Types of uncertainty in palaeobiological information

Rory E. Browne*[1], Zena Wood[1] and Michael Worboys[1]

[1]Department of Computing and Information Systems, University of Greenwich

Summary

Palaeobiological information has a variety of characteristics that increase the associated uncertainties and we briefly describe some of these along with context for how they arise. We explore the types of uncertainty commonly encountered in palaeobiological information, giving examples. Concentrating on the temporal uncertainties, we give an example of palaeobiological temporal relations as might be found in the real world. We then suggest a technique by which temporal uncertainty might be included in analysis, a modified version of Allen's temporal relations. Palaeobiological studies increasingly make use of GIS tools to explore the spatial relationships of ancient organisms. The inherent uncertainties of palaeobiological information constrain analysis. Various types of uncertainty present in palaeobiological information are identified and we discuss some approaches to addressing them. We concentrate on temporal uncertainty, as this is a central issue in geo-spatial analysis of palaeobiological information.

KEYWORDS: Uncertainty, Allen's Temporal Relations, Palaeobiology

1 Introduction

Palaeobiological studies increasingly make use of GIS tools to explore the spatial relationships of ancient organisms (Hendricks et al., 2008; Rode and Lieberman, 2004). The inherent uncertainties of palaeobiological information constrain analysis. Various types of uncertainty present in palaeobiological information are identified below and we discuss some approaches to addressing them. We concentrate on temporal uncertainty, as this is a central issue in geo-spatial analysis of palaeobiological information.

Palaeobiological information originates from many sources, predominately fossils and their associated sediments. Sediments can be analysed to reveal a variety of environmental signals, provide age information, and give context for fossils. Fossils provide taxonomic information, as well as taphonomic (type of fossilisation) and location information. The nature of the preservation has implications for the completeness of the fossil record in a rock unit, and informs the reliability of any analysis conducted using either the fossil or the rock unit.

*Rory.E.Browne@greenwich.ac.uk

Palaeobiology emphasises the biological components of fossil material, concentrating on reconstructing the creature that formed the fossil. Spatial-temporal questions asked by palaeobiologists include: did climate drive the evolution or control the distribution of a particular group or species? (Carvalho et al., 2010; Mayhew et al., 2008; Rode and Lieberman, 2004); did species or group X co-evolve with species or group Y? (Butler et al., 2009).

In this paper we classify types of uncertainty encountered in palaeobiological information and provide palaeobiological examples for each. We describe a hypothetical geological section (sequence of rocks in a location) and use this to give examples of geological temporal relations and consequent kinds of uncertainty. We demonstrate the use of Allen's temporal relations (Allen, 1983) and describe modifications to address palaeobiological uncertainties.

2 Types of uncertainty

There are six main types of uncertainty encountered in palaeobiologial information.

- Precision: how tightly data are constrained.
- Vagueness: where boundaries are unclear, so we cannot determine which group a point belongs to with certainty.
- Accuracy: how close to reality a value is.
- Reliability: how much we trust our data, this can result from unreliable data sources or a lack of knowledge about information.
- Incompleteness: where we know that we do not have all the information on a phenomenon, only a subset, and may not know how representative that subset is.
- Ambiguity: where there are disagreements as to how to identify a data object.

Methodologies for dating fossil material can be precise, e.g., radio-isotope dating, but are more commonly imprecise. The possibility of time gaps (explained below) between continuous strata increases the vagueness of dating of layers. Accuracy is hard to test, but through comparing results from different methods, we increase our confidence in the accuracy of data. Particular samples may be especially unreliable, as poor provenance inhibits the use of contextual information. These issues can sometimes be sidestepped through use of a coarser granularity, with consequential loss of precision.

Temporal granularity generally has similar issues to that of spatial granularity (Laube and Purves, 2011). In palaeobiology temporal granularity commonly takes the form of time ranges. Problems arise when time ranges are sufficiently long that the position of continents, coastlines, etc., have changed.

As the likelihood of any particular creature being fossilised is extremely low, the selection of creatures present in the fossil record is biased towards those species that are more numerous or robust (Brown et al., 2013). The limited volume of rock of any particular time-period mean that parts of the fossil

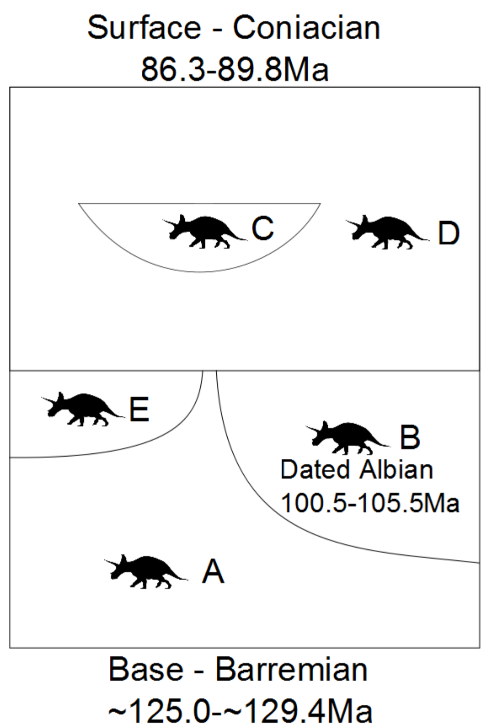

Figure 1: Stratigraphic Sequence showing the known information for our example fossils. Letters refer to the rock unit that contains the fossil. Rock units contain the fossils: A:Alice, B:Bob, C:Charlie, D:David, E:Edward. Ma = megaannus (million years).

record are under-represented. There have been numerous attempts to quantify this incompleteness, both on regional (Dunhill et al., 2014) and global (Benton et al., 2013) scales. These suggest that it is possible to quantify the completeness of the fossil record, although questions remain as to the correctness of such statements.

Ambiguity (Fisher, 1999) is a major limitation in palaeobiological information, as many factors confuse the fossil record including: disagreements about how to identify species; re-assessment of evidence (Tschopp et al., 2015); alterations from post-depositional processes (Sansom et al., 2011).

Table 1: Notation used in this paper, excluding Allen's temporal relations

Notation	Explanation
F	fossil (any uppercase letter)
$F_{interval}$ or F_i	temporal interval of fossil
f	rock unit (any lowercase letter, the same as the fossil within it)
$f_{interval}$ or f_i	temporal interval of rock unit
$tgap_n$	Time gap n
[ ]	interval from oldest start date to youngest end date of all intervals within brackets

3 Palaeobiological Temporal Relations

Palaeobiologists are interested in the *lifetime* of the *creature* that forms the *fossil*. The fossil is dated to the interval in which the rock containing it was deposited. The sequence of rocks at a location produces a relative order of deposition (*relative dating*). Some rocks can be analysed using quantitative methods to establish the date of deposition (*absolute dating*). By combining both forms of dating rock layers can be dated.

Figure 1 demonstrates a sequence containing five fossils with date information: Alice, Bob, Charlie, Dave and Edward. Table 1 explains the notation used to refer to the components of this sequence. The interval in which fossil F was deposited is $F_{interval}[F_{start},\ F_{end}]$. The rock unit that contains fossil F is f and has the temporal interval $f_{interval}[f_{start},\ f_{end}]$.

Below are statements about when each fossil existed (see also Figure 2). Geological constraints imply that e and b have erosive bases so must postdate the deposition of a.

1. All fossils were deposited between Barremian (~129.4Ma – ~125.0Ma) and Coniacian (89.8Ma – 86.3Ma).
2. Alice, Bob and Edward were deposited before Dave and Charlie.
3. Bob was deposited after Alice and before Dave, during the Albian (105.5Ma – 100.5Ma).
4. Alice must have been deposited between the Barremian and Albian.
5. Dave was deposited between the Albian and Coniacian.
6. Charlie was deposited within the interval that Dave was, between the Albian and Coniacian.
7. Edward is younger than Alice and older than Dave and Charlie, we do not know his relationship to Bob.

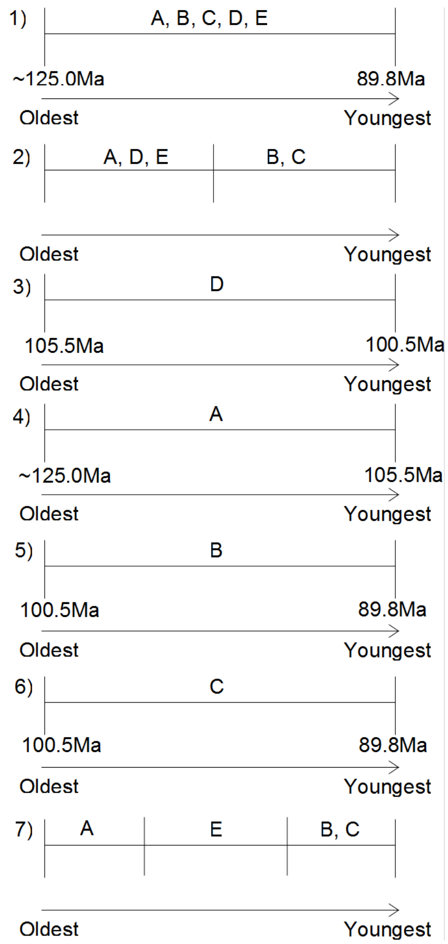

Figure 2: Timelines for the statements about Figure 1

3.1 Uncertainties present in example

Statements 1, 3, 4 and 5 are precise, as they refer to fixed temporal intervals. Statement 6 is imprecise although it refers to a fixed temporal interval, as it is constrained by that interval, rather than defined by it. All the statements include vagueness, but some statements are more vague than others. Statements 1, 3, 4 and 5 are the least vague. Statement 6 is vague, but is constrained by that vagueness. Statements 2 and 7 are vague as they provide a precise sequence, but do not provide any information on the duration of that sequence or the elements of that sequence.

4 Approaches to Uncertainty: Allen's temporal relations

Allen (1983) proposed a system for reasoning about temporal intervals that allows for uncertainty by relying on relative timescales. Using his notation, we define the following notation for geological and palaeobiological temporal relations (Table 2). Allen's statement *meets* can only apply to time gap intervals as we cannot assume that two rock units form temporally continuous sequences.

Time gaps are the result of *unconformities*; "A break in the stratigraphic record which represents a period of no sediment deposition" (Kearey, 2001) and are a major source of temporal uncertainty in geological data. This occurs as a result of no deposition or sediment removal through erosion (Figure 3).

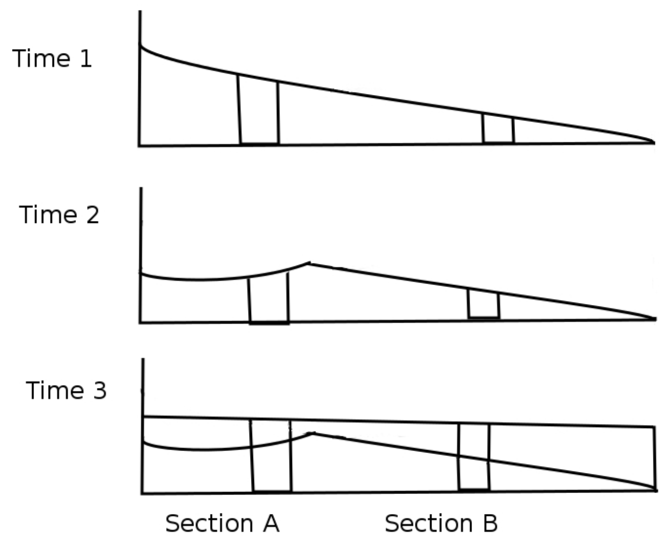

Figure 3: Diagrammatic example of time gaps and how they form. Section A shows the formation of a time gap through erosion, Section B shows the formation of a time gap through no deposition. There is no geological record of events that happened in Time 2.

For example: rock unit x has a unconformable boundary with the younger unit y that lies above it. The time gap interval *tgap* can be depicted using Allen's temporal relations as:

$$x_i \quad \mathrm{m} \quad tgap \quad \mathrm{m} \quad y_i \tag{1}$$

By distinguishing time gap intervals from other intervals we are able to represent a known temporal uncertainty in the geological record. We choose to do this so time gaps may be maintained, regardless of any refinements to the sequence. This highlights the existence of the uncertainty to non-specialists and allows for refinement of the sequence to account for it. Time gaps are the only intervals in the geological record that can use Allen's Temporal relation meets, as they must meet the preceding and succeeding intervals (as they represent the unknown period between two known intervals).

Using this system we can re-write our statements about the fossils in Figure 1 as follows:

Relation	Notation
X before Y	X < Y
X equals Y	X = Y
X after Y	X > Y
X overlaps Y	X o Y
X during Y	X d Y
Only applies to *tgap* intervals	
X meets Y	X m Y

Table 2: Allen's temporal relations as redefined for geological relations

$$Barremian < a_i, b_i, c_i, d_i, e_i < Coniacian \tag{2}$$

$$a_i, e_i, b_i < d_i, c_i \tag{3}$$

$$a_i < b_i < d_i \tag{4}$$

$$Barremian < a_i < Albian \tag{5}$$

$$Albian < d_i < Coniacian \tag{6}$$

$$c_i \quad \mathrm{d} \quad d_i \tag{7}$$

$$a_i < e_i < d_i, c_i \tag{8}$$

We can also add those relations with a known time gap deduced from geological expertise as:

$$a_i \quad \mathrm{m} \quad tgap_1 \quad \mathrm{m} \quad d_i \tag{9}$$

$$a_i \quad \mathrm{m} \quad tgap_2 \quad \mathrm{m} \quad b_i \tag{10}$$

$$a_i \quad \mathrm{m} \quad tgap_3 \quad \mathrm{m} \quad e_i \tag{11}$$

We can re-write equation 5 as follows, including the time gap.

$$Barremian < a_i \quad \mathrm{m} \quad tgap_2 \quad \mathrm{m} \quad Albian \tag{12}$$

By signalling the presence of time-gaps, we enable non-specialists to use the interpreted data without time-gaps being ignored or underestimated.

Figure 4 demonstrates the value of this approach; two new fossils, F and G are to be related to the main sequence. Using Allen's temporal relations we can describe their relations as follows:

$$a_i < b_i, c_i, d_i, e_i, f_i, g_i \tag{13}$$

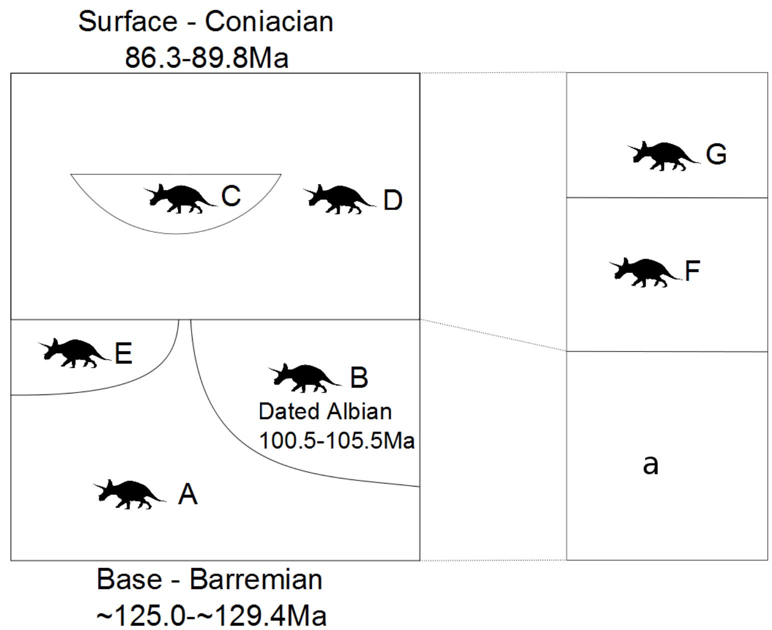

Figure 4: Expanded example of geological temporal relations

$$a_i < d_i, c_i, f_i, g_i < Coniacian \tag{14}$$

$$[d_i, c_i] = [f_i, g_i] \tag{15}$$

$$e_i, b_i < d_i, c_i, f_i, g_i \tag{16}$$

We prove that f and g were deposited after e and b. Equally, having proved the order of deposition, we can deduce:

$$a_i \quad \text{m} \quad tgap_4 \quad \text{m} \quad f_i \tag{17}$$

This is $tgap_4$ as we do not know how it might relate to the other time gaps ($tgap_1, tgap_2, tagp_3$) but do know there must be a time gap. Geologically, this could happen, although the spatial extent of the rock units may become relevant. Should the distance between the two sections be large enough, deposition might have stopped in one area and continued in another. In this scenario, however, the temporal extent of a has changed between sections, highlighting the complexity of the geological record.

5 Further Work and Conclusions

We have described the six types of uncertainty that are encountered in palaeobiological information: vagueness, accuracy, precision, reliability, incompleteness and ambiguity, concentrating on those that are relevant to determining temporal position. We have suggested that there is scope for Allen's temporal relations to reduce the uncertainties encountered in palaeobiological information, or at least to allow the correlation of unrelated temporal intervals.

Future work will concentrate on the application of the above relations to palaeobiological data. Vagueness is the primary temporal limitation of palaeobiological data, therefore the modified Allen's relations should be tested in comparison/concert with fuzzy sets (see Kauppinen et al. (2010) for examples of fuzzy sets applied to past temporal records). Additionally, we wish to consider the volume of space-time occupied by prehistoric organisms when exploring their lifetimes and relationships.

References

Allen, J. F. (1983). Maintaining Knowledge About Temporal Intervals. *Commun. ACM*, 26(11):832–843.

Benton, M. J., Ruta, M., Dunhill, A. M., and Sakamoto, M. (2013). The first half of tetrapod evolution, sampling proxies, and fossil record quality. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 372:18–41.

Brown, C. M., Evans, D. C., Campione, N. E., O'Brien, L. J., and Eberth, D. A. (2013). Evidence for taphonomic size bias in the Dinosaur Park Formation (Campanian, Alberta), a model Mesozoic terrestrial alluvial-paralic system. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 372:108–122.

Butler, R. J., Barrett, P. M., Kenrick, P., and Penn, M. G. (2009). Testing co-evolutionary hypotheses over geological timescales: interactions between Mesozoic non-avian dinosaurs and cycads. *Biological Reviews*, 84(1):73–89.

Carvalho, I. d. S., de Gasparini, Z. B., Salgado, L., de Vasconcellos, F. M., and Marinho, T. d. S. (2010). Climate's role in the distribution of the Cretaceous terrestrial Crocodyliformes throughout Gondwana. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 297(2):252–262.

Dunhill, A. M., Hannisdal, B., and Benton, M. J. (2014). Disentangling rock record bias and common-cause from redundancy in the British fossil record. *Nature Communications*, 5:4818.

Fisher, P. F. (1999). Models of uncertainty in spatial data. In Longley, P. A., Goodchild, M. F., Maguire, D. J., and Rhind, D. W., editors, *Geographical Information Systems*, volume 1, pages 191–205. Wiley.

Hendricks, J. R., Lieberman, B. S., and Stigall, A. L. (2008). Using GIS to study palaeobiogeographic and macroevolutionary patterns in soft-bodied Cambrian arthropods. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 264(12):163–175.

Kauppinen, T., Mantegari, G., Paakkarinen, P., Kuittinen, H., Hyvnen, E., and Bandini, S. (2010). Determining relevance of imprecise temporal intervals for cultural heritage information retrieval. *International Journal of Human-Computer Studies*, 68(9):549–560.

Kearey, P. (2001). *The New Penguin Dictionary of Geology*. Penguin Reference. Penguin Books Ltd, second edition.

Laube, P. and Purves, R. S. (2011). How fast is a cow? Cross-Scale Analysis of Movement Data. *Transactions in GIS*, 15(3):401–418.

Mayhew, P. J., Jenkins, G. B., and Benton, T. G. (2008). A long-term association between global temperature and biodiversity, origination and extinction in the fossil record. *Proceedings of the Royal Society of London B: Biological Sciences*, 275(1630):47–53.

Rode, A. L. and Lieberman, B. S. (2004). Using GIS to unlock the interactions between biogeography, environment, and evolution in Middle and Late Devonian brachiopods and bivalves. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 211(34):345–359.

Sansom, R. S., Gabbott, S. E., and Purnell, M. A. (2011). Decay of vertebrate characters in hagfish and lamprey (Cyclostomata) and the implications for the vertebrate fossil record. *Proceedings of the Royal Society of London B: Biological Sciences*, 278(1709):1150–1157.

Tschopp, E., Mateus, O., and Benson, R. B. (2015). A specimen-level phylogenetic analysis and taxonomic revision of Diplodocidae (Dinosauria, Sauropoda). *PeerJ*, 3:e857.