

Supplement of Geosci. Model Dev. Discuss., 8, 2313–2367, 2015
<http://www.geosci-model-dev-discuss.net/8/2313/2015/>
doi:10.5194/gmdd-8-2313-2015-supplement
© Author(s) 2015. CC Attribution 3.0 License.



Supplement of

Development and application of the WRFPLUS-Chem online chemistry adjoint and WRFDA-Chem assimilation system

J. J. Guerrette and D. K. Henze

Correspondence to: J. J. Guerrette (jonathan.guerrette@colorado.edu)

1 Introduction

WRFPLUS-Chem contains forward (FWM), adjoint (ADM), and tangent linear (TLM) models. The FWM is comparable to the equivalent version of WRF-Chem, with some modifications made to enable scaling of emissions. The ADM and TLM advection and diffusion are identical to those provided in WRFPLUS, which correspond to the adiabatic physics solver in the FWM. WRFPLUS-Chem additionally includes ADM and TLM modules corresponding to specific FWM namelist options for planetary boundary layer (PBL) mixing, surface layer, land surface model (LSM), dry deposition, emissions, and chemistry. For a description of options available, refer to the main GMD manuscript. In this release, ADM and TLM emissions, chemistry, and dry deposition are limited to the “GOCART only” aerosol option (`chem_opt = 300`).

The purpose of this manual is to describe compilation and verification procedures for WRFPLUS-Chem. The reader should already be familiar with WRF and WRF-Chem. Users should also be familiar with running WRF-Chem preprocessors for emissions. It is expected that users have some familiarity with calculating sensitivities, even if only by finite difference approximations. Refer to the user guides and online documentation for those tools as necessary while learning how to use WRFPLUS-Chem.

If you are interested in 4D-Var chemical data assimilation using WRFDA-Chem, that development and user guide will be available at a later date. Integration between WRFPLUS-Chem and WRFDA-Chem is still in its early stages, and at this time only allows sensitivity studies. Contact the author for more information.

2 Software Requirements

The minimum software requirements of WRFPLUS-Chem are the combined requirements of WRFPLUS and WRF-Chem. Of primary concern, the user will need a functional netCDF library for I/O. Although WRF-Chem and the forward chemistry model in WRFPLUS-Chem have the choice of using KPP chemistry code, this option is not yet available in the AD/TL models. Thus the user may decide to have the fast lexical analyzer (FLEX) library or not. WRFPLUS-Chem has only been tested on a single platform, however, the source code differences from WRFPLUS and WRF-Chem are minimal.

3 Installing WRFPLUS-Chem

To install WRFPLUS-Chem:

- Obtain WRFPLUS-Chem one of two ways:

1. From the NCAR code repository:

`https://svn-wrf-model.cgd.ucar.edu/branches/WRFPLUSV3-Chem`

A username and password is required to access the code here. In the future, a zipped tar file may be available for download.

2. Contact the author.

This package contains the entire WRFPLUS-Chem model, including chemistry.

- Set the environment variables
 - > `setenv WRF_CHEM 1`
 - > `setenv WRFIO_NCD_LARGE_FILE_SUPPORT 1`

The last of these you will only need if you plan to

- Run the configure script
 - > `cd WRFPLUSV3-Chem`
 - > `./configure wrfplus`

Select from options given the closest configuration to your own. Because of the high computational demands of the ADM, `dmpar` is Distributed Memory Parallel

- Modify `configure.wrf`

Compiling requires much longer than WRFPLUS or WRF-Chem would on their own, because the dependency list is longer. If you are having trouble getting the compile to succeed, consider reducing the optimization “-O3” in `configure.wrf` to “-O0.” Additionally, the CLM LSM option is not available in the TLM or ADM. Thus, also consider changing the following `configure.wrf` line from

```
ARCH_LOCAL = -DNONSTANDARD_SYSTEM_FUNC -DWRF_USE_CLM
```

to

```
ARCH_LOCAL = -DNONSTANDARD_SYSTEM_FUNC #-DWRF_USE_CLM
```

An “-O3” optimization level compile without the CLM package may take over an hour, whereas using “-O0” may take only 15 minutes. In the latter case, run time is approximately 3 x longer.

- Compile WRFPLUS-Chem
 - > `./compile -j 1 em_real >& compile.out`
 - > `ls -ls main/*.exe`

You should see the following files:

```
52456 -rwxrwx— 1 user users 53686631 Dec 2 02:59 main/ndown.exe*
51240 -rwxrwx— 1 user users 52443161 Dec 2 02:59 main/nup.exe*
52476 -rwxrwx— 1 user users 53704095 Dec 2 03:00 main/real.exe*
51324 -rwxrwx— 1 user users 52530145 Dec 2 02:59 main/tc.exe*
67848 -rwxrwx— 1 user users 69441571 Dec 2 02:56 main/wrf.exe*
```

The compilation above uses a single core (“-j 1”) because problems have been encountered compiling in parallel. This does not affect the parallel or serial choice during configure.

4 Generating Emissions

Only anthropogenic and biomass burning emissions are enabled in the ADM. A separate directory containing compiled WRF-Chem and `convert_emiss.exe` executables will be required to prepare emissions input files for WRFPLUS-Chem. The input files are generated using exactly the same methods described in the WRF-Chem 3.6 user guide, emission guide, and online tutorials (<http://ruc.noaa.gov/wrf/WG11/>). Constructing emission datasets is non-trivial, and is a required step before conducting meaningful WRFPLUS-Chem and WRFDA-Chem analyses.

Once the emission files are generated, either link them or copy them to `WRFPLUSV3-Chem/test/em_real`. Additionally, emissions should be distributed to separate files corresponding to each hour of the simulation. If you are familiar with NCL, the script `disperse_anthro_emissions.ncl` will do just that.

5 Running WRFPLUS-Chem

In its standalone form, without WRFDA-Chem, the primary purpose for running WRFPLUS-Chem is to verify the derivatives, or sensitivities, calculated by the AD/TL models. The model is distributed with an example `namelist.input` file to use for verification. You will need to specify the domain parameters, including I/O variables, as would normally be done in any WRF or WRF-Chem simulation. In order to carry out a verification, set the namelist variables `dyn_opt=102` and `scenario_type=1` or `2` (default is `0`).

The verification procedure considers derivatives of the form

$$\chi_{p,q} = \frac{\partial J_{p,f}}{\partial x_{q,i}}. \quad (1)$$

J is a cost function at location p and time step f , while x is a control variable (CV) at location q and the initial time i . J is a placeholder for a predicted model state, such as temperature (T), water vapor mixing ratio (q_v), zonal velocity (u), meridional velocity (v), or a chemical species concentration ($[c]$). x is a placeholder for any user prescribed variable, such as an initial condition for the above state variables, or emission scaling factor of a chemical species ($\alpha_{c,i_s c}$). You can select which of the three models to use to calculate these derivatives by setting the `check_NL`, `check_AD`, and `check_TL` namelist options to `.true.` or `.false..` The `check_NL` variable is used to turn on nonlinear (NL) finite difference approximation. If this option is turned on, you should also set `nL_pert` to the desired perturbation size (e.g., `0.1 = 10%`).

You must specify to WRFPLUS-Chem values for J , p , x , and q . J and x are specified in the namelist variables `numer_vars` and `denom_vars`, respectively. The selectable indices of the numerator and denominator variables can be seen in the log file by running a sample verification. The log file will include a list of variables like this:

```

1 U
2 V
3 W
4 T
5 PH
6 MU
7 DUMMY
8 MOIST 1
9 DUMMY
10 TRACER 1
11 DUMMY
12 CHEM 1
13 CHEM 2
14 CHEM 3
15 CHEM 4

```

16 CHEM 5
 17 CHEM 6
 18 CHEM 7
 19 CHEM 8
 20 CHEM 9
 21 CHEM10
 22 CHEM11
 23 CHEM12
 24 CHEM13
 25 CHEM14
 26 CHEM15
 27 CHEM16
 28 CHEM17
 29 CHEM18
 30 CHEM19
 31 DUMMY
 32 SCALEANT 1
 33 SCALEANT 2
 34 SCALEANT 3
 35 SCALEANT 4
 36 SCALEANT 5
 37 SCALEANT 6
 38 DUMMY
 39 SCALEBB 1
 40 SCALEBB 2
 41 SCALEBB 3

The first six variables in this list are fixed. The “MOIST”, “TRACER”, “CHEM”, “SCALEANT”, and “SCALEBB” indices will change based on your choices for `mp_physics`, `tracer_opt`, `chem_opt`, `scale_emiss_opt`, and `scale_burn_opt`. The number on the left is the index that should be used in the `numer_vars` and `denom_vars` lists. If `scenario_type` is set to 1, then these lists are a one-to-one matching of the variables that you would like to test. If `scenario_type` is set to 2, then the variables in the lists will be matched in all possible combinations. `max_numer` and `max_denom` should also be set accordingly to the number of variables in each list.

As an example for the “CHEM” array, you can view the list of chemical tracers in `Registry/registry.chem` under the selected value for `chem_opt`. For option 300 (GOCART aerosols), the list is as follows:

```
# GOCART Packages (no Ozone chemistry in this choice)
package gocart_simple chem_opt==300 -
chem:so2,sulf,dms,msa,p25,bc1,bc2,oc1,oc2,dust_1,dust_2,dust_3,dust_4,dust_5,
seas_1,seas_2,seas_3,seas_4,p10
```

So for example, dimethyl sulfide (dms) and hydrophilic black carbon (bc2) correspond to “CHEM 3” and “CHEM 7,” respectively (variable indices 14 and 18 in the list above). However, if either `mp_physics` or `tracer_opt` are altered, their variable indices may change.

p and q are specified as domain indices in ASCII text files `locations_i` and `locations_f` in

the `test/em_real` directory. `adtl_forc_locations.ncl` can be used to generate these files, or you can devise your own method. The initial and final locations are specified in a way that reduces the number of ADM simulations in the verification procedure, because it is the most expensive of the three models. So for each final location, you can specify multiple initial locations. The primary requirement of `locations_i` and `locations_f` is that the number of initial locations is some integer multiple of the number of final locations. For example:

```
grep '' locations_f (2 locations)
20    35    1
14    48    9
```

```
grep '' locations_i (14 locations)
12    35    1
18    38    1
14    36    1
15    37    1
20    34    11
16    33    7
22    35    3
10    48    5
12    51    5
8     49    9
9     50    7
14    47    19
10    46    15
16    48    11
```

In this case, the first 7 initial locations correspond to the first final location. `adtl_forc_locations.ncl` specifies the initial locations using identical relative coordinates from each final location, although more complex approaches could be used.

Important Notes: The emission scaling options `scale_emiss_opt` and `scale_burn_opt` must correspond to existing emission options (see `Registry/registry.chem` and `Registry/registry.chemplus` and the namelist options of this manual). For verification purposes, `num_ant_steps` and `num_bb_steps` should both be set to 1. `scaleant_frq` and `scalebb_frq` should both be set to a time (in minutes) longer than the simulation time. `nta_chem_opt` should be identical to `chem_opt`.

6 Memory Requirements

As is described in the main GMD text, running either the TLM or ADM requires storing the trajectory of state variables generated by the nonlinear model at each and every time step. The nonlinear trajectory is stored in memory. The required memory increases linearly with the number of 3D state variables, the simulation duration, and the number of model levels. The total memory required also increases by 50% for every factor of 2 increase in the number of cores (i.e., number of tiles in WRF). Finally, for a given domain extent, halving the grid cell width (doubling the res-

olution) quadruples the memory required. Keep all of these relationships in mind when specifying a domain and the number of cores per node used when submitting a job to queue.

The memory scaling with simulation duration has been minimized by adding a second order checkpointing scheme as described for the ADM in Fig. 1. A checkpointed adjoint simulation begins with a full FWM simulation beginning at the initial time, t_0 , and ending at the final time, t_f . WRF restart files are written at time intervals equal to the namelist variable `checkpoint_interval`, or Δt_c . Once the simulation is completed, the FWM is restarted at initial time equal to $t_f - \Delta t_c$. During that simulation, the trajectory is stored in memory. The trajectory is then recalled in an adjoint simulation that proceeds backward toward the current initial time. The checkpoint system alternately calls the FWM and ADM until returning to t_0 . A similar checkpointing system is also implemented for the TLM.

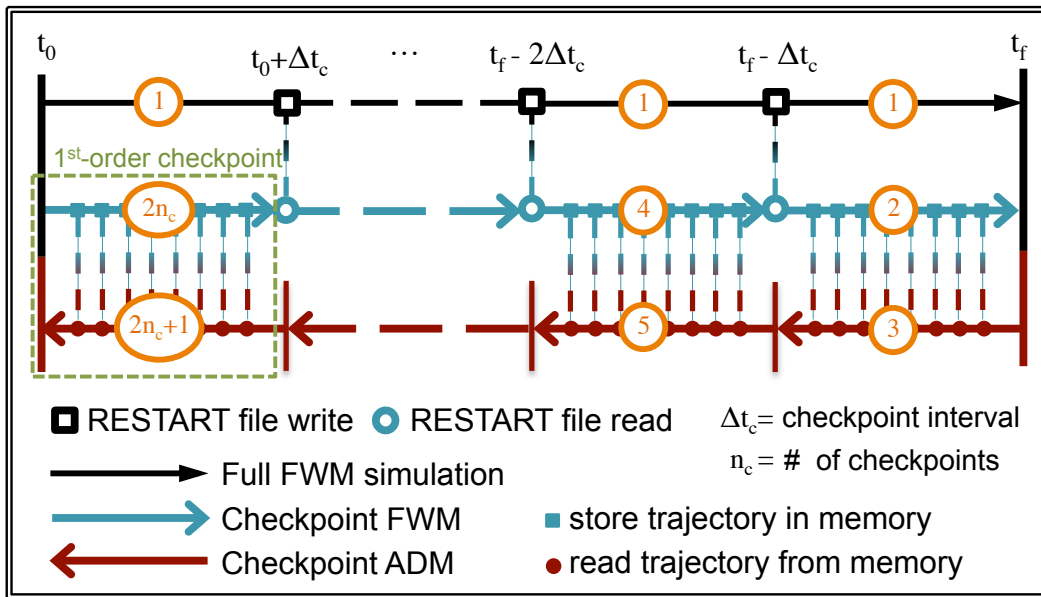


Figure 1: Second order checkpointing scheme implemented in WRFPLUS-Chem.

The total memory requirement can be reduced by setting `checkpoint_interval > 0`. If you do this, you must set it equal to the number of hours between boundary condition reading. Due to the way boundary value tendencies are calculated from meteorological input files in WPS, the minimum simulation duration and boundary reading interval (`interval_seconds`) is 3 hours. You may elect to turn boundary reading off for the verification procedure, in which case, set `boundary_io = .false..` If you are running WRFPLUS-Chem for the first time, it is suggested you start with a 3 hour simulation for your domain, and monitor the memory usage versus memory available. Use your best judgment moving forward from there.

7 New WRFPLUS-Chem namelist options

This user guide has referenced several namelist options introduced by WRFPLUS-Chem, and some that are already available in WRF-Chem. Here they are listed for your reference:

namelist variable	Description	Valid values (default)
<code>dyn_opt</code>	Select the dynamic core	(2) Forward model 102 Verification 202 Tangent linear 302 Adjoint
<code>chem_opt</code>	Forward model chemistry option	(0) no chemistry 300 GOCART aerosols only; many others available. See WRF-Chem user guide.
<code>nta_chem_opt</code>	TL and AD chemistry option	(0) no chemistry 300 GOCART aerosols only
<code>scale_emiss_opt</code>	Anthropogenic emission scaling option (should match <code>emiss_opt</code>)	
<code>num_ant_steps</code>	Number of intervals to allocate in anthropogenic scaling factors	
<code>scaleant_frq</code>	Frequency of anthropogenic emission scaling factors (minutes)	
<code>scale_burn_opt</code>	Biomass burning emission scaling option (should match <code>biomass_burn_opt</code>)	
<code>num_bb_steps</code>	Number of intervals to allocate in biomass burning scaling factors	
<code>scalebb_frq</code>	Frequency of biomass burning emission scaling factors (minutes)	
<code>scenario_type</code>	Chooses how verification scenarios are specified	(0) Carry out full domain test instead of χ verification; not available for chemistry variables 1 χ verification: list of J and x variable combinations 2 χ verification: Use all combinations of J and x variable lists
<code>check_NL</code>	Select whether to calculate χ^{NL}	.false., .true.
<code>check_TL</code>	Select whether to calculate χ^{TL}	.false., .true.
<code>check_AD</code>	Select whether to calculate χ^{AD}	.false., .true.
<code>nl_pert</code>	Nonlinear perturbation for finite difference test	$0 < \delta x < 1$

<code>numer_vars</code>	Variable list for the cost function (J)	See text.
<code>max_numer</code>	Number of numerator variables to include in verification	See text.
<code>denom_vars</code>	Variable list for the control variables (x)	See text.
<code>max_denom</code>	Number of denominator variables to include in verification	See text.
<code>checkpoint_interval</code>	Second order checkpointing interval (hours)	(0), 3, 6, 9 , etc...; Must be equal to boundary read interval.