Estimating Near-Saturated Soil Hydraulic Conductivity Based on Its Scale-Dependent Relationships with Soil Properties

Yang Yang,* Ole Wendroth, Sleem Kreba, Baoyuan Liu

Soil hydraulic conductivity near saturation ($K_{ns}$) is affected by various soil properties operating at different spatial scales. Using noise-assisted multivariate empirical mode decomposition (NA-MEMD), our objective was to inspect the scale-dependent interactions between $K_{ns}$ and various soil properties and to estimate $K_{ns}$ based on such relationships. In a rectangular field evenly across cropland and grassland, a total of 44 sampling points separated by 5 m were selected and measured for $K_{ns}$ at soil water pressure heads of $-1$, $-5$ and $-10$ cm. At each point, the saturated conductivity $K_s$ was estimated using Gardner’s exponential function, and six soil structural and textural properties were investigated. Decomposed into four intrinsic mode functions (IMFs) and a residue by NA-MEMD, each $K$ was found to significantly correlate with all six properties at one spatial scale at least. The variations in $K$ were primarily regulated by soil structure, especially at the relatively small scales. Multiple linear regression (MLR) failed to regress either IMF1 or IMF2 of each $K$ from the soil properties of the equivalent scales and only accounted for 13.7 to 43.6% of the total variance in calibration for the remaining half of the IMFs. An artificial neural network was then adopted to estimate IMF1 and IMF2, and the corresponding results were added to the MLR estimates at other scales for which each $K$ was estimated at the measurement scale. This prediction greatly outperformed the MLR modeling before NA-MEMD and, on average, accounted for additional 74.4 and 73.4% of the total variance in calibration and validation, respectively. These findings suggest nonlinear correlations between $K$ and the soil properties investigated at the small scales and hold important implications for future estimations of $K_{ns}$ and $K_s$ as well as other hydraulic properties.

Knowledge of the soil hydraulic conductivity near saturation ($K_{ns}$) is required for precise description of soil-structure-related and macropore flow (Baird, 1997; Alaoui et al., 2011), which can be a predominant pathway for soil water and solute transport even though the large voids contribute only a small portion to the total soil porosity (Beven and Germann, 1982). The direct measurement of $K_{ns}$, however, is tedious and time consuming, either in the laboratory or in the field. One feasible solution to this challenge is to make statistical estimates of $K_{ns}$ based on soil properties that are easier to acquire or are included in routine soil surveys (Wösten et al., 2001; Patil and Singh, 2016; Yang et al., 2018). The corresponding statistical models, i.e., established among soil properties, have been formally named pedotransfer functions (PTFs) since 1989 (Bouma, 1989), but their applications for hydraulic conductivity can be traced back to the early 1970s (Anderson and Bouma, 1973). To attain a reliable PTF for $K_{ns}$ estimation, the characterization of its interactions with relevant soil properties is highly critical.
Many previous studies have demonstrated the scale-dependent relationships of soil hydraulic properties including $K_{ns}$ with various soil properties (e.g., Crawford, 1994; Sobieraj et al., 2004; She et al., 2015). Summarizing these studies, any statistically significant correlation observed at a particular scale is not necessarily found at the investigation scale, which is usually designed arbitrarily. In other words, decomposing the spatial series of a soil hydraulic property into different frequency components, each representing distinctive spatial scale (Kachanoski and de Jong, 1988), the significant one observed at one particular scale can be compromised by the insignificant ones or likewise significant but opposite ones at other scales. The result is no valid correlation at the scale of investigation (Nielsen et al., 1983; Hu and Si, 2013; Yang et al., 2019). To sufficiently describe the scale-dependent relationships between $K_{ns}$ and particular soil properties, the empirical mode decomposition (EMD) has been recently recommended as an effective tool to extract their frequency components at different scales (Biswas and Si, 2011).

Spectral and wavelet analyses decompose variance components into different scales. Likewise, EMD is a data analysis tool for frequency-domain-based scale component separation. However, in contrast to the other two methods, EMD does not require stationarity or linearity for the data series (Huang et al., 1998). Instead, EMD is an adaptive, fully data-driven technique that decomposes a spatial or temporal series into a set of intrinsic oscillatory modes named intrinsic mode functions (IMFs). These functions reflect characteristic oscillation scales and a residue that manifests the monotonic trend outside the domain (Huang et al., 1998). Moreover, if the scales corresponding to IMFs are clearly separated, they are physically meaningful (Looney et al., 2015). In view of these great merits, EMD has been widely applied in geophysics (e.g., Salisbury and Wimbush, 2002; Huang and Wu, 2008; Chen et al., 2015) and neuroscience (e.g., Liang et al., 2000; Acharya et al., 2017) since its development in 1998, yet only a few applications have been reported in soil science.

To our knowledge, Biswas and Si (2011) were the first who used EMD in soil science, by partitioning soil water storage (SWS) in a hummocky landscape in Canada into different IMFs. With such SWS decomposition, they found that the variations in SWS at different spatial scales were linearly correlated with distinctive soil properties at the measurement scale. The sum of MLR estimates of all the IMFs and the residue provided a much better prediction of SWS than the one generated directly by MLR of the undecomposed data series. Nevertheless, Hu and Si (2013) emphasized that similar to SWS varying at different scales, the associated soil properties oscillated at multiple scales as well. They then applied the multivariate EMD (MEMD) (Rehman and Mandic, 2010) to separate both SWS and the soil properties synchronously, i.e., into similar scales. At each scale, a SWS IMF or its residue was then related to the soil properties using MLR. This estimation scheme was demonstrated to further improve the prediction of SWS in the study of Hu and Si (2013) and has been successfully applied to the spatial simulation of other soil hydraulic properties in different study regions (e.g., She et al., 2015, 2017).

Notwithstanding, analyzing saturated $K$ ($K_s$) spatial series along an 860-km transect across the Loess Plateau of China, Yang et al. (2019) found that at relatively small spatial scales, $K_s$ tended to be nonlinearly correlated with soil properties such as bulk density and environmental factors such as elevation. As a consequence, MLR would not have provided satisfactory estimations. In view of the obscured small-scale interrelationships, the artificial neural network (ANN) method inspired by the neural networks of the human brain was recommended (Pachepsky et al., 1996; Schaap and Bouten, 1996). This method assigns coefficients to quantify the impact of input variables and establishes network structures to describe the interconnections among all input and output variables, exhibiting an effective ability to simulate complex systems (Hecht-Nielsen, 1990; Donner and Barbosa, 2008). Yet as a non-parametric data analysis tool, ANN provides no particular function, coefficient, or connecting structure manifesting the associations among soil properties, which limits its application to datasets of the same variables sampled in other domains.

It is worth noting that both MEMD and the univariate EMD manifest a major shortcoming of mode mixing, which can become obvious in two typical ways (Huang and Shen, 2005; Rehman et al., 2013). One is a single IMF exhibiting fluctuations at several scales and the other occurs when the same scale of oscillation is displayed in more than one IMF. Therefore, the noise-assisted MEMD (NA-MEMD) was developed to minimize this problem within the extracted IMFs (Rehman and Mandic, 2011). Using the quasi-dyadic filter bank structure of MEMD, this method adds uncorrelated white Gaussian noise (WGN) in separate channels adjoining the signal channels as a decomposition reference and shows a remarkable ability to alleviate mode mixing in both synthetic and real data (Looney et al., 2015; Yang et al., 2019).

The goal of the current study was to evaluate NA-MEMD for characterizing the scale-dependent relationships of $K_{ns}$ with different soil properties and to estimate $K_{ns}$ based on these scale-specific relationships. In a rectangular field divided evenly into two land use systems, i.e., cropland and grassland, in situ hydraulic conductivities $K$ were measured at three pressure heads $b$ of $-1$, $-5$, and $-10$ cm at 44 points separated by a constant interval of 5 m. The $K_s$ was also estimated by the exponential function according to Gardner (1958). In addition to soil bulk density (BD), particle-size composition, and soil organic C (SOC) content, which are common covariates in PTFs for hydraulic property estimation, two other soil properties describing soil structure, i.e., mean weight diameter (MWD) of dry aggregates and wet aggregate stability (WAS), were furthermore investigated because $K_{ns}$ and $K_s$ strongly depend on soil structure (Beven and Germann, 1982). The specific objectives were to: (i) detect multiscale variability of $K_{ns}$, $K_s$, and their associated soil properties; (ii) analyze the correlations of $K_{ns}$ and $K_s$ with various soil properties at different spatial scales; and (iii) estimate $K_{ns}$ and $K_s$ based on their multiscale relationships examined in (ii).
Materials and Methods

Experimental Design

In the fall of 2010, the experiment was performed in a rectangular field located at the Spindletop Research Farm affiliated with the University of Kentucky, Lexington, KY. This field is 75 m long and 55 m wide and was divided into two land use systems of cropland and grassland, each covering half of the total area (Fig. 1). The cropland half was planted with no-till winter wheat (*Triticum aestivum* L.) following the local customs as described by Kreba et al. (2013). The other half was dominated by three typical grass species including bluegrass (*Poa pratensis* L.), red clover (*Trifolium pratense* L.), and tall fescue (*Festuca arundinacea* Schreb.) and was mowed approximately once a month from May to November each year. Note that before such establishment, the entire field had been used for tobacco (*Nicotiana tabacum* L.) production with conventional tillage for several years. The soil is classified as a mixed, semiactive, mesic Typic Paleudalf (Soil Survey Staff, 1999). The mean annual precipitation and temperature are 1148 mm and 13°C, respectively.

A total of 44 sampling points were laid over the experimental field (Fig. 1). To avoid the impact exerted by the surrounding land use systems, i.e., grassland without regular mowing, these points were arranged on four transects at an interval of 5 m. These transects ran parallel to the rectangular outer boundaries at a 5-m distance toward the inside. At each point, a tension infiltrometer (Soil Measurement Systems) was used to measure in situ infiltration rates at different *h*. After cutting the grass or wheat stubble and clearing the soil surface, a ring with the same inner diameter as the infiltrometer base was placed and filled with silica sand to provide a good contact between the infiltrometer’s base plate membrane and the soil below. Each water infiltration measurement was terminated when a constant volume flow rate *Q* was obtained for three consecutive time intervals. Adopting Wooding’s equation (Wooding, 1968), which assumes a simplified exponential relationship of *K* with *h* for unconfined steady-state infiltration (Ankeny et al., 1991; Thony et al., 1991), the slope of ln(*K*) vs. pressure head was derived as

\[
\alpha_{h_i, h_j} = \frac{\ln(Q_{h_j}) - \ln(Q_{h_i})}{h_j - h_i}
\]

where *h* and *h* denote two different pressure heads (cm), *Q_{h_i}* and *Q_{h_j}* are the final infiltration rates measured at *h* and *h*, respectively (cm³ min⁻¹), and \(\alpha_{h_i, h_j}\) is the dimensionless slope of ln(*K*) derived between *h* and *h*. The hydraulic conductivity *K* is calculated as

![Fig. 1. Experimental layout displaying the 44 sampling points within the 75-by-55-m rectangular field divided evenly between cropland and grassland. The points (open circles) were distributed at an interval of 5 m along four transects parallel to and 5 m away from the four sides of the field.](image-url)


\[ K_{hi} = \frac{Q_h}{\pi r^2 \left(1 + 4/j \pi \alpha_{hi},b_j \right)} \]  \hspace{1cm} \text{[2]}

\[ K_{bj} = \frac{Q_{b_j}}{\pi r^2 \left(1 + 4/j \pi \alpha_{hi},b_j \right)} \]  \hspace{1cm} \text{[3]}

where \( K_{hi} \) and \( K_{bj} \) are the hydraulic conductivities measured at \( hi \) and \( bj \), respectively (cm min\(^{-1}\)) and \( r \) is the radius of the infiltration area (cm). The saturated hydraulic conductivity can be estimated by either \( K_{hi} \) or \( K_{bj} \) (Gardner, 1958):

\[ K_s = \frac{K_{hi}}{\exp(\alpha_{hi},b_j)} \]  \hspace{1cm} \text{[4]}

\[ K_s = \frac{K_{bj}}{\exp(\alpha_{hi},b_j)} \]  \hspace{1cm} \text{[5]}

where \( K_s \) is the saturated hydraulic conductivity (cm min\(^{-1}\)), and the values of \( K_s \) generated from Eq. [4] and [5] are identical because the same slope of \( \alpha_{hi,bj} \) was used.

Using a tension infiltrometer with a 20-cm base, the radius \( r \) was equal to 10 cm in the current study. The infiltration rates were measured at \( h \) of −10, −5, and −1 cm, successively, at each sampling point. On average, 34, 20, and 6 min, respectively, were required to reach steady-state infiltration at the three pressure heads. Using Eq. [1], three slopes \( \alpha \) were calculated, i.e., between −1 and −5 cm, between −5 and −10 cm, and between −1 and −10 cm. Accordingly, for the hydraulic conductivity \( K \) at −1, −5, or −10 cm, two different values were calculated from the two slopes involving the corresponding \( Q \) and \( b \) values. For example, \( K_{−1} \) was derived from \( Q_{−1} \) and \( \alpha_{−1,−5} \) and from \( Q_{−1} \) and \( \alpha_{−1,−10} \) using Eq. [2]. Each two values were then arithmetically averaged as the final result of \( K \) at −1, −5, or −10 cm. Similarly, the three estimates of \( K_s \) corresponding to the three different values of \( \alpha \) were averaged.

At each sampling point, an undisturbed soil core with 8.6-cm diameter and 6-cm height was collected at the depth of 4 to 10 cm and oven dried for bulk density (BD) determination. Meanwhile, a disturbed soil sample was excavated from each point at approximately the 0- to 10-cm depth. After air drying, each sample was analyzed using an Analytette 3 vibratory sieve shaker (Fritsch GmbH) operated with four different sieves to obtain the fractions of five aggregate sizes, i.e., <0.05, 0.05 to 0.25, 0.25 to 1, 1 to 2, and >2 mm. Based on the resulting dry aggregate size distribution, the mean weight diameter (MWD) was calculated to reflect the average size (Kemper and Rosenau, 1986):

\[ \text{MWD} = \sum_{j=1}^{n} W_j X_j \]  \hspace{1cm} \text{[6]}

where \( X_j \) and \( W_j \) denote the mean diameter (mm) and the weight percentage of the aggregates of the \( j \)th size class, respectively, and \( n \) is the total number of the size classes analyzed and equaled 5 in this study. Utilizing a wet sieving apparatus (Eijkelkamp), the dry aggregates of 1- to 2-mm size separated above were further evaluated for wet aggregate stability (WAS), which is defined as the percentage of wet stable aggregates >0.26 mm relative to the sum of stable and unstable aggregates (Kemper and Rosenau, 1986; Nimmo and Perkins, 2002).

In another sampling campaign in 2010, at each point, a disturbed soil core was taken from the 0- to 10-cm depth. These samples were air dried and divided into two subsamples. One was analyzed with the sieving and pipette methods (Gee and Bauder, 1986) for particle-size composition, from which the fractions of sand (0.05–2 mm, SAND) and clay (<0.002 mm, CLAY) were derived (Soil Survey Staff, 1999). The other one was examined for soil organic C (SOC) using the dry combustion method.

**Noise-Assisted Multivariate Empirical Mode Decomposition**

Noise-assisted multivariate empirical mode decomposition is a multivariate extension of EMD. For a multivariate dataset with \( n \) spatial series, \( X(s) = (x_1(s), x_2(s), ..., x_n(s)) \), where \( s \) refers to the spatial position, the IMFs are obtained with NA-MEMD according to the following steps (Rehman and Mandic, 2011):

1. Generate an uncorrelated WGN dataset with \( m \) series of the same length as \( X(s) \) and add it in separate channels adjoining the \( n \) signal channels, composing a new dataset \( X'(s) \).
2. Apply a low discrepancy Hammersley sequence to derive an appropriate sampling point set on an \((n + m + 1)\) hypersphere, thereby ascertaining the direction vector for \( K \) directions and to calculate the projection of the input dataset with \( n + m \) series.
3. Interpolate the extrema of projections using cubic spline interpolation for the multivariate envelopes of all directions and calculate the corresponding means \( M(s) \).
4. Calculate \( C(s) \) as \( X'(s) - M(s) \). If \( C(s) \) fulfills the stopping criterion for a multivariate IMF, execute the steps above on \( X'(s) - C(s) \); otherwise perform them on \( C(s) \).
5. Eliminate the resulting IMFs associated with WGN, obtaining a set of IMFs for the input signal dataset \( X(s) \).

For more details regarding the principles and implementation of NA-MEMD, see Rehman and Mandic (2011) and Looney et al. (2015). Because NA-MEMD is a data processing tool applicable only to one-dimensional multivariate datasets measured at uniform intervals, the 44 points were arrayed following the numbers presented by the side of the corresponding open circles in Fig. 1, i.e., 1, 2, ..., 44. The midpoint between Sampling Points 1 and 44, which was located on the border between the cropland and grassland, was treated as the origin of this 220-m “virtual” transect. Therefore, the spatial distance of Point 1 was 2.5 m and that of Point 44 was 217.5 m.

Following the spatial arrangement above, each of the four spatial series of \( K_1, K_{−1}, K_{−5}, \) and \( K_{−10} \) was combined with the six soil properties of BD, MWD, WAS, SAND, CLAY, and SOC to form four multivariate datasets. Because some of the variables varied by one or more magnitudes, e.g., WAS vs. SAND, all data series were scaled prior to NA-MEMD to prevent numerical problems that
may arise in the following estimations of $K$ or $K$ components, i.e., IMFs and residues. For a data series $x_i$ with $i$ referring to the spatial position, the normalized $x_i^*$ is defined as (Nielsen and Wendroth, 2003)

$$x_i^* = \frac{x_i - (\bar{x} - 2\sigma_x)}{4\sigma_x}$$

where $\bar{x}$ and $\sigma_x$ denote the mean and standard deviation, respectively, of $x_i$. The $x_i^*$ after normalization would have a mean of 0.5 and a standard deviation of 0.25.

Each of the four multivariate datasets was analyzed with NA-MEMD to extract the corresponding IMFs and residues using the codes provided by Looney et al. (2015) in MATLAB 2018a. The numbers of directions and WGN channels were set at the default values of 64 and 10, respectively. The stopping criterion of sifting recommended by Huang et al. (1999, 2003) was adopted, i.e., sifting was terminated when the same number of zero crossings and extrema was obtained for three consecutive sifting steps. The oscillation scale of each IMF was identified via the periodogram function built into MATLAB 2018a (Biswas et al., 2013). The relative importance of each IMF or residue to the corresponding spatial series was calculated as the percentage contribution of variance, i.e., dividing the total variance of each spatial series by the variance of each IMF or residue (Hu and Si, 2013).

**Correlation Analysis**

Pearson correlation analysis was performed both before and after NA-MEMD using IBM SPSS Statistics 20. In the former, each of the four undecomposed $K$ data series was examined for its associations with the six soil properties at the scale of measurement. In the latter, the same index IMFs or the residue derived from each multivariate dataset were analyzed for their linear associations at different spatial scales.

**Hydraulic Conductivity Estimation**

For the estimation of each $K$, 70% (31) of the 44 sampling points were randomly selected for calibration and the remaining 30% (13) were used for validation. Both multiple linear regression (MLR) and artificial neural network (ANN) were applied to estimate $K$ of different $h$ after decomposing $K$ and the soil properties into different frequency components with NA-MEMD. Each $K$ component was then predicted from the same index IMFs or residue of the soil properties for each multivariate dataset. The estimated IMFs and residue of each $K$ were summed as the prediction at the scale of measurement.

Using IBM SPSS Statistics 20, MLR was accomplished with a stepwise procedure and the significance levels of $F$-to-enter and $F$-to-remove were set at 0.05 and 0.10, respectively. The ANN modeling was accomplished with the Neural Network Toolbox built in MATLAB 2018a. The ANN model adopted here consists of an input layer carrying independent variables, an output layer carrying dependent one(s), and a hidden layer with six nodes connecting input and output layers (Schaap and Leij, 1998). Following Yang et al. (2018), 100 ANN model runs were performed for each input–output pair, and the optimal model providing the least square error was selected. In addition, because the ANN modeling itself is not capable of choosing input variables, all possible combinations of no more than three variables were considered, and the optimal ANN model with the least square error was treated as the final result.

The quality of $K$ prediction was evaluated by two widely used parameters, i.e., the root mean square error (RMSE) and the adjusted coefficient of determination ($R^2_{adj}$). The former is defined as the square root of the mean squared differences between the estimates $\hat{x}_i$ and the measurements $x_i$ for all $i$th positions:

$$RMSE = \sqrt{\frac{1}{N}\sum(x_i^* - \hat{x}_i)^2}$$

where $N$ denotes the total number of measurements. The latter considers the number of regression variables $k$ and reflects the proportion of variance explained by the prediction:

$$R^2_{adj} = 1 - \left(1 - R^2\right)\left(\frac{N - k}{N - 1}\right)$$

where $R^2$ is the coefficient of determination, calculated as

$$R^2 = \frac{\sum(x_i^* - \bar{x}_i)^2}{\sum(x_i^* - \bar{x}_i)^2}$$

where $\bar{x}_i$ denotes the mean of the $x_i$ measurements.

**Results and Discussion**

**Spatial Distributions of Hydraulic Conductivity and Soil Properties**

According to the Shapiro–Wilk normality test, no $K$ series was normally distributed but the distributions of ln($K_{-5}$) and ln($K_{-10}$) were significantly normal. Therefore all $K$ series were natural-logarithm transformed and plotted against the spatial distance in Fig. 2a. For the six soil properties investigated, four (MWD, BD, SAND, and SOC) were normally distributed, and the remaining two (WAS and CLAY) were closer to normal distribution than the ln-transformed ones. The original ones are depicted (Fig. 2b–2g) and were used in the following analyses. As $b$ approached zero, ln($K$) increased accordingly (Fig. 2a). The values of ln($K$) consistently followed the sequence of ln($K_{0}$) > ln($K_{-1}$) > ln($K_{-5}$) > ln($K_{-10}$), except slightly higher ln($K_{-10}$) than ln($K_{0}$) spotted at three out of the total 44 sampling points, i.e., at the spatial distances of 7.5, 92.5, and 147.5 m. The unexpected behavior of $K$ at these three points was mainly caused by the small slope of ln($K$) between the pressure heads of −5 and −10 cm at these points, which strongly affected the calculation of final $K_h$ but not $K_{-1}$. In addition, the values of ln($K_{0}$) and ln($K_{-5}$) were relatively lower at these three points compared with the others, suggesting that fewer macropores of the corresponding sizes occurred here than at other locations.
In general, $K_s$ values were similar to $K_{-1}$ at all the sampling points and therefore their spatial distributions were rather alike (Fig. 2a). The corresponding means of ln($K_s$) and ln($K_{-1}$) across the field were $-1.22$ (equivalent to 0.296 cm min$^{-1}$) and $-1.51$ (0.220 cm min$^{-1}$), respectively, and the standard deviations were 1.39 (4.02 cm min$^{-1}$) and 1.23 (3.42 cm min$^{-1}$), respectively. On the other hand, ln($K_{-5}$) and ln($K_{-10}$) were much smaller, i.e., with means of $-5.0$ (0.007 cm min$^{-1}$) and $-5.8$ (0.003 cm min$^{-1}$), respectively, and their spatial distributions behaved similarly, while they both differed from those of ln($K_s$) and ln($K_{-1}$). Simply applying the capillary rise equation (Childs and Collis-George, 1950; Mualem, 1976), the pressure heads of $-1$, $-5$, and $-10$ cm studied here correspond to pores with approximate diameters <3000, <600, and <300 μm. The comparisons among ln($K$) at different pressure heads therefore suggested little presence of macropores with diameters >3000 μm but a larger number of pores between 600 and 3000 μm in the soil investigated. Besides, none of the four ln($K$) exhibited obvious spatial trends and no significant difference ($p<0.05$) was observed for any ln($K$) between the two land use systems, i.e., cropland and grassland, according to the independent sample t-test.

In contrast, however, all six soil properties except CLAY were significantly different between cropland and grassland. Significantly lower BD values were observed in the grassland than the cropland at the level of 0.05 (Fig. 2b). On the contrary, the mean sizes of dry aggregates were significantly larger in the grassland than those in the cropland at the level of 0.01 (Fig. 2c). The corresponding average MWDs were 2.5 and 2.2 mm in the grassland and cropland, respectively. The significant difference ($p<0.01$) in WAS was more apparent between these two land use systems (Fig. 2d). In the cropland, WAS ranged from 44.4 to 86.0% and averaged 66.8%, whereas WAS in the grassland were much larger, i.e., it averaged 94.6%, and fluctuated within a quite smaller range between 86.0 and 98.0% than in the cropland. The comparisons of BD, MWD, and WAS suggested more developed soil structure in the grassland than in the cropland, which probably occurred due to the typically denser
root systems and higher organic matter content under grass (Six et al., 2000; Yang et al., 2013). This reasoning was to some extent supported by SOC being significantly higher in the grassland than in the cropland ($p < 0.01$) (Fig. 2g). The corresponding means were 20.8 and 17.2 g kg$^{-1}$, and the standard deviations were 3.5 and 2.3 g kg$^{-1}$, respectively. Besides, the grassland also revealed significantly more sand than the cropland ($p < 0.01$). However, the difference was rather small quantitatively, i.e., only 0.02 g g$^{-1}$ on average (Fig. 2c).

**Multivariate Empirical Mode Decomposition on Hydraulic Conductivity and Soil Properties**

After normalization with respect to the mean and standard deviation using Eq. [7], the spatial series of the four different ln($K_s$) as well as the six soil properties were combined into four multivariate datasets and each decomposed into four IMFs and one residue with the aid of NA-MEMD (Fig. 3 and 4). The spatial series of the six properties were, of course, identical for the four datasets. Nevertheless, the resulting IMFs and residues turned out slightly different, as the sifting process for a multivariate dataset was controlled by all the spatial series combined rather than a single variable involved (Rehman and Mandic, 2010).

Based on their periodograms, the primary oscillation scale of each spatial series was identified and exhibited an increasing trend with the mode index (Table 1; Fig. 3 and 4). For example, the scales of ln($K_s$) IMFs were 14, 48, 71, and 142 m as the mode index increased from 1 to 4, successively. Comparing the oscillations for ln($K_s$) at different $h$, the spatial scales were rather similar for the IMF1s and IMF2s, i.e., around 15 and 48 m, respectively. Whereas, the IMF3 of ln($K_s$) possessed a much larger spatial scale of 128 m, i.e., nearly twice those for the IMF3s of ln($K_5$), ln($K_{10}$), and ln($K_{-10}$). The spatial scales identified for IMF4s, relatively larger ones were observed for ln($K_{10}$) and ln($K_{-10}$) compared with those for ln($K_5$) and ln($K_{-5}$). These results imply that ln($K_s$) at different $h$ may have been regulated by different soil properties operating at disparate scales and resulting in diverse spatial behaviors, especially at the relatively larger scales corresponding to the IMF3s and IMF4s.

It is also worth noting that the spatial scales of the same index IMFs more or less varied among the variables involved in each multivariate dataset, suggesting that there is no presence of exactly common scales. The CVs were typically <25% except one at 31% for the IMF3s of the multivariate dataset accommodating ln($K_{-10}$). Despite the scales not being exactly identical, however very similar, for each mode index, the scales of all the variables were averaged across each multivariate dataset (Hu and Si, 2013), and the results are depicted as the means in Table 1.

Not only did ln($K_s$) and ln($K_{10}$) behave spatially similarly to each other, as obvious from Fig. 2, their variance distributions among IMFs were also alike (Fig. 3). Around 50% of the total variance was manifested in the IMF1s, suggesting predominant fluctuations at the smallest scale of only 12 m for these two $K$ series (Fig. 3a). The second largest proportions of variance were observed in the IMF3s associated with the spatial scale around 70 m. The corresponding percentages were 11.4% for ln($K_s$) and 17.0% for ln($K_{10}$) (Fig. 3i). Each of the remaining two IMFs contributed slightly more than 5% to the total variance (Fig. 3e and 3m). Less than 3% of the variance was reflected in the residue (Fig. 3q) for either ln($K_5$). When adding up the percentage contributions of all the IMFs and the residue for either ln($K_5$) or ln($K_{10}$), however, the sum was not equal to 100%. The same outcome was observed for the other two ln($K_s$) datasets, as well as for the six soil properties discussed below. This result implies that the components extracted by NA-MEMD were not perfectly orthogonal. Because orthogonality is not a condition for EMD or NA-MEMD, the lack of fulfilling it therefore does not compromise the physical scale-specific meanings of the IMFs and the residue (Huang et al., 1998).

By comparison, ln($K_{10}$) and ln($K_{-10}$) mainly fluctuated at the smallest scale of ~12 m as well, but the corresponding percentages of variance were relatively lower than those for ln($K_5$) and ln($K_{10}$), i.e., 40.2% for ln($K_{10}$) and 43.1% for ln($K_{-10}$) (Fig. 3a). The IMF2s at the scale of 34 m contributed second, i.e., 16.8 and 20.3%, respectively, to the total ln($K$) variance (Fig. 3e). The percentages of variance attributed to IMF3 and IMF4 for both variables were relatively small, i.e., ~5%, and the residue contributed even less to the total variance. Besides the primary oscillation scales of the same index IMFs and the variance distributions among the IMFs and the residue, the components of ln($K_{10}$) and ln($K_{-10}$) exhibited quite different spatial behaviors from those of ln($K_5$) and ln($K_{-5}$). The difference in spatial patterns between both pairs of ln($K$) datasets (Fig. 2) is manifested in this result. As the most prominent contrast, the residues of ln($K_{10}$) and ln($K_{-10}$) exhibited monotonic trends opposite to the ones of ln($K_5$) and ln($K_{-5}$), suggesting distinct soil properties operating or the same soil property but operating in the opposite way.

In contrast to ln($K_s$), generally more variance of the three properties describing soil structure was attributed to the IMFs of higher indices and to the residues, indicating greater fluctuations at larger spatial scales (Fig. 3). The IMF1s contributed only ~30% to the total variances of both BD and MWD, no matter with which ln($K$) they were combined in the analysis (Fig. 3b and 3c). The variance distributions differed for the other IMFs between BD and MWD. The former exhibited nearly equivalent proportions of variance in the IMF2s, IMF3s, and IMF4s, i.e., ranging from 10.3 to 18.8% (Fig. 3f, 3j, and 3n), whereas little variance was attributed to the IMF2s (Fig. 3g) for the latter MWD but more to the IMF3s, IMF4s, and residues (Fig. 3k, 3o, and 3s). The variance distributions were even more disparate for WAS, which mainly fluctuated at a scale larger than the study domain, as 35.4 to 49.4% of variance was attributed to the residues (Fig. 3s).

For the other three soil properties, i.e., SAND, CLAY, and SOC, the largest proportions of the variances were all attributed to the IMF1s and hence occurred at the smallest spatial scales (Fig. 4a–4c). Nevertheless, not only did the corresponding scales differ, i.e., 18 m for SAND vs. 10 m for both CLAY and SOC, but also the particular percentages varied: ~45, ~35,
Fig. 3. Intrinsic mode functions (IMFs) and residues of $\ln(K)$, bulk density (BD), mean weight diameter (MWD), and wet aggregate stability (WAS) extracted by noise-assisted multivariate empirical mode decomposition (NA-MEMD). The hydraulic conductivity $K$ was investigated under four conditions, including saturated ($K_s$) and at three pressure heads of $-1$ ($K_{-1}$), $-5$ ($K_{-5}$), and $-10$ cm ($K_{-10}$). The IMFs for each variable are arranged following the mode index from 1 to 4 and ending with the residue. The percentage of variance accounted for by each IMF or residue is presented in each subplot of each variable in the color corresponding to the $\ln(K)$ analyzed.
and ~25% for SAND, CLAY, and SOC, respectively. Moreover, the second-strongest SAND and SOC fluctuations were found at the scales associated with the IMF2s, followed by the residues, whereas for CLAY, generally greater proportions of the variances were accounted for by the IMF3s in addition to the IMF1s. It is interesting that the two soil properties describing soil particle-size composition, i.e., SAND and CLAY, revealed more oscillations at the relatively smaller spatial scales than the three indicators of soil structure, i.e., BD, MWD, and WAS. This behavior was not expected because soil particle size usually fluctuates gradually while abrupt spatial changes are commonly observed in properties that are linked with soil structure. However, depending on

Fig. 4. Intrinsic mode functions (IMFs) and residues of sand content (SAND), clay content (CLAY), and soil organic C (SOC) extracted by noise-assisted multivariate empirical mode decomposition (NA-MEMD) corresponding to the hydraulic conductivities $K$ investigated under four conditions, including saturated ($K_s$) and at three pressure heads of $-1$ ($K_{-1}$), $-5$ ($K_{-5}$), and $-10$ cm ($K_{-10}$). The IMFs for each variable are arranged following the mode index from 1 to 4 and ending with the residue. The percentage of variance accounted for by each IMF or residue is presented in each subplot of each variable in the color corresponding to the $\ln(K)$ analyzed.
Linear Correlations of Hydraulic Conductivity with Different Soil Properties at Diverse Scales

The Pearson correlation coefficients of ln($K_s$), ln($K_{-5}$), ln($K_{-10}$) with various soil properties at different scales are summarized in Table 2. Based on the undecomposed data series, ln($K_s$) and ln($K_{-10}$) were only significantly correlated with SOC ($p < 0.01$) and CLAY ($p < 0.05$) at the scale of measurement, whereas for ln($K_{-5}$) and ln($K_{-10}$), significant correlations were only detected with MWD ($p < 0.01$).

After decomposing the data series using NA-MEMD, each ln($K_s$) was found to significantly correlate with every soil property investigated at various numbers of spatial scales (Table 2). For example, ln($K_{-5}$) was significantly correlated with BD at the largest scale corresponding to the residue only ($p < 0.01$) but possessed significant correlations with MWD at the three scales associated with IMF3, IMF4, and the residue. Moreover, in the latter case, the signs of significant correlations between ln($K_s$) and MWD reversed for different scales, which is also manifested in some other pairs of variables. These results agreed with those of She et al. (2015, 2017) and Yang et al. (2019), who also detected insignificant linear correlations of $K_s$ with associated soil properties at the measurement scale but significant correlations for one or more scale components extracted by MEMD or NA-MEMD. A possible reason for this phenomenon could be that the significant interactions detected at some scales associated with particular IMFs or the residue tend to be diminished by the insignificant correlations and/or by likewise significant but opposite ones at other scales, resulting in a poor correlation at the measurement scale (Si, 2003; Hu and Si, 2013).

The behavior of the scale-dependent correlations of ln($K_s$) with WAS stands in strong contrast to the relationships with MWD. At the scales corresponding to IMF3, IMF4, and the residue, ln($K_{-5}$) was significantly correlated with both MWD and WAS ($p < 0.01$), and the absolute values of corresponding coefficients were larger for the latter. However, owing to the statistically insignificant interactions at the other two scales identified, i.e., IMF1 and IMF2, WAS exhibited no significant correlation with ln($K_s$) at the measurement scale. By comparison, although an insignificant correlation was detected in the IMF2s between ln($K_{-5}$) and MWD, the IMF1 of ln($K_{-5}$), which accounted for the largest proportion of the total variance (Fig. 3a), was significantly correlated with the IMF1 of MWD ($p < 0.05$), leading to a uniquely significant correlation at the scale of measurement (Table 2).

According to the magnitude of the correlation coefficients, the small-scale variations of ln($K_s$) and ln($K_{-10}$) manifested in the IMF1s, IMF2s, IMF3, IMF4, and the residue. Moreover, in the latter case, the signs of significant correlations between ln($K_s$) and MWD reversed for different scales, which is also manifested in some other pairs of variables. These results agreed with those of She et al. (2015, 2017) and Yang et al. (2019), who also detected insignificant linear correlations of $K_s$ with associated soil properties at the measurement scale but significant correlations for one or more scale components extracted by MEMD or NA-MEMD. A possible reason for this phenomenon could be that the significant interactions detected at some scales associated with particular IMFs or the residue tend to be diminished by the insignificant correlations and/or by likewise significant but opposite ones at other scales, resulting in a poor correlation at the measurement scale (Si, 2003; Hu and Si, 2013).

The behavior of the scale-dependent correlations of ln($K_s$) with WAS stands in strong contrast to the relationships with MWD. At the scales corresponding to IMF3, IMF4, and the residue, ln($K_{-5}$) was significantly correlated with both MWD and WAS ($p < 0.01$), and the absolute values of corresponding coefficients were larger for the latter. However, owing to the statistically insignificant interactions at the other two scales identified, i.e., IMF1 and IMF2, WAS exhibited no significant correlation with ln($K_s$) at the measurement scale. By comparison, although an insignificant correlation was detected in the IMF2s between ln($K_{-5}$) and MWD, the IMF1 of ln($K_{-5}$), which accounted for the largest proportion of the total variance (Fig. 3a), was significantly correlated with the IMF1 of MWD ($p < 0.05$), leading to a uniquely significant correlation at the scale of measurement (Table 2).

According to the magnitude of the correlation coefficients, the small-scale variations of ln($K_s$) and ln($K_{-10}$) manifested in the IMF1s, IMF2s, IMF3, IMF4, and the residue. Moreover, in the latter case, the signs of significant correlations between ln($K_s$) and MWD reversed for different scales, which is also manifested in some other pairs of variables. These results agreed with those of She et al. (2015, 2017) and Yang et al. (2019), who also detected insignificant linear correlations of $K_s$ with associated soil properties at the measurement scale but significant correlations for one or more scale components extracted by MEMD or NA-MEMD. A possible reason for this phenomenon could be that the significant interactions detected at some scales associated with particular IMFs or the residue tend to be diminished by the insignificant correlations and/or by likewise significant but opposite ones at other scales, resulting in a poor correlation at the measurement scale (Si, 2003; Hu and Si, 2013).
IMF2s, and IMF3s were dominated by SOC. The corresponding coefficients were all positive, suggesting the enhancement of soil structure and soil macropores by elevated SOC content. The larger scale variations manifested in the IMF4s were most strongly correlated with CLAY (Table 2), while the negative correlations might be explained by the decrease in pore size with increasing fine particle contents. It is interesting that no significant correlation was observed for either ln(K) with BD, which has been widely adopted in the PTFs for K_s (Wösten et al., 2001; Vereecken et al., 2010). Because hydraulic conductivity depends not only on the number and size of pores but also on their geometry and connectivity, BD reflects only the total porosity but not the relevant geometric properties and therefore does not necessarily regulate K at or near saturation.

The case for ln(K−5) and ln(K−10) is more complicated than that for ln(K) closer to saturation. The variations at the smallest scale of ~12 m decomposed in the IMF1s were dominated by MWD (Table 2), but the corresponding negative correlations were unexpected because the greater values of MWD usually indicate larger sizes of aggregates and interaggregate pores and thereby higher K near saturation. The dry aggregates >2 mm contributed >50% of the total soil weight (data not shown) and largely determined the magnitude of the MWD. A probable explanation for the negative association between Kns and MWD is the occurrence of slaking within these relatively large and unstable aggregates during the infiltration experiments for K measurement, which tended to reduce the pore space and to impair hydraulic conductivity (Lado et al., 2004; Ben-Hur et al., 2009). The variations manifested in the other IMF5s were mainly controlled by WAS or SAND (Table 2), except that no significant correlation was detected between the IMF2 of ln(K−5) and any soil property investigated. This result suggested little or no probability of a linear regression for the corresponding component. The residual ln(K), no matter at which h it was measured, was significantly associated with all six soil properties (p < 0.01), and the absolute values of the corresponding coefficients were all close to 1.

In general, soil structure had a greater impact on the hydraulic conductivities at and near saturation than soil texture. The small-scale variations that dominated the spatial ln(K) distributions were regulated either by SOC indirectly, reflecting soil structure, or by MWD and WAS, which directly represented soil structure. These findings emphasize the necessity of including soil structural

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Intrinsic mode function</th>
<th>BD</th>
<th>MWD</th>
<th>WAS</th>
<th>SAND</th>
<th>CLAY</th>
<th>SOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln(K_s) Before NA-MEMD</td>
<td>-0.142</td>
<td>-0.125</td>
<td>0.147</td>
<td>0.157</td>
<td>-0.313*</td>
<td>0.522**</td>
<td></td>
</tr>
<tr>
<td>IMF1</td>
<td>-0.245</td>
<td>0.166</td>
<td>0.030</td>
<td>0.128</td>
<td>-0.203</td>
<td>0.360*</td>
<td></td>
</tr>
<tr>
<td>IMF2</td>
<td>-0.250</td>
<td>0.077</td>
<td>0.238</td>
<td>0.264</td>
<td>-0.218</td>
<td>0.653**</td>
<td></td>
</tr>
<tr>
<td>IMF3</td>
<td>0.201</td>
<td>-0.785**</td>
<td>0.049</td>
<td>0.397**</td>
<td>-0.038</td>
<td>0.863**</td>
<td></td>
</tr>
<tr>
<td>IMF4</td>
<td>-0.169</td>
<td>-0.545**</td>
<td>-0.540**</td>
<td>-0.616**</td>
<td>-0.708**</td>
<td>0.499**</td>
<td></td>
</tr>
<tr>
<td>Residue</td>
<td>-0.997**</td>
<td>0.997**</td>
<td>0.994**</td>
<td>0.995**</td>
<td>-0.990**</td>
<td>1.000**</td>
<td></td>
</tr>
<tr>
<td>ln(K−1) Before NA-MEMD</td>
<td>-0.119</td>
<td>-0.165</td>
<td>0.123</td>
<td>0.142</td>
<td>-0.305*</td>
<td>0.499**</td>
<td></td>
</tr>
<tr>
<td>IMF1</td>
<td>-0.275</td>
<td>0.157</td>
<td>0.057</td>
<td>0.107</td>
<td>-0.217</td>
<td>0.365*</td>
<td></td>
</tr>
<tr>
<td>IMF2</td>
<td>-0.169</td>
<td>-0.328*</td>
<td>0.147</td>
<td>0.200</td>
<td>-0.222</td>
<td>0.555**</td>
<td></td>
</tr>
<tr>
<td>IMF3</td>
<td>-0.028</td>
<td>-0.646**</td>
<td>0.166</td>
<td>0.386**</td>
<td>-0.107</td>
<td>0.895**</td>
<td></td>
</tr>
<tr>
<td>IMF4</td>
<td>0.107</td>
<td>-0.531**</td>
<td>-0.630**</td>
<td>-0.591**</td>
<td>-0.664**</td>
<td>0.327*</td>
<td></td>
</tr>
<tr>
<td>Residue</td>
<td>-0.993**</td>
<td>0.989*</td>
<td>0.979**</td>
<td>0.983**</td>
<td>-0.967**</td>
<td>0.996**</td>
<td></td>
</tr>
<tr>
<td>ln(K−5) Before NA-MEMD</td>
<td>0.113</td>
<td>-0.426**</td>
<td>-0.158</td>
<td>-0.057</td>
<td>-0.015</td>
<td>0.010</td>
<td></td>
</tr>
<tr>
<td>IMF1</td>
<td>-0.124</td>
<td>-0.377**</td>
<td>0.028</td>
<td>-0.003</td>
<td>-0.031</td>
<td>-0.052</td>
<td></td>
</tr>
<tr>
<td>IMF2</td>
<td>0.035</td>
<td>-0.189</td>
<td>0.019</td>
<td>0.148</td>
<td>0.089</td>
<td>-0.040</td>
<td></td>
</tr>
<tr>
<td>IMF3</td>
<td>0.596**</td>
<td>-0.389**</td>
<td>-0.633**</td>
<td>0.052</td>
<td>0.132</td>
<td>0.433**</td>
<td></td>
</tr>
<tr>
<td>IMF4</td>
<td>0.629**</td>
<td>-0.899**</td>
<td>-0.925**</td>
<td>-0.959**</td>
<td>-0.823**</td>
<td>-0.455**</td>
<td></td>
</tr>
<tr>
<td>Residue</td>
<td>0.993**</td>
<td>-1.000**</td>
<td>-1.000**</td>
<td>0.999**</td>
<td>-0.998**</td>
<td>0.997**</td>
<td></td>
</tr>
<tr>
<td>ln(K−10) Before NA-MEMD</td>
<td>0.353</td>
<td>-0.448**</td>
<td>-0.182</td>
<td>0.023</td>
<td>0.052</td>
<td>-0.048</td>
<td></td>
</tr>
<tr>
<td>IMF1</td>
<td>-0.249</td>
<td>-0.378*</td>
<td>0.058</td>
<td>0.012</td>
<td>0.051</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>IMF2</td>
<td>0.075</td>
<td>-0.397**</td>
<td>0.090</td>
<td>0.566**</td>
<td>0.228</td>
<td>0.077</td>
<td></td>
</tr>
<tr>
<td>IMF3</td>
<td>0.557**</td>
<td>-0.172</td>
<td>-0.711**</td>
<td>0.033</td>
<td>0.223</td>
<td>0.159</td>
<td></td>
</tr>
<tr>
<td>IMF4</td>
<td>0.756**</td>
<td>-0.974**</td>
<td>-0.907**</td>
<td>-0.980**</td>
<td>-0.873**</td>
<td>-0.492**</td>
<td></td>
</tr>
<tr>
<td>Residue</td>
<td>0.997**</td>
<td>-1.000**</td>
<td>-0.998**</td>
<td>-0.999**</td>
<td>0.994**</td>
<td>-1.000**</td>
<td></td>
</tr>
</tbody>
</table>

* Correlation is significant at the 0.05 level.
** Correlation is significant at the 0.01 level.
Hydraulic Conductivity Estimations before and after Multivariate Empirical Mode Decomposition

Using undecomposed data series, the stepwise MLR procedure suggested SOC or MWD for \( \ln(K) \) estimation at different \( h \) (Table 3). Owing to the rather poor linear correlations of \( \ln(K) \) with soil properties (Table 2), all the resulting MLR models accounted for no more than 25\% of the total \( \ln(K) \) variance during calibration and an even smaller proportion of variance for the subsequent validation. An extreme case exhibited in the estimation of \( \ln(K_{-5}) \). The MLR model derived from the calibration set resulted in a negative \( R^2_{adj} \) in validation, suggesting that the MLR estimates deviated from the primary trend of the \( \ln(K) \) distribution.

The same procedure was adopted to derive MLR models for each \( \ln(K) \) component at different spatial scales. As the intercepts obtained in the regression equations were not statistically significant (\( p < 0.05 \)) for all the IMFs, the corresponding regressions were forced through the origin of (0, 0). As indicated in Table 3, the MLR models almost perfectly regressed \( \ln(K) \) at the two largest spatial scales corresponding to IMF4 and the residue, no matter at which \( h \). The corresponding \( R^2_{adj} \) for both calibration and validation were typically >0.98 except one of 0.902 for \( \ln(K_{-5}) \) in validation, and the RMSE values were all <0.015. The performance of the MLR models for the IMF3 was not as satisfactory as those at the larger scales. The corresponding \( R^2_{adj} \) ranged from 0.7 to 0.9 and the RMSE fell between 0.01 and 0.04 for both calibration and validation except two \( R^2_{adj} \) of 0.496 and 0.597 obtained for the \( \ln(K_{-5}) \) and \( \ln(K_{-10}) \) validations, respectively.

At the remaining two smallest scales, the MLR models were only slightly better or even worse than the one at the measurement scale for each \( \ln(K) \), as only 13.7 to 43.6\% of the variance was explained by such models in calibration. A negative \( R^2_{adj} \) was also obtained, i.e., in the validation of \( \ln(K_{-5}) \) at the smallest scale.
associated with IMF1. Moreover, MLR failed to regress either IMF1 or IMF2 of ln($K$) based on the soil property components of the same index and basically precluded the MLR estimation of ln($K$) at the scale of measurement via incorporating NA-MEMD.

In other words, none of the six soil properties satisfied the entering criterion of the MLR stepwise procedure, i.e., $F$-to-enter of 0.05. A possible explanation for the poor performance of MLR in estimating the IMF1s and IMF2s was that the corresponding small-scale variation of each ln($K$) was presumably non-linearly, rather than linearly, correlated with the soil properties of the same scale (Yang et al., 2019). If this explanation could not be verified, the estimation of ln($K$) at such small scales may need the involvement of additional soil processes such as pore continuity, macroporosity, and/or environmental factors such as terrain attributes (Agyare et al., 2007).

To verify the reasoning above, the artificial neural network (ANN) method that relates predictors to predictands, typically via nonlinear functions, was applied to estimate the IMF1 and IMF2 of each ln($K$) from the same index IMF1s of the six soil properties. It is obvious that the ANN approach remarkably improved the prediction quality of both IMF1 and IMF2 of each ln($K$) (Table 4). Considering all the combinations of three soil properties, the optimal ANN model was able to explain at least 80% of the total variance in both calibration and validation for either IMF. Two exceptions lay in the validations of the IMF1s of ln($K_s$) and ln($K_{-10}$), which had $R^2_{adj}$ values of 0.554 and 0.742, respectively.

The ANN estimates of IMF1 and IMF2 were then added to the MLR estimates of IMF3, IMF4 and the residue to obtain another estimate of each ln($K$) at the measurement scale. Such prediction apparently surpassed the one made by MLR at the scale of measurement before NA-MEMD (Fig. 5). The former was capable of explaining additional ~74% of the total ln($K$) variance in calibration, whereas in the validation, the proportions accounted for by the former ranged from 67.8% for ln($K_s$) to 93.0% for ln($K_{-5}$), i.e., on average 73.4% more than those by the latter. When plotted against the measurements, the results were distributed more closely to the 1:1 line (Fig. 5). These findings imply that the six soil properties were capable of simulating the spatial variation of hydraulic conductivities near and at saturation, i.e., not via simple linear functions but through more complex networks involving typically nonlinear ones, especially at the relatively small scales. The impact of other soil properties and environmental factors that have been identified by previous process-based studies might have already been embedded in the scale-dependent relationships of different $K$ with the six properties investigated.

**Fig. 5.** Measurements vs. estimates made by multiple linear regression (MLR) before noise-assisted multivariate empirical mode decomposition (NA-MEMD) and by a combination of MLR and an artificial neural network (ANN) after NA-MEMD for calibration and validation of (a,b) saturated hydraulic conductivity [ln($K_s$)], (c,d) $K$ at the pressure head of −1 cm [ln($K_{-1}$)], (e,f) $K$ at the pressure head of −5 cm [ln($K_{-5}$)], and (g,h) $K$ at the pressure head of −10 cm [ln($K_{-10}$)]. With regard to the combination of MLR and ANN after NA-MEMD, the intrinsic mode functions IMF1 and IMF2 were estimated by ANN and added to the estimates of IMF3, IMF4, and the residue made by MLR for each ln($K$). The adjusted coefficient of determination for the estimations made before ($R^2_{adj, be}$) and after NA-MEMD ($R^2_{adj, af}$) are displayed in each subplot.

**Conclusions**

Noise-assisted multivariate empirical mode decomposition was applied to decompose and examine the scale-dependent
relationships of soil $K$ at and near water saturation, i.e., $K_s$, $K_{-1}$, $K_{-5}$, and $K_{-10}$, with different soil properties on a rectangular field evenly distributed across cropland and grassland. Applying NA-MEMD, the spatial variability of each $K$ was decomposed into four IMFs and a residue and was found to be significantly associated with all six soil properties investigated at various numbers of spatial scales, i.e., ranging from one scale between $K_s$ or $K_{-1}$ and BD to all five scales including the one for residue between $K_s$ or $K_{-1}$ and SOC, although most of the soil properties exhibited no significant correlation with $K$ at the measurement scale. In general, $K$ measured at all four pressure heads were primarily affected by soil structure rather than soil texture, as the relatively small-scale variations that dominated $K$ distributions were mainly regulated by the indicators of soil structure.

Based on these scale-dependent relationships, each $K$ component was predicted by the soil property components of the equivalent spatial scale. However, MLR failed to regress either IMF1 or IMF2 of each $K$ using the current stepwise procedure. The model, if derived at these two smallest scales, performed poorly, i.e., accounting for no more than 50% of the $K$ variance in calibration and <30% of variance in validation. An ANN was then adopted to replace MLR for the estimation of these two components and yielded substantially enhanced results for the corresponding IMFs. When adding the ANN estimates of IMF1 and IMF2 together with the MLR ones for the other components, the estimation of each $K$ at the measurement scale was apparently enhanced in both calibration and validation by accounting for additional 74.4 and 73.4%, on average, of the total variance, respectively. These findings, on one hand, demonstrate the superiority of NA-MEMD in characterizing scale-dependent relationships among soil properties and in assisting the estimation of soil hydraulic properties. On the other hand, they reveal the nonlinear interactions between $K$ and associated soil properties at relatively small spatial scales and the necessity of incorporating ANN or some rather simple nonlinear functions with NA-MEMD in related estimations.

Acknowledgments

We thank Riley J. Walton and Ann Freytag for their technical assistance in the laboratory and in the field. This study was funded by the National Natural Science Foundation of China (no. 41601277, 41571130082), USDA–NIFA 6440–32630–001–035 and Hatch Project KY006045–0210417, and the Project supported by State Key Laboratory of Earth Surface Processes and Resource Ecology (no. 2017–ZY–09).

References


Table 4. Root mean square error (RMSE) and adjusted coefficient of determination ($R^2_{adj}$) of optimal artificial neural network (ANN) models for intrinsic mode functions IMF1 and IMF2 of saturated hydraulic conductivity and conductivity at pressure heads of −1, −5, and −10 cm ($ln(K_s)$, $ln(K_{-1})$, $ln(K_{-5})$ and $ln(K_{-10})$, respectively) in calibration and validation.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>IMF1</th>
<th>IMF2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Calibration</td>
<td>Validation</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>$R^2_{adj}$</td>
</tr>
<tr>
<td>$ln(K_s)$</td>
<td>0.021</td>
<td>0.987</td>
</tr>
<tr>
<td>$ln(K_{-1})$</td>
<td>0.028</td>
<td>0.975</td>
</tr>
<tr>
<td>$ln(K_{-5})$</td>
<td>0.019</td>
<td>0.982</td>
</tr>
<tr>
<td>$ln(K_{-10})$</td>
<td>0.022</td>
<td>0.978</td>
</tr>
</tbody>
</table>