Quantifying how uncertainty in fuel loading maps propagates to uncertainty in emissions estimates: A sensitivity analysis of the CONSUME and FOFEM models.

Introduction:

Emissions estimates, fuel loading maps, applications

Background on database and uncertainty in fuel loading maps

Background on SA, including local v. global and define uncertainty and sources

What about correlated variables? Can’t sample independently.

Flaming v. smoldering

Objectives:

1. Quantify how uncertainty in fuel loading propagates to uncertainty in emission predictions
2. Identify the fuel types for which emission predictions are most sensitive

Objective 1 helps to inform planning, enabling the estimation of a prediction interval for emissions for a given management action. This would be a, say, 95% interval around which we believe plausibly the actual emissions will be. Objective 2, in tandem with the fuels loading database, informs where resources would be well-spent to improve precision in emissions input data. Resources would be better spent on more sensitive data inputs that are missing coverage in the database. Furthermore, more precise measurements of those fuels types for a given application (time and place) would improve emission predictions.

Methods

Database development and reference

Consume and FOFEM

Test Case EVTs (from Database paper) and variable groupings; baseline fuelbeds

Target outputs, and environmental variable settings.

Global SA: Sobol

Sampling method: correlated inputs

Let N be the number of random samples (N=1000), and K be the number of fuel types that are targeted for the SA (e.g., K = 6 for flaming fuel types). First N independent samples for each fuel type are generated from the best fitting marginal empirical distribution from the Prichard et al. fuels database for each target EVT, to create a matrix X (NxK). We then estimate Spearman’s rank correlation structure on the group of fuel types for the test case EVT groups, using only complete cases to estimate the correlation matrix (only cases for which all variables of interest are entered). We use the matrix method of (Iman & Conover, 1982) to approximate a given rank correlation structure for sampled input data, resulting in an NxK sampled input matrix X\* with correlation structure similar to the empirical correlation structure (see supplementary methods).

Quantifying uncertainty in emissions estimates

For the sampled matrix X\* we use the baseline fuelbed to fill in fuel types required for each model (see supplementary) and create an input file suitable for each model. We then use the model to estimate fuel consumption and emissions for each row in our input file, for the given environmental conditions. From the resulting model output file we characterize the distribution of predicted emissions (using boxplots and 95% intervals). This provides a link between the uncertainty in fuel loading as represented by the database, and the uncertainty in emission predictions.

Sensitivity Analysis

The goal of SA is to identify the fuel inputs with the strongest impact on emissions estimates. We will use both Sobol variance partitioning (vp) and partial rank correlation coefficients (prcc).

For the Sobol variance partitioning procedure we divide the NxK matrix X\* into two new (N/2)xK matrices, X1 and X2, where X1 is the first N/2 rows in X\* and X2 is the last N/2 rows in X\*. We then use the sobolEff function in the R sensitivity package (Pujol et al., 2017), where the function rearranges the X1 and X2 to create a new data input matrix. For example, for N=1000, the final data input X has 3500 rows.. Note that this exchanges the columns in X1 and X2, and the rank correlation structure of the final data input matrix has a poorer approximation of the original correlation matrix (C; Supplementary; might want to look into this, might have dampened correlations relative to empirical).

Variance partitioning may not provide valid partitions in the case of correlated inputs (ref, one of the Saltelli books). We complement the variance partitioning results with partial rank correlation coefficients (prcc; ref), which provide a non-parametric estimate of the correlation between each input and each output, accounting for any relationships among the inputs and between the remaining inputs and the output. As with standard correlations, the prcc takes values between -1 and 1. We will rank the inputs by their importance in variance partitioning, and by those with the greatest magnitude prcc. We will apply a bootstrap procedure for both indices to provide bootstrap standard errors and confidence intervals for the index value for each model input. Any inconsistent rankings between the two procedures will be evaluated further.

Environmental inputs?

Results

Correlation structures

95% prediction intervals for each output based on global sampling

Sensitivity indices

Discussion

Point estimates and predictions are insufficient

Much uncertainty in fuel loading data gives uncertainty in emissions predictions

Emissions are sensitive to all fuel loading categories when considered individually, but when taken as a whole the most sensitive fuel inputs are \_\_\_\_ for flaming and \_\_\_\_ for smoldering.

Sensitivities are themselves sensitive to environmental

Require better data to more precisely predict emissions. So many holes! Particularly to estimate correlation matrices

If OAT nearly all correlations are near 1—want to consider the system of fuels, not just one at a time.

**Supplementary Methods**

We adapt the method of (Iman & Conover, 1982) by first estimating Spearman’s rank correlation structure on the group of fuel types for the test case EVT groups, using only complete cases to estimate the correlation matrix (only cases for which all variables of interest are entered). Let C be the empirical correlation matrix (KxK). P’ is the Cholesky factorization of C (generated by the R chol function), and P is the transpose of the resulting matrix.

Define the scores a(i) (i=1,…,N):

Where φ-1 is the inverse of the cumulative distribution function for the standard normal distribution. Then we create a new NxK matrix R, where each column is an independent random sample (without replacement) from the vector of scores A (comprised of a(i) above). We then generate the matrix R\* as RP’. The matrix R\* will have a rank correlation structure similar to C.

To reduce the variability in the correlation in the sampled data input matrix, a new matrix correlation matrix (T) is generated as the Spearman rank correlation matrix for the matrix R\*. We then calculate Q’ as the Cholesky factorization of T, with Q the transpose of Q’. We solve for the matrix S = PQ-1, and R\*b = RS’. We then rearrange each column in the sampled data input matrix X to match the rankings in the matrix R\*b. For example, if the first row in R\*b is the 5th ordered statistic, then we place the 5th ordered statistic for that column in X in the first row (call the resulting matrix X\*). This preserves the overall marginal distribution for each column X\*, while approximating the rank correlation structure across each row in X\*.