

**BDSNP Soil NO module
(In-line with CMAQ v5.0.2)**

User's operational guidance manual

**Supplementary Material for the GMD manuscript “Enhanced
representation of soil NO emissions in the Community Multi-
scale Air Quality (CMAQ) model”**

Table of Contents

1. Introduction	4
1.1 BDSNP module in-line implementation with CMAQ: Overview	4
1.2 BDSNP module: software and operating system prerequisites.....	6
2. In-line BDSNP module implementation	7
2.1 Introduction.....	7
2.2 Program structure for in-line BDSNP module implementation	7
2.3 Build new CCTM executable with in-line BDSNP module.....	10
3. Key I/O environmental variables specific to BDSNP module	11
3.1 Key I/O environment variables in BDSNP module.....	11
3.1.1 LANDFRACFILE: soil biome type map.....	12
3.2.2 CLIMAFILE/CLIMNAFILE: arid/non-arid climate map	13
3.2.3 FERTRESFILE: fertilizer map.....	13
3.2.4 NDEPRES: archive nitrogen deposition file.....	14
3.2.5 GRIDDESC: domain description.....	14
3.2.6 METCRO2D: meteorology file.....	16
3.2.7 SOILINSTATE/SOILOUT: restart file.....	17
3.2.8 B3GTS_S: soil NO emission output file	18
4. Create specific case run using BDSNP module.....	18
5. Enhanced BDSNP module (EPIC with new biome or BDSNP ('new')).....	19
5.1 Change emission factor for soil biome	19
5.2 Replace fertilizer file with EPIC outputs.....	19
Reference	21
Appendix	22
Appendix A. Spatial map of five climate zone over CMAQ CONUS 12km domain based on Köppen-Geiger climate classification.....	22
Appendix B. Sample run script modifications for BDSNP soil NO module	23
Appendix C. Sample of soil biome emission factor setting include file	27

List of Tables

Table 1 <i>List of output variables from SOILINSTATE (restart file)</i>	20
-------------------------------------------------------------------------------	----

1. Introduction

This guide instructs potential users to build and run the Berkeley-Dalhousie Soil NO_x Parameterization (BDSNP) module in-line with CMAQ version 5.0.2, to simulate NO 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 BDSNP module apt for a first time user. Comments on how to improve the documentation are encouraged.

1.1 BDSNP module in-line implementation with CMAQ: Overview

The BDSNP previously has been implemented in a global chemical transport model (GEOS-Chem) to update the soil nitrogen oxide emission estimates (Hudman et al., 2010, 2012). Compared to the soil NO emission module (Yienger and Levy 1995; YL95) currently coded in biogenic emission models such as BEIS3 and MEGAN, the BDSNP scheme has a more continuous soil temperature/moisture dependence function.

Based on the BDSNP scheme, soil NO emissions at each modeling grid and time-step are determined by the base emission varied by the biome type and the available soil nitrogen (either form the fertilizer implementation during the growing season or the nitrogen dry/wet deposition from the atmosphere), the soil temperature and moisture response function, the soil pulsing after precipitation, and the canopy reduction due to the resistance. The algorithm for canopy reduction was proposed by Wang et al. (1998) in GEOS-Chem to consider a more physical basis based on wind speed, turbulence, canopy structure, deposition constants and other physical variables.

$$NO_{emission\ flux} = A(\text{biome, soil nitrogen}) \times f(T) \times g(\theta) \times pluse(\text{dry period}) \times \text{Canopy Reduction (LAI, biome, met)}$$

This work proposes an implementation of the BDSNP code originally implemented in GEOS-Chem, into the Community Multiscale Air Quality (CMAQ) modeling system and

produced the online soil NO 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 3.6.1. However, the inline BDSNP module proposed here is specifically implemented with CMAQ, is computationally expensive and is not cross-platform applicable. For example, BDSNP has not been implemented in the CAMx air quality model.

From the software engineering perspective, Figure 1 in the main paper provides the flow chart of the BDSNP scheme implementation. Static input files such as arid/non-arid climate zone, soil biome type (consistent with the type of the soil emission factors given by Steinkamp and Lawrence (2011)) and fertilizer pool (Potter et al. (2010) global static data) are needed to determine the soil base emission value at each modeling grid.

Soil moisture and temperature as well as some meteorological variables such as radiation, wind and air pressure are needed (provided from WRF (ARW v 3.6.1) model) to drive the BDSNP to represent the occurrence of pulsing and the process of canopy reduction.

For the in-line BDSNP module, 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. An offline (stand-alone) version of BDSNP module (stand-alone BDSNP) is also implemented to give quick diagnostics (Supplementary material Section S.3 and Figure S4).

The major difference with the option of in-line and standalone or offline BDSNP model is the approach to deal with the available nitrogen pool from the soil. Since soil nitrogen pool needs a long time to build up to reach the quasi-steady state in the model, a new series of SOILINSTATE files are needed in the stand-alone BDSNP model to provide the generic daily variation of N deposition. Here, we performed a full year CMAQ simulation over the 12 km continental U.S. modeling domain in 2005 and recorded their daily total soil N reservoir from deposition for further usage. Simulating soil NO emissions for a region outside North America or for a period with sharply different N

deposition rates from 2005 will require development of a new set of SOILINSTATE files. However, it should be noted that in our in-line CMAQ simulations of soil NO, the soil N reservoir from deposition is responsible for only about 10 percent of emissions.

The inline BDSNP module with CMAQ initially ('old') used the soil biome map directly re-gridded from global atmospheric chemistry model GEOS-Chem, which is too coarse for regional model implementation. The biome map is related with the land use/ land cover (LU/LC) classification and Köppen-Geiger climate zone definition (Kottek et al., 2006) and will determine the base soil NO emission strength. A 'new' soil biome spatial map based on 4-km CONUS 40-category 2006 NLCD-MODIS land use classification (NLCD40) and climate zone definition is created to replace the GEOS-Chem biome map to better represent the up-to-date LU/LC change with finer details. The required mapping the 40 categories to the categories for which emission rates were reported by Steinkamp and Lawrence (2011).

1.2 BDSNP module: software and operating system prerequisites

The prerequisites in terms of compilers and libraries for running BDSNP module in-line with CMAQ are the same as that for CMAQ version 5.0.2 itself, mentioned in detail in the model's instructional page

(http://www.airqualitymodeling.org/cmaqwiki/index.php?title=CMAQv5.0.2_Readme_file#CCTM). The WRF/MCIP-BDSNP-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 BDSNP in-line with CMAQ v 5.0.2 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 understand basic Shell script to invoke FORTRAN executables).

For any other custom run re-gridding of BDSNP input files to a preferred simulation domain will be required. For this knowledge and access to programming tools like, Python and/or software like ESRI ArcMAP software or an open-source software such as the spatial allocator (www.ie.unc.edu/cempd/projects/mims/spatial/) will be required.

2. In-line BDSNP module implementation

2.1 Introduction

The BDSNP module is simultaneously compiled with other executables of the CMAQ Chemistry-Transport Model (CCTM). CCTM integrates all 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 BDSNP module in-line with CCTM ensures that soil NO as per Hudman et al. (2012) with suitable enhancements (BDSNP ('new')) is also accounted for generating gridded and temporally resolved air pollutant concentrations as output. This is fairly straightforward on the supported Linux platforms. To satisfy the pre-requisites for libraries and packages on which CMAQ depends and their installation first step should be referring to the simple steps given in CMAQ's technical documentation ([http://www.airqualitymodeling.org/cmaqwiki/index.php?title=CMAQ_version_5.0_\(February_2010_release\)_OGD#Installing_CMAQ_on_your_system](http://www.airqualitymodeling.org/cmaqwiki/index.php?title=CMAQ_version_5.0_(February_2010_release)_OGD#Installing_CMAQ_on_your_system)).

2.2 Program structure for in-line BDSNP module implementation

First step is to unzip and untar the compressed file containing CMAQ version 5.0.2 source code (CMAQv5.0.2.Apr2014.tar.gz) available at <https://www.cmascenter.org/> with applicable free registration.

```
>$tar -xzf CMAQv5.0.2.Apr2014.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 EMPOL) 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, TASFC, EMPOL )
```

In this work, script tmpbeis.F has been modified to give the option to calculate soil NO emission either from Yienger and Levy (1995) parametrization or from BDSNP. For this purpose a new BDSNP_MOD_EPIC.F source code has been added as a global attribute in the CCTM source code repository, which basically gives the soil NO parametrization as given by Hudman et al. (2012) with options of enhancement in terms of new soil biome and dynamic fertilizer N pool as well.

Canopy reduction factor was missing in the original implementation of Yienger-Levy (1995) scheme in CMAQ. Hence, another new globally defined source code

canopy_nox_mod.F has also been created for incorporating Canopy reduction of soil NO as proposed by Hudman et al. (2012).

Similarly, we modified the deposition (wet and dry) source codes in \$BLD to include wet and dry deposition of N species for our application. 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. BDSNP_MOD_EPIC.F encompasses the parameterization that accounts for emissions from biome classification, fertilizer, deposition, 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) and model-ready temporal biogenic emission (tmpbeis.F) source codes and including new ones for the BDSNP module (BDSNP_MOD_EPIC.F) and canopy reduction factor (canopy_nox_mod.F) 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 and new BDSNP_MOD_EPIC.F and canopy_nox_mod.F 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 BDSNP module is computationally extensive as it will require a wide array of inputs (http://www.airqualitymodeling.org/cmaqwiki/index.php?title=CMAQ_version_5.0_%28February_2010_release%29_OGD#CMAQ_Input_Files).

But there are input files specific for BDSNP module, broadly four of them, defined in sub-routines provided for their inclusion in BDSNP_MOD_EPIC.F: Deposition, Fertilizer, 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 BDSNP 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 BDSNP 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 BDSNP:

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

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

Change the lines in 'Makefile' to include the path of modified deposition and biogenic emission schemes to call the new 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 \
              VDIFF_MAP.F  UTILIO_DEFN.F  BIDI_MOD.F  LSM_MOD.F
BDSNP_MOD_EPIC.F \
              $(BASE_INC)/$BLD /CONST.EXT \
              $(BASE_INC)/$BLD /FILES_CTM.EXT $(BASE_INC)/$BLD/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 \
              $(BASE_INC)/$BLD /CONST.EXT \
              $(BASE_INC)/$BLD /FILES_CTM.EXT
```

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

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

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

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_BDSNP &
```

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_BDSNP 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 BDSNP module

3.1 Key I/O environment variables in BDSNP module

Understanding the input/output (I/O) environment variables defined in BDSNP module is crucial for the real case implementation and further model development if needed. There are seven key input environment variables and two key output environment variables in current version of soil NO BDSNP module implemented in CMAQ v 5.0.2. Table S.1 (Supplementary material) summarizes their names and corresponding functionalities. Each I/O environment variable is generally elaborated in sequence for following subsections.

3.1.1 LANDFRACFILE: soil biome type map

Based on several field studies for soil NO flux (SNO_x) measurement (Stehfest and Bouwman, 2006), different biome type are found to have different NO emission factors under different soil conditions (wet or dry) over different sampling geographic locations (or climate zones). Steinkamp and Lawrence (2011) compiled those worldwide emission factors (a dataset consisting of 112 articles with 583 field measurements of SNO_x covering the period from 1976 to 2010) and regrouped them into 24 soil biome type base on MODIS land cover category as well as Köppen climate zone classifications (Kottek et al., 2006) and reported them into the mean values (arithmetic or geometric). Those soil NO emission factors (wet) implemented in the current version of BDSNP module are listed in Table A1 (Appendix in main paper). For old GEOS-Chem biome map i.e. BDSNP ('old') implementation in the paper, EF1 from Table A1 are to be embedded in the BDSNP_MOD_EPIC.F code as A_BIOME(NSOIL) for each of the 24 soil biome types.

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 (Appendix A2 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 5.0.2 source code release (http://www.airqualitymodeling.org/cmaqwiki/index.php?title=CMAQv5.0.2_CSQY_DATA_Files). To enable the biome type classification based on soil NO emission factors, a model resolution compatible Köppen climate zone classification (Kottek et al., 2006) is utilized to allocate appropriate emission factor for the same biome type (e.g. 'grassland') at different locations. 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>).

Figure 2 in main paper provides the spatial pattern of the 24 soil biome map in 12 km resolution over CONUS domain. The corresponding 1-24 category type names are provided in Table A2 (Appendix in main paper). For specific modeling domain implementation in future, regridding will be required due to the fine texture details.

3.2.2 CLIMAFILE/CLIMNAFILE: arid/non-arid climate map

The input files \$CLIMAFILE and \$CLIMNAFILE provide the arid and non-arid region classification respectively over the simulation domain, which are used in the BDSNP module to determine the soil moisture scaling factor $g(\theta)$ for soil NO emissions. For the modeling grid classified as ‘arid’ region, the maximum moisture scaling factor corresponds with the water-filled pore space (θ) value equal to 0.2; while for the ‘non-arid’ modeling grid, the maximum moisture scaling factor corresponds with $\theta = 0.3$ (Hudman et al., 2012).

The users can regrid these static field into their requisite designed modeling domain. Figure B1 (Appendix in main paper) shows the distribution of arid/non-arid climate map over continental US with the red color corresponding with the arid region while blue color corresponding with the non-arid region.

3.2.3 FERTRESFILE: fertilizer map

In the BDSNP module (BDSNP_MOD_EPIC.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 soil biome (which is provided as environmental variable A_BIOME). Hudman et al. (2012) used a global fertilizer database from Potter et al. (2010) and assumed 37% of fertilizer and manure N is available (1.8 Tg N yr^{-1}) for potential emission. \$FERT environmental variable provides that regridded for our 12 km CONUS domain for BDSNP (‘old’) implementation. Figure B2 (Appendix in main paper) provides the day-by-day variation of total fertilizer implementation over CONUS during a year and shows the typical cycle between growing season and non-growing season.

Without updating, any parameterization of soil NO based on fertilizer will be out of date within a decade. Hence, integration of the BDSNP with the dynamic fertilizer fields from Environmental Policy Integrated Climate (EPIC) fertilizer model (Cooter et al. 2012) is 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 5.2).

3.2.4 NDEPRES: archive nitrogen deposition file

To provide continuous deposition N fields, the wet and dry deposition routines (cloudproc_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 Hudman et al. (2012). Subroutine \$GET_N_DEP estimates the deposition rate for N species specific to a grid and time-step. \$NDEPRES is added to \$FERT to give total N available apart from base biome contribution, defined as \$A_FERT. \$A_FERT is used in calculating soil NO flux as per the algorithm by Hudman et al. (2012). \$SOILINSTATE file also stores the resulting \$NDEPRES as a diagnostic and can be used as a restart file to avoid cold run of model.

For the stand-alone module, the nitrogen available from wet or dry deposition cannot be calculated instantaneously through the air quality model but needs to be determined from offline, pre-computed files. The daily available soil N reservoir from deposition was recorded in the restart file \$SOILINSTATE. The path of environment variable \$NDEPRES should be set as the path of previous day's \$SOILINSTATE archived deposition file. For real case usage, the users need to regrid the deposition fields into the designed simulation domains and over the selected period as desired.

3.2.5 GRIDDESC: domain description

The environment 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 BDSNP with CMAQ is executed with the CCTM executable only. But for stand-alone BDSNP, it has to be provided separately like deposition fields.

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

```
''  
  
'12CONUS'  
  2   33.000   45.000  -97.000  -97.000   40.000  
''  
  
'12US2'  
'12CONUS' -2412000.000 -1620000.000  12000.000  12000.000 396 246 1  
''
```

where in line 2, '12CONUS' 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 '12US2' 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 '-2412000.000' and '-1620000.000' 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 grid in meters. The values '396', '246', '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.

3.2.6 METCRO2D: meteorology file

For our in-line BDSNP implementation, \$METCRO2D that provides the path of meteorological variables is already a part of CCTM inputs ([http://www.airqualitymodeling.org/cmaqwiki/index.php?title=CMAQ_version_5.0_\(February_2010_release\)_OGD](http://www.airqualitymodeling.org/cmaqwiki/index.php?title=CMAQ_version_5.0_(February_2010_release)_OGD)). Environment variable \$METCRO2D is however needed to be explicitly provided for the stand-alone BDSNP module. The meteorology file is in NetCDF format and is normally run by the meteorological model (e.g., WRF) and MCIP (see section 4.2). A total of nine meteorological variables are needed in the BDSNP module:

- (1) Surface temperature (in K) at 2-m, used to calculate canopy reduction factor: *Canopy Reduction (LAI, biome, met)*
- (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 to calculate the dry period and pulsing strength *pulse(dry period)* and scale the base emission by soil moisture response function: $g(\theta)$
- (5) Soil temperature (in K), used to scale the base emission by soil temperature response function: $f(T)$
- (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 noticed 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.

3.2.7 SOILINSTATE/SOILOUT: restart file

Environment variables \$SOILINSTATE and \$SOILOUT are identical and both point to the path of the restart file generated by the BDSNP module (see Appendix B in user's guide for the demo run script setting changes). It contains the previous day's soil moisture, pulsing factor, and dry period information for continuous run as well as several diagnostic variables for further model verification and analysis. Table 1 lists all four restart variables and nine diagnostic variables and their corresponding attributes. The output variables provides the breakdown of each process in stand-alone BDSNP to impact the final soil NO emission estimates, namely the biome base emission (A_DIAG), fertilizer implement amount (AFERT_DIAFG), soil nitrogen pool from deposition (NDEPRES), pulse factor (PULSEAVG) and the canopy reduction factor (CRFAVG). For each modeling grid, higher biome base emission plus higher fertilizer implementation and higher available N from deposition will result in higher potential to produce soil NO. The pulse factor depends on soil moisture and the antecedent dry period. The scaling factor CRFAVG represents the loss of NO to the plant canopy before entering the broader atmosphere.

One of the analyses that can be done with the diagnostic output variables is to calculate the fraction of NO emissions resulting from soil biome base emissions, fertilizer application, and N deposition. These fractions sum to 1.

The following formulas can be used to calculate those fractions:

$$F_{BIOME} = A_DIAG / (A_DIAG + AFERT_DIAG)$$

$$F_{FERT} = \frac{NRES_FERT_DIAG}{(NDEPRES + NRES_FERT_DIAG)} \times AFERT_DIAG / (A_DIAG + AFERT_DIAG)$$

$$F_{NDEP} = \frac{NDEPRES}{(NDEPRES + NRES_FERT_DIAG)} \times AFERT_DIAG / (A_DIAG + AFERT_DIAG)$$

3.2.8 B3GTS_S: soil NO 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 estimation from BDSNP module to update the total NO estimates. The user can choose the output unit for the soil NO emission rate either 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 July 2011, the pollutant concentration hot spots correlated well with locations that have either biome types with high base emission factors (e.g. evergreen broadleaf forest around Mississippi), high fertilizer application rates (e.g. cropland region near Iowa and Kansas), or high N deposition (e.g. central California).

4. Create specific case run using BDSNP module

To run a specific case using the BDSNP module, the user needs to prepare the entire CMAQ model domain consistent input files including but not just limited to those listed in Table S.1 (Supplementary material). 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 BDSNP 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.CMAQwithBDSNP_soil_NO_[your_case].csh & C script name is user defined
```

If it runs successfully, the diagnostic outputs \$SOILOUT or \$SOILINSTATE 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.

5. Enhanced BDSNP module (EPIC with new biome or BDSNP ('new'))

5.1 Change emission factor for soil biome

The default emission factors for soil biomes used in the current version of BDSNP module in-line with CMAQ are the Global geometric mean values reported by Steinkamp and Lawrence (2010) (see Appendix in main paper, 'EF1' in Table A1). If we need to use another set of emission factors like North American geometric mean values ('EF3' in Table A1) for sensitivity study, \$A_BIOME(NSOIL) needs to be edited and executables are to be recompiled. Another way without changing the source code and need to recompile the executables is to provide an external text include file named 'EF_BIOME.EXT' (see Appendix C in user's guide) for the ease of sensitivity study by simply changing the corresponding value for biome category.

5.2 Replace fertilizer file with EPIC outputs

The default daily fertilizer map is generated from the global database reported by Potter et al. (2010), which is relatively coarse resolution and the out-of-date information (national-level fertilizer data from 1994 to 2001). The alternatives we provided in our BDSNP('new') implementation is to use the Fertilizer Emission Scenario Tool for CMAQ (FEST-C, <http://www.epa.gov/amad/Tools/festC.html>) to consider the different cropland farm management system scenarios and to drive the Environmental Policy Integrated Climate model (EPIC) simulations. This provides for dynamic day-to-day fertilizer N fields over a finer user-defined resolution. Figure 4 (see main paper) provides the comparison of default fertilizer map and N fertilizer map from 12km EPIC outputs (considering the weighted sum of fertilizer application rate for ammonia 'T1_ANH3', nitrate 'T1_ANO3' and organic nitrogen 'T1_AON', provided as a single variable \$T1_EPIC or \$EPICN in BDSNP module) over continental US. Clearly, it has the similar pattern as Potter N inputs but represents more details and more up-to-date information.

Table 1 List of output variables from SOILINSTATE (restart file)

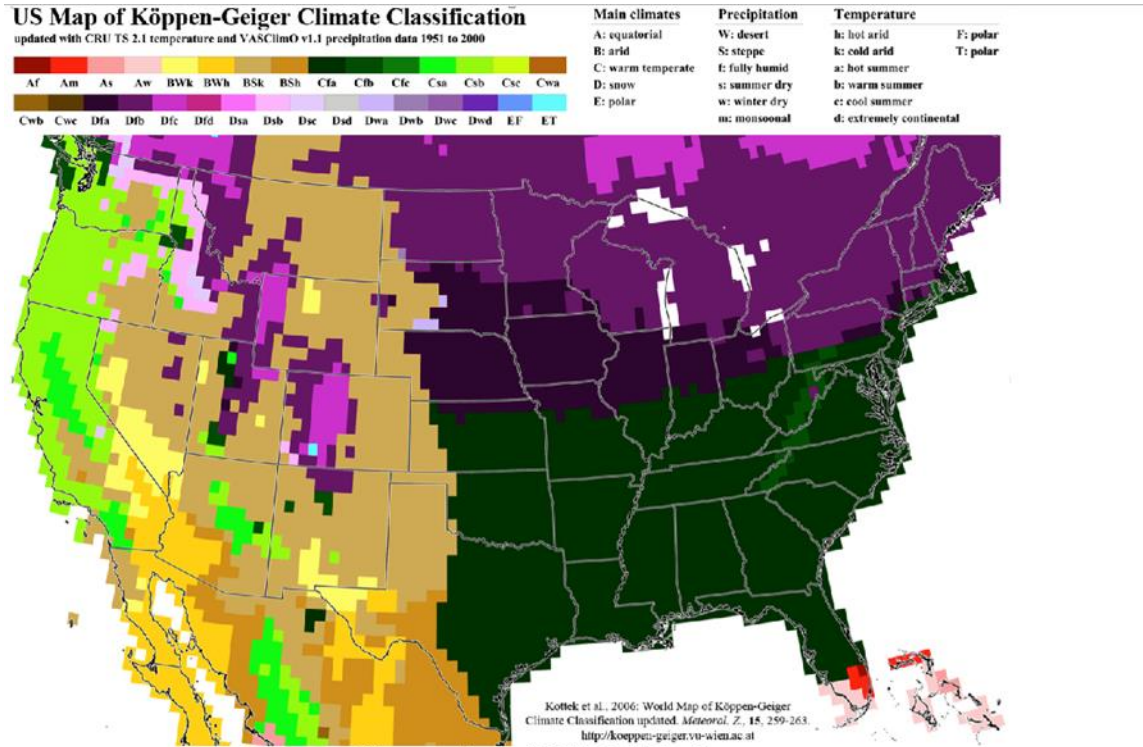
Output Variable	Unit	Variable Description	Notes
PFACTOR	--	NO emission current pulse factor	
DRYPERIOD	hr	Length of the dry period	*used for continuous BDSNP run
NDEPRES	Ng/m2	Soil N reservoir from deposition	*used for continuous BDSNP run
SOILMPREV	m3/m3	Soil moisture for previous time step	*used for continuous BDSNP run
THETA_DIAG	--	Moisture WFPS with value [0 1]	*diagnostic parameter
WET_TERM_DIAG	--	Moisture scale factor	*diagnostic parameter
TEMP_DIAG	K	Temperature in last simulation hour	*diagnostic parameter
A_DIAG	--	Base emission from soil biome type	*diagnostic parameter
NRES_FERT_DIAG	--	NRES fertilizer only	*diagnostic parameter
AFERT_DIAG	--	Fertilizer emission factor	*diagnostic parameter
NDEPRATE_DIAG	--	Daily average N deposition rate	*used for continuous BDSNP run
CRFAVG	--	Daily average canopy reduction factor	*diagnostic parameter
PLUSEAVG	--	Daily average pulse factor	*diagnostic parameter

Reference

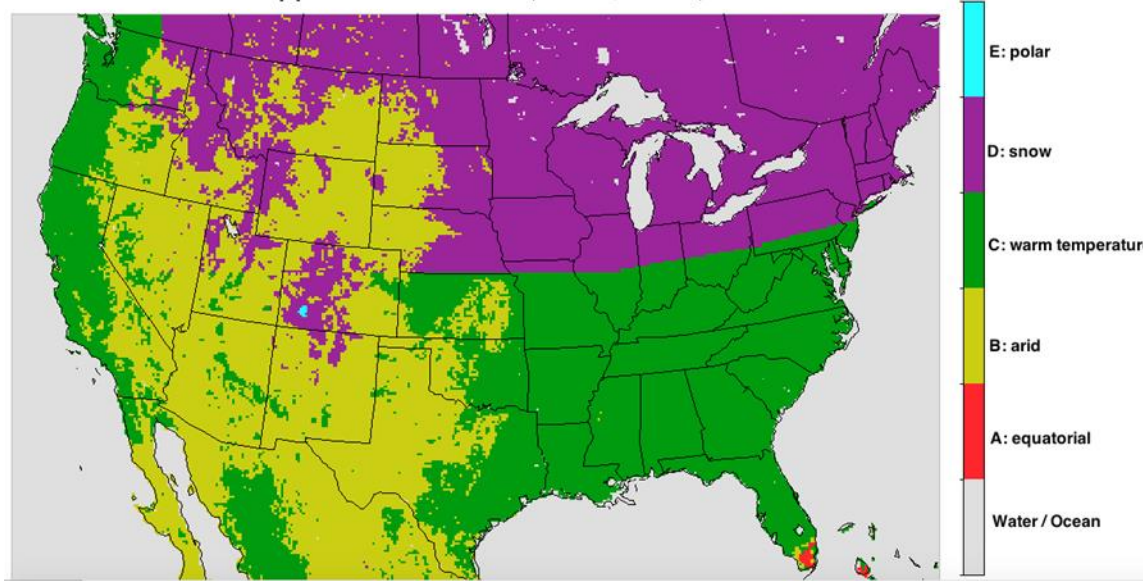
- Cooter, E. J., Bash, J.O., Benson, V., and Ran, L. (2012) Linking agricultural crop management and air quality models for regional to national-scale nitrogen assessments. *Biogeosciences*, 9, no. 10: 4023-4035.
- Hudman, R. C., Russell, A.R., Valin, L.C., and Cohen, R.C. (2010) Interannual variability 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 Zeitschrift*, 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.
- Stehfest, E. and Bouwman, L. (2006) N₂O 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-NO_x-hydrocarbon chemistry: 1. Model formulation. *Journal of Geophysical Research-Atmospheres*, 103, 10713–10725.
- Yienger, J. J. and Levy, H. (1995) Empirical-model of global soil-biogenic NO_x emissions. *Journal of Geophysical Research-Atmospheres*, 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



Köppen Climate Zone (CMAQ 12km)



Appendix B. Sample run script modifications for BDSNP soil NO module

```
#!/bin/csh
#
# In-line BDSNP Soil NO module implemented in CMAQ v 5.0.2
#

# CHANGE user-defined run.cctm script to include BDSNP module
# added flexibility to choose different input files and implementation schemes
# [ ] indicate default prompt for an environmental variable i.e. Y|N

#> Berkely-Dalhousie Soild NOx Parameterization [N|F ]
    setenv BDSNP Y
#> use EPIC fertilizer data (BDSNP 'new') or not (BDSNP 'old') with the flag BDSNP
#turn ON [N|F ]
    setenv EPIC Y

# Other USER DEFINED! Same as in base CMAQ model#

# Grid definition
setenv GRIDDESC $GRIDDESCHOME/GRIDDESC
setenv GRID_NAME 12US2
# Episode chosen in Julian date
set STJD = 201172 #Start Date
set ENJD = 201173 #End Date

## Loop between the start and end of episode period for in-line BSNP 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 BDSNP is used

#Outputs biogenic total NO and soil NO diagnostics
if ( $biogon ) then
    if ( $B3GTS_DIAG == 'Y' || $B3GTS_DIAG == 'T' ) then
        setenv B3GTS_S $OUTDIR/$B3GTS_S
    endif
    setenv SOILOUT $OUTDIR/$SOILOUT
    # Biogenic NO soil output file
endif
```

```

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

    # Soil biome and arid/non-arid map data inputs # NOTE: Use OLDBIOME and
    NEWBIOME files for respective cases of BDSNP ('old') and ('new') alternatively for
    $CLIMAFILE, $CLIMNAFILE and $LANDFRACFILE
    setenv CLIMAFILE $BDSNPINPUTS/*SOILCLIMA12US2.CONUS

    setenv CLIMNAFILE $BDSNPINPUTS /*SOILCLIMNA12US2.CONUS

    setenv LANDFRACFILE $BDSNPINPUTS /*SOILLANDFRAC12US2.CONUS
    # (OLDBIOME) - GEOS-Chem for BDSNP ('old'), updated NLCD for BDSNP
    ('new') - (NEWBIOME).

    # Fertilizer input for BDSNP ('old') - Potter et al. (2010)

    setenv FERTRESFILE $BDSNPINPUTS/*SOILFERT12US2.CONUS

if ( $?EPIC ) then
  if ( $EPIC == 'N' || $EPIC == 'F' ) then
    setenv EPICRESFILE $BDSNPINPUTS/*SOILFERT12US2.CONUS
  endif
endif

endif
endif

# Restart file SOILINSTATE/SOILOUT
#> Set Soil NO to zero? [N|F]
  setenv ZEROSOIL N #This will require an intial condition/restart SOILINSTATE file
#from previous day, in this case 20110721 or 2011172

  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 SOILINSTATE $OUTDAT/SOILINSTATE.${DATE}
    # Biogenic NO soil input file
    setenv INITIAL_RUN N
  else if ( $IMONTH == 12 ) then

```



```

    setenv SOILINSTATE $OUTDAT/SOILINSTATE.${DATE}
    # Biogenic NO soil input file

    setenv INITIAL_RUN N

else
    setenv SOILINSTATE $OUTDAT/SOILINSTATE.${DATE}
    # Biogenic NO soil 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 (in $SOILINSTATE RESTART FILE),
changes in CCTM source code handles it

# Fertilizer input for BDSNP ('new') - EPIC dynamic fertilizer fields

if ( $?BDSNP ) then
    if ( $BDSNP == 'Y' || $BDSNP == 'T' ) then
        echo "BDSNP true, choose to use EPIC fertilizer data or not"
        if ( $?EPIC ) then
            if ( $EPIC == 'Y' || $EPIC == 'T' ) then

                setenv EPICRESFILE $BDSNPINPUTS/*SOILFERT_EPIC_Potter_${APPL}.nc
                #EPIC daily fertilizer N, Canadian and Mexican regions masked from Potter
            endif
        endif
    endif
endif

#####

```


Appendix C. Sample of soil biome emission factor setting include file

#{BLD}/EF_BIOME.EXT

SPC,CATEGORY,EF

W,1,0.00

PW,2,0.00

SI,3,0.00

B_DE,4,0.00

U,5,0.00

B_ABC,6,0.06

CS,7,0.09

OS_ABC,8,0.09

OS_DE,9,0.01

G_DE,10,0.84

S_DE,11,0.84

S_ABC,12,0.24

G_ABC,13,0.42

WS,14,0.62

MF,15,0.03

EBF_CDE,16,0.36

DBF_CDE,17,0.36

DNF_AB,18,0.35

ENF,19,1.66

DBF_AB,20,0.08

EBF_AB,21,0.44

C,22,0.57

UBL,23,0.57

CVM,24,0.57