   min. c.i.  max. c.i.
## X1X2  0.062394996 -0.0046244998 0.04004442 -0.00397531 0.14262566
## X1X3 -0.000967171  0.0037616905 0.04013951 -0.09297931 0.06384114
## X1X4 -0.011540303  0.0011455646 0.03296566 -0.08591573 0.05362705
## X2X3 -0.045766702  0.0039409514 0.03696695 -0.12675988 0.02054095
## X2X4 -0.007286591  0.0001570271 0.03309789 -0.07795060 0.05714003
## X3X4 -0.007752883  0.0005991450 0.03304242 -0.08117017 0.05554077
```

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

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

3. Plot conductance estimates in a way that accounts for parameter uncertainty

4. Plot conductance estimates against windspeed use the parameter that is 2nd in terms of total effect on response

5. Estimate the Sobel Indices for your outut

6. 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 Rmarkdown 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 Sobel (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 Sobel Indicators (10 pts) * Discussion (10pts) * correctly identifying how sensitivity to windspeed changed with setting (5pts)