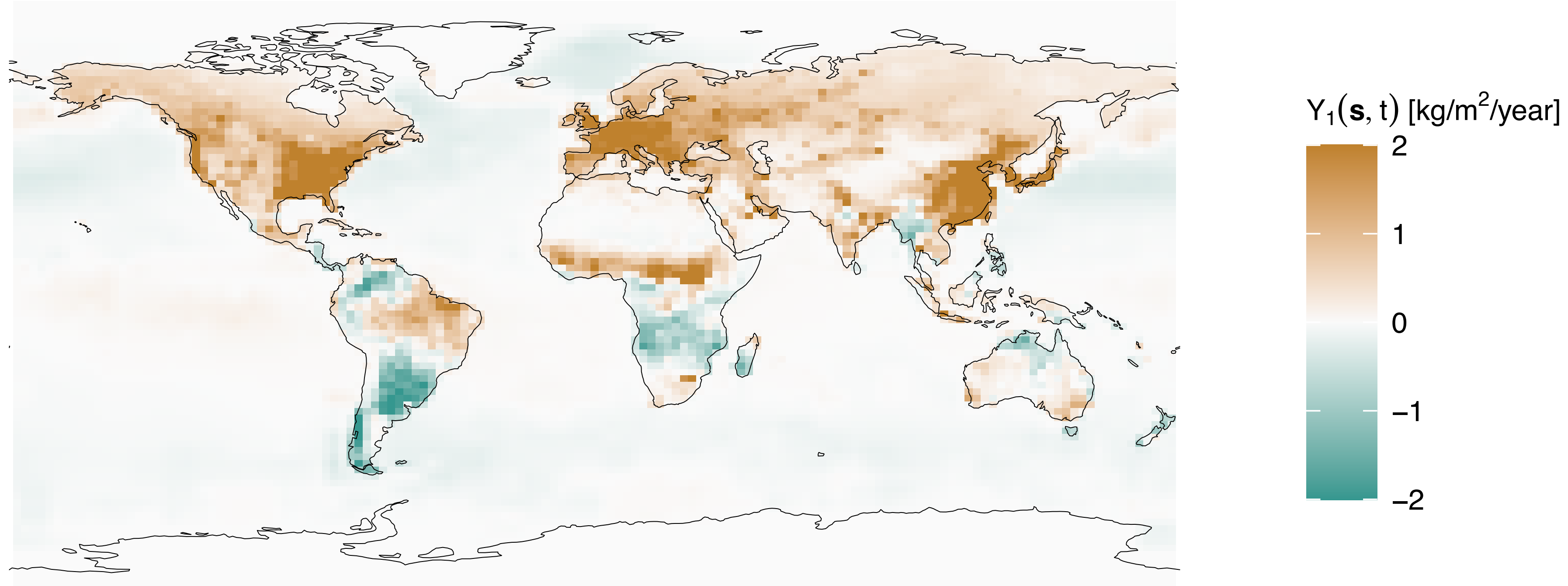
WOMBAT: A fully Bayesian global flux-inversion framework

The WOllongong Methodology for Bayesian Assimilation of Trace-gases



Michael Bertolacci
(michael_bertolacci@uow.edu.au)

National Institute for Applied Statistics Research Australia (NIASRA)
University of Wollongong, Australia

arXiv preprint: https://arxiv.org/abs/2102.04004
Contact: michael_bertolacci@uow.edu.au

UNIVERSITY OF WOLLONGONG AUSTRALIA
NIASRA
NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

Key members and their disciplines:

- Andrew Zammit-Mangion (Statistics)
- Michael Bertolacci (Statistics)
- Jenny Fisher (Atmospheric Chemistry)
- Yi Cao (IT)
- Noel Cressie (Statistics)

Other key members:

Matt Rigby, University of Bristol, UK (Atmospheric Chemistry)
Ann Stavert, CSIRO, Australia (Atmospheric Chemistry)

Valuable input/feedback:

Several others including Andrew Schuh, Anita Ganesan, Peter Rayner, Beata Bukosa, and members of the OCO-2 Flux Group.

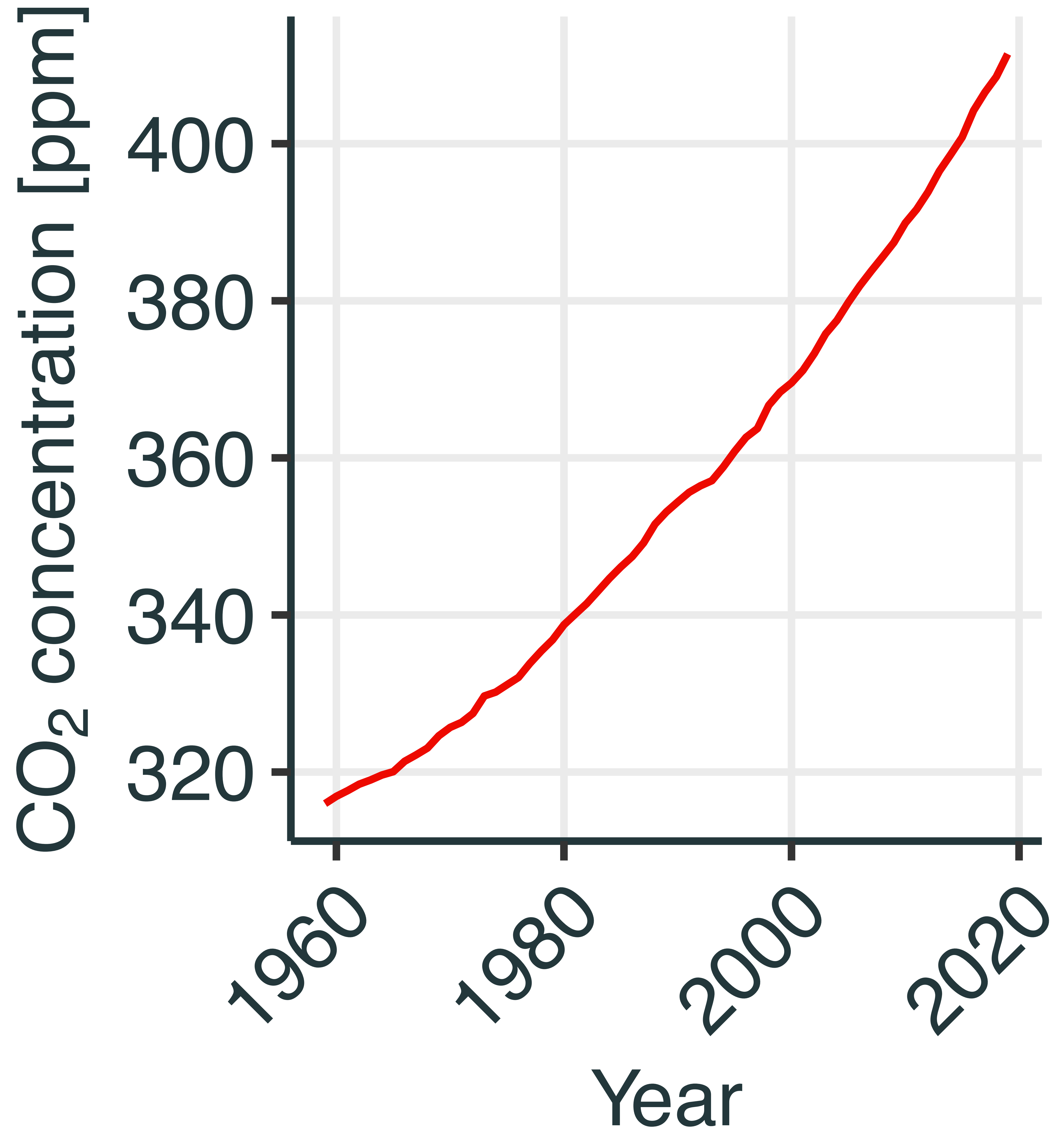
arXiv preprint: https://arxiv.org/abs/2102.04004

UNIVERSITY OF WOLLONGONG AUSTRALIA

NIASRA NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

“Human influence on the climate system is clear, and recent anthropogenic emissions of greenhouse gases are the highest in history. [...]

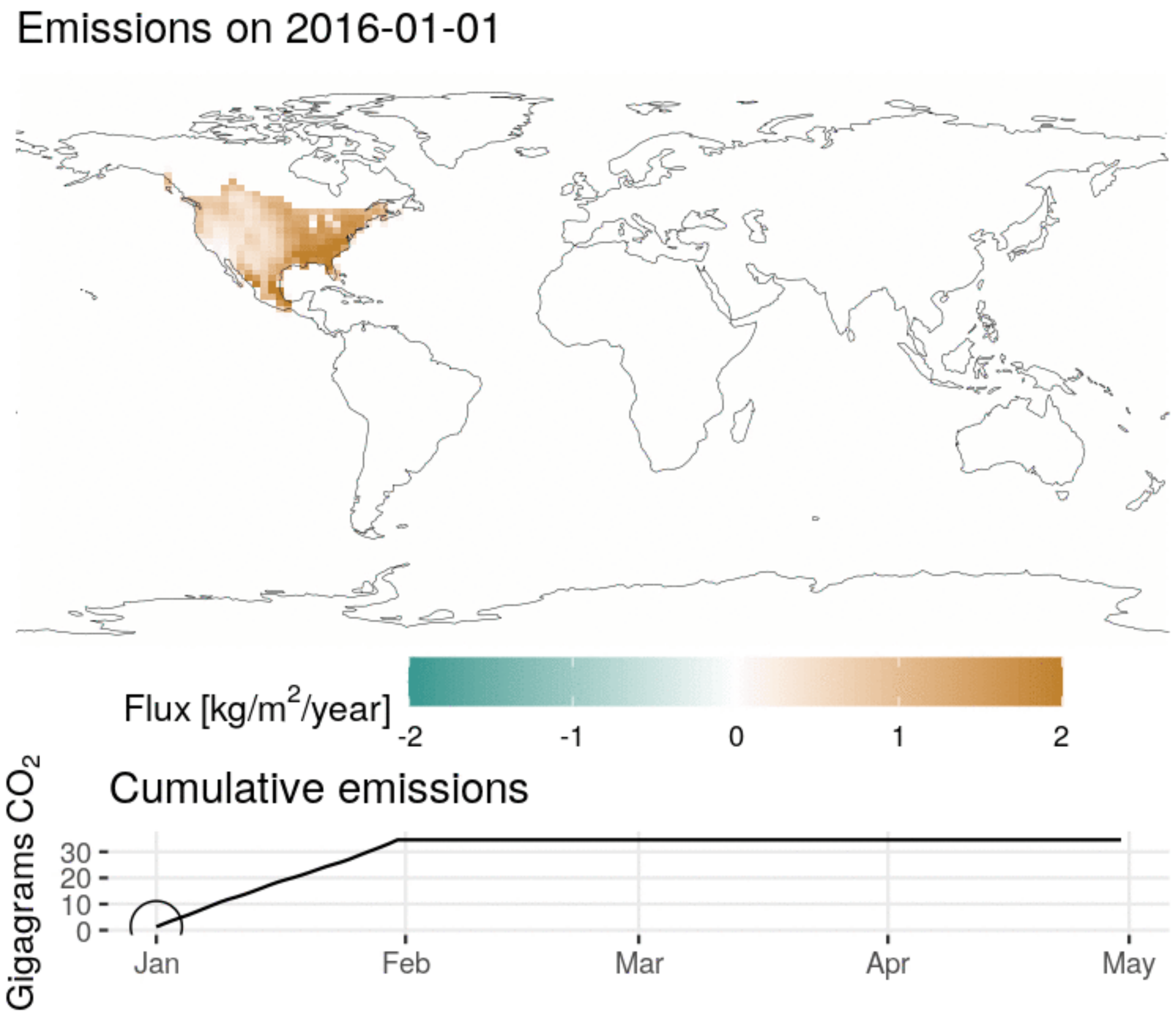
Continued emission of greenhouse gases will cause further warming and long-lasting changes in all components of the climate system.”

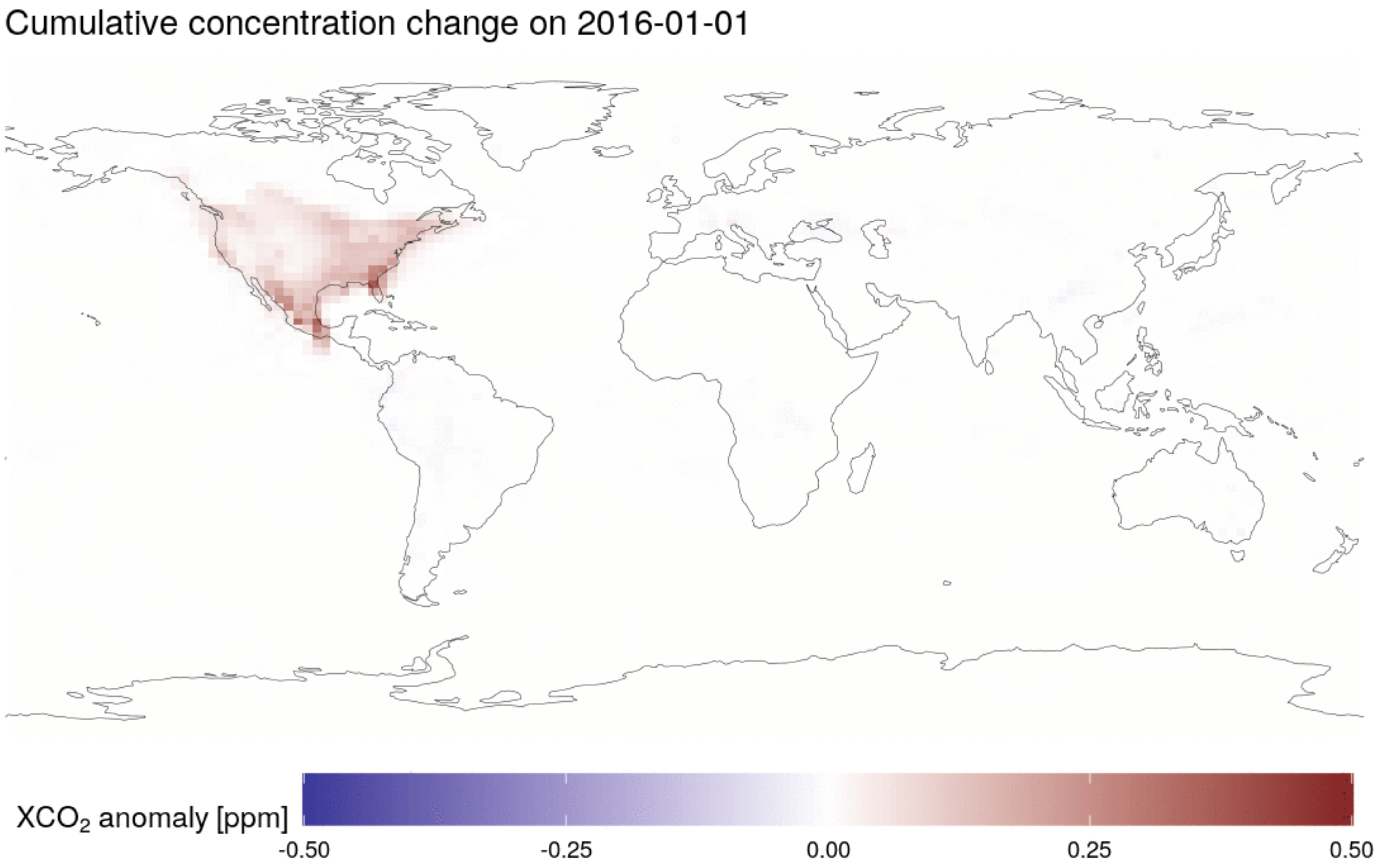


Quote: IPCC (2014), *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

Data: Dr. Pieter Tans, NOAA/GML (www.esrl.noaa.gov/gmd/ccgg/trends/) and Dr. Ralph Keeling, Scripps Institution of Oceanography (scrippsco2.ucsd.edu/)

Estimated change in the CO_2 field in response to a month of emissions in North America





Efforts to reduce net CO_2 emissions must be global.

This requires a detailed understanding of where CO_2 is:

- emitted (sources)
- absorbed (sinks)

CO_2 fluxes are hard to measure directly

For manmade sources there are good proxy measurements for emissions, such as electricity usage

This is not true for natural sources and sinks!

However, CO_2 concentrations are (relatively) easy to observe

Working backwards from CO_2 concentrations to fluxes is called flux inversion

[Images by joelbeeb (top) and Hansueli Krapf (bottom) licensed under CC BY-SA 3.0]

WOMBAT: The underlying statistical model

UNIVERSITY OF WOLLONGONG AUSTRALIA

NIASRA
NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

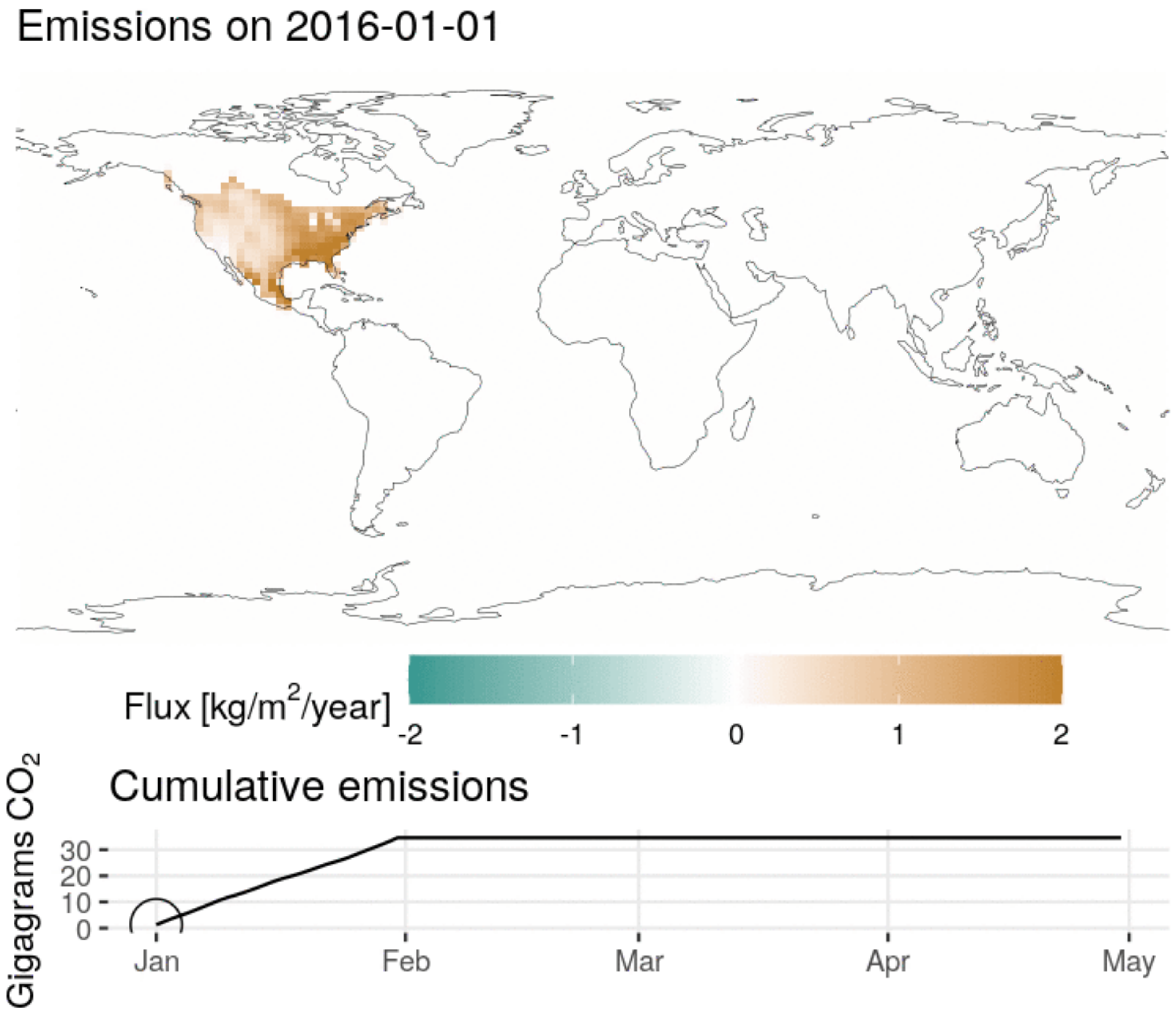
Flux field: $Y_1(\mathbf{s}, t),\ \mathbf{s} \in \mathbb{S}^2,\ t \in \mathbb{R}$

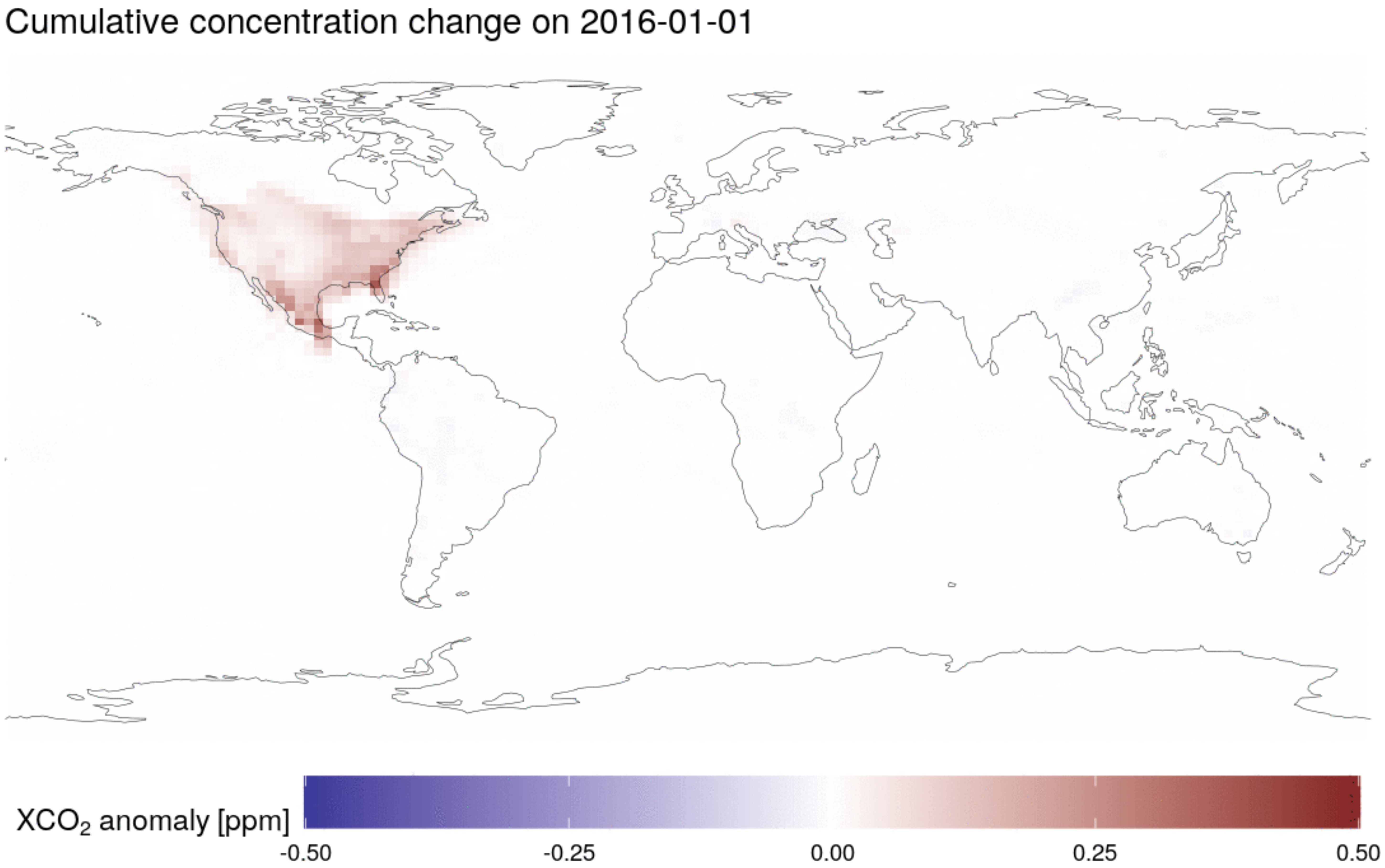
(Plausible values averaged over the month of January 2016)

Mole-fraction field: $Y_2(\mathbf{s}, h, t),\ \mathbf{s} \in \mathbb{S}^2,\ h \geq 0,\ t \in \mathbb{R}$

The flux field drives the mole-fraction field:

$$Y_2(\mathbf{s}, h, t) = \int G(\mathbf{s}, h, t; \mathbf{u}, r) Y_1(\mathbf{u}, r) d(\mathbf{u}, r) + v_2(\mathbf{s}, h, t)$$

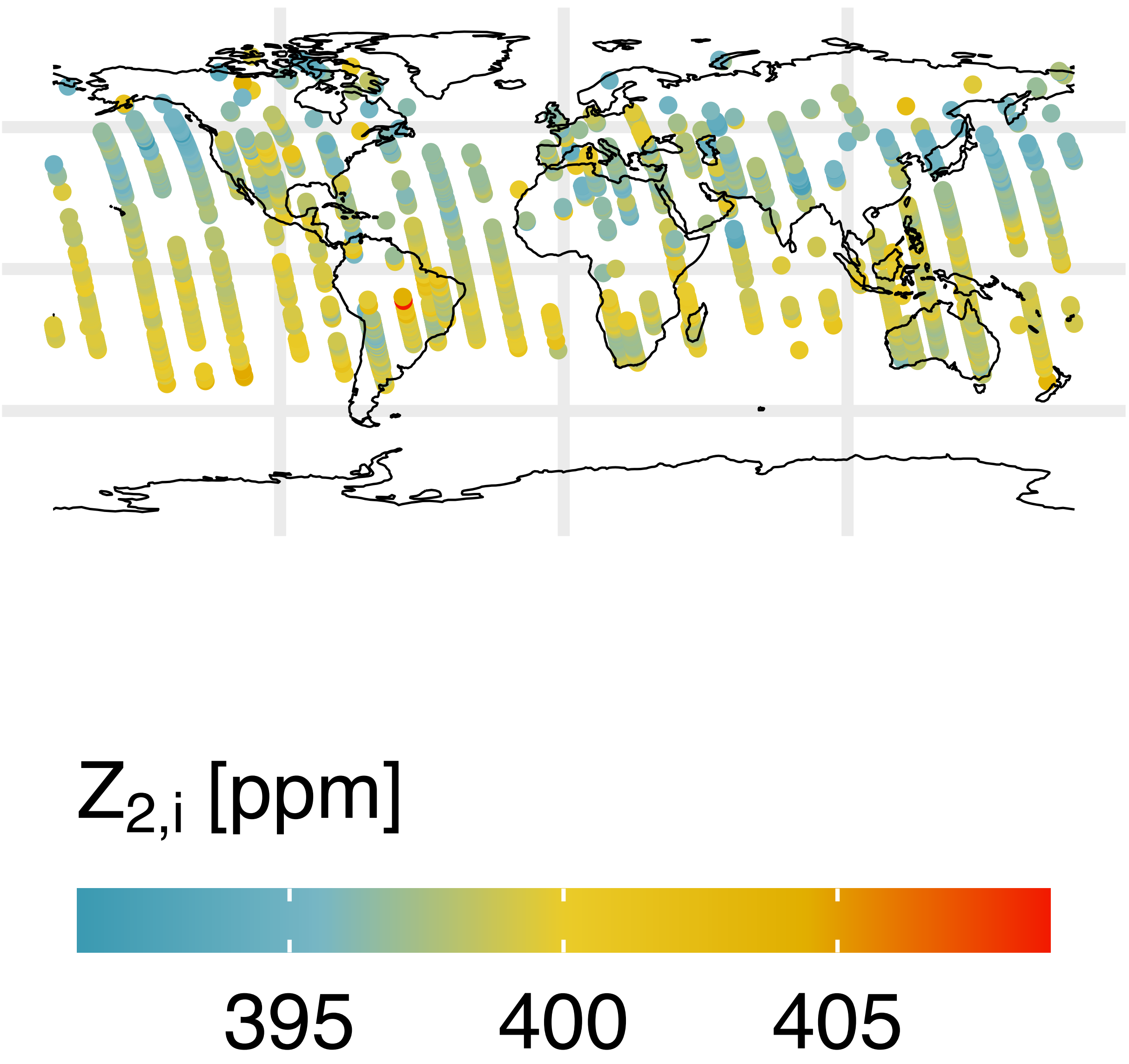




Data: $Z_{2,i}$, for $i \in \{1, \ldots, m\}$

(Observations over the period 2016-01-01 to 2016-01-07, inclusive)

A (partial) hierarchical Bayesian framework for flux inversion



Data model

$$Z_{2,i} = \mathscr{A}_i(Y_2(\mathbf{s}_i, \cdot, t_i)) + \underbrace{\mathbf{x}_i'\boldsymbol{\beta}}_{\text{Bias correction}} + \underbrace{\xi_i + \epsilon_i}_{\text{Correlated + uncorrelated errors}}$$

Mole-fraction process model:

$$Y_2(\mathbf{s}, h, t) = \int \underbrace{G(\mathbf{s}, h, t; \mathbf{u}, r)}_{\text{Atmospheric transport}} Y_1(\mathbf{u}, r) d(\mathbf{u}, r) + \underbrace{\nu_2(\mathbf{s}, h, t)}_{\text{Mole-fraction field error}}$$

The goal in flux inversion is to infer the flux field, $Y_1(\mathbf{s}, t)$, and quantify uncertainty in:

- the bias-correction coefficients, $\boldsymbol{\beta}$, and errors
- atmospheric transport via transport error
- the flux field $Y_1(\mathbf{s}, t)$ itself

Gridded flux field:
hourly at 2° lat × 2.5° lon with polar cells at half-height

⇒ 314,496 unknowns per day

Study period:
2 years with 314,496 unknowns per day

⇒ around 220 million unknowns

We decompose the flux field into prior mean, basis functions, and fine-scale term:

$$Y_1(\mathbf{s}, t) = Y_1^0(\mathbf{s}, t) + \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} \phi_{i,j}(\mathbf{s}, t)\alpha_{i,j} + \nu_1(\mathbf{s}, t)$$

Flux prior mean

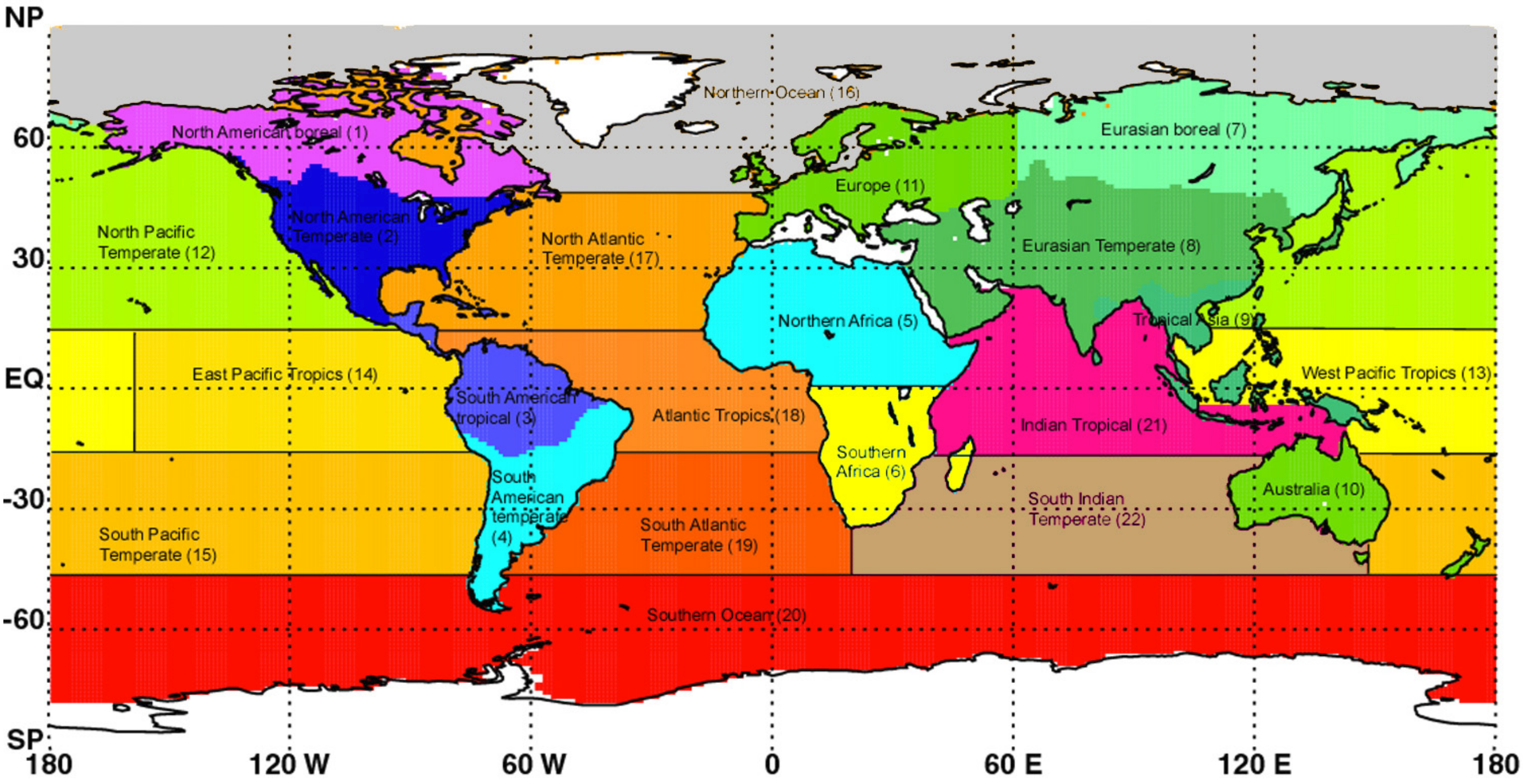
Flux basis functions and scalings

Fine-scale flux

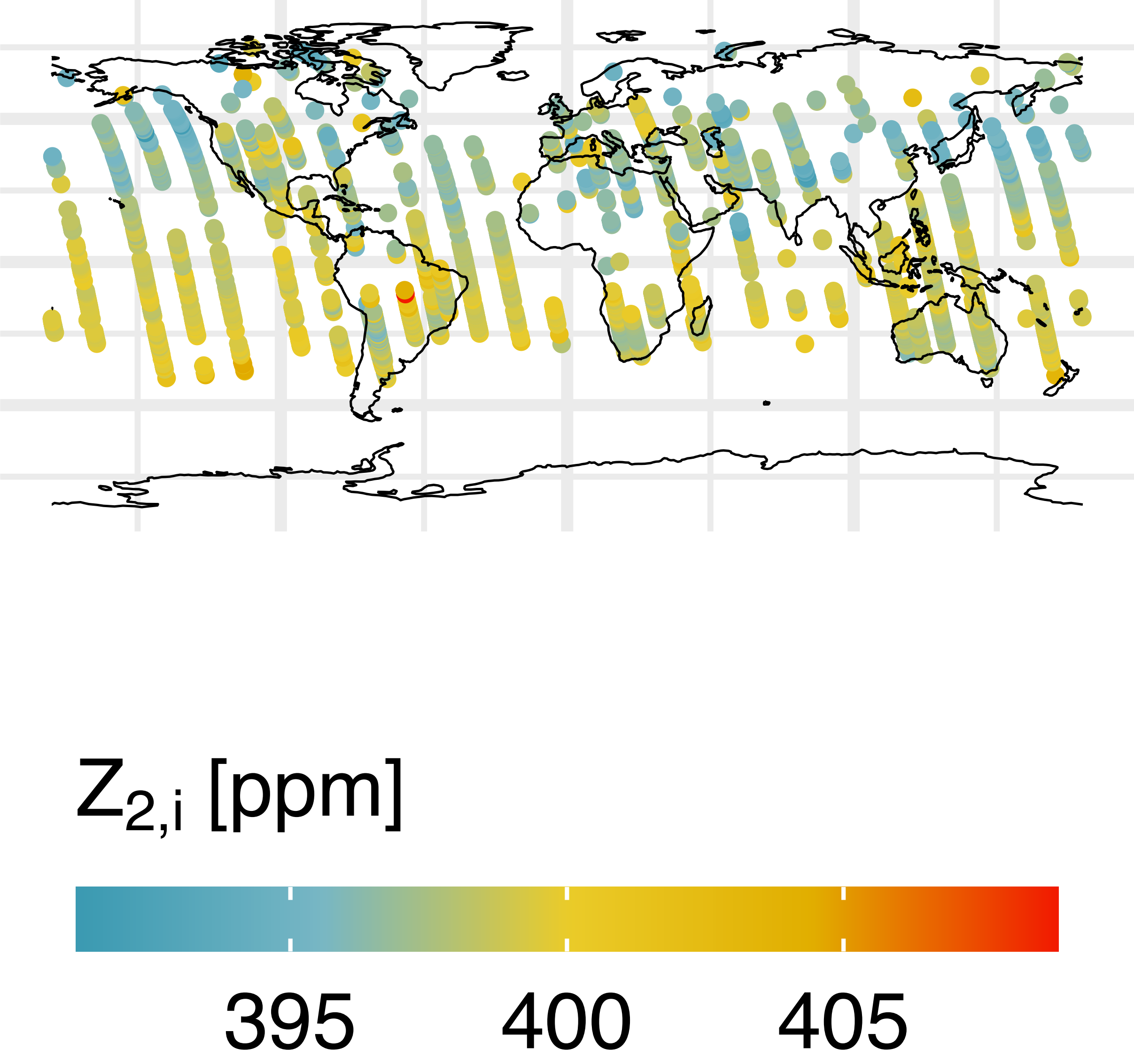
where j ranges over n_t months, i ranges over n_s regions, and $E[\alpha_{i,j}] = 0$

The fluxes evolve through $\alpha_{i,j} = \rho_i \alpha_{i,j-1} + v_{i,j}$, $v_{i,j} \sim N(0,\sigma_i^2)$, which introduces correlation from month-to-month.

The 22 TransCom3 regions



WOMBAT's hierarchical Bayesian framework for flux inversion



Data model

$$Z_{2,i} = \mathscr{A}_i(Y_2(\mathbf{s}_i, \cdot, t_i)) + \mathbf{x}_i'\boldsymbol{\beta} + \underbrace{\xi_i + \epsilon_i}$$

Bias correction

Correlated + uncorrelated errors

Mole-fraction process model:

$$Y_2(\mathbf{s}, h, t) = \int G(\mathbf{s}, h, t; \mathbf{u}, r) Y_1(\mathbf{u}, r) d(\mathbf{u}, r) + \nu_2(\mathbf{s}, h, t)$$

Atmospheric transport

Mole-fraction field error

Flux process model:

$$Y_1(\mathbf{s}, t) = Y_1^0(\mathbf{s}, t) + \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} \varphi_{i,j}(\mathbf{s}, t)\alpha_{i,j} + \nu_1(\mathbf{s}, t)$$

Flux prior mean

Flux basis functions and scalings

Fine-scale flux

$$\alpha_{i,j} = \kappa_i \alpha_{i,j-1} + \eta_{i,j}, \quad \eta_{i,j} \sim N(0, \sigma_i^2)$$

Dynamic model for scalings

Recall that the flux field has a basis function representation

$$Y_1(\mathbf{s}, t) = Y_1^0(\mathbf{s}, t) + \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} \varphi_{i,j}(\mathbf{s}, t)\alpha_{i,j} + \nu_1(\mathbf{s}, t)$$

And that

$$Y_2(\mathbf{s}, h, t) = \int G(\mathbf{s}, h, t; \mathbf{u}, r) Y_1(\mathbf{u}; r) d(\mathbf{u}, r) + \nu_2(\mathbf{s}, h, t).$$

Interchange integration and summation and assume $\nu_1(\mathbf{s}, t) = 0$:

$$Y_2(\mathbf{s}, h, t) = Y_2^0(\mathbf{s}, h, t) + \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} \psi_{i,j}(\mathbf{s}, h, t)\alpha_{i,j} + \nu_2(\mathbf{s}, h, t),$$

or more succinctly

$$Y_2(\mathbf{s}, h, t) = Y_2^0(\mathbf{s}, h, t) + \boldsymbol{\psi}(\mathbf{s}, h, t)'\boldsymbol{\alpha} + \nu_2(\mathbf{s}, h, t)\,.$$

That is, the mole-fraction field has a corresponding basis function representation.

Based on several “inventories”: bottom-up estimates of CO_2 sources and sinks due to fossil fuels, fires, biofuels, the oceans, and the biosphere

$$Y_1(\mathbf{s}, t) = Y_1^0(\mathbf{s}, t) + \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} \varphi_{i,j}(\mathbf{s}, t)\alpha_{i,j} + \nu_1(\mathbf{s}, t)$$

$$Y_2(\mathbf{s}, h, t) = Y_2^0(\mathbf{s}, h, t) + \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} \psi_{i,j}(\mathbf{s}, h, t)\alpha_{i,j} + \nu_2(\mathbf{s}, h, t)$$

Computed conditional on $Y_1(\,\cdot\,,\,\cdot\,)$ by an atmospheric transport model (GEOS-Chem), driven by estimated meteorological fields (GEOS-FP)

Estimating the model

WOMBAT performs inference using Markov chain Monte Carlo (MCMC):

⇒ Posterior distributions on all unknowns.

⇒ Can estimate posterior of functionals of unknowns (e.g. of flux field).

A simulation study, and flux estimates

UNIVERSITY OF WOLLONGONG AUSTRALIA

NIASRA
NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

Simulation study: the importance of bias correction and correlated errors

WOMBAT's data model

Bias correction

$$Z_{2,i} = \mathscr{A}_i(Y_2(\mathbf{s}_i, \cdot, t_i)) + \mathbf{x}_i'\boldsymbol{\beta} + \underbrace{\xi_i + \epsilon_i}$$

Correlated + uncorrelated errors

Model performance in an OSSE experiment when estimating fluxes at the TransCom3 level when the data are biased and have correlated errors:

	RMSE [PgC/mo]	CRPS
Inversion configuration	LG	LG
Bias correction/correlated	0.023	0.010
Bias correction/uncorrelated	0.038	0.015
No bias correction/correlated	0.045	0.016
No bias correction/uncorrelated	**0.092**	**0.034**

Lower is better

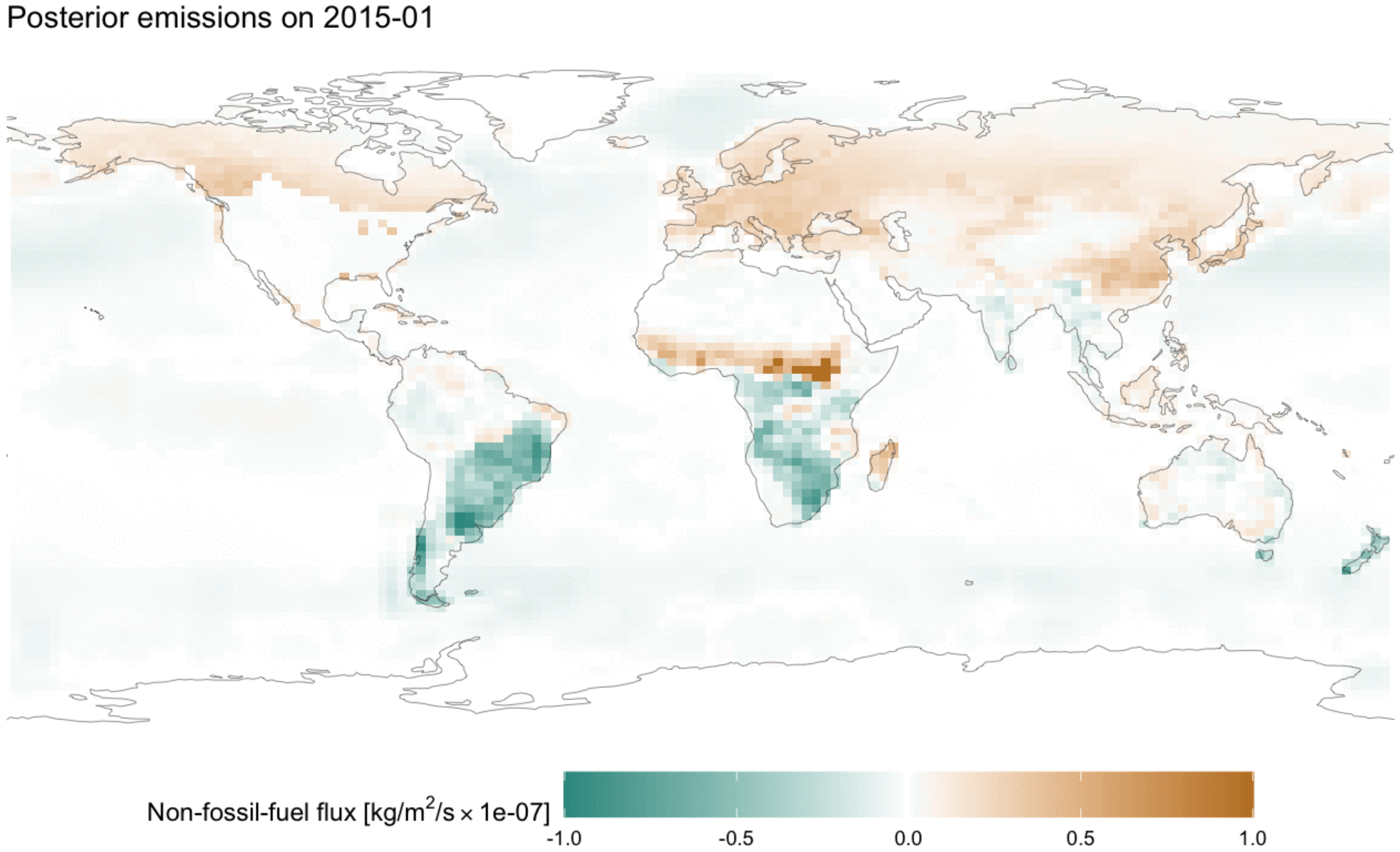
- WOMBAT's data model has both bias correction and a correlated error term
- We investigated the importance of these features in a simulation study where the simulated data are biased and have correlated errors
- **Flux estimation performance is severely impacted if these are ignored**

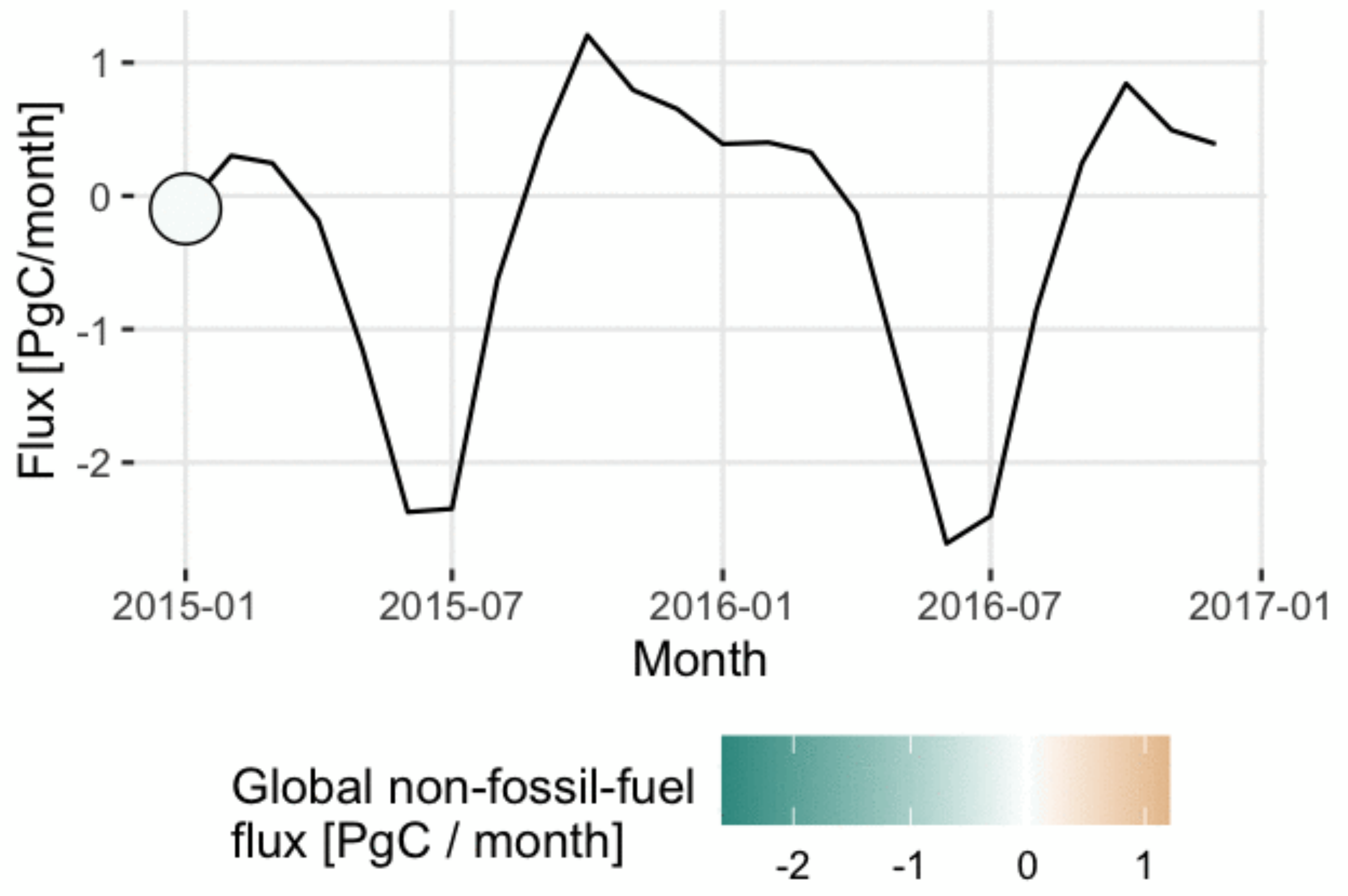
Real data results

We used WOMBAT with two years of OCO-2 data to estimate fluxes for the period of January, 2015 through to December, 2016

Fossil-fuel fluxes were assumed known, and non-fossil-fuel fluxes were estimated.

The breathing Earth: WOMBAT's posterior CO_2 fluxes for 2015–2016





(Estimated monthly posterior natural flux for each grid cell)

Conclusions, and the future

- WOMBAT is a flux-inversion system used for inferring the sources and sinks CO_2 using atmospheric concentration data
- It provides estimates of fluxes with uncertainty quantification
- Some future planned work:
  - Use multiple transport models to investigate errors in transport
  - With New Zealand's NIWA, investigate the use of WOMBAT's posterior fluxes as boundary conditions for a regional CO_2 inversion over parts of Oceania
  - Gases other than CO_2

(Photo taken on the Eyre Highway in the remote Nullarbor region of Australia)

arXiv preprint: https://arxiv.org/abs/2102.04004

UNIVERSITY OF WOLLONGONG AUSTRALIA

NIASRA
NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA