ACCEPTED MANUSCRIPT
Originally published in:
European Journal for Security Research 143(11) (2019).
https://doi.org/10.1007/s41125-019-00061-w
The original publication is available at https://link.springer.com/article/10.1007%2Fs41125-019-00061-w

Aishvarya Kumar Jain[1], Christian Grumber[1], Patrick Gelhausen[1], Ivo Häring[1], Alexander Stolz[1]

A Toy Model Study for Long-Term Terror Event Time Series Prediction with CNN

aishvarya.kumar.jain@emi.fraunhofer.de
[1] Fraunhofer Institute for High-Speed Dynamics, Ernst-Mach-Institut, Am Klingelberg 1, 79588 Efringen-Kirchen, Germany

Abstract

Followed by the 9/11 attacks in 2001 and the subsequent events, terrorism and other asymmetrical threat situations became increasingly important for security-related efforts of most western societies. In a similar period, the development of data gathering and analysis techniques especially using the methods of machine learning has made rapid progress. Aiming to utilize this development, this paper employs artificial neural networks (ANN) for long-term time series prediction of terrorist event data. A major focus of the paper lies on the specific use of convolutional neural networks (CNN) for this task, as well as the comparison to the performance of classical methods for (long-term) time series prediction. As the database like Global terrorist database (GTD) and Fraunhofer's terrorist event database (TED) are not extensive enough to train a deep learning method, a simple toy model for the generation of time-series data from one or more terrorist groups with defined properties is established. Metrics for comparison of the different approaches are collected and discussed, and a customized sliding window metric (SWM) is introduced. The study shows the principle applicability of CNNs for this task and offers constraints as well as possible extensions for future studies. Based on these results, continuation and further extension of data collection efforts and ML optimization techniques are encouraged.

Keywords: multi-step, long-term, time series prediction, artificial neural network, convolutional network, terrorism

1 Introduction

The dynamics of conflicts have changed during the course of the last century. While the security situation, especially in western societies, traditionally was dominated by scenarios of potential or actual symmetrical warfare, modern conflicts often evolve around asymmetric situations. This development has been emphasized and further accelerated by the infamous 9/11 attacks and the subsequent events. Consequently, terrorism has become a topic of increasing importance for considerations on governmental, economic and societal levels.

A major characteristic of terror is its unpredictability. To unfold its negative effects, it needs to generate a ubiquitous danger to health and material assets. This yields a strategy that actively tries to hide patterns or early warnings. Naturally to prevent or mitigate the impact, identifying such patterns, i.e. via time series analysis, ever since has become a field of high interest for science and authorities. A huge variety of instruments in classical data science and statistics has been employed with varying success (Chen and Chen 2015; Wei 2016; Bista et al. 2017; Lee et al. 2018).

The era of information provides new possibilities in this context by adding two new developments to the field: On the one hand, improved data collection and recording capabilities boost the amount and quality of available information. On the other hand, increasing computational power enables algorithms to handle emerging complexity. Combining these two aspects, Machine Learning (ML) methods gained large attention in the field of Data Science and already proved their advantages in many different fields of applications including speech-driven assistants[1].

However, when it comes to terror event prediction (Snehanshu Saha et al.), thorough data collections have just started to be established in the last decades. Databases include the Global Terrorism Database (GTD 2018), Fraunhofer's terrorist event database (TED) and a comprehensive list of other databases (Bowie 2018) covering different regions, intensities, and timeframes of terror attacks around the globe. Yet a centralized and normalized collection with a strong level of detail is missing that considers not only terror (related) events.

When it comes to time series analysis and event prediction, usually short-term predictions, e.g. stock market developments, gain the most attention. As time passes, the information is added directly and the next short span is predicted. However, for some applications, it might be useful to make a long-term prediction over several weeks or even months without adding information after the initial prediction. One might consider for example long-term planning for resource deployment on an organizational level, which cannot be adjusted on a daily or even weekly basis.

Addressing this gap, this paper aims to study the potential use of artificial neural networks (ANN), specifically convolutional neural networks (CNN) for long-term terror event prediction to enable strategic planning for counterterrorism measures on a large scale. The results are then compared exemplarily with the performance of classical statistical time series analysis methods. Additionally, the performance of the CNN is also compared to that of a recurrent neural network (RNN), which currently is the default choice for time series prediction with ANN (Akram and El 2016; Shi et al. 2018; Madan and SarathiMangipudi 2018; Sak et al. 2015).

This article tries to prove the principal competitiveness of the proposed approach to classical methods if a sufficient amount and quality of data are provided. This is realized by using data generated by a toy model, which is designed to mimic the plausible behavior of terror groups, as available real event data at this point is not yet sufficient for supporting an ANN training. Therefore, this study should not only prove the principle feasibility of the approach but also encourage the relevant organizations to further continue and extend the current processes of data recording and collection in the future. Future studies will be performed to advise the more concrete dataset preparation strategies.

The structure of the paper is as follows: After this introduction, Section 2 discusses some classical methods for long-term time series prediction and introduces the two selected ANN methods for later comparison. Section 3 sets up the toy model, where an artificial time series is created from probabilistic properties. Section 4 introduces evaluation metrics and classical methods for long-term predictions, the ANN methods are applied, and the obtained results are collected and discussed. The paper closes with a summary and outlook in Section 5.

2 ANN methods for long-term time series prediction

This section explains the methodology used in the time series prediction and gives a short overview of the concept of neural networks with the main focus on the use of convolutional neural networks (CNN) for time series forecasting.

[1] https://developer.amazon.com/alexa and https://www.google.com/intl/de/landing/now/

Time series prediction is an extensively studied task and is applied in countless fields, which have already been mentioned in Section 1. In the domain of terrorist attack prediction, time series analysis has already been applied before (Sheehan 2009; Li et al. 2017). The current work primarily discusses the application of deep learning for time series prediction of terrorist attacks.

As discussed in Section 1, this article focuses on the application area of long-term predictions for terror events. That is to predict the value for an extended period as a whole, rather than doing the short-term prediction and updating it regularly after the initial prediction using additionally available data or already predicted data. Another requirement is the feasibility to include multiple features (i.e. using a multivariate model), as a terrorist event database can assume a very complex structure and might have several relevant entries beyond terror events only depending on various other features.

As can be seen in the example of Figure 1, the data until the point of "Today" is considered as available. Using the information of the last 100 days as an input, the method is tasked to predict for instance the "Number of events" for an extended period. In this case, exemplarily a period of 30 days is chosen, because this corresponds to a real-lifetime span that allows officials to prepare countermeasures.

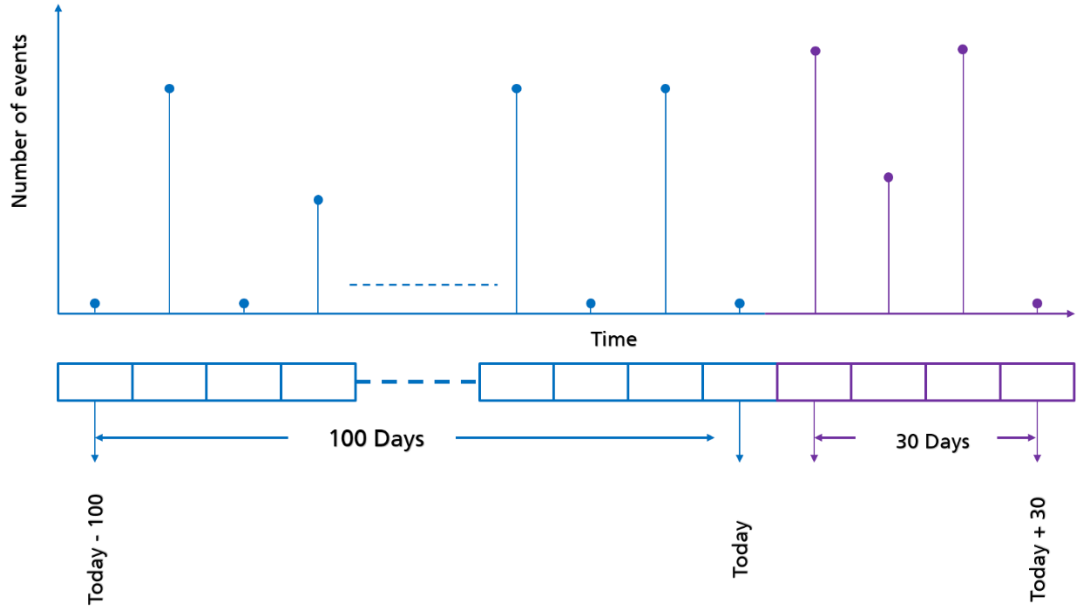

Figure 1: Illustration of long-term time series prediction. The blue time series is used as input, the purple curve is predicted as a whole.

Correspondingly, several classical statistical forecasting methods are available and employed in various application areas. Some of these methods include (Robert Nau):

- Naive estimators (using the average, the most probable value or the last value),
- Averaging and smoothing (e.g. exponential smoothing),
- Linear regression, and
- Autoregressive integrated moving average models (ARIMA).

The classical methods are very effective when small amounts of data are available. The world of today, however, becomes increasingly data-driven, following ever-improving capabilities for data recording and storage. Having access to huge amounts of data, deep learning methods have shown remarkable performance in the field of time series prediction. Currently, recurrent neural networks (RNN) are the state of the art for time series prediction (Shi et al. 2018; Madan and SarathiMangipudi 2018; Sak et al. 2015). However, according to the recent research in this field, CNNs are also very well suited for this task (Borovykh et al. 2017).

In spite of having existing techniques for time series prediction, none of them is a true long-term time series predictor. For instance, classical time series can be used to predict the event on the next day based on the inputs of the previous 100 days and that predicted value can be assumed to be the same for the next 30 days correspondingly making it a pseudo-long-term predictor (Figure 1).

The available data can be multivariate or univariate. RNN can be modified for usage in long-term time series prediction if univariate data is used (Shi et al. 2018; Zhang and Xiao), but for doing long-term time series prediction on multivariate data, a unified RNN approach is missing. Correspondingly, a new method is devised to do the long-term prediction using CNN, which can easily be adapted for multivariate data. Traditionally, CNNs are extensively used for image recognition and analysis, but with little modification, they can also be employed for time series prediction (Borovykh et al. 2017).

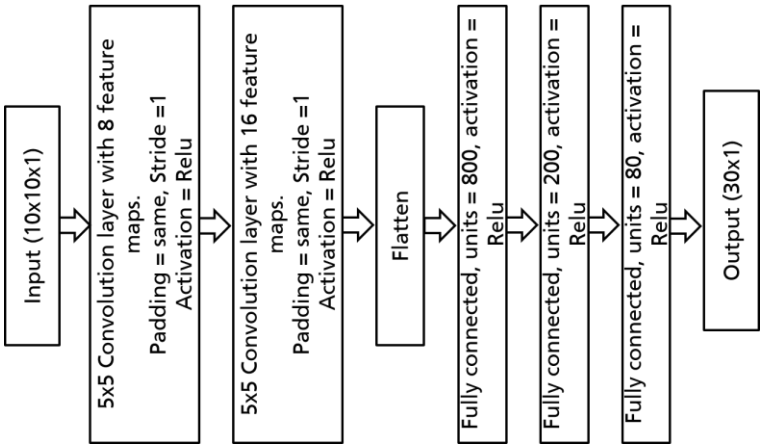

Figure 2: Architecture of the convolutional neural network (CNN) used.

For illustration purposes, the CNN architecture of this study is shown in Figure 2. Correspondingly, Figure 3 illustrates the shape of data in the successive stages of the CNN network. As illustrated in Figure 3, the information of "Number of events" which is here assumed to be the number of casualties (each day) for the past 100 days, is fed as input. This information is then reshaped into a 2D array of dimension 10 times 10 (10x10). Following this, two convolution operations are performed to increase the feature space as the two intermediate outputs contain 8 and 16 feature maps (Bengio and Lecun 1997).

This convolution output is then flattened into a 1D array of 1600 units and fed to a fully connected neural network with three hidden layers having 800, 200 and 80 units respectively. The output layer contains 30 units, which represent e.g. "Number of events" or "Number of casualties" in the next 30 days. Throughout, the network uses the rectified linear unit (ReLU) activation function.

During the training, the task is to minimize the mean square error between the number of events predicted by the network for 30 days and the actual number of events in this period.

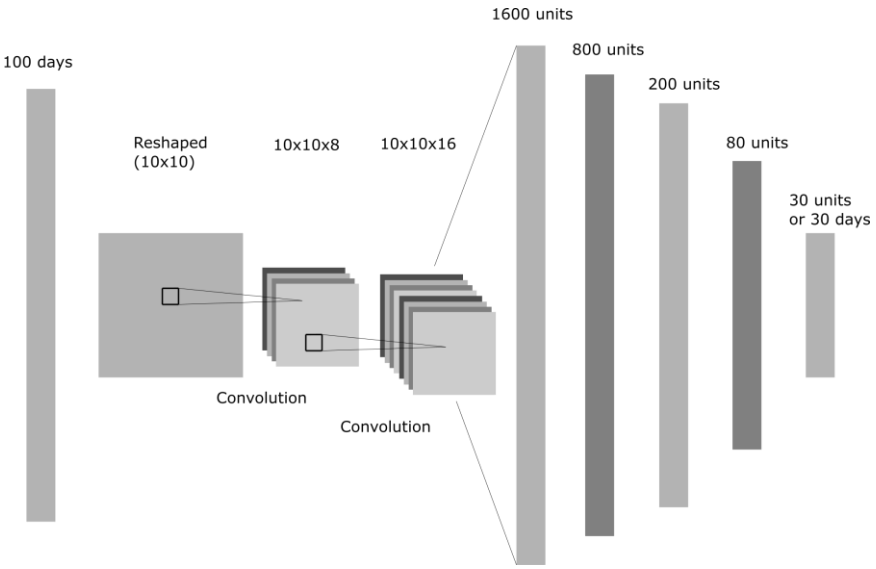

Figure 3: Reshaping the input data and architecture of the CNN used.

Reshaping the 1D array of attacks for 100 days in 2D as shown in Figure 4 ensures that convolution kernels have a wider coverage of the attack dependency. As can be seen in Figure 4, a 1D kernel centered at T_2 with a kernel width of three can only include the information from T_1 to T_3 while a 2D kernel of dimension 3x3 and centered at T_{13} not only includes the information from T_{12} to T_{14} but also from T_2 to T_4 and T_{22} to T_{24}. Therefore, the CNN is not only able to include patterns which occurred over the last few days but is also able to scan for periodic patterns over a longer time span. For example, for the prediction of attacks on a Sunday, the information what happened last Sunday could be as relevant as the information what happened the day right before.

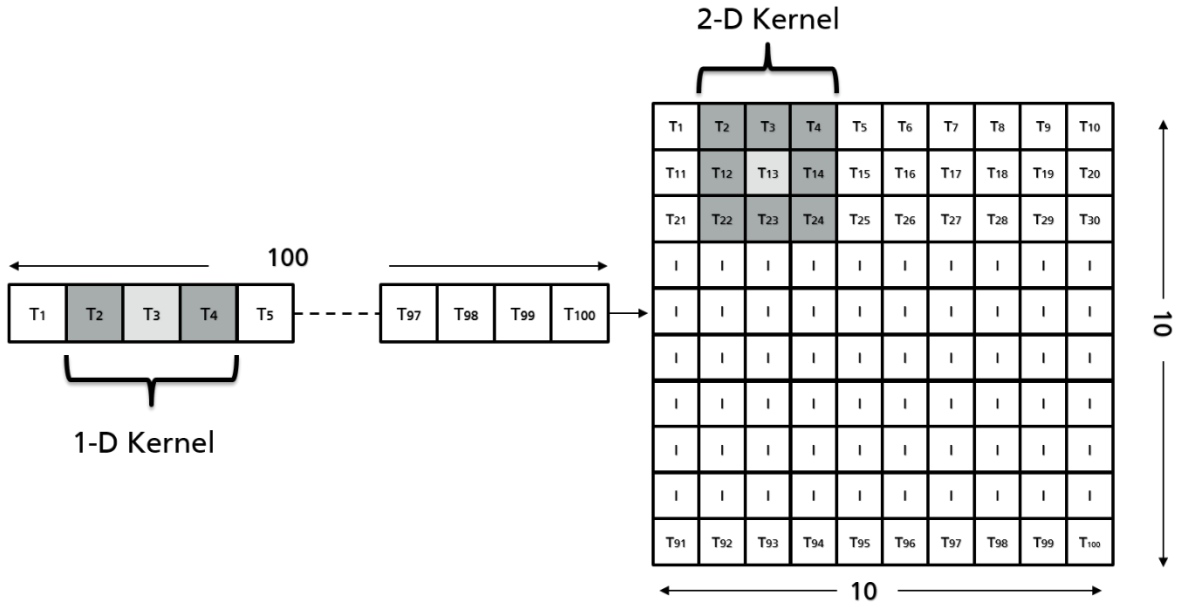

Figure 4: Modeling the attacks in 2D and coverage of the 1D kernel and 2D kernel.

For illustration purposes, Figure 5 shows the performance of both RNN and CNN for a time series generated using a simple exponential sinusoidal chirp. CNN has the same architecture as described in Figure 2. For RNN a sequence to sequence architecture is used which is built using a single-layered LSTM cell with 30 units, and a time span of 100 units. It can be seen that CNN provides a reasonable result on this task, further encouraging the application to more complex data. It should also be noted that the CNN and RNN architectures for Figure 5 are not tuned, respectively, and can produce even better results. A major benefit of CNNs compared to RNNs is the capability to include multiple features, which can be mapped to various channels of a CNN network. This especially applies to the field of long-term time series prediction, where the time series data can assume a very complex structure, which could be aperiodic.

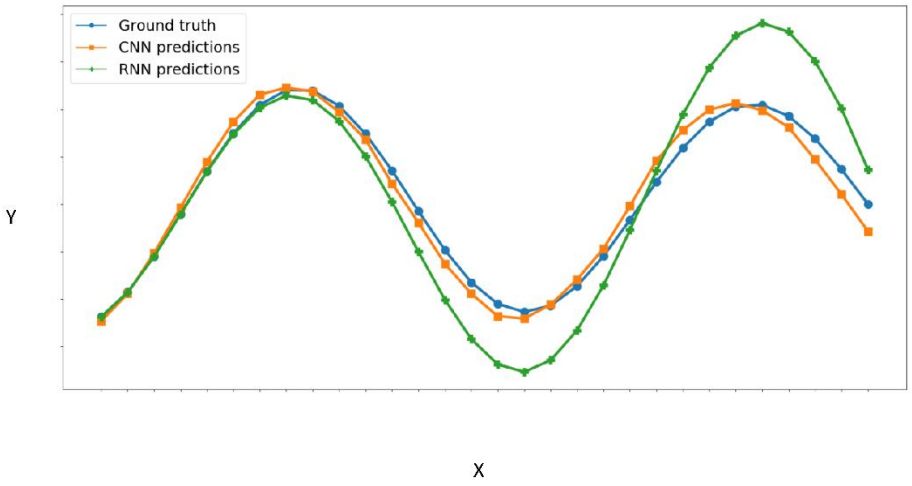

Figure 5: Comparison of long-term prediction for an exponential sinusoidal chirp using RNN (green) and CNN (orange). The actual sinusoidal signal ("ground truth") is depicted with a blue line.

3 Probabilistic toy model setup

As discussed in Section 1 and 2, the current amount of terror (related) event data in most country cases is not sufficient to train an ANN/CNN with competitive prediction accuracy. Still, for future applications and recommendations, it is reasonable to explore the potential and the limitations of ANNs in comparison to classical methods.

For this purpose, in this section, a probabilistic toy model is established for further study. Although the toy model is of very limited fashion, its adjustable properties also allow studies that are more complex in the future. It should, however, be mentioned, that this model should only illustrate the feasibility of the ANN approach – it is therefore reduced to mathematical simplicity. It can and does not claim to be accurate mapping of actual terror processes – for this task real analytical and sociological models available in the research field should be employed (Serra and Subrahmanian 2014).

3.1 General toy model properties

The foundation of the model is the stochastic behavior of one or more imaginary terror groups, executing attacks according to a probabilistic model in an imaginary country, thus generating a time series of terror attacks in a specified span of time.

To be used for testing, the model should allow for parameter-controlled patterns covering a large variety of scenarios for study. Furthermore, its complexity should be controllable and adjustable to study the relationship between the degree of complexity of the data and the performance of the ANN.

In the course of this paper, three pattern types are included as a starting point:

- **Stochastic noise:** Randomized noise is added to the data to include single terror attacks from actors, which are not associated with the actual group or groups (“Lone Wolfs”). While the behavior of the terror group determines the actual pattern of the data, these single attacks somehow “pollute” this signal. The ANN should be able to distinguish this noise from the actual signal. Such noise can also be due to other empirical data acquisition errors or other forms of attacks attributed to terrorism.
- **Constant interval patterns:** Often, terror groups show some preference towards certain dates for their attacks. This can refer to an especially suitable opportunity and therefore a higher attack probability, e.g. a holiday where the expected impact maximizes. Alternatively, it can refer to an unfavorable setting for attacks reducing their probability, where for example routinely increased security efforts hinder attacks. If such special points in time are periodically recurring, the ANN should be able to recover the period of this pattern. Examples include workdays, public holidays and religiously determined dates.
- **Variable interval patterns:** Sometimes a pattern is recurring, maintaining its principle shape, but changing some properties. For example, a sine wave could change amplitude and period, yet the pattern would remain the same. In the same sense, a recurring pattern in the time series may occur. For example, this can reflect a daily increasing resource pool of the terror organization, with the attack probability correlating with the amount of available resources. The ANN should be able to recover the average period of this pattern.

Already by combining and varying the defining parameters of these patterns, a huge range of scenarios can be generated. The studies in this paper, however, will be limited to the analysis of a time series created from a simple pattern. Future studies might follow taking up the proposed modeling framework.

3.2 Specific time series construction from the toy model

After introducing the principle structure of the toy model, this section illustrates the construction of an explicit time series for evaluation.

The series is constructed through probabilistic principles. For every day, the number of terroristic events is sampled through predefined attack probabilities. As discussed in Section 3.1, the model distinguishes between “noise” (“Lone Wolf” attacks) and actual “signal” (group attacks). Therefore, every day separately the attack probability for the noise P_L is calculated, as well as the attack probability P_G for a group attack. In principle, any number of active groups can be modeled, yet for reasons of simplicity, only one active group is assumed.

Thus, a unique time series is created for every scenario and can also be statistically evaluated accordingly. The terror model pattern types explained in Section 3.1 are implemented in the following fashion:

Stochastic noise: A fixed daily probability P_L for an attack is attributed to ”Lone Wolf” attackers which for the mentioned case is 0.01. If an attack occurs, the number of casualties is determined by a probability distribution PD_L. In the case at hand, PD_L is simply an array of discrete probability values as shown in Figure 6. The amplitude, as well as the frequency of the stochastic noise, can be manipulated via the corresponding sets of parameters.

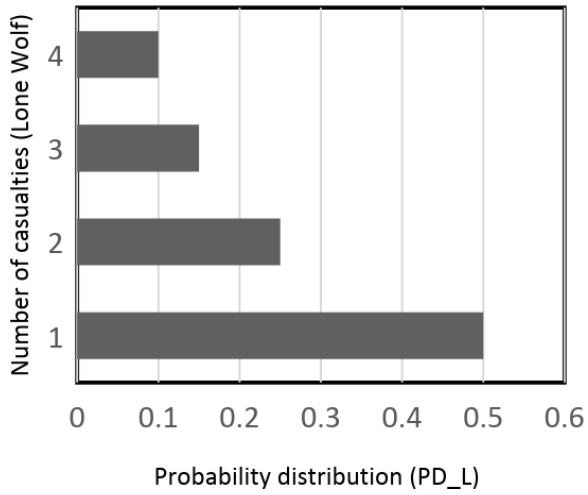

Figure 6: Sample probability distribution of the casualties created by the "Lone Wolf" attacks. The probability to create one casualty is highest and to create four is the lowest.

Constant interval patterns: The periodically recurring attack pattern is implemented with an array of fixed values peaking around a monthly fixed day. This monthly fixed day can, for example, correspond to a holiday. For each day, a probability P_R is selected from the array, depending on its distance to the peaked day (see Figure 7). Note, that the sign of these probability distributions could also be chosen as negative in a sense of a suppressing factor in the combined probability P_G.

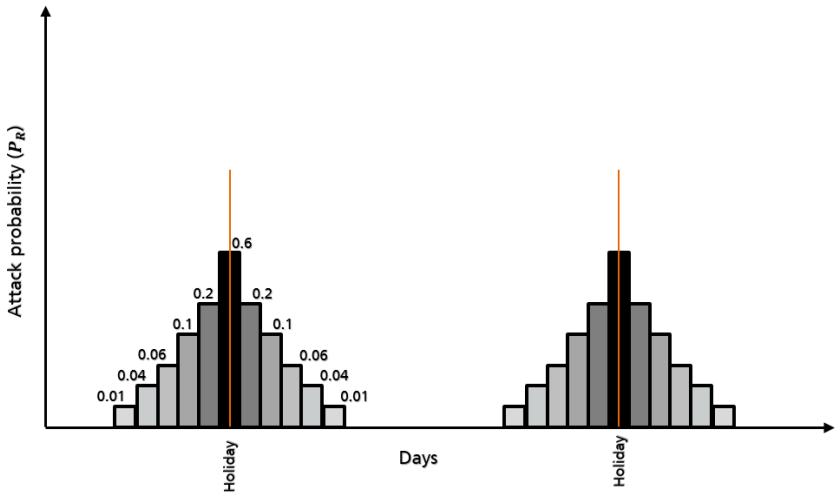

Figure 7: Attack probability distribution PD_H around the holidays. In this picture, the same distribution for recurring holidays is assumed.

Variable interval patterns: The variable interval pattern is governed by the parameters of the resource pool. The daily attack probability is enhanced by a pool value P_{Pool}, which directly corresponds to the amount of available resources u:

$$P_{Pool} = (u - u_{thres}) \cdot F_{Pool}, \tag{1}$$

with the threshold parameter u_{thres} and an amplitude parameter F_{Pool}. This implicates that the terror group requires a minimum amount of resources u_{thres}, otherwise P_{Pool} is negative and acts as a suppressive factor to the overall probability. The amplitude parameter just scales the relation between the amount of resources and attack probabilities.

Besides the explicit parameters in Eq. (1), also the daily increase of the resource pool Δu is used to control the shape of the data. The decrease of the resource pool after an attack is handled by subtracting the total cost of the attack. This cost is assumed to correspond to the amount of inflicted damage. It is calculated by multiplying the

total number of casualties in that attack by a constant impact factor C_t. Having a high number of casualties introduces a "cool down" period until the resource pool is replenished. The corresponding quantities are collected in Table 1. Within the sample toy model, the parameters are chosen such that a rather large number of terror events is generated, which also can be interpreted as terror-related events. Correspondingly, the casualties generated can be interpreted as terror-related quantifiable issues related to certain terror events. An alternative is to consider the model as an example of modeling a certain well-defined type of terror groups in a certain region.

Table 1: Collection of parameters controlling the resource pool pattern.

Resource pool value (P_{Pool})	Eq. (1)
Daily deposit in the resource pool (Δu)	500
Threshold parameter (u_{thres})	5000
Amplitude parameter (F_{Pool})	3/100000
Impact factor (C_t)	100

Following this collection of partial probabilities, the total group attack probability for each day is now defined as

$$P_G = \sum_i P_i\,, \qquad i = \{R, Pool\}, \tag{2}$$

Eq. (2) is written in a sum to indicate, that this model can be expanded with additional effects to customize a terror group. For illustration, in the present sample case $P_R = P_H$ as given by PD_H of Figure 7 and P_{Pool} as described along with Eqn. (1) and Table 1.

In the next step, the casualties for the group attack events are calculated. The group is assumed to have a much higher damage potential than the "Lone Wolf" attacks (see above), as it is more organized. The amount of inflicted casualties N_G is determined by drawing a random uniformly distributed sample from the range of 20 to 50. The sampled number of casualties is than used to calculate the decrease of the resource pool.

It should be noted that to keep the calculation simple and understandable, very simple distributions are used, but to build a realistic scenario much more complex distributions could be used.

Figure 8 gives an impression of the emerging time series by showing for one month the number of inflicted casualties.

The dataset generation is now ready for training the two ANN time series approaches and the application of classical time series approaches and their evaluation.

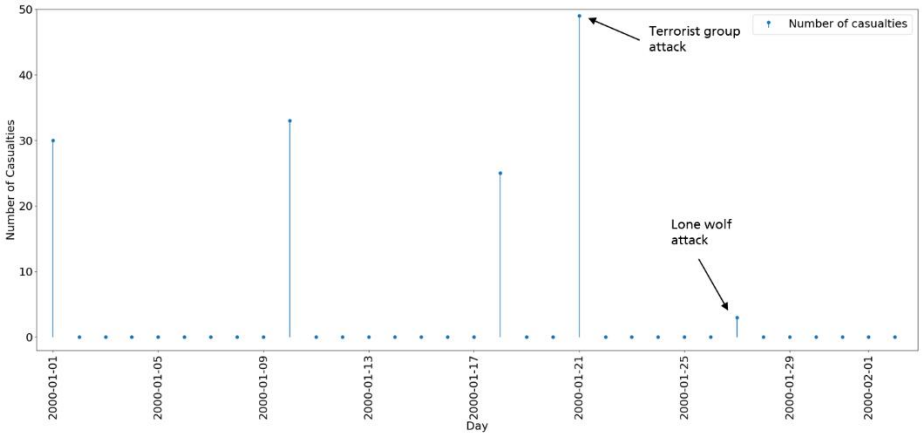

Figure 8: Generated time series over a duration of approximately one month using the terrorism model.

4 Model evaluation and results

In this section, the generated data from the toy model established in Section 3 is evaluated. First, evaluation metrics are presented in Section 4.1, followed by a list of common analysis methods for comparison in Section 4.2. The Section closes with a short discussion of the results in Section 4.3.

4.1 Evaluation metrics

The toy model is evaluated concerning different information types:

- Regression: The model is trained to predict the attack time and inflicted casualties. The result is a continuous output. Similarly, the number of attacks or the total number of persons affected could be computed.
- Classification: The model is trained to predict only whether there will be an attack or not (regardless of its intensity). The output, therefore, is binary. Similarly, it could e.g. be computed whether there are fatalities or not.

For regression tasks, two different loss metric types are adapted. The normed loss gives the usually summed deviation of the prediction from the actual value, referring to an L2-norm. With a prediction span of 30 days as used in this study, this is defined as

$$l_{norm} = \left(\sum_{i=1}^{30} (L_i - X_i)^2 \right)^{1/2} . \tag{3}$$

On the other hand, one can also evaluate the prediction according to an integrated loss, defined as

$$l_{int} = abs\left(\sum_{i=1}^{30} L_i - \sum_{i=1}^{30} X_i \right). \tag{4}$$

In Eq. (3) and (4), L_i marks the actual value of the i-th data point ("Label"), and the prediction is given with X_i. As can be seen in the definition of Eq. (3) and Eq. (4) for a good predictor, a small loss is desired. While the normed loss in Eq. (3) is sensitive to daily deviations of the prediction, the integrated loss of Eq. (4) offers a metric for a wider picture: The prediction obtains a good scoring if the overall number of casualties in the period at hand is equal to the actual data. In this sense, the integrated loss describes an overall threat situation for the time-span, while the normed loss covers a daily resolution.

The foundations of the classification metrics are the values of the confusion matrix, which lists the number of correctly and incorrectly classified events on both, positive and negative sides. The structure of the confusion matrix is given in Figure 9.

		real value	
		positive	negative
predicted value	positive	TP (true positive)	FP (false positive)
	negative	FN (false negative)	TN (true negative)

Figure 9: Schematic representation of a confusion matrix used for classification tasks.

The importance of the selected quantities varies according to the field of application. They may be combined to establish new metrics matching these requirements. Popular examples, which are also used in this paper, are the *recall* (also: *true positive rate TPR*) and the *precision* (also: *positive predictive value PPV*), which for ideal results should approach 1:

$$TPR = \frac{TP}{TP + FN}, \tag{5}$$

and

$$PPV = \frac{TP}{TP + FP}, \tag{6}$$

with the number of correctly predicted positive events TP (true positive), the number of misclassified positive events FP (false positive) and the number of misclassified negative events FN (false negative).

For terror event prediction, one would usually prefer an algorithm with a high TPR (see Eq. (5)), as one wishes to correctly identify upcoming attacks (TP) as well as avoiding misclassifying an actual attack (FN). However, the daily evaluation regarding these metrics would not be reasonable, given the natural uncertainty of the events.

Therefore, additional to this classic evaluation metrics a newly defined *sliding-window metric* (SWM) is adopted in this paper. The reason for this is motivated by the practical applications of the predictors: If the algorithm predicts an attack on a certain day, but the real attack occurs one day later, in terms of classification metrics this situation will obtain a poor scoring. Yet, in the context of the natural uncertainty of terror events, a one-day-off prediction would be a solid result for any long-term planning – one would rather be interested in the prediction of a *threat situation*, rather than exactly predicting the correct day. This threat situation might have a span of a few days.

For the regression task, the integrated loss is somehow able to cover this by accumulating the numbers over the whole month. Yet, this would only allow covering a monthly threat situation. For this reason, a certain window with the corresponding width w for scoring is established. The current work does not include the sensitivity study for the window size.

This SWM could assess one specific threat situation within these days, see Figure 10. The window is moved with a step size s throughout the whole month and every window obtains a score. In this study, s is equal to one and w is chosen as one, two and five. It should be noted, that this, of course, introduced a kind of double counting if $s < w$. But as this metric is only used for the comparison of methods, this is not an issue. As a metric in the sliding window the integrated loss of Eq. (4) is applied, resulting in the SWM loss metric

$$l_{win} = \frac{1}{N_{win}} \left(\sum_{N_{win}} abs \left(\sum_{i=win_{start}}^{w_{end}} L_i - \sum_{i=win_{start}}^{w_{end}} X_i \right) \right) \tag{7}$$

Note, that the sliding window loss metric l_{win} is defined not only for one window but as the sum over all windows (in total N_{win} windows) after sliding through the whole month. Additionally, the SWM allows for the use of the introduced classification metrics in the context of threat situations, i.e. for every window, an attack is predicted if the number of casualties in that window is greater than one. Based on these evaluated values, precision and recall are calculated.

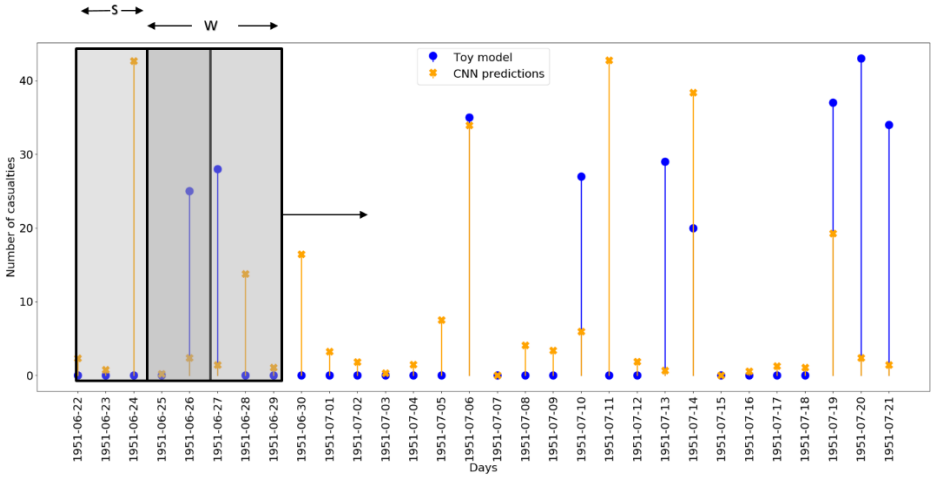

Figure 10: The sliding-window metric (SWM) for time series analysis. It is defined by the window width w and the sliding parameter s.

4.2 Analysis methods for comparison

This paper aims to discuss the feasibility of ANN methods for terror event time series prediction. Two types of ANNs are used: CNN and the RNN (see Section 2).

However, for comparison also "classical" prediction methods need to be addressed. In this study, only long-term predictions without regular updates are covered. Usual analytical methods like exponential smoothing or ARIMA

can in principle be applied, but as they rely on the input of close data points to predict the next one, their performance for long-term prediction is not reliable. Instead, naive predictors have been chosen for scoring and comparison.

In this context, two methods have been selected:

- Averaging: In terms of long-term prediction, a simple and efficient way to obtain a good result is to take the average of all previous data and set it as a default prediction for the future. Consequently, this only applies to regression tasks.
- Last year scaling: Another simple prediction method is to predict a value from a similar previous time-span, assuming some periodic nature of the data. In this case, the value from the same day of the previous year is selected. This predictor holds for both regression and classification tasks.

This in total amounts to four methods to be compared regarding their performance, two ANN and two classical ones.

4.3 Discussion of results

The training data is generated with the terrorism toy model of section 3.2 for a total of 47 years. For training 100 days are chosen as input and the next 30 days as the prediction span. For the scoring and comparison, for every method, a set of 100 prediction spans is randomly chosen and averaged over.

The results for the different methods are summarized in Table 2. Figure 11 shows one single prediction span for CNN and RNN as illustration.

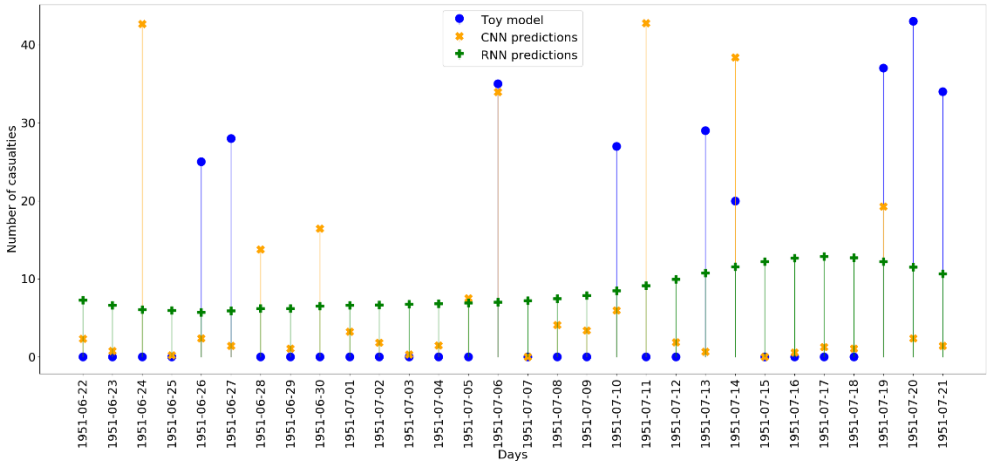

Figure 11: Predicted number of casualties by CNN and RNN methods for data generated by the toy model.

Table 2: Evaluation results for the different methods regarding the introduced metrics. The window size is only relevant for the classification metrics and the SWM.

Metrics / Method	Normed loss	Integrated loss	TPR (Recall)	PPV (Precision)	SWM integrated loss
$w = 1$					
Average	87.38	55.99	1.00	0.28	12.99
Last year scaling	119.33	46.11	0.64	0.29	13.28
RNN	85.94	32.28	1.00	0.28	12.96
CNN	111.52	53.59	0.84	0.28	13.05
$w = 2$					
Average	87.38	55.99	1.00	0.47	20.09

Last year scaling	119.33	46.11	0.74	0.48	22.03
RNN	85.94	32.28	1.00	0.47	19.33
CNN	111.52	53.59	0.96	0.47	21.66
$w = 5$					
Average	87.38	55.99	1.00	0.81	29.84
Last year scaling	119.33	46.11	0.91	0.81	35.85
RNN	85.94	32.28	1.00	0.80	27.77
CNN	111.52	53.59	1.00	0.80	36.12

For the loss metrics, the ANN methods achieve comparable or better results than the naive predictors (see Table 2). In the case of integrated loss metrics, in particular, averaging procedures naturally achieve a good scoring, while performing badly with the normed loss metrics. Regarding the periodic recurring estimate, the naive predictor achieves overall comparable scorings, in the integrated loss even outperforming CNN. This is however expected, as the data of the toy model was constructed with a strong periodic pattern (see Section 3). A global trend throughout several years (e.g. a steady rise due to growing structures or a change of attack periods) would have a heavy impact on the periodic recurring estimator, while the ANNs in principle can include it.

The confusion matrix related metrics within the SWM (see Section 4.1) show a similar picture: Overall, the quantities in terms of recall and precision are of similar size (see the fourth and fifth column of Table 2). There are still some differences: Regarding the recall, RNN and averaging procedures outscore both periodic recurring estimator and the CNN. This again is a feature of the dataset which is used for the sake of simplicity and should not be regarded as a general property of the methods. As the RNN generates a smooth and a rather flat line without updated daily input, it works similarly as the averaging estimator. That means that in the context of the SWM, always a threat situation is predicted – obviously detecting all actual events. On the other hand, this yields a high rate of false positives. For this reason, CNN and periodic recurring estimator display a significantly better accuracy for daily comparison (w=1 in Table 2), with this effect fading for larger window sizes. The reason for the latter is the high number of attacks in the data set – using a set with rare events will reproduce a similar result even for larger windows.

Indeed, when comparing the SWM matrix on a dataset with rare events, the performance shifts towards CNN as they are efficient to predict the attacks with much precision within a window which has a higher probability of getting an attack. The corresponding comparison is shown in Figure 12 where the performance indicator is the ratio of SWM for RNN to SWM for CNN. The datasets are arranged in the decreasing order of monthly attacks. It is visible in the figure that the performance of CNN increases as the number of monthly attacks decreases.

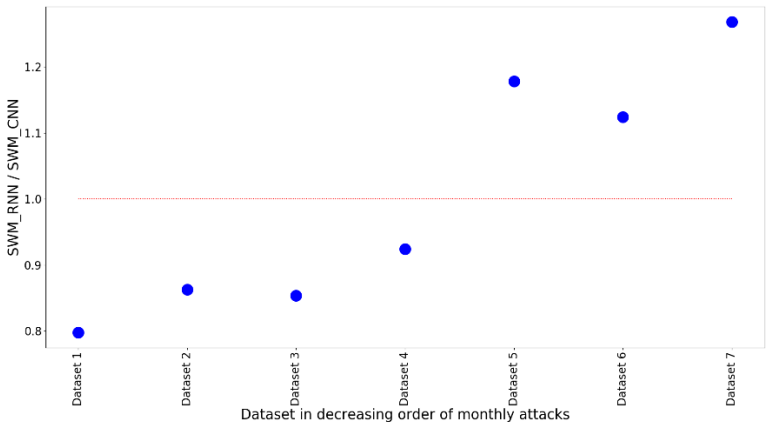

Figure 12: Comparison of SWM for RNN vs. CNN. On the X-axis the datasets are ordered with a decreasing number of monthly attacks. The Y-axis plots the ratio of SWM for RNN to SWM for CNN for a window size 5.

Regarding SWM, similar results are generated for all methods with a slight improvement of ANN methods for larger sliding window sizes.

The detailed discussions show that the ANN approaches are of similar predictive power as the selected classical time-series approaches in case of the representative terrorism toy model approach and when investigating long term predictions. It is found that the ANN approach behaves similar to the averaging naive estimator, whereas the periodic naive estimator is more accurate regarding the prediction of sparse events, which also holds true for the CNN approach. The CNN approach is better in predicting periodically recurring events and is also expected to further unfold its strengths when moving from univariate to multivariate time series.

Taking account of the generic nature of the toy model design, i.e. similar data structures as can be expected in the case of univariate times series in other application domains, and the option to consider additional time-series data, the present CNN time series approach can be expected to advance the time series analysis also in other domains of application. Such domains include minor disaster event distributions, congestions on roads and transport systems, stock market predictions, human monitoring, and health data.

5 Summary and Outlook

Due to the increased activity of terror groups worldwide, an important research topic is to investigate whether terrorist events can be predicted with sufficient accuracy, which could enable law enforcement agencies to much better prevent or mitigate the effects of attacks. Due to its complex nature and the lack of information, the accurate prediction of terror events will most likely not be possible in the near future. However, machine learning methods and ANNs, in particular, might be able to improve the prediction of threat situations by resorting to increasing input data sources, in particular, multivariate time series on terrorism event data and data on events relevant for terrorism prediction, e.g. major-related political events influencing the motivation of actors.

The paper used a simple toy model assuming general patterns of human behavior and logistics to compare the performance of ANNs and classic prediction methods for long-term time series prediction (Section 3). In the current study, the toy model generates imaginary attacks with varying intensity and is constructed to emulate several general properties of terror time series like random attacks (or data noise), certain favored attack times and a logistical pool of the corresponding terror organization. In principle, various other patterns or different groups with varying strategies can be included.

While training a deep learning model, the quantity of data plays a very important role. The presented toy model was an example of how to assess systematically the suitability of classical and ANN time series prediction methods using generated data. This allows determining the strengths and weaknesses of different approaches even before large-scale real terror and terror-related event time series data were available. In a similar way, in other fields of application, typical known features of data can be modeled and even before sufficient data is available the best time series prediction approaches can be selected. In addition, in all application domains, toy models can be used to analyze which type and quantity of data would suffice for modeling certain types of expected behaviors. Furthermore, phenomena that are not yet known in one field can be shown to be detectable by transferring data generation models from one application domain to another application domain.

Section 4 employed assessment regarding both regression and classification metrics. For the regression tasks, normed and integrated loss metrics are applied. The classification accuracy of the methods was addressed with the confusion matrix related quantities i.e. recall and precision. It was discussed that in case of terror events, pinpointing the exact time of attack holds minor importance, rather the approximate duration in which the attack will happen and also its intensity is of utmost priority. The introduced *sliding-window metric* (SWM) covered this and determines the capability to determine a certain threat situation in a given time-spans. This metric is also expected to be relevant for other domains, e.g. stock market developments.

As ANN models for analysis, RNN and CNN as specified in Section 2 were used, especially focusing on CNN. ARIMA and exponential smoothing being established time series analysis methods were not used because long-term prediction without step-wise updates cannot be applied appropriately to them, instead two naive predictors: setting the average value as a default for upcoming values (averaging estimator) and using the value of the same date from the previous year (periodic recurring estimator) are used. The analysis in Section 4 showed that overall ANNs reproduce comparable or slightly better results with SWM in this simple univariate test setup. Although RNNs are the default choice for time series prediction tasks, due to their properties this does not necessarily apply to long-term predictions without an update. Within the recent research work in the field of time series prediction, CNNs are gaining popularity over RNNs. This study showed in addition to a selected application domain that CNNs are a reasonable tool for long-term predictions without an update.

In the analysis at hand, it might not be evident why one should use ANNs for this kind of prediction if naive estimators achieve similar results. ANNs require large amounts of training data (which is not easy to obtain and at this point is not available at all) and are, in general, very resource-intensive compared to statistic data analysis methods. Still, the merits of the approach lie in its perspective usage:

- Feasibility of adding more features (e.g. country indicators like inflation rate, average costs of living or political stability indicators) and conducting a multivariate time series analysis is possible with CNNs. This will enhance the results significantly even in comparison with RNN. A further study could be conducted to perform multivariate time series analysis using the proposed architecture of CNNs and is expected to generate the advice which kind of additional time-series events would improve terrorist event prediction.
- More complex data patterns, e.g. rare events or long-term trends, will challenge the naive predictors significantly. The ANNs, on the other hand, are able to reproduce these patterns. They are also efficient to suppress the noise generated through individual attacks.
- With little adaptations, the method is also able to produce short-term predictions on a daily base as well as for longer time durations as often used in the domain.

However, the future usage of ANN methods in the field of terror event prediction should be approached with caution, as for this application several conditions need to be fulfilled:

- A sufficient amount of data should be available, which would include at least several thousand data points. This can be challenging for most of today's applications, as databases are restricted to event types that are very close to real terror events. This emphasizes the importance of comprehensive data recording and collection as well as the need to consider terror-related data.
- The prediction on the base of a pattern requires the existence of an actual pattern. Moreover, the pattern should be somehow stable throughout a sufficiently large period. Sudden paradigm changes, advances in technology or political events on a large scale will challenge the prediction abilities of all methods. This should be kept in mind when training data is prepared and results are interpreted.
- Any implementation of advanced time series prediction needs to consider laws regarding privacy and data protection.

The CNN long-term univariate time series prediction method presented in this study could also be applied in other domains. It is applicable to all fields that so far use classical times series prediction methods or RNN time series prediction methods. In addition, the discussed extension to multivariate time series prediction is also expected to be transferable to other domains.

Similarly, the toy data-model approach, i.e. modeling the most likely patterns and ensuring that they can be detected with the ANN time series prediction method selected, can be transferred to other domains. For a selected domain, it might be interesting to use data modeling approaches also from other application domains, in particular, to identify potentially unknown (unknown) features.

In summary, the paper showed that under the introduced assumptions, ANN methods and CNN, in particular, can be used for univariate long-term prediction of terror events. The results for the simple setup are comparable to standard statistical methods. Future studies may attend to situations with more complex multivariate time series data (e.g. more features, more patterns or more terror groups with different properties), optimizing the ANN

structure and quantifying uncertainties as well as applying CNNs (as a simple special case) also to short-term predictions with daily update.

References

Akram, Mohamed; El, Chaker (2016): Sequence to Sequence Weather Forecasting with Long Short-Term Memory Recurrent Neural Networks. In *IJCA* 143 (11), pp. 7–11. DOI: 10.5120/ijca2016910497.

Bengio, Y.; Lecun, Yann (1997): Convolutional Networks for Images, Speech, and Time-Series.

Bista, Iliana; Carvalho, Gary R.; Walsh, Kerry; Seymour, Mathew; Hajibabaei, Mehrdad; Lallias, Delphine et al. (2017): Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. In *Nature communications* 8, p. 14087. DOI: 10.1038/ncomms14087.

Borovykh, Anastasia; Bohte, Sander; Oosterlee, Cornelis W. (2017): Conditional Time Series Forecasting with Convolutional Neural Networks. Available online at http://arxiv.org/pdf/1703.04691v5.

Bowie, Neil G. (2018): 30 Terrorism Databases and Data Sets: a New Inventory. In *Perspectives on Terrorism* 12(5). Available online at http://bit.ly/2N3AnN4.

Chen, Mu-Yen; Chen, Bo-Tsuen (2015): A hybrid fuzzy time series model based on granular computing for stock price forecasting. In *Information Sciences* 294, pp. 227–241. DOI: 10.1016/j.ins.2014.09.038.

GTD. Global Terrorism Database (2018). Available online at https://www.start.umd.edu/gtd.

Lee, Nam-Uk; Shim, Jae-Sung; Ju, Yong-Wan; Park, Seok-Cheon (2018): Design and implementation of the SARIMA–SVM time series analysis algorithm for the improvement of atmospheric environment forecast accuracy. In *Soft Comput* 22 (13), pp. 4275–4281. DOI: 10.1007/s00500-017-2825-y.

Li, Shuying; Zhuang, Jun; Shen, Shifei (2017): Dynamic Forecasting Conditional Probability of Bombing Attacks Based on Time-Series and Intervention Analysis. In *Risk analysis : an official publication of the Society for Risk Analysis* 37 (7), pp. 1287–1297. DOI: 10.1111/risa.12679.

Madan, Rishabh; SarathiMangipudi, Partha (2018): Predicting Computer Network Traffic: A Time Series Forecasting Approach Using DWT, ARIMA and RNN. In : 2018 Eleventh International Conference on Contemporary Computing (IC3). 2018 Eleventh International Conference on Contemporary Computing (IC3). Noida, 8/2/2018 - 8/4/2018: IEEE, pp. 1–5.

Robert Nau: Statistical forecasting. notes on regression and time series analysis. Fuqua School of Business, Duke University. Available online at http://people.duke.edu/~rnau/411home.htm.

Sak, Haşim; Senior, Andrew; Rao, Kanishka; Beaufays, Françoise (2015): Fast and Accurate Recurrent Neural Network Acoustic Models for Speech Recognition, 7/24/2015. Available online at https://arxiv.org/pdf/1507.06947.

Serra, Edoardo; Subrahmanian, V. S. (2014): A Survey of Quantitative Models of Terror Group Behavior and an Analysis of Strategic Disclosure of Behavioral Models. In *IEEE Trans. Comput. Soc. Syst.* 1 (1), pp. 66–88. DOI: 10.1109/TCSS.2014.2307454.

Sheehan, Ivan Sascha (2009): Has the Global War on Terror Changed the Terrorist Threat? A Time-Series Intervention Analysis. In *Studies in Conflict & Terrorism* 32 (8), pp. 743–761. DOI: 10.1080/10576100903039270.

Shi, Heng; Xu, Minghao; Li, Ran (2018): Deep Learning for Household Load Forecasting—A Novel Pooling Deep RNN. In *IEEE Trans. Smart Grid* 9 (5), pp. 5271–5280. DOI: 10.1109/TSG.2017.2686012.

Snehanshu Saha; Harsha Aladi; Abu Kurien; Aparna Basu (2017): Future Terrorist Attack Prediction using Machine Learning Techniques.

Wei, Liang-Ying (2016): A hybrid ANFIS model based on empirical mode decomposition for stock time series forecasting. In *Applied Soft Computing* 42, pp. 368–376. DOI: 10.1016/j.asoc.2016.01.027.

Zhang, Jia-Shu; Xiao, Xian-Ci (1999): Predicting Chaotic Time Series Using Recurrent Neural Network. In *Chinese Phys. Lett.* 17 (2), p. 88. DOI: 10.1088/0256-307X/17/2/004.