Check for updates

ADVANCING EARTH AND SPACE SCIENCE

Geophysical Research Letters

RESEARCH LETTER
10.1029/2022GL099632

Forecasting Surface Velocity Fields Associated With Laboratory Seismic Cycles Using Deep Learning

G. Mastella[1], F. Corbi[2], J. Bedford[3], F. Funiciello[1], and M. Rosenau[4]

[1]Dip Scienze, Laboratory of Experimental Tectonics, Università "Roma TRE", Rome, Italy, [2]Istituto di Geologia Ambientale e Geoingegneria – CNR c/o Dipartimento di Scienze della Terra, Sapienza Università di Roma, Rome, Italy, [3]Institut für Geologie, Mineralogie und Geophysik, Ruhr-Universität Bochum, Bochum, Germany, [4]GFZ Helmholtz Centre Potsdam, German Research Centre for Geosciences, Potsdam, Germany

Key Points:

- A spatiotemporal regression framework is used to forecast surface velocity associated with seismic cycles reproduced in the laboratory
- The onset, magnitude, and propagation of analog earthquakes can be predicted up to a temporal horizon of the order of their duration
- Deep Learning outperforms standard Machine Learning, although their performances depend on data characteristics and model configurations

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

G. Mastella,
giacomo.mastella@uniroma3.it

Citation:

Mastella, G., Corbi, F., Bedford, J., Funiciello, F., & Rosenau, M. (2022). Forecasting surface velocity fields associated with laboratory seismic cycles using Deep Learning. *Geophysical Research Letters*, *49*, e2022GL099632. https://doi.org/10.1029/2022GL099632

Received 17 MAY 2022
Accepted 13 JUL 2022

Author Contributions:

Conceptualization: F. Corbi, J. Bedford, F. Funiciello
Data curation: F. Corbi
Formal analysis: F. Corbi
Funding acquisition: F. Funiciello
Investigation: F. Corbi, J. Bedford, M. Rosenau
Methodology: F. Corbi, J. Bedford
Project Administration: F. Funiciello
Resources: F. Funiciello
Supervision: F. Corbi, J. Bedford, F. Funiciello, M. Rosenau
Validation: F. Corbi
Writing – original draft: F. Corbi
Writing – review & editing: F. Corbi, J. Bedford, F. Funiciello, M. Rosenau

© 2022. American Geophysical Union. All Rights Reserved.

Abstract It has been recently demonstrated that Machine Learning (ML) can predict laboratory earthquakes. Here we propose a prediction framework that allows forecasting future surface velocity fields from past ones for analog experiments of megathrust seismic cycles. Using data from two types of experiments, we explore the prediction performances of multiple Deep Learning (DL) and ML algorithms. In such a self-supervised regression, no feature extraction is required and the entire seismic cycle is forecasted. The onset, magnitude, and propagation of analog earthquakes can thus be predicted at different prediction horizons. From all architectures tested in this study, convolutional recurrent neural networks (CNN-LSTM and CONVLSTM) provide the best predictions although their performances depend on experiment characteristics and hyperparameters tuning. Analog earthquakes can be successfully anticipated up to a horizon of the order of their duration. This laboratory-based study may open new avenues for transfer learning applications with data from natural subduction zones.

Plain Language Summary In the last few years scientists have shown their ability to predict the occurrence of earthquakes simulated in the laboratory using Machine Learning, a group of algorithms useful to learn hidden structure in data, complete tasks, and make predictions. By applying methods inspired by the structure and function of the brain (so-called "Neural Networks"), we here introduce an approach aiming to forecast the temporal evolution of the surface velocity field of analog experiments of the subduction megathrust seismic cycle. We show that our approach allows forecasting not only the onset and the size of laboratory earthquakes but also their preparatory phase and their propagation. Our success in laboratory earthquake forecasting provides optimism that one day similar results may be achieved with natural earthquakes.

1. Introduction

Over the last few years, Machine Learning (ML) has been used for predicting laboratory earthquakes in diverse experimental settings (Bergen et al., 2019; Ren et al., 2020). The majority of these studies have used ML to decrypt the acoustic signals emitted by a laboratory fault analog in double direct shear experiments (e.g., Bolton et al., 2019; Johnson et al., 2021; Lubbers et al., 2018; Rouet-Leduc et al., 2017, 2018; Wang et al., 2022). The predictive capability of ML seems to be related to asperity-scale processes (Shreedharan et al., 2021), which result in an exponential acoustic power increase observed prior to laboratory failures (Trugman et al., 2020). ML approaches applied to laboratory shearing experiments generally employed domain-specific features calculated in shifting windows (e.g., Rouet Leduc et al., 2017) or in temporal sequences (Shokouhi et al., 2021) to learn the evolution of acoustic properties during laboratory earthquake cycles. In these experiments, features (e.g., Jasperson et al., 2021) are extracted from one-dimensional arrays of acoustic emissions or shear stress measurements, thus no spatial information is taken into account. The prediction of laboratory earthquakes (i.e., labquakes) has been achieved through regression by predicting the time to failure and/or the instantaneous fault friction of the rock sample as a prediction label, that is, a one-dimensional target. Similarly, for seismotectonic analog subduction models, the time to failure has been successfully predicted (Corbi et al., 2019). In contrast to shear tests, analog models are three-dimensional (3D), downscaled replicas of the natural prototype which provide real-world-equivalent observational data by means of laboratory geodesy and seismology (Rosenau et al., 2017). When monitored by digital image correlation techniques (e.g., particle image velocimetry, PIV) a geodetic-like network is simulated (e.g., Kosari et al., 2020). Such a network tracks the spatial deformation of the model surface quantitatively in 2D (and potentially also in 3D) over time. To incorporate the resulting spatiotemporal patterns

into a prediction framework, Corbi et al. (2019) used both cumulative features and spatially-averaged instantaneous features from the deformation field to predict the time to the next labquake. Later, Corbi et al. (2020) proposed an alternative framework with a binary classification, where ML determines if failure occurs or not within an imminence time window.

These studies, both in shear experiments and in seismotectonic models, shed light and shadows on ML in labquake prediction. On one hand they highlight the capability of ML to predict different features of an upcoming labquake, providing new insights into fault physics (Hulbert et al., 2018, 2019). On the other hand, they rely on the subjective preparation of features that may depend on analog fault micromechanics and materials (Van Klaveren et al., 2020), and are potentially biased by the need of setting thresholds such as velocity for defining the time to failure or imminence window for the binary classification.

A breakthrough step in labquake prediction has been recently proposed by Laurenti et al. (2022), who introduced a forecasting procedure to infer the future shear stress of a laboratory fault. This strategy allows forecasting indefinitely in the future without employing labeling. By following a similar procedure, we profit here from geodetic-like surface deformation data from two different analog setups of megathrust seismic cycles to introduce a different forecasting scheme: a spatiotemporal regression that forecasts surface velocity fields from past ones. This method does not need to define an arbitrary threshold to identify labquakes, allowing simultaneously to forecast the onset of labquakes, their magnitude and to illuminate their space-time propagations at different prediction horizons. Moreover, the whole seismic cycle surface deformation is forecasted. The forecasting of velocity fields needs to incorporate memory effects in a spatial context. This task is a spatiotemporal problem and can be treated as a video sequence prediction problem (Reichstein et al., 2019). To perform the forecasting, we employ Deep Learning-based (DL) models from the Computer Vision field, a subfield of ML dedicated to interpreting images and video (Voulodimos et al., 2018). These algorithms are semi-supervised, meaning that feature extraction is not required. By comparing the performances obtained with baselines, such as the Persistence model and a Random Forest (RF) regressor, we quantify forecasting improvements that can be achieved with DL.

2. A Testbed for Labquakes Forecasting: Experimental Setups, Data and Methods

2.1. Experimental Setups and Data

In this study, we employ 4 datasets derived from two different experimental setups that mimic the key features of megathrust seismic cycles in a scaled fashion (see Rosenau et al. (2017) for a review on analog seismotectonic models): Gelquake (Corbi et al., 2013) and Foamquake (Mastella et al., 2022) (Figures 1a and 1b). Both setups share the same following characteristics: the subduction forearc is represented by a homogeneous and isotropic elastic wedge (of gelatin and foam rubber respectively, analogous of the upper plate) which is compressed by continuous kinematic loading at its base (subduction interface). Spontaneously nucleating frictional instabilities that propagate along the wedge base mimic megathrust earthquakes. The analog megathrust frictional properties can be tuned in different configurations such as double asperities (i.e., velocity weakening behaviors) along the strike. Asperities' interaction controls the recurrence behavior, causing sequences of full and partial ruptures (single or double asperities ruptures) with nested rupture cascades and superimposed cycles, similar to patterns observed at natural megathrusts (Philibosian & Meltzner, 2020). Given their remarkable capability to reproduce key features of natural megathrust seismic cycles (e.g., Corbi et al., 2022), Foamquake and Gelquake represent an ideal tool to test new prediction strategies.

From Gelquake we use one data set from a model with two equal (in terms of size and normal load) asperities. From Foamquake we use three datasets, one from a single asperity setup, one from a double-asperity model with two equal-sized asperities with the same load, and one from a model with the same configuration but different loads above asperities. Gelquake and Foamquake differ in wedge sizes, wedge properties, and deformation rates, but are both monitored with high-resolution (1,600 × 1,200 and 2,048 × 1,536 px^2, respectively) top-view video cameras. Imaging frequencies are 7.5 Frames Per Second (fps) for Gelquake and 50 fps for Foamquake. Data sets consist of instantaneous surface velocity maps obtained through image cross-correlation between consecutive images using the particle image velocimetry (PIV) technique (Sveen, 2004). PIV analysis resulted in 30 × 45 measurement ("GNSS") points for Gelquake and 18 × 29 for Foamquake. PIV mimics a homogeneously distributed geodetic network with 5 and 17 km spaced stations for Gelquake and Foamquake, respectively. Foamquake experiments run for 2 min, during which between 92 and 155 labquakes of 0.02–0.08 s duration (1–4 frames)

19448007, 2022, 15, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022GL099632 by CochraneItalia, Wiley Online Library on [11/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

10.1029/2022GL099632

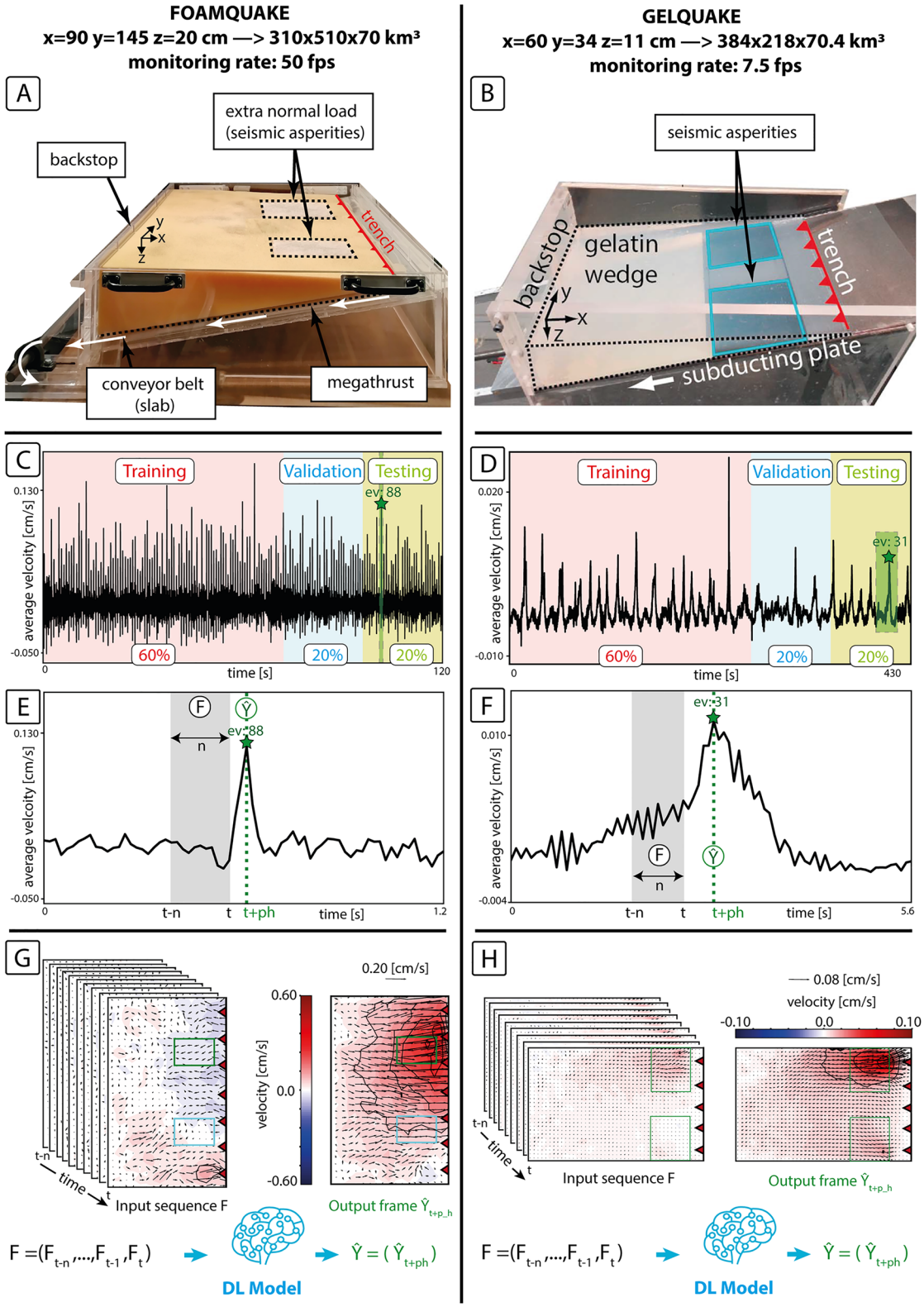

Figure 1. Experimental setups, data, and prediction strategy: oblique view of Foamquake (a) and Gelquake (b). Time series of the average velocity field (trench-orthogonal component) (c, d). Colored shades depict the fraction of data used for training, validation, and testing. Dashed gray rectangles highlight 80 frames represented in subplots (e, f) which include the labquakes marked by green stairs. Subplots e and f represent an outline of our forecasting strategy. The gray shaded area is the image sequence $F = (F_{t-n}, ..., F_{t-1}, F_t)$, with n = number of frames, used as an input to forecast the frame Y at the time (t + ph). Panels (g) and (h) represent the temporal sequence of velocity frames F used to train Deep Learning (DL) models to forecast the frame Y.

19448007, 2022, 15, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022GL099632 by CochraneItalia, Wiley Online Library on [11/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

occur. The Gelquake experiment was 7 min-long during which 40 labquakes occur with an average duration of around 1.5 s (10 frames). The time ratio between average interseismic and coseismic phases is therefore relatively different between Foamquake (∼17:1) and Gelquake (∼5.25:1) allowing us to test the sensitivity of our approach to different model behaviors.

2.2. Forecasting Strategies

Because our data consist of sequential maps (frames), in this study we exploit techniques derived from the frame predictive learning, a tool to understand and model the dynamics of natural scenes (Zhou et al., 2020). Predictive learning foresees plausible future outcomes by learning meaningful representations of the underlying patterns in a set of historical frame inputs (e.g., Oprea et al., 2020).

Our regression task of predicting future velocity maps can be formally defined as follows (Figures 1e–1h). Consider $F_t \in R^{X \times Y \times c}$ as the *t-th* frame in the image sequence $F = (F_{t-n}, ..., F_{t-1}, F_t)$ with n = number of frames, where X and Y denote the width and length of surface velocity maps, while c represents the components, that is, trench parallel and trench orthogonal (e.g., the number of channels in computer vision nomenclature). The target is to predict only one frame in the future at a given Prediction Horizon (ph), $\hat{Y} = (\hat{Y}_{t+1}, \hat{Y}_{t+2}, ..., \hat{Y}_{t+p_h})$ from the input F. In this study we employ 3 different Deep Neural Networks (DNNs) for predicting $\hat{Y}$: Convolutional Neural Network (CNN), Convolutional Long-Short-Term Memory Network (ConvLSTM) and a combination of CNN with LSTM (CNN-LSTM).

CNNs are a class of DNNs commonly applied to images to automatically learn hidden spatial patterns through backpropagation (Goodfellow et al., 2016). CNNs have been recently applied in seismological problems (e.g., Blank & Morgan, 2021; Jozinović et al., 2020; Münchmeyer et al., 2021; Rouet-Leduc et al., 2021; Zhu and Beroza, 2019) as well as in predictive learning tasks in meteorology and climatology (Chattopadhyay et al., 2020; Ham et al., 2019). We developed a CNN network expanding the LeNet architecture (LeCun et al., 1989) (Figure S1a, Text S1 in Supporting Information S1). As predictive CNNs, our network concatenates the previous n frames (each of which includes the 2 velocity components) as a sequence of tensors on the channel dimension (dimension = 2n): the convolution is thus applied both in the temporal and spatial domain to capture correlations in past inputs (e.g., Zhou et al., 2020).

ConvLSTM is a Recurrent Neural Network (RNN) developed to solve spatiotemporal forecasting problems. RNNs recursively update an internal status to process sequences of inputs. In labquake prediction, this internal status may represent the instantaneous state of the analog fault (Laurenti et al., 2022). ConvLSTM redesigns the architecture of fully-connected LSTM (Long Short-Term Model) (Hochreiter & Schmidhuber, 1997), allowing to receive and output multidimensional tensors. We adopted an architecture (Figure S1b, Text S1 in Supporting Information S1) akin to the encoding-forecasting model originally proposed by Shi et al. (2015) for precipitation forecasting.

The CNN-LSTM integrates the visual learning ability of CNNs with the effectiveness of LSTMs in learning short- and long-term temporal dependencies. In that sense, CNN-LSTM is "doubly deep," both in space and time and can be applied to a variety of visual problems (Donahue et al., 2017). This architecture can be conveniently thought of as the union of two sub-models: an encoder CNN model for feature extraction and a LSTM model for interpreting features across time steps (Vinyals et al., 2015). Our CNN-LSTM model applies a simple CNN architecture independently to each input frame, to build a sequence of internal vector representations that track the kinematic evolution of the wedge. This sequence is subsequently used by a two-layer LSTM that learns temporal dependencies thanks to feedback connections and gate techniques (Figure S1c, Text S1 in Supporting Information S1).

The predictive performances obtained with the three DL models are compared to those obtained with forecasting baselines: the Random Forest (Breiman, 2001) (Text S2 in Supporting Information S1) and the Persistence and Moving Average Models (MAV) (Text S2 in Supporting Information S1). RF is chosen as a baseline because it has shown promising results in labquakes prediction (e.g., Rouet-Leduc et al., 2017). Unlike the DNNs (which have tensor inputs), our RF is fed by 1D arrays representing single flattened velocity frames: each PIV velocity vector is an input feature. Therefore, our RF does not receive information related to the evolution of surface velocities.

19448007, 2022, 15, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022GL099632 by CochraneItalia, Wiley Online Library on [11/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

For each tested algorithm, we define a custom model architecture independently on Foamquake and Gelquake data sets, setting the training hyperparameters after extensive tuning (Text S1 in Supporting Information S1). We train a separate model for each prediction horizon. Each experimental data set is split into 60% for training, 20% for validation, and 20% for testing (Figures 1c and 1d). Training data sequences have been randomly shuffled to prevent overfitting and promote generalization, without breaking the temporal coherence (Text S1 in Supporting Information S1). For all DNNs we test model performances as a function of the temporal lengths of input sequences (i.e., number of input frames). During the model training the mean squared error (MSE) has been set as the loss function applied for gradient descent (Text S1 in Supporting Information S1), with the Adam optimizer (Kingma & Ba, 2015).

Forecasting performances are quantified through the normalized root mean squared error and the structural similarity index (NRMSE and SSIM, Text S3 in Supporting Information S1). Both metrics are useful being complementary: NRMSE represents an absolute error (i.e., point by point comparison) while SSIM quantifies the similarity between spatial patterns (Sara et al., 2019). Both metrics are scale-independent allowing to equally evaluate the forecasts of interseismic and coseismic frames despite their different magnitudes and to compare performances of different datasets. Forecasting results are calculated considering both validation and testing sets. The forecasts have been additionally evaluated using classification metrics (Text S3 in Supporting Information S1). These classification metrics are not central for our study but facilitate the quantification of prediction performances. We did not apply any independent normalization on the two velocity components: this was tried but was found to increase ∼10% the NRMSE.

3. Results

3.1. Comparison Between Different ML and DL Algorithms and Between Different Analog Experiments

For all datasets, DL algorithms outperform forecasting baselines. The forecasting performances of each tested algorithm are represented for the Foamquake experiment characterized by different normal loads over the two seismic asperities (Figures 2a and 2b). This experiment has been selected as a reference being the most similar with respect to interplate seismicity in nature and ideally the most difficult to be predicted due to the larger variability of rupture sizes, locations, and recurrence time (Mastella et al., 2022). The same analysis has been performed for all other experiments: (a) experiments from the same setup (i.e., Foamquake), but with different configurations; (b) experiments from a different setup (i.e., Gelquake). Results are summarized in Tables S1, S2 and Figures S2, S3 in Supporting Information S1.

We implement two prediction performance evaluation scenarios: scenario C where only coseismic frames predictions are evaluated and scenario IC where both interseismic and coseismic are considered.

For each prediction horizon, we report the average NRMSEs of the model with the input sequence length that provides the best performance for each scenario (Figures 2a and 2b). All algorithms show degrading prediction performances with increasing prediction horizons. Generally, in scenario IC, the performances of the different DL algorithms are similar, while differences become clear when considering only scenario C. For a ph = 1 frame, the ConvLSTM reveals the best forecasting prowess. Going further into the future, the ConvLSTM tends to produce overly smoothed predictions and the CNN-LSTM instead shows the best predictions (Figures 2a and 2b). Coseismic errors obtained with the CNN are significantly higher when compared with CNN-LSTM ones. Compared to all tested DL algorithms, RF shows generally the worst predictions (Figures 2a and 2b).

The CNN-LSTM is capable of forecasting analog earthquakes up to 3–4 frames in advance (i.e., 0.06–0.08 s): until this horizon, coseismic predictions coincide with real velocity maps both in amplitude, direction, and spatial distribution (Figure 3). The number of frames used as input influences the forecasting performances but without any clear relations (Figure S4 in Supporting Information S1). The NRMSE is similar considering both scenarios (Figures 2a and 2b) at short horizons (1–2–3 frame). The discrepancy between coseismic and interseismic predictions becomes clear at longer horizons (i.e., 4–5 frames), where the error is correlated with velocity amplitudes. Indeed, at ph > 3–4 frames all architectures prefer "conservative" (i.e., slower) predictions for coseismic frames (Figure S5 in Supporting Information S1). Generally, interseismic velocities are not well predicted, owing to fluctuations of spatial noise through time (Movie S1). Only trench orthogonal velocities are correctly predicted while for trench-parallel velocities the forecasts fail.

19448007, 2022, 15, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022GL099632 by CochraneItalia, Wiley Online Library on [11/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

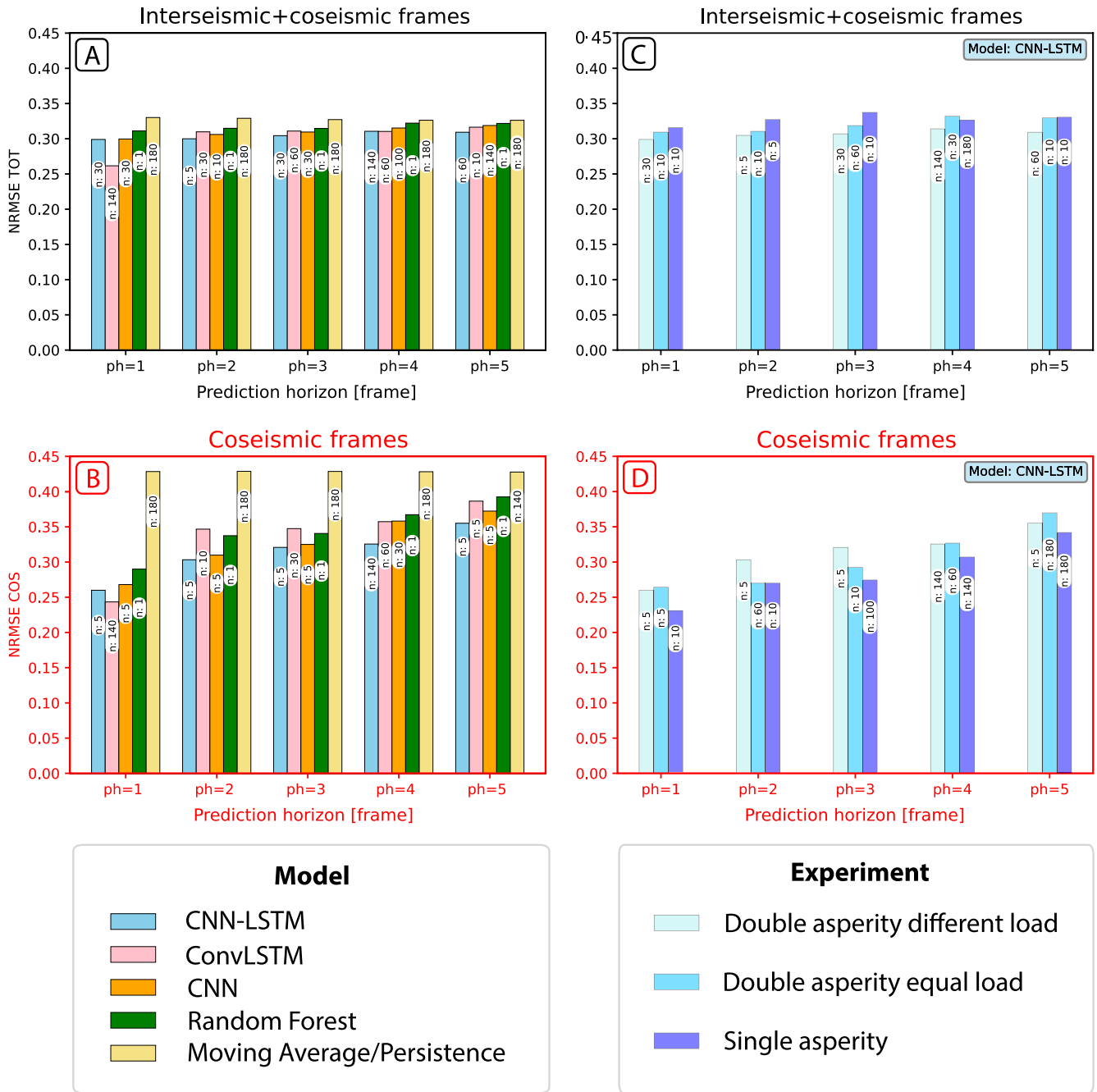

Figure 2. Prediction errors (NRMSE) for all employed algorithms as a function of the prediction horizon (a, b). The results from the double-asperity Foamquake experiment with different normal loads above asperities are reported. Comparison of errors between different experimental configurations of Foamquake obtained with the CNN-LSTM architecture (c, d). The first row refers to errors calculated considering all frames (a–c) while the second row shows only coseismic frames errors (b–d). For each horizon, the forecasts are obtained with the best performing model which uses a number n of frames as input.

The forecasting performances achieved with the Gelquake experiment are shown in Figure S6 in Supporting Information S1. The ConvLSTM is the more efficient architecture in both scenario IC and C for short horizons, while for scenario C and long horizons, the CNN-LSTM outperforms all the other architectures. The RF shows a larger error, but predictions are still acceptable. As for Foamquake, predictions degrade with the prediction horizon, but in this experiment not only the trench-orthogonal velocity component but also the trench-parallel one is well predicted (Figure S7 in Supporting Information S1). The maximum predictable horizon of gelquakes is around 6–7 frames (0.8–0.93 s) (Figure S8 in Supporting Information S1). Performances are influenced by the

19448007, 2022, 15, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022GL099632 by CochraneItalia, Wiley Online Library on [11/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

10.1029/2022GL099632

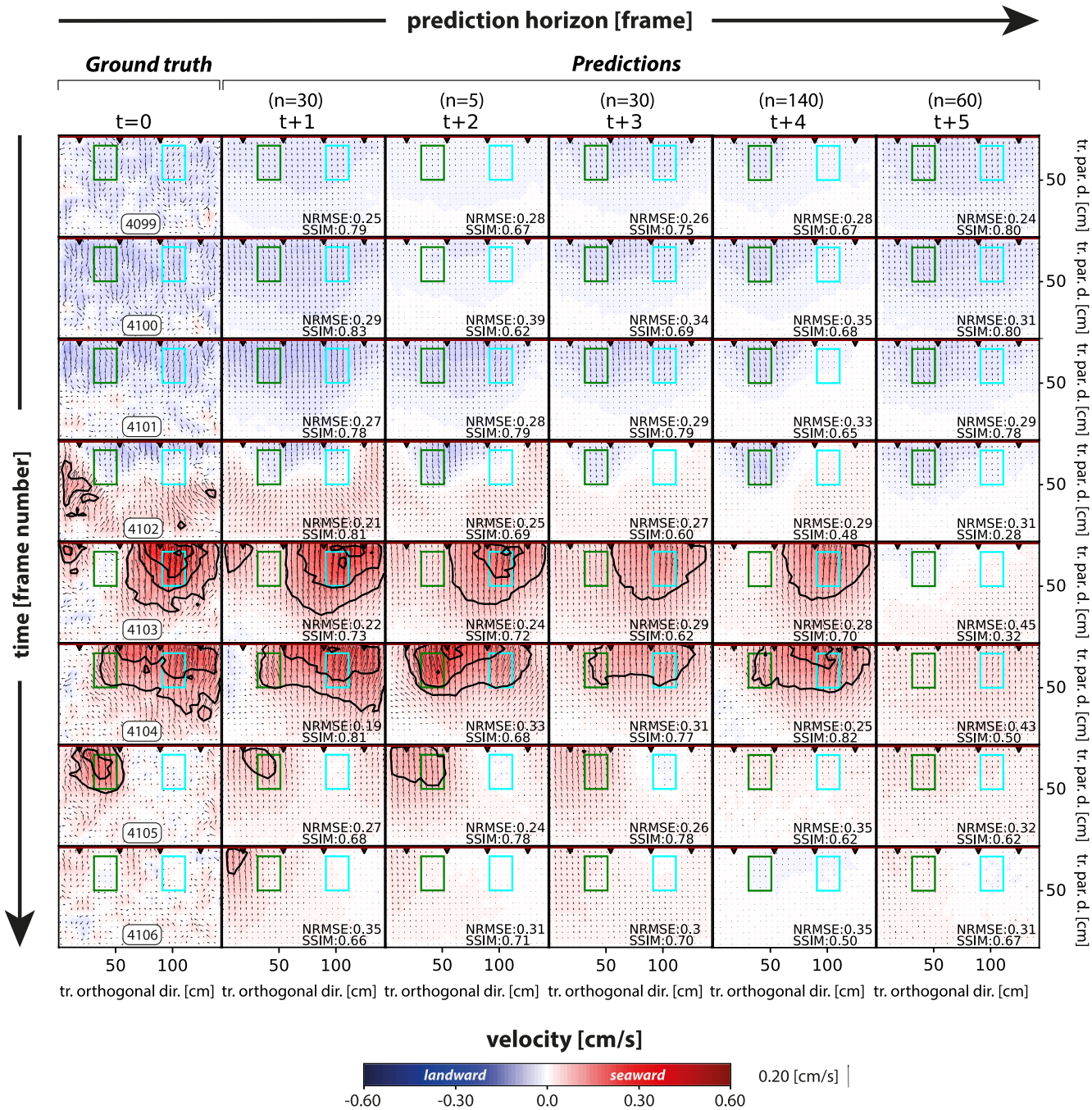

Figure 3. Velocity fields forecasting example obtained with the CNN-LSTM architecture. Reference Foamquake experiment. Asperities' projections are represented by colored rectangles (in cyan is the asperity with $\sigma_n = 40$ Pa, in green is the asperity with $\sigma_n = 100$ Pa). Each column shows the forecast of 8 frames at different prediction horizons. For each horizon, the forecasts are obtained with the best performing model which uses a number n of frames as input. Forecasting errors (NRMSE and SSIM) are reported for each frame.

number of frames used as input but without any direct relations. Coseismic predictions show smaller NRMSEs than interseismic ones (Video S2).

Figures 2c and 2d show the NRMSE obtained with the CNN-LSTM for the three different Foamquake experiments. Considering scenario IC (Figure 2c), the lowest error is observed for the experiment with two asperities with different normal loads, while the highest error is observed for the single-asperity experiment. In contrast, in scenario C, the single-asperity experiment shows the lowest NRMSE (Figure 2d). With respect to

19448007, 2022, 15, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022GL099632 by CochraneItalia, Wiley Online Library on [11/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Foamquake, the Gelquake experiment shows a considerably lower error (i.e., difference between NRMSE of the two models≃0.08), especially for scenario C.

3.2. Forecasting Labquake Rupture Characteristics

Our regression method allows us to assess the onset and evolution of slip in advance. Slip distribution at depth can be constrained by inversion (e.g., Kosari et al., 2020) but to avoid additional inversion-related bias, we estimate analog earthquake source parameters, (i.e., average slip, rupture area and therefore magnitude) directly from the surface measurements in analogy with Mastella et al. (2022) and Corbi et al. (2013).

For both Foamquake and Gelquake we compare predicted versus observed source parameters (Figures S9a and S9b in Supporting Information S1). For short horizons, foamquakes are almost perfectly predicted with a percentage higher than the 90%, and source parameters are adequately estimated (details in Text S3 in Supporting Information S1). The percentage of successful predictions of labquakes decreases to 36% for a 5-frames horizon, and it is primarily the labquakes from the asperity with the highest load that are forecasted less well, likely due to the paucity of this type of events in the data set. Gelquakes are predicted with a percentage higher than the 60% up to a horizon of 6 frames.

For both Gelquake and Foamquake, Accuracy and Specificity remain higher than 0.90 for each of the investigated prediction horizons (Figure S10 in Supporting Information S1). Recall and Precision, which measure the coverage of the coseismic minority class, oscillates from 0.2 to 0.9 depending on the prediction horizon. Regarding Foamquake experiments, coherent with regression results, the double-asperity configuration with different loads depicts the highest scores until a ph = 3 frames. Around ph = 4–5, scores suffer a sharp drop reaching values lower than 0.2. With Gelquake, Recall and Precision are higher (always >0.5), and no clear drops are observed.

4. Discussion

Aware that prediction performances obtained with different algorithms depend on hyperparameters exploration, our analysis shows that computer vision DNNs significantly outperform RF and Moving Average Models, which provide from the 10% to the 50% higher NRMSEs depending on the experiment. Why is this the case? The DNNs are different to RF in that the convolutional operations are tailored toward learning spatial patterns of analog seismic cycles. Furthermore, the DNNs are fed by a series of input frames (rather than just one frame for the RF) and so there is additional sequential information that can potentially be improving the prediction. The discrepancy in prediction performances between the two experimental setups suggests that the choice of the DL algorithm as well as the tuning of the hyperparameters depends on the experimental data set. With Gelquake, RF shows a predictive power almost as good as DL algorithms despite the lower computational effort. For other systems, such as Foamquake, an overly-simplified representation may limit the prediction of challenging spatiotemporal data. The application of recurrent convolutional neural networks, networks that emphasize the spatiotemporal sequentiality of the input, strongly improve predictions. This suggests that the predictive knowledge gained by the DL architecture is embedded in the temporal sequence structure rather than in its individual frames. Our spatiotemporal approach allows us to forecast all phases of laboratory seismic cycles: interseismic, coseismic and postseismic (if present). Our DNNs forecast with reliable fidelity the magnitude and evolution of analog earthquakes based on the kinematic deformation of analog wedge surfaces, without the need to extract features. Labquakes can be predicted up to a horizon on the order of their durations, similar to the maximum predictability estimated for slow earthquakes along the Cascadia margin (Gualandi et al., 2020). The maximum predictable horizon in Gelquake is around 6–7 frames (equivalent to 0.8/0.93 s, or around 10% of the interseismic phase duration) while for Foamquake is only around 3 frames (equivalent to 0.06 s, or ∼4% of the interseismic time length). Upscaling these temporal horizons to the natural prototype is a critical and at the same time not straightforward task. This is because the choice of the time scaling factor in analog experiments depends on whether the increment is considered to be part of the dynamic coseismic rupture phase or instead of the quasi-static interseismic phase, which are differently scaled (Rosenau et al., 2009, 2017). Considering the frames as part of the short-term nucleation phase (i.e., as part of the coseismic phase and scaled as such), the maximum predictable horizons of foamquakes and gelquakes upscale to ∼36 and 745 s, respectively. If the frames are instead considered as part of the long-term loading phase and the interseismic scaling is employed, prediction horizons of ∼6 and 220 years could be expected.

19448007, 2022, 15, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022GL099632 by CochraneItalia, Wiley Online Library on [11/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Regarding Foamquake, we observe differences between individual experimental datasets, suggesting that the spatial variability of the properties of physical systems under consideration influence the predictive performance of DNNs. Foamquakes are easier to forecast in single-asperity configurations than in double-asperity due to complexity created by asperities interaction. In contrast, the interseismic deformation is more confidently predicted in the double-asperity experiment with different loads as this experiment is characterized by the highest load (with respect to the other experiments) and the higher load causes larger interseismic displacement, and in turn less noisy time series. On the contrary, in Gelquake the interseismic deformation is successfully predicted and predictions are accurate considering both the two components of the velocity field. These results could seem counterintuitive given that for Gelquake the seismic cycles employed during the training are around ⅓ of cycles used for Foamquake experiments, but we must also consider the experimental monitoring rate and the duration of interseismic phases: despite Gelquake and Foamquake being interseismically deformed with the same rate, the lower frame rate of Gelquake allows PIV to retrieve velocities more precisely than in Foamquake.

Differently from classic ML (e.g., RF), DNNs are often considered black boxes, due to the difficulty to identify features that allowed successful predictions. For this reason, we detailed the interseismic phase, aiming to recognize possible informative patterns. We found that foamquakes are anticipated by a 2-stage preparatory phase: late interseismic phases characterized by subtle landward acceleration followed by a reversal during the last frames before the occurrence of main slip episodes. Reversals have "slower than coseismic" seaward velocities and appear at greater depths than asperities. Such slow, deep slip loads the shallower interseismically locked regions, triggering the subsequent shallow slip within the asperities (Figure 3), like what is proposed for real megathrust earthquakes (Burgmann et al., 2005; Moreno et al., 2014). For Gelquake, the late interseismic landward acceleration, a transient increase in locking, is not observed. However, also for Gelquake anticipatory "slower than coseismic" slip nucleates in the deeper asperity region, which in fact is the most informative for labquake prediction (Corbi et al., 2020). The presence of a preparatory phase to analog ruptures is compatible with the loading view of earthquake nucleation, which suggests that earthquakes are anticipated by tectonic loading processes associated with aseismic slip and shear localization (e.g., Kato & Ben-Zion, 2021; Ohnaka, 1992). This observation opens the subsequent question: do DNNs recognize, learn, and exploit only such a preparatory phase to forecast future surface velocities? Or alternatively, is it the entire training data set that allows successful predictions? To tackle this problem, we select the reference Foamquake experiment as a testbed and we train a CNN-LSTM model excluding the 5 frames that anticipate each rupture (i.e., the average duration of the preparatory phase). This model can still forecast labquakes, but with less precision on their location, magnitude, and propagation (i.e., NRMSE increases of ∼15% with respect to the original model). This result underlines that successful predictions are not achieved only due to the peculiarity of the late preparatory phase.

5. What Can We Learn From Analog Earthquakes Forecasting?

The models introduced here are purely deterministic as is the case for most existing DL-based models proposed in literature for video prediction (Oprea et al., 2020) and also as generally proposed for labquake prediction (e.g., Ren et al., 2020). Deterministic prediction would suffice for well-controlled laboratory settings. These experiments lack the geological complexity of the Earth (Beroza et al., 2021), when multiple predictions could be equally likely due to epistemic and/or aleatory uncertainties (Kiureghian et al., 2009). For this reason, the proposed methods can be considered as baseline approaches, which are improvable by providing confidence intervals on predictions, for example, by introducing Bayesian Neural Networks (Shridhar et al., 2019). Incorporating uncertainty becomes even more essential if learning capabilities developed originally with laboratory data are extrapolated to the natural prototype through Transfer Learning applications. Pioneering studies of Transfer Learning have recently been proposed using numerical simulations to predict labquakes (Wang et al., 2021) and for predicting off-fault deformation of natural faults from strike-slip analog experiments (Chaipornkaew et al., 2021). Our study paves the path to the application of transfer learning from analog experiments to the natural prototype. Given the straightforward analogies between space-time rupture behaviors observed along natural subduction zones with the patterns generated by megathrust analog models (e.g., Caniven & Dominguez, 2021; Corbi et al., 2022), learning in the laboratory how to predict fast and/or slow earthquakes from geodetic data is the challenging evolution of this study. Cascadia (North America) and Hikurangi (New Zealand) subduction zones, where the geodetic records contain multiple cycles of slow earthquakes (e.g., Michel et al., 2019; Wallace, 2020), are potential candidates to test our approach on, even though the physical mechanism of slow earthquakes might

19448007, 2022, 15, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022GL099632 by CochraneItalia, Wiley Online Library on [11/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

be different from large megathrust earthquakes. Our methods can be potentially applied in the Continuous Training context, where models are continuously retrained to respond to changes in input data. In that sense, our models could enable the spatiotemporal forecasting of the geodetic deformation in real time.

Data Availability Statement

All input data used for the analysis are available through GFZ Data Services and published open access in Corbi et al. (2019) and Mastella et al. (2021). Models' implementations are developed with the Python packages Tensorflow 2 (Abadi et al., 2016) and Scikit-learn (Pedregosa et al., 2011). The training is run on a computer with a single NVIDIA 435-21 GeForce RTX2070 GPU. All figures were made using the Python package Matplotlib v.3.1.0 (Hunter, 2007).

Acknowledgments
This paper has been financially supported by the EPOS Research Infrastructure through the contribution of the Italian Ministry of University and Research (MUR) - EPOS ITALIA Joint Research Unit. The Grant to the Department of Science, Roma Tre University (MIUR-Italy Dipartimenti di Eccellenza, ARTICOLO 1, COMMI 314–337 LEGGE 232/2016) is gratefully acknowledged. Matthias Rosenau has been supported by Deutsche Forschungsgemeinschaft (DFG) through grant CRC 1114 "Scaling Cascades in Complex Systems," Project Number 235221301, Project B01. We thank R. Minzoni (Avanade Spa) for the initial suggestions about Computer Vision algorithms.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint. https://doi.org/10.48550/arXiv.1603.04467

Bergen, K. J., Johnson, P. A., Maarten, V., & Beroza, G. C. (2019). Machine learning for data-driven discovery in solid Earth geoscience. *Science*, *363*, 6433. https://doi.org/10.1126/science.aau0323

Beroza, G. C., Segou, M., & Mostafa Mousavi, S. (2021). Machine learning and earthquake forecasting—Next steps. *Nature Communications*, *12*(1), 4761. https://doi.org/10.1038/s41467-021-24952-6

Blank, D. G., & Morgan, J. K. (2021). Can deep learning predict complete ruptures in numerical megathrust faults? *Geophysical Research Letters*, *48*(18), e2021GL092607. https://doi.org/10.1029/2021GL092607

Bolton, D. C., Shokouhi, P., Rouet-Leduc, B., Hulbert, C., Riviere, J., Marone, C., & Johnson, P. A. (2019). Characterizing acoustic signals and searching for precursors during the laboratory seismic cycle using unsupervised machine learning. *Seismological Research Letters*, *90*(3), 1088–1098. https://doi.org/10.1785/2F0220180367

Breiman, L. (2001). "Random forests". *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/a:1010933404324

Bürgmann, R., Kogan, M. G., Steblov, G. M., Hilley, G., Levin, V. E., & Apel, E. (2005). Interseismic coupling and asperity distribution along the Kamchatka subduction zone. *Journal of Geophysical Research*, *110*(B7), B07405. https://doi.org/10.1029/2005JB003648

Caniven, Y., & Dominguez, S. (2021). Validation of a multilayered analog model integrating crust-mantle visco-elastic coupling to inves-tigate subduction megathrust earthquake cycle. *Journal of Geophysical Research: Solid Earth*, *126*(2), e2020JB020342. https://doi.org/10.1029/2020JB020342

Chaipornkaew, L., Elston, H., Cooke, M. L., Mukerji, T., & Graham, S. A. (2021). Prediction of off-fault deformation from experimental strike-slip fault structures using the convolutional neural networks. *Earth and Space Science Open Archive*, *12*. https://doi.org/10.1002/essoar.10507909.1

Chattopadhyay, A., Hassanzadeh, P., & Pasha, S. (2020). Predicting clustered weather patterns: A test case for applications of convolutional neural networks to spatio-temporal climate data. *Scientific Reports*, *10*(1), 1317. https://doi.org/10.1038/s41598-020-57897-9

Corbi, F., Bedford, J., Poli, P., Funiciello, F., & Deng, Z. (2022). Probing the seismic cycle timing with coseismic twisting of subduction margins. *Nature Communications*, *13*(1), 1911. https://doi.org/10.1038/s41467-022-29564-2

Corbi, F., Bedford, J., Sandri, L., Funiciello, F., Gualandi, A., & Rosenau, M. (2020). Predicting imminence of analog megathrust earthquakes with machine learning: Implications for monitoring subduction zones. *Geophysical Research Letters*, *47*(7), e2019GL086615. https://doi.org/10.1029/2019GL086615

Corbi, F., Funiciello, F., Moroni, M., Van Dinther, Y., Mai, P. M., Dalguer, L. A., & Faccenna, C. (2013). The seismic cycle at subduction thrusts: 1. Insights from laboratory models. *Journal of Geophysical Research: Solid Earth*, *118*(4), 1483–1501. https://doi.org/10.1029/2012JB009481

Corbi, F., Sandri, L., Bedford, J., Funiciello, F., Brizzi, S., Rosenau, M., & Lallemand, S. (2019). Machine learning can predict the timing and size of analog earthquakes. *Geophysical Research Letters*, *46*(3), 1303–1311. https://doi.org/10.1029/2018GL081251

Donahue, J., Hendricks, L. A., Rohrbach, M., Venugopalan, S., Guadarrama, S., Saenko, K., & Darrell, T. (2017). Long-term recurrent convolutional networks for visual recognition and description. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(4), 677–691. https://doi.org/10.1109/TPAMI.2016.2599174

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning*. MIT Press.

Gualandi, A., Avouac, J.-P., Michel, S., & Faranda, D. (2020). The predictable chaos of slow earthquakes. *Science Advances*, *6*(27), 5548. https://doi.org/10.1126/sciadv.aaz5548

Ham, Y. G., Kim, J. H., & Luo, J. J. (2019). Deep learning for multi-year ENSO forecasts. *Nature*, *573*(7775), 568–572. https://doi.org/10.1038/s41586-019-1559-7

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hulbert, C., Rouet-Leduc, B., Johnson, P. A., Ren, C. X., Riviere, J., Bolton, D. C., & Marone, C. (2018). Laboratory earthquake prediction illuminates connections between the spectrum of fault slip modes. *Nature Geoscience*, *12*(1), 69–74. https://doi.org/10.1038/s41561-018-0272-8

Hulbert, C., Rouet-Leduc, B., Johnson, P. A., Ren, C. X., Riviere, J., Bolton, D. C., & Marone, C. (2019). Similarity of fast and slow earthquakes illuminated by machine learning. *Nature Geoscience*, *12*(1), 69–74. https://doi.org/10.1038/2Fs41561-018-0272-8

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Jasperson, H., Bolton, D. C., Johnson, P., Guyer, R., Marone, C., & Hoop, M. V. (2021). Attention network forecasts time-to-failure in laboratory shear experiments. *JGR Solid Earth*, *126*(11), e2021JB022195. https://doi.org/10.1029/2021JB022195

Johnson, P. A., Rouet-Leduc, B., Pyrak-Nolte, L. J., Beroza, G. C., Marone, C. J., Hulbert, C., et al. (2021). Laboratory earthquake forecasting: A machine learning competition. *Proceedings of the National Academy of Sciences*, *118*(5), e2011362118. https://doi.org/10.1073/pnas.2011362118

19448007, 2022, 15, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022GL099632 by CochraneItalia, Wiley Online Library on [11/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Jozinović, D., Lomax, A., Štajduhar, I., & Michelini, A. (2020). Rapid prediction of earthquake ground shaking intensity using raw waveform data and a convolutional neural network. *Geophysical Journal International*, *222*(Issue 2), 1379–1389. https://doi.org/10.1093/gji/ggaa233

Kato, A., & Ben-Zion, Y. (2021). The generation of large earthquakes. *Nature Reviews Earth & Environment*, *2*(1), 26–39. https://doi.org/10.1038/s43017-020-00108-w

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In *Proceedings of the International conference on learning representations* (p. 141). Retrieved from https://arxiv.org/abs/1412.6980

Kiureghian, A. D., & Ditlevsen, O. (2009). "Aleatory or epistemic does it matter?" *Structural Safety*, *31*(2), 105–112. https://doi.org/10.1016/j.strusafe.2008.06.020

Kosari, E., Rosenau, M., Bedford, J., Rudolf, M., & Oncken, O. (2020). On the relationship between offshore geodetic coverage and slip model uncertainty: Analog megathrust earthquake case studies. *Geophysical Research Letters*, *47*(15), e2020GL088266. https://doi.org/10.1029/2020gl088266

Laurenti, L., Tinti, E., Galasso, F., Franco, L., & Marone, C. (2022). Deep learning for laboratory earthquake prediction and autoregressive forecasting of fault zone stress. arXiv preprint arXiv:2203.13313.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). "Backpropagation applied to handwritten zip code recognition". *Neural Computation*, *1*(4), 541–551. https://doi.org/10.1162/neco.1989.1.4.541

Lubbers, N., Bolton, D. C., Mohd-Yusof, J., Marone, C., Barros, K., & Johnson, P. A. (2018). Earthquake catalog-based machine learning identification of laboratory fault states and the effects of magnitude of completeness. *Geophysical Research Letters*, *45*(24), 13–269. https://doi.org/10.1029/2F2018GL079712

Mastella, G., Corbi, F., Funiciello, F., & Matthias, R. (2021). *Particle image correlation data from foamquake: A novel seismotectonic analog model mimicking the megathrust seismic cycle*. GFZ Data Services. https://doi.org/10.5880/fidgeo.2021.046

Mastella, G., Corbi, F., Funiciello, F., & Rosenau, M. (2022). Foamquake: A novel analog model mimicking megathrust seismic cycles. *Journal of Geophysical Research: Solid Earth*, *127*(3). https://doi.org/10.1029/2021JB022789

Michel, S., Gualandi, A., & Avouac, J. P. (2019). Similar scaling laws for earthquakes and Cascadia slow-slip events. *Nature*, *574*(7779), 522–526. https://doi.org/10.1038/s41586-019-1673-6

Moreno, M., Haberland, C., Oncken, O., Rietbrock, A., Angiboust, S., & Heidbach, O. (2014). Locking of the Chile subduction zone controlled by fluid pressure before the 2010 earthquake. *Nature Geoscience*, *7*(4), 292–296. https://doi.org/10.1038/ngeo2102

Münchmeyer, J., Bindi, D., Leser, U., & Tilmann, F. (2021). The transformer earthquake alerting model: A new versatile approach to earthquake early warning, geophys. *Journal of Intelligence*, *225*(1), 646–656. https://doi.org/10.1093/gji/ggaa609

Ohnaka, M. (1992). Earthquake source nucleation: A physical model for short-term precursors. *Tectonophysics*, *211*(1–4), 149–178. https://doi.org/10.1016/0040-1951(92)90057-d

Oprea, S., Martinez-Gonzalez, P., Garcia-Garcia, A., Castro-Vargas, J. A., Orts-Escolano, S., Garcia-Rodriguez, J., & Argyros, A. (2020). A review on deep learning techniques for video prediction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(6), 2806–2826. https://doi.org/10.1109/TPAMI.2020.3045007

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., & Thirion, B. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Philibosian, B., & Meltzner, A. J., (2020). Segmentation and supercycles: A catalog of earthquake rupture patterns from the Sumatran Sunda megathrust and other well-studied faults worldwide. https://doi.org/10.1016/j.quascirev.2020.106390

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Ren, C. X., Hulbert, C., Johnson, P. A., & Rouet-Leduc, B. (2020). Machine learning and fault rupture: A review. *Advances in Geophysics*, *61*, 57–107. https://doi.org/10.1016/2Fbs.agph.2020.08.003

Rosenau, M., Corbi, F., & Dominguez, S. (2017). Analogue earthquakes and seismic cycles: Experimental modelling across timescales. *Solid Earth*, *8*(3), 1–12. https://doi.org/10.5194/se-8-597-2017

Rosenau, M., Lohrmann, J., & Oncken, O. (2009). Shocks in a box: An analogue model of subduction earthquake cycles with application to seismotectonic forearc evolution. *Journal of Geophysical Research*, *114*(B1), 1–20. https://doi.org/10.1029/2008JB005665

Rouet-Leduc, B., Hulbert, C., Bolton, D. C., Ren, C. X., Riviere, J., Marone, C., et al. (2018). Estimating fault friction from seismic signals in the laboratory. *Geophysical Research Letters*, *45*(3), 1321–1329. https://doi.org/10.1002/2F2017GL076708

Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C. J., & Johnson, P. A. (2017). Machine learning predicts laboratory earthquakes. *Geophysical Research Letters*, *44*(18), 9276–9282. https://doi.org/10.1002/2F2017GL074677

Rouet-Leduc, B., Jolivet, R., Dalaison, M., Johnson, P. A., & Hulbert, C. (2021). Autonomous extraction of millimeter-scale deformation in InSAR time series using deep learning. *Nature Communications*, *12*(1), 6480. https://doi.org/10.1038/s41467-021-26254-3

Sara, U., Akter, M., & Uddin, M. S. (2019). Image quality as-sessment through FSIM, SSIM, MSE and PSNR—A comparative study. *Journal of Computer and Communications*, *7*(03), 8–18. https://doi.org/10.4236/jcc.2019.73002

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., & Woo, W.-C. (2015). Convolutional lstm network: A ma-chine learning approach for precipitation nowcasting. *News in Physiological Sciences*, 802–810.

Shokouhi, P., Girkar, V., Rivière, J., Shreedharan, S., Marone, C., Giles, C. L., & Kifer, D. (2021). Deep learning can predict laboratory quakes from active source seismic data. *Geophysical Research Letters*, *48*(12), e2021GL093187. https://doi.org/10.1029/2021GL093187

Shreedharan, S., Bolton, D. C., Rivière, J., & Marone, C. (2021). Machine learning predicts the timing and shear stress evolution of lab earthquakes using active seismic monitoring of fault zone processes. *Journal of Geophysical Research: Solid Earth*, *126*(7), e2020JB021588. https://doi.org/10.1029/2020JB021588

Shridhar, S., Laumann, F., & Liwicki, M. (2019). A comprehensive guide to bayesian convolutional neural network with variational inference. CoRR, abs/1901.02731.

Sveen, J. K. (2004). *An introduction to matpiv v.1.6.1 Eprint no. 2*. Department of Mathematics, University of Oslo. ISSN 0809-4403. Retrieved from http://urn.nb.no/URN:NBN:no-27806

Trugman, D. T., McBrearty, I. W., Bolton, D. C., Guyer, R. A., Marone, C., & Johnson, P. A. (2020). The spatiotemporal evolution of granular microslip precursors to laboratory earthquakes. *Geophysical Research Letters*, *47*(16), e2020GL088404. https://doi.org/10.1029/2020GL088404

Van Klaveren, S., Vasconcelos, I., & Niemeijer, A. (2020). Predicting laboratory earthquakes with machine learning. arXiv preprint arXiv:2011.06669.

Vinyals, O., Toshev, A., Bengio, S., & Erhan, D. (2015). Show and tell: A neural image caption generator. In *2015 IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 3156–3164).

Voulodimos, A., Doulamis, N., Doulamis, A., Protopapadakis, E., & Andina, D. (2018). Deep learning for computer vision: A brief review. *Computational Intelligence Neuroscience*, *2018*, 1–13. https://doi.org/10.1155/2018/7068349

19448007, 2022, 15, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022GL099632 by CochraneItalia, Wiley Online Library on [11/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Wallace, L. M. (2020). Slow slip events in New Zealand. *Annual Review of Earth and Planetary Sciences*, *48*(1), 175–203. https://doi.org/10.1146/annurev-earth-071719-055104

Wang, K., Johnson, C. W., Bennett, K. C., & Johnson, P. A. (2021). Predicting fault slip via transfer learning. *Nature Communications*, *12*(1), 7319. https://doi.org/10.1038/s41467-021-27553-5

Wang, K., Johnson, C. W., Bennett, K. C., & Johnson, P. A. (2022). The temporal limits of predicting fault failure. arXiv preprint. https://doi.org/10.48550/arXiv.2202.0389

Zhou, Y., Dong, H., & El Saddik, A. (2020). Deep learning in next-frame prediction: A benchmark review. *IEEE Access*, *8*, 69273–69283. https://doi.org/10.1109/ACCESS.2020.2987281

Zhu, W., & Beroza, G. C. (2019). PhaseNet: A deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, *216*(Issue 1), 261–273. https://doi.org/10.1093/gji/ggy423

References From the Supporting Information

Atluri, G., Karpatne, A., & Kumar, V. (2018). "Spatio-temporal data mining: A survey of problems and methods". *ACM Computing Surveys*, *51*(4).

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. In *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council* (Vol. *5*, pp. 157–66). https://doi.org/10.1109/72.279181

Breiman, L. (1996). "Bagging predictors". *Machine Learning*, *24*(2), 123–140. CiteSeerX 10.1.1.32.9399. S2CID 47328136. https://doi.org/10.1007/BF00058655

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise In *Proceedings of the second international conference on knowledge discovery and data mining* (Vol. 96, pp. 226–231). AAAI Press.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *"Elements of Statistical Learning"* (Edn. 2, pp. 592–593). Springer.

Ho, T. K. (1998). "The random subspace method for constructing decision forests,". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(8), 832–844. https://doi.org/10.1109/34.709601

Keskar, N. S., Mudigere, D., Nocedal, J., Smelyanskiy, M., & Tang, P. T. P. (2017). On large-batch training for deep learning: Generalization gap and sharp minima. In *ICLR*. https://doi.org/10.48550/arXiv.1609.04836

Lu, L., Shin, Y., Su, Y., & Karniadakis, G. E. (2020). Dying ReLU and initialization: Theory and numerical examples. *Communications in Computational Physics*, *28*(5), 1671–1706. https://doi.org/10.4208/cicp.OA-2020-0165

Pascanu, R., Mikolov, T., & Bengio, Y. (2013). On the difficulty of training recurrent neural networks. In *Proceedings of the 30th International conference on machine learning 16-21 June, 2013*. (pp. 1310–1318). ICML 2013.

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE*, *10*(3), e0118432. https://doi.org/10.1371/journal.pone.0118432

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*(56), 1929–1958.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). ''Image quality assessment: From error visibility to structural similarity". *IEEE Transactions on Image Processing*, *13*, 600–612.

Xiao, C., Chen, N., Hu, C., Wang, K., Xu, Z., Cai, Y., et al. (2019). A spatiotemporal deep learning model for sea surface temperature field prediction using time-series satellite data. *Environmental Modelling and Software*, *120*, 104502. https://doi.org/10.1016/j.envsoft.2019.104502

19448007, 2022, 15, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022GL099632 by CochraneItalia, Wiley Online Library on [11/09/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License