

DATA-INTENSIVE ECOLOGICAL MODELING AND INFORMATICS TOOLS

By Xin Huang

A Dissertation

Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy
in Informatics and Computing

Northern Arizona University

May 2023

Approved:

Andrew D. Richardson, Ph.D., Co-chair

Michael Gowanlock, Ph.D., Co-chair

Christopher E. Doughty, Ph.D.

Yiqi Luo, Ph.D.

ABSTRACT

DATA-INTENSIVE ECOLOGICAL MODELING AND INFORMATICS TOOLS

XIN HUANG

The advent of enormous observations and measurements offers tremendous potential to improve our predictive understanding of ecological processes under global climate change. However, despite the abundance of data, the ability of ecological prediction is limited. It remains challenging to translate the ultimate value of big data into actionable insights. The overall research question in my dissertation is how to develop informatics frameworks that can better integrate big data with process-based models to improve predictive understanding of ecological processes and ecological forecast of future ecosystem states. In Chapter 1, I introduce the motivation, data diversity, and current research state on ecological prediction. In Chapter 2, I present a model-independent data assimilation (MIDA) module. The easy-to-use capability of MIDA enables ecologists to conduct data model fusion in various applications. In Chapter 3, I apply MIDA in an 8-yr tundra warming experiment to advance our mechanistic understanding of how carbon cycle in the tundra ecosystem would change with global warming. I found that without considering elevated water table depth, the Terrestrial ECOsystem (TECO) model will underestimate carbon emissions. In Chapter 4, I develop an assessment method to evaluate the effects of forcing corrections on the improvement of forecast accuracy. Compared to the conventional method, this new assessment method can identify the decrease in forecast accuracy for pool-based variables (e.g., Leaf). I also found that only response variables with memory (e.g., carbon pools) can benefit from periodic forcing correction. In Chapter 5, I conclude all findings in my dissertation and provide my perspectives on data-model integration to advance our predictive understanding of land carbon cycle. Overall, my dissertation work develop new

informatic tools and method to tackle three uncertainty sources (i.e., empirical parameter values, incomplete model structures, unrealistic future forcings) which impede realistic ecological predictions. My dissertation indicates an avenue about how to translate big data to improvements in the predictive capability of terrestrial ecosystem carbon cycles. Furthermore, it will benefit policy making about mitigating global warming and the associated climate change.

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation and gratitude to my mentors Drs. Yiqi Luo, Andrew Richardson, Christopher Dougherty, and Michael Gowanlock for their invaluable guidance, support, and encouragement throughout my dissertation. Their expertise, wisdom, and patience have been instrumental in shaping my research and professional development. I am also grateful to my lab members, Drs. Lifen Jiang, Xingjie Lu, Shuang Ma, Chang Gyo Jung, Yuanyuan Huang, Feng Tao, Yu Zhou, and Jian Zhou, for their contributions to my research, insightful discussions, and friendship. Their enthusiasm, dedication, and hard work have made my journey in Luo Lab more fulfilling and enjoyable. I would like to thank my collaborators from the Department of Energy's Spruce and Peatland Responses Under Changing Environments (SPRUCE) project, Drs. Paul Hanson, Daniel Ricciuto, Dan Lu, for their valuable inputs and constructive feedback that greatly contributed to my research findings. I am also grateful to Dr. Ted Schuur and other lecturers who introduced me to the fascinating world of ecology and kindled my interest in this field. I would like to extend my sincere thanks to Dr. Xuehe Lu. Dr. Lu taught me how to conduct my first independent experiment, which was pivotal to my research success. Special thanks to my friends in China who have been my pillars of support: Jingsong Cai, Haiyan Zhang, Yuxin Shi, Ning Wei, and Yanling Tang. I also would like to thank my friends in Flagstaff who helped me prepare for my defense: Jon Wells, Aaron Teets, Oscar Robert Zimmerman, Megan Foley, Natasha Wesely, Yuejie Liu, and Kathleen Orndahl. Finally, I would like to express my deepest gratitude to my parents for their unwavering love. Their love has always been my source of strength, inspiration, and motivation.

TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGEMENTS	iv
PREFACE	viii
CHAPTER 1: INTRODUCTION	1
References	5
CHAPTER 2: A MODEL-INDEPENDENT DATA ASSIMILATION (MIDA) MODULE AND ITS APPLICATION IN ECOLOGY	9
Abstract	10
1. Introduction.....	11
2. Model-independent data assimilation (MIDA).....	15
2.1 Bayes’s theorem and DA	15
2.2 An overview of MIDA.....	16
2.3 Step 1: Data preparation.....	16
2.4 Step 2: Execution of data assimilation.....	18
2.5 Step 3: Visualization	20
2.6 Implementation and architecture of MIDA.....	21
2.7 User information of MIDA	22
3. Applications of MIDA	24
3.1 Case 1: Independent data assimilation with DALEC	24
3.2 Case 2: Application of MIDA to a surrogate land surface model.....	27
3.3 Case 3: Evaluation of multiple phenological models	28
3.4 Case 4: Supporting data assimilation with a dynamic vegetation model.....	29
4. Discussion	30
5. Conclusions.....	33
Appendix A: Nine phenological models.....	35
Appendix B: An example of output configuration file	41
Appendix C: An example of the namelist.txt file	42
References	45
Tables	51
Figures.....	55
CHAPTER 3: ELEVATED WATER TABLE DEPTH UNDER 8-YEAR WARMING TURNS CARBON SINK TO SOURCE IN PERMAFROST	67

Abstract	68
1. Introduction.....	69
2. Method	71
2.1 Study site.....	71
2.2 Model description and parameters	72
2.3 Data assimilation.....	76
2.4 Assessment of the responses of C fluxes to changes in TD and WTD.....	77
3. Results.....	79
3.1 Optimized simulations by data assimilation	79
3.2 Extended simulations of TD and WTD	80
3.3 Responses of C fluxes to variations in TD or WTD	81
4. Discussion	81
4.1 Simulations constrained by data assimilation	82
4.2 Responses of C cycles to alterations in water table depth	83
4.3 The elevated thaw depth promotes carbon sequestration in permafrost ecosystem.....	85
4.4 Limitations of this study	86
5. Conclusions.....	86
Appendix A: Comparison between model simulations and observations	88
Appendix B: Meteorology variables, C fluxes, and.....	90
References.....	91
Tables.....	96
Figures.....	97
CHAPTER 4: A NEW ASSESSMENT OF FORCING CORRECTION AND ITS POTENTIAL TO GUIDE ITERATIVE PREDICTIONS IN A BIG-DATA EAR	106
Abstract	107
1. Introduction.....	108
2. Method	111
2.1 Site description and measurements	111
2.2 Overview of model structure and calibrated parameters.....	113
2.3 Ensemble prediction and forcing correction	115
2.4 Assessment of forcing correction with hindcast	117
3. Results.....	120
3.1 Comparison between conventional assessment and the new assessment ..	120
3.2 Improvement in forecast accuracy due to forcing correction.....	120

3.3	Improvement in forecast accuracy due to individual forcing correction ...	121
3.4	Weekly forcing correction improves forecast accuracy	121
4.	Discussion	122
	Appendix A: Data assimilation	125
	Appendix B: Comparison between ensemble forecast and hindcast	131
	Appendix C: Comparison between historical forcing and ensemble forcing	132
	Appendix D: description of forcing variables.....	134
	References	135
	Figures.....	141
CHAPTER 5:	CONCLUSIONS AND PERSPECTIVES	148
5.1	Model-independent data assimilation module (MIDA)	148
5.2	New formula in water-dependent cycles.....	149
5.3	A new assessment of forcing correction in terms of forecast accuracy	151
	References	153

PREFACE

The dissertation chapters presented here are written in a format for publication in peer-reviewed scientific journals. Chapter 2, entitled “A Model-Independent Data Assimilation (MIDA) module and its applications in ecology”, was published in *Geoscientific Model Development* in 2021. Chapter 3, entitled “Elevated water table depth under 8-year warming turns carbon sink to source in permafrost”, is formatted for *Journal of Geophysical Research*. Chapter 4, entitled “A new assessment of forcing correction and its potential to guide iterative predictions in a big-data era”, is formatted for *Geophysical Research Letters*. Some redundancy will result from combining these manuscripts within the university formatting requirements. Please note that the appendix, tables, figures, and references are listed separately at the end of each chapter.

CHAPTER 1: INTRODUCTION

The Earth that we are living on is under unprecedented threat of climate change. The Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report shows that global warming is expected to surpass 1.5°C over the next 20 years with more severe climate extremes (IPCC, 2021). For 1.5°C of global warming, there will be widespread and intensifying climate change, which is affecting every region across the globe, resulting in irreversible impacts, such as continued sea level rise, from centuries to millennia (Scheffers et al., 2016). Climate mitigation requires us to improve our mechanistic understanding of responses of earth system to climate change and reduce uncertainties in predictions of the future states (Eyring et al., 2019). To mitigate climate change, we need to integrate information contained in big data and knowledge offered from models, which are widely becoming available in ecology and environmental sciences.

Over the past decades, ecology has been evolving into a data-intensive scientific discipline (Kelling et al., 2009; Xia et al., 2020) due to a massive volume and diversity of data collected by remote sensing, observation networks, and extensive global change experiments. For example, the Landsat 8 satellite provides global coverage of the land ecosystems at a spatial resolution of 30 meters (Wulder et al., 2012). Observation networks have been actively and extensively collecting data, e.g., US National Ecological Observatory Network (NEON), FLUXNET, Global Lake Ecological Observatory Network (GLEON), Long Term Ecological Research (LTER). Taking FLUXNET as an example, it provides opportunities to study across multiple ecosystems with worldwide measures from over 100 tower sites. Networks of observations have generated many databases, such as TRY plant trait database (Kattge et al., 2011), global forecast carbon (C) database (ForC) (Anderson-Teixeira et al., 2018), Biofuel

Ecophysiological Traits and Yields database (BETYdb) (LeBauer et al., 2018), and PROFOUND European forest database (Reyer et al., 2020). Moreover, thousands of field global change experiments have been conducted, yielding a deluge of data for scientific research (Song et al., 2019; Xia et al., 2020). The explosion of big data provides massive information to tackle the challenge of climate change.

In addition to data from spatial and temporal observations and field experimental studies, hundreds of process-based models have been developed in ecology and environmental sciences (Manzoni and Porporato, 2007). A process-based model is an unified system comprising hypotheses for essential process components of an ecosystem (Walker et al., 2018). Models have been continuously improved through inclusion of more processes (e.g., nutrient limitations on terrestrial C cycles) (Norby et al., 2010), improvements in parameterizations (Luo and Schuur, 2020), and increases in spatial resolution (Lawrence et al., 2019) . At the same time, these efforts make the models more and more complex and the synchronous increase of model complexity challenges tracing and quantifying sources of uncertainties (Bonan et al., 2019). Improved process representation alone may not make model predictions realistic. Instead, we need to ultimately use diverse and abundant data to improve and constrain models.

With the advent of big data, ecology has become a more and more interdisciplinary field of research, which requires collaboration of interdisciplinary teams (Rineau et al., 2019). For example, the synthesis centers (e.g., NCEAS, SESYNC, NESCent, etc.) bring together fellowships from theory, data management, informatics, and computer science to advance innovative and high-impact research in ecology and environment sciences (Baron et al., 2017). Specifically, to advance our predictive understanding of ecosystems' responses to climate change, (Seidl, 2016) argued that modeling and field experiments are complementary.

Experimental studies can be greatly enhanced by process-based modeling, and if supported by experimental data, models can predict ecosystem responses more realistically (Rineau et al., 2019; Medlyn et al., 2015; Hanson and Walker, 2020). However, interactions between modelers and experimenters are largely curbed by technical barriers (Huang et al., 2019; Gil et al., 2016; Dietze et al., 2013). Thus, we need to develop tools to facilitate knowledge integration from modelers and experimenters.

With continuous model improvements and extensive data volume, our capability to understand and predict ecosystem dynamics and states has not increased accordingly (Bonan and Doney, 2018; Reichstein et al., 2019; Lovenduski and Bonan, 2017). For example, Friedlingstein et al. (2006) analyzed the projections of 11 earth system models (ESMs) and the wide spread of simulated land carbon exchange indicated a broad divergence in the capacity of future terrestrial ecosystem to act as a C sink or source by the end of 21st century (2006). Although more and more processes have been incorporated into ESMs, there is still unanimous agreement among these projections (Friedlingstein et al., 2014). Meanwhile, there is a generally bad fitting between model simulations and observations. For example, Luo et al. (2015) reported that none of 11 models participating in CMIP5 can predict the global soil carbon density compared with observations.

To advance data-intensive science in ecology and environmental sciences, we must integrate big data with models to improve predictive understanding of ecological processes and ecological forecast of future ecosystem states. Moreover, we also need to develop ecoinformatics tools to facilitate the integration of data with models. In this dissertation research, I sought to address three questions: (1) how can we build a data assimilation (DA) tool to facilitate integrating data into models for ecologists without enough programming knowledge? (2) How

can we apply the DA tool to advance mechanistic understanding of ecosystem responses to climate change? And (3) how can we quantitatively compare the forecast accuracy in ecological predictions with data-constrained models?

References

- Anderson-Teixeira, K. J., Wang, M. M. H., McGarvey, J. C., Herrmann, V., Tepley, A. J., Bond-Lamberty, B., and LeBauer, D. S.: ForC: a global database of forest carbon stocks and fluxes, *Ecology*, 99, 1507–1507, <https://doi.org/10.1002/ecy.2229>, 2018.
- IPCC, 2021: <https://www.ipcc.ch/report/sixth-assessment-report-working-group-i/>, last access: 27 February 2023.
- Baron, J. S., Specht, A., Garnier, E., Bishop, P., Campbell, C. A., Davis, F. W., Fady, B., Field, D., Gross, L. J., Guru, S. M., Halpern, B. S., Hampton, S. E., Leavitt, P. R., Meagher, T. R., Ometto, J., Parker, J. N., Price, R., Rawson, C. H., Rodrigo, A., Sheble, L. A., and Winter, M.: Synthesis centers as critical research infrastructure, *BioScience*, 67, 750–759, <https://doi.org/10.1093/biosci/bix053>, 2017.
- Bonan, G. B. and Doney, S. C.: Climate, ecosystems, and planetary futures: The challenge to predict life in Earth system models, *Science*, 359, <https://doi.org/10.1126/science.aam8328>, 2018.
- Bonan, G. B., Lombardozzi, D. L., Wieder, W. R., Oleson, K. W., Lawrence, D. M., Hoffman, F. M., and Collier, N.: Model Structure and Climate Data Uncertainty in Historical Simulations of the Terrestrial Carbon Cycle (1850–2014), *Global Biogeochemical Cycles*, 33, 1310–1326, <https://doi.org/10.1029/2019GB006175>, 2019.
- Dietze, M. C., Lebauer, D. S., and Kooper, R.: On improving the communication between models and data, *Plant, Cell & Environment*, 36, 1575–1585, <https://doi.org/10.1111/pce.12043>, 2013.
- Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., Collins, W. D., Gier, B. K., Hall, A. D., Hoffman, F. M., Hurtt, G. C., Jahn, A., Jones, C. D., Klein, S. A., Krasting, J. P., Kwiatkowski, L., Lorenz, R., Maloney, E., Meehl, G. A., Pendergrass, A. G., Pincus, R., Ruane, A. C., Russell, J. L., Sanderson, B. M., Santer, B. D., Sherwood, S. C., Simpson, I. R., Stouffer, R. J., and Williamson, M. S.: Taking climate model evaluation to the next level, *Nature Clim Change*, 9, 102–110, <https://doi.org/10.1038/s41558-018-0355-y>, 2019.
- Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., Cadule, P., Doney, S., Eby, M., Fung, I., Bala, G., John, J., Jones, C., Joos, F., Kato, T., Kawamiya, M., Knorr, W., Lindsay, K., Matthews, H. D., Raddatz, T., Rayner, P., Reick, C., Roeckner, E., Schnitzler, K.-G., Schnur, R., Strassmann, K., Weaver, A. J., Yoshikawa, C., and Zeng, N.: Climate–Carbon Cycle Feedback Analysis: Results from the C4MIP Model Intercomparison, *Journal of Climate*, 19, 3337–3353, <https://doi.org/10.1175/JCLI3800.1>, 2006.
- Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., and Knutti, R.: Uncertainties in CMIP5 Climate Projections due to Carbon Cycle Feedbacks, *Journal of Climate*, 27, 511–526, <https://doi.org/10.1175/JCLI-D-12-00579.1>, 2014.
- Gil, Y., David, C. H., Demir, I., Essawy, B. T., Fulweiler, R. W., Goodall, J. L., Karlstrom, L., Lee, H., Mills, H. J., Oh, J.-H., Pierce, S. A., Pope, A., Tzeng, M. W., Villamizar, S. R., and Yu, X.: Toward the Geoscience Paper of the Future: Best practices for documenting and sharing research from data to software to provenance, *Earth and Space Science*, 3, 388–415, <https://doi.org/10.1002/2015EA000136>, 2016.

- Hanson, P. J. and Walker, A. P.: Advancing global change biology through experimental manipulations: Where have we been and where might we go?, *Global Change Biology*, 26, 287–299, <https://doi.org/10.1111/GCB.14894>, 2020.
- Huang, Y., Stacy, M., Jiang, J., Sundi, N., Ma, S., Saruta, V., Jung, C. G., Shi, Z., Xia, J., Hanson, P. J., Ricciuto, D., and Luo, Y.: Realized ecological forecast through an interactive Ecological Platform for Assimilating Data (EcoPAD, v1.0) into models, *Geosci. Model Dev.*, 12, 1119–1137, <https://doi.org/10.5194/gmd-12-1119-2019>, 2019.
- Kattge, J., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., Bönisch, G., Garnier, E., Westoby, M., Reich, P. B., Wright, I. J., Cornelissen, J. H. C., Violle, C., Harrison, S. P., Van BODEGOM, P. M., Reichstein, M., Enquist, B. J., Soudzilovskaia, N. A., Ackerly, D. D., Anand, M., Atkin, O., Bahn, M., Baker, T. R., Baldocchi, D., Bekker, R., Blanco, C. C., Blonder, B., Bond, W. J., Bradstock, R., Bunker, D. E., Casanoves, F., Cavender-Bares, J., Chambers, J. Q., Chapin Iii, F. S., Chave, J., Coomes, D., Cornwell, W. K., Craine, J. M., Dobrin, B. H., Duarte, L., Durka, W., Elser, J., Esser, G., Estiarte, M., Fagan, W. F., Fang, J., Fernández-Méndez, F., Fidelis, A., Finegan, B., Flores, O., Ford, H., Frank, D., Freschet, G. T., Fyllas, N. M., Gallagher, R. V., Green, W. A., Gutierrez, A. G., Hickler, T., Higgins, S. I., Hodgson, J. G., Jalili, A., Jansen, S., Joly, C. A., Kerkhoff, A. J., Kirkup, D., Kitajima, K., Kleyer, M., Klotz, S., Knops, J. M. H., Kramer, K., Kühn, I., Kurokawa, H., Laughlin, D., Lee, T. D., Leishman, M., Lens, F., Lenz, T., Lewis, S. L., Lloyd, J., Llusià, J., Louault, F., Ma, S., Mahecha, M. D., Manning, P., Massad, T., Medlyn, B. E., Messier, J., Moles, A. T., Müller, S. C., Nadrowski, K., Naeem, S., Niinemets, Ü., Nöllert, S., Nüske, A., Ogaya, R., Oleksyn, J., Onipchenko, V. G., Onoda, Y., Ordoñez, J., Overbeck, G., et al.: TRY – a global database of plant traits, *Global Change Biology*, 17, 2905–2935, <https://doi.org/10.1111/j.1365-2486.2011.02451.x>, 2011.
- Kelling, S., Hochachka, W. M., Fink, D., Riedewald, M., Caruana, R., Ballard, G., and Hooker, G.: Data-intensive Science: A New Paradigm for Biodiversity Studies, *BioScience*, 59, 613–620, <https://doi.org/10.1525/bio.2009.59.7.12>, 2009.
- Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., Collier, N., Ghimire, B., van Kampenhout, L., Kennedy, D., Kluzek, E., Lawrence, P. J., Li, F., Li, H., Lombardozzi, D., Riley, W. J., Sacks, W. J., Shi, M., Vertenstein, M., Wieder, W. R., Xu, C., Ali, A. A., Badger, A. M., Bisht, G., van den Broeke, M., Brunke, M. A., Burns, S. P., Buzan, J., Clark, M., Craig, A., Dahlin, K., Drewniak, B., Fisher, J. B., Flanner, M., Fox, A. M., Gentine, P., Hoffman, F., Keppel-Aleks, G., Knox, R., Kumar, S., Lenaerts, J., Leung, L. R., Lipscomb, W. H., Lu, Y., Pandey, A., Pelletier, J. D., Perket, J., Randerson, J. T., Ricciuto, D. M., Sanderson, B. M., Slater, A., Subin, Z. M., Tang, J., Thomas, R. Q., Val Martin, M., and Zeng, X.: The Community Land Model Version 5: Description of New Features, Benchmarking, and Impact of Forcing Uncertainty, *Journal of Advances in Modeling Earth Systems*, 11, 4245–4287, <https://doi.org/10.1029/2018MS001583>, 2019.
- LeBauer, D., Kooper, R., Mulrooney, P., Rohde, S., Wang, D., Long, S. P., and Dietze, M. C.: BETYdb: a yield, trait, and ecosystem service database applied to second-generation bioenergy feedstock production, *GCB Bioenergy*, 10, 61–71, <https://doi.org/10.1111/gcbb.12420>, 2018.

- Lovenduski, N. S. and Bonan, G. B.: Reducing uncertainty in projections of terrestrial carbon uptake, *Environmental Research Letters*, 12, <https://doi.org/10.1088/1748-9326/aa66b8>, 2017.
- Luo, Y. and Schuur, E. A. G.: Model parameterization to represent processes at unresolved scales and changing properties of evolving systems, *Global Change Biology*, 26, 1109–1117, <https://doi.org/10.1111/gcb.14939>, 2020.
- Luo, Y., Keenan, T. F., and Smith, M.: Predictability of the terrestrial carbon cycle, *Global Change Biology*, 21, 1737–1751, <https://doi.org/10.1111/gcb.12766>, 2015.
- Manzoni, S. and Porporato, A.: A theoretical analysis of nonlinearities and feedbacks in soil carbon and nitrogen cycles, *Soil Biology and Biochemistry*, 39, 1542–1556, <https://doi.org/10.1016/j.soilbio.2007.01.006>, 2007.
- Medlyn, B. E., Zaehle, S., De Kauwe, M. G., Walker, A. P., Dietze, M. C., Hanson, P. J., Hickler, T., Jain, A. K., Luo, Y., Parton, W., Prentice, I. C., Thornton, P. E., Wang, S., Wang, Y.-P., Weng, E., Iversen, C. M., McCarthy, H. R., Warren, J. M., Oren, R., and Norby, R. J.: Using ecosystem experiments to improve vegetation models, *Nature Climate Change*, 5, 528–534, <https://doi.org/10.1038/nclimate2621>, 2015.
- Norby, R. J., Warren, J. M., Iversen, C. M., Medlyn, B. E., and McMurtrie, R. E.: CO₂ enhancement of forest productivity constrained by limited nitrogen availability, *Proceedings of the National Academy of Sciences*, 107, 19368–19373, <https://doi.org/10.1073/PNAS.1006463107>, 2010.
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat: Deep learning and process understanding for data-driven Earth system science, *Nature* 2019 566:7743, 566, 195–204, <https://doi.org/10.1038/s41586-019-0912-1>, 2019.
- Reyer, C. P. O., Silveyra Gonzalez, R., Dolos, K., Hartig, F., Hauf, Y., Noack, M., Lasch-Born, P., Rötzer, T., Pretzsch, H., Meesenburg, H., Fleck, S., Wagner, M., Bolte, A., Sanders, T. G. M., Kolari, P., Mäkelä, A., Vesala, T., Mammarella, I., Pumpanen, J., Collalti, A., Trotta, C., Matteucci, G., D’Andrea, E., Foltýnová, L., Krejza, J., Ibrom, A., Pilegaard, K., Loustau, D., Bonnefond, J.-M., Berbigier, P., Picart, D., Lafont, S., Dietze, M., Cameron, D., Vieno, M., Tian, H., Palacios-Orueta, A., Cicuendez, V., Recuero, L., Wiese, K., Büchner, M., Lange, S., Volkholz, J., Kim, H., Horemans, J. A., Bohn, F., Steinkamp, J., Chikalanov, A., Weedon, G. P., Sheffield, J., Babst, F., Vega del Valle, I., Suckow, F., Martel, S., Mahnken, M., Gutsch, M., and Frieler, K.: The PROFOUND Database for evaluating vegetation models and simulating climate impacts on European forests, *Earth Syst. Sci. Data*, 12, 1295–1320, <https://doi.org/10.5194/essd-12-1295-2020>, 2020.
- Rineau, F., Malina, R., Beenaerts, N., Arnauts, N., Bardgett, R. D., Berg, M. P., Boerema, A., Bruckers, L., Clerinx, J., Davin, E. L., De Boeck, H. J., De Dobbelaer, T., Dondini, M., De Laender, F., Ellers, J., Franken, O., Gilbert, L., Gudmundsson, L., Janssens, I. A., Johnson, D., Lizin, S., Longdoz, B., Meire, P., Meremans, D., Milbau, A., Moretti, M., Nijs, I., Nobel, A., Pop, I. S., Puetz, T., Reyns, W., Roy, J., Schuetz, J., Seneviratne, S. I., Smith, P., Solmi, F., Staes, J., Thiery, W., Thijs, S., Vanderkelen, I., Van Landuyt, W., Verbruggen, E., Witters, N., Zscheischler, J., and Vangronsveld, J.: Towards more predictive and interdisciplinary climate change ecosystem experiments, *Nature Climate Change*, 9, 809–816, <https://doi.org/10.1038/s41558-019-0609-3>, 2019.
- Scheffers, B. R., De Meester, L., Bridge, T. C. L., Hoffmann, A. A., Pandolfi, J. M., Corlett, R. T., Butchart, S. H. M., Pearce-Kelly, P., Kovacs, K. M., Dudgeon, D., Pacifici, M.,

- Rondinini, C., Foden, W. B., Martin, T. G., Mora, C., Bickford, D., and Watson, J. E. M.: The broad footprint of climate change from genes to biomes to people, *Science*, 354, aaf7671, <https://doi.org/10.1126/science.aaf7671>, 2016.
- Seidl, R.: To Model or not to Model, That is no Longer the Question for Ecologists, *Ecosystems* 2016 20:2, 20, 222–228, <https://doi.org/10.1007/S10021-016-0068-X>, 2016.
- Song, J., Wan, S., Piao, S., Knapp, A. K., Classen, A. T., Vicca, S., Ciais, P., Hovenden, M. J., Leuzinger, S., Beier, C., Kardol, P., Xia, J., Liu, Q., Ru, J., Zhou, Z., Luo, Y., Guo, D., Adam Langley, J., Zscheischler, J., Dukes, J. S., Tang, J., Chen, J., Hofmockel, K. S., Kueppers, L. M., Rustad, L., Liu, L., Smith, M. D., Templer, P. H., Quinn Thomas, R., Norby, R. J., Phillips, R. P., Niu, S., Fatichi, S., Wang, Y., Shao, P., Han, H., Wang, D., Lei, L., Wang, J., Li, X., Zhang, Q., Li, X., Su, F., Liu, B., Yang, F., Ma, G., Li, G., Liu, Y., Liu, Y., Yang, Z., Zhang, K., Miao, Y., Hu, M., Yan, C., Zhang, A., Zhong, M., Hui, Y., Li, Y., and Zheng, M.: A meta-analysis of 1,119 manipulative experiments on terrestrial carbon-cycling responses to global change, *Nat Ecol Evol*, 3, 1309–1320, <https://doi.org/10.1038/s41559-019-0958-3>, 2019.
- Walker, A. P., Ye, M., Lu, D., De Kauwe, M. G., Gu, L., Medlyn, B. E., Rogers, A., and Serbin, S. P.: The multi-assumption architecture and testbed (MAAT v1.0): R code for generating ensembles with dynamic model structure and analysis of epistemic uncertainty from multiple sources, *Geoscientific Model Development*, 11, 3159–3185, <https://doi.org/10.5194/gmd-11-3159-2018>, 2018.
- Wulder, M. A., Masek, J. G., Cohen, W. B., Loveland, T. R., and Woodcock, C. E.: Opening the archive: How free data has enabled the science and monitoring promise of Landsat, *Remote Sensing of Environment*, 122, 2–10, <https://doi.org/10.1016/j.rse.2012.01.010>, 2012.
- Xia, J., Wang, J., and Niu, S.: Research challenges and opportunities for using big data in global change biology, *Global Change Biology*, 26, 6040–6061, <https://doi.org/10.1111/GCB.15317>, 2020.

CHAPTER 2: A MODEL-INDEPENDENT DATA ASSIMILATION (MIDA) MODULE AND ITS APPLICATION IN ECOLOGY

Xin Huang^{1,2}, Dan Lu³, Daniel M. Ricciuto⁴, Paul J. Hanson⁴, Andrew D. Richardson^{1,2}, Xuehe Lu⁵, Ensheng Weng^{6,7}, Sheng Nie⁸, Lifan Jiang¹, Enqing Hou¹, Igor F. Steinmacher², Yiqi Luo^{1,2,9}

1 Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, AZ, USA

2 School of informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ, USA

3 Computational Sciences and Engineering Division, Climate Change Science Institute, Oak Ridge National Laboratory, Oak Ridge, TN, USA

4 Environmental Sciences Division, Climate Change Science Institute, Oak Ridge National Laboratory, Oak Ridge, TN, USA

5 International Institute for Earth System Science, Nanjing University, Nanjing, China

6 Center for Climate Systems Research, Columbia University, New York, USA

7 NASA Goddard Institute for Space Studies, New York, USA

8 Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China

9 Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ, USA

Correspondence to: Xin Huang (xh59@nau.edu)

Abstract

Models are an important tool to predict Earth system dynamics. An accurate prediction of future states of ecosystems depends on not only model structures but also parameterizations. Model parameters can be constrained by data assimilation. However, applications of data assimilation to ecology are restricted by highly technical requirements such as model-dependent coding. To alleviate this technical burden, we developed a model-independent data assimilation (MIDA) module. MIDA works in three steps including data preparation, execution of data assimilation, and visualization. The first step prepares prior ranges of parameter values, a defined number of iterations, and directory paths to access files of observations and models. The execution step calibrates parameter values to best fit the observations and estimates the parameter posterior distributions. The final step automatically visualizes the calibration performance and posterior distributions. MIDA is model-independent and modelers can use MIDA for an accurate and efficient data assimilation in a simple and interactive way without modification of their original models. We applied MIDA to four types of ecological models: the data assimilation linked ecosystem carbon (DALEC) model, a surrogate-based energy exascale earth system model: the land component (ELM), nine phenological models and a stand-alone biome ecological strategy simulator (BiomeE). The applications indicate that MIDA can effectively solve data assimilation problems for different ecological models. Additionally, the easy implementation and model-independent feature of MIDA breaks the technical barrier of applications of data-model fusion in ecology. MIDA facilitates the assimilation of various observations into models for uncertainty reduction in ecological modeling and forecasting.

Keywords:

Parameter uncertainty quantification, Data assimilation, Modules, Ecological models

1. Introduction

Ecological models require a large number of parameters to simulate biogeophysical and biogeochemical processes (Friedlingstein et al., 2006a; Ciais et al., 2013; Bonan, 2019), and specify model behaviors (Luo et al., 2016; Luo and Schuur, 2020). Parameter values in ecological models are mostly determined in some *ad hoc* fashions (Luo et al., 2001), leading to considerable biases in predictions (Tao et al., 2020). The situation becomes even worse when more detailed processes are incorporated into models (De Kauwe et al., 2017; Lawrence et al., 2019). Data assimilation (DA), a statistically rigorous method to integrate observations and models, is gaining increasing attention for parameter estimation and uncertainty evaluation. It has been successfully applied to many ecological models (Fox et al., 2009; Williams et al., 2005; Keenan et al., 2012; Wang et al., 2009; Zobitz et al., 2011; Richardson et al., 2010a; Safta et al., 2015). However, almost all those DA studies require model-dependent, invasive coding (Walls et al., 2005). This requires a DA algorithm to be programmed for a specific model. Such model-dependent coding creates a large technical barrier for ecologists to use DA to solve prediction and uncertainty quantification problems in ecology. Thus, a model-independent DA toolkit is required to facilitate the use of DA technique in ecology.

DA is a powerful approach to combine models with observations and can be used to improve ecological research in several ways (Luo et al., 2011a). First, DA can be used for parameter estimation (Bloom et al., 2016; Hararuk et al., 2015; Ise and Moorcroft, 2006; Ricciuto et al., 2011; Scholze et al., 2007; Ma et al., 2017b; Hou et al., 2019). It enables the optimization of parameter values across sites, time and treatments (Li et al., 2018; Luo and Schuur, 2020a). For example, Hararuk and his colleagues applied DA to a global land model and substantially improved the explainability of the global variation in soil organic carbon (SOC)

from 27% to 41% (Hararuk et al., 2014a). When DA was combined with deep learning to improve spatial distributions of estimated parameter values, for example, the Community Land Model version 5 (CLM5) predicted the SOC distribution in the US continent with much higher R^2 of 0.62 than CLM5 with default parameters ($R^2 = 0.32$) (Tao et al., 2020). Second, DA can be used to select alternative model structures to better represent ecological processes (Van Oijen et al., 2011; Williams et al., 2009; Smith et al., 2013; Shi et al., 2018; Liang et al., 2018a). In the study by Liang et al. (2018), DA was used to evaluate four models. And a two-pool interactive model was selected after DA to best represent SOC decomposition with priming. Additionally, DA can be applied to locate the most informative data to reduce uncertainty, thus guiding the sensor network design. (Keenan et al., 2013; Raupach et al., 2005; Williams et al., 2005; Shi et al., 2018). One DA study at Harvard Forest (Keenan et al., 2013) indicated that only a few data sources contributed to the significant reduction in parameter uncertainty. In spite of powerful applications of DA to ecological research, computational cost is a major hurdle, especially with complex models. Fer et al. (2018) developed a Bayesian model emulation to reduce the time cost of DA from 112h to 6h with the simplified Photosynthesis and Evapotranspiration model. Overall, DA is essential for ecological modeling and forecasting (Jiang et al., 2018a) and is helpful for evaluation of different inversion methods (Fox et al., 2009).

Applications of traditional DA to ecological research require highly technical skills of users. A successful DA application usually involves model-dependent coding to integrate observations into models. This requires users to have knowledge about model programming. For example, if a complex model (e.g., the community land model) is used in DA, users need to know the programming language (e.g., Fortran) of the model and its internal content to write DA algorithm into the model source code before DA can be conducted. The learning curve for model

programming is steep for general ecologists. Furthermore, users often need to update the programming knowledge when a different model is used in DA. For example, scientists who implemented the DA algorithm coded in MATLAB (Xu et al., 2006a) to an ecosystem carbon cycle model programmed in Fortran (e.g., TECO) need to understand both MATLAB and Fortran (Ma et al., 2017b). Moreover, DA often involves reading observation files about a specific study site. As a result, users usually have to update the codes of model-dependent DA to read new observations from every new study site.

A number of tools have been developed to facilitate DA applications (Table 1) but many of them are model dependent, such as the Carbon Cycle Data Assimilation Systems (CCDAS) (Scholze et al., 2007; Rayner et al., 2005), the Carbon Data Model Framework (CARDAMOM) (Bloom et al., 2016), the Ecological Platform for Assimilating Data (EcoPAD) into model (Huang et al. 2019) and Predictive Ecosystem Analyzer (PEcAn) (LeBauer et al., 2013). These tools combine DA algorithms with a specific model. For example, CCDAS specified the DA algorithm to the Biosphere Energy Transfer Hydrology (BETHY) model (Rayner et al., 2005). The hardcoding feature of aforementioned tools make them inflexible to be applied to different models.

There are some model independent DA tools that are not tailored to a specific model, such as Data Assimilation Research Testbed (DART) (Anderson et al., 2009), the open Data Assimilation library (openDA) (Ridler et al., 2014), the Parallel Data Assimilation Framework (PDAF) (Nerger and Hiller, 2013) and Parameter Estimation & Uncertainty Analysis software suit (PEST) (Doherty, 2004).

However, these model-independent tools suffer from some limitations for a general and flexible DA application. For example, openDA requires users to code three functions to initialize

a Java class (Ridler et al., 2014) (Table 1). DART enables incorporating a new model through a range of interfaces (Anderson et al., 2009). It has been successfully applied to atmospheric and oceanic models with currently available interfaces (Raeder et al., 2012; Anderson et al., 2009) and recently to the community land model (Fox et al., 2018). It is likely that users may need to prepare new interfaces for new ecological models to use DART. DART and PDAF adopted the Ensemble Kalman Filter (EnKF) method (Evensen, 2003), which may makes it difficult to obey mass conservation for biogeochemical models. This is because the parameter values estimated by EnKF change each time when new data sets are assimilated (Allen et al., 2003; Trudinger et al., 2007; Gao et al., 2011). The sudden changes in estimated parameter values at time points when data are assimilated by EnKF usually do not reflect reality of biogeochemical cycles in the real world. PEST utilizes Levenberg-Marquardt method (Levenberg, 1944) which is a local optimization method for parameter estimation. If the relationship between simulation outputs and parameters are highly nonlinear, which is common in ecological models, this method may trap into a locally optimization solution (Doherty, 2004; Trudinger et al., 2007).

In this work, we developed a model-independent DA module (MIDA) to enable a general and flexible application of DA in ecology. MIDA was designed as a highly modular tool, independent of specific models, and friendly to users with limited programming skills and/or technical knowledge of DA algorithms. Additionally, MIDA implemented advanced Markov Chain Monte Carlo (MCMC) algorithms for DA analysis which can accurately quantify the parameter uncertainty with informative posterior distribution. The anticipated user community in this initial phase of MIDA development is the biogeochemical modelers who are looking for appropriate parameter estimation methods. In the following Section 2, we first introduce the development details of MIDA and its usage. In Section 3, we demonstrate the application of

MIDA to four different types of ecological models. In Section 4, we discuss the strengths and weaknesses of MIDA in ecological modelling and lastly we give our concluding remarks in Section 5.

2. Model-independent data assimilation (MIDA)

2.1 Bayes's theorem and DA

Based on Bayes' theorem, DA is a statistical approach to constrain parameter values and estimate their posterior density distributions through assimilating observations into a model. The posterior density distributions $p(C|Z)$ of parameters C for a given observation Z can be obtained from *prior* density distributions $p(C)$ and the likelihood function $p(Z|C)$:

$$p(C|Z) \propto p(Z|C)p(C) \quad (1)$$

The *prior* density distribution $p(C)$ is assumed as a uniform distribution over the parameter range. And the likelihood function is negatively proportional to a cost function, J as:

$$p(Z|C) \propto \exp(-J) \quad (2)$$

The cost function measures the misfit between simulation outputs and observations and is described in more detail in section 2.4. The posterior density distributions $p(C|Z)$ is estimated from sampling parameter values to maximize the likelihood function $p(Z|C)$ or minimize the cost function J . DA usually uses a sampling technique, such as Markov chain Monte Carlo (MCMC) in this MIDA. The MCMC algorithm successively generates a new set of parameter values from the prior parameter ranges and requires model run with these new parameter values. Then the cost function is calculated to determine whether this new set of parameter values will be accepted or not according to the Metropolis-Hastings criterion (see more description in section 2.4). All accepted parameter values are used to generate posterior distributions where the

distinctive mode indicates the parameter uncertainty is well constrained. Meanwhile, we derive maximum likelihood estimates (MLEs) of parameters from the posterior density distributions.

MIDA realizes model-independent Bayesian-based DA to estimate posterior density distributions and MLEs of parameters via data exchanges between a given model and DA algorithm.

2.2 An overview of MIDA

MIDA is a module that allows for automatic implementation of data assimilation without intrusive modification or coding of the original model (<https://doi.org/10.5281/zenodo.4762725>, last access: May 2021). Its workflow includes three steps: data preparation, execution of data assimilation, and visualization (Fig. 1). Step 1 (data preparation) is to establish the standardized data exchange between DA algorithm and the model. Step 2 (execution of data assimilation) is to run DA as a black box independent of the model. Step 3 (visualization) is to diagnose parameter uncertainty after DA. The modularity of the 3-step workflow is designed to enable MIDA for a rapid DA application and adaption to a new model. In the following, we introduce the three-step workflows of MIDA, its technical implementation and usage in detail.

2.3 Step 1: Data preparation

Step 1 is designed to initialize data exchange to transfer parameter values, model outputs, observations and their variances between DA algorithm and the model to be used. Four types of information are required either from interactive input or by modifying the ‘namelist.txt’ file (Fig. 1). The first type is about DA configuration, including the number of sampling series in DA and the working path where the outputs of DA will be saved. The number of a sampling series is essential in a DA task to define how many times parameter values are sampled to run the model. The second type of information is about parameter ranges and their covariance. The third is the

model executable file. Finally, the fourth type is an output configuration file which contains the file paths of model outputs, observations, and their variance. This file also instructs how to read model outputs and compare each output with corresponding observation.

Traditional DA requires users to modify the code of model to incorporate the process of data exchange between DA algorithm and the model. Therefore, the program of data exchange in traditional DA is model-specific and users need to repeat such program when a new model comes. In MIDA, the process of data exchange calls a model executable file which hides the details of model code. When applied to a new model, MIDA only requires users to provide a different model executable file in the 'namelist.txt' file and does not involve any additional coding in either the model or MIDA. Thus, MIDA lowers the technical barrier for general ecologists to conduct DA.

Traditional DA usually preset the number of parameters and the model outputs according to a specific model before initializing the data exchange. This is because data exchange between DA algorithm and model uses memory to transfer items such as parameter values. Instead, MIDA organizes items in data exchange using different files. Items in data exchange are decided by the data file loaded when MIDA is running. The number of parameter values, for example, will be decided after the file of parameter range is read in MIDA. Through modifying files, MIDA allows making efficient choices about the model-related items in data exchange. Thus, MIDA is highly flexible and modular for DA with different models.

Traditional DA also preset observation types in the data exchange according to a specific study before the data exchange. For example, if the traditional DA uses carbon flux observation, it cannot switch to satellite remote sensing products without additional coding. MIDA uses the concepts of object-orient programming (Mitchell and Apt, 2003) and dynamic initialization

(Cline et al., 1998) in computer science to provide a homogenous way to create various observation types from a unified prototype class. A prototype class includes variables to store observations and their variance and functions (e.g., read from observation files). The values in variables are dynamically decided after the observation files are loaded when MIDA is running. Different observation types derive from the prototype class with a high degree of reusability of most functions. In such way, MIDA only requires users to provide different filenames of the observations to be integrated in DA. Therefore, MIDA is highly flexible and modular for DA to assimilate various observations.

2.4 Step 2: Execution of data assimilation

After the establishment of the standardized data exchange (step 1), step 2 is to run DA as a black box for users without knowledge of DA itself. Notwithstanding the black-box goal, this section provides a general description of DA below.

Data assimilation as a process integrates observations into a model to constrain parameters and estimate parameter uncertainties. Data assimilation usually uses some types of sampling algorithms, such as Markov chain Monte Carlo (MCMC), to generate posterior parameter distribution under a Bayesian inference framework (Box and Tiao, 1992). As mentioned in section 2.1, DA with MCMC algorithm estimates the posterior density distributions through sampling to maximize likelihood function $p(Z|C)$ or minimize the misfit J between simulation outputs and observations. This version of MIDA uses MCMC algorithm implemented by the Metropolis-Hasting (MH) sampling method (Hastings, 1970; Metropolis et al., 1953a). The future version of MIDA could incorporate other data assimilation algorithms. Each iteration in the Metropolis-Hasting sampling includes a proposing phase and a moving phase. The proposing phase generates a new set of parameter values based on the starting point for the first

iteration or current accepted parameter values in the following iterations. If parameter covariance (cov_{param}) is specified in step 1 on data preparation, this proposing phase will draw new parameter values (C_{new}) within the prior ranges from a Gaussian distribution $N(C_{old}, cov_{param})$ where C_{old} is the predecessor set of parameter values. Without parameter covariance, new set of parameter values will be generated from a uniform distribution within the prior ranges (Xu et al., 2006b).

The moving phase first calculates mismatches between observations and the model simulation with the new set of parameter values as a cost function (J_{new} in Eq.3) (Xu et al. 2006):

$$J_{new} = \sum_{i=1}^n \frac{\sum_{t \in obs(z_i)} [Z_i(t) - X_i(t)]^2}{2\sigma_i^2} \quad (3)$$

Where n is the number of observations, $Z_i(t)$ is the i^{th} observation at time t , $X_i(t)$ is the corresponding simulation, σ_i^2 is the variance of the observation. The error is assumed to independently follow a Gaussian distribution. This new set of parameter values will be accepted if J_{new} is smaller than J_{old} , the cost function with the previous set of accepted parameter values, or the value, $\exp\left(-\frac{J_{new}}{J_{old}}\right)$, is larger than a random number selected from a uniform distribution from 0 to 1 according to the Metropolis criterion (Shi et al., 2018; Liang et al., 2018b; Luo et al., 2011a; Xu et al., 2006a). Once the new set of parameter values is accepted, J_{new} becomes J_{old} . Those two phases of sampling will be iteratively executed until the number of sampling series set in step 1 on preparation of DA is reached. Finally, the posterior density distributions can be generated from all the accepted parameter values.

MIDA realizes the execution of data assimilation according to the procedure described above. First, MIDA uses a ‘call’ function to execute model simulations to get values of $X_i(t)$.

Observations $Z_i(t)$ and their variance σ_i^2 are already provided via the standardized data exchange as described in step 1. Then, MIDA calculates J_{new} according to Eq. 3 to decide the acceptance of the current parameter values used in this simulation. If accepted, MIDA saves this set of parameter values and associated J_{new} values in $C_{accepted}$ and $J_{accepted}$ arrays respectively and triggers new proposing phrase based on this set of accepted parameter values. If not, MIDA discards this set of parameter values and generates another new set of parameter values. MIDA saves the new parameter values generated in the proposing phrase to “ParameterValue.txt”, from which the model reads before execution of the next model simulation. MIDA repeats the proposing and moving phases until the number of sampling series is reached. At the end, MIDA selects the best parameter values through maximum likelihood estimation and run model again using this set of values to get optimized simulation outputs $X_i(t)$. Then MIDA saves the arrays of accepted parameters, associated errors, maximum likelihood estimates (MLEs), and optimized state variables $X_i(t)$ to four files, “parameter_accepted.txt”, “J_accepted.txt”, “MLE.txt”, and “OptimizedSimu.txt”, respectively.

This execution of DA algorithm in MIDA enables users to conduct DA as a black box and is independent of any particular model.

2.5 Step 3: Visualization

Step 3 is to visualize the results of DA in step 2. The end products of DA are accepted parameter values, their associated J_{new} values, the maximum likelihood estimates, and optimized simulation results as saved in the output files. MIDA enables visualization of parameter posterior density distributions with a Python script. In the script, MIDA first read accepted parameter values from “parameter_accepted.txt” file. Then, MIDA generates posterior probabilistic density function (PPDF) for each parameter via ‘kdeplot’ function in the ‘seaborn’ package. The

maximum likelihood estimates of parameters correspond to the peaks of PPDF. The distinctive mode of PPDF indicates how well the parameter uncertainty is constrained. Finally, MIDA visualizes the PPDF for all parameters in a figure using the ‘matplotlib’ package.

2.6 Implementation and architecture of MIDA

MIDA is equipped with a graphical user interface (GUI) and users can easily execute it through an interactive window. Users can also run MIDA as a script program without the GUI. MIDA is written in Python (version 3.7). For the GUI-version, all relevant Python packages used in MIDA are compiled together, thus users do not need to install them by themselves. For the non-GUI version, users need to install Python 3.7 and relevant packages (i.e., numpy, pandas, shutil, subprocess, matplotlib, math, os, and seaborn). MIDA is compatible with model source codes written in multiple programming language (e.g., Fortran, C/C++, C#, MATLAB, R, or Python). It is also independent of multiple operation systems (e.g., Windows, Linux, MacOS). In addition, MIDA is also able to run on high-performance computing (HPC) platforms via task management systems (e.g., Slurm).

The architecture of MIDA is class-based and each class is designed to describe an object (e.g., parameter, observations, etc.) with variables and operations. Five classes are defined in MIDA: parameter, observation, initialization, MCMC algorithm and the main program. The main program is the start of MIDA execution. It calls functions from all other classes to finish three-step workflow. As described in section 2.2, parameter and observation classes contain variables to be transferred in data exchanges via file I/O operations. These operations are implemented using the ‘numpy’ package. The initialization class is to read ‘namelist.txt’ in step 1 on data preparation and to assign values for the variables in all other classes. Then the class of MCMC algorithm conducts DA as described in step 2. In this step, the simulation operation uses

a ‘call’ function in ‘subprocess’ package to call model executable. At the start of model simulation, MIDA writes new parameter values to the ‘ParameterValue.txt’ file in the ‘working path’ directory specified in step 1 on data preparation. Then the model executable read parameter values from the ‘ParameterValue.txt’ file and run. After model simulation, DA algorithm can read the model outputs by the output filenames indicated in the output configuration file. After DA, step 3 executes an additional Python script to read accepted parameter values and plot the posterior density distributions of parameters. The plotting operations uses ‘matplotlib’ and ‘seaborn’ packages. The implementation of GUI uses PyQt5 toolkit to support interactive usage of MIDA. Users can also run MIDA in a non-interactive way with a ‘main.py’ script to trigger the three-step workflows.

2.7 User information of MIDA

In order to use MIDA, users need to prepare data and a model. The data to be used in MIDA are prior ranges and default values of parameters, parameter covariances, output configuration file, observations and their variances. They are organized in different files. Before running MIDA, users need to specify their filenames as suggested in step 1. When users want to use different data sets in DA, they can simply change filenames with the new data sets via GUI or in the ‘namelist.txt’ file. Figure C1 is an example of the ‘namelist.txt’ file for a data assimilation study with the DALEC model. The model to be used in MIDA should have those to-be-estimated parameter values not fixed in model source code rather than changeable through ‘ParameterValue.txt’ file. MIDA writes new parameter values in each proposing phase during DA to the ‘ParameterValue.txt’ file, from which the model reads the parameter values to run the simulation.

To calculate the cost function, J , we have to have a one-to-one match between observations and model outputs. For example, phenology models in one of the application cases of MIDA below generate discrete dates of leaf onset, which is a one-to-one match to the observations of spring leaf onset. In this case, observation $Z_i(t)$ and model output $X_i(t)$ to be used in calculation of J is straightforward. In the application case for dynamic vegetation, the data to be used are leaf area in six layers in a forest of 302 years old whereas the model simulates leaf areas in eight layers from 0 to 800 years. To match observation, the model generates outputs of leaf areas in six layers when simulated forest age reaches 302 years. This requires users to prepare an output configuration file to instruct MIDA to read model outputs and re-organize their outputs to match observation. The output configuration file starts with a single line listing an observation filename and its corresponding output filenames. Content after the directories in the output configuration file are instructions to map model outputs with the observation signified in the first line. Each instruction is to match one or continuous elements in observation with elements in outputs with the same length. A blank line means there are no further instructions. Then a new matching between another observation and model outputs starts. An example of output configure file is available in Appendix B.

Once MIDA finishes the execution of data assimilation, users may need basic knowledge to assess the performance of DA. For example, the acceptance rate, which is given by MIDA, is the fraction of proposed parameter values that is accepted. Ideally, the acceptance rate should be about 20 ~ 50% (Xu et al., 2006a). A very low acceptance rate indicates that many new proposed parameter values (C_{new}) are rejected because C_{new} jumps too far away from the previously accepted parameter values (Robert and Casella, 2013; Roberts et al., 1997). In this case, users are suggested to reduce a jump scale in the proposing phase. On the other hand, a very high

acceptance rate is likely because C_{new} moves slowly from the previously accepted parameter values. Users may increase the jump scale.

In addition, DA usually requires a convergence test to examine whether posterior distributions from different sampling series converge or not. Convergence test requires running DA parallelly or in multiple times with different initial parameter values. MIDA provides a Gelman-Rubin (G-R) test (Gelman and Rubin, 1992) for this purpose. To use the G-R test, users need to prepare a file containing initial parameters values in different sampling series and indicate its filename in the ‘namelist.txt’ file as described in step 1. If the G-R statistics approaches one, the sampling series in DA is converged. When sampling series is converged, all accepted parameter values are used to generate the posterior distributions.

There are three types of posterior distributions: bell-shape, edge-hitting, and flat. The bell-shaped posterior distributions indicate that these parameters are well constrained. Their peak values are the maximum likelihood estimates of parameter values. The flat posterior distributions suggest that the parameters are not constrained due to the lack of relevant information in data. The edge-hitting posterior distributions result from complex reasons, such as improper prior parameter range. Users may change the prior ranges to examine if those posterior distributions can be improved or examine correlations among estimated parameters.

3. Applications of MIDA

We applied MIDA to four groups of models, which are an ecosystem carbon cycle model, a surrogate-based land surface model, nine phenology models, and a dynamic vegetation model, respectively. These four cases demonstrate that MIDA is effective for stand-alone DA, flexible to be applied to different models, and efficient for multiple model comparison.

3.1 Case 1: Independent data assimilation with DALEC

The first case study is to demonstrate that MIDA can be effective for independent data assimilation with the data assimilation linked ecosystem carbon (DALEC) model (Lu et al., 2017). DALEC has been used for data assimilation in several studies (Bloom et al., 2016; Safta et al., 2015; Lu et al., 2017; Williams et al., 2005; Richardson et al., 2010a). Previous studies all incorporated data assimilation algorithms into DALEC, which requires invasive coding. This case study is focused on reproducing the data assimilation results as in the study (Lu et al., 2017) but with MIDA.

The version of DALEC used in this study is composed of six submodels (i.e., photosynthesis, phenology, autotrophic respiration, allocation, litterfall, and decomposition) to simulate the carbon exchanges among five carbon pools (i.e., leaf, stem, root, soil organic matter and litter) (Ricciuto et al., 2011). There are 21 parameters in DALEC, of which, 17 parameters are derived from the six submodels and four parameters serve to initialize the carbon pools. Table 2 summarizes the names, prior ranges and nominal values of these 21 parameters. The observation is the Harvard Forest daily net ecosystem exchange (NEE) from year 1992 to 2006. DALEC is coded in Fortran. In windows system, a gfortran compiler converts the model code to an executable file (i.e., DALEC.exe).

Figure 2 is the GUI window of MIDA. We first set up a DA task as described in step 1 using the upper panel. In this application, the number of sampling series is set as 20,000. Once users click the ‘choose a directory’ or ‘choose a file’ button, a new dialog window will pop up and users are able to choose the directory or load files interactively. As describe in step 1 on preparation of DA, the working path is where the outputs of DA and ‘ParameterValue.txt’ are saved (e.g., C:/workingPath). After the output configuration file is loaded, the filenames of model outputs, observations and their variance will be displayed in the window automatically.

This application only uses a ‘NEE.txt’ observation file. Similarly, after users load parameter range file (e.g., a file named ‘ParamRange.txt’ contains three rows which are minimum, maximum and default values of parameters), the content in this file is displayed as well. To replace the current parameter range file loaded, users can simply upload another file. In this application, the executive model file is ‘DALEC.exe’ with Fortran compiler in windows system. Because we do not have parameter covariance information, this input is left blank. After ‘save to namelist file’ is clicked, a ‘namelist.txt’ file containing all the inputs will be generated in the working path.

After the DA task set up, we load the ‘namelist.txt’ file and click the ‘run data assimilation’ button in the lower panel to trigger step 2 on execution of DA. A new dialog will pop up to show the acceptance rate information and notify the termination of DA. Then we will click the ‘generate plots’ button to visualize the posterior distributions of 21 parameters as described in step 3.

Figure 3 shows that the simulation outputs using the optimized parameter values from MIDA better fit with the observations than those using default parameter values. Figure 4 depicts posterior distributions of the 21 parameters estimated from MIDA. More than half of the parameters are constrained well with a unimodal shape. $X_{stem_{init}}$ and $X_{root_{init}}$ have a wide occupation of the prior range, indicating that the observation data does not provide useful information for them. The constrained posterior distributions in this study are similar to those from the study in (Lu et al., 2017). Note that MCMC estimates have a large variance and a low convergence rate especially in high-dimensional problems, with a finite number of samples it is not expected that two simulations would give exactly the same results.

3.2 Case 2: Application of MIDA to a surrogate land surface model

This case study is to examine the applicability of MIDA to a surrogate-based land surface model. The original model is energy exascale earth system model: the land component (ELM) (Ricciuto et al., 2018). As ELM is computationally expensive (one forward model simulation takes more than one day), a sparse-grid (SG) surrogate system was developed to reduce the computational time (Lu et al., 2018). The forcing data for the surrogate model is half-hourly meteorological measurements at Missouri Ozark flux site from 2006 to 2014. The observations that were used for optimization are annual sums of net ecosystem exchange (NEE), annual averages of total leaf area index and latent heat fluxes from 2006 to 2010. The eight parameters selected (Table 3) are the most important parameters for the variations in outputs (Ricciuto et al., 2018). The model is written in Python. A ‘pyinstaller’ library packages the model code into an executable file. The iteration number in MIDA is 20,000.

Figure 5 shows posterior distributions of calibrated parameters. c_{root} , SLA_{top} , $t_{leaf\ fall}$, GDD_{onset} are constrained well with a unimodal distribution. However, the distribution of the rest 4 parameters (i.e., N_{leaf} , CN_{root} , A_{r2l} and Res_m) cluster at near the edge. These results match well with the study by (Lu et al., 2018). As shown in Fig. 6, the calibrated parameters induce a performance improvement in simulating total leaf area index and NEE. For latent heat, both the default and optimized simulation obtain good agreement with the observation. These conclusions are also similar to those in (Lu et al., 2018).

MIDA hides the detailed differences between models. For example, DALEC model in case 1 is a process-based model to simulate ecosystem carbon cycle while surrogate-based ELM in case 2 is an approximation of land surface model. They are also different in programming language, simulation time, forcing data, etc. MIDA is able to deal with models with so many

different characteristics and hides these differences from users. Users only need to indicate the filenames of the model to be used, its parameter range, the output configuration file, etc. in the ‘namelist.txt’ file. Thus, MIDA simplified the DA applications using different models.

3.3 Case 3: Evaluation of multiple phenological models

This study case uses nine phenological models (Yun et al., 2017) to demonstrate the applicability of MIDA in model comparison. Five out of the nine models predict phenological events, such as the day of leaf onset, using growing degree days, which are calculated as temperature accumulation above a base temperature. The other four models consider two processes: chilling effects of cold temperature on dormancy before budburst and forcing effects of warm temperature on plant development. Each model uses different response functions to represent chilling and forcing effects. The detailed model descriptions and associated parameter information are in supplementary table.

Data are from the Spruce and Peatland Responses Under Climatic and Environmental Change experiment (SPRUCE) (Hanson et al., 2017) located in northern Minnesota, USA. The experiment consists of five-level whole-ecosystem warming (i.e., +0, +2.25, +4.5, +6.75, +9°C) and two-level elevated CO_2 concentrations (i.e., +0, +500ppm). Dates of leaf onset were observed with PhenoCam (Richardson et al., 2018) for tree species: *Picea mariana* and *Larix laricina*. For the sake of demonstration of MIDA application, we only show DA results for *Larix laricina* with +9°C warming treatment and +0 ppm CO_2 treatment from 2016 to 2018.

MIDA was used to compare performances of the nine models in reference to the same observations of leaf onset dates after DA. We as users changed filenames of model executable file (i.e., PhenoModels.exe), defined parameter ranges, and assigned the directory of working path for each model. MIDA then estimated the optimized parameters and save the corresponding

best simulation outputs to the working path for each of the nine models. Figure 7 shows the best simulation output of these nine models. The simulation output of the 6th, 7th, 8th, and 9th models better fit the observation than the other models. It demonstrates that models that consider both chilling and heating effects can achieve good simulations of the leaf onset dates.

3.4 Case 4: Supporting data assimilation with a dynamic vegetation model

This case study is to examine the efficiency of MIDA to integrate remote sensing data into a dynamic vegetation model. The model used in this study is Biome Ecological strategy simulator (BiomeE) (Weng et al., 2019). This model simulates vegetation demographic processes with individual-based competition for light, soil water, and nutrients. Individual trees in BiomeE model are represented by cohorts of trees with similar sizes. The light competition among cohorts is based on their heights and crown areas according to the rule of perfect plasticity approximation (PPA) model (Strigul et al., 2008). Each cohort has seven pools: leaves, roots, sapwood, heartwood, seeds, nonstructural carbon and nitrogen. After carbon are assimilated into plants via photosynthesis, the assimilated carbon enters to nonstructural carbon pool and is used for plant growth (i.e., diameter, height, crown area) and reproduction according to empirical allometric equations (Weng et al., 2019). In this application, two parameters to be constrained (Table 4) are annual productivity rate and annual mortality rate of trees.

Observations to be used in DA are leaf area indexes in six vertical heights (i.e., 0-5m, 6-10m, 11-15m, 16-20m, 21-25m, and 26-30m) at Willow Creek study site, Wisconsin, USA. The forest at the site is an upland deciduous broadleaf forest of around 302 years old. The observations were from Global Ecosystem Dynamics Investigation (GEDI) acquired by a Light Detection and Ranging (Lidar) laser system, which is deployed on the International Space Station (ISS) by NASA in 2018 (Dubayah et al., 2020). The observations were first averaged

from three footprints and then leaf area indexes in the six canopy layers were standardized to be summed up as one.

To use MIDA, we reorganized the simulation outputs to match observations as suggested in section 2.6. The BiomeE model simulates leaf areas in eight layers (i.e., 0-5m, 6-10m, 11-15m, 16-20m, 21-25m, 26-30m, 31-35m, and 36-40m) from 0 to 800 years. An output configuration file was provided to post-process model outputs of leaf area indexes in six layers to match observations at the forest age of 302 years. These simulated leaf area indexes in the six canopy layers were also standardized to match standardized observations of leaf area indexes. The observations and post-processed simulation outputs were saved to 'LAI.txt' and 'simu_LAI.txt' files, respectively. The two files are used in MIDA for data assimilation to generate posterior distributions of estimated two parameters as showed in Fig. 8. The optimized parameter values through maximum likelihood estimation are different from their default values. Figure 9 compares the simulation outputs with optimized parameters estimated by MIDA to those with default parameter values. After DA with GEDI data in MIDA, the simulation accuracy of leaf area index is substantially improved especially in middle (16~20m) and highest (26~30m) layers.

4. Discussion

This study introduced MIDA as a model-independent tool to facilitate the applications of data assimilation in ecology and biogeochemistry. The potential user community is ecologists with limited knowledge of model programming and technical implementation of DA algorithms. Several model-independent DA tools have already been developed, such as DART (Anderson et al., 2009), openDA (Ridler et al., 2014), PDAF (Nerger and Hiller, 2013) and PEST (Doherty, 2004), mainly for applications in research areas of hydrology, atmosphere, and remote sensing.

These DA tools either use gradient descent method, such as Levenburg-Marquardt algorithm in PEST, or Kalman Filter methods, such as EnKF in DART, openDA, and PDAF. The Levenburg-Marquardt algorithm is a local search method, which is hard to find global optimization solution for highly nonlinear models. EnKF updates state variables and parameter values each time when observations are sequentially assimilated, resulting discrete values of estimated parameters. Jumps in estimated parameter values by EnKF make it very difficult to obey mass conservation in biogeochemical models (Gao et al., 2011). In this study, we used the MCMC method in MIDA to generate parameter values and their posterior distributions. MCMC is a widely used method in many DA studies with biogeochemical models but has been applied to individual models with invasive coding (Ricciuto et al., 2011; Hararuk et al., 2015; Bloom et al., 2016; Liang et al., 2018a; Luo and Schuur, 2020a). Compared to the other model-independent DA tools mentioned above, MIDA is the first tool that uses the MCMC method for DA.

Biogeochemical models are incorporating more detailed processes related to carbon and nitrogen cycles (Lawrence et al. 2020). Complex biogeochemical models yield predictions with great uncertainty (Frienlingstein et al. 2009 and 2014). Data assimilation has been increasingly used to estimate parameter values against observations and reduce uncertainty in model prediction (Luo et al. 2016, Luo and Schuur 2020). However, current applications of DA are almost all model dependent. It requires ecologists to write code to integrate DA algorithm with models. The coding practice is a big technical challenge for ecologists with limited program ability. The distinct advantage of MIDA is to enable ecologists to conduct model independent DA. MIDA streamlines workflow of the three-step procedure for DA to enable users to conduct DA without extensive coding. Users mainly need to provide numerical and character values for data exchanges to transfer data (i.e., parameter values, simulation outputs, observations) between

the model and MIDA by a file named ‘namelist.txt’ or by interactive inputs via a GUI window (Fig. 2).

We tested MIDA in four cases for its applicability to ecological models. The first case is applied to DALEC model, which has been used in several data assimilation studies (Bloom et al., 2016; Lu et al., 2017; Safta et al., 2015; Williams et al., 2005). The previous DA studies all used invasive coding to incorporate DA algorithm into models. As demonstrated in this study, MIDA was applied to DALEC without invasive coding but by providing the directory to save DA results and filenames of DALEC model executable, parameter prior range, and output configuration file through the ‘namelist.txt’ file or interactive inputs in the first preparation step of the workflow. Then, MIDA run DA as a black box with DALEC before visualizing the DA results. Next, we tested the applicability of MIDA a surrogate-based ELM model and a dynamic vegetation model BiomeE. To switch the test case from DALEC to the surrogate-based ELM model and the BiomeE model, we changed the filenames of model executable, parameter prior range, and output configuration file in the ‘namelist.txt’ file for MIDA. This flexibility of MIDA in switching models for DA makes it much easier for model comparisons. We tested this capability of MIDA with nine phenological models to compare alternative model structures. Similarly, MIDA enables efficient switches of observations to be assimilated into models. Users only need to change filenames of observations in the output configuration file. This feature of MIDA makes it easier to utilize abundant traits databases such as TRY (Kattge et al., 2020), FRED (Iversen et al., 2017), etc. Moreover, this feature of MIDA also helps evaluating the relative information content of different observations for constraining model parameters and prediction (Weng and Luo, 2011). Consequently, MIDA can facilitate selection of the most informative observations and then better guide data collections in field experiments. Ultimately,

MIDA can aid ecological forecasting and help reduce uncertainty in model predictions (Huang et al., 2018; Jiang et al., 2018).

Although MIDA helps users to get rid of model detail, users may still need basic knowledge about the model outputs to prepare the output configuration file which is to match model outputs to observations one-by-one (see Section 2.6). This effort of preparing the correspondence between model outputs and observations for MIDA is not that difficult because users are reading or writing a text file and most model developers will provide reference to help understanding observations or model output files.

Generally, MIDA requires longer time to run DA than the embedded DA algorithm, because MIDA calls model simulation as an external executable rather than a function embedded. Thus, we recommend MIDA for beginners of DA users with models that are less complex. Besides, the current version of MIDA only incorporates Metropolis-Hasting sampling approach. More MCMC methods (e.g., Hamiltonian Monte Carlo) may be incorporated into MIDA in the future.

5. Conclusions

We developed MIDA to facilitate data assimilation for biogeochemical models. Traditional DA studies require ecologists to program codes to integrate DA algorithms into model source codes. The easy-to-use MIDA module enables ecologists to conduct model-independent DA without extensive coding thus advancing the application of DA for ecological modeling and forecasting. We demonstrated the capability of MIDA in four cases with a total of 12 ecological models. These cases showed that MIDA is easy to perform for a variety of models and can efficiently produce accurate parameter posterior distributions. Moreover, MIDA supports flexible usage of different models and different observations in the DA analysis and allows a quick switch from

one model to another. This capability enables MIDA to serve as an efficient tool for model intercomparison projects and enhancing ecological forecasting.

Appendix A: Nine phenological models

1. Growing degree (GD)

The growing degree (GD) model is one of the most widespread phenological model to simulate the date of leaf onset ($\hat{D}$). In this study, the time scale is limited to daily based on observation records. The kernel of GD is to calculate the growing degree days (GDD, $\sum_{d=D_s}^{\hat{D}-1} \Delta d$) which is the heat accumulation above a base temperature (T_b). For simplicity, the daily temperature (T_d) can be approximated by the average of daily maximum and minimum temperatures. The heat accumulation starts at day D_s , which is empirically estimated, and ends when GDD reaches a forcing requirement threshold (R_d). Two parameters to be constrained are base temperature (T_b) and the forcing requirement (R_d). Their default values and prior range are listed in Table A1.

$$\Delta d = \begin{cases} T_d - T_b & \text{if } T_d > T_b \\ 0 & \text{otherwise} \end{cases} \quad (\text{A1})$$

$$\sum_{d=D_s}^{\hat{D}-1} \Delta d < R_d \leq \sum_{d=D_s}^{\hat{D}} \Delta d \quad (\text{A2})$$

2. Sigmoid function (SF)

Compared to the linear response function of GDD in GD model, the sigmoid function (SF) model provides a non-linear function to better represent the non-linearity of the growth response to heat accumulation. Three parameters to be constrained in DA are base temperature (T_b), the forcing requirement (R_d) and temperature sensitivity (S_t). Their default values and prior range are listed in Table A1.

$$\Delta d = \frac{1}{1 + e^{S_t(T_d - T_b)}} \quad (\text{A3})$$

$$\sum_{d=D_s}^{\hat{D}-1} \Delta d < R_d \leq \sum_{d=D_s}^{\hat{D}} \Delta d \quad (\text{A4})$$

3. Beta function (BF)

In reality, the plant growth rate, as described with Δd , gradually increases up to a specific temperature, then rapidly declines to a supra-optimal level. Such response can be well described by a beta function with uni-modality and non-symmetrical shape. Three parameters are involved in DA: minimum temperature (T_n), optimal temperature (T_o) and forcing requirement (R_d). The other parameter values are fixed with empirical values. For example, maximum growth rate (R_x) is set to one and maximum temperature (T_x) is assumed to be 45.

$$r_d = R_x \left(\frac{T_x - T_d}{T_x - T_o} \right) \left(\frac{T_d - T_n}{T_o - T_n} \right)^{\frac{T_o - T_n}{T_x - T_o}} \quad (\text{A5})$$

$$\Delta d = \begin{cases} r_d & \text{if } r_d > 0 \\ 0 & \text{otherwise} \end{cases} \quad (\text{A6})$$

$$\sum_{d=D_s}^{\hat{D}-1} \Delta d < R_d \leq \sum_{d=D_s}^{\hat{D}} \Delta d \quad (\text{A7})$$

4. Days transferred to standard temperature (DTS)

According to Arrhenius law, the relationship between growth rate and daily temperature T_d can be interpolated by the equation 8 (Ono and Konno, 1999). With a factor weighted by standard temperature, the equation for DTS (Eq. A9) can better represent growth rate dependent on temperatures. Three parameters considered in DA are: temperature sensitivity rate (E_a), standard temperature (T_s) and forcing requirement (R_d).

$$k = e^{\frac{-E_a}{R \cdot T_d}} \quad (\text{A8})$$

$$\Delta d = e^{\frac{E_a(T_d - T_s)}{R \cdot T_d \cdot T_s}} \quad (\text{A9})$$

$$\sum_{d=D_s}^{\hat{D}-1} \Delta d < R_d \leq \sum_{d=D_s}^{\hat{D}} \Delta d \quad (\text{A10})$$

5. Thermal period fixed model (TP)

The difference between GD and TP models are heat accumulation occurs in a fixed time period (D_n). The day of leaf onset is the last day ($\widehat{D_s} + D_n$) when the accumulated heat reaches the

forcing requirement. The start day ($\widehat{D}_s$) of heat accumulation begins in day one and moves one day forward each time to estimate Eq. (A12). Three parameters are involved in DA: the base temperature (T_b), the period length (D_n) and the forcing requirement (R_d).

$$\Delta d = \begin{cases} T_d - T_b & \text{if } T_d > T_b \\ 0 & \text{otherwise} \end{cases} \quad (\text{A11})$$

$$R_d \leq \sum_{d=\widehat{D}_s}^{\widehat{D}_s+D_n} \Delta d \quad (\text{A12})$$

6. Chilling and forcing (CF)

Compared to GD, there is another distinctive chilling period for dormancy. CF model sequentially calculates two accumulations in opposite directions: chilling accumulation and anti-chilling accumulation. The start day of chilling accumulation (D_s) is implicitly set as 273.0 which is October 1st. The end day of chilling accumulation (D_0) is the beginning of anti-chilling accumulation. Three parameters are considered in DA: the chilling requirement (R_d^C) and the forcing requirement (R_d^F), the temperature threshold (T_c).

$$\Delta d = \begin{cases} T_d - T_c & \text{if } T_d \geq 0 \\ -T_c & \text{otherwise} \end{cases} \quad (\text{A13})$$

$$\Delta_d^C = \begin{cases} \Delta d & \text{if } \Delta d < 0 \\ 0 & \text{otherwise} \end{cases} \quad (\text{A14})$$

$$\Delta_d^F = \begin{cases} \Delta d & \text{if } \Delta d > 0 \\ 0 & \text{otherwise} \end{cases} \quad (\text{A15})$$

$$\sum_{d=D_s}^{D_0-1} \Delta_d^C > R_d^C \geq \sum_{d=D_s}^{D_0} \Delta_d^C \quad (\text{A16})$$

$$\sum_{d=\widehat{D}_0}^{\widehat{D}-1} \Delta_d^F < R_d^F \leq \sum_{d=\widehat{D}_0}^{\widehat{D}} \Delta_d^F \quad (\text{A17})$$

7. Sequential model (SM)

The difference between CF and SM models is that SM used a beta function (Eq. A18) for the calculation of chilling accumulation and adopted a sigmoid function (Eq. A20) for anti-chilling accumulation. The detailed descriptions of these two functions can be referred to the

introductions of BF model and CF model. The maximum temperature is empirically set as 13.7695. Six parameters are constrained in DA: minimum temperature (T_n), optimal temperature (T_o), temperature sensitivity (S_t), forcing base temperature (T_b), chilling requirement (R_d^C), and forcing requirement (R_d^F).

$$r_d = \left(\frac{T_x - T_d}{T_x - T_o}\right) \left(\frac{T_d - T_n}{T_o - T_n}\right)^{\frac{T_o - T_n}{T_x - T_o}} \quad (\text{A18})$$

$$\Delta_d^C = \begin{cases} r_d & \text{if } r_d < 0 \\ 0 & \text{otherwise} \end{cases} \quad (\text{A19})$$

$$\Delta_d^F = \frac{1}{1 + e^{S_t(T_d - T_b)}} \quad (\text{A20})$$

$$\sum_{d=D_s}^{D_0-1} \Delta_d^C > R_d^C \geq \sum_{d=D_s}^{D_0} \Delta_d^C \quad (\text{A21})$$

$$\sum_{d=D_0}^{\hat{D}-1} \Delta_d^F < R_d^F \leq \sum_{d=D_0}^{\hat{D}} \Delta_d^F \quad (\text{A22})$$

8. Parallel model (PM)

Critical difference between PM and above two-step models is that the chilling and anti-chilling accumulations happen simultaneously (Fu et al., 2012). In the earlier dates during chilling period, only small fraction (K_d) of forcing (Eq. A25) will be accumulated. The maximum temperature is empirically set as 15.3. Seven parameters will be considered in DA: minimum temperature (T_n), optimal temperature (T_o), temperature sensitivity (S_t), forcing base temperature (T_b), chilling requirement (R_d^C), forcing requirement (R_d^F), and a forcing weight coefficient (K_m).

$$r_d = \left(\frac{T_x - T_d}{T_x - T_o}\right) \left(\frac{T_d - T_n}{T_o - T_n}\right)^{\frac{T_o - T_n}{T_x - T_o}} \quad (\text{A23})$$

$$\Delta_d^C = \begin{cases} r_d & \text{if } r_d < 0 \\ 0 & \text{otherwise} \end{cases} \quad (\text{A24})$$

$$K_d = \begin{cases} K_m + (1 - K_m) \frac{\sum_{i=D_s}^d \Delta_i^C}{R_d^C} & \text{if } \sum_{d=D_s}^{D_0-1} \Delta_d^C > R_d^C \\ 1 & \text{otherwise} \end{cases} \quad (\text{A25})$$

$$\Delta_d^F = \frac{K_d}{1 + e^{S_t(T_d - T_b)}} \quad (\text{A26})$$

$$\sum_{d=D_s}^{D_0-1} \Delta_d^C > R_d^C \geq \sum_{d=D_s}^{D_0} \Delta_d^C \quad (\text{A27})$$

$$\sum_{d=D_0}^{\hat{D}-1} \Delta_d^F < R_d^F \leq \sum_{d=D_0}^{\hat{D}} \Delta_d^F \quad (\text{A28})$$

9. Alternating model (AM)

AM fixes the start date of chilling period (D_s^C) as November 1st and the start date of anti-chilling period (D_s^F) as January 1st. The difference between AM and the other models above is that the forcing requirement is not a parameter value but is decided by the length of chilling days (Fu et al., 2012). Five parameters to be constrained in DA are: chilling temperature (T_c), forcing base temperature (T_b) and three coefficients (a, b, c) in calculation of forcing requirement.

$$\Delta_d^C = \begin{cases} 1 & \text{if } T_d \leq T_c \\ 0 & \text{otherwise} \end{cases} \quad (\text{A29})$$

$$\Delta_d^F = \begin{cases} T_d - T_b & \text{if } T_d > T_b \\ 0 & \text{otherwise} \end{cases} \quad (\text{A30})$$

$$R_d^C = \sum_{i=D_s^C}^d \Delta_i^C \quad (\text{A31})$$

$$R_d^F = a + b \cdot e^{-c \cdot R_d^C} \quad (\text{A32})$$

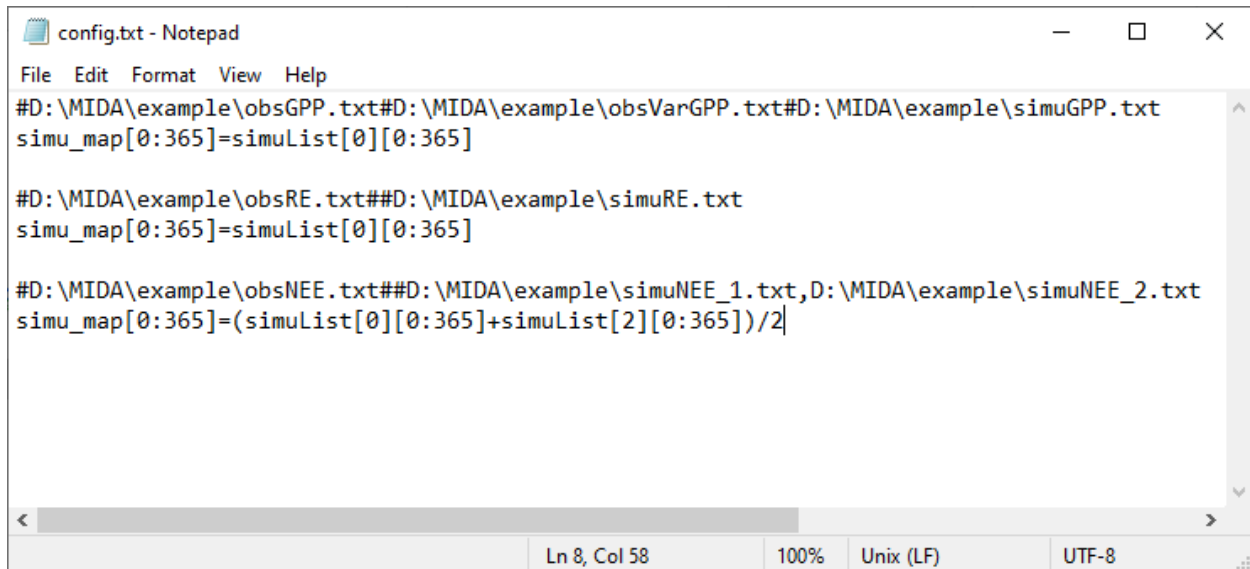
$$\sum_{d=D_s^F}^{\hat{D}-1} \Delta_d^F < R_d^F \leq \sum_{d=D_s^F}^{\hat{D}} \Delta_d^F \quad (\text{A33})$$

Table A1: A summary of parameters to be calibrated in nine phenological models. Their default parameter value and prior parameter range are shown.

Model	Parameter	Description	Unit	Default	Range
GD	T_b	Base temperature	°C	10	[-5, 25]
	R_d	Forcing requirement	°Cd	35	[0, 200]
SF	T_b	Base temperature	°C	-1.5	[-10, 25]
	R_d	Forcing requirement	°C	50	[0, 500]
BF	T_o	Optimal temperature	°C	15	[10, 35]
	T_n	Minimum temperature	°C	0	[-10, 5]
	R_d	Forcing requirement	°Cd	11	[0, 50]
DTS	E_a	Temperature sensitivity rate	-	250	[1, 1500]
	T_s	Standard temperature	°C	10	[-30, 40]
	R_d	Forcing requirement	°Cd	50	[1, 200]
TP	T_b	Base temperature	°C	12.5	[0, 30]
	D_n	Period length	d	25	[0, 50]
	R_d	Forcing requirement	°Cd	20	[0, 150]
CF	R_d^C	Chilling requirement	°Cd	-124	[-300, 0]
	R_d^F	Forcing requirement	°Cd	120	[0, 300]
	T_c	Chilling base temperature	°C	5	[0, 30]
SM	T_n	Minimum temperature	°C	-20	[-80, 0]
	T_o	Optimal temperature	°C	0	[-26, 10]
	S_t	Temperature sensitivity	-	-1.8	[-5, 0]
	T_b	Forcing base temperature	°C	5	[-5, 35]
	R_d^C	Chilling requirement	°Cd	20	[0, 80]
	R_d^F	Forcing requirement	°Cd	20	[0, 80]
PM	T_n	Minimum temperature	°C	-20	[-80, 0]
	T_o	Optimal temperature	°C	0	[-26, 10]
	S_t	Temperature sensitivity	-	-0.6	[-1, 0]
	T_b	Forcing base temperature	°C	5	[-5, 35]
	R_d^C	Chilling requirement	°Cd	11.35	[0, 80]
	R_d^F	Forcing requirement	°Cd	44.01	[0, 80]
	K_m	Forcing weight coefficient	-	0.2	[0, 1]
AM	T_c	Chilling base temperature	°C	4.6	[-10, 10]
	T_b	Forcing base temperature	°C	5	[-5, 35]
	a	Coefficient for forcing adjustment	-	11.51	[0.01, 15]
	b	Coefficient for forcing adjustment	-	88	[0, 200]
	c	Coefficient for forcing adjustment	-	-0.01	[-1, -10^{-4}]

Appendix B: An example of output configuration file

Output configuration file (e.g., config.txt) is to indicate the directories of observations and simulation output files as well as how they map to each other. Figure B1 is an example of the output configuration file. There are three blocks of functions to map simulation outputs to observed GPP, RE, and NEE. The blocks of mapping functions are separated by a blank line. Each mapping block starts with the directories of one observation, its observation variance and model outputs, which are separated by a hash key. If there is no observation variance available, users can ignore this directory. If multiple simulation outputs are used to correspond to one observation, the directories of simulation outputs are separated by a comma. The rest of the mapping block describes how to map simulation outputs to observations. The `simu_map` variable is simulation output after mapping. The `simuList` variable saves the simulation outputs specified in the first line. Taking the third mapping block in Fig. B1 as an example, `simuList[0]` saves contents in `simuNEE_1.txt` and `simuList[0][0:365]` saves the first 365 elements in this file.



```
config.txt - Notepad
File Edit Format View Help
#D:\MIDA\example\obsGPP.txt#D:\MIDA\example\obsVarGPP.txt#D:\MIDA\example\simuGPP.txt
simu_map[0:365]=simuList[0][0:365]

#D:\MIDA\example\obsRE.txt##D:\MIDA\example\simuRE.txt
simu_map[0:365]=simuList[0][0:365]

#D:\MIDA\example\obsNEE.txt##D:\MIDA\example\simuNEE_1.txt,D:\MIDA\example\simuNEE_2.txt
simu_map[0:365]=(simuList[0][0:365]+simuList[2][0:365])/2

Ln 8, Col 58 100% Unix (LF) UTF-8
```

Figure B1: An example of output configuration file

Appendix C: An example of the namelist.txt file

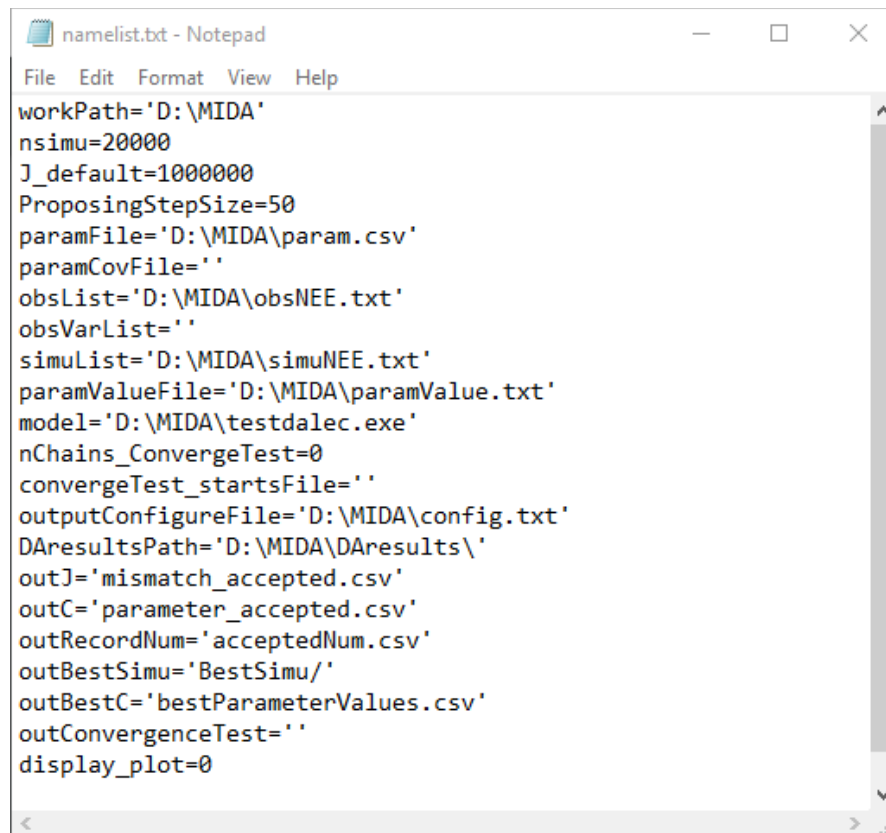
Fig. C1 shows an example of the namelist.txt for the first study case with the DALEC model.

Users need to prepare the namelist.txt before execution of data assimilation (DA) either manually or via GUI. Below describes the content in the namelist.txt. Detailed explanation or tutorials are available in the Zenodo repositories at the end of the appendixes.

‘workpath’ is the directory where the MIDA executable are saved. ‘nsimu’ is the number of iterations in execution of data assimilation. ‘J_default’ is the default mismatch (i.e., cost function) to be compared in the first moving phase of data assimilation. ‘ProposingStepSize’ controls the jump scale in the proposing phase of data assimilation. Users can increase or decrease this value to adjust the acceptance rate to be in a range from 0.2 to 0.5. ‘paramFile’ is the directory of a csv file saving parameter-related information such as parameter range. ‘obsList’ saves the directories of observations. Multiple observations are separated by semicolon. Similarly, ‘obsVarList’ saves the directories of observation variance in the same order as that of ‘obsList’. ‘simuList’ saves the directories of simulation outputs corresponding to the observations. With GUI, MIDA reads directories in the output configuration file (e.g., config.txt) which users provide and assign values for ‘obsList’, ‘obsVarList’, and ‘simuList’ in the namelist.txt automatically. In this case, if the directories of observations change, users only need to modify the output configuration file and generate the namelist.txt again with GUI-based MIDA.

‘paramValue’ is the directory of a txt file where MIDA writes out new set of parameter values for model execution in each iteration of data assimilation. Its default value is ‘ParameterValue.txt’ under the workpath specified in the first line of the namelist.txt. ‘model’ saves the directory of model executable. ‘nChains_convergeTest’ indicates whether to conduct

German-Rubin (G-R) convergence test or not. If G-R test is used, its value is the number of multiple MCMC chains. If not, its value is zero. ‘convergeTest_startsFile’ is the directory of a csv file saving default parameter values as the start points in multiple MCMC chains. ‘outConvergenceTest’ saves the results of G-R test. If ‘nChains_ConvergeTest’ is zero, both values of ‘convergeTest_startsFile’ and ‘outConvergenceTest’ are empty. ‘DAresultsPath’ is the directory saving the results of DA whose directories are also listed in the following six lines: ‘outJ’ for the accepted mismatches; ‘outC’ for the accepted parameter values; ‘outRecordNum’ for the number of accepted parameter values; ‘outBestSimu’ for the best simulation outputs with the optimal parameter values; ‘outBestC’ for the optimal parameter values. For MIDA without GUI, ‘display_plot’ indicates whether or not to visualize the posterior distributions after DA.



```
workPath='D:\MIDA'
nsimu=20000
J_default=1000000
ProposingStepSize=50
paramFile='D:\MIDA\param.csv'
paramCovFile=''
obsList='D:\MIDA\obsNEE.txt'
obsVarList=''
simuList='D:\MIDA\simuNEE.txt'
paramValueFile='D:\MIDA\paramValue.txt'
model='D:\MIDA\testdalec.exe'
nChains_ConvergeTest=0
convergeTest_startsFile=''
outputConfigureFile='D:\MIDA\config.txt'
DAresultsPath='D:\MIDA\DAresults\'
outJ='mismatch_accepted.csv'
outC='parameter_accepted.csv'
outRecordNum='acceptedNum.csv'
outBestSimu='BestSimu/'
outBestC='bestParameterValues.csv'
outConvergenceTest=''
display_plot=0
```

Figure C1. An example of the ‘namelist.txt’ file. In order to use MIDA, users need to prepare data and a model and specify their file names and directories in the ‘namelist.txt’ file.

Code and data availability. The code of MIDA is available at the Zenodo repository <https://doi.org/10.5281/zenodo.4762725> (last access: May 2021). Data used in this study are available at <https://doi.org/10.5281/zenodo.4762779>. A comparison of the time cost using the embedded DA algorithm and MIDA is available at the Zenodo repository <https://doi.org/10.5281/zenodo.4891319>.

Video supplement. Tutorial videos of how to use MIDA is available at <https://doi.org/10.5281/zenodo.4762777>

Author contributions. XH, IS, and YL designed the study. XH built the workflow of MIDA and tested its capability in four cases. DL, DMR, and PJH provided data and model for the first and second test cases. XL prepared models and ADR provided observations for the third case. EW and SN helped to prepare data and model for the fourth case. XH, LJ, EH and YL analyzed the results. All authors contributed to the preparation of the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work was funded by subcontract 4000158404 from Oak Ridge National Laboratory (ORNL) to the Northern Arizona University. ORNL is managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract DE-AC05-00OR22725.

References

- Allen, J. I., Eknes, M. and Evensen, G.: An Ensemble Kalman Filter with a complex marine ecosystem model: hindcasting phytoplankton in the Cretan Sea, *Ann. Geophys.*, 21(1), 399–411, doi:10.5194/angeo-21-399-2003, 2003.
- Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R. and Avellano, A.: The data assimilation research testbed a community facility, *Bull. Am. Meteorol. Soc.*, 90(9), 1283–1296, doi:10.1175/2009BAMS2618.1, 2009.
- Bloom, A. A., Exbrayat, J. F., Van Der Velde, I. R., Feng, L. and Williams, M.: The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times, *Proc. Natl. Acad. Sci.*, 113(5), 1285–1290, doi:10.1073/pnas.1515160113, 2016.
- Bonan, G.: *Climate Change and Terrestrial Ecosystem Modeling*, Cambridge University Press., 2019.
- Box, G. E. P. and Tiao, G. C.: *Bayesian Inference in Statistical Analysis*, John Wiley & Sons, Inc., Hoboken, NJ, USA., 1992.
- Ciais, P., Chris, S., Govindasamy, B., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., Defries, R., Galloway, J. and Heimann, M.: Carbon and other biogeochemical cycles, *Clim. Chang.* 2013 *Phys. Sci. Basis*, 465–570, 2013.
- Cline, M. P., Lomow, G. and Girou, M.: *C++ FAQs*, Pearson Education., 1998.
- Doherty, J.: *PEST: Model independent parameter estimation. Fifth edition of user manual*, Watermark Numer. Comput., doi:10.1016/B978-0-08-098288-5.00031-2, 2004.
- Evensen, G.: The Ensemble Kalman Filter: Theoretical formulation and practical implementation, *Ocean Dyn.*, 53(4), 343–367, doi:10.1007/s10236-003-0036-9, 2003.
- Fer, I., Kelly, R., Moorcroft, P. R., Richardson, A. D., Cowdery, E. M. and Dietze, M. C.: Linking big models to big data: Efficient ecosystem model calibration through Bayesian model emulation, *Biogeosciences*, 15(19), 5801–5830, doi:10.5194/bg-15-5801-2018, 2018.
- Fox, A., Williams, M., Richardson, A. D., Cameron, D., Gove, J. H., Quaife, T., Ricciuto, D., Reichstein, M., Tomelleri, E., Trudinger, C. M. and Van Wijk, M. T.: The REFLEX project: Comparing different algorithms and implementations for the inversion of a terrestrial ecosystem model against eddy covariance data, *Agric. For. Meteorol.*, 149(10), 1597–1615, doi:10.1016/j.agrformet.2009.05.002, 2009.
- Fox, A. M., Hoar, T. J., Anderson, J. L., Arellano, A. F., Smith, W. K., Litvak, M. E., MacBean, N., Schimel, D. S. and Moore, D. J. P.: Evaluation of a Data Assimilation System for Land Surface Models Using CLM4.5, *J. Adv. Model. Earth Syst.*, 10(10), 2471–2494, doi:10.1029/2018MS001362, 2018.
- Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., Cadule, P., Doney, S., Eby, M., Fung, I., Bala, G., John, J., Jones, C., Joos, F., Kato, T., Kawamiya, M., Knorr, W., Lindsay, K., Matthews, H. D., Raddatz, T., Rayner, P., Reick, C., Roeckner, E., Schnitzler, K.-G., Schnur, R., Strassmann, K., Weaver, A. J., Yoshikawa, C. and Zeng, N.: Climate–Carbon Cycle Feedback Analysis: Results from the C4MIP Model Intercomparison, *J. Clim.*, 19(14), 3337–3353, doi:10.1175/JCLI3800.1, 2006.
- Fu, Y. H., Campioli, M., Van Oijen, M., Deckmyn, G. and Janssens, I. A.: Bayesian comparison of six different temperature-based budburst models for four temperate tree species, *Ecol. Modell.*, 230, 92–100, doi:10.1016/j.ecolmodel.2012.01.010, 2012.

- Gao, C., Wang, H., Weng, E., Lakshmivarahan, S., Zhang, Y. and Luo, Y.: Assimilation of multiple data sets with the ensemble Kalman filter to improve forecasts of forest carbon dynamics, *Ecol. Appl.*, 21(5), 1461–1473, doi:10.1890/09-1234.1, 2011.
- Gelman, A. and Rubin, D. B.: Inference from Iterative Simulation Using Multiple Sequences, *Stat. Sci.*, 7(4), 457–472, doi:10.1214/SS/1177011136, 1992.
- Hanson, P. J., Riggs, J. S., Nettles, W. R., Phillips, J. R., Krassovski, M. B., Hook, L. A., Gu, L., Richardson, A. D., Aubrecht, D. M., Ricciuto, D. M., Warren, J. M. and Barbier, C.: Attaining whole-ecosystem warming using air and deep-soil heating methods with an elevated CO₂ atmosphere, *Biogeosciences*, 14(4), 861–883, doi:10.5194/bg-14-861-2017, 2017.
- Hararuk, O., Xia, J. and Luo, Y.: Evaluation and improvement of a global land model against soil carbon data using a Bayesian Markov chain Monte Carlo method, *J. Geophys. Res. Biogeosciences*, 119(3), 403–417, doi:10.1002/2013JG002535, 2014.
- Hararuk, O., Smith, M. J. and Luo, Y.: Microbial models with data-driven parameters predict stronger soil carbon responses to climate change, *Glob. Chang. Biol.*, 21(6), 2439–2453, doi:10.1111/gcb.12827, 2015.
- Hastings, W. K.: Monte carlo sampling methods using Markov chains and their applications, *Biometrika*, 57(1), 97–109, doi:10.1093/biomet/57.1.97, 1970.
- Hou, E., Lu, X., Jiang, L., Wen, D. and Luo, Y.: Quantifying Soil Phosphorus Dynamics: A Data Assimilation Approach, *J. Geophys. Res. Biogeosciences*, 124(7), 2159–2173, doi:10.1029/2018JG004903, 2019.
- Ise, T. and Moorcroft, P. R.: The global-scale temperature and moisture dependencies of soil organic carbon decomposition: An analysis using a mechanistic decomposition model, *Biogeochemistry*, 80(3), 217–231, doi:10.1007/s10533-006-9019-5, 2006.
- Iversen, C. M., McCormack, M. L., Powell, A. S., Blackwood, C. B., Freschet, G. T., Kattge, J., Roumet, C., Stover, D. B., Soudzilovskaia, N. A., Valverde-Barrantes, O. J., van Bodegom, P. M. and Violle, C.: A global Fine-Root Ecology Database to address below-ground challenges in plant ecology, *New Phytol.*, 215(1), 15–26, doi:10.1111/nph.14486, 2017.
- Jiang, J., Huang, Y., Ma, S., Stacy, M., Shi, Z., Ricciuto, D. M., Hanson, P. J. and Luo, Y.: Forecasting Responses of a Northern Peatland Carbon Cycle to Elevated CO₂ and a Gradient of Experimental Warming, *J. Geophys. Res. Biogeosciences*, 123(3), 1057–1071, doi:10.1002/2017JG004040, 2018.
- Kattge, J., Bönsch, G., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., Tautenhahn, S., Werner, G. D. A., Aakala, T., Abedi, M., Acosta, A. T. R., Adamidis, G. C., Adamson, K., Aiba, M., Albert, C. H., Alcántara, J. M., Alcázar, C. C., Aleixo, I., Ali, H., Amiaud, B., Ammer, C., Amoroso, M. M., Anand, M., Anderson, C., Anten, N., Antos, J., Apgaua, D. M. G., Ashman, T. L., Asmara, D. H., Asner, G. P., Aspinwall, M., Atkin, O., Aubin, I., Bastrup-Spohr, L., Bahalkeh, K., Bahn, M., Baker, T., Baker, W. J., Bakker, J. P., Baldocchi, D., Baltzer, J., Banerjee, A., Baranger, A., Barlow, J., Barneche, D. R., Baruch, Z., Bastianelli, D., Battles, J., Bauerle, W., Bauters, M., Bazzato, E., Beckmann, M., Beeckman, H., Beierkuhnlein, C., Bekker, R., Belfry, G., Belluau, M., Beloiu, M., Benavides, R., Benomar, L., Berdugo-Lattke, M. L., Berenguer, E., Bergamin, R., Bergmann, J., Bergmann Carlucci, M., Berner, L., Bernhardt-Römermann, M., Bigler, C., Bjorkman, A. D., Blackman, C., Blanco, C., Blonder, B., Blumenthal, D., Bocanegra-González, K. T., Boeckx, P., Bohlman, S., Böhning-Gaese, K., Boisvert-Marsh, L., Bond,

- W., Bond-Lamberty, B., Boom, A., Boonman, C. C. F., Bordin, K., Boughton, E. H., Boukili, V., Bowman, D. M. J. S., Bravo, S., Brendel, M. R., Broadley, M. R., Brown, K. A., Bruelheide, H., Brumnich, F., Bruun, H. H., Bruy, D., Buchanan, S. W., Bucher, S. F., Buchmann, N., Buitenwerf, R., Bunker, D. E., et al.: TRY plant trait database – enhanced coverage and open access, *Glob. Chang. Biol.*, 26(1), 119–188, doi:10.1111/gcb.14904, 2020.
- De Kauwe, M. G., Medlyn, B. E., Walker, A. P., Zaehle, S., Asao, S., Guenet, B., Harper, A. B., Hickler, T., Jain, A. K., Luo, Y., Lu, X., Luus, K., Parton, W. J., Shu, S., Wang, Y. P., Werner, C., Xia, J., Pendall, E., Morgan, J. A., Ryan, E. M., Carrillo, Y., Dijkstra, F. A., Zelikova, T. J. and Norby, R. J.: Challenging terrestrial biosphere models with data from the long-term multifactor Prairie Heating and CO₂ Enrichment experiment, *Glob. Chang. Biol.*, 23(9), 3623–3645, doi:10.1111/gcb.13643, 2017.
- Keenan, T. F., Davidson, E., Moffat, A. M., Munger, W. and Richardson, A. D.: Using model-data fusion to interpret past trends, and quantify uncertainties in future projections, of terrestrial ecosystem carbon cycling, *Glob. Chang. Biol.*, 18(8), 2555–2569, doi:10.1111/j.1365-2486.2012.02684.x, 2012.
- Keenan, T. F., Davidson, E. A., Munger, J. W. and Richardson, A. D.: Rate my data: Quantifying the value of ecological data for the development of models of the terrestrial carbon cycle, *Ecol. Appl.*, 23(1), 273–286, doi:10.1890/12-0747.1, 2013.
- Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., Collier, N., Ghimire, B., van Kampenhout, L., Kennedy, D., Kluzek, E., Lawrence, P. J., Li, F., Li, H., Lombardozzi, D., Riley, W. J., Sacks, W. J., Shi, M., Vertenstein, M., Wieder, W. R., Xu, C., Ali, A. A., Badger, A. M., Bisht, G., van den Broeke, M., Brunke, M. A., Burns, S. P., Buzan, J., Clark, M., Craig, A., Dahlin, K., Drewniak, B., Fisher, J. B., Flanner, M., Fox, A. M., Gentine, P., Hoffman, F., Keppel-Aleks, G., Knox, R., Kumar, S., Lenaerts, J., Leung, L. R., Lipscomb, W. H., Lu, Y., Pandey, A., Pelletier, J. D., Perket, J., Randerson, J. T., Ricciuto, D. M., Sanderson, B. M., Slater, A., Subin, Z. M., Tang, J., Thomas, R. Q., Val Martin, M. and Zeng, X.: The Community Land Model Version 5: Description of New Features, Benchmarking, and Impact of Forcing Uncertainty, *J. Adv. Model. Earth Syst.*, 11(12), 4245–4287, doi:10.1029/2018MS001583, 2019.
- LeBauer, D. S., Wang, D., Richter, K. T., Davidson, C. C. and Dietze, M. C.: Facilitating feedbacks between field measurements and ecosystem models, *Ecol. Monogr.*, 83(2), 133–154, doi:10.1890/12-0137.1, 2013.
- Levenberg, K.: A method for the solution of certain non-linear problems in least squares, *Q. Appl. Math.*, 2(2), 164–168, 1944.
- Li, Q., Lu, X., Wang, Y., Huang, X., Cox, P. M. and Luo, Y.: Leaf area index identified as a major source of variability in modeled CO₂ fertilization, *Biogeosciences*, 15(22), 6909–6925, doi:10.5194/bg-15-6909-2018, 2018.
- Liang, J., Zhou, Z., Huo, C., Shi, Z., Cole, J. R., Huang, L., Konstantinidis, K. T., Li, X., Liu, B., Luo, Z., Penton, C. R., Schuur, E. A. G., Tiedje, J. M., Wang, Y. P., Wu, L., Xia, J., Zhou, J. and Luo, Y.: More replenishment than priming loss of soil organic carbon with additional carbon input, *Nat. Commun.*, 9(1), 1–9, doi:10.1038/s41467-018-05667-7, 2018a.
- Liang, J., Zhou, Z., Huo, C., Shi, Z., Cole, J. R., Huang, L., Konstantinidis, K. T., Li, X., Liu, B., Luo, Z., Penton, C. R., Schuur, E. A. G., Tiedje, J. M., Wang, Y., Wu, L. and Xia, J.:

- organic carbon with additional carbon input, *Nat. Commun.*, 1–9, doi:10.1038/s41467-018-05667-7, 2018b.
- Lu, D., Ricciuto, D., Walker, A., Safta, C. and Munger, W.: Bayesian calibration of terrestrial ecosystem models: A study of advanced Markov chain Monte Carlo methods, *Biogeosciences*, 14(18), 4295–4314, doi:10.5194/bg-14-4295-2017, 2017.
- Lu, D., Ricciuto, D., Stoyanov, M. and Gu, L.: Calibration of the E3SM Land Model Using Surrogate-Based Global Optimization, *J. Adv. Model. Earth Syst.*, 10(6), 1337–1356, doi:10.1002/2017MS001134, 2018.
- Luo, Y. and Schuur, E. A. G.: Model parameterization to represent processes at unresolved scales and changing properties of evolving systems, *Glob. Chang. Biol.*, 26(3), 1109–1117, doi:10.1111/gcb.14939, 2020.
- Luo, Y., Wu, L., Andrews, J. A., White, L., Matamala, R., Schäfer, K. V. R. and Schlesinger, W. H.: ELEVATED CO₂ DIFFERENTIATES ECOSYSTEM CARBON PROCESSES: DECONVOLUTION ANALYSIS OF DUKE FOREST FACE DATA, *Ecol. Monogr.*, 71(3), 357–376, doi:10.1890/0012-9615(2001)071[0357:ECDECP]2.0.CO;2, 2001.
- Luo, Y., Ogle, K., Tucker, C., Fei, S., Gao, C., LaDeau, S., Clark, J. S. and Schimel, D. S.: Ecological forecasting and data assimilation in a data-rich era, *Ecol. Appl.*, 21(5), 1429–1442, doi:10.1890/09-1275.1, 2011.
- Ma, S., Jiang, J., Huang, Y., Shi, Z., Wilson, R. M., Ricciuto, D., Sebestyen, S. D., Hanson, P. J. and Luo, Y.: Data-Constrained Projections of Methane Fluxes in a Northern Minnesota Peatland in Response to Elevated CO₂ and Warming, *J. Geophys. Res. Biogeosciences*, 122(11), 2841–2861, doi:10.1002/2017JG003932, 2017.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E.: Equation of state calculations by fast computing machines, *J. Chem. Phys.*, 21(6), 1087–1092, doi:10.1063/1.1699114, 1953.
- Mitchell, J. C. and Apt, K.: Concepts in programming languages, Cambridge University Press., 2003.
- Nerger, L. and Hiller, W.: Software for ensemble-based data assimilation systems- Implementation strategies and scalability, *Comput. Geosci.*, 55, 110–118, doi:10.1016/j.cageo.2012.03.026, 2013.
- Van Oijen, M., Cameron, D. R., Butterbach-Bahl, K., Farahbakhshazad, N., Jansson, P. E., Kiese, R., Rahn, K. H., Werner, C. and Yeluripati, J. B.: A Bayesian framework for model calibration, comparison and analysis: Application to four models for the biogeochemistry of a Norway spruce forest, *Agric. For. Meteorol.*, 151(12), 1609–1621, doi:10.1016/j.agrformet.2011.06.017, 2011.
- Ono, S. and Konno, T.: Estimation of flowering date and temperature characteristics of fruit trees by DTS method, *Japan Agric. Res. Q.*, 33(2), 105–108, 1999.
- Raeder, K., Anderson, J. L., Collins, N., Hoar, T. J., Kay, J. E., Lauritzen, P. H. and Pincus, R.: DART/CAM: An ensemble data assimilation system for CESM atmospheric models, *J. Clim.*, 25(18), 6304–6317, doi:10.1175/JCLI-D-11-00395.1, 2012.
- Raupach, M. R., Rayner, P. J., Barrett, D. J., Defries, R. S., Heimann, M., Ojima, D. S., Quegan, S. and Schimmlus, C. C.: Model-data synthesis in terrestrial carbon observation: Methods, data requirements and data uncertainty specifications, *Glob. Chang. Biol.*, 11(3), 378–397, doi:10.1111/j.1365-2486.2005.00917.x, 2005.
- Rayner, P. J., Scholze, M., Knorr, W., Kaminski, T., Giering, R. and Widmann, H.: Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS),

- Global Biogeochem. Cycles, 19(2), n/a-n/a, doi:10.1029/2004GB002254, 2005.
- Ricciuto, D., Sargsyan, K. and Thornton, P.: The Impact of Parametric Uncertainties on Biogeochemistry in the E3SM Land Model, *J. Adv. Model. Earth Syst.*, 10(2), 297–319, doi:10.1002/2017MS000962, 2018.
- Ricciuto, D. M., King, A. W., Dragoni, D. and Post, W. M.: Parameter and prediction uncertainty in an optimized terrestrial carbon cycle model: Effects of constraining variables and data record length, *J. Geophys. Res.*, 116(G1), G01033, doi:10.1029/2010JG001400, 2011.
- Richardson, A. D., Williams, M., Hollinger, D. Y., Moore, D. J. P., Dail, D. B., Davidson, E. A., Scott, N. A., Evans, R. S., Hughes, H., Lee, J. T., Rodrigues, C. and Savage, K.: Estimating parameters of a forest ecosystem C model with measurements of stocks and fluxes as joint constraints, *Oecologia*, 164(1), 25–40, doi:10.1007/s00442-010-1628-y, 2010.
- Richardson, A. D., Hufkens, K., Milliman, T., Aubrecht, D. M., Chen, M., Gray, J. M., Johnston, M. R., Keenan, T. F., Klosterman, S. T., Kosmala, M., Melaas, E. K., Friedl, M. A. and Frolking, S.: Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery, *Sci. Data*, 5, 1–24, doi:10.1038/sdata.2018.28, 2018.
- Ridler, M. E., Van Velzen, N., Hummel, S., Sandholt, I., Falk, A. K., Heemink, A. and Madsen, H.: Data assimilation framework: Linking an open data assimilation library (OpenDA) to a widely adopted model interface (OpenMI), *Environ. Model. Softw.*, 57, 76–89, doi:10.1016/j.envsoft.2014.02.008, 2014.
- Robert, C. and Casella, G.: Monte Carlo statistical methods, Springer Science & Business Media., 2013.
- Roberts, G. O., Gelman, A. and Gilks, W. R.: Weak convergence and optimal scaling of random walk Metropolis algorithms, *Ann. Appl. Probab.*, 7(1), 110–120, doi:10.1214/AOAP/1034625254, 1997.
- Safta, C., Ricciuto, D. M., Sargsyan, K., Debusschere, B., Najm, H. N., Williams, M. and Thornton, P. E.: Global sensitivity analysis, probabilistic calibration, and predictive assessment for the data assimilation linked ecosystem carbon model, *Geosci. Model Dev.*, 8(7), 1899–1918, doi:10.5194/gmd-8-1899-2015, 2015.
- Scholze, M., Kaminski, T., Rayner, P., Knorr, W. and Giering, R.: Propagating uncertainty through prognostic carbon cycle data assimilation system simulations, *J. Geophys. Res.*, 112(D17), D17305, doi:10.1029/2007JD008642, 2007.
- Shi, Z., Crowell, S., Luo, Y. and Moore, B.: Model structures amplify uncertainty in predicted soil carbon responses to climate change, *Nat. Commun.*, 9(1), 1–11, doi:10.1038/s41467-018-04526-9, 2018.
- Smith, M. J., Purves, D. W., Vanderwel, M. C., Lyutsarev, V. and Emmott, S.: The climate dependence of the terrestrial carbon cycle, including parameter and structural uncertainties, *Biogeosciences*, 10(1), 583–606, doi:10.5194/bg-10-583-2013, 2013.
- Strigul, N., Pristinski, D., Purves, D., Dushoff, J. and Pacala, S.: Scaling from trees to forests: Tractable macroscopic equations for forest dynamics, *Ecol. Monogr.*, 78(4), 523–545, doi:10.1890/08-0082.1, 2008.
- Tao, F., Zhou, Z., Huang, Y., Li, Q., Lu, X., Ma, S., Huang, X., Liang, Y., Hugelius, G., Jiang, L., Dougherty, R., Ren, Z. and Luo, Y.: Deep Learning Optimizes Data-Driven Representation of Soil Organic Carbon in Earth System Model Over the Conterminous United States, *Front. Big Data*, 3(June), 1–15, doi:10.3389/fdata.2020.00017, 2020.
- Trudinger, C. M., Raupach, M. R., Rayner, P. J., Kattge, J., Liu, Q., Park, B., Reichstein, M.,

- Renzullo, L., Richardson, A. D., Roxburgh, S. H., Styles, J., Wang, Y. P., Briggs, P., Barrett, D. and Nikolova, S.: OptIC project: An intercomparison of optimization techniques for parameter estimation in terrestrial biogeochemical models, *J. Geophys. Res. Biogeosciences*, 112(2), 1–17, doi:10.1029/2006JG000367, 2007.
- Wang, Y. P., Trudinger, C. M. and Enting, I. G.: A review of applications of model-data fusion to studies of terrestrial carbon fluxes at different scales, *Agric. For. Meteorol.*, 149(11), 1829–1842, doi:10.1016/j.agrformet.2009.07.009, 2009.
- Weng, E. and Luo, Y.: Relative information contributions of model vs. data to short- and long-term forecasts of forest carbon dynamics, *Ecol. Appl.*, 21(5), 1490–1505, doi:10.1890/09-1394.1, 2011.
- Weng, E., Dybzinski, R., Farrior, C. E. and Pacala, S. W.: Competition alters predicted forest carbon cycle responses to nitrogen availability and elevated CO₂: simulations using an explicitly competitive, game-theoretic vegetation demographic model, *Biogeosciences Discuss.*, 1–35, doi:10.5194/bg-2019-55, 2019.
- Williams, M., Schwarz, P. A., Law, B. E., Irvine, J. and Kurpius, M. R.: An improved analysis of forest carbon dynamics using data assimilation, *Glob. Chang. Biol.*, 11(1), 89–105, doi:10.1111/j.1365-2486.2004.00891.x, 2005.
- Williams, M., Richardson, A. D., Reichstein, M., Stoy, P. C., Peylin, P., Verbeeck, H., Carvalhais, N., Jung, M., Hollinger, D. Y., Kattge, J., Leuning, R., Luo, Y., Tomelleri, E., Trudinger, C. M. and Wang, Y. P.: Improving land surface models with FLUXNET data, *Biogeosciences*, 6(7), 1341–1359, doi:10.5194/bg-6-1341-2009, 2009.
- Xu, T., White, L., Hui, D. and Luo, Y.: Probabilistic inversion of a terrestrial ecosystem model: Analysis of uncertainty in parameter estimation and model prediction, *Global Biogeochem. Cycles*, 20(2), 1–15, doi:10.1029/2005GB002468, 2006.
- Yun, K., Hsiao, J., Jung, M. P., Choi, I. T., Glenn, D. M., Shim, K. M. and Kim, S. H.: Can a multi-model ensemble improve phenology predictions for climate change studies?, *Ecol. Modell.*, 362, 54–64, doi:10.1016/j.ecolmodel.2017.08.003, 2017.
- Zobitz, J. M., Desai, A. R., Moore, D. J. P. and Chadwick, M. A.: A primer for data assimilation with ecological models using Markov Chain Monte Carlo (MCMC), *Oecologia*, 167(3), 599–611, doi:10.1007/s00442-011-2107-9, 2011.

Tables

Table1: Comparison among MIDA and available DA tools

DA tool	Agnostic	DA algorithms	Global optima	Posterior distribution	Visualization
CCDAS	No	Automatic differentiation from Transformation of Algorithms in Fortran (TAF)	No	No	No
CARDAMOM	No	Markov Chain Monte Carlo	Yes	Yes	No
EcoPAD	No	Markov Chain Monte Carlo	Yes	Yes	Yes
OpenDA	No	EnKF, Ensemble Square-Root Filter, Particle Filter	Yes	Yes	No
DART	Yes	EnKF	Yes	Yes	No
PDAF	Yes	EnKF	Yes	Yes	No
PEST	Yes	Levenberg-Marquardt method	Rely on initial parameter values	No	No
MIDA	Yes	Markov Chain Monte Carlo	Yes	Yes	Yes

Table 2: A summary of 21 parameters to be calibrated in DALEC model. The default parameter value and prior parameter range are shown.

Parameter	Description	Unit	Default	Range
GDD_{min}	Growing degree day threshold for leaf out	$^{\circ}C\ d$	100	[10, 250]
GDD_{max}	Growing degree day threshold for maximum LAI	$^{\circ}C\ d$	200	[50, 500]
LAI_{max}	Seasonal maximum leaf area index	-	4	[2, 7]
$T_{leaffall}$	Temperature for leaf fall	$^{\circ}C$	5	[0, 10]
K_{leaf}	Rate of leaf fall	d^{-1}	0.1	[0.03, 0.95]
NUE	N use efficiency	-	7	[1, 20]
Res_{growth}	Growth respiration fraction	-	0.2	[0.05, 0.5]
Res_m	Base rate for maintenance respiration	$\times 10^{-4}\ \mu mol\ m^{-2}d^{-1}$	1	[0.1, 100]
Q_{10mr}	Maintenance respiration T-sensitivity	-	2	[1, 4]
A_{stem}	Allocation to plant stem pool	-	0.7	[0.1, 0.95]
τ_{root}	Root turnover time	$\times 10^{-4}\ d^{-1}$	5.48	[1.1, 27.4]
τ_{stem}	Stem turnover time	$\times 10^{-5}\ d^{-1}$	5.48	[1.1, 27.4]
Q_{10hr}	Heterotrophic respiration T-sensitivity	-	2	[1, 4]
τ_{litter}	Base turnover for litter	$\times 10^{-3}\ \mu mol\ m^{-2}d^{-1}$	1.37	[0.548, 5.48]
τ_{som}	Base turnover for soil organic matter	$\times 10^{-4}\ \mu mol\ m^{-2}d^{-1}$	9.13	[0.274, 2.74]
K_{decomp}	Decomposition rate	$\times 10^{-3}\ d^{-1}$	1	[0.1, 10]
LMA	Leaf mass per area	$gC\ m^{-2}$	80	[20, 150]
$X_{stem_{init}}$	Initial value for stem C pool	$\times 10^3\ gC$	5	[1, 15]
$X_{root_{init}}$	Initial value for root C pool	gC	500	[100, 3000]
$X_{litter_{init}}$	Initial value for litter C pool	gC	600	[50, 1000]
$X_{som_{init}}$	Initial value for soil organic C pool	$\times 10^3\ gC$	7	[1, 25]

Table 3: A summary of eight parameters to be calibrated in surrogate-based ELM model. The default parameter value and prior parameter range are shown.

Parameter	Description	Unit	Default	Range
c_{root}	Rooting depth distribution parameter	m^{-1}	2.0	[0.5, 4]
SLA_{top}	Specific leaf area at canopy top	$m^2 gC^{-1}$	0.03	[0.01, 0.05]
N_{leaf}	Fraction of leaf N in RuBisCO	-	0.1007	[0.1, 0.4]
CN_{root}	Fine root C:N ratio	-	42	[25, 60]
A_{r2l}	Allocation ratio of fine root to leaf	-	1.0	[0.3, 1.5]
Res_m	Base rate for maintenance respiration	$\times 10^{-6} \mu mol m^{-2} s^{-1}$	2.525	[1.5, 4]
$t_{leaf fall}$	Critical day length for senescence	$\times 10^4 s$	3.93	[3.5, 4.5]
GDD_{onset}	Accumulated growing degree days for leaf out	$^{\circ}C d$	800	[600, 1000]

Table 4: A summary of two parameters to be calibrated in the BiomE model. The default parameter value and prior parameter range are shown.

Parameter	Description	Unit	Default	Range
V_{annual}	Annual productivity per unit leaf area	$kgC\ y^{-1}m^2$	0.4	[0.2, 2]
M_{canopy}	Annual mortality rate in canopy layer	y^{-1}	0.02	[0.01, 0.08]

Figures

Figure 1: The three-step workflow of Model Independent Data Assimilation (MIDA) module.

The workflow includes data preparation, execution of data assimilation (DA), and visualization.

The data preparation step is to provide all the formatted essential data for DA via user input. The execution step is to calibrate parameter values towards a constrained posterior distribution with the fusion of observations. The visualization step is to diagnose the effects of DA. Rhombus in orange represents user-input data. Rectangle represents procedures and document/multidocument shape is for data files in computers. Dashed lines indicate locations of data. Solids lines indicate data flow pathways. With the three-step workflow, DA is agnostic to specific models and users will be released from technical burdens.

Figure 2: the GUI-MIDA window includes two panels. The upper panel is to set up a data assimilation task. Inputs can be loaded and applied to the step 1 on data preparation for DA. The lower panel is to run DA as described in step 2 and visualize the posterior distributions of parameters in step 3.

Figure 3: Comparison between the simulated daily net ecosystem exchange (NEE) by DALEC and the observed NEE at Harvard Forest from 1992 to 2006. Red circles represent modeled NEE with the optimized parameter values and green circles represent simulated NEE with the original parameter values. Simulations of DALEC are substantially improved after data assimilation in comparison with those before data assimilation.

Figure 4: Comparison between posterior distributions (red line) and default values (gray dash line) of the 21 parameters in DALEC. The peak in posterior distribution is the constrained parameter value from maximum likelihood estimation. This distinctive mode and its divergence

from the default value indicates the effects of DA. Most parameters are well constrained, and some are far different from the original values.

Figure 5: Comparison between posterior distributions (red line) and default values (gray dash line) of the eight parameters in surrogate-based ELM. The peak in posterior distribution is the constrained parameter value from maximum likelihood estimation. This distinctive mode and its divergence from the default value indicates the effects of DA. Most parameters are well constrained, and some are far different from the original values.

Figure 6: Comparison between the simulated NEE, total leaf area index, latent heat flux by surrogate-based ELM and the observed ones at Missouri Ozark flux site from 2006 to 2014. The blue lines indicate the observations, and their 95% confidence interval is in the dashed area. The green and red lines indicate the simulations with default parameter values and optimized values respectively. Simulations are generally improved after DA for all these three variables.

Figure 7: Comparison between the simulated growth date by 9 phenology models after DA and the observed growth date for *Larix laricina* with +9°C treatment at SPRUCE site from 2016 to 2018. Colored numbers indicate different models and shape represents different year. Overall, model 6,7,8,9 achieve a better performance after DA.

Figure 8: Comparison between posterior distributions (red line) and default values (gray dash line) of the two parameters in BiomeE. The peak in posterior distribution is the constrained parameter value from maximum likelihood estimation. This distinctive mode and its divergence from the default value indicates the effects of DA. All parameters are well constrained and different from their original values.

Figure 9: Comparison between the simulated leaf area index (LAI) by BiomeE and the observed NEE at Willow Creek. Circles represent modeled NEE with the optimized parameter values and

triangles represent simulated NEE with the original parameter values. Simulations of LAI are substantially improved after data assimilation in comparison with those before data assimilation.

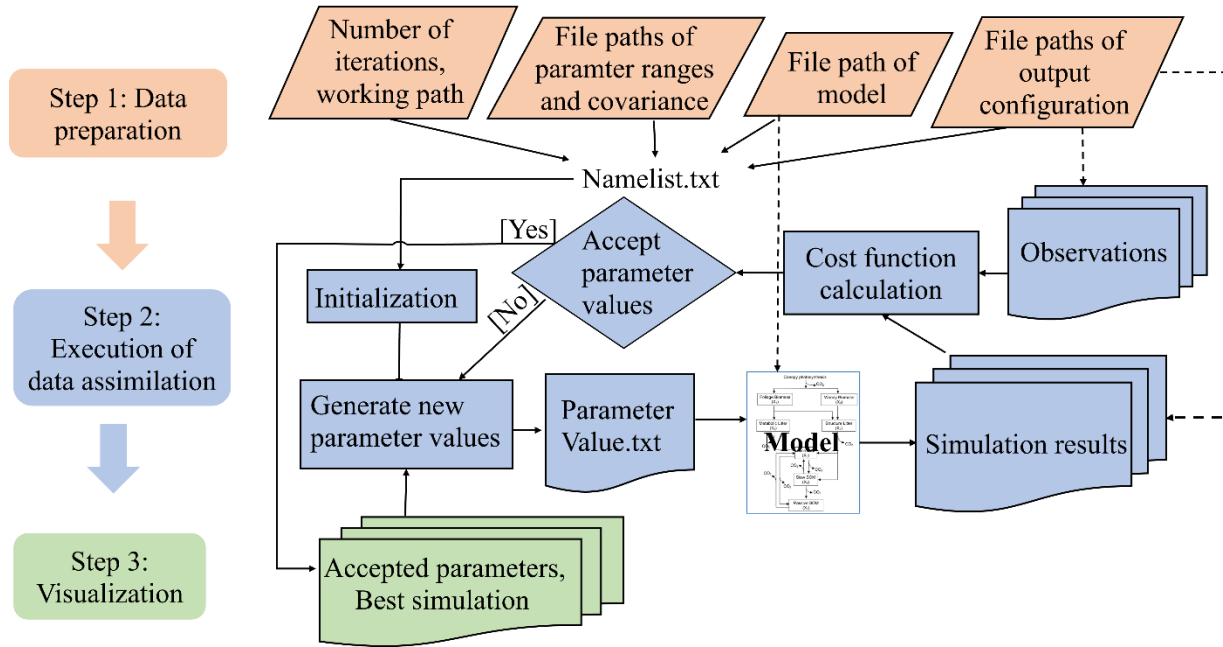


Figure 1: The three-step workflow of Model Independent Data Assimilation (MIDA) module.

The workflow includes data preparation, execution of data assimilation (DA), and visualization.

The data preparation step is to provide all the formatted essential data for DA via user input. The execution step is to calibrate parameter values towards a constrained posterior distribution with the fusion of observations. The visualization step is to diagnose the effects of DA. Rhombus in orange represents user-input data. Rectangle represents procedures and document/multidocument shape is for data files in computers. Dashed lines indicate locations of data. Solids lines indicate data flow pathways. With the three-step workflow, DA is agnostic to specific models and users will be released from technical burdens.

DAmodule - A Generic Module for Data Assimilation

Help
—
□
×

Preparation of Data Assimilation

The number of simulations
Select Work Path

Load Parameter Range

	min	max	default
1			
2			
3			
4			
5			
6			

(Optional) Load Parameter Covariance

Load Model Executable File

Observation File List

	file name
1	
2	
3	
4	
5	
6	

Observation Variance File List

	file name
1	
2	
3	
4	
5	
6	

Simulation Output File List

	file name
1	
2	
3	
4	
5	
6	

(Optional) Gelman-Rubin convergence test

Execution of Data Assimilation

Load Namelist File:

Choose variables to be print in DA:
☒ total mismatch
 ☒ acceptance rate
 ☐ delta_mismatch
 ☐ mismatch for each obs
 ☐ obs var

Figure 2: the GUI-MIDA window includes two panels. The upper panel is to set up a data assimilation task. Inputs can be loaded and applied to the step 1 on data preparation for DA. The lower panel is to run DA as described in step 2 and visualize the posterior distributions of parameters in step 3.

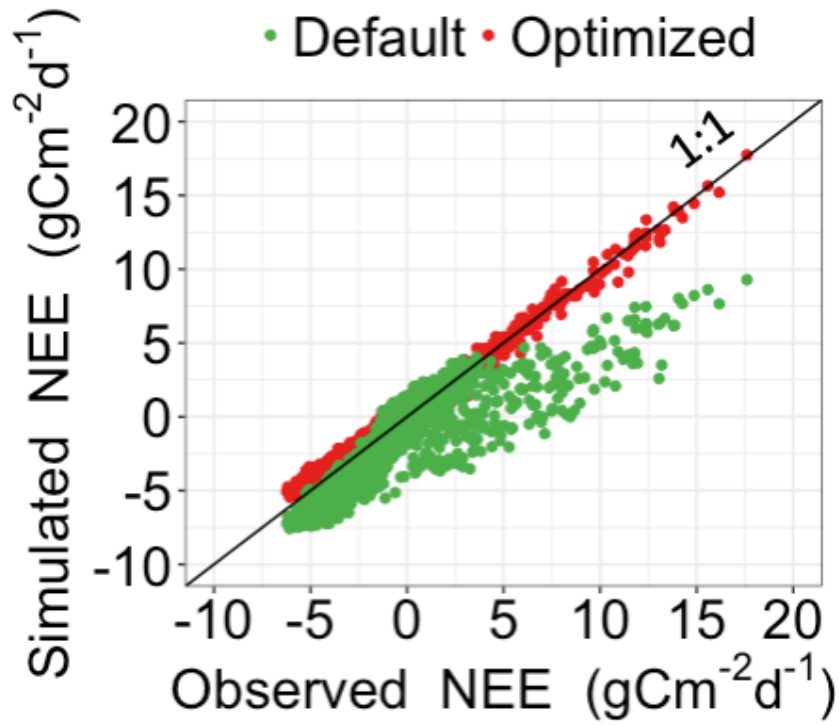


Figure 3: Comparison between the simulated daily net ecosystem exchange (NEE) by DALEC and the observed NEE at Harvard Forest from 1992 to 2006. Red circles represent modeled NEE with the optimized parameter values and green circles represent simulated NEE with the original parameter values. Simulations of DALEC are substantially improved after data assimilation in comparison with those before data assimilation.

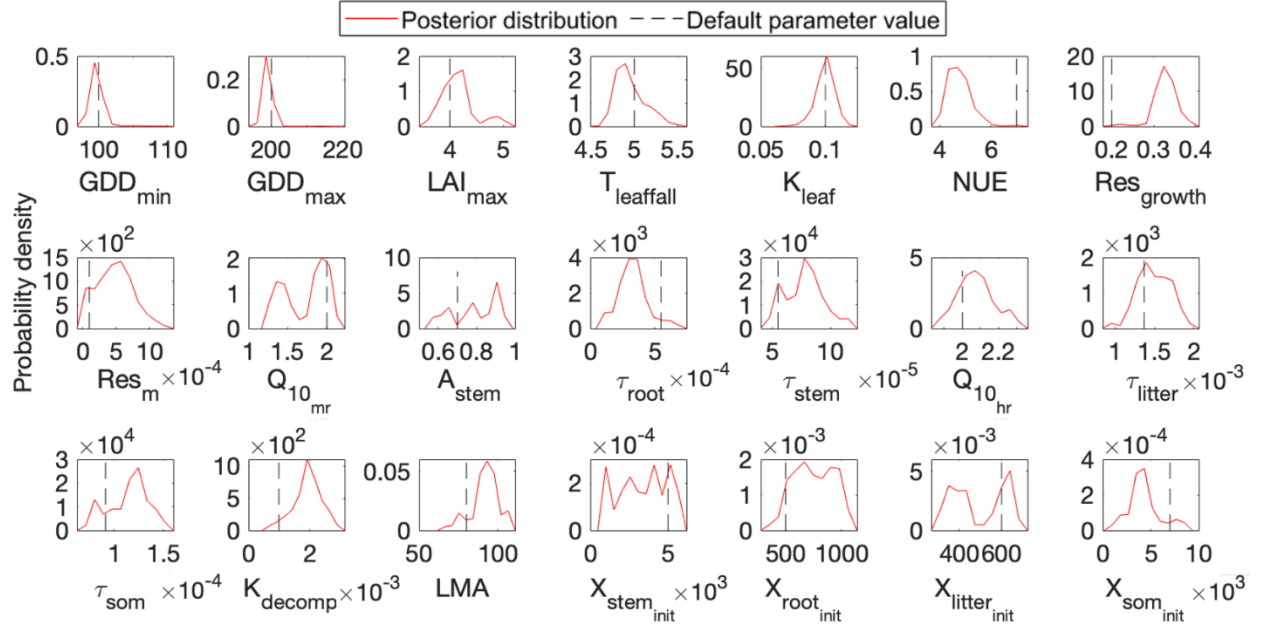


Figure 4: Comparison between posterior distributions (red line) and default values (gray dash line) of the 21 parameters in DALEC. The peak in posterior distribution is the constrained parameter value from maximum likelihood estimation. This distinctive mode and its divergence from the default value indicates the effects of DA. Most parameters are well constrained, and some are far different from the original values.

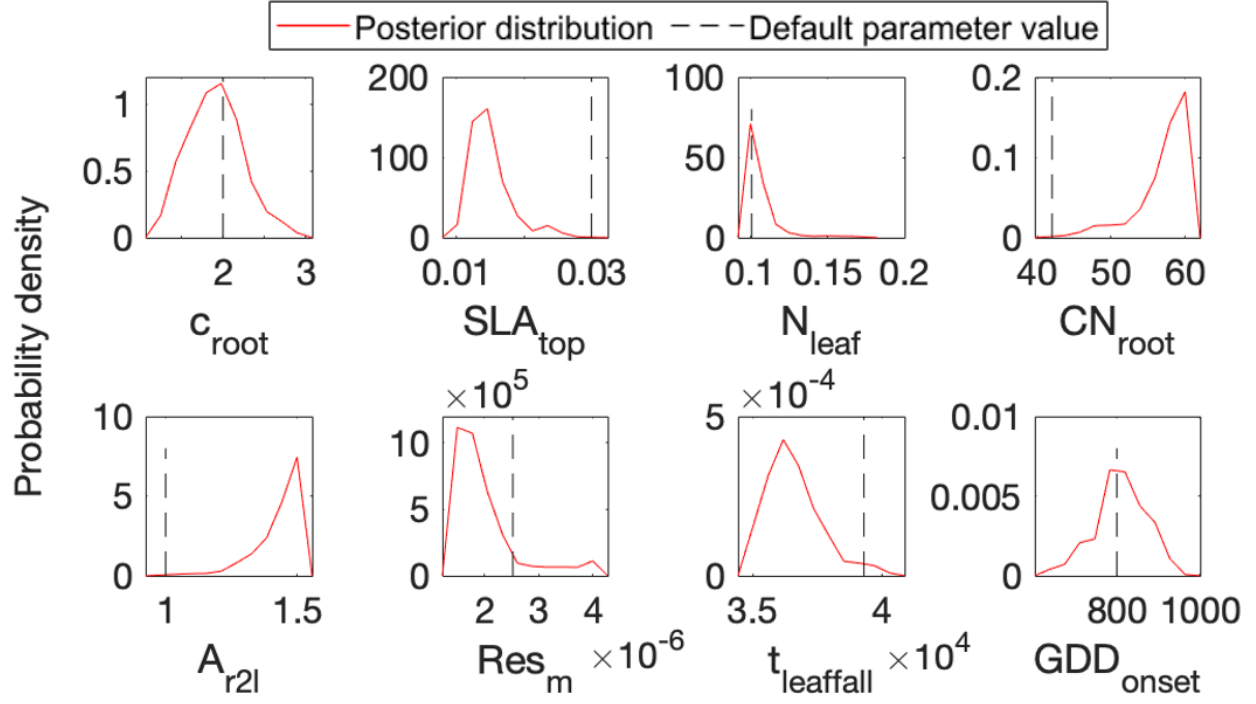


Figure 5: Comparison between posterior distributions (red line) and default values (gray dash line) of the eight parameters in surrogate-based ELM. The peak in posterior distribution is the constrained parameter value from maximum likelihood estimation. This distinctive mode and its divergence from the default value indicates the effects of DA. Most parameters are well constrained, and some are far different from the original values.

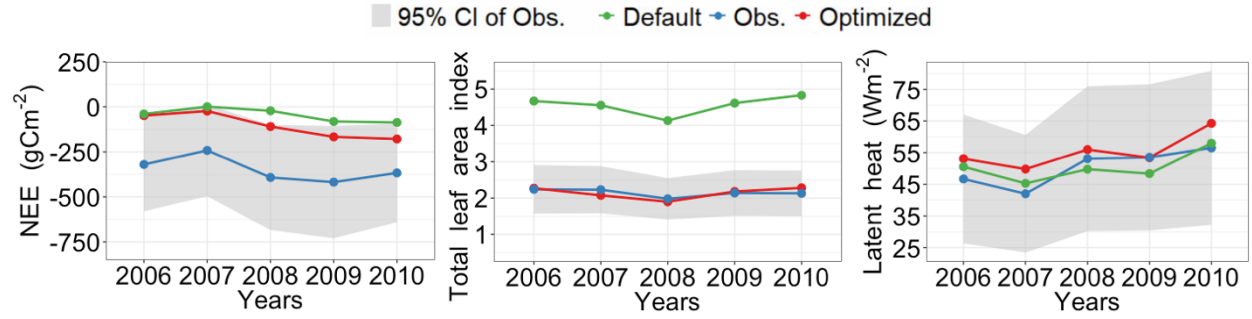


Figure 6: Comparison between the simulated NEE, total leaf area index, latent heat flux by surrogate-based ELM and the observed ones at Missouri Ozark flux site from 2006 to 2014. The blue lines indicate the observations, and their 95% confidence interval is in the dashed area. The green and red lines indicate the simulations with default parameter values and optimized values respectively. Simulations are generally improved after DA for all these three variables.

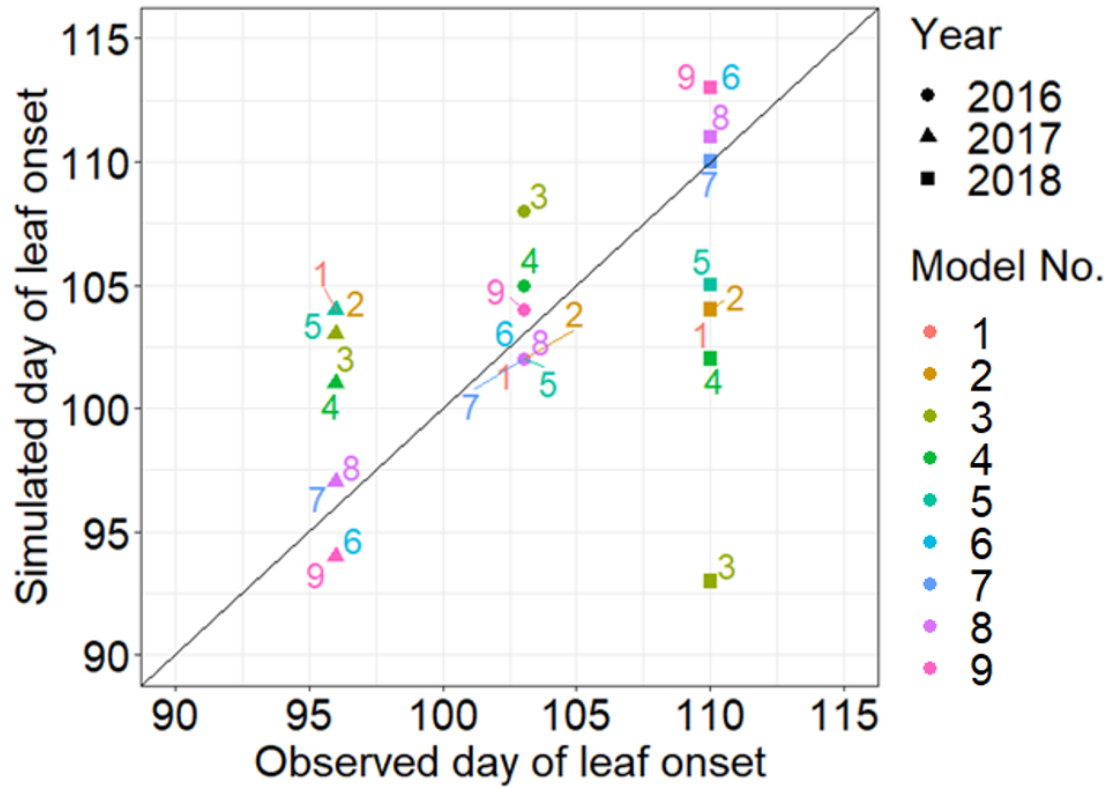


Figure 7: Comparison between the simulated growth date by 9 phenology models after DA and the observed growth date for *Larix laricina* with +9°C treatment at SPRUCE site from 2016 to 2018. Colored numbers indicate different models and shapes represent different years. Overall, model 6,7,8,9 achieve a better performance after DA.

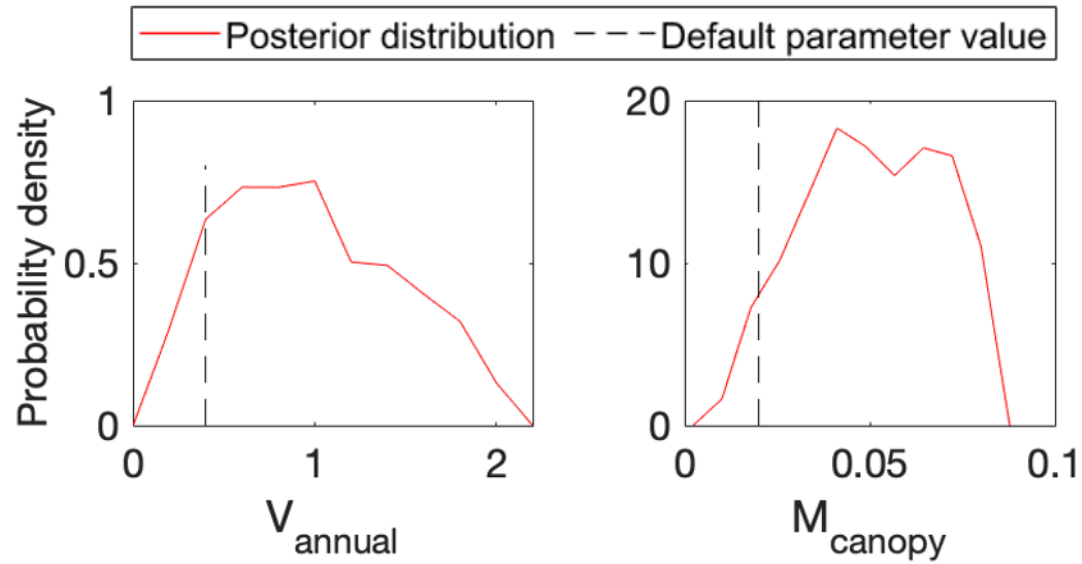


Figure 8: Comparison between posterior distributions (red line) and default values (gray dash line) of the two parameters in BiomeE. The peak in posterior distribution is the constrained parameter value from maximum likelihood estimation. This distinctive mode and its divergence from the default value indicates the effects of DA. All parameters are well constrained and different from their original values.

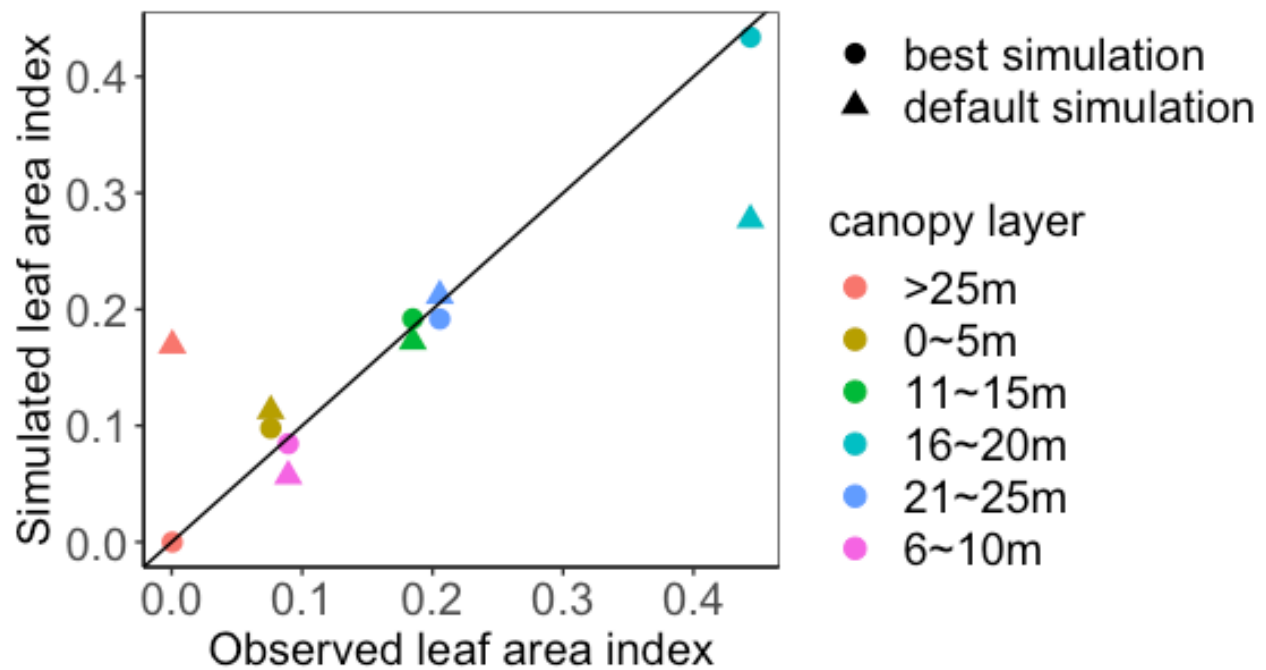


Figure 9: Comparison between the simulated leaf area index (LAI) by BiomeE and the observed NEE at Willow Creek. Circles represent modeled NEE with the optimized parameter values and triangles represent simulated NEE with the original parameter values. Simulations of LAI are substantially improved after data assimilation in comparison with those before data assimilation.

CHAPTER 3: ELEVATED WATER TABLE DEPTH UNDER 8-YEAR WARMING TURNS CARBON SINK TO SOURCE IN PERMAFROST

Xin Huang^{1,2}, Xingjie Lu³, Brian Izbicki^{1,4}, Elaine Pegoraro⁵, Ted Schuur^{1,4}, Lifen Jiang¹, Yiqi Luo^{1,2,4}

1 Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, AZ, USA

2 School of informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ, USA

3 School of Atmospheric Sciences, Sun Yat-Sen University, Guangzhou, China

4 Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ, USA

5 Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Correspondence to: Xin Huang (xh59@nau.edu)

Abstract

Permafrost soils contain an enormous amount of carbon (C) that is vulnerable to be released into the atmosphere under a warming climate. The magnitude of permafrost C release is regulated by both biotic and abiotic changes (e.g., changes in thaw depth (TD) and water table depth (WTD)) after permafrost thaws. The abiotic changes control multiple ecological processes and influence the net carbon balance of permafrost ecosystems. The abiotic changes can lead to a shift between C source or sink of permafrost, depending on their relative impacts on ecosystem productivity and microbial decomposition. However, how changes in TD and WTD control the net carbon balance has not been well examined. In this study, we combined 8-year observations from a soil warming experiment in Alaska with a terrestrial ecosystem model using a data assimilation technique. The calibrated model is used to simulate the C cycle in permafrost in response to changes in TD and WTD and the underlying mechanisms. Our results showed that net ecosystem exchange (NEE) is more sensitive to changes in WTD than those in TD, implying that hydrological dynamics have a significant impact on C cycle after permafrost thaw. Specifically, elevated water table depth (WTD) can shift permafrost from a C sink to a C source regardless of the changes in TD. Models without accounting for subsidence processes will overestimate C sink in permafrost ecosystems.

Keywords: Carbon cycle; data assimilation; permafrost thawing; process-based modeling; water table depth

1. Introduction

Permafrost in high-latitude ecosystems covers only 15% of Earth's total soil area but stores more than 30% of the global surface soil carbon (C) (Schuur et al., 2015; Hugelius et al., 2014).

Current and future warming will induce permafrost thaw and stimulate C loss via microbial decompositions (Schuur et al., 2015, Anon, 2014). Thaw depth (TD) has been identified as a key driver of cumulative C loss (Celis et al., 2017). Increased TD is always accompanied by wetting conditions or elevated water table depth (WTD) (Mauritz et al., 2017; Rodenhizer et al., 2020), especially in high-ice permafrost (Schuur and Mack, 2018a). More and more studies reveal the importance of soil moisture on C cycles. However, the effect of elevated WTD on C cycles has not ever been directly evaluated (Aerts, 2006; Berner and Schädel, 2023; Zona et al., 2023). Therefore, examination of the individual effect of WTD is essential to predict its feedback to C cycle upon permafrost thaw.

Current models usually consider top-down warming in permafrost, while the most dramatic change to tundra ecosystems is caused by abrupt thaw as a result of wildfire or other disturbance. During the abrupt thaw, the loss of soil ice structure will cause the soil surface to slump and create saturated soil conditions, such as subsidence (O'Donnell et al., 2012; Jorgenson et al., 2001). The consequences of permafrost thaw (or increased TD) and subsidence (or elevated WTD) change widely across different landscapes (Schuur and Mack, 2018b). For example, permafrost thaw in uplands may lead to drier soil (or lower WTD) because the steep slope will accelerate water drainage to streams or lakes. Permafrost thaw in lowlands, on the other hand, may result to wetter soil (or elevated WTD). Additionally, there are also environmental factors influencing TD during permafrost thaw. For example, wind and the length of snow season may influence snow accumulation above ground (Schuur et al., 2015). As a

result, TD upon permafrost thaw may become deeper or shallower. TD and WTD together influence C fluxes through plant growth and microbial activities in active layers where C content in permafrost concentrates (Schuur and Mack, 2018b).

Some studies show that tundra is a net C source (Belshe et al., 2013) while others indicate that it is a C sink (McGuire et al., 2012; Celis et al., 2017). The potential drivers of C cycle in permafrost include TD (Mauritz et al., 2017), plant community composition (Schuur et al., 2007), nutrient availability (Salmon et al., 2016), etc. However, limited studies analyze the direct influences of WTD on C cycles in permafrost regions. One possible reason is that TD and WTD are tangled in field measurements. For example, deep thaw (or increased TD) with wetter soil is often witnessed in Alaskan tundra (Rodenhizer et al., 2020). Previous studies have found a nonlinear relationship between TD and C fluxes in a permafrost warming experiment at Alaska, implying the underlying impacts of hydrology changes (Mauritz et al., 2017). Model intercomparison also indicates that the divergence in model simulations from the field measurements is caused by alterations in WTD (i.e., subsidence) and this impact of hydrology changes is not explicitly represented in most ecosystem models (Schadel et al., 2018).

Modeling provides an opportunity to assess the individual responses of C fluxes to WTD with a range of TD to depict different thawing scenarios. In such a way, we can disentangle the effects of TD and WTD on C fluxes. In this study, we use a terrestrial ecosystem model (TECO) constrained by data assimilation to simulate the effects of TD and WTD on C cycles at the Permafrost Experimental Heating Research (CiPEHR) study site in Alaska. The paper is organized as follow: in Section 2, we introduce model, data, and the methods to assess the impacts of TD and WTD on C cycles; results are represented in Section 3; in Section 4, we

discuss the model performance after data assimilation, individual impacts of TD and WTD on C cycles, and potential limitations of this study; the summary is in Section 5.

2. Method

2.1 Study site

The experimental data we used in this modeling study were collected in the Carbon in Permafrost Experimental Heating Research (CiPEHR) site. CiPEHR was initiated in 2008 and was designed to study the impacts of Arctic warming on the ecosystem C exchange by experimentally manipulating soil and air temperatures (Natali et al., 2010). It is located at an upland moist acidic tundra in the Eight Mile Lake Watershed, Alaska, USA (-149.23°W, 63.88°N) with an elevation of 670m. This site is underlain by permafrost (Osterkamp et al., 2009) but is undergoing permafrost degradation. The maximum permafrost thaw depth was approximately 0.5 m in the initial year of 2008 (Natali et al., 2010). The mean annual temperature is -1.68°C from 2009 to 2015 (Healy and McKinley Stations, Western Regional Climate Center and NOAA National Centers for Environmental Information). The vegetation community at this site is sedge (the graminoid *Eriophorum vaginatum*), deciduous shrubs (*Vaccinium uliginosum*), and mosses (Natali et al., 2012).

Air warming treatment is achieved using open top chambers ($0.36m^2 \times 0.5m$ tall). These chambers are set out from 1 May to mid-September to manipulate air temperature. Soil warming treatment is imposed using snow fences ($1.5m$ tall $\times$ $8m$ long) to accumulate snow and insulate the ground to increase soil temperature. Excess snow is removed every April to avoid delayed plant growth and additional water input. There are four treatments: Control, Air warming, Soil warming, and Air & Soil warming. There are four plots ($0.6m \times 0.6m$) on one side of the snow fence to receive soil warming treatment, and another four plots on the other side to receive

control treatment. Furthermore, half of the plots on each side of the snow fence are exposed to air warming treatment. Thus, there are two repetitions for each treatment at one snow fence. Six snow fences are distributed separately across three blocks to account for the impacts of topography. Therefore, each treatment is repeated in 12 plots. (Mauritz et al., 2017).

The air temperature at a height of 15 cm in each plot is automatically measured every 90 minutes with a shaded thermistor. Soil temperature at depths of 5 cm and 10 cm in each plot is automatically measured every 30 mins with type T copper-constantan thermocouples. Other meteorology variables such as precipitation and photosynthetically active radiation (PAR) were measured half-hourly with a HOBO Onset station (Mauritz et al., 2017). Water table depth and thaw depth are measured several times a week during the summer season. Snow depth is measured yearly in April. NEE is measured during the summer season with a closed path infrared gas analyzer (LICOR) connected to chambers. The NEE measurements are gap filled to a half-hourly resolution using a regression curve whose parameters are estimated by site-level PAR (Mauritz et al., 2017). Ecosystem respiration (ER) is calculated by soil temperature in each plot using the equations in Natali et al. (2010, 2015). GPP is estimated by subtracting NEE from ER. Aboveground biomass is measured yearly in July. More information about these measurements is available in Table B1. Overall, meteorological variables such as air temperature and soil temperature are used to drive the simulation of TECO model whereas C fluxes, pools, water table depth, etc. are used in data assimilation for constraining parameters uncertainty of the TECO model.

2.2 Model description and parameters

2.2.1 Overview of the TECO model

The model used in this study is the TECO model (Weng and Luo, 2008; Shi et al., 2016; Huang et al., 2017). The TECO model has four major components: canopy, soil water dynamics, vegetation dynamics, and soil C/nitrogen transfers. Canopy submodule is developed from a two-leaf model to simulate canopy photosynthesis, conductance, transpiration, and energy balance. The simulation of radiation transmission is based on Beer's law. The leaf photosynthesis is simulated using the model by (Farquhar et al., 1980) and the stomatal conductance is based on the model by (Ball et al., 1987). The soil water dynamic submodule has ten layers. The thickness of the soil layers is 10cm and 20cm for the first five layers and the rest five layers, respectively. Water balance is determined by water influx (i.e., precipitation and percolation) and efflux (i.e., evaporation, transpiration, and runoff). Evaporation is controlled by the water content in the surface layer and the evaporative demand of the atmosphere. Transpiration is determined by stomatal conductance, soil water content, and root distribution. When ten layers are all saturated, runoff occurs. The vegetation dynamic submodule simulates plant growth (i.e., the processes of C allocation to leaves, stems, and roots), litter production, and phenology. The soil C/nitrogen transfer submodule is developed from the models in (Yiqi and Reynolds, 1999) and (Barrett, 2002) to simulate C and nitrogen flow from plants to soils through litterfall, decomposition, and mineralization. More details about TECO can be found in (Weng and Luo, 2008; Shi et al., 2016).

2.2.2 New formulation in water-dependent dynamics

We developed and integrated new formulas into the original TECO model to account for snow accumulation, subsidence, and the impacts of subsidence and permafrost thaw on nitrogen (N) and C cycles. Generally, the snow and soil thermal dynamics are developed from the study of Huang et al. (2017) and the water table module is developed from the study of Ma et al. (2017).

Updated water table module with subsidence processes

According to the water table measurements in the study of Mauritz et al. (2017), we assumed that soils below 30 cm are saturated. Therefore, water recharge in the water table module occurs in the top three layers. Eight parameters, Δs_{yr} ($yr=1, 2, 3, \dots, 8$), were used to represent the subsidence rate in each year from 2009 to 2017. The water table depth (z_{wt}) is calculated at hourly time step using Eq.1. The total volume of water (V_{tot}) in different situations (i.e., flooding, dry soil, and other scenarios) is determined separately (Eq. 2). Given V_{tot} , we can calculate a soil moisture scalar to control photosynthesis, plant growth, and the soil C turnover time (Weng and Luo, 2008).

$$z_{wt} = z_{wt} + \sum_{yr} \Delta s_{yr} \quad (1)$$

$$V_{tot} = \begin{cases} \varphi z_{max} + z_{wt} & \text{if } z_{wt} > 0 \text{ or flooding} \\ \max \left[z_{max} \theta_{smin}, \varphi z_{max} + \frac{2z_{wt}(\varphi - \theta_{smin})}{3} \right] & \text{if } z_{wt} < -100 \text{ or dry soil} \\ \varphi z_{max} - \frac{2a_z z_{wt}^2}{3} & \text{if other cases} \end{cases} \quad (2)$$

$$a_z = \frac{\varphi - \theta_{smin}}{Z_{\theta_{smin}}} \quad (3)$$

Where φ uses the value of the maximum water content (ws_{max}) and it is estimated as a parameter. z_{max} is the maximum water table depth, which is 30 cm. θ_{smin} is the minimum soil water content (ws_{min}) available for plant survival with a value of 0.27. a_z is the gradient to account for evaporation influence (Eq. 3). $Z_{\theta_{smin}}$ is the maximum depth where evaporation can influence soil moisture, which is given a value of 100 mm.

Impacts of subsidence on C cycles

We incorporated the impacts of flooding ($z_{wt} > z_c$, z_c is a threshold parameter) by multiplying an impact factor (α_{fd}) on net photosynthesis rate (A_{day}) and leaf respiration (A_{night}) in the stomato-photosynthesis-transpiration module (Weng and Luo, 2008) as shown in Eqs. 4

and 5. Similarly, we used the impact factor to plant growth and allocations (G_p , Eq. 6) (Shevliakova et al., 2009). The value of α_{fd} is 1 when $z_{wt} < z_c$, while its value needs to be estimated as a parameter when $z_{wt} > z_c$.

$$A_{day} = \alpha_{fd} A_{day} \quad (4)$$

$$A_{night} = \alpha_{fd} A_{night} \quad (5)$$

$$G_p = \alpha_{fd} G_p \quad (6)$$

Updated snow dynamics with snow accumulation

The six snow fences in CiPEHR were used to accumulate snow during winter seasons. This snow accumulation increased snow input rate (I_s) by a certain degree (β_{si} , as a parameter) as shown in Eq. 7. This I_s was then used to simulate soil temperature by the soil thermal module and further had an impact on soil respiration (Huang et al., 2017).

$$I_s = \beta_{si} I_s \quad (7)$$

Impacts of increasing thaw depth on C/N cycles

Thaw depth (TD) is initialized as 150 cm (Mauritz et al., 2017) and is updated in an hourly time step. Its calculation depends on soil depth (D_i) and soil temperature ($T_{s,i}$) at the i -th layer ($i=1, 2, \dots, 10$) (Eq. 8). If $T_{s,i}$ at the first three layers are all below 0 °C, TD will keep constant. Otherwise, the calculation of Eq. 8 will start from the top layer to the bottom to find the maximum TD in permafrost thaw. Furthermore, the increasing TD influences heterotrophic respiration (R_h) and nitrogen immobilization (N_{active}) with a certain impact factor (γ_{td} , as a parameter) (Eqs. 9 and 10).

$$\text{If } T_{s,i} > 0 \text{ and } T_{s,i+1} < 0, TD = D_i + tk_i \min \left[\frac{T_{s,i}}{T_{s,i} - T_{s,i+1}}, \max \left[0, \frac{\log_6^1 - \log_6 \frac{T_{s,i+1}}{T_{s,i+1} - T_{s,i}}}{\log 3} \right] \right] \quad (8)$$

$$R_h = \gamma_{td} R_h \quad (9)$$

$$N_{active} = \gamma_{td} N_{active} \quad (10)$$

Where tk_i is the thickness of the i -th layer, that is, 10 cm for the first five layers and 20 cm for the last five layers.

Overall, there are 12 parameters introduced (Table 1). Their parameter values and uncertainties need to be estimated through data assimilation.

2.3 Data assimilation

Data assimilation is a technique to assimilate observational data into models to constrain parameter ranges and estimate parameter uncertainty. In this study, we used a model-independent data assimilation module (MIDA) (Huang et al., 2021) and the embedded Markov chain Monte Carlo (MCMC) method (Box and Tiao, 1992) for data assimilation. MCMC is based on Bayesian theory that the posterior distribution of model parameters ($p(\theta|O)$), given the observations (O), is determined by the product of the likelihood function ($p(O|\theta)$) and the prior distribution ($p(\theta)$) (Eq. 11).

$$p(\theta|O) \propto p(O|\theta)p(\theta) \quad (11)$$

12 parameters (Table 1) are selected for data assimilation. We assume that these parameters have a uniform prior distribution and the mismatches between model simulation and observation independently follow a Gaussian distribution with a mean of zero. Consequently, the likelihood function can be structured as follows:

$$p(O|\theta) \propto \exp \left\{ - \sum_{i=1}^7 \sum_{t \in O_i} \frac{(O_i(t) - X(t))^2}{2\sigma_i^2(t)} \right\} \quad (12)$$

Where $O_i(t)$ is the i -th of the 7 observations at time t , σ_i is the standard deviation of the i -th observation, and $X(t)$ is the corresponding simulation output variable at time t . To quantify the mismatch between $O_i(t)$ and $X(t)$, both the observations and simulation outputs were rescaled to the same time scale before comparison. The data flow is summarized in Fig. 1.

MIDA utilizes a Metropolis-Hasting algorithm to sample the posterior distribution of 12 parameters. Generally, this algorithm repeatedly generates a set of parameter values from the prior range based on the previously accepted parameter values. This set of parameter values would be accepted if it achieves a reduction in the mismatch between observations and the corresponding model simulation or through a chance selection with a possibility of 0.05. After 2,000 iterations, all accepted parameter values will be used to estimate the posterior distribution. The optimal parameter value is the average of the last 50 accepted parameter values.

As described in section 2.1, there are six snow fences, and each fence has eight plots. Within a fence, each of the four treatments is applied to two plots. We apply data assimilation to every treatment plot (48 in total) with 8-year observations from 2009 to 2016 (Table B1). To evaluate the performance of data assimilation, we select NEE, water table depth (WTD), and thaw depth (TD) as a representative, and compare these observations with corresponding simulation outputs with the optimal parameter value. Considering the visualization effect, different variables are rescaled to different time scales. In figs 2, 3, and 4, monthly NEE, daily TD, and yearly mean WTD are presented. In addition to visual comparison, we use R^2 and RMSE to quantify the consistency of model outputs and observations in each plot. We also generated the posterior distribution of the 48 R^2 and RMSEs to evaluate the overall performance of data assimilation across the 48 plots.

2.4 Assessment of the responses of C fluxes to changes in TD and WTD

According to the observations at the CiPEHR site, soil warming and air & soil warming brought in deeper permafrost thaw while air warming and control plots had similarly shallow thaw. Moreover, deep permafrost thaw often coincides with elevated water table depth (WTD) (i.e., wetter soil) (Mauritz et al., 2017). A model calibrated with observations at CiPEHR would

generate a comparable pattern between thaw depth (TD) and WTD. To gain a complete understanding of the impacts of water table depth, additional simulations should be performed to take into account warming and drying threats. Specifically, we need more simulations about shallow TD with elevated WTD and deep TD with lowered WTD.

After data assimilation, we generated 48 sets of optimal parameter values, and these parameter values imply different treatment effects of the 48 plots. For example, the eight subsidence rates (Δs_i , $i=1, 2, \dots, 8$) in soil warming plots represent the effect of wetting soil. Therefore, we use the 48 sets of parameter values to generate the two other simulation scenarios (i.e., shallow thaw with wet soil (scenario 1) and deep thaw with a drier soil (scenario 2)). With the values of all these eight subsidence rates set to zero, the TECO model can simulate the effect of dry soil /lowered WTD. Using the parameter values estimated from air warming and control plots, the TECO model can capture the effects of shallow thaw. For scenario 1, we use the optimized parameter values after data assimilation in all plots and set the eight subsidence rates to zero. For scenario 2, we use the eight subsidence rates from soil warming or air & soil warming plots and combine these with the other parameter values from air warming or control plots.

Based on the expanded simulation results which cover all relationships between TD and WTD, we visualize the responses of NEE, GPP, and ER to changes in TD and WTD respectively. In visualizing the responses of the three variables to TD, we select simulations with elevated WTD (0 – 0.15 m) and lowered WTD (-0.4 – -0.25 m) under shallow thaw (-0.38 – -0.6 m) or deep thaw (-0.95 – -1.18 m). By separating deep thaw and shallow thaw, we can identify the impacts of WTD isolated from that from TD. Similarly, when visualizing responses of WTD, we analyze the effect of TD under both elevated and lowered WTD.

3. Results

3.1 Optimized simulations by data assimilation

The simulated NEE captured the seasonal patterns in the observed NEE well (Fig. 2) for all four treatments, with R^2 ranging from 0.1 to 0.7. More than half of the 48 plots have a R^2 value greater than 0.37 (Fig. A1). The RMSE of simulated NEE in most plots centers around 25 (Fig. A2). The worst model data fittings occurred in the air warming plot (Rep 1) at fence 2, air & soil warming plot (Rep 1) at fence 6, and air warming plot (Rep 1) at fence 1. A possible reason is that the simulations fail to depict the net sources in the first three years. The net source capacity of simulated NEE in winter seasons is similar to that in observations for most plots. Except for 4 plots with air warming, most simulated NEE in summer seasons is underestimated or comparable with observation. The underestimation in simulated NEE is partly due to that simulations fail to capture the increasing sink capacity with warming, especially in some soil warming plots. Most simulated NEE can reproduce the variability of observed sink capacity. For example, in air warming (Rep 2) at fence 1, the simulated sink volume increases over years but decreases in the last year, which reflects the same pattern in observation. Generally, the control (mean $R^2=0.42$) and soil warming plots (mean $R^2=0.45$) have a better model data fitting compared to air warming (mean $R^2=0.38$) and air & soil warming (mean $R^2=0.32$) plots. Although NEE from different plots varies substantially, the soil and air & soil warming plots have a higher sink capacity in the summer and the control plots have a lower sink capacity of NEE. This variation in NEE among different treatment plots is well displayed in simulations. For most plots with soil warming, the sink capacity in the final two years declines significantly and some plots even turn into a net source (e.g., soil warming (Rep 2) at fence 5 and fence 6). This transition is also reflected in simulated NEE over years at soil warming plots.

Both TD (R^2 : 0.55 – 0.86; RMSE: 0.13) and WTD (R^2 : 0.1 – 0.96; RMSE: 0.075) have a higher model data fitting performance than NEE for most plots (Figs. 3, 4, and A1). The relatively poor fitting performance of TD is in control (mean $R^2=0.72$) and air warming (mean $R^2=0.74$) plots while the better fitting performance of TD occurs in soil warming plots with a mean R^2 of 0.76. Generally, the control and air warming treatments have a shallow TD while soil warming and air & soil warming plots have a deep thaw. Simulated TD can reproduce the development of deeper thaw over time, especially in soil warming and air & soil warming plots.

For most plots, the discrepancy between simulated WTD and observed WTD can be quite insignificant in certain years. The R^2 of simulated WTD in most plots is close to 0.75 (Fig. A1). Similarly to TD, simulated WTD in soil warming (mean $R^2=0.74$) and air & soil warming plots (mean $R^2=0.7$) have a better model data fitting performance than control (mean $R^2=0.67$) and air warming plots (mean $R^2=0.65$). Simulated WTD in most plots tends to be underestimated in the first three but ended with being overestimated compared with observations. WTD in almost all plots is observed to have an increasing trend over years and this increment is more significant in soil warming and air & soil warming plots. This divergence in increment of WTD across different treatment plots is also shown in simulations.

3.2 Extended simulations of TD and WTD

Figure 5 shows the concentration of combined TD and WTD in observations (a) and extended simulations (b). Deep thaw with wet soil and shallow thaw with dry soil are more commonly observed at the CiPEHR site (Fig. 5a). Using the optimized parameter values after data assimilation from different plots/treatments, we can extend the value space of simulated TD and WTD (Fig. 5b). For example, observed WTD crowds in the range of [-0.3, -0.15], but this range is broadened to [-0.45, 0] in extended simulations. Moreover, Fig. 5b filled up the empty grids in

bottom right (shallow TD with elevated WTD) and top left (deep TD with lowered WTD) corners of the value space of TD and WTD. Comprising a wide range of thawing scenarios, the extended simulations can give us a complete insight into how TD and WTD impact C cycles after permafrost thaw.

3.3 Responses of C fluxes to variations in TD or WTD

Figure 6 shows that NEE is more sensitive to changes in WTD (Fig. 6b) than TD (Fig. 6a). Regardless of variations in TD (Fig. 6a), NEE remains positive with wet soil and negative with dry soil most time. Additionally, the increment in WTD can cause NEE to move from a C sink to a C source, regardless of TD (Fig. 6b). All these indicate that C balance is more relevant with WTD.

Other panels in Fig. 6 further display how GPP and ER respond to variations in TD and WTD. The increase in GPP along with thaw depth (Fig. 6c) is largely offset by the increase in ER (Fig. 6e), so the impact of deepening thaw on NEE is not significant in Fig. 6a. Additionally, elevated WTD leads to a decrease in GPP (Fig. 6c) but an increase in ER (Fig. 6e) most time no matter how TD changes. This partly explained why NEE was converted to a net source with wetter soil in Fig. 6b. GPP decreases dramatically with both deep and shallow TD when WTD approaches the surface (-0.1 m) (Fig. 6d). ER with deep TD also declines when WTD reaches the surface (Fig. 6f), but the decline is slower than that of GPP in Fig. 6d. Meanwhile, ER with shallow TD remains consistent in the flooding event where WTD is above the ground (Fig. 6f). Consequently, it is GPP that dominates the shift from C sink to a C source in Fig. 6b.

4. Discussion

Warming in high-latitude ecosystems triggers permafrost thaw, leading to many abiotic and biotic alterations. Thaw-induced hydrology change can lead to a variety of alterations in water

table depth (WTD) (i.e., soil wetting or drying), depending on the landscape, soil texture, terrain, etc. WTD may have a great impact on C balance (i.e., NEE) in the high-latitude ecosystem. This effect of WTD on C cycles has not been directly estimated because of the complexity of measuring the changes in WTD across different landscapes after permafrost thaw. Moreover, the effect of WTD is often intertwined with different permafrost thawing. This data assimilation study is to investigate the effects of elevated WTD and increasing thaw depth (TD) on the C cycle in a permafrost ecosystem with 8-year experimental warming. We found that NEE is more sensitive to changes in WTD and a possible reason is the ongoing heterotrophic respirations near the surface under thermokarst.

4.1 Simulations constrained by data assimilation

The CiPEHR experiment collects measurements from three blocks located disparately, taking into account the influence of topography (Natali et al., 2010). Compared with observations at CiPEHR, the simulation of the TECO model constrained by data assimilation can achieve a R^2 of 0.42 for NEE, 0.75 for TD, and 0.70 for WTD. The simulated NEE exhibited the lowest accuracy in the air warming plots, especially in the first three years. A possible reason is that the arrangement of chambers induces disturbances that are hard for model to reproduce in simulations. Simulated TD can capture the increasing trend in observations with soil warming or air & soil warming treatments. Along with thaw deepening, an elevated WTD was also replicated in model simulations. This matches our observations at the CiPEHR site (Fig. 4) (Mauritz et al., 2017). It reveals that the simulation performance of hydrology changes improved after we incorporated subsidence processes (Eqs. 1-6) into the TECO model. On the contrary, the study by Schadel et al. (2018) employed CLM 4.5 and CLM 5 to simulate the biogeochemical processes at the CiPEHR site but it failed to predict wetting soil (i.e., elevated WTD) as

observed. Although subsidence or thermokarst is often witnessed in the permafrost region (Olefeldt et al., 2016), this alteration in WTD is often not clearly represented in many other models such as LPJGUESS (Andresen et al., 2020). The comparisons between observations (i.e., NEE, WTD, and TD) and simulation results indicate that the TECO model is well informed by observations after data assimilation and is suitable for the CiPEHR site (Figs. 2 – 4).

Due to the limitations in the observed permafrost thaw conditions in the field experiment, only the effects of warming on ecosystem carbon cycles of permafrost in deeply thawed areas with wet conditions or shallowly thawed areas with dry conditions were examined at the CiPEHR site (Fig. 5a). Taking advantage of the unique capability of models through parameterization, we extended the scenarios of WTD and TD under warming with the TECO model to include the other two conditions that were not observed in the field experiment, i.e., deep thaw areas with dry soil and shallow thaw areas with wet soil (Fig. 5b). Thus, we were able to use the extended simulations to conduct a comprehensive modeling analysis on the effects of elevated water table depth and increasing thaw depth on the C cycles in this permafrost ecosystem under different scenarios. Compared with field-based experiments, modeling-based analysis is not confined to spatial or temporal limitations.

4.2 Responses of C cycles to alterations in water table depth

We found NEE is more sensitive to changes in WTD than TD (Fig. 6a). The elevated WTD turns the permafrost from a C sink to a C source regardless of deep or shallow thaw (Fig. 6b). This finding is consistent with the conclusion of Pries et al. (2013). This research conducted at the same study site (i.e., CiPEHR) and a neighboring permafrost thaw gradient indicated that wetting soil would enhance decomposition. However, our results differ from a soil drying study in the same permafrost ecosystem (Natali et al., 2015). More experimental drying studies also

concluded that soil drying can accelerate aerobic respirations, leading to more C release (Oechel et al., 1998; Kwon et al., 2016; Lupascu et al., 2014). However, each of these previous studies was associated with one thawing scenario, that is deep thaw with drier soil or shallow thaw with wetter soil. Our study synthesizes all these scenarios together. Although there is no consensus about whether elevated WTD (i.e., soil wetting) will lead to more C sequestration or emissions, many studies have recognized the important impacts of water on C cycles (Berner and Schädel, 2023; Zona et al., 2023). For example, Aerts (2006) showed that only when there is sufficient soil moisture, warming can cause a faster litter decomposition, leading to an increase in C emission. Another model prediction study with CLM indicated that soil drying would reduce the worldwide warming potential of the C-climate feedback by half in the permafrost region (Lawrence et al., 2015).

The shift of C sink to source along with elevated WTD is mainly due to the suppressed GPP caused by flooding and subsidence (Fig. 6d). While the limited oxygen availability during flooding has been widely reported to inhibit plant growth, the dominant species in this site, *E. vaginatum*, can grow faster after thaw because of more nitrogen available (Johansson et al., 2013; Wahren et al., 2005; Schuur et al., 2007). The denser *E. vaginatum* canopy may cause self-shading and lower photosynthesis, resulting in decreases in GPP. Besides, ER suffers less compared to GPP in flooding (Fig. 6f). One possible reason is that microbes and plant roots live in the surface soil layer where elevated WTD except for flooding influences less. Meanwhile, it is easier to transfer oxygen through soil pores or vascular plants and keep a stable respiration rate. The analysis of GPP and ER responses to WTD suggests that plant photosynthesis responds more quickly to subsidence than microbial and plant respirations, which agrees with findings in previous studies in this area (Schadel et al., 2018; Mauritz et al., 2017).

The responses of GPP and ER to elevated WTD are also different from TD. In deep thaw (blue lines in figs. 6d and 6f) with dry conditions (WTD: -0.4 ~ -0.2), ER increases faster with soil wetting compared with GPP. However, in shallow thaw with wet conditions (WTD: -0.1 ~ 0.2), GPP declines faster than ER along with soil wetting. Because the rise/decline in plant respiration is comparable to the increment/decrease in GPP, the discrepancy between GPP and ER reveals stronger soil respiration in these two cases (i.e., deep thaw with dry condition and shallow thaw with wet condition).

Furthermore, an interesting question is whether older C is released Among enhanced soil respiration in these two cases. Old C emission is of particular note because it has a higher potential for sustained and strong warming feedback (Schuur and Abbott, 2011). ; Schuur et al., 2009). The study by Schuur et al. (2009) indicated areas with deep thawing released 40% more old C than that with shallow thaw. Pries et al. (2013b) also suggested that old C emissions can be intensified with rising soil moisture during deep thaw. On the contrary, Pegoraro et al. (2021) indicated that lower soil moisture in the thermokarst (WTD>0) leads to more old C emissions. We may need more isotope studies to validate the impacts of elevated WTD on old C emissions.

4.3 The elevated thaw depth promotes carbon sequestration in permafrost ecosystem

In contrast to the effect of elevated WTD on the C fluxes in permafrost, with increase in thaw depth, permafrost becomes a greater C sink (Fig. 6a). A CO₂ sink has been reported previously under a moderate thaw site in permafrost, but permafrost with extensive thaw could be a significant net source of CO₂ to the atmosphere due to the substantial increase of Reco during the spring, fall, and winter that offset the initial periods of increased GPP (Vogel et al., 2009).

The greater C sequestration with increased TD in our study can also be attributed to the different responses of GPP and Reco to changes in TD. Both GPP and Reco increase with TD

but the magnitude of the increase in GPP (Fig. 6c) is slightly larger than that in Reco (Fig. 6e). As permafrost thaws, increased nutrient availability (e.g., nitrogen availability) may promote plant productivity and increase GPP (Finger et al., 2016). At the same time, increased TD is usually associated with higher temperature, increased moisture, and availability of deep soil C, which can all stimulate Reco, especially heterotrophic respiration (Pries et al., 2013a). A rapid thaw often occurs in ice-rich soil, leading to subsidence or large variation in WTD (Mauritz et al., 2017). This alteration in WTD will reversely contribute to more emergence of rapid thaw because the thermal capacity of soil is increased by water (Romanovsky and Osterkamp, 2000).

4.4 Limitations of this study

One limitation is that our results is more associated to the CiPEHR site because the model is constrained by observations at this study site through data assimilation. We may need more test cases to verify the findings. In addition, thaw deepening releases more N for plant growth, which is represented in Eq. 10. However, we did not have N observations to constrain the model in data assimilation. Thus this N impact factor may not be well constrained and the associated effects of thaw deepening may be overestimated or underestimated. Furthermore, flooding or thermokarst caused by soil wetting facilitates CH₄ emissions (Zona et al., 2009), and this portion of greenhouse gas is not considered in this study. We may need to incorporate CH₄ processes (Ma et al., 2017) into the simulation. Finally, the effect of permafrost thaw and subsidence in our study are simplified as impact factors that are constrained in data assimilation. In the future, we may need more detailed parameterization schemes.

5. Conclusions

It is predicted that the permafrost region will experience severe temperature rising and thawing by the end of 2100 (IPCC, 2021). The responses of C cycles in the permafrost ecosystem to

increasing thaw depth (TD) are highly correlated with elevated WTD upon permafrost thaw. Our study demonstrates that the altered hydrological conditions (e.g., subsidence, thermokarst, and flooding) play an important role in regulating C cycles in permafrost regions. Elevated water table depth (WTD) can shift permafrost from a C sink to a C source regardless of the changes in TD. Models without accounting for subsidence processes will overestimate C sink in permafrost ecosystems. While this study used a modeling analysis to explore the responses of C cycles to alterations in WTD and TD, more evidence from field experiments is needed to fully understand the effects of WTD and TD on C cycles under different thawing scenarios (i.e., deep thaw with soil drying and shallow thaw with soil wetting) which may not be measured in a specific permafrost warming experiment. Considering a large amount of C stored in permafrost, the responses of C cycles to elevated WTD need to be explicitly represented in ESMs.

Appendix A: Comparison between model simulations and observations

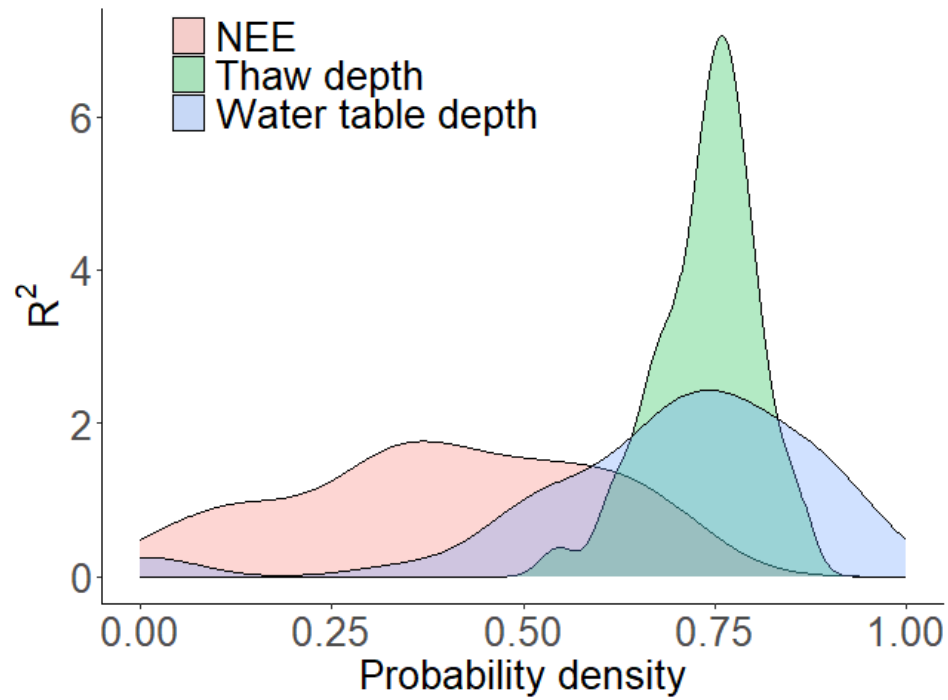


Figure A1: The probability density of R^2 of the data model fitting for net ecosystem exchange (NEE, red), thaw depth (green), and water table depth (blue) in 48 plots. For thaw depth and water table depth, most of the 48 plots have a R^2 of around 0.75. For NEE, more than half of the 48 plots have a R^2 value greater than 0.37.

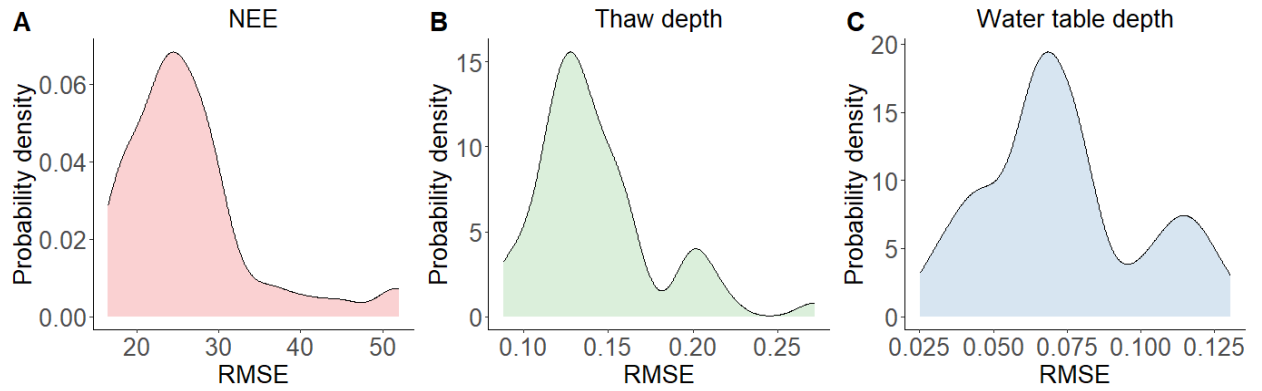


Figure A2: The probability density of root-mean-square error (RMSE) between the simulated and observed NEE (A); the simulated and observed thaw depth (B); the simulated and observed water table depth (C). Most RMSE for NEE, thaw depth, and water table depth is concentrated around 25, 0.13, and 0.075, respectively.

Appendix B: Meteorology variables, C fluxes, and

	Variable	Description	Unit	Frequency
Meteorology	Air temperature	Plot level	$^{\circ}\text{C}$	1.5 hr^{-1}
	Soil temperature	Plot level	$^{\circ}\text{C}$	Half-hourly
	Cumulative precipitation	Site level	mm	Half-hourly
	PAR	Site level	$\mu\text{mol m}^2\text{s}^{-1}$	Half-hourly
Environmental	Snow depth	Plot level	m	Yearly
	Water table depth	Plot level	cm	Daily
	Thaw depth	Plot level	cm	Daily
Carbon fluxes	NEE	Plot level	$\text{gCm}^{-2}\text{halfhour}^{-1}$	Half-hourly
	GPP	Plot level	$\text{gCm}^{-2}\text{halfhour}^{-1}$	Half-hourly
	ER	Plot level	$\text{gCm}^{-2}\text{halfhour}^{-1}$	Half-hourly
Carbon pool	Aboveground biomass	Plot-level	drygm^{-2}	Yearly

References

- Aerts, R.: The freezer defrosting: global warming and litter decomposition rates in cold biomes, *Journal of Ecology*, 94, 713–724, <https://doi.org/10.1111/j.1365-2745.2006.01142.x>, 2006.
- Andresen, C. G., Lawrence, D. M., Wilson, C. J., McGuire, A. D., Koven, C., Schaefer, K., Jafarov, E., Peng, S., Chen, X., Gouttevin, I., Burke, E., Chadburn, S., Ji, D., Chen, G., Hayes, D., and Zhang, W.: Soil moisture and hydrology projections of the permafrost region – a model intercomparison, *The Cryosphere*, 14, 445–459, <https://doi.org/10.5194/tc-14-445-2020>, 2020.
- Anon: The impact of the permafrost carbon feedback on global climate Related content, <https://doi.org/10.1088/1748-9326/9/8/085003>, 2014.
- Ball, J. T., Woodrow, I. E., and Berry, J. A.: A Model Predicting Stomatal Conductance and its Contribution to the Control of Photosynthesis under Different Environmental Conditions, in: *Progress in Photosynthesis Research*, Springer Netherlands, 221–224, https://doi.org/10.1007/978-94-017-0519-6_48, 1987.
- Barrett, D.: Steady state turnover time of carbon in the Australian terrestrial biosphere, *Global Biogeochemical Cycles*, 16, 55-1-55–21, <https://doi.org/10.1029/2002gb001860>, 2002.
- Belshe, E. F., Schuur, E. A. G., and Bolker, B. M.: Tundra ecosystems observed to be CO₂ sources due to differential amplification of the carbon cycle, *Ecology Letters*, 16, 1307–1315, <https://doi.org/10.1111/ele.12164>, 2013.
- Berner, L. T. and Schädel, C.: Soil moisture impacts the tundra carbon balance in a changing climate, *Global Change Biology*, 29, 1217–1219, <https://doi.org/10.1111/gcb.16521>, 2023.
- Box, G. E. P. and Tiao, G. C.: *Bayesian Inference in Statistical Analysis*, John Wiley & Sons, Inc., Hoboken, NJ, USA, <https://doi.org/10.1002/9781118033197>, 1992.
- Celis, G., Mauritz, M., Bracho, R., Salmon, V. G., Webb, E. E., Hutchings, J., ... & Schuur, E. A.: Tundra is a consistent source of CO₂ at a site with progressive permafrost thaw during 6 years of chamber and eddy covariance measurements, *Journal of Geophysical Research: Biogeosciences*, 122, 1471–1485, <https://doi.org/10.1002/2016JG003677>, 2017.
- Farquhar, G. D., von Caemmerer, S., and Berry, J. A.: A biochemical model of photosynthetic CO₂ assimilation in leaves of C₃ species, *Planta*, 149, 78–90, <https://doi.org/10.1007/BF00386231>, 1980.
- Finger, R. A., Turetsky, M. R., Kielland, K., Ruess, R. W., Mack, M. C., and Euskirchen, E. S.: Effects of permafrost thaw on nitrogen availability and plant–soil interactions in a boreal Alaskan lowland, *Journal of Ecology*, 104, 1542–1554, <https://doi.org/10.1111/1365-2745.12639>, 2016.
- Hicks Pries, C. E., Schuur, E. A. G., Vogel, J. G., and Natali, S. M.: Moisture drives surface decomposition in thawing tundra, *Journal of Geophysical Research: Biogeosciences*, 118, 1133–1143, <https://doi.org/10.1002/jgrg.20089>, 2013a.
- Hicks Pries, C. E., Schuur, E. A. G., and Crummer, K. G.: Thawing permafrost increases old soil and autotrophic respiration in tundra: Partitioning ecosystem respiration using $\delta^{13}\text{C}$ and $\Delta^{14}\text{C}$, *Global Change Biology*, 19, 649–661, <https://doi.org/10.1111/gcb.12058>, 2013b.
- Huang, X., Lu, D., Ricciuto, D. M., Hanson, P. J., Richardson, A. D., Lu, X., Weng, E., Nie, S., Jiang, L., Hou, E., Steinmacher, I. F., and Luo, Y.: A model-independent data

- assimilation (MIDA) module and its applications in ecology, *Geosci. Model Dev.*, 14, 5217–5238, <https://doi.org/10.5194/gmd-14-5217-2021>, 2021.
- Huang, Y., Jiang, J., Ma, S., Ricciuto, D., Hanson, P. J., and Luo, Y.: Soil thermal dynamics, snow cover, and frozen depth under five temperature treatments in an ombrotrophic bog: Constrained forecast with data assimilation, *Journal of Geophysical Research: Biogeosciences*, 122, 2046–2063, <https://doi.org/10.1002/2016JG003725>, 2017.
- Hugelius, G., Strauss, J., Zubrzycki, S., Harden, J. W., Schuur, E. A. G., Ping, C. L., Schirrmeister, L., Grosse, G., Michaelson, G. J., Koven, C. D., O'Donnell, J. A., Elberling, B., Mishra, U., Camill, P., Yu, Z., Palmtag, J., and Kuhry, P.: Estimated stocks of circumpolar permafrost carbon with quantified uncertainty ranges and identified data gaps, *Biogeosciences*, 11, 6573–6593, <https://doi.org/10.5194/bg-11-6573-2014>, 2014.
- Jiang, J., Huang, Y., Ma, S., Stacy, M., Shi, Z., Ricciuto, D. M., Hanson, P. J., and Luo, Y.: Forecasting Responses of a Northern Peatland Carbon Cycle to Elevated CO₂ and a Gradient of Experimental Warming, *Journal of Geophysical Research: Biogeosciences*, 123, 1057–1071, <https://doi.org/10.1002/2017JG004040>, 2018.
- Johansson, M., Callaghan, T. V., Bosiö, J., Åkerman, H. J., Jackowicz-Korczynski, M., and Christensen, T. R.: Rapid responses of permafrost and vegetation to experimentally increased snow cover in sub-arctic Sweden, *Environmental Research Letters*, 8, <https://doi.org/10.1088/1748-9326/8/3/035025>, 2013.
- Jorgenson, M. T., Racine, C. H., Walters, J. C., and Osterkamp, T. E.: Permafrost degradation and ecological changes associated with a warming climate in central Alaska, *Climatic Change*, 48, 551–579, <https://doi.org/10.1023/A:1005667424292>, 2001.
- Kwon, M. J., Heimann, M., Kolle, O., Luus, K. A., Schuur, E. A. G., Zimov, N., Zimov, S. A., and Göckede, M.: Long-term drainage reduces CO₂ uptake and increases CO₂ emission on a Siberian floodplain due to shifts in vegetation community and soil thermal characteristics, *Biogeosciences*, 13, 4219–4235, <https://doi.org/10.5194/bg-13-4219-2016>, 2016.
- Lawrence, D. M., Koven, C. D., Swenson, S. C., Riley, W. J., and Slater, A. G.: Permafrost thaw and resulting soil moisture changes regulate projected high-latitude CO₂ and CH₄ emissions, *Environ. Res. Lett.*, 10, 094011, <https://doi.org/10.1088/1748-9326/10/9/094011>, 2015.
- Lupascu, M., Welker, J. M., Seibt, U., Maseyk, K., Xu, X., and Czimczik, C. I.: High Arctic wetting reduces permafrost carbon feedbacks to climate warming, *Nature Clim Change*, 4, 51–55, <https://doi.org/10.1038/nclimate2058>, 2014.
- Ma, S., Jiang, J., Huang, Y., Shi, Z., Wilson, R. M., Ricciuto, D., Sebestyen, S. D., Hanson, P. J., and Luo, Y.: Data-Constrained Projections of Methane Fluxes in a Northern Minnesota Peatland in Response to Elevated CO₂ and Warming, *Journal of Geophysical Research: Biogeosciences*, 122, 2841–2861, <https://doi.org/10.1002/2017JG003932>, 2017.
- Mauritz, M., Bracho, R., Celis, G., Hutchings, J., Natali, S. M., Pegoraro, E., Salmon, V. G., Schädel, C., Webb, E. E., and Schuur, E. A. G.: Nonlinear CO₂ flux response to 7 years of experimentally induced permafrost thaw, *Global Change Biology*, 23, 3646–3666, <https://doi.org/10.1111/gcb.13661>, 2017a.
- Mauritz, M., Bracho, R., Celis, G., Hutchings, J., Natali, S. M., Pegoraro, E., Salmon, V. G., Schädel, C., Webb, E. E., and Schuur, E. A. G.: Nonlinear CO₂ flux response to 7 years of experimentally induced permafrost thaw, *Global Change Biology*, 23, 3646–3666, <https://doi.org/10.1111/gcb.13661>, 2017b.

- McGuire, A. D., Christensen, T. R., Hayes, D., Heroult, A., Euskirchen, E., Kimball, J. S., Koven, C., Lafleur, P., Miller, P. A., Oechel, W., Peylin, P., Williams, M., and Yi, Y.: An assessment of the carbon balance of Arctic tundra: Comparisons among observations, process models, and atmospheric inversions, *Biogeosciences*, 9, 3185–3204, <https://doi.org/10.5194/bg-9-3185-2012>, 2012.
- Natali, S. M., Schuur, E. A. G., Trucco, C., Pries, C. E. H., Crummer, K. G., and Lopez, A. F. B.: Effects of experimental warming of air, soil and permafrost on carbon balance in Alaskan tundra, *Global Change Biology*. 17(3): 1394–1407, 17, 1394–1407, <https://doi.org/10.1111/j.1365-2486.2010.02303.x>, 2010.
- Natali, S. M., Schuur, E. A. G., and Rubin, R. L.: Increased plant productivity in Alaskan tundra as a result of experimental warming of soil and permafrost, *Journal of Ecology*, 100, 488–498, <https://doi.org/10.1111/j.1365-2745.2011.01925.x>, 2012.
- Natali, S. M., Schuur, E. A. G., Mauritz, M., Schade, J. D., Celis, G., Crummer, K. G., Johnston, C., Krapek, J., Pegoraro, E., Salmon, V. G., and Webb, E. E.: Permafrost thaw and soil moisture driving CO₂ and CH₄ release from upland tundra, *Journal of Geophysical Research: Biogeosciences*, 120, 525–537, <https://doi.org/10.1002/2014JG002872>, 2015.
- O'Donnell, J. A., Jorgenson, M. T., Harden, J. W., McGuire, A. D., Kanevskiy, M. Z., and Wickland, K. P.: The Effects of Permafrost Thaw on Soil Hydrologic, Thermal, and Carbon Dynamics in an Alaskan Peatland, *Ecosystems*, 15, 213–229, <https://doi.org/10.1007/s10021-011-9504-0>, 2012.
- Oechel, W. C., Vourlitis, G. L., Hastings, S. J., Ault Jr., R. P., and Bryant, P.: The effects of water table manipulation and elevated temperature on the net CO₂ flux of wet sedge tundra ecosystems, *Global Change Biology*, 4, 77–90, <https://doi.org/10.1046/j.1365-2486.1998.00110.x>, 1998.
- Olefeldt, D., Goswami, S., Grosse, G., Hayes, D., Hugelius, G., Kuhry, P., McGuire, A. D., Romanovsky, V. E., Sannel, A. B. K., Schuur, E. A. G., and Turetsky, M. R.: Circumpolar distribution and carbon storage of thermokarst landscapes, *Nat Commun*, 7, 13043, <https://doi.org/10.1038/ncomms13043>, 2016.
- Osterkamp, T. E., Jorgenson, M. T., Schuur, E. A. G., Shur, Y. L., Kanevskiy, M. Z., Vogel, J. G., and Tumskey, V. E.: Physical and ecological changes associated with warming permafrost and thermokarst in Interior Alaska: Physical and Ecological Changes Associated with Thermokarst, *Permafrost Periglac. Process.*, 20, 235–256, <https://doi.org/10.1002/ppp.656>, 2009.
- Pegoraro, E. F., Mauritz, M. E., Ogle, K., Ebert, C. H., and Schuur, E. A. G.: Lower soil moisture and deep soil temperatures in thermokarst features increase old soil carbon loss after 10 years of experimental permafrost warming, *Global Change Biology*, 27, 1293–1308, <https://doi.org/10.1111/gcb.15481>, 2021.
- Rodenhizer, H., Ledman, J., Mauritz, M., Natali, S. M., Pegoraro, E., Plaza, C., Romano, E., Schädel, C., Taylor, M., and Schuur, E.: Carbon Thaw Rate Doubles When Accounting for Subsidence in a Permafrost Warming Experiment, *Journal of Geophysical Research: Biogeosciences*, 125, 1–16, <https://doi.org/10.1029/2019JG005528>, 2020.
- Romanovsky, V. E. and Osterkamp, T. E.: Effects of unfrozen water on heat and mass transport processes in the active layer and permafrost, *Permafrost and Periglacial Processes*, 11, 219–239, [https://doi.org/10.1002/1099-1530\(200007/09\)11:3<219::AID-PPP352>3.0.CO;2-7](https://doi.org/10.1002/1099-1530(200007/09)11:3<219::AID-PPP352>3.0.CO;2-7), 2000.

- Salmon, V. G., Soucy, P., Mauritz, M., Celis, G., Natali, S. M., Mack, M. C., and Schuur, E. A. G.: Nitrogen availability increases in a tundra ecosystem during five years of experimental permafrost thaw, *Global Change Biology*, 22, 1927–1941, <https://doi.org/10.1111/gcb.13204>, 2016.
- Schadel, C., Koven, C. D., Lawrence, D. M., Celis, G., Garnello, A. J., Hutchings, J., Mauritz, M., Natali, S. M., Pegoraro, E., Rodenhizer, H., Salmon, V. G., Taylor, M. A., Webb, E. E., Wieder, W. R., and Schuur, E. A. G.: Divergent patterns of experimental and model-derived permafrost ecosystem carbon dynamics in response to Arctic warming, *Environmental Research Letters*, 13, <https://doi.org/10.1088/1748-9326/aae0ff>, 2018.
- Schuur, E. A. G. and Abbott, B.: High risk of permafrost thaw, *Nature*, 480, 32–33, <https://doi.org/10.1038/480032a>, 2011.
- Schuur, E. A. G. and Mack, M. C.: Ecological Response to Permafrost Thaw and Consequences for Local and Global Ecosystem Services, *Annual Review of Ecology, Evolution, and Systematics*, 49, 279–301, <https://doi.org/10.1146/annurev-ecolsys-121415-032349>, 2018a.
- Schuur, E. A. G. and Mack, M. C.: Ecological Response to Permafrost Thaw and Consequences for Local and Global Ecosystem Services, 279–301, 2018b.
- Schuur, E. A. G., Crummer, K. G., Vogel, J. G., and MacK, M. C.: Plant species composition and productivity following permafrost thaw and thermokarst in Alaskan tundra, *Ecosystems*, 10, 280–292, <https://doi.org/10.1007/s10021-007-9024-0>, 2007.
- Schuur, E. A. G., Vogel, J. G., Crummer, K. G., Lee, H., Sickman, J. O., and Osterkamp, T. E.: The effect of permafrost thaw on old carbon release and net carbon exchange from tundra, *Nature*, 459, 556–559, <https://doi.org/10.1038/nature08031>, 2009.
- Schuur, E. A. G., McGuire, A. D., Schädel, C., Grosse, G., Harden, J. W., Hayes, D. J., Hugelius, G., Koven, C. D., Kuhry, P., Lawrence, D. M., Natali, S. M., Olefeldt, D., Romanovsky, V. E., Schaefer, K., Turetsky, M. R., Treat, C. C., and Vonk, J. E.: Climate change and the permafrost carbon feedback, *Nature*, 520, 171–179, <https://doi.org/10.1038/nature14338>, 2015.
- Shevliakova, E., Pacala, S. W., Malyshev, S., Hurtt, G. C., Milly, P. C. D., Caspersen, J. P., Sentman, L. T., Fisk, J. P., Wirth, C., and Crevoisier, C.: Carbon cycling under 300 years of land use change: Importance of the secondary vegetation sink, *Global Biogeochemical Cycles*, 23, <https://doi.org/10.1029/2007GB003176>, 2009.
- Shi, Z., Yang, Y., Zhou, X., Weng, E., Finzi, A. C., and Luo, Y.: Inverse analysis of coupled carbon-nitrogen cycles against multiple datasets at ambient and elevated CO₂, *Journal of Plant Ecology*, 9, 285–295, <https://doi.org/10.1093/jpe/rtv059>, 2016.
- Vogel, J., Schuur, E. A. G., Trucco, C., and Lee, H.: Response of CO₂ exchange in a tussock tundra ecosystem to permafrost thaw and thermokarst development, *Journal of Geophysical Research: Biogeosciences*, 114, <https://doi.org/10.1029/2008JG000901>, 2009.
- Wahren, C. H. A., Walker, M. D., and Bret-Harte, M. S.: Vegetation responses in Alaskan arctic tundra after 8 years of a summer warming and winter snow manipulation experiment, *Global Change Biology*, 11, 537–552, <https://doi.org/10.1111/j.1365-2486.2005.00927.x>, 2005.
- Weng, E. and Luo, Y.: Soil hydrological properties regulate grassland ecosystem responses to multifactor global change: A modeling analysis, *Journal of Geophysical Research: Biogeosciences*, 113, 1–16, <https://doi.org/10.1029/2007JG000539>, 2008.

- Yiqi, L. and Reynolds, J. F.: Validity of extrapolating field CO₂ experiments to predict carbon sequestration in natural ecosystems, *Ecology*, 80, 1568–1583, [https://doi.org/10.1890/0012-9658\(1999\)080\[1568:voefce\]2.0.co;2](https://doi.org/10.1890/0012-9658(1999)080[1568:voefce]2.0.co;2), 1999.
- Zona, D., Oechel, W. C., Kochendorfer, J., Paw U, K. T., Salyuk, A. N., Olivas, P. C., Oberbauer, S. F., and Lipson, D. A.: Methane fluxes during the initiation of a large-scale water table manipulation experiment in the Alaskan Arctic tundra: WATER TABLE IMPACTS ON METHANE FLUXES, *Global Biogeochem. Cycles*, 23, n/a-n/a, <https://doi.org/10.1029/2009GB003487>, 2009.
- Zona, D., Lafleur, P. M., Hufkens, K., Gioli, B., Bailey, B., Burba, G., Euskirchen, E. S., Watts, J. D., Arndt, K. A., Farina, M., Kimball, J. S., Heimann, M., Göckede, M., Pallandt, M., Christensen, T. R., Mastepanov, M., López-Blanco, E., Dolman, A. J., Commane, R., Miller, C. E., Hashemi, J., Kutzbach, L., Holl, D., Boike, J., Wille, C., Sachs, T., Kalhori, A., Humphreys, E. R., Sonnentag, O., Meyer, G., Gosselin, G. H., Marsh, P., and Oechel, W. C.: Pan-Arctic soil moisture control on tundra carbon sequestration and plant productivity, *Global Change Biology*, 29, 1267–1281, <https://doi.org/10.1111/gcb.16487>, 2023.

Tables

Table 1: Parameter descriptions and their range in the TECO model

Parameter	Description	Unit	Default	Range
β_{si}	A scaler to represent snow accumulation effect because of snow fences in soil warming plots	-	2	[1, 6]
γ_{td}	A scaler to represent TD effect	-	0.2	[0.01, 0.5]
z_c	A critical threshold to decide whether WTD is above the ground or not	mm	-100	[-200,100]
α_{fd}	A scaler to represent the flooding effect	-	0.2	[0, 0.9]
Δs_1	The subsidence rate for year 2009	cm/hr	0	[-20, 70]
Δs_2	The subsidence rate for year 2010	cm/hr	0	[-20, 70]
Δs_3	The subsidence rate for year 2011	cm/hr	0	[-20, 70]
Δs_4	The subsidence rate for year 2012	cm/hr	0	[-20, 70]
Δs_5	The subsidence rate for year 2013	cm/hr	0	[-20, 70]
Δs_6	The subsidence rate for year 2014	cm/hr	0	[-20, 70]
Δs_7	The subsidence rate for year 2015	cm/hr	0	[-20, 70]
Δs_8	The subsidence rate for year 2016	cm/hr	0	[-20, 70]

Figures

Figure 1: Workflow of data assimilation (DA) at CiPEHR site. Arrows represent data flow. The input of model simulation is meteorological forcing. The outputs of model simulation are carbon (C) flux, soil moisture, etc. Data assimilation is to integrate observations about water, C pools and fluxes, snow depth, soil temperature, and thaw depth with model.

Figure 2: Comparison between simulated monthly mean net ecosystem exchange (NEE) by TECO model and the observed NEE from 2009 to 2016 under control, air warming, soil warming, and air & soil warming across 48 plots. Negative NEE means C sink while positive NEE means C source. Blue line represents observed NEE and red line represents simulated NEE with optimized parameter values after DA. Every treatment plot repeats two times in each of the six fences that are organized across three blocks.

Figure 3: Comparison between simulated daily thaw depth (TD) by TECO model and the observed TD from 2009 to 2016 under control, air warming, soil warming, and air & soil warming treatments across 48 plots. Blue circles represent observed TD and red lines represent simulated TD with optimized parameter values after DA. Every treatment plot repeats two times in each of the six fences that are organized across three blocks.

Figure 4: Comparison between simulated annual mean water table depth (WTD) by TECO model and the observed WTD from 2009 to 2016 under control, air warming, soil warming, and air & soil warming treatments across 48 plots. Positive WTD means that flooding occurs. Blue lines represent observed WTD, and red lines represent simulated WTD with optimized parameter values after DA. Every treatment plot repeats two times in each of the six fences that are organized across three blocks.

Figure 5: Frequency of the value combination of WTD and TD in (a) observations, (b) extended simulations with the TECO model after DA. The size of a grid is $0.1m \times 0.1m$. The color in each grid represents the frequency of a particular value combination of WTD and TD. White area outside the blue border in (a) means there is no observation. Compared to observations, the extended simulations include more value combinations of WTD and TD after permafrost thaw, such as shallower TD with elevated WTD and deeper TD with lower WTD.

Figure 6: Responses of NEE to (a) thaw depth (TD) in wet (blue) and dry soil (red) and (b) water table depth (WTD) with deep (blue) and shallow thaw (red); responses of GPP to (c) TD in wet (blue) and dry soil (red) and (d) WTD with deep (blue) and shallow thaw (red); and responses of ecosystem respiration (ER) to (e) TD in wet (blue) and dry soil (red) and (f) WTD with deep (blue) and shallow thaw (red). All variables are from the extended simulations (Fig. 5) by TECO model. The wet and dry soils in (a, c, e) indicate that WTD ranges from 0 to 0.15 and -0.4 to -0.25, respectively. The deep and shallow thaws in (b, d, f) mean TD ranges from -1.18 to -0.95 and -0.6 to -0.38, respectively. The shadow represents a 95% confidence interval. Positive WTD means that flooding occurs. Positive NEE means C source while negative NEE means C sink.

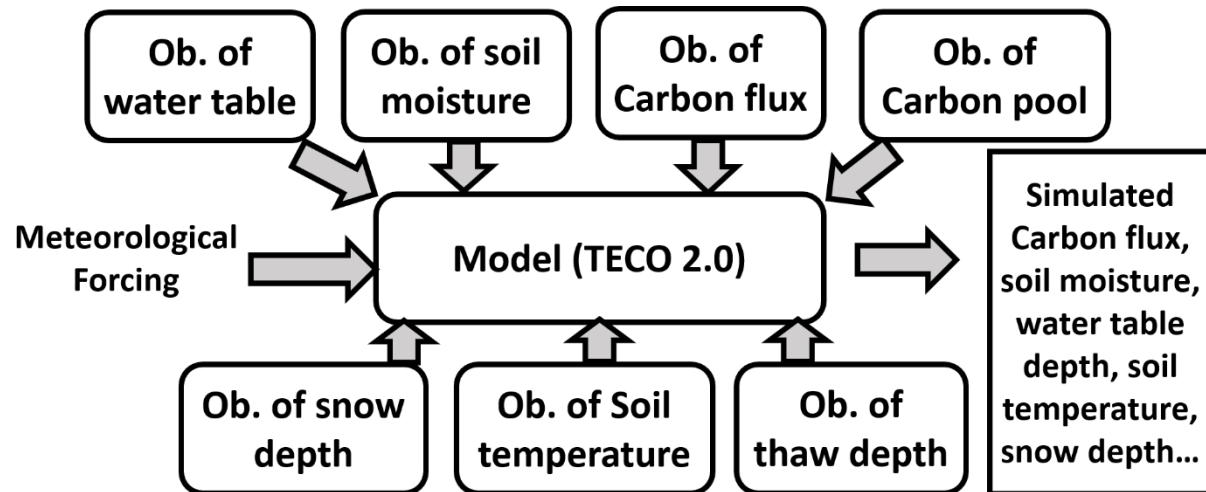


Figure 1: Workflow of data assimilation (DA) at CiPEHR site. Arrows represent data flow. The input of model simulation is meteorological forcing. The outputs of model simulation are carbon (C) fluxes, soil moisture, etc. Data assimilation is to integrate observations of water, C pools and fluxes, snow depth, soil temperature, and thaw depth with model.

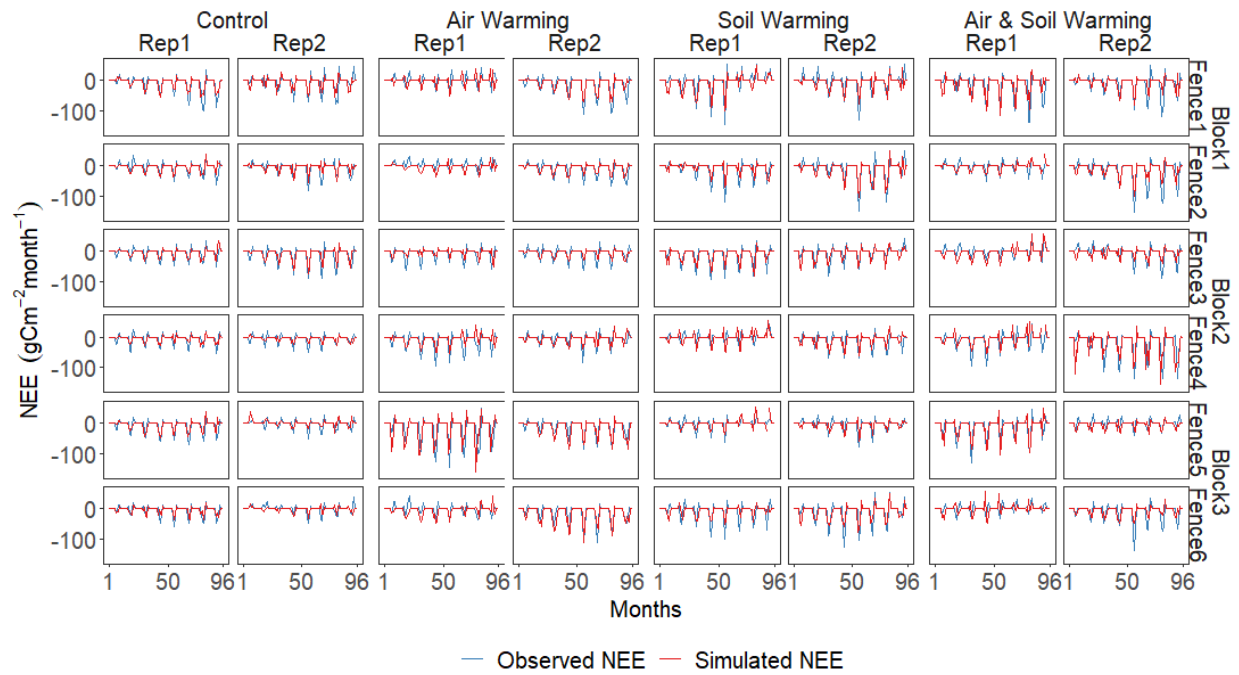


Figure 2: Comparison between simulated monthly mean net ecosystem exchange (NEE) by TECO model and the observed NEE from 2009 to 2016 under control, air warming, soil warming, and air & soil warming treatments across 48 plots. Negative NEE means C sink while positive NEE means C source. Blue lines represent observed NEE and red lines represent simulated NEE with optimized parameter values after DA. Each treatment plot repeats two times in each of the six fences that are organized by three blocks.

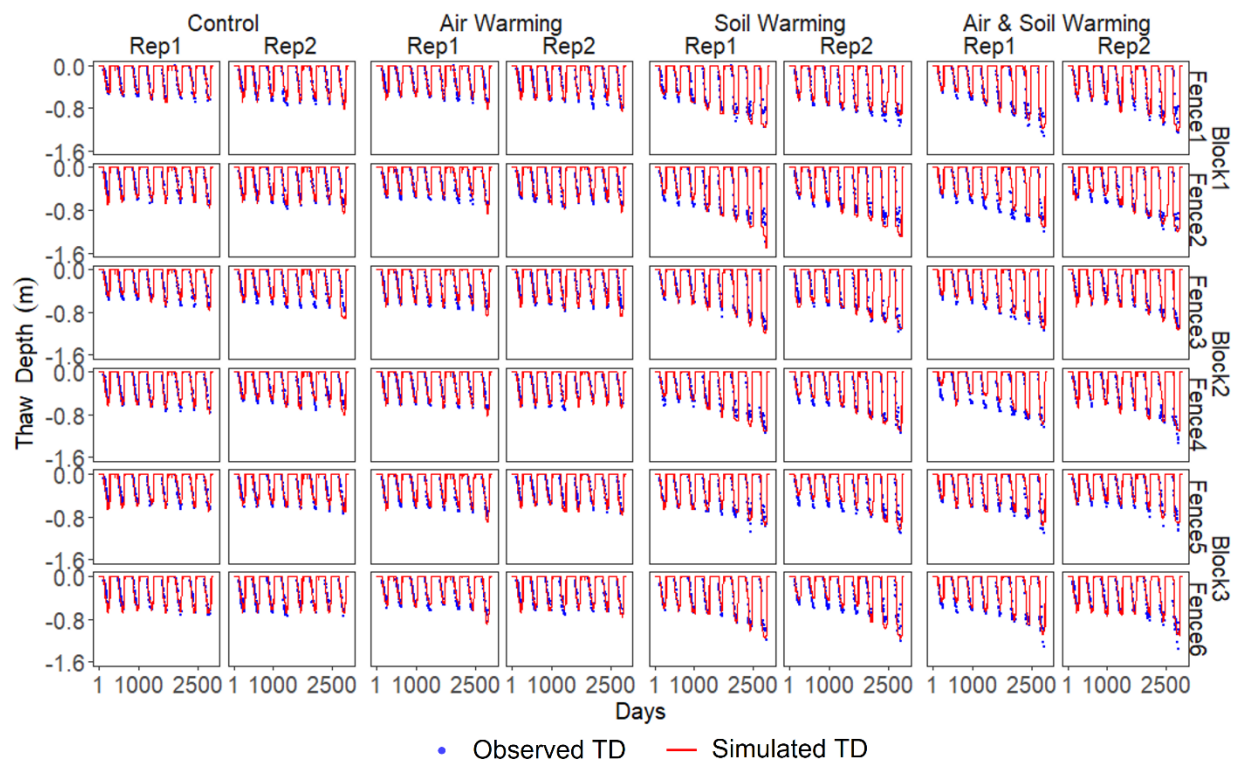


Figure 3: Comparison between simulated daily thaw depth (TD) by TECO model and the observed TD from 2009 to 2016 under control, air warming, soil warming, and air & soil warming treatments across 48 plots. Blue circles represent observed TD and red lines represent simulated TD with optimized parameter values after data assimilation. Every treatment plot repeats two times in each of the six fences that are organized across three blocks.

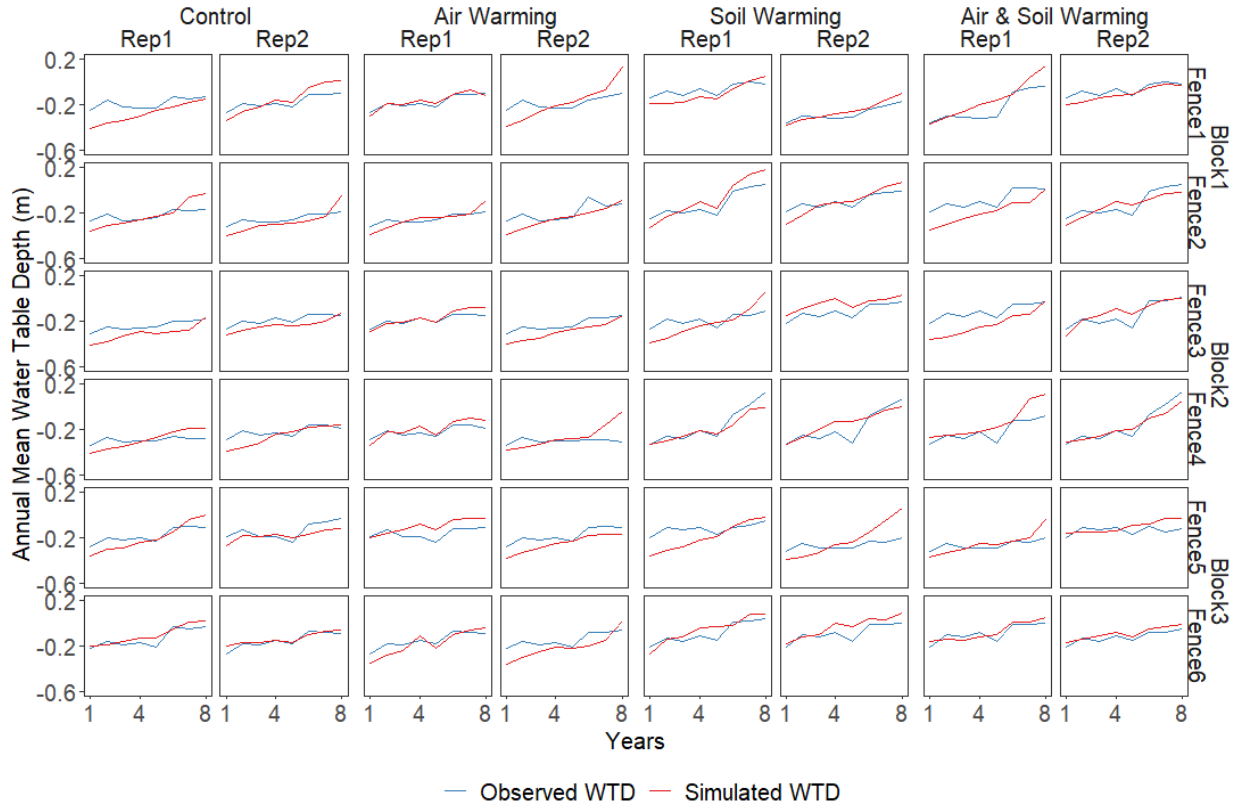


Figure 4: Comparison between simulated yearly water table depth (WTD) by TECO model and the observed WTD from 2009 to 2016 under control, air warming, soil warming, and air & soil warming treatments across 48 plots. Positive WTD means that flooding occurs. Blue lines represent observed WTD, and red lines represent simulated WTD with optimized parameter values after data assimilation. Every treatment plot repeats two times in each of the six fences that are organized across three blocks.

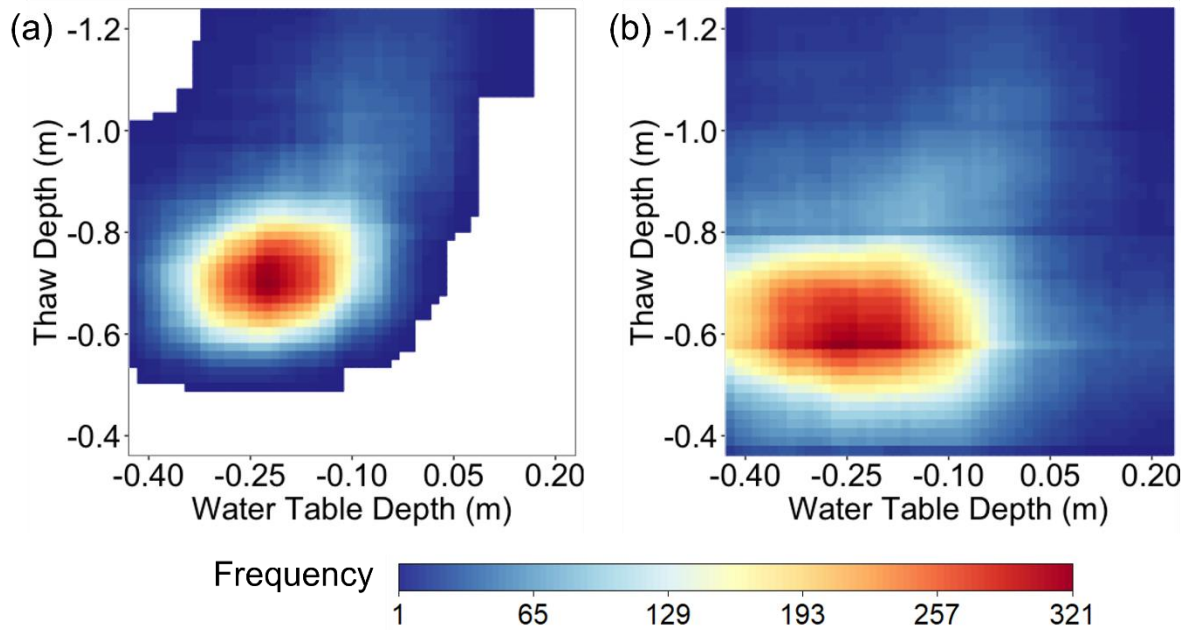


Figure 5: Frequency of the value combinations of water table depth (WTD) and thaw depth (TD) in (a) observations and (b) extended simulations with TECO model after data assimilation. The size of each grid is $0.1m \times 0.1m$. The color in each grid represents the frequency of a particular value combination of WTD and TD. White area outside the blue border in (a) means there is no observation. Compared to observations, the extended simulations include more value combinations of WTD and TD after permafrost thaw, such as shallower TD with elevated WTD and deeper TD with lowered WTD.

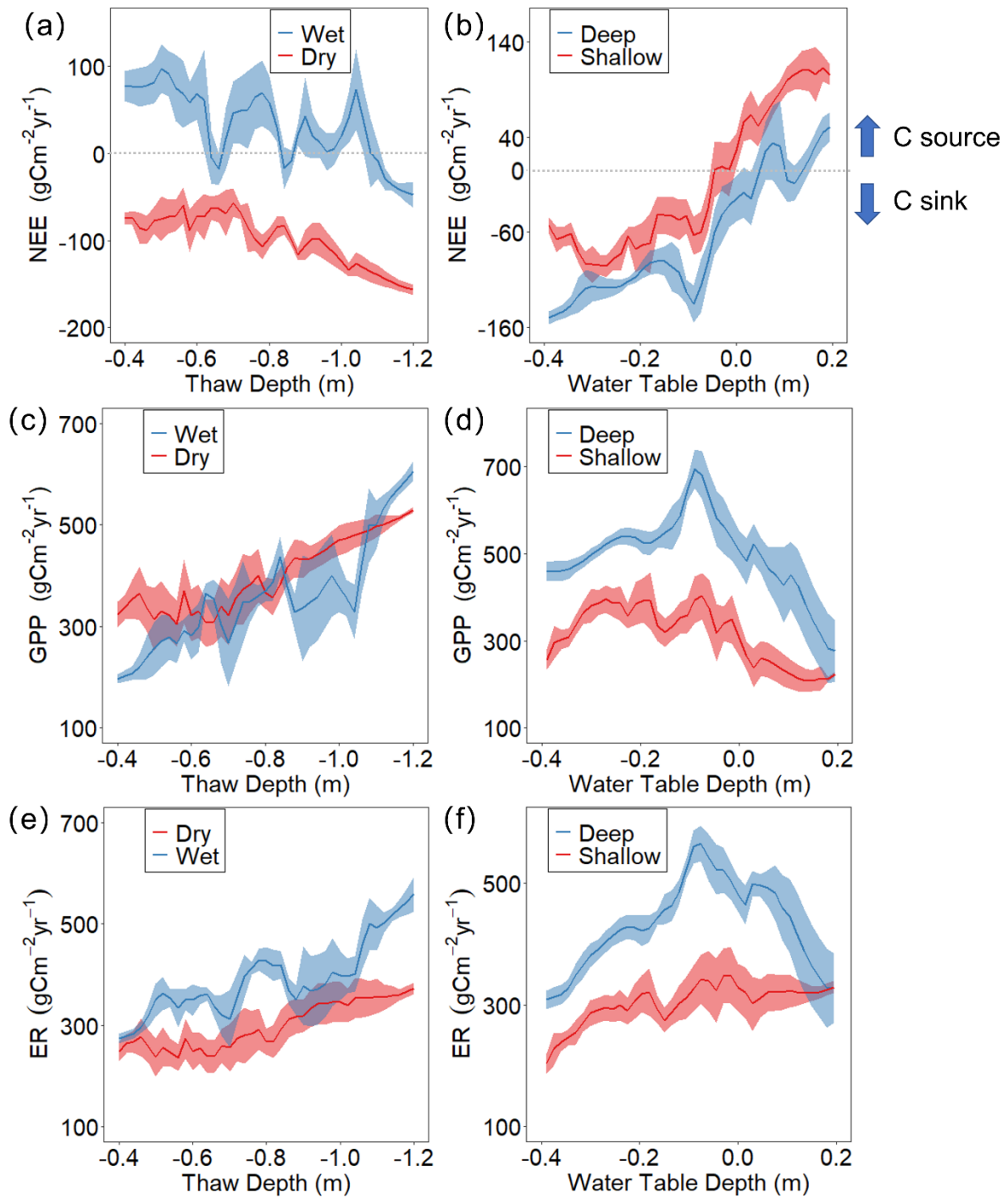


Figure 6: Responses of NEE to (a) thaw depth (TD) in wet (blue) and dry soil (red) and (b) water table depth (WTD) with deep (blue) and shallow thaw (red); responses of GPP to (c) TD in wet (blue) and dry soil (red) and (d) WTD with deep (blue) and shallow thaw (red); and

responses of ecosystem respiration (ER) to (e) TD in wet (blue) and dry soil (red) and (f) WTD with deep (blue) and shallow thaw (red). All variables are from the extended simulations (Fig. 5) by TECO model. The wet and dry soils in (a, c, e) indicate that WTD ranges from 0 to 0.15 and -0.4 to -0.25, respectively. The deep and shallow thaws in (b, d, f) mean that TD ranges from -1.18 to -0.95 and -0.6 to -0.38, respectively. The shadow represents a 95% confidence interval. Positive WTD means that flooding occurs. Positive NEE means C source while negative NEE means C sink.

CHAPTER 4: A NEW ASSESSMENT OF FORCING CORRECTION AND ITS POTENTIAL TO GUIDE ITERATIVE PREDICTIONS IN A BIG-DATA EAR

Xin Huang^{1,2}, Lifen Jiang³, Shuang Ma⁴, Tianan Zhang⁵, Andrew Richardson^{1,2}, Yiqi Luo³

1 Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, AZ, USA

2 School of informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ, USA

3 School of Integrative Plant Science, Cornell University, Ithaca, NY, USA

4 Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

5 School of Computer, Jiangsu University of Science & Technology, Zhenjiang, Jiangsu, China

Correspondence to: Xin Huang (xh59@nau.edu)

Abstract

Reliable ecological predictions are essential to inform decision-makers in order to mitigate climate change. One of the largest uncertainties in ecological prediction is what future forcing may be. Constraining forcing uncertainty and assessing the changes in prediction accuracy are imperative to understand model performances and forecast accuracy. However, until now, forcing correction and associated assessments have not been explicitly explored. In this paper, we proposed a new assessment of forcing correction in terms of forecast accuracy. It is based on comparisons of forecast with future forcing and hindcast with real forcing. We applied this assessment to predict the carbon © cycles at The Spruce and Peatland Responses Under Changing Experiments (SPRUCE) from 2015 to 2024. We iteratively conduct a forcing correction every week from 2015 to 2022 and use the assessment method to evaluate its effect on the forecast accuracy of C-flux and C-pool response variables. We found that the conventional assessment of forcing uncertainty is not proper to evaluate forecast accuracy. Forcing correction on a weekly basis improves the forecast accuracy of pool-based variables but does not impact the forecast accuracy of flux-based variables. Overall, this study provides important insights into the evaluation of forcing correction on forecast accuracy and can inform future research in ecological predictions.

Keywords: forcing uncertainty, forcing correction, assessment, ensemble prediction, C cycles

1. Introduction

With ongoing rapid climate change, the sustainability of human society is under unprecedented thread. A central challenge is predicting the future ecosystem states, ecosystem services, and the feedback of human decisions on climate change. Addressing this question relies on the predictive capability of Earth System Models (ESMs) (Clark et al., 2001). However, to date, the predictions of ESMs spread to a great extent, which precludes a robust assessment of future states on Earth. For example, the predicted soil carbon (C) storage in the next 100 years ranges from a C loss of 72 Pg to a C gain of 253 Pg by ESMs participating in Coupled Model Intercomparison Project Phase 5 (CMIP5) (Todd-Brown et al., 2013). This wide spread of predicted land C storage has not changed significantly over the past decades (Friedlingstein et al., 2014; Arora et al., 2020; Wei et al., 2022). As one of key uncertainty sources in prediction, forcing is gaining increasing attention (Bonan and Doney, 2018; Bonan et al., 2019; Friedlingstein et al., 2006). To constrain forcing uncertainty, some studies have applied forcing correction using extensive observations (Dobor et al., 2015; Ahlström et al., 2017). However, it is still unclear to us how well the forecast accuracy of ESMs has been improved due to forcing correction. Without this assessment of forcing correction, it will be difficult to gauge the effectiveness of our efforts on improving predictable capability, leading to poor decision-making and costly mistakes. Thus, it is imperative to develop an assessment to evaluate the effects of forcing correction on improvements in forecast accuracy.

The predictable capability of ESMs depends on the uncertainty sources in predictions, which can be broadly classified into initial conditions, model structure, parameters, and external forcing (Bonan and Doney, 2018; Luo et al., 2016; Dietze, 2017b). Uncertainty in initial conditions (i.e., uncertain initial values of state variables), which is more crucial in a chaotic

system (e.g., weather prediction), can be estimated by long model simulations (i.e., spin-up) (Xia et al., 2012). Different model structures or individual processes of the C cycle will lead to divergent results (Jiang et al., 2015; Wieder et al., 2015). Traceability analysis provides an opportunity to systematically diagnose model uncertainty propagating from each component (Zhou et al., 2021). Empirical parameter values and *ad hoc* parameterization also result in considerable uncertainty in predictions (Luo et al., 2011). Data assimilation (DA) is widely used to constrain parameter uncertainty and improve prediction performance (Richardson et al., 2010; Hararuk et al., 2014). External forcing (or model input, driver) has different categories including meteorological variables, disturbances (e.g., fire), environmental variables towards different scenarios (e.g., elevated atmospheric CO_2), etc. (Luo et al., 2015).

Forcing uncertainty can strongly affect projects/predictions among multiple models (Scholze et al., 2006; Hardouin et al., 2022; Ahlström et al., 2015) or within a single model (Berthelot et al., 2003; Schaphoff et al., 2006). Meanwhile, some studies indicated that forcing uncertainty dominates forecast uncertainty in the C cycles (Bonan et al., 2019; Zhao et al., 2012). For example, the study of Hardouin et al. (2022) analyzed the relative uncertainty contributions of three models and three forcings. They found forcings uncertainty plays a dominant role in gross primary productivity (GPP) explaining 75% of the variability. Thomas et al., (2018) reported the effect of forcing uncertainty is similar to that of parameter uncertainty in predicting the increases in loblolly pine forest productivity.

Compared to the significance of forcing uncertainty, little effort has been made to correct forcing bias and to assess the effect of forcing correction on forecast accuracy in the ecological field. Future forcings for ecological prediction are often sampled from historical climate databases (White et al., 2019; Cook et al., 2010) or derived from simulations of ESMs/general

circulation models (GCMs) (Heilman et al., 2022; Thomas et al., 2018). The simulated forcing variables by ESMs/GCMs contain substantial bias compared with historical meteorological or climate measurements (Li et al., 2013; Mueller and Seneviratne, 2014). For example, Ahlström et al., (2013) indicated that forcing variations simulated by GCMs is more influential than different CO_2 concentrations in predicting future terrestrial C balance. One reason is that GCMs run at a coarse spatial resolution and it is difficult for simulated forcings to capture meteorological features at a higher resolution (Castro et al., 2005). A corresponding forcing correction is to use regional climate models (RCMs) to generate forcings at higher resolution in the spatial scale (Wang et al., 2004). This bias in simulated forcings will inevitably translate into predictions of carbon and water cycles (Berthelot et al., 2005; Ukkola et al., 2016). Based on an assumption that the future climatology will probably resemble the historical climatology, the past climate records can provide a basis of forcing correction. For example, the study of Ahlström et al. (2017) used 30-year climatology (annual and seasonal) of observations from CRU TS 3.0 database for forcing correction. After that, the divergence in the predicted C cycle responses significantly declined. However, this forcing correction cannot use historical climatology on longer timescales (e.g., decadal).

To tackle forcing bias, another important aspect is to design an assessment of forcing correction in terms of forecast accuracy. Conventional assessments of forcing uncertainty are based on the concept of an ensemble. That is, the spread of model output (e.g., C flux) with ensemble forcings represents the contribution of forcing uncertainty or sensitivity of prediction to the variation of forcings. After forcing correction, a reduction in the spread of the model output indicates that forcing uncertainty is well constrained (Ahlström et al., 2017). However, this assessment does not compare predictions with reality. Thus, it does not directly measure

forecast accuracy. On the other hand, comparison with reality requires collecting observations according to the frequency of predictions, which is not realistic, for example in the tundra of Alaska. It is also inherently difficult to assess the accuracy of a long-term prediction before the predicted date arrives. Instead of observations, in this study, we proposed a new assessment based on a comparison of forecast and hindcast. Forecasting is a prediction with future forcing, while hindcasting is a prediction with real forcing (Jolliffe and Stephenson, 2012). Because hindcast eliminates forcing uncertainty, a comparison of forecast and hindcast will be used to estimate the improvement in forecast accuracy due to forcing correction (i.e., replace the random forcing with real forcing).

In this study, we developed a hindcast-based assessment method to evaluate the effect of forcing correction on the forecast accuracy. This paper is organized as follows. In Section 2, we first introduce the method details, underlying methodology, and how to apply this method to a study site. In Section 3, results will be presented. In section 4, we discuss what we have found in applications of the new assessment method and discuss potential concerns.

2. Method

2.1 Site description and measurements

We took the Spruce and Peatland Responses Under Climatic and Environmental change (SPRUCE) experiment as a case study. The SPRUCE experiment is designed to assess the combined responses of a northern peatland ecosystem to multiple-level whole-ecosystem warmings (+0, +2.25, +4.5, +6.75, and +9 °C) at ambient (~400 ppm) and elevated atmospheric CO₂ concentrations (~900 ppm) (Hanson et al., 2017). This experiment site (N47°30.476', W 93°27.162', 418 m above mean sea level) is located on the 8.1 ha S1 bog in the U.S. Department of Agriculture (USDA) Forest Service Marcell Experiment Forest (MEF), Northern Minnesota,

USA. The climate is subhumid continental with rapid diurnal and seasonal temperature fluctuations (Verry et al., 2011). The mean annual temperature at the SPRUCE site during 1961-2009 was 3.3 °C with an increase of ~0.4 °C per decade. The mean annual precipitation was 785 mm and one third of that occurs as snowfall (Sebestyen et al., 2011). Dominant vegetation at the SPRUCE site include two tree species (*Picea mariana* and *Larix laricina*), ericaceous shrubs, and *Sphagnum* sp. moss. The mean peat depth is around 2-3m with a maximum depth of 11m (Parsekian et al., 2012). Soil is mainly composed of organic matter and the water table fluctuates within the top 30 cm (Verry et al., 2011; Shi et al., 2015).

Whole-ecosystem warming operates from deep soil to tree canopy with exposure to elevated CO_2 in open-top enclosure infrastructures (12m × 8m). Deep soil warming was initiated in June 2014, and above-ground warming started in August 2015. Elevated CO_2 began in some plots in June 2016. Because our research goals are not to explore warming or elevated CO_2 effects, we will only focus on the ambient plots and the target variables are C fluxes (i.e., NEE, GPP, and ER) and C pools (i.e., Leaf, Wood, Fine root, and Soil summarizing fast, slow, and passive soil pools) during 2011-2019.

The community-level flux measurement systems collect CO_2 exchange using open-path analyzers in an area of 1.13 m² within each treatment plot. Flux measurements with a temporary black plastic cover are to estimate pseudo-nighttime respiratory (ER). Uncovered measurements are to estimate NEE and then GPP is calculated as the sum of NEE and ER. During 2011-2019, C flux measurements were collected periodically during the growing season in the S1-Bog. Measurements from 2011-2013 are at two collars outside SPRUCE and measurements under ambient treatment in 2014 are missing (Hanson et al., 2016). Thus, C fluxes measurements for ambient plots span from 2015 - 2019, whereas measurements for warming plots with/without

CO₂ elevations range from 2013 to 2019. C pool measurements are collected annually. In the summer of each year, vegetation was harvested from the plots and then processed for estimation of C pool measurements. Details about other measurements (e.g., water table depth, CH₄) can be found in Ma et al. (2022). Measurements during 2011-2017 are used to constrain model's parameter uncertainty through data assimilation and measurements from 2015 - 2019 are to validate the general patterns of forecasting (forecast period is 2015-2024).

Environmental measurements we collected at the SPRUCE site are mainly model forcings including air temperature (T_{air}), soil temperature (T_{soil}), relative humidity (RH), precipitation, wind speed (WS), and photosynthetically active radiation (PAR) (Table D1). These environmental measurements are collected at different soil layers (half hourly at 7 depths), heights above the bog surface (0.5, 1, and 2 m), and locations (away from shading, near *Sphagnum* surface, or ~+30ms). We generally average relevant environmental measurements to match the requirements of model simulations. There is also a MEF meteorological station at southern S1-Bog to provide long-term historical environmental measurements (Sebestyen et al., 2011). We will use these historical environmental measurements from 1961 to 2014 monitored at the MEF meteorological station to generate future forcings (2015-2024) for ensemble prediction, and then use measurements from 2015 to 2021 collected at the SPRUCE site for forcing correction.

2.2 Overview of model structure and calibrated parameters

Terrestrial ECOsystem (TECO) model is a process-based ecosystem model, which has been used at the SPRUCE site in several studies (Ma et al., 2022; Huang et al., 2019; Jiang et al., 2018; Ma et al., 2017). TECO simulates processes of canopy photosynthesis, vegetation dynamics, soil water dynamics, soil thermal dynamics, soil Carbon/Nitrogen (C/N) transfer, and CH₄ dynamics.

Canopy photosynthesis, which integrates surface energy, carbon and water fluxes, is based on a two-big-leaf model (Wang and Leuning, 1998). Leaf photosynthesis is simulated based on a biochemical model (Farquhar et al., 1980) and a stomatal conductance model (Ball et al., 1987). Vegetation dynamic processes include vegetation growth, phenology, and C allocations among different C pools. Phenology is manifested by the seasonal changes in leaf area index (LAI), where LAI values below 0.1 indicate the end of a growing season. Leaf onset is determined by growing degree days (GDD). Turnover rates of C pools or C fluxes leaving pools are determined by residence time and C pool size. Soil water dynamics process divides soil into 10 layers (10 cm thick for the first five layers and 20 cm thick for the rest five layers). Soil water content is regulated by precipitation, snowmelt, evapotranspiration, and runoff. Transpiration and its associated latent heat loss are determined by stomatal conductance and soil water content in layers where root exists. Evapotranspiration and its associated latent heat loss are regulated by first-layer soil moisture and atmospheric demand. Soil thermal processes simulate snow cover, melting, and changes in soil temperature across the 10 layers (Huang et al., 2017). Soil C/N transfer happens through litterfall, litter decomposition, and SOM mineralization (Shi et al., 2016). CH_4 dynamics is about CH_4 production, oxidation, transport within the 10 layers, and emissions (Ma et al., 2017).

Among the above biogeophysical and biogeochemical processes, there are a great number of parameters whose values are empirically estimated, resulting in considerable biases (Luo and Schuur, 2020). To reduce parameter uncertainty, we select 18 parameters related to key C cycle processes (Table A1) and conducted data assimilation (DA) with flux-based and pool-based observations during 2011-2017. The default parameter values and parameter range derive from a study by (Jiang et al., 2018). We applied a Markov Chain Monte Carlo (MCMC) method

(Metropolis et al., 1953) to calibrate parameter values and constrain parameter uncertainty (Figs A2-5). Generally, forecast/hindcast with uncalibrated parameters before DA will incorporate parameter uncertainty, while forecast/hindcast with calibrated parameters after DA will eliminate parameter uncertainty to a great extent. Both forecastings with and without parameter uncertainty will be considered in quantitative assessments of forcing correction. Meanwhile, a qualitative assessment of weekly forcing correction will use optimized parameter values in an iterative forecasting system.

2.3 Ensemble prediction and forcing correction

2.3.1 Stochastic weather generation for ensemble prediction

Ensemble prediction is to forecast future ecological states with a set of future forcings as model inputs. External forcings of the TECO model are six hourly meteorological variables of T_{air} , T_{soil} , Rain, RH, WS, and PAR (Table D1) as mentioned in Section 2.1. Future forcing can be predicted through stochastic weather generation using historical forcing measurements. In this study, we set the forest period as 2015-2024. 100 sets of future forcings (2015-2024) are generated from historical environmental measurements (1961-2014) collected at MEF Southern Meteorological station near the SPRUCE site at S1-Bog.

The 54-year historical measurements provide a long-term climate pattern for the S1-Bog. We first build a vector autoregressive model (VAR) to generate future temperatures and precipitations using an R package RMAWGEN (Pfaff, 2008). Without exogenous variables, daily precipitation is generated from a VAR with an order of three. To match with the time step in the TECO model, we divided daily precipitation by 24 and get hourly precipitation (Waichler and Wigmosta, 2003). With daily precipitation as an exogenous variable, daily minimum and maximum temperatures are generated from a VAR with an order of ten. We then estimate hourly

temperature by interpolating from the range of daily air temperature. Figure C1 shows the characteristics of future temperatures and precipitations. They have the same distribution as the historical temperatures and precipitations. We then generate the other three forcing variables (RH, WS, and PAR) by stochastic sampling from a frequency distribution. For each forcing variable (e.g., RH), its frequency distribution is estimated from historical measurements at a given hour every month. Thus, there are $24 \times 12 \times 3$ frequency distributions. The above procedure generates future ensemble forcings separately, which may introduce biased covariance among forcing variables. The impacts of this biased covariance can be found in the study by Jiang et al. (2018).

2.3.2 An iterative forecasting system with weekly forcing correction

Sampling from historical environmental measurements, the 100 sets of future forcings from stochastic weather generation preserve their own climate trends and variability at the SPRUCE site. However, these future forcings are unknown and this inherent uncertainty in the future forcings will propagate through model simulation, leading to considerable bias (Ahlström et al., 2017). To reduce forcing uncertainty in prediction, we need to do forcing correction.

We will iteratively select a period of available observed meteorology measurements to replace part of the future forcing in the same period. This forcing correction is repeated when new observations become available (e.g., weekly). This correction will be applied to every set of future forcings. Forcing correction will keep continuing until the end date of observed meteorology measurements. If the whole forcing is already known/observed, this prediction is called hindcast (Jolliffe and Stephenson, 2012). In this study, a hindcast is model prediction using real forcing from 2015 to 2022.

In this study, we developed an iterative forecasting system for this weekly forcing correction. Below is a brief introduction to the procedure of an iterative forecast system. The first iteration is a typical model forecast with 100 sets of random forcing, bringing in 100 sets of model outputs $f(0)_1, f(0)_2, \dots, f(0)_{100}$. The 2nd iteration is a model forecast with 100 sets of forcing which is corrected in the first week of 2015. Corresponding model outputs are $f(1)_1, f(1)_2, \dots, f(1)_{100}$. In the 3rd iteration, with 100 sets of forcing which is corrected in the first 2 weeks of 2015, model forecast will generate $f(2)_1, f(2)_2, \dots, f(2)_{100}$. The iterative forecast will end when all real forcings are used for correction. There will be 365 iterations in total.

The forecast system is executed in a high-performance computing (HPC) cluster. Slurm is used to schedule and manage forecast jobs. Each job is a repetitive forecast over 2015-2024 using one set of continuously corrected forcing. Given 100 sets of future forcing, 100 jobs will run simultaneously on different computation nodes. All model outputs will be saved for assessment in the future.

2.4 Assessment of forcing correction with hindcast

Conventional assessments analyze the spread of ensemble forecasts with ensemble forcings (e.g., coefficient variance/CV). Its objective is to assess the contribution of forcing uncertainty to the overall uncertainty or sensitivity of model output to forcing uncertainty. However, given the ensemble spread, we still cannot tell how well the forecasting is (forecast accuracy) unless we compare it to observations. In this study, instead of using observations to assess forecast accuracy, we employ hindcast. The objective is to (1) compare the conventional assessment for forcing uncertainty and the new assessment for forecast accuracy; (2) quantitatively assess the improvement in forecast accuracy due to forcing correction in hindcast; (3) quantitatively assess

which forcing variable dominates this improvement (4) assess whether weekly forcing correction improves forecast accuracy.

2.4.1 Quantitative assessment: KL divergence

Root mean square error (RMSE) is a widely used metric to measure the difference (or errors) between predictions and observations. However, RMSE has many limits. For example, it does not consider the direction of errors, it is sensitive to outliers, and it only focuses on specific data points where observations are available. To provide a more comprehensive assessment in terms of forecast accuracy, we quantify the difference between prediction and observation using Kullback-Leibler (KL) divergence. KL divergence is a measure of the difference between two probability distributions (i.e., simulated GPP and observed GPP) (Kullback and Leibler, 1951) (Eq. 2). A smaller value of KL indicates the two distributions are more similar to each other. KL divergence is often used as a loss function in machine learning to measure how well the learned distribution matches the true observation distribution (Kingma and Welling, 2013).

$$KL = -\sum_{i=1}^n p_i \log \frac{p_i}{q_i} \quad (1)$$

Where p_i and q_i respectively represent the probability distribution of observation and the corresponding simulated variable.

In this study, we calculated KL divergence between forecasting with random forcings and hindcast with real forcings for each flux-based (GPP, NEE, ER) and pool-based variable (Leaf, Wood, Root, Soil) to assess the improvement in forecast accuracy (Objective 2) after the forcing correction. The calculation of KL is repeated in two scenarios where parameter uncertainty is constrained after DA and parameter uncertainty is not constrained before DA. The final quantitative assessment is a sum of KL estimates with and without parameter uncertainty. In addition, this assessment (KL) is repeated with each set of future forcing (100 in total) to

generate the distribution of KL due to the variation in future forcing. A higher value of KL indicates a greater improvement in forecast accuracy of the target model output (i.e., simulated GPP). The higher variance of KL indicates the improvement in forecast accuracy of the target model output is more sensitive to the variation in forcings.

We also conduct a correction in each of the six forcings and keep the others the same, and then repeat the above calculations to quantitatively assess correction in individual forcing will bring in considerable improvement in forecast accuracy (Objective 3).

2.4.2 Qualitative assessment of weekly forcing correction

We developed a graphical evaluation to assess the improvement in forecast accuracy with more reliable forcings through forcing correction every week (Objective 4). Similarly, the assessment in terms of improvements in forecast accuracy is from a comparison with hindcast.

We first selected a target prediction date (i.e., 6/30/2017) from the forecast period (2015-2017). Then we use forecast/hindcast results on that target date from the iterative forecasting system and plot these results against the date to make predictions (weekly from 2015-2022). The predictions made before that target date (2015-6/29/2017) are forecast. Predictions made on and after that date (6/30/2017-2022) are hindcast because this target date is a past day for the date to make predictions and forcing on this target date is already available for forcing correction. The assessment is to qualitatively compare forecasts before the target date and hindcasts after that target date.

Here we will use equations to explain the procedure to select forecast/hindcast results on the target date. As mentioned in Section 2.3.2, the ensemble predictions with continuous i -week forcing correction are $f(i)_1, f(i)_2, \dots, f(i)_{100}$ and $f(i)_j$ is a forecast over 2015-2024 with the j -th set of future forcing. Taking 6/30/2017 as an example of the target date, this target date is the

6948-th day starting from 1/1/2015 (i.e., the beginning of forecast period). We first select $f(1,6948)_1, f(1,6948)_2, \dots, f(1,6948)_{100}$ as the first forecast on the target date, then $f(2,6948)_1, f(2,6948)_2, \dots, f(2,6948)_{100}$ as the second forecast on the target date and this forecast have forcing correction with 1-week observed forcing. Similarly, $f(3,6948)_1, f(3,6948)_2, \dots, f(3,6948)_{100}$ is the third forecast on the target date using a continuous 2-week forcing correction. This selection will continue until 2022 (i.e., the end date of observed forcing).

3. Results

3.1 Comparison between conventional assessment and the new assessment

Figure 1 illustrates the differences between the conventional assessment and the new assessment. The ensemble prediction of a flux-based response variable, such as GPP, has a larger CV (52%) while the ensemble prediction of a pool-based response variable, such as wood, has a smaller CV (1%). The conventional assessment identified that the prediction of GPP is more unreliable compared to the prediction of wood pool. However, the prediction of wood pool is more unreliable as its prediction is far away from the true values by hindcast. Thus, the conventional assessment leads to a contrary conclusion to the new assessment. The conventional assessment is not proper to evaluate forecast accuracy. The forecast accuracy of more C variables is available in Fig. B1.

3.2 Improvement in forecast accuracy due to forcing correction

Improvements in forecast accuracy due to forcing correction were quite different between three flux-based variables (Fig. B1 (A-C)) and four pool-based variables (Fig. B1 (D-G)). After forcing correction, the hindcast result of flux-based variables is still within the value range estimated by the ensemble prediction (Fig. B1 (A-C)). The quantitative assessment in (Fig. 2)

shows that the effect of forcing correction on gross primary production (GPP) and ecology respiration (ER) is less than one but the effect on net ecosystem exchange (NEE) is the greatest with a value of 2.7. Different from flux-based variables, forcing correction can shift the prediction of pool-based variables outside the value range of ensemble prediction to a significantly different value (Fig. B1 (D-G)). This shift increases with C storage/pool size except for the soil pool (Fig. B1 (G)). The quantitative assessment in Fig. 2 matches this shift pattern in Fig. B1 (D-G). The effect of forcing correction on forecast accuracy increases with the size of vegetation pools. Although the spread of ensemble prediction (forcing uncertainty) does not change much, the new assessment indicates the decreasing forecast accuracy, and this decreasing speed is relevant to pool sizes except for the tremendously large soil C pool.

3.3 Improvement in forecast accuracy due to individual forcing correction

Figure 3 decouples the effect of the forcing correction into individual forcing variables. No matter the flux-based or pool-based variables, the effects of individual forcing correction on its forecasting accuracy (Fig. 3) mostly match with the effect estimations in Fig. 2. The corrections in most forcing variables have an impact on flux-based variables, while only corrections of Rain and Tair influence the prediction of wood and root pools. The forcing group that impacts all flux-based variables is PAR, Rain, and Tair, while that kind of forcing group for all pool-based variables is Rain and Tair.

3.4 Weekly forcing correction improves forecast accuracy

Predictions in all panels of Fig. 4 have a similar pattern. Taking Fig. 4A as an example, forecasts before 6/30/2017 are stably divergent. A weekly forcing correction does not impact the ensemble predictions of flux-based variables much during this forecast time. In other words, the forcing uncertainty stays the same regardless of more and more observed forcing measurements being

involved. There is a shift on the target date (6/30/2017). Predictions after that date are hindcast. Given TECO is deterministic, forecasting uncertainty decreases to zero during hindcast time.

Predictions of pool-based variables have a significantly different response to weekly forcing corrections (Fig. 5). With more and more observed forcing, the forecasting of pool-based variables will gradually reach its hindcast result. Meanwhile, the spread in ensemble prediction during forest time does not change much, which means the forecasting uncertainty does not decrease with more observation involved. Another finding is that the hindcast results are much similar to each other regardless of the target date. It is because the C pool sizes are more stable than the C fluxes.

4. Discussion

The predictive capability of ecological C cycles is crucial to help mitigate climate change. Uncertainty in future forcing undermines this predictive capability of C-climate feedback (Huntzinger et al., 2017). The importance of forcing correction is widely acknowledged but it is unclear how well the forecast accuracy becomes after forcing uncertainty is constrained. We developed a hindcast-based assessment to evaluate the effect of forcing correction on forecast accuracy and applied it to the SPRUCE experimental site. We found that the ensemble-based assessment of forcing uncertainty is not proper for evaluating forecast accuracy. Weekly forcing correction improves the forecast accuracy of pool-based variables but does not impact that of flux-based variables.

The concept of forecast accuracy arises from (Dietze, 2017a) and it is defined as the discrepancy between the forecast and a ‘naive’ model result such as an autoregressive process ($y_{t+1} = y_t + \varepsilon$). The forecast accuracy is to evaluate whether the model can provide a more informative prediction different from a random guess. Here, I expanded this concept of forecast

accuracy to evaluate whether the prediction is more reliable. Thus, it is quantified as the discrepancy between the forecast and the true values (i.e., observations).

Hindcast provides an opportunity to evaluate forecast accuracy more efficiently. For example, we do not need to wait until the observations become available. Instead, we can estimate the forecast accuracy whenever the real forcing at the prediction period is ready. The underlying methodology is that hindcast with real forcings will always lead to a more accurate prediction result compared to predictions with random forcings. It is because there is only one true value for future forcing, that is, the observed forcing. With observed forcings, forcing uncertainty can diminish to zero. Similarly, the study of Jiang et al. (2018) used a certain forcing trajectory as model input to constrain the forcing uncertainty to zero. Thus, hindcast can be used to assess the improvement in forecasting accuracy although the hindcast results may not be accurate enough compared with observations. However, different models may produce similar results and different sets of parameters may also lead to similar model outputs. In other words, there is no true value to reduce model uncertainty or parameter uncertainty to zero.

The relatively smaller contribution of parameter calibration in Fig. 2 is partly because the prior value and the parameter range are already well constrained from other DA studies (Jiang et al., 2018; Ma et al., 2022). NEE benefits the most from forcing correction. Compared with the other two C fluxes (i.e., GPP and Reco), plant pools (i.e., Leaf, Wood, Root) benefit more from forcing correction (Fig. 2). These findings in Fig. 2 are also reflected in Fig. 3 with individual forcing correction. PAR has a generally dominant contribution for all C flux variables. It is because the organic matter produced by photosynthesis supports all plant activities. PAR impacts all C cycles though impacting photosynthesis. Because NEE is determined by subtracting the ER from GPP, forcings which are influential on GPP and ER will also impact NEE to a larger extent.

Similarly, the study of (Rafat et al., 2021) found that radiation, soil temperature, wind speed, and soil moisture are governing forcing variables for net C flux in peatlands from Canada. However, soil temperature is not identified as influential in our study. A possible reason is the lack of covariance between different forcing variables.

The different responses of C fluxes and C pools to weekly forcing correction are partly due to ‘ecological memory’. The terminology of ‘memory effect’ is usually used to evaluate the legacy effect that is attributed to a previous climate extreme event (Walter et al., 2013; Frank et al., 2015). Here we use this terminology of ‘ecology memory’ to represent the ability of a response variable (e.g., wood pool) to save and accumulate information inside weekly forcing correction. With higher ‘memory’, pools-based variables can sustain the signals inside real forcing during forcing correction (Fig. 5). On the contrary, C fluxes are sensitive to forcings, and they will fluctuate on a short time scale. As a result, the prediction of C fluxes is decided by forcings nearest the target date in Fig. 4.

This study utilized stochastic weather generator algorithms to generate an ensemble of future forcings that may not be good enough to capture the realistic meteorological features at the SPRUCE site. However, our goal is not to improve the precision of hourly meteorological variables but to identify the discrepancy between forecast and hindcast as the improvement in forecast accuracy due to forcing correction. What we found emphasizes the value of applying hindcast for measuring forecast accuracy, despite bias in ensemble forcings. Overall, this study provides important insights into the evaluation of forcing correction on forecast accuracy and can inform future research in this area.

Appendix A: Data assimilation

Table A1: A summary of the parameters to be calibrated in the TECO model.

Parameter	Description	Unit	Default	Range
SLA	Specific leaf area	$cm^{-2}g^{-1}$	18.52	[10, 150]
$g_{maxleaf}$	Maximum growth rate of foliage	day^{-1}	37.12	[10, 50]
$g_{maxroot}$	Maximum growth rate of root	day^{-1}	19.88	[10, 30]
$g_{maxstem}$	Maximum growth rate of stem	day^{-1}	10.5	[10, 30]
V_{cmax}	Maximum rate of carboxylation	$\mu mol\ m^{-2}s^{-1}$	60.41	[14, 80]
τ_{leaf}	Residence time of C in leaf	year	0.57	[0.5, 3]
τ_{wood}	Residence time of C in wood	year	5.13	[1, 100]
τ_{root}	Residence time of C in root	year	1.3	[0.3, 2]
$\tau_{finelitter}$	Residence time of C in fine litter	year	0.33	[0.1, 0.5]
$\tau_{coarselitter}$	Residence time of C in coarse litter	year	3.9	[1, 20]
$\tau_{fastSOM}$	Residence time of C in fast turnover soil organic matter	year	0.42	[0.05, 0.5]
$\tau_{slowSOM}$	Residence time of C in slow turnover soil organic matter	year	844.09	[5, 1000]
$\tau_{passiveSOM}$	Residence time of C in passive soil organic matter	year	1243.56	[190, 4000]
gdd_{onset}	Minimum degree-days for plant leaf out	$^{\circ}C$	123.26	[100, 160]
Q_{10}	Sensitivity of respiration to temperature	-	3.55	[1, 4]
$r_{0\ leaf}$	Baseline respiration of foliage	$\mu mol\ m^{-2}s^{-1}$	10.33	[10, 45]
$r_{0\ stem}$	Baseline respiration of stem	$\mu mol\ m^{-2}s^{-1}$	6.76	[4.5, 10.5]
$r_{0\ root}$	Baseline respiration of root	$\mu mol\ m^{-2}s^{-1}$	15.45	[10, 45]

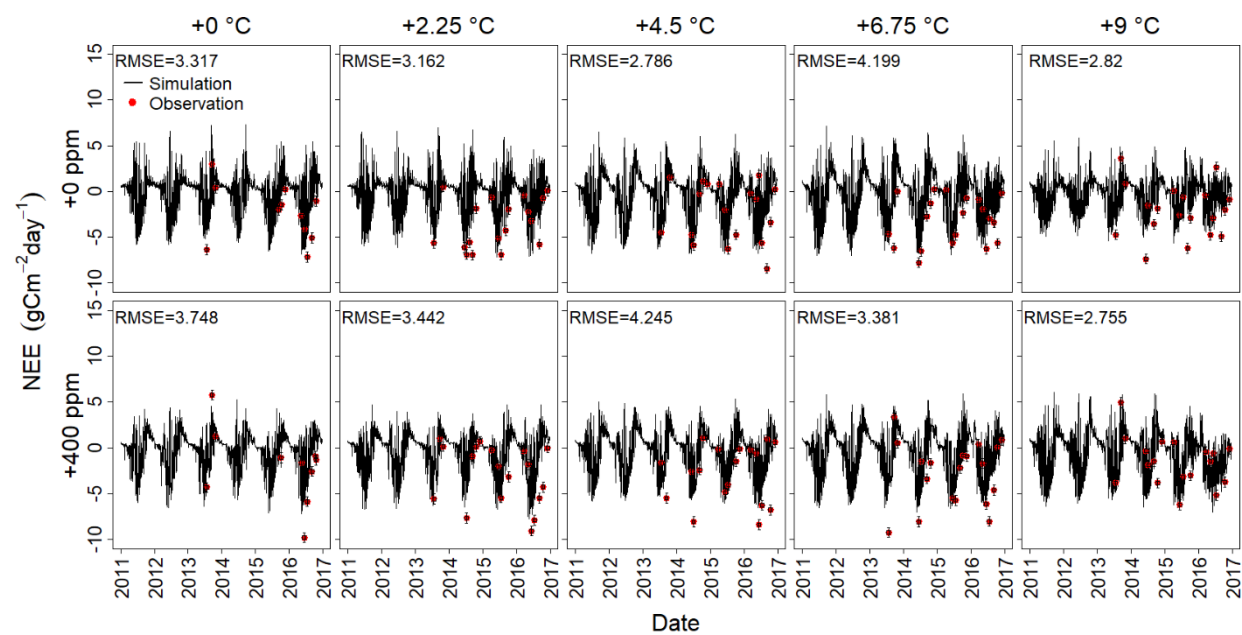


Figure A1: Comparison between the simulated NEE with optimized parameters after DA and the observed NEE under five warming treatments and two elevated CO_2 treatments from 2011 to 2017 at the SPRUCE site.

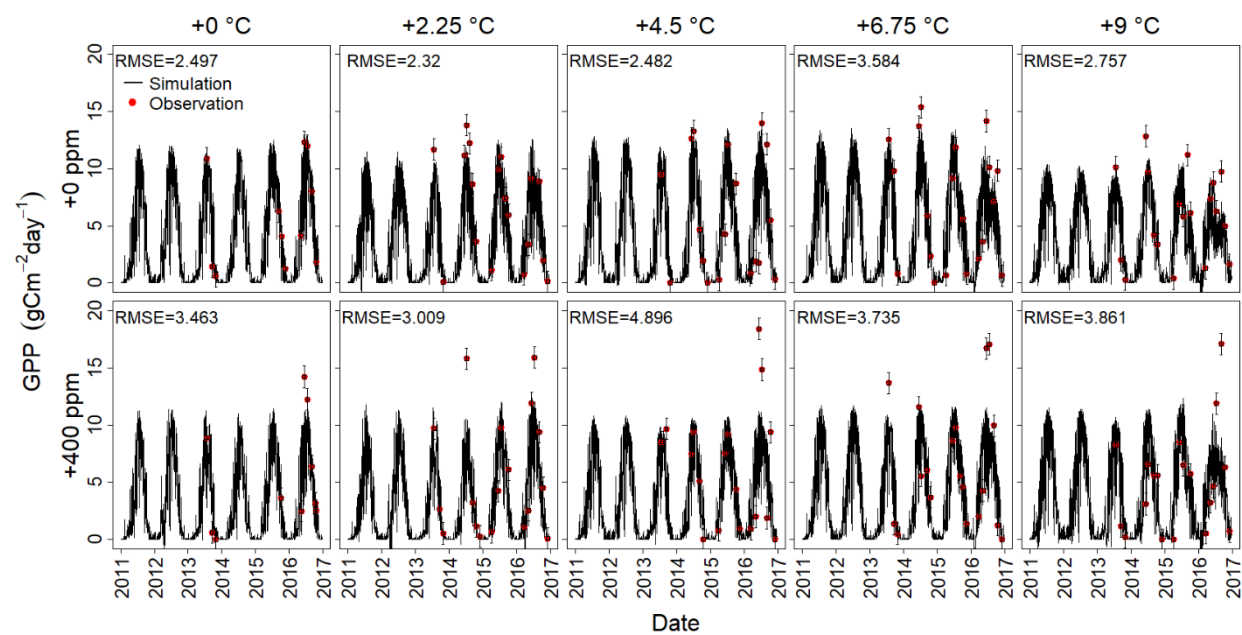


Figure A2: Comparison between the simulated GPP with optimized parameters after DA and the observed GPP under five warming treatments and two elevated CO_2 treatments from 2011 to 2017 at the SPRUCE site.

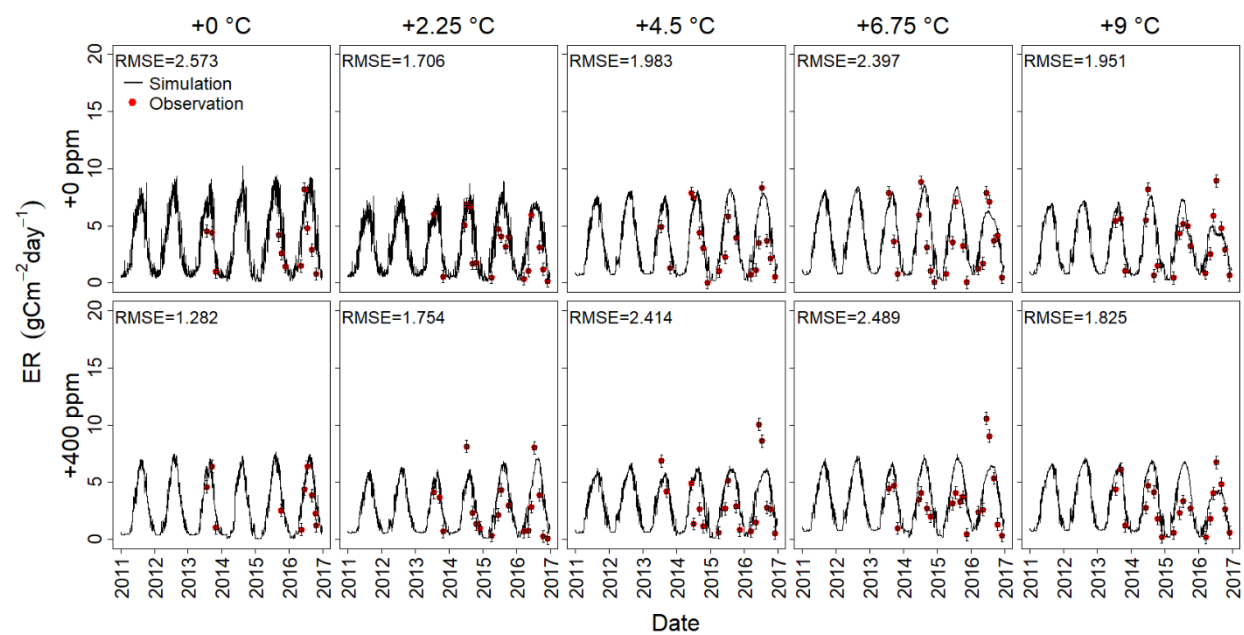


Figure A3: Comparison between the simulated ER with optimized parameters after DA and the observed ER under five warming treatments and two elevated CO_2 treatments from 2011 to 2017 at the SPRUCE site.

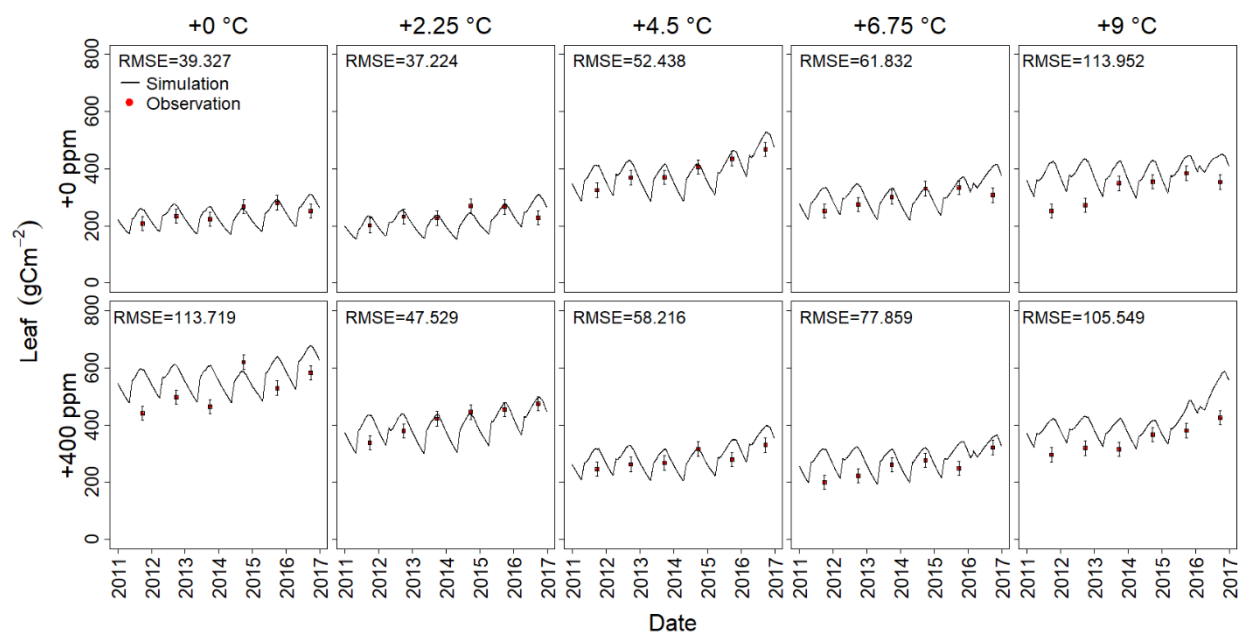


Figure A4: Comparison between the simulated leaf carbon pool with optimized parameters after DA and the observed leaf carbon pool under five warming treatments and two elevated CO_2 treatments from 2011 to 2017 at the SPRUCE site.

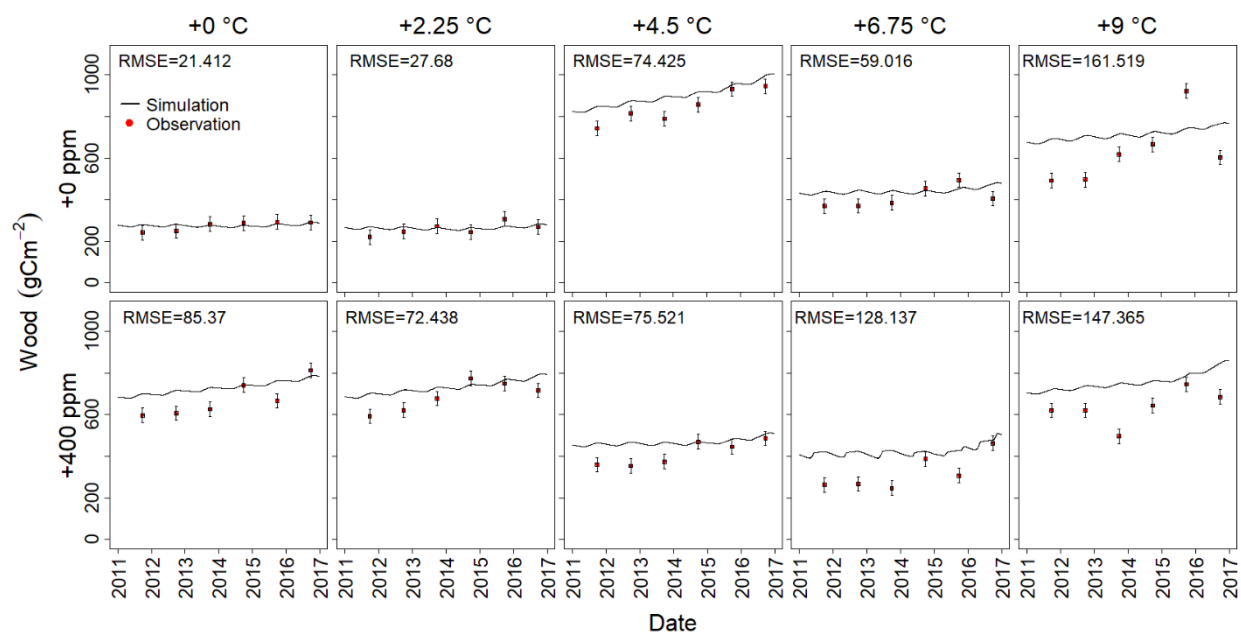


Figure A5: Comparison between the simulated wood carbon pool with optimized parameters after DA and the observed wood carbon pool under five warming treatments and two elevated CO_2 treatments from 2011 to 2017 at the SPRUCE site.

Appendix B: Comparison between ensemble forecast and hindcast

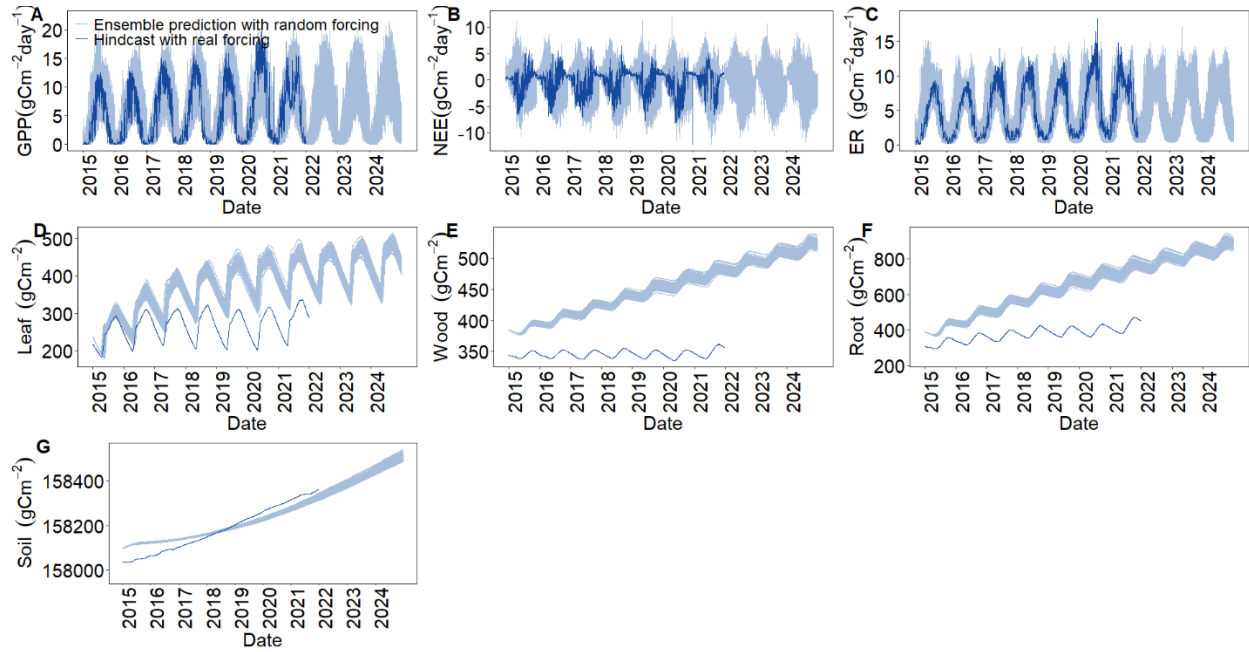


Figure B1: Comparison between forecast with ensemble future forcing (light blue, 2015 to 2024) and hindcast with real forcing (dark blue, 2015 to 2022) for flux-based (A: GPP, B: NEE, and C: ER) and pool-based (D: Leaf, E: Wood, F: Root, and G: Soil) variables in the ambient plot at SPRUCE site.

Appendix C: Comparison between historical forcing and ensemble forcing

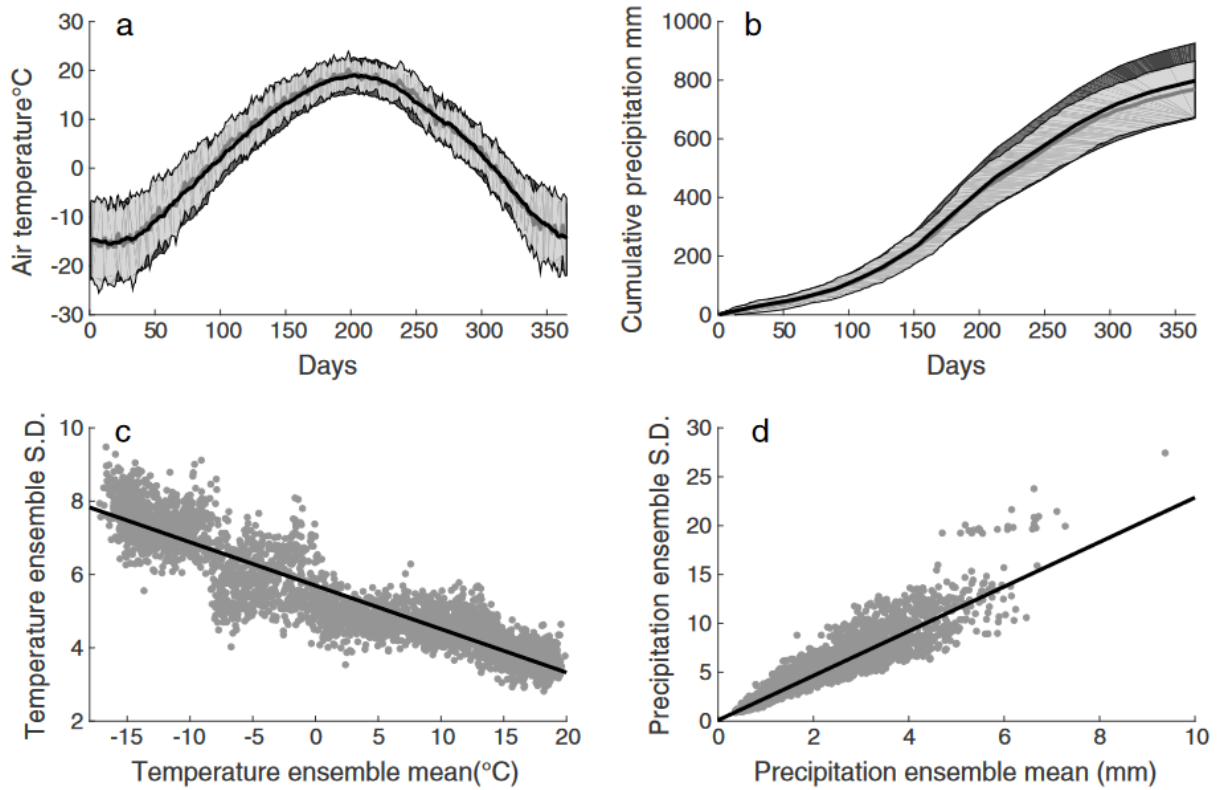


Figure C1: Comparison between historical forcing from the USDA MEF site during 1961 – 2014 and ensemble forcing generated for 2015 – 2024. The gray shaded area represents historical forcing, and the black area represents ensemble forcing in (a, b). (a) Distribution of daily air temperature. (b) cumulative precipitation. (c) and (d) standard deviation of the ensemble against means of daily air temperature and precipitation, respectively. (Credits from Jiang et al., 2017)

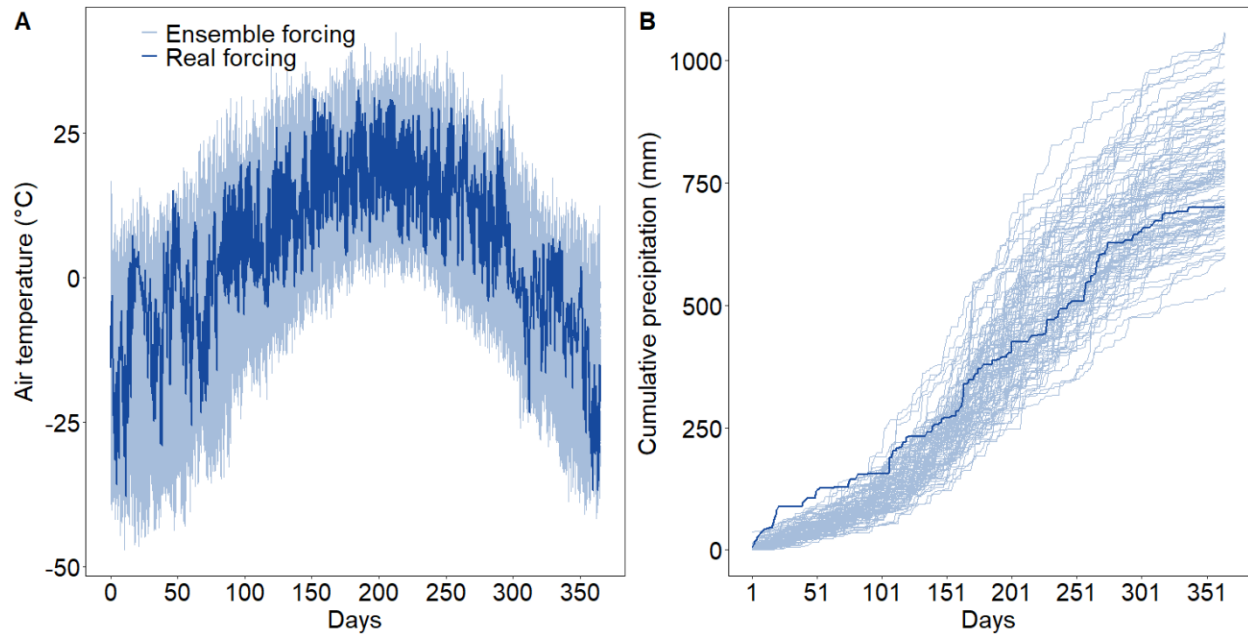


Figure C2: Comparison between historical forcing (dark blue) and ensemble forcing (light blue) in 2017 in the ambient plot at the SPRUCE site. (A) daily air temperature. (B) cumulative precipitation.

Appendix D: description of forcing variables

Table D1: A summary of forcing variables used in the TECO model

Variable	Description	Unit
T_{air}	Air temperature	$^{\circ}C$
T_{soil}	Soil temperature	$^{\circ}C$
RH	Relative humidity	%
Rain	Precipitation including winter snow	mm
WS	Wind speed	m/s
PAR	Photosynthetically active radiation	$\mu mol\ m^{-2}s^{-1}$

References

- Ahlström, A., Smith, B., Lindström, J., Rummukainen, M., and Uvo, C. B.: GCM characteristics explain the majority of uncertainty in projected 21st century terrestrial ecosystem carbon balance, *Biogeosciences*, 10, 1517–1528, <https://doi.org/10.5194/bg-10-1517-2013>, 2013.
- Ahlström, A., Xia, J., Arneeth, A., Luo, Y., and Smith, B.: Importance of vegetation dynamics for future terrestrial carbon cycling, *Environ. Res. Lett.*, 10, 054019, <https://doi.org/10.1088/1748-9326/10/5/054019>, 2015.
- Ahlström, A., Schurgers, G., and Smith, B.: The large influence of climate model bias on terrestrial carbon cycle simulations, *Environ. Res. Lett.*, 12, 014004, <https://doi.org/10.1088/1748-9326/12/1/014004>, 2017.
- Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedlingstein, P., Schwinger, J., Bopp, L., Boucher, O., Cadule, P., Chamberlain, M. A., Christian, J. R., Delire, C., Fisher, R. A., Hajima, T., Ilyina, T., Joetzjer, E., Kawamiya, M., Koven, C. D., Krasting, J. P., Law, R. M., Lawrence, D. M., Lenton, A., Lindsay, K., Pongratz, J., Raddatz, T., Séférian, R., Tachiiri, K., Tjiputra, J. F., Wiltshire, A., Wu, T., and Ziehn, T.: Carbon–concentration and carbon–climate feedbacks in CMIP6 models and their comparison to CMIP5 models, *Biogeosciences*, 17, 4173–4222, <https://doi.org/10.5194/bg-17-4173-2020>, 2020.
- Ball, J. T., Woodrow, I. E., and Berry, J. A.: A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental conditions, in: *Progress in photosynthesis research: volume 4 proceedings of the VIIth international congress on photosynthesis providence, Rhode Island, USA, august 10–15, 1986*, 221–224, 1987.
- Berthelot, M., Friedlingstein, P., Ciais, P., Dufresne, J. L., and Monfray, P.: How uncertainties in future climate change predictions translate into future terrestrial carbon fluxes, 5706, 2003.
- Berthelot, M., Friedlingstein, P., Ciais, P., Dufresne, J.-L., and Monfray, P.: How uncertainties in future climate change predictions translate into future terrestrial carbon fluxes, *Global Change Biology*, 11, 959–970, <https://doi.org/10.1111/j.1365-2486.2005.00957.x>, 2005.
- Bonan, G. B. and Doney, S. C.: Climate, ecosystems, and planetary futures: The challenge to predict life in Earth system models, *Science*, 359, eaam8328, <https://doi.org/10.1126/science.aam8328>, 2018.
- Bonan, G. B., Lombardozzi, D. L., Wieder, W. R., Oleson, K. W., Lawrence, D. M., Hoffman, F. M., and Collier, N.: Model Structure and Climate Data Uncertainty in Historical Simulations of the Terrestrial Carbon Cycle (1850–2014), *Global Biogeochemical Cycles*, 33, 1310–1326, <https://doi.org/10.1029/2019GB006175>, 2019.
- Castro, C. L., Pielke Sr., R. A., and Leoncini, G.: Dynamical downscaling: Assessment of value retained and added using the Regional Atmospheric Modeling System (RAMS), *Journal of Geophysical Research: Atmospheres*, 110, <https://doi.org/10.1029/2004JD004721>, 2005.
- Clark, J. S., Carpenter, S. R., Barber, M., Collins, S., Dobson, A., Foley, J. A., Lodge, D. M., Pascual, M., Pielke, R., Pizer, W., Pringle, C., Reid, W. V., Rose, K. A., Sala, O., Schlesinger, W. H., Wall, D. H., and Wear, D.: Ecological Forecasts: An Emerging Imperative, *Science*, 293, 657–660, <https://doi.org/10.1126/science.293.5530.657>, 2001.

- Cook, B. I., Terando, A., and Steiner, A.: Ecological forecasting under climatic data uncertainty: a case study in phenological modeling, *Environ. Res. Lett.*, 5, 044014, <https://doi.org/10.1088/1748-9326/5/4/044014>, 2010.
- Dietze, M.: Ecological Forecasting, Princeton University Press, <https://doi.org/10.1515/9781400885459>, 2017a.
- Dietze, M. C.: Prediction in ecology: a first-principles framework, *Ecological Applications*, 27, 2048–2060, <https://doi.org/10.1002/eap.1589>, 2017b.
- Dobor, L., Barcza, Z., Hlásny, T., Havasi, Á., Horváth, F., Ittész, P., and Bartholy, J.: Bridging the gap between climate models and impact studies: the FORESEE Database, *Geoscience Data Journal*, 2, 1–11, <https://doi.org/10.1002/gdj3.22>, 2015.
- Farquhar, G. D., von Caemmerer, S., and Berry, J. A.: A biochemical model of photosynthetic CO₂ assimilation in leaves of C₃ species, *Planta*, 149, 78–90, <https://doi.org/10.1007/BF00386231>, 1980.
- Frank, D., Reichstein, M., Bahn, M., Thonicke, K., Frank, D., Mahecha, M. D., Smith, P., van der Velde, M., Vicca, S., Babst, F., Beer, C., Buchmann, N., Canadell, J. G., Ciais, P., Cramer, W., Ibrom, A., Miglietta, F., Poulter, B., Rammig, A., Seneviratne, S. I., Walz, A., Wattenbach, M., Zavala, M. A., and Zscheischler, J.: Effects of climate extremes on the terrestrial carbon cycle: concepts, processes and potential future impacts, *Glob Chang Biol*, 21, 2861–2880, <https://doi.org/10.1111/gcb.12916>, 2015.
- Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., Cadule, P., Doney, S., Eby, M., Fung, I., Bala, G., John, J., Jones, C., Joos, F., Kato, T., Kawamiya, M., Knorr, W., Lindsay, K., Matthews, H. D., Raddatz, T., Rayner, P., Reick, C., Roeckner, E., Schnitzler, K.-G., Schnur, R., Strassmann, K., Weaver, A. J., Yoshikawa, C., and Zeng, N.: Climate–Carbon Cycle Feedback Analysis: Results from the C4MIP Model Intercomparison, *Journal of Climate*, 19, 3337–3353, <https://doi.org/10.1175/JCLI3800.1>, 2006.
- Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., and Knutti, R.: Uncertainties in CMIP5 Climate Projections due to Carbon Cycle Feedbacks, *Journal of Climate*, 27, 511–526, <https://doi.org/10.1175/JCLI-D-12-00579.1>, 2014.
- Hanson, P. J., Gill, A. L., Xu, X., Phillips, J. R., Weston, D. J., Kolka, R. K., Riggs, J. S., and Hook, L. A.: Intermediate-scale community-level flux of CO₂ and CH₄ in a Minnesota peatland: putting the SPRUCE project in a global context, *Biogeochemistry*, 129, 255–272, <https://doi.org/10.1007/s10533-016-0230-8>, 2016.
- Hanson, P. J., Riggs, J. S., Nettles, W. R., Phillips, J. R., Krassovski, M. B., Hook, L. A., Gu, L., Richardson, A. D., Aubrecht, D. M., Ricciuto, D. M., Warren, J. M., and Barbier, C.: Attaining whole-ecosystem warming using air and deep-soil heating methods with an elevated CO₂ atmosphere, *Biogeosciences*, 14, 861–883, <https://doi.org/10.5194/bg-14-861-2017>, 2017.
- Hararuk, O., Xia, J., and Luo, Y.: Evaluation and improvement of a global land model against soil carbon data using a Bayesian Markov chain Monte Carlo method, *Journal of Geophysical Research: Biogeosciences*, 119, 403–417, <https://doi.org/10.1002/2013JG002535>, 2014.
- Hardouin, L., Delire, C., Decharme, B., Lawrence, D. M., Nabel, J. E. M. S., Brovkin, V., Collier, N., Fisher, R., Hoffman, F. M., Koven, C. D., Séférian, R., and Stacke, T.: Uncertainty in land carbon budget simulated by terrestrial biosphere models: the role of

- atmospheric forcing, *Environ. Res. Lett.*, 17, 094033, <https://doi.org/10.1088/1748-9326/ac888d>, 2022.
- Heilman, K. A., Dietze, M. C., Arizpe, A. A., Aragon, J., Gray, A., Shaw, J. D., Finley, A. O., Klesse, S., DeRose, R. J., and Evans, M. E. K.: Ecological forecasting of tree growth: Regional fusion of tree-ring and forest inventory data to quantify drivers and characterize uncertainty, *Global Change Biology*, 28, 2442–2460, <https://doi.org/10.1111/gcb.16038>, 2022.
- Huang, Y., Stacy, M., Jiang, J., Sundi, N., Ma, S., Saruta, V., Jung, C. G., Shi, Z., Xia, J., Hanson, P. J., Ricciuto, D., and Luo, Y.: Realized ecological forecast through an interactive Ecological Platform for Assimilating Data (EcoPAD, v1.0) into models, *Geosci. Model Dev.*, 12, 1119–1137, <https://doi.org/10.5194/gmd-12-1119-2019>, 2019.
- Huntzinger, D. N., Michalak, A. M., Schwalm, C., Ciais, P., King, A. W., Fang, Y., Schaefer, K., Wei, Y., Cook, R. B., Fisher, J. B., Hayes, D., Huang, M., Ito, A., Jain, A. K., Lei, H., Lu, C., Maignan, F., Mao, J., Parazoo, N., Peng, S., Poulter, B., Ricciuto, D., Shi, X., Tian, H., Wang, W., Zeng, N., and Zhao, F.: Uncertainty in the response of terrestrial carbon sink to environmental drivers undermines carbon-climate feedback predictions, *Sci Rep*, 7, 4765, <https://doi.org/10.1038/s41598-017-03818-2>, 2017.
- Jiang, J., Huang, Y., Ma, S., Stacy, M., Shi, Z., Ricciuto, D. M., Hanson, P. J., and Luo, Y.: Forecasting Responses of a Northern Peatland Carbon Cycle to Elevated CO₂ and a Gradient of Experimental Warming, *Journal of Geophysical Research: Biogeosciences*, 123, 1057–1071, <https://doi.org/10.1002/2017JG004040>, 2018.
- Jiang, L., Yan, Y., Hararuk, O., Mickle, N., Xia, J., Shi, Z., Tjiputra, J., Wu, T., and Luo, Y.: Scale-Dependent Performance of CMIP5 Earth System Models in Simulating Terrestrial Vegetation Carbon*, *Journal of Climate*, 28, 5217–5232, <https://doi.org/10.1175/JCLI-D-14-00270.1>, 2015.
- Jolliffe, I. T. and Stephenson, D. B.: *Forecast verification: a practitioner's guide in atmospheric science*, John Wiley & Sons, 2012.
- Kingma, D. P. and Welling, M.: Auto-encoding variational bayes, arXiv preprint [arXiv:1312.6114](https://arxiv.org/abs/1312.6114), 2013.
- Kullback, S. and Leibler, R.: On information and sufficiency *Annals of mathematical statistics*, 22, 79–86, MathSciNet MATH, 1951.
- Li, J.-L. F., Waliser, D. E., Stephens, G., Lee, S., L'Ecuyer, T., Kato, S., Loeb, N., and Ma, H.-Y.: Characterizing and understanding radiation budget biases in CMIP3/CMIP5 GCMs, contemporary GCM, and reanalysis, *Journal of Geophysical Research: Atmospheres*, 118, 8166–8184, <https://doi.org/10.1002/jgrd.50378>, 2013.
- Luo, Y. and Schuur, E. A. G.: Model parameterization to represent processes at unresolved scales and changing properties of evolving systems, *Global Change Biology*, 26, 1109–1117, <https://doi.org/10.1111/gcb.14939>, 2020.
- Luo, Y., Ogle, K., Tucker, C., Fei, S., Gao, C., LaDeau, S., Clark, J. S., and Schimel, D. S.: Ecological forecasting and data assimilation in a data-rich era, *Ecological Applications*, 21, 1429–1442, <https://doi.org/10.1890/09-1275.1>, 2011.
- Luo, Y., Keenan, T. F., and Smith, M.: Predictability of the terrestrial carbon cycle, *Global Change Biology*, 21, 1737–1751, <https://doi.org/10.1111/gcb.12766>, 2015.
- Luo, Y., Ahlström, A., Allison, S. D., Batjes, N. H., Brovkin, V., Carvalhais, N., Chappell, A., Ciais, P., Davidson, E. A., Finzi, A., Georgiou, K., Guenet, B., Hararuk, O., Harden, J. W., He, Y., Hopkins, F., Jiang, L., Koven, C., Jackson, R. B., Jones, C. D., Lara, M. J.,

- Liang, J., McGuire, A. D., Parton, W., Peng, C., Randerson, J. T., Salazar, A., Sierra, C. A., Smith, M. J., Tian, H., Todd-Brown, K. E. O., Torn, M., van Groenigen, K. J., Wang, Y. P., West, T. O., Wei, Y., Wieder, W. R., Xia, J., Xu, X., Xu, X., and Zhou, T.: Toward more realistic projections of soil carbon dynamics by Earth system models, *Global Biogeochemical Cycles*, 30, 40–56, <https://doi.org/10.1002/2015GB005239>, 2016.
- Ma, S., Jiang, J., Huang, Y., Shi, Z., Wilson, R. M., Ricciuto, D., Sebestyen, S. D., Hanson, P. J., and Luo, Y.: Data-Constrained Projections of Methane Fluxes in a Northern Minnesota Peatland in Response to Elevated CO₂ and Warming, *Journal of Geophysical Research: Biogeosciences*, 122, 2841–2861, <https://doi.org/10.1002/2017JG003932>, 2017.
- Ma, S., Jiang, L., Wilson, R. M., Chanton, J. P., Bridgman, S., Niu, S., Iversen, C. M., Malhotra, A., Jiang, J., Lu, X., Huang, Y., Keller, J., Xu, X., Ricciuto, D. M., Hanson, P. J., and Luo, Y.: Evaluating alternative ebullition models for predicting peatland methane emission and its pathways via data–model fusion, *Biogeosciences*, 19, 2245–2262, <https://doi.org/10.5194/bg-19-2245-2022>, 2022.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E.: Equation of State Calculations by Fast Computing Machines, *J. Chem. Phys.*, 21, 1087–1092, <https://doi.org/10.1063/1.1699114>, 1953.
- Mueller, B. and Seneviratne, S. I.: Systematic land climate and evapotranspiration biases in CMIP5 simulations, *Geophysical Research Letters*, 41, 128–134, <https://doi.org/10.1002/2013GL058055>, 2014.
- Parsekian, A. D., Slater, L., Ntarlagiannis, D., Nolan, J., Sebesteyen, S. D., Kolka, R. K., and Hanson, P. J.: Uncertainty in Peat Volume and Soil Carbon Estimated Using Ground-Penetrating Radar and Probing, *Soil Science Society of America Journal*, 76, 1911–1918, <https://doi.org/10.2136/sssaj2012.0040>, 2012.
- Pfaff, B.: VAR, SVAR and SVEC models: Implementation within R package vars, *Journal of statistical software*, 27, 1–32, 2008.
- Rafat, A., Rezanezhad, F., Quinton, W. L., Humphreys, E. R., Webster, K., and Van Cappellen, P.: Non-growing season carbon emissions in a northern peatland are projected to increase under global warming, *Commun Earth Environ*, 2, 111, <https://doi.org/10.1038/s43247-021-00184-w>, 2021.
- Richardson, A. D., Williams, M., Hollinger, D. Y., Moore, D. J. P., Dail, D. B., Davidson, E. A., Scott, N. A., Evans, R. S., Hughes, H., Lee, J. T., Rodrigues, C., and Savage, K.: Estimating parameters of a forest ecosystem C model with measurements of stocks and fluxes as joint constraints, *Oecologia*, 164, 25–40, <https://doi.org/10.1007/s00442-010-1628-y>, 2010.
- Schaphoff, S., Lucht, W., Gerten, D., Sitch, S., Cramer, W., and Prentice, I. C.: Terrestrial biosphere carbon storage under alternative climate projections, *Climatic Change*, 74, 97–122, <https://doi.org/10.1007/s10584-005-9002-5>, 2006.
- Scholze, M., Knorr, W., Arnell, N. W., and Prentice, I. C.: A climate-change risk analysis for world ecosystems, *Proceedings of the National Academy of Sciences*, 103, 13116–13120, <https://doi.org/10.1073/pnas.0601816103>, 2006.
- Sebestyen, S. D., Dorrance, C., Olson, D. M., Verry, E. S., Kolka, R. K., Elling, A. E., and Kyllander, R.: Long-term monitoring sites and trends at the Marcell Experimental Forest, Peatland biogeochemistry and watershed hydrology at the Marcell Experimental Forest, 1, 15–71, 2011.

- Shi, X., Thornton, P. E., Ricciuto, D. M., Hanson, P. J., Mao, J., Sebestyen, S. D., Griffiths, N. A., and Bisht, G.: Representing northern peatland microtopography and hydrology within the Community Land Model, *Biogeosciences*, 12, 6463–6477, <https://doi.org/10.5194/bg-12-6463-2015>, 2015.
- Thomas, R. Q., Jersild, A. L., Brooks, E. B., Thomas, V. A., and Wynne, R. H.: A mid-century ecological forecast with partitioned uncertainty predicts increases in loblolly pine forest productivity, *Ecological Applications*, 28, 1503–1519, <https://doi.org/10.1002/eap.1761>, 2018.
- Todd-Brown, K. E. O., Randerson, J. T., Post, W. M., Hoffman, F. M., Tarnocai, C., Schuur, E. a. G., and Allison, S. D.: Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations, *Biogeosciences*, 10, 1717–1736, <https://doi.org/10.5194/bg-10-1717-2013>, 2013.
- Ukkola, A. M., Keenan, T. F., Kelley, D. I., and Prentice, I. C.: Vegetation plays an important role in mediating future water resources, *Environ. Res. Lett.*, 11, 094022, <https://doi.org/10.1088/1748-9326/11/9/094022>, 2016.
- Verry, S. E., Boelter, D. H., Päivänen, J., Nichols, S. D., Malterer, T., and Gafni, A.: Physical properties of organic soils.[In:] *Peatland biogeochemistry and watershed hydrology* (Kolka RK, Sebestyen SD, Verry ES, Brooks KN, Editors), 2011.
- Waichler, S. R. and Wigmosta, M. S.: Development of Hourly Meteorological Values From Daily Data and Significance to Hydrological Modeling at H. J. Andrews Experimental Forest, *Journal of Hydrometeorology*, 4, 251–263, [https://doi.org/10.1175/1525-7541\(2003\)4<251:DOHMFV>2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)4<251:DOHMFV>2.0.CO;2), 2003.
- Walter, J., Jentsch, A., Beierkuhnlein, C., and Kreyling, J.: Ecological stress memory and cross stress tolerance in plants in the face of climate extremes, *Environmental and Experimental Botany*, 94, 3–8, <https://doi.org/10.1016/j.envexpbot.2012.02.009>, 2013.
- Wang, Y., Leung, L. R., McGREGOR, J. L., Lee, D.-K., Wang, W.-C., Ding, Y., and Kimura, F.: Regional Climate Modeling: Progress, Challenges, and Prospects, *Journal of the Meteorological Society of Japan*, 82, 1599–1628, <https://doi.org/10.2151/jmsj.82.1599>, 2004.
- Wang, Y.-P. and Leuning, R.: A two-leaf model for canopy conductance, photosynthesis and partitioning of available energy I: Model description and comparison with a multi-layered model, *Agricultural and Forest Meteorology*, 91, 89–111, [https://doi.org/10.1016/S0168-1923\(98\)00061-6](https://doi.org/10.1016/S0168-1923(98)00061-6), 1998.
- Wei, N., Xia, J., Zhou, J., Jiang, L., Cui, E., Ping, J., and Luo, Y.: Evolution of Uncertainty in Terrestrial Carbon Storage in Earth System Models from CMIP5 to CMIP6, *Journal of Climate*, 35, 5483–5499, <https://doi.org/10.1175/JCLI-D-21-0763.1>, 2022.
- White, E. P., Yenni, G. M., Taylor, S. D., Christensen, E. M., Bledsoe, E. K., Simonis, J. L., and Ernest, S. K. M.: Developing an automated iterative near-term forecasting system for an ecological study, *Methods in Ecology and Evolution*, 10, 332–344, <https://doi.org/10.1111/2041-210X.13104>, 2019.
- Wieder, W. R., Cleveland, C. C., Lawrence, D. M., and Bonan, G. B.: Effects of model structural uncertainty on carbon cycle projections: biological nitrogen fixation as a case study, *Environ. Res. Lett.*, 10, 044016, <https://doi.org/10.1088/1748-9326/10/4/044016>, 2015.
- Xia, J. Y., Luo, Y. Q., Wang, Y.-P., Weng, E. S., and Hararuk, O.: A semi-analytical solution to accelerate spin-up of a coupled carbon and nitrogen land model to steady state,

- Geoscientific Model Development, 5, 1259–1271, <https://doi.org/10.5194/gmd-5-1259-2012>, 2012.
- Zhao, Y., Ciais, P., Peylin, P., Viovy, N., Longdoz, B., Bonnefond, J. M., Rambal, S., Klumpp, K., Oliso, A., Cellier, P., Maignan, F., Eglin, T., and Calvet, J. C.: How errors on meteorological variables impact simulated ecosystem fluxes: a case study for six French sites, *Biogeosciences*, 9, 2537–2564, <https://doi.org/10.5194/bg-9-2537-2012>, 2012.
- Zhou, J., Xia, J., Wei, N., Liu, Y., Bian, C., Bai, Y., and Luo, Y.: A traceability analysis system for model evaluation on land carbon dynamics: design and applications, *Ecol Process*, 10, 12, <https://doi.org/10.1186/s13717-021-00281-w>, 2021.

Figures

Figure 1: Comparison between the conventional assessment of forecast accuracy and the new assessment of forcing uncertainty for a flux-based response variable (i.e., GPP) and a pool-based response variable (i.e., Wood pool). The assessment of forecast accuracy is the discrepancy between ensemble forecast (light blue) and hindcast (dark blue). The ensemble prediction of wood pool has a lower accuracy compared with that of GPP. The assessment of forcing uncertainty is to quantify the spread of ensemble predictions at the final year (i.e., 2024) using coefficient variance (CV). The ensemble prediction of wood pool has a smaller forcing uncertainty with a CV value of 1%.

Figure 2: Distribution of KL divergence for flux-based (GPP, net ecosystem exchange (NEE), and ecological respiration (Reco)) and pool-based (Leaf, Wood, Root, and Soil) variables caused by DA (light green) and forcing correction (dark green). The lines in each box show, from top to bottom, the third quartile, median, and the first quartile values of the KL divergence distribution. The ends of the whisker extending from the box are the minimum and maximum values excluding outliers. Outlier points are values significantly different from the remainders of the distribution. Compared quantitatively with DA, forcing correction improves forecasting skill the greatest.

Figure 3: Distribution of KL divergence for flux-based (GPP, NEE, and Reco) and pool-based (Leaf, Wood, Root, and Soil) variables caused by correction in individual forcing (photosynthetically active radiation (PAR, red), rainfall (Rain, brown), relative humidity (RH, green), air temperature (Tair, light blue), soil temperature (Tsoil, dark blue), wind speed (WS, purple)) while other forcings vary in ensemble predictions.

Figure 4: Convergence of three predicted carbon fluxes (GPP: green; ecological respiration (ER): blue, NEE: red) for a certain date (A: 6/30/2017; B: 6/30/2018; C: 6/30/2019; D: 6/30/2020) across time (A: 1/2017-12/2017; B: 1/2017-4/2019; C: 2017-2020; D: 2017-2021). Each line depicts predictions for a given date with weekly corrected forcing. Line segment before the expected date is forecast with future forcing, and line segment after that data is hindcast with real forcing. Ensemble prediction with future forcings is represented by lines of the same color (light green, light blue, or light red), and ensemble mean is shown by lines of a darker hue (dark green, dark blue, or dark red). With weekly forcing updates, flux-based variables do not change much until very close to the expected date.

Figure 5: Convergence of three predicted carbon pools (leaf: purple; wood: orange; root: grey; soil: brown) for a specific date (A: 6/30/2017; B: 6/30/2018; C: 6/30/2019; D: 6/30/2020) over time (A: 1/2017-12/2017; B: 1/2017-4/2019; C: 2017-2020; D: 2017-2021). Each line represents predictions for the specific date with weekly corrected forcing. Line segment before the expected date is forecast with future forcing, and line segment after that data is hindcast with real forcing. Ensemble prediction with future forcings is represented by lines of the same color (light purple, light orange, light grey, or light brown), and ensemble mean is shown by lines of a darker hue (dark purple, dark orange, dark grey, or dark brown). Pool-based variables fluctuate with weekly forcing updates and converge before the expected date.

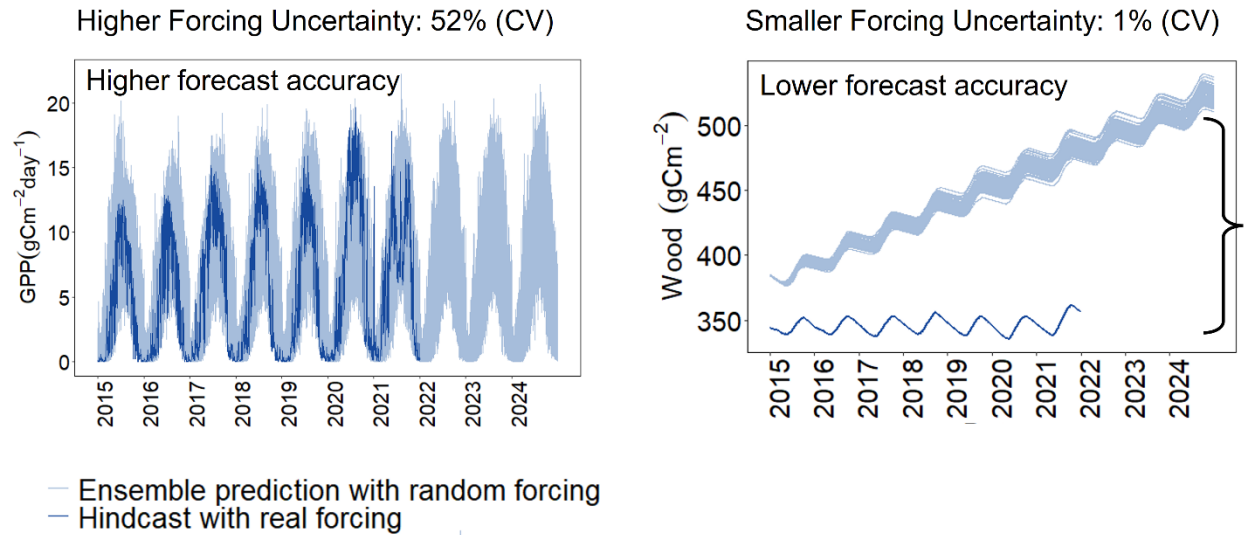


Figure 1: Comparison between the conventional assessment of forecast accuracy and the new assessment of forcing uncertainty for a flux-based response variable (i.e., GPP) and a pool-based response variable (i.e., Wood pool). The assessment of forecast accuracy is the discrepancy between ensemble forecast (light blue) and hindcast (dark blue). The ensemble prediction of wood pool has a lower accuracy compared with that of GPP. The assessment of forcing uncertainty is to quantify the spread of ensemble predictions at the final year (i.e., 2024) using coefficient variance (CV). The ensemble prediction of wood pool has a smaller forcing uncertainty with a CV value of 1%.

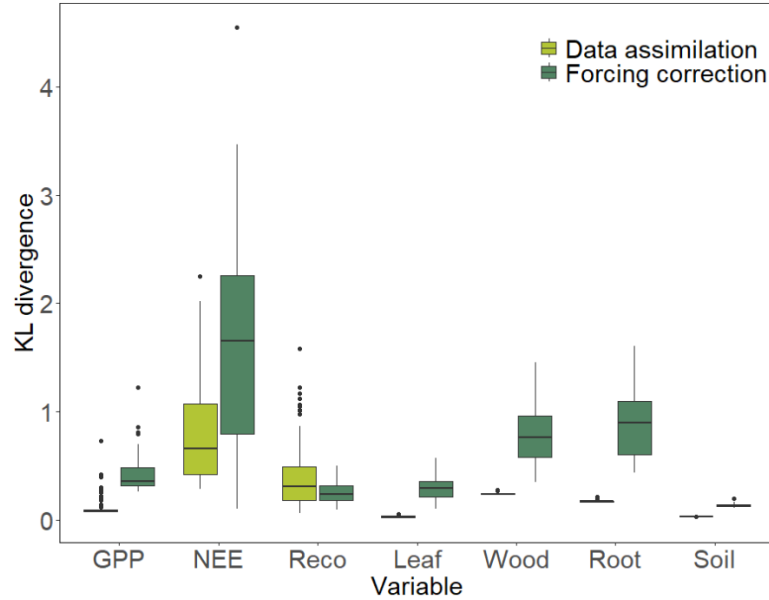


Figure 2: Distribution of KL divergence for flux-based (GPP, net ecosystem exchange (NEE), and ecological respiration (Reco)) and pool-based (Leaf, Wood, Root, and Soil) variables caused by DA (light green) and forcing correction (dark green). The lines in each box show, from top to bottom, the third quartile, median, and the first quartile values of the KL divergence distribution. The ends of the whisker extending from the box are the minimum and maximum values excluding outliers. Outlier points are values significantly different from the remainders of the distribution. Compared quantitatively with DA, forcing correction improves forecasting skill the greatest.

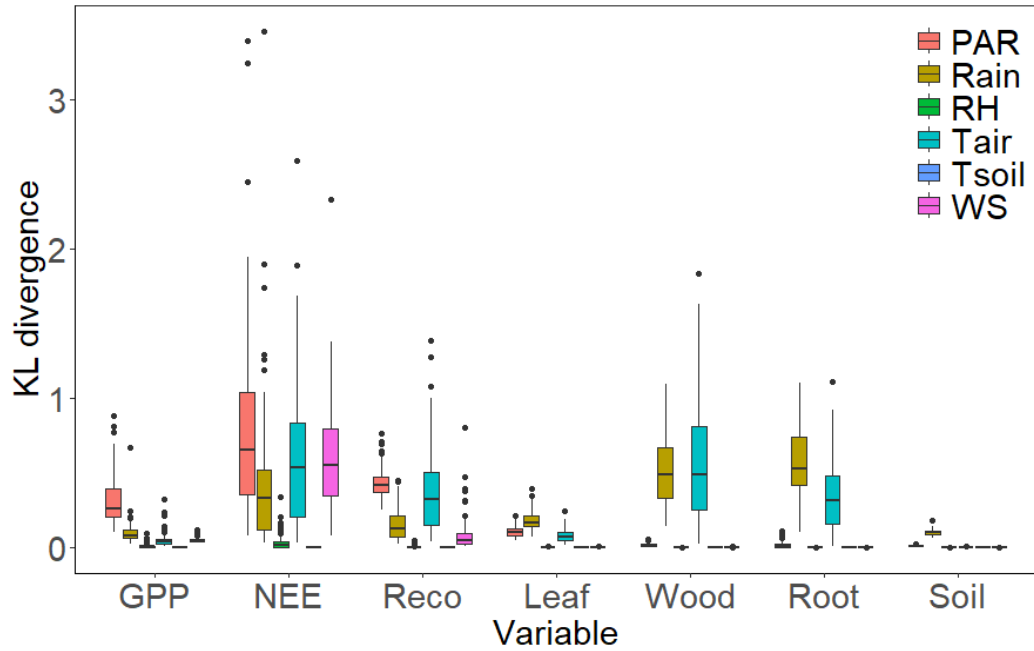


Figure 3: Distribution of KL divergence for flux-based (GPP, NEE, and Reco) and pool-based (Leaf, Wood, Root, and Soil) variables caused by correction in individual forcing (photosynthetically active radiation (PAR, red), rainfall (Rain, brown), relative humidity (RH, green), air temperature (Tair, light blue), soil temperature (Tsoil, dark blue), wind speed (WS, purple)) while other forcings vary in ensemble predictions.

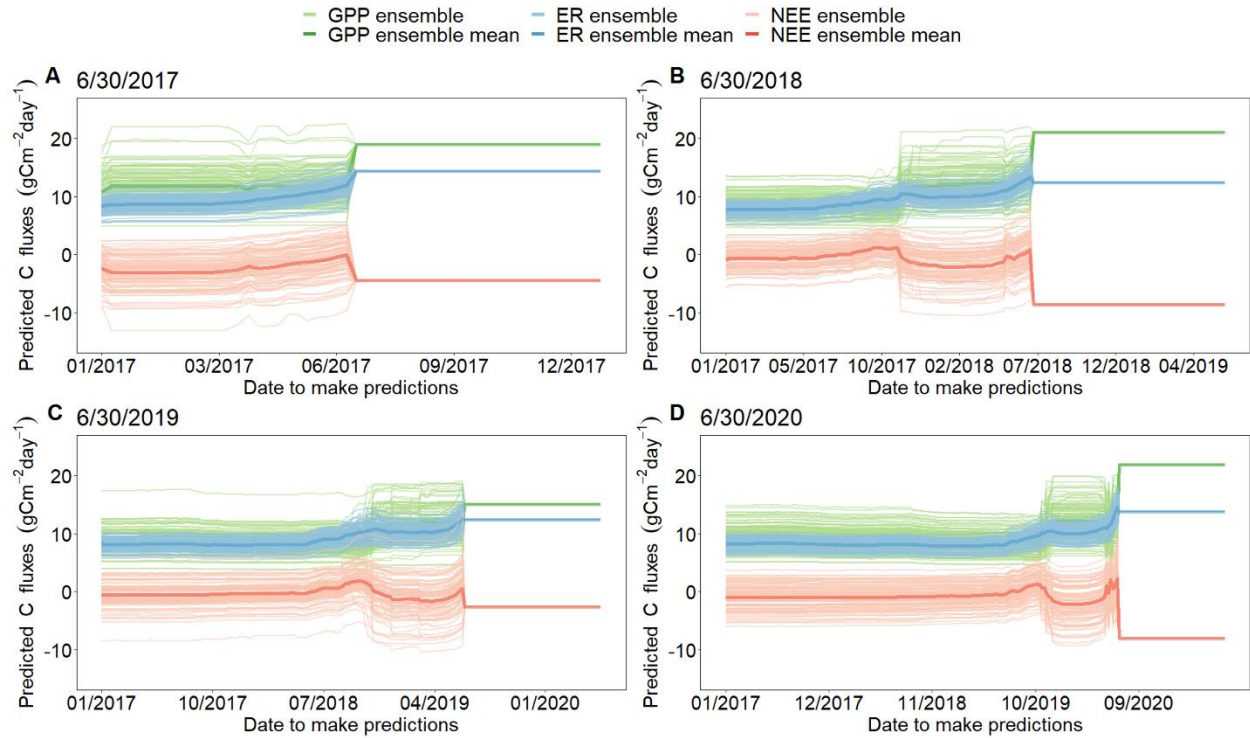


Figure 4: Convergence of three predicted carbon fluxes (GPP: green; ecological respiration (ER): blue, NEE: red) for a certain date (A: 6/30/2017; B: 6/30/2018; C: 6/30/2019; D: 6/30/2020) across time (A: 1/2017-12/2017; B: 1/2017-4/2019; C: 2017-2020; D: 2017-2021). Each line depicts predictions for a given date with weekly corrected forcing. Line segment before the expected date is forecast with future forcing, and line segment after that data is hindcast with real forcing. Ensemble prediction with future forcings is represented by lines of the same color (light green, light blue, or light red), and ensemble mean is shown by lines of a darker hue (dark green, dark blue, or dark red). With weekly forcing updates, flux-based variables do not change much until very close to the expected date.

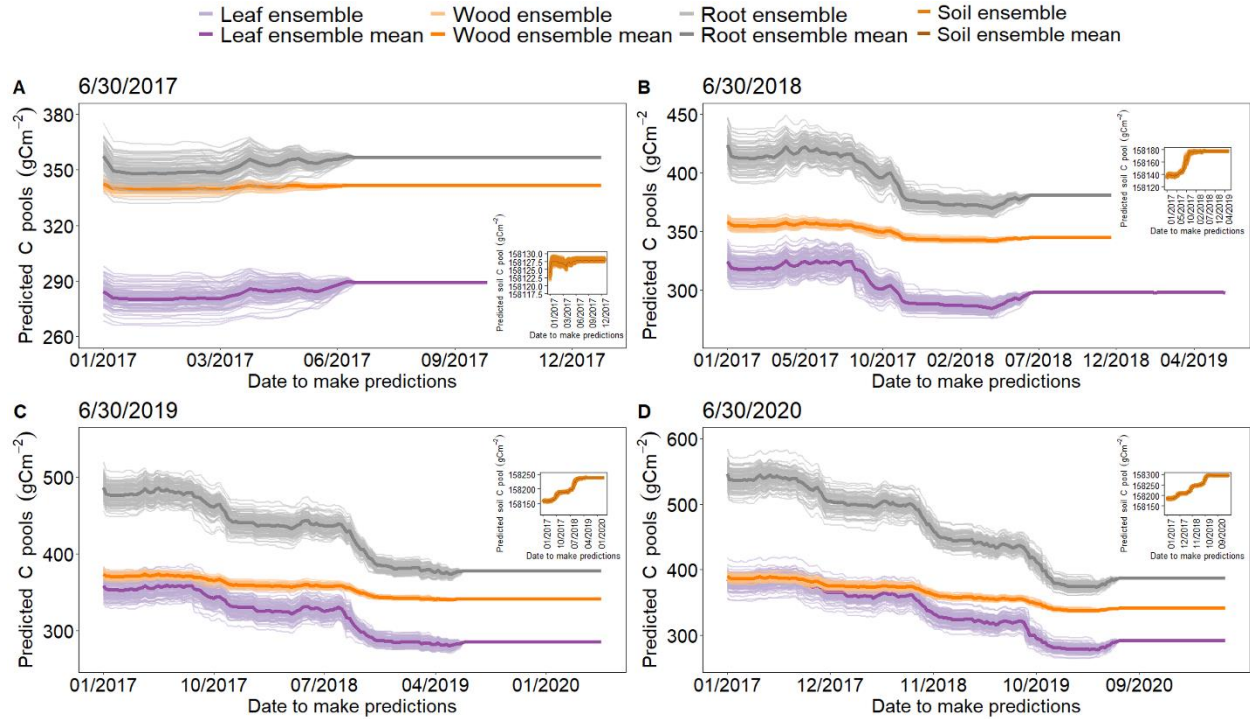


Figure 5: Convergence of three predicted carbon pools (leaf: purple; wood: orange; root: grey; soil: brown) for a specific date (A: 6/30/2017; B: 6/30/2018; C: 6/30/2019; D: 6/30/2020) over time (A: 1/2017-12/2017; B: 1/2017-4/2019; C: 2017-2020; D: 2017-2021). Each line represents predictions for the specific date with weekly corrected forcing. Line segment before the expected date is forecast with future forcing, and line segment after that data is hindcast with real forcing. Ensemble prediction with future forcings is represented by lines of the same color (light purple, light orange, light grey, or light brown), and ensemble mean is shown by lines of a darker hue (dark purple, dark orange, dark grey, or dark brown). Pool-based variables fluctuate with weekly forcing updates and converge before the expected date.

CHAPTER 5: CONCLUSIONS AND PERSPECTIVES

With the advent of extensive observation networks and new measurement systems, ecology is entering into a big data era (Kelling et al., 2009; Baldocchi, 2020; Song et al., 2019). The ultimate value of the enormous data will depend on how well the big data can be utilized to advance our comprehension of ecosystem responses to climate change and our ability to predict future ecological functions and states (Luo et al., 2011). In this dissertation, I developed novel informatics tools and methods to provide new solutions and endeavors for data-intensive ecological modeling and forecasting. These tools and methods have been developed particularly to tackle three uncertainty sources that hinder accurate ecological predictions. The first uncertainty is related to improper parameter values in a model. The second uncertainty comes from inadequate model structures to represent the underlying ecological processes of individual ecosystems. The final uncertainty arises from unclear future climate forcings to act as model input. I will summarize the conclusions of my dissertation work and potential future research directions based on my findings in the following three sections separately.

5.1 Model-independent data assimilation module (MIDA)

In chapter 2, I developed MIDA to facilitate the application of data assimilation (DA) in land C cycle modeling. Compared to traditional DA which requires invasive coding, MIDA only requires users to prepare data (e.g., parameter prior range) and model, and provide directories of the parameter file, model executable file, output configuration file, and DA results by a file named `namelist.txt` or by interactive inputs via a window. This greatly alleviates users from model-dependent coding in conducting DA research. We demonstrated the applicability of MIDA to three ecological models (i.e., the DALEC model, the surrogate-based ELM model and the BiomeE model). To switch the test case from one model to another model, MIDA only

requires users to change information in the `namelist.txt` file. This flexibility in switching models enables the capability of MIDA to support model comparison. We tested this flexibility with nine phenological models. All these test cases demonstrated the capability of MIDA in constraining parameter uncertainty, enhancing ecological forecasting, and facilitating model intercomparison.

MIDA provides a way to standardize the data exchange between model and DA algorithms. More essential components underlying DA, such as data preparation, can be expedited with emerging techniques and tools. A future research direction of Chapter 2 is to develop more open-access, easy-to-use software for the integration of big data and models. Taking data preparation as an example, in Chapter 2, we collected various observations and models for different test cases and compared the results of MIDA with previous DA studies. Collecting data manually will be challenging if we are faced with a sheer volume and diversity of datasets (Waide et al., 2017). To facilitate locating and obtaining data in a big data era, we suggest the utility of data repositories such as DataONE (Data Observation Network for Earth | DataONE, 2023). Meanwhile, sharing DA results via such data repositories can reduce redundant efforts and boost the efficiency of DA studies (Fer et al., 2021). For example, the parameter posterior distribution estimated in one DA study can be used as a prior distribution in another DA study, thus avoiding the need to repeat the same DA experiment. Finally, such data repositories, once documented well, can synthesize what has been found in thousands of publications. For example, there are many microbial models to simulate potential microbial and mineral involvements in soil biogeochemical processes (Wieder et al., 2015). The data synthesis then can provide modelers necessary information to determine which process and associated parameters should be considered in DA to answer a specific research question (Luo et al., 2016).

5.2 New formula in water-dependent cycles

In chapter 3, I updated the structure of the Terrestrial ECOSystem model (TECO) to incorporate ground hydrology changes (i.e., subsidence), snow accumulation, and associated impacts on carbon (C) and nitrogen (N) cycles after permafrost thaw. After constraining this updated TECO model with observations from a permafrost warming study site (CiPEHR) (Natali et al., 2011), I found that TECO can simulate elevated water table depth (WTD, or soil wetting) under soil warming and air & soil warming treatments. I then used this updated TECO model to extend simulations and include more thawing scenarios (i.e., deep thaw with dry soil and shallow thaw with wet soil), which were not observed in the CiPEHR study site. With the extended simulations, I evaluated the impacts of TD and WTD on net ecosystem exchange (NEE), gross primary production (GPP), and ecology respiration (ER). The most important finding is that elevated WTD shifts the permafrost ecosystem from a C sink to source. Without incorporating the subsidence processes, models may overestimate the C sequestration capability in permafrost regions.

A widespread soil wetting at CiPEHR (Mauritz et al., 2017) will increase CH₄ emissions, which is another important C emission source (Natali et al., 2015). One of future work is to take CH₄ into account when assessing the responses of C emissions to elevated WTD. The TECO model has been applied in a peatland warming study (Ma et al., 2017). We may need to integrate CH₄ chamber measurements into model to conduct another data assimilation at the permafrost ecosystem. Another future work is to explicitly represent the impacts of subsidence and permafrost thaw. In this study, I simplify these impacts with different parameters whose values are estimated by data assimilation. The relevant hydrological processes underlying subsidence include thawing of excess snow aboveground (Lee et al., 2014) and water redistributions among different soil layers (vertical and horizontal movements) (Chang et al., 2019). Specifically, many

studies imply that differences in landscapes will lead to a variety of hydrological changes. For example, thaw-induced hydrological changes in lowlands can lead to subsidence or thermokarst lake while permafrost thaw in uplands can increase water drainage and result in water table descending (Osterkamp et al., 2009; Hinzman et al., 2005). The study of Shi et al. (2015) developed new hydrological dynamics for the microtopography in peatlands, which can be a start for my future work.

5.3 A new assessment of forcing correction in terms of forecast accuracy

In chapter 4, I developed a new assessment method to evaluate forecast accuracy based on hindcast. Forecast accuracy is defined as the discrepancy between the forecasting result and a ‘I’ model result such as an autoregressive process ($y_{t+1} = y_t + \varepsilon$) (Dietze, 2017). Forecast accuracy is used to evaluate whether a model can provide a more informative prediction than a random guess. Here, I expand this concept of forecast accuracy and define it as the discrepancy between the forecasting result and the true values (i.e., hindcast values). To quantify this discrepancy, I use the Kullback-leibler (KL) divergence, which synthesizes more information about the discrepancy than RMSE or variance.

The test case is to predict the future C cycles (2015-2024) at a peatland warming experiment site (SPRUCE) with the TECO model. I first demonstrated that the current assessment method of forcing uncertainty is not able to reflect the forecast accuracy, especially in predicting pool-based variables (e.g., leaf pool). I then used the true forcing (2015-2022) to conduct forcing correction and quantified the effect of forcing correction on forecast accuracy using KL divergence. I found that forcing correction contributed more than parameter calibration for almost all C variables. To further diagnose why the evaluation of forecast accuracy is different from that of forcing uncertainty, I further conducted a weekly forcing correction and

analyzed how the prediction of a future C variable (e.g., NEE or leaf pool) changed from this forcing correction. I found pool-based variables benefit more from the weekly forcing correction because of its inherent memory feature.

Many forecasting applications use future forcing generated by Global Circulation Models (GCMs) without explicitly considering the impacts of forcing uncertainty on predictions (Teckentrup et al., 2022). This not only includes (1) how to assess the impact of forcing uncertainty (e.g., forecast accuracy) but also (2) how to reduce this uncertainty impact (e.g., weekly forcing correction). My work in chapter 4 is just the very beginning and there is a lot of potential work to move this topic forward.

For (1), we may need a variety of models and forcings to validate the feasibility of defining forecast accuracy based on hindcast. An interesting question will be how to define quantitative metrics to quantify the forecast accuracy for most ecosystem models regardless of the variety of output variables. For example, the simulation output of phenological models (e.g., yearly leaf-out date) is sparse in number compared with hourly C fluxes. Thus, a KL divergence is not a good choice to quantify the forecast accuracy. As for (2), cyberinfrastructure is needed to implement a periodic assessment of forecast accuracy in an iterative prediction platform such as EcoPAD (Huang et al., 2019) and PEcAn (LeBauer et al., 2013). If we need to evaluate the forecast accuracy of a specific future date with periodic forcing correction (e.g., weekly), it will be both data-intensive (e.g., weekly updated ensemble forcing with a high resolution in temporal and spatial scales) and computation-intensive (e.g., weekly estimates of the improvements in forecast accuracy towards that date). Accordingly, cloud storage and computing clusters may be needed.

References

- Data Observation Network for Earth | DataONE: <https://www.dataone.org/>, last access: 5 March 2023.
- Baldocchi, D. D.: How eddy covariance flux measurements have contributed to our understanding of Global Change Biology, *Global Change Biology*, 26, 242–260, <https://doi.org/10.1111/gcb.14807>, 2020.
- Chang, K.-Y., Riley, W. J., Crill, P. M., Grant, R. F., Rich, V. I., and Saleska, S. R.: Large carbon cycle sensitivities to climate across a permafrost thaw gradient in subarctic Sweden, *The Cryosphere*, 13, 647–663, <https://doi.org/10.5194/tc-13-647-2019>, 2019.
- Dietze, M.: *Ecological Forecasting*, Princeton University Press, <https://doi.org/10.1515/9781400885459>, 2017.
- Fer, I., Gardella, A. K., Shiklomanov, A. N., Campbell, E. E., Cowdery, E. M., De Kauwe, M. G., Desai, A., Duveneck, M. J., Fisher, J. B., Haynes, K. D., Hoffman, F. M., Johnston, M. R., Kooper, R., LeBauer, D. S., Mantooth, J., Parton, W. J., Poulter, B., Quaife, T., Raiho, A., Schaefer, K., Serbin, S. P., Simkins, J., Wilcox, K. R., Viskari, T., and Dietze, M. C.: Beyond ecosystem modeling: A roadmap to community cyberinfrastructure for ecological data-model integration, *Global Change Biology*, 27, 13–26, <https://doi.org/10.1111/gcb.15409>, 2021.
- Hinzman, L. D., Bettez, N. D., Bolton, W. R., Chapin, F. S., Dyurgerov, M. B., Fastie, C. L., Griffith, B., Hollister, R. D., Hope, A., Huntington, H. P., Jensen, A. M., Jia, G. J., Jorgenson, T., Kane, D. L., Klein, D. R., Kofinas, G., Lynch, A. H., Lloyd, A. H., McGuire, A. D., Nelson, F. E., Oechel, W. C., Osterkamp, T. E., Racine, C. H., Romanovsky, V. E., Stone, R. S., Stow, D. A., Sturm, M., Tweedie, C. E., Vourlitis, G. L., Walker, M. D., Walker, D. A., Webber, P. J., Welker, J. M., Winker, K. S., and Yoshikawa, K.: Evidence and Implications of Recent Climate Change in Northern Alaska and Other Arctic Regions, *Climatic Change*, 72, 251–298, <https://doi.org/10.1007/s10584-005-5352-2>, 2005.
- Huang, Y., Stacy, M., Jiang, J., Sundi, N., Ma, S., Saruta, V., Jung, C. G., Shi, Z., Xia, J., Hanson, P. J., Ricciuto, D., and Luo, Y.: Realized ecological forecast through an interactive Ecological Platform for Assimilating Data (EcoPAD, v1.0) into models, *Geosci. Model Dev.*, 12, 1119–1137, <https://doi.org/10.5194/gmd-12-1119-2019>, 2019.
- Kelling, S., Hochachka, W. M., Fink, D., Riedewald, M., Caruana, R., Ballard, G., and Hooker, G.: Data-intensive Science: A New Paradigm for Biodiversity Studies, *BioScience*, 59, 613–620, <https://doi.org/10.1525/bio.2009.59.7.12>, 2009.
- LeBauer, D. S., Wang, D., Richter, K. T., Davidson, C. C., and Dietze, M. C.: Facilitating feedbacks between field measurements and ecosystem models, *Ecological Monographs*, 83, 133–154, <https://doi.org/10.1890/12-0137.1>, 2013.
- Lee, H., Swenson, S. C., Slater, A. G., and Lawrence, D. M.: Effects of excess ground ice on projections of permafrost in a warming climate, *Environ. Res. Lett.*, 9, 124006, <https://doi.org/10.1088/1748-9326/9/12/124006>, 2014.
- Luo, Y., Ogle, K., Tucker, C., Fei, S., Gao, C., LaDeau, S., Clark, J. S., and Schimel, D. S.: Ecological forecasting and data assimilation in a data-rich era, *Ecological Applications*, 21, 1429–1442, <https://doi.org/10.1890/09-1275.1>, 2011.
- Luo, Y., Ahlström, A., Allison, S. D., Batjes, N. H., Brovkin, V., Carvalhais, N., Chappell, A., Ciais, P., Davidson, E. A., Finzi, A., Georgiou, K., Guenet, B., Hararuk, O., Harden, J.

- W., He, Y., Hopkins, F., Jiang, L., Koven, C., Jackson, R. B., Jones, C. D., Lara, M. J., Liang, J., McGuire, A. D., Parton, W., Peng, C., Randerson, J. T., Salazar, A., Sierra, C. A., Smith, M. J., Tian, H., Todd-Brown, K. E. O., Torn, M., van Groenigen, K. J., Wang, Y. P., West, T. O., Wei, Y., Wieder, W. R., Xia, J., Xu, X., Xu, X., and Zhou, T.: Toward more realistic projections of soil carbon dynamics by Earth system models, *Global Biogeochemical Cycles*, 30, 40–56, <https://doi.org/10.1002/2015GB005239>, 2016.
- Ma, S., Jiang, J., Huang, Y., Shi, Z., Wilson, R. M., Ricciuto, D., Sebestyen, S. D., Hanson, P. J., and Luo, Y.: Data-Constrained Projections of Methane Fluxes in a Northern Minnesota Peatland in Response to Elevated CO₂ and Warming, *Journal of Geophysical Research: Biogeosciences*, 122, 2841–2861, <https://doi.org/10.1002/2017JG003932>, 2017.
- Mauritz, M., Bracho, R., Celis, G., Hutchings, J., Natali, S. M., Pegoraro, E., Salmon, V. G., Schädel, C., Webb, E. E., and Schuur, E. A. G.: Nonlinear CO₂ flux response to 7 years of experimentally induced permafrost thaw, *Glob Change Biol*, 23, 3646–3666, <https://doi.org/10.1111/gcb.13661>, 2017.
- Natali, S. M., Schuur, E. a. G., Trucco, C., Hicks Pries, C. E., Crummer, K. G., and Baron Lopez, A. F.: Effects of experimental warming of air, soil and permafrost on carbon balance in Alaskan tundra, *Global Change Biology*, 17, 1394–1407, <https://doi.org/10.1111/j.1365-2486.2010.02303.x>, 2011.
- Natali, S. M., Schuur, E. A. G., Mauritz, M., Schade, J. D., Celis, G., Crummer, K. G., Johnston, C., Krapek, J., Pegoraro, E., Salmon, V. G., and Webb, E. E.: Permafrost thaw and soil moisture driving CO₂ and CH₄ release from upland tundra, *Journal of Geophysical Research: Biogeosciences*, 120, 525–537, <https://doi.org/10.1002/2014JG002872>, 2015.
- Osterkamp, T. E., Jorgenson, M. T., Schuur, E. A. G., Shur, Y. L., Kanevskiy, M. Z., Vogel, J. G., and Tumskey, V. E.: Physical and ecological changes associated with warming permafrost and thermokarst in Interior Alaska: Physical and Ecological Changes Associated with Thermokarst, *Permafrost Periglac. Process.*, 20, 235–256, <https://doi.org/10.1002/ppp.656>, 2009.
- Shi, X., Thornton, P. E., Ricciuto, D. M., Hanson, P. J., Mao, J., Sebestyen, S. D., Griffiths, N. A., and Bisht, G.: Representing northern peatland microtopography and hydrology within the Community Land Model, *Biogeosciences*, 12, 6463–6477, <https://doi.org/10.5194/bg-12-6463-2015>, 2015.
- Song, J., Wan, S., Piao, S., Knapp, A. K., Classen, A. T., Vicca, S., Ciais, P., Hovenden, M. J., Leuzinger, S., Beier, C., Kardol, P., Xia, J., Liu, Q., Ru, J., Zhou, Z., Luo, Y., Guo, D., Adam Langley, J., Zscheischler, J., Dukes, J. S., Tang, J., Chen, J., Hofmockel, K. S., Kueppers, L. M., Rustad, L., Liu, L., Smith, M. D., Templer, P. H., Quinn Thomas, R., Norby, R. J., Phillips, R. P., Niu, S., Fatichi, S., Wang, Y., Shao, P., Han, H., Wang, D., Lei, L., Wang, J., Li, X., Zhang, Q., Li, X., Su, F., Liu, B., Yang, F., Ma, G., Li, G., Liu, Y., Liu, Y., Yang, Z., Zhang, K., Miao, Y., Hu, M., Yan, C., Zhang, A., Zhong, M., Hui, Y., Li, Y., and Zheng, M.: A meta-analysis of 1,119 manipulative experiments on terrestrial carbon-cycling responses to global change, *Nat Ecol Evol*, 3, 1309–1320, <https://doi.org/10.1038/s41559-019-0958-3>, 2019.
- Teckentrup, L., De Kauwe, M. G., Abramowitz, G., Pitman, A. J., Ukkola, A. M., Hobeichi, S., François, B., and Smith, B.: Opening Pandora’s box: How to constrain regional projections of the carbon cycle, *EGUsphere*, 1–51, <https://doi.org/10.5194/egusphere-2022-623>, 2022.

- Waide, R. B., Brunt, J. W., and Servilla, M. S.: Demystifying the Landscape of Ecological Data Repositories in the United States, *BioScience*, 67, 1044–1051, <https://doi.org/10.1093/biosci/bix117>, 2017.
- Wieder, W. R., Allison, S. D., Davidson, E. A., Georgiou, K., Hararuk, O., He, Y., Hopkins, F., Luo, Y., Smith, M. J., Sulman, B., Todd-Brown, K., Wang, Y.-P., Xia, J., and Xu, X.: Explicitly representing soil microbial processes in Earth system models, *Global Biogeochemical Cycles*, 29, 1782–1800, <https://doi.org/10.1002/2015GB005188>, 2015.