Supporting Information for

Widespread Drainage Increases Carbon Emissions in Southeast Asian Peatlands

Nathan C Dadap[1], Alison M Hoyt[2,3], Alexander R Cobb[4], Doruk Oner[5], Mateusz Kozinski[5], Pascal V Fua[5], Krishna Rao[1], Charles F Harvey[6], and Alexandra G Konings[1]

[1] Department of Earth System Science, Stanford University, Stanford, CA
[2] Max Planck Institute for Biogeochemistry, Jena, Germany
[3] Lawrence Berkeley National Laboratory, Berkeley, CA
[4] Center for Environmental Sensing and Modeling, Singapore-MIT Alliance for Research and Technology, Singapore, Singapore
[5] École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
[6] Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA

Contents of this file

Introduction

We provide additional details regarding the carbon flux calculations and subsidence model analyses performed. Additionally, we present supporting figures related to model assessment, drainage density category examples, subsidence-drainage spatial patterns, subsidence-drainage behavior in the ex-Mega Rice Project area, and the relationship between the drainage canal metrics used within the subsidence model. Finally, we provide a data table summarizing the results of the subsidence model.

Text S1.

      Carbon flux estimation. Extension of the subsidence rate to carbon flux estimate was determined using the approach outlined by Hoyt et al. (2020). Equation 1 assumes that carbon loss is equivalent to the volume of peat subsidence multiplied by mean peat carbon density. Furthermore, this approach has been found to yield similar results to an alternate flux estimation method based on near surface peat characteristics.

$$CO_2 \text{ emissions rate} = \text{Subsidence rate} \times \text{Dry Bulk Density} \times \text{Carbon Density} \quad (1)$$

      Peat carbon density range (53-57%) and dry bulk density range (0.07-0.09 g/cm^3) were based on a review of in situ measurements by Couwenberg & Hooijer (2013). Uncertainty estimates were calculated by aggregating uncertainties around the mean in the dry bulk density, carbon density, and subsidence estimates (Equation 2). This approach, as outlined by Hoyt et al. (2020), accounts for the fact that errors in dry bulk density and carbon concentration are likely correlated with each other but not with subsidence.

$$\partial C = \text{mean } C \sqrt{\left(\frac{\partial \text{subsidence}}{\text{mean subsidence}}\right)^2 + \left(\frac{\partial \text{DBD}}{\text{mean DBD}} + \frac{\partial \text{CD}}{\text{mean CD}}\right)^2} \quad (2)$$

      Here, ∂C is the calculated uncertainty in the mean carbon emissions rate, C represents carbon emissions, and $\partial subsidence$, ∂DBD, and ∂CD represent the uncertainties in subsidence, dry bulk density, and carbon density respectively. Standard error of the mean was used as the uncertainty in subsidence, while half the range of the measurements in peat carbon density and dry bulk density was used as the uncertainty for those quantities. Uncertainties in the difference between any two uncertain quantities were calculated using the root sum of squares method.

Text S2.

    Subsidence modeling. To assess the value of drainage metrics in predicting subsidence, we created a model of subsidence. Here, we chose to use a random forest model because it captures the non-linear relationship between the variables, minimizes the impact of correlated predictor variables, and easily assesses predictor importance. Here, two models were compared: one that included land use (from 2007 and 1990), distance from the peatland boundary, active fire counts using VIIRS Active Fire product from 2015-2019, and drainage canal metrics as predictor variables, and another that omitted the drainage metrics of canal length and distance to canal. Following the method outlined by Hoyt et al. (2020), distance from peatland boundary was also included in the model as a proxy for both peat thickness and distance from the nearest river; distance from peatland boundary was calculated for all points using the distance transform function in Scipy python bindings applied to a map of SEA peatlands by Miettinen et al. (2016). Approximate time since forest conversion was also calculated from the land use datasets and included as a predictor; a given location was categorized as changing from forest to non-forest in categories of pre-1990, 1990-2007, 2007-2015, or not converted as of 2015. Feature importance was calculated using the Gini importance metric. These values were then summed for related features – for example, the importance of land use information overall was taken as the sum of the 2007 land use, 1990 land use, and time since forest conversion predictors.

    Random forest hyperparameters were optimized separately for the two models using cross-validation with an 80-20 train-test split. The fitting procedure sought to maximize the coefficient of determination R^2 in the training data. To ensure that the model was not overfitting, we iterated over a randomized distribution of hyperparameter sets as to minimize the train-test difference. Training data were removed for areas where the land use changed between 2007 and 2015, to eliminate potential noise in the model due to the temporal differences between the InSAR and canal datasets. This model was also applied only to areas outside of the ex-Mega Rice Project area, which was found to diverge from the subsidence-drainage density relationship (Figure S5 and Discussion). Uncertainty in the model results were estimated by 5-fold cross validation (with 80-20 train-test split), measuring the standard deviation of the cross-validated training set R^2, test set R^2 values, and variable importance values.

a Classification **b** Label

1 km 1 km

Figure S1. Validation area results. a) Canals classified by model are shown next to **b)** "true" labels in the model validation area.

a

Labeled Canals   Classified Canals

b
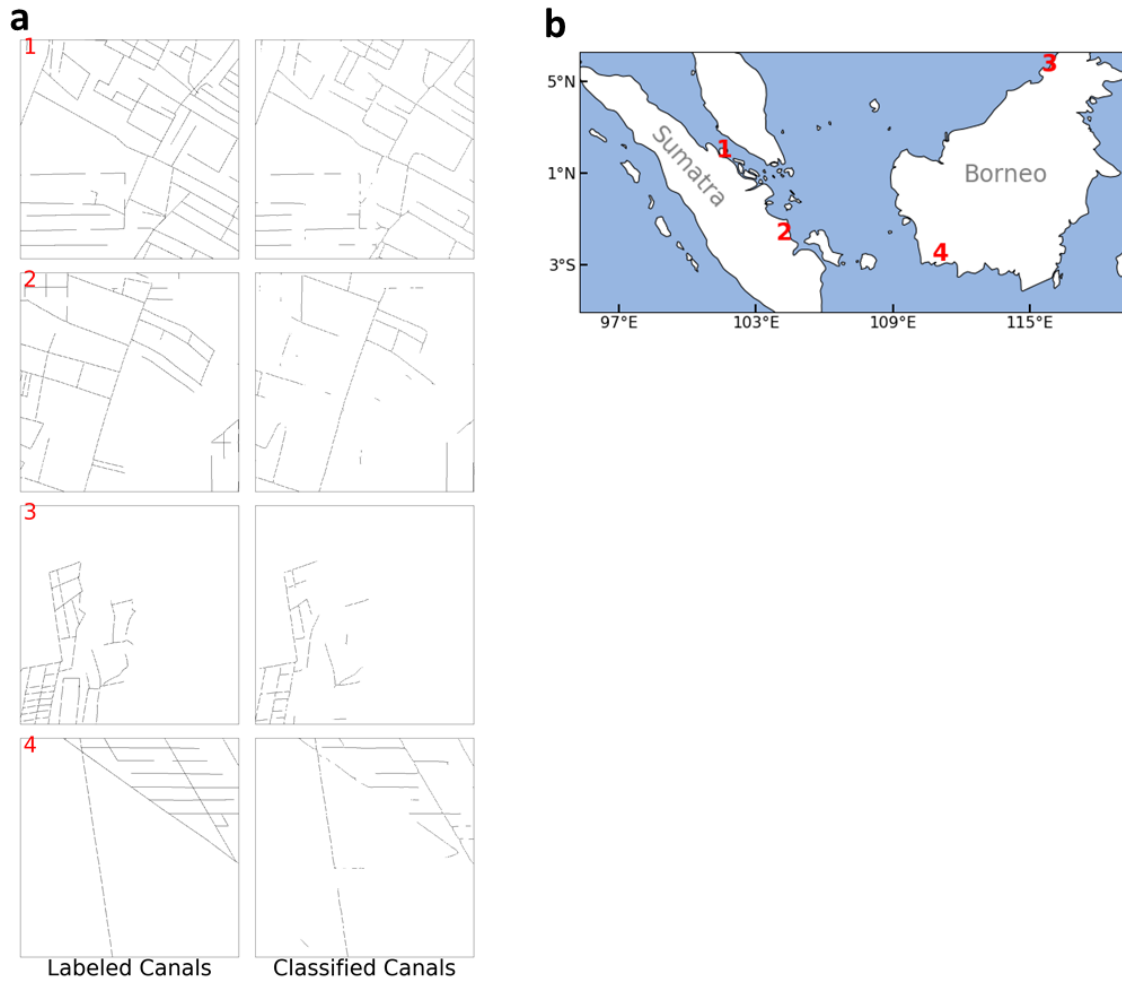
Figure S2. Test area results. a) Canals classified by model are shown next to "true" labels in the model test areas at four sample locations. Each location is a 16 km x 16 km square. Test locations were labeled by a different person than the person who labeled training and validation areas. **b)** Map showing test data locations. One location was randomly chosen from each geographic quadrant of the study area.
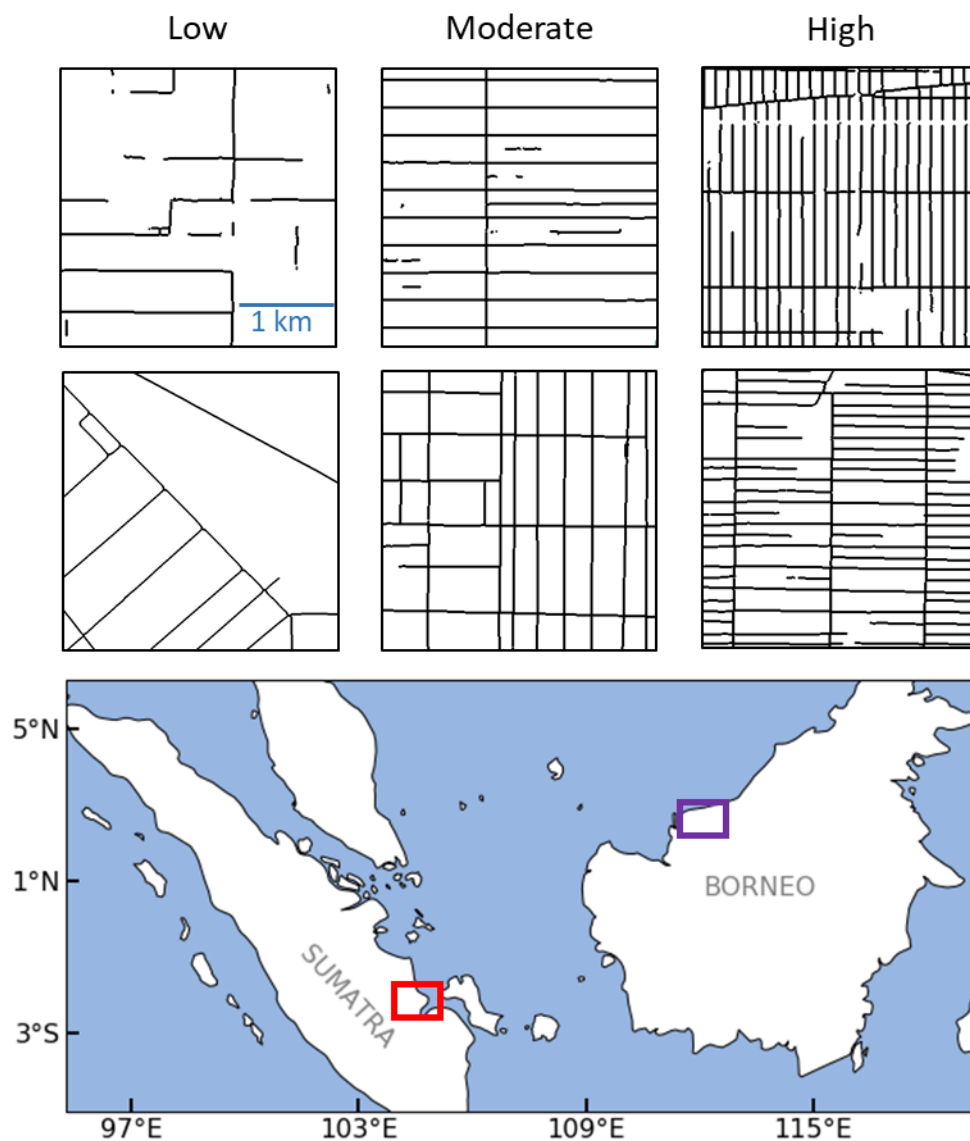
Figure S3. Examples of classified drainage density categories. Each example is a 3 km x 3 km square. Examples in top row are from the purple box in Sarawak, Malaysia. Examples in the second row are from the red box in South Sumatra, Indonesia.
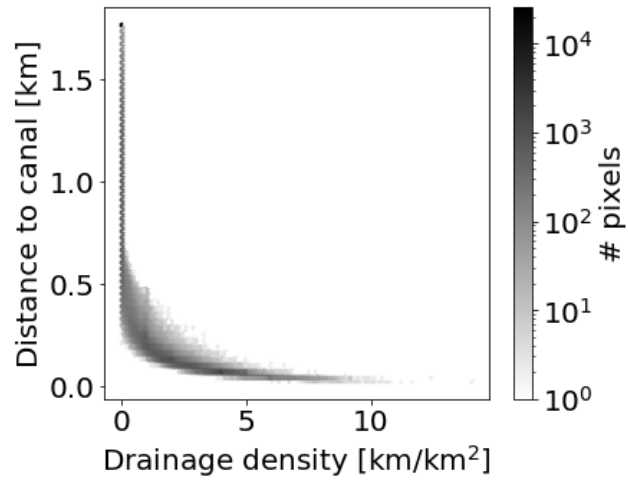
Figure S4. Relationship between drainage density and distance to canals. Both quantities were used as predictor variables in the random forest subsidence model as metrics of drainage.
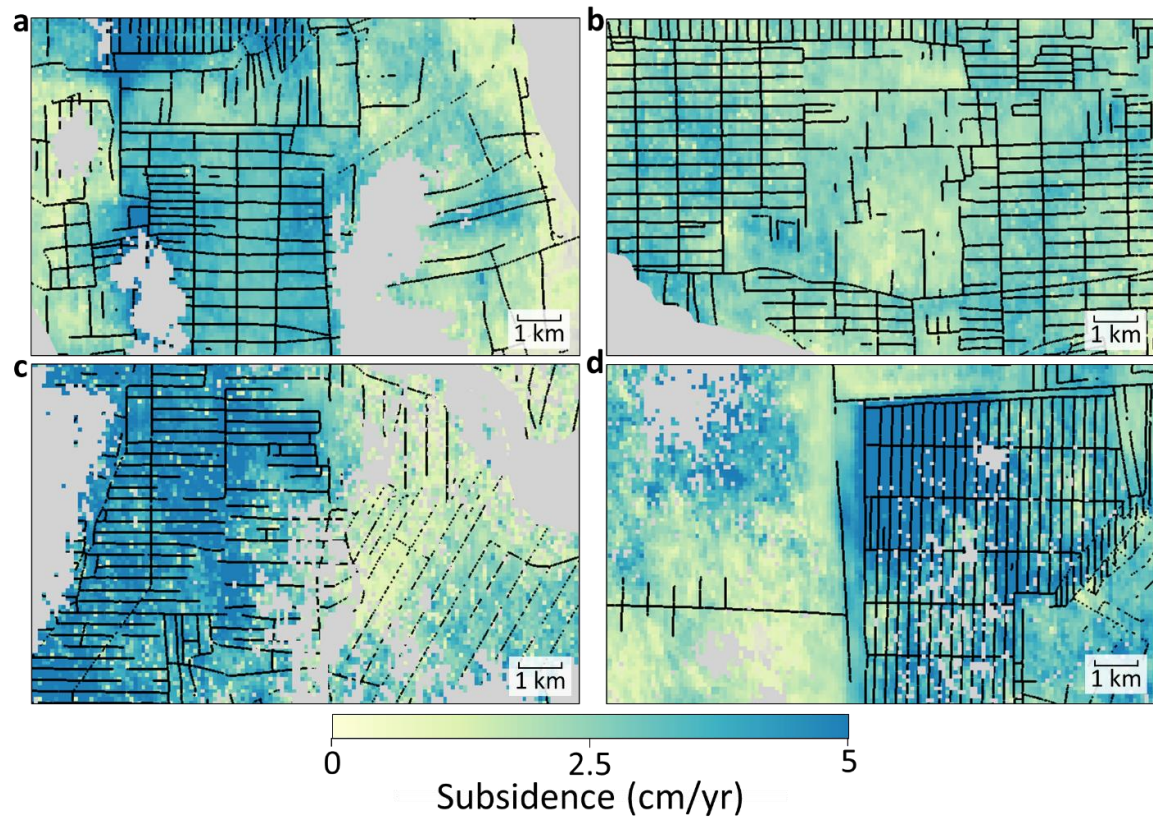
Figure S5. Example maps of subsidence rates and drainage canals. Colors denote subsidence rate while black lines denote drainage canals. Gray denotes either non-peatland areas or areas where subsidence data is unavailable. **a)** and **b)** Labuhan Batu Regency, North Sumatra, Indonesia, **c)** Indragiri Hilir Regency, Riau, Indonesia, **d)** Kapuas Regency, Central Kalimantan, Indonesia.
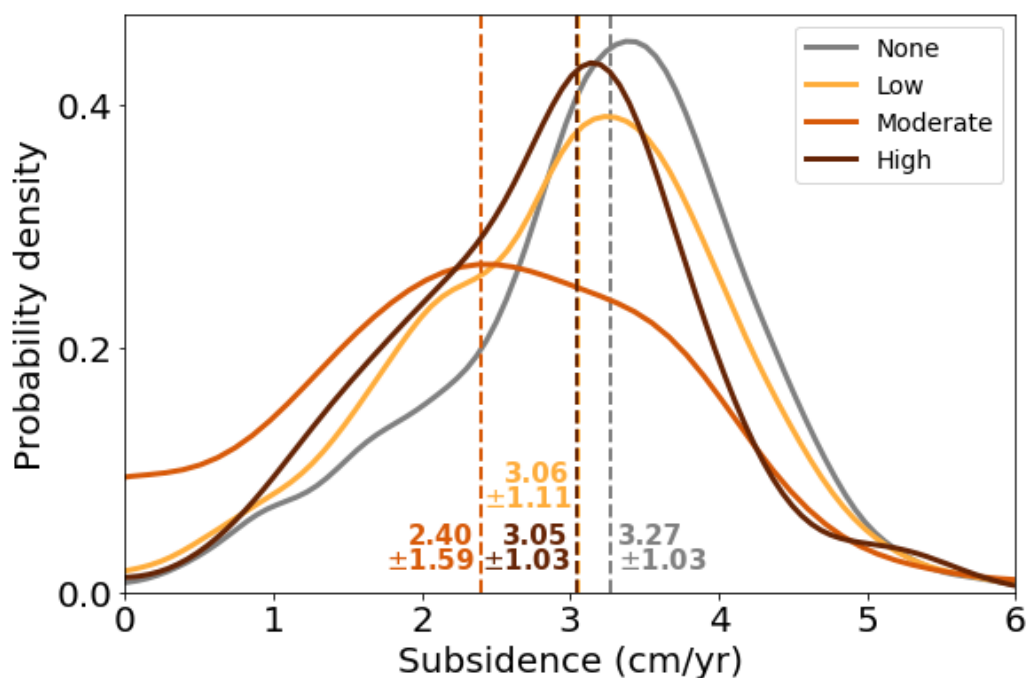
Figure S6. Distribution of subsidence grouped by drainage density in ex-Mega Rice Project area. Dashed vertical lines show medians for each distribution. Median +/- the standard deviation is shown for each distribution next to median line. Positive subsidence denotes downward ground surface displacement and CO_2 release to the atmosphere. Subsidence was measured by Hoyt et al. (2020) using ALOS-1 PALSAR interferometric data from 2007-2011.

	With Drainage Metrics	No Drainage Metrics
R^2 training set	0.47 ± 3e-3	0.37 ± 4e-3
R^2 test set	0.39 ± 2e-2	0.30 ± 2e-2
Feature Importance	• Land use 2007 (0.32 ± 3e-3) • Distance from peat edge (0.22 ± 6e-3) • Mean distance to canal (0.20 ± 3e-3) • Land use 1990 (0.11 ± 4e-3) • Canal density (0.06 ± 5e-3) • Active fire counts 2015-2019 (0.06 ± 8e-3) • Time since forest conversion (0.03 ± 7e-3)	• Land use 2007 (0.40 ± 7e-3) • Distance from peat edge (0.36 ± 1e-2) • Land use 1990 (0.12 ± 2e-3) • Active fire counts 2015-2019 (0.10 ± 1e-2) • Time since forest conversion (0.03 ± 7e-3)
Feature importance categorical sum	• Land use: 0.46 ± 9e-3 • Canals: 0.26 ± 5e-3 • Distance from peat edge : 0.22 ± 6e-3 • Fire: 0.06 ± 8e-3	• Land use: 0.55 ± 8e-3 • Distance from peat edge : 0.36 ± 1e-2 • Fire: 0.10 ± 1e-2

Table S1. Subsidence random forest model results. Uncertainty in the model results were estimated by calculating the standard deviation of the average cross-validated training set R^2, test set R^2 values, and variable importance values. Colors in the Feature Importance row denote related features, as categorized in the last row. The last row of the table denotes the sum of the feature importance values across related predictor features. Uncertainties were propagated to the categorical sums via the root sum of squares method.