Predicting rainfall induced slope stability using Random Forest regression and synthetic data

Jamalinia, Elahe; Sadeghi Tehrani, Faraz; Steele-Dunne, Susan; Vardon, Phil

DOI
10.1007/978-3-030-60713-5_24

Publication date
2020

Document Version
Final published version

Published in
Understanding and Reducing Landslide Disaster Risk.

Citation (APA)
Jamalinia, E., Sadeghi Tehrani, F., Steele-Dunne, S., & Vardon, P. (2020). Predicting rainfall induced slope stability using Random Forest regression and synthetic data. In Ž. Arbanas , P. T. Bobrowsky , K. Konagai , K. Sassa , & K. Takara (Eds.), *Understanding and Reducing Landslide Disaster Risk. : Volume 6. Specific Topics in Landslide Science and Applications* (Vol. 6, pp. 223-229). Springer. https://doi.org/10.1007/978-3-030-60713-5_24

Predicting Rainfall Induced Slope Stability Using Random Forest Regression and Synthetic Data

Elahe Jamalinia, Faraz S. Tehrani, Susan C. Steele-Dunne, and Philip J. Vardon

Abstract

Water fluxes in slopes are affected by climatic conditions and vegetation cover, which influence the effective stress and stability. The vegetation cover is the intermediate layer between the atmosphere and the slope surface that alter water balance in the slope through evapotranspiration and leaf interception. This paper studies the data-driven approach for predicting the macro stability of an example grass-covered dike based on actual data and also synthetic data provided by numerical modelling. Two numerical models are integrated in this study. The water balance in the root zone is simulated through a crop model, whereas the hydro-mechanical and safety analysis of the example dike is done using a two-dimensional Finite Element model. The considered period for these analyses is 10 years (3650 daily instances) which will be used to generate a time-series dataset for a secondary dike in the Netherlands. The features included in the dataset are parameters that (i) have a meaningful relationship with the dike Factor of safety (FoS), and (ii) can be observed using satellite remote sensing. The output dataset is used to train a Random Forest regressor as a supervised Machine Learning (ML) algorithm. The results of this proof-of-concept study indicate a strong correlation between the numerically estimated FoS and the ML-predicted one. Therefore, it can be suggested that the utilized parameters can be used in a data-driven predictive tool to identify vulnerable zones along a dike without a need for running expensive numerical simulations.

Keywords

Slope stability • Vegetation • Machine learning

Introduction

The main components of flood protection system in the Netherlands are primary and secondary dikes with the total length of more than 18,000 km. The condition of these engineering structures is assessed based on the infrequent visual inspections usually through ground-based observations. This current method can be systematically augmented by using Earth Observation (EO) data to evaluate the dike condition (Jamalinia et al. 2019a, b; Özer et al. 2018). One crucial aspect of slope stability analysis is the identification of critical points along the slope.

In geotechnical engineering, the analysis and prediction of (in)stabilities is of great importance; however, often little attention is paid to the transient conditions due to vegetated cover and interaction with the environment. This is due to the computational intensity and difficulty in collecting in situ information on the condition of the slope.

Synthetic data driven approaches based on Machine Learning (ML) can be used to develop an efficient estimation of the slope condition and speed up the assessment process, even at the regional scale. In recent years, ML methods have been used in several studies for predicting slope (in)stability (Ada and San 2018; Ghorbanzadeh et al. 2019; Lin et al. 2018; Pourghasemi and Rahmati 2018).

In this research, a Random Forest (RF) approach is used to build and train an ML model on 3650 synthetic data points produced by an integrated crop-geotechnical model on an example geometry. The results show the potential

E. Jamalinia (✉) · S. C. Steele-Dunne · P. J. Vardon
Faculty of Civil Engineering and Geoscience, Delft University of Technology, Stevinweg 1, Delft, 2600, The Netherlands
e-mail: e.jamalinia@tudelft.nl

S. C. Steele-Dunne
e-mail: s.c.steele-dunne@tudelft.nl

P. J. Vardon
e-mail: p.j.vardon@tudelft.nl

F. S. Tehrani
Geo Department, Deltares, Delft, 2600, The Netherlands
e-mail: Faraz.Tehrani@deltares.nl

Ž. Arbanas et al. (eds.), *Understanding and Reducing Landslide Disaster Risk*, ICL Contribution to Landslide Disaster Risk Reduction,
https://doi.org/10.1007/978-3-030-60713-5_24

application of Earth Observation (EO) data for identifying the vulnerable slopes (locations along the slope) without the need for repeating expensive numerical simulations.

Methodology

An integrated crop-geotechnical model (Jamalinia et al. 2020a) is used in this research to calculate the Factor of Safety (FoS) of a dike under realistic climate and vegetation conditions for ten years (daily analysis). The results are used to study the possibility of using ML algorithms to forecast a slope condition based on the observable data from climate, vegetation and slopes.

Numerical Method

In Fig. 1 the geometry of the example dike is shown. This idealised dike is a typical regional Dutch dike (de Vries 2012), which is covered by permanent grass over the surface of the dike with a fixed depth of root zone, 40 cm (shown as green area in Fig. 1). It is assumed that the base boundary is an impermeable layer, while other boundaries of the dike are assumed to be permeable, meaning that the left and right sides have a fixed phreatic surface and the top boundary has a temporal precipitation/evaporation flux applied.

Since standard geotechnical models do not simulate the (dynamic) effects of vegetation, i.e. evapotranspiration and leaf interception, on mass balance and thereby slope stability, the current research utilises an integrated crop-geotechnical model developed by the authors by integrating two existing models (Jamalinia et al. 2020a), although other academic modelling approaches have considered various aspects of the impact of vegetation (Elia et al. 2017). Using this numerical approach enables the study of climatic and vegetation conditions on the stability. The influence of the soil cracking, due to droughts and reducing shear strength, is included in our previous studies (Jamalinia et al. 2019b, 2020a, b). The workflow (Fig. 2) is controlled by Python and explained in detail.

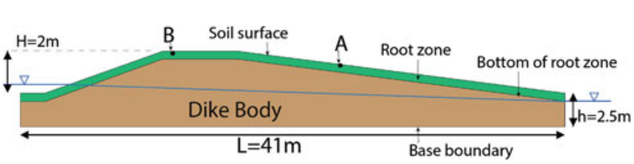

Fig. 1 Geometry representing boundary, root zone layer, and the analysis point

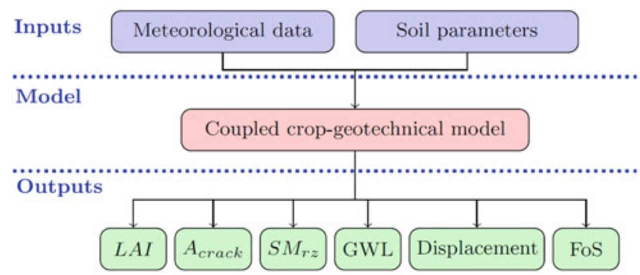

Fig. 2 Flow chart of numerical modeling procedure

The meteorological data (e.g. rain and temperature) and soil parameters are inputs for the integrated crop-geotechnical model. The climate data was obtained from the Royal Netherlands Meteorological Institute (KNMI) at Schiphol Airport station (Amsterdam), which is close (circa 9 km) to the location of the actual dike. The major outputs from the 1D crop model (LINGRA) (Bouman et al. 1996; Rodriguez et al. 1999), shown in Fig. 2 are: Leaf Area Index (LAI), area of leaves divided by the area of ground; crack area (A_{crack}); average soil moisture in the root zone (SM_{rz}). The major outputs from the 2D geotechnical model, Plaxis (2018), are the ground water level (GWL), surface displacement and FoS.

The input parameters for the crop model and the geotechnical model are listed in Tables 1 and 2, respectively.

Data Driven Method

In this study the results of 3650 realisations (simulations) from the integrated crop-geotechnical model, each simulating a period of 10 years from 2009 to 2019, are used in training and testing a RF regressor to predict the safety condition of the example dike. The Random Forest approach is one of the most widely used ensemble learning algorithms. The RF (Breiman 2001) constructs individual Decision Trees (DTs) based on bagging, using bootstrap sampling where samples are taken randomly with replacement from the training set (Qi and Tang 2018). In the DT method the data is divided into smaller subsets and a tree is expanded until the leaf node, where the decision is made about the target value or class in DT regression or DT classification. As the RF method uses the training dataset to create multiple decision trees, the variance of the final model is reduced and then it is less sensitive to over-fitting (Burkov 2019). Each decision tree of the RF predicts an output and RF regression models take the average of all the individual decision tree estimates.

Table 1 Input parameters used for the crop model, modified after Jamalinia et al. (2020a)

Parameters		Value	Unit
Soil	Water content at field capacity (prior to cracking)	0.29	$cm^3 water/cm^3 soil$
	Water content at the wilting point below that wilting starts	0.12	$cm^3 water/cm^3 soil$
	Critical water content below that transpiration is reduced	0.05	$cm^3 water/cm^3 soil$
	Maximum drainage from root zone to lower layers	50	mm/day
Vegetation	Specific Leaf Area: leaf area over leaf mass	0.025	m^2/g
	Remaining LAI after mowing	0.8	$m^2 leaf/m^2 soil$
	Critical leaf area beyond that self-shading occurs	4	$m^2 leaf/m^2 soil$

Table 2 Input parameters for the geotechnical model, modified after Jamalinia et al. (2020a)

Parameters		Value		Unit
		Root zone	Dike body	
Constitutive model (Mohr-Coulomb)	Saturated unit weight	20	12	kN/m^3
	Friction angle (prior to cracking)	23	23	°
	Cohesion (prior to cracking)	2	2	kPa
	Dilatancy angle	0	0	°
	Young's modulus	10	20	MPa
	Poisson's ratio	0.3	0.2	–
	Initial void ratio	0.67	1.2	–
Hydraulic model (van Genuchten*)	Hydraulic conductivity	0.14	0.03	m/day
	Scale parameter (α)	1.47	1.38	m^{-1}
	Fitting parameter (n)	1.97	1.32	–
	Fitting parameter (m)	0.87	−1.24	–

*Hysteresis is not considered

Results

Numerical Simulations

Example temporal inputs and outputs of the numerical simulations are shown in Fig. 3. Time-series of rainfall and temperature (TMP) as climate data is shown in Fig. 3a, b, respectively.

In Fig. 3c–e the variation of crop model outputs over time is shown. Considering the worst-case scenario, it is assumed that cracks do not close during the wet periods, but only expand during unprecedented drier conditions (Fig. 3c). The percentage of the cracked soil area increases in such dry periods and its area remains constant until the next drier period. It is assumed that the cracking happens only in the root zone area with the maximum depth of 40 cm, equal to the root zone depth. The sudden decrease in LAI on 15 June and 15 August annually in Fig. 3d shows mowing events, which were imposed in the crop model based on the mowing schedule of secondary dikes in the Netherlands (Jamalinia et al. 2019a). A higher presence of cracks causes a reduction in the rate of LAI growth after mowing. In the summer of 2018, according to Fig. 3e, the root zone experienced the driest condition during the previous 10 years, and the crack area reached the maximum value during the simulation period. In Aug. 2018, the root zone soil moisture (SM_{rz}) reached its minimum value, and it can be seen that vegetation could not easily re-grow after mowing.

The temporal variation of absolute surface displacement $|U_A|$ at point A (Fig. 1) and FoS are selected as outputs of the FEM model (2D geotechnical model), shown in Fig. 3f, g, respectively. Displacement at point A follows the variation of SM_{rz}, which reflects the response of the $|U_A|$ to the climate and vegetation conditions. The combined effect of rainfall, LAI variations, and crack area influence the water flux into the dike which caused temporal variations of FoS. The maximum crack area in August 2018 and very low LAI

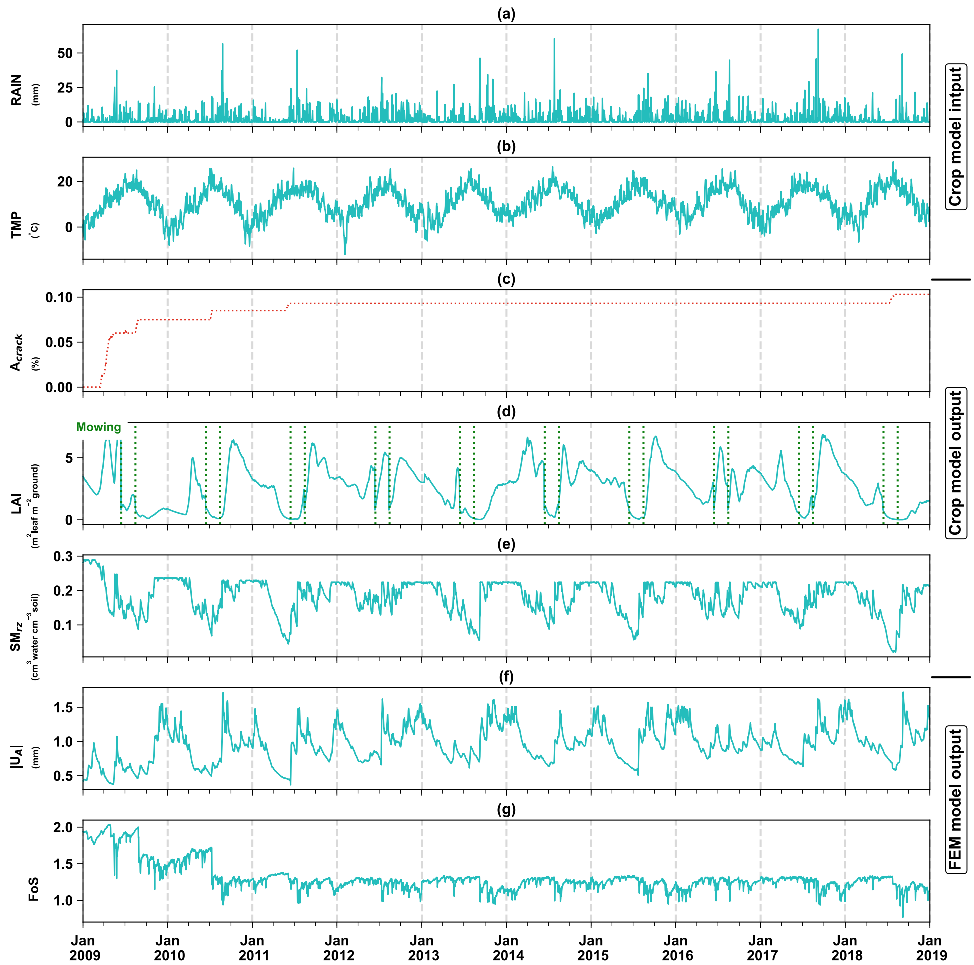

Fig. 3 Time-series of inputs and outputs from the developed integrated models for 10 years

(almost bare soil) caused an increased infiltration due to heavy precipitation event. In addition, the soil had the lowest shear strength due to the maximum crack area, which together led to the minimum FoS.

Feature Selection

There are two criteria to select the features in this study to train the RF regressor: the feature (1) has a strong, meaningful relationship with the FoS; (2) is observable remotely, so experts can monitor these parameters and assess the slope condition based on that feature. Therefore, the features in this study are from (i) climate: rainfall and temperature, (ii) vegetation: LAI, observing anomalies in vegetation could be used as an indicator to distinguish whether a dike is significantly cracked; (iii) slope surface displacement: it can be used as a proxy for both saturation (short term changes) and for accumulation of cracks (long term changes), although long term changes may also indicate subsidence or other processes (Jamalinia et al. 2020a). Using the PSInSAR method Ferretti et al. (2001), it is possible to map surface deformation with millimetre precision.

The lag correlation between pair of key parameters is plotted in Fig. 4. A positive lag means the second term causes the first one. There is a 15 days lag between LAI and Sat_{rz}, which means that root zone saturation affects vegetation growth most after 15 days. There is a strong correlation between saturation at point A (Sat_A) and $|U_A|$, which shows that surface displacement is responsive to the available water in the root zone, which is mentioned in the time-series result as well. Existing correlation in Fig. 4a, b suggest that using

LAI and $|U_A|$ could be good indicators for available water near to the dike surface, and both are reasonably easy to monitor remotely, unlike the SM. The negative correlations between (Fos, SM_{rz}) and (Fos, $|U_A|$) shown in Fig. 4c, d, suggest using $|U_A|$ as an indicator to estimate safety. The cumulative rainfall during the 35 days before an event day, Rain.cu_35 (Fig. 4f), has a stronger correlation with FoS than rainfall on the same day (Fig. 4e). This period has showed the best predicted FoS among other periods (Jamalinia et al. 2020c). Therefore, in the RF analysis a history of rainfall is considered.

Random Forest Regression

The 10-year simulation results from 2009 to 2019 are used to build a predictive model using the mentioned features in previous section. The data set is split to training set (70% of dataset) and testing set (30%), and the number of trees in the RF algorithm set to be 1000.

Here the ability of RF for real time prediction is tested. The features are selected from the same day at which the FoS is calculated, except for rainfall that accumulation during last 35 days is considered. The feature importance values are plotted in Fig. 5 which are derived from the RF regressor as a result of training processes. It turns out that the absolute surface displacement $|U_A|$ has the highest importance (0.52). LAI and cumulative precipitation during the last 35 days (Rain.cu-35) have almost the same feature importance of 0.2, and daily temperature (TMP), has the least effect on FoS and therefore its prediction. As mentioned before, according to time-series results and correlation, vegetation growth and displacement are affected by precipitation, so precipitation impact is embedded in LAI and $|U_A|$.

The predicted FoS from the RF method is plotted against the calculated FoS from the FEM model in the numerical

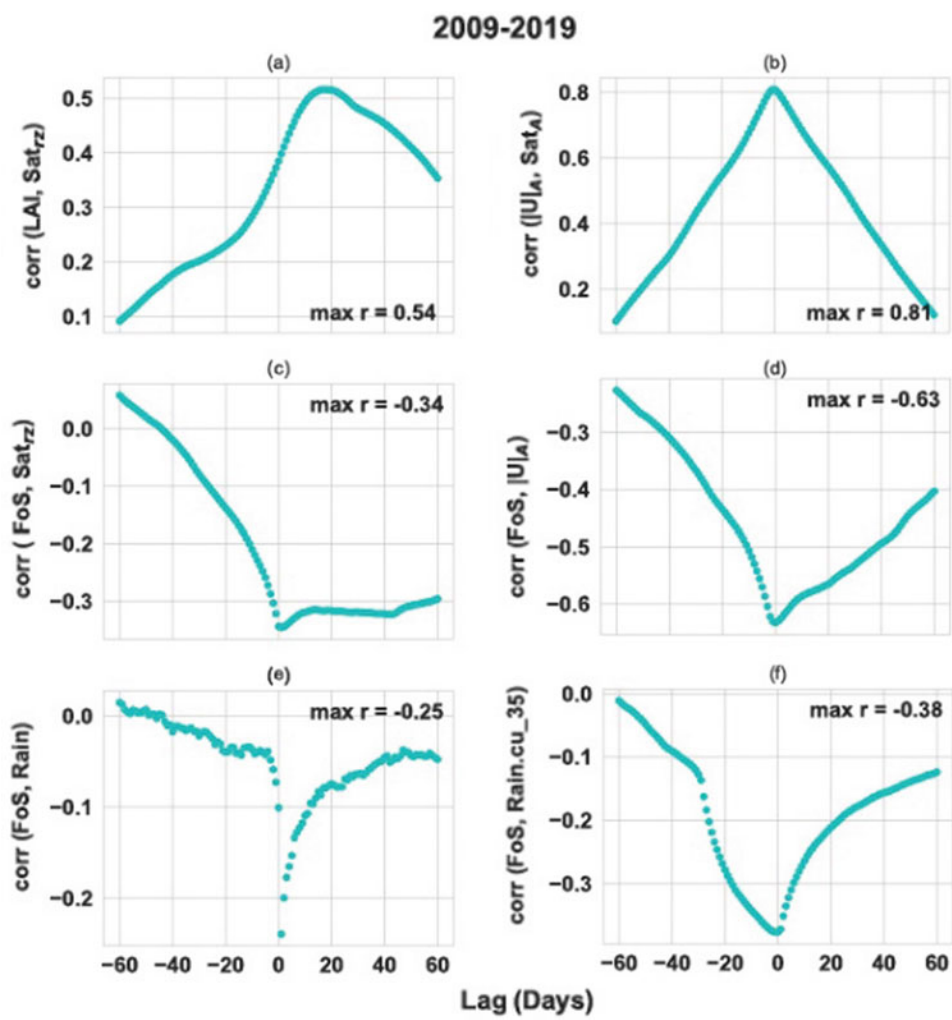

Fig. 4 Lag correlation between pair of key parameters

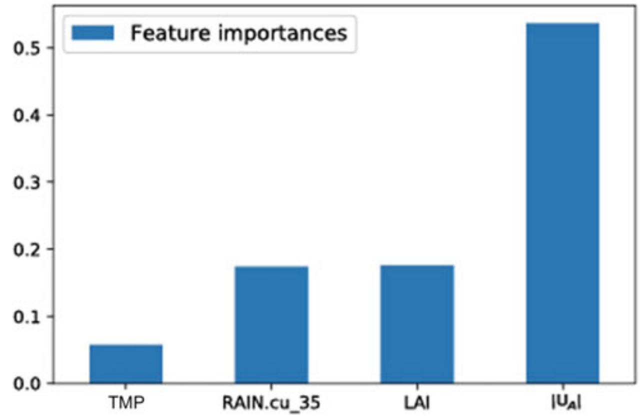

Fig. 5 Feature importance out of RF regression

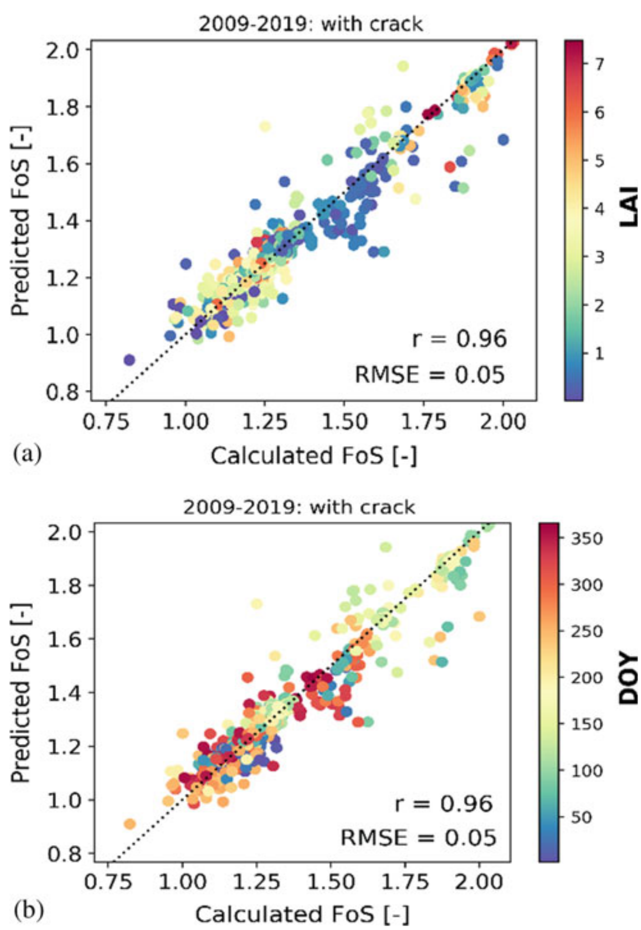

Fig. 6 Correlation between real time predicted FoS and calculated FoS. Scatters are colorized by (a) LAI, (b) Day of Year (DOY)

analysis in Fig. 6. The results show that the RF model yields precise estimation for assessing dike safety only from the observable data. The scatters are colorized by LAI values and day of year (DOY) which suggests that usually at low LAI and winter period (e.g. when LAI is lower than 2 during November to February) outliers occur, where there is not enough energy available for vegetation to grow.

In another analysis, the time-series prediction has been carried out to investigate the ability of the RF method to predict the future FoS from historical data. The training data set has been collected from the first 70% of the dataset and the remaining 30% used as the test set. So, the model is built based on avoiding random train, test split selection. The predicted and calculated FoS are plotted against time in Fig. 7. The temporal trends are well represented, with deviation in the low values and after the cracking event during the summer of 2018. The results demonstrate that RF can be used as a promising method to predict slope condition using observable input data: meteorological data, vegetation and surface displacements. Therefore, doing a numerical analysis for a slope and calculating FoS for a time period would help experts to assess the condition of the slope in future using these observable parameters, without the need to repeat time-consuming simulations.

Conclusion

This proof of concept study investigates the potential use of observable data in predicting slope condition. A one-way coupled model framework composed of a crop model and a geotechnical model was used to calculate the factor of safety of an idealised dike covered with grass for 10-year period simulation. The existing correlation between selected parameters assisted in the feature selection for this data-driven study approach, as well as an assessment of whether they are remotely observable. The supervised ML algorithm, Random Forest (RF), has been used for predicting FoS using key parameters such as: precipitation, temperature, LAI and surface displacement at a selected point on the example dike. The RF algorithm results in a prediction with high accuracy (RMSE = 0.05). Among the features, surface displacement shows the highest feature importance. It is shown that displacement is responsive to the amount of water in the root zone which is affected by the climate and vegetation condition. The results of this study show the potential use of EO data for real time monitoring of slopes

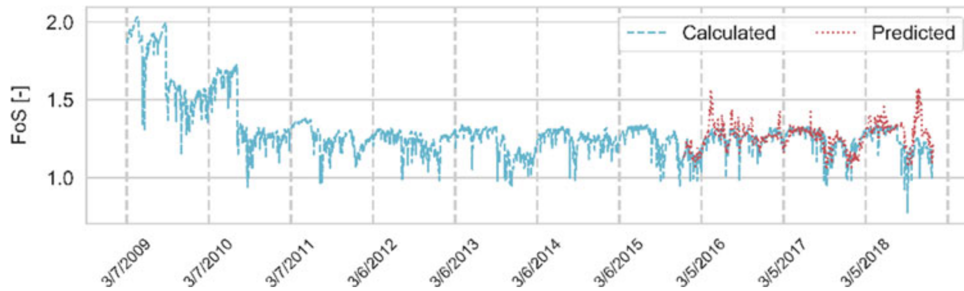

Fig. 7 Time-series prediction of FoS

and detecting the vulnerable locations along the slopes. The results show some deviation, probably due to the strong non-linearities in the physical model, therefore the worth of the RF model is to identify weak areas and allow further detailed investigation.

Acknowledgements This work is part of the research program Reliable Dikes with project number 13864 which is financed by the Netherlands Organisation for Scientific Research (NWO).

References

Ada M, San BT (2018) Comparison of machine-learning techniques for landslide susceptibility mapping using two-level random sampling (2LRS) in Alakir catchment area, Antalya, Turkey. Nat Hazards 90 (1):237–263

Bouman BAM, Schapendonk AHCM, Stol W, van Kraalingen DWG (1996) Description of the growth model LINGRA as implemented in CGMS. In: Quantitative Approaches in System Analysis, vol 7. DLO Research Institute for Agrobiology and Soil, Wageningen, the Netherlands

Breiman L (2001) Random forests. Mach Learn 45(1):5–32. Springer

Burkov A (2019) The hundred—page machine learning. In: (part of title) And add publication details, vol 1. ndriy Burkov Quebec City, Canada

de Vries G (2012) Monitoring droogteonderzoek veenkaden, Report no. 1203255-006-GEO-0001-gbh, Deltares, Delft

Elia G et al (2017) Numerical modelling of slope–vegetation–atmosphere interaction: an overview. Q J Eng Geol Hydrogeol 50:249–270

Ferretti A, Prati C, Rocca F (2001) Permanent scatterers in SAR interferometry. IEEE Trans Geosci Remote Sens 39(1):8–20

Ghorbanzadeh O, Blaschke T, Gholamnia K, Meena SR, Tiede D, Aryal J (2019) Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. Remote Sens 11(2):196

Jamalinia E, Vardon PJ, Steele-Dunne SC (2019a) Can vegetation indices predict slope (stability) conditions? In: Geophysical Research Abstracts, vol 21, pp 55–64

Jamalinia E, Vardon PJ, Steele-Dunne SC (2019b) The effect of soil–vegetation–atmosphere interaction on slope stability: a numerical study. Environ Geotech ahead of print 1–12

Jamalinia E, Vardon PJ, Steele-Dunne SC (2020a) The impact of evaporation induced cracks and precipitation on temporal slope stability. Comput Geotech 122:103506

Jamalinia E, Vardon P, Steele-Dunne S (2020b) Use of displacement as a proxy for dike safety. Proc Int Assoc Hydrol Sci 382(1):481–485

Jamalinia E, Tehrani FS, Steele-Dunne SC, Vardon PJ (2020c) A data-driven approach for stability forecasting of dikes. in prepara

Lin Y, Zhou K, Li J (2018) Prediction of slope stability using four supervised learning methods. IEEE Access 6:31169–31179

Özer IE, van Leijen FJ, Jonkman SN, Hanssen RF (2018) Applicability of satellite radar imaging to monitor the conditions of levees. J Flood Risk Manag 12(Suppl. 2):e12509

Plaxis BV (2018) PLAXIS reference manual 2018. Delft, Netherlands

Pourghasemi HR, Rahmati O (2018) Prediction of the landslide susceptibility: which algorithm, which precision? Catena 162:177–192

Qi C, Tang X (2018) Slope stability prediction using integrated metaheuristic and machine learning approaches: a comparative study. Comput Ind Eng 118(February):112–122

Rodriguez D, van Oijen M, Schapendonk AHMC (1999) LINGRA-CC: a sink-source model to simulate the impact of climate change and management on grassland productivity. New Phytol v 144(2):359–368