A Hierarchical Pattern Learning Framework for Forecasting Extreme Weather Events

Dawei Wang, Wei Ding[1]
Department of Computer Science
University of Massachussets Boston, USA
dawei.wang@umb.edu, ding@cs.umb.edu

Abstract—**Extreme weather events, like extreme rainfalls, are severe weather hazards and also the triggers for other natural disasters like floods and tornadoes. Accurate forecasting of such events relies on the understanding of the spatiotemporal evolution processes in climate system. Learning from climate science data has been a challenging task, because the variations among spatial, temporal and multivariate spaces have created a huge amount of features and complex regularities within the data. In this study we developed a framework for learning patterns from the spatiotemporal system and forecasting extreme weather events. In this framework, we learned patterns in a hierarchical manner: in each level, new features were learned from data and used as the input for the next level. Firstly, we summarized the temporal evolution process of individual variables by learning the location-based patterns. Secondly, we developed an optimization algorithm for summarizing the spatial regularities, *SCOT*, by growing spatial clusters from the location-based patterns. Finally, we developed an instance-based algorithm, *SPC*, to forecast the extreme events through classification. We applied this framework to forecasting extreme rainfall events in the eastern Central Andes area. Our experiments show that this method was able to find climatic process patterns similar to those found in domain studies, and our forecasting results outperformed state-of-art models.**

I. INTRODUCTION

Climate dynamics have a wide range of impacts across every region on many sectors of our society. Reliable knowledge about current and future changes in the climate system, in both the short and long term, is of vital importance for the economy and for our society. With the advance of data collecting techniques (model simulation, remote sensing, and *in situ* observations), climate science has become one of the most data-rich fields with regard to data volume [1]. However, comparing with other domains like electronic advertising, the success of big data-induced process in studying climate system is limited [2]. Among all reasons behind the slow progress, for methods trying to model the climate dynamics, the main difficulty lies in addressing several key challenges:

- The extremely large feature spaces. Climate science data typically have a spatiotemporal structure. From the perspective of data science, every variable containing certain spatiotemporal information can be considered as an individual attribute, or a feature of the data set. The number of features in a climate science data set can easily reach tens of thousands, or even millions.

- The complex regularities (or patterns) within the system. The structures and functions in the climate system have a hierarchical modularity [3]: the temporal evolution of variables spreads across regions and falls into spatially local subsystems (spatial clusters), and the whole system is a composition of crosstalks among the subsystems. The data representation for studying such dynamics needs to be able to abstract all the temporal, spatial and among-variable relationships (feature interactions).

- The requirement on model interpretablity, because of the greater interest in the climate science field to understand rather than to simply predict.

Frequent pattern-based data representations have been used in various studies for abstracting climatic phenomena [4], [3]. The descriptive nature of such pattern provides a very intuitive interpretation for the physical process. For example, the atmospheric phenomenon, "the transport of moist air by low-level trade winds from the tropical Atlantic Ocean to the Amazon Basin along the Intertropical Convergence Zone" [5], plays a key role in controlling the rainfalls in the eastern Central Andes (*ECA*) area of South America during monsoon season. This atmospheric phenomenon may be described as a data pattern: {*at time 0, atmospheric pressure is high at location l_1 and low at location l_2, relative humidity is high at location l_1; at time 1, zonal wind is high at location l_3;...*}.

In this paper, we developed a novel hierarchical framework for learning patterns with a level-wise manner from spatiotemporal data (Fig. 1). Specifically,

- First, we summarized the temporal evolutions of individual variables. For every location,each variable's temporal changing process is generalized into one single feature transformed from the learned patterns.

- Second, to summarize the spatial relationships, we developed an optimization algorithm, *SPC* (**S**patial **C**luster **O**ptimization **T**ool), to grow the single location features learned from the first level into spatial clusters.

- Finally, we summarized the inter-variable relationships through developing an instance-based classifier, *SPC* (**S**patial cluster **P**attern-based **C**lassifier),for forecasting the target extreme events.
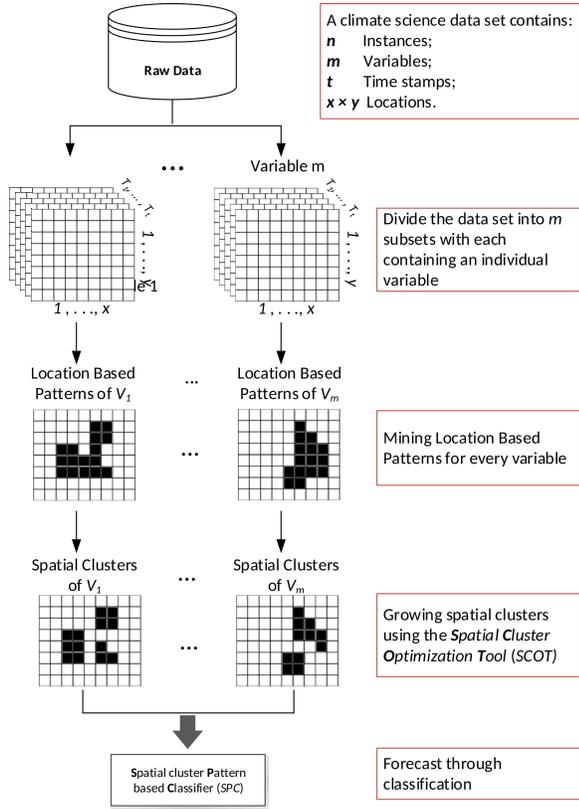
[1]Corresponding Author

Fig. 1. Flow chart of the proposed learning framework, in which patterns are summarized in a hierarchical manner.

By partitioning the climate science data set and learning/generalizing the discovered patterns according to the data set's inherent infrastructure in a level-wise manner, we made it possible to learn patterns with global coverage from the extremely large feature space. The patterns we found were highly interpretable and from empirical studies we showed that our model outperformed the state-of-art method [6] for forecasting extreme rainfall events.

II. RELATED WORK

The climate system is dynamic and its components are very different in their composition, physical and chemical properties, structure and behavior. Different data representations have been implemented for studying the spatiotemporal interactions. In Lozano et al. [7] a graph model was built for causal inference of extreme climate events. The authors used location and scale parameters in their model to incorporate spatial and temporal correlations, and only local neighbors (3*3 grids) were considered for the model. In Chen et al. [8], the same data set was used, while graph models were constructed without assuming parametric underlying distributions, which was important for studying climate data. But the graphs were only built on individual locations. The above two studies investigated the climate data set from spatially sliced pieces instead of an entirety. Ensemble phase detection and Least Absolute Deviation (LAD) regression techniques were combined in Gonzalez et al. [9] for climate system response prediction, and the climate system phases were modeled based on temporal intervals. In Wang et al.

[10] a Bayesian network-based streaming feature selection algorithm was implemented to deal with the high dimensional climate data set.

Frequent pattern based-methods have been a focused theme in data science research [11]. Co-location patterns [12] for investigating the spatial relations of objects, spatial emerging patterns for extracting discriminative information from spatial data [13], and sequential pattern for studying time series data [14]. There are two major challenges that need to be addressed in frequent pattern applications. The first challenge lies in the fact that frequent pattern mining is an NP-complete problem. The computational cost increases exponentially with the number of features contained in a pattern [15].Most applications on climate data either sliced the data set into temporal/spatial profiles [16], or set constrains on the number of features contained in a pattern [3], to alleviate the computational demand. The approach named "Progressive Refinement", which performs rough computation at a smaller (coarser) resolution for candidate identification and refines the results at larger resolutions, was firstly proposed in [17] to tackle the challenge of mining patterns from spatial data sets. The second is that the mining process usually results in a huge number of patterns which hinders a method's interpretability and application. Pattern summarization techniques aiming at reducing the number of resulting patterns by building smaller representative pattern sets have been proposed in Liu et al. [18] and Wang et al. [19].

III. FORMAL DEFINITIONS

Definition 1 *Feature and Feature Set*: A *feature* in a climate science data set D is a tuple of the form: $\{L, T, V\}$, where V is a domain variable, while L and T indicate the location (including the vertical level) and time where and when the variable was sampled, respectively. For example, a variable sampled from time stamps 1 to 4 can be considered as a feature set, which is written as $\{V^1, V^2, V^3, V^4\}$ without specifying the spatial information.

Definition 2 *Pattern and Location Based Pattern*: A pattern is a set of feature-value pairs corresponding to a feature set. $X = \{< V^1, 1 >, < V^2, 0 \ to \ 4 >, < V^3, 1 \ or \ 2 >, < V^1, 3 >\}$ is a pattern of the example feature set in Definition 1. It is a rule of possible feature values in a feature set. Notice that in our definition a feature may have different values ($< V_2, 0 \ to \ 4 >, < V_3, 1 \ or \ 2 >$) in the pattern. The tuple of the form: $\{X, L\}$, which contains a pattern (X) and its spatial information (L), is called a location-based pattern.

Definition 3 *Support and Growth Ratio*: A pattern is said to be supported by an instance I from data set D if the values of features in the instance conform to the rule specified by the pattern. The support of a pattern X is the number of instances supporting X in a data set D divided by the total number of instances in D. If we divide D into two partitions, $\{D_p \ and \ D_n\}$, the growth ratio δ of a pattern X is the ratio of $X's$ support in one partition D_p to its support in the other partition D_n.

Definition 4 *Feature of Pattern*: The feature of a pattern X is a binary variable (0 or 1) indicating whether the

	V^1	V^2	V^3	V^4	CL
I_1	1	0	1	3	P
I_2	1	3	2	3	P
I_3	2	1	2	1	N
I_4	1	4	2	1	P
I_5	1	1	1	3	N

Pattern X: [<V_1,1>, <V_2,0 to 4>,<V_3,1 or 2>,<V_4,3>.]

P-support=2/3=0.67

N-support=1/2=0.5

Growth Ratio = 0.67/0.5=1.34

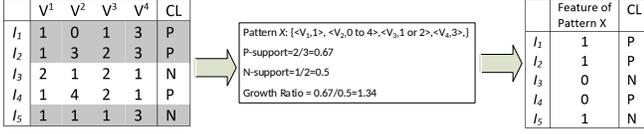|	Feature of Pattern X	CL
I_1	1	P
I_2	1	P
I_3	0	N
I_4	0	P
I_5	1	N

Fig. 2. Examples of pattern, pattern's support and growth ratio, and the pattern's representing feature.

regularity is present or not in an instance. A pattern's feature also has the same support (sum of 1s) and growth ratio (ratio of the sum of 1s on different partitions of the data set) values as the pattern.

Fig.2 gives examples of pattern, pattern's support and growth ratio, and the pattern's representing feature from an example data set with 5 instance and 4 features. The growth ratios were calculated based on the partitions using class labels (CL).

IV. METHOD

Our framework was built based on the hierarchical modularity of the climate system: the temporal evolution of individual variables in a location (level 1, location-based patterns) spreads across the spatial space (level 2, spatial clusters) and consists of spatiotemporal subsystems; these subsystems interact with each other and fall into a global system which determines the weather conditions (level 3, multiple spatial cluster patterns).

A. Learning Location-Based Patterns

We firstly developed a contrast pattern-mining algorithm (Algorithm 1) for learning the location-based patterns. In detail, by partitioning the climate science data set into m subsets, each of which contains one individual variable (line 3), we learned the location-based patterns from the subsets. Specifically, for each subset, we learned the sets of frequent contrast patterns on every location separately (lines 4-8). Here the extents of "frequent" and "contrast" were defined through two thresholds for patterns' support (ρ) and growth ratio (δ), relatively. For every location, we generalized the learned set of patterns into one single representative pattern and transformed it into a new feature (line 9).

In our framework, the location-based patterns learned from climate science data were temporal processes of climate variables. For example, for the same location the set of patterns $\{p_1, p_2, ...\}$ of a variable V was the set of indicative temporal changes of V that happened much more often in one partition (according to CL) of the data. The generalized pattern should include all such processes of V for distinguishing purpose. Thus here we summarized the patterns learned from one location using alternation: for an instance the generalized binary pattern was 1 if any of the location patterns existed, and 0 otherwise.

B. Growing Spatial Clusters

In the second step we summarized spatial regularities in the system by growing the single location-based patterns learned from step 1 into spatial clusters.

Algorithm 1: *Learning Location-Based Patterns*

Data:
ρ : the support threshold
δ : the growth ratio threshold
D : the climate science data set with:
m variables; s sampling locations;
t sampling time stamps; class labels CL.
Result:
LPF : the set of features from location-based patterns

1   $LPF = \emptyset$;
2   **for** $i = 1$ *to* m **do**
3     $D_{V_i} \leftarrow D$;
4   **for** $j = 1$ *to* s **do**
5       $f\_set = Temoral\_Feature\_Set(j, V_i)$;
6   (Retrieve the temoral feature set of V_i in location j);
7   (t features in f_set);
8       $lp = Learn\_Pattern(f\_set, CL, \rho, \delta)$;
9   (Learn location-based patterns using the class label and thresholds);
10      $X = Generalize(lp)$;
11   (Generalize the lp set into one single pattern);
12      $LPF = LPF \cup ToFeature(X)$;
13   **end**
14 **end**

Here we developed an optimization algorithm, *SCOT* (the **S**patial **C**luster **O**ptimization **T**ool, Algorithm 2). In *SCOT* we treated each variable separately (lines 2-3). For a variable V_i, we first retrieved a feature F from its location-based pattern feature set (line 5), and constructed the set N containing all the spatial neighbors of F (line 6). For every feature f in N, we tested two conditions on the joined pattern $f \cap F$: (1) if the support of $f \cap F$ is larger than $\alpha \times$ the support of F, where α is a user specified parameter with $0 < \alpha < 1$; (2) if the growth ratio of $f \cap F$ is greater than or equal to the growth ratio of F (lien 9). If $f \cap F$ satisfied both conditions, we would combine them into one new pattern feature and update the neighbor set by adding the neighbors of f into N (lines 10-12). We repeated the above process until all features in $LPFi$ were checked (line 4), and this was done for all variables (line 2).

The key idea of *SCOT* was to check the two conditions (line 9) based on location adjacency. The first condition helped to ensure that the grown out clusters are meaningful, because infrequent patterns may just happen by chance. The second condition helped to make sure the spatial clusters were as discriminative as the single location-based patterns. In practice the clusters were usually much more informative than the single location-based patterns with respect to classification tasks (see the Experiment section for details). The output from *SCOT* wouldl be a set of binary features indicating the existence (with the value of 1s) of all the spatial clusters.

C. Forecasting through Classification

Finally, we investigated the interactions among different variables by developing a **S**patial cluster **P**attern-based

Algorithm 2: *SCOT: The **S**patial **C**luster **O**ptimization Tool*

Data:
α : the support threshold parameter
LPF : the set of LP features for variables
$V_1, V_2, ..., V_m$
Result:
SCF : the set of spatial cluster features
1 $SCF = \emptyset$;
2 **for** $i = 1$ *to* m **do**
3      $LPF_i \leftarrow LPF$ (Retrieve the LP features of V_i);
4   **while** $LPF_i \neq \emptyset$ **do**
5          $F \leftarrow LPF_i$ (Retrieve a feature from LPF_i);
6          $N = get\_neighbors(F, LPF_i)$;
7   **while** $N \neq \emptyset$ **do**
8              $f \leftarrow N$;
9   **if** *support*$(f \cap F) \geq \alpha \times$ *support(F) and growthratio*$(f \cap F) \geq$ *growthratio(F)* **then**
10                $F = f \cap F$;
11                $N = N \cup get\_neighbors(f, LPF_i)$;
12   Delete f from LPF_i;
13   **end**
14   **end**
15         $SCF = SCF \cup F$;
16   **end**
17 **end**

Algorithm 3: *SPCL The **S**patial **C**luster **P**attern-based Classifier*

Data:
$Train$: the training data set with spatial cluster features
$Test$: the testing data set with spatial cluster features
δ : the growth ratio threshold
Result:
Y : the predicted labels
1 $Y = \{0\}$
2 **for** $i = 1$ *to size(N)* **do**
3      $I \leftarrow Test$ (Retrieve an instance from $Test$);
4      $P = get\_Pattern(I)$;
5      $GR = pos\_growthratio(P, Train)$;
6   (calculate the positive growth ratio of P in $Train$)
7   **if** $GR \geq \delta s$ **then**
8          $Y_i = 1$;
9   **end**
10 **end**

TABLE I.   THE STUDIED EASTERN CENTRAL ANDES (*ECA*) AREA

Name	Spatial Extension
Box 1	$26°$S to $29°$S, $63°$W to $66°$W
Box 2	$23°$S to $26°$S, $63°$W to $66°$W
Box 3	$20°$S to $23°$S, $63°$W to $66°$W
Box 3	$17°$S to $20°$S, $66°$W to $69°$W

Classifier (*SPC*, Algorithm 3). *SPC* is an instance-based learning algorithm. We classified an instance by querying its spatial cluster patterns. For example, if we found 5 spatial clusters ($\{C^1, C^2, C^3, C^4, C^5\}$) using *SCOT* and an testing instance only containing the first and the second clusters ($\{<C^1, 1>, <C^2, 1>, <C^3, 0>, <C^4, 0>, <C^5, 0>\}$, line 3 and line 4 in Algorithm 3) we would calculate the growth ratio of this pattern in the training data set and make predictions based on this information (lines 5-9). The classification process in *SPC* is also a pattern learning process, which summerized the interactions of spatial clusters.

As an instance-based learning algorithm, the main advantage of *SPC* is that it can be approximated locally, which is very important for studying weather events from climate science data. Because same weather events (for example, extreme cold weather studied in [20]) may have different predecessors.

V. EXPERIMENT

We applied our framework in a real world data set to study an extreme weather event, i.e., extreme rainfalls, which were usually trigger events of natural hazards like floods. In the work of Boers et al. [6] the authors have shown that such extreme rainfall events in the eastern Central Andes (*ECA*, Fig. 3) were associated with certain propagation patterns. In the present paper we applied our framework to learning such atmospheric patterns and to forecasting the upcoming extreme precipitation events in *ECA*. The results show that the patterns captured through our data-based process matched well with domain knowledge-based studies and that our forecasting results outperformed the state-of-art models.

A. Experimental Setup

1) Problem Definition: With the purpose of comparison, we adopted the same experimental settings as in [6] on target area locations and extreme event definitions. The studied area (*ECA*) was defined by 4 rectangular boxes (Table I, Fig. 3). The rainfall data within *ECA* was retrieved from TRMM 3B42V7 [21], at a spatial resolution of $0.25° \times 0.25°$, and temporal resolution of every three hours for all core monsoon seasons (December through February) from 1998 to 2013. The TRMM 3B42V7 rainfall data were remote sensing-derived and gauge-calibrated, while the spatial sampling locations were distributed evenly with 144 grids within each box in *ECA* .

The extreme rainfalls were defined as time stamps when the rainfall values were above the $99th$ percentile of all historical records (1998-2013) at a location. If more than 100 extreme rainfalls happened within a 48 hours interval (16 time stamps) in any of the 4 boxes, we defined it as an **extreme rainfall event**. If there was going to be an extreme event in the next 48 hours, the class label was 1, otherwise the label was 0. With this setting, for the 16-year period the ratio of positive and negative events was $797 : 10355$.

2) The Predictor Variables: We chose 5 fields of meteorological predictor variables with certain spatial and temporal information at constant pressure surfaces (3 vertical levels) from the Modern-Era Retrospective Analysis for Research and Applications (MERRA) dataset [22] (Table II). We also used precipitation data from TRMM 3B42V7 [21]. All predictor variable data were retrieved for the same time period of monsoon seasons (December through February) from 1998 to 2013 and were employed at a spatial resolution of

Name	Level(hPa)
Temperature	300,500,850
Geopotential Height	300,500,850
Meridional Wind	300,500,850
Zonal Wind	300,500,850
Specific Humidity	300,500,850
Precipitation	–

$1.25° \times 1.25°$ and a temporal resolution of 3-hour intervals. To remove the seasonal effect, we subtracted the monthly means from the data. Then We discretized the features by categorizing the normalized data into 3 intervals, abnormally high (above 90th percentile), abnormally low (below 10th percentile), and normal (between 10th and 90th percentiles).

3) Parameters: In the experiments we used data from past 8 time stamps (24 hours) to forecast the extreme rainfall events for the upcoming 16 time stamps and divided the whole data set into a training (1998-2009,12 years) set and a testing (2010-2013, 4 years) set. We ran $SCOT$ with a support threshold parameter $\alpha = 0.9$. We chose a growth ratio threshold of 10 for classification using SPC. We also ran SPC using LPF from algorithm 1 directly for comparison. Due to the unbalanced setting we evaluated our forecasting results using the F1 Score ($F1 = 2TP/2TP + FP + FN$, where TP is the number of true positives, FP is the number of false positives, TN is the number of true negatives, and FN is the number of false negatives).

B. Results and Discussion

In the study of [6] the authors used overlapped time periods (but different precipitation data sets) for training(1998-2013) and testing (2001-2013), and their complex network model achieved an F1 of 0.50. Our model outperformed it with an F1 of 0.745 by using exactly the same setting. We achieved a best prediction result of F1=0.47 when using non-overlapped periods as training and testing (Table III).

The numbers of location-based patterns and spatial clusters were very sensitive to the initial parameters. The F1 values increased with the increase of initial support and growth ratio thresholds. This is reasonable because the higher the thresholds, the better quality the learned patterns/spatial clusters should have. Low thresholds resulted in more noisy patterns which hurt the classifier's performance. But on the other hand, when the thresholds were too high, the learned patterns/spatial clusters became too few for representing the distributions in the data, and the F1 value dropped (last row in Table III).

Forecasts using spatial cluster features (SCF) achieved better results on testing set for all parameter settings than forecasts using location-based pattern features (LPF). Also, SCF had a much higher growth ratio than LPF, with the sacrifice of support. Single LPF represents a variable's temporal evolution in single location, which may have very little effect on the target events. Therefore, SCF will be more informative in distinguishing extreme weather events due to its larger size. The final inter-cluster pattern (or inter-variable pattern) learned from individual testing instance's SCF by SPC usually has an infinit growth ratio.

a. Positive Location Based Patterns of H850

b. Positive Spatial Cluster 1 to 6 of H850 (left->right, top->bottom)

c. Positive Location Based Patterns of QV850

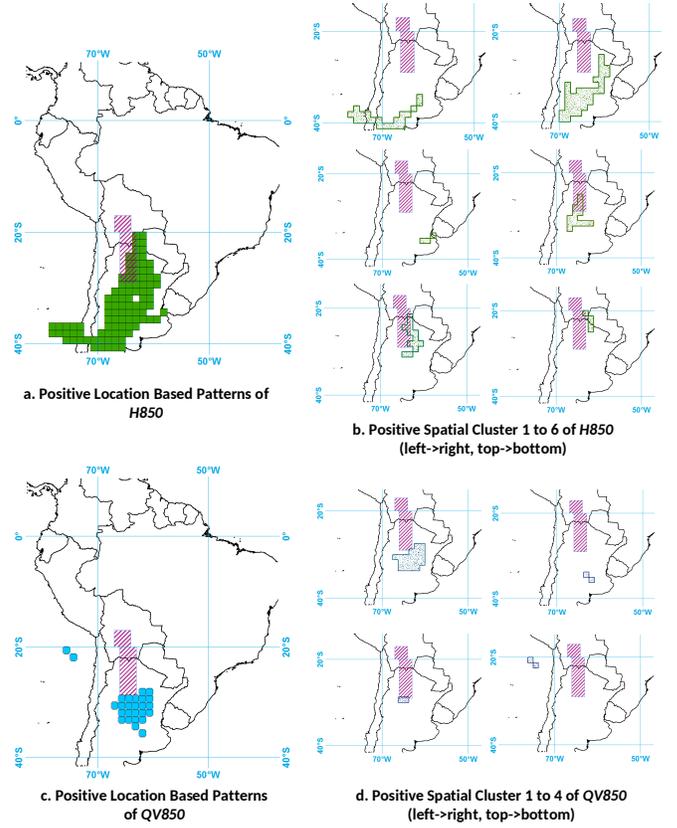d. Positive Spatial Cluster 1 to 4 of QV850 (left->right, top->bottom)

Fig. 3.   Experimental results using ρ=0.10 and δ=5: (1)location based patterns for Geopotential Height at 850hpa (H850, a) and Specific Humidity at 850hpa (QV850, c); (2) spatial clusters learned from the patterns (b,d). The hatched regions are the studied eastern Central Andes (ECA) area.

Another positive effect for learning SCF is to reduce the risk of over-fitting. Variables' behaviors in single location (LPF) may only happen by chance, but such similar behaviors observed in a large area (SCF) is rarely a random event. Classification using LPF achieved a near perfect F1 (0.981) on training data but a lower F1 in testing data.

The location-based patterns and spatial clusters for the variables of H850 (Geopotential Height at 850hpa) and QV850 (Specific Humidity at 850hpa) (Fig.4) we found conformed with domain study findings [6], [5]. For example, the H850 sensitive areas found in [6] were included in the H850 cluster 3 (Fig.4b) of our results. In Fig.4, we also demonstrated the H850 and QV850 maps for an example positive event (December 5 to 7, 2010) to show how the learned patterns captured the system dynamics.

VI. CONCLUSION AND FEATURE WORK

Extreme weather events such as extreme rainfalls have a wide negative impact on our society and economy, and there is room for improvements on forecasting such events in current weather models [1]. In this study we presented a framework for learning patterns from climate science data and forecasting extreme weather events. We adopted our framework for real wold studies on forecasting extreme rainfall events. Our model outperformed the state-of-art methods. The patterns we found using data-based approach matched

TABLE III.   Pattern Statistics

| Sup.	GR	No.LPF(Pos/Neg)	Ave.Sup.of LPF	Ave. GR of LPF	F1 (LPF)	No.SCF (Pos/Neg)	Ave. Sup.of SCF	GR of SCF	F1(SCF)
0.05	3	1801/4303	0.04	3.06	0.26	334/338	0.03	5.41	0.31
	4	1322/3252	0.04	3.81	0.29	287/226	0.03	7.58	0.34
	5	784/2383	0.04	5.14	0.31	194/76	0.03	11.46	0.39
0.10	3	1238/2305	0.08	3.11	0.27	292/101	0.06	5.61	0.37
	4	935/1800	0.09	3.91	0.34	203/81	0.06	10.03	0.42
	5	436/1122	0.09	5.07	**0.37**	112/61	0.07	15.32	**0.47**
0.15	3	725/1351	0.11	2.88	0.31	199/72	0.11	7.13	0.40
	4	475/780	0.14	4.13	0.36	94/54	0.14	10.44	0.43
	5	113/477	0.13	5.13	0.29	37/24	0.13	13.46	0.32
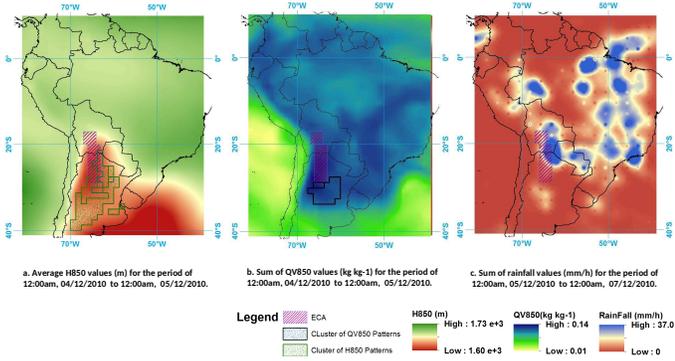
Fig. 4.   An example extreme rainfall event (c) happens in *ECA* in the 48 hours period from 12:00am Dec.5, 2010 to 12:00am Dec.7, 2010. It also demonstrates that the spatial clusters shown in Fig. 3 captured the dynamics in the fields of Geopotential Height at 850hpa (H850, a) and Specific Humidity at 850hpa (QV850, b).

well with domain studies. In our future work, we want to explore the possibility of using different data representations in our framework and extend our analyses to other fields. The framework is potentially applicable to other geographic areas and other spatiotemporal events like tornadoes.

REFERENCES

[1] J. T. Overpeck, G. A. Meehl, S. Bony, and D. R. Easterling, "Climate data challenges in the 21 st century," *Science(Washington)*, vol. 331, no. 6018, pp. 700–702, 2011.

[2] J. H. Faghmous and V. Kumar, "A big data guide to understanding climate change: The case for theory-guided data science," *Big data*, vol. 2, no. 3, pp. 155–163, 2014.

[3] D. L. G. II, S. V. Pendse, K. Padmanabhan, M. P. Angus, I. K. Tetteh, S. Srinivas, A. Villanes, F. Semazzi, V. Kumar, and N. F. Samatova, "Coupled heterogeneous association rule mining (charm): Application toward inference of modulatory climate relationships," in *Data Mining (ICDM), 2013 IEEE 13th International Conference on*. IEEE, 2013, pp. 1055–1060.

[4] V. Kumar, M. Steinbach, P.-N. Tan, S. Klooster, C. Potter, and A. Torregrosa, "Mining scientific data: Discovery of patterns in the global climate system," in *Joint Statistical Meeting*, 2001.

[5] C. Vera, W. Higgins, J. Amador, T. Ambrizzi, R. Garreaud, D. Gochis, D. Gutzler, D. Lettenmaier, J. Marengo, C. Mechoso *et al.*, "Toward a unified view of the american monsoon systems," *Journal of Climate*, vol. 19, no. 20, pp. 4977–5000, 2006.

[6] N. Boers, B. Bookhagen, H. Barbosa, N. Marwan, J. Kurths, and J. Marengo, "Prediction of extreme floods in the eastern central andes based on a complex networks approach," *Nature communications*, vol. 5, 2014.

[7] A. C. Lozano, H. Li, A. Niculescu-Mizil, Y. Liu, C. Perlich, J. Hosking, and N. Abe, "Spatial-temporal causal modeling for climate change attribution," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009, pp. 587–596.

[8] X. Chen, Y. Liu, H. Liu, and J. G. Carbonell, "Learning spatial-temporal varying graphs with applications to climate data analysis." in *AAAI*, 2010.

[9] D. L. Gonzalez, Z. Chen, I. K. Tetteh, T. Pansombut, F. Semazzi, V. Kumar, A. Melechko, and N. F. Samatova, "Hierarchical classifier-regression ensemble for multi-phase non-linear dynamic system response prediction: Application to climate analysis," in *Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on*. IEEE, 2012, pp. 781–788.

[10] D. Wang, W. Ding, K. Yu, X. Wu, P. Chen, D. L. Small, and S. Islam, "Towards long-lead forecasting of extreme flood events: a data mining framework for precipitation cluster precursors identification," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2013, pp. 1285–1293.

[11] J. Han, H. Cheng, D. Xin, and X. Yan, "Frequent pattern mining: current status and future directions," *Data Mining and Knowledge Discovery*, vol. 15, no. 1, pp. 55–86, 2007.

[12] H. Xiong, S. Shekhar, Y. Huang, V. Kumar, X. Ma, and J. S. Yoo, "A framework for discovering co-location patterns in data sets with extended spatial objects." in *SDM*, 2004, pp. 78–89.

[13] D. Wang, W. Ding, H. Lo, M. Morabito, P. Chen, J. Salazar, and T. Stepinski, "Understanding the spatial distribution of crime based on its related variables using geospatial discriminative patterns," *Computers, Environment and Urban Systems*, vol. 39, pp. 93–106, 2013.

[14] T.-c. Fu, "A review on time series data mining," *Engineering Applications of Artificial Intelligence*, vol. 24, no. 1, pp. 164–181, 2011.

[15] M. J. Zaki, "Generating non-redundant association rules," in *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2000, pp. 34–43.

[16] P. Tan, M. Steinbach, V. Kumar, C. Potter, S. Klooster, and A. Torregrosa, "Finding spatio-temporal patterns in earth science data," in *KDD 2001 Workshop on Temporal Data Mining*, vol. 19, 2001.

[17] K. Koperski and J. Han, "Discovery of spatial association rules in geographic information databases," in *Advances in spatial databases*. Springer, 1995, pp. 47–66.

[18] G. Liu, H. Zhang, and L. Wong, "Finding minimum representative pattern sets," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012, pp. 51–59.

[19] C. Wang and S. Parthasarathy, "Summarizing itemset patterns using probabilistic models," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2006, pp. 730–735.

[20] R. H. Grumm and R. Hart, "Standardized anomalies applied to significant cold season weather events: Preliminary findings," *Weather and forecasting*, vol. 16, no. 6, pp. 736–754, 2001.

[21] G. J. Huffman, D. T. Bolvin, E. J. Nelkin, D. B. Wolff, R. F. Adler, G. Gu, Y. Hong, K. P. Bowman, and E. F. Stocker, "The trmm multisatellite precipitation analysis (tmpa): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales," *Journal of Hydrometeorology*, vol. 8, no. 1, pp. 38–55, 2007.

[22] M. Bosilovich, S. Schubert, M. Rienecker, R. Todling, M. Suarez, J. Bacmeister, R. Gelaro, G. Kim, I. Stajner, and J. Chen, "Nasas modern era retrospective-analysis for research and applications (merra)," *US CLIVAR Variations*, vol. 4, no. 2, pp. 5–8, 2006.