

Water Resources Research

RESEARCH ARTICLE

10.1029/2022WR034022

Key Points:

- Soil water is used as a proxy for oxygen, which may lead to incorrect oxygen predictions (e.g., oxygen remains high under high moisture)
- A machine learning technique, the Self-Organizing Map, was used to assess the relationship between soil moisture and oxygen in riparian soil
- Soil and hydrologic conditions complicated the relationship between soil water content and oxygen, which led to anomalous oxygen conditions

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

B. V. Lancellotti,
Brittany.lancellotti@uvi.edu

Citation:

Lancellotti, B. V., Underwood, K. L., Perdrial, J. N., Schroth, A. W., Roy, E. D., & Adair, C. E. (2023). Complex drivers of riparian soil oxygen variability revealed using self-organizing maps. *Water Resources Research*, 59, e2022WR034022. <https://doi.org/10.1029/2022WR034022>

Received 2 NOV 2022
Accepted 30 MAY 2023

Author Contributions:

Conceptualization: Brittany V. Lancellotti, Julia N. Perdrial, Andrew W. Schroth, Eric D. Roy, Carol E. Adair
Data curation: Brittany V. Lancellotti
Methodology: Kristen L. Underwood
Software: Brittany V. Lancellotti, Kristen L. Underwood
Supervision: Kristen L. Underwood, Julia N. Perdrial, Carol E. Adair
Visualization: Brittany V. Lancellotti
Writing – original draft: Brittany V. Lancellotti

© 2023 The Authors.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial License](#), which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Complex Drivers of Riparian Soil Oxygen Variability Revealed Using Self-Organizing Maps

Brittany V. Lancellotti^{1,2} , Kristen L. Underwood³ , Julia N. Perdrial⁴ , Andrew W. Schroth⁴ , Eric D. Roy^{1,3}, and Carol E. Adair¹

¹Rubenstein School of Environment and Natural Resources, University of Vermont, Burlington, VT, USA, ²Now at Virgin Islands EPSCoR, University of the Virgin Islands, Kingshill, VI, USA, ³Civil & Environmental Engineering, University of Vermont, Burlington, VT, USA, ⁴Department of Geology, University of Vermont, Burlington, VT, USA

Abstract Oxygen (O₂) regulates soil reduction-oxidation processes and therefore modulates biogeochemical cycles. The difficulties associated with accurately characterizing soil O₂ variability have prompted the use of soil moisture as a proxy for O₂, as O₂ diffusion into soil water is much slower than in soil air. The use of soil moisture alone as a proxy measurement for O₂ could result in inaccurate O₂ estimations. For example, O₂ may remain high during cool months when soil respiration rates are low. We analyzed high-frequency sensor data (e.g., soil moisture, CO₂, gas-phase soil pore O₂) with a machine learning technique, the Self-Organizing Map, to pinpoint suites of soil conditions associated with contrasting O₂ regimes. At two riparian sites in northern Vermont, we found that O₂ levels varied seasonally, and with soil moisture. For example, 47% of low O₂ levels were associated with wet and cool soil conditions, whereas 32% were associated with dry and warm conditions. Contrastingly, the majority (62%) of high O₂ conditions occurred under dry and warm conditions. High soil moisture levels did not always lead to low O₂, as 38% of high O₂ values occurred under wet and cool conditions. Our results highlight challenges with predicting soil O₂ solely based on water content, as variable combinations of soil and hydrologic conditions can complicate the relationship between water content and O₂. This indicates that process-based ecosystem and denitrification models that rely solely on soil moisture to estimate O₂ may need to incorporate other site and climate-specific drivers to accurately predict soil O₂.

Plain Language Summary Oxygen in soil is an important variable to monitor, as it affects how nutrients, like nitrogen and carbon, cycle through the soil. However, oxygen is difficult to measure in soil, since it's spatially and temporally variable. Based on the general rule that soil oxygen decreases as soil moisture increases, moisture is sometimes used to approximate soil oxygen levels. In nature, however, the relationship between these variables is not simple, and sometimes oxygen behaves in a way that is the exact opposite of this general rule. Using specialized soil sensors that record data at 15-minute intervals, we monitored gas-phase soil pore oxygen, soil moisture, and other soil conditions to better understand why these unexpected soil oxygen levels occur. We found that most high oxygen levels occurred during warm months, when soils were driest, and that most low oxygen levels occurred during cool months, when soils were wettest. However, we found that 38% of high oxygen levels occurred under conditions that, according to the general rule, should have resulted in low oxygen. We think these unexpected oxygen levels are caused by complex interactions between soil processes that produce and consume oxygen in soil, and we therefore caution against using the general rule to predict oxygen.

1. Introduction

Oxygen (O₂) is the quintessential electron acceptor and therefore drives biogeochemical cycling on Earth. Its availability within soil pores strongly modulates soil oxidation-reduction (redox) potential, thereby controlling which energy yielding soil biogeochemical reactions proceed (M. Hefting et al., 2004; Silver et al., 1999). For example, decreases in dissolved O₂ and/or gas-phase soil pore O₂ reduce redox potential, causing facultative and obligate anaerobic microorganisms to shift their energy-yielding respiration processes to utilize alternative electron acceptors. For example, the transformation of nitrate (NO₃⁻), a soluble form of N that is limiting to primary productivity in freshwater and marine ecosystems, to gaseous forms of N (N₂O and N₂) via denitrification, is a major N removal pathway (M. M. Hefting et al., 2003). This process can reduce N loading to water bodies, but it will not proceed if O₂ is abundant. Further, the production of methane (i.e., methanogenesis) plays a key role in global C cycling and is completed by strictly anaerobic archaea that use oxidized C (e.g., CO₂) as an electron

Writing – review & editing: Kristen L. Underwood, Julia N. Perdrial, Andrew W. Schroth, Eric D. Roy, Carol E. Adair

acceptor (Lyu et al., 2018; Smyth et al., 2019). Soil O₂ therefore controls rates of redox-sensitive biogeochemical reactions that have important implications on global nutrient cycling (Bouwman et al., 2013; Groffman et al., 1988; Hanson & Hanson, 1996; Liptzin et al., 2011).

Soil O₂ levels are regulated by the diffusion of O₂ into and displacement of O₂ out of soil pores by water (physical processes), and the consumption of O₂ via soil respiration (a biological process, i.e., aerobic microbial, plant root, and faunal respiration; Moyano et al., 2013; Neira et al., 2015; Ponnampertuma, 1972). Because O₂ diffusion into soil water is much slower than in soil air (Moldrup et al., 2000), the presence of water inhibits O₂ diffusion from the atmosphere to soil pores (Skopp et al., 1990). Thus, the combined effects of inhibited O₂ diffusion and displacement, and soil respiration typically result in O₂ depletion (Neira et al., 2015; Ponnampertuma, 1972).

Our ability to predict soil O₂ concentrations across spatial and temporal gradients is limited. This is a result of the complex network of biotic and abiotic soil factors, as well as climatic conditions, that interact to modulate soil O₂ and moisture dynamics and create widespread spatial and temporal soil O₂ variability (Silver et al., 1999). For example, soil water inputs (precipitation and groundwater) vary seasonally, and with site-specific characteristics, and are modulated by hydraulic conductivity. Water demand (i.e., vegetation water uptake) also fluctuates seasonally and varies by plant species (Ewe et al., 2007). Furthermore, O₂ depletion by plant and microbial respiration is primarily controlled by soil temperature and soil water content and thus exhibits seasonal fluctuations (Chen et al., 2010; Kang et al., 2003; Lavigne et al., 2004).

Soil O₂ variability is difficult to manually monitor in-situ (i.e., using handheld soil probes or gas chromatography), and the collection of high spatial and temporal resolution O₂ data is costly, as it requires soil probes and data logging capabilities. However, our ability to predict soil O₂ concentrations across spatial and temporal gradients is limited. The challenges associated with measuring soil O₂ have led to the use of soil moisture as a proxy measurement for O₂ under the assumption that soil moisture is inversely proportional to O₂ concentration (Heinen, 2005; Ridolfi et al., 2003; Rubol et al., 2013). This assumption has been implemented in many simplified process-based denitrification sub-models embedded in N-cycling and ecosystem models (i.e., those that do not account for microbial processes or gaseous diffusion). Some of these models utilize bivariate nonlinear power functions that are modeled after an inverse relationship between O₂ and soil moisture to predict O₂ depletion based solely on water-filled pore space. Examples include the NEMIS model (Hénault & Germon, 2000), the LEACHMN model (Sogbedji et al., 2001), and the SHETRAN model (Birkinshaw & Ewen, 2000). Simplified process-based denitrification models that exclude direct O₂ measurements have been found to exhibit high sensitivity to formulations that represent soil moisture (Hénault & Germon, 2000).

Indeed, because a combination of multiple factors modulates the physical and biological mechanisms that control soil moisture dynamics and soil O₂ depletion (Silver et al., 1999), the use of soil moisture as a proxy measurement for O₂ can result in inaccurate O₂ estimations. An initial analysis using our high-frequency riparian soil sensor data shows a nonlinear relationship between O₂ and soil moisture, and other O₂ covariates (Figure 1 and Figure S1 in Supporting Information S1). The results show that most soil moisture levels are associated with a wide range of O₂ concentrations (Figure 1a and Figure S2 in Supporting Information S1), which suggests the relationship between these variables must be explored and defined using empirical data and nontraditional statistical methods.

Because riparian zones are located at the interface of terrestrial and aquatic ecosystems, they can function as hot spots for anaerobic biogeochemical soil processes (Vidon et al., 2010) and are therefore ideal study systems for soil O₂ dynamics. Due to their unique position on the landscape, riparian soils experience frequent hydrologic changes that alter soil moisture content, which can modify soil O₂ availability (Burgin & Groffman, 2012). Changes in soil moisture are triggered by hydrologic fluctuations, and the magnitude of these shifts depends on site-specific riparian zone characteristics, such as topography, proximity to surface and groundwater flows, the size and depth of the upland aquifer, and soil hydraulic properties. Riparian zones can also experience seasonal hydrologic fluctuations resulting from changes in connectivity with the upland aquifer (Vidon & Hill, 2004) and variability in water inputs due to seasonal precipitation patterns. Furthermore, site-specific dominant vegetation types have unique water requirements, which could impact the physical soil wetting process. The diverse potential combinations of soil O₂ drivers suggest that the response of soil O₂ to fluctuations in soil moisture is a result of multivariate interactions that are highly dependent on site-specific soil conditions, seasonal fluctuations in environmental conditions, and ecosystem water and O₂ demands.

Recent advancements in sensor technologies facilitate the simultaneous collection of multiple soil parameters at high temporal resolution, including O₂ and its relevant covariates (e.g., soil moisture, soil temperature, redox,

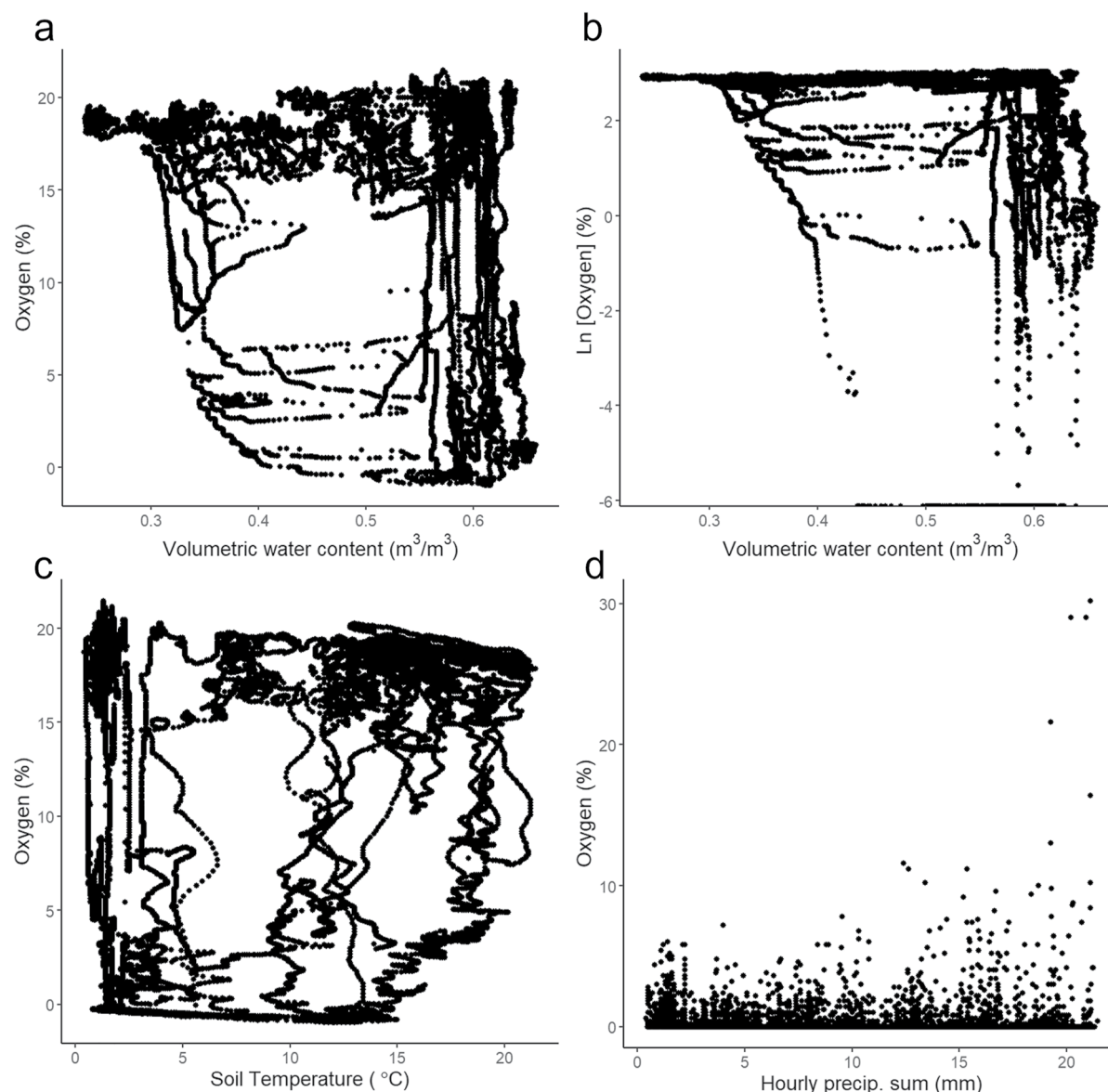


Figure 1. Scatter plots displaying nonlinear relationships between (a) O_2 and volumetric water content (VWC), (b) log transformed O_2 and original VWC values, (c) O_2 and soil temperature, and (d) O_2 and hourly precipitation sum. High frequency soil sensor data were collected in a frequently inundated position within a riparian area in northern Vermont, USA from 2017 to 2020.

CO_2 , precipitation). This enables us to comprehensively assess potential drivers of soil moisture variability and the related O_2 response. Importantly, two types of O_2 can be measured in soil: gaseous O_2 (in air-filled pore spaces between soil particles) and dissolved O_2 (in soil water and water films surrounding soil particles; Neira et al., 2015). We measured gaseous O_2 in the present study. Continuous in-situ monitoring of O_2 in soil is most commonly monitored as gaseous O_2 , as measurement of dissolved O_2 in soil requires extraction of soil water and O_2 measurement via sensor (optical or electrochemical), titrimetric, or colorimetric methods. While high-frequency data for multiple parameters is advantageous for ecosystem monitoring, it requires the use of tools that are specifically suited to analyze multivariate and nonlinear data. The Kohonen unsupervised self-organizing map (SOM), a type of artificial neural network, is a powerful clustering tool that can reliably analyze such multivariate and nonlinear data (Rivera et al., 2015), making it an ideal approach for detecting patterns in large environmental datasets. The SOM can overcome limitations of traditional statistical methods, as it can tolerate nonlinearity, temporal and serial autocorrelation, and multicollinearity (e.g., Figure 1; Kundu et al., 2013; Merdun, 2011). The

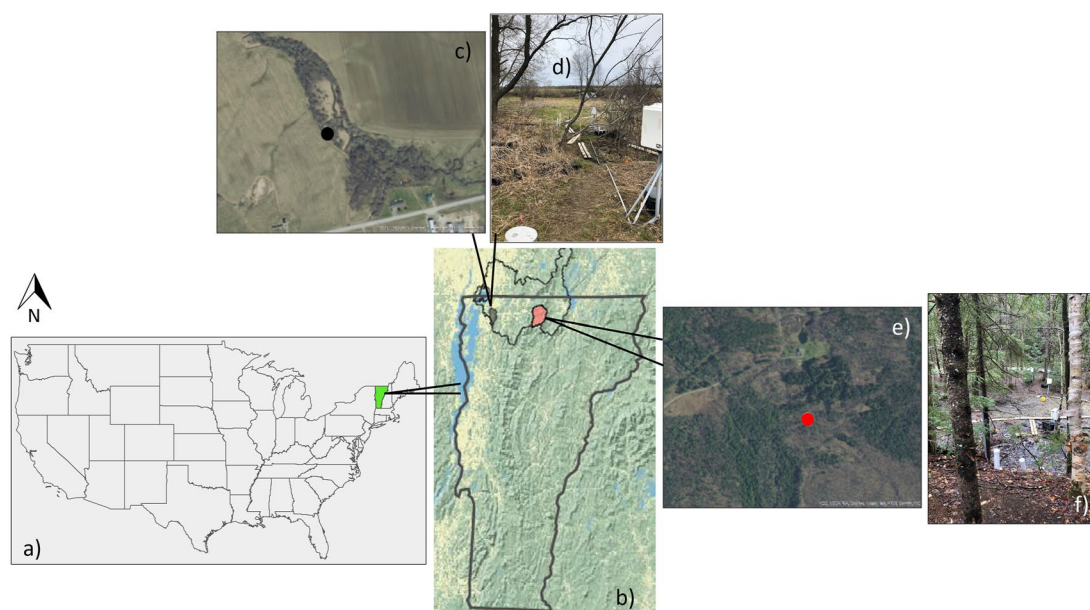


Figure 2. (a) Map of the USA with the State of Vermont highlighted in green. (b) Map of the State of Vermont, USA, and the province of Quebec, Canada, with The Missisquoi basin, a subbasin of the Lake Champlain Basin, outlined in black. The Champlain Valley (CV) site, which is located within the Hungerford Brook subwatershed, is shaded in gray. The Green Mountains (GM) site is located within the Trout River subwatershed. (c) Satellite image of the CV site (Sheldon, VT) with a black circle indicating where sensors are installed. (d) Photograph of the CV site riparian transect. (e) Satellite image of the GM site (Montgomery, VT) with a red circle indicating where soil sensors are installed and (f) photograph of the GM transect.

SOM is often applied in exploratory data analysis to leverage the temporal (or spatial) autocorrelation that may exist in the data and identify clusters of like observations (Kalteh et al., 2008; Kohonen, 2013).

The SOM maps multivariate data to a two-dimensional map/lattice, where similar data points are situated in proximity. In contrast to other, more traditional clustering algorithms, (e.g., k-means), the SOM approach enables visualization of variables that drive clustering, and thus, is a potentially powerful statistical tool for leveraging the capacity of high frequency sensor networks to monitor physical and biogeochemical parameters. SOMs have been successfully applied to resolve spatial and temporal heterogeneity in complex systems within large soil and water quality databases (Liao et al., 2019; Obach et al., 2001; Wu et al., 2008), as well as to classify sediment (Alvarez-guerra et al., 2008) and soil types (Tissari et al., 2007). Additionally, the SOM approach has been utilized to address questions concerning water resources and hydrology, such as rainfall-runoff relationships (Lin & Chen, 2006), precipitation dynamics (Kalteh et al., 2008), and links between physical soil properties and hydrologic soil processes (Merdun, 2011). However, to our knowledge, these tools have not yet been used to detect patterns in high frequency soil sensor time series.

Using the SOM approach, we addressed two main questions: (a) Can the SOM approach provide additional insight on the O_2 /soil moisture relationship by incorporating multiple predictors, and (b) Which combinations of variables lead to high versus low soil O_2 ? We clustered high frequency soil and meteorological data collected over 3 years from a poorly drained wetland position within two riparian sites located in northeastern Vermont, USA. We studied two riparian soil environments with contrasting site characteristics (e.g., adjacent land use, vegetative cover, site elevation), allowing us to address our questions in two different riparian lowland settings.

2. Methods

2.1. Study Sites

To investigate and better characterize O_2 variability within riparian soils, we studied two riparian soil transects with contrasting catchment characteristics. Both transects are located within Lake Champlain's Missisquoi Watershed in Vermont, USA (Figure 2) and are part of a larger soil monitoring network. This high frequency soil sensor network continuously measures physical and chemical soil conditions 15 cm below the soil surface

Table 1
Meteorological Conditions Observed Throughout the Sampling Period at the Champlain Valley (CV) and Green Mountains (GM) Site, Separated by Meteorological Season

Season	Months	Site	Cumulative precip. (cm)	Mean ambient T (°C)	Mean PAR (μE)	Mean solar radiation (W/m ²)	Mean relative humidity (%)
Winter	December–February	CV	15.06	−5.93	265.53	162.91	80.50
		GM	18.34	−6.48	46.89	22.64	85.44
Spring	March–May	CV	23.52	5.23	326.26	256.78	75.12
		GM	27.86	4.00	196.72	100.30	75.78
Summer	June–August	CV	25.65	19.66	345.58	305.67	79.86
		GM	31.82	18.02	241.16	137.78	84.25
Fall	September–November	CV	33.55	7.87	276.45	201.72	83.80
		GM	38.16	6.63	71.47	40.50	88.19

Note. The data were collected from meteorological stations installed at each site. PAR stands for photosynthetically active radiation.

along a gradient of landscape positions (i.e., spanning upland, wetland, and near-stream locations). Included in this study are data collected from one low-lying, poorly drained position within each transect, where soil O₂ concentrations ranged from anoxic to near atmospheric. One transect is situated within the Champlain Valley (CV) physiographic province (hereafter referred to as “CV” site), a primarily agricultural catchment in Sheldon, VT. The other transect is located within a 95% forested catchment with minimal anthropogenic impact that is located within the Northern Green Mountain physiographic province (hereafter referred to as “GM” site), approximately 7 km north of the town of Montgomery, VT (Landsman-Gerjoi et al., 2020).

The elevation range of the CV site spans 101–106 m above sea level. Vegetation at this site includes American beech (*Fagus grandifolia*) trees, various fern species (group *Pteridophytes*), and nettles (*Urtica dioica*) (Landsman-Gerjoi et al., 2020). Soil types at the CV site include Inceptisols (Aquic Dystric Eutrudepts) and Entisols from glaciolacustrine and glaciofluvial fluvial material (i.e., Aquic Udipsamments and Fluvaquentic Dystrudepts; Soil Survey Staff et al., 2019). The Green Mountains (GM) site is higher in elevation (350–365 m above sea level) and its vegetation is characteristic of a secondary growth northern hardwood forest, including sugar maple (*Acer saccharum*), yellow birch (*Betula alleghaniensis*), white ash (*Fraxinus Americana*) and red spruce (*Picea rubens*) (Landsman-Gerjoi et al., 2020). Soils at the GM site are Inceptisols (i.e., Fluvaquentic Dystrudepts and Fluvaquentic Endoaquepts; Soil Survey Staff et al., 2019). Both study sites experience a temperate climate with four distinct seasons, including snow-dominated winters (22% and 24% of annual precipitation at CV and GM site, respectively), a snow melt period, temperate summers with occasional rains, and a fall season with high litter input (Landsman-Gerjoi et al., 2020). Differences in elevation result in contrasting meteorological conditions between the two sites (Table 1). Groundwater table depth was measured continuously at 15-min intervals at both sites using HOBO water level loggers (Onset Corporation, Cape Cod, MA, USA) that were installed inside hand dug wells in both upland and frequently inundated areas.

2.2. Soil Monitoring Network

We assessed ecosystem water delivery by measuring soil volumetric water content (VWC) and precipitation. Water demand (i.e., vegetation water uptake) was indirectly assessed by measuring ambient temperature (TA). Though soil carbon dioxide (CO₂) measurements are not flux measurements and cannot be used as a direct indicator of soil respiration or O₂ demand, CO₂ measurements were used as a proxy for O₂ demand. Previous studies (e.g., Jarecke et al., 2016) have documented strong inverse relationships between subsurface (10 cm) gas-phase soil pore O₂ and CO₂. Soil volumetric water content (VWC; measuring range = 0–1.0 m³/m³; accuracy = ±0.03 m³/m³), temperature (T; measuring range = −40 to +60°C; accuracy = ±1°C), and electrical conductivity (EC; measuring range = 0–23 dS/m; accuracy = ±10% from 0 to 7 dS/m) were monitored at 15-min intervals using 5TE sensors (Meter Group, Pullman, WA).

Carbon dioxide in soil air was measured using GMT221 sensors (Vaisala, Helsinki, Finland; measuring range = 0–20%; accuracy = ±1.5% of range and ±2% of reading). Before deployment, the CO₂ sensors were calibrated with a three-point adjustment and corrections for barometric pressure (based on elevation) and humidity. Pure N₂ gas was used as a 0% CO₂ standard. The internal temperature sensor on the CO₂ probes was used for the temperature correction. Oxygen in soil air (i.e., gas-phase soil pore O₂; hereafter referred to as soil O₂) was monitored at 15-min intervals using Soil Response O₂ sensors (Apogee instruments, Logan, UT; measuring range = 0–100%; sensitivity = 2.6 mV per % O₂). Before deployment, the O₂ sensors were calibrated based on an open-air oxygen reading with corrections for barometric pressure (based on elevation), temperature, and humidity. The gas permeable membrane inlet of the O₂ sensors was encased in a porous diffusion head and was equipped with a heating element to prevent water from pooling and blocking O₂ diffusion. The O₂ sensors are subject to long-term drift (1 mV/year).

Oxygen and CO₂ sensors were installed within 45 cm of each other and were deployed inside polyvinyl chloride (PVC) pipe (with membrane inlets exposed to the soil) to secure their position in the soil. Silicone sealant was applied between the O₂ sensor diffusion head and the PVC pipe to prevent

water from entering the electrical connections. For water-proofing purposes, CO₂ probes were deployed inside a gas-permeable membrane sleeve. We filled a portion of the membrane sleeve with silicone sealant to prevent moisture from entering the electrical connections between the probe and the data logger cable. Because the sensors were shielded from the soil and enclosed within either a diffusion head (for O₂) or a membrane (for CO₂), it is probable that the air inside the chambers took some time to reach the same concentration as the surrounding soil. Each site (CV and GM) was equipped with a meteorological station that measured TA, precipitation, photosynthetically active radiation, solar radiation, relative humidity, wind direction, wind speed, and dew point at 5-min intervals. Data included in this study were collected from July 2017 to June 2020.

2.3. Data Analysis

2.3.1. Approach

To identify suites of soil conditions (clusters) associated with various soil O₂ regimes, we employed a Kohonen unsupervised SOM approach using the *kohonen* package in R (Wehrens & Buydens, 2007). We used an “unsupervised approach,” meaning we fed independent variables to the model and excluded the response variable, O₂. Furthermore, we did not constrain the number of outcome clusters (i.e., soil condition descriptors), so that we could empirically determine the most suitable number of clusters for our dataset. We first used exploratory data analysis to identify independent variables with a potential to be linked to O₂ variability within our sites, and to examine meaningful ranges of O₂ as the response variable. We also selected input variables by examining component planes generated by early iterations of the SOM, which allow for the visualization of clustering according to each independent variable. We then ran a SOM analysis approach, which involved several iterative steps, to optimize SOM execution and validate clusters, as described below. The SOM mapped our multivariate dataset to a two-dimensional map/lattice, where observations linked to similar combinations of values for input variables were situated in proximity to each other. Finally, we compared O₂ values across clusters, *post hoc*, to better understand drivers of specific O₂ ranges.

2.3.2. Overview of Kohonen Unsupervised Self-Organizing Map

A detailed description of the SOM algorithm can be found in Kohonen (2013) and Underwood et al. (2021). To summarize, the method clusters multivariate observations onto a reduced-dimension lattice. Each lattice node is first assigned a vector of random values (weights) ranging from 0 to 1. The length of this vector is equal to the number of input variables in each observation. A single vector of input values (observed data) is simultaneously presented to each node's weight vector. The vector of input values is compared to the weight vector using the Euclidian distance formula, and the lattice node with the closest matching weight vector is designated as the best matching unit (BMU). The weight vector of the BMU, in addition to nodes surrounding the BMU (defined as a “neighborhood”), is updated to resemble the input vector more closely. The neighborhood function is unique to the SOM, as other clustering tools (e.g., K-means) only update the weight vector of a single node (Merdun, 2011). The user customizes the learning rate, α , which controls the amount by which weights are adjusted for both the BMU and nodes within the neighborhood around the BMU. This process is repeated until all observations have been presented to the lattice, which constitutes one iteration of the SOM algorithm. Both the learning rate, α , and the neighborhood size are decreased as subsequent iterations of the algorithm are executed, by relying on user-defined functions. The size of the neighborhood is eventually reduced to one node—the BMU. Multiple iterations are executed until the algorithm converges. Once the algorithm converges, the adjusted weight vectors will have self-organized across the lattice such that similar observations will be aggregated together. To define clusters of observations (i.e., nodes of the lattice containing similar weight vectors), the distance between weight vectors is calculated using a hierarchical agglomerative clustering method (Underwood et al., 2021).

2.3.3. Data Analysis: Initial Data Conditioning, Exploratory Analysis, and Selection of Independent Variables

To prepare our high frequency soil and meteorological observation data for input to the SOM, data conditioning steps were required. Data from 27 June 2017 to 27 June 2020 from both study sites were selected for analysis. Some observations were missing from the time series due to sensor and data logger dysfunction caused by inadequate solar power, mostly during winter and early spring. Timesteps with missing O₂ or precipitation observations were excluded from the dataset, leaving 91% of the CV time series and 27% of the GM time series to use as SOM input data. Carbon dioxide was not included as an input variable to the GM SOM, as all CO₂ observations

were missing from the dataset. After removing timesteps with missing O_2 or precipitation observations, missing values of other input variables (ranging from 0% at the CV site to 0.9% of observations at the GM site; Table S1 in Supporting Information S1) were replaced with the overall median value using the Hmisc package's impute function in R (Harrell, 2020; Table S1 in Supporting Information S1). We then reduced the volume of data by downscaling our 15-min observations to an hourly frequency using hourly median values for each variable. We derived additional time series to potentially include in our dataset by computing rolling averages (for VWC, soil T, and ambient T) and rolling sum (for precipitation) ranging from 12 hr to 2 weeks prior to each observation to investigate the effects of antecedent soil and meteorological conditions on soil O_2 dynamics. We calculated the Julian date of each observation to include in the SOM as an independent variable. We then performed Principal Component Analyses (PCA), a dimensionality reduction tool, on a correlation matrix of all available independent variables to confirm that the O_2 data clustered distinctly, based on the available parameters. PCA analyses were also used to establish the dimensions of the SOM lattice best suited to our dataset (Underwood et al., 2021).

We selected a suite of independent variables to include in our final SOMs (Table S2 in Supporting Information S1) by first running provisional SOMs with all available independent variables (listed in Supporting Information S1). We observed the resulting component planes and plotted the distribution of each variable, separated by cluster, using box and whisker plots. We examined the component planes (Figures S7–S10 in Supporting Information S1) and box and whisker plots to manually identify variables that lacked distinct variability across clusters and excluded those variables from the input dataset (e.g., wind speed). A clear pattern or structure revealed by a component plane indicates that the variable is an important contributor to the dataset's variability. A uniform or random pattern suggests that the variable does not explain much variability.

2.3.4. Data Analysis: SOM Data Preprocessing

We prepared the input data for the SOM by normalizing each independent variable to a value between 0 and 1 using a range normalization technique. Normalization improves model performance by ensuring different measurement units and magnitudes do not influence the weight of each independent variable. In particular, range normalization has been found to result in optimal SOM performance, as it minimizes the distance between the data points and the nodes to which they are matched (measured as quantization error (QE)), thus improving model fit. QE represents the magnitude of error between the input vector and its closest match on the SOM. Thus, a lower QE indicates the SOM is more effectively representing the input data in a lower-dimensional space. The QE can be used to compare the performance of different iterations of the SOM trained with the same input dataset (not to compare between SOMs with different input datasets; C  r  ghino & Park, 2009). Additionally, range normalization reduces discontinuities between the structure of the input data and the map layout (measured as topographic error; Alvarez-Guerra et al., 2008; Breard, 2017). The response variable, O_2 , was not presented as an input to the SOM.

We ran SOMs on different groupings of the CV observations to evaluate which driving features were revealed when differing amounts of data were fed to the SOM. We initially ran one SOM for all the soil O_2 data from the CV site and found that the very abundant high O_2 values masked the signal of the less frequent low O_2 values. This prevented identification of the nuanced conditions or factors that led to low O_2 . We therefore performed SOMs on separate subsets of the CV data for a more refined assessment of factors associated with O_2 dynamics, and to ensure that factors associated with high O_2 conditions could be parsed from factors associated with low O_2 conditions. We subdivided the associated multivariate time series observations into sets associated with distinct ranges of O_2 values (high and low), using the Jenks Natural Breaks optimization method (Jenks, 1967; Khamis et al., 2018) via the BMMtools package in R (Rabosky et al., 2014). The Jenks Natural Breaks optimization method is a data clustering technique designed to classify data into a user-defined number of ranges in a way that minimizes within-group distances between values. We ran only one SOM for all O_2 observations from the GM site, as O_2 values ranged from 0% to 0.6%, a range of values that corresponded to low O_2 conditions determined for the CV site.

2.3.5. Data Analysis: SOM Computation and Model Optimization

After selecting a suite of independent variables to include in each dataset (CV all observations (CVAO), CV high O_2 , CV low O_2 , GM), we adjusted the number of lattice nodes, the lattice dimensions, and the value of k to maximize between-cluster variance and minimize within-cluster variance. It is important to optimize the dimensions of the SOM lattice for a given dataset, as an unsuitable lattice configuration can distort the distribution of the

input variables across the map. For each dataset, we followed Vesanto's rule (Vesanto et al., 2000) to pinpoint the optimal number of lattice nodes. We also approximated the column-to-row ratio of the lattice as the ratio of the two largest eigenvalues from PCA using the original, non-normalized values (C  r  ghino & Park, 2009). A hexagonal lattice arrangement was used. The lattice for the CVAO, CV high O₂, CV low O₂, and GM SOM contained 16 rows and 57 columns, 17 rows and 34 columns, 13 rows and 33 columns, and 15 rows and 27 columns, respectively. The CVAO, CV high O₂, CV low O₂, and GM SOM lattice configuration contained 912, 578, 429, and 405 nodes, respectively.

SOM training was performed over 20,000 iterations, and α was set to decrease linearly from 0.05 to 0.01. The neighborhood size started at a radius from the BMU of two-thirds the lattice size and decreased linearly to a value of 0, at which point only the weights of the BMU were being updated. For a given dataset, the SOM iteration that maximized the nonparametric F statistic (ratio of within-cluster to between-cluster variance) and minimized QE (measure of map resolution) was selected as the final model run (Underwood et al., 2021). The nonparametric F-statistic was calculated using the *adonis* function in R's *Vegan* package (Oksanen et al., 2019). The F-statistic values for the CVAO, CV high O₂, CV low O₂, and GM SOM were 63,801, 30,891.5, 31,395.1, and 32,232.6, respectively. Quantization error values (unitless) for the CVAO, CV high O₂, CV low O₂, and GM SOM were 0.002152, 0.000505, 0.000305, and 0.000337, respectively.

For a given dataset, to identify unique attributes of each cluster, we plotted the (range normalized) intra-cluster mean, relative to the overall mean, for each input variable. This metric was used to visualize suites of variables and their relative value ranges that constituted different environmental conditions associated with various O₂ levels. Mixed-effect models that incorporated random effects for temporal autocorrelation were used to test for differences in each input variable according to cluster assignment using the *lme* function in R's *nlme* package (Pinheiro et al., 2014). Standard errors of the mean of original (i.e., not range-normalized) values were compared to assess whether inter-cluster means were meaningfully different. The cluster assignments for each observation were then plotted onto an O₂ time series figure (original values) to display temporal fluctuations in classes of environmental conditions.

3. Results

3.1. Champlain Valley Site: Champlain Valley All Observations (CVAO) SOM

Because the processes that control meteorological and soil conditions exhibit patterns that repeat over time, high frequency environmental observations are usually temporally autocorrelated. We detected temporal autocorrelation in our input data for each SOM (Figures S3–S6 in Supporting Information S1). At the CV site, we first performed a SOM on all observations (O₂ = 0–21.47%, n = 22,427) to assess whether our questions could be adequately addressed at this scale (Figure 3). After examining the component planes generated by the input variables (Figures S7–S10 in Supporting Information S1), we selected key input variables for the CVAO SOM (mean O₂ = 11.14%): soil T (mean = 8.6  C), 1-week cumulative antecedent precipitation (hereafter referred to as “1-week precipitation”) (16.28 cm), VWC (0.51 m³/m³), Julian date (183), CO₂ (2,805 ppm), EC (0.19 dS/m), and dew point (2.41  C). The CVAO SOM assigned observations to six different clusters (Figure 3c and Table 2). Most (48.4% of observations) were categorized by the SOM as wet and cool, with low (below average) CO₂, EC, and 1-week precipitation (clusters 5 and 6). These clusters occurred during early spring (cluster 5) and winter (cluster 6). Clusters 2 and 3 (21.2% of observations) fell into a dry and warm category with high (above average) CO₂ and dew point, and low 1-week precipitation and EC. Both clusters occurred during summer months, but cluster 3 occurred during early spring, too. Cluster 1 (17.5% of observations) included dry and warm soil conditions, with high 1-week precipitation, CO₂, and dew point, and low EC. Cluster 1 occurred from mid-summer to fall. Some (12.9%) observations were grouped into a wet category with high 1-week precipitation, EC, and dew point, and low CO₂ (cluster 4). Cluster 4 occurred during winter months.

While the dry and warm (July–October) and wet and cool (October–April) categorization revealed a seasonal pattern in the variability of the key controls on soil O₂, clusters derived from all CV observations did not adequately differentiate between conditions that led to high versus low O₂. This is evidenced by the wide range of O₂ values included in clusters 1 (11%–20%), 2 (2%–20%), 3 (0.1%–20%), 4 (1%–20%), 5 (0.1%–21%), and 6 (0%–21%). To more effectively address our questions, further sub-setting of our data into observations associated with low O₂ versus high O₂ was therefore needed to parse the unique factors (and possibly different combinations

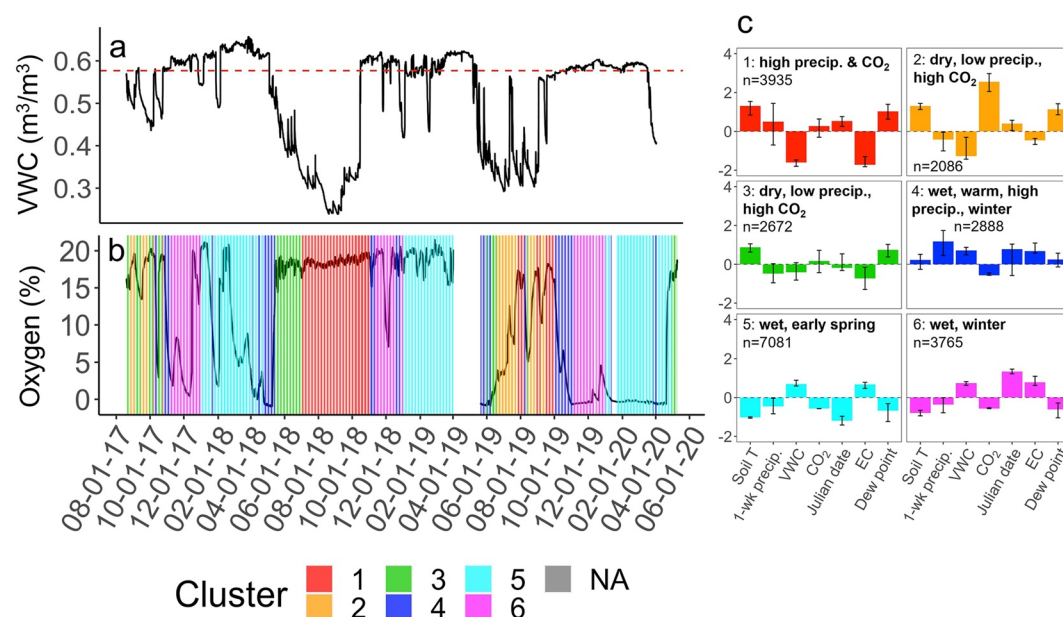


Figure 3. Results from the Champlain Valley all observations (CVAO) self-organizing map (SOM), including (a) volumetric water content (VWC) time series with dashed line representing mean VWC for the CVAO dataset, (b) O_2 time series highlighted with the six clusters identified by the CVAO SOM, (c) bar plots displaying range normalized intra-cluster means of each input variable (n = number of observations per cluster). “NA” represents O_2 values that do not have corresponding independent variable values. 1-wk precip., one-week antecedent precipitation; EC, electrical conductivity.

of factors) within or across seasons that led to anoxic and oxic conditions. Based on Jenks Natural Breaks, we therefore split the CV observations into two datasets associated with high (12.9%–21.5%, $n = 12,593$) and low (0%–4.3%, $n = 7,043$) O_2 to differentiate conditions that led to contrasting O_2 regimes. We performed separate SOMs (CV high O_2 and CV low O_2) on those subsets.

3.2. Champlain Valley Site: High O_2 SOM

Key input variables selected for the CV high O_2 SOM (mean $\text{O}_2 = 18\%$) were determined by examining component planes and were identified as soil T (mean = 10.3°C), 2-week cumulative antecedent precipitation (hereafter referred to as “2-week precipitation”) (2.74 cm), VWC ($0.46 \text{ m}^3/\text{m}^3$), CO_2 (3,215.2 ppm), and Julian date (192) (Figure S8 in Supporting Information S1). The high O_2 SOM identified four distinct clusters of multivariate time series observations (Figure 4c). High O_2 events, which made up 49% of the CV O_2 values, were somewhat evenly distributed among winter (26% of data points) spring (17%), summer (25%), and fall (32%). Oxygen values were consistently high from May 2018 through April 2019 (Figure 4b), and further research is required to identify what prevented O_2 depletion under wet and cool soil conditions during this period. At the CV sampling location, groundwater levels generally decreased during the growing season (early May to early October; average = 0.82 m below the soil surface) and increased during cooler months (average = 0.38 m below the soil surface).

Soil conditions within clusters 1 and 4, which made up 62% of the CV high O_2 data points, were generally dry and warm (Table 3). Cluster 1, which included 53% of data points, is associated with average (compared to the overall mean) 2-week precipitation and average CO_2 values. Contrastingly, cluster 4, which encompassed 8.6% of data points and had below average 2-week precipitation, was associated with the highest average CO_2 of all four clusters. There was overlap of the Julian date ranges of clusters 1 and 4 during warm months (Table 3). Wet and cool soil conditions (clusters 2 and 3) described 38% of data points within the CV high O_2 dataset and were associated with the highest average O_2 of all clusters. Key differences between clusters 2 and 3 include Julian date range (October–January, and January–April, respectively) and 2-week precipitation (39.9 and 19.2 cm, respectively) (Table 3).

3.3. Champlain Valley Site: Low O_2 SOM

Low O_2 values were distributed somewhat uniformly among the four seasons (winter = 32%, spring = 35%, summer = 12%, and fall = 21%). Low O_2 values represented 27% of the O_2 dataset and occurred intermittently

Table 2
Mean Value of Each Input Variable and O₂ Across Six Clusters Identified by the Champlain Valley All Observations Self-Organizing Map

Cluster	O ₂ (%)	Soil temp. (°C)	VWC (m ³ /m ³)	CO ₂ (ppm)	1-week cumulative antecedent precip. (cm)	EC (dS/m)	Dew point (°C)	Conditions	Julian date ranges
1	18.1 ^a	16.6 ^a	0.3 ^a	4,290.1 ^a	24.7 ^a	0.04 ^a	13.4 ^a	High precip. & CO ₂	July–November
2	11.7 ^b	17.6 ^b	0.4 ^b	15,042.5 ^b	10.7 ^b	0.13 ^b	15.4 ^b	Dry, low precip., high CO ₂	June–September
3	14.3 ^c	14.5 ^c	0.5 ^c	4,023 ^a	9.2 ^c	0.13 ^b	10.5 ^c	Dry, low precip., high CO ₂	April–October
4	5.8 ^d	9.2 ^d	0.6 ^d	776.7 ^c	35.4 ^d	0.26 ^c	4.6 ^d	Wet, warm, high precip., winter	January–December
5	9.6 ^c	1.8 ^e	0.6 ^d	424.9 ^d	9.7 ^c	0.23 ^d	−7.4 ^e	Wet, early spring	January–April
6	7.6 ^f	3.4 ^f	0.6 ^d	582.8 ^d	10.3 ^b	0.27 ^e	−6.1 ^f	Wet, winter	October–December

Note. Within the same column, different letters represent significant differences between clusters ($p < 0.05$).

throughout fall 2017 and winter 2018. Low O₂ values were observed consistently from fall (September, October, and November) 2019 through spring (March, April, and May) 2020. There were consistent periods of low O₂ during the spring snowmelt period (April–May) of 2018, 2019, and 2020 (Figure 5a). Based on an examination of component planes generated by the potential input variables (Figure S9 in Supporting Information S1), we included the same suite of input variables in the CV low O₂ SOM as the CV high O₂ SOM (soil T, 2-week precipitation, VWC, CO₂, and Julian date). Average values for these variables were 5.8°C, 3.52 cm, 0.57 m³/m³, 1,635.2 ppm, and 175.2, respectively. Median values of each input variable were significantly different between the CV high and low O₂ datasets ($p < 0.001$). The SOM identified five distinct clusters for the CV low O₂ SOM (Table 4). In contrast to the CV high O₂ SOM results, the majority of low O₂ data points (69%) could be categorized as wet and cool with low 2-week precipitation and CO₂ (clusters 1 and 3). Clusters 1 and 3 differed in Julian date ranges (Figure 5c). Clusters 4 and 5 included 9.3% of data points, which fell into a very dry and warm category, with average (cluster 4) and above average (cluster 5) CO₂ and 2-week precipitation (Figure 5c). Unique to the CV low O₂ SOM, an additional cluster was identified (cluster 2), which encompassed 21.5% of data points. Cluster 2 occurred throughout October and November, and May and June, and could be described as warmer than average, with average soil moisture, despite above average antecedent precipitation and below average CO₂ (Table 4).

3.4. Green Mountains Site SOM

We included all values from the GM dataset in one SOM, as O₂ values at the GM site were all low, ranging from 0% to 0.6% ($n = 6,921$). A unique set of input variables was chosen for the GM model based on the component planes (Figure S10 in Supporting Information S1): cumulative 2-week antecedent ambient temperature (hereafter referred to as 2-week TA), VWC, 1-week precipitation, and Julian date. Average values for these variables were 11.0°C, 0.53 m³/m³, 1.99 cm, and 186.6, respectively. The low O₂ SOM grouped the observations into 5 different clusters. Observations from the GM dataset can generally be described as dry and warm, wet and cool, or wet and warm (Table 5). Average groundwater levels at the GM sampling location were significantly higher ($p < 0.001$; average = 0.05 m below the soil surface) and remain elevated throughout the year.

Cluster 2 encompassed the most data points (46.8%), which can be described as dry and warm, with below average 1-week antecedent precipitation, occurring during early spring/summer. Clusters 3 and 5 included 28% of data points, which can be summarized as warmer and wetter than average conditions, with differing 1-week precipitation and Julian date ranges (Figures 6b and 6c). Cluster 4 (15.4% of data points) can be described as cool, and drier than average, with low 1-week precipitation. Encompassing 9.3% of data points, cluster 1 can be described as wet and cool, with low 1-week precipitation (Figure 6c).

Oxygen values included in the GM SOM remained consistently at or below zero percent throughout the entire sampling period, with the exception of an O₂ event in late December 2017 that reached 0.6% (Figure 6c). As such, it should be noted that the clustering results of the GM dataset incorporate the single O₂ event (Cluster 1) and therefore may not be representative of typical conditions observed at the GM site. Clusters 1 and 4 occurred throughout winter 2017/2018 and were associated with wet and cool soil conditions. Due to the interference of winter weather with our instrumentation, we were not able to monitor winter 2018/2019 or 2019/2020. Clusters 2 (warm/dry), and 3 (warm/wet) occurred intermittently from May to October 2018, and from June to September in 2019. O₂ values within cluster 5 (warm/wet) occurred as isolated events each summer during the months of July (2018 and 2019) and May and June 2020.

4. Discussion

4.1. SOM Results Confirm Multivariate and Complex Controls on O₂

We addressed the question: Can the SOM approach provide additional insight into the O₂/soil moisture relationship by incorporating multiple predictors? Our results confirm that using a unique combination of high frequency, multiparameter soil sensor data collected over multiple years, the SOM identified combinations of variables that control soil O₂ levels. Importantly, and in contrast to traditional ecological assumptions, low O₂ levels did not correspond solely with increasing soil moisture. Indeed, high O₂ levels persisted in both high and low moisture conditions. Low O₂ conditions instead depended on temporally varying

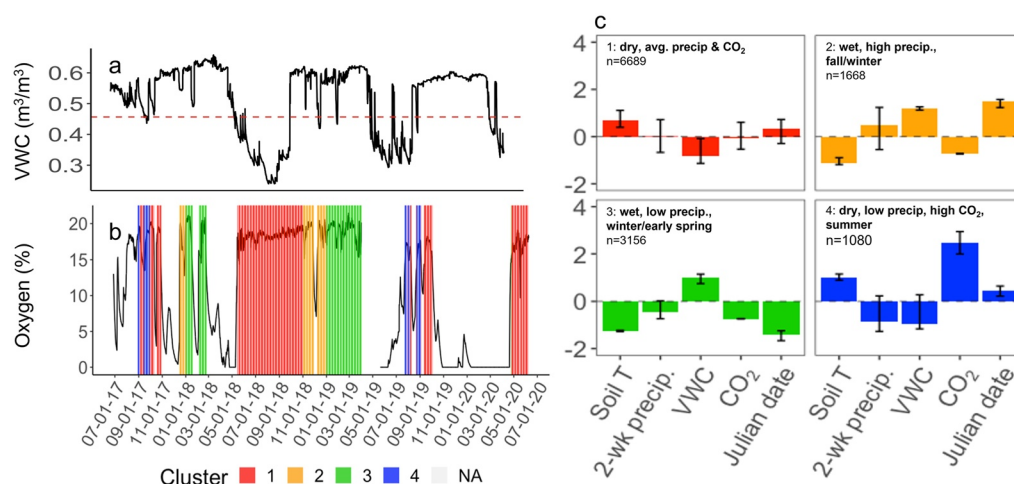


Figure 4. High O_2 self-organizing map (SOM) results for the Champlain Valley (CV) site, including (a) volumetric water content (VWC) time series with dashed line representing mean VWC for the CV high O_2 dataset, (b) O_2 time series highlighted with the four clusters identified by the high O_2 SOM, (c) bar plots displaying range normalized intra-cluster means of each input variable (n = number of observations per cluster). Clusters that are not shaded (represented by “NA”) correspond to O_2 values outside of the high O_2 range. 2-wk precip., two-week antecedent precipitation.

combinations of water inputs, water demand, and O_2 demand. The output of our unsupervised clustering analysis from both field sites (i.e., all four SOMs) revealed a pronounced seasonal signal. This is evidenced by clustering according to dry soil conditions during warm months and wet soil conditions during cool months, with varying antecedent precipitation, soil CO_2 , and soil temperature conditions.

Our results show that the information gained from the SOM clustering changes with the magnitude of variability in the input data. When all observations from the CV site were included in one SOM, we were able to detect the seasonal pattern driving fluctuations in the controls on soil O_2 but not the nuanced conditions that led to oxic versus anoxic O_2 . We thus ran separate SOMs on observations associated with high and low O_2 and hereafter refer to the results from those models.

At the CV site, the majority (69%) of low O_2 values were associated with wet and cool soil conditions with below average CO_2 and 2-week antecedent precipitation. These observations occurred from November to May (Figure 5, cluster 3). Under these conditions, a physical saturation process (as opposed to biological O_2 consumption) dominated, as the combined effects of decreased water demand (Sevanto et al., 2006) and sufficient water inputs from precipitation likely prevented the reaeration of soil pores (Neira et al., 2015). Furthermore, in Northern temperate climates, ground frost can persist during this period (November–May), especially in areas of open land without significant canopy cover (Shanley & Chalmers, 1999), which could have prevented the reaeration of soil pores. All low O_2 values at the CV site that were observed under wet and cool conditions were associated with below average subsurface CO_2 (used here as an indicator of soil O_2 demand), indicating that O_2 consumption rates were relatively low under the majority (69%) of low O_2 conditions. This finding is consistent with those of Davidson et al. (1998) and Moyano et al. (2013) who observed decreased soil respiration rates under wet and cool soil conditions during the non-growing season, which resulted from reduced plant respiration and high soil moisture levels impeding O_2 diffusion, and thus, decomposition and CO_2 production (Doran et al., 1990; Moyano et al., 2013; Skopp et al., 1990). It is therefore possible that a restriction of air exchange between the atmosphere and soil pores is necessary in order for low levels of biological soil respiration to markedly deplete O_2 before it is replenished. We note that the effects of subtle shifts in CO_2 on O_2 depletion under wet and cool soil conditions may have occurred at time scales that were finer in resolution than our hourly SOM input data. Therefore, analyzing a dataset of finer temporal resolution, or one that encompasses a shorter time period, may help detect a more significant impact of biological O_2 consumption under wet and cool conditions.

Contrastingly, most (62%) high O_2 values from the CV site were associated with dry and warm conditions that occurred during warmer months (April–October, Figure 4). This indicates that dry and warm soil conditions inhibited soil O_2 depletion by allowing O_2 to readily exchange with the atmosphere via increased air-filled pore

Table 3
Mean Value of Each Input Variable and O_2 Across Four Clusters Identified by the Champlain Valley High O_2 Self-Organizing Map

Cluster	O_2 (%)	Soil temp. ($^{\circ}C$)	VWC (m^3/m^3)	CO_2 (ppm)	2-week cumulative antecedent precip. (cm)	Conditions	Julian date ranges
1	17.8 ^a	15.2 ^a	0.4 ^a	3,570.4 ^a	30.2 ^a	Dry, avg. precip. & CO_2	April–October
2	18.2 ^b	3.2 ^b	0.6 ^b	493.5 ^b	39.9 ^b	Wet, high precip., fall/winter	October–January
3	18.6 ^c	1.3 ^c	0.6 ^c	525.7 ^b	19.2 ^c	Wet, low precip., winter/early spring	January–April
4	16.8 ^d	17.7 ^d	0.4 ^a	13,077.4 ^c	15.6 ^d	Dry, low precip., high CO_2 , summer	July–September

Note. Within the same column, different letters represent significant differences between clusters ($p < 0.05$).

space. However, as 62% of high O_2 values were associated with above-average soil CO_2 levels (clusters 1 and 4), our findings suggest that water limitation did not suppress soil respiration. Furthermore, these findings contradict previous studies by Doran et al. (1990) and Orchard and Cook (1983) that documented decreased soil respiration rates resulting from elevated soil temperatures and low soil matric potential during warm months. It is therefore likely that, under dry and warm soil conditions at our sites, sufficient soil moisture is required to block O_2 diffusion in order for elevated soil respiration rates to sufficiently deplete O_2 . We note that VWC may need to decrease below field capacity (not measured in this study) in order for water limitation to significantly reduce soil respiration rates (Davidson et al., 1998). Additionally, dry and warm soil conditions coincide with the growing season in temperate systems where the combined effect of elevated plant water uptake and negative soil matric potential can lower hydraulic conductivity (i.e., inhibit additional water inputs from percolating through the soil matrix) (Hardie et al., 2012). Under this scenario, additional precipitation inputs may not have resulted in increased VWC and therefore, despite elevated CO_2 levels (i.e., high O_2 demand) at this time, O_2 usually remained near atmospheric levels.

4.2. High Soil Moisture Levels Do Not Always Lead to Low O_2

Although the majority of high O_2 levels occurred when soils were dry and warm, and most low O_2 levels occurred under wet and cool conditions, we also intermittently observed the opposite behavior, indicating soil moisture alone is not always an accurate proxy for soil O_2 . Similar to our findings, Burgin and Groffman (2012) observed low gas-phase soil pore O_2 under dry VWC conditions in a riparian wetland during summer. These exceptions hold key insights into the importance of antecedent conditions (i.e., time lag effects) in determining soil O_2 conditions. For example, low O_2 occurred in June and July under dry and warm (below average VWC) conditions with above-average 2-week antecedent precipitation inputs. This is evidenced by clusters 4 and 5 from the CV low O_2 SOM, which accounted for 9.3% of low O_2 observations (Figure 5). In this case, high antecedent precipitation inputs could have temporarily saturated the soil, thus stimulating soil respiration, while simultaneously blocking O_2 diffusion. This could have triggered a significant O_2 depletion that persisted even after soils dried back down. This scenario agrees with the preceding warm and above-average 2-week precipitation conditions typical for cluster 2. Our findings agree with those of Silver et al. (1999), who found that forest soil O_2 concentrations were negatively correlated with cumulative rainfall for up to 4 weeks preceding O_2 measurements. The impact of antecedent conditions on O_2 dynamics is also evidenced by the findings of Smyth et al. (2019), who found that high temporal resolution sensor data, especially gas-phase soil pore O_2 , revealed distinct lag periods between changes in soil conditions and subsequent biogeochemical activity.

It is important to note that clusters 4 and 5 were present only once throughout the entire sampling period (in June–July 2019) and occurred together in quick succession, which indicates these conditions were unusual for our site, at least within our three study years. The anomalous low O_2 conditions in summer (June–July 2019) were followed by a rapid increase in soil O_2 . Burgin and Groffman (2012) and Liptzin et al. (2011) documented rapid increases after low O_2 conditions in summer and attributed these events to dry soil macropores and plant senescence.

Another example of counterintuitive patterns is the occurrence of observed high O_2 levels during fall, winter, and spring months (October–April), when conditions were wet and cool (38.3% of high O_2 data points). Two distinct near-atmospheric O_2 events occurred (one in December–February 2017/2018, the other in November–April 2018/2019) under such conditions (clusters 2 and 3 from CV high O_2 SOM, Figure 4), and the latter event had a relatively long duration. This suggests that the high O_2 levels were not a result of a brief O_2 transition period, but instead reflect the absence of an O_2 depletion mechanism in response to increased soil moisture. Interestingly, these conditions mirror those that resulted in low O_2 levels (Table 4), indicating that decreased soil respiration rates characteristic of clusters 2 and 3 from the high O_2 SOM likely did not prevent O_2 depletion. The only difference between wet and cool conditions that resulted in low O_2 (Table 4) and clusters 2 and 3 from the high O_2 CV SOM (Table 3) were higher 2-week precipitation values associated with low O_2 during January–May, which further emphasizes the important role of antecedent conditions in O_2 depletion. If O_2 consumption rates did not exceed the rate of O_2 delivery, it is possible that these distinctive O_2 levels were a result of oxygenated water inputs originating from downward flow of oxygenated surface water, or oxygenated groundwater recharge (Lahiri & Davidson, 2019; Nelson, 2002). Regardless, our results show that very similar soil conditions can result in distinctive O_2 levels. A better understanding of the drivers of soil O_2 is therefore required to investigate such heterogeneity.

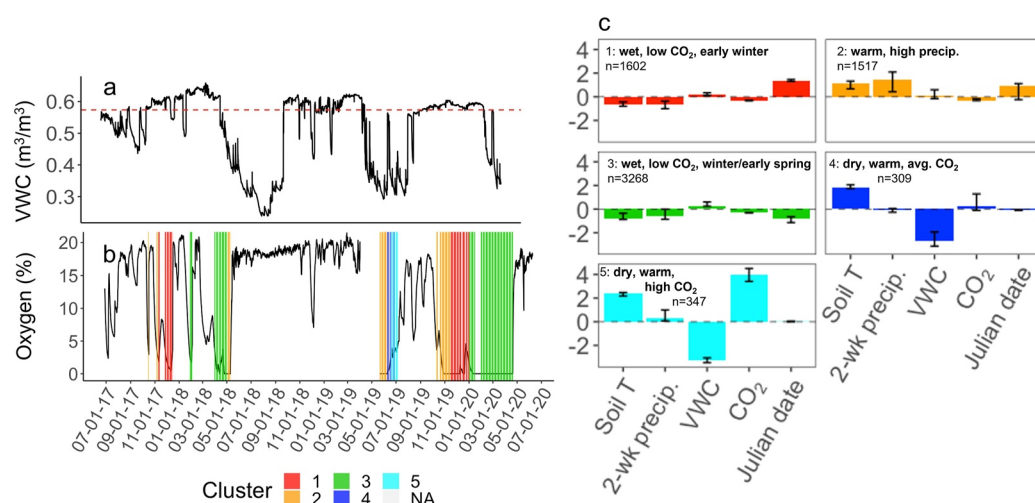


Figure 5. Low O_2 self-organizing map (SOM) results for the Champlain Valley (CV) site, including: (a) volumetric water content (VWC) time series with dashed line representing mean VWC for the CV low O_2 dataset, (b) O_2 time series highlighted with the five clusters identified by the low O_2 SOM, (c) bar plots displaying range normalized intra-cluster means of each input variable (n = number of observations per cluster). Clusters that are not shaded (represented by “NA”) correspond to O_2 values outside of the low O_2 range. 2-wk precip., two-week antecedent precipitation.

Our results also illustrate the utility of high frequency in-situ time series in capturing infrequent and unanticipated events, especially in cases when antecedent conditions may alter the O_2 response. Indeed, intermittent manual sampling campaigns could either miss these events entirely, mischaracterize the commonality of their occurrence, or have limited capacity to identify event drivers. As in-situ sensing networks become more commonplace in soil science research, we expect increased detection of these counterintuitive events. This will ultimately change how we understand the drivers of fluctuating O_2 conditions in the soil environment, and in particular, the role of antecedent conditions.

4.3. Site-Specific Controls on the Drivers of O_2 Regimes

While the key controls on soil O_2 were constant across sites, site-specific characteristics modulated the relative rates and impacts of ecosystem water inputs, water demand, and O_2 demand, which could have led to variable O_2 regimes across sites. Indeed, the constant anoxia (O_2 = 0–0.6%) and elevated VWC values (0.48–0.61 m^3/m^3) observed at the GM site were likely due to unique site-specific features. The combined effects of topography and groundwater hydrology dynamics provided a steady water supply that created consistently saturated soil conditions. This is further evidenced by wet and warm soil conditions unique to the GM site, indicating that soils did not dry out under increased ambient temperatures. This finding is consistent with those of Silver et al. (1999), who found that soil O_2 levels were sensitive to hydrologic inputs and were significantly correlated with a topographic gradient spanning ridge, slope, and valley locations. The consistently high water inputs at the GM site generated constant anoxia by preventing the re-aeration of soil pores, and/or displacing O_2 . As expected, we observed seasonal fluctuations in key O_2 controls (2-week antecedent ambient T, 1-week antecedent precipitation, VWC), but in contrast to the CV site, this resulted in steadily low O_2 concentrations. These findings highlight a disconnection between the controls on O_2 and O_2 dynamics. This suggests that a physical soil wetting process is the primary mechanism controlling O_2 dynamics at the GM site, and that the prevention of soil pore re-aeration, or O_2 diffusion prevails, thus creating a low O_2 environment, regardless of seasonal fluctuations in O_2 controls. Our results show that relationships between O_2 and soil moisture may be reliably predicted by a traditional bivariate nonlinear power function at some sites (e.g., our GM site) but a multivariate regression model may be required at sites with higher O_2 variability (e.g., our CV site). Our findings indicate that the chosen regression approach should consider the environment in which the data were collected, and our SOM analysis has identified key features to be monitored and considered in such a multivariate regression approach.

The topography, groundwater hydrology, and vegetation characteristics unique to the CV site resulted in seasonal VWC fluctuations. For example, low VWC values observed throughout the growing season at the CV site were

Table 4
Mean Value of Each Input Variable and O_2 Across Five Clusters Identified by the Champlain Valley Low O_2 Self-Organizing Map

Cluster	O_2 (%)	Soil temp. (°C)	VWC (m^3/m^3)	CO_2 (ppm)	2-week cumulative antecedent precip. (cm)	Conditions	Julian date ranges
1	1.1 ^a	2.8 ^a	0.59 ^a	611.0 ^a	22.6 ^a	Wet, low CO_2 , early winter	November–January
2	0.8 ^b	10.6 ^b	0.58 ^b	864.0 ^a	64.5 ^b	Warm, high precip.	October–November & May–June
3	0.5 ^c	2.9 ^a	0.60 ^c	470.4 ^a	27.1 ^c	Wet, low CO_2 , winter/early spring	January–May
4	1.4 ^d	15.3 ^c	0.42 ^d	4,247.5 ^b	33.3 ^d	Dry, warm, avg. CO_2	June
5	3.5 ^e	17.6 ^d	0.37 ^d	18,378.9 ^c	42.5 ^e	Dry, warm, high CO_2	June–July

Note. Within the same column, different letters represent significant differences between clusters ($p < 0.05$).

likely the result of depleted groundwater levels in combination with high plant water uptake by abundant sedges and nettles (water demand), and elevated soil respiration rates (O_2 demand). These conditions facilitated O_2 diffusion and re-aeration of soil pores, thus restoring soil O_2 to near-atmospheric levels. Under dry and warm conditions, antecedent precipitation plays an important role in O_2 depletion, as soils at the CV site can become too dry to displace O_2 or block O_2 diffusion. The significant impact of antecedent precipitation conditions on soil O_2 is also highlighted by Silver et al. (1999), who found soil O_2 levels at ridge locations to be significantly correlated with cumulative 4-week antecedent precipitation.

Low O_2 values ($n = 7,043$) occurred less frequently than high O_2 values ($n = 12,593$) at the CV site, which has important implications for nutrient cycling. This finding suggests that the process of O_2 depletion requires the convergence of a more specific suite of soil conditions than high O_2 levels do. Due to seasonal fluctuations in the controls on soil O_2 at our sites, low O_2 values also occurred less frequently (25% of low O_2 values) during the growing season (generally early May to early October in Vermont), compared to high O_2 values (61%). Nutrient cycling processes that require anaerobic soil conditions or anaerobic microsites, such as denitrification, sulfate reduction, or methanogenesis, will not proceed if the soil environment is well aerated (Sextstone et al., 1982). Furthermore, the growing season is a critically important time for nutrient sequestration and transformation within agricultural watersheds (Wang et al., 2014), as fertilizer is generally applied to agricultural fields in early spring. High soil O_2 levels during this critical period for N mobilization could reduce soil denitrification rates, which could have detrimental effects on nearby aquatic ecosystems.

4.4. Complications Associated With Predicting Soil O_2 Based Solely on Soil Moisture

We asked the question: Which combinations of variables lead to high versus low soil O_2 ? The results of our clustering analysis suggest that riparian soil O_2 dynamics are controlled by a network of seasonally variable, rate-dependent, and location-dependent parameters, and as such, the relationship between O_2 and soil moisture is more complex than represented by traditional ecological models. In contrast to traditional ecological thought, high soil moisture does not always result in low O_2 levels, and vice versa. Reliance on these more traditional reduction functions would have yielded contrasting results at our two study sites. If we had predicted soil O_2 levels at the CV site solely based on the commonly assumed negative correlation between moisture and O_2 (i.e., VWC values that typically result in low O_2 ; $VWC = 0.5\text{--}0.6\text{ m}^3/\text{m}^3$), 30.6% of O_2 values would have been incorrectly predicted as low, and 6.7% would have been incorrectly predicted as high ($VWC \leq 0.4\text{ m}^3/\text{m}^3$). In contrast, consistently high VWC observed at the GM site resulted in consistently low O_2 . Therefore, although we did not observe a significant negative correlation between O_2 and soil moisture (data not shown), our predictions based on soil moisture conditions alone would have been reasonably accurate at this site.

Our findings have important implications for nutrient cycling models that rely solely on soil moisture measurements to predict soil O_2 , and for empirical studies that make inferences about soil biogeochemical processes based on O_2 estimations (Rubol et al., 2012). Soil O_2 dynamics strongly modulate the rate and efficiency of microbially-mediated soil elemental (e.g., C, N, S) cycling through shifts in redox potential. Incorrect estimations of soil O_2 can therefore result in inaccurate predictions of critical N, C, etc., process rates. For example, much of the literature involving the measurement of soil O_2 and its relationship to soil moisture is within the context of climate-change driven shifts in soil moisture regimes and the subsequent effects on C storage and soil respiration (O'Connell et al., 2018; Santiago et al., 2005). These changes in C storage are modulated by confounding effects of seasonally variable soil moisture and temperature, as well as O_2 (Davidson et al., 1998). Critical soil biogeochemical processes not only impact watershed nutrient mobilization and downstream water quality, but also soil greenhouse gas production. It is therefore imperative to continue to improve our understanding of soil O_2 dynamics, as they are likely to increase in complexity as we face complications linked to a changing climate. Models that incorporate mass balance approaches are a potential alternative to those that rely on traditional reduction functions to predict soil O_2 . For example, Parolari et al. (2021) applied mass balance models to power spectra quantified from soil sensor timeseries and found accurate predictability of O_2 variability using temperature and soil moisture as inputs.

Our analysis uniquely incorporated multivariate data of high temporal resolution, which allowed us to investigate and provide new insight into the mechanisms controlling O_2 dynamics within our study sites. However, the limitations of our analysis are highlighted by our spatially constrained dataset, as we included observations from one landscape position within two different riparian soil sites of contrasting adjacent land use. Therefore,

Table 5

Inter-Cluster Means of Each Input Variable and O₂ Across Five Clusters Identified by the Green Mountains Self-Organizing Map

Cluster	O ₂ (%)	2-week TA (°C)	VWC (m ³ /m ³)	1-week cumulative antecedent precip. (cm)	Conditions	Julian date ranges
1	0.02 ^a	−5.4 ^a	0.5 ^a	6.2 ^a	Wet, cool, low precip.	October–December
2	0 ^b	17.1 ^b	0.5 ^b	15.5 ^b	Dry, warm, low precip.	May–September
3	0 ^b	17.3 ^b	0.5 ^a	44.2 ^c	Wet, warm, high precip.	May–October
4	0 ^b	−9.9 ^c	0.5 ^c	8.6 ^d	Drier, cool, low precip.	January–February
5	0 ^b	18.1 ^d	0.6 ^d	8.7 ^d	Wet, warm, low precip.	May–July

Note. Within the same column, different letters represent significant differences between clusters ($p < 0.05$).

our results cannot be directly scaled up to predict O₂ regimes across more expansive ecosystem scales. It is also important to consider how the limited spatial resolution of our dataset (within a given sensor location) impacts our ability to assess the suitability of soil conditions for biogeochemical processes that occur in deeper soil depths, especially soil denitrification (Groffman et al., 2009).

Furthermore, the inherent spatial heterogeneity of the environmental factors that control soil redox-sensitive processes results in hot spots and hot moments of biogeochemical activity, which are very difficult to predict. We also note that our sensor network was unable to detect O₂ within soil solution, or anaerobic microsites within soil aggregates, where hotspots of soil denitrification, for example, can occur (Parkin, 1987). These sources of spatial heterogeneity could result in rates of biogeochemical processes that are not reflective of the soil conditions measured by our instrumentation.

However, our results provide information about riparian soil O₂ dynamics that can be used for larger scale pattern analysis. As the factors that control soil O₂ were similar across the two sites, the seasonal variability we observed in the key O₂ controls may also apply to other riparian soil environments located in temperate climates. This seasonal O₂ framework could be an effective tool as a first pass prediction of whether O₂ conditions are conducive to aerobic or anaerobic soil processes. However, we also must consider that various site-specific characteristics

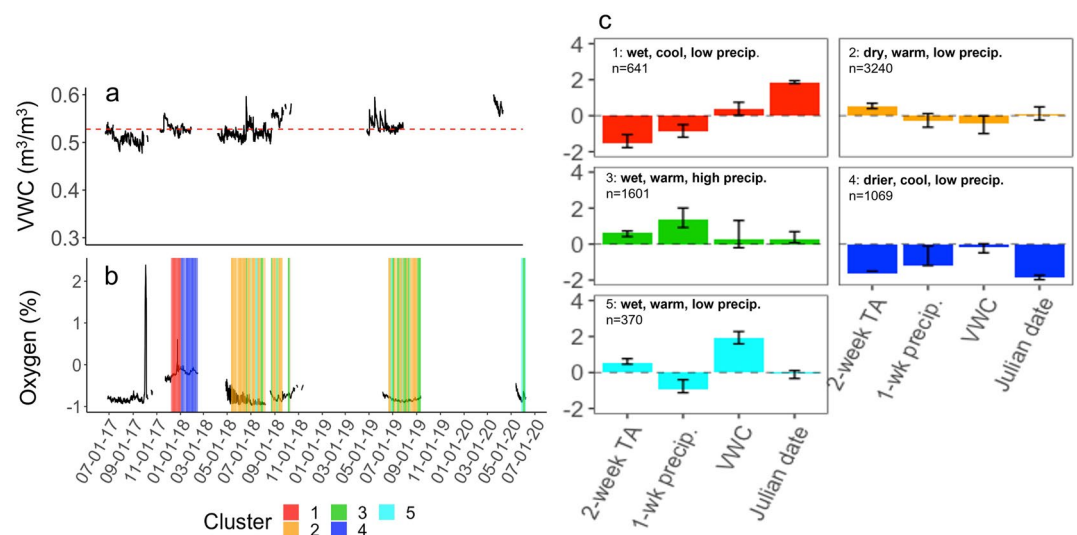


Figure 6. Self-organizing map (SOM) results for the Green Mountains (GM) site, including: (a) volumetric water content (VWC) time series with dashed line representing mean VWC for the GM dataset, (b) O₂ time series highlighted with the five clusters identified by the GM SOM. Negative O₂ values, which were set to zero when fed to the SOM, are included in this time series to show O₂ variability. (c) Bar plots displaying range normalized intra-cluster means of each input variable resulting from the GM SOM (n = number of observations per cluster). 1-wk precip., one-week antecedent precipitation.

will likely affect water inputs, water demand, and O_2 demand in ways that uniquely affect O_2 regimes. We posit that a fruitful next step would be to conduct comparable analyses that leverage more spatially expansive soil sensor networks that include other depths within the biologically active soil profile, as well as across variable climate, topographic, hydrologic, and geologic riparian soil environments (e.g., NEON, Critical Zone Observatories, LTER) to improve our understanding of the drivers of soil O_2 dynamics and our capacity to systematically model soil O_2 behavior and associated soil biogeochemical cycles.

It is also important to consider the implications of measuring gas-phase soil pore O_2 (measured in this study), compared to dissolved O_2 , especially in the context of biogeochemical process rates. Microbially mediated oxidation processes require dissolved-phase O_2 , whereas other abiotic O_2 oxidation processes can occur with gas-phase soil pore O_2 . Gas-phase soil pore O_2 may have a more direct impact on abiotic oxidation processes, like chemical weathering, photo-oxidation of organic compounds (e.g., lignin) electrochemical oxidation (i.e., the transfer electrons between minerals and change oxidation states), and abiotic oxidation of pollutants (e.g., heavy metals). However, the role of gas-phase soil pore O_2 as a predictor of microbially mediated processes is well-established. For example, Smyth et al. (2019) found gas-phase soil pore O_2 to be the best predictor for soil CH_4 fluxes, suggesting the type of soil O_2 measured in our study is linked to soil biogeochemical process rates. Rubol et al. (2013) also demonstrated the important role of soil pore O_2 in predicting rates of dissimilatory reduction of nitrate to ammonium. Thus, without labor-intensive methods needed to measure dissolved O_2 , our high frequency, continuous data provide information on the status of O_2 in soils, which has implications for both abiotic and microbially mediated processes.

5. Conclusions

We used a SOM approach to address key drivers of the spatial and temporal variability exhibited by riparian soil O_2 levels within the top 15 cm of the soil profile. Our results show that, in contrast to traditional ecological assumptions, O_2 cannot be accurately predicted based only on an inverse relationship between O_2 and soil moisture. Further, we found the SOM approach to be useful for identifying drivers of soil O_2 at intermediate soil moisture levels. However, in inundated soils, where O_2 is consistently depleted (i.e., our GM site), the reasoning that high soil moisture results in low soil O_2 holds true, eliminating the need for a multivariate approach to predict soil O_2 conditions. Our findings indicate that soil O_2 is controlled by a diverse set of seasonally variable parameters (antecedent precipitation, soil T, VWC, soil CO_2) and location-dependent conditions (topography and groundwater hydrology) that interact to result in a complex and nonlinear relationship between O_2 and soil moisture. Importantly, our results reveal that increases in soil moisture do not always trigger O_2 depletion, indicating that process-based ecosystem and denitrification models that rely on soil moisture alone to estimate soil O_2 may over-estimate denitrification, or other anaerobic process rates (e.g., iron reduction or methanogenesis). A more nuanced understanding of soil O_2 dynamics would therefore lead to improved predictions of temporal variability in redox-controlled nutrient cycling processes.

Data Availability Statement

The high frequency soil sensor data used to assess the relationship between soil moisture and soil oxygen (Lancellotti et al., 2021) are available in the Environmental Data Initiative repository under Creative Commons license via [10.6073/pasta/f62c630ff7a917eae8d320f8400ad2ec](https://doi.org/10.6073/pasta/f62c630ff7a917eae8d320f8400ad2ec) without a registration requirement. The R scripts required to perform the Self-Organizing Map (SOM) analysis and visualize the SOM results (Lancellotti et al., 2022) are available on the Zenodo repository under the Creative Commons license via [10.5281/zenodo.7271714](https://doi.org/10.5281/zenodo.7271714) at <https://doi.org/10.5281/zenodo.7271714> without a registration requirement. The software required to perform the analyses presented in this study (R and R studio) are open-source and are available for free at <https://posit.co/download/rstudio-desktop/>.

References

- Alvarez-Guerra, M., González-Piñuela, C., Andrés, A., Galán, B., & Viguri, J. R. (2008). Assessment of self-organizing map artificial neural networks for the classification of sediment quality. *Environment International*, 34(6), 782–790. <https://doi.org/10.1016/j.envint.2008.01.006>
- Birkinshaw, S. J., & Ewen, J. (2000). Nitrogen transformation component for SHETRAN catchment nitrate transport modelling. *Journal of Hydrology*, 230(1–2), 1–17. [https://doi.org/10.1016/S0022-1694\(00\)00174-8](https://doi.org/10.1016/S0022-1694(00)00174-8)

Acknowledgments

This work was supported by the National Science Foundation under Vermont EPSCoR Grant NSF OIA 1556770, as well as Grants EAR 2012123, and EAR 1724171. We thank Dustin Kincaid and Saul Blocher, as well as the Vermont EPSCoR summer interns, for help installing and maintaining the soil sensor network. We also thank the contributors of Vermont EPSCoR's Basin Resilience to Extreme Events (BREE) project for their expertise and assistance with this study. The authors do not have any financial or non-financial conflicts of interest associated with this publication to declare.

- Bouwman, A. F., Beusen, A. H. W., Griffioen, J., Van Groenigen, J. W., Hefting, M. M., Oenema, O., et al. (2013). Global trends and uncertainties in terrestrial denitrification and N_2O emissions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1621), 20130112. <https://doi.org/10.1098/rstb.2013.0112>
- Breard, G. T. (2017). Evaluating self-organizing map quality measures as convergence criteria (Master's thesis). University of Rhode Island. Retrieved from URI Digital Commons <https://digitalcommons.uri.edu/cgi/viewcontent.cgi?article=2013&context=theses>
- Burgin, A. J., & Groffman, P. M. (2012). Soil O_2 controls denitrification rates and N_2O yield in a riparian wetland. *Journal of Geophysical Research*, 117(G1), 1–10. <https://doi.org/10.1029/2011jg001799>
- Céréghino, R., & Park, Y. S. (2009). Review of the self-organizing map (SOM) approach in water resources: Commentary. *Environmental Modelling & Software*, 24(8), 945–947. <https://doi.org/10.1016/j.envsoft.2009.01.008>
- Chen, Q., Wang, Q., Han, X., Wan, S., & Li, L. (2010). Temporal and spatial variability and controls of soil respiration in a temperate steppe in northern China. *Global Biogeochemical Cycles*, 24(2), 1–11. <https://doi.org/10.1029/2009GB003538>
- Davidson, E. A., Belk, E., & Boone, R. D. (1998). Soil water content and temperature as independent or confounded factors controlling soil respiration in a temperate mixed hardwood forest. *Global Change Biology*, 4(2), 217–227. <https://doi.org/10.1046/j.1365-2486.1998.00128.x>
- Doran, J. W., Mielke, L. N., & Power, J. F. (1990). Microbial activity as regulated by soil water-filled pore space. In *Paper presented at 14th international congress of soil science, Kyoto, Japan*.
- Ewe, S. M., Sternberg, L. D. S., & Childers, D. L. (2007). Seasonal plant water uptake patterns in the saline southeast Everglades ecotone. *Oecologia*, 152(4), 607–616. <https://doi.org/10.1007/s00442-007-0699-x>
- Groffman, P. M., Butterbach-Bahl, K., Fulweiler, R. W., Gold, A. J., Morse, J. L., Stander, E. K., et al. (2009). Challenges to incorporating spatially and temporally explicit phenomena (hotspots and hot moments) in denitrification models. *Biogeochemistry*, 93(1), 49–77. <https://doi.org/10.1007/s10533-008-9277-5>
- Groffman, P. M., Tiedje, J. M., Robertson, G. P., & Christensen, S. (1988). Denitrification at different temporal and geographical scales: Proximal and distal controls. *Advances in Nitrogen Cycling in Agricultural Ecosystems*, 174–192.
- Hanson, R. S., & Hanson, T. E. (1996). Methanotrophic bacteria. *Microbiological Reviews*, 60(2), 439–471. <https://doi.org/10.1128/mr.60.2.439-471.1996>
- Hardie, M. A., Doyle, R. B., Cotching, W. E., Mattern, K., & Lisson, S. (2012). Influence of antecedent soil moisture on hydraulic conductivity in a series of texture-contrast soils. *Hydrological Processes*, 26(20), 3079–3091. <https://doi.org/10.1002/hyp.8325>
- Harrell, F. E. (2020). Hmisc: Harrell Miscellaneous. R package version 5.0-1. [Software]. Retrieved from <https://CRAN.R-project.org/package=Hmisc>
- Hefting, M., Clement, J. C., Dowrick, D., Cosandey, A. C., Bernal, S., Cimpian, C., et al. (2004). Water table elevation controls on soil nitrogen cycling in riparian wetlands along a European climatic gradient. *Biogeochemistry*, 67(1), 113–134. <https://doi.org/10.1023/b:biog.0000015320.69868.33>
- Hefting, M. M., Bobbink, R., & de Caluwe, H. (2003). Nitrous oxide emission and denitrification in chronically nitrate-loaded riparian buffer zones. *Journal of Environmental Quality*, 32(4), 1194–1203. <https://doi.org/10.2134/jeq2003.1194>
- Heinen, M. (2005). Simplified denitrification models: Overview and properties. *Geoderma*, 133(3–4), 444–463. <https://doi.org/10.1016/j.geoderma.2005.06.010>
- Hénault, C., & Germon, J. C. (2000). NEMIS, a predictive model of denitrification on the field scale. *European Journal of Soil Science*, 51(2), 257–270. <https://doi.org/10.1046/j.1365-2389.2000.00314.x>
- Jarecke, K. M., Loecke, T. D., & Burgin, A. J. (2016). Coupled soil oxygen and greenhouse gas dynamics under variable hydrology. *Soil Biology and Biochemistry*, 95, 164–172. <https://doi.org/10.1016/j.soilbio.2015.12.018>
- Jenks, G. F. (1967). The data model concept in statistical mapping. *International Yearbook of Cartography*, 7, 186–190.
- Kalteh, A. M., Hjorth, P., & Berndtsson, R. (2008). Review of the self-organizing map (SOM) approach in water resources: Analysis, modelling and application. *Environmental Modelling & Software*, 23(7), 835–845. <https://doi.org/10.1016/j.envsoft.2007.10.001>
- Kang, S., Doh, S., Lee, D., Jin, V. L., & Kimball, J. S. (2003). Topographic and climatic controls on soil respiration in six temperate mixed-hardwood forest slopes, Korea. *Global Change Biology*, 9(10), 1427–1437. <https://doi.org/10.1046/j.1365-2486.2003.00668.x>
- Khamis, N., Sin, T. C., & Hock, G. C. (2018). Segmentation of residential customer load profile in peninsular Malaysia using Jenks natural breaks. In *Paper presented at the IEEE 7th international conference on power and energy (PECon), Kuala Lumpur, Malaysia*.
- Kohonen, T. (2013). Essentials of the self-organizing map. *Neural Networks*, 37, 52–65. <https://doi.org/10.1016/j.neunet.2012.09.018>
- Kundu, P., Debsarkar, A., & Mukherjee, S. (2013). Artificial neural network modeling for biological removal of organic carbon and nitrogen from slaughterhouse wastewater in a sequencing batch reactor. *Advances in Artificial Neural Systems*, 2013, 1–15. <https://doi.org/10.1155/2013/268064>
- Lahiri, C., & Davidson, G. R. (2019). Heterogeneous oxygenation of wetland soils with increasing inundation: Redox potential, water depth, and preferential flow paths. *Hydrological Processes*, 34(6), 1350–1358. <https://doi.org/10.1002/hyp.13654>
- Lancellotti, B. V., Underwood, K. L., Perdrial, J. N., Adair, C. E., Schroth, A. W., & Roy, E. D. (2021). High frequency soil sensor data for SOM input - Complex drivers of riparian soil oxygen variability revealed using self-organizing maps ver 1 [Dataset]. Environmental Data Initiative. <https://doi.org/10.6073/pasta/f62c630ff7a917eae8d320f8400ad2ec>
- Lancellotti, B. V., Underwood, K. L., Perdrial, J. N., Adair, C. E., Schroth, A. W., & Roy, E. D. (2022). Complex drivers of riparian soil oxygen variability revealed using self-organizing maps: Relevant R codes for data analysis and visualization ver 1. [Data analysis programming]. Zenodo Open Repository. <https://zenodo.org/record/7271714>
- Landsman-Gerjoi, M., Perdrial, J. N., Lancellotti, B., Seybold, E., Schroth, A. W., Adair, C., & Wymore, A. (2020). Measuring the influence of environmental conditions on dissolved organic matter biodegradability and optical properties: A combined field and laboratory study. *Biogeochemistry*, 149(1), 37–52. <https://doi.org/10.1007/s10533-020-00664-9>
- Lavigne, M. B., Foster, R. J., & Goodine, G. (2004). Seasonal and annual changes in soil respiration in relation to soil temperature, water potential and trenching. *Tree Physiology*, 24(4), 415–424. <https://doi.org/10.1093/treephys/24.4.415>
- Liao, X., Tao, H., Gong, X., & Li, Y. (2019). Exploring the database of a soil environmental survey using a geo-self-organizing map: A pilot study. *Journal of Geographical Sciences*, 29(10), 1610–1624. <https://doi.org/10.1007/s11442-019-1644-8>
- Lin, G. F., & Chen, L. H. (2006). Identification of homogeneous regions for regional frequency analysis using the self-organizing map. *Journal of Hydrology*, 324(1–4), 1–9. <https://doi.org/10.1016/j.jhydrol.2005.09.009>
- Liptzin, D., Silver, W. L., & Detto, M. (2011). Temporal dynamics in soil oxygen and greenhouse gases in two humid tropical forests. *Ecosystems*, 14(2), 171–182. <https://doi.org/10.1007/s10021-010-9402-x>
- Lyu, Z., Shao, N., Akinyemi, T., & Whitman, W. B. (2018). Methanogenesis. *Current Biology*, 28(13), R727–R732. <https://doi.org/10.1016/j.cub.2018.05.021>
- Merdun, H. (2011). Self-organizing map artificial neural network application in multidimensional soil data analysis. *Neural Computing & Applications*, 20(8), 1295–1303. <https://doi.org/10.1007/s00521-010-0425-1>

- Moldrup, P., Olesen, T., Gamst, J., Schjønning, P., Yamaguchi, T., & Rolston, D. E. (2000). Predicting the gas diffusion coefficient in repacked soil water-induced linear reduction model. *Soil Science Society of America Journal*, 64(5), 1588–1594. <https://doi.org/10.2136/sssaj2000.6451588x>
- Moyano, F. E., Manzoni, S., & Chenu, C. (2013). Responses of soil heterotrophic respiration to moisture availability: An exploration of processes and models. *Soil Biology and Biochemistry*, 59, 72–85. <https://doi.org/10.1016/j.soilbio.2013.01.002>
- Neira, J., Ortiz, M., Morales, L., & Acevedo, E. (2015). Oxygen diffusion in soils: Understanding the factors and processes needed for modeling. *Chilean Journal of Agricultural Research*, 75, 35–44. <https://doi.org/10.4067/S0718-58392015000300005>
- Nelson, D. (2002). Natural variations in the composition of groundwater. Drinking Water Program. In *Paper presented at Groundwater Foundation annual meeting, Springfield, Oregon, USA*. Oregon Department of Human Services.
- Obach, M., Wagner, R., Werner, H., & Schmidt, H. H. (2001). Modelling population dynamics of aquatic insects with artificial neural networks. *Ecological Modelling*, 146(1–3), 207–217. [https://doi.org/10.1016/S0304-3800\(01\)00307-6](https://doi.org/10.1016/S0304-3800(01)00307-6)
- O'Connell, C. S., Ruan, L., & Silver, W. L. (2018). Drought drives rapid shifts in tropical rainforest soil biogeochemistry and greenhouse gas emissions. *Nature Communications*, 9(1), 1–9. <https://doi.org/10.1038/s41467-018-03352-3>
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., & Wagner, H. (2019). *vegan: Community ecology package*. [Software]. R package version 2.5-6.
- Orchard, V. A., & Cook, F. J. (1983). Relationship between soil respiration and soil moisture. *Soil Biology and Biochemistry*, 15(4), 447–453. [https://doi.org/10.1016/0038-0717\(83\)90010-X](https://doi.org/10.1016/0038-0717(83)90010-X)
- Parkin, T. B. (1987). Soil microsites as a source of denitrification variability. *Soil Science Society of America Journal*, 51(5), 1194–1199. <https://doi.org/10.2136/sssaj1987.03615995005100050019x>
- Parolari, A. J., Sizemore, J., & Katul, G. G. (2021). Multiscale legacy responses of soil gas concentrations to soil moisture and temperature fluctuations. *Journal of Geophysical Research: Biogeosciences*, 126(2), e2020JG005865. <https://doi.org/10.1029/2020jg005865>
- Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D., & R Core Team. (2014). *Nlme: Linear and nonlinear mixed effects models*. [Software]. R Package Version 3, 1–117.
- Ponnamperuma, F. N. (1972). The chemistry of submerged soils. *Advances in Agronomy*, 24, 29–96.
- Rabosky, D. L., Grudler, M., Anderson, C., Title, P., Shi, J. J., Brown, J. W., & Larson, J. G. (2014). BAMM tools: An R package for the analysis of evolutionary dynamics on phylogenetic trees. [Software]. *Methods in Ecology and Evolution*, 5(7), 701–707. <https://doi.org/10.1111/2041-210x.12199>
- Ridolfi, L., D'Odorico, P., Porporato, A., & Rodriguez-Iturbe, I. (2003). Stochastic soil moisture dynamics along a hillslope. *Journal of Hydrology*, 272(1–4), 264–275. [https://doi.org/10.1016/S0022-1694\(02\)00270-6](https://doi.org/10.1016/S0022-1694(02)00270-6)
- Rivera, D., Sandoval, M., & Godoy, A. (2015). Exploring soil databases: A self-organizing map approach. *Soil Use and Management*, 31(1), 121–131. <https://doi.org/10.1111/sum.12169>
- Rubol, S., Manzoni, S., Bellin, A., & Porporato, A. (2013). Modeling soil moisture and oxygen effects on soil biogeochemical cycles including dissimilatory nitrate reduction to ammonium (DNRA). *Advances in Water Resources*, 62, 106–124. <https://doi.org/10.1016/j.advwatres.2013.09.016>
- Rubol, S., Silver, W. L., & Bellin, A. (2012). Hydrologic control on redox and nitrogen dynamics in a peatland soil. *Science of the Total Environment*, 432, 37–46. <https://doi.org/10.1016/j.scitotenv.2012.05.073>
- Santiago, L. S., Schuur, E. A., & Silvera, K. (2005). Nutrient cycling and plant–soil feedbacks along a precipitation gradient in lowland Panama. *Journal of Tropical Ecology*, 21(4), 461–470. <https://doi.org/10.1017/S0266467405002464>
- Sevanto, S., Suni, T., Pumpanen, J., Grönholm, T., Kolari, P., Nikinmaa, E., & Vesala, T. (2006). Wintertime photosynthesis and water uptake in a boreal forest. *Tree Physiology*, 26(6), 749–757. <https://doi.org/10.1093/treephys/26.6.749>
- Sextstone, A. J., Revsbech, N. P., Parkin, T. B., & Tiedje, J. M. (1982). Direct measurement of oxygen profiles and denitrification rates in soil aggregates. *Soil Science Society of America Journal*, 49(3), 645–651. <https://doi.org/10.2136/sssaj1985.03615995004900030024x>
- Shanley, J. B., & Chalmers, A. (1999). The effect of frozen soil on snowmelt runoff at Sleepers River, Vermont. *Hydrological Processes*, 13(12–13), 1843–1857. [https://doi.org/10.1002/\(SICI\)1099-1085\(199909\)13:12<13%3C1843::AID-HYP879%3E3.0.CO;2-G](https://doi.org/10.1002/(SICI)1099-1085(199909)13:12<13%3C1843::AID-HYP879%3E3.0.CO;2-G)
- Silver, W. L., Lugo, A. E., & Keller, M. (1999). Soil oxygen availability and biogeochemistry along rainfall and topographic gradients in upland wet tropical forest soils. *Biogeochemistry*, 44(3), 301–328. <https://doi.org/10.1007/BF00996995>
- Skopp, J., Jawson, M. D., & Doran, J. W. (1990). Steady-state aerobic microbial activity as a function of soil water content. *Soil Science Society of America Journal*, 54(6), 1619–1625. <https://doi.org/10.2136/sssaj1990.03615995005400060018x>
- Smyth, A. R., Loeck, T. D., Franz, T. E., & Burgin, A. J. (2019). Using high-frequency soil oxygen sensors to predict greenhouse gas emissions from wetlands. *Soil Biology and Biochemistry*, 128, 182–192. <https://doi.org/10.1016/j.soilbio.2018.10.020>
- Sogbedji, J. M., Van Es, H. M., & Hutson, J. L. (2001). N fate and transport under variable cropping history and fertilizer rate on loamy sand and clay loam soils: I. Calibration of the LEACHMN model. *Plant and Soil*, 229(1), 57–70. <https://doi.org/10.1023/a:1004875116644>
- Soil Survey Staff, Natural Resources Conservation Service, & United States Department of Agriculture. (2019). *Web soil survey*. [Dataset]. Available online.
- Tissari, S., Nykänen, V., Lerssi, J., & Kolehmainen, M. (2007). Classification of soil groups using weights-of-evidence-method and RBFLN-neural nets. *Natural Resources Research*, 16(2), 159–169. <https://doi.org/10.1007/s11053-007-9040-y>
- Underwood, K. L., Rizzo, D. M., Dewoolkar, M. M., & Kline, M. (2021). Analysis of reach-scale sediment process domains in glacially-conditioned catchments using self-organizing maps. *Geomorphology*, 382, 107684. <https://doi.org/10.1016/j.geomorph.2021.107684>
- Vesanto, J., Himberg, J., Alhoniemi, E., Parhankangas, J., Team, S. O. M. T., & Oy, L. (2000). SOM toolbox for Matlab 5 (Vol. 57), No. 2.
- Vidon, P., Allan, C., Burns, D., Duval, T. P., Gurwick, N., Inamdar, S., et al. (2010). Hot spots and hot moments in riparian zones: Potential for improved water quality management. *Journal of the American Water Resources Association*, 46(2), 278–298. <https://doi.org/10.1111/j.1752-1688.2010.00420.x>
- Vidon, P., & Hill, A. R. (2004). Denitrification and patterns of electron donors and acceptors in eight riparian zones with contrasting hydrogeology. *Biogeochemistry*, 71(2), 259–283. <https://doi.org/10.1007/s10533-005-0684-6>
- Wang, Y., Li, Y., Liu, X., Liu, F., Li, Y., Song, L., et al. (2014). Relating land use patterns to stream nutrient levels in red soil agricultural catchments in subtropical central China. *Environmental Science and Pollution Research*, 21(17), 10481–10492. <https://doi.org/10.1007/s11356-014-2921-9>
- Wehrens, R., & Buydens, L. M. (2007). Self- and super-organizing maps in R: The Kohonen package. [Software]. *Journal of Statistical Software*, 21(5), 1–19. <https://doi.org/10.18637/jss.v021.i05>
- Wu, T., Chellemi, D. O., Graham, J. H., Martin, K. J., & Roskopf, E. N. (2008). Comparison of soil bacterial communities under diverse agricultural land management and crop production practices. *Microbial Ecology*, 55(2), 293–310. <https://doi.org/10.1007/s00248-007-9276-4>