ANUGA Internal Tools Manual

Release 1.3.0-beta

•	A . II	1.00	A	KI OF T	The Control of the	
Geoscience	Australia	and the	Australian	National	University	V

Tuesday March, 2013, 5th minutes past in the afternoon

Copyright © 2004, 2005, 2006 Australian National University and Geoscience Australia. All rights reserved.

Permission to use, copy, modify, and distribute this software for any purpose without fee is hereby granted under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2 of the License, or (at your option) any later version, provided that this entire notice is included in all copies of any software which is or includes a copy or modification of this software and in all copies of the supporting documentation for such software.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License (http://www.gnu.org/copyleft/gpl.html) for more details.

You should have received a copy of the GNU General Public License along with this program; if not, write to the Free Software Foundation, Inc., 59 Temple Place, Suite 330, Boston, MA 02111-1307

This work was produced at Geoscience Australia and the Australian National University funded by the Commonwealth of Australia. Neither the Australian Government, the Australian National University, Geoscience Australia nor any of their employees, makes any warranty, express or implied, or assumes any liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately-owned rights. Reference herein to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the Australian Government, Geoscience Australia or the Australian National University. The views and opinions of authors expressed herein do not necessarily state or reflect those of the Australian Government, Geoscience Australia or the Australian National University, and shall not be used for advertising or product endorsement purposes.

This document does not convey a warranty, express or implied, of merchantability or fitness for a particular purpose.

ANUGA

Manual typeset with LATEX

Credits:

• ANUGA was developed and is maintained by Stephen Roberts, Ole Nielsen, Duncan Gray and Jane Sexton.

License:

• ANUGA is freely available and distributed under the terms of the GNU General Public Licence.

CONTENTS

1	Inter	rnal Tools 1									
	1.1	Introd	uction .								
		1.1.1	accepta	nce_tests							
			1.1.1.1	Using acceptance_tests							
		1.1.2	cmpsw	w							
			1.1.2.1	Using cmpsww							
			1.1.2.2	Bugs							
		1.1.3	event_s	election							
			1.1.3.1	Using event_selection							
		1.1.4	Installi	ng event_selection							
			1.1.4.1	Requirements							
			1.1.4.2	Bugs							
		1.1.5	mk_dig	est							
			1.1.5.1	Using mk_digest							
			1.1.5.2	Installing mk_digest							
		1.1.6	plotcsv	12							
			1.1.6.1	Using plotcsv							
			1.1.6.2	Installing plotcsv							
			1.1.6.3	Building plotesy for Windows							
			1.1.6.4	Bugs							
		1.1.7	tar_file								
			1.1.7.1	Using tar_file							
			1.1.7.2	Using untar_file							
			1.1.7.3	Installing tar_file							
		1.1.8	update_	DVD_images							
			1.1.8.1	Using update_DVD_images							
			1.1.8.2	Configuration							
			1.1.8.3	extra_files							
		1.1.9	update_	lic_checksum							
			1.1.9.1	Using update_lic_checksum							
			1.1.9.2	Using create_lic_file							
		1.1.10	write_la	arge_files							
			1.1.10.1	Using write_large_files							
			1.1.10.2	Installing write_large_files							
			1.1.10.3	Bugs							
Inc	dex			29							

CHAPTER

ONE

Internal Tools

1.1 Introduction

This document describes the tools written for internal ANUGA use at Geoscience Australia.

These tools are necessarily ad-hoc in nature and of possibly limited general use. If a tool becomes useful to a wider audience it may be moved into the 'ANUGA Tools Manual'.

The tools documented below are:

- acceptance_tests
- cmpsww
- event_selection
- mk_digest
- plotcsv
- tar_file
- update_DVD_images
- write_large_files

1.1.1 acceptance_tests

This collection of tests is designed to speed up and automate acceptance testing of a 'cluster' of compute servers. The tests are highly dependent on the installed software environment, so may have limited use outside Geoscience Australia, though the system design does lend itself to change.

The suite of tests checks:

- installed software, such as python installed packages
- availability of NFS mounted filesystems
- ability to ssh to each compute node from the master
- various aspects of parallel computation

The tests are a collection of self-contained acceptance 'testlets' that will be usually run from a controlling master program, but may be run individually. This is very useful when developing a new test, as it can be run by itself until correct.

1.1.1.1 Using acceptance_tests

The acceptance tests are designed to be run from the cluster 'master node', so you must ssh to that machine. It is assumed the acceptance tests code suite itself has been installed on the node it is being run from and other required code has been installed on all nodes.

Before running the acceptance tests you must prepare some environment variables:

PYTHON Defines the path to the python executable to use for the sub-tests.

PYTHONPATH The path to the **ANUGA** source directory.

EQRMPATH The path to the EQRM source directory. If not set, EQRM is not tested.

The first sub-test run dumps the testing environment to the screen as a check.

To run the acceptance tests, do the following:

```
export PYTHON=python2.5 \# we want to run python 2.5 in the tests export PYTHONPATH=/home/r-w/sandpit/ga/anuga_core/source/ \# EQRMPATH not set python test_all.py
```

While the tests are running, you will see the results of each test listed to the screen. Don't worry about catching this output; everything is written to a log file anuga.log.

1.1.2 cmpsww

The cmpsww program is used to compare two SWW files for some approximation of *equality*. The user must be able to define what to compare in the two files, as well as set tolerances for 'how close is close'.

1.1.2.1 Using cmpsww

The usage printed by the program is:

```
Usage: cmpsww.py <options> <file1> <file2>
where <options> is zero or more of:
                   -h
                            print this help
                   -a <val> set absolute threshold of 'equivalent'
                   -r <val> set relative threshold of 'equivalent'
                   -g <arg> check only global attributes specified
                             <arg> has the form <globname>[, <globname>[,...]]
                   -t <arg> check only timesteps specified
                             <arg> has the form <starttime>[,<stoptime>[,<step>]]
                   -v <arg> check only the named variables
                             <arg> has the form <varname>[, <varname>[,...]]
and <file1> and <file2> are two SWW files to compare.
The program exit status is one of:
     the two files are equivalent
  else the files are not equivalent.
```

Note that if no globals, variable or timesteps are specified, the program checks all globals and all variables for all timesteps.

1.1.2.2 Bugs

The cmpsww program is still being developed and needs to change in concert with the methodology of determining if an SWW file is as expected.

1.1.3 event_selection

event_selection is a graphical program used to select earthquake events.

It designed to run under both Windows and Linux.

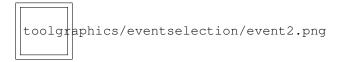
1.1.3.1 Using event_selection

Once you start the event_selection program you will see:



Before using the program, you need to set the *output base directory* field at the top of the window. The program needs to write some data files and this field tells the program where to write them. Just click in the box to select a directory somewhere in your filesystem.

We set the directory to C:\temp. Next, you need to select the *Region* from the drop-down list:



At this point you have set the values that will probably never change. If you close the program at this point, then the two values set (base directory and Region) and the three fields below (Hazard index, Minimum wave height and Maximum wave height) will be remembered and restored the next time you run the program. This data is stored in a file event_selection.cfg in the event_selection install directory.

Now you need to enter data specific to a particular event you are going to model. Fill in the *Hazard index* (location in the database of the point where the hazard is measured), *Minimum wave height* and *Maximum wave height* values and click on the *List* button:

toolgraphics/eventselection/event3.png

The program has filled the text box below the *List* button with events that satisfy your listed requirements. You need to select one of these events, which puts the *Quake_JD* number into the *Quake_JD* textbox below:

toolgraphics/eventselection/event4.png

Now you can click on either the *Multimux* or *Grid* buttons. Clicking on the *Multimux* button gives us:



If you now look in the output directory C:\temp you will see that two directories have been created:

```
10959
Results_Australia_1009_0.50_1.00
```

The Results_Australia_1009_0.50_1.00 directory contains the fault.xy and quake_prob.txt files used during the calculation of the multimux results. The Results directory name contains the region name, hazard index and minimum and maximum wave heights in an encoded form.

The 10959 directory contains the multimux data for the selected Quake_ID in a file called event_list.

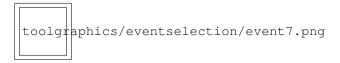
The Grid button was installed to allow the selection of seafloor deformation grid data. Clicking on this button shows:



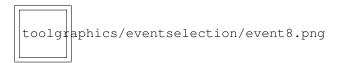
and writes some extra files into the Results_Australia_1009_0.50_1.00 directory:

event_010959.list
faults_010959.params

The event_010959.list file contains:



The faults_010959.params file contains:



1.1.4 Installing event_selection

There is an installer program used to install event_selection on a Windows machine (usually found at georisk\downloads\event_selection). The installer is generated by moving into the installer directory and right-clicking on the EventSelection.nsi file and selecting Compile NSIS script. You must have installed the NSIS package for this to work. Get it from http://nsis.sourceforge.net/Main_Page.

Once you have installed event_selection on your Windows machine you will have a desktop icon and Start menu entry to start the program with.

Under Linux just execute the event_selection.py program, either from the console or from a desktop icon or menu entry you created.

1.1.4.1 Requirements

Various pieces of python software must be installed before event_selection can be used. These are:

- wxpython a python package
- NSIS a Windows installer generator (required if creating a Windows installer)

1.1.4.2 Bugs

The look of event_selection under Linux is wrong - it needs to be rewritten using sizers for GUI layout.

1.1.5 mk_digest

mk_digest.py is a small program used to create an MD5 digest of a file. The digest string is written into a file. This program is used in the Patong Beach validation file refresh process.

1.1.5.1 Using mk_digest

1.1.5.2 Installing mk_digest

Installation is not required, just run the program.

1.1.6 plotcsv

plotesv is a GUI program to quickly plot selected columns of one or more CSV files onto a graph screen. Once the desired graph is plotted you may save the plot as a picture file.

The program is designed to run under both Windows and Linux.

The CSV files used *must* have column header information as the first line as the column header values are used during the plotting process.

1.1.6.1 Using plotcsv

Start the program by selecting it from the Start menu or double-clicking on the desktop icon. You will see the following window:



The first thing you must do is select one or more CSV files to plot. The files you are going to plot are listed in the textbox at the top of the screen. There is nothing there because this is the first time you have run plotcsv. Note that plotcsv will remember the selected files, as well as other information, when you next start the program. This files is plotcsv.cfg and it is stored in the plotcsv install directory.

This screen shot shows what happens when you click on the Add button - you get a file selector that lets you choose the CSV files to plot.

toolgraphics/plotcsv/2.png

In this example we selected both test.csv and test2.csv.

toolgraphics/plotcsv/3.png

You must now set the column data to display the X and Y axis of your plot. The *X-Column* and *Y-Column* listboxes are used to set which column data to use. In this example we are going to plot Stage versus Time, so we select the appropriate columns below:

toolgraphics/plotcsv/5.png

Note that choosing a column to plot also sets the text in the *X-Label* and *Y-Label* textboxes. You can change this text and, in this example, we want to change the stage axis text to *Stage (meters)*. We also add some title text and turn on the graph legend:



Finally, once we are ready, we click on the *Plot* button and see our plot:



You are free to configure the plot, make it larger, save a picture file, etc. Closing the plot window shuts down the application (see Bugs section below).

1.1.6.2 Installing plotcsv

For Windows execute the plotcsv_X.X.exe file in N:\georisk\downloads\plotcsv. This will install plotcsv into your C:\Program Files directory and create a desktop icon.

Linux needs no installation, just run the program.

1.1.6.3 Building plotcsv for Windows

The source directory for plotcsv contains an installer directory. Just right-click on the plotcsv.nsi file and select "Compile NSIS Script". You must ihave the NSIS installer installed, of course. Get it from http://nsis.sourceforge.net/Main_Page.

1.1.6.4 Bugs

The mixture of matplotlib and wxpython isn't successful - you only get one plot and then you must close the application. Using the wx_mpl_bars.py example from http://eli.thegreenplace.net/2008/08/01/matplotlib-with-wxpython-guis/, rewrite plotcsv to have the parameter changes (such as title text) show up immediately in the current plot.

The look of plotcsv under Linux is wrong – it needs to be rewritten using sizers for GUI layout.

1.1.7 tar_file

The tar_file.py program is used to tar and compress a file or directory into a *.tgz file. We have a python function to do this as we can't use a local *tar* program, as this wouldn't work under Windows.

The associated untar_file.py program reverses the above process.

These two programs are used in the Patong Beach validation suite.

1.1.7.1 Using tar_file

```
tar_file.py <tarfile> <file1> [<file2>, ...]
```

where tarfile is the path to the tar file to create, and file? is the path to a file or directory to include.

1.1.7.2 Using untar_file

```
untar_file.py <tarfile> [<output_directory>]
```

where *tarfile* is the path to the file to untar, and *output_directory* is the directory to write the results into.

If output_directory is not specified then the compressed file is unpacked into the current directory.

1.1.7.3 Installing tar_file

No installation is required, just run the program.

1.1.8 update_DVD_images

update_DVD_images is a program used to create the DVD image filesystems that were burnt to DVD for the 2009 East Coast Tsunami Inundation study.

1.1.8.1 Using update_DVD_images

To use the update_DVD_images program, just execute the program:

```
python update_DVD_images.py <name of jurisdiction>
```

Currently, the jurisdiction names are:

- BatemansBay
- GoldCoast
- Gosford
- Hobart

So to recreate the GoldCoast DVD image sub_directory, do:

```
python update_DVD_images.py goldcoast
```

Note that the case of the jurisdiction name doesn't matter.

The program will create a new sub-directory with the *formal* jurisdiction name (see below) in the current directory. The old jurisdiction sub-directory is deleted first.

1.1.8.2 Configuration

Here we discuss how to configure update_DVD_images to handle a new jurisdiction or change what files/directories are copied.

In update_DVD_images.py there are a set of dictionaries that control what is done for each jurisdiction.

The first dictionary is source_jurisdiction_path which maps the lowercase jurisdiction name to the dictionary defining that particular jurisdiction:

If you create a new jurisdiction, you need to add another line to the above dictionary.

In the case of the GoldCoast jurisdiction, we see that the dictionary for the GoldCoast is gold_coast_data:

```
gold_coast_data = \
{'jurisdiction': 'GoldCoast',
                                         # jurisdiction name
 # paths to various source directories
 'data_src_path': 'data/queensland/gold_coast_tsunami_scenario_2009/anuqa',
 'arcqis_src_path': 'data/queensland/gold_coast_tsunami_scenario_2009/ArcGIS',
 'proj_src_path': 'sandpits/lfountain/anuga_work/production/gold_coast_2009',
 # paths to destination directories (under 'jurisdiction' root)
 'data_dst_path': 'anuga',
 'proj_dst_path':
                    'project',
 'arcgis_dst_path': 'ArcGIS',
 # copy or create whole directories
 'make_dst_dirs': ['outputs'],
 'copy_data_dirs': ['boundaries'],
 # copy 'data' files or directories
 'copy_data_files': ['outputs/Event1_HAT', 'outputs/Event1_MSL',
                     'outputs/Event2_HAT', 'outputs/Event2_MSL',
'outputs/Event3_HAT', 'outputs/Event3_MSL'
                    1,
 # copy 'project' files or directories
 'copy_proj_files': ['build_elevation.py', 'export_results_max.py',
                      'get_runup.py', 'project.py', 'run_model.py',
                      'setup_model.py', 'build_urs_boundary.py',
                     'combine_gauges.py', 'get_timeseries.py',
                     'run_multiple_events.py'
 # copy 'arcgis' files or directories
 'copy_arc_files': ['MainBeach.mxd', 'SurfersParadise.mxd', 'GoldCoast.mxd',
                     'PalmBeach.mxd', 'Collangatta.mxd'
}
```

The first key is jurisdiction, which maps to a string defining the jurisdiction formal name. This name is used to create the output DVD staging directory.

```
'jurisdiction': 'GoldCoast',
```

The next three key values define the complete paths to source directories in the production filesystem:

```
# paths to various source directories
'data_src_path': 'data/queensland/gold_coast_tsunami_scenario_2009/anuga',
'arcgis_src_path': 'data/queensland/gold_coast_tsunami_scenario_2009/ArcGIS',
'proj_src_path': 'sandpits/lfountain/anuga_work/production/gold_coast_2009',
```

These key values are used along with a master path variable defined earlier in update_DVD_images.py to create the complete paths to source directories:

```
main_path = '/nas/gemd/georisk_models/inundation'
```

For example, the full path to the 'data' source directory would be:

```
data_src_path = os.path.join(main_path, j_dict['data_src_path'])
```

where j_dict would be a reference to the jurisdiction dictionary controlling the process ($gold_coast_data$ in this case).

The next three definitions define the names of output directories in the staging directory:

```
# paths to destination directories (under 'jurisdiction' root)
'data_dst_path': 'anuga',
'proj_dst_path': 'project',
'arcgis_dst_path': 'ArcGIS',
```

These three names are combined with the current directory and the jurisdiction staging directory name to produce the full path to output directories:

```
data_dst_path = os.path.join(os.getcwd(), j_name, j_dict['data_dst_path'])
proj_dst_path = os.path.join(os.getcwd(), j_name, j_dict['proj_dst_path'])
arcgis_dst_path = os.path.join(os.getcwd(), j_name, j_dict['arcgis_dst_path'])
```

Note that j_name is the jurisdiction name. So in this case, we would create the output directories:

```
./GoldCoast/anuga  # data directory
./GoldCoast/project  # project files directory
./GoldCoast/ArcGIS  # ArcGIS files
```

The next two key values define the names of empty directories to create or names of complete directories to copy to the data_dst_path directory:

```
# copy or create whole directories
'make_dst_dirs': ['outputs'],
'copy_data_dirs': ['boundaries'],
```

The values here are lists of one or more directories to create or copy. If there are no directories to create/copy, just use an empty list.

Next, we define which individual files we copy to the destination data directory:

Again we have a list of files to copy. Note that we must specify the path following the data_dst_path variable (anuga in this case), so we specify the directory under anuga and then the source file (or directory). Also note that we can copy a simple file or complete directory here.

You *must* create each target directory as an empty directory before copying files. That is why outputs appears in the make_dst_dirs key-value definition above.

Similarly, we now define 'project' files to copy:

These files (or directories) will be copied from the path defined in the proj_src_path variable to the path defined in the proj_dst_path variable.

Finally, we define 'arcgis' files or directories to copy:

These files (or directories) will be copied from the path defined in the arcgis_src_path variable to the path defined in the arcgis_dst_path variable.

1.1.8.3 extra files

In the same directory as update_DVD_images there must be a directory extra_files. This directory contains 'scaffolding' files that must exist on the DVD as well as jurisdiction-specific files that may be modifications of project files that replace those files on the DVD.

All files in the extra_files directory are copied to each jurisdiction DVD staging directory. All top-level directories that *aren't* named for a jurisdiction are also copied to each staging directory.

Each directory named for a jurisdiction will be copied to the staging directory if the directory has the same name as the jurisdiction staging directory we are creating. This jurisdiction directory would normally contain jurisdiction-specific scaffolding files, such as index.html, etc, as well as modified project files.

1.1.9 update_lic_checksum

The update_lic_checksum program is used to update all licence files (*.lic) in a filesystem sub_tree.

The create_lic_file program is used to create a licence file that controls one or more data files.

1.1.9.1 Using update_lic_checksum

The program is used:

```
update_lic_checksum.py [-m <lic_mask>] <directory>
```

where *directory* is the path to the sub_directory containing licence files to update. Normally, update_lic_-checksum would search for and update all *.lic files. If you want to update licence files that have a filename form of *.txt then use the -m *.txt option.

Note that the licence files being updated must contain well-formed XML data.

1.1.9.2 Using create_lic_file

 $\verb|create_lic_file| is a program used to create licence files from scratch. It is used so:$

```
usage: create_lic_file.py <options> <lic_file> [<filename> ...]
where <options> is zero or more of:
                  --author <name>
                  -w <name>
                                        - name of the author
                  --publishable [Yes|No]
                  -p [Yes|No]
                                        - is document publishable
                  --accountable <name>
                                        - name of person accountable for file
                  -a <name>
                  --source <string>
                  -s <string>
                                        - source of controlled file
                  --owner <name>
                                        - IP owner name
                  -o <name>
                  --info <string>
                  -i <string>
                                        - IP extra information
      <lic_file> is the name of the licence file to create.
      <filename> is one or more files to control.
```

If the file to be created (*lic_file*) already exists, the program aborts; it will not overwrite any existing file.

You must use the options to specify author name, etc. If these are not overridden the generated licence file will contain default values. For example, if you did this:

```
python create_lic_file.py test.lic README
```

then the output file test.lic would contain:

In particular, the *author* and *accountable* values are defaulted with the username from the environment.

Note the default values for these fields:

```
<publishable>Y</publishable>
<source>Generated by ANUGA development team</source>
<IP_owner>Geoscience Australia</IP_owner>
<IP_info>For use with the ANUGA test suite</IP_info>
```

1.1.10 write_large_files

This program is actually a suite of programs used to exercise the NetCDF file I/O code.

The NetCDF I/O model has three models of the way data is written:

- The 'classic' model
- The 'large' model
- The 'NetCDF4' model

The *classic* model is usually dismissed as the '2GiB limit' model, but this is an over-simplification. Chunks of data wriiten to a file in one write will contain an offset to the next related chunk 'upstream' in the file. This offset has a 2GiB limit, hence the '2Gib' oversimplification.

The *long* model relaxes some of the limits in the *classic* model.

The NetCDF4 model allows much larger datasets to be written to a file, along with compression, more than one unlimited dimension, etc.

Some effort is made to simulate the way an **ANUGA** program would write data. In particular, the variables written are interleaved in the way **ANUGA** would write them.

Also, each data value written to the file is a floating point number which encodes variable number, variable 'slice' and index into each slice. This is to ensure that each variable value written is unique and to allow for checking that what we read is what we wrote.

1.1.10.1 Using write_large_files

There are three programs in the write_large_files suite:

```
rwil_big_file.py writes using the classic model writes using the large model writes using the large model writes using the NetCDF4 model
```

Each of the three programs is used in the same way:

For instance, if we wanted to write a 3GiB 'large' file containing 6 variables we would do:

```
python rwil_big_file.py 512M 6
```

1.1.10.2 Installing write_large_files

No installation is necessary, just execute the programs.

1.1.10.3 Bugs

Instead of having three files, one to test each NetCDF model, just add an extra option to a single program:

- classic model (default)
- -1 large model -4 NetCDF4 model

INDEX

```
acceptance_tests, 2
ANUGA
     credits, i
     licence, i
cmpsww.py, 3
create_lic_file.py, 24
event_selection.py, 4
mk_digest.py, 11
plotcsv.py, 12
rwi4_big_file.py, 26
rwi_big_file.py, 26
rwil\_big\_file.py,\,26
tar_file, 19
tar_file.py, 19
untar\_file.py,\,19
update_DVD_images.py, 20
update_lic_checksum.py, 24
write_large_files, 26
```