NAEFS
Multi-center data exchange

NAEFS workshop May 17-19, 2010 Cuernavaca
MSC: Lewis Poulin, lewis.poulin@ec.gc.ca

Acknowledgements to
Rebecca Cosgrove (NCEP)
Mike Sestak (FNMOC)

Overview

- A NAEFS production run
- Summary NAEFS variables and file sizes
  - Raw, Bias Corrected (BC)
- How we exchange NAEFS data
- Learning from our experience
- Future considerations
- Recommendations

Overview of a NAEFS [00,12]Z production run

COMPONENT

Analysis

- Make own analysis
- Get other center analysis
- Get NCEP re-analysis

GEPS Prog data

- Generate own EPS Raw/Bc data
- Exchange Raw/Bc data

Products Raw (BC)

- Using own data
- Using other center data
- Using all center data

Distribute products

A NAEFS [00,12]Z production run
Overview of timelines and contingencies

COMPONENT	TARGETS	CONTINGENCY
Analysis		
• Make own analysis	• +4:30	full support
• Get other center analysis	• +4:30	in real time, else later
• Get NCEP re-analysis	• +4:45	in real time, else later
GEPS Prog data		
• Generate own EPS Raw/BC data	• +4:30-5:30	ASAP, full support
• Exchange Raw/BC data	• +4:30-6:30	ASAP, min # mbrs
Products Raw (BC)		
• Using own data	• +5:15 - +7:00 (8:00)	ASAP, QC
• Using other center data	• +5:00 - +7:00 (8:00)	ASAP, QC
• Using all center data	• +6:00 - +8:00 (9:00)	fn priority, QC
Distribute products	**Distribute products**	
	• Monitor distribution, QC	

NAEFS Raw variables
May 2010, 80 variables, GRIB2

HGT 10,50,100, 200, 250, 500,700, 850 , 925, 1000 mb
TEMP 10,50,100, 200, 250,500,700, 850 , 925, 1000 mb
UGRD 10,50,100, 200, 250, 500,700, 850 , 925, 1000 mb
VGRD 10,50,100, 200, 250, 500,700, 850 , 925, 1000 mb
RH 10,50,100, 200, 250, 500, 700, 850, 925, 1000 mb
VVEL 850 mb

UGRD, VGRD, 10m
TEMP, DEPR, 2m
TMAX, TMIN 2m
HGT Surface as seen by the model

TCDC Total cloud
PWAT Surface (Precipitable Water)

PMSL MSL
PRES Surface

Precipitation
APCP Surface Total Preciptation
Categorical Rain, Snow, Ice, Freezing – NCEP
Accumulated Rain, Snow, Ice, Freezing rain – CMC

CAPE

Snow water equivalent Surface
Snow depth Surface
Temperature, Soil moisture 0-10 cm below ground

CINH Convective Inhibition

Latent heat flux Surface
Sensible heat flux Surface
Downward short wave radiation Surface
Downward long wave radiation Surface
Upward short wave radiation Surface
Upward long wave radiation Surface
Outgoing long wave top of atmosphere

Black – Original NAEFS variables 2004
Red – Variables added in fall 2009 to GRIB2 files

NAEFS Bias Corrected (BC) variables
May 2010 [47,49] variables, GRIB2

HGT 10,50,100, 200, 250, 500,700, 850 , 925, 1000 mb
TEMP 10,50,100, 200, 250,500,700, 850 , 925, 1000 mb
UGRD 10,50,100, 200, 250, 500,700, 850 , 925, 1000 mb
VGRD 10,50,100, 200, 250, 500,700, 850 , 925, 1000 mb
VVEL 850 mb – NCEP only

UGRD, VGRD, 10m
TEMP 2m
TMAX, TMIN 2m

PRES Surface
PMSL MSL

Upward long wave radiation Surface – NCEP only

Notes – (Bias Correction w.r.t. originating center analysis)

- Centers now bias correct their own data and share it with other centers
- CMC using NCEP generated bias-corrected data
- NCEP soon to be using CMC generated bias corrected data

Bias correction against NCEP reanalysis still done by each center

May 2010 NAEFS GEPS datasets

	CMC	NCEP	FNMOC*	CMC	NCEP	FNMOC*
	RAW	RAW	RAW	Bias Corr.	Bias Corr.	Bias Corr.
Number of variables	80	80	69	47	49	
Number of members	21	21	21	21	21	
Start forecast hour	0	0	0	0	0	
End forecast hour	384	384	384	384	384	
Hours per time step	6	6	6	6	6	
Number of time steps	65	65	65	65	65	
Number of files per run	1365	1365	1365	1365	1365	
Size of NAEFS dataset by center GB / [00,12]Z run	**5.1**	**4.1**	**3.5**	**2.7**	**3.2**	
Format	grib2	grib2	grib2	grib2	grib2	grib2

Total GB: **CMC 7.8 GB** **NCEP 7.3 GB** **FNMOC*=3.5 GB (est.)**

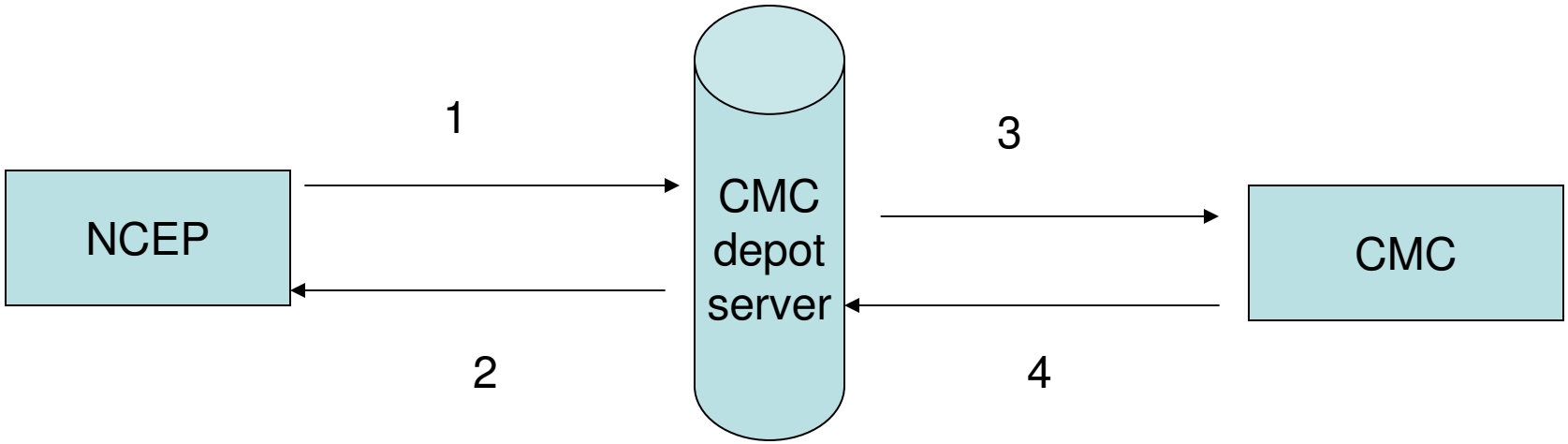
Now
The actual data exchange

NAEFS (Fall 2004 to Fall 2009)
GRIB1, shared via one server, Internet (ftp, sftp)

NCEP
(1) puts its (prog, reanalysis) data onto (A-1)
(2) pulls CMC prog data from (A-2)

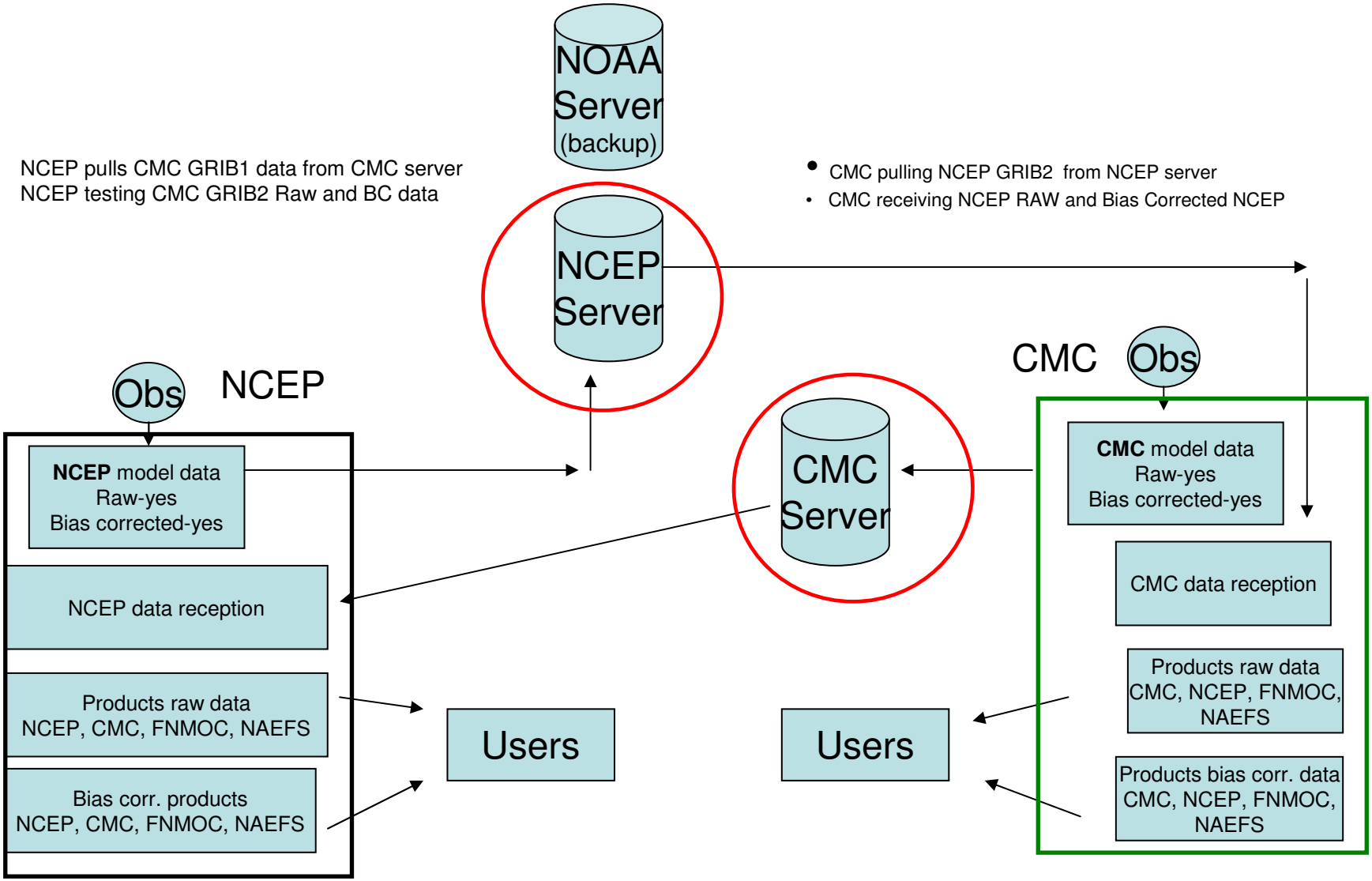
CMC
(3) pulls NCEP (prog, reanal.) data from (A-1)
(4) puts CMC prog data onto (A-2)

(A-1) ftp://depot.cmc.ec.gc.ca:naefs/ncep/grib/

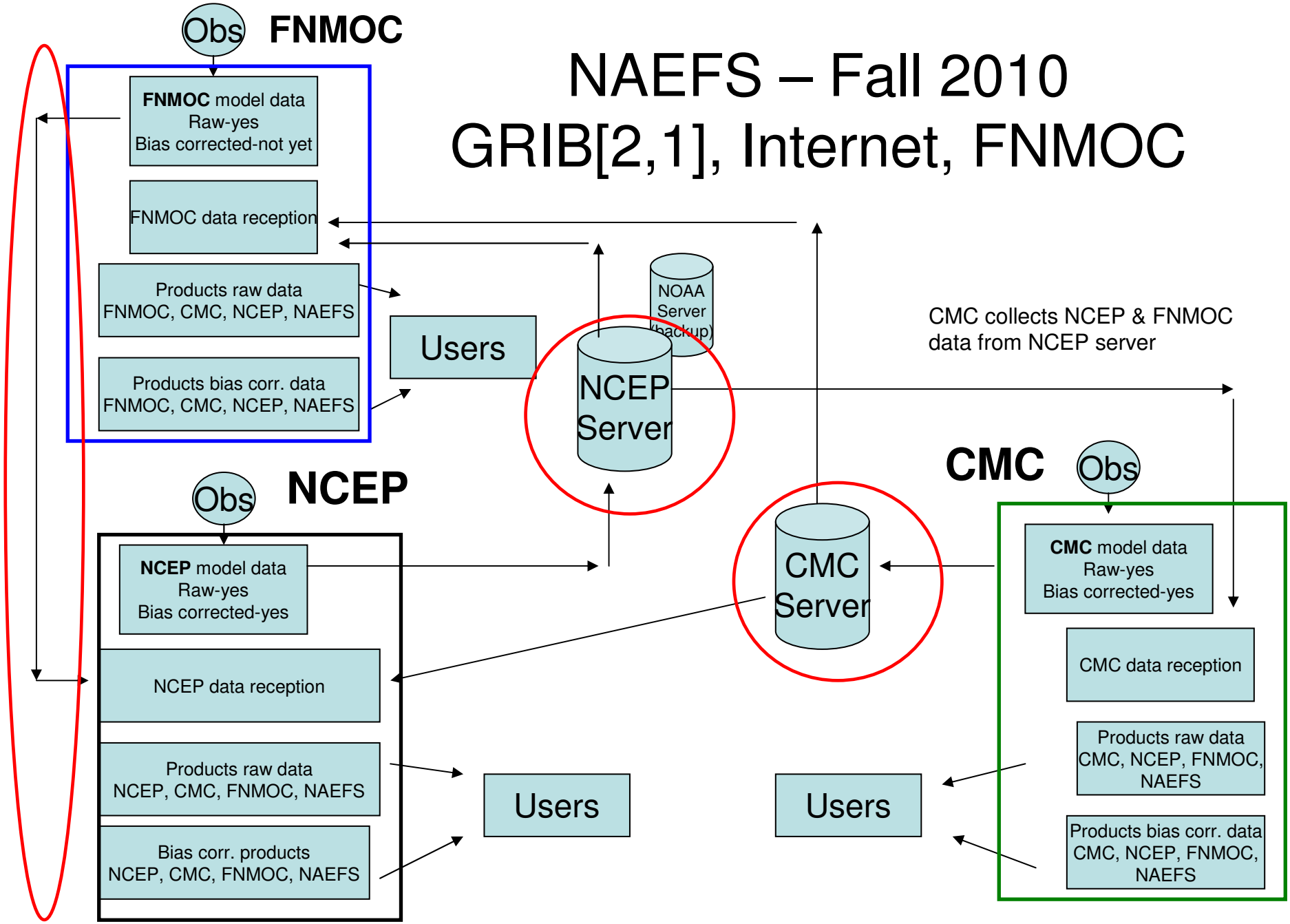


(A-2) ftp://depot.cmc.ec.gc.ca:naefs/cmc/grib/

NAEFS - Fall 2009 to present
2 servers, GRIB[1,2], Internet

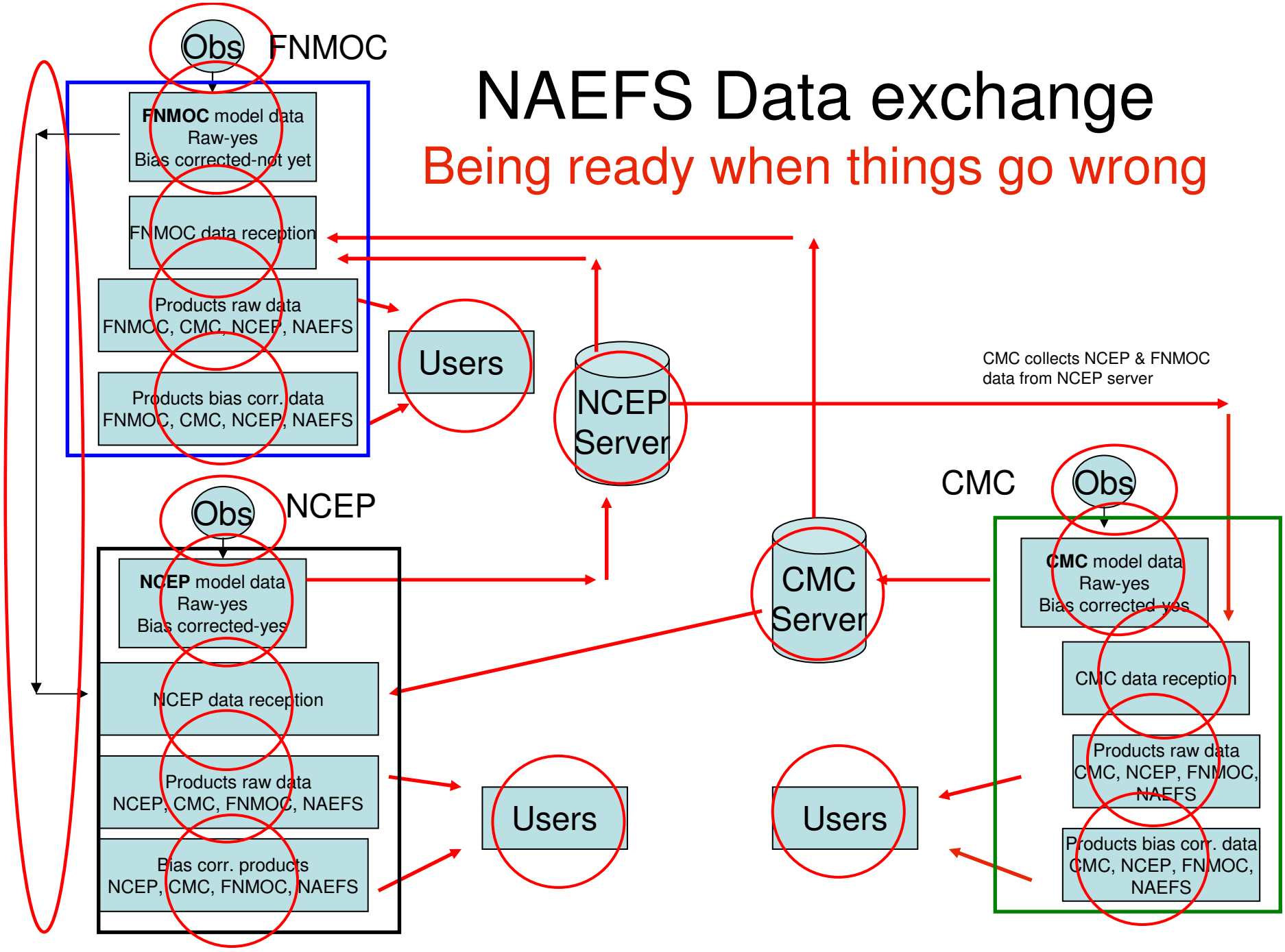


NAEFS – Fall 2010
GRIB[2,1], Internet, FNMOC



NAEFS Data exchange

Being ready when things go wrong



Observed data exchange problems / solutions

Problems

Data exchange
- Missing files - recoverable?
- Missing files – Un-recoverable
- File not named correctly
- Models running a little late
- Models running VERY late

Reanalysis data
- Missing files
- Duplicate file, incorrect

Server problems
- Server down

GRIB files
- Random corruption in files

Products
- Missing products at destination

Solutions

Data exchange
- Recovered files in real time
- If possible, substitute member. Use min mbrs
- Don't rely on file name, use meta info
- If possible accommodate late data
- Start early with min required data

Problem - Reanalysis data
- Recover next business day
- Identify missing data, recover

Server problems
- Recover systems ASAP

GRIB files
- Use secure ftp
- Add Q/C to grib production

Products
- Add Qc to monitor product distribution

Considerations for the future

Adding data to NAEFS?
Possible future NAEFS [G,R]EPS [00,06,12,18]Z runs
Increase resolution, add REPS? GEPS Day 17-35?, more data

COMPONENT

Analysis [G,R]EPS

- Make own analysis
- Get other center analysis
- Get NCEP re-analysis

[G,R]EPS Prog data

- Generate own [G,R]EPS Raw/BC data
- Exchange Raw/BC data

Products Raw (BC)

- Using own data
- Using other center data
- Using all center data

Distribute products

Expanding datasets is easier if NAEFS formats and mechanics are used

Summary possible growth - NAEFS data

GEPS day 1-16 RAW Bias Corrected (BC)	**Fall 2010**	**Fall 2011**	**Fall 2012**	**Fall 2013**
GEPS CMC NCEP	**Raw+BC 5.1+2.7 = 7.8 4.1+3.2 = 7.3**	**Raw+BC 6.5+2.7 = 9.2 9.2?**	**Raw+BC 51.2+25.1 =76.3 76?**	**Raw+BC 64.1+27.4 = 91.5 91?**
REPS CMC NCEP	**1.5+0.8 = 2.3 2.4+0.5 = 2.9**	**1.5+0.8 = 2.3 3.6+0.8= 4.4**	**1.9+1.0 = 2.9 3.6+1.0= 4.6**	**2.5+1.3 = 3.8 7.0+1.8= 8.8**
GEPS day 17-35 CMC / NCEP	**1.5+0.8 = 2.3**	**1.9+0.8=2.7**	**7.5+3.7=11.2**	**9.4+4.0=13.4**
Total GB / Run CMC NCEP ** Excluding day 17-35	**10.1 10.2**	**11.5 13.6**	**79.2 80.9**	**95.3 99.8**
Assume ideal transfers? Mbps estimated ➔	**10GB / 60 min 22.7 Mbps**	**12GB / 60 min 27.3 Mbps**	**80GB / 60 min 182 Mbps**	**100 GB / 60 min 227 Mbps**

Exchanging larger datasets?

Observed data transfer times
Internet: 5-7 GB transferred in ~60 minutes

Transfer	Raw	Bias Corrected	Total	Minutes	Mbps equivalent**
NCEP gets CMC GRIB1	**6.5**	**0**	**6.5 GB**	**~60**	**14.8 ** **
CMC gets NCEP GRIB2	**4.1**	**3.2**	**7.3 GB**	**~50**	**19.9****
NCEP gets CMC GRIB2	**5.1**	**0**	**5.1 GB**	**~55**	**12.7****
NCEP gets FNMOC GRIB?	**3.2**	**0**	**3.2 GB**	**120**	**3.6 ** **
CMC gets FNMOC	**TBD**				

GB/minutes ➔ provides Mbps equivalent of the Internet ➔ 13-20 Mbps **

** (# of GBx1024x1024x1024x8)/(50minx60sec)/(1024x1024) ➔Mbps equivalent

How much data could be exchanged in 60 minutes with ideal 45 Mbps link?

- Has been interest in a 45 Mbps dedicated link
- What if?
  - We had an ideal 45 Mbps dedicated link
  - We wanted to transfer complete datasets in ~ 60 minutes
- What would be the maximum GB that could be transferred ?
  - (45 Mbps*(10^6bits/Mb)*(60min*60s/min)/(8bits/byte)/(1024*1024*1024 bytes/GB)
  - ➔**~19GB** likely maximum data transferred under ideal conditions
  - At 50% performance, max dataset = 8-10 GB possible
  - **8-19 GB approximate maximum GB transferrable**

Recall previous estimate of datasets (2010)10-12 GB then 80-100 GB (2012+)

GEPS day 1-16 RAW Bias Corrected (BC)	**Fall 2010**	**Fall 2011**	**Fall 2012**	**Fall 2013**
GEPS CMC NCEP	**Raw+BC 5.1+2.7 = 7.8 4.1+3.2 = 7.3**	**Raw+BC 6.5+2.7 = 9.2 9.2?**	**Raw+BC 51.2+25.1 =76.3 76?**	**Raw+BC 64.1+27.4 = 91.5 91?**
REPS CMC NCEP	**1.5+0.8 = 2.3 2.4+0.5 = 2.9**	**1.5+0.8 = 2.3 3.6+0.8= 4.4**	**1.9+1.0 = 2.9 3.6+1.0= 4.6**	**2.5+1.3 = 3.8 7.0+1.8= 8.8**
GEPS day 17-35 CMC / NCEP	**1.5+0.8 = 2.3**	**1.9+0.8=2.7**	**7.5+3.7=11.2**	**9.4+4.0=13.4**
Total GB / Run CMC NCEP ** Excluding day 17-35	**10.1 10.2**	**11.5 13.6**	**79.2 80.9**	**95.3 99.8**
Assume ideal transfers? Mbps estimated ➔	**10GB / 60 min 22.7 Mbps**	**12GB / 60 min 27.3 Mbps**	**80GB / 60 min 182 Mbps**	**100 GB / 60 min 227 Mbps**

Transferring future NAEFS datasets

- Under present data transfer arrangement
  - Internet likely too slow for datasets that grow beyond 10-20 GB?
- Will require adjustment to strategy for transferring large datasets
- Explore choices
  - Clarify when we want to increase to much larger datasets
  - Find a faster link to continue the transfer of complete datasets
  - Or transfer of files in order of priority, high priority in first 60 minutes, use Internet
  - Creative reduction of NAEFS exchange dataset, for example:
    - Highest resolution (/3hr, .5x.5 lat-lon grid), Day 1-5
    - Moderate resolution (/6hr, 1x1 lat lon grid) Day 6-10
    - Low resolution (/12hr, 1x1 lat lon grid) Day 11-16?
  - Evaluate carefully to ensure size of dataset is matched to what link can deliver

Recommendations

- NAEFS production run, timelines, contingencies
  - Continue clarifying NAEFS operational timelines, contingencies
- Data exchange
  - 2012 - NAEFS datasets may grow beyond internet capacity
  - Maintain ongoing evaluation of 1-5 year options for data exchange
  - Setting up new network solutions takes time, plan well, start early
  - Carefully plan growth of dataset with capacity of Internet or dedicated link to transfer data as per timelines
  - Plan early for when adding new models (REPS), new partners
- Troubleshooting procedures
  - Ensure personnel have troubleshooting info and training, simulators help
  - Explore how centers may respond more dynamically to occasional missing data
  - Adopt ISO-based procedures if possible
- Regularly update main NAEFS document with the above information

Thank you - Gracias

Appendix 1
Brief summary of future GEPS

Possible future changes to CMC GEPS NAEFS dataset

CMC NAEFS –**GEPS**- RAW/BC datasets

Day 1 to day 16

GEPS 1-16 days RAW Bias Corrected (BC)	GEPS Fall 2010	GEPS Fall 2011	GEPS Fall 2012	GEPS Fall 2013
Number of variables Raw BC	80 47	100 50	100 55	125 60
Number of time steps 65 = 6hr time steps 129 = 3hr time steps	65	65	129	129
Grid resolution lat-lon	1 x 1	1 x 1	0.5 x 0.5	0.5 x 0.5
# of members	21	21	21	21
Estimated GB/run Raw Bc GRIB2	5.1 2.7	6.5 2.7	51.2 25.1	64.1 27.4

Possible future changes to NCEP GEPS NAEFS dataset

NCEP NAEFS –**GEPS**- RAW/BC datasets

Day 1 to day 16

GEPS day 1-16 RAW BC	GEPS Fall 2010	GEPS Fall 2011	GEPS Fall 2012	GEPS Fall 2013
Number of variables Raw BC	 80 47	 100 50	 100 55	 125 60
Number of time steps 65 = 6hr time steps 129 = 3hr time steps	 65	 65	 129	 129
Grid resolution lat-lon	1 x 1	1 x 1	0.5 x 0.5	0.5 x 0.5
# of members	21	21	21	21
Estimated GB/run Raw Bc GRIB2	 4.1 3.2			

Appendix 2
More detailed information on timeline and contingencies

Discussion document
More detailed NAEFS timelines, contingencies
Could be finalized at workshop

NAEFS 00Z timeline	Early target	Contingency		Final target	Contingency
Get other center analysis	by +4:45	best effort		? by +18 if possible	
Get NCEP reanalysis	use latest available	best effort		by +18 or next business day	
Generate own center raw grib2	by +5:45 or earlier	**If late inform others**		?	?
Get other center RAW grib2 data	by +6:30 or earlier	**All members if possible, else all members by Second target time**		by +12:00? or earlier	?
Make own center RAW products	by +7:00 or earlier	**Must have 50%+1 of own center members?**		by +12:00? or earlier	?
Make other centers RAW product	by +7:30 or earlier	**Must have 50%+1 other center membrs?**		by +13:00? or earlier	?
Make all centers RAW products	by +7:30 or earlier	**Must have 50%+1 of allcentter members?**		by +13:00? or earlier	?
Generate own center BC grib2	by +6:00 or earlier	**All members if possible, else all members by Second target time**		by +14:00? or earlier	?
Get other centers BC grib2 data	by +7:00 or earlier	**All members if possible, else all members by Second target time**		by +15:00? or earlier	?
Make own center BC products	by +8:00 or earlier	**Must have 50%+1 of own center members?**		by +15? or earlier	?
Make other centers BC products	by +8:30 or earlier	**Must have 50%+1 of other center membrrs?**		by +16:00 ? or earlier	?
Make all centers BC products	by +8:30 or earlier	**Must have 50%+1 of allcentter members?**		by +16:00? or earlier	?

Discussion document
NAEFS - Analysis / Re-analysis
Could be finalized at workshop

NAEFS 00Z timeline	Early target	Contingency		Final target	Contingency
Make own center Analysis					
Get other center Analysis	by +4:45	best effort		? by +18 If possible	
Get NCEP reanalysis	use latest available	best effort		by +18 or next business day	

Discussion document
NAEFS – Raw/Bc prog data timeline
Could be completed at workshop

NAEFS 00Z timeline	Early target	Contingency		Final target	Contingency
Own center GRIB2 **Raw** **BC**	+5:45 or earlier +6:00 or earlier	**If late inform Others**		? 14:00	?
Other center GRIB2 **Raw** **BC**	+6:30 or earlier +7:00 or earlier	**Min. number of members else all members by final target**		+12:00? or earlier +15:00	?
Own center Products **Raw** **BC**	+7:00 or earlier +8:00 or earlier	**50%+1 of own center membrs?**		+12:00? or earlier +15:00	?
Other centers Products **Raw** **BC**	+7:30 or earlier +8:00 or earlier	**50%+1 of other center membrs?**		+13:00? or earlier +16:00	?
All centers Products **Raw** **BC**	+7:30 or earlier +8:30 or earlier	**50%+1 of Allcentter members?**		+13:00? or earlier +16:00	?

Discussion document
NAEFS – Raw/Bc prog data timeline
Could be completed at workshop

NAEFS 00Z timeline	Early target	Contingency	Final target	Contingency
Own center GRIB2 **Raw** **BC**	+5:45 or earlier +6:00 or earlier	**If late inform** **Others**	? 14:00	?
Other center GRIB2 **Raw** **BC**	+6:30 or earlier +7:00 or earlier	**Min. number of members else all members by final target**	+12:00? or earlier +15:00	?
Own center Products **Raw** **BC**	+7:00 or earlier +8:00 or earlier	**50%+1 of own center members?**	+12:00? or earlier +15:00	?
Other centers Products **Raw** **BC**	+7:30 or earlier +8:00 or earlier	**50%+1 of other center members?**	+13:00? or earlier +16:00	?
All centers Products **Raw** **BC**	+7:30 or earlier +8:30 or earlier	**50%+1 of All centter members?**	+13:00? or earlier +16:00	?

Appendix 3
Additional troubleshooting details

Insights - NAEFS data exchanges

- Centers anticipate & monitor for problems ahead of time
- Centers are ready when problems occur
- Each failure point should have procedures to correct problem, and knowledgeable staff ready to act
- Problem simulators are helpful for staff training

- Scheduling/timelines can help design coping strategies
  - Allows support staff design procedures for objectives
  - Allows participants to design systems based on o guidelines

Centers can keep track of problems and solutions to learn from experience

Event	Contingency	Recommendations
2008-01-22 / Unrecoverable prog file at one time step gep15.t00z.pgrbaf384 unrecoverable	CMC uses member 7 data instead Time step missing is end fhr	
2008-02-[01,02] / 2 reanalysis missing for Feb 1 06Z	cdas data only recovered on next business day if event is on weekend	
2008-02-02/ Missing control file: gec00.t18z.pgrbaf00	Improved CMC documentation on recovery actions	
2008-02-[05,06] / Incorrect reception of 2 reanalysis files Received 2 reanalysis for Feb 5 00Z, but both contained content of *flux* files.	CMC requested re-transmission	
2008-02-18 / Incorrect push of control file gec00.t18z.pgrbaf00 pushed to depot but not renamed correctly, can't be 'seen' by CMC	Several re-transmission requests had to be made, Corrections to ncep script that pushes files, Problem caused by migration to new machines left double versions of scripts running	
2008-02-22 / Problems with a 00Z ensemble run Feb 22 NCEP naefs file missing	CMC received empty ncep file 5:15 Z CMC debiasing bombs naefs file missing Took 3+ [calls, re-transmits] to get correct file File eventually received at about 12Z (+7hrs)	
2008-01-[27,28] / Difficulty accessing NCEP reanalysis data valid for previous day 2008-02-27 00Z	Requested re-transmission	
2008-03-27 / Inadvertent loss by CMC re-analysis file	CMC requests re-transmission of reanalysis files	
2008-06-02 / NCEP server down	CMC has difficulty receiving NCEP files	
January - March 2010 / Late data delivery possible resulting from system upgrades at generating center		

Sample CMC operator simulator
What would you do if?

(1) NCEP calls NETOPS reporting CMC naefs ensemble member #13 at forecast hour 84 is missing.

Answer Question 1:
NETOPS recognizes this is is a problem where CMC NAEFS data is missing.
NETOPS asks NCEP which RUN is missing, ex: Is it missing RUN=00Z or RUN-12Z?
NETOPS uses the NETOPS data recovery checklist procedure 1 to recover the CMC NAEFS data

(2) A&P calls NETOPS to report some NCEP NAEFS RAW PROG data may be missing from today's 00Z run

Answer Question 2:
NETOPS diagnoses this is a problem with missing ncep naefs RAW PROG data for today's 00Z run
NETOPS uses procedure 2 from the NETOPS data recovery checklist to recover the missing NCEP NAEFS data.

(3) A&P calls NETOPS to report some NCEP NAEFS BIAS CORRECTED PROG data may be missing from today's 12Z run

Answer Question 3:
NETOPS diagnoses this is a problem with missing ncep naefs BIAS CORRECTED PROG data for today's 12Z run
NETOPS uses procedure 2 from the NETOPS data recovery checklist to recover the missing NCEP NAEFS BIAS CORRECTED PROG data

(4) NCEP calls NETOPS to report there is no CMC NAEFS data at all on depot server for today's 00Z run.

Answer Question 4:
NETOPS notes this is a case of missing cmc naefs data for RUN= 00Z.
NETOPS uses procedure 1 from the NETOPS data recovery checklist to correct this cmc naefs problem.

(5) **A&P calls NETOPS to ask NETOPS to recover reanalysis files of the type *cdas* missing**

Answer Question 5:
NETOPS diagnoses this is a problem with missing ncep naefs reanalysis data
NETOPS uses procedure 2 from the NETOPS data recovery checklist making sure to follow procedure for reanalysis data recovery
In the end, NETOPS informs A&P to log in shift report missing naefs reanalysis data and ask CMOI to investigate next business day.
No further action required for NETOPS.