# A Spatiotemporal Sequence Forecasting Platform to Advance the Prediction of Changing Spatiotemporal Patterns of $CO_2$ Concentration by Incorporating Human Activity and Hydrological Extremes

Melissa Allen-Dumas[1], Kuldeep Kurte[1], Haowen Xu[1], Jibonananda Sanyal[1], and Guannan Zhang[1]

[1]*Oak Ridge National Laboratory, Oak Ridge, TN, USA*

## Focal Areas

**(2)** Predictive modeling using AI techniques and AI-derived model components. It describes the use of AI and other tools to design a prediction system comprising a hierarchy of models. Specifically, this white paper focuses on the development of a collaborative deep learning platform for evaluating and predicting spatiotemporal relationships between the hydrological and carbon cycles and includes capabilities for considering biogenic and anthropogenic inputs to these systems.

## Science Challenge

How much of the shift in spatiotemporal $CO_2$ patterns is due to large-scale and gradual environmental change, and how much is due to increases in proximal and concomitant human activity (i.e., $CO_2$ emissions and water consumption)? Additionally, how do extreme hydrological events affect the near- and long-term carbon cycle responses of these natural systems?

## Rationale

Recent atmospheric measurements show that higher concentrations of $CO_2$ are in areas previously thought to be net sinks of $CO_2$ and that lower concentrations appear in areas that were previously thought to absorb less $CO_2$. This observation suggests that the spatial pattern of biogenic $CO_2$ sources and sinks are changing as the climate changes (NASA 2019), and that temporal predictions of the spatial distribution of $CO_2$ concentration cannot be deterministically projected due to the nonstationarity of the environmental forces. If these spatial patterns of $CO_2$ can be forecasted by incorporating this nonstationarity, predictions for future $CO_2$ concentrations and climate conditions can be improved. However, surface-atmosphere carbon flux has two opposing components, each relatively large compared with the net sink. These include: (1) a release of carbon to the atmosphere related to land use change and (2) inter-annual variability in the uptake of $CO_2$ in tropical, subtropical, and boreal terrestrial ecosystems (Joiner et al. 1999, Stephens et al. 2007, Schurgers et al. 2018) that pertain primarily to a lack of equilibrium between photosynthesis and plant and soil respiration (Ballantyne et al. 2015, Schurgers et al. 2018).

## Narrative

We propose a simulation platform based on Spatiotemporal Sequence Forecasting (STSF) that will allow researchers to run experiments to learn how spatial patterns of $CO_2$ concentration evolve under varying meteorological and hydrological and conditions with and without local anthropogenic forcing. We propose this platform with an eye toward developing an integrated methodology for incorporating locally resolved human and environmental processes into regional and global climate models in cooperation with community needs and relevant science drivers. However, the deliverable here will be a stand-alone file generator for spatiotemporal-gridded $CO_2$ concentration.



DOI: https://doi.org/10.2172/1769653
DOE BER, AI4ESP 2021 #AI4ESP1001 (https://www.ai4esp.org)

For the STSF analysis, we will implement a deep learning method called *recurrent neural networks* (RNN) within an Encoder-Forecaster structure. This scheme first encodes various spatiotemporal combinations of variables into a finite-dimensional vector or a probability distribution, then passes the vector to a RNN, which dynamically determines, in a hidden layer, recurrent connections among locations and timestamps of the input data. Finally, the method generates multistep-ahead predictions by using the forecaster, which includes an additional RNN step and final deconvolution (Shi, 2018). Some input variables to the analysis could include $CO_2$ emissions (e.g., US Environmental Protection Agency sources [Gurney et al. 2020]), landcover, temperature, direct and diffuse radiation, latent heat flux, wind, humidity, precipitation, consumptive water use (e.g., McManamay et al. [2021]), and soil moisture data observations. To handle uncertainty in the method, the encoder and forecaster will contain a probabilistic component informed by convolutional long short-term memory so that long-term dependencies of spatial $CO_2$ concentration on the temporal patterns of the input variables are captured as well as spatiotemporal correlations in the data. In addition, to avoid over-confident predictions made by an RNN model, we will simultaneously train an ensemble of RNN models, each of which is trained with a different initial condition. Due to the non-convexity of the RNN's loss landscape, each RNN will explore a different sub-region in the parameter space. Then, we use the average of the RNNs' outputs to make predictions and use the variance of RNNs' outputs to quantify the confidence/uncertainty of the predictions. In this way, an over-confident prediction made by one RNN can be corrected by the rest of RNNs in the ensemble.

To test the reliability of this method for producing sub-kilometer spatiotemporally explicit $CO_2$ concentrations, we will encode local and regional time series parameters for specific locations in three climate zones—tropical, midlatitude, and boreal—by using data available from the Atmospheric Radiation Measurement (ARM) data center. These tests will be conducted within the proposed platform, which will access the ARM data through a built-in application programming interface. If ARM data are unavailable for a needed parameter, other data feeds will be established. Additionally, provision will be made for the ingestion of user-selected datasets. These data access functions will be incorporated into the platform similarly to those incorporated in the Eagle-I[1] and North American Energy Resilience Model[2] platforms.

The results of the simulations at each resolution will be compared with measured spatiotemporal $CO_2$ patterns. Additionally, we will determine the relationship between the values obtained with highly resolved input and values obtained with more regional and long-term input by using traditional differencing and statistical methods and/or deep learning to relate values in high-resolution grid cells to those in low-resolution grid cells (e.g., Rodrigues et al. [2018]).

To evaluate the validity and estimate the uncertainty in the terrestrial ecosystem sink identification, we will track the times and locations at which $CO_2$ is significantly higher or lower than in the surrounding areas. Then, we will use receiver operating characteristic curves (ROC) to diagnose the predictive fitness of this parameter. This is done by plotting the true positive rate of prediction as a function of the false positive rate compared with solar-induced fluorescence measurements (i.e., indicators of the rate at which plants convert sunlight and $CO_2$ into chemical energy) and

---

[1]https://www.energy.gov/oe/articles/doe-announces-transition-eagle-i-oak-ridge-national-laboratory-ornl-taking-advantage
[2]https://www.energy.gov/sites/prod/files/2019/07/f65/NAERM_Report_public_version_072219_508.pdf

identifying the level at which the function is maximized or the highest true positive to total diagnoses (i.e., sensitivity) is met (Murphy 1996, Peres and Cancelliere 2014, Rottman 2019). A gridded graphical representation of the ROC curves will be displayed in the user interface for interactive evaluation.

***Predictive Use***

Once the system has been tested for predictive capability and its uncertainty is characterized using historical data, capabilities for predicting the impacts of future human activity on future climate will be integrated. For example, projected emissions datasets based on Shared Socioeconomic Pathways (O'Neill et al. 2014, 2016, 2017) population and energy use, disaggregated to 1 km and hourly resolution (based on previous work, such as McManamay et al. [2019]) can be provided. Accommodation can also be made for user-defined scenarios.

***Future Extreme Hydrological Event Impact***

The capability for superimposing a drought scenario on a specific location under future climate will be provided (e.g., Allen et al. [2017]) by using the Standardized Precipitation Index (SPI) and the relationship between the lack of precipitation and water scarcity during relevant historical periods for locations of interest as a basis for incremental adjustments in water availability estimates for future dates. Furthermore, the cumulative effects of evapotranspiration and projected consumptive water use can be included to modify surface water levels. Machine learning techniques for predicting SPI (e.g., Belayneh and Adamowski [2012]) can be tested for each location to determine which method might work best for such predictions. The best method will be provided as a component of the overall platform. For the test cases, gridded SPI datasets will be generated and compared with calculated SPI from measured data. Visualizations of predicted and calculated SPI along with root mean square error and $R^2$-values will be shown in the user interface. The resulting gridded SPI calculations can then be used as inputs at the encoder step for the STSF simulations.

***FAIR Data Use***

To allow results from this platform to be reproducible, in accordance with FAIR (Findable, Accessible, Interoperable, Reusable) principles, the platform will track and save data provenance and experimental output. This output, along with model codes, will be made available to the user during platform access and to the scientific community upon request.

## References


Allen, M. R., Wilbanks, T. J., Preston, B. L., Kao, S. C., and Bradbury, J. 2017. *Assessing the Costs and Benefits of Resilience Investments: Tennessee Valley Authority Case Study* (No. ORNL/TM-2017/13). Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Ballantyne, A. P., Andres, R., Houghton, R., Stocker, B. D., Wanninkhof, R., Anderegg, W., Cooper, L. A., DeGrandpre, M., Tans, P. P., Miller, J. B., and Alden, C. 2015. Audit of the global carbon budget: estimate errors and their impact on uptake uncertainty. *Biogeosciences (Online)*, *12*(8).

Belayneh, A., and Adamowski, J. 2012. Standard precipitation index drought forecasting using neural networks, wavelet neural networks, and support vector regression. *Applied Computational Intelligence and Soft Computing*.

Gurney, K. R., Liang, J., Patarasuk, R., Song, Y., Huang, J., and Roest, G. 2020. The Vulcan version 3.0 high-resolution fossil fuel CO2 emissions for the United States. *Journal of Geophysical Research: Atmospheres*, *125*(19), e2020JD032974.

Joiner, D. W., Lafleur, P. M., McCaughey, J. H., and Bartlett, P. A. 1999. Interannual variability in carbon dioxide exchanges at a boreal wetland in the BOREAS northern study area. *Journal of Geophysical Research: Atmospheres*, *104*(D22), 27,663–27,672.

McManamay, R. A., DeRolph, C. R., Surendran-Nair, S., and Allen-Dumas, M. 2019. Spatially explicit land-energy-water future scenarios for cities: Guiding infrastructure transitions for urban sustainability. *Renewable and Sustainable Energy Reviews*, *112*, 880–900.

McManamay, R. A., Binita, K. C., Allen-Dumas, M. R., Brelsford, C. M., Sanyal, J., Stewart, R. N., and Bhaduri, B. L. 2021. Reanalysis of water use for irrigation, electric power, and public supply sectors in the conterminous United States, 1950–2016. *Water Resources Research*, In Press.

Murphy, A. H. 1996. The Finley Affair: A Signal Event in the History of Forecast Verification. Weather and Forecasting, 11(1), 3–20.

NASA. 2019. The Atmosphere: Getting a Handle on Carbon Dioxide, Part Two, by Alan Buis. https://climate.nasa.gov/news/2915/the-atmosphere-getting-a-handle-on-carbon-dioxide/

O’Neill, B. C., Kriegler, E., Riahi, K., Ebi, K. L., Hallegatte, S., Carter, T. R., Mathur, R., and van Vuuren, D. P. 2014. A new scenario framework for climate change research: the concept of shared socioeconomic pathways. *Climatic change*, *122*(3), 387–400.

O'Neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J. F., Lowe, J., and Meehl, G .A. 2016. The scenario model intercomparison project (ScenarioMIP) for CMIP6.

O'Neill, B. C., Kriegler, E., Ebi, K. L., Kemp-Benedict, E., Riahi, K., Rothman, D. S., van Ruijven, B. J., van Vuuren, D. P., Birkmann, J., Kok, K., and Levy, M. 2017. The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century. *Global Environmental Change*, 42, 169–180.

Peres, D. J., and Cancelliere, A. 2014. Derivation and evaluation of landslide-triggering thresholds by a Monte Carlo approach. *Hydrology and Earth System Sciences*, *18*(12), 4,913.

Rodrigues, E. R., Oliveira, I., Cunha, R., and Netto, M. 2018. DeepDownscale: A Deep Learning Strategy for High-Resolution Weather Forecast. In *2018 IEEE 14th International Conference on e-Science (e-Science)*, 415–422. IEEE.

Rottmann, M., Maag, K., Chan, R., Hüger, F., Schlicht, P., and Gottschalk, H. 2019. Detection of False Positive and False Negative Samples in Semantic Segmentation. *arXiv preprint arXiv:1912.03673*.

Schurgers, G., Ahlström, A., Arneth, A., Pugh, T. A., and Smith, B. 2018. Climate sensitivity controls uncertainty in future terrestrial carbon sink. *Geophysical Research Letters*, *45*(9), 4,329–4,336.

Shi, X. 2018. Exploring deep learning architectures for spatiotemporal sequence forecasting (Doctoral dissertation).

Stephens, B. B., Gurney, K. R., Tans, P. P., Sweeney, C., Peters, W., Bruhwiler, L., Ciais, P., Ramonet, M., Bousquet, P., Nakazawa, T., and Aoki, S. 2007. Weak northern and strong tropical land carbon uptake from vertical profiles of atmospheric CO2. *Science*, *316*(5832), 1,732–1,735.