Analyzing moisture-heat coupling in a wheat-soil system using data-driven vector autoregression model

Xiaohang Feng [Corresp., 1, 2] , Xia Zhang [1] , Zhenqi Feng [1] , Yichang Wei [1]

[1] North China University of Water Resources Electric Power, Zhengzhou, Henan, China

[2] Peking University, Beijing, China

Corresponding Author: Xiaohang Feng
Email address: 1600018809@pku.edu.cn

Soil temperature and moisture have a close relationship, the accurate controlling of which is important for crop growth. Mechanistic models built by previous studies need exhaustive parameters and seldom consider time stochasticity and lagging effect. To circumvent these problems, this study designed a data-driven stochastic model analyzing soil moisture-heat coupling. Firstly, three vector autoregression models are built using hourly data on soil moisture and temperature at the depth of10 cm, 30 cm, and 90 cm. Secondly, from impulse response functions, the time lag and intensity of two variables' response to one unit of positive shock can be obtained, which describe the time length and strength at which temperature and moisture affect each other, indicating the degree of coupling. Thirdly, Granger causality tests unfold whether one variable's past value helps predict the other's future value. Analyzing data obtained from Shangqiu Experiment Station in Central China, we obtained three conclusions. Firstly, moisture's response time lag is 25 h, 50 h, and 120 h, while temperature's response time lag is 50 h, 120 h, and 120 h at 10 cm, 30 cm, and 90 cm. Secondly, temperature's response intensity is 0.2004°C, 0.0163°C and 0.0035°C for 1% variation in moisture, and moisture's response intensity is 0.0638%, 0.0163% and 0.0050% for 1°C variation in temperature at 10 cm, 30 cm and 90 cm. Thirdly, the past value of soil moisture helps predict soil temperature at 10 cm, 30 cm, and 90 cm. Besides, the past value of soil temperature helps predict soil moisture at 10 cm and 30 cm, but not at 90 cm. We verified this model by using data from a different year and linking it to Soil Plant Atmospheric Continuum model.

1   **Manuscript Title**

2   Xiaohang Feng[1,2*] Xia Zhang[1] Zhenqi Feng[1] Yichang Wei[1]

3   [1]North China University of Water Resources Electric Power, Zhengzhou, Henan, China; [2]Peking University, Beijing,

4   China

5   Corresponding Author:

6   Xiaohang Feng[1,2*]

7   Email address: 1600018809@pku.edu.cn; Tel.: +86-18811776315

8 **Analyzing Moisture-heat Coupling in a Wheat-soil System Using Data-**
9 **driven Vector Autoregression Model**

10 Xiaohang Feng[1,2]* Xia Zhang[1] Zhenqi Feng[1] Yichang Wei[1]
11 [1]North China University of Water Resources Electric Power, Zhengzhou, Henan, China; [2]Peking University, Beijing,
12 China
13 *Corresponding Author, Email: 1600018809@pku.edu.cn; Tel.: +86-18811776315

14 **Abstract:** Soil temperature and moisture have a close relationship, the accurate controlling of which is important for
15 crop growth. Mechanistic models built by previous studies need exhaustive parameters and seldom consider time
16 stochasticity and lagging effect. To circumvent these problems, this study designed a data-driven stochastic model
17 analyzing soil moisture-heat coupling. Firstly, three vector autoregression models are built using hourly data on soil
18 moisture and temperature at the depth of 10 cm, 30 cm, and 90 cm. Secondly, from impulse response functions, the
19 time lag and intensity of two variables' response to one unit of positive shock can be obtained, which describe the
20 time length and strength at which temperature and moisture affect each other, indicating the degree of coupling.
21 Thirdly, Granger causality tests unfold whether one variable's past value helps predict the other's future value.
22 Analyzing data obtained from Shangqiu Experiment Station in Central China, we obtained three conclusions. Firstly,
23 moisture's response time lag is 25 h, 50 h, and 120 h, while temperature's response time lag is 50 h, 120 h, and 120
24 h at 10 cm, 30 cm, and 90 cm. Secondly, temperature's response intensity is 0.2004°C, 0.0163°C and 0.0035°C for
25 1% variation in moisture, and moisture's response intensity is 0.0638%, 0.0163% and 0.0050% for 1°C variation in
26 temperature at 10 cm, 30 cm and 90 cm. Thirdly, the past value of soil moisture helps predict soil temperature at 10
27 cm, 30 cm, and 90 cm. Besides, the past value of soil temperature helps predict soil moisture at 10 cm and 30 cm,
28 but not at 90 cm. We verified this model by using data from a different year and linking it to Soil Plant Atmospheric
29 Continuum model.

30 **Introduction**

31   The hydrothermal conditions of a wheat-soil system are essential to the dynamic balance of heat, moisture and
32 organic matter within the entire system and the thriving of wheat (Yang et al. 2012; Sun et al. 2018). Because of
33 complicated biological, physical and chemical processes like soil respiration, soil evaporation, plant transpiration and
34 so on, soil temperature and moisture have a close dynamic relationship with each other. Therefore, in order to better
35 control and predict the hydrothermal conditions of a wheat-soil system, models simulating soil moisture-heat
36 coupling need to be established (Perez et al. 2013; Whelan et al. 2015).
37   Most of the models employed to simulate moisture-heat coupling of soil are mechanistic models which call for
38 complex parameters on soil properties and neglect time stochasticity. The following are some examples. Philip and
39 de Vries (1957) took soil evaporation under different temperature into consideration and brought up the theory of
40 moisture-gas-heat coupling transport under mass and energy balance. However, they did not consider the time lag
41 effect and temporal heterogeneity. Based on the Philip model, Nassar and Horton (1989) used the water, heat, and
42 salt transport equations which are based on Darcy's, Fourier's and Fick's laws to establish a model for the coupled
43 transport of water, heat, and solute. But the model required a bunch of complicated parameters. Based on the
44 theoretical integrity, Liu et al. (2005) established a model for describing the migration of heat, moisture, and gas in
45 arid surface porous soil composed of a wet unsaturated layer and a dry but saturated layer. Bittelli et al. (2008)
46 established a fully coupled numerical model to solve the governing equations for liquid water, water vapor and heat
47 transfer in bare soil. Whelan et al. (2015) studied the impact of temperature and moisture on soil water repellency
48 by designing and conducting experiments and factorial ANOVAs. However, the study fails to further explore
49 characteristics like time lag and intensity of temperature and moisture's influence on each other. Lu and Dong (2015)
50 set up a closed-formed equation for the thermal conductivity of unsaturated soils and shows that the soil water
51 retention curve (SWRC) can be used to predict the thermal conductivity of sands. Striegl and Loheide (2012)
52 developed a distributed soil moisture sensing system that addressed the difficulty of characterizing both spatial and
53 temporal soil moisture dynamics at site scales. However, under wet conditions insensitivity of the instrument
54 response curve adversely affected accuracy. Steele-Dunne et al. (2010) used distributed temperature sensing (DTS)
55 to obtain simultaneous measurements of soil moisture over large areas, but they fail to address the complexity of

56 deriving soil moisture due to the uncertainty and non-uniqueness in the relationship between thermal conductivity
57 and soil moisture.
58   Most of the models mentioned above are deterministic. Since in reality soil moisture and soil temperature
59 constantly vary with disturbances from an assortment of factors like weather and soil, stochastic models are more
60 accurate in terms of prediction than deterministic models (Bolin et al. (in press)). Moreover, mechanistic models
61 mentioned above require a lot of parameters which are difficult to obtain in some cases and some models can only
62 be applied to a certain situation which is quite limited (Pan et al. 2018). In comparison, time series models prove to
63 have a wider application since they can be easily established even if nothing but the hydrological time series data
64 are in hand. In addition, although the hysteresis effect has been included in previous models like Soil Plant
65 Atmospheric Continuum (SPAC), few indices have been established to quantify the lagging effect.
66   To take random variation into consideration, and to deal with the situation when only time series data are in
67 hand, we designed a stochastic modeling technique for accurately analyzing moisture-heat coupling within a wheat-
68 soil system. The model consists of three two-dimensional vector autoregression models, impulse response functions
69 and Granger causality tests.

70 **Methods**

71 *Models*

72 Vector Autoregression Model

73   Vector autoregression model treats all variables as endogenous variables, accounting for Sims' critique that the
74 exogeneity[1] assumptions of some of the variables in simultaneous equations models are ad hoc and often not backed
75 by fully developed theories. Impulse response analysis and Granger causality tests are tools which have been
76 proposed for disentangling the relations between the variables in a vector autoregression model (Droumaguet et al.
77 2017; Bouri et al. 2018). The process of establishing a vector autoregression system is as follows.
78   First, we remove the seasonality of the time series x_t, $t=1, 2, …, n$. Seasonality in a time series is a regular
79 pattern of changes that repeats over S time periods, where S defines the number of time periods until the pattern
80 repeats again. In this case, S = 24 (hours per day) is the span of the periodic seasonal behavior. Seasonal differencing
81 is defined as a difference between a value and a value with lag that is a multiple of S. With S = 24, a seasonal
82 difference is $y_t=(1 - B^{24})x_t = x_t - x_{t-24}$, $t = 25, 26, …, n$, where B is the lag operator.
83   Next, we check the stationarity of the time series after seasonal adjustment, because average values of non-
84 stationary time series cannot be used if the time period is not set since they are influenced by changes in time. A
85 stochastic process is called stationary when the average and variance values are constant in the corresponding
86 period and covariance values between any two time points depend not on the specific time point but on the lag
87 between these two time points. The three conditions stated above for a time series $\{y_t\}$ to be stationary can be
88 expressed in the following way:

$$E(y_t) = \mu, \ Var(y_t) = \sigma^2, \ Cov(y_t, y_{t-k}) = \gamma_k, \tag{1}$$

89 where μ, σ^2, and γ are average, variance, and auto-covariance, respectively.
90   If the time series has a unit root, then it is not stationary. Under such circumstances, we restore the stationarity
91 of the time series by making a difference of it. We use the Augmented Dickey-Fuller (ADF) test method to check the
92 stationarity of the time series by testing whether the time series has a unit root. There are three model variants
93 corresponding to three couples of hypothesis in the ADF test.
94   The first is an autoregressive model variant (referred to as *AR*), which specifies a test of the null model:

$$H_0: y_t = y_{t-1} + \beta_1\Delta y_{t-1} + \beta_2\Delta y_{t-2} + … + \beta_p\Delta y_{t-p} + \varepsilon_t, \tag{2}$$

95 against the alternative model:

---

[1] Exogenous variables are also called "input variables", thus "exogeneity" describes whether the variable is exogenous or not. An exogenous variable is completely determined by the external part of the system and are input into the system. It only affects the system and is not affected by the system.

$$H_1: y_t = \emptyset y_{t-1} + \beta_1 \Delta y_{t-1} + \beta_2 \Delta y_{t-2} + \ldots + \beta_p \Delta y_{t-p} + \varepsilon_t, \qquad (3)$$

96   with $\emptyset < 1$, where $\beta_1, \beta_2, \ldots, \beta_p$ are regression coefficients, and ε_t is the error term.

97   The second is an autoregressive model with drift variant (referred to as *ARD*), which specifies a test of the null
98   model:

$$H_0: y_t = y_{t-1} + \beta_1 \Delta y_{t-1} + \beta_2 \Delta y_{t-2} + \ldots + \beta_p \Delta y_{t-p} + \varepsilon_t, \qquad (4)$$

99   against the alternative model:

$$y_t = c + \emptyset y_{t-1} + \beta_1 \Delta y_{t-1} + \beta_2 \Delta y_{t-2} + \ldots + \beta_p \Delta y_{t-p} + \varepsilon_t, \qquad (5)$$

100   with $\emptyset < 1$, where c is the constant term, $\beta_1, \beta_2, \ldots, \beta_p$ are regression coefficients, and ε_t is the error term.
101   The third is a trend-stationary model variant (referred to as *TS*), which specifies a test of the null model:

$$H_0: y_t = c + y_{t-1} + \beta_1 \Delta y_{t-1} + \beta_2 \Delta y_{t-2} + \ldots + \beta_p \Delta y_{t-p} + \varepsilon_t, \qquad (6)$$

102   against the alternative model:

$$H_1: y_t = c + \delta t + \emptyset y_{t-1} + \beta_1 \Delta y_{t-1} + \beta_2 \Delta y_{t-2} + \ldots + \beta_p \Delta y_{t-p} + \varepsilon_t, \qquad (7)$$

103   with $\emptyset < 1$, where c is the constant term, δ is the coefficient of the trend term, $\beta_1, \beta_2, \ldots, \beta_p$ are regression
104   coefficients, and ε_t is the error term (Chen et al. 2018).
105   Then, we have to determine the optimal lag order for the vector autoregression model. The proper selection
106   of lag is important because long lag structures reduce the error term's correlation yet they may lack efficiency. We
107   use the estimators AIC (Akaike information criterion) and BIC (Bayesian information criterion), both of which are
108   founded on information theory and estimate relative goodness of fit of given models, as standards of choosing the
109   optimal lag of the vector autoregression model (Mao et al. 2018).
110   AIC value of a model is calculated as follows:

$$AIC = 2k - 2\ln(\hat{L}), \qquad (8)$$

111   where k is the number of estimated parameters in the model and $\hat{L}$ is the maximum value of the likelihood function
112   for the model.
113   BIC value of a model is calculated as follows:

$$BIC = \ln(n)k - 2\ln(\hat{L}), \qquad (9)$$

114   where n is the number of data points in x or the number of observations.
115   When fitting the models, it is possible to increase the simulation accuracy by adding parameters, but doing so
116   may result in overfitting. Both BIC and AIC attempt to solve this problem by introducing a penalty term for the
117   number of parameters in the model. And the lag order corresponding to the lowest AIC value or BIC value is taken
118   as the optimal lag order. We use maximum likelihood estimation[2] to estimate the model parameters.
119   The stability test of the established model is required because if the model is unstable, some results will not be
120   valid (such as the standard error of the impulse response function). In this paper, we use the AR root test and if all
121   the roots of the vector autoregression model estimated based on empirical data have a reciprocal of less than 1 (that
122   is, if they are within the unit circle), they are stable.
123   The way to check whether a vector autoregression model is stationary is as follow. For vector autoregression
124   model

$$(y_t - \mu) = \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + \ldots + \phi_p(y_{t-p} - \mu) + \varepsilon_t$$

126   where y_t is a vector in R^n, we first write it in the form of deviation. Then we define matrix F:

---

[2] The method of maximum likelihood is based on the likelihood function. Suppose we are given a family of distributions $\{f(\cdot; \theta) | \theta \in \Theta\}$, where θ denotes the parameters (possibly multi-dimensional) for the model. The method defines a maximum likelihood estimate: $\hat{\theta} \in \{\underset{\theta \in \Theta}{\mathrm{argmax}}\ \mathcal{L}(\theta; x)\}$, where $\mathcal{L}(\theta; x)$ denotes the likelihood function, because intuitively this selects parameter values that make the data most probable.

127
$$F = \begin{bmatrix} \phi_1 & \phi_2 & \phi_3 & \cdots & \phi_{p-1} & \phi_p \\ I_n & 0 & 0 & \cdots & 0 & 0 \\ 0 & I_n & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & I_n & 0 \end{bmatrix}$$

128 where I_n is a unit matrix of n dimension. If all characteristic roots of matrix F fall within a unit circle, then the vector
129 autoregression model is stable.

130 Impulse Response Function

131   The impulse response function is used to characterize the influence of a standard deviation shock of random
132 perturbation terms on the current and future values of other variables. It can visually describe the interactions and
133 effects between variables (Schoukens et al. 2018). We take vector autoregression that has one lag order and contains
134 tow variables (y and z) as an example. The matrix form of structure a vector autoregression model can be written as
135 follow.

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix},$$
(10)

136 Then we write it in the standard form as follow.

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} a_{10} \\ a_{20} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix},$$
(11)

137 Moreover, we can write it in the following form if the vector autoregression model is stable.

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \overline{y_t} \\ \overline{z_t} \end{bmatrix} + \Sigma_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} e_{1t-i} \\ e_{2t-i} \end{bmatrix},$$
(12)

138 For the transformation from (10) to (11), the error term can be transformed as follow.

$$\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} = 1/(1 - b_{12}b_{21}) \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix},$$
(13)

139 By combining equation (12) and (13), we obtain

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \overline{y_t} \\ \overline{z_t} \end{bmatrix} + 1/(1 - b_{12}b_{21}) \Sigma_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{yt-i} \\ \varepsilon_{zt-i} \end{bmatrix},$$
(14)

140 In order to simplify the equation, we define matrix ϕ_i, and denote its elements as $\phi_{jk}(i)$.

$$\phi_i = A_1^i/(1 - b_{12}b_{21}) \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix},$$
(15)

141 Thus, the moving average of the equation (11) and (12) can be expressed as

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \overline{y_t} \\ \overline{z_t} \end{bmatrix} + \Sigma_{i=0}^{\infty} \begin{bmatrix} \phi_{11}(i) & \phi_{12}(i) \\ \phi_{21}(i) & \phi_{22}(i) \end{bmatrix} \begin{bmatrix} \varepsilon_{yt-i} \\ \varepsilon_{zt-i} \end{bmatrix},$$
(16)

142 Moving average is a useful tool for studying the mutual influence between the two series $\{y_t\}$ and $\{z_t\}$, for the
143 impulse response function of $\{y_t\}$ and $\{z_t\}$ to one unit shock of ε_{yt} and ε_{zt} can be built using the coefficients of ϕ_i.
144 Here $\phi_{11}(i)$, $\phi_{12}(i)$, $\phi_{21}(i)$ and $\phi_{22}(i)$ are called impulse response functions. And the graphs of these functions
145 show the impulse response of the time series to one unit of positive shock.
146   To be more specific, coefficient $\phi_{12}(0)$ indicates the current impact of one unit positive change in ε_{zt} have on
147 y_t. Similarly, after one period of calibration, $\phi_{11}(1)$ and $\phi_{12}(1)$ also denote the impact of one unit positive change
148 in ε_{yt} and ε_{zt} have on y_{t+1}. The cumulative effect of ε_{yt} and ε_{zt}'s unit impulse response can be obtained by the
149 appropriate add up of the coefficients of the impulse response functions. For instance, after n periods the impact of

150 ε_{zt} on y_{t+n} is $\phi_{12}(n)$. Therefore, after n periods the accumulative impact of ε_{zt} on y_t is $\sum_{i=0}^{n}\phi_{12}(i)$. If n tends to
151 be positive infinite, the accumulative effect can be obtained. Since we have assumed that y_t and z_t are stationary,
152 when j and k approximate infinite, the value of $\phi_{jk}(i)$ is 0. Since all parameters of the vector autoregression system
153 can be calculated, it is totally possible to track every value of ε_{yt} and ε_{zt}'s impact on $\{y_t\}$ and $\{z_t\}$ can be calculated.
154 The time when $\phi_{jk}(i)$ approximates 0 is the time lag of impulse response, and the gap between the positive largest
155 value and negative largest value of $\phi_{jk}(i)$ can be used to describe the intensity of the impulse response.

156 Granger Causality Test

157   The Granger causality test is used to test if the variable y can be used to predict the variable x. That is when the
158 variable x is regressed according to the past value of the variable y, the explanatory ability of the regression can be
159 significantly enhanced. It should be noted that two variables have a temporal "causal relationship" but do not
160 necessarily have a logical causal relationship.
161   In an attempt to test whether past values of y help predict x, we first perform an ordinary least squares (OLS)
162 estimation:

$$x_t = c_1 + a_1 x_{t-1} + a_2 x_{t-2} + \cdots + a_p x_{t-p} + b_1 y_{t-1} + b_2 y_{t-2} + \cdots + b_p x_{t-p} + u_t, \qquad (17)$$

163 And we propose the null hypothesis:

$$H_0: b_1 = b_2 = \cdots = b_p = 0, \qquad (18)$$

164 which means that y is not the Granger reason for x. To conduct an F-test for this hypothesis, first, we estimate the
165 equation without restriction H_0, and obtain the sum of residual:

$$RSS_0 = \sum_{t=1}^{T}\hat{u}_t^2, \qquad (19)$$

166 Then, we estimate the equation with restriction H_0, and obtain the sum of residual:

$$RSS_1 = \sum_{t=1}^{T}\hat{e}_t^2, \qquad (20)$$

167 We build the F statistic using the expression below:

$$\frac{(RSS_1 - RSS_0)/p}{RSS_0/(T-2p-1)} \sim F(p, T-2p-1), \qquad (21)$$

168 where T is the size of the sample. Based on the data, if the value of F statistic is higher than the critical value, then
169 we reject the null hypothesis and admit that increasing the lag order of y can significantly improve the explanatory
170 ability of the model. Under such circumstances, we say that the past value of y helps predict x (Stokes & Purdon
171 2018).

172 Stochastic Model System for Soil Moisture-heat Coupling

173   First, we build three vector autoregression models at the depth of 10 cm, 30 cm, and 90 cm. We choose these
174 three depths because, during the greening period which is around 15th March, the diameter of wheat root reaches
175 its maximum approximately at the depth of 10 cm, 30 cm, and 90 cm, respectively, thus resulting in more water
176 absorption at these three depths. In addition, usually, the soil is divided into topsoil, subsoil, and substratum, the
177 representative depths of which approximate 10 cm, 30 cm, and 90 cm. In addition, we select hourly data on soil
178 moisture and temperature from December to May because this period covers the main growing stages of winter
179 wheat. This time period can be adjusted according to the growing stages of different wheat species (Awad et al.
180 2018).
181   Second, we build six impulse response functions for these three vector autoregression systems. Based on the
182 results of impulse response function analysis, we are able to obtain the time lag and intensity of impulse response.
183 For example, if one unit of positive shock is given to soil temperature at the depth of 10 cm, we calculate the time
184 length during which obvious deviation from zero can be seen in soil moisture's response to 1°C of positive variation.
185 This provides us with information on how long soil moisture will be influenced if there is a sudden change in

186 temperature. In addition, the gap between the highest point and lowest point of the response is the intensity of
187 impulse response. The larger the intensity, the stronger the influence of soil temperature has on moisture.
188   Third, we conduct six Granger causality tests between two variables of soil moisture and temperature at three
189 depths. According to the result of the Granger causality test, we are able to decide a more proper way of forecasting
190 soil moisture and temperature by taking the Granger causal relationship of moisture and temperature into
191 consideration. If, for example, at the depth of 10 cm the past value of soil temperature helps predict soil moisture,
192 it means that the past value of soil temperature influences soil moisture. Therefore, when establishing empirical
193 models forecasting soil moisture, we should choose the ARMAX model which is an ARMA model with an exogenous
194 variable in an attempt to take into consideration of temperature's influence on moisture. On the other hand, if the
195 past value of soil temperature does not help predict soil moisture, it means that the past value of soil temperature
196 does not have a significant influence on soil moisture. If soil temperature and moisture are Granger reasons for each
197 other, it suggests that the present value of two variables are related to each other's past value. Thus, a vector
198 autoregression model treating all variables as endogenous variables will be the optimal choice.

199 *Field experiment on soil moisture and temperature*

200   The experiment on the wheat of Bainong Dwarf Anti-floating No.58[3] was carried out at Shangqiu Experimental
201 Station of Farmland Irrigation Institute of Chinese Academy of Agricultural Sciences with a latitude of 34° 35.222'N,
202 longitude of 115° 34.515'E, and elevation of 55.6 m. The average annual temperature there is 13.9°C, and the frost-
203 free period is 180 d to 230 d. In addition, the average annual precipitation is 708 mm with the precipitation from
204 July to September accounting for 65%~75% of the total annual precipitation, and the average annual evaporation
205 being 1735 mm. The soil of the test site is a loam, which refers to soil consists of clay (30%~40%), silt (30%~40%),
206 and sand (30%~40%).
207   The Irrigation method adopted is drip irrigation at the depth of 30 cm under soil surface. The sensor we used
208 for soil water monitoring is called "Soil-Water", which was produced in Australia. The basic components of this soil
209 monitoring system are data collectors, solar modules, mounting components, and soil sensors. The accuracy for
210 temperature is <0.1°C, and but the accuracy of soil moisture sensor was not provided by the manufacturer. We
211 installed the instrument probes of the moisture meter at the depth of 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100 cm,
212 and measured soil moisture and soil temperature at these depths simultaneously. The instrument's temperature
213 measurement range is from -20°C to 60°C.
214   Among these ten sets of data, we select hourly data on soil moisture and soil temperature from December 19th,
215 2017 to May 30th, 2018 at the depth of 10 cm, 30 cm and 90 cm, which is in accordance with the model designed
216 above. In addition, the selection of the measurement interval takes into account both the memory capacity of the
217 instrument and the minimum scale of moisture transport, since one-hour interval is sufficient to show the footprint
218 of water movement.

Results

220   Figure 1 shows data on soil moisture and temperature measured at the depth of 10, 20, 30, 40, 50, 60, 70, 80,
221 90 and 100 cm. We can see that both soil temperature and soil moisture content exhibits periodical variations within
222 a day and trends changing with the alternation of seasons. Soil temperature decreases with the cooling trend of
223 climate from December to January and increases with the overall warming of the climate from February to May in
224 next year. Soil moisture content, however, experiences a sudden hike around 2nd March and 19th March from 10 cm
225 to 50 cm due to irrigation, and gradually decreases from March to May due to evapotranspiration, indicating an
226 active water absorption process of winter wheat root system. Soil moisture content exhibits obvious differences
227 among different depths.

228   **Figure 1.** Soil moisture (a) and temperature (b) at different depths.

---

[3] Bainong Dwarf Anti-floating No.58 is a new epoch-making wheat variety developed in 2003 by the Wheat Breeding Center in
Henan University of Science and Technology. This variety belongs to the semi-winter mid-maturing variety. The plant height is
about 70 cm, with high resistance to lodging and good fullness. It has high resistance to powdery mildew, stripe rust, leaf blight,
and medium resistance to sheath blight, strong root activity, good ripening, and yellowing. It also has the preponderance of high
yield with the average yield being 7500-8250 kg/ha and the maximum being 10500 kg/ha.

229   We use hourly data from December 19th, 2017 to May 30th, 2018 on soil moisture content at the soil depths of
230   10 cm, 30 cm and 90 cm (hereinafter referred to as *MC10, MC30* and *MC90*, respectively), soil temperature at the
231   depth of 10 cm, 30 cm and 90 cm (hereinafter referred to as *ST10, ST30* and *ST90*, respectively).
232   First, we remove the seasonality of the time series of *MC10, ST10, MC30, ST30, MC90* and *ST90*, since it is
233   known from the above figures that the six hourly time series have obvious diurnal pattern due to the periodical
234   variations of meteorological factors, evapotranspiration, and root water uptake within a day, especially in shallow
235   layers. After seasonal difference, the period of the data is from December 20th, 2017 to May 30th, 2018. The
236   corresponding time series of the six detrended variables are referred to as *MCSD10, STSD10, MCSD30, STSD30,*
237   *MCSD90,* and *STSD90*, respectively (Figure 2).

238   **Table 1.** Unit root test results of the deseasonalized soil moisture and temperature series.

239   [1] *T stat.* is the t-statistic for performing the ADF test. [2] *Model* indicates the type of hypothesis selected and the lag period which
240   is added to make the residual become white noise. [3] *P-Value* represents the corresponding p-value. [4] *C.V.* indicates critical value.[5]
241   * indicates a significant level of 10%, ** indicates a significant level of 5%, and *** indicates a significant level of 1%. [6] *Stationary*
242   indicates that the time series is stable.

243   **Figure 2.** Deseasonalized soil temperature (a) and moisture (b) at the depth of 10cm, 30cm, and 90cm.

244   According to the above results of the ADF test (Table 1), the variables of *MCSD10, STSD10, MCSD30, STSD30,*
245   and *MCSD90* are all stationary under the significance level of 0.01. The variable of STSD90 is stationary under the
246   significance level of 0.05. Thus, it is appropriate to use the variables of *MCSD10, STSD10, MCSD30, STSD30, MCSD90*
247   and *STSD90* to build vector autoregression models since they are all stationary.

248   *Vector Autoregression Model Establishment*

249   The values of AIC and BIC for vector autoregression models of different lags are calculated (Table 2) to
250   determine the optimal lag orders for the vector autoregression models.

251   **Table 2.** Calculated $\ln\left(\hat{L}\right)$, AIC and BIC for vector autoregression model of different lags at the depth of 10 cm, 30
252   cm, and 90 cm.

253   For soil depth of 10 cm, among the eight lag orders, lag order 8 has the lowest AIC value and lag order 5 has
254   the lowest BIC value. Since model with a lag order of 5 has fewer parameters to estimate which leads to smaller
255   estimation error for the whole model, we choose 5 as the optimal lag order (Table 2). Based on the optimal lag
256   structure and the time series of *MCSD10* and *STSD10*, we estimate the parameters of the vector autoregression
257   model using maximum likelihood estimation. Most of the t-statistic of the estimated coefficients in the above
258   regression model is significant at the 10% significance level. Although some of the coefficients are not significant, it
259   may because multiple hysteresis values with the same variables in the same equation result in multiple collinearities.
260   Thus, finally, we obtained the estimated vector autoregression model:

$$
\begin{bmatrix} MCSD10_t \\ STSD10_t \end{bmatrix} = \begin{bmatrix} -0.002 \\ 0.005 \end{bmatrix} + \begin{bmatrix} 1.593 & -0.026 \\ 0.167 & 1.754 \end{bmatrix}\begin{bmatrix} MCSD10_{t-1} \\ STSD10_{t-1} \end{bmatrix} + \begin{bmatrix} -0.631 & 0.019 \\ -0.205 & -0.816 \end{bmatrix}\begin{bmatrix} MCSD10_{t-2} \\ STSD10_{t-2} \end{bmatrix}
$$
$$
+ \begin{bmatrix} 0.274 & 0.012 \\ -0.061 & 0.026 \end{bmatrix}\begin{bmatrix} MCSD10_{t-3} \\ STSD10_{t-3} \end{bmatrix} + \begin{bmatrix} -0.526 & 0.024 \\ 0.094 & -0.060 \end{bmatrix}\begin{bmatrix} MCSD10_{t-4} \\ STSD10_{t-4} \end{bmatrix} + \begin{bmatrix} 0.257 & -0.031 \\ 0.008 & 0.068 \end{bmatrix} \quad (22)
$$
$$
\begin{bmatrix} MCSD10_{t-5} \\ STSD10_{t-5} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}
$$

261   For soil depth of 30 cm, Among the eight lag orders, lag order 8 has the lowest AIC value and the lowest BIC
262   value. Thus, we choose 8 as the optimal lag order and build a vector autoregression model (Table 2). We obtained
263   the estimated vector autoregression model as follow:

$$\begin{bmatrix} MCSD30_t \\ STSD30_t \end{bmatrix} = \begin{bmatrix} -0.002 \\ 0.001 \end{bmatrix} + \begin{bmatrix} 1.182 & 0.117 \\ -0.006 & 0.559 \end{bmatrix} \begin{bmatrix} MCSD30_{t-1} \\ STSD30_{t-1} \end{bmatrix} + \begin{bmatrix} -0.173 & -0.090 \\ -0.022 & 0.344 \end{bmatrix} \begin{bmatrix} MCSD30_{t-2} \\ STSD30_{t-2} \end{bmatrix}$$

$$+ \begin{bmatrix} -0.016 & 0.020 \\ 0.042 & 0.228 \end{bmatrix} \begin{bmatrix} MCSD30_{t-3} \\ STSD30_{t-3} \end{bmatrix} + \begin{bmatrix} -0.002 & -0.054 \\ -0.003 & 0.149 \end{bmatrix} \begin{bmatrix} MCSD30_{t-4} \\ STSD30_{t-4} \end{bmatrix} + \begin{bmatrix} 0.000 & 0.013 \\ -0.022 & 0.012 \end{bmatrix}$$

$$\begin{bmatrix} MCSD30_{t-5} \\ STSD30_{t-5} \end{bmatrix} + \begin{bmatrix} -0.006 & -0.014 \\ 0.019 & -0.074 \end{bmatrix} \begin{bmatrix} MCSD30_{t-6} \\ STSD30_{t-6} \end{bmatrix} + \begin{bmatrix} 0.012 & 0.040 \\ -0.010 & -0.082 \end{bmatrix} \begin{bmatrix} MCSD30_{t-7} \\ STSD30_{t-7} \end{bmatrix} + \quad (23)$$

$$\begin{bmatrix} -0.027 & -0.029 \\ -0.002 & -0.148 \end{bmatrix} \begin{bmatrix} MCSD30_{t-8} \\ STSD30_{t-8} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

264 For soil depth of 90 cm, Among the eight lag orders, lag order 8 has the lowest AIC value and lag order 7 has
265 the lowest BIC value. Since model with a lag order of 7 has fewer parameters to estimate which leads to smaller
266 estimation error for the whole model, we choose 7 as the optimal lag order (Table 2). We obtained the estimated
267 vector autoregression model as follow:

$$\begin{bmatrix} MCSD90_t \\ STSD90_t \end{bmatrix} = \begin{bmatrix} 0.690 & 0.058 \\ 0.133 & 0.345 \end{bmatrix} \begin{bmatrix} MCSD90_{t-1} \\ STSD90_{t-1} \end{bmatrix} + \begin{bmatrix} 0.336 & 0.025 \\ 0.026 & 0.200 \end{bmatrix} \begin{bmatrix} MCSD90_{t-2} \\ STSD90_{t-2} \end{bmatrix} +$$

$$\begin{bmatrix} 0.130 & -0.001 \\ 0.045 & 0.191 \end{bmatrix} \begin{bmatrix} MCSD90_{t-3} \\ STSD90_{t-3} \end{bmatrix} + \begin{bmatrix} -0.024 & -0.010 \\ 0.079 & 0.103 \end{bmatrix} \begin{bmatrix} MCSD90_{t-4} \\ STSD90_{t-4} \end{bmatrix} + \begin{bmatrix} -0.061 & -0.031 \\ -0.166 & 0.039 \end{bmatrix} \quad (24)$$

$$\begin{bmatrix} MCSD90_{t-5} \\ STSD90_{t-5} \end{bmatrix} + \begin{bmatrix} -0.004 & -0.008 \\ -0.132 & 0.056 \end{bmatrix} \begin{bmatrix} MCSD90_{t-6} \\ STSD90_{t-6} \end{bmatrix} + \begin{bmatrix} -0.071 & -0.032 \\ 0.011 & 0.058 \end{bmatrix} \begin{bmatrix} MCSD90_{t-7} \\ STSD90_{t-7} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

268 Then we check the stability of these three vector autoregression models. Figure 3 shows that all unit roots fall
269 within the unit root circle, so it is reasonable to believe that the vector autoregression models at these three soil
270 depths are stable, indicating that there is a long-term stable relationship between the variables selected which can
271 be further analyzed.

272 **Figure 3.** Inverse roots of AR characteristic polynomial at 10 cm (a), 30 cm (b) and 90 cm (c).

273 *Impulse Response Function Analysis*

274 The impulse response results at three different depths are shown in Table 3.

275 **Table 3**. Impulse response results of *MCSD10, STSD10, MCSD30, STSD30, MCSD90* and *STSD90*.

276 [1]RP represents retrospective periods, its unit is an hour.

277 The impulse response diagrams at three different depths are shown in Figure 4. The horizontal axis represents
278 the number of retrospective periods from 0 h to 120 h, while the vertical axis represents the response of the
279 dependent variable to the shock. The confidence interval is used because the vector autoregression model
280 coefficients have some error in the estimation. Setting the confidence interval can accommodate the inherent
281 uncertainty of the parameters.

282 **Figure 4.** Impulse response of *MCSD10* to shock in *STSD10* (a), impulse response of *STSD10* to shock in *MCSD10* (b),
283 impulse response of *MCSD30* to shock in *STSD30* (c), impulse response of *STSD30* to shock in *MCSD30* (d), impulse
284 response of *MCSD90* to shock in *STSD90* (e), impulse response of *STSD90* to shock in *MCSD90* (f). The solid line of blue
285 fluorescence represents the impulse response and the red dashed line represents the confidence interval of the
286 impulse response with a confidence level of 0.90.

287 When *STSD10* is given one unit of positive shock, *MCSD10* reacts to 0.0312% in the current period, and then
288 monotonically increases until it reaches a positive maximum of 0.0638% in the ninth period. And then it gradually
289 decreases and falls near zero in the long run indicating that the vector autoregression system is stable. It can be seen
290 that the impact of *STSD10* on *MCSD10* has a lagging effect since its impulse response is only close to zero after
291 around 25 periods. Moreover, the intensity of the influence which can be quantified as 0.0638% is relatively high
292 (Figure 4. (a)).
293 When *MCSD10* is given one unit of positive shock, *STSD10* reacts to 0.0453°C in the current period, and then
294 monotonically increases until it reaches a positive maximum of 0.2004°C in the fourth period. And then it gradually
295 decreases, and fall near zero in the long run indicating that the vector autoregression system is stable. It can be seen
296 that the impact of *MCSD10* on *STSD10* has a lagging effect since its impulse response never falls below zero and is
297 only close to zero after 50 periods. Moreover, the intensity of the influence which can be quantified as 0.2004°C is
298 quite high (Figure 4. (b)).
299 When *STSD30* is given one unit of positive shock, *MCSD30* reacts to -0.0077% in the current period, and then
300 monotonically increases until it reaches a positive maximum of 0.0086% in the 22nd, 23rd, 24th period. And then it
301 gradually decreases and falls near zero in the long run indicating that the vector autoregression system is stable.
302 Besides, the impact of *STSD30* on *MCSD30* has a lagging effect based on the fact that it is only close to zero after
303 around 50 periods. Moreover, the intensity of the influence which can be quantified as 0.0163% is lower than the
304 intensity of *MCSD10*'s response to *STSD10* (Figure 4. (c)).
305 When *MCSD30* is given one unit of positive shock, *STSD30* reacts to -0.0045°C in the current period, and then
306 monotonically decreases until it reaches a negative maximum of -0.0163°C at the 36th, 37th, 38th, 39th period. And
307 then it gradually increases, and fall near zero in the long run indicating that the vector autoregression system is
308 stable. It can be seen that the impact of *MCSD30* on *STSD30* has a lagging effect since its impulse response is only
309 close to zero after the 120th period. Moreover, the intensity of the influence which can be quantified as 0.0163°C is
310 lower than the intensity of *STSD10*'s response to *MCSD10* (Figure 4. (d)).
311 When *STSD90* is given one unit of positive shock, *MCSD90* reacts to -0.0033% in the current period, and then
312 monotonically increases until it reaches a positive maximum of 0.0017% from the 17th to 30th period. And then it
313 gradually decreases and falls near zero in the long run indicating that the vector autoregression system is stable. It
314 can be seen that the impact of *STSD90* on *MCSD90* has a lagging effect based on the fact that it is only close to zero
315 after 120 periods. Moreover, the intensity of the influence which can be quantified as 0.0050% is lower than the
316 intensity of *MCSD30*'s response to *STSD30* (Figure 4. (e)).
317 When *MCSD90* is given one unit of positive shock, *STSD90* reacts to -0.0107°C in the current period, and then
318 monotonically increases and reaches a relatively stable value of -0.0035°C at around the 120th period. It can be seen
319 that the impact of *MCSD90* on *STSD90* has a lagging effect. Moreover, the intensity of the influence which can be
320 quantified as 0.0035°C is lower than the intensity of *STSD30*'s response to *MCSD30* (Figure 4. (f)).

321 *Granger Causality Test*

322 By carrying out Granger causality test, we figured out whether the past value of *STSD10* (*STSD30, STSD90*) helps
323 predict *MCSD10* (*MCSD30, MCSD90*) and whether the past value of *MCSD10* (*MCSD30, MCSD90*) helps predict
324 *STSD10* (*STSD30, STSD90*).
325 From the results in Table 4 and Table 5, it is known that the past value of *MCSD10* helps predict *STSD10*, but
326 the past value of *STSD10* does not help predict *MCSD10*. This means that the past value of soil moisture content is
327 helpful in terms of predicting the present value of soil temperature at the depth of 10 cm, while soil temperature
328 has a small influence in terms of predicting soil moisture content at the depth of 10 cm. Besides, the results also
329 indicate soil moisture content varies ahead of soil temperature.
330 From Table 6 and Table 7, we can see that at the significance level of 0.05 and 0.10, the past value of *MCSD30*
331 helps predict *STSD30* and the past value of *STSD30* helps predict *MCSD30*. This suggests that the past value of soil
332 moisture content is helpful in terms of predicting the present value of soil temperature at the depth of 30 cm, and
333 the past value of soil temperature is helpful in terms of predicting the present value of soil moisture content at the
334 depth of 30 cm.
335 From Table 8 and Table 9, we can see that at the significance level of 0.01, 0.05 and 0.10, the past value of
336 *MCSD90* helps predict *STSD90* and the past value of *STSD90* helps predict *MCSD90*. This suggests that the past value
337 of soil moisture content is helpful in terms of predicting the present value of soil temperature at the depth of 90 cm,

338 and the past value of soil temperature is helpful in terms of predicting the present value of soil moisture content at
339 the depth of 90 cm.

340 **Table 4.** Granger causality test determining whether *STSD10* helps predict *MCSD10*.

341 **Table 5.** Granger causality test determining whether *MCSD10* helps predict *STSD10*.

342 **Table 6.** Granger causality test determining whether *STSD30* helps predict *MCSD30*.

343 **Table 7.** Granger causality test determining whether *MCSD30* helps predict *STSD30*.

344 **Table 8.** Granger causality test determining whether *STSD90* helps predict *MCSD90*.

345 **Table 9.** Granger causality test determining whether *MCSD90* helps predict *STSD90*.

346 **Discussions**

347 *Specific results*

348 Firstly, the time lag of soil temperature's response to shock in soil moisture is about 25 h at 10 cm, 50 h at 30
349 cm, and 120 h at 90 cm, while the time lag of soil moisture's response to shock in soil temperature is for about 50 h
350 at 10 cm and longer than 120 h at 30 cm and 90 cm.
351 Secondly, the intensity of soil temperature's impulse response to shock in soil moisture decreases from
352 0.2004°C at the depth of 10 cm, to 0.0163°C at the depth of 30 cm, and finally to 0.0035°C at the depth of 90 cm.
353 Similarly, the intensity of soil moisture's impulse response to soil temperature also decreases from 0.0638% at the
354 depth of 10 cm, to 0.0163% at the depth of 30 cm, and finally to 0.0050% at the depth of 90 cm.
355 Thirdly, soil moisture is the Granger reason of soil temperature at the depth of 10 cm, 30 cm, and 90 cm. This
356 means that the past value of soil moisture is helpful in terms of predicting the present value of soil temperature, and
357 probably is the actual logical reason for changes in soil temperature. However, the causal relationship does not
358 necessarily mean that there exists a direct connection between soil temperature and moisture, since a series of
359 physical and biological processes may complicate the relations between soil temperature and moisture. While soil
360 temperature does not help predict soil moisture at the depth of 10 cm, it helps predict soil moisture at the depth of
361 30 cm and 90 cm. Therefore, according to the basic principles of Granger causality test, when predicting the dynamic
362 variation of soil temperature and moisture at the depth of 30 cm and 90 cm, the other variable's past values should
363 be taken into consideration, which means that a vector autoregression model may be a better choice. In comparison,
364 at the depth of 10 cm, the past value of soil temperature is not helpful for predicting the present value of soil
365 moisture. However, it may be more accurate to incorporate soil moisture as an exogenous variable when predicting
366 soil temperature at 10 cm.

367 *Verification of the model*

368 To begin with, the model we proposed can be supported and explained by established mechanistic models. In
369 1966, Philip proposed the concept of a complete Soil-Plant-Atmosphere Continuum (SPAC), laying the theoretical
370 foundation for modern farmland water research. Based on this, Meng and Xia (2005) established a dynamic coupling
371 model by calibrating parameters that describes the hydrothermal conditions during crop growth and the law for
372 crop transpiration. Through this coupling model, they hope to reveal the law for moisture and heat transfer of the
373 SPAC. The SPAC is divided into three layers, namely, the atmosphere at a high altitude, the plant canopy which is
374 simplified into one layer at the momentum transfer junction, and the soil layer. To be more specific, the top of the
375 soil layer is set to be the soil surface, and the bottom of the soil layer is set to be at the groundwater level. According
376 to the mathematical expression of total latent and sensible heat consumption of SPAC, the mathematical expression
377 of plant transpiration latent and sensible heat consumption, the mathematical expression of soil evaporation latent
378 and sensible heat consumption, and the model of soil moisture migration and heat transfer, the results obtained
379 from data-driven vector autoregression models can be reasonably explained.

380   To begin with, the lagging effect of impulse response between soil temperature and moisture can be explained
381   as follows. Moisture's lagging effect on temperature can be partly interpreted by the basic heat transfer equation:

$$C_{\text{v}}\frac{\partial T}{\partial t} = \frac{\partial}{\partial z}\left(K_{\text{h}}\frac{\partial T}{\partial z}\right), \tag{25}$$

382   where T is soil temperature, C_{v} is soil volumetric heat capacity, and K_{h} is soil thermal conductivity. C_{v} and K_{h} are
383   close correlated with soil moisture content θ and can be expressed in the single factor function form of θ. From the
384   equation above, we can deduce that when there is a fluctuation in soil moisture content, the first-order derivative
385   of temperature T's function on time t will change. Thus, although the contemporary value of temperature will not
386   change, it's successive value will change gradually in accordance with the change of the first-order derivative. Finally,
387   at some point after the interference point, the deviation of temperature will reach its maximum, accounting for
388   moisture's lagging effect on temperature. Temperature's lagging effect on moisture is to some extent related to the
389   relationship between water vapor relative saturation and soil temperature.

$$h_2 = exp\left(\frac{Mg\psi_2}{R(T_2 + 273.16)}\right), \tag{26}$$

390   where h_2 is the water vapor relative saturation of the air at the soil surface, ψ_2 represents water potential at the soil
391   surface, R is the universal gas constant, and T_2 denotes soil surface temperature. Though other possible reasons for
392   the lagging effect may exist, one major process can be deduced from the expression above, where a fluctuation in
393   soil temperature will first cause deviation in water vapor relative saturation, and then change in water vapor relative
394   saturation will cause the corresponding variation in soil moisture through a series of complicated processes. In this
395   way, a time lag exists.
396   The result that the time lag and intensity of the impulse response change with depth can be explained by several
397   processes, some of which can be expressed as follows.

$$S_{\text{w}} = [(4m - 1)/lr(t) - (8m - 4)/lr^2(t)]E_{\text{v}}(t) \tag{27}$$

$$E_{\text{v}}(t) = \int_0^{lr(t)} S_{\text{w}}(z,t)dz \tag{28}$$

398   Where S_{w} is the water absorption intensity of roots, m denotes the ratio of water absorption rate of the upper part
399   to that of the lower part of roots, $lr(t)$ represents root depth, and E_{v} is crop transpiration rate. Equation (27)
400   describes how S_{w} is influenced by depth and time. Equation (28) describes the way in which plant transpiration is
401   influenced by S_{w} at different soil depths (Lei et al., 1988). Given that there exists a close relationship between soil
402   temperature and plant transpiration latent and sensible heat consumption, it can be deduced that soil temperature
403   varies with depth.
404   In addition to physical explanations, we used data from a different year and partly verified the universality of
405   the model. We collected data on soil temperature and moisture content at the same test site from December 19th,
406   2016 to May 30th, 2017 at the depth of 10 cm, 30 cm, and 90 cm, during which the same irrigation method, i.e., drip
407   irrigation, has been adopted. Then, we used the same method stated above to build three vector autoregression
408   models. Next, we built impulse response functions and conducted Granger causality tests. The results we obtained
409   are as follows. First, the time lag of soil moisture's impulse response is 30 h, 70 h, and longer than 120 h, while the
410   time lag of temperature's impulse response is about 60 h, longer than 120 h, and longer than 120 h at 10 cm, 30 cm,
411   and 90 cm. Second, temperature's response intensity is 0.1989°C, 0.0157°C and 0.0031°C for 1% of variation in soil
412   moisture, and moisture's response intensity is 0.0578%, 0.0169% and 0.0057% for 1°C of variation in soil
413   temperature at 10 cm, 30 cm and 90 cm. Third, soil moisture is helpful in terms of predicting soil temperature at the
414   depth of 10 cm, 30 cm, and 90 cm. Besides, soil temperature is helpful in terms of predicting soil moisture at the
415   depth of 10 cm and 30 cm but has no obvious relationship with soil moisture at 90 cm.
416   By comparing the results with those of 2017-2018, we can discover that the ratios for the intensity of soil
417   temperature's impulse response to moisture among three depths during both periods approximate 200:16:3. And
418   the ratios for the intensity of moisture's impulse response to temperature during both periods approximate 60:17:6.
419   Moreover, the results of Granger causality tests are the same, indicating certain stability in terms of the physical and
420   biological processes that involve the variation of soil temperature and moisture in the same area. This demonstrates
421   that for the same area the data-driven model we proposed has consistency and the results we obtained are not out

422 of a sudden. Thus, the method of building data-driven vector autoregression models can be used to study the
423 characteristics of soil heat-moisture coupling. Although the numerical results of this model will change with different
424 soil properties in different areas, the method can be applied in the same way. The data and MATLAB code for training
425 and testing the models have been included in the supplementary files for readers to verify our findings.

426 **Conclusions**

427 We designed a purely data-driven stochastic model for analyzing moisture-heat coupling of a wheat-soil
428 system, which consists of three vector autoregression models built at the depth of 10 cm, 30 cm, and 90 cm, impulse
429 response functions, and Granger causality tests. For the empirical test of this method, we use the hourly data on soil
430 moisture and soil temperature at the depth of 10 cm, 30 cm, and 90 cm obtained at Shangqiu Experiment Station.
431 Following conclusions can be drawn from the models.
432 Firstly, the time lag of soil temperature's influence on soil moisture is for about 25 h at 10 cm, 50 h at 30 cm,
433 and 120 h at 90 cm, while the time lag of soil moisture's influence on soil temperature is for about 50 h at 10 cm and
434 longer than 120 h at 30 cm and 90 cm. Secondly, the intensity of soil temperature's impulse response to shock in soil
435 moisture is 0.2004°C, 0.0163°C and 0.0035°C for 1% variation in soil moisture respectively at the depth of 10 cm, 30
436 cm, and 90 cm. Similarly, the intensity of soil moisture's impulse response is 0.0638%, 0.0163% and 0.0050% for 1°C
437 of variation in soil temperature respectively at the depth of 10 cm, 30 cm, and 90 cm. Thirdly, soil moisture is helpful
438 in terms of predicting soil temperature at the depth of 10 cm, 30 cm, and 90 cm. While soil temperature helps predict
439 soil moisture at the depth of 10 cm and 30 cm, it has no obvious correlation with soil moisture at 90 cm.
440 Although the proposed method for analyzing moisture-heat coupling in a wheat-soil system has some
441 advantages such as data-driven and easy to achieve, there are limitations. One is that the vector autoregression
442 model is based on the linear hypothesis, so other nonlinear data-driven models should be further studied. Also, since
443 the soil of our experiment site is loam, we ought to further explore whether such a method can be applied to other
444 kinds of soil using data from different experiment sites. We are currently working on this.

445 **References**

446 1. Awad, W., Byrne, P.F., Reid, S.D., Comas, L.H. & Haley, S.D 2018 Great Plains Winter Wheat Varies for Root Length and
447 Diameter under Drought Stress. Agronomy Journal, **110**(1), 226-235, doi: 10.2134/agronj2017.07.0377.
448 2. Bittelli M., Ventura F., Campbell G. S., Snyder R. L., Gallegati F. & Pisa P. R. 2008 Coupling of heat, water vapor, and liquid
449 water fluxes to compute evaporation in bare soils. *Journal of Hydrology*, **362**(3-4), 191–205, doi:
450 10.1016/j.jhydrol.2008.08.014.
451 3. Bolin D., Wallin J. & Lindgren F. (in press) Latent Gaussian random field mixture models. *Computational Statistics & Data
452 Analysis*, **130**, 80-93, doi: 10.1016/j.csda.2018.08.007.
453 4. Bouri E., Gupta R., Hosseini S. & Lau C. K. M. 2018 Does global fear predict fear in BRICS stock markets? Evidence from a
454 Bayesian graphical structural VAR model. *Emerging Markets Review*, **34**, 124-142, doi: 10.1016/j.ememar.2017.11.004.
455 5. Chen L.-J., Ho Y.-H., Hsieh H.-H., Huang S.-T., Lee H.-C. & Mahajan S. 2018 ADF: an anomaly detection framework for large-
456 scale PM2.5 sensing systems. *IEEE Internet of Things Journal*, **5**(2), 559-570, doi: 10.1109/JIOT.2017.2766085.
457 6. Droumaguet M., Warne A. & Wozniak T. 2017 Granger Causality and Regime Inference in Markov Switching VAR Models
458 with Bayesian Methods. *Journal of Applied Econometrics*, **32**(4), 802-818, doi: 10.1002/jae.2531.
459 7. Lei Z. D., Yang S. X., Xie S. C. 1988 Soil Water Dynamics (in Chinese). *Tsinghua University Press*, **314**, doi:
460 10.13243/j.cnki.slxb.1988.05.001.
461 8. Liu B. C., Liu W. & Peng S. W. 2005 Study of heat and moisture transfer in soil with a dry surface layer. *International Journal
462 of Heat and Mass Transfer*, **48**(21-22), 4579–4589, doi: 10.1016/j.ijheatmasstransfer.2005.06.004.
463 9. Lu N. & Dong Y. 2015 Closed-Form Equation for Thermal Conductivity of Unsaturated Soils at Room Temperature. *Journal
464 of Geotechnical and Geoenvironmental Engineering,* **141**(6), doi:10.1061/(ASCE)GT.1943-5606.0001295.
465 10. Mao X. & Shang P. 2018 Extended AIC model based on high order moments and its application in the financial market.
466 *Physical A-statistical Mechanics and Its Applications*, **501**, 264-275, doi: 10.1016/j.physa.2018.02.159.
467 11. Meng C.H. & Xia J. 2005 Research on water and heat transfer in soil-plant-atmosphere system. *Journal of Hydrodynamics*,
468 **20**(3), 307-312, doi: 10.16076/j.cnki.cjhd.2005.03.005.
469 12. Nassar I. N. & Horton, R. 1989 Water transport in unsaturated nonisothermal salty soil: II. Theoretical Development. *Soil
470 Science Society of America Journal*, **53**(5), 1330, doi: 10.2136/sssaj1989.03615995005300050005x.
471 13. Pan Y., Shi G., Liu Y. & Lee F. 2018 Effect of spatial variability on performance of cement-treated soil slab during deep
472 excavation. *Construction and Building Materials*, **188**, 505-519, doi: 10.1016/j.conbuildmat.2018.08.112.

473 14. Philip JR. & Vries DA. 1957 Moisture movement in porous material sunder temperature gradients. *Trans Am Geophy Unionk*,
474   **38**, 22-231, doi: 10.1029/tr038i002p00222.
475 15. Perez E. H., Paez M. I. & Figueroa A. 2013 Effect of moisture and temperature on dissipation of chlorpyrifos and diazinon in
476   andic soils, Cauca, Colombia. *Asian Journal of Chemistry*, **25**(16), 9208-9212, doi: 10.14233/ajchem.2013.15166.
477 16. Schoukens J., Godfrey K. & Schoukens M. 2018 Nonparametric data-driven modeling of linear systems estimating the
478   frequency response and impulse response function. *IEEE Control Systems Magazines,* **38**(4), 49-88, doi:
479   10.1109/MCS.2018.2830080.
480 17. Striegl & Loheide 2012 Feasibility of soil moisture estimation using passive distributed temperature sensing. *Ground Water,*
481   **50**(3), 340-347, doi:10.1111/j.1745-6584.2012.00928.x
482 18. Steele-Dunne S. C., Rutten M. M., Krzeminska D. M., Hausner M., Tyler S. W., Selker J., Bogaard T. A. & van de Giesen N. C.
483   2010 Feasibility of soil moisture estimation using passive distributed temperature sensing. *Water Resources*, **46**, W03534,
484   doi:10.1029/2009WR008272.
485 19. Stokes P. A. & Purdon P. L. 2018 Regarding interpretation of Granger causality analyses reply. *Proceedings of the National*
486   *Academy of Sciences of the United States of America*, **115**(29), E6678-E6679, doi: 10.1073/pnas.1807099115.
487 20. Sun Z., Chen Q., Han X., Bol R., Qu B. & Meng F. 2018 Allocation of photosynthesized carbon in an intensively farmed winter
488   wheat-soil system as revealed by (CO2)-C-14 pulse labelling. *Scientific Reports*, **8**, doi: 10.1038/s41598-018-21547-y.
489 21. Whelan A., Kechavarzi C., Coulon F. & Doerr S. H. 2015 Experimental characterization of the impact of temperature and
490   moisture on the breakdown of soil water repellency in sandy soils and composts. *Hydrological Processes*, **29**(8), 2065-2073,
491   doi: 10.1002/hyp.10305.
492 22. Yang Y., Shang S. & Guan H. 2012 Development of a soil-plant-atmosphere continuum model (HDS-SPAC) based on hybrid
493   dual-source approach and its verification in wheat field. *Science China Technological Sciences*, **55**(10), 2671–2685, doi:
494   10.1007/s11431-012-4974-7.

Figure 1

Soil temperature (a) and soil humidity (b) at different depths.

Soil temperature (a) and soil humidity (b) at the depth of 10 cm-100 cm obtained at the Shangqiu test site, China.

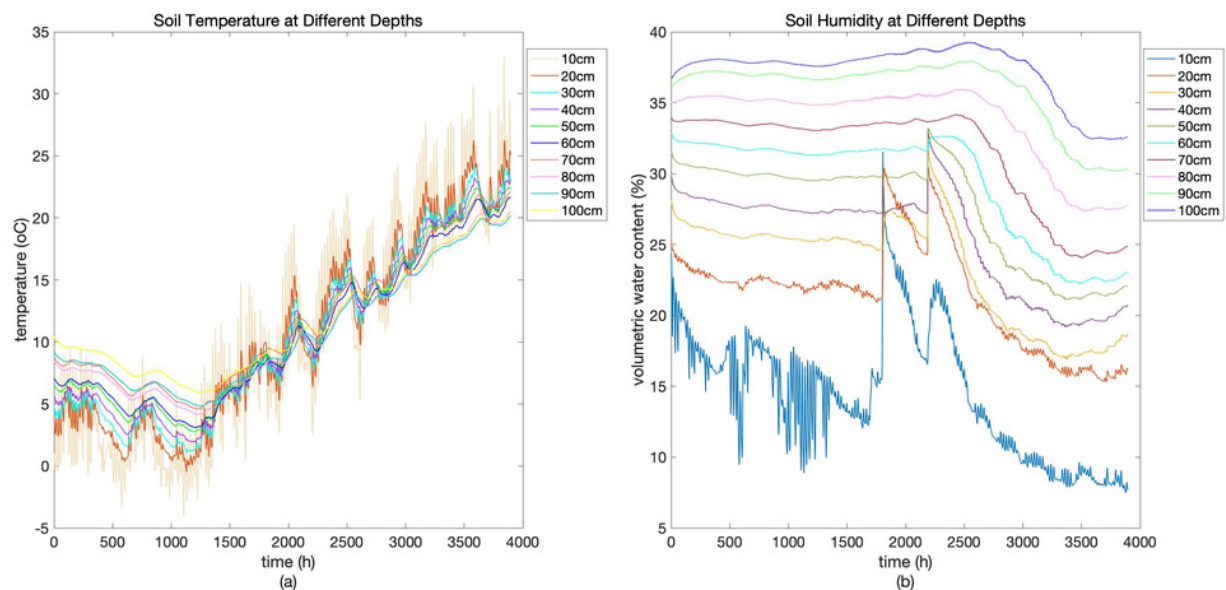



Figure 2

Deseasonalized soil temperature (a) and moisture (b) at the depth of 10 cm, 30 cm, and 90 cm.

Soil temperature (a) and moisture (b) after deseasonalization at the depth of 10 cm, 30 cm, and 90 cm.

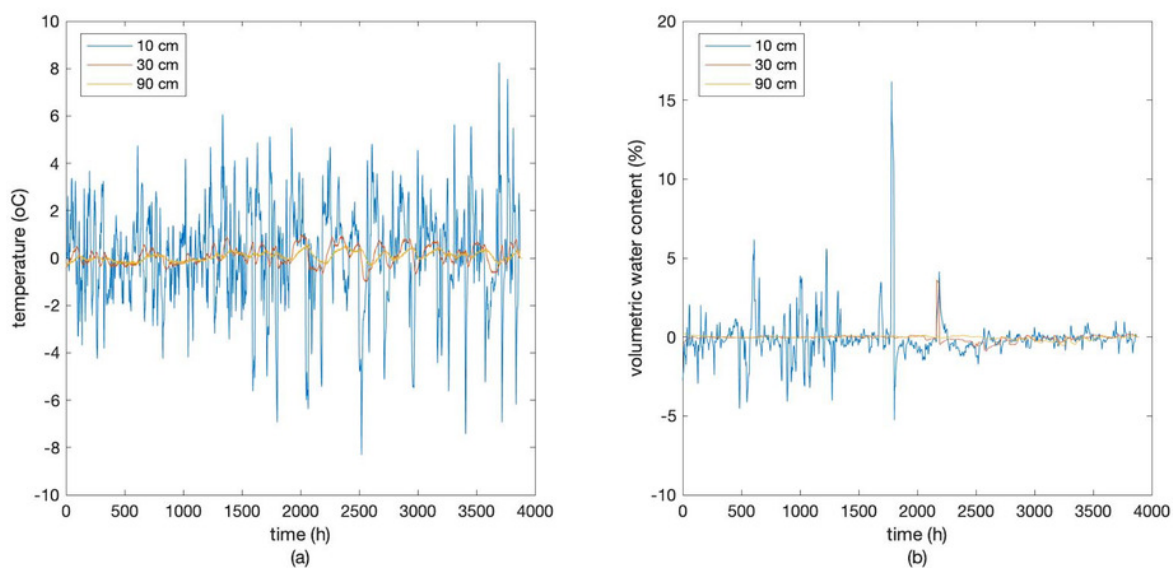



Figure 3

Inverse roots of AR characteristic polynomial at 10 cm (a), 30 cm (b) and 90 cm (c).

In this complex plane, the horizontal axis is the real axis, and the vertical axis is the imaginary axis. Inverse roots of AR characteristic polynomial at 10 cm (a), 30 cm (b) and 90 cm (c) show that all three models are stationary.

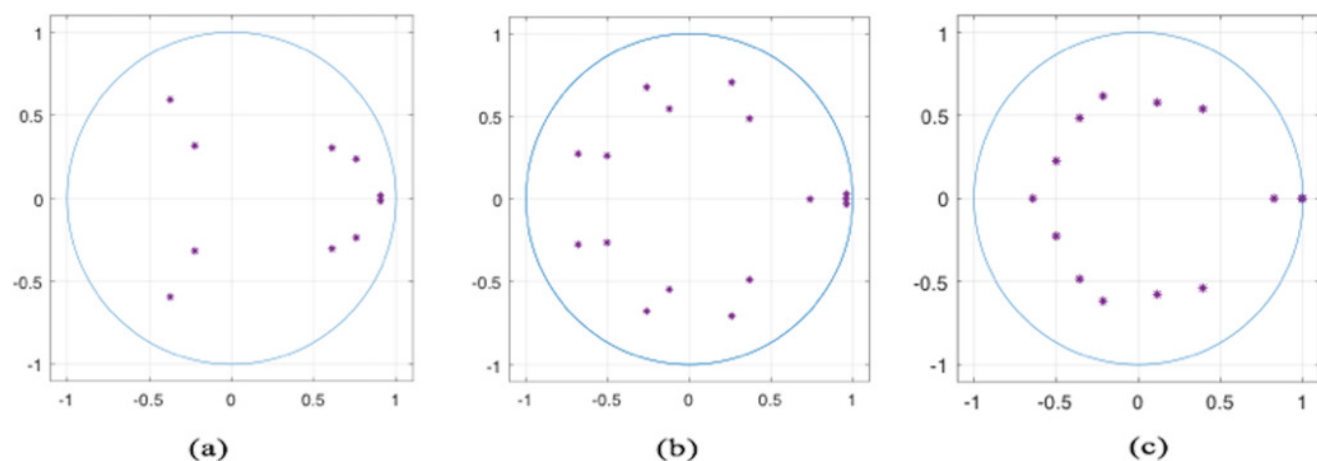
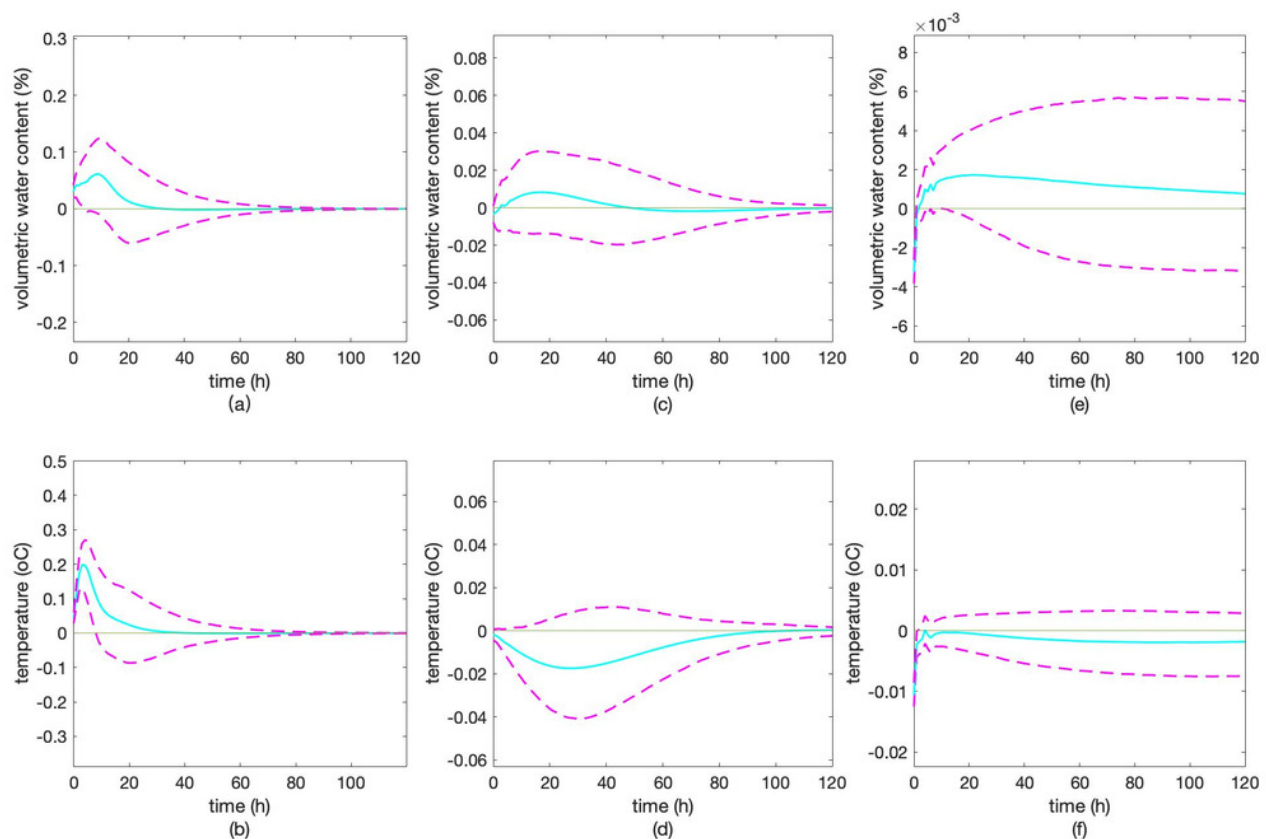



Figure 4

Impulse response results.

Impulse response of *MCSD10* to shock in *STSD10* (a), impulse response of *STSD10* to shock in *MCSD10* (b), impulse response of *MCSD30* to shock in *STSD30* (c), impulse response of *STSD30* to shock in *MCSD30* (d), impulse response of *MCSD90* to shock in *STSD90* (e), impulse response of *STSD90* to shock in *MCSD90* (f). The solid line of blue fluorescence represents the impulse response and the red dashed line represents the confidence interval of the impulse response with a confidence level of 0.90.

Unit root test results of the deseasonalizedsoil moisture and temperature time series.

[1]*T stat.*is the t-statistic for performing ADF test. [2]*Model*indicates the type of hypothesis selected and the lag period which is added to make the residual become white noise. [3]*P-Value*represents the corresponding p value. [4]*C.V.*indicates critical value.[5]* indicates a significant level of 10%, ** indicates a significant level of 5%, and *** indicates a significant level of 1%. [6]*Stationary* indicates that the time series is stable.

1   **Table 1.** Unit root test results of the deseasonalized soil moisture and temperature time series.

Variable	T stat.[1]	Model[2]	P-Value[3]	1%C.V.[4]	5%C.V.	10%C.V.	Conclusion
MCSD10	-11.273	AR(8)	0.001***[5]	-2.568	-1.942	-1.617	Stationary[6]
STSD10	-11.689	AR(5)	0.001***	-2.568	-1.942	-1.617	Stationary
MCSD30	-8.590	AR(3)	0.001***	-2.568	-1.942	-1.617	Stationary
STSD30	-7.423	AR(8)	0.001***	-2.568	-1.942	-1.617	Stationary
MCSD90	-3.927	AR(8)	0.001***	-2.568	-1.942	-1.617	Stationary
STSD90	-2.055	AR(8)	0.039**	-2.568	-1.942	-1.617	Stationary

2   [1] *T stat.* is the t-statistic for performing ADF test. [2] *Model* indicates the type of hypothesis selected and the lag period which is
3   added to make the residual become white noise. [3] *P-Value* represents the corresponding p value. [4] *C.V.* indicates critical value.[5] *
4   indicates a significant level of 10%, ** indicates a significant level of 5%, and *** indicates a significant level of 1%. [6] *Stationary*
5   indicates that the time series is stable.

Calculated In(L), AIC and BIC for vector auto-regression models of different lags at the depth of 10 cm, 30 cm, and 90 cm.

The first row of the table denotes the possible lag orders of the model. The table is used to choose the optimal lag order for each model by comparing the value In(L), AIC and BIC.

Table 2. Calculated In(L), AIC and BIC for vector auto-regression models of different lags at the depth of 10cm, 30 cm, and 90 cm.

Lag	10 cm			30 cm			90 cm		
	$\ln(\hat{L})$	AIC	BIC	$\ln(\hat{L})$	AIC	BIC	$\ln(\hat{L})$	AIC	BIC
1	-2298.5	4609.1	4646.5	11500.7	-22989.4	-22951.8	19089.2	-38166.3	-38128.8
2	242.7	-465.5	-402.9	11758.8	-23497.7	-23435.1	19625.2	-39230.3	-39167.7
3	277.9	-527.8	-440.1	11762.2	-23496.5	-23408.8	19814.5	-39600.9	-39513.3
4	327.8	-619.6	-507.0	11826.4	-23616.9	-23504.2	19927.4	-39818.8	-39706.1
5	478.2	-912.4	-774.7	11942.4	-23840.7	-23703.0	20013.4	-39982.9	-39845.1
6	493.0	-934.1	-771.3	12050.5	-24048.9	-23886.2	20035.8	-40019.7	-39856.9
7	493.2	-926.4	-738.6	12102.0	-24144.1	-23956.3	20054.4	-40048.7	-39860.9
8	508.4	-948.8	-736.0	12143.2	-24218.3	-24005.5	20068.5	-40069.0	-39856.2

Impulse response results of *MCSD10, STSD10, MCSD30, STSD30, MCSD90* and *STSD90*.

The first column of the table denotes retrospective periods with its unit being hour. And the next six columns display the results of response of *MCSD* to one unit of positive shock in *STSD* (%) and response of *MCSD* to one unit of positive shock in *STSD* (%).

Table 3. Impulse response results of *MCSD10, STSD10, MCSD30, STSD30, MCSD90* and *STSD90*.

RP[1] (h)	Response of *MCSD10* to one unit of positive shock in *STSD10*(%)	Response of *STSD10* to one unit of positive shock in *MCSD10*(°C)	Response of *MCSD30* to one unit of positive shock in STSD30(%)	Response of *STSD30* to one unit of positive shock in *MCSD30*(°C)	Response of *MCSD90* to one unit of positive shock in *STSD90*(%)	Response of *STSD90* to one unit of positive shock in *MCSD90*(°C)
0	0.0312	0.0453	-0.0077	-0.0045	-0.0033	-0.0107
1	0.0426	0.1112	-0.0046	-0.0029	-0.0003	-0.0024
2	0.0410	0.1695	-0.0051	-0.0051	0.0002	-0.0018
3	0.0434	0.1942	-0.0033	-0.0044	0.0004	-0.0015
4	0.0449	0.2004	-0.0033	-0.0048	0.0010	-0.0000
5	0.0470	0.1876	-0.0019	-0.0064	0.0009	-0.0008
6	0.0530	0.1581	-0.0011	-0.0063	0.0013	-0.0014
7	0.0584	0.1268	0.0011	-0.0070	0.0009	-0.0009
8	0.0620	0.0973	0.0019	-0.0073	0.0013	-0.0006
9	0.0638	0.0718	0.0025	-0.0081	0.0014	-0.0006
10	0.0626	0.0543	0.0036	-0.0085	0.0015	-0.0005
11	0.0586	0.0433	0.0045	-0.0091	0.0015	-0.0004
12	0.0531	0.0374	0.0054	-0.0096	0.0016	-0.0005

2   [1]RP represents retrospective periods, its unit is hour.

Granger causality test determining whether *STSD10* Granger causes *MCSD10*.

The table displays the results of Granger causality test of the depth of 10 cm. By comparing the value of F-Statistics and critical values at the significance level of 0.01, 0.05 and 0.10, we can deduce the Granger causal relationship between soil moisture and temperature.

1 **Table 4.** Granger causality test determining whether *STSD10* Granger causes *MCSD10*.

Significance	F-Statistic	Critical Value	Conclusion
0.01	0.5694	6.6415	NO
0.05	0.5694	3.8439	NO
0.10	0.5694	2.7068	NO

Granger causality test determining whether *MCSD10* Granger causes *STSD10*.

The table displays the results of Granger causality test of the depth of 10 cm. By comparing the value of F-Statistics and critical values at the significance level of 0.01, 0.05 and 0.10, we can deduce the Granger causal relationship between soil moisture and temperature.

1  **Table 5.** Granger causality test determining whether *MCSD10* Granger causes *STSD10*.

Significance	F-Statistic	Critical Value	Conclusion
0.01	18.3573	3.3240	YES
0.05	18.3573	2.3742	YES
0.10	18.3573	1.9463	YES

Granger causality test determining whether *STSD30* Granger causes *MCSD30*.

The table displays the results of Granger causality test of the depth of 30 cm. By comparing the value of F-Statistics and critical values at the significance level of 0.01, 0.05 and 0.10, we can deduce the Granger causal relationship between soil moisture and temperature.

1   **Table 6.** Granger causality test determining whether *STSD30* Granger causes *MCSD30*.

Significance	F-Statistic	Critical Value	Conclusion
0.01	12.7695	4.6107	YES
0.05	12.7695	2.9981	YES
0.10	12.7695	2.3040	YES

Granger causality test determining whether *MCSD30* Granger causes *STSD30*.

The table displays the results of Granger causality test of the depth of 30 cm. By comparing the value of F-Statistics and critical values at the significance level of 0.01, 0.05 and 0.10, we can deduce the Granger causal relationship between soil moisture and temperature.

1   **Table 7.** Granger causality test determining whether *MCSD30* Granger causes *STSD30*.

Significance	F-Statistic	Critical Value	Conclusion
0.01	4.2014	6.6415	NO
0.05	4.2014	3.8439	YES
0.10	4.2014	2.7068	YES

Granger causality test determining whether *STSD90* Granger causes *MCSD90*.

The table displays the results of Granger causality test of the depth of 90 cm. By comparing the value of F-Statistics and critical values at the significance level of 0.01, 0.05 and 0.10, we can deduce the Granger causal relationship between soil moisture and temperature.

1 **Table 8.** Granger causality test determining whether *STSD90* Granger causes *MCSD90*.

Significance	F-Statistic	Critical Value	Conclusion
0.01	40.9051	2.6439	YES
0.05	40.9051	2.0120	YES
0.10	40.9051	1.7183	YES

Granger causality test determining whether *MCSD90* Granger causes *STSD90*.

The table displays the results of Granger causality test of the depth of 90 cm. By comparing the value of F-Statistics and critical values at the significance level of 0.01, 0.05 and 0.10, we can deduce the Granger causal relationship between soil moisture and temperature.

1   **Table 9.** Granger causality test determining whether *MCSD90* Granger causes *STSD90*.

Significance	F-Statistic	Critical Value	Conclusion
0.01	10.5937	3.0220	YES
0.05	10.5937	2.2164	YES
0.10	10.5937	1.8488	YES