

Forensic Reasoning about Paleoclimatology

Laura Rassbach de Vesine*, Ken Anderson*, Marek Zreda†,
Chris Zweck†, and Liz Bradley*^o

* University of Colorado, Dept of Computer Science ^o Santa Fe Institute

† University of Arizona, Dept of Hydrology

Abstract

Human experts in many scientific fields routinely employ heuristics that are unproven and possible conclusions that are contradictory. We present a deployed software system for cosmogenic isotope dating, a domain that is fraught with these difficult issues. This system, which is called ACE (“age calculation engine”), takes as inputs the nuclide densities in a set of rock samples taken from a landform. It reasons from these data—which capture how long those rocks have been exposed to the sky—to answer the scientific question “What geological processes could have produced this distribution of nuclide concentrations, and over what time scales?” To do this, ACE employs an encoded knowledge base of the possible processes that may have acted on that landform in the past, complete with the mathematics of how those processes can affect samples, and it uses a custom workflow system to encode the computations associated with this scientific analysis. The system remains in active use to this day; the project website (ace.hwr.arizona.edu) has received over 17,000 hits since 2008 and the software (~20,000 lines of python code) has been downloaded nearly 600 times as of April 2013, which is a significant number in a research community of $O(10^2)$ PI-level scientists.

Introduction

Scientific discovery is becoming increasingly challenged by complex and sometimes contradictory reasoning about noisy, heterogeneous data—often too much data, but sometimes not enough. Helping scientists manage and make sense of those data is an important challenge for computer science in general and discovery informatics in particular. The ACE software tool, which was developed during a tight, five-year collaboration between geoscientists and computer scientists, supports a specific geological analysis task: the dating of landforms using cosmogenic isotope data. ACE uses modern software-engineering techniques to handle the associated data, computation, and interface issues. Its argumentation-based engine, which captures expert reasoning about this scientific domain, explores the hypothesis space automatically. This unified solution facilitates the doing of science

in this challenging domain, and holds promise for other domains as well.

The software-engineering and geoscience aspects of ACE are covered in (Anderson et al. 2007) and (Zweck et al. 2012), respectively. Our goal in this paper is to offer a few of the insights that we gained, as a result of this project, into the representation and reasoning challenges that arise in forensic science, where timelines are unknown, empirical data are scarce, and controlled experiments are impossible. Technical details about the representation and reasoning process, which are necessarily missing from a short workshop paper like this one, are provided in (Rassbach, Anderson, and Bradley 2011). A full discussion of the entire Calvin project—including the development and evaluation process that is described later in this paper—can be found in (Rassbach 2009).

Dating landforms is very much like investigating a crime scene: from the information that is available on the surface today, experts must deduce what happened in the past. Many landforms are created by single events that happen almost instantaneously in geological time. They then evolve over time in ways that are known, at least in general. Terminal moraines, for example, are formed when glaciers recede. As these landforms age, subsurface rocks are exposed to the sky—and thus to cosmogenic isotopes—as the fine matrix around them erodes. See Figure 1 for a diagram. A geoscientist sees the situation shown on the right. If she wants to know when that moraine was formed, she needs to reason backwards to determine the initial state of the system—the situation on the left—in order to deduce the timeline. This entails figuring out what processes were involved in the evolution of the landform. Geoscientists tackle that problem by making some assumptions about those processes, projecting those assumptions backwards through time and space to the putative formation time of the landform, and iterating the process until the modeling results are consistent with the observations.

ACE was designed to assist scientists in carrying out the complex reasoning task sketched in the previous paragraph. Working with a collection of exposure ages derived via radioisotope dating from rock samples, its first task was to “calibrate” the isotope dating method. That entailed determining a production rate for a particular nuclide from a set of rock samples of known ages—the “calibration set.” It then

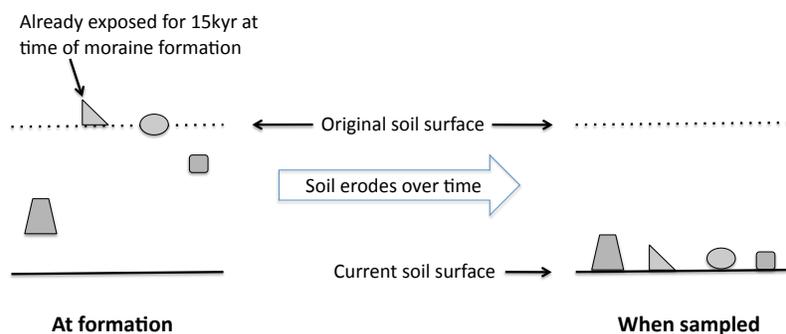


Figure 1: The evolution of a moraine

used that production rate to deduce the ages of the new, undated samples. ACE incorporated a number of different calibration sets: e.g., for different cosmogenic nuclides and different scientific situations. It used a custom-made workflow system to make it easy to create, edit, run, and evaluate new cosmogenic dating algorithms. The design of this workflow system stressed simplicity over complex functionality in order to keep the scientists engaged in the development process and to present a lower entry barrier compared with existing systems such as Kepler. This design was refined iteratively over the course of the project, as the workflow system played an important role in getting geoscientists to detangle their monolithic spreadsheets into highly reusable components.

ACE's user began each run by creating a new *experiment*, the name for the basic software construct that captured all of the information about a particular run, then specified a nuclide of interest for that experiment and a calibration set for that nuclide. S/he then ran the calibration workflow in the workflow engine to create the age estimates for the samples. ACE's reasoning engine—the topic of the rest of this paper—then took over, using an encoded knowledge base of rules about mathematical geoscience to work through scenarios about what processes could have produced that set of sample ages. Finally, ACE reported these possible scenarios to its scientist-user via a custom GUI, together with a narration of its reasoning about each one.

Automated Reasoning about Cosmogenic Isotope Dating

Cosmogenic nuclide dating techniques are ideal for dating surface features such as meteor impact sites, earthquake ruptures, lava flows, alluvial fans, terraces and landforms associated with the retreat of glaciers (Desilets and Zreda 2003). The key challenge faced by geoscientists who take this approach is to reason from the raw results—that is, the exposure times of the samples—in order to understand the overall history of the landform. This generally requires weeks or months of effort on the part of a highly trained expert. If all of the sample exposure ages overlap, the problem is comparatively easy: the true age is somewhere in that overlap. This rarely happens, however; rather, the spread of the exposure ages is generally broad and uneven. Laboratory costs

are high and good samples can be hard to come by, so sample sets are small and noisy. Furthermore, samples may have been disturbed since the formation of the landform, and calibration curves do not represent ground truth. To deduce the true age of the landform, the expert must face these daunting issues and construct a geologically meaningful and defensible explanation for the spread.

ACE's reasoning engine, Calvin, automates this complicated, subtle reasoning process (Anderson et al. 2010; Rassbach 2009; Rassbach, Anderson, and Bradley 2011; Rassbach and Bradley 2008; Rassbach et al. 2007). This engine is an iterative argumentation system that is based primarily on the Logic of Argumentation of Krause *et al.* (Krause et al. 1995). It was hand-written for this application, can handle both discrete symbols and continuous values, and consists of roughly 500 lines of Python code. Its knowledge base incorporates more than 100 rules, gleaned from an extended knowledge-engineering process involving dozens of geoscientists. Its input is a table of exposure times of a set of samples of a landform. Its goal is to abduce what process(es), acting over what time periods, could have produced those exposure times. Calvin explores this forensic scenario space by enumerating all possible hypotheses about the processes that may have affected the landform—snow cover, erosion, etc.—then considers all the evidence for and against each one. Testing of an individual hypothesis (e.g., that the landform was covered by snow, hiding it from the sky and thereby making surface rocks “look” younger) involves generating all possible arguments for and against it. To do this, Calvin first finds all of the rules in its knowledge base that apply to that hypothesis. One of the rules about snow cover¹, for instance, expresses the knowledge that snow happens at high latitudes and altitudes:

```
latitude < -60 OR
  latitude > 60 OR
  altitude > 1000
  ==> snow-cover
```

Calvin then unifies those rules with the sample ages and uses that unification to construct a collection of arguments about the associated conclusion (viz., moraine X has been affected by processes Y and Z).

The design of this reasoning engine was guided by the na-

¹This is a simplification; this rule has various other attributes as well, including a number that captures the expert's confidence in the data and the reasoning, as discussed later in this section.

ture of the problem at hand, which involves heuristic reasoning with partial support, frequent contradictions, and sparse, noisy data. Most of the explanations that human experts find for the apparent age spread of a set of samples come from a short, known list of geologic processes, most commonly exhumation/erosion and inheritance—when a landform ‘inherits’ one or more older rocks. It is also important to consider the possibility that no process was at work. Despite the relatively small number of candidate processes, constructing these explanations is not a simple matter. Available data are noisy and may not be trustworthy. Different processes may have similar (or cancelling) effects, and multiple processes may be at work. The reasoning involved is heuristic and conclusions do not have absolute confidence; stronger arguments against them may be found, and the current best hypothesis overturned. Moreover, these heuristics are often vague or only slightly supportive of their conclusions. For example, several experts informed us that they “prefer the explanation that requires throwing out the least data.” This heuristic expresses a preference, not a certainty, but it obviously lends some weight to the discussion. Implementing partial support of this nature has been a traditionally slippery problem for AI. Reasoning about sample ages generally involves a fair amount of evidence both for *and against* several processes. Handling this type of contradiction is another well-known problem for automated reasoning.

Contradiction is a particularly important issue in this application. The heuristics used in reasoning about cosmogenic isotope dates frequently contradict one another, and different experts also hold contradictory opinions about the correct heuristics. Especially interesting cases of contradiction arose when experts contradicted themselves. During the knowledge-engineering phase of the ACE project, one geologist said

“The thing about inheritance is, it’s usually thought about as quantized, not incremental. So in Antarctica, say ice advanced every 100k, so a sample is 20k or 120k, not 21k. So that is one thing that should be commonly true about inheritance, it should reflect events, should be things that date from past advances, should be quantized.”

And in the next breath, he said

“However, you can convince me you would see a continuum, [...] the glacier advanced and quarried exposed material to different depths, so delivering stuff with 100k age but could be at depth and look like only 70k. So if you have stuff with one past event but chop it up and deliver different parts you might not see the quantized aspect.”

That is, not only do experts disagree with each other, they sometimes disagree *with themselves*.

Effective interaction with users was also an important consideration in Calvin’s design. Conversations between the AI and geoscience members of the team made it clear that the latter would find most useful a system that not only presented its reasoning in full, but also reasoned as they did, using the same knowledge and information. In addition, we learned in later interactions that educating new students in

this reasoning process is a major challenge for geoscience professors, further indicating the usefulness of a system that provides a complete chain of its reasoning.

Calvin’s answers to these challenges begin with the design of its engine. Geoscientists, we have found, find it natural to reason via arguments: *for* results with which they agree, and *against* those with which they disagree. (One landform dating expert even told us “Well, mostly what we do is argue with each other.”) This maps naturally onto the venerable method of multiple simultaneous hypotheses (Chamberlain 1965): scientists construct a list of possible scenarios, then attempt to form complete arguments for and against every hypothesis, and find themselves convinced by the argument with the strongest support. Argumentation, which has a long—albeit mostly theoretical—history in AI, was an obvious choice for ACE’s reasoning engine because of its natural match to this process. Other traditional AI strategies cannot handle some of the unique ways in which geoscientists reason about cosmogenic isotope data. Experts work with multiple contradictory heuristics at the same time, for instance, and several weak arguments can weaken and/or defeat a strong one. Instantiating this kind of reasoning requires some novel modifications to traditional argumentation strategies—and completely rules out traditional knowledge-based or “expert” systems). Diagnosis systems, model based (Lucas 1997; Santos 1991; Struss 2004) and otherwise (Doyle 1983; Gaines 1996), are inappropriate for Calvin because *complete* models of most geologic processes do not exist and the rules involved are not absolute.

Calvin’s ruleset was the keystone of the AI effort in this project. The design of these rules was critical; noise and partial support, for instance, were addressed by allowing support for hypotheses to have variable strength. Another critical insight in Calvin’s rule-design strategy is that not only can specific *knowledge* be more or less certain, but the *evidence used to apply the knowledge in a specific case* may be of variable suitability. For example, the statistical significance of a sample property might cross the 0.1σ , 0.05σ , or 0.01σ boundary; each of these indicates a stronger case for an argument based on that property holding true. And a general dogma (“moraines tend to erode”) carries less weight than a specific observation (“moraine X has a rounded top²”). For this reason, Calvin uses a rich, multi-level representation to capture experts’ confidence in the data and in the conclusions drawn from it, and to propagate that knowledge through the forensic reasoning chain. Please see (Rassbach, Anderson, and Bradley 2011) for more details and (Rassbach 2009) for a full technical discussion.

Using a list of the processes that are known to operate upon landforms—erosion, snow cover, and so on—Calvin begins by generating a set of candidate hypotheses about what have affected the samples. It then considers these hypotheses one at a time, building arguments for and against each one using backwards chaining. The first step in constructing an argument involves finding all the rules that apply to that hypothesis—i.e., those that refer to the same con-

²This is a direct indication of erosion

clusion. The engine then applies unification to each of these rules, which either produces a new conclusion to consider or generates a comparison to input data.

Every rule in Calvin contains both a conclusion and a template for evidence that supports that conclusion. The primary portion of a rule is an implication of the form $A \Rightarrow C$, where A may be either a single literal or the conjunction (or disjunction) of several literals, and C is the conclusion that A supports. The two contradictory statements on the previous page, for example, became two different rules in Calvin's knowledge base: one that looks for a smooth increase in sample ages as evidence for inheritance, and one that looks for "quantized" inheritance (defined as highly clustered ages). These rules directly contradict each other. Calvin's rich representation of confidence allows it to sort out this contradiction and thereby reproduce this expert reasoning accurately—i.e., to disagree with itself.

Figure 2 shows a schematic of Calvin's logic flow in arguing about these contradictory statements. Beside each rule is a quality rating that records the expert's expressed confidence in that piece of knowledge: the quality of the rule in which the expert expressed more confidence is higher. Calvin considers each of these rules in turn (and any other rules in its knowledge base about inheritance). If a rule's premises cannot be found in the input data, Calvin then argues about those premises as new conclusions. As these rules are applied, the engine assigns the resulting arguments a confidence rating based on the quality of the rule and the applicability of the evidence to the rule (e.g. statistical significance).

Once Calvin has constructed complete arguments for every hypothesis about processes and conditions that might have affected the landform, and assigned a confidence value to each of the considered conclusions, it builds an overall argument—in favor of inheritance, in the case of Figure 2. This final argument includes any detractors found during the process, so the user can see that they have been taken into account. The engine also reports an overall confidence in whether inheritance has occurred based on all the contributing arguments. This information is displayed to the user in the form of a tree. The conclusion is at the top of the tree; supporting rules are displayed as children. Specific evidence from the inputs and information about Calvin's confidence accompanies each rule. Calvin also displays its overall confidence in each top-level conclusion, as well as in cases of controversy (viz., strong evidence both for and against). See Figure 3 for an set of annotated screenshots.

Development & Evaluation

Interdisciplinary research is always challenging—all the more so when one is trying to capture and automate all the explicit and implicit richnesses of expert reasoning. Members of the AI team spent a total of roughly 30 days onsite with the geoscience team members over the course of the first four years of the project. The goals of this phase were to understand the science and design an appropriate knowledge representation and reasoning framework for that domain. The biggest initial hurdle was the desire of experts on the isotope team to present outsiders with an idealized,

and arguably more interesting, version of their reasoning process. Early discussion was limited to (what we later discovered were) extremely unusual projects dealing with large numbers of samples and uncommon processes, without explicit discussion that these were unusual cases. This led to an initial AI design aimed at statistical analysis—something that the average dating project has far too few samples to do reliably—and spatial reasoning. It was only several months later, when the first design had been completed and the AI team was presenting it to the geoscience team, that the AI team learned that their understanding was fundamentally flawed. Planning for these kinds of setbacks is absolutely critical in discovery informatics.

Once the basic architecture and knowledge representation structures were in place, the lead AI team member (de Vesine) then conducted two onsite demo/interviews—each spanning almost two days—with other geoscientists and then iterated on the design. These experts were recommended by the lead ACE geoscientist (Zreda) for this purpose, based on his knowledge of the field. In these two long-form interviews, these experts made many statements that helped to confirm the validity of Calvin's basic design: i.e., that the results were similar in both structure and content to the reasoning of domain experts. In contrast to the serious difficulties involved in learning enough about the field to design a useful reasoning system, we found that it was almost always a trivial exercise for the AI team to turn a piece of expert knowledge into a system rule once that system was designed. Most of the kinds of statements made by experts, such as "when you are working with a moraine your first thought is erosion or inheritance," translate naturally and immediately into Calvin's rule structure, which made the encoding of knowledge a comparatively easy task. There were a few surprises: for instance, that visual observations in the field lend significantly more confidence to conclusions than the AI team had expected. Finally, we learned from these interviews that teaching students to reason about landforms is one of the hardest challenges for senior scientists in this field. Based on this, the AI team altered Calvin's interface to attempt to make results more obvious to students, who could then use it as a learning tool.

Finally, de Vesine attended the main professional meeting of the geoscience research community—the fall 2008 meeting of the American Geophysical Union (AGU)—and conducted 11 one-hour demo/interviews with a range of people working actively in isotope dating³. Some spoke English as a first language, some did not; some were veterans of the field and others were graduate students or new postdocs. In these interviews, too, experts made several general statements that emphasized the appropriateness of Calvin's basic design, such as the following exchange:

de Vesine: So the paper has a verified explanation?

Geologist: we found some evidence that supported this [...] You look for evidence that supports or falsifies individual hypotheses.

And later in the same interview:

³Specifically, those whose AGU topics and recent publications in the field's journals concerned cosmogenic isotope dating.

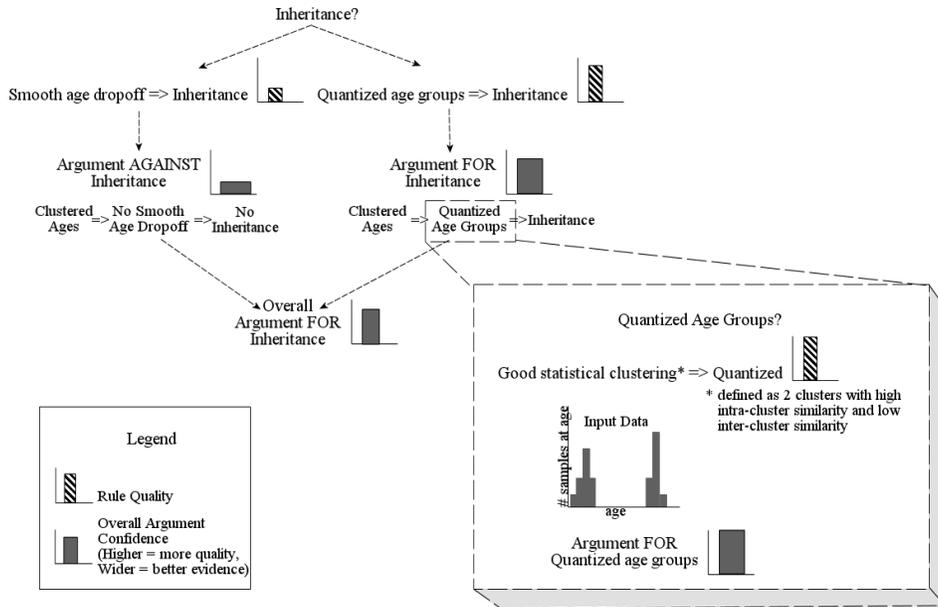


Figure 2: An example of Calvin reasoning about inheritance. Each rule is shown with a quality rating; better knowledge has a higher rating. Arguments are shown with a complete two-dimensional confidence, taking into account both rule and evidence quality. The inset box shows a sub-argument used to extract the conclusion that the input data fall into quantized age groups, which is then included as part of the argument about inheritance. The two conflicting arguments are combined to form one overall argument about inheritance, with a confidence level in the final conclusion that reflects the confidence in both the proponent and the detractors.

de Vesine: I wanted to ask about how you teach this analysis to new grad students and what they get wrong while learning

Geologist: [long pause], that's not an easy question actually. [...] it's very ad hoc [...] they don't understand [some data] at all so you sit them down and talk about them and give them things they could do to test hypotheses: graphs to make and data to collect and we work through it until we are satisfied we have come up with the most reasonable hypothesis

A persistent theme in all of these interviews was how proactively experts acknowledged and emphasized the level of disagreement in the field. Not only were they anxious to point out likely rebuttals that other experts would make to their theories, but they also introduced, without prompting, scenarios where they would need sufficient evidence to overrule a colleague's conclusions about a landform. All of the data obtained in these interviews was incorporated into Calvin's knowledge base, which currently includes 108 rules that represent approximately 50 hours (almost 100 transcribed pages) of direct, intensive interviews with more than two dozen experts in cosmogenic isotope dating. These transcripts can be found in their entirety in (Rassbach 2009).

Our final assessment of Calvin involved using it to reproduce published work: that is, feeding it the data in a published paper and comparing its results against the claims made in that paper. Experts include a self-chosen portion of their qualitative reasoning about a landform when they

publish a new dataset. While this presentation is usually incomplete due to space limitations and the desire to maintain reader interest, it typically includes information about both rejected and accepted conclusions. This material is useful for determining if Calvin's recall is sufficiently high for the most important arguments. For this comparison, we found twenty-five randomly selected papers that appeared from their titles to deal with cosmogenic isotope dating. Of these, eighteen actually discussed one or more isotope dating problems in any detail. These publications included a large cross-section of authors and different isotopes, and spanned about ten years, providing a broad basis of comparison. For each of these papers, we extracted every statement that made an assertion, such as the following example from (Jackson et al. 1997):

Erratic A (sample AE95110101) yielded an age... almost four times older than the next oldest age. This age is clearly anomalous... [t]he most likely explanation for this anomalous age is exposure to cosmic radiation prior to glacial transportation.

We then converted these statements into a form that more closely matched Calvin's terminology. This involved identifying the conclusion argued for in the statement, estimating a level of confidence from terms such as 'clearly' and 'possibly,' and extracting the evidence used in the statement to support the conclusion. Then, we converted the terms in the evidence and conclusion into Calvin's terminology: for example, 'inheritance' instead of 'prior exposure to cosmic

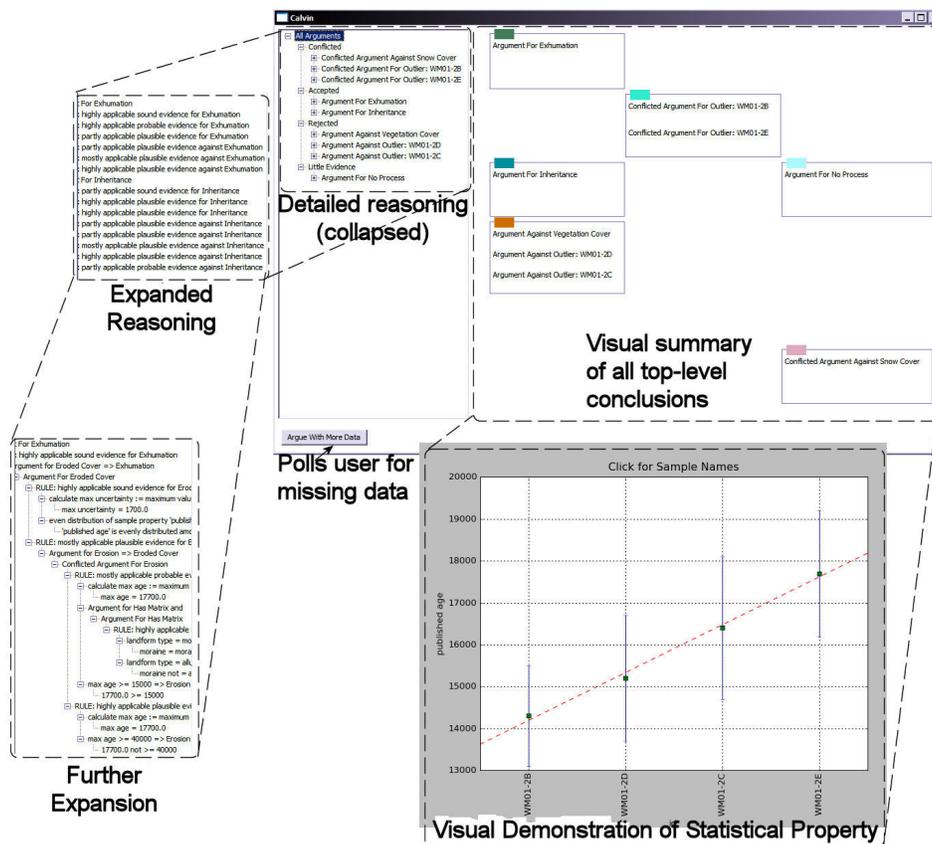


Figure 3: A collection of screenshots from Calvin’s interface [after (Rassbach 2009)], with explanatory annotations. Users can drill down to any level of reasoning, allowing them to explore and understand the system’s arguments around any given conclusion.

radiation’ (which is the definition of inheritance). We cross-validated these results by asking two other people—not experts in isotope dating—to perform the same conversions. We then entered all of the data in the paper, ran Calvin, and compared its output to the converted arguments. It closely reproduced the authors’ arguments 62.7% of the time and produced similar arguments a further 26.1% of the time. In many cases, the similarity was striking, especially when the authors of the paper expressed significant doubt about their conclusions. Again, see (Rassbach 2009) for detailed results.

Perhaps the most interesting cases were when Calvin produced an argument that did not appear in the original paper. When examining (Ballantyne, Stone, and Fifield 1998), for instance, Calvin argued that exhumation was at work. The main evidence for this was a disagreement with ages determined for this landform via other methods. To judge these results, we asked a domain expert to assess Calvin’s new argument. He responded:

I think I see both sides here. From the results, the fact that the ages are younger than the C14 data means that exhumation should be taken very seriously (...) there is not much in the way of material that could bury them. However the peaks themselves are eroding...

In this expert’s opinion, then, the lack of explicit discussion of exhumation in (Ballantyne, Stone, and Fifield 1998) was a major oversight. Although Calvin does not produce exactly the same argument, it found a major gap in the reasoning published by these authors.

Conclusion

In all, Calvin provides several contributions to AI, to geoscience, and to the larger discovery informatics community:

- Its rule base is an explicit representation of the knowledge of two dozen experts in landform dating.
- It incorporates a rich system of confidence that captures the reasoning of real scientists in a useful way.
- It is a fully implemented and deployed system—a surprisingly rare thing in the argumentation literature.
- It is a real tool that is in daily use by real scientists.

There are several software programs available to date landforms using cosmogenic nuclides: hess.ess.washington.edu/math, www.cronuscalculators.nmt.edu, and cosmo-calc.googlepages.com. None of them use modern workflow-based software-engineering techniques, and to our knowledge, only one very recent entry in that list (Applegate et al.

2012) does any kind of automated reasoning—and only for the specific case of moraines.

Overall, our assessment of Calvin indicates that its knowledge base is largely complete and that missing data is easy to add. Moreover, we found that Calvin’s reasoning process is very similar, in both structure and content, to the reasoning process of domain experts. While its design is similar to (and inspired by) previous work, it is unique—and, we feel, more suitable than those previous frameworks are to implementing the multiple-hypothesis reasoning that is employed by experts in cosmogenic isotope dating. Calvin successfully solves this challenging problem in an intuitive and natural way, following the structure of methodology already in use by domain experts.

The computer scientists in the ACE team have now moved on to applying this multidisciplinary research approach to a different domain: the analysis of ice and ocean-sediment cores. From these cores—depth-wise sequences of information—a geoscientist interested in a past climate event must first deduce the timeline for the data: that is, a curve called an *age model* that relates the depth in the core to the age of the material at that point. This is the first critical step in reasoning about the science of the events that produced the core. Like ACE, this new project (entitled CSCIENCE) brings together computer scientists and geoscientists around the goal of producing a software system that enables scientific progress in a challenging application domain. There are many challenges in the CSCIENCE project, some familiar from the previous pages and some requiring fundamental new work in the areas of big data (storage and processing) and automated reasoning. Assumptions about how ice and ocean-sediment cores are created have multiple permutations, for instance, leading to a potential explosion in the number of age models to generate and evaluate. And the data involved are very different. Calvin worked with two numbers (mean and standard deviation) for each of a few dozen rock samples taken from a single landform that was formed instantaneously in geological time, then influenced by a small list of candidate processes that involved no unknown parameters. CSCIENCE’s data sets are thousands or millions of times larger, and their ordered nature allows reasoning about *continuous* events, not just episodic ones, which is a much harder and more general problem.

Acknowledgments. This material is based upon work sponsored by the NSF under Grants ATM-0325812 and ATM-0325929. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

Anderson, K.; Bradley, E.; Zreda, M.; Rassbach, L.; Zweck, C.; and Sheehan, E. 2007. ACE: Age calculation engine: A design environment for cosmogenic dating techniques. In *Proceedings of the International Conference on Advanced Engineering Computing and Applications in Sciences (ADVCOMP)*.

Anderson, K.; Bradley, E.; Rassbach, L.; Zweck, C.; and Zreda, M. 2010. End-to-end support for paleolandform dating. In Adams, N.; Berthold, M.; and Cohen, P., eds., *Advances in Intelligent Data*

Analysis IX, volume 6065. Springer Lecture Notes in Computer Science. Proceedings of the 9th International Symposium on Intelligent Data Analysis (IDA).

Applegate, P.; Urbana, N.; Kellera, K.; Lowell, R.; Laabs, B.; Kelly, M.; and Alley, R. 2012. Improved moraine age interpretations through explicit matching of geomorphic process models to cosmogenic nuclide measurements from single landforms. *Quaternary Research* 77:293–304.

Ballantyne, C.; Stone, J.; and Fifield, L. 1998. Cosmogenic Cl-36 dating of postglacial landsliding at The Storr, Isle of Skye, Scotland. *The Holocene* 8:347–351.

Chamberlain, T. 1965. The method of multiple working hypotheses. *Science* 148:754–759. Reprint of 1890 *Science* article.

Desilets, D., and Zreda, M. 2003. Spatial and temporal distribution of secondary cosmic-ray nucleon intensities and applications to in-situ cosmogenic dating. *Earth and Planetary Science Letters* 206:21–42.

Doyle, J. 1983. Methodological simplicity in expert system construction: The case of judgments and reasoned assumptions. *AI Magazine* 4:39–43.

Gaines, B. 1996. Transforming rules and trees into comprehensible knowledge structures. In *Advances in Knowledge Discovery and Data Mining*, 205–226.

Jackson, L.; Phillips, F.; Shimamura, K.; and Little, E. 1997. Cosmogenic ³⁶Cl dating of the Foothills erratics train, Alberta, Canada. *Geology* 25:195–198.

Krause, P.; Ambler, S.; Elvang-Goransson, M.; and Fox, J. 1995. A logic of argumentation for reasoning under uncertainty. *Computational Intelligence* 11:113–131.

Lucas, P. 1997. Symbolic diagnosis and its formalisation. *The Knowledge Engineering Review* 12:109–146.

Rassbach, L., and Bradley, E. 2008. Challenges in presenting argumentation results. In *Proceedings of the International Workshop on Qualitative Reasoning about Physical Systems*.

Rassbach, L.; Anderson, K.; and Bradley, E. 2011. Providing decision support for cosmogenic isotope dating. *AI Magazine* 32:69–78.

Rassbach, L.; Bradley, E.; Anderson, K.; Zreda, M.; and Zweck, C. 2007. Arguing about radioisotope dating. In *Proceedings of the International Workshop on Qualitative Reasoning about Physical Systems*.

Rassbach, L. 2009. *Calvin: Producing Expert Arguments about Geological History*. Ph.D. Dissertation, University of Colorado.

Santos, E. 1991. On the generation of alternative explanations with implications for belief revision. In *Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence (UAI)*, 339–347.

Struss, P. 2004. Deviation models revisited. In *Proceedings of the Eighteenth International Workshop on Qualitative Reasoning*.

Zweck, C.; Zreda, M.; Anderson, K.; and Bradley, E. 2012. The theoretical basis for ACE, an age calculation engine for cosmogenic nuclides. *Chemical Geology* 291:199–205. doi:10.1016/j.chemgeo.2011.10.005.