Supporting Information

Intelligent Analysis Strategy for the Key Factor on Soil Nitrogen and Phosphorus Loss via Runoff under Simulated Karst Conditions

Yuqi Zhang, Rongchang Zeng, Tianyang Li, Lan Song, Binghui He*

College of Resources and Environment, Southwest University,

Chongqing 400715, China

* Corresponding author

E-mail: hebinghui@swu.edu.cn (B.H.)

Table of Contents

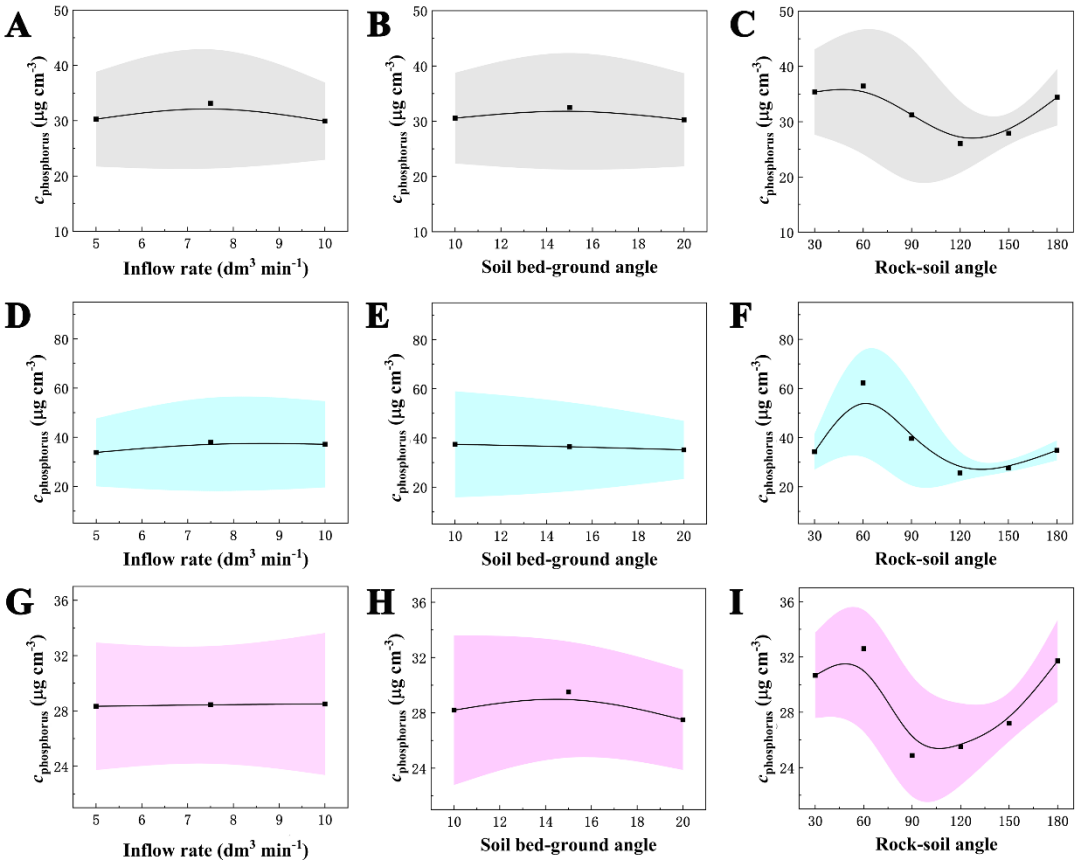


Figure S1. Characteristics relationships between the concentration of SP and inflow rate (v), soil bed-ground angle (α), and rock-soil angle (β) in soil surface runoff (A − C), subsurface runoff (D − F), and fissures runoff (G − I), respectively. The dot means the average nitrogen concentration and the graph means the standard deviation.

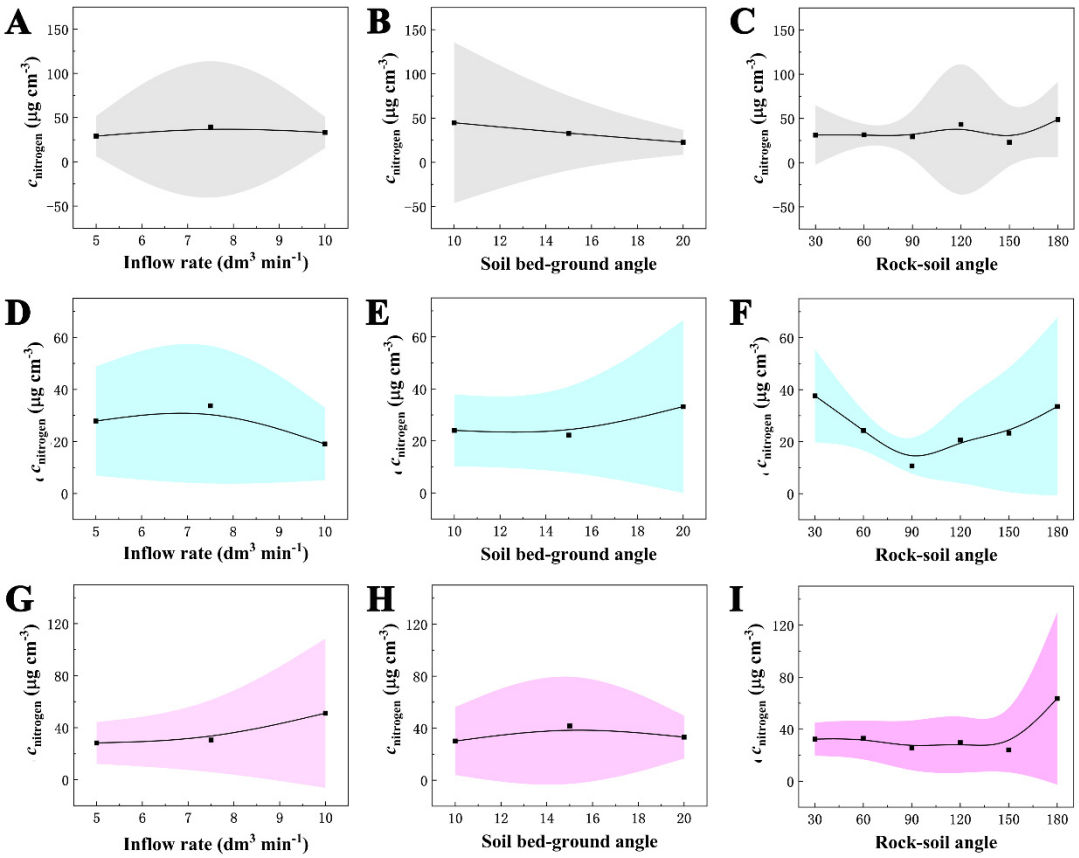


Figure S2. Characteristics relationships between the NH_4−N concentration and v, α, and β in soil surface runoff (A − C), subsurface runoff (D − F), and fissures runoff (G − I), respectively. The dot means the average nitrogen concentration and the graph means the standard deviation.

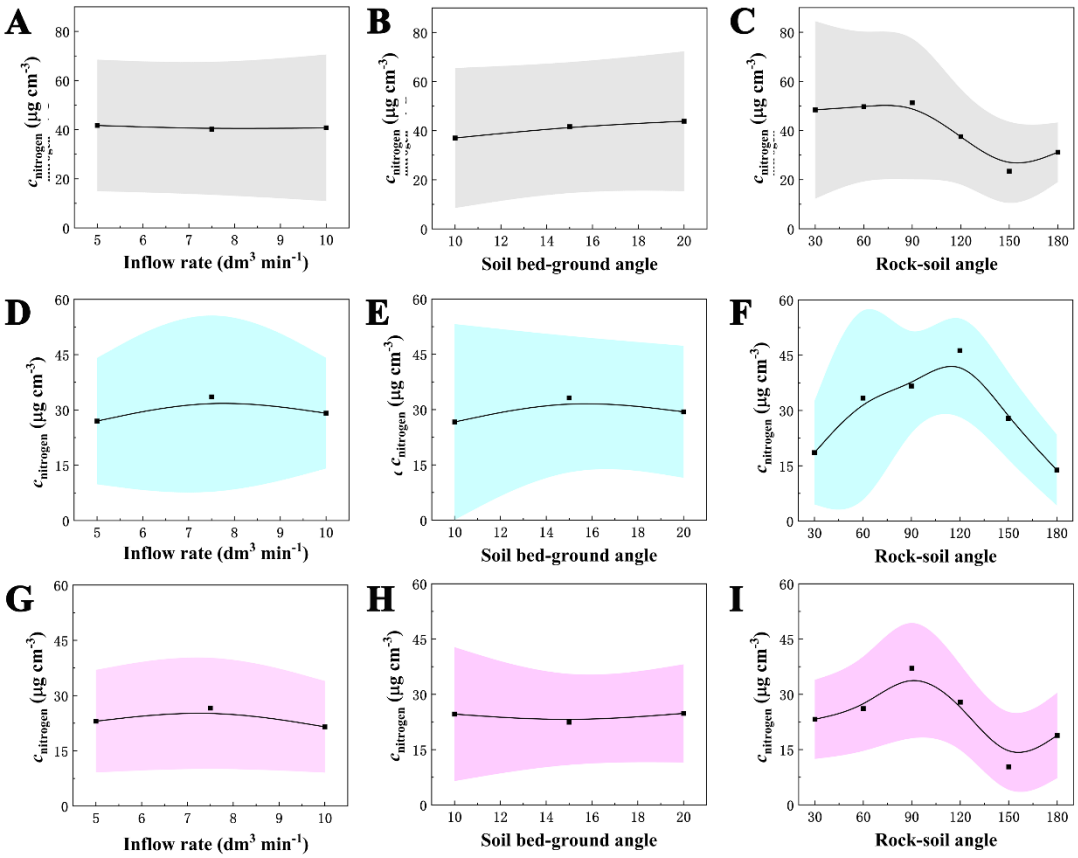


Figure S3. Characteristics relationships between the NO_3−N concentration and v, α, and β in soil surface runoff (A − C), subsurface runoff (D − F), and fissures runoff (G − I), respectively. The dot means the average nitrogen concentration and the graph means the standard deviation.

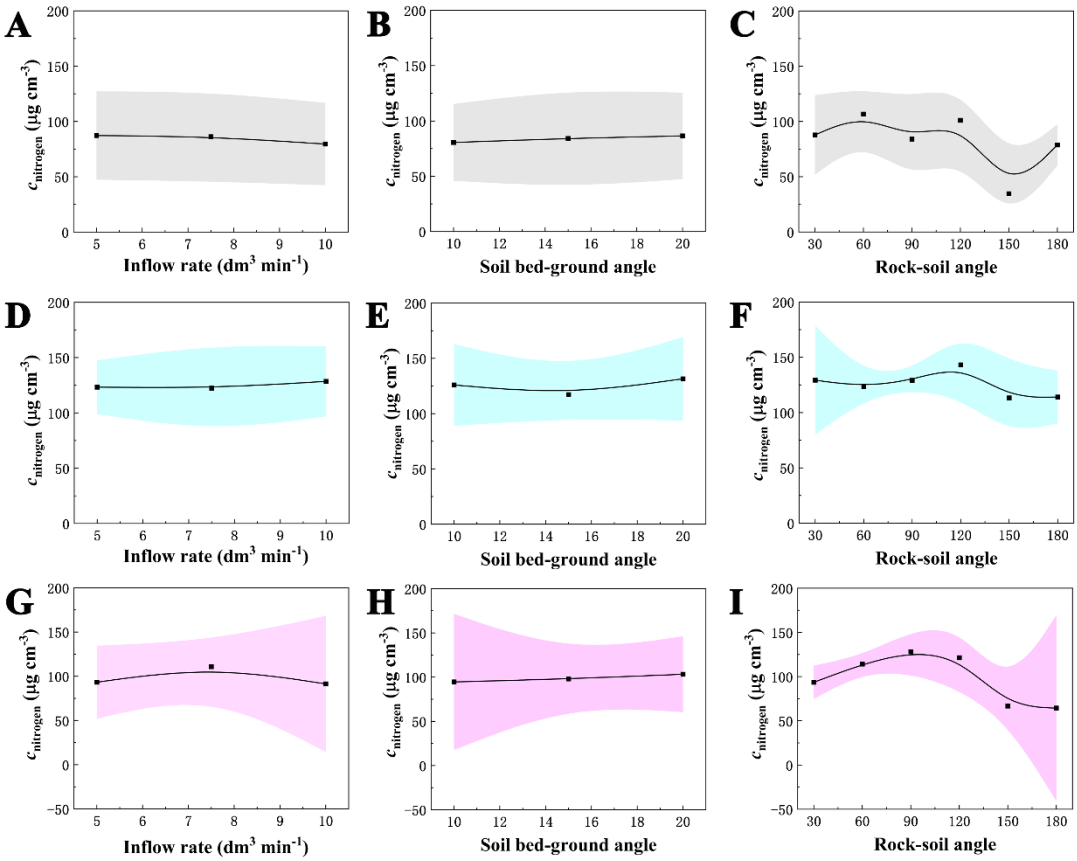


Figure S4. Characteristics relationships between the TN concentration and v, α, and β in soil surface runoff (A − C), subsurface runoff (D − F), and fissures runoff (G − I), respectively. The dot means the average nitrogen concentration and the graph means the standard deviation.

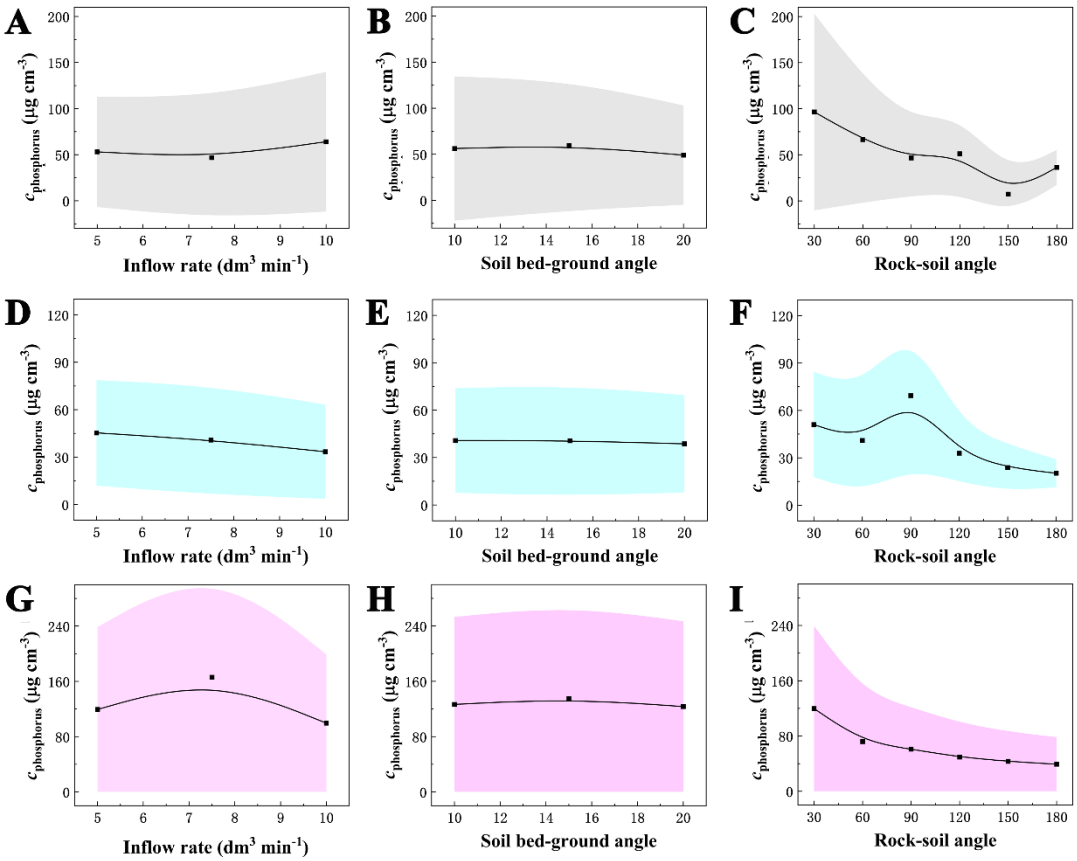


Figure S5. Characteristics relationships between the TP concentration and v, α, and β in soil surface runoff (A − C), subsurface runoff (D − F), and fissures runoff (G − I), respectively. The dot means the average nitrogen concentration and the graph means the standard deviation.

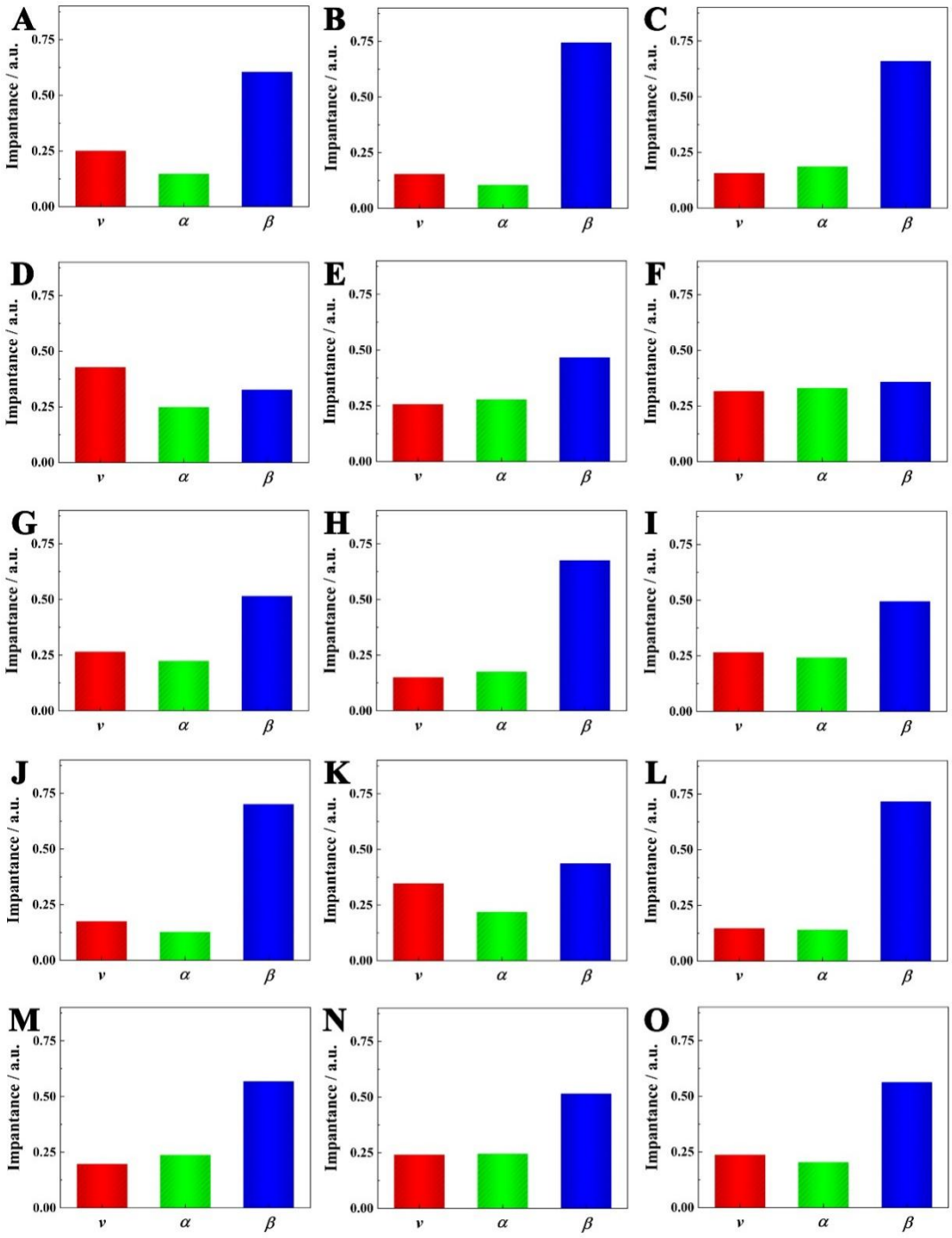


Figure S6. Characterizations of the Importance factor to estimate the effect of every feature, including the v, α, and β, on the TN (A − C), NH_4–N (D − F), NO_3–N (G − I), SP (J − L) and TP (M − O) concentration in the soil surface runoff (A, D, G, J and M), subsurface runoff (B, E, H, K and N), and fissures runoff (C, F, I, L and O), respectively.

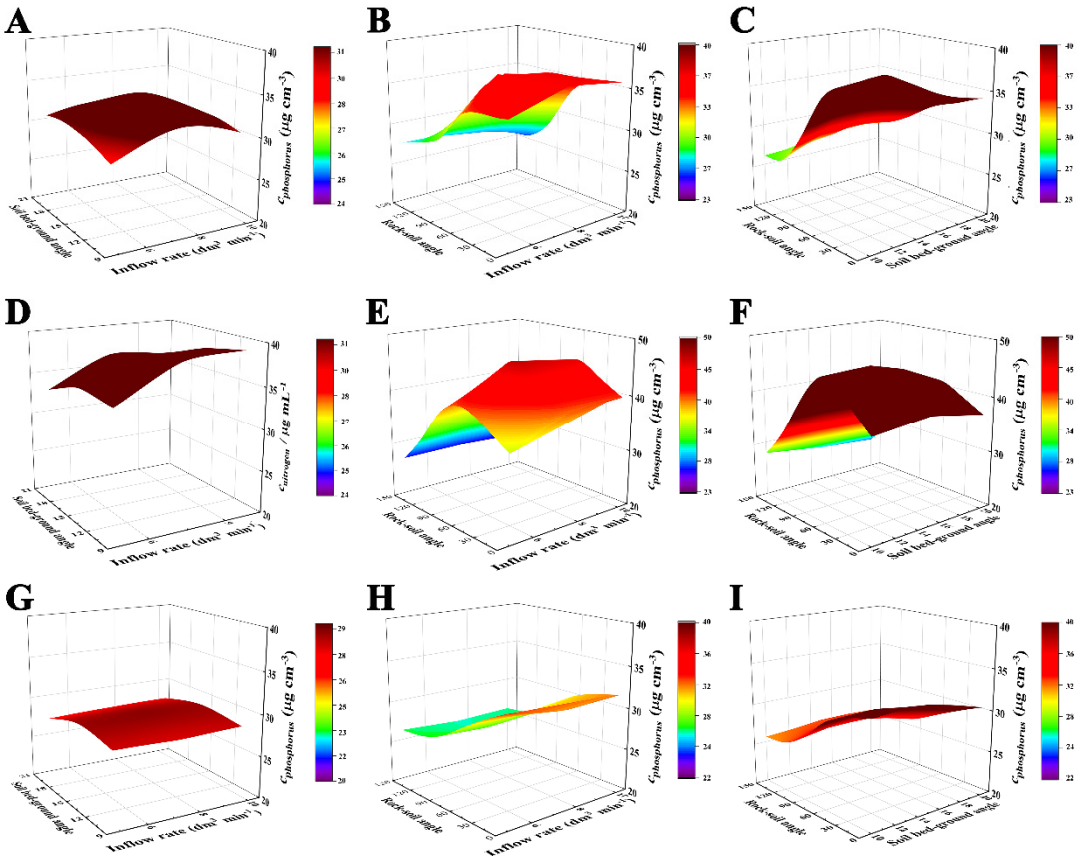


Figure S7. Characterizations of the relationship between the concentration of SP in the soil surface runoff (A − C), subsurface runoff (D − F), and fissures runoff (G − I) and the v, α, and β, respectively.

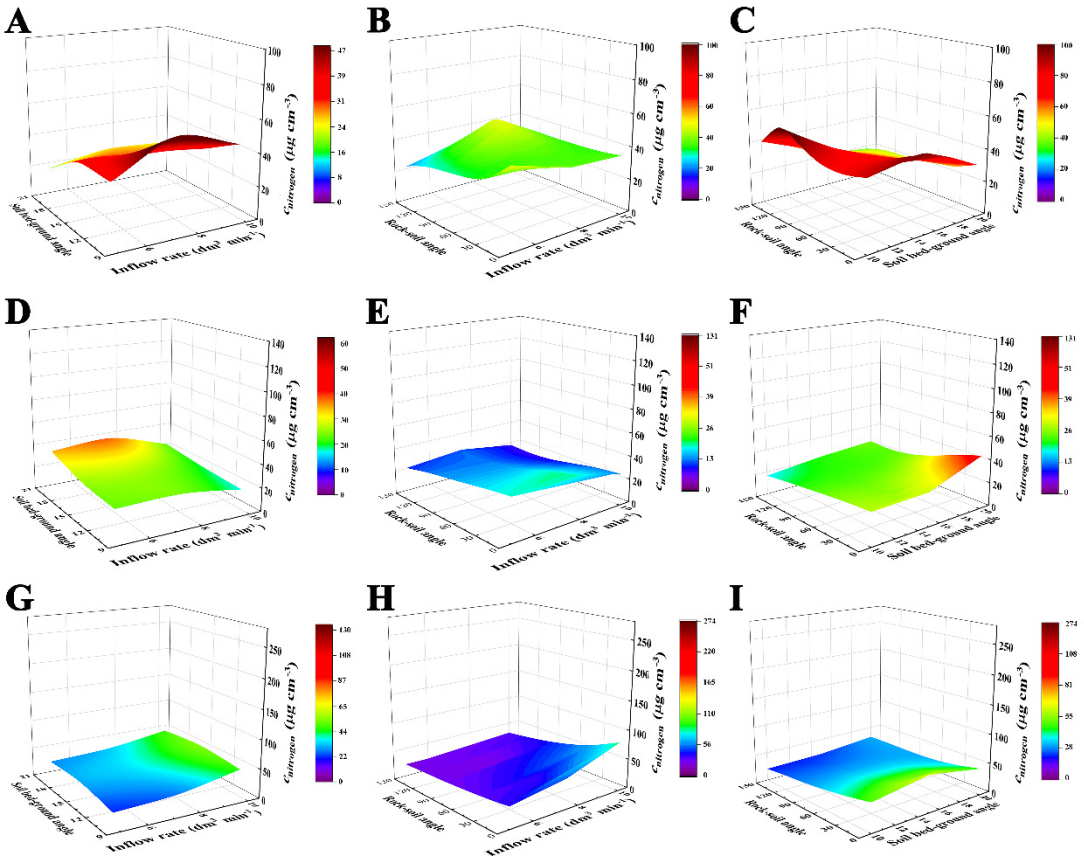


Figure S8. Characterizations of the relationship between the concentration of NH_4-N in the soil surface runoff (A − C), subsurface runoff (D − F), and fissures runoff (G − I) and the v, α, and β, respectively.

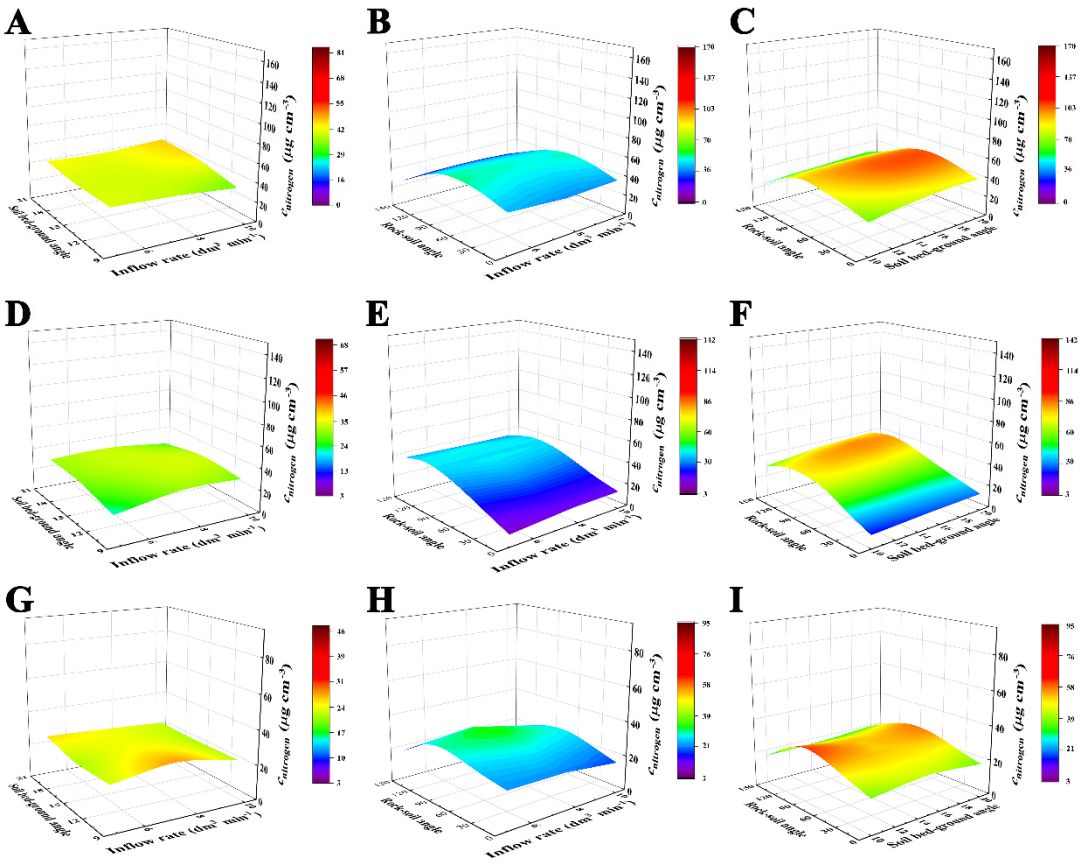


Figure S9. Characterizations of the relationship between the concentration of NO_3-N in the soil surface runoff (A − C), subsurface runoff (D − F), and fissures runoff (G − I) and the v, α, and β, respectively.

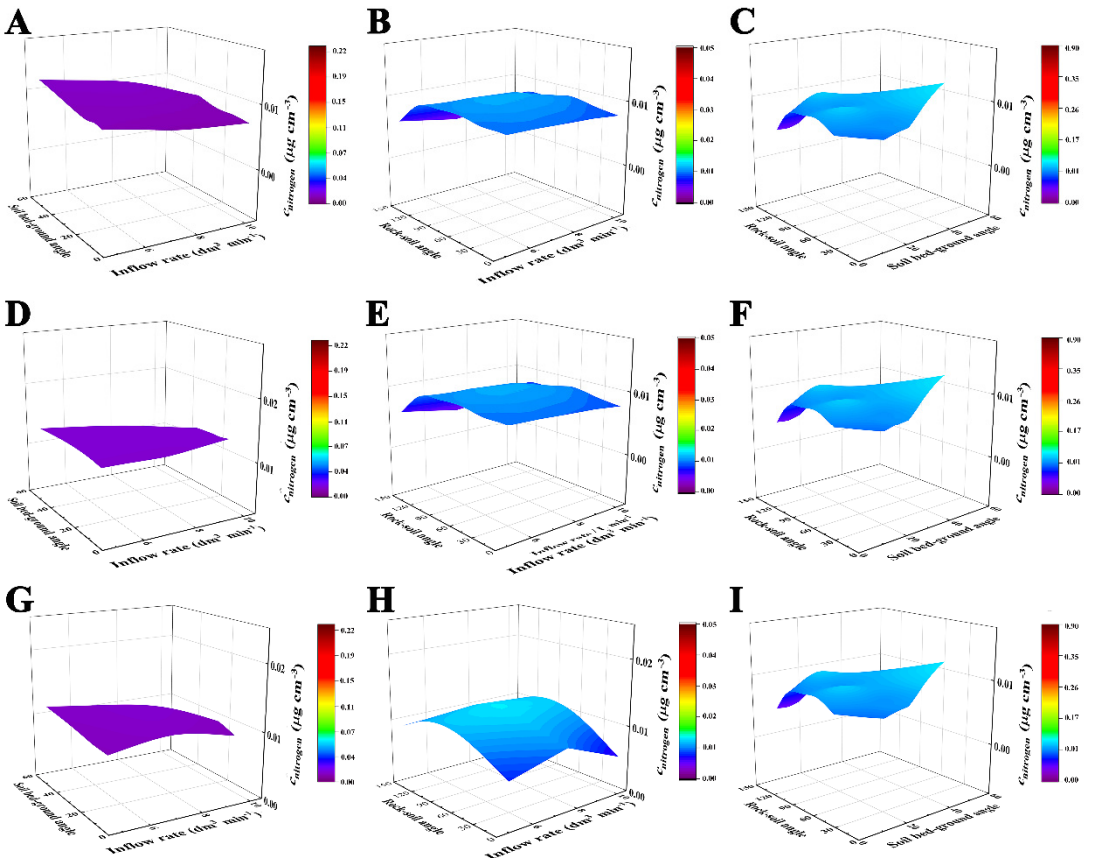


Figure S10. Characterizations of the relationship between the concentration of TN in the soil surface runoff (A − C), subsurface runoff (D − F), and fissures runoff (G − I) and the v, α, and β, respectively.

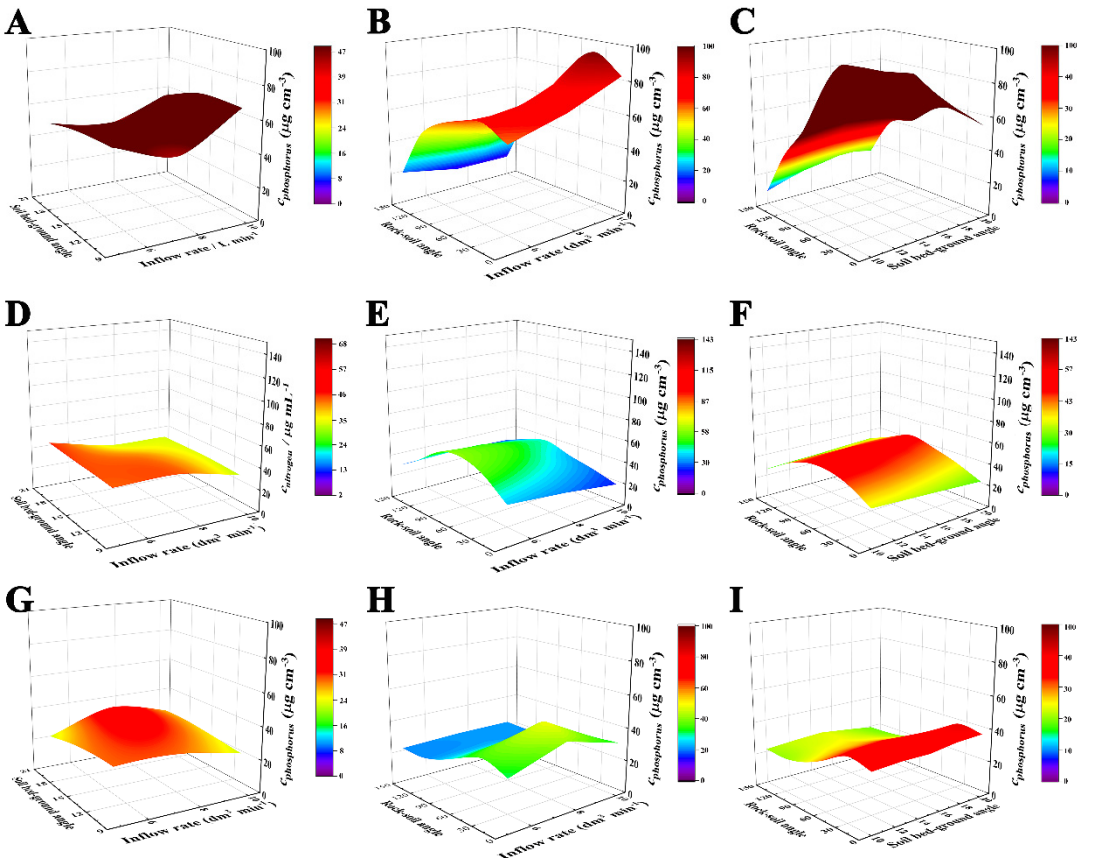
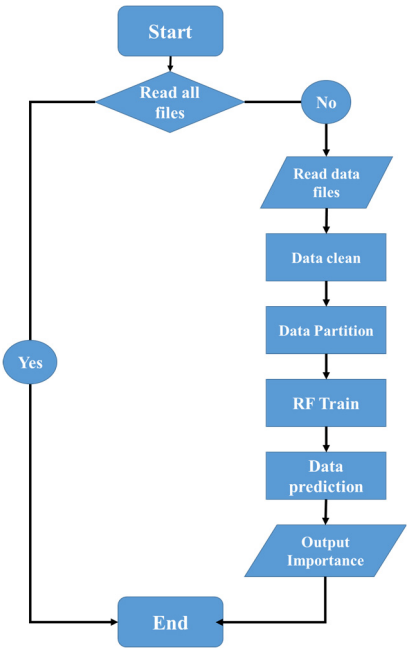


Figure S11. Characterizations of the relationship between the TP concentration in the soil surface runoff (A − C), subsurface runoff (D − F), and fissures runoff (G − I) and the v, α, and β, respectively.

1. Frame diagram of RF



2. RF code

```
import pandas as pd
from sklearn.metrics import r2_score, mean_squared_error, mean_absolute_error
from matplotlib import pyplot as plt
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split
import numpy as np
plt.rcParams['font.sans-serif']=['SimHei']
plt.rcParams['axes.unicode_minus']=False

from glob import glob
print(glob("*xlsx"))

def f(x):
    try:
        return float(x)
    except:
        return np.nan

for file_name in glob("*xlsx"):
    data = pd.read_excel(file_name,header=1)
    label_name = "Concentration"
    try:
        data = data[["Flow","Angle","RockAngle",label_name]]
    except:
        print("Wrong Name！ ")
```

```
    data["Flow"] = data["Flow"].apply(f )
    data = data.dropna()

    data_x = data[["Flow","Angle","RockAngle"]]
    data_y = data[label_name]
    X_train, X_test, y_train, y_test = train_test_split(data_x, data_y, test_size=0.25,
random_state=0)

    print("Name::",file_name)
    print("Target：",label_name)
    print("Sum of datasets：",data_x.shape[0])
    print("Training set：",X_train.shape[0])
    print("Test set：",X_test.shape[0])

    rf = RandomForestRegressor()
    rf.fit(X_train, y_train)
    predictions = rf.predict(X_test)
    print("Goodness of fit",r2_score(predictions,y_test))
    print("MSE",mean_squared_error(predictions,y_test))
    print("MAE",mean_absolute_error(predictions,y_test))
    print("Importance：")
    print(list(zip(X_train.columns,rf.feature_importances_)))
    plt.barh(X_train.columns, rf.feature_importances_)
    plt.show()
```