National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Decorating code to expose algorithmic descriptions of the code to the CIM

Simon Wilson NCAS-CMS

Introduction

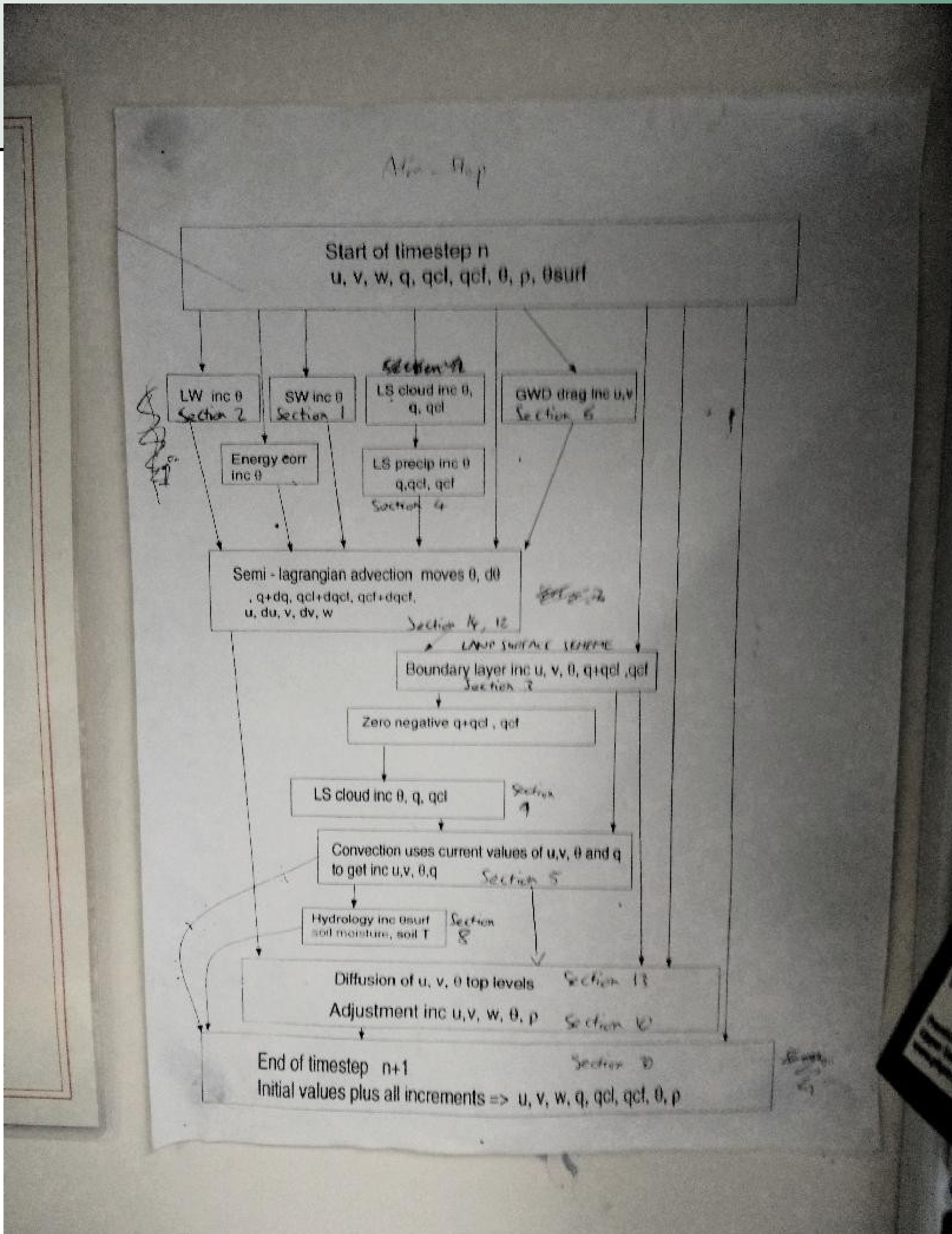
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

- Part of IS-ENES2 WP3 on developing convergent model codes.
- Comprehensive documentation and analysis of models is required so that modelling groups can be better informed of other models' formulation at a detailed level.
- Aim is to provide systems useful in developing strategies for convergent model code using the CIM (Common Information Model).
- NOTE: only concentrating on atmospheric models in the study.

Models and documentation

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

- Many ways to document models, no consistency.
- Still takes effort to “decipher” documentation to understand model structure and prognostic dependencies.
- ~20 model code bases available globally, so any model intercomparison will be a time intensive task.
- But fundamentally most (all?) models consist of a number of prognostics (state variables) such as wind, temperature, moisture, pressure/density,… altered and/or moved by a number of model components (usually) split into physical parametrisations (physics) and fluid motions and thermodynamics (dynamics).
- Should be able to clearly represent the relationship between these in a model timestep, something like this...



My Office Wall.
The anatomy of a timestep in the UM.
Original version long lost.

Requirements

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

- Investigate relationships between model components and prognostics, focusing on dependencies and flow.
- Initially a comparison of the atmosphere timestep and code structure of three different atmosphere models
  - Unified Model from the Met Office in a HadGAM3 configuration
  - HiRAM from GFDL
  - OpenIFS from ECMWF
- By component mean separate sections such as LW radiation, GWD, Convection and the Dynamics (possibly split).

How can we describe and document a model?

Fundamental equations/Model papers (UM)

National Centre for Atmospheric Science

Met Office

Fully-compressible, deep atmosphere equations

$$\frac{D_r u}{Dt} - \frac{uv}{r}\tan\phi - 2\Omega\sin\phi v + \frac{uw}{r} + 2\Omega\cos\phi w + \frac{1}{\rho r\cos\phi}\frac{\partial p}{\partial\lambda} = F_u,$$

$$\frac{D_r v}{Dt} + \frac{u^2}{r}\tan\phi + 2\Omega\sin\phi u + \frac{vw}{r} + \frac{1}{\rho r}\frac{\partial p}{\partial\phi} = F_v,$$

$$\frac{D_r w}{Dt} - \frac{u^2+v^2}{r} - 2\Omega\cos\phi u + g + \frac{1}{\rho}\frac{\partial p}{\partial r} = 0,$$

where

$$\frac{D_r}{Dt} = \frac{\partial}{\partial t} + \frac{u}{r\cos\phi}\frac{\partial}{\partial\lambda} + \frac{v}{r}\frac{\partial}{\partial\phi} + w\frac{\partial}{\partial r},$$

and

$$\nabla_r.\mathbf{u} = \frac{1}{r\cos\phi}\left[\frac{\partial u}{\partial\lambda} + \frac{\partial\left(v\cos\phi\right)}{\partial\phi}\right] + \frac{1}{r^2}\frac{\partial\left(r^2 w\right)}{\partial r}.$$

$\frac{D_r\rho}{Dt} + \rho\nabla_r.\mathbf{u} = 0,$ $\frac{D_r\theta}{Dt} = F_\theta,$ $p = R\rho T$

from Terry Davis

Technical documentation papers (IFS)

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

SURFRAD_LAYER	Computes radiative properties of the surface.
CLDPRG_LAYER	Calls CLDPP to computes cloud parameters required for the post processing (e.g. total cloud cover).
RADFLUX_LAYER	Calls RADHEATN to compute the temperature tendencies and the downward radiation fluxes at the surface with updated (every time-step) values for the zenith angle.
GWDRAG_LAYER	Computes the tendencies for u, v and T due to the parametrization of subgrid-scale orographic drag. It also computes subgrid orographic coefficients for use in VDFMAIN.
TURBULENCE_LAYER	Calls VDFOUTER/VDFMAIN in two sub time steps for numerical stability. VDFMAIN computes the vertical exchange of u, v, T, q, a, q_l, q_i by turbulence.
CLOUD_LAYER	Calls CLOUDSC as a first guess call of cloud scheme to determine preliminary entry profiles for convection.
CONVECTION_LAYER	Interface to call CUCALLN/CUMASTRN that controls the computation of the tendencies for u, v, T, q, chemical tracers and the cloud detrainment term due to the parametrization of moist convective processes.
CLOUD_LAYER	Calls CLOUDSC to compute tendencies for T, q, a, q_l, q_i, q_r and q_s due to the parametrization of cloud and precipitation processes.
GWDRAGWMS_LAYER	Calls GWDRAG_WMS to compute the tendencies for u, v and T due to the parametrization of non-orographic gravity waves.
METHOX	Computes tendencies for q due to methane oxidation and water vapour photolysis.
SURFTSTP_LAYER	Calls SURFTSTP to control the soil/surface scheme.
STOCHPERT_LAYER	Optionally add stochastic perturbations to physics tendencies.
O3CHEM	Computes tendencies for O_3 due to ozone chemistry.
SLTEND_LAYER	Optionally average tendencies from radiation, convection and cloud in time and space along the semi-Lagrangian trajectory.

Code comments and documentation (HiRAM)

NAMELIST
&coupler_nml

current_date

The date that the current integration starts with.
[integer, dimension(6), default: 0]

force_date_from_namelist

Flag that determines whether the namelist variable current_date should override the date in the restart file INPUT/coupler.res. If the restart file does not exist then force_date_from_namelist has not effect, the value of current_date will be used.
[logical, default: .false.]

calendar

The calendar type used by the current integration. Valid values are consistent with the time_manager module: 'julian', 'noleap', or 'thirty_day'. The value 'no_calendar' can not be used because the time_manager's date function are used. All values must be lowercase.
[character(maxlen=17), default: '']

months

The number of months that the current integration will be run for.
[integer, default: 0]

days

The number of days that the current integration will be run for.
[integer, default: 0]

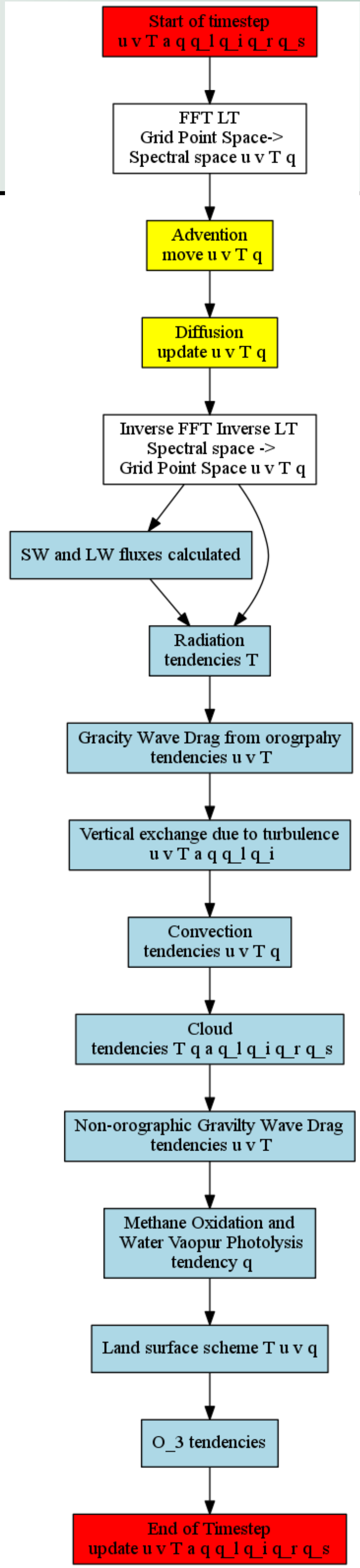
```
MAIN PROGRAM EXAMPLE
      --------------------
         DO slow time steps (ocean)
              call flux_ocean_to_ice
              call ICE_SLOW_UP
              DO fast time steps (atmos)
                   call flux_calculation
                   call ATMOS_DOWN
                   call flux_down_from_atmos
                   call LAND_FAST
                   call ICE_FAST
                   call flux_up_to_atmos
                   call ATMOS_UP
              END DO
              call ICE_SLOW_DN
              call flux_ice_to_ocean
              call OCEAN
         END DO
```

Model code (UM)

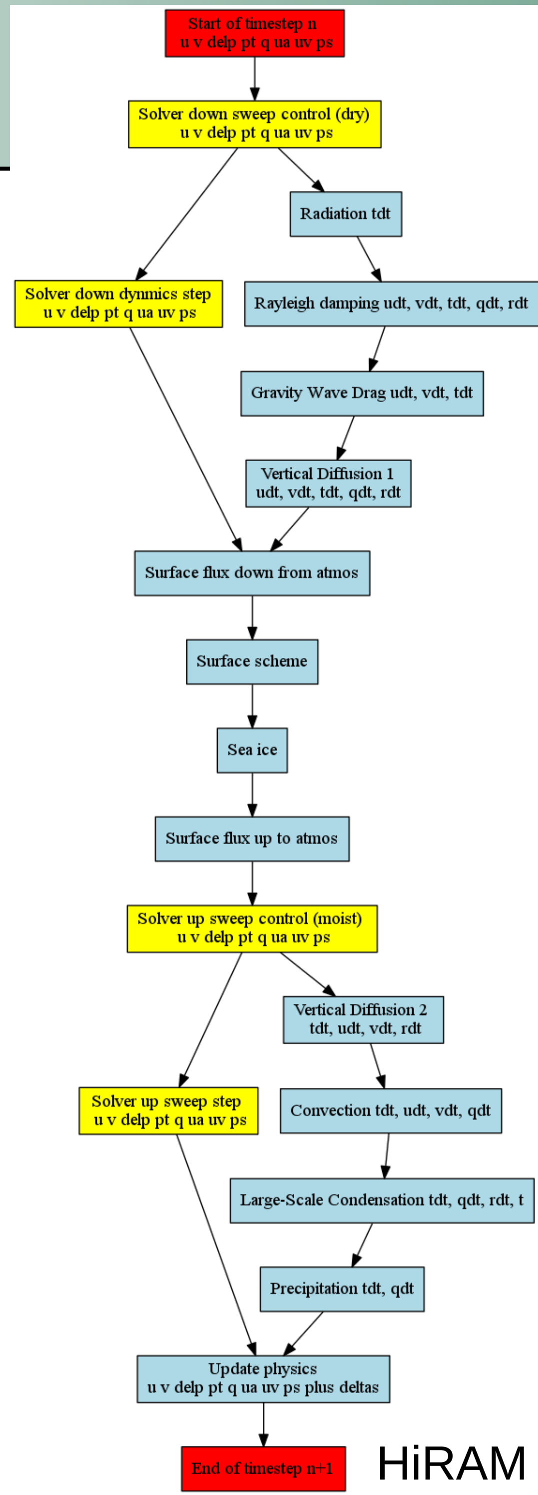
```
     Do level = 1, model_levels
! DEPENDS ON: trsrce
          Call trsrce(                                                   &
     &       rows, row_length, offx, offy                              &
     &,      halo_i, halo_j, model_levels, wet_levels                    &
     &, 0, 0                                                             &
     &,      theta, q_n, qcl_n, qcf_n, exner_rho_levels, rho             &
     &,      aerosol(:,:,level), aerosol_em (:,:,level )                 &
     &,      level, timestep, val_hour, val_minute, amp                  &
     &)
        End Do
      End If

      If ( L_murk_bdry ) Then
! DEPENDS ON: trbdry
        Call trbdry(                                                     &
     &       row_length, rows, n_rows, model_levels                      &
     &,      offx, offy, at_extremity                                  &
     &,      p, u, v                                                     &
     &,      aerosol, timestep                                           &
     &)
      End If
```

- So, by inspection, from model documentation papers (which can number in the 10s), code comments and documentation, code analysis, speaking to model gurus (and some guesswork), generated these...

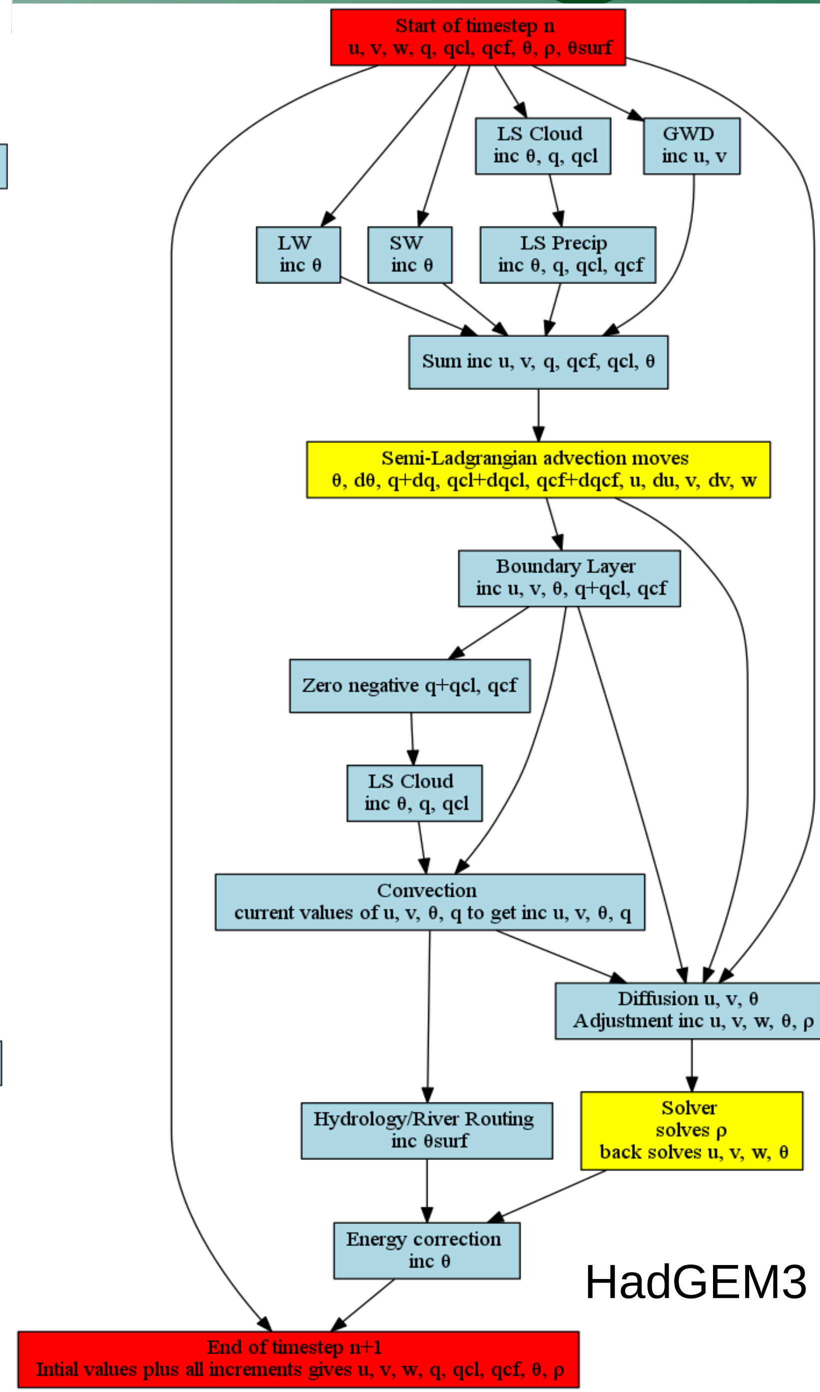
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL



IFS

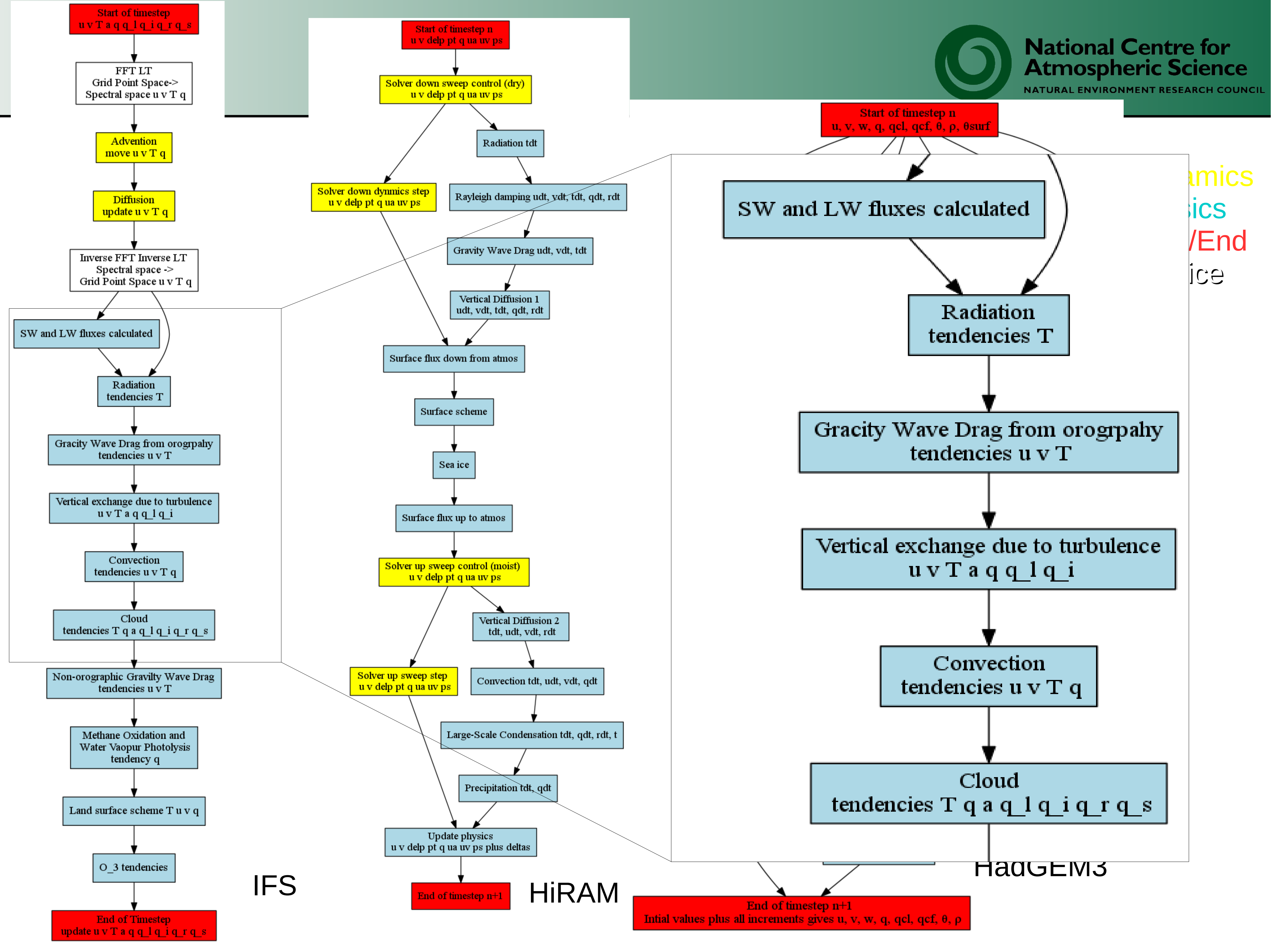


HiRAM

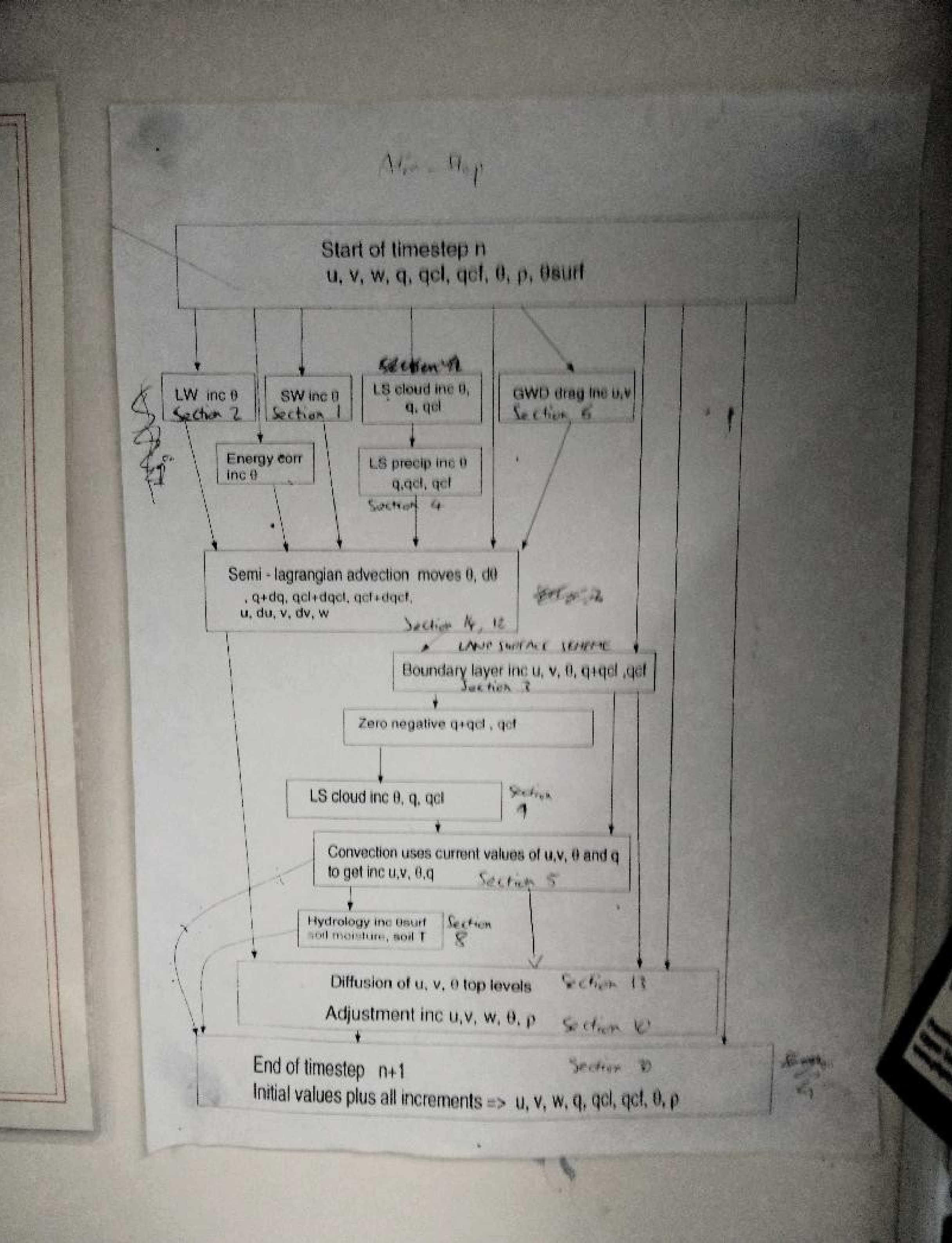


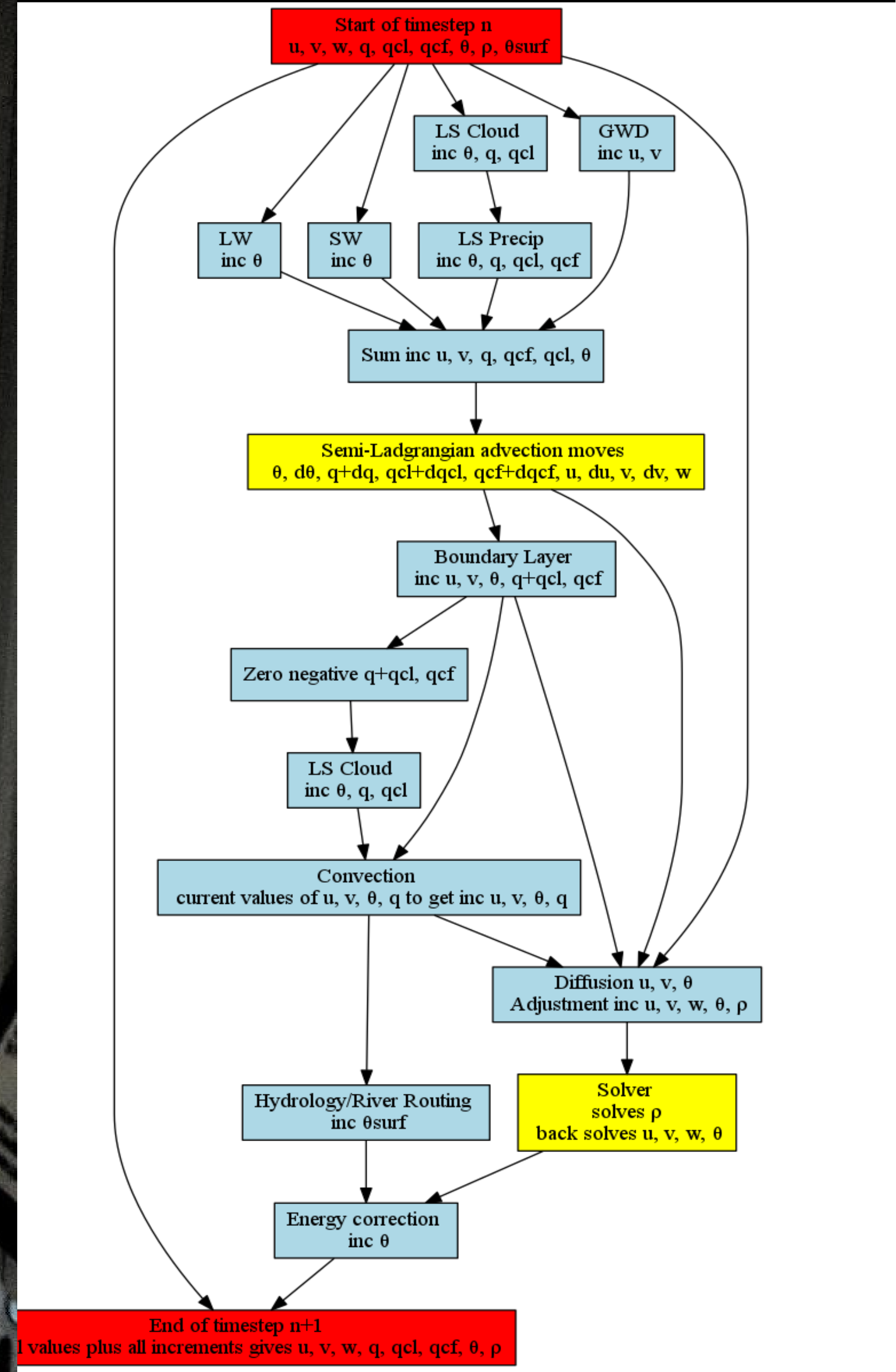
HadGEM3

Dynamics
Physics
Start/End
Service


National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL
Start of timestep
u v T a q q_l q_i q_r q_s
FFT LT
Grid Point Space->
Spectral space u v T q
Advention
move u v T q
Diffusion
update u v T q
Inverse FFT Inverse LT
Spectral space ->
Grid Point Space u v T q
SW and LW fluxes calculated
Radiation
tendencies T
Gracity Wave Drag from orogrpahy
tendencies u v T
Vertical exchange due to turbulence
u v T a q q_l q_i
Convection
tendencies u v T q
Cloud
tendencies T q a q_l q_i q_r q_s
Non-orographic Gravilty Wave Drag
tendencies u v T
Methane Oxidation and
Water Vaopur Photolysis
tendency q
Land surface scheme T u v q
O_3 tendencies
End of Timestep
update u v T a q q_l q_i q_r q_s
IFS
Start of timestep n
u v delp pt q ua uv ps
Solver down sweep control (dry)
u v delp pt q ua uv ps
Radiation tdt
Solver down dynmics step
u v delp pt q ua uv ps
Rayleigh damping udt, vdt, tdt, qdt, rdt
Gravity Wave Drag udt, vdt, tdt
Vertical Diffusion 1
udt, vdt, tdt, qdt, rdt
Surface flux down from atmos
Surface scheme
Sea ice
Surface flux up to atmos
Solver up sweep control (moist)
u v delp pt q ua uv ps
Vertical Diffusion 2
tdt, udt, vdt, rdt
Solver up sweep step
u v delp pt q ua uv ps
Convection tdt, udt, vdt, qdt
Large-Scale Condensation tdt, qdt, rdt, t
Precipitation tdt, qdt
Update physics
u v delp pt q ua uv ps plus deltas
End of timestep n+1
HiRAM
Start of timestep n
u, v, w, q, qcl, qcf, θ, ρ, θsurf
SW and LW fluxes calculated
Radiation
tendencies T
Gracity Wave Drag from orogrpahy
tendencies u v T
Vertical exchange due to turbulence
u v T a q q_l q_i
Convection
tendencies u v T q
Cloud
tendencies T q a q_l q_i q_r q_s
HadGEM3
End of timestep n+1
Intial values plus all increments gives u, v, w, q, qcl, qcf, θ, ρ

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL





Automation

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

- Surely there must be a standard way to document models.
- Yes there is, **THE CIM**.
- CIM constructs exist already:
  - Variable - prognostics
  - SoftwareComponent – model component

CIM construct descriptions

```
software.SoftwareComponent: (
        'composition',
        'connection_points',
        'coupling_framework',
        'dependencies',
        'depends_on',
        'grid',
        'language',
        'license',
        'sub_components',
        'description',
        'development_history',
        'documentation',
        'long_name',
        'name',
        'release_date',
        'repository',
        'version',
   ),

   software.Variable: (
        'description',
        'name',
        'prognostic',
   ),
```

This corresponds to the prognostic variables

The input SoftwareComponent dependencies

How to populate the CIM?

- Can use the structure of the model code.
- Automatic code analysis tools available, but these work on names and structure of functions/subroutines.
- No simple way to identify and separate top level subroutines into functional blocks of code for each model component (Radiation, GWD etc…)
- Identification of prognostics via variable name equally as difficult

Model	Number of files	Number of code statements	Number of subroutines	Number of modules
HadGEM3A	1553	217972	2217	628
HiRAM	297	372394	4505	287
OpenIFS	2776	160286	3675	624

- Direct commenting of code by model experts required.

ROBODoc

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

- After examining a number of automatic code documenting systems, ROBODoc was chosen.
- Specially formatted documentation headers are extracted from (Fortran) source files and stored in a different reformatted file.
- Metadata can be configured by user.
- ROBODoc can reformat the documentation in HTML, XML DocBook, TROFF, ASCII, LaTeX or RTF format.
- Example format...

Robodoc example

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

```
!!****f* main/main
!! NAME
!! Met Office Unified Model
!!***
...


!!****p* prog/u
!!*  NAME
!!*   U wind
!!*  DESCRIPTION
!!*   Wind blowing to the East.
!!***

!!****f* phy/Conv
!!*  NAME
!!*   Convection
!!*  DESCRIPTION
!!*   Parametrization of convective clouds, their precipitation,
!!*   phase changes and transports of heat, water and momentum.
!!*  INPUTS
!!*   u v theta q
!!*  USES
!!*   LSCloud BLayer
!!*  VERSION
!!*   A05_6A
!!***
```

Workflow

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

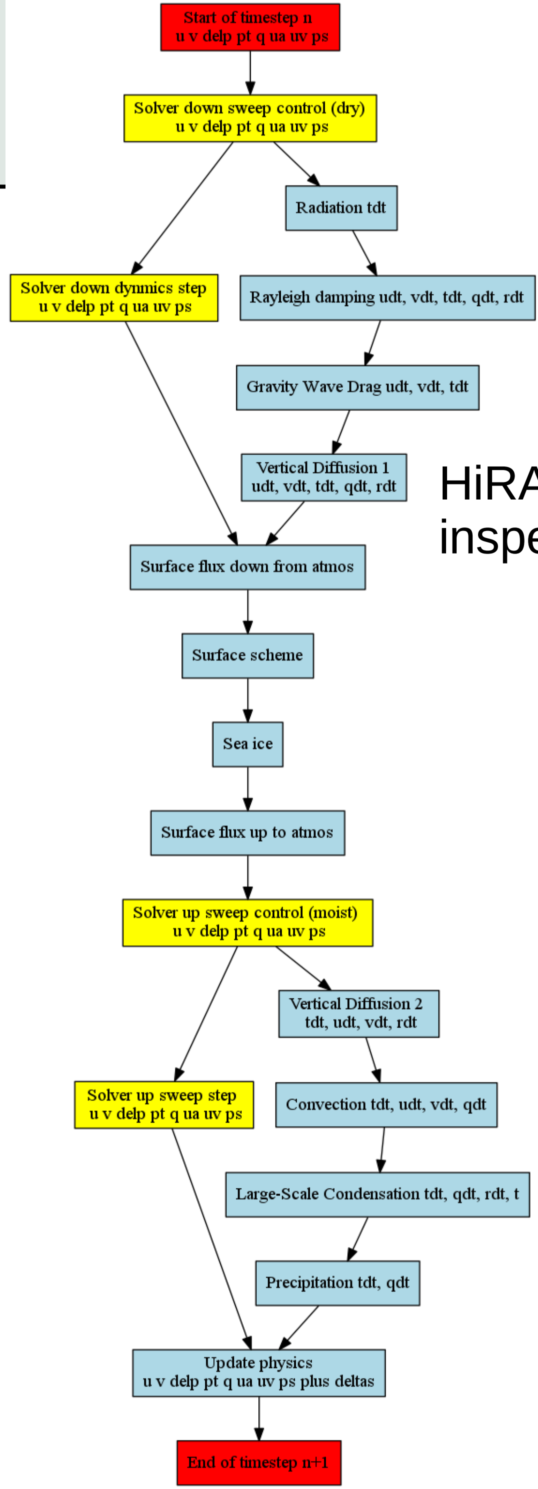
- Header block in code containing base model description and prognostic details.
- For each Software component (GWD, Convection etc) control routine, a RODODoc comment block added.
- Model code tree analysed by ROBODoc, and a DocBook XML file generated.
- Python code to read in XML and fill CIM directly from contents.
- Once model details in CIM, simple (a few lines in python) to traverse SoftwareComponent list to generate graphviz dot format flowchart.

SoftwareComponent construct

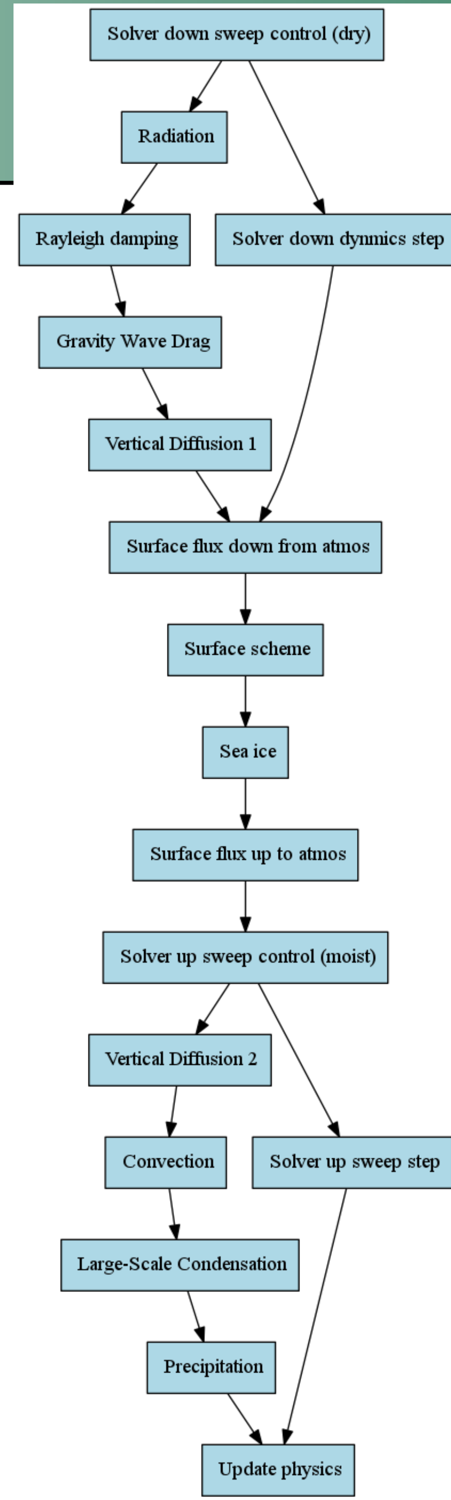

```
{'name': ' Convection',

'description': 'Parametrization of convective clouds, their precipitation, phase changes and transports of heat, water and momentum.',

'depends_on':
[<pyesdoc.ontologies.cim.v2.typeset_for_software_package.SoftwareComponent object at 0x7f8303f8e0d0>,
<pyesdoc.ontologies.cim.v2.typeset_for_software_package.SoftwareComponent object at 0x7f8303f8d0d0>],

'version': 'A05_6A',

'connection_points':
[<pyesdoc.ontologies.cim.v2.typeset_for_software_package.Variable object at 0x7f8303f6ff90>, <pyesdoc.ontologies.cim.v2.typeset_for_software_package.Variable object at 0x7f8303f6ffd0>,
<pyesdoc.ontologies.cim.v2.typeset_for_software_package.Variable object at 0x7f8303f6fed0>, <pyesdoc.ontologies.cim.v2.typeset_for_software_package.Variable object at 0x7f8303f6ff10>],
.
.
.}
```


LSCloud

u



HiRAM by inspection



HiRAM auto-generated via CIM

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Also:

Can output relational information:

```
Prognostic:  (u)  U wind
Flow:
  Vertical Diffusion 2
  Gravity Wave Drag
  Vertical Diffusion 1
  Solver up sweep step
  Solver up sweep control (moist)
  Solver down sweep control (dry)
  Update physics
  Solver down dynamics step
  Rayleigh damping

Component:  (Precip)  Precipitation
   Prognostics:
     Temperature
     Moisture
   Depends on:
     Large-Scale Condensation
```

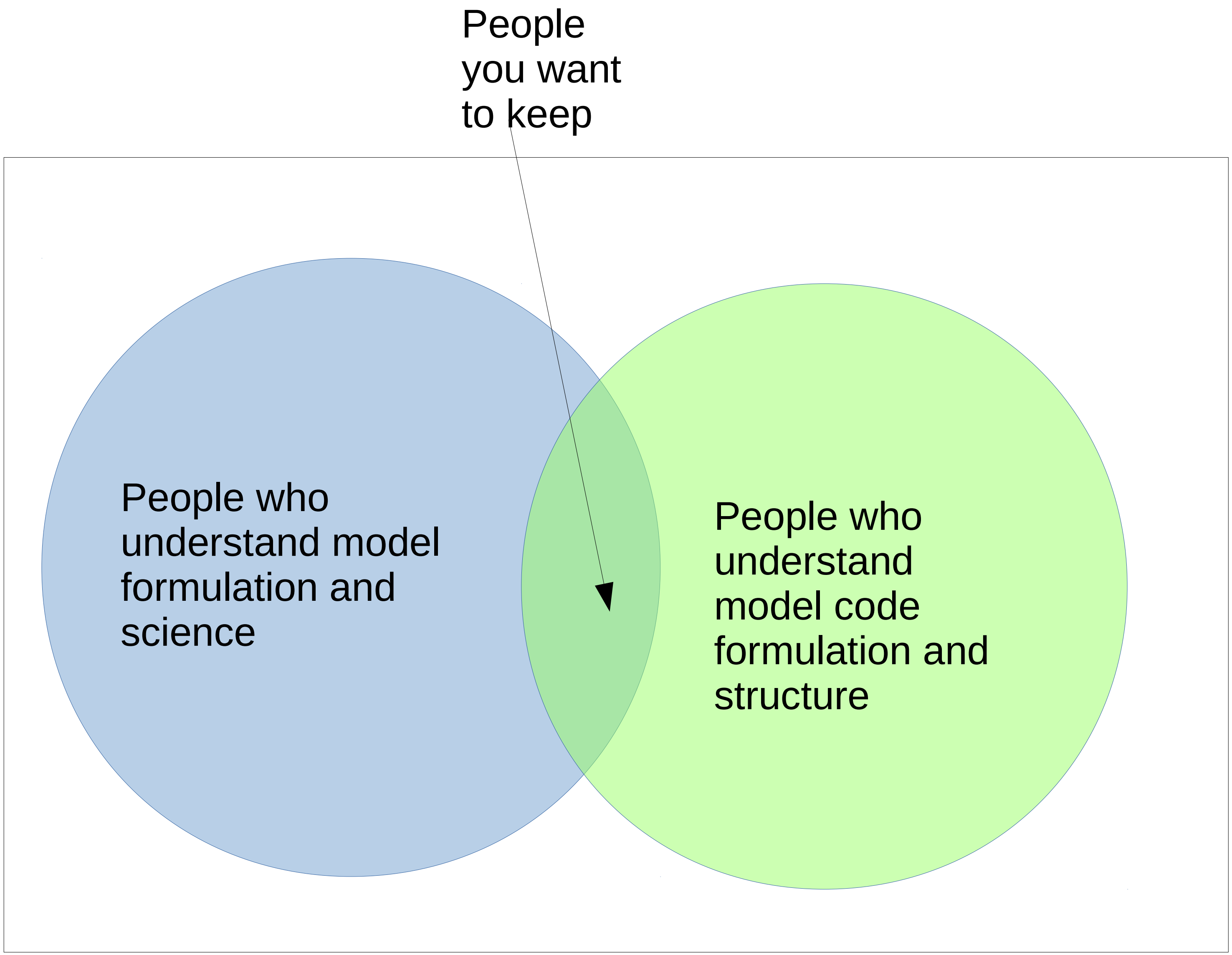
Comments

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

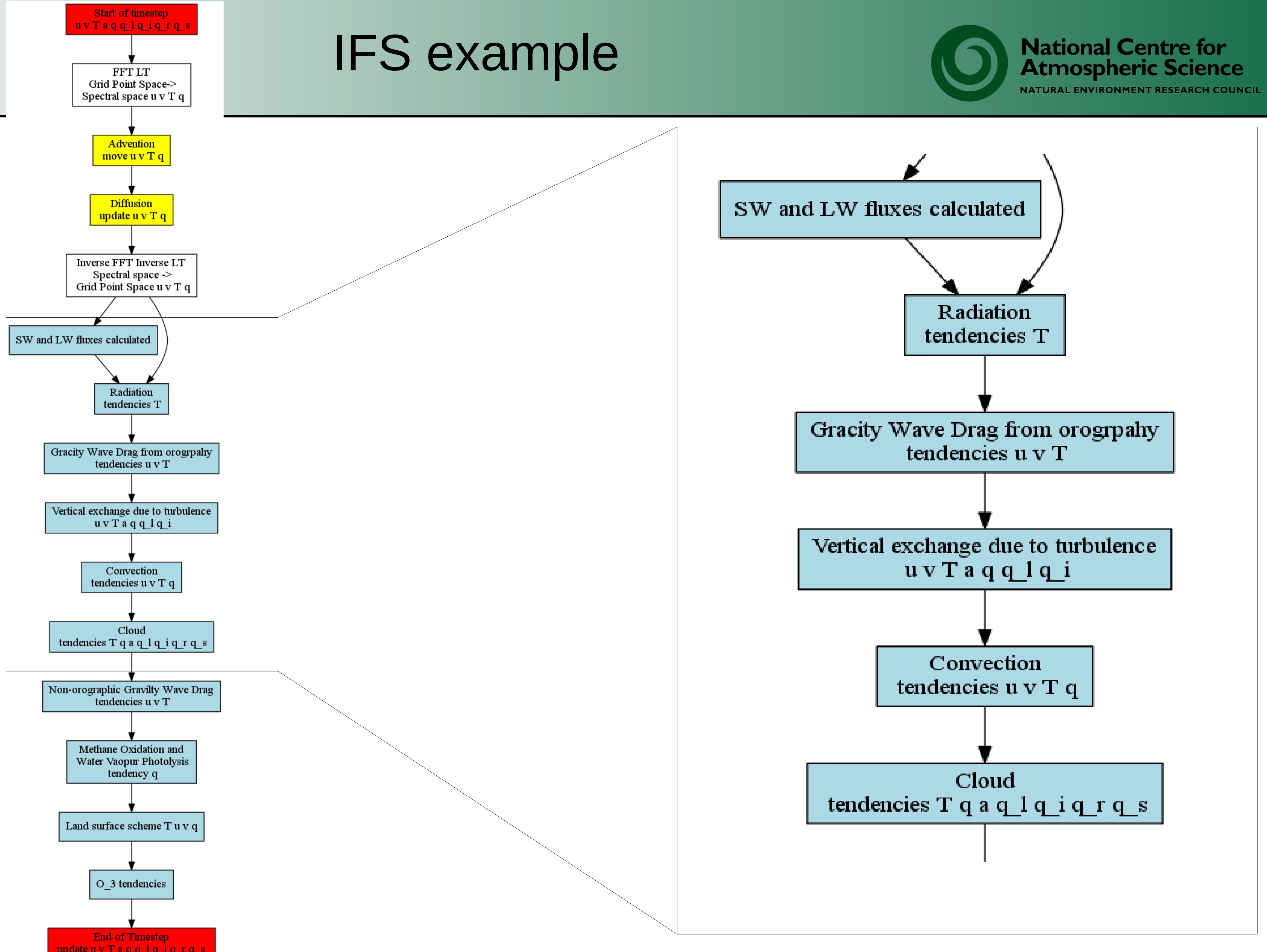
- Demonstrated a consistent way to analyse model structure directly from model code comments.
- Using the CIM.
- Very much a work in progress.
- Prognostic type (increments, deltas, timestep etc…) not defined.
- No grid info either.
- CIM construct population not fully instrumented.

Questions

A quick Venn Diagram



IFS example



Start of timestep
u v T a q q_l q_i q_r q_s
FFT LT
Grid Point Space->
Spectral space u v T q
Advention
move u v T q
Diffusion
update u v T q
Inverse FFT Inverse LT
Spectral space ->
Grid Point Space u v T q
SW and LW fluxes calculated
Radiation
tendencies T
Gracity Wave Drag from orogrpahy
tendencies u v T
Vertical exchange due to turbulence
u v T a q q_l q_i
Convection
tendencies u v T q
Cloud
tendencies T q a q_l q_i q_r q_s
Non-orographic Gravilty Wave Drag
tendencies u v T
Methane Oxidation and
Water Vaopur Photolysis
tendency q
Land surface scheme T u v q
O_3 tendencies
End of Timestep
update u v T a q q_l q_i q_r q_s
SW and LW fluxes calculated
Radiation
tendencies T
Gracity Wave Drag from orogrpahy
tendencies u v T
Vertical exchange due to turbulence
u v T a q q_l q_i
Convection
tendencies u v T q
Cloud
tendencies T q a q_l q_i q_r q_s