2950 Niles Road, St. Joseph, MI 49085-9659, USA
269.429.0300 fax 269.429.3852 hq@asabe.org www.asabe.org

An ASABE Meeting Presentation
DOI: https://doi.org/10.13031/aim.202200425
Paper Number: 2200425

Use of Self-Organizing Maps to Estimate Furrow Sediment Loss in Western U.S.

B.A. King & D.L. Bjorneberg
USDA ARS Northwest Irrigation and Soils Research Laboratory, 3793 N. 3600 E. Kimberly, Idaho 83341

Written for presentation at the
2022 ASABE Annual International Meeting
Sponsored by ASABE
Houston, Texas
July 17–20, 2022

ABSTRACT. *The area irrigated by furrow irrigation in the U.S. has been steadily decreasing but still represents about 20% of the total irrigated area in the U.S. Furrow irrigation sediment loss is a major water quality issue in the western U.S. and a method for estimating sediment loss is needed to quantify the environmental impacts and estimate effectiveness and economic value of conservation practices. The objective of the study was to investigate the use of the unsupervised machine learning technique Kohonen self-organizing maps (KSOM) to predict furrow sediment loss. Historical published and unpublished data sets containing measurements of furrow irrigation sediment loss in the western U.S. were assembled into a furrow sediment loss data set comprising over 2000 measurements. Despite the immunity of KSOMs to measurement variability, the inherent variability in measured furrow sediment loss limited the ability of a KSOM model to reliability predict furrow sediment loss. Furrow sediment loss was under predicted by 44% on average with a linear regression coefficient of determination of 0.6. The KSOM model was placing little weight on measured sediment loss in the input data set, indicating that it was clustering the data based on input parameters defining the hydraulic and soil conditions. This outcome was used to develop a transfer learning approach for predicting furrow sediment loss. The transfer learning approach used a KSOM to cluster data records of similar hydraulic and soil conditions in the data set. Mean measured sediment loss and furrow flow rate of each cluster was determined based on data set vectors assigned to a cluster by the KSOM. Furrow sediment loss prediction was obtained by applying an input vector to the KSOM to identify the cluster the input vector most closely matches. Then the mean measured sediment loss of the identified cluster was adjusted for any difference between the input vector furrow flow rate and cluster mean furrow flow rate to obtain a prediction of furrow sediment loss. Predicted furrow sediment loss was 16% less than measured sediment loss on average with a coefficient of determination of 0.82. When the data set was randomly split into model development (90%) and validation (10%) data sets the prediction results were similar.*

Keywords. *Artificial intelligence, Furrow irrigation, Model, Sediment loss, Self-organizing map, Surface irrigation, Transfer learning.*

Introduction

Surface irrigation (gravity irrigation), where the force of gravity causes water to flow downslope across the land surface and infiltrate, has been slowly decreasing in the U.S., but still represents about 36% of the total irrigated area or 8.1 Mha (USDA-NASS, 2018). Furrow irrigation is one form of surface irrigation practiced on row crops where water is introduced

at the upslope end of a field into small equally spaced, man-made channels between crop rows, which infiltrates as it flows down slope, storing water in the root zone for crop growth. In the 10 western U.S. states (AZ, CA, CO, ID, MT, NM, OR, UT, WA, WY), 1.4 Mha or 14% of the irrigated area remains under furrow irrigation. Water inflow rate to the furrow must be sufficient to ensure water advances to the end of the field in a reasonable amount of time to allow sufficient time for water to infiltrate along the full length of the furrow. Water in excess of infiltration flows off the field and often enters nearby water bodies. As water flows in furrows, it detaches soil particles which are transported down slope. Furrow flow rate decreases down slope due to infiltration such that suspended soil particles can no longer be transported and are deposited in the furrow. Water leaving the field carries sediment along with absorbed chemicals which degrades the quality of receiving water bodies. Soil loss from furrow irrigation often exceeds 2 to 11 Mg ha^{-1} (Koluvek et al., 1993) and rates of up to 100 Mg ha^{-1} have been measured in experimental studies (Berg and Carter, 1980; Evans et al., 1995; Trout, 1996; Fernández-Gómez et al., 2004). Crop yield reductions up to 25% have been documented at upslope end of fields from 80 years of furrow irrigation induced erosion (Berg and Carter, 1980; Carter et al., 1985; Carter, 1993). Furrow irrigation sediment loss is a major water quality issue and a method for estimating sediment loss is needed to quantify the environmental impacts and estimate effectiveness and economic value of conservation practices.

Acquiring field data for developing process-based furrow erosion models is challenging (Mateos and Giráldez, 2005). Physical models used to predict furrow erosion are based on predominately empirical equations used to model rainfall-induced rill erosion (Bjorneberg et al., 2000). The conditions used during experimental development of governing equations limits use of the equations to field conditions representative of the experimental conditions. The physical conditions of rill erosion under rainfall runoff differ from furrow irrigation in several aspects (Bjorneberg et al., 2000). For example, water initially flows on dry soil during furrow irrigation, but under rainfall conditions the rill surface is wet before water flow begins. Instantaneous wetting of soil aggregates replaces air absorbed on internal soil particle surfaces, which can break apart soil aggregates (Carter, 1990) increasing soil erodibility. This is one possible reason why furrow erosion often occurs with less than critical hydraulic shear (Kemper et al., 1985). Another difference between rill and furrow erosion is that relatively clean water is introduced into furrows while sediment laden water enters rills under rainfall erosion. Additionally, furrow flow rate decreases with distance, which is typically opposite for rills with rainfall runoff. Lastly, flow in furrows usually lasts 12 to 24 hours, considerably longer than a typical rainfall-runoff event. Initially, sediment detachment during furrow irrigation may be like rill erosion, but after several hours head cuts and side cuts become important mechanisms for sediment detachment (Bjorneberg et al., 2000) which are not considered in rainfall induced rill erosion. In a review of erosion and sedimentation processes in furrow irrigation from field studies, Trout and Neibling (1993) found that the process of detachment, transport, and deposition occurring in furrow irrigation are not adequately quantified by rill erosion equations based on hydraulic shear. While rill erosion-based equations do not adequately predict erosion for furrow irrigation they do provide insight into the factors and important relationships (Fernández-Gómez et al., 2004). Trout and Neibling (1993) concluded that while process-based models are important for understanding furrow erosion processes, current models can predict furrow sediment loss no better than empirical models relating sediment loss to measurable hydraulic parameters such as slope, flow rate and soil characteristics such as texture. Bjorneberg et al. (1999) evaluated furrow irrigation erosion predicted by the Water Erosion Prediction Project (WEPP) model (Nearing et al., 1989), a process-based model and concluded that the steady-state WEPP was not applicable to furrow erosion based on field evaluations for a single soil. Baseline erodibility and critical shear values developed under rainfall runoff conditions were higher than calibrated values for furrow irrigation. Sediment deposition based on transport capacity also did not match measured deposition in furrows. Bjorneberg et al. (2010) reviewed the current status of furrow sediment loss prediction and concluded that an empirical model may be as good as or better than a process-based model where the parameters cannot be quantified for field conditions.

The lack of an adequate process-based model for predicting furrow irrigation sediment loss has led to the development of empirical models for over half a century. Koluvek et al. (1993) noted that the first published equation for predicting furrow erosion was based on research in Utah U.S., which directly related erosion to exponential functions of furrow flow rate and slope with exponents greater than one (Israelsen, 1946). Regression models have been developed to predict sediment loss with limited success (Fornstrom and Borrelli, 1985; O'Donkor, 1978). More recently, an empirical surface irrigation soil loss (SISL) model was developed by the Idaho Natural Resources Conservation Service (NRCS) in 1991 to estimate annual soil loss from furrow irrigated fields to assess benefits of conservation practices, such as converting from furrow to sprinkler irrigation (Bjorneberg et al., 2007). An evaluation of the SISL model with measured furrow irrigation erosion from two studies in Idaho and one in Washington showed that the model predicted the relative effects of conservation practices of straw mulching and surge irrigation reasonably well, however, the absolute differences between measured and predicted soil loss were sometimes large. Furrow sediment loss is inherently highly variable, which makes model development challenging. Sediment loss from adjacent field furrows under identical hydraulic and field conditions can vary by a factor of eight for mean sediment losses < 20 kg and by a factor of two for mean sediment losses > 200 kg (fig.1).

Machine learning represents a collection of relatively new computer algorithms for discovering relationships or making inferences from sample data. Machine learning is finding uses in many branches of science, including agriculture. The goal of machine learning algorithms is to have the computer discover unknown relationships in a data set to optimize a feedback

function. The nature of the feedback function is used to classify machine learning algorithms in to three general categories: 1) supervised learning; 2) unsupervised learning; and 3) reinforcement learning. Supervised learning is the most common form of machine learning where the algorithm adjusts model parameters to optimize an output. Linear regression and feed-forward neural networks (FFNN) are examples of supervised learning where model parameters are adjusted to minimize the sum of square errors between model predications and a measured value. King et al. (2016) used a feed-forward artificial neural network (FFNN) to model furrow sediment loss resulting in an $R^2 = 0.71$ between predicted and measured sediment loss. The desire to maximize prediction performance resulted in an over-trained model where the model learned the data set rather than underlying physical relationships representing the furrow sediment loss process. The high degree of variability in measured furrow sediment loss prevented the model from learning underlying physical processes. The resulting NN model provided nonsensical estimates of furrow sediment loss when new input data was entered into model. For example, decreasing sediment loss with increasing flow rate. It is unlikely that any type of supervised artificial intelligence model will provide good estimates of furrow sediment loss due to the inherent natural variably in the furrow sediment loss process. Unsupervised learning algorithms optimize a measure of data structure to discover hidden relationships in the data rather than minimize prediction error of a parameter. Kohonen self-organizing maps (KSOM) are a good example of unsupervised machine learning. The KSOM is a unique type of neural network able to convert complex, nonlinear statistical relationships between high-dimensional data items into simple geometric relationships on a low-dimensional display, usually 2-dimensional. KSOMs use competitive learning related to neighborhood distances between inputs to preserve the topological properties of the measured input data set. The KSOM is a clustering algorithm primarily used for visualization of relationships between measured system variables or identification of outliers. Recently, KSOMs have been used for prediction based on measured values associated with a best matching neuron. For example, Rustum et al. (2008) adapted a KSOM as a software sensor to predict biochemical oxygen demand of municipal water streams. The KSOM provided good agreement with conventional 5-day bioassay method in much more timely fashion. Adeloye et al. (2011) used a KSOM to predict reference crop evapotranspiration (ET_r) in good agreement with the FAO Penman-Monteith model with much less meteorological input. The KSOM-based estimates of ET_r were significantly superior to recommended empirical methods for data scarce situations. Kumar et al. (2020) compared use of FFNN and KSOM to predict crop water stress index of Indian mustard using only measured air temperature, relative humidity, and measured canopy temperature. The KSOM performed much better than FFNN with the limited data set.

A KSOM is typically represented by a two-dimensional grid representing nodes or neurons tuned to different patterns of a multidimensional input vector. The principal goal of the KSOM is to transform a multidimensional input vector into a two-dimensional discrete map while preserving the most important topological and metric relationships of the input data. To train the KSOM the first step is to normalize the input data to a common range using one of several techniques so that each input variable has equal weight, regardless of numerical magnitude. Next, a normalized input vector is chosen at random and presented to each of the individual neurons for comparison with their code vectors, to identify the code vector that is the most like the chosen input vector. The identification of the most similar neuron is based on some measure of spatial similarity with Euclidian distance being the most common, which is defined as:

$$D_i = \sqrt{\sum_{j=1}^{n} (x_j - m_{ij})^2} \tag{1}$$

where
D_i = Euclidian distance between the input vector and the weight vector of neuron i
n = dimensionality of the input vector
x_{ij} = jth element of the current input vector
m_{ij} = jth element of the neuron weight vector i
M = number of neurons (nodes) in the KSOM.

The neuron whose weight vector most closely matches the input data vector (minimum D_i) is chose as the winning node and is called the best matching unit (BMU). The vector weights of the BMU neuron and those of spatially adjacent neurons are adjusted to more closely match the input vector using an adjustment algorithm that moves the neurons closer spatially (Adebayo et al., 2011). This process allows each neuron in the map to recognize input vectors like itself. This process gives rise to the term "self-organizing" because no external information is used to lead to a classification (Penn, 2005). A detailed mathematical and theoretical description of KSOM and its implementation can be found in Vesanto et al. (2000), Rustum et al. (2008) and Adeloye et al. (2012).

The development of a KSOM requires selecting the total number of neurons to include and the row and column arrangement of a two-dimensional map. There are no strict rules on the number of neurons or row and column arrangement to use for the best KSOM. Guidance for initial development of a KSOM was suggested by García and González (2004) with

the optimum total number of neurons being:

$$M = 5\sqrt{N} \tag{2}$$

where
N = total number of input vectors

and the number of rows and columns was estimated as:

$$\frac{l_1}{l_2} = \sqrt{\frac{e_1}{e_2}} \tag{3}$$

where
l_1 and l_2 are the number of rows and columns, respectively
e_1 is the biggest eigenvalue of the training data set
e_2 is the second biggest eigenvalue

The total number of nodes and row and column arrangement can be adjusted to optimize performance of the KSOM for the desired outcome.

The successful development of a KSOM results in a matrix of weight values for each neuron (node) in the two-dimensional map. This matrix is called the codebook, which contains the captured spatial and metric elements of the data set. Once the codebook is obtained, a new normalized data vector can be applied to the codebook using eqn. 1 to determine which neuron is the BMU. In the case of a classification problem, the classification of the BMU is the model predicted classification for the new input vector. A KSOM can also be used to predict unknown input vector values as indicted in figure 3. A new normalized data vector with missing value(s) is applied to the code book using eqn. 1 to find the BMU. The neuron weight values of the BMU are the estimated normalized value(s) represented by the input vector. The normalized weight value for the missing input vector parameter is the KSOM normalize estimate for the missing parameter, which then needs to be denormalized to arrive at the actual value estimate for the missing input parameter. In this study, furrow sediment loss is the missing new input vector parameter.

Table 1. Data sources used in this study along with general location where the data was collected, and number of furrows measured.

Data source	General location	Number of furrows
Fornstrom and Borrelli (1985)	Worland, Powell and Torrington, WY	745
King et al. (1984)	Royal City, WA	191
Tunio (1994)	Ontario, OR	108
O'Donkor (1978)	Kimberly, ID	130
Trout (personal communication)	Patterson, CA	318
Bjorneberg and Trout (personal communication)	Kimberly, ID	592

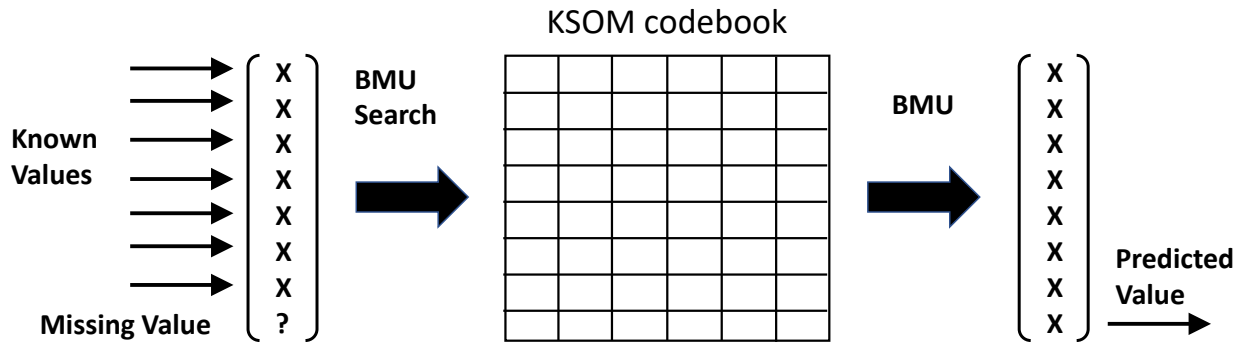

Figure 2. Process for prediction of an unknown value in an input vector using a KSOM. Adapted from Rustum and Adeloye (2007).

The high degree of variability in measured furrow sediment loss requires that modelling algorithms be immune to random

noise in the data set. The use of spatial relationships rather than a measured parameter (sediment loss) for discovering hidden relationships in a data set makes KSOM an attractive option for estimating furrow sediment loss. The objective of this study was to investigate use of KSOM to model and predict furrow sediment loss using measured data from several western U.S. states.

Methods and Materials

Database

Measured furrow irrigation sediment loss data used in this study were obtained from a combination of published research reports, theses, dissertations, and personal communications with research personnel involved in furrow irrigation research spanning four decades. Specific data sources used in this study along with general study locations are listed in Table 1. Collectively, over 2000 furrow sediment loss values were obtained covering a wide range of soil, field, and row crops with most of the data obtained from Fornstrom and Borrelli (1985), Trout (personal communication), and Bjorneberg (personnel communication). The collective data set included nine common hydraulic and field condition variables: freshly cultivated or previously eroded by prior irrigation; compacted by wheel traffic or uncompacted; irrigation duration (T, h); furrow length (L, m); furrow inflow rate (Q, L min^{-1}); furrow slope (S, %); soil sand and clay fractions (%); and furrow sediment loss (SL, kg). Additional details on the data sources are provided by King et al. (2016).

The collective furrow sediment loss data set represents furrow irrigation-induced erosion across a wide range of hydraulic, soil, field and row crop practices in the western U.S. (Table 2). Mean furrow irrigation sediment loss for the data set was 65 kg per furrow with a standard deviation of 138 kg per furrow, double the mean value, which reflects the wide range in values present in the data set (0 - 1879 kg per furrow). The data set included 1491 values for compacted (wheel) furrows and 666 values for noncompacted (nonwheel) furrows. There were 999 values for freshly tilled furrows and 1088 values for previously eroded furrows. The predominate soil type was silt loam.

Table 2. Mean, standard deviation, maximum, and minimum of furrow hydraulic and soil characteristics in the data set used in this study.

Parameter	Mean	Standard deviation	Maximum	Minimum
Furrow length (m)	197	109	549	27
Furrow slope (%)	1.5	1.1	7	0.2
Furrow inflow (L min^{-1})	27.5	14.0	118.4	3.2
Sand fraction (%)	28.8	18.2	77	11
Clay fraction (%)	18.0	7.6	48	2
Sediment loss (kg per furrow)	65	138	1879	0

KSOM modeling

Two MATLAB (MATLAB, MathWorks, Natick, Mass.) software packages were used to design and train the KSOM in this study. These were the Neural Network Toolbox (MathWorks, Natick, Mass.) and the SOM toolbox (version 2.1) (Vesanto et al., 2000; Vatanen et al., 2015). The SOM toolbox is freely available and can be downloaded from https://github.com/ilarinieminen/SOM-Toolbox. Each software package has advantages and disadvantages. For example, the SOM toolbox includes equations 2 and 3 for estimating the optimal size of the KSOM map while the Neural Network Toolbox does not. The default values for neighborhood and the learning rate of both software packages were used in this study. Euclidian distance (eqn. 1) was selected as the distance measure for determining the most similar neuron in both software packages. The data was linearly normalized to a range of 0 to 1 using the maximum and minimum values for the numerical parameters in the data set (table 2). Furrow surface condition information (compacted, noncompacted, tilled, eroded) is qualitative data, which cannot be in a utilize directly in a KSOM model as Euclidian distance (eqn. 1) cannot be calculated for qualitative parameters. To address this issue, furrow condition was converted to a single discrete value using a two-digit binary conversion as shown in table 3, which was then used in developing the KSOM model.

Table 3. Furrow surface condition coding scheme for including qualitative furrow condition information in KSOM model.

Compacted 0=no 1=yes	Eroded 0 = no 1 = yes	Furrow surface condition value
0	0	0
0	1	1
1	0	2
1	1	3

Model evaluation

Performance of prediction models were evaluated by comparing predicted furrow sediment loss versus measured furrow sediment loss. Linear regression between predicted and measured sediment loss was used as a performance measure. The linear regression slope was used to determine if predicted values over or under predicted measured sediment loss. The coefficient (R^2) of determine of the regression line was used as a measure of prediction reliability. A lower R^2 indicates higher uncertainty in predicted value.

Data Analysis

Data set management and regression analysis were conducted using MSExcel. Data set statistics were computed using PROC MEANS (SAS 9.4, SAS Institute, 2013). Graphs were constructed using Sigmaplot 14 (Systat Software, San Jose, CA).

Results and Discussion

KSOM performance

Initial investigation into the potential performance a KSOM to predict furrow sediment loss was conducted using the complete data set rather than partitioning the data set into development and validation data sets. The data base available for modeling furrow sediment loss is sparse due to the difficulty involved in collecting on-farm furrow erosion data. Developing a KSOM model using all the full data set provides an indication of the best attainable performance, since using less data for model development with a limited sparse data set will decrease model prediction performance. The size of the hexagonal KSOM based on equations 2 and 3 was 19 x 12 for a total of 228 neurons (nodes). The initial KSOM provided poor prediction of furrow sediment loss (data not shown). Since the data set was sparce, the results were grouped into regions of the hexagonal KSOM, which is ideal for classifying application but undesirable for a prediction application. To address this issue, two additionally derived parameters were included as inputs into the KSOM to disperse the data in the hexagonal map. With furrow irrigation, greater runoff generally results in greater furrow sediment loss since water is the sediment transport mechanism. In general, greater depth of water applied results in greater the runoff volume. The depth of water applied was calculated as:

$$\frac{QT}{L} \tag{4}$$

which has units of L hr min^{-1} m^{-1}. The units were not reduced as it is not necessary for KSOM development. The parameter defined by equation 4 was included as an input parameter to the KSOM.

Stream velocity is an important parameter when trying to quantify stream erosion and sediment transport. For a given flow channel geometry, flow rate and channel slope are key parameters in determining stream velocity. A measure of stream velocity effect on furrow erosion was calculated as:

$$\frac{Q^2S^2}{L} \tag{5}$$

which has units of L^2 m^{-1}. The parameter defined by equation 5 is consistent with the results reported by Israelsen (1946) where furrow erosion was an exponential function of furrow flow rate and slope with exponents greater than one. Furrow length was included in the equation because sediment loss indicated by Q^2S^2 decreases with furrow length due to infiltration decreasing Q. The parameter defined by equation 5 was included as an input parameter to the KSOM.

Sediment loss prediction performance of the KSOM was improved by inclusion of the additional parameters defined by equations 4 and 5. The KSOM model fit to measured furrow sediment loss is shown in figure 3. One notable feature in figure 3 is the horizontal plotting of data points. This is due to the data points being represented by a single node in the KSOM model. As an example, for the top horizontal set of data points in figure 3, measured sediment ranges from 134 to 1034 kg and ranges from 0.1 to 1349 for the horizontal set of data point second from the top in figure 3.

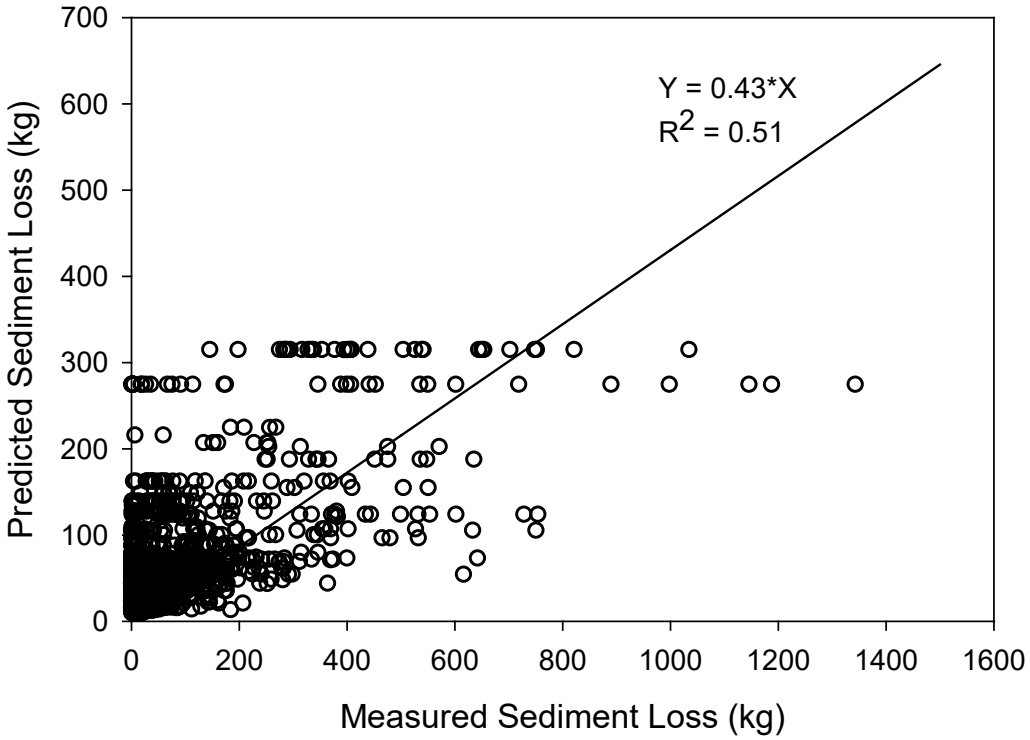

Figure 3. KSOM model predicted furrow sediment loss compared to measured furrow sediment loss.

The hit histogram for the KSOM model is shown in figure 4. The model hit histogram represents the number of times a neuron (node) in the KSOM model is the BMU for a vector in the data set. Some KSOM model nodes are the BMU for more than 35 data vectors, spanning a wide range in sediment loss value, as well as other input vector values. Not all neurons in the KSOM model were a BMU for a vector in the data set (unshaded neurons in figure 4). This is due to the data set being small and sparse.

Since flow rate is the model input parameter that has the greatest influence on furrow sediment loss, it can be used to adjust model predicted sediment loss associated with a given KSOM model neuron and overcome some of the effects of a sparse data set. Model predicted furrow sediment loss was adjusted for flow rate as:

$$SL_{pred} = SL_{BMU}\left[\frac{Q_{input}}{Q_{BMU}}\right]^{X} \tag{6}$$

where

SL_{BMU} is predicted sediment loss for the BMU in the KSOM model, kg

SL_{pred} is predicted sediment loss adjusted for furrow flow rate

Q_{input} is the model input vector value for furrow flow rate, L min^{-1}

Q_{BMU} is the furrow flow rate of the BMU neuron in the KSIM model, L min^{-1}

X is an exponent

The effect of using equation 6 to adjust KSOM predicted sediment loss is shown in figure 5. The value for X was 1.2, which was chosen to maximize the linear regression coefficient of determination (R^2). Adjusting furrow sediment loss of the KSOM model BMU nearly eliminated the presence of equal predicted furrow sediment loss over a wide range in measured sediment loss. The linear regression coefficient of determination for measured versus predicted furrow sediment loss was increased from 0.51 to 0.60 using equation 6. Despite this improvement, KSOM model predicted furrow sediment loss grossly underestimated measured sediment loss as the slope of the regression line was 0.56. A much better estimate of furrow sediment loss is needed to be a useful model.

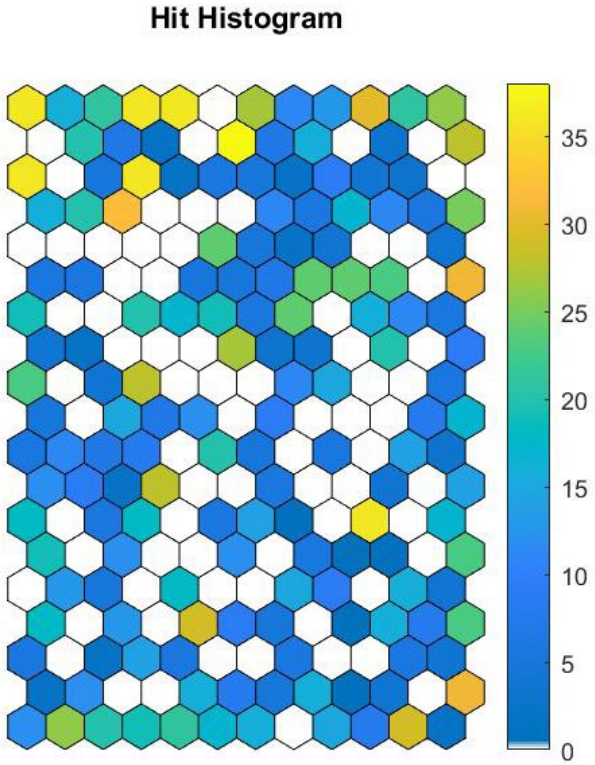

Figure 4. Hit histogram when the data set is applied to the KSOM model for predicting furrow sediment loss. Color of hexagonal node represents number of data set input vectors associated with the node.

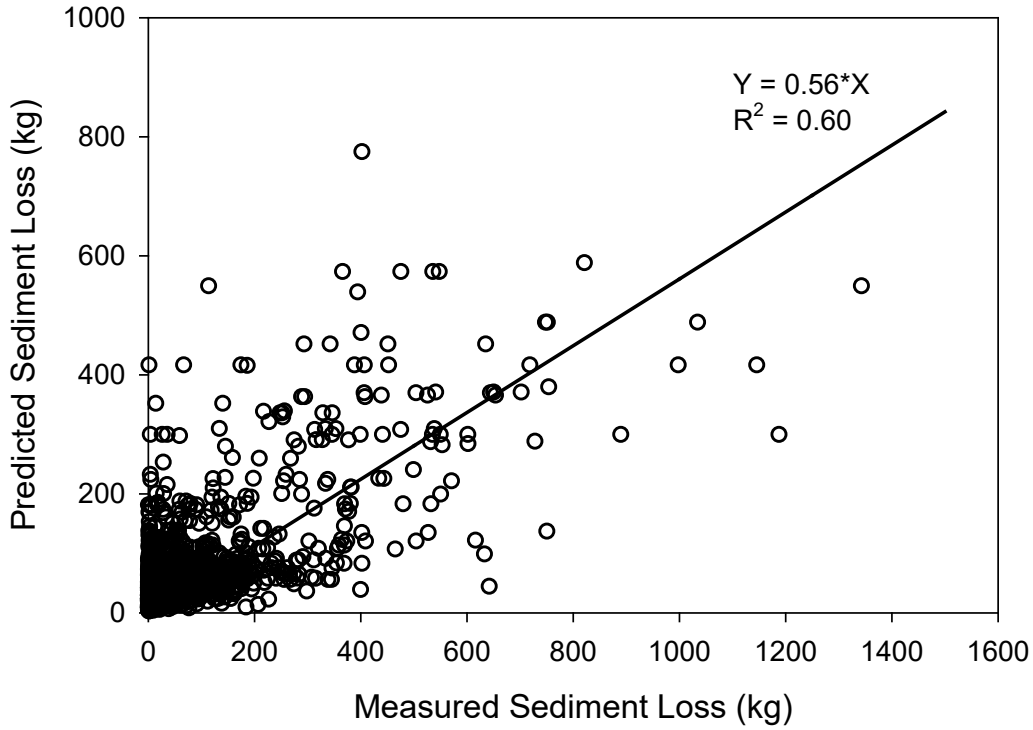

Figure 5. KSOM model predicted furrow sediment loss compared to measured furrow sediment loss fit when adjusted for measured furrow flow rate.

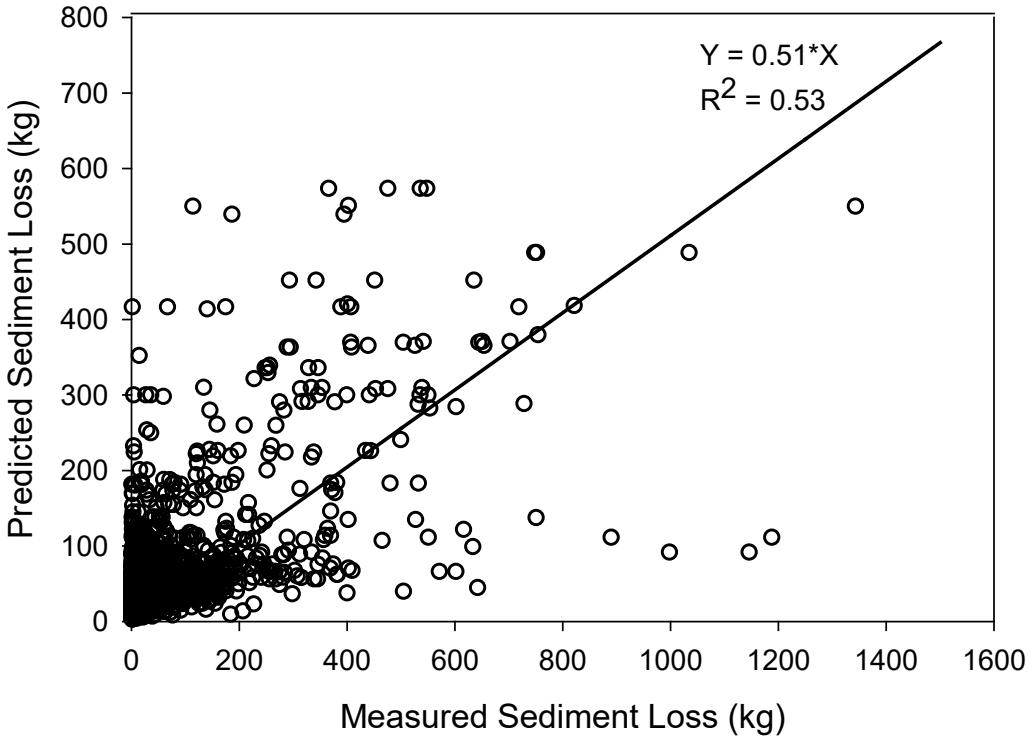

Figure 6. KSOM model predicted furrow sediment loss compared to measured sediment loss when sediment loss is not included in the model input vector.

Predicted KSOM model furrow sediment loss when each data set vector with missing furrow sediment loss (fig. 2) was input to the model compared to measured sediment loss is show in figure 6. Predicted sediment loss in figure 6 was adjusted using equation 6 with X equal to 1.2. Prediction performance of the KSOM model decreased (reduced R^2) when furrow sediment loss was not included in the input vector to the model. The decrease in prediction performance was the result of 42 different BMU neurons compared to when sediment loss was included in the input vector to the KSOM model (fig. 5). The 42 different BMUs selected represents a prediction error rate of only 2.1% for the KSOM model when furrow sediment loss was not included in the input vector to the model. This result indicates that the KSOM model placed little weight on sediment loss for selecting a BMU neuron. This is consistent with the wide range in sediment loss associated with a BMU neuron (Fig. 3). The high degree of variability in measured sediment loss impairs prediction performance of the KSOM model even though it is an unsupervised neural network model.

On approach to decrease the number of BMUs associated with a neuron in a KSOM model is to increase the number of neurons in the model. Hexagonal KSOM models with 20 x 20 and 25 x 25 nodes were tested. Increasing the number of neurons in the hexagonal map did increase KSOM model fit to the data set with the 20 x 20 map having a prediction R^2 of 0.73 and a 25 x 25 map have a prediction R^2 of 0.77. The prediction BMU error rate increased when the number of nodes in the map increased with an error rate of 3.3% and 4.0% for KSOM map sizes of 20 x 20 and 25 x 25, respectively. These small error rates further highlight that the KSOM model put little weight on sediment loss in the input vector for selecting a BMU. This outcome suggests that perhaps sediment loss need not be included as an input to a KSOM.

Transfer learning approach

Transfer learning is a machine learning method where a model trained for one task is used as a starting point for solving a different but related problem. For example, a deep learning model trained to recognize cars could be modified to recognize trucks, leveraging stored knowledge gained in recognizing cars. Since the KSOM model did not heavily weight input furrow sediment loss for predicting sediment loss, the KSOM model was relying on other input parameters to estimate sediment loss. Thus, a KSOM model could be used to cluster the input vectors of the parameters known to affect sediment loss. This knowledge could then be used as input to another data driven model to predict furrow sediment loss and perhaps providing a better estimate of furrow sediment loss.

The transfer learning concept was implemented by training a KSOM model using the complete data set with furrow sediment loss omitted from the model input vector. The resulting KSOM model was used to determine the BMU neuron for

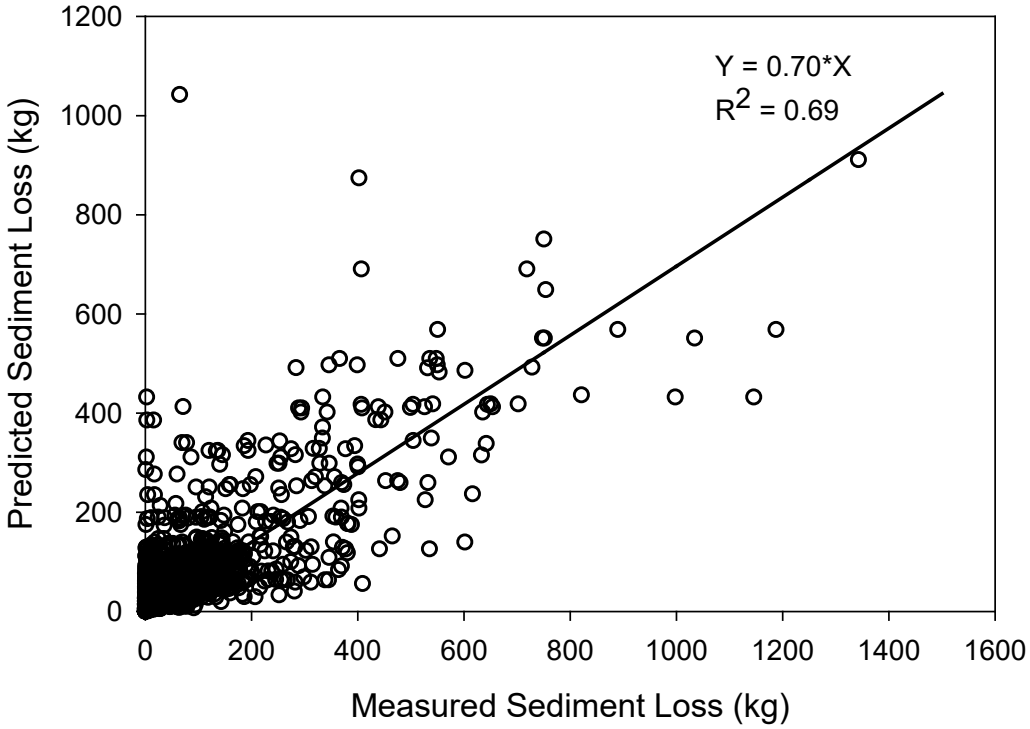

Figure 7. Transfer learning predicted furrow sediment loss versus measured sediment loss based on a 19 x 12 KSOM to cluster data set vectors without measured sediment loss.

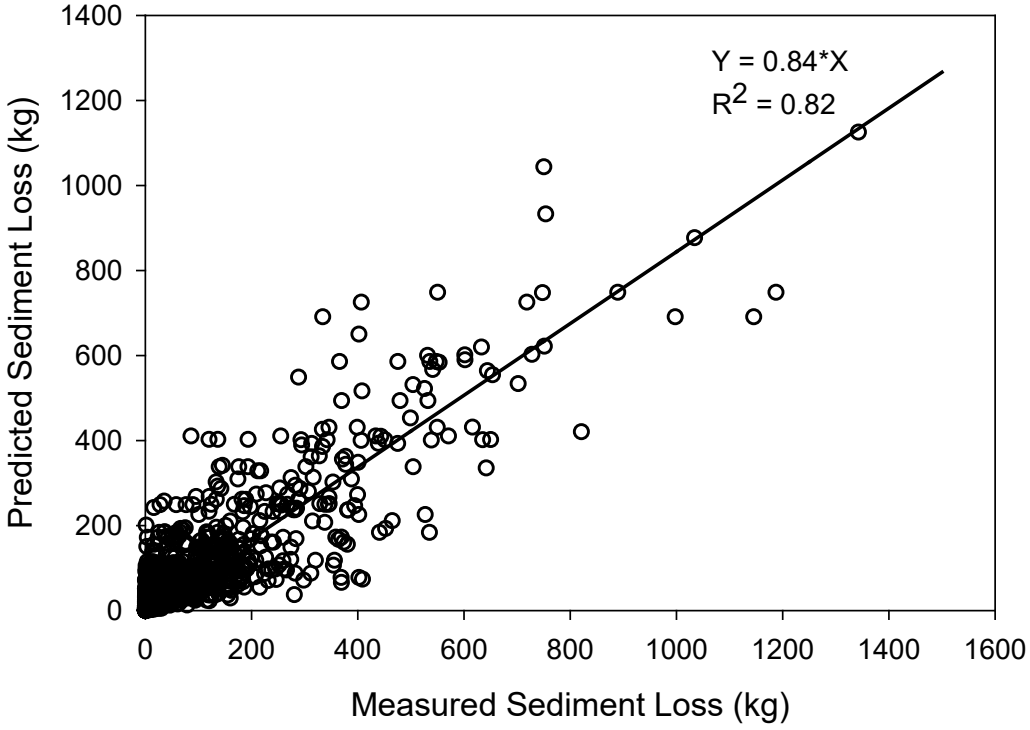

Figure 8. Transfer learning predicted furrow sediment loss versus measured sediment loss based on a 25 x 25 KSOM to cluster data set vectors without measured sediment loss.

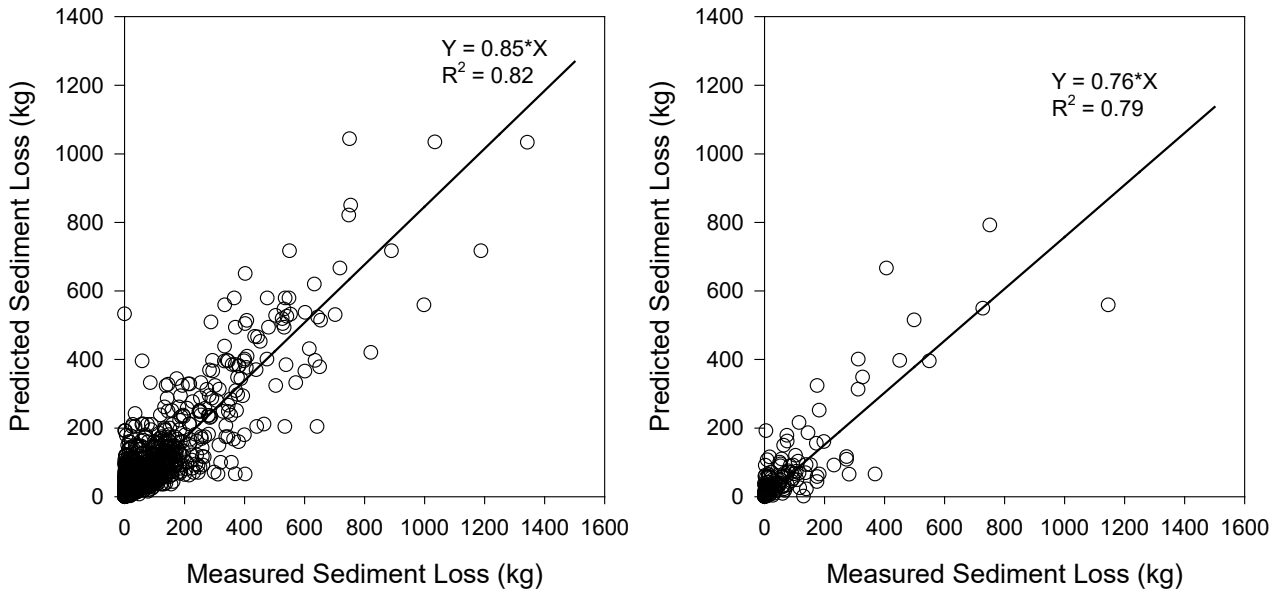

Figure 9. Predicted furrow sediment loss versus measured sediment loss for the model development data set (right) and validation data set (left) for a random split of the data set.

each data set input vector. The mean measured sediment loss and furrow flow rate of the input vectors in each BMU neuron was determined and assigned to each BMU. Mean BMU sediment loss was adjusted according to equation 6 (X = 1.2) to obtain a predicted sediment loss for each data set input vector. Predicted versus measured furrow sediment loss using a hexagonal 19 x 12 KSOM model is shown in figure 7. The transfer learning approach resulted in an increase in regression line slope and R^2 relative to a KSOM model alone (fig. 5). Based on the regression line slope predicted furrow sediment loss was 30% less than measured sediment loss. There are a few outliers where predicted sediment loss is highly overestimated. These could be true outliers in the data set or the result of the number of neurons used separate out the data set topology. The transfer learning approach was repeated using a hexagonal 25 x 25 KSOM model to provide finer resolution for clustering the data set. The greater number of neurons in the KSOM model resulted in a substantial increase in linear regression line slope and R^2 between predicted versus measured furrow sediment loss (fig. 8). The value of X in equation 6 that maximized slope and R^2 of the linear regression line was 1.9. Using a KSOM with a 25 x 25 hexagonal map resulted in a regression line slope of 0.84 indicating that on average predicted sediment loss was 16% less than measured sediment loss. This level of prediction performance makes the model useful for predicting furrow irrigation sediment loss given the natural variability in the data set but represents the upper limit of what can be achieved using the transfer learning approach since the complete data set was used for developing the KSOM model.

The data set was randomly divided into two separate data sets, one for KSOM model development and one for model validation to test the stability of the transfer learning approach. The data set division was 90% for model development and 10% for validation. The prediction results for two random data divisions are shown in figures 9 and 10. The results depend heavily upon how data vectors associated with furrow sediment losses exceeding 600 kg are split between the two data sets. For the two random divisions shown in figures 9 and 10, comparisons between measured and predicted furrow sediment loss was very similar with nearly equivalent linear regression slops and R^2 in both cases, indicating that the KSOM model provides consistent clustering of the factors affecting furrow sediment loss, regardless of small changes in the development data set. Comparisons between measured and predicted sediment loss for the validation data set varied between the two random data divisions but provided suitable to estimation of furrow sediment loss given the amount of inherent variability in the data set.

The hit histogram for a hexagonal 25 x25 KSOM for the complete data set without including measured furrow sediment loss in the input vectors. The unshaded hexagonal cells represent neurons that were not a BMU for any vector in the complete data set. Thus, with the transfer learning approach, there are no values for measured sediment loss associated with these neurons. If a new data vector input to the KSOM returns a BMU without an associate measured sediment loss value, then no predicted value of sediment loss is available. In this situation, the second best BMU, third best BMU, and so on can be determined based on distance from the first BMU using equation 1 until a neuron with associated measured sediment loss is obtained and used to estimate furrow sediment loss. Over 50 percent of the neurons (unshaded hexagonal cells) in figure 11 do not have an associated measured mean furrow sediment loss. However, these neurons are dispersed throughout the 25 x 25 hexagonal map such that the nearest neuron with measured sediment loss can be readily determined. Additional furrow

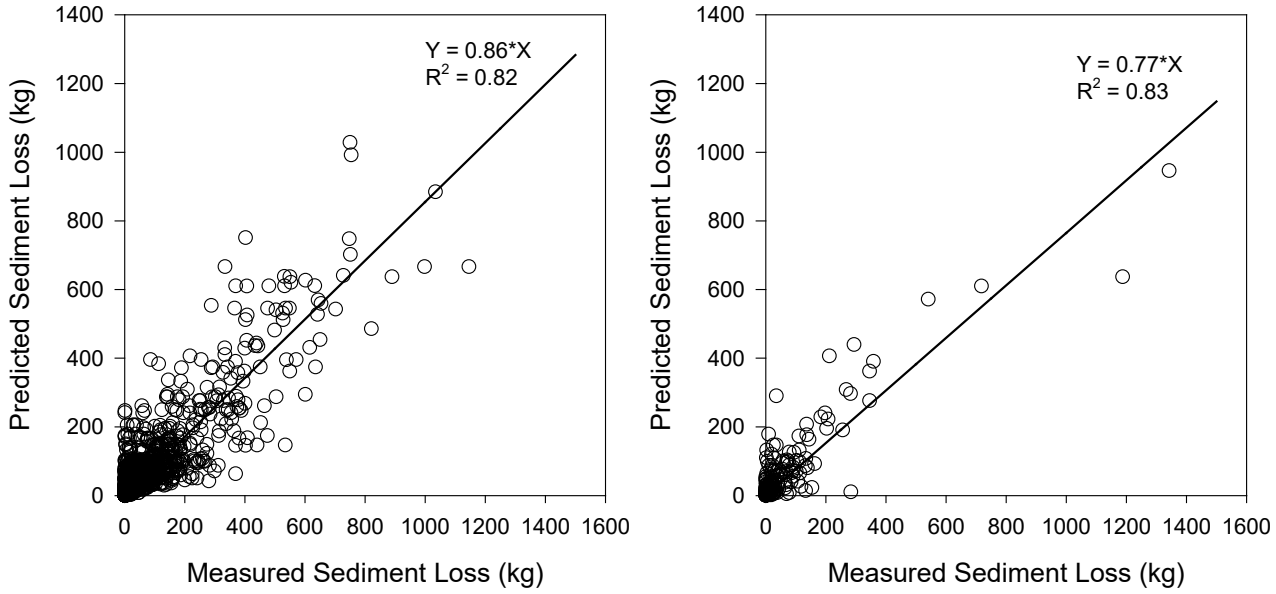

Figure 10. Predicted furrow sediment loss versus measured sediment loss for the model development data set (right) and validation data set (left) for a second random split of the data set.

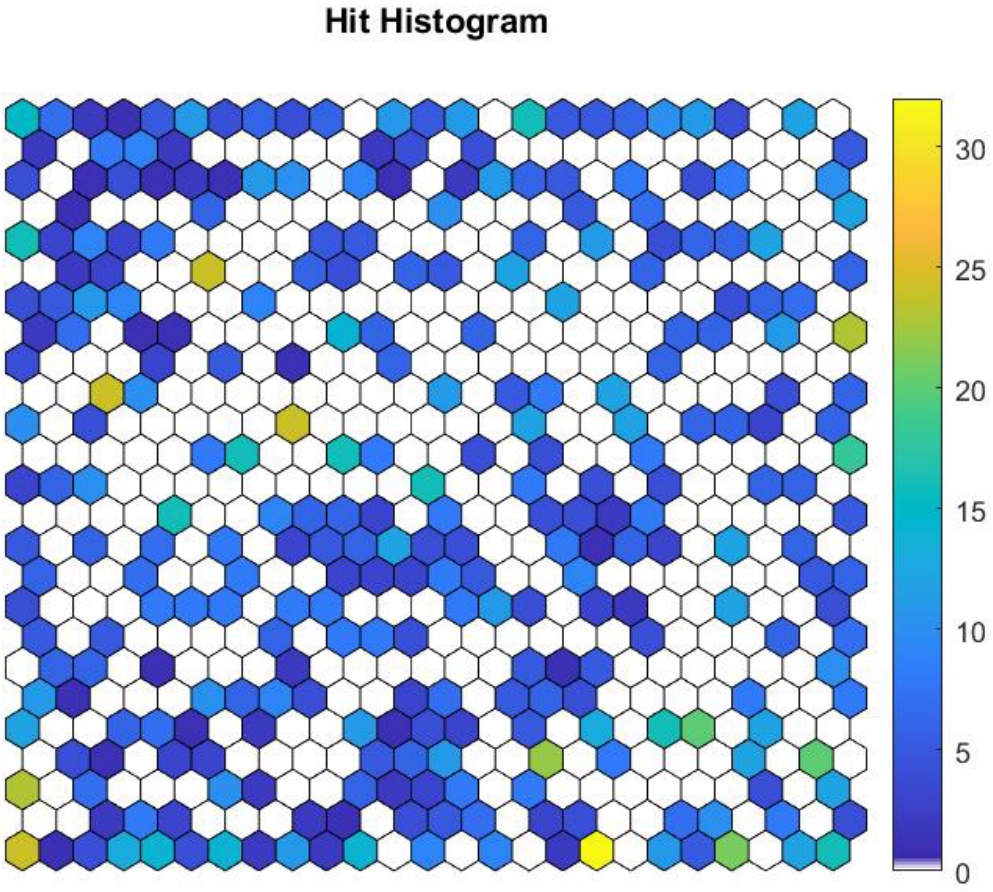

Figure 11. Hit histogram for a hexagonal 25 x25 KSOM map for the complete data without including measured sediment loss. Color of hexagonal node represents number of data set input vectors associated with the node.

erosion data sets need to be obtained to quantify the error associated finding the nearest neuron with measured mean sediment value. The KSOM model can be updated using new measured data sets when available

The transfer learning approach used in this study can be implemented in a spreadsheet. The KSOM codebook can be stored in the spread sheet and the distance between an input data vector and each neuron in the codebook can be determined using equation 1. The codebook neurons can be ranked based on distance with the nearest neuron corresponding to the 1st BMU, next nearest being the 2nd best BMU and so on. The nearest neuron with a measured mean sediment value can determined using a lookup table. The measured sediment value can be adjusted using equation 6 based on the mean flow rate of the BMU with the result being the predicted furrow sediment loss for the input data vector.

Summary

Historical published and unpublished data sets containing measurements of furrow irrigation sediment loss in the western U.S. were assembled into a furrow sediment loss data set comprising over 2000 measurements. The data set was used to evaluate the feasibility of using a KSOM to predict furrow irrigation sediment loss. Despite the immunity of KSOMs to measurement variability, the inherent variability in measured furrow sediment loss limited the ability of a KSOM model to reliability predict furrow sediment loss. When compared to measured furrow irrigation sediment loss, predicted furrow sediment loss was under predicted by 44% on average with a linear regression R^2 of 0.6. The KSOM model was placing little weight on measured sediment loss in the input data set, indicating that it was clustering the data based on input parameters defining the hydraulic and soil conditions. This outcome was used to develop a transfer learning approach to estimating furrow sediment loss. The transfer learning approach used a hexagonal 25 x 25 KSOM to cluster similar hydraulic and soil conditions in the data set. The BMU of each input vector in the data set was determined using the KSOM model and the measured furrow irrigation sediment loss associated with the input vector(s) of a BMU was averaged to obtain a mean sediment for each BMU. Similarly, the furrow flow rate associated with the input vector(s) of a BMU was averaged to obtain a mean furrow flow rate for each BMU. A predicted value of sediment loss was obtained by applying an input vector to the KSOM, which determines the BMU for the input vector, the mean sediment loss associated with the BMU is

adjusted for any difference in input vector flow rate and the mean flow rate of the BMU resulting in the predicted value. When compared to measured furrow irrigation sediment loss, predicted furrow sediment loss was under predicted by 16% on average with a linear regression R^2 of 0.82. When the data set was randomly split into model development (90%) and validation (10%) data sets the prediction results were similar. The transfer learning approach developed in this study can potentially be implemented in a spreadsheet.

References

Adeloye, A. J., Rustum, R., & Kariyama, I. D. (2011). Kohonen self-organizing map estimator for the reference crop evapotranspiration. *Water Resources Research*, *47*(8). https://doi.org/10.1029/2011WR010690

Adeloye, A. J., Rustum, R., & Kariyama, I. D. (2012). Neural computing modeling of the reference crop evapotranspiration. *Environmental Modelling & Software*, *29*(1), 61-73. https://doi.org/10.1016/j.envsoft.2011.10.012

Berg, R. D., & Carter, D. L. (1980). Furrow erosion and sediment losses on irrigated cropland. *J. Soil Water Conservation*, *35*(5), 267-270.

Bjorneberg, D. L., Kincaid, D. C., Lentz, R. D., Sojka, R. E., & Trout, T. J. (2000). Unique aspects of modeling irrigation-induced soil erosion. *Int. J. Sed. Res.*, *15*(2), 245-252.

Bjorneberg, D. L., Prestwich, C. J., & Evans, R. G. (2007). Evaluating the surface irrigation soil loss (SISL) model. *Applied Engineering in Agriculture*, *23*(4), 485-491. https://doi.org/10.13031/2013.23490

Bjorneberg, D. L., Strelkoff, T. S., Clemmens, A. J., & Lee, J. H. (2010). The current state of predicting furrow irrigation erosion. 5th National Decennial Irrigation Conference Proceedings, 5-8 December 2010, Phoenix Convention Center, Phoenix, Arizona USA, St. Joseph, MI.

Bjorneberg, D. L., Trout, T. J., Sojka, R. E., & Aase, J. K. (1999). Evaluating WEPP-predicted infiltration, runoff, and soil erosion for furrow irrigation. *Transactions of the ASAE*, *42*(6), 1733-1742. https://doi.org/10.13031/2013.13336

Carter, D. L. (1990). Soil erosion on irrigated lands. In B. A. Stewart, Nielson, D.R. (Ed.), *Irrigation of agricultural crops* (pp. 1143-1171). American Society of Agronomy.

Carter, D. L. (1993). Furrow Irrigation Erosion Lowers Soil Productivity. *Journal of Irrigation and Drainage Engineering*, *119*(6), 964-974. https://doi.org/10.1061/(ASCE)0733-9437(1993)119:6(964)

Carter, D. L., Berg, R. D., & Sanders, B. J. (1985). The Effect of Furrow Irrigation Erosion on Crop Productivity. *Soil Science Society of America Journal*, *49*(1), 207-211. https://doi.org/10.2136/sssaj1985.03615995004900010041x

Evans, R. G., Girgin, B. N., Chenoweth, J. F., & Kroeger, M. W. (1995). Surge irrigation with residues to reduce soil erosion. *Agricultural Water Management*, *27*(3), 283-297. https://doi.org/10.1016/0378-3774(95)01151-8

Fernández-Gómez, R., Mateos, L., & Giráldez, J. V. (2004). Furrow irrigation erosion and management. *Irrigation Science*, *23*(3), 123-131. https://doi.org/10.1007/s00271-004-0100-3

Fornstrom, K. J., & Borrelli, J. (1985). *Sediment losses from furrow irrigated croplands in Wyoming. Final project report*. U. o. W. Agricultural Engineering Depart.

García, H. L., & González, I. M. (2004). Self-organizing map and clustering for wastewater treatment monitoring. *Engineering Applications of Artificial Intelligence*, *17*(3), 215-225. https://doi.org/10.1016/j.engappai.2004.03.004

Israelsen, O. W. C., George D.; and Laurit, Cyril W. (1946). Bulletin No. 320 - Soil Erosion in Small Irrigation Furrows. In U. S. University (Ed.), *UAES Bulletins. Paper 281*. Logan, UT: Utah Agricultural Extension Service.

Kemper, D. W., Trout, T. J., Brown, M. J., & Rosenau, R. C. (1985). Furrow Erosion and Water and Soil Management. *Transactions of the ASAE*, *28*(5), 1564-1572. https://doi.org/10.13031/2013.32478

King, B. A., Bjorneberg, D. L., Trout, T. J., Mateos, L., Araujo, D. F., & Costa, R. N. (2016). Estimation of furrow irrigation sediment loss using an artificial neural network. *Journal of Irrigation and Drainage Engineering*, *142*(1), 04015031. https://doi.org/10.1061/(ASCE)IR.1943-4774.0000932

King, L. G., McNeal, B. L., Ziari, F. A., & Matulich, S. C. (1984). *On-farm improvements to reduce sediment and nutrients in irrigation return flow*. (Rep. No. EPA-600/2-84-044). Washington, D.C.: U.S. EPA

Koluvek, P. K., Tanji, K. K., & Trout, T. J. (1993). Overview of soil erosion from irrigation. *Journal of Irrigation and Drainage Engineering*, *119*(6), 929-946. https://doi.org/10.1061/(ASCE)0733-9437(1993)119:6(929)

Kumar, N., Adeloye, A. J., Shankar, V., & Rustum, R. (2020). Neural computing modelling of the crop water stress index. *Agricultural Water Management*, *239*, 106259. https://doi.org/10.1016/j.agwat.2020.106259

Mateos, L., & Giráldez, J. V. (2005). Suspended load and bed load in irrigation furrows. *CATENA*, *64*(2), 232-246. https://doi.org/10.1016/j.catena.2005.08.007

Nearing, M. A., Foster, G. R., Lane, L. J., & Finkner, S. C. (1989). A process-based soil erosion model for USDA-water erosion prediction project technology. *Transactions of the ASAE*, *32*(5), 1587-1593. https://doi.org/10.13031/2013.31195

O'Donkor, J. (1978). *Physical and management factors affecting sediment loss from surface irrigated fields* [Masters of Science, University of Idaho]. Moscow, ID.

Penn, B. S. (2005). Using self-organizing maps to visualize high-dimensional data. *Computers & Geosciences*, *31*(5), 531-544. https://doi.org/10.1016/j.cageo.2004.10.009

Rustum, R., Adeloye, A. J., & Scholz, M. (2008). Applying Kohonen self-organizing map as a software sensor to predict biochemical oxygen demand. *Water Environment Research*, *80*(1), 32-40. https://doi.org/10.2175/106143007X184500

Trout, T. J. (1996). Furrow Irrigation Erosion and Sedimentation: On-field Distribution. *Transactions of the ASAE*, *39*(5), 1717-1723. https://doi.org/10.13031/2013.27689

Trout, T. J., & Neibling, W. H. (1993). Erosion and Sedimentation Processes on Irrigated Fields. *Journal of Irrigation and Drainage Engineering*, *119*(6), 947-963. https://doi.org/10.1061/(ASCE)0733-9437(1993)119:6(947)

Tunio, A. F. (1994). *Evaluation of erosion and sediment loss in furrow irrigation with alternative practices* [Master of Science, Oregon State University]. Corvallis, OR.

USDA-NASS. (2018). *2018 irrigation and water managment survey*. Washington, DC Retrieved from https://www.nass.usda.gov/Publications/AgCensus/2017/Online_Resources/Farm_and_Ranch_Irrigation_Survey/index.php

Vatanen, T., Osmala, M., Raiko, T., Lagus, K., Sysi-Aho, M., Orešič, M., Honkela, T., & Lähdesmäki, H. (2015). Self-organization and missing values in SOM and GTM. *Neurocomputing*, *147*, 60-70. https://doi.org/10.1016/j.neucom.2014.02.061

Vesanto, J., Himberg, J., Alhoniemi, E., Parhankangas, J., Team, S., & Oy, L. (2000). *SOM toolbox for Matlab 5*. Espoo, Finland: Helsinki University of Technology