Article (refereed) - postprint

Hosseini, A.; Stenberg, K.; Avila, R.; Beresford, N.A.; Brown, J.E. 2013 **Application of the Bayesian approach for derivation of PDFs for concentration ratio values**.

Application of Bayesian inference for estimating parameters of probability distributions of concentration ratios

A. Hosseini[a,*] , K. Stenberg[b], R. Avila[b], N.A. Beresford[c], J.E. Brown[a]

[a]Norwegian Radiation Protection Authority, Department for Emergency Preparedness and Environmental Radioactivity, Grini næringspark 13, Post Box 55, NO-1332 Østerås, Norway

[b]Facilia AB, S-167 51, Bromma, Sweden

[c]Centre for Ecology and Hydrology, CEH-Lancaster, Lancaster Environment Centre, Library Avenue, Bailrigg, Lancaster LA 1 4AP, UK

*Corresponding author, Tel.: +47 67 16 2647; fax: +47 67 14 7407

E-mail address: ali.hosseini@nrpa.no (A. Hosseini)

Abstract

Concentration ratios (CRs) are used to derive activity concentrations in wild plants and animals. Usually, compilations of CR values encompass a wide range of element-organism combinations, extracted from different studies with statistical information reported at varying degrees of detail. To produce a more robust estimation of distribution parameters, data from different studies are normally pooled using classical statistical methods. However, there is inherent subjectivity involved in pooling CR data in the sense that there is a tacit assumption that the CRs under any arbitrarily defined biota category belong to the same population. Here, Bayesian inference has been introduced as an alternative way of making estimates of distribution parameters of CRs. This approach, in contrast to classical methods, is more flexible and also allows us to define the various assumptions required, when combining data, in a more explicit manner. Taking selected data from the recently compiled wildlife transfer database (http://www.wildlifetransferdatabase.org/) as a working example, attempts are made to refine the pooling approaches previously used and to consider situations when empirical data are limited.

1. Introduction

Quantification of risk to wildlife as a consequence of releases of radioactivity to the environment requires the determination of activity concentrations in various environmental media, e.g. soil and water, and selected plants and animals. A simple and widely used method to model radionuclide transfer from soil and water to biota is through the application of concentration ratios: $CR_{wo\text{-}media}$'s (Beresford et al., 2008; Hosseini et al., 2008), defined as concentration of a radionuclide in the whole organism divided by the concentration of the same radionuclide in the relevant environmental

media, i.e. soil or water. The whole body concentration ratio, $CR_{wo\text{-}media}$, will be abbreviated in the following text as CR. As there are many data gaps associated with CRs, a variety of extrapolation approaches have been developed to provide missing values (e.g. Brown et al. this issue). In this respect, comprehensive compilations of CRs in databases such as those provided in the ERICA Tool (Beresford et al., 2008; Hosseini et al., 2008) and the more recent wildlife transfer database as described by Copplestone et al. (this issue) (which has been used to derive an IAEA handbook (Howard et al. (in-press)) and ICRP (2009) report) provide a valuable resource.

In the course of analysing data for the development of such transfer databases, some consideration has been given to how data will be used in a more robust risk characterization where not only the severity, but also the probability of occurrence of the exposure needs to be considered (Brown et al., 2008). The requirement relates not only to detailed, site specific risk characterisation but also to screening assessments where high percentile CRs, that can only be derived from parameters with characterised probability density functions (PDFs), are often utilised in the derivation of screening criteria, such as limiting media concentrations (e.g. see Brown et al., this issue). Furthermore, by providing statistical information, the uncertainty associated with the calculation of exposure can be propagated through the assessment using approaches such as Monte Carlo simulation (e.g. see Vose, 1996). Hence, there are clear drivers to acquire statistical information, particularly in relation to the assignment of PDFs to CRs within the datasets underpinning the assessment.

While initial work in generating CR values for environmental exposure assessments, as exemplified by the ERICA Tool (Brown et al., 2008), EA R&D128 (Copplestone et al., 2001) and the USDoE's Graded Approach (USDoE, 2002) provided transfer information at generic levels, more recent efforts have aimed at providing a greater level of detail. For example, the IAEAs wildlife transfer handbook (Howard et al., in-press) provides data tables which allow for more specific information to be

accessed for some animal groups, e.g. CR values for fish by feeding group as described by Yankovich et al. (this issue). The ICRP (2009) have attempted to present CR data at the taxonomic family level. A common step for all these data collations efforts is pooling or combining data to produce more precise estimates for the parameters of interest. However, the combined data are usually extracted from different studies with variable sample sizes, different measures of central tendency and dispersion. Another common issue is the application of different extrapolation approaches to derive missing transfer parameter values.

Sheppard (2005) advised against basing assessments solely on site-specific data and argued that given the inherent large variability of transfer parameters using a few on-site data to the exclusion of many generic data may decrease accuracy due to the potential error resulting from too few measurements. However, guidance was not provided on how to combine both site-specific and generic data such that all available information could be taken into consideration without imposing unnecessary bias into any subsequent calculations.

The main objective of this paper is to provide an alternative approach on how to utilise the various related datasets/information that are often available in addition to sets of values that are specific to an organism grouping, site or element, using the wildlife transfer database of Copplestone et al. (this issue) for illustrative examples. In this way, we hope to progress beyond the tendency to select either generic or specific values that has characterised earlier discussions on radionuclide transfer and to provide refined methods to combine transfer data. This goal will be achieved by application of Bayesian statistics which allows both prior knowledge and site or study specific empirical data to be used in estimating PDF parameters Linacre et al. (2004). Given that the prior knowledge is valid, this approach provides more robust parameter estimates as compared to when only limited site or study specific empirical data are used.

In the following sections, first different approaches that can be applied for deriving PDFs of CRs for a case of interest are discussed. Thereafter, the Bayes Theorem is introduced along with a brief description of situations where it can be applied in the context of derivation of PDFs for CRs. Finally, three Bayesian techniques for derivation of PDFs are illustrated with relevant examples. More details related to these techniques along with a guidance on how to apply them and interpret outputs are provided in an appendix.

2. Derivation of a PDF of CR that is representative for a case of interest

The approach that can be applied for the derivation of a PDF characterising a CR that is representative for a case of interest will depend on the availability of representative data. By case of interest we mean the CR for a given species, family, radionuclide or site. Four typical situations with regards to the availability of representative data can be envisaged:

- Sufficient representative data are available. In this case ordinary fitting techniques, like maximum likelihood estimation (Keeping, 1995) can be directly applied to derive the distribution and obtain estimates of the distribution parameters.

  We are aware that it is difficult to define *a priori* the number of data required to obtain a reasonably good approximation of the location and scale parameters of a distribution. The number of data which are needed to derive a sound estimate of the central value of a distribution is much smaller than the number of data needed to obtain a reasonable estimate of the variability of that distribution. The number of required data points depends on the variability of the measured parameter. The higher the variability, the larger the number of data points required. Hence, there will always

be an element of judgement involved in deciding what constitute a ‘sufficient number of data’ as this depends on our knowledge base and the purpose of the study.

- Neither representative nor literature data are available. In this case suggested approaches exist for deriving the distributions, such as those based on extrapolation methods. Typical examples include using the CR distribution for an analogue such as a similar reference organism or biogeochemical element to the case of interest (Brown et al., this issue).

- Limited representative data are available for the case of interest, but no supporting data which have some relationship with the case of interest can be found. In this case all required distribution parameters are derived solely from the available data, with large associated uncertainties. It is, however, rarely the case that no other relevant data and information can be found.

- Limited data for the case of interest are available, but other relevant data can be found from the literature or as a result of using extrapolation approaches (e.g. data for the given radionuclide may be available for a similar organism). In previous studies this case has been handled by using an ordinary (classical) data pooling approach; as used in development of the ERICA (Beresford et al., 2008; Hosseini et al., 2008) and TRS CR databases (Howard et al., in press). This case can also be addressed by combining situation specific data with other relevant data using Bayesian statistics, which provides the focus for this paper. The choice of one or other Bayesian methods will depend on assumptions made which encompass our knowledge/ belief about the relevancy of the data under consideration.

3. Estimation of distribution parameters using Bayesian inference

Based on different concepts of probability, statistics may be divided into two main schools (Suter, 2007): Frequentist and Bayesian. Probability from a Frequentist (or relative-frequency) point of view is understood as an expression of frequency. Whereas, Bayesian statistics: 1) defines probability as a conditional measure of uncertainty and 2) provides a method for modification of probability in the light of new evidence. In the Bayesian approach parameters are treated as random variables. This is not a description of their variability alone, but a description of the uncertainty about their true values (Bernardo, 2003). And quantifying this uncertainty by making use of probability is the essential characteristic of all Bayesian methods (Gelman et al., 2004).

Suppose that inferences are to be drawn on the unknown parameter (or parameter vector) θ in light of vector of independent and identically distributed empirical data values $\boldsymbol{y} = y_1, \ldots, y_n$ and a prior probability distribution $p(\theta)$. Bayes' theorem (Bayes, 1763) provides the means for combining information from the prior and the likelihood to produce the posterior density of the parameter conditioned on the data:

$$p(\theta \mid y) = \frac{p(y \mid \theta)\, p(\theta)}{\int_{-\infty}^{\infty} p(y \mid \theta)\, p(\theta) d\theta} \qquad \text{or} \qquad \text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\int_{-\infty}^{\infty} (\text{likelihood} \times \text{prior}) d\theta}$$

Here, the posterior distribution describes our state of knowledge about the parameter θ after considering the data. The likelihood function describes how probable the current data are given the parameter θ. The prior represents the present state of our knowledge based on an initial consideration of the parameter θ. The denominator is the probability of the data, a normalising

constant. Hence, the combined (posterior) probability distribution of the parameter given the empirical data is proportional to the prior probability distribution times the likelihood function of the empirical data values:

$$p(\theta \mid y) \propto \prod_{i=1}^{n} p(y_i \mid \theta) \times p(\theta)$$

The power of Bayes' Theorem lies in the fact that it relates the quantity of interest (the probability of each specific value of the distribution parameter given the measured data, i.e. $p(\vartheta \mid y)$) to a term that we have a better chance of quantifying (the probability of the measured data given a specific value of the parameter; $p(y \mid \vartheta)$). Note that in this way we are implicitly recognising that the value of the distribution parameter is an uncertain quantity, meaning that different possible values have different probabilities.

The procedure for assigning a probability distribution to a model parameter, such as CR, can be divided into two main steps: i) selection of a probability model or distribution type and ii) estimation of the distribution parameters. We will assume that the CRs are log-normally distributed. This assumption is based on arguments of the log normality of concentrations of elements in the environment (Ott 1990, 1995), empirical observations of element concentrations in the environment that are used for calculation of CR values (Nordén et al., 2010) and the fact that the quotient of two log normal random variables is log normally distributed.

Assuming $\mathrm{y_i}', \mathrm{i} = 1, \ldots, \mathrm{n}$, representing a log-normally distributed dataset then we have

$$y_i \sim N(\mu, \sigma^2), \quad i = 1, \ldots, n$$

where $y_i = \ln y_i'$. That is, the original measurements transformed by the natural logarithm are normally distributed where μ and σ^2 are true mean and variance of the transformed data values. Log-normal data are often described using geometric mean and variance and are calculated from the estimates of μ and σ^2 as $GM(y') = \exp \mu$ and $GSD(y') = \exp \sigma$.

Now what remains is to discuss how to estimate the distribution parameters, i.e. the geometric mean (GM) and geometric standard deviation (GSD). For this purpose, we here discuss three methods of Bayesian inference.

3.1 A joint prior distribution for the mean and variance

The first Bayesian inference method considered here for estimation of the distribution parameters is based on the assumption that the available historical (literature) data used for deriving the prior distribution, and the available empirical data for the case of interest are exchangeable and can be considered as being from the same population. The method is attractive in cases where the prior information takes the form of a fixed number of samples with the same population variance as the case of interest (Gelman et al, 2004). In this case, the historical data carries information on both the mean and variance of the distribution and can be used to define a joint *conjugate prior distribution* (Gelman et al., 2004). A prior distribution is said to be 'conjugate' to the measurement model if the resulting posterior distribution is of the same functional form as the prior. The set of prior distributions that are jointly conjugate to the normal measurement model are expressed using a normal distribution of the prior mean and a Inverse-Chi-Square distribution for the variance: $p(\mu|\sigma^2) \sim N(\mu_0, \sigma^2/n_0)$ and $p(\sigma^2) \sim Inv - \chi^2(v_0, \sigma_0^2)$, where the parameters with subscript zero are set to the mean, variance and size of the prior sample. The two prior distributions are dependent in that the prior uncertainty of the mean is conditioned the unknown variance σ^2,

reflecting the fact that the prior and data both provide information about the same population variance.

When combined with new data using Bayes' theorem, the prior distributions are updated and the posteriors have the same form as the two priors, but with new parameters (see Appendix, Equations A4-A5). The posterior mean is then a weighted combination of the sample mean and the prior mean, with the weights being the samples sizes. The posterior variance is expressed as sum of the: i) variation within prior, ii) variation within the data and iii) the weighted squared difference between the two. The posterior variance is thus adjusted for the uncertainty provided by the difference of the prior and data means.

Bayesian inferences from the joint posterior distribution are obtained by repeatedly drawing samples from the posterior distributions. The obtained samples can then be summarized with statistics of interest, such as the mean, median or percentiles which allows the assessor to express the posterior probabilities of a distribution parameter

3.2. Independent prior distributions for the mean and variance

The second Bayesian updating technique considered here can be applied in situations where, in addition to data for the case of interest, there is available relevant historical data, but this data does not take the form of a fixed number of observations with the same variance σ^2 (Gelman et al., 2004). This would be the case if the historical data does not carry information about the variance of the case of interest or if other information than historical data is available for the variance. In such a situation, the prior distributions of μ and σ^2 are specified independently, using a prior distribution of the mean as $N(\mu_0, \tau_0^2)$ and the prior $Inv-\chi^2(\nu_0, \sigma_0^2)$ for the variance. If no prior information is available for

the variance, a so called non-informative prior for the variance can be used by setting $\nu_0 = 0$. With these prior distributions the *conditional* posterior distributions $p(\mu|\sigma^2, y)$ and $p(\sigma^2|\mu, y)$ attain the same functional form as the prior, but the joint conjugate posterior $p(\mu, \sigma^2|y)$ does not. Therefore, these prior distributions are often referred to as *semi-conjugate* prior distributions.

When combined with new data using Bayes' theorem, the posterior distributions take the same form as the priors but with updated parameters (see Appendix, Equations A9-A10). The posterior distribution of the mean reflects the combined amount of information available both from the data and the prior (historical data). Means from studies with larger sample sizes *n* carry more information, and hence the corresponding posterior will be pulled much less towards the mean of the historical data. At the limit where *n* approaches "infinity" or when the prior variance is very large, the posterior mean is simply the sample mean of the data.

The posterior distribution of the variance is expressed in terms of a weighted combination of the sample variance, an estimate of the prior variance and the squared distance between the data and posterior mean. The weights are the number of measurements n and the prior degrees of freedom ν_0 for the variance.

3.3 Hierarchical model: Prior implicitly derived from data of several similar units

Consider a number of related units (such as species or elements which share some common traits) or groups of measurements that are believed to be similar with regards to the parameter of interest. When making estimates of similar quantities, an approach called hierarchical updating can be used to

obtain estimates for all quantities simultaneously, letting the units borrow strength from the ensemble (Morris, 1983). The method offers an alternative to using either separate estimates or a complete pooled estimate and the estimates from hierarchical models are therefore sometimes called partially pooled estimates (Gelman et al., 2004; Gelman and Hill, 2007). This is accomplished by assigning a *common* prior distribution $N(\mu, \tau^2)$ for the mean $\mu_j, j = 1, \ldots, J$ of the J individual units. The two parameters of the prior distribution (often denoted hyper-parameters) are now considered unknown. When fitting the hierarchical model, posterior distributions are therefore obtained for individual units and for the two hyper-parameters (interpreted as the population mean and population variance respectively). Because the variance τ^2 is the variance of the units' means, it is also often referred to as the between-units variance. Further prior information could be included for the two hyper-parameters, but we are here assigning them non-informative prior distributions to let them be estimated hierarchically with the included data.

After fitting the hierarchical model the posterior estimates of the mean for individual cases are partially pooled towards the population mean, but only so much as allowed by the individual sample sizes and variation between cases' mean values. That is, if the estimated similarity of the means is large (indicated by a small estimated τ^2) the prior will have a strong impact and the amount of pooling of individual means will be large. If the estimated similarity of the means is small (the estimated τ^2 is large), the amount of pooling will be smaller.

The within unit variance σ^2 can be assumed as equal or different for all units. However, we consider the assumption of equal variance to be more reasonable in cases where species/elements of interest are believed to be similar. If this assumption holds, data for all units are used to estimate the variance, which is desirable when there are few data points for individual cases. Under a non-

informative prior, the full conditional posterior for the common variance is expressed as the pooled variance of all units with an additional component of uncertainty provided by the discrepancy of the updated mean and observed mean.

The posterior distribution of the unit-level parameters and the two hyper-parameters are described in terms of the full conditional posterior distributions (see Appendix, Equations A13-A17). To draw samples from the joint posterior distribution a Gibbs sampling algorithm can be implemented to iteratively draw samples from the one-dimensional full conditional posterior distributions. The posterior distributions can then be summarized with median, means and percentiles of the obtained samples. Probable values of the mean of new case (e.g a species not yet observed but assumed to belong to the same population as those included in the hierarchical model) can be predicted by drawing values from the posterior distribution of means $\mu \sim N(\mu^*, \tau^{2^*})$ where μ^*, τ^{2^*} are a set of samples from the joint posterior distribution of the hyper-parameters.

4. Illustrative examples

In the following examples CR data for 8 species of bat (Table 1) are used to illustrate the Bayesian methods described in Section 3. The underlying data for species with N>2 was assessed to be approximately log normally distributed by visual inspection of quantile-quantile plots. Based on these observations and also the theoretical consideration such as those addressed by Ott (1990, 1995) the log normal assumption was also used for the two species with N=2. All statistical analyses for this example have been conducted using the BABAR software package which can be downloaded at no charge from the website: http://facilia.se/projects/babar.asp. BABAR is a software package for

constructing and updating probability density functions of parameters for use in risk assessment models. It makes use of Bayesian statistics to derive distributions of parameters. For all calculations, inferences were based on 10 000 samples from the posterior distributions in three independently simulated chains (using different and dispersed start values) and discarding the first 500 samples in each chain (resulting in a total of 28 500 samples). Convergence was checked by inspecting graphs of the obtained simulated samples.

Let us assume that our aim is to derive a PDF for the CR of the Daubenton species of bat (Code X in Table 1). We have assumed that the CRs follow a lognormal distribution and the goal is then reduced to obtain estimates of the GM and GSD of the distribution. Since only 2 data points are available for this species, maximum likelihood estimates (e.g. sample mean and variance) will not work well. However, there are data available for other 7 bat species that can be considered more or less representative (analogues) for the species of interest (case of interest). One possibility would be to consider that one particular species is a closer analogue to species X than the others. For example, one assessor may have reasons to believe that E is a closer analogue to X than all other species in Table 1. They could then go further and assume that the CR data for E and X are exchangeable and belong to the same population. They will then use the CR data for E to specify a joint conjugate prior distribution and obtain a posterior distribution for the distribution parameters of CR representing the combined data of X and E using the method described in Section 3.1. Other assessors may have different believes about which species is the best analogue for X and will therefore obtain different posterior distributions (See Table 2). The estimates of GM, GSD were derived from the medians of the posterior distributions of μ, σ^2 ,as described in the appendix. In Figure 1 such posterior medians have been used to plot the distribution of ln(CR) for three different cases where Daubenton's bat data (X) is combined with data of various analogue.

The seven posterior distributions have GSD ranging from 2.3 to 11 with the extreme GSD obtained when X is combined with the other small data set for species G (Soprano Pipistrelle bat also with only two data points). An extreme upper GSD posterior percentile reflects the large uncertainty associated with this combination. None of the 95% probability intervals of the GMs contains the observed GM=0.01 except the one derived for species G. This reflects the fact that apart from the case of species G, the prior is allowed to dominate the data in all other cases. This is a consequence of assuming both data sets being equally representative for the case of interest. To facilitate a comparison of probable CR values under the posterior distributions, the arithmetic mean and 95% upper percentile are calculated using the formulas below. The large variation of the arithmetic means and 95% percentiles (by up to a factor of about 20) indicates that there is a large sensitivity to the choice of analogue (and hence the prior).

$$Mean = GM \cdot e^{0.5(lnGSD)^2}$$

$$Perc_{95\%} = GM \cdot GSD^{1.64}$$

Assume now that among all the considered species in Table 1 we cannot select any of the species as being a better analogue for the species of interest (X). However, we can still consider that the CR values for all species are exchangeable with each other. In this case, we can apply a hierarchical model as described in Section 3.3 to obtain posteriori distributions for the distribution parameter GM for each of the eight species (shown in Table 3) and the common GSD.

The posterior distribution of Daubenton species (with median 0.04) is partially pooled towards the population mean in a degree depending on the statistical similarity of the species (estimated with the population mean μ and between unit variance τ^2, shown in table 4). The posterior distribution of Daubenton species takes into account information from both the Daubentons data and the data of related species without weighting them equally (in contrast to the conjugate approach where the prior is seen as equally representative). This is reflected in the 95% posterior probability interval of GM, which includes the observed Daubenton species GM of 0.01.

Figure 2 shows the posterior distributions of μ (medians and 95% probability intervals) together with data (Maximum Likelihood Estimates (MLE) $\bar{y}$ and 95% confidence intervals estimated $\mu \pm 1.96 \cdot \sqrt{s^2/n}$). Making use of the hyper-parameters listed in Table 4, a new distribution has been constructed for an unobserved species belonging to the same population. The last row in Figure 2 shows the predicted distribution of this new species, denoted as “predicted”, calculated by drawing samples from the posterior population distribution of unit means as described in 3.3.

Finally, let us assume that the data in Table 1 for all species, except for X, has been pooled using the approach of combined mean and variances (see Appendix, Equations A2-A3) as used in development of the ERICA and Copplestone et al (this issue) CR databases. The pooled data does not carry information about the within species variance and therefore a joint conjugate prior cannot be specified to combine with the CR data for X using the method described in Section 3.1. Obviously, it is not possible to specify a hierarchical model either. We can, however, specify independent prior distributions for μ and σ and apply the method described in Section 3.2. The pooled distribution is

thus used as a prior for the mean with parameters GM=0.1, GSD=3.8. The prior for the variance is set as non-informative (by using $\nu_0 = 0$). The results are presented in Table 5.

Because of the large variance of Daubenton species (compared to the pooled data of all other species) and the low sample size, the posterior mean is heavily affected by the prior (left graph in Figure 3). The GSD is estimated from sample variance of the species of interest and an additional variance component accounting for the difference between the updated mean and data. The large posterior GSD (median GSD=7.0 and an extreme upper percentile) is the result of the very large uncertainty associated with the two data points and because of the discrepancy between the observed data and the prior. We compare this with the pooled distribution of all 113 data points for all species with GM=0.10, GSD=4.0 (right graph in Figure 3) which is practically equal to the distribution of the pooled data when Daubenton data was excluded. The result of the semi conjugate method represents the case of interest while accounting for prior probable values of the mean. The pooled distribution however, can be seen as representing the whole population.

4. Discussion

Bayesian approaches require elicitation of prior distributions for parameter values. In the context of the present work priors have been used to take into account the existing external information/data and also to describe our belief about the relevancy of these data for our case of interest. To individuals not familiar with Bayesian methods the inclusion of information based upon belief about the relevancy of data may seem overly subjective. However, this is exactly the process that is employed during pooling of datasets using classical statistical

methods wherein, for instance, knowledge/belief about congruity of transfer factors between taxonomically similar organism groups is used to delineate the extent to which data are combined. For example, in the wildlife transfer database (Copplestone et al., this issue), there is a tacit assumption that there is some rationale in grouping (e.g.) mammals in terms of feeding strategy, this assumption being based upon prior knowledge/belief regarding the importance of the ingestion pathway in determining internal radionuclide body burdens and similarities in physiology dictating uptake. Hence, while the 'classical' approach implicitly uses judgment, the Bayesian approach explicitly acknowledges the role of judgments made.

Furthermore, as priors represent external insight/information, that is not included in likelihood function, considering them is not only a necessary step in the process of learning and acquiring knowledge, but also a crucial element for coming to the right conclusion (Kruschke, 2010).

In the context of the present work, three methods have been used to illustrate how Bayesian statistics is able to provide the assessor with extra flexibility to combine data and information in a desired way. In updating with a joint conjugate prior we want to update a PDF with new data and the resulting GM and GSD reflects the combined data sets. The method results in point estimates similar to the approach of combined mean and variances (see Appendix, Equations A2-A3), but it also produce estimates of the uncertainties of the posterior parameter estimates. With regards to the bat data, this approach was used to combine the data for species of interest with data from each of the other species, considering each as analogue. The seven resulting posteriors were sensitive to the choice of analogue and demonstrate well the effect of this method: The resulting posterior distribution reflects the total accumulated data for both species, considering both data sets as equally representative, i.e. exchangeable.

However, when updating the mean and variance with independent (semi-conjugate) prior distributions we are interested in deriving a PDF for a given species itself by using the available empirical data and relevant information from other sources. The belief/knowledge we are expressing in using this method is that although the given species may share common traits with regards to radionuclide transfer with other species, there is a substantial likelihood that the species under consideration expresses its own unique CR values thus rendering mere pooling of data or conjugate updating inappropriate. The independent priors offers the means of placing emphasis on the species-specific data whilst linking the data to what is known about related generic datasets in a mathematically structured way. In our example, the pooled distribution of all other bat species was used as prior for the mean and was updated with the Daubenton bat data. However, since the pooled distribution did not provide any good information of the within species variance, the GSD was estimated using only the two available data points and resulted in a large posterior GSD. Nonetheless, one might expect that as more data is obtained for the Daubenton species the GSD will decrease, converging to the GSD corresponding to the natural variability of the CR for this species.

The Bayesian updating approaches demonstrated and discussed in this paper can be used to help design efficient sampling strategies (Ott, 1995). After computing the first posterior, this can be then used as a prior upon the arrival of new data. We can keep updating with new data until probabilities are adequately defined or we run out of resources. This approach has revolutionized clinical trials as it converges to desired result more quickly than conventional approaches and maximizes benefits to participants (Suter, 2007).

The hierarchical method allows for a more flexible way of weighting the means compared to when we just apply the classical combined mean method (see Appendix, Equation A2). This is achieved by allowing the weights to be derived on the basis of how "similar" the data of the considered units are. Hierarchical estimation implicitly derive prior information from other data sets and could be seen as a compromise of two extremes: i) using the combined mean of all cases, but ignoring the differences between the individual cases and ii) using the estimates for the case of interest and ignoring other information.The necessary assumption for data to be included in a hierarchical model is that of exchangeability. The concept of exchangeability is closely related to the degree of ignorance we have about a problem. This means, the less we know about a problem the more confidently we can make claims of exchangeability (Gelman et al., 2004). Here for the hierarchical model under discussion we have assumed that we do not have any information available to distinguish between the CR values of the different bat species. But if we have any other measured factor or information that can be correlated with CR and these are considered to be of importance, the hierarchical model could be expanded to take into account these factors. For more detail on this subject the reader is referred to Gelman and Hill (2007).

In cases where extrapolation approaches are used to provide missing model parameter values, sooner or later some empirical data for those cases will become available. For example, when underlying CR databases of the ERICA Tool were established for a few years ago about 60% of the CR values were derived using a variety of extrapolation approaches (Brown et al. this issue). However, now through collation of a new database (Copplestone et al., this issue) some new empirical data have become available. It is clear that application of extrapolation approaches is based on the assumption of some kind of similarity or commonality between the missing data and the data being extrapolated. This means we believe that the surrogate data contain some information about the missing value. However, this begs the question as to what status these surrogate data have once

some new empirical data for the desired parameter become available? Do they become irrelevant or can they still play a role in our estimation of the unknown parameter? These questions are especially relevant in cases where the numbers of newly acquired empirical data are scarce. In such cases ignoring the surrogate data completely amounts to ignoring relevant information and at the same time using a few data points alone may lead to a high degree of uncertainty. A compromise in such situations is to use data from both sources, using the semi-conjugate updating approach introduced in Section 3.2. However, a crucial step here is to define a suitable prior such that it allows for the external information from the extrapolation approaches to be incorporated without overruling the importance of the empirical data which is the main source of information.

For relatively simple problems such as evaluating CR values, employing non-informative priors will result in parameters estimate close to those coming from a classical analysis, but, this similarity fades away as we consider more and more complex models (Clark 2005). With complex models we mean models which are developed through the extension of the usual Bayesian structure to include more than two levels (i.e. more than just likelihood and prior). The hierarchical model introduced in Section 3.3 has three levels (likelihood, prior and hyper-parameters) and provides a simple example of such models. The flexibility and capacity of such hierarchical models in efficiently exploiting and combining information from multiple sources are demonstrated in various studies (Clark, 2005; Gelman and Hill, 2007; Kruschke et al., 2012).

Given the computational power of today's hardware and availability of modern software, Bayesian methods for data analysis are now easily accessible to everyone and getting more and more attention in different scientific disciplines (Kruschke et al., 2012; Lester et al., 2007). In this paper we hope that we have demonstrated their usefulness to radioecological studies.

Acknowledgements

This work was performed within the STAR (A Strategy for Allied Radioecology) Network of Excellence supported by the EC-EURATOM 7th Framework Programme and Norwegian Research Council. The financial support of the Norwegian Research Council is gratefully acknowledged.

5. References

Bayes, T., 1763. An Essay towards solving a Problem in the Doctrine of Chances. Philosophical Transactions of the Royal Society of London 53: 370–418.

Beresford, N.A., Barnett, C.L., Howard, B.J., Scott, W.A., Brown, J.E., Copplestone, D., 2008. Derivation of transfer parameters for use within the ERICA Tool and the default concentration ratios for terrestrial biota. Journal of Environmental Radioactivity, Volume 99, Issue 9, Pages 1393-1407.

Bernardo, J. M., 2003. Bayesian Statistics. http://www.uv.es/bernardo/BayesStat.pdf (accessed Mar 2012).

Brown, J.E., Alfonso, B., Avila, R., Beresford, N.A., Copplestone, D., Pröhl, G., Ulanovsky A., 2008. The ERICA Tool. J. Environ. Radioact. 99/9, 1371-1383.

Brown, J.E., Beresford, N.A., Hosseini, H. (this issue). Approaches to providing missing transfer parameter values in the ERICA Tool – how well do they work? Journal of Environmental Radioactivity.

Clark, J.S., 2005. Why environmental scientists are becoming Bayesians. Ecology Letters, 8, Pages 2 – 14.

Copplestone, D., Bielby, S., Jones, S.R., Patton, D., Daniel, P., Gize, I., 2001. Impact Assessment of Ionising Radiation on Wildlife. R&D Publication 128. Environment Agency, Bristol. ISBN 1-85705590.

Copplestone, D., Beresford. N.A., Brown, J.E., Yankovich T. (this issue). An international database of radionuclide Concentration Ratios for wildlife: development and uses. Journal of Environmental Radioactivity.

Gaschak, S, Beresford N.A., Maksimenko A., Vlaschenko A.S., 2010. Strontium-90 and caesium-137 activity concentrations in bats in the Chernobyl exclusion zone *Radiation Environ Biophys*., 49, (4), Pages 635-644.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian Data Analysis. Chapman & Hall, 2nd edition, 2004. ISBN 1-58488-388-X.

Gelman, A., Hill, J., 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press. ISBN 0-521-86706-1.

Gelman, A., Rubin, D.B, 1992. Inference from Iterative Simulation Using Multiple Sequences. Statistical Science, Volume 7, No. 4, Pages 457-472.

Hosseini, A., Thørring, H., Brown, J.E., Saxén, R., Ilus E., 2008.Transfer of radionuclides in aquatic ecosystems – Default concentration ratios for aquatic biota in the Erica Tool. Journal of Environmental Radioactivity, Volume 99, Issue 9, Pages 1408-1429.

Howard, B.J., Beresford, N.A., Copplestone, D., Telleria, D., Proehl, G., Fesenko, S., Jeffree, R.A., Yankovich, T.L., Brown, J.E., Higley, K., Johansen, M.P., Mulye, H.,Vandenhove, H., Gashchak, S., Wood, M.D., Takata, H., Andersson, P., Dale, P., Ryan, J., Bollhöfer, A., Doering, D., Barnett, C.L., Wells, C. (in press). The IAEA handbook on radionuclide transfer to wildlife, Journal of EnvironmentalRadioactivity (2012), doi:10.1016/j.jenvrad.2012.01.027.

ICRP, 2009. Environmental Protection: Transfer Parameters for Reference Animals and Plants. ICRP Publication 114, Ann. ICRP 39(6).

Keeping, D.E., 1995. Introduction to statistical inference, Dover Publications, New York.

Kruschke, J.K., 2010. What to believe: Bayesian methods for data analysis. Trends in Cognitive Sciences, 14, Pages 293-300.

Kruschke, J.K., Aguinis, H., Joo, H., 2012. The Time Has Come: Bayesian Methods for Analysis in the Organizational Sciences. Organizational Research Methods, 15(4), Pages 722-752.

Lester, R.R., Green, L.C., Linkov, I., 2007. Site-Specific Applications of Probabilistic Health Risk Assessment: Review of the Literature Since 2000. Society for Risk Analysis, Volume 27, No. 3, Pages 635-658.

Linacre, N.A., Oaten, A.S., Burgman, M.A., 2004. Incorporating Collateral Data in Conservation Biology.Conservation Biology, Volume 18, No. 3, Pages 768-774.

Morris, C.N., 1983. Parametric empirical Bayes inference: theory and applications. Journal of the American Statistical Association, 78, Pages 47-65.

Nordén, S., Avila, R., de la Cruz, I., Stenberg, K., Grolander, S., 2010. Element-specific and constant parameters used for dose calculations in SR-Site. SKB TR-10-07, Svensk Kärnbränslehantering AB.

Ott, W.R., 1990. A physical explanation of the lognormality of pollutant concentrations. Journal of Air Waste Manage Assoc. 1990 Oct;40(10):1378-83.

Ott, W.R., 1995. Environmental Statistics and Data Analysis. Lewis Publishers, Boca Raton.

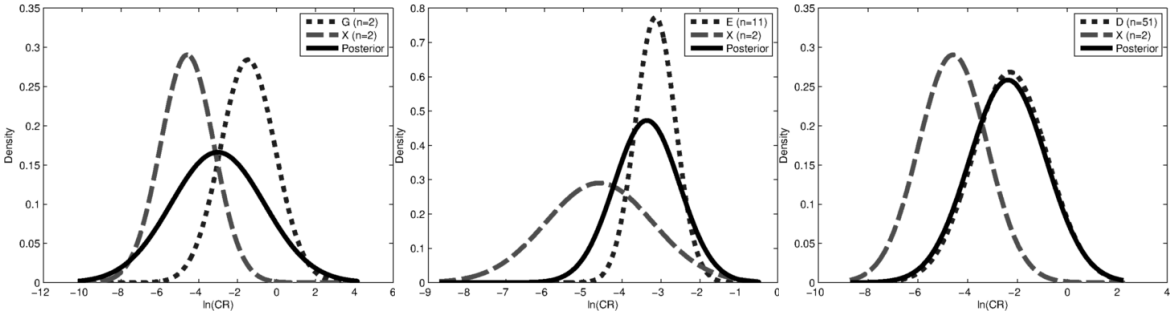
Sheppard, S.C., 2005. Transfer Parameters—Are On-Site Data Really Better? Human and Ecological Risk Assessment, 11: 939–949, 2005.

Suter II, G.W., 2007. Ecological Risk Assessment. Second Edition. CRC Press.

USDoE, 2002. A Graded Approach for Evaluating Radiation Doses to Aquatic and Terrestrial Biota. U.S. Department of Energy, Washington D.C.,U.S.A. Technical Standard DOE-STD-1153-2002.

Vose, D., 1996. Quantitative risk analysis – A guide to Monte Carlo simulation modelling. John Wiley and Sons.

Yankovich, T.L., Beresford, N.A., Fesenko, S., Fesenko, J., Phaneuf, M., Dagher, E., Outola, I., Andersson, P., Thiessen, K., Ryan, J., Wood, M.D., Bollhöfer, A., Barnett, C.L., Copplestone, D. (this issue). Establishing a database of radionuclide transfer parameters for freshwater wildlife. Journal of Environmental Radioactivity.



(a) **(b)** **(c)**

Figure 1. The distribution of ln(CR) with parameters selected as the posterior medians of the posterior distribution of μ and σ^2 when species Daubenton's bat (X) is combined with one other species seen as analogue. a) species G (N=2) as analogue leads to the largest posterior GSD among all tried analogues because of the total few observations, b) species E (N=11) as an analogue has a smaller GSD than data and leads to smaller posterior median GSD, c) species D with N=51 is used as analogue and almost completely dominates the data (N=2).

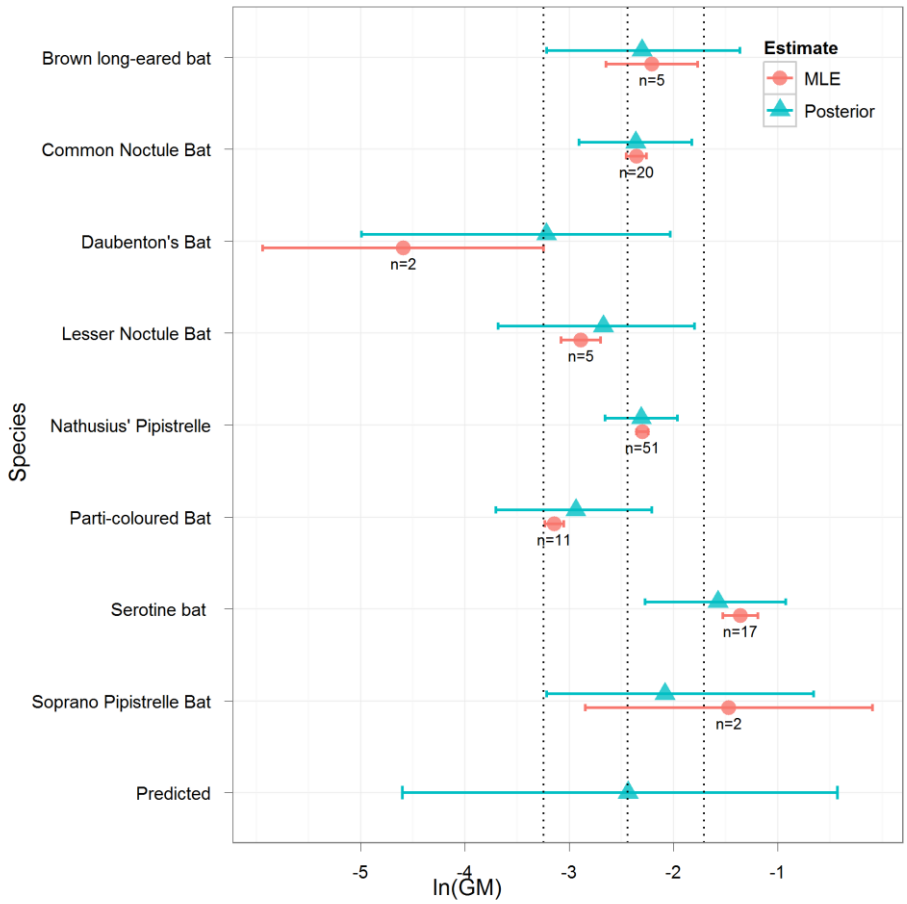
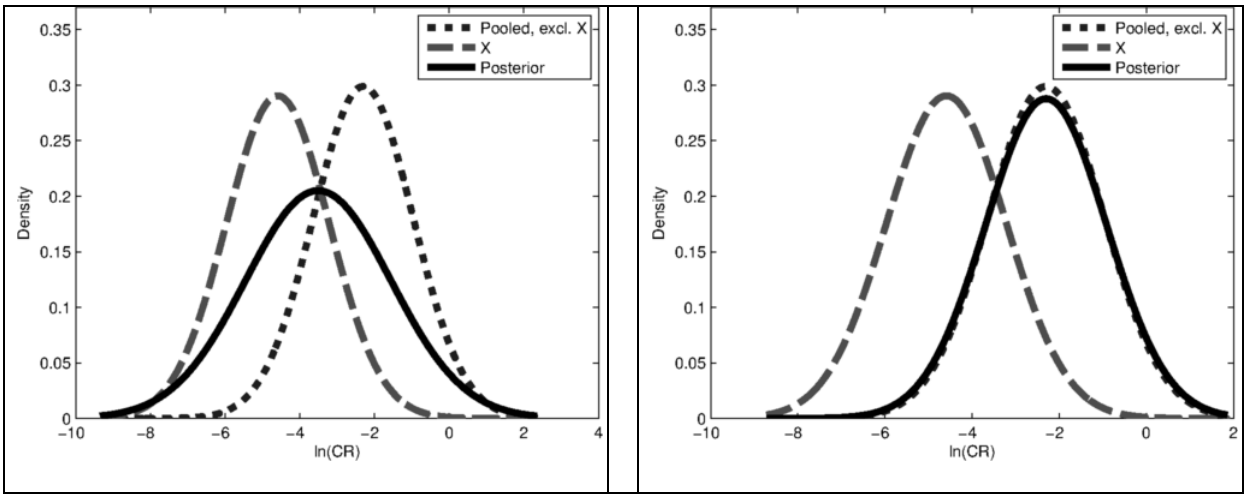


Figure 2. The posterior logarithmic mean from the hierarchical model assuming a common variance for the species. The Maximum Likelihood Estimate (MLE) is shown with 95% confidence interval ($\bar{y} \pm 1.96 \cdot \sigma\sqrt{n}$), The last bar shows the median and 95% probability interval of the predicted mean of a new species belonging to the same population, calculated from the hyper-parameters . The vertical dotted lines show the posterior median and 95% probability interval of the population mean μ. The posterior mean of the species of interest, Daubenton bat, is heavily influenced by population, but the 95% probability interval still includes the observed mean.



(a) Semi conjugate method (b) Conjugate method

Figure 3. The pooled population distribution (excluding species X) updated with species X using independent (semi conjugate) priors (left). The pooled distribution is used as a prior of the mean, but it does not provide any prior information about the variance of X. The posterior variance becomes much larger than the data variance to adjust for the updated mean. The fully conjugate approach (right) leads to a posterior very similar to the pooled data (The solid line practically overlays the dotted line).

Table 1. Data for eight bat species sampled in the Chernobyl exclusion zone (from Gaschak et al., 2010)

Code	Species	N	GM	GSD
A	Brown long-eared	5	0.11	3.1
B	Common Noctule	20	0.10	2.7
X	Daubenton	2	0.01	3.9
C	Lesser Noctule	5	0.06	1.6
D	Nathusius' Pipistrelle	51	0.10	4.4
E	Parti-coloured	11	0.04	1.7
F	Serotine	17	0.26	4.3
G	Soprano Pipistrelle	2	0.23	4.1

Table 2. Parameter values (GM and GSD) of the predictive distribution (median values of the posteriori distributions of GM and GSD) obtained for the CR of X applying the Bayesian method described in Section 3.1 and using different analogue species to specify the joint conjugate prior.

Selected Analogue species	N	GM	GM 2.50%	GM 97.50%	GSD	GSD 2.50%	GSD 97.50%	Mean	95%
A	5	5.6E-02	*1.3E-02*	2.4E-01	5.4E+00	2.8E+00	3.2E+01	2.3E-01	8.8E-01
B	20	7.7E-02	*4.6E-02*	1.3E-01	3.3E+00	2.5E+00	5.5E+00	1.6E-01	5.6E-01
C	5	3.4E-02	*1.3E-02*	9.3E-02	3.1E+00	2.0E+00	1.1E+01	6.5E-02	2.2E-01
D	51	9.2E-02	*6.0E-02*	1.4E-01	4.7E+00	3.6E+00	6.7E+00	3.0E-01	1.2E+00
E	11	3.5E-02	*2.1E-02*	5.7E-02	2.3E+00	1.8E+00	3.9E+00	4.9E-02	1.4E-01
F	17	1.8E-01	*7.9E-02*	4.2E-01	5.9E+00	3.7E+00	1.3E+01	8.7E-01	3.3E+00
G	2	4.9E-02	*1.9E-03*	1.3E+00	1.1E+01	3.3E+00	2.5E+03	8.7E-01	2.5E+00

Table 3. Median and 95 % interval of the posteriori GM obtained for each species of bats after updating with the Hierarchical model. The posteriori GSD has the same values (Median of 3.7, lower 2.5 percentile of 3.1 and upper 97.5 percentile of 4.5) for all species since it was assumed that they have the same variance. The Mean and 95-th percentile are calculated analytically from the median values of GM and GSD

Species	Posterior GM			Mean	95%
	Median	2.5 %	97.5 %		
Brown long-eared	1.0E-01	4.0E-02	2.6E-01	2.4E-01	1.3E+00
Common Noctule	9.4E-02	5.5E-02	1.6E-01	2.2E-01	1.2E+00
Daubenton	4.0E-02	6.7E-03	1.3E-01	9.5E-02	5.2E-01
Lesser Noctule	7.0E-02	2.5E-02	1.6E-01	1.6E-01	9.0E-01
Nathusius' Pipistrelle	1.0E-01	7.0E-02	1.4E-01	2.3E-01	1.3E+00
Parti-coloured	5.3E-02	2.5E-02	1.1E-01	1.2E-01	6.9E-01
Serotine	2.1E-01	1.0E-01	3.9E-01	4.8E-01	2.7E+00
Soprano Pipistrelle	1.2E-01	4.1E-02	5.2E-01	2.9E-01	1.6E+00
Predicted species	9.0E-02	1.0E-02	6.7E-01	2.1E-01	1.2E-01

Table 4. The hyper parameters and the within unit variance estimated from the hierarchical model. Median, mean and upper and lower percentile.

Parameter		Median	2.5%	97.5%	Mean
Population mean	μ	-2.4	-3.27	-1.7	-2.4
Population variance	τ^2	0.55	0.031	3.8	0.89
Within variance	σ^2	1.7	1.31	2.3	1.7

Table 5. Median and 95 % interval of the posteriori GM and GSD of the CR of X obtained after updating using a semi-conjugate prior (method in Section 3.2) specified using pooled data of all species, except for X, in Table 1.

Distribution	**Statistics**	**Value**
Posterior GM	Median	0.030
	2.5 %	0.0051
	97.5 %	0.47
Posterior GSD	Median	7.0
	2.5 %	1.9
	97.5 %	204493
Posterior Mean		0.20
Posterior 95 %		0.72
Prior GM		0.10
Prior GSD		3.8

Appendix

It is here assumed that the probability model of measurements $y'_i, i = 1, \dots, n$ is log normal:

$$y_i \sim N(\mu, \sigma^2), \quad i = 1, \dots, n \qquad \text{A 1}$$

Where N denotes the normal distribution and y_i is the natural logarithm of the original log normal measurements, $y_i = \ln y_i'$ with unknown mean μ and variance σ^2. For estimates of μ and σ^2, the geometric mean and variance of the log normal variable are calculated as $GM = e^{\mu}$ and $GSD = e^{\sigma}$.

The sample mean and variance of the logarithmic values are:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (y_i - \bar{y})^2$$

Other estimates of the mean and variance are the combined mean and variance which pool data from K multiple data sets (Hosseini et al., 2008):

$$\mu_{comb} = \frac{1}{n_{comb}}\sum_{k=1}^{K} n_k \bar{y}_k \qquad \text{A 2}$$

$$\sigma^2_{comb} = \frac{1}{n_{comb}-1}\sum_{k=1}^{K}\left((n_k - 1)s_k^2 + n_k(\bar{y}_k - \mu_{comb})^2\right) \qquad \text{A 3}$$

$$n_{comb} = \sum_{k=1}^{K} n_k$$

1. Joint conjugate prior distribution for the mean and variance

Prior distributions

If the prior information consists of a fixed number of measurements from the normal measurement model with variance σ^2, a set of prior distributions for the mean and variance of the normal measurement model (A 1) can be defined as (Gelman et al., 2004):

$$\mu|\sigma^2 \sim Normal(\mu_0, \frac{\sigma^2}{n_0})$$

$$\sigma^2 \sim Inv - \chi^2(\nu_0, \sigma_0^2)$$

The prior distribution of the mean is expressed in terms of the unknown variance σ^2 and is therefore called prior *conditioned* on σ^2. The prior specification can also be written as a joint prior distribution of μ and σ^2 using the relation $p(\mu, \sigma^2) = p(\mu|\sigma^2)p(\sigma^2)$ and is often denoted the Normal-Inverse Chi squared distribution, $N - Inv - \chi^2(\mu_0, \sigma_0^2/n_0; \nu_0, \sigma_0^2)$.

Prior information about the mean is defined by setting μ_0 equal to the sample mean and n_0 to the sample size. Prior information of the variance are included by setting σ_0^2 to the sample variance and prior degrees of freedom ν_0 to $n_0 - 1$.

$Inv-\chi^2(\nu,\sigma^2)$ denotes the Scaled Inverse Chi Square distribution with ν degrees of freedom and scale parameter σ^2. This distribution is derived from the standard $\chi^2(\nu)$ (Chi-Square) distribution: A sample from Inv-$\chi^2(\nu,\sigma^2)$ $Inv-\chi^2(\nu,\sigma^2)$ is obtained as$Y = \nu\sigma^2/X$ $\nu\sigma^2/X$ where X is a sample from $\chi^2(\nu)$. Its probability density function is

$$p(\theta) = \frac{(\nu/2)^{\frac{\nu}{2}}}{\Gamma(\nu/2)} s^{\nu}\, \theta^{-(\nu/2+1)} e^{\nu s^2/(2\theta)}$$

Posterior distributions

The joint (two-dimensional) posterior distribution is obtained by applying Bayes' theorem and multiplying the prior probability density functions with the joint likelihood of the observed data points:

$$p(\mu,\sigma^2|y) \propto N(\mu|\mu_0,\sigma^2/n_0)Inv-\chi^2(\sigma^2|\nu_0,\sigma_0^2)\prod\prod_{i=1}^{n} N(y_i|\mu,\sigma^2)$$

which can be shown (Gelman et al, 2004) to also be proportional to a $N-Inv-\chi^2$ probability density function. The Normal-Inverse Chi squared distribution is a joint conjugate prior distribution to the normal measurement model (A 1) for μ and σ^2. A conjugate prior will, when combined with data using Bayes' theorem, result in a *joint* posterior $p(\mu,\sigma^2|y)$ that is on the same functional form as joint prior. This means, that with the choice of prior distributions (**Feil! Fant ikke referansekilden.**) we expect the joint posterior to also attain the same form as $p(\mu|\sigma^2)$ and $p(\sigma^2)$ but with new parameters.

The posterior distributions can be shown to be (Gelman et al., 2004):

$$\mu|\sigma^2, y \sim N(\mu_n, \frac{\sigma^2}{n_0 + n}) \quad \text{A 3}$$

$$\sigma^2|y \sim Inv - \chi^2(\nu_n, \sigma_n^2) \quad \text{A 4}$$

where parameters with subscript *n* reflect the combined data:

$$\mu_n = \frac{n_0\mu_0 + n\bar{y}}{n_0 + n}$$

$$\sigma_n^2 = \frac{1}{\nu_n}(\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n_0 n}{n_0 + n}(\bar{y} - \mu_0)^2)$$

$$\nu_n = \nu_0 + n$$

The posterior mean, μ_n, is a weighted combination of the sample mean and the prior mean, with the weights being the individual samples sizes**.** The posterior variance is expressed as sum of the i) variation within prior, ii) variation within the data and iii) the weighted squared difference between the data mean and prior mean. The posterior variance is thus adjusted for the uncertainty provided by the difference of the prior and data means. The prior degrees of freedom ν_0 could be set to $n_0 - 1$ since the prior variance is estimated from a sample.

Making inferences

The expressions for the posterior mean μ_n and variance σ_n^2 in A5 can be shown (Gelman et al., 2004) to be equal to the expressions for the combined mean and variance (A 2) of two data sets. Therefore, point estimates of the posterior mean and variance can be obtained from μ_n and σ_n^2 directly.

Full Bayesian inferences from the joint posterior distributions are obtained by repeatedly sampling from the posterior distributions, first by drawing a sample of σ^2 from (A 4) and then sample μ from (A 3) using the previously drawn value of σ^2. The obtained posterior samples of the parameters can then be summarized with statistics of interest, such as the mean or median. The posterior uncertainty of each parameter can be reported with $100 \cdot (1-\alpha)\%$ probability intervals, constructed from the $\alpha/2$ and $1-\alpha/2$ percentiles of posterior samples. A posterior probability interval calculated for an unknown quantity of interest can be directly regarded as having a high probability of containing the unknown quantity (Gelman et al., 2004).

2. Independent prior distributions for the mean and variance

Prior distributions

A set of independent prior distributions for the mean and variance are:

$$\mu \sim N(\mu_0, \tau_0^2) \qquad \text{A 5}$$

$$\sigma^2 \sim Inv-\chi^2(\nu_0, \sigma_0^2) \qquad \text{A 6}$$

The prior distribution of the mean is a normal distribution with specified mean μ_0 and variance τ_0^2. The prior distribution of the variance is Scaled-Inverse-Chi-squared. Its parameters specify the prior estimate σ_0^2 of the variance and its degrees of freedom ν_0. A non-informative prior of the variance is reflected by setting $\nu_0 = 0$, which is equal to using the prior density function $p(\sigma^2) \propto \frac{1}{\sigma^2}$ (Gelman et al., 2004).

Posterior distributions

The joint posterior distribution is derived by multiplying the independent prior distributions with the data likelihood function:

$$p(\mu, \sigma^2|y) \propto p(\mu|\mu_0, \tau_0^2)p(\sigma^2|\nu_0, \sigma_0^2)\prod_{i=1}^{n} p(y_i|\mu, \sigma^2) \propto$$

$$\propto N(\mu|\mu_0, \tau_0^2)Inv-\chi^2(\sigma^2|\nu_0, \sigma_0^2)\prod_{i=1}^{n} N(y_i|\mu, \sigma^2) \qquad \text{A 7}$$

Here the notations N and $Inv-\chi^2$ refers to the probability density functions for the normal and Scaled-Inverse-Chi-Squared distribution. The joint posterior distribution does not take the form of any standard distribution function and cannot be sampled from directly. The conditional posterior distributions $p(\mu|\sigma^2, y)$ and $p(\sigma^2|\mu, y)$ however are still on the same form as the *conditional* prior distributions (A 5 and A 6). Under the normal measurement model (A 1) these prior distributions are therefore sometimes called semi-conjugate. The conditional posterior distributions can then be used to obtain samples from the joint posterior.

The conditional posterior distribution for the mean μ is derived by keeping only terms in the joint posterior (A 8) that depend on μ (treating terms that do not depend on μ as constants). The conditional posterior is then the normal prior distribution function of the mean multiplied by the likelihood function:

$$p(\mu|\sigma^2, y) \propto N(\mu|\mu_0, \tau_0^2) \prod_{i=1}^{n} N(y_i|\mu, \sigma^2)$$

The resulting expression can be recognized as being proportional to a normal distribution with new parameters (Gelman et al., 2004):

$$\mu|\sigma^2, y \sim N(\mu_n, \tau_n^2) \qquad \text{A 8}$$

$$\mu_n = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}$$

$$\tau_n^2 = \frac{1}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}$$

The updated parameter of the posterior mean μ_n reflects the relative amount of information available for the case of interest with observed mean $\bar{y}$ and for the selected prior μ_0.

The conditional posterior distribution of the variance σ^2 is derived by multiplying the distribution function of the prior with the likelihood function as

$$p(\sigma^2|\mu,y) \propto Inv-\chi^2(\sigma^2|\nu_0,\sigma_0^2)\prod_{i=1}^{n} N(y_i|\mu,\sigma^2)$$

The resulting expressions can be recognized as being proportional to the $Inv-\chi^2$ distributions function with updated parameters (Gelman et al., 2004):

$$\sigma^2|\mu,y \sim Inv-\chi^2(\nu_n,\sigma_n^2) \qquad \text{A 9}$$

$$\sigma_n^2 = \frac{\nu_0\sigma_0^2 + (n-1)s^2 + n(\bar{y}-\mu)^2}{\nu_0 + n}$$

$$\nu_n = \nu_0 + n$$

The parameter σ_n^2 is a weighted combination of the sample variance s^2, the estimate of the prior variance σ_0^2 and the squared distance between the data and posterior mean. The weights are the number of measurements n and the prior degrees of freedom ν_0 for the variance.

Making inferences

A set of samples from the joint posterior distribution of the parameters (μ,σ^2) can be obtained from the conditional posterior distributions (A 8) and (A 9), first by drawing a sample of σ^2 from A 9 and then of μ (from A 8) in using the previously drawn value of σ^2. When drawing the first sample of σ^2, a crude starting value can be used for μ, such as the sample mean. The posterior distribution of each of the parameter is summarized with statistics of interest (such

as mean or medians) and the posterior uncertainty is reported with posterior probability intervals, calculated from an lower and upper percentile of the obtained posterior samples. To diminish the effect of the arbitrary start values, the first obtained sample must be discarded before drawing inferences from the posterior distributions. This procedure of sampling from a multidimensional distribution using the full conditional distributions is called a Gibbs Sampling algorithm (Gelman et al., 2004).

3. Prior distribution implicitly derived from data of several similar units

Prior distributions

The prior distribution of the means of J units (e.g. J species or elements) is normal with mean μ and variance τ^2:

$$\mu_j|\tau^2 \sim N(\mu, \tau^2), \quad j=1,\ldots,J \qquad \text{A 10}$$

Here, the parameters of the prior distributions μ, τ^2 (also called hyper-parameters), are also unknowns, to be estimated from the data of similar units. The mean μ is often referred to as the population mean. The variance τ^2 is often referred to as the population variance or the variance between units. The hyper-parameters are here assigned the non-informative prior distributions $p(\mu) \propto 1$ and $p(\tau^2) \propto 1/\tau$ (Gelman et al., 2004), allowing them to be estimated from data of the J units without including any explicit prior knowledge about them.

The within-unit variance is here assumed equal among units (denoted σ^2) and assigned a non-informative to let it be estimated from the available data, equal to using the prior $p(\sigma^2) \propto 1/\sigma^2$ (Gelman et al., 2004).

Posterior distributions

The joint posterior distribution of all J+3 parameters (the J units' means, the common variance within units and the two hyper-parameters) is expressed by multiplying the prior distributions of the hyper-parameters, the prior distribution of the within unit variance σ^2, the joint prior distribution of the unit means μ_j and the joint likelihood function of all observations from the J units:

$$p(\mu_1, \dots, \mu_J, \sigma^2, \mu, \tau^2 | y) \propto$$
$$\propto \tau^{-1}\sigma^{-2} \prod_{j=1}^{J} N(\mu_j|\mu, \tau^2) \prod_{j=1}^{J} \prod_{i=1}^{n} N(y_{ij}|\mu_j, \sigma^2) \qquad \text{A 11}$$

To obtain samples from the multidimensional joint posterior distributions (A 11), a Gibbs Sampler is implemented by sampling iteratively from the full conditional posterior distributions. The full conditional posterior distribution of a parameter θ is one-dimensional and is the posterior distribution of θ conditioned on all the other parameters in the model. The full conditional distributions are derived from the joint posterior (A 11) by treating terms not depending on θ as constants.

The full conditional posterior distributions of the mean of unit j is derived by keeping terms in the joint posterior distributions that depends on μ_j (and treating functions not depending on μ_j as constant):

$$p(\mu_j|\sigma^2,\mu,\tau^2,y_j) \propto N(\mu_j|\mu,\tau^2)\prod\prod_{i=1}^{n} N(y_{ij}|\mu_j,\sigma^2)$$

The expression can be shown (Gelman et al., 2004) to be proportional to a normal distribution function with parameters depending on both the data and the prior:

$$\mu_j|\sigma^2,y \sim N(\mu_n,\tau_n^2) \qquad \text{A 12}$$

$$\hat{\mu}_n = \frac{\frac{1}{\tau^2}\mu + \frac{n_j}{\sigma^2}\bar{y}_j}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}} \qquad \text{A 14}$$

$$\tau_n^2 = \frac{1}{\frac{1}{\tau^2}+\frac{n_j}{\sigma^2}}$$

The amount of partial pooling from the unit mean $\bar{y}_j$ to the population mean μ is determined by population variance τ^2 and the within unit variance scaled by the number of measurements n_j, with complete pooling as special case when the population variance is small compared to the within unit variation (σ^2/n_j). The expression (A 13) is thus similar to the posterior mean (A 8) when using independent prior distributions, but with unspecified parameters of the prior.

Under a non-informative prior, the full conditional posterior for the common variance σ^2 is obtained by treating factors not depending on σ^2 as constants in the joint posterior distribution. The conditional posterior is then obtained by multiplying the non-informative prior of the variance, $1/\sigma^2$ with the joint likelihood function of the observations of all units:

$$p(\sigma^2|\mu_1,\dots,\mu_J,y) \propto \sigma^{-2}\prod_{j=1}^{J}\prod_{i=1}^{n_j} N(y_{ij}|\mu_j,\sigma^2)$$

The expression can be shown to be proportional to a $Inv-\chi^2$ distribution (Gelman et al, 2004):

$$\sigma^2|\mu_1,\dots,\mu_J,y \sim Inv-\chi^2(\nu_n,\sigma_n^2) \qquad \text{A 14}$$

$$\nu_n = \sum_{j=1}^{J} n_j$$

$$\sigma_n^2 = \frac{1}{\nu_n}\sum_{j=1}^{J}\left((n_j-1)s_j^2 + n_j\left(\bar{y}_j-\mu_j\right)^2\right)$$

The posterior variance is expressed as the pooled variance of all units adjusted for the discrepancy of the observed means $\bar{y}_j$ and updated means μ_j (the squared difference in the last term of σ_n^2).

The full conditional posterior of the population mean μ is obtained by treating factors not depending on μ in the joint posterior as constants. It is the result of multiplying the prior of μ (which is proportional to 1) with the joint probability of the J unit means:

$$p(\mu|\mu_1,\dots,\mu_J,\sigma^2,\tau^2) \propto \prod_{j=1}^{J} N(\mu_j|\mu,\tau^2)$$

The expression can be shown to be proportional to the normal distribution (Gelman et al., 2004):

$$\mu|\mu_1,\dots\mu_J,\tau^2,\sigma^2,y \sim N(\hat{\mu},\frac{\tau^2}{J})$$

$$\hat{\mu} = \frac{1}{J}\sum_{j=1}^{J}\mu_j \qquad \text{A 15}$$

That is, the posterior population mean is expressed in terms of the average of the units' posterior means and variance given by the variance of posterior means.

The full conditional posterior distribution of the population variance τ^2 is obtained by treating factors in the joint posterior not depending on τ^2 as constants, resulting in the multiplication of the prior τ^{-1} with the joint probability of the J unit means:

$$p(\tau^2|\mu_1,\dots,\mu_J,\sigma^2,\mu,y) \propto \tau^{-1}\prod_{j=1}^{J} N(\mu_j|\mu,\tau^2)$$

The expression is proportional to a $Inv-\chi^2$ distribution (Gelman et al., 2004):

$$\tau^2|\mu_1,\dots,\mu_J,y \sim Inv-\chi^2(J-1,\tilde{\tau}^2)$$

$$\tilde{\tau}^2=\frac{1}{J-1}\sum_{j=1}^{J}(\mu_j-\mu)^2 \qquad \text{A 16}$$

Making inferences

Samples from the joint posterior distribution (A 11) of all parameters are obtained by iteratively sampling from each of the conditional posterior distributions (A 12 to A 16). In each iteration $k=1,\dots,K$, the Gibbs Sampling procedure samples one value for each parameter of the model, conditioning on the latest values drawn from of the other parameters (either from the last iteration k-1 or the current iteration k). Each iteration in the Gibbs Sampler is here described as sampling from the parameters in four steps:

1) Individual unit means (from A 12): $\mu_{j,k} \sim p\left(\mu_j \left| \mu_{j,k-1}, \sigma^2_{k-1}, \mu_{k-1}, \tau^2_{k-1}, y_j\right.\right), j=1,\dots,J$
2) The variance within units (from A 14): $\sigma^2_k \sim p(\sigma^2|\mu_{1,k},\dots,\mu_{J,k},y)$
3) Population mean (from A 15): $\mu_k \sim p(\mu|\mu_{1,k},\dots,\mu_{J,k},\tau^2_{k-1})$
4) Population variance (from A 16): $\tau^2_k \sim p\left(\tau^2 \left| \mu_{1,k},\dots,\mu_{J,k},\mu_k\right.\right)$

In the first iteration (k=1), parameters that have index k-1 must be assigned start values which can be derived as crude estimate of the sample means, the pooled sample variance and the population mean and variances. To diminish the impact of the somewhat arbitrary start values,

the first simulated samples should be removed before inference (Gelman et al., 2004). The posterior distribution of each of the parameter is summarized with statistics of interest (such as mean or median) and the posterior uncertainty is reported with posterior probability intervals, calculated from an lower and upper percentile of the obtained posterior samples.

A mean from a new case (e.g a species not yet observed but assumed to belong to the same population as those included in the hierarchical model) can be predicted by sampling from the posterior population distribution as $\mu_{j,k}^{*} \sim N(\mu_k, \tau_k^2)$, using a set of draws μ_k, τ_k^2 from the posterior distribution of the hyper-parameters. The obtained samples are then used to draw inferences for the new case.

Convergence of the Gibbs Sampler

The arbitrary start values of the first iteration will influence the distribution of the collected samples and make the first samples not truly representative to the true posterior distributions. To check the convergence of the obtained samples, we have adopted the technique of (Gelman and Rubin, 1992) by running the sampler multiple times with different and dispersed start values. The Gelman-Rubin convergence statistic (Gelman and Rubin, 1992; Gelman et al., 2004) is then used to quantify the degree to which the posterior variance can be reduced by running the sampler for more iterations. For all simulations with the Gibbs Sampler, we run three independent simulations with different and randomly dispersed start values (adding a random component to the crude estimates) for 10 000 iterations each. We required the convergence statistic to be below 1.001 for all parameters which suggests that the probability interval of the parameters cannot be reduced with more than 0.1% if the simulations were to continue (Gelman and Rubin, 1992). The first five hundred samples of each chain were then

discarded before drawing inferences from the posterior distributions to diminish the effect of the arbitrary start values.