PEcAn v1.4.8 Hands-On Demo

In the following demos, URL is the web address of a PEcAn server and will refer to one of the following:
- **If you are doing a live demo with the PEcAn team, URL was provided**
- **If you downloaded the PEcAn VM, URL = localhost:6480**
- **If you are running PEcAn on Amazon AWS, URL is the Public IP**
- **If you installed PEcAn on your server, URL is your server's IP**

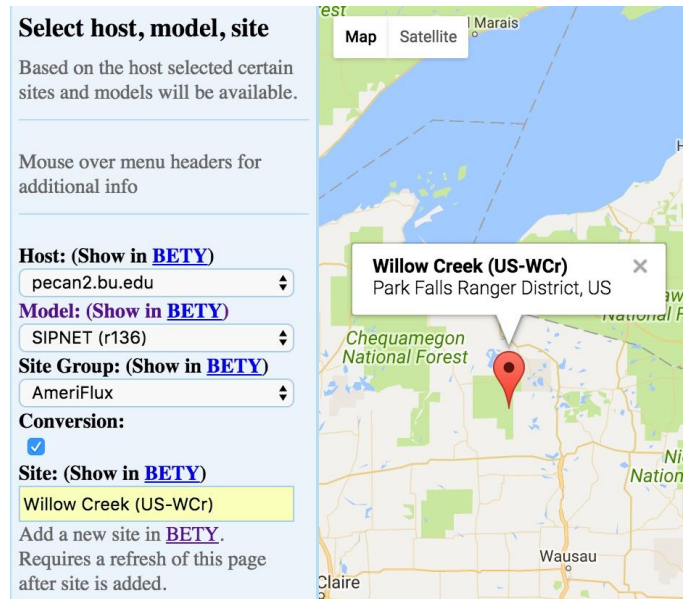## Demo 1: Single-site model run

Objective:

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:**
   1. **Enter URL in your web browser**
   2. **Click "Run Models"**
   3. **Click the 'Next' button** to move to the "Site Selection" page.

2. **Site Selection**
   1. **Host:** Select the local machine "pecan64" Other options exist if you've configured PEcAn to run models on remote servers.
   2. **Model:** Select SIPNET from the available models because it is quick & simple.
   3. **Site Group:** To filter sites, you can select a specific group of sites. For this tutorial we will use Ameriflux.
   4. **Conversion:** Select the conversion check box, to show all sites that PEcAn is capable of generating model drivers for automatically. By default (unchecked), PEcAn only displays sites where model drivers already exist in the system database
   5. **Site:** For this tutorial, type US-WCr in the search box to display the Willow Creek Ameriflux site (US-WCr), and then click on the pin icon. When you click on a site's flag on the map, it will give you the name and location of the site and put that site in the "Site:" box on the left hand site, indicating your current selection.
   6. Click "Next"

Select host, model, site

Based on the host selected certain sites and models will be available.

Mouse over menu headers for additional info

Host: (Show in BETY)
pecan2.bu.edu

Model: (Show in BETY)
SIPNET (r136)

Site Group: (Show in BETY)
AmeriFlux

Conversion:
☑

Site: (Show in BETY)
Willow Creek (US-WCr)

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

Willow Creek (US-WCr)
Park Falls Ranger District, US

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**)

luquillo.late_tropical
populus
soil
temperate.coniferous
temperate.deciduous

Start Date*

2003/01/01

End Date*

2006/12/31

Sipnet.climna*

Use Ameriflux

Willow Creek (US-WCr) ✕
Park Falls Ranger District, Wisconsin, US
Mean Annual Temp : 4.00
Mean Annual Precip : 815
Greenhouse Study :
Local Time : America/Chicago
Sand Pct : 42.52000
Clay Pct : 20.17000
Soil :
Notes : MF
Soil Notes :

1. **PFT (Plant Functional Type):** Willow Creek is temperate deciduous. Available PFTs will vary by model and some models allow multiple competing PFTs to be selected. Also select soil to control the soil parameters
2. **Start/End Date**: Select 2003/01/01 to 2006/12/31. In general, be careful to select dates for which there is available driver data.
3. **Weather Data:** Select "Use Ameriflux" from the available meteorological drivers
4. **Optional Settings:** leave all blank for demo run
   1. **Email** sends a message when the run is complete.
   2. **Use Brown Dog** will use the Brown Dog web services in order to do input file conversions. (*Note: Required if you select Use NARR for Weather Data*)
   3. **Edit pecan.xml** allows you to configure advanced settings via the PEcAn settings file
   4. **Edit model config** pauses the workflow after PEcAn has written all model specific settings but before the model runs are called and allows users to configure any additional settings internal to the model.
   5. **Advanced Setup** controls ensemble and sensitivity run settings discussed in Demo 2.
5. Click **Next** to start the model run.
6. 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.

If you get an error in your run today it is probably simplest to start over and try changing options and re-running (e.g. with a different site or PFT), as time does not permit detailed debugging.

4. Model Run Workflow

Job Finished

Job has finished successfully.
Press Finished to see the results.

| Start Over | Finished |

Not Logged in.login

Execution Status

Stage Name	Start Time	End Time	Status
MET Process	2016-05-05 11:07:12	2016-05-05 11:07:43	DONE
TRAIT	2016-05-05 11:07:43	2016-05-05 11:07:44	DONE
META	2016-05-05 11:07:44	2016-05-05 11:07:55	DONE
CONFIG	2016-05-05 11:07:55	2016-05-05 11:07:56	DONE
MODEL	2016-05-05 11:07:56	2016-05-05 11:07:57	DONE
OUTPUT	2016-05-05 11:07:57	2016-05-05 11:07:57	DONE
ENSEMBLE	2016-05-05 11:07:57	2016-05-05 11:08:00	DONE
SENSITIVITY	2016-05-05 11:08:00	2016-05-05 11:08:00	DONE
FINISHED	2016-05-05 11:08:00	2016-05-05 11:08:00	DONE

Workflow Log

Last 10 ▾ lines of the workflow.Rout

```
Created 10 connections and executed 228 queries
2016-05-05 11:08:00 DEBUG [db.print.connections] :
No open database connections.
> print("---------- PEcAn Workflow Complete ----------")
[1] "---------- PEcAn Workflow Complete ----------"
>
> proc.time()
  user system elapsed
45.919  1.544 52.060
```

1. **MET Process:** First, PEcAn will download meteorological data based on the type of the Weather Data you chose, and process it into the specific format for the chosen model
2. **TRAIT / META**: PEcAn then estimates model parameters by performing a meta-analysis of the available trait data for a PFT. TRAIT will extract relevant trait data from the database. META performs a hierarchical Bayes meta-analysis of available trait data (1). The output of this analysis is a probability distribution for each model parameter. PEcAn selects 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).
3. **CONFIG**: writes model-specific settings and parameter files
4. **MODEL:** runs model.
5. **OUTPUT:** All model outputs are converted to standard netCDF format
6. **ENSEMBLE & SENSITIVITY**: If enabled post-process output for these analyses

If at any point a **Stage Name** has the **Status "ERROR"** please notify the PEcAn team member that is administering the demo or feel free to do any of the following: refer to the PEcAn Gitbook for documentation , post the end of your workflow log on our Gitter chat, or post an issue on Github. The entire PEcAn team welcomes any questions you may have!

If the Finished Stage has a Status of "DONE", 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.

FYI, **adding a new model** to PEcAn does not require modification of the model's code, just the implementation of a wrapper function.

5. **Output and Visualization**

   For now focus on graphs, we will explore all of PEcAn's outputs in more detail in Demo 2.

   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. Variable names and units are based on a standard netCDF format.
      4. Try looking at a number of different output variables over different years.
      5. 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. **Alternative Visualization: Shiny**
      1. Click on Open SHINY, which will open a new browser window.
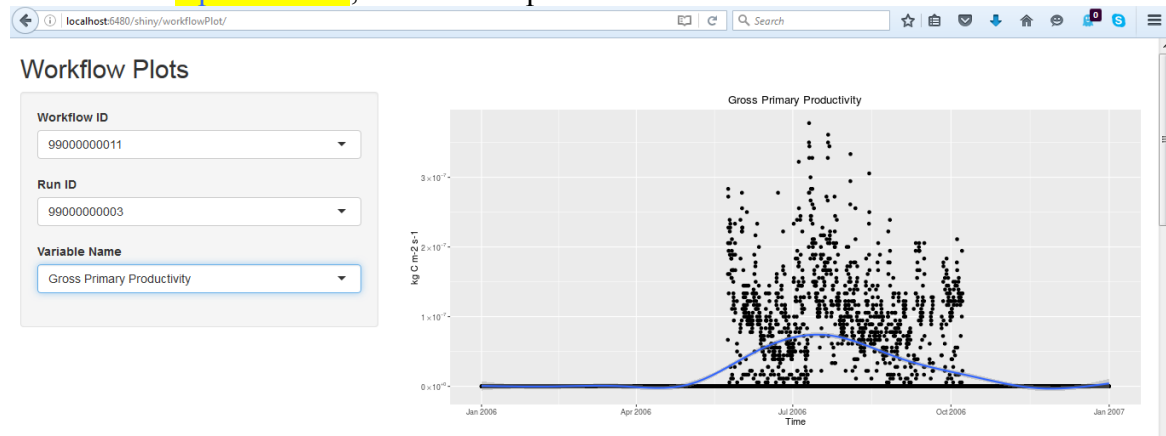
   The shiny app will automatically access your run's output files and allow you to visualize all output variables as a function of time. Use the pull down menu under **Variable Name** to choose whichever output variable you wish to plot.

6. **Model Run Archive**
   Return to the output window and Click on the HISTORY button. Click on any previous run in the "ID" column to go to the current state of that run's execution -- you can always return to old runs and runs 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**
   Follow this tutorial to learn how to open model output in R and compare to observed data

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, 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 to determine the contribution of each model parameter to the uncertainty in model outputs. Allow us to identify which parameters are driving model uncertainty.

1. Return to the main menu for the PEcAn web interface: URL > Run Models

2. Repeat the steps for site selection and run specification, **click on "Advanced setup"**

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

1. For an ensemble analysis, increase the number of runs in the ensemble, in this case to 50. *In practice you would want to use a larger ensemble size (100-5000) than we are using in the demo.* The ensemble analysis samples parameters from their posterior distributions to propagate this uncertainty into the model output.

2. 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 (e.g. 1 and -1 are converted to 0.157 and 0.841). In this example set Sensitivity to -2,-1,1,2 (the median value, 0, occurs by default).

   We also can tell PEcAn which variable to run the sensitivity on. Here, replace NPP with

NEE, so we can compare against flux tower NEE observations.

Click Next

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 outputs that were present earlier that we'll explore in greater details:

1. **Run ID**:
   While the sensitivity and ensemble analyses synthesize across runs, you can also select individual runs from the Run ID menu. You can use the Graphs menu to visualize each individual run, or open individual runs in Shiny

2. **Inputs:**
   This menu shows the contents of /run which lets you look at and download:
   1. A summary file (README.txt) describing each run: location, run ID, model, dates, whether it was in the sensitivity or ensemble analysis, variables modifed, etc.
   2. The model-specific input files fed into the model
   3. The jobs.sh file used to submit the model run

3. **Outputs:**
   This menu shows the contents of /out. A number of files generated by the underlying ecosystem model are archived and available for download. These include:
   1. Output files in the standardized netCDF ([year].nc) that can be downloaded for visualization and analysis (R, Matlab, ncview, panoply, etc)
   2. Raw model output in model-specific format (e.g. sipnet.out).
   3. Logfile.txt contains job.sh & model error, warning, and informational messages

4. **PFTs**:
   This menu shows the contents of /pft. There is a wide array of outputs available that are related to the process of estimating the model parameters and running sensitivity/uncertainty analyses for a specific Plant Functional Type.
   1. **TRAITS**: 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. The list of **variables** that are queried is determined by what variables have **priors** associated with them in the definition of the **PFTs**. Priors are output into **prior.distns.Rdata**. 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. To look at these files in RStudio 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.

2. **META-ANALYSIS:**
   1. ***.bug**: The evaluation of the meta-analysis is done using a Bayesian statistical

summary plots of beta.o for sipnet.temperate.coniferous SLA
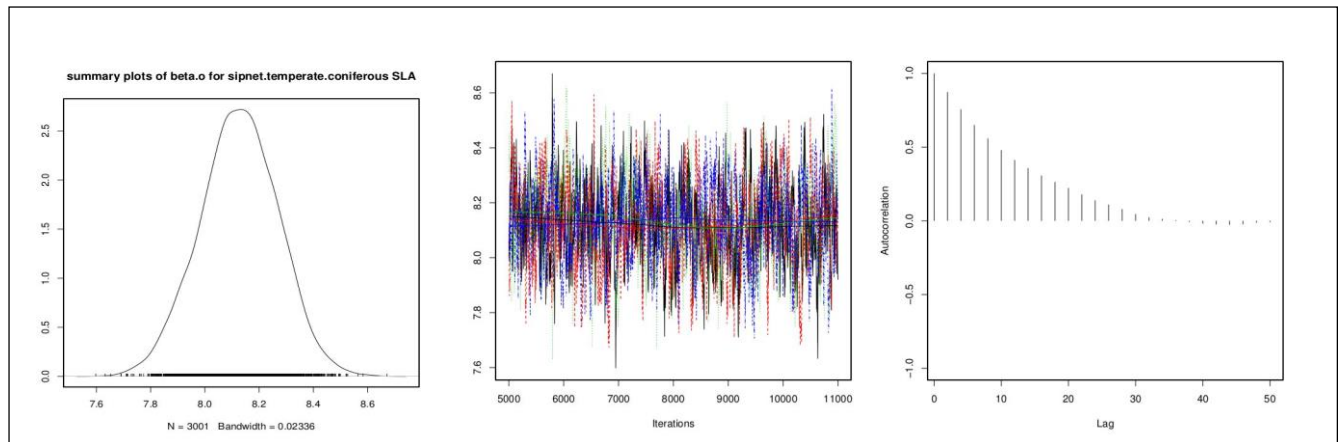
   software package called JAGS that is called by the R code. For each trait, the R code will generate a [trait].model.bug file that is the JAGS code for the meta-analysis itself. This code is generated on the fly, with PEcAn adding or subtracting the site, treatment, and greenhouse terms depending upon 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.
   2. **meta-analysis.log** contains 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).
   3. **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. 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).

   4. **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. **post.distns.Rdata** stores a simple tables of the posterior distributions for all model parameters in terms of the name of the distribution and its parameters.
   6. **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*). Open posteriors.pdf and compare the posteriors to the priors and data.

3. **SENSITIVITY ANALYSIS**
   1. **sensitivity.analysis.[RunID].[Variable].[StartYear].[EndYear].pdf** shows the raw data points from univariate one-at-a-time analyses and spline fits

through the points. ==Open this file== to determine which parameters are most and least sensitive

4. **UNCERTAINTY ANALYSIS**
   1. **variance.decomposition.[RunID].[Variable].[StartYear].[EndYear].pdf**, contains three columns, the coefficient of variation (normalized posterior variance), the elasticity (normalized sensitivity), and the partial standard deviation of each model parameter. ==Open this file for BOTH the soil and conifer PFTS and answer the following questions:==
   2. 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.
   3. 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).
   4. 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.
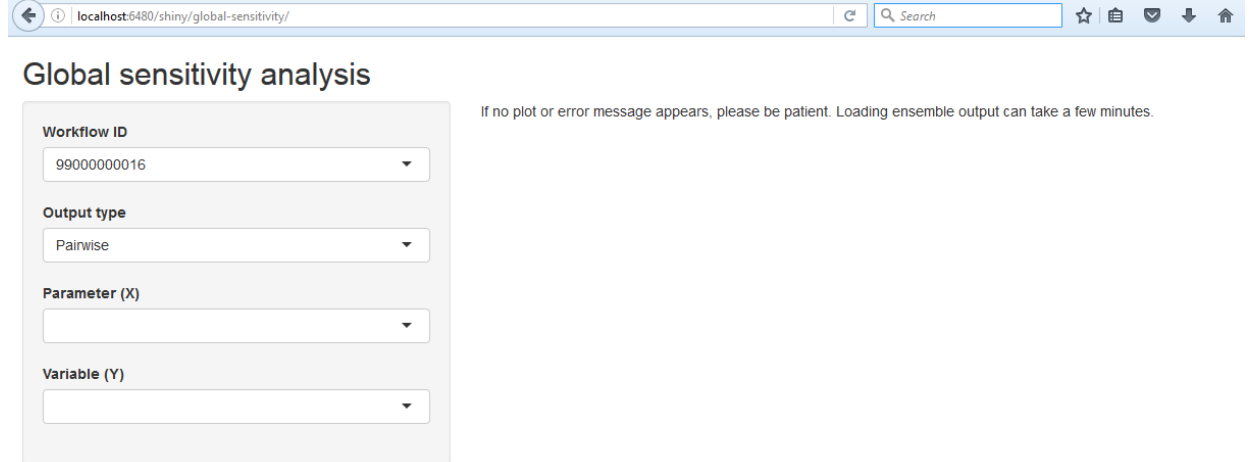
5. **PEcAn Files:**
   This menu shows the contents of the root workflow folder that are not in one of the folders indicated above. It mostly contains log files from the PEcAn workflow that are useful if the workflow generates an error, and as metadata & provenance (a detailed record of how data was generated).
   1. **STATUS** gives a summary of the steps of the workflow, the time they took, and whether they were successful
   2. **pecan.*.xml** are PEcAn settings files
   3. **workflow.R** is the workflow script
   4. **workflow.Rout** is the corresponding log file
   5. **samples.Rdata** contains the parameter values used in the runs. This file contains two data objects, sa.samples and ensemble.samples, that are the parameter values for the sensitivity analysis and ensemble runs respectively
   6. **sensitivity.output.[RunID].[Variable].[StartYear].[EndYear].Rdata** contains the object sensitivity.output which is the model outputs corresponding to the parameter values in sa.samples.
   7. **ENSEMBLE ANALYSIS**
      1. **ensemble.Rdata** contains contains the object ensemble.output, which is the model predictions at the parameter values given in ensemble.samples.
      2. **ensemble.analysis.[RunID].[Variable].[StarYear].[EndYear].pdf** contains the ensemble prediction as both a histogram and a boxplot.
      3. **ensemble.ts.[RunID].[Variable].[StartYear].[EndYear].pdf** contains a time-series plot of the ensemble mean, median, and 95% CI

6. **Global Sensitivity**: Shiny
   Navigate to <mark>URL/shiny/global-sensitivity</mark>.

This app uses the output from the ENSEMBLE runs to perform a global Monte Carlo sensitivity analysis. There are three modes controlled by **Output type**:
1. **Pairwise** looks at the relationship between a specific parameter (X) and output (Y)
2. **All parameters** looks at how all parameters affect a specific output (Y)
3. **All variables** looks at how all outputs are affected by a specific parameter(X)

In all of these analyses, the app also fits a linear regression to these scatterplots and reports a number of summary statistics. Among these, the slope is an indicator of **global sensitivity** and the R^2 is an indicator of the contribution to **global uncertainty**

Next Steps

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.

1. Explore how model error changes as a function of parameter value (i.e. data assimilation 'by hand')

2. Explore Bayesian MCMC concepts using the photosynthesis module

More info about tools, analyses, and specific tasks…

Additional information about specific tasks (adding sites, models, data; software updates; etc.) and analyses (e.g. data assimilation) can be found in the PEcAn documentation

If you encounter a problem with PEcAn that's not covered in the documentation, or if PEcAn is missing functionality you need, please search known bugs and issues, submit a bug report, or ask a question in our chat room. Additional questions can be directed to the project manager