Geological prior information, and its applications to geoscientific problems

RACHEL WOOD[1,2] & ANDREW CURTIS[1,3]

[1] *Schlumberger Cambridge Research, High Cross, Madingley Road, Cambridge CB3 0EL, UK (e-mail: rwood@cambridge.oilfield.slb.com)*

[2] *Dept. of Earth Sciences, University of Cambridge, Downing Street, Cambridge, CB2 3EQ, UK*

[3] *Grant Institute of Earth Science, School of GeoSciences, University of Edinburgh, West Mains Rd., Edinburgh EH9 3JW, UK*

Abstract. Geological information can be used to solve many practical and theoretical problems, both within and outside of the discipline of geology. These include analysis of ground stability, predicting sub-surface water or hydrocarbon reserves, and assessment of risk due to natural hazards. In many cases, geological information is provided as an *a priori* component of the solution (that is, information that existed before the solution was formed and which is incorporated into the solution). Such information is termed 'geological prior information'.

The key to the successful solution of such problems is to use only *reliable* geological information. In turn this requires that (a) multiple geological experts are consulted and any conflicting views reconciled, (b) all prior information includes measures of confidence or uncertainty (without which their reliability and worth is unknown), and (c) as much information as possible is quantitative, and qualitative information or assumptions are clearly defined so that uncertainty or risk in the final result can be evaluated. This paper discusses each of these components, and proposes a probabilistic framework for the use and understanding of prior information

We demonstrate the methodology implicit within this framework with an example: this shows how prior information about typical sequence stacking patterns allows aspects of 2-D platform architecture to be derived from 1-D geological data alone, such as that obtained from an outcrop section or vertical well. This example establishes how the extraction of quantitative, multi-dimensional, geological interpretations is possible using lower dimensional data. The final probabilistic description of the multi-dimensional architecture could then be used as prior information sequentially for a subsequent problem using exactly the same method.

Introduction

For decades, specific geological information has been transferred to other domains, helping solve applied and theoretical problems. These include, for example:

- information about typical geological architectures in a particular environment allows sub-surface properties to be estimated away from drilled observation wells in order to assess hydrocarbon or water reservoir potential (e.g. Mukerj *et al.* 2001; Schon 2004);
- assessments of predicted geohazzard risks contributes to the pricing of life and property insurance (e.g. Rosenbaum & Culshaw 2003)
- the distribution of ancient reef corals, combined with the relationship between coral growth and sea-level, allows prediction of both past sea-level oscillations, and how future sea-level change might affect the distribution of modern reefs (e.g. Potts 1984; Budd *et al.* 1998);
- historical knowledge of slope stability in different geological environments allows proposed building projects to include appropriate remedial action against such risk (e.g. Selby 1982), and
- estimates of regional tectonic stability contribute to allowing toxic waste to be stored underground with minimal risk of future leakage (e.g. Pojasek 1980).

For each of these problems, geological information is not only provided as an *a priori* component of the solutions, but is central to their creation. This information pre-existed to the formation of the solution, and so in this context is termed 'geological prior information' henceforth, *GPI*.

The special volume from the Geological Society of London, of which this paper is a part, is devoted to understanding the use of geological prior information in order to solve problems in both geological and other domains. As such, it spans research in several aspects of GPI: how GPI can be captured, quantified, ascribed an associated uncertainty or reliability, and then how this might be used to provide solutions to other problems. The purpose of research into the creation and use of GPI is therefore to make geology more accurate, useful and transferable to scientists and problem solvers.

In this paper, we begin by constructing a Bayesian, probabilistic framework that precisely defines GPI. We then define the various components of information that must be either assembled or assumed in order to solve problems using GPI, and demonstrate how uncertainty in such assembled information propagates to create uncertainty in the solutions found. This framework also allows a useful categorisation of methods that use GPI to solve geoscientific problems: we demonstrate this by placing the papers of this volume into this framework and stating the principal ways in which each contributes to the field. We then present an example of the use of GPI: in this case, how it might be used to extract and extend geological information away from a single point in space at which data was collected. This example serves to illustrate definitions given in the probablistic framework, and highlights some of the issues and difficulties involved in using GPI.

A framework for using and analysing geological prior information

In all that follows we use *P(Q)* as a general notation for the probability of an event or hypothesis *Q* being true (or the value of the probability density function at *Q* if the event space is continuous). We also make use of conditional probability notation $P(Q|R)$, meaning the probability that *Q* is true given that we know *R* is true, for some event or hypothesis *R*. Finally, we will make use of Bayes Rule which is obeyed by all probability distributions. For any events or hypotheses *Q* and *R* it may be stated:

$$P(Q \mid R) = \frac{P(R \mid Q)P(Q)}{P(R)} . \qquad \text{……………………. (1)}$$

Now, say we would like to answer a geoscientific question upon which geological information might have some bearing, and assume further that the question can be posed in such a way that its answer is given by constraining some model vector **m**. Model **m** might represent a predicted sea level curve, a set of charges for insurance against geohazzards, or the distribution of fluids in a sub-surface reservoir, for instance. Examples of the types of pertinent background knowledge might include: knowledge of regional geologic, diagenetic, tectonic or climatic history; inferences from previously acquired data; typical problems experienced when acquiring data; expected data uncertainties; generally accepted geological theory which might be assumed in a geological model, and existing methods that may be used to tackle different types of problems. We represent all such background knowledge by symbol *B*.

We will assume that *B* can be decomposed into geological knowledge *G* and other knowledge G', and further that the geological knowledge can be decomposed into quantitative and non-quantitative components, G_q and $G_{q'}$ respectively. This results in a description of all background knowledge by the partition $B = [G_q\ ,\ G_{q'}\ ,\ G']$.

Say a new data set **d** is acquired which was designed to place more focussed constraints on model **m**. Bayes Rule allows constraints on **m** both from the new data **d** and from pre-existing background information *B* to be quantified. This is useful because the relationships between **m**, **d** and *B* are complicated in which case it is not easy to calculate or weight the relative constraints offered by **d** and *B* in other ways. Since all data and background information will be uncertain to some extent, so our knowledge of **m** will always be uncertain. One way to express available information about **m** is to use a probability distribution. Substituting **m**, **d** and *B* appropriately into equation (1) we obtain:

$$P(\mathbf{m} \mid \mathbf{d}, B) = \frac{P(\mathbf{d} \mid \mathbf{m}, B)}{P(\mathbf{d} \mid B)} P(\mathbf{m} \mid B) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

The totality of information about the model given data sources **d** and *B* is called the *posterior distribution* and is described by the probability distribution on the left of equation (2). This can be calculated for any particular model **m** by the product of two terms on the right of equation (2). The first is called the *relative likelihood* term, and is the ratio of the probability of measuring data **d** if model **m** was true (numerator) to the probability when no knowledge about **m** is available (denominator). Notice that both the numerator and denominator must therefore implicitly incorporate information about the uncertainty in the measured data **d**. The second term on the right is called the *prior distribution* of model **m**, and describes all information about **m** that existed prior to data **d** having been acquired (since data **d** does not feature in this term). Hence, this term includes all information about **m** from background information *B* alone.

At this point it helps to have a particular example in mind, and the following example will be illustrated in detail later in this paper. Consider the situation in which we would like to constrain the stacking pattern of a parasequence set observed on a laterally constrained but vertically extensive outcrop exposure surface in some particular location (assuming that the exposure is too restricted to reveal the stacking patterns directly). Let model **m** represent the various possible stacking patterns of the observed parasequence set. Hence, **m** may represent one of the following patterns: regressive (R), progradational (P), aggradational (A), or retrogradational (G). Pertinent background information *B* might include: regional geological history and age of the sequence, approximate location of the parasequence set within the overall succession, and the sedimentary composition. The prior information term $P(\mathbf{m}|B)$ then represents all possible constraints on the stacking pattern given only this pre-existing background information.

Now say the relative vertical thicknesses of the parasequence set observed on the vertical outcrop section were measured to provide a data set **d**. In principle, this data set contains some information about the stacking pattern that would be observed laterally (a particular, precise relationship is given later). The numerator of the relative likelihood term $P(\mathbf{d}|\mathbf{m},B)$ is the probability that we could have recorded data set **d** if we knew that a particular value of model **m** was true (e.g., if we know that the stacking pattern was regressive) and given background information *B*. The denominator $P(\mathbf{d}|B)$ is the probability that data **d** could have been recorded with no knowledge of the true stacking pattern but still given background information *B*, and is a normalising term that ensures that the integral of the posterior distribution equals unity.

To give a numerical demonstration, say we know from background information *B* that the age of deposition occurred during a period of general sea level rise. Then *a priori* we may wish to assign a low probability that regression G occurred (i.e., **m**=‘G’ in reality) since this is often associated with a relative sea level fall, but equal probability that progradation P, aggradation A or retrogradation R occurred. Hence, we might define the following prior distribution:

$$P(\mathbf{m}=\text{‘R’} \mid B) = \frac{1}{10} \quad \text{and} \quad P(\mathbf{m}=\text{‘P’} \mid B) = P(\mathbf{m}=\text{‘A’} \mid B) = P(\mathbf{m}=\text{‘G’} \mid B) = \frac{3}{10}.$$

Say that the probabilities of recording data **d** given each of the possible values for **m** are:

$$P(\mathbf{d} \mid \mathbf{m}=\text{‘R’},B) = P(\mathbf{d} \mid \mathbf{m}=\text{‘P’},B) = P(\mathbf{d} \mid \mathbf{m}=\text{‘G’},B) = \frac{1}{5} \quad \text{and}$$

$$P(\mathbf{d} \mid \mathbf{m}=\text{‘A’},B) = \frac{2}{5}.$$

That is, the probability of the data having been recorded in an aggradational setting is twice that in all of the other settings.

Thus far we have completely defined the likelihood in the numerator, and the prior information term on the right of equation (2). It is now possible to work out the other two terms in this equation: since the left of equation (2) is a probability distribution it must integrate to unity (that is, the sum of the probabilities for each of the possible values of model **m** - R, P, A and G - must be one). Hence, the denominator on the right must assume a value that normalises the right side of the equation similarly. This gives (with terms in order **m** = R, P, A, then G):

$$P(\mathbf{d}|B) = P(\mathbf{d} \mid \mathbf{m}=\text{'R'},B)\, P(\mathbf{m}=\text{'R'}|B) + P(\mathbf{d} \mid \mathbf{m}=\text{'P'},B)\, P(\mathbf{m}=\text{'P'}|B) + P(\mathbf{d} \mid \mathbf{m}=\text{'A'},B)\, P(\mathbf{m}=\text{'A'}|B) + P(\mathbf{d} \mid \mathbf{m}=\text{'G'},B)\, P(\mathbf{m}=\text{'G'}|B)$$

$$= \frac{1}{5}\cdot\frac{1}{10} + \frac{1}{5}\cdot\frac{3}{10} + \frac{2}{5}\cdot\frac{3}{10} + \frac{1}{5}\cdot\frac{3}{10}$$

$$= \frac{14}{50}.$$

Finally, we can work out the posterior distribution on the left of equation (2),

$$P(\mathbf{m}=[\text{R,P,A,G}] \mid \mathbf{d},B) = \left[\frac{1}{14}, \frac{3}{14}, \frac{6}{14}, \frac{3}{14}\right],$$

where the numbers in the square brackets on the right are the probabilities of each of the cases for **m** listed in the square brackets on the left. Thus, we show that the probability of aggradation A is still twice that of progradation P or regression G, as was the case above when we considered only the likelihood distribution (data information). However, in the posterior distribution when prior information is also taken into account, the probability of aggradation A is shown to be *three* times that of retrogradation. This is due to the relatively low probability of retrogradation given our prior knowledge that the sequence was deposited in a period of general sea level rise.

Let us now expand the prior distribution in equation (2) using the partition of background information *B* made above, to obtain,

$$P(\mathbf{m} \mid B) = P(\mathbf{m} \mid G_q, G_{q'}, G') \quad \text{……………………. (3)}$$

Since probabilities are inherently quantitative, it is often difficult to use uncertain, qualitative information or hypotheses to constrain their distributions. For instance, in the prior information term (equation (3)), the value of the perceived probability of model **m** being true will be influenced by both quantitative and qualitative geological information G_q and $G_{q'}$, respectively. However, it will be more difficult to assess the exact, quantitative influence of $G_{q'}$ on this probability than the influence of G_q. Returning to the stacking pattern example for illustration, it is reasonably easy to assess the probability of each stacking pattern **m** given quantitative information G_q about the relative proportion of the various stacking patterns observed in other outcrops if we assume the truth of the qualitative hypothesis $G_{q'}$, "that these are good analogues for the outcrop in question". It would be far more difficult to assess, however, exactly how that information on stacking patterns might be enshrined within a prior probability distribution if we also allowed within $G_{q'}$ that there is a certain positive probability that this hypothesis might be false. Although we have only discussed the prior distribution here, such arguments also apply to each probability term in equation (2).

As a result, it is always the case that in practise it will be necessary to make an approximation of the posterior distribution by limiting the uncertainty in some qualitative information and hypotheses in $G_{q'}$ (i.e., by adding extra, assumed information) to give approximation $\hat{G}_{q'}$. This results in an approximation of the background information

$$\hat{B} = [G_q , \hat{G}_{q'} , G'] \qquad \text{.......................} (4)$$

and the following application of Bayes Rule:

$$P(\mathbf{m} \mid \mathbf{d}, B) \;\cong\; P(\mathbf{m} \mid \mathbf{d}, \hat{B}) \;=\; \frac{P(\mathbf{d} \mid \mathbf{m}, \hat{B})}{P(\mathbf{d} \mid \hat{B})} P(\mathbf{m} \mid \hat{B}) \qquad \text{...............} (5)$$

The accuracy with which $P(\mathbf{m} \mid \mathbf{d}, \hat{B})$ approximates $P(\mathbf{m} \mid \mathbf{d}, B)$ depends on the nature of the approximations made in the background information $\hat{B}$, and in particular on the truth of the assumptions added.

To summarise the discussion so far, all practical, calculable posterior distributions are approximations (equation (5)), all are conditional on background information, and in practise it is also always the case that some of this background information must be assumed to be true for expediency rather than because it is known to be so (equation (4)). This implies that all estimates of prior and posterior uncertainty are likely to be optimistic, as they cannot account for the full, true range of uncertainty.

We differentiate here between *static* and *dynamic* quantitative prior information. The former denotes quantitative information about the present day (static) observed or inferred geology; the latter denotes quantitative information about either time-dependent (dynamic) geological processes, or geologies that existed at some point in the past (which implicitly requires the use of a dynamic model).

Static prior information about a succession might comprise statistics of the geological geometries observed in one or more outcrops from formations of analogous settings and origin. In the example given later we use a Uniform prior distribution of typical parasequence architectures of platform carbonate formations, but any other, more accurate prior distribution could also be used.

Dynamic prior information might be derived, for example, from climatological data since climate affects parasequence formation by influencing carbonate production rate, and by affecting sea level oscillations and hence accommodation space. Consequently, some have postulated that fourth-order parasequence geometries are observed to be less ordered in icehouse compared to greenhouse episodes (Lehrmann & Goldhammer 1999). An example of prior information derived from this argument is therefore: given that a specific formation is carbonate (with no other information, e.g., of the geological age), the prior probability that deposition occurred during icehouse, transitional or greenhouse episodes is approximately 0.31, 0.19 or 0.5, respectively (calculated by integrating global carbonate accumulation rates provided by Bosscher & Schlager 1993) over time spanned by each episode (as defined by Sandberg 1985). In the example given later we demonstrate the method using only static information. Note, however, that dynamic information can be treated in exactly the same way.

If prior information comes from multiple independent sources (e.g., static and dynamic) it can be combined as follows: let B_1 represent background information from, for example, static sources as described above, and B_2 represent background information from dynamic sources derived, for example, from a geological forward process model (e.g., Tetzlaff & Priddy 2001; Tetzlaff this volume; Burgess & Emery this volume). Let us assume that the sources of background information and hence B_1 and B_2 are independent. Then, $P(\mathbf{m} \mid B_1)$ denotes static prior information and $P(\mathbf{m} \mid B_2)$ denotes dynamic prior information on model **m**. The correct way to combine these to derive the total resultant prior information $P(\mathbf{m} \mid B)$, where B is the total state of background information [B_1 , B_2], is to use Bayes Rule again :

$$P(\mathbf{m} \mid B) = \frac{P(\mathbf{m} \mid B_1)\, P(\mathbf{m} \mid B_2)}{P(\mathbf{m})}. \qquad \text{……………………. (6)}$$

Here $P(\mathbf{m})$ is called the null probability distribution and is the state of information about **m** when no background information at all is available. The null distribution describes the minimum possible information about the model **m** (Tarantola & Valette 1982). We assume that $P(\mathbf{m})$ is a Uniform distribution and hence is constant with respect to **m**.

Curtis & Lomax (2001) have shown that even weak prior information is often sufficient to make computationally tractable those problems that would otherwise be impossible to solve. Notice the implication of equation (6); that no matter how weakly constrained the dynamic process model used, adding the extra information cannot create more uncertainty than existed using static information alone. The worst possible situation is that the dynamic model provides the minimum possible information so $P(\mathbf{m} \mid B_2) = P(\mathbf{m})$, in which case equation (6) gives $P(\mathbf{m} \mid B) = P(\mathbf{m} \mid B_1)$. Thus, by using our method the addition of even weak dynamic (or additional static) information always improves knowledge of geological architecture, which in turn can render significantly improved computational efficiencies in constraining models from measured data.

A pitfall that might occur in practise when combining information using equation (6) is that incompatible approximations to background knowledge B_1 and B_2 are made. For example, if the dynamic information in B_2 is produced using a geological process model, then that information may be predicated on the assumption that the process model is sufficiently detailed to represent reality. Thus, an approximation $\hat{B}_2$ to background information B_2 would be used (see Curtis & Wood this volume, for further discussion). Such an assumption would not be required in order to incorporate geostatistical, static information in $P(\mathbf{m} \mid B_1)$, which could be measured directly from outcrops in the field; such information would therefore be conditional on different and incompatible approximations, $\hat{B}_1$, to the background information, B_1. Combining information $P(\mathbf{m} \mid \hat{B}_1)$ and $P(\mathbf{m} \mid \hat{B}_2)$ using equation (6) directly would be incorrect since the two distributions are based on incompatible information. Typically this would again result in an optimistic assessment of the uncertainty in $P(\mathbf{m} \mid B)$. An example of one correct way of combining such sources of information would be to use the static, statistical information to constrain the range of process model outputs that should be considered (Curtis & Wood this volume; Tetzlaff this volume; Wijns *et al.* this volume). Thus, both static and dynamic sources of information are combined using the more restrictive assumptions implicit in the conjunction of $\hat{B}_1$ and $\hat{B}_2$ - that the outcrops used to measure static information were valid analogues, *and* that the geological process model assumptions were valid.

Towards a probabilistic framework

Geological Prior Information, can now be defined more precisely as the field devoted to making geological background information G_q and $G_{q'}$ explicitly, or at least practically available, and to using such information to solve geoscientific problems. As such, this field makes geology a useful tool to solve problems in other fields of geoscience.

We may use equations (4) and (5) to create a two-step framework within which most work in the field can be described. Step 1 is defined as the quantification of geological knowledge, and is implicit within the partition on the right of equation (4). As explained above, this is often necessary in order for uncertain, yet correctly calculated inferences to be made. Research in this area quantifies previously qualitative information, and thus, information is moved from $G_{q'}$ to G_q. Step 2 is defined to be the use of such information to solve other geoscientific problems, and is associated with the use of equation (5) for Bayesian inference (or with the use of some other system of inference).

Table 1 summarises how each of the papers within this volume contributes to these two steps within this probabilistic framework, and whether a paper presents a new method, application, or general discussion. We also differentiate between papers that contribute to either *static* or *dynamic* quantitative prior information within Step 1. The papers in this volume are therefore ordered roughly so as to begin with two introduction and discussion papers, of which this is one, then to progress logically through the various stages of Step 1 and then Step 2.

Notice that only two technical papers (Pschenichny this volume; Stephenson *et al.* this volume) address the issue of calculating distribution $P(\mathbf{d} \mid B)$ in the denominator of equation (2), and of these only Stephenson et al. (this volume) addresses this directly. In most applications, this term, called the *evidence* or *marginal likelihood*, serves merely to normalise the posterior distribution and is ignored in the calculation. This is usually because relative rather than absolute values of the posterior distribution are perceived to be sufficient. However, Malinverno (2000) and Malinverno & Briggs (2004) show how the value of this term can be used to determine the complexity of model **m** that is justified by the data available. In other words, this term quantifies the 'evidence' for any particular model parameterisation, and allows the evidence for different possible model parameterisations to be compared.

Example: Extraction of 2-D Stratigraphic Information from 1-D Data

Parasequences forming over ~ 100 k.y. periods, nested within sequences that form over ~ 1-3 myr, are usually considered to be the fundamental depositional units of marine platform systems, especially in carbonate successions. Yet the shape, size, and stacking pattern of parasequences and sequences within any given basin, as well as details of their internal architectures and facies distribution, are usually poorly known in two- or three-dimensions. This is due either to the sparsity of outcrop, or to lack of sufficient resolution afforded by remote sensing techniques, such as seismic data (a good example of the latter is given in Tinker *et al*. 1999). It has been postulated, however, that under certain conditions the stratal order observed within any given parasequence includes vertical and lateral lithofacies changes that are at least semi-predictable at certain scales (e.g. Lehrmann & Goldhammer 1999), although this is far from accepted (see Wright 1992; Drummond & Wilkinson 1993; Wilkinson *et al.* 1999; Burgess 2001). There are, however, many general principles that govern degrees of predictability in sedimentary successions, and these are enshrined within the discipline of stratigraphy (e.g. Walther's Law).

In the example given here, a method is described that uses GPI gained from sedimentary successions (with some additional, explicit assumptions) to obtain probabilistic information about the stratigraphic architecture of platform successions at the parasequence scale. We demonstrate the method on a simple model of Transgressive System Tracts (TSTs) and show that (i) data describing only the depths of parasequence or sequence boundaries intersected by a 1-D data profile may be sufficient to provide significant spatial constraints on the far-field (2- or 3-D) geology when combined with prior information, and (ii) the probabilistic results obtained are consistent with those derived from independent geological reasoning.

Methodology

The method combines (1) data from a single 1-D profile, that could be derived from either an outcrop stratigraphic section or well (henceforth, a *well*), (2) prior information about the 2-D (or 3-D) stratigraphic architecture that might be expected, and (3) knowledge of the relationships between the 1-D data and the 2-D architecture of which it is a part. Components (1) to (3) include all possible information pertinent to the problem considered – that is, of estimating 2-D parasequence architecture from 1-D data profiles alone.

Using some additional assumptions, information in (1), (2) and (3) individually may be described by independent probability distributions. These distributions are combined using Bayes Rule. The result is a distribution that represents the state of knowledge that includes all possible information about the 2-D parasequence architecture conditional on the assumptions made.

In the example given here we will demonstrate the method on a 2-D representation of shallow platform parasequence architecture formed during a TST, as defined by a simple model (Fig. 1). Although highly simplified, this model may approximate the architecture of at least some real examples, and the probabilistic methodology established can be applied similarly to more complex, 2- or 3-D geologic models, and may also be extended to include a further variety of data types and distributions. That this model *may be* sufficient to explain architectures in question is a qualitative hypothesis, and hence is in $G_{q'}$. The explicit assumption that this *is* the case (i.e., that we may neglect the possibility that the model is insufficient) is made for convenience so that we need not consider the range of all other possible types of models, and results in the approximation $\hat{G}_{q'}$, and hence $\hat{B}$ in equation (4).

Probability distributions

We now describe the required probability distributions (1)-(3) and the various components of the method:

1. Data distribution: Data from the well must include an estimate of the data uncertainty. This is used to define the *data probability distribution* (or simply the *data distribution*) $\rho(\mathbf{d})$ where $\mathbf{d}$ is the vector of data. In our example, $\mathbf{d}$ contains only the observed depths of intersection of a well with parasequence boundaries (Fig. 1); $\rho(\mathbf{d})$ describes uncertainties in those depth observations. Note that $\rho(\mathbf{d})$ is not explicitly included in equation (5), but will be used below to construct the likelihood term.

2. Models and the Prior Distribution: We call the range of all possible geologies the *model space M*. Any particular geology is called a *model*, which we assume can be described by a vector **m**. In our example, **m** comprises the 7 parameters in Fig. 1 (ramp dip is not included in **m**), and although this model is particularly simple, it suffices to illustrate our method. Geological prior information includes only information about models that is independent of the current well data $\rho(\mathbf{d})$. Prior information is described by the *prior (probability) distribution* $P(\mathbf{m} \mid \hat{B})$ that describes the uncertainty in model parameter values given only the available, approximate, background information $\hat{B}$.

3. Model-data relationship and the likelihood function: In theory, the relationships between data $\mathbf{d}$ and models $\mathbf{m}$ can be described by an independent probability distribution (e.g., Tarantola & Valette 1982; Tarantola 1987). In practise, however, the relationship is usually used in the form of a *likelihood function*, a non-normalised probability distribution since normalisation can be difficult. This function describes the relative probability of occurrence of any geological model in the model space given the information contained in the current data set alone, and therefore incorporates the data distribution $\rho(\mathbf{d})$. Below this function will be used to construct the relative likelihood in equation (5).

The likelihood is often calculated as follows: for any geological model $\mathbf{m}$ we assume that we can calculate synthetically data $\mathbf{d}_0 = \mathbf{f}(\mathbf{m})$ that would have been recorded if $\mathbf{m}$ represented the true geology ($\mathbf{d}_0$ is the *expected data*). These are calculated using modelling techniques or assumptions represented here by function **f**. In our example, given particular model parameter values **m** we can define a set of parasequence boundaries, as shown in Figure 1. Function **f(m)** represents the calculation of the intersection depths $\mathbf{d}_0$ between the well and these 'synthetic' parasequences. We calculate the consistency of $\mathbf{d}_0$ with the measured current data distribution $\rho(\mathbf{d})$ by evaluating $\rho(\mathbf{d}_0)$. In turn, variations in $\rho(\mathbf{d}_0) = \rho(\mathbf{f}(\mathbf{m}))$ reflect the relative probability of occurrence of different models **m** given the current data distribution alone, and hence $\rho(\mathbf{f}(\mathbf{m}))$ defines the likelihood function.

Bayesian Inference

We use Bayes Rule in equation (5) as follows: term $P\left(\mathbf{d} \mid \mathbf{m}, \hat{B}\right)$ is the (normalised) likelihood function described above, and $P\left(\mathbf{m} \mid \hat{B}\right)$ is the prior distribution in the model space. In this example we estimate neither the value of $P\left(\mathbf{d} \mid \hat{B}\right)$ (the probability of the measured data occurring at all under the assumptions inherent in our modelling) nor the normalisation factor for the likelihood. That is because these are both constant with respect to model **m**. Hence, in this example (as is commonly the case), Bayes Rule is used to estimate the posterior distribution $P\left(\mathbf{m} \mid \mathbf{d}, \hat{B}\right)$ up to an unknown multiplicative constant. This is sufficient to find the best-fitting model, and the range of possible 2-D models that fit the 1-D data to any desired accuracy.

Application

In this example, we consider the placement of hypothetical vertical sections or wells within a succession of parasequences. Figure 2 shows the intersection of each successive parasequence boundary given at three locations, for both progradational and regressive phases of a cycle. Notice that, due to the similarity of shape and progressive horizontal offset of successive parasequences, a single, vertical section through multiple parasequences is similar to multiple sections through a *single* parasequence at different horizontal locations (Fig. 3). Intuitively then, intersection depths alone from a single, 1-D vertical well contain information about 2-D parasequence shape.

Also, different stacking patterns result in different patterns of depths of intersection between a vertical well and parasequence boundaries. For example, in the progradational situation in Fig. 2A, parasequences become relatively thinner (that is, intersection points become increasingly dense) towards the base of a section taken on the flank (right hand well). In the regressive situation in Fig. 2B, the reverse is true. Hence, vertical intersection depth data also contain information about the parasequence stacking patterns. The proposed methodology allows the degree and nature of these constraints to be explored and quantified.

In this example the specific data distribution used was constructed by placing uniform uncertainties of +/-0.5 m on the synthetic intersection depth data from the right well of the regressive representation (Fig. 2B), representing, for example, uncertainty in the interpretation of parasequence boundaries given available well data. Our static prior information consists of a joint Uniform distribution over the parameters given in Figure 1, and this distribution is defined in Table 2. This includes examples of all possible stacking patterns (regressive, aggradational, progradational, or retrogradational), and the parameters in Table 2 are derived from background information $\hat{B}$ that includes case studies in published literature (Bosscher & Schlager 1993; Lehrmann & Goldhammer 1999).

An example of the type of far field information embedded intrinsically within the prior distribution is constructed as follows: the prior distribution is simply a Uniform distribution with parameters given in Table 2. A Monte Carlo procedure was used to sample the prior distribution $P\left(\mathbf{m} \mid \hat{B}\right)$ of equation (5) across model parameters **m**. Each Monte Carlo sample is a model consisting of seven parameters. For each, the parasequence flank gradient *g*, defined to be flank dz/dx (Fig. 1), is calculated, and a histogram of the gradient values is created. Fig. 5A shows this histogram, which is an approximation to the (non-normalised) prior marginal probability distribution of parasequence flank gradient *g*. Gradient *g* is defined using parameters related to the geological architecture up to hundreds or thousands of metres laterally away from the well since it depends on the lateral extent (length) of the parasequences. This length is not observable at the well. Hence, the non-Uniform nature of the distribution in Figure 5 shows that the prior distribution contains significant far-field geological information that cannot be derived from well data alone.

A Monte Carlo procedure was then used to sample the posterior distribution of equation (5) across model parameters **m** using Bayes Rule. Marginal histograms of these samples over each model parameter are shown in Figure 4. These are (non-normalised) approximations to the posterior probability distribution of each parameters given all prior and data-derived information. These histograms are highly non-uniform despite the fact that both prior and data distributions were Uniform. This indicates that there is significant nonlinearity in the model-data relationship since linear relationships would result in Uniform posterior distributions.

Also, for at least half of the parameters, both mean and the maximum likelihood parameter values proved to be poor indicators of the 'true' model parameter values used in Figure 2B. Hence, there is no easy way to infer a single, best estimate for each parameter: the complete 7-D posterior distribution should be considered in any subsequent interpretation.

The posterior marginal histogram over flank gradient g can be estimated by calculating flank dx/dz for each Monte Carlo sample of the posterior distribution (Fig. 5B). Differences between the prior and posterior histograms represent exactly the information gained by adding the well intersection depth data from the right well in Figure 2B to the prior information (compare Figs 5A and 5B). Clearly, adding only the depth data from a single well has reduced uncertainty in the distribution considerably since the distribution in Figure 2B is narrower than that in Figure 2A (the range of possible gradients has been reduced). Also, the addition of the well data has shifted the maximum likelihood estimate of the gradient close to the value of 0.07 - the correct value for the true architecture as shown in Figure 2B.

Similarly, marginal prior and posterior distributions over the various stacking patterns (retrogradational, progradational, aggradational, or regressive,) can be calculated (Fig. 6). The stacking pattern is intrinsically a global parameter, which is not observable at the well since it is defined by geometries of the entire succession. Vertical depth data almost excludes the possibility that the true pattern was either progradational or aggradational. The result, that either regressive or retrogradational stacking patterns are almost equally likely.

Discussion of the Example

In this example we have combined simple geological prior information (geometrical information about the possible distribution of parasequence or sequence shapes) with 1-D depth data to obtain significantly improved constraints on the parasequence architecture and stacking pattern away from a single outcrop or well (Figs 5 and 6). These constraints were impossible to obtain from either the well data or the prior information alone, and the results in Figure 6 were shown to be consistent with the experience of field geologists.

The method in the example could be used similarly with a combination of more complex data types, and is easily generalised to include multiple outcrop sites or wells. For example, the use of vector data (e.g., dip and azimuth) along the vertical profile would constrain 3-D models in a similar way to the 2-D examples above.

In order to use Bayes Rule to combine information about the data with prior information about the model it was necessary to construct a model relating data and model parameters. The model used here was certainly an over simplification for many geological scenarios, but serves as an example for the Bayesian method. In principle, it is straightforward to use a more realistic and complex model. For example, the introduction of more detailed stratal shape models or parameter prior information could extend the examples to Low Stand and High Stand successions and result in significantly improved far-field interpretations. However, several studies have suggested that stratal patterns may be far less ordered that those predicted by our assumed model (see Wright 1992; Drummond &Wilkinson 1993; Wilkinson *et al.* 1999; Burgess 2001; Burgess *et al.* 2001; Burgess & Emery this volume). This is clearly a contentious topic, and hence in principle our prior information $\hat{B}$ (that the model given in Fig. 1 holds true) might better be regarded as prior assumption, and should properly be assigned a significant uncertainty. This uncertainty would be manifest in the likelihood function, creating more broad distributions. Thus, the example above also illustrates the common phenomenon in practical problems of under-estimating uncertainty in the posterior distribution.

Bayesian inference is explicitly a means to incorporate the prior beliefs of whoever is solving the problem, in a probabilistic, quantitative, and rigorous way. The method forces the scientist to be as thorough and as objective as possible in specifying his/her uncertainties, but also implicitly recognises the fact that all solutions to inference or inverse problems contain subjectivities (that is, assumptions). It is false to assume that one can ever fully overcome this problem in any study. Provided that assumptions are made clear, however, the solution must simply be interpreted with their full acceptance.

While the assumptions made in our example may be overly stringent, the principle of obtaining information away from the well remains proven as long as one has a model relating data and model parameters that can be believed with less than infinite uncertainty. If no such model exists then Bayes Rule simply results in a posterior distribution that equals the prior distribution (i.e., the model and data added nothing to our prior knowledge) as shown above using equation (6). Replacing our simple model with a more realistic one in a particular situation would not, therefore, invalidate the method; it would almost certainly enhance it.

Despite the uncertainties and approximations discussed above, the conclusions of this example are that a) geological prior information and Bayes Rule allows data from a 1-D profile to be propagated laterally both quantitatively and probabilistically, and b) quantitative, multi-dimensional, but uncertain far-field log interpretation is possible from a single 1-D profile of scalar data.

Conclusions

In this contribution, we have constructed a Bayesian, probabilistic framework that clarifies the definition and applications of Geological Prior Information. We introduce Bayesian inference as an explicit means to incorporate prior beliefs in a probabilistic, quantitative, and rigorous way. This method forces a scientist to be as objective as possible in specifying uncertainties, but also implicitly recognises that while all solutions to inverse problems contain assumptions that can never be fully overcome, it is possible to interpret these in the light of their full acceptance.

Various components of information must be assembled or assumed in order to solve problems using Geological Prior Information, and uncertainty in such assembled information will propagate to create uncertainty in the solutions found. We have created a two-step framework with which most work in the field can be described. Step 1 is defined as the quantification of geological knowledge; Step 2 is defined as the use of such information to solve other geoscientific problems, using Bayesian inference or another system of inference.

We also present an example of how Bayes Rule can be used to combine simple geological prior information with limited data to obtain significantly improved constraints on a solution by the construction of a model relating data and model parameters that can be believed with less than infinite uncertainty. These constraints were impossible to obtain from either data or prior information alone.

Acknowledgement

We thank Schlumberger for permission to publish this work.

References

Baddeley, M. C., Curtis, A., & Wood, R.A. 2004. An introduction to prior information derived from probabilistic judgements: Elicitation of knowledge, cognitive bias and herding. *This volume.*

Bosscher, H., & Schlager, W. 1993. Accumulation rates of carbonate platforms. *Journal of Geology*, **10,** 345-55.

Bowden, R.A. 2004. Building confidence in geological models. *This volume.*

Potts, D.C. 1984. Generation Times And Quaternary Evolution Of Reef-Building Corals. *Paleobiology*, **10**, 48-58.

Budd, A.F., Peterson, R.A., & Mcneill, D.F. 1998. Stepwise faunal change during evolutionary turnover: a case study from the Neogene of Curacao, Netherlands Antilles. *Palaios*, **13**, 170-188.

Burgess, P.M. 2001. Modeling carbonate sequence development without relative sea-level oscillations. *Geology*, **29**, 1127-1130.

------------------ Wright, V.P., & Emery, D. 2001. Numerical forward modelling of peritidal carbonate parasequence development: implications for outcrop interpretation. *Basin Research*, **13**, 1-16.

---------------- & Emery, D.J. 2004. Sensitive Dependence, Divergence and Unpredictable Behaviour in a Stratigraphic Forward Model of a Carbonate System. *This volume.*

Curtis, A. & Lomax, A. 2001. Prior information, sampling distributions and the curse of dimensionality. *Geophysics*, **66,** 372-378.

-------------- & Wood, R. 2004. Optimal elicitation of probabilistic information from experts. *This volume.*

Drummond, C.N. & Wilkinson, B.H. 1993. On the use of cycle thickness diagrams as records of long-term sealevel change during accumulation of carbonate sequences: *Journal of Geology*, **101**, 687-702.

Hodgetts, D., Drinkwater, N.J., Hodgson, D.M., Kavanagh, J., Flint, S.S., & Keogh, K.J. 2004. Three-dimensional geological models from outcrop data using digital data collection techniques: an example from the Tanqua Karoo depocentre, South Africa. *This volume.*

Jones, R.R., Mccaffrey, K.J.W., Wilson, R.W., & Holdsworth, R.E. 2004. Digital field data acquisition: towards increased quantification of uncertainty during geological mapping. *This volume.*

Lehrmann, D.J., & Goldhammer, R.K. 1999. Secular variation in parasequence and facies tacking patterns of platform carbonates: A guide to application of stacking pattern analysis in strata of diverse ages and settings. *In:* Harris, P.M., *et al.*, (eds) *Advances in carbonate sequence stratigraphy: Applications to reservoirs, outcrops and models*. SEPM (Society for Sedimentary Geology) Special Publication **63**, 187-225.

Malinverno, A. 2000. A Bayesian criterion for simplicity in inverse problem parametrization. Geophys. J. Int., **140**, 267-285.

------------------ & Briggs, V. A. 2004. Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes. *Geophysics, in press*.

Mukerj, T., Jorstad, A., Avseth, P., Mavco, G. & Granli, J.R. 2001. Mapping lithofacies and pore-fluid probabilities in a North Sea reservoir: seismic inversions and statistical rock physics. *Geophysics*, **66**, 988-999.

Pschenichny, C. A. 2004. Obtaining probability estimates from concepts. *This volume.*

Pojasek, R.B. 1980. *Toxic and Hazardous Waste Disposal.* Butterworth-Heinemann.

Rosenbaum, M.S. & Culshaw, M.G. 2003. Communicating the risks arising from geohazards *Journal of the Royal Statistical Society: Series A (Statistics in Society),* **166**, 261 – 288.

Sandberg, P.A. 1985. An oscillating trend in Phanerozoic non-skeletal carbonate mineralogy. *Nature*, **305**, 19-22.

Schon, J. H. 2004. *Physical Properties of Rocks: Fundamentals and Principles of Petrophysics* Handbook of Geophysical Exploration: Seismic Exploration Series). Pergamon Press.

Selby, M.J. 1982. *Hillslope materials and processes.* Oxford University Press, Oxford, 264 pp.

Shapiro, N. M., Ritzwoller, M. H., Mareschal, J. C., & Jaupart, C. 2004. Lithospheric structure of the Canadian Shield inferred from inversion of surface wave dispersion with thermodynamic a priori constraints. *This volume.*

Stephenson, D., Gallagher, K., & Holmes, C. 2004. Beyond kriging-dealing with discontinuous spatial data fields using adaptive prior information and Bayesian partition modelling. *This volume.*

Tarantola, A., & Valette, B. 1982. Inverse problems = Quest for information. Journal of Geophysics, **50**, 159-170.

---------------- 1987. Inverse problem theory. Elsevier Health Sciences (publisher); 4th reprint 2002 edition (September 1, 2002).

Tetzlaff, D. M. & Priddy, G. 2001. Sedimentary Process Modeling, from Academia to Industry. *In:* Merriam, D. F., & Davis, J. C. (eds) *Geologic Modeling and Simulation, Sedimentary Systems.* Kluwer Academic/Plenum Publishers, New York, pp. 45-69.

------------------. 2004. Input uncertainty and conditioning in siliciclastic process modelling. *This volume.*

Tinker, S., Caldwell, D., Brinton, L., Brondos, M., Carlson, J., Cox, D., Demis, W., Hamman, J., Laskowski, L., Miller, K., & Zahm, L. 1999. Sequence stratigraphy and 3-D modelling of a Pennsylvanian, distally steepened ramp reservoir: Canyon and Cisco formations, South Dagger Draw field, New Mexico, USA: GCSSEPM Foundation 19th Annual Research Conference, Advanced Reservoir Characterization, December 5-8, p. 214-232.

Verwer, K., Kenter, J.E.M., Maathuis, B., & Della Porta, G. 2004. Stratal patterns and lithofacies of an intact seismic-scale Carboniferous carbonate platform (Asturias, NW Spain): A virtual outcrop model. *This volume.*

White, N. 2004. Using prior subsidence data to infer basin evolution. *This volume.*

Wijns, C., Poulet, T., Boschetti, F., Dyt, C., & Griffiths, C. M. 2004. Interactive inverse methodology applied to stratigraphic forward modelling. *This volume.*

Wilkinson, B.H., Drummond, C.N., Rothman, E.D., & Diedrich, N.W. 1997. Stratal order in peritidal carbonate sequences. *Journal of Sedimentary Research*, **67**, 1068 - 1082.

Wright, V.P. 1992. Speculations on the controls on cyclic peritidal carbonates: Icehouse versus greenhouse eustatic controls. *Sedimentary Geology*, **76**, 1-5.

Tables

Table 1. The contribution of each paper in this volume placed within a geological prior information framework (see main text for details).

Contribution	STEP 1		STEP 2				Contribution Type		
	Quantify $G_{q'} \rightarrow G_q$ Static	Quantify $G_{q'} \rightarrow G_q$ Dynamic	Calculate Prior $P(\mathbf{m} \mid \hat{B})$	Calculate Evidence $P(\mathbf{d} \mid \hat{B})$	Calculate Likelihood $P(\mathbf{d} \mid \mathbf{m}, \hat{B})$	Calculate Posterior $P(\mathbf{m} \mid \mathbf{d}, B)$	Method	Application	Discussion
Wood & Curtis	*	*	*	*	*	*	*		*
Baddeley et al.	*		*						*
Verwer et al.	*							*	
Jones et al.	*							*	
Hodgetts et al.	*		*					*	
Burgess & Emery		*					*		
Tetzlaff		*					*		
Curtis & Wood	*	*	*				*		
Wijns et al.	*	*						*	
Pshecnichny	*	*	*	*	*	*	*		
Bowden	*	*	*		*	*	*		
Shapiro et al.	*		*		*	*	*	*	
Stephenson et al.	*	*	*	*	*	*	*		
White					*		*	*	

Table 2. Typical ranges of model parameter distributions appropriate for carbonate formations. Minimum and maximum values for each parameter are sufficient to define the joint Uniform prior probability distribution. Distributions for all parameters other than ramp dip were used to illustrate the prior and posteriori distributions in Figs 5 and 6. Published sources: Bosscher & Schlager (1993); Lehrmann & Goldhammer (1999).

Parameter	**Min Value**	**Max Value**
flank dx	25 m	1 km
flank dz	0 m	50 m
basindz	0 m	2 m
lagoondz	1 m	10 m
ramp dip	1 degree	5 degrees
trend	-50 degrees	180 degrees
wloc	0 km	1 km
number of boundaries	1	10

Figures

*Fig. 1. Model and parameters describing the boundaries of parasequences within a TST, using half a sine curve with horizontal terminations. All parameters are defined graphically except for nc (the total number of boundaries), and wloc (horizontal location of well relative to geological architecture). The depositional trend is given as an angle. All parameters other than ramp dip are included in model vector **m**.*

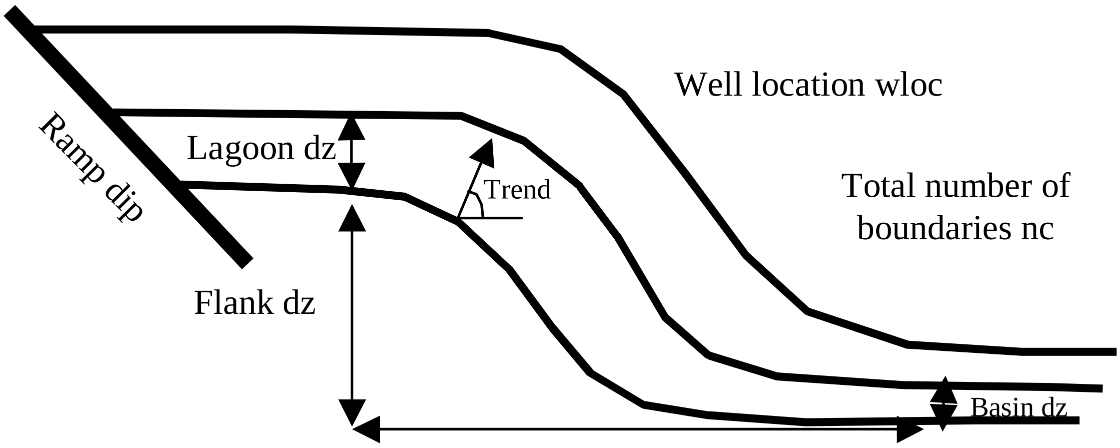

Fig. 2. Modelled parasequence successions. The uppermost models show the parasequence boundaries and three vertical sections or well locations. Dashed lines on the lower six plots show the depths of intersection of each well with parasequence boundaries (top and bottom figures are left and right wells, respectively). All dimensions are given in metres. A: Progradational sequence where the depositional trend (trajectory) is 10 degrees basinwards. B: Regressive sequence where the depositional trend (trajectory) is -10 degrees basinwards.

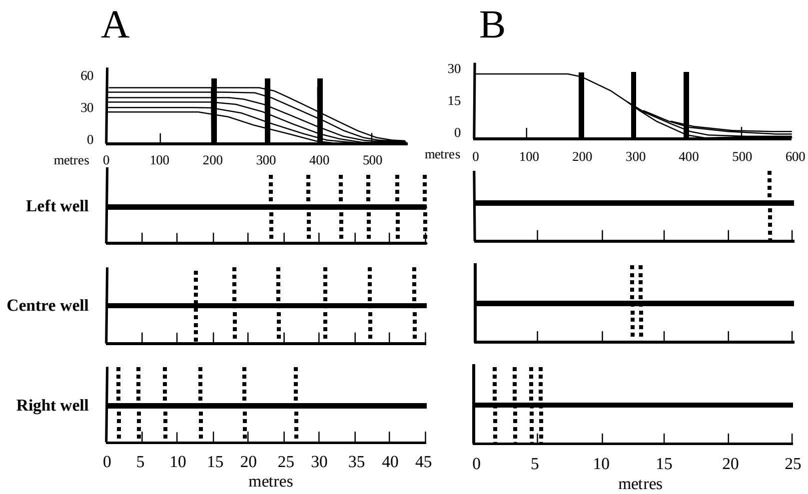

Fig. 3. Sections of different parasequences sampled at a well are approximately equivalent to sections of the same (e.g., lowermost) parasequence taken at successively horizontally offset locations. Intersections refer to the parasequence boundaries; Available cross sections are horizontally offset points where parasequence thicknesses occur that are equivalent to those found in the well. The double arrows indicate to which well section each available cross section is similar.

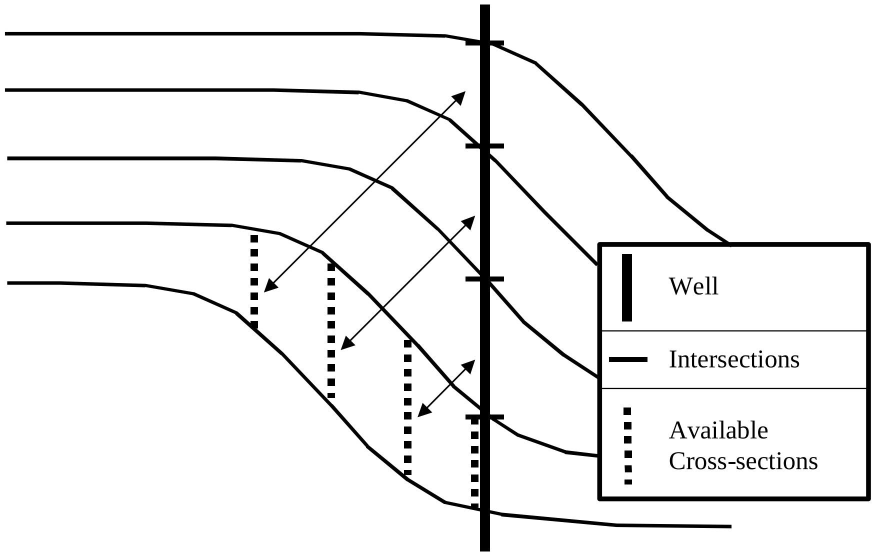

Fig. 4. 1-dimensional marginal histograms for each of the parameters controlling both well location and stratigraphic scenario. Diamonds indicate the mean value in each graph; circles represent the 'true' value in Figure 2.

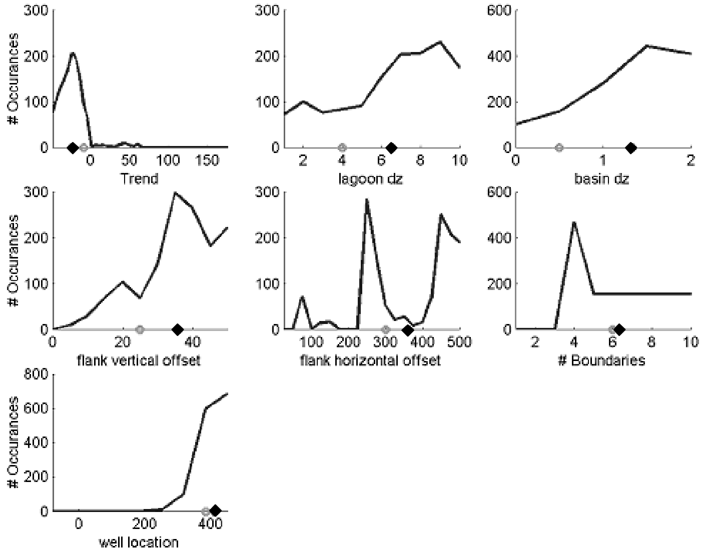

Fig. 5. Histograms over flank gradient g, defined in the text, for A: Prior distribution; B: Posterior distribution.

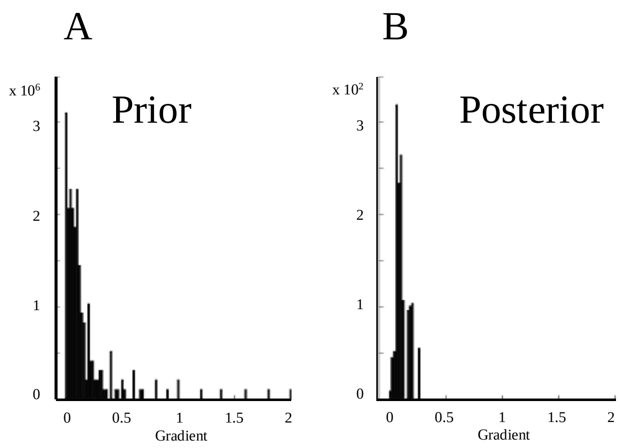

Fig. 6. Histograms over different depositional regimes (R: Retrogradation; P: Progradation; A: Aggradational; B: Regressive; arrows show corresponding trend directions) for A: Prior distribution; B: Posterior distribution.