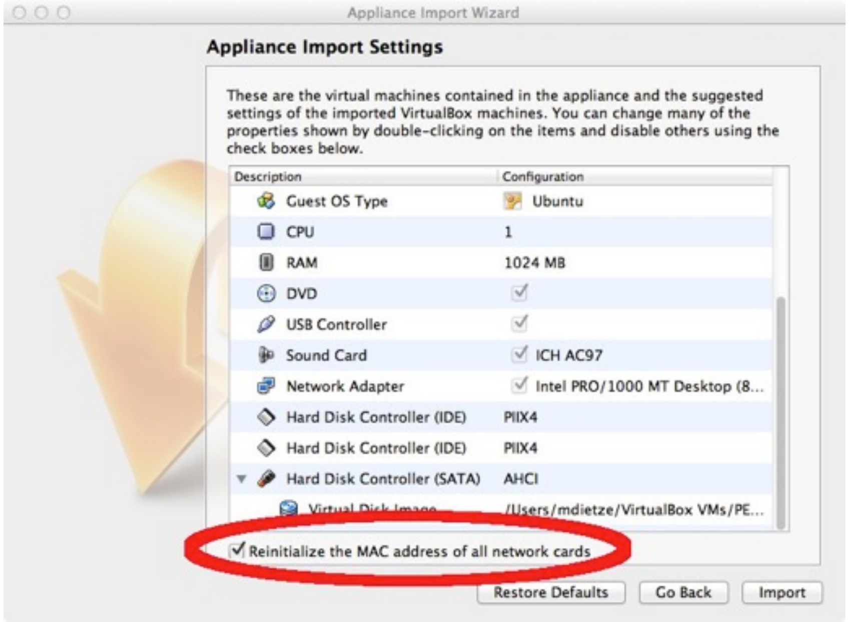
PEcAn v1.4.3 Hands-On Demo

Option 1: Installing PEcAn VM on your local computer:

1. PEcAn consists of a set of scripts and code that is compiled within a Linux operating system and saved in a "virtual machine (VM)". Virtual machines allow for running consistent set-ups without worrying about differences between operating systems, library dependencies, compiling the code, etc.
2. To run the PEcAn VM you will need to install VirtualBox, the program that runs the virtual machine (http://www.virtualbox.org). On Windows you may see a warning about Logo testing, it is okay to ignore the warning.
3. After you have Virtual Box installed you'll need to download the PEcAn virtual machine: http://opensource.ncsa.illinois.edu/projects/artifacts.php?key=PECAN. The virtual machine is available under the tag vm. Download the 32 bit version and note that the download is ~4 GB so will take from several minutes to hours depending on the connection speed.
4. To open up the virtual machine you'll first want to open up VirtualBox



Appliance Import Settings

These are the virtual machines contained in the appliance and the suggested settings of the imported VirtualBox machines. You can change many of the properties shown by double-clicking on the items and disable others using the check boxes below.

Description	Configuration
Guest OS Type	Ubuntu
CPU	1
RAM	1024 MB
DVD	✓
USB Controller	✓
Sound Card	✓ ICH AC97
Network Adapter	✓ Intel PRO/1000 MT Desktop (8...
Hard Disk Controller (IDE)	PIIX4
Hard Disk Controller (IDE)	PIIX4
Hard Disk Controller (SATA)	AHCI
Virtual Disk Image	/Users/mdietze/VirtualBox VMs/PE...

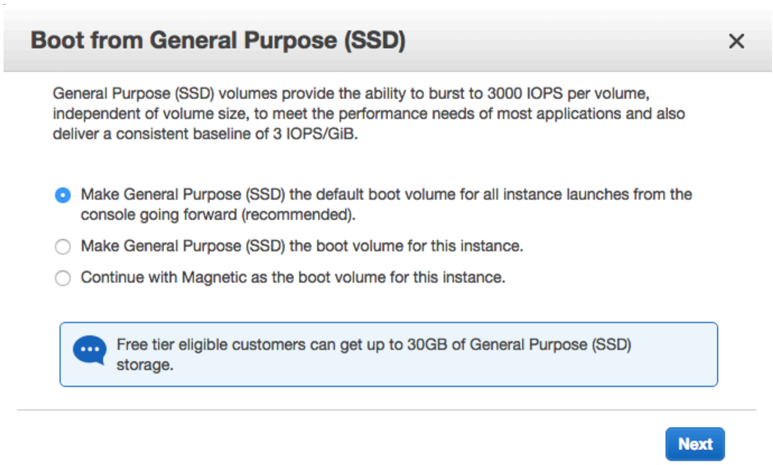
☑ Reinitialize the MAC address of all network cards

5. The first time you use the VM you'll want to use File → Import Appliance in VirtualBox in order to import the VM. This will create a virtual machine from the disk image. When asked about the Appliance Import Settings **make sure you select "Reinitialize the MAC address of all network cards"**. This is not selected by default and can result in networking issues since multiple machines might claim to have the same network MAC Address. That said, users who have experienced network connection difficulties within the VM have sometimes had better luck after reinstalling without reinitializing.
6. Next, click "Import". You only have to do this Import step once, in the future you can just skip to the next step.
7. Start Pecan by double clicking on the icon for the VM. A terminal window will pop up showing the machine booting up which may take a minute. It is done booting when you get to the "pecan32 login:" prompt. You do not need to login as the VM behaves like a server that we will be accessing through you web browser. Feel free to minimize the VM window.

Option 2: Using the Amazon Cloud

Login to Amazon Web Services (AWS), http://console.aws.amazon.com/ and select the EC2 Dashboard. If this is your first time using AWS you will need to set up an account before you are able to access the EC2 Dashboard. Important: You will need a credit card number and access to a phone to be able to verify AWS account registration. AWS is free for one year.

1. **Choose AMI**
   a. On the top right next to your name, make sure the location setting is on U.S. East (N. Virginia), not U.S. West (Oregon)
   b. On the left click, click on EC2 (Virtual servers), then click on "AMIs", also on the left
   c. In the search window toggle to change "Owned by me" to "Public images"
   d. Type "pecan" into the search window
   e. Click on the toggle button on the left next to PEcAn1.4.3
   f. Click on the "Launch" button at the top
2. **Choose an Instance Type**
   a. Select what type of machine you want to run. For this demo the default, **t2.micro**, will be adequate. Be aware that different machine types incur very different costs, from 1.3 cents/hour to over $5/hr https://aws.amazon.com/ec2/pricing/
   Select t2.micro, then click "Next: Configure Instance Details"
3. **Configure Instance Details**
   a. The defaults are OK. Click "Next: Add Storage"
4. **Add Storage**
   a. The defaults are OK. Click "Next: Tag Instance"
5. **Tag Instance**
   a. You can name your instance if you want. Click "Next: Configure Security Group"
6. **Configure Security Group**
   a. You will need to add two new rules:
      i. Click "Add Rule" then select "HTTP" from the pull down menu. This rule allows you to access the webserver on PEcAn.
      ii. Click "Add Rule", leave the pull down on "Custom TCP Rule", and then change the Port Range from 0 to 8787. Set "Source" to Anywhere. This rule allows you to access RStudio Server on PEcAn.
   b. Click "Review and Launch" . You will then see this pop-up:



Boot from General Purpose (SSD)

General Purpose (SSD) volumes provide the ability to burst to 3000 IOPS per volume, independent of volume size, to meet the performance needs of most applications and also deliver a consistent baseline of 3 IOPS/GiB.

● Make General Purpose (SSD) the default boot volume for all instance launches from the console going forward (recommended).

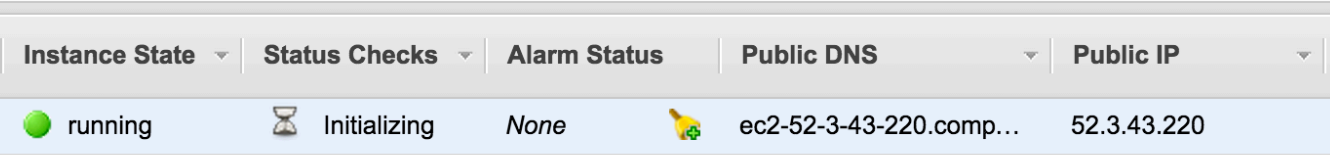
○ Make General Purpose (SSD) the boot volume for this instance.

○ Continue with Magnetic as the boot volume for this instance.

Free tier eligible customers can get up to 30GB of General Purpose (SSD) storage.

Select the default drive volume type and click Next

7. **Review and Launch**
    a. Review the settings and then click “Launch”, which will pop up a select/create Key Pair window.
8. **Key Pair**
    a. Select “Create a new key pair” and give it a name. You won’t actually need this key unless you need to SSH into your PEcAn server, but AWS requires you to create one. Click on “Download Key Pair” then on “Launch Instances”. Next click on “View Instances” at the bottom of the following page.
9. **Instances**



Instance State	Status Checks	Alarm Status	Public DNS	Public IP
● running	⌛ Initializing	*None*	ec2-52-3-43-220.comp…	52.3.43.220

    a. You will see the status of your PEcAn VM, which will take a minute to boot up. Wait until the Instance State reads “running”. The most important piece of information here is the Public IP, which is the URL you will need in order to access your PEcAn instance from within your web browser (see Demo 1 below).
    b. Be aware that it often takes ~1 hr for AWS instances to become fully operational, so if you get an error when you put the Public IP in you web browser, most of the time you just need to wait a bit longer.

Congratulations! You just started a PEcAn server in the “cloud”!

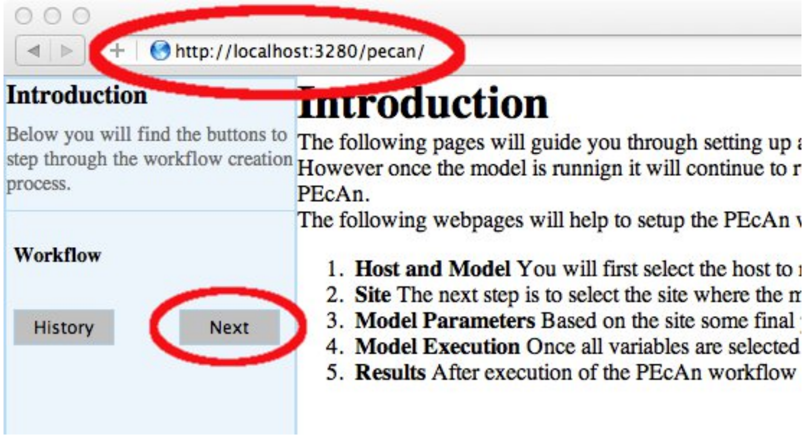
10. When you are done using PEcAn, you will want to return to the “Instances” menu to turn off your VM.
    a. To STOP the instance (which will turn the machine off but keep your work), select your PEcAn instance and click Actions > Instance state > Stop. Be aware that a stopped instance will still accrue a small storage cost on AWS. To restart this instance at any point in the future you do not want to repeat all the steps above, but instead you just need to select your instance and then click Actions > Instance state > Start
    b. To TERMINATE the instance (which will DELETE your PEcAn machine), select your instance and click Actions > Instance state > Terminate. Terminated instances will not incur costs. In most cases you will also want to go to the Volumes menu and delete the storage associated with your PEcAn VM.Remember, AWS is free for one year, but will automatically charge a fee in second year if account is not cancelled.

DEMO 1: Single Run

We will begin by exploring a set of web-based tools that are designed to run single-site model runs. A lot of the detail about what's going on under the hood, and all the outputs that PEcAn produces, are left to Demo 2. This demo will also demonstrate how to use PEcAn outputs in additional analyses outside of PEcAn.

1. **Start PEcAn**

   To access the web pages you can use a web browser on your own computer's desktop and connect to the following address http://localhost:3280/pecan
   Note: If you downloaded the 64-bit version you will want to connect to port 6480
   If you are working on AWS you will use the provided IP address as the URL, but add /pecan to the address (e.g., http://52.4.1.158/pecan)



http://localhost:3280/pecan/

Introduction

Below you will find the buttons to step through the workflow creation process.

Workflow

Introduction

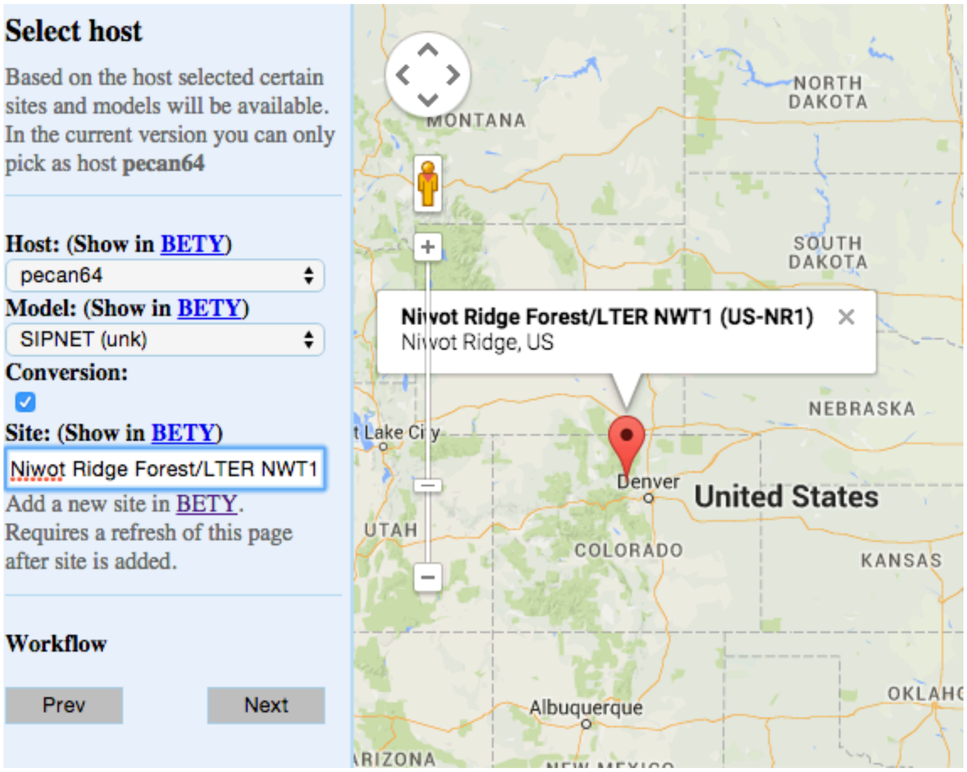
The following pages will guide you through setting up a
However once the model is runnign it will continue to r
PEcAn.
The following webpages will help to setup the PEcAn w

1. **Host and Model** You will first select the host to r
2. **Site** The next step is to select the site where the m
3. **Model Parameters** Based on the site some final
4. **Model Execution** Once all variables are selected
5. **Results** After execution of the PEcAn workflow

From the Introduction page, click the 'Next' button to move to the "Site Selection" page.

There are three steps to running PeCAn: selecting a site, specifying drivers, and running the model, as detailed below.

2. Site Selection



Select host

Based on the host selected certain sites and models will be available. In the current version you can only pick as host **pecan64**

Host: (Show in BETY) pecan64

Model: (Show in BETY) SIPNET (unk)

Conversion: ☑

Site: (Show in BETY) Niwot Ridge Forest/LTER NWT1

Add a new site in BETY. Requires a refresh of this page after site is added.

Workflow

Niwot Ridge Forest/LTER NWT1 (US-NR1)
Niwot Ridge, US

1. **Select Host:** For this exercise you will want to select the local machine "pecan32" or "pecan64" to see the sites that are stored locally on the VM and have driver data. PEcAn is designed to conduct runs both on the local VM and on remote machines. By selecting different servers from the "HOST" pull down menu you will be able to see what sites exist within the PEcAn databases on different servers. As the database also records sites where data was collected, not all sites will have the necessary driver files to run every model.
2. **Select Model:** Next, you'll want to select the ecosystem model you want to run. We'll begin by using SIPNET because it is a quick, simple model.
3. **Conversion:** By default, PEcAn will only display sites where all model drivers already exist in the system database. If you select the conversion check box, the view will switch to all sites that PEcAn is capable of generating model drivers for automatically. Go ahead and click on it.
4. **Select Site:** Finally, you'll want to select the site where you want to run the model by clicking on a pin icon on the Google Map. You can also search and filter sites by typing in a site name or part of a site name (e.g. typing US- will filter the Ameriflux sites, which all have ID's that start US-). Feel free to zoom out to examine sites outside the initial frame, or to zoom in to examine clusters of sites or to see exactly where a site is located. When you click on a site's flag on the map, it will give you the name and location of the site. It will also put that site in the "Site:" box on the left hand site, which indicated its your current site selection. For this tutorial, type US-NR1 in the search box to display the Niwot Ridge Ameriflux site (US-NR1), and then click on the pin icon.

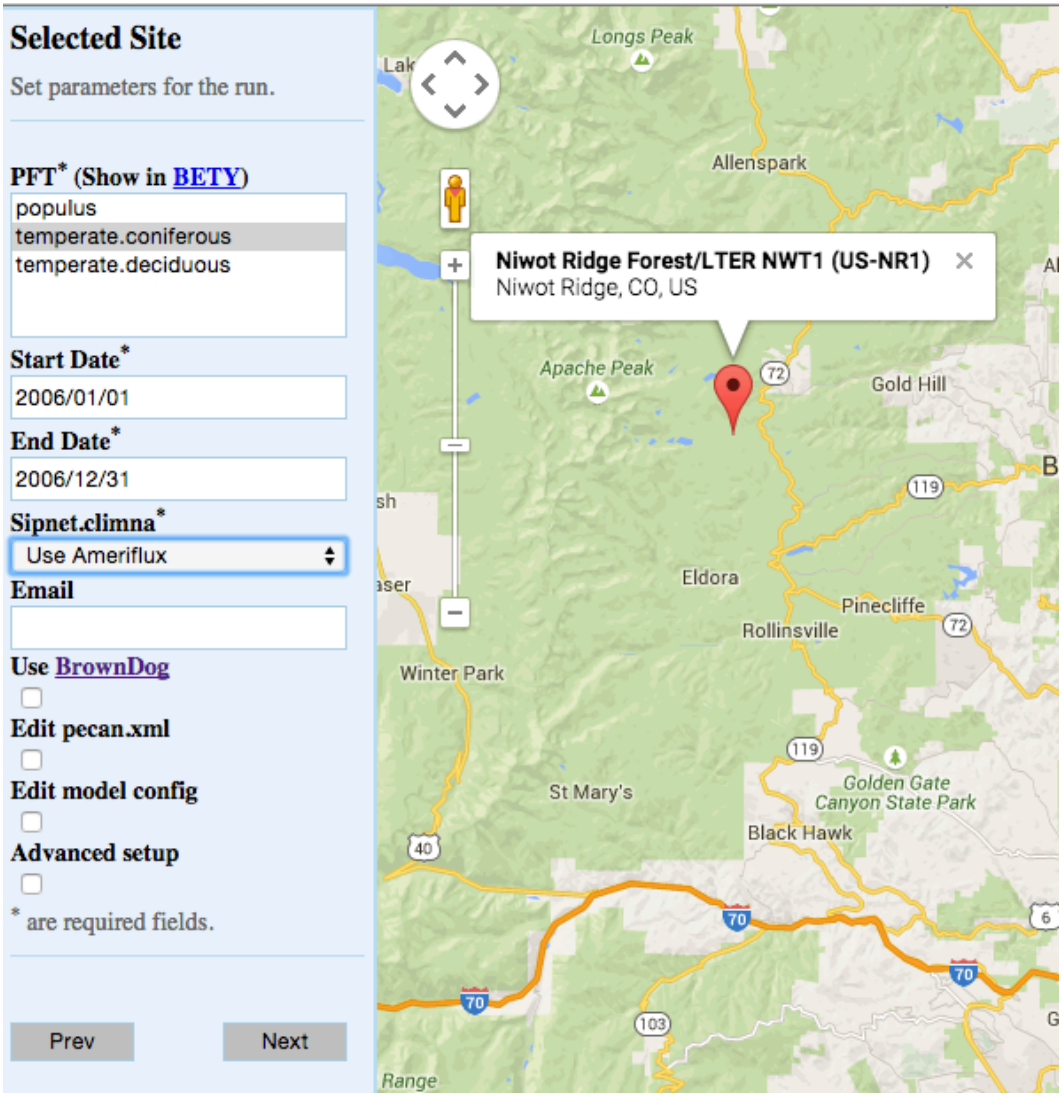
5. **Advanced features (Optional):** leave all blank for demo run

    1. Throughout PEcAn the **Show in BETY** links will redirect you to the web front end for the underlying database where you can see the detailed entries for various selections (e.g. Host server, Model, Site, PFT, etc.)
    2. The **Add a new site in BETY** will redirect you to the new site page in the database. This will allow you to add more pins to the map for your own study sites. See Demo 3 for more details.

6. Once you have selected a Host, Site, and Model, click “Next”

3. Run Specification

Next we will specify settings required to run the model. Be aware that the inputs required for any particular model may vary somewhat so there may be addition optional or required input selections available for other models.



Selected Site

Set parameters for the run.

PFT* (Show in BETY)

populus
temperate.coniferous
temperate.deciduous

Start Date*

2006/01/01

End Date*

2006/12/31

Sipnet.climna*

Use Ameriflux

Email

Use BrownDog

☐

Edit pecan.xml

☐

Edit model config

☐

Advanced setup

☐

* are required fields.

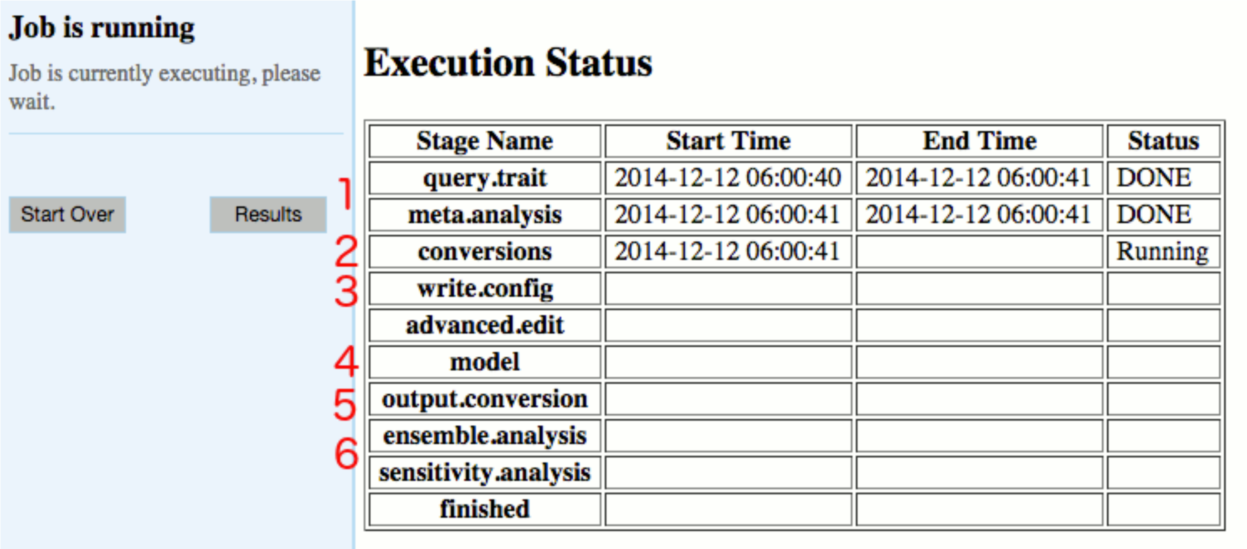
Niwot Ridge Forest/LTER NWT1 (US-NR1)
Niwot Ridge, CO, US

1. **PFT:** You'll want to select the Plant Functional Type that you want to run from the PFT menu. The list of available PFTs will vary depending upon which model you selected. Some models, such as ED2 and LINKAGES, allow for competition between PFTs within a site and thus selecting multiple PFTs is allowed (and in fact encouraged). SIPNET and DALEC currently only runs one PFT at a time so you will want to only select a single PFT. Niwot

Ridge is temperate conifer.

2. **Start/End Date**: Select the dates you want to run the model. At the moment PEcAn does not automatically enforce that selected dates correspond to the date ranges of available inputs (e.g. weather data) so be aware that this may lead to errors later in the workflow. For this tutorial, select 2006/01/01 to 2006/12/31.
3. **Weather Data:** This pull-down (Sipnet.climna for SIPNET) will allow you to select the meteorological driver dataset that you want to use by name and time span. Data sets that are already stored locally will have a name and data range. Data sets that can be automatically downloaded and processed will say "Use" (e.g. Use Ameriflux, Use NARR). For this analysis select "Use Ameriflux"
4. **Optional Settings:** leave all blank for demo run
    1. **Email** will send a message to the specified address when the run is complete. This is handy for slower model or for more complex analyses that involve many model runs, but is not required for this demo.
    2. **Use Brown Dog** will use the Brown Dog web services in order to do input file conversions. (*Note: this is required if you select Use NARR for Weather Data*)
    3. **Edit pecan.xml** will allow you to modify the PEcAn settings file before the PEcAn workflow is executed. This feature is used to configure advanced settings and turn on additional analyses that are not web-based (e.g. data assimilation). Full details on the PEcAn.xml can be found at (**https://github.com/PecanProject/pecan/wiki/PEcAn-Configuration**).
    4. **Edit model config** will pause the workflow after PEcAn has written all model specific settings but before the model runs are called. This allows the user to configure any additional settings internal to the model. Please check with individual modeling teams for information about the details and documentation of such files. Finally, this would also allow users to modify model parameters by hand, however as we'll discuss below this is not encouraged since PEcAn is estimating model parameters on demand based on it's internal trait databases. Adding empirical data to this trait database is the preferred methods of informing model parameters.
    5. **Advanced Setup** will open another window to select ensemble and sensitivity run settings. We will revisit this setting in part 2.
5. Finally, click **Next** to start the model run. You will see a data policy statement if you selected a data source with a policy. Agreeing to the policy is required prior to starting the run.

4. Model Run Workflow



Job is running

Job is currently executing, please wait.

Execution Status

Stage Name	Start Time	End Time	Status
query.trait	2014-12-12 06:00:40	2014-12-12 06:00:41	DONE
meta.analysis	2014-12-12 06:00:41	2014-12-12 06:00:41	DONE
conversions	2014-12-12 06:00:41		Running
write.config			
advanced.edit			
model			
output.conversion			
ensemble.analysis			
sensitivity.analysis			
finished			

SIPNET will execute the model run fast enough that you are liable to miss some of the following steps, but the PEcAn workflow will automatically run the following, indicating model progress at each step

1. **query.trait / meta.analysis**: PEcAn estimates model parameters by performing a meta-analysis of the available trait data for a PFT. Query.trait will extract relevant trait data from the database. Meta.analysis performs a hierarchical Bayes meta-analysis of available trait data (1). The output of this analysis is a probability distribution for each model parameter. For default single runs PEcAn will select the median value of this parameter as the default, but in Demo 2 we will see how PEcAn can use this parameter uncertainty to make probabilistic forecasts and assess model sensitivity and uncertainty. Errors at this stage usually indicate errors in the trait database or incorrectly specified PFTs (e.g. defining a variable twice).
2. **Conversions:** PEcAn next downloads and processes and model input data that was selected but not available locally.
3. **write.config**: writes model-specific settings and parameter files, as well as the script (jobs.sh) that PEcAn uses to initiate a model run.
4. **model:** run model.
5. **output.conversion**: All model outputs are converted from their model-specific format to a general format (netCDF files, MsTMIP variables & units)
6. **ensemble.analysis & sensitivity analysis**: If these analyses are enabled in Advanced Setup or in the pecan.xml, PEcAn will post-process the output across numerous model runs in order to generate a model projection mean & confidence interval, and a sensitivity & uncertainty analysis, respectively.

   FYI, **adding a new model to PEcAn** does not require modification of the model's code, just the implementation of two modules, write.configs and output.conversion. See the PEcAn wiki (https://github.com/PecanProject/pecan/wiki/Adding-an-Ecosystem-Model) for details.

   If, by chance, you get an error in your run today it is probably simplest to go back and try changing options and re-running (e.g. with a different site or PFT), as time does

not permit detailed debugging.

Congratulations! If you got this far, you have managed to run an ecosystem model without ever touching a line of code! Now it's time to look at the results. When the workflow is done click Finished.

5. **Output and Visualization**

   1. **Graphs**
      1. Select a **Year** and **Y-axis Variable**, and then **click 'Plot run/year/variable'**. Initially leave the X-axis as time.
      2. Within this figure the **points indicate the daily mean** for the variable while the **envelope encompasses the diurnal variability** (max and min).
      3. Units, variable names, and descriptions are all drawn automatically from the meta-data in a standardized, model-independent netCDF output format (derived from NACP, LBA, PalEON, and MsTMIP model-data inter-comparison projects). See http://nacp.ornl.gov/MsTMIP_variables.shtml for variable names and units.
      4. Figures are *.png files that can be saved by right-clicking and selecting "Save Image As...". However, as will be described below it is easy to go back to past runs in PEcAn so you don't have to download figures or files in order to keep a record of your results.
      5. Try looking at a number of different output variables over different years.
      6. Try changing the X-axis to look at bivariate plots of how different output variables are related to one another. Be aware that PEcAn currently runs a moving min/mean/max through bivariate plots, just as it does with time series plots. In some cases this makes more sense than others.

   2. **Inputs**
      This menu lets you look at and download:
      1. A summary file (README.txt) describing the run (location, run ID, model, dates, etc)
      2. The model-specific input files fed into the model
      3. The jobs.sh file used to submit the model run

   3. **Outputs:**
      A number of files generated by the underlying ecosystem model are archived and available for download. are archived and available for download. These include:
      1. Output files in the standardized netCDF ([year].nc)
      2. Raw model output in model-specific format (e.g. sipnet.out).
      3. Logfile.txt contains error, warning, and informational messages generated by PEcAn while running the workflow.

      For example, one could easily download model output and plot model versus data in Matlab or R, or visualize netCDF output in ncview or panoply.

   4. **PFTs:**
      There is a wide array of outputs available that are related to the process of estimating the model parameters for a specific Plant Functional Type. We will revisit many of these files in Demo 2 when we look at the meta-analysis in more detail, but the following is the general idea of what is contained in different files:
      1. *.bug files provide the statistical code used to estimate each parameter in JAGS.
      2. The ma.summaryplots.*.pdf are collections of diagnostic plots produced in R after the

above JAGS code is run that are useful in assessing the statistical model.

3. The Rdata files trait.data.Rdata and madata.Rdata are, respectively, the available trait data extracted from the database that was used to estimate the model parameters and that same data cleaned and formatted for the statistical code.
4. The file traits.mcmc.Rdata contains the raw output from the statistical code. This includes samples from all of the parameters in the meta-analysis model, not just those that feed forward to the ecosystem, but also the variances, fixed effects, and random effects.
5. The files prior.distns.Rdata and post.distns.Rdata store simple tables of the prior and posterior distributions for all model parameters in terms of the name of the distribution and its parameters.
6. Finally, posteriors.pdf provides graphics showing, for each model parameter, the prior distribution, the data, the smoothed histogram of the posterior distribution (labeled *post*), and the best-fit analytical approximation to that smoothed histogram (labeled *approx*)

5. **PEcAn Files:**
   Likewise, a number log files from the PEcAn workflow are available from the execution of the workflow. These include:
   1. The STATUS file giving a summary of the steps of the workflow, the time they took, and whether they were successful
   2. The PEcAn settings file (pecan.xml),
   3. The workflow script (workflow.R),
   4. The corresponding log file (workflow.Rout),
   5. The Monte Carlo samples used in the runs (samples.Rdata),
   6. Various ensemble analysis files, which are not terribly interesting in the current context since by default the web interface only does one run (ensemble size, n=1). We'll come back to this in Demo 2

   These files are primarily useful if the workflow generates an error, and as metadata & provenance (a detailed record of how data was generated), but also includes useful diagnostics from the meta-analysis.

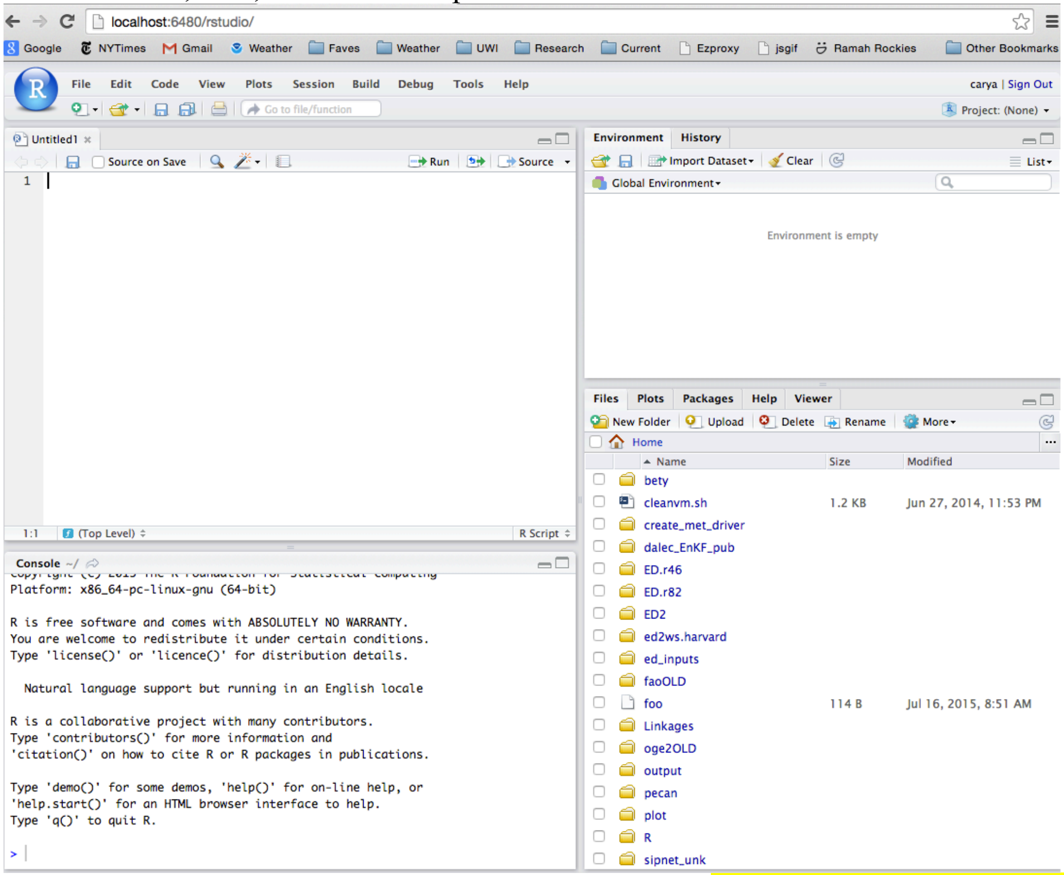
6. **Model Run Archive**

   Click on the HISTORY button at the start or end of the workflow to go to a table recording past model runs. Click on any previous run in the "ID" column to go to the current state of that run's execution. This is most commonly the output and visualization page, but if you are doing a long run you can always return to that run in-progress this way. The run you just did should be the more recent entry in the table. For the next analysis, make note of the ID number from your run.

7. **Analyzing model output**

   In this final section we will compare model outputs to data outside of the PEcAn web interface. The goal of this is to demonstrate that it is straightforward to perform additional analyses using PEcAn's outputs. To do this one can download each of the output files using the Output and Visualization menu described above, and then perform the analyses using whatever software you prefer. You can also perform analyses directly on the PEcAn server itself, which is the approach we will take in this demo so that you can also learn how PEcAn organizes outputs. Specifically we'll be analyzing model outputs in R using a browser-based version of RStudio that's installed on the server

1. Open RStudio Server, which provides a development environment with access to all of PEcAn's code, data, and model output from the web runs



```
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

>
```

Name	Size	Modified
bety		
cleanvm.sh	1.2 KB	Jun 27, 2014, 11:53 PM
create_met_driver		
dalec_EnKF_pub		
ED.r46		
ED.r82		
ED2		
ed2ws.harvard		
ed_inputs		
faoOLD		
foo	114 B	Jul 16, 2015, 8:51 AM
Linkages		
oge2OLD		
output		
pecan		
plot		
R		
sipnet_unk		

1. Open up a second window on your web browser to http://localhost:3280/rstudio
   If you closed your previous PEcAn session, make sure the first window is open and points to http://localhost:3280/pecan
   Use port 6480 for 64-bit, or the AWS IP address without a port to open rstudio on the server.

2. The username for the VM is *carya* and the password is *illinois*.

3. To open a new R script click File > New File > R Script

4. Use the Files browser on the lower right pane to find where your run(s) are located
   1. All PEcAn outputs are stored in the **output** folder. Click on this to open it up.
   2. Within the outputs folder, there will be one folder for each workflow execution. For example, if your workflow ID from the History table was 99000000001 then you will want to click to open the folder PEcAn_99000000001
   3. A workflow folder will have a few log and settings files (e.g. pecan.xml) and the following subfolders (which map directly to the subsections in the PEcAn Output menu)

run	contains all the inputs for each run
out	contains all the outputs from each run
pft	contains the parameter information for each PFT

        Within both the run and out folders there will be one folder for each unique model run, where the folder name is the run ID. Click to open the *out* folder. For our simple case we only did one run so there should be only one folder (e.g. 99000000001). Click to open this folder.

    4. Within this folder you will find, among other things, files of the format <year>.nc. Each of these files contains one year of model output in the standard PEcAn netCDF format.

2. Open and plot the raw model outputs
    1. We will use the PEcAn.utils library to read model output so add the following at the top of your script

       ```
       library(PEcAn.utils)
       ```

    2. Load the modeled NEE using the *read.output* function. The arguments to this function are the run ID, the folder where the run is located, the start year, the end year, and the variable being read.

       ```
       arun <- read.output(99000000001, "output/PEcAn_99000000001/out/99000000001/"
       ,2006,2006,c("time","NEE"))
       ```

    3. Plot the raw model output

       ```
       plot(arun$time,arun$NEE,type='l',ylab="kg ha-1 yr-1",xlab="day of year")
       ```

       Go ahead and highlight the three lines you added to the script and click on the “Run” button on the top of the script pane. You should see a plot of day of year vs. NEE

    4. Compare model to flux observations

       First we want to load up the observations and take a look at the contents of the file

       ```
       library(PEcAn.assim.batch)
       obs <- load.L2Ameriflux.cf("~/output/dbfiles/Ameriflux_site_0-772/US-NR1.2006.nc")
       names(obs)
       ```

       Here 0-722 is the database site ID, a folder that was created when you selected the site and PEcAn downloaded the Ameriflux NetCDF (.nc) file from the Ameriflux website. The names function lists all the variables within the list “obs”. Add these lines to your script, highlight them, and click run.

       Second, we need to change all bad values from -9999 to not a number (NA) and also u* filter the fluxes. Here we will apply a conservative u* of 0.2. Do the same as above for running.

       ```
       obs[obs==-9999]=NA
       obs$NEE[obs$UST<0.2]=NA
       ```

       Third, we need to convert the model to have the same units as the data, which we do with the udunits package. The observations are in umol m-2 s-1. The model output is read in kg C ha-1 yr-1. udunits can convert among many units but cannot do the conversion from grams carbon to moles carbon, so we do that separately, with the factor of 12 g C per mol.

```
modNEE <- ud.convert(arun$NEE,"kg ha-1 yr-1","ug m-2 s-1")/12.0
```

Fourth, we can plot model predictions vs. observations and compare this to a 1:1 line (which we add using the abline function, where the default arguments are intercept and slope)

```
plot(modNEE,obs$NEE)
abline(0,1,col="red")
```

Fifth, we can calculate the Root Mean Square Error (RMSE) between the model and the data

```
rmse = sqrt(mean((modNEE-obs$NEE)^2,na.rm = TRUE))
rmse
```

Finally, we can plot time-series of both the model and data together

```
plot(arun$time,modNEE,type='l',ylab="umol m-2 s-1",xlab="day of year")
lines(arun$time,obs$NEE,col="green")
```

DEMO 2: Ensemble & Sensitivity Analyses

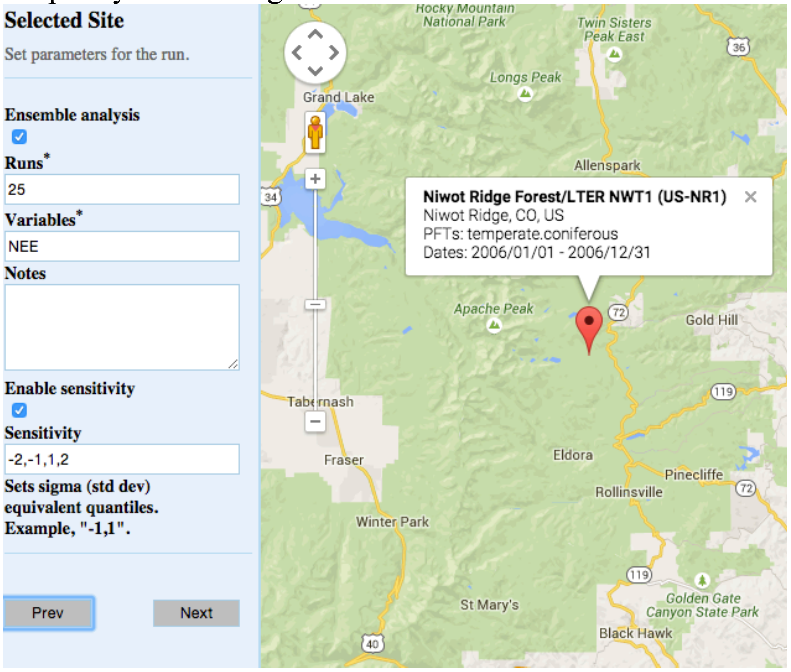
In Demo 2 we will be looking at how PEcAn can use information about parameter uncertainty to perform three automated analyses:

- **Ensemble Analysis:** Repeat numerous model runs, each sampling from the parameter uncertainty, in order to generate a probability distribution of model projections. Allows us to put a confidence interval on the model
- **Sensitivity Analysis:** Repeats numerous model runs to assess how changes in model parameters will affect model outputs. Allows us to identify which parameters the model is most sensitive to.
- **Uncertainty Analysis:** Combines information about model sensitivity with information about parameter uncertainty in order to determine the contribution of each model parameter to the uncertainty in model outputs seen in the Ensemble Analysis. Allow us to identify which parameters are driving model uncertainty.

1. Return to the main menu for the PEcAn web interface (e.g. click Start Over or type in the main pecan server address again (e.g., http://localhost:3280/pecan))

2. Repeat the steps for site selection and run specification, however this time **click to turn on "Edit pecan.xml" and also "Advanced setup"**

3. By clicking Advanced setup, PEcAn will first show an Analysis Menu, where we are going to specify new settings.



Selected Site

Set parameters for the run.

Ensemble analysis
☑

Runs*
25

Variables*
NEE

Notes

Enable sensitivity
☑

Sensitivity
-2,-1,1,2

Sets sigma (std dev) equivalent quantiles. Example, "-1,1".

Niwot Ridge Forest/LTER NWT1 (US-NR1)
Niwot Ridge, CO, US
PFTs: temperate.coniferous
Dates: 2006/01/01 - 2006/12/31

   1. In order to run an ensemble analysis, we need to tell how many runs we want in our ensemble. We do this by changing the value set in the Ensemble tag from 1 run to 25 runs. As a reminder, the ensemble analysis will sample all model parameters randomly from their posterior distributions in order to propagate this uncertainty into the model output. *In practice you would want to use a larger ensemble size (100-5000) than we are*

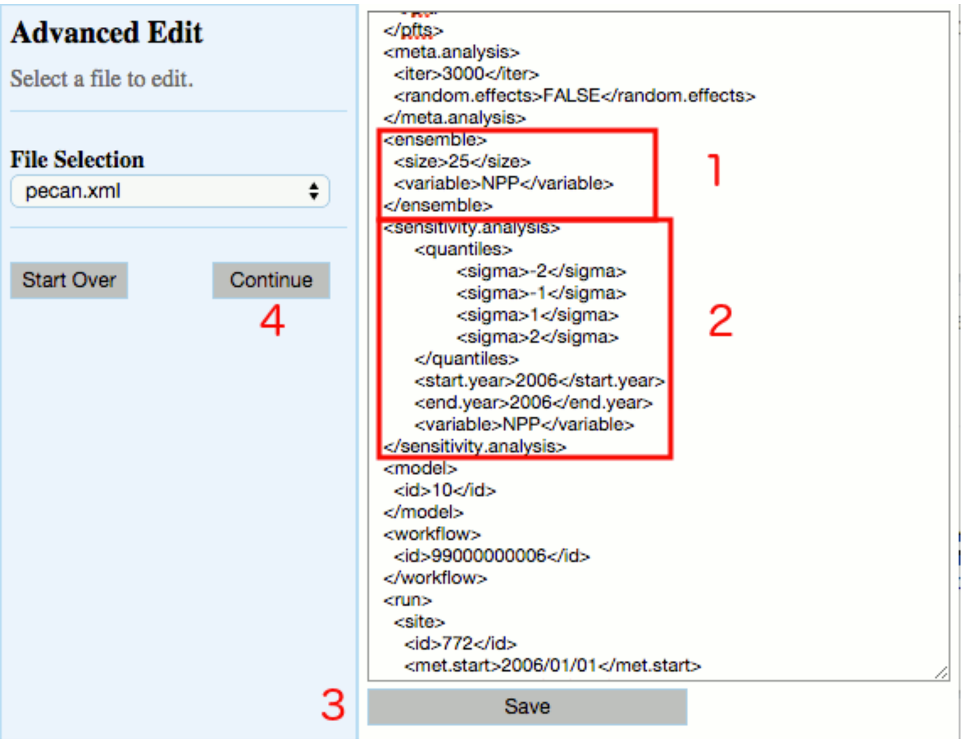
using in the demo.

2. In order to run the sensitivity analysis, we likewise need tell PEcAn how we want the parameters varied. PEcAn's sensitivity analysis holds all parameters at their median value and then varies each parameter one-at-a-time based on the *quantiles* of the posterior distribution. PEcAn also includes a handy shortcut, which is the default behavior for the web interface, that converts a specified standard deviation into its Normal quantile equivalent. For example, 1 and -1 are converted to 0.157 and 0.841 (a 68.2% CI). In this example we will look at samples from -2 to 2 sigma, which we specify in a comma-separated list as -2,-1,1,2 (the median value, 0, occurs by default).

   We also can tell PEcAn which variable to run the sensitivity on. By default PEcAn will calculate sensitivity for NPP and over the full length of the run. Here, replace NPP with NEE, so we can compare against flux tower NEE observations.

   Click Next

3. Advanced Edit (optional)
   If you want to change the output variable analyzed, the start year, or the end year you can do this by editing the pecan.xml file. To specify time you use the <start.year> and <end.year> tags. The <variable> tag is a character string that follows the output standards used by recent inter-comparison projects (MsTMIP, NACP, LBA, PalEON), see http://nacp.ornl.gov/MsTMIP_variables.shtml for variable names and units.



Advanced Edit

Select a file to edit.

File Selection

pecan.xml

```
</pfts>
<meta.analysis>
  <iter>3000</iter>
  <random.effects>FALSE</random.effects>
</meta.analysis>
<ensemble>
  <size>25</size>
  <variable>NPP</variable>
</ensemble>
<sensitivity.analysis>
    <quantiles>
            <sigma>-2</sigma>
            <sigma>-1</sigma>
            <sigma>1</sigma>
            <sigma>2</sigma>
    </quantiles>
    <start.year>2006</start.year>
    <end.year>2006</end.year>
    <variable>NPP</variable>
</sensitivity.analysis>
<model>
  <id>10</id>
</model>
<workflow>
  <id>99000000006</id>
</workflow>
<run>
  <site>
    <id>772</id>
    <met.start>2006/01/01</met.start>
```

The latest version of the pecan.xml documentation, and further documentation of the PEcAn system in general, can be found in the wiki section of our GitHub website (https://github.com/PecanProject/pecan/wiki/PEcAn-Configuration). Examples of other advanced settings include turning on the data assimilation modules and connecting

PEcAn to a cluster (https://github.com/PecanProject/pecan/wiki/Connecting-pecan-to-a-remote-server).

"Save" you changes then click "Continue"

Note this time, the met process will run faster since PEcAn's database knows that the Niwot Ridge Ameriflux file has already been downloaded. However, the model run will take much longer, since both ensemble and sensitivity analyses call the model multiple time (in our case, 100s).

4. Additional Outputs:
   The PEcAn workflow will take considerably longer to complete since we have just asked for over a hundred model runs. Once the runs are complete you will return to the output visualization page were there will be a few new outputs to explore, as well as a few outputs that were present earlier that we'll explore in greater details:

   1. **Run ID**:
      While a number of the PEcAn outputs synthesize across each of the runs, you can also select individual runs from the Run ID menu. For an individual run you can then use the Graphs menu to visualize each individual run. The README.txt file can be used to get a quick view of whether each run was part of the sensitivity or ensemble analysis, as well as what variables are being modified in the sensitivity analysis.

   2. **PFTs & Trait Data**: The list of variables that are queried is determined by what variables have priors associated with them in the definition of the PFTs. Likewise, the list of species that are associated with a PFT determines what subset of data is extracted out of all data matching a given variable name. Demo 3 will demonstrate how a PFT can be created or modified. The trait output is written to two Rdata files, **trait.data.Rdata** and **prior.distns.Rdata**, To look at these in R you could use the menu to download the file and then within R click on these files to load them into your workspace. You can further examine them in the **Environment** window or accessing them at the command line. For example, try typing `names(trait.data)` as this will tell you what variables were extracted, `names(trait.data$Amax)` will tell you the names of the columns in the Amax table, and `summary(trait.data$Amax)` will give you summary data about the Amax values.

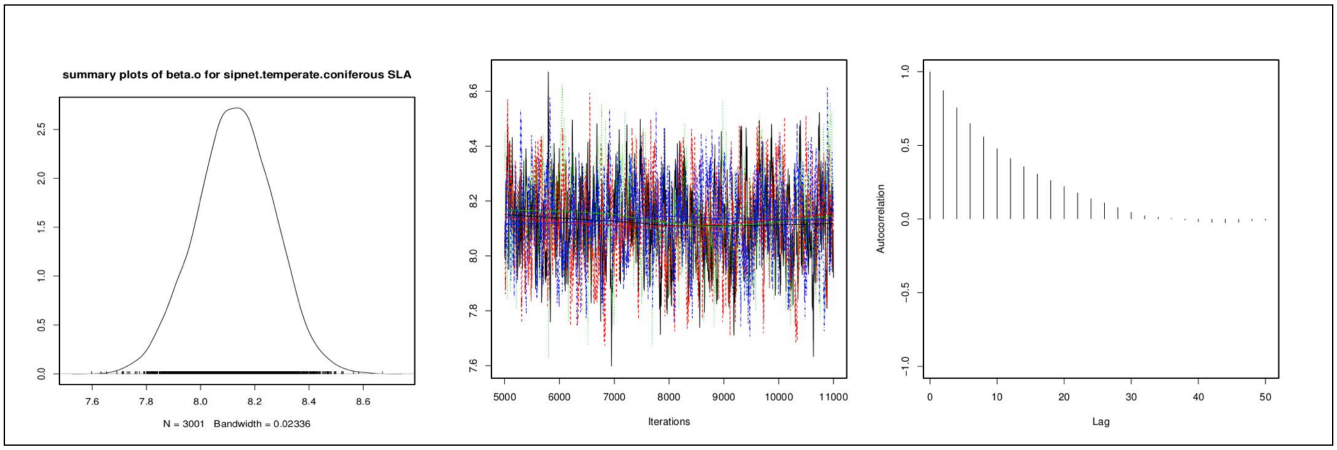
   3. **Meta-analysis**

      1. The meta-analysis code begins by distilling the trait.data to just the values needed for the meta-analysis statistical model, with this being stored in `madata.Rdata`. This reduced form includes the conversion of all error statistics into precision (1/variance), and the indexing of sites, treatments, and greenhouse. In reality, the core meta-analysis code can be run independent of the trait database as long as input data is correctly formatted into the form shown in madata.

      2. The evaluation of the meta-analysis is done using a Bayesian statistical software package called JAGS that is called by the R code. For each trait, the R code will generate a `[trait].model.bug` file that is the JAGS code for the meta-analysis

    itself. This code is generated on the fly, with PEcAn adding or subtracting the site, treatment, and greenhouse terms depending upon the presence of these effects in the data itself. If the <random.effects> tag is set to FALSE then all random effects will be turned off even if there are multiple sites and treatments.

    3. The MCMC outputs a number of diagnostics, including the summary statistics of the model, an assessment of whether the posterior is consistent with the prior, and the status of the Gelman-Brooks-Rubin convergence statistic (which is ideally 1.0 but should be less than 1.1). All of this output can be found in the meta-analysis.log file.
    4. The code generates a number of diagnostic figures in a ma.summaryplots.[pft].[trait].pdf plot for each pft and trait. Open up one of these pdfs to evaluate the shape of the posterior distributions (they should generally be unimodal), the convergence of the MCMC chains (all chains should be mixing well from the same distribution), and the autocorrelation of the samples (should be low).
    5. Finally, the code attempts to fit an analytical approximation to the MCMC posterior. The outputs of this are post.distns.Rdata, which is structured identically to the prior.distns but updated for the meta-analysis, and posteriors.pdf, which is a series of graphs comparing the MCMC posterior, the analytical approximation, the data, and the priors. Open posteriors.pdf and compare the posteriors to the priors and data.



4. **Model Parameters** for the various runs are stored in samples.Rdata. This file contains two data objects, sa.samples and ensemble.samples, that are the parameter values for the sensitivity analysis and ensemble runs respectively.

5. **Ensemble Analysis**:
    1. Raw results are stored in ensemble.Rdata, found in the PEcAn Files list. It contains the object ensemble.output, which is the model predictions at the parameter values given in ensemble.samples.
    2. Open ensemble.analysis.[RunID].[Variable].[StarYear].[EndYear].pdf contains the ensemble prediction as both a histogram and a boxplot.

3. Open ensemble.ts.[RunID].[Variable].[StartYear].[EndYear].pdf, which contains a time-series plot of the ensemble mean, median, and 95% CI

6. **Sensitivity Analysis**:
   1. Raw results are stored in sensitivity.output.[RunID].[Variable].[StartYear].[EndYear].Rdata, located in the PEcAn Files list. It contains the object sensitivity.output which is the model outputs corresponding to the parameter values in sa.samples.
   2. Sensitivity plots are in **sensitivity.analysis.[RunID].[Variable].[StartYear].[EndYear].pdf**, found in the Output list under PFT. It shows the raw data points from univariate one-at-a-time analyses and spline fits through the points

7. **Uncertainty Analysis**:
   The variance decomposition plot, **variance.decomposition.[RunID].[Variable].[StartYear].[EndYear].pdf**, is found in the Output list under PFT. It contains three columns, the coefficient of variation (normalized posterior variance), the elasticity (normalized sensitivity), and the partial standard deviation of each model parameter.

QUESTIONS:

The Variance Decomposition graph is sorted by the variable explaining the largest amount of variability in the model output (right hand column). From this graph identify the top-tier parameters that you would target for future constraint.

A parameter can be important because it is highly sensitive, because it is highly uncertain, or both. Identify parameters in your output that meet each of these criteria. Additionally, identify parameters that are highly uncertain but unimportant (due to low sensitivity) and those that are highly sensitive but unimportant (due to low uncertainty).

Parameter constraints could come from further literature synthesis, from direct measurement of the trait, or from data assimilation. Choose the parameter that you think provides the most efficient means of reducing model uncertainty and propose how you might best reduce uncertainty in this process. In making this choice remember that not all processes in models can be directly observed, and that the cost-per-sample for different measurements can vary tremendously (and thus the parameter you measure next is not always the one contributing the most to model variability). Also consider the role of parameter uncertainty versus model sensitivity in justifying your choice of what parameters to constrain.

The next set of tutorials will focus on the process of data assimilation and parameter estimation. The next two steps are in ".Rmd" files which can be viewed online at:
https://github.com/PecanProject/pecan/tree/master/documentation/tutorials

Additional material

In addition to the materials below you are encouraged to take a look at the **Site-Level Runs** section on the PEcAN wiki (https://github.com/PecanProject/pecan/wiki/Getting-started#site-level-runs) which describes the process of setting up a new site, new meteorology, new PFT, etc.

Additional useful information on the wiki includes: how to report bugs, how to sync your PEcAn database with others, how to grab the latest "development" branch of the PEcAn code, how to install PEcAn on a server, how to add new models and data processing modules, and how to add new code to PEcAn.

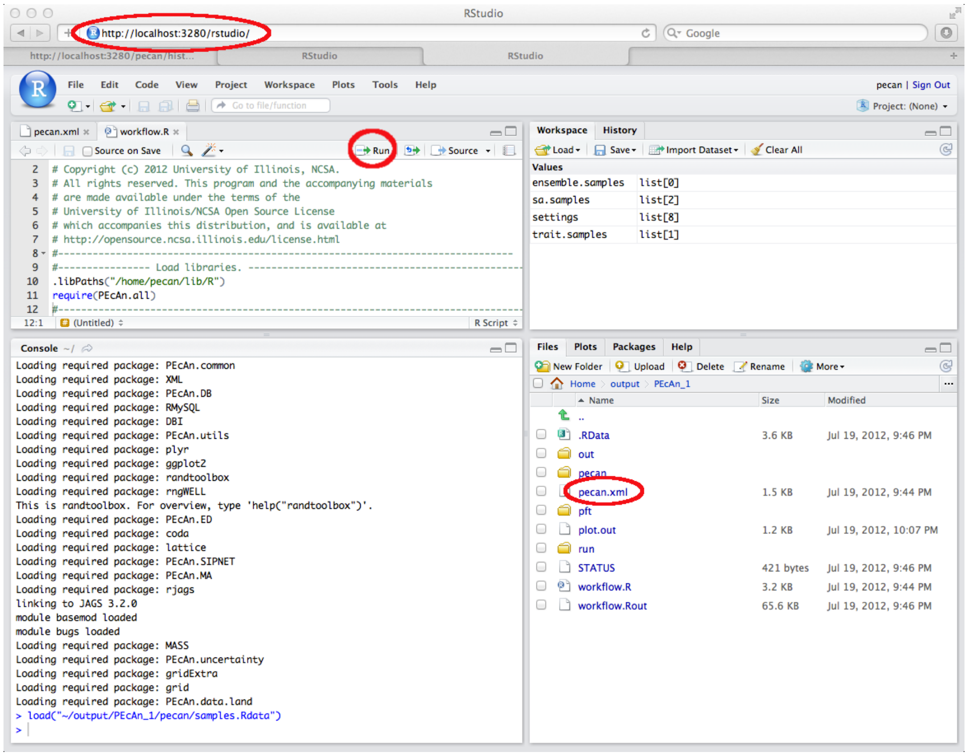
Currently, the web interface does not have access to all of PEcAn's functionality. For additional tasks we will need to work in the statistical programming language R. To make this easier we have installed a brower-based R interface, RStudio Server (http://rstudio.org), on the VM.

1. Open RStudio Server:

    1. Open up a second window on your web browser to http://localhost:3280/rstudio
       If you closed your previous PEcAn session, make sure the first window is open and points to http://localhost:3280/pecan

    2. The username for the VM is *carya* and the password is *illinois*.

    3. You do not need to do this now, but FYI it is also possible to ssh into the virtual machine using the following command in terminal:
       ssh -l pecan -p 3222 localhost



```
# Copyright (c) 2012 University of Illinois, NCSA.
# All rights reserved. This program and the accompanying materials
# are made available under the terms of the
# University of Illinois/NCSA Open Source License
# which accompanies this distribution, and is available at
# http://opensource.ncsa.illinois.edu/license.html
#-------------------------------------------------------------------------------
#-------------- Load libraries. -------------------------------------------------
.libPaths("/home/pecan/lib/R")
require(PEcAn.all)
#--------------------------------------------------------------------------------
```

```
Loading required package: PEcAn.common
Loading required package: XML
Loading required package: PEcAn.DB
Loading required package: RMySQL
Loading required package: DBI
Loading required package: PEcAn.utils
Loading required package: plyr
Loading required package: ggplot2
Loading required package: randtoolbox
Loading required package: rngWELL
This is randtoolbox. For overview, type 'help("randtoolbox")'.
Loading required package: PEcAn.ED
Loading required package: coda
Loading required package: lattice
Loading required package: PEcAn.SIPNET
Loading required package: PEcAn.MA
Loading required package: rjags
linking to JAGS 3.2.0
module basemod loaded
module bugs loaded
Loading required package: MASS
Loading required package: PEcAn.uncertainty
Loading required package: gridExtra
Loading required package: grid
Loading required package: PEcAn.data.land
> load("~/output/PEcAn_1/pecan/samples.Rdata")
> 
```

Values	
ensemble.samples	list[0]
sa.samples	list[2]
settings	list[8]
trait.samples	list[1]

Name	Size	Modified
..		
.RData	3.6 KB	Jul 19, 2012, 9:46 PM
out		
pecan		
pecan.xml	1.5 KB	Jul 19, 2012, 9:44 PM
pft		
plot.out	1.2 KB	Jul 19, 2012, 10:07 PM
run		
STATUS	421 bytes	Jul 19, 2012, 9:46 PM
workflow.R	3.2 KB	Jul 19, 2012, 9:44 PM
workflow.Rout	65.6 KB	Jul 19, 2012, 9:46 PM

2. Open PEcAn settings file

    1. Go back to http://localhost:3280/pecan and note the Run ID for your model run. In the visualization window this is listed in the top left pull-down menu

3. PEcAn Workflow

    1. In the **Files** window, go to the **Home > pecan > web** directory

    2. Open the file **workflow.R**. This script provides an example of some of the high-level modules within PEcAn and we will use it to run the meta-analysis, sensitivity analysis, ensemble analysis, and variance decomposition.

1. Re-run the sensitivity analysis and variance decomposition for a different output variable, time period, site, or PFT and compare results with first site.
    1. If you are only changing the output variable or time period, you don't need to re-run the whole workflow since all the output variables are in the model output files. You can modify the pecan.xml settings object within R and run the workflow from the extraction of the output from the model runs onward (from get.model.output on)
    2. If you are running a new site, it is generally a good idea to do a single test run first to make sure the model behaves sensibly before conducting hundreds of (potentially uninformative) runs.
2. Use the web database interface to change priors, define a new site, or define a new PFT
    1. Open up a new browser window to http://localhost:3280/bety . This page is the web-interface to the database underlying PEcAn – it started as an independent project as part of our biofuel work, hence the name BETY-db, but is now much more generic (apologies for the fact that web design has not kept up). For those that are interested the full version of the database is at betydb.org.
    2. To modify or add a new Plant Functional Type (PFT), or to change a PFT's priors, navigate on the grey menu bar to Data > PFTs
        1. To add a new pft, click "new PFT" at the top and enter a name and description.
        2. To add new species to a PFT click on [+] View Related Species and type the species, genus, or family you are looking for into the Search box. Click on the + to add.
        3. To remove a species from a PFT, click on [+] View Related Species and click on the X of the species you want to remove from the PFT.
        4. To remove a prior, click [-] View Related Prior and click on the X of the variable who's prior you want to remove. This will cause the parameter to be excluded from all analyses (meta-analysis, sensitivity analysis, etc) and revert to its default value.
        5. To add a prior, choose one from the white box of priors on the right to choose.
        6. To view the specification of a prior, or to add a new prior, click BETY-DB > Priors and enter the information on the variable, distribution name, distribution parameters, etc. N is the sample size underlying the prior specification (0 is ok for uninformative priors).
        7. You can also got to Data > Variables in order to use the search function to find an existing variable (or create a new one). Please try not to create new variables unnecessarily (e.g. changes of variable name or units to what your model uses is handled internally, so you want to find the trait with the correct MEANING).
    3. To add a new site
        1. The database currently requires that a site be associated with a citation. Navigate to > CITATIONS and select a reference for the study site by clicking the check mark, or

create a new citation (creating a 'pers. communication' or 'unpublished data' citation IS allowed)

2. Click on the **New Site** button. Fill out at least a name and lat/lon (which can be done by hand or by clicking on the map).
3. To upload model driver data [or any other input data] for the site (currently required in order to run the model at a site)
    1. Select RUNS > INPUTS
    2. At the bottom of the page click on New Input
    3. Choose your site from the pulldown menu
    4. Set the date range, file format, and other meta-data
    5. Click “Create”
    6. To associate the Input record in the database, click “Edit” and then “New File”
    7. Select the machine that the file is on (pecan32 is the VM itself) and fill in the directory path where the file is located and the file name
    8. Click Update

Citation:

LeBauer D, Wang D, Richter KT, Davidson CC, Dietze MC (2013) Facilitating feedbacks between field measurements and ecosystem models. *Ecol Monogr* 83:133–154.