Unevenly Spaced Spatio-Temporal Time Series Analysis in Context of Volcanoes Eruptions

Grigory Trifonov

Moscow State University, Moscow

trifonov.grigory@gmail.com

Abstract

Paper presents a simplistic approach towards the detection and dynamics analysis of volcanic eruptions represented as unevenly spaced spatio-temporal time series of satellite retrieved hot spots. The paper discusses isolation and interpolation of hot spots data produced by Nightfire algorithm for the purposes of short-term volcanic activity ARIMA based forecasting. The case study for Chirpoi Snow volcano is presented.

The work is supported by RFBR (grants 14-07-00548, 16-07-01028, 15-29-06045).

1 Introduction

Volcanic eruptions are one of well-known types of natural hazard. Some volcanoes are better studied than the others, this mostly depends on their location, and those located on uninhabited islands may erupt without people in field observing them, while the others may erupt nearby cities and attract thousands of people. There are different kinds of volcanic activity such as: gas emissions, ash plumes, lava flows and domes, etc. These events may accompany one another. Some volcanoes demonstrate regular patterns in their activity others not. Events listed may be tracked by instrumental networks consisting of different instruments like seismographs, regular and thermal cameras, gas analyzers and others. Such networks sometimes are impossible to deploy due to various reasons. However it is possible to utilize satellites based instruments; such instruments do not require any special onsite installation and provide coverage for the entire planet.

1.1 Volcanoes satellite observations

As of 2016 there is no single satellite developed specifically for volcanology purposes. Fortunately there are a lot of meteorological satellites suitable for volcanology purposes. There are basically three kinds of satellite data which has a widespread use for volcanology purposes, these are:

1. visual data — which is basically just an image;
2. infrared data — most of volcanic events are accompanied with heat emission, hence they are visible in infrared spectra;
3. radar readings — lava flows and domes may be observed as they change the landscape.

This paper describes an approach towards infrared data analysis.

1.2 Current problem state

There are at least two global monitoring systems of volcanic activity based on hot spots detection in satellite data, namely: MODVOLC [4], MIROVA [3]. The oldest one (MODVOLC) has been in service for more than a decade by now [8]. Both systems are using MODIS sensor data as an input, which is provided by two NASA's polar-orbiters Terra and Aqua, launched in 1999 and 2002 respectively. Each service is using hot spots detection algorithm of its own. Essentially any such algorithm is based on the Plank curves fitting with some variations and modifications, book "Thermal Remote Sensing of Active Volcanoes" by Andrew Harris [2] provides a good introduction on how does it work. Both services provide real time and historical hot spots attributed to volcanoes and alerts based on some radiant heat threshold value. Neither of services does any additional hot spot analysis.

1.3 Suggested approach

This paper suggests an alternative approach towards volcanic activity monitoring and dynamics analysis. Instead of just tracking hot spots it is suggested to analyze hot spots as time series for each volcano. Such an approach will let to better adjust alert thresholds as well as to look up behavioral patterns and anomalies for all the existing volcanoes rather than just to detect potential eruptions.

2 Data

This study is using hot spots detected by a Nightfire algorithm [1]. Hot spots come in a form of a separate csv file for each 24 hours; each csv file contains around 100 fields out of which the most important and suitable for analysis are: pixel longitude and latitude, radiant heat, estimated black body temperature, estimated black body area, cloud conditions, satellite angles, detection quality flags and original image geometry. Due to its nature the data used have some specificity, which are worth consideration.

2.1 Satellite imaging

Nightfire is using VIIRS sensor nighttime data to detect hot spots. VIIRS sensor is basically a next improved

---

Proceedings of the XVIII International Conference «Data Analytics and Management in Data Intensive Domains» (DAMDID/RCDL'2016), Ershovo, Russia, October 11 - 14, 2016

generation of MODIS sensor, as of April 2016 there is one polar-orbiting satellite carrying VIIRS sensor operating (Suomi-NPP launched in 2011). Satellite continuously goes from one Earth pole to another following sun terminator. Every full circle satellite is crossing equator once in a day time and once in a night time. Satellite is constantly reading pixels in a line orthogonal to its path, each pixel size at Nadir (directly under the satellite) is 742x776 m, this size is non linearly growing towards the edge of scan, scan width is 3040 km, the satellite does 14 full revolutions each 24 hours, which essentially means that every point at equator is seen at least once every day and once every night. However for higher latitudes due to a considerable swath overlap locations are seen several times each day and each night.

2.2 Hot spot sources

There are several major sources of hot spots, anthropogenic ones, such as: gas flares over oil fields, high temperature manufacturing, thermal power stations as well as natural ones of which the most significant are forest fires and volcanic activity. Coincidentally hot spots of volcanic and forest fire origins have closely overlapping temperature ranges.

Nightfire has been originally geared towards the detection and tracking of gas flares over oil fields. Gas flares are smaller than VIIRS pixel (they either lay inside one pixel or between several neighboring pixels); hence Nightfire performs subpixel analysis, to estimate heat source size and energy. In the case of volcanic events it is no rare occasion to have dozens of hot-spots per image, which are all part of the same lava field for instance [7]. Thus it is crucial to somehow group such hot spots as the ones attributed to a single event.

2.3 Missing data

Another problem apart from separation of hot spots of volcanic and non-volcanic origins is a high number of missing or incomplete readings. There are several reasons for volcanic hot spots to become corrupted or even missing:

- clouds – may cover the area of interest, this leads to either missing hot spots or hot spots with a lower radiant heat than it should be;
- volcanic gasses and ash – these act more or less the same as clouds;
- volcano being too far off nadir – cauldron like volcanoes with lava lake inside crater may produce hot spots only at angles close to nadir, since otherwise satellite's sensor can't see lava lake;
- polar day – since Nightfire works with nighttime data only there will be no hot spots detected for the entire duration of polar day.

2.4 Data specifics summary

To summarize, there are three key points about data involved:

1. there are hot spots readings spanning since the early 2012 available for each volcano;
2. readings are unevenly spaced in time with a distance varying between 2-24 hours (discounting polar day/night cases);
3. each day of the observations comes in a form of a set of hot-spots with varying coordinates, scan time, radiant heat and temperature.

The three key points mentioned allowed to classify data as a set of unevenly spaced spatio-temporal time series.

3 Analysis

To observe volcano as a process it is first necessary to bisect the data related to the volcano in question. The simplest method to attribute hot spot to a volcano is to check how far it is from it. Basically the largest thermal anomalies, which can be seen from space, are lava flows and these typically would not reach further than 20 km from the volcano summit, hence distance based hot spots filtering leaves out most of the hot spots of non-volcanic origins. This simplistic approach does not protect from false attribution of forest fires to a volcanic activity, however as it has been already mentioned forest fires temperature range closely overlaps with that of volcanic events, hence such discrimination is a subject of future research.

3.1 Interpolating data

To interpolate time series for a specific volcano it is first necessary to somehow regularize the data presented within a single reading. As was already mentioned each reading may consist of several hot spots of varying radiant heat, temperature, area, cloud conditions all of them with different coordinates and satellite angles. There could be no more than a single hot spot for each pixel of the original satellite image. Knowing satellite altitude, nadir and azimuth angles as well as how does internally VIIRS sensor work [6] it is possible to calculate the area of the original pixel for each hot spot. A pixel area may be calculated with the following formula:

$$\alpha = asin\left(\frac{R \times sin(\pi - \beta)}{H + R}\right)$$

$$px_s = \Lambda R \frac{px_{ns}}{H}\left(\frac{cos(\alpha)}{\sqrt{\left(\frac{R}{R+H}\right)^2 - sin(\alpha)^2}}\right)$$

$$px_t = \Lambda \frac{px_{nt}}{H}\left(cos(\alpha) - \sqrt{\left(\frac{R}{R+H}\right)^2 - sin(\alpha)^2}\right)$$

Where R – earth radius, β – satellite zenith angle, H – satellite height, px_{ns} – pixel along scan size at nadir (776m), px_t – pixel along track size at nadir (742m), Λ – aggregation group which is $\frac{1}{3}$ if $\alpha \geq 44.68º$, $\frac{2}{3}$ if $\alpha \geq 31.589º$ and 1 otherwise.

This formula produces an upper boundary area estimate, while hot spot area calculated by Nightfire itself could be used as a lower boundary estimate. Next step would be to aggregate both boundaries within any given reading:

$$A_{RU} = \sum_{p} A_{PU}$$

$$A_{RL} = \sum_{P} A_{PL}$$

$$RH_R = \sum_{p} RH_P$$

$$T_R = \sum_{p} A_P \times \frac{T_P}{A_R}$$

Where A_{RU} – reading active area estimated upper bound, A_{PU} – hot spot area upper bound, A_{RL} - reading active area estimated lower bound, A_{PL} - hot spot area lower bound, RH_R – reading's aggregate radiant heat, RH_P – hot spot radiant heat, T_R – estimated reading temperature, T_P – estimated pixel temperature.

The majority of the time series analysis theory is geared towards the analysis of regularly spaced time series; hence there are essentially two ways to analyze an unevenly spaced time series. The first one would be to interpolate an unevenly spaced time series and proceed with analysis of a regularly spaced time series. The second method would be to analyze unevenly spaced time series without performing such a transformation. For this paper the first approach as the simpler one has been chosen. Thus the resulting unevenly-spaced time series are linearly interpolated into their evenly-spaced form.

3.2 Analyzing interpolated time series

It comes as no surprise that in the majority of cases it would be impossible to forecast eruption dynamics due to the lack of data and high irregularity of the processes involved. However some volcanoes may show some regularity in their activity, which may signalize an applicability of classic time series forecasting techniques. One such popular technique ARIMA model widely used in econometric will be used in the case study to attempt to forecast volcanic activity of a selected volcano.

ARIMA (autoregressive integrated moving average) is widely used in econometrics. The model uses an initial differencing step to reduce non-stationarity of data combined with both autoregressive and moving-average models. Model is usually denoted as $ARIMA(p,d,q)$ where p – the order of autoregressive model, d – is the degree of differencing, q – is the order of moving-average model. $ARIMA(p,d,q)$ model is given by:

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right)(1-L)^d X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right)\epsilon_t$$

3.3 Forecasting Chirpoi Snow activity

Snow is a stratovolcano located on an uninhabited volcanic island Chirpoi. The island is located in the Sea of Okhotsk between Simushir and Urup in the Kuril island chain. Snow has been continuously erupting for a since April 2012, its eruption seems to follow a pattern, where periods of activity are interleaved with short periods of low to no activity. There is no other dedicated observing equipment but the only unmanned seismic station, which may be disabled for weeks due to battery discharge, located on the island. Nightfire has the full data for this eruption and there are no other potential sources of hot spots, which could have added additional noise. Factors mentioned make Snow a good candidate for a case study.

This case study attempts to answer the question: "Is it possible to forecast eruption power output, at least for some volcanoes?" Power and radiant heat are interchangeable in this context. Power poses the most interest since there is a direct relation between radiant heat observed and the volume of lava being erupted.

Radiant heat readings Figure 1 exhibit non stationary behavior, however after differencing step **Ошибка! Источник ссылки не найден.** situation gets better. For a difference time series mean is 0 standard deviation is 11.38. In fact early weeks of eruption impact standard deviation. Trimming first 6 weeks off time series brings standard deviation to about 8 and this value is valid for almost every segment of time series. Thus after differencing step process shows strong signs of being near stationary, which means that d parameter of ARIMA model will be 1 in this case. For p and q start parameters 3 will be taken.

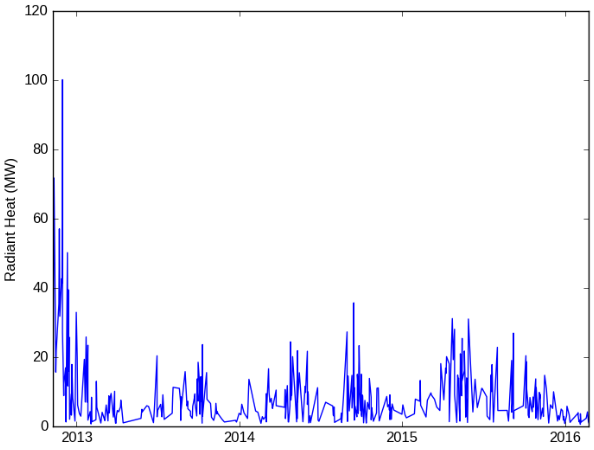

Figure 1 Chirpoi Snow Radiant Heat (MW) December 2012 - February 2016

To perform ARIMA forecast Pythons StastModels [5] package has been used. Model has been fit on 100 readings and attempted to predict the next 24 hours, next the closest to 101 reading hour prediction has been used to cross validate. One hundred retrains in the middle of the data set were made. Even though in most cases ARIMA managed to perform forecast sometimes it was impossible due to random shocks, which disrupted data strong enough to make data set non stationary, hence

ARMA step is not applicable.

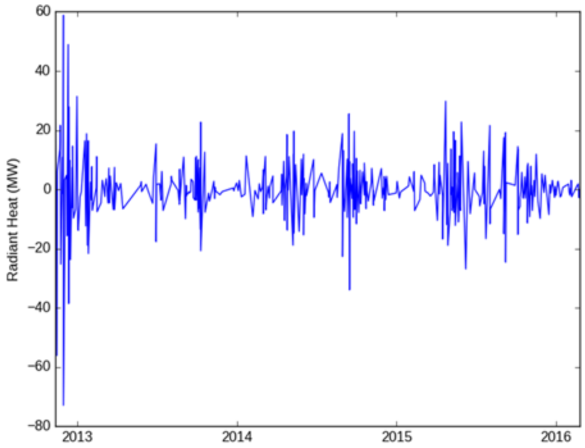

Figure 2 Chirpoi Snow Radiant Heat increment (MW) December 2012 - February 2016

Conclusion

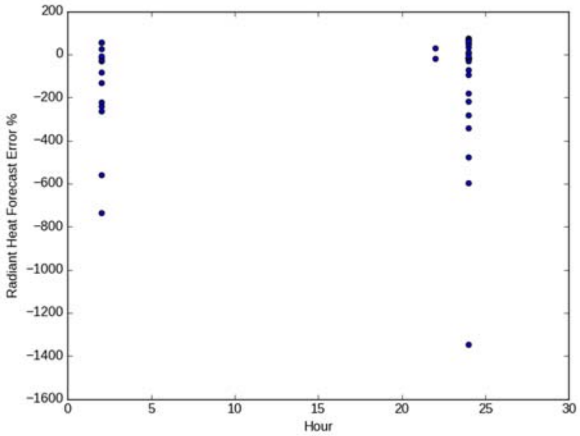

Figure 3 Relative error of ARIMA forecast

Unfortunately due to persistently poor weather conditions, there were too many gaps in data with a temporal distance between two readings coming up to several weeks. Thus, ARIMA did not manage to do reasonably accurate forecasts. As it could be seen on Fugure 3 relative error comes up to hundreds of percent. The situation could've been improved via introduction of new data sources, however many volcanoes are not easily accessible leaving little to no choice but to use satellite data only. For instance, Chirpoi volcano reviewed in this article is located at an uninhabited island. Moreover even if there does exist an instrumental network for a specific volcano it still requires a lot of work, both organizational and technical, to integrate its data, whereas satellite imagery serves as a universal data source. Hence the most reasonable approach seems to be to integrate data from different classes and generations of satellites to improve its temporal resolution and shorten the gaps between readings.

Future work

Satellite infrared imagery proves itself to be an interesting source of data for volcano research. Current plans involve, although not limited to:

1. data delivery latency improvements – via software installation at receiving station around the world (currently installed on Kamchatka);
2. data volume improvements – via incorporation of older satellites data and SAR satellites data;
3. better classification – via various clustering algorithms application and evaluation, preliminary results of a graph based hierarchical clustering have been very promising;
4. better insight – via neural networks based algorithms application, such an approach seem to be more stable in the presence of large data gaps.

This would not only allow better analysis of activity patterns in global volcanism, but will improve quality of neighboring data products such as the forest fire monitoring product.

References

[1] Christopher D. Elvidge et. al., VIIRS Nightfire: Satellite Pyrometry at Night, ISSN 2072-4292, http://www.mdpi.com/2072-4292/5/9/4423/pdf

[2] Andrew Harris, Thermal Remote Sensing of Active Volcanoes A USER'S MANUAL, ISBN 978-0-521-85945-5

[3] MIROVA Near Real Time Volcanic HotSpot Detection System http://www.mirovaweb.it

[4] MODVOLC Near-real-time thermal monitoring of global hot-spots http://modis.higp.hawaii.edu/

[5] Skipper Seabold, Jonathan Taylor, StatsModels statistical package, Josef Perktold, http://statsmodels.sourceforge.net/devel/index.html

[6] Curtis Seaman, Beginner's Guide to VIIRS Imagery Data, http://rammb.cira.colostate.edu/projects/npp/Beginner_Guide_to_VIIRS_Imagery_Data.pdf

[7] Grigory Trifonov (1), Mikhail Zhizhin (2), and Dmitry Melnikov, Nightfire method to track volcanic eruptions from multispectral satellite images, http://meetingorganizer.copernicus.org/EGU2016/EGU2016-5409-1.pdf

[8] Robert Wright, MODVOLC: 14 years of autonomous observations of effusive volcanism from space, http://www.higp.hawaii.edu/~wright/geol_soc_426.pdf