



HAL
open science

Heterogeneous Data Mining for a Semi-Quantitative Risk Assessment of Oil Contamination from Multiple-Sources in The Ecuadorian Amazon

Juan Durango, Mehdi Saqalli, Laurence Maurice, Emilie Lerigoleur, Nicolas
Maestriperi, Arnaud Elger

► **To cite this version:**

Juan Durango, Mehdi Saqalli, Laurence Maurice, Emilie Lerigoleur, Nicolas Maestriperi, et al.. Heterogeneous Data Mining for a Semi-Quantitative Risk Assessment of Oil Contamination from Multiple-Sources in The Ecuadorian Amazon. 8th International Congress on Environmental Modelling and Software, Jul 2016, Toulouse, France. hal-01446829

HAL Id: hal-01446829

<https://hal.science/hal-01446829>

Submitted on 26 Jan 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

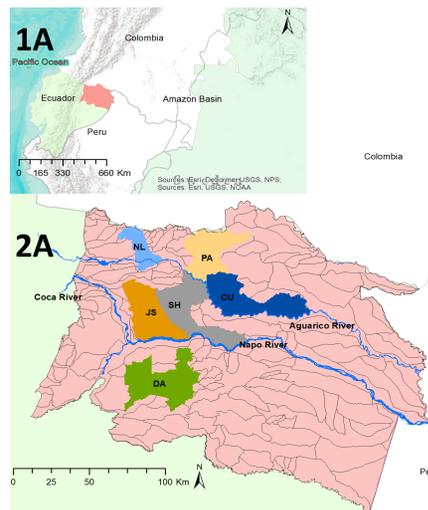
Introduction

In Ecuador oil activities began in 1964 under no environmental policy implementation in one of the most biodiverse biomes in the world. In 1993, communities of the Amazon claimed tons of millions of barrels have been unartfully dropped to the environment. Lack or non-regulated standards for industry, resulted in an infamous heretofore-unresolved litigation upon Texaco (now Chevron) by local population. Pollutant emissions have caused environmental degradation and diseases to populations (Buccina et al. 2013, San Sebastián and Hurtig 2005, Center for Economic and Social Rights 1994). Three periods of oil activity management in the Northeastern Ecuadorian Amazon can be differentiated (Juteau et al. 2014):

- T1 (1972-1991): Foreign Texaco and state CEPE (Ecuadorian Petroleum Ecuadorian Corporation) engaged in oil activities.
- T2 (1992-2001): State managed Petro-Ecuador has taken over the oil activities.
- T3 (2001-2014): After decreed of the hydrocarbon oil activity regulations.

We aim to set the framework for quantifying the emissions at source to air, soil and water. The spatial statistical analysis of these data will allow identifying patterns for a spatially-explicit model of risk.

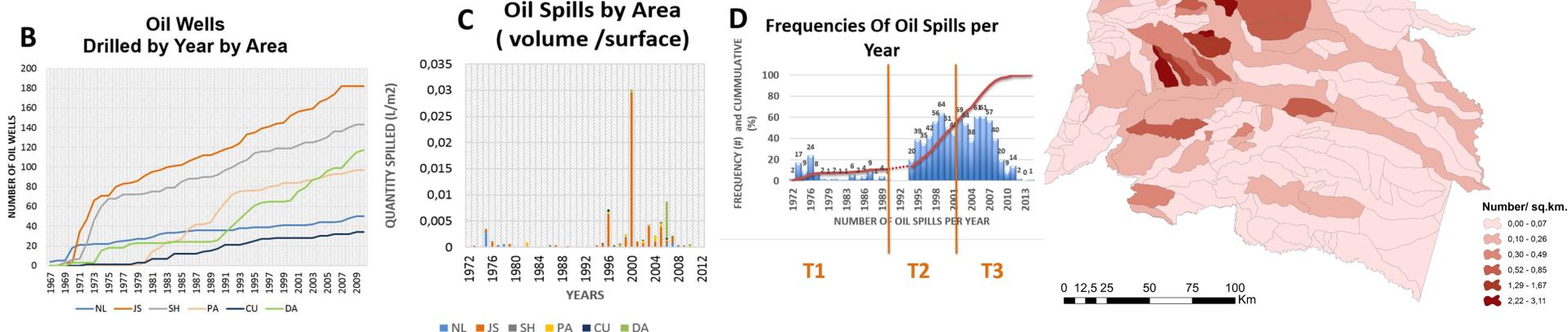
Materials & Methods



1A) The Northeastern Ecuadorian Amazon (NEA) located in Eastern Ecuador close to the borders of Colombia and Peru.
2A) The NEA with parishes studied: Green Polygon=Dayuma (DA), Orange = Joya de Los Sachas (JS), Grey=Shushufindi and Limoncocha (SH), Dark Blue = Part of Cuyabeno Reserve (CU), Light Brown = Pacayacu (PA), Light Blue = Nueva Loja (NL). In grey lines and pink color filled is shown the watershed network.

- GIS data and inventories of potential polluting infrastructures due to spilling and infiltration (i.e. drill cuttings pits, oil wells, flare stacks and pipes) for more than 45 years, were obtained from different Ecuadorian institutions. These data were grouped together and expressed in terms of density for each watershed.
- Programa de Reparación Ambiental y Social (PRAS) from Ecuador, provided historical oil spills data from environmental impact assessments. Threshold of variable “*n*- oil wells” resulting in *n*- “oil spills volume” is analyzed with regression analysis. Occurrences/quantity of oil spills vs. time is plotted for different parishes to compare practices in T1, T2 and T3.

Results & Discussion



B) Oil infrastructure density and population occur in same areas. JS and SH are the most threatened due to the amount of oil wells drilled. DA and PA oil wells reach higher numbers with time.

C) People settle along roads built for developing oil extraction. Contamination differs geographically, JS and DA account for highest volume spilled registered from 1995 to 2006.

D) Lower occurrences of oil spills during T1 challenges people claims of environmental contamination at this period. Data is missing from 1990 to 1994, which may be explained by switch in managerial operators, and T3 suggests an improvement of management activities through time. However, uncertainty of data seems to be high due to lack of disclosure in management practices.

E) Total oil infrastructure density may indicate that at least 5 watersheds are at highest risk (1.29-3.11 #/sq.km)

Perspectives

A linear regression between cumulative oil contamination and number of oil wells is observed ($r^2=0.96$). Two thresholds at 450 & 600 oil wells plead for the implementation of two sub-models. Future investigation should focus on prospective analysis and spatial analysis, considering probabilities of occurrences at each individual oil well at each watershed within the NEA. Heavy metals such as V and Ni could be estimated from crude oil characteristics (e.g. API gravity index). Upper streams aspects, data harmonization and building indicators is a next step. To solve problems with uncertainty, selecting areas with spatially homogenous distribution of oil spills to extrapolate to other areas from periods with better disclosure of oil spills records should be done. These findings could help defining key parameters to improve data and indicators measured by PRAS (e.g. *substance spilled*) to facilitate future decision-making processes for environment and population protection in the NEA.

