

Qualitative Spatial Reasoning

Extracting and Reasoning with Spatial Aggregates

Chris Bailey-Kellogg and Feng Zhao

■ Reasoning about spatial data is a key task in many applications, including geographic information systems, meteorological and fluid-flow analysis, computer-aided design, and protein structure databases. Such applications often require the identification and manipulation of qualitative spatial representations, for example, to detect whether one object will soon occlude another in a digital image or efficiently determine relationships between a proposed road and wetland regions in a geographic data set. Qualitative spatial reasoning (QSR) provides representational primitives (a spatial “vocabulary”) and inference mechanisms for these tasks. This article first reviews representative work on QSR for data-poor scenarios, where the goal is to design representations that can answer qualitative queries without much numeric information. It then turns to the data-rich case, where the goal is to derive and manipulate qualitative spatial representations that efficiently and correctly abstract important spatial aspects of the underlying data for use in subsequent tasks. This article focuses on how a particular QSR system, *SPATIAL AGGREGATION*, can help answer spatial queries for scientific and engineering data sets. A case study application of weather analysis illustrates the effective representation and reasoning supported by both data-poor and data-rich forms of QSR.

The ability to perceive spatial objects and reason about their relations seems effortless for humans but has proved so difficult for computers that they have yet to attain the capabilities of a five-year-old child. Part of the computational problem lies in the difficulty of identifying and manipulating qualitative spatial representations. For example, although the pixels in a digital image implicitly define the locations of spatial objects, the task at

hand might require a more qualitative characterization of the configuration of these objects, say, whether one object will soon occlude another. Handling spatial data is a key task in many applications, including geographic information systems (GISs), meteorological and fluid flow analysis, computer-aided design (CAD) systems, and protein structure databases (figure 1) (Zhao et al. 1999). For example, a GIS system might have large amounts of numeric information about spatial features such as highways and terrain but require query mechanisms to efficiently determine qualitative relationships such as those between a proposed route and wetlands regions.

Qualitative spatial reasoning (QSR) addresses these problems with representational primitives (a spatial “vocabulary”) and inference mechanisms. QSR approaches can be characterized by two important and complementary classes of problems. Problems in the first class are data poor, and the goal is to design representations that can answer qualitative queries without much numeric information. The goal of answering qualitative queries addresses an important aspect of common-sense reasoning by humans and can be found in many practical applications such as computer-aided tutoring or diagram understanding. Because of the lack of detailed numeric information, representations used by the approaches to data-poor problems are often carefully designed by hand with respect to a task at hand. Problems in the second class (for example, scientific and engineering applications from fluid-flow analysis to distributed

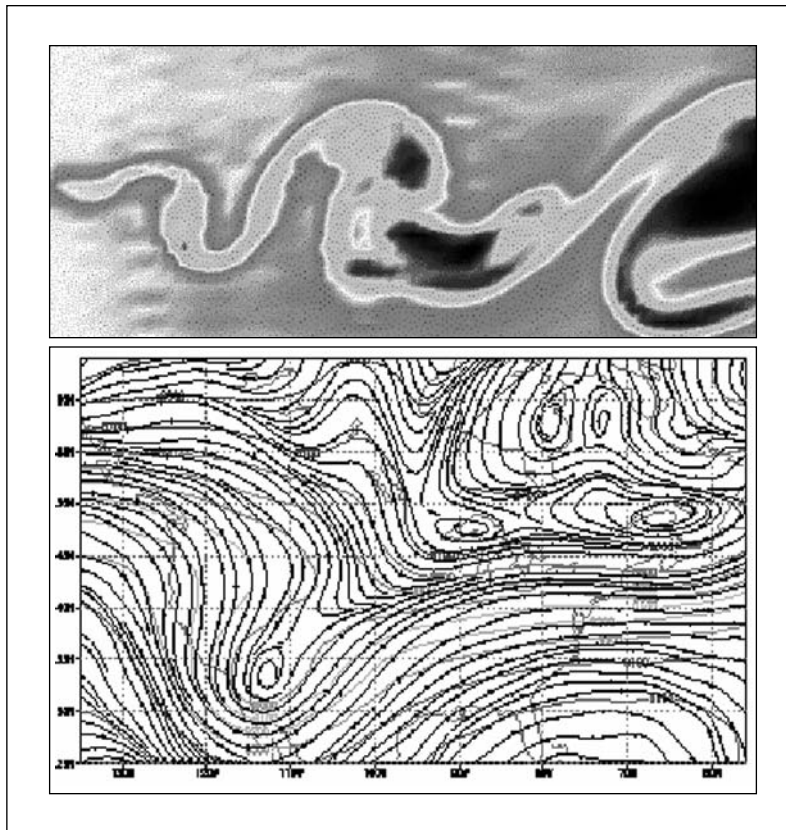


Figure 1. Spatial Objects in Physical Fields.

Top: A fluid flow. The fluid field describes how objects such as high-density regions and large vortex structures are spatially distributed (shown here) and temporally evolving (not shown). *Bottom:* A 300-megabyte weather map over North America. The data in a typical meteorological map include pressure, temperature, and wind velocities on a spatial grid. An experienced meteorologist could identify qualitatively important weather features such as the location of a cold front and the direction of its movement by extracting and correlating geometric features such as pressure troughs and thermal packing.

control design) are data rich, and the goal is to derive and manipulate qualitative spatial representations that efficiently and correctly abstract important spatial aspects of the underlying data and can be used for subsequent tasks. The approaches to data-rich problems are complementary to those for data-poor problems in that they can automatically construct spatial representations. Computational efficiency in reasoning arises from appropriate qualitative spatial representations; for example, a qualitative description of a temperature distribution as a configuration of iso-contours focuses the search for good thermal control designs. Similarly, qualitative representations allow efficient access and manipulation of data, for example, correlating maps (for example, a road map, a utilities map, and a forestry map) in a GIS system, determining the interaction of parts in a CAD design, and planning paths for a robotics application. An important feature of qualita-

tive spatial representations is their ability to relate reasoning results to underlying data (rich or poor) and domain knowledge provided by the user.

In this article, we first review the representative work on QSR for data-poor scenarios. We then turn to the data-rich case and focus on how a particular QSR system, *SPATIAL AGGREGATION*, can help answer spatial queries for scientific and engineering data sets. Finally, we present a particular application that illustrates the effective representation and reasoning supported by both forms of QSR.

Qualitative Spatial Reasoning for Data-Poor Problems

Qualitative reasoning research uses high-level representations of physical systems and domain knowledge for tasks such as prediction, diagnosis, reconfiguration, and tutoring (de Kleer and Brown 1984; Falkenhainer and Forbus 1991; Forbus 1984; Kuipers 1986; Weld and de Kleer 1990) without requiring significant amounts or quality of data. Classical qualitative reasoning work deals primarily with temporal aspects of a system, abstracting away its spatial properties. Following the spirit of this qualitative reasoning research, work in QSR for data-poor domains has focused on similarly “minimalist” spatial representations and inference mechanisms.

Much QSR work has studied purely topological descriptions of spatial regions and their relationships. These approaches often seek to generalize Allen’s temporal interval calculus (for example, *a* before *b*, *a* overlaps *b*, and so on) (Allen 1983) into higher-dimensional, spatial relationships. One representative approach, the region-connection calculus (RCC) (Cui, Cohn, and Randell 1992), provides predicates for expressing and reasoning about the relationships among topological regions (arbitrarily shaped chunks of space). One version, the RCC-8, provides eight jointly exhaustive and pairwise disjoint predicates (figure 2): (1) disconnected from (DC), (2) externally connected to (EC), (3) partially overlaps (PO), (4, 5) tangential proper part of (TPP and TPPi), (6, 7) nontangential proper part of (NTPP and NTPPi), and (8) identical with (EQ). The axioms specifying these relationships provide rigorous underpinnings to support spatial reasoning. For example, Boolean functions (for example, union, intersection, and difference) allow composition of complex spatial objects (that is, topological shapes). Additional predicates can then test, using theorem proving, topological features of these objects (for example, connect-

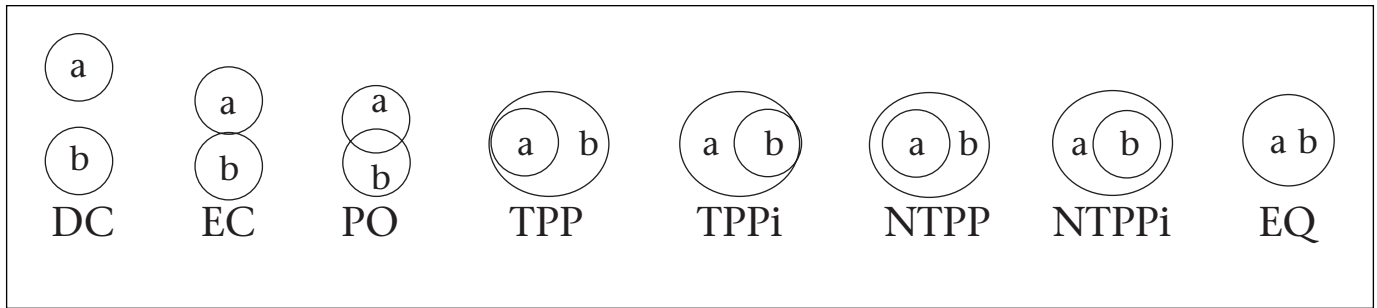


Figure 2. The Region-Connection Calculus (RCC) Represents and Manipulates Pairwise Relationships between Regions of Space.

ed, number of holes), and feasibility of additional relationships. An important characteristic of the predicates is that they support reasoning about continuity—a temporal process between two regions must pass through the possible relationships in a well-defined way (for example, they can't go directly from being disconnected to identical). Although the most general RCC theories are undecidable, tractable subsets suitable for various domains have been identified and applied to applications ranging from GIS to visual programming languages.

In addition to topological relationships, QSR researchers have also studied other key qualitative aspects of spatial objects, such as size and shape, and relationships, such as orientation and distance. A full review of these representations is beyond the scope of this article (see, for example, Cohn and Hazarika [2001]). Shape representations typically go beyond pure topology to specify some amount of geometric information, such as convex-concave portions of a boundary, and use multiscale representations similar to those described in the next section. Uncertainty in shape can be handled with coupled RCC-like predicates specifying an object's certain interior and uncertain exterior. Distance, orientation, and size can be represented relatively, for example, indicating order-of-magnitude relationships and rules for combining them (for example, distances sum when oriented properly). Of particular interest is that these representations must often bring in some amount of metric information to make significant inference possible.

In fact, Forbus, Nielsen, and Faltings (1991) hypothesized that some metric information is necessary in general for QSR. More precisely, the poverty conjecture states that "there is no problem-independent, purely qualitative representation of space or shape." *Purely qualitative* means essentially that no detailed metric information supporting perceptual-like processing is available. Problem independence means that the representation must be general—a small set of spatial objects constrained to

only specific interactions in a specific domain might indeed admit a purely qualitative representation, but this representation might then break down with the addition or modification of a single part.

To balance the tension between qualitative-ness and generality, the metric diagram/place vocabulary (MD/PV) theory (Forbus, Nielsen, and Faltings 1991) takes a hybrid approach, linking metric information supporting quantitative queries with sets of special qualitatively important entities (places) in a domain. For example, in the analysis of clock mechanisms (Forbus, Nielsen, and Faltings 1991), the place vocabulary is computed for pairs of interacting parts, which are specified with CAD-like metric diagrams that can determine such interactions. This approach addresses a key concern in qualitative reasoning—ensuring the appropriateness of a choice of qualitative vocabulary because the place vocabulary is computed for a specific problem. It also ensures tight coupling between qualitative and quantitative aspects of the reasoning. In an approach that is similar at a high level, although different in application details, the spatial semantic hierarchy (Kuipers 2000; Kuipers and Levitt 1988) discovers "interesting" locations in the construction of mappings between topological and metric maps for robot navigation.

Qualitative physical fields (Lundell 1996) capture spatially distributed qualitative parameters; that is, each spatial region consists of a uniformly valued qualitative feature. For example, a model of heating might describe regions as being warm or cold as well as regions that are sunny or shaded. Note that as opposed to pairwise interactions, this representation is, at its very heart, continuous. This representation supports reasoning about spatiotemporal processes in an extension of qualitative process theory (Forbus 1984). To continue the example, a qualitative heat flow would be established between topologically adjacent regions of different qualitative temperature (from warm to cold). This heat flow would establish

a temporal process changing the “front” between the two regions and ultimately the associated temperatures. This approach is discussed in more detail in the case study section.

Qualitative Spatial Reasoning for Data-Rich Problems

In contrast with the data-poor application domains discussed earlier, many important science and engineering applications are characterized by massive amounts of spatially distributed numeric data (figure 1). For example, to predict the weather, meteorologists use pressure, temperature, and wind velocity data collected from a large number of spatially distributed weather sensors. Similarly, in designing aircraft with minimal drag, engineers study wind tunnel and simulation data specifying airflow over a body at many points and over many instants of time. In these applications, geometric as well as topological characterizations are necessary; for example, a temperature field is influenced by the geometry of the domain, spatial variations in material property, and boundary conditions.

The massive amount of data, either collected from experiments or produced by simulations, poses significant computational challenges that can be addressed by QSR. In particular, a central problem is the automatic construction of qualitative spatial representations from a given data set to focus the search space for data interpretation and design tasks. QSR approaches to these problems are often built on a sound mathematical theory of geometric and topological analysis, for example, the theory of cell complexes from algebraic topology that naturally defines “closeness,” “composition,” and “abstraction” (Munkres 1984). As discussed earlier, the availability of data provides us the opportunity to automate the construction of qualitative spatial representations. QSR differs from traditional numeric methods for spatial data analysis problems that also abstract numeric data at multiple levels of resolution. For example, engineers use multigrid methods (Briggs 1987) to analyze numeric properties of physical phenomena using a hierarchy of grid discretization. The main difference between QSR and numeric methods lies in the ontological abstraction that QSR adopts. QSR supports more abstract, qualitative reasoning by introducing notions of objects that explicitly encapsulate key spatial properties of a physical domain. For example, meteorologists use abstract structures such as isobars, pressure troughs, and pressure cells to reason about the underlying pressure data at a higher level of ab-

straction. This key insight—physical properties such as continuity and locality give rise to regions of uniformity in spatially distributed data—enables QSR to overcome the challenge of massive data. In fact, this insight is similar to that underlying the MD/PV approach described in the previous section. Domain-specific physical knowledge justifies the extraction of qualitative information in support of more abstract reasoning processes.

QSR in data-rich domains has many connections and parallels to work in scientific visualization (Rosenblum et al. 1994) and scientific data mining (Ramakrishnan and Grama 2001). For example, weather data can be visualized using pseudocolor to represent temperature, isocontours to connect points of equal pressure, needle diagrams to indicate directions of wind flow with arrows, streamlines to show connected flows, and animations of these to show changes over time. Interactive visualizations allow scientists to explore, focus, filter, project, and transform large data sets. Feature-detection algorithms (for example, for vortexes in fluid data) both identify and track spatial structures over time (Junker and Braunschweig 1995; Ordóñez and Zhao 2000; Samtaney et al. 1994; Yip 1995). Similarly, in scientific data mining, algorithms seek to cluster, generalize, and classify patterns and correlations in databases. For example, mining such patterns can allow identification of general climate patterns across regions (Lu, Han, and Ooi 1993), automatic cataloging of sky images (Fayyad, Weir, and Djorgovski 1993), recognition of volcanoes in images of the surface of Venus (Burl et al. 1994), and tracking of cyclones in weather data (Stolorz et al. 1995).

Spatial Aggregation

Spatial aggregation (Yip and Zhao 1996) is a particular QSR approach for data-massive domains that follows an imagistic reasoning (Yip, Zhao, and Sacks 1995) style, applying vision-like routines to manipulate multilayer geometric and topological structures in spatially distributed data. Thus, it can leverage the connections described earlier with visualization and data mining. In the spirit of qualitative reasoning, however, it focuses on explicit representation and manipulation of objects, explainability of results, and utilization of explicitly encoded domain knowledge. Spatial aggregation is partially motivated by some of the spatial reasoning problems raised by Abelson et al. (1989). The Abelson paper describes a number of approaches to interpreting numeric results of simulations of dynamic systems. These problems often possess a set of geometric and topological

constraints that can be exploited to significantly cut down the search space and can be used to communicate the interpretation results to human experts. For example, in interpreting the qualitative behaviors of a nonlinear dynamic system, one can describe the set of trajectories that share the same asymptotic behaviors as a flow pipe, a geometric object that can easily be visualized (Yip and Zhao 1996). Several early examples of using geometric and spatial reasoning to aid in scientific computation include *KAM* (Yip 1991), which interprets the behaviors of Hamiltonian systems; *MAPS* (Zhao 1994), which designs control laws based on a geometric analysis of the state equations of a dynamic system; and *HIPAIR* (Joskowicz and Sacks 1991), which analyzes the kinematics of fixed-axis mechanisms.

Spatial aggregation organizes computation around image-like representations of spatially distributed data (figure 1). In the field ontology, the input is a field mapping from one continuum to another. For example, a two-dimensional (2-D) temperature field associates a temperature with each point, mapping from \mathbb{R}^2 to \mathbb{R}^1 ; a 2-D fluid field associates a velocity with each point, mapping from \mathbb{R}^2 to \mathbb{R}^2 . A field is information rich in that its representation requires many bits. The identification of structures in a field (for example, iso-bars, pressure cells, and fronts) is a form of data reduction: The data-rich field representation is abstracted into a more concise structural representation. For example, a set of points on a curve can be described more compactly by a parameterized spline—the spline parameters are a much more concise representation than the enumeration of points. Note that the qualitative physical field approach described in the previous section starts with an abstract description of a field (qualitative domain and range in the field); the synergy between the data-rich and data-poor fields are explored further in the case study section.

Spatial aggregation uncovers structures in fields at multiple levels of abstraction, with the structures uncovered at one level becoming the input to the structure-discovery process at the next level. For example, in a weather data analysis application (Huang and Zhao 2000), spatial aggregation could extract from pressure data the isobars, pressure cells, and pressure troughs. As discussed earlier, continuities in a field give rise to regions of uniformity that can be abstracted as spatial structures (for example, isothermal contours are connected curves of equal [or similar enough] temperature). Similarly, these structures exhibit their own continuities; therefore, multilayer structures arise from continuities in fields at multiple

scales. Spatial objects are introduced as primitives in QSR to encapsulate the geometric and topological properties of these points, curves, regions, or volumes. Mathematically, a spatial object is a *cell*—a portion of space topologically equivalent to a ball (Munkres 1984). Adjacency between the objects is defined by the contiguity of their cells. Navigating the mapping from field to abstract description through multiple layers rather than in one giant step allows the construction of modular programs with manageable pieces that can use similar processing techniques at different levels of abstraction. The multilevel mapping also allows higher-level layers to use global properties of lower-level objects as local properties of the higher-level objects. For example, the average temperature in a region is a global property when considered with respect to the temperature data points but a local property when considered with respect to a more abstract region description.

Spatial aggregation provides a set of data types and operators for constructing the spatial aggregate hierarchy. The data types and operators make explicit use of domain-specific knowledge (figure 3), in particular, the similarity and closeness of both field objects and their features that are encoded with metrics, adjacency relations, and equivalence predicates. Yip and Zhao (1996) present a number of application programs, ranging from dynamic systems analysis to mechanical mechanism analysis, in terms of the same set of generic operators parameterized by different such domain knowledge. The central data type of spatial aggregation, the neighborhood graph, is an explicit representation of an object-adjacency relation. The definition of adjacency is domain specific and depends on the metric properties of the input field. Common adjacency relations include Delaunay triangulations, minimal spanning trees, and uniform grids. The neighborhood graph serves as computational glue, localizing interactions between neighboring objects. The main *SPATIAL AGGREGATION* operators aggregate objects into neighborhood graphs satisfying an adjacency predicate, classify neighboring nodes into equivalence classes with respect to an equivalence predicate specifying domain-specific feature similarity, and redescribe equivalence classes into higher-level objects with respect to a domain-specific abstraction mechanism. Additional operators search through neighborhood graphs, check consistency of objects, extract geometric properties, and so forth. By instantiating these operators with proper knowledge at different levels of abstraction, spatial aggregation allows specification of a variety of application programs.

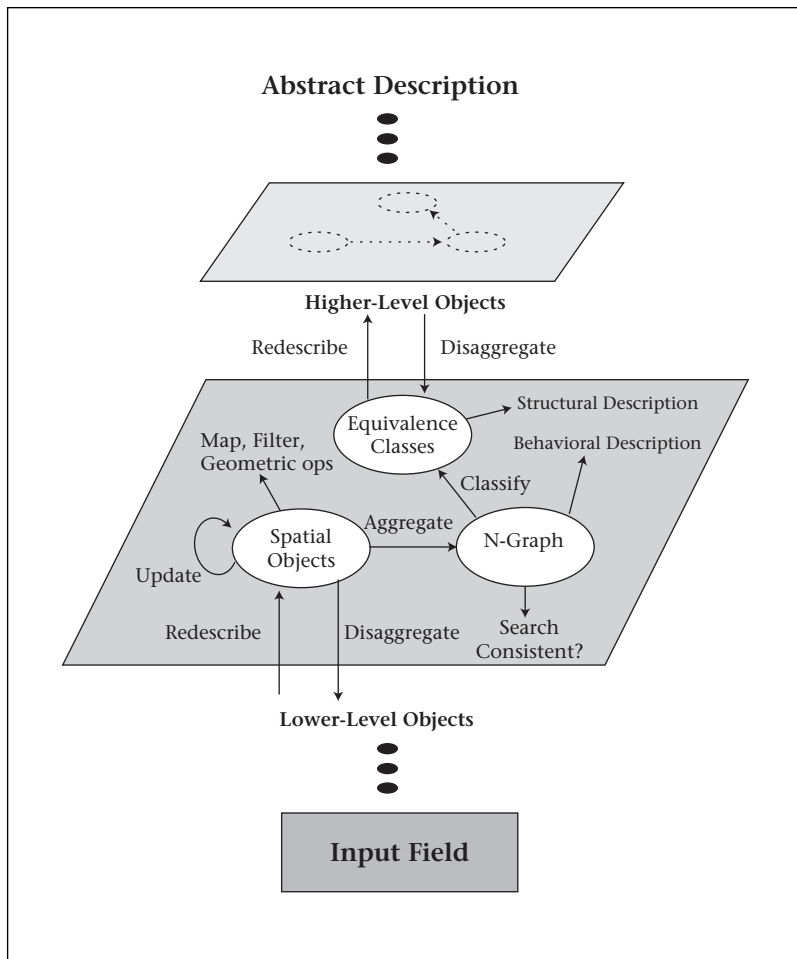


Figure 3. SPATIAL AGGREGATION—Objects Are Created, Transformed, and Abstracted Using a Set of Generic Operators.

The SPATIAL AGGREGATION language (sal) (Bailey-Kellogg, Zhao, and Yip 1996; Yip and Zhao 1996), summarized in table 1, implements the spatial aggregation theory with a C++ library of data types and operators and an interpreted, interactive environment layered over the library. The library supports construction of efficient C++ programs, and the interpreter supports rapid programming of modeling tasks by providing a convenient, high-level interface. SAL lets programmers conveniently explore trade-offs in the specification of domain knowledge such as neighborhood relations and equivalence predicates and interactively and graphically examine and modify the results.¹

Recent work on spatial aggregation (Bailey-Kellogg and Ramakrishnan 2001; Ramakrishnan and Bailey-Kellogg 2002) has moved to bridge the quantitative-qualitative gap in the opposite direction, using qualitative structures to guide sample selection in the underlying field. This approach of sparse data mining is particularly important for applications where

very expensive data collection must be carefully planned (for example, for fluid dynamics simulation and aircraft design). In particular, the iterative approach performs spatial analysis of data in one iteration; identifies ambiguities arising in the analysis; and focuses sample selection in the next iteration to clarify the ambiguities, maximizing information content and improving the analysis. This approach has been shown to make highly effective, explainable sampling decisions in several case studies, including discovery of “pockets” in n -dimensional space by aggregation of gradient vector fields in an interpolated representation derived from a minimal number of targeted samples (Bailey-Kellogg and Ramakrishnan 2001); analysis of matrixes using a perturbation sampling approach (Ramakrishnan and Bailey-Kellogg 2002) that utilizes consistent correspondence of features to determine properties such as the Jordan form from a small number of samples; and influence-based model decomposition for decentralized control design (Bailey-Kellogg and Ramakrishnan 2001; Bailey-Kellogg and Zhao 2001, 1999, 1998), where locality of a few sampled control effects supports high-quality decomposition of problem domains and reasoning about trade-offs among computation, communication among decentralized controls, and resulting control quality.

Case Study: Reasoning with Weather Data

Consider the approach taken by meteorologists interpreting weather data to make predictions about future conditions. They make sense of large, multicategory data sets by recognizing and explicitly labeling aggregate weather features such as high-low pressure centers, pressure troughs, thermal packings, fronts, and jet streams (figure 4) (Huang and Zhao 2000). They then use weather rules, such as the following, to correlate these features and establish prediction patterns:

1. “Major and minor 500mb troughs are good indicators of existing or potential adverse weather” (Air Weather Service 1975).
2. “At 850mb, the polar front is located parallel to and on the warm side of the thermal packing” (Air Weather Service 1975).
3. “A front lies in a pressure trough and the iso-bars make an abrupt change in direction at the front” (Blair and Fite 1965).
4. “A front moves slowly when it is nearly parallel to the iso-bars and increases in velocity as the number of iso-bars intersecting it increases” (Blair and Fite 1965).
5. “A strong high east of a low, especially if the

Primitive Objects represent locations and structures in spatial data.

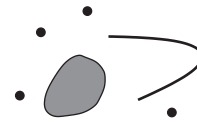
Example:



Compound Objects combine primitive objects

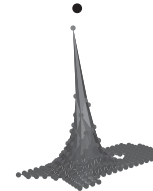
Spaces group objects.

Example (points and curves):



Fields associate objects and features.

Example (point \mapsto temperature):



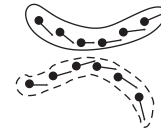
Ngraphs relate nearby objects.

Example (Delaunay triangulation):



Equivalence classes group similar objects.

Example (by vector direction):



Means of Abstraction connect compound objects at one level of abstraction and primitive objects at the next.

Example (by convex hull):

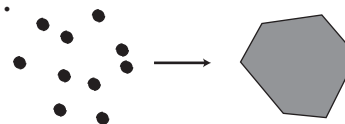


Table 1. Components of the SPATIAL AGGREGATION LANGUAGE (SAL).

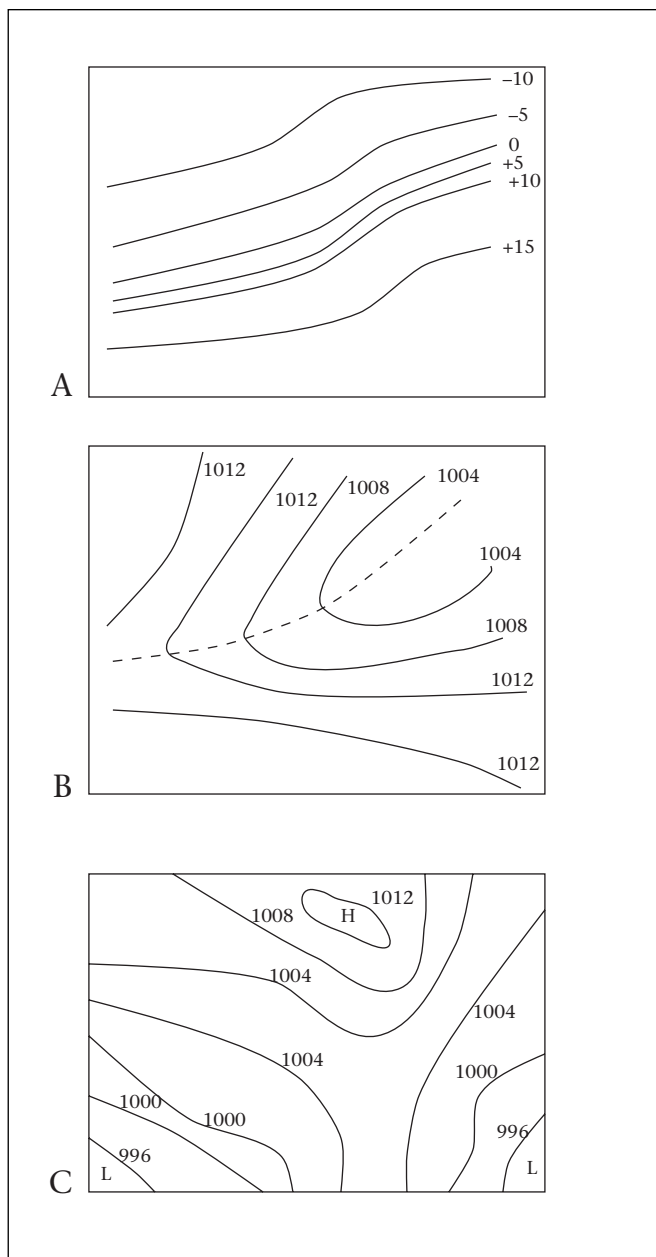


Figure 4. Weather Features.

A. A thermal packing. B. A pressure trough, where solid lines representing iso-bars and dashed line trough position. C. Highs “H” and lows “L” in a pressure map.

high is increasing in intensity or is nearly stationary, will retard the low or deflect it to the left or right. Two lows close together tend to unite” (Blair and Fite 1965).

These rules are intuitively expressed in terms of animated, interacting objects that have rich spatial and physical properties and often defy concise mathematical characterization.

This weather data analysis application illustrates a more general class of practical modeling and design problems requiring an under-

standing of distributed parameter fields. The task of modeling such fields is important by itself, for example, in ecology applications where researchers desire to understand interactions among different parameters (for example, shade and temperature profile) (Lundell 1996). As discussed earlier, modeling also leads directly to prediction based on features extracted from a field at one point in time or, even better, the history of such features over time (Ordóñez and Zhao 2000; Yip 1997). Although the weather data application isn’t amenable to control (yet!), modeling is also important when engineers desire to regulate a physical field with some set of controls (Bailey-Kellogg and Zhao 2001), for example, guiding a robotic laser welding arm in response to temperature data from an infrared camera (Doumanidis 1997) or maintaining a uniform temperature profile with a set of concentric circles of heat lamps (Kailath et al. 1996). In the following two subsections, we discuss how data-poor QSR supports reasoning with models of such heat flows and how data-rich QSR supports extraction and manipulation of such models.

Data-Poor Reasoning

This section follows the approach taken by Lundell (1996, 1995) for qualitative physical fields. In reasoning about temperature fields in ecology, dense, precise numeric data and corresponding models are often not available. However, it is desirable to envision qualitative differences in the temporal evolution for a given (qualitative) model. As described earlier, the model is defined in terms of a field—an association of some parameter (here, temperature) with spatial objects in the domain. Composite fields are defined by spatial interactions of fields with overlapping domains and different parameters (for example, temperature and pressure). The static definition of a field’s domain and involved parameters is specified separately from the dynamic process capturing the interactions among the parameters over time.

The value of a field parameter belongs to some qualitative space of possible values (such as “warm” or “cold”). At a point in time, the field domain is partitioned into maximal regions of the same value (for example, isothermal regions), which is essentially a place vocabulary (see the earlier discussion of MD/PV). The resulting qualitative spatial representations are then amenable to the general QSR techniques described in the preceding section. In particular, composite fields (for example, temperature and shade) are naturally computable using the intersection of regions in the separate fields. Figure 5 illustrates such qualita-

tive physical fields for the temperature modeling application in both the underlying geometric domain and a diagrammatic representation that captures the relevant topological connectivity and continuity.

Dynamic changes in fields (for example, because of heat flow) are captured with a spatiotemporal extension to the qualitative process theory (Forbus 1984). In particular, spatiotemporal processes are captured as interactions between spatial and temporal processes. In the case study application, each region is warmed by a temporal process that takes into account its irradiation and temperature differences, and the regions of boundaries are adjusted by a spatial process that considers the differences in temperatures. More precisely, the temporal process manipulates variables for heating rate in the regions, with a negative qualitative relationship between the heating rate and the amount of shade (that is, more shade means less heating from the sun) and a monotonic influence between the temperature in the region and the heating rate. The spatial process then distributes heat using flow between adjacent regions of different temperatures. It is specified in terms of an “expansion region” of applicability, starting at the boundary between two regions and spreading at a rate proportional to the temperature difference until equilibrium is reached. Temporal processes within the expansion region cool one part and heat the other. Simulation of this set of processes yields a vision of the qualitatively interesting state transitions in the evolution of a temperature-shade field.

Data-Rich Modeling

This section illustrates that qualitative reasoning about physical fields can extract rich structure from large spatial datasets, in support of tasks such as prediction and analysis. In particular, we focus on the identification of troughs and ridges in weather data. We provide here a high-level discussion of a spatial aggregation-based approach; interested readers are referred to Huang and Zhao (2000) for details about the approach and results.

Troughs and ridges are important features in weather analysis; for example, high-altitude troughs give rise to the bending of jet streams and are important for extended weather forecasts, but surface troughs are useful for locating weather fronts. Trough features are only qualitatively understood (see the sample weather prediction rules at the start of this section); sometimes experts even give different answers about the existence of a trough in a weather map. The key to identifying troughs lies in the

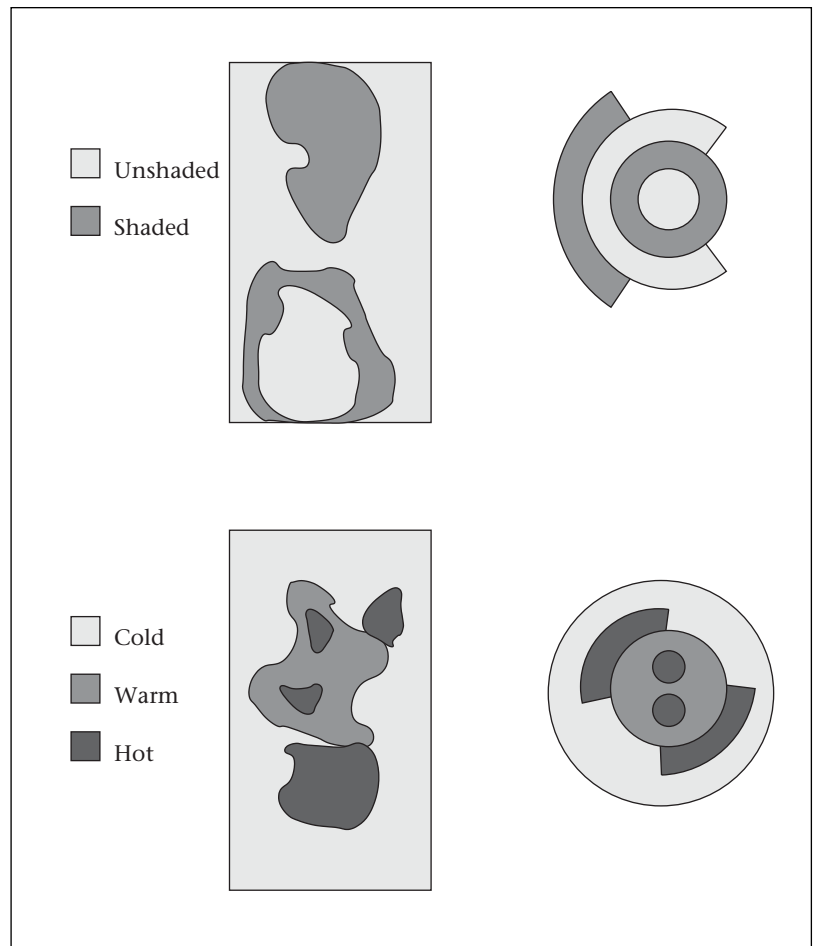


Figure 5. Example Qualitative Physical Fields.

qualitative structures of a field of atmospheric pressure data. In particular, the shape and configuration of iso-bars—the iso-contours of the pressure field, collecting points of the same pressure value—indicate areas where troughs are likely present. Visually, troughs and ridges are stacks of iso-bar segments bending sharply and consistently to one direction, with troughs pointing away from lower iso-bars and ridges away from higher iso-bars. Figure 4b shows a trough. Because of the Coriolis force, winds tend to follow iso-bars; so, sharply bending iso-bars indicate sharp change of wind direction, which usually causes more advection; more mixing of warm air with cold air; and therefore, deteriorating weather.

As previously discussed, data-rich QSR focuses on the extraction and manipulation of structures in spatial data. These structures arise as equivalence classes of neighboring objects according to some similarity measure, re-described as primitive objects at a higher level of abstraction for further analysis. A spatial aggregation-based trough-finding algorithm uses

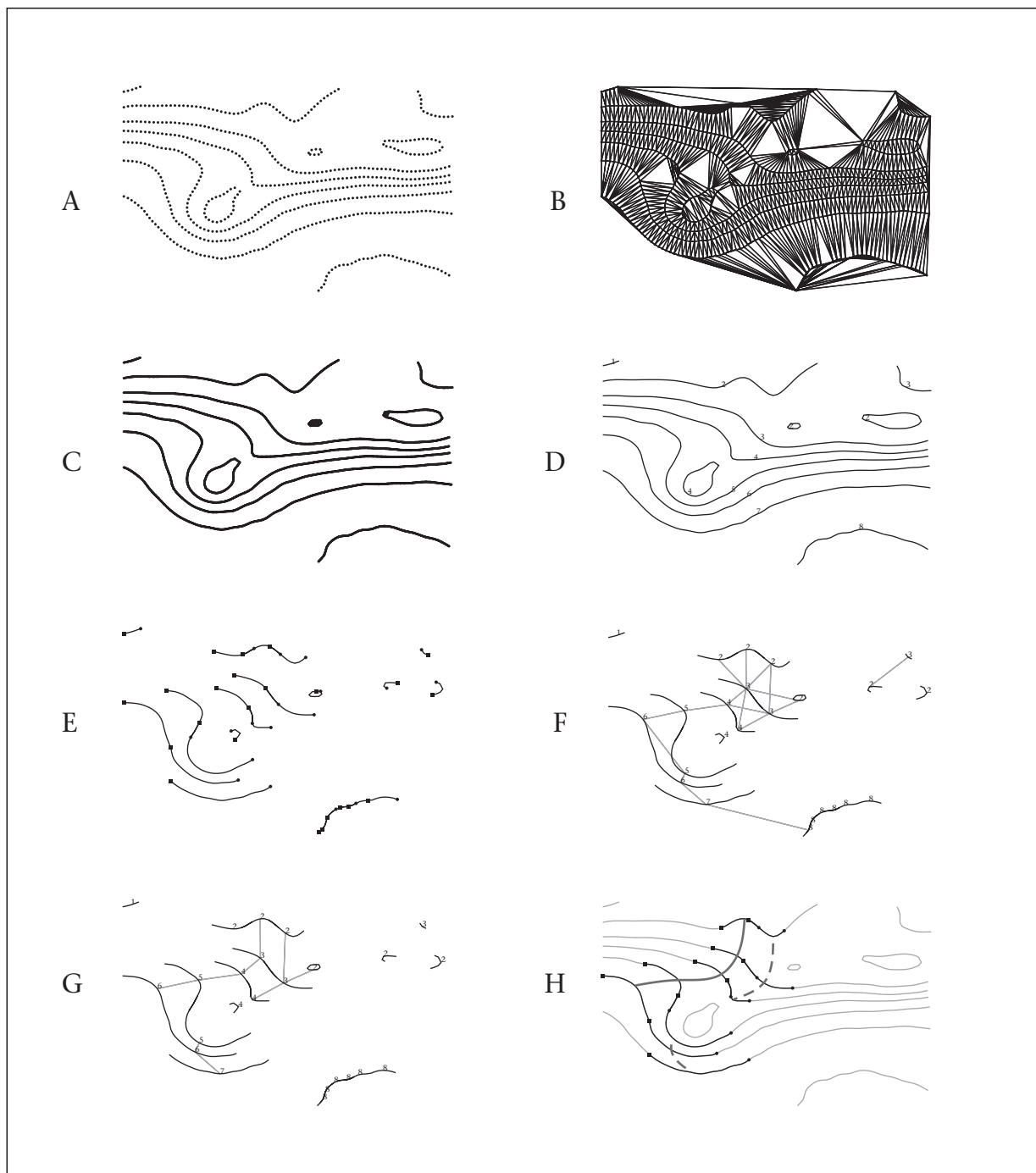


Figure 6. Example of a Spatial Aggregation Approach to Extraction of Pressure Troughs and Ridges.

A. Preprocessed input: iso-points. B. Aggregate (I): point triangulation. C. Classify (I): pressure classes. D. Redescribe (I \rightarrow II): iso-bars. E. Filter (II): high-curvature segments. F. Aggregate (II): constituent point adjacency. G. Classify (II): similar curvature. H. Redescribe (II): troughs and ridge (troughs are dashed; ridge is solid).

this approach to extract the same qualitative spatial features that experts do—sharply bending segments of iso-curves of pressure data. The input to the algorithm is a gridded pressure data set, and the output is a contoured pressure chart with troughs labeled. In a preprocessing

step, iso-bar points are interpolated from the gridded data, yielding a set of “iso-points,” with pressure at specified contour levels. The algorithm then proceeds through two levels of aggregation (see again the section on spatial aggregation for a description of operations in a

typical level), the first to group points into iso-bars and the second to group segments of iso-bars into troughs and ridges. We briefly describe the spatial aggregation steps in this process (using slightly different wording from the original paper); figure 6 (Huang and Zhao 2000) illustrates with data from a 500-megabyte pressure data set from a National Weather Service data server.

Level 1 At the first level of aggregation, the algorithm groups points into iso-bars.

Aggregate: Build a Delaunay triangulation neighborhood graph for the iso-points. The Delaunay triangulation has a number of important geometric properties. Most importantly here, although the triangulation only acts on iso-points, its edges are “well-behaved” with respect to the aggregated iso-curves, in that its edges connect only points within a single curve or in two topologically adjacent curves.

Classify: Form equivalence classes of neighboring points sharing the same pressure value. At the same time, classify graph edges into *strong adjacencies*, connecting same-class points, and *weak adjacencies*, connecting points in different classes.

Redescribe: Abstract each class of same-pressure points into an iso-curve object. As previously discussed, forming a higher-level object allows the computation of aggregate properties; here, curvature is an especially important property. We note that although other algorithms (for example, marching cubes [Lorenson and Cline 1987]) can also be used to contour a pressure data set, the approach taken here yields more structure in the spatial objects, and this structure proves useful in later steps.

Level 2 At level 2, the algorithm groups segments of iso-bars into troughs and ridges.

Filter: Segment the curves and extract high-curvature curve segments. Curve segmentation breaks a curve into piece-wise simple parameterized curves (for example, straight lines or circular arcs); for example, a split-and-merge algorithm (Pavlidis and Horowitz 1974) splits curves at places with high approximation errors and merges segments to avoid oversegmentation. Simple thresholding then allows extraction of high-curvature segments.

Aggregate: Build a neighborhood graph for the high-curvature curve segments based on the between-class (“weak”) adjacencies from level 1. That is, two curves are adjacent if and only if a constituent point in one is adjacent to a constituent point in the other in the original Delaunay triangulation.

Classify: Form equivalence classes of neigh-

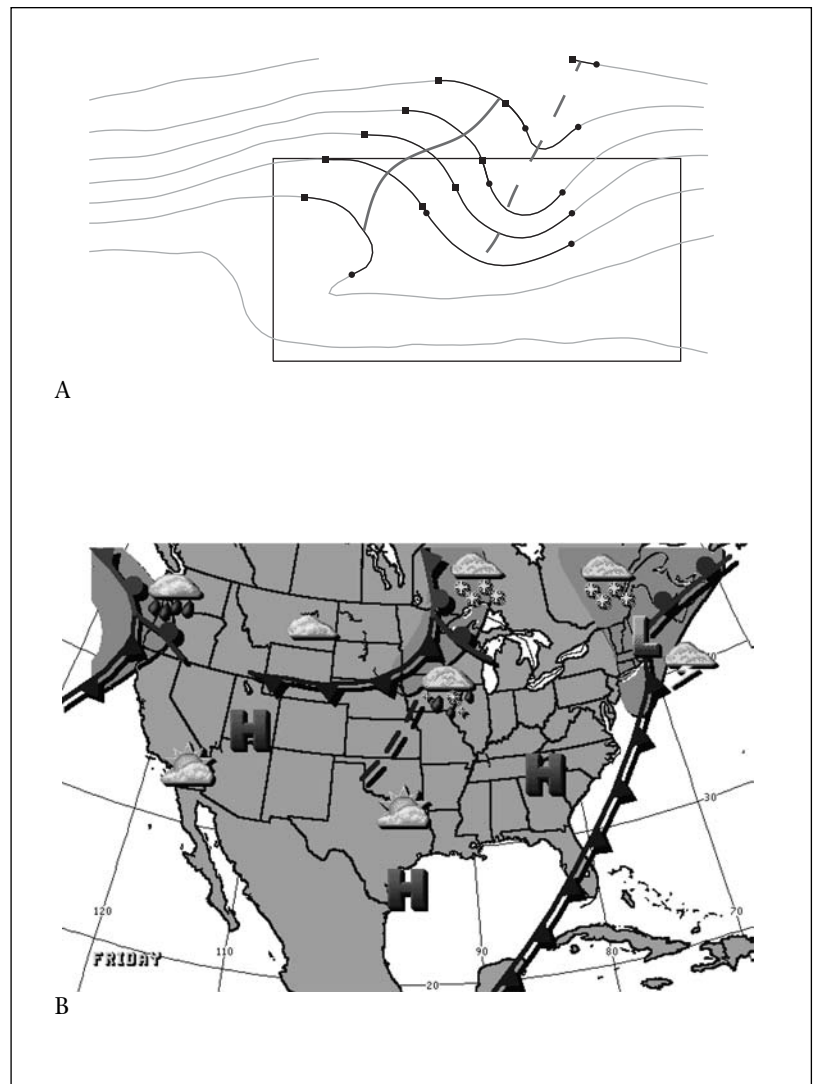


Figure 7. Labeling a Weather Chart.

A. The high-altitude trough (dashed line) detected by the spatial aggregation-based algorithm. B. The corresponding trough (parallel dashed line) drawn by meteorologists for the National Weather Service forecast map for roughly the same area as the small box in A.

boring curve segments that bend in similar directions, within some tolerance.

Redescribe: Abstract equivalence classes of similar-direction high-curvature curve segments into troughs. The abstraction process constructs a curve (for example, B-spline) through the stack of iso-curves (for example, through a representative point on each).

As figure 7 (Huang and Zhao 2000) demonstrates, the results of this algorithm are in qualitative agreement with those of professional meteorologists. Although the expert-drawn trough seems smoother and more visually pleasing, the exact shape and position are not as important for a synoptic map at this scale.

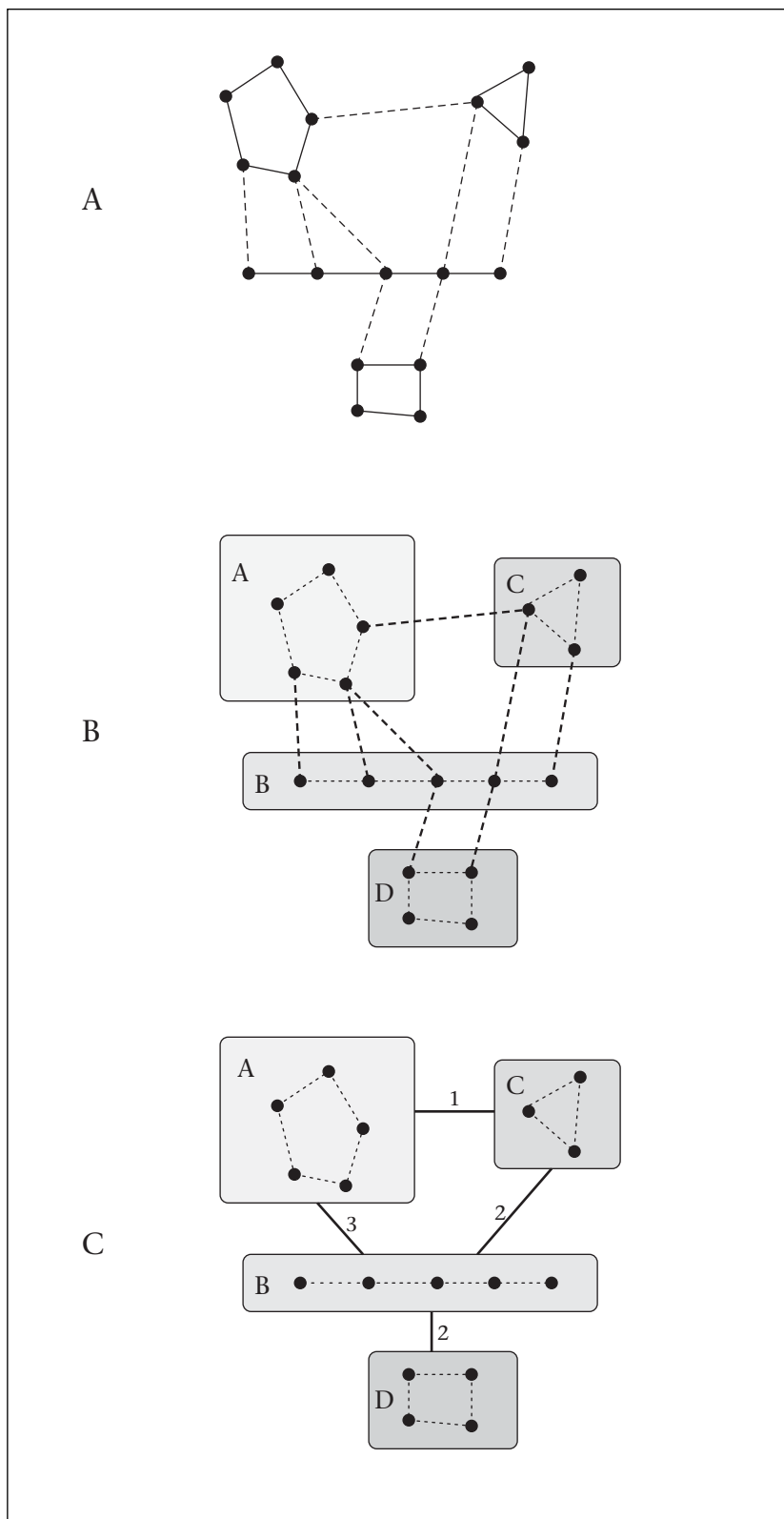


Figure 8. Explicit Representation and Aggregation of Spatial Adjacencies.

A. Neighborhood graph edges are classified into strong (solid) and weak (dashed) adjacencies. B. Strong adjacencies connect internal structure of aggregate objects A, B, and C. C. A higher-level neighborhood graph is constructed from lower-level weak adjacencies between constituent objects.

This trough-finding algorithm illustrates the importance of explicitly representing and reasoning with multilevel spatial structures. The edges in a neighborhood graph play two distinct roles in the aggregation process. The classification process in level 1 uses domain-specific knowledge to distinguish these roles as strong and weak adjacencies (figure 8) (Huang and Zhao 2000). Strong adjacencies carry information about the interactions and connections among the constituent parts of an aggregate spatial object and are abstracted as structural information of the object. Weak adjacencies carry information about the interactions and connections between aggregate objects and are abstracted into a higher-level neighborhood graph. Explicitly representing these adjacencies allows the programmer to use a natural encoding of domain knowledge, in terms of equivalence predicates, to identify objects that are internally connected, externally bounded, and related at multiple levels of abstraction.

Conclusions and Future Research Directions

We described several representative approaches to data-poor and data-rich problems in QSR. In many applications dealing with spatial data, qualitative spatial representations and inferences are preferable because either detailed numeric information is not available for the domain, or existing numeric methods are unable to describe the kinds of geometric and topological structures in data sets that can help answer high-level spatial queries. As the sample applications demonstrate, QSR is an important aspect of commonsense reasoning and can have a significant impact on many technical and scientific applications.

QSR is a rich problem domain for qualitative reasoning research. To fully realize the potential of QSR, we need to address a number of open research issues. For example, spatial aggregation introduces a number of spatial primitives for describing structures in a data-rich physical field. What are other formulations of the problem, using, say, a primal-dual space representation? What are additional primitives and inference operators in spatial aggregation that might be appropriate? How can probabilistic information be incorporated? An important problem in synthesizing the approaches to data-poor and data-rich problems is to use data-rich approaches to automatically build models for data-poor approaches. Here, the data used to build a model for a data-poor problem could perhaps come from a domain where the physics constraints are similar enough, and

numeric information is readily available. For example, in an ecology application, although detailed numeric information for a particular region might not be available because of a lack of instrumentation for that region, the model-building process could leverage data from another similar region where sensors have already been deployed.

Although each of the approaches we have described is to some degree based on mathematical theories of topology and geometry, we have yet to develop a rigorous and formal basis for a general theory of qualitative spatial reasoning that can unify the different approaches. Equally important is the development of a set of problem characterizations that can aid in transforming a general theory into an efficient algorithm for a task.

Acknowledgments

Figure 1, from Zhao et al. (1999), is included courtesy of Ohmsha, Ltd. Figures 4, 6, 7, and 8, from Huang and Zhao (2000), are included courtesy of IOS Press. The spatial aggregation research described here was conducted at The Ohio State University and Palo Alto Research Center by the authors and Xingang Huang and Iván Ordóñez. It was supported by National Science Foundation grant CCR-9457802, Office of Naval Research grant N00014-97-1-0599, and an Alfred P. Sloan research fellowship.

Note

1. For those interested, the SAL source code can be downloaded from www.cs.purdue.edu/homes/cbk/sal.html or www.parc.com/zhao/sal.html.

References

- Abelson, H.; Eisenberg, M.; Halfant, M.; Katzeelson, J.; Sacks, E.; Sussman, G. J.; Wisdom, J.; and Yip, K. 1989. Intelligence in Scientific Computing. *Communications of the ACM* 32(5): 546–562.
- Air Weather Service. 1975. Back to Basics. AWS FOT Seminar, STT-Q9-0004, Air Weather Service, Scott Air Force Base, Illinois.
- Allen, J. 1983. Maintaining Knowledge about Temporal Intervals. *Communications of the ACM* 26(11): 832–843.
- Bailey-Kellogg, C., and Ramakrishnan, N. 2001. Ambiguity-Directed Sampling for Qualitative Analysis of Sparse Data from Spatially Distributed Physical Systems. In Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence, 43–50. Menlo Park, Calif.: International Joint Conferences on Artificial Intelligence.
- Bailey-Kellogg, C., and Zhao, F. 2001. Influence-Based Model Decomposition. *Artificial Intelligence* 130(2): 125–166.
- Bailey-Kellogg, C., and Zhao, F. 1999. Influence-Based Model Decomposition. In Proceedings of the Sixteenth National Conference on Artificial Intelligence, 402–409. Menlo Park, Calif.: American Association for Artificial Intelligence.
- Bailey-Kellogg, C., and Zhao, F. 1998. Qualitative Analysis of Distributed Physical Systems with Applications to Control Synthesis. In Proceedings of the Fifteenth National Conference on Artificial Intelligence, 232–239. Menlo Park, Calif.: American Association for Artificial Intelligence.
- Bailey-Kellogg, C.; Zhao, F.; and Yip, K. 1996. Spatial Aggregation: Language and Applications. In Proceedings of the Thirteenth National Conference on Artificial Intelligence, 517–522. Menlo Park, Calif.: American Association for Artificial Intelligence.
- Blair, T., and Fite, R. 1965. *Weather Elements*. New York: Prentice-Hall.
- Briggs, W. 1987. *A Multigrid Tutorial*. Richmond, Va.: Lancaster.
- Burl, M.; Fayyad, U.; Perona, P.; Smyth, P.; and Burl, M. 1994. Automating the Hunt for Volcanoes on Venus. In Proceedings of Computer Vision and Pattern Recognition Conference (CVPR-94), 302–309. Washington, D.C.: IEEE Computer Society.
- Cohn, A., and Hazarika, S. 2001. Qualitative Spatial Representation and Reasoning: An Overview. *Fundamenta Informaticae* 46(1-2): 1–29.
- Cui, Z.; Cohn, A.; and Randell, D. 1992. Qualitative Simulation Based on a Logical Formalism of Space and Time. In Proceedings of the Tenth National Conference on Artificial Intelligence, 679–684. Menlo Park, Calif.: American Association for Artificial Intelligence.
- de Kleer, J., and Brown, J. 1984. A Qualitative Physics Based on Confluences. *Artificial Intelligence* 24(1-3): 7–83.
- Doumanidis, C. 1997. In-Process Control in Thermal Rapid Prototyping. *IEEE Control Systems* 17(4): 46–54.
- Falkenhainer, B., and Forbus, K. 1991. Compositional Modeling: Finding the Right Model for the Job. *Artificial Intelligence* 51(1-3): 95–143.
- Fayyad, U.; Weir, N.; and Djorgovski, S. 1993. SKI-CAT: A Machine Learning System for the Automated Cataloging of Large-Scale Sky Surveys. In *Proceedings of the Tenth International Conference on Machine Learning*, 112–119. San Francisco, Calif.: Morgan Kaufmann.
- Forbus, K. 1984. Qualitative Process Theory. *Artificial Intelligence* 24(1-3): 85–168.
- Forbus, K.; Nielsen, P.; and Faltings, B. 1991. Qualitative Spatial Reasoning: The CLOCK Project. *Artificial Intelligence* 51(1-3): 417–471.
- Huang, X., and Zhao, F. 2000. Relation-Based Aggregation: Finding Objects in Large Spatial Datasets. *Intelligent Data Analysis* 4(2): 129–147.
- Joskowicz, L., and Sacks, E. 1991. Computational Kinematics. *Artificial Intelligence* 51(1-3): 381–416.
- Junker, U., and Braunschweig, B. 1995. History-Based Interpretation of Finite-Element Simulations of Seismic Wave Fields. In Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95), 1789–1797. Menlo Park, Calif.: International Joint Conferences on Artificial Intelligence.

- Kailath, T.; Schaper, C.; Cho, Y.; Gyugyi, P.; Norman, S.; Park, P.; Boyd, S.; Franklin, G.; Sarasunt, K.; Maslehi, M.; and Davis, C. 1996. Control for Advanced Semiconductor Device Manufacturing: A Case History. In *The Control Handbook*, ed. W. Levine, 1243–1259. Boca Raton, Fla.: CRC.
- Kuipers, B. 2000. The Spatial Semantic Hierarchy. *Artificial Intelligence* 119(1–2): 191–233.
- Kuipers, B. 1986. Qualitative Simulation. *Artificial Intelligence* 29(3): 289–338.
- Kuipers, B., and Levitt, T. 1988. Navigation and Mapping in Large-Scale Space. *AI Magazine* 9(2): 25–43.
- Lorenson, W., and Cline, H. 1987. Marching Cubes: A High Resolution 3D Surface Construction Algorithm. *Computer Graphics* 21(4): 163–169.
- Lu, W.; Han, J.; and Ooi, B. 1993. Discovery of General Knowledge in Large Spatial Databases. Paper presented at the Far East Workshop on Geographic Information Systems, 21–22 June, Singapore.
- Lundell, M. 1996. A Qualitative Model of Physical Fields. In Proceedings of the Thirteenth National Conference on Artificial Intelligence, 1016–1021. Menlo Park, Calif.: American Association for Artificial Intelligence.
- Lundell, M. 1995. A Qualitative Model of Gradient Flow in a Spatially Distributed Parameter. Paper presented at the International Qualitative Reasoning Workshop, 6–19 May, Amsterdam, The Netherlands.
- Munkres, J. 1984. *Elements of Algebraic Topology*. Reading, Mass.: Addison-Wesley.
- Ordóñez, I., and Zhao, F. 2000. STA: Spatio-Temporal Aggregation with Applications to Diffusion-Reaction Phenomenon. In Proceedings of the Seventeenth National Conference on Artificial Intelligence, 517–523. Menlo Park, Calif.: American Association for Artificial Intelligence.
- Pavlidis, T., and Horowitz, S. 1974. Segmentation of Plane Curves. *IEEE Transactions on Computers* C-22(8): 860–870.
- Ramakrishnan, N., and Bailey-Kellogg, C. 2002. Sampling Strategies for Mining in Data-Scarce Domains. *IEEE Computing in Science and Engineering (CiSE)* 4(4): 31–43.
- Ramakrishnan, N., and Grama, A. 2001. Mining Scientific Data. *Advances in Computers, Volume 55*, ed. M. Zekowitz, 119–169. San Diego, Calif.: Academic.
- Rosenblum, L., et al., eds. 1994. *Scientific Visualization: Advances and Challenges*. San Diego, Calif.: Academic Press.
- Samtaney, R.; Silver, D.; Zabusky, N.; and Cao, J. 1994. Visualizing Features and Tracking Their Evolution. *IEEE Computer* 27(7): 20–27.
- Stolorz, P.; Nakamura, H.; Mesrobian, E.; Muntz, R.; Shek, E.; Santos, J.; Yi, J.; Ng, K.; Chien, S.; Mechoso, R.; and Farrara, J. 1995. Fast Spatiotemporal Data Mining of Large Geophysical Datasets. In *Proceedings of the First International Conference on Knowledge Discovery and Data Mining*, 300–305. Menlo Park, Calif.: AAAI Press.
- Weld, D., and de Kleer, J., eds. 1990. *Readings in Qualitative Reasoning about Physical Systems*. San Francisco, Calif.: Morgan Kaufmann.
- Yip, K. 1997. Structural Inferences from Massive Datasets. In Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence (IJCAI-95), 534–541. Menlo Park, Calif.: International Joint Conferences on Artificial Intelligence.
- Yip, K. 1995. Reasoning about Fluid Motion: Finding Structures. In Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95), 1782–1788. Menlo Park, Calif.: International Joint Conferences on Artificial Intelligence.
- Yip, K. 1991. *KAM: A System for Intelligently Guiding Numerical Experimentation by Computer*. Cambridge, Mass.: MIT Press.
- Yip, K., and Zhao, F. 1996. Spatial Aggregation: Theory and Applications. *Journal of Artificial Intelligence Research* 5:1–26.
- Yip, K.; Zhao, F.; and Sacks, E. 1995. Imagistic Reasoning. *ACM Computing Surveys* 27(3): 363–365.
- Zhao, F. 1994. Extracting and Representing Qualitative Behaviors of Complex Systems in Phase Spaces. *Artificial Intelligence* 69(1–2): 51–92.
- Zhao, F.; Bailey-Kellogg, C.; Huang, X.; and Ordóñez, I. 1999. Intelligent Simulation Tools for Mining Large Scientific Data Sets. *New Generation Computing* 17(4): 333–347.



Chris Bailey-Kellogg is an assistant professor of computer science at Purdue University. He earned a BS/MS in electrical engineering and computer science from the Massachusetts Institute of Technology and a Ph.D. in computer and information science with Feng Zhao at Ohio State and Xerox PARC. He conducted postdoctoral research in computational biology with Bruce Donald at Dartmouth University. His research combines geometric, symbolic, and probabilistic approaches for data analysis and experiment planning in scientific and engineering domains. He received a National Science Foundation CAREER award in 2003. His e-mail address is cbk@cs.purdue.edu.



Feng Zhao is a principal scientist at Palo Alto Research Center (PARC), where he directs the Embedded Collaborative Computing Area in the Systems and Practices Laboratory. He is also a consulting associate professor of computer science at Stanford University. His research interests include sensor networks, diagnostics, qualitative reasoning, and control of dynamic systems. Zhao received his Ph.D. in electrical engineering and computer science from the Massachusetts Institute of Technology in 1992. From 1992 to 1999, he was an assistant and associate professor of computer and information science at The Ohio State University. He received the Office of Naval Research Young Investigator Award and the National Science Foundation Young Investigator Award and was an Alfred P. Sloan research fellow in computer science. He has authored or coauthored over 70 peer-reviewed technical papers and is a coinventor of 3 U.S. patents and 5 pending patent applications. His e-mail address is zhao@parc.com.