



Integrated Arctic Observation System

Research and Innovation Action under EC Horizon2020
Grant Agreement no. 727890

Project coordinator:
Nansen Environmental and Remote Sensing Center, Norway

Deliverable 6.7


GHG budgets – atmosphere

Quantification of GHG budgets for selected regions in the Arctic based on the assimilation of multi-disciplinary data layers into an atmospheric modelling framework, including the identification of key processes across disciplines that govern Arctic greenhouse gas cycles and their links to climate change

Start date of project:	01 December 2016	Duration:	63 months
Due date of deliverable:	31 May 2021	Actual submission date:	28 October 2021
Lead beneficiary for preparing the deliverable:	MPG		
Person-months used to produce deliverable:	28 pm		

Authors: Mathias Göckede, Friedemann Reum, Georg Heygster

Version	DATE	CHANGE RECORDS	LEAD AUTHOR
		Delivery date reset to <u>31.10.2021</u>	Executive Board
1.0	01.10.2021	Template	
1.1	15.10.2021	1 st Draft, submitted for internal review	M. Göckede
1.2	20.10.2021	Reviewed	T. Sachs
1.3	24.10.2021	Final version	M. Göckede
1.4	28.10.2021	Technical review and submission	K. Lygre

Approval	Date:	Sign.
X	28 October 2021	 Coordinator

USED PERSON-MONTHS FOR THIS DELIVERABLE					
No	Beneficiary	PM	No	Beneficiary	PM
1	NERSC		24	TDUE	
2	UiB		25	GINR	
3	IMR		48	UNEXE	
4	MISU		27	NIVA	
5	AWI		28	CNRS-LOCEAN	
6	IOPAN		29	U Helsinki	
7	DTU		30	GFZ	
8	AU		31	ARMINES	
9	GEUS		32	IGPAN	
10	FMI		33	U SLASKI	
11	UNIS		34	BSC	
12	NORDECO		35	DNV GL	
13	SMHI		36	RIHMI-WDC	
14	USFD		37	NIERSC	
15	NUIM		38	WHOI	
16	IFREMER		39	SIO	
17	MPG	28	40	UAF	
18	EUROGOOS		41	U Laval	
19	EUROCEAN		42	ONC	
20	UPM		43	NMEFC	
21/49	(UB) / Georg-Lab	x	44	RADI	
22	UHAM		45	KOPRI	
23	NORCE		46	NIPR	
			47	PRIC	

DISSEMINATION LEVEL		
PU	Public, fully open	X
CO	Confidential, restricted under conditions set out in Model Grant Agreement	
CI	Classified, information as referred to in Commission Decision 2001/844/EC	

EXECUTIVE SUMMARY

This document, *Deliverable 6.7 - Quantification of GHG budgets for selected regions in the Arctic based on the assimilation of multi-disciplinary data layers into an atmospheric modelling framework, including the identification of key processes across disciplines that govern Arctic greenhouse gas cycles and their links to climate change*, describes the work conducted within the atmospheric component of INTAROS Task 6.5. The main contributions for the work reported herein were produced by partner MPG, with additional contributions from UB.

The East Siberian Arctic Shelf hosts vast carbon reservoirs at risk of degradation and may be a strong emitter of methane to the atmosphere. Yet, estimates of its annual methane emissions and their key controls are highly uncertain. In the presented project, we estimated these emissions with a geostatistical inverse model from atmospheric observations over seventeen months in Tiksi (Russia), Barrow (Alaska) and Ambarchik (Russia). Our simulations yielded annual methane emissions of 0.3 – 1.5 Tg CH₄, which is on the low end of previously reported estimates (0 – 17 Tg CH₄ yr⁻¹). Our geostatistical approach allows us to test the compatibility of a large number of spatiotemporal emissions patterns with the atmospheric signals. In this context, we specifically tested the suitability of novel data products from the INTAROS database to improve model performance. Our model attributes highest emissions to shallow waters and to ice-free and potentially freeze-up periods, but also finds substantial emissions during the ice-covered period. We do not detect substantial emissions of stored methane during ice breakup. Our results suggest that mixing and stratification of the water column and cracks in sea ice could be among the dominant controls of methane emissions from the shelf to the atmosphere. Other explanations are possible and discussed, including limitations of our study. The information provided by the INTAROS database led to minor improvements in the explained variability of atmospheric greenhouse gas time series, indicating the high quality of the novel products. However, since parameter selection basically replaced existing oceanic variables by better-performing new ones for the same parameter but did not add a previously omitted parameter to the highest-ranking models, we could not gain novel process insights. Our study suggests that the relevance of the shelf for the global atmospheric methane burden is currently small, but also reveals limitations of the Arctic atmospheric greenhouse gas observation network.

Table of Contents

1.	Introduction	4
2.	Data and methods	6
2.1.	Geostatistical inverse modeling in the ESAS domain	6
2.1.1.	Model domain, transport model and boundary conditions	6

2.1.2.	Atmospheric data	7
2.1.3.	Inverse modeling framework	7
2.1.4.	Covariance matrices	8
2.1.5.	Selecting regression models and evaluating auxiliary variables	8
2.2.	Auxiliary variables	9
2.2.1.	Auxiliary variables used in the reference framework	9
2.2.2.	Additional auxiliary variables retrieved from the INTAROS database	12
3.	Results and discussion	14
3.1.	Results from the reference modeling framework	14
3.1.1.	Terrestrial and oceanic carbon budgets based on the reference framework	15
3.1.2.	Dominant auxiliary data layers, and associated processes	16
3.2.	Results with an extended database of auxiliary layers	19
3.2.1.	Terrestrial and oceanic carbon budgets with extended auxiliary layers	19
3.2.2.	Effect of additional auxiliary data layers from the INTAROS database	19
4.	Summary	22
5.	Exploitation plan	23
6.	Contributions to the roadmap towards an integrated Arctic Observing System (iAOS)	24
7.	Literature	25

1. Introduction

Carbon emissions from the Arctic Ocean domain into the atmosphere are poorly constrained to date, and particularly fluxes from the large shelf areas are subject to high uncertainties. At the same time, vast subsea carbon reservoirs in the Arctic are at risk of degradation (James et al., 2016), which may lead to increasing venting of greenhouse gases to the atmosphere under warming conditions. Insight into the mechanisms that govern the sustainability of these reservoirs and greenhouse gas exchange with the atmosphere is therefore essential for the assessment of Arctic carbon-climate feedbacks and the simulation of accurate future climate trajectories. At the same time, for the Arctic, and particularly for the Arctic Ocean, available observational databases are limited, compared to other global regions. Therefore, new data assimilation schemes need to be developed that optimize the integration of existing data sources across disciplines, thus helping to improve our understanding of carbon budgets and underlying carbon cycle processes.

One of these Arctic shelf regions, the East Siberian Arctic Shelf (ESAS), has in recent years attracted interest as a potential globally important source of CH₄ (Shakhova et al., 2010;

2014). The CH₄ cycle in the shelf is a complex system of a variety of sources, sinks, reservoirs and transport mechanisms. Accordingly, the net methane budget of the ESAS domain, and its potential reaction to Arctic climate change, is complicated to assess, and substantial uncertainties remain between existing studies that aimed at constraining these processes. Carbon sources in the ESAS seabed can be generally differentiated into sediments deposited by rivers and coastal erosion within the recent past (e.g. Vonk et al., 2012), subsea permafrost and gas hydrates originating from past ice ages (e.g. Romanovskii et al., 2005), as well as natural gas (Cramer and Franke, 2005). These reservoirs can act as primary sources of methane into the ESAS water, and eventually into the atmosphere, with higher emission rates expected in case warming destabilizes the carbon pools. Such warming can e.g. be linked to natural warming of the seafloor permafrost that has been induced from the top by ocean waters since the Holocene transgression (e.g. Dmitrenko et al., 2011), by warm water input from Arctic rivers (e.g. Romanovskii et al., 2005), or by a continuous geothermal heat flux from the bottom (e.g. Dmitrenko et al., 2011).

Constraints on present-day ESAS-wide methane emissions to the atmosphere are uncertain, with budget estimates in the literature ranging between virtually zero fluxes (e.g. Berchet et al., 2016) and emissions as high as 17 Tg CH₄ yr⁻¹ (Shakhova et al., 2014). In addition, the dominant patterns and mechanisms of methane release, including the role of sea ice and the spatial distribution of emissions, are uncertain and have not been assessed in a single framework with data that cover both the whole shelf and at least a full year. Part of the controversy regarding the net emission budgets are linked to the different methods employed to constrain these numbers; however, a dominant source for uncertainties in all studies focusing on this subject so far is the scarcity of observations, and therefore the need to extrapolate information to very distant, and often also different parts of the domain.

Our work conducted in the context of INTAROS WP 6.5 aims at an integrated assessment of the methane emissions from the shelf to the atmosphere, including estimating their magnitude, finding spatiotemporal patterns and, if possible, linking them to controlling mechanisms. Our primary approach, so-called geostatistical inverse modeling (GIM), combines observations from tower based atmospheric greenhouse gas mole fractions with simulations of atmospheric transport in the polar region and ancillary data layers that help to describe the spatio-temporal variability of greenhouse gas processes within the target domain. In this context, one major objective was to exploit new data sources that have become available in the INTAROS database and integrate them into an improved modeling framework that produces better matches between observed and simulated greenhouse gas patterns.

2. Data and methods

2.1. Geostatistical inverse modeling in the ESAS domain

Our method constrains methane emissions from surface sources to the atmosphere based on measurements of the atmospheric methane mole fraction observed at a distributed network of monitoring towers. The link between these quantities is atmospheric transport and solving for fluxes means inverting the transport equation. In other words, atmospheric transport modeling allows us to identify which areas within the target domain influenced an atmospheric GHG observation taken at a specific time and place, and therefore facilitates to attribute sink and source strengths to these areas that correlate with the observed patterns in atmospheric GHG mole fractions. Since observational datasets are scarce, usually the problem is under-constrained, and thus requires additional information sources to allow for non-equifinality across large subsets of potential solutions. In this context, valuable information can be assimilated from data layers that describe the structure of sources and sinks within the target domain, and their temporal variability – termed auxiliary variables in our framework. An objective evaluation of such auxiliary variables is possible using geostatistical inverse modeling.

2.1.1. Model domain, transport model and boundary conditions

We simulated surface-atmosphere fluxes in a domain of $3200 \text{ km} \times 1600 \text{ km}$ on 32 km resolution (see also Figure 1), comprising the East Siberian Arctic Shelf ($2.0 \times 10^6 \text{ km}^2$), other ocean surfaces ($1.2 \times 10^6 \text{ km}^2$), as well as adjacent tundra and taiga regions ($1.9 \times 10^6 \text{ km}^2$). The model was run for a period of 17 months from July 2014 to November 2015.



Figure 1: Modeling domain of the presented study, including the location of the three atmospheric monitoring stations Tiksi, Ambarchik and Barrow. The ESAS domain is separated here into the three subregions Laptev Sea (green shading), East Siberian Sea (blue) and Chukchi Sea (red). The terrestrial section of the model domain is actually about the same size as the ESAS domain (see also above).

We linked spatially distributed surface fluxes to atmospheric data via the Lagrangian transport model STILT in the setup customized for the Arctic domain by Henderson et al. (2015). STILT computes the surface influence on a measurement (“footprint”) based on an ensemble of back trajectories, and multiplication with surface fluxes yields methane enhancements over background. We compute the background by initializing the trajectories at the domain boundaries (or end of simulation) using a global methane transport model (Nunez Ramirez, pers. comm.), with fluxes optimized for atmospheric

data including the stations used in this study. Transport simulations accounted for chemical methane loss in both boundary conditions and footprints.

2.1.2. Atmospheric data

Methane emissions were optimized based on data from the Arctic monitoring stations Ambarchik (Reum et al., 2019), Barrow (Dlugokencky et al., 1995) and Tiksi (Ivakhov et al., 2019), i.e. all greenhouse gas monitoring stations operating continuously at the coast of the East Siberian Arctic Ocean in the study period (see also Figure 2). Data were selected following criteria used similarly in other inverse modeling studies. These criteria included requiring wind speeds of at least 3 ms^{-1} for Ambarchik and Barrow (following Sweeney et al., 2016) and 2.5 ms^{-1} for Tiksi (following station operator recommendations), and a temperature inversion filter for Ambarchik and Barrow.

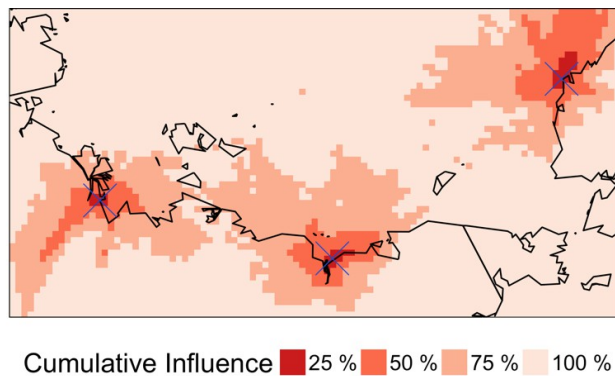


Figure 2: Footprints of the atmospheric data used for optimization in this study, aggregated over the whole study period and sorted into bins covering 25 % of the cumulative influence each.

2.1.3. Inverse modeling framework

Since inverting the transport equation is in general an under-constrained problem, meaningful results in atmospheric inverse modeling can only be obtained with some form of additional regularization. In greenhouse gas modeling, the Bayesian inverse modeling approach is often applied, where fluxes are nudged towards a prior estimate, and spatial and temporal covariances are imposed. Here, we employed the geostatistical inverse modeling approach as developed by Michalak et al. (2004) in the implementation by Miller et al. (2014a), which can be thought of a Bayesian model where prior and covariances are inferred from the atmospheric data. The inverse problem is formulated as minimizing the cost function $J_{inv}(\mathbf{f}, \boldsymbol{\beta})$:

$$J_{inv}(\mathbf{f}, \boldsymbol{\beta}) = (\mathbf{z} - \mathbf{H}\mathbf{f})^T \mathbf{R}^{-1} (\mathbf{z} - \mathbf{H}\mathbf{f}) + (\mathbf{f} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{Q}^{-1} (\mathbf{f} - \mathbf{X}\boldsymbol{\beta}) \quad (1)$$

Here, \mathbf{H} is the atmospheric transport model, \mathbf{z} observed atmospheric methane mole fractions and \mathbf{f} the posterior, optimized fluxes. \mathbf{R} and \mathbf{Q} are covariance matrices representing the so-called model-data-mismatch and the prior flux uncertainties, respectively. The regression model $\mathbf{X}\boldsymbol{\beta}$ consists of the regressor matrix \mathbf{X} , the columns of which are here called auxiliary variables X_i , and the regression model coefficients $\boldsymbol{\beta}$. Prior to optimizing fluxes, regression model coefficients are computed that satisfy:

$$(\mathbf{HX})^T \Psi^{-1} (\mathbf{HX}) \hat{\boldsymbol{\beta}} = (\Psi^{-1} \mathbf{HX})^T \mathbf{z}, \quad (2)$$

where

$$\Psi = \mathbf{HQH}^T + \mathbf{R}. \quad (3)$$

With \mathbf{Y} in equation (3) used as a summary parameter to simplify the other equations. The optimal fluxes are then

$$\hat{\mathbf{f}} = \mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{QH}^T \Psi^{-1} (\mathbf{z} - \mathbf{HX} \hat{\boldsymbol{\beta}}). \quad (4)$$

Our optimization algorithm allows only non-negative solutions (similar to Miller et al., 2014b), which is another useful regularization tool for inferring methane emissions (Miller et al., 2014a). Studies that use Bayesian inversion techniques usually report the statistical uncertainty of the optimized result. However, in our case, optimized flux budgets are primarily determined by the regression model, and this is not well accounted for in the Bayesian uncertainty estimate. Therefore, we focus on uncertainties due to varying model setups instead.

2.1.4. Covariance matrices

The covariance matrices \mathbf{R} and \mathbf{Q} determine the relative weights of data and regression model in the flux estimate, as well as time and length scales on which the optimized fluxes deviate from the regression model. We assumed \mathbf{R} to be diagonal with the same model-data mismatch standard deviation for all observations. We parameterized \mathbf{Q} by decomposing it into a spatial and a temporal component characterized by a prior uncertainty, a correlation length and a correlation time (Yadav and Michalak, 2013). In the spatial model, we removed correlations between land and ocean pixels. We optimized model-data mismatch, prior uncertainty, correlation length and correlation time with a restricted maximum likelihood approach following Michalak et al. (2004).

2.1.5. Selecting regression models and evaluating auxiliary variables

We evaluated the compatibility of regression models with atmospheric data via the Bayesian information criterion, BIC, following the description by Gourdji et al. (2012):

$$BIC = J_{BIC}(\hat{\boldsymbol{\beta}}) + p \cdot \ln(n), \quad (5)$$

where p is the number of auxiliary variables (columns of \mathbf{X}), and n is the number of observations. The first addend, $J_{BIC}(\hat{\boldsymbol{\beta}})$, is the minimum value of the cost function

$$J_{BIC}(\boldsymbol{\beta}) = \ln|\Psi| + (\mathbf{z} - \mathbf{HX}\boldsymbol{\beta})^T \Psi^{-1} (\mathbf{z} - \mathbf{HX}\boldsymbol{\beta}). \quad (6)$$

The BIC approach allows to find the best-fitting combinations of explanatory variables for the problem in question while at the same time aiming at keeping the degrees of freedom of the final equation as low as possible. In other words, a low BIC score indicates a regression model $\mathbf{X} \hat{\boldsymbol{\beta}}$ that fits the data well with a small number of explanatory variables, penalizing both overfitting and collinearity.

Since computing BIC is computationally cheap, in this study we used it to evaluate a large number of regression models. In the reference version of the modeling framework, i.e. in simulations that were generated prior to adding new information from the INTAROS database, we allowed combinations of up to seven auxiliary variables out of a pool of 33 (see Section 2.2.1 below for details). We used this as an objective way to find an unbiased prior flux estimate and as a tool for finding the dominant spatiotemporal patterns in the emissions. In the extended version, i.e., those simulations that included additional information extracted from the INTAROS database, five more data layers, and additional derived products, were considered (Section 2.2.2), while keeping the total number of allowed auxiliary variables allowed within a single model at the same level.

Kass and Raftery (1995) suggested a scale for the interpretation of the difference between BIC for different models. On this scale, a difference of less than 2 indicates that there is no evidence for the better performance of a model, while a difference of more than 10 is called "decisive" evidence. Following their suggestion, in our analyses we included all models with all-positive regression coefficients $\hat{\beta}$ and BIC scores of less than 10 over that of the best-scoring model and call them "well-scoring" regression models. For some analyses, we included only models that satisfied stricter criteria: being among the best 30 models in terms of BIC, RMSE, correlation and mean bias between modeled and observed atmospheric data. This applied to 3 – 8 models per parameter set, and we use these "best-scoring" models to verify the results of the well-scoring models. In these sets of models, we assigned each auxiliary variable a rank that indicates the rank of the best-scoring model that contains the variable. Thus, the best rank of 0 means the variable is contained in the top-scoring model, and a score of 1 indicates the variable is contained in the worst of the well-scoring models. Variables that are not featured in the well-scoring models do not receive a score. We used this scale to rank individual auxiliary variables according to their compatibility with atmospheric data.

2.2. Auxiliary variables

2.2.1. Auxiliary variables used in the reference framework

Based on the available literature on methane emissions from the East Siberian Arctic Shelf and adjacent land areas, we created a set of 33 auxiliary variables, each of which can be linked to a specific surface emission process. Each of these data layers consists of a gridded map representing a unique spatial or spatio-temporal, respectively, variability that holds the potential to explain emission patterns.

Within the oceanic part of the domain, some static data layers describe possible spatial distribution of emissions related e.g., to emission hot spots, permafrost cover, the shelf edge, and attenuation of emissions with increasing ocean depth. Temporarily variable data layers include, e.g., variables that are related to wind speed, among them diffusive gas transfer and attempts to isolate the role of storms. Next, the role of sea ice was tested by variables representing open ocean fraction, ice growth and ice retreat. We also

considered multiplicative combinations of the variables that are related to ocean depth with the sea ice cover and diffusive gas transfer variables.

In addition, we developed variables that describe terrestrial emissions in the model domain. Most importantly, these represent the spatial distribution of wetlands and lakes within the model domain. We described these with results from complex process-based models from the literature as well as simple models for wetland and zero curtain emissions. Fire and anthropogenic emissions were considered small in our domain, and thus ignored.

Table 1: Auxiliary variables used to describe methane emissions in the reference model framework.

Label	Description
ESAS spatial distribution of subsea emissions	
Hot spots	Hot spot map based on Shakhova et al. (2010)
Const. ESAS	Constant ESAS emissions
ESAS<30m	Area of ESAS shallower than 30m
ESAS>30m	Area of ESAS deeper than 30m
Edge	ESAS edge (100 m – 500 m depth)
Ocean depth	
exp-10m	Exponential attenuation with ocean depth ($\alpha = 0.1 \text{ m}^{-1}$)*
exp-20m	Exponential attenuation with ocean depth ($\alpha = 0.05 \text{ m}^{-1}$)
exp-100m	Exponential attenuation with ocean depth ($\alpha = 0.01 \text{ m}^{-1}$)
Kolyma mouth	
Kolyma mouth	Location of Kolyma mouth
Diffusive gas transfer	
k	Gas transfer velocity
Storms	
Storms-0.5-10	Storm occurrence ($d = 0.5 \text{ m}^{-1}\text{s}$, $w_{v,0} = 10 \text{ ms}^{-1}$)**
Storms-1-10	Storm occurrence ($d = 1 \text{ m}^{-1}\text{s}$, $w_{v,0} = 15 \text{ ms}^{-1}$)
Storms-0.5-15	Storm occurrence ($d = 0.5 \text{ m}^{-1}\text{s}$, $w_{v,0} = 10 \text{ ms}^{-1}$)
Storms-1-15	Storm occurrence ($d = 1 \text{ m}^{-1}\text{s}$, $w_{v,0} = 15 \text{ ms}^{-1}$)
Sea ice	
Const. ESAS-Ice	ESAS fraction not covered by sea ice
Ice growth	Sea ice growth
Ice retreat	Sea ice retreat
Combinations of variables	
exp-10m-ice	“exp-10m” modulated by sea ice cover
exp-20m-ice	“exp-20m” modulated by sea ice cover
exp-100m-ice	“exp-100m” modulated by sea ice cover
ESAS<30m-ice	“ESAS<30” modulated by sea ice cover
ESAS>30m-ice	“ESAS>30” modulated by sea ice cover
Edge-ice	“Edge” modulated by sea ice cover
Hot spots-ice	“Hot spots” modulated by sea ice cover
Ice growth-edge	“Ice growth” multiplied with “Edge”
k-exp-10m	“k” multiplied with “exp-10m”
k-exp-20m	“k” multiplied with “exp-20m”
k-exp-100m	“k” multiplied with “exp-100m”
Process-based models for terrestrial emissions	
WSL	Climatology of CH ₄ emissions based on model LPJ-WSL
bLake4Me	Climatology of CH ₄ emissions based on model bLake4Me
Simple models for terrestrial emissions	
Const. land	Constant land emissions
SWM	Simple Wetland Model
ZCM	Zero Curtain Model

*The attenuation factor α describes all mechanisms that control how much CH₄ dissolves on the way through the water column from ocean bottom to the atmosphere.

**Different combinations of a scaling coefficient d and a cutoff windspeed w were used to determine which pixels within the model domain were affected by storms.

A detailed description of the derivation of the auxiliary data layers listed in Table 1 is omitted here for means of brevity, but will be made available in the supporting information of a publication in preparation (Reum et al., 2022). To facilitate a direct comparison to the additional data layers obtained from the INTAROS database (see Section 2.2.2), some details on sea ice products used are summarized here. To describe the physical sea ice barrier, the remote-sensing sea ice concentration product ASI-SSMI (Kaleschke et al., 2001; Spreen et al., 2008) was used, which is distributed with a resolution of 12.5 km, spatially and temporally interpolated, and with a 5-day median filter applied. Sea ice growth was described by the differences between consecutive time steps of the sea ice concentration product. This approach, postulating a quasi-instantaneous link between ice formation and CH₄ emissions, was not suitable for describing mixing of the complete water column to the average depth of the ESAS, which occurs on a timescale of months (Janout et al., 2016). However, it may capture mixing of upper water layers or CH₄ release from brine. Finally, methane accumulated below or in sea ice in winter may be released upon ice breakup, a process that may influence any region of the temporarily ice-covered ocean in case the sea ice can be regarded as a mostly impermeable cap. This process was modeled as parameter ‘ice retreat’ similarly to sea ice growth based on differences between consecutive time steps of the sea ice concentration product. Shortcomings of this approach are that it may be sensitive to noise in the sea ice concentration product, accumulation times are ignored, and that the resulting variable is sensitive to both sea ice melt and sea ice drift.

2.2.2. Additional auxiliary variables retrieved from the INTAROS database

At the time of writing, the INTAROS Data Catalogue (<https://catalog-intaros.nersc.no>) comprised 137 datasets provided by 38 contributing organizations. These datasets cover a wide variety of spatial and temporal resolution, target domains and timeframes, and research fields. The range of contributions included, e.g., site-level data on vegetation characteristics or carbon and energy flux rates, transects of ocean observations based on various commercial and research vessels, or spatially explicit gridded datasets on ocean and land surface properties based on remote sensing information.

For the presented data assimilation study, the basic requirement for the provision of new auxiliary data layers was a continuous spatial grid of information that covered the entire target domain. As described above, such data layers may both be static or time-varying. In the latter case, a continuous record of observations within the timeframe set for this experiment was required. Since most of the information provided by the INTAROS database was either at site-level or episodic coverage of ocean transects, only two datasets remained that could be tested for our data assimilation approach. Descriptions in the following were largely taken over from the INTAROS database:

Atmospheric Total Water Vapor over ice and open ocean (University of Bremen, UB). This dataset merged the precipitable water vapor content from the microwave imager AMSR-E/2 over open water and from the microwave sounder AMSU-B and MHS over ice, creating an Arctic-wide daily dataset for total water vapor (TWV) of 50 km

resolution with seamless coverage from the high Arctic to mid latitudes from 2002 to date (see Figure 3 for an example). This new approach, described in detail in Scarlat et al. (2018), allowed to apply the method also to regions where previously no data were available, and ensured a more consistent physical analysis of the satellite measurements by taking into account the contribution of the surface emissivity to the measured signal.

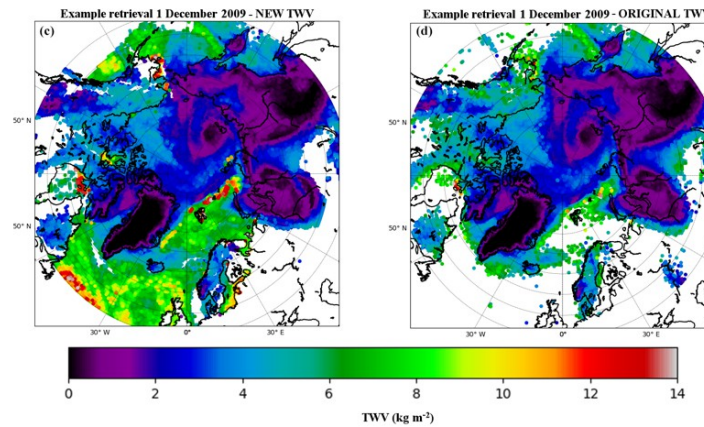


Figure 3: Daily TWV maps of the Northern Hemisphere obtained from the new algorithm (c) and compared to the original AMSU-B algorithm (d) on December 01, 2009. Figure taken from Scarlat et al. (2018), modified.

Ocean-Sea Ice Synthesis from 2007-2016 (University of Hamburg, UHH). Using the Massachusetts Institute of Technology general circulation model (MITgcm, Marshall et al., 1997) and its adjoint, in this project both in situ and remote sensing observations were used to produce new synthesis products for the Arctic. The model domain covers the entire Arctic Ocean, north of the Bering Strait and ~44N in the Atlantic Ocean. The data assimilation produced a substantially improved representation of the daily mean state of Arctic sea-surface temperature (SST), sea ice concentration (SIC), and sea ice thickness (SIT) (Figure 4). Datasets also include continuous grids of e.g. ocean potential temperature, salinity, zonal and meridional velocity, freshwater, or heat fluxes at the sea surface (Lyu et al., 2021).

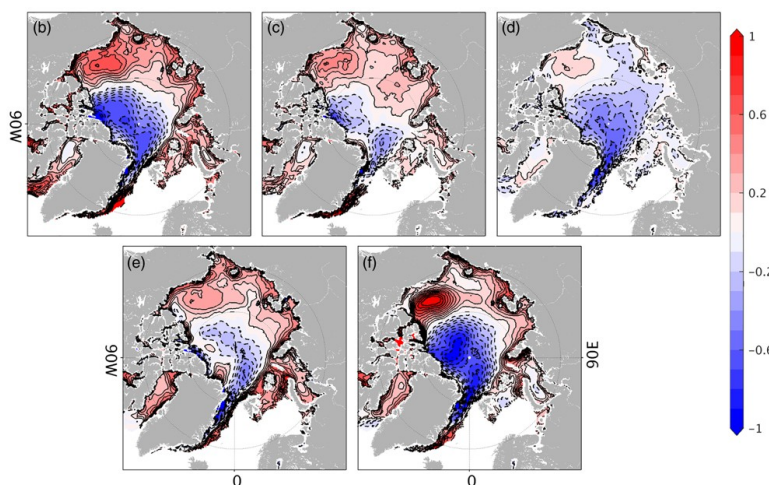


Figure 4: Mean sea-ice thickness (SIT) differences between CryoSat2-SMOS merged data and (b) INTAROS-ctrl, (c) INTAROS-opt, (d) TOPAZ4, (e) PIOMAS, and (f) ECCOv4r4. The contour interval is 0.1 m. Figure taken from Lyu et al. (2021).

For integration into the presented data assimilation study, all new datasets were reduced to the selected simulation timeframe, and re-projected onto the target grid with a regular 32 km resolution that is described above in Section 2.1.1. From the UB dataset on total water vapor (TWV), only the original product was used. From the UHH dataset, a subset of the available list of parameters was chosen, including additional derived parameters as described below in Table 2.

Table 2: Additional primary auxiliary variables based on information from the INTAROS database, which were added to form the extended version of the model framework.

Label	Description
UB: Total water vapor dataset	
UB_TWV	Total water vapor concentration
UHH: Arctic synthesis products based on MITgcm	
ESAS Ice	Open ocean fraction (1 – Sea Ice Concentration)
mitgcm_SIT	Sea Ice Thickness
mitgcm_SITinv	Inverse Sea Ice Thickness (1/SIT)
mitgcm_salt	Ocean salinity
mitgcm_qnet	Net surface heat flux
mitgcm_qnetCool	Positive surface heat flux (-1*Qnet, positive values only)

In the extended version of the data assimilation framework, besides considering the original data layers also possible derived products made use of new information provided by the added data layers from the INTAROS database. This included e.g. the derivation of a diffusive transfer coefficient k using the UHH salinity product, or the replacement of the previous open ocean fraction layer by a new one based on the MITgcm synthesis product.

3. Results and discussion

The key results summarized in the following sections are not based on a single model run, therefore they do not provide single numbers as e.g. flux budgets for the target domain, or a unique ranking of processes influencing the methane emissions. Instead, we provide statistics across multiple inversion results that reflect both the diverse modeling setups as well as the lists of combinations of auxiliary variables that e.g. constitute the well-scoring models optimized for each of these combinations. Accordingly, budgets are given as ranges, and the evaluation of the roles of specific processes is based on their relative importance across multiple model runs.

3.1. Results from the reference modeling framework

The simulations based on the reference modeling framework, i.e. the base runs excluding the new data layers from the INTAROS database, indicated an enhanced role of bathymetry, i.e. dominant emissions from shallow water regions of the ESAS domain. In addition, we were able to derive indications for flux contributions linked to sea ice dynamics, where our results suggested strong emissions during sea ice formation in fall, sustained emissions throughout the winter in spite of a seemingly

closed sea ice cover, and only marginal emissions during sea ice retreat. Several emission patterns that were developed based on the literature are not supported by our results. This applies to – besides the case of sea ice melt – all variables that relate shelf emissions to wind speed, enhanced methane release linked to geologic fault lines, and emissions from the shelf edge. This suggests that these patterns and associated processes, even though they may contribute to the total budgets, do not dominate the spatiotemporal distribution of emissions.

3.1.1. Terrestrial and oceanic carbon budgets based on the reference framework

Inversion results demonstrate that terrestrial emissions clearly dominate the overall methane budget within the study domain (Figure 5). Auxiliary variables associated with land variables yield, depending on the regression model, an average signal strength of 7 – 16 ppb at the stations. Average emission rates after optimization are 4.6 – 6.8 mg CH₄ m⁻² d⁻¹, which sums up to an annual budget of 3.2 – 4.7 Tg CH₄ yr⁻¹. These results are in good agreement with previously published budgets of process-based models for wetlands (0.4 – 2.0 Tg CH₄ yr⁻¹) and the lake model (2.1 Tg CH₄ yr⁻¹), which can be combined to 2.5 – 4.1 Tg CH₄ yr⁻¹. This estimate may be too high due to a potential overlapping of wetland and lake areas in the models, and associated double accounting of emissions (Thornton et al., 2016). At the same time, the budgets produced within the context of this study may be biased by the missing restriction of wetland area in our Simple Wetland Model and Zero Curtain Model. Although not the focus of this study, it is interesting to note that the majority of the terrestrial emission budget is attributed to the ‘simple’ variables such as ZCM. This may imply that the rather complex process-based wetland and lake methane emission models do not capture the full spatial extent of emissions and potentially underestimate cold season emissions.

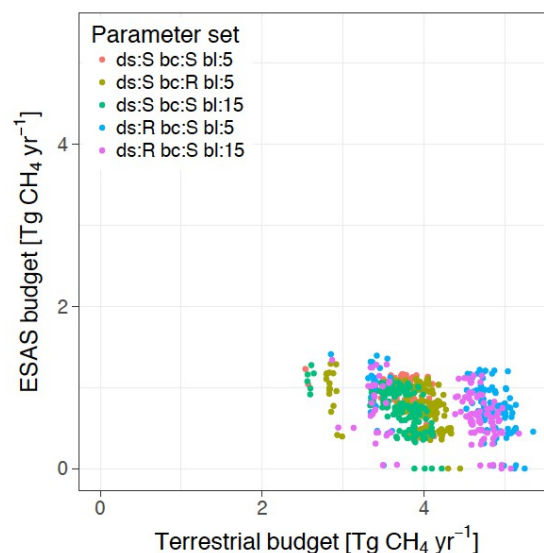


Figure 5: Annual CH₄ budgets assigned by well-scoring regression models to ESAS and land regions (Oct 2014 – Sep 2015). Different colors reflect the range of sensitivity studies performed to test e.g. the influence of background concentration fields, simulation times for backtrajectories, or methane mole fraction data filtering.

In comparison to terrestrial results, the average signal strength of ESAS auxiliary variables was smaller at 0 – 7.5 ppb, with average emission rates after optimization of 0.5 – 2.0 mg CH₄ m⁻² d⁻¹. This imbalance implies that ESAS signals are difficult to quantify at the existing coastal sites because of the influence of the larger terrestrial emissions on the atmospheric signals. Summing up these fluxes to annual numbers, our regression models yielded ESAS emission budgets of 0 – 1.4 Tg CH₄ yr⁻¹. We conducted several tests to analyze the sensitivity of these estimates to settings of the inversion framework, each of which revealed only small sensitivity of the ESAS budget to model uncertainties. Moreover, simulated atmospheric methane time series (Figure 6) matched very well with the observations across the three observation sites. In summary, even though our estimates for cumulative annual ESAS methane emissions were on the low end of literature estimates, we are confident in the quality of our findings.

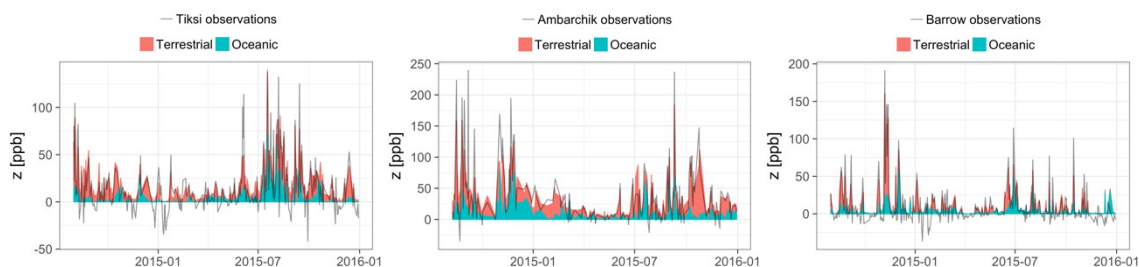


Figure 6: Example of modeled atmospheric CH₄ mole fractions at Tiksi (left), Ambarchik (center) and Barrow (right) based on posterior emission estimates of a single inversion scenario (“Coastal flux + Sea ice growth”, with relaxed atmospheric data selection).

3.1.2. Dominant auxiliary data layers, and associated processes

The best-scoring regression models in the reference simulations tend to feature the shallowest ocean variable not modulated by sea ice cover (exp-10m), and, depending on the filter applied to the atmospheric data, the sea ice growth variable. Terrestrial variables in the best-scoring models included, e.g., the constant land flux and the Zero Curtain Model (ZCM).

Regarding the influence of bathymetry, our inversion results indicated that the more emissions were attributed to shallow ESAS waters, the better the respective emission pattern was compatible with the atmospheric observations, and the more successful its performance was in model selection. This observation may be explained by the fact that methane released from the seafloor is more efficiently trapped in deeper water, as the shelf water is stably stratified most of the year (Janout et al., 2016) and the pycnocline depth is roughly on the order of 10 m in the ice-free season (Thornton et al., 2016). Furthermore, more bubbles could dissolve when traversing a deeper water column (e.g. Leifer and Patro, 2002). Other possible causes for higher seafloor emissions close to the coast may be a restriction of the permafrost layer to shallow waters (Ruppel, 2015), or fluxes linked to the degradation of recently deposited terrestrial carbon that is primarily accumulated in the near-shore region via erosion and rivers (e.g. Vonk et al., 2012). In summary, while our results do not provide direct insight into the processes below the

water surface, they are in close agreement with previous studies that also found larger emissions in shallower waters.

Regarding the influence of sea ice variables on the output of the reference inversion framework, without additional data filters the sea ice growth variable was selected in 70 – 83% of well-scoring models, with an average budget of $0.4 \text{ Tg CH}_4 \text{ yr}^{-1}$ assigned to this parameter. This makes it the best-scoring variable among all variables that describe shelf emissions. Regarding temporal variability, the inversions attributed the strongest emissions to fall, (October and November) especially in 2014 (Figure 7). Average emission rates estimated by inversions for fall were $0.9 - 3.8 \text{ mg CH}_4 \text{ m}^{-2} \text{ d}^{-1}$, with best-scoring regression models on the high end. This by far exceeds the estimates for the period of lowest ice cover (August and September), which are only $0.8 - 1.6 \text{ mg CH}_4 \text{ m}^{-2} \text{ d}^{-1}$. This finding suggests that mixing of the water column due to brine rejection may be a major factor driving annual emissions (Damm et al., 2015). Although no wind speed-related variables performed well in our regression model selection, we do not rule out the possibility that storms in the fall season could also contribute to the result.

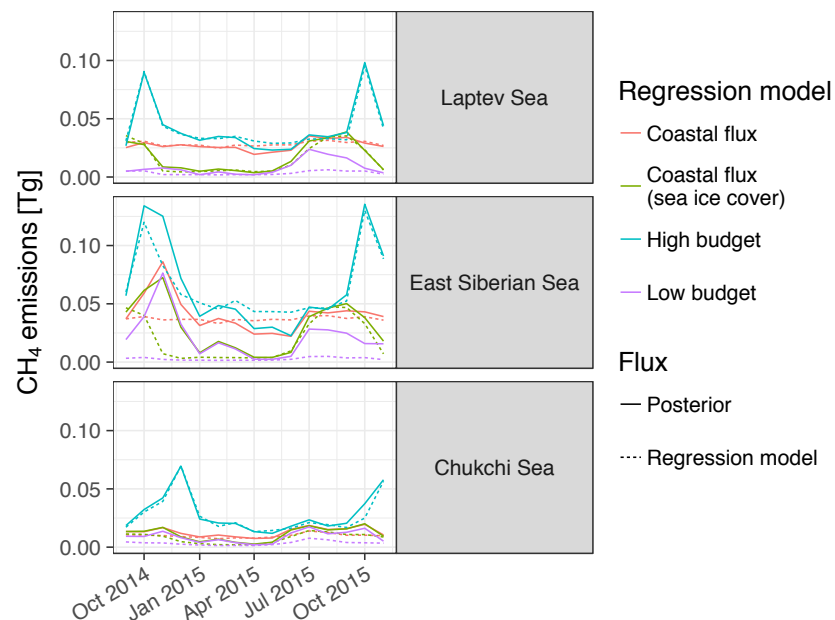


Figure 7: Monthly methane emissions from the three parts of the ESAS for four implementations of the inversion framework. The results shown here represent the range of budget estimates and the influence of varying the regression model with sea ice cover.

The elevated role of fall time emissions linked to sea ice growth was substantially reduced when applying a special data filter to the Ambarchik time series that aimed at removing transport situations potentially influenced by recirculating air masses between terrestrial and oceanic domains. In this case, the sea ice growth variable was not contained in the well performing regression models (selected in only 2% of well-scoring models, with a budget of $<0.1 \text{ Tg CH}_4 \text{ yr}^{-1}$), and in the corresponding inversions emissions in fall were reduced to the magnitude of summer emissions. This substantial difference in both model selection and assigned flux rates indicates that the attribution of flux signals between terrestrial and oceanic domains was highly sensitive to the data filter applied to

atmospheric observations from coastal sites, and accordingly uncertainties in the findings are considerable. Without additional data constraints such as e.g., isotopic signatures, a clear differentiation of high oceanic emissions from recirculated terrestrial air masses remains difficult.

In regression model selection, variables that did not use sea ice cover as a modulating parameter for flux patterns performed slightly worse than those that use that parameter. In other words, even though the role of the sea ice growth parameter seems to be dependent on the filter applied to the atmospheric GHG mole fraction data, the consideration of an open ocean fraction (derived from sea ice concentration) generally improved the explanatory power of other ocean parameters such as e.g., bathymetry. In inversions, the emissions in the period with highest ice-cover (December – May, 0.3 – 1.7 mg CH₄ m⁻² d⁻¹) rivaled those of the period of lowest ice cover (August – September, 0.8–1.6 mg CH₄ m⁻² d⁻¹). These findings could imply an enhanced role of methane emissions through cracks in the ice cover, as e.g. observed previously from aircraft data (Kort et al., 2012). The comparatively high and sustained wintertime emissions, despite the much smaller emission area of open water areas (e.g. leads, polynias) during that period, could be explained by continuous mixing of the water column due to refreezing and convective cooling, which can result in trace gas emission rates up to two orders of magnitude above those in open water (e.g. Lowry et al., 2018).

One hypothesis postulated when entering this project was that there is a potential for enhanced methane emissions associated with sea ice retreat, since during ice melt, CH₄ accumulated under or stored in ice can be released to the atmosphere (e.g. Zhou et al., 2014). However, in regression model selection, the variable describing sea ice retreat was rarely selected and ranked low among the well-scoring models. The annual budget attributed to the pattern was less than 0.1 Tg CH₄ yr⁻¹ in all cases. Consistent with this, optimized emission rates during sea ice retreat (June–July) were lower (standard data selection: 0.7 – 1.4 mg CH₄ m⁻² d⁻¹) than those of the period of lowest ice cover (standard data selection: 0.8 – 1.6 mg CH₄ m⁻² d⁻¹). Although the effective emission area is smaller, which may point to slightly elevated emission rates compared to open water, these results suggest that stored methane is not vented in amounts significant for the annual budget during ice-breakup (see also Figure 7). Instead, it points to limited accumulation below and in sea ice, and therefore indicates a marginal role of the sea ice cover to act as an effective trapping mechanism (see also previous paragraph). One aspect that could not be considered by our analysis in this context is the different characteristics of first year ice vs. second year ice. The latter remains quite impermeable even during periods of temperature variability, while the former quickly becomes permeable during warm spells. At the same time, during the melting process the formation of a melt water layer that caps the methane-rich lower ocean water layers can form an effective trapping mechanism that restricts emissions (Damm et al., 2015), which may also contribute to keeping net emissions low during that time.

3.2. Results with an extended database of auxiliary layers

All evaluations including the new INTAROS data layers were performed with the additional filter to the atmospheric dataset that aimed at excluding transport situations with air masses re-circulating between terrestrial and oceanic domains at the Ambarchik site (see above). As a consequence of this strict data filter, the influence of a sea ice growth control parameter was strongly reduced, as were the overall methane emission signals attributed to the ESAS domain in fall.

Adding the new information for the extended version of the GIM framework (Section 2.2.2), we specifically checked the following candidate processes that may be associated with the newly added auxiliary data layers from the INTAROS database:

1. Effect of the replacement of the previously used dataset describing sea ice concentration dynamics with the new UHH product.
2. Use of the inverse of sea ice thickness ($1/SIT$) as a proxy for air-ice exchange (e.g. Loose et al., 2011), which may be correlated with enhanced methane emissions in the marginal ice zone along the ice edge.
3. Use of a 'cooling product', based on the UHH surface heat flux data layer, which may indicate an enhanced mixing of the surface ocean layers associated with negative heat fluxes, and related emissions of methane into the atmosphere.
4. Use of the UHH salinity dataset for calculation of more accurate estimates of the diffusive gas transfer coefficient k , which previously was computed with a constant salinity value.
5. Inclusion of the total water vapor product by UB as a general proxy for surface-atmosphere exchange processes, and atmospheric transport patterns.

The total number of candidate auxiliary layers was reduced for this exercise, i.e. we limited the selection process to those layers that had been identified as the most important in the reference modeling framework, and added the new layers based on information from the INTAROS database. The list of auxiliary layers that contributed to the model selection is provided in Table 3 below, which already shows the ranking after the optimization process.

3.2.1. Terrestrial and oceanic carbon budgets with extended auxiliary layers

As will be shown in more detail below, the addition of new auxiliary data layers led to only marginal differences in the optimization output. Even though certain improvements to explained variance in atmospheric methane mole fractions were obtained, there was no substantial shift in the assignment of net flux budgets from terrestrial or oceanic domains. The numbers provided in Section 3.1.1 thus also apply to the extended output discussed here.

3.2.2. Effect of additional auxiliary data layers from the INTAROS database

As mentioned above, due to the consideration of different scenarios in model setup, but also due to the reason that there are often only marginal differences in the performance

of different combinations of auxiliary variables forming the top-ranked models within this data assimilation, the approach applied here does not provide a single best solution. Instead, we resorted to evaluating the data layer usage within the family of best-scoring or well-scoring models, respectively. In this context, the importance of an auxiliary variable is primarily indicated by the number of times it gets included into models, and the highest-ranking model it contributes to.

Table 3 and Figure 8 summarize the statistics for the auxiliary data layers that were selected for the well-scoring models of the extended atmospheric inversion framework. In total, 15 different data layers, or derived products reflecting combinations of multiple data layers, were considered for the models that produced the best agreement between simulations and observations. As for the reference run, the parameters that determine the terrestrial methane emissions (WSL, constant land, ZCM) clearly stood out of the most frequently selected pieces of information, and the combination of those three actually also formed the highest-ranked model. This reflects the fact that oceanic methane emissions are of significantly lower magnitude compared to the terrestrial emissions, so that the performance of land fluxes dominates the cost function of the optimization. The only ocean parameters that played a considerable role in the model formulation, i.e. those that were selected for more than two models, are simple bathymetry, sea ice cover, and the UHH surface cooling flux. All other parameters play only a marginal role.

Table 3: Ranking of auxiliary variables chosen in the well-scoring models produced by the extended version of the geostatistical atmospheric inversion framework. n: number of models including a specific parameter; rank: highest-ranking model including the specific parameter.

Order	Label	Description	New layer	n	rank
1	WSL	Climatology of CH ₄ emissions, model LPJ-WSL	No	34	1
2	Const. land	Constant land emissions	No	30	1
3	ZCM	Zero Curtain Model	No	21	1
4	exp-10m	Exponential attenuation with ocean depth	No	6	3
5	ESAS Ice_mitgcm*	Open ocean fraction (1 – Sea Ice Concentration)	Yes	6	4
6	mitgcm_qnetCool*	Surface cooling flux (-1*Qnet, positive values)	Yes	5	7
7	ESAS Ice_SSMI*	Open ocean fraction (1 – Sea Ice Concentration)	No	4	5
8	k	Gas transfer velocity, constant salinity	No	2	10
9	k coastal*	Gas transfer velocity, constant salinity	No	2	16
10	k coastal, ice*	Gas transfer velocity, constant salinity	No	2	12
11	k_mitgcm	Gas transfer velocity, using salinity data layer	Yes	2	17
12	k coastal_mitgcm*	Gas transfer velocity, using salinity data layer	Yes	2	15
13	k coastal_mitgcm, ice*	Gas transfer velocity, using salinity data layer	Yes	2	13
14	UB_TWV	Total water vapor concentration	Yes	2	11
15	EI_cool*	ERA interim cooling flux	No	1	19

*parameter additionally scaled with exp-10m data layer

Regarding the integration of the novel data layers from INTAROS into the data assimilation scheme, and the hypotheses about the impact on the model performance listed above, the following findings were obtained:

1. Replacement of the sea ice cover product, from formerly SSMI to the newly generated product provided by UHH, resulted in a slightly better model

performance (see also Figure 8, exp-10m-ice-mitgcm vs. exp-10mic-ssmi). The derived open ocean fraction, scaled with bathymetry, ranked 5th in the parameter list shown in Table 3, two positions above the SSMI product that was used in the reference framework. While the new data layer yielded a marginal improvement in the correlation between simulated and observed atmospheric methane mole fractions, the net emissions assigned to the ESAS domain by the deterministic model were virtually identical between the respective model versions.

2. The inverse of the UHH sea ice concentration product was not included in a well-scoring model for a single time. Accordingly, it does not play a role for the definition of spatio-temporal methane flux patterns that influence the atmospheric observations used within this modeling study. This may imply that either air-ice-exchange in the marginal ice zone is very low within the target domain, or alternatively the chosen product is only a weak proxy it, or that such processes are not picked up by the coastal observation sites used within the context of this study.
3. The surface heat flux from the UHH dataset (rank 6), converted to a cooling flux product, was clearly superior to the previously used cooling flux provided by ERA interim (rank 15), shown also in Figure 8 (cooling-mitgcm-exp-10m vs. cooling-erai-exp-10m). The additional deterministic flux assigned to the ESAS domain, linked to this parameter alone, sums up to 0.2 Tg CH₄ yr⁻¹.
4. The use of a spatio-temporally varying salinity product taken from the UHH dataset to compute improved gas transfer velocity coefficients k did not make a difference for the inversion. In fact, the coefficient of correlation between k computed using a constant salinity or k computed using the data fields from UHH was >0.999, so differences are negligible.
5. The total water vapor product provided by UB ranked similarly low as the gas transfer velocity parameters (see also Figure 8), i.e. did not play a substantial role for the generation of successful deterministic models. The highest-ranking model including this parameter was listed at position 11, and only two models in total made use of this information. TWV basically competes with the other low-ranking ocean variables (its major flux contribution is over the ocean).

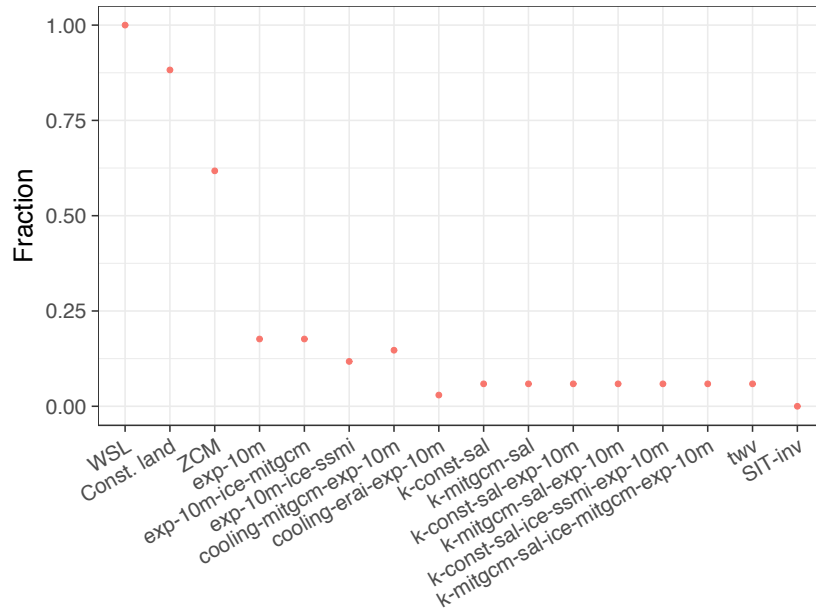


Figure 8: Fraction of top ranked regression models containing the auxiliary variables listed in Table 3. A fraction of 1 implies that all models contained a specific parameter, while a fraction of 0 implies that none used it. Results correspond to the inverse of parameter n given in Table 3.

For all models produced by the reference and extended inversion frameworks, the fraction of unexplained variability in the atmospheric methane mole fraction time series remained high even after optimization. Due to the dominance of terrestrial emissions within the chosen study domain, the addition of ocean parameters mostly led to only marginal improvements in the explanatory power of the deterministic models. This is owing to the fact that the observational database for Arctic Ocean atmospheric inversion studies can to date only be provided by coastal observation sites. Even though the integrated field of view of these coastal networks combines a considerable fraction of ocean influence to the observed dataset, the total influence of terrestrial emissions is still much higher, thus their influence dominates the cost function.

4. Summary

This report summarized a study using a data assimilation scheme based on a geostatistical inverse modeling framework to constrain methane emissions over the East Siberian Arctic Shelf, identify environmental conditions that explain spatio-temporal patterns in surface-atmosphere emissions, and link the latter to biogeochemical and biogeophysical processes governing the methane cycle in the target domain. After establishing a reference model framework and using it for in-depth model sensitivity studies and process investigation, in a second step we added new data layers from the INTAROS database, and tested their effect on model performance, and selection of dominant auxiliary data layers.

Based on the reference modeling framework, we estimated annual methane emissions from the East Siberian Arctic Shelf to the atmosphere at $0 - 1.4 \text{ Tg CH}_4 \text{ yr}^{-1}$, which is on the low end of existing literature estimates. Highest emissions were attributed to shallow

waters, while no emission spike was observed during sea ice retreat, indicating low accumulation of methane under the ice in winter. We also found potentially substantial emissions in fall and sustained emissions during winter, but these findings were sensitive to filters applied to atmospheric observation data, thus we have lower confidence in them. All results could be explained by two underlying processes: first, trapping of methane below the pycnocline could be responsible for a dominance of emissions from shallow waters, the potential emissions in fall (sea-ice growth, storms) and the missing emission spike during sea ice retreat (melt-water barrier). Second, significant emissions through cracks in sea ice could explain winter emission estimates and the missing methane spike during sea ice retreat (low accumulation). However, even though we interpret these processes as the most likely explanation of our findings, the results could also be explained by other factors, and limitations of our method could have contributed to several results.

Addition of new data layers from the INTAROS database focused on two products, provided by partners UHH and UB, which fulfilled the format requirements for data assimilation into the chosen atmospheric inversion framework, i.e. provision of spatially continuous, gridded information. In addition to the five parameters that could be extracted from these new datasets, several derived products were developed, all of which were tested as potential auxiliary data layers to explain spatio-temporal variability in methane emissions. Our results revealed that the improved sea ice concentration product provided by UHH was preferred over the previously used dataset. Moreover, the surface heat fluxes from that same dataset were clearly superior over the ERA interim cooling flux data product that had previously been used in the inversion. Other new parameters played only a marginal role (e.g., the total water vapor product by UB), or were not used in model selection at all (e.g., the sea ice thickness parameter from UHH). Accordingly, the assimilation of new data improved model performance, while no major effects on the total emissions in simulated flux fields of methane was found. Since the overall structure of chosen data layers combined in the top-ranking models did not substantially change, and successfully chosen new data layers mostly replaced similar products from other sources that were previously used, no new insights about processes and controls governing the shelf area methane cycle could be derived.

5. Exploitation plan

The presented study sheds new light on disagreements on the magnitude of methane emissions from the East Siberian Arctic Shelf to the atmosphere in the literature and provides insights into the controls of the emissions. Our study suggests that in the selected model domain, which is supposed to be representative for the high Arctic coastal region, terrestrial emissions from northern wetlands are at present more important than shelf emissions for the atmospheric methane burden. Our inversion results placed shelf emissions predominantly into shallow water areas, which is compatible with the

assumption that pycnocline inhibition strongly limits emissions from deeper areas, among other factors.

Sensitivity studies revealed the importance for future studies focusing on improving inaccuracies in atmospheric transport modeling in the Arctic especially in winter, which is particularly relevant for coastal sites that may temporarily be affected by small-scale circulation systems that complicate the separation between terrestrial and oceanic air masses. In this context, there is a general need for expanding the atmospheric greenhouse gas observation network in the Arctic, with a particular demand in observations representing the oceanic domain. The full potential of the presented data assimilation framework to deliver novel insights into controls and mechanisms governing large-scale greenhouse gas exchange processes over the Arctic Ocean can only be reached through installing new monitoring stations, and provision of additional parameters such as e.g., stable isotope signatures of carbon species that allow better constraining the source processes.

Alleviating these constraints, the data assimilation approach presented herein could be demonstrated to be a powerful research tool for exploiting additional data sources to generate new information on carbon budgets and underlying processes in previously understudied, data-poor domains. Future research should focus on assimilating additional datasets that e.g., take into account the role of transport with ocean currents, or provide novel links to oceanic biogeochemistry, in order to provide a more comprehensive view of the role of ocean shelves in the Arctic carbon cycle. Interdisciplinary collaboration with data providers from different scientific areas, but also a with statisticians and database managers who convert products into the requested formats, such as organized within the INTAROS project, are a prerequisite to achieve these goals.

6. Contributions to the roadmap towards an integrated Arctic Observing System (iAOS)

Our study demonstrated the unique potential of atmospheric inverse modeling as an important component of a pan-Arctic greenhouse gas monitoring system. Using the atmosphere as an integrator of heterogeneous surface-atmosphere exchange processes within structured Arctic landscapes, inverse modeling can provide data-driven constraints of greenhouse gas budgets at regional to pan-Arctic scales. Moreover, the application of geostatistical inverse modeling allows to assimilate multi-disciplinary data layers into the framework, test their explanatory power towards observed spatio-temporal patterns in atmospheric greenhouse gas mole fractions, and this way derive novel insights into effective control mechanisms for Arctic greenhouse gas processes at larger scales, including both terrestrial and oceanic domains.

As clearly demonstrated by the recent 6th report of working group 1 of the IPCC, the Northern circumpolar permafrost region is a key player within the global carbon budget,

and degradation of existing permafrost carbon pools may have catastrophic impacts on future global climate trajectories. Next to site level and local scale observation tools such as e.g. flux chambers and eddy-covariance flux monitoring, atmospheric inverse modeling needs to be established as an operational, large-scale monitoring tool capable of providing pan-Arctic budgets of greenhouse gas exchange, broken up into sub-regional and also sub-monthly resolution. Provided a sufficiently large database of atmospheric observations (see also below), inverse modeling assessments therefore form the best tool currently available for providing a comprehensive Arctic greenhouse gas budget at regular intervals (e.g. annual). Operated over longer timeframes, changes in budgets and long-term trends may indicate altered feedbacks between permafrost carbon and climate change, so the tool may serve as an early warning system for upcoming tipping points.

To maximize the impact of inverse atmospheric modeling as a component within an integrated Arctic observing system, the following investments would need to be implemented:

- Improve network coverage of towers that continuously monitor well-calibrated in-situ greenhouse gas mole fractions of important greenhouse gases
- Improve the network coverage of tower sites that, besides continuously monitoring the major greenhouse gases, also capture air samples for analysis of minor greenhouse gases and isotopic information, which can be instrumental for constraining source processes and regions.
- Improve the coverage of satellite remote sensing based greenhouse gas observations within the Arctic domain. Due to the special Arctic conditions affecting data quality (cloud cover, light availability), active sensors should particularly be promoted in this context.
- Provision of additional data layers from multiple geoscientific disciplines that describe the static or dynamic state of Arctic ecosystems, and which can be integrated into data assimilation frameworks as presented herein to describe and explain spatio-temporal variability in greenhouse gas emission patterns. Such information would be particularly valuable to enhance insights into the carbon cycle processes of the Arctic Ocean domain, which to date are poorly constrained.
- Finally, more resources are required to improve the accuracy and resolution of atmospheric transport modeling within the polar region.

Given these upgrades can be provided, geostatistical inverse modeling can constitute one of the central elements of a future iAOS, providing policy-relevant information as an operational monitoring tool, and integrating information sources across disciplines.

7. Literature

Berchet, A., Bousquet, P., Pison, I., Locatelli, R., Chevallier, F., Paris, J. D., Dlugokencky, E. J., Laurila, T., Hatakka, J., Viisanen, Y., Worthy, D. E. J., Nisbet, E., Fisher, R., France, J.,

- Lowry, D., Ivakhov, V., and Hermansen, O.: Atmospheric constraints on the methane emissions from the East Siberian Shelf, *Atmos. Chem. Phys.*, 16, 4147-4157, 2016.
- Cramer, B., and Franke, D.: INDICATIONS FOR AN ACTIVE PETROLEUM SYSTEM IN THE LAPTEV SEA, NE SIBERIA, *Journal of Petroleum Geology*, 28, 369-384, 2005.
- Damm, E., Rudels, B., Schauer, U., Mau, S., and Dieckmann, G.: Methane excess in Arctic surface water- triggered by sea ice formation and melting, *Sci. Rep.*, 5, 16179, 2015.
- Dlugokencky, E. J., Steele, L. P., Lang, P. M., and Masarie, K. A.: Atmospheric methane at Mauna Loa and Barrow observatories: Presentation and analysis of in situ measurements, *J. Geophys. Res.-Atmos.*, 100, 23103-23113, 1995.
- Dmitrenko, I. A., Kirillov, S. A., Tremblay, L. B., Kassens, H., Anisimov, O. A., Lavrov, S. A., Razumov, S. O., and Grigoriev, M. N.: Recent changes in shelf hydrography in the Siberian Arctic: Potential for subsea permafrost instability, *J. Geophys. Res.-Oceans*, 116, 2011.
- 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 CO₂ exchange: inter-comparison of modeled estimates with results from a fine-scale atmospheric inversion, *Biogeosciences*, 9, 457-475, 2012.
- Henderson, J. M., Eluszkiewicz, J., Mountain, M. E., Nehrkorn, T., Chang, R. Y. W., Karion, A., Miller, J. B., Sweeney, C., Steiner, N., Wofsy, S. C., and Miller, C. E.: Atmospheric transport simulations in support of the Carbon in Arctic Reservoirs Vulnerability Experiment (CARVE), *Atmos. Chem. Phys.*, 15, 4093-4116, 2015.
- Ivakhov, V. M., Paramonova, N. N., Privalov, V. I., Zinchenko, A. V., Loskutova, M. A., Makshtas, A. P., Kustov, V. Y., Laurila, T., Aurela, M., and Asmi, E.: Atmospheric Concentration of Carbon Dioxide at Tiksi and Cape Baranov Stations in 2010–2017, *Russian Meteorology and Hydrology*, 44, 291-299, 2019.
- James, R. H., Bousquet, P., Bussmann, I., Haeckel, M., Kipfer, R., Leifer, I., Niemann, H., Ostrovsky, I., Piskozub, J., Rehder, G., Treude, T., Vielstädte, L., and Greinert, J.: Effects of climate change on methane emissions from seafloor sediments in the Arctic Ocean: A review, *Limnology and Oceanography*, 61, S283-S299, 2016.
- Janout, M. A., Hölemann, J., Waite, A. M., Krumpen, T., von Appen, W.-J., and Martynov, F.: Sea-ice retreat controls timing of summer plankton blooms in the Eastern Arctic Ocean, *Geophys. Res. Lett.*, 43, 12,493-412,501, 2016.
- Kaleschke, L., Lüpkes, C., Vihma, T., Haarpaintner, J., Bochert, A., Hartmann, J., and Heygster, G.: SSM/I Sea Ice Remote Sensing for Mesoscale Ocean-Atmosphere Interaction Analysis, *Canadian Journal of Remote Sensing*, 27, 526-537, 2001.
- Kass, R. E., and Raftery, A. E.: Bayes Factors, *Journal of the American Statistical Association*, 90, 773-795, 1995.
- Kort, E. A., Wofsy, S. C., Daube, B. C., Diao, M., Elkins, J. W., Gao, R. S., Hintsä, E. J., Hurst, D. F., Jimenez, R., Moore, F. L., Spackman, J. R., and Zondlo, M. A.: Atmospheric observations of Arctic Ocean methane emissions up to 82 degrees north, *Nat. Geosci.*, 5, 318-321, 2012.
- Leifer, I., and Patro, R. K.: The bubble mechanism for methane transport from the shallow sea bed to the surface: A review and sensitivity study, *Continental Shelf Research*, 22, 2409-2428, 2002.

- Loose, B., Schlosser, P., Perovich, D., Ringelberg, D., Ho, D. T., Takahashi, T., Richter-Menge, J., Reynolds, C. M., McGillis, W. R., and Tison, J. L.: Gas diffusion through columnar laboratory sea ice: implications for mixed-layer ventilation of CO₂ in the seasonal ice zone, *Tellus B: Chemical and Physical Meteorology*, 63, 23-39, 2011.
- Lowry, K. E., Pickart, R. S., Selz, V., Mills, M. M., Pacini, A., Lewis, K. M., Joy-Warren, H. L., Nobre, C., van Dijken, G. L., Grondin, P.-L., Ferland, J., and Arrigo, K. R.: Under-Ice Phytoplankton Blooms Inhibited by Spring Convective Mixing in Refreezing Leads, *J. Geophys. Res.-Oceans*, 123, 90-109, 2018.
- Lyu, G., Koehl, A., Serra, N., Stammer, D., and Xie, J.: Arctic ocean–sea ice reanalysis for the period 2007–2016 using the adjoint method, *Q. J. Roy. Meteorol. Soc.*, 147, 1908-1929, 2021.
- Marshall, J., Adcroft, A., Hill, C., Perelman, L., and Heisey, C.: A finite-volume, incompressible Navier Stokes model for studies of the ocean on parallel computers, *J. Geophys. Res.-Oceans*, 102, 5753-5766, 1997.
- Michalak, A. M., Bruhwiler, L., and Tans, P. P.: A geostatistical approach to surface flux estimation of atmospheric trace gases, *J. Geophys. Res.-Atmos.*, 109, D14109, 2004.
- Miller, S. M., Michalak, A. M., and Levi, P. J.: Atmospheric inverse modeling with known physical bounds: an example from trace gas emissions, *Geosci. Model Dev.*, 7, 303-315, 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, *Glob. Biogeochem. Cy.*, 28, 146-160, 2014b.
- Reum, F., Göckede, M., Lavric, J. V., Kolle, O., Zimov, S., Zimov, N., Pallandt, M., and Heimann, M.: Accurate measurements of atmospheric carbon dioxide and methane mole fractions at the Siberian coastal site Ambarchik, *Atmos. Meas. Tech.*, 12, 5717-5740, 2019.
- Romanovskii, N. N., Hubberten, H. W., Gavrillov, A. V., Eliseeva, A. A., and Tipenko, G. S.: Offshore permafrost and gas hydrate stability zone on the shelf of East Siberian Seas, *Geo-Marine Letters*, 25, 167-182, 2005.
- Ruppel, C.: Permafrost-Associated Gas Hydrate: Is It Really Approximately 1 % of the Global System?, *Journal of Chemical & Engineering Data*, 60, 429-436, 2015.
- Scarlat, R. C., Melsheimer, C., and Heygster, G.: Retrieval of total water vapour in the Arctic using microwave humidity sounders, *Atmos. Meas. Tech.*, 11, 2067-2084, 2018.
- Shakhova, N., Semiletov, I., Leifer, I., Salyuk, A., Rekant, P., and Kosmach, D.: Geochemical and geophysical evidence of methane release over the East Siberian Arctic Shelf, *J. Geophys. Res.-Oceans*, 115, n/a-n/a, 2010.
- Shakhova, N., Semiletov, I., Leifer, I., Sergienko, V., Salyuk, A., Kosmach, D., Chernykh, D., Stubbs, C., Nicolsky, D., Tumskey, V., and Gustafsson, O.: Ebullition and storm-induced methane release from the East Siberian Arctic Shelf, *Nat. Geosci.*, 7, 64-70, 2014.
- Spreen, G., Kaleschke, L., and Heygster, G.: Sea ice remote sensing using AMSR-E 89-GHz channels, *J. Geophys. Res.-Oceans*, 113, 2008.

Sweeney, C., Dlugokencky, E., Miller, C. E., Wofsy, S., Karion, A., Dinardo, S., Chang, R. Y. W., Miller, J. B., Bruhwiler, L., Crotwell, A. M., Newberger, T., McKain, K., Stone, R. S., Wolter, S. E., Lang, P. E., and Tans, P.: No significant increase in long-term CH₄ emissions on North Slope of Alaska despite significant increase in air temperature, *Geophys. Res. Lett.*, 43, 6604-6611, 2016.

Thornton, B. F., Wik, M., and Crill, P. M.: Double-counting challenges the accuracy of high-latitude methane inventories, *Geophys. Res. Lett.*, 43, 12,569-512,577, 2016.

Vonk, J. E., Sanchez-Garcia, L., van Dongen, B. E., Alling, V., Kosmach, D., Charkin, A., Semiletov, I. P., Dudarev, O. V., Shakhova, N., Roos, P., Eglinton, T. I., Andersson, A., and Gustafsson, O.: Activation of old carbon by erosion of coastal and subsea permafrost in Arctic Siberia, *Nature*, 489, 137-140, 2012.

Yadav, V., and Michalak, A. M.: Improving computational efficiency in large linear inverse problems: an example from carbon dioxide flux estimation, *Geosci. Model Dev.*, 6, 583-590, 2013.

Zhou, J., Tison, J. L., Carnat, G., Geilfus, N. X., and Delille, B.: Physical controls on the storage of methane in landfast sea ice, *Cryosphere*, 8, 1019-1029, 2014.

----- END of DOCUMENT-----



INTAROS

This report is made under the project
Integrated Arctic Observation System (INTAROS)
funded by the European Commission Horizon 2020 program
Grant Agreement no. 727890.



Project partners:

