

Enhanced spatiotemporal relational probability trees and forests

Amy McGovern · Nathaniel Troutman ·
Rodger A. Brown · John K. Williams ·
Jennifer Abernethy

Received: 9 April 2011 / Accepted: 16 February 2012 / Published online: 14 March 2012
© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract Many real world domains are inherently spatiotemporal in nature. In this work, we introduce significant enhancements to two spatiotemporal relational learning methods, the spatiotemporal relational probability tree and the spatiotemporal relational random forest, that increase their ability to learn using spatiotemporal data. We enabled the models to formulate questions on both objects and the scalar and vector fields within and around objects, allowing the models to differentiate based on the gradient, divergence, and curl and to recognize the shape of point clouds defined by fields. This enables the model to ask questions about the change of a shape over time or about its orientation. These additions are validated on several real-world hazardous weather datasets. We demonstrate that these additions enable the models to learn robust classifiers that outperform the versions without these new additions. In

Responsible editor: Eamonn Keogh.

A. McGovern (✉) · N. Troutman
School of Computer Science, University of Oklahoma, 100 W Boyd St., Norman, OK 73019, USA
e-mail: amcgovern@ou.edu

N. Troutman
e-mail: nathanieltroutman@gmail.com

R. A. Brown
NOAA/National Severe Storms Laboratory, 120 David L Boren Blvd, Norman, OK 73072, USA
e-mail: Rodger.Brown@noaa.gov

J. K. Williams · J. Abernethy
Research Applications Laboratory, National Center for Atmospheric Research,
P.O. Box 3000, Boulder, CO 80307, USA
e-mail: jkwillia@ucar.edu

J. Abernethy
e-mail: jabernet75@gmail.com

addition, analysis of the learned models shows that the findings are consistent with current meteorological theories.

Keywords Spatiotemporal relational learning · Statistical relational learning · Hazardous weather

1 Introduction

The task of understanding and predicting hazardous weather phenomena presents a variety of interesting challenges to data mining researchers. Weather is inherently spatiotemporal in nature, which means that any method that aims to understand the underlying causes of a particular weather phenomenon must be able to reason with both spatial and temporally varying data. In addition to reasoning about the spatial and temporal aspects of the data, we need to learn with high-level features. For example, meteorologists already reason about weather using high-level features, such as updrafts (regions where air is moving upward) and downdrafts (regions where air is moving downward) in severe thunderstorms. The statistical relational learning paradigm (e.g., [Friedman et al. 1999](#); [Getoor et al. 2001, 2002](#); [Jensen and Getoor 2003](#); [Richardson and Domingos 2005](#); [Fern et al. 2006](#)) provides such an approach.

In this work, we present a set of significant enhancements to our earlier spatiotemporal relational probability trees (SRPTs) and spatiotemporal relational random forests (SRRFs) ([McGovern et al. 2008, 2010, 2011a](#); [Supinie et al. 2009](#); [Gagne et al. 2010](#)). In our prior work, we demonstrated the effectiveness of these algorithms for classifying spatiotemporal relational data. However, both the prior SRPT and SRRF used a version of the SRPT that was primarily focused on temporal distinctions. In this work, we introduce an enhanced version that can make both spatial and true spatiotemporal distinctions.

In our work on hazardous weather, we deliberately focus on human-readable models such as decision trees (e.g., [Quinlan 1993](#)). This is a significant difference from non-relational spatiotemporal data mining methods such as [Allcroft et al. \(2001\)](#), [Glasbey and Allcroft \(2007\)](#), [Miller and Han \(2009\)](#), [Valdes-Sosa \(2004\)](#), and [Weber and Talkner \(1993\)](#). Human readability allows for verification of the knowledge learned by the model by domain experts and for sanity checks during initial experiment setup. For meteorological applications, if the knowledge we discover while mining the data is to be useful to domain scientists, it needs to be in a form that is understandable.

Our work in mining storms using automated techniques is unique in the meteorology domain. The primary approaches used to study severe weather, such as tornadoes, are through the analysis of data collected within individual supercell thunderstorms (the type that produce tornadoes) and through the interpretation of processes taking place within individual coarsely resolved numerically modeled storms. Doppler weather radars, and especially those mounted on flatbed trucks that can get close to storms, can identify regions of circulation developing within the storms (e.g., [Wurman et al. 1996](#); [Bluestein et al. 2007](#)). However, only a minority of storms that exhibit circulations go on to produce tornadoes (e.g., [Trapp et al. 2005](#)). Meteorologists have developed various theories for tornado formation (e.g., [Markowski et al. 2003](#),

Davies-Jones 2008), but they are not able to identify the specific storm that will produce a tornado.

Researchers also consider the numerical modeling of supercell storms in the hopes of resolving the matter (e.g., Klemp and Rotunno 1983; Wicker and Wilhelmson 1995; Snook and Xue 2008). Though only approximating what occurs in nature, numerical models have the advantage of simulating all of the dynamical processes involving temperature, pressure, moisture, types of precipitation, wind fields, etc. that take place within a storm. Unfortunately, the modeling undertaken to date is rather coarse (owing to computer processing constraints) and the environmental conditions chosen have been rather limited. By mining a large number of more finely resolved numerically modeled storms produced under a variety of realistic environmental conditions, we hope to discover the basic processes that lead to tornado formation that have been elusive using the traditional approaches.

Many data mining methods use an incomplete approach to modeling the spatio-temporal aspects of data, resulting in models that are focused primarily on the spatial or temporal features alone. For example, although Sharan and Neville (2007, 2008) introduce the ability to reason with temporally varying relationships, they do not allow the algorithm to use temporally varying attributes, nor do they study spatially or spatiotemporally varying data. Likewise, Bodenhamer et al. (2009) primarily focus on the spatial aspect of the data. Our research focuses on improving the modeling of the *combination* of spatial and temporal features. By creating a method that has the ability to reason with temporal, spatial, and spatiotemporal relational data, we improve the ability of spatiotemporal relational models to accurately model real-world phenomena including hazardous weather.

Other than our prior work, the most closely related work we are aware of is that of Neville et al. (2003) and Breiman (2001). Neville et al. introduced relational probability trees (RPTs), which are a form of probability estimation trees (PETs) (Provost and Domingos 2000) that work with relational data. These trees have been proven to be successful in a wide variety of domains (Neville et al. 2005; Fast et al. 2007; Neville and Jensen 2007) but they do not address spatial, temporal, or spatiotemporally varying data. We use the Random Forest framework (Breiman 2001) in creating the SRRFs.

Our major contributions to the SRPT/SRRF models are as follows. First, we introduce the ability to learn directly with spatially and spatiotemporally varying attribute fields within objects. These *fielded objects* can contain 2D or 3D temporally varying scalar or vector fields. By retaining the fields, we enable the trees to distinguish data using two new classes of distinctions. The first is based on applying mathematical functions to the fields. For scalar fields, we can apply the gradient function. For vector fields, we can reason about the curl and divergence of the fields. The second type of distinction that the fielded objects enable is reasoning about the shape of features defined by the fields. The SRPT/SRRF models can examine the shape at a particular moment in time, how it changes over time, and how it changes in orientation over time. In addition to the new field-based distinctions, we added a distinction that selects a single object and applies another distinction to only that object, and a conjugate Boolean distinction that combines two distinctions with Boolean-and or Boolean-or.

In [McGovern et al. \(2011a\)](#), we presented work on the application of the newly enhanced SRPT/SRRFs to a wide variety of severe weather applications. The contributions of this paper over [McGovern et al. \(2011a\)](#) are as follows. First, this paper introduces the concept of fielded objects and discusses how we can make use of these fields within the trees. This paper also introduces the single object and Boolean distinctions. Second, this paper focuses on a full parameter analysis. Third, our application domains differ. While both papers focus on real-world severe weather data, including convectively-induced turbulence hazardous to aviation, this paper primarily focuses on simulated supercell thunderstorms with the goal of improving our understanding of tornadoes. Both papers use data about frontal passages in Oklahoma combined with observations of tornadoes but the type of data for the experiments differs. Last, the [McGovern et al. \(2011a\)](#) paper had an additional weather domain (drought) while this work focuses on the simulated points domain for a full sensitivity analysis.

We next describe the SRPT algorithm followed by the details of our enhancements. We then present extensive empirical results on one simulated domain and two real-world hazardous weather domains.

2 Enhanced SRPTs

2.1 Spatiotemporal relational data

In order to learn from spatiotemporally varying relational data, we need an efficient representation of such data. We make use of the attributed graph representation first used by [Neville et al. \(2003\)](#) and modified by our previous research on spatiotemporal relational data ([McGovern et al. 2008](#); [Supinie et al. 2009](#); [Gagne et al. 2010](#); [McGovern et al. 2010](#)). We again modify this approach to represent fielded objects.

Figure 1 shows a schema of our spatiotemporal relational data for the severe storms domain that we describe in detail in the experimental results. A schema shows the types of objects that are possible in a graph as well as the possible relationships between the objects. The example schema shows that meteorological features, such as regions of hail and updrafts and downdrafts, are represented as objects. Any such regions that appear in the simulation will appear as objects within the graph representing that simulation. In this example, the possible relationships are all spatial (nearby, equals, contains, and overlaps) and can be between any two objects.

Objects and relations can be either static (existing for the duration of the graph) or dynamic. If they are dynamic, they are tagged with a start and an end time. In the meteorological example, all objects are dynamic. Temporal consistency is enforced for all dynamic objects or relations. This ensures that objects or relations only exist during the stated lifetime of the graph and that relations can only exist while the objects they connect exist.

In addition, a schema shows the possible attributes on both objects and relations. Both objects and relations are required to be assigned a type and graphs are required to be labeled (the label does not have to be binary). The types in the example are shown in bold. Attributes can also be static or dynamic. If an attribute is static, its value stays the same throughout the lifetime of the object or relation. If it is dynamic, it can vary

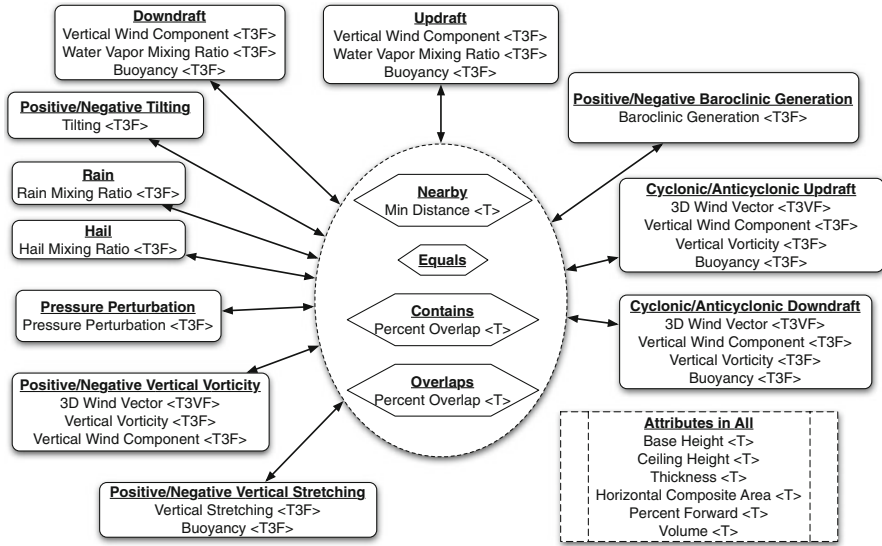


Fig. 1 Schema for the simulated supercell storms domain. The rounded rectangles represent objects and the hexagons show the relations. The type of each object or relation is *bolded* and the arrows show the directionality of the relationships. Attributes are listed inside both objects and relations. *T* denotes a temporally varying attribute; *T3F* denotes a 3D spatiotemporally varying field of scalars and *T3VF* denotes a 3D spatiotemporally varying field of vectors

as a function of time or with space and time. In the example, two attributes associated with a region of hail would be a temporal array representing the height of the hail region over time (ceiling height) and a 3D fielded attribute showing the quantity of hail at each grid point within the object (volume).

2.2 Fielded objects

The high-level objects used by the SRPT are usually derived from raw data that have been aggregated in some way. For instance, in the supercell storms dataset, a storm might contain a vorticity object representing an area of strong rotational winds. There are a variety of attributes that could be measured inside this object. For example, the current average wind speed and direction could be saved as a single vector. The benefit of this type of high level representation is its compactness and simplicity. However, information is lost since the vorticity object being represented occupies a volume of space with wind speed and direction (amongst other variables) available at many points within the region and we would prefer to maintain all of that information.

In our previous work, all attributes were reduced to a single discrete, scalar or vector value. These values could vary over time, producing temporal-discrete, temporal-scalar, and temporal-vector valued attributes. One of the main contributions of the work described in this paper is the use of the underlying raw data of such high level objects. In geographic information systems (GIS), these data are called fields, and the addition of fields to the high-level objects will produce *fielded objects* (Goodchild

et al. 2007; Cova and Goodchild 2002). By maintaining access to the full underlying field, the SRPT is able to learn directly using the spatial data. Note that each object can have multiple spatiotemporal fields associated with it, and the fields may encompass some of the area surrounding an object.

2.3 Shape detection

In addition to reasoning about the values of the fields within objects or relations, we can reason about the spatial shapes of point clouds defined by the fields. For 3D fields, the possible shapes we identify are: spheres, boxes, cylinders, and cones. In 2D, the possible shapes are: circles, ellipses, rectangles, and isosceles triangles. These shapes are shown in Fig. 2. We chose these shapes because they represent the basic shapes of many objects and each has a known geometric method for determining how far a point lies from the surface and determining the normal surface vector at any point on

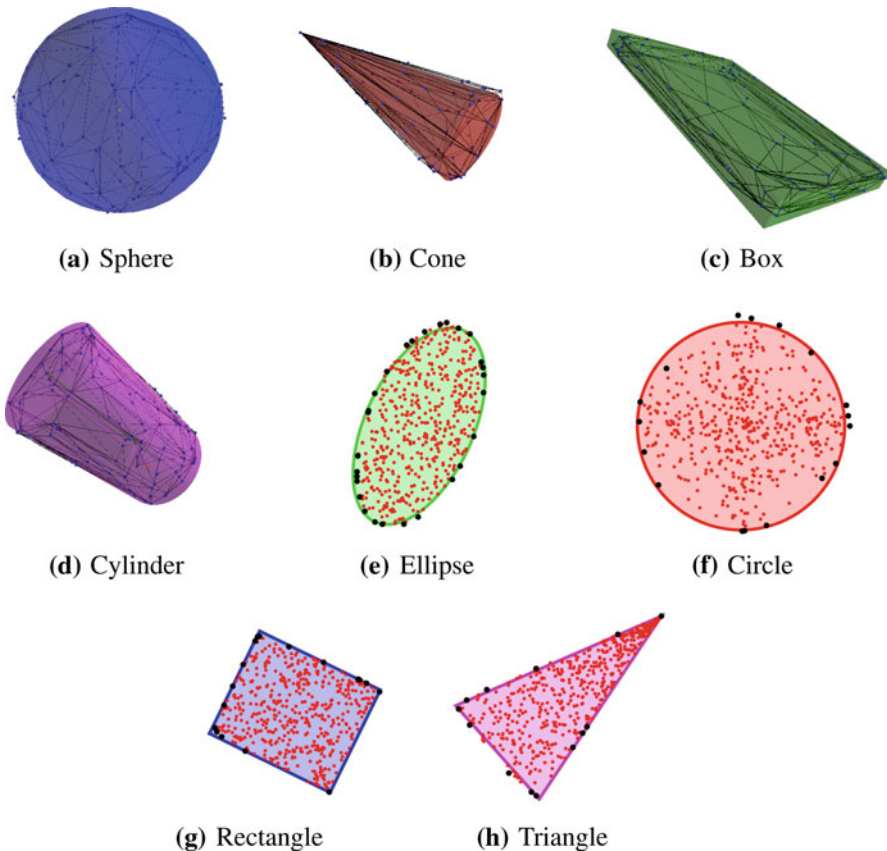


Fig. 2 a–d Examples of detecting each of the primitive 3D shapes. The convex hull is drawn as a wire frame with the points on the hull shown in blue. e–h Examples of detecting each of the primitive 2D shapes. Red points are the full field with black points being on the convex hull

the surface of the shape. We briefly describe the shape detection algorithm below and a full, formal description can be found in Troutman (2010).

Schnabel et al. (2007) presented a stochastic shape detection algorithm that identified multiple shapes within a point cloud. We initially implemented this approach but found it to be both ineffective at identifying the correct shape and overly computationally complex. Because we are only interested in detecting primary shapes, we devised a deterministic algorithm based on the known geometry of the candidate primitive shapes. First, the entire point cloud was reduced to a convex hull using Qhull (Barber et al. 1996) (note that the primitive shapes we use are all convex). Qhull also calculates the face normals, allowing us to calculate the vertex normal¹ for each point on the convex hull. Our algorithm also relies on the minimum-volume bounding box of the convex hull, which is also a difficult problem. We used a variant of the approach described by O'Rourke (1985).

To decide which primitive shape matches most closely to the data, we used the vertex normals calculated by Qhull. The error for a proposed shape is the sum of the differences between the observed vertex normals and the expected vertex normals for that shape, all divided by the number of sample points. The angle between the sample normal and the shape normal must be less than a given threshold (set to 25°). If the number of normals whose angles don't pass the threshold is less than half the total number of sample points, the shape is automatically rejected. The primitive shape that is selected as the best match is the one with the lowest error. This method of error estimation greatly improved the quality of fit for detected shapes. Figure 2 shows an example of point clouds each created from a shape and the resulting matched shape for all of our 2D and 3D primitive shapes.

2.4 SRPTs

SRPTs are PETs (Provost and Domingos 2000) for use on spatiotemporal relational data (McGovern et al. 2008). Instead of the usual attribute-based questions used by standard decision trees such as C4.5 (Quinlan 1993) or the questions about the relational structure of the data (such as the degree of a node) used by RPT (Neville et al. 2003), SRPTs split the data using new questions about spatial, temporal, and spatiotemporal characteristics. The prior version of the SRPT had eight classes of distinctions, or questions, used to split the data. They were divided into two categories: non-temporal and temporal. We describe each of these distinctions below. Each class of distinction functions as a template and the SRPT learning algorithm fills in the parameters while growing the trees. More detail about the learning algorithm is given below.

When a distinction has a set of fixed options, these are listed in brackets. For example, the *Attribute* distinction can make use of several statistics functions to split the data. Some distinctions are meta-distinctions that ask questions about the results of other distinctions. An example of a meta-distinction is the *count* distinction. It takes another (non-meta) distinction and counts the number of matching items in the graph.

¹ http://en.wikipedia.org/wiki/Vertex_normal.

To simplify the notation in the distinctions described below, if the question can apply to either an object or a relation, we call it an *item*.

The non-temporal distinctions are:

- *Exists*: Does an item of type t exist?
- *Attribute (Temporal-Scalar)*: Does an item of type t have a temporal-scalar attribute a with a [MIN, MAX, MEAN, STD] value $\geq v$?
- *Attribute (Discrete)*: Does an item of type t with discrete attribute a have a value $= v$?
- *Count (Meta)*: Are there at least n yes answers to non-meta distinction d ?
- *Structural (Meta)*: Is an answer to non-meta distinction d related to an object of type t through a relation of type r ?

The temporal distinctions are:

- *Temporal Exists*: Does an item of type t exist for at least k timesteps?
- *Temporal Ordering*: Do the matching items from base distinction a occur in a temporal relationship with the matching items from a base distinction b ? The seven types of temporal ordering are: before, meets, overlaps, equals, starts, finishes, and during (Allen 1991).
- *Temporal Partial Derivative (Non-Field)*: Is the partial derivative with respect to time on an attribute a of an item of type $t \geq v$?

The enhancements we introduce to the SRPT algorithm allow for the addition of a number of new spatial and spatiotemporal distinctions based on the fielded objects. The field-based distinctions are as follows. These distinctions vary in their behavior based upon the type of field (vector or scalar) and the field function applied (gradient, curl, or divergence). For clarity, these are listed as separate distinctions and the attributes are all assumed to be temporal.

- *Field Gradient*: Does the angle between a vector \mathbf{v} and the [MIN, MAX, MEAN, STD] vector of the gradient of a scalar-field attribute a on an item of type t ever have an angle $\geq \theta$?
- *Field Curl*: Does the angle between a vector \mathbf{v} and the [MIN, MAX, MEAN, STD] vector of the curl of a vector-field attribute a on an item of type t ever have an angle $\geq \theta$?
- *Field Divergence*: Is the [MIN, MAX, MEAN, STD] of the divergence of a vector-field attribute a on an item of type t ever $\geq v$?
- *Field Attribute Gradient*: Does an item of type t with a field attribute a have a [MIN, MAX, MEAN, STD] of the magnitude of the vector of the gradient $\geq v$?
- *Field Attribute Curl*: Does an item of type t with a field attribute a have a [MIN, MAX, MEAN, STD] of the magnitude of the vector of the curl $\geq v$?
- *Field Attribute Divergence*: Does an item of type t with a vector-field attribute a have a [MIN, MAX, MEAN, STD] of its [DIVERGENCE] $\geq v$?
- *Temporal Partial Derivative (Field)*: Is the partial derivative with respect to time of the magnitude of the [MIN, MAX, MEAN, STD] of the [CURL, DIVERGENCE, GRADIENT] of the field-attribute $a \geq v$?

In addition to examining the values of the field attributes, we examine the shape of point clouds defined by the fields. The shape-based distinctions are listed below.

- *Shape (3D)*: Is the primary 3D shape of field f a rectangular-prism, sphere, cylinder, or cone?
- *Shape (2D)*: Is the primary 2D shape of field f a rectangle, circle, isosceles triangle, or ellipse?
- *Shape Change*: Has the shape of field f changed from one of the primary shapes over to a new shape at any point?
- *Shape Tilt*: Has the primary axis of the shape of field f , at any timestep, tilted by at least θ compared to the vector \mathbf{v} ?

We also added several distinctions that allow the SRPT to combine two base distinctions and to select, and reason on, a single object. These are listed below.

- *Boolean AND Conjugate*: Does a graph end up in the yes branches of both base distinction a and base distinction b ?
- *Boolean OR Conjugate*: Does a graph end up in either of the yes branches of base distinction a or base distinction b ?
- *Single Item Select By Attribute*: Does the single item of type t with the [MIN, MAX] value of the attribute a pass the base distinction b ?
- *Single Item Select Oldest*: Does the oldest item of type t pass the base distinction b ?
- *Single Item Select List First*: Does the first item of type t pass the base distinction b ? (Ordering is arbitrary, but consistent,)

The construction of an SRPT follows the standard greedy approach used by most decision tree algorithms (Russell and Norvig 2009). We outline the process in Algorithms 1 and 2. The tree grows recursively until adding a new node, or distinction, fails to yield a statistically significant change in the classification distribution or is stopped by the parameters to the learning algorithm. Growth of the trees is controlled by the following user-supplied parameters:

- α : p value threshold
- S : number of distinctions examined at each node of tree growth
- D : max tree depth

Growing an SRPT requires a set of labeled training graphs. As the tree is built, the set of graphs is recursively split by each distinction. Choosing these distinctions is the main work of the learning algorithm. Distinctions are sampled stochastically from the set of all possible distinctions and the “best” distinction is taken. To determine which distinction is best, we calculate the χ^2 statistic of the contingency table where the rows are the yes and no branches and the columns are the class distributions. As long as the p value is below the user-specified threshold, growth continues along that branch of the tree. We do not currently use a Bonferroni correction (Jensen and Cohen 2000) on the p value threshold, although that can easily be added in future work.

Since the space of possible distinctions is extremely large, stochastic sampling is used to find the best distinction. To do this, the distinction is treated as a template. Fixed values, such the choice of statistical test, are chosen randomly. Specific values, such as a splitting point, are chosen from a graph chosen at random. This ensures the values are realistic. Because so many distinctions are sampled during learning, there is no optimization of the parameters to each template. In early versions of the

Algorithm 1: GrowSRPT: Algorithm used for growing an SRPT

Input: S = Number of samples, D = Maximum depth of tree, $\alpha = p$ value threshold, G = training data, d = current tree depth

Output: A SRPT

```

if  $d \leq D$  then
  node = FindBestSplit( $G, S, \alpha$ )
  if node  $\neq \emptyset$  then
    yesSubset, noSubset = SplitData(node,  $G$ )
    node.addYesBranch(GrowSRPT( $S, D, \alpha, \text{yesSubset}, d + 1$ ))
    node.addNoBranch(GrowSRPT( $S, D, \alpha, \text{noSubset}, d + 1$ ))
    Return node
  end
end
Return new LeafNode( $G$ )

```

Algorithm 2: FindBestSplit: Algorithm for finding the best split in a set of graphs

Input: G = training data, S = number of samples, $\alpha = p$ value threshold

Output: An internal Node or \emptyset

```

best =  $\emptyset$ 
for  $i = 1$  to  $S$  do
  distinction = new InternalNode(RandomDistinction( $S, G$ ))
   $fit, p = \text{Calc}\chi^2(\text{distinction}, G)$ 
  if  $p \leq \alpha$  and  $fit > \text{best.fit}$  then best = distinction
end
Return best

```

SRPT algorithm, the sample count was chosen according to a heuristic that relates the number of samples to a desired confidence of finding the best distinction (Srinivasan 1999). Because the SRPT does not meet his requirements of uniform sampling over the distinction space, we allow the user to supply a parameter, S , that directly specifies the number of distinctions to sample at each level.

With no controls, a tree can potentially grow quite deep, resulting in over-fitting and unnecessary complexity. The maximum size of the tree is controlled by D or the *max-depth* parameter, which limits the tree's depth by forcing it to terminate early. The depth can range from zero, which means the tree will only have a root distinction, to a depth limited only by machine memory (which is also limited by the size of the training data, G). At times, the p value threshold (α) will cut off growth along a branch before the maximum depth is reached; on the other hand, *max-depth* can cut off the growth even if there still remain distinctions that would meet the p value requirement. At each node of the tree, the best distinction is chosen from among all distinctions sampled. In order to study the SRPT algorithm under various constraints, the algorithm can be limited to using only a subset of the possible distinctions.

Tree growth terminates when the tree has reached its maximum depth, no statistically significant split can be found, the number of graphs in the current branch has dropped below 20, or the graphs in the current branch comprise only one class. If the number of graphs is below 20, then the χ^2 test is not stable. We use the Laplace corrected estimate of class probability at each leaf node (Provost and Domingos 2000).

A common difficulty when dealing with real-world data is handling missing values (Liu et al. 1997). In the spatiotemporal framework, there are three cases of missing values. The first occurs when a non-temporal attribute is missing its value, which is analogous to missing values in propositional data. This is handled easily because the data are assumed to be heterogeneous and our algorithms do not require each object or relation to have every attribute that is possible for that type. The second and third cases deal with temporal and spatial attributes. Temporal attributes could be missing time-steps within the temporal stream, but still have value for other timesteps. Likewise, spatial attributes could be missing some values, but still have many values present. Each of these cases needs to be handled in a different manner.

The simplest method is to discard objects that contain missing values (White 1987). This can have detrimental effects in propositional data, but is likely to have even larger effects in relational data as the removal of an object would also remove all the relationships connected to it. The power of relational algorithms comes from the ability to leverage the structure imposed by relationships. In the case of SRPTs, it would be possible to simply remove the attribute from the single object, leaving the rest of the attributes and the relations unchanged. While the SRPT can handle a missing attribute, it is likely that the removal of the good values would be detrimental.

The alternative to removal is to replace the missing values through imputation. Many different techniques have been suggested, each with strengths and weaknesses, such as inferring the value from objects in the same class (Kononenko et al. 1984) or using decision trees to predict the value (Quinlan 1993). Many other techniques for handling missing data have also been proposed (Breiman et al. 1984; White 1987; Friedman et al. 1996). Previous SRPT work made no attempts to handle missing data and required all data to be preprocessed such that the data given to the algorithm were pristine. We have added the ability to interpolate missing spatial and temporal data, as described below.

To handle cases two and three of missing values, we replace the missing value with the mean of the non-missing nearest neighbors. For case two, a 1D temporal attribute, the nearest neighbors are temporally local. Hence if a_t is missing, it is replaced with the linear interpolation from the two temporally closest existing values, a_{t-1} and a_{t+1} , on either side of a_t . For case three, the nearest neighbors are spatially local, regardless of whether the attribute is temporal or not. In 2D, the nearest neighbors are the eight touching values (counting diagonal as touching), and in 3D they are the 26 touching values. The missing value is replaced by the mean of the nearest neighbors that are not missing. Missing values are replaced in order of increasing indices of the z , y , and x dimension ($\langle 0, 0, 0 \rangle$, $\langle 0, 0, 1 \rangle$, ..., $\langle 0, 1, 0 \rangle$, ..., $\langle 1, 0, 1 \rangle$, ..., $\langle n_x, n_y, n_z \rangle$).

The use of a simple mean of nearest neighbors introduces some degree of temporal (case two) and spatial (case three) autocorrelation. Research has shown that real-world data are frequently autocorrelated (O'Sullivan and Unwin 2002; Longley et al. 2005; Neville and Jensen 2005). This autocorrelation is not always a problem unless an algorithm makes underlying assumptions of independent data. Hence, the deliberate use of the autocorrelation to improve missing values should significantly improve the results over simply removing the attribute.

In this work, we use the new SRPTs described above to create the SRRFs as well. A SRRF is a random-forest based SRPT, constructed using bagging. The learning

algorithm for the SRRFs is described in [McGovern et al. \(2010\)](#). By giving the SRRFs more powerful spatiotemporal trees, the forest also becomes more powerful in making spatiotemporal distinctions. In our experimental results, we examine both individual SRPTs and SRRFs. When the SRRF is used to classify, each SRPT in the forest votes and the majority vote is used. Other voting schemes are possible and left for future work. When the SRRF is used to produce probabilities, the average probability from each SRPT in the forest is used.

3 Empirical results

In order to fully explore each of the parameters and datasets, we performed a factorial experiment across the three main parameters used in the construction of SRRFs on each of the three datasets. We varied the number of trees from 1 to 100 by testing [1, 10, 50, 100]. Note that a SRRF containing a single tree is equivalent to an SRPT. The number of distinctions sampled at each node were [10, 100, 500, 1000], and the maximum depth of the tree was [1,3,5]. In total, this makes 48 parameter sets. Each set was repeated 30 times for statistical analysis.

We measure the classification power of both the SRPT and SRRFs using the Gerrity skill score (GSS). We chose GSS because it was designed for measuring performance on imbalanced multi-class classification problems. GSS is Gandin and Murphy’s equitable skill score ([Gandin and Murphy 1992](#)) with a scoring matrix derived to satisfy the constraints of symmetry and equitability as stated by Gandin and Murphy. To calculate GSS, let S be an equitable scoring matrix whose calculation is given below and E be a matrix (contingency table) such that $e(i, j)$ is the relative frequency of instances with the true class i being classified as class j . From S and E , $GSS = \text{trace}(S^T E)$. To calculate S , following [Gerrity \(1992\)](#), let $P(r)$ be the relative frequency of class r and define the following:

$$D(n) = \frac{1 - \sum_{r=1}^n P(r)}{\sum_{r=1}^n P(r)}, \text{ and}$$

$$R(n) = \frac{1}{D(n)}.$$

Let K be the number of classes and $\kappa = \frac{1}{K-1}$. The elements of S are calculated as:

$$s_{n,n} = \kappa \left[\sum_{r=1}^{n-1} R(r) + \sum_{r=n}^{K-1} D(r) \right] \quad n = (1, 2, \dots, K)$$

$$s_{m,n} = \kappa \left[\sum_{r=1}^{m-1} R(r) + \sum_{r=m}^{n-1} (-1) + \sum_{r=n}^{K-1} D(r) \right] \quad 1 \leq m < K, m < n \leq K$$

$$s_{n,m} = s_{m,n} \quad 2 \leq n \leq K, 1 \leq m < n$$

GSS varies from -1 to 1 ([Gerrity 1992](#)). A value of -1 indicates “intentionally” classifying incorrectly; a value of 0 is equivalent to the performance of a random classifier; and a value of 1 indicates perfect classification.

We considered a variety of skill metrics including Heidke's skill score (HSS) and Matthew's correlation coefficient (MCC). Previous work on SRPTs and SRRFs used true skill score (TSS) (aka Hanssen–Kuipers discriminant, true skill statistic, and Peirces's skill score); however, we chose GSS to replace TSS as it is believed to be a better skill indicator without the bias introduced by rare events that TSS exhibits (Jolliffe and Stephenson 2003; Marzban 1998). GSS and TSS are identical in the case of two classes. In addition to GSS, we also measure area under the receiver operator characteristic curve (AUC) on domains that only contain two classes. AUC has a long, respected history in evaluating machine learning algorithms (Bradley 1997; Egan 1975). AUC measures the robustness of the classifier across the various probability thresholds (Provost and Fawcett 2001). AUC varies from 0 to 1, with 1 indicating a perfect classifier and 0.5 indicating a random classifier.

To show the improvement arising from the addition of the fielded objects and the new distinctions, we also generate a non-fielded version of each dataset. This was done by reducing each field into four new attributes based on the minimum, maximum, mean, and standard deviation of the field within the object. This is how the data would have been processed before the addition of fields. We repeated the same parameter search over the non-field datasets as we did for the fielded datasets. This allows for comparison of the performance between fields and non-fields.

In order to further study the results, we also calculated variable importance using a SRRF trained with 100 trees, 1000 distinction samples and a max depth of 5 for each domain. The method for calculating variable importance in the SRRF is described in McGovern et al. (2011a) and follows the same approach as Breiman (2001). As the largest and deepest forests among the parameters considered, these forests appear to provide the most robust and reliable information about variable importance. In addition to calculating variable importance, we also calculated the frequency with which distinctions were picked, as well as the frequency with which objects, relations, and attributes were used in those distinctions. Distinctions that referenced an attribute counted toward the object or relation counts. Also, relations were counted toward both the source and destination object counts.

3.1 Synthetic domain: points

Points is a synthetic dataset designed to test the new capabilities of SRPTs, particularly shape detection. Each graph is comprised of 3–10 objects each having two attributes: a discrete color and a 3D scalar field of {0, 1} values. This field represents a point cloud with a 1 indicating a point at that location and 0 indicating empty space. A single relationship type is used with a random number of relations within each graph. There are three classes in the Points dataset: *change*, *grow*, and *flip*. Each of the classes is based only on color and/or the field. *Change* has a box turn into a sphere and a cone turn into a cylinder. *Grow* has three blue spheres that grow. *Flip* has a cone that flips 180° along its axis. To generate the data, we randomly assigned a uniform distribution of class labels from the three classes. Using the class labels, we filled in the required number of objects with the correct attributes to meet the definition of each class label. Additional randomly generated objects were added to the graph along with relations

Fig. 3 Schema for points dataset: attributes tagged $\langle T \rangle$ are temporal, $\langle T3F \rangle$ are temporal 3D fields, and untagged attributes are discrete

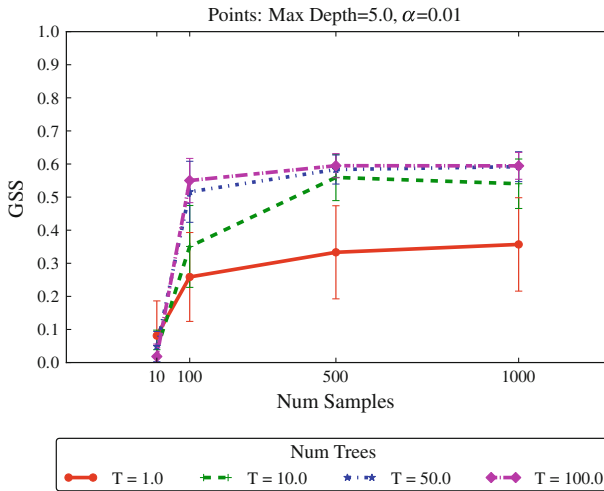
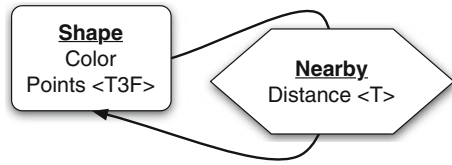


Fig. 4 GSS on the test set for points varying across S and T with $D = 5$. Error bars are one standard deviation

that randomly related objects within the graph. These extra objects and relations served to add noise. The schema for Points is shown in Fig. 3.

Figure 4 shows the performance of the SRRFs on Points for a maximum tree depth $D = 5$. Increasing the number of distinctions sampled, S (x -axis), increases performance initially but it quickly levels out. Increasing the number of trees, T , also increases performance, though it also levels out between 50 and 100 trees. Figure 5 shows the GSS of the SRRFs as a function of the maximum depth of the trees. There is not much visual difference between depth 5 and 3 but between 3 and 1 there appears to be some difference. This trend exists in the other domains as well and we examine depth 5 results for the other domains.

Table 1 contains an N-way analysis of variance (ANOVA) for the factors: number of samples (S), number of trees (T), and max depth (D) versus GSS. The first three columns show the significance of single factor effects and the remaining four show interaction effects between parameters. Each of the three factors, S , T , D individually show a statistically significant effect on GSS. This means that a change in the factor produces a statistically significant change in GSS. These results are consistent with common assumptions on decision trees and random forests. Increasing sampling should allow the SRPTs to find better distinctions, and this is supported by statistics. Increasing the number of trees and max depth should also increase performance by

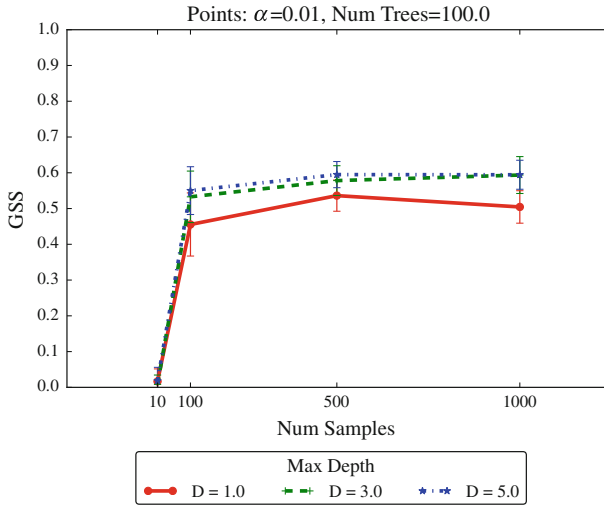


Fig. 5 GSS for on the test set for points varying across S and D with T = 100. Error bars are one standard deviation

Table 1 N-way ANOVA for points across all parameters

GSS	T	S	D	T × S	T × D	S × D	T × S × D
<i>p</i> value	≈0	≈0	≈0	≈0	0.541	4.98e−4	0.004

Values in *bold* are significant with $\alpha = 5\%$. ≈0 implies a value ≤ 0

allowing the SRRFs to capture more complex concepts. Again, the effect of each of these parameters is supported by the statistics.

The interaction between number of samples and number of trees means that varying the two together produces more of a change in GSS than varying either one alone. The interaction effect between samples and depth is also significant. While ANOVA reports that increasing the number of trees (and number of samples) is statistically significant, if we look at a single slice, such as the right-most set of points in Fig. 4, it is unclear whether varying the number of trees for 1000 samples and max depth 5 produces a significant change in GSS. To test if there is a difference, we performed an unpaired bootstrap randomization test with 10,000 samples to compare the mean GSS. The results are shown in Table 2. The tests indicate that there is a statistically significant difference between 1 and 10 trees, between 10 and 50 trees, but not between 50 and 100 trees. We also applied the bootstrap randomization to examine the effect of varying the number of distinctions sampled at 100 trees. It shows a significant difference between 10 and 100 samples and between 100 and 500 samples, but no significant difference between 500 and 1000 samples. These results confirm the asymptotic behavior visually observed in the learning curves.

Table 3 contains the distinction counts and the counts of the attributes used by the distinctions. From the table, we can see that the SRRFs are mostly picking distinctions that make sense given the definition of each class. *Count* is the most frequently

Table 2 Unpaired bootstrap randomization on GSS for $S = 1000$ with $D = 5$ between various T ; and for $T = 100$ with $D = 5$ between various S

Number of trees	1 and 10	10 and 50	50 and 100
p value	≈ 0	0.003	0.819
Number of samples	10 and 100	100 and 500	500 and 1000
p value	≈ 0	0.003	0.956

Values in *bold* are significant with $\alpha = 5\%$

Table 3 Distinction counts for the top 10 distinctions and top 2 attributes

	%Total	Total	Shape.Points	Shape.Color
% Total	100		32	15
Total		398,531	125, 699	60, 422
Count	20	78,484		
Attribute.Exact	10	39,879		39, 879
Shapes.Tilt	9	34,195	34, 195	
Field.Gradient	7	28,660	28, 660	
Attribute.LessThan	5	20,543		20, 543
Shapes.Match	5	18,766	18, 766	
FieldAttribute.Gradient	4	16,927	16, 927	
Shapes.Change	4	16,614	16, 614	
Exist	4	15,783		
TemporalExists	3	13,274		

chosen distinction. Counting is an important element of the definition of the *grow* class. The second most frequently chosen distinction is *Attribute.Exact* on the color attribute. Color is critical for successful classification, so the frequent use of matching color makes sense. *Shapes.Tilt*, *Shapes.Match* and *Shapes.Change* also show up as we would expect given the definition of each class.

The Points domain is designed specifically to highlight the power of fielded objects. Figure 6 shows the results of training SRRFs on a non-fielded version of the Points dataset. The resulting performance shows that the SRRFs retained some ability to learn despite the loss of fields. However, having the fields in the data does produce statistically significantly higher GSS, according to an unpaired bootstrap randomization test with 10,000 resamplings, with a p value of ≈ 0 .² The variable importance analysis for this non-fielded case shows that *shapes.color* is by far the most important attribute. The next closest attribute's variable importance isn't statistically significantly different from 0. After examining some of the trees and returning to the definition of the classes, the requirement that the *grow* class must contain three blue spheres makes

² Despite the algorithm for bootstrap randomization returning ≈ 0 , with 10,000 resamplings the strongest statement that can actually be made about p is $p < 10^{-4}$.

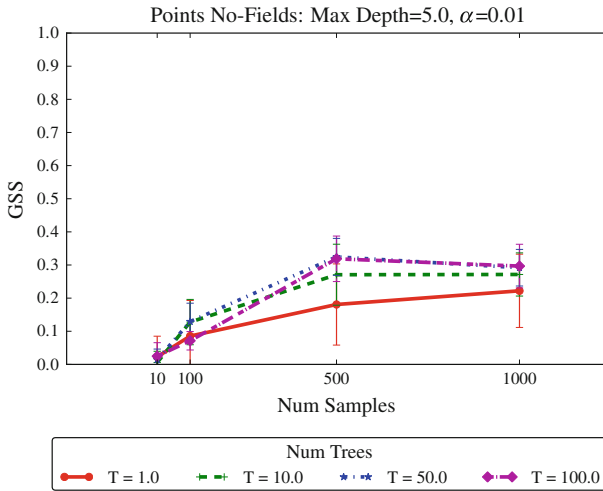


Fig. 6 GSS on the test set for no-fields points varying across S and T with $D = 5$. Error bars are one standard deviation

that class predictable even without the ability to check the other requirements of the class definition.

3.2 Understanding convectively induced turbulence

Turbulence is a major hazard for aviation that causes flight delays, routing changes and bumpy rides for passengers. In addition, the turbulence encountered by aircraft can cause structural damage, serious injuries and even fatalities. Providing better information about probable locations of turbulence to airline dispatchers, air traffic control systems, and pilots will enable them to avoid unnecessary ground delays and plan efficient routes that mitigate or avoid turbulence encounters. The National Center for Atmospheric Research (NCAR) is involved in ongoing work to develop turbulence nowcast and forecast products for the Federal Aviation Administration's Aviation Weather Research Program. An NCAR-developed clear-air turbulence forecast product, graphical turbulence guidance (GTG), is currently in operational use by the National Oceanic and Atmospheric Administration's Aviation Weather Center. In accordance with plans for the FAA's Next Generation Air Transportation System, additional capabilities for GTG are under development to enhance its ability to nowcast convectively-induced turbulence (CIT), that is, turbulence in and around storms, using data sources such as radar, satellite and numerical weather prediction (NWP) models (Sharman et al. 2006; Williams et al. 2008). In this section, we specifically focus on developing improved prediction of CIT.

Because pilot reports of turbulence are subjective and lack highly accurate time and position, NCAR developed an automated "in-situ" turbulence reporting system for aircraft (Cornman et al. 1995, 2004). Several major airlines have outfitted planes with this system, which measures eddy dissipation rate (EDR), an

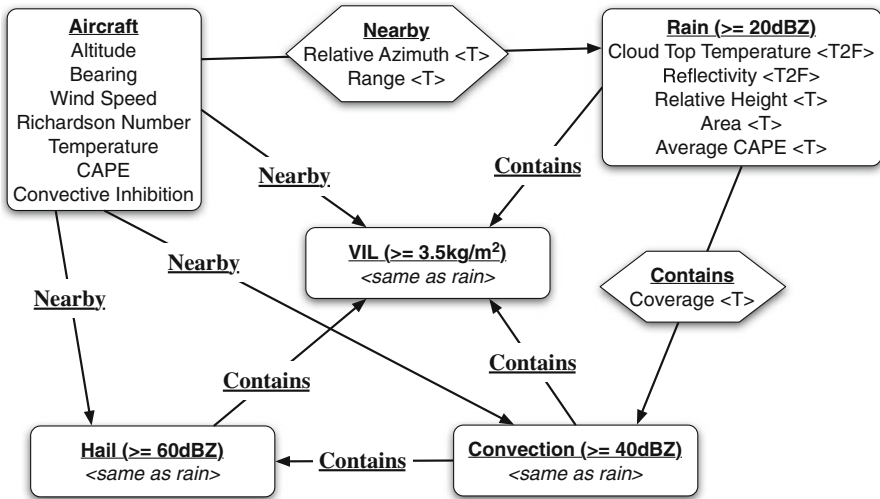


Fig. 7 Schema for the convectively induced turbulence data. *CAPE* is convective available potential energy, a measure of the potential storm severity

aircraft-independent measure of atmospheric turbulence. In the data used for this study, the peak EDR values for one-minute flight segments are used. Our EDR measurements come from United Airlines Boeing 757 aircraft that flew in the summer of 2007 over the continental United States (CONUS). We combine these reports with archived radar and satellite weather observations and archived real-time NWP forecast data. We focus on a summer period because it has a higher rate of convective activity and associated air traffic delays.

Figure 7 shows the schema for the spatiotemporal relational data we created from the raw turbulence dataset. The aircraft is the key object since it is co-located with the turbulence measurement. Starting with the aircraft’s location, we examined the atmosphere within a 40 nautical mile radius. We applied a threshold to the composite radar reflectivity to outline 2D areas of “rain” (≥ 20 dBZ), “convection” (≥ 40 dBZ), and “hail” (≥ 60 dBZ). Note that these names are meant to be suggestive rather than meteorologically precise. We also created areas of vertically integrated liquid (VIL) $\geq 3.5 \text{ kg m}^{-2}$. VIL is a radar-derived estimate of the total amount of liquid water in a column, and these regions of high VIL are generally indicative of convective storms. Each of the regions has several 2D temporally-varying spatial fields derived from archived satellite and radar data and each is spatially related to the aircraft object. In order to isolate convection as the most likely source of the turbulence, we focused only on data measured above a flying level of 15,000 feet.

Turbulence is a rare phenomenon, and commercial pilots do their best to avoid it. Thus, our EDR data had only a 1% rate of light-to-moderate or greater (LMOG) turbulence. Rare events can complicate learning (Provost and Fawcett 2001). We under-sampled the null and light turbulence data to increase the proportion of the rare LMOG reports. The final class distribution used for our study was 1514 negative examples

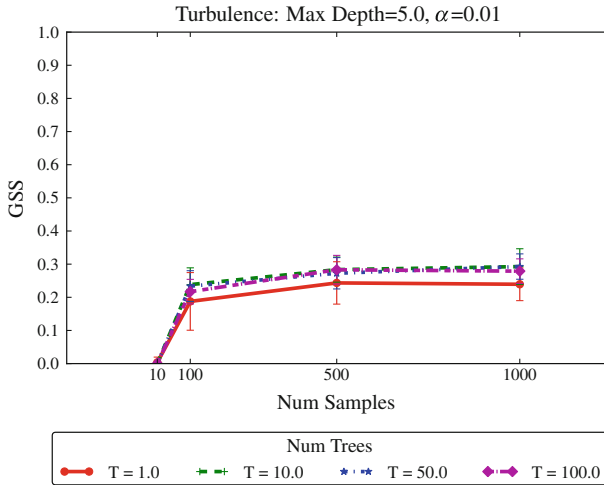


Fig. 8 Average GSS over 30 runs for the turbulence test data as a function of the number of samples at each level of tree growth and the number of trees in the forest. The maximum depth is fixed at 5 for this graph. The *error bars* show one standard deviation

(74%) and 541 positive examples (26%). We split these data into a training and a test set, using cross-validation.

We again performed a full parameter search using the SRRFs and measured performance using GSS. Figure 8 shows the average GSS as a function of the number of distinctions sampled and the number of trees in the forest. Although the skill level is lower than with the Points domain discussed above, the forests are still showing skill at this challenging task. Because this was a two-class problem, we also examined the AUC of the forest as a function of the same parameters. These results, shown in Fig. 9, indicate that the SRRFs are a very robust classifier. Both skill score analyses demonstrate that the forests with 50 and 100 trees are more robust and accurate than the smaller forests.

We compared the effects of varying the main SRRF parameters using ANOVA and these results are shown in Table 4. ANOVA indicates that each individual parameter is statistically significant for GSS. It also shows strong interaction effects between parameters. However, examining the graphs for specific sets of parameters indicates that not all parameter sets may be different. For example, it visually appears that there is no difference between 10, 50, and 100 trees when 1000 distinctions are sampled. To analyze this statistically, we performed an unpaired bootstrap randomization test with 10,000 resamplings to compare the mean GSS. These results are shown in Table 5 and they confirm that there is no statistically significant difference for this set of parameters. The test also confirms an asymptote when increasing the number of distinctions sampled. Going from 10 to 100 samples on 100 trees and depth 5 is significant, but 100 to 500 and 500 to 1000 are not significant.

We also reduced the fields into temporal-scalars as described in the previous section and again trained the SRRFs across all parameter variations. Figure 10 shows the GSS for the turbulence data without fields, and Fig. 11 shows the corresponding

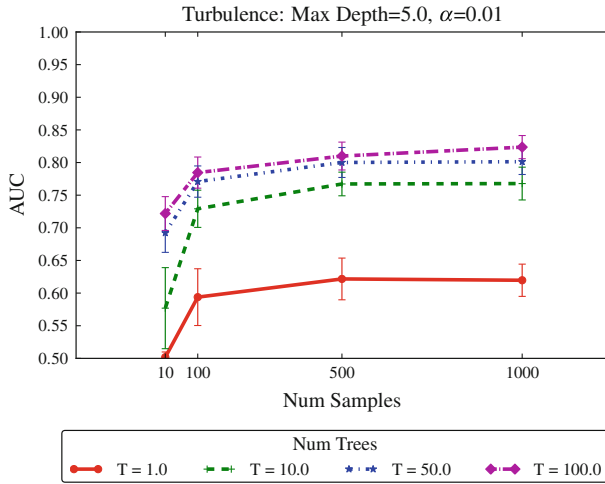


Fig. 9 AUC for the testing set for turbulence across S and T with D = 5. Error bars are one standard deviation

Table 4 N-way ANOVA on turbulence GSS

GSS	S	T × D	T	D	S × D	T × S	T × S × D
p value	0.0118	≈0	≈0	8.71e−6	9.13e−5	≈0	2.52e−8

Values in *bold* are significant with $\alpha = 5\%$

Table 5 Top: unpaired bootstrap randomization on AUC and GSS for 1000 samples (S) with a maximum depth of 5 (D) between varying numbers of tree sizes (T) and bottom: unpaired bootstrap randomization for a forest with 100 trees (T) with maximum depths 5 (D) between varying numbers of samples (S)

Number of trees	1 and 10	10 and 50	50 and 100
AUC	≈0	≈0	≈0
GSS	5e−04	0.996	0.164
Number of samples	10 and 100	100 and 500	500 and 1000
AUC	≈0	≈0	0.011
GSS	≈0	≈0	0.687

Values in *bold* are significant with $\alpha = 5\%$

AUCs. The resulting GSS appears identical to the fielded data and an unpaired bootstrap randomization test with 10,000 resamplings confirms that there is no statistically significant difference between the two with a p value of 0.943. Comparing AUC using the same test gives the same result with no significant difference with a p value of 0.069. Although the fields did not improve performance in this case, they also did not harm the performance.

In order to better understand the results and the effect of including fielded data for this problem, we examined the importance of the variables in both the fielded and

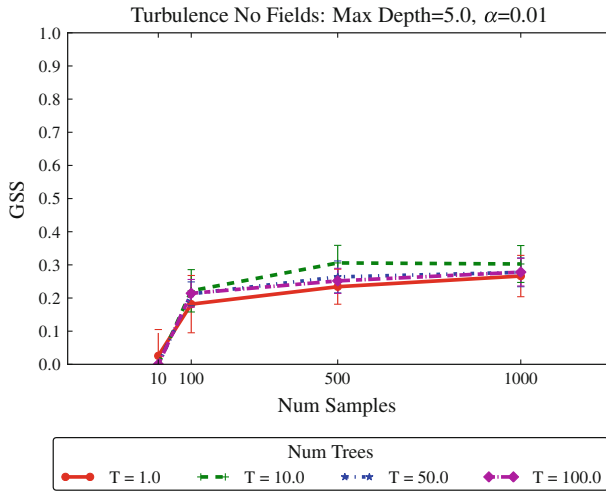


Fig. 10 GSS for turbulence without fields varying across S and T with D = 5. Error bars are one standard deviation

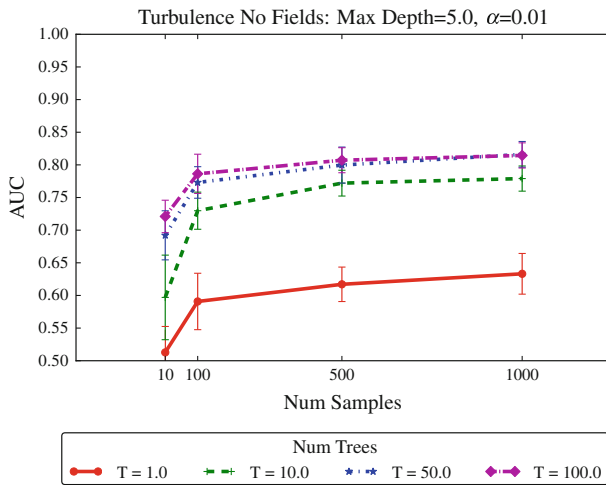


Fig. 11 AUC for turbulence without fields varying across S and T with D = 5. Error bars are one standard deviation

non-fielded versions of the data. Table 6 shows the importance of the top 10 most statistically significant variables in both the fielded and non-fielded data. Only one non-field attribute, the NWP model-derived Richardson number at the plane’s location, shows up as being significantly important for both fields and non-fields. This makes sense as the Richardson number captures the relationship between static stability of atmospheric layers and wind shear between them that can result in turbulence (e.g., Dutton and Panofsky 1970). A stable atmosphere suppresses turbulence and vertical wind shear (the change in wind speed with height) can break up stable layers. This leads to breaking waves, causing vertical mixing and turbulence. As shear increases or

Table 6 Top 10 statistically significant ($\alpha = 0.05$) attributes according to variable importance on the Turbulence dataset including fielded objects (left) and with fielded objects omitted (right)

Fields	Mean variable importance	Non-fields	Mean variable importance
VIL.Area	0.198	Plane.ConvectiveInhibition	0.153
Plane.RichardsonNumber	0.106	Plane.RichardsonNumber	0.128
Rain→Contains.Coverage→VIL	0.085	Conv.StdDev_CloudTopTemperature	0.124
Rain→Contains.Coverage→Conv	0.084	Conv.Area	0.110
Conv→Contains.Coverage→VIL	0.061	Conv.Max_CloudTopTemperature	0.105
Rain.Area	0.056	Plane→Nearby.Azimuth→Conv	0.101
Hail.CloudTopTemperature	0.054	Conv.Mean_Reflectivity	0.090
Plane→Nearby.Range→Rain	0.053	VIL.StdDev_Reflectivity	0.081
VIL.CloudTopTemperature	0.052	Rain.StdDev_Reflectivity	0.080
Plane→Nearby.Range→VIL	0.048	Conv.StdDev_Reflectivity	0.071

For non-field attributes, the name of the field is prefixed with the statistic used in the reduction from a field to a scalar, e.g., VIL.StdDev_Reflectivity is the standard deviation of the VIL's Reflectivity field

stability decreases, the Richardson number falls and the likelihood of turbulence rises. Thus, Richardson number diagnoses the potential for turbulence in both convective and non-convective environments, though it is limited by the resolution and accuracy of the NWP model forecast.

Other important variables for the fielded data include VIL and Rain areas, which are related to the size of the storm and may be correlated with both its intensity and ability to influence the environment. VIL area, in particular, is a good representation of the portion of the storm with strong enough updrafts to loft significant amounts of water into the atmosphere; note that it is most important by an impressive margin. The third through fifth variables serve a somewhat similar purpose, indicating the extent of the storm or the fraction of it that is severe. The CloudTopTemperature field within Hail and VIL objects provides a representation of the height of the tallest clouds. In general, storms with higher clouds are more severe; moreover, the texture of the clouds may offer information about the storm's development. For instance, isolated cold tops may indicate updrafts that are vigorous enough to overshoot into the lower stratosphere (Bedka et al. 2010), indicating active convection. Finally, the eighth and tenth variables represent the proximity of the aircraft to the rain and VIL objects, which is a strong factor in being influenced by them.

For the non-fielded data, the presence of the NWP model-derived convective inhibition as the most important variable may be due to its ability to distinguish regions susceptible to convection from those with a capping layer that prevents convective growth. The third and fifth variables are the standard deviation and maximum values of cloud top temperature within convection objects, which could include information about both the severity of the storms and the variability of the clouds top heights, possibly including the presence of overshooting tops (a feature in a cloud that indicates a strong updraft and thus a strong storm). The area of convection objects is the fourth

most important variable, representing the storm size as the VIL and rain areas did for the fielded objects. The azimuth angle from the plane to convection objects comes next. Location relative to large convective storms has been shown to be a significant factor in the occurrence of turbulence, as the anvil outflow interacts with ambient winds to locally enhance or reduce vertical shear (Trier and Sharman 2009). The final four variables in the list include the mean composite reflectivity within convection objects and its standard deviation within VIL, rain and convection objects. Mean composite reflectivity is likely indicative of the convective storm's strength, though it doesn't distinguish between widespread moderate convection and much more intense, localized pockets. We hypothesize that the variability of reflectivity within the objects, captured by the standard deviation, may assist the SRRF in making this distinction, helping to identify storms with vigorous updrafts that may be particularly conducive to the production of turbulence. If this is true, it may serve much the same role as the fielded cloud top temperature did in the fielded data.

We also examined the types of objects and attributes chosen by the trees in the forests. Table 7 shows the top 10 distinctions that appear for the 100 tree, 1000 sample forest across all 30 runs. This table represents a summary of the best among the randomly-sampled candidate distinctions as the constituent SRPTs were being trained; thus, these distinctions are significant but may not be the highest contributors to the trees' classification performance, unlike the variable importances presented above. Variables for which simple distinctions are highly predictive may not be selected as often, because once selected in a tree they might not be needed again. On the other hand, this table does complement the variable importances by supplying valuable information about what particular distinctions were made for various attributes.

According to Table 7, the CloudTopTemperature fields within Convection, Rain and VIL objects are selected frequently, with its gradient being selected most often and summary statistics somewhat less. This indicates that not only the coldest cloud top within the storm but also the cloud top texture has predictive value, confirming the interpretation of the importance results discussed previously. The gradient of the Reflectivity field within these objects, which gives information about the variation in precipitation intensity, is also chosen frequently; among the Reflectivity field statistics, the mode and count conjugate appear most often, suggesting that typical storm reflectivity values and the number of certain features within the graph are significant. The aircraft-relative heights of the echo top within Rain and VIL objects come next, along with the aircraft altitude and its collocated temperature, which contains similar information. These did not show up in the importance results, but their relationship to turbulence has been established in the literature, e.g., Wolff and Sharman (2008).

3.3 Understanding strong low-altitude rotations (SLAR) in supercell thunderstorms

Tornadoes are one of the most violent forms of hazardous weather, causing millions of dollars of damage and loss of life and property every year. In 2011, tornadoes caused 550 deaths in the United States (Storm Prediction Center 2012). While improvement in the radar capabilities and in prediction techniques have dramatically improved the ability to predict tornadoes, the false alarm rate remains unacceptably high (Markowski

Table 7 Top 10 distinctions and attributes/objects/relations for turbulence with 100 trees, 1000 samples, depth 5 set of trees

All	% Total	Total	Conv. CloudTop Temp	Rain. Reflectivity	VIL. Reflectivity	Rain. CloudTop Temp	VIL. CloudTop Temp	Conv. Reflectivity	Rain. AcRel Height	Plane. Altitude	VIL. AcRel Height	Plane. Temp
% Total	100		7	6	6	6	6	5	5	4	4	4
Total		110746	7851	7043	6629	6333	6316	5998	5642	4799	4625	3974
Attribute:LessThan	11	12659.0	—	—	—	—	—	—	—	3356	—	2530
Field.Gradient	9	10367.0	1645	2588	1566	1460	1175	1610	—	—	—	—
Attribute.Min	7	8071.0	613	175	138	402	337	142	1270	—	1081	—
Attribute:Median	7	7889.0	418	168	212	411	333	176	1129	—	879	—
Attribute:Mean	7	7837.0	534	139	173	395	313	172	1186	—	826	—
Attribute:Mode	7	7685.0	342	341	423	314	344	397	187	—	216	—
Count	6	6825.0	410	326	465	283	381	375	370	242	292	206
Attribute:Max	6	6510.0	464	104	118	210	280	135	696	—	542	—
Attribute:Exact	6	6133.0	—	—	—	—	—	—	—	676	—	888
FieldAttribute.Gradient	5	5726.0	1143	889	864	740	878	843	—	—	—	—

Table 8 Fundamental meteorological variables saved each time step in the ARPS simulations

Cloud Water Mixing Ratio
Cloud Water Number Concentration
Graupel Mixing Ratio
Graupel Number Concentration
Hail Mixing Ratio
Hail Number Concentration
Ice Mixing Ratio
Ice Number Concentration
Pressure
Pressure Perturbation
Potential Temperature
Potential Temperature Perturbation
Rain Mixing Ratio
Rain Number Concentration
Snow Mixing Ratio
Snow Number Concentration
Water Vapor Mixing Ratio
Water Vapor Number Concentration
3D Wind Field (u, v, w)

and Richardson 2009). In order to improve the prediction of tornadoes, we need to dramatically improve our understanding of how they form.

In this work, we study hundreds of simulated supercell thunderstorms—the most severe type of thunderstorm that produces the most devastating tornadoes. We use 263 simulations of such storms that were generated by Rosendahl (2008) using the advanced regional prediction system (ARPS) model (Xue et al. 2000, 2001, 2003). This is one of the best state-of-the-art systems for numerically modeling mesoscale storms. We have previously studied this dataset using temporal approaches to data mining (Rosendahl 2008; McGovern et al. 2011b) and identified patterns that were consistent with current meteorological theories. We have also studied these data using the original SRPT implementation, which had no fielded objects and very few spatial distinctions (McGovern et al. 2008). While that study was illustrative, we felt that examining the full field of variables within the objects would provide more information.

Rosendahl's model used a 500 m horizontal grid spacing over a 100 km × 100 km domain with 50 vertical levels resulting in a domain of 200 × 200 × 50 points. The initial conditions were varied across a set of parameter values consistent with supercell storms. Each simulation was run for three hours with snapshots of the domain's state saved every 30 s. The saved data consists of time series of 200 × 200 × 50 grids of meteorological variables, which are listed in Table 8.

We tracked individual storms within each simulation and labeled the storms as to whether they produced SLAR. Because the grid spacing is too coarse to resolve tornadoes, we study strong rotations at the ground level. We labeled each storm as either

Table 9 Class distribution with and without under-sampling for the SLAR domain

	Total	Negative (%)	Intermediate (%)	Positive (%)
Original	1057	846 (80)	124 (12)	87 (8)
Under-sampled	575	364 (63)	124 (22)	87 (15)

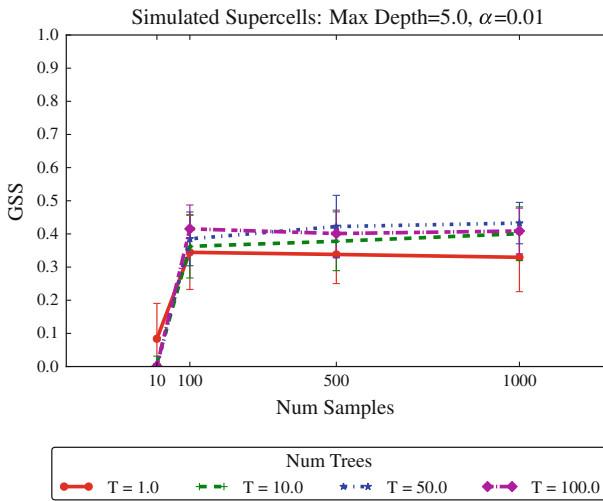


Fig. 12 GSS for the SLAR varying across S and T with D = 5. Error bars are one standard deviation

a positive, an intermediate, or a negative storm. Negative storms produced no strong rotations, intermediate storms produced a possibly strong rotation, and positive storms produced strong low altitude rotations. A given simulation could produce positive as well as negative storms. Tornadoic storms are rare in real life and positive storms were also rare in this dataset, producing the very uneven distribution shown in Table 9. To produce a more balanced distribution, we removed all simulations that produced only negative storms. From a meteorological point of view, a simulation that only produced negative storms is uninteresting; hence, we used this approach instead of randomly under-sampling. This reduced the number of negative storms from 846 to 364, yielding the class distribution in Table 9.

Figure 1 (in Sect. 1) shows the schema we used to create the relational data from the simulations. Each object was created using thresholds on the fielded variables, and multiple fields were saved for each object. Each object can be in one of four spatial relationships with any other object. All of the objects and relations are dynamic.

Figure 12 shows the average GSS for all parameters as a function of the number of distinctions sampled at each level of tree growth and as a function of the number of trees in the forest. The best GSS is ≈ 0.4 for 100 trees and 100 samples at depth 5. The ANOVA calculations in Table 10 show that each parameter individually has a statistically significant effect on GSS. Table 11 shows that the appearance of 50 trees out performing 100 trees is not statistically significant. In fact 10, 50, and 100 trees are

Table 10 N-way Anova on GSS

GSS	S	T × D	T	D	S × D	T × S	T × S × D
<i>p</i> value	1.70e-7	≈0	≈0	≈0	0.772	≈0	0.786

Values in *bold* are significant with $\alpha = 5\%$

Table 11 Unpaired bootstrap randomization on GSS for S = 1000 with D = 5 between various T values; and for T = 100 with D = 5 between various S values

Number of trees	1 and 10	10 and 50	50 and 100
<i>p</i> value	0.006	0.095	0.172
Number of samples	10 and 100	100 and 500	500 and 1000
<i>p</i> value	≈0	0.443	0.675

Values in *bold* are significant with $\alpha = 5\%$

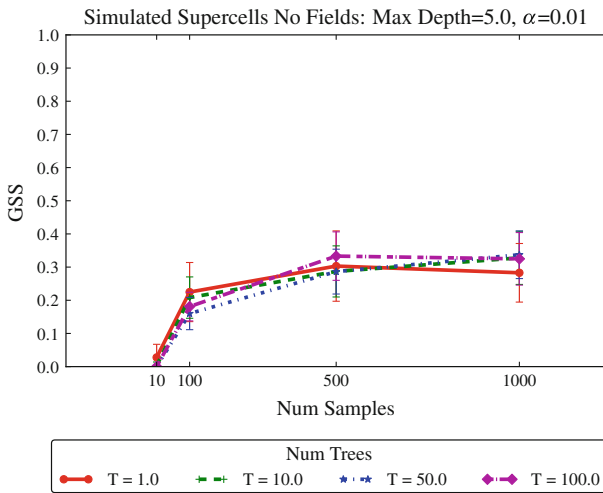


Fig. 13 GSS for the SLAR without fields varying across S and T with D = 5. Error bars are one standard deviation

statistically the same based on an unpaired bootstrap randomization test with 10,000 resamplings to compare the mean GSS. However, they all are statistically higher than 1 tree.

We also performed the full parameter search on this dataset without the fielded objects, and the results are shown in Fig. 13. There is a noticeable drop in GSS between the fielded and non-fielded versions of the data. Taking all the data together and running ANOVA indicates that this difference is statistically significant with a *p* value of ≈0.³ We also compared non-fields to fields on a single set of parameters with 100 trees, 1000 sampled distinctions and max depth 5. An unpaired bootstrap

³ Floating-point precision cannot represent values $\leq 2.2e-16$.

Table 12 Top 10 statistically significant attributes according to variable importance on the fielded simulated storms data

Fields	Mean variable importance
PressurePerturbation.PressurePerturbationField	0.251
PressurePerturbation → Overlap.PercentOverlap → Rain	0.188
Hail.HorizontalCompositeArea	0.114
PressurePerturbation.Volume	0.107
PressurePerturbation.Thickness	0.101
PressurePerturbation.BaseHeight	0.099
Hail.Volume	0.089
Updraft.Volume	0.086
Updraft.Thickness	0.061
Updraft.Buoyancy	0.060

Attributes in common between this and Table 13 are shown in *bold*

Table 13 Top 10 statistically significant attributes according to variable Importance on the non-fielded simulated storms data

Fields	Mean variable importance
PressurePerturbation.Thickness	0.264
PressurePerturbation.Volume	0.179
Hail.HorizontalCompositeArea	0.177
PressurePerturbation → Overlap.PercentOverlap → Rain	0.156
Hail.Volume	0.134
PressurePerturbation→Nearby.MinDistance →PressurePerturbation	0.105
Updraft.Volume	0.067
PressurePerturbation.PercentForward	0.065
Hail.BaseHeight	0.060
Hail→Nearby.MinDistance →PressurePerturbation	0.060

Attributes in common between this and Table 12 are shown in *bold*

randomization test with 10,000 resamplings still shows a strong significance with a p value of $4.05e-5$.

Given the strong performance in this domain, it makes sense to analyze the SRRF results meteorologically. We first examine the statistically important variables using the variable importance analysis. Table 12 lists the top 10 statistically significant attributes for the fielded data and Table 13 lists the same for the unfielded data. Attributes in common between the two tables are shown in bold. Because rotation is associated with pressure drops, the inclusion of multiple attributes of the pressure perturbation object seems quite reasonable. Also, updrafts are the thermodynamic engines that power supercell thunderstorms and they play a major role in the concentration of rotation into tornado-scale vortices. Hail is an ever-present feature of supercell storms, whether or not low-altitude rotation is present. As such, it is not involved in the production of

low-altitude rotation and it should be noted that the fielded data in Table 12 properly minimizes the role of hail in the development of low-altitude rotation.

Table 14 (continued in Table 15) lists the distinction counts of the top 10 distinctions and top 10 attributes chosen by all the trees in the 100 tree forest with 1000 samples per node and with trees of maximum depth 5. The counts of the chosen attributes are consistent with the results of the variable importance analysis. Again, perturbation pressure is prominent among the distinctions. The tilting of horizontal vorticity (localized rotation about a horizontal axis induced by increasing horizontal winds with height above the earth's surface) into the vertical by an updraft ("positive tilting") is thought to produce updraft rotation. The subsequent development of a localized downdraft adjacent to the updraft then is thought to play an important role in the development of the tornado-scale vortex within a supercell storm. We expect that our current storm simulations using much finer horizontal resolution will start to clarify the processes associated with these updraft-downdraft interactions.

Figure 14 shows the best SRRF with one tree, i.e., an SRPT. In this example we can see several uses of the top attributes (by variable importance and distinction count). Note that the probabilities of the formation of SLAR occur only along the left portion of the tree and all of probabilities of no low-altitude rotation occur along the right portion of the tree. The category "maybe" is found throughout the tree.

4 Conclusions

We have presented and empirically evaluated several major enhancements to the SRPT algorithm and to SRRFs based on them. The addition of both 2D and 3D fields to objects and relations, either scalar or vector, greatly increases the information available to the SRPTs. To make use of this new information, we added two types of distinctions increasing the number and type of questions that SRPTs can ask of the data. First, we added shape detection in both 2D and 3D. Second, we provided the means to reason about the actual value of data in the fields by providing distinctions that deal with the gradient, curl, and divergence of fields. Our hypothesis was that the addition of fielded objects would noticeably increase the performance of the SRPT algorithm.

We demonstrated that these additions significantly enhanced the performance of the SRPT and SRRFs on one synthetic domain and one real-world domain. In addition to demonstrating the enhanced performance, we validated the results found by the SRRFs on the two real-world domains. These results are consistent with current meteorological theories.

Not all relational domains will benefit from the addition of spatiotemporal fields. In particular, this approach is best suited for data that evolves over space and time, such as the severe weather data presented here. This might also include domains such as Robocup.⁴ Other relational domains such as fraud (e.g., Neville et al. 2005), social networks (e.g., Eldardiry and Neville 2011), or large network data such as the music networks in KDD Cup 2011, are unlikely to benefit from this type of analysis as they do not contain distinct volumes of objects that rapidly change as a function of time or space.

⁴ <http://www.robocup.org/>.

Table 14 Top 1–5 distinctions and top 10 attributes/objects/relations for SLAR with 100 trees, 1000 sampled distinctions, and depth 5 trees (Continued in Table 15)

All	% Total	Total	Exist	Attribute.StdDev	Attribute.Max	Attribute.Mean	Attribute.Median
% Total	100		25	12	11	9	8
Total		25, 204	6339	3127	2875	2383	2027
PressurePerturbation → Overlap.PercentOverlap → Rain	4	938	–	189	215	209	192
PressurePerturbation.PressurePerturbation	3	814	–	340	31	3	5
Hail	2	630	516	–	–	–	–
PressurePerturbation → Contains.PercentOverlap → Downdraft	2	573	–	108	100	99	89
PosTilting	2	512	510	–	–	–	–
Hail → Overlap → PressurePerturbation	2	482	476	–	–	–	–
Hail Mixing Ratio	2	474	–	16	14	10	10
PressurePerturbation	2	465	33	–	–	–	–
PressurePerturbation → Overlap.PercentOverlap → Updraft	2	403	–	85	87	90	60
PosTilting.Tilting	2	382	–	10	20	11	16

Table 15 Top 6–10 distinctions and top 10 attributes/objects/relations for SLAR with 100 trees, 1000 samples, depth 5 set of trees (Continued from Table 14)

All	% Total	Total	Attribute..Min	Attribute..Mode	FieldAttribute.Gradient	TemporalExists	Field.Gradient
% Total	100		8	7	5	4	3
Total		25, 204	1938	1862	1264	922	652
PressurePerturbation → Overlap.PercentOverlap → Rain	4	938	111	15	—	—	—
PressurePerturbation, PressurePerturbation	3	814	11	17	312	—	33
Hail	2	630	—	—	—	112	—
PressurePerturbation → Contains.PercentOverlap → Downdraft	2	573	101	71	—	—	—
PosTilting	2	512	—	—	—	2	—
Hail → Overlap → PressurePerturbation	2	482	—	—	—	—	—
Hail Mixing Ratio	2	474	22	11	140	—	71
PressurePerturbation	2	465	—	—	—	429	—
PressurePerturbation → Overlap.PercentOverlap → Updraft	2	403	56	24	—	—	—
PosTilting.Tilting	2	382	21	8	109	—	47

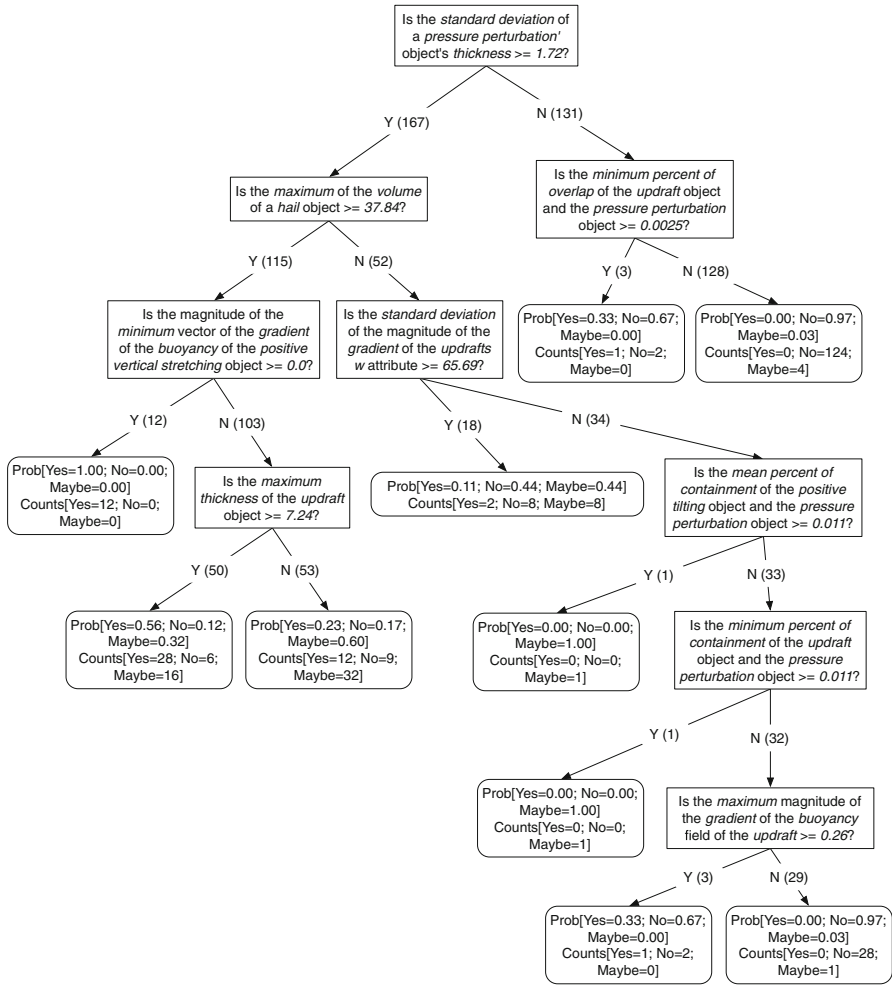


Fig. 14 Highest scoring (by GSS) single tree in the simulated storms domain

In fields with spatiotemporal data, such as the two weather domains shown here, the addition of the fields was able to significantly improve the results for the simulated storms while not changing the results of the turbulence prediction. This is most likely because the fields for the simulated storm domain were generated based on current theories about the generation of storms. In the turbulence domain, the most relevant data for fields is not available. For example, the strength of the updraft at the tropopause is critical to knowing how strong the storm is and how far its effects will spread in space. However, that data is not available because current numerical weather prediction models cannot resolve the updraft at that level and that resolution across the entire CONUS domain.

In current work on the supercell domain, we are generating simulations with a 75 m horizontal resolution. This is sufficient to resolve most tornadic vortices and

we believe that mining these simulations will produce significant new results in our understanding of the formation of tornadoes.

Autocorrelation was mentioned briefly in the discussion of missing values. While autocorrelation is often considered a nuisance because of its effects on traditional statistics (Jensen and Neville 2002), it is an accepted fact in geographical (spatial) studies and can even be exploited to advantage (O'Sullivan and Unwin 2002). Thus, under certain circumstances, the autocorrelation becomes part of the structure being studied and should be left alone or even used (as we did for the missing values). For example, O'Sullivan and Unwin (2002) states, "Geographical data are often not samples in the sense meant in standard statistics. Frequently, geographical data represent the whole population. Often, we are only interested in understanding the study region, not in making wider inferences about the whole world, so that the data *are* the entire population of interest."

Research Reproducibility Further details of the implementation, source code, and experimental results can be found at: <http://idea.cs.ou.edu/theses/ntroutman>.

Acknowledgements The authors thank Jason Craig, David J. Gagne II, Nathan Hiers, Gregory Meymaris, Timothy Supinic, and Derek Rosendahl for their work in generating some of the data used in this research. This research was supported by the National Science Foundation under Grant No. NSF/IIS/0746816 and by NASA under Grant No. NNX08AL89G. Much of the computing for this project was performed at the OU Supercomputing Center for Education & Research (OSKER) at the University of Oklahoma (OU).

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Allcroft DJ, Glasbey C, Durban M (2001) Modelling weather data. In: SCRI annual report 2001, pp 192–195
- Allen JF (1991) Time and time again: the many ways to represent time. *Int J Intell Syst* 6(4):341–355
- Barber C, Dobkin D, Huhdanpaa H (1996) The quickhull algorithm for convex hulls. *ACM Tran Math Softw* 22(4):469–483. <http://www.qhull.org>. Accessed 4 March 2012
- Bedka K, Brunner J, Dworak R, Fletz W, Otkin J, Greenwald T (2010) Objective satellite-based detection of overshooting tops using infrared window channel brightness temperature gradients. *J Appl Meteorol Climatol* 49:181–202
- Bluestein H, Weiss C, French M, Holthaus E, Tanamachi R, Frasier S, Pazmany A (2007) The structure of tornadoes near Attica, Kansas, on 12 May 2004: high-resolution, mobile, Doppler radar observations. *Mon Weather Rev* 135(2):475–506
- Bodenhamer M, Bleckley S, Fennelly D, Fagg AH, McGovern A (2009) Spatio-temporal multi-dimensional relational framework trees. In: Proceedings of the 2009 IEEE international conference on data mining (ICDM) workshop on spatiotemporal data mining, Miami, electronically published
- Bradley AP (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit* 30(7):1145–1159
- Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
- Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees. Wadsworth, Belmont
- Cornman L, Morse C, Cuning G (1995) Real-time estimation of atmospheric turbulence severity from in situ aircraft measurements. *J Aircr* 32:171–177
- Cornman L, Meymaris G, Limber M (2004) An update on the FAA aviation weather research program's in situ turbulence measurement and reporting system. In: Preprints, 11th conf on aviation, range and aerospace meteorology, Hyannis, p P4.3
- Cova T, Goodchild M (2002) Extending geographical representation to include fields of spatial objects. *Int J Geogr Inform Sci* 16(6):509–532

- Davies-Jones R (2008) Can a descending rain curtain in a supercell instigate tornadogenesis barotropically?. *J Atmospheric Sci* 65:2469–2497
- Dutton J, Panofsky HA (1970) Clear air turbulence: a mystery may be unfolding. *Science* 167:937–944
- Egan JP (1984) Signal detection theory and ROC analysis. Series in cognition and perception. Academic Press, New York
- Eldardiry H, Neville J (2011) Across-model collective ensemble classification. In: Proceedings of the 25th conference on artificial intelligence AAAI 2011, San Francisco, electronically published
- Fast A, Friedland L, Maier M, Taylor B, Jensen D, Goldberg H, Komoroske K (2007) Relational data pre-processing techniques for improved securities fraud detection. In: Proceedings of the 13th international conference on knowledge discovery and data mining, San Jose, pp 941–949
- Fern A, Getoor L, Milch B (2006) SRL2006: open problems in statistical relational learning. <http://www.cs.umd.edu/projects/srl2006/>. Accessed 4 March 2012
- Friedman HF, Kohavi R, Yun Y (1996) Lazy decision trees. In: Proceedings of the 13th national conference on artificial intelligence, Portland, pp 717–724
- Friedman N, Getoor L, Koller D, Pfeffer A (1999) Learning probabilistic relational models. In: Proceedings of the international joint conference on artificial intelligence, Stockholm, pp 1300–1309
- Gagne II DJ, Supinie T, McGovern A, Basara J, Brown RA (2010) Analyzing the effects of low level boundaries on tornadogenesis through spatiotemporal relational data mining. In: Presented at the 8th conference on artificial intelligence applications to environmental science, Atlanta, electronically published
- Gandin LS, Murphy AH (1992) Equitable skill scores for categorical forecasts. *Mon Weather Rev* 120(2):361–370
- Gerrity JP (1992) A note on Gandin and Murphy's equitable skill score. *Mon Weather Rev* 120(11):2709–2712
- Getoor L, Friedman N, Koller D, Taskar B (2001) Learning probabilistic models of relational structure. In: Proceedings of the eighteenth international conference on machine learning, Montreal, pp 170–177
- Getoor L, Friedman N, Koller D, Taskar B (2002) Learning probabilistic models of link structure. *J Mach Learn Res* 3:679–707
- Glasbey C, Allcroft DJ (2007) Spati-temporal weather models. Talk at 39th French statistical association (SFDS) congress, Angers. <http://www.bioss.sari.ac.uk/staff/chris/angers07.pdf>. Accessed 4 March 2012
- Goodchild M, Yuan M, Cova T (2007) Towards a general theory of geographic representation in GIS. *Int J Geogr Inform Sci* 21(3):239–260
- Jensen DD, Cohen PR (2000) Multiple comparisons in induction algorithms. *Mach Learn* 38(3):309–338
- Jensen D, Getoor L (2003) IJCAI 2003 workshop on learning statistical models from relational data. <http://kdl.cs.umass.edu/srl2003/>. Accessed 4 March 2012
- Jensen D, Neville J (2002) Linkage and autocorrelation cause feature selection bias in relational learning. In: Proceedings of the international conference on machine learning, Sydney, pp 259–266
- Jolliffe IT, Stephenson DB (2003) Forecast verification: a practitioner's guide in atmospheric science. Wiley, New York
- Klemp JB, Rotunno R (1983) A study of the tornadic region within a supercell thunderstorm. *J Atmospheric Sci* 40(2):359–377
- Kononenko I, Bratko I, Roskar E (1984) Experiments in automatic learning of medical diagnostic rules. Tech. rep., Jozef Stefan Institute, Ljubljana, Yugoslavia
- Liu WZ, White AP, Thompson SG, Bramer MA (1997) Techniques for dealing with missing values in classification. *Advances in intelligent data analysis*. Springer, Berlin, pp 527–536
- Longley PA, Goodchild M, Maguire DJ, Rhind DW (2005) Geographic information systems and science. Wiley, New York
- Markowski P, Richardson Y (2009) Tornadogenesis: our current understanding, forecasting considerations, and questions to guide future research. *Atmospheric Res* 93:3–10
- Markowski PM, Straka JM, Rasmussen EN (2003) Tornadogenesis resulting from the transport of circulation by a downdraft: Idealized numerical simulation. *J Atmospheric Sci* 60(6):295–823
- Marzban C (1998) Scalar measures of performance in rare-event situations. *Weather Forecast* 13(3):753–763
- McGovern A, Hiers N, Collier M, Gagne II DJ, Brown RA (2008) Spatiotemporal relational probability trees. In: Proceedings of the 2008 IEEE international conference on data mining, Pisa, pp 935–940

- McGovern A, Supinie T, Gagne DJ II, Troutman N, Collier M, Brown RA, Basara J, Williams J (2010) Understanding severe weather processes through spatiotemporal relational random forests. In: Proceedings of the 2010 NASA conference on intelligent data understanding, Mountain View, pp 213–227
- McGovern A, Gagne DJII, Troutman N, Brown RA, Basara J, Williams J (2011a) Using spatiotemporal relational random forests to improve our understanding of severe weather processes. *Stat Anal Data Min* 4(4):407–429
- McGovern A, Rosendahl DH, Brown RA, Droegemeier KK (2011b) Identifying predictive multi-dimensional time series motifs: an application to understanding severe weather. *Data Min Knowl Discov* 22(1):232–258
- Miller HJ, Han J (eds) (2009) *Geographic data mining and knowledge discovery*, 2nd edn. Chapman and Hall/CRC Press, Boca Raton
- Neville J, Jensen D (2005) Leveraging relational autocorrelation with latent group models. In: Proceedings of the international conference on data mining, Houston, pp 322–329
- Neville J, Jensen D (2007) Relational dependency networks. *J Mach Learn Res* 8:653–692
- Neville J, Jensen D, Friedland L, Hay M (2003) Learning relational probability trees. In: Proceedings of the 9th ACM SIGKDD international conference on knowledge discovery and data mining, San Diego, pp 625–630
- Neville J, Şimşek Ö, Jensen D, Komoroske J, Palmer K, Goldberg H (2005) Using relational knowledge discovery to prevent securities fraud. In: Proceedings of the 11th ACM SIGKDD international conference on knowledge discovery and data mining, San Diego, pp 449–458
- O'Rourke J (1985) Finding minimal enclosing boxes. *Int J Comput Inform Sci* 14(3):17
- O'Sullivan D, Unwin DJ (2002) *Geographic information analysis*. Wiley, Hoboken
- Provost F, Domingos P (2000) Well-trained PETs: improving probability estimation trees. University of Washington; CDER working paper 00-04-is, Stern School of Business, NYU, electronically published
- Provost F, Fawcett T (2001) Robust classification for imprecise environments. *Mach Learn* 42(3):203–231
- Quinlan JR (1993) C4.5 programs for machine learning. Morgan Kaufmann, Santa Mateo
- Richardson M, Domingos P (2005) Markov logic networks. *Mach Learn* 62(1-2):107–136
- Rosendahl DH (2008) Identifying precursors to strong low-level rotation within numerically simulated supercell thunderstorms: a data mining approach. Master's thesis, School of Meteorology, University of Oklahoma
- Russell S, Norvig P (2009) *Artificial intelligence: a modern approach*, 3rd edn. Prentice Hall, Englewood Cliffs
- Schnabel R, Wahl R, Klein R (2007) Efficient RANSAC for point-cloud shape detection. *Comput Graphics Forum* 26:214–226
- Sharan U, Neville J (2007) Exploiting time-varying relationships in statistical relational models. In: Proceedings of the 1st SNA-KDD workshop, 13th ACM SIGKDD conference on knowledge discovery and data mining, San Jose
- Sharan U, Neville J (2008) Temporal-relational classifiers for prediction in evolving domains. In: Proceedings of the IEEE international conference on data mining, Pisa
- Sharman R, Tebaldi C, Wiener G, Wolff J (2006) An integrated approach to mid- and upper-level turbulence forecasting. *Weather Forecast* 21:268–287
- Snook N, Xue M (2008) Effects of microphysical drop size distribution on tornadogenesis in supercell thunderstorms. *Geophys Res Lett* 35(24):L24,803
- Srinivasan A (1999) A study of two probabilistic methods for searching large spaces with ILP. Tech. rep., PRG-TR-16-00 Oxford University Computing Laboratory, University of Oxford
- Storm Prediction Center (2012) Annual fatal tornado summaries. <http://www.spc.noaa.gov/climo/torn/fataltorn.html>. Accessed 4 March 2012
- Supinie T, McGovern A, Williams JK, Abernethy J (2009) Spatiotemporal relational random forests. In: Proceedings of the 2009 IEEE international conference on data mining workshops, Miami, pp 630–635
- Trapp RJ, Stumpf GJ, Manross KL (2005) A reassessment of the percentage of tornadic mesocyclones. *Weather Forecast* 20(4):680–687
- Trier SB, Sharman RD (2009) Convection-permitting simulations of the environment supporting wide-spread turbulence within the upper-level outflow of a mesoscale convective system. *Mon Weather Rev* 137:1972–1990
- Troutman N (2010) Enhanced spatiotemporal relational probability trees and forests. Master's thesis, School of Computer Science, University of Oklahoma

- Valdes-Sosa PA (2004) Spatio-temporal autoregressive models defined over brain manifolds. *Neuroinformatics* 2(2):239–250
- Weber RO, Talkner P (1993) Some remarks on spatial correlation function models. *Mon Weather Rev* 121(9):2611–2617
- White AP (1987) Probabilistic induction by dynamic path generation in virtual trees. *Research and development in expert systems III*. Cambridge University Press, Cambridge, pp 34–46
- Wicker LJ, Wilhelmson RB (1995) Simulation and analysis of tornado development and decay within a three-dimensional supercell thunderstorm. *J Atmospheric Sci* 52(15):2675–2703
- Williams JK, Sharman R, Craig J, Blackburn G (2008) Remote detection and diagnosis of thunderstorm turbulence. In: *Proceedings of SPIE, remote sensing applications for aviation weather hazard detection and decision support*, vol 7088, San Diego
- Wolff J, Sharman R (2008) Climatology of upper-level turbulence over the continental United States. *J Appl Meteorol Climatol* 47:2198–2214
- Wurman J, Straka JM, Rasmussen EN (1996) Fine-scale doppler radar observations of tornadoes. *Science* 272(5269):1774–1777
- Xue M, Droegemeier KK, Wong V (2000) The advanced regional prediction system (ARPS)—a multiscale nonhydrostatic atmospheric simulation and prediction model. Part I: model dynamics and verification. *Meteorol Atmospheric Phys* 75:161–193
- Xue M, Droegemeier KK, Wong V, Shapiro A, Brewster K, Carr F, Weber D, Liu Y, Wang D (2001) The advanced regional prediction system (ARPS)—a multiscale nonhydrostatic atmospheric simulation and prediction tool. Part II: model physics and applications. *Meteorol Atmospheric Phys* 76:143–165
- Xue M, Wang D, Gao J, Brewster K, Droegemeier KK (2003) The advanced regional prediction system (ARPS), storm-scale numerical weather prediction and data assimilation. *Meteorol Atmospheric Phys* 82:139–170