Machine Learning for Source Identification of Dust on the Chinese Loess Plateau

Xin Lin1,2, Hong Chang3, Kaibo Wang3, Guishan Zhang1, and Ganggang Meng1

1School of Earth Sciences and Resources, Chang’an University, Xi'an, China, 2Harquail School of Earth Sciences, Laurentian University, Sudbury, Ontario, Canada, 3Key Laboratory of Loess and Quaternary Geology, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an, China

Abstract The provenance of voluminous eolian dust on the Chinese Loess Plateau (CLP) is still highly debated. Here we apply machine learning methods of support vector machine and convolutional neural network to train models using element compositions of surface sediments from eight potential source regions, accordingly, to determine the dust sources and contributions by classifying the last glacial loess and present interglacial sediments on the CLP. The trained models succeed in differentiating major secondary sources and quantitatively estimating the contributions of both primary and secondary sources at least during the last glacial-interglacial cycle. The understanding that a constant dust source despite changing climate conditions agrees with those derived from Sr-Nd isotopes and U-Pb age spectra. Our observations demonstrate that big geochemical data sets coupled with machine learning technology are fully capable of tracing sources.

Plain Language Summary Central northern China is occupied by a unique landscape called the Loess Plateau which is formed by accumulation of extensive eolian dust (640,000 km² in area and 105 km³ in volume). To understand this majestic Earth surface process, it is of great significance to study the dust sources. In the present study, we use a big geochemical data set and machine learning algorithms to decipher the mystery of where the dust on the Loess Plateau originates. The advantage of the new methods is that they can fully exploit the used big data set. Our results indicate that the contributions of major secondary sources are dominated (approximately 50%) by the recycled Yellow River sediments from the Hetao Graben, followed by materials from inland basins in the northwest of China (>15% for eastern Tibetan Plateau, approximately 15% for Qaidam Basin, and >5% for Tarim Basin, Junggar Basin, and Alxa Plateau). Two primary sources, that is, northern Tibetan Plateau and Central Asian Orogenic Belt, thus contribute about 80% and 20%, respectively, to the dust deposition. An unchanged dust source structure during the last glacial-present interglacial cycle structure is also recognized. These findings strengthen our understanding of the formation of the Loess Plateau.

1. Introduction

The alternating deposition of loess and paleosols during glacial-interglacial cycles on the Chinese Loess Plateau (CLP) in Central China serves as one of the most complete archives of regional climate change (An, 2000; An et al., 2001; Guo et al., 2002). Constraining the dust provenance is thus of profound paleoclimatic significance in Central Asia (Licht, Dupont-Nivet, et al., 2016; Liu & Ding, 1998). Despite its importance, the dust provenance remains highly debated and is further complicated because the dust sedimentation has been demonstrated to be episodic, not continuous (Stevens et al., 2016), which relates to the spatial-temporal variability (Kapp et al., 2015) at both tectonic and glacial-interglacial timescales (Sun et al., 2020).

Persistent endeavors have been made to determine the dust sources primarily by investigating (1) distribution of loess thickness and grain size (e.g., Liu, 1985; Yang & Ding, 2017); (2) compositions of major, trace, rare earth elements (REEs) and Sr-Nd, oxygen, and uranium isotopes (e.g., Chen & Li, 2013; Jahn et al., 2001; Li et al., 2018; Yan et al., 2014); (3) zircon U-Pb age spectra (e.g., Licht, Dupont-Nivet, et al., 2016; Nie et al., 2015; Pullen et al., 2011; Xiao et al., 2012); (4) ferromagnetic and carbonate mineralogy (Jia et al., 2019; Li et al., 2013; Meng et al., 2019); and (5) electron spin resonance (ESR) signal intensity of quartz (Sun et al., 2013, 2008). Consequent sources have been primarily ascribed to the Qaidam Basin and northern
Tibetan Plateau (NTP), and the Tarim Basin to the west of the CLP (Liu et al., 1994; Pullen et al., 2011), the Alxa Plateau, Gobi Desert, and adjacent sandy deserts to the north (Li et al., 2009; Sun et al., 2008), and recycling of the Yellow River sediments (Nie et al., 2015). Instead of a single source, multiple ones have been claimed by many studies (Jia et al., 2019; Li et al., 2018; Licht, Pullen, et al., 2016). However, no matter which source tracer is deployed, the fundamental principle is to identify unique similarity between the potential source regions (PSRs) and the CLP.

After years of study, a major consensus that the NTP and the Central Asian Orogenic Belt (CAOB) are two primary sources has been reached (Sun et al., 2020). Nevertheless, major secondary sources such as the Gobi Desert, Qaidam Basin, and the river sediments cannot be clearly discriminated. Meanwhile, the processes that control aridification in Central Asia and concurrent sedimentation of eolian dust are at a continental scale (Guo et al., 2002; Licht, Dupont-Nivet, et al., 2016). To our knowledge, however, there is no spatial assessment of compositions of dust both on the CLP and from the PSRs and their relation at this scale. If any, our empirical knowledge might be inadequate to reckon with resultant big geodata. In this sense, our understanding of the provenance and quantitative estimation of the source contributions (e.g., Jia et al., 2019; Licht, Pullen, et al., 2016) is still incomplete. Consequently, large-scale spatial geochemical assessment coupled with big-data approaches is immediately required.

Machine learning allowing computers to handle new situations via analysis, self-training, observation, and experience has been successfully applied in the geosciences (Bergen et al., 2019). Zhao et al. (2019) use random forest and deep neural network to predict the origin of Cenozoic basalts in Northeast China. Mousavi and Beroza (2020) design a regressor composed of convolutional and recurrent neural networks to estimate the earthquake magnitude. Using an artificial neural network, Withers et al. (2020) successfully build a ground motion model. In this study, we apply two supervised machine learning algorithms, support vector machine (SVM) and convolutional neural network (CNN), to determine the dust sources simply by classifying the last glacial loess (LGL) and present interglacial sediments (PIS) on the CLP. To build the classifying models, a continental-scale geochemical mapping data set covering the PSRs (Figure 1, see section 2) is utilized as training data. The features learned by the models indicate that the PSRs are geochemically separable, which means even though fed by similar primary sources, the secondary sources have their own unique geochemical signatures. The trained models are thus capable of predicting which PSRs the LGL and PIS on the CLP resemble the most. Principal component analysis (PCA) is used to aid in understanding the probable dominant geochemical features of the PSRs extracted by SVM and CNN. Based on the classification, we identify the real secondary sources and estimate the contributions of both primary and secondary sources to the dust deposition. Our study also demonstrates the feasibility of applying CNN to big geodata processing as well as the bright future of using element geochemistry and big-data approaches (e.g., machine learning technology) to solve problems in land surface and solid-earth geoscience.

**Figure 1.** Map showing the shaded relief of central East Asia and the sampled sites. The continental-scale geochemical mapping covers the Chinese Loess Plateau (CLP) and eight potential source regions (PSRs). Glacial dust storm track and modern surface winds are according to Pullen et al. (2011). Circled H means high pressure over Central Asia. DT R. and HS R. are Datong River and Huangshui River, respectively.
2. Data and Methods

A large number \((n = 737)\) of surface catchment, fluvial, and/or alluvial sediments at a depth of \(-25\) cm are collected from eight PSRs (Figure 1), including Junggar Basin \((n = 222)\), Tarim Basin \((n = 170)\), eastern Tibetan Plateau (ETP, \(n = 39)\), Qaidam Basin \((n = 63)\), South (S.) Alxa Plateau \((n = 50)\), North (N.) Alxa Plateau \((n = 89)\), Hetao Graben \((n = 51)\), and northeastern sandy deserts \((n = 53)\). Besides, both the LGL \((n = 65)\) corresponding to the L1 layer of Quaternary loess and PIS \((n = 65)\) equivalent to the S0 layer of Quaternary paleosol are sampled (Figure 1). The whole sampling is deployed at the continental scale, that is, to document the abundance of chemical elements in materials occurring at the Earth’s surface often with a density and coverage larger than 1 sample/1,000 km² and 0.5 million km², respectively (Darnley et al., 2005). Concentrations of 67 elements, 11 oxides, \(H_2O^{+}\), total carbon, and organic carbon are determined (Supporting Information Text S1) and used in the machine learning modeling. The data are first screened to account for missing values (unreported, often appear as blank or null) and censored values (greater or less than the detection limits).

Considering that geochemical data are compositional (closed) data, centered log-ratio (clr) transformation (Aitchison, 1986) is used to resolve the closure problem. Zero-mean normalization is applied next (Text S2). Supervised machine learning algorithms can “learn” to recognize a pattern and build models from known examples (training set), test and optimize the trained models via labeled validation set, and ultimately make predictions for previously unseen data (Bergen et al., 2019). SVM (Cortes & Vapnik, 1995) and CNN (LeCun et al., 1998) are used in this study to build the classifying models.

The SVM method is applicable to both binary and multi-class classifications by defining the optimal boundary (ies) between (among) classes with either linear or such nonlinear kernels as radial basis function (Chang & Lin, 2011; Cracknell & Reading, 2013; Shahnas et al., 2018). The biggest advantage of SVM is its effectiveness in high-dimensional spaces. In this work, hyper-parameters of the SVM including the penalty to avoid overfitting the model and kernel function are finely tuned by using grid search based on a tenfold cross-validation to find the best estimators (Text S3). Balanced class weight is used to address the imbalance among the sample numbers.

The CNN algorithm is a popular class of deep neural networks (LeCun, Bengio, & Hinton, 2015; LeCun, Bottou et al., 2015) and has been widely applied in computer vision and natural language processing (Krizhevsky et al., 2012; LeCun et al., 1998). The CNN network structure is composed of convolutional layers for feature extraction, pooling layers for feature compression, and fully connected layers for classification (Bergen et al., 2019). The neural network is trained using backward propagation of errors with stochastic gradient descent (LeCun et al., 1998). In this research, the cross-entropy (Shore & Johnson, 1980) and adaptive moment estimation (Adam) optimizer (Kingma & Ba, 2014) are implemented to optimize the weights and minimize the mean loss. A large number of parameters such as the number of convolutional layers (or the network structure), learning rate, and dropout probability are varied to find the best estimators and prevent overfitting the model (Text S4). Figure 2 shows the CNN network structure used in this contribution.

The performance of the two models is estimated by four metrics, that is, accuracy, precision, recall, and F1 score. Accuracy is the ratio of correct predictions to all predictions. Precision refers to the ratio of true positives to true and false positives in one class. Recall denotes the ratio of true positives to false negatives.
reflecting the classifier’s completeness. F1 score, with the formula of \((2 \times \text{precision} \times \text{recall})/(\text{precision} + \text{recall})\), balances precision and recall and is usually regarded as the most effective metric.

PCA, which highlights inter-group differences and within-group similarities in a data set, is a powerful method for process discovery and pattern identification (Zhao et al., 2019). It discovers new linear combinations of the variables (elements) based on measures of association (correlation). A compositional biplot is usually used to describe PC loadings for each element (Aitchison & Greenacre, 2002) which displays their contribution to the PCs. The larger the absolute value, the greater the contribution. PCA is used in the present study only for extracting geochemical features and illustrating variations of element associations of the PSRs to aid in understanding the machine learning modeling.

3. Results

3.1. Machine Learning Modeling

The input data are chemical compositions of 67 elements, 11 oxides, \(\text{H}_2\text{O}^+\), total carbon, and organic carbon of the LGL and PIS on the CLP. The target output is one of the eight PSRs. Both machine learning algorithms present good performance. For SVM model, the kernel type of radial basis function (rbf), kernel coefficient \((\gamma)\) of 0.01, and penalty \((C)\) of 26 are used (Text S3), and the resultant four metrics, that is, accuracy, precision, recall, and F1 score, are 0.97, 0.98, 0.97, and 0.97, respectively (Table S1). The CNN model is composed of three convolutional layers with 32, 64, and 32 neurons; that is, a neural network of “Input (81)–32–64–32–Output (8)” is built (Figure 2 and Text S4), and the four measuring metrics are all 0.94 (Table S2). The two well-performed models suggest that the differentiation of major secondary PSRs is successful, effective, and reliable. This means that even though fed by similar primary sources (Sun et al., 2020), these secondary sources have their own unique geochemical signatures identified by the two used algorithms, although what features extracted by the two algorithms may have played a role is unclear because of the black-box effect of the used methods. Nonetheless, we speculate that the two models may have recognized valid geochemical patterns related with regional differences of the PSRs as is shown by the PCA biplots. In Figure 3a, the LGL and PIS are well characterized not only in the direction of the first principal component (PC1) with large positive loadings on REEs and high-strength field elements (HFSEs) but also in the direction of the second principal component (PC2) with large negative loadings on large ion lithophile elements (LILEs) as well as total carbon. A small part of LGL and PIS is featured by negative PC1 and positive PC2 showing mafic signatures (e.g., Cr, Co, Ni, MgO, and Pt). Compared with Figures 3b and 3c, it is clear that the Qaidam Basin, Tarim Basin, ETP, and Hetao Graben (termed as NTP-related secondary sources) share similar distribution on the PC biplot with the LGL and PIS than the Junggar Basin, N. Alxa, S. Alxa, and northeastern sandy deserts do. However, by examining Figure 3d, only ETP and a small part of N. Alxa and S. Alxa fall into the CLP domain with negative PC3 (e.g., MgO and CaO) as well as negative PC6 loadings (e.g., organic carbon and N). Meanwhile, even though fed by similar sources (i.e., the CAOB, Zhang et al., 2016), the Junggar Basin and N. Alxa can be distinguished by felsic signature demonstrated by positive PC1 loadings (Figure 3c). Comparably, SVM and CNN can discover valid geochemical patterns of the PSRs based on the used 81 elements.

3.2. Classification of the Dust on the CLP

Figure 4 shows the classification results by the two models. Two distinct similarities can be observed: (1) The two algorithms deliver generally consistent classification of both LGL and PIS into the PSRs, with clear dominance of the Hetao Graben, Qaidam Basin, and ETP (Figures 4a–4d), and (2) the classification patterns for LGL and PIS resemble each other (Figures 4a–4d). With the accuracy of each model being the weights, weighted average proportions of each PSR in LGL and PIS are obtained (Figures 4e and 4f). For both LGL and PIS, almost half is grouped into the Hetao Graben, around 15% and 15–25% into the Qaidam Basin and ETP, respectively. The remainder is roughly equally (about 5–6%) categorized into the Junggar Basin, Tarim Basin, and S. Alxa Plateau. The N. Alxa Plateau accounts for a small proportion less than 1%. There are no LGL and PIS grouped into the northeastern sandy deserts. More importantly, from LGL to PIS, slight variation (<10%) in the same PSR is detected (Figures 4e and 4f). In this regard, the classification of LGL and PIS exhibits an overall consistency.
We also verify the spatial agreement of the classification of LGL or PIS at the same sites by the two models. It turns out that 65% of all the dust samples have the identical predicted outputs in both models, illustrating a relatively high spatial consistency. Although the remaining 35% are not identical in the two models, confusion mostly occurs between the Hetao Graben and the other PSRs like the Qaidam Basin and ETP. The 35% discordant classification by the two models shows no clear spatial patterns and might be due to three reasons: (1) the common source region for some PSRs, for example, northeastern Tibetan Plateau for the Qaidam Basin, ETP, and Hetao Graben (Li et al., 2009; Zhang et al., 2016); (2) dust homogenization or mixing on the CLP (Kapp et al., 2015; Li et al., 2018; Licht, Pullen, et al., 2016); and (3) slightly different importance of elements during the classification by the two models. However, this could also indicate that the Hetao Graben is probably a storage of materials from other PSRs (Nie et al., 2015) considering its geographic location.

3.3. Dust Provenance

Our results show that the majority (approximately 80%) of LGL and PIS bear resemblance to the surface sediments from the Hetao Graben, ETP, and Qaidam Basin. Besides our results, zircon U-Pb age spectra also
emphasize the significance of ETP and Qaidam Basin (Pullen et al., 2011) and the recycled fluvial deposits from the Hetao Graben in the eolian sedimentary budget on the CLP as well (Licht, Pullen, et al., 2016; Nie et al., 2015). Therefore, based on the machine learning modeling, we propose that reworked fluvial sediments from the Hetao Graben contribute the most (approximately 50%) to the dust sediment on the CLP during the last glacial-interglacial cycle, followed by ETP (14–24%), Qaidam Basin (14–16%), Tarim Basin, Junggar Basin, and S. Alxa (5–7%), and N. Alxa (<1%).

Nevertheless, the Hetao Graben, whose materials are derived largely from the northeastern Tibetan Plateau (Lin et al., 2001; Nie et al., 2015) and adjacent highlands, still serves as a "transitional storage." This is revealed by the confusion between the Hetao Graben and the rest of the PSRs during the classification. Consequently, we apply the above-described two machine learning models to identify the provenance of the Hetao Graben. The two models perform well, and the four measuring metrics are all 0.93 for SVM and 0.92 for CNN (Tables S3 and S4). The corresponding weighted average proportions are then obtained (Figures 4g and 4h). Clearly, the S. Alxa, N. Alxa, and ETP contribute greatly to the Hetao Graben materials, which agrees with former studies. Given that the S. Alxa is composed mainly of the Qilian piedmonts on the NTP (Zhang et al., 2016), we combine the S. Alxa, ETP, Qaidam Basin, and Tarim Basin as the NTP-sourced PSR which explains more than 65% of the contribution to the Hetao Graben (Figures 4g and 4h). Some tributaries of the Yellow River like the Datong River and Huangshui River draining the Qilian Mountains (Figure 1) could account for the similarity between the S. Alxa and Hetao Graben. Contributions from the Junggar and Tarim Basins are likely to be resulted from the westerly and northwesterly winds and dust storms over the Hexi Corridor (Roe, 2009). It has been suggested that element ratios/compositions are not reliable geochemical tracers (Chen & Li, 2011; Sun et al., 2020) because changes in element compositions by either fluvial or eolian processes would complicate source attribution. However, the decomposition of the Hetao Graben demonstrates that big geochemical data sets coupled with machine learning technology are fully capable of tracing sources.

In accordance with the results of the two-stage machine learning modeling, we propose that the eolian dust sedimentary budget on the CLP is dominated by the NTP-sourced materials (approximately 75–80%) and minor by the CAOB-sourced materials (approximately 20–25%), at least during the last glacial-present...
Acknowledgments
This study was jointly supported by the Key Research & Development Project of the Ministry of Science and Technology of the People’s Republic of China (MOST) (2016YFC0600601) and National Natural Science Foundation of China (NSFC) (41671511 and 41702213). This study was completed while the first author was a visiting scholar at Laurentian University, which was financially supported by the China Scholarship Council, Grant 201906565036. The authors would like to thank the Editor-in-Chief, Prof. Harihar Rajaram and two anonymous reviewers for their handling and reviewing of our manuscript.

4. Conclusions
We investigate the sources of the dust on the CLP by applying two supervised machine learning algorithms. SVM and CNN models are trained by using 81 element compositions of 737 surface sediments from eight PSRs. The trained models are used to determine the dust sources by classifying the LGL and PIS into the PSRs. Classifications of the two models are found to be consistent. We show that the sedimentary budget on the CLP is dominated (approximately 50%) by reworked fluvial sediments from the Hetao Graben during the last glacial interglacial cycle, followed by ETP (14–24%), Qaidam Basin (14–16%), Tarim Basin, Junggar Basin, and S. Alxa (5–7%), and N. Alxa (1%). Through decomposing the Hetao Graben, two primary sources, that is, the NTP-sourced and CAOB-sourced materials, contribute 75–80% and 20–25% to the dust sedimentation, respectively. Moreover, a constant dust source despite changing atmospheric circulations is identified. Our results agree with those derived from Sr-Nd isotopes, quartz ESR-CI-$\delta^{18}O$ data, and zircon age spectra. We quantitatively estimate the contributions of both primary and secondary sources based on big geochemical data sets and accurate machine learning classification.

Data Availability Statement
All data used in this study have been archived and published at 4TU.Centre for Research Data (http://doi.org/10.4121/uuid:6cb0bf79-7467-4e78-a531-cc91655d9fd0).

References
Aitchison, J. (1986). The statistical analysis of compositional data. London: Chapman & Hall. https://doi.org/10.1007/978-94-009-4109-0
Aitchison, J., & Greenacre, M. (2002). Biplots of compositional data. Journal of the Royal Statistical Society: Series C: Applied Statistics, 51(4), 375–392. https://doi.org/10.1111/1467-9876.00275
An, Z. (2000). The history and variability of the East Asian paleomonsoon climate. *Quaternary Science Reviews*, 19(1–5), 171–187. https://doi.org/10.1016/S0277-3794(99)00060-8

An, Z., Kutzbach, J. E., Prell, W. L., & Porter, S. C. (2001). Evolution of Asian monsoons and phased uplift of the Himalayan-Tibetan plateau since Late Miocene times. *Nature*, 411(6833), 62–66. https://doi.org/10.1038/35075055

Bergen, K. J., Johnson, P. A., de Hoop, M. V., & Beroza, G. C. (2019). Machine learning for data-driven discovery in solid Earth geoscience. *Science*, 363, eaau0323. https://doi.org/10.1126/science.aau0323

Bird, A., Millar, I., Rodenburg, T., Stevens, T., Rittner, M., Vermeesch, P., & Lu, H. (2020). A constant Chinese Loess Plateau dust source since the late Miocene. *Quaternary Science Reviews*, 227, 106642. https://doi.org/10.1016/j.quascirev.2020.106642

Chang, C., & Lin, C. (2011). LIBVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2, 27. https://doi.org/10.1145/1961189.1961199

Chen, J., & Li, G. (2011). Geochemical studies on the source region of Asian dust. *Science China Earth Sciences*, 54(9), 1279–1301. https://doi.org/10.1007/s11430-011-4269-z

Chen, Z., & Li, G. (2013). Evolving sources of eolian detritus on the Chinese Loess Plateau since early Miocene: Tectonic and climatic controls. *Earth and Planetary Science Letters*, 371, 220–225. https://doi.org/10.1016/j.epsl.2013.03.044

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. https://doi.org/10.1023/A:1022627414111

Cracknell, M. J., & Reading, A. M. (2013). The upside of uncertainty: Identification of lithology contact zones from airborne geophysics and satellite data using random forest340s and support vector machines. *Geophysics*, 78, WB113. https://doi.org/10.1190/geo2012-0411.1

Darnley, A. G., Bjorklund, A., Bolviken, B., Gustavsson, N., Koval, P. V., Plant, J. A., et al. (2005). A global geochemical database for environmental and resource management: recommendations for International Geochemical Mapping, final report of IGCP Project 259.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324. https://doi.org/10.1109/5.726791

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324. https://doi.org/10.1109/5.726791

Li, G., Chen, J., & Chen, Y. (2013). Primary and secondary carbonate in Chinese loess discriminated by trace element composition. *Geochemistry and Geochimica Acta*, 103, 26–35. https://doi.org/10.1016/j.jgca.2012.10.049

Li, G., Chen, J., Ji, J., Yang, J., & Conway, T. M. (2009). Natural and anthropogenic sources of East Asian dust. *Geology*, 37(8), 727–730. https://doi.org/10.1130/G30031A.1

Li, L., Chen, J., Chen, Y., Hedding, D. W., Li, T., Li, L., et al. (2018). Uranium isotopic constraints on the provenance of dust on the Chinese Loess Plateau. *Geology*, 46(9), 747–750. https://doi.org/10.1130/G45130.1

Licht, A., Dupont-Nivet, G., Pullen, A., Kapp, P., Abels, H. A., Lai, Z., et al. (2016). Resilience of the Asian atmospheric circulation shown by paleolimnological evidence from lake sediment in northeastern Chinese Loess Plateau. *Quaternary Research*, 86(11), 2278–2324. https://doi.org/10.1109/5.726791

Licht, A., Pullen, A., Kapp, P., Abels, H. A., Lai, Z., et al. (2016). Resilience of the Asian atmospheric circulation shown by paleolimnological evidence from lake sediment in northeastern Chinese Loess Plateau. *Quaternary Research*, 86(11), 2278–2324. https://doi.org/10.1109/5.726791

Lin, A., Yang, Z., Sun, Z., & Yang, T. (2001). How and when did the Yellow River develop its square bend. *Quaternary Research*, 55(2), 260–268. https://doi.org/10.1016/S0033-5894(01)00034-7

Mousavi, S. M., & Beroza, G. C. (2020). A machine-learning approach for earthquake magnitude estimation. *Geophysical Research Letters*, 47, e2020GL085976. https://doi.org/10.1029/2020GL085976

Nie, J., & Peng, W. (2014). Automated SEM-EDS heavy mineral analysis reveals no provenance shift between glacial loess and interglacial paleo Loess Plateau. *Aeolian Research*, 13, 71–75. https://doi.org/10.1016/j.aeolre.2014.03.005

Nie, J., Pullen, A., Garzione, C. N., Peng, W., & Wang, Z. (2018). PreQuaternary decoupling between Asian aridification and high dust accumulation rate. *Science Advances*, 4, eaau9677. https://doi.org/10.1126/sciadv.aau9677

Nie, J., Stevens, T., Rittner, M., Stockli, D., Garzanti, E., Limonta, M., et al. (2015). Loess Plateau storage of Northeastern Tibetan Plateau-derived Yellow River sediment. *Nature Communications*, 6, 8511. https://doi.org/10.1038/ncomms8511

Pullen, A., Kapp, P., McCallister, A. T., Chang, H., Gehrels, E. G., Garzione, C. N., et al. (2011). Qaidam Basin and northern Tibetan Plateau as dust sources for the Chinese Loess Plateau and paleoclimatic implications. *Geology*, 39(11), 1031–1034. https://doi.org/10.1130/g32261.1

Roe, G. (2009). On the interpretation of Chinese loess as a paleoclimate indicator. *Quaternary Research*, 71(2), 150–161. https://doi.org/10.1016/j.yqres.2008.09.004

Shahna, M. H., Yuen, D. A., & Pysklywec, R. N. (2018). Inverse problems in geodynamics using machine learning algorithms. *Journal of Geophysical Research: Solid Earth*, 123, 296–310. https://doi.org/10.1002/2017JB014846
Shore, J. E., & Johnson, R. W. (1980). Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross entropy. *IEEE Transactions on Information Theory*, 26(1), 26–37. https://doi.org/10.1109/tit.1980.11056144

Stevens, T., Buylaert, J. P., Lu, H., Thiel, C., Murray, A., Frechen, M., et al. (2016). Mass accumulation rate and monsoon records from Xifeng, Chinese Loess Plateau, based on a luminescence age model. *Journal of Quaternary Science*, 31(4), 391–405. https://doi.org/10.1002/jqs.2848

Sun, J., & Zhu, X. (2010). Temporal variations in Pb isotopes and trace element concentrations within Chinese eolian deposits during the past 8 Ma: Implications for provenance change. *Earth and Planetary Science Letters*, 290(3–4), 438–447. https://doi.org/10.1016/j.epsl.2010.01.001

Sun, Y., Chen, H., & Tada, R. (2013). ESR signal intensity and crystallinity of quartz from Gobi and sandy deserts in East Asia and implication for tracing Asian dust provenance. *Geochemistry, Geophysics, Geosystems*, 14, 2615–2627. https://doi.org/10.1002/ggge.20162

Sun, Y., Tada, R., Chen, J., Liu, Q., Toyoda, S., Tani, A., et al. (2008). Tracing the provenance of fine-grained dust deposited on the central Chinese Loess Plateau. *Geophysical Research Letters*, 35, L01804. https://doi.org/10.1029/2007GL031672

Sun, Y., Yan, Y., Nie, J., Li, G., Shi, Z., Qiang, X., et al. (2020). Source-to-sink fluctuations of Asian aeolian deposits since the late Oligocene. *Earth-Science Reviews*, 206, 102963. https://doi.org/10.1016/j.earscirev.2019.102963

Vandenberghhe, J., Renissen, H., Huissleden, K. V., Nugteren, G., Konert, M., Lu, H.-Y., et al. (2006). Penetration of Atlantic westerly winds into Central and East Asia. *Quaternary Sciences Reviews*, 25(17–18), 2380–2389. https://doi.org/10.1016/j.quascirev.2006.02.017

Withers, K. B., Moschetti, M. P., & Thompson, E. M. (2020). A machine learning approach to developing ground motion models from simulated ground motions. *Geophysical Research Letters*, 47, e2019GL086690. https://doi.org/10.1029/2019GL086690

Xiao, G., Zong, K., Li, G., Hu, Z., Dupon-Nivet, G., Peng, S., & Zhang, K. (2012). Spatial and glacial-interglacial variations in provenance of the Chinese Loess Plateau. *Geophysical Research Letters*, 39, L20715. https://doi.org/10.1029/2012GL053304

Yan, Y., Sun, Y., Chen, H., & Ma, L. (2014). Oxygen isotope signatures of quartz from major Asian dust sources: Implications for changes in the provenance of Chinese loess. *Geochimica et Cosmochimica Acta*, 139, 399–410. https://doi.org/10.1016/j.gca.2014.04.043

Yang, S., & Ding, Z. (2017). Spatial changes in grain size of loess deposits in the Chinese Loess Plateau and implications for palaeoenvironment. *Quaternary Sciences*, 37, 934–944. https://doi.org/10.11928/jissn.1001-7410.2017.05.02

Zhang, H., Lu, H., Xu, X., Liu, X., Yang, T., Stevens, T., et al. (2016). Quantitative estimation of the contribution of dust sources to Chinese loess using detrital zircon U-Pb age patterns. *Journal of Geophysical Research: Earth Surface*, 121, 2085–2099. https://doi.org/10.1002/2016JF003936

Zhao, Y., Zhang, Y., Geng, M., Jiang, J., & Zou, X. (2019). Involvement of slab-derived fluid in the generation of Cenozoic basalts in Northeast China inferred from machine learning. *Geophysical Research Letters*, 46, 5234–5242. https://doi.org/10.1029/2019GL082322