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

Brittany Victoria Lancellotti (brittany.lancellotti@live.uvi.edu)
University of Vermont Rubenstein School of Environment and Natural Resources
https://orcid.org/0000-0003-4148-944X

Kristen Underwood
University of Vermont

Julia Perdrial
University of Vermont

Carol Adair
University of Vermont

Andrew Schroth
University of Vermont

Eric Roy
University of Vermont

Research Article

Keywords: Self-organizing map, clustering algorithm, riparian zone, soil O2, soil moisture, high-frequency data

Posted Date: December 15th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1108620/v1

Abstract

Oxygen (O_2) is a key regulator of soil reduction-oxidation processes and therefore modulates biogeochemical cycles. The difficulties associated with accurately characterizing soil O_2 variability have prompted the use of soil moisture as a proxy for soil O_2, based on the low solubility of O_2 in water. Due to seasonal shifts in soil O_2 depletion mechanisms, the use of soil moisture alone as a proxy measurement could result in inaccurate O_2 estimations. For example, soil O_2 may remain high during cool months when soil respiration rates are low. We analyzed high-frequency sensor data (e.g., soil moisture, temperature, CO_2, O_2) with a machine learning technique, the Self-Organizing Map, to pinpoint suites of soil conditions that are associated with contrasting O_2 regimes. At two low-lying riparian sites in contrasting land use and topographic settings of northern Vermont, we found that soil O_2 levels varied seasonally, and with soil moisture. For example, forty-seven percent of low O_2 levels were associated with cool and wet soil conditions, whereas 32% were associated with warm and dry conditions. Contrastingly, the majority (62%) of high O_2 conditions occurred under warm and dry conditions. High soil moisture levels did not always lead to low O_2, however, as 38% of high O_2 values occurred under cool and wet conditions. Our results highlight challenges associated with predicting soil O_2 solely based on soil moisture, as variable combinations of soil and site-specific hydrologic conditions can complicate the relationship between soil water content and O_2. This indicates that process-based ecosystem and denitrification models that rely solely on soil moisture to estimate O_2 availability will, in some cases, need to incorporate other site and climate-specific drivers to accurately predict soil O_2.

1 Introduction

Oxygen (O_2) 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 (Hefting et al., 2004; Silver, et al., 1999). For example, decreases in soil O_2 reduce redox potential, causing facultative and obligate anaerobic microorganisms to shift their energy-yielding respiration processes to utilize alternative electron acceptors. Oxygen availability therefore impacts the capacity of soils to transform nutrients, such as carbon (C) and nitrogen (N) via processes like heterotrophic aerobic respiration and denitrification, respectively (Liptzin et al., 2011). For example, the transformation of nitrate (NO_3^-), a soluble form of N that is limiting to primary productivity in freshwater and marine ecosystems, to gaseous forms of N (N_2O and N_2) via denitrification, is a major N removal pathway (Hefting, 2003). This process can reduce N loading to water bodies, but it will not proceed if O_2 is abundant. Oxygen availability therefore controls denitrification rates and efficiency (Groffman et al., 1988; Bouwman et al., 2013), playing a critical role in creating conducive environments for soil N removal. As such, soil O_2 regulates N_2O emissions and N mobility (Groffman et al., 1988), highlighting the importance of accurately characterizing shifts in soil O_2.

Our ability to predict soil O_2 concentrations across spatial and temporal gradients is limited, however. This is a result of the complex network of biotic and abiotic soil factors, as well as climatic conditions, that interact to modulate soil O_2 dynamics and create widespread spatial and temporal soil O_2 variability (Silver et al., 1999). Soil O_2 levels are regulated by the diffusion of O_2 into and displacement of O_2 out of soil pores by water (physical processes), and the consumption of O_2 via soil respiration (a biological process, i.e., aerobic microbial, plant root, and faunal respiration) (Moyano et al., 2013; Neira et al., 2015; Ponnamperuma, 1972). Because O_2 has a very low solubility in water (Moldrup et al., 2000), the presence of water inhibits O_2 diffusion from the atmosphere to soil pores (Skopp et al., 1985). Thus, the combined effects of inhibited O_2 diffusion and displacement, and soil respiration typically result in O_2 depletion if reaeration of soil pores is prevented (Neira et al., 2015; Ponnamperuma, 1972). Furthermore, variability in soil O_2 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 O_2 data is costly, as it requires soil probes and data logging capabilities.

The challenges associated with measuring gaseous O_2 in soils have led to the use of soil moisture as a proxy measurement for O_2 under the assumption that soil moisture is inversely proportional to O_2 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 O_2 and soil moisture to predict O_2 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 O_2 measurements have been found to exhibit high sensitivity to formulations that represent soil moisture (Hénault & Germon, 2000), which highlights that the relationship between these variables must be validated and defined using empirical data.

Indeed, due to seasonal shifts in the mechanisms that control soil O_2 depletion (Silver et al., 1999), the use of soil moisture as a proxy measurement for O_2 could result in inaccurate O_2 estimations. For example, soil water inputs (precipitation or groundwater) vary seasonally 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, O_2 depletion by plant and microbial respiration is primarily controlled by soil temperature and soil water content and thus exhibits seasonal fluctuations (Kang et al., 2003; Lavigne et al., 2004; Chen et al., 2010).

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 O_2 dynamics. Due to their unique position on the landscape, riparian soils experience frequent hydrologic changes that alter soil moisture content, which can modify O_2 availability.

Changes in soil moisture are triggered by hydrologic fluctuations, and the magnitude of these shifts depend 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 O_2 drivers suggest that the response of soil O_2 to soil moisture fluctuations is a result of multi-variate interactions that are highly dependent on site-specific soil conditions, seasonal fluctuations in environmental conditions, and ecosystem water and O_2 demands.

Recent advancements in sensor technologies facilitate the simultaneous collection of multiple soil parameters at high temporal resolution, including O_2 and its relevant controls (e.g., soil moisture, soil temperature, redox, CO_2, precipitation). This enables us to comprehensively assess potential drivers of soil moisture variability and the related O_2 response. 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 outliers, non-normally distributed, non-continuous data, and multicollinearity (Kundu et al., 2013; Merdun, 2011).

The SOM maps multivariate data to a two-dimensional map/lattice, where similar data points are situated in close 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 (Obach et al., 2001; Wu et al., 2008; Xiaoyong et al., 2019), 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.

We applied SOMs to test our overarching hypothesis that shifts in riparian soil O_2 levels are predictively driven by combinations of key environmental controls, such as ecosystem water delivery, and water and O_2 demand. We hypothesize that, while these controls are ubiquitous, the way they interact to impact soil O_2 will be modulated by site-specific characteristics and seasonal variability in their relative importance.

To test our hypotheses, we used the SOM to identify key drivers of variability in riparian soil O_2 dynamics and combinations of conditions that lead to low and high O_2 levels. We clustered high frequency soil and meteorological data collected over three 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 test our hypotheses 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 1) 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 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 O_2 concentrations ranged from anoxic to near atmospheric. One transect is situated within the Champlain Valley 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 to 106 m above sea level. Vegetation at this site includes American beech (*Fagus grandifolia)* trees, and 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) (Ross 2019; Soil Survey Staff, 2019). The 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) (Ross, 2019; Soil Survey Staff, 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 levels at the CV sampling location generally decrease during the growing season (early May - early October; average = 0.82 m below the soil surface) and increase during cooler months (average =

0.38 m below the soil surface). Contrastingly, average groundwater levels at the GM sampling location are significantly higher ($p<0.001$; average = 0.05 m below the soil surface) and remain elevated throughout the year. The combination of distinctive elevation, atmospheric characteristics (Table 1), and groundwater characteristics between sites could alter water inputs, water demand, and O_2 demand in ways that result in contrasting O_2 dynamics.

Table 1
Meteorological conditions observed throughout the sampling period at the Champlain Valley (CV) and Green Mountains (GM) site, separated by meteorological season. The data were collected from meteorological stations installed at each site. PAR stands for photosynthetically active radiation

Season	Months	Site	Cumulative precip. (cm)	Mean Ambient T (°C)	Mean PAR (uE)	Mean Solar radiation (W/m^2)	Mean Relative humidity (%)
Winter	Dec. – Feb.	CV	15.06	-5.93	265.53	162.91	80.50
		GM	18.34	-6.48	46.89	22.64	85.44
Spring	Mar. – May	CV	23.52	5.23	326.26	256.78	75.12
		GM	27.86	4.00	196.72	100.30	75.78
Summer	Jun. – Aug.	CV	25.65	19.66	345.58	305.67	79.86
		GM	31.82	18.02	241.16	137.78	84.25
Fall	Sept. – Nov.	CV	33.55	7.87	276.45	201.72	83.80
		GM	38.16	6.63	71.47	40.50	88.19

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 assessed by measuring ambient temperature (TA), and O_2 demand was measured via soil carbon dioxide (CO_2). Soil volumetric water content (VWC), temperature (T), and electrical conductivity (EC) were monitored at 15-minute intervals using 5TE sensors (Meter Group, Pullman, WA). Soil CO_2 was measured using GMT221 sensors (Vaisala, Helsinki, Finland), and soil O_2 was monitored using a Soil Response O_2 sensor (Apogee instruments, Logan, UT) (15-minute intervals). Each site (CV and GM) was equipped with a meteorological station that measured TA, precipitation, photosynthetically active radiation (PAR), solar radiation, relative humidity, and dew point at 5-minute intervals. Data included in this study were collected from July 2017 to June 2020.

Data analysis: Approach

To identify suites of soil conditions (clusters) associated with various soil O_2 regimes, we employed a Kohonen unsupervised Self-Organizing Map (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, O_2. 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 O_2 variability within our sites, and to examine meaningful ranges of O_2 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 an 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 close proximity. Finally, we compared O_2 values across clusters, *post hoc*, to better understand drivers of specific O_2 ranges.

2.3 Data analysis: Overview of Kohonen Unsupervised Self-Organizing Map

A detailed description of the SOM algorithm can be found in Kohonen (2013) and Underwood and others (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"), are updated to more closely resemble the input vector. 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).

2.4 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. We first replaced missing observations (except for precipitation) with the overall median value for each independent variable using the *impute* function in the R package Hmisc (Harrell et al., 2020). We then reduced the volume of data by downscaling our 15-minute 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 hours to 2 weeks prior to each observation to investigate the effects of antecedent soil and meteorological conditions on soil O_2 dynamics. 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 best suited to our particular data set (Underwood, et al., 2021).

We selected a suite of independent variables to include in our final SOMs by first running provisional SOMs with all available independent variables. We observed the resulting component planes and plotted the distribution of each variable, separated by cluster, using box and whisker plots. These component planes (Figure s-1) and box and whisker plots were visually inspected to exclude variables from the input dataset, such as wind speed, relative humidity, and EC, that lacked distinct variability across clusters.

2.5 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 topographic error and quantization error (Alvarez-guerra et al., 2008). The response variable, O_2, was not presented as an input to the SOM. However, we did examine O_2 data during data preprocessing to subdivide the associated multivariate timeseries observations into sets associated with distinct ranges of O_2 values, using the Jenks Natural Breaks optimization method via the BAMMtools package in R (Rabosky et al., 2014). We ran separate SOMs for independent multivariate data observations associated with high and low O_2 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.

2.6 Data analysis: SOM computation and model optimization

After selecting a suite of independent variables to include in each dataset (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, as an unsuitable lattice configuration can distort the distribution of the input variables across the map. We followed Vesanto's rule (Vesanto et al., 2000) to pinpoint the optimal number of lattice nodes. For each dataset, we 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 (Cereghino & Park, 2009). A hexagonal lattice arrangement was used. The lattice for the CV high O_2, CV low O_2, and GM SOM contained 17 rows and 34 columns, 13 rows and 33 columns, and 15 rows and 27 columns, respectively. The CV high O_2, CV low O_2, and GM SOM lattice configuration contained 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 quantization error (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 CV high O_2, CV low O_2, and GM SOM were 30891.5, 31395.1, and 32232.6, respectively. Quantization error values for the CV high O_2, CV low O_2, and GM SOM were 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 identify suites of variables and their value ranges that constituted different classes of environmental conditions associated with various O_2 levels. Kruskal-Wallis tests and *post hoc* Dunn's tests (with a Bonferroni adjustment) were completed to compare inter-cluster medians of original (i.e., not range-normalized) values of each input variable across clusters. The cluster assignments for each observation were then plotted onto an O_2 time series figure (original values) to display temporal fluctuations in classes of environmental conditions.

3 Results

3.1 Champlain Valley site

For the CV dataset, we performed separate SOMs for observations associated with high (12.9%-21.5%, n = 12,593) and low (0-4.3%, n = 7043) O_2. The SOMs (i) identified sets of input variables that accounted for variability within the dataset and (ii) classified suites of conditions that led to different O_2 regimes. After visually inspecting component planes of all available input variables, key input variables selected for the CV high O_2 SOM (mean O_2 = 18%) were soil T (mean = 10.3°C), 2-week cumulative antecedent

precipitation (hereafter referred to as “2-week precipitation”) (2.74 cm), VWC (0.46 m^3/m^3), CO_2 (3215.2 ppm), and Julian date (192) (Figure s-1). Four clusters maximized the nonparametric F statistic and minimized quantization error.

3.2 Champlain Valley site: High O_2 SOM

The High O_2 SOM identified four distinct clusters of multivariate time series observations (Figure 2c). The clustering analysis grouped CV high O_2 data points into two main categories: warm and dry (i.e., above average soil T and below average VWC), and cool and wet (below average soil T and above average VWC). Two-week precipitation and CO_2 values drove sub-clustering within the warm and dry category, while Julian date, along with two-week precipitation, drove sub-clustering within the cool and wet observations. Oxygen at the CV site ranged from 0 to 21.5%. 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 2b), and further research is required to identify what prevented O_2 depletion under cool and wet soil conditions during this period.

Soil conditions within clusters 1 and 4, which made up 62% of the CV high O_2 data points, were generally warm and dry (Table 2). Cluster 1, which includes 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 (Table 2). Cool and wet 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 (Oct.-Jan., and Jan.-Apr., respectively) and 2-week precipitation (39.9 cm and 19.2 cm, respectively) (Table 2).

Table 2
Mean value of each input variable and O2 across four clusters identified by the Champlain Valley high O2 SOM. Within the same column, different letters represent significant differences between clusters (p<0.001)

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	17.8a	15.2a	0.4a	3570.4a	30.2a	Warm/dry	Apr. – Oct.
2	18.2b	3.2b	0.6b	493.5b	39.9b	Cool/wet	Oct. – Jan.
3	18.6c	1.3c	0.6c	525.7b	19.2c	Cool/wet	Jan. – Apr.
4	16.8d	17.7d	0.4a	13077.4c	15.6d	Warm/dry	Jul. – Sept.

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 throughout fall 2017 and winter 2018. Low O_2 values were observed consistently from fall (September, October November) 2019 through spring (March, April, May) 2020. There were consistent periods of low O_2 during the spring snowmelt period (April-May) of 2018, 2019, and 2020 (Figure 3a).

Based on a visual inspection of the SOM component planes (Figure s-2), we included the same suite of input variables in the CV low O_2 SOM as the CV high O_2 SOM (soil T, 2-week precipitation, VWC, CO_2, and Julian date). Average values for these variables were 5.8°C, 3.52 cm, 0.57 m^3/m^3, 1635.2 ppm, and 175.2, respectively. Median values of each input variable were significantly different between the CV high and low O_2 datasets ($p < 0.001$). The SOM identified 5 distinct clusters for the CV low O_2 SOM: like the CV high O_2 dataset, observations fell into warm and dry, or cool and wet categories (Table 3). The majority of low O_2 data points (69%) could be categorized as cool and wet (clusters 1 and 3), in contrast to results for the high O_2 SOM. Data points within clusters 1 and 3 had below average 2-week precipitation and CO_2, and they differed in Julian date ranges (Figure 3c). Clusters 4 and 5 included 9.3% of data points, which fell into a warm and dry category, with above average CO_2. Two-week precipitation differed between these two clusters (Figure 3c), as cluster 4 had average, and cluster 5 had above average 2-week precipitation. Unique to the CV low O_2 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 can be described as warmer than average, with average soil moisture, despite above average antecedent precipitation and below average CO_2 (Table 3).

Table 3
Mean value of each input variable and O2 across five clusters identified by the Champlain Valley low O2 SOM. Within the same column, different letters represent significant differences between clusters ($p<0.001$)

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	1.1a	2.8a	0.59a	611.0a	22.6a	Cool/wet	Nov. – Jan.
2	0.8b	10.6b	0.58b	864.0b	64.5b	Warm/high precip.	Oct. – Nov. & May – Jun.
3	0.5c	2.9c	0.60c	470.4c	27.1c	Cool/wet	Jan. – May
4	1.4d	15.3d	0.42d	4247.5d	33.3d	Warm/dry	Jun.
5	3.5e	17.6d	0.37d	18378.9e	42.5e	Warm/dry	Jun. – Jul.

3.4 Green Mountains site SOM

We included all values from the GM dataset in one SOM, because O_2 values at the GM site ranged from 0 to 0.6% (n=6921). Based on a visual inspection of the component planes (Figure s-3), a unique set of input variables was chosen for the GM model: cumulative 2-week antecedent ambient temperature (hereafter referred to as 2-week TA), VWC, 1-week cumulative antecedent precipitation (hereafter referred to as 1-week precipitation), and Julian date. Average values for these variables were 11.0°C, 0.53 m^3/m^3, 1.99 cm, and 186.6, respectively. The low O_2 SOM grouped the observations into 5 different clusters. Similar to results of the CV low O_2 SOM, data points within the GM dataset can be described as warm and dry, cool and wet, or warm and wet (Table 4).

Table 4
Inter-cluster means of each input variable and O2 across five clusters identified by the Green Mountains SOM. Within the same column, different letters represent significant differences between clusters ($p<0.001$)

Cluster	O_2 (%)	2-week TA (°C)	VWC (m^3/m^3)	1-week cumulative antecedent precip. (cm)	Conditions	Julian date ranges
1	0.02a	-5.4a	0.5a	6.2a	Cool/wet	Oct. – Dec.
2	0b	17.1b	0.5b	15.5b	Warm/dry	May – Sept.
3	0b	17.3c	0.5c	44.2c	Warm/wet	May – Oct.
4	0b	-9.9d	0.5d	8.6d	Cool/wet	Jan. – Feb.
5	0b	18.1c	0.6e	8.7d	Warm/wet	May – Jul.

Encompassing 25% of data points, clusters 1 and 4 can be described as cool and wet, with differing Julian date ranges and VWC (Figure 4c). Clusters 3 and 5 (warm and wet) included 28% of data points, which can be summarized as warm and wet conditions. However, they differed in terms of 1-week precipitation and Julian date, which implies increased soil moisture was caused by differing mechanisms (Figure 4b and c). Cluster 2 encompassed 46.8% of the data points, which can be described as warm and dry, with below average 1-week antecedent precipitation and average Julian date.

Oxygen values included in the GM SOM remained consistently at or below zero percent throughout the entire sampling period, with the exception of an O_2 event in late December 2017 that reached 0.6% (Figure 4c). Clusters 1 and 4 occurred throughout winter 2017/2018 and were associated with cool and wet 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. O_2 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 O_2

Using a unique combination of high frequency, multiparameter soil sensor data collected over multiple years, coupled with machine learning methods, we were able to pinpoint key drivers of riparian soil O_2 variability and identify the complex combinations of variables that control soil O_2 levels. Importantly, and

in contrast to traditional ecological assumptions, low O_2 levels did not correspond solely with increasing soil moisture. Indeed, high O_2 levels persisted in both high and low moisture conditions. As hypothesized, low O_2 conditions instead depended on temporally varying combinations of water inputs, water demand, and O_2 demand. As a result, the output of our unsupervised clustering analysis from both field sites (i.e., all three SOMs) could be placed into general categories of “warm and dry”, “cool and wet”, and “warm and wet”. Within those distinct categories, antecedent precipitation, soil CO_2, and Julian date (i.e., season) drove further clustering.

At the CV site, the majority (69%) of low O_2 values were associated with cool and wet soil conditions (i.e., below average soil T, above average VWC) and occurred from November to May (Figure 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 precluded the reaeration of soil pores. All low O_2 values at the CV site that were observed under cool and wet conditions were associated with below average subsurface CO_2 (used here as indicator of aerobic soil respiration), 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 and others (1998) and Moyano and others (2013) who observed decreased soil respiration rates under cool and wet 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., 1985). 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 cool and wet 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 cool and wet conditions.

Contrastingly, most (62%) high O_2 values from the CV site were associated with warm and dry conditions that occurred during warmer months (Apr.-Oct., Figure 2). This is consistent with our hypothesis that warm and dry soil conditions would inhibit soil O_2 depletion by allowing O_2 to readily exchange with the atmosphere via increased air-filled pore 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, which contradicts our hypothesis. Furthermore, these findings contradict previous studies by Doran and others (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 warm and dry 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, warm and dry 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 under warm and dry conditions, and most low O_2 levels occurred under cool and wet soil conditions, we also intermittently observed the opposite behavior. These exceptions hold key insights into the important role of antecedent conditions and lag effects in determining soil O_2 regimes. For example, low O_2 events occurred in June and July under warm and dry (below average VWC) soil 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 3). 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 is in agreement with the preceding warm and above average 2-week precipitation conditions typical for cluster 2. Similar to our findings, Silver and others (1999) found that forest soil O_2 concentrations were negatively correlated with cumulative rainfall for up to four weeks preceding O_2 measurements. However, 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 that these conditions were unusual for our site, at least within our two study years.

Another example of counterintuitive patterns are occurrences of observed high O_2 levels during fall, winter, and spring months (Oct.-Apr.), when conditions were cool and wet (38.3% of high O_2 data points). Two distinct near-atmospheric O_2 events occurred (one in Dec.-Feb. 2017/2018, the other in Nov.-Apr. 2018/2019) under such conditions (clusters 2 and 3 from CV high O_2 SOM, Figure 3), 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 3), 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 cool and wet conditions that resulted in low O_2 (Table 3) and clusters 2 and 3 from the high O_2 CV SOM (Table 2) 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. It is also possible that the distinctive O_2 levels resulted from oxygenated subsurface water inputs originating from oxygenated groundwater recharge (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.

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

We also hypothesized that, while the key controls on soil O_2 are constant across sites, site-specific characteristics modulate the relative rates and impacts of ecosystem water inputs, water demand, and O_2 demand, which could lead 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 warm and wet 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 and others (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 major 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 reaeration or O_2 diffusion prevails, thus creating a low O_2 environment, regardless of seasonal fluctuations in O_2 controls.

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 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 reaeration of soil pores, thus restoring soil O_2 to near-atmospheric levels. Under warm and dry 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 and others (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 = 7043) occurred less frequently than high O_2 values (n = 12593) 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–early Oct. in Vermont), compared to high O_2 values (61%). Nutrient cycling processes that require anaerobic soil conditions or anaerobic microsites, such as denitrification, will not proceed if the soil environment is well aerated (Sexstone 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

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= 0.5-0.6 m^3/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 m^3/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., 2015; 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.

Our analysis uniquely incorporated multivariate data of high temporal resolution, which allowed us to investigate and provide new insight about 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, our results cannot be directly scaled up to predict O_2 regimes across more expansive ecosystem scales. However, our results provide information about riparian soil O_2 dynamics that can be used for larger scale pattern analysis. As the factors that control soil O_2 were similar across the two sites, the seasonal variability we observed in the key O_2 controls may also apply to other riparian soil environments located in temperate climates. This seasonal O_2 framework could be an effective tool as a first pass prediction of whether O_2 conditions are conducive to aerobic or anaerobic soil processes. However, we also must consider that various site-specific characteristics 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 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.

5 Conclusions

We used a self-organizing map approach to address the widespread spatial and temporal variability exhibited by riparian soil O_2 levels. Our results show that, in contrast to traditional ecological assumptions, O_2 cannot reliably be predicted solely based on an inverse relationship between O_2 and soil moisture. Soil O_2 is instead 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 availability 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.

Declarations

Funding

This work was supported by the National Science Foundation under Vermont EPSCoR Grant No. NSF OIA 1556770, as well as grants EAR 2012123, and EAR 1724171.

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

Author contributions

Conceptualization, and review and editing of the manuscript: all authors. Writing-original draft preparation: Brittany Lancellotti. Methodology: Kristen Underwood and Brittany Lancellotti.

Data availability

The datasets generated and analyzed during the current study are available in the Environmental Data Initiative repository (https://portal.edirepository.org/nis/mapbrowse?scope=edi&identifier=1034&revision=1)

Acknowledgements

This work was supported by the National Science Foundation under Vermont EPSCoR Grant No. 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.

References

1. Alvarez-Guerra M, González-Piñuela C, Andrés A, Galán B, Viguri JR (2008) Assessment of self-organizing map artificial neural networks for the classification of sediment quality. Environ Int 34(6):782–790
2. Birkinshaw SJ, Ewen J (2000) Nitrogen transformation component for SHETRAN catchment nitrate transport modelling. J Hydrol 230(1–2):1–17
3. Bouwman AF, Beusen AHW, Griffioen J, Van Groenigen JW, Hefting MM, Oenema O, Stehfest E (2013) Global trends and uncertainties in terrestrial denitrification and N2O emissions. Philosophical Transactions of the Royal Society B: Biological Sciences 368(1621):20130112
4. Céréghino R, Park YS (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

5. 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. Glob Biogeochem Cycles 24(2). https://doi.org/10.1029/2009GB003538
6. Davidson EA, Belk E, Boone RD (1998) Soil water content and temperature as independent or confounded factors controlling soil respiration in a temperate mixed hardwood forest. Glob Change Biol 4(2):217–227
7. Doran JW, Mielke LN, Power JF (1990) Microbial activity as regulated by soil water-filled pore space. In Transactions 14th International Congress of Soil Science, Kyoto, Japan, August 1990 3:94-99
8. Ewe SM, da Sternberg SL, Childers L 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
9. Groffman PM, Tiedje JM, Robertson GP, Christensen S (1988) Denitrification at different temporal and geographical scales: Proximal and distal controls. Advances in nitrogen cycling in agricultural ecosystems 174-192
10. Hardie MA, Doyle RB, Cotching WE, Mattern K, Lisson S (2012) Influence of antecedent soil moisture on hydraulic conductivity in a series of texture-contrast soils. Hydrol Process 26(20):3079–3091. https://doi.org/10.1002/hyp.8325
11. Hefting MM, Bobbink R, de Caluwe H (2003) Nitrous oxide emission and denitrification in chronically nitrate-loaded riparian buffer zones. J Environ Qual 32(4):1194–1203
12. Heinen M (2006) Simplified denitrification models: overview and properties. Geoderma 133(3–4):444–463. https://doi.org/10.1016/j.geoderma.2005.06.010
13. Hénault C, Germon JC (2000) NEMIS, a predictive model of denitrification on the field scale. Eur J Soil Sci 51(2):257–270
14. Kalteh AM, 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
15. Kang S, Doh S, Lee D, Jin VL, Kimball JS (2003) Topographic and climatic controls on soil respiration in six temperate mixed-hardwood forest slopes, Korea. Glob Change Biol 9(10):1427–1437
16. Kohonen T (2013) Essentials of the self-organizing map. Neural Netw 37:52–65
17. 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
18. Landsman-Gerjoi M, Perdrial JN, Lancellotti B, Seybold E, Schroth AW, 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
19. Lavigne MB, Foster RJ, Goodine G (2004) Seasonal and annual changes in soil respiration in relation to soil temperature, water potential and trenching. Tree Physiol 24(4):415–424

20. Lin GF, Chen LH (2006) Identification of homogeneous regions for regional frequency analysis using the self-organizing map. J Hydrol 324(1–4):1–9. https://doi.org/10.1016/j.jhydrol.2005.09.009
21. Liptzin D, Silver WL, 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
22. Merdun H (2011) Self-organizing map artificial neural network application in multidimensional soil data analysis. Neural Comput Appl 20(8):1295–1303. https://doi.org/10.1007/s00521-010-0425-1
23. Moldrup P, Olesen T, Gamst J, Schjønning P, Yamaguchi T, Rolston DE (2000) Predicting the gas diffusion coefficient in repacked soil water-induced linear reduction model. Soil Sci Soc Am J 64(5):1588–1594
24. Moyano FE, Manzoni S, Chenu C (2013) Responses of soil heterotrophic respiration to moisture availability: An exploration of processes and models. Soil Biol Biochem 59:72–85. https://doi.org/10.1016/j.soilbio.2013.01.002
25. 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
26. Nelson D (2002) Natural variations in the composition of groundwater. Drinking Water Program, Oregon Department of Human Services
27. O'Connell CS, Ruan L, Silver WL (2018) Drought drives rapid shifts in tropical rainforest soil biogeochemistry and greenhouse gas emissions. Nat Commun 9(1):1–9. https://doi.org/10.1038/s41467-018-03352-3
28. Obach M, Wagner R, Werner H, Schmidt HH (2001) Modelling population dynamics of aquatic insects with artificial neural networks. Ecol Model 146(1–3):207–217
29. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Wagner H (2020) vegan: Community Ecology Package. R package version 2.5-6. 2019
30. Orchard VA, Cook FJ (1983) Relationship between soil respiration and soil moisture. Soil Biol Biochem 15(4):447–453
31. Ponnamperuma FN (1972) The chemistry of submerged soils. In Advances in agronomy 24:29–96
32. Rabosky DL, Grundler M, Anderson C, Title P, Shi JJ, Brown JW, Larson JG (2014) BAMM tools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. Methods Ecol Evol 5(7):701–707
33. Ridolfi L, D'Odorico P, Porporato A, Rodriguez-Iturbe I (2003) Stochastic soil moisture dynamics along a hillslope. J Hydrol 272(1–4):264–275
34. Rivera D, Sandoval M, Godoy A (2015) Exploring soil databases: a self-organizing map approach. Soil Use Manag 31(1):121–131. https://doi.org/10.1111/sum.12169
35. Ross DS (2019) Soil classification for BREE sites. Plant and soil science. University of Vermont, Vermont
36. 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). Adv Water

Resour 62:106–124. https://doi.org/10.1016/j.advwatres.2013.09.016

37. Rubol S, Silver WL, Bellin A (2012) Hydrologic control on redox and nitrogen dynamics in a peatland soil. Sci Total Environ 432:37–46. https://doi.org/10.1016/j.scitotenv.2012.05.073
38. Santiago LS, Schuur EA, Silvera K (2005) Nutrient cycling and plant–soil feedbacks along a precipitation gradient in lowland Panama. J Trop Ecol 21(4):461–470
39. https://doi.org/10.1017/S0266467405002464
40. 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 Physiol 26(6):749–757
41. Sexstone AJ, Revsbech NP, Parkin TB, Tiedje JM (1985) Direct measurement of oxygen profiles and denitrification rates in soil aggregates. Soil Sci Soc Am J 49:645–651
42. Shanley JB, Chalmers A (1999) The effect of frozen soil on snowmelt runoff at Sleepers River, Vermont. Hydrol Process 13(12–13):1843–1857
43. Silver WL, Lugo AE, Keller M (1999) Soil oxygen availability and biogeochemistry along rainfall and topographic gradients in upland wet tropical forest soils. Biogeochemistry 44(3):301–328
44. Skopp J, Jawson MD, Doran JW (1990) Steady-state aerobic microbial activity as a function of soil water content. Soil Sci Soc Am J 54(6):1619–1625
45. Sogbedji JM, Van Es HM, Hutson JL (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 Soil 229(1):57–70
46. Soil Survey Staff (2019)National Resources Conservation Service United States Department of Agriculture (Web
47. Soil, Survey)
48. Tissari S, Nykänen V, Lerssi J, Kolehmainen M (2007) Classification of soil groups using weights-of-evidence-method and RBFLN-neural nets. Nat Resour Res 16(2):159–169
49. https://doi.org/10.1007/s11053-007-9040-y
50. Underwood KL, Rizzo DM, Dewoolkar MM, 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
51. Vesanto J, Himberg J, Alhoniemi E, Parhankangas J (2000) SOM toolbox for Matlab 5. Technical Report 57:1-59
52. Vidon P, Allan C, Burns D, Duval TP, Gurwick N, Inamdar S, Sebestyen S (2010) Hot spots and hot moments in riparian zones: Potential for improved water quality management. J Am Water Resour Assoc 46(2):278–298. https://doi.org/10.1111/j.1752-1688.2010.00420.x
53. Vidon P, Hill AR (2004) Denitrification and patterns of electron donors and acceptors in eight riparian zones with contrasting hydrogeology. Biogeochemistry 71(2):259–283
54. Wang Y, Li Y, Liu X, Liu F, Li Y (2014) Relating land use patterns to stream nutrient levels in red soil agricultural catchments in subtropical central China. Environ Sci Pollut Res 10481–10492.

https://doi.org/10.1007/s11356-014-2921-9

55. Wehrens R, Buydens LM (2007) Self-and super-organizing maps in R: the Kohonen package. J Stat Softw 21(5):1–19
56. Wu T, Chellemi DO, Graham JH, Martin KJ, Rosskopf EN (2008) Comparison of Soil Bacterial Communities Under Diverse Agricultural Land Management and Crop Production Practices. Microb Ecol 55(2):293–310. https://doi.org/10.1007/s00248-007-9276-4
57. Xiaoyong L, Huan TAO, Xuegang G, You LI (2019) Exploring the database of a soil environmental survey using a geo-self-organizing map:A pilot study. J Geog Sci 29(10):1610–1624

Figures

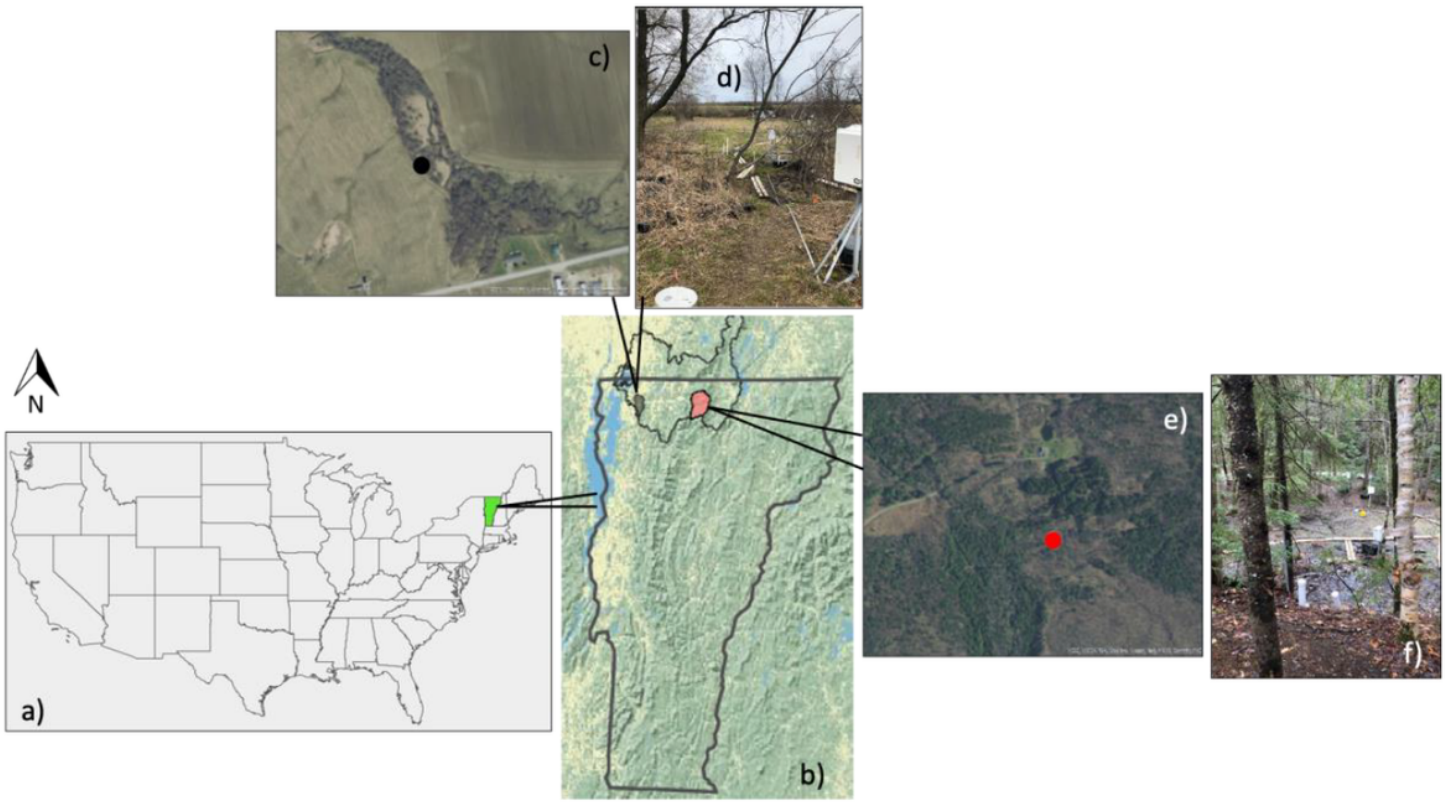

Figure 1

(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 site, which is located within the Hungerford Brook subwatershed (shaded in gray), is represented by a black circle. The Green Mountains site, located within the Trout River subwatershed (shaded in red) is represented by a red circle. (c) Aerial image of the Champlain Valley (CV) site (Sheldon, VT) with a black circle indicating where sensors are installed. (d) Photograph of the CV site riparian transect. (e) Aerial image of the Green Mountains (GM) site

(Montgomery, VT) with a red circle indicating where soil sensors are installed and (f) photograph of the GM transect

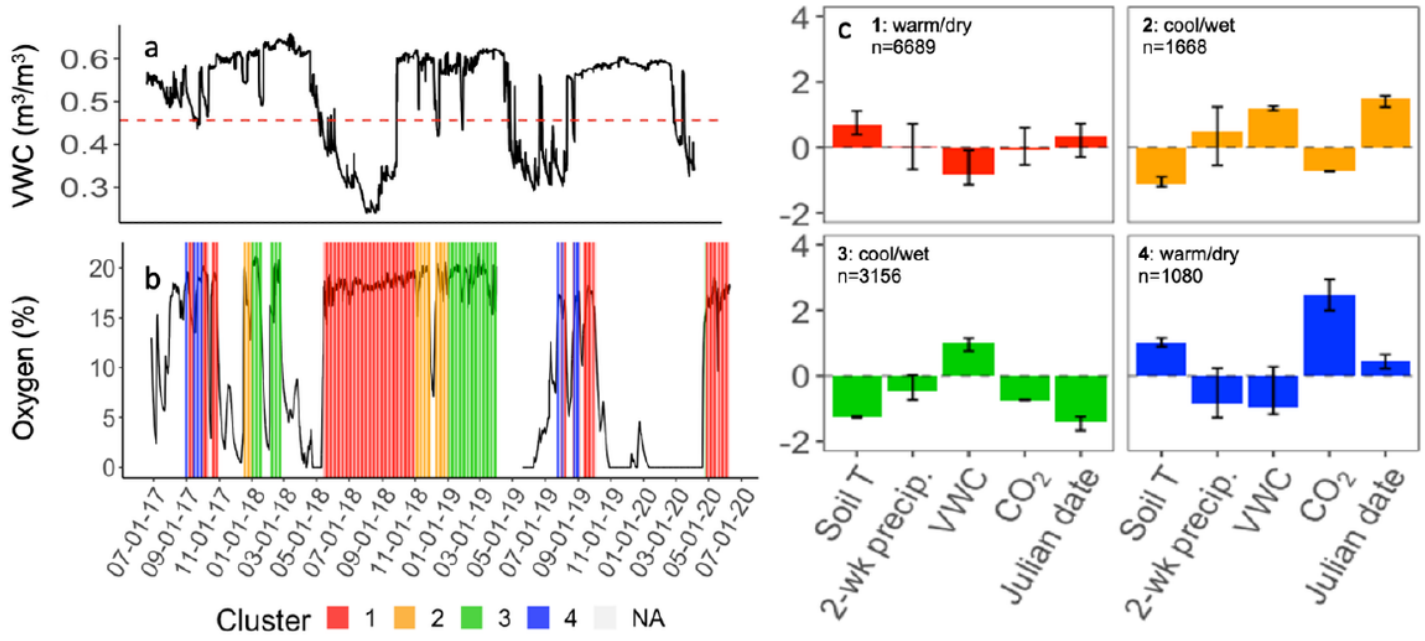

Figure 2

High O2 SOM results for the Champlain Valley (CV) site, including (a) volumetric water content (VWC) time series with dashed line representing mean VWC, (b) O2 time series highlighted with the four clusters identified by the high O2 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 O2 values outside of the high O2 range. 2-wk precip. is two-week antecedent precipitation

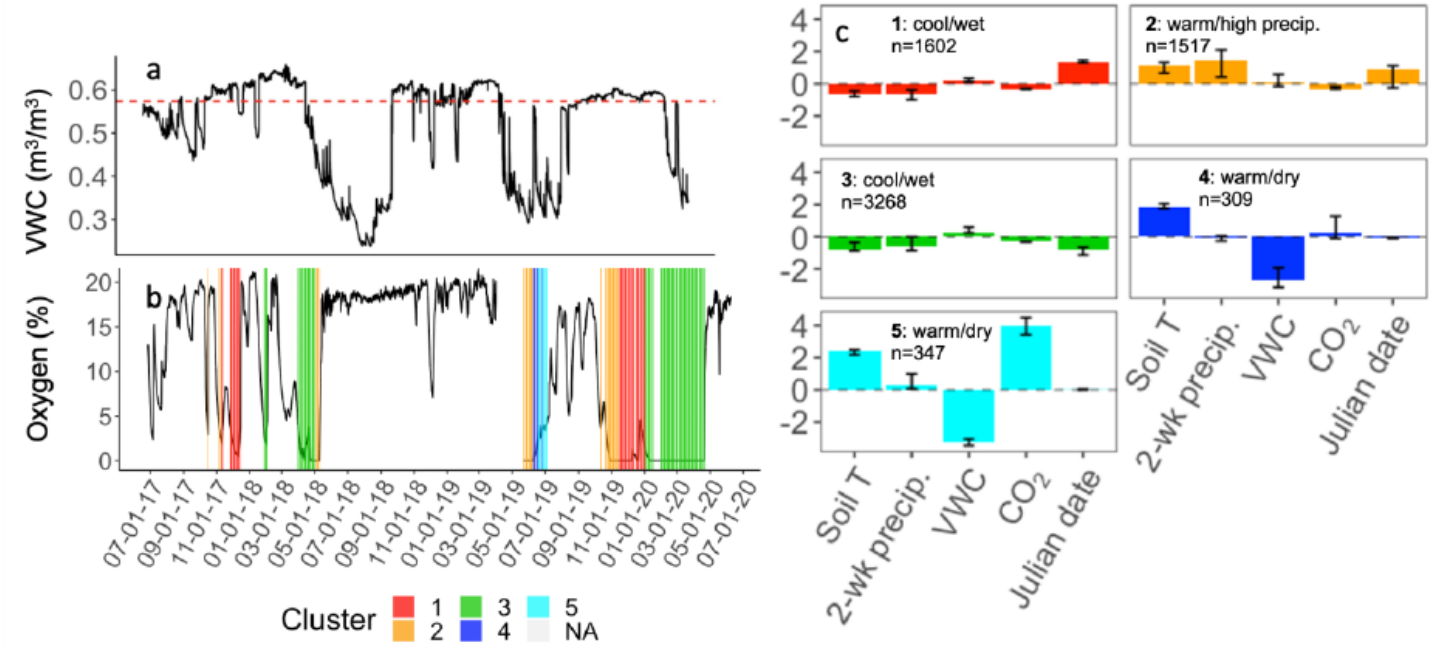

Figure 3

Low O2 SOM results for the Champlain Valley (CV) site, including: (a) volumetric water content (VWC) time series with dashed line representing mean VWC, (b) O2 time series highlighted with the five clusters identified by the low O2 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 O2 values outside of the low O2 range. 2-wk precip. is two-week antecedent precipitation

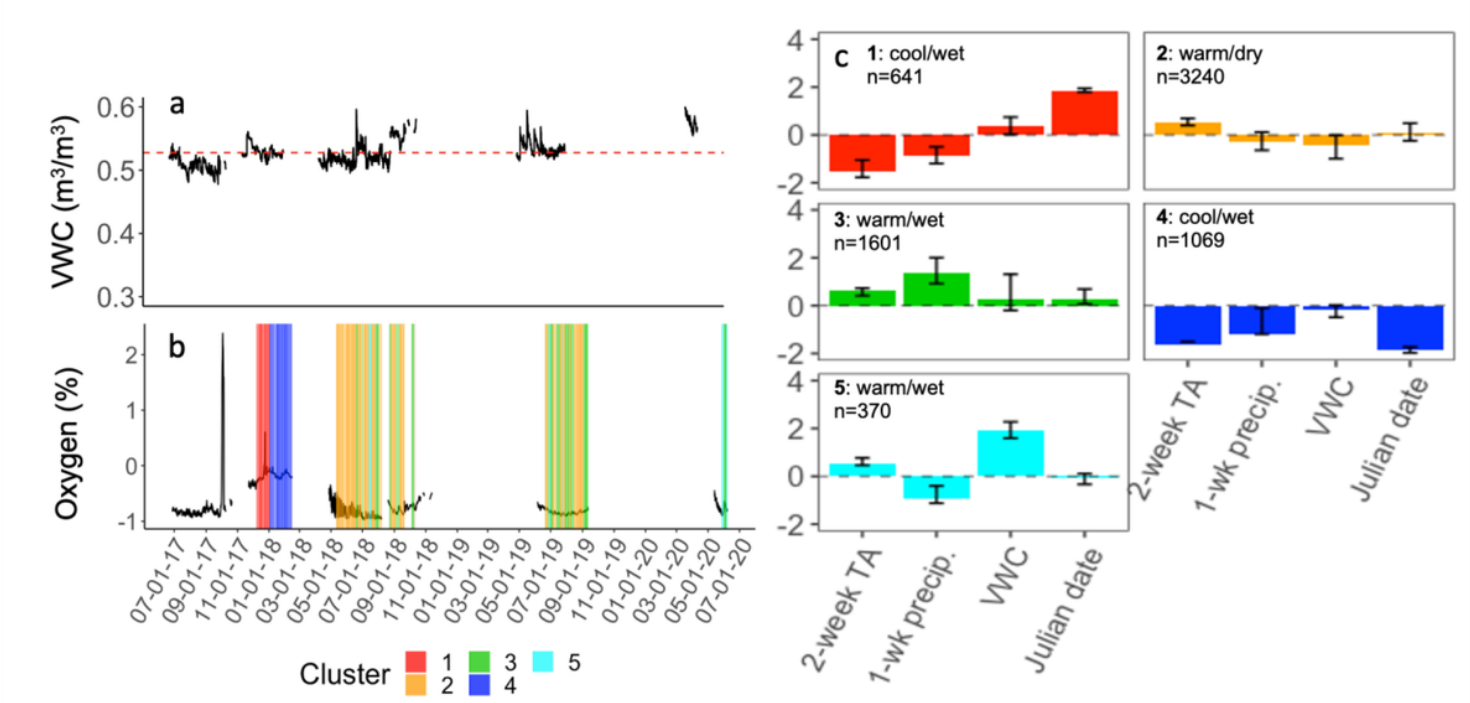

Figure 4

SOM results for the Green Mountains (GM) site, including: (a) volumetric water content (VWC) time series with dashed line representing mean VWC, (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. is one-week antecedent precipitation

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SI.docx