CALIFORNIA
AIR RESOURCES BOARD

Testing Internally Compressed NetCDF-4 File Format within SMOKE-IOAPI Framework

Roger Kwok and Sarika Kulkarni
CMAS Conference 2021

Motivation

- 2016. Climate change study (Zhan et al, 2020; 2021). WRF netcdf classic too big. Nccopy converts netcdf-3 to netcdf-4 classic, size halved.

  wrfout_d01_2013-01-09_00:00:00
  ➢Original size: **2.7G**
  ➢Updated size (after Nccopy): **1.3G**, *50% reduction*

CARB

Motivation

- 2017. CMAQ5.1 released. Nccopy converts mcip, emis, icon/bcon etc into Netcdf-4 classic. Steeper size reduction observed, as shown below.

  st_2k.Rx.v0011..2017.2017067...14jun2019
  ➢Original size: **81G**
  ➢Updated size (after Nccopy): **85M,** *99% reduction*
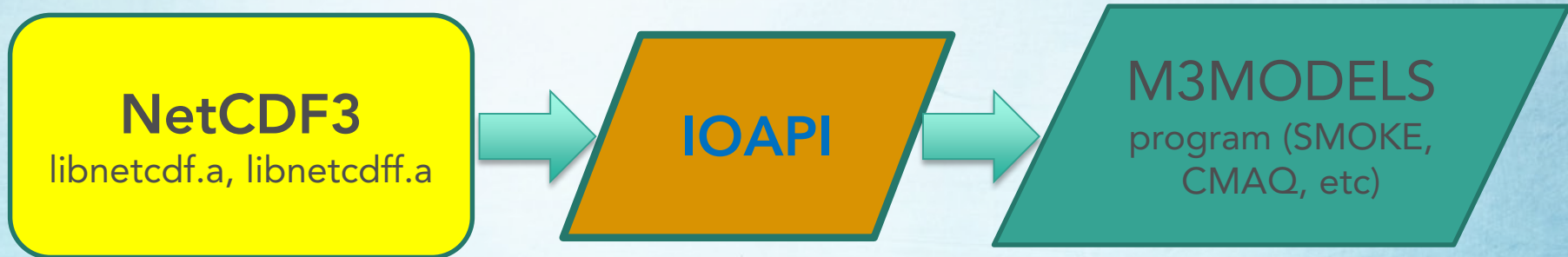
CARB

Motivation

2017

- CMAQ on statewide 2km, file size four times more than 4km.

- Gzip, Pbzip2 take up more time

- Still need nccopy for nc4 conversion

- Necessitates development of compressible IOAPI.

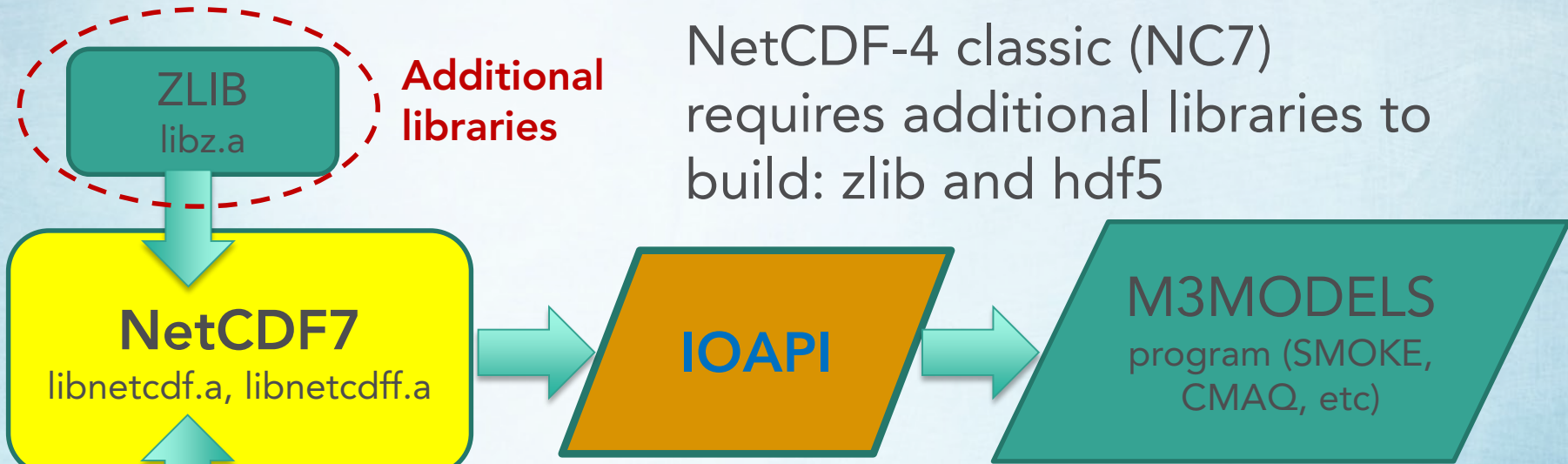| CMAQ v5.1 benchmark case: Netcdf-3 vs Netcdf-4 (both classic)			
File Type	nc3 (MB)	nc4 (MB)	ratio (nc3/nc4)
bcon	106	97	1.1
bidi	91	3.8	23.9
crop	1.7	0.38	4.5
dust	3.4	0.872	3.9
emis	4200	627	6.7
icon	200	168	1.2
mcip	1100	682	1.6
ocean	0.072	0.044	1.6
cctm	6600	5500	1.2
Input (GB)	5.7	1.6	3.6
output (GB)	6.6	5.5	1.2
Total (GB)	**12.3**	**7.1**	**1.7**

CARB

How were netCDF and IOAPI built?

- This is how to build NetCDF classic (NC3) libraries from which IOAPI and subsequent Models3 programs are compiled (BAMS recommended).
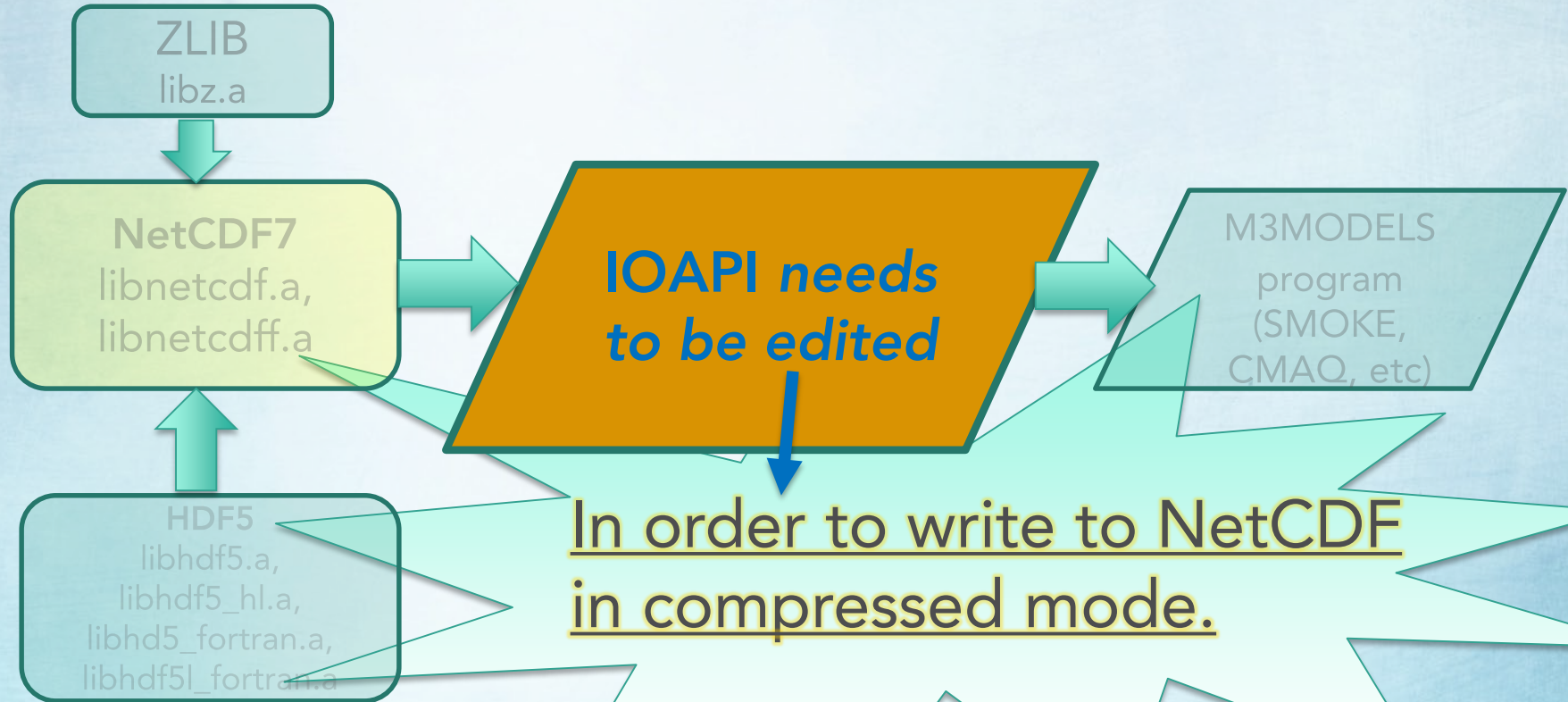
- Does not support NetCDF-4 I/O

NetCDF3
libnetcdf.a, libnetcdff.a

IOAPI

M3MODELS
program (SMOKE, CMAQ, etc)

CARB

How were netCDF and IOAPI built?

ZLIB
libz.a

Additional libraries

NetCDF-4 classic (NC7) requires additional libraries to build: zlib and hdf5

NetCDF7
libnetcdf.a, libnetcdff.a

IOAPI

M3MODELS
program (SMOKE, CMAQ, etc)

HDF5
libhdf5.a,
libhdf5_hl.a,
libhd5_fortran.a,
libhdf5l_fortran.a

- **Can read** any format including Netcdf-4
- **Cannot write** Netcdf-4 compressed (NC7)
- Still need nccopy to convert NC3 or 64-bit into NC7

How were netCDF and IOAPI built?

```
ZLIB
libz.a
```

```
NetCDF7
libnetcdf.a,
libnetcdff.a
```

IOAPI *needs to be edited*

```
M3MODELS
program
(SMOKE,
CMAQ, etc)
```

```
HDF5
libhdf5.a,
libhdf5_hl.a,
libhd5_fortran.a,
libhdf5l_fortran.a
```

In order to write to NetCDF in compressed mode.

CARB

IOAPI Changes made

- The key change lies in the subroutine "crtfil3.F90" that has been called mostly by "open3.F90".

- Original version has "NF_DEF_VAR" statements each of which creates a netCDF variable.

- Compression is enabled by adding a "NF_DEF_VAR_DEFLATE" statement after each "NF_DEF_VAR" call.

CARB

Ncdump: check netCDF file version

- Ncdump –k ${Nc file}
- Returns message "netCDF-4 classic model" if ${Nc file} is NC7
- Returns message "netCDF-4" if ${Nc file} is regular netCDF 4
- Returns message "classic" if ${Nc file} is netCDF 3

Models3 programs tested well

- Reads in any netCDF format
- Writes out NC7 compressed or uncompressed
- Writes out 64bit-offset if users choose to (see later)

NC3, NC7, or mixed → {Models3 with IOAPI/NC7} → NC7 ✔

NC3, NC7, or mixed → {Models3 with IOAPI/NC7} → ✘ NC3
(*can always use "nccopy" to convert*)

CARB

C-shell script environment setting: COMPRESS_NC

setenv COMPRESS_NC N   #Disables compression while writing out Nc7

setenv COMPRESS_NC Y   #Y means compression
setenv COMPRESS_NC   #Empty environment also means compression

If ($?COMPRESS_NC == 0) #Not setting environment means compression

For emission inventory preparation, we don't worry about this environment variable.

CARB

SMOKE file size, runtime: NC7 vs NC3

		NC3	NC7	NC7-to-NC3 ratio	Remarks
			SMOKE area run for 365 days		
SMOKE	Per day file size (MB)	633.0	55.0	8.7%	NC3 uncompressed
	Run time (min)	30.6	41.6	135.9%	Time taken by SMOKE
GZIP	Per day file size (MB)	55.0	N\A	100.0%	NC3 compressed
	Runtime (min)	29.6	N\A	69.2%	Time taken by SMOKE + gzip
PBZIP2	Per day file size (MB)	52.0	N\A	105.8%	NC3 compressed
	Runtime (min)	85.6	N\A	35.8%	Time taken by SMOKE + pbzip2

CARB

Verdi visualization

- VERDI not only visualizes both NC7 and NC3 on the same interactive console, but also plots difference plots across the files of different formats.

- VERDI 2.1 visualizes NC7 without issuing any warning messages. Previous versions complain presence of HDF5.

- Other visualizing tools also support NC7; e.g. Ncview, matplotlib

64bit Offset

- SMOKE sparse data structure → compression effective

- CMAQ outputs continuous structure → less effective

- With IOAPI/NC7, users still can choose to revert to default CMAQ output format: IOAPI_OFFSET_64

- Can always use "nccopy" to convert to NC3

CARB

64bit Offset settings

setenv IOAPI_OFFSET_64 #Y or N   (Default **N**)

setenv USR_DFLAT_LVL   #1, 2,…,9 (Default **2**)

setenv COMPRESS_NC   #Y or N   (Default **Y**)

❖ Setting IOAPI_OFFSET_64 to **Y** *overrides the COMPRESS_NC option*; in other words, no compression whenever the 64-bit offset option is turned on.

❖ Without explicitly setting IOAPI_OFFSET_64, the environment variable tacitly assumes a number. Then Nc7 compression resumes.

❖ Default USR_DFLAT_LVL is 2 unless specified otherwise by end-users.

CARB

Environment Settings Summary

Environment variable	Default value	Other values	Effects
COMPRESS_NC	Y	N	• Y to compress output • N to disable compression • *Ignored if IOAPI_OFFSET_64 Y*
USR_DFLAT_LVL	2	1 thru 9 except 2	• 1 mildest, 9 heaviest • Ignored if COMPRESS_NC N
IOAPI_OFFSET_64	N	Y	• *Y to output 64-bit offset and disable compression* • N to resume compression.

CARB

Summary

- IOAPI with compressible option offers a solution to streamline SMOKE processing and data storage

- Greatly reduces sizes of files with sparse data structures

- Eliminates the needs for
  - ➢ Offline zipping/unzipping
  - ➢ Nccopy to convert other formats to NC7

- When compression is less effective, users still have a choice to write CMAQ output in default 64bit-OFFSET

Disclaimer

- IOAPI/NC7 was implemented based on the IOAPI developed and maintained by Baron Advanced Meteorological Systems (BAMS)

- This feature offers a solution to address the disk space issue, as end-users see fit. It is not related to any policy endorsed by USEPA or BAMS

- The emission compression was also applied in CMAQ runs (See S. Kulkarni's poster)

- To request for the library package, contact Roger Kwok via roger.kwok@arb.ca.gov

CARB

Reference

- Zhan et al 2020: Assessment of climate change impact over California using dynamical downscaling with a bias correction technique: method validation and analyses of summertime results (https://link.springer.com/article/10.1007/s00382-020-05200-x)

- Zhan et al 2021: Assessment of climate change impact on wintertime meteorology over California using dynamical downscaling method with a bias correction technique (https://link.springer.com/article/10.1007/s00382-021-05718-8)

CARB