Rank aggregation of local expert knowledge for conservation planning of the critically endangered saola

Nicholas M Wilkinson

Department of Geography

University of Cambridge

Downing Place
Cambridge CB2 3EN

United Kingdom, email nicholas.m.wilkinson@gmail.com

Luong Van Duc*

Faculty of Environmental Science

Hue University of Science

77 Nguyen Huệ street,

Huế city,

Vietnam

*Current address: Graduate School of Science and Technology, Kumamoto University, 2-39-1 Kurokami, Chuo-ku, Kumamoto 860-8555, Japan

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/cobi.12853.

Running head: Rank aggregation of expert knowledge

Keywords: aggregation, expert, local knowledge, saola, Vietnam, Rapid Rural Appraisal, spatial data, pebble distribution method

Abstract

There has been much recent interest in using local knowledge and expert opinion for conservation planning, particularly for hard-to-detect species. However experts' knowledge is often geographically restricted relative to the area of interest. While it is possible to ask for direct estimation of quantities such as population size, relative abundance is easier to estimate. Here we present a new approach combining a Rapid Rural Appraisal method based on community maps with a rank aggregation procedure employing Google's PageRank algorithm. We apply this technique to conservation planning for the poorly-known saola (*Pseudoryx nghetinhensis*) across a conservation landscape in Central Vietnam. Using assessments of relative abundance by local people across the knowledge domain of each village, we produce a prioritization map of an area larger than the knowledge domain of any one village. Because direct information about the saola is so difficult to obtain, these results are valuable for conservation planning and the targeting of field surveys.

Introduction

Spatial prioritization decisions for threatened species are frequently limited by poor and biased data on distribution and relative abundance (Pressey 2004; Rondinini et al. 2006). Field surveys to eliminate this uncertainty may not be worth the cost (Chades et al. 2008; Anadón et al. 2009; Meijaard et al. 2011) and, for some species, extinction may arrive before the survey data (VanderWerf et al. 2006; Jaramillo-Legorreta et al. 2007). Expert opinion, which can include local people's knowledge (Burgman et al. 2011), is increasingly popular in ecology (Drescher et al. 2013). It is therefore unsurprising that conservation scientists increasingly turn to local people to estimate population size (van der Hoeven et al. 2004; Anadón et al. 2009) or occurrence probability (Murray et al. 2009; James et al. 2010).

Several Rapid Rural Appraisal (RRA) techniques exist to elicit rankings (Newing et al. 2010) and can be used to rank sites by the perceived abundance or probability of occurrence of a species. However the problem of combining alternative input rankings into a single aggregate ranking is not trivial and has a long history in the design of electoral systems (Condorcet 1785; Llull, 1299 cited in Hägele & Pukelsheim 2001) and social choice theory (Arrow 1953; Young 1988). Following the work of Dwork et al. (2001) on aggregation of search engine results, various computer science approaches have emerged (Lin 2010). These 'rank aggregation' approaches resemble 'psychological scaling methods' for aggregating expert opinions (Cooke 1991) but make no distributional assumptions (Lin 2010).

In rank aggregation, various procedures, called 'aggregators', are used to produce a single ranking $\mathbf{R}^{\mathbf{A}}$ from a set of input rankings $\{\mathbf{R}^{\mathbf{1}},\ldots,\mathbf{R}^{\mathbf{N}}\}$ of a series of objects. If all rankings include all objects, it may be appropriate to use mean object ranks. However, where the input rankings are incomplete or 'partial', this can be inappropriate as illustrated by the hypothetical example in Table 1. In this simplistic case the four sites ($\mathbf{o_1}$ to $\mathbf{o_4}$) can be objectively ranked in the order $\mathbf{o_1}<\mathbf{o_2}<\mathbf{o_3}<\mathbf{o_4}$. The five experts ($\mathbf{R^1}$ to $\mathbf{R^5}$) never disagree and would all rank the sites in this order but have limited knowledge. Therefore three experts rank site $\mathbf{o_3}$ highest, unaware that $\mathbf{o_2}$ is better. One ($\mathbf{R^1}$) knows both sites and ranks $\mathbf{o_2}$ above $\mathbf{o_3}$, but $\mathbf{o_3}$ gets the higher mean rank because it is more widely known.

It is better to define an optimal $\mathbf{R^A}$ as a ranking with the smallest possible number of disagreements with the inputs in the pairwise comparisons made between objects (Kemeny 1959). The number of disagreements between two rankings $\mathbf{R^i}$ and $\mathbf{R^j}$ is called the Kendall distance ($K(\mathbf{R^i}, \mathbf{R^j})$). We consider two alternative rankings of the objects in Table 1: $\mathbf{R_1^A}$ denotes the ranking $\mathbf{o_1}<\mathbf{o_3}<\mathbf{o_2}<\mathbf{o_4}$ while $\mathbf{R_2^A}$ denotes $\mathbf{o_1}<\mathbf{o_2}<\mathbf{o_3}<\mathbf{o_4}$. The distance $K\left(\mathbf{R_1^A}, \mathbf{R^1}\right)$ is 1 because expert $\mathbf{R^1}$ disagrees with $\mathbf{R_1^A}$ about the pairwise comparison between $\mathbf{o_2}$ and $\mathbf{o_3}$. Conversely, $K\left(\mathbf{R_2^A}, \mathbf{R^1}\right) = \mathbf{0}$ because there are no such disagreements. The other experts do not disagree with either $\mathbf{R_1^A}$ or $\mathbf{R_2^A}$ so the total Kendall distance from the input rankings to $\mathbf{R_1^A}$ is one ($\sum_{x=1}^{N} K(\mathbf{R}_1^{\mathrm{A}}, \mathbf{R^x}) = 1$) while the distance to $\mathbf{R_2^A}$ is zero. $\mathbf{R_2^A}$, is a ‘Kemeny optimal’ ranking because no possible ranking is closer to the inputs. Sometimes there is more than one such optimum. When experts disagree there is no perfect solution with zero distance to the input rankings but a Kemeny optimal ranking can be seen as a maximum likelihood estimate of a true ranking ($\mathbf{R^T}$) which the experts noisily reflect (Young 1988; Adali et al. 2007).

However, when experts disagree, the Kemeny optimal ranking cannot be computed for over 3 objects and algorithms are needed to produce approximations (Dwork et al. 2001). Markov chain (MC) algorithms are based on a process of moving between the ranked objects according to certain rules. For example, the MC4 algorithm of (Dwork et al. 2001) will move from the current object $\mathbf{o_i}$ to a randomly-selected new object $\mathbf{o_j}$ if $\mathbf{o_j}$ is ranked above $\mathbf{o_i}$ by a majority of the rankers. Full details are given in the article but the process can be seen as navigating a graph where the nodes represent objects and the directed edges between them indicate majority preferences (Dwork et al. 2001). Figure 1 shows the graph for the example in Table 1. Certain features of graph structure cause MC4 to get stuck (Dwork et al. 2001) so Adali et al. (2006) employ Google's PageRank algorithm instead. PageRank works in a similar way but adds a 'damping factor' which is usually set to 0.85 to indicate a 15% probability of the algorithm 'jumping' to a random node (Brin & Page 1998). Adali et al further develop their PageRank aggregator to incorporate additional information not used by MC4. Like many rank aggregation algorithms, this PageRank aggregator was developed for 'top-*k* lists'; i.e. those produced by rankers who know all the objects but only rank the *k* best ones.

Dwork et al. (2001) employ a two stage process, feeding the MC output ranking into a second algorithm for 'local Kemenization'. This further reduces the distance from the inputs but the main purpose is to ensure the final ranking meets the 'extended Condorcet criterion' of Truchon (1998) i.e. $\mathbf{o_j}$ is ranked above $\mathbf{o_i}$ if this is the case in the majority of input rankings.

If expert opinion on abundance or probability of presence can be interpreted as rankings of a series of sites then rank aggregation techniques can rank all sites even if no one expert knows them all. This paper applies these approaches to the real-world problem of fine-scale prioritization of conservation action for a poorly-known species.

Specifically we:

a) develop an approach to elicit spatially-explicit expert opinion on relative quantities which can be applied in a rural village setting,

b) aggregate expert opinion on relative abundance for experts with different but overlapping knowledge areas.

Study species

The saola (*Pseudoryx nghetinhensis*) is a distinctive bovid related to buffaloes and cattle (Hassanin 1999), and endemic to the Annamite Mountains of Vietnam and Lao PDR. First described for western science in 1992 (Vu Van Dung et al. 1993), the species remains very poorly known and threatened by indiscriminate hunting (Timmins et al. 2008b). No standardized field survey has ever recorded a saola and they have been detected only five times by camera traps. Direct information on their abundance and distribution has all come from local people. Effective conservation actions, including field surveys, will be difficult and expensive so their spatial targeting appears essential to the species' survival (SWG 2013).

Methods

Study area

The 'Thừa Thiên Huế-Quảng Nam Saola Landscape' (TTH-QN-SL) is considered the most promising site in Vietnam for saola conservation (SWG 2013). Since 2000, the World Wildlife Fund (WWF) and provincial Forest Protection Departments (FPDs) have aimed to target conservation actions to areas where saola are most likely to occur. A single WWF camera-trap image is the only recent field data to support this aim.

The TTH-QN-SL comprises three contiguous new protected areas (PAs) (

Figure 2); the Saola Nature Reserves in Thừa Thiên Huế and Quảng Nam provinces (TTH-SLNR and QN-SLNR) and the recent extension to Bạch Mã National Park (BMNP-E). Saola have also been recorded in adjacent areas including part of the Xe Xap NBCA in Lao. The habitat is exclusively semi-evergreen forest with a topography dominated by the steep-sided valleys of numerous small streams.

Most local people around the TTH-QN-SL belong to the Katu ethnic group with Tà Ôi people forming the majority in A Roàng and A Đớt communes. Both groups are traditionally swidden agriculturalists but post-war government programmes have moved villages closer to roads and striven to eliminate swidden (Lưu Hùng 2007; McElwee 2008). Hunting is common and is mostly now for the restaurant trade although traditional sharing of game meat still occurs in some villages (MacMillan & Nguyen Anh Quoc 2013). There is also an increasing local population of the Vietnamese ethnic majority, the Kinh.

Apart from saola, the TTH-QN-SL also supports sambar (*Rusa unicolor*), serow (*Capricornis milneedwardsi*), wild pig (*Sus scrofa*) and at least one species of chevrotain (*Tragulus*) of uncertain taxonomic status (Meijaard & Groves 2004). Local Katu and Tà Ôi people refer to these taxa with single-word ‘unitary lexemes’ (Conklin 1968) or ‘productive primary’ composite lexemes (Berlin 1992) based on the head term *xoong* or ‘beast’. Some synonyms exist as do distinct names for life stages of *S. scrofa*. Among muntjacs (*Muntiacus*), local people typically describe a larger, red-coloured ‘young forest muntjac’ (YFM) and a smaller, darker, ‘old-forest muntjac’ (OFM). These are usually referred to *Muntiacus vaginalis* and *M. truongsonensis* respectively (Giao et al. 1998) but there are taxonomic issues with the latter identification and more species may be present (Timmins et al. 2008a). The endangered large-antlered muntjac (*M. vuquangensis*) is represented by hunters’ trophies but apparently localized and possibly extirpated. Available evidence suggests all large mammals have declined (Long 2005; Turvey et al. 2015).

Data were collected in 2010 and 2011, prior to PA establishment. Although national law proscribes most forest resource extraction, the *de facto* situation was open access. However, as FPD staff could apprehend hunters transporting or selling game, interviewees could be reluctant to discuss hunting.

Village and expert selection

Local experts comprised any inhabitants of villages using the forests of the TTH-QN-SL who were considered by their peers to be knowledgeable about forest animals. We compiled a shortlist of 120 villages based on geographical proximity and previous interview data (Wilkinson 2007) and this was reduced to a sample of 58 (

Figure 2, Appendix S1). Of these, 17 were predominantly ethnic Tà Ôi and the remainder Katu. Although village selection was based on knowledge of the 3 PAs, the knowledge areas of the selected villages also included adjoining areas, including some in Lao.

One mapping workshop was conducted per village by one of two teams of two young researchers from the Hue University of Science. On arrival, the team introduced themselves to the village headman in accordance with local procedure, showing their permission letter from the FPD. The headman was then asked to invite 7-10 people "who knew about the forest and wild animals" to meet in the village communal house or commune Learning Centre. Only men of Katu or Tà Ôi ethnicity were selected, knowledge of forest animals and hunting being seen as a male domain (Lưu Hùng 2007).

Invitees were made aware that the purpose of the work was nature conservation and that data would be shared with FPD and a wildlife protection organization. No government staff were present. Invitees were not required to participate but were compensated VND 20,000 (ca USD 1.00) for a half-day's work if they did.

Elicitation method – the bean data

Community maps were produced by eliciting local names of landscape features appearing on published maps. We used black and white, 841 x 1189 mm base maps derived from publicly available GIS data on streams, major roads and commune People's Committee offices. We added important, ground-truthed place names where possible. The only administrative boundaries included were provincial and national boundaries following mountain ridges. We developed this approach after a pilot survey showed most people could easily interpret published maps and that stream valleys were their primary geographical reference points with even very small streams having names. Base map extents were based on information from pilot surveys and previous community mapping by Thừa Thiên Huế FPD.

Scale varied between 1:15,000 and 1:30,000. Map extent was increased as a precautionary measure when information about a village's use area was insufficient.

Mapping workshops lasted half a day. Interview teams began by indicating the village location and other key features. An initial period of orientation and naming of major streams usually led onto rapid identification of small streams and other features with some being added to the map. Simultaneously, a separate interview group from the same village produced a map of the same area for a separate study of forest product use. The two maps were compared to resolve discrepancies.

We entered place names into a GIS, checking for agreement among villages and resolving the infrequent discrepancies through recourse to other data sources or individual local experts. Streams added by participants were matched to valleys on published maps. Unrelated fieldwork trips over most of the TTH-QN-SL confirmed the maps' accuracy. From the GIS, we produced a revised series of base maps showing all well-supported stream names.

In a second half-day group interview, experts were asked to place beans on revised base-maps to represent relative quantities, an approach based on the RRA 'Pebble Distribution Method' (Lynam et al. 2007; Jones et al. 2008). More beans in an area indicated a higher value and few beans a lower value. This method was used to enquire about relative abundance of all artiodactyl species (species data) as well as frequency of forest visits by villagers (visit data). Questions were in the following style:

1. "Where do people from this village go, when they go to the forest?"
2. "Where, in the forest, do you think there are most wild pigs?"

3. “Where, in the forest, would you have most chance of finding a saola?”

Style 3 questions were used for rarer species, usually saola or sambar, when participants responded that the equivalent style 2 question was meaningless, the species being everywhere rare. For muntjacs, locally recognised ‘ethnospecies’ were discussed with the interview group and mapped separately.

As in the mapping interviews, participants generally took some time to understand the method but responded quickly thereafter. However, researchers were not always successful in discouraging inappropriate clustering of beans around place names.

Analytical Methods

Interpreting bean data

Photographs of the maps with beans were geo-registered with an affine transformation in ArcGIS (ESRI 2011) and bean locations manually entered as points. We scored each government forest unit (*tieu khu*) based on the density of beans within an area comprising the unit and a 200m buffer. Forest units are used in conservation planning and their boundaries follow the streams and ridgelines used by local people to describe the landscape. No such divisions exist in Lao so we used similar landscape features to divide the Lao territory covered by our survey into units of equivalent size. Buffering reduced the problem of arbitrary assignment of beans along their boundaries. For saola alone we assigned an alternative subjective score based on visual inspection of the bean maps by one researcher (NMW). .

We assumed each village had a discrete knowledge area comprising a number of units. We delineated these areas subjectively for each village on the basis of visit data supported by species and forest product data. Four sets of criteria produced alternative nested estimates of knowledge area numbered Kn1 to Kn4. Kn1 , based only on the visit data, is the most conservative while Kn4 is the least, , including any units in which beans were placed in response to any question. Kn2 and Kn3 are intermediate with full criteria and example maps given in Appendix S1. Data for each of the 7 species plus the subjective scores for saola were assessed under each set of criteria, giving a total of 32 species-knowledge combinations.

For each combination, a set of s partial rankings of N units was derived for the s villages. For each village the highest-scoring unit was ranked 1 but data were considered missing for units not known to that village. Villages which did not rate at least two units differently were excluded, as were units not known to any of this reduced set of villages. Values of N and s therefore vary among species-knowledge combinations (Appendix S1).

Aggregators

We used the 'igraph' package (Csardi & Nepusz 2006) to create novel routines in R 3.1.1 (R Development Core Team 2013) for the MC4 algorithm and a PageRank aggregator . Code is presented in Appendix S2.Unlike that of Adali et al. (2006), our simpler PageRank aggregator makes use of only the information used by MC4. More complex aggregators were abandoned due to poor performance in simulations (see below) but code is included in Appendix S2.

Adali et al. use PageRank to aggregate top-k lists, assuming that input rankings exclude because they are not ranked highly enough. PageRank becomes a biased aggregator when objects are excluded because they are unknown to the rankers. We therefore derive two alternative aggregate rankings with opposite bias and take the average rank of each unit across both. These are the alternative ranked PageRank scores from a graph equivalent to that in Fig. 1 and the reverse-ranked scores from the converse graph. Simulations detected no bias and results were similar to those from MC4. Simulations indicated no appreciable effect from changing the PageRank damping factor to 0.15 or 0.50 (c.f. Adali et al. 2006) so we retained the standard value of 0.85. We trialled two local Kemenization procedures, the algorithm of Dwork et al. (2001) and the slower but more powerful 'Iterative Best Flip' (IBF) method of Adali et al. (2006).

Before conducting the aggregation we constructed graphs of majority preference (c.f. Figure 1) to check for features which might limit algorithm performance. We checked for separate weakly connected components (WCCs); i.e. distinct sets of nodes which are not interconnected and therefore cannot be aggregated together. Graphs based on criteria Kn3 and Kn4 comprised a single WCC but the 18 units from Đồng Giang district formed a separate WCC under criteria Kn1 and Kn2 and were excluded from further analyses.

Strongly connected components (SCCs) are sets of nodes within which it is possible to reach any node from any other by following the directed edges. The MC4 algorithm will get stuck in SCCs which are sinks, i.e. which have edges leading into them but none leading out (Dwork et al. 2001). Such 'sink-SCCs' were present in all graphs making PageRank a more appropriate aggregator than MC4 (Adali et al. 2006).

Simulations

We conducted simulations to test algorithm performance. In each run a set of s partial rankings $\{\mathbf{R}^{1},\ldots,\mathbf{R}^{s}\}$ was generated for a 'true' ranking ($\mathbf{R}^{\mathbf{T}}$) of N units. A new, random $\mathbf{R}^{\mathbf{T}}$ was used in each run but the data structure, including village knowledge areas, matched that of the real saola data under Kn3. We simulated the rankers' imperfect knowledge by adding a random Normal error term (ϵ) to the rank of each unit. The standard deviation of ϵ was equal to the number of units known to that ranker multiplied by a constant (α). We conducted four alternative sets of 100 simulation runs where α = {0.0, 0.1, 0.2, 0.5}. In two further sets we used positively and negatively heteroscedastic noise terms where α decreased or increased linearly with unit rank between extremes of 0.0 and 0.5.

For each simulation run we derived an aggregate ranking $\mathbf{R}^{\mathbf{A}}$ from our PageRank aggregator and calculated the total Kendall distance of $\mathbf{R}^{\mathbf{A}}$ from the true ranking $K(\mathbf{R}^{\mathbf{A}}, \mathbf{R}^{\mathbf{T}})$ and from the input ranking($\sum_{i=1}^{s} K(\mathbf{R}^{\mathbf{A}}, \mathbf{R}^{\boldsymbol{i}})$). Alternative versions of Kendall distance differ in the way they treat tied ranks and 'void comparisons' for which data are missing in one or both rankings (Fagin et al. 2004). We exclude void comparisons and use the number of non-void comparisons to Normalize Kendall distances. Comparisons involving ties are treated as agreements.

We also calculated the mean rank error $\sum_{j=1}^{N}\left|r_{\mathbf{R}^{\mathrm{A}}}(\mathbf{o_j}) - r_{\mathbf{R}^{\mathrm{T}}}(\mathbf{o_j})\right|$ where $r_{\mathbf{R}^{i}}(\mathbf{o_j})$ indicates the rank of unit $\mathbf{o_j}$ by village $\mathbf{R}^{\boldsymbol{i}}$. The top-5 precision is the proportion of the top 5 objects in $\mathbf{R}^{\mathbf{T}}$ which are also in the top 5 in $\mathbf{R}^{\mathbf{A}}$. We calculated the top-*k* precision for *k*= {5,10,20}. For each of the six sets of simulations, we took the mean of each of these six performance measures across 100 runs.

To check whether the aggregator was biased towards particular units, we took the mean of *b* over 100 runs, where $b = 1$ when $r_{\mathbf{R}^{\mathrm{A}}}(\mathbf{o_j}) > r_{\mathbf{R}^{\mathrm{T}}}(\mathbf{o_j})$ and 0 otherwise. We also tested for correlations between a unit's mean rank error in the first 50 simulation runs and the same measure in the second 50.

We used the same simulations and performance measures to assess the value of local Kemenization but found only minor incremental improvement on the PageRank output from either LK procedure ($K(\mathbf{R}^{\mathbf{A}}, \mathbf{R}^{\mathbf{T}})$ never reduced by > 0.01). As the Extended Condorcet Criterion is not crucial, we therefore abandoned local Kemenization.

Aggregation of real data.

Using the PageRank aggregator, we produced complete aggregate rankings from the set of partial rankings for each species-knowledge combination and mapped the results.

To assess the sensitivity of the final aggregate rankings to estimate of knowledge area we measured the Kendall distance and top-*k* precision between all equivalent pairs of output rankings under criteria Kn1 to Kn4 and inspected a matrix of scatterplots.

Results

Data collected

Under the Kn4 criteria, knowledge areas ranged from 3 to 66 units (mean =23.2, variance=352.8) with similarly large variation under the other criteria. Some units were known by to up to 25 villages, others only to 1. The dataset comprised between 105 and 143 units depending on knowledge criteria.

A total of 42 of 58 villages reported two kinds of muntjac assignable to the YFM-OFM types. Although a third ethnospecies was described in two communes, no description reliably matched *M. vuquangensis*. Numbers of villages reporting on other species were: saola 48, serow 57, sambar 47, wild pig 57 and chevrotain 28. Under more conservative knowledge estimates data from up to four villages was dropped for certain species.

For the saola data under the Kn4 criteria, some comparative information was given for N=135 units. Comparing two rankings of 135 units would ordinarily involve (N (N -1)/2)=9,045 pairwise comparisons while 434,160 such comparisons (48 × 9,045) would be made while calculating the total Kendall distance of $\mathbf{R}^{\mathbf{A}}$ from the input rankings. In fact, even under Kn4, most comparisons were void and only 24,134 (5.6%) were actually made. This is the number used to Normalize the Kendall distance from aggregate ranking to input rankings. Normalized Kendall distances range from 0 to 1 with 0.5 meaning that the two rankings disagree 50% of the time and are therefore no closer than would be expected by chance.

Of the 24,134 non-void comparisons, 57.1% involved ties in the input rankings. For density scores, ties arose only when both units were scored zero. Subjective saola scores also included ties between non-zero scores and the equivalent figure was 59.8%. There were fewer zero scores, and hence fewer ties, for other knowledge criteria and other species (Appendix S1).

Performance in simulations

Performance statistics from the simulations are given in Table 2. When rankers make no mistakes (α=0.0), the PageRank aggregator performs well. $K(\mathbf{R}^{\mathbf{A}}, \mathbf{R}^{\mathbf{T}})$ =0.06, meaning only 6% of the pairwise comparisons made in the output disagree with those in the true ranking . However, this figure rises to almost 25% when α =0.5. For reference, α =0.5 indicates a situation where rankers are highly inaccurate, with a 32% chance of falsely ranking the best unit known to them below the worst. The aggregator performs worse when error is greatest for the highest-ranked objects (negative heteroscedasticity) than in the reverse situation.

PageRank outperformed MC4 on all metrics in 5 of the 6 sets of simulations while when MC4 performed very slightly better when α =0.5 (reduction in $K(\mathbf{R}^{\mathbf{A}}, \mathbf{R}^{\mathbf{T}})$ and $\sum_{i=1}^{s} K(\mathbf{R}^{\mathbf{A}}, \mathbf{R}^{i})$ of < 0.003 when $\mathbf{R}^{\mathbf{A}}$ derived from MC4).

Aggregation Results

Results from our PageRank aggregator did not resemble a naïve interpretation based on ranked mean scores (

Figure 3), while MC4 and PageRank outputs were similar in spatial pattern (Appendix S3).

Results were sensitive to the criteria used for delineating village knowledge areas (

Figure 4). For subjective saola scores, the 6 possible comparisons between 4 alternative knowledge estimates had a mean Kendall distance from one another of 0.208 (20.8% disagreement) and the top-5 precision was always <0.5. Kendall distance between the subjective and density scores for saola ranged between 0.161 under Kn4 to 0.234 under Kn1. However maps reveal robustness in the broad spatial pattern (

Figure 4).

Discussion

For the cryptic and little known saola, we have used local expert opinion to answer a non-local question. By combining RRA-based techniques with rank aggregation we have mapped perceptions of relative abundance across a conservation landscape, despite the inherent spatial limitations of local knowledge. Our results have direct conservation application and earlier versions of the maps in Figs 3 and 4 have been used in the prioritization of sites for snare removal and camera trapping. The mapping and aggregation approaches we develop represent a considerable advance for other cryptic and threatened species and future developments may further increase their utility.

Our aggregate rankings model a 'consensus view' among the villages. We investigated this approach precisely because direct survey data were unavailable and we cannot therefore compare our results with a true ranking. However the pattern is believable as the aggregate ranking mainly prioritizes areas further from villages which are likely to have been less heavily hunted. The biased mean ranks give the opposite pattern (

Figure 3).

The figures for saola in Table 3 show that our aggregate ranking is a fairly good reflection of the consensus, disagreeing with only 10-15% of the pairwise comparisons made by the experts. This does not imply 85-90% agreement, however, as ties are prevalent. If one ranking ranks unit $\mathbf{o_i}$ and $\mathbf{o_j}$ differently while another scores both as zero, this is not counted as a disagreement. Fewer disagreements will therefore occur for a species considered absent across much of the landscape. This likely explains the lower figures in Table 3 for saola compared to the commoner pigs and muntjacs.

The simulation results suggest that low distance to the input rankings indicates a similarly low distance to the true ranking if the different villages really are observing the same pattern with a degree of error. However, even where this assumption is met, top-*k* precision is low. If prioritization must select a small number of forest units, then there is a clear value of additional information.

Our PageRank aggregator outperforms MC4 and avoids the necessity of identifying ‘sink-SCCs’. Our use of PageRank is based on the work of Adali et al. (2006) but their more complex version performed poorly on our data, probably because of overfitting (c.f. Adali et al. 2007). Due to limited knowledge of the rankers, the amount of missing data is high in our example compared to the datasets investigated by Dwork et al. (2001), Adali et al. (2006) and others. The standard PageRank aggregator was biased in this situation and, while our solution performs well in simulations, we offer no proof that it is unbiased. Testing this and other algorithms on a range of realistic simulations would be a valuable focus of future work. A more statistical approach to the problem would increase its utility in decision-making as would methods to identify differences of opinion among experts (Adali et al. 2007).

A more complex challenge would be to develop a method able to use the spatially explicit information gathered by our map-based elicitation method. The maps suggest disagreement among villages at the level of forest units masks agreement about wider spatial patterns. Inspection of the raw data shows that certain hills, seen as wildlife hotspots, are divided among several units which receive high scores relative to other areas but are scored arbitrarily relative to one another. This partially explains the high sensitivity to knowledge area and scoring system and shows that disagreement about the relative ranking of forest units may mask agreement at a broader scale.

Aggregation of expert opinion is an underexplored topic in conservation biology even though decisions frequently depend on expert judgement (Burgman 2005). Given that the most valuable expert knowledge is localized relative to the area of interest, the general approach we develop here should be widely applicable.

Acknowledgments

Special thanks to all villagers who participated in our survey and the village headmen and commune officials who organized meetings and facilitated our work. Thanks to WWF Vietnam and the FPDs of Thừa Thiên Huế and Quảng Nam Provinces for supporting us and handling paperwork and logistics. At WWF, particular thanks go to O.D.B. Tri and L.V. Hung for helping our teams on their first surveys. Data collection was made possible through the support of colleagues from the Faculty of Environmental Science at Huế University of Science. T.D. Thanh, N.T.T. Hieu and N.T. Hoang collected much of the data. P.K. Lieu and H.D.T. Hoang helped design and facilitate the survey. Finally, thanks to S. Adali of Rensselaer Polytechnic Institute for advice on rank aggregation, M. Hall of Imperial College London for coding advice and R. Puri of the University of Kent for advice on field methods.

T. Amano of Cambridge University's Zoology Department and M. Smith of Microsoft Research in Cambridge both offered comments on a draft manuscript. Special thanks to N. Leader-Williams of Cambridge for his support as PhD supervisor to NMW. This work was supported by the UK Government through the Darwin Initiative Project 17-008.

Supporting Information

Criteria and example maps showing knowledge areas (Appendix S1), R code used in the analysis (Appendix S2) and output maps for all species (Appendix S3) are available online. The authors are solely responsible for the content and functionality of these materials. Queries (other than absence of the material) should be directed to the corresponding author.

Literature cited

Adali, S., B. Hill, and M. Magdon-Ismail. 2006. The Impact of Ranker Quality on Rank Aggregation Algorithms: Information vs. Robustness. Pages 37–3722nd International Conference on Data Engineering Workshops, 2006. Proceedings.

Adali, S., B. Hill, and M. Magdon-Ismail. 2007. Information vs. Robustness in Rank Aggregation: Models, Algorithms and a Statistical Framework for Evaluation. Journal of Digital Information Management **Special Issue on Web Information Retrieval 5**(5).

Anadón, J. D., A. Giménez, R. Ballestar, and I. Pérez. 2009. Evaluation of Local Ecological Knowledge as a Method for Collecting Extensive Data on Animal Abundance. Conservation Biology **23**:617–625.

Arrow, K. J. 1953. Social Choice and Individual Values. Cowles Foundation Monographs; New York: Wiley 1964.

Berlin, B. 1992. Ethnobiological Classification: Principles of Categorization of Plants and Animals in Traditional Societies. Princeton University Press.

Brin, S., and L. Page. 1998. The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems **30**:107–117.

Burgman, M. 2005. Risks and Decisions for Conservation and Environmental Management. Cambridge University Press.

Burgman, M., A. Carr, L. Godden, R. Gregory, M. McBride, L. Flander, and L. Maguire. 2011. Redefining expertise and improving ecological judgment. Conservation Letters **4**:81–87.

Chades, I., E. McDonald-Madden, M. A. McCarthy, B. Wintle, M. Linkie, and H. P. Possingham. 2008. When to stop managing or surveying cryptic threatened species. Proceedings of the National Academy of Sciences **105**:13936–13940.

Condorcet, J.-A.-N. de C., marquis de. 1785. Essai sur l’application de l’analyse à la probabilité des décisions rendues à la pluralité des voix. Imprimerie royale, Paris.

Conklin, H. C. 1968. Lexicographical treatment of folk taxonomies. Readings in the Sociology of Language **1968**:414–433.

Cooke, R. M. 1991. Experts in uncertainty: opinion and subjective probability in science. Oxford University Press, New York.

Csardi, G., and T. Nepusz. 2006. The igraph software package for complex network research. InterJournal, Complex Systems **1695**:1–9.

Drescher, M., A. H. Perera, C. J. Johnson, L. J. Buse, C. A. Drew, and M. A. Burgman. 2013. Toward rigorous use of expert knowledge in ecological research. Ecosphere **4**:1–26.

Dwork, C., R. Kumar, M. Naor, and D. Sivakumar. 2001. Rank aggregation methods for the Web. Pages 613–622 Proceedings of the 10th international conference on World Wide Web. ACM, New York. DOI. 10.1145/371920.372165

ESRI. 2011. ArcGIS Desktop. Environmental Systems Research Institute, Redlands, CA.

Fagin, R., R. Kumar, M. Mahdian, D. Sivakumar, and E. Vee. 2004. Comparing and Aggregating Rankings with Ties. Pages 47–58 Proceedings of the 23rd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems. ACM, New York, NY. DOI 10.1145/1055558.1055568

Giao, P. M., D. Tuoc, V. V. Dung, E. D. Wikramanayake, G. Amato, P. Arctander, and J. R. MacKinnon. 1998. Description of Muntiacus truongsonensis, a new species of muntjac (Artiodactyla: Muntiacidae) from Central Vietnam, and implications for conservation. Animal Conservation **1**:61–68.

Hägele, G., and F. Pukelsheim. 2001. Llull's writings on electoral systems. Studia Lulliana **41**:3–38.

Hassanin, A. 1999. Evolutionary affinities of the enigmatic saola (Pseudoryx nghetinhensis) in the context of the molecular phylogeny of Bovidae. Proceedings of the Royal Society B: Biological Sciences **266**:893–900.

James, A., S. L. Choy, and K. Mengersen. 2010. Elicitator: An expert elicitation tool for regression in ecology. Environmental Modelling & Software **25**:129–145.

Jaramillo-Legorreta, A., L. Rojas-Bracho, R. L. B. Jr, A. J. Read, R. R. Reeves, K. Ralls, and B. L. Taylor. 2007. Saving the Vaquita: Immediate Action, Not More Data. Conservation Biology **21**:1653–1655.

Jones, J. P. G., M. M. Andriamarovololona, N. Hockley, J. M. Gibbons, and E. J. Milner-Gulland. 2008. Testing the use of interviews as a tool for monitoring trends in the harvesting of wild species. Journal of Applied Ecology **45**:1205–1212.

Kemeny, J. G. 1959. Mathematics without Numbers. Daedalus **88**:577–591.

Lin, S. 2010. Rank aggregation methods. Wiley Interdisciplinary Reviews: Computational Statistics **2**:555–570.

Long, B. 2005. Identification of priority areas for integrated conservation in Quang Nam province, Vietnam. PhD. Durrell Institute of Conservation and Ecology, University of Kent.

Lưu Hùng. 2007. A Contribution to Katu Ethnography. The Gioi Publishers, Ha Noi, Vietnam.

Lynam, T., W. De Jong, D. Sheil, T. Kusumanto, and K. Evans. 2007. Review of Tools for Incorporating Community Knowledge, Preferences, and Values into Decision Making in Natural Resources Management. Ecology and Society **12**. http://www.ecologyandsociety.org/vol12/iss1/art5/.

MacMillan, D. C., and Nguyen Anh Quoc. 2013. Factors influencing the illegal harvest of wildlife by trapping and snaring among the Katu ethnic group in Vietnam. Oryx **48**:304–312.

McElwee, P. 2008. "Blood relatives" or uneasy neighbors? kinh migrant and ethnic minority interactions in the Trường Sơn mountains. Journal of Vietnamese Studies **3**:81–116.

Meijaard, E., and C. P. Groves. 2004. A taxonomic revision of the Tragulus mouse-deer (Artiodactyla). Zoological Journal of the Linnean Society **140**:63–102.

Meijaard, E., K. Mengersen, D. Buchori, A. Nurcahyo, M. Ancrenaz, S. Wich, S. S. . Atmoko, A. Tjiu, D. Prasetyo, and Y. Hadiprakarsa. 2011. Why Don't We Ask? A Complementary Method for Assessing the Status of Great Apes. PloS one **6**:e18008.

Murray, J. V., A. W. Goldizen, R. A. O'Leary, C. A. McAlpine, H. P. Possingham, and S. L. Choy. 2009. How useful is expert opinion for predicting the distribution of a species within and beyond the region of expertise? A case study using brush-tailed rock-wallabies Petrogale penicillata. Journal of Applied Ecology **46**:842–851.

Newing, H., C. Eagle, R. Puri, and C. W. Watson, editors. 2010. Conducting Research in Conservation: Social Science Methods and Practice. Routledge, UK.

Pressey, R. L. 2004. Conservation planning and biodiversity: assembling the best data for the job. Conservation Biology **18**:1677–1681.

R Development Core Team. 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Rondinini, C., K. A. Wilson, L. Boitani, H. Grantham, and H. P. Possingham. 2006. Tradeoffs of different types of species occurrence data for use in systematic conservation planning. Ecology Letters **9**:1136–1145.

SWG. 2013. Conservation Through Collaboration: Proceedings of the 3rd Meeting of the Saola Working Group. Saola Working Group, Asian Wild Cattle Specialist Group of the IUCN Species Survival Commission, Vientiane, Lao PDR.

Timmins, R. J., J. W. Duckworth, and B. Long. 2008a. Muntiacus truongsonensis. Page IUCN Red List of Threatened Species. http://www.iucnredlist.org/details/18597

Timmins, R. J., W. G. Robichaud, B. Long, S. Hedges, R. Steinmetz, A. Abramov, Do Tuoc, D. P. Mallon, and IUCN. 2008b. Pseudoryx nghetinhensis. Page 2008 IUCN Red List of Threatened Species. http://www.iucnredlist.org/details/18597

Truchon, M. 1998. An Extension of the Condorcet Criterion and Kemeny Orders. cahier 98-15. Centre de Recherche en Économie et Finance Appliqueé, University of Luxembourg.

Turvey, S. T. et al. 2015. Interview-based sighting histories can inform regional conservation prioritization for highly threatened cryptic species. Journal of Applied Ecology **52**:422–433.

van der Hoeven, C. A., W. F. de Boer, and H. H. T. Prins. 2004. Pooling local expert opinions for estimating mammal densities in tropical rainforests. Journal for Nature Conservation **12**:193–204.

VanderWerf, E. A., J. J. Groombridge, J. Scott Fretz, and K. J. Swinnerton. 2006. Decision analysis to guide recovery of the po ‘ouli, a critically endangered Hawaiian honeycreeper. Biological Conservation **129**:383–392.

Vu Van Dung, Pham Mong Giao, Nguyen Ngoc Chinh, D. Tuoc, P. Arctander, and J. MacKinnon. 1993. A new species of living bovid from Vietnam. Nature **363**:443–445.

Wilkinson, N. M. 2007. Status of the southernmost population of Saola Pseudoryx nghetinhensis in Thừa Thiên Huế and Quảng Nam provinces, Vietnam. A critical review of existing evidence with recommendations for future conservation and research. Unpublished report. WWF, Ha Noi, Vietnam.

Young, H. P. 1988. Condorcet’s Theory of Voting. American Political Science Review **82**:1231–1244.

Table 1: Hypothetical ranking of four sites $\{\mathbf{o_1}$ to $\mathbf{o_4}\}$, by five experts $\{\mathbf{R^1}$ to $\mathbf{R^5}\}$, each with limited knowledge

	$\mathbf{R^1}$	$\mathbf{R^2}$	$\mathbf{R^3}$	$\mathbf{R^4}$	$\mathbf{R^5}$	mean rank
$\mathbf{o_1}$	1	NA*	NA	NA	1	1
$\mathbf{o_2}$	2	NA	NA	NA	2	2
$\mathbf{o_3}$	3	1	1	1	NA	1.5
$\mathbf{o_4}$	NA	2	2	2	3	2.25

*"NA" indicates missing data due to limited knowledge of the individual experts.

Table 2: Mean performance measures for each of 6 sets of 100 simulations for PageRank aggregator
Set of simulations

Heteroscedasticity:	none				+ve[a]	-ve
Noise parameter (α)[b]	0.0	0.1	0.2	0.5	0≤ α ≥0.5	0≤ α ≥0.5
Kendall distance to input rankings[c]	0.012	0.087	0.168	0.299	0.013	0.071
Kendall distance to true ranking	0.060	0.089	0.138	0.247	0.060	0.113
Mean rank error	5.39	7.72	11.84	20.85	5.44	9.53
Top-*k* Precision[d]						
k=5	0.73	0.51	0.39	0.21	0.71	0.52
k=10	0.79	0.67	0.54	0.33	0.79	0.62
k=20	0.87	0.79	0.68	0.51	0.87	0.76

[a] for heteroscedastic error terms, α varies between 0 and 0.5. Under negative heteroscedasticity it is 0.5 for the highest ranked object and 0 for the lowest. Positive heteroscedasticity is the reverse.
[b] determines standard deviation of error term; see text for details
[c] Normalized Kendall distances are from 0 (perfect agreement) to 1 with 0.5 indicating no correlation. See text for details of all performance metrics.
[d] measures whether the aggregate ranking correctly identifies the top *k* objects in the true ranking. High values indicate good performance in contrast to other measures.

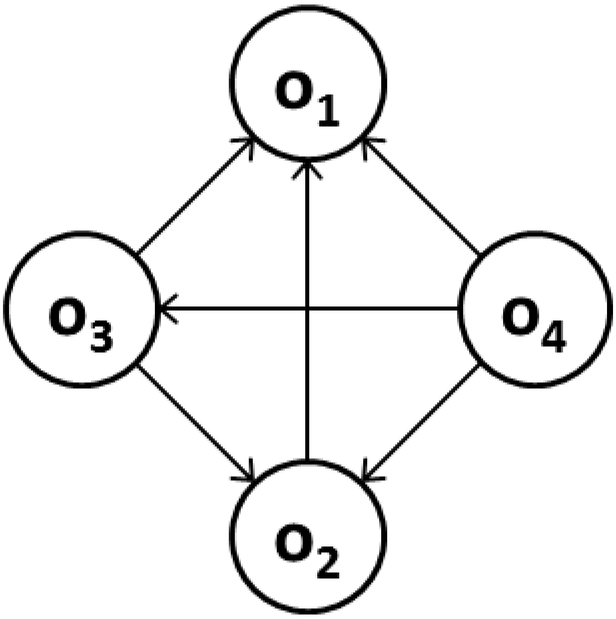
Table 3: Total Kendall distance between partial expert rankings and PageRank aggregate ranking.

	Kn1[e]	Kn2	Kn3	Kn4
Saola	0.126	0.112	0.111	0.138
Saola (subj scores)	0.118	0.108	0.104	0.143
Serow	0.222	0.207	0.198	0.216
Sambar	0.148	0.180	0.150	0.160
OFM (muntjac sp)	0.220	0.204	0.197	0.256
YFM (muntjac sp)	0.306	0.236	0.263	0.300
Wild Pig	0.281	0.267	0.261	0.301
Chevrotain	0.194	0.200	0.203	0.208

Distance measure = Total Kendall distance from set of input rankings, Normalized by number of non-void comparisons between items.

[e] Four alternative sets of estimates of village knowledge area, from most to least conservative.

Figure 1: Graphical representation of the comparisons between sites in Table 1



Nodes represent sites. Directed edges (arrows) represent majority preferences between sites so the edge ($\mathbf{o}_2$, $\mathbf{o}_1$) indicates that a majority of experts rank $\mathbf{o}_1$ above $\mathbf{o}_2$.

Figure 2 Map of the 'Saola Landscape' in Thừa Thiên Huế and Quảng Nam Provinces, Vietnam

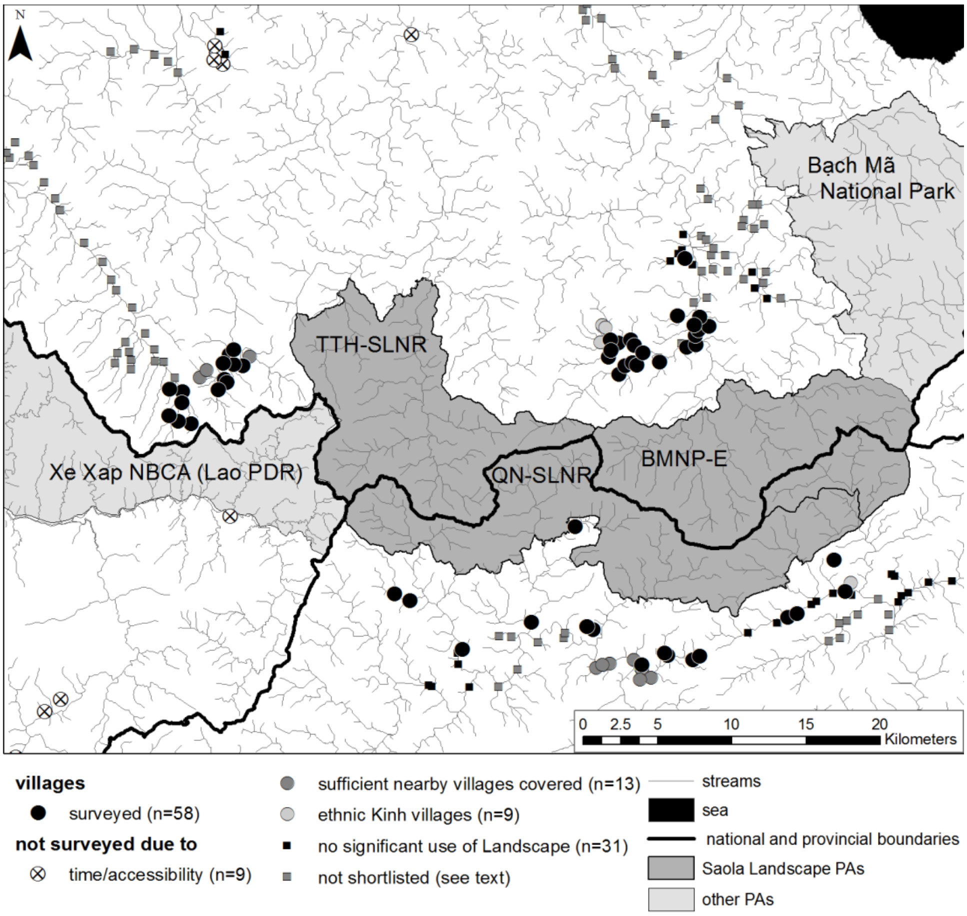
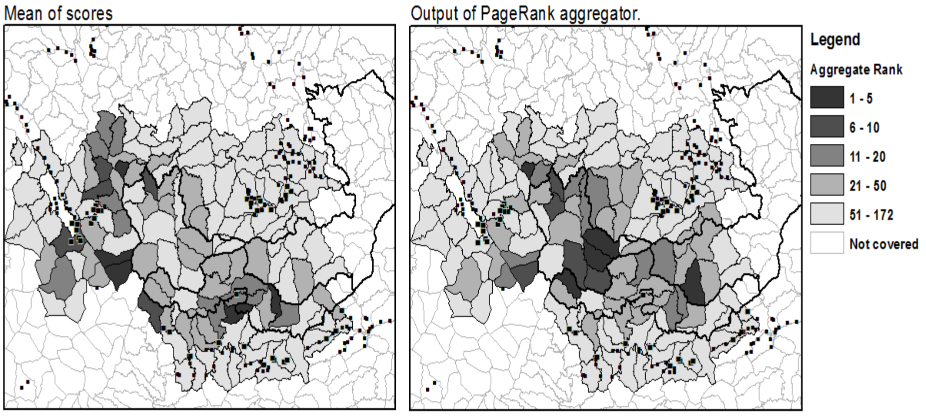
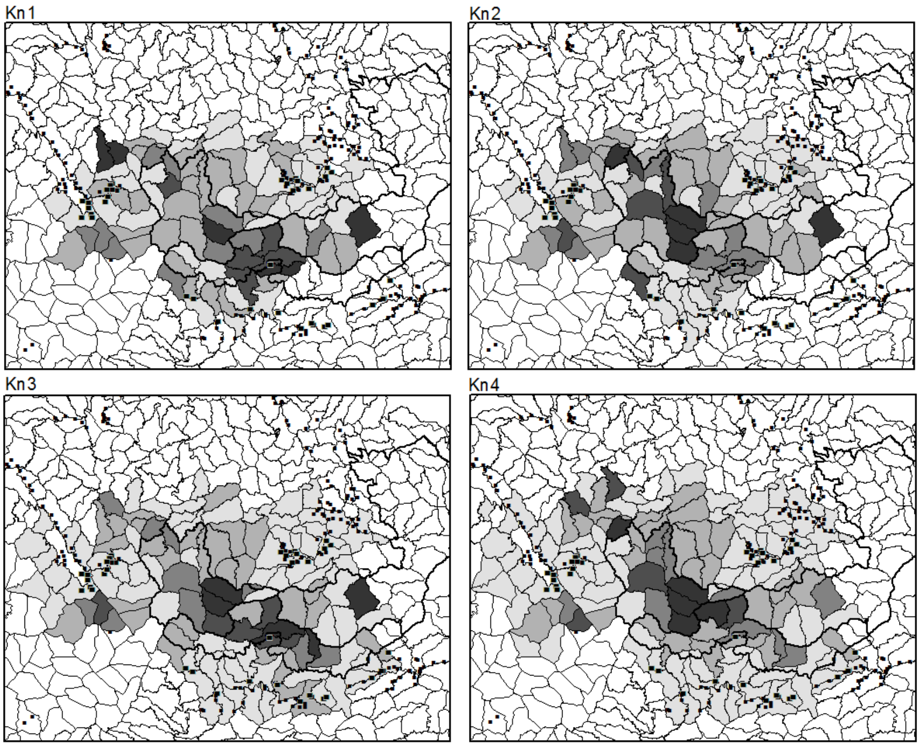


Figure 3: Aggregated local expert opinion on relative abundance of saola in forest units based on PageRank output and a naïve aggregator (mean of scores).



For description of algorithms see text. Aggregated data are densities of beans placed within buffered forest units on community maps (density score); the least conservative estimate of village knowledge areas is used (Kn4).

Figure 4: Mapped rank aggregation results for Saola (subjective rankings) for 4 alternative knowledge estimates



For detailed description of knowledge criteria, see Appendix S1. Kn1 is most conservative. Aggregated data are subjective scores derived from beans placed on community maps by expert groups.