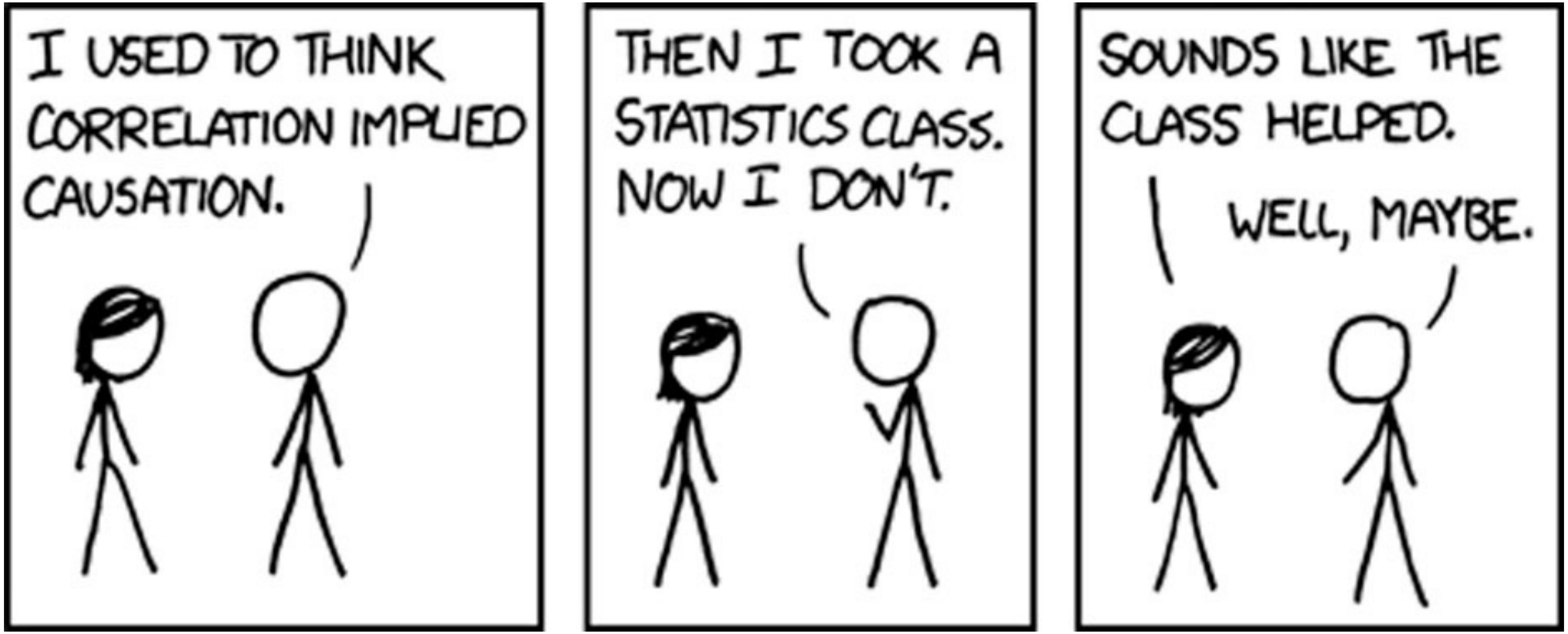
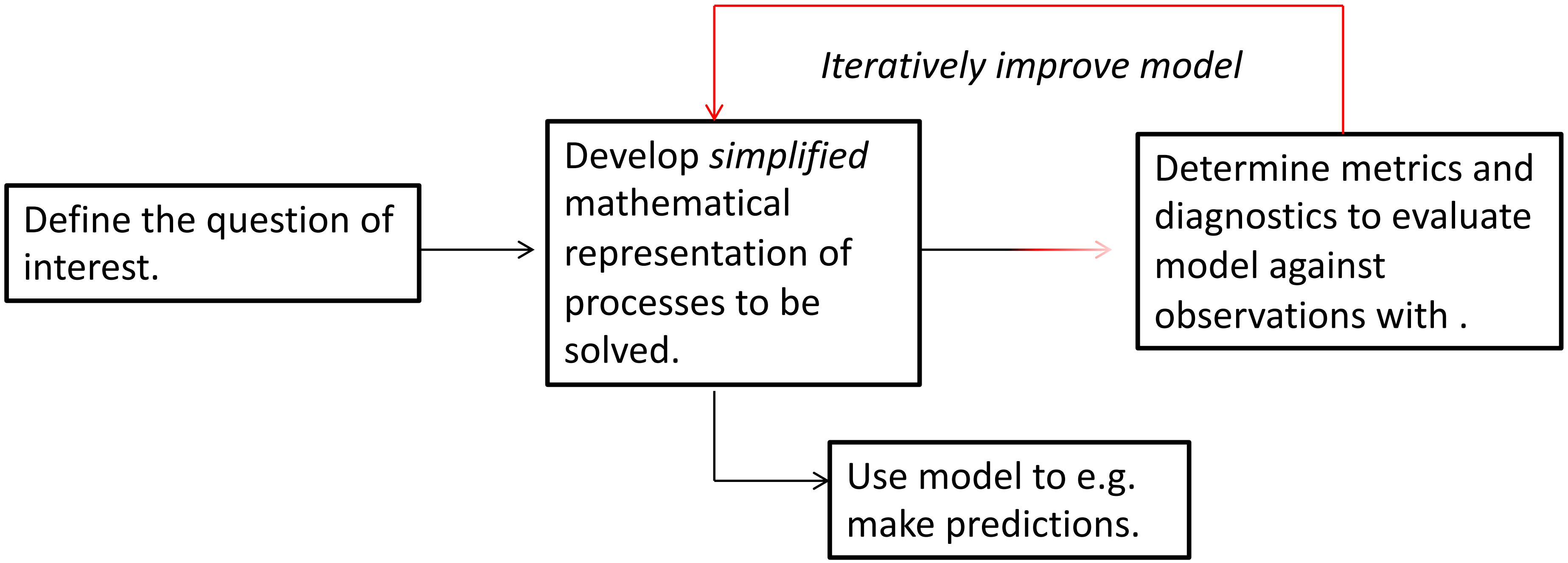
Model Evaluation



The challenge: *We want to understand how and why the atmosphere works.*

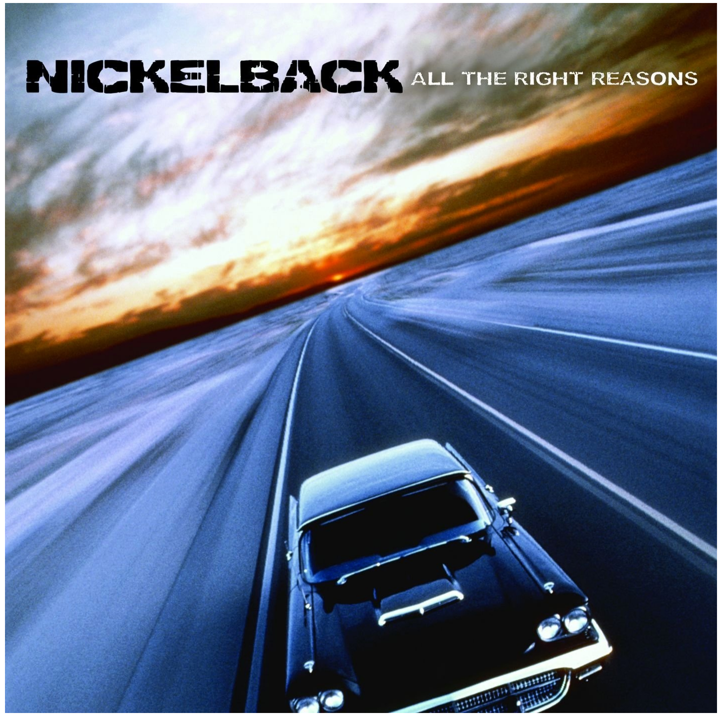
The problem: *The atmosphere is hideously complex.*

The solutions: *We can observe it in its natural state (field observations), we can test behavior under controlled situations (laboratory studies) or we can develop mathematical representations and* ***model*** *it.*


Iteratively improve model
Define the question of interest.
Develop *simplified* mathematical representation of processes to be solved.
Determine metrics and diagnostics to evaluate model against observations with .
Use model to e.g. make predictions.

But how do we know if our model is right for the right reasons?

If you Google this question this is, apparently, the answer:



But how do we know if our model is right for the right reasons?

We evaluate our model against other models (model intercomparisons or beauty contests), reanalyses and observations.

We may want to evaluate lots of aspects of our model simulation, but generally we will look at model bias and correlation as two key measures (metrics).

Increasingly, we must also look not only at the model predictions but also dig into the processes (process-based model evaluation).

Model evaluation can mean many things.

Lets define what we mean by model evaluation to be multi component. Model evaluation includes:

Model calibration – where we identify how to refine parameters/inputs into our model through comparison of model output with observations/model data.
This can be manual (i.e. one at a time “tuning”) or automated (i.e. using stochastic procedures)

Model evaluation can mean many things.

Lets define what we mean by model evaluation to be multi component. Model evaluation includes:

Model verification – where we quantify the predictive capability of our model. Again we compare the model and observations but this is different to calibration as we will not be using the results of these comparisons to modify the model logic/parameters.

For simple models (and for code) verification may include checking the logic of the model. This is increasingly difficult for the complex models we use like UKCA.

Model evaluation can mean many things.

Lets define what we mean by model evaluation to be multi component. Model evaluation includes:

Model verification – where we quantify the predictive capability of our model. Again we compare the model and observations but this is different to calibration as we will not be using the results of these comparisons to modify the model logic/parameters.

It is vital that the observational data used in model verification is distinct from the data used in calibration. NB this is not always the case or even possible.

Model evaluation can mean many things.

Lets define what we mean by model evaluation to be multi component. Model evaluation includes:

Model validation – all models are wrong, some models are useful. Not to get too bogged down by philosophical argument but from a technical perspective, a valid model is one in which the scientific or conceptual output is acceptable for its purpose.

For those wanting to think more meta: Can you ever validate a model?

Model evaluation can mean many things.

Lets define what we mean by model evaluation to be multi component. Model evaluation includes:

Sensitivity analysis – where the response of the model to changes in inputs/parameters is quantified. This understanding is important for:

1) **The range of suitability of the model**
2) **Identifying “key” parameters/inputs**
3) **Understanding behavior at critical points**

We will touch on perturbed parameter ensembles (PPEs – a type of sensitivity analysis) later.

Model evaluation can mean many things.

Lets define what we mean by model evaluation to be multi component. Model evaluation includes:
Model calibration
Model verification
Model validataion
Sensitivity analysis

And it requires some objective measures of “goodness of fit”

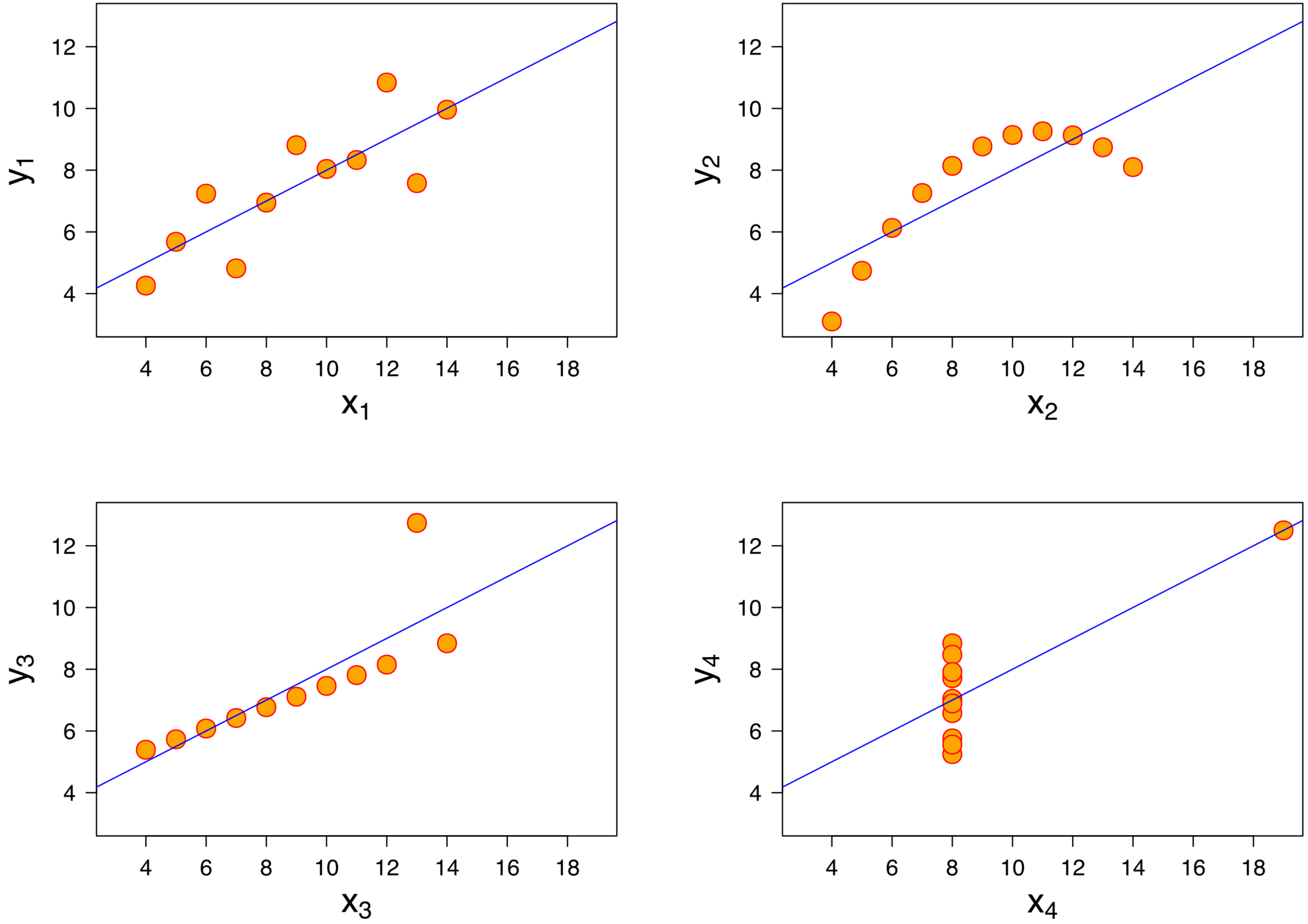
How can I tell if my model is good or bad?

First, don’t forget to focus on what you are comparing! Integral quantities? Hourly/high time frequency data? Other model data? What are the biases in the observational data? How are the characterized?

There are many, many, many, statistical measures that we can use and software like R and Python make it easy to abuse them.

How can I tell if my model is good or bad?

Nothing beats a visual inspection in my opinion.



Anscombe's quartet: all have the same mean, variance and correlation coefficient.

Air Quality Model Performance Metric Definitions

Common Variables:

M = predicted concentration
O = observed concentration
X = predicted or observed concentration
σ = standard deviation

I. Mean Bias, Mean Error, and Root Mean Square Error (ppb)

Mean Bias =

$$\frac{1}{n}\sum_{1}^{n}(M - O)$$

Mean Error =

$$\frac{1}{n}\sum_{1}^{n}\left|M - O\right|$$

Root Mean Square Error =

$$\sqrt{\frac{\sum_{1}^{n}(M - O)^{2}}{n}}$$

Air Quality Model Performance Metric Definitions

Common Variables:

M = predicted concentration
O = observed concentration
X = predicted or observed concentration
σ = standard deviation

II. Normalized Mean Bias and Error (unitless)

Normalized Mean Bias =

$$\frac{\sum_{1}^{n}(M-O)}{\sum_{1}^{n}(O)}$$

Normalized Mean Error = $$\frac{\sum_{1}^{n}\left|M-O\right|}{\sum_{1}^{n}(O)}$$

Air Quality Model Performance Metric Definitions

Common Variables:

M = predicted concentration
O = observed concentration
X = predicted or observed concentration
σ = standard deviation

III. Fractional Bias and Error (unitless)

Fractional Bias =

$$\frac{1}{n}\left(\frac{\sum_{1}^{n}(M-O)}{\sum_{1}^{n}\left(\frac{(M+O)}{2}\right)}\right)$$

Fractional Error = $$\frac{1}{n}\left(\frac{\sum_{1}^{n}\left|M-O\right|}{\sum_{1}^{n}\left(\frac{(M+O)}{2}\right)}\right)$$

Air Quality Model Performance Metric Definitions

Common Variables:

M = predicted concentration
O = observed concentration
X = predicted or observed concentration
σ = standard deviation

IV. Correlation Coefficient (unitless)
Correlation =

$$\frac{1}{(n-1)}\sum_{1}^{n}\left(\left(\frac{O-\overline{O}}{\sigma_o}\right)*\left(\frac{M-\overline{M}}{\sigma_m}\right)\right)$$

VI. Coefficient of Variation (unitless)
Coefficient of Variation =

$$\frac{\sigma}{\overline{X}}$$

Air Quality Model Performance Metric Definitions

Common Variables:

M = predicted concentration
O = observed concentration
X = predicted or observed concentration
σ = standard deviation

IV. Correlation Coefficient (unitless)
Correlation =

$$\frac{1}{(n-1)}\sum_{1}^{n}\left(\left(\frac{O-\overline{O}}{\sigma_o}\right)*\left(\frac{M-\overline{M}}{\sigma_m}\right)\right)$$

V. Coefficient of Variation (unitless)
Coefficient of Variation =

$$\frac{\sigma}{\overline{X}}$$

VI. Index of Agreement (unitless)
Index of Agreement =

$$1-\left[\frac{\sum_{1}^{n}(O-M)^2}{\sum_{1}^{n}\left(\left|M-\overline{O}\right|+\left|O-\overline{O}\right|\right)^2}\right]$$

Error:

Mean Absolute Error

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} = \frac{\sum_{i=1}^{n} |e_i|}{n}.$$

is a straightforward measure of how far away our model simulation (y) was from our observations (x) on average. It takes the modulus of the absolute error (bias) and so is always positive.

Error:

Mean Squared Error

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2.$$

is measure of both the bias and the variance of the model. The variance is the expectation of the squared deviation of a random variable from its mean. It measures the spread from the average.

New approaches to evaluation:

The MSE is the squared difference of the modelled (mod) and observed (obs) values:

$$\mathrm{MSE} = E(\text{mod-obs})^2 = \frac{\sum_{i=1}^{nt}(\mathrm{mod}_i - \mathrm{obs}_i)^2}{n_t}, \quad (1)$$

where $E(\cdot)$ denotes expectation and n_t is the length of the time series. The bias is

$$\mathrm{bias} = E\,(\text{mod-obs}) \quad (2)$$

i.e. $\mathrm{bias} = \overline{\mathrm{mod}} - \overline{\mathrm{obs}}$. Thus, the following relationship holds:

$$\mathrm{MSE} = \mathrm{var}\,(\text{mod-obs}) + \mathrm{bias}^2, \quad (3)$$

which is a well-known property of the MSE, ($\mathrm{var}(\cdot)$ is the variance operator). By using the property of the variance for correlated fields:

$$\mathrm{var}\,(\text{mod-obs}) = \mathrm{var}\,(\mathrm{mod}) + \mathrm{var}\,(\mathrm{obs}) - 2\mathrm{cov}(\mathrm{mod,obs}), \quad (4)$$

the final formulation for the MSE components reads as follows:

$$\mathrm{MSE} = \mathrm{bias}^2 + \mathrm{var}\,(\mathrm{mod}) + \mathrm{var}(\mathrm{obs}) - 2\mathrm{cov}(\mathrm{mod,obs}), \quad (5)$$

where the covariance term (last term on the right-hand side of Eq. 5) accounts for the degree of correlation between the modelled and observed time series. When the covariance term is zero, var(obs) is referred to as the *incompressible part of the error* and represents the lowest limit that the MSE of the model can achieve. When dealing with model evaluation, the modelled and observed time series are typically highly correlated and therefore, within the limits of the perfect match (correlation coefficient of unity), $\mathrm{cov(mod,obs)} = \mathrm{cov(obs,obs)} = \mathrm{cov(mod,mod)} = \mathrm{var(mod)} = \mathrm{var(obs)}$ and the MSE can be reduced to only the bias term. That implies that the development of a high-quality model needs to ensure

a. the highest possible precision in order to maximise the cov(mod, obs) term;

b. the highest possible accuracy, in order to minimise the bias.

Elaborating on Eq. (5), Theil (1961) derived the following:

$$\mathrm{MSE} = (\overline{\mathrm{mod}} - \overline{\mathrm{obs}})^2 + (\sigma_{\mathrm{mod}} - \sigma_{\mathrm{obs}})^2 + 2(1-r)\sigma_{\mathrm{mod}}\sigma_{\mathrm{obs}}. \quad (6)$$

Solazzo et al **Atmos. Chem. Phys., 16, 6263–6283, 2016**

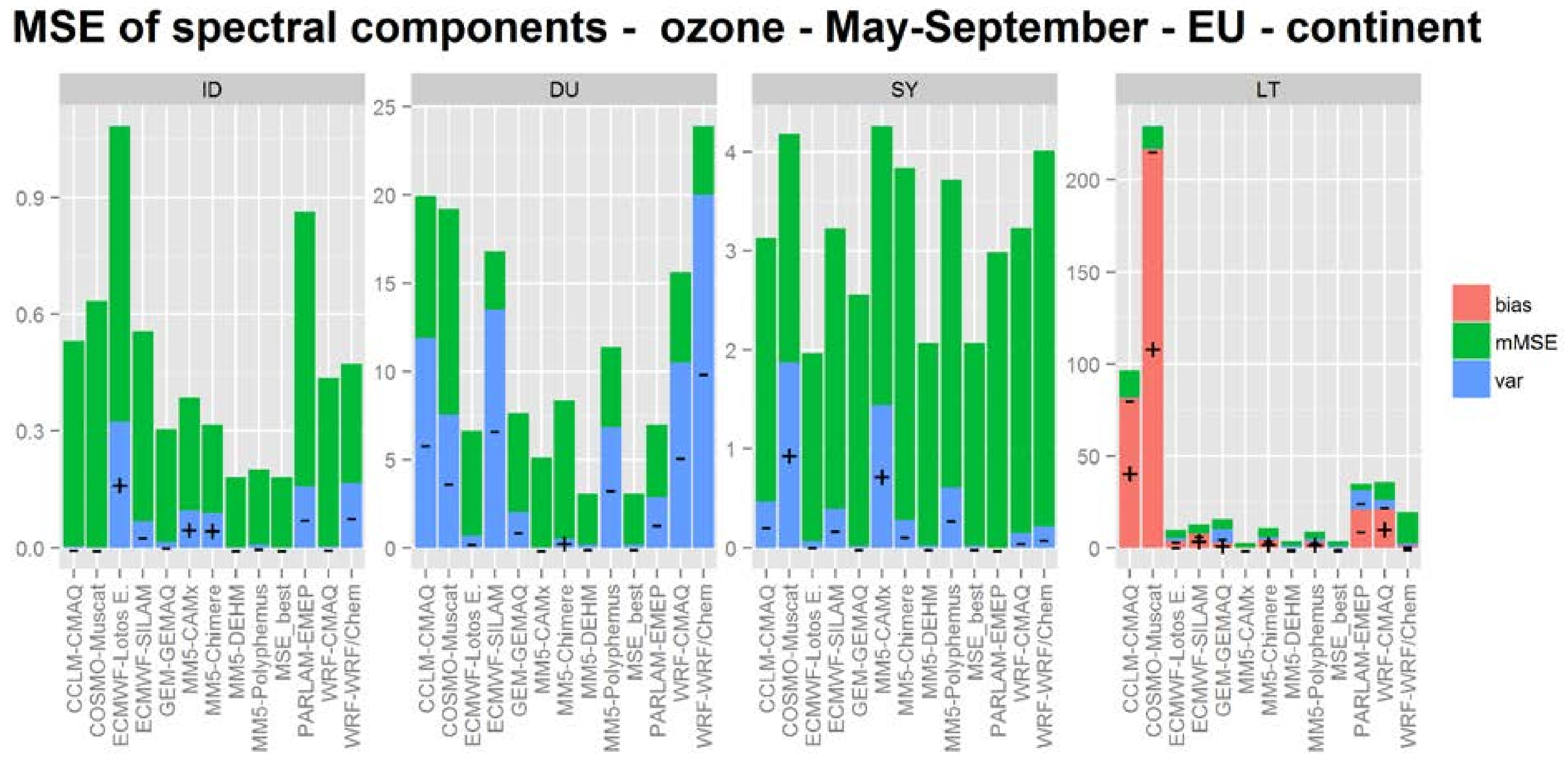
New approaches to evaluation:

mMSE is the minimum achievable Mean Square Error
mMSE = $\sigma^2_{obs}(1-r^2)$

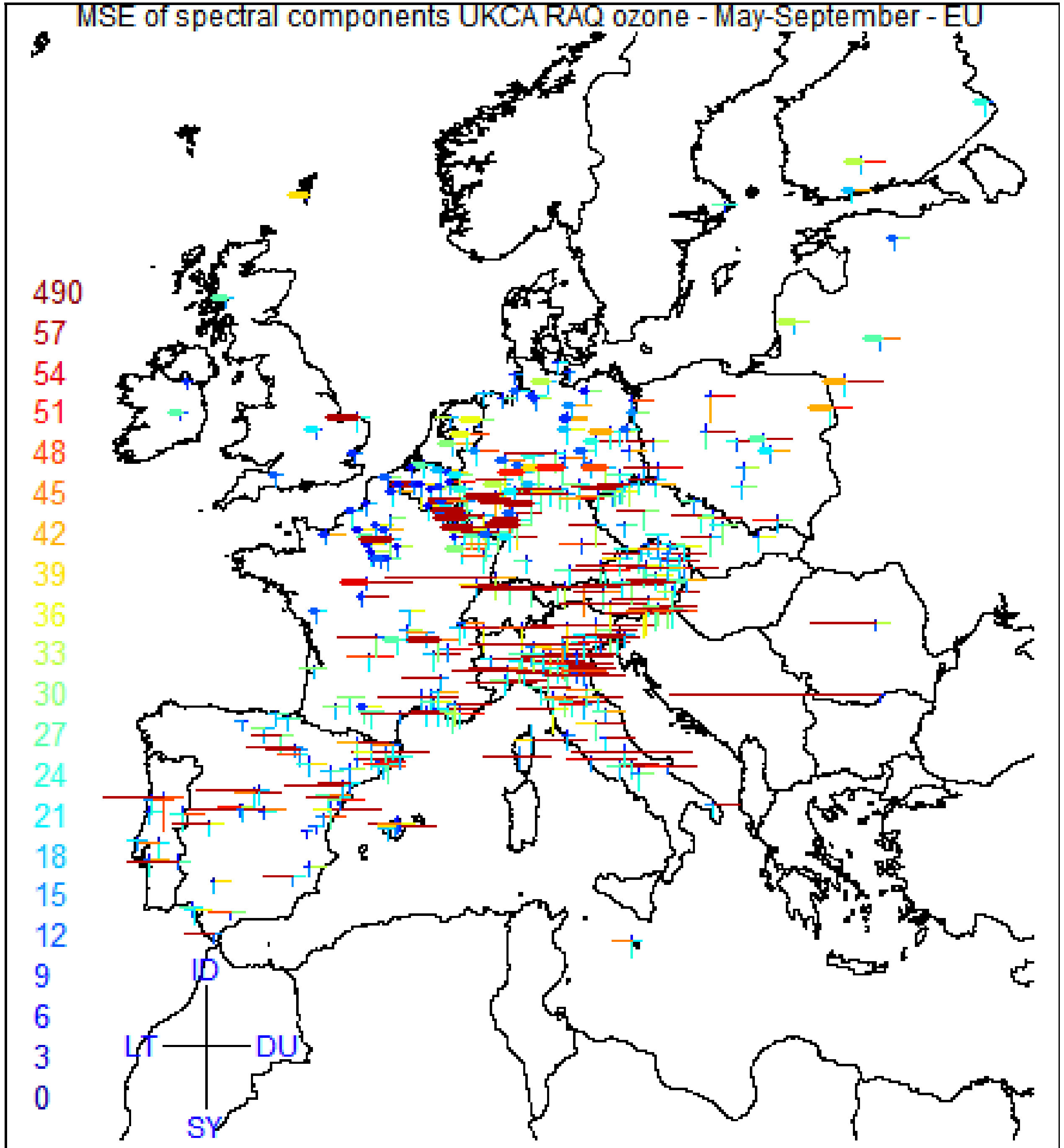
Solazzo and Galmarini suggest:
MSE = $(\langle mod \rangle - \langle obs \rangle)^2 + (\sigma_{mod} - r\sigma_{obs})^2 + mMSE$

As this metric allows for quantification of accuracy (bias), precision (variance) and associativity (unexplained portion through the correlation coefficient – r)

AQMEII1



MSE of spectral components - ozone - May-September - EU - continent
ID
DU
SY
LT
bias
mMSE
var
CCLM-CMAQ
COSMO-Muscat
ECMWF-Lotos E.
ECMWF-SILAM
GEM-GEMAQ
MM5-CAMx
MM5-Chimere
MM5-DEHM
MM5-Polyphemus
MSE_best
PARLAM-EMEP
WRF-CMAQ
WRF-WRF/Chem


MSE of spectral components UKCA RAQ ozone - May-September - EU
490
57
54
51
48
45
42
39
36
33
30
27
24
21
18
15
12
9
6
3
0
ID
LT
DU
SY
f)

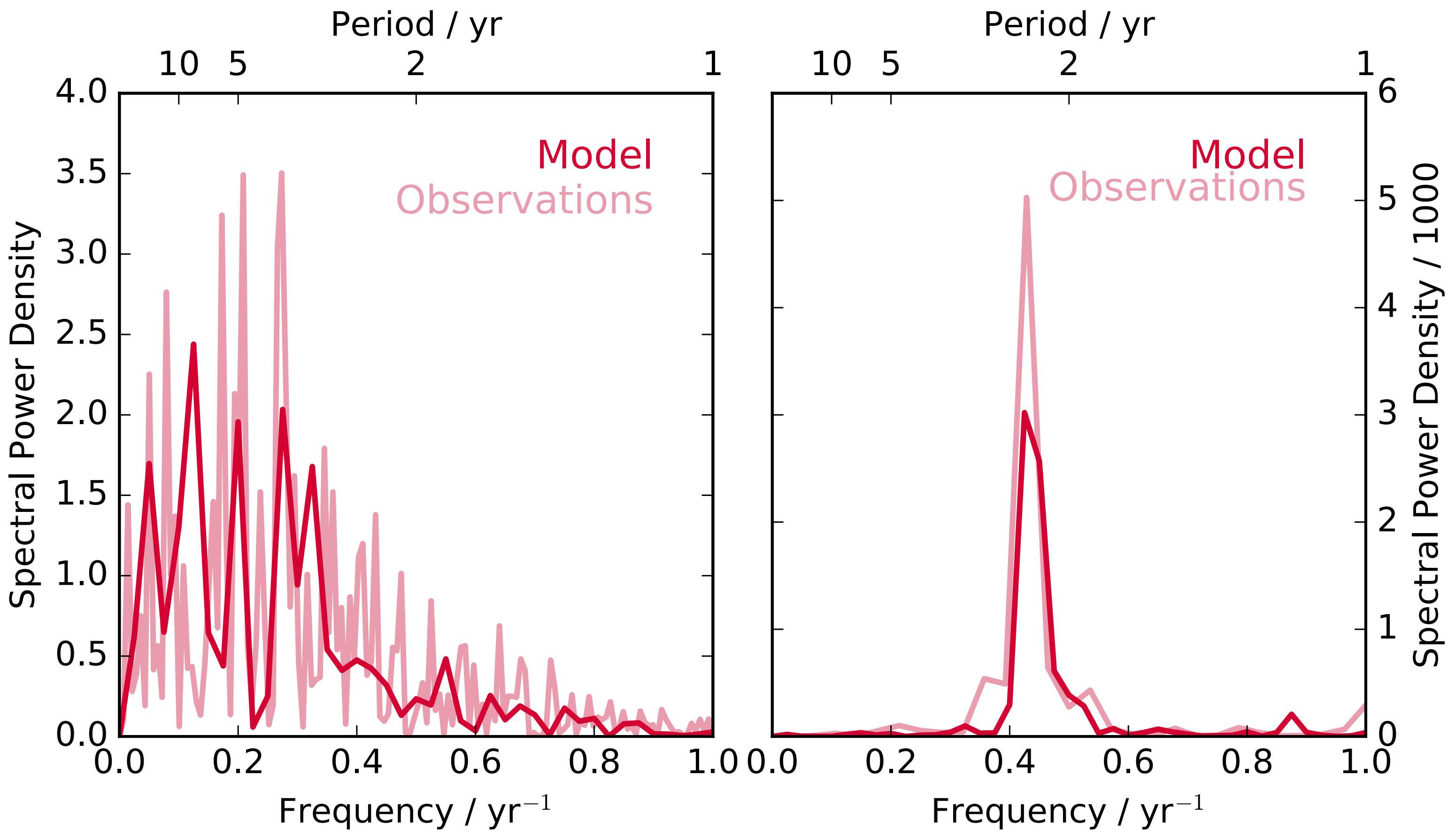
New approaches to evaluation:

When the analytical decomposition of the error into bias, variance and *mMSE* is applied to the decomposition of the signals into long-term, synoptic, inter-diurnal and diurnal components, information can be gathered that helps reduce the spectrum of possible sources of errors and pinpoint the processes that are most active at a particular scale which need to be improved. The procedure is denoted here as *error apportionment* and provides an improved and more powerful capacity to identify the nature of the error and associate it with a specific part of the spectrum of the model/measurement signal. The AQMEII set of models and measurements have been used in the evaluation procedure.

New approaches to evaluation:

Spectral decomposition of modelled and observed time series

Spectral decomposition is not new and is widely used in other fields of physical science but has been used less in evaluating composition.



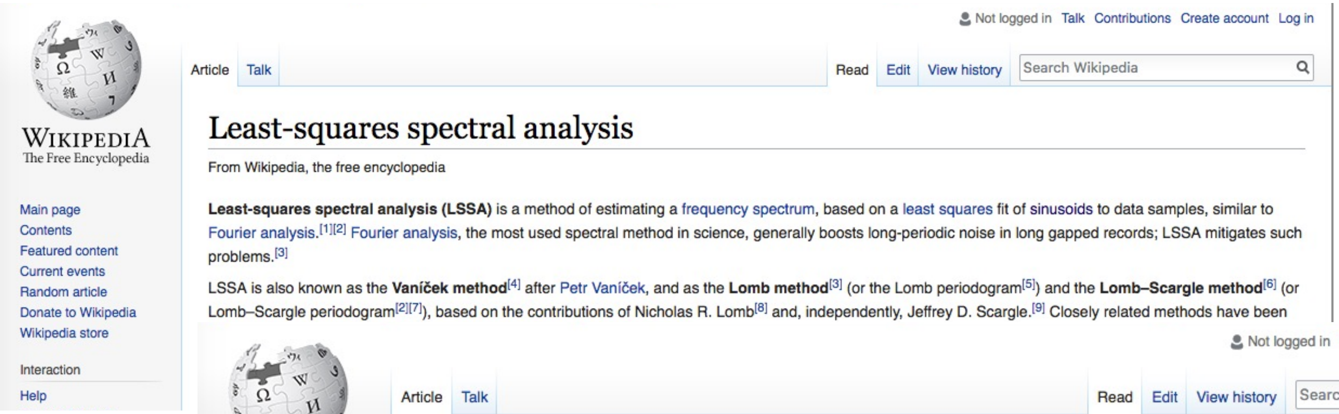
Courtesy of David Wade

New approaches to evaluation:

Spectral decomposition of modelled and observed time series

$$O_3 = LT(O_3) + SY(O_3) + DU(O_3) + ID(O_3)$$

Spectral decomposition is not new and is widely used in other fields of physical science but has been used less in evaluating composition.



Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help

Least-squares spectral analysis

From Wikipedia, the free encyclopedia

Least-squares spectral analysis (LSSA) is a method of estimating a frequency spectrum, based on a least squares fit of sinusoids to data samples, similar to Fourier analysis.[1][2] Fourier analysis, the most used spectral method in science, generally boosts long-periodic noise in long gapped records; LSSA mitigates such problems.[3]

LSSA is also known as the **Vaníček method**[4] after Petr Vaníček, and as the **Lomb method**[3] (or the Lomb periodogram[5]) and the **Lomb–Scargle method**[6] (or Lomb–Scargle periodogram[2][7]), based on the contributions of Nicholas R. Lomb[8] and, independently, Jeffrey D. Scargle.[9] Closely related methods have been

Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction

Kolmogorov–Zurbenko filter

From Wikipedia, the free encyclopedia

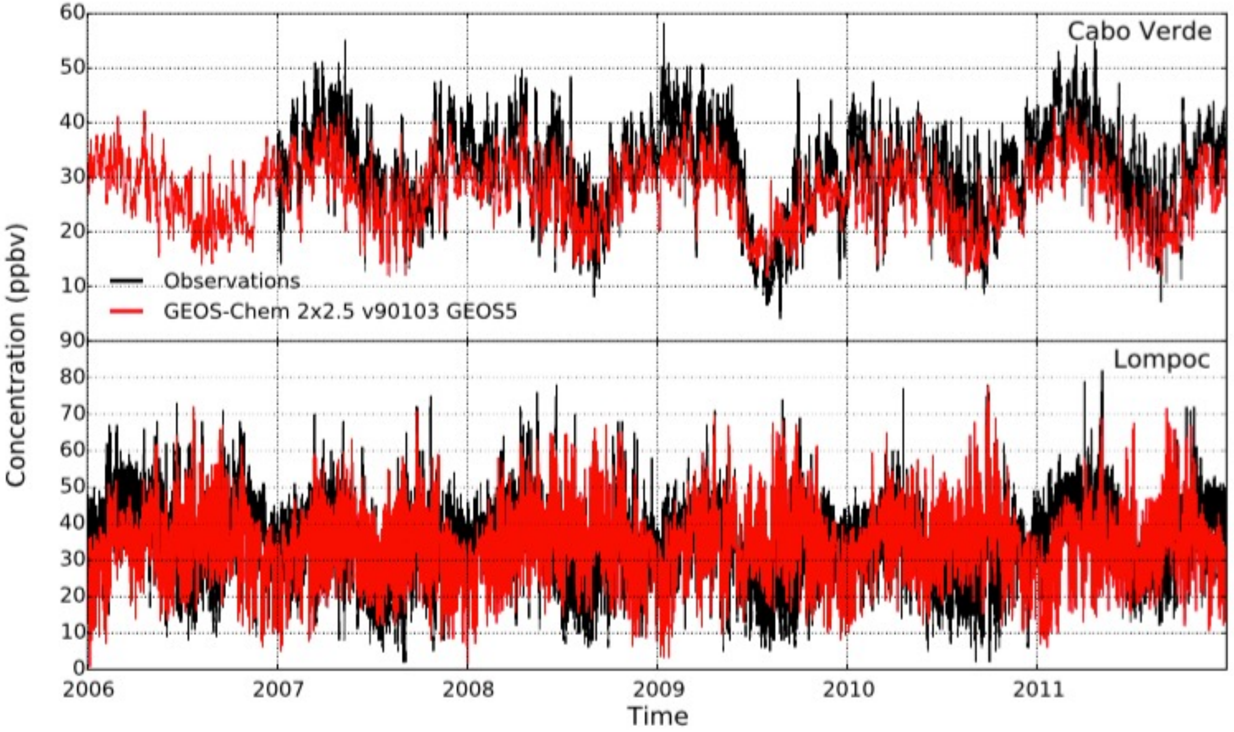
This article **provides insufficient context for those unfamiliar with the subject.** Please help improve the article with a good introductory style. *(January 2012) (Learn how and when to remove this template message)*

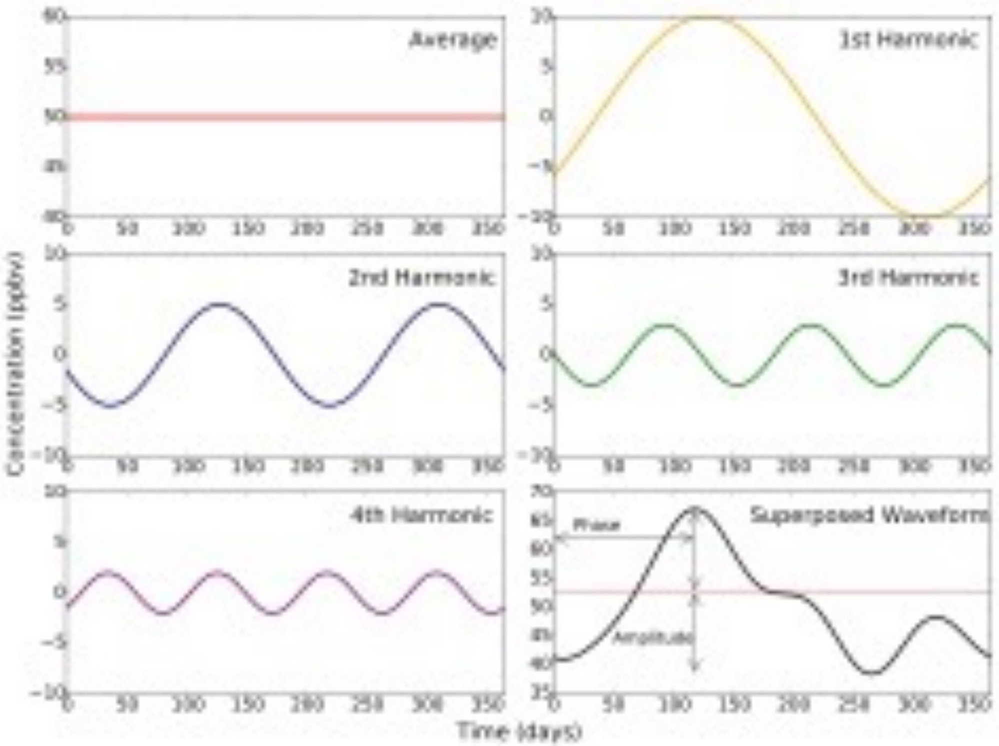
The **Kolmogorov–Zurbenko (KZ) filter** was first proposed by A. N. Kolmogorov and formally defined by Zurbenko.[1] It is a series of iterations of a moving average filter of length *m*, where *m* is a positive, odd integer. The KZ filter belongs to the class of low-pass filters. The KZ filter has two parameters, the length *m* of the moving average window and the number of iterations *k* of the moving average itself. It also can be considered as a special window function designed to eliminate spectral leakage.

New approaches to evaluation:

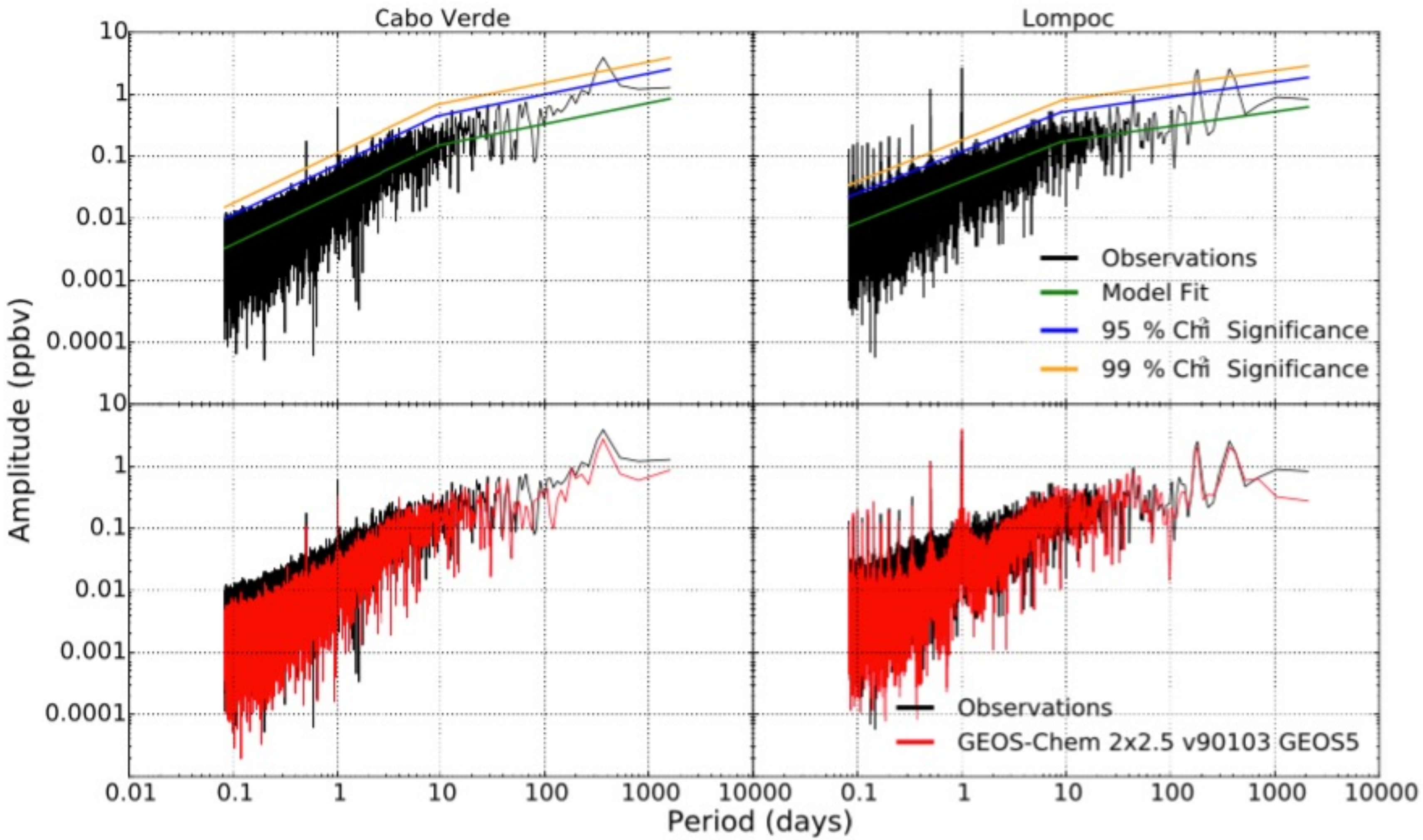
Spectral decomposition of modelled and observed time series

$$O_3 = LT(O_3) + SY(O_3) + DU(O_3) + ID(O_3)$$





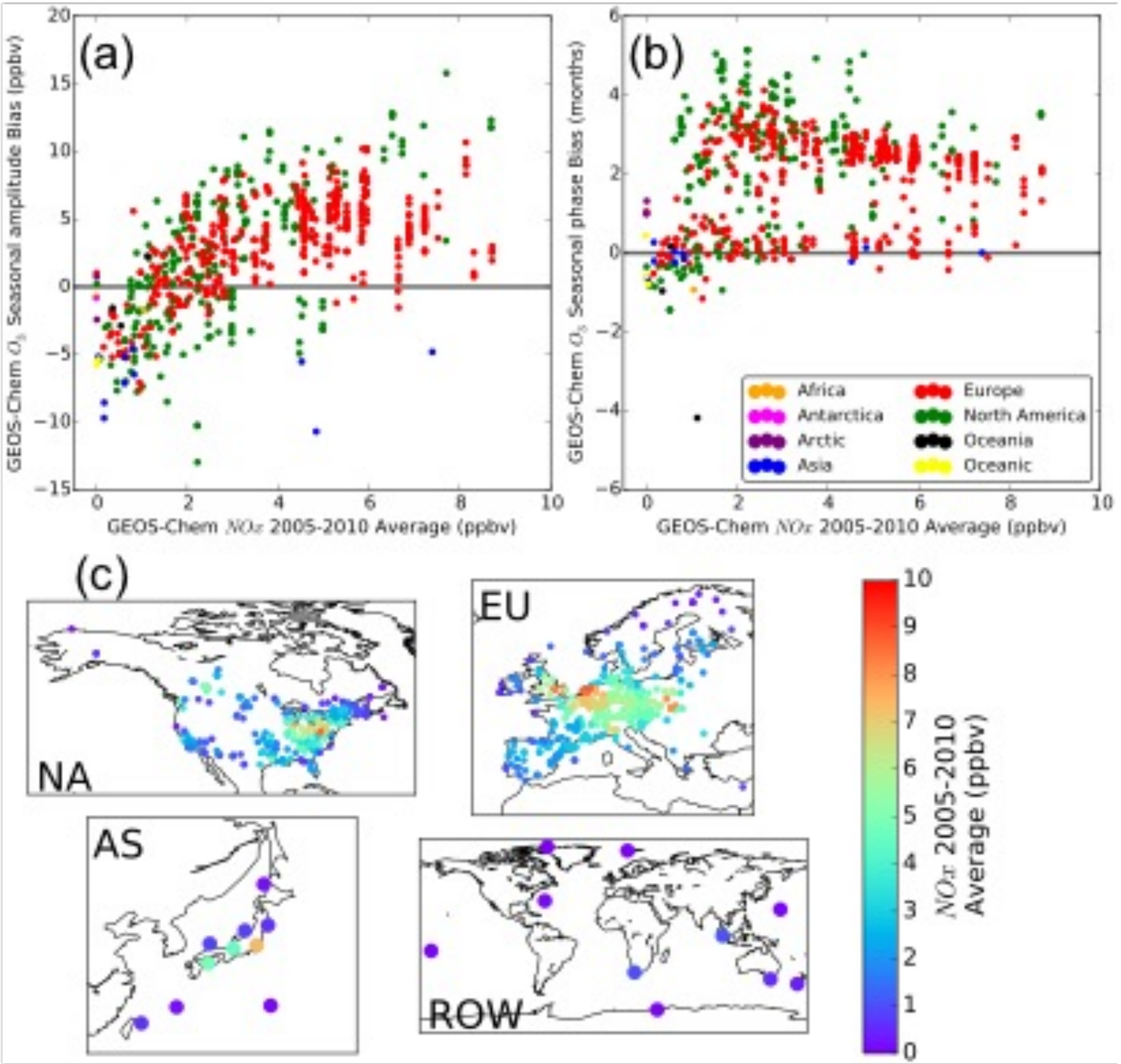
New approaches to evaluation:



www.atmos-chem-phys.net/16/8295/2016/

Atmos. Chem. Phys., 16, 8295–8308, 2016

New approaches to evaluation:



www.atmos-chem-phys.net/16/8295/2016/

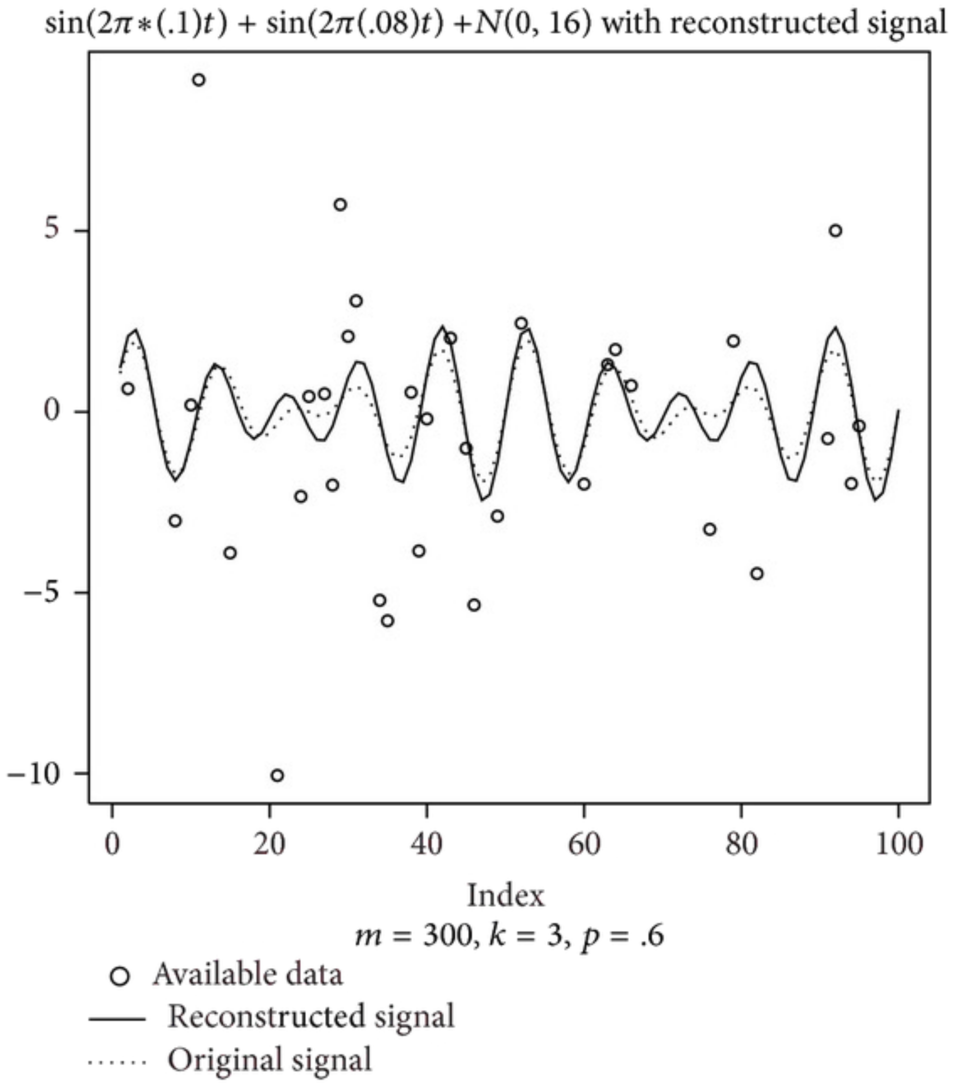
New approaches to evaluation:

Atmos. Chem. Phys., 17, 3001–3054, 2017
www.atmos-chem-phys.net/17/3001/2017/
doi:10.5194/acp-17-3001-2017

Atmospheric Chemistry and Physics
Open Access
EGU

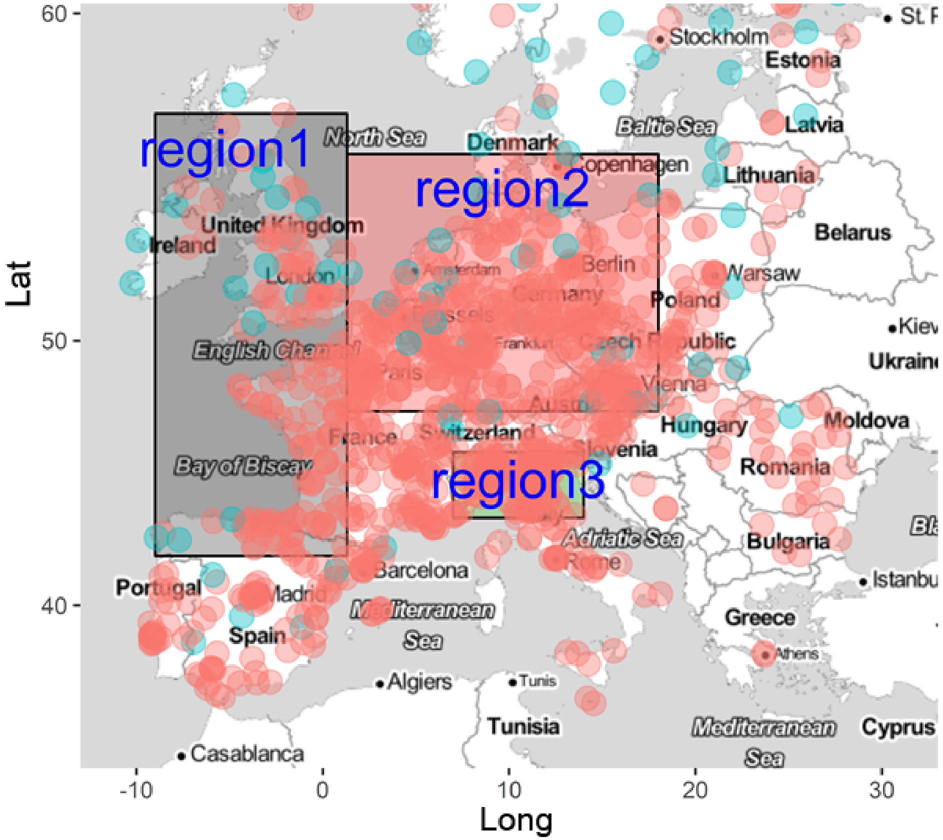
Evaluation and error apportionment of an ensemble of atmospheric chemistry transport modeling systems: multivariable temporal and spatial breakdown

Efisio Solazzo[1], Roberto Bianconi[2], Christian Hogrefe[3], Gabriele Curci[4,5], Paolo Tuccella[5], Ummugulsum Alyuz[6], Alessandra Balzarini[7], Rocío Baró[8], Roberto Bellasio[2], Johannes Bieser[9], Jørgen Brandt[10], Jesper H. Christensen[10], Augistin Colette[11], Xavier Francis[12], Andrea Fraser[13], Marta Garcia Vivanco[11,14], Pedro Jiménez-Guerrero[8], Ulas Im[10], Astrid Manders[15], Uarporn Nopmongcol[16], Nutthida Kitwiroon[17], Guido Pirovano[7], Luca Pozzoli[6,1], Marje Prank[18], Ranjeet S. Sokhi[12], Alper Unal[6], Greg Yarwood[16], and Stefano Galmarini[1]

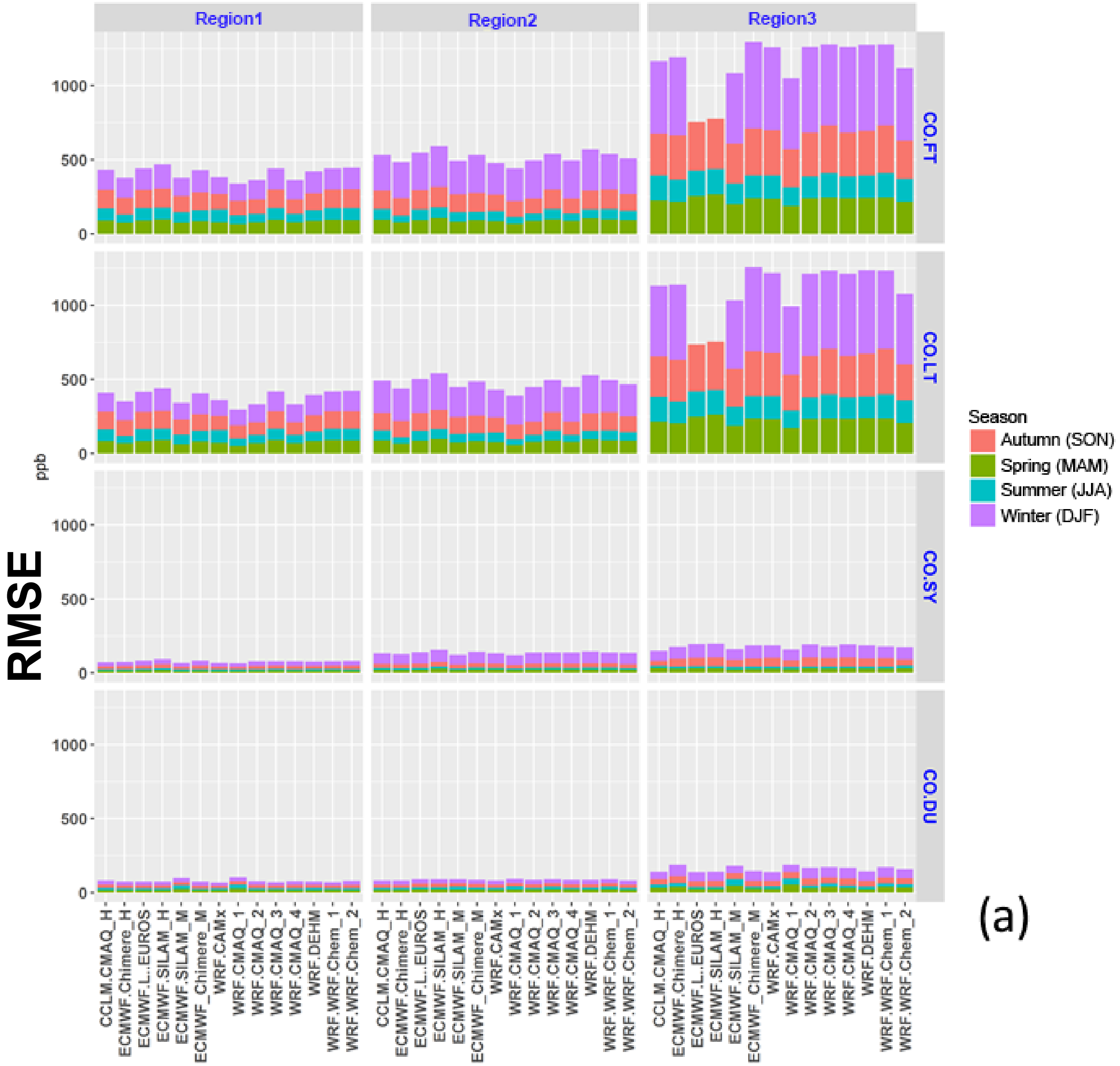


New approaches to evaluation: CO

the cause of model bias for CO is most probably attributable to the emissions and to a lesser extent the generally overestimated surface wind speed (Sect. 3.1.1). Sensitivity of the



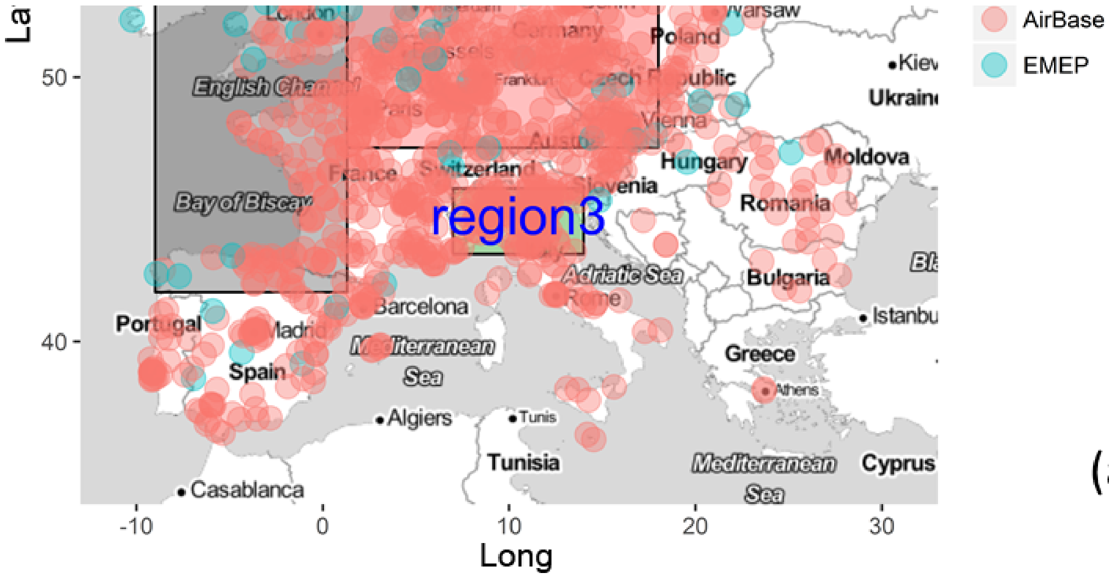
(a)



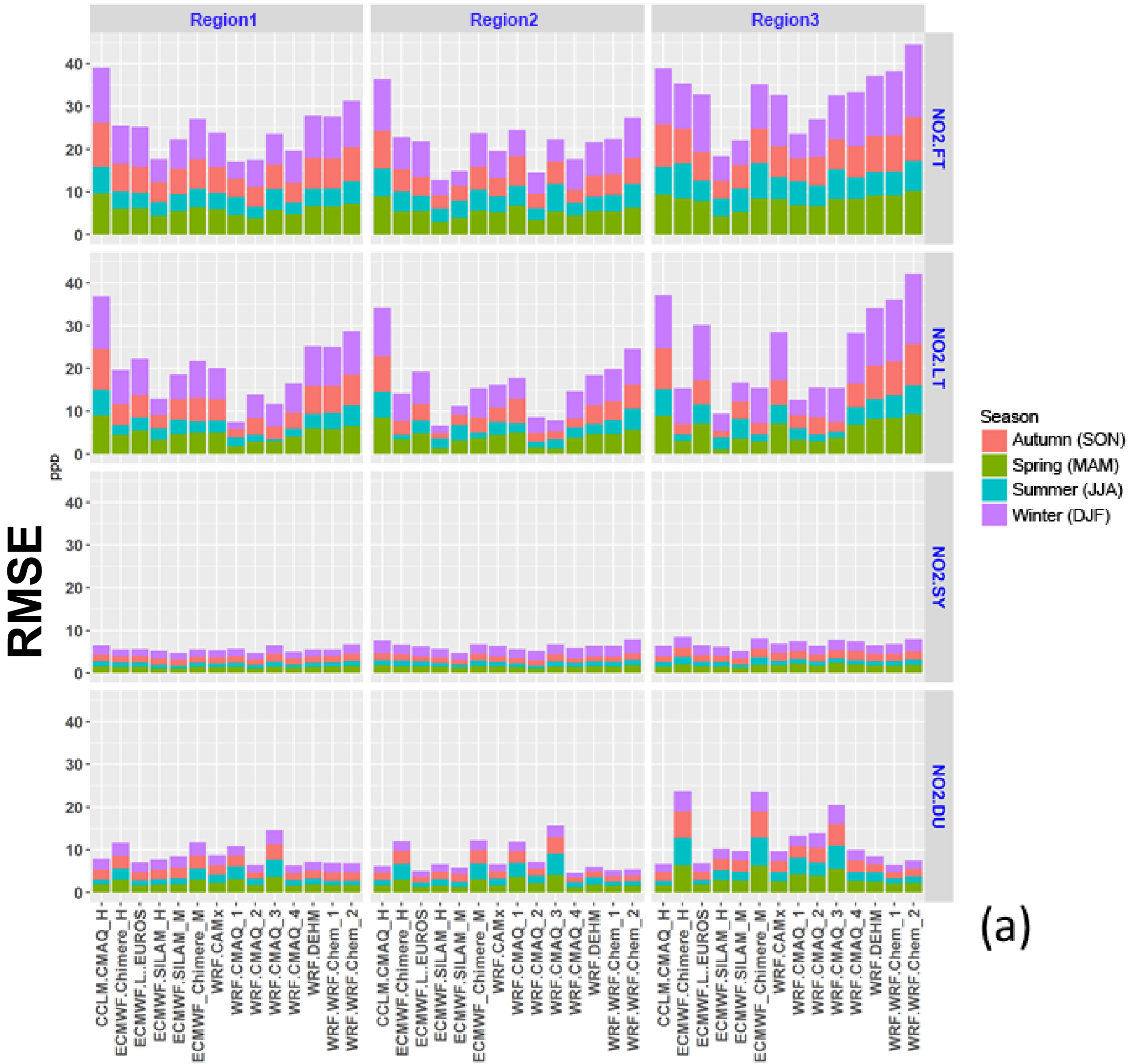
(a)

New approaches to evaluation: NO_2

The bias is the main contributor to the NO_2 error and stems from a model underprediction of the mean observed concentration during the entire year (but, with the exception of the winter season, it is positive for WRF-CMAQ in NA and WRF-CMAQ1 in EU; Table S7). The bias is probably caused by a combination of factors, including emission estimates (e.g., underestimation of residential combustion), PBL height and vertical mixing at night (when wood com-



(a)

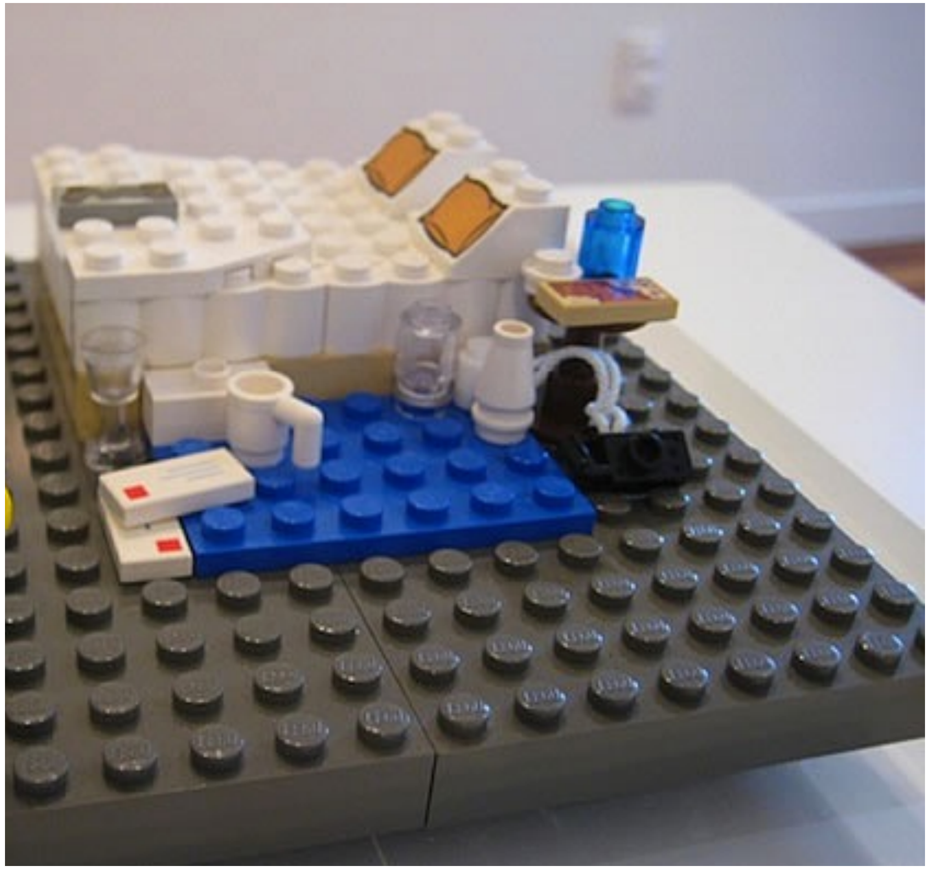


(a)

New approaches to evaluation: A success?

Although remarkable progress has been made since the first phase of AQMEII, both in terms of model performance and in terms of developing a more versatile and robust evaluation procedure, results of AQ model evaluation and intercomparison remain generic since they fail to associate errors with processes, or at least to narrow down the list of processes responsible for model error. AQ models are meant to be applicable to a variety of geographic (and topographic) scenarios under almost any type of weather, season, and emission conditions. For such a wide range of conditions the inherent nonlinearity among processes is difficult to disentangle, and specifically designed sensitivity runs seems to be the only viable alternative. A model evaluation strategy relying solely on the comparison of modeled vs. observed time series would never be able to quantify exactly the error induced by biogenic emissions, vertical emission profiles, or their dependence on temperature, deposition, and vertical mixing, for example, and the analyses presented in this work are no exception. In fact, the methodology devised to carry out the evaluation activity in this study has not succeeded in determining the actual causes of model error, although it does provide much clearer indications of the processes responsible for the error with respect to conventional operational model evaluation.

Comparing models and reality?



One of these images shows a Turner nominated art piece, which sold for £150,000.

Perturbed parameter ensembles

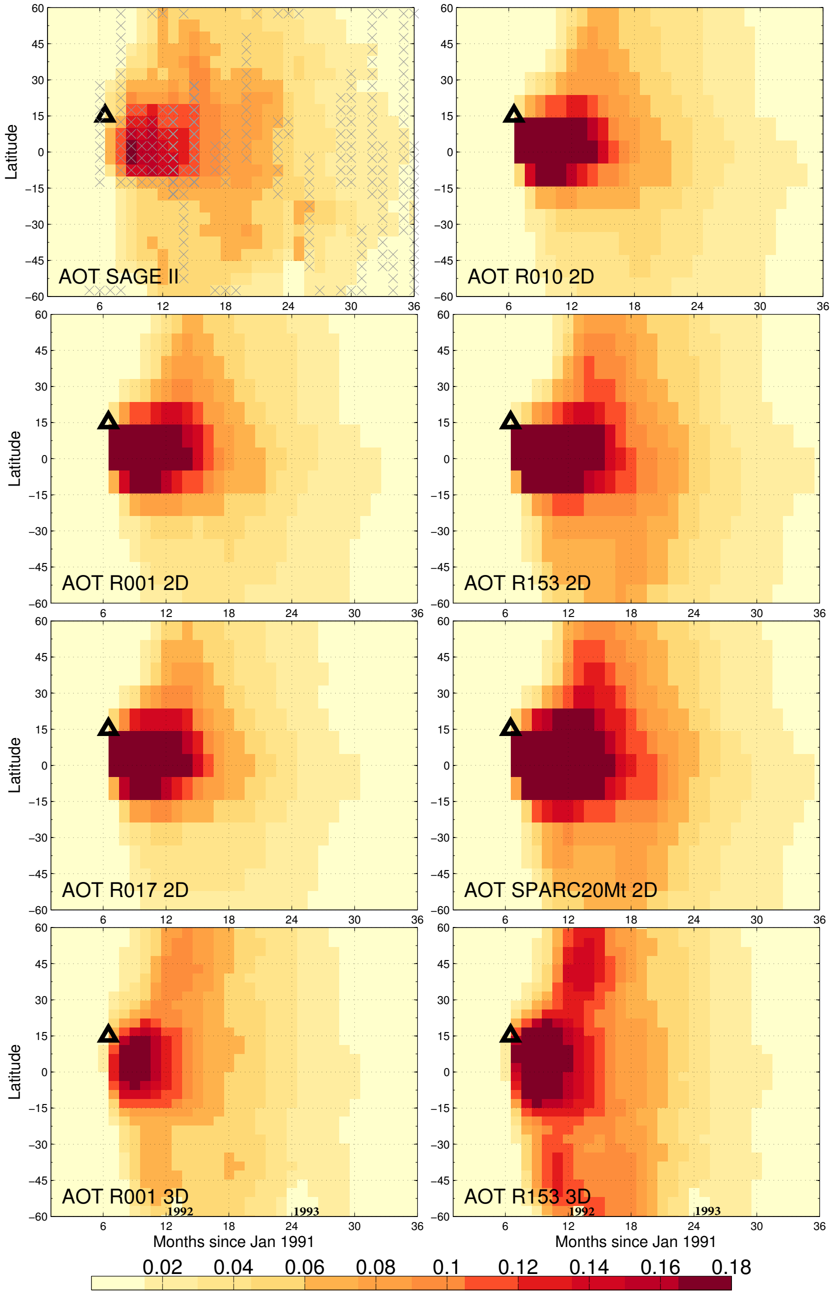
Atmos. Chem. Phys., 15, 11501–11512, 2015
www.atmos-chem-phys.net/15/11501/2015/
doi:10.5194/acp-15-11501-2015

Atmospheric Chemistry and Physics
Open Access

A perturbed parameter model ensemble to investigate Mt. Pinatubo's 1991 initial sulfur mass emission

J.-X. Sheng[1,a], D. K. Weisenstein[2], B.-P. Luo[1], E. Rozanov[1,3], F. Arfeuille[4,b], and T. Peter[1]

[1]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland
[2]School of Engineering and Applied Science, Harvard University, Cambridge, MA, USA
[3]Physikalisch-Meteorologisches Observatorium Davos and World Radiation Center, Davos, Switzerland
[4]Oeschger Centre for Climate Change Research and Institute of Geography, University of Bern, Bern, Switzerland
[a]now at: School of Engineering and Applied Science, Harvard University, Cambridge, MA, USA
[b]now at: Empa, Swiss Federal Laboratories for Materials Testing and Research, Dübendorf, Switzerland

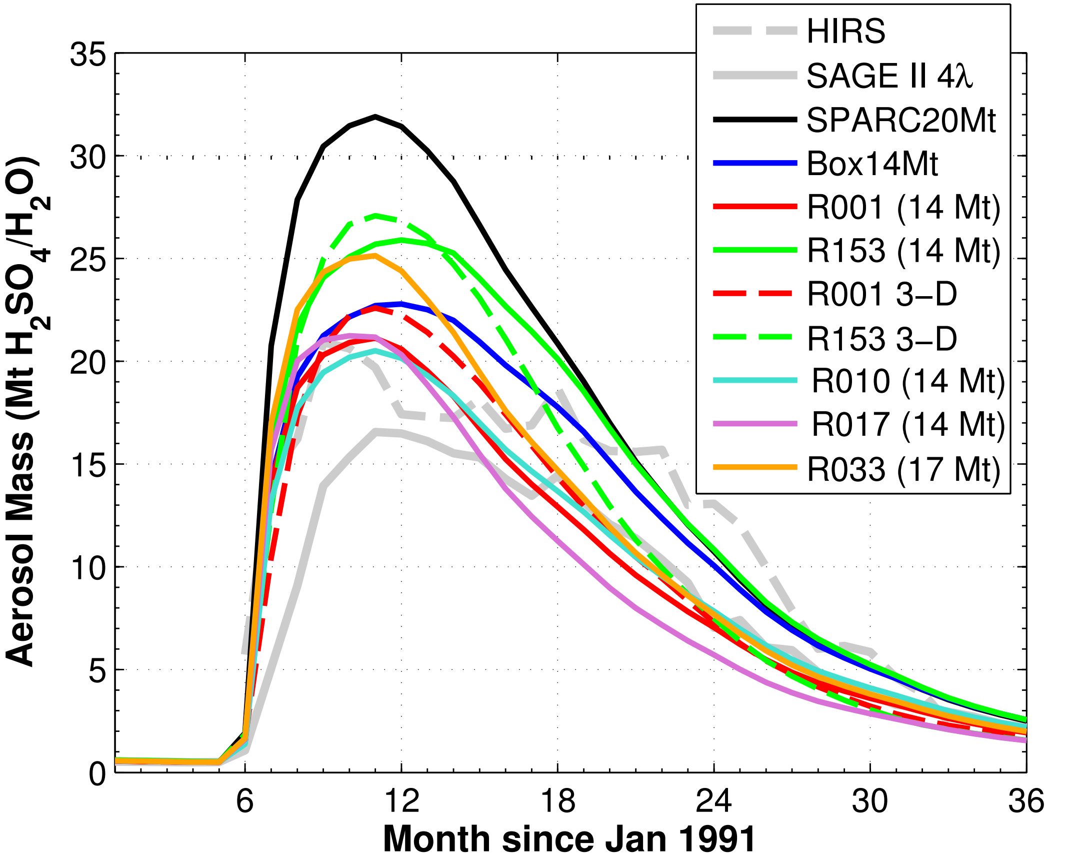


$$||X_{\mathrm{SO_2,model}} - X_{\mathrm{SO_2,MLS}}||/||X_{\mathrm{SO_2,MLS}}||$$

$$\mathrm{errOPC} = ||N_{\mathrm{model}} - N_{\mathrm{OPC}}||/||N_{\mathrm{OPC}}||$$

$$\frac{1}{2}(||\boldsymbol{B}^{t_1}_{\mathrm{model}} - \boldsymbol{B}^{t_1}_{\mathrm{HIRS}}||/||\boldsymbol{B}^{t_1}_{\mathrm{HIRS}}|| + ||\boldsymbol{B}^{t_2}_{\mathrm{model}} - \boldsymbol{B}^{t_2}_{\mathrm{SAGE}}||/||\boldsymbol{B}^{t_2}_{\mathrm{SAGE}}||)$$

Took the average of these to determine which of the 8 simulations was best



Perturbed parameter ensembles

Atmos. Chem. Phys., 11, 12253–12273, 2011
www.atmos-chem-phys.net/11/12253/2011/
doi:10.5194/acp-11-12253-2011

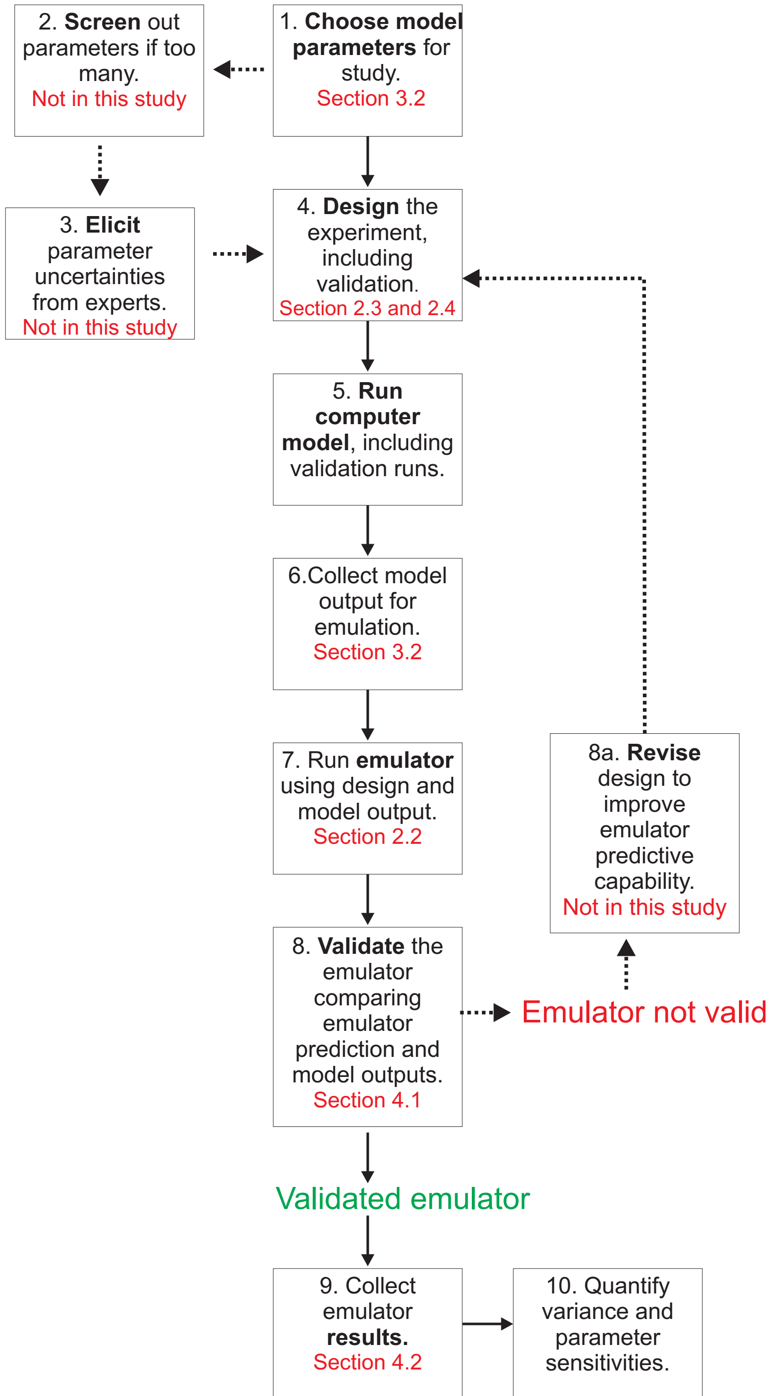
Atmospheric Chemistry and Physics

Emulation of a complex global aerosol model to quantify sensitivity to uncertain parameters

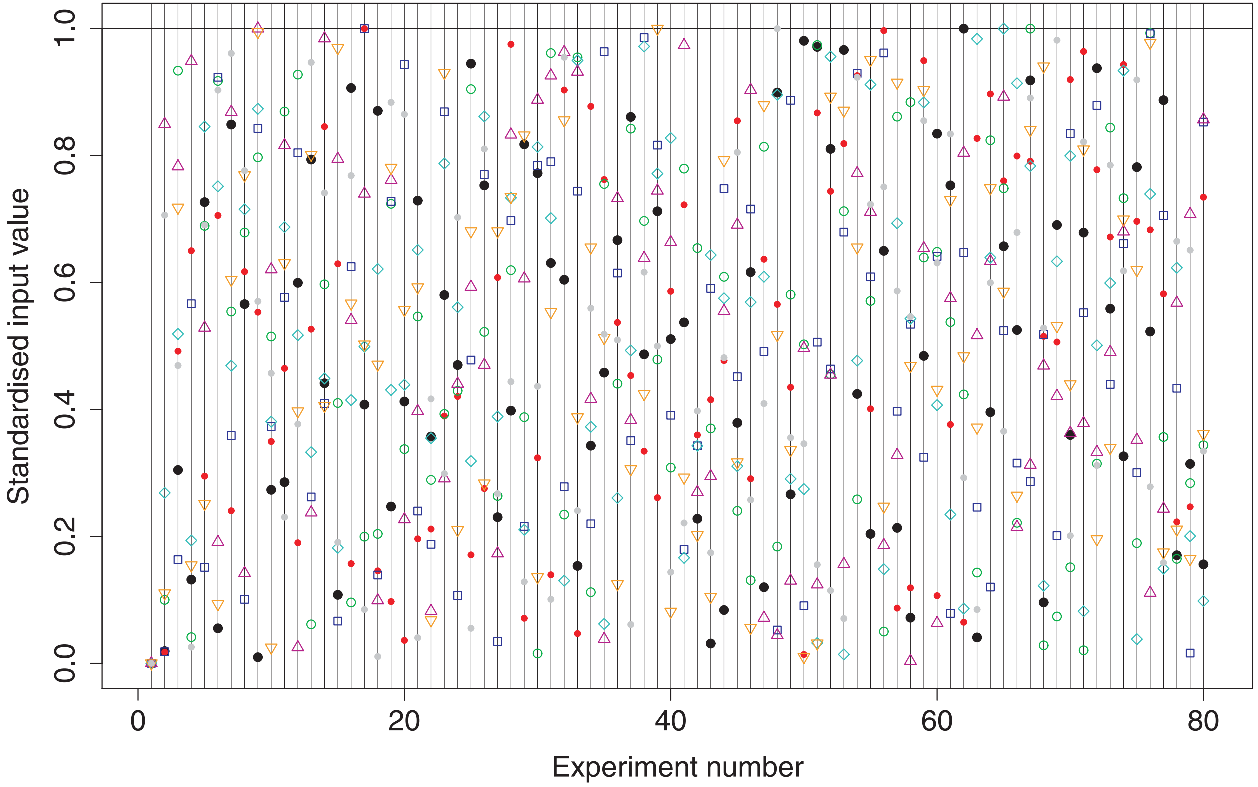
L. A. Lee, K. S. Carslaw, K. J. Pringle, G. W. Mann, and D. V. Spracklen

Institute for Climate and Atmospheric Science, University of Leeds, UK

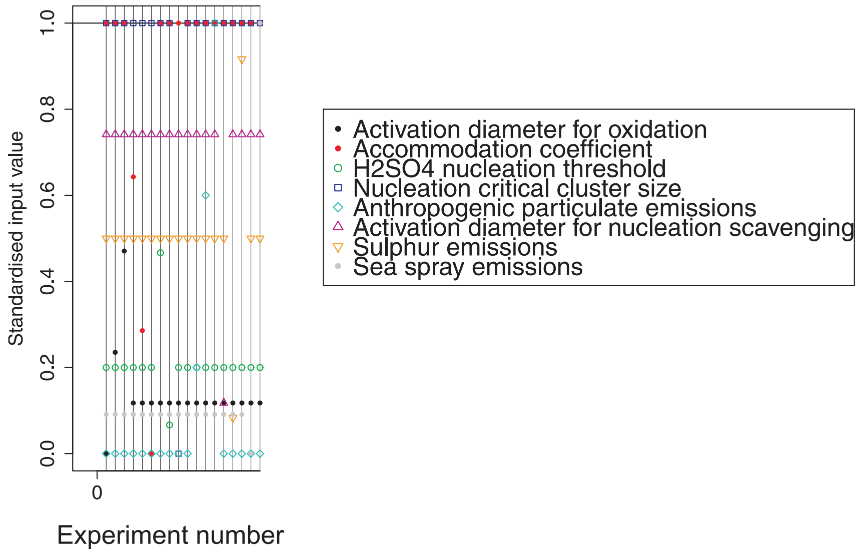
Received: 7 July 2011 – Published in Atmos. Chem. Phys. Discuss.: 19 July 2011
Revised: 11 November 2011 – Accepted: 16 November 2011 – Published: 8 December 2011



1. **Choose model parameters** for study.
Section 3.2
2. **Screen out** parameters if too many.
Not in this study
3. **Elicit** parameter uncertainties from experts.
Not in this study
4. **Design** the experiment, including validation.
Section 2.3 and 2.4
5. **Run computer model**, including validation runs.
6.Collect model output for emulation.
Section 3.2
7. Run **emulator** using design and model output.
Section 2.2
8. **Validate** the emulator comparing emulator prediction and model outputs.
Section 4.1
8a. **Revise** design to improve emulator predictive capability.
Not in this study
Emulator not valid
Validated emulator
9. Collect emulator **results.**
Section 4.2
10. Quantify variance and parameter sensitivities.

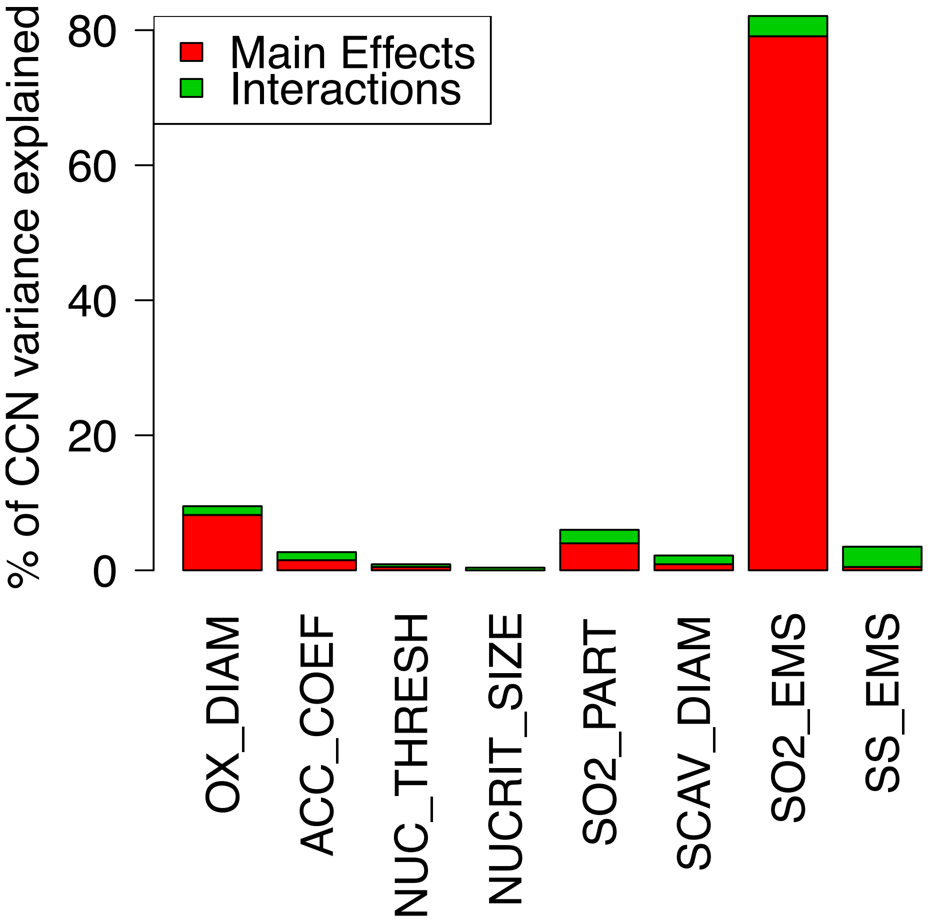


Standardised input value
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
Experiment number

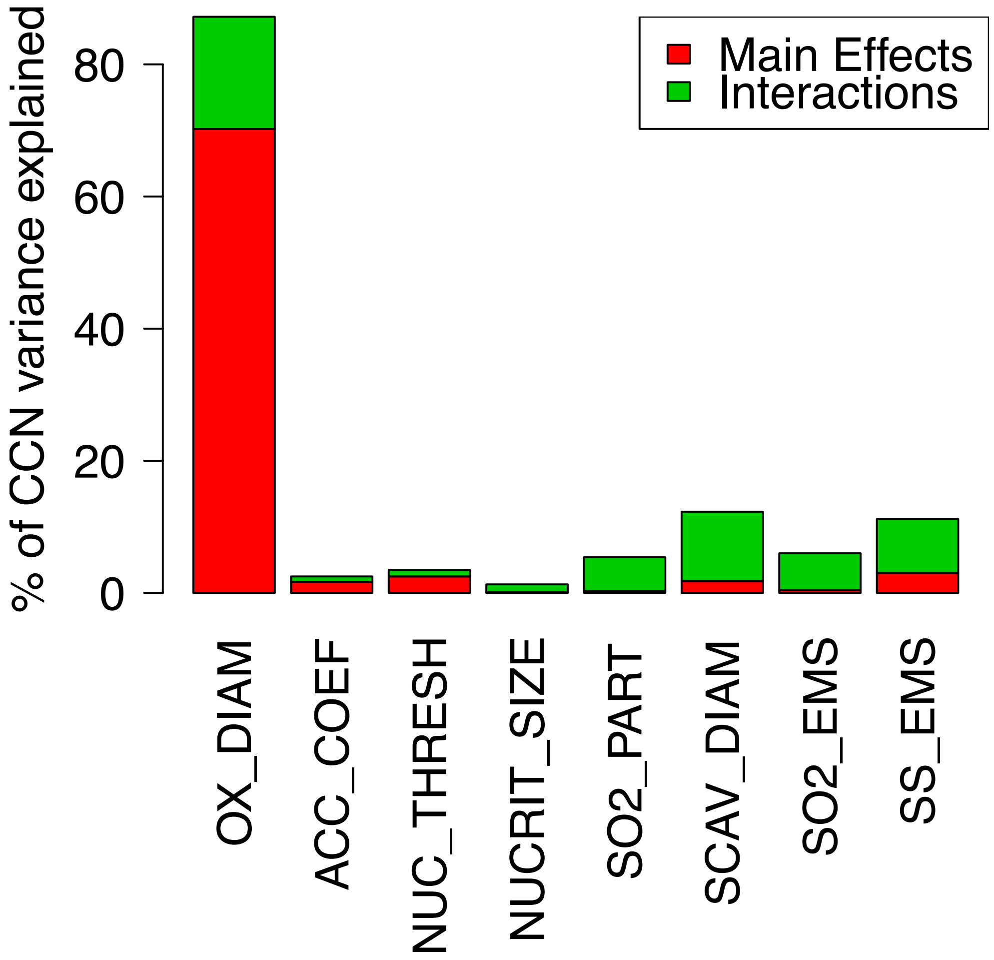


Standardised input value
0.0
0.2
0.4
0.6
0.8
1.0
0
Experiment number
Activation diameter for oxidation
Accommodation coefficient
H2SO4 nucleation threshold
Nucleation critical cluster size
Anthropogenic particulate emissions
Activation diameter for nucleation scavenging
Sulphur emissions
Sea spray emissions

Understanding of uncertainty for different areas (polluted and rural)



(a)



(b)

Everything should be made as simple as possible, but not simpler. **Albert Einstein**

Einstein clearly never used STASH

Practical steps:
How to “play with model data”

By now you have probably (hopefully) worked out where to find the results of your UKCA runs. Sorry about the file structure!

There are lots and lots of runs that are available for analysis and that have been archived on the Met Office MASS archive. To get access you will need an account but you can get access from MONSooN or JASMIN.

Practical steps:
How to “play with model data”

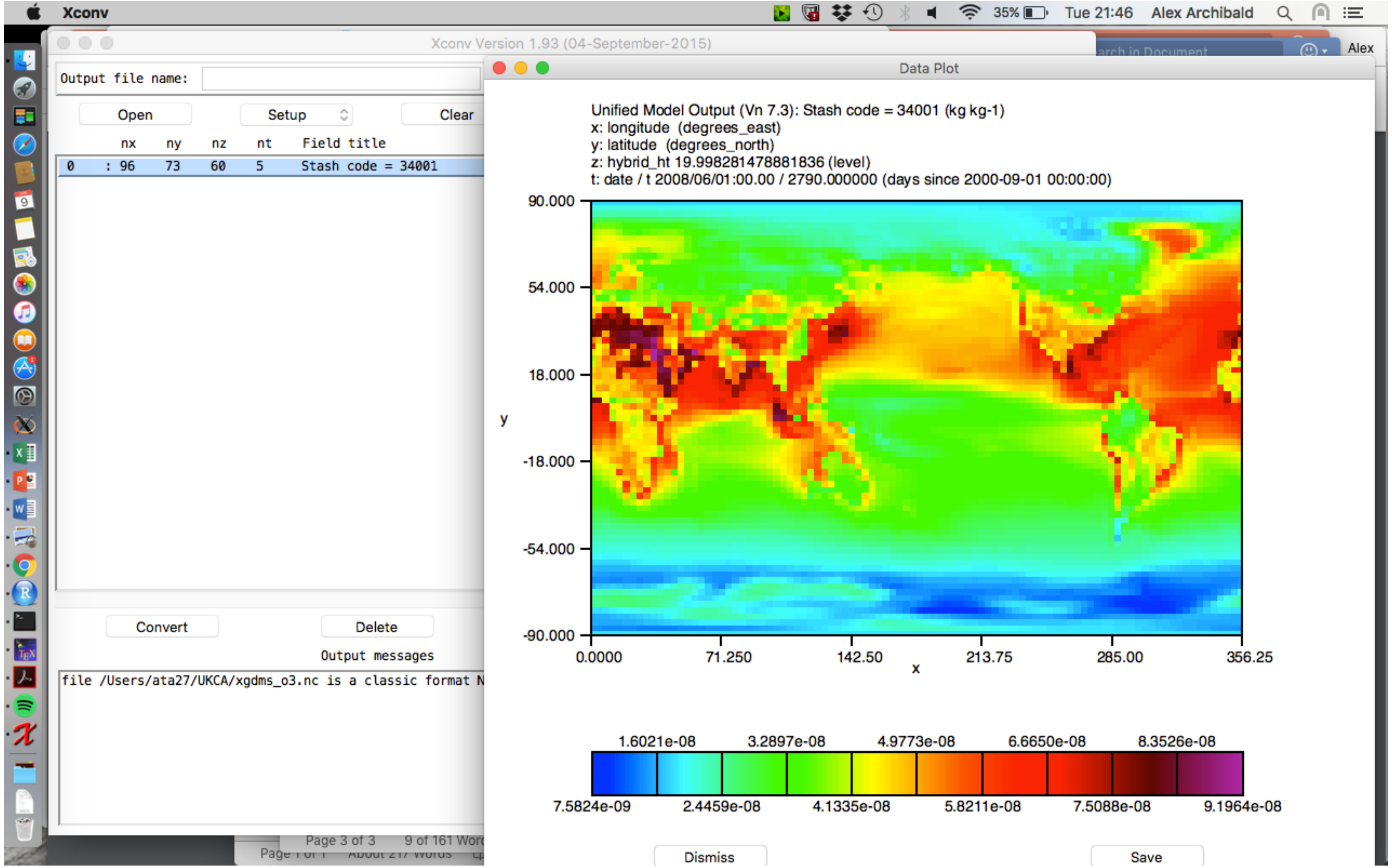
You will then need to make use of moo

moo ls –l :crum/xgywn

Will list all the archived model data from the (old) UKCA run xgywn.

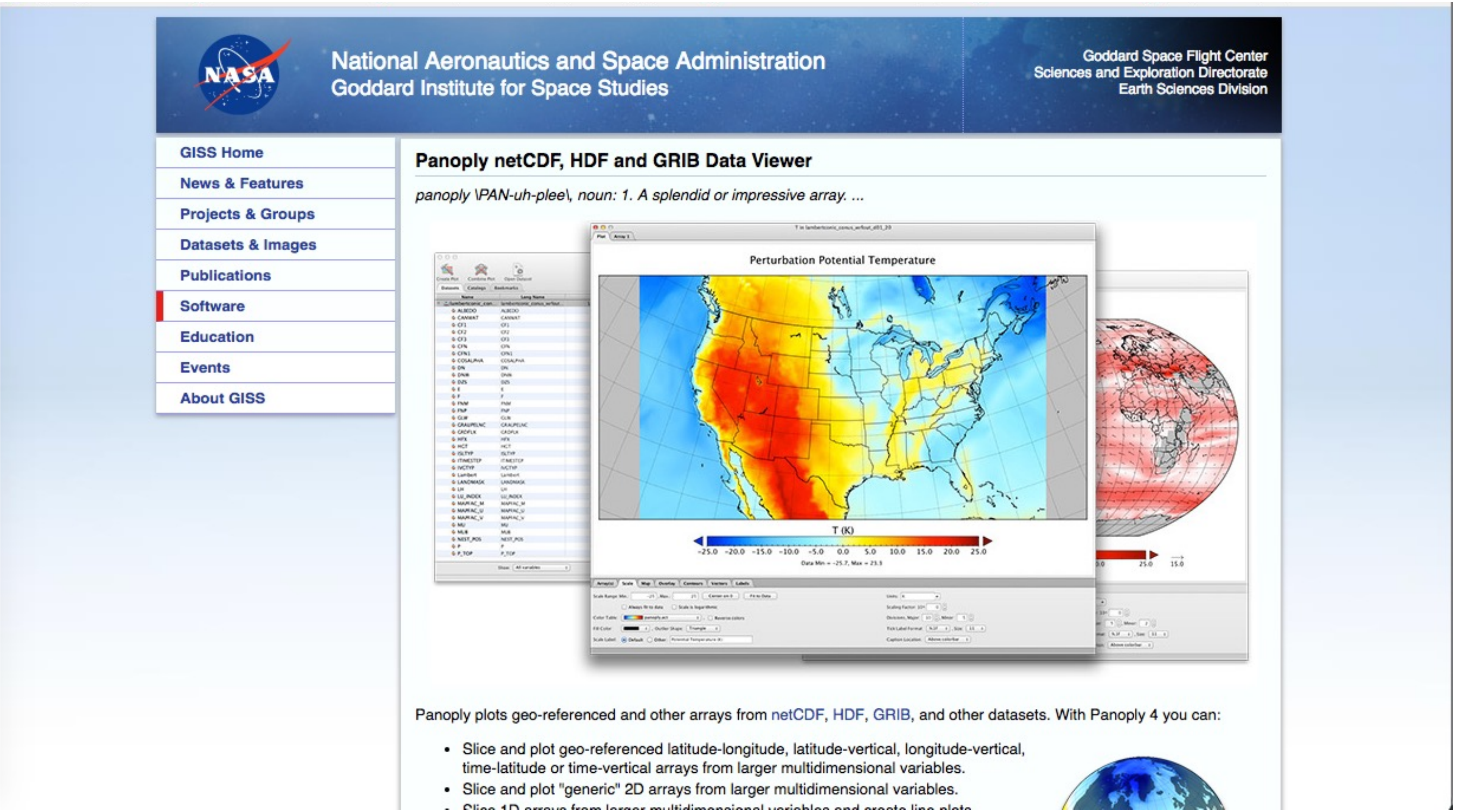
You will then be able to extract and save data (as PP files).
See here for more details:
http://cms.ncas.ac.uk/wiki/UM/GettingInitialData

Practical steps:
How to “play with model data”

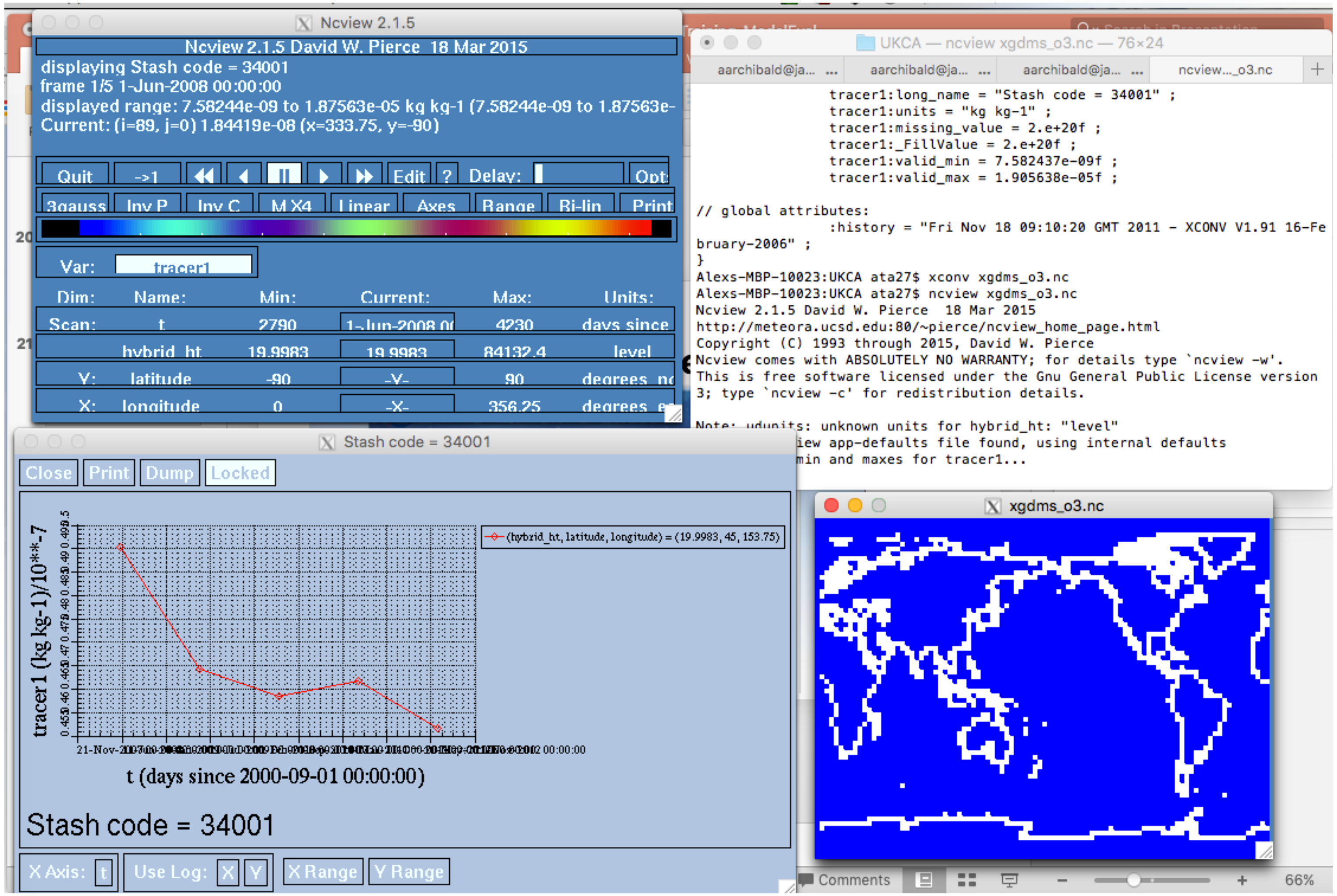


Xconv is very handy!
Especially because you can use it to convert PP to netCDF!

Practical steps:
How to “play with model data”



Practical steps:
How to "play with model data"



Practical steps:
How to evaluate your UKCA runs

Evaluation suite2

Contents [hide]

Navigation
Home
Recent changes
Random page
Help using the Wiki
Request an Account
Log in
Log out

Tools
What links here
Related changes
Special pages
Printable version
Permanent link

UKCA Evaluation Suite - Version 2

This is the description for Version 2. Instructions for Version-1 can be found here

The UKCA Evaluation suite is a collection of basic assessment methods being used at various partner institutions. The package can be divided largely into three categories:

- Tropospheric chemistry evaluation
- Stratospheric chemistry and dynamics evaluation
- Aerosol chemistry and processes evaluation --*Not yet available for V2, use from V1*

The evaluation suite can currently analyse the outputs from UM-UKCA configurations against specific observation datasets. To use the evaluation suite, the model output needs to be in the form of 12 x monthly mean pp files (UM *pm* stream) conforming to the UM pp file naming format (jobida.somename.pp). The *Stratospheric* as well as *Tropospheric* Chemistry suites can carry out multi-annual meaning if data for more than one year are specified as input. Each type of analysis requires a specified list of fields to be present in the pp files and there are UM STASH macro handedits to ensure that these fields are requested in the output.

For the Tropospheric and Stratospheric Chemistry part of evaluation, these fields can be added to your job by using the hand_edit:

- */home/h02/hadzm/umui_jobs/hand_edits/vn8.2/add_ukca_eval1_diags_lxx.ed* (where xx is the number of vertical levels) on the Met Office systems and
- */home/mdalvi/umui_jobs/hand_edits/vn8.2/add_ukca_eval1_diags_lxx.ed* on the Puma server.

For UM versions vn8.5-8.6, use the hand_edits:

- */home/h02/hadzm/umui_jobs/hand_edits/vn8.5/add_ukca_eval_diags_vn85_l85.ed* on the Met Office systems and

The UKCA Evaluation suite V2 is available on the Met Office Desktop, MONSooN postprocessor, ARCHER, and JASMIN (sci1,lotus).

Tropospheric Chemistry package

This currently performs the following analysis/ evaluations:

- CO against CMDL obs
- ClO against MLS data
- CO, HNO3, NOx against Emmons et al dataset
- OH using Lawrence etal (1991) method, vs ACCMIP and Parta et al Reference values.
- O3 against Tilmes ozonesonde data at multiple locations
- Ox budget (sources/sinks, deposition)
- OH/CH4 ratio and CH4 lifetime vs ACCMIP multi-model values
- Profiles of O3,HNO3,NO2,H2O2,water vap against ACE,UARS,.. data
- Age of Air against SF6 obs

The tool will automatically carry out multi-annual meaning on-the-fly if more than 12 (and a multiple of 12) files are detected.

Usage:

A. Met Office desktop:

/home/h02/hadzm/eval_v2/camb_chem/eval_tropchem_spice.py -i <ppfiles> [-s STASHlist] [-m trmap] [--eval_only] [--noclean] [--nocopy]

Options

-h, --help	show this help message and exit
-i	Required: ppfiles (12) from the year to analyse -full path-
-s STASHLIST	Optional: STASHcodes list, e.g. if using pre-vn8.5 output (diags in Section 34 vs 50)
-m TRMAPS	Optional :Var<->STASH mapping file, e.g. if using pre-vn8.5 output
-f SCALE_FAC	Optional : Flux multiplication factor, to account for difference in UM:UKCA call frequency (default=3.0 for 1 UM: 3 UKCA timesteps)
--eval_only	Optional: Only carry out Evaluation, skipping the extraction. Useful when extract is ok but evaluation has previously failed.
--noclean	Optional: Do not delete extracted NetCDF data after completion
--nocopy	By default the input files will be copied temporarily to /scratch for processing. Use this option *only if model output is in $DATADIR*

B. MONSooN Postprocessor:

/home/mdalvi/eval_v2/camb_chem/eval_tropchem.py -i <ppfiles> [-s STASHlist] [-m trmap] [--eval_only] [--noclean]

ARCHER

The scripts have been modified slightly to work on ARCHER. They are designed such that output from a UM vn10.6 job or above will not require any additional arguments, other than the location of the *.pp files. Also, all Stratospheric output is saved as .pdf.

While you can use the login nodes for this, you can also log-in to the post-processing nodes by

```
ssh -Y esppl
```

Required modules

To be able to use Iris (required for both Stratospheric and Tropospheric packages), you will need to:

```
module load anaconda/2.2.0-python2
```

To run the Tropospheric chemistry package you will need to load R by

```
module load R
```

and load the required netCDF libraries by

```
module load cray-netcdf-hdf5parallel
```

Running the packages

For UM versions vn10.6 and above you will just need to:

- **Tropospheric Chemistry**: `/work/n02/n02/ukca/Eval/eval_v2/camb_chem/eval_tropchem.py -i /path/to/pp/files/*.pp [--eval_only]`
- **Stratospheric Chemistry**: `/work/n02/n02/ukca/Eval/eval_v2/toms_haloe/compare_toms_haloe.py /path/to/pp/files/*.pp`

You can use `evince` to view the outputted .pdf files.

If you need to run these on pre-vn10.6 versions, the equivalent to *<mohit home>* is ***/work/n02/n02/ukca/Eval*** for the paths to particular STASH maps etc.

The Stratospheric chemistry package takes a few minutes for a single years-worth of data. The Tropospheric chemistry package will take about 50 minutes to extract the data to netCDF and then an additional 15-20 minutes to produce the plots.

Example Data

Example data from vn10.9 can be found at

```
/work/n02/n02/ukca/Eval/ExampleData/u-as022
```

and an empty directory with the data already extracted to netCDF can be found here:

```
/work/n02/n02/ukca/Eval/WorkedExample/u-as022
```

This directory can be `rsync`-d to a working directory and the evaluation suite can be run using the `--eval_only` command which will save some time.

Required Chemistry Diagnostics

The UKCA chemistry evaluation packages require the following diagnostics (STASH section/item numbers from vn10.3 onwards):

STASH Section	STASH Item	STASH Name
0	010	SPECIFIC HUMIDITY AFTER TIMESTEP
0	408	PRESSURE AT THETA LEVELS AFTER TS
16	004	TEMPERATURE ON THETA LEVELS
30	451	Pressure at Tropopause Level
30	453	Height at Tropopause Level
34	001	O3 MASS MIXING RATIO AFTER TIMESTEP
34	002	NO MASS MIXING RATIO AFTER TIMESTEP
34	004	NO2 MASS MIXING RATIO AFTER TIMESTEP*
34	007	HONO2 MASS MIXING RATIO AFTER TSTEP
34	009	CH4 MASS MIXING RATIO AFTER TSTEP
34	010	CO MASS MIXING RATIO AFTER TSTEP
34	042	ClO MASS MIXING RATIO AFTER TSTEP
34	049	N2O MASS MIXING RATIO AFTER TIMESTEP
34	081	OH MASS MIXING RATIO AFTER TIMESTEP

50	022	Ox BUDGET: NOy DRY DEPOSITION (3D)
50	031	Ox BUDGET: NOy WET DEPOSITION (3D)
50	041	RXN FLUX: OH+CH4 (CH4 LIFETIME) TROP
50	051	STE: O3
50	061	AIR MASS DIAGNOSTIC (TROP ONLY)
50	062	TROPOSPHERIC MASK
50	063	AIR MASS DIAGNOSTIC (WHOLE ATMOS)
50	219	Ozone column in Dobson Units

Notes:

* NO2 is not available in s34i004 in StratTrop/CheST
** HCl is only available in s34i992 in StratTrop/CheST
** NO2 is only available in s34i996 in StratTrop/CheST

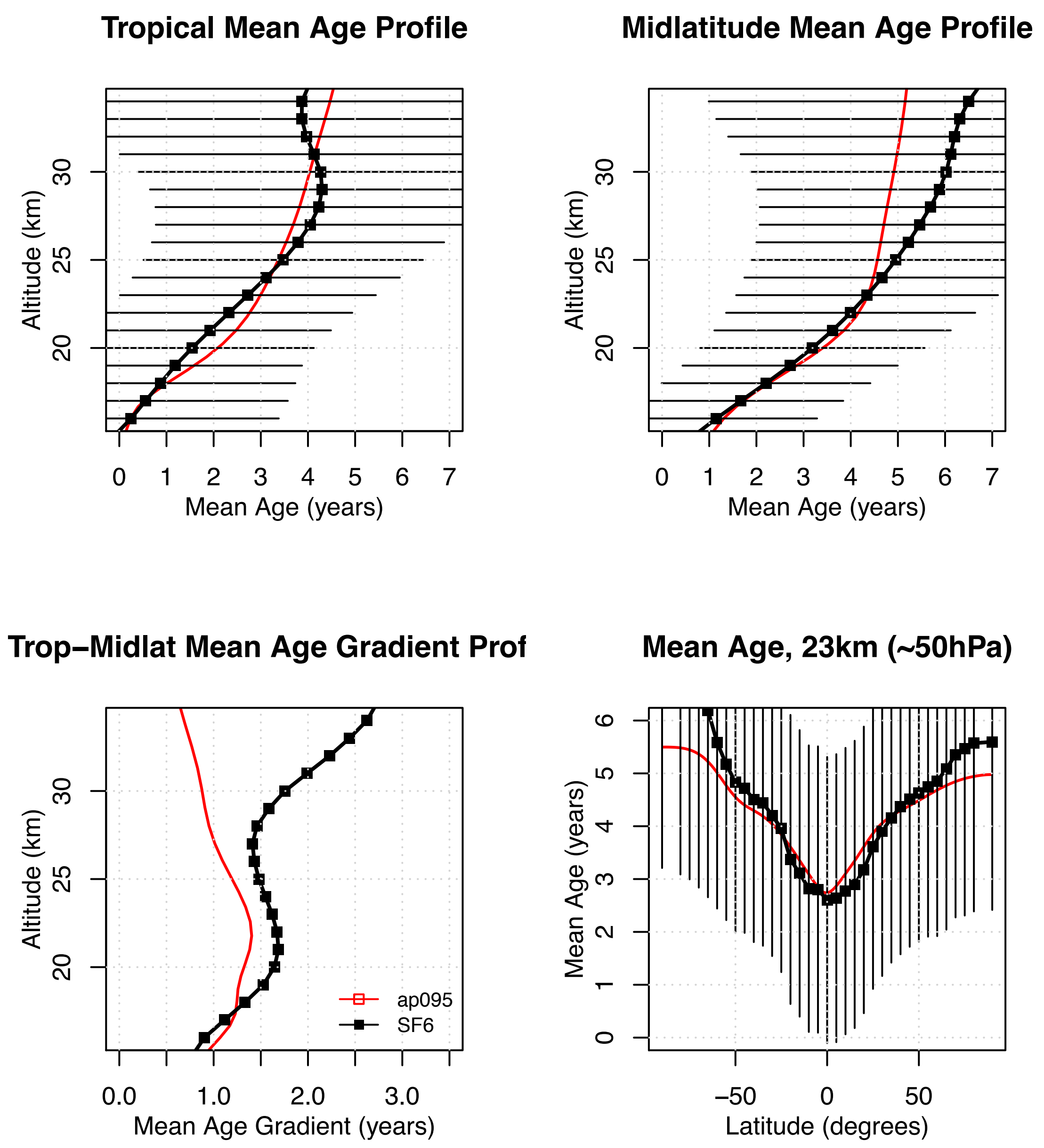
If you are extracting these from MASS, you can use the following file with the `moo select` command:

```
begin
   stash=(10, 408, 16004, 30451, 30453, 34001, 34002, 34004, 34007, 34009, 34010, 34042, 34049, 34081, 34150, 34992,
     34996, 50001, 50002, 50003, 50004, 50005, 50006, 50007, 50011, 50012, 50013, 50014, 50015, 50016, 50017, 50021,
     50022, 50031, 50041, 50051, 50061, 50062, 50063, 50219)
end
```

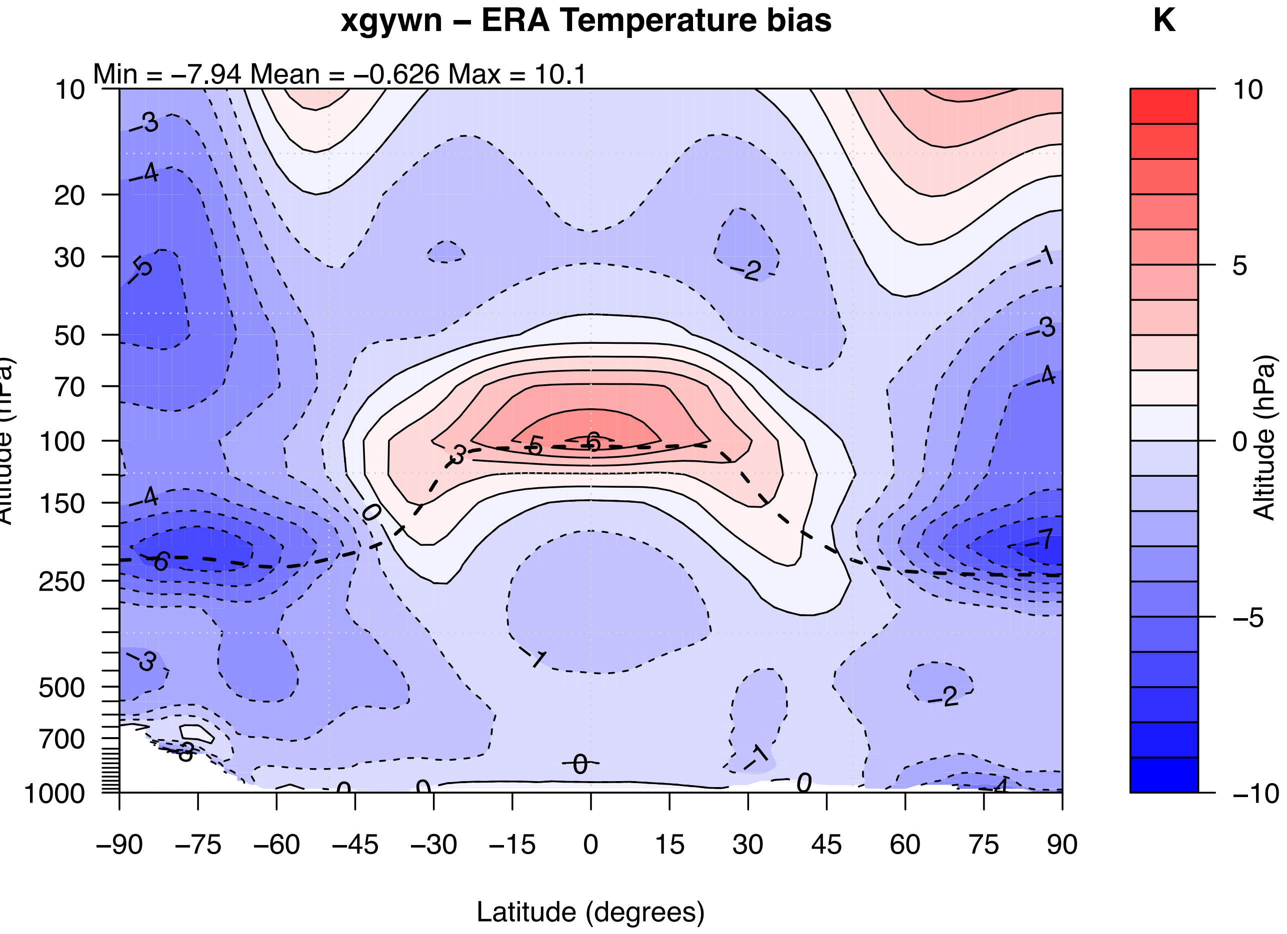
If you want to extract e.g. only a particular year, you can specify additional ranges, e.g.:

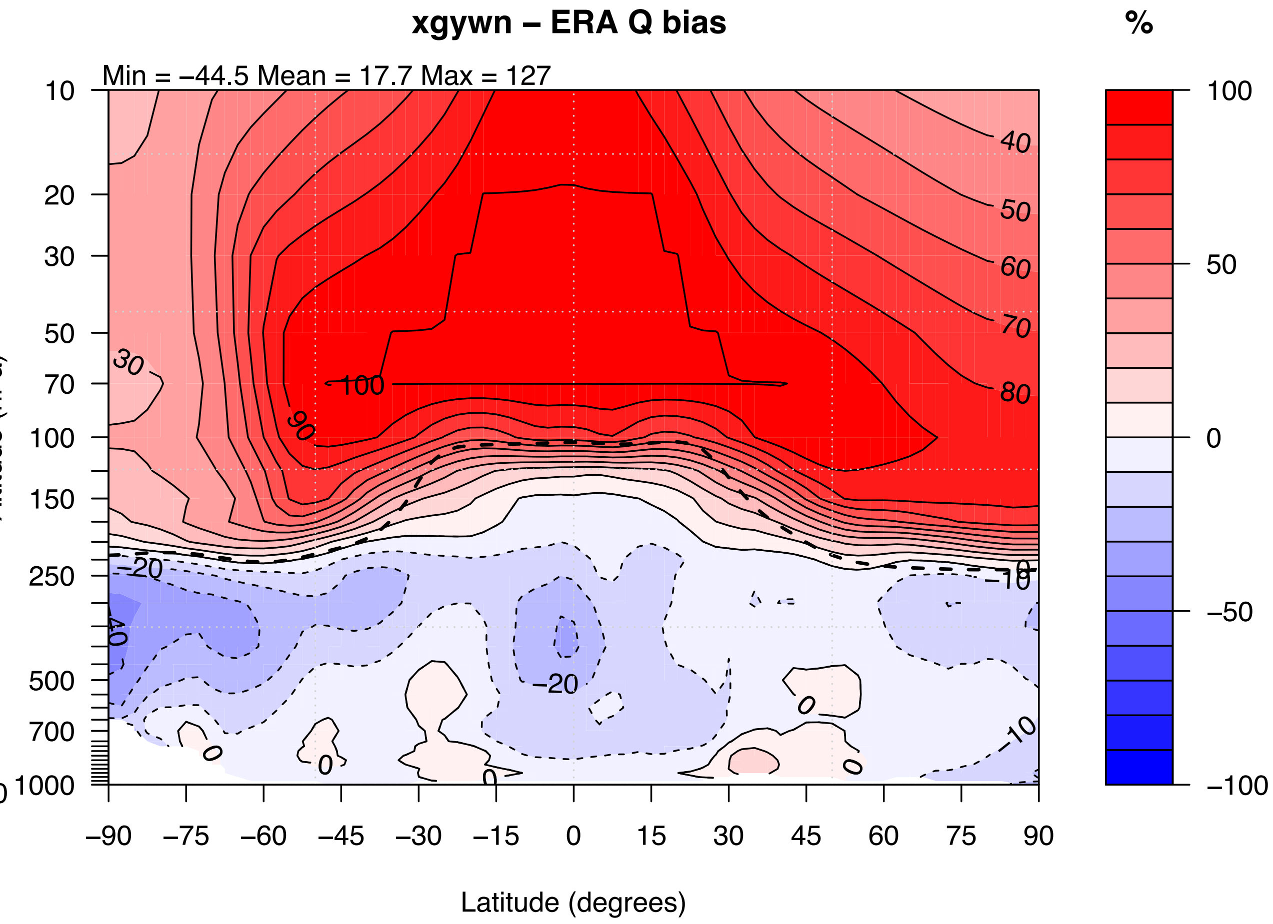
```
   T1>={2008/01/01 00:00}
   T1<={2008/12/30 23:59}
```

Comparison of the age of air against satellite SF_6 data

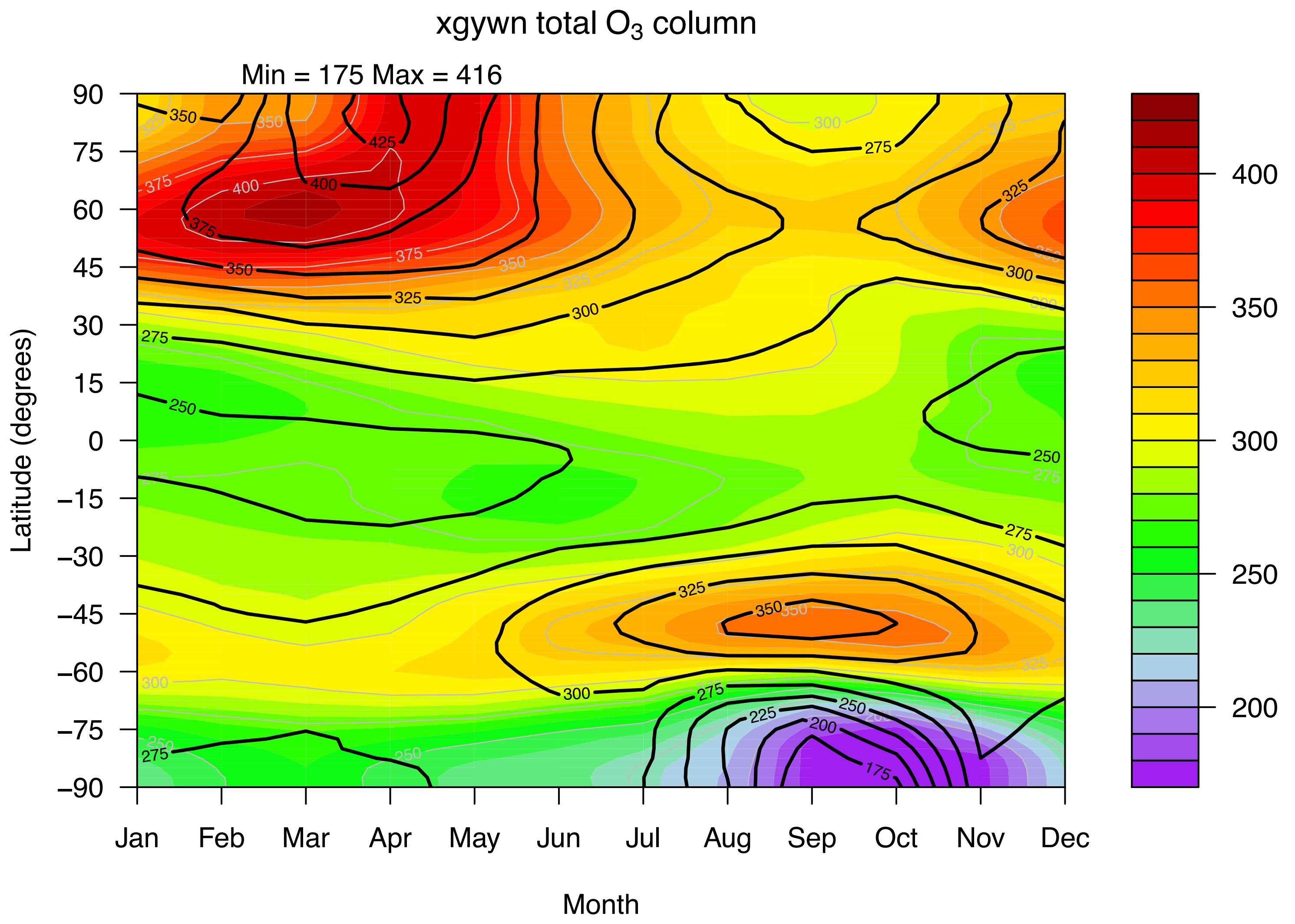


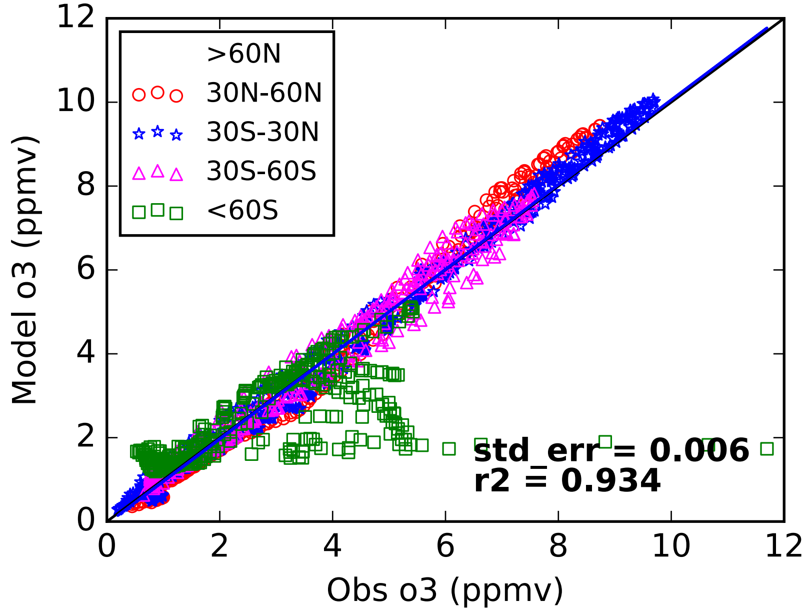
Comparison of temperature and humidity against ECMWF reanalysis

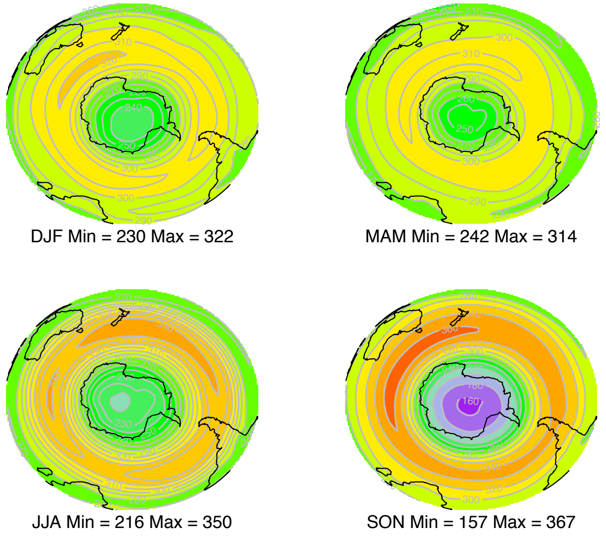




Comparison of total ozone column against Bodeker dataset

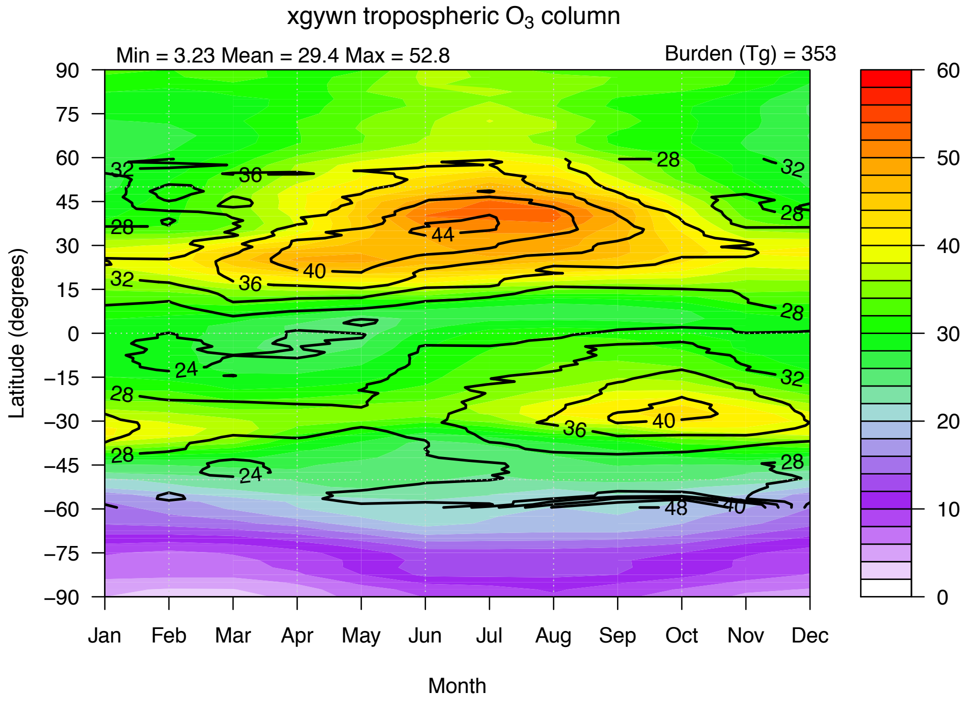


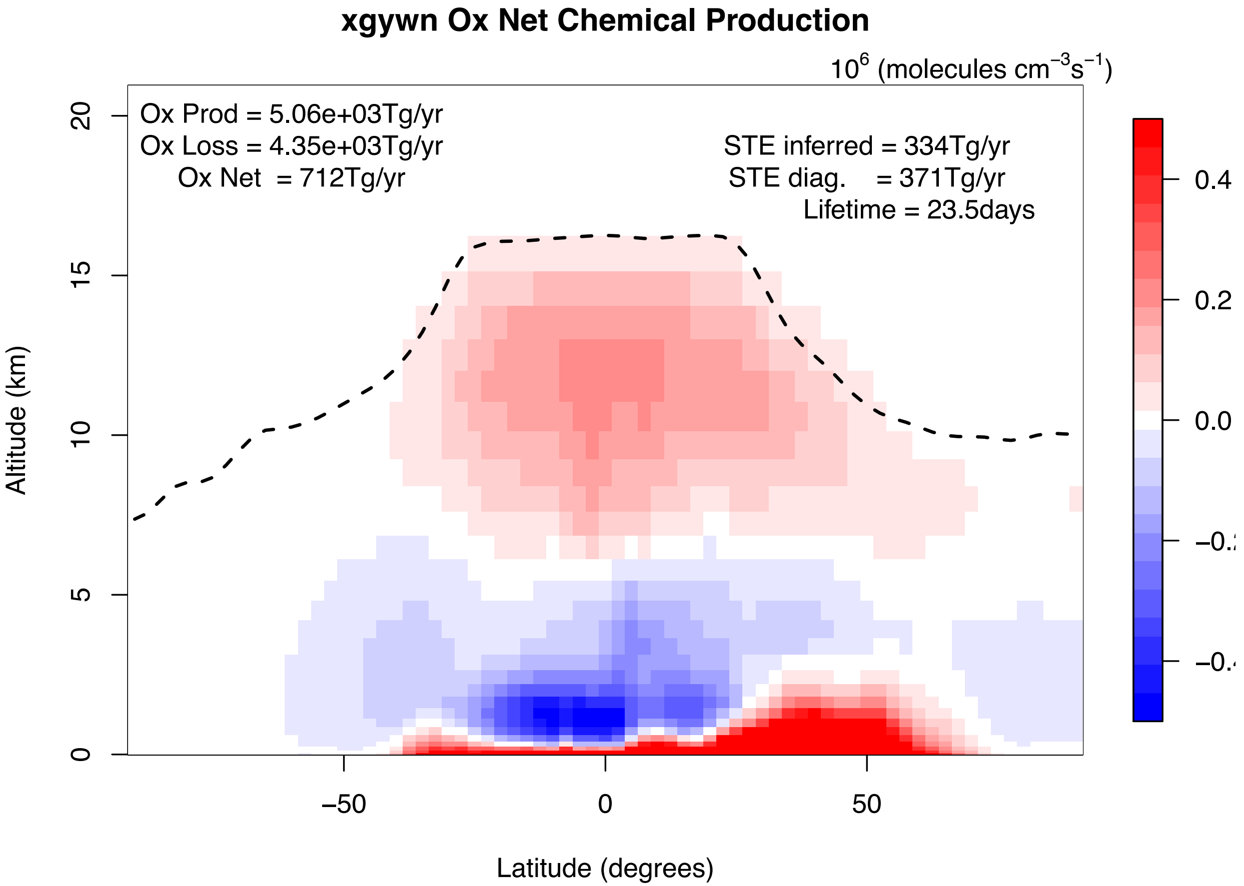


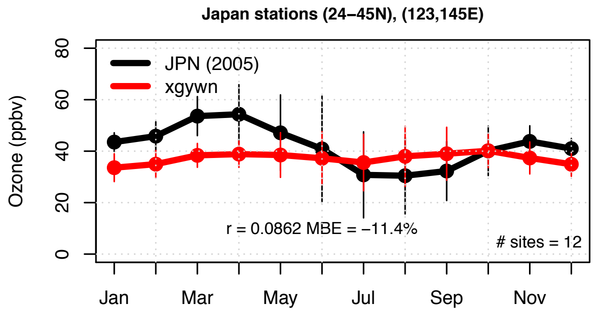


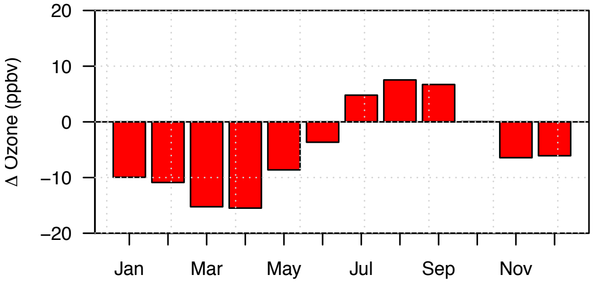
Comparison of tropospheric ozone and budget

MIPs have quantified the tropospheric ozone budget and we can evaluate our model against these.









Towards process based model evaluation

Given the huge number of diagnostics its getting harder and harder to evaluate models and determine their validity and there is a movement towards process based evaluation. This requires evaluation of the processes or diagnostics/prognostics in the model that contribute to the e.g. tracer. Once these are identified it is common to compare to obs and grade using the following:

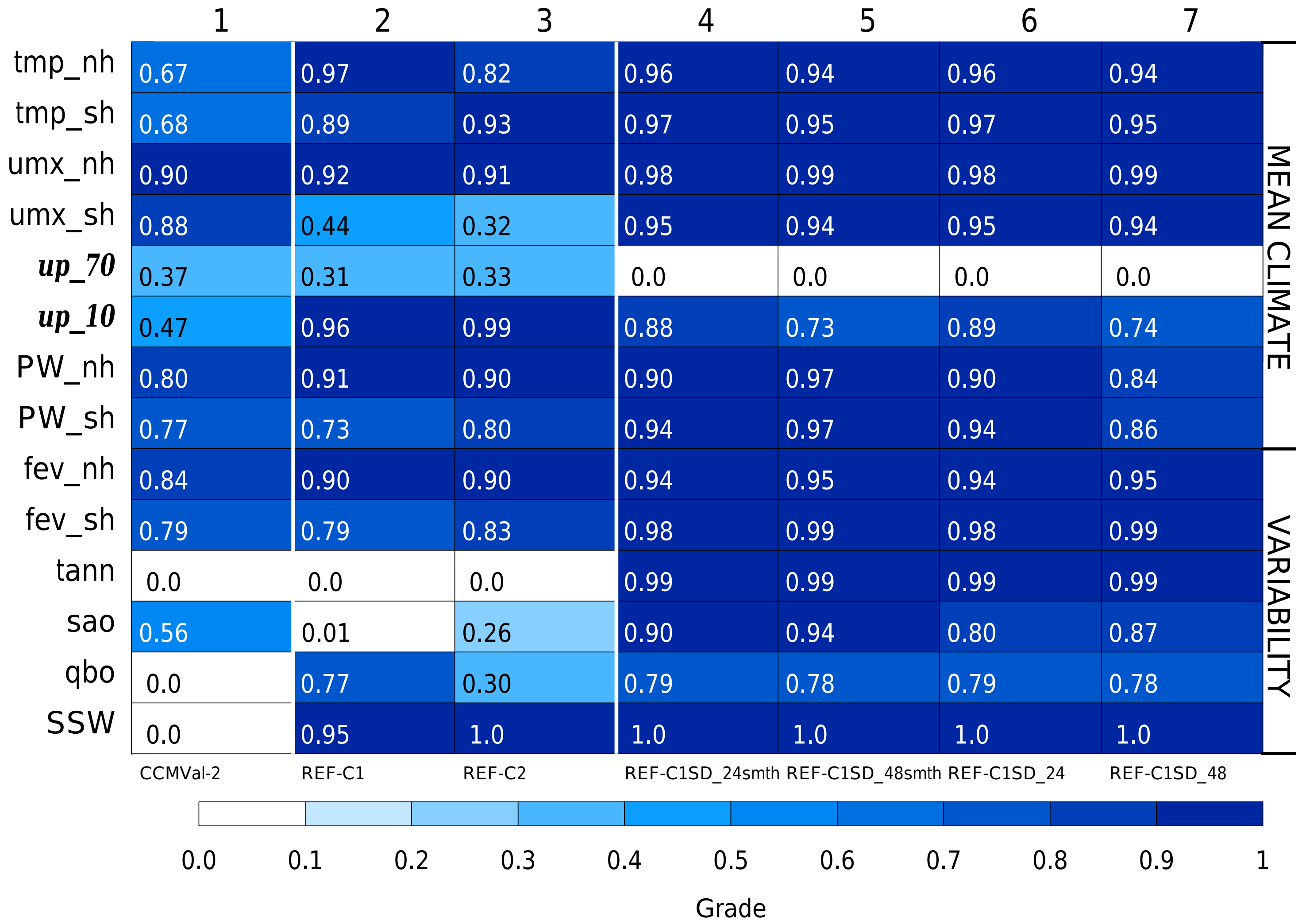
$$g = 1 - \frac{1}{3}\frac{|\mu_{\mathrm{model}} - \mu_{\mathrm{obs}}|}{\sigma_{\mathrm{obs}}}$$

Name	Description
Mean climate	
tmp_nh	60–90° N December–January–February temperatures at 50 hPa
tmp_sh	60–90° S September–October–November temperatures at 50 hPa
umx_nh	Maximum Northern Hemisphere eastward wind in December–January–February at 10 hPa
umx_sh	Maximum Southern Hemisphere eastward wind in June–July–August at 10 hPa
up_70	Tropical upwelling mass flux at 70 hPa
up_10	Tropical upwelling mass flux at 10 hPa
PW_nh	Slope of the regression of the February and March 50 hPa temperatures 60–90° N on the 100 hPa January and February heat flux 40–80° N
PW_sh	Slope of the regression of the August and September 50 hPa temperatures 60–90° S on the 100 hPa July and August heat flux 40–80° N
Variability	
fev_nh	Amplitude of the leading mode of variability (EOF) of the 50 hPa zonal-mean zonal wind for the Northern Hemisphere, poleward of 45° (EOFs are scaled to have the same standard deviation as the original data)
fev_sh	Amplitude of the leading mode of variability (EOF) of the 50 hPa zonal-mean zonal wind for the Southern Hemisphere, poleward of 45° (EOFs are scaled to have the same standard deviation as the original data)
tann	Amplitude of the annual cycle at 2 hPa in the zonal-mean zonal wind, 10° S–10° N
SAO	Amplitude of the semi-annual oscillation at 1 hPa in the zonal-mean zonal wind, 10° S–10° N
QBO	Amplitude of the quasi-biennial oscillation at 20 hPa in the zonal-mean zonal wind, 10° S–10° N
SSW	Frequency per year of major sudden stratospheric warmings, defined using reversal of the zonal-mean zonal wind at 10 hPa, 60° N

Evaluation of UKCA CCMI set up

Geosci. Model Dev., 10, 1209–1232, 2017
www.geosci-model-dev.net/10/1209/2017/
doi:10.5194/gmd-10-1209-2017

Geoscientific Model Development
Open Access



The Met Office HadGEM3-ES chemistry–climate model: evaluation of stratospheric dynamics and its impact on ozone

Table 1. Model simulations.

Name	Time period	Coupled ocean?	Nudging time-scale	Smoothing?
REF-C1	1960–2010	No	N/A	N/A
REF-C2	1960–2100	Yes	N/A	N/A
REF-C1SD-24 h	1980–2010	No	24 h	No
REF-C1SD-48 h	1980–2010	No	48 h	No
REF-C1SD-24 h, smoothed	1980–2010	No	24 h	Yes
REF-C1SD-48 h, smoothed	1980–2010	No	48 h	Yes
CCMVal-2 (UMUKCA-METO)	1960–2005	No	N/A	N/A