Summary of the revisions:

We have re-written many sections of the manuscript and have modified or replaced many of the figures. We discuss many of these changes in detail in the replies to the reviewers. Below is a list of several large-scale changes to the manuscript. We have also attached two pdf files that detail changes to the manuscript and to the supplement.

- We have re-designed the synthetic data experiments. The new setup provides analysis on why atmospheric observations can or cannot detect wetland fluxes. This revision adds depth to the analysis, as requested by reviewer #1. We also cut synthetic experiment #2, as suggested by reviewer #1. Furthermore, the revised approach addresses the concerns raised by reviewer #2. In particular, we use all of the WETCHIMP models in the analysis. We rely more heavily on the existing literature to support the methodology. Furthermore, we show that the method would always select a wetland flux model if there were no measurement or modeling errors. These changes show that the result of the analysis are not spurious and that the methodology is not un-tested.

- We have added additional information on atmospheric transport and modeling errors throughout the manuscript. For example, we have added a section to the supplement on atmospheric transport and modeling errors. We also provide more extensive references to existing literature that directly addresses this topic. The re-designed synthetic data experiments further highlight the effect of atmospheric transport and modeling errors. These changes address suggestions from both reviewers.

- We have re-written the conclusions section based on the general comments from reviewer #1. The revised manuscript also provides more analysis on why discrepancies might exist between bottom-up models and results based upon atmospheric data. We discuss advantages and disadvantages of both bottom-up and top-down approaches, a more equitable balance than in the previous manuscript (as suggested by reviewer #1).

- We changed the wording and structure of the article based upon the general comments from reviewer #1. Among other changes, we provide more discussion of the observations and modeling framework. We also removed repeated/redundant material and made other stylistic edits recommended by reviewer #1.

Manuscript prepared for Biogeosciences Discuss.
with version 2014/05/30 6.91 Copernicus papers of the LaTeX class copernicus.cls.
Date: 19 November 2015

The ability of atmospheric data to resolve ~~discrepancies~~ differences in wetland methane estimates over North America

S. M. Miller[1,*], R. Commane[1], J.R. Melton[2], A. E. Andrews[3], J. Benmergui[4], E.J. Dlugokencky[3], G. Janssens-Maenhout[5], A.M. Michalak[6], C. Sweeney[7], and D.E.J. Worthy[8]

[1]Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA, USA
[2]Climate Processes Section, Environment Canada, Victoria, Canada
[3]Global Monitoring Division, Earth Systems Research Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO, USA
[4]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA
[5]Institute for Environment and Sustainability, European Commission Joint Research Centre, Ispra, Italy
[6]Department of Global Ecology, Carnegie Institution for Science, Stanford, CA, USA
[7]Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO, USA
[8]Climate Research Division, Environment Canada, Toronto, Canada
[*]now at: Department of Global Ecology, Carnegie Institution for Science, Stanford, CA, USA

Correspondence to: S. M. Miller (scot.m.miller@gmail.com)

Abstract

Existing estimates of methane (CH_4) fluxes from North American wetlands vary widely in both magnitude and distribution. In light of these disagreements, this study uses atmospheric ~~methane~~ CH_4 observations from the US and Canada to analyze seven different bottom-up, wetland ~~methane~~ CH_4 estimates reported in a recent model comparison project. We first use synthetic data to explore ~~how well atmospheric observations can constrain wetland fluxes~~whether wetland CH_4 fluxes are detectable at atmospheric observation sites. We find that ~~observation sites can identify an atmospheric pattern from Canadian wetlands but not reliably from US wetlands. The network can also identify the spatial distribution of fluxes in Canada at multi-province spatial scales.~~ the observation network can detect aggregate wetland fluxes from both eastern and western Canada but generally not from the US. Based upon these results, we then use real data to ~~evaluate~~ analyze the magnitude, ~~temporal distribution~~seasonality, and spatial distribution of each model estimate. ~~Most models overestimate the magnitude of fluxes across Canada. Most~~ The magnitude of Canadian fluxes in many models is larger than indicated by atmospheric observations. Many models predict a seasonality that is ~~too narrow, potentially~~ narrower than implied by atmospheric CH_4 data, possibly indicating an over-sensitivity to air or soil temperatures. ~~In addition, the~~ The LPJ-Bern and SDGVM models have a ~~spatial~~ geographic distribution that is most consistent with atmospheric observations, depending upon the ~~season and region. Unlike most models , LPJ-Bern and SDGVM~~ region and season. These models utilize land cover maps ~~, not just remote sensing inundation data,~~ or dynamic modeling to estimate wetland coverage ~~. A flux model with a constant spatial distribution outperforms all other existing flux estimates across Canada.~~

while most other models rely primarily on remote sensing inundation data.

1 Introduction

~~Methane~~ CH_4 fluxes from wetlands play a critical role in global climate change. ~~Methane~~ CH_4 is the second-most important long-lived greenhouse gas~~;~~, and the radiative forcing of the current atmospheric burden is approximately 26 % of carbon dioxide (Butler, 2014) . Wetlands are possibly the largest single source of this gas to the atmosphere and account for roughly 30 % of global emissions (Ciais et al., 2013).

Despite the important role of wetland ~~methane~~ CH_4 fluxes in climate change, existing estimates of this source disagree ~~markedly~~ on the magnitude, seasonality, and spatial distribution of fluxes, from regional to global scales. In fact, a recent global model comparison project named WETCHIMP (Wetland and Wetland ~~CH_4~~ CH_4 Inter-comparison of Models Project) found large ~~discrepancies among existing methane wetland models (Fig. 1, Melton et al., 2013; Wania et al., 2013)~~ among existing CH_4 wetland models (Fig. 1, Melton et al., 2013; Wania et al., 2013) . For example, existing estimates of maximum global wetland coverage differ by over a factor of 6 – from ~~4.1×10^6 to 26.9×10^6 km^2~~ 4.1×10^6 to $26.9 \times 10^6\,km^2$. Furthermore, estimates of global natural wetland fluxes range from 92–264~~$Tg\,CH_4\,yr^{-1}$~~ $Tg\,CH_4\,yr^{-1}$. The relative magnitude of these uncertainties increases at sub-global spatial scales. As a case in point, ~~methane~~ CH_4 estimates for Canada's Hudson Bay Lowlands (HBL) range from 0.2 to 11.3~~$Tg\,CH_4\,yr^{-1}$~~ $Tg\,CH_4\,yr^{-1}$. These disagreements in current ~~methane~~ CH_4 estimates do not bode well for scientists' abilities to accurately predict future changes in wetland fluxes due to climate change (Melton et al., 2013).

A number of studies have used chamber measurements of ~~methane~~ CH_4 to parameterize or evaluate biogeochemical ~~methane~~ CH_4 models (e.g., Livingston and Hutchinson, 2009). ~~However, these~~ These measurements usually encompass fluxes from a ~~very small spatial scale~~relatively small area, and fluxes can ~~vary by an order of magnitude over one meter or less (Waddington and Roule~~ ~~Methane~~ often vary greatly with landscape heterogeneity at these spatial scales (Waddington and Roule CH_4 data collected in the atmosphere ~~, by contrast,~~ sees the cumulative effect of ~~methane~~ CH_4 fluxes across a ~~much broader region (e.g., Kort et al., 2008; Pickett-Heaps et al., 2011; Miller et al., 20~~ region (e.g., Winderlich et al., 2010; Pickett-Heaps et al., 2011; Miller et al., 2014) . Hence, atmospheric data can provide ~~an important~~ a unique tool for evaluating existing ~~methane~~ CH_4 flux estimates across different countries or continents.

The present study compares the WETCHIMP ~~methane~~ CH_4 flux estimates against atmospheric ~~methane~~ CH_4 data from 2007–2008 through two sets of analyses. First, we construct

~~progressively demanding~~ a set of synthetic data experiments to ~~explore how well available data can constrain wetland fluxes. Can the atmospheric data identify methane patterns from wetlandsover distracting patterns in the atmosphere? These patterns include methane from anthropogenic sources or random noise due to model and measurement errors. If yes, can the observation sites detect spatial variability in the wetland fluxes? We seek to understand whether large uncertainties in wetland methane estimates point to a paucity of methane data – data capable of calibrating or evaluating the models. In the alternative, perhaps these disagreements would be much smaller if existing biogeochemical models leveraged all available data.~~ understand whether the atmospheric CH_4 observation network can detect CH_4 fluxes from wetlands. We also explore the factors that might prevent the network from detecting wetland fluxes. To answer these questions, we utilize a ~~modeling approach~~ model selection procedure based upon the Bayesian Information Criterion (BIC) ~~, described in greater detail in Sect. **??** (Shiga et al., 2014; Fang e~~

~~Based on the~~ (Sect. 2.2 Shiga et al., 2014; Fang et al., 2014; Fang and Michalak, 2015) . This procedure determines whether wetland fluxes from different regions and seasons are necessary to describe variability in synthetic atmospheric CH_4 observations. Based on these synthetic experiments, we conduct a second set of analyses using real atmospheric data. We use this data to ~~evaluate~~ analyze the magnitude, seasonal cycle, and spatial distribution of each WETCHIMP ~~methane estimate. Of the seven available models, which have a magnitude, seasonal cycle, or spatial distribution that is most consistent with the available data? We investigate this question over the United States~~ CH_4 estimate. We investigate these questions over the US and Canada using ~~methane~~ CH_4 data collected from towers and regular aircraft flights operated by NOAA and its partners and from towers operated by Environment Canada.

2 Methods

This section first describes the atmospheric ~~methane~~ CH_4 data and the atmospheric model that allows direct comparison between the data and various flux estimates. Subsequent sections describe ~~how we use these tools to construct~~ both the synthetic and real data experiments outlined in the introduction (Sect. 1).

2.1 Data and atmospheric model

The present study utilizes atmospheric ~~methane observations at Environment Canadaand NOAA observation sites~~ CH_4 observations from aircraft and tower platforms across the US and Canada, a total of 14 703 observations from 2007–2008. These observation sites include four towers operated by Environment Canada, and 10 towers in the US operated by NOAA and its partners. Observations at the NOAA towers consist of daily flasks (occasionally weekly), and observations at the Environment Canada sites are continuous measurements. As in Miller et al. (2014) , we use afternoon averages of this continuous data. In addition to these towers, we utilize observations from 17 regular NOAA aircraft monitoring locations in the US and Canada (Fig. 2). ~~These include regular measurements from tower and aircraft platforms, a total of 14,703 observations from 2007-2008. The~~ We incorporate aircraft data up to 2500m altitude as was done in Miller et al. (201 Observations above that height are usually representative of the free troposphere with limited sensitivity to surface fluxes. The tower and aircraft observations used here are ~~identical to~~ the same as those in Miller et al. (2013) and Miller et al. (2014).

We then employ an atmospheric transport model to relate ~~methane~~ CH_4 fluxes at the Earth's surface to atmospheric concentrations at the observation sites. The modeling approach here combines the Weather Research and Forecasting (WRF) meteorological model and a particle-following model known as STILT, the Stochastic Time-Inverted Lagrangian Transport model (e.g., Lin et al., 2003; Nehrkorn et al., 2010; Hegarty et al., 2013). WRF-STILT generates a set of footprints; these footprints quantitatively estimate the sensitivity of each observation to fluxes at each surface location (with units of ppb per unit surface flux). We multiply the footprints by a flux model and add this product to an estimate of the ~~'background'~~ "background" concentration – the ~~methane~~ CH_4 concentration of air entering the North American regional domain. We estimate this background concentration using CH_4 observations collected near or over the Pacific Ocean and in the high Arctic, a setup described in detail by Miller et al. (2013) and Miller et al. (2014) . The resulting modeled concentrations can be compared directly against atmospheric ~~methane~~ CH_4 observations. This modeling setup is ~~identical to~~ the same as in Miller et al. (2013) and Miller et al. (2014). ~~Both the observationsand~~ The observations, the WRF-STILT model, background concentrations, and uncertainties in the modeling framework are described in greater detail in ~~those papers and in the supplement~~the Supplement and in those papers.

Using this setup, we can ~~compare predicted methane~~ evaluate predicted CH_4 concentrations using the WETCHIMP flux estimates (Fig. 1) against observed atmospheric concentrations. The WETCHIMP project was designed to compare simulated wetland distributions and modeled CH_4 fluxes at multi-year, continental scales (Melton et al., 2013; Wania et al., 2013) . The project entailed several sets of model runs, but Melton et al. (2013) primarily reported on one set of runs – runs for 1901–2009 that used the same observed climate and CO_2 concentration datasets to force all models. Each CH_4 model utilized its own parameterization for wetland area and distribution. We use the outputs from this set of model runs in the present study. Of the WETCHIMP models, seven provide a~~flux estimate~~ flux estimate on a suitable time step for boreal North America and six provide an estimate for temperate North America. These models include CLM4Me (Riley et al., 2011), DLEM (Tian et al., 2010), LPJ-Bern (Spahni et al., 2011), LPJ-WHyMe (Wania et al., 2010), LPJ-WSL (Hodson et al., 2011), ORCHIDEE (Ringeval et al., 2010), and SDGVM (Singarayer et al., 2011). All ~~model outputs have a~~ flux model outputs used from the WETCHIMP study have a temporal resolution of one month. These models are described in ~~Melton et al. (2013) , Wania et al. (2013) ,~~ Melton et al. (2013); Wania et al. (2013) , and the Supplement.

2.2 ~~Model selection framework~~Synthetic data experiments

~~This study employs two synthetic data experiments to explore the sensitivity of atmospheric observations to wetland fluxes: can the observations detect an atmospheric pattern from wetlands fluxes over distracting patterns from anthropogenic emitters? If yes, can the observations detect spatial variability in wetland fluxes from different regions? We build a modeling approach based upon the BICto answer these questions.~~

~~The BIC is a model selection technique, and various forms of the BIC are used widely in statistical regression analysis (e.g., ?Ramsey and Schafer, 2012) . It scores all possible combinations of explanatory variables based on model-data fit, and it penalizes combinations that have a greater number of variables~~ We assess the ability of the CH_4 observation network to detect wetland fluxes and use a model selection framework adapted from the BIC. ~~The best combination or candidate model has the lowest BIC score~~A model selection framework can sort through a large number of potential explanatory variables and will choose the smallest set of variables that best describe the dataset of interest (e.g., Ramsey and Schafer, 2012) . In the current setup, we generate synthetic atmospheric CH_4 observations. The model selection framework then indicates whether a wetland model and/or an anthropogenic emissions inventory are necessary to describe variability in these observations. In this way, model selection can indicate the sensitivity of the observation network to wetland CH_4 fluxes.

We use a form of the BIC that has been adapted for use within a geostatistical inverse modeling framework~~(e.g., ?Miller et al., 2014) .~~. This setup has previously been used to select either bottom-up models or environmental drivers of CO_2 and CH_4 fluxes (e.g., Mueller et al., 2010; Yadav e The implementation here ~~parallels~~ mirrors that of Fang et al. (2014)~~and Shiga et al. (2014) :~~, Shiga et al. (2014) , and Fang and Michalak (2015) :

$$\mathrm{BIC} = \underbrace{\ln|\boldsymbol{\Psi}| + (\boldsymbol{z} - \mathbf{H}\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{\Psi}^{-1}(\boldsymbol{z} - \mathbf{H}\mathbf{X}\boldsymbol{\beta})}_{\text{negative log-likelihood}} + \underbrace{p\ln(n)}_{\text{penalty term}} \qquad (1)$$

The first ~~term~~ two terms in Eq.~~1 is~~ (1) are the negative log-likelihood, a measure of how well the model fits the data. ~~In that term ,~~The last term penalizes a particular model based upon the number of explanatory variables (p). The best combination or candidate model has the lowest BIC score.

The variable $\boldsymbol{z}$ $(n \times 1)$ represents the observations minus background concentrations, $\mathbf{H}$ $(n \times m)$ the footprints, ~~$\mathbf{X}$~~ and $\boldsymbol{\Psi}$ (~~$m \times p$) a matrix of p explanatory variables, $\boldsymbol{\beta}$ $(p \times 1)$ a set of coefficients assigned to those variables, and $\boldsymbol{\Psi}$ (~~$n \times n$) a covariance matrix derived from an atmospheric inversion framework. ~~The data ($\boldsymbol{z}$), footprints ($\mathbf{H}$), and parameters that define the covariance matrix ($\boldsymbol{\Psi}$) are taken from~~ These variables are based upon two existing inverse modeling studies by Miller et al. (2013) and Miller et al. (2014) (refer to the Supplement). The ~~second term in Eq. 1 penalizes the BIC score of a particular model based upon the number of explanatory variables~~ ~~(p) .~~

~~We employ this model selection framework to understand which explanatory variables~~matrix $\mathbf{X}$ ($m \times p$) contains p explanatory variables. In the current setup, $\mathbf{X}$ can include a wetland flux estimate and/or individual emissions sources from an anthropogenic ~~emissions inventoryand from the WETCHIMP ensemble are required to describe either synthetic or real methane data at North American observation sites.~~

2.3 ~~Synthetic data experiments~~

~~The experiments described in this section use synthetic~~ inventory. $\boldsymbol{\beta}$ ($p \times 1$) is a set of coefficients that scale the variables in $\mathbf{X}$. We set these coefficients to one in the the synthetic data ~~generated at each of the observation sites. We use~~ experiments. As a result, the model selection framework cannot scale other variables in $\mathbf{X}$ to reproduce the atmospheric CH_4 signal from wetlands.

The first experiments described here use synthetic atmospheric CH_4 data. We generate the synthetic data using one of the WETCHIMP models and the anthropogenic emissions estimates ~~for the US and Canada~~ from Miller et al. (2013) and Miller et al. (2014)~~, respectively, and use one of the WETCHIMP models as the wetland flux estimate~~. For consistency among the synthetic datasets, we scale the annual HBL CH_4 budget in each WETCHIMP model to match the overall magnitude estimated by several top-down studies (Pickett-Heaps et al., 2011; Miller et al., 2 We then multiply these fluxes by the footprints ($\mathbf{H}$~~to create the synthetic data at the measurement locations. We further add in~~) and add error that is randomly generated from the covariance matrix ~~$\boldsymbol{\Psi}$~~ ($\boldsymbol{\Psi}$). This covariance matrix represents errors in atmospheric transport and in the measurements – ~~error that represents uncertainties in the fluxes, the measurements, and the atmospheric transport model, among other error sources (refer to the Supplement).~~ collectively referred to as model-data mismatch. This matrix also represents uncertainties in the prior flux estimate. In a geostatistical inverse model, this prior flux model is given by $\mathbf{X}\boldsymbol{\beta}$ (refer to the supplement for more detail).

~~The synthetic experiments ask progressively demanding questions that test the limits of available data. In experiment one, we examine whether methane observations can detect patterns in the atmosphere due to wetland fluxes from different regions. When given multiple possible explanatory variables (including data from the EDGAR anthropogenic emissions inventory), will the model selection framework choose a wetland estimate? If yes, the observations can identify a pattern in atmospheric methane due to wetland fluxes and that pattern is large enough to be visible over other signals in the atmosphere. If not, then either the contribution of wetlands at that site is small, or the observations cannot differentiate atmospheric patterns due to wetlands over other atmospheric patterns due to anthropogenic sources or model-measurement errors. This setup follows Shiga et al. (2014) , who explored the detectability of atmospheric patterns from anthropogenic CO_2 emissions.~~

~~For this test, we generate the synthetic data using one of the WETCHIMP models. We then allow the model selection framework to select wetland fluxes and/or the EDGAR data used to generate the synthetic fluxes.~~ We divide the WETCHIMP wetland fluxes into four regions (Fig. 2) and four seasons (~~winter, spring, summer, fall~~DJF, MAM, JJA, and SON). The model selection ~~can choose none, some, or all of these sixteen wetland variables~~ framework then chooses variables that are necessary to reproduce the synthetic data, variables that include EDGAR and the 16 wetland flux maps. We run this model selection experiment 1000 times, generating new synthetic data each time, and calculate the percentage of all trials in which the model selection chooses a wetland model. ~~In this experiment, the coefficients (β)are fixed to one. Note that several of the WETCHIMP models overestimate the magnitude of fluxes (Sect. 4.2), so we only use models with asmaller magnitude to generate the synthetic data in this experiment (CLM4Me, DLEM, SDGVM, andLPJ-WSL).~~

~~In experiment two, we investigate whether the observation network is sensitive to spatial variability in the wetland fluxes, independent of magnitudeor seasonality~~ The 1000 repeats are needed due to the random or stochastic nature of the synthetic data experiment; the results of the model selection can vary slightly, depending on the particular random errors that we generate based upon the covariance matrix (Ψ). ~~In this setup , we do not fix the~~ This procedure ensures that the model selection results are not the output of a single realization. We then report on how frequently each of the 16 wetland flux maps is chosen. If a wetland flux map is chosen with high frequency, then a wetland flux map is necessary to describe variability in the synthetic CH_4 observations, and the synthetic observation network can detect aggregate wetland CH_4 fluxes from the given region and season. This setup mirrors that of Shiga et al. (2014) , who employed a model selection framework to explore the detectability of anthropogenic CO_2 emissions.

We also explore why the synthetic CH_4 observations may not be able to detect wetland fluxes. We run a series of case studies and in each case remove a different confounding factor that might prevent the network from detecting wetland CH_4 fluxes. In one case, we remove anthropogenic emissions. In subsequent cases, we remove model-data mismatch errors and/or prior flux errors. In each case, we re-run the model selection experiment and examine whether the results improve when each of these confounding factors is removed.

2.3 Real data experiments

This paper subsequently compares the spatial distribution, magnitude, and seasonality of each WETCHIMP estimate against real atmospheric CH_4 observations, using the synthetic experiments to guide the analysis.

We first explore the spatial distribution of the WETCHIMP flux estimates. We modify the model selection setup in Sect. 2.2 to focus on the spatial distribution of each estimate using a procedure developed by Fang et al. (2014) and Fang and Michalak (2015) . Instead of fixing the coefficients ($\boldsymbol{\beta}$) ~~but rather estimate coefficients that minimize the log-likelihood in Eq. 1.~~ to one, we instead estimate the coefficients using real atmospheric CH_4 observations. We also include ~~a spatial constant or intercept term in **X** that can change by month.~~ an intercept term that can vary by month; the intercept for each month is represented by a vector of ones in the matrix $\mathbf{X}$, and this intercept is always included within $\mathbf{X}$. We then run model selection using real observations. As a result of this setup, ~~the magnitude and~~ a wetland model is not necessary to reproduce either the magnitude or seasonality of the ~~intercept can be adjusted to match the data, but any spatial variability in the fluxes~~ atmospheric CH_4 data; the model selection framework can simply scale the intercept term or scale EDGAR to reproduce the magnitude or seasonality of the observations. The spatial distribution of wetland fluxes, however, can only come from ~~the~~ a wetland model. ~~As in experiment one, the model selection framework can choose among 16 wetland variables – variables that represent different regions and seasons. If model selection chooses a wetland variable, then the spatial distribution in that variable is necessary to reproduce the synthetic data. If not , then the observations are not sensitive to~~ The model selection procedure will only choose a wetland model if the spatial distribution of that model describes sufficient, additional variability in the observations (e.g., Fang et al., 2014) .

Model selection can therefore indicate which WETCHIMP models have the best spatial distribution relative to the atmospheric observations; any WETCHIMP model chosen by model selection must have a spatial ~~variability in wetland fluxes for that region /season. This approach follows that of Fang et al. (2014) , who employed a model selection framework to evaluate~~ distribution that improves model-data fit, and the model must improve that fit more than the penalty term in Eq. 1. A negative result does not necessarily indicate that a WETCHIMP model has a poor spatial distribution. In that case, the observations may not be very sensitive to the spatial distribution of fluxes for the given region or given season. Similarly, the spatial distribution ~~of biospheric CO_2 flux models.~~ in a WETCHIMP model may improve model-data fit but not by more than the penalty term in Eq. 1. By contrast, a positive result indicates that a WETCHIMP model has a particularly good spatial distribution. As in Sect. 2.2, we divide the wetland fluxes into four sub-continental regions and four seasons. The Supplement describes this setup in greater detail.

2.4 ~~Real data experiments~~

~~If experiment two is successful on synthetic data, we then apply the experiment to real data. We use the model selection framework to determine which, if any, bottom-up models have a spatial distribution that can describe the methane observations more effectively than a spatial constant.~~

We ~~also include a number of model-data time series to evaluate both~~ then analyze the magnitude and seasonality of the ~~fluxes~~WETCHIMP fluxes using a number of model-data time series. We model ~~methane~~ CH_4 concentrations at a number of US and Canadian observation sites using the WRF-STILT ~~, WETCHIMP, and~~ model, the WETCHIMP estimates, and the EDGAR v4.2FT2010 ~~(Olivier and Janssens-Maenhout, 2012; European Commission, Joint Research Centre (J~~ inventory (Olivier and Janssens-Maenhout, 2012; European Commission, Joint Research Centre (JRC)/ 2013). We average the observations and model output at the monthly scale and then compare the magnitude of these model estimates for each month against the averaged observations.

~~Several studies indicate that EDGAR may underestimate emissions in certain regions of the US and Canada (e.g., Kort et al., 2008; Miller et al., 2013, 2014; Wecht et al., 2014) . Therefore, we scale the magnitude of EDGAR v4.2FT2010 to match wintertime observations (November–April) at each site using a standard major axis (SMA) regression. During those months, fluxes from wetlands are small and any model biases are likely due to anthropogenic emissions. We then apply this scaling factor, estimated for each site from winter data, to anthropogenic emissions in all seasons. Miller et al. (2013) found that anthropogenic emissions in the US lack significant seasonality, so the wintertime scaling factors should be applicable to other seasons.~~

We further compare the seasonality of existing bottom-up models against the seasonality of a recent inverse modeling estimate by Miller et al. (2014). We plot the monthly ~~methane budget as a fraction of the annual total~~ budgets for both the bottom-up models and the inversion estimate~~. We only conduct this analysis for wetland flux regions that are visible to the observation network (synthetic experiments one and two)~~, and we plot the monthly CH_4 budget as a fraction of the annual total.

Note that inter-annual variability in existing ~~methane~~ CH_4 flux models is small relative to the differences among these models; as a result, conclusions from the ~~2-year~~ 2 year study period (2007–2008) likely hold for other years. For example, the inter-annual variability in the total US/Canadian budget is ~~$\pm 7.3 - 9.7\%$~~ ± 7.3–$9.7\,\%$ (standard deviation), depending upon the model in question (Note that LPJ-Bern has even larger inter-annual variation due to an issue with model spin-up described in ~~Wania et al. (2013) .~~ Wania et al., 2013).

3 Results and discussion: synthetic experiments

The synthetic experiments presented here explore the limits of existing atmospheric data for constraining wetland fluxes. ~~We first leverage synthetic data to examine whether the atmospheric~~

~~observation sites can distinguish an atmospheric pattern from wetland fluxes above other patterns due anthropogenic emissions or simulated model, measurement, and emissions uncertainties.~~ If atmospheric observations are to constrain wetland ~~methane~~ CH_4 fluxes, those observations must ~~, at minimum, identify an atmospheric pattern from wetland fluxes from other distracting patterns in the model and/or data~~be able to detect wetland CH_4 fluxes above errors in the transport model and above other emissions sources such as fossil fuels and agriculture.

The ~~results of this experiment are summarized in Fig. 3a. The~~ four columns in Fig.~~3a~~ 3a display the results from an individual season in each of four geographic regions. In this experiment, the ~~observation network can detect a summertime methane pattern from wetlands in both Eastern and Western Canada in~~ synthetic CH_4 observations can detect aggregate wetland CH_4 fluxes from Eastern Canadian wetlands in greater than 75 % of all trials for the summer and fall seasons. In the eastern US, the model selection framework chooses a wetland model in ~~50–75~~25–50 % of all trials in multiple different seasons. By contrast, the ~~observations~~ synthetic CH_4 data are least sensitive to wetland fluxes in the western US, and the model selection framework chooses wetland fluxes from that region in ~~less~~ fewer than 25 % of all trials irrespective of the season. This result may be due, in part, to the ~~relatively dry climate and scant wetlands~~ scant wetlands and sparse atmospheric observations in much of the west. ~~The methane signal from resource extraction and/or agriculture may also overshadow any patterns from wetlands.~~

The results also ~~contain a number of seasonal trends~~vary by season. Of any region, the ~~observation~~ atmospheric CH_4 network is best able to constrain ~~the seasonal cycle~~ fluxes across multiple seasons in eastern Canada. The largest wetland fluxes ~~estimated for the US and Canada~~ in the WETCHIMP models are in Ontario and Quebec ~~.~~ (Fig. 1). It is therefore unsurprising that the network is ~~so sensitive to fluxes from this~~ best able to detect wetland fluxes in that region, even though there are relatively few observation sites in the area. In other regions, the ~~observation~~ atmospheric CH_4 network is less sensitive to wetlands during the winter, fall, and spring ~~seasons. For example, the model selection framework chooses a wetland model in less than 25~~shoulder seasons.

We run several additional test scenarios to explore why the synthetic observations may not always be able to detect wetland CH_4 fluxes. We remove anthropogenic emissions from the synthetic dataset for the experiment in Fig. 3b. We remove all model data mismatch errors in Fig. 3c; model-data mismatch encompasses errors in atmospheric transport and in the measurements. Subsequently, we remove all errors due to the prior flux estimate in Fig. 3d. In Fig. 3e, we remove both types of errors. In each case, we re-run the model selection experiment to see if the sensitivity of the atmospheric CH_4 network to wetland fluxes improves.

Anthropogenic emissions have only a modest effect on the results in specific regions and seasons. In case (b) without anthropogenic emissions, the results improve by ∼25–50% ~~of all trials during the winter in all regions.~~ in the fall and spring shoulder seasons for several geographic regions.

~~The density of the observation network may also play a role in these results. Wetlands in the Eastern US are sparse relative to Canada, but the higher density of observations in the Eastern US may contribute to a moderate success rate (> 50%) for that region. A recent observation network expansion could play a key role in future efforts to constrain wetland fluxes across these regions . Environment Canada has recently been expanding their observation network across western and Arctic Canada (i. e. , Saskatchewan, Alberta, Northwest Territories, and Nunavut). In addition, Earth Networks has begun to install new observation sites across the eastern US in a privately-funded initiative.~~ By contrast, the model-data mismatch and prior flux errors have a much larger effect on the model selection results. The results improve incrementally across many regions and seasons when we remove model-data mismatch errors in case (c). The results improve across the spring, summer, and fall seasons and improve across all four geographic regions. However, the magnitude of this improvement is never more than 25%. Model-data mismatch errors are likely dominated by errors in modeled atmospheric transport. These results imply that transport errors play an incremental yet pervasive role in the utility of the atmospheric observations.

~~Compared to experiment one, the second experiment asks a more demanding question of the observation network: is the observation network sensitive to spatial variability in~~ The prior flux errors have the largest effect on the results, particularly during the warmest seasons. In case (d), the results show great improvement during fall, spring, and summer and show little improvement during winter or in the western US. In the setup here, the ~~wetland fluxes from~~

~~each region? Alternately, can a spatially-constant model reproduce the synthetic atmospheric observations as well? Existing bottom-up estimates disagree markedly on the spatial distribution of wetland fluxes, but perhaps atmospheric data can provide guidance~~prior flux uncertainties scale with the seasonal magnitude of the fluxes. When we remove the prior flux errors, the results concomitantly show the greatest improvement in seasons that have larger overall CH_4 fluxes. These results indicate that the prior estimate greatly impacts the utility of the atmospheric CH_4 observations. A geostatistical inverse model can leverage any combination of land surface maps, meteorological maps, and/or anthropogenic inventory estimates in the inversion prior. These maps or estimates are incorporated into the $\mathbf{X}$ matrix in Eq. 1. If accurate maps or estimates are not available, then the prior flux errors will be large, and the model selection framework will be less likely to choose any particular variable. If these maps or estimates have high explanatory power, then the prior flux errors will be small, and the model selection framework will be more likely to choose any one variable. As a result, the detectability of wetland CH_4 fluxes partly depends on the availability of land surface or meteorological data that matches those fluxes. The atmospheric network can better detect wetland CH_4 fluxes when accurate prior information can guide that identification.

~~Figure 3b displays the results of this experiment for each region and season. The available data is only sensitive to spatial variability in certain cases. The model selection framework~~ Case (e) (no model-data mismatch errors and no errors in the prior flux estimate) shows large, ubiquitous improvements; the model selection chooses a wetland model ~~in $>75\%$ of all trials in eastern Canada during summer and fall and in western Canada during summer. In remote regions of northern Ontario and Quebec, large wetland fluxes dominate variability in atmospheric methane. Hence, it is understandable that observations are most sensitive to the spatial distribution of fluxes in this region. By contrast, the observation network is largely insensitive to spatial variability in wetland fluxes across the US; in most instances, the model selection framework favors aspatially-constant model over a wetland model for the two US~~ 100% of the time in almost all regions and seasons. The results for Eastern Canada during winter are the exception. In winter, the wetland model cannot always explain enough variability in the synthetic observations to overcome the BIC penalty term in Eq. 1.

The density of the atmospheric CH_4 network may also play a role in these results. Wetlands in the Eastern US are sparse relative to eastern Canada, but the higher density of observations in the Eastern US may contribute to a moderate success rate (25 – 50 %) for that region. Recent and planned network expansions in the eastern US and in Canada could play a key role in future efforts to constrain wetland fluxes across these regions.

~~These~~ Overall, the synthetic experiment results indicate that the observation network ~~has limited capacity to evaluate wetland fluxes over the United States.~~cannot detect wetland fluxes from the US (i.e., model selection has a success rate <50%). Across Canada, the results are ~~far more promising ,~~ more promising (i.e., near 100% success rate in some regions/seasons), despite the relative sparsity of the observation network there. ~~Existing bottom-up methane estimates are highly uncertain across Canada, and the synthetic experiments indicate that atmospheric observations can reduce these uncertainties.~~

4 Results and discussion: comparisons with atmospheric data

4.1 Spatial ~~flux patterns~~distribution of the fluxes

We ~~first~~ compare the spatial distribution of the ~~existing wetland~~ WETCHIMP flux estimates against ~~methane~~ CH_4 data from the atmospheric observation network. ~~We apply experiment two to real data and report the results for regions and seasons that had a high success rate in the synthetic experiment. That experiment examined whether the spatial variability in a wetland model is more useful at describing the atmospheric data than a spatial constant. We now apply this question to real data: do the WETCHIMP models have spatial variability that describe the real data better than a spatial constant? If so, which models? This approach indicates whether each model contributes positive information on the location of wetland fluxes~~To this end, we use a version of the model selection framework that chooses wetland models based upon their spatial distribution (Fang et al., 2014; Fang and Michalak, 2015) . WETCHIMP models that are chosen by the framework have a spatial distribution that is more consistent with atmospheric observations relative to those that are not selected.

The results of this ~~real data experiment~~ model selection analysis are displayed in Table 1. This table ~~only~~ lists the regions and seasons that had a~~success rate >75in~~ success rate $> 50\,\%$ in the synthetic data experiment~~two. If a wetland model describes the distribution of fluxes better than a spatial constant~~ ; the atmospheric CH_4 network is most sensitive to wetland CH_4 fluxes in those regions ~~/seasons, then the model selection framework should select that model.~~ ~~Only a small number of WETCHIMP models are able to describe the distribution of wetland fluxes (as seen via the atmospheric observations) better than a spatial constant – between 0 – 28of the available models depending upon the region and season. The~~ and seasons. Two of the WETCHIMP models were chosen by the model selection framework ~~chooses~~ – LPJ-Bern (in eastern Canada~~and LPJ-Bern and SDGVM in western Canada~~) and SDGVM (in eastern and western Canada). The spatial ~~patterns in the remaining WETCHIMP models do not perform better than a spatial constant when compared to atmospheric data.~~ distribution of these models improve the model-data fit more than the penalty term in Eq. 1.

The LPJ-Bern and SDGVM models have several unique spatial characteristics that could explain these results. Over eastern Canada, ~~the~~ LPJ-Bern ~~model concentrates the largest~~ and SDGVM concentrate the large fluxes in the HBL. Other models, by contrast, often distribute the fluxes more broadly across Ontario and Quebec or put the largest fluxes in Ontario outside of the HBL. In western Canada, ~~the LPJ-Bern and SDGVM models distribute fluxes broadly across both northern~~SDGVM distributes fluxes across northern, boreal Saskatchewan and Alberta. ~~A number of other estimates like DLEM or CLM4Me assign relatively small fluxes in these provinces relative to other regions.~~

The LPJ-Bern and SDGVM models share another common characteristic: both model wetland area independently instead of relying solely on remote sensing inundation datasets. ~~LPJ-WSL~~LPJ-ORCHIDEE, DLEM, and CLM4Me use remote sensing inundation datasets like ~~GEIMS~~ GIEMS (Global Inundation Extent from Multi-Satellites, Prigent et al., 2007) to construct a wetland map. Other models, like LPJ-Bern ~~,~~ and LPJ-WHyMe ~~, and SDGVM~~ also use land cover maps and/or land surveys to estimate wetland (or at least ~~methane-producing) area~~CH_4-producing) area. SDGVM estimates this area dynamically as a function of soil moisture (Melton et al., 2013; Wania et al., 2013). Wetland maps generated using ~~the two~~ these different approaches show substantial differences. Remote sensing datasets estimate relatively high levels of inundation in regions of Canada that are ~~non-forested~~ not forested or have many small lakes ~~(see further discussion in Melton et al., 2013; Bohn et al., 2015) . Independently-generated wetland maps, by contrast,~~ (see further discussion in Melton et al., 2013; Bohn et al., 2015) . By contrast, modeling approaches that dynamically generate wetland area or use land cover maps assign more wetlands over regions with high water tables but little surface water as seen by remote sensing based inundation datasets). As a result of these differences, models like LPJ-Bern assign more wetlands and ~~methane~~ CH_4 fluxes in the HBL relative to other regions of eastern Canada.

4.2 Flux magnitude

We next compare the magnitude of predicted concentrations using the WETCHIMP models against atmospheric observations at individual locations. Unlike previous sections that utilized model selection, this section employs several ~~simple~~ model-data ~~timeseries~~time series, displayed in Fig.~~4. We model methane concentrations at a number of US and Canadian observation sites using WRF-STILT, the WETCHIMP flux estimates, and anthropogenic emissions from the EDGAR v4.2FT2010 inventory.~~ 4. This model estimate consists of several components: the background (in green) is the estimated ~~concentration of methane~~ background concentration of CH_4 in clean air before entering the model domain as in ~~Miller et al. (2013) and Miller et al. (2014)~~ Mi The estimated contribution of anthropogenic emissions from EDGAR v4.2FT2010 is added to this background (in red). ~~Note that the estimated scaling factors for EDGAR (Sect. 2.3) are 1.7 ± 0.3 at Chibougamau, 5.6 ± 0.5 at East Trout Lake, 2.4 ± 0.3 at Fraserdale, and 2.5 ± 0.3 at Park Falls.~~ The contribution of wetland fluxes from the WETCHIMP models is then added to the previous inputs, and the sum of all components (blue lines) can be compared directly against measured concentrations.

The various WETCHIMP flux estimates produce very different modeled concentrations at the atmospheric observation sites (Fig. 4). Overall, modeled concentrations with the WETCHIMP fluxes usually exceed the ~~methane~~ CH_4 measurements during summer. At Chibougamau, Fraserdale, and Park Falls in early summer, all six WETCHIMP models predict ~~methane~~ CH_4 concentrations that equal or exceed the observations. The ORCHIDEE, LPJ-WHyMe, and LPJ-Bern models always exceed the measurements during summer while DLEM and SDGVM better match the observations at these sites. ~~In contrast to these results, a recent study by Bohn et al. (2015) found that the ensemble average is not biased over the Western Siberian Lowlands relative to inverse modeling estimates. The models also show a large spread in that region.~~ Notably, a number of previous studies report that the EDGAR inventory may underestimate US anthropogenic CH_4 emissions (e.g., Kort et al., 2008; Miller et al., 2013; Wecht et al., 2014; Turner et al., 2015) . If EDGAR underestimate emissions, then the WETCHIMP models would be an even larger overestimate relative to the atmospheric data.

~~Methane~~ CH_4 models that overestimate fluxes in North America do not always compensate with smaller fluxes elsewhere. For example, the ORCHIDEE model not only estimates large fluxes over North America but also estimates higher fluxes over the tropics than any other model (Melton et al., 2013). Hence, the disagreement in magnitude over North America not only reflects uncertainty in the global distribution of wetland fluxes but also reflects uncertainty in the global wetland budget.

4.3 Seasonal cycle

Bottom-up ~~methane~~ CH_4 flux estimates show variable ~~performance when compared against~~ features when compared to atmospheric observations, and the ~~temporal distribution~~ seasonal cycle of these estimates is no exception. Figure 5 compares the seasonal cycle of the existing estimates over Canada's HBL. Eastern Canada is one of the largest wetland regions in North America (Fig. 1), and ~~unlike other regions, the observation network there can detect a clear wetland signal through most of the seasonal cycle~~ nearby atmospheric observation sites see a much larger CH_4 enhancement from wetlands relative to other regions (Fig. ~~3~~4 and S4).

In this region, the bottom-up estimates diverge on the seasonal cycle of fluxes. Most estimates predict peak fluxes in July or August, though two variations of the LPJ model predict seasonal peaks in September and October ~~(~~ – LPJ-WHyMe and LPJ-Bern, respectively~~). Discrepancies~~ . LPJ-WHyMe is a module inside of LPJ-Bern, a possible explanation for the similar seasonal cycle in these two models. Differences among models are also notable during the fall and spring seasons. For example, fluxes in June account for anywhere between 6 and 21 % of the annual ~~methane~~ CH_4 budget, depending upon the model. Fluxes in October account for between 1 and 23 % of the annual budget ~~.~~ (Fig. 5b).

~~The figure~~ Figure 5 also displays the seasonality of an inverse modeling estimate from Miller et al. (2014) for comparison. That estimate incorporates observations from Chibougamau, Quebec, and Fraserdale, Ontario, atmospheric measurement sites that are strongly influenced by fluxes from the HBL. The discrepancies among the WETCHIMP models often exceed the 95 % confidence interval of the inversion estimate. The WETCHIMP estimates are often comparable to Miller et al. (2014) in magnitude during fall and spring months but exceed the inversion estimate in summer months (Fig. 5a). On whole, the WETCHIMP ~~estimates have anarrower~~ models have a narrower relative seasonal cycle than the inversion estimate ~~, which assigns comparatively larger fluxes~~ (Fig. 5b). That estimate assigns a greater portion of the annual budget to the fall and spring shoulder seasons. ~~A~~

Additional top-down studies exist for the HBL, but those studies use a seasonal cycle drawn from an existing bottom-up model and do not estimate the seasonal cycle independently from CH_4 observations (Pickett-Heaps et al., 2011; Wecht et al., 2014; Turner et al., 2015) . By comparison, a recent inverse modeling study of the Western Siberian Lowlands found parallel results for that region – existing models also ~~under-predict the shoulder seasons relative to summer months~~ predict a seasonal cycle that is narrower than the seasonality implied by atmospheric observations (Winderlich, 2012; Bohn et al., 2015).

Numerous possible explanations could underly ~~this discrepancy~~differences in the seasonal cycle of CH_4 fluxes. For example, the ~~bottom-up models could be too sensitive to soil/air temperature , and may therefore shut off methane emissions too early. Compared to the inversion estimate~~temperature threshold for CH_4 production may be too high in some models. Relative to summer months, the bottom-up models predict small ~~or minimal~~ fluxes during fall/spring months when air temperatures are near freezing but soils are still unfrozen (Fig.~~S3~~ S3). According to estimates from the North American Regional Reanalysis (NARR) (Mesinger et al., 2006), surface soils in the HBL (0 and ~~10cm~~ 10 cm depth) begin to thaw in April and are largely unfrozen in May (Fig.~~S3~~ S3). In the fall, surface soils (0~~cm~~ cm depth) begin to freeze in November, but deeper soils (~~10cm and 40cm~~10 and 40 cm) remain largely unfrozen until December. Compared to the bottom-up models, the inversion estimate predicts a wider seasonal window, a result that ~~is consistent with soil freeze/~~would be consistent with dates of deep soil freeze and thaw.

5 Conclusions

A recent model comparison study revealed ~~substantial differences in existing~~ wide differences among several estimates of wetland ~~methane~~ CH_4 fluxes – differences at global to regional scales. In the first component of this study, we use ~~two increasing stringent synthetic data experiments to understand how sensitive~~ a synthetic data experiment to understand whether the atmospheric observation network ~~is to regional-scale wetland~~ can detect wetland CH_4 fluxes. We find that the network can reliably identify ~~an atmospheric pattern from Canadian wetlands~~aggregate wetland fluxes from both eastern and western Canada. The network can ~~identify a methane pattern~~ detect wetland fluxes from the eastern US in ~~50 – 75%~~ a smaller fraction of trials and rarely from the western US. ~~The network can also detect spatial variability in the Canadian wetland source but rarely in the US wetland source.~~ This analysis also accounts for distracting signals ~~or patterns~~ in the atmosphere from anthropogenic sources or simulated ~~modeling errors. These results indicate that uncertainties in current methane models can be reduced, if those models begin to leverage available methane data. Furthermore, these discrepancies indicate a disconnect between scientists who build process-based and/or biogeochemical models and scientists who collect or use atmospheric methane data. Improved collaboration between these two groups could help reduce present uncertainties in natural methane fluxes, at least over Canada.~~ atmospheric transport errors.

In a second component of the study, we ~~evaluate each existing~~ analyze each bottom-up ~~methane model at regional scale~~ CH_4 model from the WETCHIMP study using real atmospheric data. We find that ~~only 0–28of all models have a spatial pattern that describes the atmospheric data more effectively than a constant. The~~ the LPJ-Bern and SDGVM models have spatial distributions that are most consistent with atmospheric observations, depending upon the region and season of interest. In addition, almost all ~~existing~~ models overestimate the magnitude of wetland ~~methane~~ CH_4 fluxes when compared against atmospheric data at individual observation sites. The ~~ensemble of models also appears to estimate aseasonal cycle~~ model ensemble may also estimate a seasonal cycle for eastern Canada that is too narrow ~~across the HBL, a large region of methane fluxesin North America. Overall, this study indicates numerous areas for improvement in existing~~ (i.e., places too much of the total annual flux in the summer relative to the fall and spring shoulder seasons).

The results of this paper suggest possible pathways to improve future top-down estimates of wetland CH_4 fluxes. The ability of the atmospheric observation network to detect wetland fluxes depends in large part upon the prior flux model. In a geostatistical inverse model, this model can incorporate land surface maps – wetland maps, estimates of land surface processes, and maps of anthropogenic emissions sources. This information plays a large role in whether atmospheric observations can detect wetland fluxes; the observations can more adeptly identify wetland fluxes when accurate land surface maps are available to guide that identification. By contrast, atmospheric transport and measurement errors (i.e., model-data mismatch errors) have a ubiquitous but smaller effect on the utility of atmospheric CH_4 observations.

The results presented here also hold a number of suggestions for future bottom-up ~~wetland methane estimates~~ modeling efforts:

1. Spatial distribution: Bottom-up estimates that use surface water inundation as the only proxy for wetland area do not perform as well relative to atmospheric observations. Bottom-up models that use satellite inundation data should incorporate additional tools like wetland mapping or dynamic modeling to capture wetlands covered by vegetation.

2. Magnitude: Existing top-down studies that use a diverse array of in situ and satellite CH_4 observations show good agreement on the magnitude of CH_4 fluxes from the Hudson Bay Lowlands (HBL) region (e.g., Pickett-Heaps et al., 2011; Miller et al., 2014; Wecht et al., 2014; These studies could be used to calibrate the magnitude of future bottom-up estimates, at least over the HBL where CH_4 observations provide a strong constraint on wetland fluxes.

3. Seasonal cycle: Bottom-up models do not show consensus on the seasonal cycle of wetland fluxes across Canada. Few top-down studies estimate the seasonal cycle independently using atmospheric observations. Additional top-down studies would indicate the range of seasonal cycle estimates that are consistent with atmospheric observations, particularly studies that use a diverse set of atmospheric models and/or diverse observational datasets. These efforts could help reconcile differences in the seasonal cycle among bottom-up models and between bottom-up models and the few, existing top-down studies.

These steps will hopefully lead to better convergence among wetland CH_4 estimates for North America.

The Supplement related to this article is available online at doi:10.5194/bgd-0-1-2015-supplement.

Acknowledgements. We thank Marc Fischer and Sebastien Biraud of Lawrence Berkeley Labs and John Miller of NOAA; these collaborators operate several sites in the US greenhouse gas observation network. We also thank Steven Wofsy of Harvard University and Thomas Nehrkorn of Atmospheric and Environmental Research. The National Aeronautics and Space Administration (NASA) Advanced Supercomputing Division provided key computing resources and assistance. This work was supported by the Department of Energy Computational Science Graduate Fellowship Program of the Office of Science and National Nuclear Security Administration in the Department of Energy under contract DE-FG02-97ER25308. This work was also supported by a National Science Foundation Graduate Research Fellowship.

~~urlstyle~~

References

Andrews, A. E., Kofler, J. D., Trudeau, M. E., Williams, J. C., Neff, D. H., Masarie, K. A., Chao, D. Y., Kitzis, D. R., Novelli, P. C., Zhao, C. L., Dlugokencky, E. J., Lang, P. M., Crotwell, M. J., Fischer, M. L., Parker, M. J., Lee, J. T., Baumann, D. D., Desai, A. R., Stanier, C. O., De Wekker,~~S.~~

S. F. J., Wolfe, D. E., Munger, J. W., and Tans, P. P.: CO_2, CO, and CH_4 measurements from tall towers in the NOAA Earth System Research Laboratory's Global Greenhouse Gas Reference Network: instrumentation, uncertainty analysis, and recommendations for future high-accuracy greenhouse gas monitoring efforts, ~~Atmospheric Measurement Techniques~~Atmos. Meas. Tech., 7, 647–687, doi:10.5194/amt-7-647-2014, 2014.

Bohn, T. J., Melton, J. R., Ito, A., Kleinen, T., Spahni, R., Stocker, B. D., Zhang, B., Zhu, X., Schroeder, R., Glagolev, M. V., Maksyutov, S., Brovkin, V., Chen, G., Denisov, S. N., Eliseev, A. V., Gallego-Sala, A., McDonald, K. C., Rawlins, ~~M.~~ M.A., Riley, W. J., Subin, Z. M., Tian, H., Zhuang, Q., and Kaplan, J. O.: WETCHIMP-WSL: intercomparison of wetland methane emissions models over West Siberia, Biogeosciences, 12, 3321–3349, doi:10.5194/bg-12-3321-2015, 2015.

Butler, J.: The ~~W~~NOAA ~~est~~ annual greenhouse gas index ~~S~~(AGGI)~~iberia, BiogeosciencesDiscussions~~, ~~12~~http://www.esrl.noaa.gov/gmd/aggi/, ~~1907–1973, , 2015.~~ 2014.

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., and Canadell, J.: Carbon and Other Biogeochemical Cycles ~~-~~ – Final Draft Underlying Scientific Technical Assessment, chap. 6, IPCC Secretariat, Geneva, 2013.

European Commission, Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL): Emission Database for Global Atmospheric Research (EDGAR), release EDGARv4.2 FT2010, available at: http://edgar.jrc.ec.europa.eu (last access: 14 June 2015), 2013.

Fang, Y. and Michalak, A. M.: Atmospheric observations inform ~~CO_2~~ CO_2 flux responses to enviroclimatic drivers, Global Biogeochem. Cy., ~~2014GB005034, ,~~ 29, GB005034, doi:10.1002/2014GB005034, 2015.

Fang, Y., Michalak, A. M., Shiga, Y. P., and Yadav, V.: Using atmospheric observations to evaluate the spatiotemporal variability of ~~CO~~CO_2 fluxes simulated by terrestrial biospheric models, Biogeosciences, 11, 6985–6997, doi:10.5194/bg-11-6985-2014, 2014.

Gourdji, S. M., Mueller, K. L., ~~Schaefer, K., and Michalak,~~ Yadav, V., Huntzinger, D. N., Andrews, A. ~~M.: Global monthly averaged~~ E., Trudeau, M., Petron, G., Nehrkorn, T., Eluszkiewicz, J., Henderson, J., Wen, D., Lin, J., Fischer, M., Sweeney, C., and Michalak, A. M.: North American CO~~2 fluxes recovered using a geostatistical inverse modeling approach: 2. Results including auxiliary environmental data, J. Geophys.Res.-Atmos., 113, D21115, , 2008~~$_2$ ~~.~~ exchange: inter-comparison of modeled estimates with results from a fine-scale atmospheric inversion, Biogeosciences, 9, 457–475, doi:10.5194/bg-9-457-2012, 2012.

Hegarty, J., Draxler, R. R., Stein, A. F., Brioude, J., Mountain, M., Eluszkiewicz, J., Nehrkorn, T., Ngan, F., and Andrews, A.: Evaluation of Lagrangian ~~Particle Dispersion Models with Measurements~~

~~from Controlled Tracer Releases,~~particle dispersion models with measurements from controlled tracer releases, J. Appl. Meteorol. Clim., 52, 2623–2637, 2013.

Hendriks,~~D.~~ D. M. D., van Huissteden, J., and Dolman, A. J.: Multi-technique assessment of spatial and temporal variability of methane fluxes in a~~peat meadow, Agriculture and Forest Meteorology,~~ peat meadow, Agr. Forest Meteorol., 150~~,~~, 757–774~~, 2010.~~, doi:10.1016/j.agrformet.2009.06.017, 2010.

Hodson, E. L., Poulter, B., Zimmermann, N. E., Prigent, C., and Kaplan, J. O.: The El Niño–Southern Oscillation and wetland methane interannual variability, Geophys. Res. Lett., 38, ~~L08810,~~ L08810, doi:10.1029/2011GL046861, 2011.

Kort, E. A., Eluszkiewicz, J., Stephens, B. B., Miller, J. B., Gerbig, C., Nehrkorn, T., Daube, B. C., Kaplan, J. O., Houweling, S., and Wofsy, S. C.: Emissions of ~~CH_4~~CH_4 and ~~N_2O~~N_2O over the ~~United States and Canada~~ United States and Canada based on a receptor-oriented modeling framework and COBRA-NA atmospheric observations, Geophys. Res. Lett., 35, ~~L18808,~~ L18808, doi:10.1029/2008gl034031, 2008.

Lin, J., Gerbig, C., Wofsy, S., Andrews, A., Daube, B., Davis, K., and Grainger, C.: A near-field tool for simulating the upstream influence of atmospheric observations: ~~The~~ the Stochastic Time-Inverted Lagrangian Transport (STILT) ~~model,~~model, J. Geophys. Res.-Atmos., 108, 4493, doi:10.1029/2002jd003161, 2003.

Livingston, G. and Hutchinson, G.: Enclosure-based measurement of trace gas exchange: applications and sources of error, chap. 2, ~~Methods in ecology~~in: Methods in Ecology, Wiley, London, 14–51, 2009.

Melton, J. R., Wania, R., Hodson, E. L., Poulter, B., Ringeval, B., Spahni, R., Bohn, T., Avis, C. A., Beerling, D. J., Chen, G., Eliseev, A. V., Denisov, S. N., Hopcroft, P. O., Lettenmaier, D. P., Riley, W. J., Singarayer, J. S., Subin, Z. M., Tian, H., ~~Zürcher,~~Zürcher, S., Brovkin, V., van~~Bodegom,~~ Bodegom, P. M., Kleinen, T., Yu, Z. C., and Kaplan, J. O.: Present state of global wetland extent and wetland methane modelling: conclusions from a model inter-comparison project (~~WETCHIMP~~WETCHIMP), Biogeosciences, 10, 753–788, doi:10.5194/bg-10-753-2013, 2013.

~~Mesinger,~~

Mesinger, F., ~~Dimego,~~Dimego, G., ~~Kalnay,~~Kalnay, E., ~~Mitchell,~~Mitchell, K., ~~Shafran,~~Shafran, P. C., ~~Ebisuzaki,~~Ebisuzaki, W., ~~Jovi,~~Jovi, D., ~~Woollen,~~Woollen, J., ~~Rogers,~~Rogers, E., ~~Berbery,~~Berbery, E. H., ~~Ek,~~Ek, M. B., ~~Fan,~~Fan, Y., ~~Grumbine,~~Grumbine, R., ~~Higgins,~~Higgins, W., ~~Li,~~Li, H., ~~Lin,~~Lin, Y., ~~Manikin,~~Manikin, G., ~~Parrish,~~Parrish, D., and ~~Shi,~~Shi, W.: ~~North American~~

~~Regional Reanalysis.,~~North American regional reanalysis, B. Am. Meteorol. Soc., 87, 343–360, doi:10.1175/BAMS-87-3-343, 2006.

Miller, S. M., Wofsy, S. C., Michalak, A. M., Kort, E. A., Andrews, A. E., Biraud, S. C., Dlugokencky, E. J., Eluszkiewicz, J., Fischer, M. L., Janssens-Maenhout, G., Miller, B. R., Miller, J. B., Montzka, S. A., Nehrkorn, T., and Sweeney, C.: Anthropogenic emissions of methane in the ~~United States,~~United States, P. Natl. Acad. Sci. USA, 110, ~~20018–20022,~~ 20018–20022, doi:10.1073/pnas.1314392110, 2013.

Miller, S. M., Worthy,~~D.~~ D. E. J., Michalak, A. M., Wofsy, S. C., Kort, E. A., Havice, T. C., Andrews, A. E., Dlugokencky, E. J., Kaplan, J. O., Levi, P. J., Tian, H., and Zhang, B.: Observational constraints on the distribution, seasonality, and environmental predictors of ~~North American~~ North American boreal methane emissions, Global Biogeochem. Cy., 28, 146–160, doi:10.1002/2013GB004580, 2014.

Mueller, K. L., Yadav, V., Curtis, P. S., Vogel, C., and Michalak, A. M.: Attributing the variability of eddy-covariance CO_2 flux measurements across temporal scales using geostatistical regression for a mixed northern hardwood forest, Global Biogeochem. Cy., 24, doi:10.1029/2009GB003642, 2010.

Nehrkorn, T., Eluszkiewicz, J., Wofsy, S. C., Lin, J. C., Gerbig, C., Longo, M., and Freitas, S.: Coupled Weather Research and Forecasting-Stochastic Time-Inverted Lagrangian Transport (WRF-STILT) model, Meteorol. Atmos. Phys., 107, 51–64, doi:10.1007/s00703-010-0068-x, 2010.

Olivier, J. and Janssens-Maenhout, G.: ~~CO~~CO_2 emissions from fuel combustion, 2012 ~~edition~~edn., chap. III~~: Greenhouse-gas emissions~~, in: Greenhouse-Gas Emissions, International Energy Agency, Paris, III.1 – III.51, 2012.

~~Pan,~~

Pan, L. L., ~~Bowman,~~Bowman, K. P., ~~Atlas,~~Atlas, E. L., ~~Wofsy,~~Wofsy, S. C., ~~Zhang,~~Zhang, F., ~~Bresch,~~Bresch, J. F., ~~Ridley,~~Ridley, B. A., ~~Pittman,~~Pittman, J. V., ~~Homeyer,~~Homeyer, C. R., ~~Romashkin,~~Romashkin, P., and ~~Cooper,~~Cooper, W. A.: ~~The Stratosphere-Troposphere Analyses of Regional Transport~~ The stratosphere-troposphere analyses of regional transport 2008 ~~Experiment,~~experiment, B. Am. Meteorol. Soc., 91, 327–342, doi:10.1175/2009BAMS2865.1, 2010.

Pickett-Heaps, C. A., Jacob, D. J., Wecht, K. J., Kort, E. A., Wofsy, S. C., Diskin, G. S., Worthy,~~D.~~ D. E. J., Kaplan, J. O., Bey, I., and Drevet, J.: Magnitude and seasonality of wetland methane emissions from the Hudson Bay Lowlands (Canada), Atmos. Chem. Phys., 11, 3773–3779, doi:10.5194/acp-11-3773-2011, 2011.

Prigent, C., Papa, F., Aires, F., Rossow, W. B., and Matthews, E.: Global inundation dynamics inferred from multiple satellite observations, 1993–2000, J. Geophys. Res.-Atmos., 112, ~~D12107,~~ D12107, doi:10.1029/2006JD007847, 2007.

Ramsey, F. and Schafer, D.: The Statistical Sleuth: A Course in Methods of Data Analysis, Cengage Learning, Boston, ~~3~~ 3rd edn., 2012.

Riley, W. J., Subin, Z. M., Lawrence, D. M., Swenson, S. C., Torn, M. S., Meng, L., Mahowald, N. M., and Hess, P.: Barriers to predicting changes in global terrestrial methane fluxes: analyses using CLM4Me, a methane biogeochemistry model integrated in ~~CESM~~CESM, Biogeosciences, 8, 1925–1953, doi:10.5194/bg-8-1925-2011, 2011.

Ringeval, B., de Noblet-Ducoudré, N., Ciais, P., Bousquet, P., Prigent, C., Papa, F., and Rossow, W. B.: An attempt to quantify the impact of changes in wetland extent on methane emissions on the seasonal and interannual time scales, Global Biogeochem. Cy., 24, GB2003, doi:10.1029/2008GB003354, 2010.

~~Schwarz, G.: Estimating the dimension of a model, Ann. Stat., 6, 461–464, , 1978.~~

Shiga, Y. P., Michalak, A. M., Gourdji, S. M., Mueller, K. L., and Yadav, V.: Detecting fossil fuel emissions patterns from subcontinental regions using ~~North American~~ North American in situ ~~CO_2~~CO_2 measurements, Geophys. Res. Lett., 41, 4381–4388, doi:10.1002/2014GL059684, 2014.

Singarayer, J. S., Valdes, P. J., Friedlingstein, P., Nelson, S., and Beerling, D. J.: Late Holocene methane rise caused by orbitally controlled increase in tropical sources, Nature, 470, 82–85, doi:10.1038/nature09739, 2011.

Spahni, R., Wania, R., Neef, L., van~~Weele,~~ Weele, M., Pison, I., Bousquet, P., Frankenberg, C., Foster, P. N., Joos, F., Prentice, I. C., and van~~Velthoven,~~ Velthoven, P.: Constraining global methane emissions and uptake by ecosystems, Biogeosciences, 8, 1643–1665, doi:10.5194/bg-8-1643-2011, 2011.

Tian, H., Xu, X., Liu, M., Ren, W., Zhang, C., Chen, G., and Lu, C.: Spatial and temporal patterns of ~~CH_4 and N_2O~~ CH_4 and N_2O fluxes in terrestrial ecosystems of ~~North America during 1979-2008~~North America during 1979–2008: application of a global biogeochemistry model, Biogeosciences, 7, 2673–2694, doi:10.5194/bg-7-2673-2010, 2010.

Turner, A. J., Jacob, D. J., Wecht, K. J., Maasakkers, J. D., Lundgren, E., Andrews, A. E., Biraud, S. C., Boesch, H., Bowman, K. W., Deutscher, N. M., Dubey, M. K., Griffith, D. W. T., Hase, F., Kuze, A., Notholt, J., Ohyama, H., Parker, R., Payne, V. H., Sussmann, R., Sweeney, C., Velazco, V. A., Warneke, T., Wennberg, P. O., and Wunch, D.: Estimating global and North American methane emissions with high spatial resolution using GOSAT satellite data, Atmos. Chem. Phys., 15, 7049–7069, doi:10.5194/acp-15-7049-2015, 2015.

Waddington, J. and Roulet, N.: Atmosphere-wetland carbon exchanges: ~~Scale~~ scale dependency of ~~CO_2~~CO_2 and ~~CH_4~~CH_4 exchange on the developmental topography of a~~peatland,~~ peatland, Global Biogeochem. Cy.~~,~~, 10~~,~~, 233–245~~, , 1996.~~, doi:10.1029/95GB03871, 1996.

Wania, R., Ross, I., and Prentice, I. C.: Implementation and evaluation of a new methane model within a dynamic global vegetation model: LPJ-WHyMe v1.3.1~~, Geoscientific Model Development~~, Geosci. Model Dev., 3, 565–584, doi:10.5194/gmd-3-565-2010, 2010.

Wania, R., Melton, J. R., Hodson, E. L., Poulter, B., Ringeval, B., Spahni, R., Bohn, T., Avis, C. A., Chen, G., Eliseev, A. V., Hopcroft, P. O., Riley, W. J., Subin, Z. M., Tian, H., van~~Bodegom,~~ Bodegom, P. M., Kleinen, T., Yu, Z. C., Singarayer, J. S., ~~Zürcher,~~Zürcher, S., Lettenmaier, D. P., Beerling, D. J., Denisov, S. N., Prigent, C., Papa, F., and Kaplan, J. O.: Present state of global wetland extent and wetland methane modelling: methodology of a model inter-comparison project (~~WETCHIMP), Geoscientific Model Development~~WETCHIMP), Geosci. Model Dev., 6, 617–641, doi:10.5194/gmd-6-617-2013, 2013.

Wecht, K. J., Jacob, D. J., Frankenberg, C., Jiang, Z., and Blake, D. R.: Mapping of ~~North American~~ North American methane emissions with high spatial resolution by inversion of ~~SCIAMACHY~~ SCIAMACHY satellite data, J. Geophys. Res.-Atmos., 119, 7741–7756, doi:10.1002/2014JD021551, 2014.

Winderlich, J.: Setup of a~~CO_2 and CH_4~~ CO_2 and CH_4 measurement system in Central Siberia and modeling of its results, ~~Ph.D.~~ PhD thesis, University of Hamburg, Hamburg, Germany, available at: http://ediss.sub.uni-hamburg.de/volltexte/2012/5533/pdf/Dissertation.pdf (last access: 14 June 2015), 2012.

Winderlich, J., Chen, H., Gerbig, C., Seifert, T., Kolle, O., Lavrič, J. V., Kaiser, C., Höfer, A., and Heimann, M.: Continuous low-maintenance CO_2/CH_4/H_2O measurements at the Zotino Tall Tower Observatory (ZOTTO) in Central Siberia, Atmospheric Measurement Techniques, 3, 1113–1128, doi:10.5194/amt-3-1113-2010, 2010.

Yadav, V., Mueller, K. L., Dragoni, D., and Michalak, A. M.: A geostatistical synthesis study of factors affecting gross primary productivity in various ecosystems of North America, Biogeosciences, 7, 2655–2671, doi:10.5194/bg-7-2655-2010, 2010.

Table 1. Spatial flux patterns chosen by the model selection framework.

Region	~~season~~ Season	~~Models chosen over a constant model name(s)~~ Chosen models
E. Canada	summer	~~1 of 7~~ LPJ-Bern, SDGVM
E. Canada	fall	~~0 of 7~~ LPJ-Bern
W. Canada	summer	~~2 of 7 LPJ-Bern,~~ SDGVM

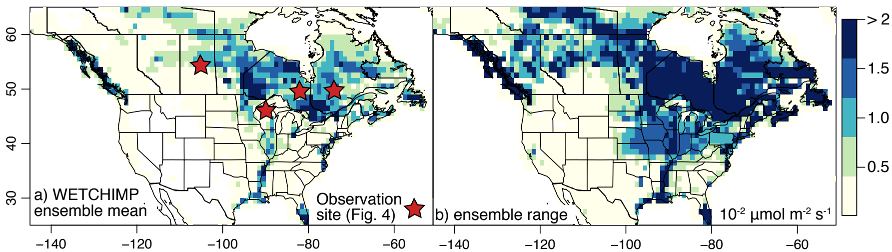

Figure 1. Mean of the annual methane fluxes estimated by the WETCHIMP models ~~(a)~~ **(a)** and the range of fluxes estimated by the ensemble ~~(b)~~**(b)**. Note that the range in estimates is larger than the mean. The fluxes shown above are ~~averaged over an entire grid cell~~the average flux per m^2 of land area, not per m^2 of ~~wetlands~~wetland area.

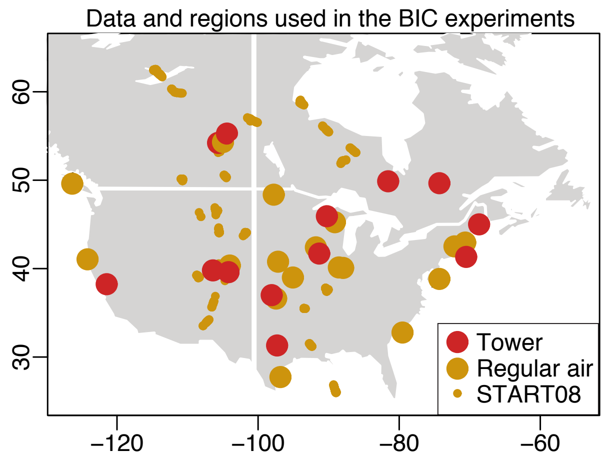

Figure 2. The US and Canadian atmospheric methane observation network for 2007–2008 (14~~,~~703 total observations). Small yellow dots indicate observations from the START08 measurement campaign (Pan et al., 2010). Larger dots indicate tower and aircraft sites with regular observations over the two year period (Andrews et al., 2014). The grey background delineates the four regions used in the synthetic data experiments (~~sect~~Sect. 2.2).

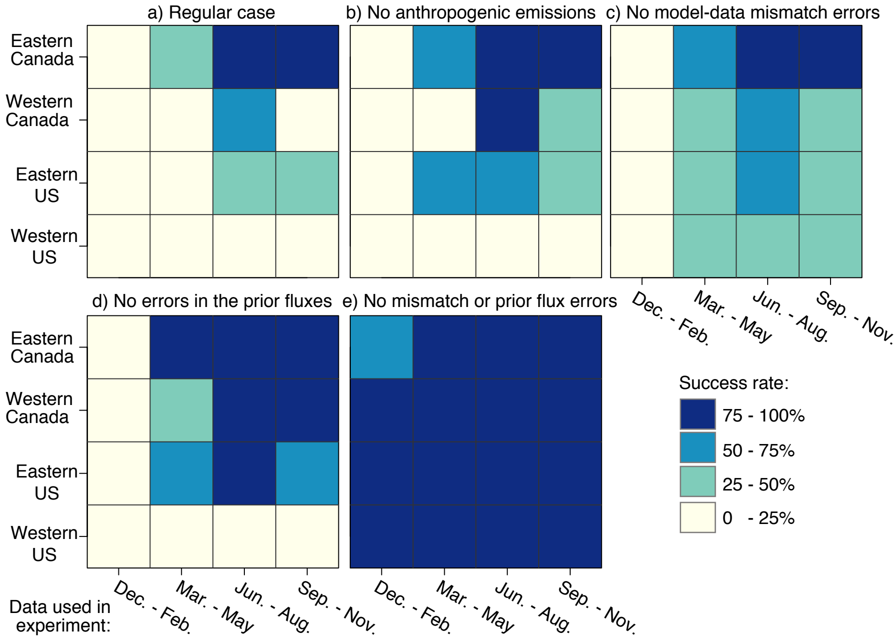

Figure 3. This figure displays the results of the synthetic data experiments. These experiments examine whether the observation network can ~~(a) identify a methane pattern from~~ detect aggregate wetland CH_4 fluxes~~and (b) identify spatial variability in the wetland fluxes~~. The figure shows the percentage of trials that are successful. Darker shades indicate that the network is more sensitive to wetland fluxes in the given region and season. Panel **(a)** shows the results for the standard setup while panels **(b-e)** show the results of several test cases in which anthropogenic emissions or different errors are set to zero.

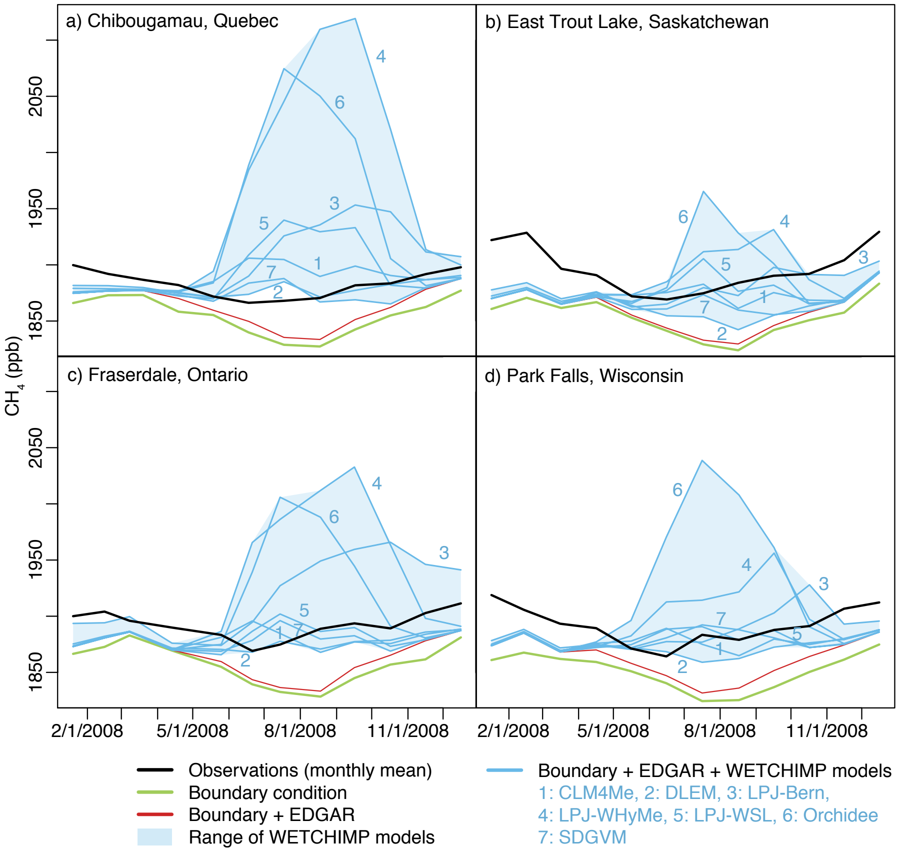

Figure 4. These time series compare atmospheric methane measurements at several observation sites against model estimates using the WETCHIMP ensemble and the EDGAR v4.2FT2010 anthropogenic emissions inventory. ~~The range of estimates~~ Refer to Fig. S4 for model-data time series at additional sites, particularly sites that are distant from ~~the various WETCHIMP models is~~ large wetlands.

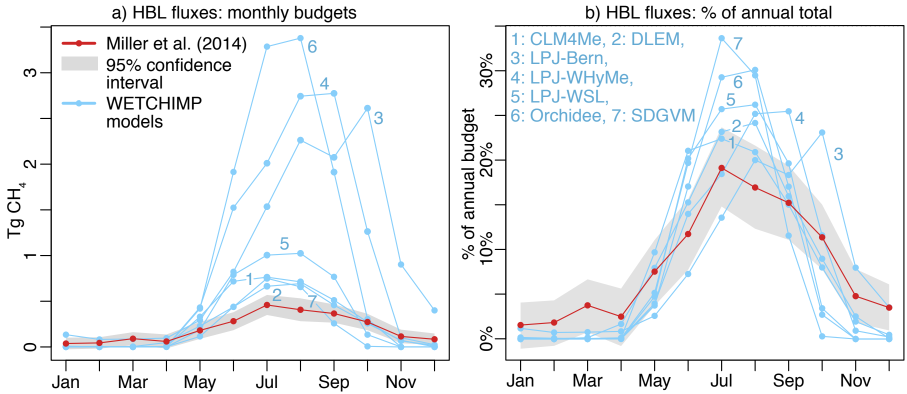

Figure 5. The seasonal cycle in methane fluxes estimated for the ~~Hudson Bay Lowlands (~~HBL ~~;~~ (50–60~~°~~° N, 75–96~~°~~° W). We include both the WETCHIMP estimates and an inverse modeling estimate from Miller et al. (2014). ~~Each~~ Panel **(a)** displays the monthly budget from each estimate while **(b)** displays each month ~~is displayed~~ as a percentage of the annual budget estimated by a given model. ~~This approach highlights differences in the seasonality of the models and controls for differences in magnitude. In general, the WETCHIMP models estimate a narrower seasonal cycle relative to Miller et al. (2014) .~~

This supplement provides more detail on the atmospheric observations, the wetland methane (CH_4) flux estimates, and the statistical methods used throughout the paper.

S1 Atmospheric observation sites

Here we describe, in greater depth, the atmospheric ~~methane~~ CH_4 observations collected across the US and Canada in 2007–2008. The observations used here are ~~identical to~~ the same as those in Miller et al. (2013) and Miller et al. (2014b), and the discussion below summarizes the data descriptions in those papers.

The ~~methane~~ CH_4 analysis in the main article uses either real or synthetic data at US and Canadian observation sites – a total of 14~~,~~703 observations. Of those measurements, 2~~,~~009 are from observation towers in Canada. These towers (from east to west) include Chibougamau, Quebec (CHM, 50°N, 74°W, 30m above ground level); Fraserdale, Ontario (FSD, 50°N, 83°W, 40m agl); East Trout Lake, Saskatchewan (ETL, 54°N, 104°W, 105m agl); and Candle Lake, Saskatchewan (CDL, 54°N, 105°W, 30m agl, 2007 only). These sites, operated by Environment Canada, measure ~~methane~~ CH_4 continuously. In this study, as in Miller et al. (2014b), we use only afternoon averages of the ~~methane~~ CH_4 data and WRF-STILT model output (1pm - 7pm local time); small scale heterogeneities in the continuous data caused by turbulent eddies and incomplete mixing make it difficult to model finer scale temporal patterns in the data. The 2~~,~~009 observations at these Canadian sites represent the total after averaging.

An additional 4~~,~~984 ~~methane~~ CH_4 observations were collected from US towers operated by the National Oceanic and Atmospheric Administration (NOAA) and its partners. These observations include daily flask samples from a number of tower sites (weekly at Argyle and Ponca City): Argyle, Maine (AMT, 45 °N, 69 °W, 107m above ground level (agl)); Erie, Colorado (BAO, 40 °N, 105°W, 300m agl); Park Falls, Wisconsin (LEF, 46°N, 90°W, 244m agl), Martha's Vineyard, Massachusetts (MVY, 41°N, 71°W, 12m agl); Niwot Ridge and Niwot Forest, Colorado (NWF, NWR, 40°N, 105°W, 2,3,23m agl); Ponca City, Oklahoma (SGP, 37°N, 97°W, 60m agl); West Branch, Iowa (WBI, 42°N, 93°W, 379m agl); Walnut Grove, California (WGC, 38°N, 121°W, 91m agl), and Moody, Texas (WKT, 31°N, 97°W, 122, 457m agl).

A further ~~7710 methane~~ 7 710 CH_4 measurements were obtained from flask samples on regular NOAA aircraft flights and from the START08 (Stratosphere-Troposphere Analyses of Regional Transport 2008) measurement campaign (Pan et al., 2010). As in Miller et al. (2013), we only use aircraft observations up to ~~2500m~~ 2 500m above ground level. Observations at higher altitudes are less sensitive to surface emissions and were instead used by Miller et al. (2013) to optimize the ~~empirical methane boundary condition~~ estimated CH_4 boundary condition or background concentrations. In this study, we only use aircraft and tower-based observations collected during daytime hours.

We further screen the data for biomass burning influence at the Canadian sites and at Park Falls, Wisconsin. At the these sites, we remove all days with CO that peaks above 200 ppb, as was done in Miller et al. (2014b). When these sites see influence from distant anthropogenic emissions, CO is often elevated, but it rarely exceeds 200 ppb except during time periods with known fires (Miller et al., 2008).

S2 WETCHIMP ~~methane~~ CH_4 flux models

This section of the supplement details the WETCHIMP ~~methane~~ CH_4 estimates from Melton et al. (2013) and Wania et al. (2013). The seven ~~methane~~ CH_4 estimates used in this study are shown in Fig. S1. The wetland ~~methane~~ CH_4 fluxes estimated by these models varies widely – both in magnitude and in spatial distribution. For example, the SDGVM model places large fluxes over the US Corn Belt relative to other regions while other models like Orchidee place large fluxes in Northern Canada that extent far into the Northwest Territories. For a more in-depth inter-comparison of these flux estimates, refer to Melton et al. (2013) and Wania et al. (2013).

S3 ~~The synthetic data~~Additional information on the model selection setup

In the main article, we use synthetic ~~methane data to explore the sensitivity of atmospheric observations to wetland fluxes (sections~~ CH_4 data and a model selection framework to examine whether atmospheric observations can detect aggregate wetland CH_4 fluxes (Sect. 2.2 and 3). This section ~~describes in greater detail how we construct this synthetic data .~~ first describes the synthetic data experiments (Sect. 2.2) followed by additional detail on the model selection runs that use real data (Sect. 2.3). The methods described here are adapted from Fang et al. (2014)~~and~~, Shiga et al. (2014), and ? , and the discussion below parallels the descriptions in those studies.

The synthetic observations include contributions from anthropogenic sources, from wetlands, and from simulated model and measurement errors:

$$\boldsymbol{z}_{\text{synthetic}} = \mathbf{H}(\boldsymbol{s}_{\text{anthro}} + \boldsymbol{s}_{\text{wetland}}) + \boldsymbol{\epsilon} \tag{S1}$$

In this equation, $\boldsymbol{z}_{\text{synthetic}}$ $(n \times 1)$ represents the synthetic observations generated for an observation site. The vector $\boldsymbol{s}_{\text{anthro}}$ $(m \times 1)$ represents emissions from anthropogenic sources, and $\boldsymbol{s}_{\text{wetland}}$ $(m \times 1)$ represents wetland fluxes. The footprint or sensitivity matrix $\mathbf{H}$ $(n \times m)$, generated from WRF-STILT, models the impact of these emissions at the observation sites.

In this study, we use the a priori anthropogenic emissions estimates from Miller et al. (2013) and Miller et al. (2014b) for $\boldsymbol{s}_{\text{anthro}}$. Those studies used activity data from the EDGAR inventory and a model selection framework to construct a prior anthropogenic emissions estimate. These EDGAR activity datasets include economic or demographic data that may predict the spatial distribution of ~~methane~~ CH_4 emissions (e.g., human or ruminant population maps).

The wetland fluxes ($\boldsymbol{s}_{\text{wetland}}$) in Eq. S1 are taken from the WETCHIMP ~~methane~~ CH_4 flux models (experiment two in Melton et al. (2013)). ~~We use only four of the seven WETCHIMP models to generate synthetic data : CLM4Me, DLEM, LPJ-WSL, and SDGVM. These models have an overall magnitude that most closely matches the methane budgets estimated by three recent top-down studies over Canada's~~ For the synthetic data experiments, we scale these models to match the Hudson Bay Lowlands (HBL) ~~(Pickett-Heaps et al., 2011; Miller et al., 2014b; Wecht et al., 2014) . The magnitude of these four models is likely the most realistic among the WETCHIMP flux estimates. The other WETCHIMP models, in contrast, predict much higher fluxes (Fig. 4).~~ budget estimated by Pickett-Heaps et al. (2011) , Miller et al. (2014b) , and Wecht et al. (2014) . This scaling ensures more consistent or representative results. The larger the wetland flux, the more likely that the observation network can detect a CH_4 fluxes from wetlands. Therefore, if we conduct the synthetic data experiment using a flux model that has an anomalously large magnitude, we would concomitantly obtain anomalously optimistic results.

As in Miller et al. (2013) and Miller et al. (2014b), the emissions ($\boldsymbol{s}_{\text{anthro}}$ and $\boldsymbol{s}_{\text{wetland}}$) are regridded to a spatial resolution of 1° latitude by 1° longitude. The EDGAR activity data do not have any seasonality, so the anthropogenic emissions ($\boldsymbol{s}_{\text{anthro}}$) are seasonally invariant. The WETCHIMP models have a monthly temporal resolution, as in Melton et al. (2013). That study provides flux estimates for the years 1993-2004; we use the mean of these ten years for all analysis in this study.

The final term in equation S1, $\boldsymbol{\epsilon}$ $(n \times 1)$, represents simulated errors in the measurements, in WRF-STILT, and in the fluxes ($\boldsymbol{s}_{\text{anthro}}$ and $\boldsymbol{s}_{\text{wetland}}$). The ~~magnitude and spatial/temporal~~

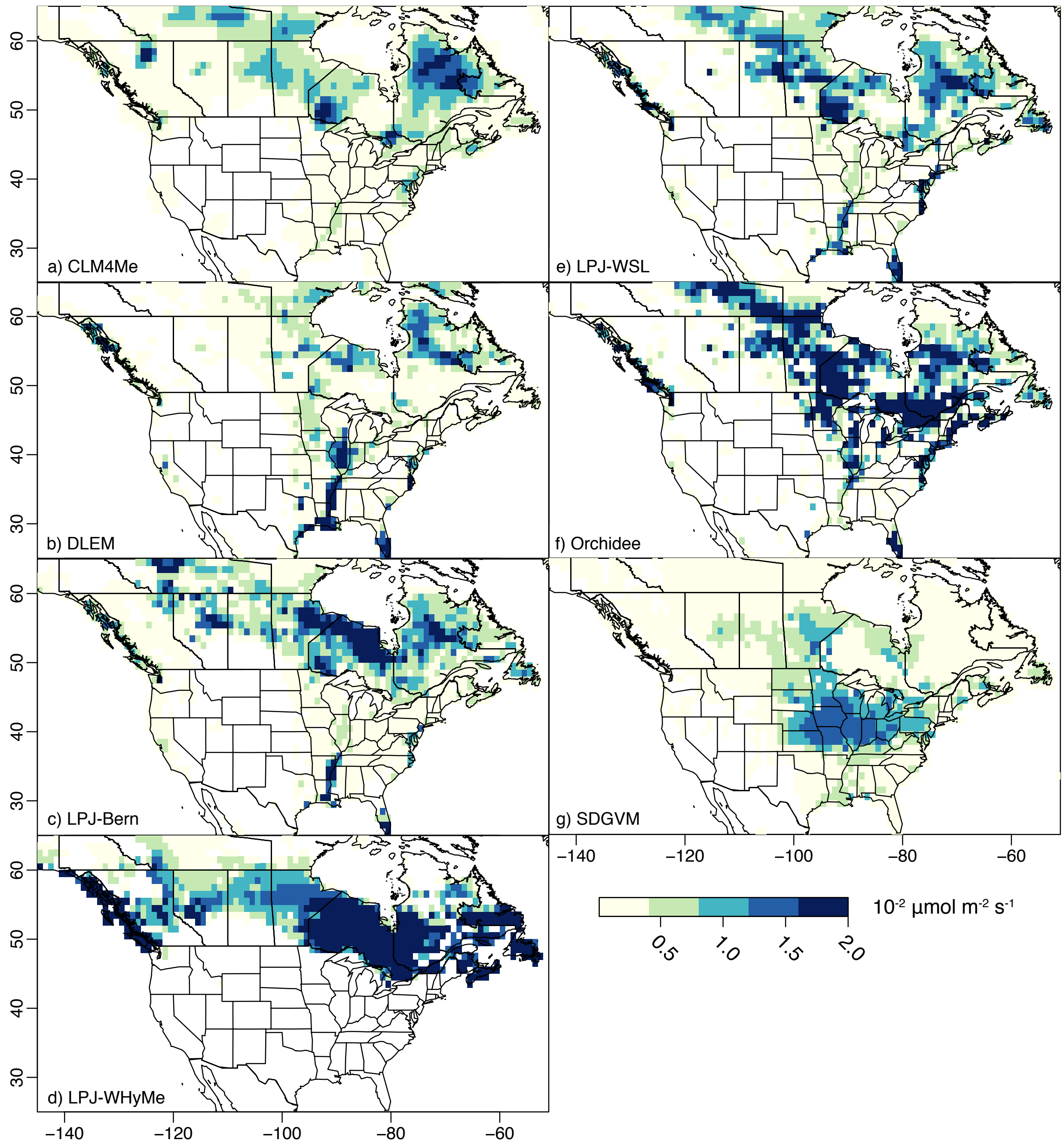

Figure S1: Annual mean wetland ~~methane~~ CH_4 fluxes from seven different WETCHIMP estimates (Melton et al., 2013; Wania et al., 2013). The fluxes shown here are averaged over the 1993-2004 study period. Note that the fluxes shown above are averaged over the entire grid cell, not per m^2 of wetlands.

~~structure of these errors were estimated in Miller et al. (2013) for the US and Miller et al. (2014b) for Canada. The remainder of this section details the specific calculations for simulating~~ errors in $\boldsymbol{\epsilon}$ ~~.~~

~~The errors in $\boldsymbol{\epsilon}$~~ are distributed according to the covariance matrix $\boldsymbol{\Psi}$ $(n \times n)$ (Eq. 1):

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}) \tag{S2}$$

$$\boldsymbol{\Psi} = \mathbf{H}\mathbf{Q}\mathbf{H}^T + \mathbf{R} \tag{S3}$$

The variances and covariances within $\boldsymbol{\Psi}$ fall into two different categories. The first category are errors due to imperfect emissions, described by covariance matrix $\mathbf{Q}$ $(m \times m)$. In atmospheric inversion studies, this matrix is typically termed the a priori covariance matrix. The diagonal elements of $\mathbf{Q}$ describe a set of variances – differences between the prior fluxes and the unknown true emissions over long spatial or temporal scales. The off-diagonal elements of $\mathbf{Q}$ describe any spatial and/or temporal covariances in these differences. In Eq. S3, the footprint or sensitivity matrix ($\mathbf{H}$) projects $\mathbf{Q}$ from units of $(\text{flux})^2$ into units of parts per billion squared, $(\text{ppb})^2$.

We refer to the second type of errors as model-data mismatch errors, denoted by covariance matrix $\mathbf{R}$ $(n \times n)$. These include all errors in the WRF-STILT model or the measurements that are unrelated to an imperfect flux estimate. Examples of model-data mismatch errors include measurement error, atmospheric transport error, or errors due to the spatial or temporal resolution of WRF-STILT. ~~Over the United States, we~~

The synthetic data simulations in this study use values of $\mathbf{Q}$ and $\mathbf{R}$ estimated in Miller et al. (2013) and Miller et al. (2014b) . In the synthetic data studies, we construct a statistical model that is representative of a prototypical real data inverse model. Similarly, we want to use values for ~~$\mathbf{R}$ and~~ $\mathbf{Q}$ ~~that were estimated by Miller et al. (2013) using WRF-STILT and the same atmospheric methane observations~~ and $\mathbf{R}$ that are representative of what one would likely encounter in a real-data setup. Miller et al. (2013) and Miller et al. (2014b) constructed real data inverse models over the US and Canada, respectively, using the same atmospheric observations and WRF-STILT simulations used in this study. ~~Similarly, we use values for $\mathbf{R}$ and~~ Those studies used a model selection framework to find prior models that show optimal fit against available observations. In each study, the authors then estimated the elements of $\mathbf{Q}$ and $\mathbf{R}$ using that prior model. The resulting estimates of $\mathbf{Q}$ ~~over Canada that were estimated in Miller et al. (2014b) , a parallel inverse modeling study over that country~~are representative of prior models that shows optimal agreement with atmospheric observations. For case study (b) (no anthropogenic emissions), we estimate $\mathbf{Q}$ using the same approach as in Shiga et al. (2014) . In that study, the authors used the estimated variances and covariances of the remaining fluxes (in this case wetland fluxes) to populate $\mathbf{Q}$.

In ~~order to simulate~~ the real data experiments (Sect. 2.3), we estimate unique values of $\mathbf{Q}$ and $\mathbf{R}$ each time we run the model selection framework. We estimate these parameters using Restricted Maximum Likelihood (RML) (Corbeil and Searle, 1976; Kitanidis, 1995; Michalak et al., 2004; Gourd the same procedure used in Miller et al. (2013) and Miller et al. (2014b) .

We use these covariance matrices to compute $\boldsymbol{\epsilon}$ ~~, we next~~ through several steps. First, we compute the Cholesky decomposition of the combined covariance matrix $\boldsymbol{\Psi}$:

$$\boldsymbol{\Psi} = \mathbf{C}\mathbf{C}^T \tag{S4}$$

The covariance matrix $\boldsymbol{\Psi}$ has units of $(\text{ppb})^2$, but its Cholesky decomposition ($\mathbf{C}$) has units of ppb~~, a fact that will become useful in the next step~~. With this decomposition in hand, we next simulate a set of errors, $\boldsymbol{\epsilon}$ (e.g., Fang et al., 2014; Shiga et al., 2014):

$$\boldsymbol{\epsilon} = \mathbf{C}\boldsymbol{u} \tag{S5}$$

$$\boldsymbol{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{1}) \tag{S6}$$

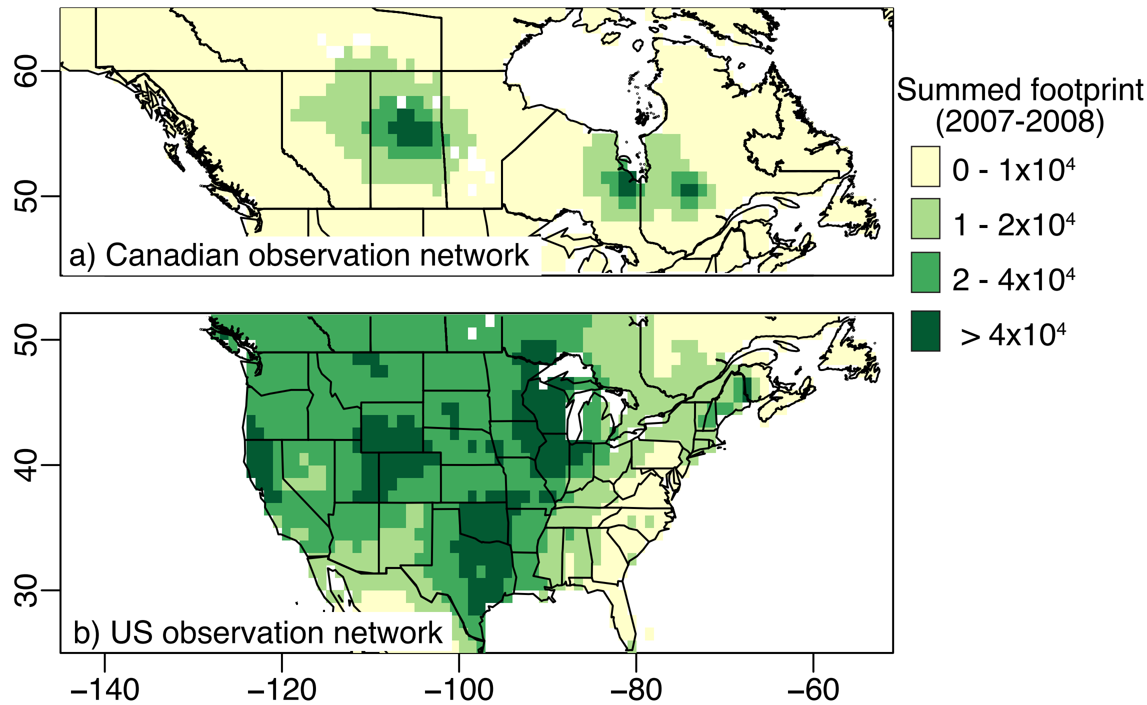

Figure S2: Total, summed footprint from the ~~(a)~~ **(a)** Canadian and ~~(b)~~ **(b)** US observation networks. The observation sites incorporated into this figure are shown in Fig. 2. Each individual footprint (associated with an individual atmospheric observation) has units of concentration per unit flux (ppb per μmol m^{-2} s^{-1}). In this figure, we sum all footprints for 2007–2008.

where $\boldsymbol{u}$ represents a set of randomly-generated numbers with a mean of zero and variance of one.

We simulate 1000 synthetic datasets for each experiment to adequately sample the random errors in $\boldsymbol{\epsilon}$. We then use the model selection framework to find the optimal candidate model for each of these datasets. The results presented in Fig. 3 are therefore the composite of thousands of model selection runs: one model selection run for each synthetic dataset. We use a branch and bound algorithm from Yadav et al. (2013) to improve the computational efficiency of these model selection runs. Furthermore, we estimate the coefficients ($\boldsymbol{\beta}$) in Eq. 1 using Lagrange multipliers to ensure that none of the estimated coefficients have unrealistic negative values (e.g., Miller et al., 2014a).

In the real data setup (Sect. 2.3), we run the model selection procedure once for each of the seven WETCHIMP flux estimates. We only include one of the seven WETCHIMP flux models in each model selection run. As a result, the WETCHIMP models do not compete against one another for selection. In each run, the model selection framework can select the given WETCHIMP model in any of the four geographic regions and any of the four seasons.

S4 ~~Sensitivity~~ Overall sensitivity of the observation network to ~~surface~~ CH_4 fluxes

In this section, we describe the overall footprint or sensitivity of the observation network to ~~methane~~ CH_4 fluxes. This sensitivity will play at least some role in network's ability to ~~identify a signal from wetlands~~detect wetland CH_4 fluxes. The WRF-STILT model quantifies this sensitivity in terms of a footprint. Each row the matrix $\mathbf{H}$ is the footprint associated with a

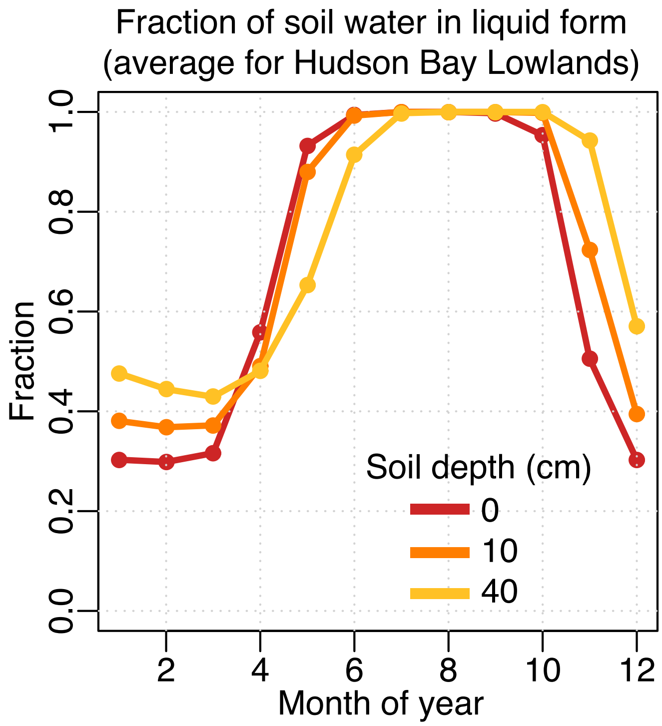

Figure S3: This figure displays the fraction of soil water that is unfrozen for the HBL in different seasons and at different soil depths. Estimates are taken from NARR (Mesinger et al., 2006).

different atmospheric ~~methane~~ CH_4 observation. In Fig. S2, we plot these footprints, summed over all of 2007–2008.

This figure show several distinctive patterns. First, the US network has a higher sensitivity than the Canadian network. This pattern is due to the larger number of observation sites over the US. Second, the highest sensitivities are clustered in distinctive regions with multiple observation sites – Wisconsin, Texas/Oklahoma, and California, among other regions.

S5 Soil freeze/thaw estimates from NARR

Figure S3 shows the soil freeze/thaw cycle at different depths averaged across the HBL. These estimates are taken from North American Regional Reanalysis (NARR) (Mesinger et al., 2006), and the values shown in Fig. S3 are average values for each month. The main article references this figure in a discussion of the ~~methane~~ CH_4 flux seasonal cycle (Sect. 4.3).

S6 Additional model-data time series

This section includes additional model-data time series analogous to those in Fig. 4. That figure compares averaged concentrations modeled by WRF-STILT against monthly-averaged observations at four different observation sites. The sites displayed in that figure are located near large wetlands and in regions where the synthetic data experiments had a high success rate (Fig. 3). The sites displayed in Fig. S4 in this section are located further from wetlands and in regions that had a low success rate in the BIC experiments. At many of the sites in Fig. S4, the modeled wetland signal is difficult to distinguish. These sites contrast with those in Fig. 4 which "see" a relatively large CH_4 increment from wetlands.

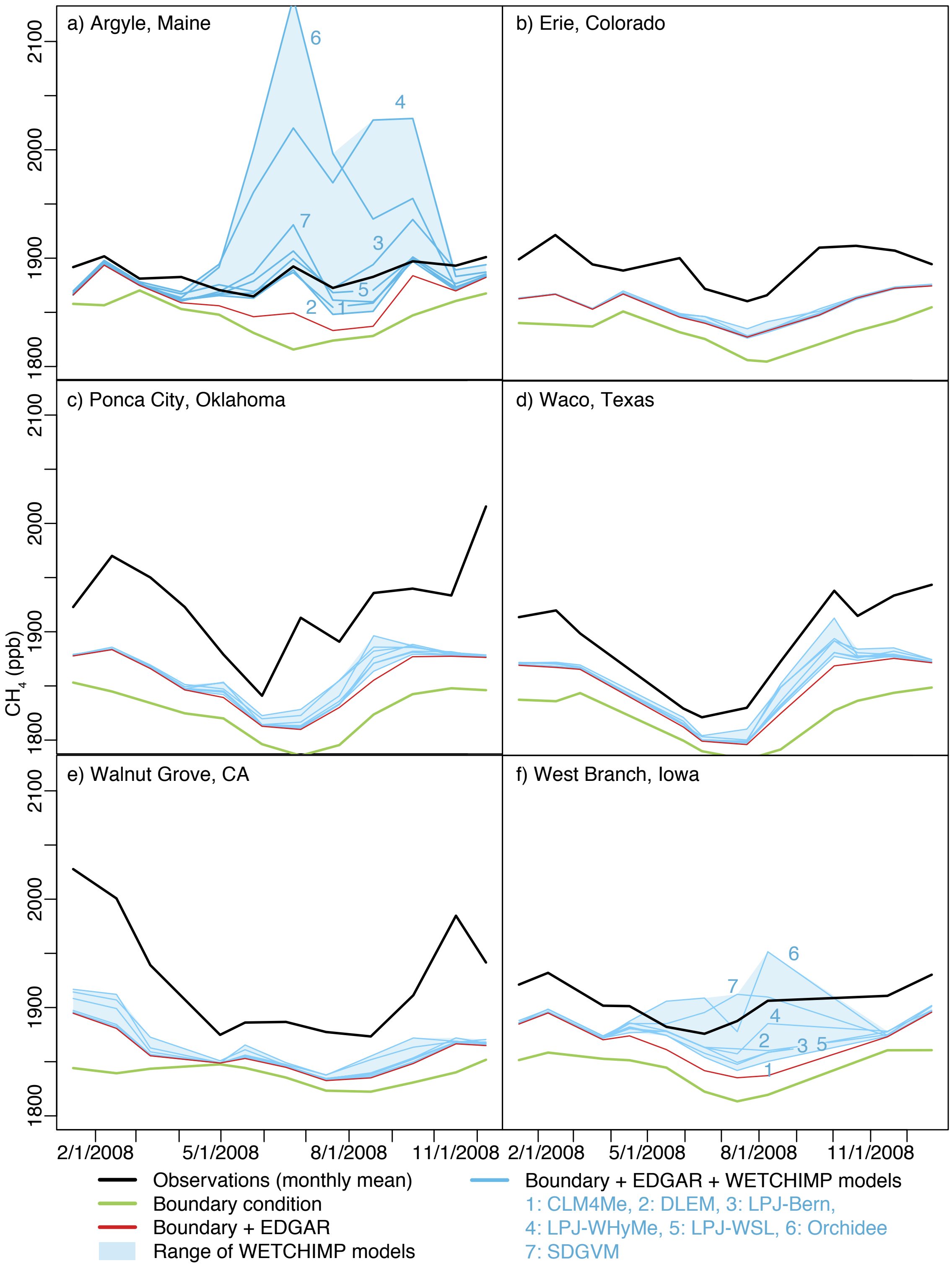

Figure S4: This figure is analogous to Fig. 4 and displays monthly-averaged model-data time series for additional atmospheric observation sites.

S7 Uncertainty in the model-data framework

A number of modeling and measurement uncertainties influence the results presented in this paper. These uncertainties are discussed in detail by Miller et al. (2013) , Miller et al. (2014b) , and Miller et al. (2014a) . This section provides a summary of those discussions.

The model selection framework in this study accounts for modeling or measurement errors in the **R** covariance matrix (Eq. S3). This covariance matrix is typically included in a Bayesian synthesis or geostatistical inverse model (e.g., Michalak et al., 2004) . The errors described by **R** are collectively referred to as model-data mismatch – any errors in the model-data framework that are unrelated to an imperfect flux estimate. This mismatch includes errors in the modeled winds, errors in the CH_4 boundary condition, and any errors due to the finite spatial or temporal resolution of the model, among other possible error sources. This section of the supplement first discusses the overall magnitude of these model-data mismatch errors and then discusses individual components of the model-data mismatch, including potential errors in the estimated winds and in the boundary condition.

Both Miller et al. (2013) and Miller et al. (2014b) estimate the magnitude of model-data mismatch errors for observation sites in the US and Canada, respectively. These studies used a procedure known as Restricted Maximum Likelihood (RML) to estimate the parameters that define both the **R** and **Q** covariance matrices (e.g., Corbeil and Searle, 1976; Kitanidis, 1995; Michalak et al., 200 The estimated mismatch errors range in magnitude from 12-13 ppb (standard deviations) at Canadian tower sites to 20-30 ppb at tower sites near oil and gas operations in the southern US (refer to Fig. S2 in Miller et al. (2013) and Fig. S6 in Miller et al. (2014b)). This magnitude (12-30 ppb) is equivalent to 25–70% of the average CH_4 signal from North American emissions as seen at the various observation sites.

These model-data mismatch errors encompass numerous sources of error, but these errors are likely dominated by uncertainties in atmospheric transport. Nehrkorn et al. (2010) generated WRF meteorology for use in STILT and compared the estimated winds against US and Canadian radiosondes. They computed a root mean squared error (RMSE) of 2.5–4 m s^{-1} in the horizontal winds and found no change in error statistics at the top of the boundary layer. Hegarty et al. (2013) further coupled the STILT model with several weather models and found that simulations with WRF produced lower error statistics relative to other weather models.

Several existing studies have shown consistent results between WRF-STILT and other atmospheric models; this consistency may indicate a lack of large-scale bias in atmospheric transport estimated by WRF-STILT. For example, constraints on summertime US carbon monoxide emissions estimated with STILT and the GEOS-Chem model match to within 10% (Miller et al., 2008; Hudman et al., 2008) . CH_4 budgets estimated for the HBL in Canada using WRF-STILT and GEOS-Chem are similar to within 10% (Pickett-Heaps et al., 2011; Miller et al., 2014b; W Furthermore, CH_4 budgets estimated for the US with WRF-STILT and GEOS-Chem match to within ∼10% (Miller et al., 2013; Turner et al., 2015) .

The CH_4 boundary condition is another, potentially large source of uncertainty in the CH_4 modeling framework. To create this boundary condition, we interpolate atmospheric CH_4 observations near or over the Pacific and Arctic Oceans to create a boundary "curtain." This curtain estimates CH_4 concentrations over the Pacific and Arctic; it varies by latitude, altitude, and time (see Fig. S4 in Miller et al. (2014b)). We then sample concentrations along this curtain depending upon the ending latitude, altitude, and time of each WRF-STILT trajectory. These sampled concentrations become the boundary condition – an estimate of the CH_4 concentration in air before that air reaches North America. Miller et al. (2013) and Miller et al. (2014b) describe this setup in greater detail along with the associated uncertainties. For example, Miller et al. (2013) compared the boundary condition estimate against aircraft data collected above 3000m over the United States. They found an average difference of 2.7 ppb between the aircraft observations and boundary condition estimate. Miller et al. (2013) then adjusted the boundary condition based upon this aircraft data. They subsequently estimated a total US CH_4 budget using boundary conditions with and without the aircraft adjustment. The total CH_4 budget using the aircraft-corrected boundary condition was approximately 5% higher than the unadjusted boundary condition estimate. This result indicates the possible effects of boundary condition uncertainties on a national-scale CH_4 budget estimate.

References

Corbeil, R. R. and Searle, S. R.: Restricted maximum likelihood (REML) estimation of variance components in the mixed model, Technometrics, 18, pp. 31–38, 1976.

Fang, Y., Michalak, A. M., Shiga, Y. P., and Yadav, V.: Using atmospheric observations to evaluate the spatiotemporal variability of CO_2 fluxes simulated by terrestrial biospheric models, Biogeosciences, 11, 6985–6997, doi:10.5194/bg-11-6985-2014, 2014.

Gourdji, S. M., Mueller, K. L., Yadav, V., Huntzinger, D. N., Andrews, A. E., Trudeau, M., Petron, G., Nehrkorn, T., Eluszkiewicz, J., Henderson, J., Wen, D., Lin, J., Fischer, M., Sweeney, C., and Michalak, A. M.: North American CO_2 exchange: inter-comparison of modeled estimates with results from a fine-scale atmospheric inversion, Biogeosciences, 9, 457–475, doi:10.5194/bg-9-457-2012, 2012.

Hegarty, J., Draxler, R. R., Stein, A. F., Brioude, J., Mountain, M., Eluszkiewicz, J., Nehrkorn, T., Ngan, F., and Andrews, A.: Evaluation of Lagrangian particle dispersion models with measurements from controlled tracer releases, J. Appl. Meteorol. Clim., 52, 2623–2637, 2013.

Hudman, R. C., Murray, L. T., Jacob, D. J., Millet, D. B., Turquety, S., Wu, S., Blake, D. R., Goldstein, A. H., Holloway, J., and Sachse, G. W.: Biogenic versus anthropogenic sources of CO in the United States, Geophys. Res. Lett., 35, doi:10.1029/2007GL032393, l04801, 2008.

Kitanidis, P.: Quasi-linear geostatistical theory for inversing, Water Resour. Res., 31, 2411–2419, doi:10.1029/95WR01945, 1995.

Melton, J. R., Wania, R., Hodson, E. L., Poulter, B., Ringeval, B., Spahni, R., Bohn, T., Avis, C. A., Beerling, D. J., Chen, G., Eliseev, A. V., Denisov, S. N., Hopcroft, P. O., Lettenmaier, D. P., Riley, W. J., Singarayer, J. S., Subin, Z. M., Tian, H., Zürcher, S., Brovkin, V., van Bodegom, P. M., Kleinen, T., Yu, Z. C., and Kaplan, J. O.: Present state of global wetland extent and wetland methane modelling: conclusions from a model inter-comparison project (WETCHIMP), Biogeosciences, 10, 753–788, doi:10.5194/bg-10-753-2013, 2013.

Mesinger, F., Dimego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., Jovi, D., Woollen, J., Rogers, E., Berbery, E. H., Ek, M. B., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D., and Shi, W.: North American Regional Reanalysis, B. Am. Meteorol. Soc., 87, 343–360, doi:10.1175/BAMS-87-3-343, 2006.

Michalak, A., Bruhwiler, L., and Tans, P.: A geostatistical approach to surface flux estimation of atmospheric trace gases, J. Geophys. Res.-Atmos., 109, doi:10.1029/2003JD004422, 2004.

Miller, S. M., Matross, D. M., Andrews, A. E., Millet, D. B., Longo, M., Gottlieb, E. W., Hirsch, A. I., Gerbig, C., Lin, J. C., Daube, B. C., Hudman, R. C., Dias, P. L. S., Chow, V. Y., and Wofsy, S. C.: Sources of carbon monoxide and formaldehyde in North America determined from high-resolution atmospheric data, Atmos. Chem. Phys., 8, 7673–7696, doi: 10.5194/acp-8-7673-2008, 2008.

Miller, S. M., Wofsy, S. C., Michalak, A. M., Kort, E. A., Andrews, A. E., Biraud, S. C., Dlugokencky, E. J., Eluszkiewicz, J., Fischer, M. L., Janssens-Maenhout, G., Miller, B. R., Miller, J. B., Montzka, S. A., Nehrkorn, T., and Sweeney, C.: Anthropogenic emissions of methane in the United States, P. Natl. Acad. Sci. USA, 110, 20 018–20 022, doi:10.1073/pnas.1314392110, 2013.

Miller, S. M., Michalak, A. M., and Levi, P. J.: Atmospheric inverse modeling with known physical bounds: an example from trace gas emissions, Geoscientific Model Development, 7, 303–315, doi:10.5194/gmd-7-303-2014, 2014a.

Miller, S. M., Worthy, D. E. J., Michalak, A. M., Wofsy, S. C., Kort, E. A., Havice, T. C., Andrews, A. E., Dlugokencky, E. J., Kaplan, J. O., Levi, P. J., Tian, H., and Zhang, B.: Observational constraints on the distribution, seasonality, and environmental predictors of North American boreal methane emissions, Global Biogeochem. Cy., 28, 146–160, doi:10.1002/2013GB004580, 2014b.

Nehrkorn, T., Eluszkiewicz, J., Wofsy, S. C., Lin, J. C., Gerbig, C., Longo, M., and Freitas, S.: Coupled Weather Research and Forecasting-Stochastic Time-Inverted Lagrangian Transport (WRF-STILT) model, Meteorol. Atmos. Phys., 107, 51–64, doi:10.1007/s00703-010-0068-x, 2010.

Pan, L. L., Bowman, K. P., Atlas, E. L., Wofsy, S. C., Zhang, F., Bresch, J. F., Ridley, B. A., Pittman, J. V., Homeyer, C. R., Romashkin, P., and Cooper, W. A.: The Stratosphere-Troposphere Analyses of Regional Transport 2008 Experiment, B. Am. Meteorol. Soc., 91, 327–342, doi:10.1175/2009BAMS2865.1, 2010.

Pickett-Heaps, C. A., Jacob, D. J., Wecht, K. J., Kort, E. A., Wofsy, S. C., Diskin, G. S., Worthy, D. E. J., Kaplan, J. O., Bey, I., and Drevet, J.: Magnitude and seasonality of wetland methane emissions from the Hudson Bay Lowlands (Canada), Atmos. Chem. Phys., 11, 3773–3779, doi:10.5194/acp-11-3773-2011, 2011.

Shiga, Y. P., Michalak, A. M., Gourdji, S. M., Mueller, K. L., and Yadav, V.: Detecting fossil fuel emissions patterns from subcontinental regions using North American in situ CO_2 measurements, Geophys. Res. Lett., 41, 4381–4388, doi:10.1002/2014GL059684, 2014.

Turner, A. J., Jacob, D. J., Wecht, K. J., Maasakkers, J. D., Lundgren, E., Andrews, A. E., Biraud, S. C., Boesch, H., Bowman, K. W., Deutscher, N. M., Dubey, M. K., Griffith, D. W. T., Hase, F., Kuze, A., Notholt, J., Ohyama, H., Parker, R., Payne, V. H., Sussmann, R., Sweeney, C., Velazco, V. A., Warneke, T., Wennberg, P. O., and Wunch, D.: Estimating global and North American methane emissions with high spatial resolution using GOSAT satellite data, Atmos. Chem. Phys., 15, 7049–7069, doi:10.5194/acp-15-7049-2015, 2015.

Wania, R., Melton, J. R., Hodson, E. L., Poulter, B., Ringeval, B., Spahni, R., Bohn, T., Avis, C. A., Chen, G., Eliseev, A. V., Hopcroft, P. O., Riley, W. J., Subin, Z. M., Tian, H., van Bodegom, P. M., Kleinen, T., Yu, Z. C., Singarayer, J. S., Zürcher, S., Lettenmaier, D. P., Beerling, D. J., Denisov, S. N., Prigent, C., Papa, F., and Kaplan, J. O.: Present state of global wetland extent and wetland methane modelling: methodology of a model inter-comparison project (WETCHIMP), Geoscientific Model Development, 6, 617–641, doi:10.5194/gmd-6-617-2013, 2013.

Wecht, K. J., Jacob, D. J., Frankenberg, C., Jiang, Z., and Blake, D. R.: Mapping of North American methane emissions with high spatial resolution by inversion of SCIAMACHY satellite data, J. Geophys. Res.-Atmos., 119, 7741–7756, doi:10.1002/2014JD021551, 2014.

Yadav, V., Mueller, K. L., and Michalak, A. M.: A backward elimination discrete optimization algorithm for model selection in spatio-temporal regression models, Environ. Modell. Softw., 42, 88 – 98, doi:10.1016/j.envsoft.2012.12.009, 2013.

Reply to reviewer #1

S. M. Miller, et. al.

We would like to thank the reviewer for suggestions and comments on the manuscript. The reviewer's detailed suggestions have been very helpful in improving the manuscript. Below, we have included the reviewers comments (in bold) along with our reply and the associated changes/updates to the manuscript.

1 Overall comments:

- **Concerning the language, the authors frequently make use of question sentences, often of rhetorical nature, which I personally find quite annoying within the context of a journal article. So I highly recommend rewriting these passages. Particularly these question sentences are often unnecessarily repeated throughout the paper.**

  We have revised this language accordingly.

- **The structure is clear and concise, though at some sections slightly unbalanced – in some cases (e.g. Section 2.1) you skip over many essential details and refer to existing manuscripts, while in others (e.g. 2.2) you provide many details where you could also have used citations instead.**

  We have added more detail to Section 2.1 on WRF-STILT and on the WETCHIMP model comparison project. We have edited Section 2.2 to rely more heavily on the existing literature.

- **The authors claim in the conclusions section (p.9357, ll.3ff) that bottom-up and top-down modelers should do a better job in joining forces to arrive at more solid estimates of methane emission budgets, a statement that I fully support (even though I don't think it belongs into a conclusion section in the way it is presented here ..). At the same time, they work out various systematic differences between bottom-up and top-down products within the context of this study, and attribute all the 'blame' to the bottom-up models, without even starting to discuss shortcomings in the inverse modeling approach which also lead to (well known) large uncertainties. Stated a bit provocative, it sounds like the authors' intention is to tell bottom-up modelers that they need to do a much better job, and better ask the top-down crowd how to do things right**

  The reviewer makes a reasonable point. Both bottom-up models and top-down, inverse models have respective strengths and weaknesses/uncertainties; neither provides the final word on greenhouse gas fluxes. We certainly do not want to 'blame' anyone or any previous research effort, and we have revised the manuscript wording where possible to make this point clearer. Both top-down and bottom-up modelers often express a desire to meld efforts in a way that would leverage the respective strengths of each approach. This goal is easier said than done. As the reviewer points out, this statement in the conclusions section is likely too cursory or too provocative, and we have removed it. In the revised manuscript, we not only suggest future improvements to bottom-up models but also to top-down estimates.

- **Moreover, I was disappointed to find that the authors don't really make an effort to explain where such differences might stem from.**

  In the revised manuscript, we have augmented this discussion of 'why' or 'how' in sections 4.1 – 4.3. Atmospheric data can often provide useful information on the magnitude, location, or timing of fluxes, but it is usually much more difficult to infer how or why these fluxes occurred. In several instances, we can hypothesize why model-data differences occur. For example, existing bottom-up flux estimates exhibit different spatial distributions over North America, and many of those differences appear to stem from the underlying wetland distribution. In the paper, we discuss this difference in context of atmospheric data; bottom-up models that are most consistent with the atmospheric data use wetland distributions that are based, at least in part, on land cover mapping. We also discuss discrepancies in seasonality and why these discrepancies may occur (Section 4.3).

- **However, a comprehensive interpretation of the observed differences as presented herein needs to include an extensive section that discusses the uncertainties and potential biases that stem from the atmospheric inversion part of the comparison.**

  We have added a section to the supplement that highlights the largest sources of uncertainties in the top-down analysis conducted in the paper.

  The analysis in this paper is based upon inversion frameworks developed in Miller et al. (2013), Miller et al. (2014b), and Miller et al. (2014a). Those papers discuss, in detail, the uncertainties and potential biases that stem from inverse modeling. For example, Miller et al. (2013) explore uncertainties in the estimated methane boundary condition, uncertainties in the estimated covariance matrix parameters (the parameters that define $\mathbf{\Psi}$), uncertainties due to atmospheric transport in WRF-STILT, uncertainties due to geological CH_4 sources, and uncertainties in the attribution of CH_4 to individual sources. Miller et al. (2014b) discuss uncertainties due to the nested meteorology domains in WRF-STILT, uncertainties in the methane boundary condition, uncertainties due to the sparsity of the CH_4 observation network, uncertainties in atmospheric transport estimated by WRF-STILT, and uncertainties in the covariance matrix parameters. Finally, Miller et al. (2014a) discuss uncertainties in the flux estimate due to the assumptions made by the statistical modeling framework. In addition to these inverse modeling papers, Nehrkorn et al. (2010) and Hegarty et al. (2013) also discuss atmospheric transport uncertainties in the WRF-STILT modeling framework.

  The uncertainties that were estimated/developed in those existing studies are used throughout the current manuscript. For example, these uncertainties form the basis of the $\mathbf{R}$ and $\mathbf{Q}$ covariance matrices which are an integral part of the model selection analysis (see Eq. 1 and Fig. 3). In addition, the inversion estimates in Fig. 5 of the revised manuscript are shown in context of the confidence interval estimated from the inversion. This confidence interval accounts for limited data coverage, transport model errors, the finite resolution of the inverse model, and other error sources (as discussed in Miller et al. (2014b)).

2 Detailed comments:

- **P.9345f, Section 2.1: As mentioned above, I think this is extremely short. Even though the details might be given somewhere else, the reader needs more information to understand what approaches you used in the context of this study.**

  We have expanded this section to describe the data, atmospheric model, and WETCHIMP methane models in greater detail.

- **p.9346, l.16f: You never explain and/or discuss how the low temporal resolution (monthly) of emission fluxes is actually coupled to your mixing ratio observations, which probably have a temporal resolution of 1-3hrs (details also not given in the text)? Do you assume flat temporal trends in emission rates over the course of one month, then a step change to the rates of the next month? If so, you should add a sensitivity study how this low temporal resolution in the bottom-up products affect your inversion results. Didn't you have access to bottom-up products with a higher temporal resolution?**

  We have clarified this point in the manuscript. We are limited by the temporal resolution of the bottom-up products from the WETCHIMP study; those methane flux simulations have a monthly temporal resolution. With that said, observations during the first ten days of each month have footprints that extend into the previous month. As a result, the model estimate at any given site is often based upon wetland flux estimates from multiple time periods. In this way, the WRF-STILT model estimate is not a step change from one month to the next.

- **p.9346f.: Section 2.1 is almost of equal length compared to 2.1, even though also here you could refer many of the details to e.g. the Gourdji et al. reference. These are just minor details, but they make the paper appear unbalanced in parts.**

  We have expanded Section 2.1 to describe the data, atmospheric model, and WETCHIMP flux products in greater detail. Reviewer #2 asked a number of questions about the model selection framework, so we have also expanded Section 2.2 to better explain the statistical approach for a non-technical audience.

- **p.9347ff.: the strategy of the synthetic modeling setup needs to be rewritten in some parts. Some details are only given in the last paragraph, which are required earlier to understand the concept. For example, you only mention in the past paragraph that the 16 combinations of regions/seasons are optimised separately. Also, one thing that is not clear to me: in 1000 repeats different combinations of turning the 16 options for regions/seasons are randomly created. If each region/season gets an individual model fit through the BIC approach, why do you need the repeats?**

  The reviewer makes an astute point here. We have re-arranged the information in Section 2.3 as the reviewer suggested.

  The 1000 repeats are needed due to the random or stochastic nature of the synthetic data simulations. We add random noise to the synthetic data to simulate the effect of real-world modeling and measurement errors. We do not know the exact value of these modeling or measurement errors. Instead, we have an estimate of the properties of these errors (i.e., their variances and covariances), and we can simulate a plausible set of errors using these estimated properties and a random number generator. The results of the model selection can vary slightly, depending on the particular random numbers that we draw. Hence, we repeat the synthetic data experiments over and over again (1000 times in total) and average the results across all 1000 repeats. This procedure ensures that the model selection results are not the output of a single random number draw. We have clarified this setup in the revised manuscript.

- **p.9349, ll.25ff: Your assigned scaling factors for EDGAR emissions should be discussed in more detail as sources of uncertainty in the simulated mixing ratio time series! What about the influence of boundary layer height, which is certainly shallower in winter, and might thus exaggerate the influence of ground sources on mixing ratio changes in the atmosphere.**

  We do not scale the EDGAR emissions inventory in the revised manuscript. Many atmospheric CH_4 observation sites near wetlands are also located far from large anthropogenic emissions. As a result, any effort to scale the EDGAR inventory at these sites could be error-prone. Instead, we present the inventory as is.

  Miller et al. (2013) and Miller et al. (2014b) explore in detail the possible influence of boundary layer height. Miller et al. (2013), for example, found no significant seasonality in their US CH_4 emissions estimate (Fig. S8 in that paper). Seasonal bias in the estimated boundary layer height could manifest as erroneous seasonality in the emissions estimate. The absence of seasonality in estimated US emissions suggests an absence of bias in estimated boundary layer heights.

- **p.9350ff, Section 3, first part: I'm lacking a summarising conclusion/discussion here. To what extent does the ratio of natural to anthropogenic emissions influence the detectability of wetland fluxes? And to what extent is the network configuration responsible?**

  We have expanded the first synthetic data study to explore these questions in greater depth. For example, we explore how these results change if we set anthropogenic emissions to zero. Similarly, we explore how these results change if we set atmospheric transport errors to zero. These expanded results are summarized in Fig. 3 and section 3.

- **p.9351f, Section 3, second part: since the patterns displayed in Figs. 3a and 3b are virtually the same, the question arises whether you can truly separate the 2 effects you are looking after. After all, it boils down to the same question: What is the ratio of natural and anthropogenic emissions in a certain region/season, and how well is the observation network designed to capture these signals. I therefore strongly recommend to explain better if the 2 steps of virtual experiments truly provide different answers!**

  Part one of the synthetic data experiments asks a question of magnitude and part two asks a question related to the spatial distribution fluxes. In particular, part one investigates whether the observation network can detect any kind of atmospheric pattern from wetlands. Part one asks a basic question about about the detectability of wetlands over patterns from anthropogenic sources or from model errors. Part two investigates whether it matters to the observation network where those fluxes are located. Part two is prerequisite for the real data experiments in section 4.1.

  In many ways, it makes sense that the first and second synthetic data experiments produce similar results, but that result is not necessarily guaranteed. In regions with large wetland fluxes, those fluxes often display high spatial variance. The first synthetic data experiment often produces positive results when the wetland fluxes are large. The second synthetic data experiment often produces positive results when the fluxes display high spatial variance. Hence, the first and synthetic data experiments often produce similar results. Furthermore, we show in the revised manuscript that these results are not necessarily due to the ratio of natural and anthropogenic emissions (refer to Section 3 of the revised manuscript).

Fang et al. (2014) validated and tested the model selection framework used in the second synthetic experiment. That study used atmospheric observations and model selection to differentiate among spatial patterns in CO_2 flux estimates for North America. We have elaborated on this discussion in the revised manuscript.

- **p.9351ff, Section 4.1: I think it is a very important finding that plausible spatial patterns in CH4 emissions from bottom-up models are only based on land cover maps, not on the remotely sensed inundation maps. Here, you provide the only detailed recommendation to the bottom-up community how their model estimates can be improved! So this definitely deserves a more detailed discussion, and a more prominent place in the conclusions.**

  We have expanded this discussion in section 4.1 and have featured this result more prominently in the conclusions.

- **p.9354ff, Section 4.2: these results basically indicate that none of the bottom-up models is useful for North American regional simulations ... the summertime emissions seem to be extremely overestimated, so that the resulting seasonal courses in modeled data are opposite of what the observations show. This isn't discussed at all ... ??? I think what definitely needs to be added here is an uncertainty estimate of the background data set, and the scaling factors of the EDGAR emissions. Given the substantial overestimates in summertime emissions by virtually all models, it's hard to imagine how these models could be re-calibrated to match the observations, given that the other modeling components are correct ...**

  We would hesitate to say that none of the bottom-up models is "useful" for North American regional simulations. We would argue that there is an opportunity to tune the seasonal and spatial patterns in these bottom-up estimates. Similarly, the revised manuscript also offers several suggestions for improving future top-down emissions studies.

  We do not scale the EDGAR emissions inventory in the revised manuscript. We have also added a discussion on boundary condition uncertainties to the supplement. This discussion mirrors the boundary condition uncertainty analysis in Miller et al. (2013) and Miller et al. (2014b).

- **p.9354, l.12f: you need to provide an explanation why you restricted your time series analysis to only a few sites, and why you chose those 4.**

  We chose those four sites because they are located near large wetlands and in regions where the synthetic data experiments produced positive results; the wetland methane signal is easier to distinguish at these sites relative to others. In the revised manuscript, we have added plots of all remaining US and Canadian sites in the supplement (Fig. S4).

- **p.9355f, Section 4.3: I think it's not enough to base the seasonality analysis on relative flux contributions from each month alone. Since most of the bottom-up models (as shown in Fig.4) have very high flux emissions rates in**

summertime, it may well be that the shoulder season fluxes are matching the inverse product quite well, while there are simply way off in the peak of the warm season.

We have added a second panel to Fig. 5. That panel shows the absolute budget in Tg CH_4 per month. The revised version of Fig. 5 therefore allows the viewer to compare both the relative and absolute seasonal cycle in each model.

- **p.9356, ll.10ff: It's a plausible explanation that air temperatures are significantly decoupled from the conditions in the soil (where CH4 is produced) for fall, but not for spring ... even if you can show through NARR that soils start thawing in April, this isn't possible without air temperatures that are appropriately high ...**

  Another possible explanation is that the temperature threshold for CH_4 production may be too high in some existing models. Most models predict relatively small fluxes when soils are cold but still above freezing. And most predict dramatically larger fluxes in the summer when both soil and air temperatures are at their peak. Our analysis suggests that the shape of the seasonal cycle may be broader than that predicted by many models; the relative difference between cool-season and warm-season fluxes may not be as great as predicted by many bottom-up estimates. This conclusion is supported by flux measurements taken across the North Slope of Alaska by Donatella Zona (University of Sheffield). Her paper is currently under review. We have added to this discussion in section 4.3 of the revised manuscript.

- **p.9357, ll.3ff: As mentioned already above, this hasn't been discussed earlier, and I don't think this is the proper place to start with this kind of agenda. I agree with the general statement, but if you want to place it in a publication you need to be more constructive. Your results show that there are obviously still large discrepancies between the methane signal that is simulated by WETCHIMP models, and the methane signal as seen from the atmospheric observations. Still, you don't offer any conclusions how information from atmospheric methods might be used to improve the bottom-up models ...**

  The reviewer makes a good suggestion here. We have removed this statement from the conclusions.

- **p.9357, 2nd paragraph (ll.8ff): I think this part of the conclusions needs more details. You just list your basic findings, without even attempting to interpret where these differences come from. Also, you seem to assume that any atmospheric inverse modeling product (or the approach to link tower observations to surface fluxes through atmospheric inverse modeling) can be regarded as the 'truth', and all discrepancies with bottom-up products can be attributed to shortcomings in the latter.**

  We have revised the wording of the article to make the top-down, atmospheric analysis sound less absolute (relative to bottom-up modeling). We have also included more discussion in Sections 4.1-4.3 and in the Conclusions, discussion that emphasizes plausible reasons for any discrepancies between the top-down analysis in the paper and existing bottom-up estimates of wetland fluxes.

References

Fang, Y., Michalak, A. M., Shiga, Y. P., and Yadav, V.: Using atmospheric observations to evaluate the spatiotemporal variability of CO_2 fluxes simulated by terrestrial biospheric models, Biogeosciences, 11, 6985–6997, doi:10.5194/bg-11-6985-2014, 2014.

Hegarty, J., Draxler, R. R., Stein, A. F., Brioude, J., Mountain, M., Eluszkiewicz, J., Nehrkorn, T., Ngan, F., and Andrews, A.: Evaluation of Lagrangian particle dispersion models with measurements from controlled tracer releases, J. Appl. Meteorol. Clim., 52, 2623–2637, 2013.

Miller, S. M., Wofsy, S. C., Michalak, A. M., Kort, E. A., Andrews, A. E., Biraud, S. C., Dlugokencky, E. J., Eluszkiewicz, J., Fischer, M. L., Janssens-Maenhout, G., Miller, B. R., Miller, J. B., Montzka, S. A., Nehrkorn, T., and Sweeney, C.: Anthropogenic emissions of methane in the United States, P. Natl. Acad. Sci. USA, 110, 20 018–20 022, doi:10.1073/pnas.1314392110, 2013.

Miller, S. M., Michalak, A. M., and Levi, P. J.: Atmospheric inverse modeling with known physical bounds: an example from trace gas emissions, Geoscientific Model Development, 7, 303–315, doi:10.5194/gmd-7-303-2014, 2014a.

Miller, S. M., Worthy, D. E. J., Michalak, A. M., Wofsy, S. C., Kort, E. A., Havice, T. C., Andrews, A. E., Dlugokencky, E. J., Kaplan, J. O., Levi, P. J., Tian, H., and Zhang, B.: Observational constraints on the distribution, seasonality, and environmental predictors of North American boreal methane emissions, Global Biogeochem. Cy., 28, 146–160, doi:10.1002/2013GB004580, 2014b.

Nehrkorn, T., Eluszkiewicz, J., Wofsy, S. C., Lin, J. C., Gerbig, C., Longo, M., and Freitas, S.: Coupled Weather Research and Forecasting-Stochastic Time-Inverted Lagrangian Transport (WRF-STILT) model, Meteorol. Atmos. Phys., 107, 51–64, doi:10.1007/s00703-010-0068-x, 2010.

Reply to reviewer #2

S. M. Miller, et. al.

We would like to thank the reviewer for suggestions and comments on the manuscript. The reviewer's detailed suggestions have been very helpful in improving the manuscript. Below, we have included the reviewers comments (in bold) along with our reply and the associated changes/updates to the manuscript.

1 Summary

- **In general, a new method of statistical analysis (BIC for CH4 wetland flux model selection) needs to be tested fully to understand the abilities and limitations to give confidence to the results.**

  The statistical methods and associated applications in this paper are by no means new. Model selection based upon the BIC has been used widely across statistics and has been used in numerous top-down studies of CO_2 and CH_4 to select a flux model. Schwarz (1978) originally developed the Bayesian Information Criterion (BIC), and BIC-based model selection has subsequently been used widely in statistical modeling (Ramsey and Schafer, 2012). In more recent years, several studies have used BIC-based model selection to explain patterns in eddy-covariance CO_2 flux measurements (Mueller et al., 2010; Yadav et al., 2010). A number of atmospheric inverse modeling studies have also employed BIC-based model selection to choose bottom-up models or environmental datasets for the prior flux estimate; these inverse modeling studies have applied BIC-based model selection to both CO_2 (Gourdji et al., 2012) and CH_4 (Miller et al., 2013, 2014a) flux problems. In addition to the aforementioned applications, BIC-based model selection has also been used to investigate the detectability of atmospheric patterns from surface sources (Fang et al., 2014; Shiga et al., 2014; Fang and Michalak, 2015). The model selection methods outlined in those three papers are identical to those used in the present manuscript.

  In the revised manuscript, we have included more references to these sources to bolster the methods outlined in Section 2.2.

- **Examine the uncertainties for each model components, such as sensitivity of the methodology to transport errors, flux errors, background concentration estimates, other assumptions/approximations and their interactions, as well as the stability or robustness of the analysis. If a thorough analysis is not done, at least a full discussion of the potential problems is needed to put the preliminary results in the proper perspective.**

  The analysis in this paper is based upon several existing inverse modeling studies, studies that use the same WRF-STILT model simulations, background concentration estimate, covariance matrix parameters, etc. (Miller et al., 2013, 2014a,b). Those studies discuss potential modeling and measurement errors in great detail. For example, Miller et al. (2013) explore uncertainties in the estimated methane boundary condition, uncertainties in the estimated covariance matrix parameters (the parameters that define $\mathbf{\Psi}$), uncertainties due to atmospheric transport in WRF-STILT, uncertainties due to geological CH_4 sources, and uncertainties in the attribution of CH_4 to individual sources. Miller et al. (2014a) discuss uncertainties due to the nested meteorology domains in WRF-STILT, uncertainties in the methane boundary condition, uncertainties due to the sparsity of the CH_4 observation network, uncertainties in atmospheric transport estimated by WRF-STILT, and uncertainties in the covariance matrix parameters. Finally, Miller et al. (2014b) discuss uncertainties in the flux estimate due to the assumptions made by the statistical modeling framework.

We have added a new section to the supplement (S7) that summarizes and highlights the discussion in these previous papers.

2 General comments

- **In Section 3: 4 of the 7 models were selected from WETCHIMP for the synthetic data experiment to evaluate the BIC method. Presumably the authors expected (without sufficient reasons) the 3 remaining models to be 'unacceptable', including LPJ-Bern. Yet BIC results in section 4.1 indicate LPJ-Bern is selected most often when using real observations, bringing into question the value and correctness of this BIC method (in its current form). This suggests the rejection of LPJ-Bern could be a problematic assumption in section 3 and/or the BIC methodology is not working. Possibly other assumptions and approximations (such as uncertainty estimates, model transport errors, etc.) are incorrect, the whole new method needs to be evaluated much more thoroughly as noted above. In general, much stronger justifications are needed for model selection. The BIC analysis should be done for all WETCHIMP models in the synthetic data experiment. If the range of flux model variations is too large, then what is the range of applicability for the BIC method?**

  We have revised the manuscript to include all seven WETCHIMP models in the synthetic data experiment.

  In the original manuscript, we did not include three models in the synthetic experiments because those models had an anomalously large magnitude relative to the other model estimates. The larger the wetland flux, the more likely that the observation network can detect a methane pattern from wetlands. Therefore, if we conduct the synthetic data experiment using a flux model that has an anomalously large magnitude (see Fig. 4), we would concomitantly obtain anomalously optimistic results. Thus, we had not included these three models in the synthetic data experiments to avoid biasing the experiment results.

  The analysis in Section 4.1 examines a completely different question, one that is independent of the flux model's magnitude. Rather, that experiment examines which models have spatial distributions that are consistent with the data. That analysis disentangles the spatial distribution of the fluxes from other confounding questions of magnitude or seasonality. Unlike the first synthetic data experiment, the analysis in Section 4.1 is insensitive to the magnitude of the flux models.

  We have clarified the experimental design in sections 2.3 – 2.4 to prevent any confusion. In the revised manuscript, we have also included all seven WETCHIMP models in the synthetic data experiments. We have scaled the magnitude of those models to match Pickett-Heaps et al. (2011), Miller et al. (2014a), and Wecht et al. (2014); this step ensures that the BIC results in Section 3 will not be biased by anomalously large WETCHIMP flux models, as explained above.

- **The authors stated 'By contrast, the observation network is largely insensitive to spatial variability in wetland fluxes across the US; in most instances, the model selection framework favors a spatially-constant model over a wetland model for the two US regions.' (page 9352, lines 9-12). Was 'a spatially-constant model' included in the work and the BIC model 'favors' or selected it over the other models? The comparison to 'a spatially-constant model' in general needs to be explained and documented more clearly. If the BIC method actually selects the 'spatially-constant model' (a clearly wrong model), maybe this is an indication of the lack of ability of the method.**

  We have added additional explanation to the manuscript to clarify this concept. The 'spatially-constant model' is analogous to an intercept term in a regression. It is standard practice to include an intercept in regression modeling; this intercept is always included in the model, irrespective of the availability or suitability of other explanatory variables. In spatial statistics, that intercept term is a spatial constant, hence the phrase 'spatially-constant model.'

  The BIC-based model selection framework examines every possible combination of explanatory variables. The BIC-based framework will then score each combination of explanatory variables based upon how well each fits the observations and based upon the number of variables in each combination; combinations with a large number of variables receive a larger penalty (Eq. 1). Each candidate combination always includes an intercept term. It is plausible that none of the variables explain substantial variability in the observations. In such case, the best-scoring model may be one that only includes an intercept (the most basic possible model). The intercept term is almost certainly not the "true", correct model; the intercept does not represent the unknown methane flux processes that occur within soils. Similarly, the intercept almost never explains all variability in the observations. Rather, the model selection framework may choose the intercept-only model if other candidate variables have limited ability to explain the atmospheric data.

  A hypothetical example helps illustrate this concept. In theory, we could use model selection to explore whether observations at the LEF tower in Wisconsin can detect a methane pattern from wetland fluxes in Siberia. We could examine this question using a synthetic data experiment like the first experiment in Section 2.3. Any variability in the observations at LEF would be unrelated to methane fluxes in Siberia. Therefore, a model selection framework would not choose a Siberian flux model as an explanatory variable. That result is not necessarily a judgement on the accuracy of the Siberian flux model. Similarly, that result does not indicate that the intercept is a 'perfect model.' Rather, that result simply indicates that methane observations at LEF cannot be explained by fluxes in Siberia.

  In addition, it is important to note that BIC-based analysis in sections 3 and 4.1 address different questions, and we do not expect any one model to be selected by the model selection framework in all cases. The first synthetic data experiment explores whether the atmospheric observation network can detect an atmospheric pattern from wetland sources. The second experiment asks a more detailed question – whether the network is sensitive to the location of wetland fluxes. The network could select both the intercept and a wetland model in the first experiment but may only select the intercept (no wetland model) in the second experiment. In that case, the network can detect an atmospheric

CH_4 pattern from wetlands but cannot pinpoint the geographic origin of that atmospheric pattern.

For additional discussion, refer to Shiga et al. (2014); the authors used model selection to explore the detectability of anthropogenic CO_2 fluxes across North America. Our experimental design parallels that study.

- **The 3 subsections, 4.1, 4.2, 4.3 seem to be unrelated and somewhat conflict with each other. Sections 4.2 and 4.3 did not make use of the information from 4.1 that LPJ-Bern and SDGVM performed best in the BIC analysis. In fact, the best model from 4.1, LPJ-Bern performed poorly in the flux magnitude and seasonal maximum comparison in section 4.2. There is no explanation on this discrepancy. This could be an indication of the problem in the BIC analysis in selecting a poorly performing flux model.**

  Subsections 4.1, 4.2, and 4.3 analyze the spatial distribution, magnitude, and seasonality of each WETCHIMP model, respectively. Those sections examine each of these qualities individually; each section should be unrelated to the others. A model that best matches the magnitude of the observations will not necessarily have a spatial distribution that best matches the observations (and vice versa). Similarly, a model that best matches the magnitude of the observations at annual scale will not necessarily have a seasonality that best matches the observations (and vice versa). For example, the spatial distribution of LPJ-Bern compares favorably against the data, but its magnitude must be scaled downward to match the observations. This result is neither a discrepancy in the methodology nor a problem in the model selection framework.

- **Section 4.3 is a comparison of all 7 flux models to the results of another inversion model, Miller et al. 2014. Since inversion model results are highly uncertain, this simple model to model type comparison has little useful information unless the authors can show independently that the Miller et al. 2014 results are good and can serve as a benchmark for comparison. At the minimum, the WETCHIMP fluxes should be compared to a variety of inversion model results to see the uncertainty possible for the inversion model results. The authors should note in the manuscript that section 4.3 is only a comparison of models, which is different than comparison to 'available data' (page 9345, lines 6-9, see above).**

  We disagree with the premise that an observation-based estimate like this one provides 'little useful information.' Fig. 5 displays the uncertainties in the inverse modeling estimate. Refer to the supplement and Miller et al. (2014a) for a more detailed discussion of the uncertainties.

  There are few regional-scale inversions that estimate the seasonal cycle of wetland fluxes from the HBL using in situ or satellite methane observations. Miller et al. (2014a) is the only study that we are aware of. Three other top-down methane studies, Pickett-Heaps et al. (2011), Wecht et al. (2014), and Turner et al. (2015), examine methane fluxes from the HBL. However, those studies use the seasonality from a bottom-up model and do not estimate the seasonal cycle independently using atmospheric or satellite data. For reference, Melton et al. (2013) provides a list of existing methane inversion studies that are relevant to wetland fluxes.

3 Specific comments

- **Page 9345, line 1: 'biogeochemical models leveraged all available data', what is this vague 'all available data' referring to?**

  We have modified this line to be more specific.

- **Page 9346: "background" concentration – the methane concentration of air entering the North American regional domain? Do all STILT particles always leave the North American regional domain during the model simulations? Provide more information on the estimation of the background concentration and the whole model settings to enable other scientists to check and/or compare results.**

  Refer to bullet point #2 above and to the reply to Reviewer #1. Both discuss modeling uncertainties, including uncertainties in the estimated background concentrations.

- **Page 9346, lines 21-24: repeating lines 23-27 on page 9344.**

  We have condensed the text to remove this repetition.

- **Page 9347, line 7: change 'first term in Eq. (1)' to 'first two terms in Eq. (1)'.**

  We have updated the text accordingly.

- **Page 9348, line 29: 'We also include a spatial constant or intercept term in X'. What is the physical or numerical significance of the 'spatial constant or intercept term'. How do the results compared with or without the 'spatial constant or intercept term'?**

  Refer to the earlier discussion on the intercept for an explanation of this point.

- **Page 9353, line 26: change 'GEIMS' to 'GIEMS'.**

  We have changed the text accordingly.

- **Page 9354, lines 16-20: 'The estimated contribution of anthropogenic emissions from EDGAR v4.2FT2010 is added to this background (in red). Note that the estimated scaling factors for EDGAR (Sect. 2.4) are 1.7±0.3 at Chibougamau, 5.6±0.5 at East Trout Lake, 2.4±0.3 at Fraserdale, and 2.5±0.3 at Park Falls.' Explain the meaning of the different scaling factors and how realistic are they (up to 5.6x)? What are the spatial regions these scaling factors are applied to?**

  In the revised manuscript, we do not scale the EDGAR inventory. Many atmospheric CH_4 observation sites near wetlands are also located far from large anthropogenic emissions. As a result, any effort to scale the EDGAR inventory at these sites could be error-prone. Instead, we present the inventory as is.

- **Page 9364: left figure contains wrong information 'Observation site (Fig. 4)'. Units in the right figure conflicts with caption.**

  The figure should be correct as is.

- **Page 9368: Label each curve in Fig. 5 as in Fig. 4.**

  We have updated the figure accordingly.

4 Supplement

- **Page 2, lines 56-57: 'In this equation, zsynthetic (n x 1) represents the synthetic observations generated for an observation site'. What is the number of sites (k) and does n vary for each of the k sites?**

  We have added more detail to the supplement on the observation sites.

  The study includes 15 total tower-based observation sites and 17 regular aircraft-based observation sites. Refer to Miller et al. (2013), Miller et al. (2014a), or the NOAA program websites (`http://www.esrl.noaa.gov/gmd/ccgg/insitu/`, `http://www.esrl.noaa.gov/gmd/ccgg/aircraft/`) for more detail individual observation sites. The observations at each tower site are either daily flasks (most US sites) or daily averages of continuous data (Canadian sites). A few sites, including the tower sites in Oklahoma and in Maine only have weekly flask measurements. The frequency of the regular aircraft observations varies depending upon the site (see `http://www.esrl.noaa.gov/gmd/ccgg/aircraft/`).

- **Page 2, lines 71-72: 'The other WETCHIMP models, in contrast, predict much higher fluxes (Fig. 4)'. Fig. 4 only showed concentrations or mole fractions, change 'much higher fluxes' to 'much higher concentrations'.**

  We have changed the text accordingly.

- **Page 2, 80-82: 'The magnitude and spatial/temporal structure of these errors were estimated in Miller et al. (2013) for the US and Miller et al. (2014b) for Canada.' The referenced works were for different prior fluxes. Authors need to show why new error estimates are not needed.**

  There are two covariance matrices used in the paper (Fig. S3). The first matrix, $\mathbf{R}$ describes model-data mismatch errors – errors in the measurements, atmospheric transport, and errors due to the finite spatial or temporal resolution of WRF-STILT. These errors should be invariant to the choice of prior model.

  The second matrix, $\mathbf{Q}$, describes the residuals between the true fluxes (denoted $\boldsymbol{s}$) and prior model estimate ($\mathbf{X}\boldsymbol{\beta}$). The diagonal terms quantify the variance of these residuals and the off-diagonal terms quantify the spatial and temporal covariances in these residuals. The variances and covariances in $\mathbf{Q}$ can change depending upon the choice of prior model.

  In the synthetic data studies, we construct a statistical model that is representative of a prototypical real data inverse model. Similarly, we want to use values for $\mathbf{Q}$ and $\mathbf{R}$ that are representative of what one would likely encounter in a real-data setup. Miller et al. (2013) and Miller et al. (2014a) constructed real data inverse models over the US and Canada, respectively. Those studies used a model selection framework to find prior models that show optimal fit against available observations. In each study, the authors then estimated the elements of $\mathbf{Q}$ and $\mathbf{R}$ using that prior model. The resulting estimates of $\mathbf{Q}$ are representative of prior models that shows optimal agreement with atmospheric observations. We use these values of $\mathbf{Q}$ and $\mathbf{R}$ in the synthetic data studies. We also use these prototypical variances and covariances to generate random errors in the synthetic data (see Section 2.3).

  We have revised the treatment of the covariance matrix parameters in the real data experiments (Section 2.4). Previously, we had used covariance matrix parameters from Miller et al. (2013) and Miller et al. (2014a) in these experiments. In the revised manuscript, we re-estimate the covariance matrix parameters for each prior model. We estimate these parameters using Restricted Maximum Likelihood (RML) (Corbeil and Searle, 1976; Kitanidis, 1995; Michalak et al., 2004; Gourdji et al., 2012), the same procedure used in Miller et al. (2013) and Miller et al. (2014a).

We have added an explanation of these concepts in the supplement.

- **Page 3: Correct the difference in units in the Figure and caption.**

  The figure should be correct as is. We could not find any difference or discrepancy in the units.

References

Corbeil, R. R. and Searle, S. R.: Restricted maximum likelihood (REML) estimation of variance components in the mixed model, Technometrics, 18, pp. 31–38, 1976.

Fang, Y. and Michalak, A. M.: Atmospheric observations inform CO_2 flux responses to enviroclimatic drivers, Global Biogeochem. Cy., p. 2014GB005034, doi:10.1002/2014GB005034, 2015.

Fang, Y., Michalak, A. M., Shiga, Y. P., and Yadav, V.: Using atmospheric observations to evaluate the spatiotemporal variability of CO_2 fluxes simulated by terrestrial biospheric models, Biogeosciences, 11, 6985–6997, doi:10.5194/bg-11-6985-2014, 2014.

Gourdji, S. M., Mueller, K. L., Yadav, V., Huntzinger, D. N., Andrews, A. E., Trudeau, M., Petron, G., Nehrkorn, T., Eluszkiewicz, J., Henderson, J., Wen, D., Lin, J., Fischer, M., Sweeney, C., and Michalak, A. M.: North American CO_2 exchange: inter-comparison of modeled estimates with results from a fine-scale atmospheric inversion, Biogeosciences, 9, 457–475, doi:10.5194/bg-9-457-2012, 2012.

Kitanidis, P.: Quasi-linear geostatistical theory for inversing, Water Resour. Res., 31, 2411–2419, doi:10.1029/95WR01945, 1995.

Melton, J. R., Wania, R., Hodson, E. L., Poulter, B., Ringeval, B., Spahni, R., Bohn, T., Avis, C. A., Beerling, D. J., Chen, G., Eliseev, A. V., Denisov, S. N., Hopcroft, P. O., Lettenmaier, D. P., Riley, W. J., Singarayer, J. S., Subin, Z. M., Tian, H., Zürcher, S., Brovkin, V., van Bodegom, P. M., Kleinen, T., Yu, Z. C., and Kaplan, J. O.: Present state of global wetland extent and wetland methane modelling: conclusions from a model inter-comparison project (WETCHIMP), Biogeosciences, 10, 753–788, doi:10.5194/bg-10-753-2013, 2013.

Michalak, A., Bruhwiler, L., and Tans, P.: A geostatistical approach to surface flux estimation of atmospheric trace gases, J. Geophys. Res.-Atmos., 109, doi:10.1029/2003JD004422, 2004.

Miller, S. M., Wofsy, S. C., Michalak, A. M., Kort, E. A., Andrews, A. E., Biraud, S. C., Dlugokencky, E. J., Eluszkiewicz, J., Fischer, M. L., Janssens-Maenhout, G., Miller, B. R., Miller, J. B., Montzka, S. A., Nehrkorn, T., and Sweeney, C.: Anthropogenic emissions of methane in the United States, P. Natl. Acad. Sci. USA, 110, 20 018–20 022, doi:10.1073/pnas.1314392110, 2013.

Miller, S. M., Worthy, D. E. J., Michalak, A. M., Wofsy, S. C., Kort, E. A., Havice, T. C., Andrews, A. E., Dlugokencky, E. J., Kaplan, J. O., Levi, P. J., Tian, H., and Zhang, B.: Observational constraints on the distribution, seasonality, and environmental predictors of North American boreal methane emissions, Global Biogeochem. Cy., 28, 146–160, doi:10.1002/2013GB004580, 2014a.

Miller, S. M., Michalak, A. M., and Levi, P. J.: Atmospheric inverse modeling with known physical bounds: an example from trace gas emissions, Geoscientific Model Development, 7, 303–315, doi:10.5194/gmd-7-303-2014, 2014b.

Mueller, K. L., Yadav, V., Curtis, P. S., Vogel, C., and Michalak, A. M.: Attributing the variability of eddy-covariance CO_2 flux measurements across temporal scales using geostatistical regression for a mixed northern hardwood forest, Global Biogeochem. Cy., 24, doi: 10.1029/2009GB003642, 2010.

Pickett-Heaps, C. A., Jacob, D. J., Wecht, K. J., Kort, E. A., Wofsy, S. C., Diskin, G. S., Worthy, D. E. J., Kaplan, J. O., Bey, I., and Drevet, J.: Magnitude and seasonality of wetland methane emissions from the Hudson Bay Lowlands (Canada), Atmos. Chem. Phys., 11, 3773–3779, doi:10.5194/acp-11-3773-2011, 2011.

Ramsey, F. and Schafer, D.: The Statistical Sleuth: A Course in Methods of Data Analysis, Cengage Learning, Boston, 3 edn., 2012.

Schwarz, G.: Estimating the dimension of a model, Ann. Stat., 6, 461–464, doi:10.1214/aos/1176344136, 1978.

Shiga, Y. P., Michalak, A. M., Gourdji, S. M., Mueller, K. L., and Yadav, V.: Detecting fossil fuel emissions patterns from subcontinental regions using North American in situ CO_2 measurements, Geophys. Res. Lett., 41, 4381–4388, doi:10.1002/2014GL059684, 2014.

Turner, A. J., Jacob, D. J., Wecht, K. J., Maasakkers, J. D., Lundgren, E., Andrews, A. E., Biraud, S. C., Boesch, H., Bowman, K. W., Deutscher, N. M., Dubey, M. K., Griffith, D. W. T., Hase, F., Kuze, A., Notholt, J., Ohyama, H., Parker, R., Payne, V. H., Sussmann, R., Sweeney, C., Velazco, V. A., Warneke, T., Wennberg, P. O., and Wunch, D.: Estimating global and North American methane emissions with high spatial resolution using GOSAT satellite data, Atmos. Chem. Phys., 15, 7049–7069, doi:10.5194/acp-15-7049-2015, 2015.

Wecht, K. J., Jacob, D. J., Frankenberg, C., Jiang, Z., and Blake, D. R.: Mapping of North American methane emissions with high spatial resolution by inversion of SCIAMACHY satellite data, J. Geophys. Res.-Atmos., 119, 7741–7756, doi:10.1002/2014JD021551, 2014.

Yadav, V., Mueller, K. L., Dragoni, D., and Michalak, A. M.: A geostatistical synthesis study of factors affecting gross primary productivity in various ecosystems of North America, Biogeosciences, 7, 2655–2671, doi:10.5194/bg-7-2655-2010, 2010.