Common Community Physics Package Single Column Model (SCM)

User and Technical Guide
v4.0

March 2020

Grant Firl, Laurie Carson, Michelle Harrold

National Center for Atmospheric Research and Developmental Testbed Center

Ligia Bernardet

NOAA/ESRL Global Systems Laboratory and Developmental Testbed Center

Dom Heinzeller

NOAA/ESRL Global Systems Laboratory, Developmental Testbed Center and CIRES/CU

Acknowledgement

If significant help was provided via the helpdesk for work resulting in a publication, please acknowledge the Developmental Testbed Center team.

For referencing this document please use:

Firl, G., L. Carson, L. Bernardet, D. Heinzeller, and M. Harrold, 2020. Common Community Physics Package Single Column Model v4.0 User and Technical Guide. 38pp. Available at https://dtcenter.org/GMTB/v4.0/scm-ccpp-guide-v4.pdf

Contents

Preface v

1 Introduction 1
- 1.1 Version Notes . . . 1
  - 1.1.1 Limitations . . . 2

2 Quick Start Guide 3
- 2.1 Obtaining Code . . . 3
  - 2.1.1 Release Code . . . 3
  - 2.1.2 Development Code . . . 4
- 2.2 System Requirements, Libraries, and Tools . . . 4
  - 2.2.1 Compilers . . . 5
  - 2.2.2 Installing Libraries on Supported Platforms . . . 6
  - 2.2.3 Using Existing Libraries on Preconfigured Platforms . . . 7
- 2.3 Compiling SCM with CCPP . . . 8
- 2.4 Run the SCM with a supplied case . . . 9
  - 2.4.1 Single Run Script Usage . . . 9
  - 2.4.2 Multiple Run Script Usage . . . 11
  - 2.4.3 Batch Run Script . . . 12
- 2.5 Creating and Using a Docker Container with SCM and CCPP . . . 12
  - 2.5.1 Building the Docker image . . . 13
  - 2.5.2 Running the Docker image . . . 13

3 Repository 16
- 3.1 What is included in the repository? . . . 16

4 Algorithm 18
- 4.1 Algorithm Overview . . . 18
- 4.2 Reading input . . . 18
- 4.3 Setting up vertical grid and interpolating input data . . . 19
- 4.4 Physics suite initialization . . . 19
- 4.5 Time integration . . . 20
- 4.6 Writing output . . . 20

5 Cases 22
- 5.1 How to run cases . . . 22
  - 5.1.1 Case configuration namelist parameters . . . 22
  - 5.1.2 Case input data file . . . 24
- 5.2 Included Cases . . . 26
- 5.3 How to set up new cases . . . 27

5.4 Using other LASSO cases . . . 29
5.5 Using UFS Initial Conditions . . . 29

6 CCPP Interface **32**
6.1 Setting up a suite . . . 32
6.1.1 Preparing data from the SCM . . . 32
6.1.2 Editing and running `ccpp_prebuild.py` . . . 33
6.1.3 Preparing a suite definition file . . . 33
6.2 Initializing/running a suite . . . 34
6.3 Changing a suite . . . 34
6.3.1 Replacing a scheme with another . . . 34
6.3.2 Modifying "groups" of parameterizations . . . 35
6.3.3 Subcycling parameterizations . . . 35
6.4 Adding variables . . . 35
6.4.1 Adding a physics-only variable . . . 35
6.4.2 Adding a prognostic SCM variable . . . 37

Preface

Meaning of typographic changes and symbols

Table 1 describes the type changes and symbols used in this book.

Typeface or Symbol	Meaning	Example
`AaBbCc123`	The names of commands, files, and directories; on-screen computer output	Edit your `.bashrc` Use `ls -a` to list all files. `host$ You have mail!`.
`AaBbCc123`	What you type, contrasted with on-screen computer output	`host$ su`
`AaBbCc123`	Command line placeholder: replace with a real name or value	To delete a file, type `rm` *`filename`*

Table 1: Typographic Conventions

1 Introduction

A single column model (SCM) can be a valuable tool for diagnosing the performance of a physics suite, from validating that schemes have been integrated into a suite correctly to deep dives into how physical processes are being represented by the approximating code. This SCM has the advantage of working with the Common Community Physics Package (CCPP), a library of physical parameterizations for atmospheric numerical models and the associated framework for connecting potentially any atmospheric model to physics suites constructed from its member parameterizations. In fact, this SCM serves as perhaps the simplest example for using the CCPP and its framework in an atmospheric model. This version contains all parameterizations of NOAA's evolved operational GFS v15.2 suite (implemented in 2019), plus additional developmental schemes. The schemes are grouped in four supported suites described in detail in the CCPP Scientific Documentation (GFS_v15p2, GFS_v16beta, csawmg, and GSD_v1). Two additional suites without the near sea surface temperature scheme are available to match the first Unified Forecast System (UFS) public release.

This document serves as both the User and Technical Guides for this model. It contains a Quick Start Guide with instructions for obtaining the code, compiling, and running a sample test case, an explanation for what is included in the repository, a brief description of the operation of the model, a description of how cases are set up and run, and finally, an explanation for how the model interfaces with physics through the CCPP infrastructure.

Please refer to the release web page for further documentation and user notes:
https://dtcenter.org/community-code/common-community-physics-package-ccpp/download

1.1 Version Notes

The CCPP SCM v4.0 contains the following major and minor changes since v3.0.

Major

- Codebase updated to work with latest ccpp-framework (v4.0) and ccpp-physics (v4.0), allowing use of 6 supported physics suites
- Support added for NOAA's Hera HPC platform and Docker containers
- Support added for CCPP's "static" build. This release no longer supports the "dynamic" build.

- Added capability to utilize UFS initial conditions and initialize the Noah and NoahMP LSMs (without advective forcing)

Minor

- Integrated CCPP code generation with the CMake step
- Adopted the new CCPP metadata format
- Name change: this code is now known as the CCPP SCM instead of the "GMTB" SCM, although filenames in the code have not been changed yet.

1.1.1 Limitations

This release bundle has some known limitations:

- The provided cases over land points cannot use an LSM at this time due to the lack of initialization data for the LSMs. Therefore, for the provided cases over land points (ARM_SGP_summer_1997_* and LASSO_*, where sfc_type = 1 is set in the case configuration file), prescribed surface fluxes must be used:
  - surface sensible and latent heat fluxes must be provided in the case data file
  - sfc_flux_spec must be set to true in the case configuration file
  - the surface roughness length in cm must be set in the case configuration file
  - the suite defintion file used (physics_suite variable in the case configuration file) must have been modified to use prescribed surface fluxes rather than an LSM. An example is included in the development repository
  - NOTE: If one can develop appropriate initial conditions for the LSMs for the supplied cases over land points, there should be no technical reason why they cannot be used with LSMs, however.
- As of this release, using the SCM over a land point with an LSM is possible through the use of UFS initial conditions (see section 5.5). However, advective forcing terms are unavailable as of this release, so only short integrations using this configuration should be employed. Using dynamical tendencies (advective forcing terms) from the UFS will be part of a future release.

2 Quick Start Guide

This chapter provides instructions for obtaining and compiling the CCPP SCM. The SCM code calls CCPP-compliant physics schemes through the CCPP framework code. As such, it requires the CCPP framework code and physics code, both of which are included as submodules within the SCM code. This package can be considered a simple example for an atmospheric model to interact with physics through the CCPP.

Alternatively, if one doesn't have access or care to set up a machine with the appropriate system requirements but has a working Docker installation, it is possible to create and use a Docker container with a pre-configured computing environment with a pre-compiled model. This is also an avenue for running this software with a Windows PC. See section 2.5 for more information.

2.1 Obtaining Code

The source code for the CCPP and SCM is provided through GitHub.com. This release branch contains the tested and supported version for general use, while a development branch is less stable, yet contains the latest developer code. Instructions for using either option are discussed here.

2.1.1 Release Code

Clone the source using

```
git clone --recursive -b v4.0.0 https://github.com/NCAR/gmtb-scm
```

Recall that the `recursive` option in this command clones the main gmtb-scm repository and all subrepositories (ccpp-physics and ccpp-framework). Using this option, there is no need to execute **`git submodule init`** and **`git submodule update`**.

The CCPP framework can be found in the ccpp/framework subdirectory at this level. The CCPP physics parameterizations can be found in the ccpp/physics subdirectory.

2.1.2 Development Code

If you would like to contribute as a developer to this project, please see (in addition to the rest of this guide) the scientific and technical documentation included with this release:

https://dtcenter.org/community-code/common-community-physics-package-ccpp/documentation

There you will find links to all of the documentation pertinent to developers.

For working with the development branches (stability not guaranteed), check out the **dtc/develop** branches of the repository (and submodules):

```
git clone --recursive -b dtc/develop https://github.com/NCAR/gmtb-scm
```

You may want to double-check that the dtc/develop branch of the SCM is pointing to the latest commits of the dtc/develop branches of ccpp-physics and ccpp-framework. While we update the submodule pointers often, it is occasionally forgotten. To ensure that you have the latest development code for the submodules, execute the following:

1. Navigate to the ccpp-physics directory.

   ```
   cd gmtb-scm/ccpp/physics
   ```

2. Check out the right branch (cloning recursively as instructed above creates a "detached head" state for the submodules by default).

   ```
   git checkout dtc/develop
   ```

3. Pull down the latest changes just to be sure.

   ```
   git pull
   ```

4. Do the same for ccpp-framework

   ```
   cd ../framework
   git checkout dtc/develop
   git pull
   ```

5. Change back to the main directory for following the instructions in section 2.3 assuming system requirements in section 2.2 are met.

   ```
   cd ../..
   ```

2.2 System Requirements, Libraries, and Tools

The source code for the SCM and CCPP component is in the form of programs written in FORTRAN, FORTRAN 90, and C. In addition, the I/O relies on the netCDF libraries. Beyond the standard scripts, the build system relies on use of the Python scripting language, along with cmake, GNU make and date.

The basic requirements for building and running the CCPP and SCM bundle are listed below. The versions listed reflect successful tests and there is no guarantee that the code will work with different versions.

- FORTRAN 90+ compiler
  - ifort 18.0.5.274, 19.0.2 and 19.0.5
  - gfortran 6.2, 8.3, and 9.2
- C compiler
  - icc 18.0.5.274, 19.0.2 and 19.0.5
  - gcc 6.2, 8.3, and 9.2
  - Apple clang 11.0.0.11000033, LLVM clang 9.0.0
- cmake 2.8.12.1, 2.8.12.2, 3.6.2, 3.16.3, 3.16.4
  - NOTE: Version 3.15+ is required if installing NCEPLIBS
- netCDF 4.3.0, 4.4.0, 4.4.1.1, 4.5.0, 4.6.1, 4.6.3, 4.7.0, 4.7.3 (not 3.x) with HDF5 and ZLIB
- Python 2.7.5, 2.7.9, 2.7.13, and 2.7.16 (not 3.x) with f90nml module (and Shapely if using the `UFS_IC_generator.py` script)

Because these tools and libraries are typically the purview of system administrators to install and maintain, they are considered part of the basic system requirements. The Unified Forecast System (UFS) release v1.0.0 of March 11, 2020, provides software packages and detailed instructions to install these prerequisites and the NCEPlibs on supported platforms (see section 2.2.2).

Further, there are several utility libraries as part of the NCEPlibs package that must be installed with environment variables pointing to their locations prior to building the SCM.

- bacio - Binary I/O Library
- sp - Spectral Transformation Library
- w3nco - GRIB decoder and encoder library

The following environment variables are used by the build system to properly link these libraries: `BACIO_LIB4`, `SP_LIBd`, and `W3NCO_LIBd`. These libraries are prebuilt on most NOAA machines using the Intel compiler and on Cheyenne for the Intel and GNU compilers. The machine setup scripts mentioned in section 2.3 load these libraries (which are identical to those used by the UFS Medium Range Weather Application on those machines) and set these environment variables for the user automatically. For installing the libraries and its prerequisites on supported platforms, existing UFS packages can be used (see section 2.2.2).

2.2.1 Compilers

The CCPP and SCM have been tested on a variety of computing platforms. Currently the CCPP system is actively supported on Linux and MacOS computing platforms using the Intel or GNU Fortran compilers. Please use versions listed in

the previous section as unforeseen build issues may occur when using older compiler versions. Typically the best results come from using the most recent version of a compiler. If you have problems with compilers, please check the "Known Issues" section of the release website (https://dtcenter.org/community-code/common-community-physics-package-ccpp/download).

2.2.2 Installing Libraries on Supported Platforms

For users on supported platforms such as generic Linux or macOS systems, the UFS v1.0.0 release provides software packages and detailed setup instructions at https://github.com/NOAA-EMC/NCEPLIBS-external/tree/ufs-v1.0.0 and https://github.com/NOAA-EMC/NCEPLIBS/tree/ufs-v1.0.0. UFS users who already installed the `NCEPLIBS` package only need to set the compiler environment variables as indicated in https://github.com/NOAA-EMC/NCEPLIBS-external/tree/ufs-v1.0.0/doc/README_*.txt and source the shell script that is created by the `NCEPLIBS` install process to set the required environment variables for compiling the SCM.

The SCM uses only a small part of the UFS `NCEPLIBS` package and has fewer prerequisites (i.e. no `ESMF` or `wgrib2` needed). Users who are not planning to use the UFS can follow the machine setup instructions in https://github.com/NOAA-EMC/NCEPLIBS-external/tree/ufs-v1.0.0/doc/README_*.txt and, instead of installing `NCEPLIBS-external` and `NCEPLIBS`, install only NetCDF/NetCDF-Fortran manually or using the software package manager (`apt`, `yum`, `brew`).

Note. On macOS systems, it may be necessary to add the (future) location of the NetCDF libraries `libnetcdf.dylib` and `libnetcdff.dylib` to the `rpath` linker flags before compiling the NetCDF/NetCDF-Fortran libraries. Execute the following command before running `configure` and `make` for `netcdf-c` and `netcdf-fortran`:

```
export LDFLAGS="-L/dir/where/libnetcdf/and/libnetcdff/live -Wl,-rpath,/
   dir/where/libnetcdf/and/libnetcdff/live"
```

Users need to set the compiler enviroment variables `CC`, `CXX`, `FC` and the environment variable `NETCDF` for compiling the three NCEP libraries (instead of the `NCEPLIBS` umbrella build referred to in the `NCEPLIBS-external` instructions) and the SCM.

Installing the NCEP libraries: The SCM repository contains a bash installation script in `gmtb-scm/contrib/build_nceplibs.sh` that will fetch the source code of the three required NCEP libraries from their authoritative repositories on GitHub and install them locally for the SCM to use. To execute this script, perform the following step from the top level directory (`gmtb-scm`).

```
./contrib/build_nceplibs.sh /path/to/nceplibs
```

Following successful execution of this script, the commands to set the proper environment variables mentioned above will be written to the terminal as output. One must execute the correct set for the active shell to finish the installation, e.g., for bash

```
export BACIO_LIB4=/path/to/nceplibs/lib/libbacio_v2.2.0_4.a
export SP_LIBd=/path/to/nceplibs/lib/libsp_v2.1.0_d.a
export W3NCO_LIBd=/path/to/nceplibs/lib/libw3nco_v2.1.0_d.a
```

and for t/csh

```
setenv BACIO_LIB4 /path/to/nceplibs/lib/libbacio_v2.2.0_4.a
setenv SP_LIBd /path/to/nceplibs/lib/libsp_v2.1.0_d.a
setenv W3NCO_LIBd /path/to/nceplibs/lib/libw3nco_v2.1.0_d.a
```

The installation of NCEPLIBS requires `cmake` v3.15+. There are many ways to obtain the required version, either by following instructions provided by `cmake` (https://cmake.org/install/), or by following the instructions provided for the UFS release (https://github.com/NOAA-EMC/NCEPLIBS-external/tree/ufs-v1.0.0). Prepend this installation directory of `cmake` to your path environment variable to use it for building the NCEPLIBS.

The Python environment must provide the `f90nml` module for the SCM scripts to function. Users can test if f90nml is installed using this command in the shell:

```
python -c "import f90nml"
```

If `f90nml` is installed, this command will succeed silently, otherwise an `ImportError: No module named f90nml` will be printed to screen. To install the `f90nml` Python module, use the install method preferred for your Python environment (`easy_install`, `pip`, `conda`, ...) or perform the following steps to install it manually:

```
cd scm/etc/scripts/f90nml-0.19/
python setup.py build
python setup.py install --prefix=/my/install/directory
```

The directory `/my/install/directory` must exist and its subdirectory `/my/install/directory/lib/python2.7/site-packages` (or `lib64` instead of `lib`, depending on the system) must be in the `PYTHONPATH` environment variable.

2.2.3 Using Existing Libraries on Preconfigured Platforms

Platform-specific scripts are provided to load modules and set the user environment for preconfigured platforms: These script loads compiler modules (Fortran 2003-compliant), netCDF module, python environment, etc. and set compiler and NCEPlibs environment variables. From the top-level code directory (`gmtb-scm` by default), source the correct script for your platform and shell. For *t/csh* shells,

```
source scm/etc/Hera_setup_intel.csh
source scm/etc/Cheyenne_setup_gnu.csh
source scm/etc/Cheyenne_setup_intel.csh
```

For bourne/bash shells,

```
. scm/etc/Hera_setup_intel.sh
. scm/etc/Cheyenne_setup_gnu.sh
. scm/etc/Cheyenne_setup_intel.sh
```

2.3 Compiling SCM with CCPP

The first step in compiling the CCPP and SCM is to properly setup your user environment as described in sections 2.2.2 and 2.2.3. The second step is to download the lookup tables (large binaries, 324 MB) for the Thompson microphysics package and place them in the correct directory: From the top-level code directory (`gmtb-scm` by default), execute the following script:

```
./contrib/get_thompson_tables.sh
```

If the download step fails, make sure that your system's firewall does not block access to GitHub. If it does, download the file `thompson_tables.tar` from the GitHub release website using your browser and manually extract its contents in the directory `scm/data/physics_input_data/`.

Following this step, the top level build system will use `cmake` to query system parameters, execute the CCPP prebuild script to match the physics variables (between what the host model – SCM – can provide and what is needed by physics schemes in the CCPP), and build the physics caps needed to use them. Finally, `make` is used to compile the components.

1. From the top-level code directory (`gmtb-scm` by default), change directory to the top-level SCM directory.

   ```
   cd scm
   ```

2. Make a build directory and change into it.

   ```
   mkdir bin && cd bin
   ```

3. Invoke **`cmake`** on the source code to build using one of the options below.
   - Default mode

     ```
     cmake ../src
     ```

   - The statements above can be modified with the following options (put before `../src`):
     - Use threading with openmp (not for macOS with clang+gfortran)

       ```
       -DOPENMP=ON
       ```

     - Debug mode

       ```
       -DCMAKE_BUILD_TYPE=Debug
       ```

   CMake automatically runs the CCPP prebuild script to match required physics variables with those available from the dycore (SCM) and to generate physics caps and makefile segments. It generates software caps for each physics group defined in the supplied SDFs and generates a static library that becomes part of the SCM

executable. Appropriate software caps **will be generated for all suites defined in the `gmtb-scm/ccpp/suites` directory automatically.**
If necessary, the CCPP prebuild script can be executed manually from the top level directory (`gmtb-scm`). The basic syntax is

```
./ccpp/framework/scripts/ccpp_prebuild.py --config=./ccpp/config/
   ccpp_prebuild_config.py --static --suites=SCM_GFS_v15p2,
   SCM_GFS_v16beta,SCM_GSD_v1[...] --builddir=./scm/bin [--debug]
```

where the argument supplied via the `--suites` variable is a comma-separated list of suite names that exist in the `./ccpp/suites` directory. Note that suite names are the suite definition filenames minus the **`suite_`** prefix and **`.xml`** suffix.

4. Compile. Add `VERBOSE=1` to obtain more information on the build process.

```
make
```

The resulting executable may be found at `./gmtb-scm` (Full path of `gmtb-scm/scm/bin/gmtb-scm`).

Although `make clean` is not currently implemented, an out-of-source build is used, so all that is required to clean the build/run directory is (from the `bin` directory)

```
pwd #confirm that you are in the gmtb-scm/scm/bin directory before
   deleting files
rm -rfd *
```

Note: This command can be dangerous (deletes files without confirming), so make sure that you're in the right directory before executing!

If you encounter errors, please capture a log file from all of the steps, and contact the helpdesk at: gmtb-help@ucar.edu

2.4 Run the SCM with a supplied case

There are several test cases provided with this version of the SCM. For all cases, the SCM will go through the time steps, applying forcing and calling the physics defined in the chosen suite definition file using physics configuration options from an associated namelist. The model is executed through one of two Python run scripts that are pre-staged into the `bin` directory: `run_gmtb_scm.py` or `multi_run_gmtb_scm.py`. The first sets up and runs one integration while the latter will set up and run several integrations serially.

2.4.1 Single Run Script Usage

Running a case requires three pieces of information: the case to run (consisting of initial conditions, geolocation, forcing data, etc.), the physics suite to use (through a CCPP suite

definition file), and a physics namelist (that specifies configurable physics options to use). As discussed in chapter 5, cases are set up via their own namelists in `../etc/case_config`. A default physics suite is provided as a user-editable variable in the script and default namelists are associated with each physics suite (through `../src/default_namelists.py`), so, technically, one must only specify a case to run with the SCM. The single run script's interface is described below.

```
./run_gmtb_scm.py -c CASE_NAME [-s SUITE_NAME] [-n
   PHYSICS_NAMELIST_WITH_PATH] [-g] [-d]
```

When invoking the run script, the only required argument is the name of the case to run. The case name used must match one of the case configuration files located in `../etc/case_config` (*without the .nml extension!*). If specifying a suite other than the default, the suite name used must match the value of the suite name in one of the suite definition files located in `../../ccpp/suites` (Note: not the filename of the suite definition file). As part of the fourth CCPP release, the following suite names are valid:

1. SCM_GFS_v15p2
2. SCM_GFS_v16beta
3. SCM_GFS_v15p2_no_nsst
4. SCM_GFS_v16beta_no_nsst
5. SCM_csawmg
6. SCM_GSD_v1

Note that using the Thompson microphysics scheme (as in SCM_GSD_v1) requires the computation of look-up tables during its initialization phase. As of the release, this process has been prohibitively slow with this model, so it is HIGHLY suggested that these look-up tables are downloaded and staged to use this scheme (and the SCM_GSD_v1 suite) as described in section 2.3.

Also note that some cases require specified surface fluxes. Special suite definition files that correspond to the suites listed above have been created and use the `*_prescribed_surface` decoration. It is not necessary to specify this filename decoration when specifying the suite name. If the `spec_sfc_flux` variable in the configuration file of the case being run is set to `.true.`, the run script will automatically use the special suite definition file that corresponds to the chosen suite from the list above.

If specifying a namelist other than the default, the value must be an entire filename that exists in `../../ccpp/physics_namelists`. Caution should be exercised when modifying physics namelists since some redundancy between flags to control some physics parameterizations and scheme entries in the CCPP suite definition files currently exists. Values of numerical parameters are typically OK to change without fear of inconsistencies. Lastly, the `-g` flag can be used to run the executable through the **`gdb`** debugger (assuming it is installed on the system), and the `-d` flag is required when running this command in a Docker container in order to successfully mount a volume between the host machine and the Docker container instance and to share the output and plots with the host machine.

A netCDF output file is generated in the location specified in the case configuration file, if the `output_dir` variable exists in that file. Otherwise an output directory is constructed

from the case, suite, and namelist used (if different from the default). All output directories are placed in the `bin` directory. If using a Docker container, all output is copied to the `/home` directory in container space for volume-mounting purposes. Any standard netCDF file viewing or analysis tools may be used to examine the output file (ncdump, ncview, NCL, etc).

2.4.2 Multiple Run Script Usage

A second Python script is provided for automating the execution of multiple integrations through repeated calling of the single run script. From the run directory, one may use this script through the following interface.

```
./multi_run_gmtb_scm.py {[-c CASE_NAME] [-s SUITE_NAME] [-f
   PATH_TO_FILE]} [-v{v}] [-t] [-d]
```

No arguments are required for this script. The `-c or -case`, `-s or -suite`, or `-f or -file` options form a mutually-exclusive group, so exactly one of these is allowed at one time. If `-c` is specified with a case name, the script will run a set of integrations for all supported suites (defined in `../src/supported_suites.py`) for that case. If `-s` is specified with a suite name, the script will run a set of integrations for all supported cases (defined in `../src/supported_cases.py`) for that that suite. If `-f` is specified with the path to a filename, it will read in lists of cases, suites, and namelists to use from that file. An example for this file's syntax can be found in `../src/example_multi_run.py`. If multiple namelists are specified in the file, there either must be one suite specified *or* the number of suites must match the number of namelists. If none of the `-c or -case`, `-s or -suite`, or `-f or -file` options group is specified, the script will run through all permutations of supported cases and suites (as defined in the files previously mentioned).

In addition to the main options, some helper options can also be used with any of those above. The `-vv or -verbose` option can be used to output more information from the script to the console and to a log file. If this option is not used, only completion progress messages are written out. If one `-v` is used, the script will write out completion progress messages and all messages and output from the single run script. If two `-vv` are used, the script will also write out all messages and single run script output to a log file (`multi_run_gmtb_scm.log`) in the `bin` directory. The option, `-t or -timer`, can be used to output the elapsed time for each integration executed by the script. Note that the execution time includes file operations performed by the single run script in addition to the execution of the underlying (Fortran) SCM executable. By default, this option will execute one integration of each subprocess. Since some variability is expected for each model run, if greater precision is required, the number of integrations for timing averaging can be set through an internal script variable. This option can be useful, for example, for getting a rough idea of relative computational expense of different physics suites. Finally, the `-d` flag is required when running this command in a Docker container in order to successfully mount a volume between the host machine and the Docker container instance and to share the output and plots with the host machine.

2.4.3 Batch Run Script

If using the model on HPC resources and significant amounts of processor time is anticipated for the experiments, it will likely be necessary to submit a job through the HPC's batch system. An example script has been included in the repository for running the model on Hera's batch system (SLURM). It is located in `gmtb-scm/scm/etc/gmtb_scm_slurm_example.py`. Edit the `job_name`, `account`, etc. to suit your needs and copy to the `bin` directory. The case name to be run is included in the `command` variable. To use, invoke

```
./gmtb_scm_slurm_example.py
```

from the `bin` directory.

Additional details regarding the SCM may be found in the remainder of this guide. More information on the CCPP can be found in the CCPP Technical Documentation available at https://ccpp-techdoc.readthedocs.io/en/v4.0/.

2.5 Creating and Using a Docker Container with SCM and CCPP

In order to run a precompiled version of the CCPP SCM in a container, Docker will need to be available on your machine. Please visit https://www.docker.com to download and install the version compatible with your system. Docker frequently releases updates to the software; it is recommended to apply all available updates. NOTE: In order to install Docker on your machine, you will be required to have root access privileges. More information about getting started can be found at https://docs.docker.com/get-started

The following tips were acquired during a recent installation of Docker on a machine with Windows 10 Home Edition. Further help should be obtained from your system administrator or, lacking other resources, an internet search.

- Windows 10 Home Edition does not support Docker Desktop due to lack of "Hyper-V" support, but does work with Docker Toolbox. See the installation guide (https://docs.docker.com/toolbox/toolbox_install_windows/).
- You may need to turn on your CPU's hardware virtualization capability through your system's BIOS.
- After a successful installation of Docker Toolbox, starting with Docker Quickstart may result in the following error even with virtualization correctly enabled: `This computer doesn't have VT-X/AMD-v enabled. Enabling it in the BIOS is mandatory`. We were able to bypass this error by opening a bash terminal installed with Docker Toolbox, navigating to the directory where it was installed, and executing the following command:

  ```
  docker-machine create default --virtualbox-no-vtx-check
  ```

2.5.1 Building the Docker image

The Dockerfile builds CCPP SCM v4.0 from source using the GNU compiler. A number of required codes are built and installed via the DTC-supported common community container. For reference, the common community container repository can be accessed here: https://github.com/NCAR/Common-Community-Container.

The CCPP SCM has a number of system requirements and necessary libraries and tools. Below is a list, including versions, used to create the the GNU-based Docker image:

- gfortran - 8.3.1
- gcc - 8.3.1
- cmake - 3.16.5
- netCDF - 4.6.2
- HDF5 - 1.10.4
- ZLIB - 1.2.7
- SZIP - 2.1.1
- Python - 2.7.5
- NCEPLIBS subset: bacio v2.2.0_4, sp v2.1.0_d, w3nco v2.1.0_d

A Docker image containing the SCM, CCPP, and its software prerequisites can be generated from the code in the software repository obtained by following section 2.1 by executing the following steps:

NOTE: Windows users can execute these steps in the terminal application that was installed as part of Docker Toolbox.

1. Navigate to the `gmtb-scm/docker` directory.
2. Run the `docker build` command to generate the Docker image, using the supplied Dockerfile.

   ```
   docker build -t ccpp-scm .
   ```

   Inspect the Dockerfile if you would like to see details for how the image is built. The image will contain SCM prerequisite software from DTC, the SCM and CCPP code, and a pre-compiled executable for the SCM with the 6 supported suites for the SCM. A successful build will show two images: dtcenter/common-community-container, and ccpp-scm. To list images, type:

   ```
   docker images
   ```

2.5.2 Running the Docker image

NOTE: Windows users can execute these steps through the Docker Quickstart application installed with Docker Toolbox.

1. Set up a directory that will be shared between the host machine and the Docker

container. When set up correctly, it will contain output generated by the SCM within the container for manipulation by the host machine. For Mac/Linux,

```
mkdir -p /path/to/output
```

For Windows, you can try to create a directory of your choice to mount to the container, but it may not work or require more configuration, depending on your particular Docker installation. We have found that Docker volume mounting in Windows can be difficult to set up correctly. One method that worked for us was to create a new directory under our local user space, and specifying the volume mount as below. In addition, with Docker Toolbox, double check that the mounted directory has correct permissions. For example, open VirtualBox, right click on the running virtual machine, and choose "Settings". In the dialog that appears, make sure that the directory you're trying to share shows up in "Shared Folders" (and add it if it does not) and make sure that the "auto-mount" and "permanent" options are checked.

2. Set an environment variable to use for your SCM output directory. For *t/csh* shells,

   ```
   setenv OUT_DIR /path/to/output
   ```

   For bourne/bash shells,

   ```
   export OUT_DIR=/path/to/output
   ```

   For Windows, the format that worked for us followed this example: `/c/Users/my username/path/to/directory/to/mount`

3. To run the SCM, you can run the Docker container that was just created and give it the same run commands as discussed in sections 2.4.1 and 2.4.2. **Be sure to remember to include the `-d` option for all run commands**. For example,

   ```
   docker run --rm -it -v ${OUT_DIR}:/home --name run-ccpp-scm ccpp-
       scm ./run_gmtb_scm.py -c twpice -d
   ```

   will run through the TWPICE case using the default suite and namelist and put the output in the shared directory. NOTE: Windows users may need to omit the curly braces around environment variables: use `$OUT_DIR` instead of `${OUT_DIR}`. For running through all supported cases and suites, use

   ```
   docker run --rm -it -v ${OUT_DIR}:/home --name run-ccpp-scm ccpp-
       scm ./multi_run_gmtb_scm.py -d
   ```

   The options included in the above `run` commands are the following:
   - `--rm` removes the container when it exits
   - `-it` interactive mode with terminal access
   - `-v` specifies the volume mount from host directory (outside container) to inside the container. Using volumes allows you to share data between the host machine and container. For running the SCM, the output is being mounted from `/home` inside the container to the `OUT_DIR` on the host machine. Upon exiting the container, data mounted to the host machine will still be accessible.
   - `--name` names the container. If no name is provided, the daemon will auto-generate a random string name.

4. To use the SCM interactively, run non-default configurations, create plots, or even develop code, issue the following command:

   ```
   docker run --rm -it -v ${OUT_DIR}:/home --name run-ccpp-scm ccpp-
       scm /bin/bash
   ```

You will be placed within the container space and within the `bin` directory of the SCM with a pre-compiled executable. At this point, one could use the run scripts as described in previous sections (remembering to include the `-d` option on run scripts if output is to be shared with the host machine). To create plots, from within the `bin` directory of the SCM in container space, issue the following command, with an appropriately configured `*.ini` file, .i.e.

```
./gmtb_scm_analysis.py twpice_all_suites.ini -d
```

NOTE: If developing, since the container is ephemeral, one should push their changes to a remote git repository to save them (i.e. a fork on GitHub.com).

3 Repository

3.1 What is included in the repository?

The repository contains all code and data required to run the CCPP SCM (with the exception of large initialization tables for the Thompson microphysics scheme discussed in subsection 2.4.1). It is functionally separated into 3 subdirectories representing the SCM model infrastructure (`scm` directory), the CCPP infrastructure (`ccpp/framework` directory), and the CCPP physics schemes (`ccpp/physics` directory). The entire `gmtb-scm` repository resides on Github's NCAR space, and the `ccpp/framework` and `ccpp/physics` directories are git submodules that point to repositories `ccpp-framework` and `ccpp-physics` on the same space. The structure of the entire repository is represented below. Note that the `ccpp-physics` repository also contains files needed for using the CCPP with the UFS Atmosphere host model that uses the Finite-Volume Cubed-Sphere (FV3) dynamical core.

```
gmtb-scm/
└── ccpp/
    ├── config/ ..... contains the CCPP prebuild configuration file
    ├── framework/
    │   ├── cmake/ ..... custom cmake code for building ccpp-framework
    │   ├── CMakeLists.txt ..... cmake configuration file for ccpp-framework
    │   ├── CODEOWNERS ..... list of GitHub users with permission to merge
    │   ├── doc/ ... doxygen configuration, output, and Technical Documentation (obsolete)
    │   ├── LICENSE
    │   ├── README.md
    │   ├── schemes/ ..... contains schemes used for testing
    │   ├── scripts/ contains ccpp_prebuild and other Python scripts for parsing metadata
    │   ├── src/ ..... contains CCPP framework code
    │   ├── test/ ..... contains scripts and configurations for testing
    │   └── tests ..... contains next-generation files for testing
    ├── physics/ ..... contains all physics schemes
    │   ├── CMakeLists.txt ..... cmake configuration file for ccpp-physics
    │   ├── CODEOWNERS ..... list of GitHub users with permission to merge
    │   ├── LICENSE
    │   ├── pgifix.py
    │   ├── physics/ ..... contains all CCPP physics and interstitial schemes
    │   │   └── docs/ ..... contains CCPP physics doxygen documentation
    │   └── README.md
    └── physics_namelists ..... contains physics namelist files associated with suites
```

- `suites/` ..... contains suite definition files
- `CODEOWNERS` ..... list of GitHub users with permission to merge
- `contrib/`
  - `build_nceplibs.sh` ..... script for installing prerequisite NCEPLIBS locally
  - `get_thompson_tables.sh` . script for downloading/extracting the Thompson lookup tables
- `docker/`
  - `Dockerfile` ..... contains Docker instructions for building the CCPP SCM image
- `README.md`
- `scm/`
  - `bin/` ..... build directory (initially empty; populated by cmake)
  - `data/`
    - `comparison_data/` ..... contains data with which to compare SCM output
    - `physics_input_data/` ..... contains data needed by the CCPP physics
    - `processed_case_input/` ..... contains initialization and forcing data for cases
    - `raw_case_input/` ..... contains case data to be processed by scripts
    - `vert_coord_data/` ..... contains data to calculate vertical coordinates (from GSM-based GFS only)
  - `doc/` ..... contains this User's/Technical Guide
    - `TechGuide/` ..... contains LaTeX for this User's Guide
  - `etc/` ..... contains case configuration, machine setup scripts, and plotting scripts
    - `case_config/` ..... contains case configuration files
    - `CENTOS_docker_setup.sh` ..... contains machine setup for Docker container
    - `Cheyenne_setup_gnu.csh` ..... setup script for Cheyenne HPC for csh, tcsh
    - `Cheyenne_setup_gnu.sh` ..... setup script for Cheyenne HPC for sh, bash
    - `Cheyenne_setup_intel.csh` ..... setup script for Cheyenne HPC for csh, tcsh
    - `Cheyenne_setup_intel.sh` ..... setup script for Cheyenne HPC for sh, bash
    - `gmtb_scm_slurm_example.py` ..... example QSUB run script
    - `Hera_setup_intel.csh` ..... setup script for Theia HPC for csh, tcsh
    - `Hera_setup_intel.sh` ..... setup script for Theia HPC for sh, bash
    - `scripts/` ..... Python scripts for setting up cases and plotting
      - `f90nml.0.19/` ..... f90nml Python package
      - `Shapely-1.7.0` ..... Shapely Python package
      - `plot_configs/` ..... plot configuration files
  - `LICENSE.txt`
  - `src/` ..... source code for SCM infrastructure and Python run scripts

4 Algorithm

4.1 Algorithm Overview

Like most SCMs, the algorithm for the CCPP SCM is quite simple. In a nutshell, the SCM code performs the following:

- Read in an initial profile and the forcing data.
- Create a vertical grid and interpolate the initial profile and forcing data to it.
- Initialize the physics suite.
- Perform the time integration, applying forcing and calling the physics suite each time step.
- Output the state and physics data.

In this chapter, it will briefly be described how each of these tasks is performed.

4.2 Reading input

The following steps are performed at the beginning of program execution:

1. Call `get_config_nml()` in the `gmtb_scm_input` module to read in the `case_config` and `physics_config` namelists. This subroutine also sets some variables within the `scm_state` derived type from the data that was read.
2. Call `get_case_init()` in the `gmtb_scm_input` module to read in the case input data file. This subroutine also sets some variables within the `scm_input` derived type from the data that was read.
3. Call `get_reference_profile()` in the `gmtb_scm_input` module to read in the reference profile data. This subroutine also sets some variables within the `scm_reference` derived type from the data that was read. At this time, there is no "standard" format for the reference profile data file. There is a `select case` statement within the `get_reference_profile()` subroutine that reads in differently-formatted data. If adding a new reference profile, it will be required to add a section that reads its data in this subroutine.

4.3 Setting up vertical grid and interpolating input data

The CCPP SCM uses pressure for the vertical coordinate (lowest index is the surface). There are two choices for generating the vertical coordinate corresponding to a) the 2017 operational GFS v14 based on the Global Spectral Model (GSM) (set `model_name = 'GFS'` in the `case_config` file), and b) the FV3-based GFS v15 (set `model_name = 'FV3'` in the `case_config` file). For both methods, the pressure levels are calculated using the surface pressure and coefficients (a_k and b_k). For the GSM-based vertical coordinate, the coefficient data is read from an external file. Only 28, 42, 60, 64, and 91 levels are supported. If using the FV3-based vertical coordinate, it is possible to use potentially any (integer) number of vertical levels. Depending on the vertical levels specified, however, the method of specification of the coefficients may change. Please see the subroutine `get_FV3_vgrid` in the source file `gmtb-scm/scm/src/gmtb_scm_vgrid.F90` for details. This subroutine was minimally adapted from the source file `fv_eta.F90` from the v0 release version of the FV3GFS model.

After the vertical grid has been set up, the state variable profiles stored in the `scm_state` derived data type are interpolated from the input and reference profiles in the `set_state` subroutine of the `gmtb_scm_setup` module.

4.4 Physics suite initialization

With the CCPP framework, initializing a physics suite is a 5-step process:

1. Call `ccpp_init()` with the name of the suite and the CCPP derived data type (`cdata`) as arguments. This call will read and parse the suite definition file and initialize the `cdata` derived data type.
2. Initialize variables needed for the suite initialization routine. For suites originating from the GFS model, this involves setting some values in a derived data type used in the initialization subroutine. Call the suite initialization subroutine to perform suite initialization tasks that are not already performed in the `init` routines of the CCPP-compliant schemes (or associated initialization stages for groups or suites listed in the suite definition file). Note: As of this release, this step will require another suite intialization subroutine to be coded for a non-GFS-based suite to handle any initialization that is not already performed within CCPP-compliant scheme initialization routines.
3. Associate the `scm_state` variables with the appropriate pointers in the `physics` derived data type. Note: It is important that this step be performed before the next step to avoid segmentation faults.
4. Execute the `ccpp_field_add()` calls for the remaining variables to be used by the physics schemes. This step makes all physics variables that are exposed by the host application available to all physics schemes in the suite. This is done through an inclusion of an external file, `ccpp_fields.inc` that is automatically generated from the `ccpp_prebuild.py` script using the metadata contained in the host application cap (`gmtb-scm/scm/src/gmtb_scm_type_defs.f90` in the current implementation).

5. Call `ccpp_physics_init` with the `cdata` derived data type as input. This call executes the initialization stages of all schemes, groups, and suites that are defined in the suite definition file.

4.5 Time integration

Two time-stepping schemes have been implemented within the CCPP SCM: forward Euler (`time_scheme = 1` in the `case_config` namelist) and filtered leapfrog (`time_scheme = 2` in the `case_config` namelist). If the leapfrog scheme is chosen, two time levels of state variables are saved and the first time step is implemented as forward time step over $\Delta t/2$.

During each step of the time integration, the following sequence occurs:

1. Update the elapsed model time.
2. Calculate the current date and time given the initial date and time and the elapsed time.
3. If the leapfrog scheme is used, save the unfiltered model state from the previous time step.
4. Call the `interpolate_forcing()` subroutine in the `gmtb_scm_forcing` module to interpolate the forcing data in space and time.
5. Recalculate the pressure variables (pressure, Exner function, geopotential) in case the surface pressure has changed.
6. Call `do_time_step()` in the `gmtb_scm_time_integration` module. Within this subroutine:
   - Call the appropriate `apply_forcing_*` subroutine from the `gmtb_scm_forcing` module.
   - If using the leapfrog scheme, transfer the model state from one memory slot to the other.
   - For each column, call `ccpp_physics_run()` to call all physics schemes within the suite (this assumes that all suite parts are called sequentially without intervening code execution)
7. If using the leapfrog scheme, call `filter()` in the `gmtb_scm_time_integration` module to time filter the model state.
8. Check to see if output should be written during the current time step and call `output_append()` in the `gmtb_scm_output` module if necessary.

4.6 Writing output

As of this release, the SCM output is only instantaneous. Specifying an `output_frequency` in the case configuration file greater than the timestep will result in data loss. Prior to the physics suite being initialized, the `output_init()` subroutine in the `gmtb_scm_output` module is called to create the netCDF output file and define all dimensions and variables.

Immediately after the physics suite initialization and at the defined frequency within the time integration loop, the `output_append()` subroutine is called and instantaneous data values are appended to the netCDF file. Any variables defined in the `scm_state` and/or `physics` derived data types are accessible to the output subroutines. Writing new variables to the output involves hard-coding lines in the `output_init()` and `output_append()` subroutines.

5 Cases

5.1 How to run cases

Only two files are needed to set up and run a case with the SCM. The first is a configuration namelist file found in `gmtb-scm/scm/etc/case_config` that contains parameters for the SCM infrastructure. The second necessary file is a netCDF file containing data to initialize the column state and time-dependent data to force the column state. The two files are described below.

5.1.1 Case configuration namelist parameters

The `case_config` namelist expects the following parameters:

- `model_name`
  - This controls which vertical coordinates to use. Valid values are `'FV3'` or `'GFS'`. Here, `'GFS'` refers to vertical coordinates used in the GSM.
- `n_columns`
  - The code can be used to run a single column or multiple *independent* columns using the same or different physics suites. Specify an integer, `n`. NOTE: As of this release, only `n_columns` = 1 is supported.
- `case_name`
  - Identifier for which dataset (initialization and forcing) to load. This string must correspond to a dataset included in the directory `gmtb-scm/scm/data/processed_case_input/` (without the file extension).
- `dt`
  - Time step in seconds (floating point)
- `time_scheme`
  - Specify 1 for the forward-Euler time-stepping scheme or 2 for the filtered leapfrog scheme.
- `runtime`
  - Specify the model runtime in seconds (integer). This should correspond with the forcing dataset used. If a runtime is specified that is longer than the supplied forcing, the forcing is held constant at the last specified values.
- `output_frequency`
  - Specify the frequency of the model output in seconds (floating point). Currently, no averaging of the output fields is done if `output_frequency` $\neq$ `dt`; only instantaneous output at the supplied frequency is implemented.

- `n_levels`
  - Specify the integer number of vertical levels. If `model_name='GFS'`, only values of 28, 42, 60, 64, 91 are supported.
- `output_dir`
  - A string representing the path (relative to the build directory) to which output should be written. (OPTIONAL)
- `output_file`
  - A string representing the name of the netCDF output file to be written (no `.nc` extension expected).
- `case_data_dir`
  - A string representing the path (relative to the build directory) where case initialization and forcing data files can be found.
- `vert_coord_data_dir`
  - A string representing the path (relative to the build directory) where vertical coordinate data files can be found (for `model_name='GFS'` only).
- `thermo_forcing_type`
  - An integer representing how forcing for temperature and moisture state variables is applied (1 = total advective tendencies, 2 = horizontal advective tendencies with prescribed vertical motion, 3 = relaxation to observed profiles with vertical motion prescribed)
- `mom_forcing_type`
  - An integer representing how forcing for horizontal momentum state variables is applied (1 = total advective tendencies; not implemented yet, 2 = horizontal advective tendencies with prescribed vertical motion, 3 = relaxation to observed profiles with vertical motion prescribed)
- `relax_time`
  - A floating point number representing the timescale in seconds for the relaxation forcing (only used if `thermo_forcing_type` = `3` or `mom_forcing_type` = `3`)
- `sfc_flux_spec`
  - A boolean set to `.true.` if surface flux are specified from the forcing data (there is no need to have surface schemes in a suite definition file if so)
- `sfc_roughness_length_cm`
  - Surface roughness length in cm for calculating surface-related fields from specified surface fluxes (only used if `sfc_flux_spec` is True).
- `sfc_type`
  - An integer representing the character of the surface (0 = sea surface, 1 = land surface, 2 = sea-ice surface)
- `reference_profile_choice`
  - An integer representing the choice of reference profile to use above the supplied initialization and forcing data (1 = "McClatchey" profile, 2 = mid-latitude summer standard atmosphere)
- `year`
  - An integer representing the year of the initialization time
- `month`
  - An integer representing the month of the initialization time
- `day`
  - An integer representing the day of the initialization time
- `hour`

  - An integer representing the hour of the initialization time
- `column_area`
  - A list of floating point values representing the characteristic horizontal domain area of each atmospheric column in square meters (this could be analogous to a 3D model's horizontal grid size or the characteristic horizontal scale of an observation array; these values are used in scale-aware schemes; if using multiple columns, you may specify an equal number of column areas)
- `model_ics`
  - A boolean set to `.true.` if UFS atmosphere initial conditions are used rather than field campaign-based initial conditions
- `C_RES`
  - An integer representing the grid size of the UFS atmosphere initial conditions; the integer represents the number of grid points in each horizontal direction of each cube tile

5.1.2 Case input data file

The initialization and forcing data for each case is stored in a netCDF (version 4) file within the `gmtb-scm/scm/data/processed_case_input` directory. Each file has two dimensions (`time` and `levels`) and is organized into 3 groups: scalars, initial, and forcing. Not all fields are required for all cases. For example the fields `sh_flux_sfc` and `lh_flux_sfc` are only needed if the variable `sfc_flx_spec` = `.true.` in the case configuration file and state nudging variables are only required if `thermo_forcing_type` = `3` or `mom_forcing_type` = `3`. Using an active LSM (Noah, NoahMP, RUC) requires many more variables than are listed here. Example files for using with Noah and NoahMP LSMs are included in `gmtb-scm/scm/data/processed_case_input/fv3_model_point_noah[mp].nc`.

Listing 5.1: example netCDF file header for case initialization and forcing data

```
netcdf arm_sgp_summer_1997 {
dimensions:
 time = UNLIMITED ; // (233 currently)
 levels = UNLIMITED ; // (35 currently)
variables:
 float time(time) ;
  time:units = "s" ;
  time:description = "elapsed time since the beginning of the simulation" ;
 float levels(levels) ;
  levels:units = "Pa" ;
  levels:description = "pressure levels" ;

// global attributes:
  :description = "GMTB SCM forcing file for the ARM SGP Summer of 1997 case" ;

group: scalars {
  } // group scalars

group: initial {
  variables:
   float height(levels) ;
    height:units = "m" ;
    height:description = "physical height at pressure levels" ;
   float thetail(levels) ;
    thetail:units = "K" ;
    thetail:description = "initial profile of ice-liquid water potential temperature" ;
   float qt(levels) ;
    qt:units = "kg kg^-1" ;
```

```
    qt:description = "initial profile of total water specific humidity" ;
   float ql(levels) ;
    ql:units = "kg kg^-1" ;
    ql:description = "initial profile of liquid water specific humidity" ;
   float qi(levels) ;
    qi:units = "kg kg^-1" ;
    qi:description = "initial profile of ice water specific humidity" ;
   float u(levels) ;
    u:units = "m s^-1" ;
    u:description = "initial profile of E-W horizontal wind" ;
   float v(levels) ;
    v:units = "m s^-1" ;
    v:description = "initial profile of N-S horizontal wind" ;
   float tke(levels) ;
    tke:units = "m^2 s^-2" ;
    tke:description = "initial profile of turbulence kinetic energy" ;
   float ozone(levels) ;
    ozone:units = "kg kg^-1" ;
    ozone:description = "initial profile of ozone mass mixing ratio" ;
  } // group initial

group: forcing {
  variables:
   float lat(time) ;
    lat:units = "degrees N" ;
    lat:description = "latitude of column" ;
   float lon(time) ;
    lon:units = "degrees E" ;
    lon:description = "longitude of column" ;
   float p_surf(time) ;
    p_surf:units = "Pa" ;
    p_surf:description = "surface pressure" ;
   float T_surf(time) ;
    T_surf:units = "K" ;
    T_surf:description = "surface absolute temperature" ;
   float sh_flux_sfc(time) ;
    sh_flux_sfc:units = "K m s^-1" ;
    sh_flux_sfc:description = "surface sensible heat flux" ;
   float lh_flux_sfc(time) ;
    lh_flux_sfc:units = "kg kg^-1 m s^-1" ;
    lh_flux_sfc:description = "surface latent heat flux" ;
   float w_ls(levels, time) ;
    w_ls:units = "m s^-1" ;
    w_ls:description = "large scale vertical velocity" ;
   float omega(levels, time) ;
    omega:units = "Pa s^-1" ;
    omega:description = "large scale pressure vertical velocity" ;
   float u_g(levels, time) ;
    u_g:units = "m s^-1" ;
    u_g:description = "large scale geostrophic E-W wind" ;
   float v_g(levels, time) ;
    v_g:units = "m s^-1" ;
    v_g:description = "large scale geostrophic N-S wind" ;
   float u_nudge(levels, time) ;
    u_nudge:units = "m s^-1" ;
    u_nudge:description = "E-W wind to nudge toward" ;
   float v_nudge(levels, time) ;
    v_nudge:units = "m s^-1" ;
    v_nudge:description = "N-S wind to nudge toward" ;
   float T_nudge(levels, time) ;
    T_nudge:units = "K" ;
    T_nudge:description = "absolute temperature to nudge toward" ;
   float thil_nudge(levels, time) ;
    thil_nudge:units = "K" ;
    thil_nudge:description = "potential temperature to nudge toward" ;
   float qt_nudge(levels, time) ;
    qt_nudge:units = "kg kg^-1" ;
    qt_nudge:description = "q_t to nudge toward" ;
   float dT_dt_rad(levels, time) ;
    dT_dt_rad:units = "K s^-1" ;
    dT_dt_rad:description = "prescribed radiative heating rate" ;
   float h_advec_thetail(levels, time) ;
    h_advec_thetail:units = "K s^-1" ;
```

```
    h_advec_thetail:description = "prescribed theta_il tendency due to horizontal
        advection" ;
   float v_advec_thetail(levels, time) ;
    v_advec_thetail:units = "K s^-1" ;
    v_advec_thetail:description = "prescribed theta_il tendency due to vertical
        advection" ;
   float h_advec_qt(levels, time) ;
    h_advec_qt:units = "kg kg^-1 s^-1" ;
    h_advec_qt:description = "prescribed q_t tendency due to horizontal advection" ;
   float v_advec_qt(levels, time) ;
    v_advec_qt:units = "kg kg^-1 s^-1" ;
    v_advec_qt:description = "prescribed q_t tendency due to vertical advection" ;
  } // group forcing
}
```

5.2 Included Cases

Several cases are included in the repository to serve as examples for users to create their own and for basic research. All case configuration namelist files for included cases can be found in `gmtb-scm/scm/etc/case_config` and represent the following observational field campaigns:

- Tropical Warm Pool – International Cloud Experiment (TWP-ICE) maritime deep convection
- Atmospheric Radiation Measurement (ARM) Southern Great Plains (SGP) Summer 1997 continental deep convection
- Atlantic Stratocumulus Transition EXperiment (ASTEX) maritime stratocumulus-to-cumulus transition
- Barbados Oceanographic and Meteorological EXperiment (BOMEX) maritime shallow convection
- Large eddy simulation ARM Symbiotic Simulation and Observation (LASSO) for May 18, 2016 (with capability to run all LASSO dates - see 5.4) continental shallow convection

For the ARM SGP case, several case configuration files representing different time periods of the observational dataset are included, denoted by a trailing letter. The LASSO case may be run with different forcing applied, so three case configuration files corresponding to these different forcing are included. In addition, two example cases are included for using UFS Atmosphere initial conditions:

- UFS initial conditions for 38.1 N, 98.5 W (central Kansas) for 00Z on Oct. 3, 2016 with Noah variables on the C96 FV3 grid (`fv3_model_point_noah.nc`)
- UFS initial conditions for 38.1 N, 98.5 W (central Kansas) for 00Z on Oct. 3, 2016 with NoahMP variables on the C96 FV3 grid (`fv3_model_point_noahmp.nc`)

See 5.5 for information on how to generate these files for other locations and dates, given appropriate UFS Atmosphere initial conditions.

5.3 How to set up new cases

Setting up a new case involves preparing the two types of files listed above. For the case initialization and forcing data file, this typically involves writing a custom script or program to parse the data from its original format to the format that the SCM expects, listed above. An example of this type of script written in Python is included in `/gmtb-scm/scm/etc/scripts/twpice_forcing_file_generator.py`. The script reads in the data as supplied from its source, converts any necessary variables, and writes a netCDF (version 4) file in the format described in subsection 5.1.2. For reference, the following formulas are used:

$$\theta_{il} = \theta - \frac{\theta}{T}\left(\frac{L_v}{c_p}q_l + \frac{L_s}{c_p}q_i\right) \tag{5.1}$$

$$q_t = q_v + q_l + q_i \tag{5.2}$$

where θ_{il} is the ice-liquid water potential temperature, θ is the potential temperature, L_v is the latent heat of vaporization, L_s is the latent heat of sublimation c_p is the specific heat capacity of air at constant pressure, T is absolute temperature, q_t is the total water specific humidity, q_v is the water vapor specific humidity, q_l is the suspended liquid water specific humidity, and q_i is the suspended ice water specific humidity.

As shown in the example netCDF header, the SCM expects that the vertical dimension is pressure levels (index 1 is the surface) and the time dimension is in seconds. The initial conditions expected are the height of the pressure levels in meters, and arrays representing vertical columns of θ_{il} in K, q_t, q_l, and q_i in kg kg^{-1}, u and v in m s^{-1}, turbulence kinetic energy in m^2 s^{-2} and ozone mass mixing ratio in kg kg^{-1}.

For forcing data, the SCM expects a time series of the following variables: latitude and longitude in decimal degrees [in case the column(s) is moving in time (e.g., Lagrangian column)], the surface pressure (Pa) and surface temperature (K). If surface fluxes are specified for the new case, one must also include a time series of the kinematic surface sensible heat flux (K m s^{-1}) and kinematic surface latent heat flux (kg kg^{-1} m s^{-1}). The following variables are expected as 2-dimensional arrays (vertical levels first, time second): the geostrophic u (E-W) and v (N-S) winds (m s^{-1}), and the horizontal and vertical advective tendencies of θ_{il} (K s^{-1}) and q_t (kg kg^{-1} s^{-1}), the large scale vertical velocity (m s^{-1}), large scale pressure vertical velocity (Pa s^{-1}), the prescribed radiative heating rate (K s^{-1}), and profiles of u, v, T, θ_{il} and q_t to use for nudging.

Although it is expected that all variables are in the netCDF file, only those that are used with the chosen forcing method are required to be nonzero. For example, the following variables are required depending on the values of `mom_forcing_type` and `thermo_forcing_type` specified in the case configuration file:

- `mom_forcing_type` = `1`
  - Not implemented yet
- `mom_forcing_type` = `2`
  - geostrophic winds and large scale vertical velocity

- `mom_forcing_type` = `2`
  - u and v nudging profiles
- `thermo_forcing_type` = `1`
  - horizontal and vertical advective tendencies of θ_{il} and q_t and prescribed radiative heating (can be zero if radiation scheme is active)
- `thermo_forcing_type` = `2`
  - horizontal advective tendencies of θ_{il} and q_t, prescribed radiative heating (can be zero if radiation scheme is active), and the large scale vertical pressure velocity
- `thermo_forcing_type` = `2`
  - θ_{il} and q_t nudging profiles and the large scale vertical pressure velocity

For the case configuration file, it is most efficient to copy an existing file in `gmtb-scm/scm/etc/case_config` and edit it to suit one's case. Recall from subsection 5.1.1 that this file is used to configure the SCM framework parameters for a given case. Be sure to check that model timing parameters such as the time step and output frequency are appropriate for the physics suite being used. There is likely some stability criterion that governs the maximum time step based on the chosen parameterizations and number of vertical levels (grid spacing). The `case_name` parameter should match the name of the case input data file that was configured for the case (without the file extension). The `runtime` parameter should be less than or equal to the length of the forcing data unless the desired behavior of the simulation is to proceed with the last specified forcing values after the length of the forcing data has been surpassed. The initial date and time should fall within the forcing period specified in the case input data file. If the case input data is specified to a lower altitude than the vertical domain, the remainder of the column will be filled in with values from a reference profile. There is a tropical profile and mid-latitude summer profile provided, although one may add more choices by adding a data file to `gmtb-scm/scm/data/processed_case_input` and adding a parser section to the subroutine `get_reference_profile` in `gmtb-scm/scm/src/gmtb_scm_input.f90`. Surface fluxes can either be specified in the case input data file or calculated using a surface scheme using surface properties. If surface fluxes are specified from data, set `sfc_flux_spec` to `.true.` and specify `sfc_roughness_length_cm` for the surface over which the column resides. Otherwise, specify a `sfc_type`. In addition, one must specify a `column_area` for each column.

To control the forcing method, one must choose how the momentum and scalar variable forcing are applied. The three methods of Randall and Cripe (1999, JGR) have been implemented: "revealed forcing" where total (horizontal + vertical) advective tendencies are applied (type 1), "horizontal advective forcing" where horizontal advective tendencies are applied and vertical advective tendencies are calculated from a prescribed vertical velocity and the calculated (modeled) profiles (type 2), and "relaxation forcing" where nudging to observed profiles replaces horizontal advective forcing combined with vertical advective forcing from prescribed vertical velocity (type 3). If relaxation forcing is chosen, a `relaxation_time` that represents the timescale over which the profile would return to the nudging profiles must be specified.

5.4 Using other LASSO cases

In order to use other LASSO cases than the one provided, perform the following steps:

1. Access http://archive.arm.gov/lassobrowser and use the navigation on the left to choose the dates for which you would like to run a SCM simulation. Pay attention to the "Large Scale Forcing" tab where you can choose how the large scale forcing was generated, with options for ECMWF, MSDA, and VARANAL. All are potentially valid, and it is likely worth exploring the differences among forcing methods. Click on Submit to view a list of simulations for the selected criteria. Choose from the simulations (higher skill scores are preferred) and check the "Config Obs Model Tar" box to download the data. Once the desired simulations have been checked, order the data (you may need to create an ARM account to do so).
2. Once the data is downloaded, decompress it. From the `config` directory, copy the files `input_ls_forcing.nc`, `input_sfc_forcing.nc`, and `wrfinput_d01.nc` into their own directory under `gmtb-scm/scm/data/raw_case_input/`.
3. Modify `gmtb-scm/scm/etc/scripts/lasso1_forcing_file_generator_gjf.py` to point to the input files listed above. Execute the script in order to generate a case input file for the SCM (to be put in `gmtb-scm/scm/data/processed_case_input/`):

   ```
   ./lasso1_forcing_file_generator_gjf.py
   ```

4. Create a new case configuration file (or copy and modify an existing one) in `gmtb-scm/scm/etc/case_config`. Be sure that the `case_name` variable points to the newly created/processed case input file from above.

5.5 Using UFS Initial Conditions

A script exists in `scm/etc/scripts/UFS_IC_generator.py` to read in UFS Atmosphere cold start initial conditions and generate a case input data file that the SCM can use. Since the Noah LSM is the operational LSM, it is assumed that initial variables for it exist in the UFS Atmosphere initial condition files. Although NoahMP is not a member of any officially supported suite as of this release, if NoahMP is to be used, its initial conditions are generated from the Noah initial conditions using the same algorithm used in the UFS Atmosphere. Note that the script requires a few python packages that may not be found by default in all python installations: **`argparse`**, **`fnmatch`**, **`logging`**, **`netCDF4`**, **`numpy`**, **`shapely`**, **`f90nml`**, and **`re`**.

NOTE: If using NOAA's Hera HPC, the **`shapely`** python package does not seem to be installed with the version of Anaconda used by the rest of this software package. Users can test if **`shapely`** is installed using this command in the shell:

```
python -c "import shapely"
```

If **`shapely`** is installed, this command will succeed silently, otherwise an **`ImportError: No module named shapely`** will be printed to screen. To install the **`shapely`** Python module,

use the install method preferred for your Python environment (`easy_install`, `pip`, `conda`, ...) or perform the following steps to install it manually:

```
cd scm/etc/scripts/Shapely-1.7.0/
python setup.py build
python setup.py install --prefix=/my/install/directory
```

The directory `/my/install/directory` must exist and its subdirectory `/my/install/directory/lib/python2.7/site-packages` (or `lib64` instead of `lib`, depending on the system) must be in the `PYTHONPATH` environment variable.

The `UFS_IC_generator.py` script usage is as follows:

```
./UFS_IC_generator.py [-h] (-l LOCATION LOCATION | -ij INDEX INDEX) -d
DATE -i IN_DIR -g GRID_DIR [-t {1,2,3,4,5,6}]
[-a AREA] [-mp] -n CASE_NAME [-oc]
```

Mandatory arguments:

1. **`--location (-l)`** OR **`--index (-ij)`**: Either longitude and latitude in decimal degrees east and north of a location OR the UFS grid index with the tile number
   - -l 261.51 38.2 (two floating point values separated by a space)
   - -ij 8 49 (two integer values separated by a space; this option must also use the **`--tile (-t)`** argument to specify the tile number)
2. **`--date (-d)`** YYYYMMDDHHMM: date corresponding to the UFS initial conditions
3. **`--in_dir (-i)`**: path to the directory containing UFS initial conditions
4. **`--grid_dir (-g)`**: path to the directory containing the UFS supergrid files (AKA "fix" directory)
5. **`--case_name (-n)`**: what to call the output netCDF file

Optional arguments:

1. **`--tile (-t)`**: if one already knows the correct tile for the given longitude and latitude OR one is specifying the UFS grid index (**`--index`** argument)
2. **`--noahmp (-mp)`**: flag to generate cold-start initial conditions for NoahMP LSM from Noah LSM initial conditions
3. **`--area (-a)`**: area of grid cell in m^2 (if known or different than the value calculated from the supergrid file)
4. **`--old_chgres (-oc)`**: flag if UFS initial conditions were generated using older version of chgres (global_chgres); might be the case for pre-2018 data

The following commands were used from within the **`scm/etc/scripts`** directory to generate the example UFS Atmosphere initial condition case input files:

```
./UFS_IC_generator.py -l 261.51 38.2 -d 201610030000 -i ../../data/
   raw_case_input/FV3_C96_example_ICs -g ../../data/raw_case_input/
   FV3_C96_example_ICs -n fv3_model_point_noah -oc

./UFS_IC_generator.py -l 261.51 38.2 -d 201610030000 -i ../../data/
   raw_case_input/FV3_C96_example_ICs -g ../../data/raw_case_input/
   FV3_C96_example_ICs -n fv3_model_point_noahmp -mp -oc
```

Note that the **--in_dir (-i)** and **--grid_dir (-g)** arguments are the same in this case (since the supergrid files were copied to the same directory as the initial conditions files for point of example), but they will not in general be the same. Also note that the default behavior of the script is to expect that the netCDF initial condition files were generated from `chgres_cube` and not the older `global_chgres`. If they were generated from the older version (which is likely for pre-2018 data), they will have a slightly different format requiring the **--old_chgres (-oc)** option to be set in order for the files to be read properly by the script. If you try without the **--old_chgres (-oc)** flag and receive a "IndexError: t not found" error, try the script again with the flag.

In addition to the case input files generated by this script, one will need appropriate case configuration files. Make sure that the **model_ics** variable is set to **.true.** and that the **C_RES**, **year**, **month**, **day**, and **hour** are all set to the appropriate values that match the UFS Atmosphere initial conditions used. See `scm/etc/case_config/fv3_model_point_noah.nml` and `scm/etc/case_config/fv3_model_point_noahmp.nml` for examples.

Running the model is the same as for observational field campaign cases:

```
./run_gmtb_scm.py -c fv3_model_point_noah -s SCM_GFS_v15p2
```

Although not officially supported, one may run with the NoahMP LSM too. The `dtc/develop` branch contains a SDF and namelist such that one could execute the following to exercise NoahMP:

```
./run_gmtb_scm.py -c fv3_model_point_noahmp -s SCM_GFS_v15_noahmp -n
   input_GFS_v15_noahmp.nml
```

6 CCPP Interface

Chapter 6 of the CCPP v4 Technical Documentation (https://ccpp-techdoc.readthedocs.io/en/v4.0) provides a wealth of information on the overall process of connecting a host model to the CCPP framework for calling physics. This chapter describes the particular implementation within this SCM, including how to set up, initialize, call, and change a physics suite using the CCPP framework.

6.1 Setting up a suite

Setting up a physics suite for use in the GMTB SCM with the CCPP framework involves three steps: preparing data to be made available to physics through the CCPP, running the `ccpp_prebuild.py` script to reconcile SCM-provided variables with physics-required variables, and preparing a suite definition file.

6.1.1 Preparing data from the SCM

As described in sections 6.1 and 6.2 of the CCPP Technical Documentation a host model must allocate memory and provide metadata for variables that are passed into and out of the schemes within the physics suite. As of this release, in practice this means that a host model must do this for all variables needed by all physics schemes that are expected to be used with the host model. For this SCM, all variables needed by the physics schemes are allocated and documented in the file `gmtb-scm/scm/src/gmtb_scm_type_defs.f90` and are contained within the `physics` derived data type. This derived data type initializes its component variables in a `create` type-bound procedure. As mentioned in section 6.2 of the CCPP Technical Documentation, a table containing all required metadata was constructed for describing all variables in the `physics` derived data type. The standard names of all variables in this table must match with a corresponding variable within one or more of the physics schemes. A list of all standard names used can be found in `ccpp/framework/doc/DevelopersGuide/CCPP_VARIABLES_SCM.pdf`. The `local_name` for each variable corresponds to how a variable is referenced from the point in the code where `ccpp_field_add()` statements are made. For this SCM, then, all `local_name`s begin with the `physics` derived data type. Nested within most of the `local_name`s is also the name of a derived data type used within the UFS Atmosphere cap (re-used here for expediency). Since the `ccpp_field_add()` statements are made within a loop over all columns within `gmtb_scm.F90`, most `local_name`s are also referenced with `i` as an array index.

6.1.2 Editing and running `ccpp_prebuild.py`

General instructions for configuring and running the `ccpp_prebuild.py` script can be found in chapter 8 of the CCPP Technical Documentation. The script expects to be run with a host-model-dependent configuration file, passed as argument `-config=path_to_config_file`. Within this configuration file are variables that hold paths to the variable definition files (where metadata tables can be found on the host model side), the scheme files (a list of paths to all source files containing scheme entry points), the auto-generated physics schemes makefile snippet, the auto-generated physics scheme caps makefile snippet, the file where `ccpp_modules.inc` and `ccpp_fields.inc` are included, and the directory where the auto-generated physics caps should be written out to. Other variables less likely to be modified by a user are included in this configuration file as well, such as code sections to be included in the auto-generated scheme caps. As mentioned in section 2.3, this script must be run to reconcile data provided by the SCM with data required by the physics schemes before compilation by following step 1 in that section.

6.1.3 Preparing a suite definition file

The suite definition file is a text file read by the model at compile time. It is used to specify the physical parameterization suite, and includes information about the number of parameterization groupings, which parameterizations that are part of each of the groups, the order in which the parameterizations should be run, and whether subcycling will be used to run any of the parameterizations with shorter timesteps.

In addition to the six or so major parameterization categories (such as radiation, boundary layer, deep convection, resolved moist physics, etc.), the suite definition file can also have an arbitrary number of additional interstitial schemes in between the parameterizations to prepare or postprocess data. In many models, this interstitial code is not known to the model user but with the suite definition file, both the physical parameterizations and the interstitial processing are listed explicitly.

The suite definition file also invokes an initialization step, which is run only once when the model is first initialized. Finally, the name of the suite is listed in the suite definition file. By default, this suite name is used to compose the name of the shared library (.so file) that contains the code for the physical parameterizations and that must be dynamically linked at run time.

For this release, supported suite definition files used with this SCM are found in `gmtb-scm/ccpp/suites`. For all of these suites, the physics schemes have been organized into 3 groupings following how the physics are called in the UFS Atmosphere model, although no code is executed in the SCM time loop between execution of the grouped schemes. Several “interstitial” schemes are included in the suite definition file to execute code that previously was part of a hard-coded physics driver. Some of these schemes may eventually be rolled into the schemes themselves, improving portability.

6.2 Initializing/running a suite

The process for initializing and running a suite in this SCM is described in sections 4.4 and 4.5, respectively. A more general description of the process for performing suite initialization and running can also be found in sections 6.4 and 6.5 of the CCPP Technical Documentation.

6.3 Changing a suite

6.3.1 Replacing a scheme with another

When the CCPP has reached a state of maturity, the process for modifying the contents of an existing physics suite will be a very straightforward process, consisting of merely changing the name of the scheme in the suite definition file. As of this release, which consists of one scheme of each "type" in the pool of CCPP-compliant physics schemes with many short interstitial schemes, the process requires some consideration. Of course, prior to being able to swap a scheme within a suite, one must first add a CCPP-compliant scheme to the pool of available schemes in the CCPP physics repository. This process is described in chapter 2 of the CCPP Technical Documentation.

Once a CCPP-compliant scheme has been added to the CCPP physics repository, the process for modifying an existing suite should take the following steps into account:

- Examine and compare the arguments of the scheme being replaced and the replacement scheme.
  - Are there any new variables that the replacement scheme needs from the host application? If so, these new variables must be added to the host model cap. For the SCM, this involves adding a component variable to the `physics` derived data type and a corresponding entry in the metadata table. The new variables must also be allocated and initialized in the `physics%create` type-bound procedure.
  - Do any of the new variables need to be calculated in an interstitial scheme? If so, one must be written and made CCPP-compliant itself. The CCPP Technical Documentation will help in this endeavor, and the process outlined in its chapter 2 should be followed.
  - Do other schemes in the suite rely on output variables from the scheme being replaced that are no longer being supplied by the replacement scheme? Do these output variables need to be derived/calculated in an interstitial scheme? If so, see the previous bullet about adding one.
- Examine existing interstitial schemes related to the scheme being replaced.
  - There may be scheme-specific interstitial schemes (needed for one specific scheme) and/or type-generic interstitial schemes (those that are called for all schemes of a given type, i.e. all PBL schemes). Does one need to write analogous scheme-specific interstitial schemes for the replacement?

  - Are the type-generic interstitial schemes relevant or do they need to be modified?
- Depending on the answers to the above considerations, edit the suite definition file as necessary. Typically, this would involve finding the `<scheme>` elements associated with the scheme to be replaced and its associated interstitial `<scheme>` elements and simply replacing the scheme names to reflect their replacements. See chapter 4 of the CCPP Technical Documentation for further details.

6.3.2 Modifying "groups" of parameterizations

The concept of grouping physics in the suite definition file (currently reflected in the `<group name="XYZ">` elements) enables "groups" of parameterizations to be called with other computation (perhaps related to the dycore, I/O, etc.) in between. In the suite definition file included in this release, three groups are specified, but currently no computation happens between `ccpp_physics_run` calls for these groups. However, one can edit the groups to suit the needs of the host application. For example, if a subset of physics schemes needs to be more tightly connected with the dynamics and called more frequently, one could create a group consisting of that subset and place a `ccpp_physics_run` call in the appropriate place in the host application. The remainder of the parameterizations groups could be called using `ccpp_physics_run` calls in a different part of the host application code.

6.3.3 Subcycling parameterizations

The suite definition file allows subcycling of schemes, or calling a subset of schemes at a smaller time step than others. The `<subcycle loop = n>` element in the suite definition file controls this function. All schemes within such an element are called `n` times during one `ccpp_physics_run` call. An example of this is found in the `suite_SCM_GFS_v15p2.xml` suite definition file, where the surface schemes are executed twice for each timestep (implementing a predictor/corrector paradigm). Note that no time step information is included in the suite definition file. If subcycling is used for a set of parameterizations, the smaller time step must be an input argument for those schemes.

6.4 Adding variables

6.4.1 Adding a physics-only variable

Suppose that one wants to add the variable `foo` to a scheme that spans the depth of the column and that this variable is internal to physics, not part of the SCM state or subject to external forcing. Here is how one would do so:

1. First, add the new variable to the `physics` derived data type definition in `gmtb-scm/scm/src/gmtb_scm_type_defs.f90`. Within the definition, you'll notice that there are nested derived data types (which contain most of the variables needed by the physics and are used for mainly legacy reasons) and several other integers/reals/logicals. One could add the new variable to one of the nested GFS derived data types if the variable neatly fits inside one of them, but it is suggested to bypass the GFS derived data types and add a variable directly to the `physics` type definition:

   ```
   real(kind=kind_phys), allocatable :: foo(:,:)
   ```

2. Second, within the `physics_create` subroutine, add an allocate and initialization statement.

   ```
   allocate(foo(n_columns, n_levels))
   physics%foo = 0.0
   ```

   Note that even though foo only needs to have the vertical dimension, it is also allocated with the `n_columns` dimension as the first dimension since this model is intended to be used with multiple independent columns. Also, the initialization in this creation subroutine can be overwritten by an initialization subroutine associated with a particular scheme.
3. At this point, these changes are enough to allocate the new variable (`physics%create` is called in the main subroutine of `gmtb_scm.F90`), although this variable cannot be used in a physics scheme yet. For that, you'll need to add an entry in the long metadata table entry that precedes the `physics` type definition in `gmtb_scm_type_defs.F90`. This entry looks like:

   ```
   !! | physics%foo(i) | foo | description of foo | units of foo |
      rank of foo | data type of foo| kind of data type (if real) |
      intent (none) | whether foo is optional (T or F) |
   ```

   The elements of the metadata entry are all on one line (following the same format as other entries) and include the variable's "local name" or how it is referenced from `gmtb_scm.F90/main`, its "standard name" or how it is referenced by both the host model and the physics code, its units, its rank (dimensionality), Fortran intrinsic data type, the real kind if necessary, its intent (must be none for this host-side table), its optionality (must be F for this host-side table). This metadata entry is parsed by the CCPP framework and makes this variable available for any CCPP-compliant physics schemes to use.
4. On the physics scheme side, there will also be a metadata table entry for **`foo`** preceding the subroutine in which it is used. For example, say that scheme **`bar`** uses **`foo`**. If **`foo`** is further initialized in **`bar`**'s `_init` subroutine, a metadata entry for **`foo`** must precede the `bar_init` subroutine's code. If it is used in **`bar`**'s run subroutine, a metadata entry for foo must also appear in the preceding metadata table for `bar_run`. The metadata entry on the physics scheme side has the same format as the one on the host model side described above. The standard name, units, rank, type, and kind must match the entry from the host model table. Others attributes (local name, description, intent, optional) can differ. The local name corresponds to the name of the variable used within the scheme subroutine, and intent and optional attributes should reflect how the variable is actually used within the scheme.
   Note: In addition to the metadata table, the argument list for the scheme subroutine must include the new variable (i.e., **`foo`** must actually be in the argument list for `bar_run` and be declared appropriately in regular Fortran).

If a variable is declared following these steps, it can be used in any CCPP-compliant physics scheme and it will retain its value from timestep to timestep. A variable will ONLY be zeroed out (either every timestep or periodically) if it is in the `GFS_interstitial` or `GFS_diag` data types. So, if one needs the new variable to be 'prognostic', one would need to handle updating its value within the scheme, something like:

$$\text{foo}^{t+1} = \text{foo}^{t} + \Delta t * \text{foo\_tendency} \tag{6.1}$$

Technically, the host model can "see" foo between calls to physics (since the host model allocated its memory at initialization), but it will not be touching it.

6.4.2 Adding a prognostic SCM variable

The following instructions are valid for adding a passive, prognostic tracer to the SCM. Throughout these instructions, the new tracer is called 'smoke'.

1. Add a new tracer to the SCM state. In `gmtb-scm/scm/src/gmtb_scm_type_defs.f90` do the following:
   - Add an index for the new tracer in the `scm_state_type` definition.
   - Do the following in the `scm_state_create` subroutine:
     - Increment `scm_state%n_tracers`
     - Set `scm_state%smoke_index` = **`(next available integer)`**
     - Set `scm_state%tracer_names(scm_state%smoke_index)` = **`'smoke'`**
     - Note: `scm_state%state_tracer` is initialized to zero in this subroutine already, so there is no need to do so again.
2. Initialize the new tracer to something other than zero (from an input file).
   - Edit an existing input file (in `gmtb-scm/scm/data/processed_case_input`): add a field in the 'initial' group of the netCDF file(s) (with vertical dimension in pressure coordinates) with an appropriate name in one (or all) of the input netCDF files and populate with whatever values are necessary to initialize the new tracer.
   - Create a new input variable to read in the initialized values. In `gmtb-scm/scm/src/gmtb_scm_type_defs.f90`:
     - Add a new input variable in `scm_input_type`

       ```
       real(kind=dp), allocatable              :: input_smoke(:)
       ```

     - In `scm_input_create`, allocate and initialize the new variable to 0.
   - Read in the input values to initialize the new tracer. In `gmtb-scm/scm/src/gmtb_scm_input.f90/get_case_init`:
     - Add a variable under the initial profile section:

       ```
       real(kind=dp), allocatable  :: input_smoke(:) !< smoke
          profile (fraction)
       ```

     - Add the new input variable to the allocate statement.
     - Read the values in from the file:

```
call check(NF90_INQ_VARID(grp_ncid,"smoke",varID))
call check(NF90_GET_VAR(grp_ncid,varID,input_smoke))
```

  - set `scm_input%input_smoke` = **`input_smoke`**
- Interpolate the input values to the model grid. Edit `gmtb_scm_setup.f90/set_state`:
  - Add a loop over the columns to call `interpolate_to_grid_centers` that puts **`input_smoke`** on grid levels in `scm_state%state_tracer`

```
do i=1, scm_state%n_cols
       call interpolate_to_grid_centers(scm_input%
          input_nlev, scm_input%input_pres, scm_input%
          input_smoke, scm_state%pres_l(i,1,:), &
          scm_state%n_levels, scm_state%state_tracer(i
             ,1,:,scm_state%smoke_index,1),
             last_index_init, 1)
end do
```

- At this point, you have a new tracer initialized to values specified in the input file on the model vertical grid, but it is not connected to any physics or changed by any forcing.

3. For these instructions, we'll assume that the tracer is not subject to any external forcing (e.g., horizontal advective forcing, sources, sinks). If it is, further work is required to:
   - One needs to provide data on how tracer is forced in the input file, similar to specifying its initial state, as above.
   - Create, allocate, and read in the new variable for forcing (similar to above).
   - Add to `interpolate_forcing` (similar to above, but interpolates the forcing to the model grid and model time).
   - Add statements to time loop to handle the first time step and different time-advancing schemes.
   - Edit `apply_forcing_forward_Euler` and `apply_forcing_leapfrog` in `gmtb-scm/scm/src/gmtb_scm_forcing.f90`.
4. In order to connect the new tracer to the CCPP physics, perform steps 1-4 in section 6.4.1 for adding a physics variable. In addition, do the following in order to associate the `scm_state` variable with variables used in the physics through a pointer:
   - Point the new physics variable to `scm_state%state_tracer(:,:,:,scm_state%smoke_index)` in `gmtb-scm/scm/src/gmtb_scm_type_defs.f90/physics_associate`.
5. There may be additional steps depending on how the tracer is used in the physics and how the physics scheme is integrated with the current GFS physics suite. For example, the GFS physics has two tracer arrays, one for holding tracer values before the physics timestep (`gmtb-scm/scm/src/GFS_typedefs.F90/GFS_statein_type/qgrs`) and one for holding tracer values that are updated during/after the physics (`gmtb-scm/scm/src/GFS_typedefs.F90/GFS_stateout_type/gq0`). If the tracer needs to be part of these arrays, there are a few additional steps to take.