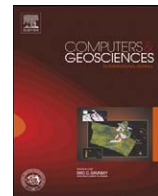




Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

The event bush as a semantic-based numerical approach to natural hazard assessment (exemplified by volcanology)

C.A. Pshenichny^{a,b,*}, S.I. Nikolenko^c, R. Carniel^{d,1}, P.A. Vaganov^e, Z.V. Khrabrykh^f,
V.P. Moukhachov^g, V.L. Akimova-Shterkhun^h, A.A. Rezyapkin^h

^a Geological Mapping Division, VNII Okeangeologia, Angliiskiy Prospekt, 1, St. Petersburg 190121, Russian Federation

^b Levinson-Lessing Earthcrust Research Institute, St. Petersburg State University, Universitetskaya Naberezhnaya, 7/9, St. Petersburg 199034, Russian Federation

^c St. Petersburg Department of the Steklov Mathematical Institute, Naberezhnaya Reki Fontanki, 27, St. Petersburg, Russian Federation

^d Dip. Georisorse e Territorio, Università di Udine, Via Cotonificio 114, I-33100 Udine, Italy

^e Ecological Geology Department, Faculty of Geology, St. Petersburg State University, Universitetskaya Naberezhnaya, 7/9, St. Petersburg 199034, Russian Federation

^f Laboratory "Magma and Volcanoes", University of Blaise Pascal, 24 Avenue des Landais 63177-AUBIERE Cedex, Clermont-Ferrand, France

^g Logic Department, Faculty of Philosophy, St. Petersburg State University, Mendelevskaya Liniya, 5, St. Petersburg 199034, Russian Federation

^h Karpinsky All-Russia Geological Research Institute (VSEGEI), Sredny Prospekt, 74, St. Petersburg 191126, Russian Federation

ARTICLE INFO

Article history:

Received 31 October 2006

Received in revised form

22 December 2007

Accepted 14 January 2008

Keywords:

Event bush method

Natural hazard

Bayesian belief network

Reasoning research

Complex formal approach

Semantic

Graph

Logic

Volcanology

ABSTRACT

The event bush is a new formalism for organizing knowledge in various fields of geoscience, particularly suitable for hazard assessment purposes. Acting as an intermediary between expert knowledge and the well-established field of Bayesian belief networks, the event bush allows at the same time a variety of other applications, linking geoscientific knowledge to the field of artificial intelligence and uniting probabilistic, deterministic, and fuzzy approaches. In this paper, we present basic principles, mathematical formulation, guidelines for application, and examples, including the connection with Bayesian belief networks. Further development of the method will include spatial and temporal modelling, implementation in mapping in GIS medium, formalization by means of predicate logic, definition of variable states in BBNs by membership functions based on the event bush semantics, and other applications.

© 2008 Elsevier Ltd. All rights reserved.

* Corresponding author at: Levinson-Lessing Earthcrust Research Institute, St. Petersburg State University, Universitetskaya Naberezhnaya, 7/9, St. Petersburg 199034, Russian Federation. Tel.: +79213175686, +78127144042; fax: +78127141470.

E-mail addresses: pshenich@kp1306.spb.edu, pshenich@pochtamt.ru (C.A. Pshenichny), sergey@logic.pdmi.ras.ru, smartnik@inbox.ru (S.I. Nikolenko), roberto.carniel@uniud.it, carniel65@gmail.com (R. Carniel), paganov@hotmail.com (P.A. Vaganov), Z.Khrabrykh@opgc.univ-bpclermont.fr, zkh@mail.ru (Z.V. Khrabrykh), mukhachev3@mail.ru (V.P. Moukhachov), Victoriya_Akimova@vsegei.ru, logika81@yandex.ru (V.L. Akimova-Shterkhun), alex_rez@pisem.net (A.A. Rezyapkin).

¹ Temporary address: Instituto de Geofísica, UNAM—Universidad Nacional Autónoma de México, C.U. Coyoacán 04510 D.F. México, México.

1. Introduction

1.1. Evolution of graphic representation of reasoning in hazard assessment

The assessment of natural hazards benefits much from application of formal graphic notations showing the semantics of the field of knowledge, structure of data, and paths of inference. Purely qualitative hierarchically built event trees (Fig. 1a) used since the 1970s were replaced by probability trees (Newhall and Hoblitt, 2002) enabling numerical assessment (Fig. 1b; MVO, 2006), and

these, in turn, are giving way to quantitative multi-root notations, or polytrees. Polytrees may have more than one starting point and, hence, possess more complicated (not necessarily hierarchical) structure, which is in most cases much more adequate to the real structure of a field of knowledge, structure of data, and lines of reasoning.

1.2. State of the art of hazard assessment: Bayesian belief networks

A Bayesian belief network (BBN) is an artificial intelligence apparatus that performs Bayesian inference on polytrees (Pearl, 1988). The mathematics underlying BBNs is presented in Appendix A, here we simply touch some basic notions.

Nodes in a BBN denote possible processes, events, facts, laws, opinions, etc., going under the general term “variables”. In Fig. 2, variables are “Season”, “Seismicity”, “Extrusion rate”, “Rainfall”, “Pressurization”, “Dome collapse”, and “Dome size”. Arcs of the BBN graph mean dependencies between the variables. Once nodes A and B are linked by an arc, one may, according to the task and common sense, assign the meaning “parent” to node A, and “child”, to node B, or vice versa. The “parent–child” relation usually represents the cause–effect dependence explained, or implied, by some more or less objective rationale (“If there is a breakdown at the local power station, food may rot in my fridge”), noticeable coincidence (“Every time something goes wrong at the local power station, I have my food rotten in the fridge”), or just

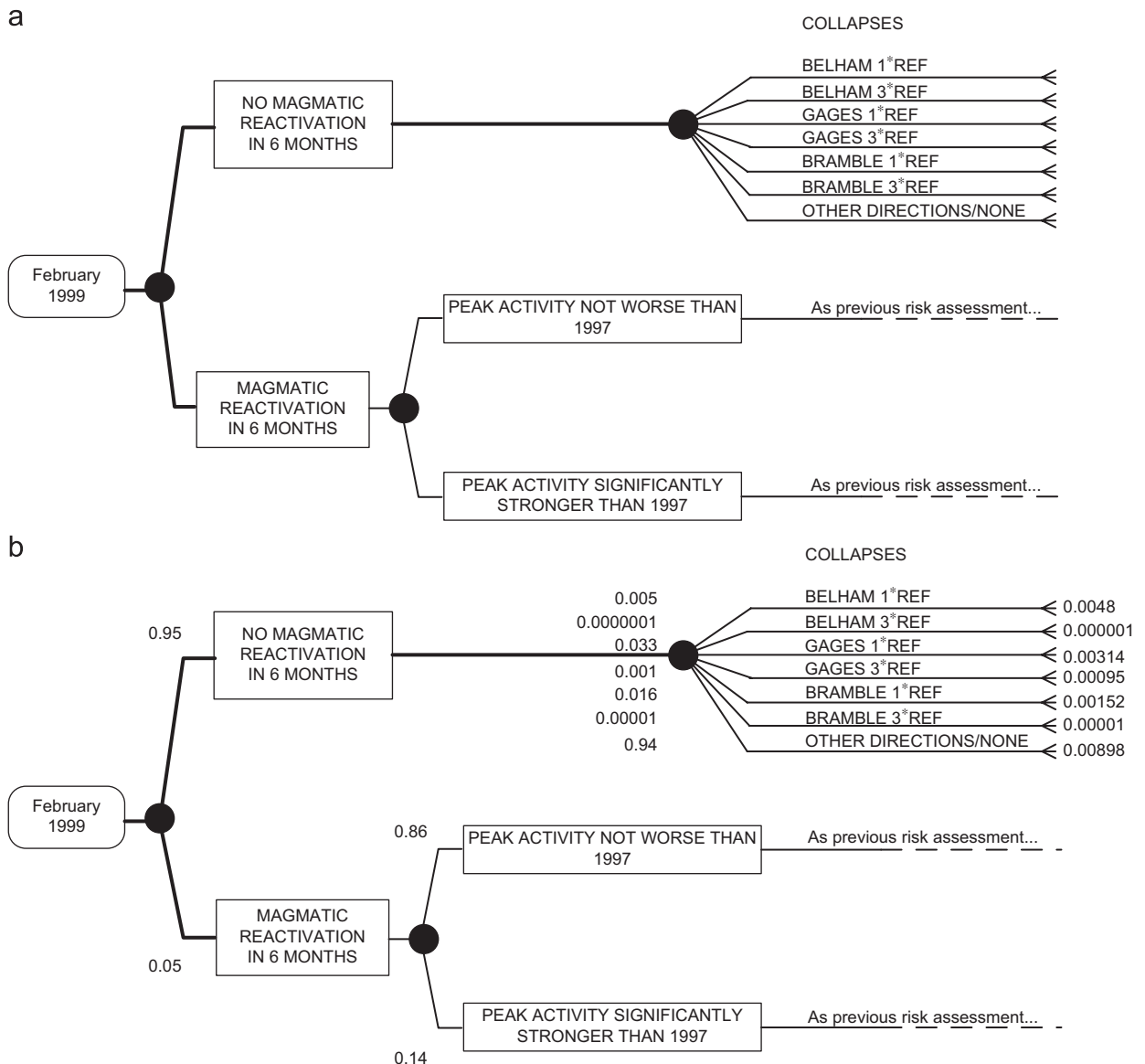


Fig. 1. Examples of event (a) and probability (b) trees for a period of 6 months from July 1998 of continuous eruption of Soufriere Hills volcano (modified from MVO, 2006).

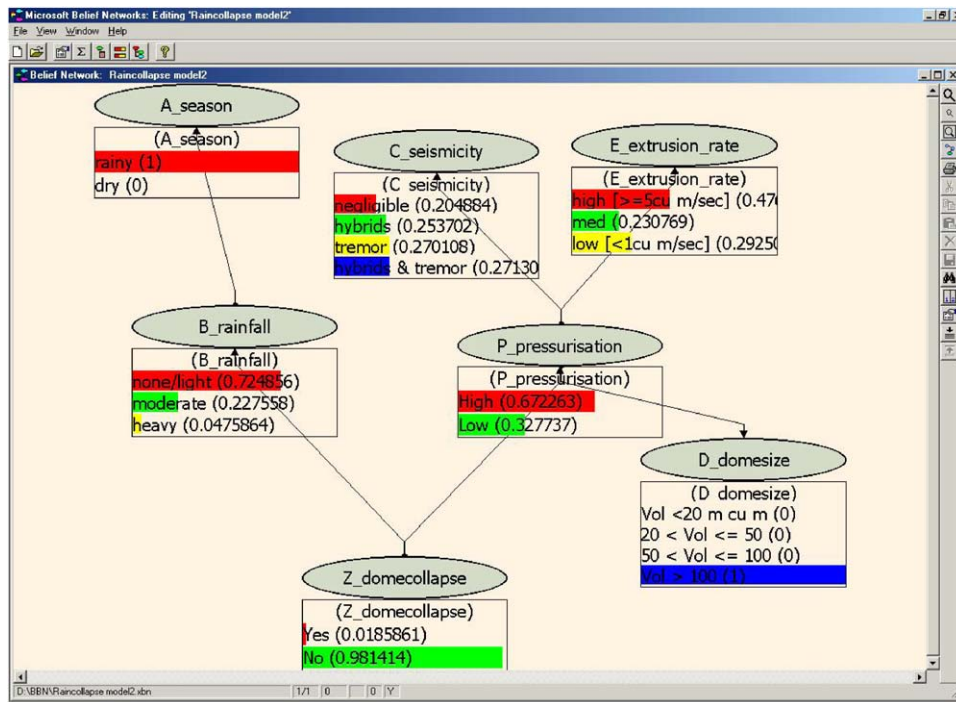


Fig. 2. Example of a Bayesian Belief Network (from Aspinall et al., 2003).

the way things are subjectively taken by human(s) (“If the food is rotten in my fridge, wasn’t there a breakdown at the local power station?”).

A node has a finite number of *states* that represent possible mutually exclusive outcomes of the corresponding variable/event. For example, in Fig. 2 the states of the variable “Rainfall” are “none/light”, “moderate”, and “heavy”, while they could be alternatively defined as “none”, “light to moderate”, “heavy”, and “shower”, or somehow else. In order to set up a BBN, one needs to specify conditional probabilities of a node with respect to its parents. For example, in Fig. 2 “Rainfall” (let us denote it *B*) is a child of “Domecollapse” (denoted below as *Z*), and therefore one should specify the following probabilities:

$$\begin{aligned} & p(B = \text{none/light} | Z = \text{yes}), p(B = \text{none/light} | Z = \text{no}), \\ & p(B = \text{moderate} | Z = \text{yes}), p(B = \text{moderate} | Z = \text{no}), \\ & p(B = \text{heavy} | Z = \text{yes}), p(B = \text{heavy} | Z = \text{no}). \end{aligned}$$

Obviously, the sum of probabilities over the whole family of mutually exclusive states should be equal to 1, e.g.,

$$p(B = \text{none/light} | Z = \text{yes}) + p(B = \text{moderate} | Z = \text{yes}) + p(B = \text{heavy} | Z = \text{yes}) = 1.$$

If a node has several parents, then conditional probabilities over the whole family of possible outcomes over these probabilities should be considered (see details and examples in Section 3.2 and in Appendix A). Parentless nodes, or the “roots” of the graph (“Dome collapse” in Fig. 2), have their probabilities set based either on

statistics (if appropriate data are available) or on intuitive expert judgments expressed as “subjective probabilities” (see, e.g., Woo, 1999). Pshenichny (2004) discusses a perspective of setting probabilities based on a family of alternative strict logical theories. Initial (*a priori*) probabilities of other nodes can be easily calculated by going top-down on the graph with the total probability formula.

Bayesian networks are commonly used for making predictions and diagnostics. A Bayesian network with all conditional probabilities set defines a (single) probability distribution over all of its nodes. It serves as a model of the body of knowledge about the nodes and their relationships. To use a Bayesian network, one usually sets some of its nodes to be known (observed, postulated) and recalculates the rest of the probabilities. This is precisely the diagnostic feature of BBNs: given some of the corollaries (symptoms), find out which of the possible causes have higher probabilities. At present, BBN is seen as the ground for a new technology of geohazard mapping in GIS medium (Felpeto et al., 2007).

Usually, the subject area imposes no formal restrictions on the structure of the graph apart from the lack of directed cycles. For instance, another researcher speculating about the weather, seismicity, magma extrusion and dome collapse would deny the link between the rainfall and dome collapse but insist on a link between the dome size and dome collapse. Moreover, someone may need, for instance, to take the rock composition into consideration deeming this variable valuable for other nodes, or the number of tourists having visited the lava dome, or their age and gender, and so forth. Again, no formal constraint

is put on the contents of the BBN. Thus, the structure of the BBN and many of its numerical parameters are usually left entirely for the experts.

1.3. Purpose and tasks of this paper

Such formal flexibility of the BBN calculus is very useful, because it allows us to incorporate any kind of information—data, intuitive expert judgments, models, and, possibly, other types of information. Thus, on one hand, we may use all the information that can give us a clue of future events in the hazard assessment procedure. However, on the other hand, these advantages turn to be shortcomings that erode the result of Bayesian computation. Indeed, one phenomenon in question can be described by infinitely many BBNs that differ either in the number of nodes, or sets of states of similar nodes, or graph structure, or the unsaid, but implied, meaning of the arcs (cause-effect, coincidence, or convenience of consideration), or in all these. The choice between alternative BBNs becomes purely the matter of personal taste and intuition. It is entirely possible to have two BBNs that are both plausible for experts but produce completely different results in similar situations. Definitions of nodes and their states, routing of arcs and convention on the nodes' meaning constitute the *semantics* of the domain of knowledge covered by a BBN (e.g., the domain “Lava dome collapse” in Fig. 2). Obviously, the semantics underlying known applications of BBN remains largely chaotic.

To cope with emerging problems and thus optimize the use of BBN, it looks promising to develop a method of handling information that would give a strict and finite but extendable display of a domain of knowledge and at the same time enable numerical assessment of hypotheses in this domain in terms of probability. In this paper, we present such a method, namely the event bush approach. For Bayesian inference, event bushes rely on BBNs, the latter being perfectly suited for calculations, with many good algorithms developed over the last 20 years (e.g., Jensen, 2001, and references therein). However, event bushes have a different structure that allows experts (especially in Earth sciences) to better understand the formalism, thus improving the chances of getting accurate judgments on numerical parameters and logical structure of reasoning.

2. The event bush approach to a domain of knowledge

2.1. General principles

Any domain of knowledge can be comprehended in a large, perhaps innumerable, number of ways. Descriptive narration, thesaurus, physical or chemical models, mathematical models are some but few of them. These approaches emphasize or obscure various properties of knowledge that refer either to its form (e.g., verbal or numerical) or to its contents (the reality it depicts). If some environment becomes an arena for various, intertangling processes, which are thought to originate from a limited number of causes or sources, tend to leave a

definable result and can be followed as more or less distinct scenarios (not necessarily complete), we believe there is a universal way to describe the contents of corresponding domain of knowledge. Apparently, most, if not all, of geohazards fall in this wide field. These are:

- earthquakes triggering slope processes, hydrologic and other effects;
- volcanic eruptions proceeding as a number of simultaneous or consequent “happenings” (phreatic explosions at the foot of volcano, a lava flow from the summit crater, a pyroclastic flow from the lava flow flank, continued ash venting, etc.);
- hydrologic and meteorologic cataclysms with diverse consequences;
- many other, especially large-scale and complex hazards.

This universal approach, henceforth denoted the *event/probability bush* (analogous to, but wider than, the well-known event/probability tree), implies organizing a domain of knowledge in a particular way (see Fig. 3). In the considered area of reality, we recognize:

- primary, non-overlapping and non-unique inputs (basic objects, processes or tendencies) that determine any further course of events (“happenings”);
- incoming circumstances that indicate the way the environment may affect basic inputs or influence their further, indirect manifestations, thus “shaping up” different “happenings”;
- secondary entities (processes or objects) that result from primary inputs with or without the contribution of incoming circumstances, and
- end results, or products, generated either by primary or by secondary processes, with or without incoming circumstances, that document the completed “happenings”.

In the domain of knowledge, (i) correspond to *primary statements*, (ii) to *environmental statements*, (iii) to *secondary*, and (iv) to *tertiary statements*. Perhaps, other types of

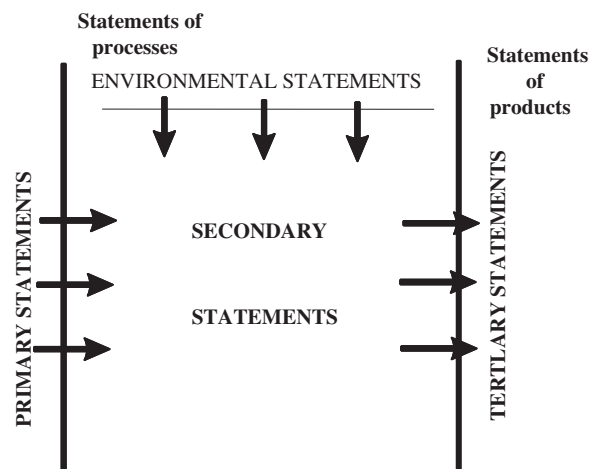


Fig. 3. Architecture of event bush.

statements (e.g., those describing some other type of results—say, quaternary ones) can be added, but the presence of the (i)–(iv) is mandatory. Their general interrelation is shown in Fig. 3. Two examples of an event bush drawn for different domains of knowledge are presented in Appendices B and C.

2.2. Organization and combination of statements

In accordance with the architecture of an event bush presented in Fig. 3, we set the following relations between (i) (ii), (iii) and (iv):

- statements (i) and (ii) must not lead to another (i) or (ii);
- statement (ii) may lead to a statement only together with (i) or (iii);
- statement (iii) must not lead to (i) or (ii);
- statement (iv) must not lead to (i), (ii), (iii), or another (iv).

If one or two statements directly lead to another statement or statements, we call this *elementary succession* (these can be grouped in complex ones). Elementary succession of statements reflects a cause-effect relation, which we believe to exist in reality, and takes the form of logical implication (an “IF... THEN...” compound statement; see, e.g., Kleene, 1952). We decide that the antecedent (the statement left to THEN) expressing the cause(s), *leads to* the consequent (the statement right to THEN) that expresses the effect(s).

In the event bush, statements (i)–(iv) are denoted as nodes (frames may be omitted for convenience), and successions, as arcs (Fig. 4). Each arc must end up either with an arrow pointing to a node, or with a rounded crossing (informally called “right turn”), or with a circle. The crossing means that two statements together (i.e., listed via AND) lead to another statement (i.e., “IF (A AND B),

THEN C”), and the circle, that the resulting statement is compound and consists of alternatives listed via OR (“IF A THEN (B OR C)”, e.g., “IF “Magma ascent (takes place)” THEN (“Lava doming” OR “Lava flowage”)”). This structure corresponds to the mathematical concept known as a circuit with corresponding graphic notation (Papadimitriou, 1994) and can be transformed into it as shown in Section 2.4. However, a circuit does not show the regularities of contents represented by the suggested architecture of event bush.

The connectives AND and OR used here cannot be directly related to conjunction (AND) and disjunction (OR) known in the classical symbolic logic (Kleene, 1952). To avoid confusion, we will term them as *combination* (AND) and *division* (OR). Only pair combination is adopted in the event bush. Aiming to combine three or more statements, one should first take any two of them and ensure this makes sense in the considered domain, otherwise the operation is not permitted. For example, combining the statements, (a) “Slump reaches the river”, (b) “Low terrace is wide”, and (c) “Organics is accumulating in the water”, one should first find out whether these lead to anything meaningful in pairs. Obviously, “(a) AND (b)” does, while “(a) AND (c)” and “(b) AND (c)” likely do not. Putting IF ((a) AND (b)) THEN (d), we mean by (d) “Flood develops”. Then, in turn, the combination “(d) AND (c)” (in this case it is equivalent to “(c) AND (d)”) starts making sense: IF ((d) AND (c)) THEN (e), (e) being “Fertile soil accumulates”. To warrant that such expressions are not nonsense, one should designate them in terms of the considered domain of knowledge. The designations here are the statements, “Flood develops” and “Fertile soil accumulates”. Thus the appearance of a new statement in succession appears dependent not only on the set of premises but also on the order of listing. Moreover, in each combination, we distinguish the main and the modifying members—e.g., in Fig. 4, “Lava flows” is the main, and the “Lateral direction”, the modifying, member. The sense of

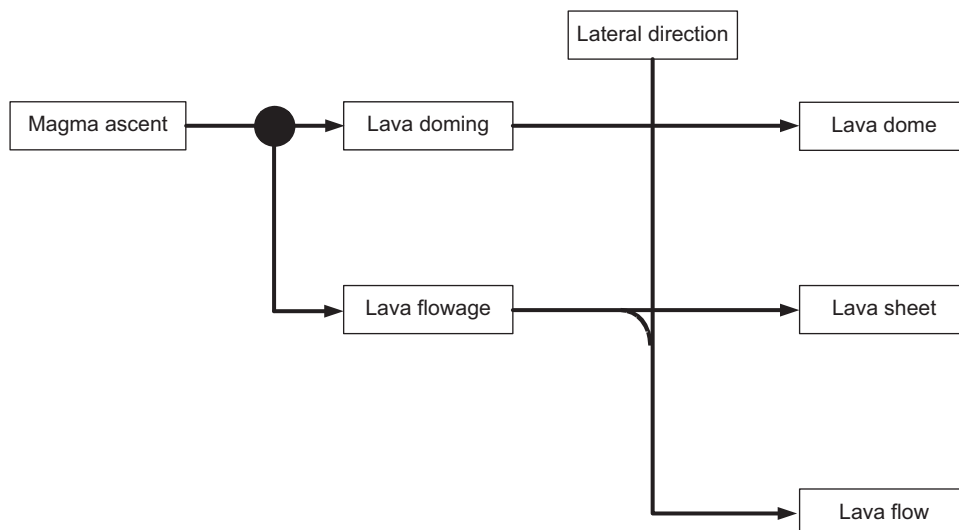


Fig. 4. A fragment of event bush describing an eruption of magma on land (see larger version in Appendix C).

this for the geoscientific contents is, “A (“Lava flows”) leads to B (“Lava sheet”), unless C (“Lateral direction”) is involved; together with C, A would lead to D (“Lava flow”)”. Obviously, all environmental statements (ii) act as modifiers. For intuitive clarity, we adopt that the rounded crossing points always down, from the modifying statement to the resulting one, which denotes the combination. The main and the modifying statements may change places, if this makes sense in the considered domain of knowledge only and no statements (ii) are involved. Then their designation may be different. Formally, this means that the combination of *A AND B* is not the same as that of *B AND A* in the event bush.

The operation OR, or division, in the event bush also has specificity dictated by the contents. Like in the probability theory, the resulting statements must comply with additivity condition, i.e., be mutually incompatible and exhaust the whole variety of possible outcomes. In Fig. 4, putting IF Magma ascends THEN (Lava flows OR Lava piles up in a dome) one postulates that no other outcomes but lava flowage or lava doming are possible if magma ascends and nothing else is given, and the same lava cannot take two options simultaneously.

Once a statement (i) or (ii) is introduced or statement (iii) appears in a succession, one should subsequently coincide it with all previously introduced statements to find out whether any new statement can result from each pair. Suggesting new statement, one should carefully manage its terms as discussed in Section 2.3.

Several arcs can leave one node. This means that a statement can lead to any number of other statements, in a combination with another one or by itself. For instance, in the event bush plotted in Appendix B, “Water contacts the interior of Quaternary deposits” simultaneously leads to “Water contacts cobbles in the interior of Quaternary deposits” and “Water contacts clayey sand in the interior of Quaternary deposits”, and in combination with “Clayey inliers form in the interior of Quaternary deposits”, it leads to “Water is retained on clayey inliers in the interior of Quaternary deposits”. However, drawing each next arc from a node, one should make sure that the newly introduced statement denotes a consequence really compatible with those denoted by the statements already derived from this node; otherwise a new arc must begin not from the node but from the existing arc, marked by division (circle).

2.3. Semantics of statements

In contrast to BBN, event trees, and other tools used in hazard assessment, the event bush method is provided with the guidelines for lingual and logical processing of terms and concepts.

From the point of view of Aristotelian logic, statements (i)–(iv) are simple, i.e., having one subject, one predicate, and no logical connective. Negation, the only connective that may apply to one pair of subject and predicate, is not directly involved in an event bush. Verbal expressions of subject and predicate, nevertheless, may include more than one word.

At present, we consider statements (i) and (ii) also *existential*, that is, their predicate is an *existential concept* like “exists”, “takes place”, “is”, etc., that can be omitted for brevity in verbal notation. For instance, verbally we may put just “thunderstorm” instead of “thunderstorm occurs”, and so forth. This might also appear useful in further logical formalization (see Section 4.2.5). Desirably, their subjects become subjects or predicates of at least some secondary and tertiary statements.

In addition, we suggest the following recommendations on the semantics of terms that join the statements of event bush.

For the sake of objectivity and transparency, terms should be descriptive, not explanatory. For example, one may put that conjunction “Loose sediments deposit” AND “Earthquake occurs” leads to “Gravity-driven current occurs”, but term “Sediment current” looks more appropriate than “gravity-driven”. In some cases, however, description can hardly be discerned from explanation (e.g., washing-out is a description of an observed process and, the same time, an explanation of changes in soil or rock).

Wording should be kept minimal in each statement. Each word must be term or preposition, irrelevant information should be rejected, and the relevant one related to particular terms sitting in particular statements. An example can be brought from volcanology. Aiming to put into an event bush the knowledge verbally expressed as “Erupted tephra settles down, and if the air moisture is high enough, accretionary lapilli form”, one should either operate with the term “tephra” generic for lapilli, ash, and bombs, and ignore accretionary lapilli, or introduce all three subtypes of tephra, which initially differ in size, i.e., quantitatively, and consider their *qualitatively* different behaviour (e.g., one settles down while another does not). Nevertheless, all the information from a “causal” statement that is important for the effect should be repeated in the newly introduced “effect” statement, irrespective of length. In the example given in Appendix B, the phrases “in the interior of Quaternary deposits” or “on the surface of Quaternary deposits” are repeated in many nodes or replaced with dots or abbreviated if no space is left.

New terms can be introduced to generalize different existing ones if the entities they describe behave in similar way—e.g., the term “lateral direction”, meaning one effect the topography and wind exert on transported material, in—see Appendix C. Also, one should not worry if some diagnostic, or classification, or just peculiar properties of processes or objects, which have been conventionally considered relevant, “escape” from an event bush (like, say, chemical composition of magma in the example in Appendix C). They should do so if they are not explicitly pronounced in the consequences. Our knowledge in many fields largely rests on tradition and convention, while the event bush, like other knowledge engineering tools, is supposed to give a more objective presentation. However, one has no limitation enriching the construction of event bush with additional signs. For instance, an information that is unused but considered interesting can be stored in annotation boxes that relate to nodes or arrows.

Synonyms should be avoided throughout the event bush. Thus, we should choose one from “current”,

“stream”, and “flow”, despite that all three are widespread in the scientific literature (see Section 2.4). Generic terms are preferred (e.g., “denudation” instead of “erosion” (by rivers), “abrasion” (by sea waves), “deflation” (by wind), and “grinding” (by glaciers), unless the domain covered by event bush includes only one particular type of denudation or, for some sake, the classification of denudation is to be incorporated into the event bush. However, the latter may cause problems better handled by another type of graphic notation, existential or conceptual graphs (Peirce, 1909; Sowa, 2000), still unused in geosciences, to the authors’ knowledge.

The occurrence of every term should be carefully traced throughout the event bush to avoid accidental use. Ideally, the list of cases in which given word/term may appear, is determined prior to composing the event bush. For instance, in Appendix C, the “lateral direction of” any material evokes the “irregular-shaped cover of”, while its absence (i.e., negation) leads to “concentric cover of”, the same material. Simultaneously, if a term appears only once or very few times, it should be well substantiated that the case it relates to is that rare or unique and cannot be named alternatively.

With further formalization of event bush (see Section 4.2.5), desirably, these recommendations will evolve into strict rules defining the normalization, or “states of maturity”, of an event bush, analogous to normal forms of databases (Codd, 1970) or propositional

logic formulae (e.g., Kleene, 1952). However, this is a subject for future research.

2.4. Mathematical definitions of the event bush formalism

Processing the natural-scientific knowledge, we distinguished between four types of statements that hold different positions in the event bush: primary, environmental, secondary, and tertiary. While this distinction is important for the considered domains of knowledge, in mathematical analysis we do not have to discern so many different entities. Mathematically, event bush is a simple *directed graph*. Both primary and environmental statements are nodes without parents, while tertiary ones are nodes without children. However, sometimes we will have to introduce new, purely operational, nodes in order to reflect the relations, or operations, of combination and division, which we previously depicted as special edge intersections.

Representing the combination, we will need two operational nodes: AND and NAND (x AND (NOT y)). The latter connective, obviously, is asymmetric with respect to its parents; in this case, the modifier (the second combined statement—e.g., an environmental statement) will always be y . This is nothing else but formal expression of the difference between the combination and logical conjunction considered in Section 2.2. With these preconditions, the event bush can be plotted as a circuit, as was mentioned in Section 2.2. Fig. 5 demonstrates as a

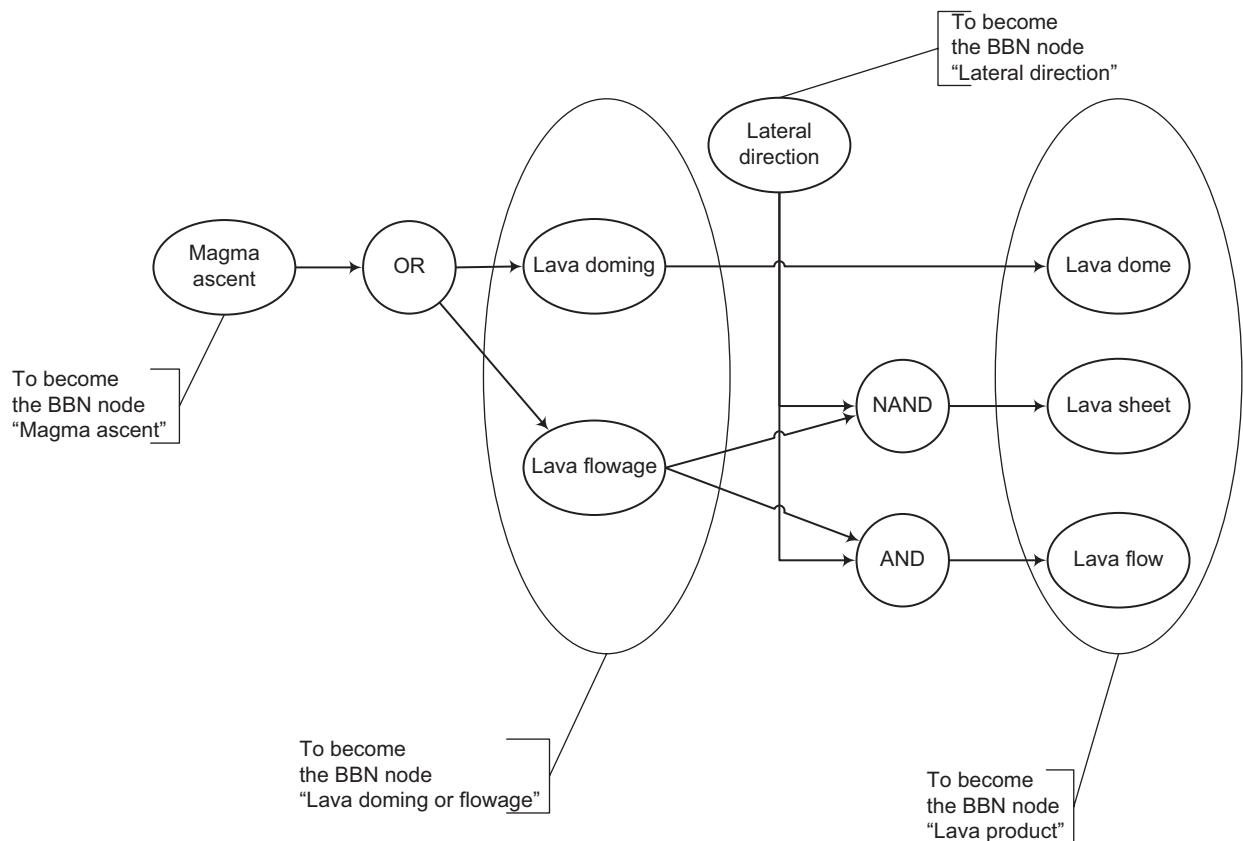


Fig. 5. Transformation of event bush into circuit for further translation into a BBN.

circuit ready to transform into BBN the example presented first as an event bush in Fig. 4.

Note that the graph became much messier. It is then advisable to design event bushes as shown before; this representation is to be used only for explaining the mathematics of an event bush.

3. Numerical assessment based on the event bush

3.1. Numerical assessment by probability bush

If weights are assigned to its nodes, the event bush transforms into *probability bush*. Weights represent conditional probabilities (which can be statistical, subjective, or logical—see Woo (1999)) of the events given their parents. Numerical assessment usually aims at calculating

probabilities of the events denoted by the participating statements, sometimes with some of the events already given (observed or postulated).

Mathematically, it makes no difference which formalism to choose for Bayesian inference: a probability bush or a BBN. However, the intuitive clarity of the event bush approach may be helpful to visualize the pathway of reasoning that underlies the computation of values. We will demonstrate an alternative computation by probability bush (Fig. 6a) and BBN (Figs. 6b and c) taking the same example as in Figs. 4 and 5.

Values of probability are initially attributed to primary and environmental statements. Arrow transmits the value from cause to effect. Combination of statements A and B is interpreted as the product of probabilities of events A and B. The probability of the consequence of event A unaffected by B event equates to the product of

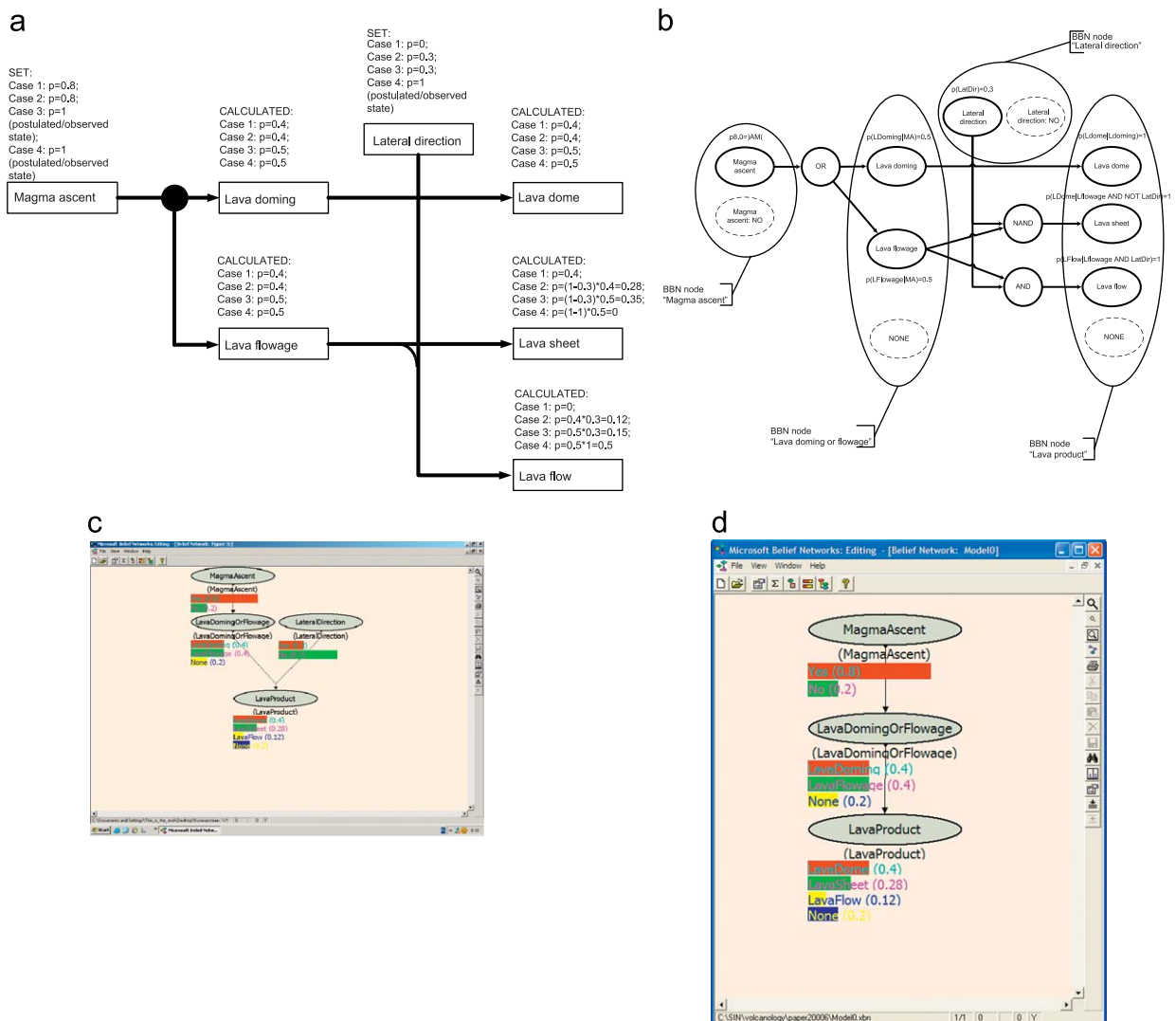


Fig. 6. Computation of conditional probabilities by probability bush (a), putting variable states and conditional probabilities in BBN built on circuit (c) and a poorer BBN with same values of states (d), which could be easily built without, but could never be built with, event bush shown in Fig. 4—mind absence of variable "Lateral Direction" in this BBN.

probability of A and probability of not-B, i.e., $\text{probability}(A) \times (1 - \text{probability}(B))$. The operation of division means subtraction of probabilities: given the probability of one alternative is p , that of the sum of probabilities of other alternatives is $1 - p$. These rules are demonstrated below (Fig. 6a).

Case 1: Magma supposedly ascends (e.g., based on seismic and geodetic data), lateral direction (i.e., topographic control) is postulated absent. The latter may mean that the surface is leveled or valleys are negligible compared with expected volume of magma.

Let the probability of magma ascent be 0.8, and the relative probability of its effusive vs. extrusive behaviour are equal for simplicity (say, the composition of magma, its viscosity, volatile content, etc., may favour either scenario). In other words, the odds ratio for both the lava doming and flowage given the magma ascent = 1. Then the probabilities of lava doming and flowage are $0.8 - 0.4 = 0.4$, and the probabilities of appearance of lava dome and lava sheet are 0.4, and the probability of lava flow is 0.

Case 2: Magma supposedly ascends, the probability estimated as 0.8, and the topography might influence the motion of lava once it appears. The probability of the lateral direction is estimated as 0.3.

(Importantly, this literally means that we consider the chance that a topographic slope will control the motion of *the entire volume* of lava that reaches it as 0.3. This *does not* mean that 30% of lava will be laterally directed in this place. The latter belongs not to probabilistic but to fuzzy reasoning, which also can be incorporated in the event bush, but in a different way that is considered in Section 4.)

Given these values, the probabilities of lava doming and flowage remain the same, 0.4 each, the probability of lava dome is also the same, 0.4, but the probability of lava flow is the product of probabilities of lava flowage and lateral direction, $0.4 \times 0.3 = 0.12$. Hence, the probability of appearance of lava sheet (considered as unaffected by lateral direction) is $(1 - \text{Probability}(\text{Lateral direction})) \times \text{Probability}(\text{Lava flowage}) = (1 - 0.3) \times 0.4 = 0.28$.

Case 3: Magma is already seen on the surface (e.g., on the crater floor), but its behaviour remains uncertain. The probability of magma ascent is 1 (observed state). Let the probability of lateral direction remain the same as in the previous case, i.e., 0.3.

Lava doming and flowage are again assumed equally probable. Then their probabilities will be 0.5, following the same rules of calculation; the probabilities of formation of lava dome, lava sheet, and lava flow will be 0.5, 0.35, and 0.15, correspondingly.

Case 4 describes the case that magma has appeared on the surface and the topography is uneven enough for us to be totally sure it *will* influence the lava motion, if only the latter “decides” to flow and, hence, reaches the slope. Here both the primary and environmental statements represent observed or postulated states and are attributed the probability 1. Then the probabilities of secondary statements are distributed like in the previous case (0.5 each), the probability of lava dome is 0.5 (if lava piles up and, hence, does not reach the topographic irregularities, it

definitely forms a dome), that of lava sheet is 0 (if lava flows, it definitely reaches the topographic irregularities, and the latter are so big that there is no way for a sub-isometric lava sheet to form), and the probability of lava flow is 0.5.

In any case, the sum of probabilities of alternatives must be 1, be these the secondary statements, Lava doming or Lava flowage, or tertiary ones, Lava dome, Lava sheet, and Lava flow. However, this condition is not always explicitly met (e.g., sum up the probabilities of Lava dome, Lava sheet, and Lava flow in Case 2). This shortcoming is handled by the BBN (see below).

3.2. Numerical assessment by Bayesian belief network

In the BBN, the operational nodes can be viewed as regular nodes with predefined conditional probability tables. For example, a binary AND node with two binary parents A , B would have a conditional probability table

$$p(C = \text{yes} | A = \text{yes}, B = \text{yes}) = 1, p(C = \text{yes} | A = \text{yes}, B = \text{no}) = 0, \\ p(C = \text{yes} | A = \text{no}, B = \text{yes}) = 0, p(C = \text{yes} | A = \text{no}, B = \text{no}) = 0.$$

We also note that the conditional probability table of every secondary event and end result except for the connectives and children of an OR node is very simple: it is 1 in case the parent event did happen, and 0 otherwise. Were it not for the OR nodes, inference in event bushes would be trivial and would only involve propositional logic. The semantics of the OR node is “either one child happens with probability p or the other”; the children of this node are not independent variables.

Except for the case when several arcs go out of one node, which point to compatible options, every set of consequences from one set of causes are mutually exclusive in the event bush. If we were to build a BBN on the same information, we would have these results as states of the same node. It should be noted here that perhaps the only semantic requirement the BBN poses on the contents of a field of knowledge is the additivity condition that should be kept in definition of variable states. It urges us to put a question to every variable/node, whether a state “no”, or “none of the alternatives”, makes sense.

Let us exemplify this mathematical description by rewriting the previously shown circuit (Fig. 5) derived from an event bush (Fig. 4) with proper variable states, the weights borrowed from Case 2 considered in Section 3.1 (Fig. 6a), but still with the connectives of combination and division (Fig. 6b). The input (parentless) nodes, similar to the event bush, are “Magma ascent” and “Lateral direction”, both binary (yes/no). The two statements “Lava doming” and “Lava flowage” are mutually exclusive, and they both follow from “Magma ascent”. In the BBN formalism, this would correspond to a single child of “Magma ascent”, but this node should have three possible states: “Lava doming”, “Lava flowage”, and “None” (no volcanic activity of this kind). “Lateral direction” comes into play on the next level, and partially determines which of the consequences we shall have. The consequences are also mutually exclusive; thus, they are also

considered as different states of the same BBN node (we now have a node with four possible states: three listed outcomes and an empty outcome “None”).

It is now rather straightforward to assign conditional probabilities. As we estimated in previous sections, $p(\text{MagmaAscent} = \text{yes}) = 0.8$. Since we considered “Lava Doming” and “Lava Flowage” to be equally likely in case of magma ascent,

$$p(\text{Lava Doming or Flowage} = \text{Lava Doming} | \text{MagmaAscent} = \text{yes}) = 0.5, \\ p(\text{Lava Doming or Flowage} = \text{Lava Flowage} | \text{MagmaAscent} = \text{yes}) = 0.5.$$

If no magma ascent happened, then certainly there will be no effect, so conditional probabilities of lava doming and lava flowage with this condition are zero, and the probability of “None” (no effect) is equal to 1:

$$p(\text{Lava Doming or Flowage} = \text{Lava Doming} | \text{MagmaAscent} = \text{no}) = 0, \\ p(\text{Lava Doming or Flowage} = \text{Lava Flowage} | \text{MagmaAscent} = \text{no}) = 0, \\ p(\text{Lava Doming or Flowage} = \text{None} | \text{MagmaAscent} = \text{no}) = 1.$$

It remains to set up the probabilities for the outcome node; this is done in the table depicted in Fig. 6b. Fig. 6c, shows Case 2 from Section 3.1 in a form of Bayesian inference. Note that in the BBN, due to the explicitly put state “None” of the node/variable “Lava product”, the sum of probabilities is equal to 1, as it used to be.

4. Discussion

4.1. Current applications of the event bush method

4.1.1. Event bush and the BBN

We see that the event bush as proposed in the previous sections is easily and completely represented as a Bayesian belief network, and we hope this task can be automated in near future. Mutually incompatible nodes of an event bush are taken as states of one node of a BBN, compatible nodes of an event bush are provided with two states (“yes” and “no”), operations of combination and division are transformed into operational nodes to determine the semantics of BBN states, and, when the latter is set, are substituted by normal BBN arcs. Other arcs of the BBN directly inherit those of an event bush.

A BBN built on the basis of event bush may have important distinctions from a “blindly built” BBN. Nonetheless, these distinctions cannot be defined formally. Let us consider our running example of the event bush (Fig. 4). In Section 3, we discussed a way to translate it into an equivalent BBN. Note that we could, for example, get rid of the “Lateral Direction” node in the BBN, recalculating the conditional probabilities accordingly to incorporate its 0.3 probability. In this case, we get a simplified BBN depicted in Fig. 6d.

Note, however, that this simplified BBN lacks the possibility to actually set the value of the “Lateral Direction” node (insert hard evidence). This means that with the BBN from Fig. 6d, one cannot make a distinction, e.g., between a volcano with barrancos on slopes controlling possible motion of lava and a volcano without barrancos; one can only reason with the general case

of 0.3 probability in mind. Thus, we can trade expressive power for BBN complexity.

An event bush better represents the logical flow of the expert knowledge and can serve as a natural and powerful intermediary between expert knowledge and BBNs: on an event bush it is easier to reason, spot mistakes, and make corrections, and then it is straightforward to transform an event bush into a BBN to make inference; the latter task may be automated. This is an important direction for our future work.

A BBN built from the event bush is warranted to have proper organization of contents—complete and non-excessive set of relevant nodes with clearly defined states, and arcs representing the cause-effect relations. Besides, event bushes may incorporate more expressive features, and corresponding BBNs may grow so complicated that they would be extremely hard and error-prone to create by hand.

4.1.2. Event bush and geoscientific knowledge

Building an event bush, a scientist is supposed, in fact, to repeat the evolution of his domain of knowledge, but in a formalized framework. It is a combination of deductive and inductive reasoning: one suggests a set of primary and environmental statements, then generates secondary and tertiary statements from them, but may inductively repeat this procedure many times and is able to retrospectively correct previously introduced nodes and arcs (arrows). We can back up getting rid of the OR nodes philosophically: if there are no OR nodes, it means that we completely understand what is going on; there is no situation of random choice in this case. Replacing an OR node with, for instance, another environmental circumstance that will allow to choose between two possibilities, brings us closer to understanding how nature works.

Deductive component ensures completeness (all consequences can be traced), inductive, the finite (“rounded”) character of the created context. Thus, the event bush can serve as a model of a field of knowledge representative for the type of knowledge we have in the Earth sciences (and alike areas). It is easier for an expert to follow the lines of reasoning, understand what is going on and spot mistakes. Besides, for this reason an event bush is good for educational and discussion purposes and possesses a great potential for generalizations.

It should be emphasized that an event bush is not once and for all limited by the construction we have described in this paper. Below are some but few ideas how this method can evolve.

4.2. Lines of future research

4.2.1. Implications for temporal modelling

We have demonstrated by the example of annotation boxes that any additional information can be assigned to the nodes of event/probability bush. In particular, these can be intervals of time expressed in any form, from seconds to biostratigraphic units, paleomagnetic epochs or phases of tectonism. Then, theoretically, nothing prohibits to make this information work—e.g., introduce

some additional rules, what information can and what cannot be attributed to particular nodes, say, to two nodes, one of which denotes the cause, and the other, the effect. For instance, the cause must begin earlier, and the time intervals of cause and effect must overlap. The operations of unit and division also can be provided the temporal sense. Then, if there are some external constraints on time intervals, this gives an independent control on the construction of event/probability bush.

Furthermore, such “enchronised” bush automatically supplies us with the definitions of states, which is a long-lasting problem in investigation of so-called processes with memory (Carniel et al., 2008) including the Markov chain modelling (Aspinall et al., 2006) and time-series analysis (Carniel and Di Cecca, 1999), which are, according to Carniel et al. (2008), of particular significance for monitoring and natural hazard assessment. Any succession of nodes in an event bush gives us a memory of the process in question, and the probabilities of transition from one state to another calculated from the observation data can enter the probability bush/BBN.

It is noteworthy that attempts to introduce time in BBN usually result just in a multitude of copies of the basic network. We have suggested a different approach, still convertible into BBN, but, hopefully, more efficient, that will be reported elsewhere.

4.2.2. Implications for spatial modelling

Similar to time (one-dimensional value), the space (two- or three-dimensional value) can be introduced into event bush. Likewise, this can impose special rules on formulation of the nodes according to their interrelation and, vice versa, setting arrows representing these interrelations according to the definition of nodes.

Simultaneously, such “regionalised” event/probability bushes could be studied by well-developed tools of geostatistics (Matheron, 1962; Chilès and Delfiner, 1999).

The result, again, is expected to be readily translatable into BBN.

4.2.3. Implications for mapping

With the co-ordinate values, the information put in the event/probability bush can be mapped. Modern means of geoinformation systems (GIS) allow layer-by-layer mapping of various information (ESRI, 1995–2006), and relations can be set between the nodes of the bush and layers or themes of the GIS. This might solve a problem widely discussed by applied specialists in GIS, whether there exist, or can be thought of, some formal constraints on the list of layers to use in a GIS project. We believe that, given an event/probability bush on the topic of the project, necessary and sufficient kit of GIS layers could be defined through the association of the layers with the nodes of the bush.

4.2.4. Implications for ontology design

The very fact that nodes of event/probability bush can act as a reference, or “raw bones” of a domain of knowledge, to which a variety of information can be attached—layers of GIS, spatial or temporal coordinates, synonymic terms, bibliography, comments,

and so forth—suggests that these nodes represent a kind of metadata and may serve to formally organize a domain of knowledge, or, in other words, to become its *ontology*. Up to now, ontologies appeared in a number of fields, from commercial activities (e.g., Omelayenko, 2001) to fundamental science, such as chemistry and biology (see, e.g., Bersini, 2000). In the geoscience, ontologies have been developed and used for hydrocarbon exploration (POSC, 1994–2006), evaluation of mineral deposits and general geological mapping purposes (Loudon, 2000; Bonham-Carter et al., 2003; GeoReference Online, 2003–2006) and geotechnical hazard assessment issues (Smyth, 2003). However, to the authors' knowledge, these ontologies represent simple classifications as the Aristotelian logic takes them and, in this sense, stay close to event trees; it seems likely, for the reasons described in Section 1, that the bush as a multi-root notation is a much better option for ontology design.

4.2.5. Implications for metadata/knowledge formalization

Admitting that statements of event bush can play the role of “crystallization nodes” for the modelled domain of knowledge, it seems reasonable to enquire how far we can and should advance in formalization of geoscientific knowledge after event bush. By absolutely formal theory we mean classical logic proper (Kleene, 1952; Hilbert and Bernays, 1934; Takeuti, 1975, and others). Mathematical theories fully recordable in terms of predicate logic and enabling deductive inference, which is the virtue of logic (Gentzen, 1934; Takeuti, 1975), may be considered the limit of formalization of non-logical contents. Such theories can be formally, or even automatically, verified to satisfy the requirements of self-consistency, completeness, deducibility and independence of axioms (Hilbert and Bernays, 1934), which are, strictly speaking, a must for any scientific knowledge.

Concepts of geosciences are still quite far away from such formalization, excluding, possibly, some physical models, which, strictly speaking, reside in the field of physics and are only applied in geoscience as well as in other fields. These issues are discussed in detail by Pshenichny (2004).

One opinion is that formalization of geosciences can be achieved through wide application of mathematical tools (Vistelius, 1992), which the BBN belongs to. However, these tools, like physical models, treat a geologic or geographic object “externally”, assuming it identical to some mathematical (physical) entity. This deprives the geoscience of its own sense and makes it rather emasculated than formalized. Alternatively, logic could be directly applied to geoscientific contents (Pshenichny et al., 2003), with mathematical, physical and other modelling playing only secondary and auxiliary role. The claim for “internal”, or direct, formalization of geoscientific contents led to the appearance of new discipline, the Reasoning Research in the geoscience (Pshenichny and Moukhachov, 2001; Pshenichny, 2002; Reasoning Research, 2003–2006). This discipline aims to create a complex formal approach to treatment of geoscientific information (Pshenichny, 2003), which is thought to be based on logical formalization of the contents and

determine the application of other formal methods, including probabilistic and possibilistic (fuzzy).

In contrast to this, most of the geoscientific knowledge, unlike mathematics, physics or chemistry, is not apt to logical formalization as is and needs preliminary reworking by means of knowledge engineering to become more compact and “solid” (Pshenichny, 2004). Knowledge representation in the form of event bushes looks an appropriate option. Hence, an important direction of future research is translation of the event bush into the language of predicate logic. We expect that primary statements with an existential predicate may provide individual variables, and environmental statements, the basic predicates for logically formalized theories. The definition of terms for such theories will hopefully evolve from its present, semi-intuitive form (see Section 2.3) into more strict and formal rules.

Along with theoretical benefits, this may have important practical implications for setting the values of prior probabilities in Bayesian computations, when these values cannot be obtained from statistics—e.g., in case of rare or never-happened events. At present, such values are provided by intuitive expert judgments. The quality of these judgments is reasonably criticized both by mathematicians and practitioners (Yudkovsky, in press). Another approach may be provided by logical formalization of expert knowledge through event bushes.

Once an event bush is meant to express an expert opinion, its logical formalization would mean translation of expert statements into a strict language. Then, assuming a number of experts have agreed on a certain kit of primary, environmental, secondary, and tertiary statements, their opinions (individual or collective) would result in a family of bushes that differ only in composition of arrows. If some statements are denied by some experts they should nevertheless be present in the corresponding event bushes but should be assigned zero probability or, in terms of logic, negation of any implication involving them.

Logical formalization transforms a family of bushes into a family of strict theories that operate with similar terms (consider geometries of Euclid, Lobachevsky, Gauss, and other possible ones that operate with similar objects—points, lines, angles and so forth, but postulate different relations between them). Pshenichny (2004) demonstrated that the Bayes rule is fully applicable to a set of premises (in the basic case, primary and environmental statements) of a given corollary in a family of strict theories operating with similar terms. Then the probability of one set of premises could be calculated by multiplying its conditional probability given the corollary A and the statistically obtained probability of A (which may be considered as a stronger or weaker support of the theories in which A is inferred).

The obtained value could be attributed to other corollaries of this set. Hence, if some corollaries of a set of premises are statistically verified, the probabilities of others, describing even rare or unique events, hopefully will become computable.

For illustration, let us consider a hypothetical assessment of probability of pyroclastic flow on an active volcano. It may suffice us to collect expert judgments of

this probability and use the subjective numbers they provide. However, an alternative that is foreseen is to treat expert statements qualitatively, convert them into event bushes, build a number of strict theories (naturally, not larger than the number of opinions, and at least some of them will imply pyroclastic flows) and try to infer the events from these theories, the probabilities of which are statistically known, e.g., those of some types of seismic unrest. Then, coming backward, one could compute the logical probabilities of particular theories implying this type of unrest and then attribute these probabilities to the corollary “Pyroclastic flow occurs”. If more than one value appear, they can be treated in a number of ways, e.g., in a belief–plausibility fashion following the Dempster–Shafer’s reasoning (Shafer, 1976).

The Dempster–Shafer theory, by the way, can be used directly to compute values of events compliant with a number of corollaries, e.g., tephra fall, which may result from eruption column rise, pyroclastic flow motion, or other causes.

In the debate on subjective probability, many scientists strive to get totally rid of subjectivity and claim for a method to provide absolutely impartial values. In the authors’ view, the subjectivity in natural sciences is rooted in the selection of basic terms (e.g., in formulation of primary and environmental statements of an event bush) and therefore cannot be removed, only reasonably decreased. This will be hopefully done by the suggested approach to expert judgment processing by means of event bush. However, this is an issue of further research.

4.2.6. Account for natural diversity

A substantial argument counter application of classical logic to geoscientific contents is wide natural diversity and variability of the object of study, which is considered by some authors a specific type of uncertainty in Earth sciences (Bardossy and Fodor, 2004). Once we put, e.g., two end results, lava flow and lava dome (see Figs. 4–6), we have to keep in mind that an infinite number of transitional varieties between the two are possible. As was pointed in Section 3.1, we should discern herewith two lines of reasoning. Saying that the probability of lateral direction (or topographic control) is estimated as 0.3, we literally mean that the chance that a slope will control the motion of *entire volume* of lava is 0.3, and the only expected result is normal lava flow. At the same time we may say, “this slope will likely affect the motion of 30% of lava”—i.e., the other 70% of lava are thought to form a sub-isometric dome. Such rationale, accounting for natural variability of objects, pertains to the fuzzy reasoning (Zadeh, 1965) that is explicitly or implicitly present in the geoscience as two-dimensional, triangular, rectangular, and tetrahedral diagrams (see Fig. 7a, with an example of triangle diagram from Irvine and Baragar, 1971), numerous “transitional cases” in classifications, non-explicit definitions and the like, often being a source of epistemic uncertainty (Pshenichny, 2004).

A typical example is a non-explicit (visual) definition widely used in petrology, e.g., that of calc-alkaline and tholeiitic magma compositions between the three endmembers represented by the apices of the triangle in

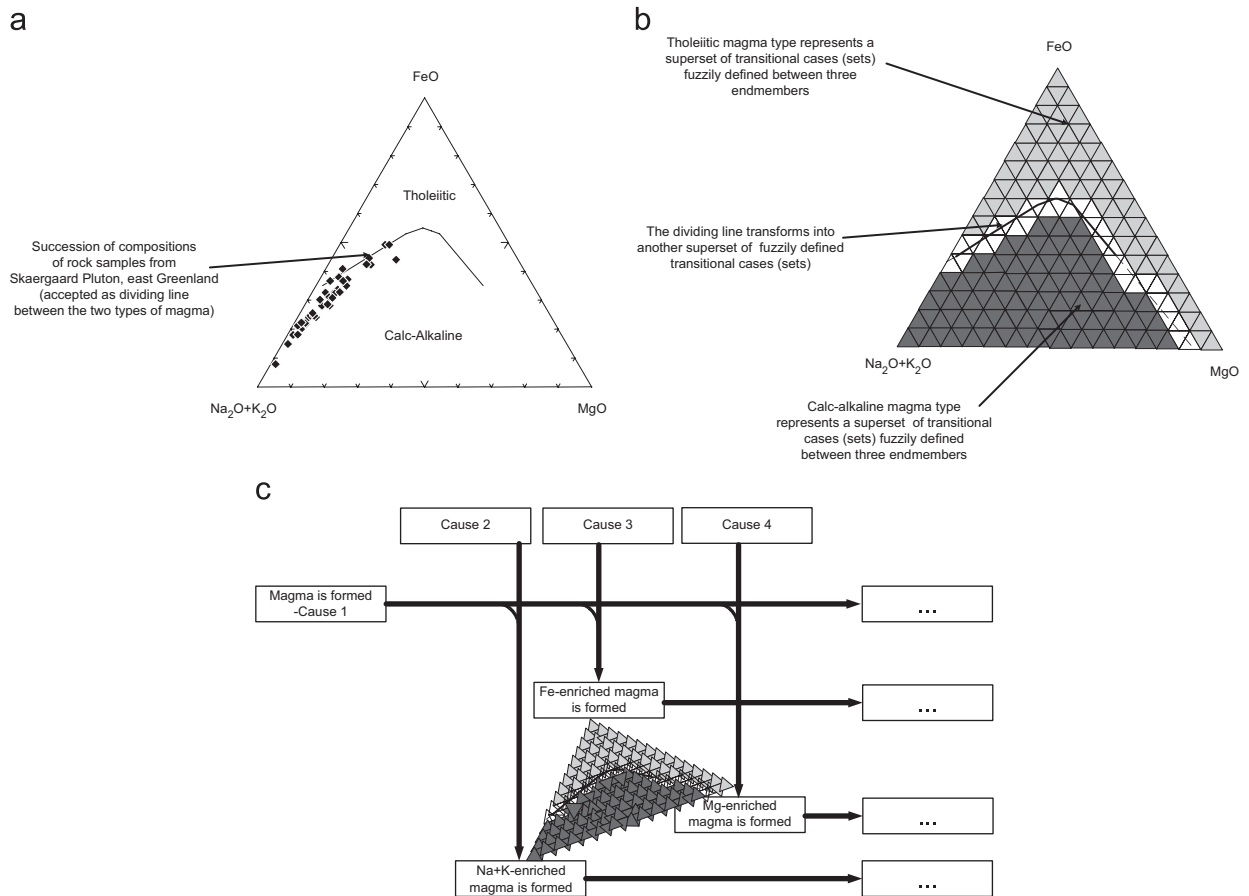


Fig. 7. Formalization of continuous spectrum of magma compositions by introducing fuzzy approach in event bush; a—initial semi-intuitive representation in a form of triangle diagram from Irvine and Baragar (1971); data from Pshenichny (1998); b—a more formalized representation by event bush with fuzzy definition of transitional cases between nodes by a membership function; c—putting this representation into a hypothetic event bush.

Fig. 7a. In practice, the diagram presented in this figure is used as a working definition. Actual chemical data obtained from magmatic rocks can be plotted on it (see Fig. 7a) to find out, which type, calc-alkaline or tholeiitic, the rocks in question belong to.

Formally, we suppose that this definition implies two, still unmade, steps, (i) highly fractional division of the entire field of compositions into transitional cases between the crisply defined endmembers (see Fig. 7b) and definition of these cases as fuzzy sets by a membership function (Cagnoli, 1998; Bardossy and Fodor, 2004, and others), which is based, in the simplest case, just on the pace of fractioning, and (ii) postulation that some group (superset) of these cases and endmembers represents one entity (e.g., calc-alkaline magma), those fuzzy sets, which include the dividing line form another entity (transitional not between the endmembers but between the calc-alkaline and tholeiitic magmas), and, hence, the remaining superset, yet another (tholeiitic magma). These entities are filled in Fig. 7b, dark grey, white and light grey, correspondingly.

Fig. 7c, shows how the information from a “fuzzified” triangle diagram can be incorporated into an event bush. Endmembers of the diagram become nodes, possibly,

undergoing reformulation as suggested in Section 2.3. Furthermore, the calculation of Bayesian probabilities can be easily extended onto the transitional cases, if only the membership function that defines them provides discrete intervals as required by the BBN. Each of these cases must be added to the set of states of corresponding node in the BBN. For instance, referring to the example with lava flowage/lava doming, we may put the states of the “Lava Flowage or Doming” node (and, consequently, the values of membership function) not just “Lava Flowage”, “Lava Doming”, and “None”, but also, say, “1–30% Lava Flowage—99–70% Lava Doming”, “30–50% Lava Flowage—70–50% Lava Doming”, “50–70% Lava Flowage—50–30% Lava Doming”, “70–99% Lava Flowage—30–1% Lava Doming”, and, possibly, a number of transitional cases between each of these endmembers and “None”. Then we may put, e.g., “the probability of that 30–50% of lava is directed by topography is 0.72” (based on statistics, modelling, or expert’s gut feeling). Getting back to the example of definition of states of the “Rainfall” node (see the BBN in Fig. 2 and comments in Section 1), now we can affirm that these states could be better defined based on a hard semantic foundation and using a membership function.

Thus, the probabilistic and possibilistic (fuzzy) approaches work together in the general framework of event bush. The usefulness of combined probabilistic–possibilistic approaches in the geoscience has been emphasized by some researchers (e.g., Bardossy and Fodor, 2004), and we believe that the method suggested here can evolve into a complex formal approach covering a wide range of tasks with further logical formalization of event bush.

4.2.7. Extended knowledge modelling

Formalization of different domains of knowledge by event bush (see Appendices B and C) puts a natural question, how different bushes can be related to each other. Obviously, each event bush is a complicated logical structure, and the interrelations between different bushes can be understood only with sufficient logical formalization. However, some clues can be already offered.

Event bushes can be united “on plane” and “in space”. “On plane”, this can be done “in line” (tertiary statements of one bush become primary or environmental ones of another), in a simple network (tertiary statements of one bush become primary, and tertiary statements of another, environmental, of the third bush), and in a complex network (tertiary statements of one bush become primary or environmental statements of more than one bush; or vice versa, one bush takes primary or environmental statements from more than one bush). For instance, an

event bush describing the sheet glaciation at the North-East of Europe in Late Pleistocene may have, inter alia, tertiary statements “Flat surface of Quaternary deposits forms” or “Layer of Quaternary clayey sand with cobbles forms”. These may be primary statements in another bush, describing the evolution of the resulting landscape in the Holocene (see Fig. 8).

“In space”, primary, environmental, secondary, or tertiary statements of one bush can become secondary statements of another. For instance, an event bush describing the mass transfer in the atmosphere can incorporate some secondary statements from the bush of a volcanic eruption (like “Fragments of lava and vapor join cloud”; see Fig. 9). Similar or other statements of the event bush from Fig. 9, e.g., “Fragments of magma and vapor join cloud” and “Flow of fragments of magma and vapor undergoes rheomorphism”, may also appear as secondary statements in a bush describing the sediment currents of various origins.

These interrelations, which still need to be explored and understood, must give us a better vision of knowledge in the geoscience and hopefully arm us with an efficient tool to handle it.

5. Conclusions

In conclusion, let us review what we have done in this paper. Our work shows that the event/probability bush

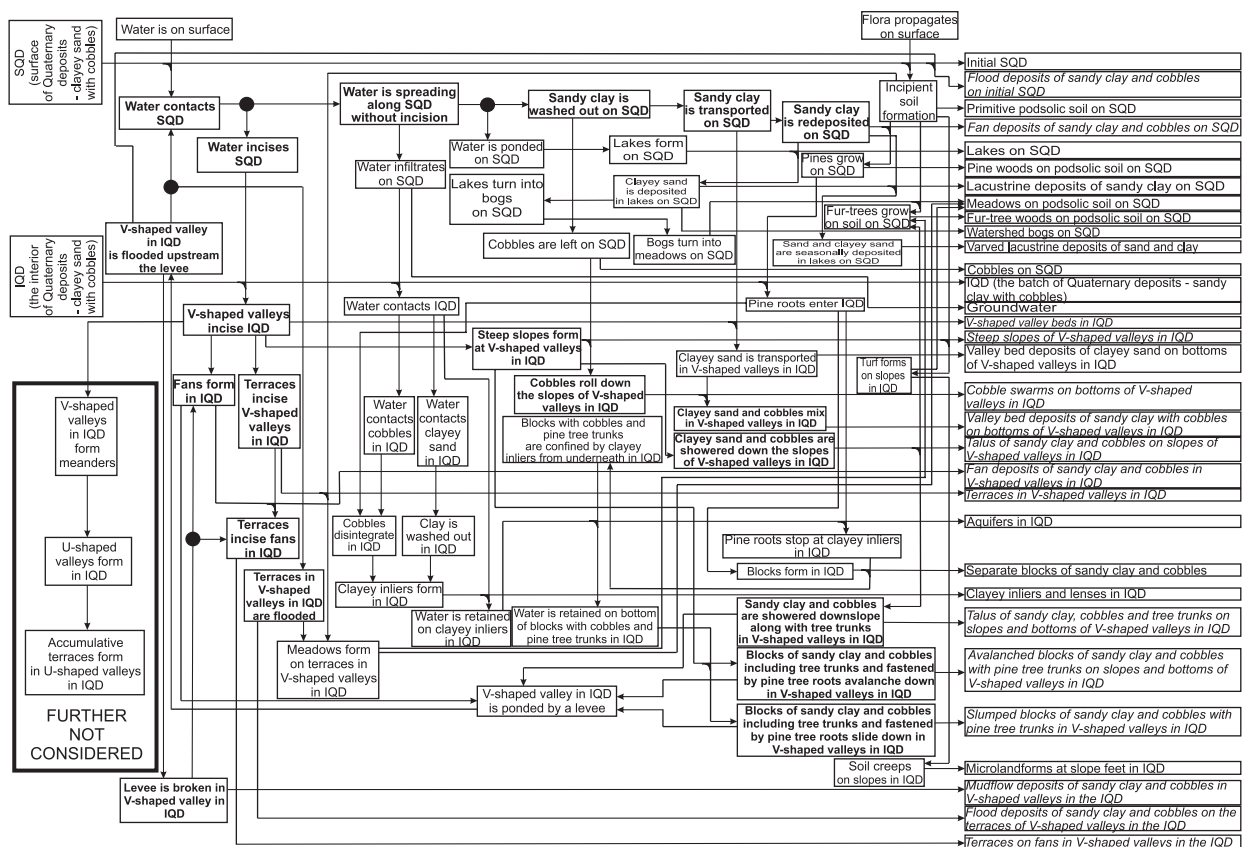


Fig. 8. (Appendix B) An event bush describing exogenic transformation of a Quaternary sequence represented by clayey sand with cobbles.

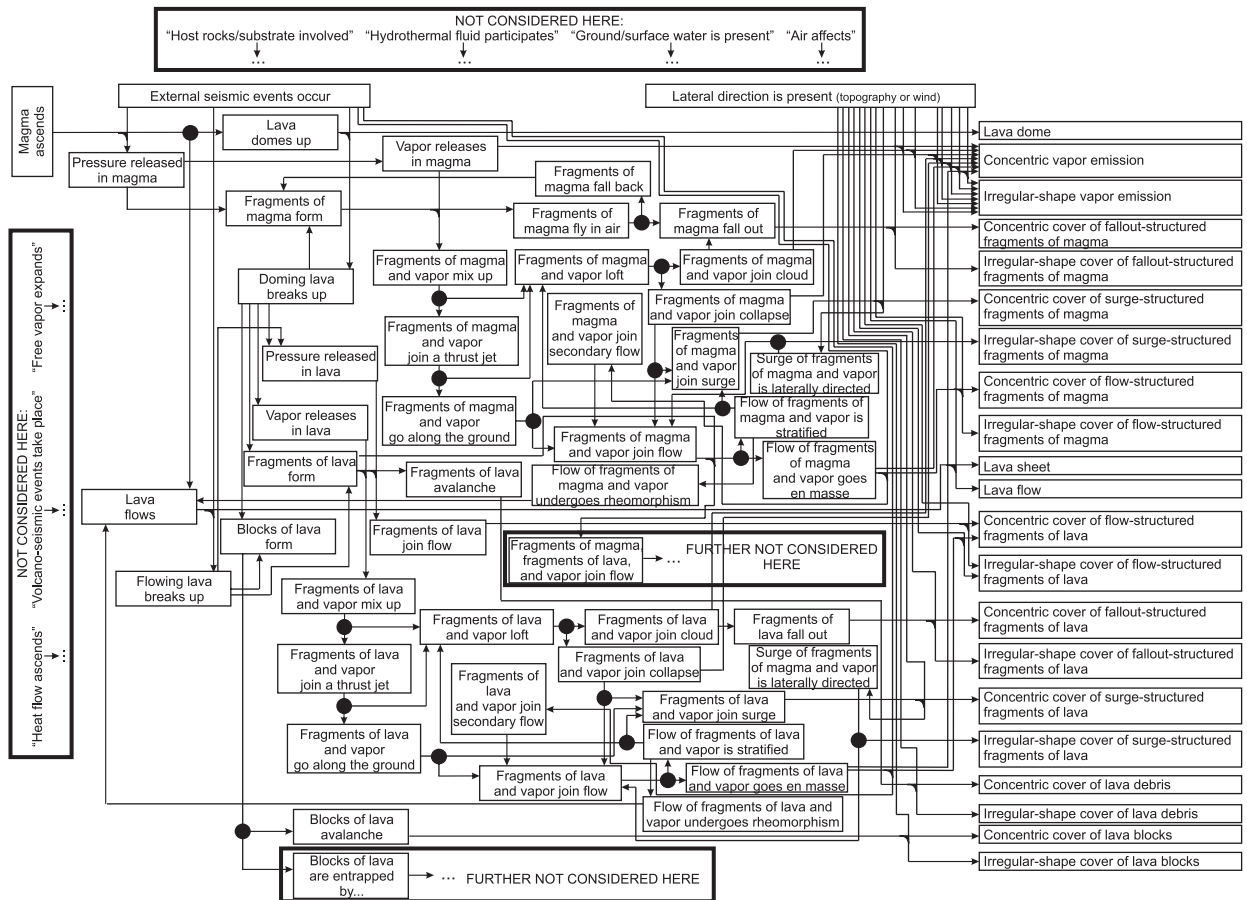


Fig. 9. (Appendix C) A fragment of event bush describing magmatic eruption on land accompanied by external seismic activity.

approach is a novel method of handling information in the geosciences. It has several advantages over the knowledge-handling techniques that are currently used in geoscientific applications.

First, the event bush gives a strict and finite but extendable display of a domain of knowledge; we can use as such various domains of knowledge in geology, geography, hazard assessment, and related fields. We have specifically designed event bushes to perfectly fit the needs of geosciences and have always kept in mind how practitioners can set up and maintain this framework.

Mathematically speaking, the event bush formalism performs Bayesian inference via the use of Bayesian belief networks. Being founded on a well-defined and deeply studied BBN formalism, the event bushes do not suffer from any common shortages of new artificial intelligence techniques: they are formally defined, mathematically solid, easy to reduce to known apparatus, and have simple and effective inference algorithms. As a result, event bushes allow for numerical assessment of hazards' probabilities (and/or degrees of belief).

Note that event bushes have a broad potential for incorporating spatial and temporal data. The formalism is

potentially applicable to mapping tasks in GIS media, and our immediate plans for further work include creating real-life event bushes, which will necessarily include spatial (and, perhaps, temporal) data.

Finally, event bushes may unite, with further development, the deterministic, probabilistic, and fuzzy approaches to geoscientific modelling and hazard assessment. Thus, we believe that event bushes are a very efficient new tool of reasoning research in the geosciences. Development of the event/probability bush method opens a new area of research at the intersection of Earth sciences and artificial intelligence.

Acknowledgements

The authors are deeply obliged to many colleagues whose timely advice and constructive criticism coupled with moral support fuelled our enthusiasm and sharpened our thought. In particular, we would like to thank Peter Kokelaar, Augusto Neri, Susanna Falsaperla, Andrew Curtis, Rachel Wood, Stephen Henley, Alex Sobishevitch, Jurgen Neuberg, Clinton Smyth, George Bardossy—and especially Willy Aspinall, whose unseen presence is felt

every time the BBN matters are concerned. We are also thankful to the reviewers and guest editors for their interest, enthusiasm, and will to make our inventions fit the publication format that sometimes was not easy at all.

The research has been funded by Project V4 of Istituto Nazionale di Geofisica e Vulcanologia, Italy, “Development of quantitative tools to evaluate volcanic hazard”, Human Potential Foundation (“Nauchny potentsial”) grant for development of information systems for geosciences, and NATO-Russia collaborative linkage Grant 982484.

Appendix A. Bayesian belief networks

Bayesian belief networks (Pearl, 1988) are widely used in different applications of artificial intelligence. In this appendix, we briefly describe the basic definitions of the BBN theory; we cover here neither inference algorithms nor any proofs; just the basic facts necessary to use BBNs in practical applications.

A Bayesian belief network is a directed graph with nodes representing variables (events), edges representing dependency between nodes, and weights on the nodes representing conditional probability tables of the node given its parents for each value of the parents. So, if a node x with values v_i , $i = 1 \dots n$, has parents $y_1, y_2, \dots, y_m$ with values u_j^i , $i = 1 \dots n_j$, $j = 1 \dots m$, then at this node we should specify $(n-1)n_1n_2 \dots n_m$ probabilities of the form

$$p(x = v_i | y_1 = u_1^i, \dots, y_m = u_m^i).$$

The first factor is $(n-1)$, not n , because the conditional probabilities $p(x = v_i | \dots)$ should sum up to 1 for each set of values of the parents.

A BBN represents a set of conditional dependencies and independencies. These independencies are formalised through a notion of d -separation (d is for “directional”). Two variables x and y are conditionally independent given z if knowledge about x does not provide any extra information about y if z is already observed. d -separation allows to formalise the independency data a network’s structure provides.

When we have set up a BBN, we can use Bayesian inference to calculate *a priori* probabilities of different events in the network. The d -separation and conditional independence definitions are set up in such a way that the overall probability distribution over all variables of the network can be calculated as the product of all local conditional probability tables. Then it is easy to project the overall distribution to get any *a priori* probability (of course, the actual algorithm does not do this directly and is much more efficient, but mathematically it is equivalent).

The main use of a Bayesian network is that it allows to use inference for calculating conditional *a posteriori* probabilities in case several of the variables have been instantiated (that is, their values have been fixed). The algorithm used in this case is similar to the one calculating *a priori* probabilities and involves passing messages between nodes of the network with

recalculated probability tables (usually called potentials in this context).

BBNs are often used for diagnostic purposes. Nodes without parents represent causes, and nodes without children represent symptoms; the task is, given several symptoms, to find the most probable cause.

Appendix B. An example of event bush: exogenic transformation of a quaternary sequence represented by clayey sand with cobbles

An event bush (Fig. 8) was composed to reveal the group of potentially hazardous geologic processes that can occur at exogenic transformation of a Quaternary sequence built of clayey sand with cobbles. Such deposits are common in many regions of the world, e.g., in Northern Europe. Hence, there is wide class of objects (i.e., localities or regions), to which this bush may apply. Secondary statements denoting potentially hazardous processes are marked bold, and tertiary statements denoting the corresponding results, italicized.

Primary statements include (i) the surface of deposits and (ii) the interior of deposits. This division appeared reasonable when the environmental circumstances were considered (see below), because the terms required to describe their influence on the surface and interior appeared quite different. However, a terminological trap occurs when we put, e.g., “Streamwater incises the interior” that, by common sense, should lead to the formation of slopes. Following the semantic rules (see Section 2.3), we should keep the relevant information from parental statements and put, “Slopes form in the interior”. This sounds somewhat unusual, but to write “Slopes form on surface”, we have to introduce the substitution “interior → surface” that would make the bush much more complicated and brings little information itself.

This consideration led us to the solution, “surface” and “interior” should denote fixed, immovable entities, and “surface” should mean the initial, “starting” surface that became the arena for modern geologic and geomorphologic processes. Any new parts of surface appear at place of the interior of Quaternary deposits.

However, a question that immediately appears is, whether it is possible to find this “starting point” marking the “initial” division between the surface and interior acceptable for our event bush? To find such point would actually mean to mark the onset of modern period of history of corresponding regions. In historical geology, this point, or period, is known at the beginning of the Holocene, when glacial or, depending on the region, another Pleistocene environment changed to more or less modern one, at least in Northern Europe. Thus, a merely terminological consideration resulted in a constraint on the contents, particularly the age of objects of considered class. Having decided this, we may put, “Slopes are formed in the interior”. This is well compliant with other effects that take place in the interior and may be pronounced on newly forming slopes, e.g., groundwater discharge that triggers rapid slope processes, and so forth.

In the variety of landscape processes, we believe that only two can be figured out, which, for given deposits, are simultaneously

- (i) basic (e.g., groundwater is not basic, for it was derived from surface water) and
- (ii) directly affect the deposits (e.g., solar radiation may affect solid rock favouring its disintegration, but unlikely it will notably affect sand or even cobbles; rather, it affects other environmental processes; fauna may affect the deposits, but mainly in the soil layer, which is essentially the product of vegetation, which, in turn, controls the presence of fauna species).

So, the two circumstances were selected and formulated as existential statements, “Water is present” and “Vegetation propagates”.

Unfortunately, the format of the paper does not allow to present the bush entirely; we refrained from consideration of many possible “happenings” (see notes in rectangle in Fig. 8). Full version of the bush and the experience of its construction will be reported elsewhere.

Appendix C. An example of event bush: magmatic eruption on land accompanied by external seismic activity

Volcanic eruption on land is a class of very complex phenomena, involving a diverse range of various processes. Aiming to describe its essential principles by event bush, we took an oversimplistic case (see Fig. 9) of magmatic eruption under conditions of external seismic activity and lateral direction of erupted material on land (by topography) and in air (by wind). Disregarded were such important primary inputs as seismic events triggered by ascending magma, heat flow from the magma chamber, and expansion of free vapour, and the environmental circumstances—involvement of host rocks, hydrothermal system of volcano, ground and surface water, and air that favours rapid cooling of erupted material.

Yet in this, unrealistically trivial, case, the diversity of possible events appeared so wide that we failed to fit the publication format describing possible scenarios of eruption and twice had to truncate the generation of new statements (see Fig. 9). Still, some observations valuable for the practice of building event bushes can be made from this graph. First, it appeared not easy to discern between the primary inputs and environmental impact: e.g., seismic events can be (i) purely external, unrelated to, but influential on, the volcanic system, and (ii) generated by this system.

Moreover, the (i) can trigger the (ii); then the volcano-seismic events cannot be considered really primary input. This point will be dealt with in detail in future publications.

Another observation is that, like in event trees (Newhall and Hoblitt, 2002), some successions of nodes have to be cloned with substitution of only one term—e.g., the succession of statements involving “fragments of lava and vapour” largely repeat those with “fragments of magma”.

However, this reflects the natural diversity of volcanic processes and their products.

The size and complexity of the event bush describing the magmatic volcanic eruption on land urges to think about appropriate forms of presentation of event bushes on computer screen, which would make them user-friendly and adaptable to different tasks and queries.

References

- Aspinall, W.P., Woo, G., Voight, B., Baxter, P.J., 2003. Evidence-based volcanology: application to eruption crises. *Journal of Volcanology and Geothermal Research* 128, 273–285.
- Aspinall, W., Carniel, R., Jaquet, O., Woo, G., Hincks, T., 2006. Using hidden multi-state Markov models with multi-parameter volcanic data to provide empirical evidence for alert level decision-support. *Journal of Volcanology and Geothermal Research* 153 (1–2), 112–124.
- Bardossy, G., Fodor, J., 2004. Evaluation of Uncertainties and Risks in Geology: New Mathematical Approaches for Their Handling. Springer, Berlin, 221pp.
- Bersini, H., 2000. Object-Oriented Chemistry <<http://iridia.ulb.ac.be/sf/abshug.html>>.
- Bonham-Carter, G.F., Smyth, C., Kerswill, J., 2003. Semantic net approach for mapping mineral potential: a feasibility study. In: Cubitt, J., Whalley, J., Henley, S. (Eds.), *Proceedings of IAMG 2003 on Modeling Geohazards*, Portsmouth, UK, on CD.
- Cagnoli, B., 1998. Fuzzy logic in volcanology. *Episodes* 21 (2), 94–96.
- Carniel, R., Di Cecca, M., 1999. Dynamical tools for the analysis of long term evolution of volcanic tremor at Stromboli. *Annali di Geofisica* 42 (3), 483–495.
- Carniel, R., Jaquet, O., Tarraga, M., 2008. Perspectives on the application of the geostatistical approach to volcano forecasting at different time scales. In: Gottsmann, J., Marti, J. (Eds.), *Caldera volcanism: Analysis, modelling and response*, *Developments in Volcanology*. Elsevier, pp. 516–531.
- Chiles, J.P., Delfiner, P., 1999. *Geostatistics: Modelling Spatial Uncertainty*. Wiley Series in Probability and Mathematical Statistics, New York, 695pp.
- Codd, E.F., 1970. A relational model of data for large shared data banks. *Communications of the ACM* 13 (6), 377–387.
- ESRI, 1995–2006. Environmental Systems Research Institute, Inc. <<http://www.esri.com/index.html>>.
- Felpeo, A., Marti, J., Ortiz, R., 2007. Automatic GIS-based system for volcanic hazard assessment. *Journal of Volcanology and Geothermal Research* 166 (2), 106–116.
- Gentzen, G., 1934. Untersuchungen über das logische Schliessen, I–II. *Mathematisches Zeitschrift* 39, 176–210, 405–431, translated as (Gentzen, 1969), *Investigations into logical deduction*. In: Szabo, M. (Ed.), *The Collected Papers of Gerhard Gentzen*. North Holland Publishing Co., Amsterdam, pp. 68–128.
- GeoReference Online, 2003–2006. GeoReference Online Ltd. <<http://www.georeferenceonline.com>>.
- Hilbert, D., Bernays, P., 1934. *Grundlagen der Mathematik* (Fundamentals of Mathematics), vol. I. Springer, Heidelberg, 322pp.
- Irvine, T.N., Baragar, W.R.A., 1971. A guide to the chemical classification of the common volcanic rocks. *Canadian Journal of Earth Sciences* 8, 523–548.
- Jensen, F.V., 2001. *Bayesian Networks and Decision Graphs*. Springer, New York, 268pp.
- Kleene, S.C., 1952. *Introduction to Mathematics*. D. Van Nostrand Co., Inc., New York, 550pp.
- Loudon, T.V., 2000. Geoscience after IT (special issue). *Computers and Geosciences* 26 (3).
- Matheron, G., 1962. *Traité de géostatistique appliquée* (Treatise of Applied Geostatistics). Tome I. Editions Technip, Paris, 334pp.
- MVO, 2006. The Montserrat Volcano Observatory <<http://www.mvo.ms>>.
- Newhall, C.G., Hoblitt, R.P., 2002. Constructing event trees for volcanic crises. *Bulletin of Volcanology* 64, 3–20.
- Omelayenko, B., 2001. Syntactic-Level Ontology Integration Rules for E-commerce <<http://citeseer.ist.psu.edu/omelayenko01syntacticlevel.html>>.
- Papadimitriou, C., 1994. *Computational Complexity*. Addison-Wesley, Reading, MA, 500pp.
- Pearl, J., 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Francisco, CA, 552pp.

- Peirce, C.S., 1909. Existential Graphs. Manuscript 514. Houghton Library, Harvard University.
- POSC, 1994–2006. Petrotechnical Open Standards Consortium, Inc. <<http://www.posc.org>>.
- Pshenichny, C.A., 1998. Petrologiya i eruptivnaya istoriya ignimbritov vostochnoi chasti Badjal'skoi vulkanicheskoi zony (Habarovskiy Krai)—(Petrology and eruptive history of ignimbrites of the eastern part of Badjal Volcanic Zone, Khabarovsk Area, Russia). Ph.D. Dissertation. St. Petersburg State University, St. Petersburg, 185pp.
- Pshenichny, C.A., 2002. Investigation of geologic reasoning as a new objective of geoscience and geohazard assessment. *Earth Science Computer Applications* 17 (11), 1–3.
- Pshenichny, C.A., 2003. A draft for complex formal approach in geoscience. In: Cubitt, J., Whalley, J., Henley, S. (Eds.), *Proceedings of IAMG 2003 on Modeling Geohazards*, Portsmouth UK, on CD.
- Pshenichny, C.A., 2004. Classical logic and the problem of uncertainty. In: Curtis, A., Wood, R. (Eds.), *Geological Prior Information*, vol. 239. Geological Society, London, pp. 111–126 Special Publications.
- Pshenichny, C.A., Moukhachov, V.P., 2001. Volcano-Logic: Reasoning Research in Volcanology, vol. 1. IAVCEI (International Association of Volcanology and Chemistry of the Earth's Interior) News, 2pp.
- Pshenichny, C.A., Moukhachov, V.P., Khrabrykh, Z.V., 2003. Logical assessment of observational knowledge in volcanology. *Journal of Volcanology and Geothermal Research* 128, 287–298.
- Reasoning Research in the Geosciences, 2003–2006. <<http://www.jiscmail.ac.uk/lists/GEO-REASONING.html>>.
- Shafer, G., 1976. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ, 312pp.
- Smyth, C., 2003. Distinguishing partonomies from taxonomies in science languages: a prerequisite for computer-aided georeasoning. In: Cubitt, J., Whalley, J., Henley, S. (Eds.), *Proceedings of IAMG 2003 on Modeling Geohazards*, Portsmouth, UK, on CD.
- Sowa, J.F., 2000. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Brooks/Cole Publishing Co., Pacific Grove, CA, 594pp.
- Takeuti, G., 1975. *Proof Theory. Studies in Logic and the Foundations of Mathematics*, vol. 81. North-Holland Publishing Co., Amsterdam, London, American Elsevier Publishing Co., New York, 372pp.
- Vistelius, A.B., 1992. *Principles of Mathematical Geology*. Kluwer Academic Publishers, Dordrecht, Boston, London, 477pp.
- Woo, G., 1999. *The Mathematics of Natural Catastrophes*. Imperial College Press, London, 292pp.
- Yudkowsky, E., in press. Cognitive bias potentially affecting judgment of global risks. In: Bostrom, N., Cirkovic, M. (Eds.), *Global Catastrophic Risks*. Oxford University Press, USA.
- Zadeh, L., 1965. Fuzzy sets. *Information Control* 8, 338–353.