Enhancing Model-Data Symbiosis to Improve Global Carbon-Sink Predictions

Stuart Muller and Stefan Gerber

To predict future levels of atmospheric carbon dioxide (CO2), and thus climate change, it is crucial to determine and understand the response of the land biosphere to global environmental changes (temperature, CO2, nitrogen deposition, land-use, etc). Dynamic global land models (DGLMs) are used to predict the long-term carbon (C) uptake by terrestrial vegetation and soil (C-sink), but considerable disparity exists between the results obtained using different models. Constraining this divergence is necessary to reduce the uncertainty in our projections. To this end, recent model refinements have attempted to account for the important role that nitrogen (N) limitation plays in primary productivity. However, DGLMs rely on vast amounts of data for tuning and testing, and thus far the focus has been on C. There is consequentially a paucity of global N data to support model benchmarking. Additionally, it remains unclear which specific benchmarks are more useful for improving long-term C-sink predictions. Ideally, we strive to use model simulations and data collection symbiotically, each benefiting the other, and reciprocating in turn. Data collection strategies form the foundation of our understanding; data therefore informs model design and is necessary to assess the validity of simulation results. Models in turn permit us to conduct artificial experiments beyond the scope of physical practicability, which elucidate gaps in our understanding or in the data resources necessary for improved simulations. We seek to optimize this exchange for our particular modeling purposes by identifying the most potent data we have, and the most valuable data we should like to acquire. It would therefore be valuable to know 1) which benchmark data offer the greatest potential for constraining C-sink predictions, and 2) which variables should be measured, or measured better, to support these benchmarks, with particular attention paid to N in both cases. This work presents a conceptual schema for framing the problem, and a practical strategy, based on sensitivity analysis, to identify the existing and potential data domains that offer optimal returns on resource investment.