



OPEN Detecting causal relationship of non-floodplain wetland hydrologic connectivity using convergent cross mapping

Sangchul Lee^{1,5}, Byeongwon Lee^{1,5}, Junga Lee^{2,5}, Jihoon Song³ & Gregory W. McCarty⁴

The hydrologic connectivity of non-floodplain wetlands (*NFWs*) with downstream water (*DW*) has gained increased importance, but connectivity via groundwater (*GW*) is largely unknown owing to the high complexity of hydrological processes and climatic seasonality. In this study, a causal inference method, convergent cross mapping (CCM), was applied to detect the hydrologic causality between upland *NFW* and *DW* through *GW*. CCM is a nonlinear inference method for detecting causal relationships among environmental variables with weak or moderate coupling in nonlinear dynamical systems. We assumed that causation would exist when the following conditions were observed: (1) the presence of two direct causal ($NFW \rightarrow GW$ and $GW \rightarrow DW$) and one indirect causal ($NFW \rightarrow DW$) relationship; (2) a nonexistent opposite causal relationship ($DW \rightarrow NFW$); (3) the two direct causations with shorter lag times relative to indirect causation; and (4) similar patterns not observed with pseudo *DW*. The water levels monitored by a well and piezometer represented *NFW* and *GW* measurements, respectively, and the *DW* was indicated by the baseflow at the outlet of the drainage area, including *NFW*. To elucidate causality, the *DW* taken at the adjacent drainage area with similar climatic seasonality was also tested as pseudo *DW*. The CCM results showed that the water flow from *NFW* to *GW* and then *DW* was only present, and any opposite flows did not exist. In addition, direct causations had shorter lag time than indirect causation, and 3-day lag time was shown between *NFW* and *DW*. Interestingly, the results with pseudo *DW* did not show any lagged interactions, indicating non-causation. These results provide the signals for the hydrologic connectivity of *NFW* and *DW* with *GW*. Therefore, this study would support the importance of *NFW* protection and management.

Non-floodplain wetlands (*NFWs*) enclosed by uplands without surface runoff outlets¹ provide hydrologic, biological, chemical, and ecological benefits to landscapes^{2,3}. Thus, understanding *NFW* hydrologic connectivity with adjacent or distant water bodies through surface runoff or subsurface flow including groundwater (*GW*) is important for demonstrating their key roles in landscapes⁴. Water mediating the transport of matter, energy, and organisms within or between elements of the hydrologic cycle refers to hydrologic connectivity⁵. The degree, type, and frequency of *NFW* connectivity (i.e., hydrologic connectivity between *NFW* and other components) differ by geomorphic characteristics⁶. For example, surface water connectivity is prevalent in the Prairie Pothole region owing to limited underground hydraulic conductivity⁶, whereas *GW* connectivity between *NFWs* and other landscape components prevails on the Coastal Plain of the Chesapeake Bay Watershed (CBW)^{7,8}.

Hydrologic connectivity between *NFWs* and *DWs* is a function of surface runoff and *GW*, and the major diver varies depending on the landscape and climatic characteristics^{1,9}. Surface runoff connectivity is often measured using in-situ observations on the Coastal Plain of the CBW^{10,11}. However, demonstrating *GW* connectivity using observational data is extremely challenging because of the inherent uncertainty in *GW* connectivity, which is characterized as a combination of nonlinear behaviors driven by interactions among climatic inputs, hydrogeologic characteristics, and human intervention^{3,12}. Uncertainty regarding *GW* connectivity is further complicated

¹Department of Environmental Engineering, University of Seoul, Dongdaemun-gu, Seoul 02504, Republic of Korea. ²Division of Environmental Science & Ecological Engineering, College of Life Sciences & Biotechnology, Korea University, Seoul 02841, Republic of Korea. ³Ojeong Resilience Institute, Korea University, Seoul 02841, Republic of Korea. ⁴USDA-ARS, Hydrology and Remote Sensing Laboratory, Beltsville, MD 20705, USA. ⁵These authors contributed equally: Sangchul Lee, Byeongwon Lee and Junga Lee. ✉email: sangchul.lee84@gmail.com; greg.mccarty@usda.gov

by climatic seasonality (e.g., evapotranspiration [ET] and precipitation [P]). Climatic seasonality often drives landscape hydrology, leading to similar temporal dynamics of landscape components⁷.

On the coastal plain of the CBW, the *NFW* water budget increases with *P* and *GW* inflow but decreases with *ET* and *GW* outflow. When the water table is higher or lower than the bottom of *NFW* during the wet and dry seasons, *GW* inflow to and outflow from *NFWs* occur, respectively¹³. Surface inflow and outflow also affect *NFW* dynamics; however, the impact of surface runoff is mostly observed during heavy rainfall events⁷. Thus, *NFW* connectivity via surface water has been determined in this region after rainfall events or wet seasons^{10,11}. In addition, *NFW* connectivity with *DW* via *GW* has been reported in this region, owing to subsurface permeable conditions that lead to the strong interactions between *NFW* and *GW*^{14–17}. However, *GW* connectivity is speculated based on regional characteristics and the modeling approach; strong evidence in this regard is still lacking in this region.

The major driving forces of the regional water budget on the CBW coastal plain are *P* and *ET*. When *P* is greater or lower than *ET*, the water budget of the landscape components, including *NFWs* increases or decreases, respectively. Landscape water storage is strongly dependent on wetlands in this region, which are characterized by low-gradient topography and poorly drained soils¹⁸. These conditions collectively lead to similar dynamics of *NFW* and *DW*, but their differences are observed after rainfall events for a short period⁷. To demonstrate *GW* connectivity between *NFW* and *DW*, the three components should exhibit a chain of causal relationships (represented as “cause” → “effect”) with different lag times (Fig. 1): *NFW* → *GW* with Lag_a , *GW* → *DW* with Lag_b , and *NFW* → *DW* with Lag_c . Any changes in *NFWs* will be reflected in *GW* and later in *DWs* along a hydrologic gradient due to lag times. A chain of causal relationships with *NFW* → *GW* and *GW* → *DW* have a direct causal relationship, while *NFW* indirectly affects *DW*. Thus, Lag_c is equal to or greater than the summation of Lag_a and Lag_b (i.e., $Lag_c \geq Lag_a + Lag_b$). In contrast, the responses of *NFW* and *DW* to climatic seasonality might be simultaneous (i.e., no lag time) when the effects in both *NFW* and *DW* are linked to a shared cause (i.e., climatic seasonality). The causal relationships between *NFW*, *GW*, and *DW* with different lag times can be indicative of *GW* connectivity. Thus, identifying *GW* connectivity requires a metric that detects the causality between two variables with time-delayed interactions within nonlinear systems.

Convergent cross mapping (CCM) is a novel approach for detecting causality in nonlinear systems^{20,21}. This causal inference method is unique because it can detect causal relationships between two variables with weak or moderate coupling. CCM has been successfully tested to demonstrate causal relationships between two time-series datasets of environmental systems: temperature impacts on greenhouse gases²², invading species and soil nitrate²³, between summer precipitation and aboveground biomass²³, population dynamics of anchovies and sardines²⁴, soil moisture impacts on precipitation²⁵, hydrologic connectivity between two reservoirs²⁶, and interactions between hydrologic and climatic variables²⁷, and surface and groundwater relationship²⁸. Extended CCM has also been used for distinguishing between uni- and bi-directional flow and detecting causal chains between entities with varying degrees of lagged behaviors²¹. Extended CCM can quantify whether two entities causally interact with each other (i.e., bi-directional) or whether only one entity affects the other entity (i.e., unidirectional)²¹.

In this study, we employed extended CCM to demonstrate the hydrologic connectivity of *NFW* with *DW* through *GW* and quantify the time delay in this causal relationships in the Greensboro watershed (GBW) within CBW in the USA⁷. We speculated that if *NFW* has any hydrologic connectivity with *DW* via *GW*, the following would be observed from CCM analysis: (1) two direct causality (*NFW* → *GW* and *GW* → *DW*) and one indirect causality (*NFW* → *DW*) would be present; (2) the opposite direction of causality would not appear, and (3) direct causality would have shorter lag time than indirect causality. Because of the high uncertainty in the causation between *GW* and *DW*, we considered pseudo *DW* (*DW_{adj}*) adjacent to the *DW* located at the outlet of

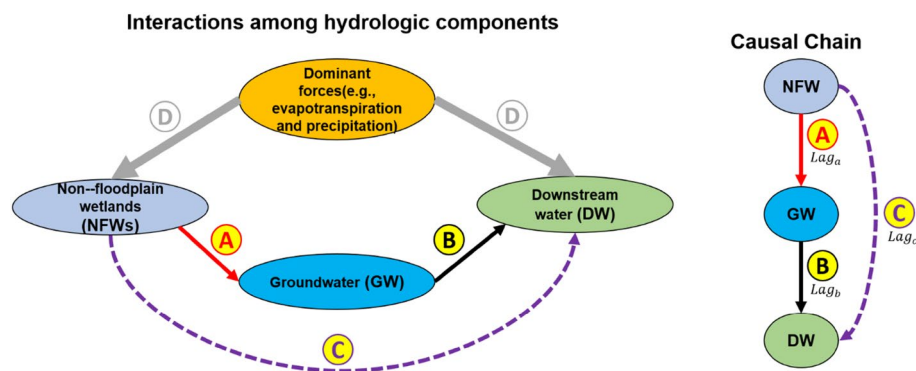


Figure 1. Schematic diagram showing interactions among *NFWs*, *GW*, *DW*, and climatic seasonality (left) and their causal chain (right). The figure is partially adopted from Pyzoha et al.¹⁹. “A”, “B”, and “C” indicate flow from *NFW* to *GW*, *GW* to *DW*, and *NFW* to *DW*, respectively. D is a causal variable shared by *NFW* and *DW*, resulting in similar patterns. In the causal chain, Lag_a , Lag_b , and Lag_c stand for the lag for flow from *NFW* to *GW*, from *GW* to *DW*, and from *NFW* to *DW*, respectively. The figure was generated by the MS Office PowerPoint.

the watershed, including the NFW (DW_{org}). It was also assumed that causality would be observed between NFW and DW_{org} but not between NFW and DW_{adj} .

Materials and methods

Study area and data

The NFW is located within the GBW, which is the drainage area of the U.S. Geological Survey (USGS) gauge station #01491000, on the coastal plain of the CBW (Fig. 2). The CBW is divided into 11 hydrogeomorphic regions (HGMRs) based on the rock type and physiographic province¹⁷. The study area is included in the coastal plain upland (CPU) with high precipitation infiltration into shallow aquifer due to well-drained soils and flat topography¹⁷. The CPU is also underlain with unconsolidated sediments with high permeability, leading to large groundwater discharge to streams¹⁷.

The sizes of NFW and its drainage area are 1 ha and 3.1 ha, respectively; and the drainage area is dominated by forest⁷. The nearest streams indicated by the USGS high-resolution National Hydrography Dataset (NHD) is 0.1 km from NFW (Fig. 2). The “Wetland” component of the U.S. Department of Agriculture Conservation Effects Assessment Project (CEAP-Wetlands) implemented a well and piezometer to explore the surface water level and groundwater level of NFW in this region, respectively²⁹. A PVC pipe with a 2.54 cm diameter was utilized to construct a well and piezometer, with well screens being positioned over the whole length of the wells and the lower 30 cm of the piezometers. The well was utilized to directly measure the height of the water column, whereas the piezometer was employed to assess the pressure of the groundwater exerted by the water column. The well was equipped with pressure transducers (Campbell Scientific CS451, Campbell Scientific, Logan, Utah, United States), while the piezometer was physically linked to the data logger (Campbell Scientific CR1000). To see the interactions between the surface and groundwater, a well and piezometer were installed side by side (Fig. S1 of the Supplementary Material) since different soil conditions under the bottom of NFW might not correctly capture their interactions³⁰.

The studied NFW was monitored from January 2016 to December 2019 (Table 1). The water levels were continuously collected by a well and piezometer at every 15-min, respectively (Fig. 2). The well and piezometer in the NFW was installed to 0.9 and 3.0 m below the wetland bottom, respectively, and they were spatially close each other³⁰. Following similar previous studies^{7,30}, this study assumed the water levels monitored by a well and piezometer indicated NFW and GW , respectively.

Baseflow from streamflow collected from the USGS gauge station was used to represent DW_{org} , and USGS gauge station #01491500 was also prepared for DW_{adj} (Fig. 2). Both DW_{org} and DW_{adj} were spatially adjacent, and thus they similarly responded to regional climatic seasonality. DW_{org} was hydrologically downstream of NFW based on the watershed boundary. Baseflow separation was calculated using the EcoHydrology package³¹ in the R programming environment. The digital filter separation described by Nathan and McMahon³² was used in this package, and the default settings were applied in our study.

15-min data were aggregated into daily values for the CCM analysis. The NFW measurements included 177 missing samples (January 11th–July 6th, 2017), and data incompleteness could cause errors in the CCM analysis. The missing samples were filled with simulations from the process-based model developed to predict the water level of NFW ³³. The process-based model was modified to simulate the hydrology of NFW , and the model

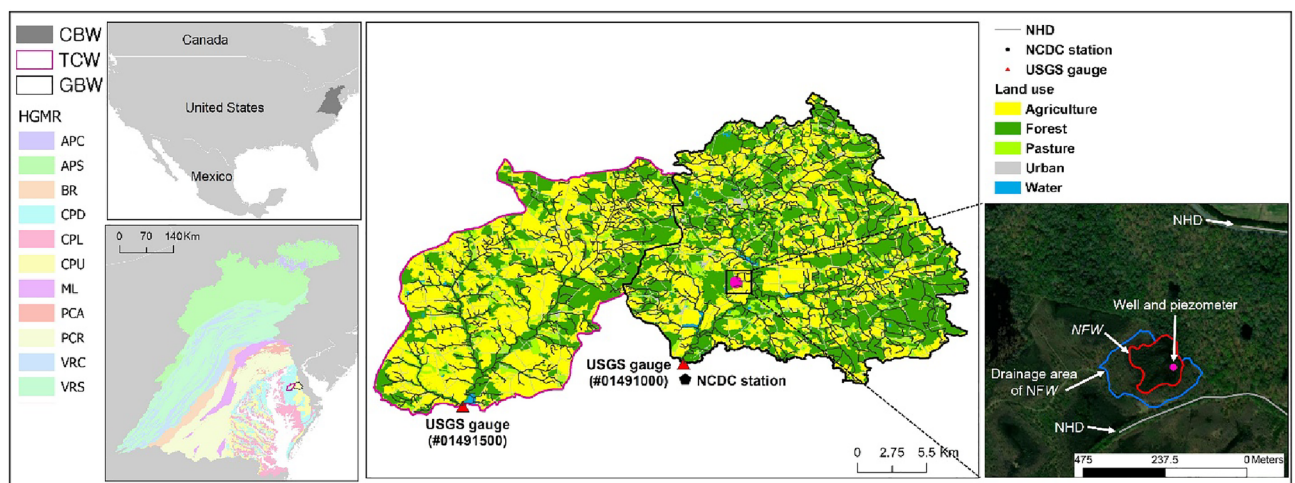


Figure 2. The spatial location of the studied wetland. CBW indicates the Chesapeake Bay watershed. GBW (Greensboro watershed) and TCW (Tuckahoe Creek watershed) are the drainage areas of the USGS gauge stations #01491000 and #01491500, respectively. NHD stands for the National Hydrography Dataset. The location of well and piezometer was further zoomed-in in Fig. S1 of the Supplementary Material and the data source used in this figure is listed in Table S1 of the Supplementary Material. The description of the hydrogeomorphic region (HGMR) is available in Table S2 of the Supplementary Material. The map was generated by ArcMap 10.7.

Attributes	Descriptions			
Location	39.053 (Latitude), −75.748 (Longitude)			
Monitoring period	January 1st, 2016 to December 31st, 2019			
Monitoring range	0.8 and 3 m below the wetland bottom (well and piezometer, respectively)			
Surface level (m)	min	max	std	Avg
	−0.90	0.44	0.31	0.14
Groundwater level (m)	min	max	std	Avg
	−1.38	0.32	0.37	−0.07
Distance to <i>DW</i> (km)	7.1			
Monitoring samples (day)	<i>NFW</i> : 1259, <i>GW</i> : 1461			
Missing samples (day)	<i>NFW</i> : 177, <i>GW</i> : 0			
Area (ha)	1.0			

Table 1. Descriptions of the studied *NFW* and its monitoring. The location indicates the specific points of wells and piezometers installed. Non-floodplain wetland (*NFW*) and groundwater (*GW*) are defined as water levels measured by wells and piezometers, respectively, in this study. *DW* indicates downstream water.

successfully predicted the daily water level of the *NFW* studied with high accuracy with the R^2 value of 0.84–0.88 and further detailed results are provided in Junyu et al.³³.

Convergent cross mapping

In this study, we introduced CCM, a method suggested by Sugihara et al.²⁰ and extended by Ye et al.²¹, to detect causal relationships between variables of interest. CCM is a handy and robust numerical tool to determine causal influence when the time series of two variables are available. The theoretical basis of this approach comes from Takens' theorem: when “*x*” (cause) influences “*y*” (effect) in dynamical systems, the value of “*x*” can be retrieved from the value of “*y*”. One critical process is to determine whether the accuracy of the reconstruction of “*x*” from “*y*” increases as the number of *L* (library vectors) of “*y*” increases. As the number of *L* means the number of different time points used for the reconstruction. In other words, it tests whether the inclusion of “*y*” values from longer periods reconstructs “*x*” better. The degree of how accurately “*y*” estimates “*x*”, called cross map skill (*p*), is determined based on Pearson's correlation between observed and estimated “*x*”²⁰. Most importantly, this test allows screening of spurious correlational relationships. If “*x*” and “*y*” have a spurious correlational relationship, an increase in *L* does not result in a more accurate reconstruction of “*x*”, because the correlation is only due to the short-time synchronization of the two variables. In CCM, this relationship between *x* and *y* is represented as “*y* xmap *x*” that means *x* is estimated from *y*. In this study, we stated “*y* (effect) xmap *x* (cause)” as “*x* → (affects) *y*” to better indicate the relationship between cause and effect. This study also tested the significance of CCM results (*p*-value < 0.01) to identify spurious causation using the method introduced in Bonotto et al.²⁸ and Ye et al.³⁴. The test included assessing whether the cross-map skill at full library size is significantly greater than the highest lagged cross-correlation and surrogate time series. The *p*-value was computed as below:

$$p\text{-value} = (n + 1)/(k + 1), \quad (1)$$

where *k* is the total number of surrogates and *n* is the number of replicates with greater cross-map skill than the actual value.

The extension and enhancement of CCM by Ye et al.³⁴ solved one of the weaknesses of the original version: overwhelming influence from “*x*” on “*y*” may lead to successful CCM in both directions (meaning that “*y*” also affects “*x*”), although this is not the case. This weakness, called ‘generalized synchrony’, occurs when “*y*” is almost only affected by “*x*”, and thus, “*x*” and “*y*” act as one system. To address this issue, Ye et al.³⁴ considered different lag times. Causal relationships signify that there exists a chronological order between the two variables. Therefore, if the causal relationship between “*x*” and “*y*” is unidirectional (here only “*x*” affects “*y*”), the lag time found in CCM recreating “*x*” from “*y*” will be negative, while the time lag found in CCM recreating “*y*” from “*x*” will be positive. The optimal time lag is decided when the cross map skill is largest. The negative optimal lag time of “*x* (cause)” → “*y* (effect)” indicates the response of “*y*” to any changes in “*x*” will appear after the optimal lag time. The zero optimal lag time of “*x* → *y*” denotes “*y*” promptly responds to “*x*” without any lag time, and the positive optimal lag time of “*x* → *y*” means “*x* (cause)” responds to “*y* (effect)” with the lag time, indicating illogical causation.

We used the rEDM package developed by Ye et al.³⁴ to implement CCM processes in R. The processes involved the reconstruction of individual system states using a time series to generate the joint state, for the estimation of one variable from another. Individual system states were reconstructed by the time-delay embedding approach, which represents the delay coordinates of each system state. The optimal embedding dimension (*E*) of each system state was calculated through the rEDM package in R³⁴.

Results

Temporal dynamics of non-floodplain wetland (NFW), groundwater (GW), and downstream water (DW)

The daily time series of *NFW*, *GW*, DW_{org} , and DW_{adj} are shown in Fig. 3a,b. Overall, the temporal dynamics of *NFW*, *GW*, DW_{org} , and DW_{adj} were highly similar in response to the seasonal trends of this area, with high and low water balance during the winter and summer seasons, respectively. *NFW* and *GW* ranged from -0.9 to 0.4 m and from -1.4 to 0.3 m, respectively, while the variation ranges of DW_{org} and DW_{adj} were $2\text{--}403$ (m^3/s) and $15\text{--}195$ (m^3/s), respectively. DW_{org} and DW_{adj} showed similar temporal dynamics (Fig. 3b). Monthly variations in ET were high during summer months (June, July, and August) and low during winter months (December, January, and February) and the monthly pattern of precipitation was overall uniform over the course of the year, indicating seasonality in this region (Fig. 3c,d).

NFW and *GW* dynamics did not comply with several peaks of *DW* owing to the vertical limits in the monitoring range of *NFW* and *GW* (dotted vertical purple arrows in Fig. 3b). When the water storage of *NFW* was filled by heavy rainfall, the fill-spill dynamics of *NFW* frequently occurred^{10,11}. Regarding the configuration of a well and piezometer (Fig. S2 of the Supplementary Material) the maximum upper water level of *NFW* and *GW* is the same as the depth of the *NFW* while the monitoring range of *DW* had no limit, leading to different dynamics

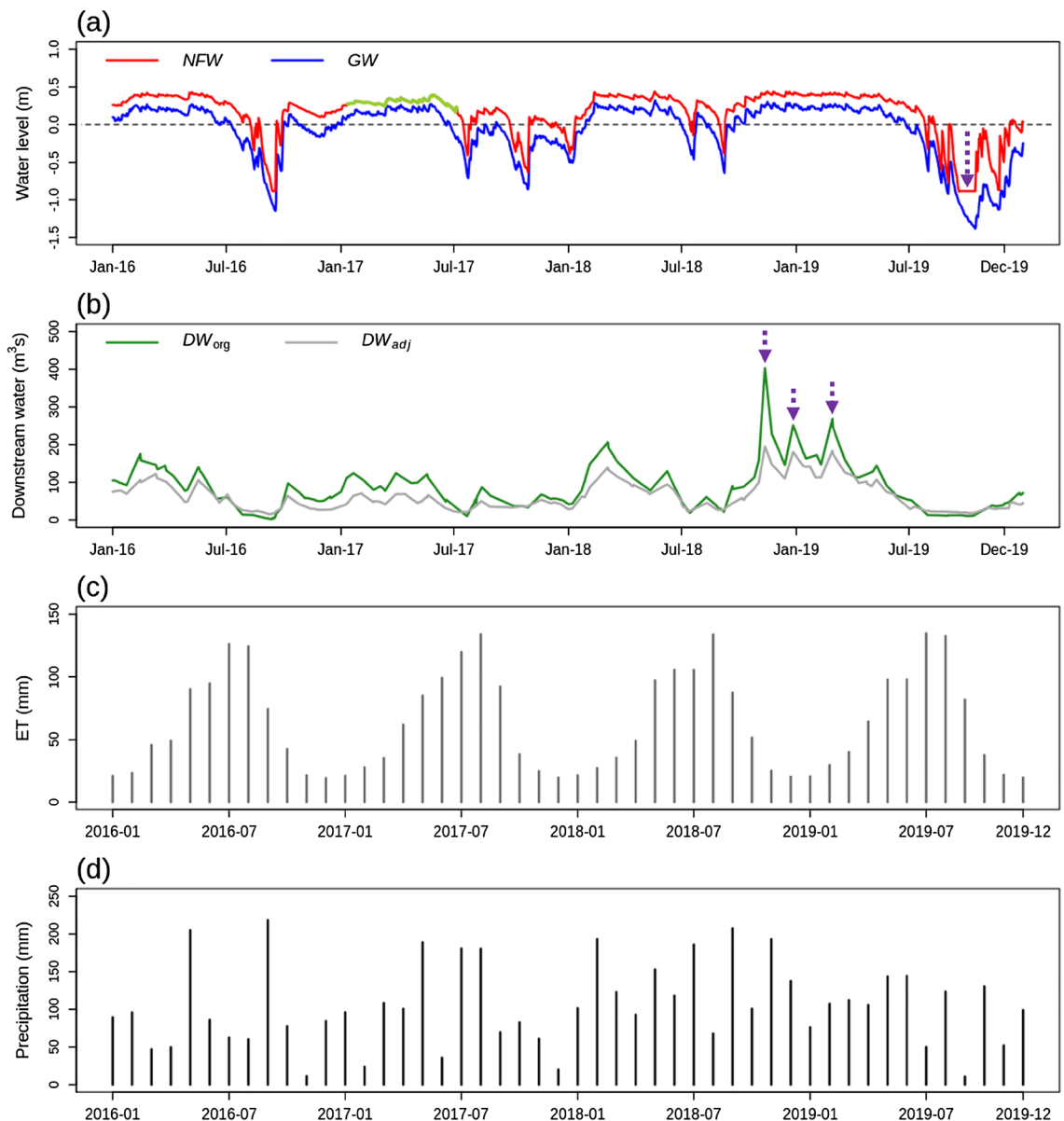


Figure 3. (a) Daily time series of non-floodplain wetlands (*NFW*) and groundwater (*GW*), (b) daily time series of two downstream waters (*DW*s), (c) monthly time series of evapotranspiration (ET), and (d) monthly time series of precipitation. The yellow-green line in (a) indicates the predicted *NFW* by the process-based model³³. DW_{org} and DW_{adj} are baseflow derived from streamflow measured at USGS gauge stations #01491000 and #01491500, respectively. The figure was generated by the R 3.6.1 program.

between NFW/GW and DW . The monitoring lower limit also caused the flat level of NFW during the summer of 2019 (dotted vertical purple arrow in Fig. 3a). In addition, DW is the summation of intra-watershed processes and thus the behavior of DW might be different from NFW and GW that are the tiny landscape component.

Embedding dimension and nonlinearity

To apply the CCM method, the embedding dimension (E) should be determined. The embedding dimension (E) is the number of time steps used for the prediction. Following previous studies, the optimal E values for NFW , GW , and DW were computed using a simplex projection (Table 2)^{24,25}. Nonlinearity was identified by observing the best prediction when different combinations of the degree of nonlinearity (θ) and embedding dimension (E) were tested. When the best prediction was shown with $\theta > 0$, the time-series variable was nonlinear. The nonlinearity was assessed by S-map with the embedding dimension (E , 1, 2, ..., 10) and the degree of nonlinearity (θ , 0, 0.25, 0.5, ..., 0).

Differentiation of causation from non-causal relationships

In the CCM method, we attempted to differentiate causation from non-causal correlations between observed and estimated variables: $NFW \rightarrow GW$, $GW \rightarrow DW$, and $NFW \rightarrow DW$, using cross map skill. Figure 4 shows the cross map skill (p) between the observed and estimated variables with different library lengths (L). Interestingly, overall relationships had high cross map skill (p) with a longer library length (L) with the significant results (p -value < 0.01). The causality was detected at $NFW \rightarrow GW$, which was driven by the vertical water flow from surface to groundwater. The high cross map skill (p) of $GW \rightarrow NFW$ was likely due to the seasonal flow from upland to NFW via GW in this region⁷. The causality results between precipitation and hydrologic variables (e.g., NFW , GW , and DW_{org}) showed precipitation exhibited a causal relationship with three hydrologic variables (p -value < 0.01 , Fig. S3 of the Supplementary Material). This finding suggested that the seasonal lateral movement of GW to NFW could be substantiated by the causative link between precipitation and these hydrologic variables.

However, the significant test showed that the cross map skill of $GW \rightarrow NFW$ at full library size was lower than the highest lagged cross-correlation, indicating non-causation (Fig. 4a). The cross map skill (p) of $GW \rightarrow DW_{org}$ and DW_{adj} was high and remained unchanged regardless of the length of the library (L). DW_{org} and DW_{adj} represents the combination results of hydrological landscape components. A tiny component (GW) could have minimal impacts on DW_{org} and DW_{adj} and thus the cross map skill (p) of $GW \rightarrow DW_{org}$ and DW_{adj} did not consistently increase with an increase in the length of the library (L). In contrast, DW_{org} and DW_{adj} complied with seasonal dynamics indicated by dominant driving forces (e.g., P and ET), leading to a relatively high value of the cross map skill (p) of DW_{org} and $DW_{adj} \rightarrow NFW/GW$ ⁷. Except for $GW \rightarrow NFW$, all interactions were statistically significant.

Detecting causality in the hydrological connectivity between NFW , GW , and DW

First, we applied extended CCM method to further identify the true causality among NFW , GW , DW_{org} and DW_{adj} . The true causations in Fig. 4 were further explored. Figure 5a shows that the optimal cross map lags of $NFW \rightarrow GW$ was not distinguishable from zero lag time. This was likely because the spatial configuration of the well and piezometer might lead to synchronization between NFW and GW . In addition, daily measurements might not capture lagged responses of GW to NFW .

The optimal lag time of $GW \rightarrow DW_{org}$ was observed at a negative one day (Fig. 5b), indicating that changes in GW would be reflected in DW_{org} one day later. Non-causation with the positive optimal lag time was found in the opposite direction ($DW_{org} \rightarrow GW$). The optimal lag time of $NFW \rightarrow DW_{org}$ was negative 3 days, and the opposite direction was zero (Fig. 5c). The results of $NFW \rightarrow DW_{org}$ informed that after 3 days, any changes in NFW would affect DW_{org} . Regarding the lag time in the causal chain of NFW , GW , and DW_{org} , the CCM results represented changes in NFW caused changes in GW and then subsequently affected DW_{org} . The lag times of among NFW , GW , and DW_{org} agreed with our assumption as the sum of the lag times of two direct causal links ($NFW \rightarrow GW$ and $GW \rightarrow DW_{org}$) was lower than the indirect causal link (the optimal lag of $NFW \rightarrow DW_{org}$). Moreover, the two lagged interactions ($GW \rightarrow DW_{org}$ and $NFW \rightarrow DW_{org}$) were found to be statistically significant (p -value < 0.01). Regarding the optimal lag times, the significance of two causal links (GW at $t \rightarrow DW_{org}$ at $t + 1$ and NFW at $t \rightarrow DW_{org}$ at $t + 3$, where t represents the time) were assessed using the method proposed by Bonotto et al.²⁸ and Ye et al.³⁴. The results from extended CCM well demonstrated the causal chain from NFW to DW_{org} via GW with different lag times.

In the case of $GW \rightarrow DW_{adj}$, the optimal lag time was a positive one day, representing that DW_{adj} responded to GW although any changes did not take place at GW (Fig. 5d). The optimal lag time of $DW_{adj} \rightarrow GW$ with the positive optimal lag time also showed non-causation. Interestingly, a negative one day was the optimal lag time of $NFW \rightarrow DW_{adj}$, indicating the changes in NFW will be reflected in DW_{adj} after one day. Regarding a transitive causal chain from NFW to DW through GW , the negative one-day optimal lag time of $NFW \rightarrow DW_{adj}$ could

Attributes	E	Nonlinearity
NFW	3	$\theta = 3.0$, $E = 4$
GW	2	$\theta = 4.0$, $E = 3$
DW_{org}	3	$\theta = 0.75$, $E = 3$
DW_{adj}	2	$\theta = 8$, $E = 10$

Table 2. The optimal embedding dimensions (E) and degree of nonlinearity (θ) of time-series observations.

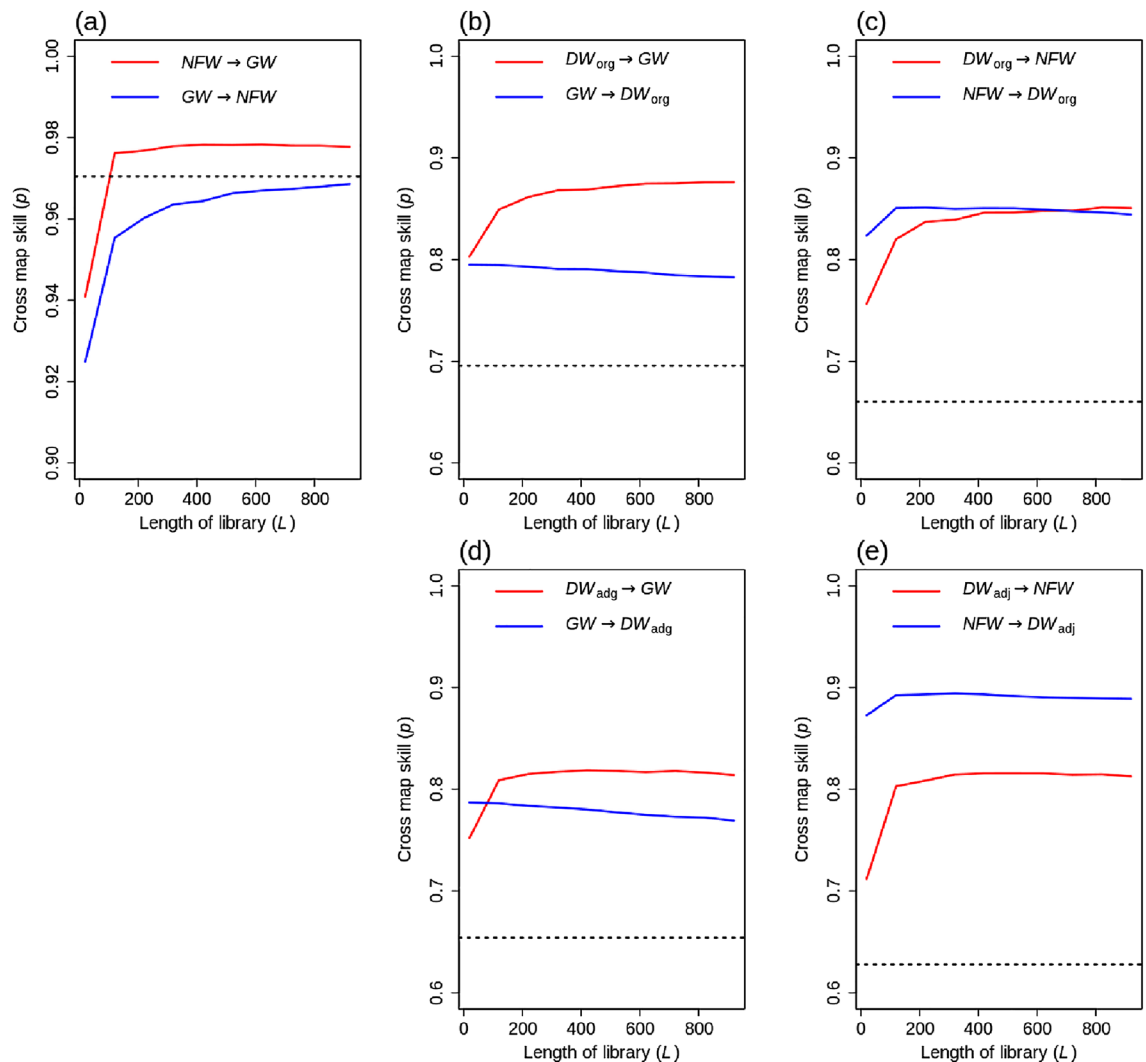


Figure 4. Cross map skill (p) of observed and estimated values as a function of the length of the library (L): (a) NFW and GW , (b) GW and DW_{org} , (c) NFW and DW_{org} , (d) GW and DW_{adj} , and (e) NFW and DW_{adj} . The dotted horizontal line is the highest lagged cross-correlation. $x \rightarrow y$ indicates x affects y . NFW and GW indicate non-floodplain wetland and groundwater, respectively. DW_{org} and DW_{adj} are baseflow derived from streamflow measured at USGS gauge stations #01491000 and #01491500, respectively. The figure was generated by the R 3.6.1 program.

not support the GW hydrologic connectivity between NFW and DW_{adj} because GW was not causal to DW_{adj} (Fig. 5e). Therefore, the extended CCM analysis showed that the evidence of causal relationships in a transitive causal chain from NFW to DW through GW was only found at DW_{org} and not at DW_{adj} based on the lagged responses among them.

Discussions and limitations

In this study, CCM was used to find causal relationships in nonlinear dynamical systems for observables with weak or moderate coupling. Previous studies explored various causal relationships between two entities in environmental systems using CCM^{22–28}. Among them, different causal methods including CCM were compared for the hydrologic variable^{26–28}. A study by Bonotto et al.²⁸ tested the impacts of data seasonality, sampling frequency, and long-term trends on the performance of CCM. Following previous studies, this study could provide additional insight on the use of CCM on the causality between one small casual variable (i.e., NFW) and an aggregated affected-variable (i.e., DW). NFW is one of water storages that drains to DW known as the summation of intra-watershed processes, and thus NFW might have trivial impacts on DW . To partially address the uncertainty on the causal impacts of NFW on DW , this study introduced pseudo DW to demonstrate the reliability of CCM results. This method could offer the potential way of using CCM to see causality between a small causal variable and the aggregated-affected variable. However, the observed causal relationship between NFW and DW was uncertain although this study adopted pseudo DW to partially address this issue. Extensive observations along the hydraulic gradient from NFW to DW might offer reliable evidence of causation, but this is

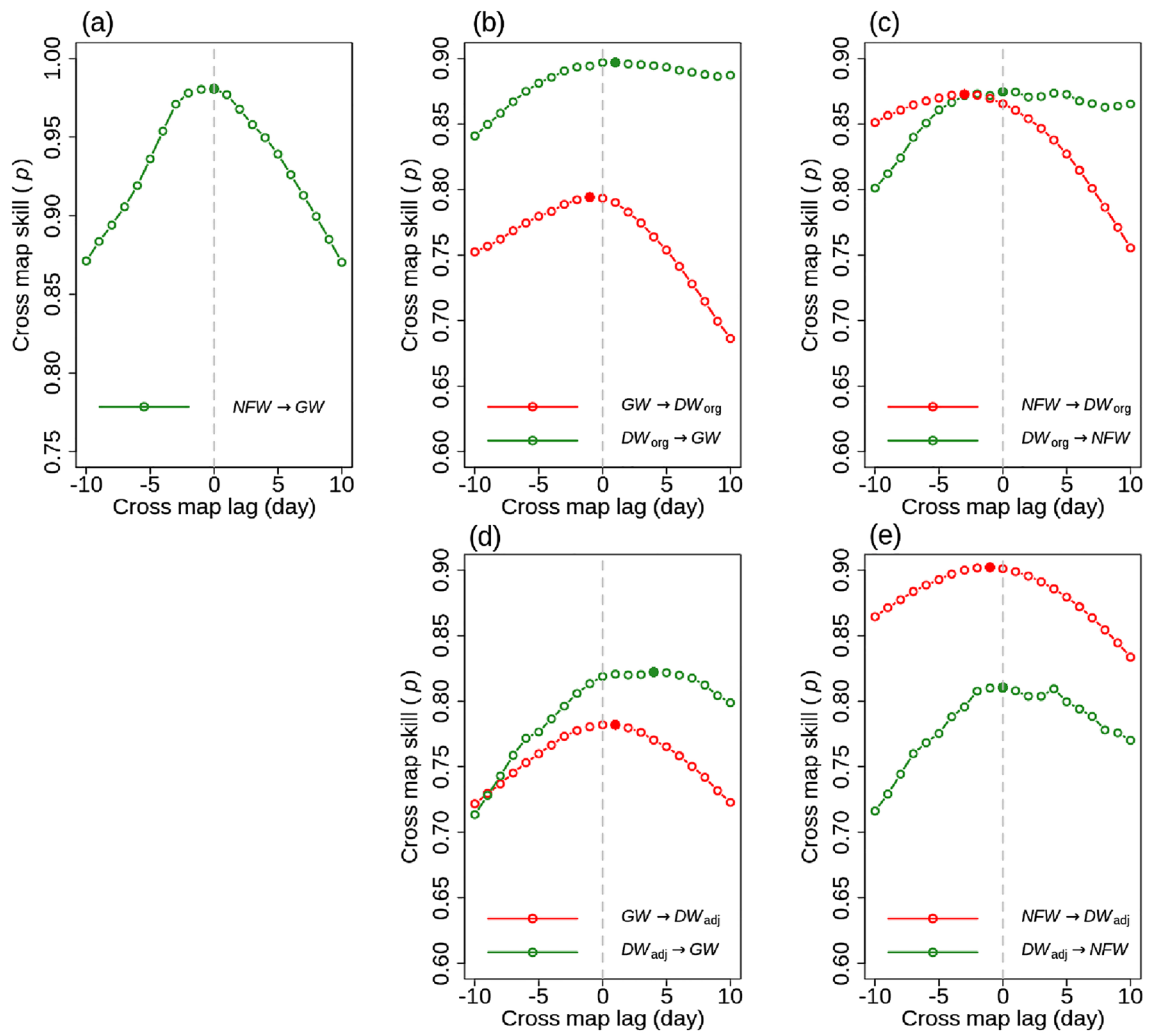


Figure 5. Cross map skill (p) with cross map lag between: (a) *NFW* and *GW*, (b) *GW* and *DW_{org}*, (c) *NFW* and *DW_{org}*, (d) *GW* and *DW_{adj}*, and (e) *GW* and *DW_{adj}*. $x \rightarrow y$ indicates x affects y . *NFW* and *GW* indicate non-floodplain wetland and groundwater, respectively. *DW_{org}* and *DW_{adj}* are baseflow derived from streamflow measured at USGS gauge stations #01491000 and #01491500, respectively. The figure was generated by the R 3.6.1 program.

challenging. To achieve dependable causality, efforts to install multiple observations along a hydrologic gradient between *NFW* and *DW* would be critical.

This study used daily time series of *NFW*, *GW* and *DW* mainly due to missing data in *NFW* (see the green line in Fig. 3a). In-situ observational data from a well and piezometer inevitably includes missing data due to uncontrollable environment conditions. The CCM results with daily time series represented the synchronized behaviors between *NFW* and *GW* (Fig. 5a), but sub-daily time series might show the lagged interaction between *NFW* and *GW*. The CCM analysis could not be performed with missing data. Accordingly, 15-min monitoring data were converted into daily data to replace missing data by the simulations from a process-based model that demonstrated decent performance measures in this region³³. The process-based model only simulated daily dynamics of *NFW*³³. Sub-daily observations data are recommended to see lagged responses of *GW* to *NFW* since the CCM results were sensitive to the sampling frequency²⁸.

To test the influence of sampling frequency on CCM results, this study applied a conversion process that transformed daily data into 3-day and 7-day intervals through the averaging of daily values (Fig. 6). Despite variations in sampling frequency, the overall trends remained consistent. The highest lagged cross-correlation, represented by the dotted horizontal line, tended to increase as the temporal frequency decreased from 1 to 7 days. Notably, the causal link between *NFW* and *GW* exhibited significance with one-day data but did not show significance with 7-day data. This discrepancy was attributed to the lower cross-map skill at the full library size, which fell below the highest lagged cross-correlation (Fig. 6f). This observation suggests that the causal interaction between *NFW* and *GW* might not be effectively captured by CCM when utilizing 7-day data, as their interactions are subtle and low-frequency data fails to depict their causal behaviors. The conversion of data from higher to lower frequencies likely resulted in smoother data patterns, allowing for the generalization of subtle behaviors observed in individual *NFW*s.

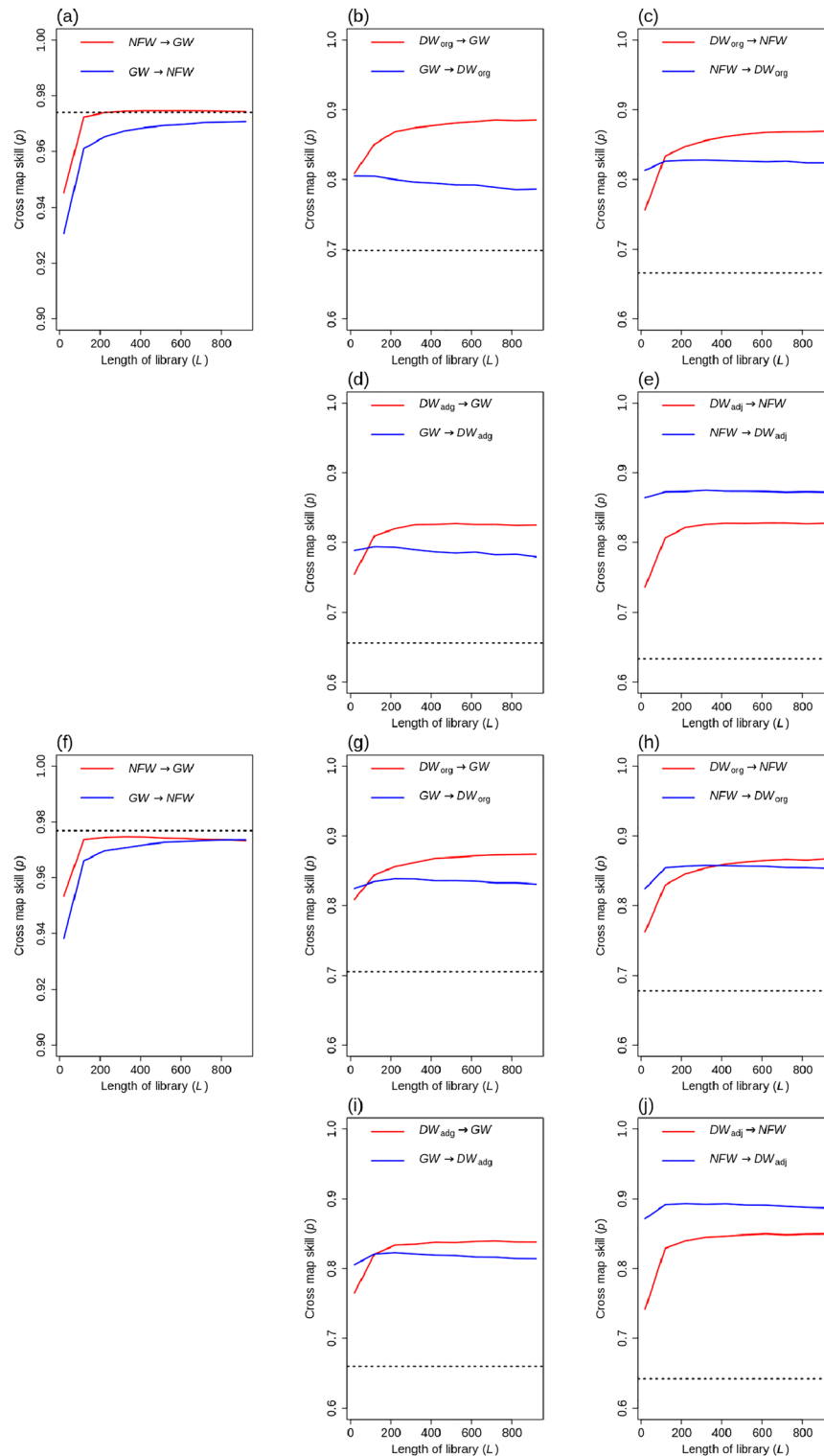


Figure 6. Cross map skill (p) of observed and estimated values as a function of the length of the library (L) with 3-day (a–e) and 7-day (f–j) data. The figure was generated by the R 3.6.1 program.

The spatial configuration of a well and piezometer also likely led to synchronization between *NFW* and *GW* (Fig. 5a). If a piezometer was distant to a well, the lagged response of *GW* to *NFW* might be observed. However, the subsurface soil characteristic greatly affected the vertical water transport from *NFW* to *GW*³⁰. When groundwater flow direction was not clear, measurements from a piezometer might not be associated with those from a well away from a piezometer. Thus, implementing a piezometer at the spot distant to a well is challenging.

Although *NFW* and *GW* were monitored at the same spot with different vertical ranges, that practices might be the best option to see hydrologic interactions of *NFW* with sub-surface systems.

Using pseudo *DW* (represented as DW_{adj}), this study showed a transitive causal chain from *NFW* to *DW* via *GW* in this region. The two direct causations ($NFW \rightarrow GW$ and $GW \rightarrow DW_{org}$) had a shorter lag time than one indirect causation ($NFW \rightarrow DW_{org}$), but the one direct causation ($GW \rightarrow DW_{org}$) showed a lower correlation than indirect one ($NFW \rightarrow DW_{org}$). Extended CCM method emphasized a direct causation with shorter lag times and stronger correlation than an indirection one²¹. However, our results were not consistent with the previous study²¹. It was speculated that: (1) a lower correlation in direction causation was likely due to the monitoring spot of *GW* spatially far away from *DW*; and (2) the two entities (*NFW* and DW_{org}) in indirect causality were directly exposure to driving forces and their behaviors might have great similarity while *GW* indirectly affected by driving forces could have less similarity.

In this region, the nutrient transport time via *GW* ranged from years to decades and this time varied by subsurface conditions³⁵. Our results estimated a 3-day lag time from *NFW* to *DW* via *GW*, which is not in agreement with a previous study³⁵. This could be explained by “old” water in long-term groundwater storage^{36,37}. The streamflow is mainly comprised of baseflow derived from long-term groundwater storage (“old” water), and the “new” precipitation has strong impacts on streamflow at storm events³⁶. The water vertical transport from *NFW*s to *GW* is the pressure head of the groundwater storage, sequentially pushing the old water from upgradient to downstream and eventually releasing “old” water closer to streamflow. As groundwater storage is influenced by various landscape components, the analysis of a 3-day lag time using only the concept of “old” water may not be adequately comprehensive. The cumulative effects of multiple landscape elements on downstream water through groundwater could lead to an extended lag time for indirect causation, compared to the total of the lag times for two direct causations. More accurate estimation of lag time would be possible using tracers³⁸ and extensive monitoring stations. This study is the first attempt to corroborate *GW* hydrologic connectivity between *NFW* and *DW* using CCM with available monitoring data and showed causal signals. This finding would provide insight on the role and *NFW* at a watershed. As emphasized by Bonotto et al.²⁸, environmental monitoring data are often limited and thus the interpretation of CCM results should be performed with detailed observations.

Conclusions

Our study applied a causal inference method (CCM) to detect causality between *NFW* and *DW* through *GW* on the Coastal Plain of the CBW. The hydrologic connectivity of *NFW* with *DW* via surface runoff was often observed, but demonstrating the connectivity via *GW* is difficult because strong climatic seasonality causes all landscape components have similar behaviors. CCM could detect causality among variables in nonlinear dynamical systems using time-series observations. Using CCM, we examined the causal relationship between *NFW* and *DW* via *GW* using daily time series. The CCM results showed a transitive causal chain from *NFW* to *DW* through *GW* with shorter lag of direct causation ($NFW \rightarrow GW$ and $GW \rightarrow DW$) relative to indirect causation ($NFW \rightarrow DW$). However, this causal chain was not observed with pseudo *DW*. These results support the notion that *NFW* is hydrologically connected with *DW* through *GW*. These findings emphasize the important role and benefits of *NFW* in landscape hydrology.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Received: 3 March 2023; Accepted: 3 October 2023

Published online: 11 October 2023

References

1. Lane, C. R., Leibowitz, S. G., Autrey, B. C., LeDuc, S. D. & Alexander, L. C. Hydrological, physical, and chemical functions and connectivity of non-floodplain wetlands to downstream waters: A review. *J. Am. Water Resour. Assoc.* **54**, 346–371 (2018).
2. Lane, C. R. *et al.* Vulnerable waters are essential to watershed resilience. *Ecosystems*. <https://doi.org/10.1007/S10021-021-00737-2/FIGURES/8> (2022).
3. Cohen, M. J. *et al.* Do geographically isolated wetlands influence landscape functions? *Proc. Natl. Acad. Sci. U.S.A.* **113**, 1978–1986 (2016).
4. Hosen, J. D., Armstrong, A. W. & Palmer, M. A. Dissolved organic matter variations in coastal plain wetland watersheds: The integrated role of hydrological connectivity, land use, and seasonality. *Hydrol. Process.* **32**, 1664–1681 (2018).
5. Pringle, C. What is hydrologic connectivity and why is it ecologically important? *Hydrol. Process.* **17**, 2685–2689 (2003).
6. van der Kamp, G. & Hayashi, M. Groundwater-wetland ecosystem interaction in the semiarid glaciated plains of North America. *Hydrogeol. J.* **17**, 203–214 (2009).
7. Lee, S. *et al.* Seasonal drivers of geographically isolated wetland hydrology in a low-gradient, Coastal Plain landscape. *J. Hydrol.* **583**, 124608 (2020).
8. Jones, C. N. *et al.* Modeling connectivity of non-floodplain wetlands: Insights, approaches, and recommendations. *J. Am. Water Resour. Assoc.* **55**, 559–577 (2019).
9. Thorslund, J. *et al.* Solute evidence for hydrological connectivity of geographically isolated wetlands. *Land Degrad. Dev.* **29**, 3954–3962 (2018).
10. Epting, S. M. *et al.* Landscape metrics as predictors of hydrologic connectivity between Coastal Plain forested wetlands and streams. *Hydrol. Process.* **32**, 516–532 (2018).
11. McDonough, O. T., Lang, M. W., Hosen, J. D. & Palmer, M. A. Surface hydrologic connectivity between Delmarva Bay wetlands and nearby streams along a gradient of agricultural alteration. *Wetlands* **35**, 41–53 (2015).
12. Sivakumar, B. & Singh, V. P. Hydrologic system complexity and nonlinear dynamic concepts for a catchment classification framework. *Hydrol. Earth Syst. Sci.* **16**, 4119–4131 (2012).
13. Denver, J. M. *et al.* Nitrate fate and transport through current and former depressional wetlands in an agricultural landscape, Choptank Watershed, Maryland, United States. *J. Soil Water Conserv.* **69**, 1–16 (2014).

14. McLaughlin, D. L., Kaplan, D. A. & Cohen, M. J. A significant nexus: Geographically isolated wetlands influence landscape hydrology. *Water Resour. Res.* **50**, 7153–7166 (2014).
15. Phillips, P. J. & Shedlock, R. J. Hydrology and chemistry of groundwater and seasonal ponds in the Atlantic Coastal Plain in Delaware, USA. *J. Hydrol. (Amst.)* **141**, 157–178 (1993).
16. Winter, T. C. & Labaugh, J. W. Hydrologic considerations in defining isolated wetlands. *Wetlands* **23**, 532–540 (2003).
17. Lindsey, B. D. *et al.* *Residence Times and Nitrate Transport in Ground Water Discharging to Streams in the Chesapeake Bay Watershed* (US Geological Survey, 2003).
18. Fenstermacher, D. E., Rabenhorst, M. C., Lang, M. W., McCarty, G. W. & Needelman, B. A. Distribution, morphometry, and land use of Delmarva Bays. *Wetlands* **34**, 1219–1228 (2014).
19. Pyzoha, J. E., Callahan, T. J., Sun, G., Trettin, C. C. & Miwa, M. A conceptual hydrologic model for a forested Carolina bay depression on the Coastal Plain of South Carolina, USA. *Hydrol. Process* **22**, 2689–2698 (2008).
20. Sugihara, G. *et al.* Detecting causality in complex ecosystems. *Science* **338**, 496–500 (2012).
21. Ye, H., Deyle, E. R., Gilarranz, L. J. & Sugihara, G. Distinguishing time-delayed causal interactions using convergent cross mapping. *Sci. Rep.* **5**, 1–9 (2015).
22. van Nes, E. H. *et al.* Causal feedbacks in climate change. *Nat. Clim. Change* **5**, 445–448 (2015).
23. Clark, A. T. *et al.* Spatial convergent cross mapping to detect causal relationships from short time series. *Ecology* **96**, 1174–1181 (2015).
24. Nakayama, S. I., Takasuka, A., Ichinokawa, M. & Okamura, H. Climate change and interspecific interactions drive species alternations between anchovy and sardine in the western North Pacific: Detection of causality by convergent cross mapping. *Fish Oceanogr.* **27**, 312–322 (2018).
25. Wang, Y. *et al.* Detecting the causal effect of soil moisture on precipitation using convergent cross mapping. *Sci. Rep.* **8**, 1–8 (2018).
26. Delforge, D., De Viron, O., Vanclooster, M., Van Camp, M. & Watlet, A. Detecting hydrological connectivity using causal inference from time series: Synthetic and real karstic case studies. *Hydrol. Earth Syst. Sci.* **26**, 2181–2199 (2022).
27. Ombadi, M., Nguyen, P., Sorooshian, S. & Hsu, K. L. Evaluation of methods for causal discovery in hydrometeorological systems. *Water Resour. Res.* **56**, 7251 (2020).
28. Bonotto, G., Peterson, T. J., Fowler, K. & Western, A. W. Identifying causal interactions between groundwater and streamflow using convergent cross-mapping. *Water Resour. Res.* **58**, 231 (2022).
29. Lee, S., McCarty, W. G., Lang, W. M. & Li, X. Overview of the USDA Mid-Atlantic regional wetland conservation effects assessment project. *J. Soil Water Conserv.* **75**, 684–694 (2020).
30. Lee, S. *et al.* Effects of subsurface soil characteristics on wetland-groundwater interaction in the coastal plain of the Chesapeake Bay Watershed. *Hydrol. Process.* <https://doi.org/10.1002/hyp.13326> (2018).
31. Fuka, D., Walter, M., Archibald, J., Steenhuis, T. & Easton, Z. Package ‘EcoHydRology’. <https://cran.microsoft.com/snapshot/2017-04-21/web/packages/EcoHydRology/EcoHydRology.pdf> (2015).
32. Nathan, R. & McMahon, T. Evaluation of automated techniques for base flow and recession analyses. *Water Resour. Res.* **26**, 1465–1473 (1990).
33. Qi, J. *et al.* A coupled surface water storage and subsurface water dynamics model in SWAT for characterizing hydroperiod of geographically isolated wetlands. *Adv. Water Resour.* **131**, 103380 (2019).
34. Ye, H., Clark, A., Deyle, E. & Munch, S. *rEDM: An R Package for Empirical Dynamic Modeling and Convergent Cross Mapping*. <https://ha0ye.github.io/rEDM/articles/rEDM.html>.
35. Sanford, W. E. & Pope, J. P. Quantifying groundwater’s role in delaying improvements to Chesapeake Bay water quality. *Environ. Sci. Technol.* **47**, 13330. <https://doi.org/10.1021/es401334k> (2013).
36. Kirchner, J. W. Quantifying new water fractions and transit time distributions using ensemble hydrograph separation: Theory and benchmark tests. *Hydrol. Earth Syst. Sci.* **23**, 303–349 (2019).
37. Hrachowitz, M. *et al.* Transit times—The link between hydrology and water quality at the catchment scale. *Wiley Interdiscip. Rev. Water* **3**, 629–657 (2016).
38. Baily, A., Rock, L., Watson, C. J. & Fenton, O. Spatial and temporal variations in groundwater nitrate at an intensive dairy farm in South-East Ireland: Insights from stable isotope data. *Agric. Ecosyst. Environ.* **144**, 308–318 (2011).

Acknowledgements

This work was supported by 1) the National Research Foundation of Korea (NRF) Grant funded by the Korean government (MIST) (No. 2021R1C1C1006030), 2) the Core Research Institute Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2021R1A6A1A10045235) and 3) the United States Department of Agriculture (USDA) Natural Resources Conservation Service, in association with the Wetland Component of the National Conservation Effects Assessment Project. This research was a contribution of the Long-Term Agroecosystem Research (LTAR) network, which is supported by the USDA.

Disclaimer

The U.S. Department of Agriculture is an equal opportunity provider and employer. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Author contributions

S.L. wrote the manuscript with support from J.L. and J.S. S.L. and B.L. carried out the experiment. S.L. and G.M. supervised the manuscript. S.L. and B.L. revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-44071-0>.

Correspondence and requests for materials should be addressed to S.L. or G.W.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023