

Research Article

Examining the Potential Environmental Controls of Underground CO₂ Concentration in Arid Regions by an SVD-PCA-ANN Preview Model

Zhikai Zhuang,¹ Xiaoqiang Li,² Wenfeng Wang^(b),^{3,4} and Xi Chen^(b),^{4,5,6,7}

¹International Energy Institute, Jinan University, Jinan 310006, China

²Department of Scientific Research, The Enlighten Academic Institute, Nanjing, Jiangsu Province 210000, China

³School of Science, Shanghai Institute of Technology, Shanghai 201418, China

⁴State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences,

Urumqi 830011, China

⁵University of Chinese Academy of Sciences, Beijing 100049, China

⁶Sino-Belgian Joint Laboratory of Geo-Information, Urumqi 830011, China

⁷CAS Research Centre for Ecology and Environment of Central Asia, Urumqi 830011, China

Correspondence should be addressed to Wenfeng Wang; wangwenfeng@nimte.ac.cn and Xi Chen; chenxi@ms.xjb.ac.cn

Received 18 July 2021; Revised 17 September 2021; Accepted 30 September 2021; Published 20 October 2021

Academic Editor: Yong-Hong Lan

Copyright © 2021 Zhikai Zhuang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study attempts to examine environmental controls of the underground CO_2 concentration, taking the CO_2 concentration 4 m beneath the soil as an example. An SVD-PCA-ANN (singular value decomposition-principal component analysis-artificial neural network) preview model is proposed with the data of underground CO_2 concentration and 12 environmental variables (the soil and meteorological data). The R^2 , RMSE, and RPD values of the proposed model are, respectively, 0.8874, 0.3351, and 2.7929, performing better than the popular preview models like SAE (stacked autoencoders), SVM (support vector machine), and LSTM (long short-term memory). It is proved that the underground CO_2 concentration can be approximated by a nonlinear function of the considered variables. Soil temperature, salinity, and wind speed are the leading environmental controls, which explain 32.04%, 13.68%, and 11.21% in the variability of the underground CO_2 concentration, respectively. Possible mechanisms associated with the environmental controls are also preliminarily discussed.

1. Introduction

In the previous studies of land-atmosphere CO_2 exchange, the measurements of CO_2 concentration mainly focused on the dynamics of soil surface CO_2 concentration, which is characterized as soil CO_2 fluxes or soil respiration [1–5]. However, the soil surface CO_2 concentration not only is determined by the atmospheric CO_2 concentration but also is potentially affected by the underground CO_2 concentration [6–8].

Especially, some recent studies of desert ecosystems revealed an unneglectable CO_2 absorption by the salinealkali soils [9–11]. This implied that the underground CO_2 concentration in arid regions is closely linked with the soil surface CO_2 concentration. However, the controls of soil absorption of CO_2 in arid regions were not well-explained until now [12–15]. One significant reason is that the CO_2 concentration beneath the soil might be influenced by many factors and up to now, and the corresponding theoretical basis has not been well-described [16]. There is a lack of research on the effects of various environmental factors on the underground CO_2 concentration in arid regions [17].

In order to further understand the soil absorption of CO_2 in arid regions, we will do research on the effects of various environmental factors on the underground CO_2 concentration. Objectives of this study are (1) to examine environmental controls of the underground CO_2 concentration, taking the CO_2 concentration 4 m beneath the soil as an example, (2) to propose a preview model for analyzing the concentration dynamics, and (3) to discuss how environmental factors are influencing the underground CO_2 concentration. The preview model will be proposed in Section 2, and in Section 3, we will examine environmental controls of the underground CO_2 concentration utilizing the proposed model and discuss the possible mechanisms.

2. Materials and Methods

2.1. Data Collection. All data used in this paper, including the underground CO_2 concentration (FX), soil temperature (T_s) , soil moisture (S_m) , atmospheric temperature (A_t) , atmospheric humidity (A_h) , soil salinity (EC), wind speed (WS), air pressure (AP), wind direction (WD), groundwater level (WL), soil alkalinity (pH), atmospheric CO_2 concentration (C_a) , and rainfall (*R*), were collected from three automatic weather stations (equipped with 12 sensors for the soil and meteorological data), where FX is the dependent variable and the other 12 environmental factors are independent variables. These three weather stations are located at the south edge of the Gurbantunggut Desert and the north of Xinjiang Uygur Autonomous Region, China, as shown in Figure 1.

2.2. The Proposed Model. Differing from the previous studies on the soil absorption of CO₂ in arid regions, we integrate two adaptive methods such as the principal component analysis (PCA) and the artificial neural network (ANN) to examine both the linear and nonlinear relationships between FX and the 12 possible environmental controls. A series of PCA algorithms were proposed in the previous studies, some of which have been widely used in the dimensionality reduction of various data [18-23]. The singular value decomposition (SVD) can improve PCA [24, 25]. Hence, we will utilize the SVD-based PCA to examine the contributions of the 12 environmental variables to FX, where SVD is used to calculate the eigenvalue and eigenvector of the covariance matrix [25]. For the statement convenience, we symbolize the proposed model as SVD-PCA-ANN throughout the paper.

We will reconstruct k-dimensional features based on the original n-dimensional features and then map the n-dimensional features to k-dimensional features (known as the main components) [19, 21, 25]. The output of the SVD-based PCA will be input into ANN—a multilayer forward back propagation network with 2 input layers, 4 hidden layers, and 1 output layer. ReLU transfer functions (nonlinear) are selected for the hidden layer and linear transfer functions for the output layer to approximate the nonlinear relationship between the ANN input and output. The structure diagram of the proposed model is shown in Figure 2.

2.3. Examination of the Controls. The following 3 indices are calculated to quantify the robustness of the SVD-PCA-ANN model in previewing the possible environmental controls of the underground CO_2 concentration, which are also utilized in the comparison with stacked autoencoders (SAE), support vector machine (SVM), and LSTM (long short-term memory) [26–30]. For a reliable comparison, the data set was uniformly divided into 3 subsets in all the experiments, one for training (half of the input data), one for validation (one quarter of the input data), and one for testing (one quarter of the input data).

RMSE =
$$\sqrt{\frac{\sum_{i} (y_{i} - y'_{i})^{2}}{N}}$$
, (1)
 $R^{2} = 1 - \frac{\sum_{i} (y_{i} - y'_{i})^{2}}{\sum_{i} (y_{i} - \overline{y})^{2}}$.

The ratio of prediction to deviation is as follows:

$$\operatorname{RPD} = \frac{\left(\sum y_i^2 - \left[\sum y_i^2/n\right]/(N-1)\right)^{1/2}}{\left(\sum_i (y_i - y_i')^2 - \left\{\left[\sum (y_i - y_i')/N\right]^2/N - 1\right\}\right)^{1/2}},$$
(2)

where y_i is the true value, y'_i is the previewed value, \overline{y} is the average of the true value, and N is the number of environmental variables.

The pseudocode for the SVD-PCA-ANN model is shown in Figure 3.

3. Experimental Results

3.1. Outputs of the SVD-Based PCA. The calculated contribution ratios of the 12 environmental variables and the explained ratios of the 9 principal components from the SVD-based PCA are shown in Figure 4. The contribution ratios of soil moisture (S_m) , soil salinity (EC), wind speed (WS), soil temperature (T_s) , groundwater level (WL), wind direction (WD), atmospheric temperature (T_a) , atmospheric humidity (H_a) , air pressure (AP), atmospheric CO₂ concentration (C_a), soil alkalinity (pH), and rainfall (R) to the underground CO₂ concentration are 32.04%, 13.68%, 11.21%, 8.38%, 8.31%, 7.87%, 7.21%, 5.22%, 3.46%, 1.23%, 0.82%, and 0.51%, respectively. The explained ratios of the 9 principal components from the SVD-based PCA are, respectively, 28.5%, 15.2%, 11.4%, 9.08%, 7.87%, 7.63%, 6.23%, 5.77%, and 5.05%, respectively.

This suggests that soil moisture, soil salinity, and wind speed are 3 leading drivers of the underground CO_2 concentration. The total contribution ratio of these 3 environmental controls to FX is 56.93%. The overall contribution ratio of soil temperature, groundwater level, wind direction, atmospheric temperature, atmospheric humidity, air pressure, atmospheric CO_2 concentration, soil alkalinity, and rainfall is 43.07%, among which the contribution ratios of soil alkalinity and rainfall are both less than 1%. According to the outputs of the SVD-based PCA, the 9 principal components can explain 96.73% of changes in the underground CO_2 concentration.



FIGURE 1: Distribution of the three automatic weather stations where the soil and meteorological data are collected in this study.

3.2. Performance of the SVD-PCA-ANN Model. The original PCA model is linear and simple, allowing a feasible calculation of the contribution ratio of each environmental variable. However, such a pure linear model is not robust enough as a preview model (RMSE = 1.2468; R^2 = 0.5028). Since the 9 principal components can explain 96.73% of changes in the underground CO₂ concentration, these 9 principal components from the SVD-based PCA can be utilized as the inputs of ANN. This step integrated the advantages of the linear model (PCA) and nonlinear model (ANN). The calculated R^2 , RMSE, and RPD of the integrated model (the SVD-PCA-ANN preview model) with the increase in model training epochs are shown in Figures 5–7, respectively.

According to the calculated R^2 , RMSE, and RPD of the SVD-PCA-ANN model, the proposed model displays a good performance with the training, validation, and test data sets. This suggests that the proposed model is robust for predicting the underground CO₂ concentration in arid regions.

3.3. Comparison with SAE, SVM, and LSTM. In comparison of the PCA, SAE, and SVD-PCA-ANN, as seen in Table 1, we find that the proposed model performs better than both PCA and SAE on the training, validation, and testing data sets. The SVD-PCA-ANN model can explain more than 88.7% of the variability in the training data set with an accuracy of RMSE = 0.3351, while the SVD-based PCA can only explain 48.9% of the variability in the training data set with an accuracy of RMSE = 1.2330. SAE can explain not more than 1% of the variability in the training data set, but RMSE from the SAE model is a half of RMSE from the PCA model. In explanation of the variability in the validation and testing data sets, the performance of the SVD-PCA-ANN model also indicates a good prediction. SAE explains about 17% and 37% of the variability in the validation and testing data sets, respectively, but the generated RMSE increased. PCA explains about 50% of the variability in the validation and testing data sets with no evident changes in the generated RMSE.



Matrix B is arranged according to the covariance matric of A by the first 9 terms in order of eigenvalues

FIGURE 2: The structure diagram of the proposed model.



FIGURE 3: Pseudocode for examining process based on the proposed model.



FIGURE 4: Outputs of the SVD-based PCA.



FIGURE 5: \mathbf{R}^2 of the SVD-PCA-ANN model at various phases for the training, validation, and test data sets.



FIGURE 6: RMSE of the SVD-PCA-ANN model at various phases for the training, validation, and test data sets.

Comparing the performance of the SVM model and the SVD-PCA-ANN model, we also find that the proposed model performs better. The SVD-PCA-ANN model explains about 80% of the variability in the training, validation, and testing data sets with a good accuracy (RMSE <0.43). The SVM model can only explain 12.1%, 48%, and 61% of the variability in the training, validation, and testing data sets, respectively, and the generated RMSE >1. The SVD-PCA-ANN model also performs better than LSTM. As seen in Table 1, LSTM explains about 33.8%, 32.7%, and 27.2% of the variability in the training, validation, and testing data sets, respectively, and the generated RMSE values are obviously bigger than the

generated RMSE from SVD-PCA-ANN. The calculated RPD further demonstrates a better prediction of PCA-ANN than SAE, SVM, and LSTM.

4. Discussion

Differing from the variability in soil surface CO_2 concentration and the atmospheric CO_2 concentration, the dynamics of CO_2 concentration beneath the ground might be influenced by many environmental factors and unknown subterranean processes [9–17]. Until now, the whole story of soil CO_2 absorption in arid regions is still a gap in our



FIGURE 7: RPD of the SVD-PCA-ANN model at various phases for the training, validation, and test data sets.

TABLE 1: Comparison of the SVD-PCA-ANN model with SAE, SVM, and LSTM.

Accuracy evaluation index		R^2	RMSE	RPD
Training	SVD-PCA-ANN	0.8874	0.3351	2.7929
	SVM	0.1209	1.0552	0.7001
	LSTM	0.3375	0.8087	0.5367
	SAE	0.0074	0.6074	0.1999
	SVD-based PCA	0.4890	1.2330	2.9422
Validation	SVD-PCA-ANN	0.8197	0.4121	1.9069
	SVM	0.4800	1.2130	0.2723
	LSTM	0.3271	0.8253	0.5366
	SAE	0.1698	0.9435	0.1289
	SVD-based PCA	0.5028	1.2468	2.9516
Testing	SVD-PCA-ANN	0.7661	0.4257	1.9043
	SVM	0.6070	1.0491	0.6071
	LSTM	0.2722	0.8096	0.5337
	SAE	0.3713	0.9435	0.2439
	SVD-based PCA	0.4931	1.2328	2.4960

knowledge [31]. The calculated contribution ratios of the considered 12 environmental variables in the presented study suggest that soil moisture, soil alkalinity, and wind direction are 3 leading controls among the considered 12 environmental variables. Possible mechanisms are as follows. Soil moisture can integrate with CO_2 under the condition that soil salinity is high. The integrating processes can be influenced by wind speed [32]. Additionally, our experimental results also imply that the wind direction is less significant than wind speed. Among the other 9 variables, the contribution ratio of groundwater level is only less than soil temperature. This presents new evidence for the necessity to take into account groundwater discharge and recharge as a factor in analyzing the underground CO_2 concentration in arid regions [33].

In the present study, we proposed the SVD-PCA-ANN model for predicting the CO_2 concentration 4 m beneath the

soil. This is a first SVD-PCA-based neural network for learning the underground CO₂ concentration in arid regions. Based on all the experimental results in the present study, the major advantage of the proposed method is that SVD-PCA-ANN integrated the linear model PCA and the nonlinear model ANN. However, according to the outputs of the SVDbased PCA, the 9 principal components can only explain 96.73% of changes in the underground CO₂ concentration. This represents one limitation of the proposed method. The SVD-PCA-ANN model introduces 9 linear components in predicting the underground CO₂ concentration in arid regions to improve the interpretability of traditional neural networks [34-36]. This is also the major reason why the proposed method has a better prediction than SAE, SVM, and LSTM. However, there are still 11.3%, 18.1%, and 23.4% in the variability of the training, validation, and testing data sets beyond the explanation of the proposed model. This represents another limitation of the SVD-PCA-ANN model.

A next research priority in the subsequent studies is to break through these two limitations of the SVD-PCA-ANN model for a fully understanding of changes in the underground CO_2 concentration in arid regions. To break through such limitations, the first principles to be considered in the subsequent studies include constructing better linear components for the preview model, finding more efficient learning systems, and introducing new environmental factors to reduce the uncertainty in the model analyses.

5. Conclusion

The considered environmental variables, including the soil and meteorological factors, can be recognized as potential controls of the underground CO_2 concentration in arid regions, among which soil moisture, soil salinity, and wind speed are leading controls. Experimental results demonstrated that the proposed method performs better than SAE, SVM, and the well-known LSTM on the training, validation, and test data sets. The relationship between the underground CO_2 concentration in arid regions and the considered environmental variables cannot be characterized as a single linear function. The SVD-PCA-ANN model can effectively predict the underground CO_2 concentration in arid regions by integrating the advantages of both linear and nonlinear models and therefore presents a novel method for studying soil CO_2 absorption in arid regions.

Data Availability

The data utilized to support the theory and models of the present study are available from the corresponding authors upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (41571299), the Shanghai High-Level Base-Building Project for Industrial Technology Innovation (1021GN204005-A06), and the Ningbo Natural Science Foundation (2019A610106).

References

- L.-d. Shen, Y.-l. Yang, J.-q. Liu et al., "Different responses of ammonia-oxidizing archaea and bacteria in paddy soils to elevated CO2 concentration," *Environmental Pollution*, vol. 286, no. 1, p. 117558, 2021.
- [2] M. Dima, D. R. Nichita, G. Lohmann, M. Ionita, and M. Voiculescu, "Early-onset of atlantic meridional overturning circulation weakening in response to atmospheric CO2 concentration," *Npj Climate and Atmospheric Science*, vol. 4, no. 1, p. 27, 2021.
- [3] I. N. Kurganova, V. O. Lopes de Gerenyu, D. A. Khoroshaev et al., "Analysis of the long-term soil respiration dynamics in the forest and meadow cenoses of the prioksko-terrasny biosphere reserve in the perspective of current climate trends," *Eurasian Soil Science*, vol. 53, no. 10, pp. 1421–1436, 2020.
- [4] I. F. Dmitry and J. Tatarinov, "Soil respiration in paludified forests of European Russia," *Journal of Forestry Research*, vol. 31, no. 5, pp. 464–473, 2020.
- [5] B. Klimek, M. Chodak, and M. Niklińska, "Soil respiration in seven types of temperate forests exhibits similar temperature sensitivity," *Journal of Soils and Sediments*, vol. 21, no. 1, pp. 338–345, 2021.
- [6] T. A. Kursar, "Evaluation of soil respiration and soil CO2 concentration in a lowland moist forest in Panama," *Plant and Soil*, vol. 113, no. 1, pp. 21–29, 1989.
- [7] A. H. Frank, R. van Geldern, A. Myrttinen, M. Zimmer, J. A. C. Barth, and B. Strauch, "Geological CO2 quantified by high-temporal resolution stable isotope monitoring in a salt mine," *Scientific Reports*, vol. 10, no. 1, p. 20671, 2020.
- [8] E. Ayres, D. Wall, B. Simmons et al., "Belowground nematode herbivores are resistant to elevated atmospheric CO2

concentrations in grassland ecosystems," Soil Biology and Biochemistry, vol. 40, no. 4, pp. 978–985, 2008.

- [9] M. H. Yim, S. J. Joo, and K. Nakane, "Comparison of field methods for measuring soil respiration: a static alkali absorption method and two dynamic closed chamber methods," *Forest Ecology and Management*, vol. 170, no. 1, pp. 189–197, 2002.
- [10] Z. Huang, Z. Yu, and M. Wang, "Environmental controls and the influence of tree species on temporal variation in soil respiration in subtropical China[J]," *Plant and Soil*, vol. 382, no. 1-2, pp. 75–87, 2014.
- [11] Y. Li, M. Xu, X. Zou, and Y. Xia, "Soil CO2 efflux and fungal and bacterial biomass in a plantation and a secondary forest in wet tropics in Puerto Rico," *Plant and Soil*, vol. 268, no. 1, pp. 151–160, 2005.
- [12] K.-Y. Fa, J.-B. Liu, Y.-Q. Zhang et al., "CO2absorption of sandy soil induced by rainfall pulses in a desert ecosystem," *Hydrological Processes*, vol. 29, no. 8, pp. 2043–2051, 2015.
- [13] W. Wang, X. Chen, L. Wang, H. Zhang, G. Yin, and Y. Zhang, "Approaching the truth of the missing carbon sink," *Polish Journal of Environmental Studies*, vol. 25, no. 4, pp. 1799–1802, 2016.
- [14] Y. Gao, P. Zhang, and J. Liu, "One third of the abioticallyabsorbed atmospheric CO2 by the loess soil is conserved in the solid phase," *Geoderma*, vol. 374, no. 6, p. 114448, 2020.
- [15] W. Wang, X. Chen, H. Zheng et al., "Soil CO2 uptake in deserts and its implications to the groundwater environment," *Water*, vol. 8, no. 9, p. 379, 2016.
- [16] W. M. P. d. M. Ivo and I. H. Salcedo, "Soil CO2 flux: a method comparison of closed static chambers in a sugarcane field," *Revista Brasileira de Ciência do Solo*, vol. 36, no. 2, pp. 421– 426, 2012.
- [17] J. I. Freijer and W. Bouten, "A comparison of field methods for measuring soil carbon dioxide evolution: experiments and simulation," *Plant and Soil*, vol. 135, no. 1, pp. 133–142, 1991.
- [18] A. Wiesel and A. O. Hero, "Decomposable principal component analysis," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4369–4377, 2009.
- [19] R. Moller and A. Konies, "Coupled principal component analysis," *IEEE Transactions on Neural Networks*, vol. 15, no. 1, pp. 214–222, 2004.
- [20] H. Li, G. C. Linderman, A. Szlam, K. P. Stanton, Y. Kluger, and M. Tygert, "Algorithm 971," ACM Transactions on Mathematical Software, vol. 43, no. 3, pp. 1–14, 2017.
- [21] Q. He, X. Shao, and W. Chen, "Adaptive multi-scale tracking target algorithm through drone," *IEICE-Transactions on Communications*, vol. E102.B, no. 10, 2019.
- [22] P. Praus, "SVD-based principal component analysis of geochemical data," *Open Chemistry*, vol. 3, no. 4, pp. 731–741, 2005.
- [23] P. Praus, "Water quality assessment using SVD-based principal component analysis of hydrological data," *Water S A*, vol. 31, no. 4, pp. 417–422, 2005.
- [24] B. Geluvaraj, "AMatrix factorization technique using parameter tuning of singular value decomposition for Recommender Systems," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 2, pp. 3313–3319, 2021.
- [25] J. H. Caltenco, J. Lopez-Bonilla, and B. E. Carvajal-Gámez, "Singular value decomposition," *The Bulletin of Society for Mathematical Services and Standards*, vol. 11, pp. 13–20, 2014.
- [26] J. Long, Z. Sun, and C. Li, "A novel sparse echo autoencoder network for data-driven fault diagnosis of delta 3-D printers,"

IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 3, pp. 683-692, 2020.

- [27] B. Wu, W. He, and J. Wang, "A convolutional-LSTM model for nitrogen oxide emission forecasting in FCC unit," *Journal* of *Intelligent and Fuzzy Systems*, vol. 40, no. 1, pp. 1537–1545, 2021.
- [28] O. Nyarko-Boateng, "Tracing the exact location of failures in underground optical networks using LSTM deep learning model [J]," *Indian Journal of Science and Technology*, vol. 14, no. 4, pp. 297–309, 2021.
- [29] Y. Mao, G. Qin, and P. Ni, "Analysis of road traffic speed in Kunming plateau mountains: a fusion PSO-LSTM algorithm," *International Journal on the Unity of the Sciences*, no. 7, pp. 1–21, 2021.
- [30] M. Sundermeyer, H. Ney, and R. Schluter, "From feedforward to recurrent LSTM neural networks for language modeling," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 3, pp. 517–529, 2015.
- [31] A. Rey, "Mind the gap: non-biological processes contributing to soil CO₂ efflux," *Global Change Biology*, vol. 21, no. 5, pp. 1752–1761, 2015.
- [32] A. Rey, L. Belelli-Marchesini, and A. Were, "Wind as a main driver of the net ecosystem carbon balance of a semiarid Mediterranean steppe in the South East of Spain," *Global Change Biology*, vol. 18, no. 2, pp. 539–554, 2012.
- [33] P. An, W. F. Wang, and X. Chen, "Introducing a chaotic component in the control system of Soil respiration," *Complexity*, vol. 2020, no. 8, pp. 1–8, 2020.
- [34] P. Petrovic, "Placebo and opioid analgesia-- imaging a shared neural network," *Science*, vol. 295, no. 5560, pp. 1737–1740, 2002.
- [35] G. Buzsáki and J. J. Chrobak, "Temporal structure in spatially organized neuronal ensembles: a role for interneural networks," *Current Opinion in Neurobiology*, vol. 5, no. 4, pp. 504–510, 1995.
- [36] J. T. Coull, C. Büchel, and K. J. Friston, "Noradrenergically mediated plasticity in a human attentional neural network," *NeuroImage*, vol. 11, no. 6, p. 822, 2000.