Atmospheric Science Data Analysis Tools

P. K. Patra

Department of Environmental Geochemical Cycle Research

JAMSTEC, Yokohama, Japan

Presented at the ACAM Training School on
"Satellite and Model Data use for Aerosols and Air Quality"
11-12 June 2015, Asian Institute of Technology (AIT), Bangkok

GRENE-Arctic

JSPS 独立行政法人 日本学術振興会 Japan Society for the Promotion of Science

D E G C R

JAMSTEC

Is there any universal tool for data analysis?

Year	Event	Research and tools adopted
1992	Joined Ph. D.	Learned instrumentation, field campaigns, little data analysis, simple software (Fortran, SM graphics, LaTeX)
1998	Joined IBM	No data, but model development and analysis (lots of Fortran, 5-dimensional data analysis, data compression technique-PCA, advanced visualisation, GrADS)
2001	Joined JAMSTEC	Apply optimization tools (Simulated annealing), Bayesian inversion, data assimilation of GHGs; Microsoft Office
2005	Become older	Apply AGCM to GHG modelling – worked for people, with people. Learned a lot in terms of science
2015	Today	Not so much research; hard to learn new tools!

Fortran: originally developed by IBM in the 1950s for scientific and engineering applications

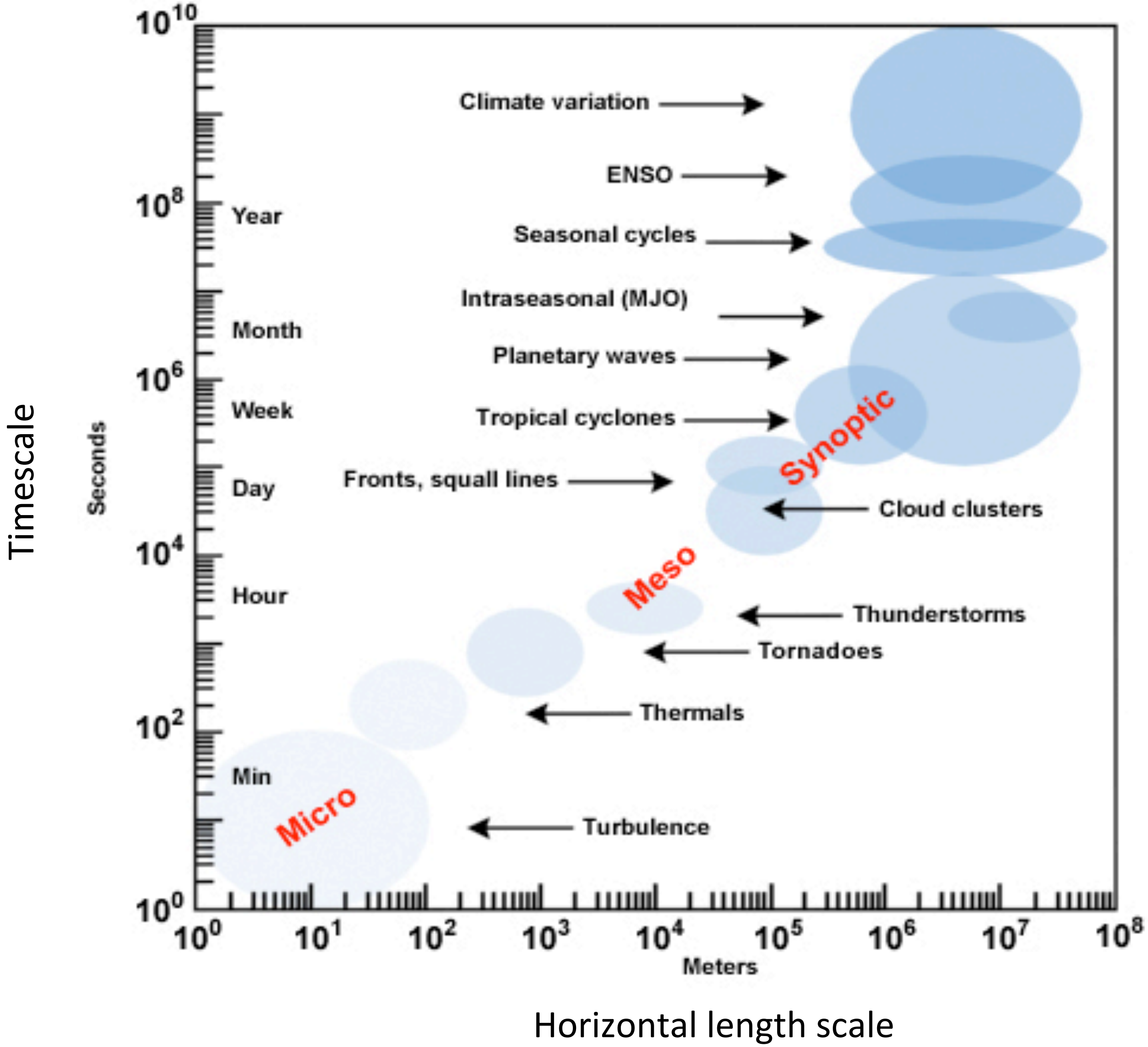
MM5,
RAMS
NCEP
FNL

$$C_S = (G^T C_D^{-1} G + C_{S_0}^{-1})^{-1}$$

$$S = S_0 + \left(G^T C_D^{-1} G + C_{S_0}^{-1}\right)^{-1} G^T C_D^{-1} (D - D_{ACTM})$$

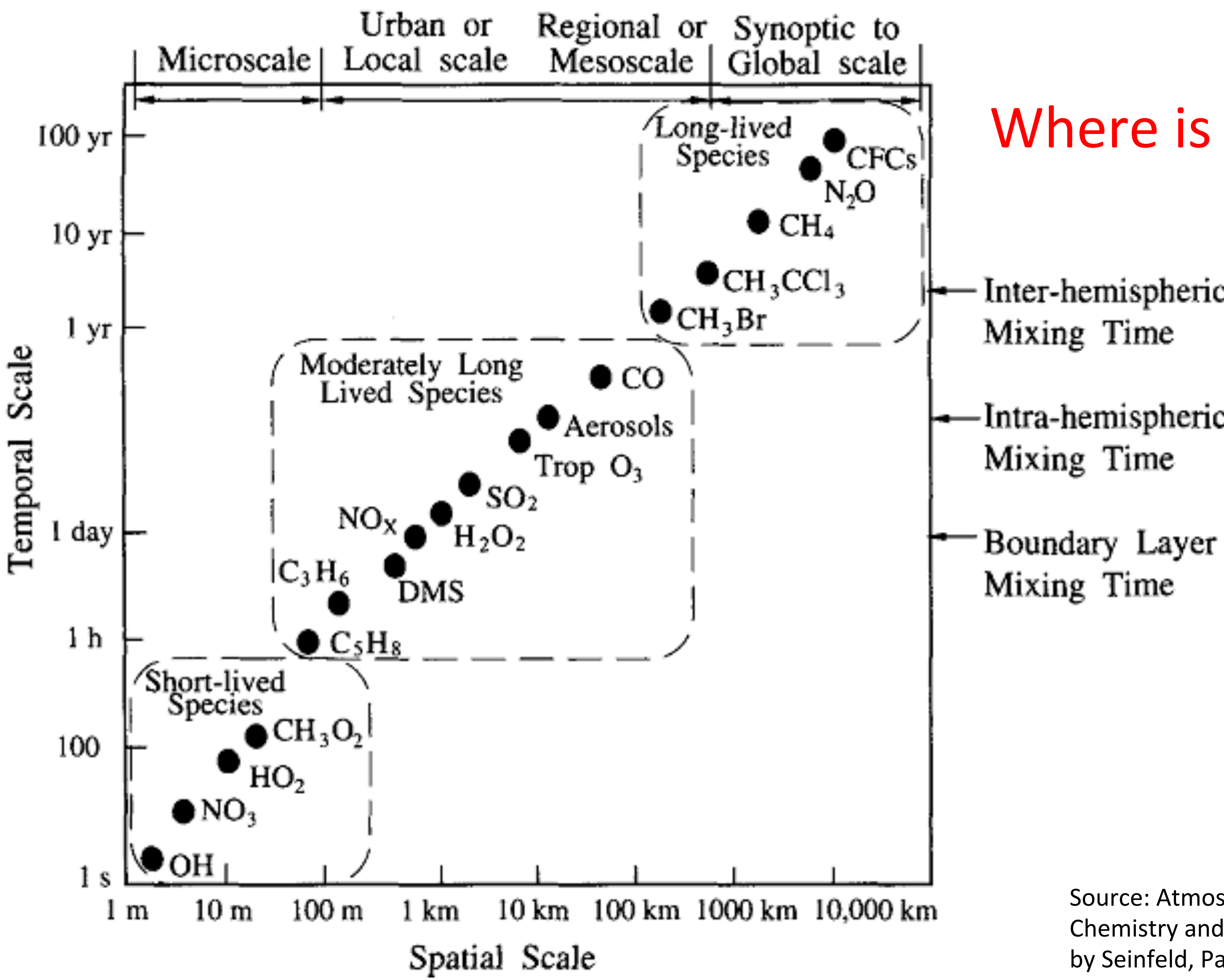
That's probably is one of the reasons I am asked to teach!

Scales of atmospheric dynamics



Source:
The Comet Program

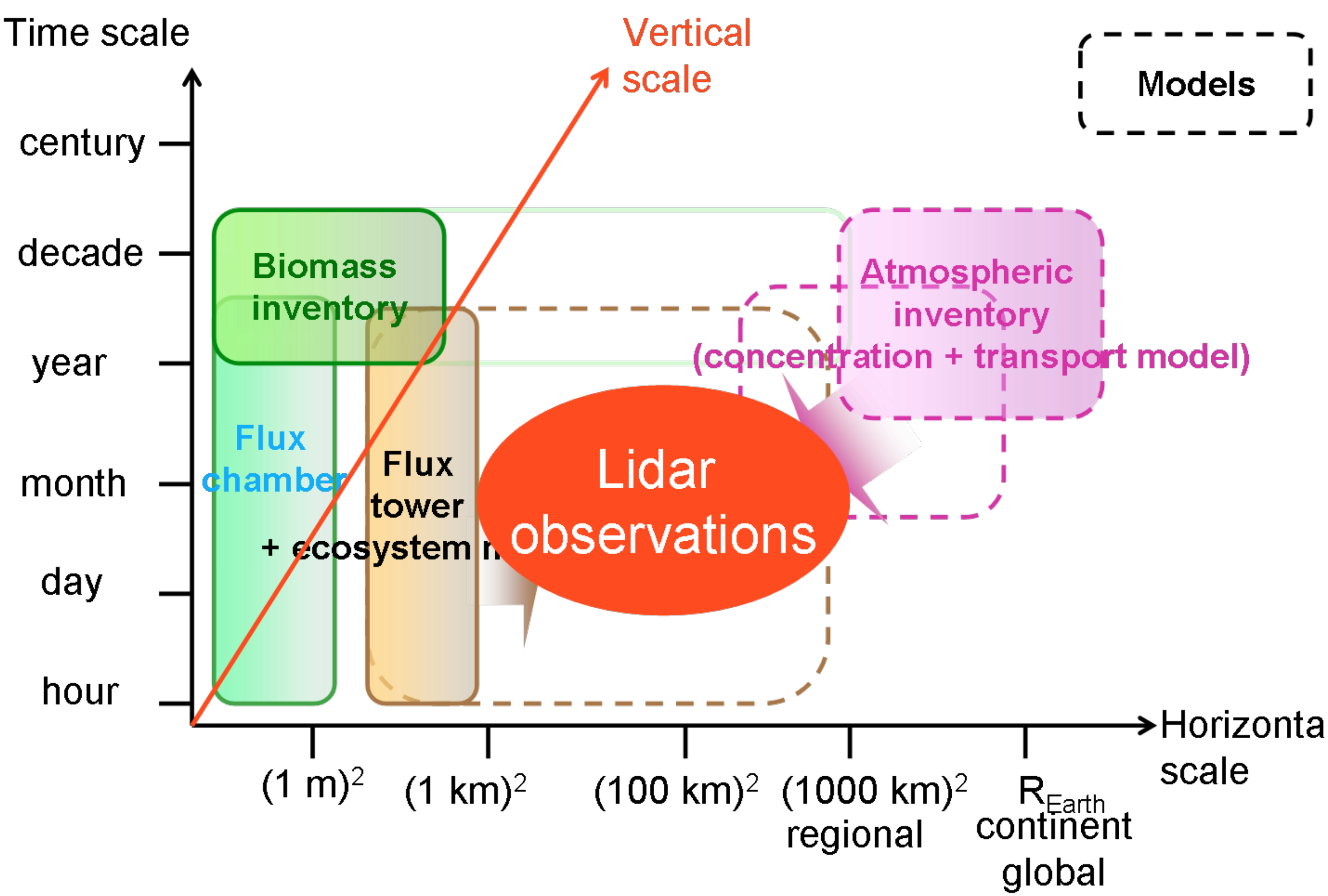
Scales of atmospheric constituents



Where is CO_2

Source: Atmospheric Chemistry and Physics by Seinfeld, Pandis

Scale gaps between CO_2 Surface flux measurements/ inventory and atmospheric measurements/models

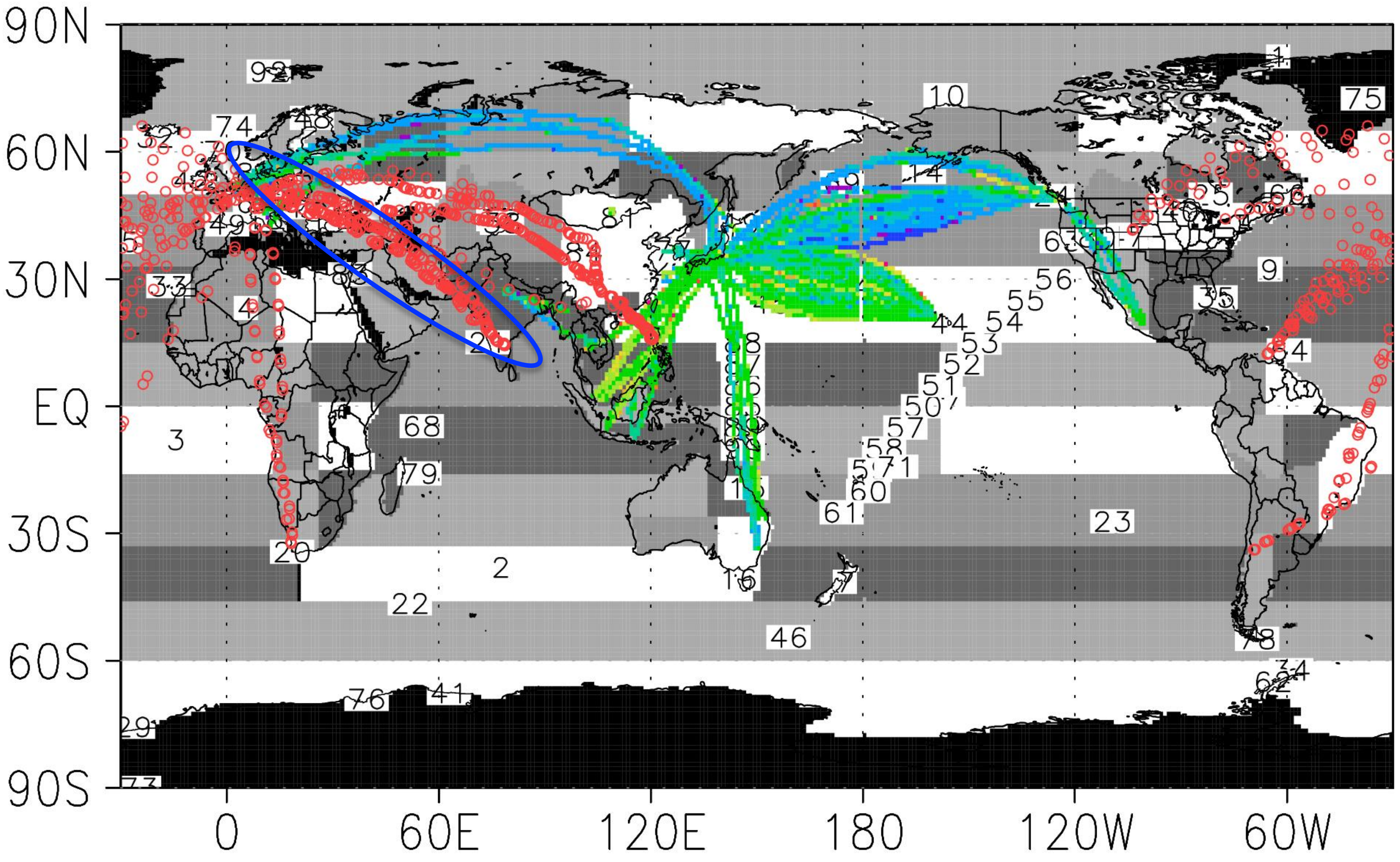


Source: http://lidar.abct.lmd.polytechnique.fr/index.php?page=carbon-cycle-2

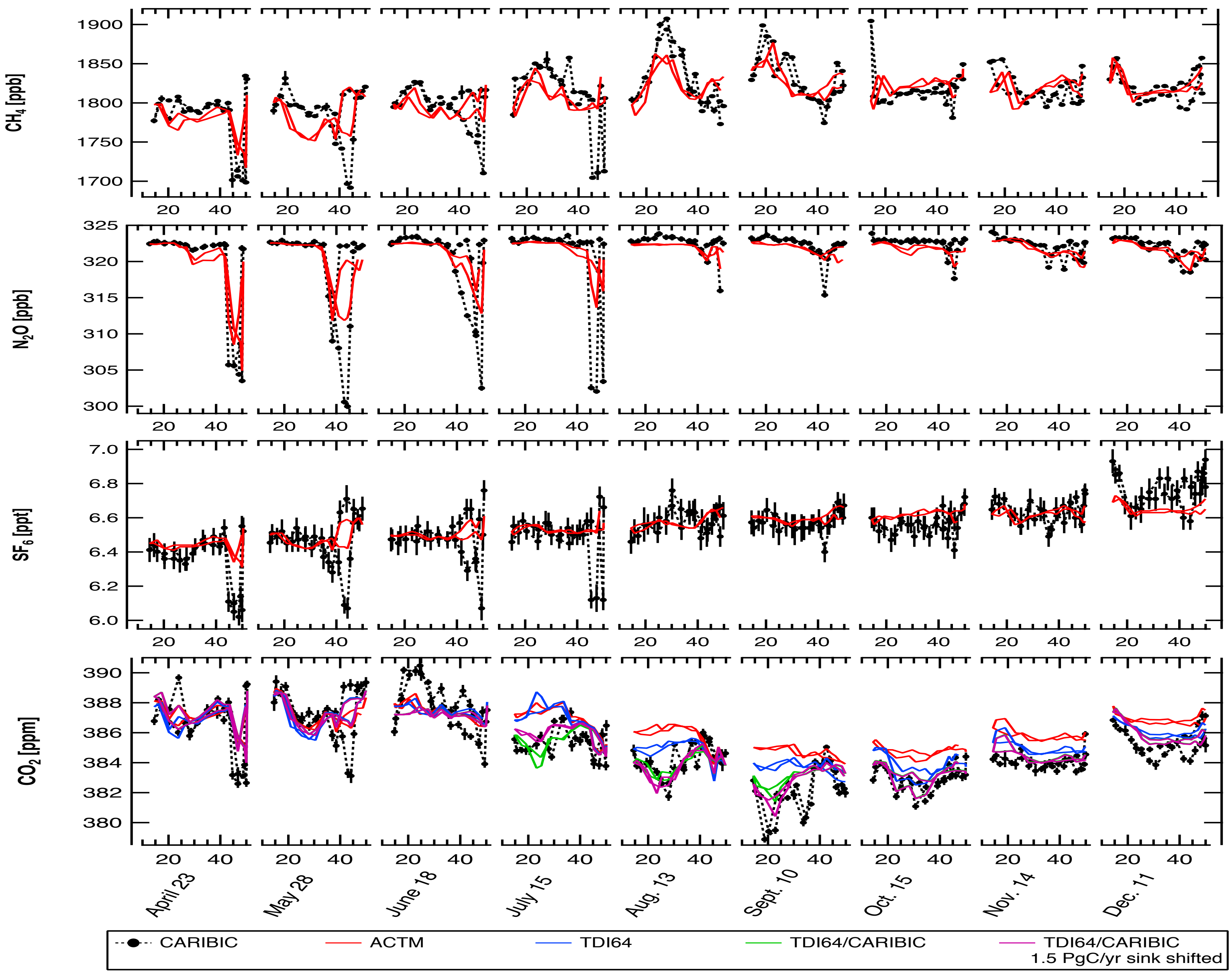
The Research Problem drives the development of tools

EXAMPLE ANALYSIS TOOLS

In situ (direct) measurement networks:
(surface + CONTRAIL + CARIBIC)



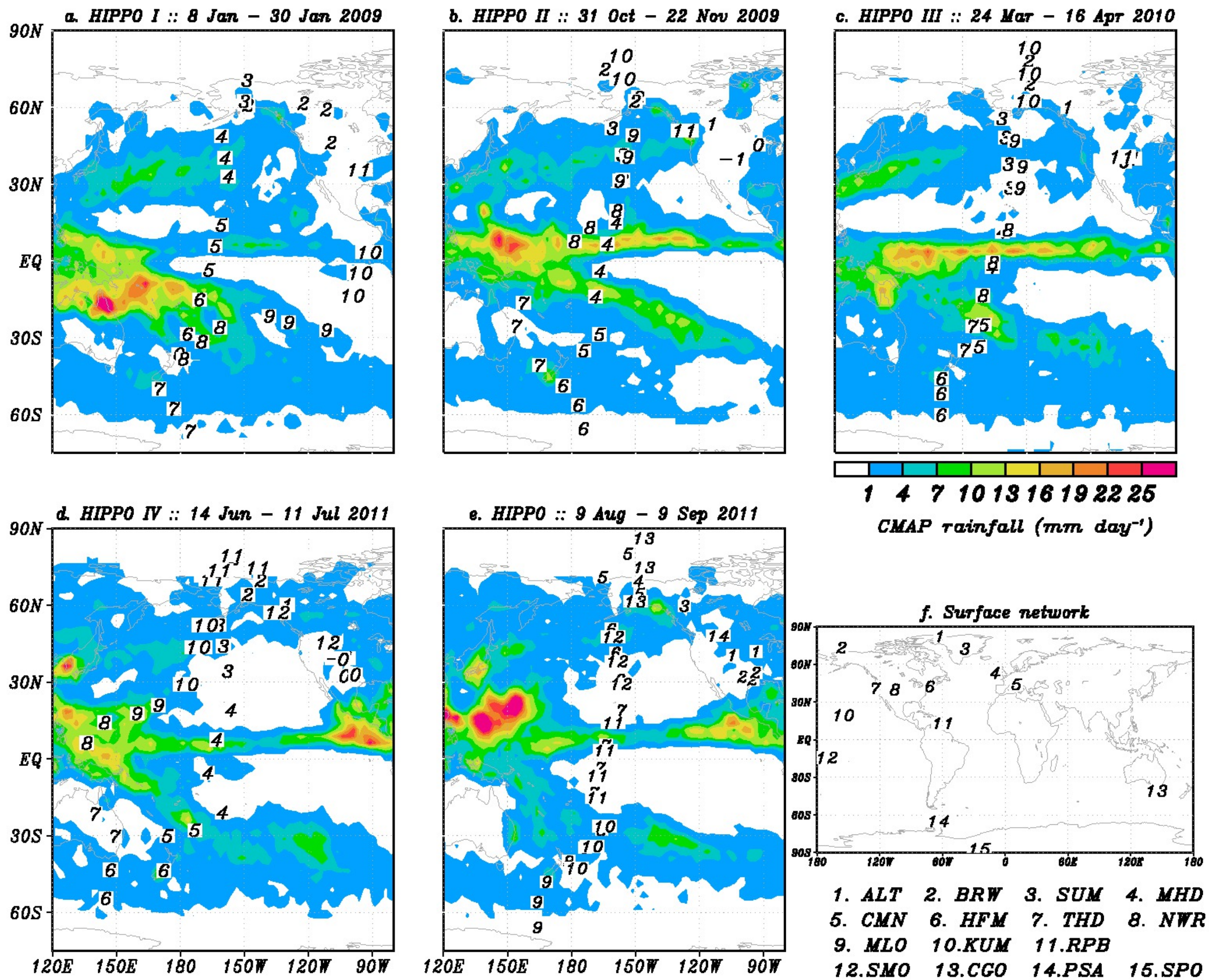
CARIBIC measurements between Frankfurt and Chennai in comparison with ACTM simulations (an early view of the South Asian GHG fluxes)



Plot for Readers

Patra et al., ACP, 2011; plotted by Tanja Schuck using IGOR;
Baker et al., ACAM, 2015

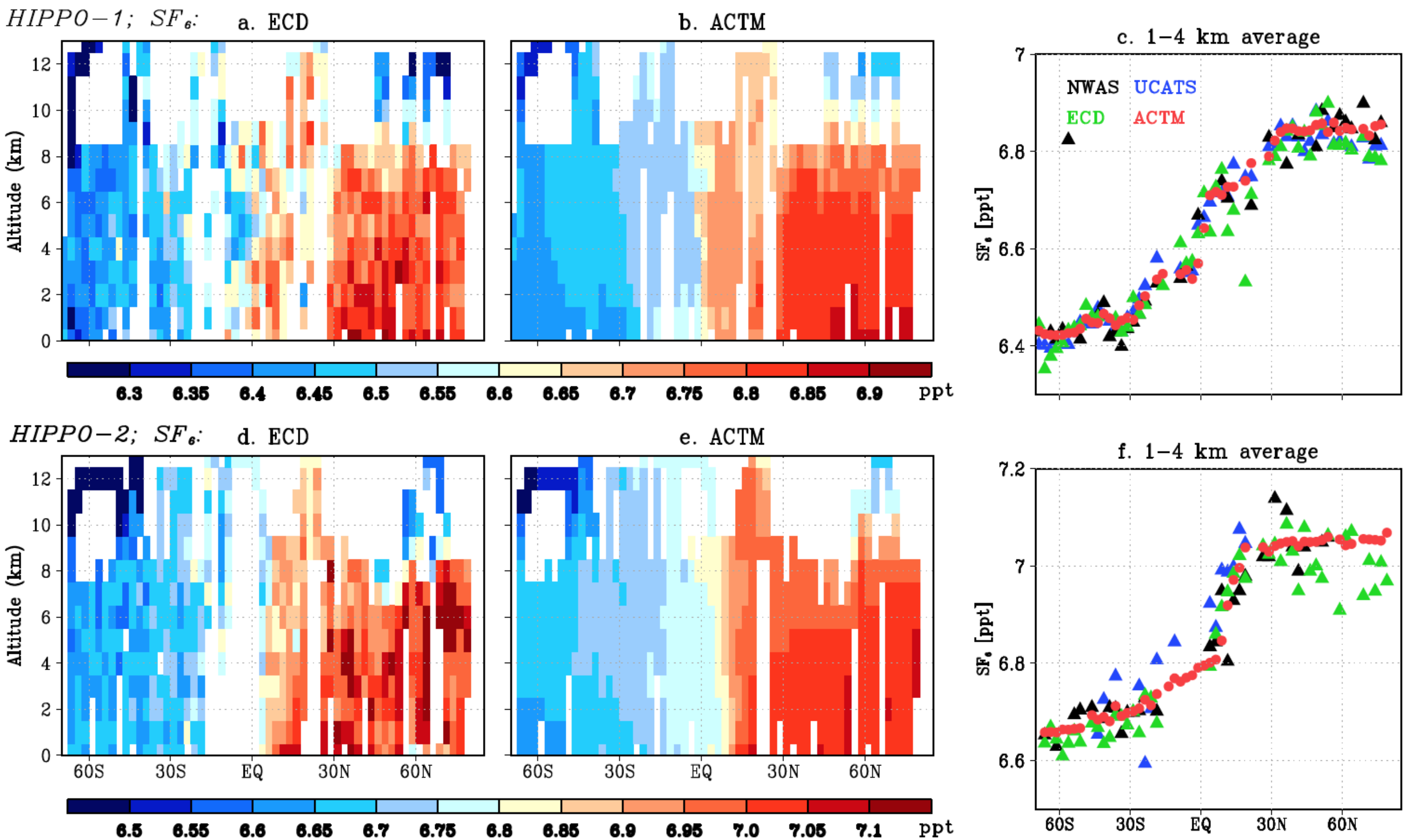
Summary of HIPPO flights and CMAP rainfall, and surface network



HIPPO data from central Pacific only

H#	Res. Flt#
1	#2-8
2	#1-7
3	#1-7
4	#1-7
5	#3-9

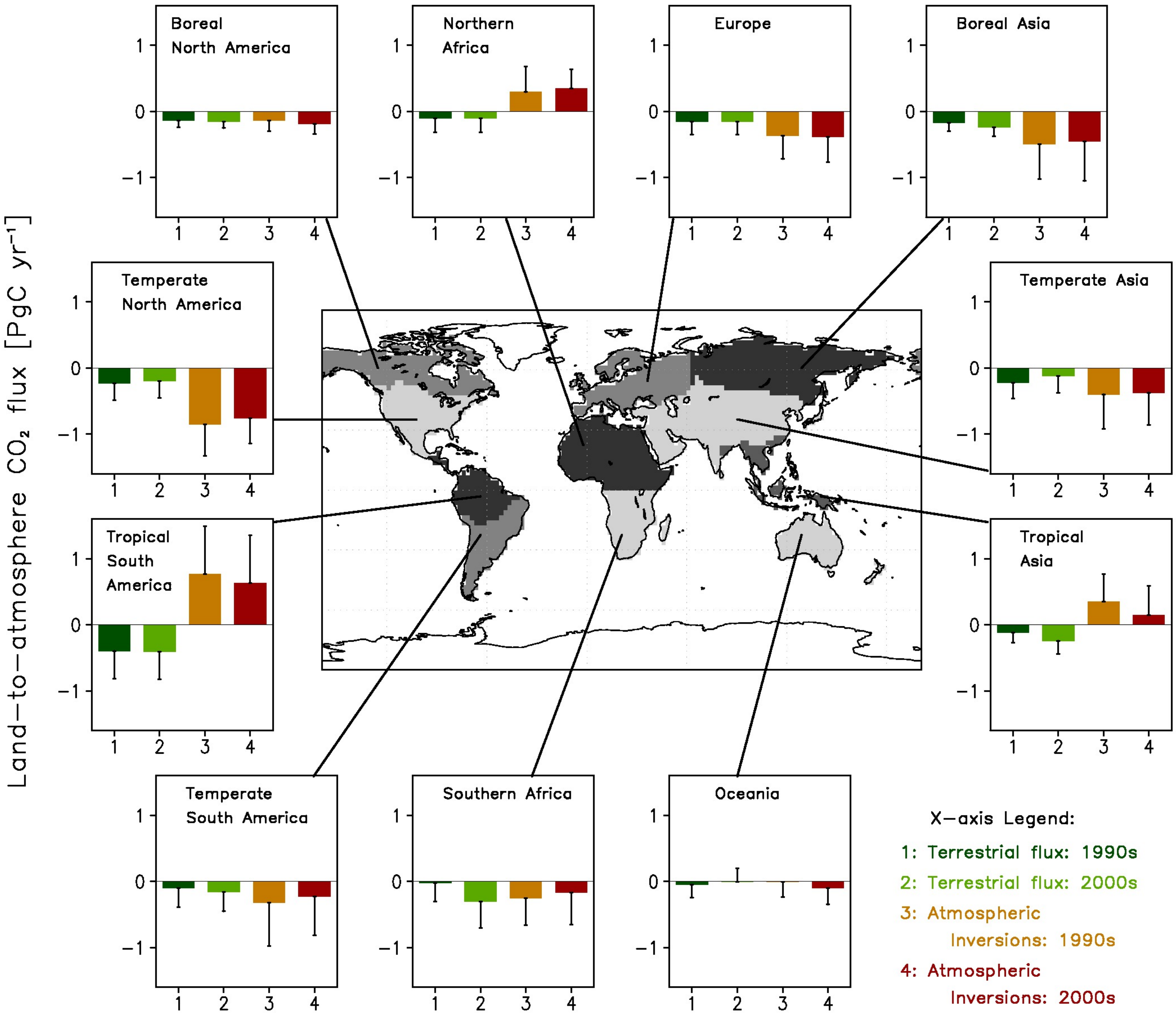
Latitude-altitude variations of HIPPO & ACTM SF_6 using EDGAR4.2 (validation of NH to SH transport in ACTM)



Obs –Mod (ppt)	-0.02 (6%)	0.02 (6%)	-0.02 (6%)	0.00 (0%)	-0.01 (3%)
NH – SH SF_6	H-1	H-2	H-3	H-4	H-5

Patra et al., 2009,2011,2014

Synthesis of Land fluxes from TDIs and DGVMs




```
'reinit';* 'set display color white';
'set xlopts 1 4 0.44'; 'Set ylopts 1 4
'set mproj off'; 'set font 0'
* Greyscale
'set rgb 50  50  50  50'; 'set rgb 51
'set rgb 52  90  90  90'; 'set rgb 53
'set rgb 54 130 130 130'; 'set rgb 5
'set rgb 56 170 170 170'; 'set rgb 5
'set rgb 58 210 210 210'; 'set rgb 5
'set rgb 60 250 250 250'
'set rgb 61  153   0   2';* warm red
'set rgb 62  196 121   0';* warm m
'set rgb 63    0  79   0';* warm gree
'set rgb 64    0 170 208';* warm bl
'set rgb 65  224   0   0';* bringt red
'set rgb 66  239  85  15';* dark ora
'set rgb 67  255 169   0';* warm ye
'set rgb 68   89 169   0';* bright gr
'set rgb 69    0  52 102';* dark blue
'set rgb 70  127   0 110';* dark pur
*
'open ./ctls/flux_summ.ctl'; 'open
'open ./ctls/flux_sum2.ctl'; 'open
'open ./ctls/flux_sum4.ctl'
* Boreal North America
'set vpage 0.5 2.5 4.2 6.2'; 'set gra
'set x 0.5 4.5'; 'set y 1'; 'set z 1'; 'se
'set gxout bar'; 'set bargap 30'; 'se
'set ccolor 63'; 'set baropts filled';
'set ccolor 68'; 'set baropts filled';
'set ccolor 62'; 'set baropts filled';
'set ccolor 61'; 'set baropts filled';
'set ccolor  1'; 'set gxout errbar'; 'c
'set strin
'set strin
'set line
'set vpag
*Temperate North America
```

Figure 6.15, IPCC-AR5-WG1, analysis by P. Patra

Large scale data analysis
using Principal Component Analysis

Correlations in time series (some preliminaries) :

Univariate time series :

$$C(\tau) = \langle (x(t) - \bar{x}) \ (x(t+\tau) - \bar{x}) \rangle$$

Bi-variate time series :

$$C_{x,y}(\tau) = \langle (x(t) - \bar{x}) \ (y(t) - \bar{y}) \rangle$$

Multivariate time series : $x_1(t), x_2(t), x_3(t) \ldots\ldots x_N(t)$

$$t = 1, 2, 3 \ldots . T$$

Pair wise correlations : $\dfrac{N(N+1)}{2}$

Principal Component Analysis (PCA/EOFs)

Correlation matrix : $$C = D^{\mathrm{T}} D$$

D is T x N data matrix, with each column representing a time series.

Then, C is a square matrix of order N.

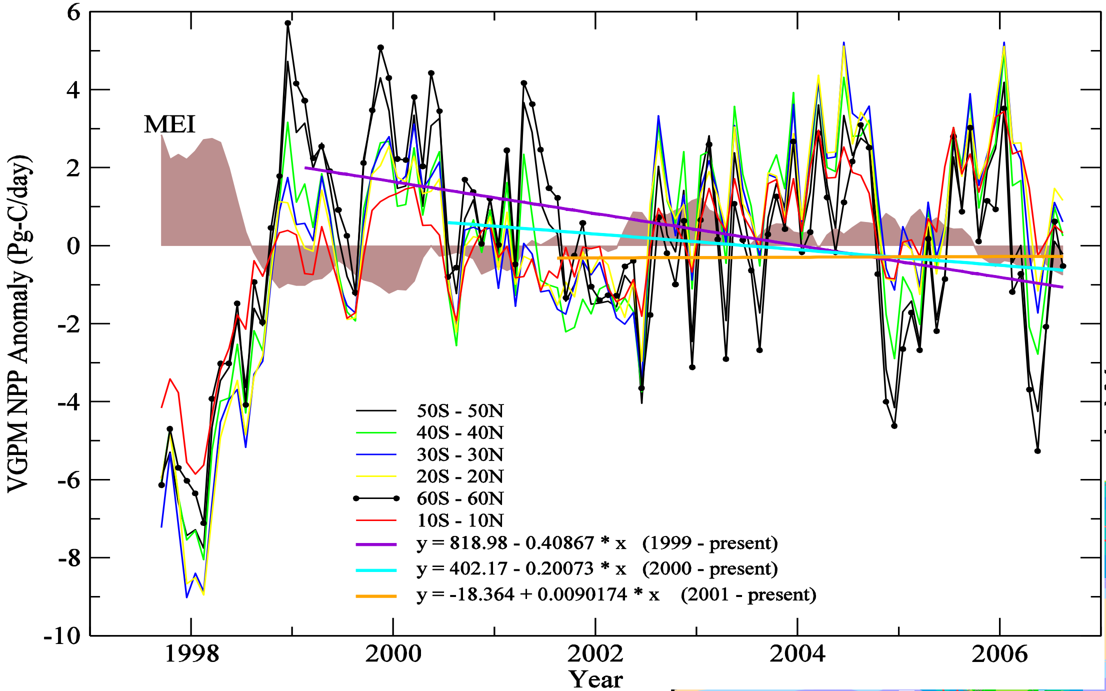
Spectra of correlation matrix C.

$$C \; u_i = \lambda_i \; u_i , \qquad i = 1, 2, \ldots N$$

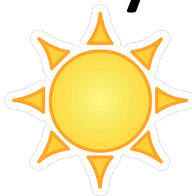
Positive semi-definite eigenvalues : $\lambda_i \geq 0$

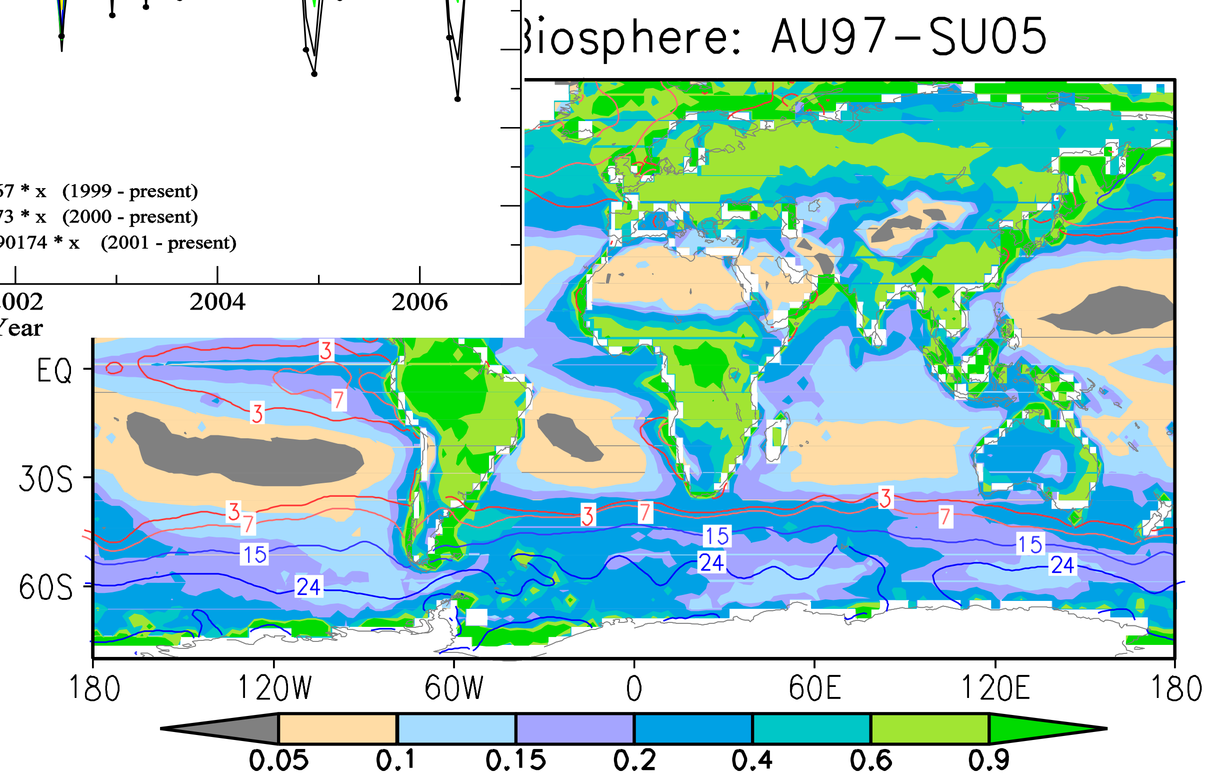
Real, symmetric : $C = C^{\mathrm{T}}$

Our biosphere: trends and variability



Behrenfeld et al., 2006 (Edited)





Patra et al., 2007

Principal component analysis: Empirical orthogonal function (EOF)

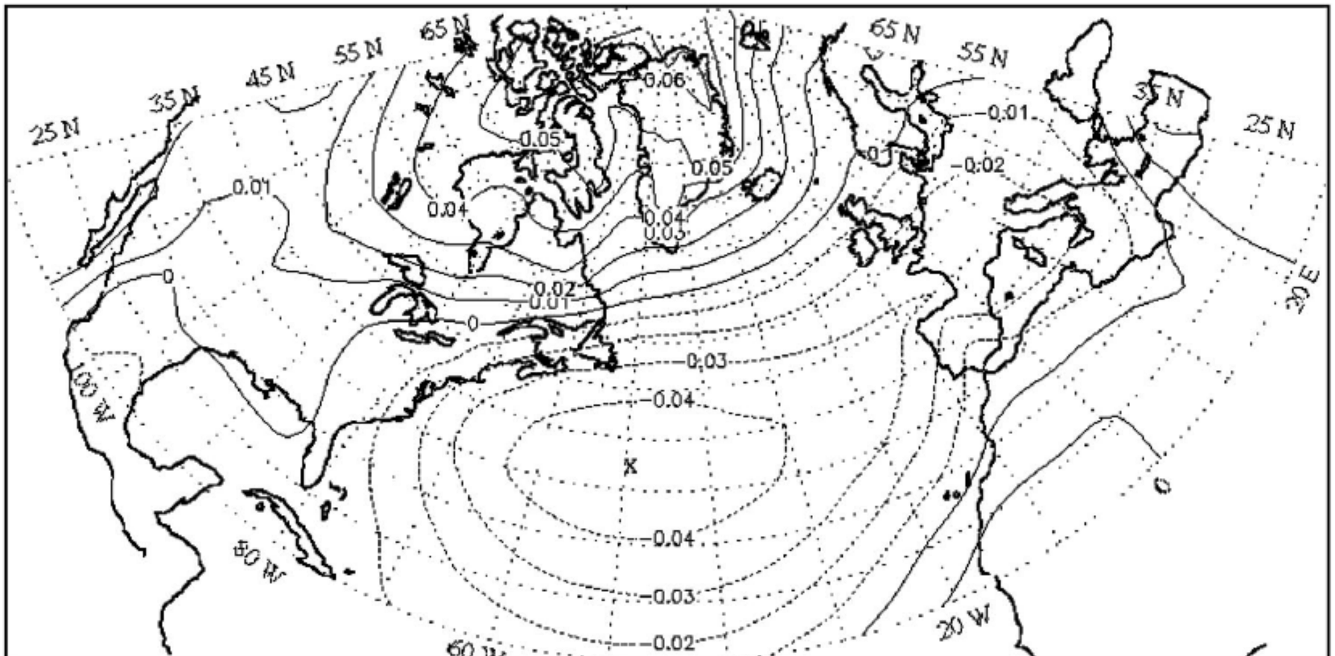


FIG. 1. The NAO pattern from the EOFs of monthly mean SLP correlation matrix with the geographical map of the domain of analysis in the background. The contours are drawn after averaging over the first two dominant EOFs. Note the north-south dipole shown as closed contours, in mid Atlantic (dotted contour) and over Greenland (solid contours).

anomaly $z'(x,t)$ will be used that will have zero mean $[\overline{z'}(x)=0]$ and is rescaled such that its variance $\langle z'(x)^2\rangle$ is unity. If the observations were taken n times at each of the p spatial locations and the corresponding anomalies $z'(x,t)$ assembled in the data matrix $\mathbf{Z}$ of order $p\times n$, then the spatial correlation matrix of the anomalies is given by

$$\mathbf{S}=\frac{1}{n}\mathbf{Z}\mathbf{Z}^{\dagger}. \qquad (2.1)$$

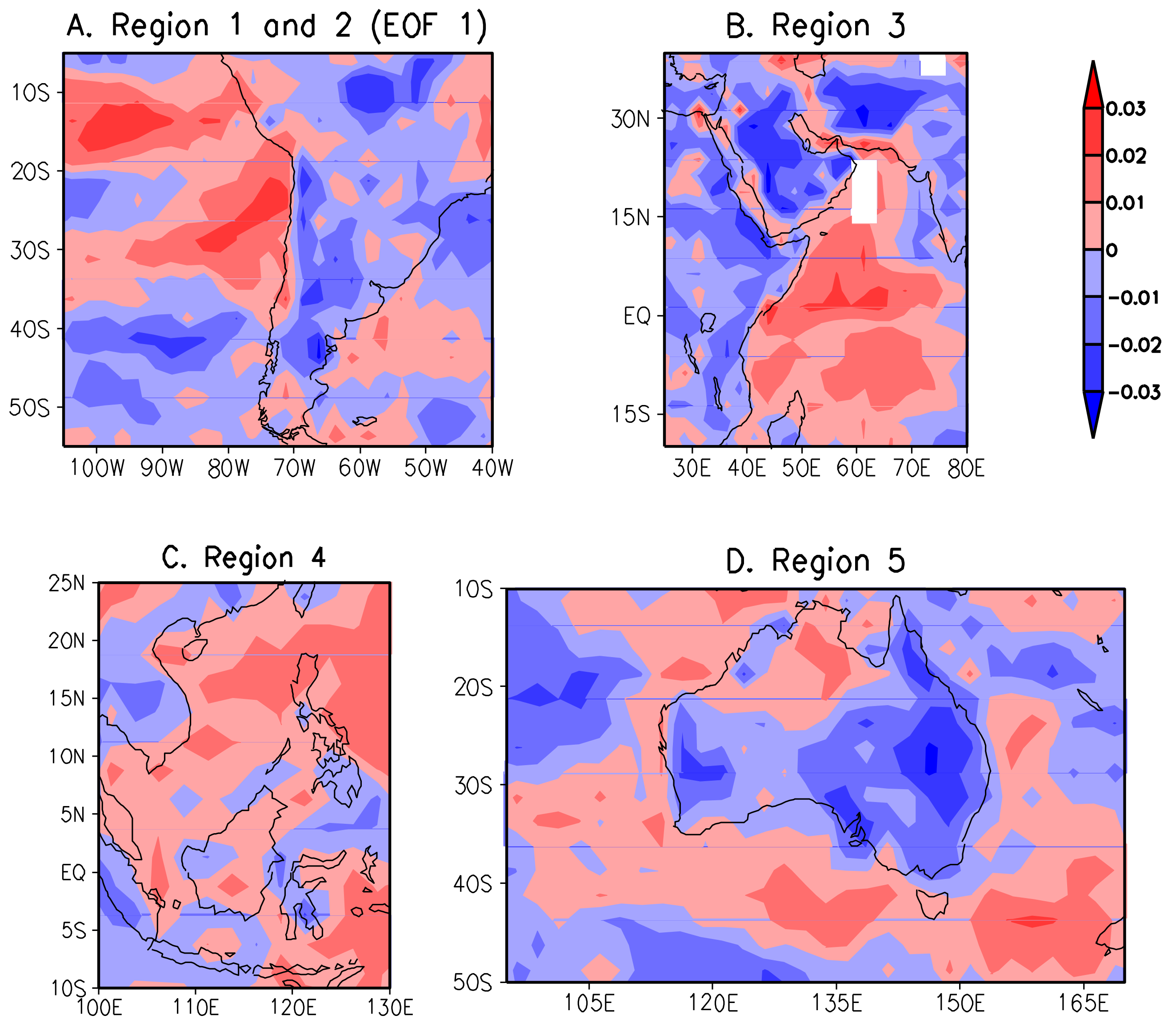
location. If the eigenvalue corresponding to the mth eigenmode is λ_m, then the percentage variance associated with that mode is given by

$$\lambda_m / \sum_{i=1}^{p} \lambda_i \times 100,$$

p denotes the total number of eigenmodes. The PCs for each modes of variability (u_m) can be calculated as

$$u_m = \sum_i e_{im} x_i .$$

EOF-1 for selected regions



Possible mechanism of strong land-ocean coupling – correlations of timeseries for 10 years

Lead-lag correlation for identifying the cause-and-effect relationship

	Chl-a/ SST	NDVI/ PCP	Chl-a/ NDVI
Region 1 (105°W-50°W, 35°S-5°S)	-0.54 (-2)	0.44 (-2)	-0.27 (0)
Region 2 (75°W-50°W, 55°S-30°S)	-0.39 (-1)	0.17 (-2)	-0.26 (-4)
Region 3 (25°E-80°E, 20°S-40°N)	-0.61 (0)	0.46 (-2)	-0.70 (1)
Region 4 (100°E-130°E, 10°S-25°N)	-0.70 (0)	0.36 (0)	-0.42 (0)
Region 5 (105°E-170°E, 50°S-25°S)	-0.44 (-1)	0.41 (-2)	-0.34 (-4)

Chl-a lags SST

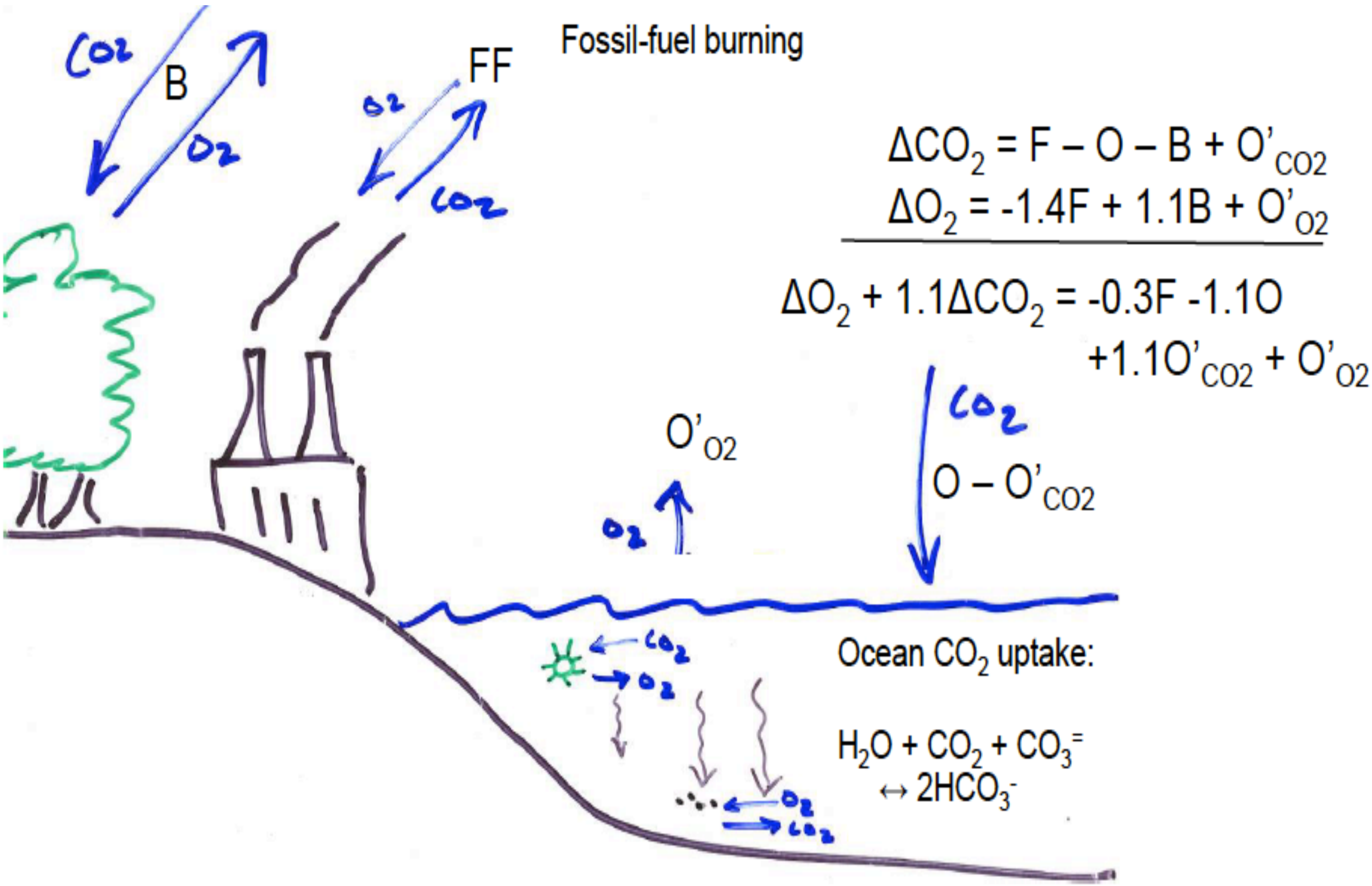
P. K. Patra, SOLA, 2007

My little Fortran code – notable is memory allocation

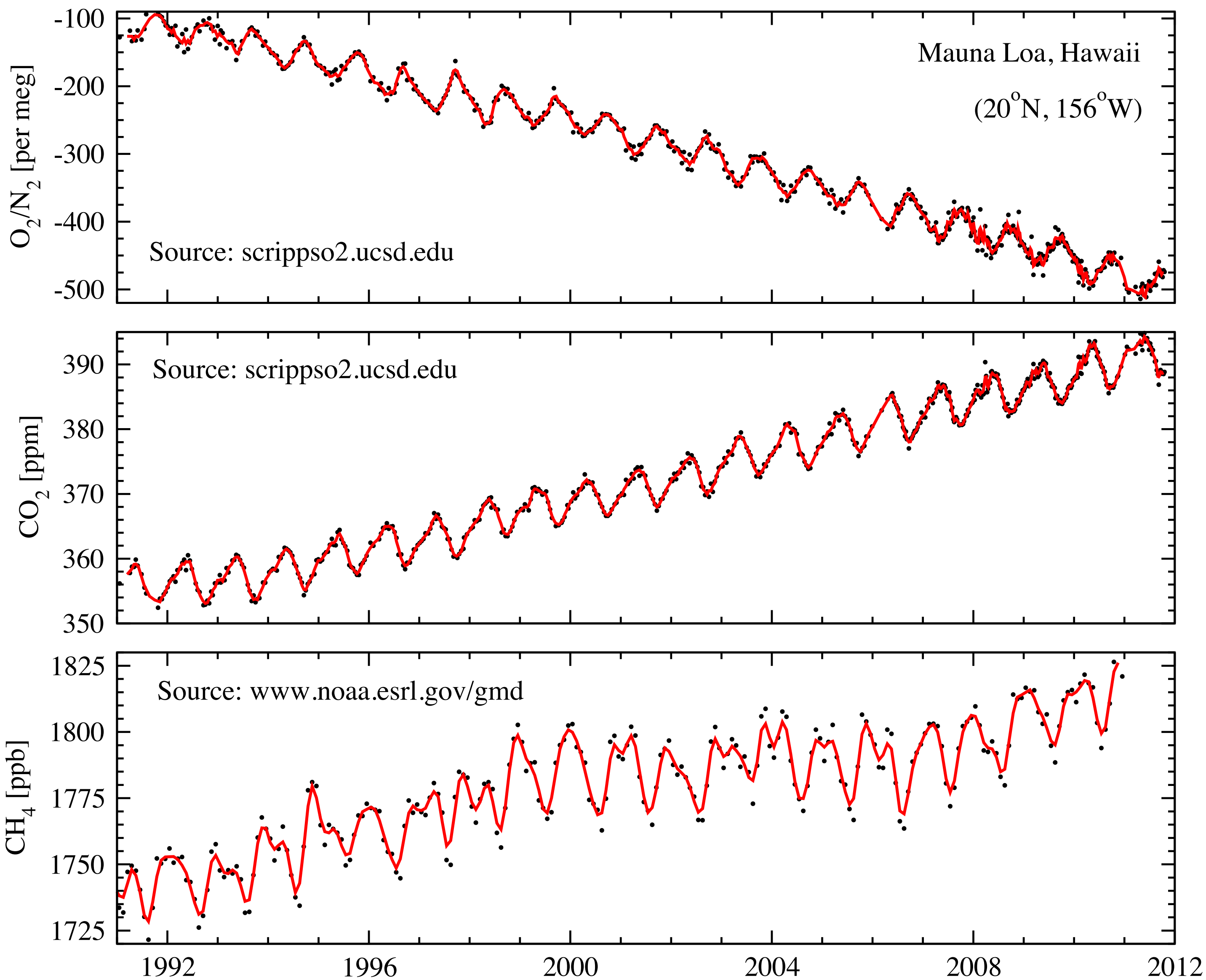
```
! compile with : ifort eofbios2p5.f90 -r8 (-L/usr/local/lib) -llapack -lblas
!
   PROGRAM EOFBIO
   implicit none
   integer :: nyear,nmon,nlat,nlon,nlat1,nlat2,nun,iun,count
   parameter (nyear=9, nmon=4, nlat=72, nlon=144,nun=20)
   integer :: i,j,k,iy,im,idim,jdim,kdim,idx,idy,irec,num0
   real   :: c1,sum,lb,ub
   real *4 :: LS(nlon,nlat)
   real,allocatable :: cov(:,:),cor(:),avg(:,:),BIOi(:,:)
   real*4,allocatable :: BIO(:,:,:,:),BIOp(:,:,:,:),BIOa(:,:,:), chla(:,:), ndvi(:,:)
   real,allocatable :: d(:),e(:),tau(:),work(:),w(:),z(:,:)
   real*4,allocatable :: evec(:,:),pca(:,:)
   integer, allocatable :: IWORK(:),IFAIL(:)
   integer :: n,ml,m5,nev,m,info,il,iu,ils,jls,iev,nls,ndim
   character*1 JOBZ,RANGE,UPLO
!
   OPEN(31,FILE='./data/sw_chlo2.grd',status='old',form='unformatted',access='direct',recl=nlon*nlat)
   OPEN(32,FILE='./data/sw_ndvi2.grd',status='old',form='unformatted',access='direct',recl=nlon*nlat)
   OPEN(33,FILE='./data/sw_bios2.grd',status='unknown',form='unformatted',access='direct',recl=nlon*nlat)
   OPEN(71,FILE='evalbios.dat',status='unknown')
   OPEN(62,FILE='pcabios.dat',status='unknown')
!
   allocate (  BIO(1:nyear,1:nmon,1:nlat,1:nlon) ); BIO  = 66.83439
   allocate ( chla(1:nlat,1:nlon) ); chla = 66.83439
   allocate ( ndvi(1:nlat,1:nlon) ); ndvi = 66.83439
   allocate ( BIOa(1:nmon,1:nlat,1:nlon) ); BIOa = 66.83439
   allocate ( BIOp(1:nyear,1:nmon,1:nlat,1:nlon) ); BIOp  = 66.83439
   !include "time.h"
   irec = 0   ! 0 (nyear=8), 4 (nyear=7)
   do iy = 1, nyear; do im = 1, nmon; irec = irec + 1 !; print *,'irec =',irec
      read (31, rec=irec) ((chla(j,i), i=1,nlon), j=1,nlat)  ! 1x1 deg Grads
      read (32, rec=irec) ((ndvi(j,i), i=1,nlon), j=1,nlat)  ! 1x1 deg Grads
      do i = 1, nlon; do j = 1, nlat
         if ( chla(j,i) .le. 60.0) then
            BIO(iy,im,j,i) = chla(j,i)
         else if ( ndvi(j,i) .le. 60.0) then
            BIO(iy,im,j,i) = ndvi(j,i)
```

EVALUATION OF MODELS – CO2 EXAMPLE

Coupled CO_2 and O_2 system (source: Keeling and Keeling)



What's the evidence of increased CO_2 loading?



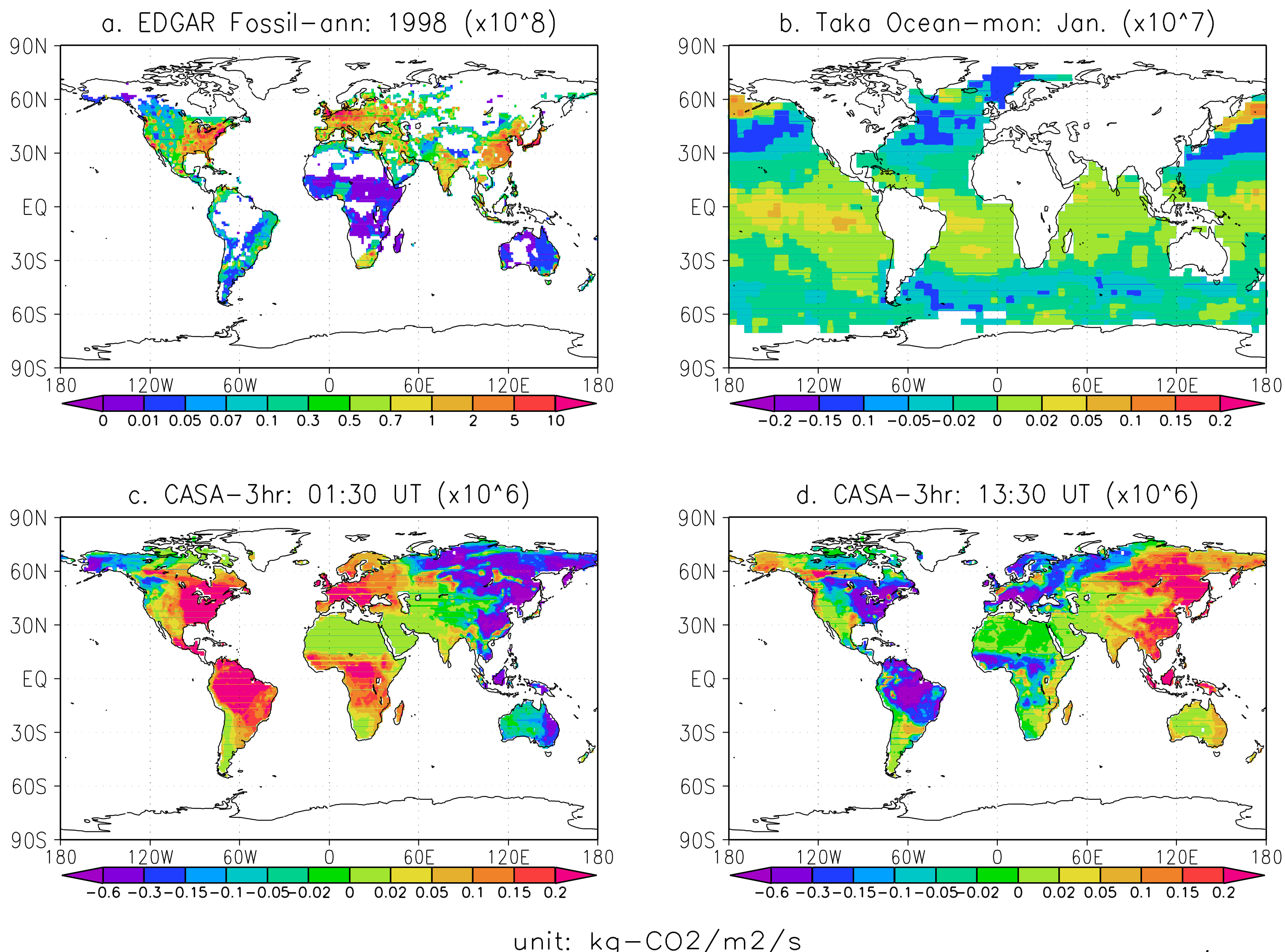
For ACTM simulations:

Ishidoya et al., 2012

Patra et al., 2011a

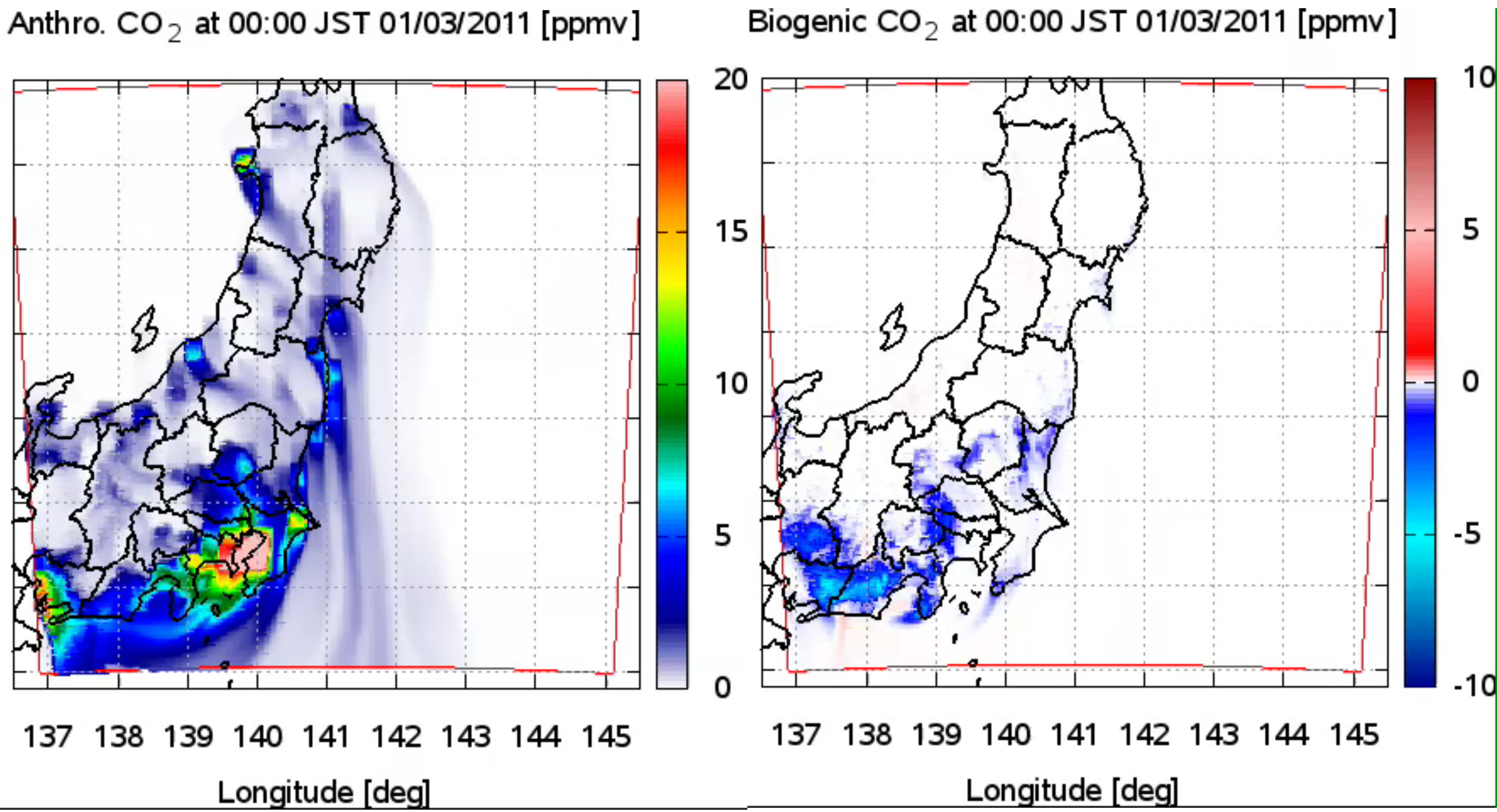
Patra et al., 2011b

Surface flux distributions – overview



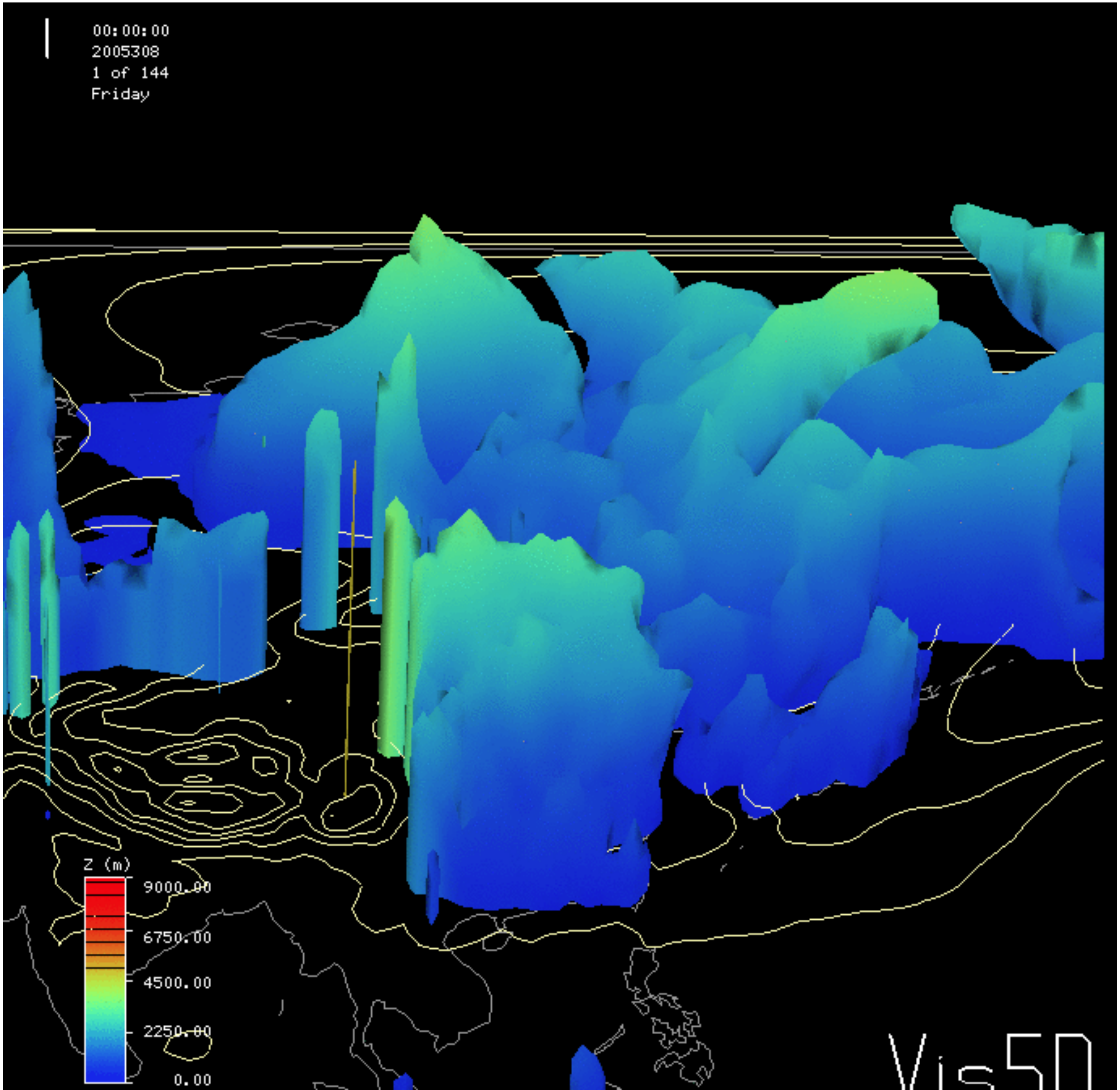
EDGAR: JRC/PBL
CASA: Randerson et al., 1997
Ocean: Takahashi et al., 2009

Movies of WRF-CO_2 simulations



Courtesy of Masayuki Takigawa, JAMSTEC; Ballav et al., JMSJ, 2012

Spatial scales of atmospheric-CO_2 hourly variations



With help from M. Takigawa

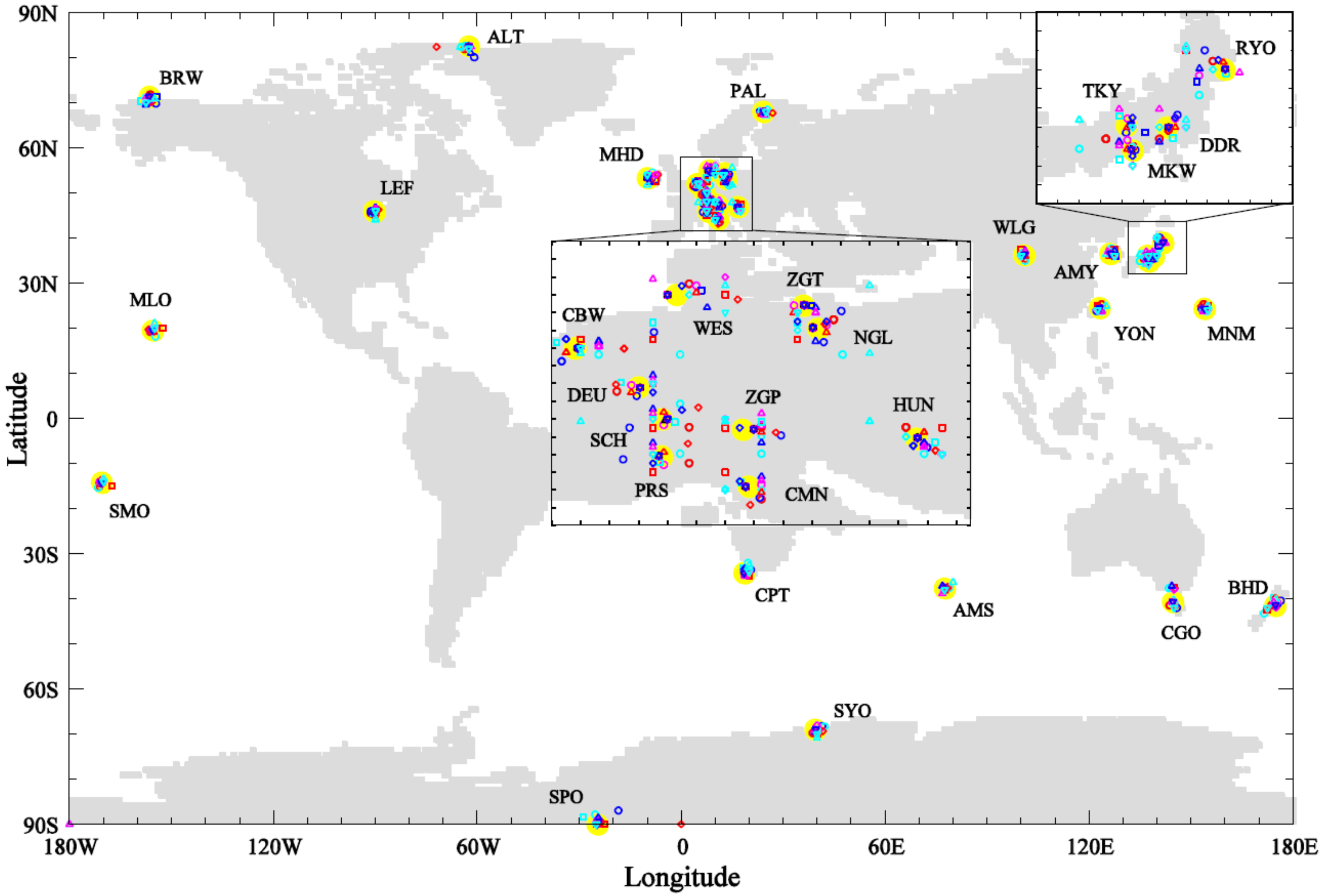
DoY 308-313; 4-9 Nov 2005

Isosurface of 380 ppm CO_2 concentrations coloured by Height (blue: ground red: ~9km height)

White contour: Sea-level Pressure

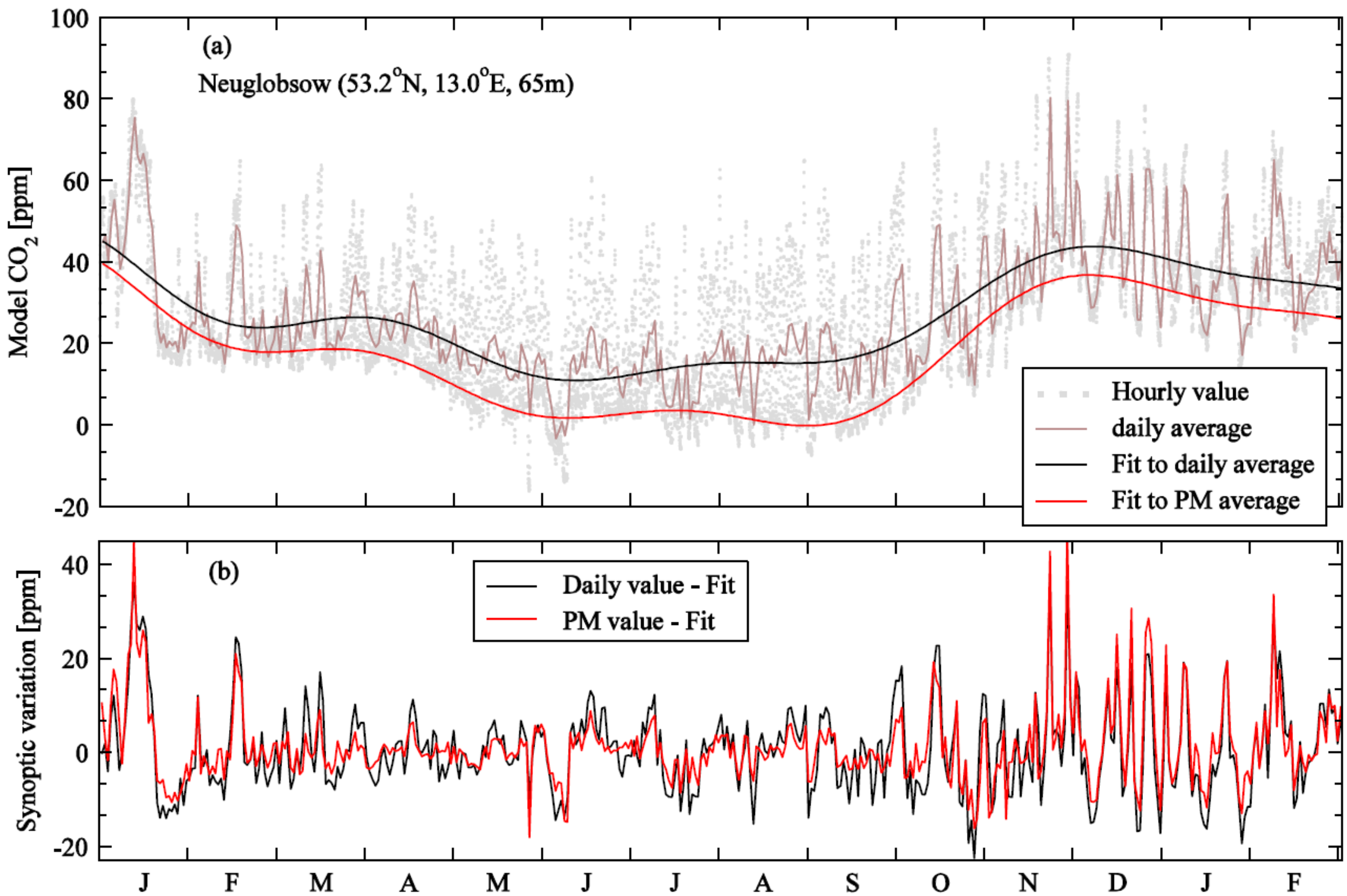
Monthly average CO_2 shows mainly inter-hemispheric difference

Measurement locations of Hourly CO_2

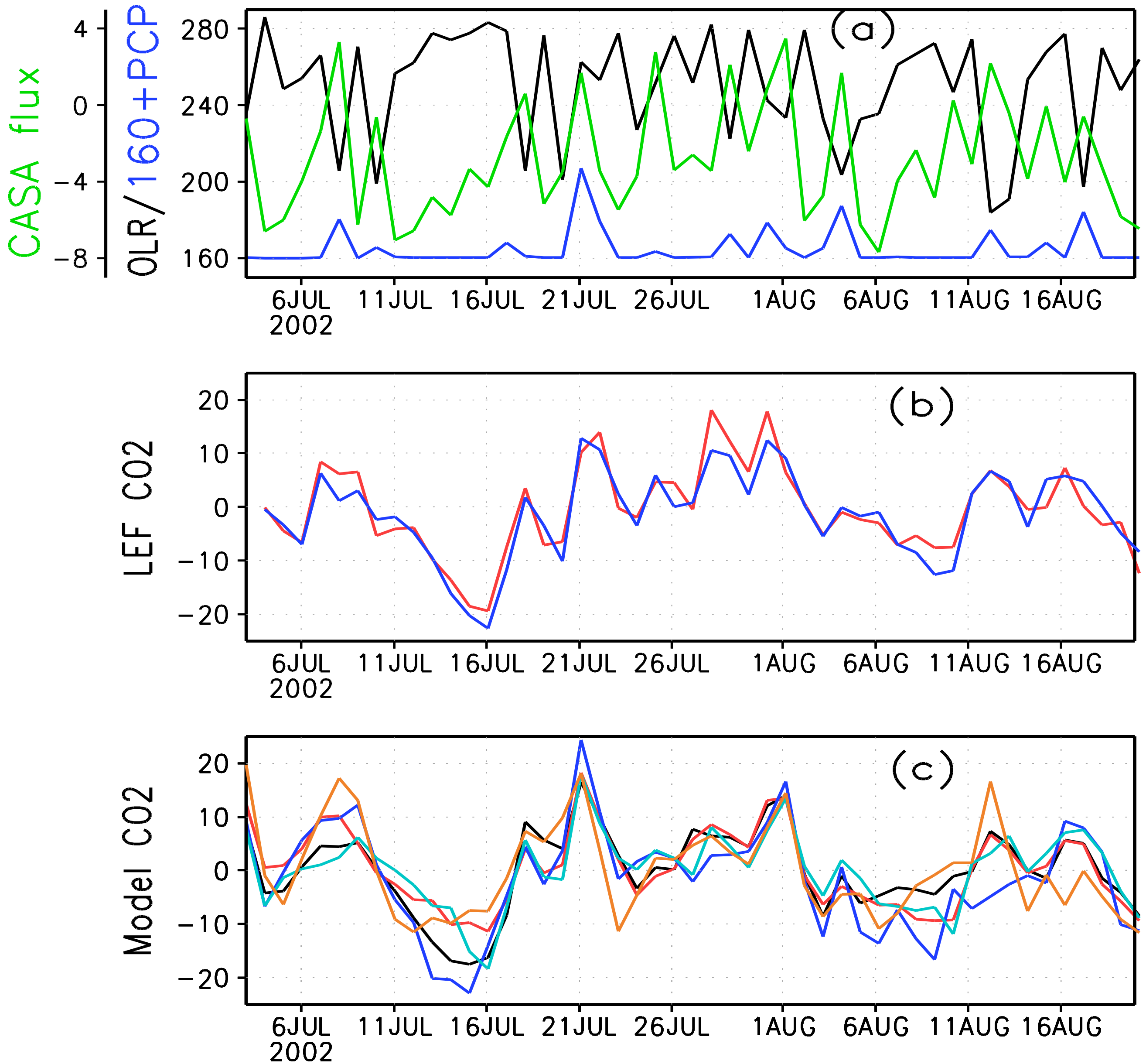


TransCom (Transport model inter-comparison project) – CO_2 continuous experiment
Law et al., GBC, 2008; Patra et al., GBC, 2008

What is synoptic variability in CO_2?



Model-data comparison



LEF Tower, 76m Wisconsin

a. Weather conditions and flux

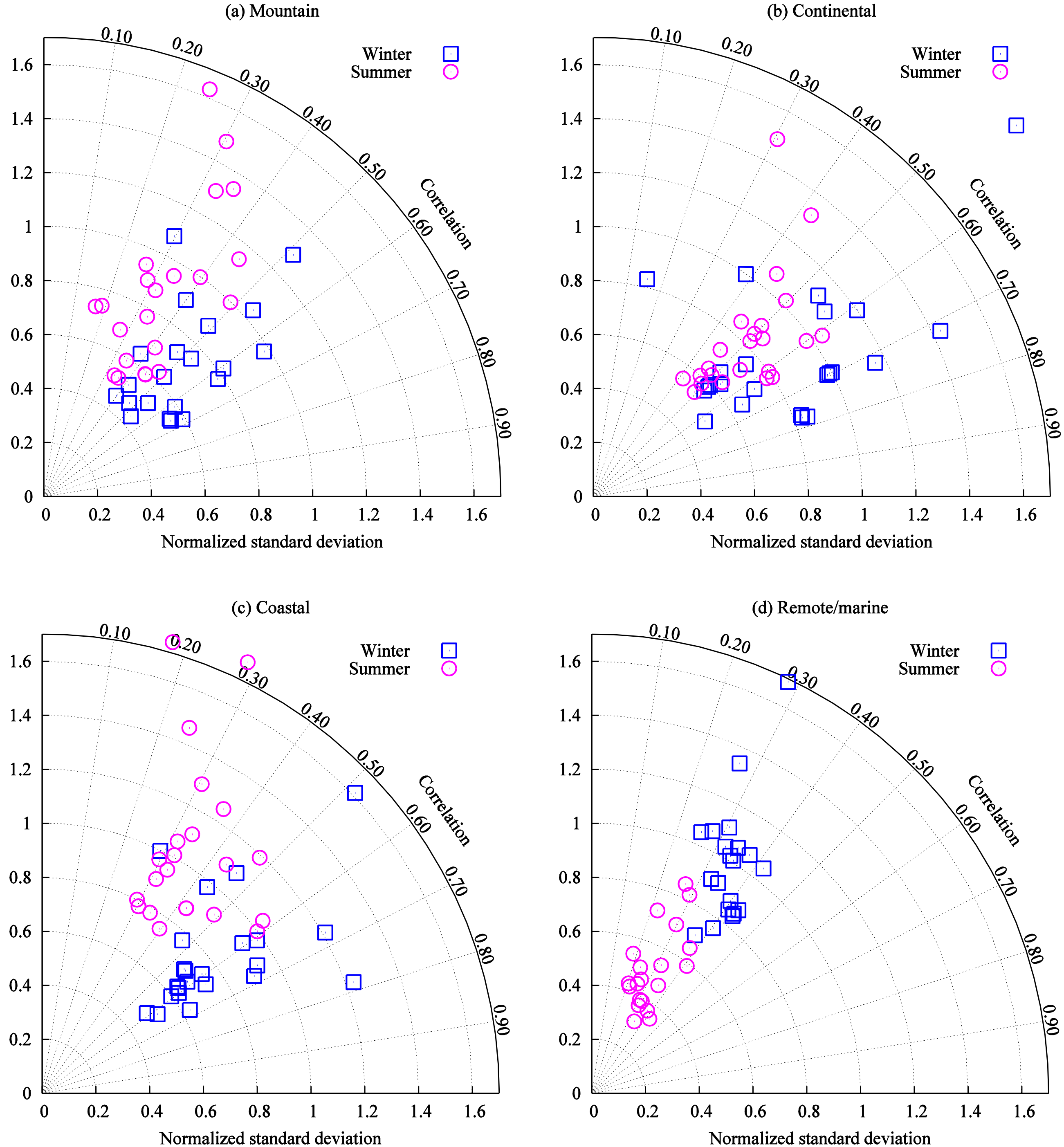
b. Observed CO_2 synoptic variability

c. Modelled variability

Correlation (R): phase of model-data comparison

NSD (Sd_{mod}/SD_{obs}) : amplitude of variability (1 ideally)

Taylor Diagrams



Salient features:

Simulations accuracy during winter is better than summer

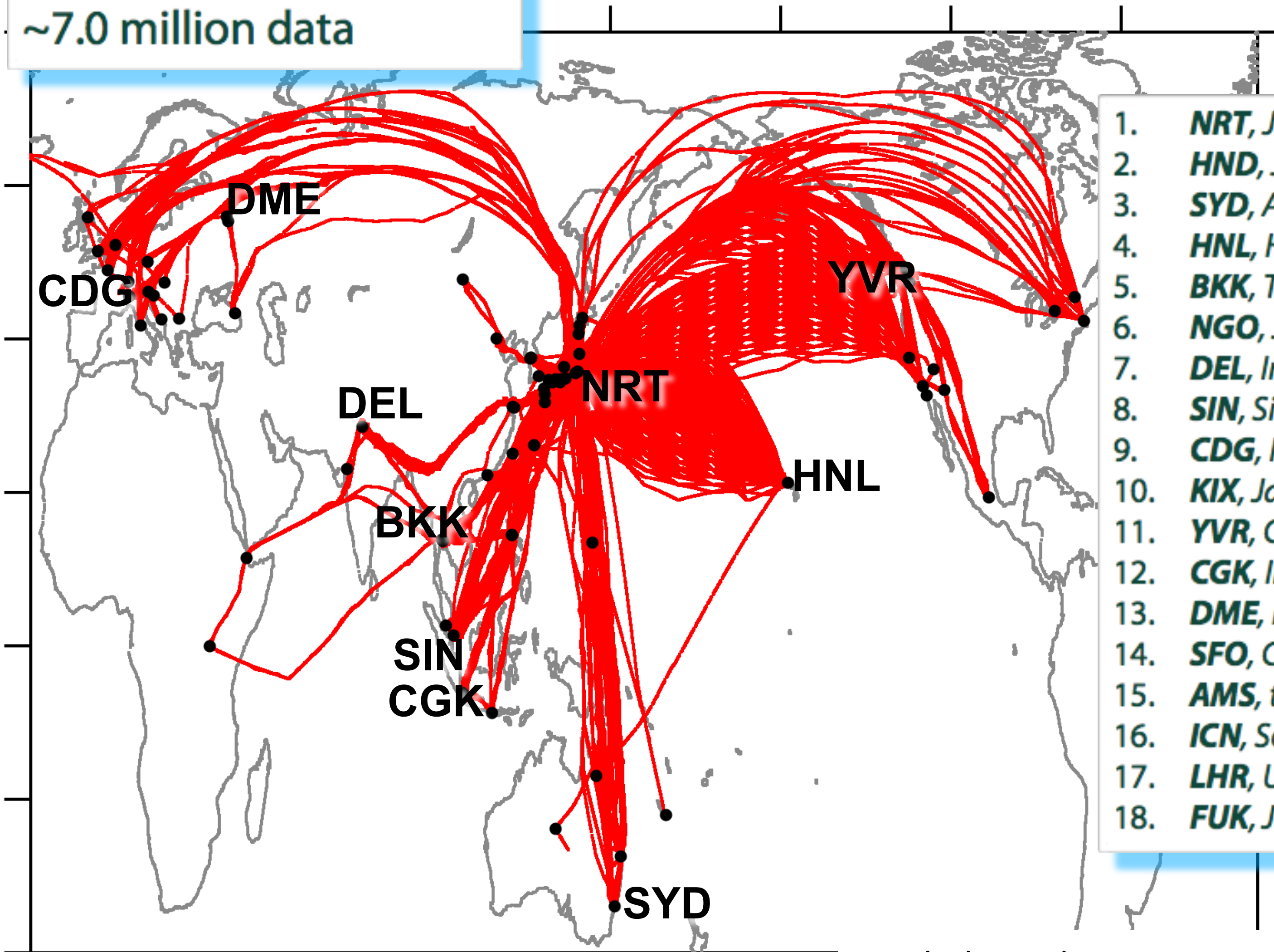
a. Sampling at mountain sites is tricky

b. Continental sites are best simulated. Note diurnally varying flux

d. Oceanic flux variability is weaker

Flight statistics CME (Nov. 2005–Dec. 2014)

~12,000 flights
~23,000 vertical profiles
~7.0 million data



1. ***NRT**, Japan* 7270
2. ***HND**, Japan* 2937
3. ***SYD**, Australia* 1581
4. ***HNL**, Hawaii* 1455
5. ***BKK**, Thailand* 1273
6. ***NGO**, Japan* 855
7. ***DEL**, India* 787
8. ***SIN**, Singapore* 686
9. ***CDG**, France* 683
10. ***KIX**, Japan* 635
11. ***YVR**, Canada* 440
12. ***CGK**, Indonesia* 422
13. ***DME**, Russia* 409
14. ***SFO**, California* 366
15. ***AMS**, the Netherlands* 238
16. ***ICN**, South Korea* 199
17. ***LHR**, UK* 196
18. ***FUK**, Japan* 189

Machida et al., 2008; Matsueda et al., 2008

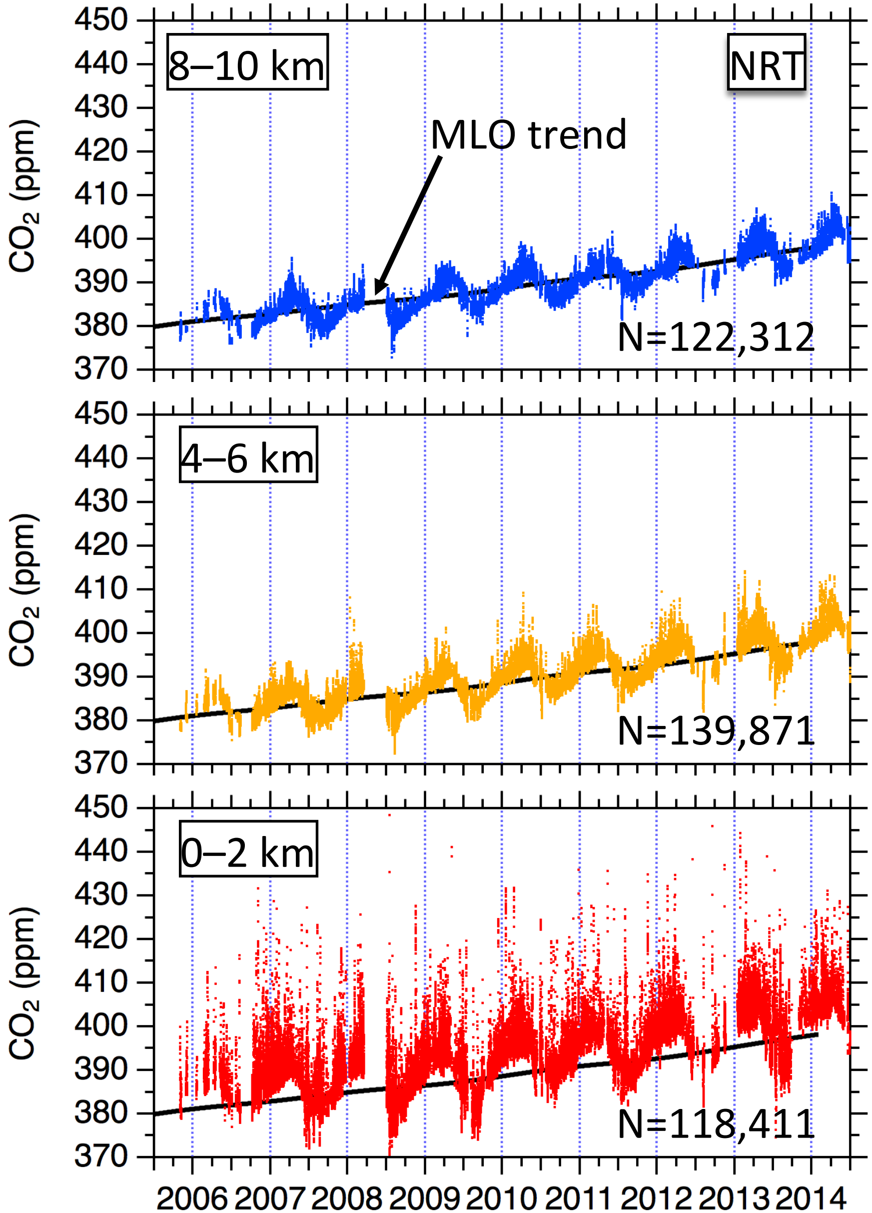
http://www.cger.nies.go.jp/contrail/

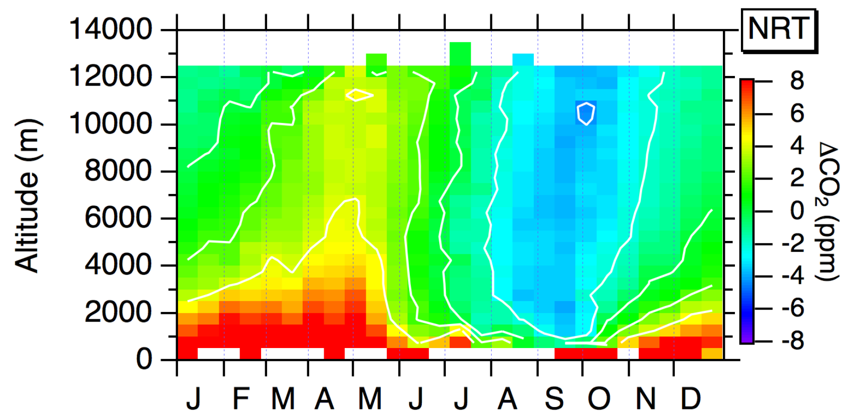
CONTRAIL
Comprehensive Observation Network for Trace gases by Airliner

CO_2 over Tokyo, Japan (NRT)

http://www.cger.nies.go.jp/contrail/

CONTRAIL
Comprehensive Observation Network for Trace gases by Airliner

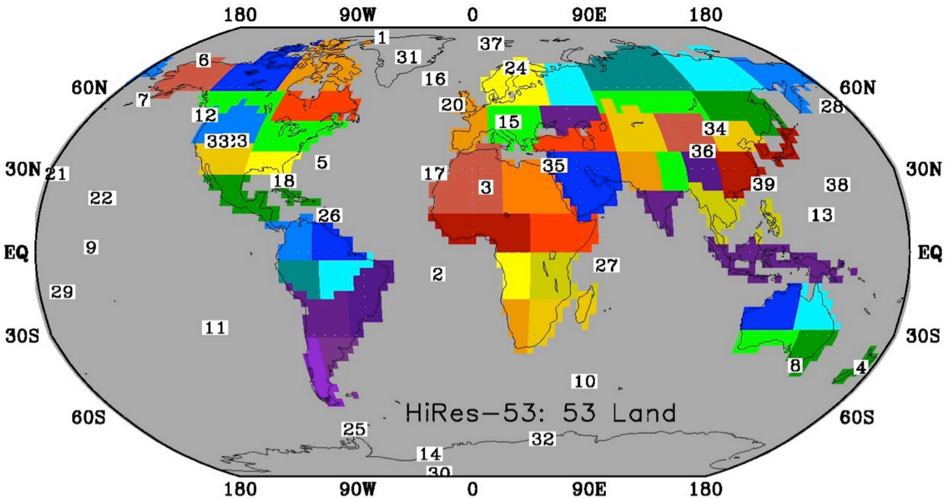


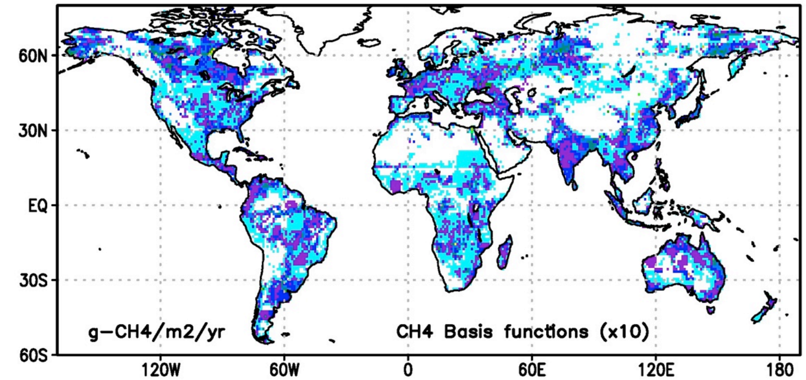


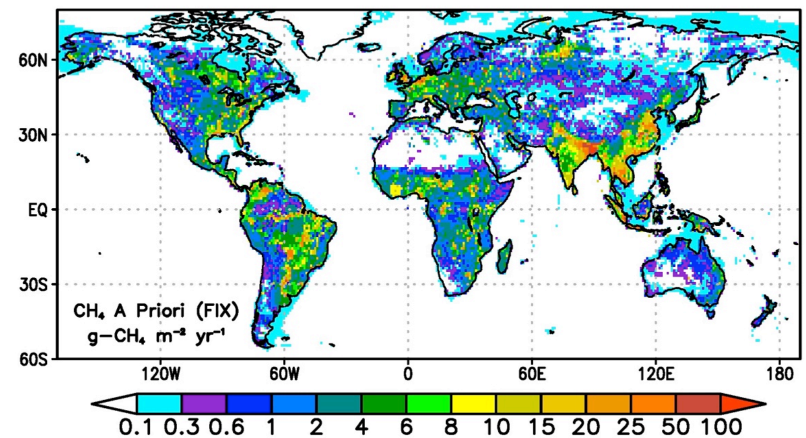
We need more and more data mining tools – data volume beyond human searching of signals

Annual mean, seasonal cycle and anomalies for climate analysis

53-Regions (land only) Inverse Model for CH_4





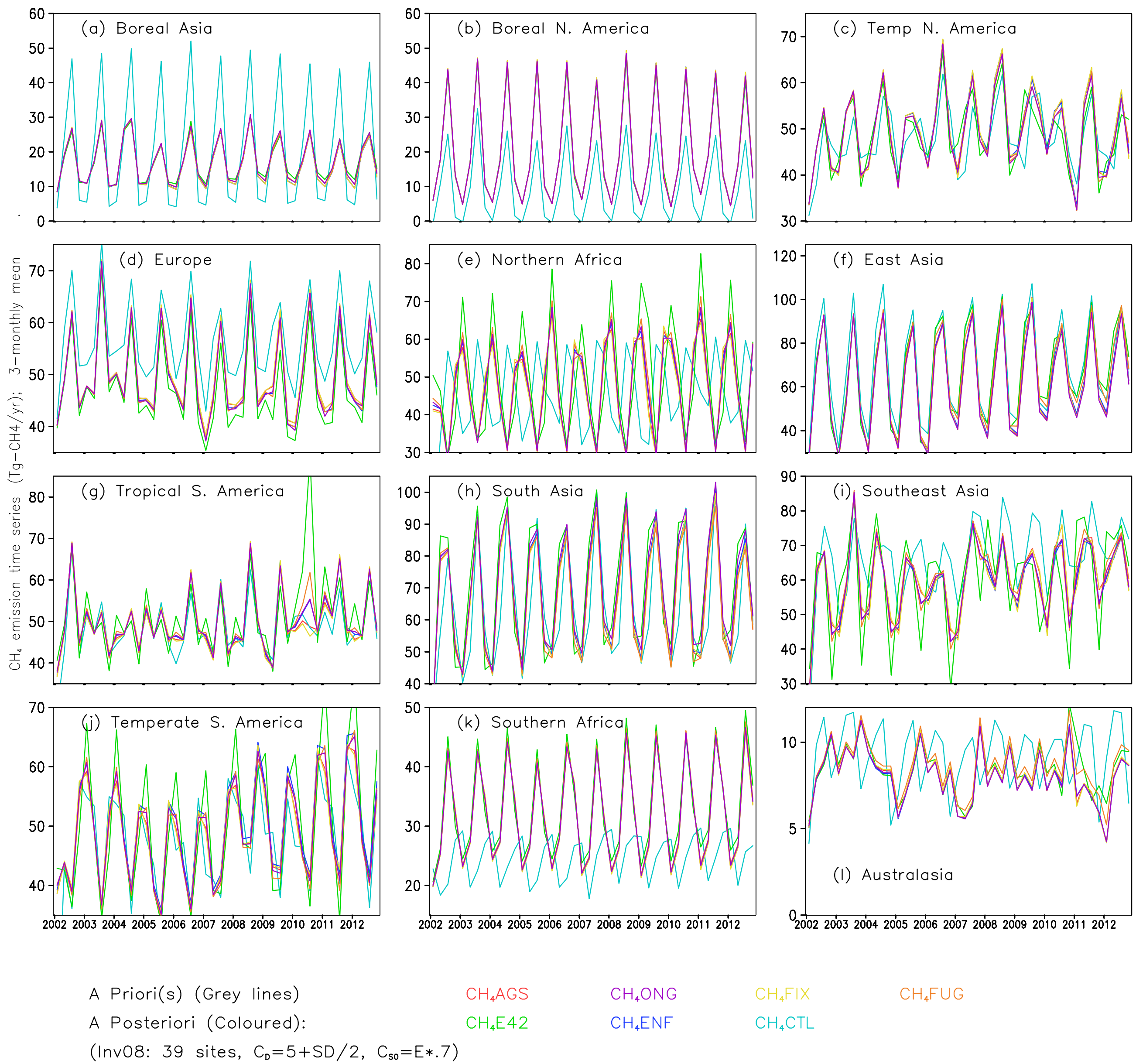


$$C_S = (G^T C_D^{-1} G + C_{S_0}^{-1})^{-1}$$

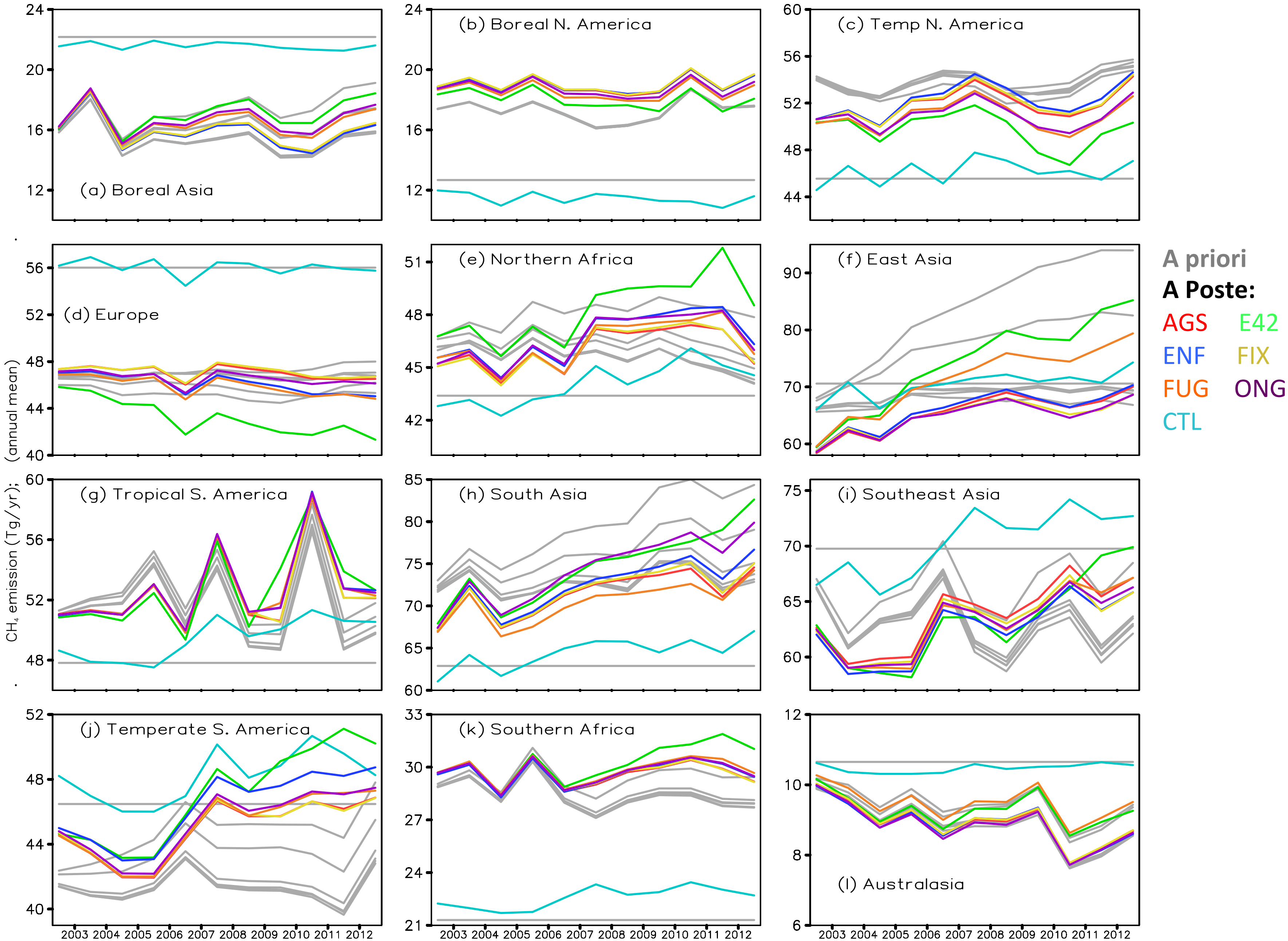
$$S = S_0 + \left(G^T C_D^{-1} G + C_{S_0}^{-1}\right)^{-1} G^T C_D^{-1} (D - D_{ACTM})$$

S_0 = regional prior sources
C_{S0} = Prior source covariance = 50% of region-total emission for each month
D = atmospheric concentration data
Data covariance ***C_D*** = 10 ppb; 5 ppb for measurements + 5 ppb for model uncertainty
D_{ACTM} = ACTM simulation using S_0
G = Green's functions for regional source-receptor relationships

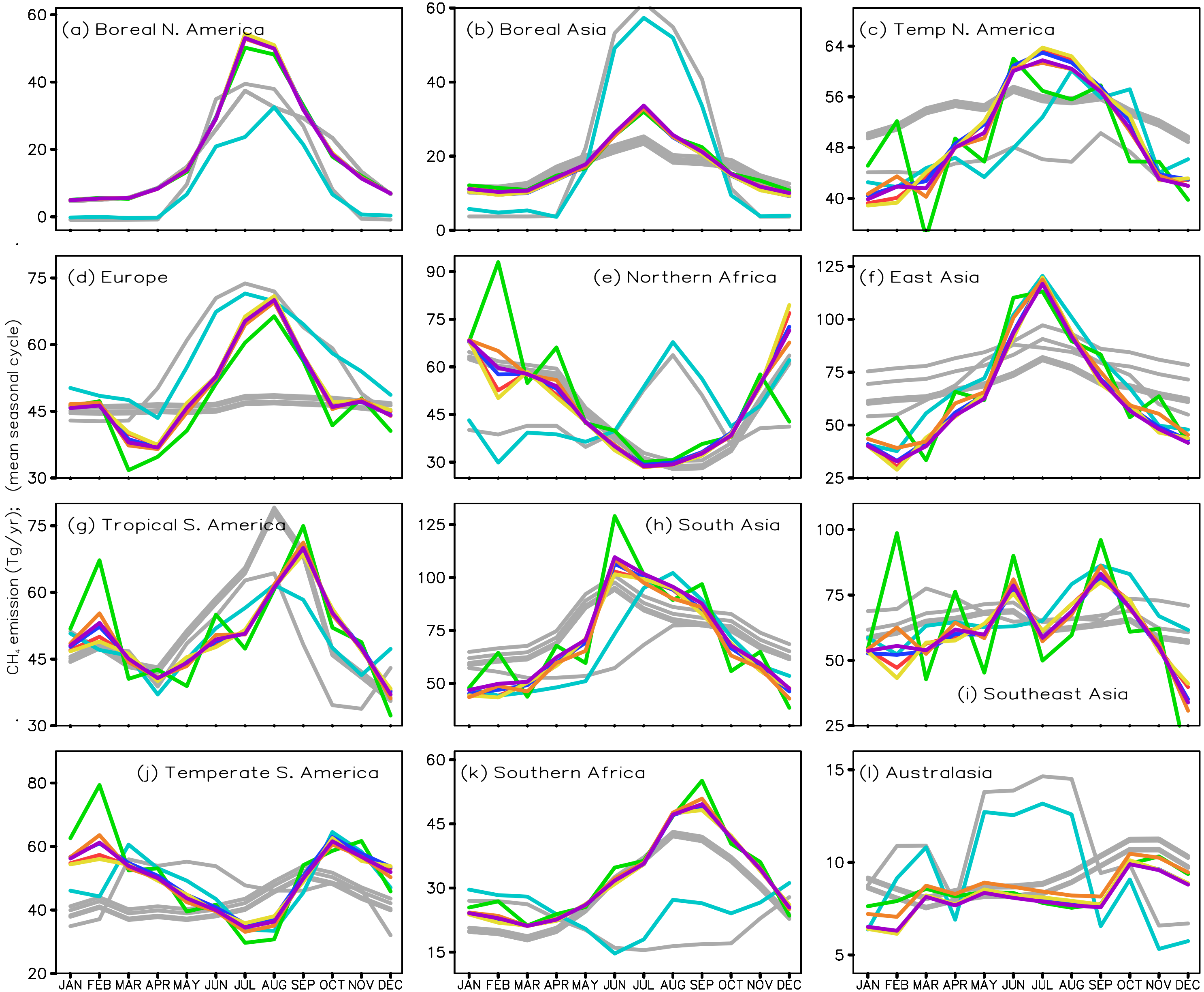
CH_4 fluxes: 3-monthly time series for 2002-2012



CH_4 fluxes : annual means for the period 2002-2012



CH_4 fluxes: Mean (2002-2012) Seasonal Cycle

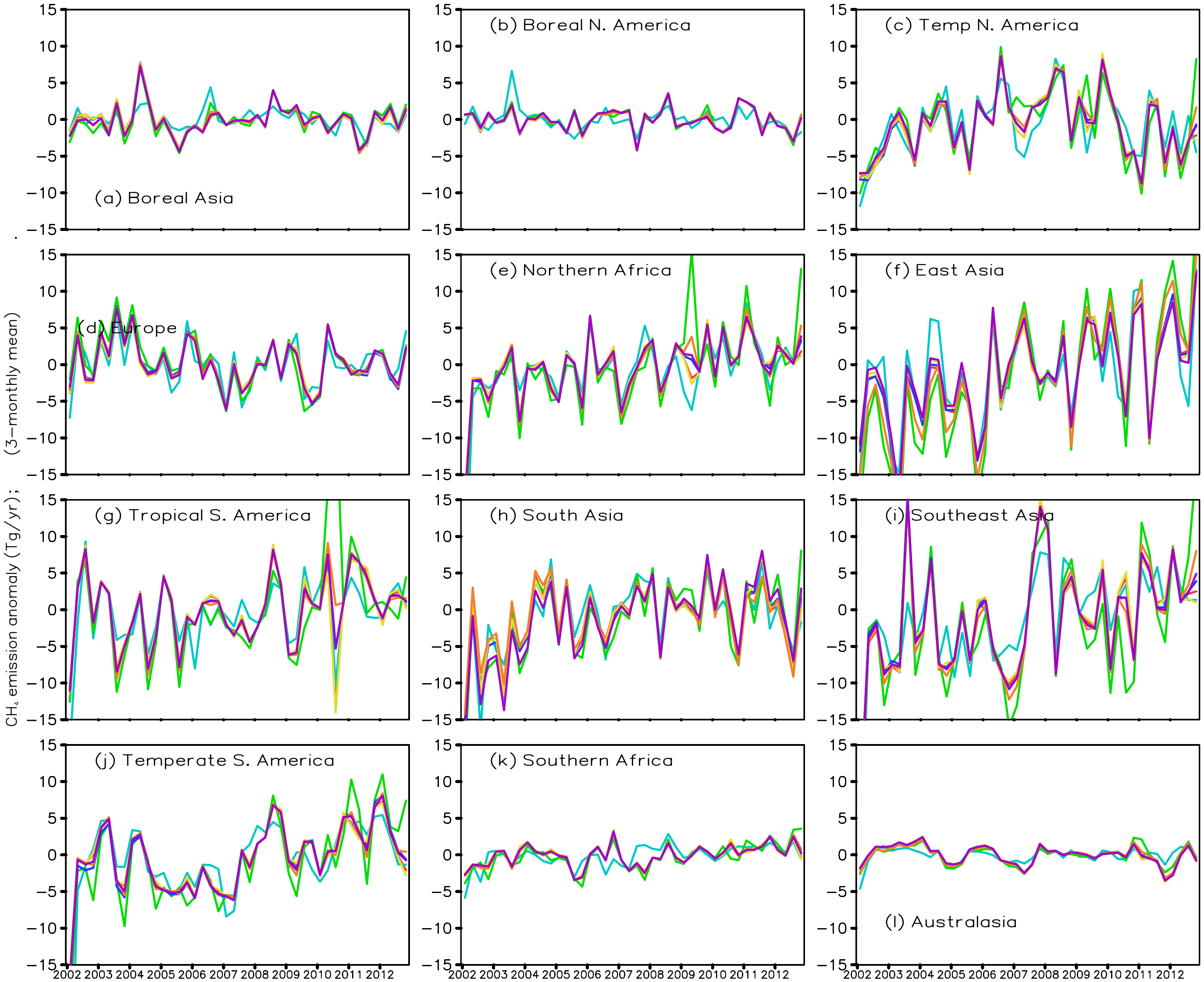


Results do not match!

Is it a good or bad thing?

Opportunity vs obviousness

CH_4 fluxes: 3-monthly anomalies (2002-2012)



Can we now conclude that we understand / estimate the CH_4 flux anomalies better then their annual means at regional scale

Outlook

- Be open minded – as Gabi said the credit for discovery of “ozone hole” probably went to wrong person/group
-

Survey

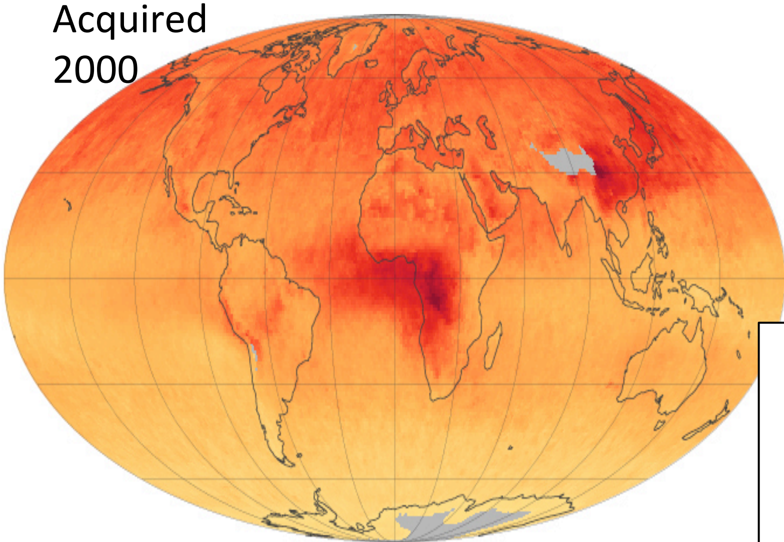
- Research interests (2-3 key words, e.g., aerosol, gases, dynamics)
- Method of research (model, satellite or in situ measurements)
- Research tools (Fortran, C, IDL, python, R, matlab)
- Years in research field
- Social media participation (Facebook, WhatsApp)
- Academic goal (optional)

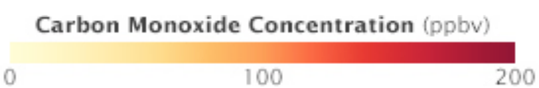
CO_2, CH_4, CO growth rates

Make plots like CH_4 by Saeki-san

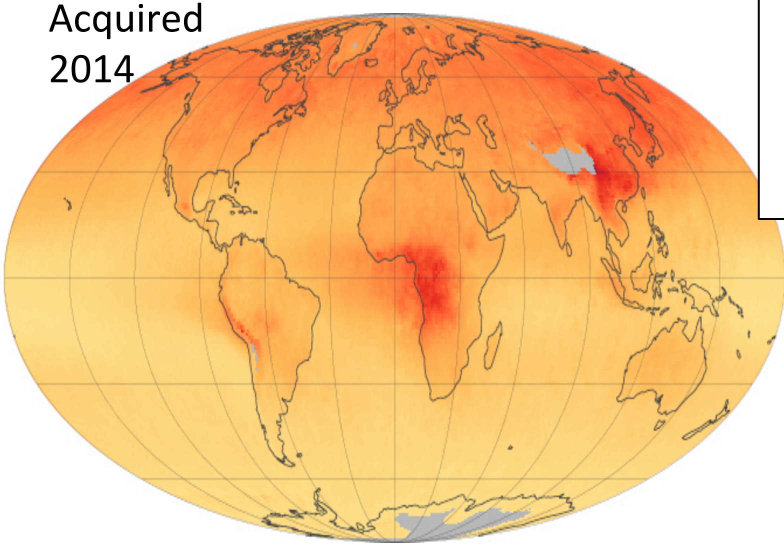
CO from MOPITT

Acquired 2000





Acquired 2014



http://earthobservatory.nasa.gov/IOTD/view.php?id=85967&src=eoa-iotd

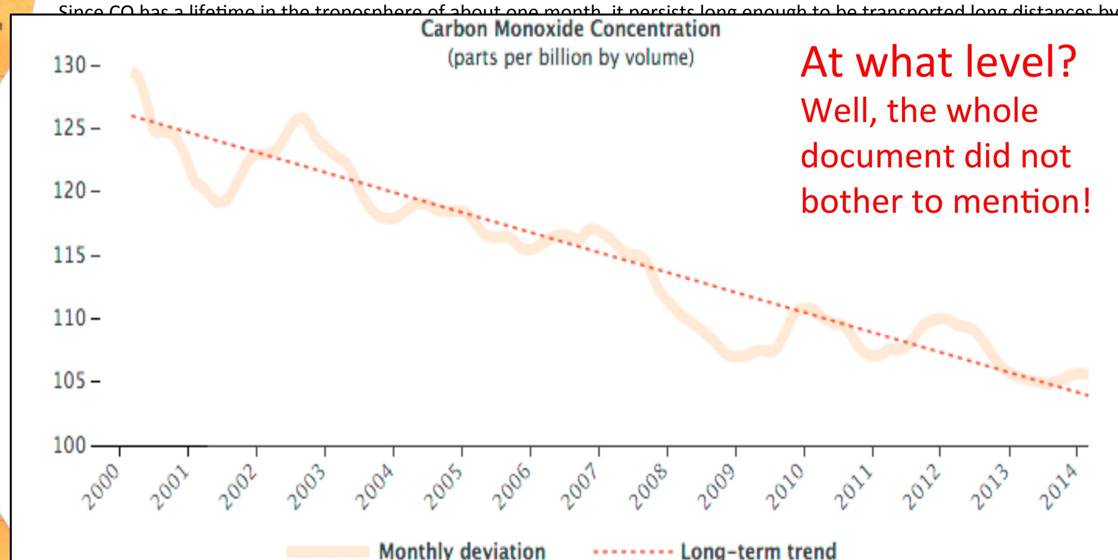
Carbon monoxide is perhaps best known for the lethal effects it can have in homes with faulty appliances and poor ventilation. In the United States, the colorless, odorless gas kills about 430 people each year.

However, the importance of carbon monoxide (CO) extends well beyond the indoor environment. Indoors or outdoors, the gas can disrupt the transport of oxygen by the blood, leading to heart and health problems. CO also contributes to the formation of tropospheric ozone, another air pollutant with unhealthy effects. And though carbon monoxide does not cause climate change directly, its presence affects the abundance of greenhouse gases such as methane and carbon dioxide.

Carbon monoxide forms whenever carbon-based fuels—including coal, oil, natural gas, and wood—are burned. As a result, many human activities and inventions emit carbon monoxide, including: the combustion engines in cars, trucks, planes, ships, and other vehicles; the fires lit by farmers to clear forests or fields; and industrial processes that involve the combustion of fossil fuels. In addition, wildfires and volcanoes are natural sources of the gas.

Little was known about the global distribution of carbon monoxide until the launch of the Terra satellite in 1999. Terra carries a sensor—Measurements of Pollution in the Troposphere (MOPITT)—that can measure carbon monoxide in a consistent fashion on a global scale. With a swath width of 640 kilometers (400 miles), MOPITT scans the entire atmosphere of Earth every three days.

Since CO has a lifetime in the troposphere of about one month, it persists long enough to be transported long distances by

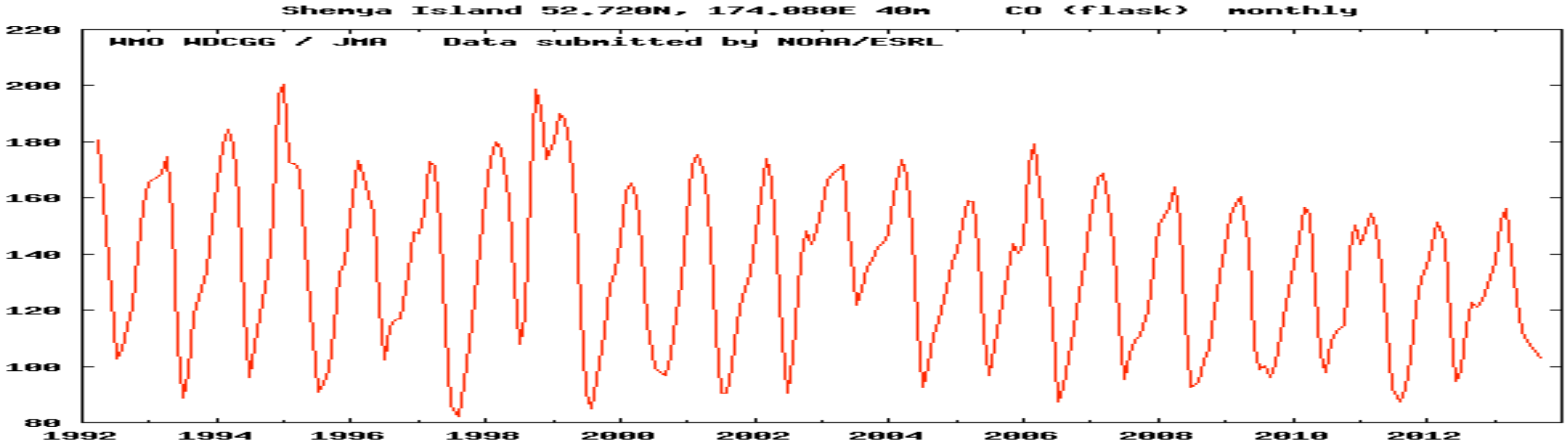


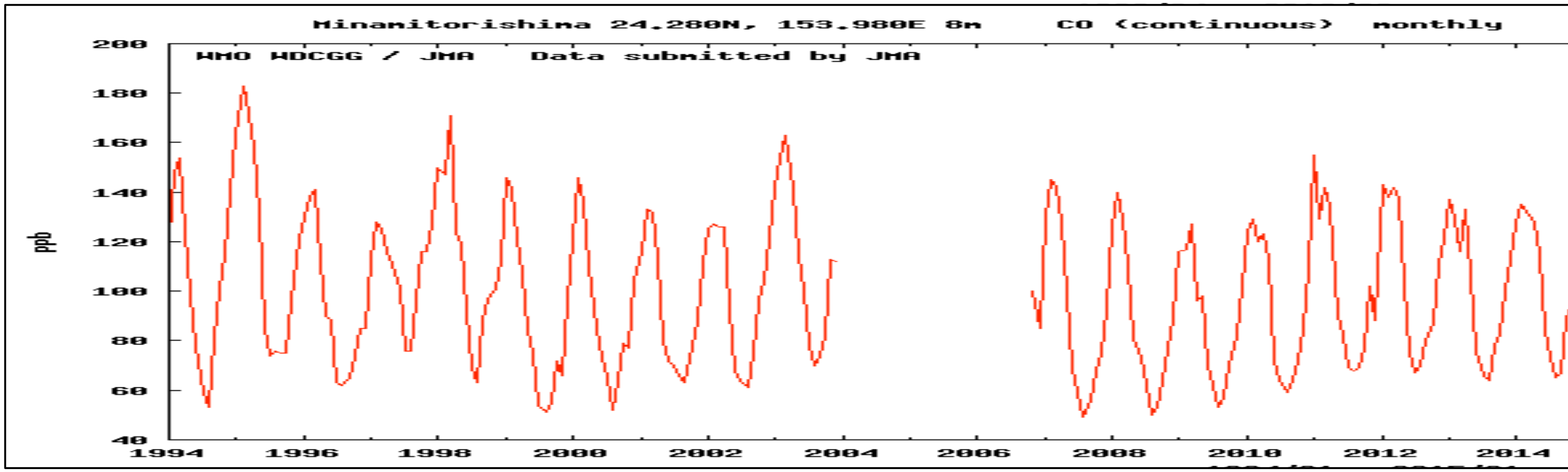
carbon monoxide over China and India, satellites and emissions inventories have shown that other pollutants like sulfur dioxide and nitrogen dioxide have risen during the same period.

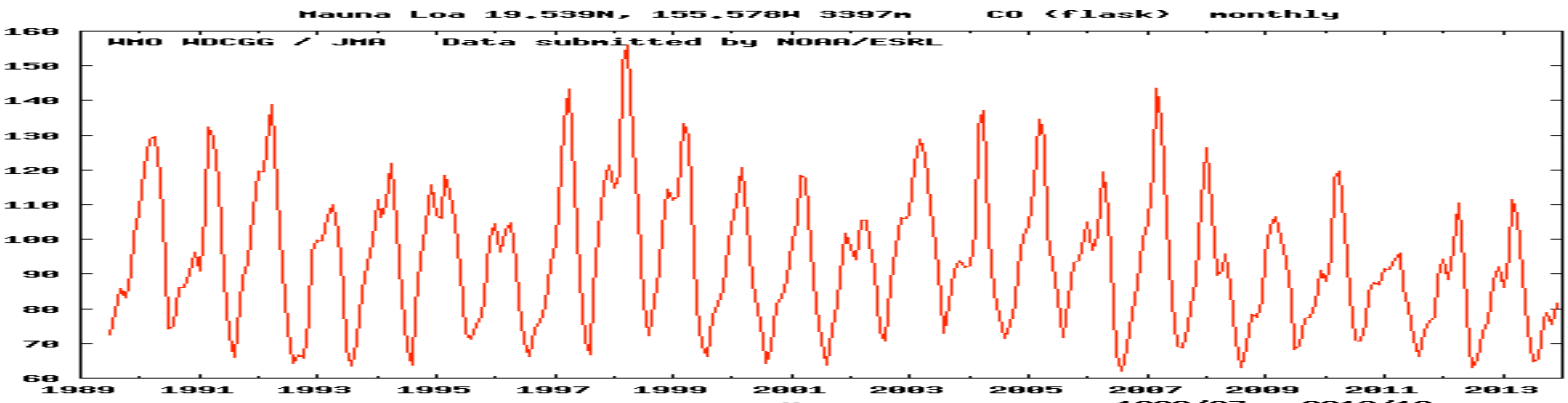
"For China, nitrogen dioxide emissions are mostly from the power and transportation sectors and have grown significantly since 2000 with the increase in demand for electricity," explained Helen Worden, an atmospheric scientist from the National Center for Atmospheric Research. "Carbon monoxide emissions, however, have a relatively small contribution (less than 2 percent) from the power sector, so vehicle emissions standards and improved combustion efficiency for newer cars have lowered carbon monoxide in the atmosphere despite the fact that there are more vehicles on the road burning more fossil fuel."

As illustrated by the maps, the news is also generally positive for the Southern Hemisphere, where deforestation and agricultural fires are the primary source of carbon monoxide. In South America, MOPITT observed a slight decrease in carbon monoxide; other satellites have observed decreases in the number of small fires and areas burned, suggesting a decrease in deforestation fires since 2005. Likewise, MOPITT has observed decreases in the amount of carbon monoxide

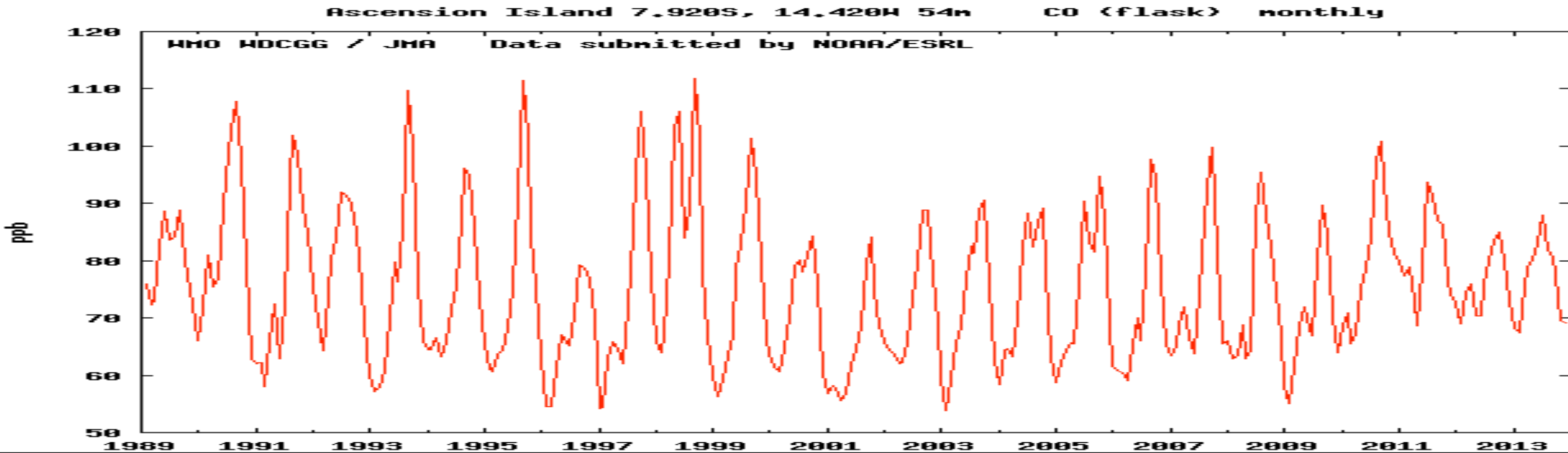
CO from Surface sites (Northern Hemisphere)

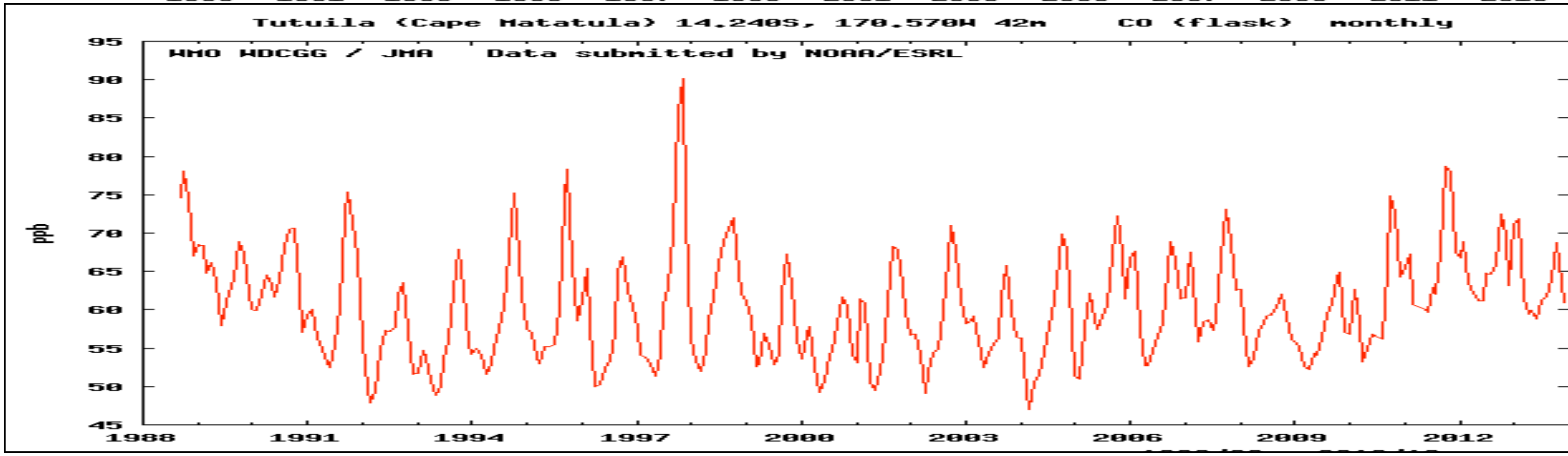


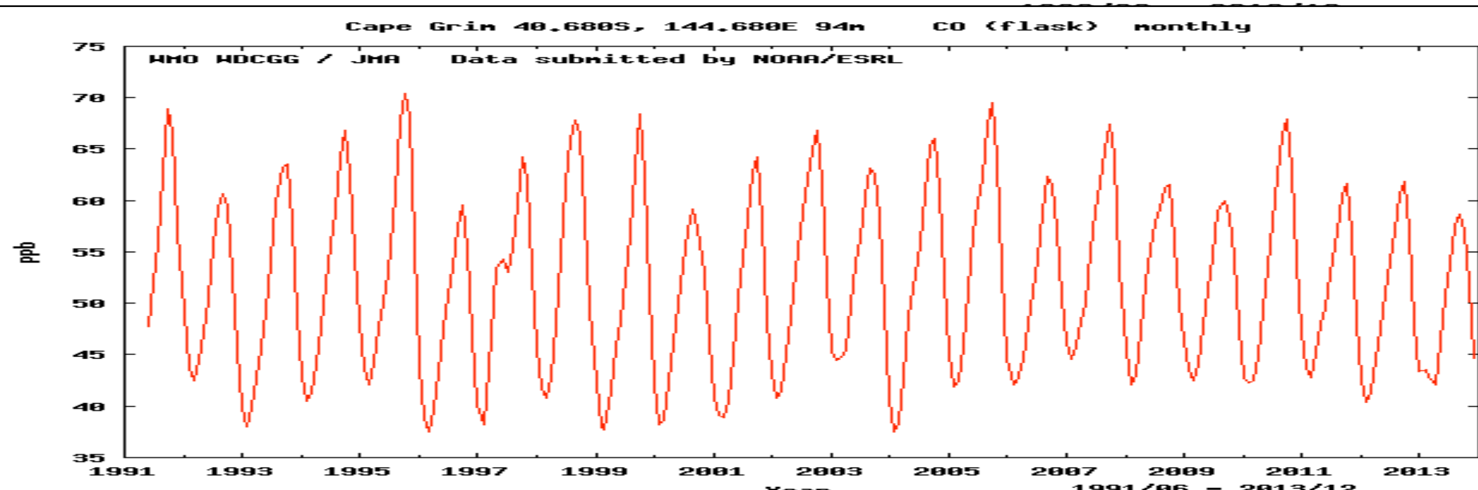




CO from Surface sites
(Southern Hemisphere)







Decadal trends in global CO emissions as seen by MOPITT

Y. Yin[1], F. Chevallier[1], P. Ciais[1], G. Broquet[1], A. Fortems-Cheiney[2], I. Pison[1], and M. Saunois[1]
[1]Laboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-UVSQ, UMR8212, Gif-sur-Yvette, France
[2]Institut National de l'Environnement Industriel et des RisqueS (INERIS), Verneuil-en-Halatte, France

Review Status
This discussion paper is under review for the journal Atmospheric Chemistry and Physics (ACP).

Received: 20 April 2015 – Accepted: 04 May 2015 – Published: 22 May 2015

Abstract. Negative trends of carbon monoxide (CO) concentrations are observed in the recent decade by both surface measurements and satellite retrievals over many regions, but they are not well explained by current emission inventories. Here, we attribute the observed CO concentration decline with an atmospheric inversion that simultaneously optimizes the two main CO sources (surface emissions and atmospheric hydrocarbon oxidations) and the main CO sink (atmospheric hydroxyl radical OH oxidation) by assimilating observations of CO and other chemically related tracers. Satellite CO column retrievals from Measurements of Pollution in the Troposphere (MOPITT), version 6, and surface in-situ measurements of methane and methyl-chloroform mole fractions are assimilated jointly for the period of 2002–2011. Compared to the prior simulation, the optimized CO concentrations show better agreement with independent surface in-situ measurements in terms of both distributions and trends. At the global scale, the atmospheric inversion primarily interprets the CO concentration decline as a decrease in the CO emissions, and finds noticeable trends neither in the chemical oxidation sources of CO, nor in the OH concentrations that regulate CO sinks. The latitudinal comparison of the model state with independent formaldehyde (CH_2O) columns retrieved from the Ozone Measurement Instrument (OMI) confirms the absence of large-scale trends in the atmospheric source of CO. The global CO emission decreased by 17% during the decade, more than twice the negative trend estimated by emission inventories. The spatial distribution of the inferred decrease of CO emissions indicates contributions from both a decrease in fossil- and bio-fuel emissions over Europe, the USA and Asia, and from a decrease in biomass burning emissions in South America, Indonesia, Australia and Boreal regions. An emission decrease of 2% yr^{-1} is inferred in China, one of the main emitting regions, in contradiction with the bottom-up inventories that report an increase of 2% yr^{-1} during the study period. A large decrease in CO emission factors due to technology improvements would outweigh the increase of carbon fuel combustions and may explain the observed decrease. In Africa, instead of the negative trend (1% yr^{-1}) reported by CO emission inventories mainly contributed by biomass burning, a positive trend (1.5% yr^{-1}) is found by the atmospheric inversion, suggesting different trends between satellite-detected burned areas and CO emissions.

Citation: Yin, Y., Chevallier, F., Ciais, P., Broquet, G., Fortems-Cheiney, A., Pison, I., and Saun
in global CO emissions as seen by MOPITT, Atmos. Chem. Phys. Discuss., 15, 14505-14547, do
15-14505-2015, 2015.