IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 15, 2021, accepted March 3, 2021, date of publication March 8, 2021, date of current version March 15, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3064264

Causal Research on Soil Temperature and Moisture Content at Different Depths

ZHIHAO CAO[1], SHAOMIN MU[2], LI XU[1], MINGFENG SHAO[1], AND HONGCHUN QU[1]
[1]College of Information Science and Engineering, Zaozhuang University, Zaozhuang 277160, China
[2]College of Information Science and Engineering, Shandong Agricultural University, Tai'an 271018, China

Corresponding author: Hongchun Qu (uzzqhc@163.com)

This work was supported by the Key Research Program of the Science Foundation of Shandong Province under Grant ZR2020KE001.

ABSTRACT The soil system is complex and dynamic, making it difficult to understand using traditional statistical approaches. In this paper, we analyze the causal relationship of soil temperature and moisture content at different depths in summer and winter based on dynamic empirical modelling. Specifically, we describe the complexity of soil temperature and moisture content system through mathematical methods. Moreover, we demonstrate the direction and magnitude of causal relationship between soil moisture content and temperature at different depths by equation-free methods. Besides, we describe the difference of soil system properties in summer and winter through causal research. The experiments show that results obtained are consistent with the actual soil environment. The causality is described by dynamic empirical modelling rather than prior soil knowledge. The paper may provide a new idea for soil dynamics research.

INDEX TERMS Machine learning, soil science, causal research, dynamic empirical modelling.

I. INTRODUCTION

Soil is considered to be the skin of the earth and the medium for crop growth. Soil science is closely related to agriculture [1], [2]. Soil temperature and moisture content are two important factors affecting crop growth. It is significant to study soil temperature and moisture content for analyzing crop growth.

Farmland is a kind of natural system, and natural systems are often complex and nonlinear, making them difficult to understand using linear statistical approaches. Linear approaches such as linear regression are fundamentally based on correlation [3], [4]. However, these are ill-posed for natural systems, where correlation can occur in the absence of causation, and causation may also occur without correlation [5]. For example, there are many microorganisms in the soil. In a certain temperature range, the activity of microorganisms will increase with the increase of ambient temperature. However, when the temperature reaches a certain level, the activity of microorganisms will decrease with the increase of temperature. In hot summer, soil moisture content and air temperature show a very obvious inverse relationship, because the higher the temperature, the faster the water evaporation. In the cold winter, this inverse relationship will become weak, because the temperature in winter is low, then the soil may freeze, and the water evaporation will be very slow. The two phenomena above belong to the very typical ''mirage correlation''. Mirage correlation means the sign and magnitude of the correlation between different variables may change with time. Mirage correlation is the hallmark of nonlinear systems that results from state dependency [6]. State dependency means that the relationships among interacting variables change with different states of the system, as the relationship between soil temperature and microbial activity mentioned above. Such state-dependent behavior is a defining hallmark of complex nonlinear systems [7]. With increasing recognition that nonlinear systems are ubiquitous, and that relationships among variables will depend on system state, the use of correlation to infer causation becomes truly difficult.

According to our daily prior knowledge, the soil heat mainly comes from solar energy, and the moisture comes from precipitation and irrigation [2]. According to the statistical data, the temperature and moisture content in the shallow layer of soil are in direct proportion to these in the deep layer, because the shallow soil will top-down transfer the heat and water to the deep layer according to general knowledge. However, these analyses are only based on the correlation between the data. For example, it can be seen

The associate editor coordinating the review of this manuscript and approving it for publication was Alexandros Iosifidis.

from the statistical data that the above two variables present a positive correlation, but this correlation does not indicate that there is a causation between the two variables [7].

In recent years, causal analysis has become one of the most challenging fields in machine learning [8]–[14]. For examples, some researchers proposed causal analysis methods in medical diagnosis [15], [16] and social sciences [17], [18]. Causal analysis plays an important role in revealing the essential relationship of things and identifying causal relationship is important for effective management recommendations on climate, agriculture, epidemiology, financial regulation, and much else [6]. At present, machine learning algorithms commonly used in artificial intelligence are based on correlation rather than causation. Causal analysis is considered to be an important factor in the realization of ''artificial general intelligence'' [5]. Causal analysis methods commonly used in natural systems include Granger causality [19], nonlinear state-space methods, causal network learning algorithms [8] and so on. Empirical dynamic modelling (EDM) is an important algorithm [3], and is especially suitable for the analysis of natural systems [4], [20].

To effectively manage available water and heat resources, it is important to understand the causal processes of water and heat in soil system [1]. In this paper, empirical dynamic modelling is used to analyze the soil temperature and moisture system at different depths. The experimental results show that the conclusions obtained by dynamic empirical modelling are consistent with the actual soil environment. The causality between soil temperature and moisture content at different depths is described by scientific methods.

To summarize, the main contributions of this work are in four-fold as follows: 1) The complexity of soil temperature and moisture content system is described through simple projection algorithm; 2) The direction and magnitude of causal relationship between soil temperature and moisture content at different depths are demonstrated by equation-free methods rather than prior soil knowledge; 3) The difference of soil properties in summer and winter is described through causal research; 4) The paper may provide a new idea for soil dynamics research.

II. RELATED WORK

A. DATA SOURCES

Farmland soil temperature and moisture content can be seen as a series of data that change with time, which can be collected by various sensors in farmland. For example, Fig.1 shows the time-varying temperature of soil surface, 10 cm depth, 20 cm depth and 30 cm depth over a period of time. It can be seen from the figure that the soil temperature changes periodically, with the highest temperature and the lowest temperature occurring once in a cycle. With the increase of depth, the change of soil temperature in the lower layer lags behind the surface temperature more obviously, and the temperature amplitude is weakening. As the energy must be conducted to the subsoil, the fluctuation in daily

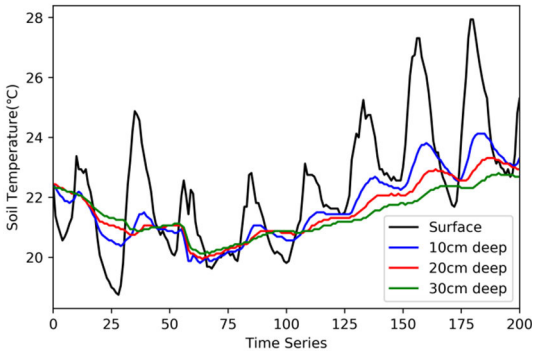


FIGURE 1. The time-varying temperature of soil surface, 10 cm depth, 20 cm depth and 30 cm depth over a period of time.

FIGURE 2. The specific location where the data were collected.

temperature is progressively less in deeper sections of the profile [1]. Research on this kind of data can be seen as a kind of time series analysis.

We used special sensors to collect soil data in the Tea Valley (36°11'N 117°06'E) of Taian, Shandong Province, China. The specific location is shown in Fig.2. The data were collected in August (representing summer) and December (representing winter). The crop grown in this area is Longjing Tea.

The acquisition interval is set to one hour, and there are about 720 sets of data in a month. It includes nine kinds of data: surface temperature (T00), 10cm soil temperature (T10), 20cm soil temperature (T20), 30cm soil temperature (T30), 40cm soil temperature (T40), 10cm soil moisture content (M10), 20cm soil moisture content (M20), 30cm soil moisture content (M30) and 40cm soil moisture content (M40). Some of the data are shown in Table 1 and Table 2. The unit of temperature is °. The soil moisture content is expressed by the volume water content, and the unit is %, which refers to the proportion of water in the total soil volume.

The statistical information of soil temperature and moisture content in the original data is shown in Table 3, including mean and standard deviation. The units in the table refer to Table 1 and Table 2. Sign (s) represents the data in summer, and sign (w) represents the data in winter in Table 3. The arrows in the table indicate the direction in which the values increase.

TABLE 1. Some examples of data used in the experiment (August, summer).

Time	T00(℃)	M10(%)	T10(℃)	M20(%)	T20(℃)	M30(%)	T30(℃)	M40(%)	T40(℃)
1	22.4375	28.17	22.4375	28.17	22.4375	26.25	22.3750	26.99	22.2500
2	21.3750	30.85	22.3125	29.56	22.4375	27.14	22.3125	27.47	22.1875
3	21.0625	30.77	22.1875	29.53	22.3750	27.56	22.3125	27.86	22.1875
4	20.8125	29.96	22.0625	29.12	22.3125	27.15	22.3125	27.70	22.1250

TABLE 2. Some examples of data used in the experiment (December, winter).

Time	T00(℃)	M10(%)	T10(℃)	M20(%)	T20(℃)	M30(%)	T30(℃)	M40(%)	T40(℃)
1	5.3125	26.51	8.8125	21.72	10.3750	21.19	11.6250	24.76	12.3125
2	5.0000	26.46	8.5625	21.71	10.3125	21.18	11.5625	24.74	12.2500
3	5.4375	26.41	8.5625	21.73	10.1875	21.19	11.5000	24.74	12.2500
4	5.8125	26.39	8.5000	21.74	10.1875	21.20	11.4375	24.75	12.2500

TABLE 3. The statistical information of soil temperature and moisture content at different depths in the original data.

Data	Mean		Standard deviation
T00(s)	22.802	▲ High	1.807
T10(s)	22.931		0.960
T20(s)	22.275		0.783
T30(s)	22.119		0.661
T40(s)	22.006		0.560
M10(s)	28.262	▲ High	1.719
M20(s)	27.754		1.072
M30(s)	25.887		1.093
M40(s)	26.621		0.955
T00(w)	7.4240		3.754
T10(w)	8.6090		1.888
T20(w)	9.3990		1.412
T30(w)	10.123		1.186
T40(w)	10.568	▼ High	1.074
M10(w)	19.847		2.969
M20(w)	21.306		0.992
M30(w)	20.829		0.751
M40(w)	24.686	▼ High	0.170

In summer, the average soil temperature at all depths is about 22 degrees, and the temperature difference between different depths is small. The average moisture content in summer is about 26%-28%, and the difference between different depths is also small. Note that surface temperature and moisture content is higher than subsoil. In winter, the average soil temperature changes greatly with the depth, the surface average temperature (T00) is only about 7.4 degrees, but the average temperature of 40 cm depth (T40) can reach 10.5 degrees. The soil moisture content also increases with the depth significantly. It can be seen that the system properties of soil temperature and moisture content are different in different seasons. Note that surface temperature and moisture content is low than subsoil. Moreover, the standard deviation of surface soil is high than subsoil, which indicates the change amplitude of temperature and moisture content is greater.

In this paper, we use dynamic empirical modelling to describe the causality between these data. Note that soil temperature and moisture content may be affected by some interference factors. For examples, The Longjing tea planted in this field is dormant in winter, but active in summer. The average root length of Longjing tea is about 15 cm which is smaller than 40cm. Microorganisms are more active in summer than winter. In this paper, we mainly study causal relationship between soil temperature and moisture content at different depths.

B. TIME SERIES ANALYSIS

Many methods can be used in time series analysis, such as linear regression, temporal convolutional networks, neural networks, RNN and empirical dynamic modelling.

Neural networks are the most basic model in deep learning [21]. Assuming that the neural networks are used to predict the soil temperature at 30 cm depth (T30) by the other three variables (T00, T10, T20), in this case, the other three temperatures at the same time will be input into the networks, as shown in Fig.3 (a). Then the predicted value will be output. The neural networks can automatically learn the relationship between the four variables, and continuously update the network parameters according to the loss function, so as to improve the prediction accuracy [22]. Although the traditional neural networks can analyze the time series, this method does not consider the time factor and ignores the time continuity between variables. These methods without considering time continuity are called time series regression.

Traditional neural networks can only input multiple variables at a certain time point [23]. The previous input and the subsequent input of the same variable are completely irrelevant, although they are related in time. However, the temperature data contains abundant time-series information, and the temperature in the former moment is closely related to that in the latter. Recurrent neural networks (RNN) and variants such as LSTM (long short term memory) are novel neural networks that can process sequence data [24]–[26], as shown

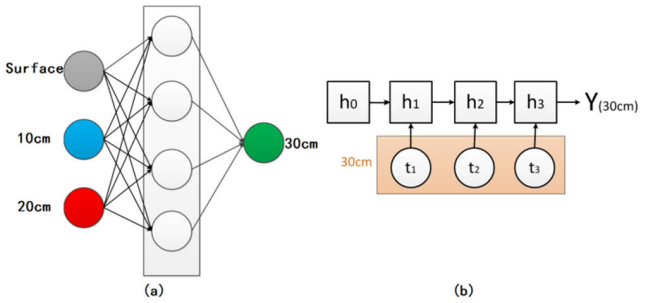


FIGURE 3. Traditional neural networks and recurrent neural networks.

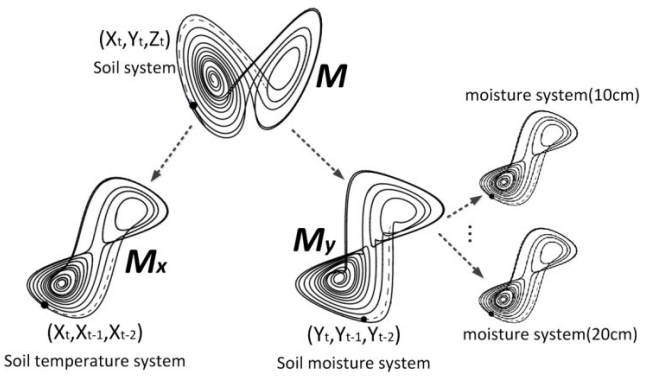


FIGURE 4. Convergent cross mapping tests for the shadow manifolds M_x (temperature system) and M_y (moisture system).

in Fig.3 (b). These methods with considering time continuity are called time series prediction. However, these algorithms only focus on curve fitting of correlation. Although these algorithms have good effects on the prediction and classification, they lack interpretability and cannot explain the causal relationship between variables in essence.

At present, machine learning algorithms used commonly in artificial intelligence are mostly based on correlation rather than causality [5]. EDM was proposed by Hao Ye *et al.* and based on the concept of attractor reconstruction to analyze time series for causation problems [19], [27]. The goal of attractor reconstruction is to approximate the originating dynamic system using time series data. EDM was developed on the basis of Takens' Theorem [28], [29]. The EDM methods mostly include convergent cross mapping (CCM) algorithm [7], simplex projection algorithm and S-Map algorithm [30].

The traditional linear statistical method is not suitable for the general nonlinear dynamic systems, because in the ecosystem dynamics, farmland is easily affected by various interference factors such as the activities of microorganisms and organic matter in the soil. In recent years, causality research has developed rapidly. The EDM method is composed of a series of nonlinear statistical methods that acknowledge state dependence. These nonlinear methods are rooted in state space reconstruction, i.e. lagged coordinate embedding of time series data [6]. These methods do not assume any set of equations governing the system but recover the dynamics from time-series data. EDM bears a variety of utilities to investigating dynamical systems [3]. Some researchers try to use the EDM to analyze the causal relationship between multiple variables in natural systems and have achieved good results [31]–[35].

CCM was proposed by Sugihara G *et al*. In CCM theory, time series variables such as the soil temperature and moisture content are causally linked because they are from the same soil dynamic system, there are sharing a common attractor manifold M, as shown in Fig.4. M_x represents shadow manifold of soil temperature and M_y represents shadow manifold of soil moisture content. This algorithm predicts the current quantity of one variable M_x using the time lags of another variable M_y and vice versa. If M_x and M_y belong to the same dynamical system, the cross-mapping between them shall be convergent. This means that each variable can identify the state of other variables. Additionally, when one variable A is a stochastic driver of one variable B, and we can express it as $A \rightarrow B$, information about the states of A can be recovered from B, but not vice-versa. CCM can test causation by measuring the extent to which the historical record of B values can reliably estimate the state of A.

III. RESULTS

A. COMPLEXITY OF SOIL TEMPERATURE DYNAMICAL SYSTEM

Soil temperature is the result of solar radiation balance, soil heat balance and soil thermal properties [1]. The most basic source of soil heat is solar radiation, and the other part mostly comes from geothermal energy. Since the experimental data acquisition area is a common field, the geothermal energy can be ignored. When the soil obtains heat from solar radiation, most of the heat is consumed by the process of water evaporation and heat exchange between soil and air, while the other part is transmitted to the subsoil through heat exchange, and a small part is consumed by soil biological activities [36]. Soil temperature affects almost every physical, chemical, and biological activity that occurs in the soil. It is very important for the production of crops.

The complexity of dynamic system can be tested using simple projection algorithm by nonparametric predictions. The complexity of a system can be practically defined as the number of independent variables needed to reconstruct the attractor [37], [38]. Based on Takens' Theorem [28], the dynamics of the system can be reconstructed from the time lags of a single time $\{x_t, x_{t-1\tau}, x_{t-2\tau}, \ldots x_{t-(E-1)\tau}\}$, as shown in Fig.5, where τ is time lag and E is the embedding dimension. Determining embedding dimension E is basic in EDM. By carrying out simplex projection using different values of E in library set, the optimal embedding dimension E can be determined according to the predictive skill ρ such as the correlation coefficient between the observations Y and the forecast result $\hat{Y}$. The optimal E is selected based on the criterion that maximizes the predictive skill ρ by evaluating the correlation coefficients. Note that the higher the best

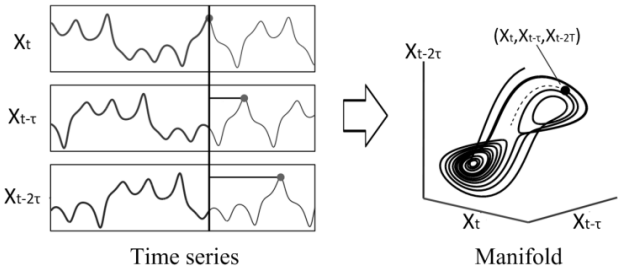


FIGURE 5. The diagram of simple projection algorithm.

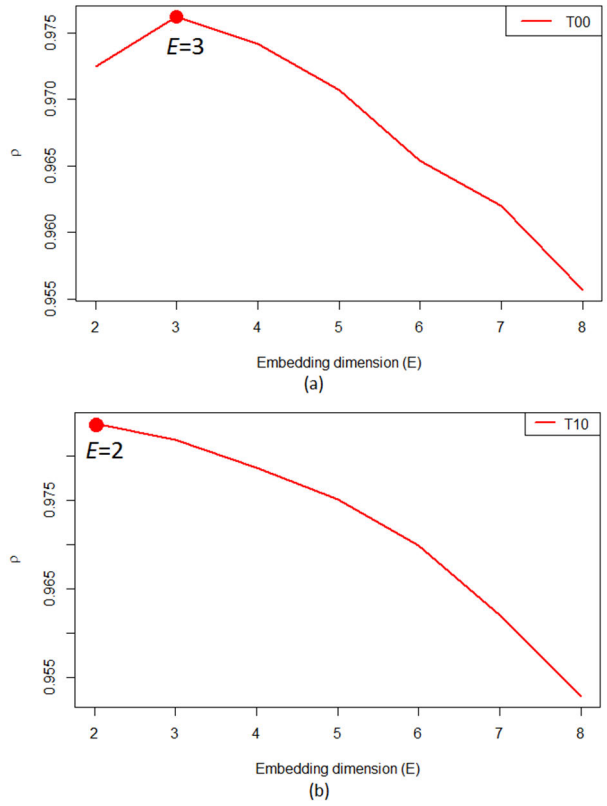


FIGURE 6. Simplex projection algorithm is used to test the complexity of soil temperature system (T00, T10) in summer.

embedding dimension E, the more complex the dynamical system is.

Simple projection algorithm is used in this section to test the complexity of soil temperature dynamical system. Some experimental results are shown in Fig.6 (summer) and Fig.7 (winter). The horizontal axis represents the embedding dimension E in the algorithm, and the vertical axis represents the corresponding predictive skill ρ. It can be seen from Fig. 6 that the best embedding dimension of surface temperature (T00) system in summer is 3, while the best embedding dimension of 10 cm (T10) depth systems is 2, indicating that the complexity of soil surface temperature system is higher than that of soil internal temperature system in summer. The detailed experimental results are shown in Table 4.

The environment of surface soil is complex because the surface soil is in direct contact with strong solar radiation,

TABLE 4. The best embedding dimension E of temperature and moisture content systems at different depths in summer and winter.

Data	Season	The Optimal E
T00	summer	3
T10	summer	2
T20	summer	2
T30	summer	2
T40	summer	2
T00	winter	2
T10	winter	2
T20	winter	2
T30	winter	2
T40	winter	2
M10	summer	2
M20	summer	16
M30	summer	17
M40	summer	18
M10	winter	13
M20	winter	2
M30	winter	5
M40	winter	2

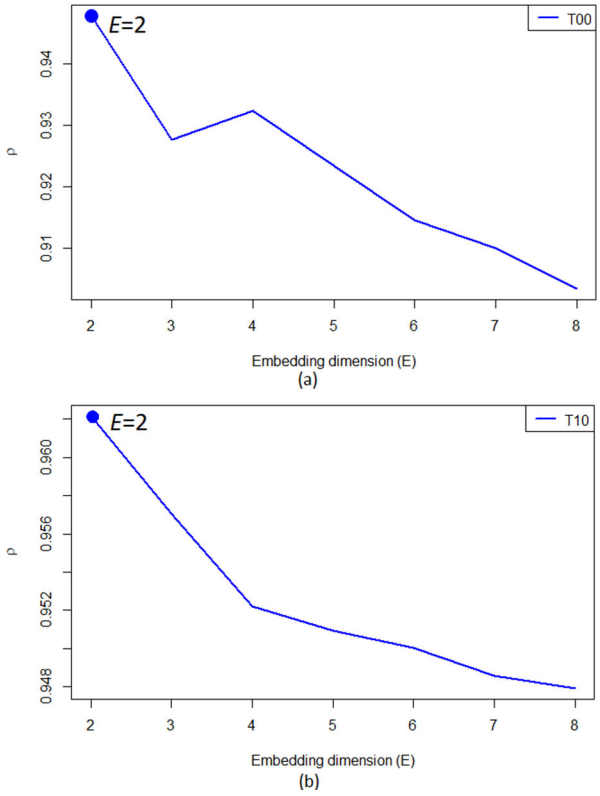


FIGURE 7. Simplex projection algorithm is used to test the complexity of soil temperature system (T00, T10) in winter.

abundant precipitation and water evaporation process in summer. The embedding dimension E of temperature system in summer is in the range of 2-3, which is relatively simple. It can be seen from Fig.7 and Table 4 that in winter the optimal embedding dimension of soil temperature system at different depths is 2, which is simpler than summer. This is because the soil receives less heat and microorganisms are inactive in winter due to low solar radiation.

B. COMPLEXITY OF SOIL MOISTURE DYNAMICAL SYSTEM

Water stored in the soil does several things. It is essential to plant growth and plays an important role in the moderation of soil temperatures. Water is one of the most important components of soil, usually accounting for 20%-30% of soil volume. Soil water is essential to all forms of life—both plants and animals. Soil water can evaporate directly from surface soil [1]. The water in the soil has two main inputs (i.e. precipitation and irrigation) and outputs (i.e. evaporation and percolation). Therefore, soil moisture content system is more complex than soil temperature system.

The simple projection algorithm is used in this section to test the complexity of soil moisture dynamical systems at different depths. Some experimental results are shown in Fig.8 (summer) and Fig.9 (winter). The detailed experimental results are shown in Table 4 above. It can be seen from Table 4 that the soil moisture system at 10 cm depth (M10) is simple in summer, while the deeper soil system is complex relatively. This is due to the mutual influence of many factors such as the root activity. Compared with soil temperature, plant roots have a greater impact on soil moisture content [2]. The average root length of Longjing tea planted in this farmland is about 15 cm, so the soil moisture content is obviously affected by plant roots at the depth of 10 cm, and water can be stored in the root zone. However, when the depth is more than 15 cm, the soil moisture content system lacks the influence of plant roots. Moreover, there is more precipitation in summer, which makes the deep soil moisture system more complex.

It can be seen from Table 4 that the soil moisture content system is different in winter from summer. In winter, the surface soil moisture content system is complex, while the deep moisture content system is simple relatively. In winter, the complexities are reverse. This is due to the different environment of the soil in summer and winter. For example, Longjing tea is dormant in winter, and metabolism activities such as water and nutrient absorption are slow. Therefore, Longjing roots have little effect on soil system. Moreover, there is less precipitation in winter than summer, and the surface soil is exposed to precipitation and water evaporation directly.

C. CAUSAL ANALYSIS OF SOIL TEMPERATURE AT DIFFERENT DEPTHS

In this section, we selected the temperature data of soil adjacent layers for causal analysis. EDM can be used to reveal causation between variables. Convergent cross mapping algorithm is used to test the causation between a pair of temperature variables in dynamical systems [20], [39]. This algorithm predicts the current quantity of one variable X using the time lags of another variable Y. If X and Y are causally linked, the cross-mapping between them will be convergent. Convergence means that the cross-mapping skill ρ improves with increasing library size. Convergence

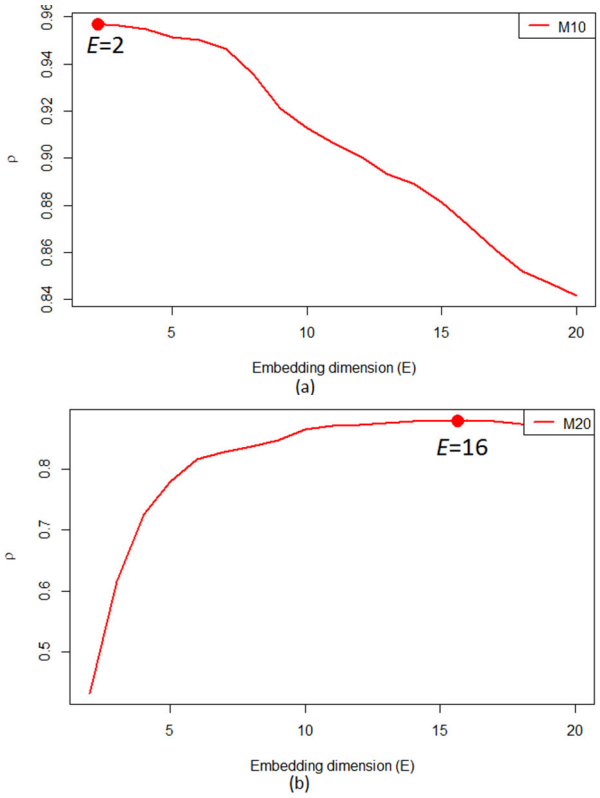


FIGURE 8. Simplex projection algorithm is used to test the complexity of soil moisture content system M10 and M20 in summer.

is the key property that distinguishes causation from simple correlation [7]. Convergence is a necessary condition for causation. In practical applications, convergence will be limited by process noise and time series length. Thus, with limited or noisy field data causation is demonstrated by predictability that increases with library size. It is generally considered that the more significant the convergence, the stronger the causal relationship.

Thermal energy is transferred as a result of a temperature difference within or between objects and heat always flows from a warm object to a cooler one. Only the surface soil is subject to solar energy input. Once the energy is absorbed by the surface soil, the soil is attempting to reach equilibrium with the topsoil above as well as the subsoil below [1].

The experimental results are shown in Fig.10 (summer) and Fig.11 (winter), where the curve "X cross mapping Y (X xmap Y)" represents testing Y as a cause of X. It can be seen from the curve "T10 xmap T00" in Fig.10(a) that the change of soil surface temperature is the cause of temperature change at 10 cm depth. This blue curve shows a high convergence magnitude ($\rho = 0.91$), while the red curve shows a weak convergence trend ($\rho = 0.85$). It can be concluded that the change of surface temperature leads to the change of temperature at the depth of 10 cm, rather than the change of temperature at the depth of 10 cm leading to the change of surface temperature.

However, the convergence of blue curve and red curve is very weak in Fig.10(b-d), and the cross-mapping skill $\rho \approx 1$,

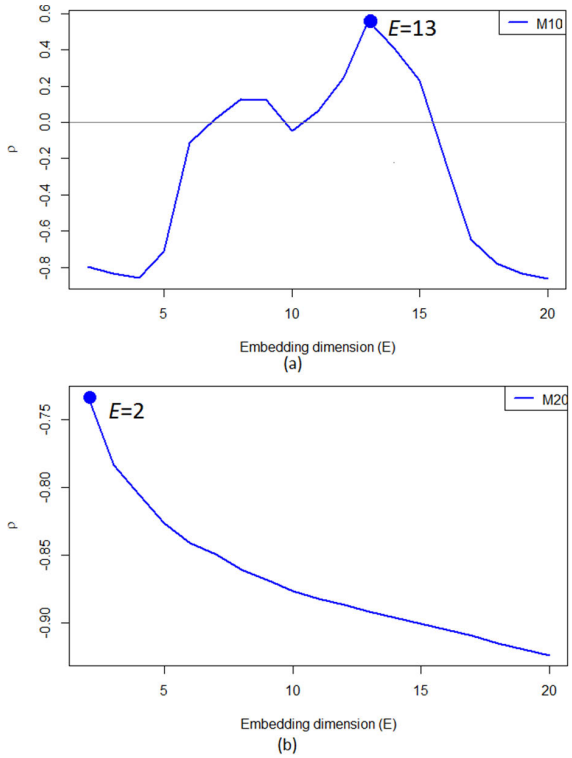


FIGURE 9. Simplex projection algorithm is used to test the complexity of soil moisture content system M10 and M20 in winter.

as show in the figures. It means that the properties of deep soil and shallow soil are very similar, and there is little causal relationship. Because the two temperature variables belong to the same system, i.e. soil temperature system. Moreover, the difference between the deep soil layers is less obvious than that between the shallow soil layers, because the environment of deep soil is simple without plants and precipitation.

In winter, the convergence at different depths is shown in Fig.11. It can be concluded that the change of T10 leads to the change of T00, and the change of T20 leads to the change of T10. The causal relationship is different from summer. This is because deep soil is warmer than shallow soil in winter, but shallow soil is warmer than deep soil in summer. Thermal energy is transferred as a result of the temperature difference between soil layers and heat always flows from a warm layer to a cooler layer. The statistical information of soil temperature at different depths can be seen in Table 3.

Detailed direction and magnitude of causal relationship between different depths are shown in Fig.12. The brown arrows in the figure show the direction of causal relationship. The thickness of the line indicates the magnitude of causal relationship. The dotted rectangle indicates that the soil temperature properties between the adjacent layers are very similar and can be considered as the same system.

D. CAUSAL ANALYSIS OF SOIL MOISTURE AT DIFFERENT DEPTHS

In this section, we selected the soil moisture data of adjacent layers for causal analysis in summer and winter. The experimental results are shown in Fig.13 (summer) and Fig.14 (winter).

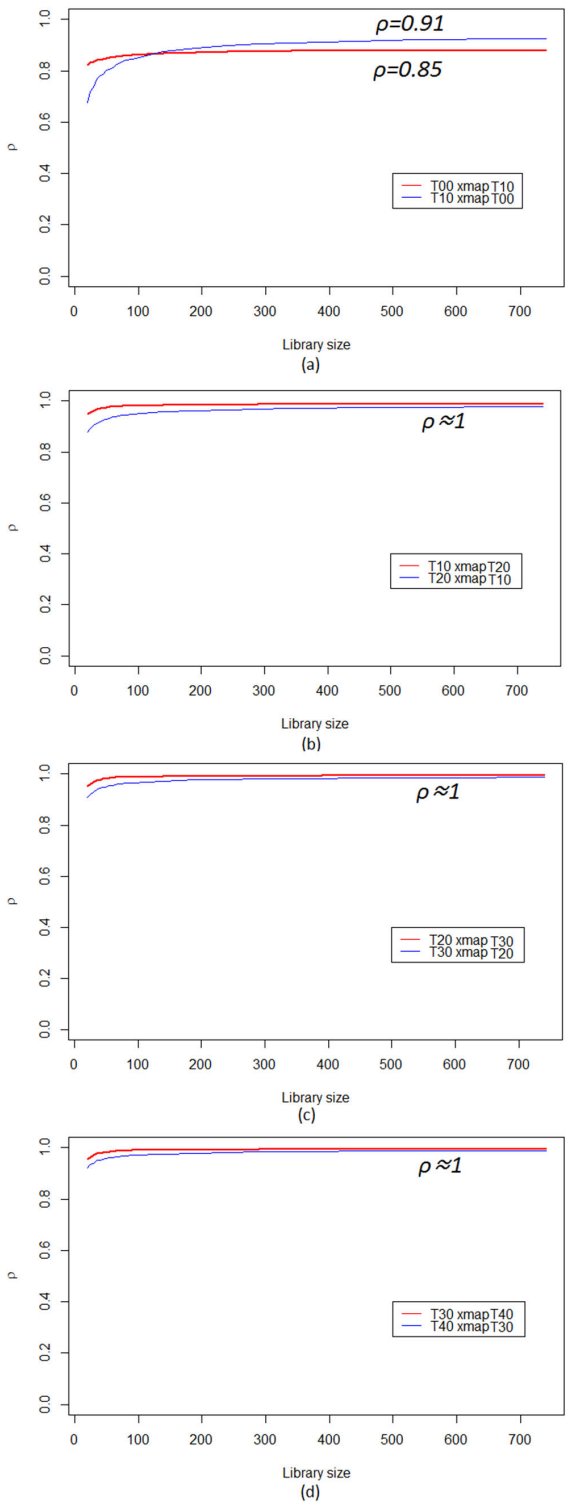


FIGURE 10. CCM algorithm is used to test the causation between soil temperatures at different depths in summer.

In general, soil water such as precipitation percolates by the force of gravity until it is adsorbed by drier soil below. Water flow in the uniform soil will be primarily downward [1].

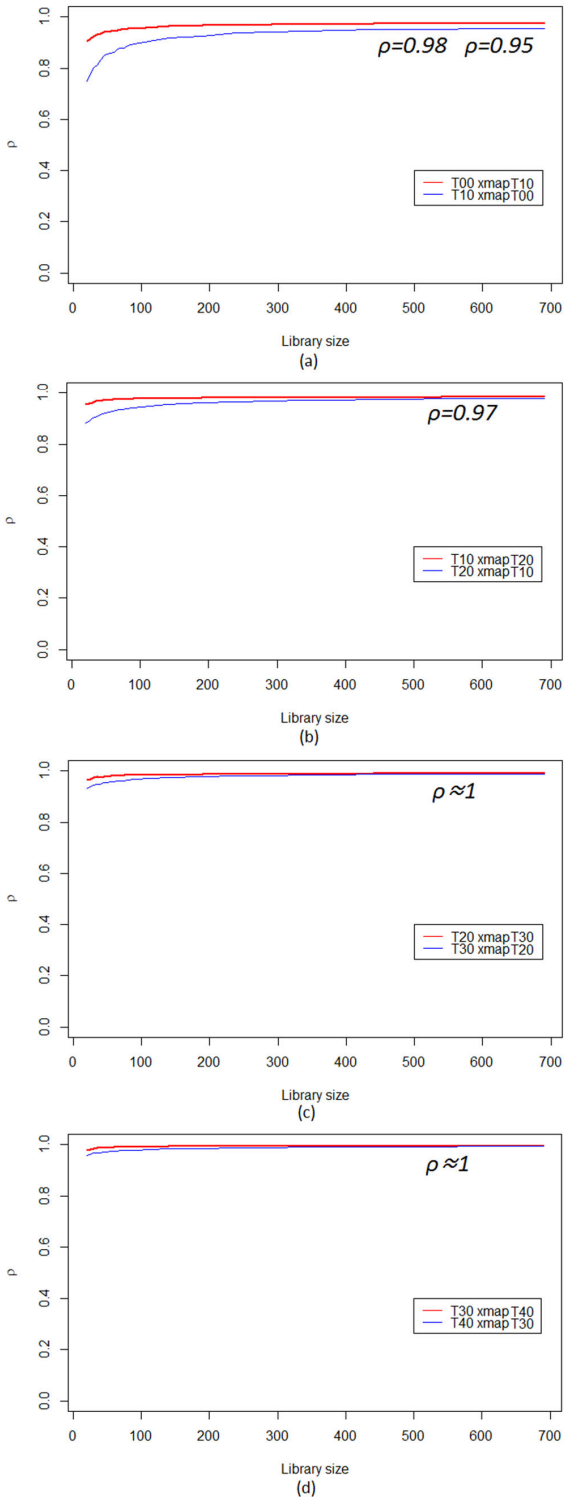


FIGURE 11. CCM algorithm is used to test the causation between soil temperatures at different depths in winter.

However, soil water will flow upward in some cases. For example, when the temperature is high in summer, the water in shallow soil evaporates quickly, and the water in subsoil will permeate upward. Soil water may also move upward or laterally in the process of furrow irrigation. Therefore, water can flow down or up in the soil, indicating the bidirectional

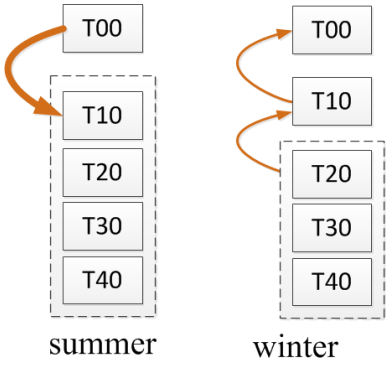


FIGURE 12. The direction and magnitude of causal relationship between different depths in summer and winter.

causal relationship. In Fig.13(a), the red and blue curve both show obvious convergence. It presents obvious bidirectional causality between M10 and M20. However, it can be seen that the blue curve shows a high convergence magnitude ($\rho = 0.93$), while the red curve shows a weak convergence magnitude ($\rho = 0.89$). It can be concluded that the change of surface moisture (M10) leads to the change of moisture at the depth of 20 cm (M20) more significantly. Besides, it can be seen from Table 4 that the moisture content of shallow soil is higher than that of deep soil in summer. The convergence magnitude in Fig.13(b-c) is very weak, and the cross-mapping skill $\rho \approx 1$. It means that the moisture properties of deep soil and shallow soil moisture content system are very similar, and there is little causal relationship.

In winter, the convergence of the red curve ($\rho = 0.99$) in Fig. 14(a) is more obvious than blue curve ($\rho = 0.97$), indicating that the change of M20 is the cause for the change of M10. This is because the moisture content of deep soil is higher than that of shallow soil in winter, as shown in Table 4. The convergence magnitude in Fig.14(b-c) is very weak ($\rho \approx 1$). It means that the properties of M20, M30, M40 are very similar, and there is little causal relationship between them. Due to the low temperature in winter, the water evaporation and penetration is weak. The curves in Fig.14(b-c) are almost straight compared with Fig.13(b-c), which indicate there is almost no causal relationship.

Detailed direction and magnitude of causal relationship between different depths are shown in Fig.15. The brown arrows in the figure show the direction of causal relationship. The dotted arrow indicates the bidirectional causal relationship between M10 and M20. The thickness of the line indicates the magnitude of causal relationship. The dotted rectangle indicates that the properties of soil moisture content inside are very similar and can be considered as the same system.

E. CAUSAL ANALYSIS BETWEEN SOIL MOISTURE CONTENTS AT DIFFERENT DEPTHS AND ENVIRONMENT TEMPERATURE

In this section, the causal relationship between soil moisture contents at different depths and soil temperature (T00 and

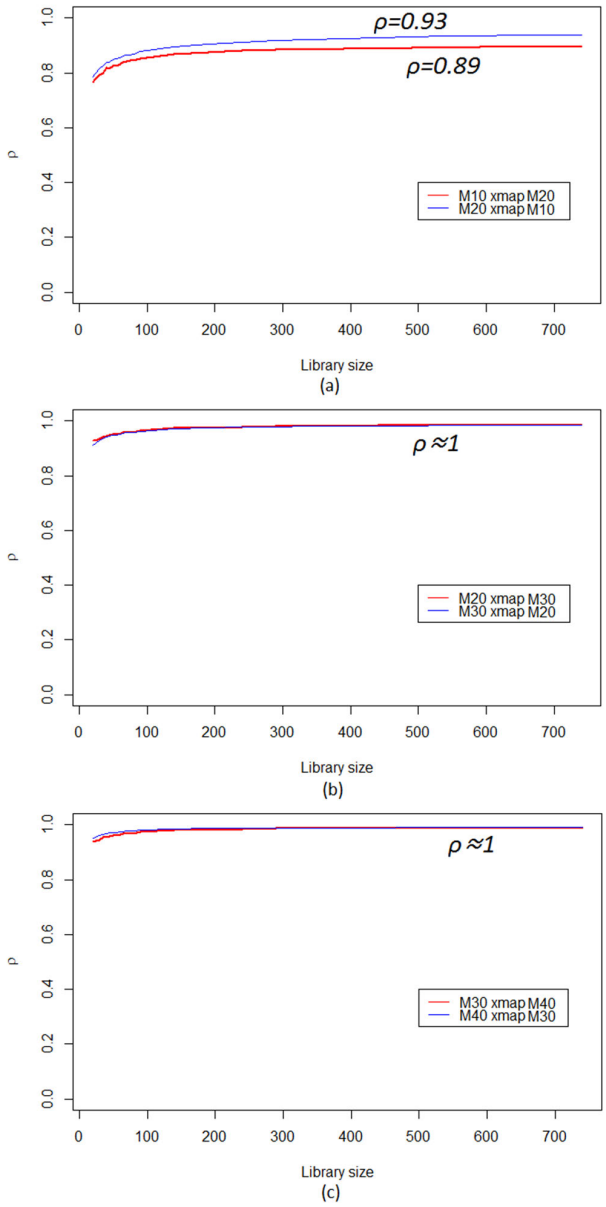


FIGURE 13. CCM algorithm is used to test the causation between soil moisture contents at different depths in summer.

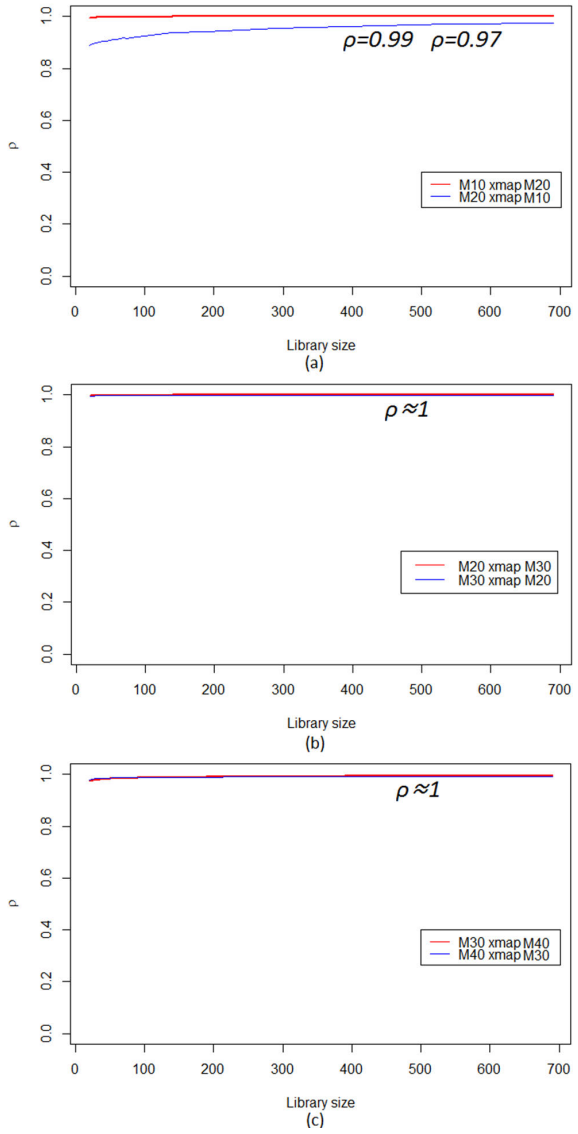


FIGURE 14. CCM algorithm is used to test the causation between soil moisture contents at different depths in winter.

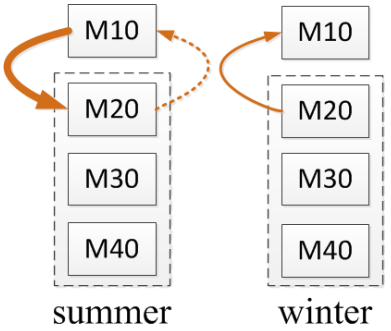


FIGURE 15. The direction and magnitude of causal relationship between different depths in summer and winter.

T40) is analysed. T00 and T40 can represent the soil surface temperature (environment temperature) and soil internal temperature respectively. The experimental results are shown in Fig.16 (summer), Fig.17 (winter) and Fig.18.

It can be seen from Fig.16(a) that the convergence of the red curve ($\rho = 0.62$) is more obvious than blue curve ($\rho = 0.39$), indicating that the soil surface temperature change (T00) is an important cause for the change of soil moisture content at 10 cm depth (M10). Moreover, the causal relationship between T00 and M10 is bidirectional and the soil moisture content (M10) can also influence soil surface temperature (T10). There is a similar phenomenon in Fig.16(b-c). This shows that the soil temperature "drives" the moisture content to change more significantly. It is because the temperature has a great influence on water activity. For example, warm soil conditions can promote water evaporation [2]. Note that the blue curve also shows a weak convergence trend, which indicates that moisture content also affects soil temperature, because the activity of water in soil

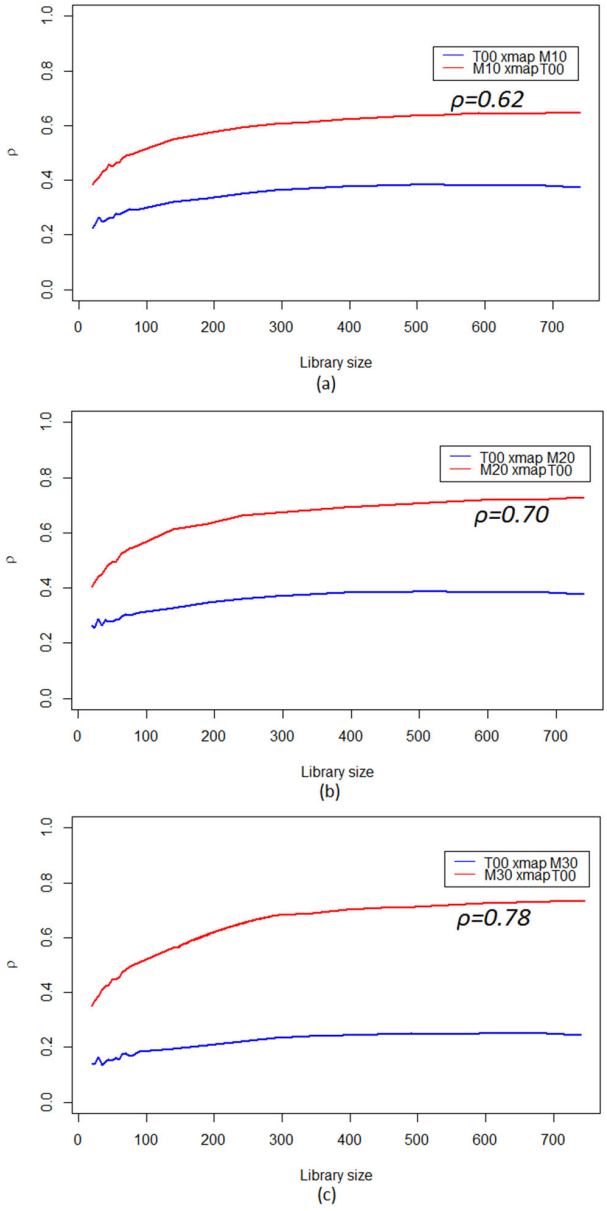


FIGURE 16. CCM is used to test the causation between soil moisture contents at different depths and soil surface temperature (T00) in summer.

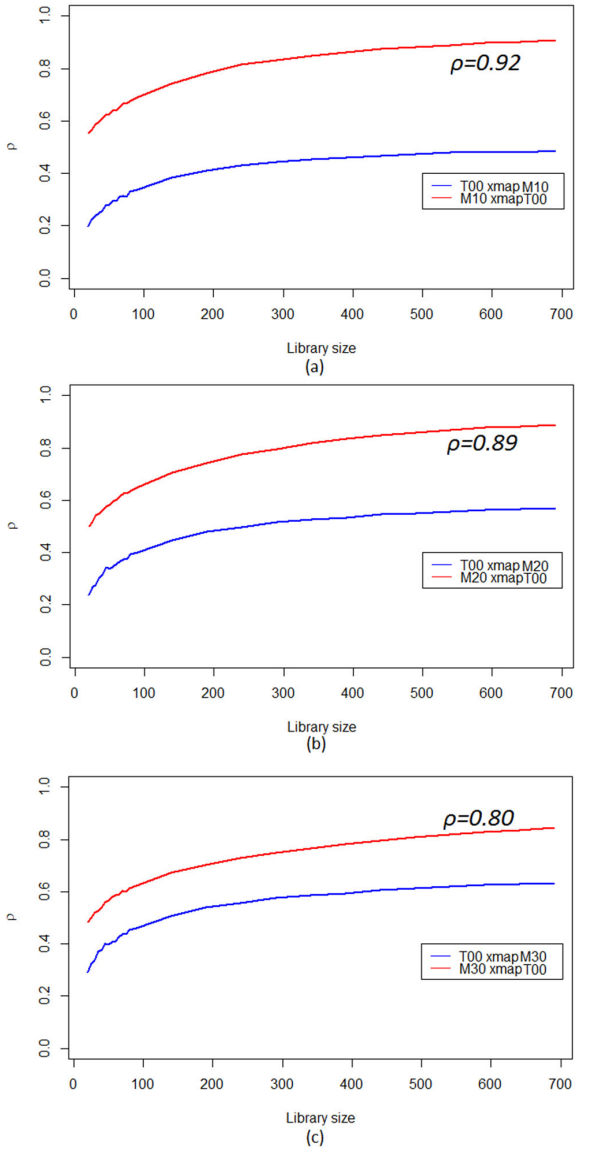


FIGURE 17. CCM is used to test the causation between soil moisture contents at different depths and soil surface temperature (T00) in winter.

is closely related to energy. For example, the process of water evaporating will absorb heat, creating feedback on the moisture content system. Therefore, soil moisture is an important factor in controlling the rate of temperature change [1].

The experimental results in winter are shown in Fig.17. The convergence of red curves is still more obvious, which indicates that soil surface temperature (T00) is also an important cause for the change of soil moisture contents at different depths. However, the convergence magnitude ρ of the red curves increases compared with that in summer. This indicates that soil temperature has a greater influence on the soil moisture content change in winter than summer. This is because the soil environment in summer is complex, and there are many extra factors affecting the soil moisture content (wind, microorganisms, plant root activity, etc.), while the soil environment in winter is simple, so the influence of soil temperature on the moisture content is more obvious, resulting in the increase of predictive skill ρ.

Then we analyse the causal relationship between soil moisture contents at different depths and soil internal temperature (T40). The detailed direction and magnitude ρ of causal relationship are shown in Fig.18.

It can be seen from Fig.18 that the causal effect of T00 on soil moisture content is more obvious than T40 in summer. The causal effect of T40 on soil moisture content is more obvious than T00 in winter by comparing predictive skill ρ. T00 and T40 can represent the soil surface temperature and internal temperature respectively. Moreover, the causal effect of soil temperature on moisture content is more obvious in winter than in summer.

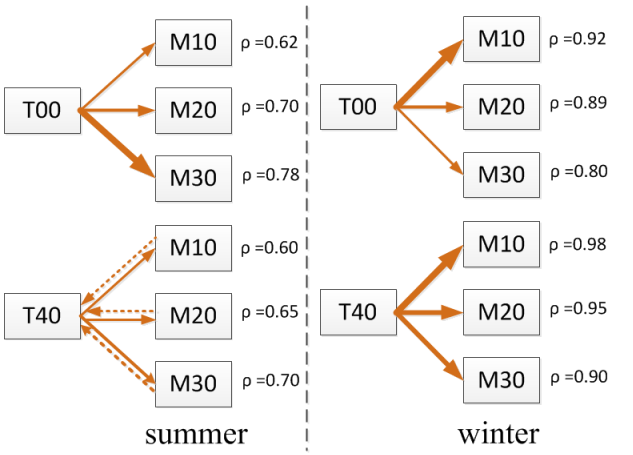


FIGURE 18. The direction and magnitude of causal relationship between different depths and soil temperature (T00, T40) in summer and winter.

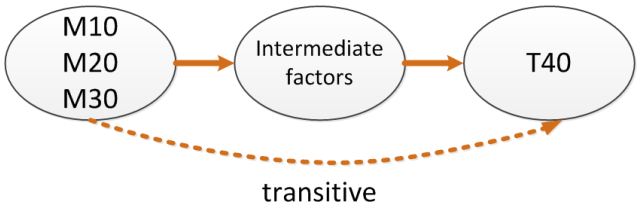


FIGURE 19. The causal relationship among moisture contents at different depths (M10, M20, M30) and T40. The causation is transitive.

It presents significant bidirectional causality between T40 and moisture contents at different depths (M10, M20, M30) in summer. This shows that soil moisture content will have a causal effect on soil internal temperature (T40) in summer. We analyze that this is partly because the soil moisture content will affect some intermediate factors such as plant root and microbial activities in summer, then these activities will affect the soil temperature. For example, M10 may influence the microbial activities in soil, and microbial activities influence soil internal temperature (T40), then M10 and T40 will be causally linked. It indicates that causality is transitive, as shown in Fig.19. The intermediate factors are recessive in this process.

IV. DISCUSSION

Soil is a complex system. Soil temperature and moisture content will be affected by many factors. We mainly study causal relationship between soil temperature and moisture content at different depths through dynamic empirical modelling.

Some causal relationships involved belong to bidirectional causality in the complex soil system. The magnitude of causal relationship is strong in one direction but weak in another direction. Note that causation is transitive (e.g., if T00 influence T10, and T10 influence T20, then T00 and T20 are causally linked, and vice versa). It is proved by real-data experiments that there is a positive correlation between predictive skill ρ and causality magnitude in CCM algorithm.

In conclusion, the causal relationship between soil temperature and moisture content system can be divided into three categories:

1) Unidirectional causality. Its performance is that the curve convergence is strong in one direction and weak in the other, such as Fig.16(a).
2) Bidirectional causality. Its performance is that the curve convergence is strong in one direction and also strong in the other, such as Fig.13(a).
3) No causality. Its performance is that the convergences in two directions are both weak, and the predictive skill $\rho \approx 1$, such as Fig.13(b).

V. CONCLUSION

In this paper, the causal relationship of soil temperature and moisture content at different depths is analyzed based on dynamic empirical modelling. We study the causal relationship and find the experimental results are not the same in summer and winter. We describe the complexity of soil temperature and moisture content system through simple projection algorithm. We demonstrate the direction and magnitude of causal relationship between soil moisture content and temperature at different depths by CCM algorithm. Besides, we describe the difference of soil system properties in summer and winter through causal research. The experimental results are consistent with the actual soil environment. The causality is described by dynamic empirical modelling rather than prior soil knowledge. The paper may provide a new idea for soil dynamics research.

REFERENCES

[1] H. Kohnke and D. P. Franzmeier, *Soil Science Simplified*. Long Grove, IL, USA: Waveland Press, 2015.
[2] C. Y. Huang and J. M. Xu, *Soil Science*. Beijing, China: China Agricultural Press, 2010.
[3] H. Ye, R. J. Beamish, S. M. Glaser, S. C. H. Grant, C.-H. Hsieh, L. J. Richards, J. T. Schnute, and G. Sugihara, "Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 13, pp. E1569–E1576, Mar. 2015, doi: 10.1073/pnas.1417063112.
[4] D. L. DeAngelis and S. Yurek, "Equation-free modeling unravels the behavior of complex ecological systems," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 13, pp. 3856–3857, Mar. 2015, doi: 10.1073/pnas.1503154112.
[5] J. Pearl and D. Mackenzie, *The Book of Why: The New Science of Cause and Effect*. New York, NY, USA: Basic Books, 2018.
[6] C.-W. Chang, M. Ushio, and C.-H. Hsieh, "Empirical dynamic modeling for beginners," *Ecol. Res.*, vol. 32, no. 6, pp. 785–796, Nov. 2017, doi: 10.1007/s11284-017-1469-9.
[7] G. Sugihara, R. May, H. Ye, C.-H. Hsieh, E. Deyle, M. Fogarty, and S. Munch, "Detecting causality in complex ecosystems," *Science*, vol. 338, no. 6106, pp. 496–500, Oct. 2012, doi: 10.1126/science.1227079.
[8] J. Runge *et al.*, "Inferring causation from time series in earth system sciences," *Nature Commun.*, vol. 10, no. 1, pp. 1–13, Dec. 2019, doi: 10.1038/s41467-019-10105-3.
[9] J. Runge, P. Nowack, M. Kretschmer, S. Flaxman, and D. Sejdinovic, "Detecting and quantifying causal associations in large nonlinear time series datasets," *Sci. Adv.*, vol. 5, no. 11, Nov. 2019, Art. no. eaau4996, doi: 10.1126/sciadv.aau4996.
[10] H. Zenil, N. A. Kiani, A. A. Zea, and J. Tegnér, "Causal deconvolution by algorithmic generative models," *Nature Mach. Intell.*, vol. 1, no. 1, pp. 58–66, Jan. 2019, doi: 10.1038/s42256-018-0005-0.
[11] Z. Shen, W.-X. Wang, Y. Fan, Z. Di, and Y.-C. Lai, "Reconstructing propagation network-ks with natural diversity and identifying hidden sources," *Nature Commun.*, vol. 5, no. 1, pp. 1–10, 2014, doi: 10.1038/ncomms5323.
[12] R. Iten, T. Metger, H. Wilming, L. del Rio, and R. Renner, "Discovering physical concepts with neural networks," *Phys. Rev. Lett.*, vol. 124, no. 1, Jan. 2020, Art. no. 010508, doi: 10.1103/PhysRevLett.124.010508.

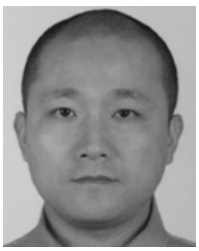
[13] E. H. van Nes, M. Scheffer, V. Brovkin, T. M. Lenton, H. Ye, E. Deyle, and G. Sugihara, "Causal feedbacks in climate change," *Nature Climate Change*, vol. 5, no. 5, pp. 445–448, May 2015, doi: 10.1038/nclimate2568.
[14] B. Schölkopf, "Causality for machine learning," 2019, *arXiv:1911.10500*. Accessed: Dec. 23, 2019. [Online]. Available: http://arxiv.org/abs/1911.10500
[15] M. Prosperi, Y. Guo, M. Sperrin, J. S. Koopman, J. S. Min, X. He, S. Rich, M. Wang, I. E. Buchan, and J. Bian, "Causal inference and counterfactual prediction in machine learning for actionable healthcare," *Nature Mach. Intell.*, vol. 2, no. 7, pp. 369–375, Jul. 2020, doi: 10.1038/s42256-020-0197-y.
[16] J. G. Richens, C. M. Lee, and S. Johri, "Improving the accuracy of medical diagnosis with causal machine learning," *Nature Commun.*, vol. 11, no. 1, pp. 1–9, Dec. 2020, doi: 10.1038/s41467-020-17419-7.
[17] H. R. Varian, "Causal inference in economics and marketing," *Proc. Nat. Acad. Sci. USA*, vol. 113, no. 27, pp. 7310–7315, Jul. 2016, doi: 10.1073/pnas.1510479113.
[18] Y. Xie, "Population heterogeneity and causal inference," *Proc. Nat. Acad. Sci. USA*, vol. 110, no. 16, pp. 6262–6268, Apr. 2013, doi: 10.1073/pnas.1303102110.
[19] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica, J. Econ. Soc.*, vol. 37, no. 3, pp. 424–438, Aug. 2069, doi: 10.2307/1912791.
[20] H. Ye and G. Sugihara, "Information leverage in interconnected ecosystems: Overcoming the curse of dimensionality," *Science*, vol. 353, no. 6302, pp. 922–925, 2016, doi: 10.1126/science.aag0863.
[21] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.
[22] G. E. Hinton, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006, doi: 10.1126/science.1127647.
[23] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, *arXiv:1803.01271*. Accessed: Apr. 19, 2018. [Online]. Available: http://arxiv.org/abs/1803.01271
[24] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent neural network regularization," 2014, *arXiv:1409.2329*. Accessed: Feb. 19, 2015. [Online]. Available: http://arxiv.org/abs/1409.2329
[25] K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017, doi: 10.1109/TNNLS.2016.2582924.
[26] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, *arXiv:1406.1078*. Accessed: Sep. 3, 2014. [Online]. Available: http://arxiv.org/abs/1406.1078
[27] D. K. Arrowsmith, C. M. Place, and C. H. Place, *An Introduction to Dynamical Systems*. Cambridge, U.K.: Cambridge Univ. Press, 1990.
[28] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick*. Berlin, Germany: Springer, 1981, pp. 366–381, doi: 10.1007/BFb0091924.
[29] E. R. Deyle, M. Fogarty, C.-H. Hsieh, L. Kaufman, A. D. MacCall, S. B. Munch, C. T. Perretti, H. Ye, and G. Sugihara, "Predicting climate effects on pacific sardine," *Proc. Nat. Acad. Sci. USA*, vol. 110, no. 16, pp. 6430–6435, Apr. 2013, doi: 10.1073/pnas.1215506110.
[30] C.-H. Hsieh, S. M. Glaser, A. J. Lucas, and G. Sugihara, "Distinguishing random environmental fluctuations from ecological catastrophes for the north pacific ocean," *Nature*, vol. 435, no. 7040, pp. 336–340, May 2005, doi: 10.1038/nature03553.
[31] H. Toju, K. G. Peay, M. Yamamichi, K. Narisawa, K. Hiruma, K. Naito, S. Fukuda, M. Ushio, S. Nakaoka, Y. Onoda, K. Yoshida, K. Schlaeppi, Y. Bai, R. Sugiura, Y. Ichihashi, K. Minamisawa, and E. T. Kiers, "Core microbiomes for sustainable agroecosystems," *Nature Plants*, vol. 4, no. 5, pp. 247–257, May 2018, doi: 10.1038/s41477-018-0139-4.
[32] M. Ushio, C.-H. Hsieh, R. Masuda, E. R. Deyle, H. Ye, C.-W. Chang, G. Sugihara, and M. Kondoh, "Fluctuating interaction network and time-varying stability of a natural fish community," *Nature*, vol. 554, no. 7692, pp. 360–363, Feb. 2018, doi: 10.1038/nature25504.
[33] S.-I. Nakayama, A. Takasuka, M. Ichinokawa, and H. Okamura, "Climate change and interspecific interactions drive species alternations between anchovy and sardine in the western north pacific: Detection of causality by convergent cross mapping," *Fisheries Oceanogr.*, vol. 27, no. 4, pp. 312–322, Jul. 2018, doi: 10.1111/fog.12254.
[34] K. Suzuki, K. Yoshida, Y. Nakanishi, and S. Fukuda, "An equation-free method reveals the ecological interaction networks within complex microbial ecosystems," *Methods Ecol. Evol.*, vol. 8, no. 12, pp. 1774–1785, Dec. 2017, doi: 10.1111/2041-210X.12814.
[35] X. Liu, G. Dur, S. Ban, Y. Sakai, S. Ohmae, and T. Morita, "Planktivorous fish predation masks anthropogenic disturbances on decadal trends in zooplankton biomass and body size structure in Lake Biwa, Japan," *Limnol. Oceanogr.*, vol. 65, no. 3, pp. 667–682, Mar. 2020, doi: 10.1002/lno.11336.
[36] R. E. White, *Principles and Practice of Soil Science: The Soil as a Natural Resource*. Hoboken, NJ, USA: Wiley, 2013.
[37] G. Sugihara and R. M. May, "Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series," *Nature*, vol. 344, no. 6268, pp. 734–741, Apr. 1990, doi: 10.1038/344734a0.
[38] G. Sugihara, "Nonlinear forecasting for the classification of natural time series," *Philos. Trans. Roy. Soc. London A, Phys. Eng. Sci.*, vol. 348, no. 1688, pp. 477–495, 2014, doi: 10.1098/rsta.1994.0106.
[39] A. T. Clark, H. Ye, F. Isbell, E. R. Deyle, J. Cowles, G. D. Tilman, and G. Sugihara, "Spatial convergent cross mapping to detect causal relationships from short time series," *Ecology*, vol. 96, no. 5, pp. 1174–1181, May 2015, doi: 10.1890/14-1479.1.

ZHIHAO CAO received the M.S. degree in computer science and technology from Shandong Agricultural University, Tai'an, in 2018. Since 2018, he has been a Lecturer with the College of Information Science and Engineering, Zaozhuang University. His research interests include artificial intelligence and complex systems.

SHAOMIN MU received the Ph.D. degree in computer science from Beijing Jiaotong University, China, in 2008. He is currently a Professor of computer science with Shandong Agricultural University, China. His primary research interests include image processing, deep learning, pattern recognition, and big data technology.

LI XU received the M.S. and Ph.D. degrees in computer science and technology from China Agricultural University, Xuzhou, in 2012 and 2018, respectively. Since 2018, she has been a Lecturer with the College of Information Science and Engineering, Zaozhuang University. Her research interests include machine learning and data mining.

MINGFENG SHAO is currently pursuing the B.E. degree with Zaozhuang University. Her research interests include artificial intelligence and complex systems.



HONGCHUN QU received the Ph.D. degree in computer science and technology from Iowa State University. Since 2019, he has been a Professor with the College of Information Science and Engineering, Zaozhuang University. His research interests include machine learning and complex systems.