Predicting Volcanic Eruptions: Case Study of Rare Events in Chaotic Systems with Delay

Justin Parra[1], Olac Fuentes[1],
Elizabeth Anthony[2], and Vladik Kreinovich[1]
Departments of [1]Computer Science and [2]Geological Sciences
University of Texas at El Paso, 500 W. University, El Paso, TX 79968, USA,
jrparra2@miners.utep.edu, ofuentes@utep.edu,
eanthony@utep.edu, vladik@utep.edu

***Abstract*—Volcanic eruptions can be disastrous; it is therefore important to be able to predict them as accurately as possible. Theoretically, we can use the general machine learning techniques for such predictions. However, in general, without any prior information, such methods require an unrealistic amount of computation time. It is therefore desirable to look for additional information that would enable us to speed up the corresponding computations. In this paper, we provide an empirical evidence that the volcanic system exhibit chaotic and delayed character. We also show that in general (and in volcanic predictions in particular), we can speed up the corresponding predictions if we take into account chaotic and delayed character of the corresponding system.**

I. Machine-Learning Base Predictions of Dynamical Phenomena: The Main Challenge

Predictions are important. In many application areas, we are interested in predicting the values $y(t_f)$ of different quantities y at different future moments of time t_f.

To predict this future value, we can use the values of the related quantities $x_1(t)$, ..., $x_n(t)$ at the present moment of time t_0 and at the previous moments of time $t < t_0$.

Machine learning as a natural tool for predictions. To be able to predict the desired future value $y(t_f)$, we need to know the dependence of $y(t_f)$ on the values $x(t) = (x_1(t), \ldots, x_n(t))$ measured at previous moments of time.

Specifically, to make a prediction m moments ahead, for the moment of time $t_f = t_0 + m$, we can need to know how the value $y(t_f)$ depends on the tuples $x(t)$ corresponding to moments $t \le t_f - m$.

In some practical situations, we know the desired dependence. For example, we know Newton's equations that describe the orbit of an asteroid. Thus, we can use these known equation to make the corresponding predictions.

In other cases, however, we do not know the desired dependence. In such situations, we can use the general techniques for determining the desired dependence from the observations – techniques of machine learning; see, e.g., [4]. Examples of such techniques include neural networks (in particular, deep learning networks [8]), support vector machines, etc.

To find out how a quantity y depends on quantities $x_i(t)$, machine learning methods use, an input, patterns (x, y) consisting of the observed values x_i and the corresponding observed value y.

In particular, for each period m, to find the dependence enabling us to predict m moments into the future, we use patterns of the type $(x, y(t_f))$, where:

- $y(t_f)$ is the observed value y at moment t_f and
- x is a collection of all the x-tuples $x(t)$ observed at moments $t \le t_f - m$.

Let us describe this idea in precise terms.

What information do we have. For several situations $k = 1, \ldots, K$, we have observed the values $x_i^{(k)}(t)$ and $y^{(k)}(t)$ of the quantities x_i and y at several consequent moments of time $t_-^{(k)}$, $t_-^{(k)} + \Delta t$, $t_-^{(k)} + 2\Delta t$, ..., $t_+^{(k)}$.

How we can predict: theoretical possibility. For each time period m, to be able to predict m moments of time ahead, we use the following patterns to train the corresponding machine learning algorithm:

$$(x_1^{(k)}(t_-^{(k)}), \ldots, x_n^{(k)}(t_-^{(k)}),$$
$$x_1^{(k)}(t_-^{(k)} + \Delta t), \ldots, x_n^{(k)}(t_-^{(k)} + \Delta t), \ldots$$
$$x_1^{(k)}(t), \ldots, x_n^{(k)}(t), y^{(k)}(t + m))$$

for all moments t for which $t + m \le t_+^{(k)}$.

From the practical viewpoint, we face a challenge. From the purely theoretical viewpoint, the above approach should work. However, in practice, the computation time needed to apply a machine learning technique grows fast with the number of unknowns. In the prediction problem, as possible inputs, we can have each of n values x_i measured at each of N_t past moments of time – i.e., we need the dependence on $N_t \cdot n$ unknowns. When N_t is large, the resulting number of unknowns is large and thus, the corresponding computation require too much computation time. And indeed, successful predictions – e.g., the ones based on deep learning [8] – require state-of-the-art high-performance computers.

To overcome this challenge, we need to limit moments of time used for training. To make predictions practically possible – without the need to use difficult-to-access state-of-the-art supercomputers – we need to limit the number of inputs in the corresponding patterns. In other words, instead of considering values of the quantities x_i at *all* past moments

of time t, we should only consider values at *some* of these moments.

The fewer moments of time we consider, the faster we train the prediction algorithm. From this viewpoint, speeding up the prediction means minimizing the number of moments of time used for training.

Examples show that such a limitation is indeed possible. Suppose that we want to predict the weather in the next hour. In most cases, the weather does not change much during an hour. Thus, the most informative information comes from the current values $x(t_0)$. Knowing weather on the same day last year will probably not help much with our predictions.

To get predictions for the next day, it may be a good idea to also look for yesterday's weather, to see if there is a tendency for the temperature to increase or to decrease.

If we are currently in spring, then, to get predictions for the next summer, today's data is probably useless, it is much more useful to get data from last summer.

In general, the more into the future we want to predict, the further in the past is the information which is the most important for the corresponding prediction.

So, how can we limit? A natural way to limit is to divide the overall time period into sub-periods, find out which subperiod is the most important, and only use moments of time from this subperiod for training.

What we do in this paper. In this paper, on the example of predicting rare events – and specifically, on the example of predicting volcanic eruptions – we show how we can limit the number of moments of time and thus, speed up the corresponding computations. This possibility is related to the notions of chaos and delay.

II. Case Study: Predicting Rare Events Such As Volcanic Eruptions

Predicting rare events: an additional challenge and how to deal with it. In some cases, our goal is to predict rare events. In this case, for each moment t_f, the value $y(t_f)$ is simply 1 or 0:

- $y(t_f) = 1$ if the event occurs at moment t_f, and
- $y(t_f) = 0$ if the event did not occur at this moment.

The additional problem with this description is that since the events are rare, most of the values $y(t_f)$ are equal to 0 – and thus, most patterns do not carry much information.

One way to avoid this problem is to consider an alternative representation of the desired output: namely, instead of the original values $y(t_f)$, we can use cumulative values $Y(t_f)$ that describe whether the event has occurred between the current moment t_0 and the future moment t_f. In this case, if the event occurred at some moment $t_e > t_0$, then we will have:

- $Y(t_f) = 0$ for $t_f < t_e$ and
- $Y(t_f) = 1$ for $t_f > t_e$

and thus, many patterns will carry some information useful for predictions.

Similarly, if the prediction is based on rare events $x(t)$, i.e., on events for which $x_i(t) = 0$ for most moments of time t, it is useful to consider instead the cumulative values $X_i(t)$ which add up the values of $x_i(t')$ for all moments of time between t' and t_0.

In the case of predicting rare events based on rare observations, we need to find the dependence of the cumulative values $Y(t_f)$ on the cumulative values $X(t) = (X_1(t), \ldots, X_n(t))$ estimated at the previous moments of time $t \le t_0$.

Volcanic eruptions: a brief description of the case study. In this paper, we consider an important problem of predicting volcanic eruptions.

An unexpected eruption can be a big disaster:

- The ancient city of Pompei was destroyed by a nearby volcano.
- The Cretan civilization was destroyed by a tsunami caused by a volcanic eruption.

Nowadays, millions of people live in the close vicinity of active volcanos: it is sufficient to name Naples in Italy and Mexico City in Mexico. This makes the task of predicting volcanic eruptions even more critical.

What information we can use to predict volcanic eruptions. When magma ascends to the surface, this massive movement causes some seismic activity (see, e.g., [13]) – and eventually leads to ground deformation (see, e.g., [3]). Also, as magma rises to the surface, volcanic gases come out, so we can also see the changes in the amount and chemical composition of the volcanic gas emissions; see, e.g., [19].

All this can be, in principle, used to predict volcanic emissions. Out of these three sources, seismic activity is the easiest to detect and provides the most information. There are several reasons for this:

- It is well known that seismic waves – even relatively weak ones – easily propagate through long distances and are relatively easy to detect and to separate from noise. Thus, they can be detected even at stations at some distance from the volcano. In contrast, surface deformations and gas emissions data require complex on-site measurements which are rarely available.
- Based on the seismic observations, we can very accurately determine the location of the corresponding activity. In contrast, based on the deformations or emissions, we can only get the general idea of the volcanic activity, without getting any information about where exactly this activity takes place.

As a result, volcanic prediction techniques are based mostly on seismic activities – with other types of information used to improve the prediction accuracy.

There are some successes in volcanic prediction, but the situation is far from perfect. Due to the importance of the volcanic prediction problem, many techniques have been developed and tested for solving this problem. These methods range from purely statistical techniques (see, e.g., [2], [9], [17]) to technique based on machine learning; see, e.g., [6],

[10], [12] and references therein. In spite of all the successes, for all these techniques, predictions are still not perfect, more accurate (and more efficient) predictions methods are needed.

III. Which Time Period Should We Keep: Chaos and Delays

What can we do? At first glance, the more information we have, i.e., the more values $x(t)$ corresponding to different moments of time t we have, the better our predictions. And in some application areas, this is indeed the case: e.g., if we want to predict the trajectory of an asteroid or of a comet, the more observations we have, the more accurate are our predictions.

However, in many other cases, not all prior information is useful. The information from the distance past may not be very useful: these events happened so long ago that the resulting effects are negligible. On the other hand, the information about the most recent events $t \approx t_0$ may not be useful either, since these events may not have yet affected the desired result $y(t_f)$.

In such situations, we do not have to consider values $x(t)$ from all possible previous moments of time: values from a distance past and/or values which are most recent can be dismissed. This will decrease the amount of possible inputs and thus, make predictions more feasible.

Comment. At first glance, all this is common sense, but in practice, taking all this into account required a lot of efforts – and is still not always done in predictions. Let us deal with these two phenomena one by one.

Distant past is sometimes useless for predictions: a phenomenon known as chaos. A historically first experimental evidence that events from a distant past may be useless for predictions came from the work of a meteorologist and mathematician Edward Lorenz; see, e.g., [16]. He studied the possibility to use differential equations describing atmospheric phenomena to predict future weather (i.e., temperature, wind speed and direction, humidity, etc.) based on the results of the current and past measurements.

If we knew the exact values of today's and past quantities, then, in principle, we could solve the corresponding equations and come up with the future predictions. In practice, however, measurements are never absolutely accurate; see, e.g., [14]. No matter how accurately we measure, there are always different values of the initial conditions consistent with the measurement results. For example, if the result of measuring temperature is 20° C, with accuracy ± 1 degree, this means that the actual (unknown) temperature can take any value between 19° C and 21° C.

For solutions of the corresponding differential equation, different initial conditions leads, in general, to different future values of the corresponding quantities. Thus, if we take into account the measurement-related uncertainty wit which we know the initial conditions, then we can conclude that instead of the *exact* values of the future quantities, we can only predict *ranges* of possible values of these quantities.

Lorenz's discover was that for the different equations corresponding to meteorology, the width of the corresponding interval of future values exponentially increases with time. As a result, after a short period of time – about a week – the predicted range becomes so wide that it includes all physically possible values of the corresponding quantity. From the practical viewpoint, this means that longer-term predictions are simply not possible. This phenomenon is called *chaos*.

Originally, chaos was discovered in meteorological phenomena, but later research showed that chaos is really ubiquitous: chaotic phenomena have bene observed in many application areas; see, e.g., [7], [16].

Recent past is also sometimes useless for predictions: a phenomenon of delays. Another phenomenon that needs to be taken into account when we make predictions is the phenomenon of delays.

When we write and solve differential equations describing a natural phenomenon, we get an impression that any change in one of the inputs immediately leads to changes in the solution. However, this impression is sometimes false.

In real life, there is often a delay between the change in the inputs and its effect on the observed system. For example, differential equations describing the spread of an epidemic assume that the emergence of sick people from other geographical regions immediately leads to people in a given area starting feeling sick. In reality, for many diseases, there is an incubation period, during which an effected person feels quite well – and only after this period, people will start feeling sick.

If the incubation period is 2 weeks, then in order to predict how many new patients will go to the doctors tomorrow, it does not help to trace how many infected outsiders came to this region today – or even yesterday, or the day before yesterday. The only information that will help in our predictions is how many infected outsiders arrived 2 weeks ago.

With natural phenomena like volcanos, the delay is inevitable. For example, a sudden earthquake at a certain depth underneath a volcano means that the magma have started moving up. However, this motion is reasonably slow, it may take days or even weeks for the magma to reach the surface and thus, cause an eruption.

IV. How to Experimentally Detect the Presence of Chaos and Delays: Case of Rare Events

Need to experimentally detect the presence of chaos and delays. In some applications – namely, in situations like meteorology or epidemic studies, where successful predictions has been made for decades – we know whether there is chaos and/or delay, and we know the parameters characterizing the corresponding chaos and/or delay phenomena.

In other applications, however, where we are still looking for good prediction techniques – like in the problem of predicting volcanic eruptions – we do not know whether there is a significant chaos and/or delay phenomena – and, even when we have an impression that such phenomena are present, we do not have a good estimate of the corresponding characteristics. In such applications, we need to detect the presence of these

phenomena based on the experimental data – and we need to determine the characteristics of these phenomena based on this same experimental data.

How to experimentally detecting the presence of chaos and delays: main idea. In the absence of chaos and delay, as we have mentioned earlier:

- for short-term predictions, with $t_f \approx t_0$, out of all possible values $x(t)$, the most important values are the most recent values $x(t)$, i.e., the values corresponding to $t \approx t_0$;
- as we increase t_f, the moments t for which $x(t)$ is most important for predictions decreases; and, in principle, there is no limit to this decrease as t_f increases.

Chaos means that for some time duration T_c – after which predictions are impossible – inputs $x(t)$ with $t < t_f - T_c$ do not affect in our predictions. The only values which affect our predictions are values $x(t)$ with $t \geq t_f - T_c$.

Thus, for short-term predictions, with $t_f \approx t_0$, out of all possible values $x(t)$, the most important values are the values t close to $t_0 - T_d$.

Similarly, delay means that for some time duration T_d – the smallest delay time after which some effects can be observed – there is no need to consider inputs $x(t)$ with $t > t_f - T_d$. The only values to be considered in our predictions are values $x(t)$ with $t \leq t_f - T_d$. Thus, as we increase t_f, the moments t for which $x(t)$ is most important for predictions decreases – but it always stays above $t_f - T_c$.

In situations when both chaos and delay phenomena are present, it is therefore sufficient to only consider the values $x(t)$ for which $t_f - T_c \leq t \leq t_f - T_d$. Thus, in the presence of chaos and/or delay, the moments of time t for which predictions of $t(t_f)$ are most important change:

- for short-term predictions, with $t_f \approx t_0$, out of all possible values $x(t)$, the most important values are the values t close to $t_0 - T_d$;
- as we increase t_f, the moments t for which $x(t)$ is most important for predictions decreases – but it always stays above $t_f - T_c$.

Case of rare events. For the case of rare events, when we predict the cumulative value $Y(t_f)$ based on the cumulative inputs $X(t)$, we should observe a similar phenomenon:

- for short-term predictions, with $t_f \approx t_0$, out of all possible values $X(t)$, the most important values are the values t close to $t_0 - T_d$;
- as we increase t_f, the moments t for which $X(t)$ is most important for predictions decreases – but it always stays above $t_0 - T_c$.

Thus, to find the corresponding values T_d and T_c, it is important to find out for which i and m_0, the values $q \stackrel{\text{def}}{=} X_i(t_f - m_0)$ – or, alternatively, which combinations q of such values – are most important for predictions.

How to determine which input is the most important: look at the root of the decision tree. Volcanic eruptions are a particular case of a problem of predicting rare events, in which the prediction $Y(t_f)$ are binary: either the event occurred during the period between t_0 and t_f or it did not.

In this case, selecting the most important variable is what is actually happening at the top level of the decision tree analysis (see, e.g., [4]). In the decision tree approach, for each of the quantities $q = X_i(t_f - m_0)$ affecting the decision and for each possible threshold q_0, we consider how well the split of all the cases into cases with $q < q_0$ and $q \geq q_0$ helps to separate positive from negative situations. Then, we select the quantity q (i.e., the pair (i, m_0)) and the threshold q_0 that lead to the best separation. The corresponding quantity q is thus the one which is the most important for predictions.

How do we know which separation is better? A natural idea is to use a separation that maximally decreases the uncertainty. We start with a situation in which, out of N observations, we have N_+ positive ones and N_- negative ones. In other words, we have:

- the proportion
$$p_+ \stackrel{\text{def}}{=} \frac{N_+}{N}$$
of positive situations and
- the proportion
$$p_- \stackrel{\text{def}}{=} \frac{N_-}{N}$$
of negative situations.

The corresponding uncertainty can be described by the expected number of binary questions that we need to ask to determine $y(t_f)$, i.e., to determine whether the situation is positive or negative. This average number of binary question is known to be equal to the *entropy*

$$S = -p_+ \cdot \ln(p_+) - p_- \cdot \log(p_-);$$

see, e.g., [11], [15].

When we select the quantity q and the threshold q_0, we thus separate all the observations into the following two groups.

First is the group of all the observations for which $q < q_0$. We will denote the overall number of such observations by $N^<$. The proportion of such observations is $p^< \stackrel{\text{def}}{=} \frac{N^<}{N}$. Among these observations, we have:

- $N_+^<$ positive ones and
- $N_-^<$ negative ones,

with proportions $p_+^< \stackrel{\text{def}}{=} \frac{N_+^<}{N^<}$ and $p_-^< \stackrel{\text{def}}{=} \frac{N_-^<}{N^<}$, and uncertainty

$$S^< = -p_+^< \cdot \ln\left(p_+^<\right) - p_-^< \cdot \log\left(p_-^<\right).$$

Second is the group of all the observations for which $q \geq q_0$. We will denote the overall number of such observations by $N^\geq$. The proportion of such observations is $p^\geq \stackrel{\text{def}}{=} \frac{N^\geq}{N}$. Among these observations, we have:

- $N_+^\geq$ positive ones and
- $N_-^\geq$ negative ones,

with proportions $p_+^{\geq} \stackrel{\text{def}}{=} \frac{N_+^{\geq}}{N^{\geq}}$ and $p_-^{\geq} \stackrel{\text{def}}{=} \frac{N_-^{\geq}}{N^{\geq}}$, and uncertainty

$$S^{\geq} = -p_+^{\geq} \cdot \ln\left(p_+^{\geq}\right) - p_-^{\geq} \cdot \log\left(p_-^{\geq}\right).$$

So, to determine whether the situation is positive or negative:

- in $p^{<}$-th portion of the cases, we need to ask, on average, $S^{<}$ binary equations, and,
- in the $p^{\geq}$-th portion of the cases, we need to ask, on average, $S^{\geq}$ binary equations.

Thus, after the separation, the resulting uncertainty – i.e., the average number of binary questions needed to separate positive form negative situations – is equal to

$$S(q, q_0) \stackrel{\text{def}}{=} p^{<} \cdot S^{<} + p^{\geq} \cdot S^{\geq}.$$

The ideal case would be if when all the positive phenomena go into one of the two classes and all the negative phenomena go into another class – in this case, we already have the ideal separation, no further questions are needed. In general, the fewer questions we need to ask, the better the separation.

Thus, as a variable which is most important for prediction, we take the variable q for which, for some threshold q_0, the remaining uncertainty $S(q, q_0)$ is the smallest possible.

Resulting algorithm. If we only consider individual values $X_i(t)$, then, for each m, we do the following. For i from 1 to n and for each combination of values $m_0 \geq m$ and q_0, we find:

- the number of patterns $N_+^{\leq}$ for which $X_i(t - m_0) < q_0$ and $Y(y) = 1$,
- the the number of patterns $N_-^{\leq}$ for which $X_i(t - m_0) < q_0$ and $Y(y) = 0$,
- the number of patterns $N_+^{\geq}$ for which $X_i(t - m_0) \geq q_0$ and $Y(y) = 1$, and
- the number of patterns $N_-^{\geq}$ for which $X_i(t - m_0) \geq q_0$ and $Y(t) = 0$,

and compute the entropy $S(q, q_0) = S(X_i(t - m_0), q_0)$. We find the combination (i, m_0, q_0) for which the entropy is the largest, and conclude that the value $X_i(t - m_0)$ is the most important for predicting the value $Y(t_0 + m)$.

Similarly, we can deal with combinations q of values $X_i(t)$.

V. Applying the Above Idea to Volcanic Prediction

What data we used. In this study, we used the volcanic chain for which the most information is available: the Aleutian chain of volcanoes that reaches from Alaska to Russia [5]. Because of their location, silicate ash erupted from them into the atmosphere impacts air traffic across major flight paths in the Pacific; as a result, they are heavily monitored, with seismic sensors near almost all of them.

Specifically, we used the data about the following volcanos (in alphabetic order): Aniakchak, Augustine, Dutton, Fisher, Fourpeaked, Gareloi, Great Sitkin, Griggs, Hague, Iliamna, Kanga, Katmai, Kliuchef, Korovin, Little Sitkin, Mageik, Makushin, Martin, Novarupta, Okmok, Pavlof, Redoubt, Semisopochnoi, Shishaldin, Snowy, Spurr, Tanaga, Trident, Ugashik-Peulik, Ukinrek Maars, Veniaminof, Westdahl, and Wrangell.

The resulting information about these earthquakes was taken from the existing databases [1], [18]. Specifically, the information about the earthquake hypocenters magnitudes was taken from the following databases:

Year	URL
1994–1999	http://pubs.usgs.gov/of/2001/0189/
2000–2001	https://pubs.er.usgs.gov/publication/ofr02342
2002	http://pubs.usgs.gov/of/2003/0267/
2003	http://pubs.usgs.gov/of/2004/1234/
2004	http://pubs.usgs.gov/of/2005/1312/
2005	http://pubs.usgs.gov/of/2006/1264/
2006	http://pubs.usgs.gov/ds/326/
2007	http://pubs.usgs.gov/ds/367/
2008	http://pubs.usgs.gov/ds/467/
2011	http://pubs.usgs.gov/ds/730/
2012	http://pubs.usgs.gov/ds/789/
General	http://earthquake.usgs.gov/earthquakes/search/
General	http://www.ncedc.org/anss/catalog-search.html

TABLE I
Sources of earthquake information

Which variables we tried. Overall, we performed four experiments, in which we tried to predict the cumulative data $Y(t_f) = Y(t_0 + m)$ for $m = 7$, 15, 30, and 180 days.

To make each of these predictions, we used the cumulative earthquake values $X_i(t_0 - m_0)$ corresponding to four similar time periods $m_0 = 7$, 15, 30, and 180. For each of these time periods m_0, we used two types of data:

- the overall number of earthquakes that occurred in a certain zone in the vicinity of the given volcano during the given period of time (i.e., 7, 15, 30, or 180 days), and
- the sum of the magnitudes of all these volcanos.

For each type of data, we also used the differences between the average values over the given period and the average values over the previous period, to gauge to what extent the seismic activity has intensified. Specifically, we used the following three differences:

$$\frac{X(t_0 - 7)}{7} - \frac{X(t_0 - 15)}{15}; \quad \frac{X(t_0 - 15)}{15} - \frac{X(t_0 - 30)}{30};$$
$$\frac{X(t_0 - 30)}{30} - \frac{X(t_0 - 180)}{180}.$$

So, for each zone, and for each the two data types, we use 7 different values:

- 4 values corresponding to 4 time periods, and
- 3 values corresponding to the 3 differences.

Thus, for each zone, we considered $2 \times 7 = 14$ values.

The overall neighborhood of each volcano was divided into $3 \times 3 = 9$ zones:

- by the distance to the volcano: 0–2.5 km, 2.5–5 km, and 5–15 km; and
- by depth: 0–5 km, 5–15 km, and 15–30 km.

For each of these 9 zones, we had 14 variables, so the overall number of variables was $9 \times 14 = 126$.

What we expected in general: reminder. As we have mentioned earlier, we expected that:

- for predictions for the nearby moment $t_f \approx t_0$, the most important variables should be $X(t)$ corresponding to $t \approx t_0 - T_d$ ($t \approx t_0$ if there is no delay), and
- for predictions for the faraway moment $t_f \gg t_0$, the most important variables should be $X(t)$ corresponding to $t \approx t_0 - T_c$ ($t \ll t_0$ if there is no chaos).

This would enable us to find the desired values of T_c and T_d.

What we observed. In our experiments, in all 4 prediction problems – predictions for 7, 15, 30, and 180 days ahead – the most important variable corresponds to:

- the value $X_i(t_0 - 30)$ corresponding to the previous 30 days, and
- the zone which is the closest to the volcano and the shallowest, i.e., the zone corresponding to:
  - distance 0–2.5 km from the volcano, and
  - depth 0–5 km.

Discussion. We expected to see two different values $t - T_d$ and $t - T_c$ – corresponding to delay and to chaos, depending on whether we want short-term or long-term predictions. Surprisingly, we got the exact same value $T_c = T_d \approx 30$ in both cases. So, it looks like volcanic eruptions are an unusual phenomenon where the delay and the chaos periods are approximately the same.

As a result, only values $X(t)$ with $t \approx t_0 - 30$ should be taken into account, while more recent and more distant values $X(t)$ do not affect the prediction.

This is bad news and good news: It is bad news because it looks like that, due to the experimentally observed chaos effect with $T_c \approx 30$ days, we cannot predict volcanic eruptions further than 30 days in the future – at least not if we only use seismic data for this prediction.

It is good news since this means that in predicting volcanic eruptions, instead of all possible earthquakes, it is sufficient to consider only the earthquakes that occurred approximately 30 days ago – in the nearby vicinity of the volcano. This is what we are working on right now. Our preliminary results with several types of neural networks shows that the possibility to not consider values $X_i(t)$ for $t \neq t_f - 30$ drastically decreases the number of inputs and thus, enables us to speed up the computations – without lowering the prediction accuracy.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721, and by an award "UTEP and Prudential Actuarial Science Academy and Pipeline Initiative" from Prudential Foundation.

The authors are thankful to the anonymous referees for valuable suggestions.

REFERENCES

[1] Advanced National Seismic System (ANSS) Database, hosted by Northern California Earthquake Data Center, http://www.ncedc.org/anss/catalog-search.html

[2] M. S. Bebbington and W. Marzocchi, "Stochastic models for earthquake triggering of volcanic eruptions", *Journal of Geophysical Research*, 2011, Vol. 116, Paper B05204.

[3] J. Biggs, E. Y. Anthony, and C. J. Ebinger, "Multiple inflation and deflation events at Kenyan volcanoes", *East African Rift, Geology*, 2009, Vol. 37, pp. 979–982.

[4] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, 2006.

[5] K. F. Bull and H. Buurman, "An overview of the 2009 eruption of Redoubt Volcano, Alaska", *Journal of Volcanology and Geothermal Research*, 2013, Vol. 259, pp. 2–15.

[6] M. Curilem, F. Huenupan, C. San Martin, G. Fuentealba, C. Cardona, L. Franco, G. Acuña, and M. Chacón, "Feature analysis for the classification of volcanic seismic events using support vector machines", In: A. Gelbukh, F. C. Espinoza, and S.-N. Galicia-Haro (eds), *Nature-Inspired Computation and Machine Learning: Proceedings of MICAI'2014*, Springer Lecture Notes in Computer Science, Vol. 8857, 2014, pp. 160–171.

[7] J. Gleick, *Chaos: Making a New Science*, Penguin Books, New York, 2008.

[8] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Leaning*, MIT Press, Cambridge, Massachusetts, 2016.

[9] W. Marzocchi and M. S. Bebbington, "Probabilistic eruption forecasting at short and long time scales", *Bulletin of Volcanology*, 2012, Vol. 74, pp. 1777–1805.

[10] M. Masotti, S. Falsaperla, H. Langer, S. Spampinato, and R. Campanini, "Application of Support Vector Machine to the classification of volcanic tremor at Etna, Italy", *Geophysical Research Letters*, 2006, Vol. 33, Paper L20304.

[11] H. T. Nguyen, V. Kreinovich, B. Wu, and G. Xiang, *Computing Statistics under Interval and Fuzzy Uncertainty*, Springer Verlag, Berlin, Heidelberg, 2012.

[12] J. Parra, O. Fuentes, E. Anthony, and V. Kreinovich, "Use of machine learning to analyze and – hopefully – predict volcano activity", *Acta Politechnica Hungarica*, to appear.

[13] J. A. Power, S. D. Stihler, B. A. Chouet, M. M. Haney, and D. M. Ketner, "Seismic observations of Redoubt Volcano, Alaska – 1989-2010 and a conceptual model of the Redoubt magmatic system", *Journal of Volcanology and Geothermal Research*, 2013, Vol. 259, p. 14.

[14] S.G. Rabinovich, *Measurement Errors and Uncertainty: Theory and Practice*, Springer Verlag, Berlin, 2005.

[15] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Chicago, Illinois, 1971.

[16] S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*, Westview Press, Boulder, Colorado, 2014.

[17] R. Tonini, L. Sandri, D. Rouwet, C. Caudron, W. Marzocchi, and Suparjan, "A new Bayesian Event Tree tool to track and quantify volcanic unrest and its application to Kawah Ijen volcano", *Geochemistry, Geophysics, Geosystems*, 2016, Vol. 17, pp. 2539–2555.

[18] United States Geological Survey (USGS) website http://pubs.usgs.gov/

[19] C. Werner, P. J. Kelly, M. Doukas, T. Lopez, M. Pfeffer, R. McGimsey, and C. Neal, "Degassing of CO_2, SO_2, and H_2S associated with the 2009 eruption of Redoubt Volcano, Alaska", *Journal of Volcanology and Geothermal Research*, 2013, Vol. 259, pp. 270–284.