Ecological Engineering, Vol. 97, December 2016, pp. 567-576
DOI:10.1016/j.ecoleng.2016.10.003

A datamining approach to identifying spatial patterns of phosphorus forms in the Stormwater Treatment Areas in the Everglades, US.

Corstanje, R.[a]*, Grafius, D.R.[a], Zawadzka, J.[a], Moreira Barradas, J.[a], Vince, G.[b], Ivanoff, D.[c], Pietro, K.[c]

*Corresponding author: roncorstanje@cranfield.ac.uk

[a]Cranfield Soil and Agrifood Institute, Cranfield University, College Road, Cranfield, Bedfordshire, MK43 0AL, UK

[b]Tetra Tech Inc., 759 S. Federal Hwy., Suite 314, Stuart, FL 34994-2936, USA

[c]South Florida Water Management District, 3301 Gun Club Road, West Palm Beach, FL 33406, USA

Abstract

The Everglades ecosystem in Florida, USA, is naturally phosphorus (P) limited, and faces threats of ecosystem change and associated losses to habitat, biodiversity, and ecosystem function if subjected to high inflows of P and other nutrients. In addition to changes in historic hydropattern, upstream agriculture (sugar cane, vegetable, citrus) and urbanization has placed the Everglades at risk due to nutrient-rich runoff. In response to this threat, the Stormwater Treatment Areas (STAs) were constructed along the northern boundary of the Everglades as engineered ecological systems designed to retain P from water flowing into the Everglades. This research investigated data collected over a period from 2002 to 2014 from the interior of the STAs using data mining and analysis techniques including a) exploratory methods such as Principal Component Analysis to test for patterns and groupings in the data, and b) modelling approaches to test for predictive relationships between environmental variables. The purpose of this research was to reveal and compare spatial trends and relationships between environmental variables across the various treatment cells, flow-ways, and STAs. Common spatial patterns and their drivers indicated that the flow-ways do not function along simple linear gradients; instead forming zonal patterns of P distribution that may increasingly align with the predominant flow path over time. Findings also indicate that the primary drivers of the spatial distribution of P in many of these systems relate to soil characteristics. The results suggest that coupled cycles may be a key component of these systems; i.e. the movement and transformation of P is coupled to that of nitrogen (N).

Keywords: phosphorus, data mining, stormwater treatment areas, constructed wetland, Everglades, water quality

1. Introduction

The Stormwater Treatment Areas (STAs), located around the northern boundary of the Everglades in Florida, USA, were constructed over a period from 1994 to 2013. As a set of engineered ecological systems, the general purpose and function of the STAs is to reduce phosphorus (P) in runoff water prior to discharging to the Everglades Protection Area. They consist of a series of shallow, freshwater marshes divided into flow-ways and treatment cells by interior levees and control structures, populated with emergent or submergent aquatic vegetation (EAV and SAV, respectively) (Chen et al., 2015). The Everglades as a system is naturally P limited (Entry, 2014; McCormick et al., 1996), and so the water it receives must meet stringent requirements for ultra-low levels of water P (Pietro and Ivanoff, 2015). Since 1995, the STAs have treated approximately 16.5 billion m^3 inflow volume, retained approximately 1,727 metric tons (mt) of total phosphorus (TP), lowering phosphorus surface water concentrations from an overall annual TP of 140 micrograms per liter ($\mu g\ L^{-1}$) to 37 $\mu g\ L^{-1}$ (flow weighted mean; South Florida Water Management District, 2015), and improving further in most recent years to exhibit outflow concentrations averaging between 15-25 $\mu g\ L^{-1}$ (South Florida Water Management District et al., 2015). STA-2 and STA-3/4 are two of the best performing STAs, and have recorded reductions in surface water P from 100 and 87 $\mu g\ L^{-1}$ at inflow structures, respectively, to 23 and 18 $\mu g\ L^{-1}$ at outflow (Pietro and Ivanoff, 2015).

The STAs are wetland systems, and the controls on the P removal process are therefore set by the internal biogeochemical, ecological and physical processes and conditions in each cell, in each STA (Ivanoff et al., 2013). Phosphorus reduction from each STA must be maximized in order to meet stringent regulatory effluent limits, which implies that these natural processes must be manipulated (engineered) to maximize P retention. Phosphorus in surface water can have various forms; from soluble reactive to forms of organic and particulate

P with varied degrees of recalcitrance (Reddy and DeLaune, 2008). The retention of P in these systems needs to therefore consider these different forms.

There are abiotic processes of P retention, including P sorption to the STA soil particulates (Reddy et al., 1999) and particulate (co)-precipitation with cations such as calcium (Ca), magnesium (Mg), iron (Fe), and aluminium (Al) (Malecki-Brown et al., 2007). Factors that influence these processes are surface flow rate and path (Kadlec and Wallace, 2009) but also water and soil chemistry (e.g. concentrations of Ca, Mg, Fe and Al), pH, and the oxidation reduction potential (Reddy et al., 1999). Ideally this P then gets buried, or retained by the sediment within the wetland, resulting in gradually lower soil P-levels as water flows from the inflow point towards the outflow points (P gradient), similar to what has been observed in the nearby Water Conservation Area 2A (DeBusk et al., 1994). There are circumstances under which P is transported along the hydrologic gradient due to sediment re-suspension, P desorption from the sediment matrix, or poor vegetation condition. In properly performing STAs, these are limited and water column P could be reduced further down the flow-way, reducing the slope of the gradient. Uptake and retention of P by plants is generally (though not exhaustively; dependent upon plant type) considered to be short-term and rapid; while abiotic/physical retention processes tend to be longer term and are considered to account for 50-70% of permanent storage (Richardson, 1999).

Biological cycling of P involves direct uptake of available P by plant and microbial communities (Newman et al., 2001) to meet their physiological requirements, action of extracellular enzymes on complex organic P to release P uptake (Corstanje et al., 2007) and the release of P from the biological decomposition of organic material. Under anaerobic environments, decomposition of organic material is slow, resulting in formation and accretion of peat; forming another sink for P as long as the peat remains intact. Biological P cycling and the resulting spatial distribution of the different forms of P is highly complex, as it is driven by coupled P, N and C cycles; determined by redox conditions and characterized by the plant ecology (Chen et al., 2015; Orem et al., 2014; Reddy et al., 2011).

Extensive sampling has been conducted over a period from 2002 to 2014, in which soil, surface water and macrophytes have been sampled within the STA cells, resulting in a large dataset of observations. Coupled with hyper-spectral measurements made through various aerial surveys, the results comprise a fairly comprehensive dataset on the spatial variation in key components of the STA ecosystem. Here, we report on a broad scale analysis of these datasets, in order to determine common trends across the various flow-ways in the STAs, and in individual STAs. The expectation here is that common biogeochemical processes will generate common multivariate patterns across STAs. We then considered, given the extent and comprehensiveness of the datasets under consideration, implications for future monitoring of these systems.

2. Materials and Methods

2.1. Study Area

The STAs, operated by the South Florida Water Management District, cover an effective treatment area of circa 230 km^2. There are five STAs: STA-1E, STA-1W, STA-2, STA-3/4, and STA-5/6 (Figure 1); STA-5/6 was formerly two separate STAs until water year (WY) 2010. The STAs vary in size and location, and each is constructed with sets of interconnected cells forming treatment 'flow-ways'. Data from surface water (sampled along internal transects within the treatment cells), floc (i.e. flocculant; loosely clumped particles either suspended in the water column or resting atop the soil, analogous to litter in terrestrial systems), and soil collected within the various cells were available for analysis, and have been previously described and used to evaluate conditions within the STAs (e.g. Pietro and Ivanoff, 2015; Reddy et al., 2009). Normalized Difference Vegetation Index (NDVI) and vegetation class and habitat maps were derived from recent-year hyper-spectral imagery at a resolution of approximately 1 square foot to represent the approximate current state of vegetation within the cells. The available datasets were diverse in spatial extents, subjects (e.g. soil samples, surface water transects, vegetation coverage) and data types (e.g. categorical vs. continuous), necessitating a data mining approach capable of addressing this diversity. Below we describe the structure of each STA; specifics of data availability are described in the sections that follow.

STA-1E began full operation in 2006-2007 and consists of three flow-ways; Eastern, Western, and Central. Due to data availability only the Central Flow-way was analyzed here. STA-1W's Eastern and Western flow-ways were in operation from 1994 as the Everglades Nutrient Removal (ENR) project, with an additional Northern flow-way constructed in 2000. All three flow-ways were analyzed. STA-

2 Cells 1-3, each single-cell flow-ways, were operational from 2000 onwards. Additional cells, 4-8, involve multi-cell flow-ways and became operational between 2008 and 2012 but were not studied here due to insufficient data availability. STA-3/4 consists of three flow-ways (Flow-ways 1, 2 and 3) and became operational in 2004; all were included in analysis. STA-5 originally consisted of three flow-ways, denoted Flow-ways 1, 2 and 3; each consisting of a combination of two cells. Flow-ways 1 and 2 became operational in 1999; Flow-way 3 in 2008. Flow-ways 4 and 5 were later added, flow-capable in 2010, but not studied here. Combination with STA-6 to form STA-5/6 added three additional flow-ways; 6, 7 and 8, of which Flow-ways 7 and 8 are single cell flow-ways (operational in 1998), and Flow-way 6 (not analyzed) couples two cells (6-4, flow-capable in 2010 and 6-2, constructed in 2006).

2.2. Data quality control

Quality control checks were performed on all datasets at various stages of the data compilation. Blank or null records were treated as no data and not zero. For soil and floc data, parameter values were reported within specific ranges of the profile, typically ranging from 0 to 10 cm. Some records included data on the upper profile (0-10 cm), lower profile (10-30 cm), and full profile combined (0-30 cm). In some cases soil nutrients within selected STA cells were measured at variable depth increments (e.g. 0-2, 2-4, 4-6 cm, etc.). In such cases, all parameters for relevant increments were averaged into a single 0-10 cm field for analysis to ensure consistency across the dataset (including bulk density). In some other cases, the sampling depth of the upper profile did not reach 10 cm, but these were still marked as the upper profile. The full profile value was very rarely given, and was calculated only for the datasets that were subsequently used in the data mining analysis. In these instances, the average of the upper and lower profile was used.

2.3. Data Analysis

2.3.1. Preparation of datasets for data mining

The following rules were applied for inclusion of the data measured within the STAs: (1) There must be at least 10 observations for a given STA cell and year (an arbitrary cutoff point but sufficient to allow the calculation of meaningful statistics) and (2) There must be at least one instance of at least 10 observations per year within all STA cells in a flow-way. Seasonality at temporal scales finer than full years was not considered here. Additionally, any GIS data with full coverage of STA cells were considered. These included vector maps of vegetation class and habitat, NDVI rasters, and topography rasters representing the elevation differences of the STA floor at various year intervals. The resulting flow-ways included in data mining and their available data are listed in Table 1.

Table 1. List of flow-ways included in interpolation and their available data including years and number of observations (n). Surface water quality data are from transects internal to each treatment cell.

STA	*Flow-way*	*Cells*	*STA Data Availability*
STA-1E	Central	3 to 4N to 4S	Soil/floc (2004, 07, 09, 10; n=97) Surface water (2013; n=16) Macrophyte nutrients (2009; n=46) Hyper-spectral imagery (2011-12)
STA-1W	Eastern	1A and 1B to 3	Soil/floc (Eastern/Western FW only: 1995-97, 99; all FW: 2003-08, 10; n=1006) Surface water (2003, 04, 09-13; n=2689)
	Western	2A and 2B to 4	

	Northern	5A to 5B	Macrophyte nutrients (Eastern/Western FW only: 1996, 97; all FW: 2003, 04, 08-10; n=262) Hyper-spectral imagery (2011-12)
STA-2	Flow-way 1 Flow-way 2 Flow-way 3	1 2 3	Soil/floc (2003, 07, 09-11; n=830) Surface water (2003-10, 13, 14; n=1126) Macrophyte nutrients (2003, 09, 10; n=91) Hyper-spectral imagery (2011-12)
STA-3/4	Flow-way 1 Flow-way 2 Flow-way 3	1A to 1B 2A to 2B 3A to 3B	Soil/floc (2004, 07, 10; n=1272) Surface water (2003-10, 13, 14; n=1134) Macropyte nutrients (2010-12; n=58) Hyper-spectral imagery (2011-2012)
STA-5/6	Flow-way 1 Flow-way 2 Flow-way 7 Flow-way 8	1A to 1B 2A to 2B 5 3	Soil/floc (FW 1/2: 2002, 03, 07-11; n=617. FW 7/8: 2003, 07-11; n=138) Surface water (FW 1/2: 2013; n=74) Macrophyte nutrients (FW 1/2: 2002, 03; n=147. FW 7/8: 2003; n=31) Hyper-spectral imagery (all FW: 2011-12)

2.3.2. Interpolation of flow-way data within STA cells

Interpolation was done using an Empirical Bayesian Kriging (EBK) algorithm. For Bayesian geostatistical analysis, we used the Gaussian Spatial Linear Mixed Model as formulated by Diggle et al. (1998) without fixed effects:

$$Y(s_i) = W(s_i) + \varepsilon$$

where the random variable $Y(s_i)$ is an $n \times 1$ vector of observed values at locations $\mathrm{s}_1, \mathrm{s}_1, \ldots, \mathrm{s}_i$; W represents the spatial random effect which is a Gaussian process with mean of 0, variance of σ^2 (partial sill) and correlation function $R(h;\varphi)$, for which we selected an exponential correlation function: $R(h;\varphi) = \exp(-\frac{h}{\varphi})$; and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of errors with mean of 0 and variance of

τ^2(nugget variance). These semivariogram parameters were estimated using restricted maximum likelihood (REML). The EBK tool produced 1137 pairs of interpolated and standard error maps which, together with other spatial datasets available (described above in 2.3.1), were sampled with 100 randomly distributed points (separated by at least fifty feet) within each STA cell.

2.3.3. Multivariate Analysis

Multivariate analysis used a combination of exploratory and modeling tools to identify underlying patterns in the data. Within each treatment flow-way, data from all available years were pooled to facilitate a single, data-rich analysis. For initial calculation of summary statistics, the record set within each cell containing the greatest number of observations for each year of coverage was selected, and the mean and standard deviation of TP measurements were calculated across all recorded years in Microsoft Excel (Microsoft, 2003). The mean and standard deviation of key soil nutrients (i.e. total phosphorus, nitrogen and carbon) were calculated for entire STAs. Principal components analysis (PCA) and clustering analysis (CA) were used in an exploratory mode using JMP (SAS, 2013); PCA to determine the main axis of variation the datasets, and CA to determine if there were any meaningful groups in the observations. The primary goals were: (a) to determine if there are any consistent main drivers of variation across the flow-ways (i.e. do the flow-ways and STAs behave consistently across the board, or is each a unique system responding to unique operational circumstances); and (b) within each flow-way, to determine if there are natural groupings of multivariate data (e.g. are observations from areas around the inflow sufficiently similar in floc, soil and vegetation characteristics to cluster, and sufficiently distinct from other areas). We used a combination of Ward's and *k*-means clustering methods (Corstanje et al., 2009). Ward's is a minimum variance, hierarchical clustering method which produces a scree plot, that in turn allows us to both identify the optimal number of clusters and establish the seeds which are then used to run the *k*-means clustering process. This was then followed by Stepwise Canonical Discriminant (SCD) analysis in JMP (SAS, 2013) to help identify the primary drivers of the clusters.

Subsequently, we applied a set of non-linear, hierarchical structured models using Statistica (StatSoft, 2014) to predict surface water TP concentrations (Classification and Regression Trees; CART). Where no surface water TP data were available (as was the case in 10 out of 24 cells: STAs 1E, 2 Cell 2 only, 3/4, and 6), floc TP was substituted as the best available indicator of TP and its drivers in the flowing system. The CART approach has a number of advantages; the method is not sensitive to non-normal data, it accepts categorical as well as continuous data (needed as soil series and soil parent material are categorical, whereas soil organic matter is continuous) and it is not confounded by the presence of non-linear relationships (Breiman et al., 1984; McCune and Grace, 2002). Bayesian Belief Networks (BBNs), having similar advantages in their ability to handle non-normal and categorical data, were also created using Netica (Norsys, 2014) to predict the most recently available NDVI and TP (preferentially in surface water if available, otherwise in floc or soil as described above) in each cell. BBNs are graphical probabilistic models; graphical in that they represent the variables that affect the response of interest (e.g. floc or surface water P) in the form of a network, and probabilistic in that the relationships between the drivers and response are conditioned by a probability (Taalab et al., 2015). Bayesian inference is thus based on a set of prior probabilities that can be updated as new information becomes available. In this case, some knowledge of potential drivers of P dynamics was available from the CART analysis and a review of the existing P process literature; the network thus consisted of those variables that the previous CART models identified as drivers. For both CART and BBN approaches, model fitness and the strongest predictor variables were of primary interest.

3. Results

3.1. Summary Statistics

Data on TP from internal surface water transects and TP, total carbon (TC) and nitrogen (TN) from soil samples in all STAs and across all available years were pooled and their summary statistics calculated (Tables 2 and 3), but distributions were highly variable in terms of timing, data type, number of observations, and data were not available or complete for all cells and flow-ways. Cell 2A in STA-5/6 achieved the highest overall mean internal surface water TP (0.216 mg L^{-1}) followed by STA-1W's Cell 5A (0.129 mg L^{-1}). The Cells with the lowest mean internal surface water TP were STA-3/4's Cell 3B (0.012 mg L^{-1}) and STA-1W's Cell 4 (0.024 mg L^{-1}). Variability was present in the data, both within sets of records and between different years and cells; most standard deviations tended to fall proportionally between 30% and 80% of their associated means. Total soluble phosphorus (TSP) and soluble reactive phosphorus (SRP) in internal surface water were variable in their proportional relationship with TP (not shown); combined across all STAs, TSP averaged roughly half of TP (59.2%) with a standard deviation of 15.4%, and SRP averaged 28.1% of TP with a standard deviation of 14.6%. As these statistics summarize the data for entire treatment cells they do not address spatial patterns within individual cells (this is explored below in section 3.3); however in flow-ways composed of multiple cells, an apparent trend of decreasing mean TP was visible along the length of the flow-ways from the summary statistics, evidencing the removal of phosphorus from surface water as it flows through the STAs. The greatest proportional drop was in Flow-way 2 in STA-5/6, where Cell 2A exhibited a mean TP of 0.216 mg L^{-1} and Cell 2B a mean of 0.062 mg L^{-1}.

Table 1: Summary statistics for all combined data on total surface water phosphorus [mg L^{-1}] sampled within the STAs (internal surface water transect). SD = Standard Deviation, N = number of observations. Values marked 'n/a' represent cells where summary data were insufficient for calculation of summary statistics.

STA	*Flow-way*	*Cells*	*Mean*	*SD*	*N*
STA-1E	Central	3	n/a	n/a	0
		4N	0.108	0.017	16
		4S	n/a	n/a	0
STA-1W	Eastern	1A	0.106	0.049	8
		1B	0.065	0.044	159
		3	0.030	0.018	95
	Western	2A	0.123	0.069	77
		2B	0.047	0.022	89
		4	0.024	0.012	70
	Northern	5A	0.129	0.051	54
		5B	0.071	0.079	699
STA-2	Flow-way 1	1	0.044	0.036	197
	Flow-way 2	2	n/a	n/a	0
	Flow-way 3	3	0.034	0.024	606
STA-3/4	Flow-way 1	1A to 1B	n/a	n/a	0
	Flow-way 2	2A to 2B	n/a	n/a	0
	Flow-way 3	3A	0.037	0.005	4

		3B	0.012	0.001	42
STA-5/6	Flow-way 1	1A	0.064	0.048	12
		1B	0.031	0.023	16
	Flow-way 2	2A	0.216	0.074	12
		2B	0.062	0.045	16
	Flow-way 7	5	n/a	n/a	0
	Flow-way 8	3	n/a	n/a	0

Table 3: Summary statistics for all combined data on total soil phosphorus [TP; mg kg^{-1}], total carbon [TC; g kg^{-1}] and total nitrogen [TN; g kg^{-1}] sampled within the STAs. SD = Standard Deviation, N = number of observations.

STA	***Soil TP (mg kg^{-1})***			***Soil TC (g kg^{-1})***			***Soil TN (g kg^{-1})***		
	Mean	***SD***	***N***	***Mean***	***SD***	***N***	***Mean***	***SD***	***N***
STA-1E	241	207	294	85.2	63.3	294	5.7	4.3	294
STA-1W	550	237	1405	432	57.6	1322	26.5	4.2	1319
STA-2	611	250	1166	392	51.2	1078	23.1	3.7	1078
STA-3/4	718	243	1858	346	74.2	1857	22.1	5.0	1857
STA-5/6	727	315	952	285	111	783	20.5	7.8	783

Data for TP, TC and TN in soil and floc across the STAs were analyzed at the STA level. STA-5/6 exhibited the highest mean levels of soil TP (727 mg kg^{-1}), while STA-1W achieved the highest values for both mean TC (432 g kg^{-1}) and mean TN (26.5 g kg^{-1}). STA-1E had the lowest mean values for all three nutrients; 241 mg kg^{-1} TP, 85.2 g kg^{-1} TC, and 5.7 g kg^{-1} TN. Variability was highest in STA-5/6 across all three nutrients; exhibiting a standard deviation of 315 mg kg^{-1} TP, 111 g kg^{-1} TC, and 7.8 g kg^{-1} TN. TP variability was lowest in STA-1E (standard deviation of 207 mg kg^{-1}), while STA-2 displayed the lowest variability for both TC (51.2 mg kg^{-1}) and TN (3.7 g kg^{-1}). Note that these statistics represent averages across entire treatment cells or STAs; Table 3 reports the associated variability (as standard deviations).

3.2. Multivariate Analysis Results

Principal Component Analysis (PCA) results are characteristically not straightforward to interpret and do not involve clear cutoffs to determine whether or not a component variable can be considered specifically important or unimportant, so focus was placed on determining and reporting those variables that were clearly the strongest drivers and/or recurred consistently across STAs. Results varied by cell, but the most commonly identified variables related to soil TC, soil TN, soil and floc bulk density (BD), soil and floc TP, and soil and floc ash-free dry weight (AFDW) as the greatest contributors to variability in the data (Table 4). Cluster analysis identified 3 or 4 clusters in most cells, with spatial structure to cluster membership apparent in some but not all cells (Table 5).

Table 4: Summary of the main outcomes from Principal Component Analysis (soil/floc/surface water parameters separated by semicolon). Abbreviations: total phosphorus (TP), total carbon (TC), total nitrogen (TN), bulk density (BD), sulfur (S), calcium (Ca), iron (Fe), macrophyte nutrient (macro), exchange capacity (exc), surface water (sw), alkalinity (Alk) ash-free dry weight (AFDW). Note that data availability was not consistent (e.g. few surface water observations in STA-1E Cells 3 and 4S) so PCA may not accurately reflect the importance of underrepresented variables in some cells.

STA	***Flow-way***	***Cell***	***PCA main variables***	***% var explained by PC1,..,PC3***
STA-1E	Central	3	Soil TC, TN, AFDW, BD, TP, Ca	80.25
		4N	Soil AFDW, BD, TC, Ca, Fe, TP	79.58
		4S	Soil AFDW, BD, TC, TN, TP, Ca, Fe	77.77
STA-1W	Northern	5A	Soil AFDW, BD, TC, TN, TP; floc AFDW; sw TP	83.23
		5B	Floc BD, TC, AFDW; sw Ca, P	68.01
	Eastern	1	n/a*	57.49
		3	Soil Al exc, Fe exc, TN, Alk, AFDW, BD, K; sw TP	76.59
	Western	2	Soil Fe, BD, TC; sw TP, Ca, AFDW	81.34

		4	Soil AFDW, Fe, TC, TN; sw Ca, TP	81.66
STA-2	1	1	Soil TC, TN, TP; floc TC; sw TP	75.09
	2	2	Floc BD, TC, TN, TP	72.19
	3	3	Soil macroDryWt, TC; sw Ca, TP	77.98
STA-3/4	1	1A	Soil BD, TN; floc BD, TP	72.54
		1B	Soil TP, TN, BD; floc TN; sw TN	71.78
	2	2A	Soil BD, TC, TN; floc BD, TC, TN	63.12
		2B	Soil TC, TP, TN, BD	71.71
	3	3A	Soil TC, TN, BD, TP; sw Ca, TP	74.21
		3B	Floc TC, TN; sw Ca, P	69.59
STA-5/6	1	1A	Soil macro TN, Fe, BD; floc DryWt; sw Ca	72.11
		1B	Soil TC, TN, S, TP; floc dryWt, AFDW; sw TP	79.28
	2	2A	Soil TC, TP, AFDW; floc BD; sw TP	77.99
		2B	Soil TN, TP, BD, TC	77.17
	7	5	Soil AFDW, TC, TN, Ca; floc moisture	77.90
	8	3	Soil AFDW, Fe, macro AFDW, TC; floc AFDW	77.87

* STA-1W Cell 1 PCA results consisted of similar and low average values, not highlighting any particular driving variables.

Table 5: List of analyzed flow-ways by age, number of clusters and observed spatial pattern of clusters (maps of cluster patterns available in supplementary material).

STA	*Flow-way*	*Oper. start year*	*Cell*	*No. clusters*	*Observed cluster pattern*
STA-1E	Central	2006/7	3	4	Zonal
			4N	3	Zonal
			4S	3	Zonal
STA-1W	Eastern	1994	1	3	Zonal gradient
			3	4	Zonal gradient
	Western	1994	2	4	Zonal gradient
			4	4	Tenuous zonal gradient
	Northern	2000	5A	4	Zonal
			5B	4	Tenuous zonal gradient
STA-2	Flow-way 1	2000	1	5	Zonal gradient
	Flow-way 2	2000	2	4	Zonal
	Flow-way 3	2000	3	4	Zonal gradient
STA-3/4	Flow-way 1	2004	1A	5	Zonal
		2004	1B	4	Zonal

	Flow-way 2	2004	2A	4	Zonal
		2004	2B	3	Zonal
	Flow-way 3	2004	3A	3	Tenuous zonal gradient
		2004	3B	5	Zonal
STA-5/6	Flow-way 1	1999	1A	6	Tenuous zonal gradient
		1999	1B	3	Tenuous zonal gradient
	Flow-way 2	1999	2A	4	Zonal
		1999	2B	3	Zonal
	Flow-way 7	1998	5	3	Zonal gradient
	Flow-way 8	1998	3	5	Zonal gradient

CART analysis consistently found the strongest predictor variables for surface water and floc TP to be other variables relating to P content (i.e. P in different forms such as SRP, etc.) in soil, floc, and surface water; soil and floc BD; and soil and floc TN. Measures relating to AFDW, TC and Ca also showed occasional influence but were less widespread. Maps of CART model standard error by location (not pictured) did not generally reveal any spatial relationships with direction of flow, but did in some cases reveal zonal structures similar to the cluster analysis (described below in 3.3).

Analysis with BBNs identified the strongest consistent predictors of recent year NDVI to be variables relating to: vegetation type and cover, NDVI from previous years, surface water TP, soil and floc TN, and soil and floc TC. BBNs predicting surface water TP were most influenced by: other forms of surface water P, soil BD, soil TN, soil TC, and soil TP.

3.3. Spatial Trends

Spatial patterns varied to a degree among treatment flow-ways. For instance, floc and macrophyte characteristics dominated the models which predicted surface water TP in STA-5/6; soil physical properties (e.g. bulk density) described many of the spatial patterns in the treatment flow-ways of STA-3/4, etc. Notwithstanding this, some general observations can be made regarding all treatment flow-ways: (1) there are clear zonal patterns consistently present in these systems that are, in many cases, independent of the direction of flow and do not exhibit a simple linear gradient (Figure 2 shows STA-3/4 Flow-way 3 as an example of purely zonal pattern; other examples include Flow-ways 1 and 2 in the same STA and STA-1E's Central Flow-way, shown in supplementary material); however these zonal patterns appear to align along the direction of flow in the case of some older STAs and flow-ways (Figure 3 shows STA-1W's Eastern flow-way as an example of zone-based gradient pattern; other examples include STA-1W's Western Flow-way, STA-2's Flow-ways 1 and 3, and STA-5/6's Flow-ways 7 and 8, shown in supplementary material and summarized in Table 5); (2) There is some consistency in the spatial arrangement of these zones over the treatment flow-ways, such as surface water TP concentration being highest close to the inflow structures and there closely associated with a zone of higher floc and soil TP concentrations. Following these points, there is rarely any further consistency in the spatial organization of zones, or in their characterization, across flow-ways; but 3) soil TN often becomes an important factor characterizing the zone around the outflow (e.g. STA-1W, STA-3/4).

4. Discussion and Conclusions

4.1. Summary Statistics

Two results stood out from the cell-wide summary statistics that were consistent with expectations. Firstly, the lowest mean values of internal surface water TP were found in flow-ways present in STAs 2 and 3/4, which have been previously cited as being two of the best-performing STAs for P removal (Pietro and Ivanoff, 2015). Secondly, all flow-ways consisting of multiple cells exhibited a trend of decreasing TP along the length of the flow-way (cell-wide summary statistics did not consider internal spatial patterns of single-cell flow-ways; these are discussed below), demonstrating the effects of P removal by the system at the STA scale. Taken broadly, this is consistent with the expectation that wetlands experiencing a uniform sheet flow should exhibit P decreases along a longitudinal flow-based gradient (Walker and Kadlec, 2011).

4.2. Multivariate Analysis

In considering the outputs from the data-mining analysis for the flow-ways; PCA is a general dimension reduction technique in which the underlying variation is maintained. It was used here because it is one of the primary steps in any multivariate data analysis as well as an effective way to represent variation in the data. Generally the PCA was successful, with an average of 75% of the variation explained. The most common variables identified as influential in the PC loadings were soil TC, soil TN, soil and floc BD, soil and floc TP and soil and floc AFDW. It should be noted that this particular analysis does not take into account non-continuous data (e.g. categorical variables such as soil series and parent material). In essence, the outcome from this analysis is an effective summarization of the data but with little further insight into drivers, mainly highlighting that most of the within cell/within flow-way variation is driven by sediment nutrient concentrations and, to a lesser degree, floc TC and nutrient content.

Cluster analysis resulted in cluster memberships that could be assigned to the original data, revealing spatial patterns and structure in the data. Of interest here were two points; do the data resolve clearly in clusters, and if so, how many (i.e. how many classes of data are there in an STA flow-way), and are these classes meaningful in any way? In general, most cells could be described by 3 to 5 clusters and only in one case (STA-5/6 Cell 1A; 6 clusters) were more clusters needed (see Table 5). Clusters consistently grouped spatially into zone features which did not appear to be tied to cell location within the flow path in many cases; however in some cells these zonal features were observed to align along the direction of flow. While not an unequivocal relationship, these 'zone-based gradient' patterns appeared more likely to occur in older STAs and flow-ways (Table 5). Patterns seemed only tenuously related to flow path at best in STAs-1E and -3/4 (completed in 2007 and 2004, respectively), and generally more obviously following the flow gradient in STA-1W (completed in 1994-2000), STA-2 (completed in 2000), and STA-5/6 (completed in 1998/9).

The CART and BBN analyses both revealed similar relationships and driving variables in the data. Surface water TP was found to share consistently strong linkages with other forms of phosphorus in surface water (e.g. SRP and TSP) as well as in floc and soil. Nitrogen, carbon, and bulk density in soil and floc also factored in frequently; this highlights the potential importance of soil properties to P dynamics in the STAs, as well as the possibility of coupled cycles wherein P, N, and possibly C dynamics share co-dependencies and interrelationships.

4.3. Observed Relationships and Drivers of P Dynamics

It is evident from studies in the Everglades and elsewhere (Bayley and Mewhort, 2004; Bostic and White, 2007; Gu and Dreschel, 2008; Riggsbee et al., 2012), that plant communities actively regulate P dynamics in wetlands. In the STAs, low levels of water column P are achieved using strategic combinations of SAV and EAV to address P in different forms and in different stages of the flow-ways (Chen et al., 2015). In projecting this fact on the data mining exercise, one would expect the spatial patterns of soil P to reflect plant community composition, and plant communities would be expected to be a strong determinant in any predictive model for soil or floc P. In our analysis this was only rarely the case; however these effects may be obscured by the fact that much of the available data on vegetation composition were categorical (e.g. vegetation class and habitat type; NDVI being the notable exception as a continuous variable), and thereby only possible to include in CART and BBN analyses. Both CARTs and BBNs modeling surface water TP did not commonly reveal vegetation-related measures as key predictors, but BBNs predicting NDVI frequently did highlight surface water TP as an important driver (i.e. TP did not appear driven by vegetation, but vegetation appeared driven by TP). Linkages between TP and vegetation therefore may not be direct or omnipresent, but our analysis shows support for some relationships.

Where P is limiting, or effectually buried, and therefore not available for the plant communities, this may be reflected as plant stress (i.e. P limitation), which can be remotely determined using NDVI (Henrik, 2012). The hypothesis is that the indication of effective functioning of an STA is that, in the lower reaches of a flow path, the vegetation may become P-limited. As a first instance, predictive modeling of NDVI should indicate whether this is responsive to floc and soil nutrient status. For BBNs predicting NDVI this indeed was the case; the strongest predictors consistently included floc and soil nutrients, along with surface water TP and other measures of vegetation health and composition. Note however that prolonged exposure to low P concentrations may trigger a shift in plant community composition to species that are more adapted to the low levels; such a shift would be reflected in categorical habitat variables but not necessarily by a decrease in NDVI. This highlights the importance of vegetation-related measures beyond NDVI, and in turn the importance of methods such as BBNs that can consider categorical expressions of vegetation community.

4.4. Spatial Patterns of P and their Implications

The observation that consistent spatial patterns appear zonal rather than based on simple gradients is probably the most significant finding of the data mining, in that the processes controlling P in these systems operate in zones in the treatment flow-way, rather than along a smooth linear gradient as would be the expectation (see Table 5). These zones are observed repeatedly across STAs and flow-ways, and are consistently present as modeling outcomes (e.g. cluster analysis and CART outputs) and as such are unlikely to be a modeling artifact. There are a number of implications from approaching the STA flow-ways as zones rather than a simple gradient. From a research perspective, the relative importance of different factors, transformation and transport pathways of P occurs in spatial patterns, and the form and shape of these patterns indicates the relative importance of particular pathways. Likewise, this affects the experimental sampling design, as these would then target zones rather than seeking to measure along a gradient (biased sampling). From a management perspective, this could simplify management options in that the operation and management strategies can be directed at particular zones within a treatment flow-way rather than an entire cell or the full flow-way, particularly once the drivers of these zones are better understood. Nevertheless, in older STAs (e.g. STA-1W, -2, and -5/6) these zonal patterns appeared to align more frequently and obviously with the direction of flow, suggesting that P dynamics may function largely in zonal patterns but slowly shift toward a zone-based gradient pattern over the operational time of an STA. Of particular note, STA-2 flow-way 3 exhibited a strong gradient pattern in the cluster analysis result and has been previously studied as one of the longest-running and best-performing treatment flow-ways (Juston and Debusk, 2011; Juston et al., 2013).

The finding of zonal patterns of P concentrations in the STAs (whether forming zone-based flow gradients or not), rather than simple uniform gradients decreasing along the axis of water flow, differs from previous findings and the usual expectation of P dynamics in wetlands (e.g. Kadlec, 1999; Walker and Kadlec, 2011). One possible explanation for this difference is that the treatment cells may be wide enough to allow partial mixing of water rather than a relatively uniform sheet flow; this would account for more complex patterns (Walker and Kadlec, 2011). If true, this would have implications for the assumptions made in future flow modeling efforts in the STAs, and require a more complex interpretation of the system than a one-dimensional sheet flow. Chen et al. (2015) cautioned that analyses focused solely on inflow and outflow P concentrations, while useful, do not consider P removal processes internal to the treatment cells, as well as recommending that future studies consider multivariate relationships. Doing so here has enabled additional findings, such as the potential importance of relationships between P and soil factors, and the possibility of P-N coupled cycles impacting dynamics. This latter result, while not widely explored previously, is consistent with previous findings in Water Conservation Area 2A (WCA 2a) on P and N functional linkages (White and Reddy, 2003). Corstanje et al. (2009, 2007) found evidence that areas enriched with P in WC-2a are mediated by N related parameters, such as potentially mineralizable N and related microbial extracellular enzymatic activities. In STA areas closest to the inflow, as P is relatively plentiful, the cycling P is likely to be co-mediated by N and its dynamics.

4.5. Data-Mining Advantages and Future Research

Previous studies have examined the extensive data now available for P dynamics in the STAs (e.g. Chen et al., 2015; Juston et al., 2013; Pietro and Ivanoff, 2015), but this is one of the first known studies to comprehensively make use of the diverse data collected in the interior treatment cells and flow-ways (e.g., soils, vegetation, internal water quality) and the first to do so at such a broad scale through a data mining approach. Doing so has facilitated new findings and understanding around the functional P dynamics of the STA systems. Approaches making use of these techniques are valuable for identifying biogeochemical relationships, and should be considered and further employed in future studies of the STAs as well as other engineered wetlands where sufficient data are available.

In addition, there remain a number of further considerations moving forward. First, many links between plant community composition and P dynamics remain unclear beyond known differences between EAV and SAV in P removal (e.g. Dierberg et al., 2002; Juston and DeBusk, 2006). In particular, we suspect there is an element of scale effect; where these processes occur and are important at scales finer than we considered in this study. Second, the approach used here focused on data mining techniques, and while effective for exploring patterns in the data it lacks a detailed process understanding of P biogeochemistry. The incorporation of process understanding and process models (e.g. first order equations) into the more stochastic modeling environment considered in this study could produce a set of hybrid models which would both reflect process knowledge and understanding but also, critically, allow for scaling and mapping. Such an approach could better explore the process-based reasons for the zonal patterns observed here and their potential relationships with flow-way age. Finally, future research should seek to effectively consider the interaction between different datasets available from the STAs in order to rigorously consider time series analysis and pulsed events. A future study which initiates with a thorough decomposition of the STA inflow and outflow data (volume and concentrations), considers the stochasticity of this data and then moves to incorporate it in the models of flow-way behavior should generate significant insights in the STA dynamics, and to what degree performance is related to stochastic events (e.g. storms or droughts) vs. deterministic processes (e.g. P biogeochemistry, SAV, periphyton). Eventually this will relate to a measure of the resilience of these systems; expressed as their capacity to withstand pressures and maintain long term performance.

4.6. Conclusions

In conclusion, the use of data mining approaches on STA treatment cell and flow-way data has identified, in a very general sense, spatial patterns in these systems. These patterns are consistently zone-based across all flow-ways, which suggests that the flow-ways function first as zonal systems rather than simple linear gradient systems. Our analysis suggests that the primary drivers of the spatial distribution of P in many of these systems are related to soil characteristics, and that the zonal patterns of P distribution may begin to follow the predominant flow path over time. The data further suggest the importance of coupled cycles in these systems; in other words, the movement and transformation of P is coupled to that of N.

Acknowledgements

This research was conducted for the South Florida Water Management District (SFWMD) in support of the Restoration Strategies Science Plan (SFWMD 2013).

References

Bayley, S.E., Mewhort, R.L., 2004. Plant community structure and functional differences between marshes and fens in the southern boreal region of Alberta, Canada. Wetlands 24, 277–294. doi:10.1672/0277-5212(2004)024

Bostic, E.M., White, J.R., 2007. Soil Phosphorus and Vegetation Influence on Wetland Phosphorus Release after Simulated Drought. Soil Sci. Soc. Am. J. 71, 238. doi:10.2136/sssaj2006.0137

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.I., 1984. Classification and regression trees. Wadsworth, Belmont, California.

Chen, H., Ivanoff, D., Pietro, K., 2015. Long-term phosphorus removal in the Everglades stormwater treatment areas of South Florida in the United States. Ecol. Eng. 79, 158–168. doi:10.1016/j.ecoleng.2014.12.012

Corstanje, R., Portier, K.M., Reddy, K.R., 2009. Discriminant analysis of biogeochemical indicators of nutrient enrichment in a Florida wetland. Eur. J. Soil Sci. 60, 974–981. doi:10.1111/j.1365-2389.2009.01186.x

Corstanje, R., Reddy, K.R., Prenger, J.P., Newman, S., Ogram, A. V., 2007. Soil microbial eco-physiological response to nutrient enrichment in a sub-tropical wetland. Ecol. Indic. 7, 277–289. doi:10.1016/j.ecolind.2006.02.002

DeBusk, W.F., Reddy, K.R., Wang, Y., Koch, M.S., 1994. Spatial Distribution of Soil Nutrients in a Northern Everglades Marsh: Water Conservation Area 2A. Soil Sci. Soc. Am. J. 58, 543. doi:10.2136/sssaj1994.03615995005800020042x

Dierberg, F.E., DeBusk, T.A., Jackson, S.D., Chimney, M.J., Pietro, K., 2002. Submerged aquatic vegetation-based treatment wetlands for removing phosphorus from agricultural runoff: Response to hydraulic and nutrient loading. Water Res. 36, 1409–1422. doi:10.1016/S0043-1354(01)00354-2

Diggle, P.J., Tawn, J.A., Moyeed, R.A., 1998. Model-based Geostatistics. Appl. Stat. 47, 299–350. doi:10.1111/1467-9876.00113

Entry, J.A., 2014. The Impact of Stormwater Treatment and Best Management Practices on Nutrient Concentration in the Florida Everglades. Water, Air, Soil Pollut. 225, 1758. doi:10.1007/s11270-013-1758-z

Gu, B., Dreschel, T., 2008. Effects of plant community and phosphorus loading rate on constructed wetland performance in Florida, USA. Wetlands 28, 81–91. doi:10.1672/07-24.1

Henrik, J.J., 2012. Utilizing NDVI and remote sensing data to identify spatial variability in plant stress as influenced by management. Iowa State University.

Ivanoff, D.B., Pietro, K., Chen, H., Gerry, L., 2013. Chapter 5: Performance and Optimization of the Everglades Stormwater Treatment Areas, in: 2013 South Florida Environmental Report - Volume I. South Florida Water Management District, West Palm Beach, FL.

Juston, J., DeBusk, T.A., 2006. Phosphorus mass load and outflow concentration relationships in stormwater treatment areas for Everglades restoration. Ecol. Eng. 26, 206–223. doi:10.1016/j.ecoleng.2005.09.011

Juston, J.M., Debusk, T.A., 2011. Evidence and implications of the background phosphorus concentration of submerged aquatic vegetation wetlands in Stormwater Treatment Areas for Everglades restoration. Water Resour. Res. 47, 1–13. doi:10.1029/2010WR009294

Juston, J.M., DeBusk, T.A., Grace, K.A., Jackson, S.D., 2013. A model of phosphorus cycling to explore the role of biomass turnover in submerged aquatic vegetation wetlands for Everglades restoration. Ecol. Modell. 251, 135–149. doi:10.1016/j.ecolmodel.2012.12.001

Kadlec, R.H., 1999. The limits of phosphorus removal in wetlands. Wetl. Ecol. Manag. 7, 165–175. doi:10.1023/A:1008415401082

Kadlec, R.H., Wallace, S.D., 2009. Treatment Wetlands, 2nd ed. CRC Press, Boca Raton, FL.

Malecki-Brown, L.M., White, J.R., Reddy, K.R., 2007. Soil Biogeochemical Characteristics Influenced by Alum Application in a Municipal Wastewater Treatment Wetland. J. Environ. Qual. 36, 1904. doi:10.2134/jeq2007.0159

McCormick, P. V., Rawlik, P.S., Lurding, K., Smith, E.P., Sklar, F.H., 1996. Periphyton-Water Quality Relationships along a Nutrient Gradient in the Northern Florida Everglades. J. North Am. Benthol. Soc. 15, 433–449. doi:10.2307/1467797

McCune, B., Grace, J.B., 2002. Analysis of Ecological Communities. MjM Software, Gleneden Beach, Oregon, USA.

Microsoft, 2003. Excel.

Newman, S., Kumpf, H., Laing, J.A., Kennedy, W.C., 2001. Decomposition responses to phosphorus enrichment in an Everglades (USA) slough. Biogeochemistry 54, 229–250. doi:10.1023/A:1010659016876

Norsys, 2014. Netica.

Orem, W., Newman, S., Osborne, T.Z., Reddy, K.R., 2014. Projecting Changes in Everglades Soil Biogeochemistry for Carbon and Other Key Elements, to Possible 2060 Climate and Hydrologic Scenarios. Environ. Manage. 55, 776–798. doi:10.1007/s00267-014-0381-0

Pietro, K.C., Ivanoff, D., 2015. Comparison of long-term phosphorus removal performance of two large-scale constructed wetlands in South Florida, U.S.A. Ecol. Eng. 79, 143–157. doi:10.1016/j.ecoleng.2014.12.013

Reddy, K.R., DeLaune, R.D., 2008. Biogeochemistry of Wetlands: Science and Applications. CRC Press, Boca Raton, FL.

Reddy, K.R., Jawitz, J.W., Paudel, R., Bhomia, R., Jerauld, M.A., 2009. Comprehensive Analysis and Evaluation of Historical Data and Information for the Stormwater Treatment Areas (STAs). Gainesville.

Reddy, K.R., Kadlec, R.H., Flaig, E., Gale, P.M., 1999. Phosphorus Retention in Streams and Wetlands: A Review. Crit. Rev. Environ. Sci. Technol. 29, 83–146. doi:10.1080/10643389991259182

Reddy, K.R., Newman, S., Osborne, T.Z., White, J.R., Fitz, H.C., 2011. Phosphorous Cycling in the Greater Everglades Ecosystem: Legacy Phosphorous Implications for Management and Restoration. Crit. Rev. Environ. Sci. Technol. 41, 149–186. doi:10.1080/10643389.2010.530932

Richardson, C.J., 1999. The role of wetlands in storage, release, and cycling of phosphorus on the landscape: a 25-year retrospective, in: Reddy, K.R., O'Connor, G.A., Schelske, C.L. (Eds.), Phosphorus Biogeochemistry of Sub-Tropical Ecosystems. CRC Press, Taylor & Francis Group, Boca Raton, USA.

Riggsbee, J.A., Wetzel, R., Doyle, M.W., 2012. Physical and plant community controls on nitrogen and phosphorus leaching from impounded riverine wetlands following dam removal. River Res. Appl. 28, 1439–1450. doi:10.1002/rra.1536

SAS, 2013. JMP.

South Florida Water Management District, 2015. South Florida Environmental Report 2015. West Palm Beach.

South Florida Water Management District, Andreotta, H., Chimney, M., DeBusk, T., Garrett, B., Gerry, L., Henry, J., Ivanoff, D., Jerauld, M., Kharbanda, M., Kirkland, M., Larson, N., Miao, S., Piccone, T., Pietro, K., Schwartz, L., Sierer-Finn, D., Toth, L., Xue, S.K., Yan, Y., Zamorano, M., Zhao., H., 2015. Chapter 5B : Performance of the Everglades Stormwater Treatment Areas, 2015 South Florida Environmental Report. West Palm Beach.

StatSoft, 2014. Statistica 12.

Taalab, K., Corstanje, R., Zawadzka, J., Mayr, T., Whelan, M.J., Hannam, J.A., Creamer, R., 2015. On the application of Bayesian Networks in Digital Soil Mapping. Geoderma 259-260, 134–148. doi:10.1016/j.geoderma.2015.05.014

Walker, W.W., Kadlec, R.H., 2011. Modeling Phosphorus Dynamics in Everglades Wetlands and Stormwater Treatment Areas. Crit. Rev. Environ. Sci. Technol. 41, 430–446. doi:10.1080/10643389.2010.531225

White, J.R., Reddy, K.R., 2003. Nitrification and Denitrification Rates of Everglades Wetland Soils along a Phosphorus-Impacted Gradient. J. Environ. Qual. 32, 2436. doi:10.2134/jeq2003.2436

Figures

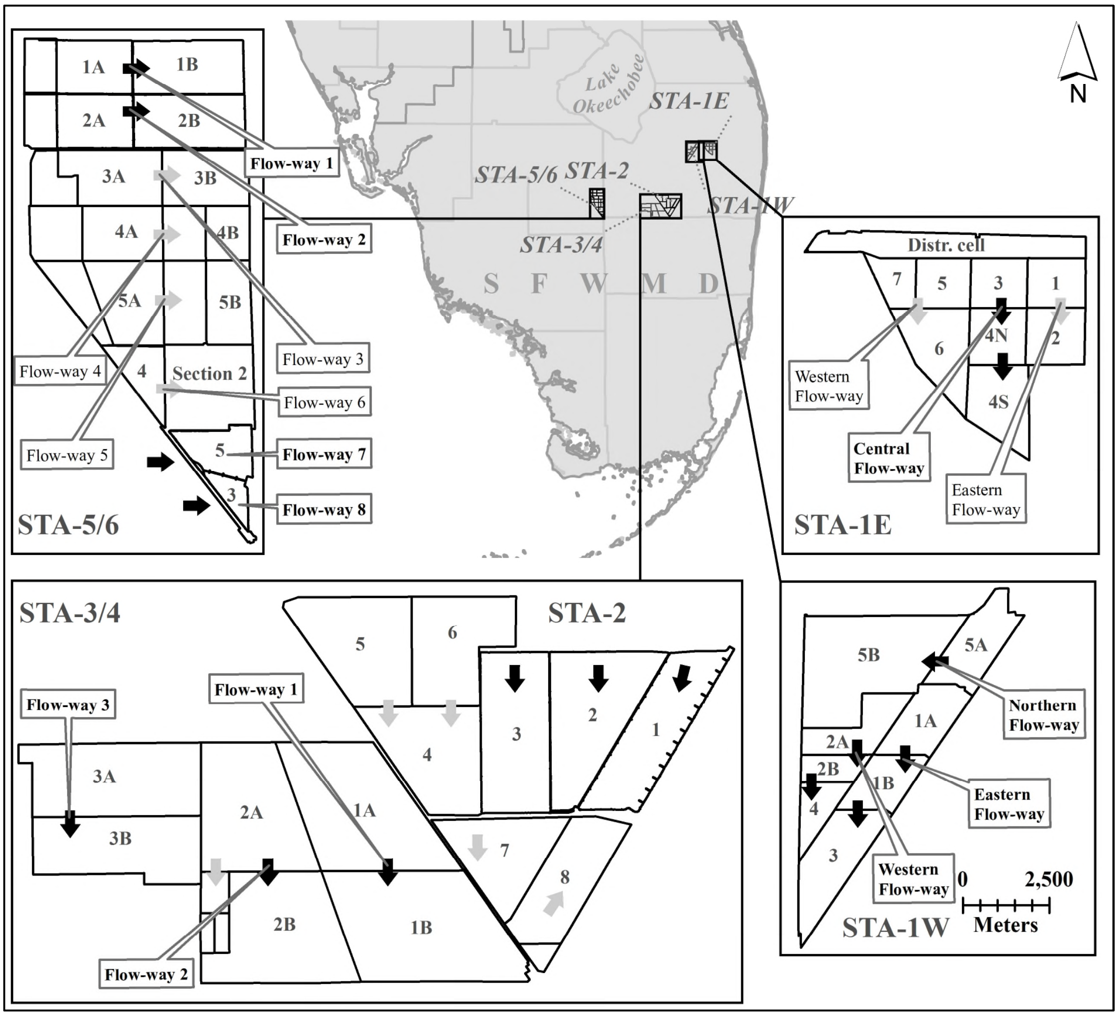


Figure 1: Locations of the Stormwater Treatment Areas in south Florida, USA, indicating individual treatment cells and direction of flow. Bolded flow-way names and darkened arrows denote flow-ways included in analysis.

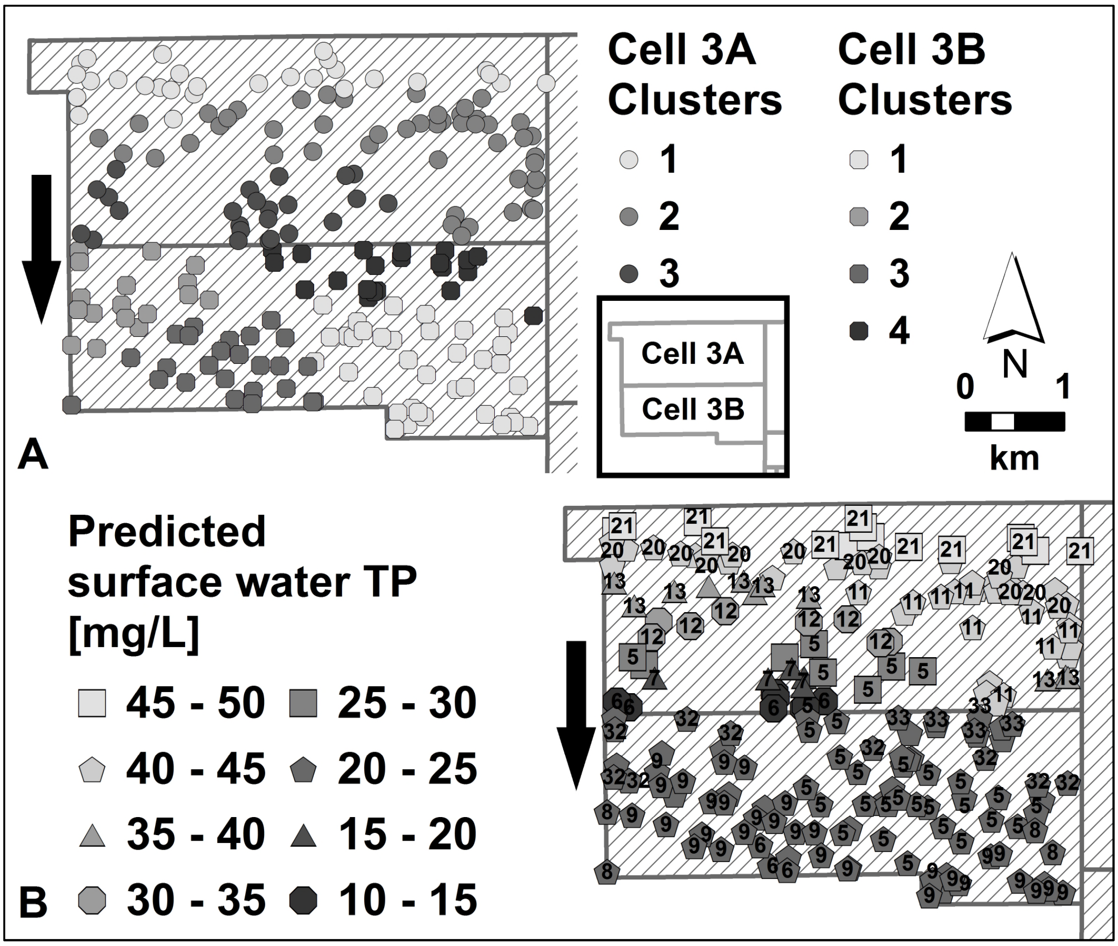


Figure 2: Spatial patterns detected by cluster (A) and CART (B) analyses – an example for STA-3/4 flow-way 3. Image B represents the distribution of CART nodes (symbol numbers represent the number of nodes in the CART model) corresponding to the prediction of surface water total P (concentration denoted by symbol color). Note that patterns are predominantly zonal and only tenuously aligned with flow direction.

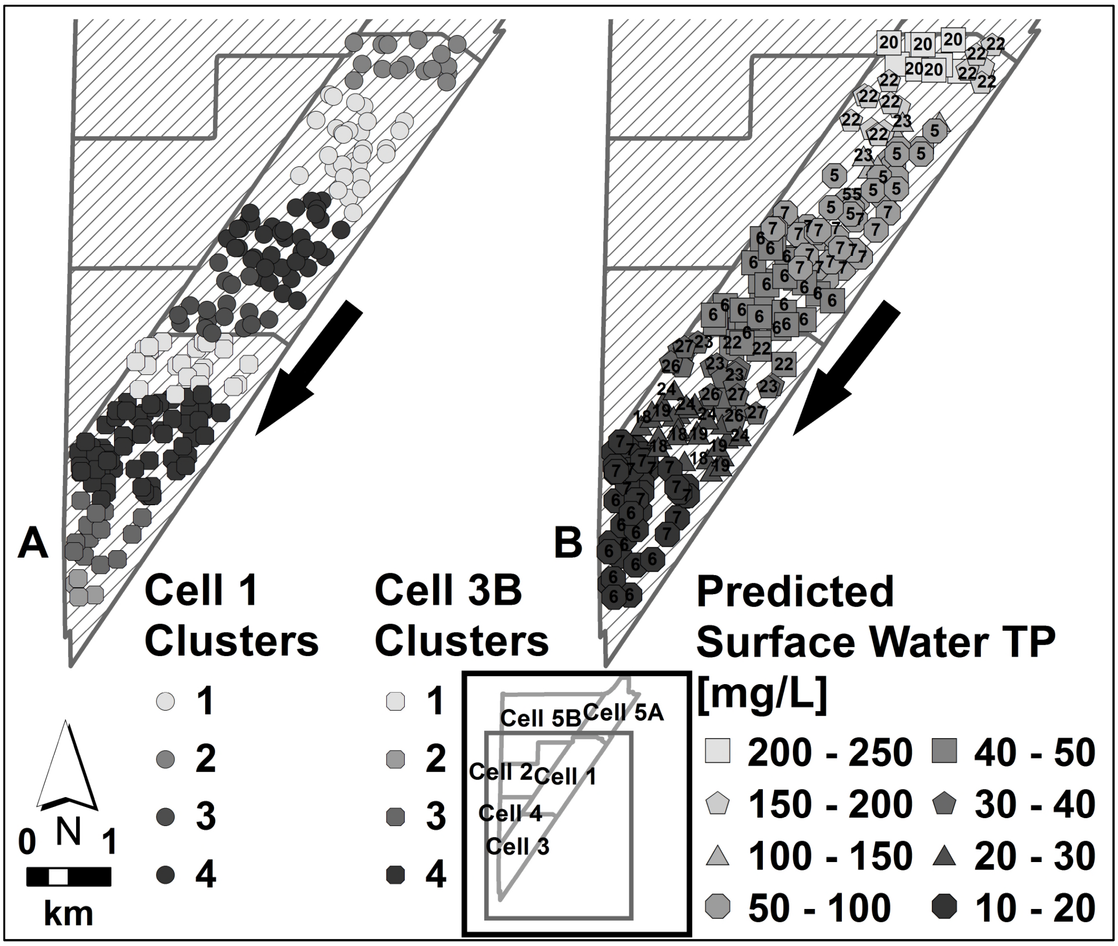


Figure 3: Spatial patterns detected by cluster (A) and CART (B) analyses – an example for STA-1W Eastern flow-way. Image B represents the distribution of CART nodes (symbol numbers represent the number of nodes in the CART model) corresponding to the prediction of surface water total P (concentration denoted by symbol color). Note that zonal patterns appear largely aligned with flow direction, indicating a gradient-based behavior to the individual zones.

A datamining approach to identifying spatial patterns of P forms in the Stormwater Treatment Areas in the Everglades, US

Corstanje, R., Grafius, D.R., Zawadzka, J., Moreira J., Vince, G., Ivanoff, D., Pietro, K.

Supplementary Materials

These maps show the K-mean clusters and tree nodes resulting from Classification and Regression Trees (CARTs) analysis performed within particular flow-ways of the Stormwater Treatment Areas (STAs) that had sufficient data to do so. Please note that, in the case of CARTs, the results are only shown for the flow-ways with availability of data on total surface water phosphorus.

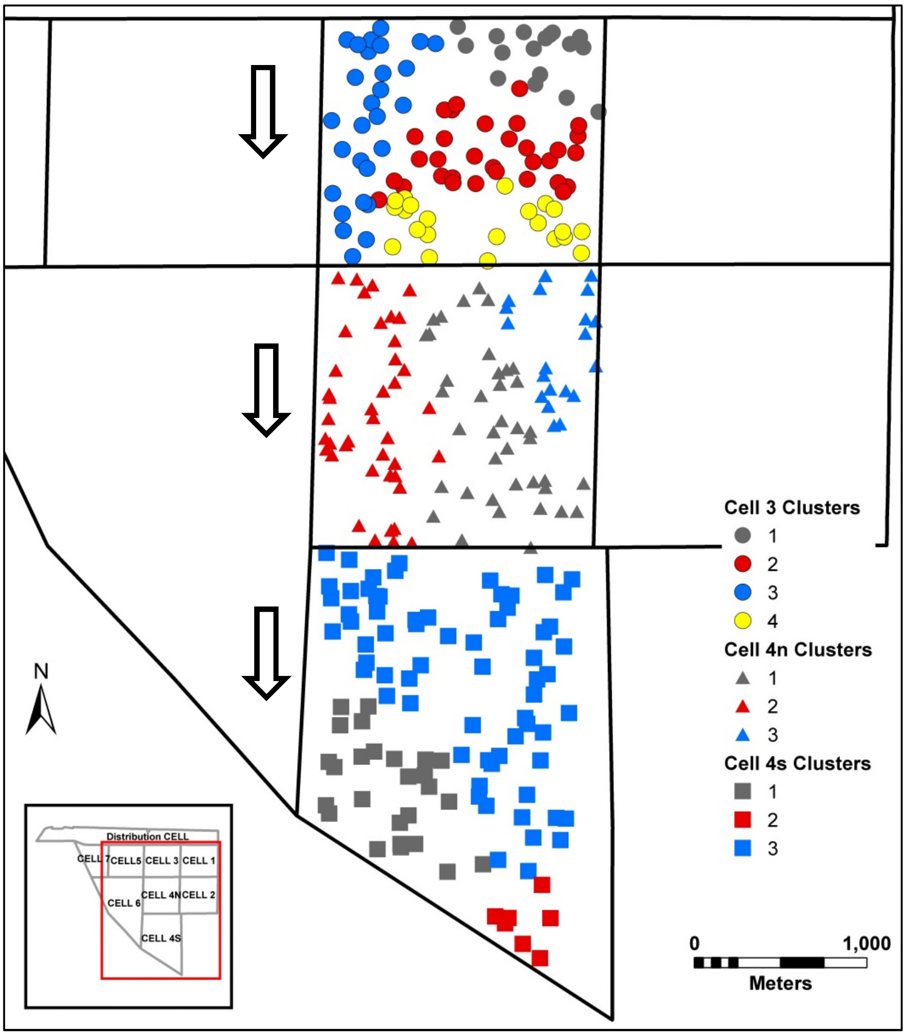
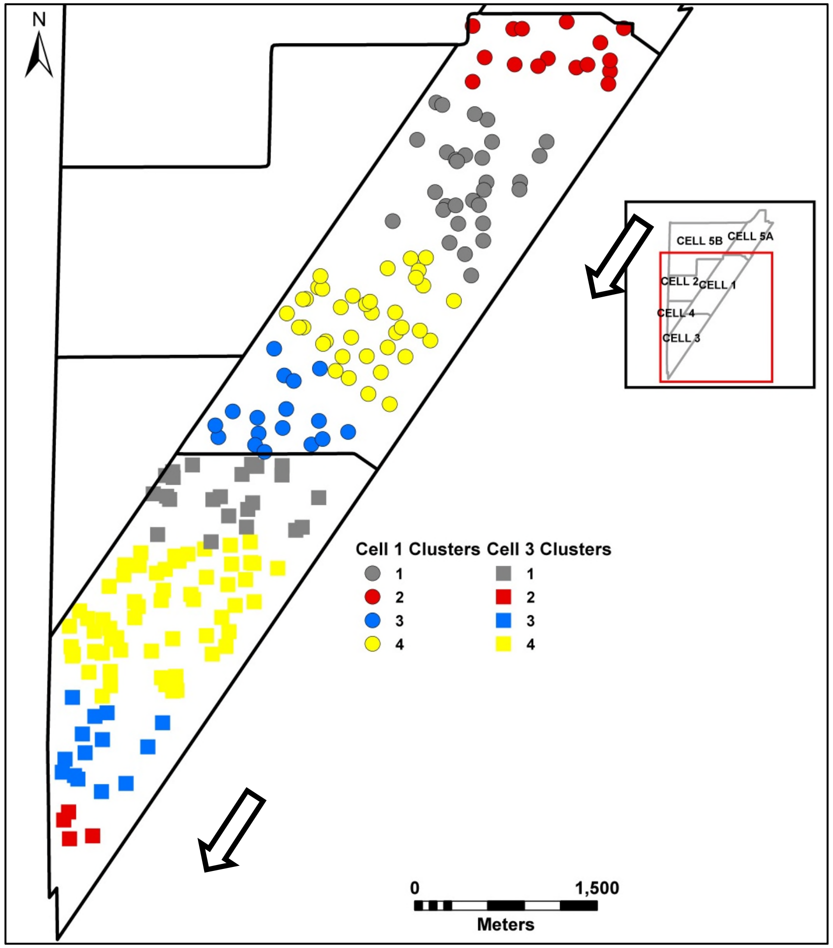
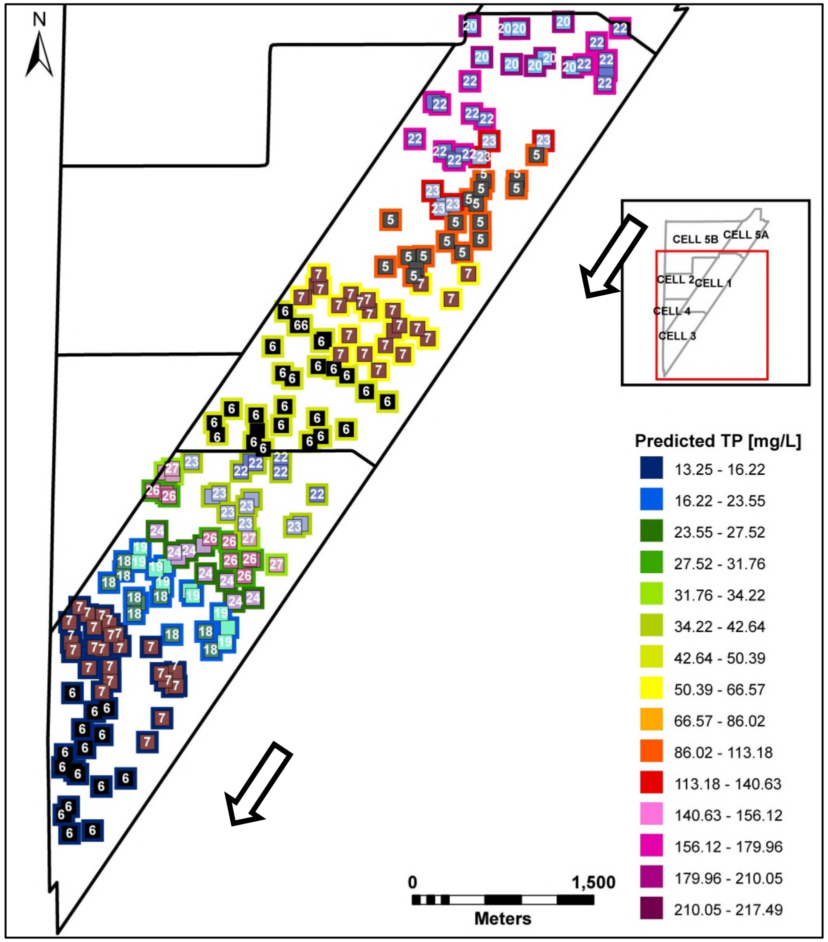


Figure S1: Cluster analysis for STA-1E Central flow-way. Arrows indicate the direction of water flow through the flow-way.

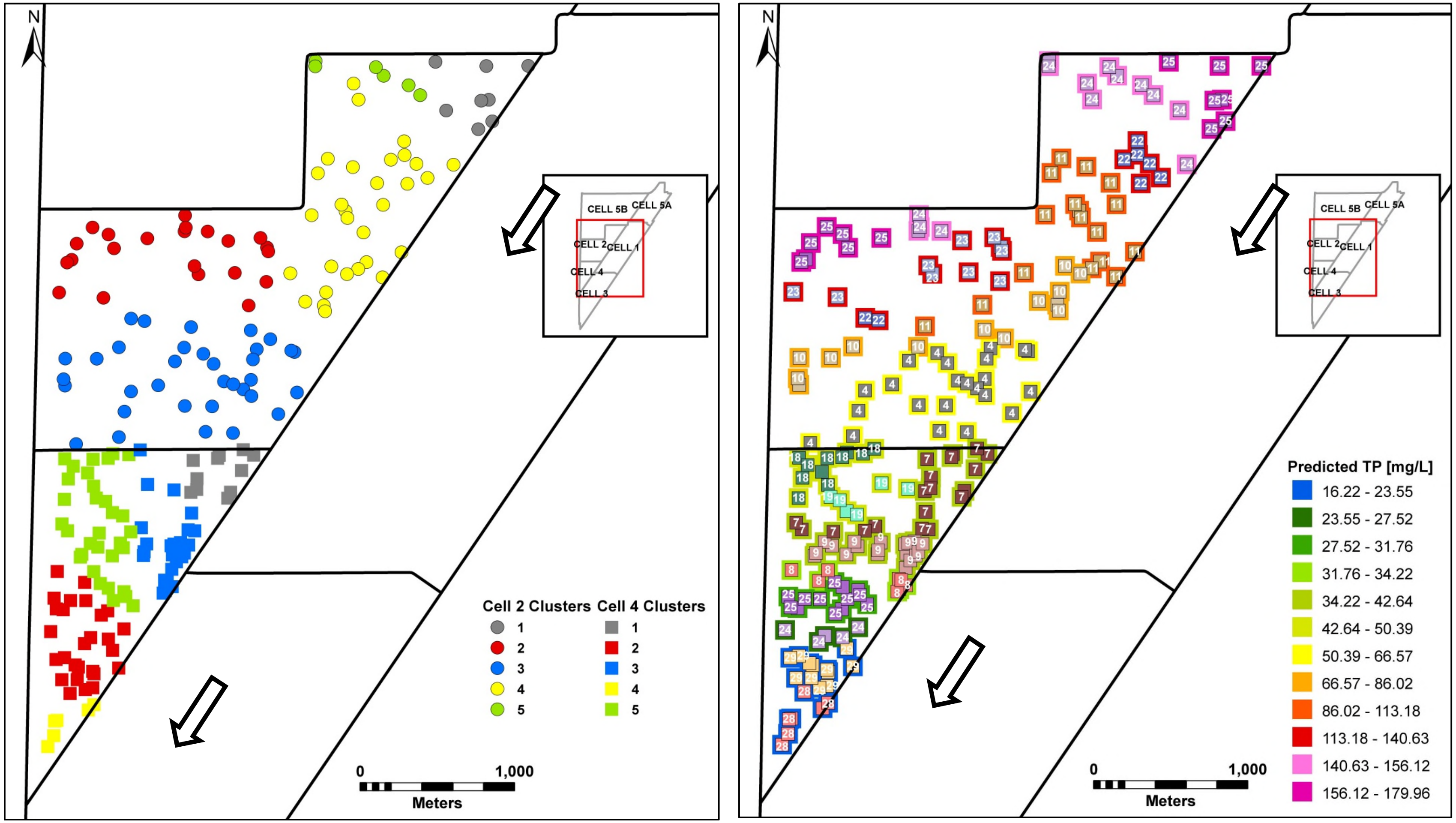


A



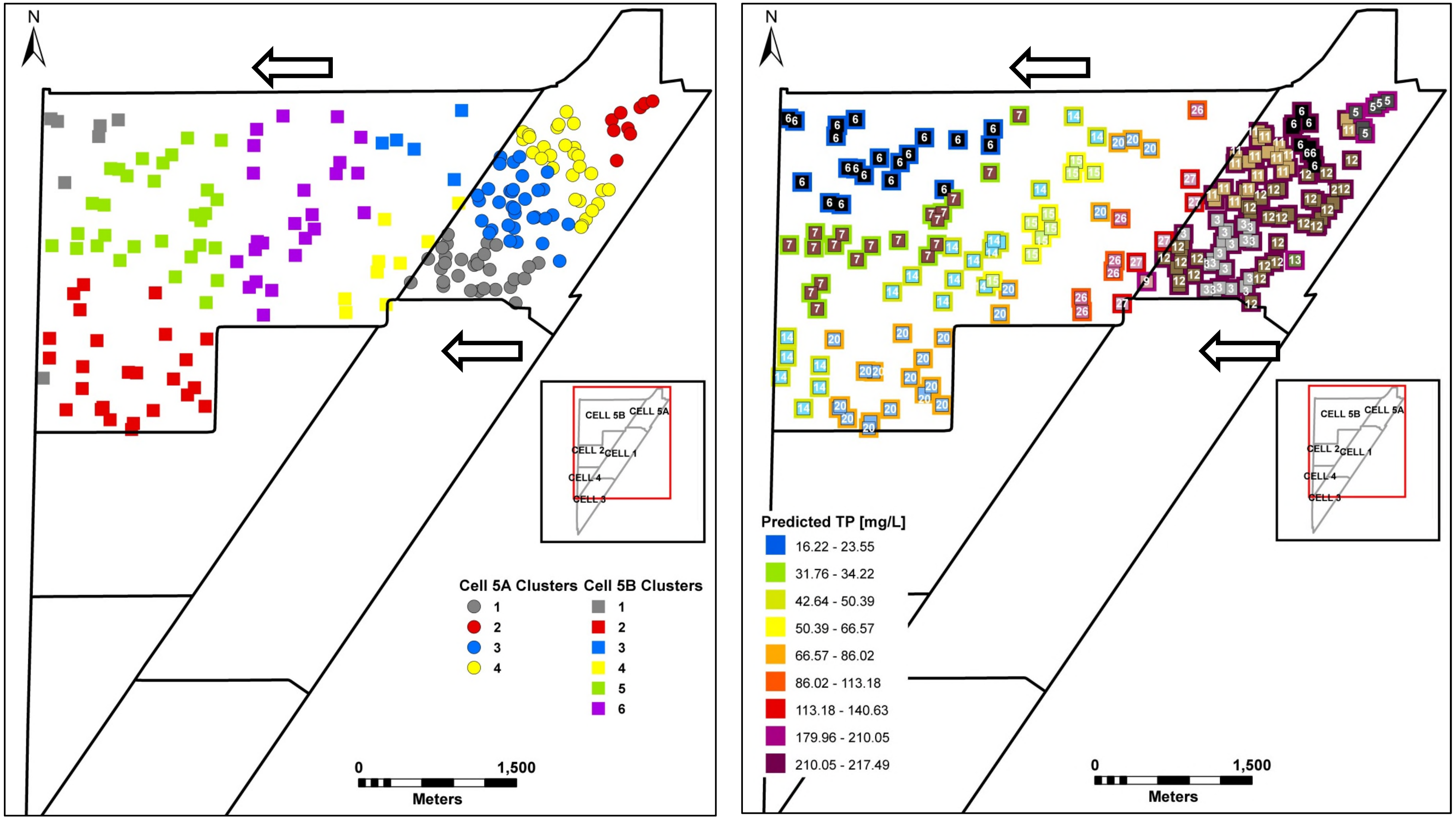
B

Figure S2: Results of A – cluster analysis, and B – CARTs analysis for STA-1W Eastern flow-way. Arrows indicate the direction of water flow through the flow-way. Numbers in CART results indicate the number of nodes in the CART model.



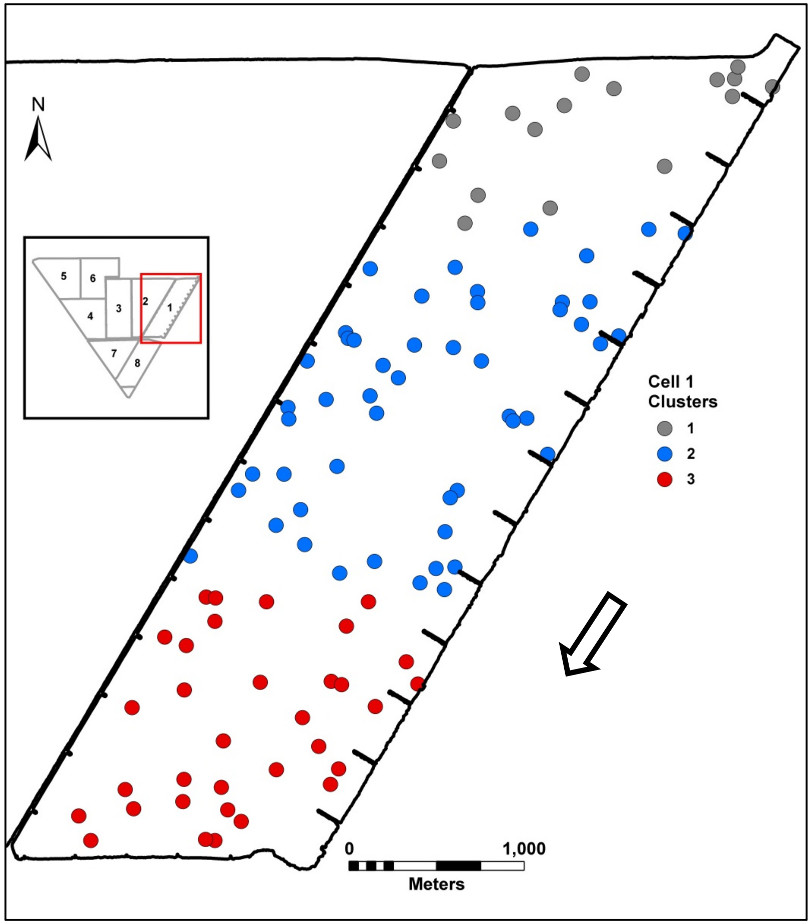
A **B**

Figure S2: Results of A – cluster analysis, and B – CARTs analysis for STA-1W Western flow-way. Arrows indicate the direction of water flow through the flow-way. Numbers in CART results indicate the number of nodes in the CART model.

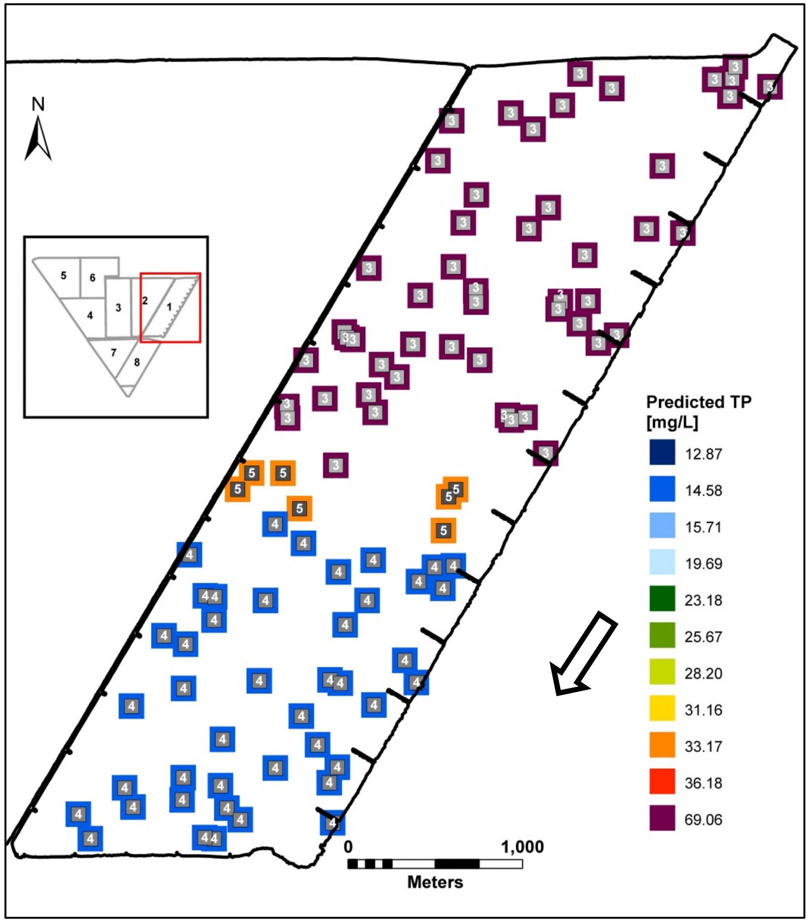


A **B**

Figure S3: Results of A – cluster analysis, and B – CARTs analysis for STA-1W Northern flow-way. Arrows indicate the direction of water flow through the flow-way. Numbers in CART results indicate the number of nodes in the CART model.



A



B

Figure S4: Results of A – cluster analysis, and B – CARTs analysis for STA-2 flow-way 1. Arrows indicate the direction of water flow through the flow-way. Numbers in CART results indicate the number of nodes in the CART model.

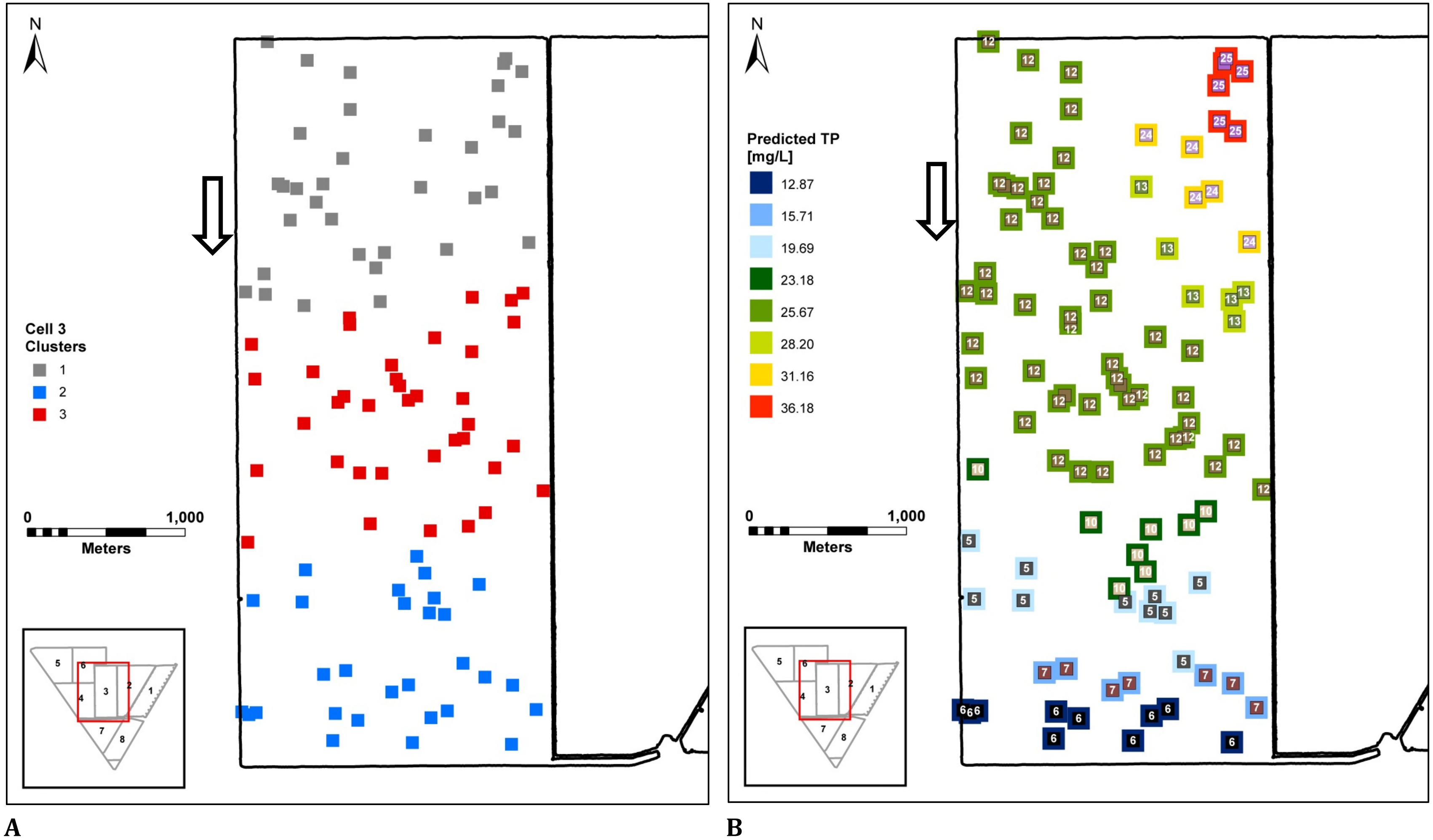


Figure S5: Results of A – cluster analysis, and B – CARTs analysis for STA-2 flow-way 3. Arrows indicate the direction of water flow through the flow-way. Numbers in CART results indicate the number of nodes in the CART model.

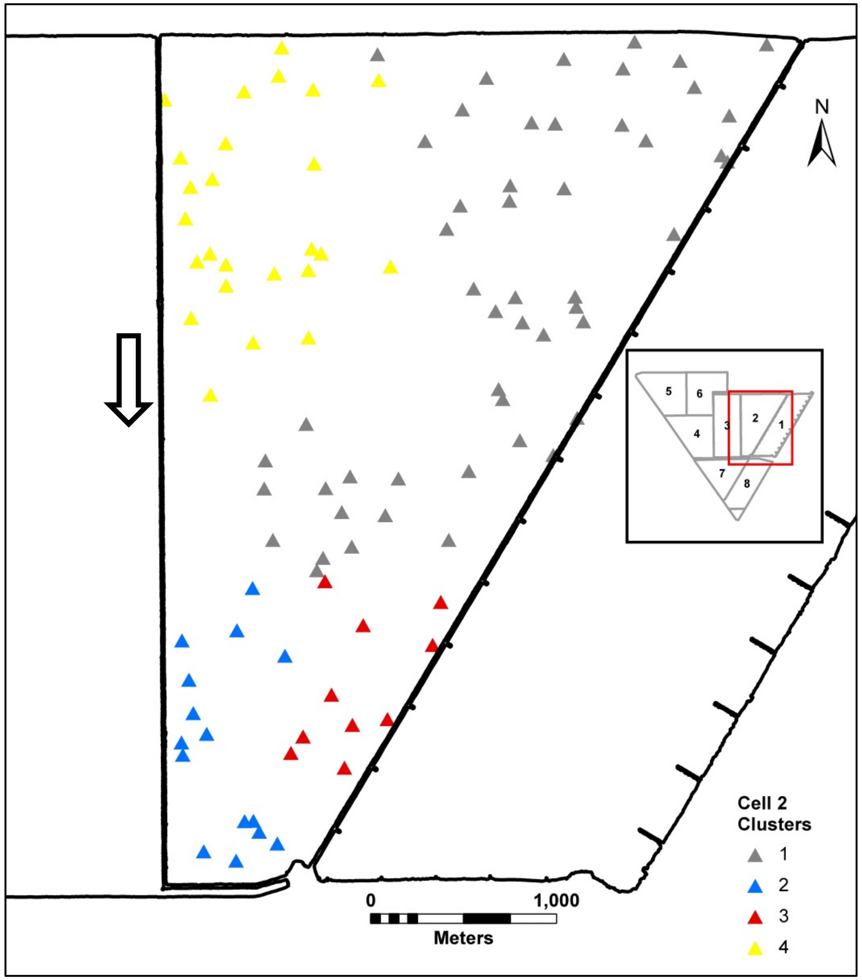
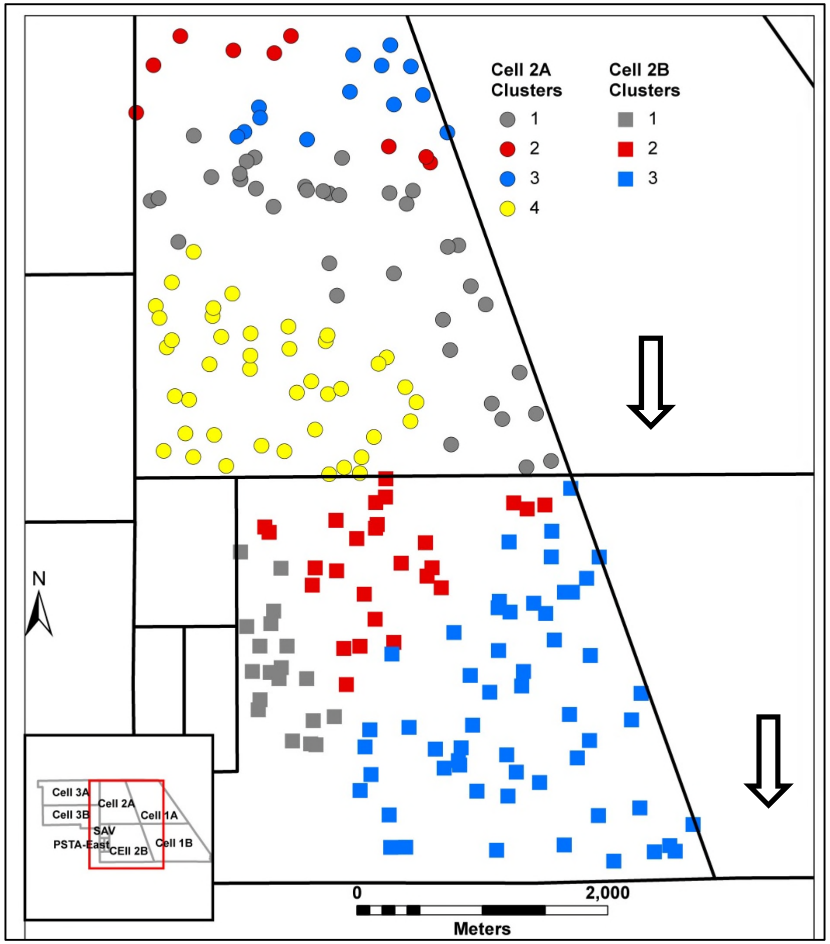
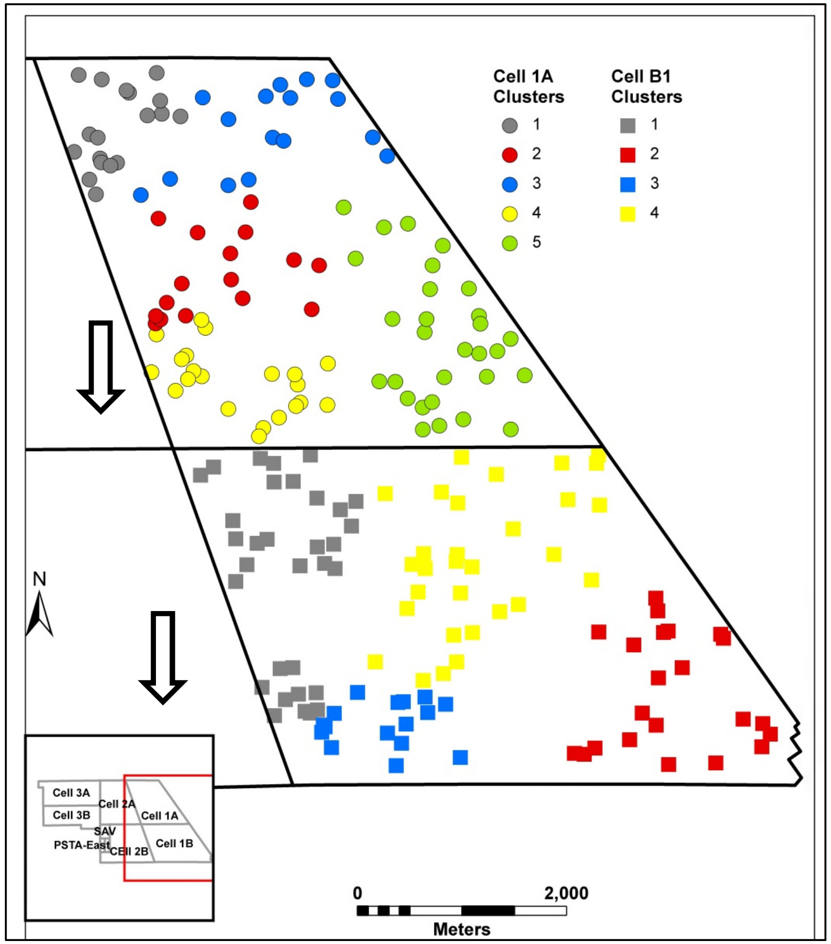


Figure S6: Cluster analysis for STA-2 flow-way 2. Arrows indicate the direction of water flow through the flow-way.



A



B

Figure S7: Results of cluster analysis for A – STA-3/4 flow-way 2 and B – STA-3/4 flow-way 1. Arrows indicate the direction of water flow through the flow-way.

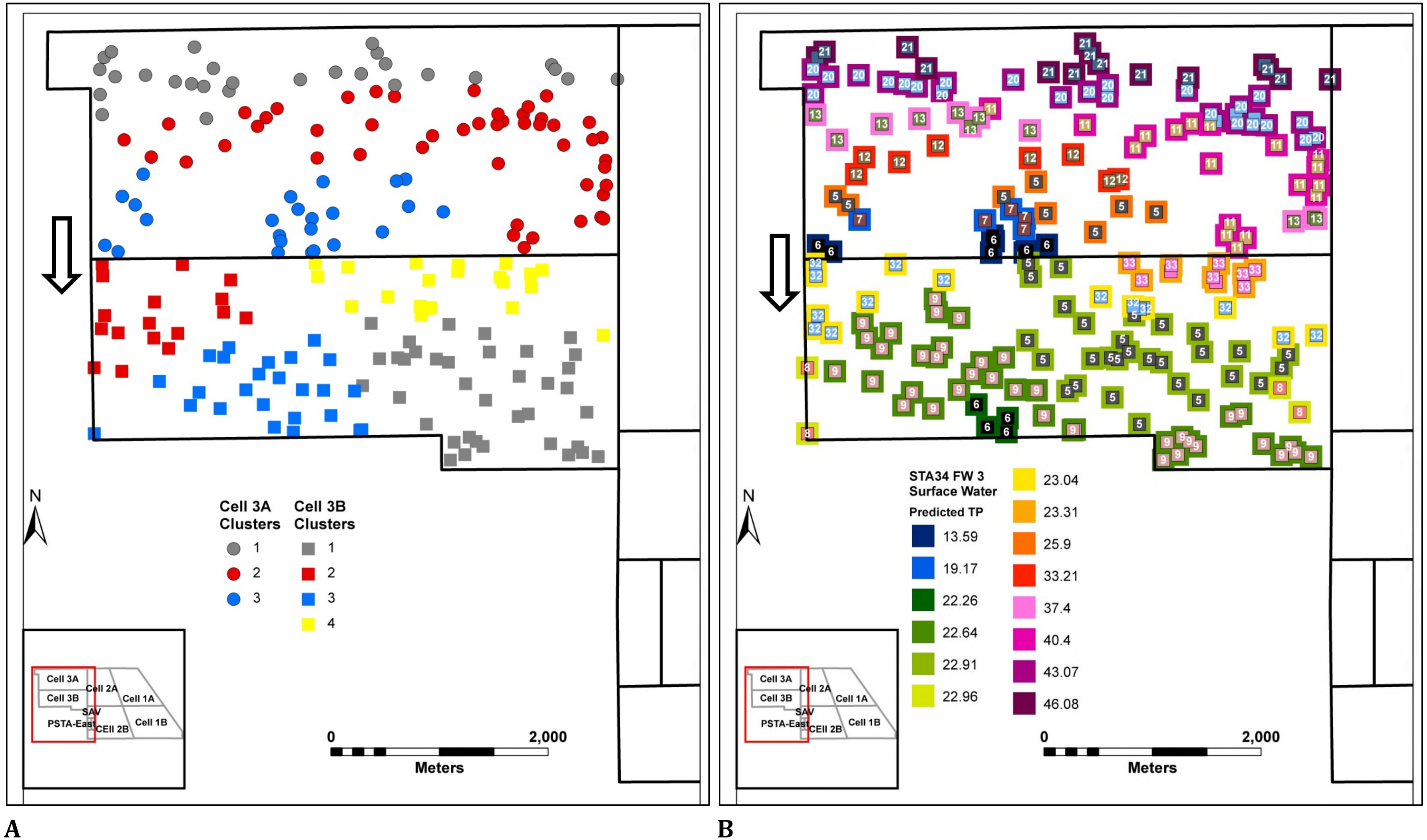
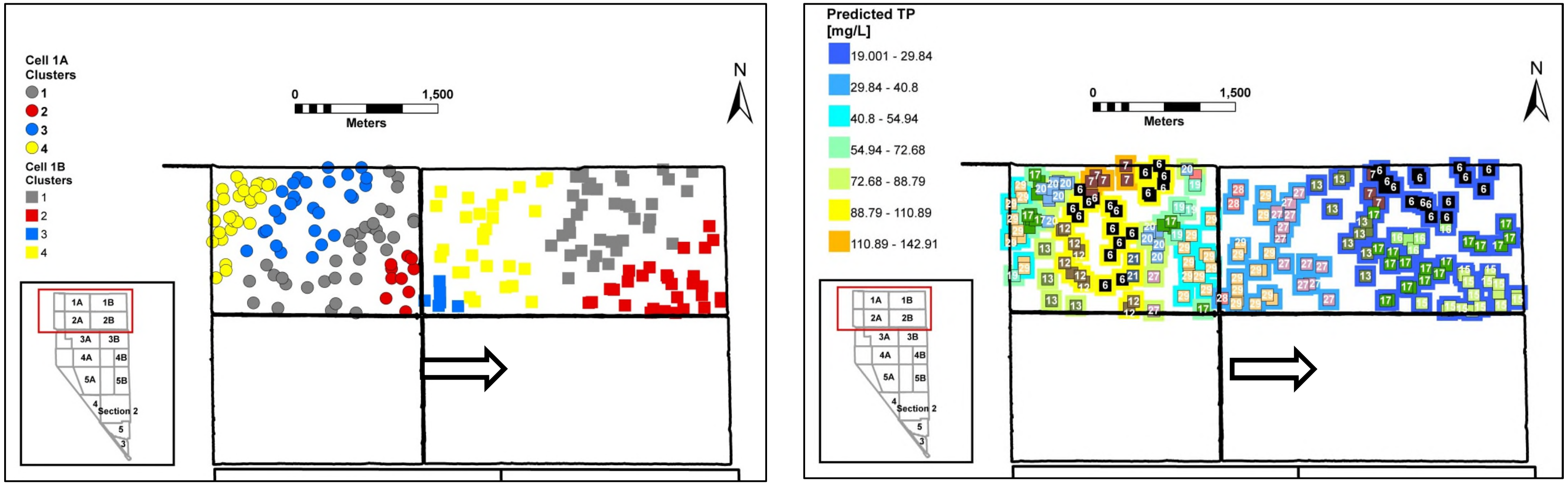


Figure S8: Results of A – cluster analysis, and B – CARTs analysis for STA-3/4 flow-way 3. Arrows indicate the direction of water flow through the flow-way. Numbers in CART results indicate the number of nodes in the CART model.



A **B**

Figure S9: Results of A – cluster analysis, and B – CARTs analysis for STA-5/6 flow-way 1. Arrows indicate the direction of water flow through the flow-way. Numbers in CART results indicate the number of nodes in the CART model.

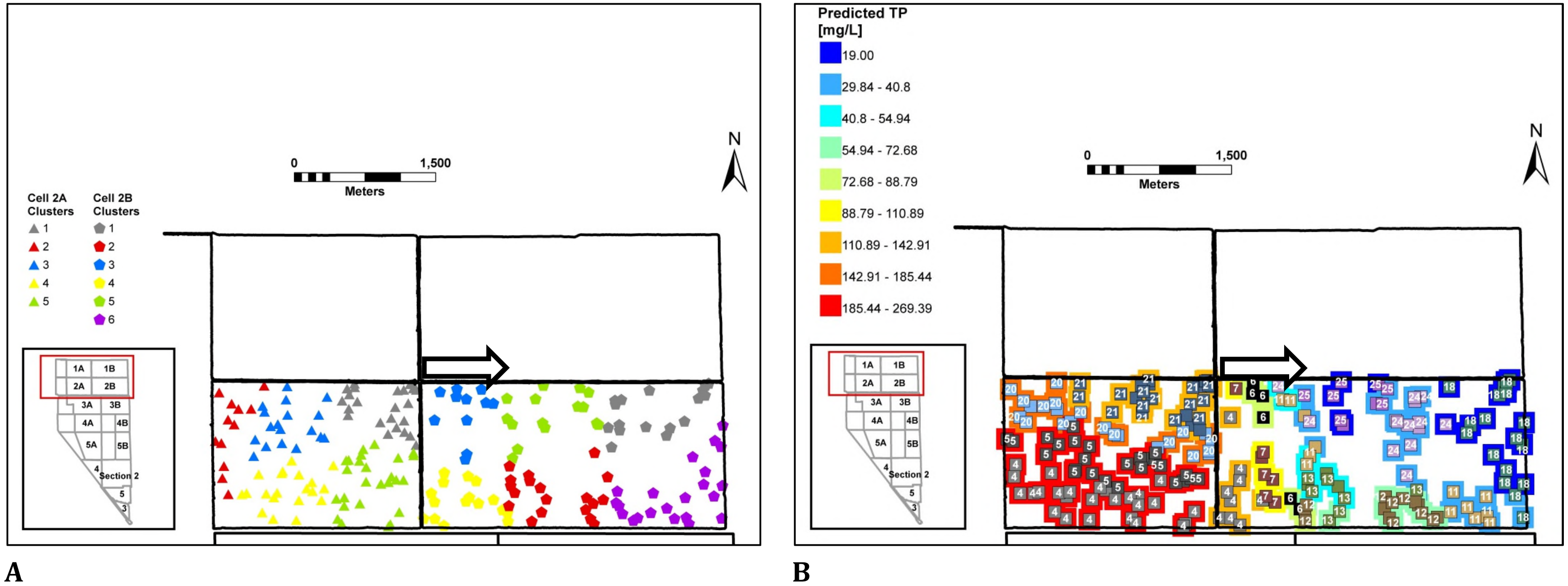
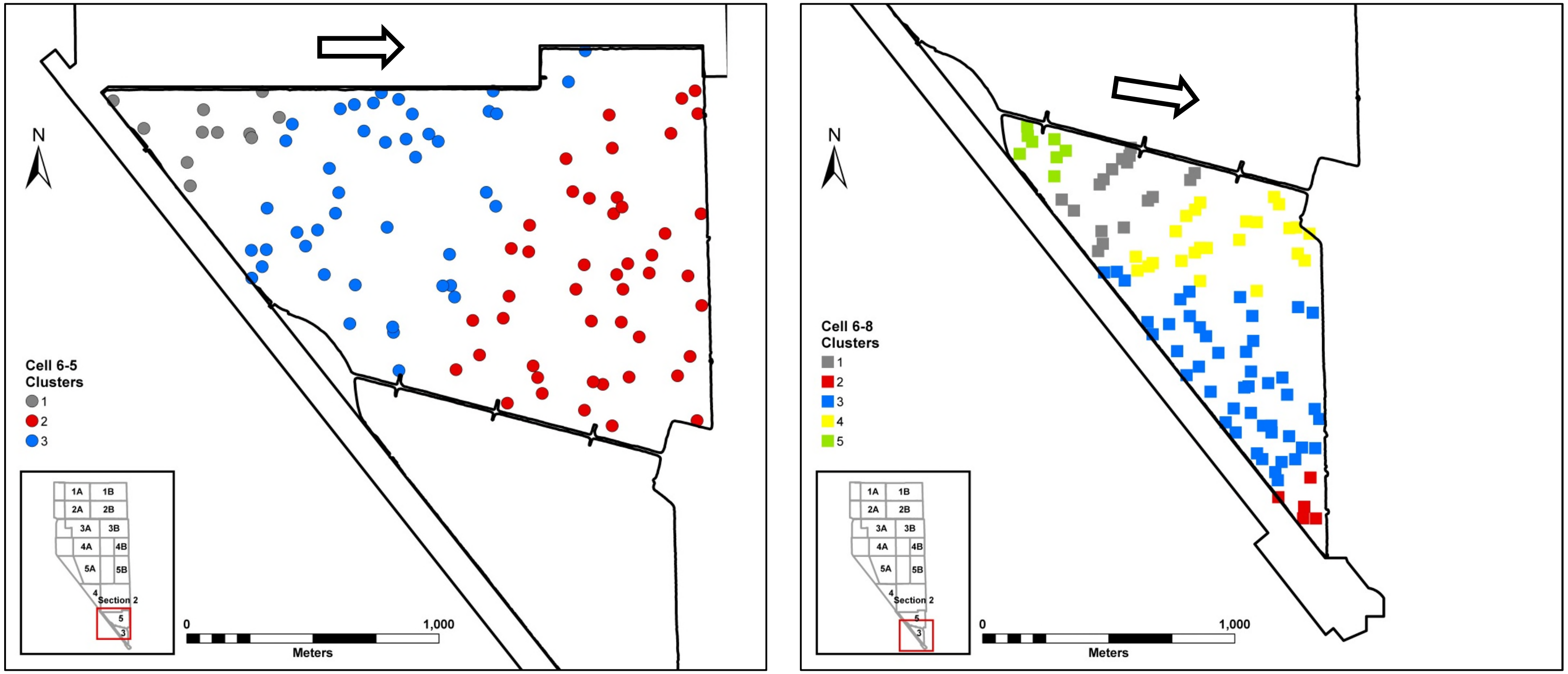


Figure S11: Results of A – cluster analysis, and B – CARTs analysis for STA-5/6 flow-way 2. Arrows indicate the direction of water flow through the flow-way. Numbers in CART results indicate the number of nodes in the CART model.



A **B**

Figure S10: Results of cluster analysis for A – STA-5/6 flow-way 7 and B – STA-5/6 flow-way 8. Arrows indicate the direction of water flow through the flow-way.

A datamining approach to identifying spatial patterns of phosphorus forms in the Stormwater Treatment Areas in the Everglades

Corstanje, Ronald

2016-12-01

R. Corstanje, D.R. Grafius, J. Zawadzka, J. Moreira Barradas, G. Vince, D. Ivanoff, K. Pietro, A datamining approach to identifying spatial patterns of phosphorus forms in the Stormwater Treatment Areas in the Everglades, US, Ecological Engineering, Volume 97, Issue December, 2016, pp.567-576

https://doi.org/10.1016/j.ecoleng.2016.10.003