NATURAL RESOURCES CANADA - INVENTIVE BY NATURE

A HYPERVOLUME APPROACH FOR ASSESSING RISK UNDER UNCERTAINY

Denys Yemshanov[1], Frank Koch[2], Bo Lu[1],
Jean Turgeon[1], Ron Fournier[1] and Gericke Cook[3]

Denys.Yemshanov@NRCan-RNCan.gc.ca

[1]NRCan, Canadian Forest Service, Great Lakes Forestry Centre, Sault Ste. Marie, ON
[2]USDA Forest Service, Southern Research Station, Research Triangle Park, NC
[3]USDA-APHIS-PPQ, Centre for Plant Health Science and Technology, Fort Collings, CO

IX International Pest Risk Mapping Workshop,
Fort Collins, CO, August 25-28, 2015

Introduction

Information about alien pest invasions is often scarce and imprecise – this makes the assessment of pest invasion risk a daunting task

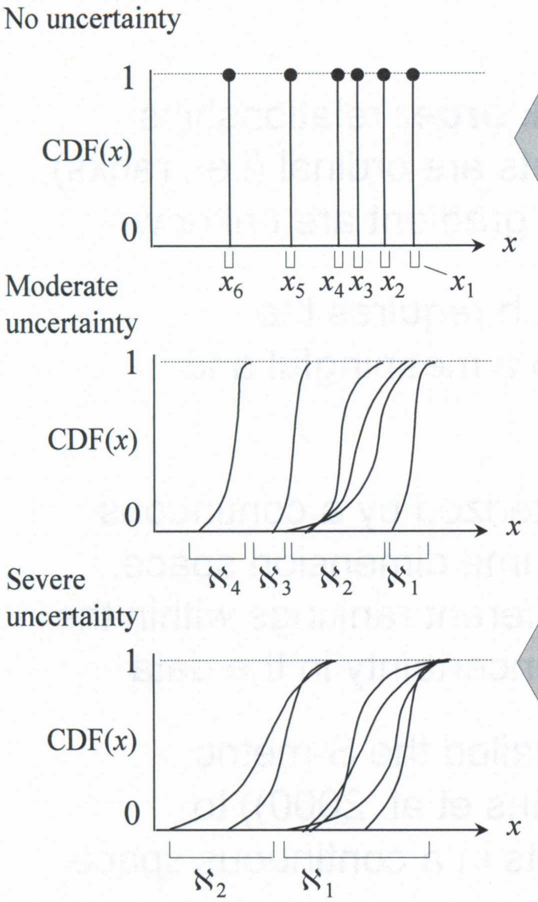
Commonly, it is impossible to derive a single estimate of risk – instead, an analyst deals with a set of plausible estimates based on ensembles of different models, expert opinions or stochastic forecasts

In this case, prioritizing and comparing risk assumes choosing between sets of plausible risk values

The ambiguity in the distributions of risk estimates affects a decision-maker's capacity to differentiate distinct levels of risk and make sensible recommendations for managing and mitigating risk

We address this problem with a continuous hypervolume measure that factors in the ambiguity in the data and can be used to compare and prioritize uncertain estimates of risk for decision-making

Certain vs. uncertain estimates of risk



$\aleph_1$ - $\aleph_4$ – non-dominant sets (by SD rule)

When risk estimates are certain, they can be ordered (or compared) via equality / inequality operators; the ordering is only limited by the measurement accuracy

Ordering uncertain estimates requires comparing their CDFs. CDFs can be compared via a pairwise stochastic dominance (SD) rule:

CDF F(x) dominates G(x) by 1st order SD rule if:

$F(x) \leq G(x)$ for all x and $F(x) < G(x)$ for one or more x.

When the dominance for *F* over *G* and *G* over *F* fail, it is impossible to establish a preference order between *F* and *G* and the CDFs of *F* and *G* become non-dominant to each other

Thus, uncertainty in the data diminishes our ability to discriminate different classes of risk and makes our assessments coarser

Issues with non-dominant sets

Furthermore, the SD rule can only establish **rank order** relationships between sets of CDFs. Distinct non-dominant sets are ordinal (i.e., ranks), so their exact positions within the "high-low" risk gradient are unknown

This limits the practical utility of the method, which requires the subsequent step of converting the order ranks to a meaningful and continuous decision-making priority measure

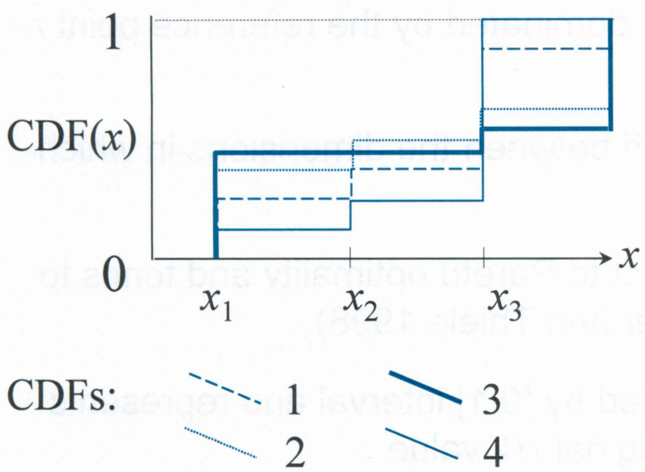
Ideally, each non-dominant set would be characterized by a continuous priority measure that defines its position in the same dimension space. Such a measure would enable comparison of different rankings within the same frame of reference, while factoring in the uncertainty in the data

We propose a hypervolume indicator (HV, also called the S-metric (Fleischer 2003) or Lebesgue measure (Laumanns et al. 2000)) to characterize the position of the non-dominant sets in a continuous space

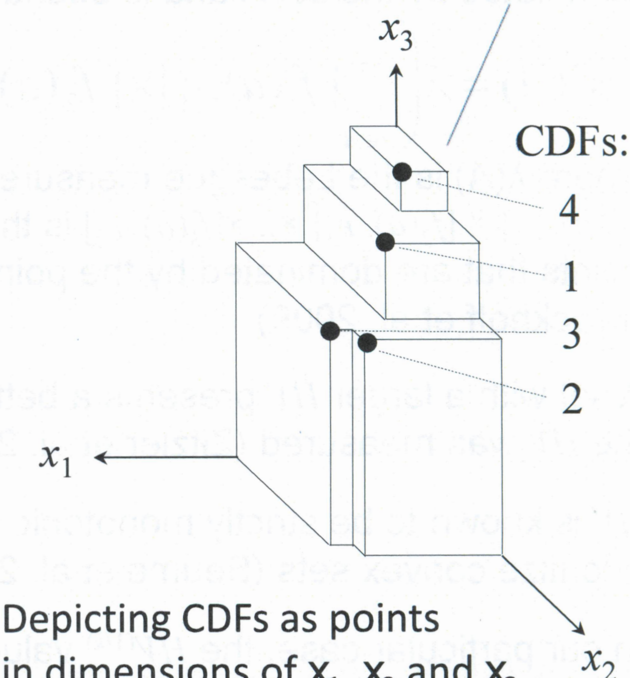
A hypervolume of a non-dominant set: A geometric illustration

Consider a set A of four CDFs sampled at discrete points x_1, x_2, x_3:

CDF 1: (0.25, 0.375, 0.875)
CDF 2: (0.375, 0.5, 0.625)
CDF 3: (0.45, 0.45, 0.55)
CDF 4: (0.125, 0.25, 1)



A hypervolume under a set of points 1-4 and a reference point r = (0,0,0)



Depicting CDFs as points in dimensions of x_1, x_2 and x_3

A hypervolume of a non-dominant set: A formal definition

The hypervolume (HV) of a non-dominant subset A, $A \subseteq N$ can be defined as the hypervolume of the k-dimensional space that is dominated by the set A and is bounded by a reference point $r = (r_1, \ldots, r_k)$:

$$HV(A) = \lambda\left(\square_{a \in A} [f(a), r_1] \times [f_2(a), r_2] \times \ldots \times [f_k(a), r_k] \right)$$

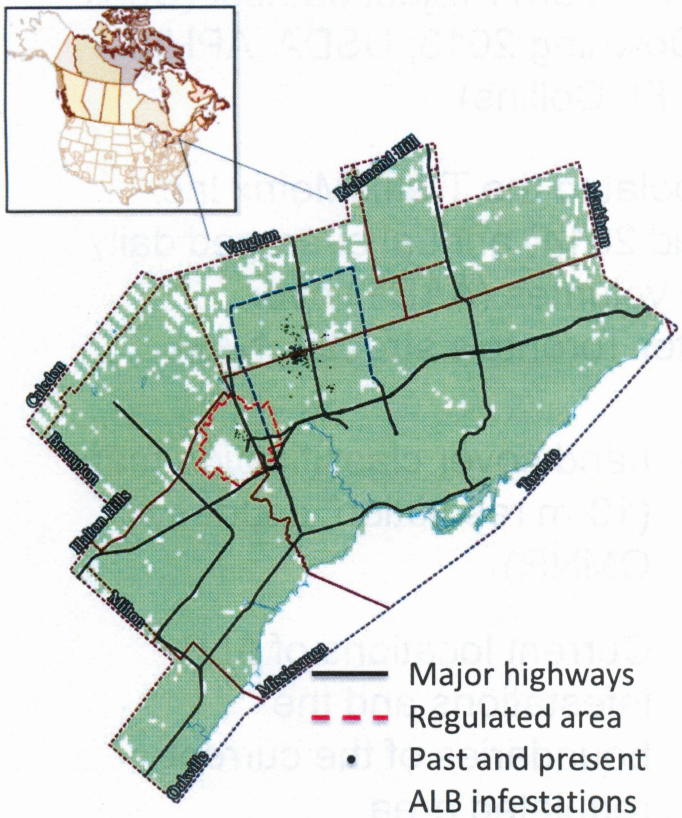
where $\lambda(A)$ is the Lebesgue measure of a set A and $[f_1(a), r_1] \times [f_2(a), r_2] \times \ldots \times [f_k(a), r_k]$ is the k-dimensional hypercuboid consisting of all points that are dominated by the point a but not dominated by the reference point r (Brockhoff et al. 2008).

A set with a larger HV presents a better trade-off between the dimensions in which the HV was measured (Zitzler et al. 2003)

HV is known to be strictly monotonic with respect to Pareto optimality and tends to prioritize convex sets (Beume et al. 2007; Zitzler and Thiele 1998)

In our particular case, the $HV^{(1/k)}$ value is bounded by [0;1] interval and represents a more convenient priority measure than the original HV value

Case study example: assessing risk of spread of Asian longhorned beetle in Greater Toronto, ON



Asian longhorned beetle (ALB)

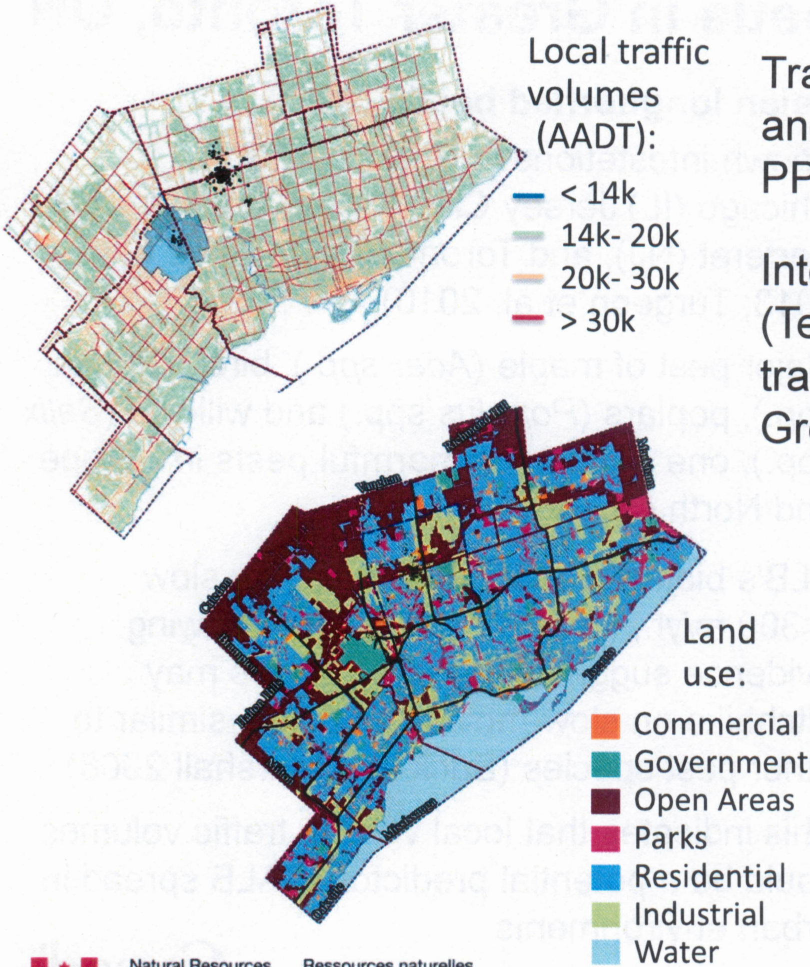
Known infestations in New York (NY), Chicago (IL) Jersey City (NJ), Worcester (MA), Carteret (NJ), and Toronto (ON) (Shatz et al., 2013; Turgeon et al. 2010)

Major pest of maple (*Acer spp.*), birch (*Betula* spp.), poplars (*Populus* spp.) and willows (*Salix* spp.), one of the most harmful pests in Europe and North America

ALB's biological spread rate is very slow (<300 m/yr., Favaro et al. 2015). Growing evidence suggests that the species may hitchhike on slow-moving vehicles, similar to other pest species (Buck and Marshall 2008)

This indicates that local vehicle traffic volumes could be a potential predictor of ALB spread in urban environments

Urban traffic and spatial data inputs



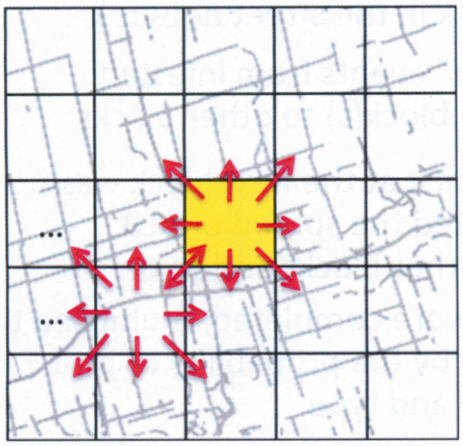
Traffic Pattern Project dataset (Cook and Downing 2013, USDA APHIS PPQ, Ft. Collins)

Interpolated the TrafficMetrix Inc. (Tetrad 2014) annual averaged daily traffic volumes (AADT) over Greater Toronto's street network

Land cover classification map (10-m resolution, courtesy of OMNR)

Current locations of ALB infestations and the boundaries of the current regulated area

Synthesizing the traffic data into a human-mediated spread matrix



The spatial resolution (400 m) of the study was dictated by current regulatory constraints and mandatory eradication protocols for ALB

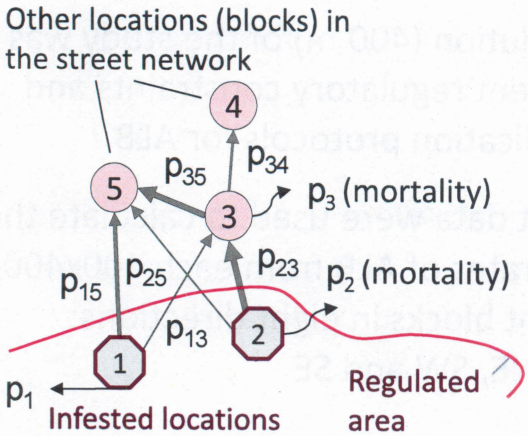
The traffic count data were used to calculate the relative spread rates of ALB from each 400x400-m block to adjacent blocks in eight directions: N, S, E, W, NW, NE, SW and SE

We used the sums of traffic volumes from one block to another in each direction to estimate the relative spread rates p_{ij} between the locations in Greater Toronto area

Divided and 400-series controlled-access highways with high-speed traffic were not included

The spread rates are relative values (i.e., $p_{ij} = \text{traffic}_{ij} / \text{traffic}_{max}$) and were further calibrated to match the historical spread rates

Generating the stochastic spread patterns with the stochastic simulations



For each 400-m block in the street network:

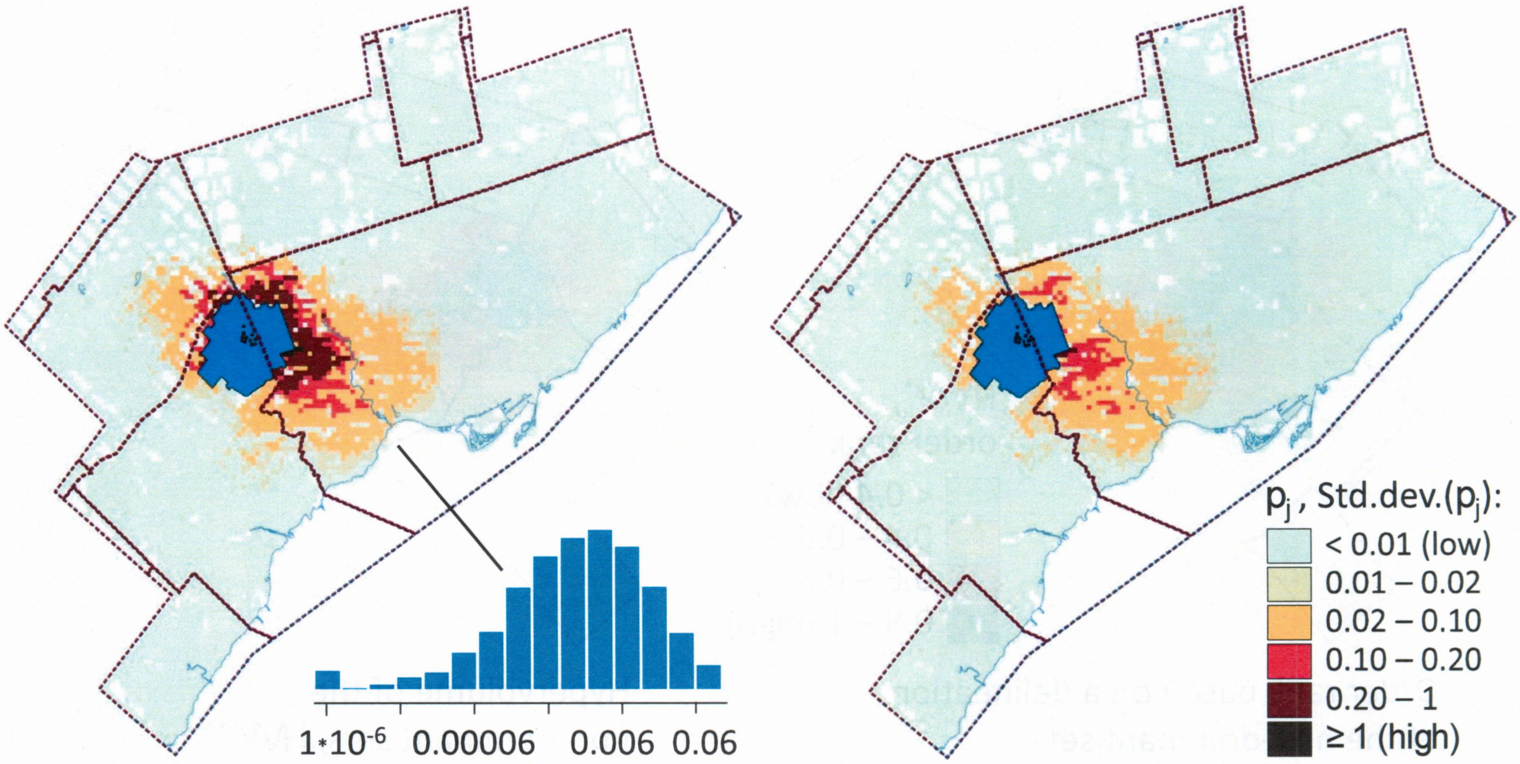
- Simulated "spread" events from infested (or likely infested) block(s) to other blocks
- If "arrival" of the pest at the next block was successful, simulated the spread of ALB from the next to other blocks and so on
- Once simulations were completed, multiplied the spread arrival rate by the probability of pest survival in a given land type

The spread matrix of ALB arrival rates, p_{ij} was calibrated so that the patterns of pest arrival recreated through simulations matched the historical spread rate prior to eradication (i.e., 12 new infestation nuclei per year)

We generated 5000 probabilistic spread scenarios from the areas infested (or likely Infested) with ALB to locations outside of the regulated area - each 400-m block was characterized by a distribution of 5000 arrival rate values, p_j

We used the distributions of the p_j values to illustrate the hypervolume technique

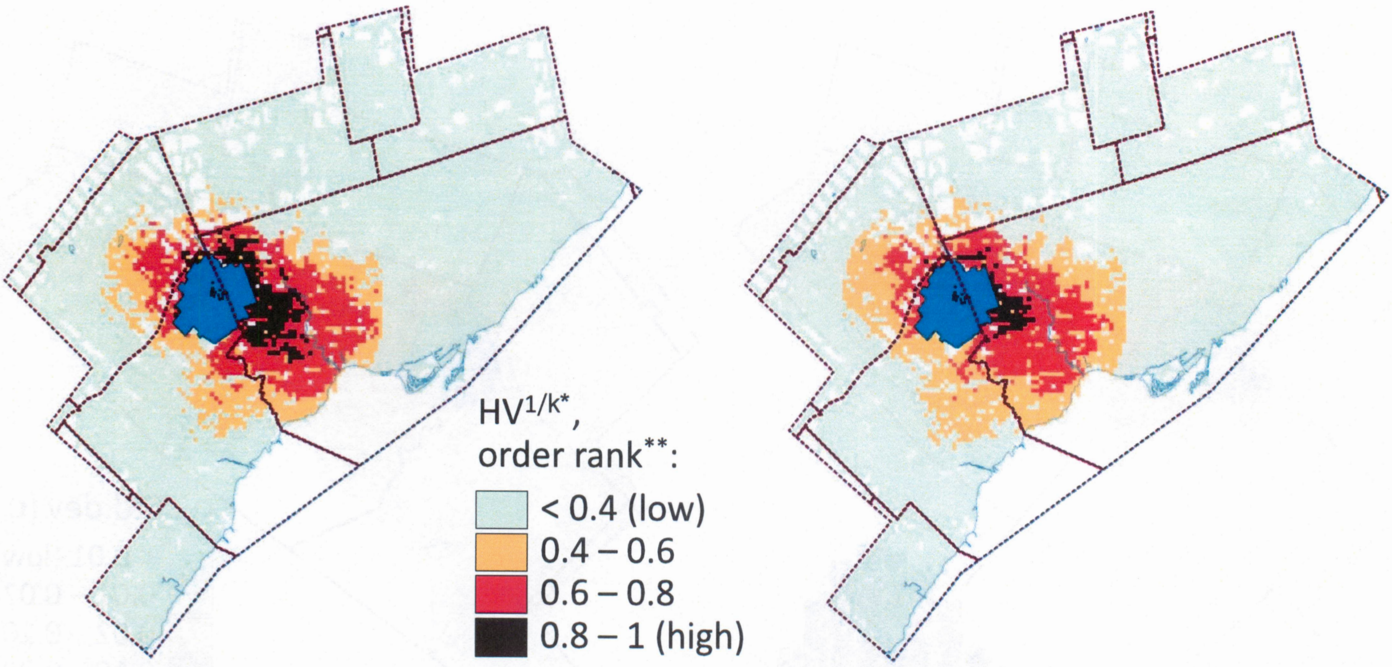
Average ALB arrival rates and their variation



Mean ALB arrival rates, p_j Std. deviations of the p_j values

(Mean and std. deviation of 5000 randomized spread scenarios)

Order rank and hypervolume estimates



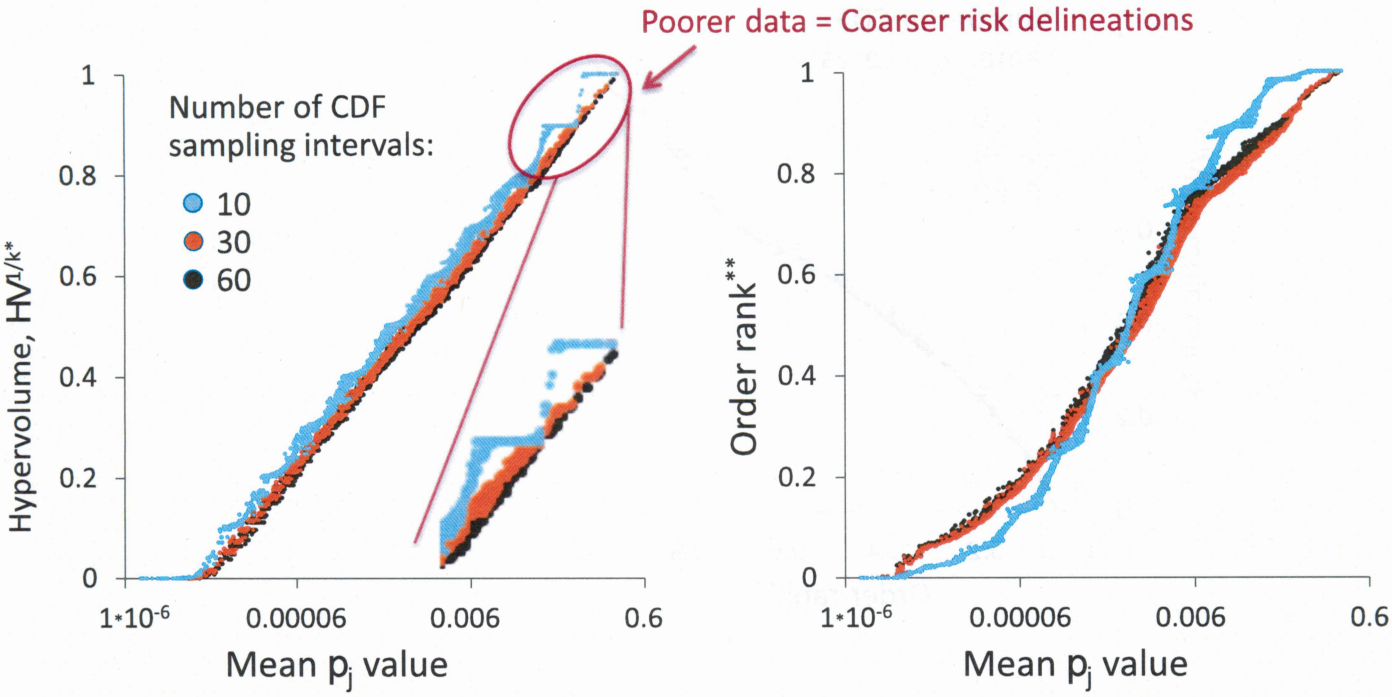
Order rank based on a delineation of the non-dominant sets

Hypervolume of the non-dominant sets, $HV^{1/k*}$

* The k^{th} root of the hypervolume value is shown so that the $HV^{1/k}$ range fits to a 0-1 interval

**Order rank is based on the delineation of nested non-dominant sets starting from the highest-risk sets. The ordinal rank values are rescaled linearly to fit to a 0-1 interval

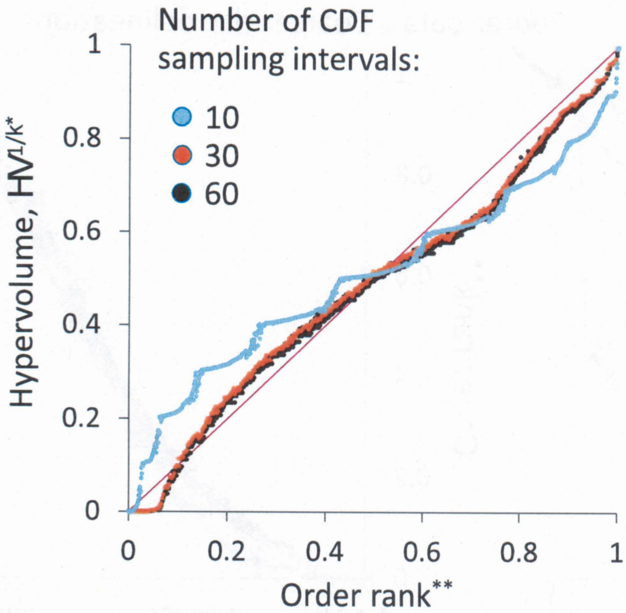
Sensitivity of the *HV* estimates to the number of CDF sampling intervals



* Order rank is based on the delineation of nested non-dominant sets starting from the highest-risk sets. The ordinal rank values are rescaled linearly to fit to a 0-1 interval

**The k^{th} root of the hypervolume value is shown so that the $HV^{1/k}$ range fits to a 0-1 interval

The hypervolume vs. order rank



* The k^{th} root of the hypervolume value is shown so that the $HV^{1/k}$ range fits to a 0-1 interval

**Order rank is based on the delineation of nested non-dominant sets starting from the highest-risk sets. The ordinal rank values are rescaled linearly to fit to a 0-1 interval

Concluding points

The HV approach addresses some important issues in assessing risk from uncertain data:

- Factors in uncertainty by finding non-dominant sets and coarsening risk gradations
- Can be applied when the source of uncertainty (e.g., Type I/II error) cannot be identified
- Can be used with other techniques, such as Pareto-based or multi-criteria rankings

Compared to other CDF tests (e.g., Kolmogorov-Smirnov's, Kuiper's or Earth Mover's Distance), the HV metric characterizes a set and also preserves preference order relationships among the non-dominant sets of distributions

HV can be used to compare different rankings as long as the same CDF sampling intervals are used

The outcomes of the SD tests depend on the number and the spacing of the CDF sampling intervals and the amount of the uncertainty in the data

The approach can help deal with ensemble modelling data - when a distribution of forecasts is generated by the models with qualitatively distinct properties, so that averaging these outcomes to a summary metric is inappropriate

Potentially wide area of applications in assessing risks of disturbances, climate change, floods, invasive pests, dealing with multiple GCM forecasts, etc.

References

Beume, N., Naujoks, B., Emmerich, M. 2007. SMS-EMOA: Multiobjective selection based on dominated hypervolume. European Journal on Operational Research, 181: 1653-1669.

Brockhoff, D., Friedrich, T., Neumann, F. 2008. Analyzing hypervolume indicator based algorithms. In: G. Rudolph et al. (Eds.): Parallel Problem Solving from Nature X, LNCS 5199, pp. 651–660. Springer-Verlag Berlin, Heidelberg.

Buck, J.H. & Marshall, J.M. (2008) Hitchhiking as a secondary dispersal pathway for adult emerald ash borer, Agrilus planipennis. The Great Lakes Entomologist, 41(1-2), 197-198.

Cook, G., Downing, M. 2013. Traffic Pattern Project Report: Methodology for Interpolating Traffic Count Data to a Road Network. USDA APHIS PPQ, Centre for Plant Health Science and Technology. Fort Collins, CO.

Favaro, R., Wichmann, L., Ravn, H.P., Faccoli, M. 2015. Spatial spread and infestation risk assessment in the Asian longhorned beetle, Anoplophora glabripennis. Entomologia Experimentalis et Applicata 155(2): 95-101.

Fleischer, M. 2003. The measure of Pareto optima. Applications to multi-objective metaheuristics. In: C.M. Fonseca, P.J. Fleming, E. Zitzler, K. Deb, and L. Thiele, Eds., Evolutionary Multi-Criterion Optimization. Second International Conference, EMO 2003, pp.519-533, Faro, Portugal. Springer. Lecture Notes in Computer Science. Volume 2632.

Laumanns, M., Zitzler, E., Thiele, L. 2000. A unified model for multi-objective evolutionary algorithms with elitism. In: 2000 Congress on Evolutionary Computation, July 2000. Vol.1, pp 46-53, Piscataway, New Jersey. IEEE Service Center.

Shatz, A.J., Rogan, J., Sangermano, F., Ogneva-Himmelberger, Y., Chen, H. 2013. Characterizing the potential distribution of the invasive Asian longhorned beetle (Anoplophora glabripennis) in Worcester County, Massachusetts. Appl. Geogr. 45: 259-268.

Tetrad 2014. TrafficMetrix Canada. Official website: http://www.tetrad.com/maps_and_data/canada/traffic/

Turgeon, J.J., Pedlar, J., de Groot, P., Smith, M.T., Jones, C., Orr, M., Gasman, B. 2010. Density and location of simulated signs of injury affect efficacy of ground surveys for Asian longhorned beetle. Canadian Entomologist 142: 80-96.

Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C. M., da Fonseca, V. G. 2003. Performance assessment of multiobjective optimizers: An analysis and review. IEEE Transactions on Evolutionary Computation, 7(2):117-132.

Zitzler, E., Thiele, L.. 1998. Multiobjective Optimization using evolutionary algorithms - a comparative case study. In: Conference on Parallel Problem Solving from Nature (PPSN V), pages 292-301, Amsterdam, 1998.