Mechanistic Soil N module (In-line with CMAQ v5.1)

User's Guide

Table of Contents

List of Tables

1. Introduction

This guide provides instructions on how to build and run the new mechanistic soil nitrogen (N) module in-line with CMAQ version 5.1, to simulate NO, HONO, and N_2O emissions from soil for climate and air quality applications. This document is written in a relatively generic order so that the reader can choose individual sections for reference based on their own purpose (e.g. background knowledge, software installation, real case implementation, future model update). However, Section 1 gives a complete overview of the implementation of mechanistic soil N module apt for a first-time user. Comments on how to improve the documentation are encouraged.

1.1 Mechanistic soil N module in-line implementation with CMAQ: Overview

This is the first time a mechanistic representation of N emissions from soil, such as- nitric oxide (NO), nitrous acid (HONO) and nitrous oxide (N_2O) is being implemented in a regional scale air quality model. Previously, more recent parametrized schemes like, BDSNP has been implemented in both CMAQ (Rasool et al., 2016) and global chemical transport models such as, GEOS-Chem (Hudman et al., 2010, 2012) to estimate soil NO emissions. Although, BDSNP has a more continuous soil temperature/moisture dependence function compared to the soil NO emission module (Yienger and Levy 1995; YL95) currently coded in biogenic emission models such as BEIS3 and MEGAN, it still ignores the complete soil N transformation cycle that produces other N emissions like HONO and N_2O.

Based on the mechanistic scheme, soil NO, HONO, and N_2O emissions at each modeling grid and time-step are determined by the accounting for soil nitrogen (either form the fertilizer application during the growing season or the soil N naturally available in non-agricultural biome or the nitrogen dry/wet deposition from the atmosphere), and the subsequent transformation of soil nitrogen either through nitrification or denitrification depending on variables such as, soil temperature and moisture, availability of soil NH_4 or NO_3, relative availability of NO_3 to C, soil pH, gas diffusivity, the soil pulsing after

precipitation ($P(l_{dry})$), and the canopy reduction due to the resistance (CRF). The algorithm for canopy reduction was proposed by Wang et al. (1998) in GEOS-Chem is detailed in section 2.5 of the paper and it considers a more physical basis based on wind speed, turbulence, canopy structure, deposition constants and other physical variables. Both Pulsing and canopy reduction functions are the same as applied for BDSNP (Rasool et al., 2016). The algorithms for nitrification and denitrification soil N emissions are based on DAYCENT biogeochemical model (Del Grosso et al., 2000; Parton et al., 2001).

$$S_{NO} = \begin{pmatrix} N_{NO_x} - S_{HONO} \\ + \\ D_{NO} \end{pmatrix} CRF(LAI, Meteorology, Biome)$$

$$\equiv \begin{pmatrix} f(NH_4, T_{soil}, \theta_{soil}, Dr, pH)P(l_{dry}) \\ + \\ f(NO_3:C, T_{soil}, \theta_{soil}, Dr) \end{pmatrix} CRF(LAI, Meteorology, Biome)$$

$$S_{HONO} = (HONO_f)(N_{NO_x})(f_{SWC})CRF(LAI, Meteorology, Biome)$$

$$\equiv (HONO_f)\left(f(NH_4, T_{soil}, \theta_{soil}, Dr, pH)P(l_{dry})\right)(f_{SWC})CRF(LAI, Meteorology, Biome)$$

$$S_{N_2O} = \begin{pmatrix} N_{N_2O} \\ + \\ D_{N_2O} \end{pmatrix} \equiv \begin{pmatrix} f(NH_4, T_{soil}, \theta_{soil}, Dr, pH) \\ + \\ f(NO_3:C, T_{soil}, \theta_{soil}, Dr) \end{pmatrix}$$

As stated in the paper, soil NH3 emission is calculated based on the bi-directional exchange scheme (Bash et al., 2013) in CMAQ.

This work proposes an implementation of the new mechanistic code for the first time into the Community Multiscale Air Quality (CMAQ) modeling system v5.1 to produce the online soil NO, HONO, and N2O emission estimates using the soil temperature/moisture information from the Pleim-Xiu land surface model (Pleim and Xiu, 2003) in the Weather Research Forecasting (WRF) model v 3.7. However, the inline mechanistic module proposed here is specifically implemented with CMAQ, and is computationally expensive and not available for cross-platform yet. For example, the mechanistic soil N scheme has not been implemented in the CAMx air quality model.

From the software engineering perspective, Figure 1c in the main paper provides the flow chart of the mechanistic soil N scheme implementation. Input files in addition to those available in the traditional WRF-CMAQ setup include:

a) Environmental Policy Integrated Climate (EPIC v.0509) agricultural model outputs made compatible for use in CMAQ through Fertilizer Emission Scenario Tool for CMAQ (FEST-C v1.2, https://www.cmascenter.org/fest-c/). These daily inputs include soil properties like pH and bulk density for agricultural soil and agricultural soil NH_4 (also utilized for NH_3 bi-directional scheme in CMAQ), NO_3 and Organic C and N pools (all in FERTRESFILE).

b) Soil biome type file (LANDFRACFILE) consistent with the type of the MODIS 24 Land Use categories as used in implementing BDSNP in-line with CMAQ (Rasool et al., 2016) and climate zone file based (KOPPENFILE) on Köppen-Geiger climate zone definition (Kottek et al., 2006) are needed to map the soil N emissions at each modeling grid relevant to different biomes.
CMAQ by default uses 40-category NLCD Land Use classification (NLCD40) with sub-grid fractions for different land use, hence in this case we map the MODIS based soil biome map to NLD40 using the climate zone definition to better represent the up-to-date LU/LC change with finer details. The required mapping to the 40 NLCD categories is carried out in-line in CMAQ itself.
c) Separate input files for soil organic carbon (SMCRESFILE), soil organic nitrogen (SMNRESFILE), Soil microbial C:N ratio (CNRATIOFILE) and soil pH for (PHTOPSOILFILE) are provided, re-gridded for the CONUS domain from a compilation of global soil dataset (Xu et al., 2014), to drive the C-N biogeochemical cycling that leads to nitrification and denitrification generating varied soil N emissions.

Soil moisture and temperature as well as some meteorological variables such as radiation, wind and air pressure are needed (provided from WRF v3.7 model) to drive the mechanistic model, and also to represent the occurrence of pulsing and the process of canopy reduction. Also, owing to implementation of EPIC-CMAQ bi-directional framework we already have

in-line soil properties such as, water contents at wilting point; field capacity; and saturation, crucial in calculating dependence of nitrification and denitrification rates on soil moisture.

For the in-line mechanistic scheme, dry or wet deposition process schemes were modified for CMAQ air quality model to include and continuously update the available nitrogen (N) from the atmosphere to the ground. N Deposition have been used earlier to add to soil N both in NH_3 bi-directional scheme (Bash et al., 2013) and parametrized schemes like BDSNP (Rasool et al., 2016) in-line in CMAQ.

1.2 Mechanistic soil N module: software and operating system prerequisites

The prerequisites in terms of compilers and libraries for running mechanistic soil N module in-line with CMAQ are the same as that for CMAQ version 5.1 itself, mentioned in detail in the model's instructional page https://www.airqualitymodeling.org/index.php/CMAQv5.1_Readme_file (https://www.airqualitymodeling.org/index.php/CMAQv5.1_Readme_file). The WRF/MCIP-EPIC-CMAQ framework used for this work consists of multiple models, and hence in practice system architecture requirements for compiling and running them varies depending on the user-end resources and requirements.

Running this version of the mechanistic model in-line with CMAQ v 5.1 model requires the working knowledge of Unix/Linux operating system and FORTRAN programming language as prerequisites (e.g, how to unzip/untar files, install libraries, link files, and implement basic commands within Unix, and an understanding of basic Shell scripting to invoke FORTRAN executables).

For any other custom run, re-gridding and default WRF-CMAQ input files to a preferred simulation domain will be required. For this knowledge, access to programming tools like Python and/or software like ESRI ArcMAP software or an open-source software such as the spatial allocator will be required.

2. In-line mechanistic module implementation

2.1 Introduction

The mechanistic module is simultaneously compiled with other executables of the CMAQ Chemistry-Transport Model (CCTM). CCTM integrates model-ready inputs such as emissions, meteorology, photochemistry, initial and boundary conditions to simulate continuous atmospheric chemical conditions at user-defined time frequency (hourly, daily etc.). Implementation of the mechanistic module in-line with CCTM ensures that soil NO, HONO and N_2O as per newly proposed mechanistic scheme is also accounted for while generating gridded and temporally resolved air pollutant concentrations as output. This is fairly straight-forward on the supported Linux platforms. To satisfy the pre-requisites for libraries and packages on which CMAQ depends and their installation the first step should be referring to the simple steps given in CMAQ's readme outline for users https://www.airqualitymodeling.org/index.php/CMAQv5.1_Readme_file.

2.2 Program structure for in-line mechanistic module implementation

The first step is to unzip and untar the compressed file containing CMAQ version 5.1 source code (CMAQv5.1.Nov2015.tar.gz) available at https://www.cmascenter.org/ with applicable free registration.
>$tar –xzf CMAQv5.1.Nov2015.tar.gz

The source code for models, tools, and libraries, and C-Shell scripts to build and execute the CMAQ model are all contained in the unzipped home folder also referred as $M3HOME. The part of the program structure of interest is the $M3HOME/scripts/cctm folder that contains all source codes driving the main core of CMAQ i.e. CCTM. To create the compiled object files (.o files) that eventually create executables, we need to create a $BLD directory in which they are stored:

>cd $M3HOME/scripts/cctm
 ./bldit.cctm

The folder $BLD so created, contains all CCTM source code repository that handles all its parametrizations.

The original includes tmpbeis.F program to create model-ready biogenic emissions from gridded land use data with from emissions inputs in moles/s (environmental variable NOPOL) from models like SMOKE, MEGAN and BEIS. This program originally forwarded soil NO emission as calculated by Yienger and Levy (1995) parameterization through subroutine HRNO:

C Calculate temporal non-speciated soil NO emissions from Yienger and Levy (1995) Scheme
    CALL HRNO(JDATE, JTIME, TSTEP, NOPOL)

script tmpbeis.F has been modified (in red font) to give the option to calculate soil NO emission either from Yienger and Levy (1995) parametrization or both soil NO and HONO from the new mechanistic scheme (Note: N_2O is not output through tmpbeis.F, since, it does not impact the secondary chemistry in CMAQ and also no emission inventory or boundary conditions are provided for it):

    USE BDSNP_MOD
    ! NOTE: BDSNP_MOD is used to pass emissions into the biogenic output file and calling canopy reduction and pulsing function, not be confused with use of BDSNP module as in Rasool et al. (2016)
! emissions of HONO need to allocated
    REAL,   ALLOCATABLE, SAVE :: HONOPOL(:,:)
  DO J = 1, NSEF-2 ! for mechanistic scheme
C   DO J = 1, NSEF-1 ! in YL

```fortran
ALLOCATE( BIPOL( NCOLS,NROWS,NSEF-2 ), STAT=IOS ) ! for mechanistic
scheme
C ALLOCATE( BIPOL( NCOLS,NROWS,NSEF-1 ), STAT=IOS ) ! in YL


        ALLOCATE( HONOPOL( NCOLS,NROWS ), STAT=IOS )
C  Choose Mechanistic or Yienger and Levy:
        MESG = 'Use Mechanistic?'
        BDSNP = ENVYN( 'BDSNP', MESG, .FALSE., IOS )
C NOTE: BDSNP Environment variable denotes Mechanistic scheme here


IF ( BDSNP ) THEN  ! Use Mechanistic scheme, if BDSNP is Y
        CALL  HRNOBDSNP( JDATE, JTIME, TSTEP, MY_NCOLS, MY_NROWS,
        COSZEN, NOPOL, HONOPOL ) !BDSNP_MOD.F


        EMPOL( :,:,NSEF )     = HONOPOL
        ELSE
C  Calculate temporal non-speciated soil NO emissions using YL95
        CALL HRNO( JDATE, JTIME, TSTEP, NOPOL )   !YL Scheme
END IF
        EMPOL( :,:,NSEF-1 )     = NOPOL
        EMPOL( :,:,1:NSEF-2 ) = BIPOL
```

 For this purpose, two source codes have been added: (1) a new Ncycling.F source code addressing the soil N emissions in agricultural biome, and (2) a new CNnonAgcycling.F source code addressing the same for non-agricultural biome. These codes were added as global attributes in the CCTM source code repository, which basically gives the mechanistic representation of soil NO, HONO, and N_2O emissions for the first time in a regional scale air quality model like, CMAQ. BDSNP_MOD_EPIC.F is used eventually to call the new Ncycling.F and CNnonAgcycling.F to finally pass the computed soil NO, HONO and N_2O into the biogenic emission outputs (B3GTS_S). Note: N_2O is not being passed to $B3GTS_S file but as diagnostic into $SOILINP (See Section 3)

9

Canopy reduction factor is globally defined in CCTM repository through source code canopy_nox_mod.F, which has also been used in in-line BDSNP in CMAQ (Rasool et al., 2016) for incorporating Canopy reduction of soil NO as proposed by Hudman et al. (2012) and described in section 2.5 in the main paper. The same approach is applied for the new mechanistic module.

Similarly, we modified the deposition (wet and dry) source codes in $BLD to include wet and dry deposition of N species for our application as we did for in-line BDSNP implementation in CMAQ in Rasool et al. (2016). The original code for dry and wet deposition, vdiffacm2.F and cldproc_acm.F respectively were modified to obtain total mass of N per area deposited to the ground from atmosphere. Ncycling.F and CNnonAgcycling.F encompasses the parameterization that accounts for soil N emissions from nitrification and denitrification accounting for biome classification, soil properties, fertilizer N, deposition N, non-agricultural N inputs, pulsing and reduction due to canopy (by calling in the canopy_nox_mod.F routine).

Re-compiling of CCTM source codes is required after making suitable changes to deposition schemes (dry deposition: vdiffacm2.F and wet deposition: cldproc_acm.F); the model-ready temporal biogenic emission (tmpbeis.F) source codes and new ones for the mechanistic module (Ncycling.F and CNnonAgcycling.F) and canopy reduction factor (canopy_nox_mod.F), to be called in the updated BDSNP_MOD_EPIC.F modified to accommodate the mechanistic scheme soil N emissions in the already available CCTM base source code repository in $M3SCRIPTS/cctm/$BLD (refer to CMAQ model's technical documentation).

Modified tmpbeis.F, vdiffacm2.F, cloudproc_acm.F, BDSNP_MOD_EPIC.F and newNcycling.F, CNnonAgcycling.F and canopy_nox_mod.F (canopy reduction code is actually implemented already in Rasool et al. (2016) implementation of BDSNP in CMAQ) codes have been separately attached in the supplementary material. The original source code repository for CCTM is open source, provided with the base CMAQ model package.

In the end, to run the new modified CCTM with an in-line new mechanistic soil N or parametrized BDSNP modules is computationally extensive as it will requires a wide array of inputs

(https://www.airqualitymodeling.org/index.php?title=CMAQ_version_5.1_%28Novembe r_2015_release%29_Technical_Documentation).

There are input files specific for the new mechanistic module defined in sub-routines provided for their inclusion in Ncycling.F, CNnonAgcycling.F and BDSNP_MOD_EPIC.F: Agricultural biome soil nutrient and properties, non-agricultural soil nutrient and properties, biome classification, and meteorology. More details about the inputs preparation for this work are given in Section 4. Although, it should be noted that running the mechanistic module means running the full CMAQ simulation dependent on meteorological and emissions inputs among others, which differ for different custom spatial domain and temporal configuration and take up a lot of space depending on user requirement.

2.3 Build new CCTM executable with in-line mechanistic scheme/module

Go the directory $M3HOME/scripts/cctm/$BLD in the CMAQ model directory structure and change the 'Makefile' file by setting the path to include new global modules created for including the new and modified mechanistic scheme source codes:

```
>$ cd $M3HOME/scripts/cctm
>$ cd $BLD
>$ vi Makefile
```

```
BDSNP_SOIL = \
  canopy_nox_mod.o \
  Ncycling.o \
  CNnonAgcycling.o \
  BDSNP_MOD_EPIC.o
```

Change the lines in 'Makefile' (Changes in red font) to include the path of modified deposition and biogenic emission schemes to call the new Ncycling.F and CNnonAgcucling.F codes, and subsequently call them in BDSNP_MOD_EPIC.F scheme as following and save the change:

```
vdiffacm2.o:    CGRID_SPCS.F GRID_CONF.F EMIS_DEFN.F DEPV_DEFN.F \
        ASX_DATA_MOD.F VDIFF_MAP.F UTILIO_DEFN.F BIDI_MOD.F \
        LSM_MOD.F BDSNP_MOD_EPIC.F VDIFF_DIAG.F \
        ./CONST.EXT ./FILES_CTM.EXT ./PA_CTL.EXT

tmpbeis.o:    HGRD_DEFN.F UTILIO_DEFN.F BIOG_EMIS.F BDSNP_MOD_EPIC.F

cldproc_acm.o:        GRID_CONF.F CGRID_SPCS.F UTILIO_DEFN.F  \
        BDSNP_MOD_EPIC.F ./CONST.EXT ./FILES_CTM.EXT

BDSNP_MOD_EPIC.o:       HGRD_DEFN.F   BIOG_EMIS.F   ASX_DATA_MOD.F
UTILIO_DEFN.F \
        canopy_nox_mod.F Ncycling.F CNnonAgcycling.F
```

Also, make sure that new mechanistic modules calls (include the dependency on) the canopy reduction scheme in Makefile as:

```
CNnonAgcycling.o:       HGRD_DEFN.F       BIOG_EMIS.F       ASX_DATA_MOD.F
UTILIO_DEFN.F LSM_MOD.F \
        canopy_nox_mod.F

canopy_nox_mod.o:        HGRD_DEFN.F    BIOG_EMIS.F    ASX_DATA_MOD.F
UTILIO_DEFN.F
```

NOTE: Make sure HONO is added as a species in BIOG_EMIS.F before compiling the new mechanistic soil N scheme in-line with CCTM

After saving these changes in 'Makefile', to successfully re-compile and generate the updated CCTM executable that accounts for these changes:

```
>make clean
>make >& log.build_Mechanistic &
```

If successful, you can find the executable, user-defined as environmental variable 'MODEL' in the 'Makefile', ready to use at the path $BLD. The log.build_Mechanistic is just a log file that is used to dump the compiling output message and see if it was successful or not, if not the errors would be highlighted in the log file and can be resolved accordingly.

3. Key I/O environmental variables specific to mechanistic module

Understanding the input/output (I/O) environment variables defined in mechanistic module is crucial for the real case implementation and further model development if needed.

3.1 LANDFRACFILE: soil biome type map

BDSNP parametrization used 24 MODIS biome specific soil NO emission factors based on several field studies compiled from world-wide (583 field measurements of SNOx covering the period from 1976 to 2010) for soil NO flux (SNO_X) measurement (Stehfest and Bouwman, 2006). Different biome types have different NO emission factors under different soil conditions (wet or dry) over different sampling geographic locations (or climate zones) (Steinkamp and Lawrence,2011). The LANDFRACFILE used was still the same as n BDSNP with 24 soil biome type base on MODIS land cover categories as well as Köppen climate zone classifications (Kottek et al., 2006). CMAQ v5.1 now has sub-grid scale representation for 40 National Land Cover dataset (NLCD) biomes, and was thus utilized rather than the MODIS 24 classification. Microbial/Organic biomass C and N % and HONO/N_{NOx} % which drive the mechanistic scheme are first defined for MODIS 24 Land Use types as listed in Table A1 (Appendix in main paper).

Section 2.2 of the main paper gives detail on the development of biome type maps for the 12 km spatial resolution simulations based in Continental US (CONUS). The mapping algorithm (Table A2 of Appendix in main paper) is determined by referring to the CMAQ mapping scheme, which is documented in each of the CSQY_DATA_* files under the MECHS/ directory at the CMAQ version 5.1 source code release. To enable the MODIS 24 biome type classification of emission factors, a model resolution compatible Köppen climate zone classification (Kottek et al., 2006) is utilized (KOPEENFILE) to allocate appropriate emission factor for the same biome type (e.g. 'grassland') at different locations (Table A1 in Appendix of main paper). There are five climate zone classifications, namely A: equatorial, B: arid, C: warm temperature, D: snow, E: polar. A 12-km CONUS model resolution climate zone classification map (see Appendix A in user's guide) is also created

using the Spatial Allocator based on the county level text file climate zone documentation as the surrogate (http://koeppen-geiger.vu-wien.ac.at/data/KoeppenGeiger.UScounty.txt).

The mapping algorithm provided in Table A2 (Appendix in main paper) helps to get Microbial/Organic biomass C and N % and HONO/N_{NOx} % for all 40 NLCD biome types (Table A3 in Appendix of main paper) corresponding to 1-24 category type names as listed in Table A1 (Appendix in main paper). For specific modeling domain implementation in future, re-gridding will be required due to the fine texture details.

3.2 SMCRESFILE/ SMNRESFILE, CNRATIOFILE and PHTOPSOILFILE: Soil organic Carbon/Nitrogen map

The input files **$SMCRESFILE, $SMNRESFILE, $CNRATIOFILE** and **$PHTOPSOILFILE** provide the soil organic carbon and nitrogen pools, microbial C:N ratio and soil pH respectively over the simulation domain. These inputs are used in the mechanistic module to drive the C and N cycling based on Schimel and Weintraub (2003) for non-agricultural regions, where EPIC is not applicable (See Section 2.6.3 in the main paper for detail). For the modeling grid non-agricultural grids can be classified from 'LANDFRFILE' and MODIS to NLCD mapping algorithm coded in mechanistic scheme source codes.

Users can re-grid these static field into their requisite designed modeling domain.

3.3 FERTRESFILE: EPIC fertilizer and soil properties data

In the mechanistic module for agricultural regions (**Ncycling.F**), the environment variable $FERTRESFILE provides the path of the file with the daily nitrogen input from fertilizer implementation, in addition to the base emission from non-agricultural soil biome (which is computed in **CNnonAgcycling.F**). Integration of the mechanistic module with the dynamic fertilizer fields from Environmental Policy Integrated Climate (EPIC) fertilizer model (Cooter et al. 2012) provides flexibility for future implementation as well, this integration serves as one possible state-of-the-science approach for enabling dynamic representation of year-specific fertilizer fields (see Section 2.6.3 in main paper).

3.4 NDEPRES: in-line nitrogen deposition

To provide continuous deposition N fields, the wet and dry deposition routines (**cldproc_acm.F** and **vdiffacm2.F**) call specific subroutines from BDSNP_MOD_EPIC.F that provide updated value for specific time step and spatial grid. In BDSNP module (BDSNP_MOD_EPIC.F), subroutine $GET_NDEPRATE estimates the deposition rate of the appropriate species for the appropriate time step, by considering both the addition to reservoir and decay as per Rasool et al. (2016). Subroutine $GET_N_DEP estimates the deposition rate for N species specific to a grid and time-step. Speciated $NDEPRES is added to corresponding speciated inputs from agricultural fertilized N inputs ($FERTRESFILE) or from non-agricultural base biome contribution (non-agricultural grids in $SMNRESFILE). This is an update over how N deposition is used to calculate soil NO flux as per the BDSNP algorithm by Hudman et al. (2012), as it took N deposition as total rather than speciated thus ignoring the actual biogeochemical transformations. $SOILINSTATE file also stores the resulting $NDEPRES as a diagnostic and can be used as a restart file to avoid cold run of model.

3.5 GRIDDESC: domain description

The environmental variable $GRIDDESC gives the paths of text input files for the grid description on the configuration of designed simulation domain. The format of $GRIDDESC is the same as CMAQ setting in the IO/API library (https://www.cmascenter.org/ioapi/documentation/3.1/html/GRIDDESC.html).
$GRIDDESC in this case of in-line mechanistic scheme with CMAQ is executed with the CCTM executable only.

For this paper, the domain configuration is set as following:
' '

'LAM_040N097W'

 2   33.000   45.000   -97.000   -97.000   40.000
' '

'12US1WRF_CMAQ'

'LAM_040N097W' -2556000.000 -1728000.000   12000.000   12000.000 459 299   1
' '

Where, in line 2, 'LAM_040N097W' is the logic name in I/O API library for the domain projection configuration. In line 3, '2' is the $GDTYP3D in I/O API library to determine the projection type, which in here is Lambert Conformal Conic. '33.000' and '45.000' are corresponding with the $P_ALP3D and $P_BET3D in I/O API library to determine the 1st and 2nd true latitude. The first value '-97.000' in line 3 is the $P_GAM3D parameter in I/O API library to determine the central meridian longitude (with the West in negative value). The second value '-97.000' in line 3 is the $XCENT3D in I/O API library, which represents the center longitude. Correspondingly, '40.000' is the $YCENT3D in I/O API library for the center latitude.

The value '12US1WRF_CMAQ' in line 5 is the logic name $GDNAM3D (or $GRID_NAME) for the simulation domain name in I/O API library. In line 6, the values '-2556000.000' and '-1728000.000' are corresponding to the $XORIG3D and $YORIG3D in I/O API library to determine the distance in meters (m) from the center point of domain the lower left grid point at x (column) and y (row) directions. The values '12000.000' and '12000.000' are corresponding to the $XCELL3D and $YCELL3D in I/O API library to set the domain resolution at each modeling gird in meters. The values '459', 299', '1' are corresponding to the $NCOLS3D, $NROWS3D, $NTHIK3D in I/O API library to set the 3D modeling grid number in the x (column), y (row), and z (layer) dimension respectively.

3.6 METCRO2D: meteorology file

For our in-line mechanistic scheme implementation, $METCRO2D that provides the path of meteorological variables is already a part of CCTM inputs (https://www.airqualitymodeling.org/index.php/CMAQ_version_5.1_(November_2015_r elease)_Technical_Documentation:

http://www.airqualitymodeling.org/cmaqwiki/index.php?title=CMAQ_version_5.0_(Febr uary_2010_release)_OGD).

Environmental variable $METCRO2D refers to the METCRO2D MCIP output file. The meteorology file is in NetCDF format and is normally run by the meteorological model (e.g., WRF) and MCIP. A total of nine meteorological variables are mainly needed in the mechanistic module:

(1) Surface temperature (in K) at 2-m, used to calculate canopy reduction factor: $Canopy\ Reduction\ (LAI, biome, met)$ and also used in DAYCENT nitrification and denitrification algorithm in mechanistic source codes

(2) Ground reaching solar radiation (in W/m^2), used to calculate canopy reduction factor

(3) Surface pressure (in Pa), used to calculate canopy reduction factor

(4) Soil moisture (in m^3/m^3), used in DAYCENT nitrification and denitrification algorithm in mechanistic source codes

(5) Soil temperature (in K), used in DAYCENT nitrification and denitrification algorithm in mechanistic source codes

(6) Surface snow cover, used to calculate canopy reduction factor

(7) Wind speed (in m/s) at 10m, used to calculate canopy reduction factor

(8) Cloud fraction in modeling grid, used to calculate canopy reduction factor

(9) Leaf area index, used to calculate canopy reduction factor

It should be noted that, in this version of the model, the Pleim-Xiu land surface model (2003) should be selected when running the meteorological model (e.g., WRF) and MCIP to provide the diagnosed soil temperature and moisture fields and other meteorological parameters to get pulsing term (PULSE) and Canopy reduction factor (CRF_TERM) for a given grid and time step.

3.7 SOILINP (SOILINSTATE in BDSNP): restart and diagnostic file

Environmental variables $SOILINP are identical to $SOILINSTATE used in BDSNP implementation in CMAQ (Rasool et al., 2016) and point to the path of the restart file generated by the mechanistic module (see Appendix B in user's guide for the demo run script setting changes). For previous day's soil moisture, pulsing factor, and dry period

information (which were also there in case of BDSNP in-line), it also contains information on nitrification rate, denitrification rate, NH_4, NO_3 and N_2O emission, all as diagnostic variables (N_2O emission specifically) for further model verification and analysis. Table 1 lists all four restart variables and nine diagnostic variables and their corresponding attributes. Some output variables provide the breakdown of each process in BDSNP (Rasool et al., 2016) to impact the final soil NO emission estimates, namely like the biome base emission (A_DIAG) and fertilizer implement amount (AFERT_DIAFG) and are obsolete for mechanistic scheme case. Variables like soil nitrogen pool from deposition (NDEPRES), pulse factor (PULSEAVG) and the canopy reduction factor (CRFAVG) are used in both in-line BDSNP and mechanistic schemes in CMAQ. The pulse factor depends on soil moisture and the antecedent dry period. The scaling factor CRFAVG represents the loss of NO and HONO to the plant canopy before entering the broader atmosphere.

3.8 B3GTS_S: soil NO and HONO emission output file

The environment variable $B3GTS_S in tmpbeis.F biogenic emission estimation routine, is used to set the output path and file name that includes the hourly soil NO and HONO estimations from mechanistic soil N module to update the total NO and HONO estimates. The user can choose the output unit for the soil NO and HONO emission rate in terms of mass (gm/s). Estimates from $B3GTS_S are eventually utilized in CMAQ to estimate secondary pollutant concentrations like ozone and particulate matter. For our simulation period of May and July 2011, the pollutant concentration hot spots correlated well with locations that have high soil NO and HONO emissions (e.g. high fertilizer application cropland region across Midwest U.S. extending from North Dakota to Texas), or high N deposition (e.g. central California).

4. Create specific case run using mechanistic soil N module

To run a specific case using the mechanistic module, the user needs to prepare the entire CMAQ model domain consistent input files including but not just limited to those required for WRF-CMAQ framework typically. After preparing all the needed input files, the user needs to modify the run script file (see Appendix B in user's guide for demo run script modifications) to accommodate the changes for those key I/O environment variable settings and make sure all the inputs are located at the correct paths. The exercise of procuring inputs for the WRF-SMOKE/MEGAN-CMAQ platform with mechanistic scheme also included in-line is quite computationally extensive and varies with different user requirement for modeling scenarios (domain and time period). The user can run the model eventually by typing:

>$ cd $M3HOME/scripts/cctm
>$./run.CMAQwithMechanisitc_soil_N_[your_case].csh &   C script name is user defined

If it runs successfully, the diagnostic outputs $SOILINP, and new HONO environment variable in $B3GTS_S along with other CCTM output files should be generated and the log files for the simulation can be checked for completion status or any warnings or errors.

Table 1 List of output variables from SOILINSTATE (as called for BDSNP)/SOILINP (as called for mechanistic scheme (diagnostic and restart file)

| Output Variable	Unit	Variable Description	Notes
PFACTOR	--	NO emission current pulse factor	
DRYPERIOD	hr	Length of the dry period	*used for continuous BDSNP and new mechanistic scheme run
NDEPRES	Ng/m2	Soil N reservoir from deposition	*used for continuous BDSNP and new mechanistic scheme run
SOILMPREV	m3/m3	Soil moisture for previous time step	*used for continuous BDSNP and new mechanistic scheme run
THETA_DIAG	--	Moisture WFPS with value [0 1]	*diagnostic parameter for BDSNP (WFPS is calculated in-line in the new mechanistic scheme)
WET_TERM_DIAG	--	Moisture scale factor	*diagnostic parameter for BDSNP
TEMP_DIAG	K	Temperature in last simulation hour	*diagnostic parameter for BDSNP
A_DIAG	--	Base emission from soil biome type	*diagnostic parameter for BDSNP
NRES_FERT_DIAG	--	NRES fertilizer only	*diagnostic parameter for BDSNP
AFERT_DIAG	--	Fertilizer emission factor	*diagnostic parameter for BDSNP
NDEPRATE_DIAG	--	Daily average N deposition rate	*used for continuous BDSNP and new mechanistic scheme run
CRFAVG	--	Daily average canopy reduction factor	*diagnostic parameter in both BDSNP and mechanistic
PLUSEAVG	--	Daily average pulse factor	*diagnostic parameter in both BDSNP and mechanistic
N2O_EMIS	g-N/hr	N_2O emission from mechanistic scheme routines	*diagnostic parameter in mechanistic only
NH4_Pool, NO3_Pool	Kg-N/ha	Updated soil NH_4 and NO_3 pools	*diagnostic parameter in mechanistic only
NITR_Rate, DENITR_Rate	Kg-N/ha/s	Nitrification and denitrification rates	*diagnostic parameter in mechanistic only
			*diagnostic parameter in mechanistic only

Reference

Bash, J., Cooter, E., Dennis, R., Walker, J., and Pleim, J.: Evaluation of a regional air-quality model with bidirectional NH 3 exchange coupled to an agroecosystem model, Biogeosciences, 10, 1635-1645, 2013.

Cooter, E. J., Bash, J.O., Benson, V., and Ran, L. (2012) Linking agricultural crop management and air quality models for regional to national-scale nitrogen assessments. *Biogeosciences*, 9, no. 10: 4023-4035.

Hudman, R. C., Russell, A.R., Valin, L.C., and Cohen, R.C. (2010) Interannual variability in soil nitric oxide emissions over the United States as viewed from space. *Atmospheric Chemistry and Physics,* 10**,** 9943-9952.

Hudman, R. C., Moore, N. E., Mebust, A. K., Martin, R. V., Russell, A. R., Valin, L. C., and Cohen, R. C. (2012) Steps towards a mechanistic model of global soil nitric oxide emissions: implementation and space based-constraints, *Atmospheric Chemistry and Physics*, 12, 7779-7795, doi:10.5194/acp-12-7779-2012.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F. (2006) World Map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitshrift*, 15, 259-263. DOI: 10.1127/0941-2948/2006/0130.

Pleim, J. E and A. Xiu (2003). Development of a land surface model. Part II: data assimilation. *Journal of Applied Meteorology*, 42(12), 1811-1822.

Potter, P., Navin, R., Elena M. B., and Simon D. D. (2010) Characterizing the spatial patterns of global fertilizer application and manure production. *Earth Interactions*, 14, no. 2: 1-22.

Rasool, Q.Z., R. Zhang, B. Lash, D.S. Cohan, E. Cooter, J. Bash, L.N. Lamsal. Enhanced representation of soil NO emissions in the Community Multi-scale Air Quality (CMAQ) model version 5.0.2. Geoscientific Model Development, 9, 3177-3197, 2016.

Schimel, J. P. and Weintraub, M. N.: The implications of exoenzyme activity on microbial carbon and nitrogen limitation in soil: a theoretical model, Soil Biology and Biochemistry, 35, 549-563, 2003.

Stehfest, E. and Bouwman, L. (2006) N_2O and NO emission from agricultural fields and soils under natural vegetation: summarizing available measurement data and modeling of global annual emissions. *Nutrient Cycling in Agroecosystems*, 74(3), 207-228. doi: 10.1007/s10705-006-9000-7.

Steinkamp, J. and Lawrence, M. G. (2011) Improvement and evaluation of simulated global biogenic soil NO emissions in an AC-GCM. *Atmospheric Chemistry and Physics*, *11*(12), 6063–6082. doi:10.5194/acp-11-6063-2011.

Wang, Y., Jacob, D. J., and Logan, J. A. (1998) Global simulation of tropospheric O3-NOx-hydrocarbon chemistry: 1. Model formulation. *Journal of Geophysical Research-Atmosphere*s, 103, 10713–10725.

Xu, X., P.E. Thornton, and W.M. Post. 2014. A Compilation of Global Soil Microbial Biomass Carbon, Nitrogen, and Phosphorus Data. Data set. Available on-line [http://daac.ornl.gov] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA. http://dx.doi.org/10.3334/ORNLDAAC/1264.

Yienger, J. J. and Levy, H. (1995) Empirical-model of global soil-biogenic NOx emissions. *Journal of Geophysical Research-Atmosphere*s, 100(D6), 11447-11464. doi: 10.1029/95jd00370.

Appendix

Appendix A. Spatial map of five climate zone over CMAQ CONUS 12km domain based on Köppen-Geiger climate classification (KOPPENFILE)
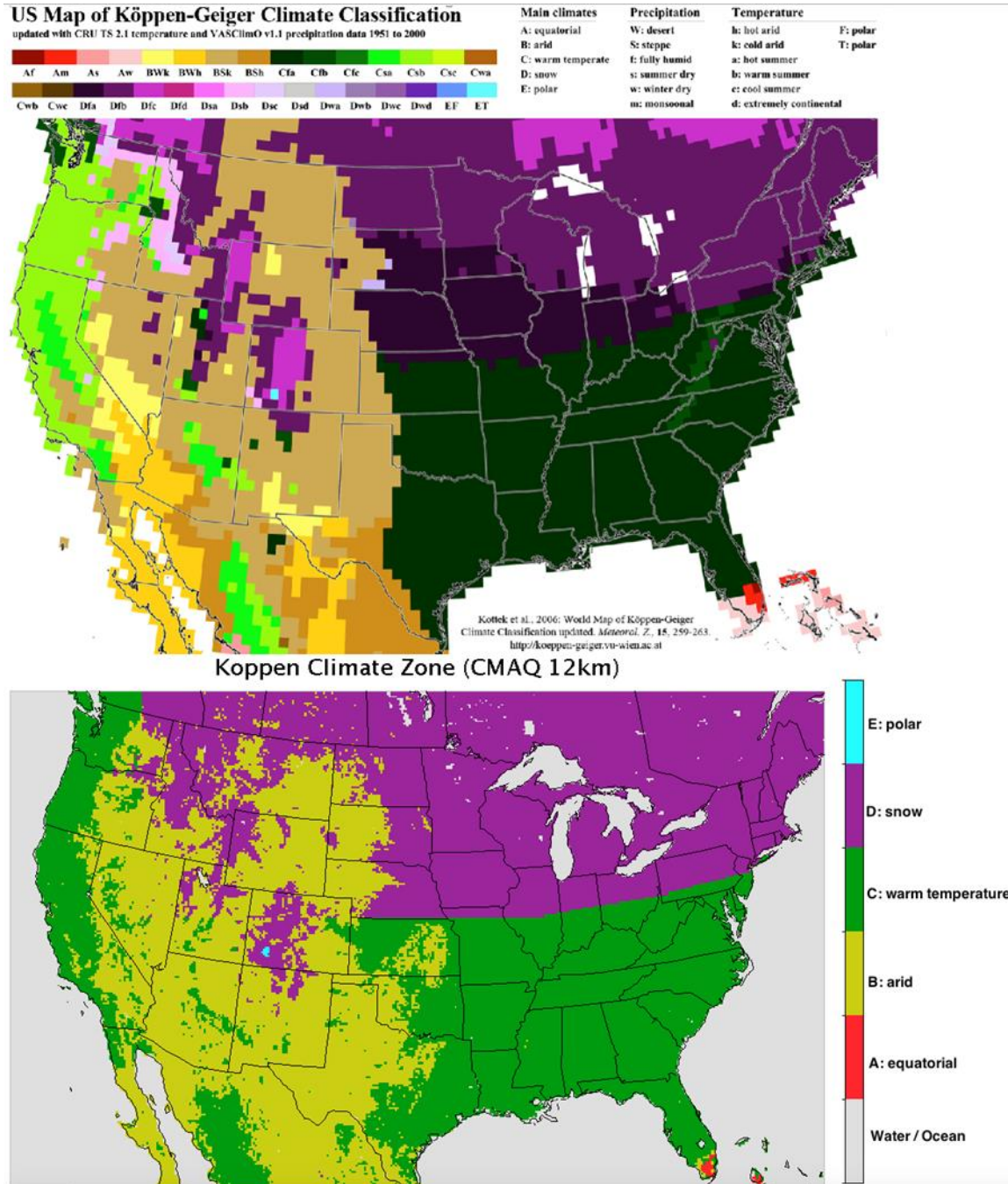
Appendix B. Sample run script modifications for mechanistic soil N module

```
#!/bin/csh
#
# In-line mechanistic Soil N module implemented in CMAQ v 5.1
#

# CHANGE user-defined run.cctm script to include mechanistic module
# added flexibility to choose different input files and implementation schemes
# [ ] indicate default prompt for an environmental variable i.e. Y|N
# Both, BDSNP and EPIC are turned Y, as mechanistic scheme also uses
BDSNP_MOD_EPIC.F source code

#> BDSNP_MOD_EPIC.F used by mechanistic scheme to pass soil NO and HONO into
B3GTS_S
#Default CMAQ-YL95 turn [N|F ]
   setenv BDSNP Y
 #> use EPIC fertilizer data # Defau;t CMAQ – YL95 turn [N|F ]
   setenv EPIC Y

# Other USER DEFINED! Same as in base CMAQ #

# Grid definition
setenv GRIDDESC  $GRIDDESCHOME/GRIDDESC
setenv GRID_NAME  12US1WRF_CMAQ
# Episode chosen in Julian date
set STJD = 201178  #Start Date
set ENJD = 201178  #End Date

## Loop between the start and end of episode period for in-line mechanistic scheme in
CMAQ

set CJD = $STJD
while ( $CJD <= $ENJD ) #demo, loop set-up is user-defined


#> beis mass emissions diagnostic file [N|F]
   setenv B3GTS_DIAG Y       # updated if mechanistic scheme or BDSNP is used

#Outputs biogenic total NO, HONO and soil NO, HONO diagnostics
if ( $biogon ) then
    if ( $B3GTS_DIAG == 'Y' || $B3GTS_DIAG == 'T' ) then
      setenv B3GTS_S $OUTDIR/$B3GTS_S
    endif
    setenv SOILINP $OUTDIR/$SOILINP
```

```
#SOILINP: Restart/diagnostic file for soil N2O in mechanistic scheme
    # B3GTS_S: Biogenic NO and HONO soil output file
   endif

 if ( $?BDSNP ) then
  if ( $BDSNP == 'Y' || $BDSNP == 'T' ) then
   echo "BDSNP true, setting mechanistic scheme data paths"

   # Soil biome and climate zone map data input

   setenv LANDFRACFILE  $MECHINPUTS /BIOME_CMAQ_LANDFRAC.nc
   setenv KOPPENFILE  $MECHINPUTS /BIOME_CMAQ_CLIMATEZONE.nc

# Soil nutrient and properties map data input for CNnonAgcycling.F

   setenv SMCRESFILE  $MECHINPUTS /SMCRESFILE.nc

   setenv SMNRESFILE  $MECHINPUTS /SMNRESFILE.nc

   setenv SMCRESFILE  $MECHINPUTS /SMCRESFILE.nc

   setenv SMCRESFILE  $MECHINPUTS /SMCRESFILE.nc



   # Fertilizer input from EPIC

    if ( $?EPIC ) then
    if ( $EPIC == 'N' || $EPIC == 'F' ) then
      setenv EPICRESFILE $MECHINPUTS/EPIC_SOILFERT12US1_$date".CONUS
     endif
   endif

  endif
 endif

# Restart file SOILINSTATE/SOILINP
#> Set Soil NO to zero? [N|F]
   setenv ZEROSOIL N

   set APPL = ${YEAR}${MONTH}${DAY1}

   if ( $APPL == $FIRSTDAY ) then   #Cold Run

   # Use archive nitrogen pool already accumulated do not from zero!!!

     echo "Firstday =  $APPL"
```

```
        setenv SOILINP $OUTDAT/SOILINP.${DATE}

        # Biogenic soil N input file

        setenv INITIAL_RUN N

    else if ( $IMONTH == 12 ) then

        setenv SOILINP $OUTDAT/SOILINP.${DATE}

        # Biogenic soil N input file

        setenv INITIAL_RUN N

    else
        setenv SOILINSTATE $OUTDAT/SOILINSTATE.${DATE}
        # Biogenic soil N input file

        setenv INITIAL_RUN N

    endif
 endif

# MCIP input directory already included in base CMAQ run.cctm script

    setenv MET_CRO_2D $METpath/METCRO2D.${EXTN}

    setenv MET_CRO_3D $METpath/METCRO3D.${EXTN}

    setenv MET_DOT_3D $METpath/METDOT3D.${EXTN}

    setenv MET_BDY_3D $METpath/METBDY3D.${EXTN}

# Nitrogen deposition not provided explicitly, changes in CCTM and mechanistic source
code, which handles it

############################################################################
```