climateprediction.net
the world's largest climate forecasting experiment for the 21st century

Environmental Change Institute · eci · University of Oxford · Oxford e-Research Centre

climate*prediction*.net and how to use it

Environmental Change Institute

Myles Allen · Friederike Otto · Karsten Haustein · Sihan Li

Volunteer Computing Group

David Wallom · Sarah Sparrow · Andy Bowery

Mamun Rashid · Sihan Li

Met Office

Richard Jones · Simon Wilson · William Ingram

Public Volunteers

Everywhere!

GOTHAM

Outline

Overview, infrastructure setup and workflow:

- Introduction to distributed computing
- Workflow
- Available models
- Monitoring progress
- Filename conventions

Submission:

- XMLS and Batches
- Ancillary Files and Namelists
- Diagnostic output or STASH

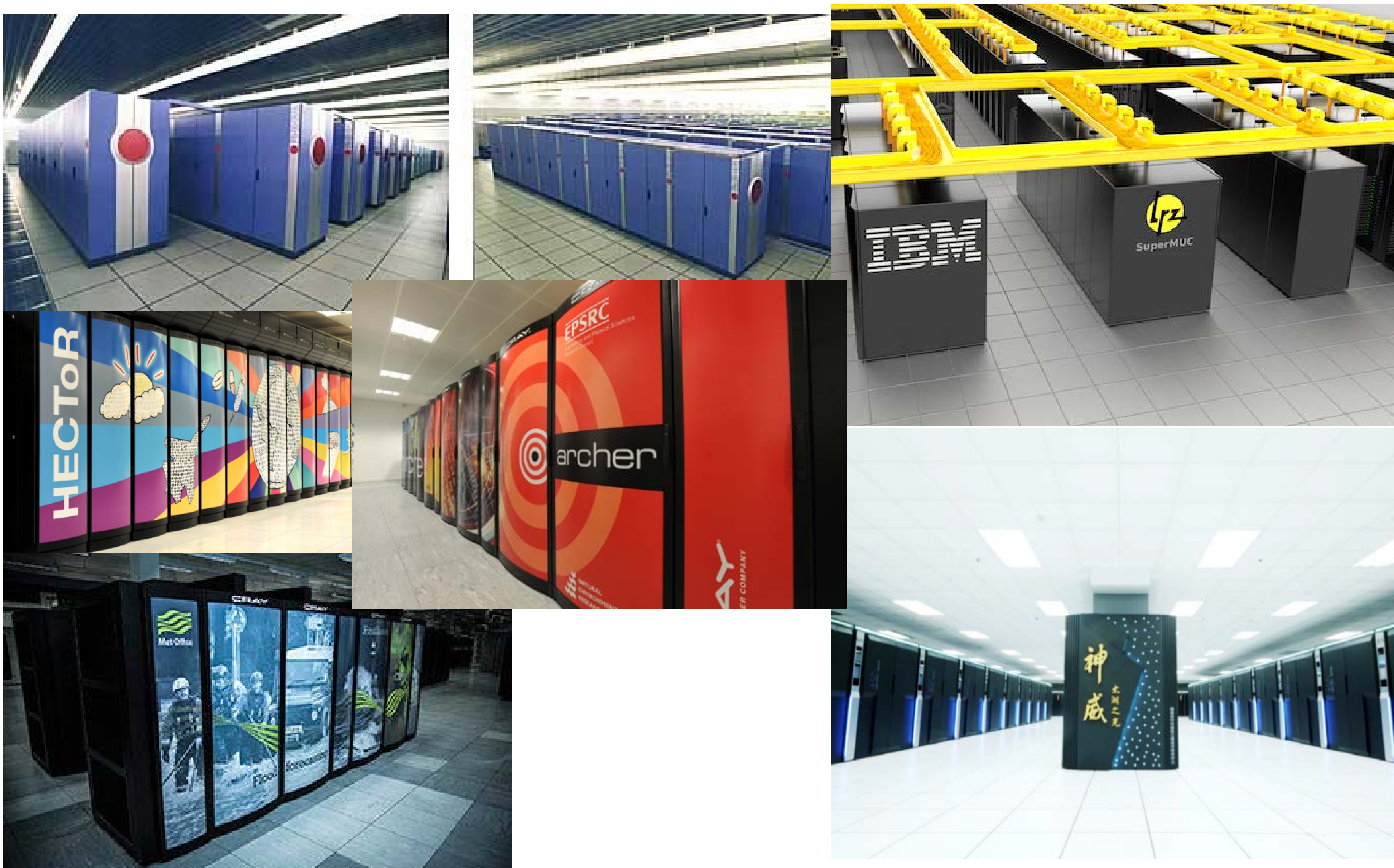
Result data handling:

- Data extraction
- Restart extraction
- Regridding regional data

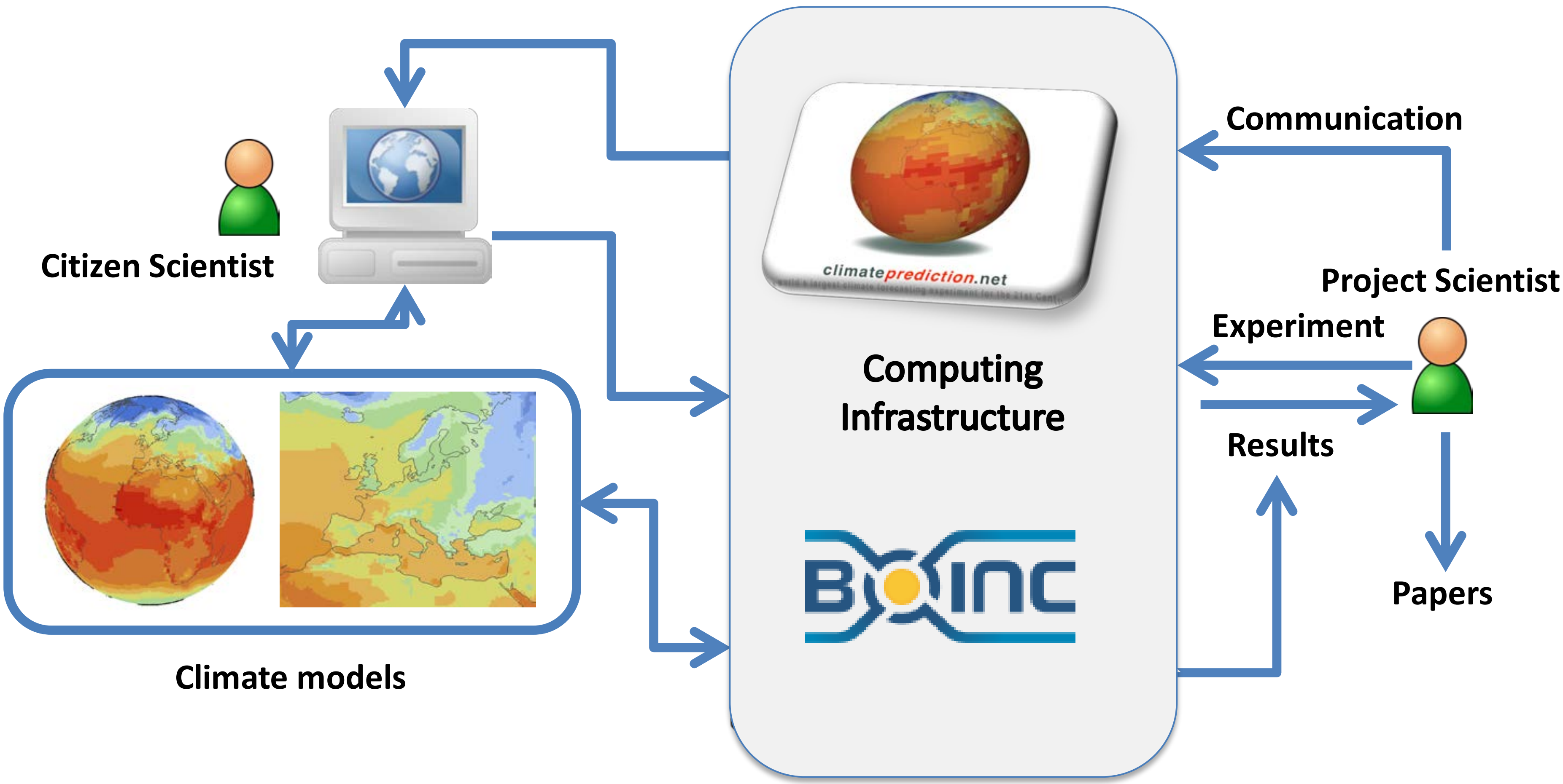
Publications

Practical Exercises

Computational challenge of climate science



Volunteer Distributed Computing



Very large ensembles of simulations can be generated by using this framework.

Models

All models are part of Hadley Centre HadCM3 family (currently)

Climate*prediction*.net

- HadCM3
  - Fully coupled free running atmosphere-ocean model

Weather@Home

- HadAM3P
  - N96 Global Atmosphere only model with prescribed SST and sea ice.
  - Mainly used as driver of regional model but capable of individual operation
- HadRM3P
  - Regional Climate Model with flexible user defined region of interest
  - Land surface scheme:

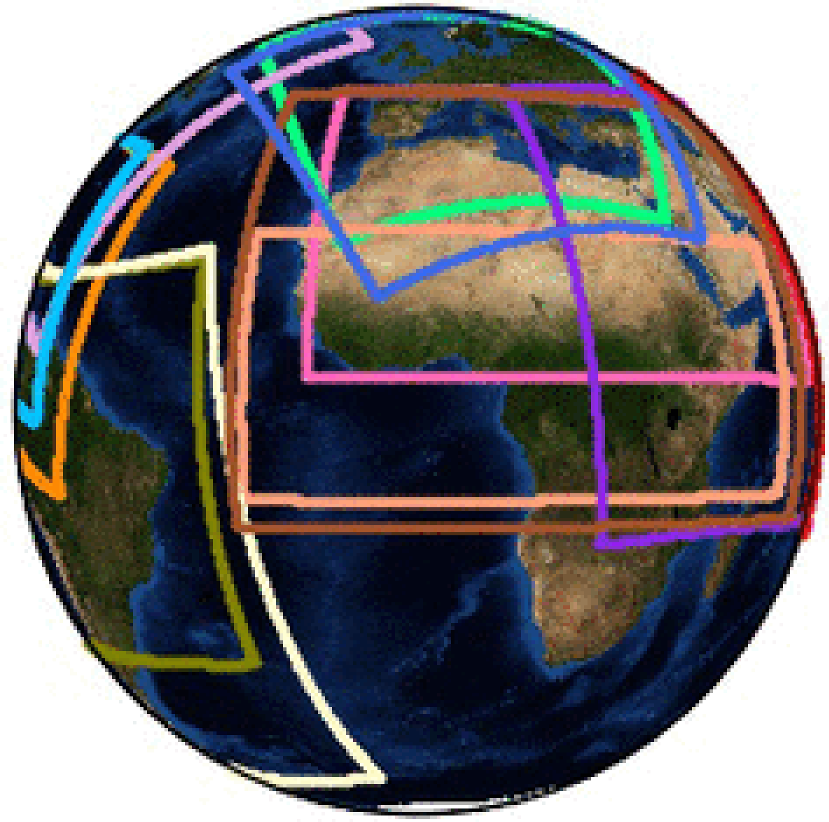
    MOSES1 in weather@home1

    MOSES2 in weather@home2
  - Optional vegetation model (TRIFFID) available in weather@home2 only.

Hadcm3: The coupled model

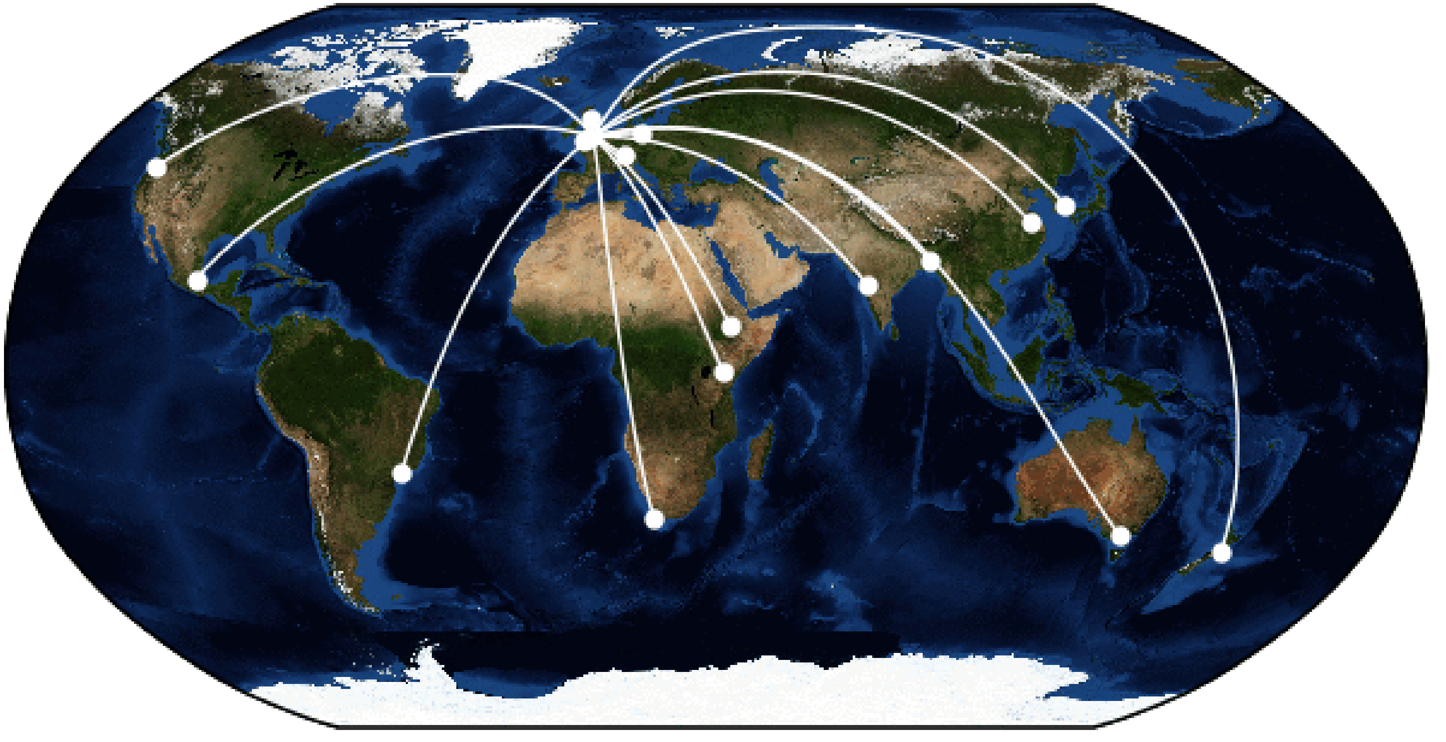
- Within CPDN referred to as **hadcm3s**.
- Capable of running with **monthly** or **yearly** upload frequency.
- Resolution: N48 L19 atmosphere, 1.25x1.25 L20 ocean.
- Sulphur and carbon cycle implemented.
- Has 73 latitude bands for volcanic emissions.
- Compatible with start dumps from the CPDN hadcm3n model (same base model with different post processing and only 4 volcanic latitude bands).
- Model can be started in any month and run for any number of integer months (participants prefer shorter runs)
- Start dumps can be output for any specified month of the run, **but** only one set of start dumps returned per simulation.
- Initial conditions perturbations (IC) are applied via a **potential temperature perturbation** to the atmosphere by changing the <dtheta> parameter.

Weather@home



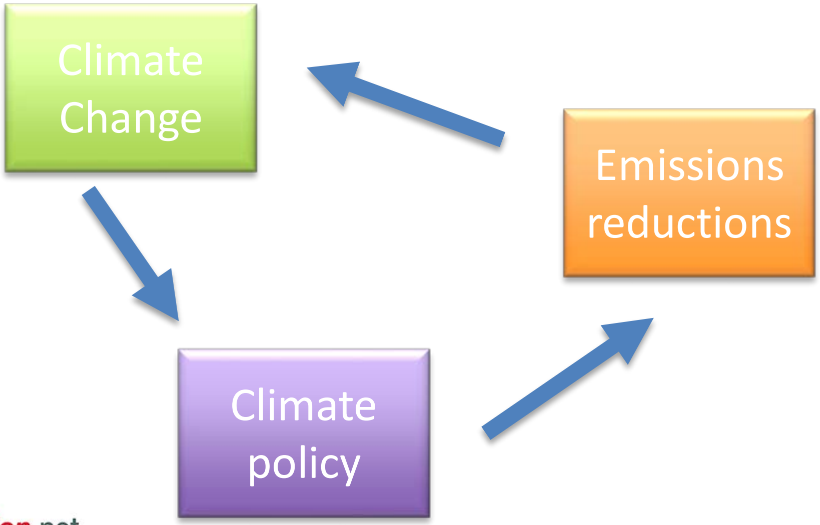
- New regions can easily be created.
- Global resolution N96, regional resolution typically **25 km** or **50 km**.
- Generic start dumps for new regions can be made by the computing team and can optionally include TRIFFID prognostic variables.
- Model can be started in any month and run for any number of integer months (participants prefer shorter runs)
- Start dumps can be output for any specified month of the run, **but** only one set of start dumps returned per simulation.
- Capable of running in **global only** mode.
- The dynamic vegetation model (TRIFFID) can be switched on or off in a region. However start dumps **must** be configured to include all necessary variables.
- Initial conditions perturbations (IC) are applied via a **potential temperature perturbation** to the atmosphere derived from a single long model run

CPDN Network



Adaptive Pathways to a 1.5ºC world (TCRE1.5)

How do we achieve the goal of the Paris Agreement to limit global warming to 1.5ºC?



- How fast should emissions be reduced to limit warming to 1.5ºC?
- How can incorporate new climate information into emission reduction plans?
- What range of 1.5ºC worlds might we expect?

climateprediction.net
the world's largest climate modelling experiment for the 21st century

UNIVERSITY OF EXETER

World Weather Attribution (WWA)

Hopes to provide real time answers about the role of human n

activit

waves for

use in

nature International weekly journal of science

Login Cart

Search go Advanced search

Journal home > Archive > Letters to Nature > Abstract

Journal content

- Journal home
- Advance online publication
- Current issue
- Nature News
- Archive
- Supplements
- Web focuses
- Podcasts
- Videos
- News Specials

Journal information

- About the journal
- For authors
- Online submission
- Nature Awards

Letters to Nature

Nature **432**, 610-614 (2 December 2004) | doi:10.1038/nature03089; Received 21 May 2004; Accepted 5 October 2004

There is a Corrigendum (25 August 2005) associated with this document.

Human contribution to the European heatwave of 2003

Peter A. Stott[1], D. A. Stone[2,3] & M. R. Allen[2]

1. Met Office, Hadley Centre for Climate Prediction and Research (Reading Unit), Meteorology Building, University of Reading, Reading RG6 6BB, UK
2. Department of Physics, University of Oxford, Oxford OX1 3PU, UK
3. Department of Zoology, University of Oxford, Oxford OX1 3PS, UK

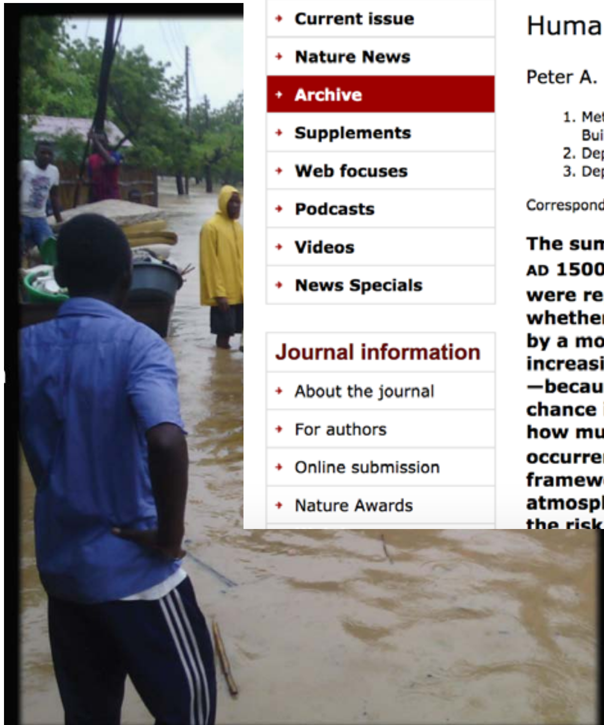
Correspondence to: Peter A. Stott[1] Email: peter.stott@metoffice.gov.uk

The summer of 2003 was probably the hottest in Europe since at latest AD 1500[1, 2, 3, 4], and unusually large numbers of heat-related deaths were reported in France, Germany and Italy[5]. It is an ill-posed question whether the 2003 heatwave was caused, in a simple deterministic sense, by a modification of the external influences on climate—for example, increasing concentrations of greenhouse gases in the atmosphere—because almost any such weather event might have occurred by chance in an unmodified climate. However, it is possible to estimate by how much human activities may have increased the risk of the occurrence of such a heatwave[6, 7, 8]. Here we use this conceptual framework to estimate the contribution of human-induced increases in atmospheric concentrations of greenhouse gases and other pollutants to the risk of the occurrence of unusually high mean summer temperatures

subscribe to nature

ABSTRACT

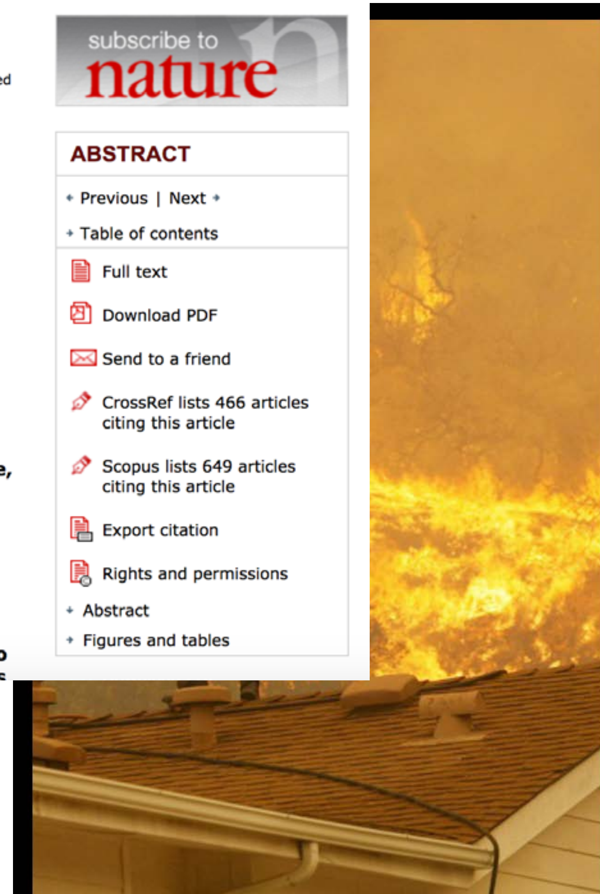
- Previous | Next
- Table of contents
- Full text
- Download PDF
- Send to a friend
- CrossRef lists 466 articles citing this article
- Scopus lists 649 articles citing this article
- Export citation
- Rights and permissions
- Abstract
- Figures and tables



Malawi flood, 2015

detectable role:

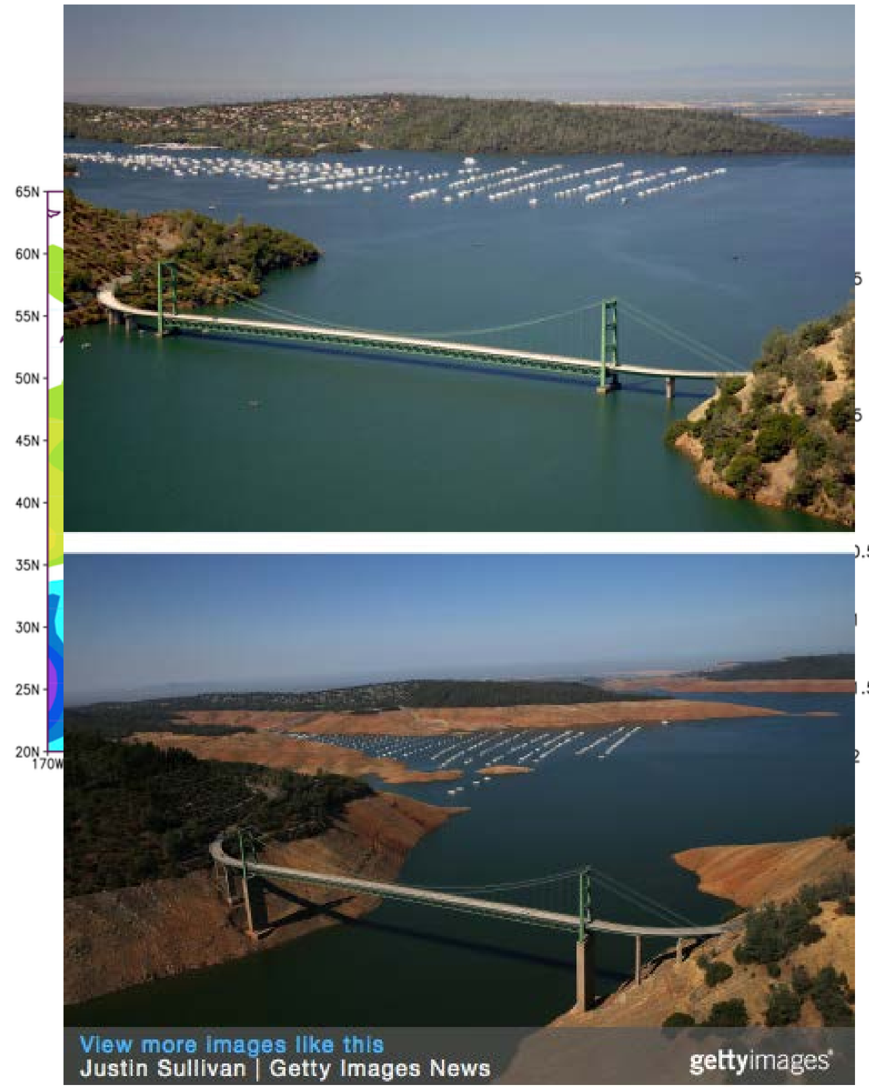
- Our analysis methods were unable to give information.



California wildfires, 2014

Californian Drought Experiment

- Investigate effect of climate change on the current drought in California
  - 5k current conditions including ‘the blob’
  - 5k current conditions with averaged SST
  - 12k natural runs
- Time relevant results



http://www.climateprediction.net/weatherathome/western-us-drought/

East Asia Summer Heatwave Attribution (EASHA)

W@H East Asia domain

Summer 2013 Heat Wave

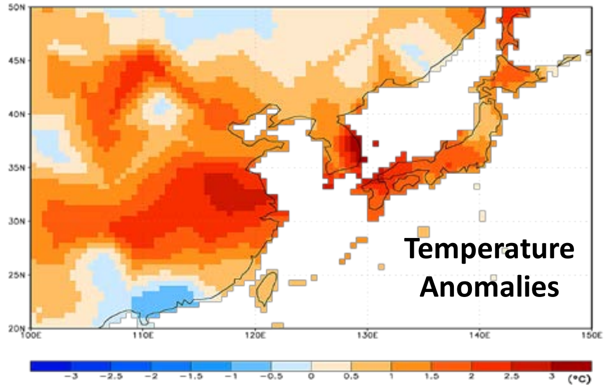


Image: CCRL, POSTECH (using CRU data)

Heat Wave Probability

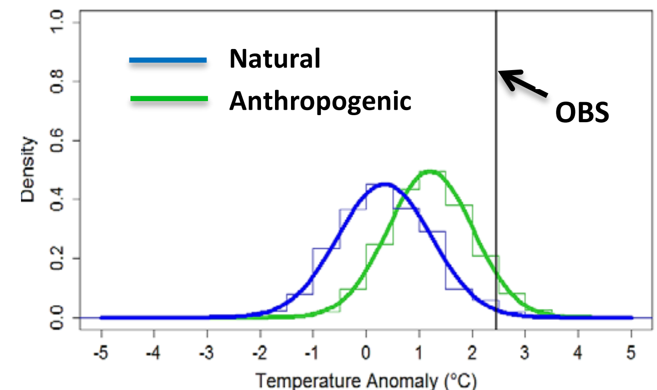


Image: CCRL, POSTECH

climateprediction.net
the world's largest climate modelling experiment for the 21st century

POSTECH CRL

Forest Mortality through Environmental Change (FMEC)

Image: Colorado State Forest Service

Image: Jamie Francis, The Oregonian

Climate Change

Bark Beetle Outbreak

Forest Dieback

Economic Impact

climateprediction.net
the world's largest climate modelling experiment for the 21st century

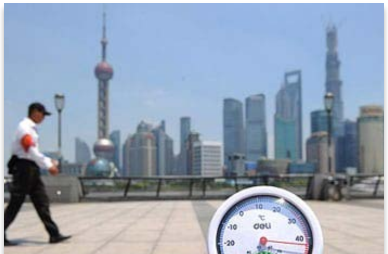
Met Office

UCSB
UNIVERSITY OF CALIFORNIA SANTA BARBARA

University of Idaho

Oregon State University

Long Term Undulations versus Secular Change in Chinese Climate (LOTUS)



Heatwaves

Droughts

Floods

Extreme Event Risk in China

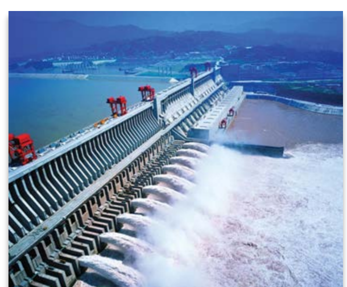
THE UNIVERSITY of EDINBURGH

University of Reading

Climate Services

Agriculture

Urban



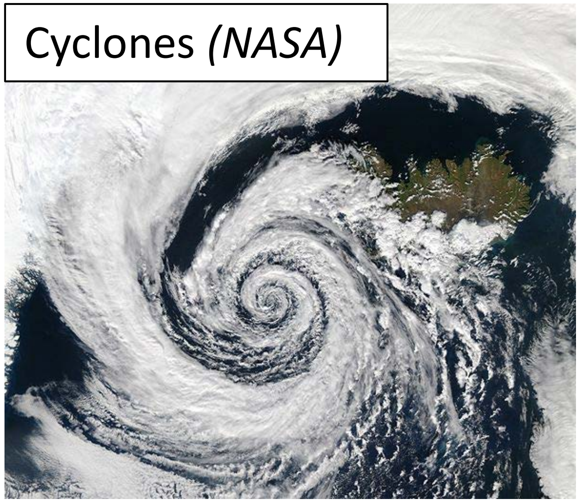
Water resources

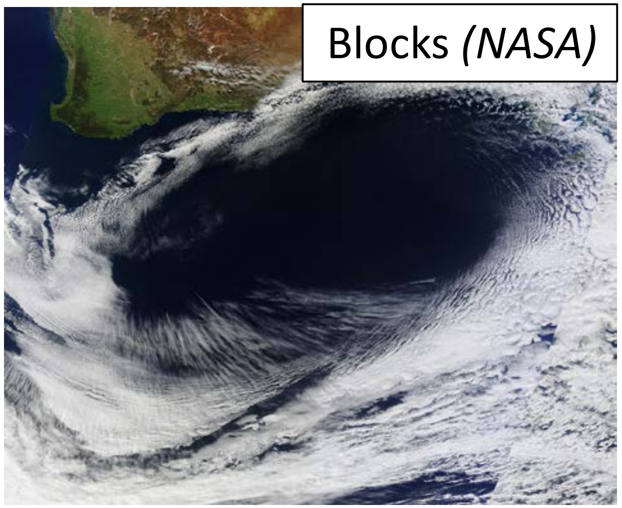
Energy

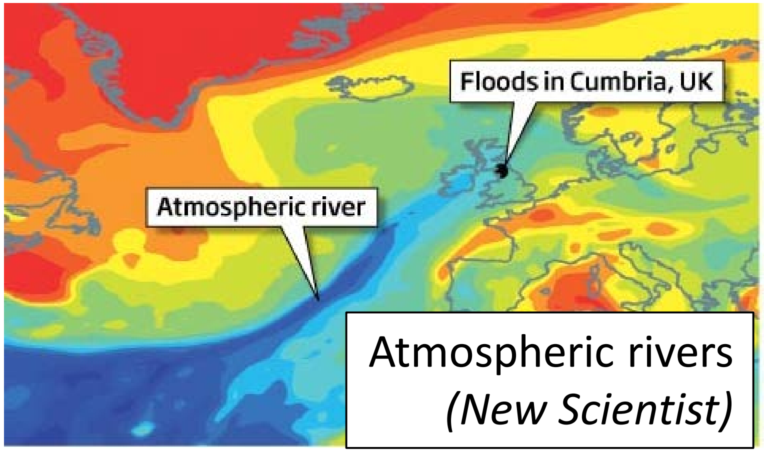
Air Quality

Drivers Of Change In mid-Latitude weather Events (DOCILE)

Advancing simulation of the weather systems behind the extremes, and how they will change under global warming:



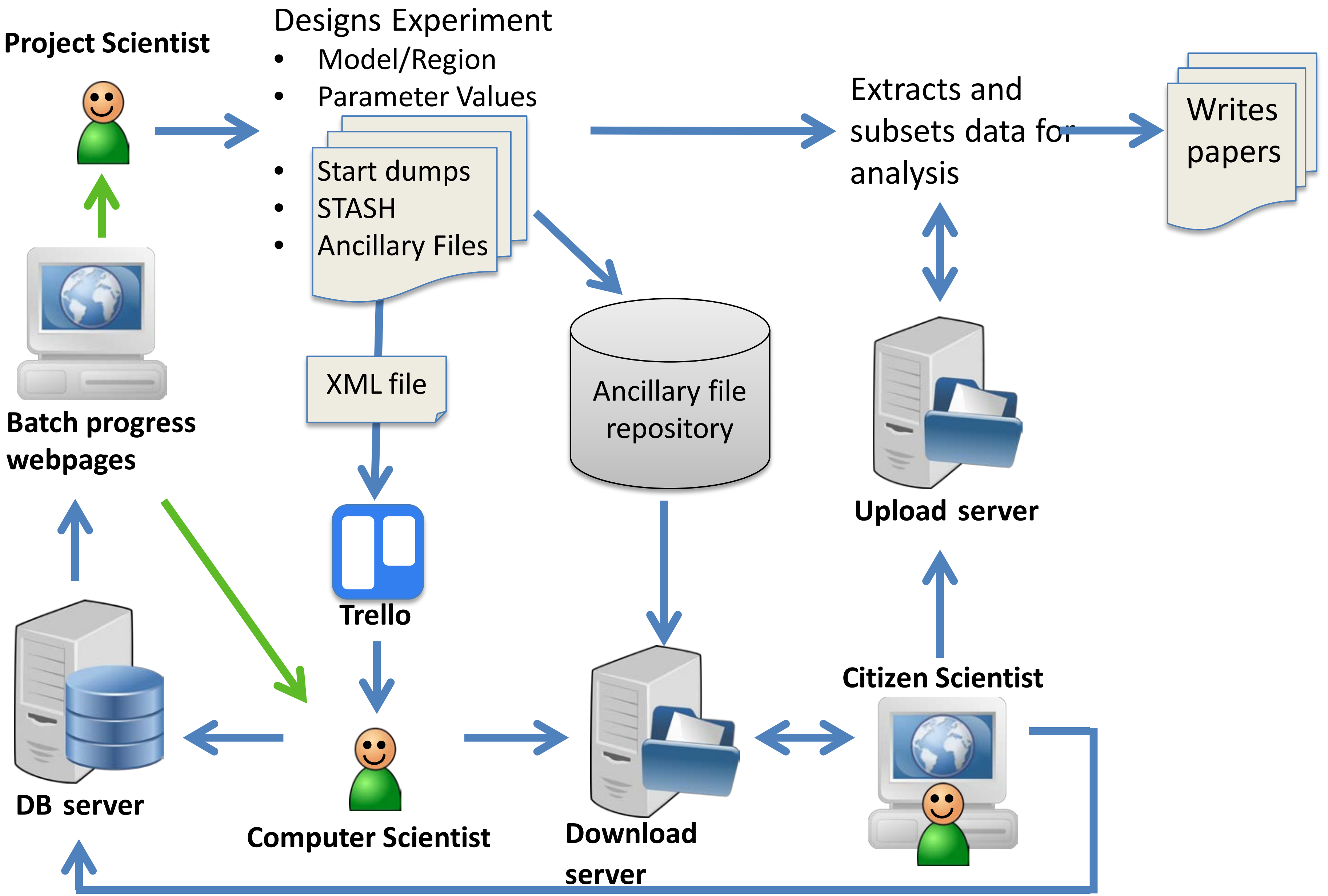




climateprediction.net
the world's largest climate modelling experiment for the 21st century

Supported by Natural Environment Research Council (NE/P002099/1)

Creating work for CPDN

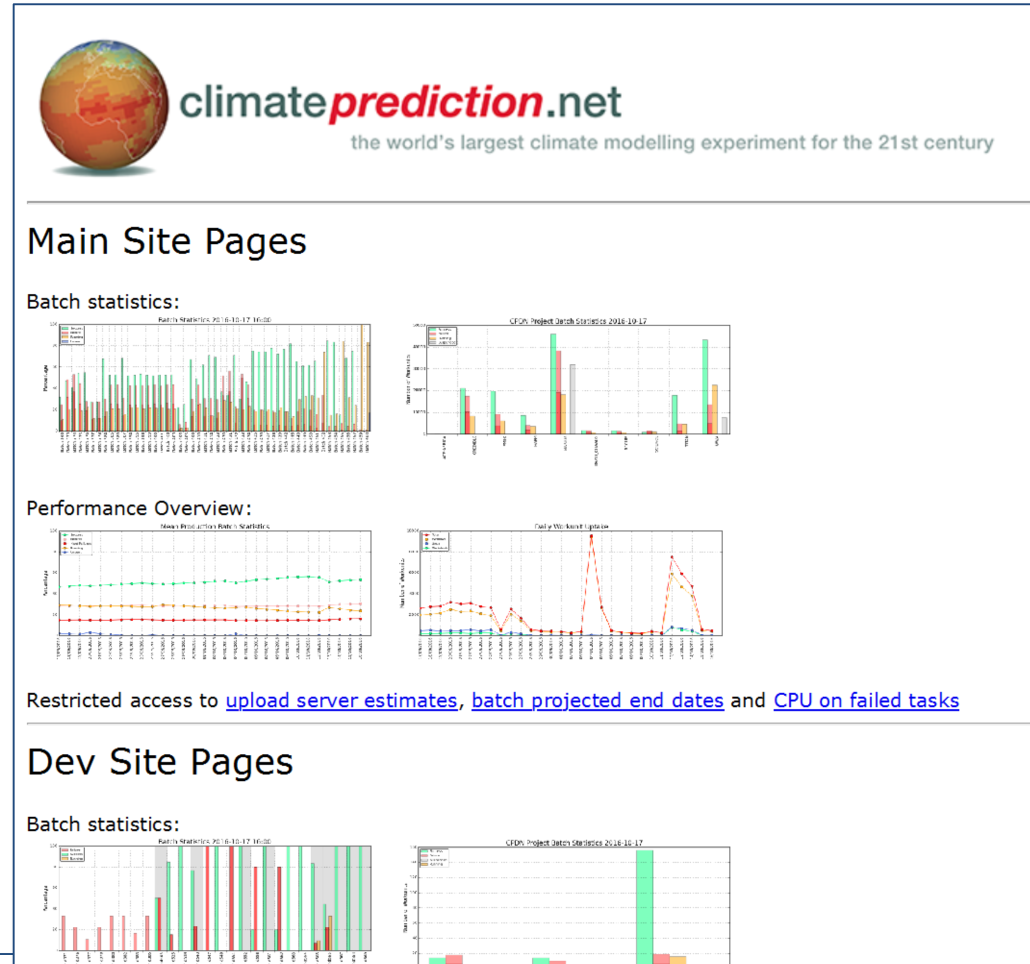


CPDN Batches

Workunits are sent out in batches

Batch progress can be monitored on the following web page for both the main and dev site batches (click on the relevant plot and drill down for more detail). Plots update 4 times daily and information live from database
http://climateapps2.oerc.ox.ac.uk/cpdnboinc/batch_analysis.php

Batch e-mails are sent daily with a wget list that can be used in conjunction with the extract scripts to pull down the desired data from a batch.



cpdn-servers@oerc.ox.ac.uk

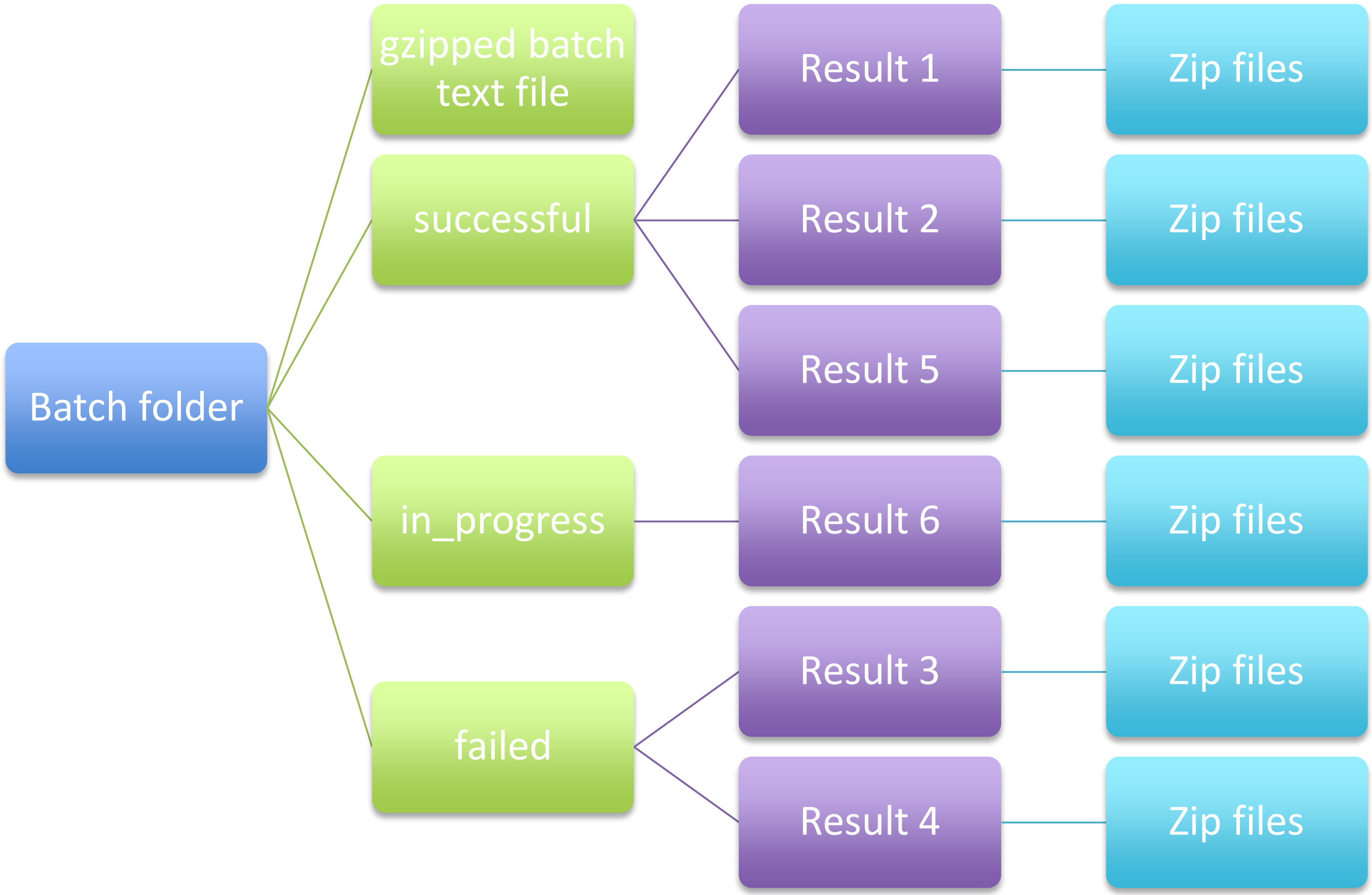
For any batch, changing the batch number appropriately:

Further batch information is available at:
http://climateapps2.oerc.ox.ac.uk/cpdnboinc/batch_info.php?batchid=627

You can also find the workunit submission xml for this batch at:
http://download.cpdn.org/download/batch_627/batch_627_workunit_submission.xml.gz

Result batch file structure

On the upload server results are sorted into batches and subcategorised into successful, running and failed, so the directory structure looks something like below.



Filename conventions

Workunits follow a naming convention as follows:
<app_name>_<umid>_<start_year><start_month>_<run_months>_<batch_number>_<workunit_id>

So for example:
wah2_eu25_2io4_198012_2_416_007302382
hadcm3s_7hpw_199012_24_412_009410187

Note for weather@home the "app_name" also includes the region being run.

Since workunits may fail for a variety of reasons that may be independent from an incorrect configuration, workunits are typically regenerated up to 3 times when they fail. Therefore results are appended with a "regeneration" number giving a result directory filename as:

<workunit_name_as_above>_<regeneration_number>

So for example:
wah2_eu25_2io4_198012_2_416_007302382_1
hadcm3s_7hpw_199012_24_412_009410187_2

Note workunits that have failed for the maximum regeneration number are said to have **"Hard Failed"** and **may** point to an incorrect setup of that particular workunit.

Filename conventions

The result directories on a given upload server contain a number of zip files. The numerically number zip files correspond to the month (or year if using hadcm3s yearly upload) that the results pertain to. In addition to the monthly result zip files there are two further files, namely the <result>_restart.zip and <result>_out.zip files.

The **restart.zip** files contain the restart files for the month specified in the xml that the restarts are to be uploaded for.

The **out.zip** contains the UM output file streams. These can be useful when failures have occurred to work out what has gone wrong with a particular workunit.

So for example within our example result directories for a 2 month run there will be:
wah2_eu25_2io4_198012_2_416_007302382_1_1.zip (results from month 1)
wah2_eu25_2io4_198012_2_416_007302382_1_2.zip (results from month 2)
wah2_eu25_2io4_198012_2_416_007302382_1_restart.zip (restart files for month specified)
wah2_eu25_2io4_198012_2_416_007302382_1_out.zip (UM output files)

And for a 2 year run of hadcm3s with yearly upload files there will be:
hadcm3s_7hpw_199012_24_412_009410187_2_1.zip (results from year 1)
hadcm3s_7hpw_199012_24_412_009410187_2_1.zip (results from year 2)
hadcm3s_7hpw_199012_24_412_009410187_2_restart.zip (restart files for **month** specified)
hadcm3s_7hpw_199012_24_412_009410187_2_out.zip (UM output files)

Outline

Overview, server setup and workflow:

- Introduction to distributed computing
- Workflow
- Available models
- Monitoring progress
- Filename conventions

Submission:

- XMLS and Batches
- Ancillary Files and Namelists
- Diagnostic output or STASH

Result data handling:

- Data extraction
- Restart extraction
- Regridding regional data

Publications

Practical Exercises

Weather@home region abbreviations

Region abbreviation	Description
afr50l	Africa 50km region
anz50	Australia/New Zealand 50km region
cafr25	Central Africa 25km region
cam25	Central America 25km region
cam50	Central America 50km region
eas50	East Asia 50km region
eu25	Europe 25km region
eu50r	Europe 50km region
nawa25	North Africa/West Asia 25km region
pnw25	Pacific North West 25km region
sam50	South America 50km region
sas50	South Asia 50km region
wus25	Western United States 25km region
sam25	South America 25km region
nam50	North America 50km region
global	Global simulation (i.e. no region running)

XML Header Information

The header part of the XML file defines a few critical details about the experiment. The wiki page http://climateprediction.net/wiki/doku.php?id=xml_header provides a checklist of details that should be verified.

Read only wiki access:
User: ITSS-USER1
Pass: apart-pigeon-key-5

The first of these is the model configuration file which should be chosen appropriately. Examples are given below. Note any weather@home region can be used:

Model Setup	Config file
Weather@home1 ANZ dyn template	config_hadam3p_anz_dyn.xml
Weather@home1 ANZ standard	config_hadam3p_anz.xml
Weather@home2 WUS with TRIFFID	config_wah2.2_wus25_triffid.xml
Weather@home2 WUS without TRIFFID	config_wah2.2_wus25.xml
Weather@home2 EU25	config_wah2.2_eu25.xml
Weather@home2 Global only	config_wah2.2_global.xml
Hadcm3s with monthly uploads	config_hadcm3s_v2.2_monthly.xml
Hadcm3s with yearly uploads	config_hadcm3s_v2.2_yearly.xml

XML Header Information

The STASH entry should be checked to ensure that it is correct.

The XML header also details where results should be uploaded to. Details of the available upload servers are as follows:

Upload Number	Upload Location	Template Directory
upload2	Oxford	upload_templates/oxford2.2/result_template_upload2
upload3	JASMIN	upload_templates/badc2.2/result_template_upload3
upload4	Hobart	upload_templates/anz2.2/result_template_upload4
upload5	Oregon	upload_templates/oregon2.2/result_template_upload5
upload6	Mexico	upload_templates/mexico2.2/result_template_upload6
upload7	Korea	upload_templates/korea2.2/result_template_upload7
upload8	India	upload_templates/india2.2/result_template_upload8

Note if using weather@home1 change omit the “2.2” from the location name.

OeRC staff will also double check the header information in xmls provided on submission.

Batches

To help with error tracking:

- Each new xml submission goes out as a new batch.

Batch tags to be entered into the XML:

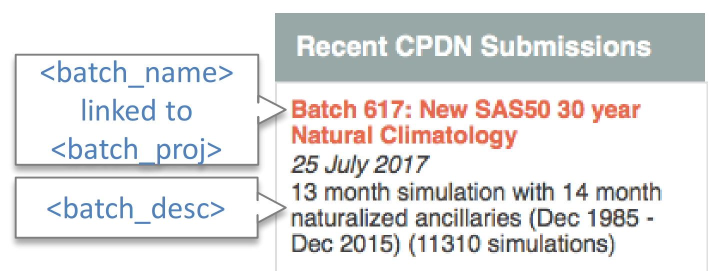
<batch_name>	Public title
<batch_desc>	Short public description
<batch_proj>	Science project associated with batch
<batch_owner>	Name of owner and e-mail address for batch e-mails. (More than one person can be listed here.)
<batch_tech_info>	Technical information associated with batch
<batch_first_year>	First start year for batch.
<batch_last_year>	Last start year for batch.
<batch_umid_start>	First umid in the batch
<batch_umid_end>	Last umid in the batch

RSS feed on the website:

- This will be taken direct from the XML from the following fields:

<batch_name>	Public title
<batch_desc>	Short public description
<batch_proj>	Science project associated with batch

SO PUT IN SOMETHING MEANINGFUL!



XML Header Information

```
<?xml version="1.0" ?>
<WorkGen>
    <!--App configuration setting wah1, wah2 and region along with dyn/triffid template
    <app_config>config_dir/config_wah2.2_sas50.xml</app_config>
    <!--Upload information-->
    <upload_handler>http://upload2.cpdn.org/cgi-bin/file_upload_handler</upload_handler>
    <result_template_prefix>upload_templates/oxford2.2/result_template_upload2</result_temp
    <!--Download information for participants to get workunits-->
    <download_url_base>http://download.cpdn.org/download/</download_url_base>
    <download_dir_base>/storage/download/</download_dir_base>
    <project_dir>/var/www/boinc/projects/cpdnboinc</project_dir>
    <namelist_dir>namelist_template_files/wah2.2</namelist_dir>
    <!--Stash to use in the simulations-->
    <global_stashc>xadae.sulph.stashc.global_jul16</global_stashc>
    <region_stashc>xacxf.sulph.stashc.region_jul16</region_stashc>
    <!--Batch information (to be updated for each batch)-->
    <batch_name>SAS50 revised SO2 12/2014-12/2015</batch_name>
    <batch_desc>South Asia experiments with updated SO2 emissions</batch_desc>
    <batch_tech_info>14 month ancillaries with 13 month simulation using OSTIA SST, cyclic sea ice and ECLIPSE SO2 emission
ancillaries. Restarts uploded after 12 month</batch_tech_info>
    <batch_proj>TESTING</batch_proj>
    <batch_owner>Sarah Sparrow &lt;sarah.sparrow@oerc.ox.ac.uk&gt;,Peter Uhe &lt;peter.uhe@ouce.ox.ac.uk&gt;</batch_owner>
    <batch_umid_start>iuk0</batch_umid_start>
    <batch_umid_end>iz13</batch_umid_end>
    <batch_first_start_year>2014</batch_first_start_year>
    <batch_last_start_year>2014</batch_last_start_year>
```

Application / Region details

Upload Information

Download Information

Global and Regional STASH

Batch Information (can also appear at the end of the XML)

XML Body Information

```
<experiment>
    <parameters>
        <file_atmos>xhjlya.start.0000.360.new</file_atmos>
        <file_region>xacxfa.start.0000</file_region>
        <file_sst>final_ancil_2year_OSTIA_sst_1997-12-01_1999-12-30</file_sst>
        <file_sice>final_ancil_2year_OSTIA_ice_1997-12-01_1999-12-30</file_sice>
        <file_so2dms>so2dms_hist_N96_1989_2000v2</file_so2dms>
        <file_sulphox>oxi.addfa</file_sulphox>
        <file_ozone>ozone_hist_N96_1989_2000v2</file_ozone
        <file_pert>ic00000000_10_N96</file_pe
        <file_solar>solar_1985_2020</file_solar>
        <file_volcanic>volc_cmip5</file_volcanic>
        <file_ghg>ghg_defaults</file_ghg>
        <run_years>0</run_years>
        <run_months>1</run_months>
        <exptid>a000</exptid>
        <model_start_year>1997</model_start_year>
    </parameters>
</experiment>
<experiment>
    <parameters>
        <file_atmos>xhjlya.start.0000.360.new</file_atmos>
        <file_region>xacxfa.start.0000</file_region>
        <file_sst>final_ancil_2year_OSTIA_sst_1997-12-01_1999-12-30</file_sst>
        <file_sice>final_ancil_2year_OSTIA_ice_1997-12-01_1999-12-30</file_sice>
        <file_so2dms>so2dms_hist_N96_1989_2000v2</file_so2dms>
        <file_sulphox>oxi.addfa</file_sulphox>
        <file_ozone>ozone_hist_N96_1989_2000v2</file_ozone>
        <file_pert>ic19611112_16_N96</file_pert>
        <file_solar>solar_1985_2020</file_solar>
        <file_volcanic>volc_cmip5</file_volcanic>
        <file_ghg>ghg_defaults</file_ghg>
        <run_years>0</run_years>
        <run_months>1</run_months>
        <exptid>a001</exptid>
        <model_start_year>1997</model_start_year>
    </parameters>
</experiment>
</WorkGen>
```

Restart Files

Ancillary Files

Initial Conditions Perturbation

Forcings via namelists

UMID, start year, start month, run length

Don’t Forget!

- Parameter perturbations can also be added in this section.
- Any parameter that has been setup in the model namelist template to take an input or default value can be varied.
- Ask OeRC staff if you want to do this – help is on hand.

Selecting the restart upload month

By default restart files will be uploaded for the **final** month that is simulated.
So for example:

Run starts in Dec 2015 and run for 12 months ➔ Restarts returned will be for 1st Dec 2016
Run starts in Mar 2016 and run for 3 months ➔ Restarts returned will be for 1st Jun 2016

Sometimes restart files want to be returned for a specified month rather than the end of the simulation. This can be defined by setting the <restart_upload_month> as a parameter in the body of the XML file. Its value should correspond to the month number from the start month that the restarts should be uploaded for. So for example:

```
<experiment>
        <parameters>
            <file_atmos>xhjlya.start.0000.360.new</file_atmos>
            <file_region>xacxfa.start.0000</file_region>
            <file_sst>final_ancil_2year_OSTIA_sst_1997-12-01_1999-12-30</file_sst>
            <file_sice>final_ancil_2year_OSTIA_ice_1997-12-01_1999-12-30</file_sice>
            <file_so2dms>so2dms_hist_N96_1989_2000v2</file_so2dms>
            <file_sulphox>oxi.addfa</file_sulphox>
            <file_ozone>ozone_hist_N96_1989_2000v2</file_ozone>
            <file_pert>ic00000000_10_N96</file_pert>
            <file_solar>solar_1985_2020</file_solar>
            <file_volcanic>volc_cmip5</file_volcanic>
            <file_ghg>ghg_defaults</file_ghg>
            <run_years>0</run_years>
            <run_months>4</run_months>
            <restart_upload_month>2</restart_upload_month>
            <exptid>a000</exptid>
            <model_start_year>1997</model_start_year>
        </parameters>
  </experiment>
```

Run starts in Dec 1997 and run for 4 months ➔ Restarts returned will be for 1st Feb 1998

Note: Generic restart files are **1st Dec** restarts valid for any year.

Outline

Overview, server setup and workflow:

- Introduction to distributed computing
- Workflow
- Available models
- Monitoring progress
- Filename conventions

Submission:

- XMLS and Batches
- Ancillary Files and Namelists
- Diagnostic output or STASH

Result data handling:

- Data extraction
- Restart extraction
- Regridding regional data

Publications

Practical Exercises

Searching the Repository of Ancillary Files

To search for a particular file, you can use this page and specify a string to search for:

Searching in the file name:
https://www.cpdn.org/cpdnboinc/ancil_search_files.php?file_name=restart_pnw

Search in the description:
https://www.cpdn.org/cpdnboinc/ancil_search_files.php?description=batch_382

The full information for each file is listed by using the ancil_file_info.php page and entering the file name.
e.g:https://www.cpdn.org/cpdnboinc/ancil_file_info.php?file_name=GloSea_sst_201506_201601_09_allforcing_ancil.gz

Downloading Ancillary files

There is an entry for the "url" of each ancillary file. Each file can be downloaded from it's url. This follows the structure:
https://download.cpdn.org/cpdn_ancil_files/restart_regional/dchapa.start.EU.b.new.0000.gz

The url for each file is listed on the ancil_file_info page.

UM Namelists: Forcings

Namelists in the UM contain the parameter settings needed to run the model. These are passed into the model at the start of the simulation.

Three types of forcing input are contained in the namelists:

- Solar forcing
- Volcanic forcing
- Well mixed greenhouse gas (GHG) forcing

In CPDN these namelist forcing elements are kept in the ancillary files repository.

The GHG file is a complete namelist section that gets substituted into a "namelist template" to form a full namelist to be input into the model on submission.

The solar and volcanic forcings are XML segments corresponding to the relevant year and input values. On submission values for the relevant year(s) are inserted into the namelist template file. Note that for HadCM3s 73 latitude bands are used for the volcanic forcing as opposed to 4 in weather@home models.

UM Namelists: Parameter Settings

A number of values in the namelist template are configured to take either a default value (specified in the template) or an input value from an appropriately named parameter tag in the XML workunit specification. The xml tag name to use is specified in the namelist template (which can be supplied on request from OeRC) and appear as $<tag_name> , for example:

Namelist entry:

ENTCOEF={$entcoef or 3.0}

```
<parameters>
   <file_atmos>xhjlya.start.0000.360.new</file_atmos>
   <file_region>xacxfa.start.0000</file_region>
   <file_sst>final_ancil_2year_OSTIA_sst_1995-12-01_1997-12-30</file_sst>
   <file_sice>final_ancil_2year_OSTIA_ice_1995-12-01_1997-12-30</file_sice>
   <file_so2dms>so2dms_hist_N96_1989_2000v2</file_so2dms>
   <file_sulphox>oxi.addfa</file_sulphox>
   <file_ozone>ozone_hist_N96_1989_2000v2</file_ozone>
   <file_pert>ic00000000_10_N96</file_pert>
   <file_solar>solar_1985_2020</file_solar>
   <file_volcanic>volc_cmip5</file_volcanic>
   <file_ghg>ghg_defaults</file_ghg>
   <run_years>1</run_years>
   <exptid>a08c</exptid>
   <model_start_year>1995</model_start_year>
   <entcoef>9.0000</entcoef>
</parameters>
```

Tag in workunit XML

If a parameter that you wish to perturb does not appear in the namelist template or you require a copy of the current namelist templates for the model please contact the OeRC applications team for assistance.

UM Namelists: Parameter Settings

In certain circumstances it may be desirable to have one parameter setting for the global model and a different setting for the regional model. To do this simply prefix the parameter tag with “global_” or “region_” in the XML file. This will apply the “global_” value to the global model namelist and the “region_” value to the regional model namelist.

```
<parameters>
   <file_atmos>xhjlya.start.0000.360.new</file_atmos>
   <file_region>xacxfa.start.0000</file_region>
   <file_sst>final_ancil_2year_OSTIA_sst_1995-12-01_1997-12-30</file_sst>
   <file_sice>final_ancil_2year_OSTIA_ice_1995-12-01_1997-12-30</file_sice>
   <file_so2dms>so2dms_hist_N96_1989_2000v2</file_so2dms>
   <file_sulphox>oxi.addfa</file_sulphox>
   <file_ozone>ozone_hist_N96_1989_2000v2</file_ozone>
   <file_pert>ic00000000_10_N96</file_pert>
   <file_solar>solar_1985_2020</file_solar>
   <file_volcanic>volc_cmip5</file_volcanic>
   <file_ghg>ghg_defaults</file_ghg>
   <run_years>1</run_years>
   <exptid>a08c</exptid>
   <model_start_year>1995</model_start_year>
   <global_entcoef>9.0000</global_entcoef>
```

NOTE: Parameter values used for a given workunit are added to the global attributes of the output NetCDF files for further traceability. No entry means the default values have been used.

Outline

Overview, server setup and workflow:

- Introduction to distributed computing
- Workflow
- Available models
- Monitoring progress
- Filename conventions

Submission:

- XMLS and Batches
- Ancillary Files and Namelists
- Diagnostic output or STASH

Result data handling:

- Data extraction
- Restart extraction
- Regridding regional data

Publications

Practical Exercises

Diagnostic output or STASH

Stash files are the files that configure the model's diagnostic output. These files are generally denoted by 'stashc' in the file name.

A good starting point would be to take a pre-existing basic stash e.g:

For weather@home:
xaakm_global_basic_2016-04-18.stashc (global stash)
xacxf_region_basic_2016-07-19_v5.stashc (regional stash)

For hadcm3:
xabnk.stashc.monthly

Stash files can be searched for in the cpdn_ancil_files repository using the link:
https://www.cpdn.org/cpdnboinc/ancil_search_files.php?file_name=stashc

We have developed a STASH translator tool to provide a more human readable version of the STASH file available at:
https://github.com/CPDN-git/cpdn_um_stash_translator

There is also a useful reference document describing STASH in the UM in more detail:
http://cms.ncas.ac.uk/documents/vn4.5/pc004.pdf

Adding new STASH items

Each diagnostic available to the UM is has a stash code associated with it.

This consists of a Model code (Atmosphere=1, Ocean=2), Section code and Item code.

The section and item of often combined into a Stash code which is just Section*1000+Item.

A list of stash codes is available on the web:
http://puma.nerc.ac.uk/STASH_to_CF/STASH_to_CF.html
Note that for CPDN we use UM version 4.5 and therefore STASH items must be valid for that version of the model.

Stash requests are made by a string e.g.

&STREQ IMOD= 1, ISEC=16, ITEM=222, IDOM=3, ITIM=4, IUSE=6 /

This is split up into:

- IMOD: Model code
- ISEC: Section code
- ITEM: Item code
- IDOM: Spatial domain code
- ITIM: Time frequency/averaging code
- IUSE: Output stream

The number of stash requests need to be counted and included in the field “NUM_REQ”

TIP: Add comments with a description and numbering

Outline

Overview, server setup and workflow:

- Introduction to distributed computing
- Workflow
- Available models
- Monitoring progress
- Filename conventions

Submission:

- XMLS and Batches
- Ancillary Files and Namelists
- Diagnostic output or STASH

Result data handling:

- Data extraction
- Restart extraction
- Regridding regional data

Publications

Practical Exercises

Data Extraction and Subsetting

We have two scripts you can run to extract particular variables out of the CPDN data zips:

wah_extract_local.py: extracts data by specifying a directory on the upload server filesystem. This is intended for use on systems such as JASMIN and **not** the OeRC subsetting server (unless you have prior agreement to do so when the scratch space should be used).

wah_extract_wget.py: extracts data by taking a list of zip file urls (from a .txt.gz file supplied in batch e-mail) and extracting to a remote server (e.g. OUCE linux)

The code is available here: https://github.com/CPDN-git/cpdn_extract_scripts
It is recommended that the repository is cloned using git, so you can pull updates.
(Ask Sarah Sparrow if you are unsure about how to do this)

There is a README file there that explains the operation, and you can also run the functions with the -h option (wah_extract_local.py -h or wah_extract_wget.py -h) for more information about the command line arguments.

Note:

- **The extraction scripts require netCDF4 python libraries. There are instructions in the readme file on how to use miniconda to install these if required.**
- **The extraction scripts require a stash code rather than a fieldname.**
- **They also require extra inputs to specify cell methods and the time frequency of the variable to select.**

Data Extraction and Subsetting

Command line arguments are:

wah_extract_local.py ONLY:

-i / --in_dir: Input directory containing
(e.g. /group_workspaces/jasmin2/got

wah_extract_wget.py ONLY:

-u / --urls_file: File containing list of u

BOTH SCRIPTS:

-o / --out_dir: Output directory for ex
-y / --year: Year to extract, set to 0 fo
-s / --start_zip: First zip to extract
-e / --end_zip: Last zip to extract
-f / --fields: List of (comma separated)

Each field entry has the form
[file_stream,stash_code,[region],proc

where:

file_stream = ga.pd (regional daily), g
stash_code = stash_section * 1000 + s
[region] = [lon_NW,lat_NW,lon_SE,lat
process = spatial post_processing: mi
time_freq = input variable data freque
cell_method = input variable time cell method: minimum,maximum,mean
vert_lev = input variable name of vertical level in netcdf file or ''

```
# Remote extraction script
# Script to extract data from BATCH to a remote server/ personal computer

#
# Set up paths
EXTRACT_SCRIPTS_DIR=..
EXTRACT_DATA_DIR=../extracted_data

BATCH=443
YEAR=0
# Get List of URLS
if [ ! -f ./batch_${BATCH}.txt.gz ]; then
  wget http://upload2.cpdn.org/results/batch_${BATCH}/batch_${BATCH}.txt.gz
fi
# Extact data from the list of URLS
$EXTRACT_SCRIPTS_DIR/wah_extract_wget.py -u ./batch_${BATCH}.txt.gz \
-f "\
['ma.pc',3236,[],'all',150,400,720,'maximum',''],\
['ma.pc',3236,[],'all',150,400,720,'minimum',''],\
['ma.pc',3236,[],'all',150,400,720,'mean',''],\
['ga.pe',5216,[],'all',-0.0001,1,720,'mean',''],\
['ga.pd',5216,[],'all',-0.0001,1,24,'mean','z0'],\
"  -o $EXTRACT_DATA_DIR/batch_${BATCH} –y $YEAR -s 1 -e 2
```

Duplicated days in CPDN global data

When a participant suspends their simulation during the regional model, the global model will repeat the day it just computed when restarted.

Here is an example of what could possibly happen where daily taxmin and tasmax are output in the same file:

Normal data:
field16 (tasmin): 30 days
field16_1 (tasmax): 30 days

With 2 duplicated days:
field16 (tasmin): 30 days
field16_1 (tasmin): 2 days
field16_2 (tasmax): 30 days
field16_3 (tasmax): 2 days

So you can see that with duplicated days, the variable names can be different from the normal data.
To determine the correct variables, you need to rely on:

1. cell methods of variable.
   e.g. field16_2: cell_method = "time: maximum"
2. meaning_period of time axis (to determine whether the data is daily or monthly)
   e.g. time0:meaning_period = "24 hours"
3. stash code (multiple different stash codes can map to the same field (e.g. T on pressure levels and T at surface).

This feature will only affect global model daily data and is handled correctly in the extraction scripts

Restart File Extraction

Code for extracting restart files from an existing batch to use for a new simulation is provided in the cpdn_xml_generation repository: https://github.com/CPDN-git/cpdn_xml_generation

This code also checks that the date of the restart file is the first of the month.

Command line arguments are:

--batch	= batch number of extract restarts from
--data_dir	= location of the batch data directory
--out_dir	= location to put the extracted restarts in
--model_type	= model type enter 'global', 'coupled' or 'nested'
--dry_run	= do a dry run without extracting files

The model_type options should be used as follows:

- 'global' when extracting weather@home2 restarts from a global only simulation
- 'coupled' when extracting hadcm3s restarts from the coupled model
- 'nested' when extracting weather@home2 restarts for the global and regional model

Note the default model_type is set to 'nested'

Regridding the Regional Data

Using CDO: REGIONAL TO LAT-LON

CDO can convert regional data files to regular lat-lon grids:

This works straight out for some regridding methods e.g. bilinear interpolation:

```
cdo remapbil,r720x360 wah_anz_regional_outpt.
```

Some forms of remapping require more
the -setgrid and specifying the rotated g

```
cdo remapcon2,r360x180 -setgrid,rot_grid_anz.tx
```

```
#
# gridID 1
#
gridtype  = lonlat
gridsize  = 31320
xname     = longitude0
xlongname = longitude in rotated pole grid
xunits    = degrees
yname     = latitude0
ylongname = latitude in rotated pole grid
yunits    = degrees
xsize     = 216
xfirst    = 138.64
xinc      = 0.44
yfirst    = 36.96
yinc      = -0.44
```

Where rot_grid_anz.txt is:

Note:

- This only works for variables on a single grid, so you should select specific variables to regrid using this method.
- If you are extracting variables on the shifted grid (e.g. winds), you will have to

Using CDO: LAT-LON TO REGIONAL

To convert lat-lon data to the regional model grid, use the same format in reverse:

```
cdo remapcon2,rot_grid_anz.txt lat-lon_data.nc anz_region_data.nc
```

Outline

Overview of server setup and workflow:

- Introduction to distributed computing
- Workflow
- Available models
- Filename conventions
- Monitoring progress

Submission:

- XMLS and Batches
- Ancillary file creation and upload
- Namelists and parameter perturbations
- Diagnostic output or STASH

Result data handling:

- Data extraction
- Restart extraction
- Regridding regional data

Publications

Practical Exercises

Publications

We ask that when you have a paper published using CPDN data you let us know:

- Publication doi
- Batch numbers for data used within that publication

In this way we can track output and enable our public volunteers to link to the research that they have actively contributed towards.

For important output data or datasets that represent a significant resource for the community we actively encourage and support production of a **data publication** in a journal such as Data in Brief or Nature Scientific Data where as part of this process the data will be lodged in a publically available well managed repository.

If you require clarification or assistance with this please contact a member of the CPDN team who will be happy to help you.

Outline

Overview of server setup and workflow:

- Introduction to distributed computing
- Workflow
- Available models
- Filename conventions
- Monitoring progress

Submission:

- XMLS and Batches
- Ancillary file creation and upload
- Namelists and parameter perturbations
- Diagnostic output or STASH

Result data handling:

- Data extraction
- Restart extraction
- Regridding regional data

Publication policy

Practical Exercises

XML Generation Exercise

Create your own xml document containing 4 two month duration workunits for either weather@home or hadcm3s and pass it to Sarah Sparrow ready for a dev test submission.

Clone the basic xml generation scripts from the following git repository:
https://github.com/CPDN-git/cpdn_xml_generation

Tip: GIT repositories can be cloned on the command line by typing:
git clone <URL>

For weather@home:

1. Choose the region, ancillary files and STASH checking for their availability in the repository as shown on:http://climateprediction.net/wiki/doku.php?id=cpdn_ancil_repo
2. Create a generic spinup XML using either create_basic_attribution.py or create_basic_spinup_climatology.py

For hadcm3:

1. Choose the ancillary files checking for their availability in the repository as shown on: http://climateprediction.net/wiki/doku.php?id=cpdn_ancil_repo
2. Use the example hadcm3 generation code create_basic_hadcm3_xml.py to generate an xml with parameter perturbations and including standard physics settings.

Data Extraction Exercise

Download the diagnostic output STASH file from the ancillary file repository.

Use the stash translator tool to examine the output and find relevant STASH codes. This is available in the following git repository: https://github.com/CPDN-git/cpdn_um_stash_translator

Tip: The csv file produced can be imported into Excel for easy viewing

Note: Please make sure that your ssh keys are lodged with JASMIN

Clone the extraction scripts from the git repository:
https://github.com/CPDN-git/cpdn_extract_scripts

Extract data from the test batch (batch XXX) using either the local server or wget extract script and extract variables relevant to your analysis. You may want to:

- Extract data from a limited range of zip files.
- Extract data for more than one variable.

Check a random sample of extracted data with xconv, ncview, panoply to check the fields look physical.

Data Extraction Exercise

Suggested data extraction:
Located on JASMIN GOTHAM group workspace in:
/group_workspaces/jasmin2/gotham/gotham/cpdn_data/gotham

Please extract data to the following directory:
/group_workspaces/jasmin2/gotham/summer_school

Group 1: From batches 627-630 extract (‘ma.pc’ filestream)

- 1.5m temperature (Stash code 3236) (Range 150, 400)
- Temperature at 300 hPa (Range 150, 400)
- u wind at 300 hPa (Range -150, 150)
- v wind at 300 hPa (Range -150, 150)

Group 2: From batch 623 (‘ma.pc’ filestream)

- Sea-ice area fraction (Range -0.0001, 1.0001)
- Geopotential height at 1000 hPa and 100 hPa (Range 2000, 20000)
- Sea level pressure (Range 40000, 150000)

Restart File Extraction Exercise

Use the extract_restarts.py script in the cpdn_xml_generation repository:
https://github.com/CPDN-git/cpdn_xml_generation

Extract restart files from the test weather@home2 and hadcm3s batch (batch XXX and batch YYY).

These can then be mass uploaded to the repository for further use. Please contact OeRC applications staff to assist with this when required.

If you have access to the subsetting server (cpdn-ppc01) please place the files to upload in a directory in the staging area:
/gpfs/projects/cpdn/cpdnboinc/tmp_sepia/staging/

EXTRA SLIDES

XML Generation Scripts

The cpdn_xml_generation repository (https://github.com/CPDN-git/cpdn_xml_generation) contains basic code for generating weather@home attribution experiments and climatologies as well as hadcm3s perturbed parameter experiments.

There are four main xml generation scripts:

(1) create_basic_attribution.py

(2) create_basic_spinup_climatology.py

(3) create_second_generation_climatology.py

(4) create_basic_hadcm3_xml.py

IMPORTANT: Remember to edit scripts to specify:

(a) The correct model and region are run

(b) The results are sent to the correct upload server

(c) The batch description is correct and a batch owner is added

XML Generation Scripts

For weather@home simulations:

All scripts:

- Edit to make sure that the correct region, upload server, restart files (or restart csv file listing) and forcing files are used.
- Edit the run length, restart upload month and number of perturbations per restart file as required

create_basic_attribution.py : Creates two xmls for actual and natural experiments

Command line options:

--site= This should contain either 'dev' or 'main' to specify whether the workunit is for the dev or main sites

--generic This means that generic restart files (specified in the script) are used throughout.

create_basic_spinup_climatology.py: Creates an xml for a climatology spinup

Command line options:

--site= This should contain either 'dev' or 'main' to specify whether the workunit is for the dev or main sites

create_second_generation_climatology.py: Creates an xml for a second generation climatology

Command line options:

--site= This should contain either 'dev' or 'main' to specify whether the workunit is for the dev or main sites

XML Generation Scripts

For HadCM3 simulations:

create_basic_hadcm3_xml.py:

Creates a hadcm3 workunit xml with parameter perturbations taken from an existing data file (this is located in param_data directory of the git repository)

Command line options:

--site=	This should contain either 'dev' or 'main' to specify whether the workunit is for the dev or main sites
--generic	This will apply the standard physics restart files throughout
--add_stdp	This will ensure that the standard physics configuration is included as a workunit in the xml
--paramids=	This should be either a comma separated list of parameter sets OR filename to read parameter sets from default will be all parameter sets in the data structure

JASMIN Login

1. Ensure SSH keys are lodged in the JASMIN/BADC website.

1. Start the SSH agent

```
exec ssh-agent $SHELL
ssh-add .ssh/id_rsa
```

1. Login to JASMIN and on the the analysis machine

```
ssh –XA <username>@jasmin-login1.ceda.ac.uk
ssh –XA <username>@jasmin-sci2.ceda.ac.uk
```

Data Extraction Exercise

Suggested data extraction:
Located on JASMIN GOTHAM group workspace in:
/group_workspaces/jasmin2/gotham/gotham/cpdn_data/gotham

Please extract data to the following directory:
/group_workspaces/jasmin2/gotham/summer_school

Group 1: From batches 627-630 extract ('ma.pc' filestream)

- 1.5m temperature (Stash code 3236)
- Temperature at 300 hPa (Stash code 16203)
- u wind at 300 hPa (Stash code 15201)
- v wind at 300 hPa (Stash code 15202)

Group 2: From batch 623 ('ma.pc' filestream)

- Sea-ice area fraction (Stash code 31)
- Geopotential height at 1000 hPa and 100 hPa (Stash code 16202)
- Sea level pressure (Stash code 16222)

Data Extraction Exercise

```
# Extraction script for Exercise 1
# Set up paths and variables
EXTRACT_SCRIPTS_DIR=../cpdn_extract_scripts/
EXTRACT_DATA_DIR=/group_workspaces/jasmin2/gotham/summer_school
BATCH=630
PROJECT=gotham
BATCH_DATA_DIR=/group_workspaces/jasmin2/gotham/gotham/cpdn_data/$PROJE
CT/batch_${BATCH}/successful/
YEAR=2010
START_ZIP=1
END_ZIP=4

# Extract data
$EXTRACT_SCRIPTS_DIR/wah_extract_local.py -i $BATCH_DATA_DIR \
-f "\
['ma.pc',15201,[],'all',-150,150,24,'mean','z0'],\
['ma.pc',15202,[],'all',-150,150,24,'mean','z0'],\
['ma.pc',16203,[],'all',150,400,24,'mean','z0'],\
['ma.pc',3236,[],'all',150,400,720,'mean','z6'],\
"  -o $EXTRACT_DATA_DIR/batch_${BATCH} -y $YEAR -s $START_ZIP -e $END_ZIP
```

Data Extraction Exercise

```
# Extraction script for Exercise 2
# Set up paths and variables
EXTRACT_SCRIPTS_DIR=../cpdn_extract_scripts/
EXTRACT_DATA_DIR=/group_workspaces/jasmin2/gotham/summer_school
BATCH=623
PROJECT=gotham
BATCH_DATA_DIR=/group_workspaces/jasmin2/gotham/gotham/cpdn_data/$PROJE
CT/batch_${BATCH}/successful/
YEAR=0
START_ZIP=10
END_ZIP=15

# Extract data
$EXTRACT_SCRIPTS_DIR/wah_extract_local.py -i $BATCH_DATA_DIR \
-f "\
['ma.pc',31,[],'all',-0.0001,1.0001,24,'mean','z0'],\
['ma.pc',16202,[],'all',-2000,20000,24,'mean','z2'],\
['ma.pc',16222,[],'all',40000,150000,24,'mean','z3'],\
['ma.pc',16222,[],'all',40000,150000,720,'mean','z3'],\
"  -o $EXTRACT_DATA_DIR/batch_${BATCH} -y $YEAR -s $START_ZIP -e $END_ZIP
```

Glossary

Workunit	A description of work given to a single volunteer's computer to compute a single ensemble run.
XML	In CPDN, this refers to the description of the workunits. Describes start date, run length, forcing files, parameter settings, upload destination etc.
STASH	Spatial and Temporal Averaging and Storage Handling. Defines the output from the model.
Ancillary file	File containing some external forcing to the model.
Start dump	The initial condition of the models (global, regional, ocean).
Namelist	Text file that defines values for variables in the model. Typically used to specify GHG, other well mixed gases, volcanics and solar cycle.