



**HAL**  
open science

## Expert-based ecosystem services capacity matrices: Dealing with scoring variability

Carole Sylvie Campagne, Philip Roche, Frédéric Gosselin, Leita Tschanz,  
Thierry Tatoni

### ► To cite this version:

Carole Sylvie Campagne, Philip Roche, Frédéric Gosselin, Leita Tschanz, Thierry Tatoni. Expert-based ecosystem services capacity matrices: Dealing with scoring variability. *Ecological Indicators*, 2017, 79, pp.63-72. 10.1016/j.ecolind.2017.03.043 . hal-01681625

**HAL Id: hal-01681625**

**<https://hal.science/hal-01681625v1>**

Submitted on 11 May 2018

**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.

# Expert-based ecosystem services capacity matrices: Dealing with scoring variability

Carole Sylvie Campagne<sup>a,b,\*</sup>, Philip Roche<sup>a</sup>, Frédéric Gosselin<sup>c</sup>, Léïta Tschanz<sup>b</sup>, Thierry Tatoni<sup>b</sup>

<sup>a</sup> Irstea, UR RECOVER, 3275 route de Cézanne, CS 40061, 13182 Aix-en-Provence Cedex 5, France

<sup>b</sup> Institut Méditerranéen de Biodiversité et d'Ecologie marine et continentale (IMBE), UMR 7263, Aix Marseille Université, CNRS, IRD, Avignon Université, FST St-Jérôme, 13397 Marseille, France

<sup>c</sup> Irstea, UR EFNO, Domaine des Barres, 45290 Nogent-Sur-Vernisson, France

---

## A B S T R A C T

Capacity matrices are widely used for assessment of ecosystems services, especially when based on participatory approaches. A capacity matrix is basically a look-up table that links land cover types to ecosystem services potentially provided. The method introduced by Burkhard et al. in 2009 has since been developed and applied in an array of case studies. Here we address some of the criticisms on the use of capacity matrices such as expert panel size, expert confidence, and scoring variability.

Based on three case-study capacity matrices derived from expert participatory scoring, we used three different approaches to estimate the score means and standard errors: usual statistics, bootstrapping, and Bayesian models. Based on a resampling of the three capacity matrices, we show that central score stabilizes very quickly but that intersample variability shrinks after 10–15 experts while standard error of the scores continues to decrease as sample size increases. Compared to usual statistics, bootstrapping methods only reduce the estimated standard errors for small samples. The use of confidence scores expressed by experts and associated with their scores on ecosystem services does not change the mean scores but slightly increases the standard errors associated with the scores on ecosystem services. Here, computations considering the confidence scores marginally changed the final scores. Nevertheless, many participants felt it important to have a confidence score in the capacity matrix to let them express uncertainties on their own knowledge. This means that confidence scores could be considered as supplementary materials in a participatory approach but should not necessarily be used to compute final scores.

We compared usual statistics, bootstrapping and Bayesian models to estimate central scores and standard errors for a capacity matrix based on a panel of 30 experts, and found that the three methods give very similar results. This was interpreted as a consequence of having a panel size that counted twice the minimal number of experts needed. Bayesian models provided the lowest standard errors, whereas bootstrapping with confidence scores provided the largest standard errors.

These conclusions prompt us to advocate when the panel size is small (less than 10 experts), to use bootstrapping to estimate final scores and their variability. If more than 15 experts are involved, the usual statistics are appropriate. Bayesian models are more complex to implement but can also provide more informative outputs to help analyze results.

---

### Keywords:

Ecosystem services  
Capacity matrix  
Confidence index  
Uncertainty  
Expert panel size  
Participatory assessment

---

## 1. Introduction

Ecosystem service (ES) is a popular and widely recognised concept (Burkhard et al., 2012; Fisher et al., 2009), and the term ‘ecosystem services’ has translated from scientific studies into the mainstream vocabulary of stakeholders and experts (Jacobs et al., 2014). Increasing demand from policymakers like the European Commission has prompted the development of an array of ES

---

*Abbreviations:* ES, ecosystem service; ET, ecosystem types; SD, standard deviation; SE, standard errors; CP, cut point; RNP, Regional Natural Park; RNPBP, “Les Baronnies Provençales” RNP; RNP-SE, “Scarpe-Escaut” RNP; RNP-SEwet, “Scarpe-Escaut RNP wetlands only”.

\* Corresponding author at: Irstea, UR RECOVER, 3275 route de Cézanne, CS 40061, 13182 Aix-en-Provence Cedex 5, France.

*E-mail addresses:* [sylvie.campagne@irstea.fr](mailto:sylvie.campagne@irstea.fr), [sylviecampagne@gmail.com](mailto:sylviecampagne@gmail.com) (C.S. Campagne), [philip.roche@irstea.fr](mailto:philip.roche@irstea.fr) (P. Roche), [frederic.gosselin@irstea.fr](mailto:frederic.gosselin@irstea.fr) (F. Gosselin), [ltschanz@imbe.fr](mailto:ltschanz@imbe.fr) (L. Tschanz), [t.tatoni@imbe.fr](mailto:t.tatoni@imbe.fr) (T. Tatoni).

assessment and mapping methods (Martnez-Harms and Balvanera, 2012; Willemen et al., 2015).

One such method that is gaining ground is the “capacity matrix approach”, which was even touted as “the most popular ES assessment technique today” (Jacobs et al., 2014). The capacity matrix is basically a look-up table that links land cover types to ecosystem services potentially provided (Burkhard et al., 2009).

Since the “matrix” first introduced by Burkhard et al. in 2009, the method has been developed and applied in an array of case studies (e.g. Hermann et al., 2013; Kroll et al., 2012; Stoll et al., 2014; Vihervaara et al., 2010; etc.). Based on experts' knowledge, it gives a quick assessment of ES potentially provided in an area (Vihervaara et al., 2012; Stoll et al., 2014). The ES concept makes the matrix method mobilize the ES concept in way that makes it easy for stakeholders to understand and appropriate. It is a pedagogical tool that has proven its utility by targeting priorities and highlighting management hotspots. The approach can be applied at different scales (e.g. Stoll et al., 2014 or Hermann et al., 2013). A lack of quantitative data and their spatial heterogeneity raises issues that can be bypassed by asking experts to estimate scores. Depending on the concertation process applied, data can be developed by consensus among the different experts of a territory. The method is also flexible enough to integrate all kinds of data—from models or measurements alike (Burkhard et al., 2014).

Some researchers have started to study the limits of this method, like Jacobs et al. (2014) who point out its poor methodological transparency, lack of reproducibility and lack of appropriate factoring on uncertainty. Hou et al. (2012) also discuss the uncertainties related to the matrix method. The uncertainty of experts judgments is often cited as a limit, but few have analyzed it or integrated it in ES studies (Seppelt et al., 2011; Hou et al., 2012; Vihervaara et al., 2012). Based on the combination of experts' judgments, the scoring in the capacity matrix may carry two sources of uncertainties:

- Variability among experts: variability of the expertise within the chosen experts and of more general knowledge (professional or personal knowledge depending on their experiences) (Hou et al., 2012).
- Variability of each expert: the confidence the expert has in his/her own scores (Jacobs et al., 2014).

The objective of our study is twofold. First, we aim to integrate the different variabilities in the final score of the capacity matrix and present and compare different approaches to computing them. Second, we aim to identify a minimal size of the expert pool needed to obtain a reliable estimate of the mean of the scores and a small SE of this mean.

In order to meet these objectives, we begin by defining our capacity matrices, the experts, and how the scoring was done. The confidence score we added on the capacity matrix is also detailed. In the second section, we present three sets of approaches to the final scores on a matrix: raw parametric approaches (mean and weighted mean), bootstrap models, and Bayesian models. For each approach, we have two calculations: means of scores that experts expressed, and means integrating a metric of the expert's confidence on his/her own scores. The final scores in the final matrix that incorporates all experts scores is thus estimated with 6 calculations. Most existing capacity matrices (e.g. Stoll et al., 2014) using expert knowledge are expressed as mean of scores of the expert panel, so we start by presenting the raw parametric approach. The bootstrap model enables to estimate different statistics by assuming an independent sampling from an unknown distribution and to integrate uncertainties. The Bayesian methods that we used are elaborate parametric statistical models that enable to integrate and estimate the different kinds of uncertainties in the statistical anal-

ysis. We restrict ourselves to these three sets of statistical methods. This work is the first comparison of three calculations applied on capacity matrix scores. We are not setting out to identify the best calculation of final score but to highlight the various possibilities and their related advantages and disadvantages. In the third section, we present the results of the three calculations on one matrix and look at the final scores and their variabilities on three capacity matrices with a growing number of experts. Finally, we conclude with recommendations on using the capacity matrix approach.

The data used in this paper came from three ES assessments: ES provided by land-cover types in the ‘Baronnies Provençales’ Regional Natural Park (RNP) (associated scores noted RNP-BP) and two ES assessments in the ‘Scarpe-Escaut’ RNP in northern France—one on ES provided by wetlands (associated scores noted RNP-SEwet) and another on ES provided by all land-cover types (associated scores noted RNP-SEall).

## 2. Data

### 2.1. Study sites

The Baronnies Provençales RNP (<http://www.baronnies-provencales.fr/>) is a sub-mountainous rural area in Southern France located at the crossroad between the Alps and Provence influences. Created in 2015, it is the latest RNP in France, taking the total to 51. The capacity matrix made in 2014 was based on the Park project of 2350 km<sup>2</sup> and 130 municipalities. This nature-preserve territory is recognised nationally for its unique landscapes, rich “terroir”, built heritage (terraces in dry stone, hilltop villages) and agriculture (orchards, olive groves, linden, lavender, thyme, rosemary, and more), as well as its remarkable geology and biodiversity.

The Scarpe-Escaut RNP (<http://www.pnr-scarpe-escaut.fr/en>) in northern France, near the Belgian border, extends over 430 km<sup>2</sup> crossed by the Scarpe and the Escaut rivers with 55 municipalities. It is the oldest of the 51 French RNP. It is also the largest European park, as together with its Belgian neighbour, the Plaines de l'Escaut Natural Park, they form the Hainaut cross-border Nature Park. The Scarpe-Escaut RNP is especially marked by the wet low-land plain around the Scarpe and the Escaut rivers. As a peri-urban area, urban pressure is high (use of space) in a landscape formed by a mosaic of agricultural and natural environments (crops, grasslands, woodlands, marshes, ponds. . .) and urbanized areas. Water is everywhere, and man has been managing it for centuries to develop key activities (drainage, land use, channeling of rivers. . .). For decades, wetlands have been considered a less attractive landscape, and wet meadows have been declining under urban pressure or exploited as profitable sites for agricultural production such as livestock and as landscaped recreational ponds. Perceptions of wetlands today are either bad for certain local stakeholders or nonexistent for the wider community, despite their importance as ES provided to the territory. This perception deficit prompted a study of the ES provided by the wetland types in 2015. After the positive local feedback on the initiative, the method was applied in 2016 to all land cover types.

### 2.2. Data

#### 2.2.1. The capacity matrices

We define ES as goods or services provided by ecosystems that directly or indirectly benefit humans (Millennium Ecosystem Assessment, 2005). For provisioning services and regulating services, the ES list has been based on the European CICES classification (Haines-Young and Potschin, 2013). We considered provisioning services, regulating services and cultural services.

Besides the score value for each ES/ecosystem type combination, the experts were asked to provide an index of confidence in their score for each ES and each ecosystem type for the last two datasets. This confidence index is used to estimate the impact of expert confidence on the capacity score and their variations.

As this paper focuses not on capacity scores to estimate ES capacity per se but on the methodology used to obtain expert-based capacity matrices and evaluate score confidence levels, detailed ES and ecosystem types for the datasets are given as supplementary materials. Our dataset is composed of three sets of expert scoring exercises used to produce three different capacity matrices.

The first set was completed in 2014 and consists of a panel of 23 experts in the RNP-BP. This capacity matrix counts 33 ecosystem types (ET) (15 agriculture habitats, 12 forest habitats and 6 aquatic and urban habitats) related to land cover types used in the regional land cover map derived from the National Forest Inventory (IFNV2), the agricultural typology of the Ministry of Agriculture, Food and Forestry, and the Agency for Services and Payment (ASP; 2010), called the Graphic Parcel Register (2011), as well as the CRIGE-PACA regional land cover classification called Ocsol (2006) (a regional implementation of Corine Land Cover with a higher spatial resolution of 1/50,000) by 22 ES (8 provisioning services, 9 regulating services and 5 cultural services), giving 726 scores (Tschanz et al., 2015). This total dataset is thus 726 ES/ET expert scores  $\times$  23 experts.

The second set was completed in 2015 and consists of a panel of 30 experts in the Scarpe-Escut RNP dealing exclusively with wetland (RNP-SEwet). In this typology, ET is associated with wetland environments based on a local wetlands map. We identified nine ET (2 forest types, 3 agricultural types, 4 semi-natural types) and 17 ES (5 provisioning services, 7 regulating services and 5 cultural services), giving 153 scores. This total dataset is 153 ES/ET expert scores  $\times$  (9 + 17) confidence scores  $\times$  30 experts.

The third set was completed in 2016 and consists of a panel of 17 experts in the Scarpe-Escut RNP as a whole (noted RNP-SEall) intended to be more exhaustive as it integrates all land cover types. This capacity matrix counts 33 ET (6 aquatic habitats, 13 agriculture habitats, 7 forest habitats and 7 anthropized habitats, corresponding to the land cover types in the ARCH map, [www.archnature.fr](http://www.archnature.fr)) by 25 ES (9 provisioning services, 11 regulating services and 5 cultural services), giving 726 final scores. The total dataset is 726 ES/ET expert scores  $\times$  (33 + 25) confidence scores  $\times$  17 experts.

These three matrices were produced in different contexts and with different objectives. While the RNP-SEwet matrix was completed in a context of experts' concertation and outreach education to local citizens, the other two were completed for a more exhaustive spatial assessment, so we did not use identical ET or ES typologies but instead ran local contextual adaptations based on the land cover maps available on each application site.

### 2.2.2. Expert-based participatory method

The capacity matrix aims to produce estimates of the biophysical capacity of the ecosystem to provide ES. We restricted the panel of eligible experts to those having both local and global ecological knowledge in order to take into account all major ecosystem types and all major activities applied on them. We generally tried to follow the recommendations of Jacobs et al. (2014) on forming a relevant sample of experts with specific affinity to their territory. Our definition of 'expert' was a person with extensive knowledge or skills based on research, experience, or occupation in a particular field.

The experts considered included researchers with expertise in ecology and/or ES, project or site managers, technicians working on environmental or ecological fields, and heads of territorial organizations such as the water agency, regional chamber of agriculture, regional professional centre for forest owners, regional environ-

mental science council, regional conservatories of natural areas, national botanical conservatory, local or regional departments of environmental affairs. The departmental federation of hunters and fishers was also represented, along with associations for environmental protection or naturalists.

For each of the ES assessments, the participants were invited to a workshop dedicated to filling out capacity matrix scores. Any participant who was unable to attend the workshop was interviewed individually. During the workshop and the personal interviews, participants were informed on the state-of-the-art in ES, the study, the methods, and the list of ES and ET. We took time to discuss and clarify all the definitions involved in the matrix and the scoring. We proposed to complete the matrix by columns and to give a score by comparing the different ES capacities of each ET. Our experience has proven that this was an effective way to fill in the matrix. The workshops lasted one day, with the morning session dedicated to presenting all definitions and the afternoon left for participants to fill in the matrices and discuss their understanding of the scorings. In both situations (workshop or personal interviews), we let people give their own score independently. The difference between the two approaches lies in the dialogue initiated in the workshop after the scoring, where everyone was given time to voice their chosen score and open a dialogue on any divergences. At the beginning of the workshop, we defined and specified the rules of dialogue: freedom of speech and to hold divergent opinions, and the possibility of constructive criticism.

### 2.2.3. Scoring

The capacities of the different ET to provide an individual ES were quantified following the scale given in Burkhard et al. (2009), i.e. "1 = low relevant capacity, 2 = relevant capacity, 3 = moderate relevant capacity, 4 = high relevant capacity, 5 = very high relevant capacity" and "0 = no relevant capacity". For the RNP-SEwet matrix, "0 = no relevant capacity" was not considered since all the selected ES were provided to some level by all the wetland habitats discussed.

In the rest of the paper, "expert score" refers to the potential ES capacity value that individual experts provided and is noted  $Y_{k,i,o}$ , where  $i$  is the ES,  $k$  is the ET and  $o$  is the experts (the total of all experts involved is noted  $n$ ), and "final score" refers to the average potential capacity value computed from a subset or all expert scores and is noted  $\mu_{k,i,o}$ .

### 2.2.4. The confidence score

In any approach based on experts' knowledge, one of the difficulties is to assess the degree of confidence of each individual expert in their own score. As Jacobs et al. (2014) stressed "confidence reporting is paramount for communication of results and quality comparison". We propose to quantify a score of the confidence the expert has on their scores, where "confidence refers to the degree of confidence in being correct" (Jacobs et al., 2014). Each expert was asked to state their confidence in their knowledge on ET and ES using a confidence score ranging from: "1 = I don't feel comfortable on my score, 2 = I feel fairly comfortable on my score and 3 = I feel comfortable on my score".

Thus, for each expert (noted  $o$ ) of the RNP-SEwet matrix, we had 17 confidence scores for each ES noted  $V_{i,o}^{ES}$  and 9 confidence scores for each ET noted  $V_{k,o}^{ET}$ . We considered these ET and ES confidence scores as margins of a table, and computed a confidence score for each ET and ES by multiplying them:  $V_{k,i,o} = V_{k,o} \times V_{i,o}$ . The confidence score obtained could take values of 1, 2, 3, 4, 6 or 9 that we recoded on a scale of 1–6 (by changing the original "6" to "5" and the original "9" to "6") to facilitate the analysis.

For the bootstrapping and the Bayesian approach incorporating the confidence score, we first estimated the variation of ES  $\times$  ET score associated with each confidence score value based on an

additional direct survey with 10 experts of the RNP-SEwet matrix (allocated within the different ecological specificities/expertises and knowledge) in order to express the uncertainty underlying their confidence scores (cited as levels of uncertainty and noted  $U_{k,i,o}$ ). Through concrete examples (combinations of ES and ET with different confidence values), we asked them to express a range of ES  $\times$  ET scores (0; 0.33; 0.5; 1 or 2 up and down their own score) that they could have considered on a set of specific ET  $\times$  ES combinations with different associated confidence scores. We tried to cover as many of the experts' questions of possible combinations of confidence scores  $V_{k,i,o}$ . From this dataset we computed the SE of ET  $\times$  ES scores associated with each confidence score.

### 3. Statistical analysis

#### 3.1. Usual statistics

Most capacity matrices using expert knowledge are expressed as mean of the score values of the expert panel.

First, a final capacity matrix can be obtained with usual direct statistics through the **mean**, also cited as average score (noted  $\hat{\mu}_{k,i}^1$ ) and its **SD** (noted  $\hat{\sigma}_{k,i}^1$ ).

From the SD, the SE of the mean score is estimated by dividing by  $\sqrt{n}$  calculated as  $\hat{s}\hat{e}_{k,i}^1$  and as follows for the SEs of the mean:

$$\hat{s}\hat{e}_{k,i}^1 = \frac{\hat{\sigma}_{k,i}^1}{\sqrt{n}}$$

We also used the confidence index score (1–6) as a weighting coefficient for the usual statistics. The **weighted mean** (noted  $\hat{\mu}_{k,i,o}^1$ ) can be calculated with different values depending on study, context, etc. Here we used confidence score as weighting factor for the expert scores, so the weighted score is the expert's score multiplied by confidence score, as follows:

$$Y'_{k,i,o} = Y_{k,i,o} \times V_{k,i,o}$$

Accordingly, the weighted mean of the scores is calculated as follows:

$$\hat{\mu}_{k,i}^1 = \frac{\sum^{i,o} Y'_{k,i,o}}{\sum^{i,o} V_{k,i,o}}$$

Then the **SD of the weighted means** (noted  $\hat{\sigma}_{k,i}^3$ ) is calculated as follows:

$$\hat{\sigma}_{k,i}^1 = \sqrt{\frac{\sum^{i,o} \left\{ (Y_{k,i,o} - \mu_i^3)^2 * V_{k,i,o} \right\}}{\sum^{i,o} V_{k,i,o}}}$$

From the SDs, the SEs of the mean of scores is estimated by dividing by  $\sqrt{n}$  and thus calculated as  $\hat{s}\hat{e}_{k,i}^1 = \frac{\hat{\sigma}_{k,i}^1}{\sqrt{n}}$  and as follows for the SEs of the weighted mean:

$$\hat{s}\hat{e}_{k,i}^1 = \frac{\hat{\sigma}_{k,i}^1}{\sqrt{n}}$$

#### 3.2. Bootstrapping statistics

Assuming an independent sampling from an unknown distribution, the bootstrap method enables to estimate different statistics in a better way. The Monte Carlo simulation approach provides approximations of different statistics by independently repeating N times a resampling with replacement from a given set of n

observations. Each resampling provides an independent realization from the initial dataset from which N realizations of the statistics are produced, providing a distribution of the statistics considered. The bootstrapped Monte Carlo simulations were done using R3.3.1, which features the "boot" package but we preferred, here, to write our own script for greater flexibility.

In our case, n, the number of observations, is equal to the number of experts. We estimated two statistics:  $\hat{\mu}_{k,i}^2$  the mean of scores and  $\hat{s}\hat{e}_{k,i}^2 = \frac{\hat{\sigma}_{k,i}^2}{\sqrt{n}}$  the SE of the mean of scores, with N the number of repetitions being equal to 500. Number of repetitions is not usually crucial once it passes 100 (Efron, 1981).

With the bootstrapping statistics, the variation in ES  $\times$  ET score associated with each confidence value was used in a bootstrap Monte Carlo approach to estimate  $\hat{\mu}_{k,i}^2$  and its SE (noted  $\hat{s}\hat{e}_{k,i}^2$ ) taking into account the confidence score associated to  $Y_{k,i}$ . We added in the bootstrap script a random error  $\varepsilon_{k,i,o}$  drawn from a normal distribution of errors associated with each level of the confidence scores, which gives  $\varepsilon_{k,i,o} \in \mathcal{N}(0, \hat{\sigma}_{k,i,o,c})$  with c being the value of the confidence score (0, ..., 6).

#### 3.3. Bayesian approach

For the Bayesian analyses, we used two versions of a heteroscedastic ordered probit model (Litchfield et al., 2012; see Supplementary Materials) for  $Y_{k,i,o}$ . Indeed, the probabilities associated with each observation  $Y_{k,i,o}$  were calculated as if the observations stemmed from a latent random variable  $\omega_{k,i,o}$  following a normal distribution according to the following formulas:

$$P(Y_{k,i,o} = 1) = P(\omega_{k,i,o} < CP_{1,i})$$

$$P(Y_{k,i,o} = 2) = P(CP_{1,i} < \omega_{k,i,o} < CP_{2,i})$$

$$P(Y_{k,i,o} = 3) = P(CP_{2,i} < \omega_{k,i,o} < CP_{3,i})$$

$$P(Y_{k,i,o} = 4) = P(CP_{3,i} < \omega_{k,i,o} < CP_{4,i})$$

$$P(Y_{k,i,o} = 5) = P(\omega_{k,i,o} > CP_{4,i})$$

where the cut-points  $CP_{j,i}$  were (i) fixed to 0 and 1 for  $CP_{1,i}$  and  $CP_{4,i}$ ; and (ii) estimated between 0 and 1 within the model for  $CP_{2,i}$  and  $CP_{3,i}$ , via the following formulas:  $CP_{2,i} = 1 / [1 + \exp(-CP_{0,2,i})]$  and  $CP_{3,i} = 1 / [1 + \exp(-CP_{0,2,i} - \exp(-CP_{0,3,i}))]$  with both  $CP_{0,2,i}$  and  $CP_{0,3,i}$  having a Gaussian prior distribution of  $\mathcal{N}(0, 5)$ .  $CP_{j,i}$  cut-points were considered to vary from one ES to another because we conjectured that the span between score values would vary from one ES to another. We checked the variations of the estimators of  $CP_{j,i}$  with i and found that, for some ES at least, the cut-point values were very significantly different from other values.

The two parameters describing the distribution of  $\omega_{k,i,o}$  were the location or mean parameter  $\hat{\mu}_{k,i,o}^3$  and the dispersion parameter  $\hat{s}\hat{e}_{k,i,o}^3$  corresponding to the SE of the latent normal distribution. After comparing four models (see Supplementary Materials), we retained two different models that contained different parametrizations of the heteroscedasticity (i.e. in parameter  $\hat{s}\hat{e}_{k,i,o}^3$ ). In the first model (model B1), no other information than basic score  $Y_{k,i,o}$  was provided to the model, whereas in the second model (model B2), both the confidence scores  $V_{k,i,o}$  and the levels of uncertainty  $U_{k,i,o}$  were taken into account. The formula for the

mean parameter  $\mu_{k,i,o}$  of the latent variable  $\omega_{k,i,o}$  was the same for all models:

$$\hat{\mu}_{k,i,o}^3 = \alpha_{k,i} + \gamma_o$$

whereas the formulas for  $\hat{s}e_{k,i,o}^3$  varied across the two models:

$$\hat{s}e_{k,i,o}^3 = \exp(\sigma_{k,i}) \quad (\text{Bayesian model 1 : B1})$$

$$\text{or : } \hat{s}e_{k,i,o}^3 = \exp(\sigma_{k,i}) \sigma'_{k,i,o} \quad (\text{Bayesian model 2 : B2})$$

where:

(i)  $\sigma_{k,i}$  is a random effect used to parametrize SD level; it followed a Gaussian prior distribution of  $\mathcal{N}(\mu_\sigma, \sigma_\sigma)$  where  $\mu_\sigma$  had a Gaussian prior centered on 0 with 5 as SD and  $\sigma_\sigma$  had a relatively non-informative prior uniform distribution between 0 and 100;

(ii)  $\alpha_{k,i}$  is the mean random effect estimated for the latent variable for each ES  $i$  and ET  $k$ , with a non-informative Gaussian prior distribution of  $\mathcal{N}(\mu_\alpha, \sigma_\alpha)$  where  $\mu_\alpha$  and  $\sigma_\alpha$  had the same priors as  $\mu_\sigma$  and  $\sigma_\sigma$ ;

(iii)  $\gamma_o$  is a random observer  $o$  effect on the mean, chosen here to be constant across ETs and ESs, and drawn from a common Gaussian distribution centered on 0 and with SD  $\sigma_{obs}$ :

$$(\gamma_o)_o \sim \mathcal{N}(0, \sigma_{obs})$$

where  $\sigma_{obs}$  has the same prior as the  $\sigma_{k,i}$ s.

(iv) confidence score  $V_{k,i,o}$  is the product of the confidence scores of each ET and ES ( $V_{k,o}^{ET} V_{i,o}^{ES}$ ), restricted to its order (from 1 to 6);

(v) the confidence multipliers  $(\sigma'_i)_{1 \leq i \leq 6}$  in model B2 were constructed with no constraint on their order, so that the vector summed to one (to prevent convergence issues with the  $\sigma_{k,i}$ s). This was done by first setting  $\sigma'_{i=1} = 1/S$ ,  $\sigma'_i = \sigma''_i/S$  for  $2 \leq i \leq 6$  with  $S = 1 + \sum_{i=2}^6 \sigma''_i$  and the  $\sigma''_i$  chosen from a uniform prior distribution between 0 and 100. We first tried a version that constrained the order of these parameters to decrease with  $i$ , but we met strong convergence issues that forced us to adopt this unconstrained structure.

Furthermore, we also integrated in model B2 the data related to the calibration of confidence scores  $U_{k,i,o}$  for specific scores  $Y_{k,i,o}$  corresponding to cases  $(k,i)$  with controlled levels of confidence  $V_{k,o}^{ET}$  and  $V_{i,o}^{ES}$ .  $U_{k,i,o}$  could take predefined values (0, 0.33, 0.5, 1, and 2) and was assumed to represent the standard variation of  $Y_{k,i,o}$ . We introduced the information provided by  $U_{k,i,o}$  through a discrete distribution on the 5 possible values proportional to Gaussian distribution  $\mathcal{N}[\sigma'(v_{k,i,o} + \rho'), \sigma_U]$ , where:

(vi)  $v_{k,i,o}$  is the estimated SD of  $Y_{k,i,o}$  as calculated from the above equations for  $\{P(Y_{k,i,o} = y)\}_{1 \leq y \leq 5}$ ;

(vii)  $\sigma'$  is a positive mean SD multiplier chosen to multiply the underlying SD from a relatively non-informative prior uniform distribution between 0.1 and 5;

(viii)  $\rho'$  is a negative number with a relatively informative uniform prior between  $-2.0$  and  $0.0$  used to scale  $v_{k,i,o}$  to make it fit better to the scale of  $U_{k,i,o}$ ;

(ix)  $\sigma_U$  is a mean SD with an informative prior distribution between 0 and 1.

We mostly used informative priors here because less-informative priors gave much noisier estimators, which was contrary to expectations (as we did not expect the mean of  $\sigma_U$  to be above 10.0 given the variation of the  $U$ s, whereas we observed such values with non-informative priors).

Model B2 therefore incorporated variations of precision and mean between ES and ET, as well as potential systematic biases between observers. This model also incorporated additional information, i.e. the confidence score provided by experts together with

their raw score, as well as a calibration of this confidence in an independent survey. In the rest of the paper, for Model B2,  $\hat{\mu}_{k,i,o}^3$  is mean parameter and  $\hat{s}e_{k,i,o}^3$  is the SE.

These two models plus three additional models (see Supplementary Materials) were fitted with Stan from R 3.3.0 through the rstan library (2.9.0-3) (Stan Development Team, 2015). We used 3 Markov Chains with a total number of iterations of 100,000 each, a burn-in length of 20,000 and a thinning parameter of 20. Mixing was correct and had 12,000 output values for each parameter, with a number of effective values (CODA R library) above 11,000 for Model 1 and above 3500 for Model 2 (for the worst parameters).

Based on the posterior samples of parameters, we calculated the posterior mean and SEs of the mean across observers of several ES (on the original scale of  $Y$ ) provided by the different land-cover types studied.

### 3.4. Resampling

In order to estimate the impact of expert panel size on the central and dispersion parameters, we used a bootstrap subsampling method. We simulated different expert panel sizes by subsampling each dataset. We fixed the size of the panel “n” between 2 and the actual number of experts (30 for the RNP-SEwet matrix, 23 for the RNP-BP matrix and 17 for the RNP-SEall matrix), and for each size we generated samples with replacement from the actual data. At each iteration, several statistical parameters were computed. This approach enables us to evaluate the variation of the statistics given smaller-than-actual samples. We used it to try to identify a minimal size of the expert pool needed to obtain a reliable estimate of the final scores and a small SE of this mean.

## 4. Results

### 4.1. Central values, capacity matrix scores

Based on the RNP-SEwet dataset that we analyzed in more detail than the other datasets, we observed that the different statistical methods used to compute central parameters (mean) provided very similar results. Table 1, which presents mean and SD of all expert scores (153 scores  $\times$  17 expert scores; first column) and of the mean scores of our approaches (means of the 153 final scores with the six different calculations), shows the proximity of their mean results clearly visible at around 3.3 with similar values to the third decimal place for all experts scores, Mean of averages scores, Mean of bootstrap scores and Mean Bayesian scores. The mean of all expert scores (calculated with normal theory statistics) is 3.363, and SD at 1.368 (Table 1) with median at 4. For the mean of average scores (mean of the 30 expert scores for each of the 153 ET  $\times$  ES), the SD is 0.967.

The SD on all means' scores has the same proximity of results as the means except for all experts scores which show higher variation.

Specific ET/ES scores (see Table S.2 in Supplementary Materials) show for example the mean  $\mu_{i,o}^1$  which ranks from 1.40 for ES10 “Contribution to freshwater” in ET8 “Crops” to 4.93 for ES17 “Knowledge, scientific and educational” in ET6 “Marsh, bogs and reed beds”. Examination of means between approaches and specific ES and ET finds no distinction between means of the 6 calculations.

### 4.2. Scores variability

We examined means of SEs of the scores (based on the 153 ES  $\times$  ET combinations) estimated from each method as presented in Fig. 1 and found that variability in mean scores was lowest for weighted scores (0.168) and highest for the Bootstrap scores with

**Table 1**  
 Statistics (mean and Standard Deviation SD) using 3 different methods: Usual, Bootstrap and Bayesian. Usual refers to the use of standard parametric statistics. Bootstrap is based on resampling and Monte-Carlo methods. Bayesian is based on heteroscedastic ordered probit models.

		Usual		Bootstrap		Bayesian	
	On all expert scores $Y_{k,i,o}$	Mean of average scores $\hat{\mu}_{k,i}^1$	Mean of weighted scores $\hat{\mu}_{k,i}^1$	Mean of Bootstrap scores $\hat{\mu}_{k,i}^2$	Mean of Bootstrap scores with confidence scores $\hat{\mu}_{k,i}^2$	Mean Bayesian scores $\hat{\mu}_{k,i}^3$	Mean Bayesian scores with confidence scores $\hat{\mu}_{k,i}^3$
Mean	3.363	3.363	3.371	3.363	3.298	3.363	3.365
SD	1.368	0.967	1.006	0.967	0.903	0.932	0.935

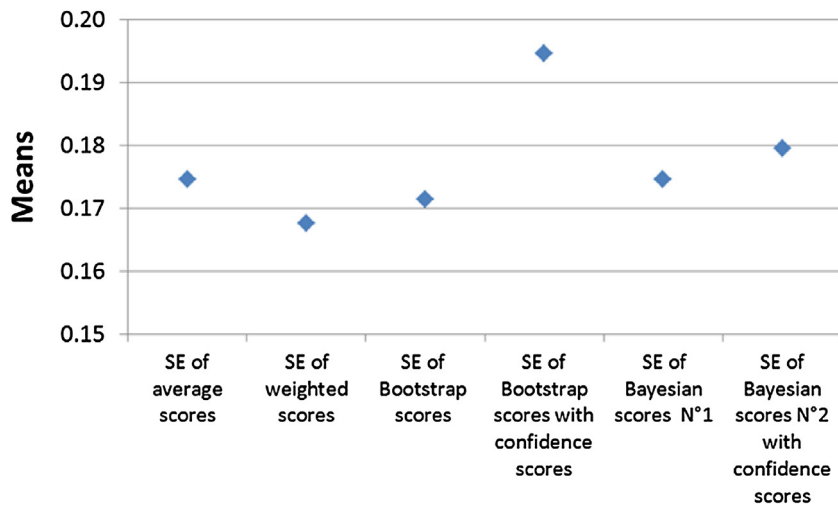


Fig. 1. Means of standard errors (SE) of the different methods.

confidence score (0.195). The three approaches without confidence score shared similar variability in mean scores, at around 0.173.

The bootstrap SE with integration of expert confidence index provided the highest values (14% higher than without confidence index), as expected by including another source of variability (i.e. the individual expert range of values based on confidence on the raw score). This does not mean that this bootstrap is less precise but that, by construct, it integrates more sources of variability than the other method. The Bayesian model with confidence scores also incorporates intra-expert variability and provided higher variability in mean scores than without confidence score (0.1794 vs 0.1746).

As examining the whole capacity matrix in detail is such a cumbersome task, we propose to focus on the SE of ES9 “Domestic food and their outputs” for each ET. The SE of Bootstrap scores with confidence scores has the highest values of 8–9 of the SE presented in Fig. 1. However, there are no clearly visible patterns to emerge between the other five approaches. The results presented in Fig. 2 for ET8 (“Crops”) are more dispersed than for other ET. For this specific ET, the SE of Bayesian models are lower and the SE of weighted scores are higher than with other approaches. When we look closer at the ET8 × ES9 scores (available in Supplementary Materials Table S.2), the mean of average score is 4.73 and the mean of confidence index is 4.37.

#### 4.3. Effect of expert confidence

During the different ES assessments, several experts accepted to fill the matrix only when offered a confidence score giving the opportunity to express their self-confidence on their scores. In the RNP-SEwet matrix, mean confidence score per expert shows a trend towards high confidence scores for 10 experts out of 30 (median at 3) while two experts tended to have low confidence score with a

median at 1. The table S.3 (Supplementary Materials) presents  $V_{k,i,o}$  in the table and the means of each  $V_{i,o}^{ES}$  and  $V_{k,o}^{ET}$  in margin.

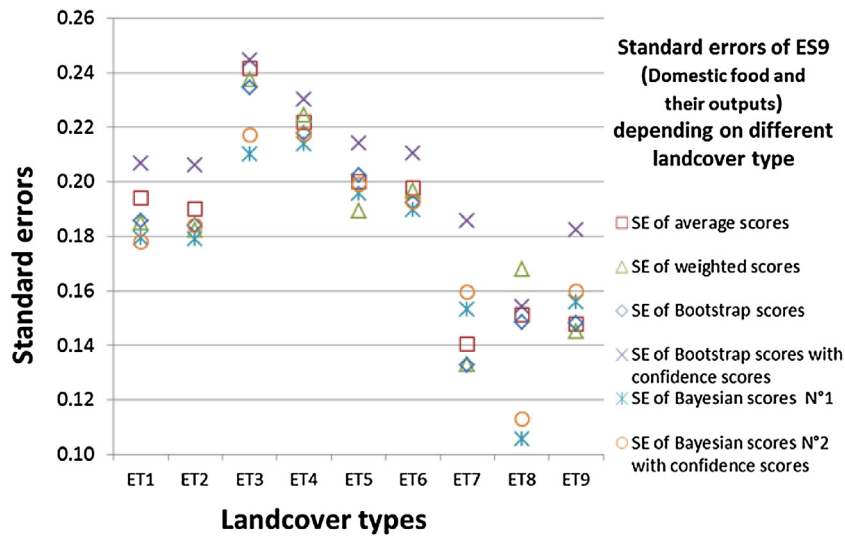
The mean confidence score of all experts for all ET and ES (noted  $V_{k,o}^{ET}$  and  $V_{i,o}^{ES}$ , respectively, and rating between 1 and 3) is 2.23 with each confidence score (1, 2 and 3) represented at 20%, 37% and 43%, respectively. Mean score of all ES was 2.24 and mean score of all ET was 2.21. A more detailed analysis shows there is no uniform profile for ET or ES except for cultural services where each confidence score (1, 2 and 3) represented at 12%, 28% and 60%, respectively.

The results from additional direct survey presented in Fig. 3 show the range and means of variation in expert scores (the levels of uncertainty  $U_{k,i,o}$ ) related to each confidence score (with  $V_{k,i,o} = V_{k,o} \times V_{i,o}$  recoded on the 1-to-6 scale used in the different models).

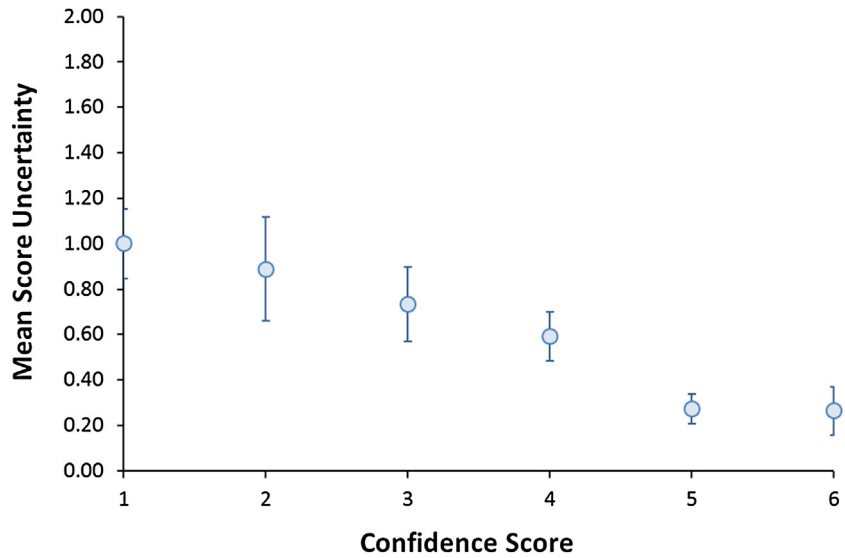
Note the gradual decrease in score variation as confidence score increases. The variability of the ES score errors associated with confidence scores was higher for lower confidence values (1–2–3) than for higher confidence values (4–5–6). The high values of  $V_{k,i,o}$  are related to high confidence of the experts on their scores and thus to lower variation in their final scores. There was no significant difference between scores errors associated with confidence score 5 and 6.

#### 4.4. Effect of expert panel size

Expert panel size had an effect on two aspects of variability. First, it had an effect on the stability of the final score obtained from different sets of resampled same-size tables: SD of resampled final scores drops sharply from 2 to 10 (Fig. 4) but then remains relatively stable before flattening after a size of 15 for the three matrices. This means that above an expert size panel of 15 experts, the variation of final score obtained from different panel compositions is stable.



**Fig. 2.** Standard errors (SE) of ES9 “Domestic food and their outputs” for different land-cover types ET1: Wet forest, ET2: Broadleaf monospecific afforestation in wetlands, ET3: Wet meadow, ET4: Ponds and pools, ET5: Mining subsidence ponds and quarry waters, ET6: Marsh, bogs and reedbed, ET7: Wetland fallows, ET8: Crops, ET9: Green space.



**Fig. 3.** Mean and standard errors of the variation in expert scores ( $U_{k,i,o}$ ) related to each confidence score ( $V_{k,i,o}$ ).

The total reduction of SD is approximately 0.2 (Fig. 4), i.e. about 16% lower when we have 15 experts involved compared to 2.

Second, we tested the effect of expert panel size on the SEs of final mean score (the precision of estimates). As expected, the SEs decrease with the sample size. The bootstrap estimates show lower values than usual statistics when the sample size is below 10–15 experts (Fig. 5). All three datasets show similar trends. However, even with an expert panel size at 30, we still do not reach stabilization.

## 5. Discussion

### 5.1. Final scores and variability

The mean is the most widely used statistic for computing a final score in the capacity matrix approach, but some authors advocate using median scores instead (Kopperoinen et al., 2014). Different kind of weights can also be applied to obtain the final score. Depending on the approach, weighting can be applied during the

extrapolation at larger geographic scale, as in Hermann et al. (2013) who chose to weight scores by the area of base unit. Koschke et al. (2012) elected to apply “explicit weights as the importance of the various ecosystem services might differ with respect to the context, the included stakeholders, and the investigated region”. These weights are linked to specific methodological choices. Here, we did not apply ES or ET dependent weightings as we wanted to equally consider every ES and ET. We used a weight based on the confidence scores, but the upshot is that a participant with more confidence than another gives a better or a more realistic estimate of the habitat’s capacity to provide the ES.

With 30 experts involved in the RNP-SEwet dataset, the different statistical methods used to compute parameters provided very similar results. The results show only marginal differences in final scores when we integrate confidence score or the uncertainty underlying confidence scores in general observation of the data (mean of means) or at an ES level. Indeed, with 30 experts involved, the final score did not change much after taking the confidence score into account. The variabilities of final scores are also



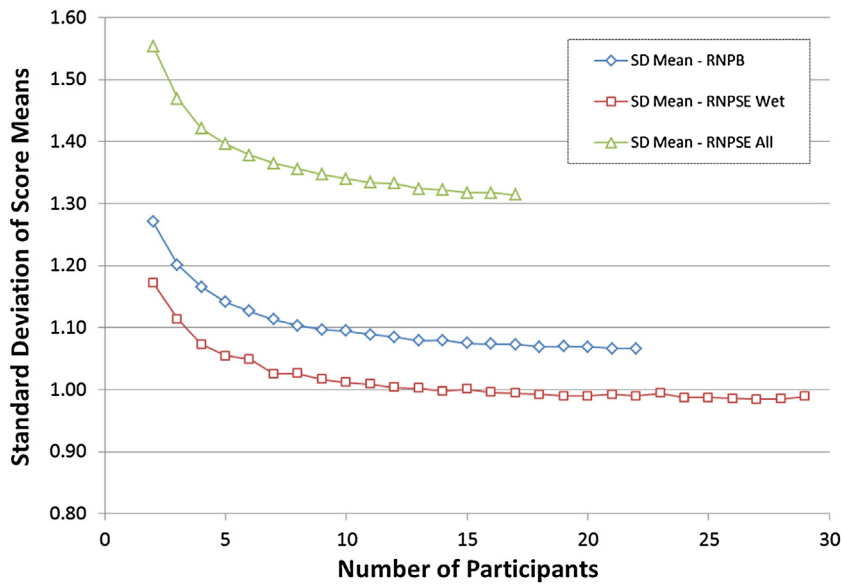


Fig. 4. Effect of expert panel size (number of participants) on the variation (SD) of means of average scores.

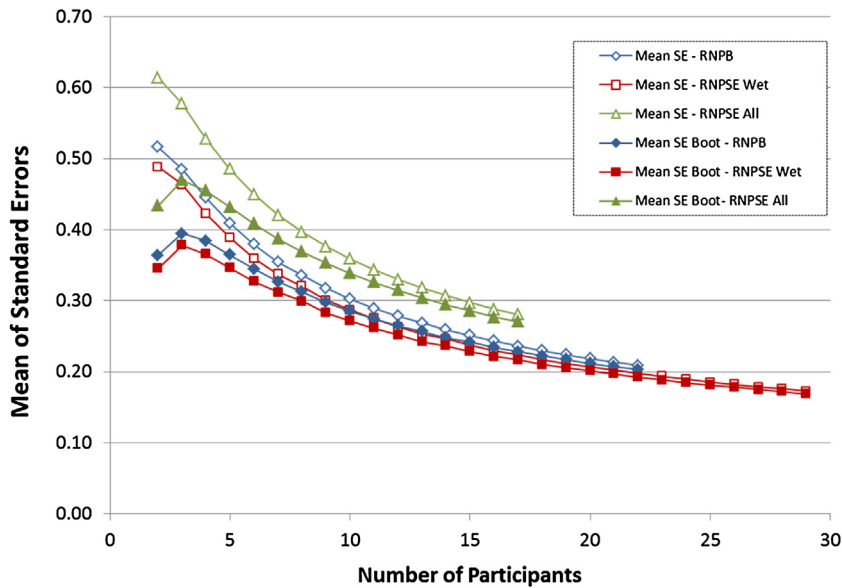


Fig. 5. Effect of expert panel size (number of participants) on the mean of SEs of ES  $\times$  ET scores (average of the whole capacity matrix) using usual statistics (SE) and bootstrapped statistics (SE Boot).

quite similar, except for bootstrap scores with confidence scores which had higher variability. The increase of variability is coherent with the addition of inner confidence or doubt of each expert in terms of variability of scores.

Regarding the Bayesian approach, the variability is similar for the two models considered, which may be due to the fact that both integrate an observer effect. The observer effect may integrate some systematic bias of each expert when providing their scores. Nevertheless, the final results of the two Bayesian models are close to those of the other methods. However, the Bayesian framework has the potential to deliver substantially more information than just central values and errors—it can clearly integrate all the information and provide estimates on these other pieces of information. To illustrate, the Bayesian method can reach an estimate of the variation of precision of the latent process as a function of confidence score. If we had observed strong systematic biases between observers—which was not the case here, as the mean estimate of

$\sigma_{obs}$  was as low as 0.10—Bayesian methods would have made it possible to provide estimates of SEs that would have removed these biases, which other methods are unable to do.

Given the amount of information contained in our capacity matrices, it was not possible to separately detail all the ES and ET combinations, so we have detailed one example combination here, and provided all the tables as Supplementary Materials for interested readers. For the specific ES  $\times$  ET combinations looked at here, the results of ET8  $\times$  ES1 scores can be explained by the nature of this score. “Crops” have a far higher potential capacity to provide “Domestic food and their outputs” than other ES types. The variability of its score is low due to the logical general agreement on this potential capacity. The Bayesian models reflect this general agreement “better” as the lowest SE. Only the weighted score has a surprising result compared to other methods, which may be explained by the fact that the confidence score is less a reflection of the variability of each participant than a reflection of the uncer-

tainty underlying confidence scores expressed through the results of the additional survey.

### 5.2. Importance of experts' confidence in their scores

Jacobs et al. (2014) recommended following the IPCC (2014) or Millennium Ecosystem Assessment (2005) methods with participant confidence reported with five levels of confidence or assessing the reliability of outcomes in a similar standardized manner. In our case, asking for such specific confidence scores was too much of a burden given the time and thought already invested to fill out the capacity matrix. Additional information like a confidence score has to be suitable, understandable, and on a different scale to the participants' score (5 levels).

Many participants expressed the importance of having a confidence score in the capacity matrix to let them moderate their self-reliability on their own knowledge. Analysis of the confidence score showed that 80% of the participants felt "comfortable or moderately comfortable on their scores". It is important to add a confidence score in a participatory process, not just as supplementary materials but also to be able to recalibrate it for integration into the analysis. We consequently chose to use three levels of confidence for ET and three levels of confidence for ES, which when combined provide a 6-level confidence score for each ET and ES. This approach proved to be flexible, workable, and easy for all participants to understand.

The additional survey supports the hypothesis of a correlation between confidence score and SD of the final score, and shows the negative relationship between the confidence score and the mean of score errors. Our confidence index can thus be used as a proxy of the errors in each expert's estimate scores. A limitation here is that we have not explored all the possibilities of the Bayesian models. We were able to analyze the noise around expert scores with other approaches but we chose not to report the results in this paper.

### 5.3. How large should an expert panel be?

Without distinction between the three matrices, the variability of the final scores is stable after 15 experts involved and the SE in the final scores decreased with increasing expert panel size. The bootstrap approach also increases the precision by reducing the variation when panel sizes are small.

These results serve to explain our low difference between means of the three approaches and with or without taking into account the confidence score. With 30 experts involved, we reach a stable mean that may thus iron out disparities.

This result is interesting in practice as it is the first time that a recommendation regarding panel size is provided for participatory ES assessment. Having a large panel is good, but smaller panels can still provide reliable estimates. It should be stressed that the case configuration here is panel composition based on experts and scientists with expertise in ecosystems and/or ES in the studied region. Other studies are needed to address the question of panels or surveys based on non-specialists.

The stability of the scores expresses a consensus of the panel on the potential capacity of a given ET to provide a given set of ES. There is still a need to further explore the link between this consensus score and actual potential capacity.

## 6. Conclusion

The two parts of this paper bring complementary analysis to help identify what methods to use or how many experts are needed. These questions also depend on the level of error accepted and of course the number of experts that can viably be involved in the participatory approach.

For a restricted number of participants, it is better to use bootstrap estimate parameters in order to increase the stability of the final score and its variability. If more than 15 experts are involved, the usual statistics is sufficient. Indeed, with 30 experts involved in our matrix, final scores showed little distinction between the three approaches, as we had reached a stable mean.

Bayesian methods are more demanding but they also give more complete results. Bayesian methods effectively require a degree of expertise in how to write, fit and interpret Bayesian models. However, they do enable to integrate different kinds of information (here, the expert scores, confidence scores and their uncertainties were highlighted) in an integrated statistical modelling framework (Gimenez et al., 2014) as well as to estimate all the parameters associated with these variables. In particular, Bayesian models hold great potential if the aim is to gain a more inferential point of view. For the sake of consistence and concision, this paper did not explore all the possibilities of the Bayesian models.

For participatory processes, we recommend adding a confidence score, on the caveat that incorporation of a confidence score or the uncertainty it carries to the final score in the capacity matrix depends on the SE accepted and the method used. We found here that considering confidence score in final score induced only marginal differences.

## Acknowledgements

We thank Jean-Michel Salles for his input and comments on the paper. This study was supported by the Natural Regional Park of Scarpe-Escaut. (PNRSE), and we thank Gerald Duhayon and Tangui Lefort from the PNRSE for their support. The RNP-BP matrix is part of a 2017 PhD project by L. Tschanz titled "Towards a socio-ecological observation for a global approach to biodiversity challenges of in regional planning" at the IMBE, Aix-Marseille University. This work was funded by Irstea, the PNRSE, the PNRBP, and the Institut INSPIRE.

## References

- Burkhard, B., Kroll, F., Müller, F., 2009. Landscapes' capacities to provide ecosystem services—a concept for land-cover based assessments. *Landsc. Online*, 1–22, <http://dx.doi.org/10.3097/LO.200915>.
- Burkhard, B., Kroll, F., Nedkov, S., Müller, F., 2012. Mapping ecosystem service supply, demand and budgets. *Ecol. Indic.* 21, 17–29, <http://dx.doi.org/10.1016/j.ecolind.2011.06.019>.
- Burkhard, B., Kandziora, M., Hou, Y., Müller, F., 2014. Ecosystem service potentials, flows and demands—concepts for spatial localisation, indication and quantification. *Landsc. Online* 32, 1–32, <http://dx.doi.org/10.3097/LO.201434>.
- Efron, B., 1981. Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. *Biometrika* 68, 589–599.
- Fisher, B., Turner, R.K., Morling, P., 2009. Defining and classifying ecosystem services for decision making. *Ecol. Econ.* 68, 643–653, <http://dx.doi.org/10.1016/j.ecolecon.2008.09.014>.
- Gimenez, O., Buckland, S.T., Morgan, B.J., Bez, N., Bertrand, S., Choquet, R., Dray, S., Etienne, M.-P., Fewster, R., Gosselin, F., Mérigot, B., Monestiez, P., Morales, J., Mortier, F., 2014. Statistical ecology comes of age. *Biol. Lett.* 10, 20140698.
- Haines-Young, R., Potschin, M., 2013. CICES V4.3 – Report Prepared Following Consultation on CICES Version 4, August–December 2012. EEA Framework Contract No EEA/IEA/09/003.
- Hermann, A., Kuttner, M., Hainz-Renetzeder, C., Konkoly-gyuró, É., Tirászi, Á., Brandenburg, C., Allex, B., Ziener, K., Wrška, T., 2013. Assessment framework for landscape services in European cultural landscapes: an Austrian Hungarian case study. *Ecol. Indic.* 37, 229–240, <http://dx.doi.org/10.1016/j.ecolind.2013.01.019>.
- Hou, Y., Burkhard, B., Müller, F., 2012. Uncertainties in landscape analysis and ecosystem service assessment. *J. Environ. Manag.* 127, S117–S131, <http://dx.doi.org/10.1016/j.jenvman.2012.12.002>.

- IPCC, 2014. *Confidence and Likelihood in the IPCC Fifth Assessment Report*. IPCC.
- Jacobs, S., Burkhard, B., Van Daele, T., Staes, J., Schneiders, A., 2014. The Matrix Reloaded: a review of expert knowledge use for mapping ecosystem services. *Ecol. Model.* 295, 21–30, <http://dx.doi.org/10.1016/j.ecolmodel.2014.08.024>.
- Kopperoinen, L., Itkonen, P., Niemelä, J., 2014. Using expert knowledge in combining green infrastructure and ecosystem services in land use planning: an insight into a new place-based methodology. *Landsc. Ecol.* 29, 1361–1375, <http://dx.doi.org/10.1007/s10980-014-0014-2>.
- Koschke, L., Fürst, C., Frank, S., Makeschin, F., 2012. A multi-criteria approach for an integrated land-cover-based assessment of ecosystem services provision to support landscape planning. *Ecol. Indic.* 21, 54–66, <http://dx.doi.org/10.1016/j.ecolind.2011.12.010>.
- Kroll, F., Müller, F., Haase, D., Fohrer, N., 2012. Rural-urban gradient analysis of ecosystem services supply and demand dynamics. *Land Use Policy* 29, 521–535, <http://dx.doi.org/10.1016/j.landusepol.2011.07.008>.
- Litchfield, J., Reilly, B., Veneziani, M., 2012. An analysis of life satisfaction in Albania: an heteroscedastic ordered probit model approach. *J. Econ. Behav. Organ.* 81, 731–741.
- Martnez-Harms, M.J., Balvanera, P., 2012. Methods for mapping ecosystem service supply: a review. *Int. J. Biodivers. Sci. Ecosyst. Serv. Manag.* 8, 17–25, <http://dx.doi.org/10.1080/21513732.2012.663792>.
- Millennium Ecosystem Assessment, 2005. *Ecosystems and Human Well-Being Synthesis*. Island Press, Washington, DC.
- Seppelt, R., Dormann, C.F., Eppink, F.V., Lautenbach, S., Schmidt, S., 2011. A quantitative review of ecosystem service studies: approaches, shortcomings and the road ahead. *J. Appl. Ecol.* 48 (3), 630–636, <http://dx.doi.org/10.1111/j.1365-2664.2010.01952>.
- Stan Development Team (Ed.), 2015. *Stan Development Team*.
- Stoll, S., Frenzel, M., Burkhard, B., Adamescu, M., Augustaitis, A., Baeßler, C., Bonet, F.J., Carranza, M.L., Cazacu, C., Cosor, G.L., Díaz-Delgado, R., Grandin, U., Haase, P., Hämäläinen, H., Loke, R., Müller, J., Stanisci, A., Staszewski, T., Müller, F., 2014. Assessment of ecosystem integrity and service gradients across Europe using the LTER Europe network. *Ecol. Model.* 295, 13, <http://dx.doi.org/10.1016/j.ecolmodel.2014.06.019>.
- Tschanz, L., Taton, T., Brun, J.-J., 2015. *Services écosystémiques. Pour une gestion durable des territoires, pensons en bouquet*. *Espac. Nat.* 52, 27–28.
- Vihervaara, P., Kumpula, T., Tanskanen, A., Burkhard, B., 2010. Ecosystem services-A tool for sustainable management of human-environment systems. Case study Finnish Forest Lapland. *Ecol. Complex.* 7, 410–420, <http://dx.doi.org/10.1016/j.ecocom.2009.12.002>.
- Vihervaara, P., Kumpula, T., Ruokolainen, A., Tanskanen, A., Burkhard, B., 2012. The use of detailed biotope data for linking biodiversity with ecosystem services in Finland. *Int. J. Biodivers. Sci. Ecosyst. Serv. Manag.* 8, 169–185, <http://dx.doi.org/10.1080/21513732.2012.686120>.
- Willemen, L., Burkhard, B., Crossman, N., Drakou, E.G., Palomo, I., 2015. Editorial: best practices for mapping ecosystem services. *Ecosyst. Serv.* 13, 1–5, <http://dx.doi.org/10.1016/j.ecoser.2015.05.008>.