Economic Activity, Species Occurrence and Spread: Evidence from Satellite Imagery Data

George Marbuah[a*], Ing-Marie Gren[a], Brendan G. McKie[b1], Laëtitia Buisson[c]

[a] Department of Economics, Swedish University of Agricultural Sciences, P.O. Box 7013, 750 07 Uppsala, Sweden

[b] Department of Aquatic Sciences and Assessments, Swedish University of Agricultural Sciences, P.O. Box 7050, 750 07 Uppsala, Sweden

[c] EcoLab, Université de Toulouse, CNRS, Toulouse, France

Abstract

In this paper, we address ecological and economic aspects in invasive species occurrence and spread. Specifically, we quantify the relative importance of the two factors driving the probability of occurrence of the aquatic invasive species *Elodea canadensis* Michx. across lakes in Sweden. We use satellite imagery to generate nighttime lights data as a proxy for economic activities to match ecological data on occurrences of the species at the catchment scale. A spatial probit model is used to explain the probability and dispersal of the species in lakes. With specific focus on the predictive ability of nighttime light on the invasion phenomenon, we find a robust positive relationship between economic activity and exotic aquatic invasion. This relationship is significantly characterized by spatial dependence.

Keywords: Elodea canadensis, invasive species, nighttime lights satellite imagery, GDP, spatial probit model, lakes, Sweden

[1] The authors acknowledge financial support for this study by The Swedish Research Council FORMAS (grant number 2011-836).

* Corresponding author: george.marbuah@slu.se

1. Introduction

The incidence and spread of biological invasion is well documented and still constitutes a major global concern. This is because invasive species can cause extensive environmental damage and impose substantial economic costs (see Pimentel *et al*., 2000; see Gren *et al*., 2009; Marbuah *et al*., 2014; Epanchin-Niell, 2017 for details on the economics of invasion).

Human-induced environmental change resulting in invasive species introduction, dominance and spread is a well-accepted fact in the literature (Taylor and Irwin, 2004; Keller *et al*., 2009). Intentional and unintentional introduction of exotic invasive species into new environments are transported through routes such as trade in exotic plants for aquaculture or horticultural production, recreational boating and fishing, swimming, organisms found in ballast water of ships or soil of potted plants, pests and viruses carried by humans, etc.

Growing empirical evidence suggests that the occurrence and spread of biological invasive species (including aquatic invaders) is mainly human-induced (see e.g. Johnson *et al*., 2001; Taylor and Irwin, 2004; Liu *et al*., 2005; Hulme, 2009; Keller *et al*., 2009; Gallardo, 2014). The role of economic activities in terms of e.g. trade and national prosperity have been examined in several studies (e.g. Vila and Pujadas, 2001; Gren *et al*., 2011). However, the lack of good quality data that adequately captures the degree of economic activities at the scale below well-known and established administrative boundaries such as country, municipality, region, etc. (e.g. at the lake-level) is a major challenge for many scientists interested in understanding this nexus. This constraint has made it daunting to model the invasion-socio-economic activity link and sometimes proven even elusive.

In this paper, we extract data on economic activities from nighttime lights (luminosity) satellite imagery data at the lake-level and link it to occurrence records of *Elodea canadensis* Michx. (hereafter *E. canadensis*) in Swedish lakes for which data

on *E. canadensis* presence-absence is available. Recent evidence show a strong correlation between gross domestic product (GDP), a measure of economic activity, and night light data at different scales and is currently being used extensively in many areas of the economics discipline (Ghosh *et al*., 2010; Chen and Nordhaus, 2011; Henderson *et al*., 2012; Michalopoulos and Papaioanna, 2014; Addison and Stewart, 2015; Mellander *et al*., 2015; Donaldson and Storeygard, 2016). To the best of our knowledge, there is not study using night light data as a proxy for economic activities to characterize pathway to invasion by non-native invasive species.

E. canadensis, native to North America, is invasive, not only in Sweden (mainly in the South), but in many regions of the world. It was first introduced into Europe in 1859 as an ornamental plant (aquarium and pond plant in botanical gardens) and rapidly spread throughout Europe (Josefsson and Andersson, 2001). It reaches very high densities in favourable environmental conditions (Tattersdill *et al*., 2017), often forming near-monocultural stands, spreads fast and can affect the entire lake ecosystem (Josefsson and Andersson, 2001). Its negative impact has been well documented and includes decrease in biodiversity (Mjelde *et al*., 2012), creation of problems for boat traffic, fishing (especially crayfish), swimming, and other recreational value of lakes where present in high abundance (Josefsson and Andersson, 2001; Zehnsdorf *et al*., 2015).

The main objective of this study is to examine the explanatory power of night light data as a proxy for economic activities in relation to climatic and habitat factors in driving the occurrence and spread of *E. canadensis* at the lake-level across Sweden[2] using spatial econometric modelling. Based on the literature, we formulate the following main research hypothesis: the presence-absence of *E. canadensis* species in lakes is closely linked to economic activities. Furthermore, since lakes are potentially networked or linked in some form, given the various vectors of spread, we postulate

[2] Sweden has over 40,000 lakes.

that the occurrence of this species can be characterized spatially. To validate our hypothesis, we use a spatial probit model to estimate the probability of *E. canadensis* occurrence in Sweden due to economic activities while controlling for ecological factors as well.

Our study is most closely related to the relatively scant literature that explicitly incorporates some measure of economic activities within a spatial setting (see Taylor and Irwin, 2004; Engler *et al*., 2004; Latimer *et al*., 2006; Carl and Kühn, 2007; Gren *et al*., 2011; Tamayo and Olden, 2014; Gallardo, 2014). However, many of these papers either ignore spatial dependence in the species data, exclude proxies or are uncertain about the appropriate measures for economic activities with the unit of analysis often at the regional or global scale. We fill the lacuna in the literature with a detailed within-country-wide lake-level analysis of aquatic invasive species incidence and dispersion while exploiting unique satellite imagery nighttime lights data found to be a strong candidate proxy for economic activities.

The remainder of the paper proceeds with the following structure. Data and methodological issues are discussed in Sections 2 and 3 respectively, while Section 4 presents a discussion of the main findings including sensitivity analysis. Concluding remarks are made in Section 5.

2. Description of Data

Nighttime lights[3] (obtained from remote sensing) provides a uniform, consistent, and independent estimate of economic activity (Doll *et al*., 2006; Elvidge *et al*., 2009). This imagery is based on the Defense Meteorological Satellite Program's Operation Linescan System (DMSP-OLS). The DMSP-OLS has a spatial resolution of 2.7km with image geo-located 30 arc-second grids (about 1 km^2 at the equator). The DMSP-

[3] Night lights data can be downloaded from http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html.

OLS data is taken from the U.S. National Geographic Data Center (NGDC) of the National Oceanic and Atmospheric Administration (NOAA). The process of data retrieval of night light from DMSP-OLS follows from Lowe (2014).

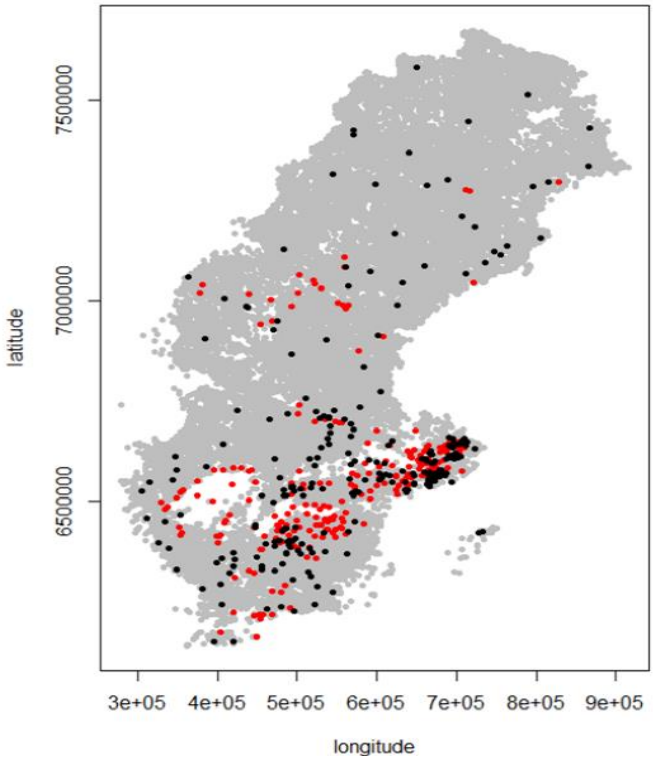
In this paper, nighttime lights imagery for the year 2013[4] is used to calculate average light for a pre-defined distance buffer (15km) around each lake as an estimate of economic activity[5]. Several distance buffers were experimented with but we found 15km distance radius to be the most appropriate since this captured the most light available around each identified lake. Average visible, stable and cloud-free lights coverage is extracted[6] and matched to the water surface polygon (i.e. at the lake-level and/or at the catchment) based on the GIS layer (geo-referenced information based on longitude and latitude coordinates) used to generate the *E. canadensis* inventory database that includes data on not only *Elodea* presence and absence but also corresponding lake-level environmental data including water chemistry, climatic conditions among others. For the purposes of modelling the probability of *E. canadensis* occurrence in at least 38,000 Swedish lakes assessed, a total matched data set of 498 observations comprising 257 recorded absences and 241 presence of this taxa of aquatic invasive species was recorded/retrieved. The distribution of *E. canadensis* in Sweden is shown on the map in Fig. 1. Also shown in the appendix is Table A1 with descriptive statistics of all the dataset.

To account for other potential drivers of this unique presence-absence species data set, following Buisson (XX) we control for cumulated length of shorelines of the water surface polygon (proxy for habitat availability of *E. canadensis*), accumulated number of days with mean daily temperature below 0°C of the water surface polygon (proxy for *E. canadensis* temperature limitation), water alkalinity and

[4] The most recent DMSP-OLS nighttime lights satellite data available.
[5] We transform the night lights data for estimation purposes as the natural log of (1+night lights), given some minimum values of zero (see Table A1).
[6] Matching of the two maps and extraction of night lights data was done using the Spatial Analyst toolbox in ArcGIS 10.3

nitrogen deposition of the water surface polygon. The motivation for selection of the identified controls is based on an on-going elaborate ecological niche study that is modelling the distribution of *E. canadensis* in Swedish lakes with the view to not only identify the key drivers/predictors of Elodea but also to assess the risk of invasion under the scourge of climate change (Buisson *et al*., 2018).

Fig. 1. Distribution of *E. canadensis* in Sweden



Note: ● Elodea present in lakes ● Elodea absent in lakes

Source: Buisson *et al*., (2018)

3. Regression Model Specification and Estimation Strategy

In order to identify the main drivers of the occurrence and dispersion of *E. canadensis* in all the lakes considered in this paper, we use a spatial probit econometric model. Specifically, we adopt the spatial autoregressive (lag) probit model (SAR) to estimate the drivers of the presence-absence aquatic invasive species data.

The general specification for a limited dependent variable SAR model is given as follows (LeSage and Pace, 2009; LeSage *et al*., 2011; and Wilhelm and de Matos, 2013). Let $\mathbf{y^*} = (y_1, ..., y_n)'$ be a vector of outcome variable (here *Elodea* presence-absence) and $\mathbf{X}(n \times k)$ be some vector of explanatory variables associated with the parameter vector $\beta(k \times 1)$. With a pre-specified $\mathbf{W}(n \times n)$ non-negative spatial weights matrix – which captures the dependence structure between neighbouring observations (i.e. lakes) –, the basic spatial probit SAR model is

$$\mathbf{y^*} = \rho \mathbf{W} \mathbf{y^*} + \mathbf{X}\beta + \varepsilon, \qquad \varepsilon \sim N(0, \sigma_\varepsilon^2 \mathbf{I}_n) \qquad (1)$$

where ρ is the spatial dependence parameter assumed in absolute value to be less than one (i.e. $|\rho| < 1$) and $\mathbf{W}\mathbf{y^*}$ a linear combination of neighbouring lakes. For identification purposes, the variance is typically set to one (i.e. $\sigma_\varepsilon^2 = 1$).

The data generating process (DGP) for the variable $\mathbf{y^*}$ is

$$\mathbf{y^*} = (\mathbf{I}_n - \rho \mathbf{W})^{-1} \mathbf{X}\beta + (\mathbf{I}_n - \rho \mathbf{W})^{-1} \varepsilon, \qquad \varepsilon \sim N(0, \mathbf{I}_n) \qquad (2)$$

If $\rho = 0$ or $\mathbf{W} = \mathbf{I}_n$, the above model reduces to an ordinary probit model. Characteristic of limited dependent variable models, $\mathbf{y^*}$ is considered as a latent (unobserved) variable. Rather, we observe only the binary variable $y_i(0,1)$ with outcome/response modality

$$y_i = \begin{cases} 1 \text{ if } y_i^* \geq 0, \\ 0 \text{ if } y_i^* < 0 \end{cases} \qquad (3)$$

where y_i reflects *E. canadensis* presence-absence in a particular lake.

As demonstrated by LeSage and Pace (2009) and noted in Wilhelm and de Matos (2013) among others, in spatial lag models in general and its probit variants, a change in a driver (x_{ir}) of the response variable (*E. canadensis* occurrence in a particular lake) will not only affect the outcome variable in the same lake y_i^* (i.e. direct effects) but may also possibly impact all other possible outcomes y_j^* (for $i \neq j$ -

i.e. indirect or spatial spillover effects - *E. canadensis* occurrence/spread into other lakes and possibly feedback into the original source). Thus consistent with the preceding, we also compute and analyze, after the main SAR probit model is estimated via Bayesian computations, summary measures of average direct, indirect and total effects as marginal effects to complement the analysis.

The specified SAR probit model of equation (1) is estimated via Bayesian techniques[7]. Bayesian estimation proceeds basically by sampling from a posterior distribution of the model parameters $p(\mathbf{y^*}, \rho, \beta | y)$ given data y and prior distributions $p(\mathbf{y^*})$, $p(\rho)$ and $p(\beta)$. Using Bayesian techniques allows the investigator to directly ask how probable are the study's hypotheses, given the data (Latimer *et al*., 2006). At the heart of Bayesian econometrics is the Bayes' rule which states that

$$p(\text{Parameters} \mid \text{Data}) = \frac{p(\text{Data} \mid \text{Parameters}) \times p(\text{Parameters})}{p(\text{Data})} \qquad (4)$$

where $p(\text{Parameters} \mid \text{Data}) \equiv p(x \mid y)$ is what is of fundamental interest (i.e. given the data, y, what do we know about x? – see Koop, 2003). In other words, $p(x \mid y)$ is the conditional probability of x on y. Importantly, the posterior probability distribution, $p(\text{Parameters} \mid \text{Data})$, gives a broad picture of what is certain or known about each relevant parameter given the model, data and prior information $p(\text{Parameters})$. Unlike the mean and confidence interval estimated in classical econometric analysis, the posterior distribution $p(\text{Parameters} \mid \text{Data})$ in Bayesian allows the analyst to explicitly interpret the results for a particular parameter in probabilistic terms (Koop, 2003 and Latimer et al., 2006 provide more intuition and technical details).

[7] Maximum likelihood and GMM estimation can also be implemented using spprobitml and spprobit in the R package **McSpatial** due to McMillen (2013). GMM works well only in very large samples albeit computationally more efficient as an IV method compared to Bayesian MCMC (Wilhelm and de Matos, 2013).

To proceed, the Markov Chain Monte Carlo (MCMC) and Gibbs sampler are used to facilitate the sampling of the posterior distribution $p(\mathbf{y}^*, \rho, \beta | y)$. The sample from the conditional densities $p(\mathbf{y}^* | \beta, \rho, y)$, $p(\beta | \mathbf{y}^*, \rho, y)$ and $p(\rho | \mathbf{y}^*, \beta, y)$ are obtained according to the following. Given the observed variable y (Elodea presence/absence data), driving forces and spatial parameters β and ρ, $p(\mathbf{y}^* | \beta, \rho, y)$ is a truncated multinormal distribution shown in equation (5)

$$\mathbf{y}^* \sim N\left((\mathbf{I}_n - \rho\mathbf{W})^{-1}\mathbf{X}\beta, \left[(\mathbf{I}_n - \rho\mathbf{W})'(\mathbf{I}_n - \rho\mathbf{W})\right]^{-1}\right) \quad (5)$$

subject to $y_i^* \geq 0$ for $y_i = 1$ and $y_i^* < 0$ for $y_i = 0$.

However, for a normal prior $\beta \sim N(\mathbf{c}, \mathbf{T})$, $p(\beta | \mathbf{y}^*, \rho, y)$ can be sampled from a multivariate normal as

$$p(\beta | \mathbf{y}^*, \rho, y) \propto N(\mathbf{c}^*, \mathbf{T}^*) \quad (6)$$

$$\mathbf{c}^* = (\mathbf{X}'\mathbf{X} + \mathbf{T}^{-1})^{-1}(\mathbf{X}'\mathbf{S}y^* + \mathbf{T}^{-1}\mathbf{c})$$

$$\mathbf{T}^* = (\mathbf{X}'\mathbf{X} + \mathbf{T}^{-1})^{-1}$$

$$\mathbf{S} = (\mathbf{I}_n - \rho\mathbf{W})$$

For the remaining conditional density, $p(\rho | \mathbf{y}^*, \beta, y)$ is distributed as

$$p(\rho | \mathbf{y}^*, \beta, y) \propto |\mathbf{I}_n - \rho\mathbf{W}| exp\left(-\frac{1}{2}(\mathbf{S}\mathbf{y}^* - \mathbf{X}\beta)'(\mathbf{S}\mathbf{y}^* - \mathbf{X}\beta)\right) \quad (7)$$

We used the MCMC and Gibbs sampler for the estimation of the SAR probit model. In all estimated models, 10,000 iterations were made and 2,000 burn-in samples drawn and discarded in each case until the model converges to the target distribution that computes the posterior means/probabilities for the parameters. For technical details, see LeSage and Pace (2009), LeSage et al., (2011), Wilhelm and de Matos (2013) and Elhorst et al., (2017). Finally, the spatial weights matrix $\mathbf{W}$ used for the estimation is the distance-based *k*-nearest neighbours approach where *k* denote the nearest number of neighbouring lakes assumed to be linked spatially that ensures dispersion/spread of the aquatic invader. Given no formal theoretical guide

in choosing the neighbours for the connectivity matrix, we experimented with different numbers as part of our sensitivity analysis but used 30-nearest lakes for the baseline estimates. All estimations were carried out using the R package **spatialprobit** (Wilhelm and de Matos, 2013).

4. Results and Discussion

Baseline

A discussion of the main SAR probit results is preceded with a brief analysis of some pre-tests for spatial autocorrelation in the *E. canadensis* species data which we argue (or suspect) might be spatially correlated or clustered. Specifically, the frequently used Moran's *I*[8] statistic was applied to the univariate species presence-absence data to detect any inherent spatial patterns. The results are shown in Table 1.

Table 1. Test for global spatial autocorrelation in *E. canadensis* species

Spatial weight matrix (*k*-nearest lakes)	Moran's *I*
k=1	0.346***
k=5	0.223***
k=10	0.192***
k=15	0.168***
k=20	0.154***
k=25	0.132***
k=30	0.112***
k=60	0.058***

Note: *** indicates statistical significance at the 1% level.

The test results are conclusive on the revealed spatial characteristics of the freshwater invader's data. Regardless of the definition of number of nearest neighbouring lakes (arbitrarily chosen numbers between 1 and 60) used to construct the spatial weights matrix, all results overwhelmingly affirm the fact that this species

[8] The global Moran's statistic (only one value calculated) is a spatial dependence measure that describes the overall spatial relationship across all the geographic units for the whole study area (Moran, 1948). The values of global Moran's *I* ranges between -1 and 1, where tends toward zero in the absence of spatial autocorrelation. In the case of positive spatial autocorrelation, the value of corresponds to a value *I* greater than zero while the reverse holds true for negative spatial autocorrelation.

is spatially interdependent. That is, we find highly significant positive spatial autocorrelation in the species occurrence data of *E. canadensis* across Swedish lakes. The implication is that there is a high probability that *E. canadensis* presence in a neighbouring lake does not provide any immunity for other lakes from nascent infection, reinfection and or further aggressive spread of *E. Canadensis* into an uninfected or already infected lake, given environmental/habitat conditions and known or unknown pathways/vectors of biological species invasion.

The rest of the analysis, given results of the preceding spatial effects tests, proceeds as follows. The results for the SAR probit model are shown in Table 2. Note that the results of the spatial lag probit are based on a 30-nearest lakes weights matrix. Further, given the main hypothesis of the paper, that the presence-absence of *E. canadensis* species in lakes is closely linked to economic activities, we focus the analysis on the outcome of this link while controlling for non-economic factors recommended by Buisson (XX). Thus, the night light variable (proxy for economic activities) is included in all estimated models.

Four SAR probit model specifications were estimated and shown in columns (1)-(4). Model comparison and goodness-of-fit assessment is made on the basis of the computed Akaike information criterion (AIC) and Schwarz Bayesian information criterion (BIC). The model with the lowest AIC and BIC is the one with the most explanatory power and hence chosen. We start with a simple bivariate model with only night light as the predictor of the probability of *E. canadensis* species occurrence. The next model (column 2) is the most over-parameterized and controls for accumulated length of shorelines, alkalinity of the water, nitrogen deposition and number of days with mean daily temperature below 0^{o}C. Specification (3) excludes the temperature variable but retains the remaining variables captured in model (2).

Table 2. MCMC SAR probit model posterior mean estimates

Variables	SAR probit				Ordinary probit
	(1)	(2)	(3)	(4)	(5)
Night light (log)	0.186***	0.104**	0.108**	0.117***	0.100**
	(0.042)	(0.045)	(0.047)	(0.045)	(0.046)
Shorelines length (log)		0.341***	0.350***	0.347***	0.380***
		(0.051)	(0.051)	(0.050)	(0.051)
Alkalinity of water		0.173	0.258***	0.257***	0.341***
		(0.106)	(0.096)	(0.096)	(0.110)
Nitrogen deposition (log)		-0.420	0.084		
		(0.310)	(0.127)		
Mean daily temperature $< 0^0 C$		-0.977*			
(# of days; log)		(0.547)			
ρ	0.751***	0.613***	0.624***	0.635***	
	(0.078)	(0.098)	(0.097)	(0.094)	
Intercept	-0.335***	5.286	-4.392***	-3.515***	-3.848***
	(0.084)	(5.615)	(1.400)	(0.457)	(0.470)
Observations	498	498	498	498	498
Log-likelihood	-355.09	-300.67	-301.04	-301.97	-301.49
AIC	716.18	615.35	614.09	613.94	610.99
BIC	728.81	644.82	639.35	634.99	627.83

Notes: Results are based on the Bayesian Monte Carlo Markov Chain (MCMC) estimation and Gibbs sampling for the spatial autoregressive (SAR) probit model. MCMC used 10,000 draws/replications/iterations and 2,000 burn-in samples to estimate the posterior probabilities (or means). The dependent variable is a dichotomous binary variable where 1 denotes Elodea presence and 0 absence in a particular lake. In all, there were 241 Elodea presence while 257 recorded absences. Rho (ρ) represents the spatial lag parameter and implies spread of Elodea within a particular lake and to other lakes. Spatial weights matrix constructed to show degree of connectivity among/between lakes is based on 30 nearest lakes using longitude and latitude coordinates at the lake level. ***, ** and * indicates statistical significance at 1%, 5% and 10% levels, respectively. In parenthesis are standard deviations of each estimated coefficient.

The model with the least AIC and BIC is specification (4) and appear to be the most parsimonious fit to the observed data. This specification includes information on night light, length of shorelines and water alkalinity. Based on the information criteria selected model (i.e. column 4), we estimated a corresponding standard probit model (column 5) which ignores any potential spatial dependencies in the invader species data within the lake network structure for comparison purposes. Even though model (1) is very parsimonious, it is not the most appropriate to form a basis for a convincing analysis and drawing conclusions. It obviously suffers from omitted

variables bias since the probability of invasion is not solely influenced by economic activities. Nonetheless, the results sure shows that even in the absence of ecological data, the impact of economic activities on the probability of invasion should not be overlooked.

We checked whether the spatial dependence parameter (ρ) added any value to our understanding of the biological invasion dynamics. Expectedly, and in agreement with our initial assessment of spatial effects in the presence-absence species data, the estimate of ρ is positive and statistically significant at the 1% level. The implication is that the threat of invasion contagion as a phenomenon is still real and spatial in nature and that our SAR model correctly describes this dynamic.

Overall, our baseline estimates show night light is positively related to the probability of occurrence of the *E. canadensis* species in Swedish lakes. These estimates are highly significant across all model specifications. This is true independent of type or number of controls in each equation. Thus consistent with our *a priori* hypothesis and the extant theoretical and empirical literature, human-induced economic activities is a major driver of the specific taxa under consideration in this paper. For example, our selected model (column 3) shows that holding all the other factors constant, economic activities has the potential to increase the probability of *E. canadensis* occurrence in lakes across Sweden by approximately 12% (see column 4). Compared to the traditional non-spatial probit model, it is evident that the probability of this effect is a bit smaller (about 10%). This potential attenuation bias (downward) might be due to the neglect of the spatial interdependence in the data. Our findings have in recent times been corroborated *inter alia* by Taylor and Irwin (2004), Liu et al., (2005), Hulme (2009) and Gallardo (2014). In particular, it compares quite well with Gallardo and Aldridge (2013) who found that the inclusion of socio-economic indicators in bioclimatic models could lead to a 20% increase in probability of biological invasion in general and a six-fold

increase in the predicted area suitable for quagga mussel (*Dreissena bugensis* or *Dreissena rostriformis bugensis*), an aquatic invader.

Table 3. MCMC SAR probit model marginal effects estimates

	(1)	(2)	(3)	**(4)**
Variable				
Direct effects				
Night light (log)	0.066***	0.034**	0.035**	0.040***
	(0.013)	(0.015)	(0.015)	(0.015)
Shorelines length (log)		0.112***	0.114***	0.117***
		(0.015)	(0.014)	(0.014)
Alkalinity of water		0.056	0.084***	0.087***
		(0.034)	(0.030)	(0.031)
Nitrogen deposition (log)		-0.136	0.027	
		(0.100)	(0.041)	
Mean daily temperature $< 0^0 C$		-0.317*		
(# of days; log)		(0.174)		
Indirect effects				
Night light (log)	0.212**	0.057	0.065	0.073*
	(0.089)	(0.036)	(0.040)	(0.043)
Shorelines length (log)		0.185**	0.206**	0.213**
		(0.077)	(0.085)	(0.089)
Alkalinity of water		0.094	0.152*	0.157*
		(0.071)	(0.084)	(0.086)
Nitrogen deposition (log)		-0.226	0.047	
		(0.200)	(0.081)	
Mean daily temperature $< 0^0 C$		-0.523		
(# of days; log)		(0.367)		
Total effects				
Night light (log)	0.278***	0.091*	0.100*	0.112**
	(0.092)	(0.048)	(0.052)	(0.053)
Shorelines length (log)		0.296***	0.320***	0.330***
		(0.079)	(0.087)	(0.091)
Alkalinity of water		0.150	0.236**	0.244**
		(0.100)	(0.104)	(0.108)
Nitrogen deposition (log)		-0.363	0.074	
		(0.286)	(0.119)	
Mean daily temperature $< 0^0 C$		-0.841*		
(# of days; log)		(0.507)		

Notes: ***, ** and * indicates statistical significance at 1%, 5% and 10% levels, respectively. In parenthesis are standard errors of each estimated coefficient.

As elaborated in the methodology section, the SAR model produces both direct and indirect (spatial spillover) effects, the sum of which gives the total effect of a change in any of the covariates in the model. Turning to the marginal effects of the SAR probit model, Table 3 sheds light on the differing nature of these impacts. Again, we focus on only model (4) and interpret only the night light parameter estimates. The results show that the direct marginal impact of a change in economic activities increases the likelihood of invasion in lake i by about 4% on the average but even more so on neighbouring lakes j and potential feedback effect by as much as 7%, giving a combined marginal effect of about 11%. This result means that the biological invasion process can be quite impactful in severity and spread facilitated by human-induced economic activities.

Robustness checks

To provide further support for the baseline outcome, model (4) in Table 2 is re-estimated under varying scenarios. The results are presented in Table 4. In column (1), the sum of night lights is used rather than average lights to proxy economic activities. The outcome is that this does not make much of a difference since the positive and significant effect of economic activities is still observed. In the remaining 10 other specifications (columns 2-11), we constructed several spatial weights matrices by changing the number of nearest lakes in the range of 1 to 70. The idea is to check whether our estimates of economic activities on *E. canadensis* occurrence is sensitive to any arbitrary choice. Again, the estimates are very much comparable to the baseline thus providing further validation to the main study hypothesis.

Table 4. SAR probit estimates based on sum of lights and varying number of nearest neighbours in the spatial weights matrix

Variables	*Light*	*Alternate choice of number of lake neighbourhood in the spatial weight matrix*									
	Sum of lights	1	5	10	15	20	25	40	50	60	70
	(1)	(2)	(3)	(4)	**(5)**	(6)	(7)	(8)	(9)	(10)	(11)
Night light (log)	0.083***	0.095**	0.100**	0.107**	0.110**	0.113**	0.117**	0.104**	0.094**	0.087*	0.086*
	(0.023)	(0.046)	(0.045)	(0.046)	(0.046)	(0.045)	(0.046)	(0.046)	(0.045)	(0.046)	(0.047)
Shoreline length (log)	0.294***	0.355***	0.345***	0.341***	0.350***	0.346***	0.346***	0.351***	0.364***	0.370***	0.378***
	(0.054)	(0.051)	(0.051)	(0.050)	(0.050)	(0.050)	(0.051)	(0.051)	(0.050)	(0.050)	(0.052)
Alkalinity of water	0.260***	0.303***	0.243**	0.253***	0.254***	0.248**	0.249**	0.293***	0.329***	0.336***	0.326***
	(0.093)	(0.102)	(0.098)	(0.095)	(0.095)	(0.096)	(0.096)	(0.097)	(0.097)	(0.098)	(0.097)
ρ	0.606***	0.305***	0.488***	0.564***	0.595***	0.620***	0.626***	0.648***	0.652***	0.661***	0.608***
	(0.097)	(0.053)	(0.069)	(0.070)	(0.073)	(0.076)	(0.084)	(0.110)	(0.125)	(0.135)	(0.161)
Intercept	-3.139***	-3.589***	-3.450***	-3.44***	-3.529***	-3.500***	-3.502***	-3.557***	-3.686***	-3.740***	-3.804***
	(0.480)	(0.467)	(0.463)	(0.456)	(0.461)	(0.462)	(0.470)	(0.467)	(0.463)	(0.459)	(0.473)
Observations	498	498	498	498	498	498	498	498	498	498	498
Log-likelihood	-296.79	-301.43	-299.27	-300.13	-298.49	-299.67	-300.48	-305.51	-308.97	-310.76	-309.13
AIC	**603.58**	612.86	608.54	610.26	**606.99**	609.35	610.96	621.03	627.94	631.52	628.26
BIC	**624.63**	633.91	629.59	631.31	**628.04**	630.40	632.01	642.08	648.99	652.58	649.32

Notes: Results are based on the Bayesian Monte Carlo Markov Chain (MCMC) estimation and Gibbs sampling for the spatial autoregressive (SAR) probit model. MCMC used 10,000 draws/replications/iterations and 2,000 burn-in samples to estimate the posterior probabilities (or means). The dependent variable is a dichotomous binary variable where 1 denotes Elodea presence and 0 absence in a particular lake. In all, there were 241 Elodea presence while 257 recorded absences. Rho (ρ) represents the spatial lag parameter and implies spread of Elodea within a particular lake and to other lakes. Spatial weights matrix constructed to show degree of connectivity among/between lakes is based on 30 nearest lakes using longitude and latitude coordinates at the lake level. ***, ** and * indicates statistical significance at 1%, 5% and 10% levels, respectively. In parenthesis are standard deviations of each estimated coefficient.

5. Conclusion

Propagule pressure associated with socio-economic activities has been suggested to affect the transportation, colonization, establishment and dispersal of biological invasive species at different scales (Gallardo, 2014). Several scientific studies have shown that inclusion of proxies for economic activities, jointly with ecological factors such as habitat, water chemistry and climate conditions significantly explain the likelihood of invasive species occurrence at multiple spatial scales (e.g. Taylor and Irwin, 2004; Gallardo, 2014). While the importance of economic activities in facilitating the process of invasion distribution in different habitats is widely acknowledged, it is sometimes omitted from risk assessments because it is challenging to measure and or quantify. This omission could then lead to substantial underestimation of the area at risk, scale of impact and hence management of these non-native species (Gallardo and Aldridge, 2013; Gallardo, 2014).

In this paper, we combine both geo-referenced ecological data and a novel proxy for economic activities in order to model and identify the main drivers of the occurrence of an aquatic invasive species *Elodea canadensis* Michx. in a spatially explicit probability model in Swedish lakes. We utilize nighttime lights luminosity imagery data to construct a proxy for economic activities by matching it to the presence-absence invasive *E. canadensis* species data across the entire Sweden. Results from our estimates based on a Bayesian spatial autoregressive probit model showed a highly significant positive relationship between economic activities and the probability of occurrence of the identified taxa of aquatic invader under investigation. The results are also highly characterized by spatial spillovers. This finding remained robust after controlling for ecological factors considered important in explaining the occurrence of *E. canadensis*. Sensitivity analysis based on different definition of spatial connectivity between lake networks and night lights further confirmed our main findings. To the extent that our results appear robust to some significant degree, satellite data capture holds promise in generating night lights to

plausibly proxy the extent of human economic activities on the introduction and spread of many invasive species at scales hitherto difficult to work with (e.g. lake, stream or river level) compared to say using well-known indicators such as gross domestic product (GDP) and human density at the country, subnational, regional or global scales.

References

Buisson, T., McKie, B.G., Tattersdil, K., Futter, M., Ecke, F. and Laffaille, P. (2018). The threat is not over yet: Towards a model of future distribution of an invasive species under climate change. Unpublished manuscript.

Carl, G. and Kühn, I. (2007). Analyzing spatial autocorrelation in species distributions using Gaussian and logit models. *Ecological Modelling*, 207, 159-170.

Chen, X. and Nordhaus, W.D. (2011). Using luminosity data as a proxy for economic statistics. *PNAS*, 108(21), 8589-8594.

Doll, C.N.H., Muller, J.P. and Morley, J.G. (2006). Mapping regional economic activity from night-time light satellite imagery. Ecological Economics, 57, 75-92.

Donaldson, D. and Storeygard, A. (2016). The view from above: Applications of satellite data in economics. *Journal of Economic Persoectives*, 30(4), 171-198.

Addison, D. and Stewart, B. (2015). Nighttime lights revisited: The use of nighttime lights data as a proxy for economic variables. Policy research working paper WPS7496, the World Bank, Washington DC.

Elvidge, C.D., Erwin, E.H., Baugh, K.E., Ziskin, D., Tuttle, B.T., Ghosh, T., Sutton, P.C. (2009). Overview of DMSP nighttime lights and future possibilities, IEEE proceedings of the 7th international urban remote sensing conference, Shanghai, China.

Elhorst, J.P., Heihnen, P., Samarina, A. and Jacobs, J.P.A.M. (20017). Transitions at different moments in time: A spatial probit approach. *Journal of Applied Econometrics*, 32, 422-439.

Engler, R., Guisan, A., and Rechsteiner, L. (2004). An improved approach for predicting of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology*, 41, 263-274.

Epanchin-Niell, R.S. (2017). Economics of invasive species policy and management. *Biological Invasions*, 19(11), 3333-3354.

Gallardo, B. (2014). Europe's top 10 invasive species: relative importance of climatic, habitat and socio-economic factors. *Ethology Ecology and Evolution*, 26(2-3), 130-151.

Gallardo, B. and Aldridge, D.C. (2013). The 'dirty dozen': Socio-economic factors amplify the invasion potential of 12 high risk aquatic invasive species in Great Britain and Ireland. *Journal of Applied Ecology* , 50, 757-766.

Ghosh, T., Powell, R., Elvidge, C.D., Baugh, K.E., Sutton, P.C. and Anderson, S. (2010). Shedding light on the global distribution of economic activity. *The Open Geography Journal*, 3, 148-161.

Gren, I.-M., Isacs, L., and Carlsson, M. (2009). Costs of alien invasive species in Sweden. *Ambio*, 38, 135–140.

Gren, I-M., Thierfelder, T., and Berglund, H., (2011). Country characteristics and non-indigenous species. *Environment and Development Economics* 16, 51-70.

Henderson, J. V., Storeygard, A. and Weil, D. (2012). Measuring growth from outer space. *American Economic Review*, 102(2), 994-1028.

Hulme, P.E. (2009). Trade, transport and trouble: Managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*, 46, 10-18.

Johnson, L.E., Ricciardi, A. and Carlton, J.T. (2001). *Ecological Applications*, 11(6), 1789-1799.

Josefsson, M and B Andersson (2001). The environmental consequences of alien species in the Swedish Lakes Mälaren, Hjälmaren, Vänern and Vättern. *AMBIO: A Journal of the Human Environment*, 30(8), 514-521.

Keller, R.P., Zu Ermgassen, P. and Aldridge, D.C. (2009). Vectors and timing of freshwater invasions in Great Britain. *Conservation Biology*, 23, 1526-1534.

Koop, G. (2003), *Bayesian Econometrics*. New York: John Wiley and Sons.

Latimer, A.M., Wu, S., Gelfand, A.E. and Silander Jr., J.A. (2006). Building statistical models to analyze species distributions. *Contemporary Statistics and Ecology*, 16(1), 33-50.

LeSage, J. and Pace, R. K. (2009). *Introduction to Spatial Econometrics*. New York: Taylor and Francis - CRC Press.

LeSage, J., Pace, R. K., Lam, N., Campanella, R. and Liu, X. (2011). New Orleans business recovery in the aftermath of Hurricane Katrina. *Journal of the Royal Statistical Society A*., 174, 1007-1027.

Liu, J., Liang, S-C., Liu, F-H., Wang, R-Q. and Dong, M. (2005). Invasive alien plant species in China: regional distribution patterns. *Diversity and Distribution*, 11, 341-347.

Lowe. M. (2014). Night lights and ArcGIS: A brief guide. Unpublished manuscript. https://economics.mit.edu/grad/mlowe.

Marbuah, G., Gren, I-M., and McKie, B. (2014). Economics of harmful invasive species: A review. *Diversity*, 6(3), 500-523.

McMillen, D. (2013). *McSpatial: Nonparametric spatial data analysis*. R package version 2.0, URL http://CRAN.R-project.org/package=McSpatial.

Michalopoulos, S. and Papaioannou, E. (2014). National institutions and subnational development in Africa. *The Quarterly Journal of Economics*, 151-213. doi:10.1093/qje/qjt029

Mellander, C., Lobo, J., Stolarick, K. and Matheson, Z. (2015). Night-time light data: A good proxy measure for economic activity? *PLoS ONE*, 10(10), 1-18. https://doi.org/10.1371/journal.pone.0139779.

Mjelde, M., Lombardo, P., Berge, D., and Johansen, S.W. (2012). Mass invasion of non-native E. canadensis Michx in a large, clear-water, species-rich Norwegian lake - impact on macrophyte biodiversity. *International Journal of Limnology*, 48, 225–240.

Moran, P.A.P. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B*, 10, 243–251.

Pimentel, D., Lach, L., Zuniga, R., and Morrison, D. (2000). Environmental and economic costs of nonindigenous species in the United States. *BioScience*. 50, 53–65.

Talyor, B.W. and Irwin, R.E. (2004). Linking economic activities to the distribution of exotic plants. *PNAS*, 101(51), 17725-17730.

Tamayo, M. and Olden, J.D. (2014). Forecasting the vulnerability of lakes to aquatic plant invasions. *Invasive Plant Science and Management*, 7(1), 32-45.

Tattersdil, K., Ecke, F., Frainer, A., and McKie, B.G. (2017). A head start for an invasive species in a strongly seasonal environment? Growth of *Elodea canadensis* in boreal lakes. *Aquatic Invasions*, 12(4), 487-498.

Vila, M., and Pujadas, J. (2001). Land use and socio-economic correlates of plant invasion in European and African countries. *Biological Conservation* 100, 397-401.

Wilhelm, S. and de Matos, M. (2013). Estimating spatial probit models in R. *R Journal* 5, 130-143.

Zehnsdorf, A., Hussner, A., Eismann, F., Rönicke, H., and Melzer, A. (2015). Management options of invasive E. nuttallii and E. canadensis. *Limnologica*, 51, 110–117.

Appendix

Table A1. Descriptive statistics of data

Variables	Mean	Std. Dev.	Min.	Max.	# of Obs
E. canadensis occurrence (*presence*=1)	0.48	0.50	0	1	498
Night lights (average)	10.84	15.44	0	63	498
Night lights (sum)	821.31	2650.84	0	26693	498
Length of shorelines	24026	71937	771	927279	498
Water Alkalinity	0.69	0.57	0.03	3.15	498
No. of days of water temperature $< 0^0 C$	89	26	45	199	498
Nitrogen deposition	31972	10565	6116	58837	498