A Metamodelling Approach to Estimate Global N\textsubscript{2}O Emissions from Agricultural Soils
Joshua Perlman\textsuperscript{1}, Robert Hijmans\textsuperscript{1}, William Horwath\textsuperscript{2}
\textsuperscript{1}Department of Environmental Science and Policy, University of California, Davis (jperlman@ucdavis.edu)
\textsuperscript{2}Department of Land, Air, and Water Resources, University of California, Davis

Introduction
Metamodels are simplifications of complex models that can be used to better understand the effect of the main driving variables and to reduce computation time. We developed metamodels of the “DeNitrification DeComposition” (DNDC) model for N\textsubscript{2}O emissions from maize and wheat. We ran DNDC for a diverse sample of global climate and soil types, and fitted the model output as a function of (sometimes simplified) model input variables, using the Random Forest machine learning algorithm.

Materials and Methods
1) Obtain global data sets and run DNDC for a sample of locations, 2) Fit Random Forest models (Breiman 2001) to DNDC input and output, 3) Use the Random Forest metamodels to make global predictions of N\textsubscript{2}O emissions at high resolution (~1km) and test the effects of different treatment and aggregation of input data (e.g. soils).

Results
- R\textsuperscript{2} between holdout data and metamodel predictions was 0.96 for maize and 0.94 for wheat.
- Predicting with metamodels is \(\sim 4.5 \times 10^4\) times faster than running DNDC simulations.
- Metamodels show that N\textsubscript{2}O emissions with DNDC are highly sensitive to soil organic carbon (SOC) content.
- Aggregation of soil data has a substantial effect on emission estimates.

Table 1: Summary of global emission estimates (Mt CO\textsubscript{2}-e)

<table>
<thead>
<tr>
<th></th>
<th>Maize</th>
<th>Wheat</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total emissions (Mt)</td>
<td>157.38</td>
<td>184.91</td>
<td>342.29</td>
</tr>
<tr>
<td>% Total global emissions</td>
<td>0.36</td>
<td>0.42</td>
<td>0.78</td>
</tr>
<tr>
<td>% Global agricultural emissions</td>
<td>2.58</td>
<td>3.03</td>
<td>5.62</td>
</tr>
<tr>
<td>% Global agricultural soil emissions</td>
<td>6.85</td>
<td>8.05</td>
<td>14.91</td>
</tr>
<tr>
<td>Average emission kg/ha</td>
<td>1038.63</td>
<td>861.81</td>
<td>950.22</td>
</tr>
<tr>
<td>Average emissions kg (grain)</td>
<td>0.23</td>
<td>0.32</td>
<td>0.28</td>
</tr>
</tbody>
</table>

Figure 2: Variable importance plots

Figure 3: Modeled maize emissions (kg N\textsubscript{2}O-N/ha) with A) uniform fertilization of 50 kg N/ha, and B) variable fertilization (shown in figure 1B)

Figure 4: Range of modeled wheat emissions (kg N\textsubscript{2}O-N/ha) across different soil types within grid cells

Conclusions
- Random forest metamodels of N\textsubscript{2}O emissions from maize and wheat effectively reproduced emission outputs obtained with DNDC.
- Metamodels provide concise insight about associations between DNDC inputs and outputs.
- Increased speed allows for high resolution modeling at a global scale.
- High resolution is important due to the strong effects of soil data aggregation on N\textsubscript{2}O estimates.
- Loss of accuracy in metamodels is more than compensated for by increase in resolution.

References