Journal of Environmental Management 193 (2017) 188–200

Contents lists available at ScienceDirect

Journal of Environmental Management

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

Research article

A new hypervolume approach for assessing environmental risks

Denys Yemshanov [a, *], Frank H. Koch [b], Bo Lu [c], Ronald Fournier [c], Gericke Cook [d], Jean J. Turgeon [c]

[a] Natural Resources Canada, Canadian Forest Service, Great Lakes Forestry Centre, 1219 Queen Street East, Sault Ste. Marie, ON, P6A 2E5, Canada
[b] USDA Forest Service, Southern Research Station, Eastern Forest Environmental Threat Assessment Center, 3041 East Cornwallis Road, Research Triangle Park, NC, 27709, USA
[c] Natural Resources Canada, Canadian Forest Service, Great Lakes Forestry Centre, Sault Ste. Marie, ON, Canada
[d] USDA Animal and Plant Health Inspection Service, Centre for Plant Health Science and Technology, Plant Protection and Quarantine, Fort Collins, CO, USA

A R T I C L E I N F O

Article history:
Received 21 October 2016
Received in revised form
6 February 2017
Accepted 11 February 2017

Keywords:
Environmental risks
Non-dominant set
Hypervolume
Uncertainty
Asian longhorned beetle
Invasive species
Stochastic dominance

A B S T R A C T

Assessing risks of uncertain but potentially damaging events, such as environmental disturbances, disease outbreaks and pest invasions, is a key analytical step that informs subsequent decisions about how to respond to these events. We present a continuous risk measure that can be used to assess and prioritize environmental risks from uncertain data in a geographical domain. The metric is influenced by both the expected magnitude of risk and its uncertainty. We demonstrate the approach by assessing risks of human-mediated spread of Asian longhorned beetle (ALB, *Anoplophora glabripennis*) in Greater Toronto (Ontario, Canada). Information about the human-mediated spread of ALB through this urban environment to individual geographical locations is uncertain, so each location was characterized by a set of probabilistic rates of spread, derived in this case using a network model. We represented the sets of spread rates for the locations by their cumulative distribution functions (CDFs) and then, using the first-order stochastic dominance rule, found ordered non-dominant subsets of these CDFs, which we then used to define different classes of risk across the geographical domain, from high to low. Because each non-dominant subset was estimated with respect to all elements of the distribution, the uncertainty in the underlying data was factored into the delineation of the risk classes; essentially, fewer non-dominant subsets can be defined in portions of the full set where information is sparse. We then depicted each non-dominant subset as a point cloud, where points represented the CDF values of each subset element at specific sampling intervals. For each subset, we then defined a hypervolume bounded by the outermost convex frontier of that point cloud. This resulted in a collection of hypervolumes for every non-dominant subset that together serve as a continuous measure of risk, which may be more practically useful than averaging metrics or ordinal rank measures.

Overall, the approach offers a rigorous depiction of risk in a geographical domain when the underlying estimates of risk for individual locations are represented by sets or distributions of uncertain estimates. Our hypervolume-based approach can be used to compare assessments made with different datasets and assumptions.

1. Introduction

Assessing risks of uncertain but potentially hazardous environmental events is a critical analytical step in deciding whether to monitor those events and, if necessary, develop appropriate mitigation strategies. Examples include introductions of unwanted insects and diseases (Aukema et al., 2011; Meyerson and Reaser, 2003), as well as negative ecological and economic impacts from fires, floods (Smith, 2013) and changing climate (Oppenheimer et al., 2014; Schneider et al., 2007). By nature, these events are uncertain and the sources of that uncertainty cannot be isolated sufficiently. In many cases, the uncertainty appears to be irreducible, such as the uncertainty associated with the spread of nonnative harmful species (Melbourne and Hastings, 2009). Regardless, when predictive models (Carrasco et al., 2010; Hester et al., 2010; Koch et al., 2009; Yemshanov et al., 2009), including

* Corresponding author.
E-mail address: denys.yemshanov@canada.ca (D. Yemshanov).

http://dx.doi.org/10.1016/j.jenvman.2017.02.021

ensemble analyses (Araújo and New, 2006; Cheung, 2001), are used to assess the potential impacts of such events, this uncertainty greatly curtails the decision support value of their results.

In general, the risk of an undesirable event can be represented by the probability of the event given a suite of conditions, along with some characterization of its consequences (Kaplan and Garrick, 1981). When knowledge about the event is poor, it may not be possible to depict the risk of the event with a single value. Instead, the risk can be depicted, at best, by a set (a distribution) of plausible values. For example, one way to estimate risks of invasion by an alien plant pest species in a spatial context is to simulate spatial stochastic scenarios of the invader's spread and calculate the probabilities of spread for each scenario, as well as the likelihoods under each scenario that populations of the pest will become established in newly-invaded locations. If the scenarios are assumed to have equal probability of occurrence, then this would essentially represent a distribution of likely outcomes with respect to the event of interest for a given site.

When multiple sites (or alternative events) must be compared, risk prioritization requires ordering the distributions of scenario-based risk estimates for each individual site (or event). In theory, if a distribution of plausible risk values can be approximated by a functional form (e.g., Gaussian), it is possible to describe that distribution by its first moments and depict these moments across the domain (e.g., as maps of mean risk values and their variance). This approach has been widely adopted in mean-variance investment analyses (Elton and Gruber, 1995; Keisler and Linkov, 2010; Linkov et al., 2006; Salo et al., 2011; Zhou et al., 2012). Indeed, these types of one- and two-dimensional estimates of risk are popular (see Sims and Finnof, 2013) because they make it possible to formulate and solve a decision-making problem with common optimization algorithms.

Unfortunately, the extent of knowledge about rare but potentially harmful events may be insufficient to properly determine the functional form of a distribution, and analysis may further be hindered by unknown limitations in the modeling technique or uncertainties in the data. Under such circumstances, comparisons of multiple observations in a geographical data set can only be done by considering the entire distribution (i.e., not just the first few moments) of estimated risk values for each of those observations. Ideally, this computationally complex task would be done using a metric that factors in the uncertainty of those multiple distributions, so that the final measure reflects the impact of both the expected magnitude of risk and its degree of variation.

In this paper, we propose a continuous metric that can be used to prioritize uncertain estimates of environmental risks that are depicted by plausible sets (distributions) of values that characterize the extent of impact or damage. Our metric is developed specifically for geographical assessments of environmental risks. It builds upon previous work (Yemshanov et al., 2012) in which we developed a risk metric utilizing the concept of stochastic dominance (SD). We first represent the sets of risk estimates for individual geographical locations in a data set (such as a collection of risk values for a given location generated under different modeling scenarios) by their cumulative distribution functions (CDF), so that every geographical location in the data set is characterized by its own CDF of risk estimates. We then find and rank distinct subsets of those spatial locations (i.e., of their CDFs) using the first-order stochastic dominance (FSD) rule. Ranking the subsets of spatial locations establishes their rank order along a gradient of risk. As described in Yemshanov et al. (2012), the subsets are termed “non-dominant” because each contains a group of CDFs that fail to dominate each other under the first-order stochastic dominance rule. These non-dominant subsets correspond to broad ordinal classes of risk that are ranked from high to low. Notably, for a rational decision-maker, a non-dominant subset would be perceived as a single risk class. This is because it is impossible to establish preference order relationships among the CDFs within a non-dominant subset (Levy, 1992; see more details in Section 2) due to uncertainty in the data, which makes the CDFs in the subset indistinguishable from each other in terms of FSD. This is why FSD is called a partial ordering approach: it facilitates ordering among, but not within, the subsets of a full set.

A relevant corollary is that when the data underlying a set of values, as represented by their CDFs, are highly uncertain, using FSD to compare all CDFs in that set will yield fewer and larger non-dominant subsets than if the underlying data were known more precisely. For decision makers, the ability to prioritize observations according to their risk will be constrained by the uncertainty in the underlying data. This restrictive behavior has important practical benefits: although FSD will yield imprecise and therefore coarse delineations of risk when applied to uncertain data, they are less likely to lead to erroneous decisions than estimates that fail to account for uncertainty and communicate false precision.

An acknowledged limitation of the SD approach is that it only provides an ordinal ranking of CDFs (and the corresponding risk gradations). As noted previously, non-dominant subsets are ordered from high to low risk under SD rules, but the actual difference in the levels of risk between any two subsets is unknown. For example, a geographical assessment of environmental risk (i.e., a risk map) developed using the SD approach would depict the risk levels of different geographic regions with ordinal ranks: 1st, 2nd, etc. Regions classified as 1st and 2nd rank might have very similar levels of risk, but in absolute terms, regions in the 2nd rank may actually be closer to regions in the 3rd rank, or even a much lower rank, than to regions in the 1st rank. Notably, an ordinal measure like SD offers only weak support for decisions where these kinds of fine-scale differences in risk may prove important, such as resource allocation to mitigate risk under tight budget constraints. A continuous measure, where the difference in the level of risk between the non-dominated subsets is quantified, would be far more suitable for such tasks.

Our objective is to describe a new approach, based on the concept of hypervolumes, which transforms the ordinal risk rank measure generated with the FSD rule into a continuous one. Briefly, we depict each non-dominant subset as a point cloud, where the points consist of the CDF values of each observation in the subset at defined sampling intervals. For each subset, we then define a hypervolume that is bounded by the outermost convex frontier of the point cloud and a chosen reference point. The result is a collection of nested hypervolumes, each of which quantifies volumetrically the region within the entire multi-dimensional risk space occupied by a non-dominant subset. Collectively, the calculated hypervolumes for these subsets act as a continuous measure of risk for the full set. As a continuous measure, it has another advantage over order-based risk metrics in that it can be used to prioritize and compare multiple assessments based on different datasets or risk assessment scenarios. We demonstrate the approach with a contemporary example: assessing the risk of human-assisted spread of the Asian longhorned beetle, an invasive forest pest (Haack et al., 2010; Nowak et al., 2001), in the Greater Toronto Area (Ontario, Canada).

2. Methodology

2.1. Assessing risks from scenario-based data

A set of uncertain risk estimates can be described by defining it as a stochastic variable, with a cumulative distribution function (CDF) of risk values x: