5th International Workshop on Climate Informatics
September 24-25, 2015
Hosted by the National Center for Atmospheric Research in Boulder, CO

ONLINE CHANGE POINT DETECTION FOR REMOTE SENSING TIME SERIES

Xi C. Chen[1], Yuanshun Yao[1], Sichao Shi[1], Vipin Kumar[1] , James H. Faghmous[1,2]

***Abstract*—Lack of the global knowledge of land-cover changes limits our understanding of the earth system, hinders natural resource management and also compounds risks. Remote sensing data provides an opportunity to automatically detect and monitor land-cover changes. Although changes in land cover can be observed from remote sensing time series, most traditional change point detection algorithms do not perform well due to the unique properties of the remote sensing data, such as noise, missing values and seasonality. We propose an online change point detection method that addresses these challenges. Using an independent validation set, we show that the proposed method performs better than the four baseline methods in both of the two testing regions, which has ecologically diverse features.**

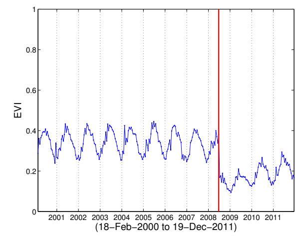
I. MOTIVATION

Land-cover change is a general term for any modification of the Earth's terrestrial surface, which includes both natural (e.g., forest fires and floods) and man-made events (e.g., urbanization and deforestation). Land cover events are associated with some of our greatest environmental concerns including climate change, biodiversity loss and the pollutions of water, soils and air [1]. Monitoring these land-cover changes on a global scale, hence, is important to researches and policymakers.

Traditionally, changes in the land-cover are manually recorded and reported mostly by government agencies. Due to the large cost of human canvassing, the massive area of the Earth's surface and varieties of land-cover change events, the records, compared with the total number of events, are very incomplete. Additionally, mostly due to political issues, most governments do not share their event records. As a result, we lack a global atlas of land cover changes.

Publicly available remote sensing data provides an opportunity to automatically detect and monitor land-cover change on a global scale. Remote sensing instruments scan the earth surface every one or two days. They continuously provide multi-dimensional reflectance signals for every location. The multi-dimensional reflectance signals captured from one remote sensing instrument within a period can be viewed as a time series. Since most changes in the land cover type can lead changes in the reflectance signals, we can detect those land cover changes as certain types of change points in the remote sensing time series. Figure 1 is an Enhanced Vegetation Index (EVI) time series for a location in the state of California (US) that was burned in year 2008. EVI values are a proxy for the greenness of a given location. When a fire burns a forest, the loss of trees causes the greenness (or EVI) of that location to drop. As a result, most of the forest fire can be detected as a sudden drop in their EVI time series.

(a) A forest fire.



(b) An EVI time series.

Fig. 1: The Basin Complex fire, which was started by lightning near Big Sur, California in June 2008, consumed more than 160,000 acres before it merged with another fire. Over $120M were spent fighting it.

Detecting change point events from remote sensing time series faces three major challenges: (i) noise; (ii) missing values; and (iii) seasonality. As a result, the performance of most existing methods, such as seasonal ARIMA [2] and Gaussian process model method [3], have limited applicability. In this work, we propose an online change point detection method that can avoid the previously mentioned challenges and performs better than most of the existing methods.

II. METHOD

The proposed method has two steps. First, an anomaly score matrix is constructed for a target time series. Then, a change score of any time step is calculated from the anomaly score matrix.

Corresponding author: Xi C. Chen, University of Minnesota chen@cs.umn [1] Department of Computer Science, University of Minnesota [2] The Icahn School of Medicine at Mount Sinai

For a time series with n time steps, its *anomaly score matrix*, we refer it as M, is a n by n matrix. Each column of the matrix contains the anomaly scores for a given time step that is calculated using different sets of historical time points. In detail, column j in M contains the anomaly scores for time step j. The entry in column j and row i, i.e., $M(i, j)$, when $i < j$, is the anomaly score for time step j based on an predictive model that was trained by time points from x_{i-w} to x_{i-1}, where w is the pre-defined length of the training window.$M(i, j)$ is empty when $i < j$. Most existing predictive models can be used here. Note that the selected predictive model, in most cases, should vary with the application.

The change score of each time step is reported from the anomaly score matrix. For any time step t and a user-defined parameter l, we define the sub-matrix $B_t = M(t-l : t-1, t-l : t-1)$ as the reference area and the sub-matrix $A_{t,p} = M(t-l : t-1, t+1 : t+p)$ as the test area, where p is a unknown parameter. The score for time step t with unknown parameter p can be obtained as

$$s(t,p) = \frac{|mean(A_{t,p}) - mean(B_t)|}{\sqrt{var(A_{t,p})/c(A_{t,p}) + var(B_t)/c(B_t)}}$$

where, for any sub-matrix, mean(.) is the mean value of all nonempty entries, var(.) is the variance of all nonempty entries and c(.) is the total number of all nonempty entries.

Then, the change score of time step t is

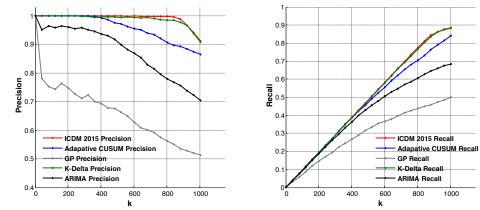
$$s(t) = \max_{p > p_{min}, t+p<n} s(t,p)$$

The change score for a time series is the maximum change score of all its time steps.

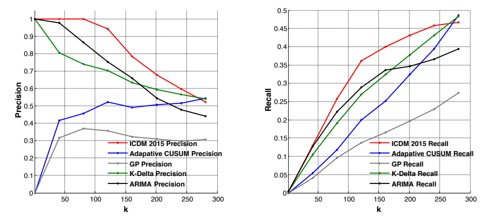
III. EVALUATION

To test the accuracy of our proposed method we attempt to autonomously identify forest fires using EVI time series, a remote sensing data that has been used for forest fire detection. Since EVI time series is periodic, the baseline algorithms we choose are (i) seasonal ARIMA [2]; (ii) adaptive CUSUM – a CUSUM algorithm [4] that is adapted for seasonal time series; (iii) Gaussian process method [3] that is designed to detect changes in periodic time series and (iv) V2Delta that is designed to specifically detect sudden drops under noisy conditions in remote sensing data.

To show the robustness and generalization ability of different algorithms, we chose two ecologically varied regions in Northern and Southern California. One is near San Diego (the region between $32.85^oN \sim 32.53^oN$ and $116.4^oW \sim 117^oW$) and the other is near Salinas (the region between $36.4^oN \sim 35.9^oN$ and $121.2^oW \sim 122^oW$). These geographic areas represent diverse regions with different variability, land cover types, geography, and noise characteristics.



(a) The test region near Salinas



(b) The test region near San Diego.

Fig. 2: The precision and recall curves as a function of k. When the size of the positive set is k, we label the time points with the top-k data change scores as positives.

We use *Precision* and *Recall* as evaluation metrics. We compare the performance of different algorithms as a function of k (*i.e.* how many points do we allow to be positives or how low a change score do we consider as a change event). Figure 2 shows precision (left panel) and recall (right panel) for different values of k. From the figure, we observe that our method has better precision and recall, especially when the amount of positive is not too large. Note that in Northern California where the data are more homogeneous and less noisy, we perform similarly to other methods. But in Southern California where the data are more challenging, our method is more robust, even compared to an application-specific algorithm such as K-Delta (green curve).

ACKNOWLEDGMENTS

This research was supported in part by the National Science Foundation Awards 1029711, 0905581, and 1464297. The NASA Award NNX12AP37G and an UMII MnDRIVE Fellowship. Access to computing facilities was provided by the University of Minnesota Supercomputing Institute.

REFERENCES

[1] E. Ellis and R. Pontius, "Land-use and land-cover change," *Encyclopedia of earth*, 2007.

[2] J. D. Hamilton, *Time series analysis*, vol. 2. Princeton university press Princeton, 1994.

[3] V. Chandola and R. R. Vatsavai, "A gaussian process based online change detection algorithm for monitoring periodic time series.," in *SDM*, pp. 95–106, SIAM, 2011.

[4] M. Basseville, I. V. Nikiforov, *et al.*, *Detection of abrupt changes: theory and application*, vol. 104. Prentice Hall Englewood Cliffs, 1993.