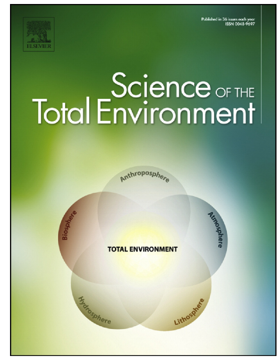
Accepted Manuscript



A web-based geovisual analytics platform for identifying potential contributors to culvert sedimentation

Haowen Xu, Ibrahim Demir, Caglar Koylu, Marian Muste

PII: S0048-9697(19)33270-X
DOI: https://doi.org/10.1016/j.scitotenv.2019.07.157
Reference: STOTEN 33351

To appear in: *Science of the Total Environment*

Received date: 22 February 2019
Revised date: 9 July 2019
Accepted date: 11 July 2019

Please cite this article as: H. Xu, I. Demir, C. Koylu, et al., A web-based geovisual analytics platform for identifying potential contributors to culvert sedimentation, Science of the Total Environment, https://doi.org/10.1016/j.scitotenv.2019.07.157

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

A Web-Based Geovisual Analytics Platform for Identifying Potential Contributors to Culvert Sedimentation

Haowen Xu [(1)], Ibrahim Demir [(2)], Caglar Koylu [(3)], Marian Muste [(2)]

[1]Computational Urban Sciences Group, Oak Ridge National Laboratory,

[2]IIHR—Hydroscience & Engineering, the University of Iowa,

[3] Geographical and Sustainability Sciences, the University of Iowa

Corresponding author: Haowen Xu (haowen-xu@uiowa.edu)

Present/permanent address: 302E C. Maxwell Stanley Hydraulics Laboratory, Iowa City, IA 52242-1585, U.S.A.

Abstract

Sediment accumulation at culverts involves large-scale and interlinked environmental processes that are difficult to address with experimental or physical modeling methods. This article presents an alternative data-driven investigation for shedding insights into these processes. Accordingly, a web-based geovisual analytics application, the IowaDOT platform, was developed, which allows users to explore the complex processes associated with the sediment deposition at culverts. The platform provides systematic procedures for (1) collecting and integrating analytical variables into a single dataset, (2) quantifying the degree of culvert sedimentation using time series of aerial images, (3) identifying drivers that contribute to culvert sedimentation processes from a variety of culvert structural and upstream landscape characteristics using a tree-based feature selection algorithm, and (4) facilitating the understanding of complex spatial and relational patterns of culvert sedimentation processes using multivariate geovisualizations supported by a self-organizing map (SOM). As the outcomes of this study, these patterns identify culvert sedimentation-prone regions in Iowa and quantify empirical relationships

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

between the drivers and culvert sedimentation degrees. A simple evaluation of the platform was performed to assess the usefulness and user satisfaction of the tool by professional users, and positive feedbacks are received.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

1. Introduction

Anthropogenic activities and climate change have significantly altered river systems around the world with immediate implications for stream channel stability and riverine infrastructure security (Larsen et al., 2007; Wohl et al., 2015). One of the direct consequences is sediment accumulation at culverts, which has been recognized as a widespread and costly infrastructure operating hazard in the Midwestern United States (Muste, 2013; Rowley, 2014). Culverts, as man-made structures, alter the cross-section of natural streams through human-induced flow expansions and contractions, providing a favorable environment for sediment deposition near culvert structures (Ho et al., 2013). Sediment blockage formed at the structural inlet impairs a culvert’s capacity to convey its design discharge, causing water crossing failures and roadway overtopping during flood events (Cafferata, 2004; Furniss et al. 1998). Since the existing culvert design specifications do not typically consider aspects of sediment transport and deposition within culvert structures, the commonly accepted strategy for mitigating culvert sedimentation is through costly and labor-intensive mechanically cleaning. This is unfortunate as sedimentation is prevalent not only in Iowa but also in many other states, and it requires considerable expense to mechanically clean culverts to avoid socio-economic losses (Cafferata, 2004; Rowley, 2014).

Culvert sedimentation is a typical coupled human-environment system problem, which involves fast-evolving anthropogenic activities interacting with natural processes occurring in culverts’ drainage areas. Although sedimentation occurs naturally in undisturbed landscapes, excessive sediment production that stems from human activities, such as agriculture, urbanization, and channelization, delivers more sediment supply to downstream culverts. Additionally, the local alteration of the stream geometry in the vicinity of a culvert is another contributing factor to the accumulation of sediment deposits near culverts (Ead et al., 2000). For better informing the culvert design and for providing a more system-based approach for planning the remedial actions, there is a need to understand the key drivers that contribute to the culvert sedimentation processes, as well as to identify the regions that are

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

prone to culvert sedimentation. Due to process-related complexity and the limited research in this area, there are many knowledge gaps in understanding these processes (Gregory 2006; Ho, 2010).

Conventionally, when solving sediment-related problems, erosion and sediment transport models would often be preferred over other approaches (Nearing et al., 2005; Merritt et al., 2013). Unlike other sediment-related problems, culvert sedimentation is more complex, as it is driven by the interaction among geomorphological (different types of erosions in watersheds), hydrological (instream sediment transport), hydrodynamics (sediment deposition near culverts), and ecological (growth of the riparian vegetation that stabilizes sediment deposition) processes (Cafferata, 2004; Haan et al, 1994; Ho, 2010; Merritt et al., 2003). Therefore, the problem cannot be easily solved using erosion and sediment transport models without considerable efforts for integrating multiple models and considering influences from human activities.

Alternatively, data-driven approaches can effectively help environmental scientists analyze and solve large-scale complex problems that involve pattern recognition, non-linear modeling, and classification, without detailed consideration of the internal structure of the physical process (Dibike and Solomatine, 2001). A comparative study conducted by Liang et al. (2019) showed that data-driven models can effectively inform and complement the simulations conducted with physically-based models. Currently, there are a relatively limited number of studies that utilize various computation-based data models (machine-learning and statistics), such as artificial neural networks (Tayfur, 2002), random forests (Francke et al. 2008; López-Tarazón et al. 2011), and M5 model trees (Onderka et al., 2012), to tackle sediment research (e.g., estimate sediment budget and transport). Despite their capability of handling large environmental datasets and making predictions, many computational methods are deemed “black-box”, as their results are difficult for humans to understand and interpret (Keim, 1997). Therefore, these methods are less intuitive for operations research (e.g., watershed and infrastructure management) that aims to support decision-making and improve understanding of the

ACCEPTED MANUSCRIPT

problem's nature, including data, dynamics, and the critical relationships involved (Andrienko et al., 2007). In contrast, geovisual analytics can foster better human interpretation of abstract spatial and relational patterns within a dataset and can effectively support decision-making through human-computer collaboration (Keim, 1997). However, design and implementation of geovisual analytical tools for environmental decision-making are challenging due to the complexity, size, and high-dimensionality of spatial environmental data (Andrienko et al., 2007; Demir et al., 2015). To address these limitations, many efforts were spent in Geographic Information Science (GIS) to integrate geovisual techniques (e.g., a series of linked maps and graphs) and computational methods (e.g., spatial statistics and spatial regression) for conducting exploratory spatial data analysis (Anselin, 2000; Bertolotto et al., 2007; Demir and Beck, 2009; Gahegan et al., 2002; Guo, 2005; Tukey 1977). Examples of visual analytical software produced from these efforts include GeoDa (Anselin, 2000), GeoVISTA (Gahegan et al., 2002, 2008), GeoXP (Laurent, et al., 2009), and Arc_Mat (LeSage and Liu, 2010). Nevertheless, web-based geovisual analytics application that improves the understanding of complex environmental processes are still rare in the water management sector, except for a few studies (e.g., Kollat et al., 2011; Leonard et al., 2016; Matrosov et al., 2015). Currently, there are no applications that address the culvert sedimentation problem using geovisual analytical techniques.

This article introduces a geovisual analytical platform that allows users to explore complex environmental processes of culvert sedimentation and provides data-driven insights for effectively mitigating sediment blockage at culverts in Iowa. The objectives of the platform design include (1) enabling quantification of the culvert sedimentation degrees using time series aerial images, (2) integrating multiple analytical variables into a single dataset from distributed sources; (3) identifying the key drivers that contribute to sediment deposition at culverts from a wider pool of variables that describe culverts' upstream landscape characteristics and structural attributes using a tree-based feature selection algorithm; and (4) facilitating the exploration of the complex spatial and relational

ACCEPTED MANUSCRIPT

patterns in culvert sedimentation processes using multivariate geovisualizations supported by a self-organizing map (SOM). The targeted patterns include the spatial variation of culvert sedimentation degrees in Iowa and the quantification of the empirical relationships between the culvert sedimentation degrees and key drivers. These specifications are then used to guide the development of a web-based application, the IowaDOT Culverts Platform, using Python and JavaScript. Developed using a data-driven architecture and open-source technologies, the platform is generalizable and can be extended to address similar environmental problems in other U.S. regions.

2. Methods

2.1 Conceptual Design

Streams, as final recipients of the water-borne material transported from local and upstream basins, are a good reflection of the hydrologic characteristics of upstream watershed areas (Allan, 2004; Auerbach et al., 2016, Hill et al., 2016). Thus, in-stream sediment conditions can be indirectly analyzed using landscape characteristics in culverts' drainage areas (watershed, catchment, riparian corridors), which describe changes in river form, substrate characteristics, and floodplain characteristics through time or with respect to reference channels and may reflect the main drivers of sedimentation processes (Olden et al., 2012; Rymszewicz et al., 2018; Wagener et al., 2007; Wohl et al., 2015). Through a data-driven investigation, the platform aims to establish the likelihood of the culvert sedimentation (dependent variable) as a function of these process drivers (independent variables) across Iowa. The analytical pipeline (Fig. 1) of the platform consists of three steps: 1) data collection & integration, 2) identification of sedimentation drivers, and 3) pattern exploration & inferences.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

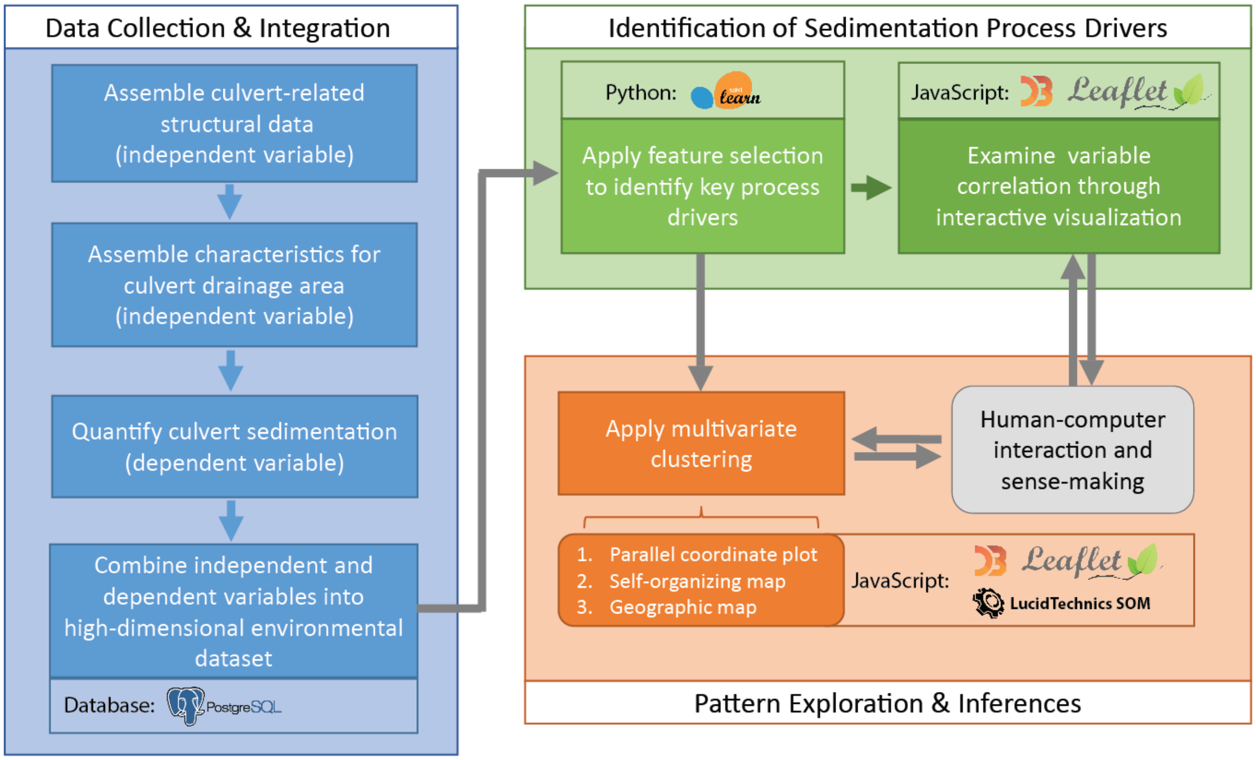


Fig. 1. The analytical pipeline of the platform.

2.2. Data Collection & Integration

This step aims to create a comprehensive culvert sedimentation dataset that contains both the dependent and independent variables. The independent variables consist of culverts' upstream landscape characteristics and structural attributes, and are collected from multiple third-party datasets. Culvert's upstream landscape characteristics describe the land cover, soils, lithology, runoff, and topography in culverts' drainage area, reflecting different environmental aspects of sediment production processes (e.g., hydrologic, geomorphological, and ecological). Culvert structural attributes describe factors that can influence the hydrodynamics processes for sediment deposition near culvert structures, including culverts' design discharge, geometry specifications, slope, and hydraulic control (Howley, 2004). Table 1 summarizes the data sources of these variables and provides a detailed data description. Afterwards, all independent variables are combined into one dataset as summary statistics (created using ArcGIS Zonal Statistics and Tabulate Area tools, ESRI, 2014) for four types of drainage

ACCEPTED MANUSCRIPT

extents that include watershed, local catchment, riparian corridor, and immediate corridor upstream of culverts (Fig. 2). These spatial extents are selected based on the principles of multi-dimensional stream connectivity (Demir and Szczepanek, 2017; Tockner and Stanford, 2002) representing the potential scales of the sediment production and transport processes. The delineation of these spatial extents is carried out using watershed delineation tools provided in the IowaDOT Culverts Platform.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 1. Summary of data sources.

Dataset	Provider	Variables & Associated Processes	Variable Type
Soil Survey Geographic Database (SSURGO)	Natural Resources Conservation Service (NRCS)	Culvert upstream basin characteristics (sediment production)	Independent variable
Revised Universal Soil Loss Equation (RUSLE)			
Stream-Catchment (StreamCat) dataset	United States Environmental Protection Agency (USEPA)		
High Resolution Land Cover (HRLC) dataset	Iowa Geological Survey (IGS)		
Iowa landforms map			
National Hydrography Dataset (NHD) plus version 2	United States Geological Survey (USGS)	Stream characteristics (in-stream sediment transport)	
Structure Inventory and Inspection Management System (SIIMS)	Iowa Department of Transportation (Iowa DOT)	Culvert structural attributes (sediment deposition at the culvert structure)	
Field inspections (at 255 culverts)	IIHR - Hydroscience & Engineering		
Culvert design discharge using Eash method (Eash, 2001)	USGS		
Quantification of culvert sedimentation degrees at 400 culverts via time-series aerial photos	IIHR - Hydroscience & Engineering	Severity of sediment accumulation at culverts	Dependent variable

ACCEPTED MANUSCRIPT

The dependent variable describes the severity of the sediment accumulation at culverts and is classified using time-series aerial images of the culvert vicinity (for the 2005-2017 time interval). The platform has a built-in map viewer that can systematically collect and store orthorectified aerial images downloaded from the EagleView Technologies, Inc., a technology provider of high-quality aerial imagery (with resolution up to 2-inch to 3-inch per pixel) (Fig. 3a). A mapping tool is then developed within the viewer to allow the delineation of sediment deposits near a culvert (shown in Fig. 3b). Based on the traced area of the sediment blockage at culverts, a classification of culvert sedimentation degrees is created consisting of 5 quantile classes, including “clean”, “low”, “med”, “high”, and “very high”.

ACCEPTED MANUSCRIPT

Finally, all independent variables are combined with the dependent variable through spatial joins to create the comprehensive culvert dataset that entails 300 independent variables. The dataset is physically stored in a PostgreSQL database deployed the server that hosts the platform.

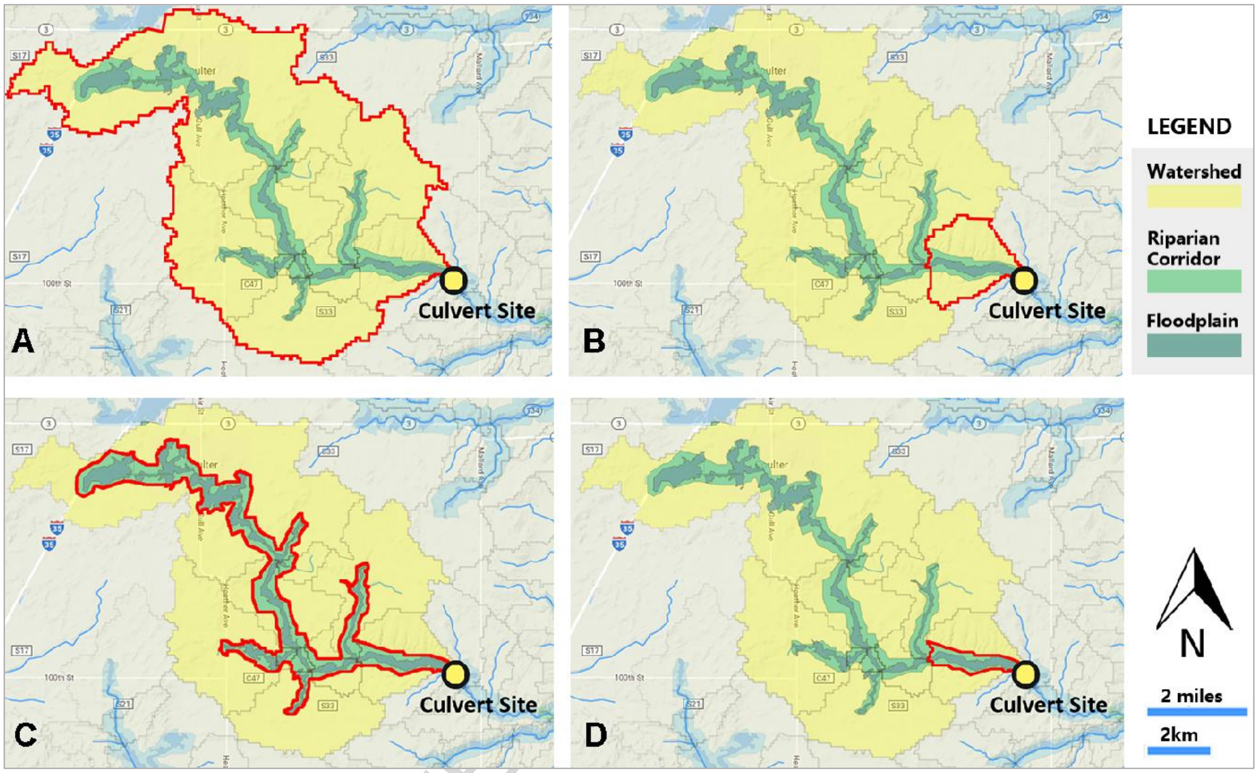


Fig. 2. Characterization of culvert upstream areas (highlight in red lines): (a) watershed, (b) local catchment, (c) riparian corridor (represented by the green buffer areas), and (d) immediate corridor upstream of culvert.

Fig. 3. The digital viewer for (a) displaying time-series aerial imagery in the culverts' vicinity and (b) delineating the area of sediment depositions at a culvert.

2.3 Identification of Sedimentation Process Drivers

Although all independent variables are pre-selected based on hydrologic and sediment transport domain knowledge, it is assumed that not all of them are equally contributing to the culvert sedimentation processes under investigation. The large numbers of the contributing variables involved in natural-scale environmental processes and their possible correlations pose a significant challenge for analysis by hiding or diluting the multifunctional relationships among the most relevant variables. To overcome this limitation, a feature selection strategy (Liu and Motoda, 1998) was adopted to identify the key drivers of culvert sedimentation processes. Specifically, a tree-based feature selection algorithm from the Python Scikit-learn library (Buitinck et al., 2013) was applied within the platform to rank the independent variables based on their relevance to the classified sedimentation severity. The algorithm utilizes a forest of decision trees to evaluate the importance of each variable based on the average

ACCEPTED MANUSCRIPT

impurity decrease (characterized by the information gain/entropy reduction). The justification for choosing this method is based on the fact that (1) the performance of decision trees is not affected by non-linear relationships, (2) the method requires less data transformation, and (3) the output of decision tree-based ranking is intuitive and easy-to-understand. The top-ranked variables are summarized in Table 2 as sedimentation process drivers.

Table 2. Summary of key process drivers

Variables	**Associated spatial extent**
Stream-to-culvert width ratio (SCW ratio)	Culvert structural attributes
Design discharge	
Landform regions	Watershed characteristics
Drainage density	
Ag land (%)	
Slope Length and Steepness Factor (LS factor)	Riparian corridor
Forest (%)	
Grass (%)	Immediate corridor upstream of culvert
Stream sinuosity	Stream
Stream width	

Following the automated feature selection step, a preliminary visual interface is developed, which allows users to visually examine the dependencies between the degrees of culvert sedimentation and the selected drivers in a Parallel Coordinates Plot (PCP) (Inselberg 1985). The PCP presents a multidimensional space through a two-dimensional display by using parallel axes to represent variables (Edsall, 2003). In the PCP, each edge links the sedimentation degree (the first axis from the left) with values of different key drivers (the rest of axes) at an individual culvert. The platform also synchronizes other visualizations, including a pie chart for displaying major sedimentation classes and a web map that displays culvert locations (Fig. 4a). The PCP also allow users to interact with the data through filtering and brushing, through which users can define criteria to filter data and test the dependencies between the culvert sedimentation degree and a specific driver. In Fig. 4b, an example is provided to demonstrate

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

the visual interface's capability of revealing the dependency between the sedimentation degree and the SCW.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

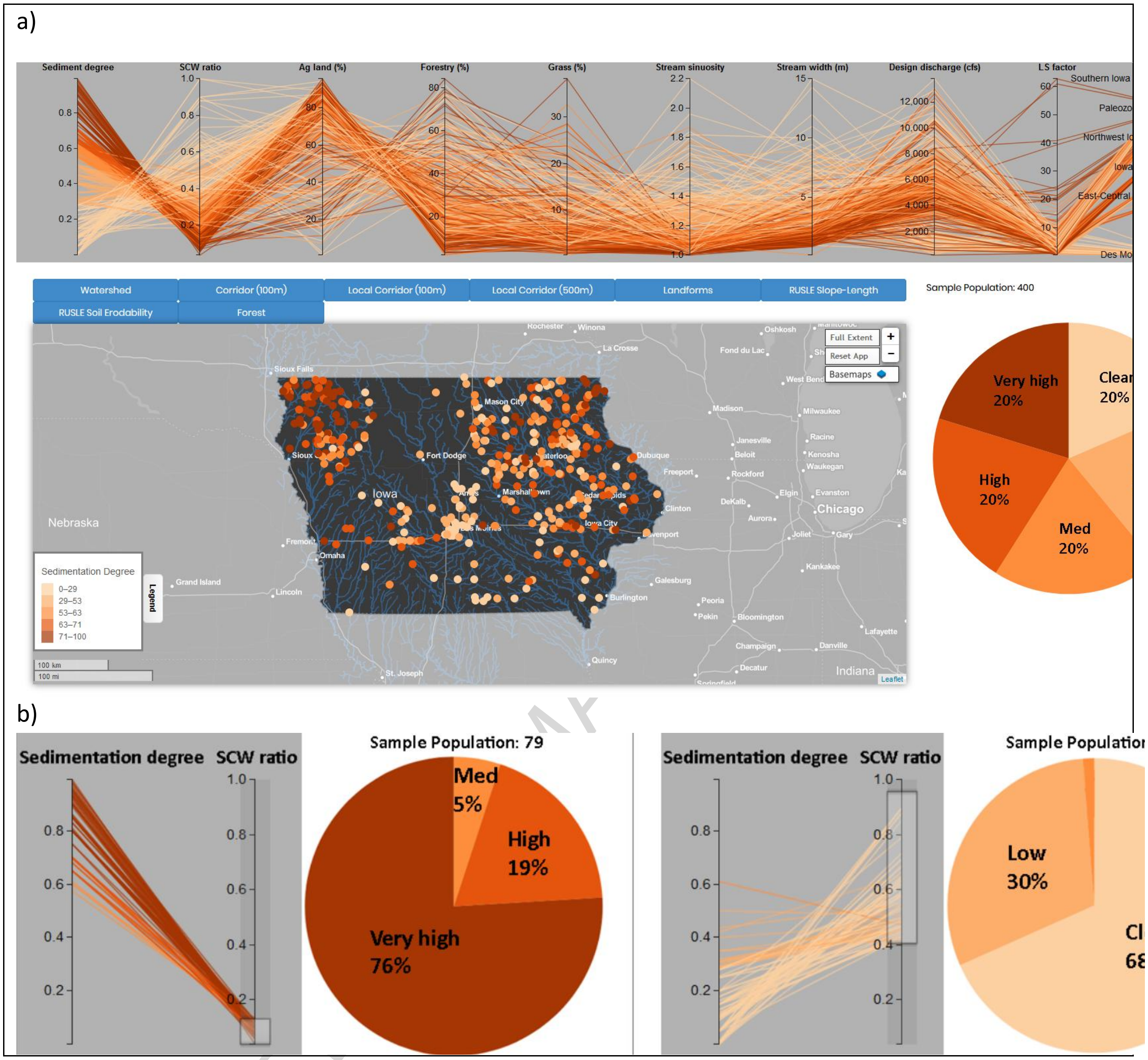


Fig. 4. Interfaces for (a) visualizing culvert sedimentation datasets in both the geographic space (web map) and the multi-dimensional attribute space (PCP and pie chart) and (b) examining the dependency between the sedimentation degree and SCW ratio.

2.4 Pattern Exploration and Inferences

The PCP alone is sensitive to visual clutter and large dataset sizes, and has limited ability to present the multivariate patterns resulting from combinations of multiple drivers (Johansson and Forsell 2015;

ACCEPTED MANUSCRIPT

Keim and Kreigel 1996) that influence culvert sedimentation degrees. Subsequently, further multivariate analysis is needed to group culverts with similar characteristics and summarize the associated multivariate patterns, which can provide additional insights for developing mitigation solutions that target specific drivers. For an enhanced *feature exploration stage*, the platform adopts a Self-Organizing Map (SOM), a multivariate clustering technique that utilizes an artificial neural network (ANN) (Kohonen 2001) to preserve and abstract patterns (Guo, 2015; Koua and Kraak 2004; Koylu 2014; Skupin and Fabrikant 2003) in the dataset. Before the analysis, the culvert data is normalized using the minimum and maximum values (a.k.a. feature scaling techniques) and a SOM network of 15x15 nodes (size) is constructed within the platform. Through a series of preliminary testing and experiments, the study discovered that the size of 15x15 allows the SOM to produce clusters with a minimum value range of the culvert sedimentation degree, indicating that each cluster is fully separated and has relatively low entropy.

In a more-advanced visual interface, SOM outputs are visualized through an interactive 2D map that adopts a hexagonal layout of SOM nodes (Fig. 5a). Each hexagon in the map represents a node, containing a cluster of culverts that are similar to each other and are different from those in other nodes. The orientation of the SOM map also indicates that neighboring nodes (hexagon) are similar to each other in multivariate space. A 2-D color scheme is them applied to the SOM map rendering each node with a discrete color (cluster-color). Then, the colored SOM map is connected to a modified PCP that bundles edges (Palmas, 2014) by the cluster-average value of each variable and renders them using their associated cluster-color (Fig. 5b). This configuration allows users not only to overview the pattern of different culvert clusters but also to explore the detailed patterns within a specific cluster (Shneiderman, 1996). To visualize the geographic distribution of different clusters, the platform synchronized the SOM outputs with a web map that can display culvert locations using the cluster-color and represent their sedimentation degrees as the radius of dot symbology. By clicking specific hexagons

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

of the 2D SOM map (also serves as a legend for the geographic map), users can zoom to specific clusters and further explore the associated relationship through the geographic map and the PCP as shown in Fig. 5b and Fig. 5c, which visualizes a 5x5 subset of the SOM outputs for demonstrating the interface. Additionally, the visual interface also allows users to explore patterns in the data by changing the SOM configurations (e.g., colors, node size, and learning rates) and provides evaluative feedback of the SOM performance (clusters size, average cluster range, and entropy).

The platform was developed using multiple open-source JavaScript libraries. The user interface is created using Bootstrap. The PCP and the SOM visualization are created using D3.js. Leaflet map engine is used to create the geographic map. The underlying SOM analysis is conducted using LucidTechnics SOM library.

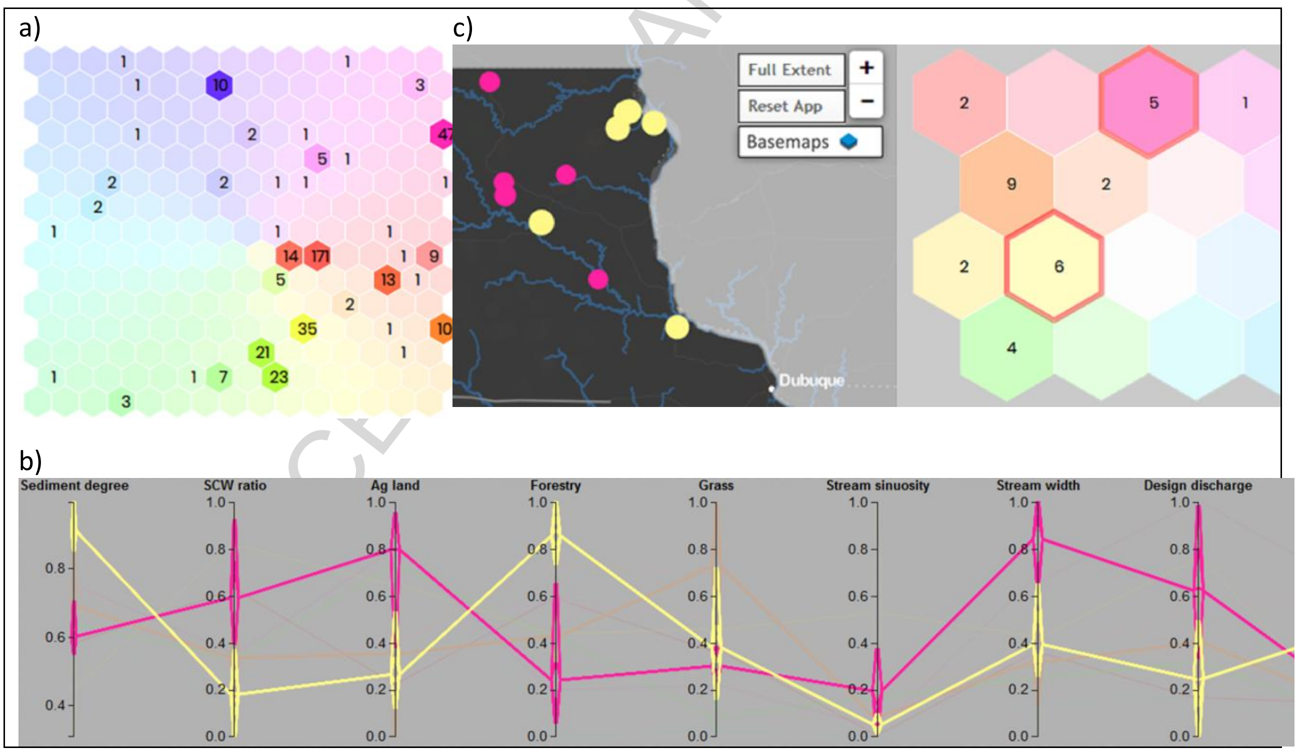


Fig. 5. The SOM visualizations: (a) hexagonal layout of SOM nodes, (b) edge-bundling PCP that displays the multivariate patterns and overall trends of each culvert clusters, and (c) spatial locations of culverts contained in individual cluster (node/hexagon).

ACCEPTED MANUSCRIPT

3. Results

3.1 Culvert Sedimentation-Prone Regions

At the overview level, this study discovers that the spatial patterns of SOM clusters are strongly correlated with the landform regions covering the state (Fig. 6). Past studies also indicate that the landform regions are good reflections of the landscapes' erosional history, as well as other synthetic factors characterizing topography, soil erodibility, and soil mobility (Scheidegger, 1973; Prior, 1991). A summary of the interactions with the applications is provided in Fig. 7 to reveal the spatial variation of the culvert sedimentation degrees across Iowa. More specifically, culverts with high sedimentation degrees are located in the Northwest Iowa Plains (red cluster) and Paleozoic Plateau (green cluster), while the degree of culvert sedimentation is considerably lower in the Des Moines Lobe (yellow clusters). In between these extremes, the culverts located in the Iowan Surface (pink cluster) and Southern Iowa Drift Plain (blue cluster) display a mixed sedimentation degrees. This discovered pattern can help transportation agencies make rapid assessments on sedimentation risk at a new culvert based its pertaining landform region, and prioritize sedimentation mitigation at existing culverts across Iowa.

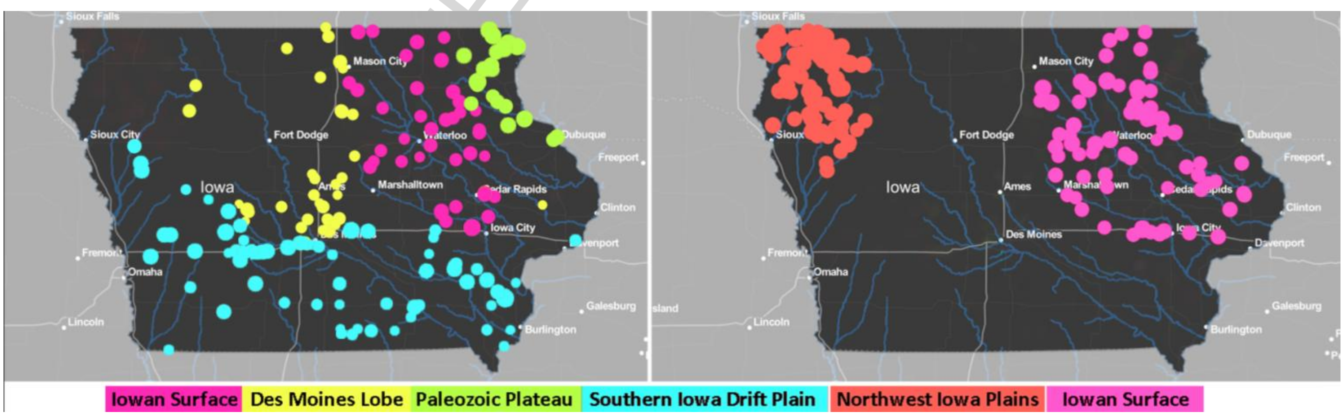


Fig. 6. SOM clusters that are spatially correlated to the landform regions.

ACCEPTED MANUSCRIPT

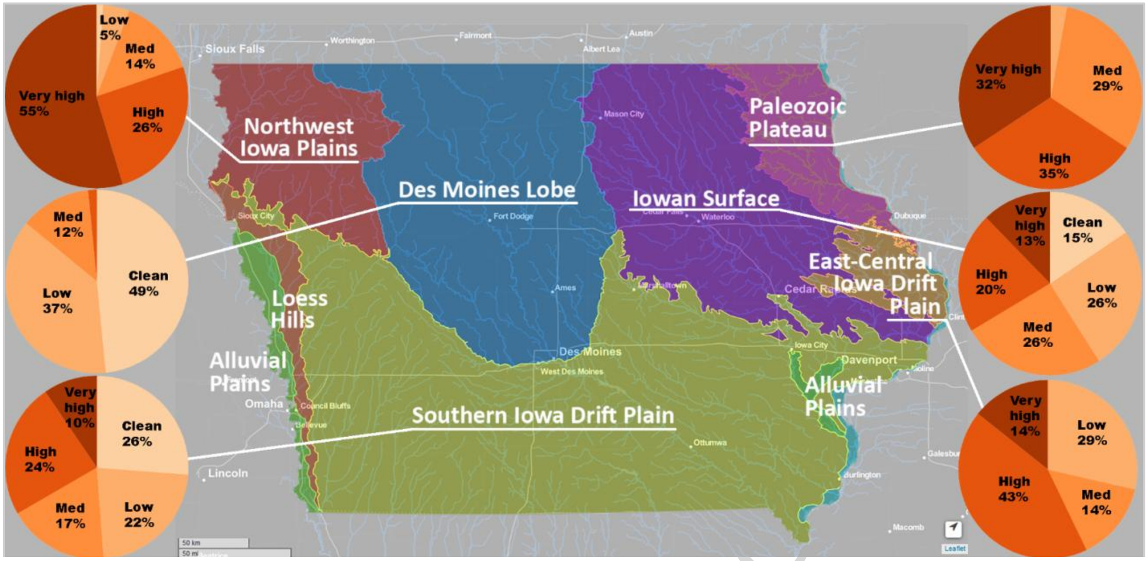


Fig. 7. Spatial variation of the culvert sedimentation degrees across different landform regions in Iowa.

3.2 Empirical Relationships between Drivers and the Culvert Sedimentation Degrees

Further exploration through the visual interface can be conducted to investigate what drivers, or combination of them, are responsible for the mix of culvert sedimentation within individual landform regions. This additional analysis is especially useful in areas such as the Iowan Surface region and Southern Iowa Drift Plain regions where culvert sedimentation potentials are mixed. Subsequently, the SOM-based multivariate clustering analysis is applied to culverts within each landform region using appropriate settings (e.g., node size and learning rate) that lead to the lowest entropy in the output clusters. Fig. 8 provides an example of results discovered through the visual interface, revealing the empirical relationships between drivers and the culvert sedimentation degrees in the Southern Iowa Drift Plain regions.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

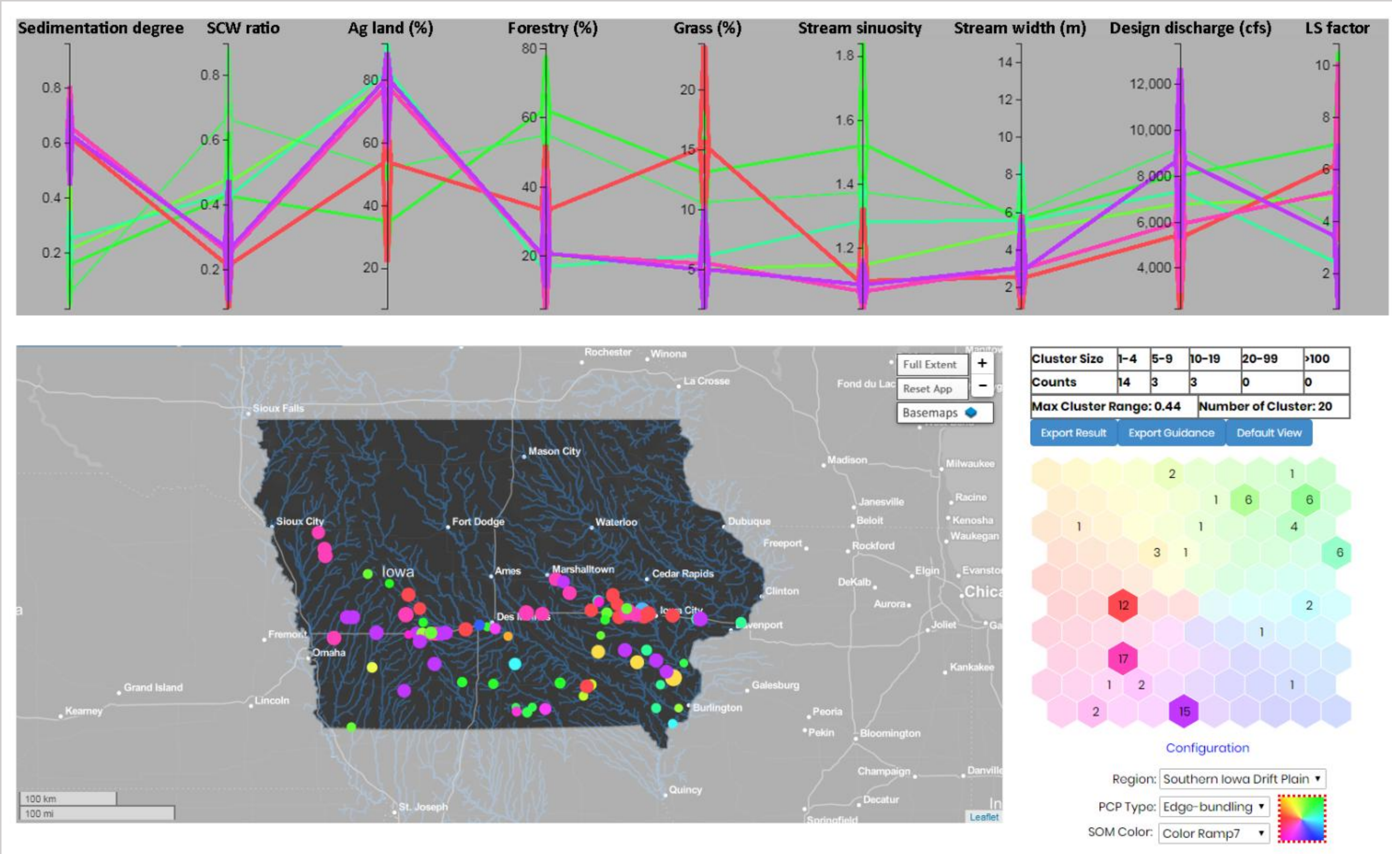


Fig. 8. Empirical relationships between drivers and the culvert sedimentation degrees in the Southern Iowa Drift Plain regions.

By clicking specific hexagons, users can zoom to specific clusters and further explore the associated relationship within the cluster through the geographic map and PCP. As an example, a comparison between two culvert clusters in the Southern Iowa Drift Plain regions was made in Fig.9 to provide an insight when other drivers are in a similar range, culverts with lower SCW ratio and smaller stream width (represented by the red cluster) are prone to high sediment deposition. This data-driven insight can help hydraulic engineers and culvert managers avoid culvert oversizing (so as to increase the SCW ratio) during culvert design and retrofitting. Such practices are also justified using the domain knowledge of culvert hydrodynamics (Ho, 2010). More insights and driver-specific solutions can be readily derived by users through interactions with the web interface. These visual-explicit presentations of complex patterns can be readily understood and interpreted by users without expertise in data science and GIS.

ACCEPTED MANUSCRIPT

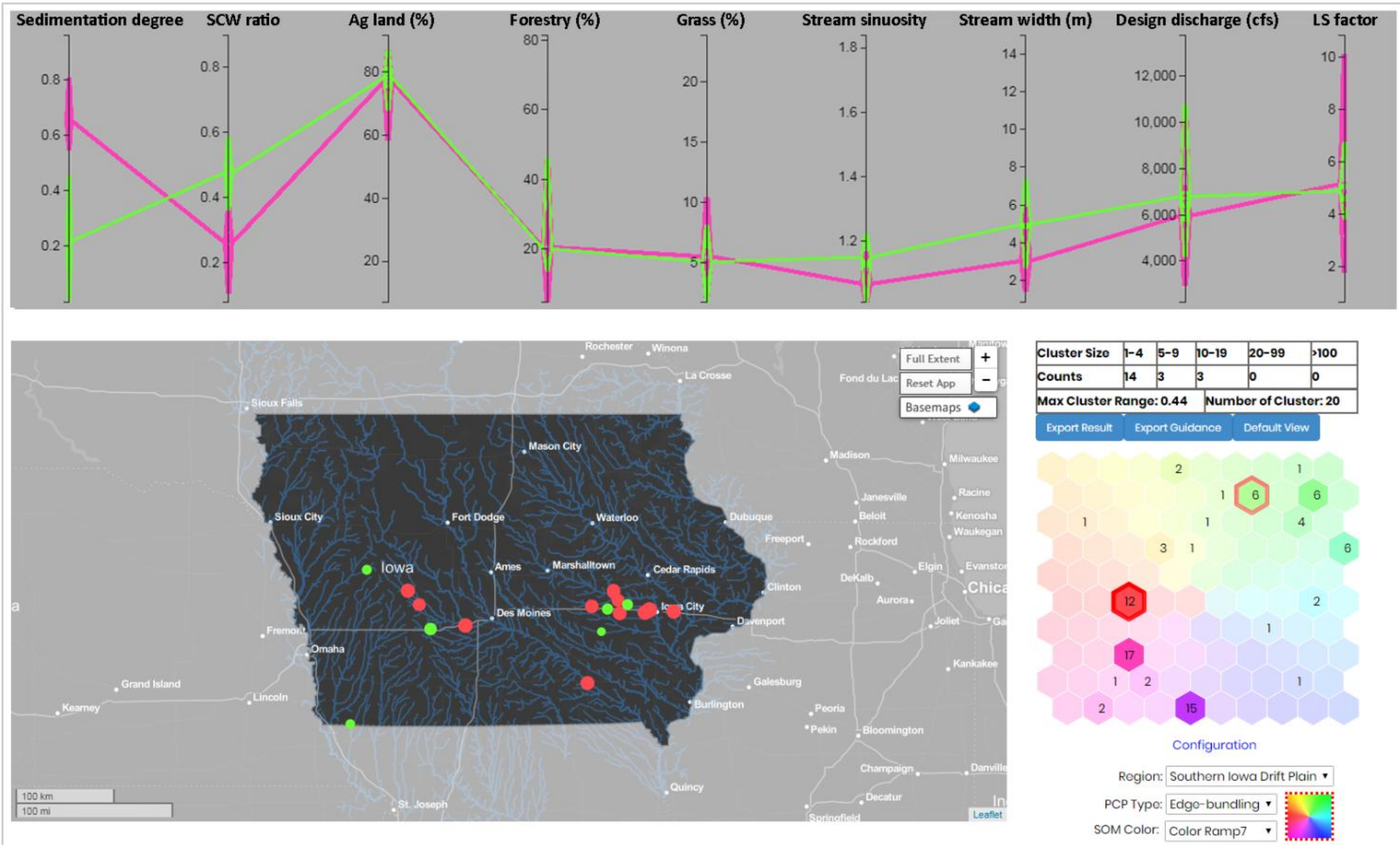


Fig. 9. Comparison of 2 culvert clusters (high sedimentation vs low sedimentation).

3.3 User evaluation

An evaluative survey of 30 hydraulic engineers and researchers was conducted to evaluate the platform. The survey consists of a series of training workshops organized by the Iowa Local Technical Assistance Program in 5 Iowa counties (Iowa LTAP, 2017). In this study, the evaluation was not designed to assess the usability of the platform using formal protocols. Instead, the intent was to receive feedback and critique from participants in terms of the results and user experiences of the geovisual analytics tool. The goal of the evaluation was to (1) gather qualitative feedback on the usefulness of the platform and user-friendliness of the visual interface design, and (2) learn whether additional design aspects should be added and improvements should be made to the platform. Specifically, the evaluation

ACCEPTED MANUSCRIPT

solicited comments for the platform functionality and for its "look and feel". After each workshop, the users were provided with an online link to the platform. Then, questionnaires were sent out to ask users to provide comments on the usefulness of the platform, user satisfaction, and visual interface design. The assembled input allowed us to quickly identify design issues and improve the functionality of the platform.

Most users commented positively on the usefulness of the platform and provided an average rating of 3.75/5 (29 users voted). Examples of some comments include: "this is a great tool that I will keep for the next time I have to work on a large box culvert", "a very handy tool for managing culverts", and "the web-based tool is very useful and can be accessed through mobile devices". Six users provided explicit comments on the visual interface design: "the visualization is easy-to-understand", "the visualization responds instantaneously to users' interaction", "the relationships in the visualization are very clear and straightforward", "the practical knowledge presented can be used in the field", "the aesthetic designs of the interface can be improved", and "a precise method for defining the brushing interval should be added". In summary, most users provided positive feedback during the evaluation, and corrective measures were applied to the platform as suggested by these users. Currently, the platform is transferred to the Iowa DOT to support the design and maintenance of the culverts on a continuous basis. After the software is widely used over a substantial time period, we plan to conduct a formal usability test, which aims to systematically measure user satisfaction, effectiveness, efficiency, and other usability metrics.

4. Discussion and Conclusion

This article introduces a visual analytics platform for investigating complex and interlinked environmental processes with a focus on sediment transport and accumulation at culverts. The platform combines data-driven analytics with human reasoning to create a friendly visual analytics application

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

that helps users explore the relational and spatial patterns of culvert sedimentation in Iowa. By taking advantage of a plethora of available data in third-party data sources, the data-driven approach is able to solve a practical problem that is difficult to tackle with conventional investigation means, i.e., experimental of numerical modeling approaches. The variables associated with the description of the sedimentation process from its original location to final deposition place (input data) and the degree of culvert sedimentation (output data) are all conveniently and efficiently linked using the same web-based platform. Special tools and cyber structure were developed to collect and integrate the data, identify the sedimentation process drivers, and to explore the spatial patterns of the assembled data. The insights provided from the visual interfaces can be applied to support decisions for culvert management and sedimentation mitigation, as well as to guide parameter selections for culvert design. Since the targeted audiences for the platform usage are culvert designers and hydraulic engineers without a computer science background, the platforms' interface design must account for the diversity of the users and the need for presenting the information in an easy-to-understand form.

Developed using a data-driven architecture and open-source technologies, the platform is generalizable and extendable to other locations with a similar degree of data availability. The platform components are modular and can be applied to investigate complex patterns of other water-driven scientific or management issues, such as water pollution or flood hazard occurrence.

Acknowledgments

Funding for the "Iowa Culverts" platform development was provided by the Iowa Highway Research Board, and Iowa Department of Transportation, Grant TR-655. Special thanks to our agency colleagues David Claman and Vanessa Goetz from the Iowa Department of Transportation for guidance and support.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Reference

Allan, J. D., 2004. Landscapes and riverscapes: the influence of land use on stream ecosystems. Annual Review of Ecology, Evolution, and Systematics. 35 (1), 257-284.

Andrienko, G., Andrienko, N., Jankowski, P., Keim, D., Kraak, M-J., MacEachren, A., Wrobel, S., 2007. Geovisual analytics for spatial decision support: Setting the research agenda, International Journal of Geographical Information Science. 21 (8), 839-857.

Anselin, L., 2000. Computing environments for spatial data analysis. Journal of Geographical Systems. 2 (3), 201–220.

Auerbach, D. A., Buchanan, B. P., Alexiades, A. V., Anderson, E. P., Encalada, A. C., Larson, E. I., Flecker, A. S., 2016. Towards catchment classification in data-scarce regions. Ecohydrology. 9 (7), 1235-1247.

Bertolotto , M., Martino S. Di., Ferrucci, F., Kechadi, T., 2007. Towards a framework for mining and analysing spatio-temporal datasets, International Journal of Geographical Information Science. 21 (8), 895-906.

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., et al., 2013. API design for machine learning software: experiences from the scikit-learn project. arXiv Preprint arXiv:1309.0238.

Cafferata, P. H., 2004. Designing watercourse crossings for passage of 100-year flood flows, wood, and sediment. California Dept. of Forestry and Fire Protection, Sacramento, CA (Updated 2017).

Demir, I., Beck, M. B., 2009. GWIS: a prototype information system for Georgia Watersheds, Paper 6.6.4. In: Proceedings Georgia Water Resources Conference: Regional Water Management Opportunities, April 27-29, 2009. UGA, Athens, GA, USA.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Demir, I., Conover, H., Krajewski, W. F., Seo, B. C., Goska, R., He, Y., McEniry, M. F., Graves, S. J. and Petersen, W., 2015. Data-enabled field experiment planning, management, and research using cyberinfrastructure. Journal of Hydrometeorology, 16 (3), 1155-1170.

Demir, I., Szczepanek, R., 2017. Optimization of river network representation data models for web-based systems. Earth and Space Science. 4 (6), 336–347.

Dibike, Y. B., Solomatine, D. P., 2001. River flow forecasting using artificial neural networks, Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere. 26 (1), 1-7.

Ead, S. A., Rajaratnam, N., Katopodis, C., Ade, F., 2000. Turbulent open-channel flow in circular corrugated culverts. Journal of Hydraulic Engineering. 126 (10), 750–757.

Edsall, R. M., 2003. An enhanced geographic information system for exploration of multivariate health statistics. The Professional Geographer 55 (2), 146-60.

ESRI, 2014. ArcGIS Desktop: Release 10.2.2. Environmental Systems Research Institute, Redlands, California.

Francke, T., López-Tarazón, J. A., Schröder, B., 2008. Estimation of suspended sediment concentration and yield using linear models, random forests and quantile regression forests. Hydrol Process. 22 (25), 4892–4904.

Furniss, M. J., Ledwith, T. S., Love, M. A., McFadin, B. C., Flanagan, S. A., 1998. Response of road-stream crossings to large flood events in Washington, Oregon, and Northern California, USDA Forest Service Technology & Development Program, 14.

Gahegan, M., Takatsuka, M., Wheeler, M., Hardisty,F., 2002, Introducing GeoVISTA Studio: an integrated suite of visualization and computational methods for exploration and knowledge construction in geography, Computers, Environment and Urban Systems, 26 (4), 267-292.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Gahegan, M., F. Hardisty, U. Demsar, and M. Takatsuka. 2008. GeoVISTA Studio: Reusability by design. In Open Source Approaches to Spatial Data Handling, edited by B. Hall and M. Leahy, 201–20. Berlin, Germany: Springer-Verlag.

Gregory, K. J., 2006. The human role in changing river channels. Geomorphology 79 (3–4), 172-191.

Guo, D., Gahega, M., MacEachren A., Zhou, B., 2005. Multivariate analysis and geovisualization with an integrated geographic knowledge discovery approach. Cartography and Geographic Information Science. 32 (2), 113–132.

Haan, C. T., Barfield, B. B., Hayes, J. C., 1994. Design hydrology and sedimentology for small catchments, San Diego Academic Press, San Diego, CA, USA.

Hill, R. A., Weber, M. H., Leibowitz, S. G., Olsen, A. R., and Thornbrugh, D. J., 2016. The Stream Catchment (StreamCat) Dataset: A Database of Watershed Metrics for the Conterminous United States. Journal of the American Water Resources Association (JAWRA) 52 (1): 120-128.

Ho, H.C., 2010. Investigation of unsteady and non-uniform flow and sediment transport characteristics at culvert sites. Ph.D Thesis, University of Iowa, Iowa City, IA, USA.

Ho, H.C., Muste, M., Ettema, R., 2013. Sediment self-cleaning multi-box culverts. Journal of Hydraulic Research. 51 (1), 92-101.

Howley, C. S., 2004. The relationships among culvert characteristics and culvert sedimentation. Master Thesis, University of Tennessee, Knoxville, TN, USA.

Inselberg, A., 1985. The plane with parallel coordinates. The Visual Computer. 1 (2), 69-97.

Iowa LTAP. 2017. The Self-Cleaning Multi-Barrel Culvert workshop. "iowaltap.iastate.edu/self-cleaning-multi-barrel-culverts". Iowa Local Technical Assistance Program.Retrieved on 11 May 2019.

Johansson, J., Forsell, C., 2015. Evaluation of parallel coordinates: overview, categorization and guidelines for future research. IEEE Transactions on Visualization and Computer Graphics. 22 (1), 579-88.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Keim D. A., 1997. Visual techniques for exploring data-bases, Invited tutorial, Int. Conference on Knowledge Discovery in Databases (KDD'97), Newport Beach, CA, USA.

Keim, D. A., Kriegel, H. P., 1996. Visualization techniques for mining large databases: a comparison, IEEE Transaction on Knowledge and Data Engineering. 8 (6), 923 – 938.

Kisi O., Karahan E. M., Sen Z., 2006. River suspended sediment modelling using a fuzzy logic approach. Hydrological Processes. 20, 4351–4362.

Kohonen, T. 2001. Self-organizing maps. Berlin, Germany, New York: Springer, 501.

Kollat, J. B., Reed, P. M., Maxwell, R. M., 2011. Many-objective groundwater monitoring network design using bias-aware ensemble Kalman filtering, evolutionary optimization, and visual analytics, Water Resources Research. 47 (2).

Koua, E. L., Kraak, M. J., 2004. Geovisualization to support the exploration of large health and demographic survey data. International Journal of Health Geographics. 3, 13.

Koylu, C., 2014. Understanding geo-social network patterns: computation, visualization, and usability. Ph.D thesis, University of South Carolina, Columbia, SC, USA.

Larsen, E. W., Girvetz, E. H., Fremier, A. K., 2007, Landscape level planning in alluvial riparian floodplain ecosystems: Using geomorphic modeling to avoid conflicts between human infrastructure and habitat conservation, Landscape and Urban Planning. 79 (4), 338-346.

Laurent, T., A. Ruiz-Gazen, and C. Thomas-Agnan. 2009. ''GeoXP: an R package for exploratory spatial data analysis.'' TSE Working Paper Series 99-099, Toulouse School of Economics. Toulouse, France.

Leonard, L., MacEachren, A. M., and Madduri, K., 2016. Graph-based visual analysis for large-scale hydrological modeling, Information Visualization. 16 (3), 205 – 21.

Liang, J., Li, W., Bradford, S.A. and Simunek, J. (2019). Physics-informed data-driven models to predict water quantity and quality in agricultural fields, Water 11, doi:10.3390/w11020200.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Liu, H., Motoda, H., 1998. Feature transformation and subset selection. IEEE Expert. 13(2), 26-28.

Liu, X., LeSage, J., 2010. Arc_Mat: a matlab-based spatial data analysis toolbox. Journal of Geographical Systems 12:69–87.

López-Tarazón J. A., Batalla R. J., Vericat D., Francke T., 2011. The sediment budget of a highly dynamic mesoscale catchment: the River Isábena. Geomorphology. 138 (1), 15–28.

Matrosov, E. S., Huskova, I., Kasprzyk, J. R., Harou, J. J., Lambert, C., Reed, P. M., 2015. Many-objective optimization and visual analytics reveal key trade-offs for London's water supply, Journal of Hydrology, 531 (3), 1040-1053.

Merritt, W. S., Letcher, R. A. and Jakeman, A. J., 2003. A review of erosion and sediment transport models. Environmental Modelling & Software, 18 (8), 761-799.

Muste, M., Constantinescu, G., 2013. Mitigation of sedimentation at multi-box culverts, Proposal for Iowa Highway Research Board (IHRB), Iowa Department of Transportation, Ames, IA, USA.

Muste, M., Xu, H., 2017. Mitigation of sedimentation at multi-box culverts, IIHR Report No. TR-655, Submitted to the Iowa Highway Research Board, Ames, IA, USA.

Nearing, M. A., Jetten, V., Baffaut, C., Cerdan, O., Couturier, A., Hernandez, M., Le Bissonnais Y., Nichols, M. H., Nunes J. P., Renschler C. S., SouchÃ¨re, V., Van Oost K., 2005. Modeling response of soil erosion and runoff to changes in precipitation and cover, CATENA, 61 (3), 131-154.

Olden, J. D., Kennard, M. J., Pusey, B. J., 2012. A framework for hydrologic classification with a review of methodologies and applications in ecohydrology. Ecohydrology 5 (4), 503-518.

Onderka, M., Krein, A., Wrede, S., Martinez-Carreras, N., Hoffmann, L., 2012. Dynamics of storm-driven suspended sediments in a headwater catchment described by multivariable modeling, Journal of Soils and Sediment. 12 (4), 620-635.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Palmas, G., Bachynskyi, M., Oulasvirta, A., Seidel, H. P., Weinkauf, T., 2014. An edge-bundling layout for interactive parallel coordinates. In IEEE Pacific Visualization Symposium (PacificVis), Yokohama, JP, 2014. 57-64.

Prior, J., 1993. Landforms of Iowa: University of Iowa Press, Iowa City, IA, USA.

Rowley, K. J., 2014. Sediment transport conditions near culverts, Master Thesis, Brigham Young University, Provo, UT, USA.

Rymszewicz, A., Bruen, M., O'Sullivan, J. J., Turner, N. J., Lawler, D. M., Harrington, J. R., Conroy, E., Kelly-Quinn M., 2018. Modelling spatial and temporal variations of annual suspended sediment yields from small agricultural catchments, Science of The Total Environment, 619-620, 672-684.

Shneiderman, B., 1996. The eyes have it: a task by data type taxonomy for information visualizations, Proceedings of IEEE Symposium on Visual Languages, Boulder, CO, USA, 1996. 336-343.

Skupin, A., Fabrikant, S., 2003. Spatialization methods: A cartographic research agenda for non-geographic information visualization. Cartography and Geographic Information Science. 30 (2), 99-119.

Solomantine, D.P. (2002). Applications of data-driven modelling and machine learning in control of water resources, Computation intellegnece in Control, M.Mohammadian, R.A. Sarker and X. Yao (eds), Idea Group Publishing, pp. 197-217.

Tayfur, G., 2002. Artificial neural networks for sheet sediment transport. Hydrological Sciences Journal. 47 (6), 879–892.

Tockner, K., Stanford, J. A., 2002. Riverine flood plains: Present state and future trends. Environmental Conservation. 29 (3), 308-330.

Tukey, J. 1977. Exploratory Data Analysis. Reading, MA: Addison Wesley.

Wagener, T., Sivapalan, M., Troch, P. A., Ross, W. A., 2007. Catchment classification and hydrologic similarity. Geography Compass. 1 (4), 901–931.

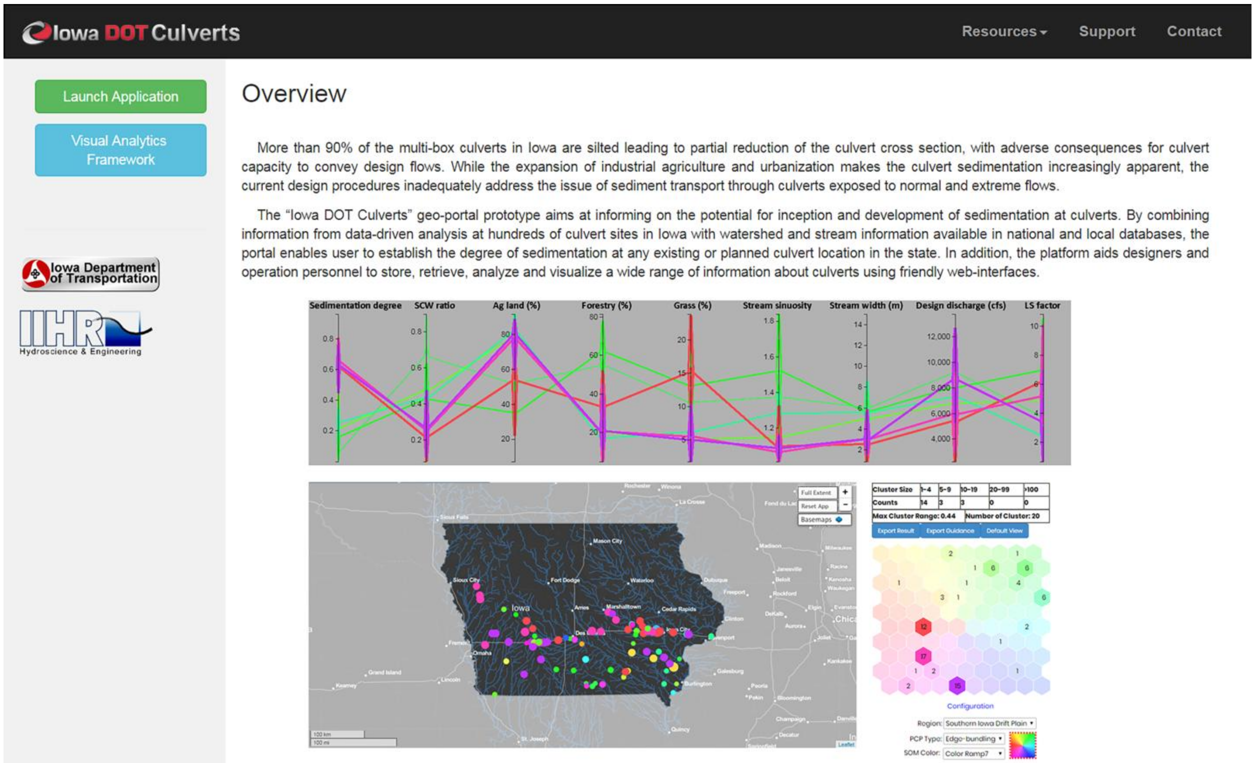
ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Wohl, E., Bledsoe, B. P., Jacobson R. B., Poff N. L, Rathburn, S. L., Walters, D. M., Wilcox, A. C., 2015. The natural sediment regime in rivers: broadening the foundation for ecosystem management. BioScience. 65 (4), 358–371.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT



Graphical abstract

Highlights

- A geovisual analytics platform is designed for exploring culvert sedimentation contributors.
- Several computational and visual analytics techniques are integrated into the web platform.
- The platform helps users identify culvert sedimentation-prone regions in Iowa.
- The platform improves the understanding of complex patterns in culvert sedimentation processes.
- The platform design is generalizable and adaptable to other environmental studies.

ACCEPTED MANUSCRIPT