SB pipe documentation

Release 1.18.0

Piero Dalle Pezze and Nicolas Le Novère

November 04, 2016

CHAPTER

ONE

USER MANUAL

Contacts: Dr Piero Dalle Pezze (piero.dallepezze AT babraham.ac.uk) and Dr Nicolas Le Novère (lenov AT babraham.ac.uk)

Affiliation: The Babraham Institute, Cambridge, CB22 3AT, UK

Mailing list: sbpipe AT googlegroups.com

Forum: https://groups.google.com/forum/#!forum/sbpipe

Introduction

This package contains a collection of pipelines for dynamic modelling of biological systems. It aims to automate common processes and speed up productivity for tasks such as model simulation, single/double parameter scan, and parameter estimation.

Requirements

In order to use SB pipe, the following software must be installed:

- Copasi 4.16 - http://copasi.org/
- Python 2.7+ - https://www.python.org/
- R 3.2.3+ - https://cran.r-project.org/

If LaTeX/PDF reports are also desired, the following software must also be installed:

- LaTeX 2013

Depending on your operating system, LaTeX can be downloaded at these websites:

- GNU/Linux: https://latex-project.org/ftp.html
- Windows: https://miktex.org/

GNU/Linux

It is advised that users install Python, R and (optionally) LaTeX packages using the package manager of their GNU/Linux distribution. Users need to make sure that the packages `python-pip` and `texlive-latex-base` (only for reports). In most cases, the installation via the package manager will automatically configure the correct environment variables.

If a local installation of Python, R, or LaTeX is needed, users need to add the following environment variables to `$PATH` in their `$HOME`/.bashrc file as follows:

```
# Path to R
export PATH=$PATH:/path/to/R/binaries/

# Path to Python. Scripts is the folder (if any) containing the Python
# script `pip`. pip must be available via command line.
export PATH=$PATH:/path/to/Python/:/path/to/Python/Scripts/

# Path to LaTeX
export PATH=$PATH:/path/to/LaTeX/binaries/
```

The correct installation of Python, R, and LaTeX can be tested by running the commands:

```
# If variables were manually exported, reload the .bashrc file
$ source $HOME/.bashrc

$ python -V
Python 2.7.12
$ pip -V
pip 8.1.2 from /home/ariel/.local/lib/python2.7/site-packages (python 2.7)

$ R --version
R version 3.2.3 (2015-12-10) -- "Wooden Christmas-Tree"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

$ pdflatex -v
pdfTeX 3.14159265-2.6-1.40.16 (TeX Live 2015/Debian)
kpathsea version 6.2.1
Copyright 2015 Peter Breitenlohner (eTeX)/Han The Thanh (pdfTeX).
```

As of 2016, Copasi is not available as a package in GNU/Linux distributions. Users must add the path to Copasi binary files manually editing their GNU/Linux `$HOME`/.bashrc file as follows:

```
# Path to CopasiSE
export PATH=$PATH:/path/to/CopasiSE/
```

The correct installation of CopasiSE can be tested by running the command:

```
# Reload the .bashrc file
$ source $HOME/.bashrc

$ CopasiSE -h
COPASI 4.16 (Build 104)
```

At this stage, Python, R, Copasi, and (optionally) LaTeX should be installed correctly. SB pipe requires the configuration of the environment variable `$SBPIPE` which must also be added in the `$HOME`/.bashrc file. The package also needs to be added to `$PATH`. To do so, users need to add the following lines to their `$HOME`/.bashrc file:

```
# SBPIPE
export SBPIPE=/path/to/sbpipe
export PATH=$PATH:$SBPIPE/scripts
```

Now you should reload the .bashrc file to make the previous change effective:

```
# Reload the .bashrc file
$ source $HOME/.bashrc
```

Before testing the correct installation of SB pipe, users need to install Python and R dependency packages used by SB pipe. Two scripts are provided to perform these tasks automatically.

To install SB pipe Python dependencies on GNU/Linux, run:

```
$ cd $SBPIPE/
$ ./install_pydeps.py
```

To install SB pipe R dependencies on GNU/Linux, run:

```
$ cd $SBPIPE/
$ R
# Inside R environment, answer 'y' to install packages locally
> source('install_rdeps.r')
```

If R package dependencies must be compiled, it is worth checking that the following additional packages are installed in your machine: `build-essential`, `liblapack-dev`, `libblas-dev`, `libcairo-dev`, `libssl-dev`, `libcurl4-openssl-dev`. After installing these packages, `install_rdeps.r` must be executed again.

The correct installation of SB pipe can be tested by running the command:

```
$ run_sbpipe.py -v
1.17.0
```

Windows

Windows users are also strongly advised to install the package:

- Cygwin 2.6.0 https://www.cygwin.com/

Cygwin offers a GNU/Linux-like shell. This makes the installation of dependencies easier as this follows the configuration for GNU/Linux users.

Windows users may need to edit the `PATH` environment variable so that the binary files for the previous packages (Copasi, Python, R, and (optionally) LaTeX) are correctly found. Specifically for Python, the python scripts `pip.py` and `easy_install.py` are located inside the folder `Scripts` within the Python root directory. The path to this folder must also be added to `PATH`.

Therefore, the following environment variables must also be added:

```
SBPIPE=\path\to\sbpipe
PATH=[previous paths];%SBPIPE%\scripts
```

NOTE for Cygwin: Environment variables can also be configured directly within the .bashrc file in cygwin/home/USERNAME/. In the beginning of this file, users should place:

```
# Path to R
export PATH=$PATH:/path/to/R/binaries/

# Path to Python
export PATH=$PATH:/path/to/Python/:/path/to/Python/Scripts/

# Path to LaTeX
export PATH=$PATH:/path/to/LaTeX/binaries/

# Path to CopasiSE
export PATH=$PATH:/path/to/CopasiSE/binaries/

# SBPIPE
export SBPIPE=/path/to/sbpipe
export PATH=$PATH:$SBPIPE/scripts
```

After configuring the environment variables directly or internally in Cygwin, the next step is to install Python and R packages used by SB pipe. Two scripts are provided to perform these tasks automatically.

To install SB pipe Python dependencies using Cygwin on Windows, run:

```
$ cd /cygdrive/PATH/TO/SBPIPE/
$ python.exe install_pydeps.py
```

To install SB pipe R dependencies using Cygwin on Windows, run:

```
$ cd /cygdrive/PATH/TO/SBPIPE/
$ R.exe
# Inside R environment, answer 'y' to install packages locally
> source('install_rdeps.r')
```

Installation

If desired, SB pipe can be installed in your system. To do so, run the command inside the sbpipe folder:

```
$ cd $SBPIPE
$ python setup.py install
```

The correct installation of SB pipe and its dependencies can be checked by running the following commands inside the SB pipe folder:

```
$ cd $SBPIPE/tests
$ ./test_suite.py
```

How to use SB pipe

Preliminary configuration steps

Pipelines using Copasi

Before using these pipelines, a Copasi model must be configured as follow using CopasiUI:

pipeline: simulation

- Tick the flag *executable* in the Time Course Task.
- Select a report template for the Time Course Task.
- Save the report in the same folder with the same name as the model but replacing the extension .cps with .csv.

pipelines: single or double parameter scan

- Tick the flag *executable* in the Parameter Scan Task.
- Select a report template for the Parameter Scan Task.
- Save the report in the same folder with the same name as the model but replacing the extension .cps with .csv.

pipeline: parameter estimation

- Tick the flag *executable* in the Parameter Estimation Task.
- Select the report template for the Parameter Estimation Task.
- Save the report in the same folder with the same name as the model but replacing the extension .cps with .csv.

Running SB pipe

SB pipe is executed via the command *run_sbpipe.py*. The syntax for this command and its complete list of options can be retrieved by running *run_sbpipe.py -h*.

As of Sep 2016 the output is as follows:

```
$ run_sbpipe.py -h
Usage: run_sbpipe.py [OPTION] [FILE]
Pipelines for systems modelling of biological networks.

List of mandatory options:
        -h, --help
                Shows this help.
        -c, --create-project
                Create a project structure using the argument as name.
        -s, --simulate
                Simulate a model.
        -p, --single-param-scan
                Simulate a single parameter scan.
        -d, --double-param-scan
                Simulate a double parameter scan.
        -e, --param-estim
                Generate a parameter fit sequence.
        -l, --license
                Shows the license.
        -v, --version
                Shows the version.
Exit status:
 0  if OK,
 1  if minor problems (e.g., a pipeline did not execute correctly),
 2  if serious trouble (e.g., cannot access command-line argument).

Report bugs to sbpipe@googlegroups.com
SB pipe home page: <https://pdp10.github.io/sbpipe>
For complete documentation, see README.md .
```

The first step is to create a new project. This can be done with the command:

```
$ run_sbpipe.py --create-project project_name
```

This generates the following structure:

```
project_name/
    | - Data/
    | - Models/
    | - Working_Folder/
```

Models must be stored in the Models/ folder. The folder Data/ is meant for collecting experimental data files and analyses in one place. Once the data files for Copasi (e.g. for parameter estimation) are generated, **it is advised** to move them into the Models/ folder so that the Copasi (.cps) file and its associated experimental data files are stored in the same folder. To run SB pipe, users need to create a configuration file for each pipeline they intend to run (see next section). These configuration files should be placed in the Working_Folder/. This folder will eventually contain all the results generated by SB pipe.

For instance, the pipeline for parameter estimation configured with a certain configuration file can be executed by typing:

```
$ run_sbpipe.py -e my_config_file.conf
```

Pipeline configuration files

Pipelines are configured using files (here called configuration files). These files are INI files and are therefore structured as follows:

```
[pipeline_name]
option1=value1
option2=value2
...
```

In SB pipe each pipeline executes three tasks: data generation, data analysis, and report generation. Each task depends on the previous one. This choice allows users to analyse the same data without having to generate it every time, or to skip the report generation if not wanted. Assuming that the configuration files are placed in the Working_Folder of a certain project, examples are given as follow:

Example 1: configuration file for the pipeline *simulation*

```
[simulate]
# True if data should be generated, False otherwise
generate_data=True
# True if data should be analysed, False otherwise
analyse_data=True
# True if a report should be generated, False otherwise
generate_report=True
# The relative path to the project directory (from Working_Folder)
project_dir=..
# The name of the configurator (e.g. Copasi)
simulator=Copasi
# The Copasi model name
model=insulin_receptor_stoch.cps
# The cluster type. pp if the model is run locally,
# sge/lsf if run on cluster.
cluster=pp
# The number of CPU if pp is used, ignored otherwise
pp_cpus=7
# The number of simulations to perform.
# n>=1 for stochastic simulations.
runs=40
# An experimental data set (or blank) to add to the
# simulated plots as additional layer
exp_dataset=insulin_receptor_dataset.csv
# True if the experimental data set should be plotted.
plot_exp_dataset=True
# The label for the x axis.
xaxis_label=Time [min]
# The label for the y axis.
yaxis_label=Level [a.u.]
```

Example 2: configuration file for the pipeline *single parameter scan*

```
[single_param_scan]
# True if data should be generated, False otherwise
generate_data=True
# True if data should be analysed, False otherwise
analyse_data=True
# True if a report should be generated, False otherwise
generate_report=True
# The relative path to the project directory (from Working_Folder)
project_dir=..
# The name of the configurator (e.g. Copasi)
simulator=Copasi
# The Copasi model name
model=insulin_receptor_inhib_scan_IR_beta.cps
# The variable to scan (as set in Copasi Parameter Scan Task)
```

```
scanned_par=IR_beta
# The number of intervals in the simulation
simulate__intervals=100
# The number of simulations to perform for each scan
single_param_scan_simulations_number=1
# True if the variable is only reduced (knock down), False otherwise.
single_param_scan_knock_down_only=True
# True if the scanning represents percent levels.
single_param_scan_percent_levels=True
# The minimum level (as set in Copasi Parameter Scan Task)
min_level=0
# The maximum level (as set in Copasi Parameter Scan Task)
max_level=100
# The number of scans (as set in Copasi Parameter Scan Task)
levels_number=10
# True if plot lines are the same between scans
# (e.g. full lines, same colour)
homogeneous_lines=False
# The label for the x axis.
xaxis_label=Time [min]
# The label for the y axis.
yaxis_label=Level [a.u.]
```

Example 3: configuration file for the pipeline *double parameter scan*

```
[double_param_scan]
# True if data should be generated, False otherwise
generate_data=True
# True if data should be analysed, False otherwise
analyse_data=True
# True if a report should be generated, False otherwise
generate_report=True
# The relative path to the project directory (from Working_Folder)
project_dir=..
# The name of the configurator (e.g. Copasi)
simulator=Copasi
# The Copasi model name
model=insulin_receptor_inhib_dbl_scan_InsulinPercent__IRbetaPercent.cps
# The 1st variable to scan (as set in Copasi Parameter Scan Task)
scanned_par1=InsulinPercent
# The 2nd variable to scan (as set in Copasi Parameter Scan Task)
scanned_par2=IRbetaPercent
# The simulation length (as set in Copasi Time Course Task)
sim_length=10
```

Example 4: configuration file for the pipeline *parameter estimation*

```
[param_estim]
# True if data should be generated, False otherwise
generate_data=True
# True if data should be analysed, False otherwise
analyse_data=True
# True if a report should be generated, False otherwise
generate_report=True
# True if a zipped tarball should be generated, False otherwise
generate_tarball=True
# The relative path to the project directory (from Working_Folder)
project_dir=..
# The name of the configurator (e.g. Copasi)
simulator=Copasi
# The Copasi model name
model=insulin_receptor_param_estim.cps
# The cluster type. pp if the model is run locally,
```

```
# sge/lsf if run on cluster.
cluster=pp
# The number of CPU if pp is used, ignored otherwise
pp_cpus=7
# The parameter estimation round which is used to distinguish
# phases of parameter estimations when parameters cannot be
# estimated at the same time
round=1
# The number of parameter estimations
# (the length of the fit sequence)
runs=250
# The threshold percentage of the best fits to consider
best_fits_percent=75
# The number of available data points
data_point_num=33
# True if 2D all fits plots for 66% confidence levels
# should be plotted. This can be computationally expensive.
plot_2d_66cl_corr=True
# True if 2D all fits plots for 95% confidence levels
# should be plotted. This can be computationally expensive.
plot_2d_95cl_corr=True
# True if 2D all fits plots for 99% confidence levels
# should be plotted. This can be computationally expensive.
plot_2d_99cl_corr=True
# True if parameter values should be plotted in log space.
logspace=True
# True if plot axis labels should be plotted in scientific notation.
scientific_notation=True
```

Additional examples of configuration files can be found in:

```
$SBPIPE/tests/insulin_receptor/Working_Folder/
```

Reporting bugs or requesting new features

SB pipe is a relatively young project and there is a chance that some error occurs. The following mailing list should be used for general questions:

```
sbpipe AT googlegroups.com
```

All the topics discussed in this mailing list are also available at the website:

https://groups.google.com/forum/#!forum/sbpipe

To help us better identify and reproduce your problem, some technical information is needed. This detail data can be found in SB pipe log files which are stored in ${HOME}/.sbpipe/logs/. When using the mailing list above, it would be worth providing this extra information.

Issues and feature requests can also be notified using the github issue tracking system for SB pipe at the web page:

https://github.com/pdp10/sbpipe/issues.

CHAPTER

TWO

DEVELOPER MANUAL

Mailing list: sbpipe AT googlegroups.com

Forum: https://groups.google.com/forum/#!forum/sbpipe

Introduction

This guide is meant for developers and contains guidelines for developing this project.

Development model

This project follows the Feature-Branching model. Briefly, there are two main branches: `master` and `develop`. The former contains the history of stable releases, the latter contains the history of development. The `master` branch contains checkout points for production hotfixes or merge points for release-x.x.x branches. The `develop` branch is used for feature-bugfix integration and checkout point in development. Nobody should directly develop in here. The `develop` branch is versionless (just call it *-dev*).

Conventions

To manage the project in a more consistent way, here is a list of conventions to follow:

- Each new feature is developed in a separate branch forked from *develop*. This new branch is called *featureNUMBER*, where *NUMBER* is the number of the GitHub Issue discussing that feature. The first line of each commit message for this branch should contain the string *Issue #NUMBER* at the beginning. Doing so, the commit is automatically recorded by the Issue Tracking System for that specific Issue. Note that the sharp (#) symbol is required.
- The same for each new bugfix, but in this case the branch name is called bugfixNUMBER.
- The same for each new hotfix, but in this case the branch name is called hotfixNUMBER and is forked from *master*.

Work flow

The procedure for checking out a new feature from the `develop` branch is:

```
$ git checkout -b feature10 develop
```

This creates the `feature10` branch off `develop`. This feature10 is discussed in *Issue #10* in GitHub. When you are ready to commit your work, run:

```
$ git commit -am "Issue #10, summary of the changes. Detailed
description of the changes, if any."
$ git push origin feature10       # sometimes and at the end.
```

As of June 2016, the branches `master` and `develop` are protected and a status check using Travis-CI must be performed before merging or pushing into these branches. This automatically forces a merge without fast-forward. In order to merge **any** new feature, bugfix or simple edits into `master` or `develop`, a developer **must** checkout a new branch and, once committed and pushed, **merge** it to `master` or `develop` using a `pull request`. To merge `feature10` to `develop`, the pull request output will look like this in GitHub Pull Requests:

```
base:develop  compare:feature10   Able to merge. These branches can be
automatically merged.
```

A small discussion about feature10 should also be included to allow other users to understand the feature.

Finally delete the branch:

```
$ git branch -d feature10      # delete the branch feature10 (locally)
```

New releases

When the `develop` branch includes all the desired feature for a release, it is time to checkout this branch in a new one called `release-x.x.x`. It is at this stage that a version is established. Only bugfixes or hotfixes are applied to this branch. When this testing/correction phase is completed, the `master` branch will merge with the `release-x.x.x` branch, using the commands above. To record the release add a tag:

```
git tag -a v1.3 -m "PROGRAM_NAME v1.3"
```

To transfer the tag to the remote server:

```
git push origin v1.3   # Note: it goes in a separate 'branch'
```

To see all the releases:

```
git show
```

Package structure

This section presents the structure of the SB pipe package. The root of the project contains general management scripts for installing Python and R dependencies (install_pydeps.py and install_rdeps.r), and installing SB pipe (setup.py). Additionally, the logging configuration file (logging_config.ini) is also at this level.

In order to automatically compile and run the test suite, Travis-CI is used and configured accordingly (.travis.yml).

The project is structured as follows:

```
sbpipe:
  | - docs/
  | - sbpipe/
        | - pl
        | - R
        | - report
        | - simul
        | - utils
  | - scripts/
  | - tests/
```

These folders will be discussed in the next sections. In SB pipe, Python is the project main language. Instead, R is essentially used for computing statistics (see section configuration file in the user manual) and for generating plots. This choice allows users to run these scripts independently of SB pipe if needed using an R environment like Rstudio. This can be convenient if further data analysis are needed or plots need to be annotated or edited.

docs

The folder `docs/` contains the documentation for this project. The user and developer manuals in markdown format are contained in `docs/source`. In order to generate the complete documentation for SB pipe, the following packages must be installed:

- python-sphinx
- pandoc
- texlive-fonts-recommended
- texlive-latex-extra

By default the documentation is generated in html and LaTeX/PDF. Instruction for generating or cleaning SB pipe documentation are provided below.

To generate the source code documentation:

```
$ cd $SBPIPE/docs
$ ./gen_doc.sh
```

To clean the documentation:

```
$ cd $SBPIPE/docs
$ ./cleanup_doc.sh
```

The complete source code documentation for this project is stored in `docs/build/html` (html format) and `docs/build/latex` (LaTeX/PDF format). A shortcut to the documentation in html format is available at the page `docs/index.html`

sbpipe

This folder contains the source code of the project SB pipe. At this level a file called `__main__.py` enables users to run SB pipe as a Python module via the command:

```
$ python sbpipe
```

Alternatively `sbpipe` can be imported as shown below:

```
$ cd $SBPIPE
$ python
# Python environment
>>> import sbpipe.main as sb
>>> sb.version()
'1.25.0 '
```

The following subsections describe sbpipe subpackages.

pl

The subpackage `sbpipe.pl` contains the class `Pipeline` in the file `pipeline.py`. This class represents a generic pipeline which is extended by SB pipe pipelines. These are organised in the following subpackages:

- `create`: creates a new project
- `ps1`: scan a model parameter, generate plots and report;
- `ps2`: scan two model parameters, generate plots and report;
- `pe`: generate a fits sequence, tables of statistics and plots.
- `sim`: simulates a model deterministically or stochastically, generate plots and report;

All these pipelines can be invoked directly via the script `$SBPIPE/scripts/run_sbpipe.py`. Each SB pipe pipeline extends the class `Pipeline` and therefore must implement the following methods:

```
def run(self, config_file)
def read_config(self, lines)
```

The former contains the procedure to execute an INI configuration file. The latter parses the pipeline options. The `Pipeline` class implements the INI parser which is therefore available to each pipeline. The INI parser returns the configuration file as a list of lines.

R

This folder contains a collection of R utility methods for plotting and generating statistics. These utilities are used by the pipelines during data analysis.

report

The subpackage `sbpipe.report` contains Python modules for generating LaTeX/PDF reports.

simul

The subpackage `sbpipe.simul` contains the class `Simul` in the file `simul.py`. This is a generic simulator interface used by the pipelines in SB pipe. This mechanism uncouples pipelines from specific simulators which can therefore be configured in each pipeline configuration file. As of 2016, the only available simulator is `Copasi` and this is available in the package `sbpipe.simul.copasi`. This implements all the method of the class `Simul`.

Pipelines can dynamically load a simulator via the class method `Pipeline.get_simul_obj(simulator)`. This method instantiates an object of subtype `Simul` by refractoring the simulator name as parameter. A simulator class (e.g. `Copasi`) must have the same name of their package (e.g. `copasi`) but start with an upper case letter. A simulator class must be contained in a file with the same name of their package (e.g. `copasi`). Therefore, for each simulator package, exactly one simulator class can be instantiated. Simulators can be configured in the configuration file using the field `simulator`.

utils

The subpackage `sbpipe.utils` contains a collection of Python utility modules which are used by sbpipe.

scripts

The folder `scripts` contains the scripts: `cleanup_sbpipe.py` and `run_sbpipe.py`. `run_sbpipe.py` is the main script and is used to run the pipelines. `cleanup_sbpipe.py` is used for cleaning the package including the test results.

tests

The package `tests` contains the script `test_suite.py` which executes all sbpipe tests. It should be used for testing the correct installation of SB pipe dependencies as well as reference for configuring a project before running any pipeline. Projects inside the folder `$SBPIPE/tests/` have the SB pipe project structure:

- `Data`: (e.g. training / testing data sets for the model);
- `Model`: (e.g. models, Copasi models, data sets directly used by Copasi models);
- `Working_Folder`: (e.g. pipelines configurations and parameter estimation results, time course, parameter scan, etc).

Examples of configuration files (*.conf) can be found in $SBPIPE/tests/insulin_receptor/Working_Folder/.

As of 2016, the repository for SB pipe source code is `github.com`. This is configured to run Travis-CI every time a `git push` into the repository is performed. The exact details of execution of Travis-CI can be found in Travis-CI configuration file `$SBPIPE/.travis.yml`. Importantly, Travis-CI runs all SB pipe tests using `nosetests`.

Miscellaneous of useful commands

Git

Startup

```
# clone master
$ git clone https://github.com/pdp10/sbpipe.git
# get develop branch
$ git checkout -b develop origin/develop
# to get all the other branches
$ for b in `git branch -r | grep -v -- '->'`; do git branch
--track ${b##origin/} $b; done
# to update all the branches with remote
$ git fetch --all
```

Update

```
# ONLY use --rebase for private branches. Never use it for shared
# branches otherwise it breaks the history. --rebase moves your
# commits ahead. For shared branches, you should use
# `git fetch && git merge --no-ff`
$ git pull [--rebase] origin BRANCH
```

File system

```
$ git rm [--cache] filename
$ git add filename
```

Information

```
$ git status
$ git log [--stat]
$ git branch       # list the branches
```

Maintenance

```
$ git fsck      # check errors
$ git gc        # clean up
```

Rename a branch locally and remotely

```
git branch -m old_branch new_branch         # Rename branch locally
git push origin :old_branch                 # Delete the old branch
git push --set-upstream origin new_branch   # Push the new branch, set
local branch to track the new remote
```

Reset

```
git reset --hard HEAD    # to undo all the local uncommitted changes
```

Syncing a fork (assuming upstreams are set)

```
git fetch upstream
git checkout develop
git merge upstream/develop
```

CHAPTER

THREE

SOURCE CODE

Python modules

sbpipe package

Subpackages

sbpipe.pl package

Subpackages

sbpipe.pl.create package

Submodules

sbpipe.pl.create.newproj module

class sbpipe.pl.create.newproj.**NewProj**(*data_folder='Data'*, *models_folder='Models'*, *working_folder='Working_Folder'*)

Bases: *sbpipe.pl.pipeline.Pipeline* (page 22)

This module initialises the folder tree for a new project.

Parameters

- **data_folder** – the folder containing the data
- **models_folder** – the folder containing the models
- **working_folder** – the folder to store the results

run(*project_name*)

Create a project directory tree.

Parameters **project_name** – the name of the project

Returns 0

Module contents

sbpipe.pl.pe package

Submodules

sbpipe.pl.pe.collect_results module

sbpipe.pl.pe.collect_results.**get_all_fits**(*path_in='.'*, *path_out='.'*, *filename_out='all_estimates.csv'*)

Collect all the parameter estimates from the Copasi parameter estimation report. Results are stored in filename_out.

Parameters

- **path_in** – the path to the input files
- **path_out** – the path to the output files
- **filename_out** – the filename to store the final estimates

sbpipe.pl.pe.collect_results.**get_best_fits**(*path_in='.'*, *path_out='.'*, *filename_out='final_estimates.csv'*)

Collect the final parameter estimates from the Copasi parameter estimation report. Results are stored in filename_out.

Parameters

- **path_in** – the path to the input files
- **path_out** – the path to the output files
- **filename_out** – the filename to store the final estimates

sbpipe.pl.pe.collect_results.**get_input_files**(*path*)

Retrieve the input files in a path.

Parameters **path** – the path containing the input files to retrieve

Returns the list of input files

sbpipe.pl.pe.collect_results.**get_params_list**(*filein*)

Return the list of parameter names from filein

Parameters **filein** – a Copasi parameter estimation report file

Returns the list of parameter names

sbpipe.pl.pe.collect_results.**write_all_fits**(*files*, *path_out*, *filename_out*)

Write all the estimates to filename_out

Parameters

- **files** – the list of Copasi parameter estimation reports
- **path_out** – the path to store the file combining all the estimates
- **filename_out** – the file containing all the estimates

sbpipe.pl.pe.collect_results.**write_best_fits**(*files*, *path_out*, *filename_out*)

Write the final estimates to filename_out

Parameters

- **files** – the list of Copasi parameter estimation reports
- **path_out** – the path to store the file combining the final (best) estimates (filename_out)
- **filename_out** – the file containing the final (best) estimates

sbpipe.pl.pe.collect_results.**write_params**(*col_names*, *path_out*, *filename_out*)

Write the list of parameter names to filename_out

Parameters

- **col_names** – the list of parameter names
- **path_out** – the path to store filename_out
- **filename_out** – the output file to store the parameter names

sbpipe.pl.pe.parest module

class sbpipe.pl.pe.parest.**ParEst**(*data_folder='Data'*, *models_folder='Models'*, *working_folder='Working_Folder'*, *sim_data_folder='param_estim_data'*, *sim_plots_folder='param_estim_plots'*)

Bases: *sbpipe.pl.pipeline.Pipeline* (page 22)

This module provides the user with a complete pipeline of scripts for running model parameter estimations

classmethod analyse_data(*model*, *inputdir*, *outputdir*, *fileout_final_estims*, *fileout_all_estims*, *fileout_param_estim_details*, *fileout_param_estim_summary*, *sim_plots_dir*, *best_fits_percent*, *data_point_num*, *plot_2d_66cl_corr=False*, *plot_2d_95cl_corr=False*, *plot_2d_99cl_corr=False*, *logspace=True*, *scientific_notation=True*)

The second pipeline step: data analysis.

Parameters

- **model** – the model name
- **inputdir** – the directory containing the simulation data
- **outputdir** – the directory to store the results
- **fileout_final_estims** – the name of the file containing final parameter sets with Chi^2
- **fileout_all_estims** – the name of the file containing all the parameter sets with Chi^2
- **fileout_param_estim_details** – the name of the file containing the detailed statistics for the

estimated parameters :param fileout_param_estim_summary: the name of the file containing the summary for the parameter estimation :param sim_plots_dir: the directory of the simulation plots :param best_fits_percent: the percent to consider for the best fits :param data_point_num: the number of data points :param plot_2d_66cl_corr: True if 2 dim plots for the parameter sets within 66% should be plotted :param plot_2d_95cl_corr: True if 2 dim plots for the parameter sets within 95% should be plotted :param plot_2d_99cl_corr: True if 2 dim plots for the parameter sets within 99% should be plotted :param logspace: True if parameters should be plotted in log space :param scientific_notation: True if axis labels should be plotted in scientific notation :return: True if the task was completed successfully, False otherwise.

classmethod generate_data(*simulator*, *model*, *inputdir*, *cluster_type*, *pp_cpus*, *nfits*, *outputdir*, *sim_data_dir*, *updated_models_dir*)

The first pipeline step: data generation.

Parameters

- **simulator** – the name of the simulator (e.g. Copasi)
- **model** – the model to process
- **inputdir** – the directory containing the model
- **cluster_type** – pp for parallel python, lsf for load sharing facility, sge for sun grid engine
- **pp_cpus** – the number of cpu for parallel python
- **nfits** – the number of fits to perform
- **outputdir** – the directory to store the results
- **sim_data_dir** – the directory containing the simulation data sets
- **updated_models_dir** – the directory containing the models with updated parameters for each estimation

Returns True if the task was completed successfully, False otherwise.

classmethod generate_report (*model*, *outputdir*, *sim_plots_folder*)
The third pipeline step: report generation.

Parameters

- **model** – the model name
- **outputdir** – the directory to store the report
- **sim_plots_folder** – the folder containing the plots

Returns True if the task was completed successfully, False otherwise.

read_config (*lines*)

run (*config_file*)

Module contents

sbpipe.pl.ps1 package

Submodules

sbpipe.pl.ps1.parscan1 module

class sbpipe.pl.ps1.parscan1.**ParScan1** (*data_folder='Data'*, *models_folder='Models'*, *working_folder='Working_Folder'*, *sim_data_folder='single_param_scan_data'*, *sim_plots_folder='single_param_scan_plots'*)
Bases: *sbpipe.pl.pipeline.Pipeline* (page 22)

This module provides the user with a complete pipeline of scripts for computing single parameter scans.

classmethod analyse_data (*model*, *scanned_par*, *knock_down_only*, *outputdir*, *sim_data_folder*, *sim_plots_folder*, *simulations_number*, *percent_levels*, *min_level*, *max_level*, *levels_number*, *homogeneous_lines*, *xaxis_label*, *yaxis_label*)
The second pipeline step: data analysis.

Parameters

- **model** – the model name
- **scanned_par** – the scanned parameter
- **knock_down_only** – True for knock down simulation, false if also scanning over expression.
- **outputdir** – the directory containing the results
- **sim_data_folder** – the folder containing the simulated data sets
- **sim_plots_folder** – the folder containing the generated plots
- **simulations_number** – the number of simulations
- **percent_levels** – True if the levels are percents.
- **min_level** – the minimum level
- **max_level** – the maximum level
- **levels_number** – the number of levels
- **homogeneous_lines** – True if generated line style should be homogeneous
- **xaxis_label** – the name of the x axis (e.g. Time [min])

- **yaxis_label** – the name of the y axis (e.g. Level [a.u.])

Returns True if the task was completed successfully, False otherwise.

classmethod generate_data(*simulator*, *model*, *scanned_par*, *sim_number*, *simulate_intervals*, *single_param_scan_intervals*, *inputdir*, *outputdir*)

The first pipeline step: data generation.

Parameters

- **simulator** – the name of the simulator (e.g. Copasi)
- **model** – the model to process
- **scanned_par** – the scanned parameter
- **sim_number** – the number of simulations (for det sim: 1, for stoch sim: n>1)
- **simulate_intervals** – the time step of each simulation
- **single_param_scan_intervals** – the number of scans to perform
- **inputdir** – the directory containing the model
- **outputdir** – the directory to store the results

Returns True if the task was completed successfully, False otherwise.

classmethod generate_report(*model*, *scanned_par*, *outputdir*, *sim_plots_folder*)

The third pipeline step: report generation.

Parameters

- **model** – the model name
- **scanned_par** – the scanned parameter
- **outputdir** – the directory containing the report
- **sim_plots_folder** – the folder containing the plots

Returns True if the task was completed successfully, False otherwise.

read_config(*lines*)

run(*config_file*)

Module contents

sbpipe.pl.ps2 package

Submodules

sbpipe.pl.ps2.parscan2 module

class sbpipe.pl.ps2.parscan2.**ParScan2**(*data_folder='Data'*, *models_folder='Models'*, *working_folder='Working_Folder'*, *sim_data_folder='double_param_scan_data'*, *sim_plots_folder='double_param_scan_plots'*)

Bases: *sbpipe.pl.pipeline.Pipeline* (page 22)

This module provides the user with a complete pipeline of scripts for computing double parameter scans.

classmethod analyse_data(*model*, *scanned_par1*, *scanned_par2*, *inputdir*, *outputdir*)

The second pipeline step: data analysis.

Parameters

- **model** – the model name

- **scanned_par1** – the first scanned parameter
- **scanned_par2** – the second scanned parameter
- **inputdir** – the directory containing the simulated data sets to process
- **outputdir** – the directory to store the performed analysis

Returns True if the task was completed successfully, False otherwise.

classmethod generate_data(*simulator*, *model*, *sim_length*, *inputdir*, *outputdir*)
The first pipeline step: data generation.

Parameters

- **simulator** – the name of the simulator (e.g. Copasi)
- **model** – the model to process
- **sim_length** – the length of the simulation
- **inputdir** – the directory containing the model
- **outputdir** – the directory to store the results

Returns True if the task was completed successfully, False otherwise.

classmethod generate_report(*model*, *scanned_par1*, *scanned_par2*, *outputdir*, *sim_plots_folder*)
The third pipeline step: report generation.

Parameters

- **model** – the model name
- **scanned_par1** – the first scanned parameter
- **scanned_par2** – the second scanned parameter
- **outputdir** – the directory containing the report
- **sim_plots_folder** – the folder containing the plots.

Returns True if the task was completed successfully, False otherwise.

read_config(*lines*)

run(*config_file*)

Module contents

sbpipe.pl.sens package

Submodules

sbpipe.pl.sens.sens module

class sbpipe.pl.sens.sens.**Sens**(*data_folder='Data'*, *models_folder='Models'*, *working_folder='Working_Folder'*, *sim_data_folder='sensitivity_data'*, *sim_plots_folder='sensitivity_plots'*)
Bases: *sbpipe.pl.pipeline.Pipeline* (page 22)

This module provides the user with a complete pipeline of scripts for computing model sensitivity analysis.

classmethod analyse_data(*outputdir*)
The second pipeline step: data analysis.

Parameters **outputdir** – the directory to store the performed analysis.

Returns True if the task was completed successfully, False otherwise.

classmethod generate_data(*simulator*, *model*, *inputdir*, *outputdir*)
The first pipeline step: data generation.

Parameters

- **simulator** – the name of the simulator (e.g. Copasi)
- **model** – the model to process
- **inputdir** – the directory containing the model
- **outputdir** – the directory to store the results

Returns True if the task was completed successfully, False otherwise.

classmethod generate_report(*model*, *outputdir*, *sim_plots_folder*)
The third pipeline step: report generation.

Parameters

- **model** – the model name
- **outputdir** – the directory to store the report
- **sim_plots_folder** – the directory containing the time courses results combined with experimental data

Returns True if the task was completed successfully, False otherwise.

read_config(*lines*)

run(*config_file*)

Module contents

sbpipe.pl.sim package

Submodules

sbpipe.pl.sim.sim module

class sbpipe.pl.sim.sim.**Sim**(*data_folder='Data'*, *models_folder='Models'*, *working_folder='Working_Folder'*, *sim_data_folder='simulate_data'*, *sim_plots_folder='simulate_plots'*)
Bases: *sbpipe.pl.pipeline.Pipeline* (page 22)

This module provides the user with a complete pipeline of scripts for running model simulations

classmethod analyse_data(*model*, *inputdir*, *outputdir*, *sim_plots_dir*, *exp_dataset*, *plot_exp_dataset*, *xaxis_label*, *yaxis_label*)
The second pipeline step: data analysis.

Parameters

- **model** – the model name
- **inputdir** – the directory containing the data to analyse
- **outputdir** – the output directory containing the results
- **sim_plots_dir** – the directory to save the plots
- **exp_dataset** – the full path of the experimental data set
- **plot_exp_dataset** – True if the experimental data set should also be plotted

- **xaxis_label** – the label for the x axis (e.g. Time [min])
- **yaxis_label** – the label for the y axis (e.g. Level [a.u.])

Returns True if the task was completed successfully, False otherwise.

classmethod generate_data(*simulator*, *model*, *inputdir*, *outputdir*, *cluster_type='pp'*, *pp_cpus=2*, *runs=1*)

The first pipeline step: data generation.

Parameters

- **simulator** – the name of the simulator (e.g. Copasi)
- **model** – the model to process
- **inputdir** – the directory containing the model
- **outputdir** – the directory containing the output files
- **cluster_type** – pp for local Parallel Python, lsf for Load Sharing Facility, sge for Sun Grid Engine.
- **pp_cpus** – the number of CPU used by Parallel Python.
- **runs** – the number of model simulation

Returns True if the task was completed successfully, False otherwise.

classmethod generate_report(*model*, *outputdir*, *sim_plots_folder*)

The third pipeline step: report generation.

Parameters

- **model** – the model name
- **outputdir** – the output directory to store the report
- **sim_plots_folder** – the folder containing the plots

Returns True if the task was completed successfully, False otherwise.

read_config(*lines*)

run(*config_file*)

Module contents

Submodules

sbpipe.pl.pipeline module

class sbpipe.pl.pipeline.**Pipeline**(*data_folder='Data'*, *models_folder='Models'*, *working_folder='Working_Folder'*, *sim_data_folder='sim_data'*, *sim_plots_folder='sim_plots'*)

Generic pipeline.

Parameters

- **data_folder** – the folder containing the experimental (wet) data sets
- **models_folder** – the folder containing the models
- **working_folder** – the folder to store the results
- **sim_data_folder** – the folder to store the simulation data
- **sim_plots_folder** – the folder to store the graphic results

config_parser(*config_file*, *section*)
Return the configuration for the parsed section in the config_file

Parameters

- **config_file** – the configuration file to parse
- **section** – the section in the configuration file to parse

Returns the configuration for the parsed section in the config_file

get_data_folder()
Return the folder containing the experimental (wet) data sets.

Returns the experimental data sets folder.

get_models_folder()
Return the folder containing the models.

Returns the models folder.

get_sim_data_folder()
Return the folder containing the in-silico generated data sets.

Returns the folder of the simulated data sets.

get_sim_plots_folder()
Return the folder containing the in-silico generated plots.

Returns the folder of the simulated plots.

classmethod get_simul_obj(*simulator*)
Return the simulator object if this exists. Otherwise throws an exception. The simulator name starts with an upper case letter. Each simulator is in a package within *sbpipe.simulator*.

Parameters **simulator** – the simulator name

Returns the simulator object.

get_working_folder()
Return the folder containing the results.

Returns the working folder.

classmethod read_common_config(*lines*)
Parse the common parameters from the configuration file

Parameters **lines** – the lines to parse.

Returns return a tuple containing the common parameters

read_config(*lines*)
Read the section lines from the configuration file. This method is abstract.

Returns a tuple containing the configuration

run(*config_file*)
Run the pipeline.

Parameters **config_file** – a configuration file for this pipeline.

Returns True if the pipeline was executed correctly, False otherwise.

Module contents

sbpipe.report package

Submodules

sbpipe.report.latex_reports module

sbpipe.report.latex_reports.**get_latex_header**(*pdftitle='SB pipe report'*, *title='SB pipe report'*, *abstract='Generic report.'*)

Initialize a Latex header with a title and an abstract.

Parameters

- **pdftitle** – the pdftitle for the LaTeX header
- **title** – the title for the LaTeX header
- **abstract** – the abstract for the LaTeX header

Returns the LaTeX header

sbpipe.report.latex_reports.**latex_report**(*outputdir*, *sim_plots_folder*, *model_noext*, *filename_prefix*, *caption=False*)

Generate a generic report.

Parameters

- **outputdir** – the output directory
- **sim_plots_folder** – the folder containing the simulated plots
- **model_noext** – the model name
- **filename_prefix** – the prefix for the LaTeX file
- **caption** – True if figure captions (=figure file name) should be added

sbpipe.report.latex_reports.**latex_report_dps**(*outputdir*, *sim_plots_folder*, *filename_prefix*, *model_noext*, *scanned_par1*, *scanned_par2*)

Generate a report for a double parameter scan task.

Parameters

- **outputdir** – the output directory
- **sim_plots_folder** – the folder containing the simulated plots
- **filename_prefix** – the prefix for the LaTeX file
- **model_noext** – the model name
- **scanned_par1** – the 1st scanned parameter
- **scanned_par2** – the 2nd scanned parameter

sbpipe.report.latex_reports.**latex_report_pe**(*outputdir*, *sim_plots_folder*, *model_noext*, *filename_prefix*)

Generate a report for a parameter estimation task.

Parameters

- **outputdir** – the output directory
- **sim_plots_folder** – the folder containing the simulated plots
- **model_noext** – the model name
- **filename_prefix** – the prefix for the LaTeX file

sbpipe.report.latex_reports.**latex_report_sim**(*outputdir*, *sim_plots_folder*, *model_noext*, *filename_prefix*)

Generate a report for a time course task.

Parameters

- **outputdir** – the output directory
- **sim_plots_folder** – the folder containing the simulated plots

- **model_noext** – the model name
- **filename_prefix** – the prefix for the LaTeX file

sbpipe.report.latex_reports.**latex_report_sps**(*outputdir*, *sim_plots_folder*, *filename_prefix*, *model_noext*, *scanned_par*)

Generate a report for a single parameter scan task.

Parameters

- **outputdir** – the output directory
- **sim_plots_folder** – the folder containing the simulated plots
- **filename_prefix** – the prefix for the LaTeX file
- **model_noext** – the model name
- **scanned_par** – the scanned parameter

sbpipe.report.latex_reports.**pdf_report**(*outputdir*, *filename*)

Generate a PDF report from LaTeX report using pdflatex.

Parameters

- **outputdir** – the output directory
- **filename** – the LaTeX file name

Module contents

sbpipe.simul package

Subpackages

sbpipe.simul.copasi package

Submodules

sbpipe.simul.copasi.copasi module

class sbpipe.simul.copasi.copasi.**Copasi**

Bases: *sbpipe.simul.simul.Simul* (page 27)

Copasi simulator.

pe(*model*, *inputdir*, *cluster_type*, *pp_cpus*, *nfits*, *outputdir*, *sim_data_dir*, *updated_models_dir*)

ps1(*model*, *scanned_par*, *sim_number*, *simulate_intervals*, *single_param_scan_intervals*, *inputdir*, *outputdir*)

ps2(*model*, *sim_length*, *inputdir*, *outputdir*)

sens(*model*, *inputdir*, *outputdir*)

sim(*model*, *inputdir*, *outputdir*, *cluster_type='pp'*, *pp_cpus=2*, *runs=1*)

sbpipe.simul.copasi.copasi_parser module

class sbpipe.simul.copasi.copasi_parser.**CopasiParser**

Retrieve information from a Copasi file.

classmethod get_param_estim_val(*file_in*)

Parse a Copasi file and retrieve information on the parameters to estimate.

Parameters **file_in** – the Copasi file including absolute path to parse

Returns a tuple containing the report file name, the parameter lower bounds, names, starting values,

and upper bounds

sbpipe.simul.copasi.copasi_utils module

sbpipe.simul.copasi.copasi_utils.**replace_str_copasi_sim_report**(*report*)

Replace a group of annotation strings from a generated copasi report file

Parameters **report** – The report file with absolute path

sbpipe.simul.copasi.randomise module

class sbpipe.simul.copasi.randomise.**Randomise**(*path*, *filename_in*)

This class generates multiple copies of a Copasi file configured for parameter estimation task, and randomises the starting values of the parameters to estimate.

Parameters

- **path** – the path to filename_in
- **filename_in** – the Copasi file to process.

get_copasi_obj()

Return the Copasi parser object

Returns the Copasi parser object

get_lower_bounds_list()

Return the list of parameter lower bounds

Returns the list of parameter lower bounds

get_param_names_list()

Return the list of parameter names

Returns the list of parameter names

get_path()

Return the path containing the template Copasi file

Returns the path to the Copasi file

get_report_filename()

Return the name of the template parameter estimation report

Returns the name of the report file name for parameter estimation

get_start_values_list()

Return the list of parameter starting values

Returns the list of parameter starting values

get_template_copasi_file()

Return the name of the template Copasi file

Returns the name of the Copasi file

get_upper_bounds_list()

Return the list of parameter upper bounds

Returns the list of parameter upper bounds

print_params_2_estim()

Print the parameter names, lower/upper bounds, and starting value, as extracted from COPASI template file

randomise (*num_files*, *idstr*)
Randomise the starting values for the parameter to estimate.

Parameters

- **num_files** – the number of files (instances) to generate
- **idstr** – an ID string to label the generated files (e.g. a timestamp)

replicate (*num_files*, *idstr*)
Generate num_files files and add an ID string to Copasi file/report names

Parameters

- **num_files** – the number of files (instances) to generate
- **idstr** – an ID string to label the generated files (e.g. a timestamp)

Module contents

Submodules

sbpipe.simul.simul module
class sbpipe.simul.simul.**Simul**
Generic simulator.

pe (*model*, *inputdir*, *cluster_type*, *pp_cpus*, *nfits*, *outputdir*, *sim_data_dir*, *updated_models_dir*)
parameter estimation.

Parameters

- **model** – the model to process
- **inputdir** – the directory containing the model
- **cluster_type** – pp for parallel python, lsf for load sharing facility, sge for sun grid engine
- **pp_cpus** – the number of cpu for parallel python
- **nfits** – the number of fits to perform
- **outputdir** – the directory to store the results
- **sim_data_dir** – the directory containing the simulation data sets
- **updated_models_dir** – the directory containing the models with updated parameters for each estimation

ps1 (*model*, *scanned_par*, *sim_number*, *simulate_intervals*, *single_param_scan_intervals*, *inputdir*, *outputdir*)
Single parameter scan.

Parameters

- **model** – the model to process
- **scanned_par** – the scanned parameter
- **sim_number** – the number of simulations (for det sim: 1, for stoch sim: n>1)
- **simulate_intervals** – the time step of each simulation
- **single_param_scan_intervals** – the number of scans to perform
- **inputdir** – the directory containing the model
- **outputdir** – the directory to store the results

ps2 (*model*, *sim_length*, *inputdir*, *outputdir*)
Double paramter scan.

Parameters

- **model** – the model to process
- **sim_length** – the length of the simulation
- **inputdir** – the directory containing the model
- **outputdir** – the directory to store the results

sens (*model*, *inputdir*, *outputdir*)
Sensitivity analysis.

Parameters

- **model** – the model to process
- **inputdir** – the directory containing the model
- **outputdir** – the directory to store the results

sim (*model*, *inputdir*, *outputdir*, *cluster_type='pp'*, *pp_cpus=2*, *runs=1*)
Time course simulator.

Parameters

- **model** – the model to process
- **inputdir** – the directory containing the model
- **outputdir** – the directory containing the output files
- **cluster_type** – pp for local Parallel Python, lsf for Load Sharing Facility, sge for Sun Grid Engine.
- **pp_cpus** – the number of CPU used by Parallel Python.
- **runs** – the number of model simulation

Module contents

sbpipe.utils package

Submodules

sbpipe.utils.io module

sbpipe.utils.io.**files_with_pattern_recur** (*folder*, *pattern*)
Return all files with a certain pattern in folder+subdirectories

Parameters

- **folder** – the folder to search for
- **pattern** – the string to search for

Returns the files containing the pattern.

sbpipe.utils.io.**get_pattern_pos** (*pattern*, *filename*)
Return the line number (as string) of the first occurrence of a pattern in filename

Parameters

- **pattern** – the pattern of the string to find
- **filename** – the file name containing the pattern to search

Returns the line number containing the pattern or "-1" if the pattern was not found

sbpipe.utils.io.**refresh**(*path*, *file_pattern*)
Clean and create the folder if this does not exist.

Parameters

- **path** – the path containing the files to remove
- **file_pattern** – the string pattern of the files to remove

sbpipe.utils.io.**replace_str_in_file**(*filename_out*, *old_string*, *new_string*)
Replace a string with another in filename_out

Parameters

- **filename_out** – the output file
- **old_string** – the old string that should be replaced
- **new_string** – the new string replacing old_string

sbpipe.utils.io.**write_mat_on_file**(*path*, *filename_out*, *data*)
Write the matrix results stored in data to filename_out

Parameters

- **path** – the path to filename_out
- **filename_out** – the output file
- **data** – the data to store in a file

sbpipe.utils.monitor module

class sbpipe.utils.monitor.**Monitor**
This is a monitor. It is a callback class for collecting information about finished processes. It is used by Parallel Python (pp).

add(*pid*, *value*)
The callback function

Parameters

- **pid** – this is callbackargs passed to parallel python *submit()* method
- **value** – the return value of the parallelised function. It is the callback value.

get_count()
Return the counter

Returns the number of running processes.

get_value()
Return the internal status.

Returns True if the counter is empty.

sbpipe.utils.parcomp module

sbpipe.utils.parcomp.**parcomp**(*cmd*, *cmd_iter_substr*, *cluster_type*, *runs*, *output_dir*, *pp_cpus=1*)
Generic funcion to run a command in parallel

Parameters

- **cmd** – the command string to run in parallel
- **cmd_iter_substr** – the substring of the iteration number. This will be replaced in a number automatically
- **cluster_type** – the cluster type among pp (multithreading), sge, or lsf
- **runs** – the number of runs

- **output_dir** – the output directory
- **pp_cpus** – the number of cpus that pp should use at most

sbpipe.utils.parcomp.**run_cmd_instance**(*cmd*)
Run a command using Python subprocess.

Parameters **cmd** – the string of the command to run

sbpipe.utils.parcomp.**run_command_pp**(*cmd*, *cmd_iter_substr*, *runs*, *server*, *monitor=<sbpipe.utils.monitor.Monitor instance>*)
Run instances of a command in multithreading using parallel python (pp).

Parameters

- **cmd** – the command string to run in parallel
- **cmd_iter_substr** – the substring of the iteration number. This will be replaced in a number automatically
- **runs** – the number of runs
- **server** – the server that pp should use
- **monitor** – the mutex object to count the jobs

sbpipe.utils.parcomp.**run_jobs_lsf**(*cmd*, *cmd_iter_substr*, *out_dir*, *err_dir*, *runs*)
Run jobs using a Load Sharing Facility (LSF) cluster.

Parameters

- **cmd** – the full command to run as a job
- **cmd_iter_substr** – the substring in command to be replaced with a number
- **out_dir** – the directory containing the standard output from bsub
- **err_dir** – the directory containing the standard error from bsub
- **runs** – the number of runs to execute

sbpipe.utils.parcomp.**run_jobs_pp**(*cmd*, *cmd_iter_substr*, *runs*, *pp_cpus=1*)
Run jobs using parallel python (pp) locally.

Parameters

- **cmd** – the full command to run as a job
- **cmd_iter_substr** – the substring in command to be replaced with a number
- **runs** – the number of runs to execute
- **pp_cpus** – The number of available cpus. If pp_cpus <=0, all the available cores will be used.

sbpipe.utils.parcomp.**run_jobs_sge**(*cmd*, *cmd_iter_substr*, *out_dir*, *err_dir*, *runs*)
Run jobs using a Sun Grid Engine (SGE) cluster.

Parameters

- **cmd** – the full command to run as a job
- **cmd_iter_substr** – the substring in command to be replaced with a number
- **out_dir** – the directory containing the standard output from qsub
- **err_dir** – the directory containing the standard error from qsub
- **runs** – the number of runs to execute

sbpipe.utils.rand module

sbpipe.utils.rand.**get_rand_alphanum_str**(*length*)
Return a random alphanumeric string

Parameters **length** – the length of the string

Returns the generated string

sbpipe.utils.rand.**get_rand_num_str**(*length*)
Return a random numeric string

Parameters **length** – the length of the string

Returns the generated string

sbpipe.utils.re_utils module

sbpipe.utils.re_utils.**nat_sort_key**(*str*)
The key to sort a list of strings alphanumerically (e.g. "file10" is correctly placed after "file2")

Parameters **str** – the string to sort alphanumerically in a list of strings

Returns the key to sort strings alphanumerically

Module contents

Submodules

sbpipe.__main__ module

sbpipe.main module

exception sbpipe.main.**Usage**(*msg*)
Bases: exceptions.Exception

This class is used for printing a generic exception

sbpipe.main.**check_args**(*args*, *msg*)
Check that at least one argument is passed.

Parameters

- **args** – the list of arguments
- **msg** – the message to print

Raise Usage exception if less than one argument is passed

Returns no output

sbpipe.main.**help**()
Return help message.

Returns the help message

sbpipe.main.**license**()
Return the license

Returns the license

sbpipe.main.**logo**()
Return sbpipe logo.

Returns sbpipe logo

sbpipe.main.**main**(*argv=None*)
SB pipe main function.

Parameters **argv** – options for sbpipe. Type python -m sbpipe -h for a full list of options.

Returns 0 if OK, 1 if minor problems, or 2 if serious trouble.

sbpipe.main.**version**()
Return the version

Returns the version

sbpipe.sb_config module

sbpipe.sb_config.**which**(*cmd_name*)
Utility equivalent to *which* in GNU/Linux OS.

Parameters **cmd_name** – a command name

Returns return the command name with absolute path if this exists, or None

Module contents

CHAPTER

FOUR

META INFORMATION

Copyright

Contacts: Dr Piero Dalle Pezze (piero.dallepezze AT babraham.ac.uk) and Dr Nicolas Le Novère (lenov AT babraham.ac.uk)

Affiliation: The Babraham Institute, Cambridge, CB22 3AT, UK

CHAPTER

FIVE

INDICES

- genindex
- modindex
- search

S

A

C

F

G

U

V

W