UNIVERSITY OF CAMBRIDGE

Cambridge Centre *for* Climate Science

Modelling chemistry in the atmosphere: A very brief introduction.

Alexander T. Archibald.

Department of Chemistry, University of Cambridge
& NCAS Climate

5th December 2022.

Talk outline.

- **Brief history of modelling:**

What do we mean by models?

Why do we model?

- **Building blocks of a model:**

Writing a chemical process in a mathematical framework.

Solving our mathematical problems.

- **Examples :**

0-D model studies.

- **Evaluation:**

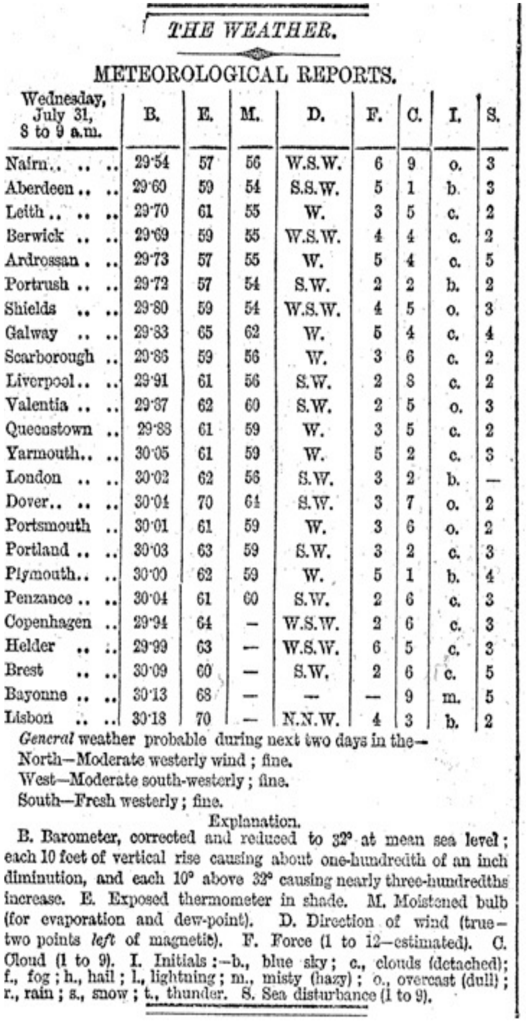
Process based/oriented model evaluation.

- **Q&A:**

What is a model?

Lot's of definitions:
"A model is as a simplified representation of a complex system that enables inference of the behavior of that system."

Modelling has been going on for a long time!



THE WEATHER.

METEOROLOGICAL REPORTS.

Wednesday, July 31, 8 to 9 a.m.	B.	E.	M.	D.	F.	C.	I.	S.
Nairn..	29·54	57	56	W.S.W.	6	9	o.	3
Aberdeen	29·60	59	54	S.S.W.	5	1	b.	3
Leith	29·70	61	55	W.	3	5	c.	2
Berwick	29·69	59	55	W.S.W.	4	4	c.	2
Ardrossan . ..	29·73	57	55	W.	5	4	c.	5
Portrush	29·72	57	54	S.W.	2	2	b.	2
Shields	29·80	59	54	W.S.W.	4	5	o.	3
Galway	29·83	65	62	W.	5	4	c.	4
Scarborough ..	20·86	59	56	W.	3	6	c.	2
Liverpool.. ..	29·91	61	56	S.W.	2	8	c.	2
Valentia	29·37	62	60	S.W.	2	5	o.	3
Queenstown ..	29·88	61	59	W.	3	5	c.	2
Yarmouth.. ..	30·05	61	59	W.	5	2	c.	3
London	30·02	62	56	S.W.	3	2	b.	—
Dover..	30·04	70	64	S.W.	3	7	o.	2
Portsmouth ..	30·01	61	59	W.	3	6	o.	2
Portland	30·03	63	59	S.W.	3	2	c.	3
Plymouth.. ..	30·00	62	59	W.	5	1	b.	4
Penzance	30·04	61	60	S.W.	2	6	c.	3
Copenhagen ..	29·94	64	—	W.S.W.	2	6	c.	3
Helder	29·99	63	—	W.S.W.	6	5	c.	3
Brest	30·09	60	—	S.W.	2	6	c.	5
Bayonne	30·13	68	—	—	—	9	m.	5
Lisbon	30·18	70	—	N.N.W.	4	3	b.	2

General weather probable during next two days in the—
North—Moderate westerly wind ; fine.
West—Moderate south-westerly ; fine.
South—Fresh westerly ; fine.

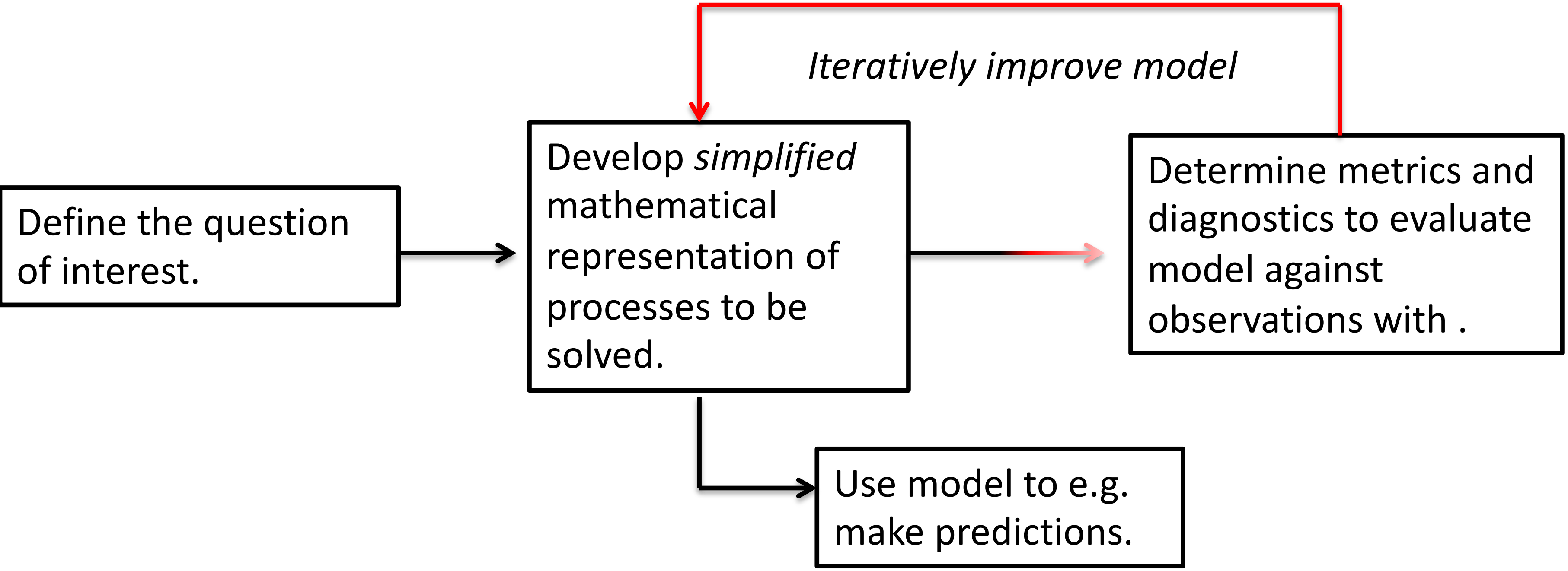
Explanation.

B. Barometer, corrected and reduced to 32° at mean sea level ; each 10 feet of vertical rise causing about one-hundredth of an inch diminution, and each 10° above 32° causing nearly three-hundredths increase. E. Exposed thermometer in shade. M. Moistened bulb (for evaporation and dew-point). D. Direction of wind (true—two points *left* of magnetic). F. Force (1 to 12—estimated). C. Cloud (1 to 9). I. Initials :—b., blue sky ; c., clouds (detached); f., fog ; h., hail ; l., lightning ; m., misty (hazy) ; o., overcast (dull) ; r., rain ; s., snow ; t., thunder. S. Sea disturbance (1 to 9).

1st August 1861, The Times

Building the model – on paper!

Most of the interesting problems you want to study will be impossible to solve exactly – need to develop a model to represent the system.



Define the question of interest.

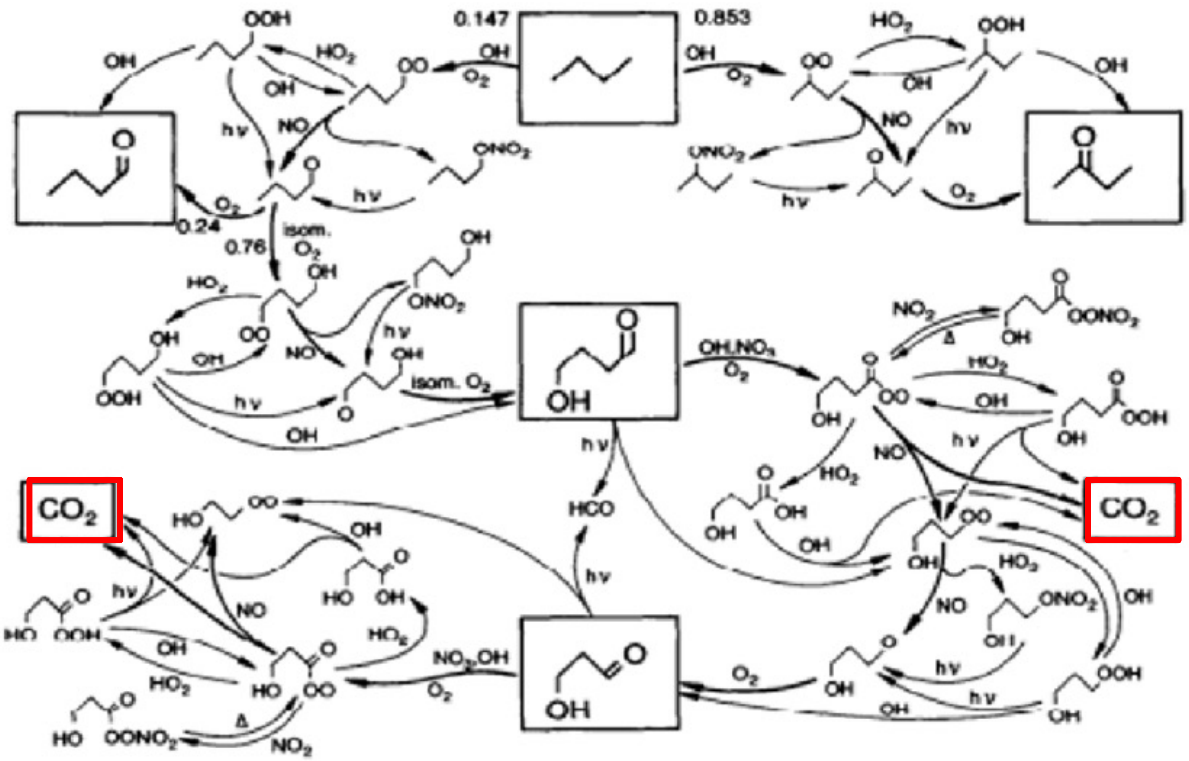
Develop *simplified* mathematical representation of processes to be solved.

Iteratively improve model

Determine metrics and diagnostics to evaluate model against observations with .

Use model to e.g. make predictions.

Chemistry in the atmosphere



Most chemically active molecules in the atmosphere arise from **emissions** (source) and the subsequent formation of **intermediate** species. The ultimate fate of emissions in the present-day atmosphere is **oxidation**. A **mechanism** (shown schematically above) represents this process.

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Chemistry in the atmosphere

$$A + B \rightarrow C + D$$

$$\frac{d[\mathrm{A}]}{dt} = -k[\mathrm{A}][\mathrm{B}] = -\frac{d[\mathrm{C}]}{dt}$$

The model we use is based on *chemical kinetics*. For every **reaction** in the **mechanism** we can write an overall **rate equation**. For each species (A, B, C etc) we can then write a **continuity equation** – an ODE which describes how the species concentration ([A], [B] etc) changes over time.

Chemistry in the atmosphere

$$A + B \rightarrow C + D$$

$$\frac{d[\mathrm{A}]}{dt} = -k[\mathrm{A}][\mathrm{B}] = L_A$$

$$\frac{d[\mathrm{C}]}{dt} = +k[\mathrm{A}][\mathrm{B}] = P_C$$

The overall rate equation can tell us if there is a **production** of a species or a **loss**.

Chemistry in the atmosphere

$$\mathrm{A + B \rightarrow C + D}$$

$$\frac{d[\mathrm{A}]}{dt} = -k[\mathrm{A}][\mathrm{B}]$$

$$k = A \times \exp\left(\frac{-E_A}{RT}\right)$$

The rate constant for a reaction is usually not constant! It usually depends on **temperature** and can also depend on pressure OR even the concentration of things like water vapour (see asad_bimol or asad_trimol).

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Chemistry in the atmosphere

A + B → C + D

$$\frac{d[\mathrm{A}]}{dt} = -k[\mathrm{A}][\mathrm{B}]$$

$$k = A \times \exp\left(\frac{-E_A}{RT}\right)$$

The rate constant is important as it measures how quickly a reaction will happen. If [B] is much greater than [A] then, k[B] will be roughly constant – in which case we can write:

National Centre for
Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Chemistry in the atmosphere

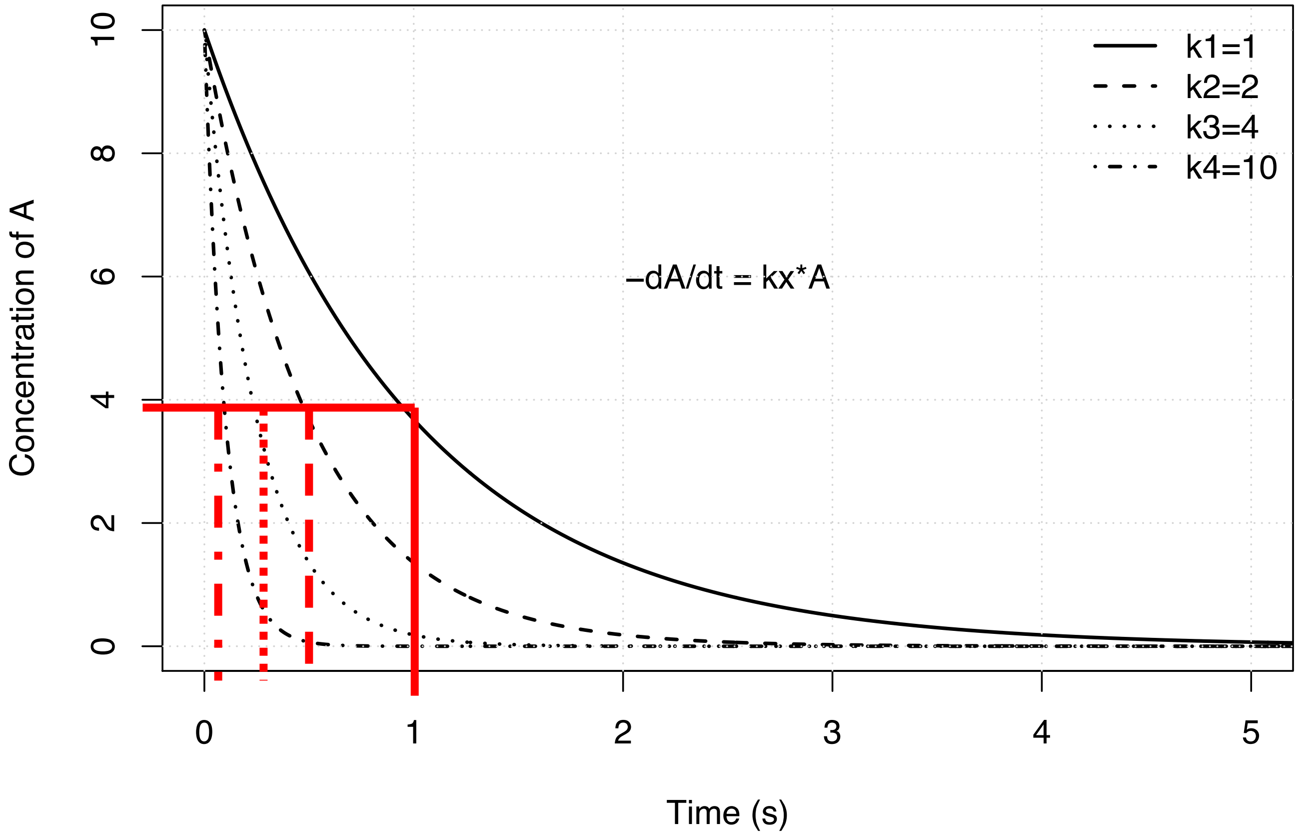
$$A \rightarrow C + D$$

$$\frac{d[\mathrm{A}]}{dt} = -k'[\mathrm{A}]$$

$$k' = A \times \exp\left(\frac{-E_A}{RT}\right) \times [\mathrm{B}]$$

k' is now a pseudo-first order rate constant. It has dimensions of 1/time and so 1/k' gives us a characteristic time constant τ, for the reaction.

Chemistry in the atmosphere



τ = 1 s
τ = 0.5 s
τ = 0.25 s
τ = 0.1 s

k’ is now a pseudo-first order rate constant. It has dimensions of 1/time and so 1/k’ gives us a characteristic time constant τ, for the reaction.

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Chemistry in the atmosphere

$$A \rightarrow C + D$$

$$\frac{d[\mathrm{A}]}{dt} = -J[\mathrm{A}]$$

$$J_{(i)} = \int_0^\infty \sigma_{(\lambda,i)} \varphi_{(\lambda,i)} F_{(\lambda)}\, d\lambda$$

$J_{(i)}$ is the photolysis coefficient (rate constant) for the reaction. It depends on the flux of photons (F), the quantum yield (φ) and the absorption cross-section of the molecule (σ).

The Continuity Equation

$$\frac{\partial[X]}{\partial t} = \boxed{E_X + P_X}\,\boxed{- L_X - D_X - \nabla \bullet (U[X])}$$

Local change in [*X*] with time.

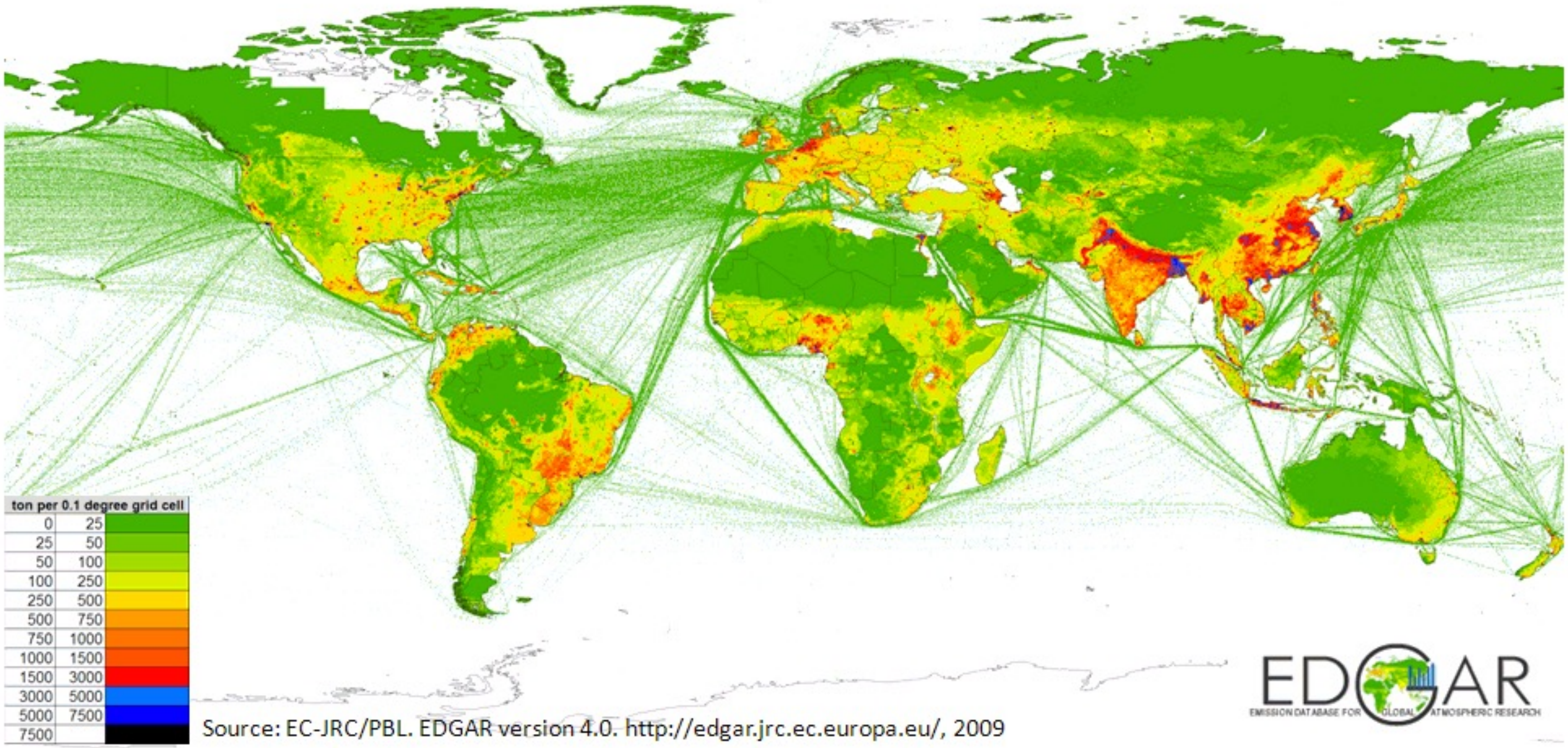
Emissions and **Chemical Production**

Chemical Loss, Deposition and transport (flux out)

This equation is at the core of all the problems we will want to study regarding chemistry in the atmosphere. The *difficulty* lies in parameterizing the individual terms.

Emissions

$$\frac{\partial[X]}{\partial t} = \boxed{E_X + P_X}\,\boxed{- L_X - D_X - \nabla\bullet(U[X])}$$

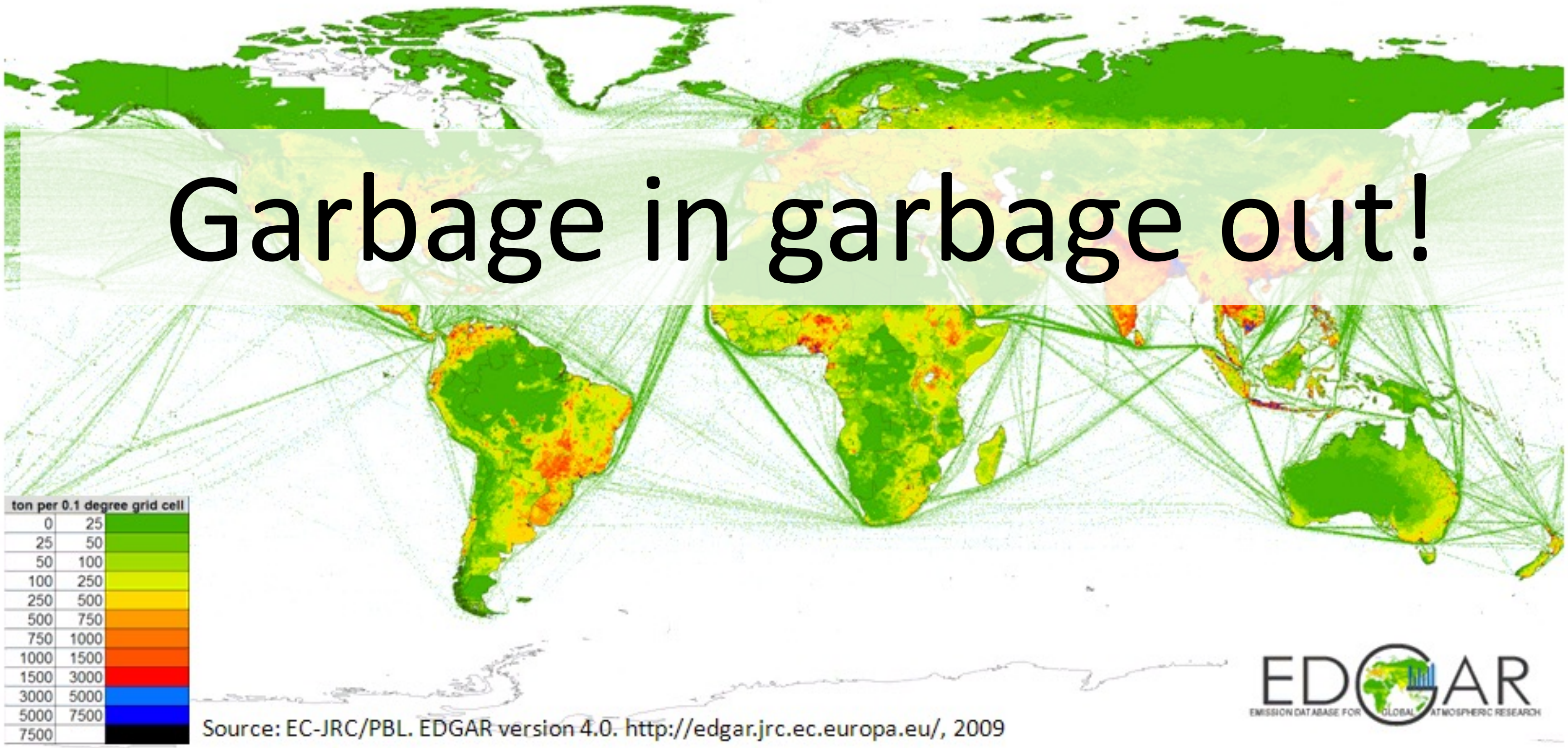


Source: EC-JRC/PBL. EDGAR version 4.0. http://edgar.jrc.ec.europa.eu/, 2009

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Emissions

$$\frac{\partial[X]}{\partial t} = \boxed{E_X + P_X} \boxed{- L_X - D_X - \nabla \bullet (U[X])}$$



Garbage in garbage out!

Source: EC-JRC/PBL. EDGAR version 4.0. http://edgar.jrc.ec.europa.eu/, 2009

Emissions

$$\frac{\partial[X]}{\partial} = \boxed{E_X + P_X} \boxed{- L_X - D_X - \nabla\bullet(U[X])}$$

JOINT RESEARCH CENTRE

EDGAR - Emission Database for Global Atmospheric Research

Welcome to the homepage of EDGAR

The Emissions Database for Global Atmospheric Research (EDGAR) provides global past and present day anthropogenic emissions of greenhouse gases and air pollutants by country and on spatial grid. The current development of EDGAR is a joint project of the European Commission JRC Joint Research Centre and the Netherlands Environmental Assessment Agency (PBL).

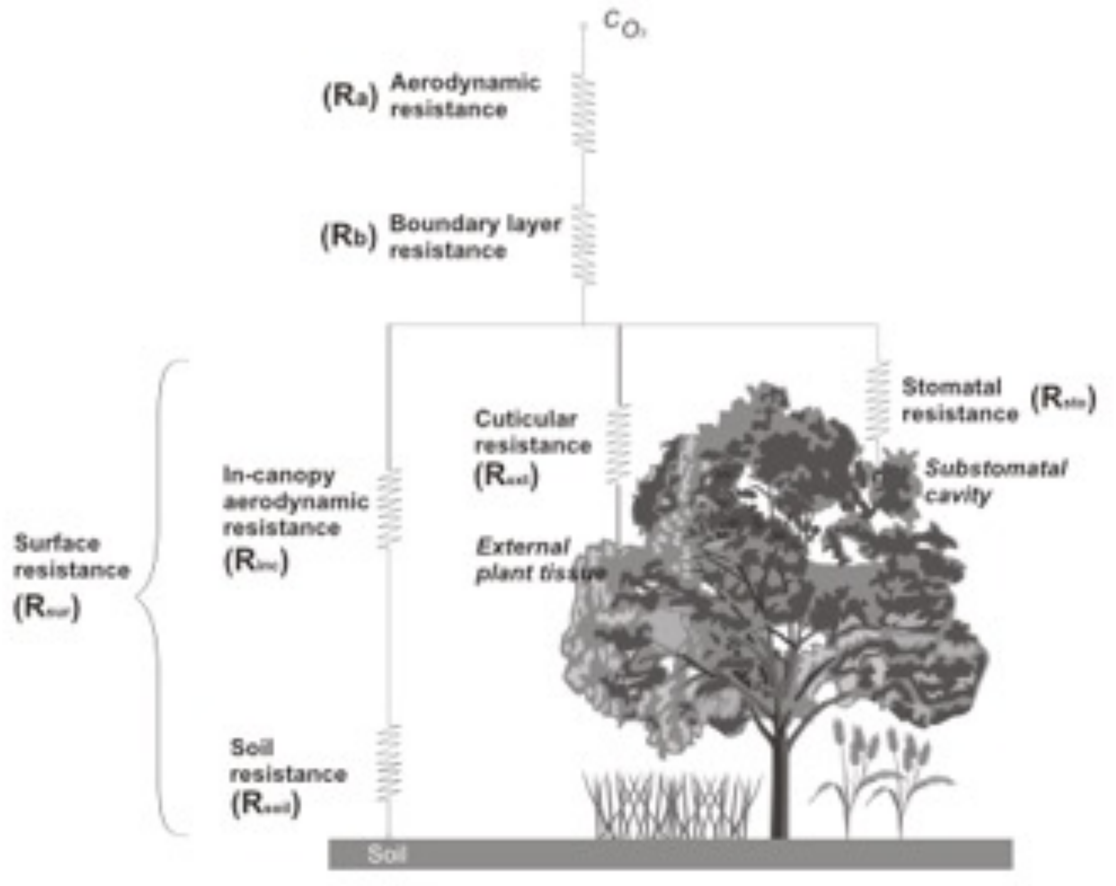
What's new?

16-12-2014
Further slowdown in the increase in global CO2 emissions in 2013.
JRC report 93171 / PBL report 1490; ISBN 978-94-91506-87-1 December 2014

2013 saw a further slowdown in the increase in global CO2 emissions from fossil fuel use and cement production that started in 2012. The emissions grew with only 0.7 billion tonnes (Gt) CO2 in 2013 to the new record of 35.3 Gt CO2. The global CO2 increased at a notably slower rate (2%) than on average in the last ten years (3.8% per year since 2003, excluding the credit crunch years). This signals a partial decoupling of global emissions and economic growth, which reflects mainly the lower emissions growth rate of China and which is also in line with a growing share of the service sector at the expense of more energy-intensive industries for the

Deposition

$$\frac{\partial[X]}{\partial t} = \boxed{E_X + P_X} \boxed{- L_X - D_X - \nabla \bullet (U[X])}$$



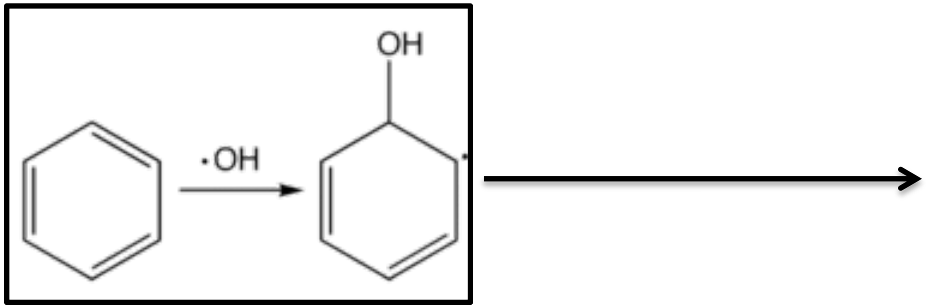
Deposition modelled using a resistors in series approach.

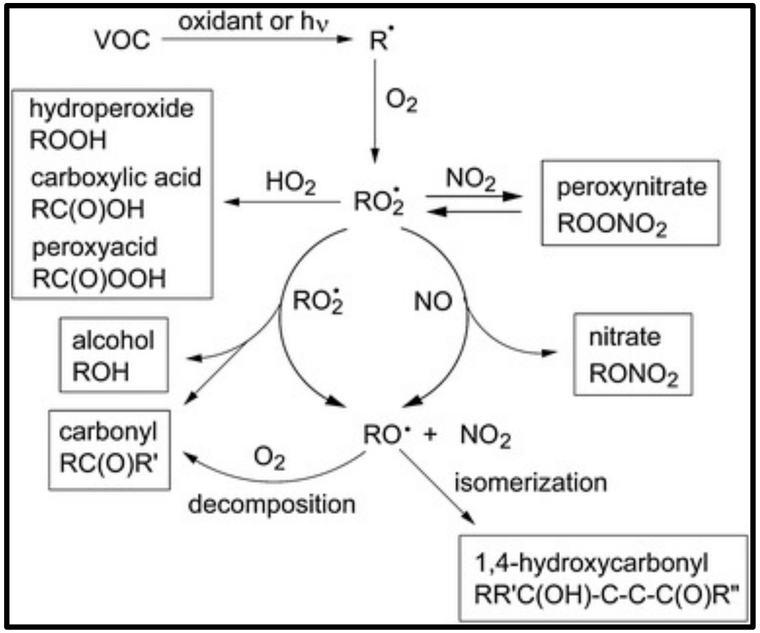
Ra in cm s^{-1}

Deposition flux = 1/Ra + 1/Rb + 1/Rc + …

Reactions (Production and Loss)

$$\frac{\partial[X]}{\partial t} = \boxed{E_X + P_X} \boxed{- L_X - D_X - \nabla \bullet (U[X])}$$



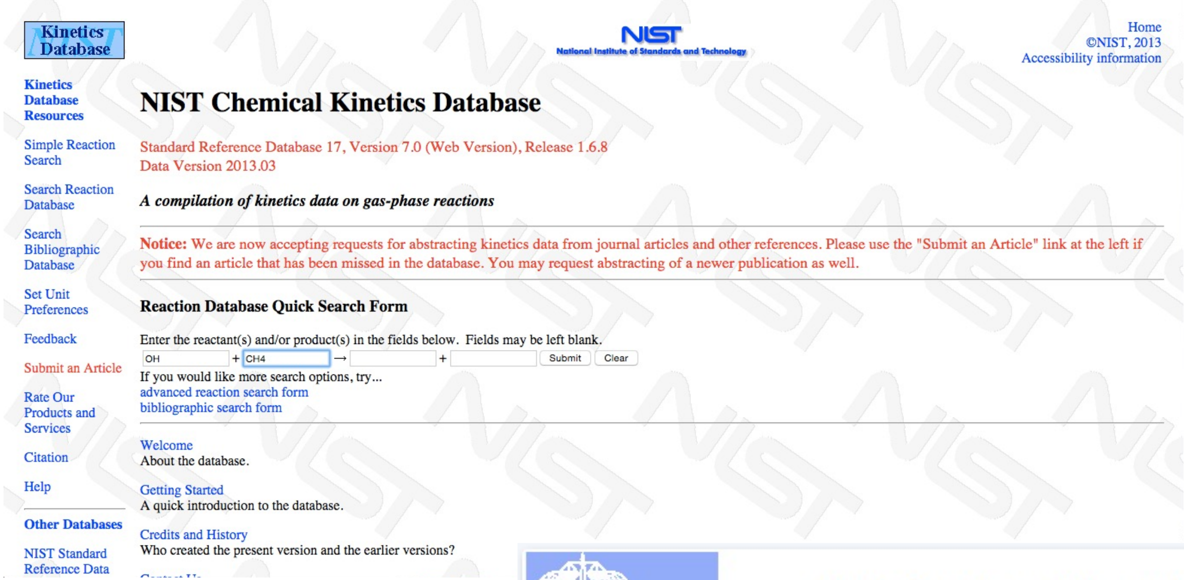


Need to build up a "picture" of the chemistry of interest.

Mechanism development is just another aspect of modelling – parameterizing the millions of possible reactions to a solvable subset.

Reactions (Production and Loss)

Kinetic data bases:



Kinetics Database

NIST
National Institute of Standards and Technology

Home
©NIST, 2013
Accessibility information

Kinetics Database Resources

Simple Reaction Search

Search Reaction Database

Search Bibliographic Database

Set Unit Preferences

Feedback

Submit an Article

Rate Our Products and Services

Citation

Help

Other Databases

NIST Standard Reference Data

NIST Chemical Kinetics Database

Standard Reference Database 17, Version 7.0 (Web Version), Release 1.6.8
Data Version 2013.03

A compilation of kinetics data on gas-phase reactions

Notice: We are now accepting requests for abstracting kinetics data from journal articles and other references. Please use the "Submit an Article" link at the left if you find an article that has been missed in the database. You may request abstracting of a newer publication as well.

Reaction Database Quick Search Form

Enter the reactant(s) and/or product(s) in the fields below. Fields may be left blank.

OH + CH4 → + Submit Clear

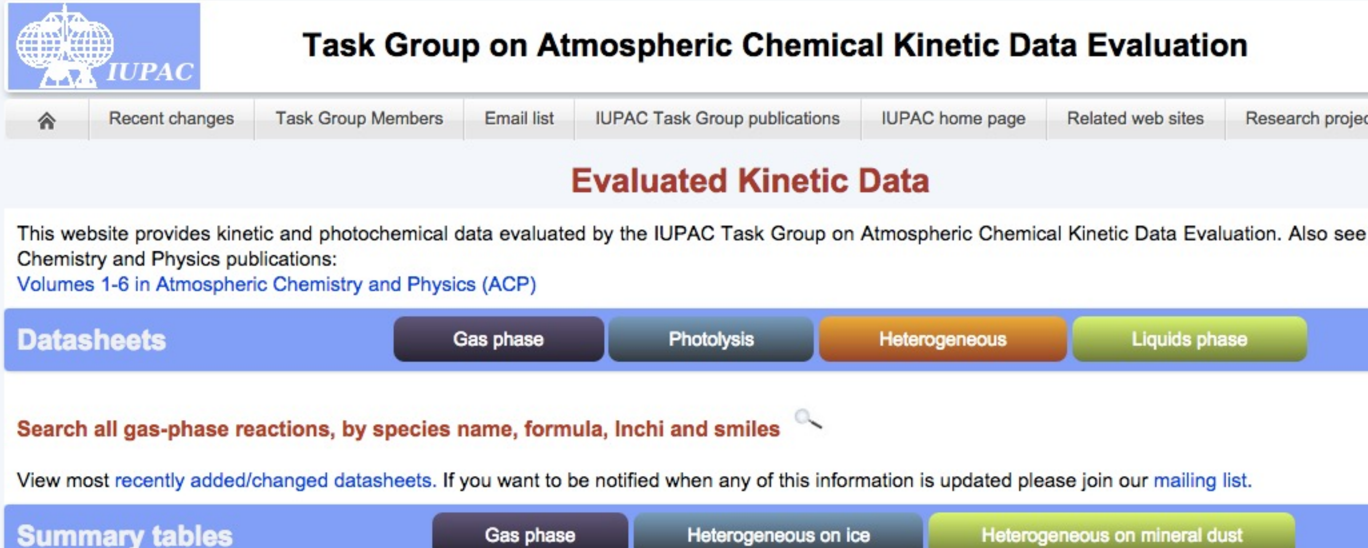
If you would like more search options, try...
advanced reaction search form
bibliographic search form

Welcome
About the database.

Getting Started
A quick introduction to the database.

Credits and History
Who created the present version and the earlier versions?

Provide archives of kinetic data and provide *evaluations* of kinetic data.



IUPAC

Task Group on Atmospheric Chemical Kinetic Data Evaluation

Evaluated Kinetic Data

This website provides kinetic and photochemical data evaluated by the IUPAC Task Group on Atmospheric Chemical Kinetic Data Evaluation. Also see Chemistry and Physics publications:
Volumes 1-6 in Atmospheric Chemistry and Physics (ACP)

Search all gas-phase reactions, by species name, formula, Inchi and smiles

View most recently added/changed datasheets. If you want to be notified when any of this information is updated please join our mailing list.

Reactions (Production and Loss)

$$
\begin{array}{llll}
\text{(i)} & \mathrm{CH_4} + \mathrm{OH} \Rightarrow \cdots \Rightarrow \mathrm{CO} & R_5 = k_5[\mathrm{CH_4}][\mathrm{OH}] & k_5 = 1.266 \times 10^{-7}\,\mathrm{s}^{-1}\,\mathrm{ppt}^{-1} \\
\text{(ii)} & \mathrm{CO} + \mathrm{OH} \Rightarrow \cdots & R_6 = k_6[\mathrm{CO}][\mathrm{OH}] & k_6 = 5.08 \times 10^{-6}\,\mathrm{s}^{-1}\,\mathrm{ppt}^{-1} \\
\text{(iii)} & \mathrm{OH} + \mathrm{X} \Rightarrow \cdots & R_7 = k_7[\mathrm{X}][\mathrm{OH}] & k_7[\mathrm{X}] = 1.062\,\mathrm{s}^{-1},
\end{array}
$$

Reactions (Production and Loss)

$$\begin{array}{llll}
\text{(i)} & CH_4 + OH \Rightarrow \cdots \Rightarrow CO & R_5 = k_5[CH_4][OH] & k_5 = 1.266 \times 10^{-7}\,\mathrm{s}^{-1}\,\mathrm{ppt}^{-1} \\
\text{(ii)} & CO + OH \Rightarrow \cdots & R_6 = k_6[CO][OH] & k_6 = 5.08 \times 10^{-6}\,\mathrm{s}^{-1}\,\mathrm{ppt}^{-1} \\
\text{(iii)} & OH + X \Rightarrow \cdots & R_7 = k_7[X][OH] & k_7[X] = 1.062\,\mathrm{s}^{-1},
\end{array}$$

$$\frac{\partial[X]}{\partial t} = E_X + P_X - L_X - D_X - \nabla\bullet(U[X])$$

$$\frac{\mathrm{d}[CH_4]}{\mathrm{d}t} = S_{CH_4} - R_5$$

$$\frac{\mathrm{d}[CO]}{\mathrm{d}t} = S_{CO} + R_5 - R_6$$

$$\frac{\mathrm{d}[OH]}{\mathrm{d}t} = S_{OH} - R_5 - R_6 - R_7.$$

National Centre for
Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Reactions (Production and Loss)

$$
\begin{array}{llll}
\text{(i)} & \mathrm{CH_4} + \mathrm{OH} \Rightarrow \cdots \Rightarrow \mathrm{CO} & R_5 = k_5[\mathrm{CH_4}][\mathrm{OH}] & k_5 = 1.266 \times 10^{-7}\ \mathrm{s}^{-1}\ \mathrm{ppt}^{-1} \\
\text{(ii)} & \mathrm{CO} + \mathrm{OH} \Rightarrow \cdots & R_6 = k_6[\mathrm{CO}][\mathrm{OH}] & k_6 = 5.08 \times 10^{-6}\ \mathrm{s}^{-1}\ \mathrm{ppt}^{-1} \\
\text{(iii)} & \mathrm{OH} + \mathrm{X} \Rightarrow \cdots & R_7 = k_7[\mathrm{X}][\mathrm{OH}] & k_7[\mathrm{X}] = 1.062\ \mathrm{s}^{-1},
\end{array}
$$

$$
\frac{\mathrm{d}[\mathrm{CH_4}]}{\mathrm{d}t} = S_{\mathrm{CH_4}} - R_5
$$

$$
\frac{\mathrm{d}[\mathrm{CO}]}{\mathrm{d}t} = S_{\mathrm{CO}} + R_5 - R_6
$$

$$
\frac{\mathrm{d}[\mathrm{OH}]}{\mathrm{d}t} = S_{\mathrm{OH}} - R_5 - R_6 - R_7.
$$

Jacobian

$$
J = \begin{array}{ccc}
\dfrac{\partial\left(\dfrac{\mathrm{d}[\mathrm{CH_4}]}{\mathrm{d}t}\right)}{\partial[\mathrm{CH_4}]} & \dfrac{\partial\left(\dfrac{\mathrm{d}[\mathrm{CH_4}]}{\mathrm{d}t}\right)}{\partial[\mathrm{CO}]} & \dfrac{\partial\left(\dfrac{\mathrm{d}[\mathrm{CH_4}]}{\mathrm{d}t}\right)}{\partial[\mathrm{OH}]} \\
\dfrac{\partial\left(\dfrac{\mathrm{d}[\mathrm{CO}]}{\mathrm{d}t}\right)}{\partial[\mathrm{CH_4}]} & \dfrac{\partial\left(\dfrac{\mathrm{d}[\mathrm{CO}]}{\mathrm{d}t}\right)}{\partial[\mathrm{CO}]} & \dfrac{\partial\left(\dfrac{\mathrm{d}[\mathrm{CO}]}{\mathrm{d}t}\right)}{\partial[\mathrm{OH}]} \\
\dfrac{\partial\left(\dfrac{\mathrm{d}[\mathrm{OH}]}{\mathrm{d}t}\right)}{\partial[\mathrm{CH_4}]} & \dfrac{\partial\left(\dfrac{\mathrm{d}[\mathrm{OH}]}{\mathrm{d}t}\right)}{\partial[\mathrm{CO}]} & \dfrac{\partial\left(\dfrac{\mathrm{d}[\mathrm{OH}]}{\mathrm{d}t}\right)}{\partial[\mathrm{OH}]}
\end{array}
$$

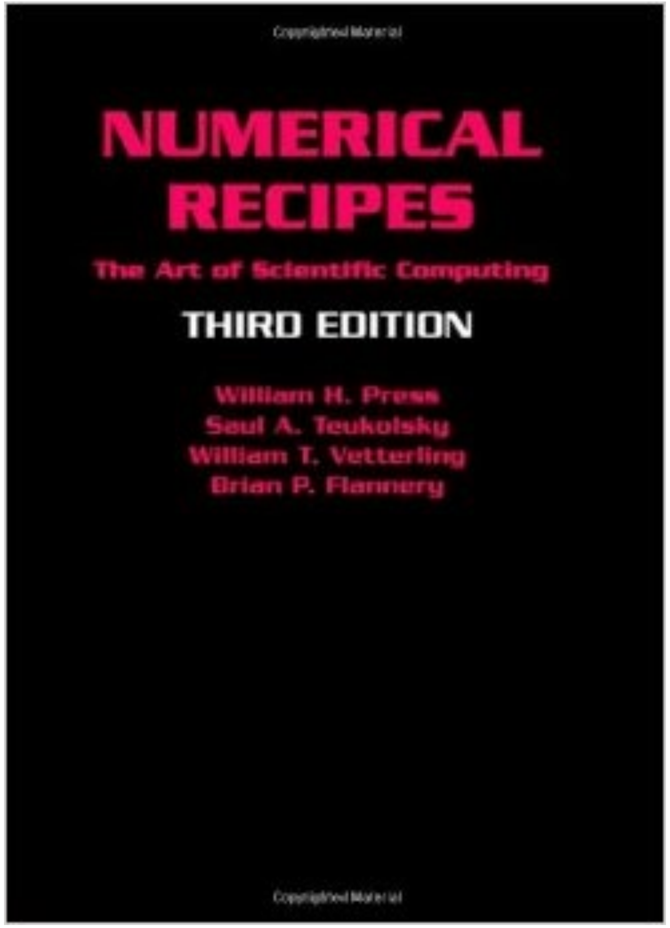
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Methane oxidation is more complex

Atmospheric oxidation of CH_4	species lifetime
$\mathbf{CH_4} + OH \longrightarrow CH_3 + H_2O$	$\tau \approx$ 10 yrs
$CH_3 + O_2 \longrightarrow CH_3O_2$	$\tau \approx$ 1 ms
$\mathbf{CH_3O_2} + NO \longrightarrow CH_3O + NO_2$	$\tau \approx$ 100 s
$CH_3O + O_2 \longrightarrow HCHO + HO_2$	$\tau \approx$ 1 s
$\mathbf{HCHO} + OH \longrightarrow CO + H + H_2O$	$\tau \approx$ 1 day
$\mathbf{HCHO} + h\nu \longrightarrow CO + 2H$	
$\mathbf{CO} + OH \longrightarrow CO_2 + H$	$\tau \approx$ 3 months
$H + O_2 \longrightarrow HO_2$	$\tau \approx$ 1 ms

Drop the very fast reactions and very short-lived species, replacing the latter with their ultimate oxidation products (e.g., HO_2 for H)

Reactions (Production and Loss)



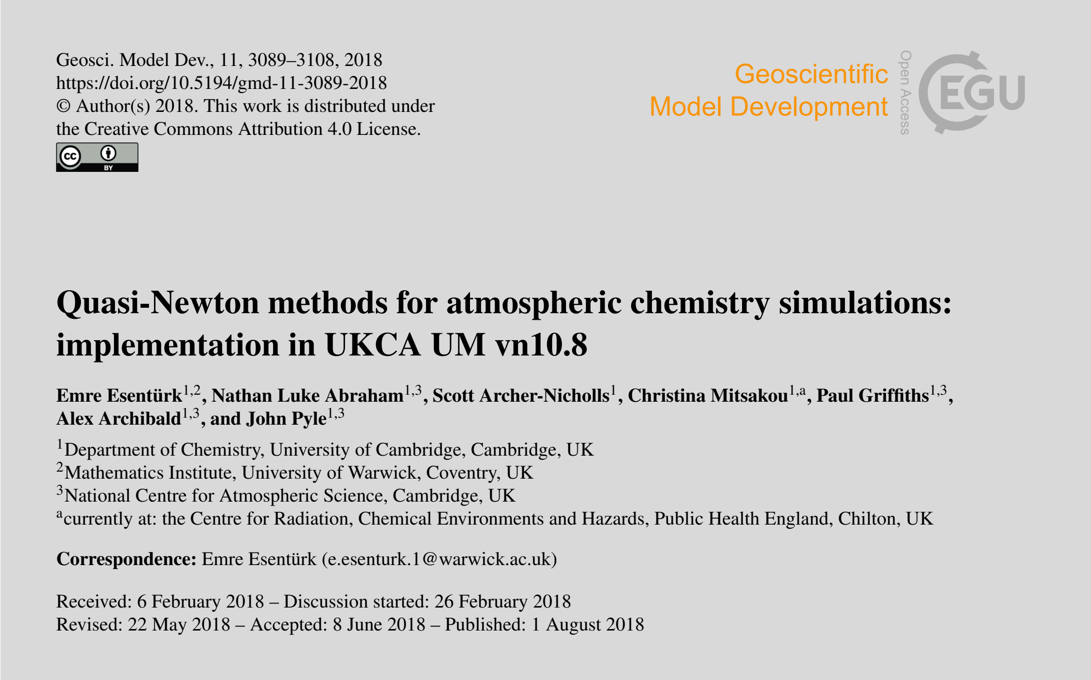
Copyrighted Material

NUMERICAL
RECIPES
The Art of Scientific Computing
THIRD EDITION

William H. Press
Saul A. Teukolsky
William T. Vetterling
Brian P. Flannery

Copyrighted Material

Let ***ASAD*** do the hard work for you!!



Geosci. Model Dev., 11, 3089–3108, 2018
https://doi.org/10.5194/gmd-11-3089-2018
© Author(s) 2018. This work is distributed under
the Creative Commons Attribution 4.0 License.

Geoscientific
Model Development
Open Access
EGU

Quasi-Newton methods for atmospheric chemistry simulations: implementation in UKCA UM vn10.8

Emre Esentürk[1,2], **Nathan Luke Abraham**[1,3], **Scott Archer-Nicholls**[1], **Christina Mitsakou**[1,a], **Paul Griffiths**[1,3], **Alex Archibald**[1,3], **and John Pyle**[1,3]

[1]Department of Chemistry, University of Cambridge, Cambridge, UK
[2]Mathematics Institute, University of Warwick, Coventry, UK
[3]National Centre for Atmospheric Science, Cambridge, UK
[a]currently at: the Centre for Radiation, Chemical Environments and Hazards, Public Health England, Chilton, UK

Correspondence: Emre Esentürk (e.esenturk.1@warwick.ac.uk)

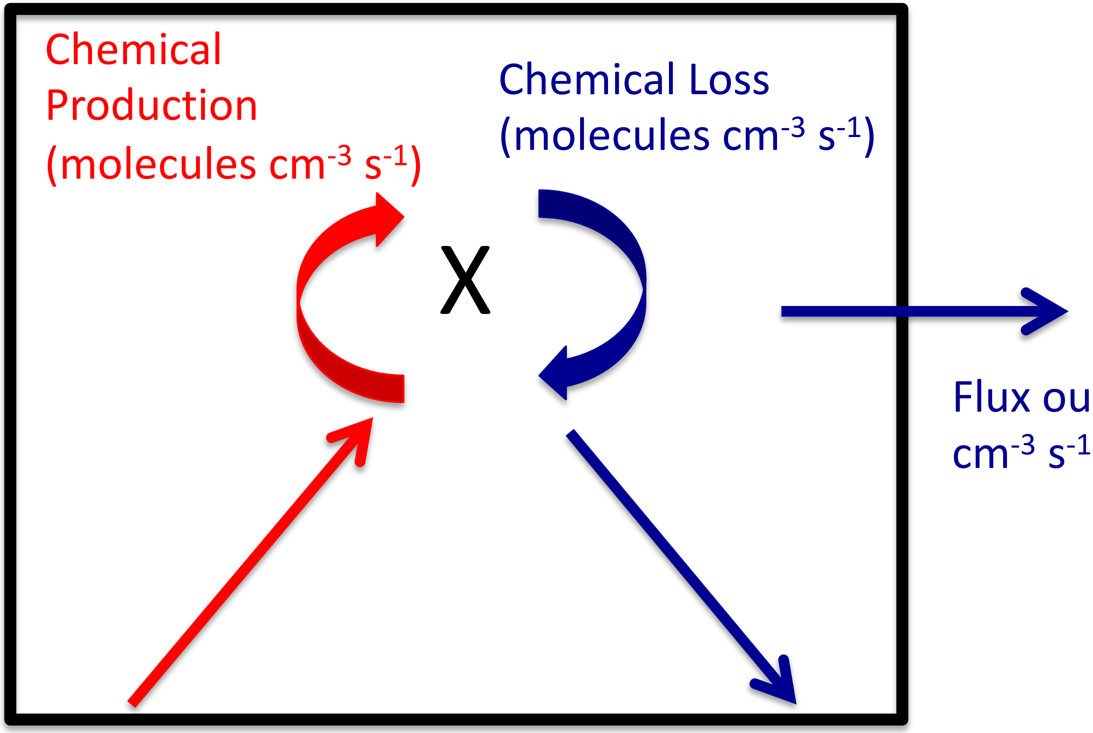
Received: 6 February 2018 – Discussion started: 26 February 2018
Revised: 22 May 2018 – Accepted: 8 June 2018 – Published: 1 August 2018

Model types: Example 1 – the box

The 0-D box model – possibly the best way to get into modelling atmospheric chemistry

$$\frac{\partial[X]}{\partial t} = \boxed{E_X + P_X} \boxed{- L_X - D_X - \nabla \bullet (U[X])}$$

When running a box model it's common to think of it as a well mixed parcel of air.
It's common to assume a fixed **temperature** and a fixed length scale (e.g. the **boundary layer height** – **NB** needed for converting emission fluxes and deposition rates).



Chemical Production (molecules cm^{-3} s^{-1})

Chemical Loss (molecules cm^{-3} s^{-1})

X

Flux out (molecules cm^{-3} s^{-1})

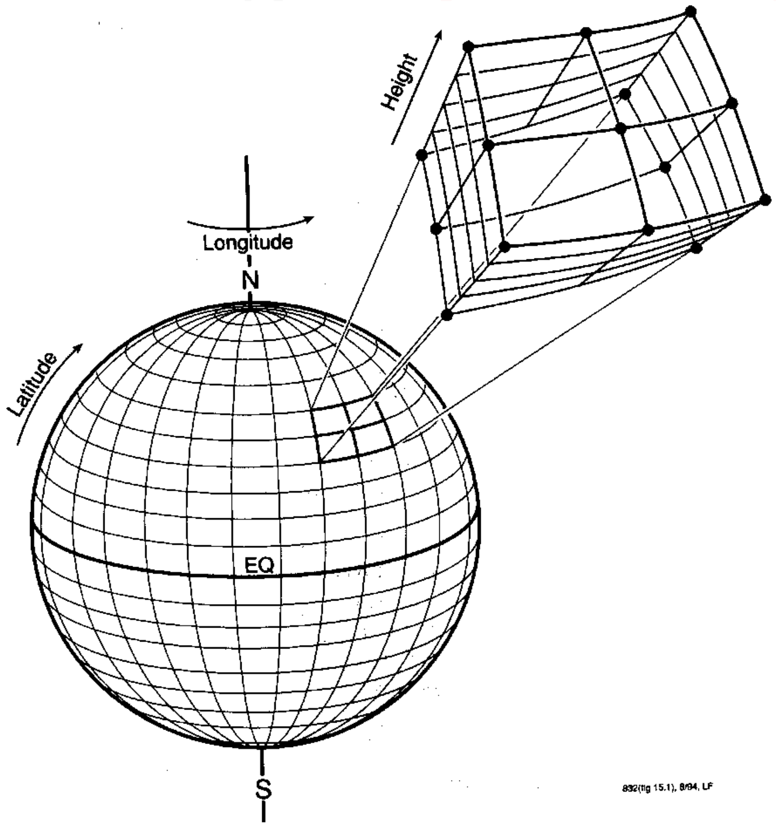
Emissions (molecules cm^{-2} s^{-1}) / "height of box" cm

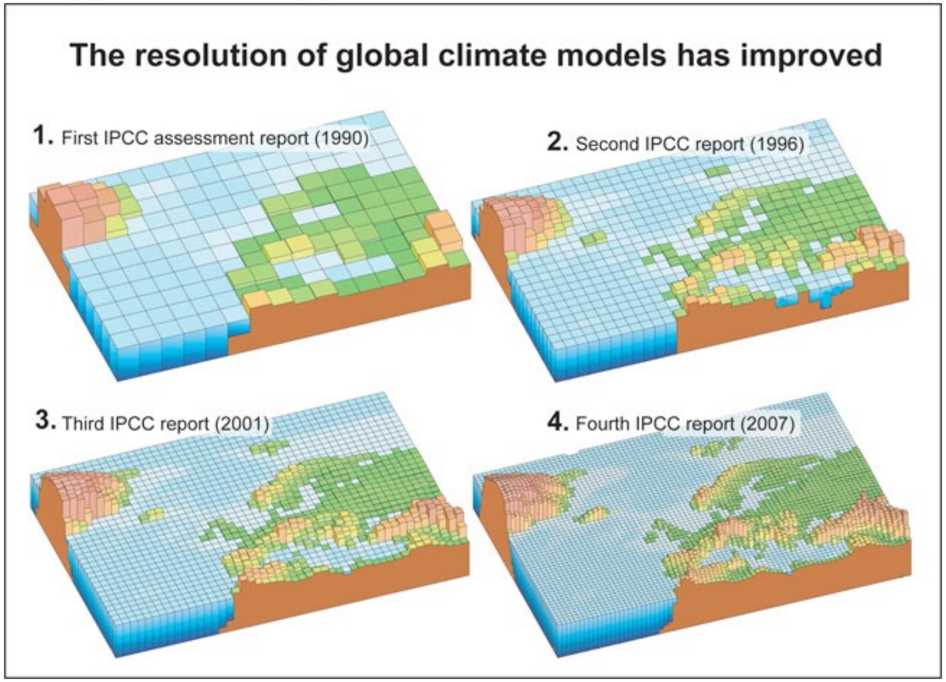
Deposition (cm s^{-1}) / "height of box" cm

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Model types: Example 2 – many boxes

The 3-D *Eulerian* model (e.g. UM-UKCA) – possibly the most widely used modelling tool. Think of it as ~ 10^6 box models!



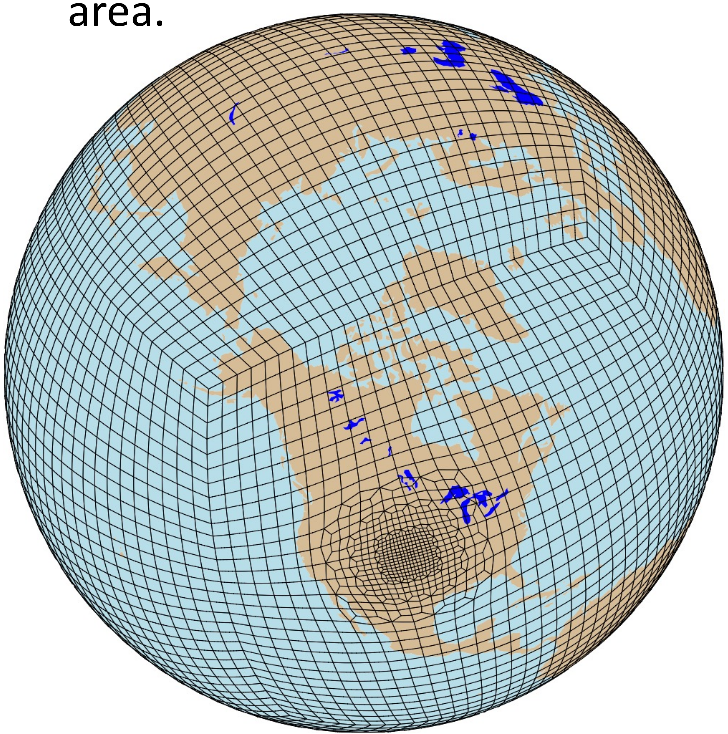


Resolution is a big problem. If you think about chemistry, it's highly dependent on concentration gradients!

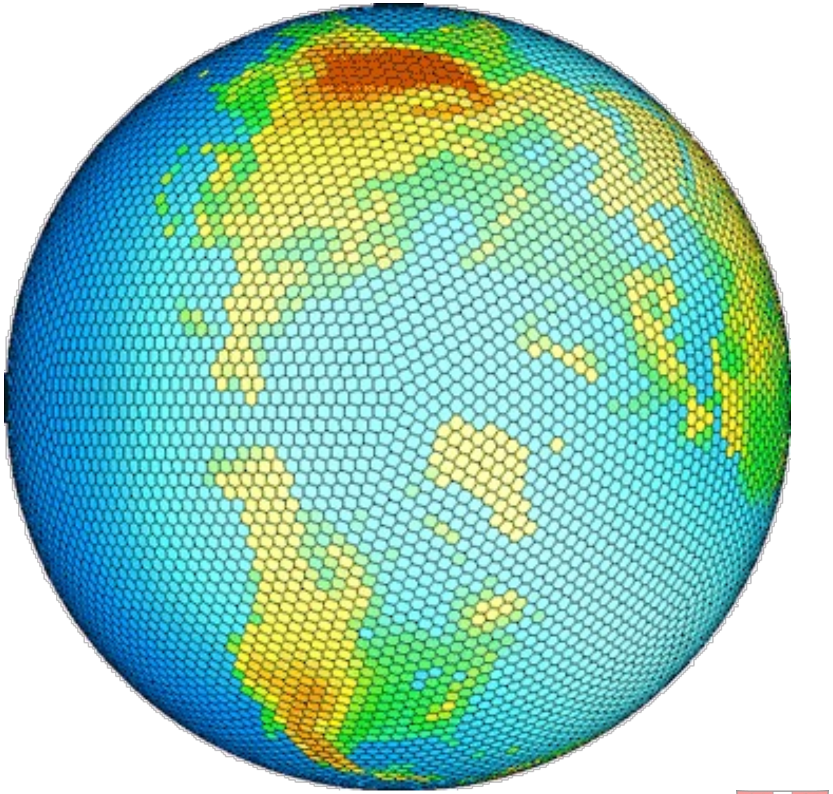
Model types: Example 2 – many boxes

The 3-D *Eulerian* model (e.g. UM-UKCA) – possibly the most widely used modelling tool. Think of it as ~ 10^6 box models!

Cubed sphere grid with adaptive mesh – nested higher resolution area.

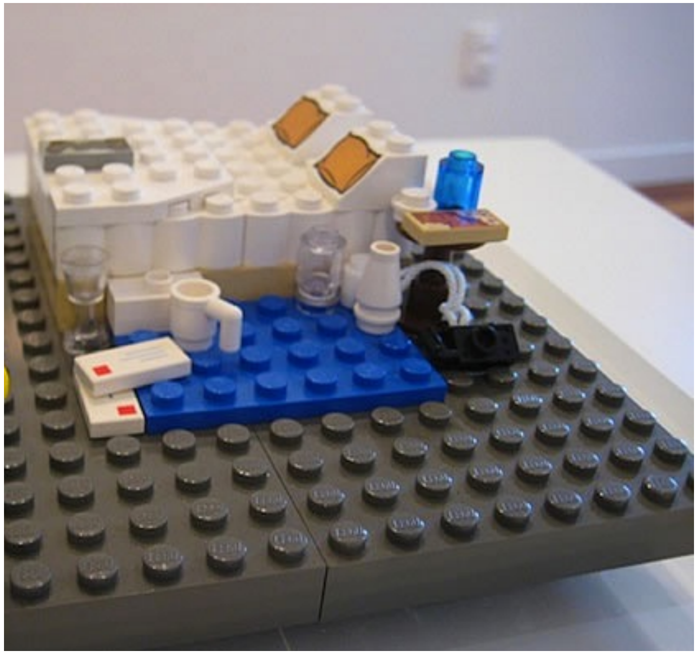


Icosahderal grid – removes pole singularities.



Model Evaluation:

Comparing models and reality.



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

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.

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.

Approaches to evaluation:

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

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

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

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

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

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

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:

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

the final formulation for the MSE components reads as follows:

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

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:

$$\begin{aligned}\mathrm{MSE} = &(\overline{\mathrm{mod}} - \overline{\mathrm{obs}})^2 + (\sigma_{\mathrm{mod}} - \sigma_{\mathrm{obs}})^2 \\ &+ 2(1-r)\sigma_{\mathrm{mod}}\sigma_{\mathrm{obs}}.\end{aligned} \tag{6}$$

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

Approaches to evaluation:

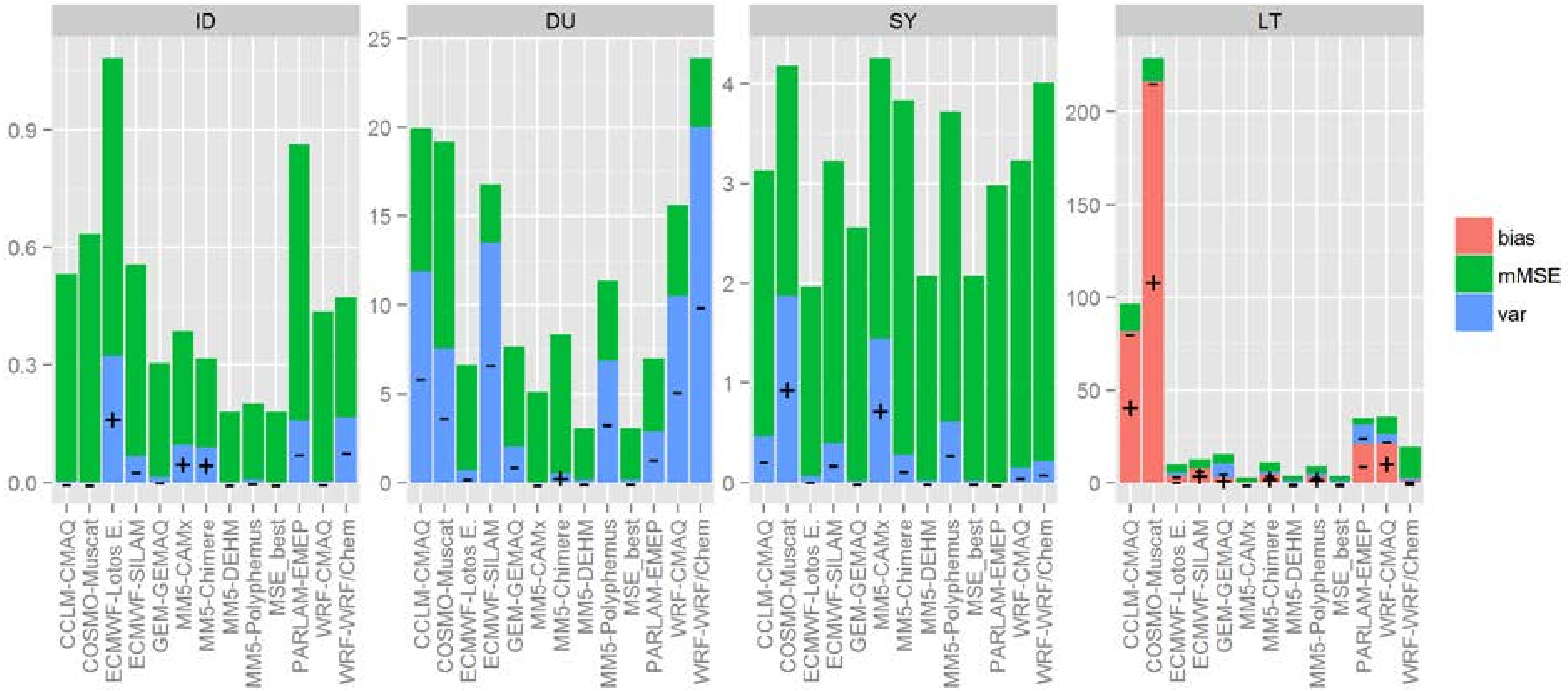
mMSE is the minimum achievable Mean Square Error:
mMSE = $\sigma^2_{obs}(1-r^2)$

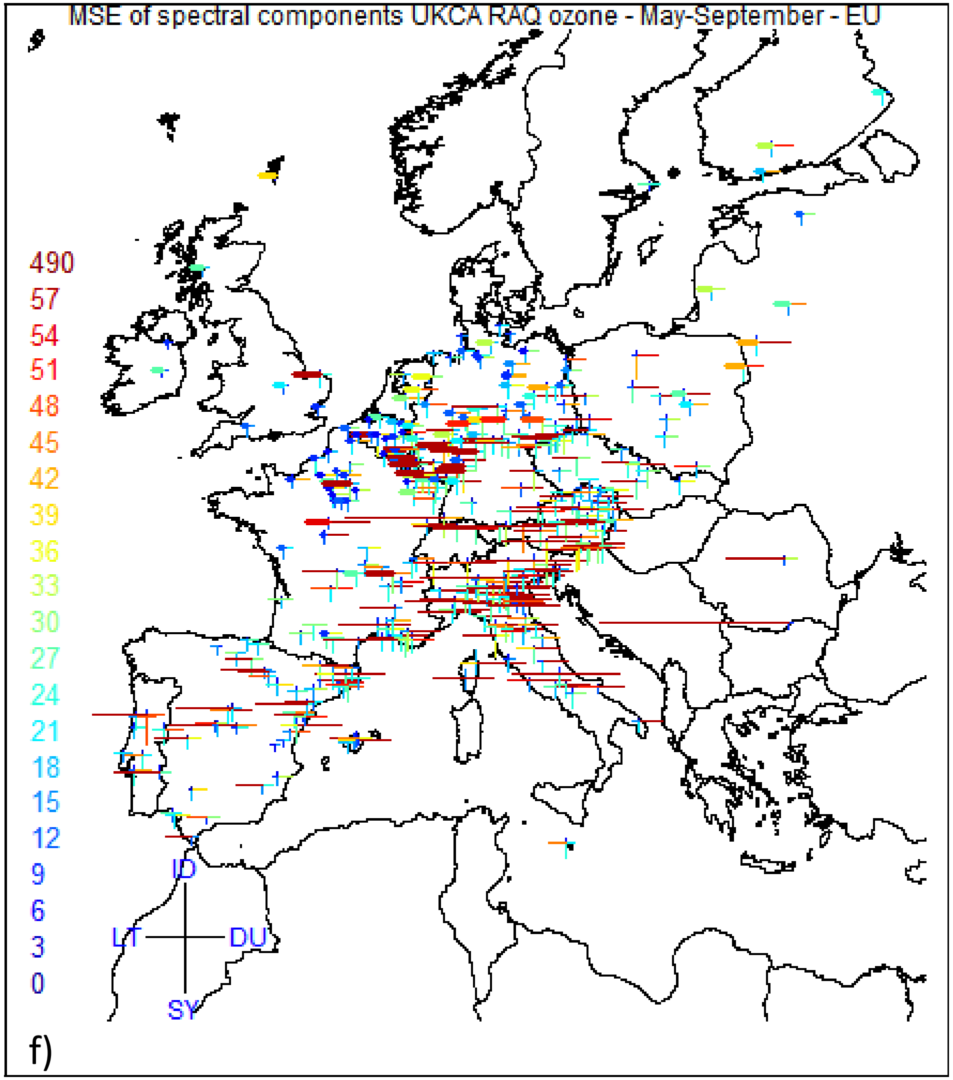
Solazzo and Galmarini suggest:
MSE = $(\langle mod \rangle - \langle obs \rangle)^2 + (\sigma_{mod} - r\sigma_{obs})^2$ + mMSE

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




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)

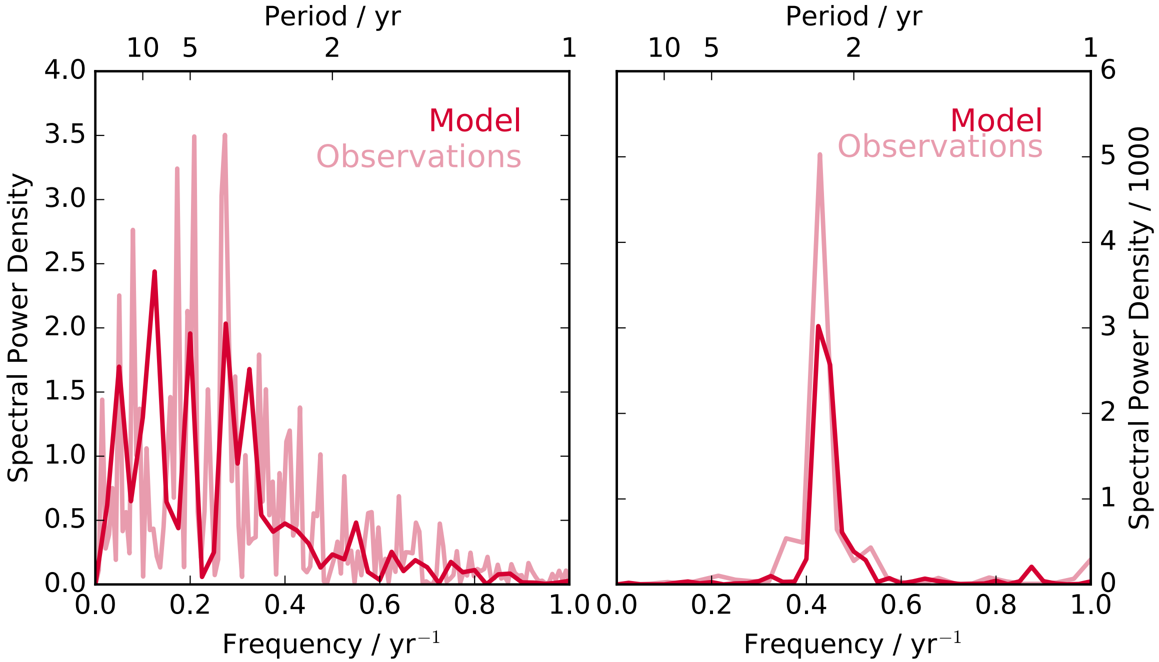
Approaches to evaluation:

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.

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.



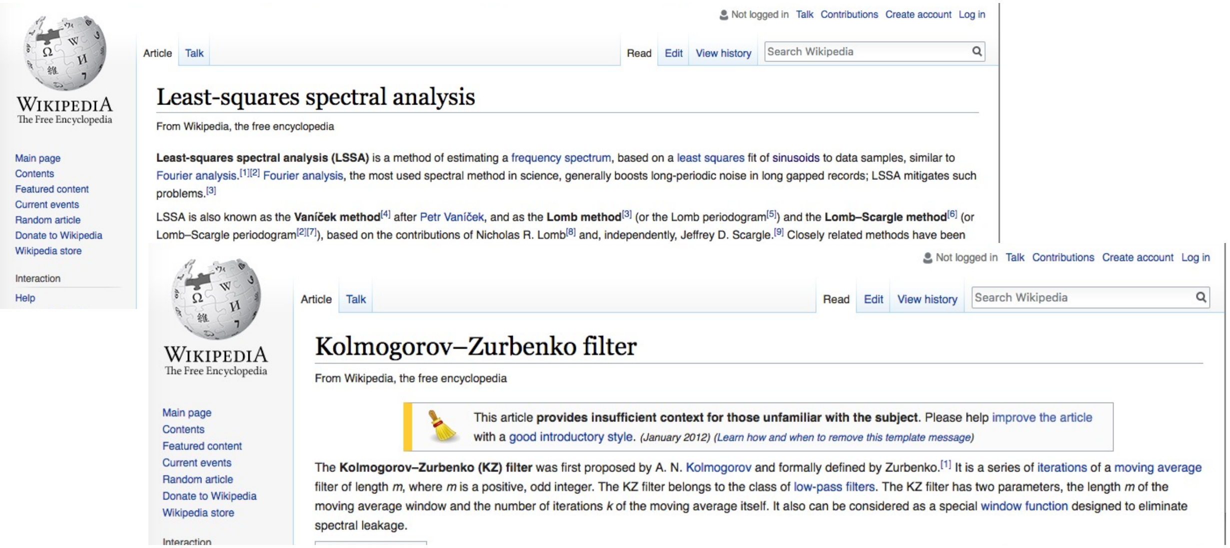
Courtesy of David Wade

New approaches to evaluation:

Spectral decomposition of modelled and observed time series

$$\mathrm{O_3} = \mathrm{LT(O_3)} + \mathrm{SY(O_3)} + \mathrm{DU(O_3)} + \mathrm{ID(O_3)}$$

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



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

Kolmogorov–Zurbenko filter

From Wikipedia, the free encyclopedia

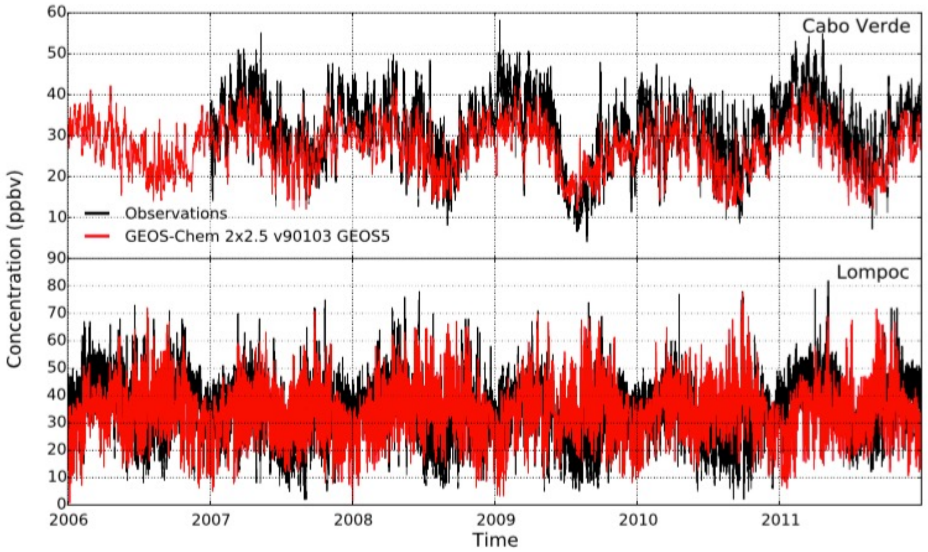
This article **provides insufficient context for those unfamiliar with the subject**. Please help improve the article with a good introductory style. *(January 2012) (Learn how and when to remove this template message)*

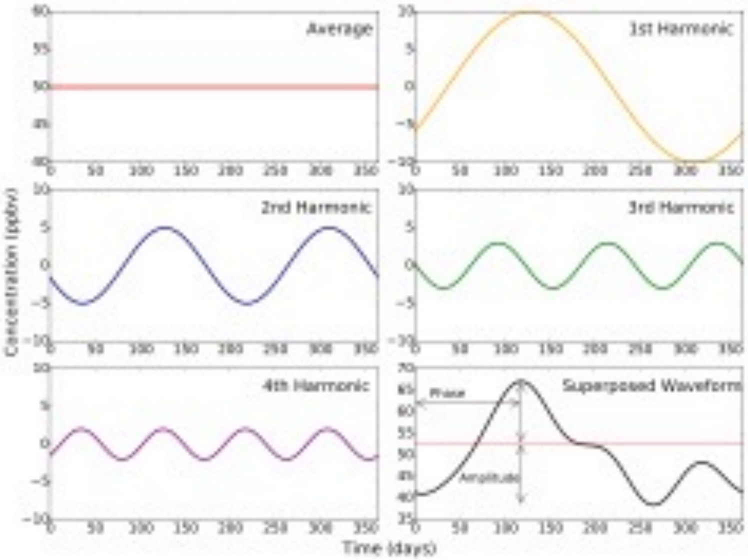
The **Kolmogorov–Zurbenko (KZ) filter** was first proposed by A. N. Kolmogorov and formally defined by Zurbenko.[1] It is a series of iterations of a moving average filter of length *m*, where *m* is a positive, odd integer. The KZ filter belongs to the class of low-pass filters. The KZ filter has two 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.

Approaches to evaluation:

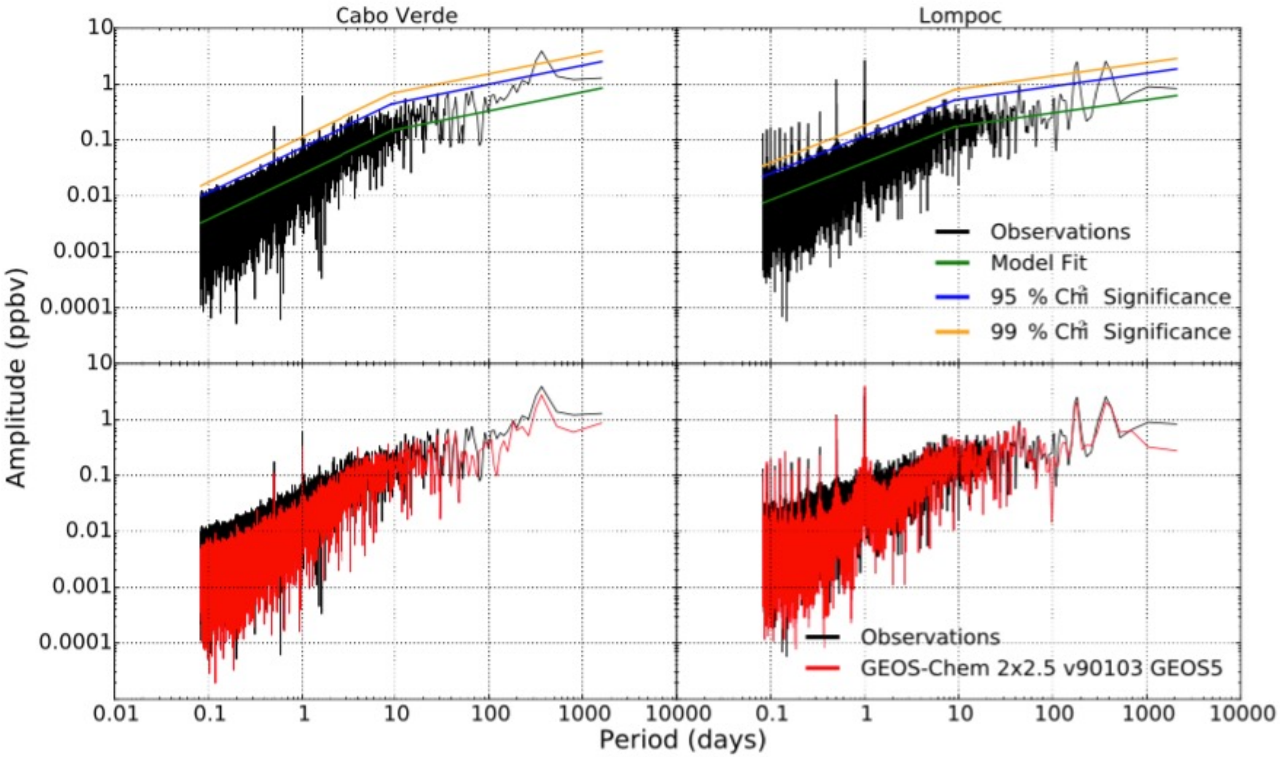
Spectral decomposition of modelled and observed time series

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



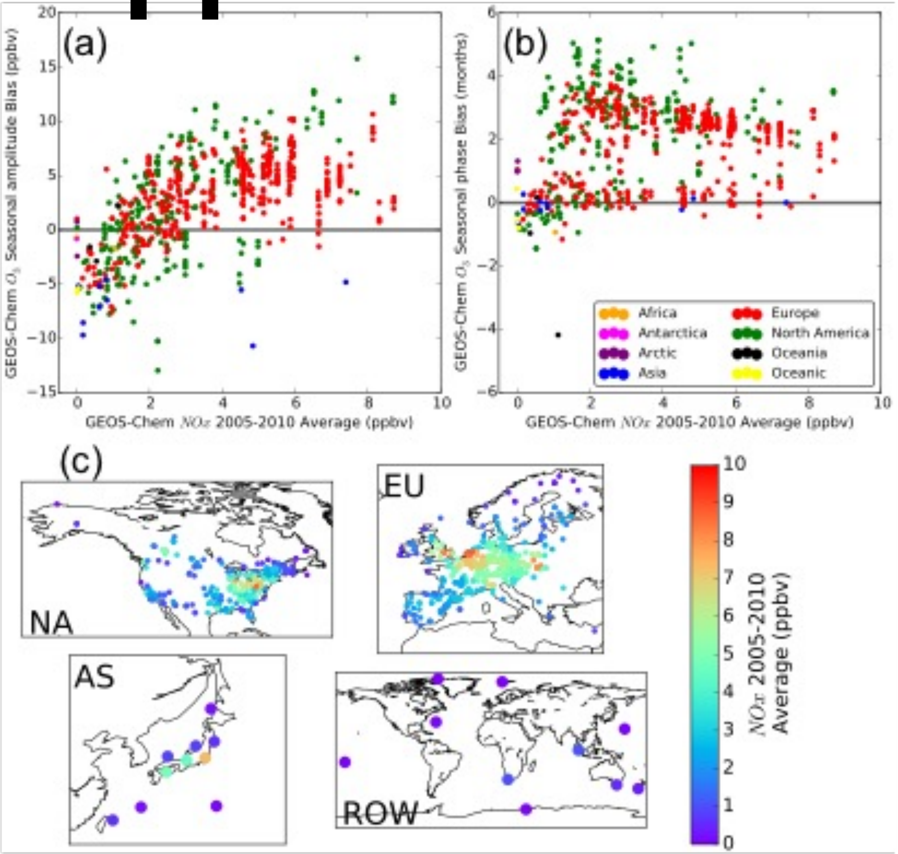


Approaches to evaluation:



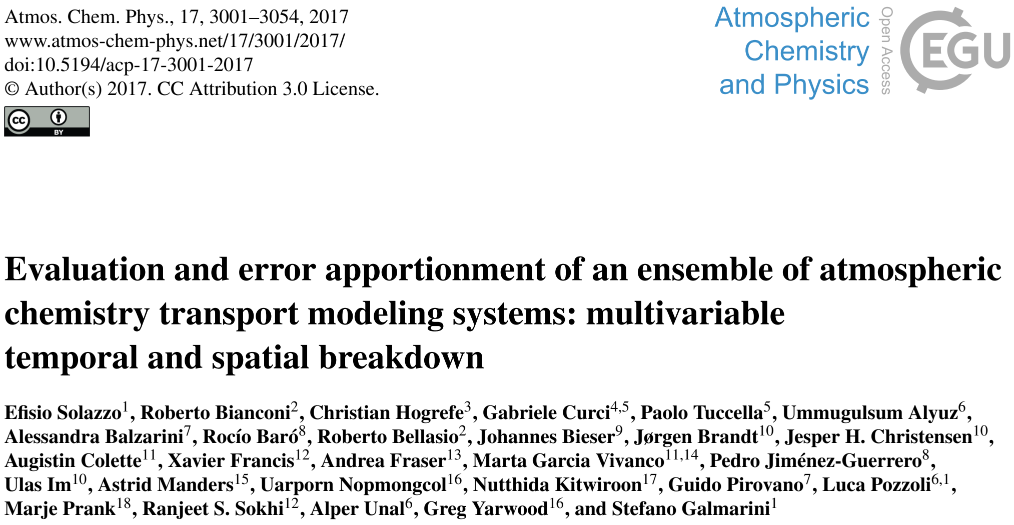
www.atmos-chem-phys.net/16/8295/2016/ Atmos. Chem. Phys., 16, 8295–8308, 2016

Approaches to evaluation:



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

Approaches to evaluation:

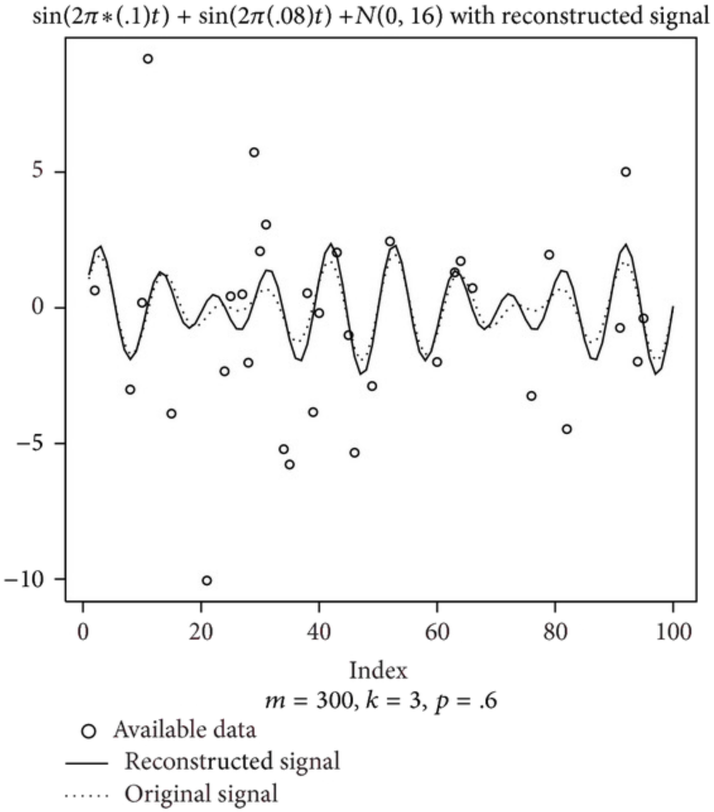


Atmos. Chem. Phys., 17, 3001–3054, 2017
www.atmos-chem-phys.net/17/3001/2017/
doi:10.5194/acp-17-3001-2017

Atmospheric Chemistry and Physics
Open Access

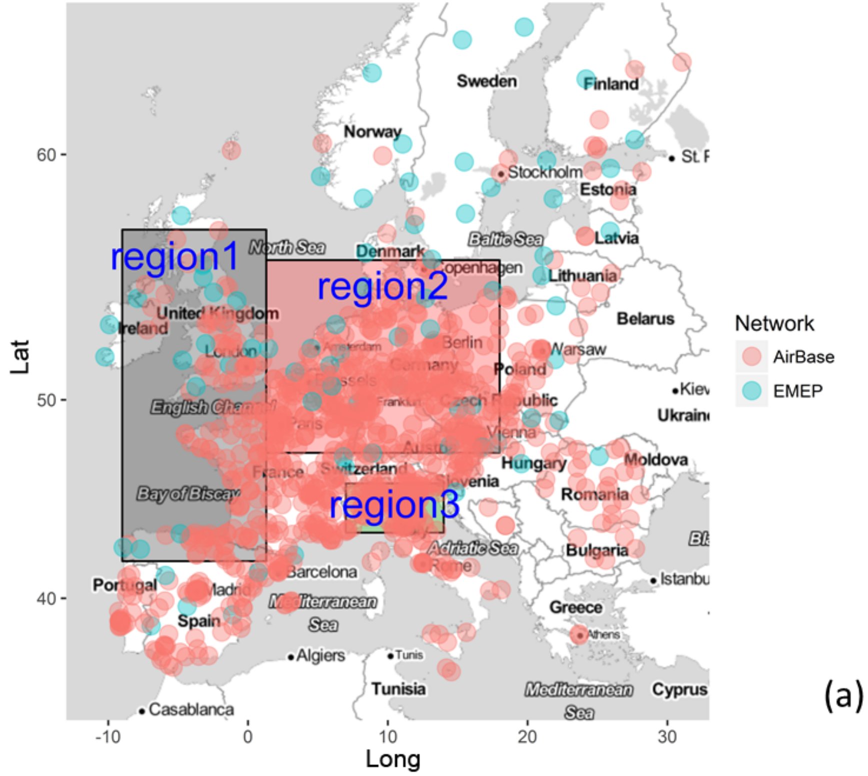
Evaluation and error apportionment of an ensemble of atmospheric chemistry transport modeling systems: multivariable temporal and spatial breakdown

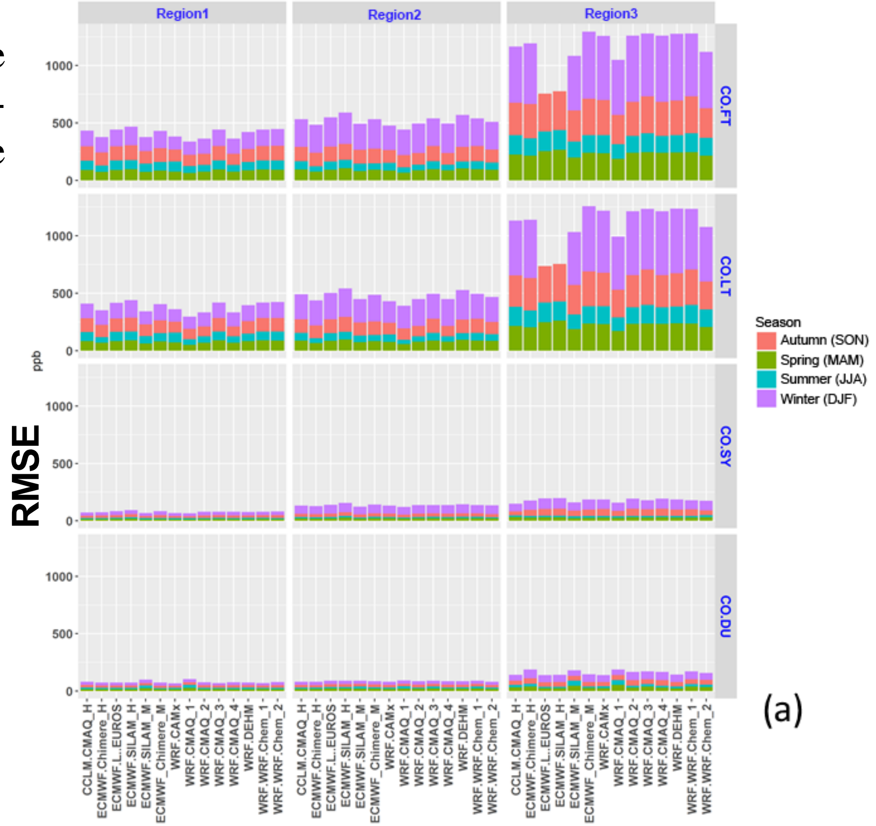
Efisio Solazzo[1], Roberto Bianconi[2], Christian Hogrefe[3], Gabriele Curci[4,5], Paolo Tuccella[5], Ummugulsum Alyuz[6], Alessandra Balzarini[7], Rocío Baró[8], Roberto Bellasio[2], Johannes Bieser[9], Jørgen Brandt[10], Jesper H. Christensen[10], Augistin Colette[11], Xavier Francis[12], Andrea Fraser[13], 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]



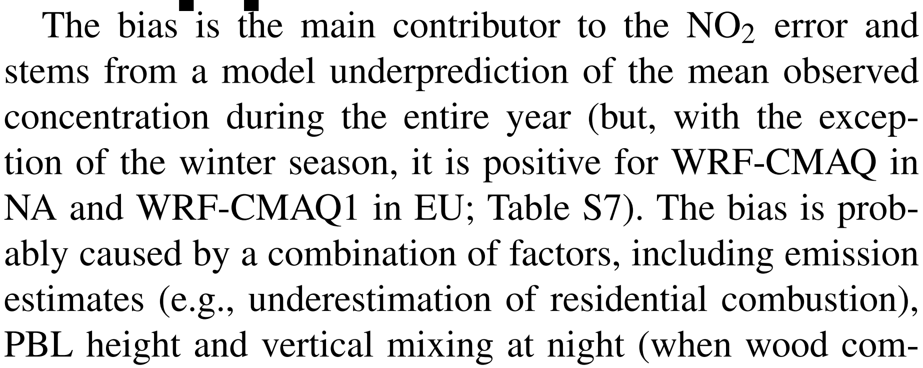
Approaches to evaluation: CO

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

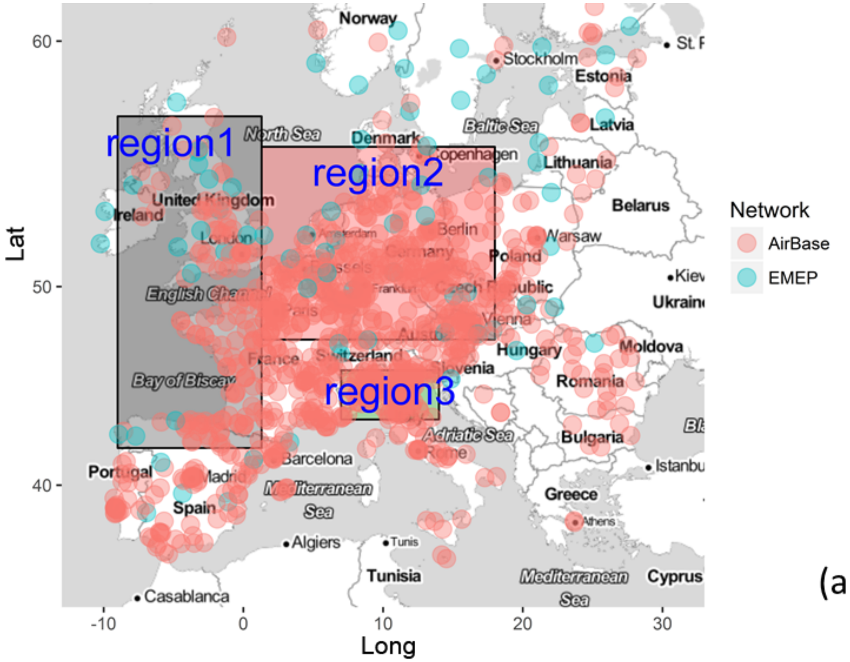




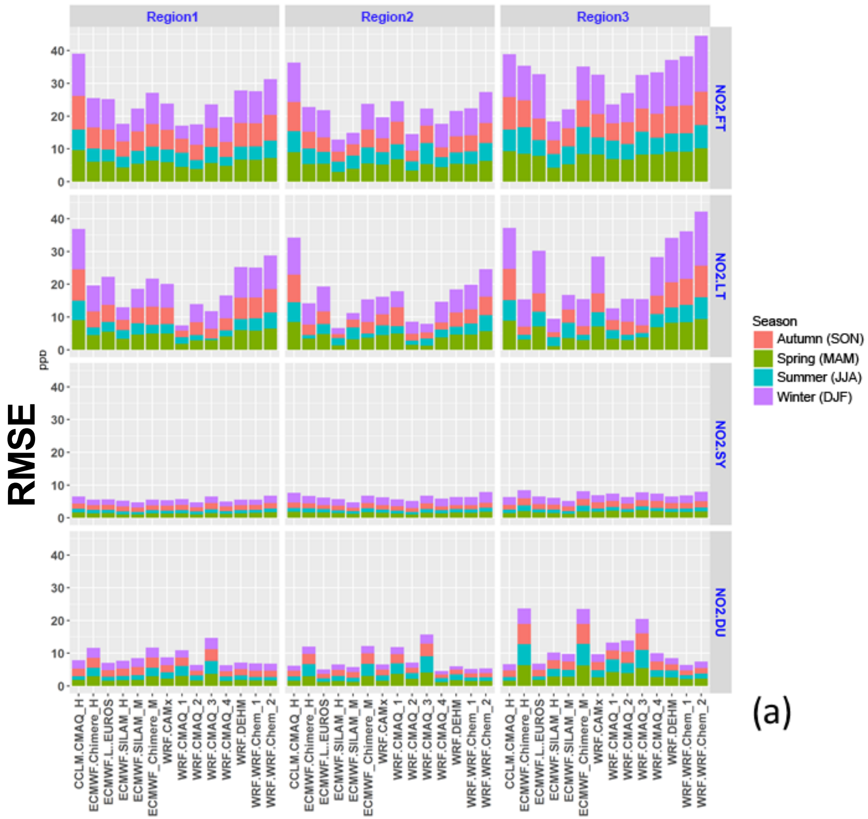
Approaches to evaluation: NO_2



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-

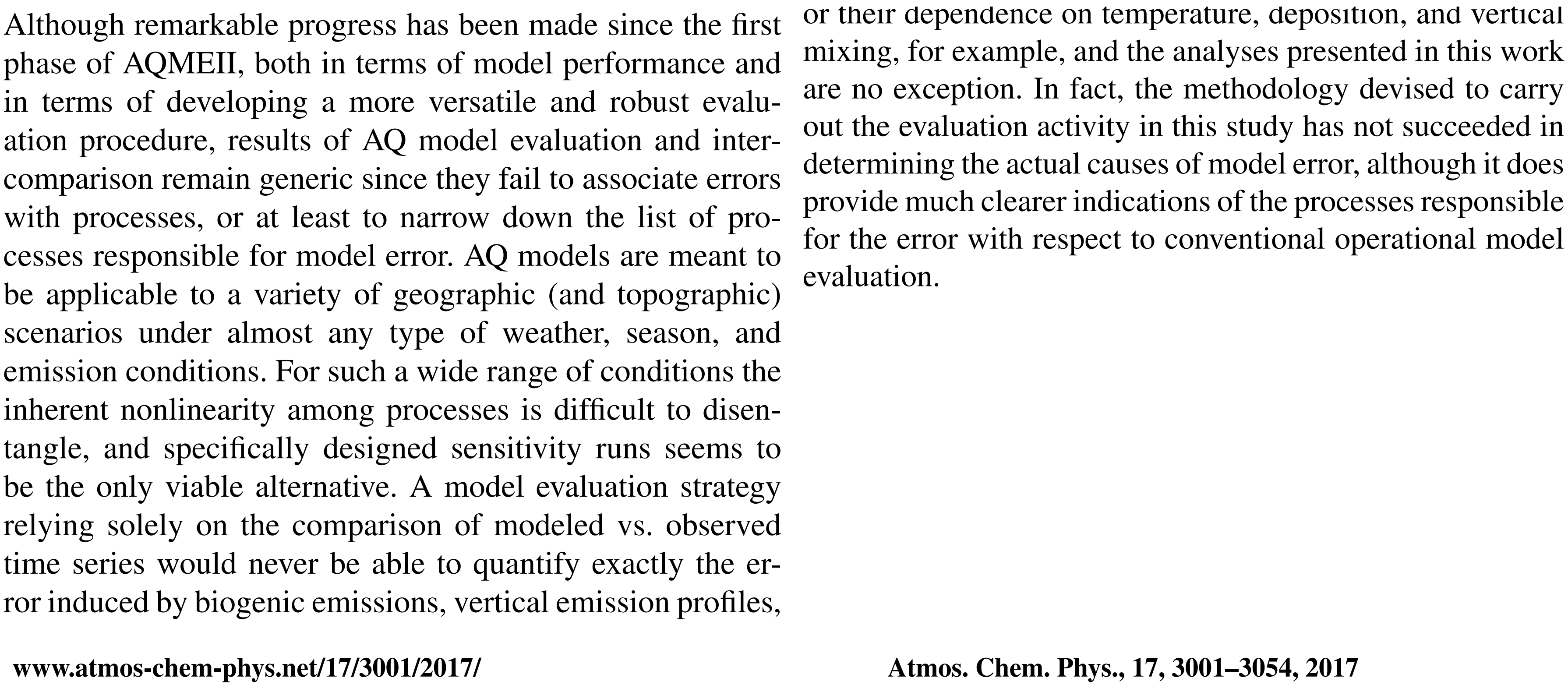


(a)



(a)

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.

www.atmos-chem-phys.net/17/3001/2017/ **Atmos. Chem. Phys., 17, 3001–3054, 2017**

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

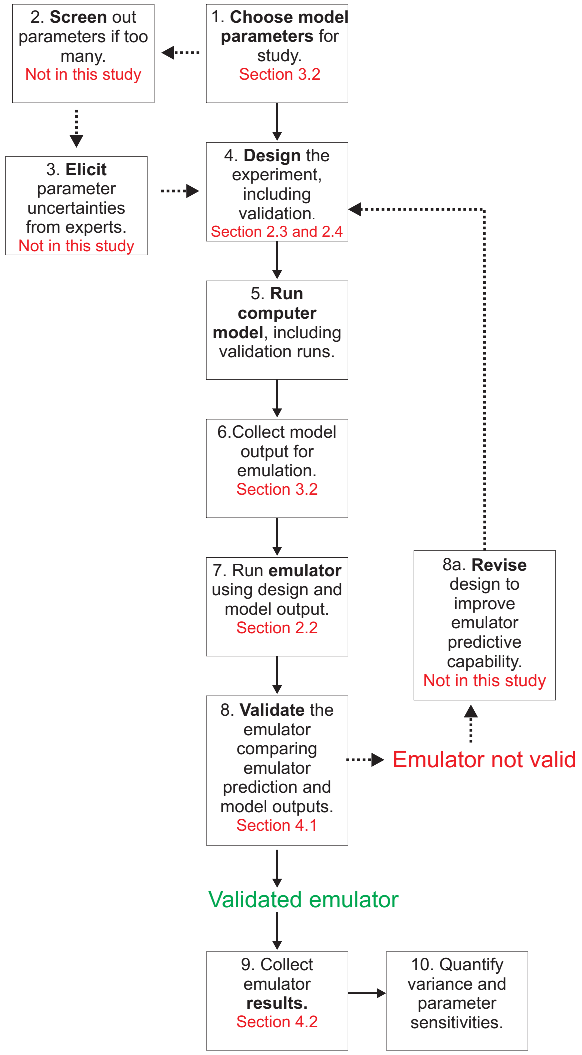
Atmospheric
Chemistry
and Physics

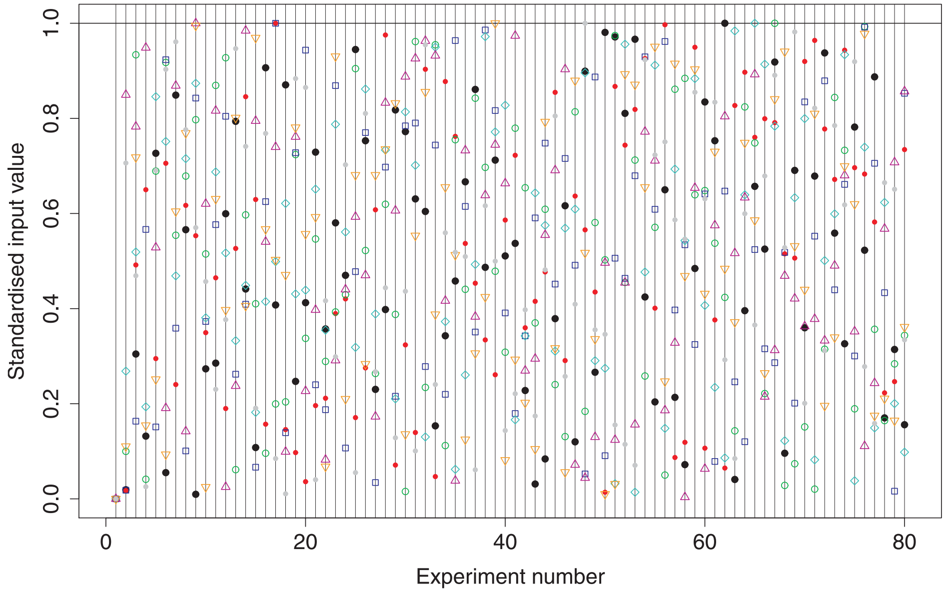
Emulation of a complex global aerosol model to quantify sensitivity to uncertain parameters

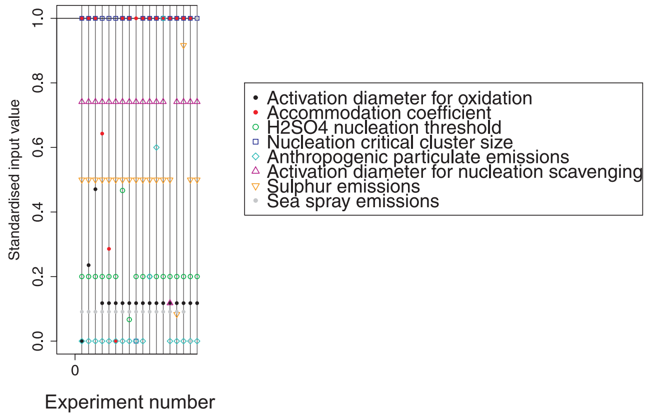
L. A. Lee, K. S. Carslaw, K. J. Pringle, G. W. Mann, and D. V. Spracklen

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

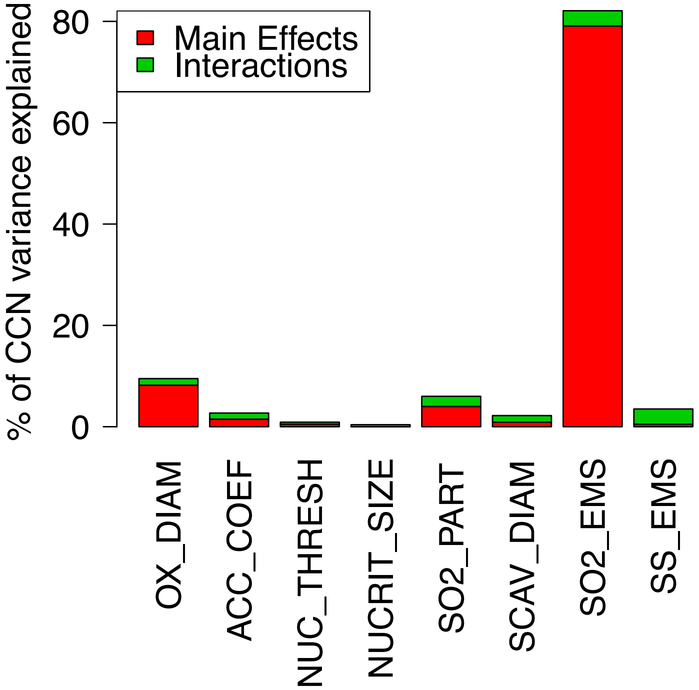
Received: 7 July 2011 – Published in Atmos. Chem. Phys. Discuss.: 19 July 2011
Revised: 11 November 2011 – Accepted: 16 November 2011 – Published: 8 December 2011


2. **Screen** out parameters if too many.
Not in this study
1. **Choose model parameters** for study.
Section 3.2
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
8a. **Revise** design to improve emulator predictive capability.
Not in this study
8. **Validate** the emulator comparing emulator prediction and model outputs.
Section 4.1
Emulator not valid
Validated emulator
9. Collect emulator **results.**
Section 4.2
10. Quantify variance and parameter sensitivities.

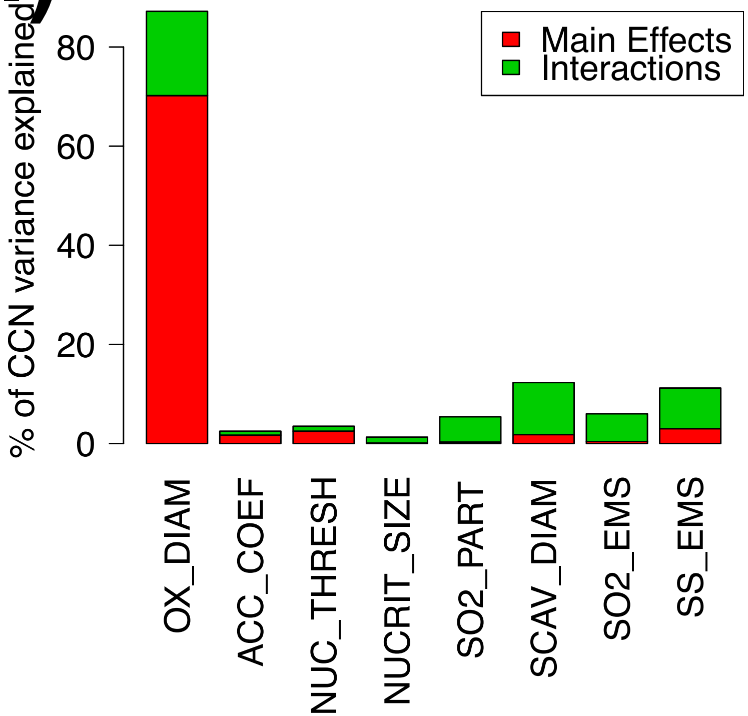

Standardised input value
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
Experiment number


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)