Testing and reviewing code for the Met Office Unified Model

N. Luke Abraham
luke.abraham@atm.ch.cam.ac.uk

With thanks to:
Glenn Greed, Stuart Whitehouse, and the MetUM Systems Team
Steven Hardiman, Nigel Wood, & Fiona O’Connor
James Mollard
Grenville Lister, Karthee Sivalingam, Simon Wilson, and NCAS CMS
Outline

• What is the Unified Model?
• Code organisation
• Making a change
  • Scientific vs. technical testing
• Examples
• Supporting the UK user community
• Take-home messages
What is the Unified Model?
Let's unmask the truth.... Why do we have a UM?
Modelling: integrating our knowledge of atmospheric behaviour forward in time

- Newton’s laws for rotating fluid
- Gas laws
- Laws of thermodynamics

Surface Processes

• The challenge:

To reproduce the behaviour of (hazardous) weather systems
**Operational forecasts**
- Mesoscale (resolution approx. 4km, 1.5km)
- Global scale (resolution approx. 17km)

**Global and regional climate predictions**
- Resolution around 120km
- Run for 10-100-… years

**Seasonal predictions**
- Resolution approx. 60km

**Research mode**
- Resolution 1km - 10m

> 25 years old
The consequence of unification

...the same scheme has to continue to work

A factor of 1000 between these

© Crown copyright   Met Office
Exoplanets: Hot Jupiters


Mayne et al. (2013b) (The UM, a fully-compressible, non-hydrostatic, deep atmosphere GCM, applied to hot Jupiters
What is the UM?

- Sophisticated numerical modelling software:

- May be run in many modes:
  - Global, Limited Area Model; Mesoscale (NWP)
  - Aquaplanet, SCM (Idealised tests)
  - Climate modelling; atmosphere only or coupled with ocean models….etc

- Exoplanet research
Development of Models (1)

Figure courtesy of UCAR
Development of Models (2)
Global model cf. other centres

Parameters: PMSL, 500hPA GPH, 250hPa/850hPA Winds; Range: T+24 to T+120

© Crown copyright Met Office

* Parameters: PMSL, 500hPA GPH, 250hPa/850hPA Winds; Range: T+24 to T+120
CODE MANAGEMENT & ORGANISATION
Developers and Users

- The Unified Model is primarily used and developed by the U.K. Met Office
- There are several hundred users in U.K. Universities
- The MetUM is also licensed abroad, and is used in Australia, New Zealand, South Korea, South Africa, and India. There are also some partners in the United States
- There are almost 200 active developers of the MetUM
Code Organisation

- The Unified Model has over 1 million lines of code, organised into 52 sections.
- Each section, e.g. convection, chemistry & aerosols, & top level control routines, have a **code owner**.
  - The code owner is responsible for that section, and must approve all changes that are made to that section.
  - The code owner also has first refusal to review the code in more detail.
- The code is managed by a Project Board, who are responsible for its delivery.
What is the UM made of?

Mainly Fortran and some C

Fortran 77 through Fortran 95 and recently the use of some Fortran 2003.

Uses FCM for code management (Trac and subversion)

To support the UM infrastructure:

bash Perl and Python

rose/cylc – python gtk
Release schedule

~3-4 release per annum
Project Board oversees the release schedule plans.
UM Release Build Cycle

Stable Release

Develop Changes
- Create Ticket
- Create branch
- Develop change
- Test change
- Document change

Science Reviews
- Scientific/Technical correctness
  - Code changes correct
  - Sufficient documentation explaining change and impacts
  - Ticket updated

Integration Testing
- Regression testing
- Fix problems

System Reviews
- Coding standards
- System impacts
- Sufficient testing?
- Ticket updated

Freeze Changes

Code Merge into Trunk
- Check Daily builds
- If ok Ticket closed

Resolve code conflicts (Branch from trunk)
Making Code changes:
Typical Workflow (as used at the Met Office)

Time

Trac
Web pages
Repository
Fcm Server
Working Copy
Desktop
Remote Machine (eg HPC)

Create a Ticket → Document changes

Create a Branch

Checkout UMx.x code → Modify Code

Build New exec and test code

Commit back to the branch
Making Code changes:
Create a ticket
Documentation

- In line with code
  - UM code in repository
- Documentation papers
  - UMDP source code in repository
- Tickets
  - TRAC pages
    - Links to UM code branch
    - Links to UMDP source branch
    - Ticket Summary
    - Standardised Testing reports
Testing, using the Test Harness.

- A single Test Harness is used by
  - code developers
  - daily builds of the UM trunk
  - Release Candidate testing, checking progression between stable UM versions and applying wider set of tests.

- The user determines which tests are performed.

- The test output provides the supporting evidence of any change; in a standard format across a UM system team controlled set of tests.
• There are over 120 separate tests that can be run using the test harness.
• These are organised into various groups, e.g. *all*, *developer*, *ukca*, *recon*, which test specific configurations.
• The minimum test that needs to be run is *developer*, but code owners will usually ask for their own tests to be run if a change is being made to their section.
Ticket summary:

• Communication of the change…
  • what is the vital information that any review needs to know?
  • formalise this communication through TRAC templates….
Ticket Summary #343

To be completed prior to tech review and updated as required during the review process.

All developers are expected to have worked through:

- Write documentation
- Add unit tests
- This ticket summary needs to be completed to provide evidence of the impact of each change.
- New to supply test evidence.
  - Met Office tickets:
    - are expected to provide evidence from rose stem tests based upon the Met Office provided standard test plan (job)
    - are expected to supply as much evidence as possible on the impact of their code based upon their rose stem output.
    - the code system reviewer will perform the Met Office rose stem tests on your behalf as part of their code review.
  - Partners without access to a rose stem suite:
    - will require a proxy at the Met Office (a collaborator) to take ownership of the ticket and push it through the Met Office rose stem tests and the subsequent review process on behalf of the partner developer.

Author: Paul Earnshaw

Branch

Code branch

Documentation branch

Testing branch Only required if you have altered mets data and/or added an upgrade macro

time

Testing

Testing should use rose stem tests which will test a selection of standard met configurations.

Rose testing summary

Please list the rose stem groups that have been tested from the available groups:

Groups:
- A1 developer
- A2 subgroups
- C1 check_c1
- C2 check_c2
- C3 check_c3
- C4 check_c4
- C5 check_c5
Sci tech review

- The code changes are understood and appropriately made.
- The documentation is sufficient to understand the code change and its impacts. (Ticket, inline code and UMDPs).
- Is familiar with the area of being code updated, preferably a code owner for that UM section.

- If results change previous KGOs, the reviewer must also check that details are provided showing the ‘science’ impact of the change on forecast performance and that it has been given approval by the configuration owners that are impacted.
Scientific Validation

Validation notes are automatically produced from standard output.

A large number of plots can be automatically generated.

Plots typically compare against re-analysis products and reference model output.
Code System review

• Coding standards are met
• Impact of the change on the UM system.
• Ensure all appropriate testing has been performed, request more if required. If the trunk at risk if this change is applied?

• Only Code system reviewers may commit changes to the trunk!
• Thus the trunk is tightly controlled by a small number of staff.
This is very draconian

• Why all the hassle to get a change on the UM trunk?

• To ensure releases are effectively managed, quality controlled and delivered on time.
Size of recent releases

Ticket, Commits per release

Sizes of releases in terms of tickets and commits

UM Release

7.3 7.4 7.5 7.6 7.7 7.8 7.9 8.0 8.1 8.2 8.3 8.4 8.5 8.6 8.7 9.0 9.1 9.2 10.0 10.1

Commits per release

Tickets per release

10.2 250 tickets
10.3 260 tickets
10.4 226 tickets
Does this process catch all problems?

- Test harness
  - Catches *unintended* errors that have been introduced by a change
- Validation notes & evaluation suites
  - Compare the code to observations/reanalysis & existing model output
  - Test that a change improves output compared to what is *currently* the best configuration
- How do you check that (new) schemes don’t break other (new) schemes?
- What can you do if a bug is already in the code?
Process Evaluation Groups

PEGs are set-up to look into particular processes in more detail.

UKCA – coupling with other components

UKCA A34 Prognostics and Chemical Diagnostics (A38 Aerosol Diagnostics)

- DUSTSOIL
- SULPEMIS
- SO2NATEM
- Chemical Tracer Surface Mass Mixing Ratios
- Photolysis Data Files (2D Scheme)
- Stratospheric Aerosol Climatologies
- Chemical Tracer Upper-Boundary Conditions

- JULES
- LANDFRAC
- USRANCIL
- USRMULTI

- UKCA Non-Advected Prognostics
- A0 Primary fields
- A8 Hydrology
- A3 Boundary layer and surface
- A15 Dynamics diagnostics
- A30 Processed climate diagnostics
- OBG
- A5 Convection
- STASH
- Atmospheric Tracer Advection
- A17 Aerosols
- A2 Long-wave radiation
- A4 Large scale precipitation
- ACTIVATE
Aerosols

Atmospheric aerosols in the MetUM are subject to a number of processes, including *wet scavenging*, where aerosols particles are removed by precipitation. Instead of being done in the UKCA code, this was moved to the convection scheme as the required diagnostics were more easily available.

**However**, unknown to the UKCA developers, the diagnostics used in convection were incorrect. This lead to incorrect aerosols, and therefore an incorrect climate.
To stop an error like this happening again, the Systems Team added the following to the Code Review template:

Does this code change make use of prognostic or diagnostic output from other sections (or model)?

If YES, have suitable inline comments been added to the other section’s source code and both UMDPs updated?
UKCA – coupling with other components

UKCA A34 Prognostics and Chemical Diagnostics (A38 Aerosol Diagnostics)

- Chemical Tracer Surface Mass Mixing Ratios
- Stratospheric Aerosol Climatologies
- Chemical Tracer Upper-Boundary Conditions
- Fast-JX Data Files
- Chemical Tracer Data Files (2D Scheme)

- A0 Primary fields
- A8 Hydrology
- A3 Boundary layer and surface
- A15 Dynamics diagnostics
- A30 Processed climate diagnostics
- OBGC
- A5 Convection
- Atmospheric Tracer Advection
- A17 Aerosols

- JULES
- LANDFRAC
- DUSTSOIL
- SULPEMIS
- SO2NATEM
- USRANCIL
- USRMULTI
- RADAER
- A1 Short-wave radiation
- A2 Long-wave radiation
- A4 Large scale precipitation

- ACTIVATE

- ≤22
- ≤340 (≤512)
- ≤150
- ≤5
- ≤5
- NEMO
- CICE
Slotted cylinder test case

Initial Conditions

Kohei Aranami (MetO/JMA)
Age of air in ENDGame far too young, due to conservation scheme. Age of air in ENDGame fixed with new conservation scheme (keeping quintic vertical interpolation for tracers).
Key Points

• Communication is key

• PEGs often require input from many people with expertise in different areas

• The problems with the wet removal of aerosols could have been reduced by opening a dialogue with the convection scheme developers at an earlier stage

• The success of the age of air test means that it is now planned to be used as a standard tracer test in the assessment the next dynamical core, currently under development.
PORTING AND USER SUPPORT
Testing and porting MetUM configurations

• For UK Universities, support for the MetUM is provided by the National Centre for Atmospheric Science, through the Computational Modelling Support section.

• CMS do a large number of different things, through installing the MetUM, porting configurations, managing resources, and developing software tools.

• While Met Office staff move up versions quite quickly, researchers at Universities often keep with the same MetUM version for many years.
  • e.g. UM4.5 is still actively used by many.
Testing a ported job

• The MetUM should (or can be made to) fulfill various criteria:

1. If you run it again a second time, you should get the same answer
2. If you start a new run from a restart file from the middle of a simulation, you should get the same answer
3. If you change the number of cores, the code should give you the same answer
4. While the answers will be different on different architectures, they shouldn’t deviate greatly

• However, scientific validity should be determined by the users
<table>
<thead>
<tr>
<th>UM Version</th>
<th>Job desc</th>
<th>UMUI</th>
<th>Installed</th>
<th>NRUN</th>
<th>CRUN</th>
<th>CHECK PORT</th>
<th>Climate Meaning</th>
<th>Optimisation</th>
<th>Archiving</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.5</td>
<td>N512 L180 GA 6.0</td>
<td>xjanp</td>
<td>Yes</td>
<td>Passed</td>
<td>Passed</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>8.5</td>
<td>N216 L180 GA 6.0</td>
<td>xjief</td>
<td>Yes</td>
<td>Passed</td>
<td>Passed</td>
<td>NA</td>
<td>NA</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>8.4</td>
<td>N512 L85 GA 6.0</td>
<td>xjanu</td>
<td>Yes</td>
<td>Passed</td>
<td>Passed</td>
<td>Passed - HECTOR to ARCHER</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>8.6</td>
<td>GA4.0 &amp; UKCA CheST &amp; GLGMAP</td>
<td>xjnjb</td>
<td>Passed - Only for the same PE decomposition</td>
<td>Failed (expect it to Not bit-compare- Mohit Dalvi)</td>
<td>~</td>
<td>NA</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>8.5</td>
<td>GC2 N96/Orca1</td>
<td>xjnja</td>
<td>Yes</td>
<td>Passed</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>8.5</td>
<td>GA6.0 N216 antib</td>
<td>xjlee</td>
<td>Yes</td>
<td>Passed</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>8.5</td>
<td>GA6.0 N95 antia</td>
<td>xjleo</td>
<td>Yes</td>
<td>Passed</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

```
outputfile.10

0-1161-14 TEMPERATURE INCR swrad - pc2 : 4.161771647161E9 5.84343092000526112E-14 2.8643678510763624E-14
9.74306052503592923E-12
0-1161-15 TEMPERATURE INCR swrad - pc2 : 3.48361591051561 7.77384719588039757E-15 4.16505621239696626E-15
1.2104067748097691E-12
0-1161-16 TEMPERATURE INCR swrad - pc2 : 2161.88171145865 6.48518961457868189E-14 1.39478322840193698E-15
1.0817291506981519E-11
0-1161-17 TEMPERATURE INCR swrad - pc2 : 18.3356237186656 6.57131210961955792E-15 1.53463200385040314E-15
1.06020053847188933E-12
0-1161-18 TEMPERATURE INCR swrad - pc2 : 0.816090978851124 4.59294806955636057E-15 5.08419555367761834E-15
7.32960567240148464E-12
1.52171678036161495E-12
```
Testing a ported job

- `O2`  
  - `Ovector1`  
  - `hfp0`  
  - `hflex_mp=strict`  

- `O2`  
  - `hflex_mp=intolerant`
Testing a ported job

IBM Power 7 (MONSooN)  Cray XC30 (ARCHER)
Take-home messages

• The Unified Model is a large code that has been developed over many years, and which has hundreds of developers who are not all located in the same place

• To ensure that code changes are managed properly a rather complicated change process has been developed

• Communication is key

• The needs of University users are often quite different to those of active developers
THANK YOU!
Met Office NWP System (Global and Regional)

- **Global 17km**: 70 levels, lid 80km
- **EURO 4km**: 70 levels
- **UKV 1.5km**: 70 levels, lid 40km
- **MOGREPS-G**: 33km, 45 member, 12 for forecast, 33 for hybrid analysis
- **MOGREPS-UK**: 2.2km, 12 members
Releases, Parallel Suites and Operational Cycles (Development Process)

- Code Development
  - UM Release: UM10.x
  - New science or technical capabilities (branches) developed
  - UM Release: UM(10.x+1)

- Operational Upgrades
  - Parallel Suite: PSxx
  - Operational Configuration Cycle
  - Improvement!

NB: May have many Operation cycles at a given UM release.

Time →
Example Run of rose-stem
Test fail when compared with KGO
UNIFIED MODEL

Forcing Evaluation

Data Assimilation

OBS Evaluation

Global

Regional

Idealised

CRM

Km Scale

NWP & THORPEX

SEASONAL DECADAL

CLIMATE

NWP REGIONAL

AQUAPLANET

JULES

SCM

DYN CORE

UNIFIED MODEL
UKCA Tracers in ENDGame: Age of Air

-Age of air → time taken for parcel of air from free troposphere into (and around) the stratosphere. ‘Observed’ age derived from stratospheric SF6, CO2 (Engel et al)

-Age-of-air tracer in UKCA passively advected, no influence of chemistry

-Initial ENDGame runs → “very young” age (heavy movement across tropopause)

-Working group formed (Summer 2013) with Dynamics Research (DR) Team

-ND, EG parallel runs with combinations of influencing processes (transport, conservation, convection) switched ON and OFF

-Profile ‘different’ to previously tested fields (Q) i.e. mass increasing with height

-ENDGame (ADAS) conservation found most influential in cause of divergence

-Priestley-like conservation scheme developed by DR for testing

-CURRENT best results – Priestley scheme, Quintic interpolation (only for UKCA tr)
\[ \frac{\partial \theta}{\partial t} \approx -w \frac{\partial \theta}{\partial z} \]

Why?


\[ z \]

\[ \theta \]

cubic Lagrange

\[ w > 0 \]

\[ w < 0 \]

cubic Hermite

\[ w > 0 \]

\[ w < 0 \]

Chris Smith (Met Office)