

Geospatial analysis of oil discharges observed by the National Aerial Surveillance Program in the Canadian Pacific Ocean



Stefania Bertazzon^{a,*}, Patrick D. O'Hara^{b,c}, Olesya Barrett^a, Norma Serra-Sogas^d

^a Department of Geography, University of Calgary, 2500 University Dr. NW, Calgary, AB T2N 1N4, Canada

^b Canadian Wildlife Service, Environment Canada, Institute of Ocean Sciences, 9860 W. Saanich Rd., Sidney, BC V8L 4B2, Canada

^c Department of Biology, University of Victoria, Box 3020, Station CSC, Victoria, BC V8W 3N5, Canada

^d Department of Geography, University of Victoria, Box 3020, Station CSC, Victoria, BC V8W 3N5, Canada

A B S T R A C T

Keywords:

Oil pollution
Canadian Pacific Ocean
Poisson regression
Spatial analysis
Aerial surveillance
Prediction

Oil pollution resulting from day to day human maritime activities contributes a high portion of the overall input into marine environments, constituting a major threat to marine ecosystems worldwide. In Canada, the National Aerial Surveillance Program (NASP) extensively monitors and collects information on oily discharges using remote sensing devices. Despite the availability of data from NASP and other surveillance programs internationally, there is a paucity of spatial analyses of oil pollution patterns, particularly in their association with human marine pursuits. The objective of this paper is to analyze the association between observed oily discharges and human maritime activities in the Canadian Pacific Ocean. This study used Poisson regression to spatially model detected oily discharges with marine traffic, coastal facilities and proximity to coast. Further, it developed localized ('regional') models to address spatial heterogeneity. The models identify recreational activities, passenger traffic, commercial traffic, fisheries, and proximity to the coast as predictors of observed oily discharges. The regional models yield more accurate and reliable estimates of local associations, and identify more parsimonious sets of predictors for each region. By identifying and accounting for human activities most associated with oily discharge patterns, the models developed in this study could be used to estimate pollution rates in areas with less surveillance, and identify areas where NASP coverage may need to be increased. Spatially explicit rates estimated by these models can be used to monitor the effectiveness of programs and policy aimed at reducing discharge rates of oily pollution. This study can be used as a model approach for extending the analysis to the other coasts of Canada, using available NASP data.

© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Introduction

Oil pollution is a major threat to marine ecosystems and coastal communities worldwide (Charlier, Fink, & Krystosyk-Gromadzinska, 2012). Extensive research has proven this true for seabirds, which are highly visible and considered one of the most vulnerable taxa to marine oil pollution (Burger & Fry, 1993; Camphuysen & Heubeck, 2001). There are many ways that oil can enter these ecosystems, including natural seeps, vessel accidents, terrestrial run-off, and as a result of day to day activities associated with thousands of human maritime pursuits (GESAMP, 2007, 83 pp.; NRC, 2003, 265 pp.). Oily discharges can occur intentionally (see Camphuysen & Heubeck, 2001; Wiese & Robertson, 2004),

through negligence (i.e., poor maintenance for example), or purely by accident. In more extreme cases, oily discharges into marine environments can occur as a result of third party activities such as malicious (i.e., "interdictions" as per Church, Scaparra, & Middleton, 2004; see Anifowose, Lawler, van der Horst, & Chapman, 2012) or accidental (see Transport Safety Board of Canada, 2007) pipeline disruptions, or acts of war (see Iraq–Kuwait Conflict, p. 135 in NRC, 2003, 265 pp.). Nevertheless, in most if not all cases, these discharges would be considered non-compliant with international standards (i.e., MARPOL) and therefore, illegal as defined by Canadian legislation. Although land-based activities (i.e., through terrestrial run-off) and natural seeps are considered the largest inputs of oil pollution into marine environments in terms of volume per year, (GESAMP, 2007, 83 pp.; NRC, 2003, 265 pp.), research and enforcement efforts have focused on human maritime activities as they are easier to quantify and associate with specifically identifiable oily discharges.

* Corresponding author. Tel.: +1 403 220 7355.

E-mail address: bertazzs@ucalgary.ca (S. Bertazzon).

Oil pollution from shipping accidents tends to be catastrophic, capturing much of the public attention; however, cumulatively oil pollution stemming from day to day activities (“operational discharges”) contributes a higher rate of input into marine environments than pollution from shipping accidents (GESAMP, 2007, 83 pp.; NRC, 2003, 265 pp.). Most studies on impacts from operational discharges have focused on operational discharges and larger catastrophic spills from marine vessels (for example, see Camphuysen & Heubeck, 2001; Piatt, Lesink, Butler, Kendziorek, & Nysewander, 1990; Wiese & Robertson, 2004). However, day to day oily discharges also can be associated with outflows from terrestrial run-off, accidental spills from fuel docks, derelict vessels, coastal transfer facilities, pleasure craft, and fishery activities (GESAMP, 2007, 83 pp.; NRC, 2003, 265 pp.; NASP crew pers. comm.). Furthermore, size of discharge is not the only predictor of impact (Burger, 1993). Indeed, operational discharges can have devastating effects on highly mobile marine taxa such as seabirds, simply by virtue of timing and location. For example, over 50% of

the global population of Cassin’s Auklet (*Ptychoramphus aleuticus*) returns to Triangle Island (see Fig. 1) to breed. These breeding auklets tend to forage in relatively small areas of ocean over the shelf break approximately 60 km southwest or 80 km northwest of Triangle Island (Boyd, McFarlane Tranquilla, Ryder, Shisko, & Bertram, 2008), which are areas also transited by vessels moving between the lower U.S. and Alaska or Prince Rupert in northern B.C. (O’Hara & Morgan, 2006). A single operational oily discharge from any of these vessels could have a major impact on Cassin’s Auklet populations on a global scale. Camphuysen (1989), 322 p. provided an empirical example of the importance of timing and location of a spill when he reported on a massive die-off of Common Scoters (*Melanitta nigra*) and Common Eiders (*Somateria mollissima*), species which tend to aggregate in flocks along the shoreline, from a small coastal spill in the North Sea. Clearly, when considered cumulatively, oil pollution from operational discharges is likely to have a big effect on marine ecosystems as these discharges occur at a much higher frequency and are extensive spatially (Serra-Sogas,

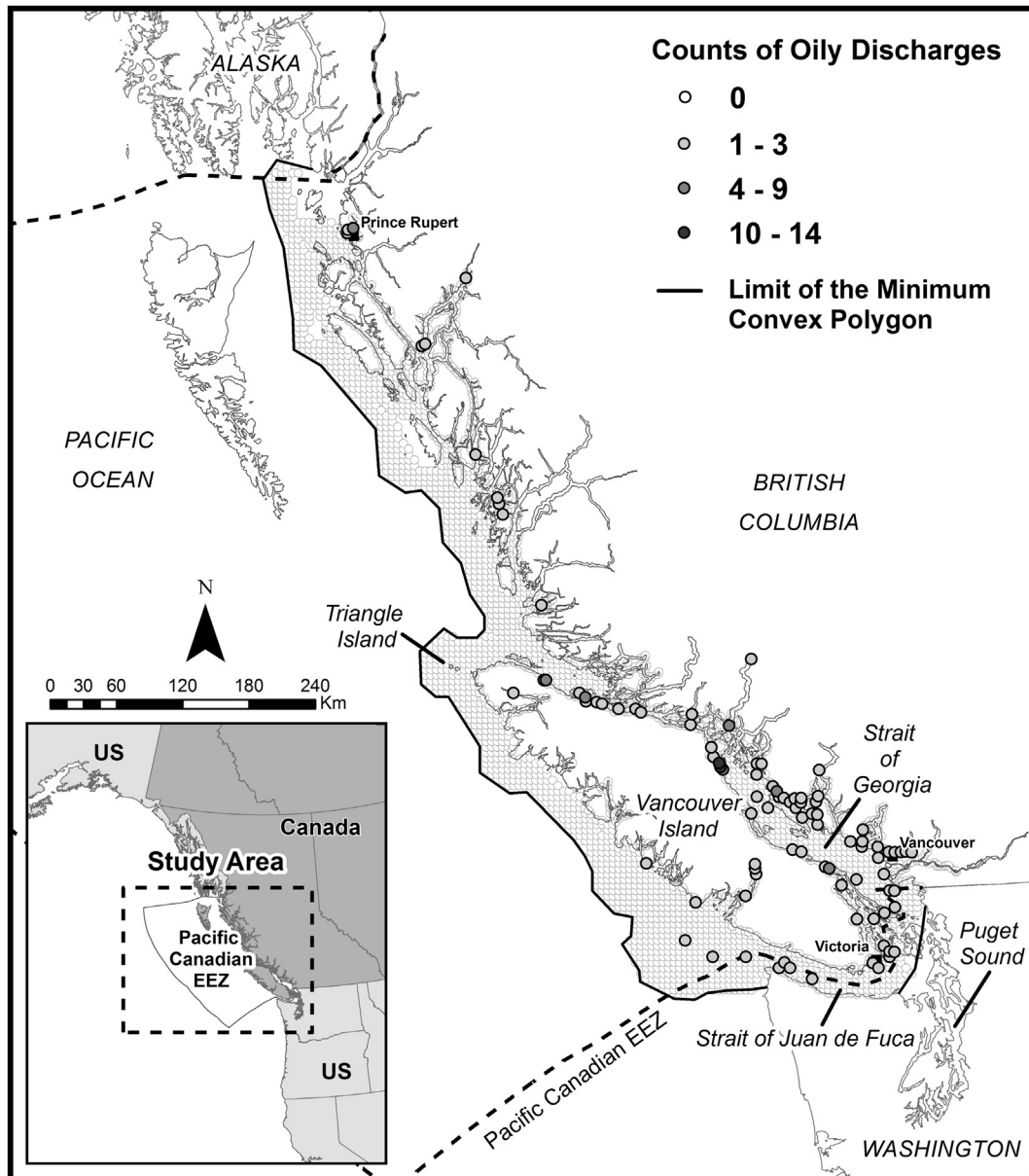


Fig. 1. Observed oily discharges in the Canadian Pacific region.

O'Hara, Canessa, Keller, & Pelot, 2008). It is, therefore, vitally important to integrate the spatial and temporal patterns of operational discharges, or any anthropogenic stressor, to fully estimate potential impacts associated with them.

To date, analyses estimating occurrence rates and impact of oil spills have focused on the larger spills (for examples on estimating occurrence see Meade, LaPointe, & Anderson, 1983; Ketkar & Babu, 1997; Vieites, Nieto-Roman, Palanca, Ferrer, & Vences, 2004; on impact estimation see Gundlach et al., 1983; Teal & Howarth, 1984; Piatt et al., 1990). It is generally difficult to model occurrence and estimate the impact of smaller oil spills because they are far more likely to go unreported and undetected. Both aerial and satellite surveillance techniques are increasingly employed and have been proven effective at monitoring and enforcing pollution law internationally (for example see Keramistoglou, Cartalis, & Kiranoudis, 2006 [Aegean Sea]; Carpenter, 2007 [Bonn agreement: North Sea]; Ferraro et al., 2007 [Adriatic Sea]; Backer et al., 2010 [HELCOM: Baltic Sea]; Wang, Gong, Pan, Hao, & Zhu, 2010 [China]; O'Hara, Serra-Sogas, Canessa, Keller, & Pelot, 2013 [Canada]). Although both monitoring techniques provide useful information for modeling the occurrence and impact of these smaller often unreported spills, there have been only a few studies analyzing these data statistically (Carpenter, 2007; Gade & Alpers, 1999; Volckaert, Kayens, Schallier, & Jacques, 2000), and fewer studies with spatially explicit components such as mapping of probability surfaces of oil pollution occurrence (Ferraro et al., 2007; Serra-Sogas et al., 2008) or association with possible sources such as vessel traffic intensity (i.e., inside or outside traffic separation schemes: Volckaert et al., 2000).

Transport Canada's National Aerial Surveillance Program (NASP) uses aircraft fitted with specialized monitoring equipment that includes remote sensing instruments to monitor and enforce pollution regulations in Canadian waters. This study analyzes oily discharges as detected by NASP in the Pacific region of Canada, using multivariate statistical methods to model these discharges with factors that may be associated with spatial patterns in their occurrence. The main objectives of this study were to: i) identify human activities that may be driving spatial patterns in oily discharges detected by NASP; ii) partition the study area into relatively homogeneous portions ('regions') and define localized models; iii) estimate the parameters linking observed oily discharges with each predictor, as these parameters can support the extrapolation of oily discharges to areas less covered by NASP. Similar modeling approaches have been proposed in the land use regression (LUR) literature (Gilbert, Goldberg, Beckerman, Brook, & Jerrett, 2005; Jerrett et al., 2005), which has emerged as a valuable technique for estimating local scale variability in air pollution, providing an accurate method for exposure analysis. LUR models estimate pollutant concentrations at fine spatial resolutions by regressing measured pollutant concentrations against land-use characteristics, such as surrounding traffic volume, potential industrial sources, and population density (Henderson, Beckerman, Jerrett, & Braurer, 2007; Jerrett et al., 2007). This study may constitute the first attempt at applying similar LUR modeling techniques in the marine environment. There is no evidence, in the surveyed literature, of use of localized modeling approaches; the definition of local (regional) models, presented in this paper, is therefore novel in this context.

Study area and data

Although the NASP covers all oceanic regions of Canada, our analyses are based on data collected in the Pacific region only, to simplify analyses and because of authors' familiarity to this area. The Canadian Pacific is a large region where oily discharges are a

generally rare occurrence, exhibiting a sparse spatial distribution and visually discernible clustering in proximity of the coast, as summarized in Fig. 1.

Analyses were based on data collected from 2008 to 2010 (inclusive) by NASP in the Pacific region. All analyses were based on oily discharge counts rather than total estimated volume, in part because quantity of discharge generally is not a good predictor of ecological impact (Burger, 1993), and because count data were considered more reliable than data on estimated volume discharged for the models discussed in this paper (NASP crew pers. comm.). For analytical purposes, the study area was divided into 5×5 km grid cells. Cell size was determined in consideration of size of the study area and spacing of observed discharges. Oil discharge counts for each year were summed up in each cell, and the total was used as regression response variable: the Pacific portion of the Canadian economic exclusive zone (EEZ) contains 37,162 grid cells with 101 observed discharges over the three years (Fig. 1).

The regression analysis (see below) was guided by a combination of expert opinion and spatial exploratory techniques (i.e., Getis-Ord G_i^* , Moran's I). Expert opinion was harnessed through consultation with members of the Oil in Canadian Waters (OCW) research working group, and members of the NASP as well as the Marine Aerial Reconnaissance Team (MART) with years of experience detecting oily discharges in the Pacific region of Canada. The MART is responsible for the detection and documentation of oily discharges using the Marine Surveillance System or MSS6000 developed by the Swedish Space Corporation. The MSS6000 has been installed on all NASP aircraft since 2007–2008. The OCW consisted of researchers with experience with the issue of smaller scale oily discharges and modeling risk associated with ship traffic from federal government departments (Transport Canada and Environment Canada)¹ and universities (University of Victoria, University of Calgary, Mount Allison University and Dalhousie University).²

Exploratory spatial data analyses indicated that the majority of observed oily discharges occur near shore, with very few occurrences farther than 20–30 km from the coastline. Expert opinion also suggested that near-shore discharges are likely determined by different processes than those governing offshore discharges. Therefore, it was decided to define a minimum bounding or minimum convex polygon (MCP) (Nilsen, Pedersen, & Linnell, 2008) around observed near-shore discharges to delineate the study area, in order to reduce the number of null observations. After experimenting with various buffer sizes, again in consultation with the same experts described above, a buffer was selected of 25 km from the coast and 50 km from the farthest near-shore observed discharge. The resulting polygon contains 3414 grid cells, and 96 discharges³ (Fig. 1).

For the multivariate regression models of oil discharges, two categories of predictor variables were considered: marine traffic (i.e., presence of different vessel types), and proximity to other human maritime activities (i.e., distance from coast, distance from ports, and number of marinas), in addition to surveillance effort (see Table 1). Surveillance data were provided by the NASP along with oily discharge data. The variable "ViewArea" represents NASP surveillance effort in square kilometers per cell. Yearly data were summed up, as with oily discharge counts. Surveillance effort is

¹ Louis Armstrong, John Heiler, Suxanne Baumeler, Kim Pearce, Ralph Hilchie, Patrick O'Hara, and Sabina Wilhelm.

² Rosaline Canessa, Norma Serra, Stefania Bertazzon, Olesya Barrett, Abul Azad, David Lieske, Ron Pelot, Andrew Szeto, and Casey Hilliard.

³ 5 of the 101 observed discharges were considered offshore and excluded from the analysis.

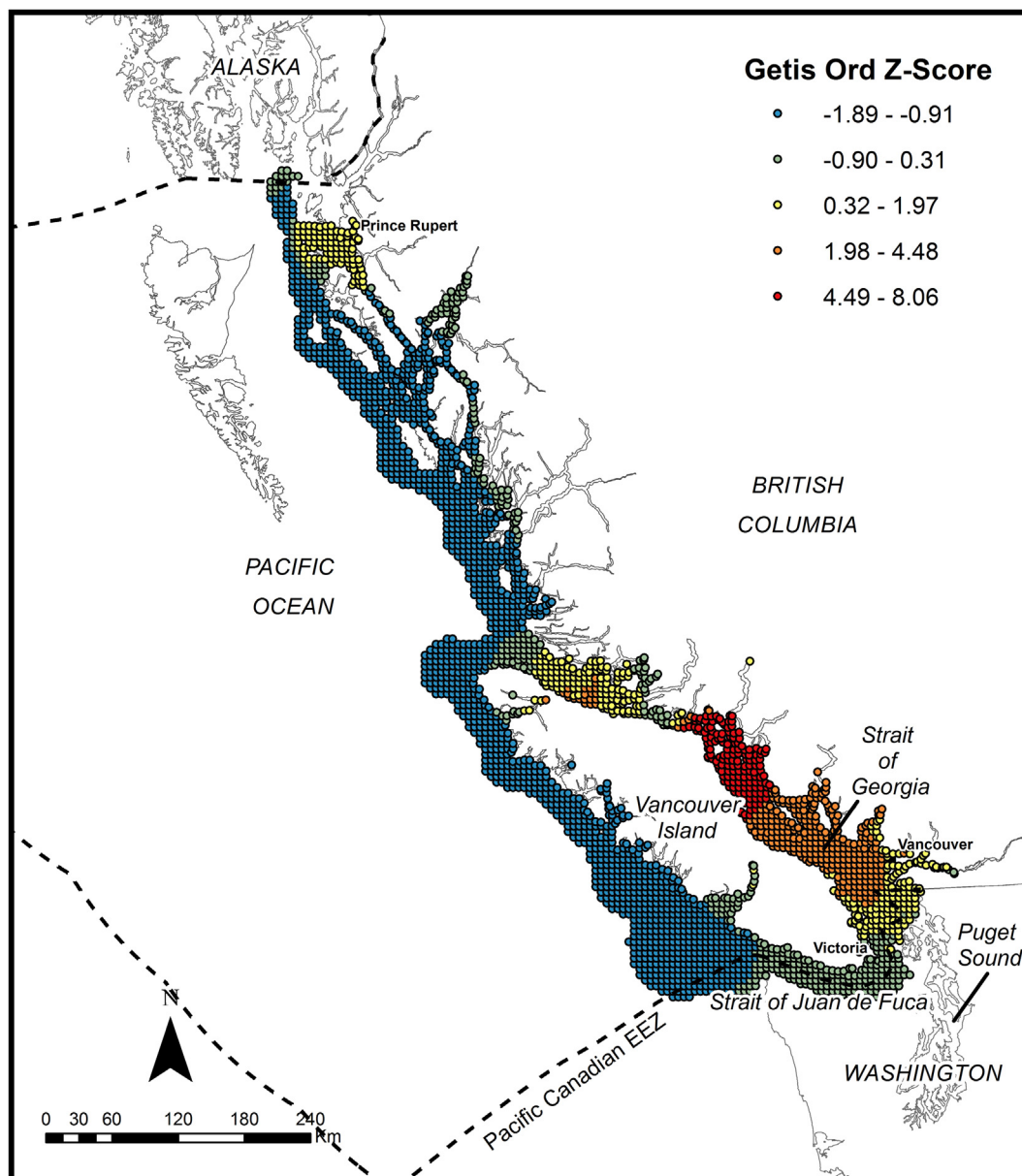


Fig. 2. Spatial clustering analysis.

viewed as a precondition for the observation of oily discharges, and therefore is treated as an offset variable in many of our regression models (see 'Methods' below).

Minimum distance or proximity to coast or port was measured as the distance from each cell centroid to the closest point on the coastline, or closest port, using the Euclidean metric (straight line) in ArcGIS (ArcMap version 10). Top commercial ports in the region were selected based on volume of domestic and international cargo handled in 2007 and 2008 (AAPA). In addition, a few U.S.A. ports were included, due to their proximity to Canadian waters. Also calculated per cell was the average distance from all ports.

The total number of marinas per cell was also determined ("Marina"). Marinas were defined as small craft harbors, which include marine recreation facilities (i.e., marinas, yacht clubs and public docks), fishing harbors, and ferry docks. Location information on these facilities was provided by the Department of Fisheries and Oceans. A simple count of the number of small harbors per cell was performed using ArcGIS (ArcMap 10). Total marinas per cell

and coastal proximity were treated as independent, or predictor variables in regression models, to analyze the association between chronic oil pollution and boating activities occurring in proximity to those facilities and the coast. These variables were specific to the process linked to near-shore oily discharges.

A rich marine traffic database, summarized in Table 1, was prepared by OCW research working group partners (Ron Pelot and Casey Hilliard, Marine Activities Risk Information Network [MARIN – <http://www.marin-research.ca/>], Dalhousie University). Each traffic variable represents the total number of hours spent in each grid cell per vessel type or category. Categories were based on ship type, ship length, flag of convenience and ship age. Flag of convenience refers to vessels that are registered to nation-states that are different from the nation-state of the ship owner, often to avoid high registration fees, taxes and tighter regulatory controls. Ship age refers to vessels built before or after 1992 – this threshold age of vessel was considered important based on a report on sub-standard shipping published by the Organization of Economic

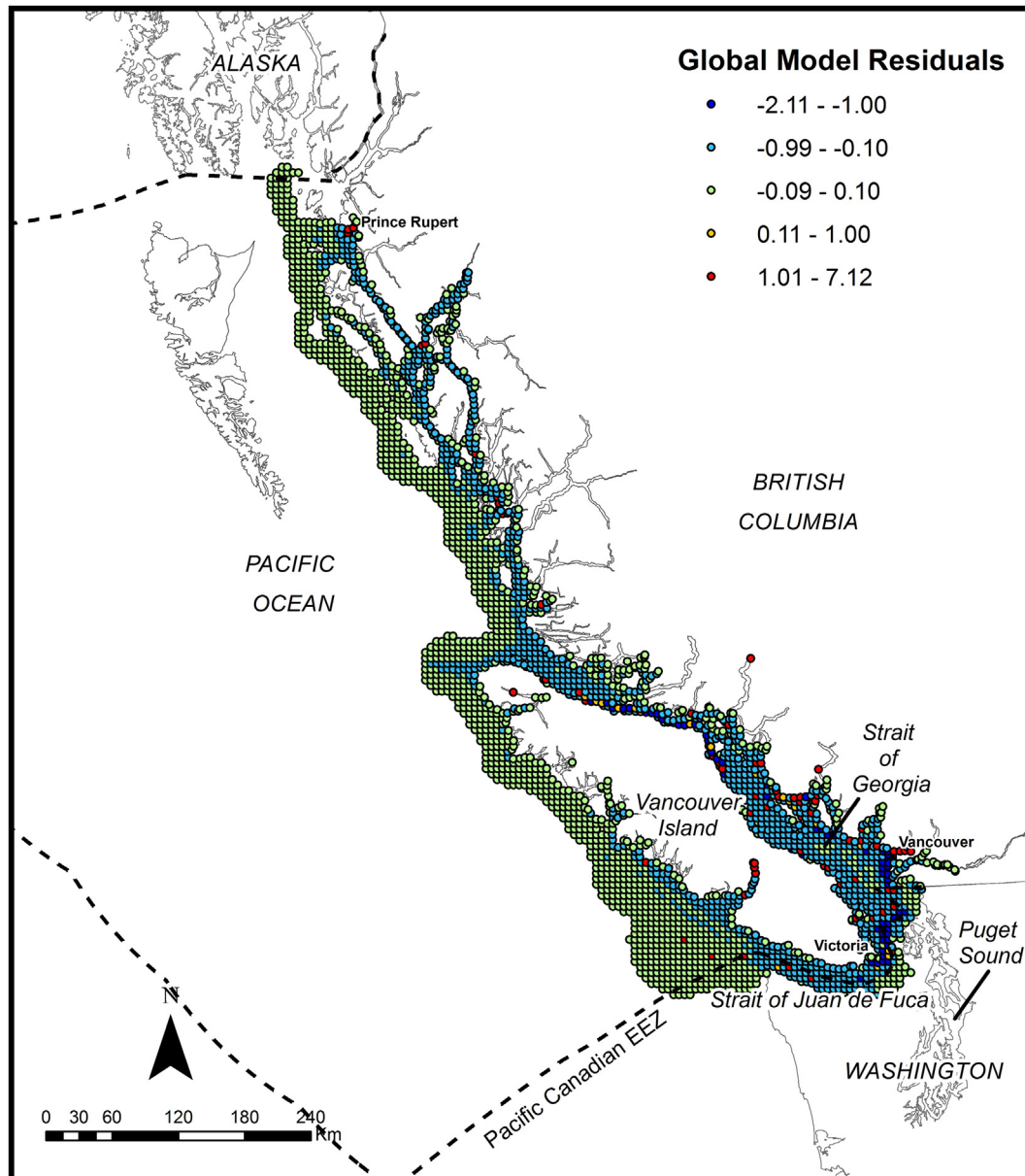


Fig. 3. Model residuals.

Cooperation and Development (OECD, 2002, 53 pp.) that stated vessels generally became less compliant with international regulations at about 20 years of age. Marine traffic data were available over the same period as NASP data (2008–2010); likewise, yearly data were summed up per grid cell.

Our multivariate analysis tested for significant association between total count of oily discharges detected (response variable) and this subset of predictor variables.

Methods

Regression analysis is a fundamental multivariate technique, where a response variable is conceptualized as a function of a set of predictor variables (Burt, Barber, & Rigby, 2009). In this study, all the variables are spatial, hence prone to spatial dependence and heterogeneity (Anselin, 1998). The dependent variable (observed oily discharges) consists of positive integers, exhibiting few non-null low values and a large proportion of null values (Fig. 1): such variables

are best modeled by Poisson regression (Burt et al., 2009). The Poisson regression model assumes equidispersion of the response variable; that is, the expected mean is equal to the expected variance. When this assumption is violated and the variance is greater than the mean, the distribution is considered overdispersed and treated accordingly (McCullagh & Nelder, 1989). It also assumes that the logarithm of the response variable changes linearly with equal increment increases in the predictor variables (McGree & Eccleston, 2012). It further assumes that changes in the dependent variable, from combined effects of different independent variables, are multiplicative. The specification of a Poisson regression model may include a so-called offset variable. As shown in Eq. (1), the offset variable (os) enters as denominator of the dependent variable.

$$\ln(y/os) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (1)$$

Model specification was guided by the same experts described above (NASP and MART members, and OCW partners). Prior to estimating any regression, cross-correlation across response and

Table 1
Predictor variables.

Category	Data field	Description
Surveillance	ViewArea	Total surveilled area (sqkm)
	Visits	Number of visits(count data)
Proximity	CoastDist	Shortest distance to shore (m)
	PortShortDist	Shortest distance to port (m)
	PortAvgDist	Average distance to ports (m)
	Marina	Harbor density (no. facilities/sqkm)
Marine Traffic	Traffic	Total hours of traffic
	Passenger	Hours of passenger and cruise vessels
	Fishing	Hours of fishing vessels
	Government	Hours of government vessels
	Recreation	Hours of pleasure vessels and yachts
	Cargo	Hours of merchant (bulk and cargo) vessels
	Tank	Hours of tanker vessels
	Research	Hours of research vessels
	Tug	Hours of tug and service vessels
	Length	L50-
L50-100		Hours of vessels of $50 < \text{length} \leq 100$ m
L100-200		Hours of vessels of $100 \text{ m} < \text{length} \leq 200$ m
L200+		Hours of vessels of length > 200 m
Flag	FOC	Hours of flag of convenience vessels
	NonFOC	Hours of non-flag of convenience vessels
Age	Built<92	Hours of vessels built before 1992
	Built>92	Hours of vessels built in 1992 and after

predictor variables was tested using Pearson's coefficient (Burt et al., 2009). In an effort to control for multicollinearity, which implies extreme instability of model estimates (Gujarati, 2003), when two predictors exhibited high cross-correlation (greater than 0.65 in absolute value), only one was maintained in the model. Whenever possible, the predictor most highly correlated with the response variable was chosen; in some cases, two or more alternative models were specified, each one containing alternative sets of uncorrelated predictors. The Pearson's correlation coefficient, however, provides only an indication of the linear correlation between predictors, but high correlation may still occur within the regression due to a non-linear relationship among predictors (Eq. (1)). Therefore, the correlation of estimated coefficients was analyzed and, likewise, in the case of highly correlated coefficients, one of the predictors was removed. Further, each regression was refined using a backward selection procedure, which excluded predictors that were not significantly associated with the response variable. Akaike information criterion (AIC) (Gujarati, 2003) was used, in conjunction with other regression diagnostics, to compare and select alternative models. McFadden pseudo R^2 (Veall & Zimmermann, 1996) was also calculated for each model; it is computed as the ratio between the log likelihood of the null model (constant only) and the log likelihood of the model with the selected set of predictors. It can be adjusted to take into consideration the number of predictors and, similar to AIC, it favors more parsimonious models. The assumption of equidispersion was assessed by calculating the difference between mean and variance, and t -tests and Wilcoxon Rank Sum tests were computed.

Lastly, the Poisson regression model assumes independence of observations. Meeting this assumption may be problematic with spatial data, which tend to exhibit spatial heterogeneity and spatial dependence (Anselin, 1998). In the practice of spatial analysis, a number of exploratory and analytical tools are generally used to test these properties on model variables, and to assess the error properties once the model is estimated. However, most of these tools employ inferential procedures that assume normally distributed variables and residuals (Anselin, 1995; Lin & Zhang, 2007). Although some procedures exist for statistical testing of non-normally distributed data (Lin & Zhang, 2007), admittedly they are generally not reliable (Augustin, McNicol, & Marriott, 2006;

Dormann et al., 2007; Dray, Legendre, & Peres-Neto, 2006). Griffith (2010) discusses the validity of computing the Moran coefficient for non-normally distributed variables; however, he notes the need to extend his results to function-based test statistics. For this reason, classical spatial exploratory tools (i.e., Moran's I and Getis-Ord G) were used, even though the variables used in this study are not distributed normally. While these spatial analytical tools cannot be used for inference on these data, they still provide an indication of those key spatial properties.

Spatial heterogeneity was analyzed using Getis-Ord general G statistic (Getis & Ord, 1992), which provides an assessment of the degree of spatial clustering. Significant positive values of the z-score associated with G indicate clustering of high values; significant negative values of the z-score indicate clustering of low values (Getis & Ord, 1992). Non-significant values of the z-score indicate spatial randomness, or independence of observations. Spatial dependence was assessed by spatial autocorrelation analysis, which indicates whether an observed spatial variable displays self-similarity over short distances. Moran's I (Getis, 2008) spatial autocorrelation index was used. Moran's I ranges from -1 for perfect negative spatial autocorrelation (dispersion, or regular pattern) to $+1$ for perfect positive spatial autocorrelation (clustering), with a value of 0 indicating absence of spatial autocorrelation, or spatial randomness (Getis, 2008). It is known that spatial autocorrelation analysis tends to produce significant results when conducted on a raster of cells where only few cells present non-null values (Boots & Tiefelsdorf, 2000); tests therefore were conducted on the whole dataset, on reduced datasets, where arbitrary portions of null values were removed, and on the reduced set containing only non-null values.

When the assumptions of identical and independent distribution of the residuals are violated by spatial heterogeneity or spatial dependence, spatial regression techniques yield more reliable models than standard regression techniques. Spatial regression techniques include geographically weighted regression and spatially autoregressive methods (Anselin, 1998; Fotheringham, Brundson, & Charlton, 1998). These modeling techniques have been extended to the Poisson model: generalized forms of Poisson include spatially weighted generalized models (Lambert, McNamara, & Garrett, 2006; Mohebbi, Wolfe, & Jolley, 2011), and spatially autoregressive generalized models (Nakaya, Fotheringham, Brundson, & Charlton, 2005). Even with generalized models, statistically assessing the spatial properties of model residuals remains problematic due to the non-normality of their distribution (Leung, Mei, & Zhang, 2000).

Because of all these difficulties, independence of observations could not be assessed conclusively. It was preferable to use a standard Poisson regression model, while the G statistic was employed to visually identify spatial heterogeneity in the response variable and in the model residuals. These analyses lead to the definition of relatively homogeneous regions, and for each region individual models were estimated. Spatial heterogeneity suggests that distinct processes are locally associated with oily discharges: individual models estimated for each region are expected to yield more accurate and reliable results.

In delineating the borders of the three regions (see 'Results' below), it was observed that no discharges were recorded on the northern portion of the west coast of Vancouver Island, located at the intersection of the three regions; for this reason, this portion of coast was omitted in the regional analyses. The three regions can be broadly labeled as: Region 1, or the northern coast; Region 2, or the Strait of Georgia; and Region 3, or the south-western coast of Vancouver Island (see Fig. 4, below). For each of the three regions, alternative models were specified, using the same methods employed for the entire study area; cross-correlation patterns

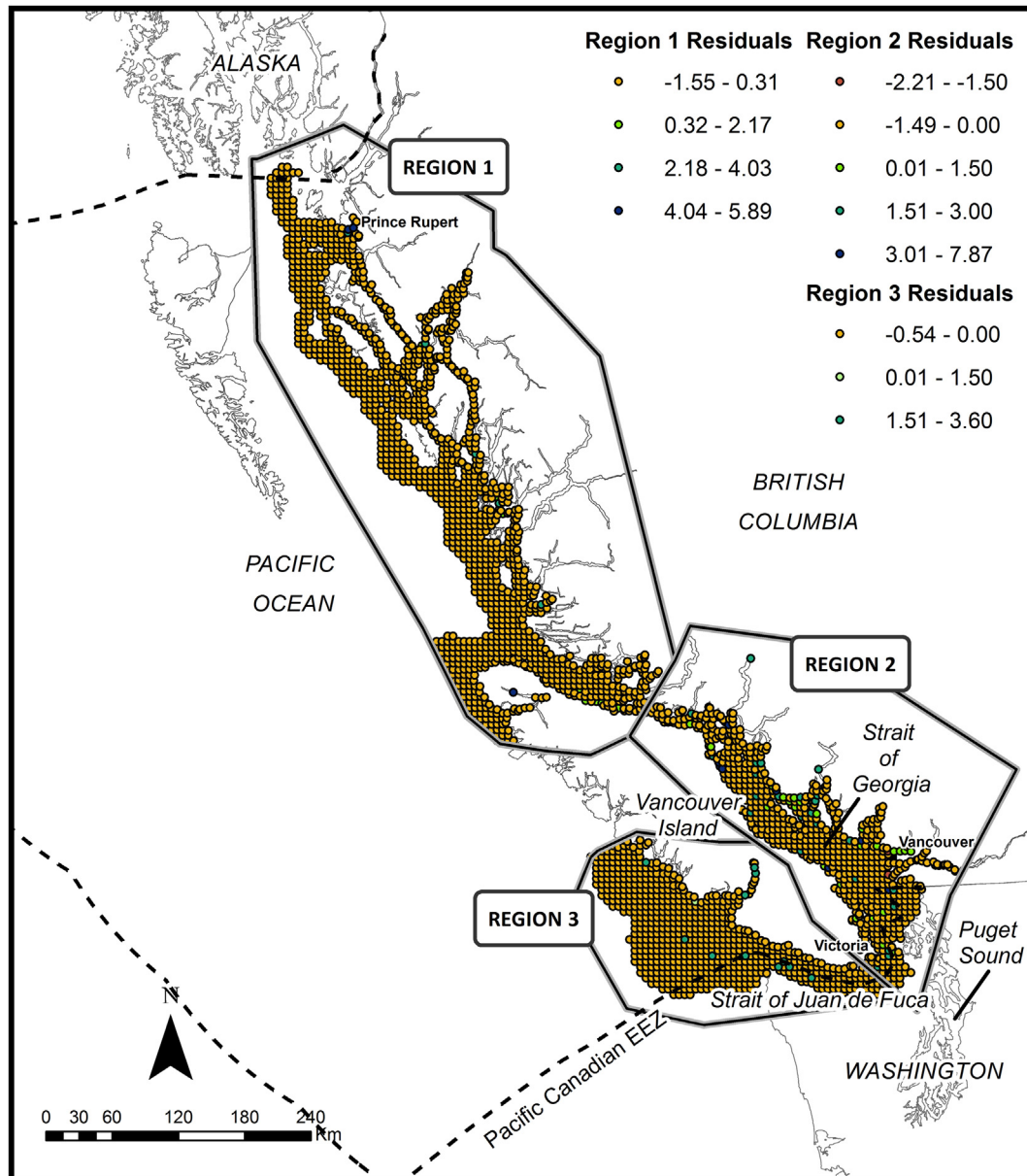


Fig. 4. Identified regions and regional model residuals.

within each region were considered in each model specification. In inspecting the three regions individually, it was observed that the surveillance effort covers the entire Regions 2 and 3, whereas its spatial coverage is much more limited for Region 1. Consequently, effort as an offset variable was considered redundant and removed from the models for Regions 2 and 3.

Unavoidably, spatial analytical results are subject to the modifiable areal unit problem (MAUP) (Fotheringham & Wang, 1991; Openshaw & Alvanides, 1999), and their interpretation is prone to ecological fallacy, i.e., relationships found at the aggregate level cannot be inferred to the individual level (Anifowose et al., 2012; Robinson, 2009). In consideration of these problems, all inferences in this study are only drawn for the spatial units defined above. Further, exploratory analyses (not reported in this paper for brevity sake) were conducted on the whole Canadian Pacific EEZ, on a minimum convex polygon of 200 km from shore, as well as on variously 0-reduced sets. While these analyses cannot be

directly compared, it can be said that all the results were relatively consistent.

Results

All the analyses reported in this study were conducted on 5×5 km grid cells on the minimum convex polygon shown in Fig. 1. At the spatial scale of analysis (Fig. 1), observed oily discharges are Poisson-distributed and equidispersed. Exploratory spatial analysis does not provide conclusive evidence of spatial heterogeneity. The G statistic indicates that clustering occurs only in parts of the region. Most notable is positive clustering in the Strait of Georgia, whereas minor areas of negative clustering are observed in the northern and southwestern parts (see Fig. 2). Spatial autocorrelation was inspected on the complete minimum convex polygon and on variously 0-reduced sets (Boots & Tiefelsdorf, 2000), and calculated for 1 to 30 nearest neighbors. The z statistic associated

Table 2
Correlation between response and predictor variables.

	Discharges	ViewArea	Traffic	Passenger	Fishing	Recreation	Cargo	Tank	Tug	Built < 92	Built > 92	Marina	CoastDist
Discharges	1	0.02	0.24**	0.14	0.16	0.34**	0.11	0.01	0.21**	0.10	0.12	0.15	-0.07
ViewArea	0.02	1	0.11	0.00	0.18*	0.05	0.32**	0.05	0.03	0.30**	0.18*	-0.03	0.52**
Traffic	0.24**	0.11	1	0.73**	0.36**	0.47**	0.56**	0.42**	0.68**	0.56**	0.57**	0.01	-0.12
Passenger	0.14	0.00	0.73**	1	0.08	0.20**	0.16	0.06	0.13	0.11	0.10	0.00	-0.07
Fishing	0.16	0.18*	0.36**	0.08	1	0.45**	0.28**	0.07	0.38**	0.26**	0.23**	-0.02	-0.06
Recreation	0.34**	0.05	0.47**	0.20**	0.45**	1	0.29**	0.18*	0.41**	0.34**	0.38**	0.08	-0.14
Cargo	0.11	0.32**	0.56**	0.16	0.28**	0.29**	1	0.28**	0.37**	0.79**	0.58**	0.01	0.01
Tank	0.01	0.05	0.42**	0.06	0.07	0.18*	0.28**	1	0.24**	0.70**	0.87**	0.00	0.00
Tug	0.21**	0.03	0.68**	0.13	0.38**	0.41**	0.37**	0.24**	1	0.35**	0.37**	0.00	-0.16
Built < 92	0.10	0.30**	0.56**	0.11	0.26**	0.34**	0.79**	0.70**	0.35**	1	0.88**	0.00	0.02
Built > 92	0.12	0.18*	0.57**	0.10	0.23**	0.38**	0.58**	0.87**	0.37**	0.88**	1	0.00	-0.01
Marina	0.15	-0.03	0.01	0.00	-0.02	0.08	0.01	0.00	0.00	0.00	0.00	1	-0.05
CoastDist	-0.07	0.52**	-0.12	-0.07	-0.06	-0.14	0.01	0.00	-0.16	0.02	-0.01	-0.05	1

*Sig. 0.01; **Sig. 0.05.

with Moran's I spatial autocorrelation index was significant up to 18 nearest neighbors, but its values were generally low: for example, values lower than 0.1 were significant. Considering the non-normal distribution of the data, the low values of Moran's I were not considered a conclusive indication of spatial dependence. Summing up, both exploratory spatial analyses suggest that spatial heterogeneity and spatial dependencies, though possibly present, are unlikely to severely impact the models. Therefore, standard Poisson regression was preferred over generalized forms of Poisson regression, such as spatially autoregressive or geographically weighted Poisson regression. Model residuals were also assessed using Getis G and Moran's I.

Global models

Table 2 summarizes the correlations of the response variable with each predictor variable, as well as the correlations between all predictors. Correlation is generally low between response and predictor variables, suggesting that there is more contributing to oily discharges than traffic and other marine activities represented by this set of variables. Notably, the highest correlations are with total traffic and with recreational activities, whereas tank and cargo vessels are among the lowest correlations.

Correlation between predictor variables is also generally low, with a few exceptions: both vessel age categories are highly correlated with tank and cargo vessels (greater than 0.65 in absolute value). Additionally, there are some expected high cross-correlations: between older and newer vessels, as well as between the variable representing total traffic and those representing traffic of some vessel types. To address these instances, two alternative models were examined, to control for multicollinearity⁴ (Table 3). The variable representing the surveillance effort by the NASP program was included in each model as an offset variable (Eq. (1)), to emphasize its role as a precondition for the observation of oily discharges.

Based on statistical indicators, including AIC, the log likelihood and McFadden pseudo R^2 , Model 1 (vessel type) performs slightly better than Model 2 (vessel age). Moreover, in comparison with Model 2, Model 1 presents a more explicit set of predictors, which potentially make this model a more effective analytical and decision support tool, therefore Model 1 was preferred.

Significant variables, rank-ordered by significance, are: presence of small harbors, proximity to coast, recreational traffic, passenger traffic, tug traffic, and fishing traffic, with McFadden pseudo R^2 just

above 0.3. Cargo and tank vessels were not significant, and therefore sequentially removed from the model. Fig. 3 depicts the spatial distribution of the model residuals over the study area. Residuals range between -2.11 and +7.12, whereas the dependent variable ranges between 0 and 14. The residuals appear to be less clustered than the observed dependent variable (see Figs. 2 and 3), and Moran's I spatial autocorrelation tests yield lower correlation values than for the dependent variable. Negative residuals and low positive residuals appear to be evenly scattered throughout the region. The highest positive residuals are observed in the northern part of the study region (near Prince Rupert), off northeastern Vancouver Island, in the Strait of Georgia (mostly north of Vancouver), and off southwestern Vancouver Island (Fig. 3). Comparing the observed pattern of the response variable with the residual pattern (Figs. 1 and 3), the model tends to slightly underpredict the response variable in most of the study region, whereas overprediction occurs mostly in the Strait of Georgia, particularly on the East side.

The model, with its set of predictors, suggests that oily discharges are mainly associated with recreational activities and passenger traffic, in close proximity to the coast. The localized pattern of positive residuals (Fig. 3), along with the pattern of observed discharges (Fig. 1) and their clustering pattern (Fig. 2), appears to reflect differences between the various parts of the study region, leading to spatial heterogeneity. Indeed, all the analyses conducted in this paper hint to the presence of a spatial pattern, which is visually apparent even though it cannot be confirmed by inferential tests.

Table 3
Alternative models.

Model 1: Vessel type			
	Std. beta	Std. error	Wald
Marina	0.45	0.04	15.97
CoastDist	-10.36	0.05	-8.45
Recreation	0.18	0	5.4
Passenger	0.19	0	4.75
Tug	0.22	0	4.44
Fishing	0.12	0	2.09
McFadden pseudo R^2	0.32	Adjusted McFadden pseudo R^2	0.31
Log-likelihood	-550.54	AIC	1115.1
Model 2: Vessel age			
	Std. beta	Std. error	Wald
Marina	0.44	0.04	16.97
Traffic	0.42	0	9.81
CoastDist	-0.43	0.05	-9.43
McFadden pseudo R^2	0.3	Adjusted McFadden pseudo R^2	0.3
Log-likelihood	-564.65	AIC	1137.3

⁴ Only one of the two vessel age variables was included, because of their high cross-correlation. Neither variable was significant in the model.

Regional models

Building on these results and visual analyses, and guided by expert opinion, the study area was subdivided into three regions, as shown in Fig. 4. Partitioning of the study area not only addresses spatial heterogeneity, but also yields more accurate and reliable models of the local processes associated with oily discharges in each region. Regional models remain based on 5×5 km grid cells; however, direct comparisons between global and regional models should be avoided, in consideration of the MAUP and ecological fallacy problems (Fotheringham & Wang, 1991; Robinson, 2009; Tranmer & Steel, 1998).

The selection criteria used for the entire region lead to prefer Model 1 for all the regions.⁵ The three regional models are summarized in Table 4.

Since each of the three regional models and the global model are estimated on a different number of observations, a comparison of the models based on AIC or other statistical indices would be problematic; therefore, the adjusted McFadden pseudo R^2 was used as an indicator of fit. The local regression for Region 1 presents the highest McFadden pseudo R^2 value, showing a remarkable improvement over the global regression, whereas Regions 2 and 3 present similar McFadden pseudo R^2 values. Notably, Region 1 contains a larger number of grid cells, approximately twice as many as each of Region 2 and Region 3 (see Fig. 4).

The three regional models appear to identify local processes associated with oil discharges more clearly than the global model. In Region 1 oily discharges are clearly associated only with recreational and passenger activities, mostly represented by recreational vessel traffic, proximity to the coast, and passenger traffic; the variable “small harbors” is borderline significant, probably due to a more sparse presence of these facilities in the northern area. Tug traffic was excluded from the model due to its high (negative) correlation with recreational traffic; cargo, tank, and marina were not significant, therefore sequentially removed in the selection process. Even though McFadden pseudo R^2 values were adjusted for sample size and number of predictors, the larger sample size may have a role in improving the goodness of fit of the model for this region. Region 2 features the largest set of predictors, reflecting the broader range of activities that take place in the area. The number of discharges is the highest in this region, at least three times as many as in each of the other regions. This is also the only region where extensive positive spatial clustering was detected (Fig. 2). This set of variables is similar to the variables significant in the global model, probably because of the intensity of the phenomenon in this region. Tug traffic is only significant in this regional model, likely due to fact that tug boats are the most common mode of commercial transportation along the coastal routes and inner passages. Cargo, passenger, and tank traffic were not significant, therefore sequentially removed from the model. Finally, Region 3, with the lowest number of observed discharges, features only two predictor variables: cargo and fishing traffic. Marinas, recreation, and tug traffic were excluded from this model because of their high cross-correlations; tank traffic, passenger traffic, and distance from coast were not significant, therefore sequentially removed through the selection process. The spatial distribution of the regional model residuals is shown in Fig. 4: only the model for Region 1 presents

Table 4
Regional regression models.

Region 1 model			
	Std. beta	Std. error	Wald
Recreation	0.19	0	6.4
CoastDist	-18.19	0	-4.48
Passenger	0.08	0	2.85
McFadden pseudo R^2 0.47		Adjusted McFadden pseudo R^2 0.45	
Log-likelihood -124.74		AIC 257.48	
Region 2 model			
	Std. beta	Std. error	Wald
Marina	0.21	0.05	6.59
Recreation	0.21	0	6.28
Tug	0.28	0	4.75
Fishing	0.28	0	4.4
CoastDist	-0.91	0	-4.37
McFadden pseudo R^2 0.2		Adjusted McFadden pseudo R^2 0.19	
Log-likelihood -335.2		AIC 682.36	
Region 3 model			
	Std. beta	Std. error	Wald
Cargo	1.76	0	3.93
Fishing	1.55	0	3.79
McFadden pseudo R^2 0.13		Adjusted McFadden pseudo R^2 0.1	
Log-Likelihood -60.05		AIC 126.10	

consistently lower residuals (in absolute value) than the global model. The residual values of the models for Regions 2 and 3 are lower than for the global model, however the intensity of oily discharges is lower in these regions, hence it cannot be concluded that these models perform better than the global model. The residuals of all regional models exhibit lower clustering than the residuals of the global model, suggesting that regional models address spatial heterogeneity and yield more reliable results.

Discussion

Regional models performed better than the global model, suggesting that oily discharges occurred in association with human maritime activities that varied among regions. Most of the detected oily discharges occurred close to shore, particularly in Regions 1 and 2. Discharges were negatively correlated with distance from shore in the best-fit global model, and in regional models for Regions 1 and 2. Oily discharge patterns are clearly associated with number of marinas in a grid cell, as the predictor variable “Marina” was retained in the best-fit global, and best-fit regional models for both regions 1 and 2. In region 1, the positive residuals in the vicinity of Prince Rupert are probably due to the anomaly of high discharge values in an area otherwise characterized by low and sparse discharge values; analogous considerations apply for areas off northeastern and southwestern Vancouver Island.

Region 2 is a very complex area, both morphologically and in terms of human use. This region is exposed to intense commercial traffic traveling to and from major ports and terminals in Vancouver, Puget Sound and along the southern coast of the Strait of Juan de Fuca. The Port of Vancouver is large by any standard, with the fifth largest container throughput in North America (AAPA). Not surprisingly, this region exhibits the highest clustering and the highest number of observed discharges, in an area that is just over half the extent of each of the other regions. The pseudo R^2 of this model is slightly lower than the global model. Compared to the global model, it contains almost the same set of predictors; however, the relative contribution of each predictor is different, as indicated by the coefficient associated with each variable. The model for this region does not include surveillance effort as offset variable, as the coverage by the NASP program is quite consistent

⁵ Notably, Model 2 produces slightly better statistical results for region 1 (AIC = 238.09; LogLikelihood = -115.04; pseudo- R^2 = 0.51; with significant variables, rank-ordered, “built before 1992”, “Distance from Coast”, and “Total Traffic Hours”); however, Model 1 is still preferred and reported in detail, as it provides a more clear set of predictors and can be more easily compared to the other regional models.

(i.e., little variability for the model to test) over the entire region, and apparently complete.

Although the models performed well, there was a significant amount of variability in detection rates that was not captured with our suite of predictor variables. The most obvious explanation would be that drift moved oily discharges before detection occurred, producing error associating these discharges with potential sources. Hind cast trajectory modeling of detected oily discharges may reduce this error, though it would be difficult to estimate the length of time between detection and the oily discharge occurrence. As well, we did not include variation in type of oil discharged or conditions that might affect weathering and thus the persistence of an oily discharge. Persistence can affect the detectability of an oily discharge by the NASP crew, biasing our results. For example, oily discharges associated with recreational activities are likely highly refined petroleum products such as diesel or gasoline. The higher volatility of these products results in a lower probability of persisting long enough to be detected by NASP when compared to heavier, less refined products such as bunker fuel spills associated with larger commercial vessels. Another potential explanation would be that important factors were not included in the models. There is a paucity of data on both the fishing and recreational vessel activities because they generally are not tracked by marine traffic control systems, such as Automated Identification System (AIS) and radar. Even if fishing activity is covered by marine traffic control systems, their fishing locations are protected by Canadian federal legislation ("Privacy Act"). There is some anecdotal evidence that suggests oily discharges are associated with these two activities, but this is not conclusive and should be modeled specifically. Associating oily discharges with marinas is consistent with this anecdotal evidence (these marinas are used by both fishery and recreational vessels), but this only provides a piece of the puzzle.

The distribution of the response variable of this study imposed a number of limitations to the use of common standard and spatial statistical methods. The variable contains a very large number of null values over the entire Canadian Pacific EEZ. In this paper we used Poisson regression; however, alternative modeling techniques can be used to model surveillance data. One such alternative is the log transformation of the dependent variable and the consequent estimation of a linear model on the transformed data (Gujarati, 2003). Our study may lend itself well to Bayesian approaches as well (for example see Szwed, van Dorp, Merrick, Mazzuchi, & Singh, 2006). The equidispersion assumption was met by restricting the analysis to near-shore discharges and reducing the study area by means of a minimum convex polygon. Further studies, considering the entire area, may need to employ other analytical methods, such as zero-inflated (Lambert, 1992; Zuur, Saveliev, & Ieno, 2012, 324 pp.) or hurdle models (Heilbron, 1994; Mullahy, 1986), or binomial models with an integrated parameter for dispersion correction (Zeileis, Kleiber, & Jackman, 2008). When modeling a data set with a large number of zeros, it is important to determine first what types of zeros are present in order to select the appropriate analytical approach (Martin et al., 2005). For example, if the excess of zeros is due to poor or inadequate sampling, it is important to collect information on factors that affect detectability: factors such as visibility, ocean surface conditions, height of aircraft, and observer for example. Much of these data are not currently being collected consistently by NASP (Serra-Sogas, 2010).

The application of spatial analytical methods to Poisson-distributed data remains problematic, despite recent progress (Griffith, 2010). Here, the use of geographically weighted or spatially autoregressive generalized Poisson models appeared unwarranted, given our Moran's I and Getis-Ord G scores. However, future work should consider the use of spatially autoregressive

models, as the autoregressive term may be a way of modeling spatial uncertainty induced by wind and currents for example. The use of regional models to address spatial heterogeneity appears to be a very valuable alternative to geographically weighted models. Although the regional approach improves the model fit only in one of the regions based on pseudo R^2 values (Region 1), this approach reduces the clustering of residuals in all regions, yielding more reliable models. These regional models constitute more effective decision support tools, because of their increased reliability and because they provide a more detailed analysis by identifying predictors that are locally associated with the response variable.

As mentioned in the introduction, most of the literature modeling oil pollution occurrence and impacts has focused on accidental discharges associated with large vessel traffic. We believe this is the first published attempt at modeling the high resolution (5×5 km) spatial probability of occurrence of smaller, often unreported oily discharges based on aerial surveillance data. Furthermore, although external predictor variables have been used to model accidental discharges from large vessel traffic (i.e., Hong, Chen, & Zhang, 2010; Meade et al., 1983; Ketkar & Babu, 1997; Pelot & Plummer, 2008), we believe our study is the first attempt to model oily discharge occurrence probability using an exhaustive suite of external predictor variables.

The reasons for predictive modeling of oily discharges are manifold. Surveillance effort is patchy in time and space, and the environmental consequences of detected oily discharges typically involve some form of extrapolation to areas not covered sufficiently by NASP, particularly to areas considered sensitive to exposure of oil pollution. The identification of factors significantly associated with the spatial-temporal patterns of oily discharges could inform the extrapolation to areas with less surveillance. Additionally, explanatory variables or predictors that are significantly associated with spatial patterns of discharge rates can be used to account for spatial variability, allowing for greater precision in estimating spatially explicit trends in discharge rates over time. Increasing the precision or reducing the coefficient of variation of estimated means increases the power of statistical test for detecting change (Gerrodette, 1987). Our approach to increasing precision in estimated discharge rate is essentially a model-based approach for increasing statistical power as described by Buckland et al. (2012). Model results can also be used to inform NASP efforts to identify human activities most associated with oily discharge patterns, and identify areas where NASP coverage may need to be increased. Finally, spatially explicit estimates of discharge rates can be used to monitor the effectiveness of programs and policy aimed at reducing the rates of discharge of oily pollution.

Conclusion

For several years the national aerial surveillance program (NASP) has detected oily discharges in Canadian waters. This study is a first effort to analyze the spatial patterns of discharge and their association with human marine activities. Discharges occurring in the Canadian Pacific Ocean were modeled using Poisson regression. To address the complexity of the study area and observed spatial heterogeneity, three partitions of the area ('regions') were identified and modeled separately. These regional models were improvements to the global model, by yielding more accurate and reliable estimates, and by more clearly identifying predictors associated with oily discharges locally. The northern coast exhibits a sparse pattern of oily discharges, associated with recreational and passenger traffic, and a similar pattern of discharges off the southwest coast of Vancouver Island is associated with commercial and fishing traffic. The highest intensity of discharges is observed in the Strait of Georgia; this region exhibits the largest set of predictors,

including commercial, fishing, and recreational traffic. All the models tend to underestimate the response variable, likely due to spatial uncertainty associated with ocean drift, insufficient surveillance coverage, and other potential predictors, not considered in this study.

As NASP data are available for all the Canadian coasts, future studies can extend the analysis to the Atlantic coast, the Great Lakes and the Arctic region, as well as compare the results obtained for each coast. These studies should also consider larger geographic areas, ideally the entire Canadian economic exclusive zone (EEZ). For such analyses, the methods used here may be limited, but promising alternatives include negative binomial models, zero-inflated and hurdle Poisson, or Bayesian models. Future directions of this study should also consider more explicit spatial modeling techniques, including geographically weighted and spatially autoregressive generalized models.

Acknowledgments

We are indebted to the Oil in Canadian Waters (OCW) research working group for facilitating this project in so many ways, including but not limited to lively discussions regarding small scale oily discharges in all regions of Canada. Transport Canada's National Aerial Surveillance Program, and Environment Canada's Marine Aerial Reconnaissance Team provided both data and key information regarding the operational details of surveillance as well as insights into the human activities associated with discharges they had detected. Transport Canada's Research and Development provided the majority of the funding for this research, and Environment Canada's Canadian Wildlife Service provided essential in-kind support. The Department of Fisheries and Ocean's Institute of Ocean Sciences and the Department of Biology also provided in-kind support to P.D. O'Hara. Finally, we would like to extend our warm thanks to the anonymous referees who so thoroughly and patiently reviewed our manuscript: their insightful comments and constructive criticism greatly helped improve the quality and presentation of our work.

References

- AAPA. American Association of Port Authorities. <http://www.aapa-ports.org>.
- Anifowose, B., Lawler, D. M., van der Horst, D., & Chapman, L. (2012). Attacks on oil transport pipelines in Nigeria: a quantitative exploration and possible explanation of observed patterns. *Applied Geography*, 32, 636–651.
- Anselin, L. (1995). Local indicators of spatial association – LISA. *Geographical Analysis*, 27, 93–115.
- Anselin, L. (1998). *Spatial econometrics: Methods and models*. New York: Kluwer Academic Publisher.
- Augustine, N. H., McNicol, J., & Marriott, C. A. (2006). Using the truncated auto-Poisson model for spatially correlated counts of vegetation. *Journal of Agricultural Biological and Environmental Statistics*, 11, 1–23.
- Backer, H., Leppanen, J.-M., Brusendorff, A. C., Forsius, K., Stankiewicz, M., Mehtonen, J., et al. (2010). HELCOM Baltic Sea action plan – a regional programme of measure for the marine environment based on the ecosystem approach. *Marine Pollution Bulletin*, 60, 642–649.
- Boots, B., & Tiefelsdorf, M. (2000). Global and local spatial autocorrelation in bounded regular tessellations. *Journal of Geographical Systems*, 2, 319–348.
- Boyd, W. S., McFarlane Tranquilla, L., Ryder, J. L., Shisko, S. G., & Bertram, D. F. (2008). Variation in marine distributions of Cassin's Auklets (*Ptychoramphus aleuticus*) breeding at Triangle Island, British Columbia. *The Auk*, 125(1), 158–166.
- Buckland, S. T., Burt, M. L., Rexstad, E. A., Mellor, M., Williams, A. E., & Woodward, R. (2012). Aerial surveys of seabirds: the advent of digital methods. *Journal of Applied Ecology*, 49, 960–967.
- Burger, A. E. (1993). Estimating the mortality of seabirds following oil-spills - effects of spill volume. *Marine Pollution Bulletin*, 26, 140–143.
- Burger, A., & Fry, D. (1993). Effects of oil pollution on seabirds in the northeast Pacific. In K. Vermeer, K. Briggs, K. Morgan, & D. Siegel-Causey (Eds.), *The status, ecology, and conservation of marine birds of the North Pacific*. Ottawa: Canadian Wildlife Service Special Publication.
- Burt, J., Barber, G., & Rigby, D. L. (2009). *Elementary statistics for geographers* (3rd ed.). New York: Guilford Press.
- Camphuysen, C. J. (1989). *Beached bird surveys in the Netherlands 1915 – 1988: Seabird mortality in the southern North Sea since the early days of oil pollution*. Technisch Rapport/Vogelbescherming. Amsterdam: Werkgroep Noordzee.
- Camphuysen, C. J., & Heubeck, M. (2001). Marine oil pollution and beached bird surveys: the development of sensitive monitoring instrument. *Environmental Pollution*, 112, 443–461.
- Carpenter, A. (2007). The Bonn agreement aerial surveillance programme: trends in North Sea oil pollution 1986–2004. *Marine Pollution Bulletin*, 54, 149–163.
- Charlier, R. H., Fink, C. W., & Krystosyk-Gromadzinska, A. (2012). Throw it overboard: a commentary on coastal pollution and bioremediation. *Journal of Coastal Research*, 28, 881–890.
- Church, R. L., Scaparra, M. P., & Middleton, R. S. (2004). Identifying critical infrastructure: the median and covering facility interdiction problems. *Annals of the Association of American Geographers*, 94, 491–502.
- Dormann, C. F., McPherson, J. M., Araujo, M. B., Bivand, R., Bolliger, J., Carl, G., et al. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30, 609–628.
- Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, 196, 483–493.
- Ferraro, G., Bernardini, A., David, M., Meyer-Roux, S., Muellenhoff, O., Perkovic, M., et al. (2007). Towards an operational use of space imagery for oil pollution monitoring in the Mediterranean basin: a demonstration in the Adriatic Sea. *Marine Pollution Bulletin*, 54, 403–422.
- Fotheringham, A., Brundson, C., & Charlton, M. (1998). Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment & Planning A*, 30, 1905–1927.
- Fotheringham, A. S., & Wang, D. W. S. (1991). The modifiable aerial unit problem in multivariate statistical analysis. *Environment and Planning A*, 23, 1025–1044.
- Gade, M., & Alpers, W. (1999). Using ERS-2 SAR images for routine observation of marine pollution in European coastal waters. *The Science of the Total Environment*, 237/238, 441–448.
- Gerrodette, T. (1987). A power analysis for detecting trends. *Ecology*, 68, 1364–1372.
- GESAMP (Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection). (2007). *Estimates of oil entering the marine environment from sea-based activities*. London: International Maritime Organization.
- Getis, A. (2008). A history of the concept of spatial autocorrelation: a geographer's perspective. *Geographical Analysis*, 40, 297–309.
- Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24, 189–206.
- Gilbert, N. L., Goldberg, M. S., Beckerman, B., Brook, J. R., & Jerrett, M. (2005). Assessing spatial variability of ambient nitrogen dioxide in Montreal, Canada, with a land-use regression model. *Journal of Air and Waste Management Association*, 55, 1059–1063.
- Griffith, D. A. (2010). The Moran coefficient for non-normal data. *Journal of Statistical Planning and Inference*, 140, 2980–2990.
- Gujarati, D. (2003). *Basic econometrics* (4th ed.). Boston: McGraw Hill.
- Gundlach, E. R., Boehm, P. D., Marchand, M., Atlas, R. M., Ward, D. M., & Wolfe, D. A. (1983). The fate of the Amoco Cadiz Oil. *Science*, 221, 122–129.
- Heilbron, D. C. (1994). Zero-altered and other regression models for count data with added zeros. *Biometrical Journal*, 36, 531–547.
- Henderson, S. B., Beckerman, B., Jerrett, M., & Brauer, M. (2007). Application of land use regression to estimate long-term concentration of traffic-related nitrogen oxides and fine particulate matter. *Environmental Science and Technology*, 41, 2422–2428.
- Hong, X., Chen, W., & Zhang, L. (2010). A probabilistic risk forecast of accidental oil spills from vessels in Luoyuan Bay, Fujian Province, PRC. *Procedia Environmental Sciences*, 2, 49–56.
- Jerrett, M., Arain, M. A., Kanaroglou, P., Beckerman, B., Crouse, D., Gilbert, N. L., et al. (2007). Modeling the intraurban variability of ambient traffic pollution in Toronto, Canada. *Journal of Toxicology and Environmental Health, Part A*, 70, 200–212.
- Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahsuvaroglu, T., et al. (2005). A review and evaluation of intraurban air pollution exposure models. *Journal of Exposure Analysis and Environmental Epidemiology*, 15, 185–204.
- Keramistoglou, I., Cartalis, C., & Kiranoudis, C. T. (2006). Automatic identification of oil spills on satellite images. *Environmental Modelling & Software*, 21, 640–652.
- Ketkar, K. W., & Babu, A. J. G. (1997). An analysis of oil spills from vessel traffic accidents. *Transport Research – D*, 2, 35–41.
- Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34, 1–14.
- Lambert, D. M., McNamara, K. T., & Garrett, M. I. (2006). An application of spatial Poisson models to manufacturing investment location analysis. *Journal of Agricultural and Applied Economics*, 38, 105–121.
- Leung, Y., Mei, C.-L., & Zhang, W.-X. (2000). Statistical tests for spatial non-stationarity based on the geographically weighted regression model. *Environment and Planning A*, 32, 9–32.
- Lin, G., & Zhang, T. (2007). Loglinear residual tests of Moran's I autocorrelation and their applications to Kentucky breast cancer data. *Geographical Analysis*, 39, 293–310.
- MARPOL. International convention for the prevention of pollution from ships (MARPOL). International Maritime Organization (IMO). <http://www.imo.org/about/conventions/listofconventions/pages/international-convention-for-the-prevention-of-pollution-from-ships-%28marpol%29.aspx>.

- Martin, T. G., Wintle, B. A., Rhodes, J. R., Kuhnert, P. M., Field, S. A., Low-Choy, S. J., et al. (2005). Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology Letters*, 8, 1235–1246.
- McGree, J. M., & Eccleston, J. (2012). Robust designs for Poisson regression models. *Technometrics*, 54, 64–72.
- Meade, N., LaPointe, T., & Anderson, R. (1983). Multivariate analysis of worldwide tanker accidents. In *International oil spill conference proceedings (Feb. 1983)* (pp. 553–557).
- Mohebbi, M., Wolfe, R., & Jolley, D. (2011). A Poisson regression approach for modelling spatial autocorrelation between geographically referenced observations. *Medical Research Methodology*, 11, 133.
- Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics*, 33, 341–365.
- Nakaya, T., Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2005). Geographically weighted Poisson regression for disease association mapping. *Statistics in Medicine*, 24, 2695–2717.
- Nilsen, E. B., Pedersen, S., & Linnell, J., D., C. (2008). Can minimum convex polygon home ranges be used to draw biologically meaningful conclusions? *Ecological Research*, 23, 635–639.
- NRC (US National Research Council). (2003). *Oil in the sea. III. Inputs, fates, and effects*. National Research Council of the National Academies (Ocean Studies Board, Marine Board, Divisions of Earth and Life Studies, and Transportation Research Board). National Academy of the Sciences. Washington, DC: National Academies Press.
- OECD (Organization for Economic Cooperation and Development). (2002). *Cost savings stemming from non-compliance with international environmental regulations in the maritime sector*. Paris, France: Organization for Economic Cooperation and Development.
- Openshaw, S., & Alvanides, S. (1999). Applying geocomputation to the analysis of spatial distributions. *Geographical Information Systems: Principles and Technical issues*, 1, 267–282.
- O'Hara, P. D., & Morgan, K. H. (2006). Do low rates of oiled carcass recovery in beached bird surveys indicate low rates of ship-source oil spills? *Marine Ornithology*, 34, 133–140.
- O'Hara, P. D., Serra-Sogas, N., Canessa, R., Keller, P., & Pelot, R. (2013). Estimating discharge rates of oily wastes and deterrence based on aerial surveillance data collected in the western Canadian marine waters. *Marine Pollution Bulletin*, 69, 157–164.
- Pelot, R., & Plummer, L. (2008). Spatial analysis of traffic and risks in the coastal zone. *Journal of Coastal Conservation*, 11, 201–207.
- Piatt, J. F., Lesink, C. J., Butler, W., Kendziorek, M., & Nysewander, D. R. (1990). Immediate impact of the 'Exxon Valdez' oil spill on marine birds. *The Auk*, 107, 387–397.
- Robinson, W. S. (2009). Ecological correlations and the behavior of individuals. *International Journal of Epidemiology*, 38(2), 337–341.
- Serra-Sogas, N. (2010). *Modelling risk of chronic oil pollution from vessel operations in Canada's West Coast*. MSc thesis. Department of Geography. University of Victoria <http://hdl.handle.net/1828/2921>.
- Serra-Sogas, N., O'Hara, P. D., Canessa, R., Keller, P., & Pelot, R. (2008). Visualization of spatial patterns and temporal trends for aerial surveillance of illegal oil discharges in western Canadian marine waters. *Marine Pollution Bulletin*, 56, 825–833.
- Szwed, P., van Dorp, J. R., Merrick, J. R. W., Mazzuchi, T. A., & Singh, A. (2006). A Bayesian paired comparison approach for relative accident probability assessment with covariate information. *European Journal of Operational Research*, 169, 157–177.
- Teal, J. M., & Howarth, R. W. (1984). Oil spill studies: a review of ecological effects. *Environmental Management*, 8, 27–44.
- Tranmer, M., & Steel, D. G. (1998). Using census data to investigate the causes of the ecological fallacy. *Environment and Planning A*, 30(5), 817–831.
- Transport Safety Board of Canada. (2007). *Pipeline investigation report (P07H0040): Crude oil pipeline – Third-party damage*. Published 24 July 2007. URL <http://www.tsb.gc.ca/eng/rapports-reports/pipeline/2007/p07h0040/p07h0040.asp>.
- Veall, M. R., & Zimmermann, K. (1996). Pseudo-R² measures for some common limited dependent variable models. *Journal of Economic Surveys*, 10, 241–259.
- Vieites, D. R., Nieto-Roman, S., Palanca, A., Ferrer, X., & Vences, M. (2004). European Atlantic: the hottest oil spill hotspot worldwide. *Naturwissenschaften*, 91, 535–538.
- Volckaert, F. A. M., Kayens, G., Schallier, R., & Jacques, T. G. (2000). Aerial surveillance of operational oil pollution in Belgium's maritime zone of interest. *Marine Pollution Bulletin*, 40, 1051–1056.
- Wang, D., Gong, F., Pan, D., Hao, Z., & Zhu, Q. (2010). Introduction to the airborne marine surveillance platform and its application to water quality monitoring in China. *Acta Oceanologica Sinica*, 29, 33–39.
- Wiese, F. K., & Robertson, G. J. (2004). Assessing seabird mortality from chronic oil discharges at sea. *Journal of Wildlife Management*, 68, 627–638.
- Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software*, 27, 1–25.
- Zuur, A., Saveliev, A., & Ieno, E. (2012). *Zero inflated models and generalized linear mixed models with R*. Newburgh, UK: Highland Statistics Ltd.