Qualitative Spatial Representation and Reasoning About Fold Strata

Yuta Taniuchi and Kazuko Takahashi
School of Engineering, Kwansei Gakuin University, 1, Gakuen, Uegahara, Sanda, 669-1330, Japan
{hcs57846, ktaka}@kwansei.ac.jp

Keywords: Qualitative Spatial Reasoning, Knowledge Representation, Logical Reasoning, Shape Information.

Abstract: We propose a method of handling strata in qualitative spatial representation. We make a model for a typical fold structure projected onto a two-dimensional plane extracted by a rectangle. It is expressed by a pair of sequences of symbols that represents the strata configuration and the shape of the layers, respectively. We define the validity required of the representation and show that the representation and the model have a one-to-one relation. Moreover, we define operations on the representation, such as rotation and symmetric transitions, and show that validity is preserved. We also show that global data can be constructed by connecting local data. This method can provide a logical explanation of the processes involved in strata-generation prediction, which in the field of structural geology have been examined manually to date, and find results that manual analysis may overlook.

1 INTRODUCTION

Temporal changes in a landscape have a strong relationship with the occurrence of natural disasters. Thus, to predict future events, such as landslides, earthquakes or river flooding, there is a need to elucidate the formation processes of specific landscapes and the causality of the changes involved. To consider morphological changes, investigation of strata is an essential component.

In structural-geology research (Kano and Murata, 1998), the shapes and structures of strata are analyzed using data at various scales, from the micro level, such as collected small sample data measured in tens of centimeters, or slices thereof that can be observed by microscopy, to the macro level at the out-crop scale of several-hundred meters, or aerial photos of larger regions. Regardless of scale, the entire shape of a stratum is estimated by integrating local data collected from multiple locations, since in a real landscape it is rare for an entire stratum to be exposed. Since human error may affect this process. a method that can overcome this shortcoming is required.

In this study, we propose a novel approach that uses qualitative spatial reasoning (QSR), which is a subfield of artificial intelligence. QSR represents spatial entities symbolically without using concrete numerical data, and enables reasoning on the representation (Cohn and Renz, 2008; Chen et al., 2013; Ligozat, 2011; Sioutis and Wolter, 2021). Representation focuses on specific aspects or properties of an object or the relation of objects, depending on the user's purpose, such as mereological relations, the relative positions or directions of objects, rough shapes, and on on. Avoiding the need for precise values enables a small computational burden, and declarative representation suits human recognition. So far, lots of works have been done depending on the focused aspects of spatial data.

To apply QSR to the shapes of strata, there are two primary requirements: one layer continues in one direction if there is no fault, and the relations of inter-layer connections remain unchanged even if a stratum rotates or bends.

Although it is rather difficult to consider shape in QSR, several researchers have proposed handling the shape of an object by projecting it onto a two-dimensional plane (Cohn, 1995; Falomir et al., 2013; Galton and Meathrel, 1999; Kulik and Egenhofer, 2003; Kumokawa and Takahashi, 2008; Leyton, 1988; Cabedo and Escrig, 2004; Cabedo et al., 2010; Pich and Falomir, 2018; Tosue and Takahashi, 2019). In most of these studies, a set of primitives was introduced and the shape of the object was represented by arranging these primitives in the order of their occurrence when tracing the outline of the object. This process indicates that the target is essentially one-dimensional spatial data.

On the other hand, for our application, we have to consider representation based on local data extracted

Taniuchi, Y. and Takahashi, K.
Qualitative Spatial Representation and Reasoning About Fold Strata.
In *Proceedings of the 15th International Conference on Agents and Artificial Intelligence (ICAART 2023) - Volume 2*, pages 211-220
ISBN: 978-989-758-623-1; ISSN: 2184-433X

from a stratum, since the entire data do not comprise a closed curve. Moreover, we have to represent not only the shapes of layers that become regions of a two-dimensional plane but also their interconnections. Therefore, we cannot apply existing methods.

In this study, we propose representation and reasoning for a fold as a relatively simple strata structure. First, we define a model for local fold data and the language needed to describe it. Next, we define the validity required of the representation, and show that the model representation is valid and that a figure can be drawn on a two-dimensional plane for the valid representation. Moreover, we define operations on the representation corresponding to rotation and symmetric transitions, and show that validity is preserved. Finally, we discuss the interconnection of models that have the same strata configurations. Our goal is to derive spatial relations among multiple local data collected in different places or at different times.

This study provides a mechanical treatment of strata using symbolic representation that focuses on their features. This approach can provide logical explanations of processes that may be involved in future morphological changes that manual analysis may overlook.

This paper is organized as follows. In Section 2, we identify our target fold and the model thereof . In Section 3, we define a description language. In Section 4, we provide an algorithm to generate a representation for a model, and show that the representation and the model have a one-to-one relation. In Section 5 we define operations on this representation. In Section 6, we discuss reasoning using these operations. In Section 7, we compare our study with related studies. Finally, in Section 8, we show our conclusions and future works.

2 MODEL

We describe a typical form of fold strata such as that shown in Figure 1(a). We assume that there is no fault or hole, and that the curvature of all the layers is the same. We model a vertical cross section of the fold projected onto a two-dimensional plane. We derive the local data extracted from the global data by a rectangle that satisfies the following conditions **[COND]**. Based on these conditions, the fold is divided into regions using multiple smooth continuous curves (called *layer-borderlines*). Pairs of layer-borderlines do not intersect and there is no self-intersection. We treat this figure as our model.

[COND]

1. All layers and any space (a region containing no layer) in the global data appear to be connected regions in the local data.
2. The end-points of each layer-borderline are not located on a corner of the rectangle.
3. Each layer-borderline is a smooth curve with neither an extremum nor an inflection point.

(a) fold form (b) model on a 2D plane

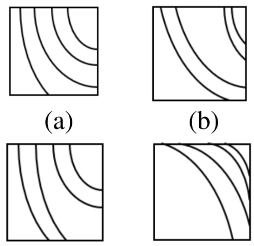
Figure 1: A model for a fold.

For example, part of the fold shown in Figure 1(a) is modeled as the figure in Figure 1(b). In the model, the bottom-left point is regarded as the origin and the inclination of the curve is determined to be either increasing or decreasing.

We refer to the borderlines between layers as *layer-borderlines* to discriminate them from the borderline of the rectangle.

Note that since this is a qualitative model, we focus only on the side on which end-points of layer-borderlines occur and the order of the locations, ignoring their precise positions. As for the shape of a layer-borderline, we focus only on its inclination and convexity, ignoring its precise shape. As a result, several figures are regarded as the same model.

Example 1. *In Figure 2, (b) is regarded as the same model as (a), whereas (c) and (d) are not.*



(a) (b) (c) (d)

Figure 2: Models.

3 DESCRIPTION LANGUAGE

3.1 Language

We define two kinds of description languages Lang1 and Lang2 to represent the model for local data.

Lang1 is used to describe the configuration of a stratum. This is defined as Lang1 $= \{A_1, \ldots, A_n\} \cup \{\theta\}$ where $A_1, \ldots, A_n$ are the names of the layers and θ denotes the outside of the stratum. $A_1, \ldots, A_n$ and θ are called *layer-symbols*.

Lang2 is used to describe the shape of a layer-borderline. This is defined as Lang2 $= \{$↱, ↴, ⬏, ↳$\}$ where ↱, ↴, ⬏ and ↳ indicate convex upward and increasing, convex upward and decreasing, convex downward and increasing, and convex downward and decreasing, respectively. ↱, ↴, ⬏ and ↳ are called *shape-symbols*. We also denote $Up = \{$↱, ⬏$\}$, $Dn = \{$↴, ↳$\}$.

Let $\sigma = e_1 \ldots e_k$ be either a sequence of symbols in Lang1 or that of those in Lang2. If σ is the null sequence, then we denote it as ε. For each i $(1 \leq i \leq k)$, we denote $e_i \in \sigma$. We also denote $first(\sigma) = e_1$, $last(\sigma) = e_k$ and $\sigma^{-1} = e_k \ldots e_1$ for $i \geq 1$.

Definition 1 (local data description, layer-sequence)**.** *Local data description is defined as a pair (L, C), where L and C are finite sequences that include symbols in* Lang1 *and* Lang2*, respectively. L consists of four segments in the form $(\sigma^1)(\sigma^2)(\sigma^3)(\sigma^4)$ with auxiliary symbols '(' and ')'. The sequence of symbols without the auxiliary symbols '(' and ')' is called* a layer-sequence *of L.*

A layer-sequence is considered cyclic data, that is, for a layer-sequence $e_1 \ldots e_k$, e_0 is considered e_k, and for all i $(1 \leq i \leq k)$, $e_i \ldots e_k e_1 \ldots e_{i-1}$ are considered the same data.

Definition 2 (sequence-of-transitions)**.** *For a local data description (L, C), let $I = e_1, \ldots, e_k$ be a layer-sequence of L, where $k \neq 1$. Then the sequence $c_1 \ldots c_k$ where for each i $(1 \leq i \leq k)$, $c_i = e_{i-1}/e_i$, $e_i \in \sigma_i$, $\sigma_i \in \{\sigma^1, \sigma^2, \sigma^3, \sigma^4\}$ is said to be a* sequence-of-transitions *of L. And we denote $chgpt(c_i, \sigma_i)$.*

Example 2. *For $L = (A\theta)()(ABC)(B)$, the layer-sequence of L is $I = A\theta ABCB$, the sequence-of-transitions of L is $B/A, A/\theta, \theta/A, A/B, B/C, C/B$, and $chgpt(A/\theta, \sigma^1)$ and $chgpt(\theta/A, \sigma^3)$ hold.*

3.2 Validity

For the local data description (L, C) where $L = (\sigma^1)(\sigma^2)(\sigma^3)(\sigma^4)$, we introduce the term 'inclination of a layer-borderline' that relates L and C.

Definition 3 (inclination of a layer-borderline)**.** *For each pair of layer-symbols X and Y, for which $chgpt(X/Y, \sigma)$ and $chgpt(Y/X, \sigma')$ where $\sigma \neq \sigma'$ hold,* the inclination of a layer-borderline C_{XY} *is defined depending on the pair of σ and σ' as follows:*

- *if (σ, σ') is either $(\sigma^1, \sigma^2), (\sigma^2, \sigma^1), (\sigma^3, \sigma^4)$ or (σ^4, σ^3), then $C_{XY} = dn$*
- *if (σ, σ') is either $(\sigma^1, \sigma^4), (\sigma^2, \sigma^3), (\sigma^3, \sigma^2)$ or (σ^4, σ^1), then $C_{XY} = up$*
- *otherwise $C_{XY} = any$.*

Definition 4 (validity)**.** *If the local data description (L, C) satisfies the following conditions, then it is said to be* a valid representation.

Let $L = (\sigma^1)(\sigma^2)(\sigma^3)(\sigma^4)$ and its layer-sequence $I = e_1 \ldots e_k$.

v1 *For any pair X and Y of layer-symbols if L includes $chgpt(X/Y, \sigma)$, then it includes exactly one $chgpt(Y/X, \sigma')$, where $\sigma \neq \sigma'$ holds.*

v2 *I is $X_n\theta$ or in the form of $X_1 \ldots X_{n-1} X_n X_{n-1} \ldots X_1 \theta$ where $X_i \neq X_j$ $(1 \leq i < j \leq n)$.*

v3 $|C| = 1$.

v4
- *If for all C_{XY}, $C_{XY} = up$* or *any, then $C \in Up$.*
- *If for all C_{XY}, $C_{XY} = dn$* or *any, then $C \in Dn$.*
- *otherwise, $C \in Up \cup Dn$.*

From [v2], the following proposition holds.

Proposition 1. *Let (L, C) be a valid representation and I be a layer-sequence of L. Then $I = I^{-1}$.*

4 REPRESENTATION FOR A MODEL

We provide a representation for a model. We show that it is valid; and that conversely there exists a model of a valid representation and we can draw a figure satisfying **[COND]**.

4.1 Representation for a Model

When a model M of local data is provided, starting from the top-left of M, trace the borderline of M in a clockwise manner to obtain a sequence of the layer-symbols that are encountered, and place parentheses around each side of the rectangle. Then we set $L = (\sigma_t)(\sigma_r)(\sigma_b)(\sigma_l)$, where $\sigma_t, \sigma_r, \sigma_b$ and σ_l are the sequence of upper side, right side, lower side and left side of the rectangle, respectively. We set C to correspond to the shape of the layer-borderline. (Note that the shape of all the layer-borderlines is the same.) Then $D = (L, C)$ is said to be *a representation for M*,

Example 3. *The representation for the model shown in Figure 3 is $(\,(A\theta)()(ABC)(B),$ ↴$)$.*

The sequence starts not from layer-symbol B, but from A, although this may seem unnatural. If the sequence were to start from B, the layer-symbol occupying the top-left corner would appear in both σ_t

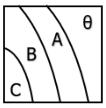


Figure 3: Representation for a model.

and σ_l. To avoid such a situation and to treat the sequence cyclically, the sequence starts from A, the layer-symbol that is encountered first on tracing.

Let (L,C) be a representation for the model M. The sequence-of-transitions $c_1 \ldots c_k$ of L shows the order of occurrence of the end-points of each layer-borderline on tracing the borderline of M. And for each i $(1 \leq i \leq k)$, $chgpt(c_i, \sigma_i)$ indicates that the end-point c_i of a layer-borderline is on the side corresponding to σ_i.

4.2 Validity and Drawability

Theorem 1 (validity of the model)**.** *The representation for the model is valid.*

Proof. For any pair X and Y of layer-symbols, $chgpt(X/Y, \sigma)$ and $chgpt(Y/X, \sigma')$ correspond to the two end-points of the layer-borderline of X and Y. From the first condition of **[COND]**, each layer-borderline of M does not intersect with itself or another layer-borderline. It has exactly two end-points on the borderlines, which are not on the same side of M, in accordance with the third condition of **[COND]**. Therefore, $\sigma \neq \sigma'$. Thus, validity [v1] holds.

The length of each layer-sequence is even, since each layer-borderline has exactly two end-points. Let $e_1 \ldots e_{2k}$ be the layer-sequence of L. If there is only one layer-borderline, then the layer-sequence of L is $X\theta$ where X is a layer-symbol. If there is more than one layer-borderline, then let $e_0 = e_{2k} = \theta$, $e_1 = X_1, \ldots, e_k = X_k$, where $X_i \in \text{Lang1}$ $(1 \leq i \leq k)$. For each i, j $(0 \leq i < j \leq k-1)$, if the end-points X_i/X_{i+1} and X_j/X_{j+1} occur in this order in L, then X_{j+1}/X_j and X_{i+1}/X_i occur in this order in L, since layer-borderline pairs should not intersect. Moreover, if we assume that $X_i = X_j$ $(i \neq j)$ holds, then the layer X_i should appear more than twice in L, indicating that it is a disconnected region; this contradicts the first condition of **[COND]**. Therefore, $X_i \neq X_j$. Thus, validity [v2] holds.

Validity [v3] holds from the assumption of the model. Therefore, C_{XY} are defined uniquely and consistently for all pairs of X and Y. Thus, validity [v4] holds. □

Theorem 2 (drawability of the representation)**.** *There exists a model for the valid representation.*

Proof. Let (L,C) be a valid representation and $L = (\sigma^1)(\sigma^2)(\sigma^3)(\sigma^4)$.

Let $c_1 \ldots c_{2k}$ be a sequence-of-transitions of L, since the lengths of sequence-of-transitions of L are even from validity [v2]. We locate each $c_i \in \sigma_i$ $(\sigma_i \in \{\sigma^1, \sigma^2, \sigma^3, \sigma^4\})$ $(1 \leq i \leq 2k)$ on the borderline of the rectangle in the clockwise direction: locate the elements σ^1, σ^2, σ^3 and σ^4 on the upper side, right side, lower side and left side, respectively, in accordance with the order of occurrence in the sequence-of-transitions. Then, we can draw each layer-borderline so that its two end-points are not on the same side, and not on a corner, for the following reason.

For any pair X and Y of layer-symbols, we can draw a line between the end-points corresponding to X/Y and Y/X in the sequence-of-transitions. Validity [v1] indicates that a line connecting the two points exists; and validity [v2] indicates that lines do not intersect, and are without extremum or inflection points. Therefore, a region encircled by layer-borderlines and the borderlines of the rectangle is a connected region.

The inclination of all layer-borderlines is the same, based on validity [v4]. Then, based on validity [v3], we can draw a smooth curve according to C.

Therefore, a model for the valid representation exists, which means that we can draw a figure corresponding to the model. □

5 OPERATION

Our goal is to derive spatial relations among multiple local data collected in different locations or at different times. To achieve this, we define operations on the local data description and check for changes in the model resulting from these operations.

Let S_0 be a set of representations for models of local data. From Theorem 1, any element of S_0 is valid.

Here, we define three operations: rotation, horizontal flip and vertical flip on S_0 (Figure 4). We define the operation o on $D = (L,C)$ as $o(D) = (o(L), o(C))$.

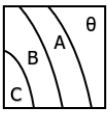
5.1 $\pi/2$ Rotation

Let f be the operation that rotates the model by $\pi/2$ clockwisely. This is defined as follows.

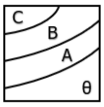
For $L = (\sigma_t)(\sigma_r)(\sigma_b)(\sigma_l)$, $f(L) = (\sigma_l)(\sigma_t)(\sigma_r)(\sigma_b)$.

For C, $f(\Rsh) = ⤵$, $f(⤵) = ⤴$, $f(⤴) = ↳$, $f(↳) = \Rsh$.

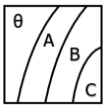
Proposition 2. *1. The model corresponding to $f(D)$ is a figure that is $\pi/2$ clockwisely rotated relative to that corresponding to D.*



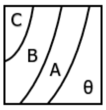
(a) original data
($(A\theta)()(ABC)(B)$, ⬎)



(b) $\pi/2$ rotation
($(B)(A\theta)()(ABC)$, ⬏)



(c) flip horizontal
($(AB)(C)(BA\theta)()$, ↱)



(d) flip vertical
($(BA\theta)()(AB)(C)$, ⬏)

Figure 4: Operations on S_0.

2. *For each D in S_0, $f(D)$ is valid.*
3. $f(f(f(f(D)))) = D$.

Proof. This can easily be proved, since the operation is only swapping segments. □

Example 4. *The representation for Figure 4(a) is $D = ((A\theta)()(ABC)(B),$ ⬎). If we draw $f(D) =$ ($(B)(A\theta)()(ABC)$, ⬏), then we can obtain the model shown in Figure 4(b), which corresponds to $\pi/2$ clockwisely rotated with respect to the original model shown in Figure 4(a).*

5.2 Horizontal Flip

Let g be the operation that flips the model horizontally.

First, we detect the layer that is encountered last on tracing the borderline of the model before applying the operation, that is, the layer which will occupy the top-left corner of the model after applying the operation. This is said to be *a delimiter* and is defined as follows.

For $L = (\sigma_t)(\sigma_r)(\sigma_b)(\sigma_l)$,

$$delimiter = \begin{cases} last(\sigma_t) & (\text{if } \sigma_t \neq \varepsilon) \\ last(\sigma_l) & (\text{if } \sigma_t = \varepsilon, \sigma_l \neq \varepsilon) \\ last(\sigma_b) & (\text{if } \sigma_t = \sigma_l = \varepsilon). \end{cases}$$

Let $I = e_1 \dots e_k$ be the layer-sequence of L and e_z be the delimiter $(1 \leq z \leq k)$. Let $I' = e_{z-1}e_{z-2}\dots e_1 e_k e_{k-1}\dots e_z$.

Then we set $g(L) = (\sigma'_t)(\sigma'_r)(\sigma'_b)(\sigma'_l)$, by dividing I' into four segments by inserting the symbols '(' and ')' so that $|\sigma'_t| = |\sigma_t|$, $|\sigma'_r| = |\sigma_l|$, $|\sigma'_b| = |\sigma_b|$ and $|\sigma'_l| = |\sigma_r|$.

For C, g(↱) = ⬎, g(⬎) = ↱, g(⬏) = ↳, g(↳) = ⬏.

Proposition 3. *1. The model corresponding to $g(D)$ is a figure that is horizontally flipped relative to that corresponding to D.*

2. *For each D in S_0, $g(D)$ is valid.*
3. $g(g(D)) = D$.

Proof. 1. Considering cyclicity, $I' = I^{-1}$. The encountered order of layers on tracing the borderline of the model for $g(L)$ is the inverse of that in the original model. Moreover, the numbers of end-points on each side of the original model are the same as on the corresponding sides of the model for $g(L)$, since $|\sigma|$ indicates the number of end-points on the side σ.

2. Assume that L is valid.
$I' = I^{-1}$. In addition, for any pair X and Y of layer-symbols $chgpt(X/Y, \sigma)$ and $chgpt(Y/X, \sigma')$ are mapped to $chgpt(Y/X, \tau)$ and $chgpt(X/Y, \tau')$, respectively, by g. Then $\tau \neq \tau'$ holds since $\sigma \neq \sigma'$ holds, from the definition of g. Therefore, validity [v1] holds.
Validity [v2] holds, since $I' = I^{-1}$.
Validity [v3] trivially holds.
We show that validity [v4] holds as follows. We show the case of $C \in Dn$. Since the inclinations of all the layer-borderlines are either dn or any, we consider a case in which $chgpt(X/Y, \sigma_b)$ and $chgpt(Y/X, \sigma_l)$ hold where the inclination is $C_{XY} = dn$. In this case, the pair of these end-points is mapped to the pair $chgpt(Y/X, \sigma'_b)$ and $chgpt(X/Y, \sigma'_r)$, respectively, by g. Their inclination is up. Similarly, for the other layer-borderlines, the inclination of dn is mapped to up, and any to any. Therefore, $g(C) \in Up$ holds. It follows that validity [v4] holds in this case. We can prove the other cases similarly.

3. $g(g(D)) = D$ holds trivially.

□

Example 5. *The representation for the model in Figure 4(a) is $D = ((A\theta)()(ABC)(B),$ ⬎). If we draw $g(D) = ((AB)(C)(BA\theta)(),$ ↱), then we can obtain the model shown in Figure 4(c), which corresponds to the horizontally flipped original model shown in Figure 4(a).*

5.3 Vertical Flip

Let h be the operation that flips the model vertically. In this case, the delimiter is defined as follows.

For $L = (\sigma_t)(\sigma_r)(\sigma_b)(\sigma_l)$,

$$delimiter = \begin{cases} last(\sigma_b) & (\text{if } \sigma_b \neq \varepsilon) \\ last(\sigma_r) & (\text{if } \sigma_b = \varepsilon, \sigma_r \neq \varepsilon) \\ last(\sigma_t) & (\text{if } \sigma_b = \sigma_r = \varepsilon). \end{cases}$$

Let $I = e_1 \dots e_k$ be the layer-sequence of L and e_z be the delimiter $(1 \leq z \leq k)$. Let $I' = e_{z-1}e_{z-2}\dots e_1 e_k e_{k-1}\dots e_z$.

Then we set $h(L) = (\sigma'_t)(\sigma'_r)(\sigma'_b)(\sigma'_l)$, by dividing I' into four segments by inserting the symbols '(' and ')' so that $|\sigma'_t| = |\sigma_b|$, $|\sigma'_r| = |\sigma_r|$, $|\sigma'_b| = |\sigma_t|$ and $|\sigma'_l| = |\sigma_l|$.

For C, $h($↱$) =$ ↳, $h($↴$) =$ ⬏, $h($⬏$) =$ ↴, $h($↳$) =$ ↱.

Proposition 4. *1. The model corresponding to $h(D)$ is a figure that is vertically flipped relative to that corresponding to D.*

2. For each D in S_0, $h(D)$ is valid.

3. $h(h(D)) = D$.

Proof. Similar to the proof of g. □

Example 6. *The representation for Figure 4(a) is $D = ((A\theta)()(ABC)(B),$ ↴$)$. If we draw $h(D) = ((BA\theta)()(AB)(C),$ ⬏$)$, then we can obtain the model shown in Figure 4(d) that corresponds to the vertically flipped original model shown in Figure 4(a).*

5.4 Combination of Operations

Proposition 5. *For $D_1, D_2 \in S_0$ where $D_1 = (L_1, C_2)$ and $D_2 = (L_2, C_2)$, if D_2 can be obtained from D_1 by applying the operations f, g and h finite times, then the layer-sequences of L_1 and L_2 are equivalent.*

Proof. Let $L_1 = (\sigma_t)(\sigma_r)(\sigma_b)(\sigma_l)$. Then layer-sequence of L_1 is $I = \sigma_t\sigma_r\sigma_b\sigma_l$.
The layer-sequence of $f(L_1) = (\sigma_l)(\sigma_t)(\sigma_r)(\sigma_b)$ is $\sigma_l\sigma_t\sigma_r\sigma_b$, which is equivalent to I because of its cyclicity.
The layer-sequences of $g(L_1)$ and $h(L_1)$ are equivalent to $I' = e_{z-1}e_{z-2}\ldots e_1 e_k e_{k-1}\ldots e_z$, where e_z is the delimiter of L_1. Therefore, they are equivalent to I. □

The following property holds with respect to the combination of the operations, which can be easily proved.

Proposition 6. $f(f(D)) = g(h(D)) = h(g(D))$.

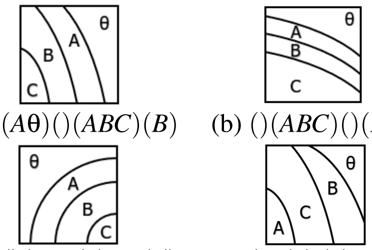
6 REASONING

6.1 Interconnection of Models

For a pair of representations for models, if the adjacency between the layers appearing in them is the same, then their configuration is the same.

Definition 5 (same configuration)**.** *For a pair of representations for models $D_1 = (L_1, C_1)$ and $D_2 = (L_2, C_2)$, let I_1 and I_2 be layer-sequences of L_1 and L_2, respectively. If $I_1 = I_2$, then it is said that D_1 and D_2 have* the same configuration.

Example 7. *In Figure 5, (a), (b) and (c) have the same configuration, whereas (d) does not.*



(a) $(A\theta)()(ABC)(B)$ (b) $()(ABC)()(BA\theta)$
(c) $()(ABC)(BA\theta)()$ (d) $(B\theta)(B)(CA)(C)$

Figure 5: Same/different configuration.

Let S_0 be a set of representations for models of local data. When $D_1, D_2 \in S_0$ have the same configuration, we make a new model D by connecting them horizontally or vertically.

First, we discuss horizontal connection.

For a pair of $D_1 = (L_1, C_1)$ and $D_2 = (L_2, C_2)$ in S_0 that have the same configuration, we can connect the right side of D_1 to the left side of D_2 if the following two conditions are satisfied.

[Conditions for horizontal connection]

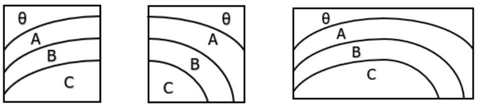
1. Let $L_1 = (\sigma_t)(\sigma_r)(\sigma_b)(\sigma_l)$ and $L_2 = (\tau_t)(\tau_r)(\tau_b)(\tau_l)$. Take elements c_i of the sequence-of-transitions of L_1 that satisfy $chgpt(c_i, \sigma_r)$, and put them in the order of their appearance to make the sequence $c_1 \ldots c_k$. Similarly, take elements c'_j of the sequence-of-transitions of L_2 that satisfy $chgpt(c'_j, \tau_l)$, and put them in their order of appearance to make the sequence $c'_1 \ldots c'_{k'}$. Then $k = k' = |L_1|/2 = |L_2|/2$, and if $c_i = e_{i-1}/e_i$ then $c'_i = e_{k+1-i}/e_{k-i}$ for each i $(1 \le i \le k)$.
2. Either of the following holds.
   (a) $last(C_1), first(C_2) \in Up$
   (b) $last(C_1), first(C_2) \in Dn$
   (c) $last(C_1) =$ ↱ and $first(C_2) =$ ↴
   (d) $last(C_1) =$ ↳ and $first(C_2) =$ ⬏

The first condition means that all the end-points occurring on two connecting sides coincide, and that all the layer-symbols appear on both connecting sides. This avoids the case in which the figure corresponding to the resulting representation could contain a disconnected region. We explain this later. The second condition means that the shapes of all the connected layer-borderlines are smooth.

If both conditions are satisfied, then we can connect D_1 and D_2 to generate the representation $D =$

(L,C). $L = (\sigma_t\tau_t)(\tau_r)(\tau_b\sigma_b)(\sigma_l)$. C is the sequence of shape-symbols obtained by replacing qq, a repetition of shape-symbols, in C_1C_2 by q for each $q \in$ Lang2.

Example 8. *The representations for the models in Figure 6(a) and (b) are $D_1 = (\ ()(ABC)()(BA\theta), ↱\ )$ and $D_2 = (\ ()(A)(BC)(BA\theta), ↴\ )$, respectively, and their horizontal connection is computed as $D_1||D_2 = (\ ()(A)(BC)(BA\theta), ↱↴)$ which is the representation for the model in Figure 6(c).*

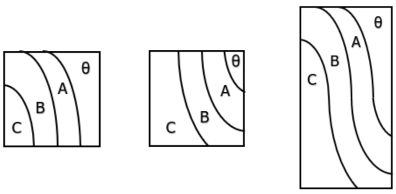


(a) D_1 (b) D_2 (c) $D_1||D_2$

Figure 6: Horizontal connection.

Vertical connection can be defined similarly; connecting the lower side of D_1 and the upper side of D_2, which is denoted by $D_1 + D_2$, when D_1 and D_2 satisfy the conditions on vertical connection.

Example 9. *The expressions for the models in Figure 7(a) and (b) are $D_1 = (\ (A\theta)()(ABC)(B), ↴)$ and $D_2 = (\ (BA\theta)(AB)(C)(), ↳\ )$, respectively, and their vertical connection can be computed as $D_1 + D_2 = (\ (A\theta)(AB)(C)(B), ↴↳\ )$ which is the expression for the model of Figure 7(c).*



(a) D_1 (b) D_2 (c) $D_1 + D_2$

Figure 7: Vertical connection.

We define S_1 as a union of the set of representations for the horizontally/vertically connected models and S_0.

$$S_1 = \{\ D \mid D = D_1||D_2, D_1, D_2 \in S_0\ \} \cup \{\ D \mid D = D_1 + D_2, D_1, D_2 \in S_0\ \} \cup S_0.$$

We repeat this process by generating S_n from S_{n-1} for $n > 1$.

$$S_n = \{\ D \mid D = D_1||D_2, D_1, D_2 \in S_{n-1}\ \} \cup \{\ D \mid D = D_1 + D_2, D_1, D_2 \in S_{n-1}\ \} \cup S_{n-1}.$$

For the representation $(L,C) \in S_n$, we can extract the characteristics of the shape of layer-borderlines from C: if C includes a sub-sequence ↱↴, then it has a maximal point; if C includes a sub-sequence ↳⬏, then it has a minimal point; if C includes a sub-sequence ↱⬏, ⬏↱, ↴↳ or ↳↴, then it has an inflection point.

In general, $D \in S_n$ is not always a valid representation, since a layer-borderline of D may include an extremum or inflection. Instead, the following properties hold.

Theorem 3 (property of extended representation)**.** *$D = (L,C) \in S_n$ satisfies the following properties.*

Let $L = (\sigma_t)(\sigma_r)(\sigma_b)(\sigma_l)$, the layer-sequence of L, $I = e_1 \dots e_k$ and $C = q_1, \dots, q_t$.

p1: *For any pair X and Y of layer-symbols, if L includes $chgpt(X/Y,\sigma)$, then it includes exactly one $chgpt(Y/X,\sigma')$, where $\sigma, \sigma' \in \{\sigma_t, \sigma_r, \sigma_b, \sigma_l\}$.*

p2: *I is $X_n\theta$ or in the form $X_1 \dots X_{n-1}X_nX_{n-1} \dots X_1\theta$, where $X_i \neq X_j$ $(1 \leq i < j \leq n)$.*

p3: *For any i $(1 \leq i \leq t-1)$, $q_i \neq q_{i+1}$ holds.*

We can apply a combination of horizontal/vertical connection to obtain the sequence $S_0, S_1, \dots, S_n$. However, we have to choose the order of application because of the conditions of the connection.

Example 10. *In Figure 8, D_1 and D_2 cannot be horizontally connected, since they do not satisfy the first condition of horizontal connection. On the other hand, $D_3 + D_1$ and $D_4 + D_2$ can be generated since the pair D_3 and D_1, and the pair D_4 and D_2 satisfy the conditions of vertical connection, respectively. In addition, $(D_3 + D_1)||(D_4 + D_2)$ can be generated since $D_3 + D_1$ and $D_4 + D_2$ satisfy the conditions of horizontal connection.*

As a result, a representation for the global data that may contain a maximal point is obtained.

This example means that the properties stated in Theorem 3 are preserved on applying the connections in an appropriate order.

6.2 Prediction of Global Data

When multiple data are collected at distant locations, if they have the same configuration, then we can find several possible ways to connect them, using the above reasoning.

Example 11. *Assume that D_1 and D_2 are given, and were collected at distant locations (Figure 9(a)), does a global stratum exist that contains both of them? One is shown in Figure 9(b), which shows that there exist D_3 and D_4 to connect D_1 and D_2. Such intermediate data are often missing, and we infer the possibility that the global stratum in Figure 9(b) exists, which changes in the long term as a result of crustal movements.*

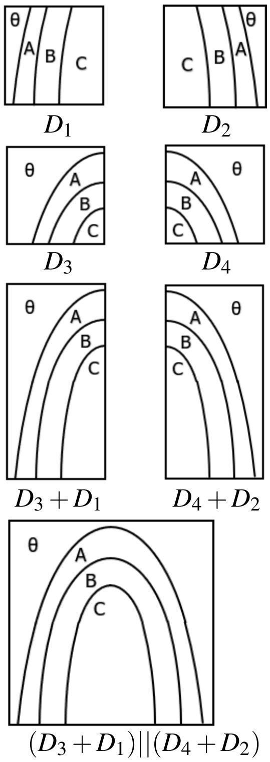


Figure 8: Operation process.

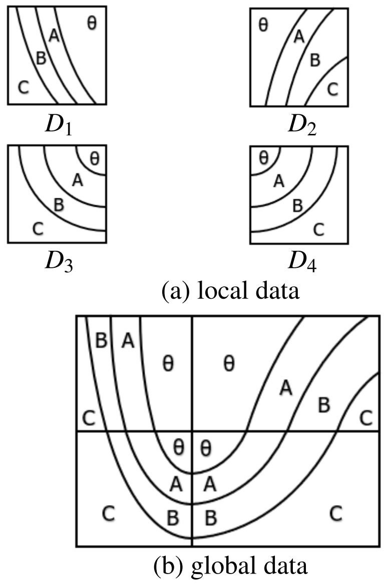


Figure 9: Prediction of globald data.

In this example, the right two rectangles are wider than the left ones in Figure 9(b). Actually, the sizes of collected local data are not always same or distances between their locations are different. Here, we focus on the end-points of layer-borderlines and their shapes both of which are treated qualitatively. Therefore, we can make a model for a global data by changing the size or ratio of sides of rectangles of the models for local data.

Simple fold forms can be generated by connecting multiple local data. This means that we can infer the shape and structure of global data from a set of local data collected in different locations. Currently, the connection conditions are too strict to generate complex fold forms. We will revise the conditions in future studies.

7 RELATED WORKS

To the best of the authors' knowledge, there has been almost no research on strata that uses symbolic representation and logical reasoning.

To combine AI techniques and structural geology, application of machine learning using big data is one possibility. However, the currently available strata-data archives are quite small and the stored data are not sufficiently categorized. Moreover, in most data archives, figures and landscapes are stored using numerical data.

On the other hand, in the QSR research field, several methods for symbolic treatment of shapes have been proposed. Almost all of them treat spatial data on a two-dimensional plane.

Leyton proposed a grammar that represents changes in the shape of a closed curve, starting from a simple smooth curve. He explained changes in shape based on a force acting from inside or outside the curve (Leyton, 1988). He showed that any shape of a smooth closed curve can be represented using language based on the proposed grammar. Tosue et al. extended the grammar so that it can represent phenomena such as the creation of a tangent point and division of the curve (Tosue and Takahashi, 2019). They applied the method to a process of organogenesis. Galton et al. proposed another grammar that can apply to not only a smooth curve but also to a straight line or a curve with cusps (Galton and Meathrel, 1999). They showed that objects of various shapes can be symbolically represented by connecting a finite number of primitive segments. They also referred to transformation between representations differing in granularity. Cabedo et al. proposed a representation for the borderline of an object with further information such as relative lengths and relative angles, and also showed the juxtaposition of objects (Cabedo and Escrig, 2004; Cabedo et al., 2010; Pich and Falomir, 2018), and Falomir et al. defined similarity between qualitative shapes described in their extended model (Falomir et al., 2013).

All of these expressions adopted methods that represent the shape of an object by connecting primitive segments when tracing its borderline. On the other hand, Cohn took a different, approach to represent a concave object (Cohn, 1995). He regarded differences in the closure and the object itself as regions and represented the spatial relations of these regions. Kumokawa et al. also proposed a different representation for a concave shape using closure (Kumokawa and Takahashi, 2008).

A study by Kulik et al. applied QSR to landscape silhouettes (Kulik and Egenhofer, 2003). They proposed a description language for the shape of an open line. They defined several primitives comprising two consecutive vectors depending on relative lengths and angles; regarded the borderline of a silhouette of a landscape as a pattern of connections between these primitives; and deduced landscape features, including mountain, valley, and plateau. They also proposed a transformation from the refined level to the abstract level. The differences between Kulik's method and ours are: first, he used straight lines as primitives, whereas we use curves; second, his target silhouette was always in the vertical direction, whereas our method can be applied to rotated forms; third, he neither formalized the method nor discussed the validity of the representation, whereas we both define the validity of the representation and prove one-to-one relation with the model.

In addition, whereas all extant studies treated the essentially one-dimensional data of a borderline, we treated the two-dimensional data of a stratum consisting of multiple regions.

8 CONCLUSIONS

We have discussed qualitative representation and reasoning for strata.

We developed a model for local data from a typical fold, and proposed its representation in the form of a pair of sequences of symbols that stand for the configuration of a layer and the shapes of the borderlines between layers. This representation is suitable to show the main features of strata: one layer extends in one direction if there is no fault, and the relations of interconnections between layers are unchanged even if the width of a layer, shape, or axis of a fold changes.

We defined the required validity of the representation, and then showed that the valid representation and that of the model have a one-to-one relation. Moreover, we defined several operations on the representation, and showed that they preserve its validity. We also showed that global data can be generated by connecting local data with the same configuration. This enables derivation of relations among multiple local data collected in different locations or at different times. Our main contribution is to show symbolic treatment of strata and provide a basis for logically explaining the process of landscape generation.

In future studies, we intend to identify sets of representations obtained from repetitive application of connections of local data. We are also considering the formalization needed to explain the strata-generation process, as well as a qualitative simulation for possible future morphological changes.

ACKNOWLEDGEMENTS

This research is supported by JSPS Kakenhi JP21K12020. The authors would like to thank Motohiro Tsuboi for giving useful advice from the field of geology.

REFERENCES

Cabedo, L. M., Abril, L. G., Morente, F. V., and Falomir, Z. (2010). A pragmatic qualitative approach for juxtaposing shapes. *Journal of Universal Computer Science*, 16(11):1410–1424.

Cabedo, L. M. and Escrig, M. T. (2004). A qualitative theory for shape representation and matching for design. *Proceedings of the Sixteenth Eureopean Conference on Artificial Intelligence, ECAI'2004*, pages 858–862.

Chen, J., Cohn, A. G., Liu, D., Wang, S., Ouyang, J., and Yu, Q. (2013). A survey of qualitative spatial representations. *The Knowledge Engineering Review*, 30:106–136.

Cohn, A. and Renz, J. (2008). Qualitative spatial representation and reasoning. In *Handbook of Knowledge Representation*, chapter 13. Elsevier.

Cohn, A. G. (1995). Hierarchical representation of qualitative shape based on connection and convexity. *Spatial Information Theory: Cognitive and Computational Foundations of Geographic Information Science (COSIT95)*, pages 311–326.

Falomir, Z., Abril, L. G., Cabedo, L. M., and Ortega, J. A. (2013). Measures of similarity between objects based on qualitative shape descriptions. *Spatial Cognition & Computation*, 13(3):181–218.

Galton, A. and Meathrel, R. (1999). Qualitative outline theory. *Proceedings of the Sixteenth Inter-*

national Joint Conference on Artificial Intelligence, pages 1061–1066.

Kano, K. and Murata, A. (1998). *Structural Geology*. Asakura Publishing Co., Ltd. (in Japanese).

Kulik, L. and Egenhofer, M. J. (2003). Linearized terrain: languages for silhouette representations. *Spatial Information Theory. Foundations of Geographic Information Science, International Conference, COSIT 2003*, pages 118–135.

Kumokawa, S. and Takahashi, K. (2008). Qualitative spatial representation based on connection patterns and convexity. *AAAI08 Workshop on Spatial and Temporal Reasoning*, pages 40–47.

Leyton, M. (1988). A process-grammar for shape. *Artificial Intelligence*, 34:213–247.

Ligozat, G. (2011). *Qualitative Spatial and Temporal Reasoning*. Wiley.

Pich, A. and Falomir, Z. (2018). Logical composition of qualitative shapes applied to solve spatial reasoning tests. *Cognitive Systems Research*, 52:82–102.

Sioutis, M. and Wolter, D. (2021). Qualitative spatial and temporal reasoning: current status and future challenges. *Proceedings of 30th International Joint Conference on Artificial Intelligence*, pages 4594–4601.

Tosue, M. and Takahashi, K. (2019). Towards a qualitative reasoning on shape change and object division. In *Spatial Information Theory. 11th International Conference, COSIT 2019*, pages 7:1–7:15.