Manuscript Title

Xiaohang Feng[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[1,2*]

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

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

Xiaohang Feng[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, Email: 1600018809@pku.edu.cn; Tel.: +86-18811776315

Abstract: 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 of 10 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.

Introduction

The hydrothermal conditions of a wheat-soil system are essential to the dynamic balance of heat, moisture and organic matter within the entire system and the thriving of wheat (Yang et al. 2012; Sun et al. 2018). Because of complicated biological, physical and chemical processes like soil respiration, soil evaporation, plant transpiration and so on, soil temperature and moisture have a close dynamic relationship with each other. Therefore, in order to better control and predict the hydrothermal conditions of a wheat-soil system, models simulating soil moisture-heat coupling need to be established (Perez et al. 2013; Whelan et al. 2015).

Most of the models employed to simulate moisture-heat coupling of soil are mechanistic models which call for complex parameters on soil properties and neglect time stochasticity. The following are some examples. Philip and de Vries (1957) took soil evaporation under different temperature into consideration and brought up the theory of moisture-gas-heat coupling transport under mass and energy balance. However, they did not consider the time lag effect and temporal heterogeneity. Based on the Philip model, Nassar and Horton (1989) used the water, heat, and salt transport equations which are based on Darcy's, Fourier's and Fick's laws to establish a model for the coupled transport of water, heat, and solute. But the model required a bunch of complicated parameters. Based on the theoretical integrity, Liu et al. (2005) established a model for describing the migration of heat, moisture, and gas in arid surface porous soil composed of a wet unsaturated layer and a dry but saturated layer. Bittelli et al. (2008) established a fully coupled numerical model to solve the governing equations for liquid water, water vapor and heat transfer in bare soil. Whelan et al. (2015) studied the impact of temperature and moisture on soil water repellency by designing and conducting experiments and factorial ANOVAs. However, the study fails to further explore characteristics like time lag and intensity of temperature and moisture's influence on each other. Lu and Dong (2015) set up a closed-formed equation for the thermal conductivity of unsaturated soils and shows that the soil water retention curve (SWRC) can be used to predict the thermal conductivity of sands. Striegl and Loheide (2012) developed a distributed soil moisture sensing system that addressed the difficulty of characterizing both spatial and temporal soil moisture dynamics at site scales. However, under wet conditions insensitivity of the instrument response curve adversely affected accuracy. Steele-Dunne et al. (2010) used distributed

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

76   **Methods**

77   *Models*

78   Vector Autoregression Model

79   Vector autoregression model treats all variables as endogenous variables, accounting for Sims'
80   critique that the exogeneity[1] assumptions of some of the variables in simultaneous equations models
81   are ad hoc and often not backed by fully developed theories. Impulse response analysis and Granger
82   causality tests are tools which have been proposed for disentangling the relations between the
83   variables in a vector autoregression model (Droumaguet et al. 2017; Bouri et al. 2018). The process of
84   establishing a vector autoregression system is as follows.
85   First, we remove the seasonality of the time series x_t, t=1, 2, …, n. Seasonality in a time series is
86   a regular pattern of changes that repeats over S time periods, where S defines the number of time
87   periods until the pattern repeats again. In this case, S = 24 (hours per day) is the span of the periodic
88   seasonal behavior. Seasonal differencing is defined as a difference between a value and a value with
89   lag that is a multiple of S. With S = 24, a seasonal difference is $y_t=(1-B^{24})x_t = x_t - x_{t-24}$, t = 25,
90   26, …, n, where B is the lag operator.
91   Next, we check the stationarity of the time series after seasonal adjustment, because average
92   values of non-stationary time series cannot be used if the time period is not set since they are
93   influenced by changes in time. A stochastic process is called stationary when the average and variance
94   values are constant in the corresponding period and covariance values between any two time points
95   depend not on the specific time point but on the lag between these two time points. The three
96   conditions stated above for a time series $\{y_t\}$ to be stationary can be expressed in the following way:

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

97   where μ, σ^2, and γ are average, variance, and auto-covariance, respectively.
98   If the time series has a unit root, then it is not stationary. Under such circumstances, we restore
99   the stationarity of the time series by making a difference of it. We use the Augmented Dickey-Fuller
100 (ADF) test method to check the stationarity of the time series by testing whether the time series has a
101 unit root. There are three model variants corresponding to three couples of hypothesis in the ADF test.
102   The first is an autoregressive model variant (referred to as *AR*), which specifies a test of the null
103 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.

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

against the alternative model:

$$H_1: y_t = \emptyset y_{t-1} + \beta_1 \Delta y_{t-1} + \beta_2 \Delta y_{t-2} + \cdots + \beta_p \Delta y_{t-p} + \varepsilon_t, \tag{3}$$

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

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

$$H_0: y_t = y_{t-1} + \beta_1 \Delta y_{t-1} + \beta_2 \Delta y_{t-2} + \cdots + \beta_p \Delta y_{t-p} + \varepsilon_t, \tag{4}$$

against the alternative model:

$$y_t = c + \emptyset y_{t-1} + \beta_1 \Delta y_{t-1} + \beta_2 \Delta y_{t-2} + \cdots + \beta_p \Delta y_{t-p} + \varepsilon_t, \tag{5}$$

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.

The third is a trend-stationary model variant (referred to as *TS*), which specifies a test of the null model:

$$H_0: y_t = c + y_{t-1} + \beta_1 \Delta y_{t-1} + \beta_2 \Delta y_{t-2} + \cdots + \beta_p \Delta y_{t-p} + \varepsilon_t, \tag{6}$$

against the alternative model:

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

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 coefficients, and ε_t is the error term (Chen et al. 2018).

Then, we have to determine the optimal lag order for the vector autoregression model. The proper selection of lag is important because long lag structures reduce the error term's correlation yet they may lack efficiency. We use the estimators AIC (Akaike information criterion) and BIC (Bayesian information criterion), both of which are founded on information theory and estimate relative goodness of fit of given models, as standards of choosing the optimal lag of the vector autoregression model (Mao et al. 2018).

AIC value of a model is calculated as follows:

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

where k is the number of estimated parameters in the model and $\hat{L}$ is the maximum value of the likelihood function for the model.

BIC value of a model is calculated as follows:

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

where n is the number of data points in x or the number of observations.

When fitting the models, it is possible to increase the simulation accuracy by adding parameters, but doing so may result in overfitting. Both BIC and AIC attempt to solve this problem by introducing a penalty term for the number of parameters in the model. And the lag order corresponding to the lowest AIC value or BIC value is taken as the optimal lag order. ~~we~~ We use maximum likelihood estimation[2] to estimate the model parameters.

The stability test of the established model is required because if the model is unstable, some results will not be valid (such as the standard error of the impulse response function). In this paper, we use the AR root test and if all the roots of the vector autoregression model estimated based on empirical data have a reciprocal of less than 1 (that is, if they are within the unit circle), they are stable.

The way to check whether a vector autoregression model is stationary is as follow. For vector autoregression model

---

[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: $\hat{\theta} \in \{\arg\max_{\theta \in \Theta} \mathcal{L}(\theta; x)\}$, where $\mathcal{L}(\theta; x)$ denotes the likelihood function, because intuitively this selects parameter values that make the data most probable.

138 $$(y_t - \mu) = \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + \cdots + \phi_p(y_{t-p} - \mu) + \varepsilon_t$$

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

140 $$F = \begin{bmatrix} \phi_1 & \phi_2 & \phi_3 & \cdots & \phi_{p-1} & \phi_p \\ I_n & 0 & 0 & \cdots & 0 & 0 \\ 0 & I_n & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & I_n & 0 \end{bmatrix}$$

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

143 Impulse Response Function

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

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

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

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

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

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \bar{y}_t \\ \bar{z}_t \end{bmatrix} + \sum_{i=0}^{\infty}\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i\begin{bmatrix} e_{1t-i} \\ e_{2t-i} \end{bmatrix}, \tag{12}$$

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

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

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

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

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

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

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

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \bar{y}_t \\ \bar{z}_t \end{bmatrix} + \sum_{i=0}^{\infty}\begin{bmatrix} \phi_{11}(i) & \phi_{12}(i) \\ \phi_{21}(i) & \phi_{22}(i) \end{bmatrix}\begin{bmatrix} \varepsilon_{yt-i} \\ \varepsilon_{zt-i} \end{bmatrix}, \tag{16}$$

155 Moving average is a useful tool for studying the mutual influence between the two series $\{y_t\}$ and
156 $\{z_t\}$, for the impulse response function of $\{y_t\}$ and $\{z_t\}$ to one unit shock of ε_{yt} and ε_{zt} can be
157 built using the coefficients of ϕ_i. Here $\phi_{11}(i)$, $\phi_{12}(i)$, $\phi_{21}(i)$ and $\phi_{22}(i)$ are called impulse
158 response functions. And the graphs of these functions show the impulse response of the time series to
159 one unit of positive shock.
160   To be more specific, coefficient $\phi_{12}(0)$ indicates the current impact of one unit positive change
161 in ε_{zt} have on y_t. Similarly, after one period of calibration, $\phi_{11}(1)$ and $\phi_{12}(1)$ also denote the
162 impact of one unit positive change in ε_{yt} and ε_{zt} have on y_{t+1}. The cumulative effect of ε_{yt} and
163 ε_{zt}'s unit impulse response can be obtained by the appropriate add up of the coefficients of the impulse
164 response functions. For instance, after n periods the impact of ε_{zt} on y_{t+n} is $\phi_{12}(n)$. Therefore,
165 after n periods the accumulative impact of ε_{zt} on y_t is $\sum_{i=0}^{n}\phi_{12}(i)$. If n tends to be positive
166 infinite, the accumulative effect can be obtained. Since we have assumed that y_t and z_t are
167 stationary, when j and k approximate infinite, the value of $\phi_{jk}(i)$ is 0. Since all parameters of the

168 vector autoregression system can be calculated, it is totally possible to track every value of ε_{yt} and
169 ε_{zt}'s impact on $\{y_t\}$ and $\{z_t\}$ can be calculated. The time when $\phi_{jk}(i)$ approximates 0 is the time
170 lag of impulse response, and the gap between the positive largest value and negative largest value of
171 $\phi_{jk}(i)$ can be used to describe the intensity of the impulse response.

172 Granger Causality Test

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

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

179 And we propose the null hypothesis:

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

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

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

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

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

183 We build the F statistic using the expression below:

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

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

188 Stochastic Model System for Soil Moisture-heat Coupling

189   First, we build three vector autoregression models at the depth of 10 cm, 30 cm, and 90 cm. We
190 choose these three depths because, during the greening period which is around 15th March, the
191 diameter of wheat root reaches its maximum approximately at the depth of 10 cm, 30 cm, and 90 cm,
192 respectively, thus resulting in more water absorption at these three depths. In addition, usually, the
193 soil is divided into topsoil, subsoil, and substratum, the representative depths of which approximate 10
194 cm, 30 cm, and 90 cm. In addition, we select hourly data on soil moisture and temperature from
195 December to May because this period covers the main growing stages of winter wheat. This time period
196 can be adjusted according to the growing stages of different wheat species (Awad et al. 2018).
197   Second, we build six impulse response functions for these three vector autoregression systems.
198 Based on the results of impulse response function analysis, we are able to obtain the time lag and
199 intensity of impulse response. For example, if one unit of positive shock is given to soil temperature at
200 the depth of 10 cm, we calculate the time length during which obvious deviation from zero can be seen
201 in soil moisture's response to 1°C of positive variation. This provides us with ~~the~~ information on how
202 long soil moisture will be influenced if there is a sudden change in temperature. In addition, the gap
203 between the highest point and lowest point of the response is the intensity of impulse response. The
204 larger the intensity, the stronger the influence of soil temperature has on moisture.
205   Third, we conduct six Granger causality tests between two variables of soil moisture and
206 temperature at three depths. According to the result of the Granger causality test, we are able to decide
207 a more proper way of forecasting soil moisture and temperature by taking the Granger causal
208 relationship of moisture and temperature into consideration. If, for example, at the depth of 10 cm the
209 past value of soil temperature helps predict soil moisture, it means that the past value of soil
210 temperature ~~has an influence on~~influences soil moisture. Therefore, when establishing empirical

models forecasting soil moisture, we should choose the ARMAX model which is an ARMA model with an exogenous variable in an attempt to take into consideration of temperature's influence on moisture. On the other hand, if the past value of soil temperature does not help predict soil moisture, it means that the past value of soil temperature does not have a significant influence on soil moisture. If soil temperature and moisture are Granger reasons for each other, it suggests that the present value of two variables are related to each other's past value. Thus, a vector autoregression model treating all variables as endogenous variables will be the optimal choice.

Field experiment on soil moisture and temperature

The experiment on the wheat of Bainong Dwarf Anti-floating No.58[3] was carried out at Shangqiu Experimental Station of Farmland Irrigation Institute of Chinese Academy of Agricultural Sciences with a latitude of 34° 35.222'N, longitude of 115° 34.515'E, and elevation of 55.6 m. The average annual temperature there is 13.9°C, and the frost-free period is 180 d to 230 d. In addition, the average annual precipitation is 708 mm with the precipitation from July to September accounting for 65%~75% of the total annual precipitation, and the average annual evaporation being 1735 mm. The soil of the test site is a loam, which refers to soil consists of clay (30%~40%), ~~powder~~ silt (30%~40%), and sand (30%~40%).

The Irrigation method adopted is drip irrigation at the depth of 30 cm under soil surface. The sensor we used for ~~acquiring data~~soil water monitoring is called "Soil-Water", which was produced in Australia. The basic components of this soil monitoring system are data collectors, solar modules, mounting components, and soil sensors. The accuracy for temperature is <0.1°C, and but the accuracy of soil moisture sensor was not provided by the manufacturer. We installed the instrument probes of the ~~tube~~moisture meter at the depth of 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100 cm, and measured soil moisture and soil temperature at these depths simultaneously. The instrument's temperature measurement range is from -20°C to 60°C.

Among these ten sets of data, we select hourly data on soil moisture and soil temperature from December 19th, 2017 to May 30th, 2018 at the depth of 10 cm, 30 cm and 90 cm, which is in accordance with the model designed above. In addition, the selection of the measurement interval takes into account both the memory capacity of the instrument and the minimum scale of moisture transport, since one-hour interval is sufficient to show the footprint of water movement.

Results

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

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

We use hourly data from December 19th, 2017 to May 30th, 2018 on soil moisture content at the soil depths of 10 cm, 30 cm and 90 cm (hereinafter referred to as *MC10, MC30* and *MC90*, respectively), soil temperature at the depth of 10 cm, 30 cm and 90 cm (hereinafter referred to as *ST10, ST30* and *ST90*, respectively).

First, we remove the seasonality of the time series of *MC10, ST10, MC30, ST30, MC90* and *ST90*, since it is known from the above figures that the six hourly time series have obvious diurnal pattern due to the periodical variations of meteorological factors, evapotranspiration, and root water uptake

---

[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.

within a day, especially in shallow layers. After seasonal difference, the period of the data is from December 20th, 2017 to May 30th, 2018. The corresponding time series of the six detrended variables are referred to as *MCSD10, STSD10, MCSD30, STSD30, MCSD90,* and *STSD90*, respectively (Figure 2).

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

[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 is added to make the residual become white noise. [3] *P-Value* represents the corresponding p-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.

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

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

Vector Autoregression Model Establishment

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

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

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

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

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

$$\begin{bmatrix} MCSD30_{\mathrm{t}} \\ STSD30_{t} \end{bmatrix} = \begin{bmatrix} -0.002 \\ 0.001 \end{bmatrix} + \begin{bmatrix} 1.182 & 0.117 \\ -0.006 & 0.559 \end{bmatrix} \begin{bmatrix} MCSD30_{t-1} \\ STSD30_{t-1} \end{bmatrix} +$$
$$\begin{bmatrix} -0.173 & -0.090 \\ -0.022 & 0.344 \end{bmatrix} \begin{bmatrix} MCSD30_{t-2} \\ STSD30_{t-2} \end{bmatrix} + \begin{bmatrix} -0.016 & 0.020 \\ 0.042 & 0.228 \end{bmatrix} \begin{bmatrix} MCSD30_{t-3} \\ STSD30_{t-3} \end{bmatrix} +$$
$$\begin{bmatrix} -0.002 & -0.054 \\ -0.003 & 0.149 \end{bmatrix} \begin{bmatrix} MCSD30_{t-4} \\ STSD30_{t-4} \end{bmatrix} + \begin{bmatrix} 0.000 & 0.013 \\ -0.022 & 0.012 \end{bmatrix} \begin{bmatrix} MCSD30_{t-5} \\ STSD30_{t-5} \end{bmatrix} + \qquad (23)$$
$$\begin{bmatrix} -0.006 & -0.014 \\ 0.019 & -0.074 \end{bmatrix} \begin{bmatrix} MCSD30_{t-6} \\ STSD30_{t-6} \end{bmatrix} + \begin{bmatrix} 0.012 & 0.040 \\ -0.010 & -0.082 \end{bmatrix} \begin{bmatrix} MCSD30_{t-7} \\ STSD30_{t-7} \end{bmatrix} +$$
$$\begin{bmatrix} -0.027 & -0.029 \\ -0.002 & -0.148 \end{bmatrix} \begin{bmatrix} MCSD30_{t-8} \\ STSD30_{t-8} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

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

$$\begin{bmatrix} MCSD90_{\mathrm{t}} \\ STSD90_{t} \end{bmatrix} = \begin{bmatrix} 0.690 & 0.058 \\ 0.133 & 0.345 \end{bmatrix} \begin{bmatrix} MCSD90_{t-1} \\ STSD90_{t-1} \end{bmatrix} + \begin{bmatrix} 0.336 & 0.025 \\ 0.026 & 0.200 \end{bmatrix} \begin{bmatrix} MCSD90_{t-2} \\ STSD90_{t-2} \end{bmatrix} +$$
$$\begin{bmatrix} 0.130 & -0.001 \\ 0.045 & 0.191 \end{bmatrix} \begin{bmatrix} MCSD90_{t-3} \\ STSD90_{t-3} \end{bmatrix} + \begin{bmatrix} -0.024 & -0.010 \\ 0.079 & 0.103 \end{bmatrix} \begin{bmatrix} MCSD90_{t-4} \\ STSD90_{t-4} \end{bmatrix} +$$
$$\begin{bmatrix} -0.061 & -0.031 \\ -0.166 & 0.039 \end{bmatrix} \begin{bmatrix} MCSD90_{t-5} \\ STSD90_{t-5} \end{bmatrix} + \begin{bmatrix} -0.004 & -0.008 \\ -0.132 & 0.056 \end{bmatrix} \begin{bmatrix} MCSD90_{t-6} \\ STSD90_{t-6} \end{bmatrix} + \qquad (24)$$
$$\begin{bmatrix} -0.071 & -0.032 \\ 0.011 & 0.058 \end{bmatrix} \begin{bmatrix} MCSD90_{t-7} \\ STSD90_{t-7} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

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

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

297   *Impulse Response Function Analysis*

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

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

300   [1]RP represents retrospective periods, its unit is an hour.

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

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

311   When *STSD10* is given one unit of positive shock, *MCSD10* reacts to 0.0312% in the current period,
312   and then monotonically increases until it reaches a positive maximum of 0.0638% in the ninth period.
313   And then it gradually decreases and falls near zero in the long run indicating that the vector

314 autoregression system is stable. It can be seen that the impact of *STSD10* on *MCSD10* has a lagging
315 effect since its impulse response is only close to zero after around 25 periods. Moreover, the intensity
316 of the influence which can be quantified as 0.0638% is relatively high (Figure 4. (a)).
317   When *MCSD10* is given one unit of positive shock, *STSD10* reacts to 0.0453°C in the current period,
318 and then monotonically increases until it reaches a positive maximum of 0.2004°C in the fourth period.
319 And then it gradually decreases, and fall near zero in the long run indicating that the vector
320 autoregression system is stable. It can be seen that the impact of *MCSD10* on *STSD10* has a lagging
321 effect since its impulse response never falls below zero and is only close to zero after 50 periods.
322 Moreover, the intensity of the influence which can be quantified as 0.2004°C is quite high (Figure 4.
323 (b)).
324   When *STSD30* is given one unit of positive shock, *MCSD30* reacts to -0.0077% in the current period,
325 and then monotonically increases until it reaches a positive maximum of 0.0086% in the 22nd, 23rd, 24th
326 period. And then it gradually decreases and falls near zero in the long run indicating that the vector
327 autoregression system is stable. Besides, the impact of *STSD30* on *MCSD30* has a lagging effect based
328 on the fact that it is only close to zero after around 50 periods. Moreover, the intensity of the influence
329 which can be quantified as 0.0163% is lower than the intensity of *MCSD10*'s response to *STSD10* (Figure
330 4. (c)).
331   When *MCSD30* is given one unit of positive shock, *STSD30* reacts to -0.0045°C in the current period,
332 and then monotonically decreases until it reaches a negative maximum of -0.0163°C at the 36th, 37th,
333 38th, 39th period. And then it gradually increases, and fall near zero in the long run indicating that the
334 vector autoregression system is stable. It can be seen that the impact of *MCSD30* on *STSD30* has a
335 lagging effect since its impulse response is only close to zero after the 120th period. Moreover, the
336 intensity of the influence which can be quantified as 0.0163°C is lower than the intensity of *STSD10*'s
337 response to *MCSD10* (Figure 4. (d)).
338   When *STSD90* is given one unit of positive shock, *MCSD90* reacts to -0.0033% in the current period,
339 and then monotonically increases until it reaches a positive maximum of 0.0017% from the 17th to 30th
340 period. And then it gradually decreases and falls near zero in the long run indicating that the vector
341 autoregression system is stable. It can be seen that the impact of *STSD90* on *MCSD90* has a lagging
342 effect based on the fact that it is only close to zero after 120 periods. Moreover, the intensity of the
343 influence which can be quantified as 0.0050% is lower than the intensity of *MCSD30*'s response to
344 *STSD30* (Figure 4. (e)).
345   When *MCSD90* is given one unit of positive shock, *STSD90* reacts to -0.0107°C in the current period,
346 and then monotonically increases and reaches a relatively stable value of -0.0035°C at around the 120th
347 period. It can be seen that the impact of *MCSD90* on *STSD90* has a lagging effect. Moreover, the
348 intensity of the influence which can be quantified as 0.0035°C is lower than the intensity of *STSD30*'s
349 response to *MCSD30* (Figure 4. (f)).

350 *Granger Causality Test*

351   By carrying out Granger causality test, we figured out whether the past value of *STSD10* (*STSD30,*
352 *STSD90*) helps predict *MCSD10* (*MCSD30, MCSD90*) and whether the past value of *MCSD10* (*MCSD30,*
353 *MCSD90*) helps predict *STSD10* (*STSD30, STSD90*).
354   From the results in Table 4 and Table 5, it is known that the past value of *MCSD10* helps predict
355 *STSD10*, but the past value of *STSD10* does not help predict *MCSD10*. This means that the past value of
356 soil moisture content is helpful in terms of predicting the present value of soil temperature at the depth
357 of 10 cm, while soil temperature has a small influence in terms of predicting soil moisture content at
358 the depth of 10 cm. Besides, the results also indicate soil moisture content varies ahead of soil
359 temperature.
360   From Table 6 and Table 7, we can see that at the significance level of 0.05 and 0.10, the past value
361 of *MCSD30* helps predict *STSD30* and the past value of *STSD30* helps predict *MCSD30*. This suggests
362 that the past value of soil moisture content is helpful in terms of predicting the present value of soil
363 temperature at the depth of 30 cm, and the past value of soil temperature is helpful in terms of
364 predicting the present value of soil moisture content at the depth of 30 cm.
365   From Table 8 and Table 9, we can see that at the significance level of 0.01, 0.05 and 0.10, the past
366 value of *MCSD90* helps predict *STSD90* and the past value of *STSD90* helps predict *MCSD90*. This
367 suggests that the past value of soil moisture content is helpful in terms of predicting the present value

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

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

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

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

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

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

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

376   **Discussions**

377   *Specific results*

378   Firstly, the time lag of soil temperature's response to shock in soil moisture is about 25 h at 10 cm,
379   50 h at 30 cm, and 120 h at 90 cm, while the time lag of soil moisture's response to shock in soil
380   temperature is for about 50 h at 10 cm and longer than 120 h at 30 cm and 90 cm.
381   Secondly, the intensity of soil temperature's impulse response to shock in soil moisture decreases
382   from 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
383   depth of 90 cm. Similarly, the intensity of soil moisture's impulse response to soil temperature also
384   decreases from 0.0638% at the depth of 10 cm, to 0.0163% at the depth of 30 cm, and finally to 0.0050%
385   at the depth of 90 cm.
386   Thirdly, soil moisture is the Granger reason of soil temperature at the depth of 10 cm, 30 cm, and
387   90 cm. This means that the past value of soil moisture is helpful in terms of predicting the present value
388   of soil temperature, and probably is the actual logical reason for changes in soil temperature. However,
389   the causal relationship does not necessarily mean that there exists a direct connection between soil
390   temperature and moisture, since a series of physical and biological processes may complicate the
391   relations between soil temperature and moisture. While soil temperature does not help predict soil
392   moisture at the depth of 10 cm, it helps predict soil moisture at the depth of 30 cm and 90 cm. Therefore,
393   according to the basic principles of Granger causality test, when predicting the dynamic variation of soil
394   temperature and moisture at the depth of 30 cm and 90 cm, the other variable's past values should be
395   taken into consideration, which means that a vector autoregression model may be a better choice. In
396   comparison, at the depth of 10 cm, the past value of soil temperature is not helpful for predicting the
397   present value of soil moisture. However, it may be more accurate to incorporate soil moisture as an
398   exogenous variable when predicting soil temperature at 10 cm.

399   *Verification of the model*

400   To begin with, the model we proposed can be supported and explained by established mechanistic
401   models. In 1966, Philip proposed the concept of a complete Soil-Plant-Atmosphere Continuum (SPAC),
402   laying the theoretical foundation for modern farmland water research. Based on this, Meng and Xia
403   (2005) established a dynamic coupling model by calibrating parameters that describes the
404   hydrothermal conditions during crop growth and the law for crop transpiration. Through this coupling
405   model, they hope to reveal the law for moisture and heat transfer of the SPAC. The SPAC is divided into
406   three layers, namely, the atmosphere at a high altitude, the plant canopy which is simplified into one
407   layer at the momentum transfer junction, and the soil layer. To be more specific, the top of the soil
408   layer is set to be the soil surface, and the bottom of the soil layer is set to be at the groundwater level.
409   According to the mathematical expression of total latent and sensible heat consumption of SPAC, the
410   mathematical expression of plant transpiration latent and sensible heat consumption, the
411   mathematical expression of soil evaporation latent and sensible heat consumption, and the model of
412   soil moisture migration and heat transfer, the results obtained from data-driven vector autoregression
413   models can be reasonably explained.

414   To begin with, the lagging effect of impulse response between soil temperature and moisture can
415 be explained as follows. Moisture's lagging effect on temperature can be partly interpreted by ~~one of~~
416 the basic equation ~~s within the model of soil moisture migration and~~ heat transfer equation:~~.~~

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

417 where T is soil temperature, C_v is soil volumetric heat capacity, and K_h is soil thermal conductivity.
418 C_v and K_h are close correlated with soil moisture content θ and can be expressed in the single factor
419 function form of θ. From the equation above, we can deduce that when there is a fluctuation in soil
420 moisture content, the first-order derivative of temperature T's function on time t will change. Thus,
421 although the contemporary value of temperature will not change, it's successive value will change
422 gradually in accordance with the change of the first-order derivative. Finally, at some point after the
423 interference point, the deviation of temperature will reach its maximum, accounting for moisture's
424 lagging effect on temperature. Temperature's lagging effect on moisture is to some extent related to
425 the relationship between water vapor relative saturation and soil temperature.

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

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

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

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

434 Where S_w is the water absorption intensity of roots, m denotes the ratio of water absorption rate
435 of the upper part to that of the lower part of roots, $lr(t)$ represents root depth, and E_v is crop
436 transpiration rate. Equation (27) describes how S_w is influenced by depth and time. Equation (28)
437 describes the way in which plant transpiration is influenced by S_w at different soil depths (Lei et al.,
438 1988). Given that there exists a close relationship between soil temperature and plant transpiration
439 latent and sensible heat consumption, it can be deduced that soil temperature varies with depth.
440   In addition to physical explanations, we used data from a different year and partly verified the
441 universality of the model. We collected data on soil temperature and moisture content at the same test
442 site from December 19th, 2016 to May 30th, 2017 at the depth of 10 cm, 30 cm, and 90 cm, during which
443 the same irrigation method, <mark>i.e.,</mark> drip irrigation, has been adopted. Then, we used the same method
444 stated above to build three vector autoregression models. Next, we built impulse response functions
445 and conducted Granger causality tests. The results we obtained are as follows. First, the time lag of soil
446 moisture's impulse response is 30 h, 70 h, and longer than 120 h, while the time lag of temperature's
447 impulse response is about 60 h, longer than 120 h, and longer than 120 h at 10 cm, 30 cm, and 90 cm.
448 Second, temperature's response intensity is 0.1989°C, 0.0157°C and 0.0031°C for 1% of variation in soil
449 moisture, and moisture's response intensity is 0.0578%, 0.0169% and 0.0057% for 1°C of variation in
450 soil temperature at 10 cm, 30 cm and 90 cm. Third, soil moisture is helpful in terms of predicting soil
451 temperature at the depth of 10 cm, 30 cm, and 90 cm. Besides, soil temperature is helpful in terms of
452 predicting soil moisture at the depth of 10 cm and 30 cm~~,~~ but has no obvious relationship with soil
453 moisture at 90 cm.
454   By comparing the results with those of 2017-2018, we can discover that the ratios for <u>the </u>intensity
455 of soil temperature's impulse response to moisture among three depths during both periods
456 approximate 200:16:3. And the ratios for <u>the </u>intensity of moisture's impulse response to temperature
457 during both periods approximate 60:17:6. Moreover, the results of Granger causality tests are the same,
458 indicating certain stability in terms of the physical and biological processes that involve the variation of
459 soil temperature and moisture in the same area. This demonstrates that for the same area the data-

460 driven model we proposed has consistency and the results we obtained are not out of a sudden. Thus,
461 the method of building data-driven vector autoregression models can be used to study the
462 characteristics of soil heat-moisture coupling. Although the numerical results of this model will change
463 with different soil properties in different areas, the method can be applied in the same way.

Conclusions

465 We designed a purely data-driven stochastic model for analyzing moisture-heat coupling of a
466 wheat-soil system, which consists of three vector autoregression models built at the depth of 10 cm,
467 30 cm, and 90 cm, impulse response functions, and Granger causality tests. For the empirical test of
468 this method, we use the hourly data on soil moisture and soil temperature at the depth of 10 cm, 30
469 cm, and 90 cm obtained at Shangqiu Experiment Station. Following conclusions can be drawn from the
470 models.
471 Firstly, the time lag of soil temperature's influence on soil moisture is for about 25 h at 10 cm, 50
472 h at 30 cm, and 120 h at 90 cm, while the time lag of soil moisture's influence on soil temperature is for
473 about 50 h at 10 cm and longer than 120 h at 30 cm and 90 cm. Secondly, the intensity of soil
474 temperature's impulse response to shock in soil moisture is 0.2004°C, 0.0163°C and 0.0035°C for 1%
475 variation in soil moisture respectively at the depth of 10 cm, 30 cm, and 90 cm. Similarly, the intensity
476 of soil moisture's impulse response is 0.0638%, 0.0163% and 0.0050% for 1°C of variation in soil
477 temperature respectively at the depth of 10 cm, 30 cm, and 90 cm. Thirdly, soil moisture is helpful in
478 terms of predicting soil temperature at the depth of 10 cm, 30 cm, and 90 cm. While soil temperature
479 helps predict soil moisture at the depth of 10 cm and 30 cm, it has no obvious correlation with soil
480 moisture at 90 cm.
481 Although the proposed method for analyzing moisture-heat coupling in a wheat-soil system has
482 some advantages such as data-driven and easy to achieve, there are limitations. One is that the vector
483 autoregression model is based on the linear hypothesis, so other nonlinear data-driven models should
484 be further studied. ~~In addition~~Also, since the soil of our experiment site is loam, we ought to further
485 explore whether such a method can be applied to other kinds of soil using data from different
486 experiment sites. We are currently working on this.

References

487
488 1.   Awad, W., Byrne, P.F., Reid, S.D., Comas, L.H. & Haley, S.D 2018 Great Plains Winter Wheat Varies for Root
489   Length and Diameter under Drought Stress. Agronomy Journal, **110**(1), 226-235, doi:
490   10.2134/agronj2017.07.0377.
491 2.   Bittelli M., Ventura F., Campbell G. S., Snyder R. L., Gallegati F. & Pisa P. R. 2008 Coupling of heat, water
492   vapor, and liquid water fluxes to compute evaporation in bare soils. *Journal of Hydrology*, **362**(3-4), 191–205,
493   doi: 10.1016/j.jhydrol.2008.08.014.
494 3.   Bolin D., Wallin J. & Lindgren F. (in press) Latent Gaussian random field mixture models. *Computational*
495   *Statistics & Data Analysis*, **130**, 80-93, doi: 10.1016/j.csda.2018.08.007.
496 4.   Bouri E., Gupta R., Hosseini S. & Lau C. K. M. 2018 Does global fear predict fear in BRICS stock markets?
497   Evidence from a Bayesian graphical structural VAR model. *Emerging Markets Review*, **34**, 124-142, doi:
498   10.1016/j.ememar.2017.11.004.
499 5.   Chen L.-J., Ho Y.-H., Hsieh H.-H., Huang S.-T., Lee H.-C. & Mahajan S. 2018 ADF: an anomaly detection
500   framework for large-scale PM2.5 sensing systems. *IEEE Internet of Things Journal*, **5**(2), 559-570, doi:
501   10.1109/JIOT.2017.2766085.
502 6.   Droumaguet M., Warne A. & Wozniak T. 2017 Granger Causality and Regime Inference in Markov Switching
503   VAR Models with Bayesian Methods. *Journal of Applied Econometrics*, **32**(4), 802-818, doi: 10.1002/jae.2531.
504 7.   Lei Z. D., Yang S. X., Xie S. C. 1988 Soil Water Dynamics (in Chinese). *Tsinghua University Press*, **314**, doi:
505   10.13243/j.cnki.slxb.1988.05.001.
506 8.   Liu B. C., Liu W. & Peng S. W. 2005 Study of heat and moisture transfer in soil with a dry surface layer.
507   *International Journal of Heat and Mass Transfer*, **48**(21-22), 4579–4589, doi:
508   10.1016/j.ijheatmasstransfer.2005.06.004.
509 9.   Lu N. & Dong Y. 2015 Closed-Form Equation for Thermal Conductivity of Unsaturated Soils at Room
510   Temperature. *Journal of Geotechnical and Geoenvironmental Engineering,* **141**(6),
511   doi:10.1061/(ASCE)GT.1943-5606.0001295.
512 10. Mao X. & Shang P. 2018 Extended AIC model based on high order moments and its application in the financial
513   market. *Physical A-statistical Mechanics and Its Applications*, **501**, 264-275, doi:
514   10.1016/j.physa.2018.02.159.

11. Meng C.H. & Xia J. 2005 Research on water and heat transfer in soil-plant-atmosphere system. *Journal of Hydrodynamics*, **20**(3), 307-312, doi: 10.16076/j.cnki.cjhd.2005.03.005.

12. Nassar I. N. & Horton, R. 1989 Water transport in unsaturated nonisothermal salty soil: II. Theoretical Development. *Soil Science Society of America Journal*, **53**(5), 1330, doi: 10.2136/sssaj1989.03615995005300050005x.

13. Pan Y., Shi G., Liu Y. & Lee F. 2018 Effect of spatial variability on performance of cement-treated soil slab during deep excavation. *Construction and Building Materials*, **188**, 505-519, doi: 10.1016/j.conbuildmat.2018.08.112.

14. Philip JR. & Vries DA. 1957 Moisture movement in porous material sunder temperature gradients. *Trans Am Geophy Unionk*, **38**, 22-231, doi: 10.1029/tr038i002p00222.

15. Perez E. H., Paez M. I. & Figueroa A. 2013 Effect of moisture and temperature on dissipation of chlorpyrifos and diazinon in andic soils, Cauca, Colombia. *Asian Journal of Chemistry*, **25**(16), 9208-9212, doi: 10.14233/ajchem.2013.15166.

16. Schoukens J., Godfrey K. & Schoukens M. 2018 Nonparametric data-driven modeling of linear systems estimating the frequency response and impulse response function. *IEEE Control Systems Magazines,* **38**(4), 49-88, doi: 10.1109/MCS.2018.2830080.

17. Striegl & Loheide 2012 Feasibility of soil moisture estimation using passive distributed temperature sensing. *Ground Water,* **50**(3), 340-347, doi:10.1111/j.1745-6584.2012.00928.x

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. 2010 Feasibility of soil moisture estimation using passive distributed temperature sensing. *Water Resources*, **46**, W03534, doi:10.1029/2009WR008272.

19. Stokes P. A. & Purdon P. L. 2018 Regarding interpretation of Granger causality analyses reply. *Proceedings of the National Academy of Sciences of the United States of America*, **115**(29), E6678-E6679, doi: 10.1073/pnas.1807099115.

20. Sun Z., Chen Q., Han X., Bol R., Qu B. & Meng F. 2018 Allocation of photosynthesized carbon in an intensively farmed winter wheat-soil system as revealed by (CO2)-C-14 pulse labelling. *Scientific Reports*, **8**, doi: 10.1038/s41598-018-21547-y.

21. Whelan A., Kechavarzi C., Coulon F. & Doerr S. H. 2015 Experimental characterization of the impact of temperature and moisture on the breakdown of soil water repellency in sandy soils and composts. *Hydrological Processes*, **29**(8), 2065-2073, doi: 10.1002/hyp.10305.

22. Yang Y., Shang S. & Guan H. 2012 Development of a soil-plant-atmosphere continuum model (HDS-SPAC) based on hybrid dual-source approach and its verification in wheat field. *Science China Technological Sciences*, **55**(10), 2671–2685, doi: 10.1007/s11431-012-4974-7.