The Seventh Asian-Pacific Symposium on Structural Reliability and Its Applications (APSSRA2020)
October 4–7 2020, Tokyo, Japan
T. Takada, I. Yoshida & T. Itoi (editors)

DTM and rainfall-based landslide susceptibility analysis using machine learning: A case study of Lantau Island, Hong Kong

H.J. Wang[1], L.M. Zhang[2] and T. Xiao[3]

[1]*Ph.D. Research Assistant, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Kowloon, Hong Kong. Email: h.wang@connect.ust.hk*

[2,*]*Corresponding Author. Chair Professor, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Kowloon, Hong Kong. Email: cezhangl@ust.hk*

[3] *Postdoctoral Researcher, Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Kowloon, Hong Kong. Email: xiaote@ust.hk*

Abstract: Landslide susceptibility analysis is an essential part of landslide risk assessment and hazard mitigation. A high-resolution digital terrain model (DTM) and its derivatives can precisely capture and characterize the ground features of historical landslide locations. Also, machine learning has been proven to be a promising tool in landslide susceptibility analysis. In this paper, a machine learning based case study at Lantau Island, Hong Kong is performed to investigate the feasibility of landslide susceptibility analysis utilizing DTM and rainfall data. A DTM with a resolution of 2 m, natural terrain landslide records and hourly rainfall data of Hong Kong from 1984 to 2010 are used in this study. Results of this study reveal excellent capability of machine learning approaches in landslide susceptibility mapping based on DTM and rainfall data. The susceptibility map produced in this study can be an important tool for identifying zones where landslide mitigation measures are needed.

Keywords: Landslide, Landslide susceptibility, Machine learning, Big data

1. Introduction

Landslide susceptibility analysis provides indispensable references for assessing landslide risk and designing hazard mitigation measures (Fell et al., 2008; van Westen et al., 2008; Reichenbach et al., 2018). Hong Kong, as a coastal city, is frequently affected by severe weather phenomena like typhoons and strong rainstorms, making it highly susceptible to natural terrain landslides (Gao et al., 2017a, b). From 1984 to 2010, the Geotechnical Engineering Office (GEO) of the Hong Kong SAR interpreted 11,382 natural terrain landslides from annual aerial photographs of Hong Kong (GEO, 1996; Maunsell-Fugro Joint Venture and GEO, 2007), of which 62% are open hillslope landslides, 37% are channelized debris flows and the remaining 1% are costal landslides.
Several pioneering researchers and engineers have investigated the susceptibility pattern of Hong Kong at different scales. Evans and King (1998) produced the first territory-wide landslide map in Hong Kong based on a correlation of landslide susceptibility with slope gradient and geology. Dai and Lee (2001) used logistic regression and geographic information system (GIS) to assess the susceptibility of Lantau Island of Hong Kong. Lee et al. (2001) investigated the feasibility of applying artificial neural network (ANN) to classify landslide susceptible and non-susceptible areas in the middle part of Lantau Island. Chau and Chan (2005) studied the regional bias of the landslide data in generating the landslide susceptibility of Hong Kong Island using logistic regression. Yao et al. (2008) utilized support vector machine (SVM) in evaluating the landslide susceptibility in a subarea of New Territories. Wang et al. (2019) proposed a novel physically-based landslide susceptibility updating method with a case study in western Lantau Island. Besides, with the advent of modern remote sensing and machine learning techniques, high-resolution digital terrain model (DTM) and modern machine learning algorithms become available and have been proven to be powerful tools in landslide studies (Li et al., 2018; Li et al., 2019; Wang and Zhang, 2019; Wang et al., 2020; Xiao and Zhang, 2020). High-resolution DTM carries the capability of charactering precious ground features of both landslide prone and non-prone areas. The application of high-precision DTM in landslide susceptibility analysis has not been fully explored inside Hong Kong. Hence, this study takes the largest outlying island of Hong Kong as the study area and aims to (a) investigate the robustness of applying only topographic and rainfall data to assess landslide susceptibility using machine learning and (b) compare the performance of different machine learning approaches and produce a susceptibility map of Lantau Island.

2. Data and methods

2.1 Data

The study area is the whole Lantau Island and the boundary is shown in Figure 1. Lantau Island is the largest outlying island at southwestern Hong Kong, covering a total area of 147 km^2. The data for this study are composed of three parts: landslide inventory from 1984 to 2010 (Figure 1), 2-m resolution DTM based on Light Detection and Ranging (Lidar) product of 2010 and hourly rainfall data from 1984 to 2010. First, the landslide inventory data consists of 4616 natural terrain landslides which are extracted from the Enhanced Natural Terrain Landslide Inventory of GEO. All landslide records are classified into three categories according to their ground characteristics; namely, open hillslope landslides (2385 cases), channelized debris flow (2192 cases) and coastal landslides (39 cases).

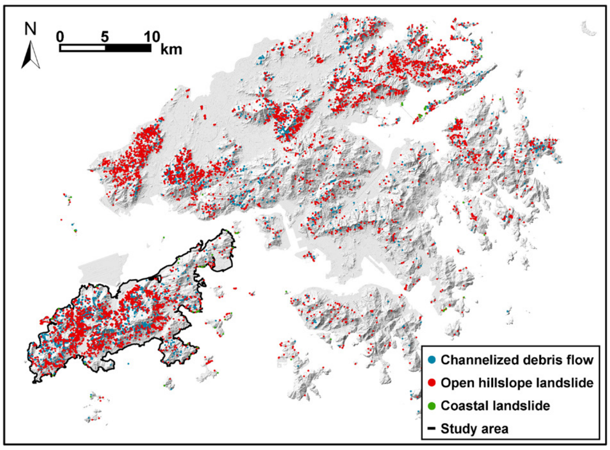

Figure 1. Spatial distribution of landslide incidents of Hong Kong from 1984 to 2010.

In addition, the yearly distribution of landslide incidents in Lantau Island is examined in Figure 2. It can be found that 1992, 1993, 1999 and 2008 are the four years which have the highest numbers of natural terrain landslide events.

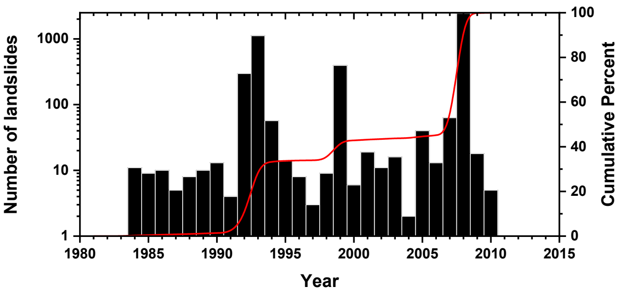

Figure 2. Yearly distribution of landslide incidents in Lantau Island.

Second, the DTM utilized in this study is resampled to a resolution of 2 m based on the original product of air-borne Lidar scan conducted from December 2010 to January 2011. As shown in Figure 3 (a-d), apart form the DTM, three more data layers are also produced with the GIS platform, namely, slope gradient (SG), aspect and curvature. These three predictors have been proven to be important factors related to landslide hazard (van Westen et al., 2008).

Third, the hourly rainfall data are provided by the Hong Kong Observatory (HKO) and GEO. With 50 rain gauges of HKO and 91 rain gauges of GEO, the total 141 rain gauges produced comprehensive historical rainfall data over the entire Hong Kong. In order to consider the worst historical rainfall conditions, the rainfall data in this study is transformed to the form of 4-hour (4h) and 24-hour (24h) maximum rolling rainfall (MRR). The 4h and 24h MRR of each rain gauge is first calculated with the original hourly rainfall data, then the inverse distance weighting approach is used to interpolate the rainfall for the entire Lantau Island. The detailed distributions of 4h and 24h MRR are presented in Figure 3 (e and f). It can be found that areas with the highest 4h MRR are located in the western and middle part of Lantau Island. The highest areas for the 24h MRR are similar with those for the 4h MRR but the middle peak has moved slightly towards the north.

In total, six data layers are generated for the susceptibility analysis of this study; namely, DTM, SG, aspect, curvature, 4h MRR and 24h MRR. The six data layers can then form a data cube, as each data layer consists of 12,637 x 9737 cells with each cell representing an area of 2 x 2 m^2, the data cube's dimensions are 12,637 x 9737 x 6.

Then, landslide and non-landslide samples can then be extracted form the data cube for learning. For landslide samples, the landslide cells are first obtained by mapping the landslides on the data layer and then finding which cells contains landslides. If a cell contains a landslide, it is then defined as a landslide cell. Then, the data of the landslide cell and its neighboring region from the same spatial location in the data cube constitute a landslide data cube sample (H x H x *6, H is the side length*). Similarly, non-landslide data cube samples are extracted from areas with no landslide records. To determine the side length H, the length records of landslide source area are summarized in Table 1, a value of 30 m is chosen which could accommodates 99.28% of landslide incidents.

Table 1. Summary of length of landslide sources.

Length intervals of landslide source area (m)	Number	Proportion
0 to 10	3599	77.97%
10 to 20	877	19.00%
20 to 30	107	2.32%
30 to 117	33	0.72%

Overall, 4589 landslide data cube samples are obtained, and the same number of non-landslide samples are also extracted. Very few costal landslides have a distance less than 30 m from the coastline, so these records are excluded from the databases. By combining both, a dataset for machine learning is constructed, the dataset is then normalized to resolve the inconsistent data layer magnitude problem.

2.2 Machine learning models

Three machine learning algorithms are used in this case study; namely, support vector machine (SVM) with Gaussian kernel, logistic regression (LR), and random forest (RF). The three machine learning algorithms have been widely adopted and accepted in geotechnical applications and landslide susceptibility studies. A 5-fold cross validation method is used for assessing the overall performance of the trained machine learning models.

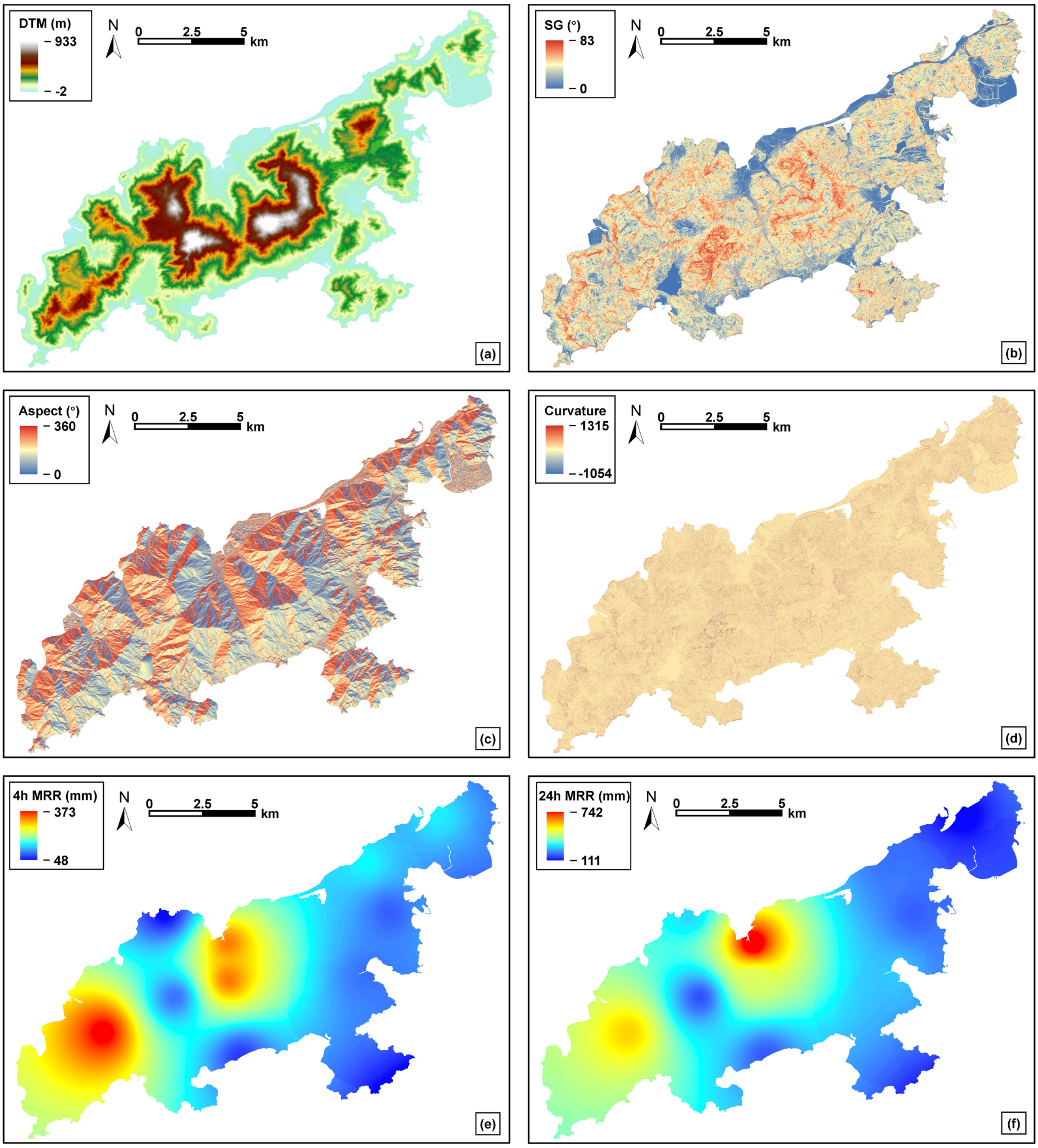

Figure 3. Data layers of the study area: (a) digital terrain model (DTM); (b) slope gradient (SG); (c) aspect; (d) curvature; (e) 4-hour maximum rolling rainfall (4h MRR) and (f) 24-hour maximum rolling rainfall (24h MRR)

2.3 Model evaluation

In this study, the problem is treated as a binary classification task to classify a sample into landslide (positive) and non-landslide (negative) categories. The probability of a sample being positive is taken as its susceptibility, ranging from 0 to 1. The accuracy of a model is given by Eq. 1:

$$Accuracy = (TP + TN) / (TP + FP + TN + FN) \quad (1)$$

where *TP* represents true positive predictions, which means the prediction and sample true category are both positive; *FP* represents the false positive predictions, indicating that the prediction is positive but the sample is indeed a negative one; *TN* are true negative predictions and *FN* are false negative predictions. In addition, a receiver and operating characteristic (ROC) curve is another popular machine learning evaluation tool. By plotting the true positive rate (TPR, Eq. 2) against the false positive rate (FPR, Eq. 3) at different threshold settings, the ROC curve is able to describe relative trade-offs between true positive and false positive of the classifier. From the ROC curve, the area under the curve (AUC) can be calculated, which is equal to the probability that the classifier will give a higher rank to a positive sample compared with a negative one, where both samples are randomly chosen.

$$TPR = TP/(TP+FN) \quad (2)$$

$$FPR = FP/(TN+FP) \quad (3)$$

3. Results

3.1 Performance of machine learning models

The performance of the three trained machine learning models are summarized in Table 2 with the corresponding ROC curves presented in Figure 4. The SVM model with gaussian kernel has the highest accuracy, reaching 86.4%, followed by the RF, 85.5%, and LR, 82.2%. According to the results of AUC, the same ranking can be observed: 0.94 for SVM, 0.92 for RF and 0.89 for LR.

Table 2. Performance summary of machine learning models.

Machine learning models	5-fold cross validation accuracy	Area under the curve
SVM	86.4%	0.94
LR	82.2%	0.89
RF	85.5%	0.92

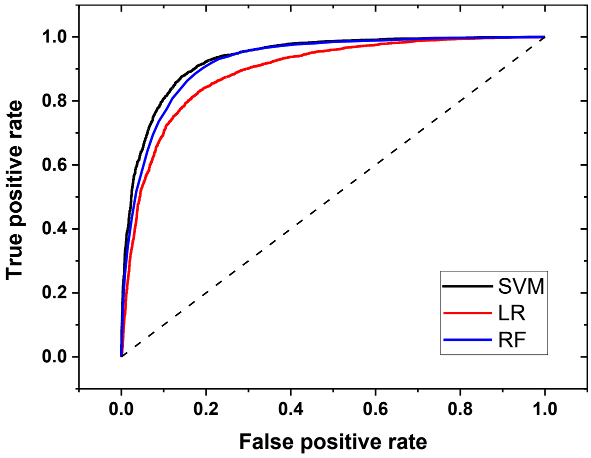

Figure 4. ROC curves for the three trained machine learning models

3.2 Landslide susceptibility map of Lantau Island

As the Gaussian SVM has the highest performance in classifying landslide and non-landslide samples, it is then used to produce the landslide susceptibility map of the entire Lantau Island. The SVM predicted susceptibility map is shown in Figure 5. With reference to Dai et al. (2001), the probability of landside occurrence given by the model is taken as the landslide susceptibility, ranging from 0 to 1. Then, the spatial landslide susceptibility is divided into five classes: (a) very low (0 to 0.1), (b) low (0.1 to 0.3), (c) moderate (0.3 to 0.55), (d) high (0.55 to 0.75) and very high (over 0.75). For historical landslide incidents, 92.42% incidents are located in "high" or "very high" area, indicating that the confidence level of SVM predicted landslide susceptibility map is high.

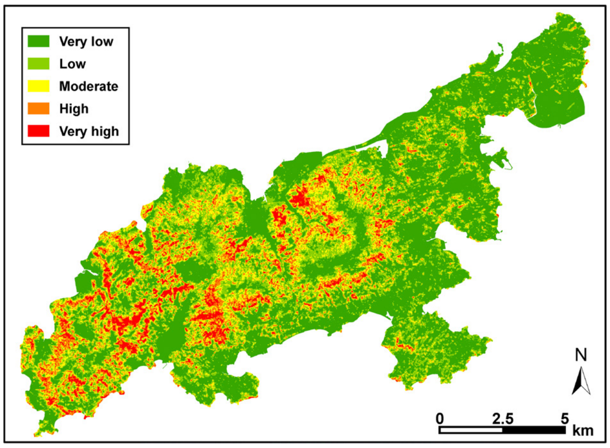

Figure 5. Landslide susceptibility map produced by trained SVM model

4. Summary and conclusions

This study investigated the performance of three machine learning models (i.e., SVM, LR and RF) in the task of landslide susceptibility assessment. The SVM model with Gaussian kernel outperforms the other two in the classification task, achieving an accuracy of 86.4%. Then, the SVM model is used to produce the landslide susceptibility map of Lantau Island. 92.42% of landslide area is successfully classified into "high" or "very high" area of susceptibility, proving the robustness of susceptibility analysis utilizing only DTM and rainfall data. The susceptibility map of Lantau Island produced by this study can serve as an important tool for identifying zones where landslide mitigation measures may be needed.

5. Acknowledgement

This research is supported by the Research Grants Council of the Hong Kong SAR Government (Nos. 16205719, AoE/E-603/18 and 16206217). The authors would also like to thank the Geotechnical Engineering Office of the Hong Kong SAR for providing the DTM and landslide inventory data.

References

Chau, K.T., Chan, J.E., 2005. Regional bias of landslide data in generating susceptibility maps using logistic regression: Case of Hong Kong Island. *Landslides*, 2(4): 280-290.

Dai, F.C., Lee, C.F., 2001. Terrain-based mapping of landslide susceptibility using a geographical information system: a case study. *Canadian Geotechnical Journal*, 38(5): 911-923.

Dai, F.C., Lee, C.F., Li, J., Xu, Z.W., 2001. Assessment of landslide susceptibility on the natural terrain of Lantau Island, Hong Kong. *Environmental Geology*, 40(3): 381-391.

Evans, N.C., King, J.P., 1998. *The Natural Terrain Landslide Study - Debris Avalanche Susceptibility*, GEO Technical Note No. TN 1/98, Geotechnical Engineering Office, Hong Kong Special Administration Region.

Fell, R. et al., 2008. Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning. *Engineering Geology*, 102(3-4): 99-111.

Gao, L., Zhang, L.M., Cheung, R.W.M., 2017a. Relationships between natural terrain landslide magnitudes and triggering rainfall based on a large landslide inventory in Hong Kong. *Landslides*, 15(4): 727-740.

Gao, L., Zhang, L.M., Lu, M.Q., 2017b. Characterizing the spatial variations and correlations of large rainstorms for landslide study. *Hydrology and Earth System Sciences*, 21(9): 4573-4589.

GEO, 1996. *Compilation of a database on landslide consequence*. Submitted by Mitchell, McFarlane, Brentnall & Partners Int. Ltd., Agreement No. GEO 8/95. Geotechnical Engineering Office, Hong Kong Special Administration Region.

Lee, C.F., Ye, H., Yeung, M.R., Shan, X., Chen, G., 2001. AIGIS-based methodology for natural terrain landslide susceptibility mapping in Hong Kong. *Episodes*, 24(3): 150-159.

Li, X.Y., Zhang, L.M., Xiao, T., Zhang, S., Chen, C., 2019. Learning failure modes of soil slopes using monitoring data. *Probabilistic Engineering Mechanics*, 56: 50-57.

Li, X.Y., Zhang, L.M., Zhang, S., 2018. Efficient bayesian networks for slope safety evaluation with large quantity monitoring information. *Geoscience Frontiers*, 9(6): 1679-1687.

Maunsell-Fugro Joint Venture, GEO, 2007. *Final report on compilation of the Enhanced Natural Terrain Landslide Inventory (ENTLI)*. Maunsell-Fugro Joint Venture & Geotechnical Engineering Office, Hong Kong Special Administration Region.

Reichenbach, P., Rossi, M., Malamud, B.D., Mihir, M., Guzzetti, F., 2018. A review of statistically-based landslide susceptibility models. *Earth-Science Reviews*, 180: 60-91.

van Westen, C.J., Castellanos, E., Kuriakose, S.L., 2008. Spatial data for landslide susceptibility, hazard, and vulnerability assessment: an overview. *Engineering Geology*, 102(3-4): 112-131.

Wang, H.J., Zhang, L.M., 2019. Landslide susceptibility updating considering real-time observations. In *Proceedings of Geo-Congress 2019*, Philadelphia, Pennsylvania, March 24-27, 2019. ASCE, Virginia, pp. 107-113.

Wang, H., Zhang, L., Yin, K., Luo, H., Li, J., 2020. Landslide identification using machine learning. *Geoscience Frontiers*, 10.1016/j.gsf.2020.02.012.

Wang, H.J., Xiao, T., Li, X.Y., Zhang, L.L., Zhang, L.M., 2019. A novel physically-based model for updating landslide susceptibility. *Engineering Geology*, 251: 71-80.

Xiao, T., Zhang, L., 2020. Evaluation of performance of engineered slopes under extreme rainstorms. In *Proceedings of Geo-congress 2020*, Minneapolis, Minnesota, February 25-28, 2020. ASCE, Virginia, pp. 737-743.

Yao, X., Tham, L.G., Dai, F.C., 2008. Landslide susceptibility mapping based on support vector machine: A case study on natural slopes of hong kong, china. *Geomorphology*, 101(4): 572-582.