Causality guided machine learning model on wetland CH$_4$ emissions across global wetlands

Kunxiaojia Yuan$^{a,}$, Qing Zhu$^{a,}$, Fa Li$^{a,}$, William J. Riley$^a$, Margaret Torn$^a$, Housen Chu$^a$, Gavin McNicol$^b$, Min Chen$^c$, Sara Knox$^{d,}$, Kyle Delwiche$^{e,}$, Huayi Wu$^f$, Dennis Baldocchi$^g$, Hongxu Ma$^g$, Ankur R. Desai$^h$, Jiqian Chen$^i$, Torsten Sachs$^j$, Masahito Ueyama$^k$, Oliver Sonnentag$^l$, Manuel Helbig$^m$, Eeva-Stiina Tuittila$^n$, Gerald Jurasinski$^o$, Franziska Koebsch$^p$, David Campbell$^q$, Hans Peter Schmid$^r$, Annalea Lohila$^s$, Mathias Goeckede$^t$, Mats B. Nilsson$^u$, Thomas Friborg$^v$, Joachim Jansen$^w$, Donatella Zona$^x$, Eugenie Euskirchen$^y$, Eric J. Ward$^z$, Gil Bohrer$^{aa,}$, Zhenong Jin$^{ab,}$, Licheng Liu$^{ab,}$, Hiroki Iwata$^{ac,}$, Jordan Goodrich$^{ad,}$, Robert Jackson$^{*ad}$

$^a$ Climate and Ecosystem Sciences Division, Climate Sciences Department, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
$^b$ Department of Earth and Environmental Sciences, University of Illinois at Chicago, Chicago, IL, USA
$^c$ Department of Forest and Wildlife Ecology, University of Wisconsin-Madison, Madison, WI, USA
$^d$ Department of Geography, The University of British Columbia, Vancouver, British Columbia, Canada
$^e$ Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA, USA
$^f$ State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China
$^g$ Department of Geography, University of California, Berkeley, CA, USA
$^h$ Department of Atmospheric and Oceanic Sciences, University of Wisconsin-Madison, Madison, WI, USA
$^i$ Department of Geography, Environment, and Spatial Sciences, Michigan State University, East Lansing, MI, USA
$^j$ GFZ German Research Centre for Geosciences, Potsdam, Germany
$^k$ Graduate School of Life and Environmental Sciences, Osaka Prefecture University, Sakai, Japan
$^l$ Departement de Geographie, Universite de Montr`eal, Montr`eal, QC, Canada
$^m$ Department of Physics and Atmospheric Science, Dalhousie University, Halifax, NS, Canada
$^n$ School of Earth Sciences, University of Eastern Finland, Joensuu, Finland
$^o$ Landscape Ecology, University of Rostock, Rostock, Germany
$^p$ Digital Forest, University of Göttingen, Göttingen, Germany
$^q$ School of Science, University of Waikato, Hamilton, New Zealand
$^r$ Karlsruhe Institute of Technology, Institute of Meteorology and Climate Research, Karlsruhe, Germany
$^s$ Institute for Atmospheric and Earth System Research/Forest Sciences, University of Helsinki, Helsinki, Finland
$^t$ Department of Biogeochemical Signals, Max Planck Institute for Biogeochemistry, Jena, Germany
$^u$ Department of Forest Ecology and Management, Swedish University of Agricultural Sciences, Umeå, Sweden
$^v$ Department of Geosciences and Natural Resource Management, University of Copenhagen, Copenhagen K, Denmark
$^w$ Department of Ecology and Genetics, Uppsala University, Uppsala, Sweden
$^x$ Department of Biology, San Diego State University, San Diego, CA, USA
$^y$ Institute of Arctic Biology, University of Alaska Fairbanks, Fairbanks, AK, USA
$^z$ U.S. Geological Survey, Wetland and Aquatic Research Center, Lafayette, LA, USA
$^{aa}$ Department of Civil, Environmental & Geodetic Engineering, Ohio State University, Columbus, OH, USA
$^{ab}$ Department of Bioproducts and Biosystems Engineering, University of Minnesota, St Paul, MN, USA
$^{ac}$ Department of Environmental Science, Faculty of Science, Shinshu University, Matsumoto, Japan
$^{ad}$ Department of Earth System Science, Stanford University, Stanford, CA, USA

$^*$ Corresponding author.
E-mail address: qzhu@lbl.gov (Q. Zhu).

https://doi.org/10.1016/j.agrformet.2022.109115
Received 11 February 2022; Received in revised form 22 July 2022; Accepted 4 August 2022
Available online 11 August 2022

A R T I C L E  I N F O

Keywords:
Eddy covariance CH$_4$ emission
Wetlands

A B S T R A C T

Wetland CH$_4$ emissions are among the most uncertain components of the global CH$_4$ budget. The complex nature of wetland CH$_4$ processes makes it challenging to identify causal relationships for improving our understanding and predictability of CH$_4$ emissions. In this study, we used the flux measurements of CH$_4$ from eddy covariance
towers (30 sites from 4 wetlands types: bog, fen, marsh, and wet tundra) to construct a causality-constrained machine learning (ML) framework to explain the regenerative factors and to capture CH$_4$ emissions at sub-seasonal scale. We found that soil temperature is the dominant factor for CH$_4$ emissions in all studied wetland types. Ecosystem respiration (CO$_2$) and gross primary productivity exert controls at bog, fen, and marsh sites with lagged responses of days to weeks. Integrating these asynchronous environmental and biological causal relationships in predictive models significantly improved model performance. More importantly, modeled CH$_4$ emissions differed by up to a factor of 4 under a -1°C warming scenario when causality constraints were considered. These results highlight the significant role of causality in modeling wetland CH$_4$ emissions especially under future warming conditions, while traditional data-driven ML models may reproduce observations for the wrong reasons. Our proposed causality-guided model could benefit predictive modeling, large-scale upscaling, data gap-filling, and surrogate modeling of wetland CH$_4$ emissions within earth system land models.

1. Introduction

Methane (CH$_4$) has been the second most important contributor to post-industrial global warming after carbon dioxide (CO$_2$), with a Global Warming Potential (GWP) of 28-34 times of CO$_2$ over a 100-year time horizon (Bergamaschi et al., 2013; IPCC, 2013). Wetland CH$_4$ emissions are the largest natural global sources, contributing around 20-30% to global emissions (Bouquet et al., 2006; Chen and Prinn, 2006; Saunois et al., 2020). Global warming (Koffi et al., 2020), anthropogenic emissions (Boothroyd et al., 2017), wetland expansion (Zhang et al., 2017), and increasing methanogenic substrate availability (Schuur et al., 2008) are expected to increase CH$_4$ emissions and thereby amplify climate warming (Tao et al., 2020). Freshwater wetlands remain the largest and most uncertain natural CH$_4$ source to the atmosphere (Peltola et al., 2019; Saunois et al., 2020), but with considerable discrepancies among bottom-up biogeochemistry models, top-down atmospheric inversion models, and data-driven machine learning models (Koffi et al., 2020; Peltola et al., 2019; Saunois et al., 2020). Therefore, improvements in the understanding of causality sources and development of robust modeling frameworks for CH$_4$ emissions are required to estimate present-day and future wetland CH$_4$ emissions (Dean et al., 2018).

Wetland CH$_4$ emissions are affected by multiple environmental (e.g., temperature, redox conditions) and biological (e.g., plant photosynthesis, microbial enzyme activity) factors (Delwiche et al., 2021; Knox et al., 2021). Wetland CH$_4$ is produced by methanogens under anaerobic conditions (Mayer and Conrad, 1990), with the production rate controlled by multiple drivers such as temperature, availability of substrate (Bergman et al., 2000; Schaufler et al., 2010; Whalen, 2005), O$_2$, and alternative electron acceptors (Pasut et al., 2021). After production, CH$_4$ can be emitted to the atmosphere through various pathways (e.g., diffusion, ebullition, plant aerenchyma transport) that are affected by temperature, water depth, air pressure, and plant aerenchyma properties (Bastviken, 2009; Knox et al., 2021; Morin et al., 2014; Rey-Sanchez et al., 2018; Villa et al., 2020). CH$_4$ can be oxidized by aerobic bacteria when passing through oxic soil or water during transport (Whalen, 1993) or even via anaerobic pathways (anaerobic oxidation of methane, AOM) (Fan et al., 2021). The impacts of environmental and biological factors on CH$_4$ emissions are often non-linear and operate over a range of time scales (Sturtevant et al., 2016). For example, the response of CH$_4$ production to temperature is observed to be hysteretic (Chang et al., 2021) due to seasonal substrate availability and microbial activity (Chang et al., 2020). The response of CH$_4$ emissions to GPP may be delayed and the relationship between them has been observed to be lagged by hours to days (Hatala et al., 2012a; Rinne et al., 2018), while CH$_4$ emission responses to water table fluctuations can be lagged by days to months (Chen et al., 2021; Goodrich et al., 2015; Sturtevant et al., 2016). The multi-driver dependency, nonlinearity, and time-lagged characteristics make it challenging to understand how CH$_4$ emissions interact with environmental and biological factors and to accurately represent them in predictive models (Kim et al., 2020; Sturtevant et al., 2016; Turner et al., 2021).

In most ecosystem biogeochemical models, wetland CH$_4$ production is represented as a function of net primary production and/or heterotrophic respiration (as a proxy for microbial activity), with both constrained by environmental scalars (Melton et al., 2013; Wania et al., 2013; Xu et al., 2016). For example, temperature sensitivity scalars have been proposed based on observed CH$_4$ emissions (Yvon-Durocher et al., 2014). However, in situ observations reveal high variability and uncertainty in CH$_4$ emissions even with nearly identical environmental conditions (Chadburn et al., 2020; Granberg et al., 1997; Hemes et al., 2018; Koch et al., 2014; Rinne et al., 2018; Villa et al., 2021; Zona et al., 2016), implying much more complex functional relationships between CH$_4$ emissions and environmental and biological factors. A few ecosystem models explicitly represent more of the underlying microbial, plant, and abiotic processes leading to wetland CH$_4$ emissions (e.g., ecosys (Grant et al., 2015; Grant et al., 2017a; Grant et al., 2017b), BAMS4 (Pasut et al., 2021), and JSBACH-methane (Castro-Morales et al., 2018)) and confirm that these nonlinear interactions should be considered to improve model predictions of methane emissions (Chang et al., 2019).

In addition to the ecosystem biogeochemical models, Machine Learning (ML) models are becoming useful tools for capturing complex nonlinear relationships, and have achieved good performance in gap filling CH$_4$ emission data (Hatala et al., 2012a; Hatala et al., 2012b; Irvin et al., 2021; Kim et al., 2020; Knox et al., 2019; Morin et al., 2014) and spatial upsampling (Peltola et al., 2019). However, widely-applied ML frameworks do not accurately represent lagged CH$_4$ emission dependencies (Kim et al., 2020). Including lagged variables as predictors may improve ML model performance, but risks overfitting, especially for multiple-driver dominated ecosystems with limited temporal observations (Kim et al., 2020). Furthermore, commonly used ML models do not consider causality constraints (Pearl, 2019; Reichstein et al., 2019). Such ML models may fit an observational dataset well while not being driven by causal relationships (Pearl, 2019; Runge et al., 2019a). In this study, we explore whether an ML model that represents lagged responses and considers underlying causal relationships can improve process understanding and wetland CH$_4$ emission predictions.

We used CH$_4$ emission measurements at 30 eddy covariance towers covering 4 wetland types (bog, fen, marsh, and wet tundra), to test three hypotheses: (1) It is possible to infer with statistical confidence causal relationships between drivers and CH$_4$ emissions. (2) The environmental drivers significantly affecting methane emissions differ among the wetlands by their type and location. (3) Future model predictions that are well calibrated based on current flux observations, but differ in their assumed causal relationships between drivers and methane emissions, will diverge significantly. To test these hypotheses, we develop an integrated framework that combines causality and ML to improve understanding of causal relationships affecting CH$_4$ emissions and modeling of wetland CH$_4$ emissions across various wetland ecosystems. In this work, a causal relationship exists between predictor (X) and CH$_4$ emissions if, when excluding the confounding effects from other predictors and from the history of CH$_4$ emissions, knowing the predictor (X) could significantly reduce the uncertainty in predicting CH$_4$ emissions (Abdul Razak and Jensen, 2014; Runge et al., 2019a). The overarching goal of this study is to develop, train, and validate a ML model to improve predictive modeling of wetland CH$_4$ emission for diverse wetlands.
2. Methodology

2.1. Study sites and data description

The dataset used in this study is from the FLUXNET-CH4 synthesis activity, which compiles, standardizes, and gap-fills available daily eddy covariance CH4 emission data, via the regional networks of AmeriFlux, EuroFlux, OzFlux, and AsiaFlux (Delwiche et al., 2021; Knox et al., 2019). We focus on four types of natural freshwater wetlands (bog, fen, marsh, and wet tundra), and use 30 wetland sites, each with at least one year of CH4 observations (Fig. 1; Table 1). The wetland classification is based on the site-specific literature (Delwiche et al., 2021). Daily CH4 emissions (FCH4) and 13 potential drivers are considered in our analysis: Air Temperature (T_a), Topsoil Temperature (T_s), Water Table Depth (Dwt), Precipitation (P), Soil Water Content (θ), Relative Humidity (RH), Vapor Pressure Deficit (VPD), Atmospheric Pressure (PA), Wind Speed (WS), and Incoming Shortwave Radiation (SW); and biological factors: Gross Primary Production (GPP), Ecosystem Respiration (RECO), and Net Ecosystem Exchange (NEE) (Runge et al., 2019b). We acknowledge as important driving factors for wetland CH4 activity, which compiles, standardizes, and gap-fills available daily eddy covariance CH4 emissions (Knox et al., 2021; Oertel et al., 2016). Details of data standardization for the FLUXNET-CH4 dataset are presented in Knox et al. (2019). In this study, we used the observed non-gap-filled measurements to maintain the original dynamic patterns and avoid potential biases from the gap-filling algorithms that have their own assumed causal relationships.

2.2. Transfer entropy analysis

We employ a transfer entropy approach with PCMCI framework (Runge et al., 2019b) to identify non-linear directional relationships between environmental and biological factors and FCH4. Transfer entropy is a powerful tool to reveal the causality for non-linear and asynchronous systems (Bouskill et al., 2020; Liu et al., 2019; Schreiber, 2000). The approach quantifies information entropy flow from source variables (e.g., T_s) to the target variable (FCH4) by measuring the information entropy reduction in the target variables when excluding effects from various confounders (Yuan et al., 2022; Li et al., 2022). If transfer entropy is statistically significant, the causal relationship from a source variable to the target variable is confirmed. For each pair of variables of interest, we calculate the transfer entropy (T) from source variable X to a target variable Y considering the confounders of Z (Schreiber, 2000):

\[
T(X\rightarrow Y) = \sum_{n} p(y_t, z_t, x_t^n) \log_2 \frac{p(y_t | z_t, x_t^n)}{p(y_t | z_t)}
\]

where \( i \) is the corresponding time lag of source variable \( X \), \( p \) is the probability density. Compared with the linear and nonlinear correlation based methods (e.g., mutual information in Knox et al. (2021)), transfer entropy can explicitly exclude confounding effects when detecting the causal strength from one variable to \( F_{CH4} \), through removing shared information between confounders (Z) and the target variable (Y).

In theory, all potential confounders should be included when identifying causal relationships. However, in practice, too many confounders will cause high dimensionality and statistical instability issues (Runge et al., 2019a; Yuan et al., 2021). However, wetland \( F_{CH4} \) can be jointly regulated by multiple factors including the history of \( F_{CH4} \). To minimize the interferences from important confounders and to avoid high dimensionality, we adaptively considered three confounders that have the strongest control on the variation of \( F_{CH4} \), through the PCMCI framework (Runge et al., 2019b). PCMCI contains two key steps: (1) PC (named after its inventors Peter and Clark) (Spirtes et al., 2000) and (2) Momentary Conditional Independence (MCI) (Runge et al., 2019b). To infer the causal strength from a source variable to the target variable, we firstly used the transfer entropy method in PC to rank the contribution of all potential confounders (e.g., air temperature, soil water content) with relative lower dimensionality (Spirtes et al., 2000), and used transfer entropy in MCI to calculate the causal strength from a source variable to the target variable by excluding the information entropy from the most important confounders (Runge et al., 2019b). We iteratively conducted the causal inference process for each variable to obtain the causal strength (Fig. S1).

The shuffled surrogate method (Kantz and Schürmann, 1996) was employed to test the statistical significance of transfer entropy. This method randomly shuffles source and target time series to destroy time correlations. Shuffled surrogate transfer entropy was computed 100 times through Monte Carlo simulations. A one-tailed significance test is then applied to determine the 95% confidence of the transfer entropy (Ruddell and Kumar, 2009).

2.3. CH4 emission predictive models

We develop a causality constrained interpretable ML model based on the Long-Short-Term-Memory framework (Guo et al., 2019a; Hochreiter and Schmidhuber, 1997; Li et al., 2020) for prediction (hereafter causal-LSTM). We compared the causal-LSTM model performance, internal functional relationships, and model sensitivity against its baseline LSTM model (described below), to illustrate the benefit of including causality constraints in prediction.

2.3.1. Baseline model

The baseline naïve LSTM model has been widely used in time sequence predictions (Alahi et al., 2016; Li et al., 2020). One of the advanced features of LSTM is the gate mechanism that controls the information flow to be memorized or forgotten, which enables capturing short-term and long-term dependencies underlying data sequences. Here, we use the LSTM model for prediction, given the lagged responses of emissions to environmental and biological factors. The recursive representations of LSTM and prediction can be represented as:

\[
h_t = f(x_t, h_{t-1}, c_{t-1})
\]

\[
\hat{Y}_{t+1} = W^y h_t + b_y
\]

where \( x_t \) (t is time step, 0 ≤ t ≤ T) is the input vector, \( c_t \) is the cell memory state vector, and \( h_t \) represents the hidden state vector with useful
information for predictions. In this study, $\mathbf{x}$ represents the biotic and abiotic drivers across sites; $h_t$ is the hidden state vector at the last time step $T$; $Y_{T+1}$ is the predicted $F_{CH_4}$ at the time step $T+1$; and $W_i$ and $b_i$ are the parameters that need to be learned. The $f$ in Eq. (2) is an integrated function that includes five individual equations:

$$ f_i = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) $$
$$ i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) $$
$$ o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) $$
$$ c_t = f_i \odot c_{t-1} + i_t \odot \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) $$
$$ h_t = o_t \odot \tanh(c_t) $$

where $f_i$, $i_t$, and $o_t$ are gating vectors that control how much information for the cell memory to forget, input/update, and output, respectively; $\sigma$ is the sigmoid activation function; $\odot$ is element-wise product; $W_{xi}$, $W_{hi}$, $W_{xo}$, $W_{ho}$, $W_{xc}$, and $W_{hc}$ are linear transformation matrices that need to be learned; and $b_i$, $b_o$, $b_c$, and $b_h$ are corresponding bias vectors obtained through model training.

For the LSTM, we used the recursive feature elimination (RFE) method (Guyon et al., 2002) to remove spurious predictors. Specifically, we iteratively removed one predictor, used the remaining predictors to train the LSTM, and calculated the correlation coefficient between

<table>
<thead>
<tr>
<th>Wetland Type</th>
<th>Site ID</th>
<th>Site name</th>
<th>IGBP</th>
<th>LAT</th>
<th>LON</th>
<th>Startyear</th>
<th>Endyear</th>
<th>Data DOI</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bog</td>
<td>CA SCB</td>
<td>Scotty Creek Bog</td>
<td>WET</td>
<td>61</td>
<td>-121</td>
<td>2014</td>
<td>2017</td>
<td>DOI: 10.18140/FLX/1669613</td>
<td>(Sonnentag and Hellwig, 2020)</td>
</tr>
<tr>
<td></td>
<td>CA SOC</td>
<td>Scotty Creek Landscape</td>
<td>ENF</td>
<td>61</td>
<td>-121</td>
<td>2013</td>
<td>2016</td>
<td>DOI: 10.18140/FLX/1669628</td>
<td>(Sonnentag and Hellwig, 2020)</td>
</tr>
<tr>
<td></td>
<td>FI-SI2</td>
<td>Siikaneva II</td>
<td>WET</td>
<td>62</td>
<td>24.2</td>
<td>2012</td>
<td>2016</td>
<td>DOI: 10.18140/FLX/1669639</td>
<td>(Vesala et al., 2020)</td>
</tr>
<tr>
<td></td>
<td>JP BBV</td>
<td>Bibai bog</td>
<td>WET</td>
<td>43</td>
<td>141.8</td>
<td>2015</td>
<td>2018</td>
<td>DOI: 10.18140/FLX/1669646</td>
<td>(Ueyama et al., 2020)</td>
</tr>
<tr>
<td></td>
<td>NZ Kopua</td>
<td>WET</td>
<td>-37</td>
<td>175.6</td>
<td>2012</td>
<td>2015</td>
<td>DOI: 10.18140/FLX/1669652</td>
<td>(Campbell and Goodrich, 2020)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DE-Zrk</td>
<td>Zamekow</td>
<td>WET</td>
<td>54</td>
<td>12.89</td>
<td>2013</td>
<td>2018</td>
<td>DOI: 10.18140/FLX/1669636</td>
<td>(Sachs and Wille, 2020)</td>
</tr>
<tr>
<td></td>
<td>FL-Lom Lompolojanka</td>
<td>WET</td>
<td>68</td>
<td>24.21</td>
<td>2006</td>
<td>2010</td>
<td>DOI: 10.18140/FLX/1669638</td>
<td>(Lohila et al., 2020)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FI-Sii</td>
<td>Siikaneva I (FI-Sii)</td>
<td>WET</td>
<td>62</td>
<td>24.19</td>
<td>2013</td>
<td>2018</td>
<td>DOI: 10.18140/FLX/1669640</td>
<td>(Vesala et al., 2020)</td>
</tr>
<tr>
<td></td>
<td>SE-Deg</td>
<td>Degero</td>
<td>WET</td>
<td>64</td>
<td>19.56</td>
<td>2014</td>
<td>2018</td>
<td>DOI: 10.18140/FLX/1669659</td>
<td>(Nilsson and Peichl, 2020)</td>
</tr>
<tr>
<td></td>
<td>SE-St1</td>
<td>Stordalen Mire</td>
<td>WET</td>
<td>65</td>
<td>19.05</td>
<td>2012</td>
<td>2014</td>
<td>DOI: 10.18140/FLX/1669660</td>
<td>(Jansen et al., 2020)</td>
</tr>
<tr>
<td></td>
<td>US-Myb</td>
<td>Mayberry Wetland</td>
<td>WET</td>
<td>30</td>
<td>-122</td>
<td>2011</td>
<td>2018</td>
<td>DOI: 10.18140/FLX/1669685</td>
<td>(Matthes et al., 2020)</td>
</tr>
<tr>
<td></td>
<td>US-Tw1</td>
<td>Twitchell Wetland West Pond</td>
<td>WET</td>
<td>38</td>
<td>-122</td>
<td>2011</td>
<td>2018</td>
<td>DOI: 10.18140/FLX/1669696</td>
<td>(Valach et al., 2020)</td>
</tr>
<tr>
<td></td>
<td>US-Tw4</td>
<td>Twitchell East End Wetland</td>
<td>WET</td>
<td>38</td>
<td>-122</td>
<td>2013</td>
<td>2018</td>
<td>DOI: 10.18140/FLX/1669698</td>
<td>(Eichelmann et al., 2020)</td>
</tr>
<tr>
<td>Wet tundra</td>
<td>RU-Ch2</td>
<td>Chersky Reference</td>
<td>WET</td>
<td>69</td>
<td>161.4</td>
<td>2014</td>
<td>2016</td>
<td>DOI: 10.18140/FLX/1669654</td>
<td>(Göckede, 2020; Göckede et al., 2019)</td>
</tr>
<tr>
<td></td>
<td>US-Beo</td>
<td>Barrow Environmental Observatory (BEO) tower</td>
<td>WET</td>
<td>71</td>
<td>-157</td>
<td>2013</td>
<td>2014</td>
<td>DOI: 10.18140/FLX/1669664</td>
<td>(Zona and Oechel, 2020a)</td>
</tr>
<tr>
<td></td>
<td>US-Bes</td>
<td>Barrow-Bes (Biocomplexity Experiment South tower)</td>
<td>WET</td>
<td>71</td>
<td>-157</td>
<td>2013</td>
<td>2015</td>
<td>DOI: 10.18140/FLX/1669665</td>
<td>(Zona and Oechel, 2020b)</td>
</tr>
<tr>
<td></td>
<td>US-ICs</td>
<td>Innavait Creek Watershed Wet Sedge Tundra</td>
<td>WET</td>
<td>69</td>
<td>-149</td>
<td>2014</td>
<td>2016</td>
<td>DOI: 10.18140/FLX/1669678</td>
<td>(Euskirchen et al., 2020)</td>
</tr>
</tbody>
</table>
observations and predictions after removing the predictor. Then, we removed the weakest predictor which showed the lowest impacts on model performance, and repeated the predictor elimination process until only one predictor was left. Finally, we present LSTM modeling results based on the subset of predictors selected by RFE method that have the highest model performance.

2.3.2. Causality constrained LSTM

Although baseline LSTM is capable of capturing short-term and long-term dependencies in the input time series, it works as a black-box and cannot explicitly select important driving variables and lacks interpretability of its predictions. Also, the dependencies identified within the LSTM model are based on correlations rather than causality (a more informative directional relationship). To this end, the LSTM model can be improved through attention mechanism, an effective weight assignment method, to increase its transparency (Alahi et al., 2016; Guo et al., 2019a; Li et al., 2020; Liang et al., 2018; Qin et al., 2017; Vaswani et al., 2017). The weight mechanism explicitly and dynamically assigns larger weights to more important variables, thereby improving model performance and interpretability (Guo et al., 2019a; Li et al., 2020). However, without the guide or constraint of causality, the correlation-based ML models may represent wrong processes (e.g., mis-capture dominant causal drivers) (Moraffah et al., 2020; Pearl, 2019; Runge et al., 2019b), making the model unreliable, especially for predictions using multiple drivers with similar seasonal trends (confounding) information under climate change (Runge et al., 2019a). In addition, we further imposed additional constraints using causal relationships from input variables to the target variable and led to the causal-LSTM model. The causal-LSTM model first calculated the causal relationship using transfer entropy. And then through optimization, it reduced the model biases on both prediction error and structure difference between model captured variable dependency and observation-based causal strength. Below, we introduce the weight assignment mechanism (attention mechanism) in the LSTM approach and describe details of how we incorporate causality constraints in the model.

Similar to the baseline LSTM, the ith driving variable at time step \( t \) can be iteratively transformed to a hidden state vector \( h^i_t \) through the gate mechanism (Guo et al., 2019a; Hochreiter and Schmidhuber, 1997; Li et al., 2020; Qin et al., 2017). To represent the importance of the ith variable at time step \( t \), a weight \( w^i_t \) or \( w^i \) is dynamically calculated through Eqs. (4) and (5), and assigned to \( h^i_t \). Then, the weighted summation \( h^i_{\text{sum}} \) of \( h^i \) across time steps is obtained to represent the summarized information for the ith driving variable:

\[
w^i_t = \tanh(W^i_t h^i_t)
\]

\[
w^i = \frac{e^{w^i}}{\sum_{l=1}^{n} e^{w^l}}
\]

\[
h^i_{\text{sum}} = \sum_{t=1}^{T} w^i_t h^i_t
\]

Where \( W^i_t \) is a parameter matrix that needs to be learned, and \( \tanh \) is the hyperbolic tangent function. \( T \) is the total number of time steps.

To further represent the relative importance of the ith driving variable compared to other driving variables, a weight, \( a_i \), is obtained and normalized as \( a_i' \):

\[
a_i = \tanh(W^i_t h^i_{\text{sum}})
\]

\[
a_i' = \frac{e^{a_i}}{\sum_{j=1}^{n} e^{a_j}}
\]

where \( W^i_t \) is a learnable parameter matrix.

Finally, using the weighted sum of all driving variables, the model generates the prediction \( \hat{Y}_{T:1} \):

\[
a_i = W^i_t [h^i_{\text{sum}} h^i_T] + b_i
\]

\[
\hat{Y}_{T:1} = \sum_{i=1}^{n} a_i' a_i
\]

where the linear function with weight \( W^i_t \) and bias \( b_i \), along with weight \( a_i' \) produce the final prediction.

To make the internal structure of the model more consistent with underlying physical processes, we use transfer entropy inferred causal relationships to constrain the variable importance (variable weight) in the predictive model. A larger transfer entropy from a driver (e.g., soil temperature) to \( F_{CH_4} \) implies variations of the driver can cause larger variations in \( CH_4 \) emissions, compared to other drivers (Ruddell and Kumar, 2009). Similarly, a larger variable weight (\( a_i' \)) indicates that the \( i \)th variable plays more important roles in modeling the target variable (Guo et al., 2019a; Li et al., 2020; Liang et al., 2018; Qin et al., 2017). To guide the model to learn dependencies between causally dominant drivers and \( F_{CH_4} \), we measure the difference between transfer entropy inferred feature importance vector \( a_{TE} \) and that of the model captured feature importance vector \( a_k \) for each sample \( k \), and integrate the difference along with modeled errors into the final loss function (Eq. (11)). In the vector \( a_{TE} \), \( a_{TE} \) represents the transfer entropy from the ith driving variable to \( F_{CH_4} \). In \( a_k \), \( a_k \) represents the ith variable weight, \( a_k' \), for a sample \( k \). Each vector is divided by its summation to obtain a probability distribution ranging from 0 to 1, and KLDivergence (Kullback and Leibler, 1951) (the second item in the loss function, Eq. (11)) is used to measure the difference distribution between the two vectors:

\[
\text{Loss} = \frac{1}{N} \sum_{k=1}^{N} \left( \tilde{Y}_k - Y_k \right)^2 + \lambda \sum_{k=1}^{N} \sum_{i=1}^{n} a_{TE} \log \frac{a_{TE}}{a_k'}
\]

where \( \lambda \) is a structural punishment parameter, and a larger \( \lambda \) means that the model puts more emphasis on structural similarity instead of errors. In Eq. (11), the first right hand side term is the errors between observations and predictions, while the second term is the structural similarity between causality inferred feature importance and importance the model captured. \( N \) is the number of predicted data samples, and \( n \) is the number of variables. The baseline LSTM uses only the first term on the right-hand side for the loss function, while the causal-LSTM has additional constraint from causal relationships via the second term (Eq. (11)).

The model parameters are learned via a back-propagation algorithm (Rumelhart et al., 1986) by minimizing the integrated loss (Eq. (11)) with a variational dropout to avoid overfitting (Gal and Ghahramani, 2016). We used the intra-site validation scheme to test model performance on capturing intra-site temporal variations of \( F_{CH_4} \). Specifically, in each experiment, for each site, we randomly sampled 80% of data as a training dataset, remained 10% as a validation dataset (used to avoid overfitting during training (Prochelt, 1998)), and retained the remaining 10% as a test dataset (a holdout dataset used to unbiasedly evaluate the final model). We repeated each experiment 20 times to reduce model bias due to random data selection. We compared the model performance with different \( \lambda \) values (Fig. S2), and selected the best one (\( \lambda = 0.005 \)) that has the lowest prediction errors. To evaluate the lag effects for model improvement, we varied the lengths (one-week vs. one-month) of time series input used in the models. In addition, we also used the leave-one-site-out scheme (here referred as inter-site validation) (Jung et al., 2011) to test model performance on spatial extrapolation of \( F_{CH_4} \) on each tested site. Other detailed experimental settings of each model are listed in Table S2.

3. Results
3.1. Causal relationships derived from transfer entropy

Transfer entropy analysis revealed that daily \( F_{CH_4} \) was most strongly
driven by soil temperature ($T_s$) in the four analyzed wetland ecosystem types (bog, fen, marsh, and wet tundra; Fig. 2a), with a range of different time lags. The statistics of dominant drivers at each individual site also showed that $T_s$ dominated in most sites (Fig. S3). Furthermore, the strength of the $T_s \rightarrow \text{FCH}_4$ relationship declined with increasing mean air temperature (slope = -0.0014, R value = -0.63, $p$ value < 0.05) (Fig. 2b). This inverse relationship suggested that $\text{CH}_4$ emissions in colder regions were more sensitive to temperature than in warmer areas. The control from air temperature ($T_a$) was weaker than that from $T_s$ and was prominent only at fen and marsh wetlands (Fig. 2a).

Two biological factors, Ecosystem Respiration (RECO) and plant Gross Primary Production (GPP), also exerted strong controls on daily $\text{FCH}_4$ in bog, fen, and marsh wetlands. These strong relationships between $\text{FCH}_4$ and vegetation carbon turnover are consistent with the findings of many previous studies (Hatala et al., 2012a; Mitra et al., 2020; Rinne et al., 2018). Plant GPP stimulates $\text{CH}_4$ production indirectly by providing carbon input, mainly via root exudates fueling microbial activity, which produces substrates (such as acetate and $\text{CO}_2$) for acetotrophic and hydrogenotrophic methanogenesis (Bastviken, 2009; Mitra et al., 2020; Strom et al., 2012; Whiting and Chanton, 1993). Additionally, GPP can be seen as a proxy of plant-mediated $\text{CH}_4$ transport via aerenchyma tissue (Bastviken, 2009; King et al., 1998; Turetsky et al., 2014). Previous studies argued that the relationship between GPP and $\text{FCH}_4$ may be due to covariation with confounding drivers (e.g., soil temperature) (Chang et al., 2021; Knox et al., 2019). In this study, we confirmed the existence of a strong coupling from GPP and RECO with $\text{FCH}_4$ by removing confounding effects when identifying the causal relationships across multiple wetland types.

Compared with temperature and biological factors, the controls from other variables on $\text{FCH}_4$ were much weaker (Figs. 2a, S3) and less consistent across wetland types. For example, VPD controlled $\text{FCH}_4$ more at bog and fen ecosystems, while PA showed weak causal relationships with $\text{FCH}_4$ across all sites. For water-related factors, significant controls on $\text{FCH}_4$ existed only at a few sites, which may be attributed to limited observations of water table depth ($D_{WT}$, 16 sites) and soil water content (0, 9 sites), and limited variations of soil wetness across studied sites (more details are discussed in Section 4.1).

3.2. $\text{FCH}_4$ predictions with causal constraints

Because causal relationships varied across wetland ecosystems, we trained independent ML models for each wetland type (bog, fen, marsh, and wet tundra). Two types models were considered: Long Short-Term Memory (LSTM) and causality-constrained interpretable LSTM (causal-LSTM). We found that Causal-LSTM performed consistently better than LSTM for all four wetland types with higher Pearson correlation coefficient (R) and lower relative MAE (mean absolute error) when inputting four weeks of historical drivers (Table S3 and S4). For example, R values in LSTM ranged from 0.861 to 0.908 and relative MAE ranged from 0.271 to 0.433, while R in causal-LSTM ranged from 0.904 to 0.921 and relative MAE ranged from 0.217 to 0.368 (Fig. 3, Tables S3 and S4). Consistently, with one week of inputs, the causal-LSTM also showed significantly higher R and lower relative MAE compared with LSTM in all wetland types ($p < 0.05$, Tables S3 and S6). We also compared the causal-LSTM approach with four other widely used ML algorithms (random forest, decision tree, artificial neural networks, and support vector machine), and found that causal-LSTM had the highest prediction accuracy (Fig. S4), with R value of 0.94 between observations and predictions of causal-LSTM across all sites (Fig. S5).

For model evaluation with the inter-site validation scheme, causal-LSTM also performed reasonably well with R value of 0.75 between observations and predictions (Fig. S6) and lower biases than that of LSTM (Table S7). However, the inter-site validation performance of causal-LSTM dropped, compared with the intra-site validation scheme.
especially for the marsh (the mean R value dropped to 0.81 in bog, 0.81 in fen, 0.86 in wet tundra, and 0.69 in marsh), which may be due to the strong spatial heterogeneity of $F_{CH_4}$ magnitude (e.g., mean $CH_4$ emission ranged from 2.706 nmol m$^{-2}$ s$^{-1}$ to 165.472 nmol m$^{-2}$ s$^{-1}$ across different sites) and environmental conditions (e.g., annual precipitation in marsh varied from ~200 to ~1400 mm/year). Overall, we conclude that the causal-LSTM provides the most effective approach to model wetland $F_{CH_4}$.

The results showed that model performance tended to be improved as the length of input time series increased from one to four weeks for both causal-LSTM (Fig. 3a and b, yellow vs. red bars) and LSTM models (Fig. 3a and b, green vs. purple bars). For R, the performance of both models at bog, and marsh was significantly (p < 0.05; Tables S8 and S9) improved as the input data length increased. Similarly, in terms of relative MAE, the causal-LSTM model showed significantly lower biases (p < 0.05; Table S10) in bog, marsh, and wet tundra ecosystems, and LSTM showed significant lower biases in bog and wet tundra (p < 0.05, Table S11). Overall, longer histories of drivers (i.e., memories) can provide additional information for predictions, especially in bog, marsh, and wet tundra.

4. Discussions

4.1. Soil temperature versus soil water control on $F_{CH_4}$

Wetland $CH_4$ emissions are regulated by multiple biotic (i.e., production, oxidation) and abiotic (i.e., advection, diffusion) processes, with each posting different dependencies on environmental factors. Therefore, the emergent relationships between wetland methane emissions and the corresponding environmental factors are expected to be complex and diverse across different wetland ecosystem types and across sites with different climate conditions (Turetsky, 2014). Among those environmental variables, previous studies have identified temperature and soil water content as major abiotic drivers for wetland $CH_4$ emissions (Knox et al., 2021; Song et al., 2011; Strachan et al., 2015) because soil water saturation and warm soil conditions are two prerequisites for anaerobic production of wetland $CH_4$ (Riley et al., 2011).

Here, we found strong soil temperature control on $CH_4$ emissions across bog, fen, marsh, and wet tundra ecosystems. The stronger causal relationship of $T_s \rightarrow F_{CH_4}$ compared to $T_a \rightarrow F_{CH_4}$ is consistent with the hypothesis that air temperature may decouple from soil temperature in colder ecosystems (e.g., wet tundra) due to snow insulation of the ground (Kim et al., 2007). Similar strong correlations between $T_s$ and wetland $F_{CH_4}$ have been reported in numerous site-level studies (Granberg et al., 1997; Knox et al., 2021; Morin, 2019).

We also found relatively weak control from soil water related variables (Fig. 2), admit low confidence because of limited data. For example, soil water content had weak control in fen ecosystems, and water table depth had moderate control in bog and marsh ecosystems, but not in fen or wet tundra ecosystems. The lack of sensitivity may partly be due to the data quality of water related variables ($D_{wt}$ is available in ~50% of our studied 30 sites, and $\theta$ is available in only ~30% of the 30 sites (Table S1)). Another potential reason is the fact that the sites used in this study all experienced relatively low variation of $D_{wt}$ (mean standard deviation is 10.6 cm). Strong seasonal fluctuations of soil water are more expected at rice paddy or tropical swamp ecosystems (Juutilainen et al., 2005; Mezbahuddin et al., 2014), which are not included in this study. For example, water table depth could vary ~80 cm at a managed rice paddy site in northern California and plays an important role in driving $CH_4$ emissions during both growing season and fallow periods (Knox et al., 2016). Although not frequently occurred, extreme droughts may result in significantly different water table at fen and bog sites that will reduce the methane emission (Brown et al., 2014; Rinne et al., 2020). However, ML model was trained with majority of the data to capture non-extreme conditions. In addition, we note that several studies reported weak dependencies between $D_{wt}$ and $F_{CH_4}$ (Jackowicz-Korczyńska et al., 2010; Rinne et al., 2007; Rinne et al., 2018). Given the limitations in sites and water-related data availability, our results highlight the need for more eddy covariance and ancillary measurements in bog, fen, marsh, and wet tundra ecosystems, particularly measurements under long-term drying and rewetting conditions, or experiencing natural flooding and water table fluctuation. These observations will facilitate a more complete picture of how various factors affect wetland $CH_4$ emissions within these wetland ecosystems.

4.2. Causal relationships inform model evaluation and development

In addition to commonly used model evaluation metrics (e.g., MAE and R), causal inference provided additional metrics to evaluate and benchmark models in terms of internal causal structures. Causal relationships may also help select process-based models with model causal structures similar to those in observations. In this analysis, we found that methane ML models can achieve comparable performance even though they have diverse causal relationships. We visualized variable importance within LSTM and causal-LSTM models and validated the modeled relationships against observed causal relationships identified by transfer entropy analysis (Fig. 4). The feature importance of causal-LSTM and LSTM were calculated according to attention weight statistics (Guo et al., 2019b; Li et al., 2020) and the feature importance derived from RFE (Guyon et al., 2002; Meyer et al., 2019) of 20 repeated experimental results, respectively, and were both normalized to 0~1. We found that LSTM mainly used dependencies from wind, atmospheric pressure, soil and air temperature, and total ecosystem respiration to estimate $F_{CH_4}$, which were different from those inferred from observations and causal-LSTM (Fig. 4). The feature importance in the causal-LSTM model is much more consistent with observations, confirming the effectiveness of the causality constraints.

The inferred causal relationships from biological and environmental variables on $CH_4$ emissions vary across different wetland types and time windows. Our results show that soil temperature dominantly controls $F_{CH_4}$ in wet tundra, while biotic variables along with soil temperature co-dominate $F_{CH_4}$ in fens, bogs, and marshes. The different controls imply that different ecosystems need to be considered separately in machine learning model development (Turetsky et al., 2014). Also for each wetland ecosystem, the responses of $F_{CH_4}$ rely on processes with short time lags (e.g., $CH_4$ transport, microbial activity) and long time lags (e.g., fine-root turnover). Integrating both short and long causal relationships may also improve model performance.
4.3. Implications of considering causal relationships in CH$_4$ emission projections

Our results, in line with previous studies, suggest that data-driven ML models may accurately reproduce observations with the wrong reasons (Pearl, 2019; Reichstein et al., 2019; Runge et al., 2019a). The different causal relationships built within predictive models are critically important for climate change projections, since the responses of CH$_4$ emissions to climate change strongly depend on the strength of the underlying causal relationships. Thus we hypothesized that although both LSTM and causal-LSTM performed reasonably well under present-day conditions, their predictions under warming climate could differ due to their differences in internal functional relationships, or altered combinations of forcing mechanisms. To test this hypothesis, we conducted a theoretical soil warming experiment (−1 °C) at all sites through modeling. We acknowledged that more complex changes can occur in a real soil warming experiment (e.g., soil drying caused by warming) (Pries et al., 2017). However, this simple soil warming experiment isolates impacts from other environmental or biological variables and focuses only on the temperature effect.

For each wetland type, we calculated the mean change in $F_{CH4}$ due to soil warming across all site years. We defined response ratio to warming by percentage change of $F_{CH4}$ under warmed and controlled conditions. Large differences between the LSTM and causal-LSTM existed in response to warming, especially for bogs (4.9% vs 21.8%) and fens (2.7% vs 10.1%) (Fig. 5). The differences in causal-LSTM predictions are significantly larger than those of LSTM for bog, fen, marsh and wet tundra sites (p < 0.05; Table S12). Overall, the LSTM model estimated lower methane emission in response to warming than causal-LSTM model, primarily due to the less important role of soil temperature in its internal model functions (Fig. 4a). Therefore, this work highlights the importance of considering causal relationships in modeling CH$_4$ emissions under a changing climate. We advocate the use of these types of causal relationship constraints for other ecosystem variables calculated through machine learning approaches (e.g., FLUXNET-MTE GPP (Jung et al., 2011)). In addition, causality constrained ML models could serve as surrogate modules for efficient parameterization and high accuracy prediction, especially for processes that lack theoretical understanding and mathematical model structures.

5. Conclusions

Based on in situ eddy covariance measurements of daily CH$_4$ emissions ($F_{CH4}$) at 30 eddy covariance sites in bog, fen, marsh, and wet tundra wetlands, we found consistent causal regulations from soil temperature on $F_{CH4}$, using a transfer entropy approach. We also confirmed important causal relationships with ecosystem respiration (RECO) and gross primary production at bog, fen, and marsh wetlands. The transfer entropy approach explicitly excludes confounding variables and therefore reduces the possibility that the observed causal relationship between $F_{CH4}$ and RECO or GPP was due to covariation with other environmental drivers, such as temperature (Chu et al., 2014; Knox et al., 2019). We then developed a predictive model that integrated the transfer entropy inferred causal relationships for $F_{CH4}$ simulations. The causality constrained model outperformed other baseline ML models in terms of accuracy (relative MAE and R); more importantly, we demonstrated that including underlying causal relationships in predicting $F_{CH4}$ under a 1 °C soil warming could differ by up to a factor of 4, compared with traditional ML models. Our results highlighted that those causal relationships can be used to benchmark, evaluate, and improve wetland methane emission models. Our proposed causality constrained model could benefit large-scale upscaling, data gap-filling, and surrogate modeling of wetland CH$_4$ emissions within earth system land models.

Fig. 5. $F_{CH4}$ response ratio to an imposed +1 °C soil warming of LSTM (green) and causal-LSTM (yellow) models. The boxes represent 25th to 75th percentiles, and the whiskers represent 5th to 95th percentiles of R or MAE for each wetland type.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research was primarily supported by NASA Carbon Monitoring System grant (#NNH20ZDA001N) and the Reducing Uncertainties in Biogeochemical Interactions through Synthesis and Computation (RUBISCO) Scientific Focus Area Project, the latter is sponsored by the Earth and Environmental Systems Modeling (EESM) Program under the Office of Biological and Environmental Research of the US Department of Energy Office of Science. Min Chen acknowledges the support from NASA Terrestrial Ecology Arctic Boreal Vulnerability Experiment Phase 2 program (#80NSSC21K1702). Margaret Torn acknowledges the support from The Next-Generation Ecosystem Experiments (NGEE Arctic) project under the Office of Biological and Environmental Research in the DOE Office of Science. We acknowledge support from John Wesley Powell Center for Analysis and Synthesis of the U.S. Geological Survey (“Wetland FLUXNET Synthesis for Methane” working group). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.agrformet.2022.109115.

References
