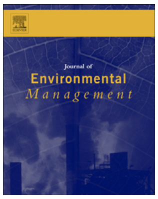
Journal of Environmental Management 188 (2017) 308–321

Contents lists available at ScienceDirect

Journal of Environmental Management

journal homepage: www.elsevier.com/locate/jenvman



Research article

Integrative analysis of the Lake Simcoe watershed (Ontario, Canada) as a socio-ecological system

Alex Neumann, Dong-Kyun Kim, Gurbir Perhar, George B. Arhonditsis*

Ecological Modelling Laboratory, Department of Physical & Environmental Sciences, University of Toronto, 1065 Military Trail, Toronto, Ontario M1C 1A4, Canada

ARTICLE INFO

Article history:
Received 29 March 2016
Received in revised form
23 November 2016
Accepted 27 November 2016

Keywords:
Self-organizing maps
Lake Simcoe
Sustainability
Urbanization
Environmental policy
Integrated watershed management

ABSTRACT

Striving for long-term sustainability in catchments dominated by human activities requires development of interdisciplinary research methods to account for the interplay between environmental concerns and socio-economic pressures. In this study, we present an integrative analysis of the Lake Simcoe watershed, Ontario, Canada, as viewed from the perspective of a socio-ecological system. Key features of our analysis are (i) the equally weighted consideration of environmental attributes with socioeconomic priorities and (ii) the identification of the minimal number of key socio-hydrological variables that should be included in a parsimonious watershed management framework, aiming to establish linkages between urbanization trends and nutrient export. Drawing parallels with the concept of Hydrological Response Units, we used Self-Organizing Mapping to delineate spatial organizations with similar socio-economic and environmental attributes, also referred to as Socio-Environmental Management Units (SEMUs). Our analysis provides evidence of two SEMUs with contrasting features, the “undisturbed” and “anthropogenically-influenced”, within the Lake Simcoe watershed. The “undisturbed” cluster occupies approximately half of the Lake Simcoe catchment (45%) and is characterized by low landscape diversity and low average population density <0.4 humans ha^{-1}. By contrast, the socio-environmental functional properties of the “anthropogenically-influenced” cluster highlight the likelihood of a stability loss in the long-run, as inferred from the distinct signature of urbanization activities on the tributary nutrient export, and the loss of subwatershed sensitivity to natural mechanisms that may ameliorate the degradation patterns. Our study also examines how the SEMU concept can augment the contemporary integrated watershed management practices and provides directions in order to promote environmental programs for lake conservation and to increase public awareness and engagement in stewardship initiatives.

1. Introduction

The worldwide degradation of shared natural resources, in general, and receiving water bodies, in particular, has led to the development of a concept related to their management, known as common-pool resources (CPR). Lakes and adjacent watersheds in the CPR concept are perceived as resources providing numerous services to multiple stakeholders with diverse background (Ostrom, 2008). CPR's shared status also implies that resource users belong both to present and future stakeholder generations, which underscores resource sustainability as one of the primary management goals (Mayer et al., 2014; Ostrom, 2009). In lake conservation, the integrated watershed management (IWM) practice is conceptually on par with a comprehensive CPR management framework - establishing a balance between the stakeholders' social and economic activities, while minimizing the environmental impact on CPR integrity (Lagutov, 2011). In human-dominated catchments, which are heavily affected by intensive farming or urbanization, achieving long-term ecosystem sustainability requires the development of interdisciplinary research methods to account for long-term evolution of social and economic pressures (Bowen and Riley, 2003). Thus, restoration programs should not solely revolve around environmental perspectives, but should also consider human population dynamics, social and economic conditions, and cultural traditions (Bowen and Riley, 2003; Liu et al., 2007). However, the implementation of IWM strategies are driven by technocratic principles and often considers the socio-economic factors as a post-hoc feature, with more emphasis

---

* Corresponding author.
E-mail address: georgea@utsc.utoronto.ca (G.B. Arhonditsis).

http://dx.doi.org/10.1016/j.jenvman.2016.11.073

placed on the environmental aspects (Lagutov, 2011). The present study, by contrast, works on leveling off this misbalance and recommends an advance on IWM that considers the feasibility of incorporating social and economic inputs on a more level playing field with environmental science and technology.

The contemporary practice of investigating cause-effect relationships between anthropogenic stressors and undesirable ecosystem changes follows the Pressure-State-Response (PSR) methodology proposed by the Organization for Economic Co-operation and Development (OECD, 1993) (Fig. 1). Models in socioeconomics, hydrology, biogeochemistry, agriculture and ecology are an integral part of this framework aiming to replicate the real-world coupling of human and environmental subsystems within a catchment of interest. Nonetheless, the rising complexity of these decision-support systems (DSS) affects their ability to identify model parameters, accentuates the problem of model equifinality (Beven, 2006), and ultimately undermines the credibility of projections of management plan outcomes (Lagutov, 2011). This technical challenge underscores the demand to control model complexity and minimize sources of uncertainty in resource management (Polasky et al., 2011). From this point of view, a parsimonious modelling strategy gradually capitalizes upon advancements in understanding of CPR dynamics and ensures that any increase in model complexity adds explanatory power (Gudimov et al., 2012). Indeed, Walker and Salt (2006) showed that socio-environmental systems could be adequately described using a small number of key controlling variables. A characteristic example is the IPAT model, in which environmental impact/pollution (I) is dependent on driving forces of population (P), affluence (A), and technology (T) (Chertow, 2000). Likewise, our study attempts to identify a minimal number of controlling variables, directly or indirectly responsible for lake impairment, and then to associate them with forecasted rates of catchment urbanization.

In this study, we present an integrative analysis of the Lake Simcoe watershed, Ontario, Canada, as viewed from the perspective of a socio-ecological system or SES (Ostrom, 2009). Key features of our analysis are the equally weighted consideration of environmental values with socioeconomic parameters, and the identification of the minimal number of key variables that should be included in a parsimonious watershed management framework, aiming to establish linkages between urbanization trends and ecosystem response with respect to nutrient export. Finally, drawing parallels with the concept of Hydrological Response Units, our study uses Self-Organizing Mapping to delineate spatial organizations with similar socio-economic and environmental attributes.

2. Methods

2.1. Case study

Lake Simcoe is the largest inland dimictic lake in Southern Ontario (44.44°N, −79.34°W) with a watershed area of 3400 km^2 and a relatively small catchment-to-surface area ratio of 5:1 (Fig. 2). The lake as a popular fishing destination contributes ~$200M Canadian Dollars (or CAD) year^{-1} to the local economy. The lake sediments have revealed a historic record of continuous anthropogenic pressure since the first European immigrants settled in the lake catchment in the 17th century (Vovik, 2014). In the pre-settlement period, the lake had provided habitat for precious cold- and cool-water fish species, including lake sturgeon (*Acipenser fulvescens*) and lake trout (*Salvelinus namaycush*). With human population growth, the lake watershed experienced severe pressure arising from massive deforestation, followed by exponential urbanization after 1945, from 60,000 to (the current) 350,000 residents. The resultant watershed export of phosphorus (P) to the lake increased from the pre-settlement baseline of 27–33 T P year^{-1} to 300 T P year^{-1} in the 1960s, which triggered excessive algal growth and hypolimnetic oxygen depletion, lethal for cold-water fish species. As an indicator of the severity of lake ecosystem impairment, the sturgeon fish population experienced a collapse, from 62,000 kg of commercial catch in the 1880s, to a total disappearance by 1956 (OMNR, 2009). Similarly, lake trout reproduction capacity has been affected due to hypolimnetic hypoxia,

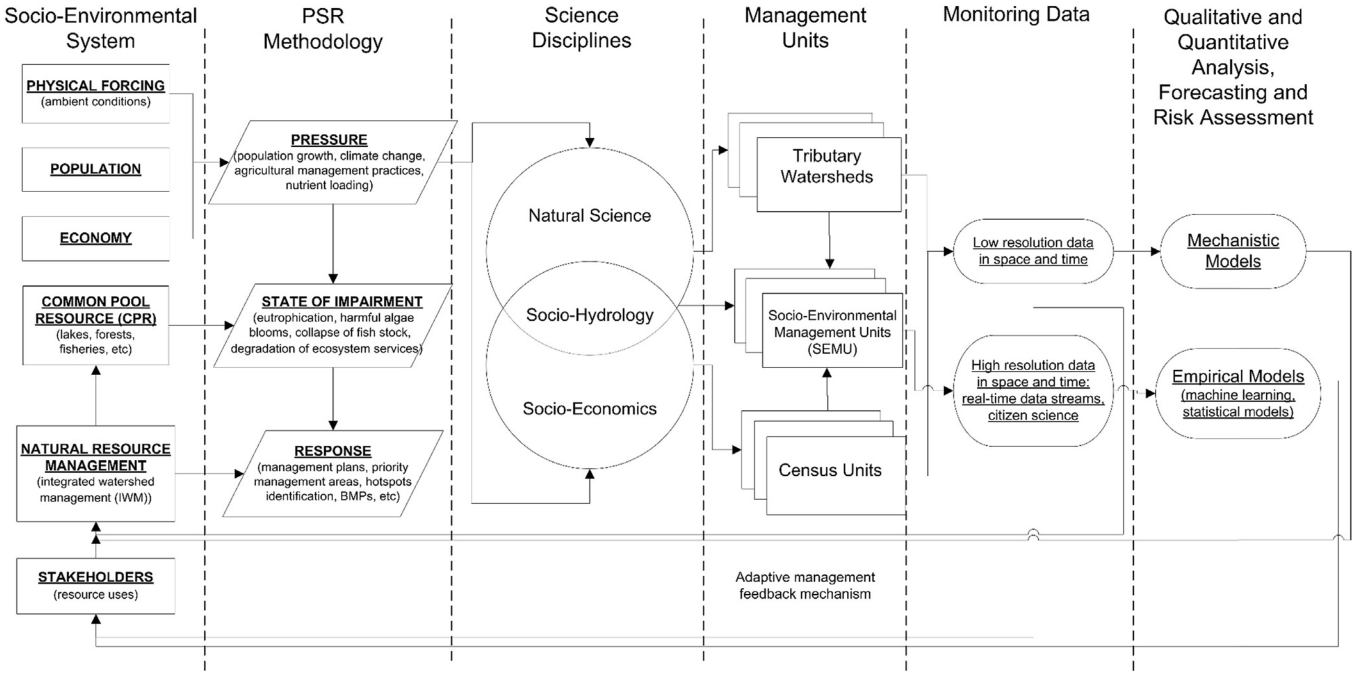


Fig. 1. Conceptual diagram of the Lake Simcoe watershed studied as a socio-environmental system.

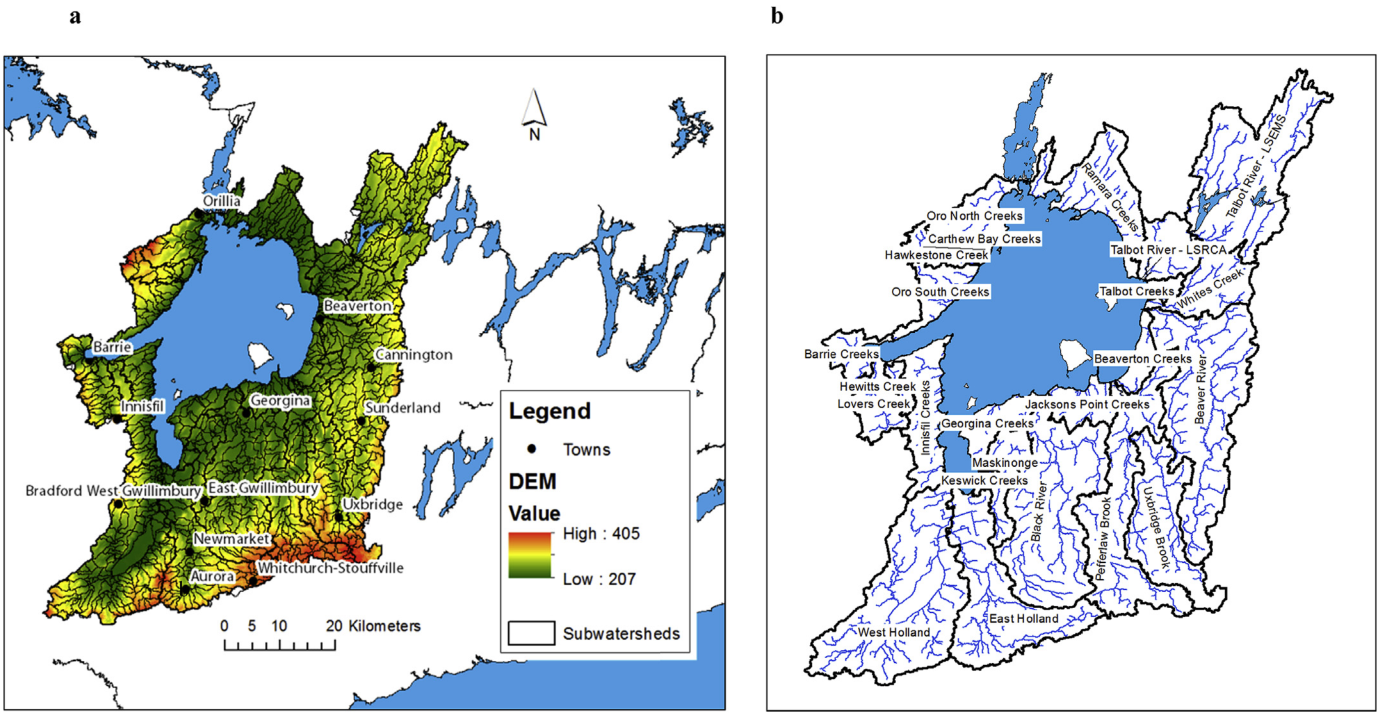


Fig. 2. (*a*) Lake Simcoe basin subdivided in 824 sub-watersheds with a mean area of 350 ha; (*b*) Lake Simcoe basin subdivided into sub-catchments based on pour points of the main tributaries. The sub-catchments are defined as the land draining to a tributary of the main watercourse.

such that its current population has to be artificially maintained through stocking with hatchery-reared yearlings (Palmer et al., 2011). Recent P loading to the lake has varied between 72 and 115 T P year^{-1} in 2007–2011 (LSRCA, 2013; MOECC, 2014), compared to a protection plan reduction goal of 44 T P year^{-1} by 2045 (MOECC, 2010). The Canadian government has invested $58.5M CAD in 2007–2012 in P reduction, with an additional $29M allocated until 2017. Complicating the issue of the lake restoration program, projected population growth in the catchment of 30 percent over the next 20 years raises concerns as to whether the Lake Simcoe watershed (LSW) can sustain this urbanization rate, meet and maintain the P reduction target in future, thereby returning to a sustainable cold-water fish habitat (Jin et al., 2013). In this context, our study pursues an interdisciplinary analysis of the LSW, explicitly accommodating the non-linear nature of hydrological-human interactions.

2.2. Environmental data

In this analysis, we examined the Lake Simcoe watershed as a system of hydrologically connected 824 subwatersheds with a mean area of 350 ha (~0.1% of total basin). We imported and “burned-in” the stream network layer from the Ontario integrated hydrology dataset into the Provincial Digital Elevation Model (DEM; 10 × 10 m resolution) to account for flat topography and multiple braided channels in agricultural plain polders (Luo et al., 2011). We acquired the soil characteristics from Agriculture and Agri-Food Canada's Detailed Soil Survey (1:50,000; http://sis.agr.gc.ca/cansis/nsdb/dss/index.html) for the upper soil layer, with an average depth (SOL_Z) of 262 mm. The list of soil characteristics included moist bulk density (SOL_BD, g cm^{-3}), available water capacity (SOL_AWC, mm H_2O mm^{-1} soil), saturated hydraulic conductivity (SOL_K, mm hr^{-1}), organic carbon content (SOL_CBN, % soil weight), and moist soil albedo (SOL_ALB, ratio of solar radiation reflectance).

For the land use/land cover (LULC) information, we relied on the Lake Simcoe Watershed Land Cover dataset (2013), which required reclassification from codes of the provincial Ecological Land Classification (Lee, 1998) to definitions of the Phosphorus Budget Tool (MOECC, 2012). The latter document offered regionalized P export coefficients (E_i, kg P ha^{-1} year^{-1}), which were used to estimate Baseline Annual Phosphorus Tributary Loading (L_{BAPTL}, kg P year^{-1}) at pour points into the lake by integrating E_i coefficients over upstream subwatershed areas A_i (ha) occupied by a specific LULC type:

$$L_{BAPTL} = \sum E_i A_i$$

where L_{BAPTL} values were estimated twice (*i*) by integrating fine-resolution LULC information; (*ii*) by considering only a dominant LULC; i.e., the most common LULC within each subwatershed, which allowed us to identify the optimal LULC characterization strategy. Shannon diversity index (H-LU) was used to determine landscape diversity (or potential patchiness) on a subwatershed scale (Eiden et al., 2000).

2.3. Socio-demographic data

Census records for the 1996–2006 period were extracted from Statistics Canada (www.12.statcan.gc.ca) for the smallest spatial resolution of data collection; namely, geographic areas with 400–700 persons in the 2006 census (dissemination areas, DA) and 125–650 dwellings in the 1996 census (enumeration areas, EA). We used socio-economic variables from the IPAT model with total population counts and average family income (in CAD). Examination of the social dynamics of the studied system was also based on the number of immigrants (cumulative for out-of-province and out-of-country over the past 15 years), age-group demographics, total labour-force counts by industry for the age group 15 years and older, average value of dwellings (CAD), and number of houses per different types of dwelling structures. The subwatershed census data were aggregated from census to subwatershed boundaries, in proportion to areas of DA or EA within each subwatershed.

2.4. Data reduction and clustering

We applied Self-Organizing Map (SOM) analysis as an amalgamation algorithm to characterize dominant patterns of data by synthesizing multidimensional socio-environmental variables. SOM is an efficient un-supervised artificial neural-network algorithm to identify and visualize patterns in multidimensional datasets (Kohonen, 2013). SOM analysis provides more flexibility to analyze non-linear dependencies in complex systems, compared to the linear limitations of standard statistical tools, such as Principal Component Analysis (Giraudel and Lek, 2001; Frey and Rusch, 2013). SOM framework has been widely applied to environmental assessment and ecosystem analysis exercises (Chon, 2011; Ha et al., 2015), as well as to demographics and political science research (Skupin and Hagelman, 2005; Niemelä and Honkela, 2009). In this paper, the multivariate data ordination was mapped onto a two-dimensional lattice, whereby subwatersheds with similar socio-environmental characteristics occupied adjacent cells after model training. Subwatershed characteristics for soil, LULC, nutrient export, and socioeconomic parameters comprised the 21×824 ($n \times N = input\ variable \times data\ sample$) size of the SOM input space. Land-use inputs were defined in relative terms as a percentage contribution of a specific LULC in a subwatershed area. The nutrient export parameters were normalized based on the catchment area to eliminate the effect of subwatershed size. All the input data have been log-transformed. SOM output layer was arranged by a hexagonal lattice, rather than by a rectangular one, as it impartially considered the distances to the neighboring neurons (Kohonen, 2001). We opted for 135 lattice cells to be close to $5\sqrt{N}$, according to the recommendations of Vesanto and Alhoniemi (2000). The map size was optimized (9×15 cells) by minimizing both quantization and topological errors (Park et al., 2014). Similar to a previous watershed analysis by Kim et al. (2016), we applied a Gaussian neighborhood function. The weight vector was linearly initialized using its greatest eigenvectors of the covariance matrix of training data. We also applied a post-hoc hierarchical cluster analysis to summarize the dominant patterns of data more quantitatively using weight vectors of U-matrix (Kohonen, 2013). As a result, SOM served to downsize the complexity of the studied SES by assigning subwatersheds with overall similar socio-environmental characteristics to neighboring cells, in contrast to subwatersheds with the greatest differentiation, which were separated by the largest distances in a 2D lattice. The SOM analysis was performed in a SOM Toolbox 2.0 for Matlab 5.0 (www.cis.hut.fi/somtoolbox; Vesanto et al., 2000).

2.5. Statistical modelling for phosphorus export

We used structural equation modelling (SEM) to delineate the key causal relationships underlying the interplay among nutrient export at pour points in Lake Simcoe and socio-environmental control variables (i.e., socio-economic census data, LULC, subwatershed latitude, and landscape diversity coefficient) from upstream watersheds within the identified SOM clusters. SEM is a multivariate statistical method that encompasses both factor and path analysis, which allows decomposing multiple causal pathways and quantifying direct and indirect relationships among variables (Arhonditsis et al., 2006; Ullman, 2006; Arbuckle, 2013). Another advantage of SEM is that it can explicitly incorporate uncertainty due to measurement error and/or accommodate the discrepancy between conceptual ecosystem properties and observed variables that can be directly measured. SEM is also an a priori statistical method whereby a hypothetical structure of the system studied, reflecting the best knowledge available, is tested against the observed covariance structure. Drawing parallels with the reverse-engineering approach (Lobo and Levin, 2015), we tested multiple a priori models until an optimal fit was achieved with minimal residuals when comparing hypothesized and observed covariance structures (Arhonditsis et al., 2006). Box-Cox power transformations were implemented to stabilize variance of data and effectively linearize the bivariate relationships examined within the SEM structure (Box and Cox, 1964).

3. Results

3.1. Self-organized map analysis

SOM maps enable the comparison of the spatial clustering of subwatersheds based on different watershed input parameters (Fig. 3a). For example, subwatersheds with rich organic carbon in the top layer (SOL_CBN) became concentrated in the upper-left corner of the grid lattice. The same aggregation pattern was found with wetland-dominated subwatersheds (WETLAND) characterized by minimal patchiness (H-LU), while a negative correlation exists with soil bulk density and albedo (SOL_BD and SOL_ALB). Watersheds with steeper slopes (SLOPE) also associated with soils of light colors and high irradiation reflectance (SOL_ALB), low available water capacity (SOL_AWC), and high values of saturated hydraulic conductivity (SOL_K); all of which suggest soil instability consistent with a history of intense erosional processes (Howard et al., 1995). Areas with low values of saturated hydraulic conductivity partly overlap with high density residential areas (RESIDEN-HD). This combination typically exacerbates the intensity of surface run-off, thereby increasing nutrient export (P_EXPORT) from urbanized watersheds (Scalenghe and Marsan, 2009). Soils with higher-than-average available water capacity (SOL_AWC) only partially overlap with cropland and hay/pasture locations, indicative of potential vulnerability to water stress in subwatersheds with high quality agricultural soils during summer (Harrison et al., 2014).

Anthropogenically transformed landscapes, primarily located at the bottom of the SOM lattice, overlap with areas of high intensity of socio-economic attributes. In particular, subwatersheds dominated by high-density residential areas, lower left corner of RESIDEN-HD map on Fig. 3a, correspond to high population records (POP). COM/IND cells representing commercial/industrial areas in terms of similarity of their socio-environmental parameters occupy the same cells as highly urbanized areas in the SOM lattice. Interestingly, both urbanized LULC types (RESIDEN-HD and COM/IND) are characterized by relatively low landscape fragmentation (H-LU index), which bears resemblance to the single-use, homogeneous land use features of the urban sprawl in Toronto region (Hess and Sorensen, 2015). Finally, several residential low-density subwatershed nodes (URB-LD), located in the right part of SOM lattice, partially overlap with areas of protected natural heritage woodlands (FOREST in the lower right corner), areas with high slope and the highest family-income. These subwatersheds meet the criteria for Toronto exurbia, defined as areas beyond the suburbs of the city, inhabited by wealthy professionals and executives who seek natural “wilderness” and equestrian life styles, in contrast to middle-class suburban areas (Cadieux and Taylor, 2013).

Based on second-level SOM training, we have identified two distinct SES groups (Fig. 3b–c), each consisting of four socio-environmental subwatershed clusters (Table 1). The first group, consisting of clusters #1,4,5, and 8 (Table 1), is primarily occupied by significant portions of woodland (10–63%), permanently or seasonally flooded wetlands (15–47%), and reservoirs

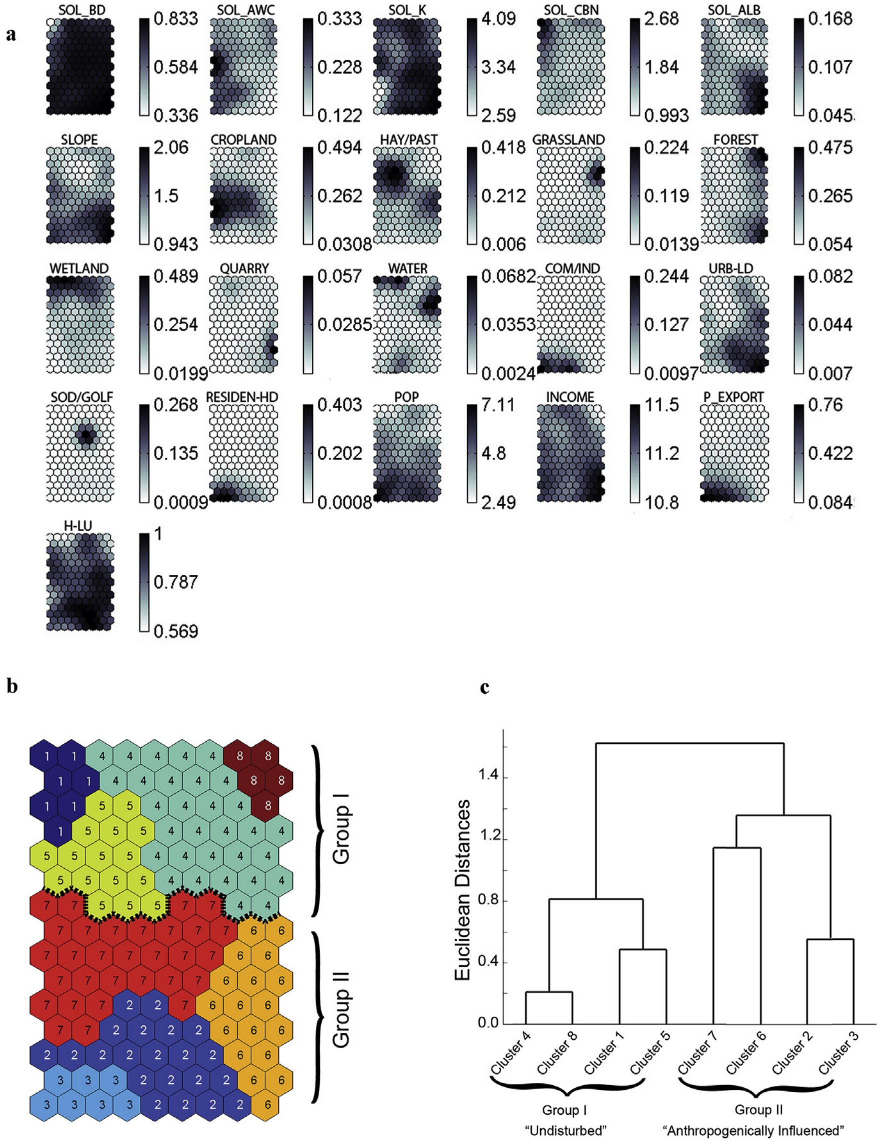


Fig. 3. Self-organizing maps (SOM) for Lake Simcoe watershed parameters in log-transformed scale, where SOL_BD - soil bulk density (g cm^{-3}); SOL_AWC - available water capacity of the top soil layer (mm H_2O mm^{-1} soil); SOL_K - saturated hydraulic conductivity (mm hr^{-1}); SOL_CNB - organic carbon content (% soil weight); SOL_ALB - moist soil albedo; SLOPE - slope; CROPLAND - various crop types (henceforth LULC as % of subwatershed); HAY/PAST - hay and pasture areas; GRASSLAND - grassland areas; FOREST - all types of forests; WETLAND - wetland areas including bogs and marsh areas; QUARRY - excavated quarry areas; WATER - all water bodies including lakes and ponds; COM/IND - commercial and industrial areas; URB-LD - urban low density areas; SOD/GOLF - sod farms and golf courses; RESIDEN-HD - residential high density areas; POP - population census (2006); INCOME - family income before taxes (2005); P_EXPORT - phosphorus export coefficient (kg P ha^{-1} year^{-1}); H-LU - Shannon index of land use diversity; b) resultant map of 8 SOM clusters; c) dendrogram of SOM clusters based on Euclidean distances.

characterized by stagnant waters (0.9–4.6%). Furthermore, this group reflects the low end of population density (<0.4 humans ha^{-1}), modest land use diversity (H-LU = 0.9–1.3), highest organic carbon content in the top soil (3.3–12.0%), and below-average family income in the context of the Lake Simcoe watershed ($52,700-$79,000). Alongside with their common characteristics, there is also a unique mix of cluster-specific features. Namely, subwatersheds in cluster #4 represent a mix of wetland (36%), forest (19%), farmland (18.5%), and grassland (8%), with the highest proportion of land allocated to sod farms and golf clubs (7.7%). Cluster #4 also has relatively extensive urban, low-density territories (3.4%), and highest family income ($79,000) in the group, which can be important for considering stewardship initiatives and implementation of best management practices. Cluster #8 is strongly homogenous (H-LU = 0.9), with woodland and wetlands (63% and 21%, respectively) comprising the majority of the area. Nearly half of cluster #1 encompasses wetlands (47%) with extremely high organic matter composition (12% of organic carbon in soil weight) and lowest bulk density (0.6 g cm^{-3}). Finally, non-intensive farming dominates cluster #5, with 49% of the cluster dedicated to hay/pasture and the rest occupied by wetlands and forests (15% and 10%).

The second group of clusters, #2,3,6, and 7 (Fig. 3b–c), incorporates landscapes mostly affected by anthropogenic transformation; i.e., highly urbanized areas (1–51% of a subwatershed), commercial/industrial land use (2–24%) and farmland with intensive or non-intensive agriculture (2–51% and 5–23%, respectively). Two additional, unifying features of these subwatersheds are their noticeable high landscape diversity (H-LU = 1.2–1.7) and, naturally, their elevated population density (≥0.4 person ha^{-1}). Regarding

Table 1
Spatial clusters derived by Self-Organizing Map analysis in Lake Simcoe watershed. Variables are input used for data-learning, while darkest cell shade and white font numbers indicate the highest value for each variable among the eight clusters.

Variable	Unit	Mean	Undisturbed clusters				Anthropogenically-influenced clusters			
			1	4	5	8	2	3	6	7
Bulk density	g cm^{-3}	1.12	0.61	1.17	1.14	1.07	1.18	1.17	1.24	1.13
Propotion of area	%	12.51	7.20	16.00	14.30	5.50	15.60	2.80	15.50	23.20
Available water capacity	mm H_2O mm^{-1} soil	0.23	0.22	0.20	0.23	0.24	0.24	0.26	0.15	0.27
Saturated hydraulic conductivity	mm hr^{-1}	38.2	33.8	40.2	26.7	27.6	43.4	31.9	53.5	37.3
Organic carbon content	% soil weight	4.0	12.0	3.2	3.9	4.5	2.9	3.0	2.5	3.5
Soil albedo		0.10	0.11	0.06	0.08	0.10	0.10	0.08	0.16	0.09
Slope	degree	3.5	2.6	2.6	2.5	2.7	4.4	4.5	5.7	3.4
Cropland	%	20.51	9.5	10.3	19.3	4.8	16.9	2.0	13.9	50.7
Hay and pasture	%	17.53	19.2	8.2	48.9	0.3	10.8	4.8	23.2	16.0
Grasssland	%	4.22	3.5	7.5	2.0	4.3	4.8	2.4	4.5	2.6
Forested areas	%	19.14	12.9	18.9	9.9	63.3	19.4	8.2	34.3	11.2
Wetlands	%	19.56	46.5	35.5	14.6	21.0	11.4	3.9	10.0	11.4
Quarries	%	0.76	0.5	0.9	0.5	0.5	0.7	0.0	2.7	0.1
Water covered areas	%	2.18	4.6	4.3	1.0	3.4	2.4	1.7	0.7	0.5
Commercial/Industrial areas	%	4.33	1.0	1.9	1.1	1.0	11.6	23.7	1.6	2.1
Residential low density areas	%	3.29	1.6	3.4	1.1	0.8	5.8	1.8	6.3	3.0
SOD/Golf farms	%	2.60	0.4	7.7	0.8	0.1	2.5	0.9	1.9	1.2
Residential areas	%	5.87	0.4	1.4	0.8	0.5	13.6	50.8	1.0	1.2
Shannon Index, H-LU		1.33	1.1	1.2	1.3	0.9	1.7	1.2	1.5	1.3
Population density	human ha^{-1}	1.31	0.2	0.3	0.3	0.1	3.4	18.5	0.4	0.5
Family Income, average	\$K family^{-1}	82	69	79	66	53	86	79	101	77
Phosphorus export	kg P ha^{-1}	0.29	0.09	0.13	0.12	0.07	0.46	1.14	0.14	0.20

the cluster-specific characteristics, cluster #7 is comprised mostly of agricultural areas with row crops and hay/pasture (51% and 16%, respectively). Small communities with the highest family income across the entire Lake Simcoe watershed ($100,500) occupy 6% of cluster #6, whereas another 23.2% are hay/pasture areas, a necessary source of forage for equestrian properties in this cluster. In contrast, cluster #3, with maximum population density of up to 18.5 person ha^{-1}, represents urbanized cores, whereby an average of 51% and 24% of the subwatersheds LULC belongs to RESIDEN-HD and COMM/IND, respectively. Cluster #2 represents suburban areas, which are adjacent to cluster #3, both in terms of SOM lattice positioning after neural-network training and actual spatial location in the catchment (Fig. 4). Cluster #2 demonstrates moderate population density of 3.4 person ha^{-1} and the second-highest average annual family income among the clusters; $85,700. As a transient zone between clusters #4,6,7 and the highly urbanized cluster #3, subwatersheds in cluster #2 incorporate a mix of different LULC types, such as farmland (28%), forest (20%), and wetlands (11%).

Overall, the SOM classification provides a convenient framework for visual detection of land cover patterns (Skupin and Hagelman, 2005; Arribas-Bel et al., 2011). Specifically, the intensively farmed agricultural lands (CROPLAND) and pastures (HAY/PAST) are expectedly associated with the low end of slope range, in contrast to the concentration of low-density urban areas (URB-LD) in high slope regions (Figs. 2 and 4). Notably, SOM classification of bio-physical and socio-economic data assigned subwatersheds to neighboring lattice nodes based on their overall parametric similarities as an integrated SES unit, although their actual locations may not be in adjacent areas. As such, croplands demonstrate socio-environmental similarities with hay/pasture land cover characteristics, whereas high-density residential areas resemble commercial, institutional and industrial locations (COM/IND). With the same reasoning, grassland areas are positioned in SOM lattice nodes close to forested areas, while subwatersheds with natural or artificial reservoirs (WATER) resemble territories dominated by wetlands.

3.2. *Phosphorus export analysis*

We examined the accuracy of P-export estimates based on multiple and dominant LULC strategies against the reported P tributary loadings at the pour points of 17 LSRCA watersheds (MOECC, 2014; Fig. 5a–d). The correspondence was evaluated for 2009 as the baseline year, with multi-year average catchment runoff of 0.8 km^3 year^{-1} between 1990 and 2009 (MOECC, 2014). While similar agreement was found for multiple LULC ($r^2 = 0.92$, Fig. 5a) and dominant LULC ($r^2 = 0.90$, Fig. 5c), we note that the dominant LULC approach provided better fit in Barrie Creek and East Holland watersheds, while the multiple LULC characterization provided better results in Black River, Pefferlaw Brook, and Ramara Creeks watersheds; the areas with the largest mix of LULC (Louis Berger and LSRCA, 2010, Fig. 5d). Along the same line of reasoning, the corresponding root mean square error values for multiple and dominant LULC characterizations were 1.39 and 1.60 T P year^{-1}, respectively, which favors the former (more detailed) specification of watershed landscape diversity. It should be noted that the LRSCA spatial delineation of Lake Simcoe

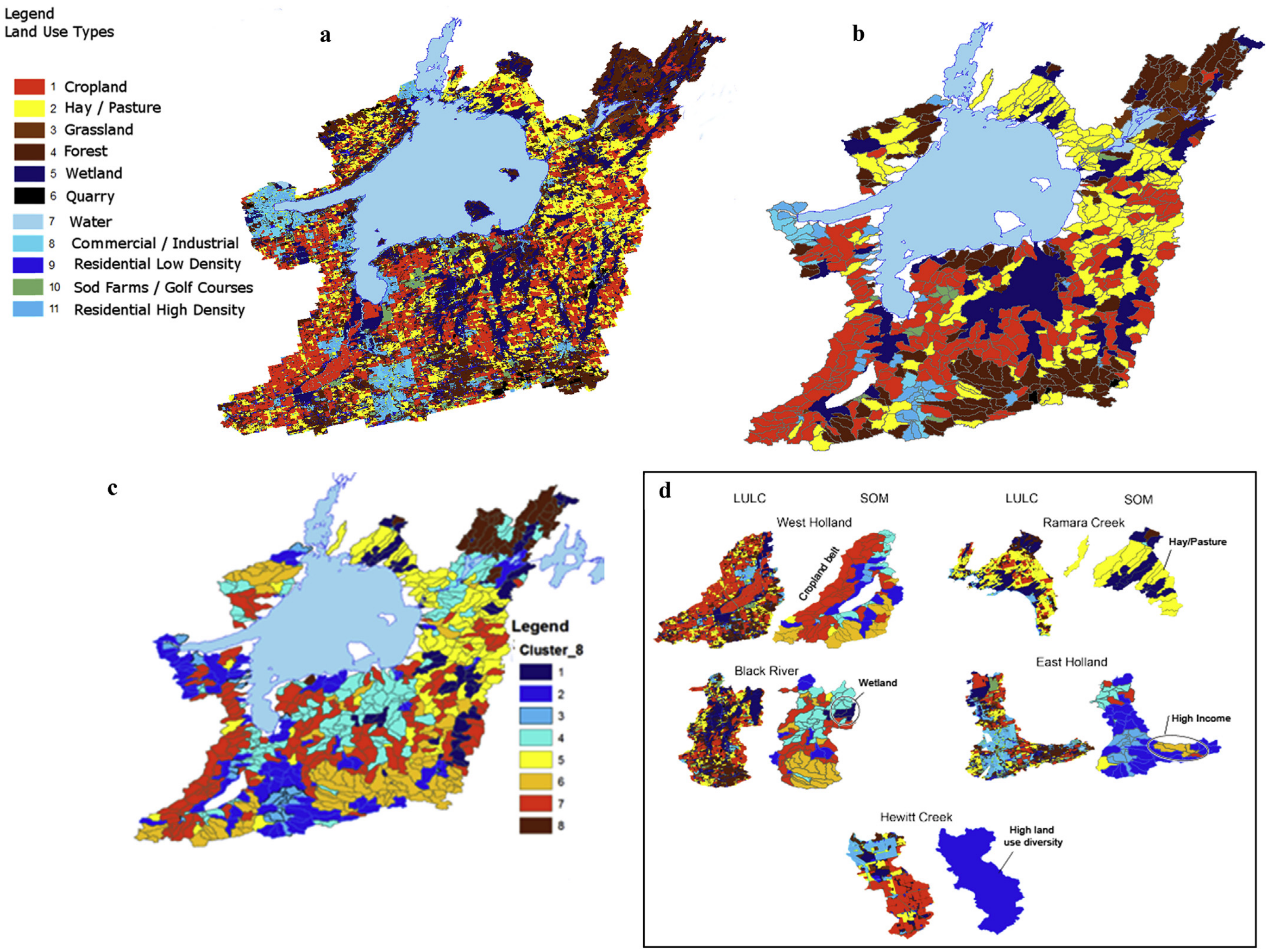


Fig. 4. LULC in Lake Simcoe watershed: (*a*) based on Ecological Land Classification in Ontario; (*b*) based on dominant LULC within each subwatershed; (*c*) according to SOM clustering; (*d*) comparison of multiple LULC vs SOM clustering within LSRCA subwatersheds: upper West Holland classified in cluster 7 due to the dominant agricultural LULC; Ramara Creek classified to cluster 5 due to the predominance of hay/pasture LULC; encircled area in the wetland-dominated Black River classified in cluster 1 due to the very high carbon content; encircled area in East Holland assigned to the exurban cluster 6 due to the high income of the local residents; Hewitt Creek classified to the suburban cluster 2 because of the high land use diversity in the area .

catchment was based on a hydrological approach, mainly driven by the local priorities such as the protection of drinking sources (surface and ground water), storm water and flood management, watershed-based fisheries management and less so the nutrient export mitigation. Fig. 6a summarizes the annual P export from individual watershed clusters as a function of the corresponding occupied area and residing population. Among the LSRCA management watersheds, East Holland, West Holland, Black River, and Pefferlaw Brook are predominantly included in the “anthropogenically-influenced” clusters (Fig. 4c–d), while the former two sites are also responsible for the highest P export (Fig. 6b). By contrast, Beaver River and Talbot River are the LSRCA watersheds with the highest areal contribution to the “undisturbed” clusters (Fig. 6c).

To elucidate the functional relationships between nutrient export from watershed and socio-environmental catchment characteristics, we identified two distinct structural equation models (SEMs) for the undisturbed cluster (Fig. 7a) and anthropogenically-influenced cluster (Fig. 7b), according to the dendrogram classification on Fig. 3c, with acceptable goodness-of-fit statistics (Table S1). The first cluster group connected areal nutrient export rates (kg P ha^{-1} year^{-1}) with the proportion of the subwatersheds occupied by different LULC types, i.e., primarily commercial/industrial and high density residential areas, followed by lakes and ponds, and then all types of forests, wetland areas including bogs and marsh areas, and various crop types. In the second group, the absolute rates of nutrient fluxes (kg P year^{-1}) were controlled by actual area sizes of human-associated LULC (m^2 subwatershed^{-1}), such as hay and pasture areas, various crop types, residential high density areas, commercial and industrial areas, and then the number of residents (persons subwatershed^{-1}), latitude, and landscape diversity index.

3.3. Temporal trends of census data in SOM clusters

Aggregation of socio-economic data into subwatershed boundaries allowed us to identify clear temporal changes, between 1996 and 2006, and differences between SES clusters regarding the age and population demographics (Figs. S1a–c), immigration (Figs. S1d–f), occupation (Figs. S1g–h), types and number of housing (Figs. S2a–c), housing density (Fig. S2d), dwelling prices (Fig. S2e), and family income (Fig. S2f). Notably, although the population concentrates in clusters 2 and 3, all clusters have

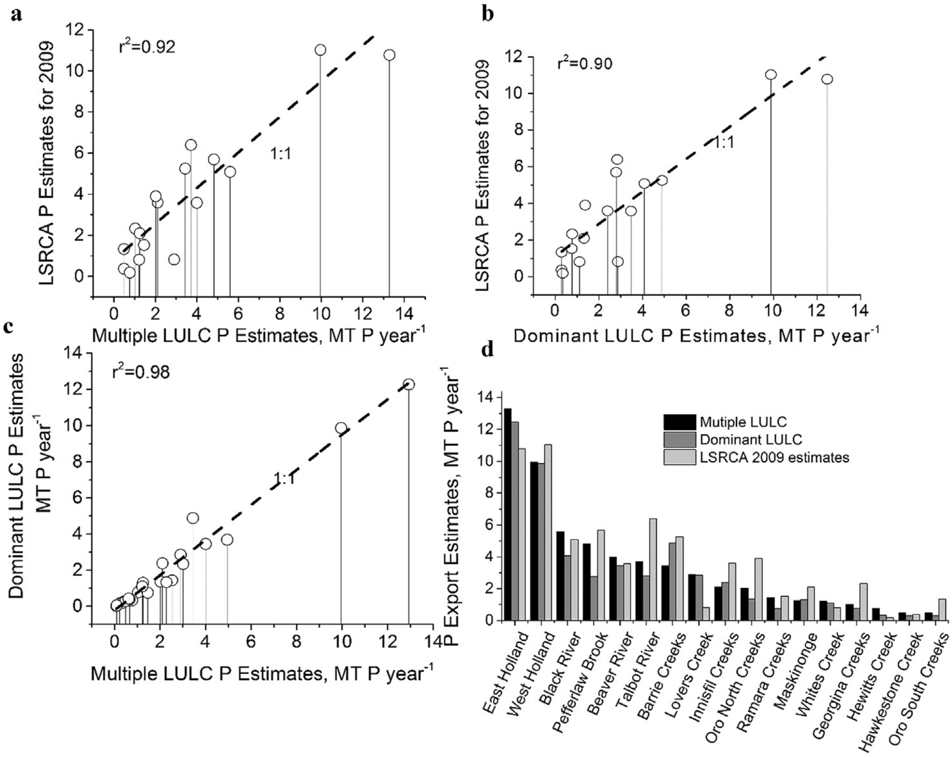


Fig. 5. Comparison of P export from tributary watersheds: (*a*) based on the multiple land use characterization versus LSRCA loading estimates for baseline year 2009; (*b*) P export estimates based on dominant land use for each subwatershed versus LSRCA loading estimates for 2009; (*c*) comparison of P export estimates for multiple and dominant LULC classifications; (*d*) histogram of P export estimates from LSRCA designated watersheds at tributary pour points for multiple and dominant land use classifications.

experienced population growth (in relative terms), which likely accentuates the overall impact of catchment urbanization on the lake ecosystem integrity.

4. Discussion

4.1. Delineation of socio-environmental management units

A holistic approach in natural resource management implies sophisticated coupling of multi-dimensional environmental and social data. This integration is typically based on complex mechanistic DSS models, which simulate social, economic and environmental state variables within a watershed context (van Delden et al., 2011). Alternatively, artificial neural networks, based on biologically inspired principles of adaptability and self-organization, provide a convenient, low-cost framework to achieve synthesis of complex datasets and to extract non-linear patterns (Chon, 2011). As such, the presented SOM approach enabled the examination of Lake Simcoe basin based on its partitioning into small-sized subwatersheds, interconnected by common environmental and socio-economic characteristics. This strategy contrasts the typical practice of basin classification into large-sized subwatersheds based on tributary downstream pour points.

In our research, after introducing the social dimension (e.g., demographics, family income) into an environmental dataset, we used SOM analysis to examine if human activity and associated catchment land use changes result in spatial organizations with common socio-economic and environmental attributes, regardless of whether these entities are hydrologically connected to the same tributaries. Drawing parallels with the concept of Hydrological Response Units (HRU), as cross-basin spatial organizations with common hydrological behaviour (Zehe et al., 2014), we can define the proposed SOM clusters as Socio-Environmental Management Units (SEMU). Interestingly, this SOM clustering bears resemblance to the subwatershed ranking based solely on the dominant LULC (Fig. 4b–c), especially in areas with high spatial homogeneity and minimal population density, such as the cropland belt in West Holland tributary or hay/pasture in Ramara Creeks. Beyond dominant LULC, other SES parameters that can shape watershed classification are the soil characteristics (low soil bulk density in Black River watershed) or family income (forested areas in southern East Holland, Fig. 4d). We also note that the inclusion of the landscape diversity index (H-LU) played an important role in the SOM decision process, e.g., assignment of Hewitt's Creek to cluster #2 (Fig. 4d). Taking all these points into account, the sensitivity of SEMU definition to socio-environmental parameters used is a critical facet of the present framework that warrants further investigation.

Our integrative socio-environmental analysis offered a preliminary tool for the identification of “hot-spots” with increased likelihood of nutrient export rates in Lake Simcoe watershed (Kovacs et al., 2012). Three subwatershed clusters, #2, 3, and 7, were identified as major P sources across the entire basin (68% of total tributary P export), with >50% contribution in 11 out of 17 LSRCA-designated subwatersheds (Fig. S3). In particular, our study pinpointed the subcatchments of West Holland, East Holland, Black River, and Pefferlaw Brook as the largest P contributors in the area (Fig. 6b). Given the ever-growing population and infrastructural expansion plans (MOI, 2011), the latter finding casts doubt on the likelihood of successful management implementation in these locations. Our spatial clustering also sheds light on the water quality impairment in Cook's Bay at the southwestern part of Lake Simcoe, as this embayment is both geographically and hydrologically connected to the previously mentioned West and East Holland Rivers

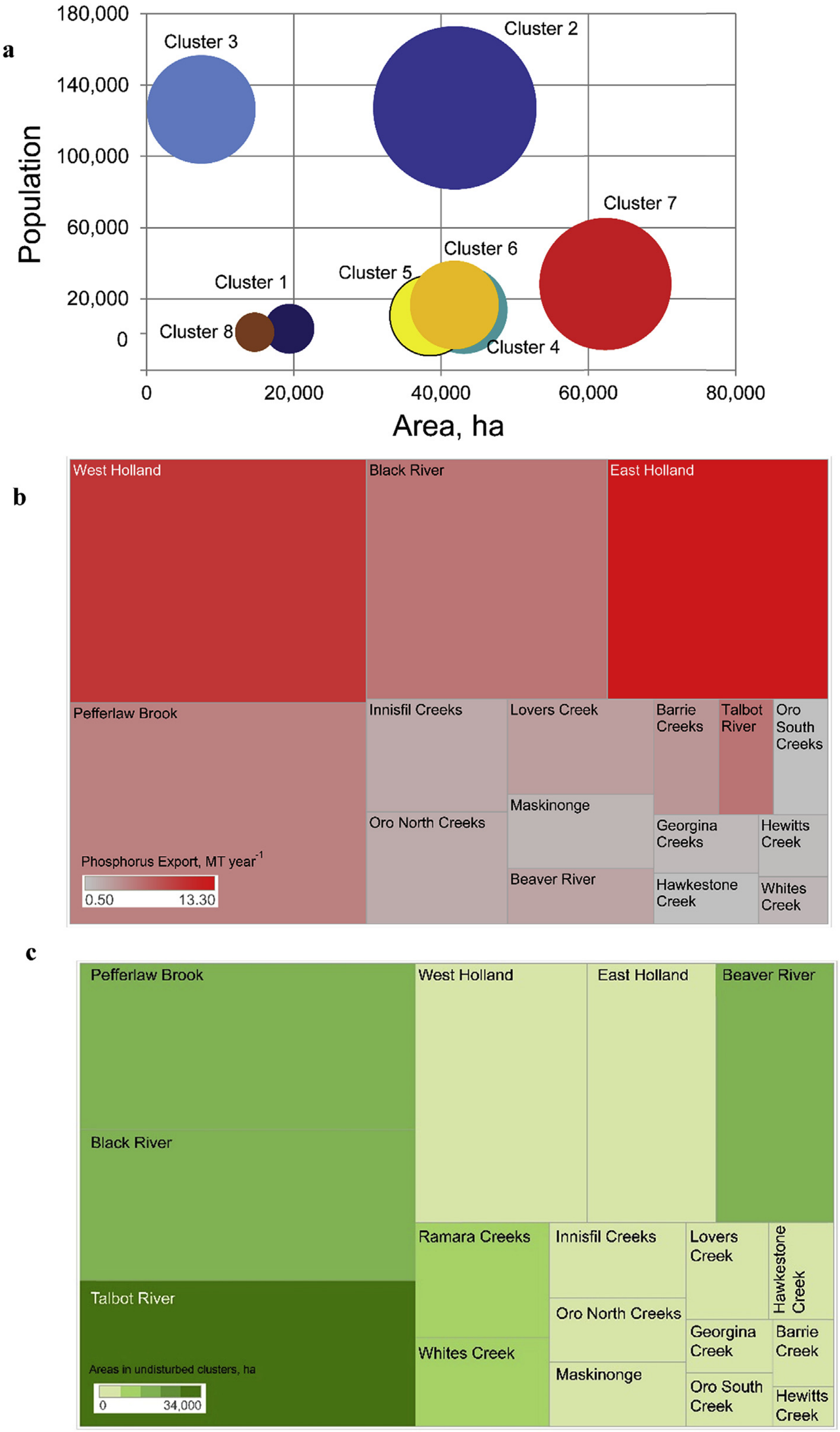


Fig. 6. (*a*) Phosphorus export by size and population of each cluster; the size of each point is proportional to the amount of phosphorous exported; (*b*) diagram of area distribution of anthropogenically-influenced clusters within LRSCA management watersheds with color intensity representing corresponding P export, MT P year^{-1}; (*c*) diagram of total area of LRSCA management watersheds with color intensity representing area contribution of undisturbed clusters within LRSCA management watersheds (ha).

(Gudimov et al., 2012). Phosphorus levels in Cook's Bay are relatively higher (14.8 ± 4.1 μg TP L^{-1}) than any other locations in Lake Simcoe, as a result of elevated P inputs (>18 T year^{-1}) from the highly agricultural Holland Marsh. It is also interesting to note that the predominant fraction of TP in the sediments is carbonate-bound P (apatite-P) mainly due to the accelerated erosion in the

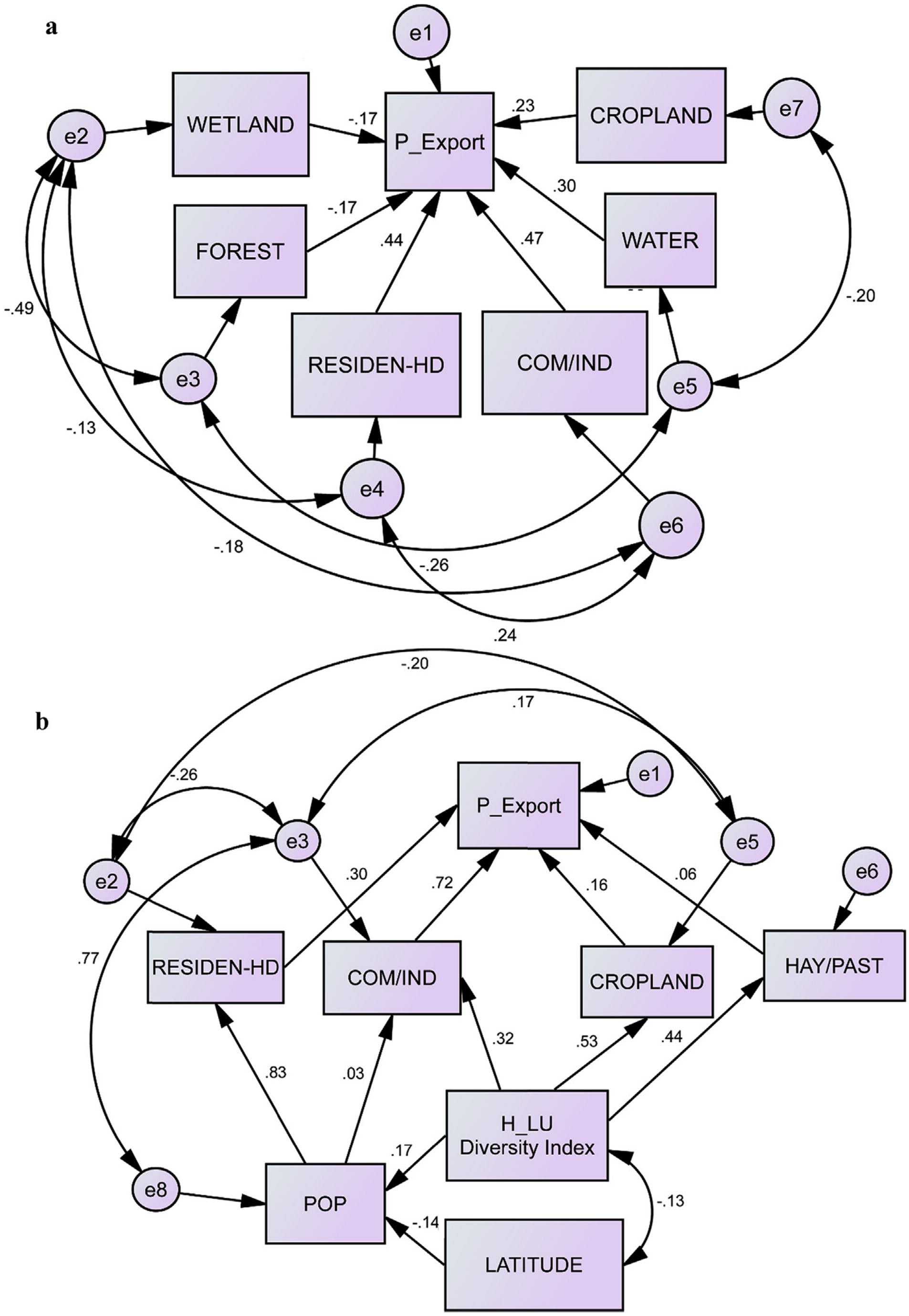


Fig. 7. Structural Equation Model (SEM) with standardized path coefficients: (*a*) for the “undisturbed” clusters, where RESIDEN-HD and COM/IND, WETLAND, FOREST, WATER, CROPLAND and HAY/PASTURE - in %; P_Export - in kg P ha^{-1} year^{-1}; (*b*) for the “anthropogenically-influenced” clusters, where RESIDEN-HD and COM/IND - in m^2; CROPLAND and HAY/PASTURE - in %; P_Export - in kg P year^{-1} subwatershed^{-1}.

Table 2
Standardized total, direct, and indirect effects in SEM model for the “undisturbed” group of clusters (#1,4,5, and 8).

Variable	Cropland	COM/IND	RESIDEN-HD	Water	Forest	Wetland
Standardized Total Effects						
P_Export	0.233	0.467	0.444	0.300	−0.171	−0.165
Standardized Direct Effects						
P_Export	0.233	0.467	0.444	0.300	−0.171	−0.165
Standardized Indirect Effects						
P_Export	N/A	N/A	N/A	N/A	N/A	N/A

adjacent catchment (Gudimov et al., 2015).

4.2. Determination of causal linkages within socio-environmental management units

Our structural equation modelling exercise enabled the identification of the important SES parameters to control nutrient export in two major SOM cluster groups. The first “undisturbed” group of clusters, characterized by modest landscape diversity index (<1.3) and low average population density <0.4 humans ha^{-1} (Table 1),

Table 3
Standardized total, direct and indirect effects in SEM model for the "anthropogenically-influenced" group of clusters (#2, 3, 6, and 7).

Variable	Latitude	H_LU	POP	HAY/PAST	CROPLAND	COM/IND	RESIDEN-HD
Standardized Total Effects							
POP	−0.145	0.167					
HAY/PAST		0.438					
CROPLAND		0.530					
COM/IND	−0.005	0.321	0.034				
RESIDEN-HD	−0.120	0.138	0.830				
P_Export	−0.040	0.383	0.277	0.060	0.157	0.721	0.304
Standardized Direct Effects							
POP	−0.145	0.167					
HAY/PAST		0.438					
CROPLAND		0.530					
COM/IND		0.315	0.034				
RESIDEN-HD			0.830				
P_Export				0.060	0.157	0.721	0.304
Standardized Indirect Effects							
POP							
HAY/PAST							
CROPLAND							
COM/IND	−0.005	0.006					
RESIDEN-HD	−0.120	0.138					
P_Export	−0.040	0.383	0.277				

occupies approximately half of the Lake Simcoe catchment (45%). In this group, the areal nutrient export is modulated by the relative contribution of different land use/land cover types, expressed as percentages of a subwatershed area. Importantly, P export is particularly sensitive to the relative amount of anthropogenically transformed areas, such as infrastructural and urbanized locations (standardized path coefficients of 0.47 and 0.44), artificial on-stream reservoirs (0.30), and croplands (0.23) (Table 2). This cluster group contains the largest inventory of forests and wetlands, the known natural landscape mediators to control P mobilization from catchments (McClain et al., 2003). Nonetheless, the standardized paths from forests and wetlands (≈-0.17) are weaker relative to those from anthropogenically affected areas (0.23–0.47). Thus, to maximize the attenuation effects of this land use coverage, it is critical not only to faithfully comply with the LSRCA stipulation that forests and wetlands should not fall below a minimum of 40% of total basin area but also that this threshold should be consistently met in all major local tributaries (MOECC, 2010).

Unlike the first group, nutrient fluxes from the "anthropogenically-influenced" group of clusters are predominantly controlled by two factors, i.e., human-induced landscape heterogeneity and subwatershed demographics with total effects equal to 0.38 and 0.28, respectively (Table 3). In terms of the importance of different LULC types, commercial/industrial and urbanized high-density areas in the "anthropogenically-influenced" clusters exert significant control over catchment P budget, whereas croplands and hay/pastures appear to play a secondary role. The extent of residential areas is directly associated with catchment demographics (pathway strength of 0.83). Further, the strong positive covariance between the residual variability in population and infrastructural LULC (correlation coefficient between residuals e3-e8 equal to 0.77) is conceptually on par with the notion that community infrastructure expansion closely follows population growth in Ontario (MOI, 2011). Additionally, the infrastructural expansion and land-use diversity (H-LU) appear to have a tight association and collectively this pathway is the second strongest modulator of P export (total effect equal to 0.23 = 0.32 × 0.72). It is therefore reasonable to assume that the effect of basin population growth and ecosystem degradation is magnified by the concurrent increase in infrastructure and industrial developments.

Interestingly, the optimal SEM structure for the "anthropogenically-influenced" clusters does not identify any pathways with negative regression coefficients, which could provide offsetting mechanisms for human-induced P mobilization and therefore could be socio-environmentally non-sustainable in the long run. This structural difference may be reflective of the presence of two alternative states within the Lake Simcoe watershed (Polasky et al., 2011); namely, the "undisturbed" vis-à-vis the "anthropogenically-influenced". The latter finding allows us to speculate about two critical conditions that may solidify the shift between the two contrasting states: (i) establishment of a distinct signature of human pressure on the tributary nutrient export, effectively accelerating nutrient export rates for every incremental addition of new residents and/or conversion to urban/infrastructural land use in the watershed; and (ii) the loss of subwatershed sensitivity to natural (counterbalancing) mechanisms that may curb nutrient export.

4.3. Population and housing trends in socio-environmental management units

The standardized paths of our structural equation models suggest that the "anthropogenically-influenced" group of clusters demonstrates weak latitudinal gradients in the population densities (−0.14), residential built-up coverage (−0.12 = −0.14 × 0.83), and weak covariance between latitude and landscape diversity indices (r = −0.13). Taking into account that our study area spans 90 km from south to north just above the City of Toronto, this finding contrasts the typically hypothesized suburban-density gradients and lends validity to the use of (not necessarily spatially connected) socio-environmental units as pockets of growth and environmental stress in the Greater Toronto Area (Bunting et al., 2002). Post-hoc, multi-layer superposition of census data with SOM clusters shows that 91% of the population tends to concentrate in "anthropogenically-influenced" clusters with 77% belonging exclusively to urbanized clusters #2–3 (Figs. S1b and c). Importantly, all corresponding decadal growth rates in clusters #2, 3, 6 and 7 (28%, 18%, 23% and 22%) were well above the concurrent provincial growth of 13% and the catchment long-term growth projection of ~17% for 2011–2031 (MOI, 2012). The highest growth rate of 28% in cluster #2 resulted from the cumulative effect of both migration of families with multiple children or out-of-province migration (Figs. S1e and f). The main local urban centers (Newmarket, Barrie, Orillia, Beaverton, Uxbridge) have demonstrated a shift from the manufacturing industry as a major employment sector, to a post-industrial service economy (Figs. S1g and h). This qualitative change in the productivity mode is conceptually on par with the urban scaling effect (Bettencourt, 2013), whereby bigger urban centers demonstrate an ability to create internal service jobs and thus multiply socio-economic interactions within the basin, which in turn serves to attract more residents in the catchment. Taking into account the basin's population growth, from 185,000 to 435,500 residents between 1981 and 2011 (MOECC, 2014), with primary settlement in "anthropogenically-influenced" clusters, these trends collectively bring to the forefront of our study the catchment's overall capacity to maintain environmental resilience without massive investment in best-management and conservation practices.

In the same context, we followed the Clark (2012) proposal to analyze housing parameters in order to unveil the critical drivers of accelerated residential mobility in the Lake Simcoe watershed and validate the MOI's (2012) population-growth projections. The majority of local families prefer to reside in single, detached houses, which also registered the largest increase (>16,000) of the new houses built between 1996 and 2006 (Figs. S2a–c). The total housing inventory correlated well with cluster demographics, with ~80% of total dwellings allocated to clusters #2–3 (Figs. S2b and c). These clusters also accounted for the widest affordable housing mix, with the largest number of apartments located in highly urbanized cluster #3, while suburban cluster #2 experienced higher development rates in row- and semi-detached dwellings (Fig. S2c). Following demographic trends, clusters #2–3 demonstrated the highest rates of residential development (Fig. S2d), given the addition of ~1 housing unit ha^{-1} in cluster #3, partly due to construction of high-rise, multi-apartment buildings. This impetus for active residential development can be attributed to disparity between average dwelling prices in the Lake Simcoe watershed ($60,000–120,000) and the Greater Toronto Area ($350,000) in 2006, respectively; with more recent prices (2016) of detached houses of $749,194 in Newmarket and $1,061,789 in Toronto. Taking into account family incomes between 1996 and 2006 were ranging from $85,000 to $125,000 in Lake Simcoe catchment (Fig. S2f) versus a contemporaneous annual average income of $97,000 in the GTA, affordable housing can, indeed, represent a strong incentive for accelerated residential mobility in the Lake Simcoe watershed.

4.4. SEMU and stewardship programs

Our integrated socio-environmental analysis can also cast light on residents' potential engagement in stewardship programs for nutrient loading reduction initiatives (Bulkley, 2011). Based on the SOM analysis, residents of cluster #6 are characterized by highest family income and dwelling values in the basin (Figs. S2e and f). These residents deliberately opt for quasi-rural settlements in order to be more connected to nature, which may imply an aptitude to support conservation programs (Cadieux and Taylor, 2013). This promotes cluster #6 as a priority target for catchment stewardship initiatives. By contrast, residents of clusters #1, 5, and 8 represent the low end of family-income spectrum (Fig. S2f), which may reflect their potential financial interest in wetland mitigation programs (Bendor, 2009), such as Water Quality Trading (WQT) and the Lake Simcoe Phosphorus Offset Program (MOECC, 2010). Specifically, while already holding a large inventory of wetlands (14.6–46.5% of subwatershed areas, Table 1), clusters #1, 5, and 8, under the proposed WQT regulations to minimize P loading from non-point sources (Bendor, 2009), can be further occupied with wetlands and forested areas by phasing out hay, pasture, and other agricultural lands.

4.5. Optimal SEMU spatial resolution

Selecting an appropriate spatial scale for natural-resource management represents an important issue for integrated socio-environmental analysis (Laniak et al., 2013). Watershed management plans are typically developed based on catchment hydrological boundaries instead of administrative ones, which enables accounting for the hydrological connectivity and therefore tracking point and non-point P sources (Laniak et al., 2013). The Lake Simcoe Phosphorus Reduction Strategy (MOECC, 2010) also considers catchment delineation for 19 tributary watersheds along main rivers (Fig. 2b), with areas ranging between 1752–44,624 ha, which roughly corresponds to the USGS 12-digit hydrological unit classification (USGS, 2009). Each LSW tributary watershed is subject to an integrated management plan with detailed specification of current and projected environmental stressors, social and economic factors, and land-use planning strategies (LSRCA, 2012). The modern IWM practice partly connects the success of watershed management programs with their finer resolution, down to a level of hectares or separate households (van Delden et al., 2011; Laniak et al., 2013). For example, the Soil and Water Assessment Tool (SWAT) modelling framework (Arnold et al., 1998) recommends 2% of watershed area as an optimal subdivision level for analyzing BMP scenarios (Arabi et al., 2006). In a recent case study of the Eagle Creek Watershed, Indiana, USA (Babbar-Sebens et al., 2015), a watershed resolution in the range of 50–700 ha was considered optimal for management purposes. Likewise, our study followed the 2% threshold with a mean subwatershed size of 350 ha to maximize the value of information for our socio-environmental analysis (Fig. 2a).

The proposed SEMU alternative as a basic unit in watershed management would also greatly benefit from the recent advances in real-time ecohydrology and continuous water-quality monitoring on a municipal scale (Krause et al., 2015), e.g., the AQUARIUS stormwater study in Lake Oswego; Southern Ontario Water Consortium project (www.sowcdata.ca). The development of technical capacity of implementing SMART (Specific Measurable Achievable Relevant Time-bound) principles at a scale of smaller subwatersheds creates a necessary feed-back management mechanism to monitor unsustainable practices or reveal stakeholders' free-riding attitude. The SMART principle would serve to increase the stakeholders' accountability for BMP implementation in a specific SEMU, since their progress can be verified qualitatively and quantitatively in a continuum (along the tributary) relative to the typical practice of sampling water-quality parameters at selective pour points downstream. As an additional feedback mechanism, volunteer capabilities in data-collection under the citizen-science initiative (Rieman and Wallenburn, 2015) and the overall involvement of watershed residents of similar SEMUs (Conrad and Hilchey, 2011) can fill the gaps in government-based monitoring networks by enhancing the temporal and spatial frequency of grab samples.

5. Conclusions

We have presented an integrated socio-environmental analysis to shed light on the present state of Lake Simcoe watershed and future perspective in achieving a significant nutrient loading reduction from 72 to 115 T P year^{-1} in 2007–2011 to 44 T P year-^{1} by 2045. Using Self- Organizing Mapping of biophysical and socio-economic data, we delineated Socio-Environmental Management Units as a basic organizational unit for integrated watershed management planning. Integrating social, economic, and environmental data enabled to distinguish between reference or "undisturbed" and "anthropogenically-influenced" subwatersheds. In the latter group, affordable housing appears to be a strong accelerating factor of residential mobility. Moreover, qualitative changes in the productivity mode, characterized by a shift from the manufacturing industry to a post-industrial service economy, appear to magnify socio-economic interactions and thus attract more residents in the Lake Simcoe basin. Bearing in mind the ever-growing urbanization pressure, we showed that the "anthropogenically-influenced" sites have functional properties that collectively cast doubt on their sustainability, such as the presence of a tight relationship between demographics and tributary nutrient export, and absence of natural mechanisms to control environmental degradation. Differences in the socio-environmental patterns within Lake Simcoe watershed can be used to gauge the residents' potential engagement in stewardship programs for nutrient loading reduction initiatives.

Acknowledgments

This project was undertaken with the financial support of the Government of Canada provided through the Department of the Environment granted to Prof. George Arhonditsis. Alex Neumann has also received additional financial support from the Mathematics of Information Technology and Complex Systems (MITACS) Elevate program (grant #IT03615).

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jenvman.2016.11.073.

References

Lake Simcoe Region Conservation Authority (LSRCA), 2012. Barrie Creeks, Lovers Creek, and Hewitt's Creek Subwatershed Plan. Retrieved at: www.lsrca.on.ca/pdf/reports/barrie_subwatershed_plan_2012.pdf.

Ministry of the Environment and Climate Change (MOECC), 2012. Phosphorus Budget Tool in Support of Sustainable Development for the Lake Simcoe Watershed. Retrieved at:. Hutchinson Environmental Sciences Ltd http://goo.gl/oFn8yR.

Ministry of the Environment and Climate Change (MOECC), 2014. Lake Simcoe Monitoring Report. Retrieved at:. Queen's Printer for Ontario www.lsrca.on.ca/Shared%20Documents/reports/moecc-lake-simcoe-monitoring.pdf.

AQUARIUS stormwater, TMDL monitoring & regulatory compliance case study: City of Lake Oswego improves urban water resource management with real-time data analysis. Retrieved at http://pages.aquaticinformatics.com/rs/aquaticinformatics/images/Case-Study-City-of-Lake-Oswego.pdf.

Arabi, M., Govindaraju, R., Hantush, M., Engel, B., 2006. Role of watershed subdivision on modelling the effectiveness of best management practices with SWAT. J. Am. Water Res. Assoc. 42 (2), 513–528.

Arbuckle, J.L., 2013. IBM SPSS Amos 22 User's Guide. Amos Development Corporation, Crawfordville, FL. Retrieved at: http://www.sussex.ac.uk/its/pdfs/SPSS_Amos_User_Guide_22.pdf.

Arhonditsis, G.B., Stow, C.A., Steinberg, L.J., Kenney, M.A., Lathrop, R.C., McBride, S.J., Reckhow, K.H., 2006. Exploring ecological patterns with structural equation modelling and Bayesian analysis. Ecol. Model. 192 (3), 385–409.

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modelling and assessment. Part I: model development. J. Am. Water Resour. Assoc. 34 (1), 73–89.

Arribas-Bel, D., Nijkamp, P., Scholten, H., 2011. Multidimensional urban sprawl in Europe: a self-organizing map approach. Comput., Environ. Urban Syst. 35 (4), 263–275.

Babbar-Sebens, M., Mukhopadhyay, S., Singh, V.B., Piemonti, A.D., 2015. A web-based software tool for participatory optimization of conservation practices in watersheds. Environ. Model. Softw. 69, 111–127.

Bendor, T., 2009. A dynamic analysis of the wetland mitigation process and its effects on no net loss policy. Landsc. Urban Plan. 89 (1), 17–27.

Bettencourt, L.M., 2013. The origins of scaling in cities. Science 340 (6139), 1438–1441.

Beven, K.J., 2006. A manifesto for the equifinality thesis. J. Hydrol. 320, 18–36.

Bowen, R.E., Riley, C., 2003. Socio-economic indicators and integrated coastal management. Ocean. Coast. Manag. 46 (3), 299–312.

Box, G., Cox, D., 1964. An analysis of transformations. J. R. Stat. Soc. Ser. B 26, 211–252.

Bulkley, J.W., 2011. Integrated watershed management: past, present, and future. J. Contemp. Wat. Res. Edu. 100 (1), 3–15.

Bunting, T., Filion, P., Priston, H., 2002. Density gradients in Canadian metropolitan regions, 1971-96: differential patterns of central area and suburban growth and change. Urb. Stud. 39 (13), 2531–2552.

Cadieux, K.V., Taylor, L., 2013. Landscape and the Ideology of Nature in Exurbia: Green Sprawl. Routledge, 312 pp.

Chertow, M.R., 2000. The IPAT equation and its variants. J. Ind. Ecol. 4 (4), 13–29.

Chon, T.-S., 2011. Self-organizing maps applied to ecological sciences. Ecol. Inf. 6 (1), 50–61.

Clark, W.A., 2012. Residential Mobility and the Housing Market. In: The SAGE Handbook of Housing Studies, pp. 66–83.

Conrad, C.C., Hilchey, K.G., 2011. A review of citizen science and community-based environmental monitoring: issues and opportunities. Environ. Monit. Assess. 176 (1–4), 273–291.

Eiden, G., Kayadjanian, M., Vidal, C., 2000. From Land Cover to Landscape Diversity in the European Union. Retrieved at: http://ec.europa.eu/agriculture/publi/reports/corine2000.pdf.

Frey, U.J., Rusch, H., 2013. Using artificial neural networks for the analysis of social-ecological systems. Ecol. Soc. 18 (2), 40.

Giraudel, J.L., Lek, S., 2001. A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination. Ecol. Model. 146 (1), 329–339.

Gudimov, A., O'Connor, E., Dittrich, M., Jarjanazi, H., Palmer, M.E., Stainsby, E., Arhonditsis, G.B., 2012. Continuous Bayesian network for studying the causal links between phosphorus loading and plankton patterns in Lake Simcoe, Ontario, Canada. Environ. Sci. Technol. 46 (13), 7283–7292.

Gudimov, A., Kim, D.-K., Young, J.D., Palmer, M.E., Dittrich, M., Winter, J.G., Stainsby, E., Arhonditsis, G.B., 2015. Examination of the role of dreissenids and macrophytes in the phosphorus dynamics of Lake Simcoe, Ontario, Canada. Ecol. Inf. 26, 36–53.

Ha, J.Y., Hanazato, T., Chang, K.H., Jeong, K.S., Kim, D.-K., 2015. Assessment of the lake biomanipulation mediated by piscivorous rainbow trout and herbivorous daphnids using a self-organizing map: a case study in Lake Shirakaba, Japan. Ecol. Inf. 29, 182–191.

Harrison, M.T., Tardieu, F., Dong, Z., Messina, C.D., Hammer, G.L., 2014. Characterizing drought stress and trait influence on maize yield under current and future conditions. Glob. Chang. Biol. 20, 3.

Hess, P.M., Sorensen, A., 2015. Compact, concurrent, and contiguous: smart growth and 50 years of residential planning in the Toronto region. Urban. Geogr. 36 (1), 127–151.

Howard, K.W., Eyles, N., Smart, P.J., Boyce, J.I., Gerber, R.E., Salvatori, S.L., Doughty, M., 1995. The Oak Ridges Moraine of southern Ontario: a groundwater resource at risk. Geosci. Can. 22, 3.

Jin, L., Whitehead, P., Baulch, H., Dillon, P., Butterfield, D., Oni, S., Futter, M., Crossman, J., O'Connor, E., 2013. Modelling phosphorus in Lake Simcoe and its subcatchments: scenario analysis to assess alternative management strategies. Inland Waters 3 (2), 207–220.

Kim, D.-K., Kaluskar, S., Mugalingam, S., Arhonditsis, G.B., 2016. Evaluating the relationships between watershed physiography, land use patterns, and phosphorus loading in the bay of Quinte basin, Ontario, Canada. J. Great Lakes Res. 42, 972–984.

Kohonen, T., 2013. Essentials of the self-organizing map. Neural Netw. 37, 52–65.

Kohonen, T., 2001. Self-organizing Maps. Springer, 501 pp.

Kovacs, A., Honti, M., Zessner, M., Eder, A., Clement, A., Blöschl, G., 2012. Identification of phosphorus emission hotspots in agricultural catchments. Sci. Total Environ. 433, 74–88.

Krause, S., Lewandowski, J., Dahm, C.N., Tockner, K., 2015. Frontiers in real-time ecohydrology–a paradigm shift in understanding complex environmental systems. Ecohydrology 8 (4), 529–537.

Lagutov, V. (Ed.), 2011. Environmental Security in Watersheds: the Sea of Azov. Springer, 468 pp.

Lake Simcoe Region Conservation Authority (LSRCA), 2013. Annual Water Balances and Total Phosphorus Loads to Lake Simcoe (2007–2009). Retrieved at: www.lsrca.on.ca/Shared%20Documents/reports/phosphorus_etal_2007_2009.pdf.

Laniak, G.F., Olchin, G., Goodall, J., Voinov, A., Hill, M., Glynn, P., Hughes, A., 2013. Integrated environmental modelling: a vision and roadmap for the future. Environ. Model. Softw. 39, 3–23.

Lee, H.T., 1998. Ecological Land Classification for Southern Ontario: First Approximation and its Application. Ontario Ministry of Natural Resources, Southcentral Science Section, North Bay.

Liu, J., Dietz, T., Carpenter, S.R., Alberti, M., Folke, C., Moran, E., Taylor, W., 2007. Complexity of coupled human and natural systems. Science 317 (5844), 1513–1516.

Lobo, D., Levin, M., 2015. Inferring Regulatory networks from experimental morphological phenotypes: a computational method reverse-engineers planarian regeneration. PLoS Comput. Biol. 11 (6), e1004295.

Louis Berger and LSRCA, 2010. Estimation of Phosphorus Loading to Lake Simcoe. 2010. Retrieved at: www.lsrca.on.ca/pdf/reports/phosphorus_estimates_2010.pdf.

Luo, Y., Su, B., Yuan, J., Li, H., Zhang, Q., 2011. GIS techniques for watershed delineation of SWAT model in plain polders. Proced. Environ. Sci. 10, 2050–2057.

Mayer, A., Donovan, R., Pawlowski, C., 2014. Information and entropy theory for the sustainability of coupled human and natural systems. Ecol. Soc. 19 (3), 11.

McClain, M.E., Boyer, E.W., Dent, C.L., Gergel, S.E., Grimm, N.B., Groffman, P.M., Hart, S.C., Harvey, J.W., Johnston, C.A., Mayorga, E., McDowell, W.H., 2003. Biogeochemical hot spots and hot moments at the interface of terrestrial and aquatic ecosystems. Ecosystem 6 (4), 301–312.

Ministry of Infrastructure (MOI), 2011. Building Together. Jobs and Prosperity for

Ontarians. Retrieved at:. Queen's Printer for Ontario http://www.moi.gov.on.ca/pdf/en/buildingtogether_en.pdf.

Ministry of Infrastructure (MOI), 2012. Growth Plan for the Greater Golden Horseshoe, 2006. Retrieved at: https://goo.gl/7LNwQl.

Ministry of the Environment and Climate Change (MOECC), 2010. Lake Simcoe Phosphorus Reduction Strategy. Retrieved at:. Queen's Printer for Ontario www.ontario.ca/document/phosphorus-reduction-strategy.

Niemelä, P., Honkela, T., 2009. Analysis of parliamentary election results and socio-economic situation using self-organizing map. In: Advances in Self-organizing Maps. Springer, Berlin Heidelberg, pp. 209–218.

Ontario Ministry of Environment (MOECC), 2010. Water Quality Trading in the Lake Simcoe Watershed -feasibility Study. Retrieved at: http://www.ontariocanada.com/registry/view.do?postingId=2982.

Ontario Ministry of Natural Resources (OMNR), 2009. The Lake Sturgeon in Ontario. Downloaded at:. Fish and Wildlife Branch. Peterborough, Ontario www.mnr.gov.on.ca.

Organization for Economic Cooperation and Development (OECD), 1993. OECD Core Set of Indicators for Environmental Performance Reviews, Paris.

Ostrom, E., 2008. The challenge of common-pool resources. Environment 50 (4), 8–21.

Ostrom, E., 2009. A general framework for analyzing sustainability of social-ecological systems. Science 325, 419–422.

Palmer, M.E., Winter, J.G., Young, J.D., Dillon, P.J., Guildford, S.J., 2011. Introduction and summary of research on Lake Simcoe: research, monitoring, and restoration of a large lake and its watershed. J. Gt. Lakes. Res. 37, 1–6.

Park, Y.S., Kwon, Y.S., Hwang, S.J., Park, S., 2014. Characterizing effects of landscape and morphometric factors on water quality of reservoirs using a self-organizing map. Environ. Model. Softw. 55, 214–221.

Polasky, S., Carpenter, S.R., Folke, C., Keeler, B., 2011. Decision-making under great uncertainty: environmental management in an era of global change. Trends Ecol. Evol. 26 (8), 398–404.

Rieman, B., Wallenburn, J., 2015. Water Quality Monitoring to Determine the Influence of Roads and Road Restoration on Turbidity and Downstream Nutrients: a Pilot Study with Citizen Science. Retrieved at: http://goo.gl/EV6NDw.

Scalenghe, R., Marsan, F.A., 2009. The anthropogenic sealing of soils in urban areas. Landsc. Urban Plan. 90, 1.

Skupin, A., Hagelman, R., 2005. Visualizing demographic trajectories with self-organizing maps. GeoInformatica 9 (2), 159–179.

Ullman, J.B., 2006. Structural equation modelling: reviewing the basics and moving forward. J. Pers. Asses. 87 (1), 35–50.

U.S. Geological Survey (USGS), 2009. Federal Guidelines, Requirements, and Procedures for the National Watershed Boundary Datasets (Chapter 3) of Section A, Federal Standards, Book 11, Collection and Delineation of Spatial Data.

van Delden, H., Seppelt, R., White, R., Jakeman, A.J., 2011. A methodology for the design and development of integrated models for policy support. Environ. Model. Softw. 26 (3), 266–279.

Vesanto, J., Alhoniemi, E., 2000. Clustering of the self-organizing map. Neural Netw. IEEE Trans. 11 (3), 586–600.

Vesanto, J., Himberg, J., Alhoniemi, E., Parhankangas, J., 2000. SOM Toolbox for Matlab 5.

Vovik, O., 2014. Non-pollen Palynomorphs and Thecamoebians as Proxies of Environmental and Anthropogenic Change: a Case Study from Lake Simcoe, Ontario, Canada. Master's thesis. Brock University. Retrieved at: http://goo.gl/VOTQ5A.

Walker, B., Salt, D., 2006. Resilience Thinking: Sustaining Ecosystems and People in a Changing World. Island Press, Washington, D.C., USA, 192 pp.

Zehe, E., Ehret, U., Pfister, L., Blume, T., Schröder, B., Westhoff, M., Jackisch, C., Schymanski, S.J., Weiler, M., Schulz, K., Allroggen, N., Tronicke, N., van Schaik, L., Dietrich, P., Scherer, U., Eccard, J., Wulfmeyer, V., Kleidon, A., 2014. HESS Opinions: from response units to functional units: a thermodynamic reinterpretation of the HRU concept to link spatial organization and functioning of intermediate scale catchments. Hydrol. Earth Syst. Sc. 11, 4635–4655.

INTEGRATIVE ANALYSIS OF THE LAKE SIMCOE WATERSHED (ONTARIO, CANADA) AS A SOCIO-ECOLOGICAL SYSTEM

Alex Neumann, Dong-Kyun Kim, Gurbir Perhar, George B. Arhonditsis*

SUPPORTING INFORMATION

Ecological Modelling Laboratory,
Department of Physical & Environmental Sciences,
University of Toronto, 1065 Military Trail, Toronto, Ontario, Canada, M1C 1A4

* Corresponding author

E-mail: georgea@utsc.utoronto.ca, Tel.: +1 416 208 4858; Fax: +1 416 287 7279.

List of abbreviations used in the paper

BMP - Best Management Practices

CAD – Canadian Dollars

CPR – Common Pool Resource

DA - Dissemination Areas in Census Canada

EA - Enumeration Areas in Census Canada

IPAT – Impact, Population, Affluence and Technology variables

IWM – Integrated Watershed Management

LSRCA – Lake Simcoe Region Conservation Authority

LSW – Lake Simcoe Watershed

LULC – Land Use/Land Cover

MOECC - The Ministry of the Environment and Climate Change in the Province of Ontario, Canada

PCA - Principal Component Analysis

PSR - Pressure-State-Response Framework

SEM – Structural Equation Model

SES - Social-Ecological System

SMART- Specific Measurable Achievable Relevant Time-bound

SOM – Self-Organizing Maps

SWAT – Soil and Water Assessment Tool

USGS - United States Geological Survey

FIGURES LEGENDS

Figure S1: Socio-economic census data for 1996 & 2006: (*a*,*d*) aggregated for the Lake Simcoe catchment; (*b*-*c*, *e*-*h*) superimposed to SOM subwatershed clusters: (*a*) demographics of age groups; (*b*) age group distributions within SOM clusters; (*c*) relative changes of age groups demographics between SOM clusters in 1996-2006; (*d*) immigrants, labor force and total population in 2006; (*e*) relative distribution of immigrants among SOM clusters; (*f*) relative changes of immigrants among SOM clusters; (*g*) census data for occupation groups for the entire catchment; and (*h*) relative changes in occupations among SOM clusters.

Figure S2: Housing and family income statistics from 1996 and 2006 census data: (*a*) types of housing statistics for the entire catchment; (*b*) distribution of dwelling types between SOM clusters; (*c*) relative changes of dwelling types between SOM clusters in 1996-2006; (*d*) housing density in SOM cluster, unit ha^{-1}; (*e*) average dwelling value in SOM clusters, \$ CAD; (*f*) average family income in SOM clusters, \$CAD.

Figure S3: Contribution (*%*) of clusters #2,3,7 in P export from LSRCA tributary watersheds.

Figure S1

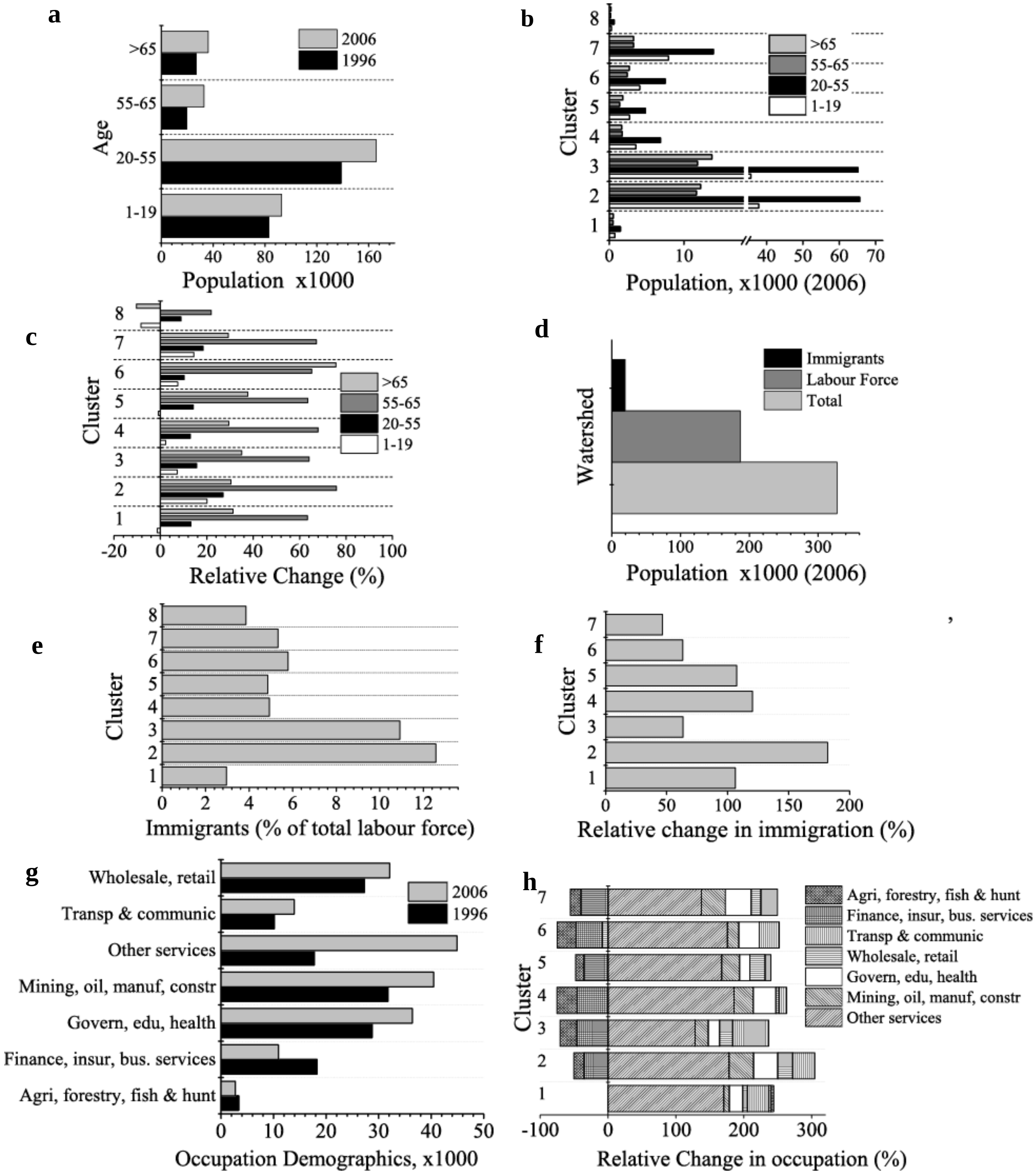

a
>65
55-65
20-55
1-19
Age
2006
1996
0 40 80 120 160
Population x1000
b
Cluster
8 7 6 5 4 3 2 1
>65
55-65
20-55
1-19
0 10 40 50 60 70
Population, x1000 (2006)
c
Cluster
8 7 6 5 4 3 2 1
>65
55-65
20-55
1-19
-20 0 20 40 60 80 100
Relative Change (%)
d
Watershed
Immigrants
Labour Force
Total
0 100 200 300
Population x1000 (2006)
e
Cluster
8 7 6 5 4 3 2 1
0 2 4 6 8 10 12
Immigrants (% of total labour force)
f
Cluster
7 6 5 4 3 2 1
0 50 100 150 200
Relative change in immigration (%)
g
Wholesale, retail
Transp & communic
Other services
Mining, oil, manuf, constr
Govern, edu, health
Finance, insur, bus. services
Agri, forestry, fish & hunt
2006
1996
0 10 20 30 40 50
Occupation Demographics, x1000
h
Cluster
7 6 5 4 3 2 1
Agri, forestry, fish & hunt
Finance, insur, bus. services
Transp & communic
Wholesale, retail
Govern, edu, health
Mining, oil, manuf, constr
Other services
-100 0 100 200 300
Relative Change in occupation (%)

Figure S2

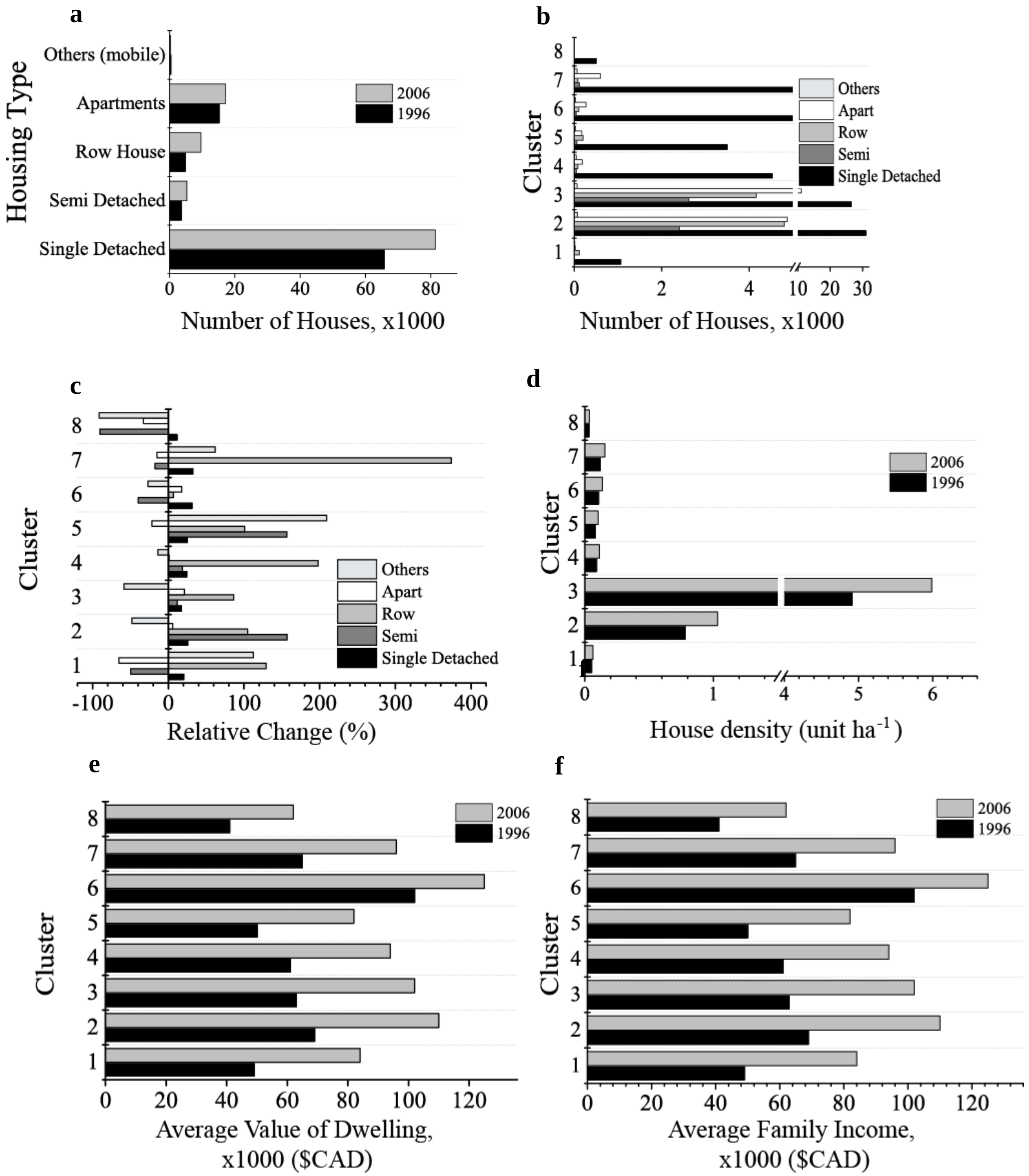

a
Housing Type
Others (mobile)
Apartments
Row House
Semi Detached
Single Detached
2006
1996
0
20
40
60
80
Number of Houses, x1000
b
Cluster
8
7
6
5
4
3
2
1
Others
Apart
Row
Semi
Single Detached
0
2
4
10
20
30
Number of Houses, x1000
c
Cluster
8
7
6
5
4
3
2
1
Others
Apart
Row
Semi
Single Detached
-100
0
100
200
300
400
Relative Change (%)
d
Cluster
8
7
6
5
4
3
2
1
2006
1996
0
1
4
5
6
House density (unit ha^{-1})
e
Cluster
8
7
6
5
4
3
2
1
2006
1996
0
20
40
60
80
100
120
Average Value of Dwelling, x1000 ($CAD)
f
Cluster
8
7
6
5
4
3
2
1
2006
1996
0
20
40
60
80
100
120
Average Family Income, x1000 ($CAD)

Figure S3

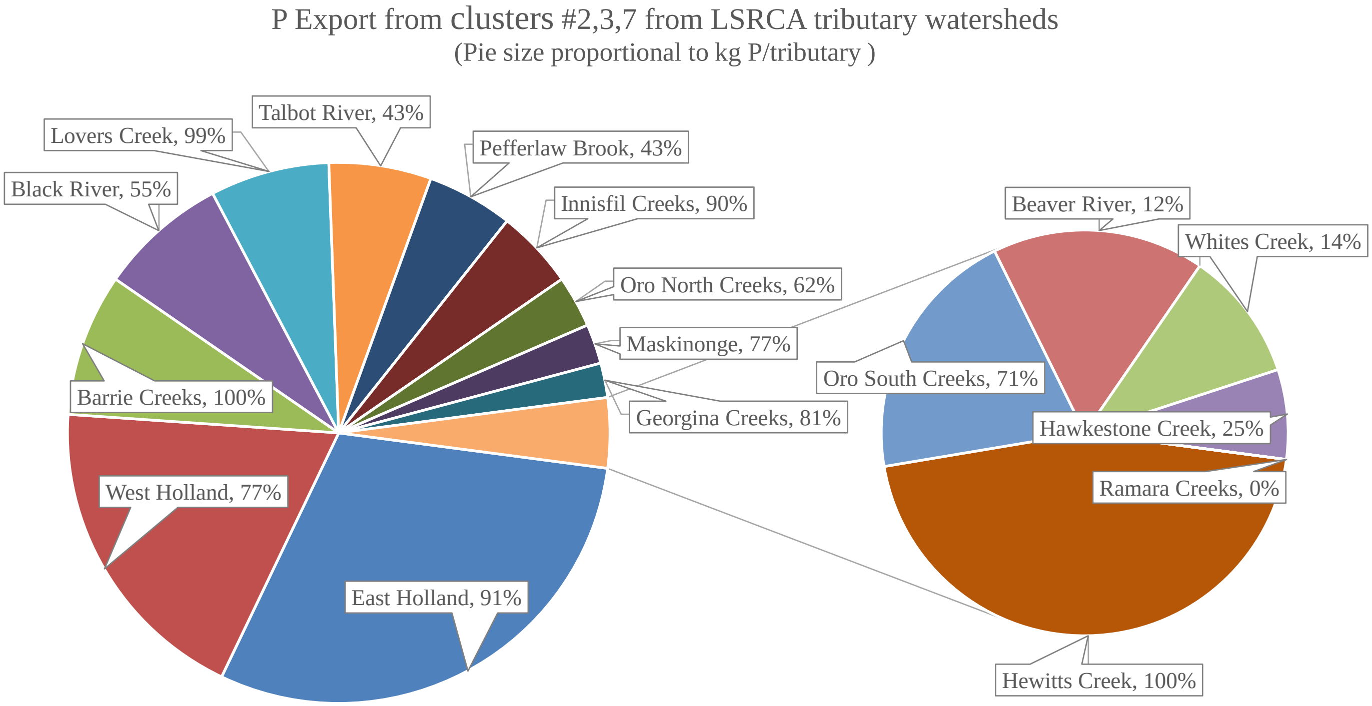

P Export from clusters #2,3,7 from LSRCA tributary watersheds
(Pie size proportional to kg P/tributary)
Talbot River, 43%
Lovers Creek, 99%
Pefferlaw Brook, 43%
Black River, 55%
Innisfil Creeks, 90%
Beaver River, 12%
Whites Creek, 14%
Oro North Creeks, 62%
Maskinonge, 77%
Oro South Creeks, 71%
Barrie Creeks, 100%
Georgina Creeks, 81%
Hawkestone Creek, 25%
Ramara Creeks, 0%
West Holland, 77%
East Holland, 91%
Hewitts Creek, 100%

Table S1. Goodness-of-fit statistics for the two structural equation models.

Goodness-of-Fit statistics	**SEM for clusters #1,4,5,8**	**SEM for clusters #2,3,6,7**
GFI	0.97 > 0.9	0.96 >0.9
CFI	0.96>0.9	0.98>0.9
NFI	0.95>0.9	0.97>0.9
NNFI /TLI	0.91	0.96
RMR	0.001	0.08
RMSEA	0.091<0.1	0.093<0.1

GFI (Goodness-of-fit index) = $1-\frac{F[S,\sum(\hat{\theta})]]}{F[S,\sum(O)]]}$, where $F[S,\sum(\hat{\theta})]]$ represent the minimum of the fit function after the model has been fitted; the denominator is the fit function before any model has been fitted (Joreskog and Sorbom, 1993). S – model sample covariance matrix S, ∑(θ)] – covariance structure for the observable variables, F[S,∑(θ)]] - fit function.

CFI (Bentler's Comparative Fit Index) = $\frac{(\chi_0 2-\mathrm{df}_0)-(\chi_t 2-\mathrm{df}_t)}{(\chi_0 2-\mathrm{df}_0)}$, where $\chi_0 2$ represents χ^2 fit statistics for the null model (Ullman, 2006) taking into account associated degrees of freedom df_0, while $\chi_t 2$ represents χ^2 fit statistics for the tested model with associated degrees of freedom df_t.

NFI (Bentler-Bonett Normed Fit Index) = $\frac{\chi_0 2-\chi_t 2}{\chi_0 2}$, where $\chi_0 2$ represents χ^2 fit statistics for the null model, while $\chi_t 2$ represents χ^2 fit statistics for the tested model.

NNFI / TLI (Non-normed NFI) = $\frac{\frac{\chi_0 2}{\mathrm{df}_0}-\frac{\chi_t 2}{\mathrm{df}_t}}{\frac{\chi_0 2}{\mathrm{df}_0}-1}$, where $\chi_0 2$ represents χ^2 fit statistics for the null model taking into account associated degrees of freedom df_0, while $\chi_t 2$ represents χ^2 fit statistics for the target model.

RMR (Root Mean Square Residual) - the average squared differences between the residuals of the sample covariances and the residuals of the estimated covariances under the assumption that the model is correct.

RMSEA (Root Mean Square Error of Approximation $=\sqrt{\frac{\frac{\chi 2}{\mathrm{df}}-1}{N-1}}$, where N – sample size, df – degrees of freedom, thus allowing to adjust for sample size in χ^2 statistics.

Jöreskog, K. G. and Sörbom, D., 1993.LISREL 8: Structural equation modeling with the SIMPLIS command language. Scientific Software International.

Ullman, J. B., 2006. Structural equation modeling: Reviewing the basics and moving forward. J. Pers. Asses., 87(1): 35-50.

Table S2: Model estimates for the structural equation model developed for clusters #1,4,5, and 8.

Parameters	**Estimate**	**Standard Error**	**Critical Ratio**
Regression Coefficients			
P_Export ⟵ WETLAND	-0.05	0.010	-5.198
P_Export ⟵ FOREST	-0.06	0.011	-5.279
P_Export ⟵ WATER	0.29	0.027	10.486
P_Export ⟵ RESIDEN-HD	0.73	0.046	16.093
P_Export ⟵ COM/IND	1.01	0.060	16.746
P_Export ⟵ CROPLAND	0.15	0.018	8.539
Covariances			
e2 ↔ e3	-0.01	0.001	-8.992
e4 ↔ e6	0.00	0.000	4.460
e5 ↔ e7	0.00	0.000	-4.073
e2 ↔ e6	0.00	0.000	-3.979
e2 ↔ e4	0.00	0.000	-2.816
e3 ↔ e5	-0.00	0.000	-5.779
Variances			
e2	0.01	0.001	13.928
e3	0.01	0.001	14.041
e5	0.00	0.000	13.837
e4	0.00	0.000	13.784
e6	0.00	0.000	13.784
e7	0.00	0.000	13.784
e1	0.00	0.000	13.784

Table S3: Model estimates for the structural equation model developed for clusters #2, 3, 6 and 7.

Parameters	**Estimate**	**Standard Error**	**Critical Ratio**
Regression Coefficients			
P_Export ← RESIDEN-HD	0.03	0.002	14.069
P_Export ← COM/IND	0.25	0.008	31.760
P_Export ←CROPLAND	10.83	1.380	7.853
CROPLAND ← H_LU	0.03	0.003	7.781
COM/IND ← H_LU	3.21	1.207	2.659
P_Export ← HAY/PAST	6.05	1.895	3.191
HAY /PAST ← H_LU	0.02	0.003	5.948
POP ← H_LU	0.60	0.284	2.078
COM/IND ← POP	0.10	1.447	0.068
RESIDEN-HD ← POP	7.935	0.429	18.514
POP ← LATITUDE	-87.288	48.426	-1.803
Covariances			
H_LU ↔ LATITUDE	0.000	0.000	-1.605
e2 ↔ e3	-78.13	16.811	-4.648
e3 ↔ e8	41.20	29.038	1.419
e3 ↔ e5	0.12	0.036	3.170
e2 ↔ e5	-0.28	0.112	-2.450
Variances			
H_LU	1.67	0.193	8.631
LATITUDE	0.000	0.000	8.631
e8	19.73	2.285	8.631
e2	595.71	69.017	8.631
e3	146.48	120.268	1.218
e5	0.003	0.000	8.631
e6	0.002	0.000	8.631
e1	0.99	0.115	8.631