

Development of a Data-Assimilation Framework for Integrating 25 Years of Surface and Airborne observations to assess patterns of net CO₂ Exchange from Arctic Ecosystems: Data Management Plan

J.W. Munger and S.C. Wofsy

School of Engineering and Applied Sciences, Harvard University

The following Data Management Plan was part of the NASA ROSES 2012 Proposal Development of a Data-Assimilation Framework for Integrating 25 Years of Surface and Airborne observations to assess patterns of net CO₂ Exchange from Arctic Ecosystems ([summary](#)) submitted to the Terrestrial Ecology Program. It is presented as an example plan.

Data Management Plan

Our group is strongly committed to providing open access to data and model products. Ongoing and past projects post all data on local ftp data servers and submit to relevant project data archives. For example, as lead PI for HIPPO, Co-I S. Wofsy led the effort to make HIPPO data universally available and to develop highly accessible merged products from many different investigators (see <http://hippo.ornl.gov>). We intend to make our collated, synthesized data sets and WRF/STILT model products available in the most user-friendly format possible, probably through the ORNL DAAC or NACP data server. Careful attention will be given to making accessible large data sets such as the WRF/STILT product, and linking seamlessly to related data portals such as the CARVE download site.

The major new data products that will be generated in this project are the WRF meteorological fields and STILT footprints. These will be formatted in flat ascii and netCDF-4 files as appropriate for their size. During the project, data files will be stored locally on servers that are routinely backed up to external locations. Active files will be protected by automated routine backups capture incremental changes and maintain stable monthly and weekly versions. Large input data files and model outputs that are generated once will be manually duplicated on external storage media before archiving at the ORNL DAAC.

All model outputs will be made publicly available as soon as they pass quality assurance checks (one year or less) on our local ftp servers as well as uploaded to an appropriate data archive center or project archive. Metadata sufficient to describe how the meteorological fields were generated and how to use the footprints will be included.

Additional data will be generated by synthesis from existing publicly available data sets. With permission from the data generators we will post on our project web site links to the original data that we use as well as provide access to any derived or value added products such as a priori flux estimations and VPRM parameters. Computer code (generally written in R) associated with the STILT-VPRM assimilation along with Readme files to document the model structure will be made available on the project ftp server. Stable releases will be made on the public site at the ORNL DAAC or NACP. Our experience in HIPPO provides the background for this: the complete, end-to-end data processing script for HIPPO (individual investigator files => 5-mission merged files) was delivered and successfully demonstrated to be environment-independent.