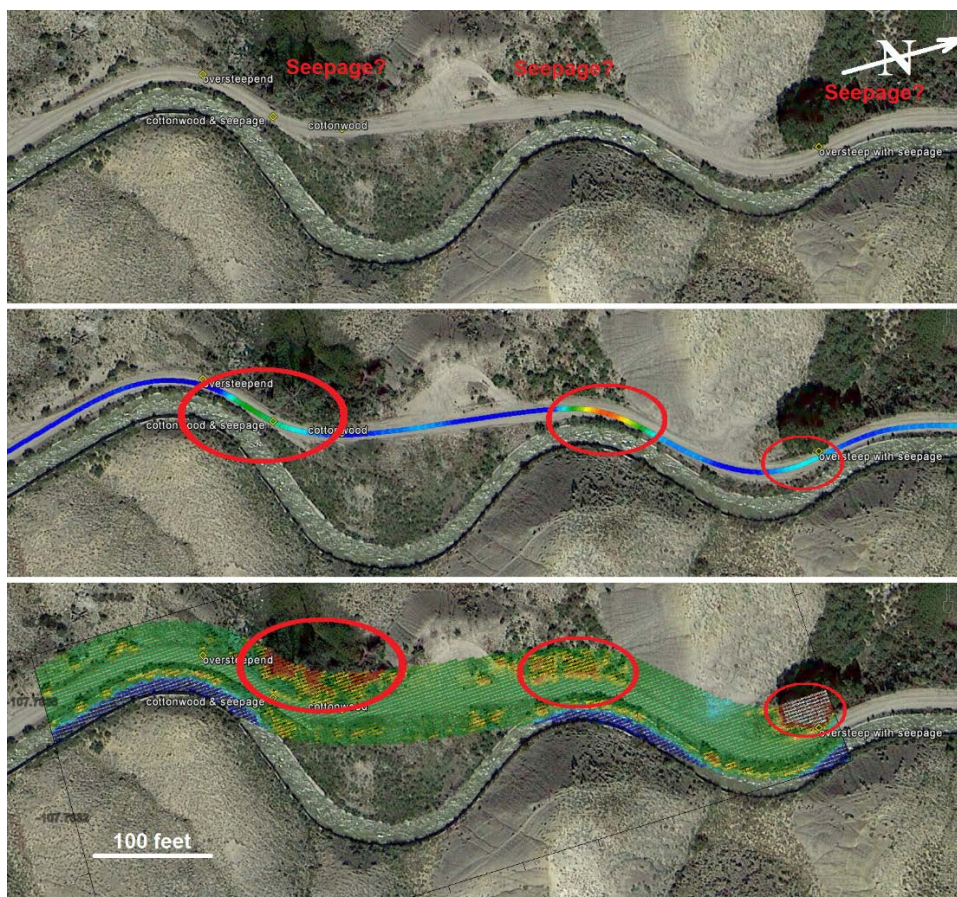


RECLAMATION

Managing Water in the West

Rapid Canal Embankment and Levee Health Assessment and Seepage Detection

Research and Development Office
Science and Technology Program
Final Report ST-2018-9918-01



U.S. Department of the Interior
Bureau of Reclamation
Research and Development Office

September, 2018

Mission Statements

Protecting America's Great Outdoors and Powering Our Future

The Department of the Interior protects and manages the Nation's natural resources and cultural heritage; provides scientific and other information about those resources; and honors its trust responsibilities or special commitments to American Indians, Alaska Natives, and affiliated island communities.

Disclaimer:

This document has been reviewed under the Research and Development Office Discretionary peer review process https://www.usbr.gov/research/peer_review.pdf consistent with Reclamation's Peer Review Policy CMP P14. It does not represent and should not be construed to represent Reclamation's determination, concurrence, or policy.

REPORT DOCUMENTATION PAGE				<i>Form Approved</i> <i>OMB No. 0704-0188</i>	
T1. REPORT DATE: SEPTEMBER 2018		T2. REPORT TYPE: RESEARCH		T3. DATES COVERED 2016 – 2018	
T4. TITLE AND SUBTITLE Evaluation and Development of Rapid Geophysical Profiling for Canal Embankment Seepage Detection				5a. CONTRACT NUMBER 16XR0680A1-RY1541IS201629918	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER 1541 (S&T)	
6. AUTHOR(S) Justin B. Rittgers				5d. PROJECT NUMBER ST-2018-9918-01	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER 86-68330	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Justin B. Rittgers, Principal Investigator, Bureau of Reclamation, Technical Services Center, 303-445-3010				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Research and Development Office U.S. Department of the Interior, Bureau of Reclamation, PO Box 25007, Denver CO 80225-0007				10. SPONSOR/MONITOR'S ACRONYM(S) R&D: Research and Development Office BOR/USBR: Bureau of Reclamation DOL: Department of the Interior	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) ST-2018-9918-01	
12. DISTRIBUTION / AVAILABILITY STATEMENT Final report can be downloaded from Reclamation's website: https://www.usbr.gov/research/					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Canal embankment health and seepage detection and monitoring is becoming increasingly important due to the Nation's vast water conveyance canals systems that are aging beyond their original design life. This research is aimed at developing new approaches for rapid data acquisition and detection and characterization of seepage with the use of multiple geophysical data types merged with satellite remote sensing data. The results of this research include developing semi-automated and robust workflows for combining and interpreting these various data types, including data fusion, attributes analysis, and machine learning, to better identify and differentiate between anomalies related to seepage, unauthorized penetrations, or other unhealthy conditions.					
15. SUBJECT TERMS geophysics, remote sensing, seepage detection, canals, earthen embankments					
16. SECURITY CLASSIFICATION OF: U			17. LIMITATION OF ABSTRACT U	18. NUMBER OF PAGES 87	19a. JUSTIN B. RITTGERS
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			19b. 303-445-3010

BUREAU OF RECLAMATION

**Research and Development Office
Science and Technology Program**

**Seismology, Geomorphology, and Geophysics Group,
Technical Services Center, 86-68330**

Final Report ST-2018-9918-01

**Evaluation and Development of Rapid Geophysical
Profiling for Canal Embankment Seepage Detection**

 9/26/18
Prepared by: Justin B. Rittgers

Geophysicist, Seismology, Geomorphology, and Geophysics Group, Technical Services Center,
86-68330

 9/26/18
Technical Approval: Richard D. Markiewicz

Geophysicist, Seismology, Geomorphology, and Geophysics Group, Technical Services Center,
86-68330

 9/26/18
Peer Review: Daniel Liechty

Geophysicist, Seismology, Geomorphology, and Geophysics Group, Technical Services Center,
86-68330

For Reclamation disseminated reports, a disclaimer is required for final reports and other research products, this language can be found in the peer review policy:

This document has been reviewed under the Research and Development Office Discretionary peer review process https://www.usbr.gov/research/peer_review.pdf consistent with Reclamation's Peer Review Policy CMP P14. It does not represent and should not be construed to represent Reclamation's determination, concurrence, or policy.

PEER REVIEW DOCUMENTATION

Project and Document Information

Project Name: Evaluation and Development of Rapid Geophysical Profiling for Canal
Embankment Seepage Detection _____ WOID: FA304

Document: Final Report

Document Author(s) Justin B. Rittgers _____ Document date: September 2018

Peer Reviewer: Daniel Liechty _____

For Reclamation disseminated reports, a disclaimer is required for final reports and other research products, this language can be found in the peer review policy:

"This information is distributed solely for the purpose of pre-dissemination peer review under applicable information quality guidelines. It has not been formally disseminated by the Bureau of Reclamation. It does not represent and should not be construed to represent Reclamation's determination or policy."

Review Certification

Peer Reviewer: I have reviewed the assigned items/sections(s) noted for the above document and believe them to be in accordance with the project requirements, standards of the profession, and Reclamation policy.

Reviewer: _____

(Signature)

Date reviewed: _____

9/26/18

Acknowledgements

The author would like to acknowledge and thank the various partnering Reclamation Area Offices and personnel that helped enable and support data collection along various canals, including Mathew Altman of the Western Colorado and Dale Hamilton of the Provo Utah Area Offices. We would also like to acknowledge and thank Gail Heath of the University of Arizona for their help and support in the design and construction of the non-magnetic cart used for data collection during this project. Finally, the author would like to thank Reclamation's project team members, notably, Audrey Rager, for her ongoing support in providing normalized difference vegetative index data throughout the development of this project.

Acronyms and Abbreviations

1D	One-dimensional
2D	Two-dimensional
DFC	Denver Federal Center
DGPS	Differential global positioning system
ERT	Electrical resistivity tomography
FDEM	Frequency-domain electromagnetics
LIDAR	Light detection and ranging
mS/m	millisiemens per meter
NAIP	National Agriculture Imagery Program
NDVI	Normalized difference vegetative index
NIR	Near Infrared
O&M	Operations and maintenance
Reclamation	Bureau of Reclamation
Red	Visible red light
SP	Self-potential
SVM	Support vector machines
USGS	United States Geological Survey

Executive Summary

Problem

Here in the United States alone, there are currently over 100,000 miles of canal and levee embankments and approximately 79,000 dams on the national inventory list. Many of these structures are reaching or have surpassed their initial design life, where most earthen embankment structures are over 50 years in age, and many are greater than 80 years in age. Development and urban encroachment along this aging infrastructure increases risk associated with seepage and internal erosion-caused failures. Additionally, water conveyance systems are known to lose significant quantities of water due to distributed seepage along canal embankments. As a result, there is a need for proactive yet efficient characterization of these vast infrastructural systems by means of rapid reconnaissance surveying technologies. Implementation of rapid reconnaissance tools of this sort could provide a list of “hot spots” that could be prioritized in terms of more focused mitigation or investigation efforts along urbanized segments of canals (e.g., using higher resolution geophysical methods or geotechnical investigation techniques at locations of elevated concern).

Research Approach

In order to help address these issues, this research has focused on developing a new approach for rapid assessment of earthen canal embankments for the detection and characterization of seepage. The investigated and developed approach involves the use of Normalized Difference Vegetative Index (NDVI) remote sensing data images, combined with continuous frequency-domain electromagnetic (FDEM) and magnetic gradiometry profiling surveys conducted along canal embankments.

Summary of Results

The main results of this research include:

- 1) Development of capabilities for rapid geophysical data collection along canal embankments with the design, construction and use of a vehicle-towed non-magnetic/non-metallic cart platform, enabling the collection of tens of miles of detailed geophysical profiling data per day.
- 2) Development of a semi-automated and robust data processing and analysis workflow for combining and interpreting these various geospatial data-types, enabling experimentation with data fusion, attributes analysis, and machine learning algorithms. This new workflow allows for better identification and differentiation between anomalies related to seepage versus the presence of metal infrastructure (e.g., known or unknown abandoned penetrations). The primary output of this workflow is a “Seepage Index,” or a numerical estimate of relative likelihood for seepage along the length of canal surveyed.

- 3) Surveying and seepage detection analysis results for the various surveys conducted along significant portions of South Canal, Government Highline Canal, and Strawberry Highline Canal.

Recommended Next Steps

This approach to identifying abnormal seepage or saturation levels within canal levees (based on electrical conductivity signatures/anomalies/trends) could help save a considerable amount of resources being spent on embankment assessment. Furthermore, adoption of this approach could help to reduce risk associated with embankment failures, and would help to characterize, prioritize, and address unwanted water conveyance seepage losses. The workflow is ready to be implemented on Reclamation's canal embankments, as needed or otherwise deemed helpful for ongoing Operations and Maintenance (O&M) activities.

Next steps could include further adaptation and testing of the automated interpretation workflows to incorporate additional data types (e.g., additional geophysical profiling and remote sensing data products beyond NDVI and FDEM, digitized United States Geological Survey (USGS) soil map data for surficial geologic and mineralogical information, cross-sectional embankment geometry information, topographic roughness, etc.). Additionally, the automated data analysis and machine learning approaches developed by this research effort for seepage detection should be further evaluated and developed for providing a robust and reliable means to predict seepage and health issues along vast lengths of canal embankments.

Contents

Executive Summary	vi
Problem.....	vi
Research Approach	vi
Recommended Next Steps	vii
Background	12
Problem	12
Conceptual Model of Canal Seepage.....	19
Frequency-Domain Electromagnetics.....	20
Enabling Rapid Data Collection: Non-magnetic Cart Design	25
Normalized Difference Vegetative Index	28
Developing a New Data Attribute: Normalized Seepage Index	29
Preliminary Approach	30
Addition of Magnetic Gradiometry Data	30
Addition of NDVI Statistics and Spatial Attributes	32
Addition of Time-lapse FDEM for Seasonal Change Detection	36
Interpreting the Seepage Index: Automatic Seepage Detection	39
Detrending and Thresholding	39
Machine learning: Clustering Analysis and Binary Classification	42
Fed Center Test Site Activities for System Evaluations	45
South Canal	49
Government Highline Canal	60
Strawberry Highline Canal	69
References	79

Figures

Figure 1. Map showing the approximate 4-acre area where preliminary non-magnetic cart configurations and geophysical system testing were performed (red polygon) located south of Building 56 and near the intersection of Main Ave. and 7 th St. on the Denver Federal Center, Colorado.....	15
Figure 2. Map showing the study area and data coverage along an approximate 9 mile segment of South Canal (red line) located southeast of Montrose, Colorado (managed by the Western Colorado Area Office).	16
Figure 3. Map showing the study area and data coverage along an approximate 9 mile segment of Government Highline Canal (red line) located north of Grand Junction, Colorado (managed by the Western Colorado Area Office).	17
Figure 4. Map showing the study area and data coverage along an approximate 13 mile segment of Strawberry Highline Canal (red line) extending between the cities of Spanish Fork and Santaquin, Utah (managed by the Provo Area Office).	18
Figure 5. A conceptual model of canal seepage and related changes in soil water saturation, electrical resistivity, and plant growth. The top and bottom images depict under-seepage and through-seepage, respectively.	19
Figure 6. Conceptual diagram showing inductive coupling of a FDEM system.	21
Figure 7. Diagram of an FDEM system's sensitivity as a function of depth (z) below ground surface for both a vertical dipole orientation ($\emptyset V(z)$ curve), and for a horizontal dipole orientation ($\emptyset H(z)$ curve). Image modified from (Geonics 2018).	22
Figure 8. Five FDEM systems tested for this research project: Geonics ltd's EM31-mk2 (A), Geonics ltd's EM34-mk2 (B), DualEM inc's DualEM-4 (C), Geonics ltd's EM38-mk2 (D), and GSSI inc's EMP-400 (E).	23
Figure 9. Photo of a Geometrics Inc. G858 magnetometer system configured in a vertical gradiometer fashion, with two cesium-vapor magnetometer sensors vertically aligned one half meter apart, and with DGPS positioning.	25
Figure 10. Two photos showing the first iteration of the non-magnetic cart design, being towed behind a truck at the DFC test-site while carrying an EM34 system with 10m coil separation and horizontal dipole orientation (top) and the EM31 FDEM system (bottom).	26
Figure 11. Photos showing the current version of the non-magnetic cart configuration, now utilizing a tethered and gimbaled min-cart for sake of carrying the quite long (~30ft) EM34 system (top two photos). The original main cart is still sufficient for carrying smaller instruments, including the EM38, EM31, and the magnetic gradiometer (bottom photo).	27
Figure 12. Example of how NDVI works to identify lush green vegetation based on different spectral reflectances of NIR and visible-red light bands. Image taken from (NASA, 2000).	29
Figure 13. Example showing a segment of extracted NDVI values along the master GPS track, a calculated vector of motion of travel along the canal embankment, and the corresponding cross-line used to calculate the local and long-range gradients of NDVI values in the direction from canal to downstream toe (e.g., in the direction of expected seepage). In this figure, the dark blue strip along the lower half of the NDVI data swath is the canal/water, and the yellow and red NDVI values that speckle the upper-half of the swath is vegetation along the canal slope and downstream toe.	33

Figure 14. Example plot showing a subset of raw NDVI data (colored scatterplot in upper-left plot) that has been extracted along a 20m wide swath that follows and is centered on a section of the master GPS track (black line in upper-left plot), and the various NDVI data metrics extracted from this subset of NDVI values along the GPS track.	34
Figure 15. Example plot showing a subset of NDVI data (upper-left plot) extracted along a 20m wide swath that follows the master GPS track, and the various NDVI data metrics extracted from this subset of NDVI values along the GPS track.	35
Figure 16. Example line graphs showing a comparison of time-lapse EM34 terrain conductivity data (mS/m) collected along Strawberry Highline Canal. Here, black circles identify segments where noticeable increases in conductivity were recorded during the second survey performed during watered-up conditions (red line).	37
Figure 17. Example X-Y scatterplot showing a comparison of time-lapse EM34 terrain conductivity data (mS/m) collected along a portion of Strawberry Highline Canal. Here, black circles identify segments where noticeable increases in conductivity were recorded during the second survey performed during watered-up conditions (data shifted to the south by 100m).	38
Figure 18. Example of the various input geophysical data-types and NDVI data metrics (top cluster of line-plots) from South Canal used as input for calculation of the resulting Seepage Index (bottom green line-plot).	40
Figure 19. Example from South Canal Depicting of the process of detrending clipping and renormalization of the Seepage Index for sake of applying thresholding for detection of most likely seepage areas.	41
Figure 20. An example scatterplot showing the results of clustering applied to define three categories of data (depicted in this example with different colors). Image modified from (Wikipedia contributors, 2018a).	42
Figure 21. Graphic showing how a support vector machine would choose a separating hyperplane for two classes of points in 2D. H_1 does not separate the classes. H_2 does, but only with a small margin. H_3 separates them with the maximum margin. Image taken from (Wikipedia contributors, 2018b).	44
Figure 22. Three ERT models for different data collection approaches (different electrode dipole configurations), and the resulting averaged “final” electrical resistivity model used for comparison to FDEM mapping results.	46
Figure 23. A map of EM31 terrain conductivity data superimposed on aerial imagery (left image), and a comparison between EM31 conductivity values and ERT modeled resistivity values collected along a “test-line” established within the DFC test-site (right image). The test-line was utilized to inspect system performance and sensitivity to subsurface electrical conductivity structure imaged by the ERT surveys.	47
Figure 24. Plots of three different frequencies of FDEM data collected across the DFC test site with an EMP-400 system.	48
Figure 25. Example from South Canal showing some of the data-types and resulting calculated “raw” Seepage Index values along the canal section surveyed for this study.	50
Figure 26. Example from South Canal results showing the calculated “raw” Seepage Index (left plot), the detrended Seepage Index (center plot), and the detrended-shifted-clipped-renormalized Seepage Index values (right plot).	51
Figure 27. Detected seepage locations along South Canal using thresholding of the detrended Seepage Index values. Here, red circles are “most likely” seepage, yellow circles are “likely”	

seepage, and black circles are “possible” seepage. The “raw” Seepage Index values are plotted under these predicted locations.	52
Figure 28. A detailed view showing an example of a successful seepage prediction using thresholding of the Seepage Index. Here, a known seepage location was provided by the Western Colorado Area Office, and the location matches almost perfectly with a predicted location of “most likely seepage” (red circle) and a surrounding band of “likely” seepage (band of adjacent yellow circles). Note that the “shallow seepage” waypoints labeled on the figure just to the north of the wet-spot location was simply surficial leakage through a crack in the canal’s liner, and likely not associated with significant subsurface saturation and corresponding conductivity anomaly.	53
Figure 29. A second detailed view near the southern end of the segment of South Canal tested. Here, various successful seepage detections and generally elevated Seepage Index values are validated with extensive observed apparent seepage (noted and coordinates recorded with a hand-held GPS during data collection).	54
Figure 30. A third detailed view along South Canal showing seemingly successful seepage detections and generally elevated Seepage Index values consistent with visible vegetation patterns.	55
Figure 31. Results of K-means clustering applied to the original and normalized (between 0 and +1) data-types and metrics along South Canal.	56
Figure 32. Stacked bar graph showing the results of K-means clustering for seepage prediction. This result is rather promising, where most seeps are identified with cluster number 4, and the majority of the other three clusters are predominantly background “non-seepage” segments (based on Seepage Index thresholding).	57
Figure 33. A spatial plot of the results of SVM prediction results.	58
Figure 34. Comparisons and evaluations of SVM prediction results relative to thresholding-based predictions along the South Canal study area. In general, these results show an excellent application of SVMs for seepage prediction, based on the nature of our input multidimensional data set.	59
Figure 35. Detected seepage locations along Government Highline Canal using thresholding of the detrended Seepage Index values. Here, red circles are “most likely” seepage, yellow circles are “likely” seepage, and black circles are “possible” seepage. The “raw” Seepage Index values are plotted under these predicted locations.	61
Figure 36. A detailed view of Government Highline Canal showing seemingly successful seepage detections and generally elevated Seepage Index values consistent with visible vegetation patterns.	62
Figure 37. A second detailed view showing various successful seepage detections and generally elevated Seepage Index values in areas of either observed apparent seepage (noted and coordinates recorded with a hand-held GPS during data collection) or elevated vegetation growth.	63
Figure 38. A third detailed view along Government Highline Canal showing successful seepage detections and generally elevated Seepage Index values consistent with observed seepage in the field and vegetation patterns.	64
Figure 39. Results of K-means clustering applied to the original and normalized (between 0 and +1) data-types and metrics along Government Highline Canal.	65
Figure 40. Stacked bar graph showing the results of K-means clustering for seepage prediction. While most of cluster 1 captures seepage (and captures all “most likely” seepage points), the	

results of clusters 2 and 3 are not as successful in differentiating between “likely” and “possible” seepage locations with background “non-seepage” locations (based on Seepage Index thresholding).	66
Figure 41. A spatial plot of the results of SVM prediction results along Government Highline Canal.	67
Figure 42. Comparisons and evaluations of SVM prediction results relative to thresholding-based predictions along the Government Highline Canal study area. In general, these results again show an excellent application of SVMs for seepage prediction, based on the nature of our input multidimensional data set.	68
Figure 43. Detected seepage locations along Strawberry Highline Canal using thresholding of the detrended Seepage Index values. Here, red circles are “most likely” seepage, yellow circles are “likely” seepage, and black circles are “possible” seepage. The “raw” Seepage Index values are plotted under these predicted locations.	70
Figure 44. A detailed view showing an example of a failed seepage prediction using thresholding of the Seepage Index (false negative at “Erik Robison’s Seep”) due to the presence of extensive reinforced concrete lining that resulted in suppression of the Seepage Index values at this location. However, two nearby general locations adjacent to this false negative were pointed out in the field by Provo Area Office personnel, and these locations appear to have successful predictions of seepage.	71
Figure 45. A second detailed view near the southern end of the segment of South Canal tested. Here, various successful seepage detections and generally elevated Seepage Index values are validated with extensive observed apparent seepage (noted and coordinates recorded with a hand-held GPS during data collection).	72
Figure 46. Results of K-means clustering applied to the original and normalized (between 0 and +1) data-types and metrics along South Canal.	73
Figure 47. Stacked bar graph showing the results of K-means clustering for seepage prediction. While most of clusters 1 and 2 contain “non-seepage” locations, and most of cluster 3 only captures seepage locations (and captures virtually all “most likely” seepage points), the results of cluster 4 is not as successful in differentiating between “likely” and “possible” seepage locations with background “non-seepage” locations (based on Seepage Index thresholding).	74
Figure 48. A spatial plot of the results of SVM prediction results along Strawberry Highline Canal.	75
Figure 49. Comparisons and evaluations of SVM prediction results relative to thresholding-based predictions. In general, these results show an excellent application of SVMs for seepage prediction, based on the nature of our input multidimensional data set. As expected, the worst fit is found with “possible” seepage location predictions, and yet 72% are still predicted correctly, with the majority of incorrect predictions being a conservative false positive prediction. This most likely has to do with the level of extensive vegetative growth along much of the length of Strawberry Highline Canal that may or may not have to do with seepage.	76

Background

Problem

Here in the United States alone, there are currently over 100,000 miles of canal and levee embankments and approximately 79,000 dams on the national inventory list. Many of these structures are reaching or have surpassed their initial design life, where most earthen embankment structures are over 50 years in age, and many are greater than 80 years in age. Development and urban encroachment along this aging infrastructure increases risk associated with concentrated seepage and internal erosion-caused failures due to the increased consequences of these failure events (i.e., greater economic damage to downhill properties and, increased costs of follow-up investigations and litigation and repairs, and a higher likelihood of injury or loss of life).

In many cases, canal embankment failures are triggered by networks of rodent burrows that can become established overtime within the embankment at or near the waterline. These burrows open conduits for water to enter and tend to “short-circuit” the hydraulic gradient by means of shortening the distance from the water to the downstream face or toe of the structure during high-stage water levels. This embankment failure process is oftentimes but not always preceded by the development of localized seeps along the embankment slope and toe.

Similarly, trees with large root systems tend to grow adjacent to canals, as the canal becomes a good source of water and nutrients. These root systems tend to grow towards the source of water (e.g., through or under an embankment and towards the canal invert). Eventually, voids and preferential seepage pathways very similar to rodent burrows can be created as the tree dies and the mature root systems begin to rot away.

In other cases, a physical or geologic condition can exist along canals and levees that leads to localized seepage and increases the chances of concentrated seepage development over time at a particular point along the embankment (e.g., sharp bends in the canal alignment, over-steepened or narrow embankment geometry, high permeability and low cohesive strength soils, shallow aquicludes such as impermeable bedrock that tends to focus otherwise distributed seepage into higher hydraulic gradients).

In addition to the risk of concentrated seepage and canal embankment failures, water conveyance systems are known to lose significant quantities of water due to distributed seepage. This distributed water loss leads to significant economic losses. “It has been estimated that, generally speaking, 30 percent of all water diverted for irrigation is lost by seepage. A way of visualizing this quantity of seepage is that the water diverted in 1946 for irrigation of 36 Bureau of Reclamation projects would have irrigated an additional 1 million acres of land, if there had been no seepage loss,” (Lancaster, 1952). These statistics are from over 70 years ago, and annual water losses from both lined and unlined canals throughout the nation’s aging inventory have likely worsened with time.

Motivation and Goals

Standard approaches to canal inspections predominantly involve visual inspections. These approaches often fail to detect problematic segments of canals, or fall short of providing a larger view of where distributed seepage is occurring, or to evaluate subsurface conditions that are not visible to the human eye. In some cases, localized monitoring sensors are being utilized (e.g., piezometers, flow-meters, and a more recently developed probes for measuring thermal profiles within the invert of canals to estimate local seepage rates). Additionally, geophysical surveys are occasionally conducted along canals. These geophysical surveys, typically electrical resistivity tomography (ERT) and self-potential (SP) surveys, involve relatively laborious fieldwork which typically takes days to collect and only provides very limited data coverage (e.g. a few thousand feet of data coverage after days of multi-person field crew efforts).

These various approaches are perfectly suited and typically applied at known problematic seepage locations already identified for the sake of better characterization, monitoring and imaging of an embankment defect or seepage zone. However, these spatially sparse and relatively expensive approaches fail to provide a realistic means for obtaining a more comprehensive view of where seepage is occurring along the entire length of a canal embankment that is several miles long.

As a result, there is a need for effective yet efficient characterization technologies that can be utilized along these vast infrastructural systems. Specifically, there is a need for rapid reconnaissance surveying technologies that will help provide system-wide canal embankment data coverage and improve our understanding of the health conditions along our aging canals. This approach would help to better inform, guide, and prioritize more focused and costly inspections, monitoring, and mitigation efforts.

The main goals of this research include the following:

- 1) Develop a means to perform rapid geophysical profiling data collection along canal embankments.
- 2) Develop a new semi-automated approach to combining and analyzing rapid ground-based geophysical profiling of earthen canal embankments with satellite remote sensing data in a manner capable of detecting and mapping seepage along large lengths of canal systems, or on a system-wide scale.

General Approach

The general approach taken in this study involved the following:

- 1) Testing of various commercially available geophysical instrumentation on the Denver Federal Center for system performance and applicability for seepage detection.
- 2) Developing the capabilities to collect geophysical profiling data along canal embankments in a very rapid manner, by means of constructing a rugged non-metallic cart platform used for mounting geophysical instruments that are integrated with mapping-grade differential global positioning systems (DGPS) and capable of collecting georeferenced data at relatively high rates of speed (e.g. towed behind an all-terrain vehicle or truck at nominally 5 to 10 miles per hour).
- 3) Develop semi-automated and robust data processing and analysis workflows capable of integrating 2D NDVI datasets with various 1D geophysical profiling datasets.
- 4) Development of a new “Seepage Index,” or a normalized data attribute capable of detecting seepage-related data patterns and anomalies and that provides a relative indication of the likelihood of seepage along vast length of canal embankments.
- 5) Preliminary testing of a trained machine learning algorithm (Support Vector Machines) and multi-dimensional data clustering to evaluate the ability for automatic prediction of where seepage may be occurring along canals.
- 6) Comparing seepage detection results with field observations and known seepage locations along canals surveyed as part of this research project.

Study Locations

This research project’s approach was developed and tested by means of system evaluations, modifications, data collection and analysis at four different field locations:

- 1) The Denver Federal Center (DFC), near the corner of Main Ave. and 7th St.
- 2) An approximate 9-mile segment of South Canal located southeast of Montrose, Colorado.
- 3) An approximate 9-mile segment of Government Highline Canal located north of Grand Junction, Colorado.
- 4) An approximate 13-mile segment of Strawberry Highline Canal located south of Provo, Utah and extending between the cities of Spanish Fork and Santaquin, Utah.

These four testing field sites are depicted in Figure 1 through Figure 4 respectively. Here, initial geophysical system evaluations and non-magnetic cart design permutations were tested at the DFC test-site.



Figure 1. Map showing the approximate 4-acre area where preliminary non-magnetic cart configurations and geophysical system testing were performed (red polygon) located south of Building 56 and near the intersection of Main Ave. and 7th St. on the Denver Federal Center, Colorado.

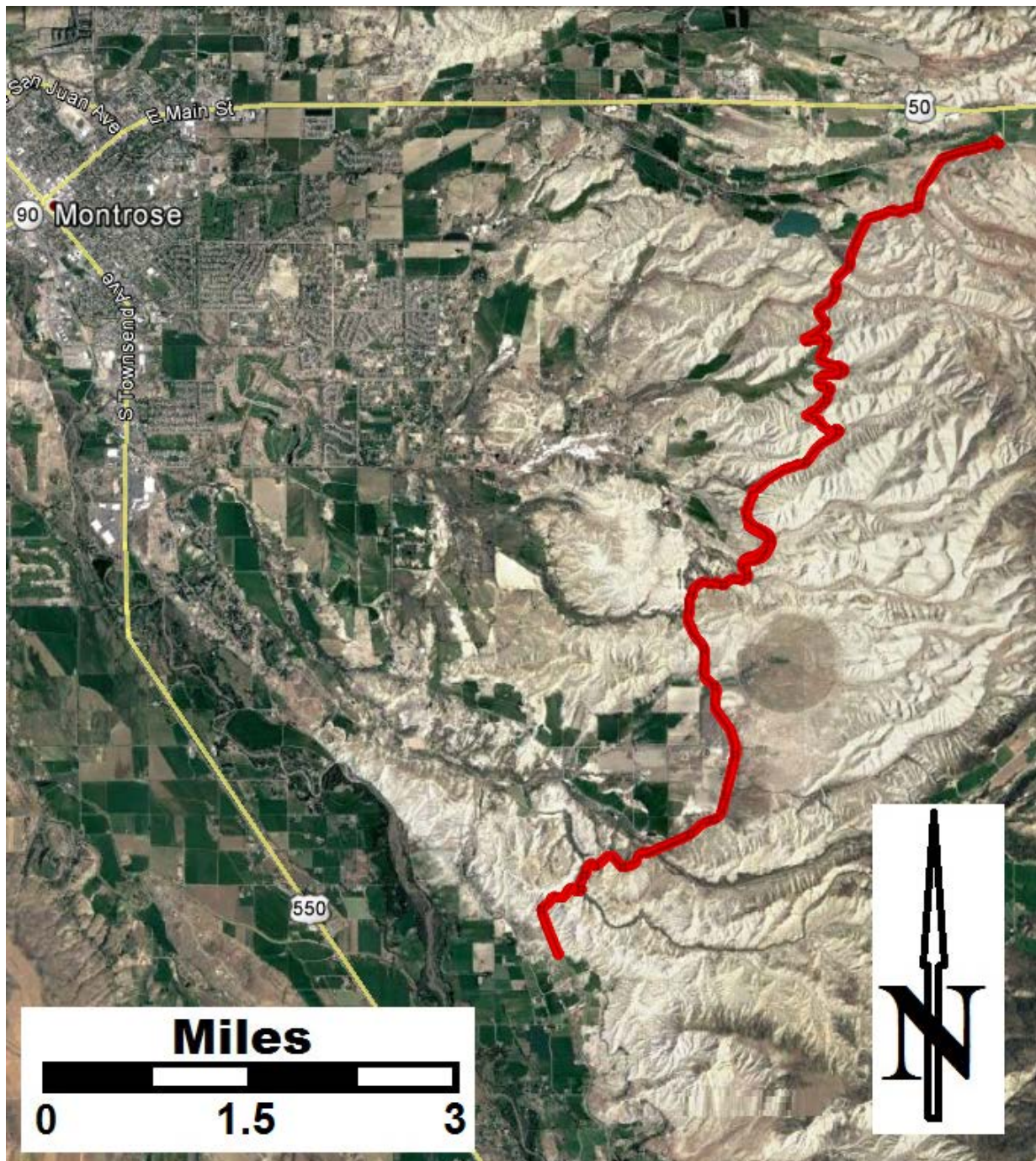


Figure 2. Map showing the study area and data coverage along an approximate 9 mile segment of South Canal (red line) located southeast of Montrose, Colorado (managed by the Western Colorado Area Office).

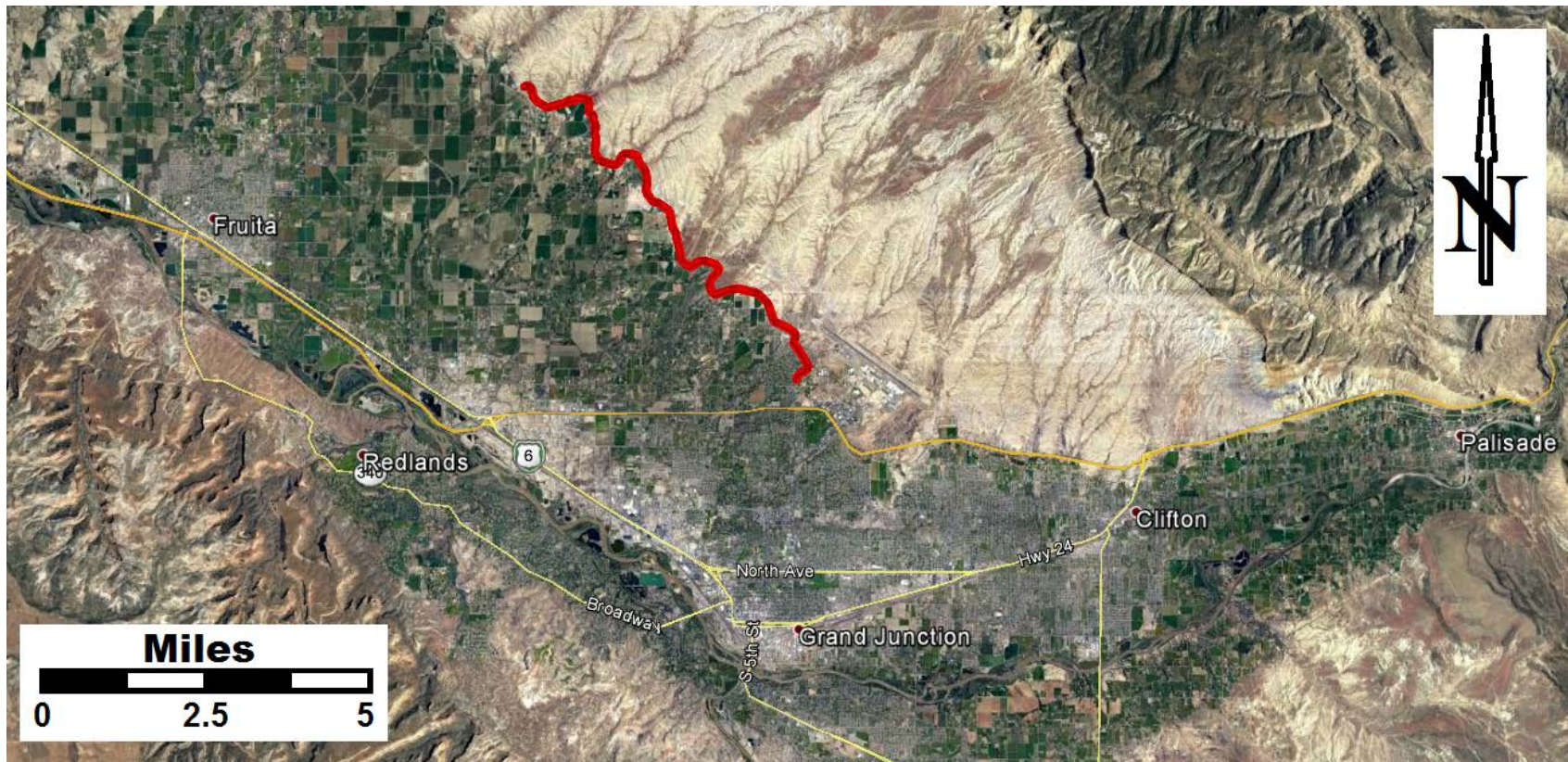


Figure 3. Map showing the study area and data coverage along an approximate 9 mile segment of Government Highline Canal (red line) located north of Grand Junction, Colorado (managed by the Western Colorado Area Office).

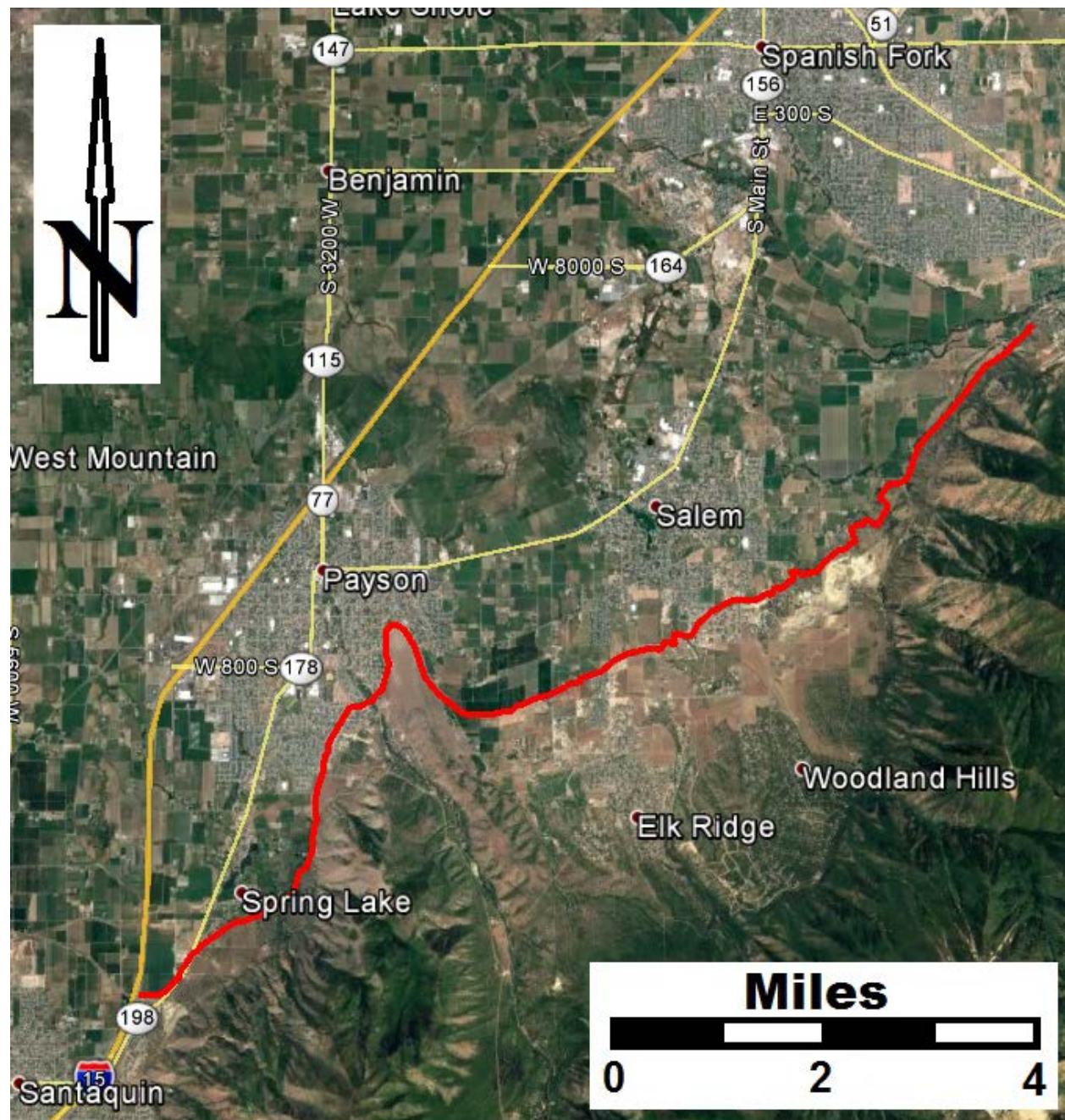


Figure 4. Map showing the study area and data coverage along an approximate 13 mile segment of Strawberry Highline Canal (red line) extending between the cities of Spanish Fork and Santaquin, Utah (managed by the Provo Area Office).

Methods

Conceptual Model of Canal Seepage

Figure 5 depicts conceptual models of through-seepage and under-seepage conditions that can occur along canals. In both cases, there is typically an increase in the water saturation within and underneath the embankment, as well as immediately downhill of the embankment with the eventual development of vegetation at a seep location. In under-seepage conditions, the embankment slope surface can be completely dry and bare soil while there is standing water and lush vegetation along the downstream toe. Conversely, through-seepage can lead to shallow wetting-fronts within the embankment, and wet-spots and vegetation can develop further up along the slope.

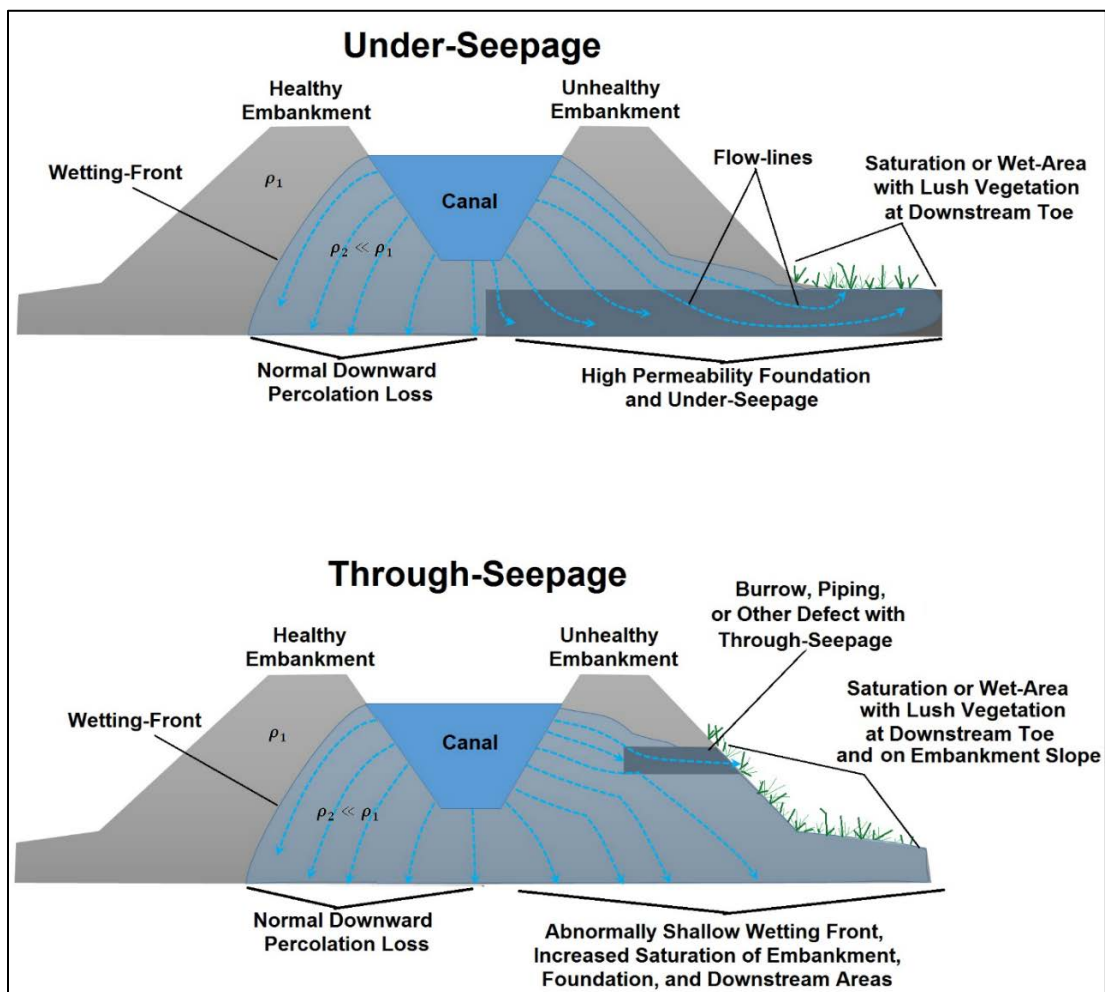


Figure 5. A conceptual model of canal seepage and related changes in soil water saturation, electrical resistivity, and plant growth. The top and bottom images depict under-seepage and through-seepage, respectively.

Conveniently, the seepage-related conditions depicted in Figure 5 can be targeted for seepage detection purposes. Specifically, the increase in electrical conductivity (a corresponding decrease in resistivity) due to increases in water saturation can be targeted with geophysical techniques, and the spatial variations in vegetation along a canal can be used as an indirect indication of seepage. Accordingly, this research has focused on developing a new approach for combining NDVI remote sensing data images with ground-based FDEM and magnetic gradiometry profiling surveys conducted along canal embankments. These data are explored and combined to create a means to more accurately detect and map embankment seepage in comparison to utilizing a single data type only.

As discussed in more detail below, magnetic gradiometry was added to the geophysical survey types utilized in this study, in order to help identify the presence of metallic infrastructure (e.g. turnout structures, weirs, metal conduit perforations, reinforced concrete lining, storm drains, guardrails at road crossings, etc.) and to help differentiate between seepage and metal-related anomalies captured in FDEM profiling data.

Frequency-Domain Electromagnetics

FDEM systems are used to measure electrical properties of the subsurface, and FDEM data can be collected using DGPS for absolute coordinate data positioning, or using an arbitrary local coordinate system (e.g., tape measure distance along lines), if needed. FDEM surveys can be conducted along one-dimensional (1D) survey profiles (e.g., independent dataset collected along a single path) or in a two-dimensional (2D) mapping fashion (e.g., across a grid of overlapping X and Y-oriented survey lines). FDEM data are then post-processed, filtered, and plotted as either 1D line graphs or 2D contour plots to indicate the average electrical conductivity of the subsurface.

As depicted in Figure 6, FDEM systems are based upon the principle of electromagnetic induction, and generally consist of two coils: a transmitter coil that energizes the subsurface materials, and a receiver coil that records the induced response from the subsurface. Here, the transmitter coil generates a time-varying “primary” magnetic field which induces eddy-currents (electrical currents) within conductive objects or zones within the subsurface. These time-varying eddy-currents generate a corresponding time-varying “secondary” magnetic field that is then recorded by a receiver coil (the time-varying secondary magnetic field induces a proportional current within the receiver coil).

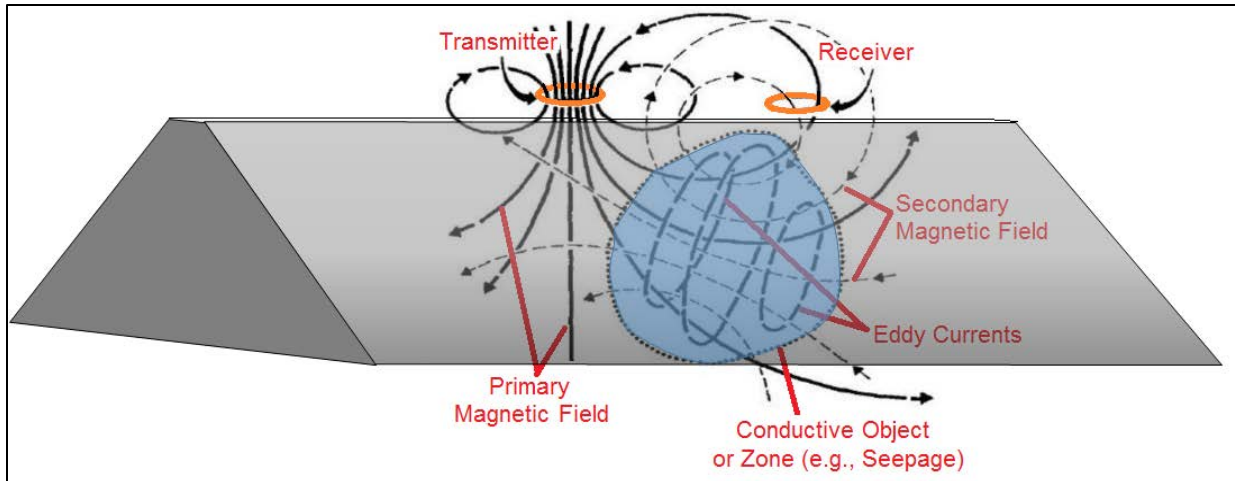


Figure 6. Conceptual diagram showing inductive coupling of a FDEM system.

In this scenario, the strength of the induced eddy currents and associated secondary magnetic field is directly proportional to the electrical conductivity of the subsurface, within some proximity to the FDEM system (e.g., within the depth of investigation for a given FDEM system). This is the main reason why FDEM systems are often referred to as “conductivity meters.” Here, the depth of investigation (DOI) is a function of four main factors or system parameters:

- 1) The operating frequency of the transmitter coil
- 2) The strength (electrical current) of the transmitter coil’s output
- 3) The orientation of the transmitter and receiver coils
- 4) The separation between the transmitter and receiver coils

Generally, increased transmitter current, increased inter-coil spacing, and lower system frequencies result in a corresponding increase in DOI. The coil orientations (horizontal or vertical) also play a role in the FDEM system’s DOI, where a vertical dipole orientation (coils oriented parallel to the ground) tends to increase the DOI. Orientation also tends to change the system’s sensitivity versus depth profile, where a vertical dipole orientation has a slightly greater DOI compared to horizontal dipole orientation of the same FDEM system. However, a vertical dipole orientation has very low sensitivity at relatively shallow depths, with a “lobe” of maximum sensitivity at approximately one half of the inter-coil spacing, as depicted in Figure 7:

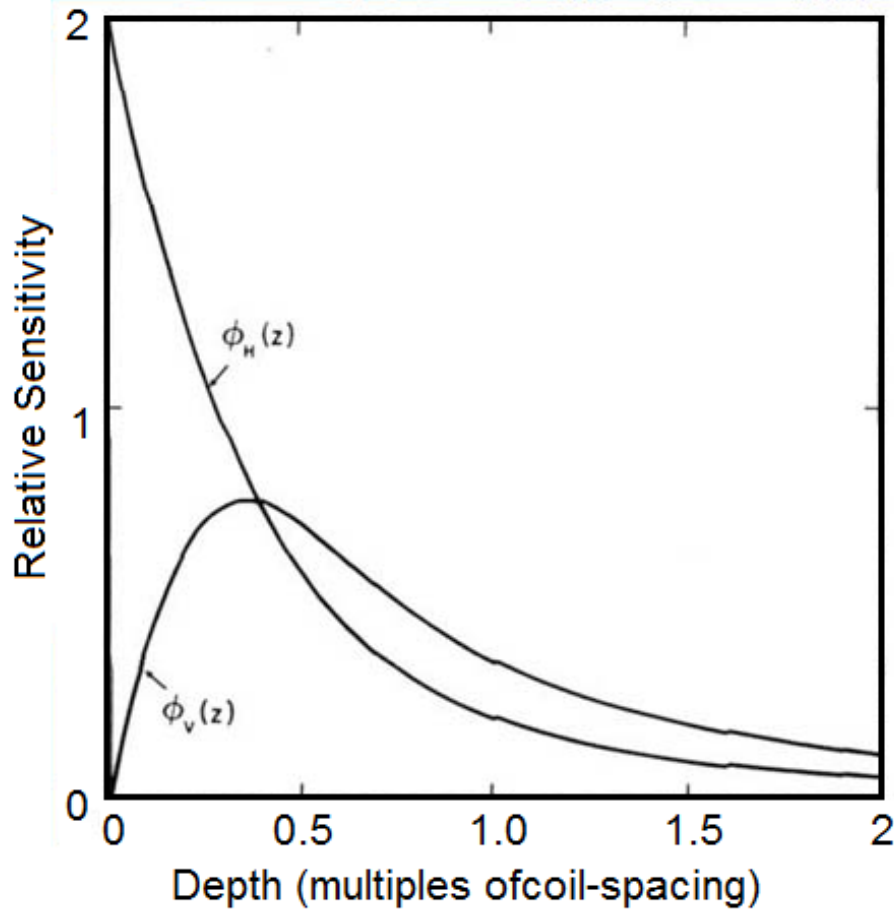


Figure 7. Diagram of an FDEM system's sensitivity as a function of depth (z) below ground surface for both a vertical dipole orientation ($\phi_V(z)$ curve), and for a horizontal dipole orientation ($\phi_H(z)$ curve). Image modified from (Geonics 2018).

For this research project several commercially available FDEM systems were evaluated¹ and utilized for mapping conductivity changes related to canal seepage. Specifically, the following FDEM systems were tested and evaluated on the Denver Federal Center for robustness, repeatability, sensitivity, and applicability to seepage detection:

- 1) GSSI Inc's EMP-400 (multi-frequency, fixed coil-spacing)
- 2) DualEM Inc's DualEM-2 (single-frequency, fixed coil-spacing)
- 3) Geonics Ltd's EM31-mk2 (single-frequency, fixed coil-spacing)
- 4) Geonics Ltd's EM34-mk2 with 10m coil separation (single-frequency, variable coil-spacing)
- 5) Geonics Ltd's EM38-mk2 (single-frequency, two fixed coil-spacings)

¹ Mention of trade names or commercial products does not constitute endorsement or recommendation for use by the U.S. Government.

Ultimately, it was decided to not utilize the DualEM-2 or the EMP-400 systems for field-scale implementation along canals. This decision was mainly based on system usability and availability, system performance (repeatability) and sensitivity to lateral changes in subsurface conditions and utilities (e.g. metallic infrastructure), and value of information added to the study, in terms of offering unique yet applicable DOI for most canal seepage conditions and embankment sizes encountered within Reclamation’s inventory and chosen test field sites.

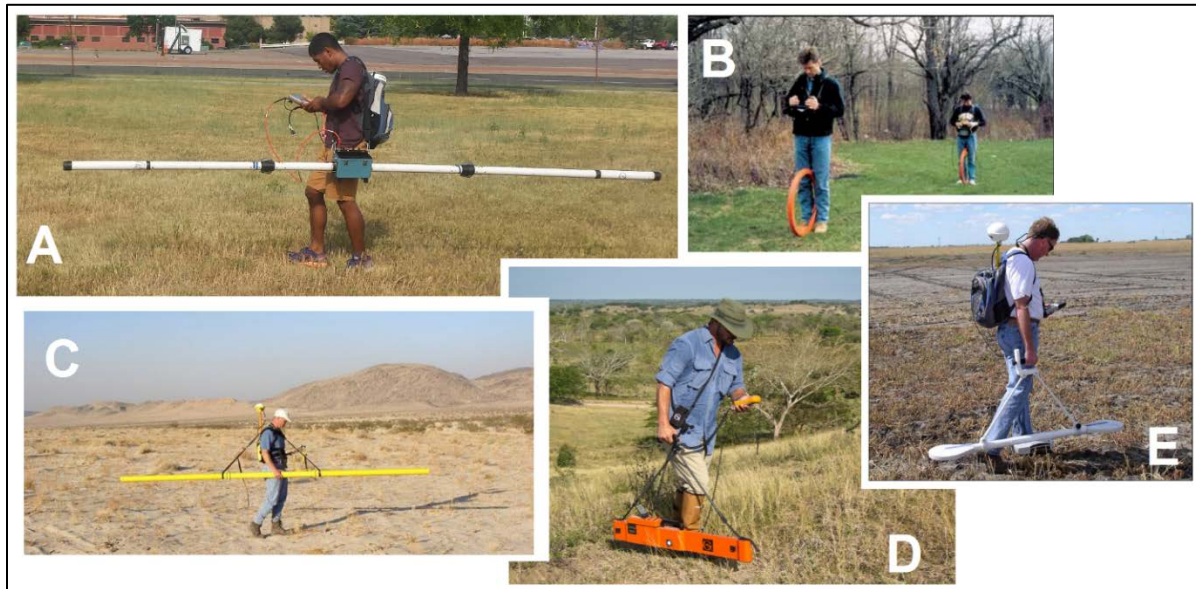


Figure 8. Five FDEM systems tested for this research project: Geonics Ltd’s EM31-mk2 (A), Geonics Ltd’s EM34-mk2 (B), DualEM inc’s DualEM-4 (C), Geonics Ltd’s EM38-mk2 (D), and GSSI inc’s EMP-400 (E).

In order to measure the electrical conductivity of the subsurface (more specifically, a depth-weighted average of the subsurface conductivity profile), FDEM systems also need to record both the amplitude of the secondary magnetic field and the phase difference between the primary and secondary fields. Here, the amplitude (i.e., strength) of “quadrature component” (the component of the secondary field that is 90 degrees out of phase relative to the transmitter waveform) is directly proportional to the subsurface “terrain conductivity.” Conversely, the amplitude of the “in-phase component” (the component of the secondary field that is zero degrees out of phase relative to the transmitter waveform) is generally indicative of the presence of nearby metallic infrastructure or subsurface buried metallic objects. Thus, the recorded in-phase component can be used to help distinguish between abrupt increases in measured conductivity (quadrature-component) as being mineralogical or hydrogeological in nature, as opposed to being the result of nearby metal.

However, in most cases, the in-phase and quadrature components are highly correlated signals. Therefore, it is usually quite difficult to distinguish between metal-related and non-metal related conductivity anomalies using the in-phase component only. This particular challenge is the main motivation for including magnetic gradiometry data for this research project, since we are most interested in identifying changes in terrain conductivity related to canal seepage.

Magnetic Gradiometry

Magnetometers are geophysical instruments that measure the strength and orientation of natural or anthropogenic magnetic fields, using various sensing technologies. In this study, we utilized cesium vapor proton-precession magnetometer sensors that measure only the strength of the total magnetic field (e.g., the sum of the X, Y, and Z components), but not the direction of the field (e.g., does not record the inclination and declination).

Similar to FDEM systems, magnetometers can be configured to collect data in 1D or 2D fashions, utilizing DGPS systems for data positioning and plotting. Magnetometers measure a combination of the following:

- 1) Earth's natural background magnetic field (e.g., what we measure as magnetic north)
- 2) Remnant magnetization of objects and materials (i.e., locked-in magnetizations of materials, similar to permanent magnets)
- 3) Induced magnetization of high magnetic permeability materials (e.g., ferrous metals such as iron and steel become magnetized by exposure to Earth's background field).

When mapping magnetic fields across a survey area, there is typically an approximately constant background magnetic field (Earth's very slowly changing dynamo), plus the occasional magnetic storm-related noise event that is generated from coronal ejecta interacting with Earth's magnetosphere and ionosphere. These background fields are superimposed by local trends and patterns caused by geologic structures, and also sharp and high-amplitude anomalies caused by induced or remnant magnetization of any nearby metal objects above ground or in the near subsurface (e.g., chain-link fences, ferrous pipes, conduits, wires, reinforced concrete, metal shrapnel, ordinances, steel-lined underground storage tanks, etc.).

Figure 9 shows a photo of a Geometrics Inc. G858 magnetometer system configured in a vertical gradiometer fashion, with two cesium-vapor magnetometer sensors vertically aligned one half meter apart, and with DGPS positioning. This is the magnetometer system utilized for this study, here both vertical and horizontal gradiometry configurations were tested. When collecting magnetic data with two or more sensors placed at some small distance apart, such as that shown in Figure 9, Earth's background magnetic field is virtually identical at each magnetic sensor, and so the resulting gradient of the magnetic field between the sensors is nearly equal to zero. However, in the presence of relatively shallow or nearby metallic objects or infrastructure with some combination of induced and remnant magnetization, there is typically a relatively strong local gradient in the magnetic field surrounding the magnetized object, and so there is a correspondingly large difference in the recorded magnetic field strengths at each sensor.

Detection of this difference in field intensity between the two sensors is the essence of magnetic gradiometry, which is particularly effective at mapping localized magnetic anomalies associated with shallow or nearby metal. Metal detection was the primary application of magnetic gradiometry profiling for this study, in order to identify and map the locations of significant metallic infrastructure along canals. This information is potentially useful by itself, in that it could be used to verify and map the locations of abandoned turnouts or illegal canal perforations, for example. Furthermore, it is useful to identify FDEM anomalies that are more likely caused inductive coupling with nearby metal, as opposed to embankment seepage.



Figure 9. Photo of a Geometrics Inc. G858 magnetometer system configured in a vertical gradiometer fashion, with two cesium-vapor magnetometer sensors vertically aligned one half meter apart, and with DGPS positioning.

Enabling Rapid Data Collection: Non-magnetic Cart Design

FDEM systems and magnetometers are affected by significant nearby metal, such that data collection and surveying with these instruments is normally performed by carrying the systems by hand. However, mounting and towing these systems on non-metallic platforms (e.g., rafts, sleds, and plastic and fiber-glass carts) is an option to avoid carrying the systems by hand. With adequate offsets, a bicycle or motorized vehicle can be utilized to tow the platform, which is the approach utilized in this study to enable rapid data collection along miles of canal embankment per day.

Figure 10 and Figure 11 each show the evolution of the cart that was designed and built for this research project. The main challenge of this design and build was to have an adequate offset from a towing vehicle as to not negatively affect system performance and resulting data while still maintaining a rugged enough of a platform to support an EM34 system (nominally 45lbs) with a 10m coil separation. The cart also needed to be rugged enough to withstand hundreds of miles of rough canal embankment road conditions.

The first design iteration involved the use of a vertical mast and rope supports for supporting the rear coil, but this was deemed too unstable and problematic, and still required the receiver coil to be very close to the vehicle (see top photo in Figure 10). Testing showed that having the transmitter close to the vehicle affected the data due to inductive coupling with the nearby metal, but having the receiver coil next to the vehicle was not a major influence on data quality.

Regardless, it was decided to instead build a tethered mini-cart that is attached and towed behind the main cart platform with a rubber pivoting arm connection to allow for system actuation over hills, dips, and around sharp turns (see top photos in Figure 11). The original main cart platform is still sufficient for carrying smaller instruments, including the EM38, the EM31, and the magnetic gradiometer (see top photos in Figure 11)



Figure 10. Two photos showing the first iteration of the non-magnetic cart design, being towed behind a truck at the DFC test-site while carrying an EM34 system with 10m coil separation and horizontal dipole orientation (top) and the EM31 FDEM system (bottom).



Figure 11. Photos showing the current version of the non-magnetic cart configuration, now utilizing a tethered and gimbaled min-cart for sake of carrying the quite long (~30ft) EM34 system (top two photos). The original main cart is still sufficient for carrying smaller instruments, including the EM38, EM31, and the magnetic gradiometer (bottom photo).

Normalized Difference Vegetative Index

NDVI is a remote sensing data metric aimed at identifying the presence and extent (e.g., density and lushness) of vegetation. NDVI values are typically calculated on a per-pixel basis using satellite remote sensing images of the ground surface's spectral reflectance of both visible-red and near-infrared bands of the electromagnetic spectrum. NDVI is calculated as follows:

$$NDVI = \frac{NIR - Red}{NIR + Red}$$

where *Red* and *NIR* stand for the spectral reflectance measurements acquired in the visible-red and near-infrared bands, respectively. Each of these input datasets or images are typically normalized between 0 and 1 first, and so NDVI data are normalized between -1 and +1. In general, lush and dense vegetation has a high NDVI value, while bare dry ground surface and bodies of water have a very low NDVI value (see Figure 12).

When utilizing more than one satellite image for a given band, each of the multiple images (e.g., a mosaic) must first be corrected for radiometric factors that affect the absolute measured spectral reflectance values. These radiometric corrections generally include absolute calibration of the satellite sensor, and accounting for the different Sun zenith angle, Earth-Sun distance, and atmospheric conditions experienced during collection of each individual data image.

For this research project, free satellite imagery from the National Agriculture Imagery Program (NAIP) was utilized for producing NDVI datasets with a spatial resolution of 1.5 meters (e.g., pixel size). An open-source remote sensing data processing software package produced by Harris Corporation and called ENVI was used for processing the NAIP imagery, and calculating NDVI values. In some cases, radiometric corrections were not correctly applied before calculating NDVI values for use in this research project. While this caused specific issues with two of the three final NDVI images used (the different satellite image areas within the NDVI image each have different dynamic ranges and averages of NDVI values), it was only a small area of two surveys performed that were slightly affected by this issue, and the overall impact on results was negligible.

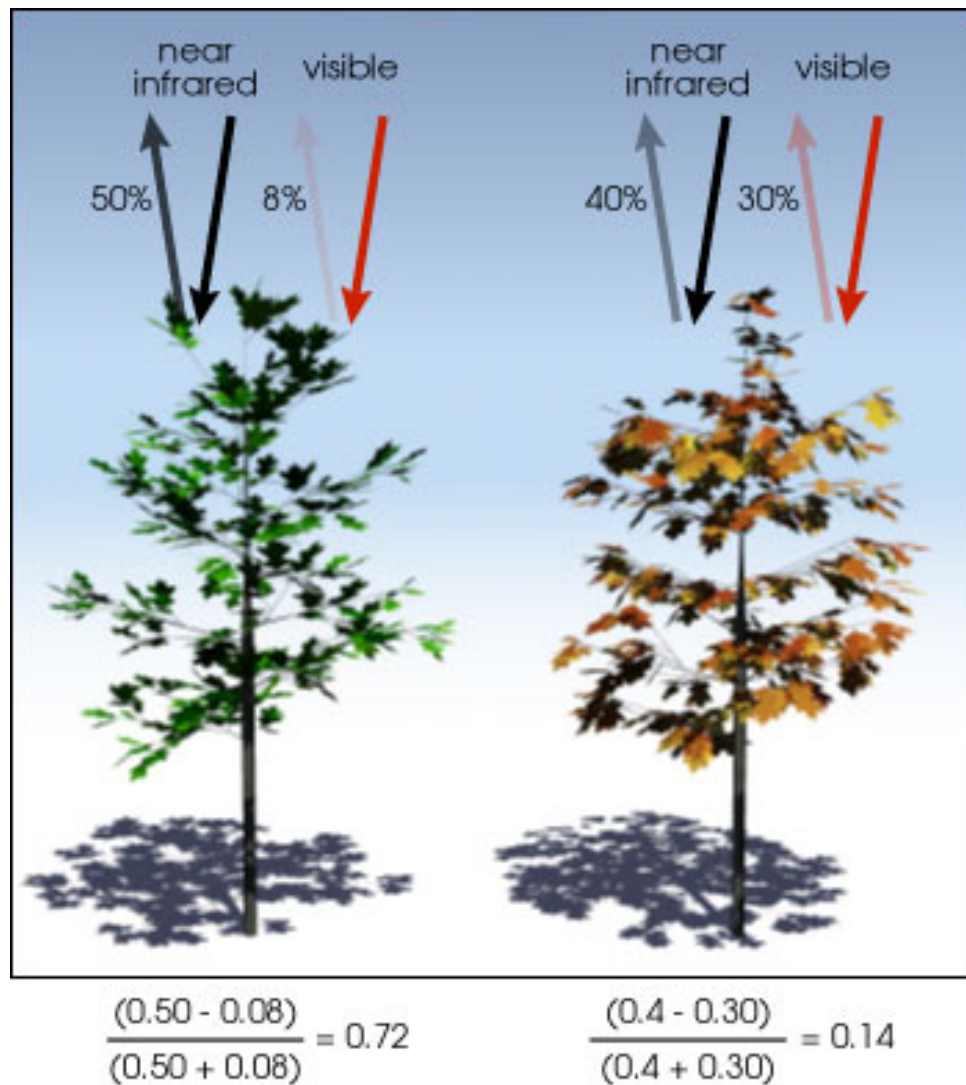


Figure 12. Example of how NDVI works to identify lush green vegetation based on different spectral reflectances of NIR and visible-red light bands. Image taken from (NASA, 2000).

Developing a New Data Attribute: Normalized Seepage Index

One of the main objectives of this research project was to develop means of combining various data types in order to better detect and map the presence of embankment seepage along canals (and earthen levees, which are essentially the same as a canal embankment except for the general lack of flowing water experienced with canal water conveyance). The underlying challenges with this objective involve the combining of multiple geophysical profiling datasets with remote sensing geospatial imagery (NDVI data) in a manner that robustly detects and maps seepage. Furthermore, this effort was aimed at making the processing, merging and interpretation of these various data types semi-automated. As a result, an effort was made to develop a “Seepage Index,” and data attribute similar to that of NDVI, where Seepage Index values are arbitrarily normalized between 0 and +1. Here, relatively large index values close to +1 indicate the likely

presence of seepage, while relatively small index values closer to zero indicate “normal” embankment performance.

Preliminary Approach

At the onset of this research project, it was first expected to use only FDEM profiling data combined with NDVI data to develop an indicator of seepage. This approach was born out of the concept that combining NDVI and FDEM data would help to prevent false negatives and false positives, where seepage areas may not always create all the conditions described in one of the conceptual seepage models shown in Figure 5. Specifically, a given seepage area may or may not be associated with anomalous vegetation growth and corresponding elevated NDVI values relative to background NDVI values and patterns along the canal. Similarly, a given seepage area may or may not be associated with an anomalous increase in electrical conductivity. However, if there is both an increase in NDVI values and an increase in electrical conductivity at a given location along a canal, this scenario would lead to a fairly compelling indication of active seepage at that location, and the magnitudes of increases in these data values relative to background values could offer some indication of seepage severity.

The following equation represents a first attempt at creating a normalized Seepage Index:

$$SeepageIndex_1 = Q_{em34} \times \frac{Q_{em31}}{IP_{em31}} \times \frac{Q_{em38}}{IP_{em38}} \times NDVI_{ave}$$

Here, Q_{em34} , Q_{em31} , and Q_{em38} are the apparent terrain conductivities recorded during EM34, EM31, and EM38 FDEM profiling surveys (each set of values first normalized from +1 to +2), respectively. Similarly, IP_{em31} , and IP_{em38} are the in-phase system responses from the EM31, and EM38 FDEM profiling surveys. In this first approach, NDVI values were simply averaged along the length of the canal, and so the $NDVI_{ave}$ term in the above equation is the running average of nearby NDVI values along the length of canal being studied. In this approach, multiplication of FDEM conductivities and averaged NDVI values should result in large values in the presence of seepage, while dividing by the in-phase components helps to compensate for increases in conductivity due to the presence of metal.

This point-wise calculation is achieved by first establishing a “master GPS track” along the length of the canal embankment being inspected. Next, a swath of nearby NDVI data within some user-define radial distance (10 meters for this study) surrounding each point along this master GPS track is extracted from a larger input NDVI image. Then, each normalized data type is interpolated to the latitude and longitude coordinates of the master GPS track, such that each required data type is available at each point along the canal. In order to avoid excessive decimation of input data resolutions, a 1m coordinate interval was used for the master track in this study.

Addition of Magnetic Gradiometry Data

After initial development and testing of this preliminary Seepage Index approach using South Canal datasets, it quickly became apparent that metallic structures, (e.g., turnouts, roadway crossings, spillways, gates, bridges, and hydroelectric power generators, etc.) were causing detrimental impacts on the analysis and automated interpretation outcomes. This challenge was mostly in the form of very large spikes in electrical conductivity values due to nearby above-

grade and subsurface metallic features. These metal-related spikes resulted in extremely large ranges in conductivity values, and subsequent normalizations of FDEM datasets for use in calculating Seepage Indices resulted in significant suppression of the seepage-related data patterns and anomalies of interest (e.g., loss of detail and contrast during compression of large dynamic-range FDEM data). As mentioned before, the high level of correlation between the quadrature and in-phase components of the FDEM data prevented the sufficient suppression of metal-related peaks in Seepage Index values.

In order to address this challenge, it was decided to also implement magnetic gradiometry in order to independently detect, map, and remove the effects of metal in the final Seepage index values. Two approaches for accomplishing this were attempted with similar results. The first involves performing an ad-hoc clipping of Seepage Index values based on thresholding of magnetic gradient data, where:

- 1) Elevated Seepage Index values associated with strong magnetic gradients above a normalized threshold of 0.5 are assumed to be severely contaminated from the presence of significant metal, and are simply set equal to zero
- 2) Elevated Seepage Index values associated with magnetic gradient values above a normalized threshold of 0.3 are assumed to be partially influenced from the presence of metal, and are scaled by a factor of 0.1
- 3) Elevated Seepage Index values associated with magnetic gradient values above a normalized threshold of 0.15 are assumed to be potentially influenced from the presence of metal, and are scaled by a factor of 0.5

The second approach involves dividing each normalized input FDEM conductivity data type by the normalized magnetic gradient data, leading to the following approach to calculating a Seepage Index:

$$SeepageIndex_2 = Q_{em34} \times \frac{1}{(|\frac{dB}{dx}| + |\frac{dB}{dz}|)} \times \frac{Q_{em31}}{IP_{em31}} \times \frac{1}{(|\frac{dB}{dx}| + |\frac{dB}{dz}|)} \times \frac{Q_{em38}}{IP_{em38}} \times \frac{1}{(|\frac{dB}{dx}| + |\frac{dB}{dz}|)} \times NDVI_{ave}$$

Here, $|\frac{dB}{dx}|$ is the absolute value of the change in the magnetic field with respect to a change in horizontal location (i.e., the horizontal magnetic gradient), and $|\frac{dB}{dz}|$ is the absolute value of the change in the magnetic field with respect to a change in vertical location (i.e., the vertical magnetic gradient). During development and testing of this improved approach, it was determined that while vertical and horizontal gradients are both sensitive to the nearby presence of metal, the sum of both terms maximizes sensitivity and detectability of metal, regardless of whether the metal feature is in the subsurface below the magnetometer, or if it is off to the side of the canal embankment yet close enough to potentially influence FDEM conductivity values.

As mentioned previously, magnetic gradiometry is additionally useful for simply mapping/detecting abandoned or unauthorized embankment perforations, so this data type not only helps to improve seepage detection, but it also helps to more fully characterize subsurface conditions along canal embankments. One limit to this approach is that magnetometers are not sensitive to the presence of non-magnetic metals such as Aluminum, and so these types of metals will still impact the FDEM measured conductivity data. Luckily, most metallic infrastructure

related to canals and civil works are composed primarily of iron and steel, and so this is likely not a common problem for this application. Regardless, the in-phase component of FDEM data will still be sensitive to non-magnetic metals, and will still help to compensate for these materials in calculating the Seepage Index.

Addition of NDVI Statistics and Spatial Attributes

The next step in improving the normalized Seepage Index involved the addition of various statistics and spatial attributes or metrics related to the surrounding NDVI data set, as opposed to only utilizing a simple running average and maximum of NDVI values along the master DGPS track. Specifically, this step involved the development of code that performs various cross-line NDVI gradient calculations as depicted in Figure 13, in order to better account for spatial patterns in NDVI values that may or may not indicate seepage (e.g., high NDVI values on the upstream side of the canal embankment may not indicate seepage as does high NDVI values on the downstream side of the embankment). This was accomplished by performing the following:

- 1) First calculate the direction of motion associated with each data point along the master GPS track (e.g., the azimuthal headings at each data point).
- 2) Calculate the cross-line azimuth and the latitude and longitude coordinates of multiple points along this cross-line, both to the left and right side of the master GPS track, out to the user-defined radial distance used in NDVI data swath extraction (10m to either side of the master GPS track was used in this study).
- 3) Calculate the local cross-line gradient in the NDVI data in the direction of assumed seepage and immediately adjacent to each point along the master GPS track.
- 4) Calculate the longer-range cross-line NDVI gradients using each pair of cross-line points centered on the master GPS track (see black dots along cross-line in Figure 13).

Once the various cross-line gradients are calculated, the sum and maximum of the long-range gradients are calculated and saved for each point along the GPS track (see Figure 14). The local cross-line gradient values (e.g., immediately adjacent to the GPS track) are also saved as an additional data layer at each point along the GPS track for use in point-wise Seepage Index calculations. As indicated by the black circle labeled “A” on Figure 15, this local cross-line NDVI gradient (upper-right plot in Figure 15) is very sensitive to any narrow canal embankment sections where the canal water is close to the GPS track on one side and vegetation growing on the embankment slope or near the crest on the other side of the GPS track (e.g., is indicative of potential through-seepage or “weak points” along the canal embankment). Similarly, the sum and maximum of the long-range cross-line gradients help to capture patterns in vegetative growth further from the GPS track that are potentially related to under-seepage (black circle labeled “B” on Figure 15).

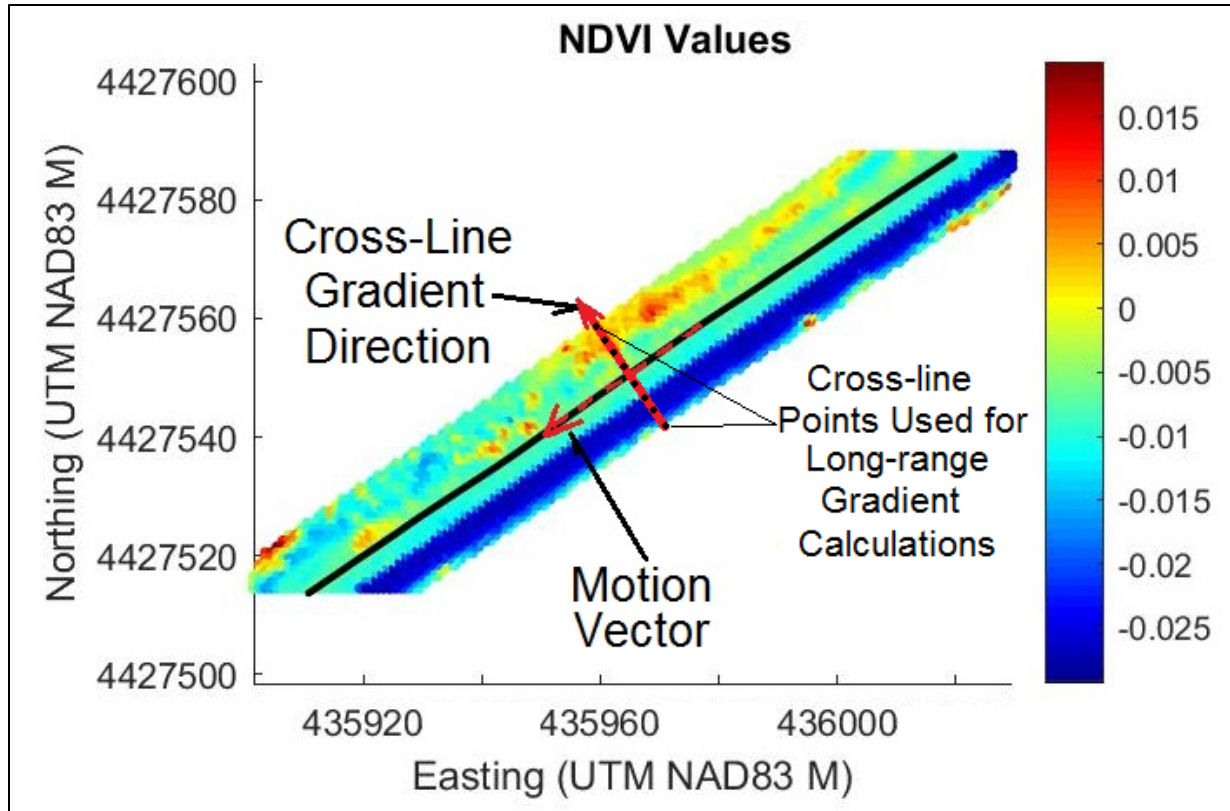


Figure 13. Example showing a segment of extracted NDVI values along the master GPS track, a calculated vector of motion of travel along the canal embankment, and the corresponding cross-line used to calculate the local and long-range gradients of NDVI values in the direction from canal to downstream toe (e.g., in the direction of expected seepage). In this figure, the dark blue strip along the lower half of the NDVI data swath is the canal/water, and the yellow and red NDVI values that speckle the upper-half of the swath is vegetation along the canal slope and downstream toe.

The directionality of the gradient calculations are defined by the analyst, where the canal is identified as being on the right or left side of the GPS track. As depicted by black circle labeled “C” on Figure 15, this directionality is important, as it helps to compensate for vegetation along the canal that may or may not be indicative of seepage (e.g., vegetation on the upstream side of the embankment normally does not indicate seepage). Here, the resulting negative NDVI gradients compensate for elevated running NDVI average and sums. In this fashion, segments of a survey that switch sides of the canal need to be processed and analyzed separately, in order to account for the expected cross-line direction of seepage and corresponding NDVI gradients.

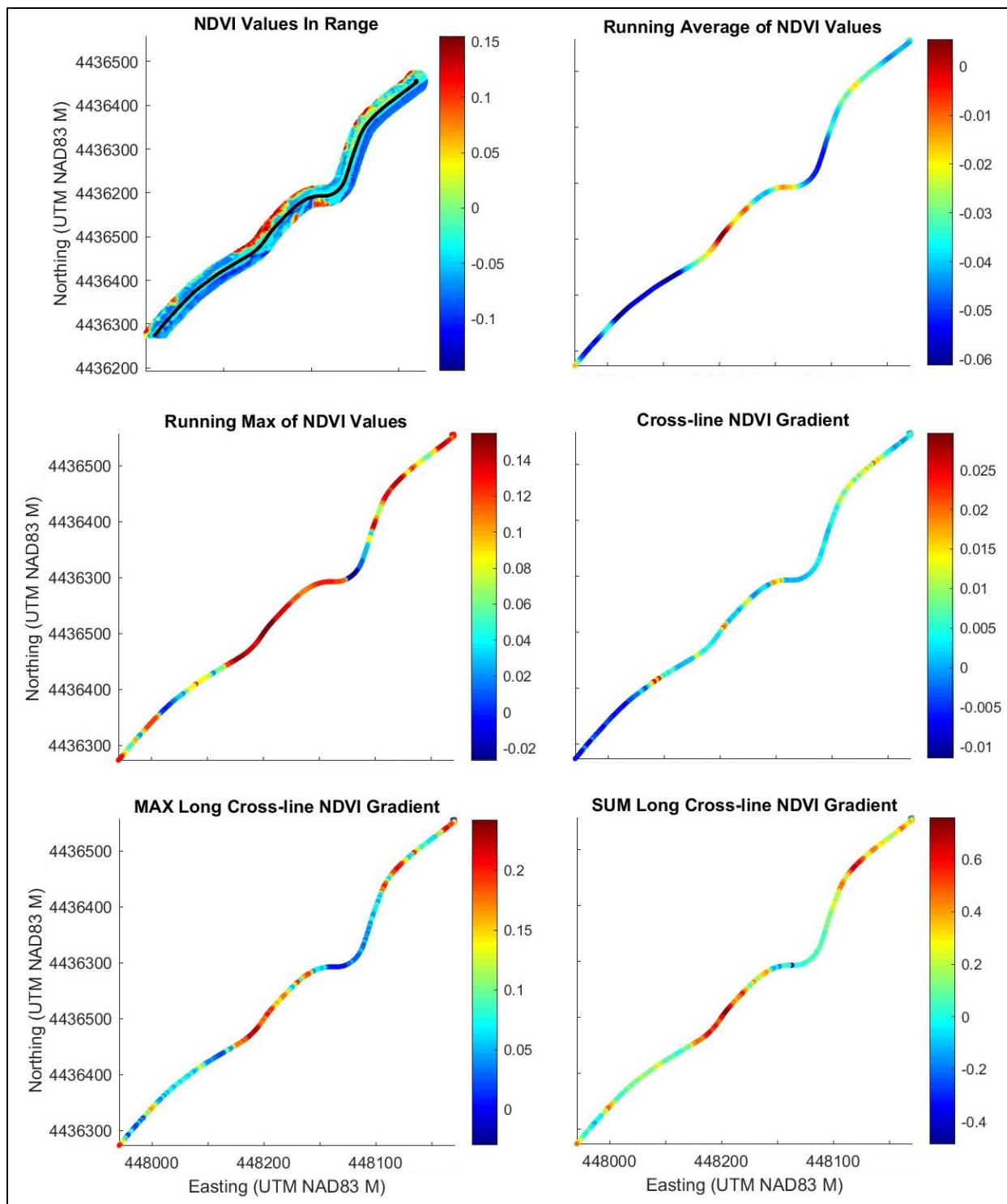


Figure 14. Example plot showing a subset of raw NDVI data (colored scatterplot in upper-left plot) that has been extracted along a 20m wide swath that follows and is centered on a section of the master GPS track (black line in upper-left plot), and the various NDVI data metrics extracted from this subset of NDVI values along the GPS track.

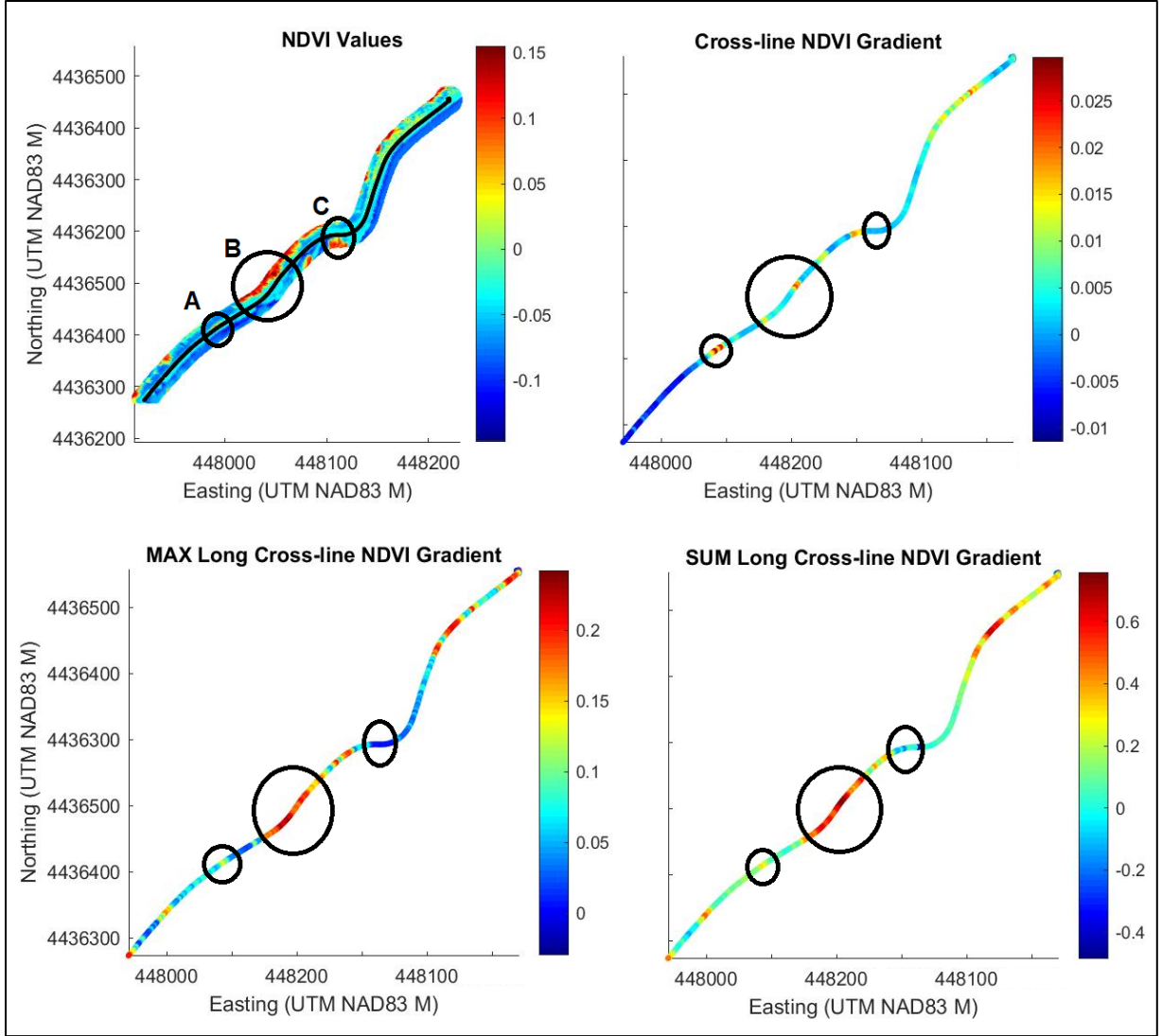


Figure 15. Example plot showing a subset of NDVI data (upper-left plot) extracted along a 20m wide swath that follows the master GPS track, and the various NDVI data metrics extracted from this subset of NDVI values along the GPS track.

With these NDVI data metrics (e.g., spatial gradients) calculated and saved for each point along the master GPS track, we can expand the calculation of the Seepage Index as follows:

$$SeepageIndex_3 = SeepageIndex_2 \times NDVI_{Max} \times NDVI_{cl}^{grad} \times NDVI_{LongSum}^{grad} \times NDVI_{LongMax}^{grad}$$

Here, $NDVI_{Max}$ is the running maximum of nearby NDVI values, $NDVI_{cl}^{grad}$ is the local cross-line NDVI gradients (immediately adjacent to GPS track), $NDVI_{LongSum}^{grad}$ is the sum of all long-range cross-line NDVI gradients, and $NDVI_{LongMax}^{grad}$ is the maximum long-range cross-line NDVI gradient.

Addition of Time-lapse FDEM for Seasonal Change Detection

For this study, both static and time-lapse data attributes were explored for their usefulness in seepage detection. Specifically, time-lapse FDEM data were collected at Strawberry Highline Canal by acquiring EM34 and EM31 profiling data during both watered-up and dewatered conditions: the first FDEM surveys were conducted during 2018 pre-irrigation season conditions while the canal had been dewatered for several months, and these same FDEM surveys were then repeated late in the 2018 irrigation season, while the canal had been watered-up for several months and presumably allowing for any existing seepage areas to become fully established. This was one of the last additions to calculation of a Seepage Index, and was only performed at Strawberry Highline Canal.

Figure 16 shows line plots of the pre-irrigation season baseline EM34 conductivity dataset (blue line) and the watered-up time-lapse EM34 conductivity dataset (red line) for the entire length of the master GPS track (approximately 45,000 data points). Similarly, Figure 17 presents an X-Y scatterplot of these two datasets zoomed in to better show a subset of the data coverage. Here, a color-scale is applied to the terrain conductivity values, and the time-lapse (watered-up) data are shifted to the south by 100m to prevent overlap and reveal time-lapse changes.

Both figures contain black circles that indicate the locations of noticeable (e.g., spatially coherent and significant amplitude) time-lapse changes. These changes are typically seen as increases in measured conductivity during watered-up conditions, suggesting the development of seepage at these locations during the irrigation season. In addition to the additional time-lapse conductivity and in-phase component datasets, the calculated time-lapse differences in apparent terrain conductivity offer another data-type for incorporation into the calculation of a more robust Seepage Index.

Now that we have these time-lapse data sets and measures of change in conductivity, the final version of the Seepage index developed in this study can be expressed as follows:

$$\begin{aligned}
 SeepageIndex = & Q_{em34} \times \frac{1}{\left(\left|\frac{dB}{dx}\right| + \left|\frac{dB}{dz}\right|\right)} \times \frac{Q_{em31}}{IP_{em31}} \times \frac{1}{\left(\left|\frac{dB}{dx}\right| + \left|\frac{dB}{dz}\right|\right)} \times \frac{Q_{em38}}{IP_{em38}} \times \frac{1}{\left(\left|\frac{dB}{dx}\right| + \left|\frac{dB}{dz}\right|\right)} \times NDVI_{ave} \\
 & \times NDVI_{Max} \times NDVI_{cl}^{grad} \times NDVI_{LongSum}^{grad} \times NDVI_{LongMax}^{grad} \times Q_{em34}^{TL} \times \frac{1}{\left(\left|\frac{dB}{dx}\right| + \left|\frac{dB}{dz}\right|\right)} \times \frac{Q_{em31}^{TL}}{IP_{em31}^{TL}} \\
 & \times \frac{1}{\left(\left|\frac{dB}{dx}\right| + \left|\frac{dB}{dz}\right|\right)} \times \frac{Q_{em38}^{TL}}{IP_{em38}^{TL}} \times \frac{1}{\left(\left|\frac{dB}{dx}\right| + \left|\frac{dB}{dz}\right|\right)} \times \nabla Q_{em34} \times \nabla Q_{em31} \times \nabla Q_{em38}
 \end{aligned}$$

Here, Q_{em34}^{TL} and Q_{em31}^{TL} and Q_{em38}^{TL} and IP_{em34}^{TL} and IP_{em31}^{TL} and IP_{em38}^{TL} are the second time-lapse EM34 and EM31 and EM38 conductivity and in-phase component FDEM datasets, respectively, and ∇Q_{em34} and ∇Q_{em31} and ∇Q_{em38} are the calculated time-lapse changes in terrain conductivities as measured by the EM34 and EM31 and EM38 FDEM systems (if available). Note that each contribution of conductivity is being proportionally scaled by repeatedly dividing by the total magnetic gradient. This helps compensate for the effects of metallic structures on the data, suppressing large conductivity spikes added from each FDEM survey data-type. The code developed for this study was designed to make it trivial to add or remove any of these data-types in calculating the Seepage Index, and it should be relatively trivial to include additional data-types or metrics in future efforts to further improve and test this seepage detection algorithm.

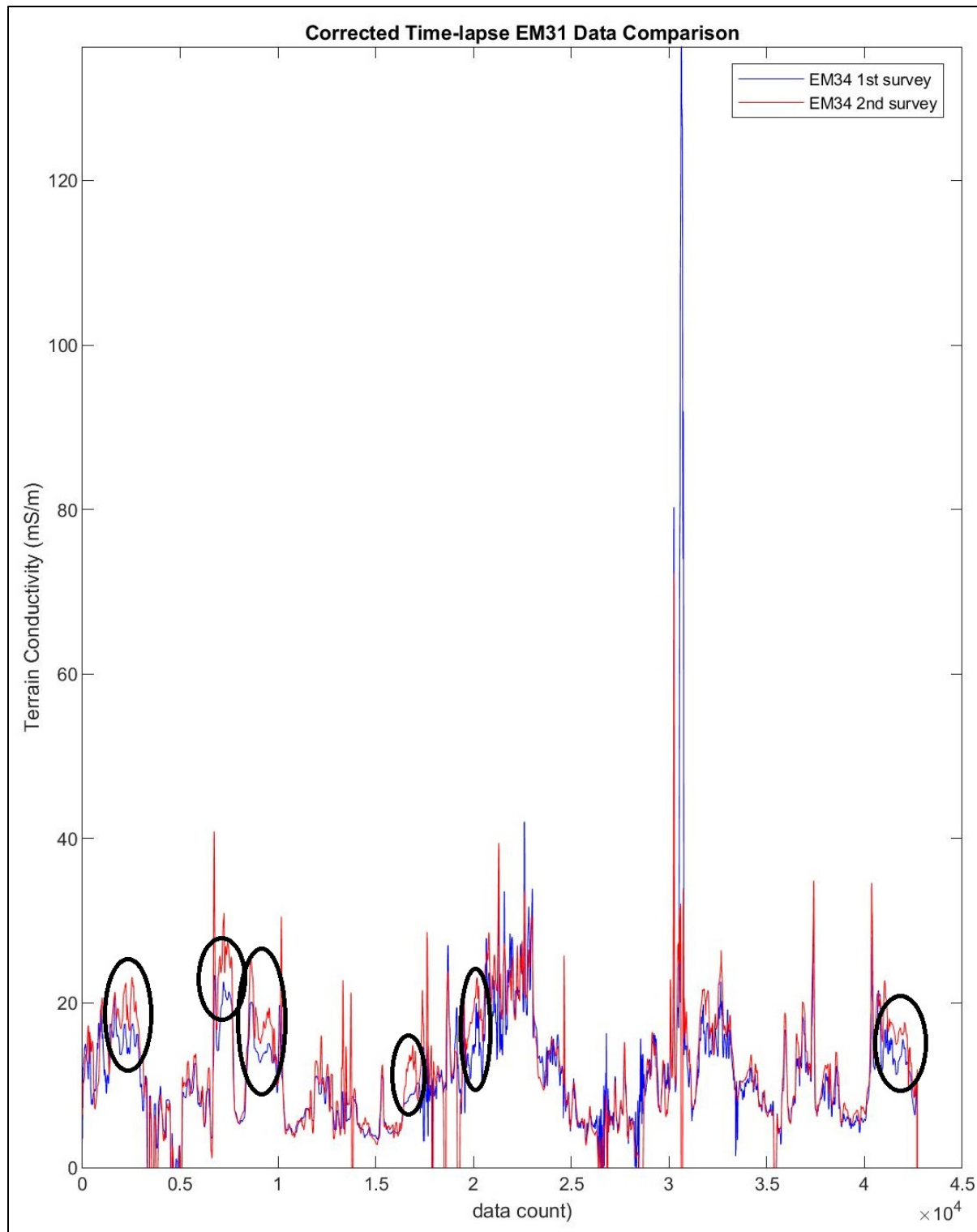


Figure 16. Example line graphs showing a comparison of time-lapse EM34 terrain conductivity data (mS/m) collected along Strawberry Highline Canal. Here, black circles identify segments where noticeable increases in conductivity were recorded during the second survey performed during watered-up conditions (red line).

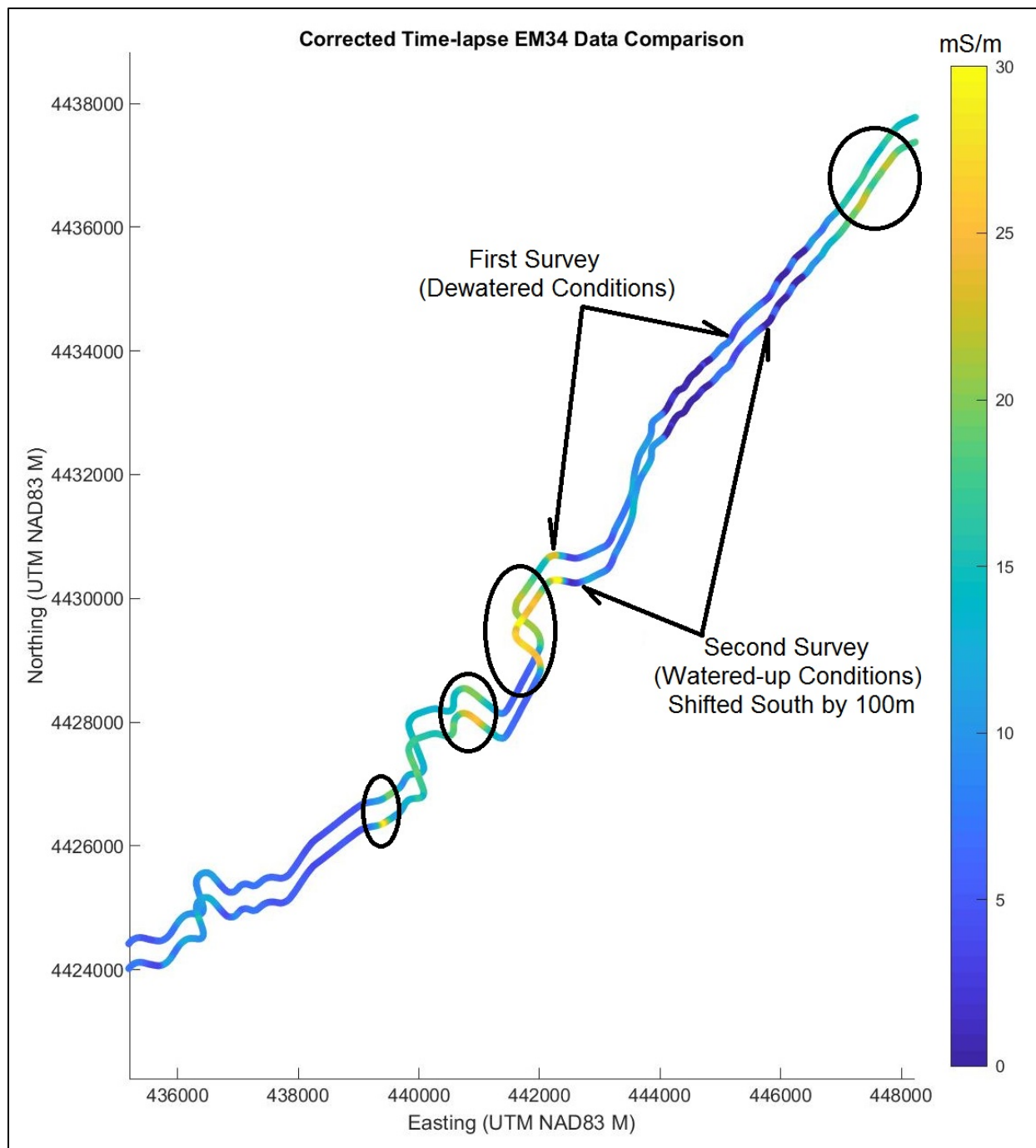


Figure 17. Example X-Y scatterplot showing a comparison of time-lapse EM34 terrain conductivity data (mS/m) collected along a portion of Strawberry Highline Canal. Here, black circles identify segments where noticeable increases in conductivity were recorded during the second survey performed during watered-up conditions (data shifted to the south by 100m).

Interpreting the Seepage Index: Automatic Seepage Detection

Now that all data types have been calculated, interpolated to the master GPS track, and a Seepage Index values calculated along the GPS track, we could simply plot and visually interpret these values for potential seepage areas. In general, these values can be thought of as a probability of seepage along the GPS track. However, this study takes on the challenge of automatically detecting the most likely seepage areas/features to help attempt automation of the entire analysis workflow. Two approaches were taken in this study. The first approach involves thresholding of Seepage Index values to detect 1) “most-likely seepage,” 2) “likely seepage,” and 3) “potential seepage” areas. The second approach involves the initial testing of machine learning and clustering to detect seepage areas, using all data-types and Seepage Index values as input data-layers (i.e., multi-dimensional data).

Detrending and Thresholding

Figure 18 shows a series of line-plots of the various input geophysical data-types and NDVI data metrics (top cluster of line-plots) from South Canal’s master GPS track that were then used as inputs for calculation of corresponding Seepage Index values (bottom green line-plot). Here, the resulting “raw” Seepage Index values show both sharp spikes (possibly indicating more focused seepage areas) and longer trends (possibly indicating more distributed seepage areas) along the GPS track.

The underlying challenge is to use simple thresholding to extract the locations of spikes likely associated with focused seepage areas or possibly even source points for the larger trends observed in the seepage index values. To accomplish this, the Seepage Index values must first be detrended (i.e., high-pass filtered) and then renormalized from 0 to +1. This process is depicted in Figure 19, where the following is performed:

1. The trend of the Seepage Index values is calculated (black-dashed line on top plot of Figure 19), and this trend is then subtracted from the Seepage Index values.
2. The mean of detrended values is then subtracted to shift the detrended values to zero (middle plot of Figure 19).
3. These detrended and shifted values are then clipped at zero to remove negative values and then renormalized between 0 and +1 (bottom plot of Figure 19).

Finally, user-define thresholds can be applied to the resulting values to identify the locations of potential focused seepage. In the example of South Canal shown in Figure 19, threshold values of 0.5, 0.2, and 0.1 were utilized for detecting “most-likely seepage,” “likely seepage,” and “potential seepage” areas, respectively (horizontal dashed-lines on bottom plot of Figure 19). Again, this approach of detrending and renormalizing the Seepage Index values helps identify locations of possible “concentrated” seeps or focused sources for larger saturated/standing water areas adjacent to canal that only NDVI or visual inspections may not otherwise be able to detect or isolate.

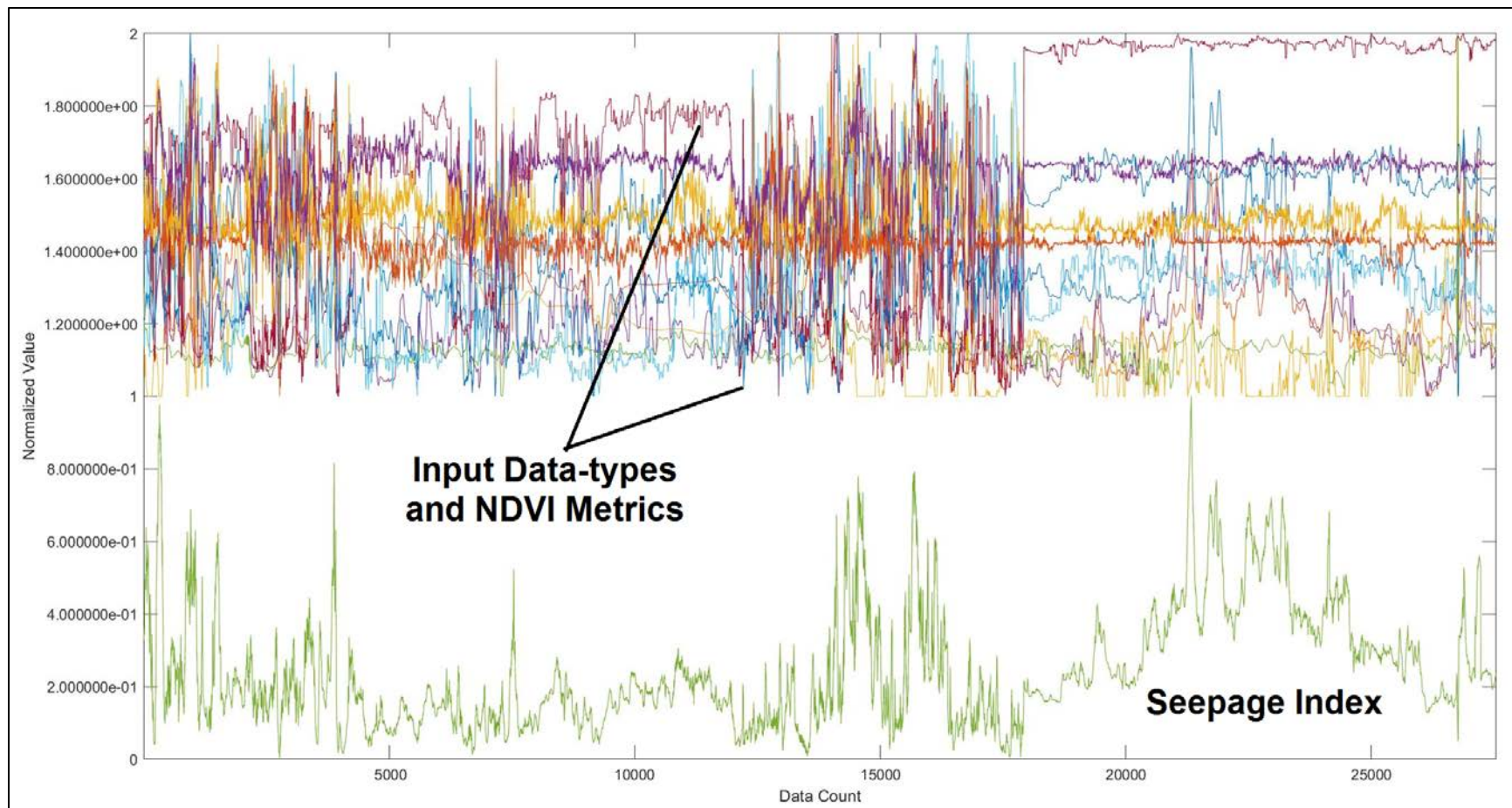


Figure 18. Example of the various input geophysical data-types and NDVI data metrics (top cluster of line-plots) from South Canal used as input for calculation of the resulting Seepage Index (bottom green line-plot).

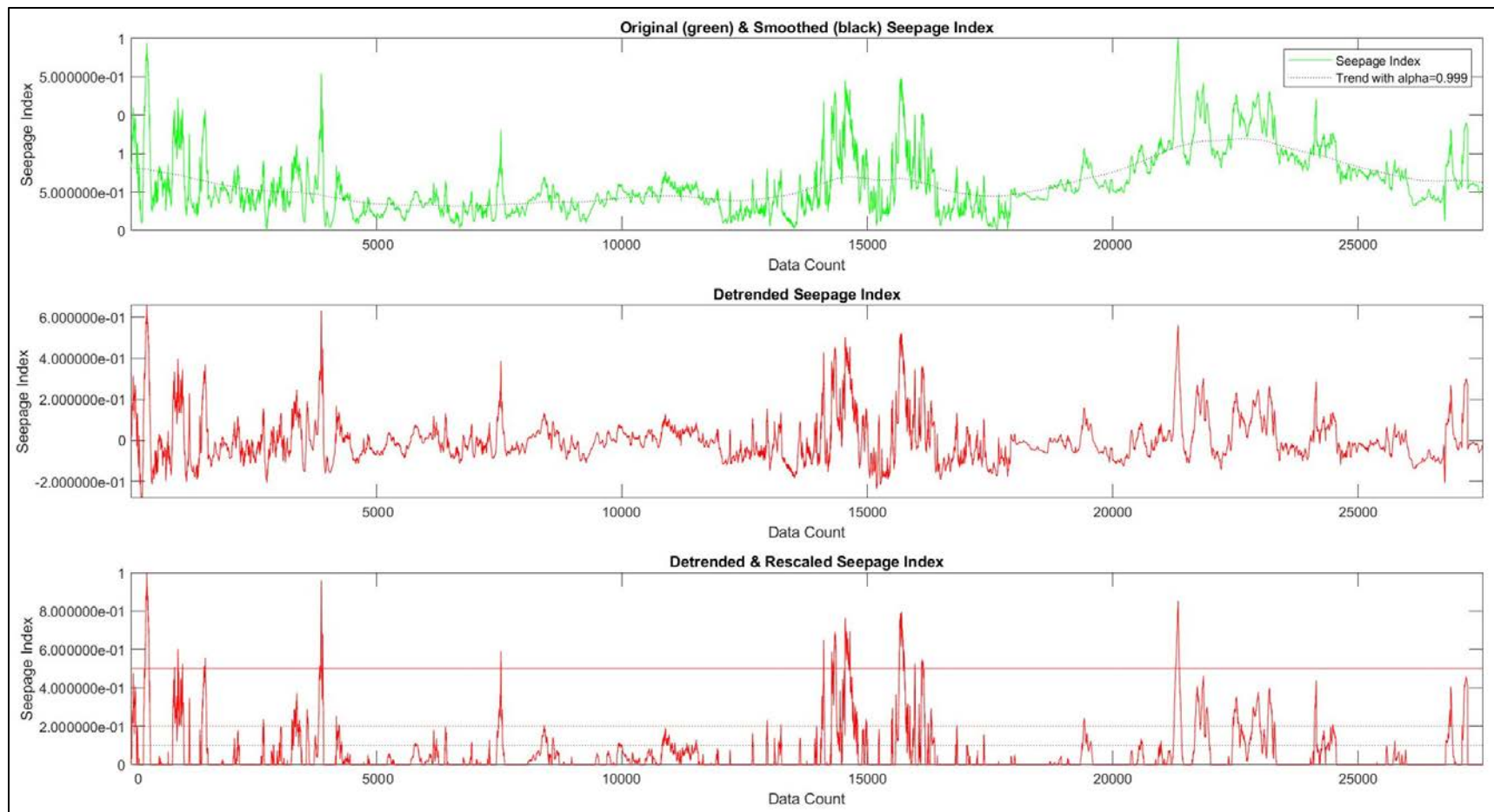


Figure 19. Example from South Canal Depicting of the process of detrending clipping and renormalization of the Seepage Index for sake of applying thresholding for detection of most likely seepage areas.

Machine learning: Clustering Analysis and Binary Classification

In machine learning and statistics, trained classification is the problem of identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations whose category membership is known. This is the fundamental challenge of clustering analysis, which is a subset of various data mining techniques that are implemented to separate data into unique categories, where each data point is most similar to others in the same cluster and most different from points in other clusters. In untrained classification, various algorithms are used to cluster data in an automated fashion, without the use of a training set.

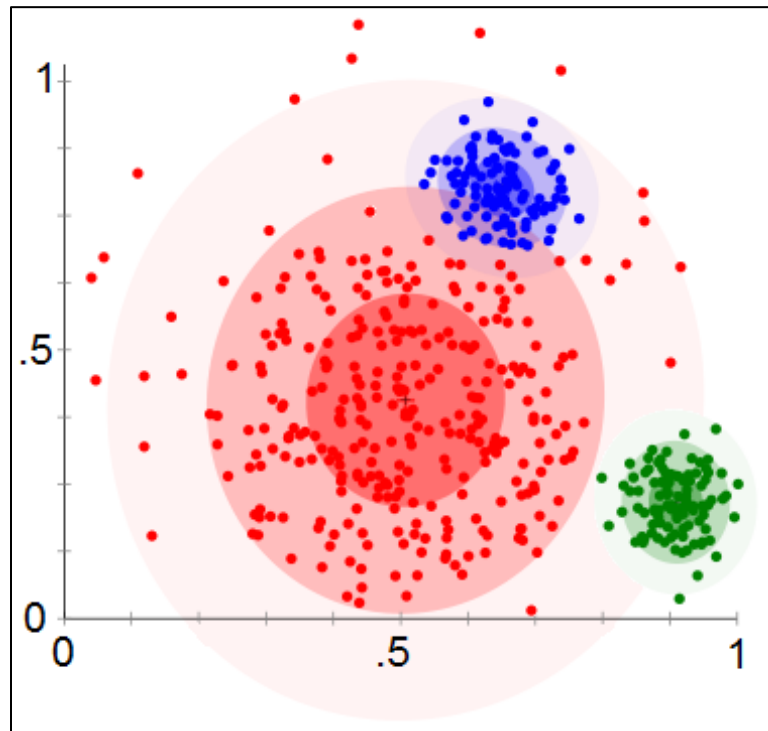


Figure 20. An example scatterplot showing the results of clustering applied to define three categories of data (depicted in this example with different colors). Image modified from (Wikipedia contributors, 2018a).

As depicted in Figure 20, this concept is most easily thought of as a 2D cross-plot of two data-types with various clusters of data that capture some set of common relationships found between the two data-types (three common relationships in this example). Here, a variety of clustering algorithms can be used to identify which cluster each data point most likely belongs to, based on some measure of similarity and difference from other points. In this 2D scatterplot example, a Euclidean measure of distance between each pair of data points is the most apparent choice. However, various machine learning and clustering/classification algorithms can be implemented using non-Euclidean measures of distance (i.e., similarity) between data points that are multi-dimensional (e.g., data points that contain multiple channels or layers of data values).

The latter is the case in this study, where we are dealing with multiple types of geospatial data (e.g., ground-based geophysical data sets and NDVI data images) all collected or calculated along the same path of a canal embankment. For this study, we performed preliminary testing of a K-means clustering algorithm, setting a predefined number of clusters to search for equal to four. Four clusters was selected, in order to attempt to capture the following conditions along a given canal:

1. “Most likely” seepage
2. “Likely” seepage
3. “Possible” seepage
4. No seepage

Support vector machines (SVMs) are most typically implemented as supervised (i.e., trained) machine learning algorithms that analyze data for classification and regression analysis. SVMs are referred to as “non-probabilistic binary linear classifiers,” (Wikipedia contributors, 2018b). Here, binary classification means that SVMs split data into only two groups or classifications: one subset of the input data that match a given pattern of interest, and the remaining data that do not fit the pattern or classification of interest. More specifically, as shown in Figure 21, SVMs seek to find a “hyperplane” (a line in the case of 2D data, a 3D surface in the case of 3D datasets, and a multidimensional surface in the case of multidimensional data) that provide the greatest separation between the two identified subsets.

SVMs are typically implemented as trained classification algorithms, where a set of “training data” from a given known category of interest is provided to the SVM. This training set of data is then randomly sampled in order to split it into a training subset and a validation subset. An iterative training algorithm then builds a hyperplane model by using the training subset to build trial models, and the validation subset is used to evaluate the quality of the trial models. Once the model converges, it is then used to classify new data (new data that are not of a known classification).

For this study, preliminary testing of SVM’s ability to identify locations of likely seepage versus non-seepage was performed. Unfortunately, not a significant number specific locations of known seeps were available along the canals tested. This information would ideally have served as the training dataset (e.g., the various data metrics and Seepage Index values at these locations would be used as a multidimensional training set). Instead, the results of the Seepage Index thresholding process discussed above were used as training sets, and SVM was implemented to predict where other threshold exceedances would be encountered.

For example, a small subset (only the first 10%) of the identified “most likely seepage” areas (based on the above thresholding approach) were extracted for use in training an SVM model. This extracted set of training data was then divided again, such that 85% were used to train the SVM model and the other 15% were retained as the validation subset. The SVM model was used to predict the other locations of “most likely seepage” along the entire length of the canal surveyed. This approach was also implemented for predicting areas of “likely seepage” and “potential seepage,” as identified from the thresholding approach described above.

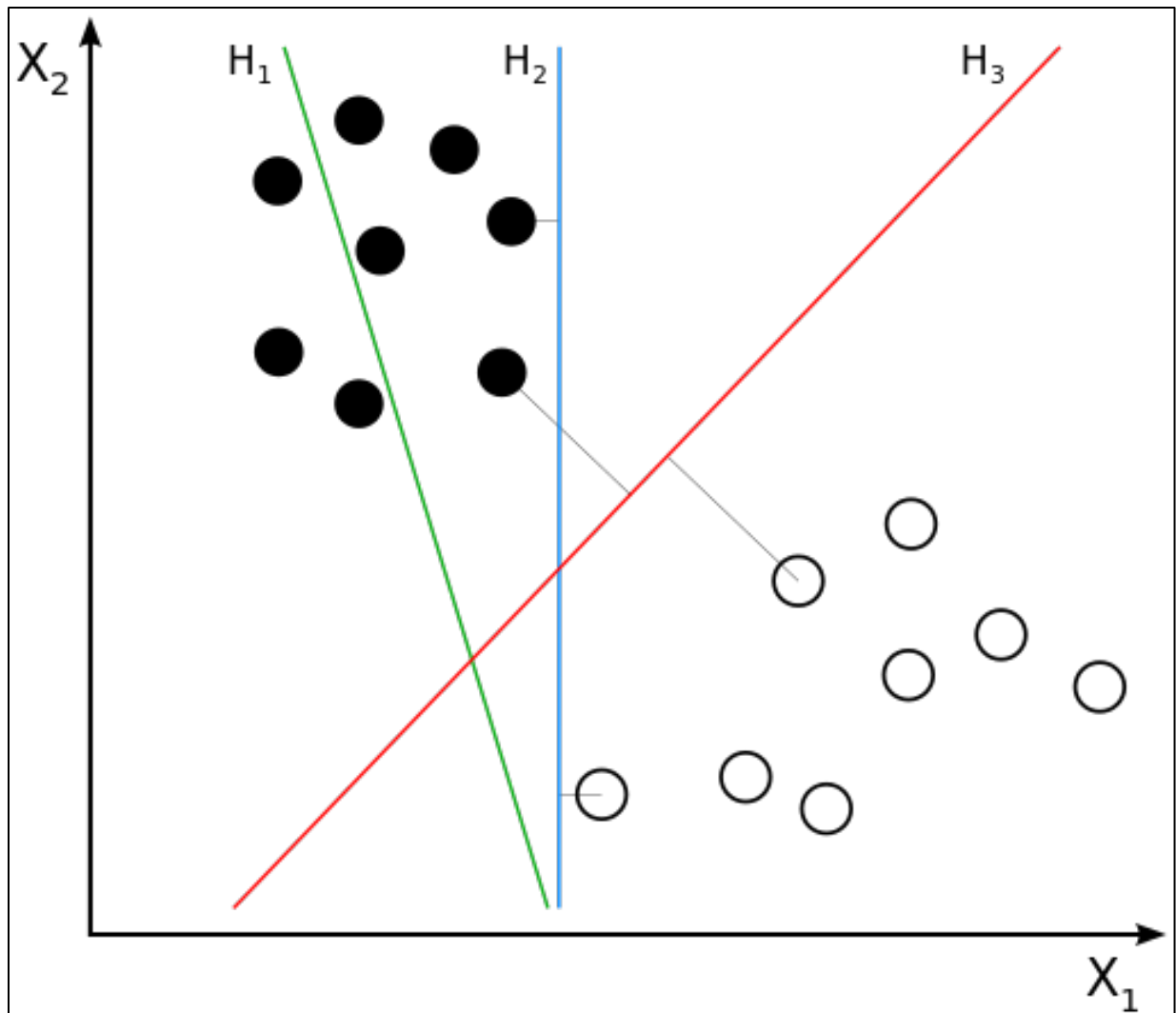


Figure 21. Graphic showing how a support vector machine would choose a separating hyperplane for two classes of points in 2D. H_1 does not separate the classes. H_2 does, but only with a small margin. H_3 separates them with the maximum margin. Image taken from (Wikipedia contributors, 2018b).

Results

The results from this study are presented in a series of subsections below, starting with the DFC test site, and then sections for each of the three test canals (South, Government Highline, and Strawberry Highline Canals). These subsections each present only a select subset of all results for the sake of discussion, mainly due to the vast length of data collected along the test canals, and the details captured in the results. Additional electronic deliverables, including Google Earth™ .KMZ files that contain data layers, calculated Seepage Index values, predicted seepage locations, and .CSV files of corresponding values are also available, as describe in more detail at the end of this report.

We encourage the reader to access and explore the additionally available Google Earth™ .KMZ files in order to fully peruse, understand and potentially utilize the results of this research.

Fed Center Test Site Activities for System Evaluations

Various experiments were conducted using the EM31, EM34, EM38, and the EMP-400 FDEM systems, as well as the non-magnetic cart configurations, as described above. These tests included:

1. 2D mapping of the test site using each FDEM system for comparison of results, and to determine system sensitivities to various utilities observed to be running through the test site area.
2. Testing for azimuth-dependent FDEM variations (generally not observed)
3. Testing inter-system interference while collecting data simultaneously (generally not observed)
4. System sensitivity to towing vehicle offset and coil orientation (transmitter should be the far-coil to avoid coupling and saturation)
5. Comparison of FDEM mapping results to 2D resistivity tomography models produced along a “test-line” in the middle of the FDEM test area (see Figure 22 and Figure 23).

Part of this latter effort involved efforts to calculate depth-weighted averages of ERT resistivity model values for comparison with FDEM terrain conductivities by using the horizontal dipole sensitivity curve in Figure 7 and fitting of an exponential curve to calculate depth weights. The comparisons of the ERT and FDEM data are best depicted as shown in Figure 23. Here, various similar patterns in the tomogram can be seen in the EM31 terrain conductivity map at the same locations and within the nominally expected 0-15ft depth of investigation for an EM31 system (right side of Figure 23). Various lineation patterns can be seen in the EM31 map that show the locations of buried electrical, drainage, and water-main utilities that cross the test site. Similarly, Figure 24 shows the mapping results using an EMP400 system at three different frequencies, starting with the lowest available setting (corresponding to the greatest DOI). While the overall patters of the EMP-400 were agreeable with the other systems, it appeared to have less DOI and sensitivity to lateral changes in terrain conductivity, and so the EM38 was chosen as an alternative system for similar DOI.

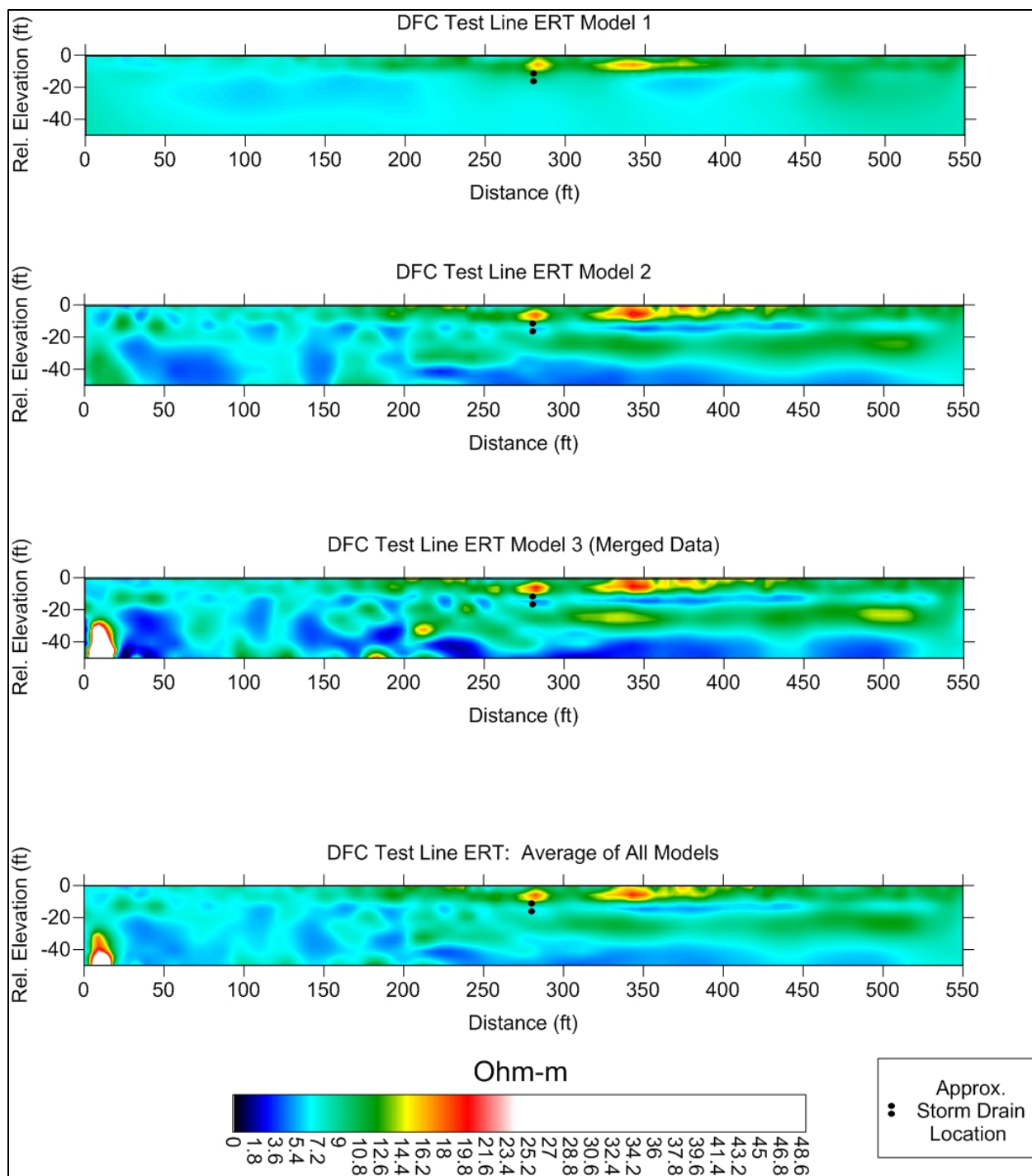


Figure 22. Three ERT models for different data collection approaches (different electrode dipole configurations), and the resulting averaged “final” electrical resistivity model used for comparison to FDEM mapping results.

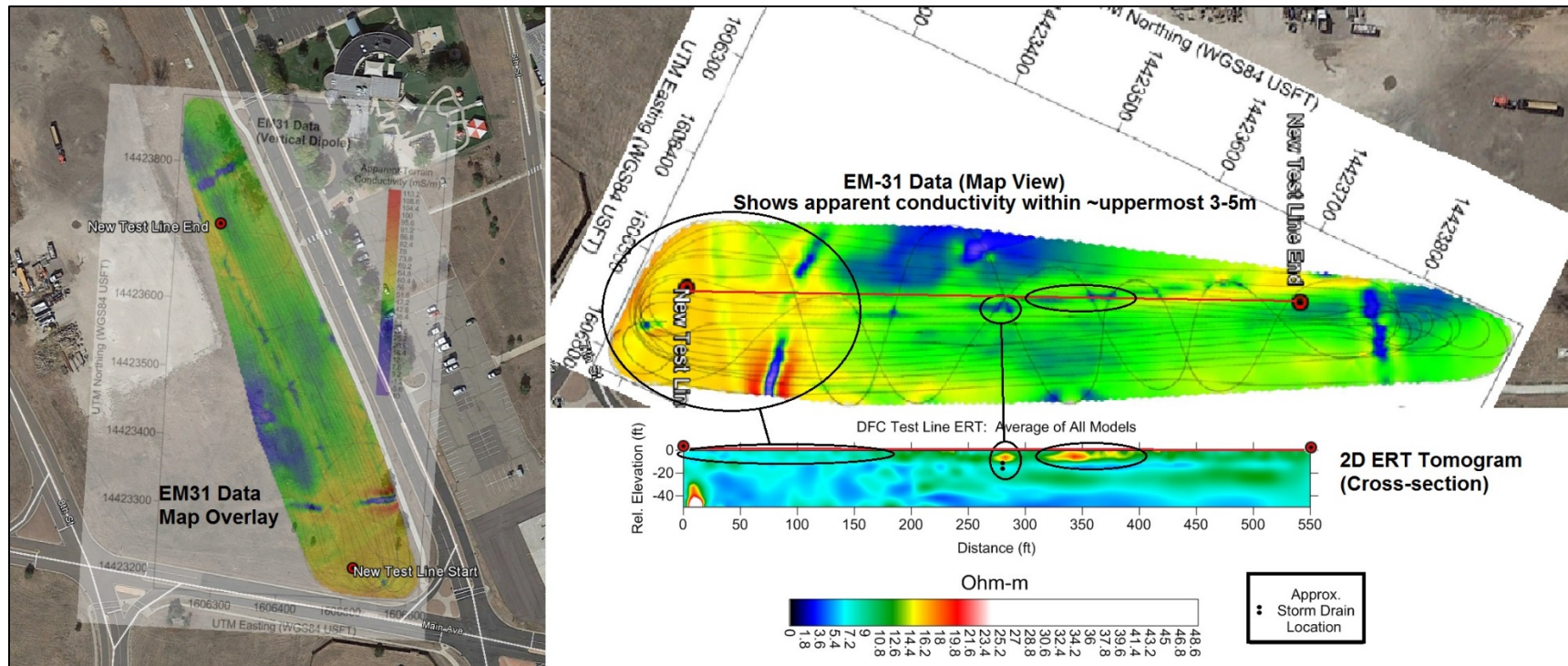


Figure 23. A map of EM31 terrain conductivity data superimposed on aerial imagery (left image), and a comparison between EM31 conductivity values and ERT modeled resistivity values collected along a “test-line” established within the DFC test-site (right image). The test-line was utilized to inspect system performance and sensitivity to subsurface electrical conductivity structure imaged by the ERT surveys.

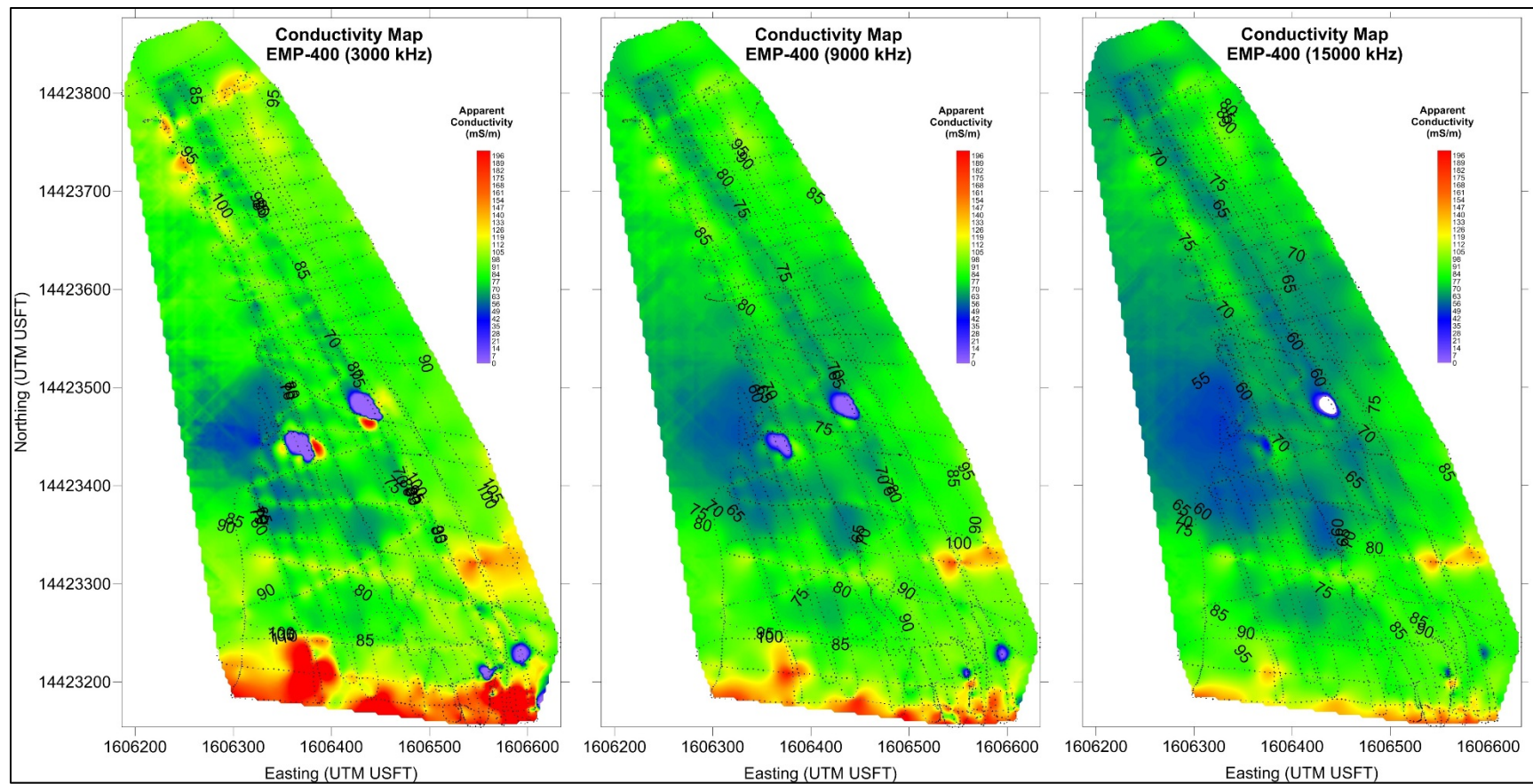


Figure 24. Plots of three different frequencies of FDEM data collected across the DFC test site with an EMP-400 system.

South Canal

Results from testing along South Canal are presented below in Figure 25 through Figure 34. Generally, results of Seepage Index calculations and seepage predictions (using the thresholding approach described above) are corroborated nicely by field observations of wet areas, lush wetland vegetation such as cattails and reeds, as well as a few known seepage locations from field observations provided by partnering Reclamation Area Offices. Generally, there are several somewhat discrete locations of seepage predicted along the northern-half of the canal segment surveyed for this study. The lower third of this canal segment has several extensive observed wet areas and lush vegetation adjacent to the embankment, and corresponding elevated Seepage Index values and predicted seeps (see Figure 25 through Figure 27).

Detailed examples of successful seepage predictions are provided in the detailed views shown in Figure 28 through Figure 30 below. These figures were generated from screen-shots of Google Earth™.KMZ overlays with aerial imagery. Here, Figure 28 shows an almost perfect prediction of a known seepage location provided by the Western Colorado Area Office (also observed during FDEM data collection). Figure 29 and Figure 30 show other examples of successful seepage predictions in areas where likely seepage was observed during FDEM data collection. In some cases, false positives are observed at the locations of infrastructure (magnetic gradiometry was not performed here), but most elevated Seepage Index values and corresponding seepage predictions tend to not be associated with these features.

Finally, Figure 31 and Figure 32 present the results of applying the rather simple K-means clustering algorithm in an attempt to differentiate between seepage and non-seepage areas of the canal. Similarly, Figure 33 and Figure 34 present results of applying SVMs for prediction of seepage locations (based on Seepage Index thresholding). In general, K-means clustering was deemed somewhat successful, and thus clustering approaches are deemed promising for automated seepage detection (e.g., trained or untrained machine learning). Further research in the application of more advanced clustering algorithms is deemed worthy for this application, based on the results presented herein.

SVM classification results are observed to be extremely promising in identifying data patterns related to seepage, based on the results presented here. SVM results were observed to be almost identical regardless of the use of normalized or original “raw” data-types and various spatial attributes as inputs. The metrics calculated for correct predictions versus false positives and false negatives is quite compelling for this type of machine learning classification technique applied to seepage detection. The incorporation of additional data types should only further improve Seepage Index calculations and seepage prediction results.

While results of all three classification prediction sets reflect a good match, the SVM predictions are observed to be most accurate for the “most likely seepage” classification, and least accurate for the “possible seepage” classification, as expected. This result may change slightly, depending on which training data are chosen. Further testing will be required to evaluate SVM classification as a reliable approach to predicting seepage using this approach and additional known seepage location data for use as training sets.

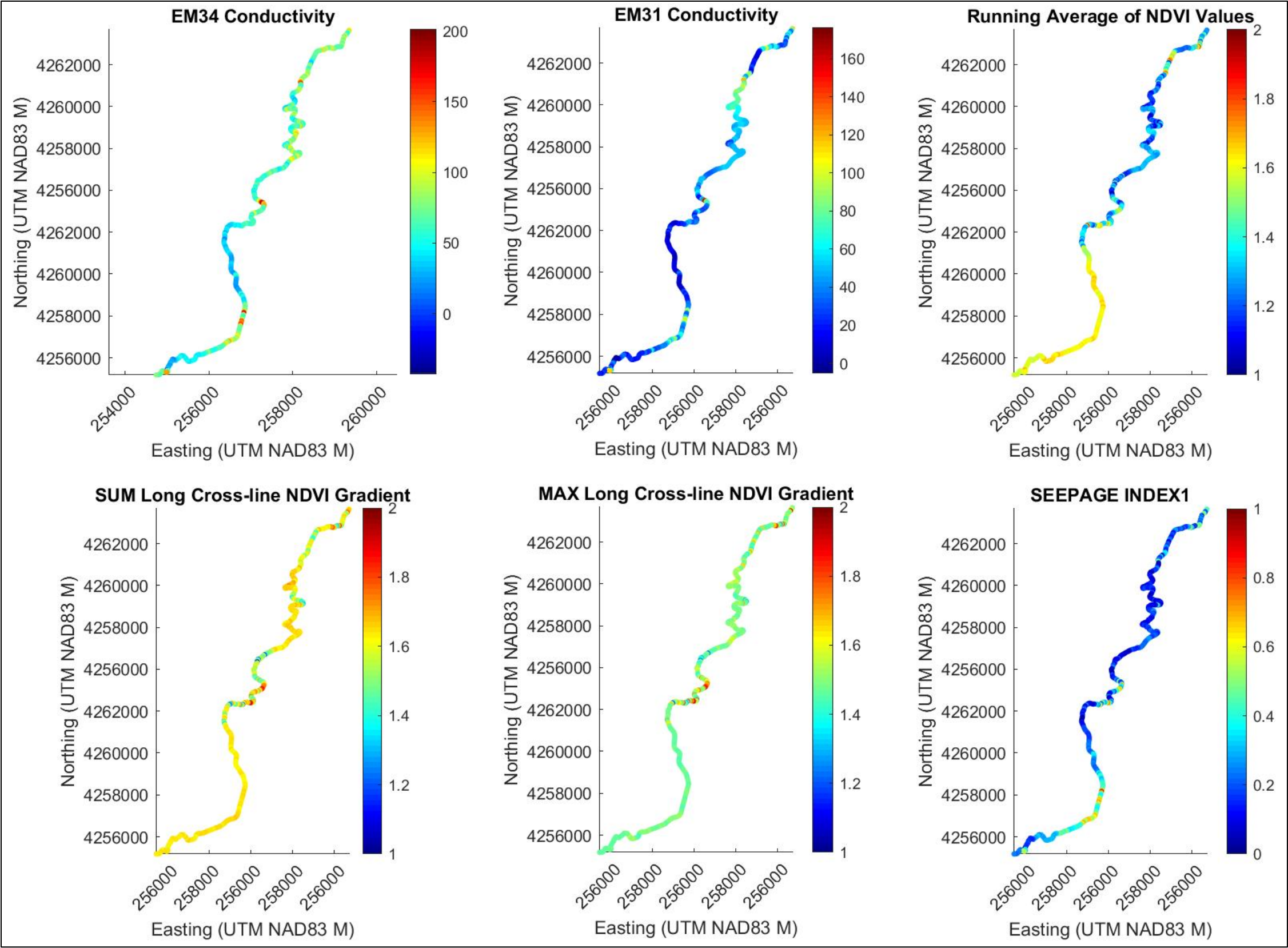


Figure 25. Example from South Canal showing some of the data-types and resulting calculated “raw” Seepage Index values along the canal section surveyed for this study.

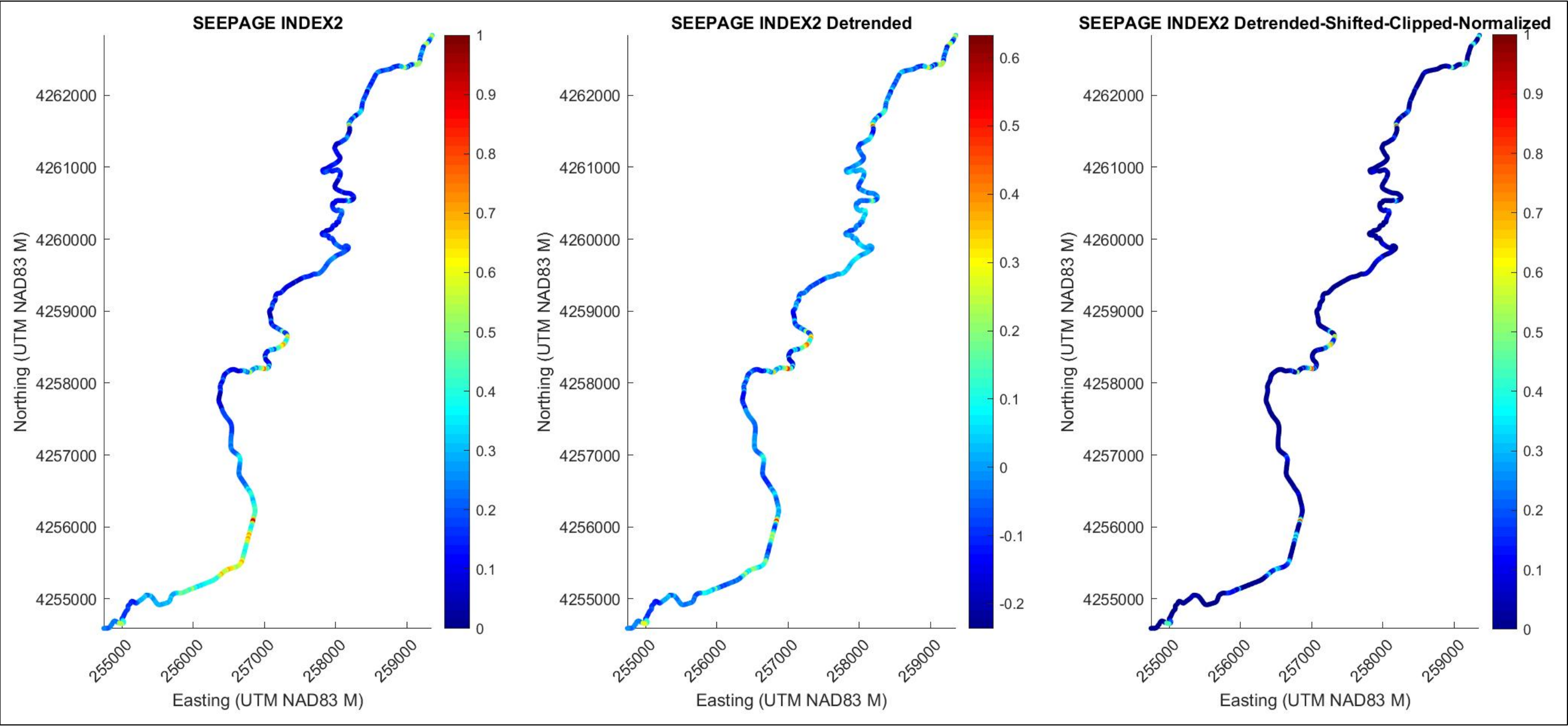


Figure 26. Example from South Canal results showing the calculated “raw” Seepage Index (left plot), the detrended Seepage Index (center plot), and the detrended-shifted-clipped-renormalized Seepage Index values (right plot).

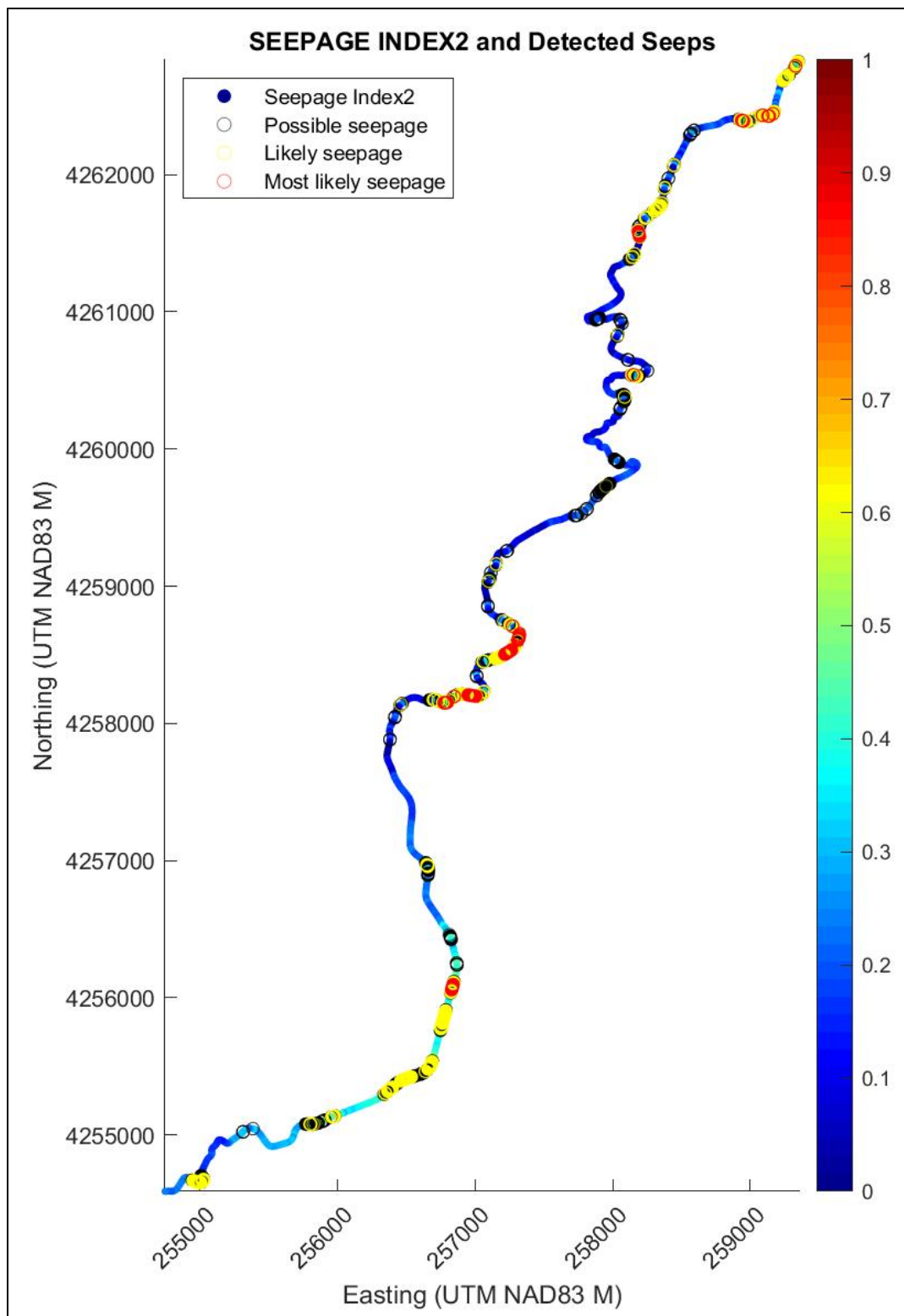


Figure 27. Detected seepage locations along South Canal using thresholding of the detrended Seepage Index values. Here, red circles are “most likely” seepage, yellow circles are “likely” seepage, and black circles are “possible” seepage. The “raw” Seepage Index values are plotted under these predicted locations.

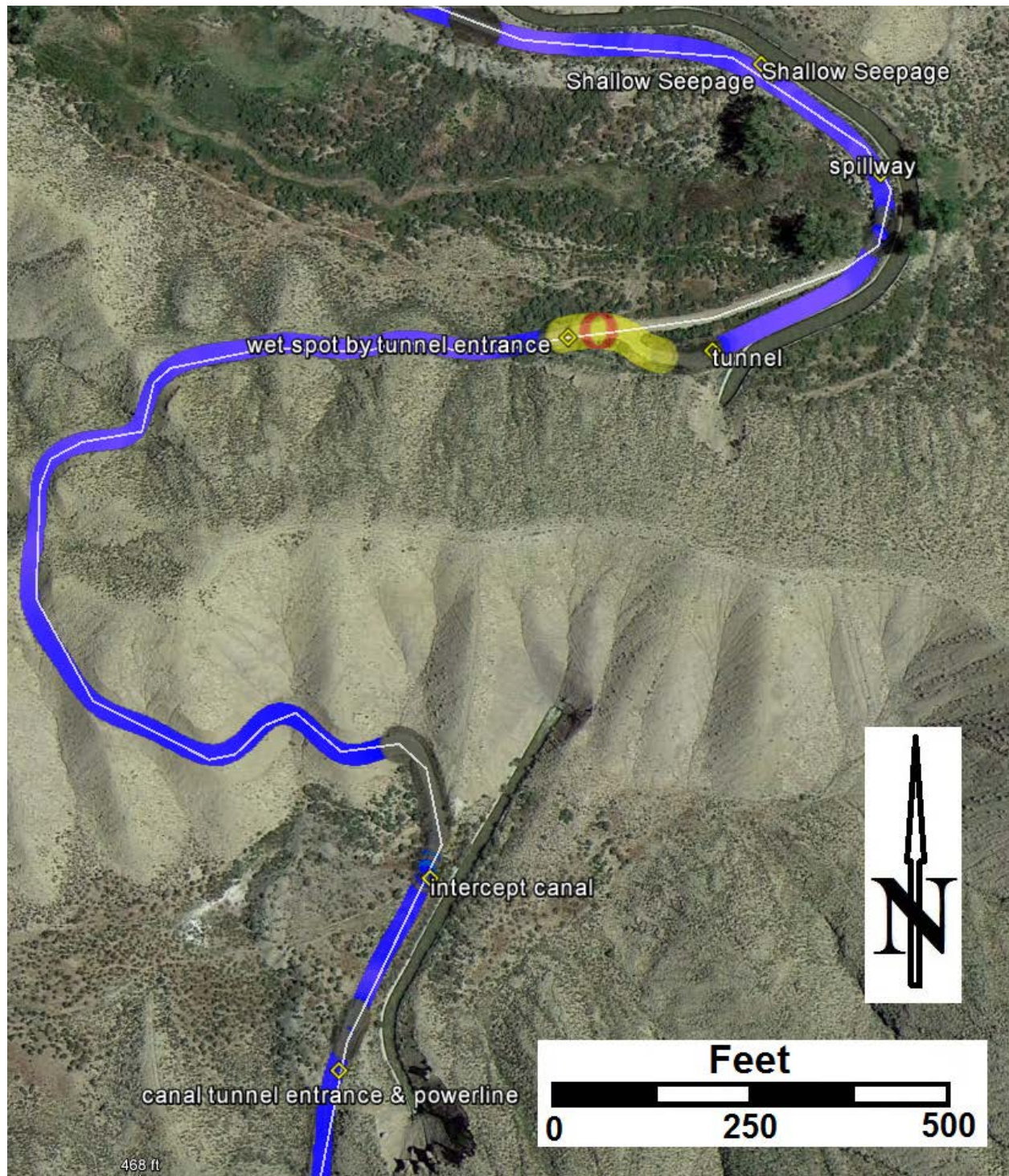


Figure 28. A detailed view showing an example of a successful seepage prediction using thresholding of the Seepage Index. Here, a known seepage location was provided by the Western Colorado Area Office, and the location matches almost perfectly with a predicted location of “most likely seepage” (red circle) and a surrounding band of “likely” seepage (band of adjacent yellow circles). Note that the “shallow seepage” waypoints labeled on the figure just to the north of the wet-spot location was simply surficial leakage through a crack in the canal’s liner, and likely not associated with significant subsurface saturation and corresponding conductivity anomaly.

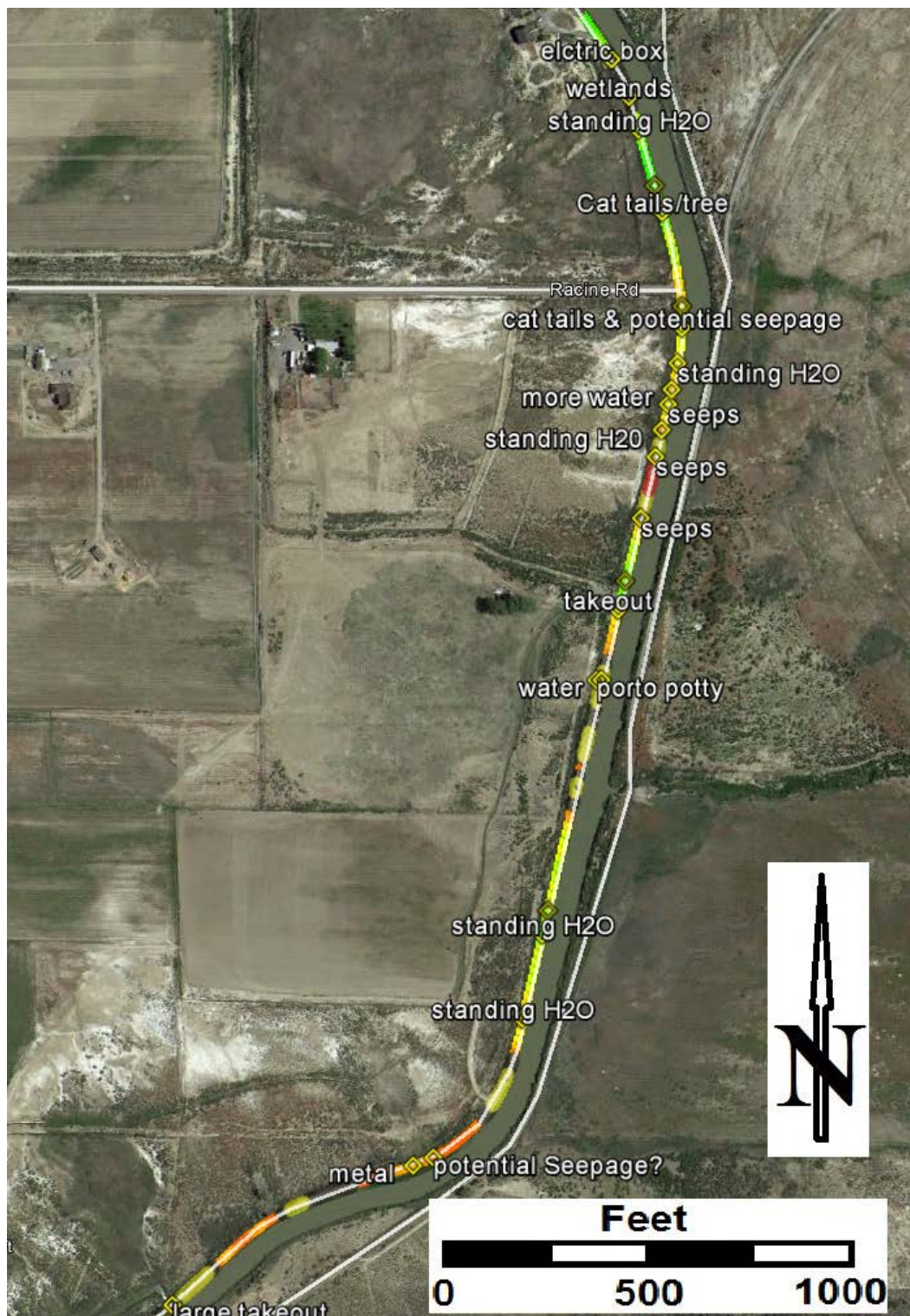


Figure 29. A second detailed view near the southern end of the segment of South Canal tested. Here, various successful seepage detections and generally elevated Seepage Index values are validated with extensive observed apparent seepage (noted and coordinates recorded with a hand-held GPS during data collection).

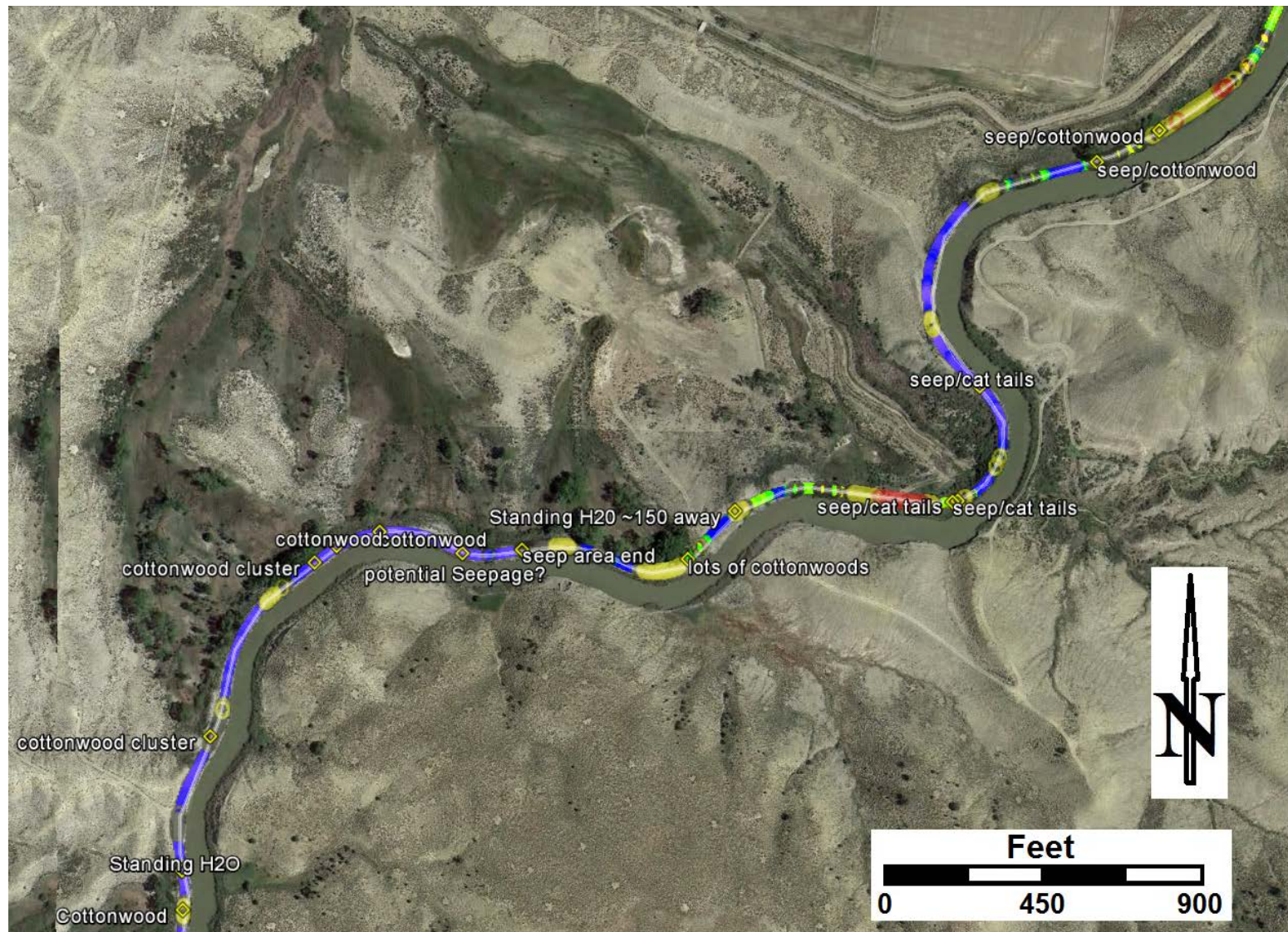


Figure 30. A third detailed view along South Canal showing seemingly successful seepage detections and generally elevated Seepage Index values consistent with visible vegetation patterns.

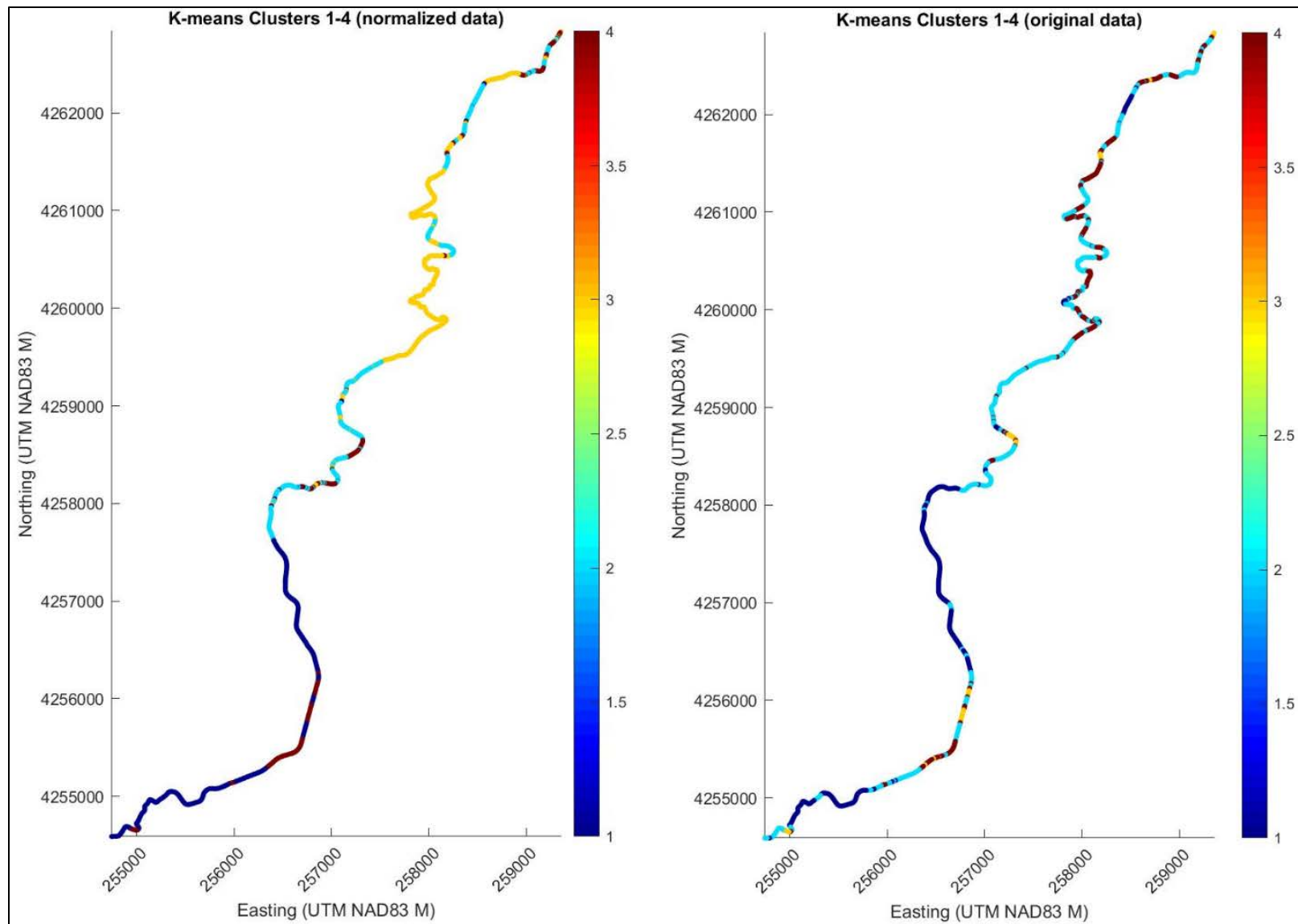


Figure 31. Results of K-means clustering applied to the original and normalized (between 0 and +1) data-types and metrics along South Canal.

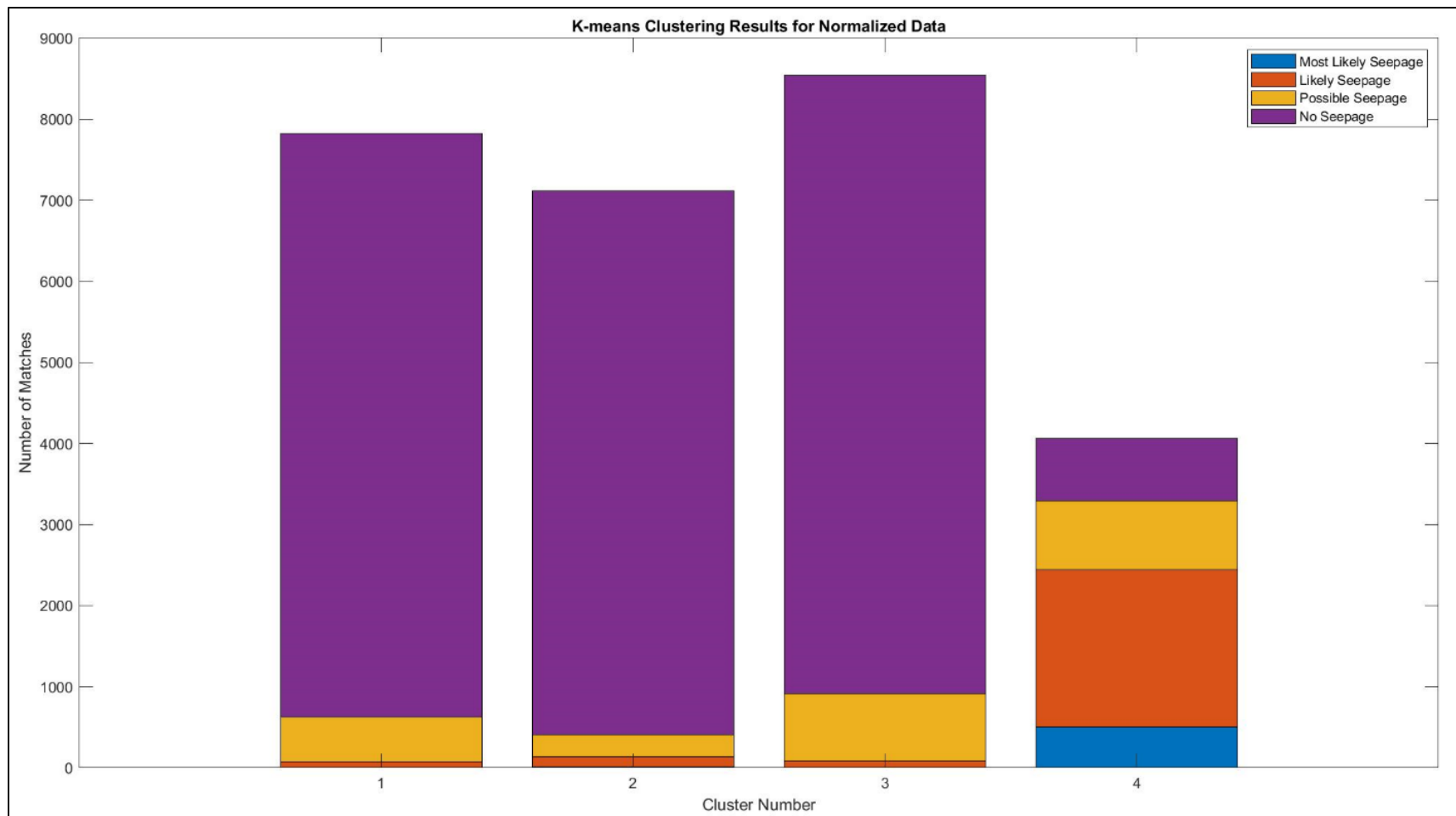


Figure 32. Stacked bar graph showing the results of K-means clustering for seepage prediction. This result is rather promising, where most seeps are identified with cluster number 4, and the majority of the other three clusters are predominantly background “non-seepage” segments (based on Seepage Index thresholding).

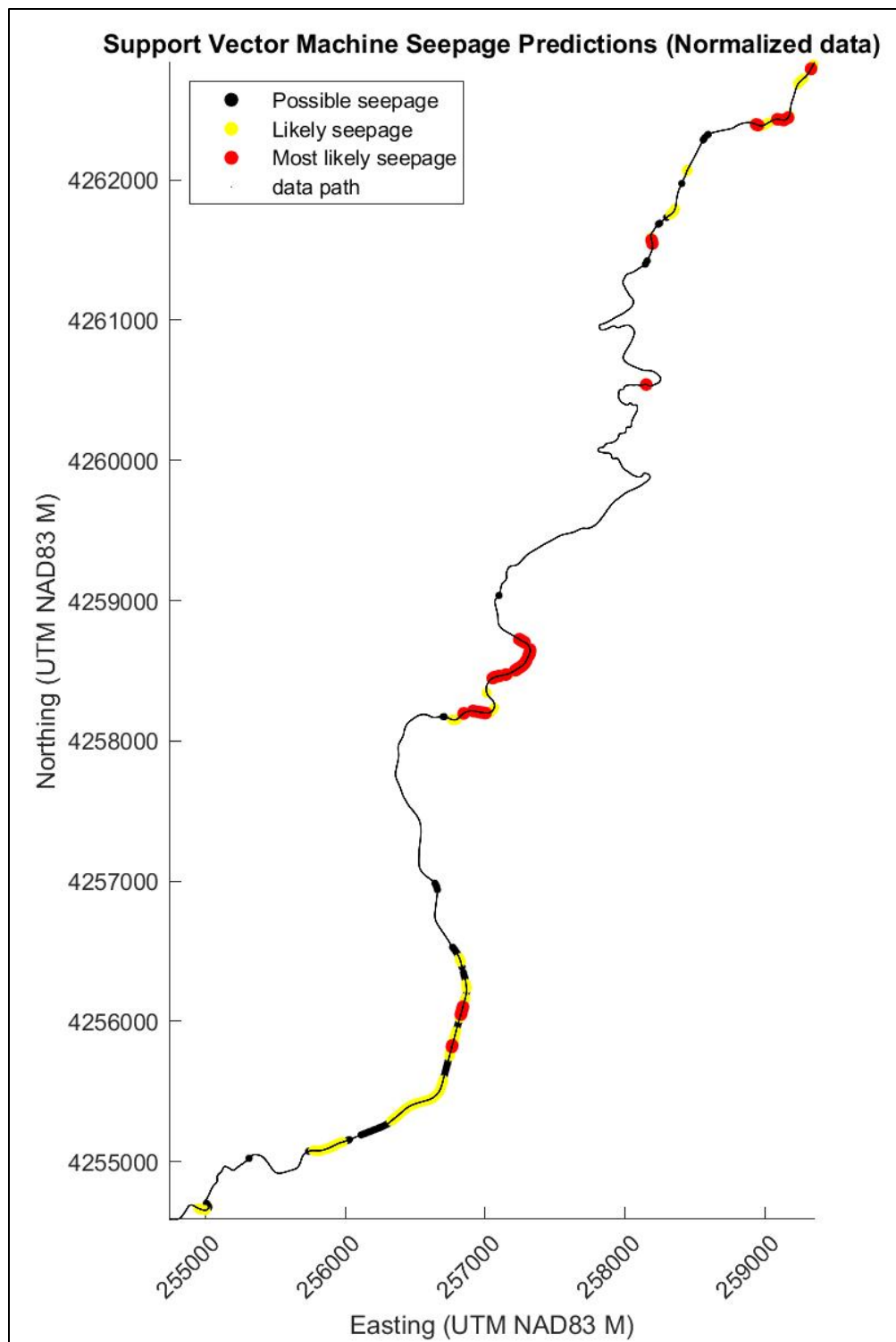


Figure 33. A spatial plot of the results of SVM prediction results.

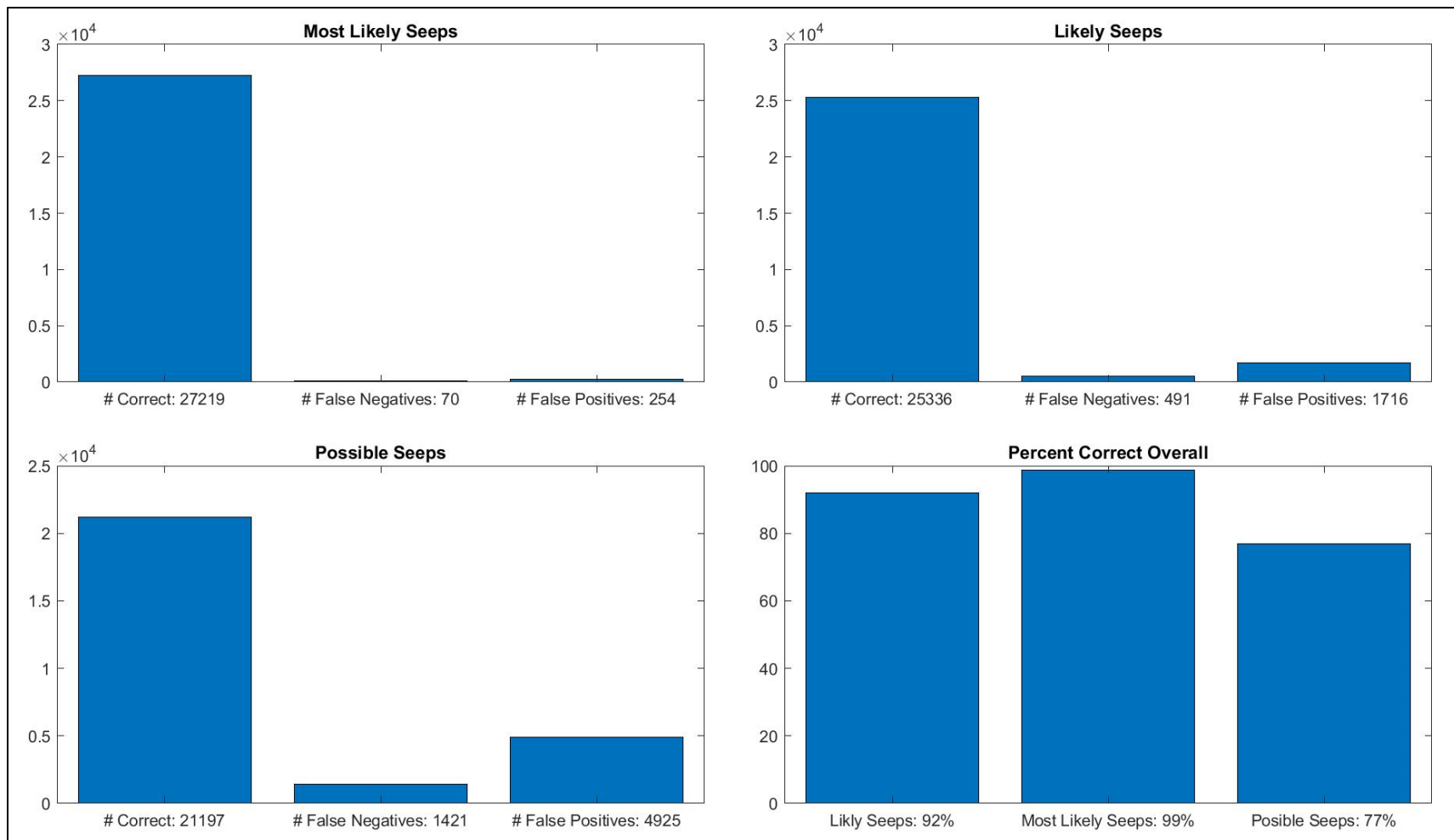


Figure 34. Comparisons and evaluations of SVM prediction results relative to thresholding-based predictions along the South Canal study area. In general, these results show an excellent application of SVMs for seepage prediction, based on the nature of our input multidimensional data set.

Government Highline Canal

Results from testing along Government Highline Canal are presented below in Figure 35 through Figure 42. Generally, results of Seepage Index calculations and seepage predictions (using the thresholding approach described above) are again corroborated nicely by field observations of wet areas, lush wetland vegetation such as cattails and reeds along the canal. Generally, there are several somewhat discrete locations of seepage predicted along the length of canal segment surveyed for this study (see Figure 35).

Detailed examples of successful seepage predictions are provided in the detailed views shown in Figure 36 through Figure 38 below. These figures were generated from screen-shots of Google Earth™ .KMZ overlays with aerial imagery, and show various examples of successful seepage predictions in areas where likely seepage was observed during FDEM data collection. In some cases, false positives are observed at the locations of infrastructure (magnetic gradiometry was not performed here), but most elevated Seepage Index values and corresponding seepage predictions tend to not be associated with these features.

Figure 39 and Figure 40 present the results of applying the rather simple K-means clustering algorithm in an attempt to differentiate between seepage and non-seepage areas of the canal. While the K-means clustering results are somewhat successful in clustering seepage areas from non-seepage areas, the results are not as successful as in the case of South Canal, and future work is needed to explore other clustering algorithms for improved seepage detection.

Similarly, Figure 41 and Figure 42 present results of applying SVMs for prediction of seepage locations (based on Seepage Index thresholding). SVM classification results are observed to be extremely promising in identifying data patterns related to seepage, based on the results presented here. The metrics calculated for correct predictions versus false positives and false negatives is again quite compelling for this type of machine learning classification technique applied to seepage detection.

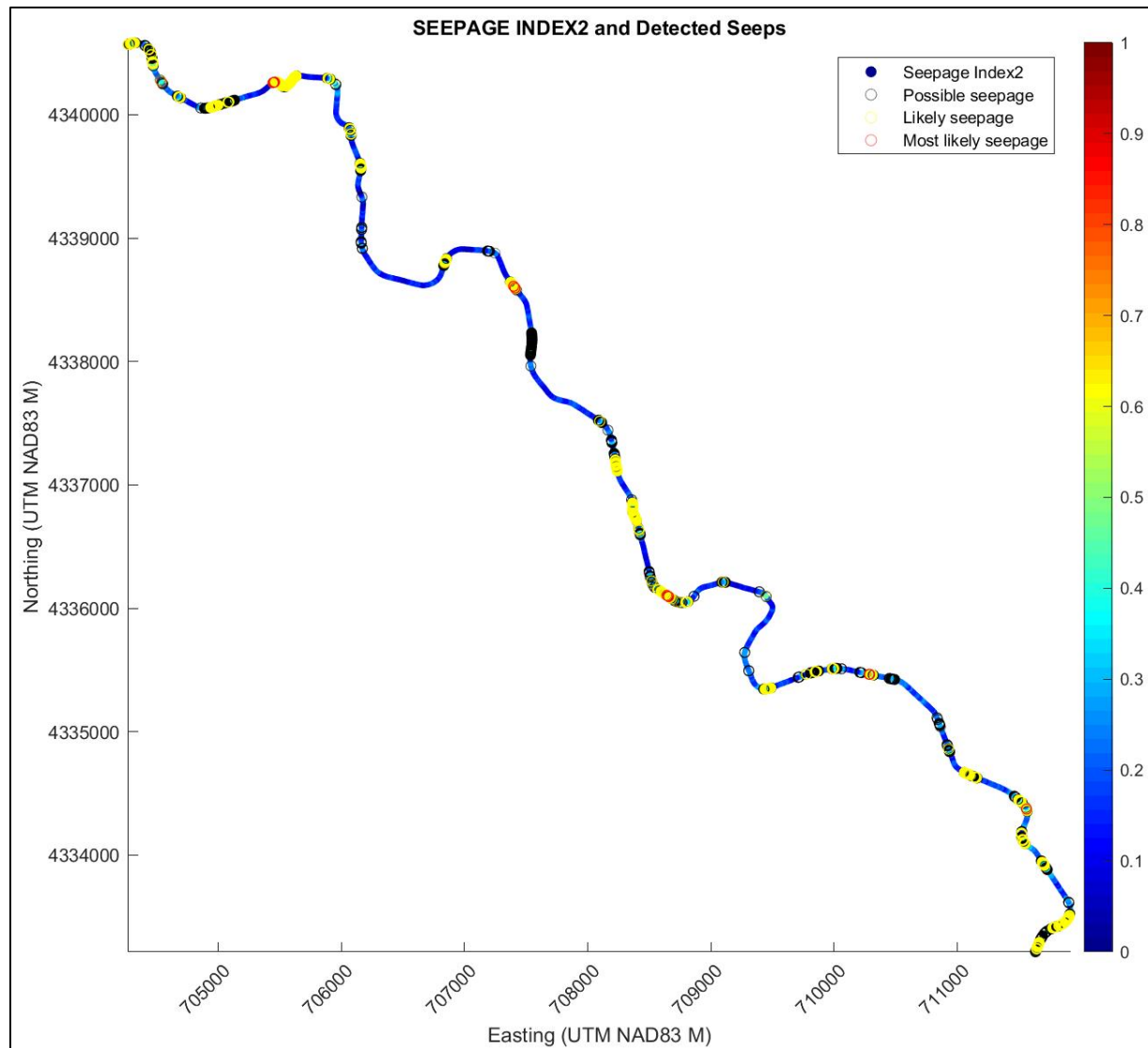


Figure 35. Detected seepage locations along Government Highline Canal using thresholding of the detrended Seepage Index values. Here, red circles are “most likely” seepage, yellow circles are “likely” seepage, and black circles are “possible” seepage. The “raw” Seepage Index values are plotted under these predicted locations.

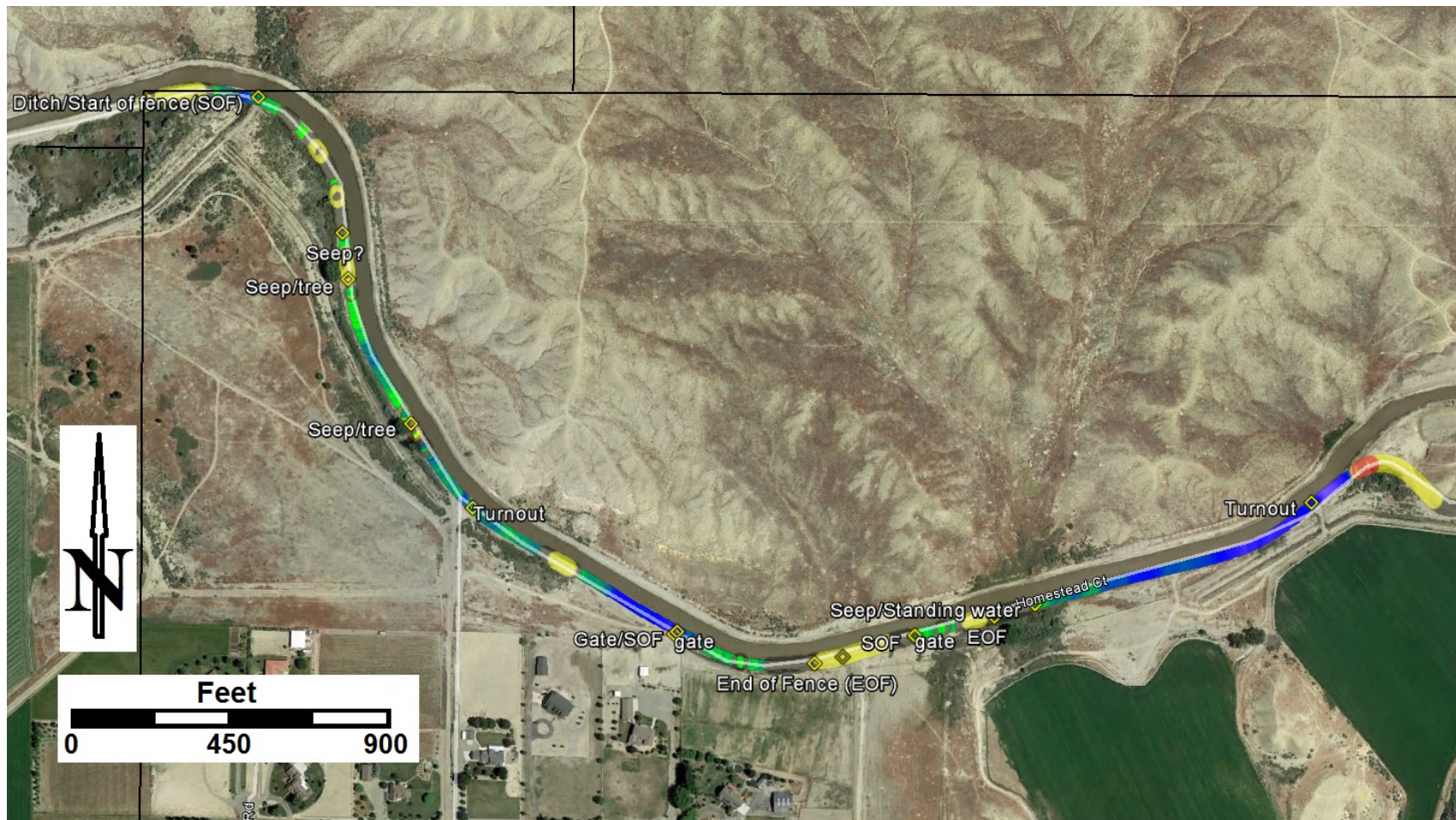


Figure 36. A detailed view of Government Highline Canal showing seemingly successful seepage detections and generally elevated Seepage Index values consistent with visible vegetation patterns.

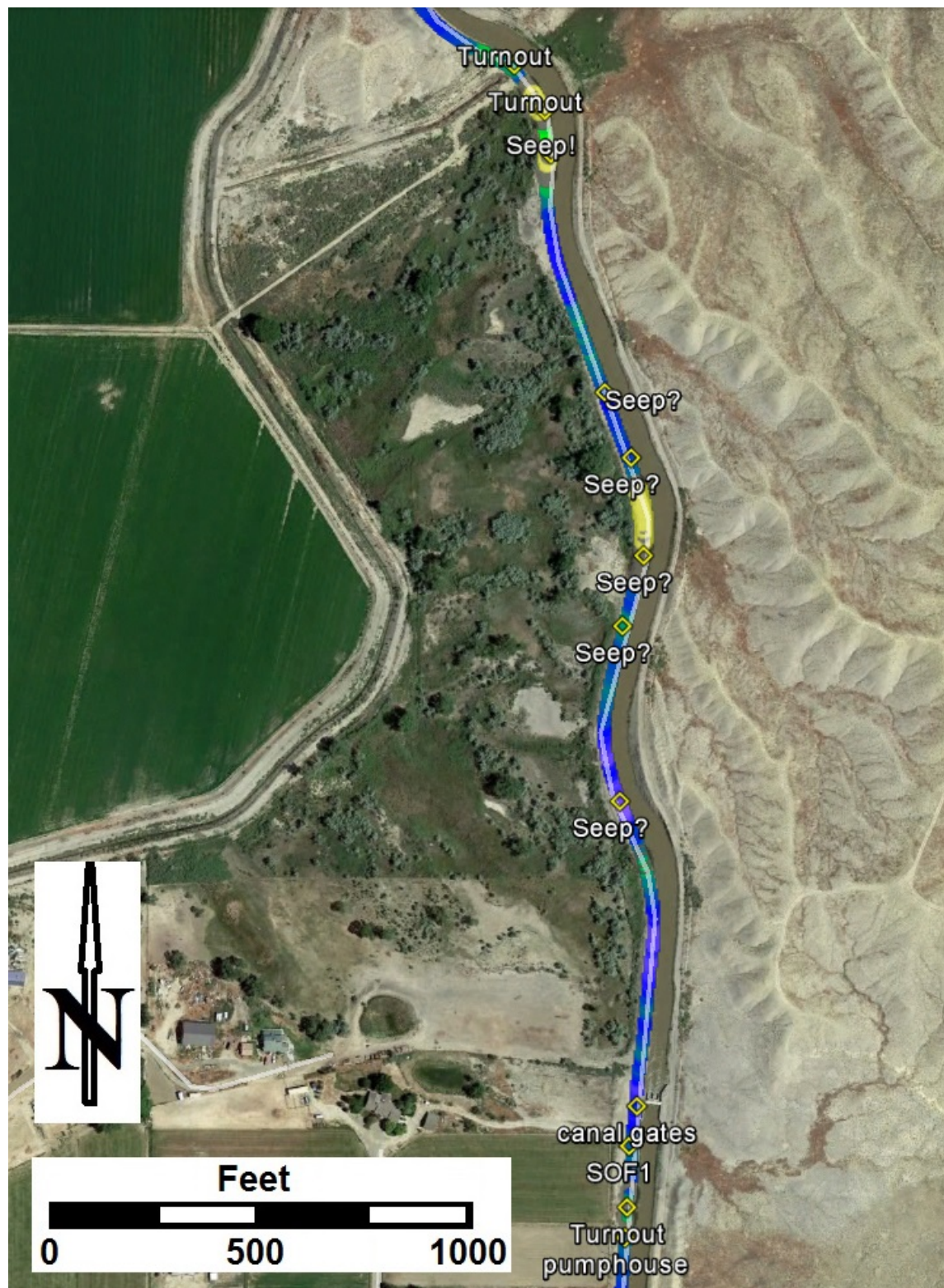


Figure 37. A second detailed view showing various successful seepage detections and generally elevated Seepage Index values in areas of either observed apparent seepage (noted and coordinates recorded with a hand-held GPS during data collection) or elevated vegetation growth.

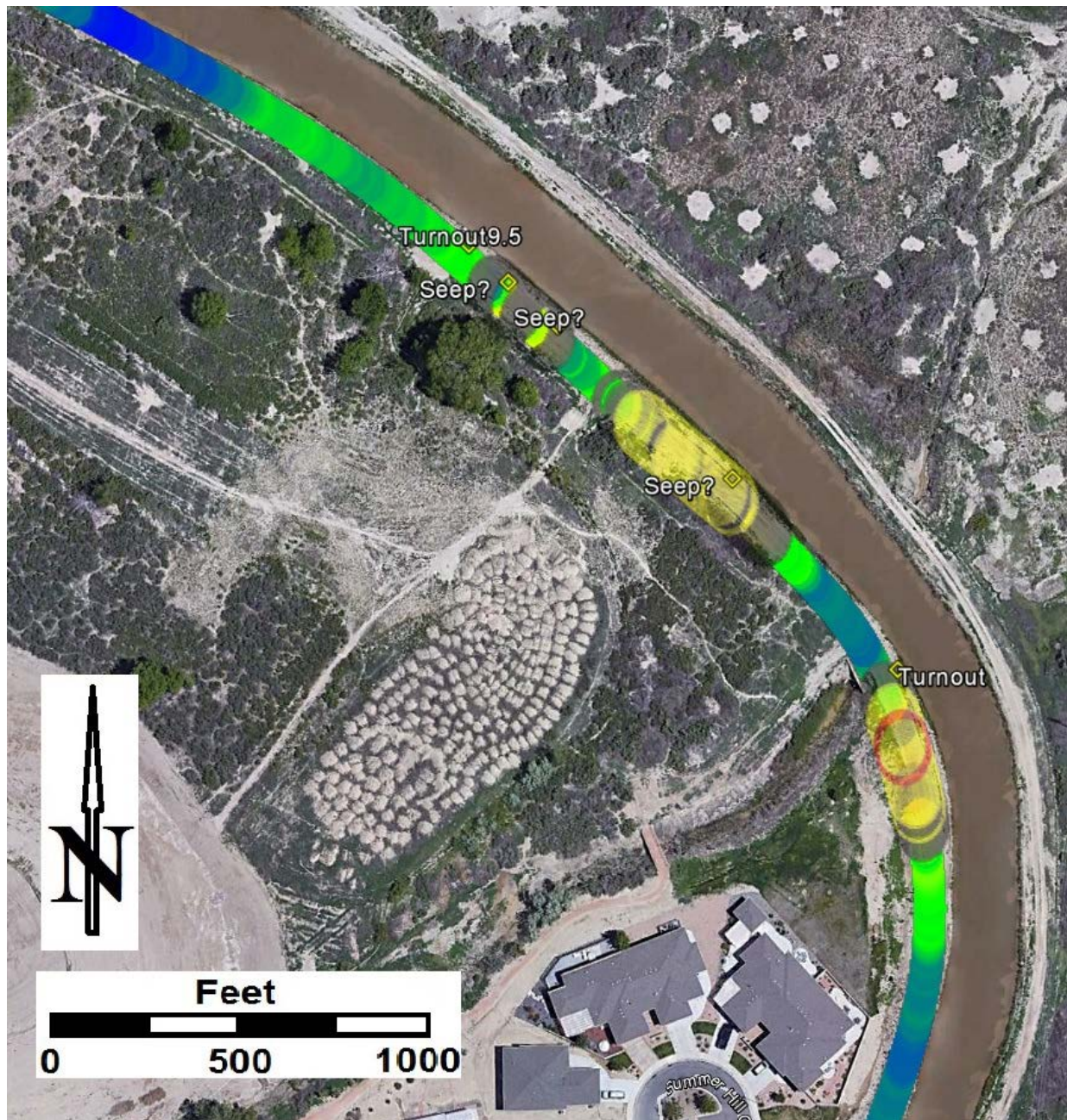


Figure 38. A third detailed view along Government Highline Canal showing successful seepage detections and generally elevated Seepage Index values consistent with observed seepage in the field and vegetation patterns.

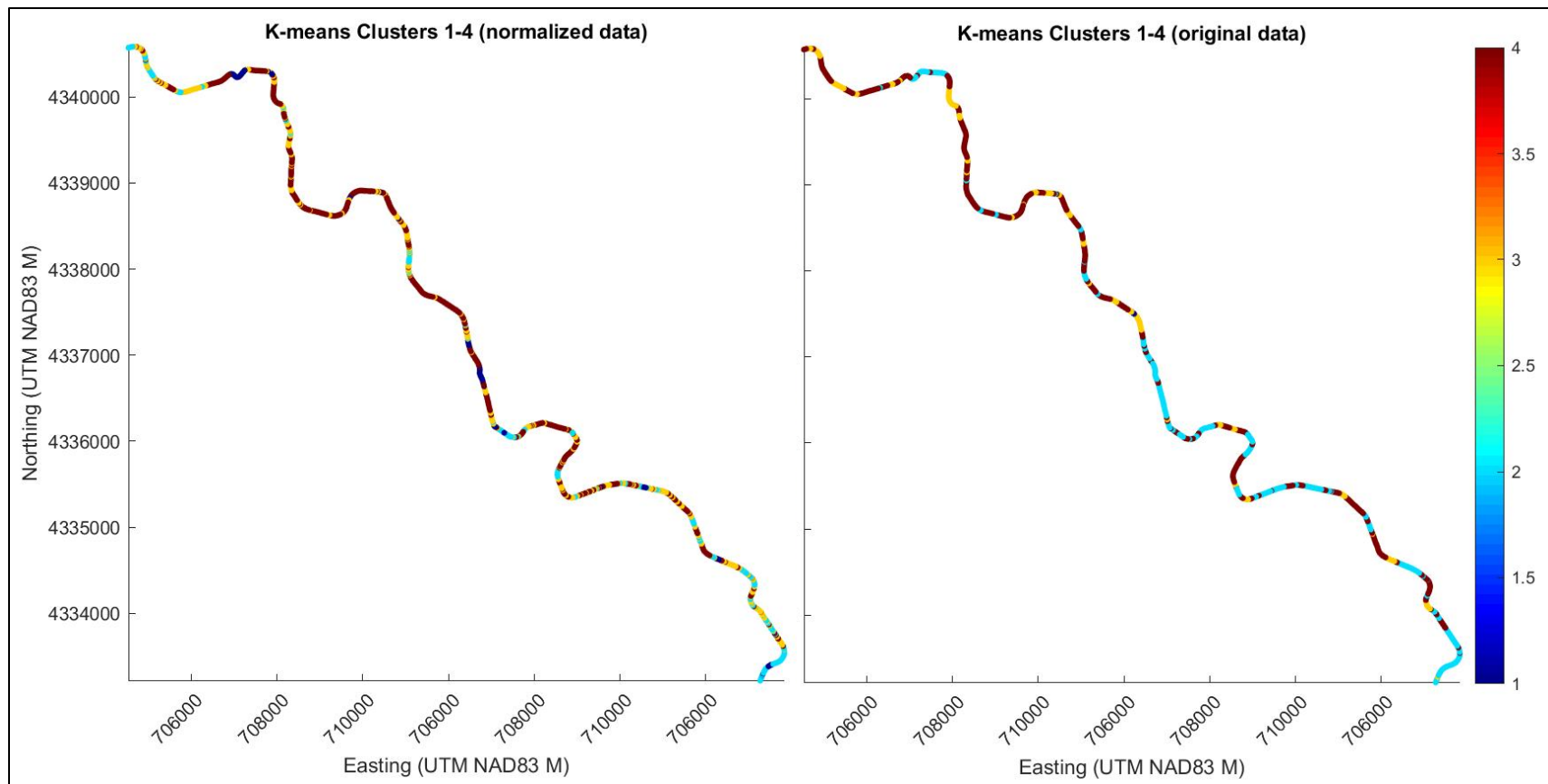


Figure 39. Results of K-means clustering applied to the original and normalized (between 0 and +1) data-types and metrics along Government Highline Canal.

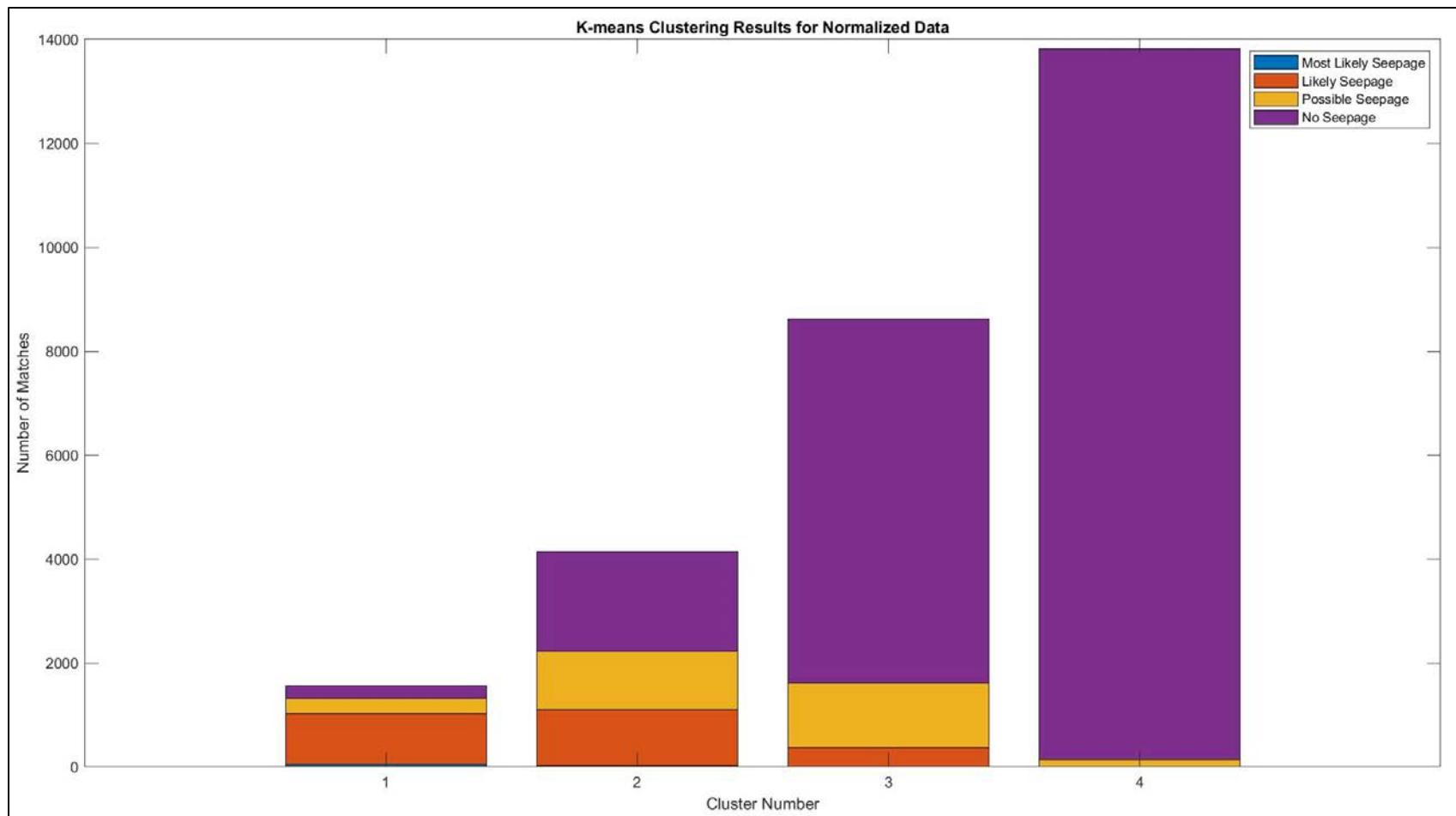


Figure 40. Stacked bar graph showing the results of K-means clustering for seepage prediction. While most of cluster 1 captures seepage (and captures all “most likely” seepage points), the results of clusters 2 and 3 are not as successful in differentiating between “likely” and “possible” seepage locations with background “non-seepage” locations (based on Seepage Index thresholding).

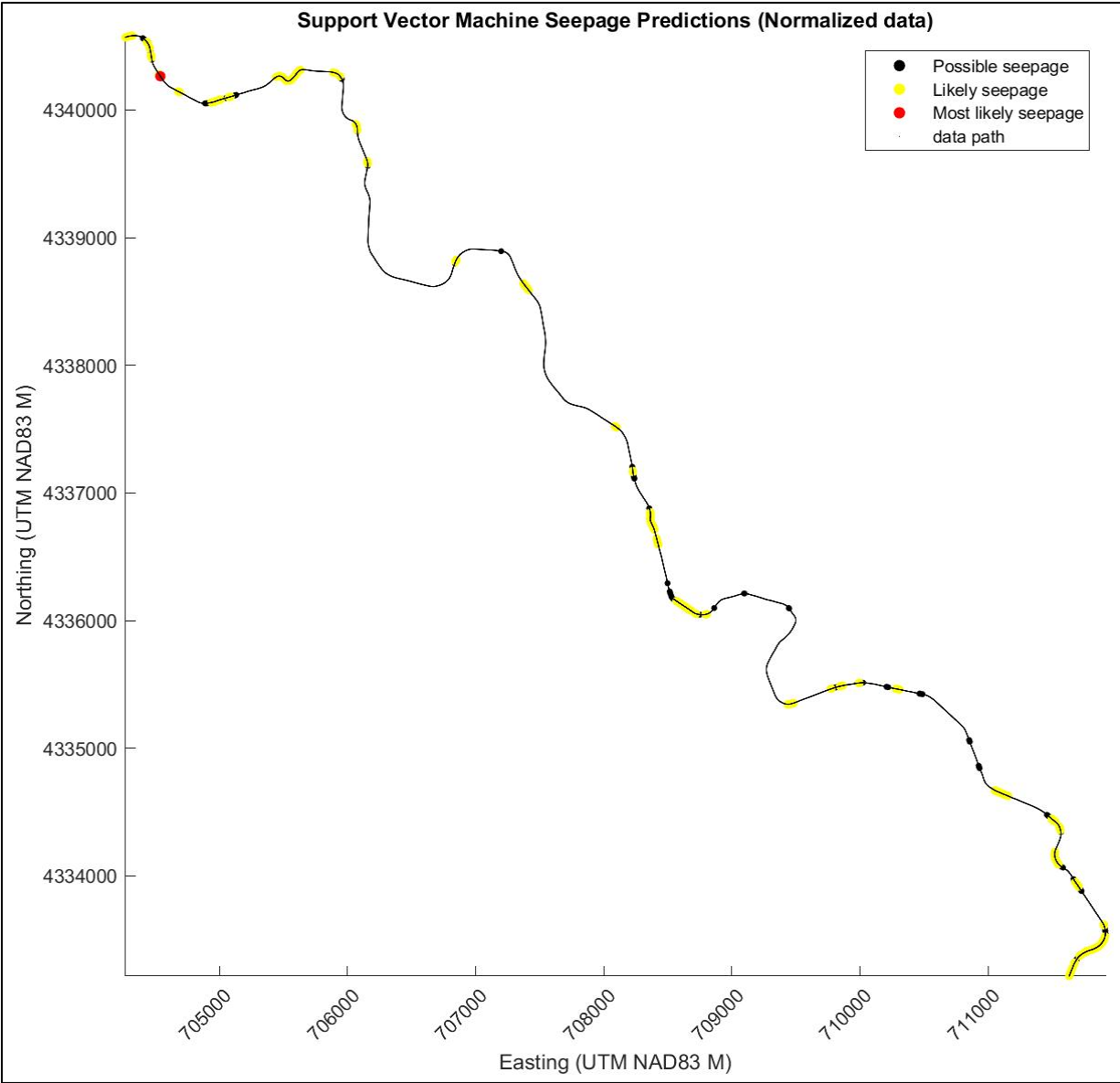


Figure 41. A spatial plot of the results of SVM prediction results along Government Highline Canal.

Rapid Canal Embankment and Levee Health Assessment and Seepage Detection

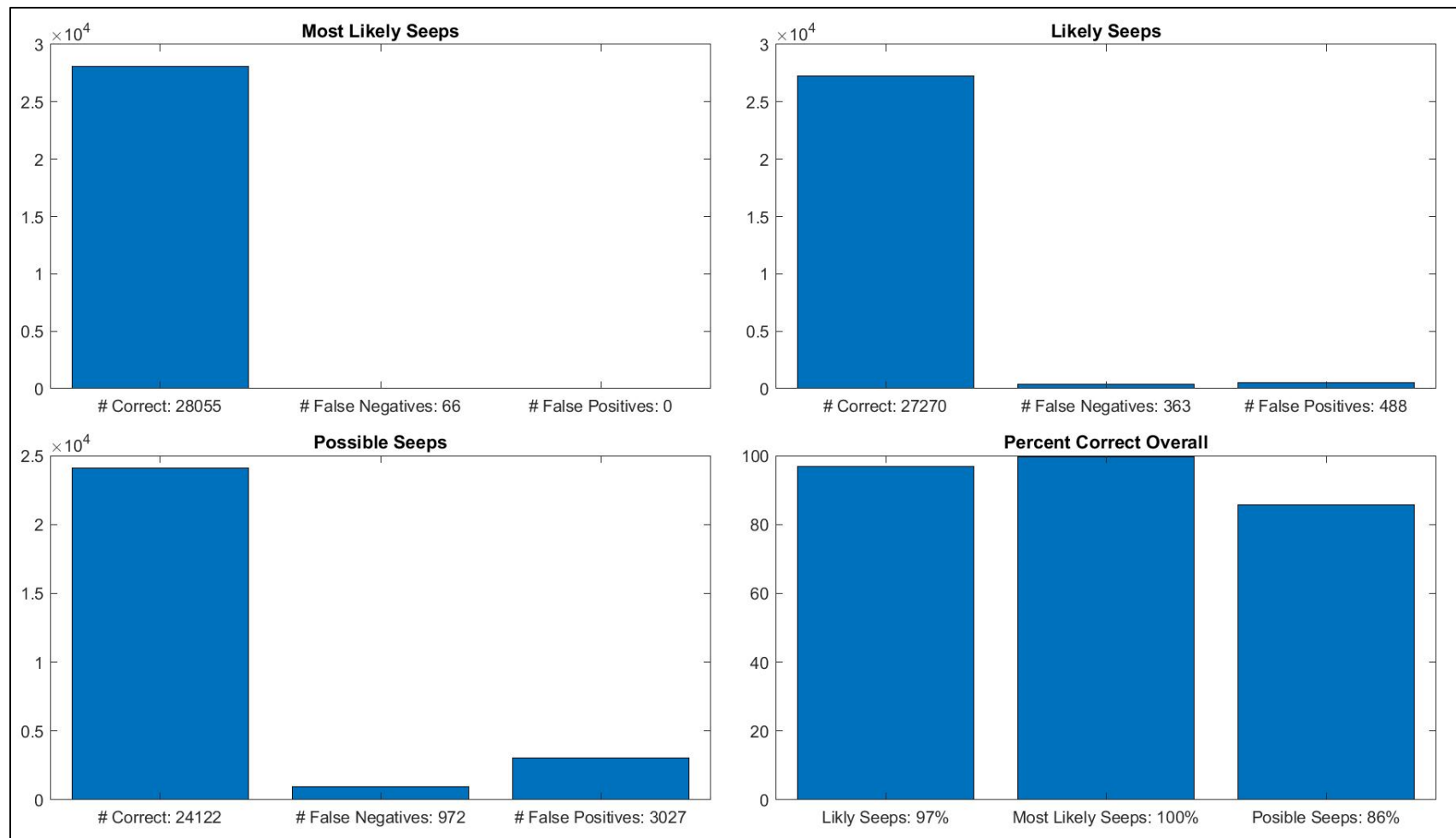


Figure 42. Comparisons and evaluations of SVM prediction results relative to thresholding-based predictions along the Government Highline Canal study area. In general, these results again show an excellent application of SVMs for seepage prediction, based on the nature of our input multidimensional data set.

Strawberry Highline Canal

Results from testing along Government Highline Canal are presented below in Figure 43 through Figure 49. Generally, results of Seepage Index calculations and seepage predictions (using the thresholding approach described above) are again corroborated nicely by field observations of wet areas, lush wetland vegetation such as cattails and reeds along the canal. Generally, there are several segments and also several discrete locations of seepage predicted along the length of canal segment surveyed for this study (Figure 43).

Detailed examples of successful seepage predictions are provided in the detailed views shown in Figure 44 and Figure 45 below. These figures were generated from screen-shots of Google Earth™ .KMZ overlays with aerial imagery, and show various examples of successful seepage predictions in areas where likely seepage was observed during FDEM data collection, and also at locations where known seepage locations were provided by Provo Area Office personnel. Figure 45 shows successful prediction (or identification) of a recent embankment failure that occurred during the previous irrigation season. This location also has nearby predicted seepage locations that may or may not warrant closer inspection (e.g. this section of the canal may be experiencing poor performance due to a local change in geology or gradation of embankment materials).

In few cases, false positives are observed at the locations of infrastructure. However, magnetic gradiometry was performed at this last test-canal and integrated into the data analysis workflow. As a result, most of the Seepage Index values at these infrastructural features were successfully suppressed by incorporation of the magnetic data into the workflow. Incorporation of the magnetic data was so effective overall, that it fully suppressed at least one known seepage location due to the presence of reinforced concrete linings at the same location. This approach could prove problematic in identifying seeps along reinforced concrete lining. While this seepage still poses the issue of unnecessary water conveyance loss, these lined and reinforced sections of canal are at least not expected to catastrophically fail.

Figure 46 and Figure 47 present the results of applying the rather simple K-means clustering algorithm in an attempt to differentiate between seepage and non-seepage areas of the canal. While the K-means clustering results are somewhat successful in clustering seepage areas from non-seepage areas, the results are again not as successful as in the case of South Canal.

Similarly, Figure 48 and Figure 49 present results of applying SVMs for prediction of seepage locations (based on Seepage Index thresholding). SVM classification results are observed to be extremely promising in identifying data patterns related to seepage, based on the results presented here. The metrics calculated for correct predictions versus false positives and false negatives is again quite compelling for this type of machine learning classification technique applied to seepage detection. As expected, the worst fit is found with “possible” seepage location predictions, and yet 72% are still predicted correctly, with the majority of incorrect predictions being a conservative false positive prediction. This most likely has to do with the level of extensive vegetative growth along much of the length of Strawberry Highline Canal that may or may not have to do with seepage.

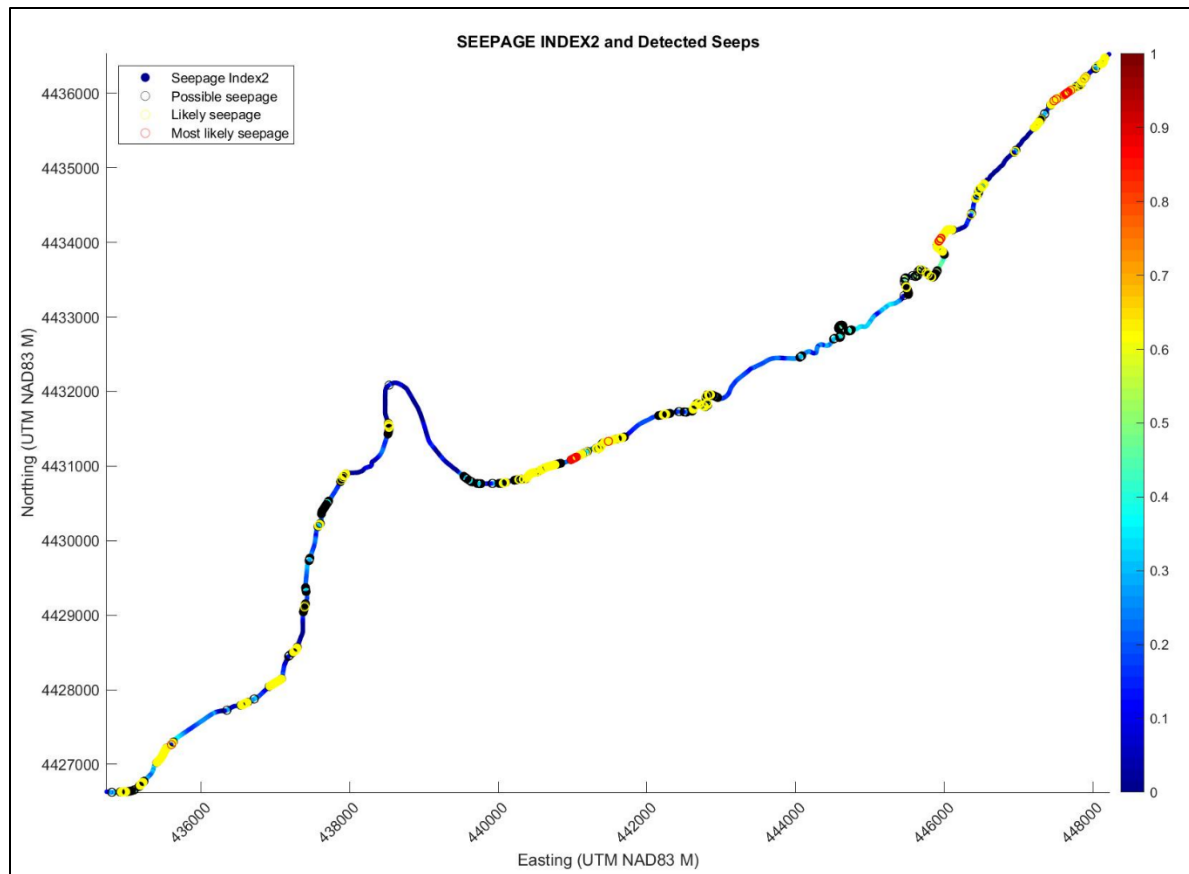


Figure 43. Detected seepage locations along Strawberry Highline Canal using thresholding of the detrended Seepage Index values. Here, red circles are “most likely” seepage, yellow circles are “likely” seepage, and black circles are “possible” seepage. The “raw” Seepage Index values are plotted under these predicted locations.

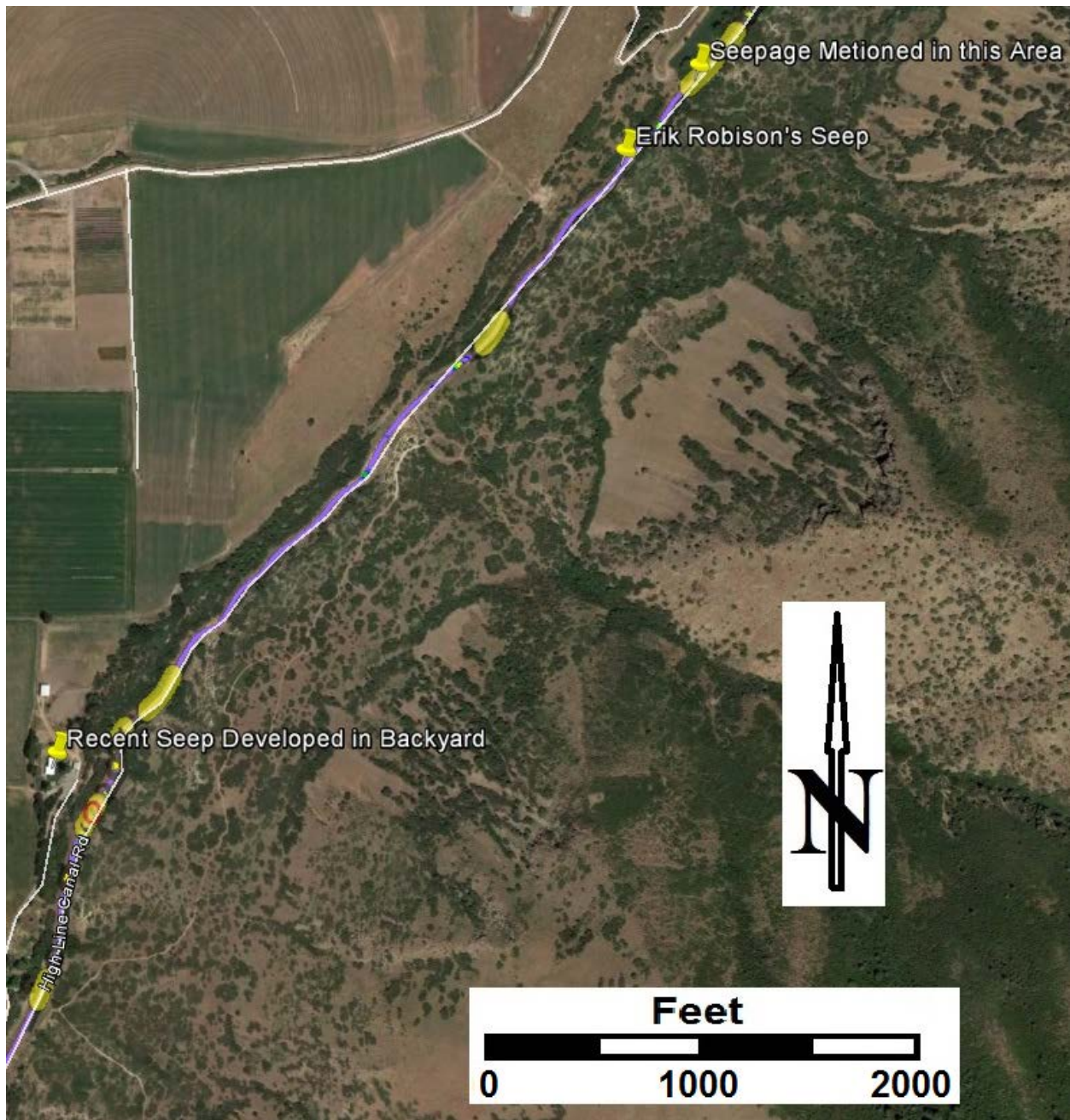


Figure 44. A detailed view showing an example of a failed seepage prediction using thresholding of the Seepage Index (false negative at “Erik Robison’s Seep”) due to the presence of extensive reinforced concrete lining that resulted in suppression of the Seepage Index values at this location. However, two nearby general locations adjacent to this false negative were pointed out in the field by Provo Area Office personnel, and these locations appear to have successful predictions of seepage.

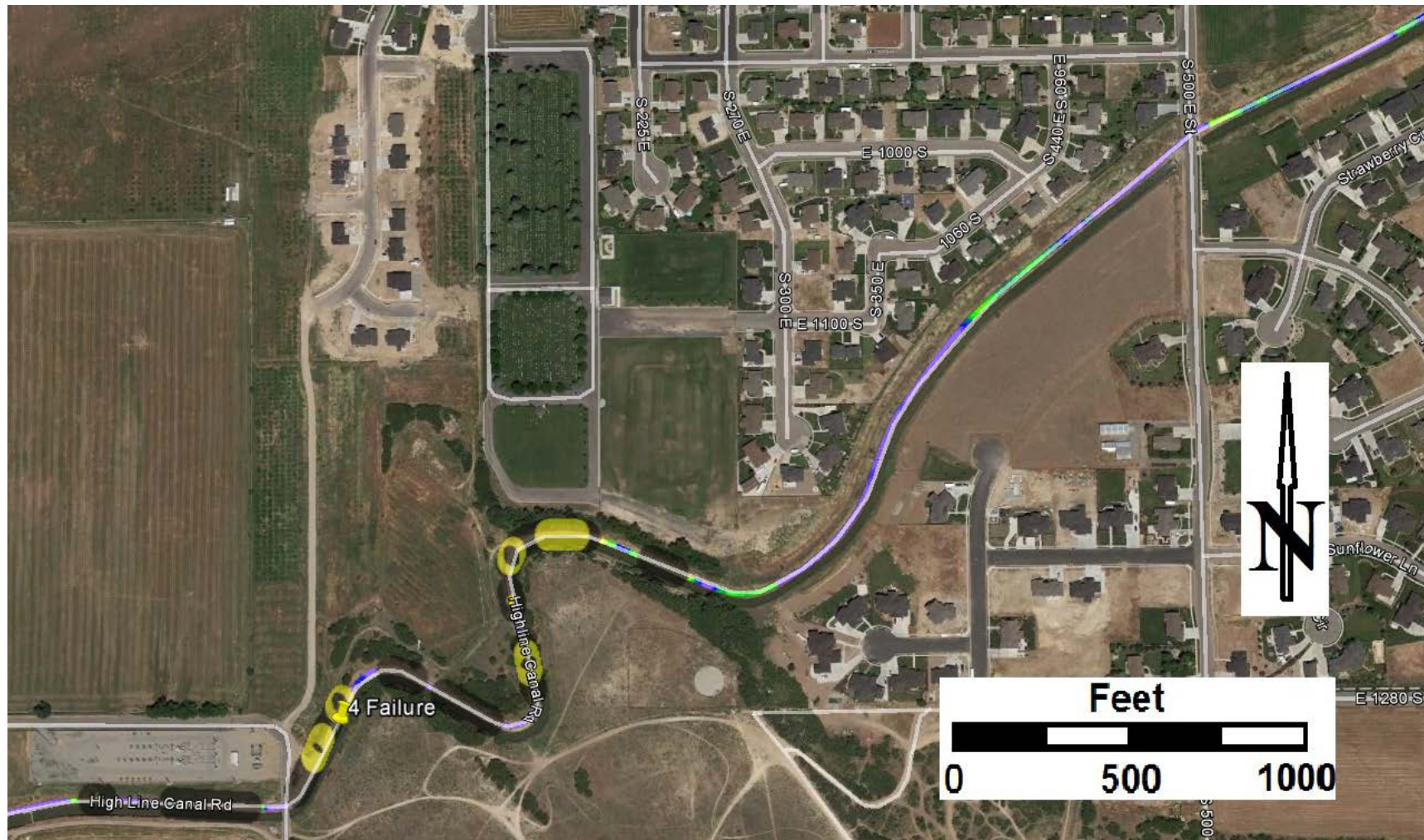


Figure 45. A second detailed view near the southern end of the segment of South Canal tested. Here, various successful seepage detections and generally elevated Seepage Index values are validated with extensive observed apparent seepage (noted and coordinates recorded with a hand-held GPS during data collection).

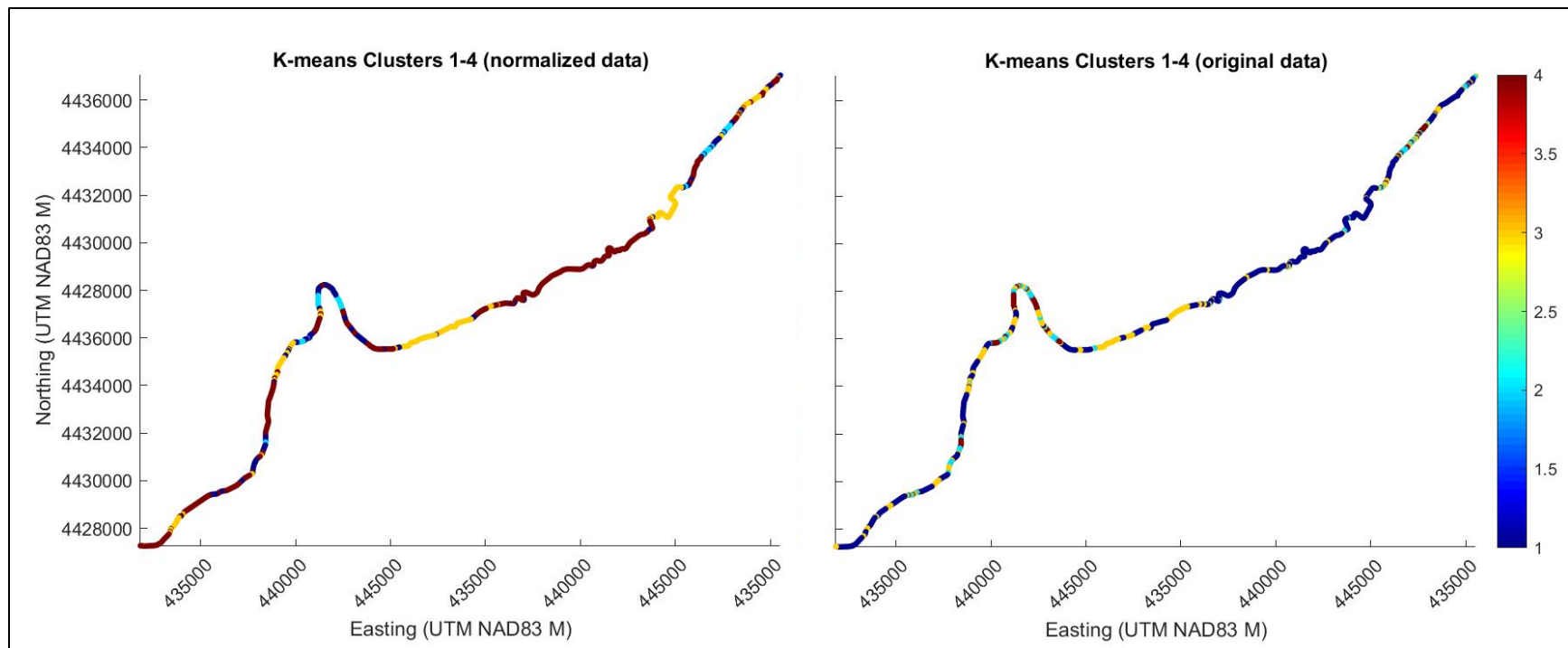


Figure 46. Results of K-means clustering applied to the original and normalized (between 0 and +1) data-types and metrics along South Canal.

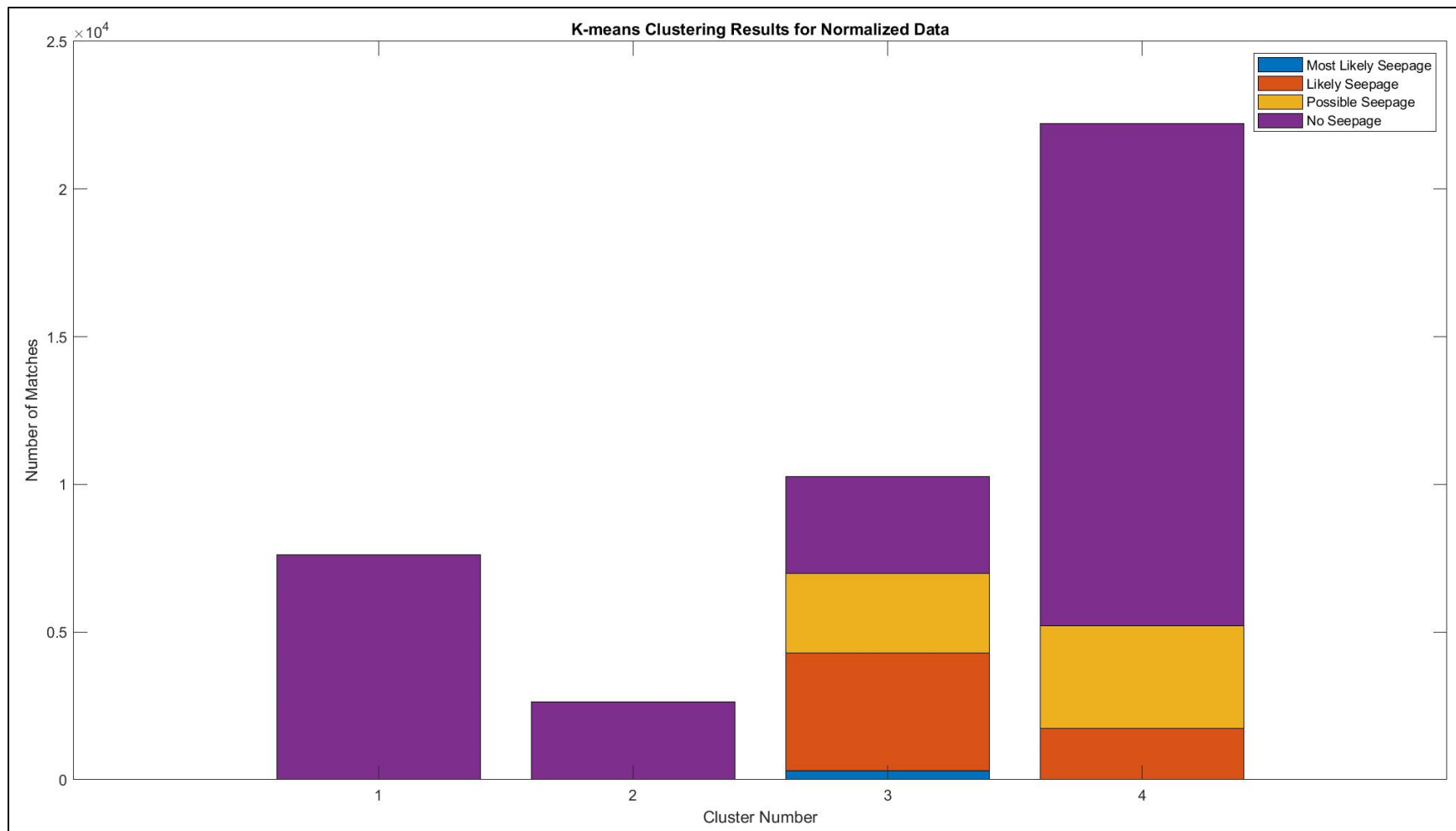


Figure 47. Stacked bar graph showing the results of K-means clustering for seepage prediction. While most of clusters 1 and 2 contain “non-seepage” locations, and most of cluster 3 only captures seepage locations (and captures virtually all “most likely” seepage points), the results of cluster 4 is not as successful in differentiating between “likely” and “possible” seepage locations with background “non-seepage” locations (based on Seepage Index thresholding).

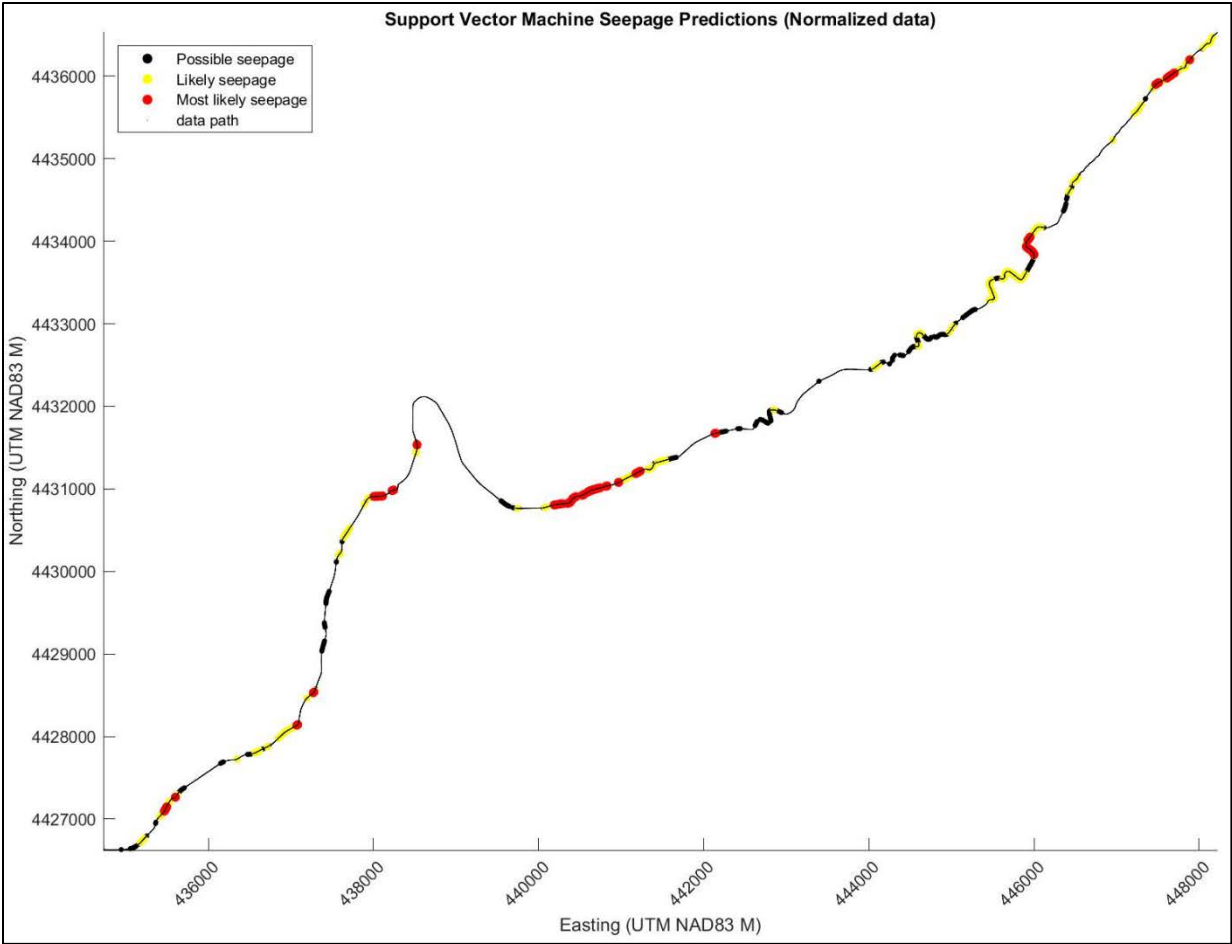


Figure 48. A spatial plot of the results of SVM prediction results along Strawberry Highline Canal.

Rapid Canal Embankment and Levee Health Assessment and Seepage Detection

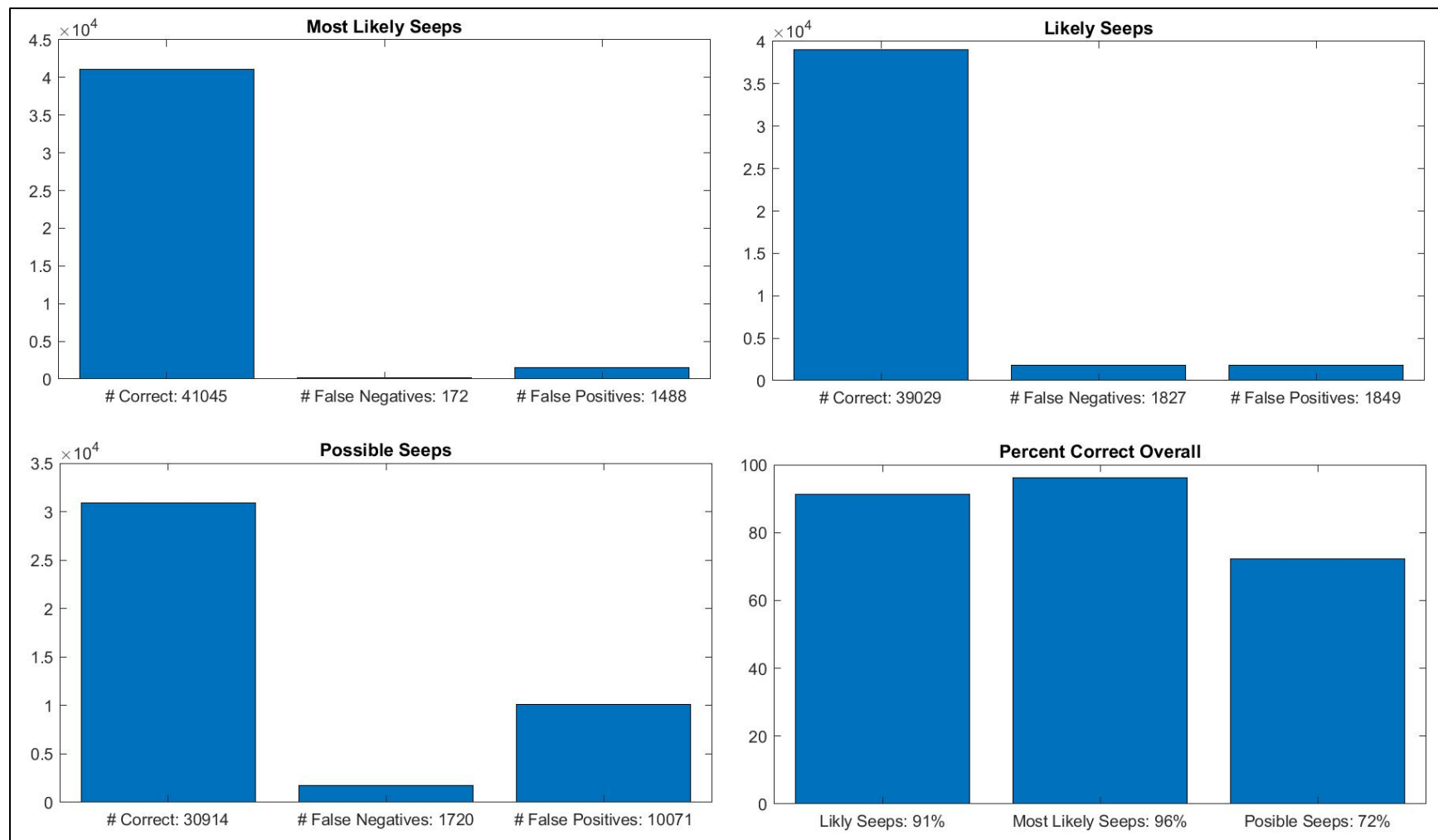


Figure 49. Comparisons and evaluations of SVM prediction results relative to thresholding-based predictions. In general, these results show an excellent application of SVMs for seepage prediction, based on the nature of our input multidimensional data set. As expected, the worst fit is found with “possible” seepage location predictions, and yet 72% are still predicted correctly, with the majority of incorrect predictions being a conservative false positive prediction. This most likely has to do with the level of extensive vegetative growth along much of the length of Strawberry Highline Canal that may or may not have to do with seepage.

Conclusions

This research serves the interests of multiple agencies and stakeholders across the United States, including Reclamation, the US Army Corps of Engineers, State-level and private canal and levee owners, as well as nearby public residents and property owners. Overall, this research effort was successful at meeting project objectives, including the following:

1. Evaluation of the capabilities of various FDEM systems for rapid collection of reliable geophysical profiling data on a large-scale (along hundreds of miles of canal embankments) that is sensitive to the presence of seepage.
2. Evaluation and development of algorithms for performing general embankment condition assessment and seepage detection by means of integration of FDEM and NDVI data analysis.
3. Automation of data analysis and interpretation workflow in a manner that makes production of results and time-lapse monitoring a more feasible option to system-wide canal embankment interrogation.

This research has provided as basic foundation for seepage detection and general embankment health assessment with the integration of geophysical profiling and satellite remote sensing data analysis. Here, the use of several data-types has been shown to help minimize the impact of random data errors on the resulting “Seepage Index” values and corresponding predictions of seepage locations. Challenges with simple averages of NDVI data due to extensive mountainside vegetation are compensated by addition of multiple NDVI spatial attributes (e.g. local cross-line gradients). Incorporating NDVI data attributes also helped to resolve issues with DC shifts in NDVI normalizations resulting from inconsistencies in input remote sensing data layers and radiometric corrections (i.e., averages shift, but gradients are similar). The addition of magnetic gradiometry was shown to be quite successful at suppressing metal-related FDEM data anomalies that affected the overall results via Seepage Index thresholding. Similarly, the addition of time-lapse EM data collection and analysis for watered-up/dewatered conditions was shown to be useful in highlighting segments of canal that are most likely seeping.

The figures and discussions presented in this report only provide a small subset of the overall value of results from the research effort, mainly due to the vast amount of data coverage along the test canals surveyed. The reader is encouraged to access the electronic deliverables described at the end of this report, including .CSV files of georeferenced computed values, and Google EarthTM .KMZ files containing overlays of data, Seepage Index values, and predicted seepage locations.

Although initial efforts toward automation of seepage prediction with clustering and classification algorithms have proved to be promising in this study, future work is still needed to explore potentially more robust alternatives to predicting seepage (e.g., incorporating additional data-types, implementing more advanced machine learning algorithms such as Fussy Logic, etc.). This research is considered to be mutually complimentary with standard and recently developed approaches to embankment inspection and monitoring, and has a lot of opportunity for future testing, refinement, and unique field applications.

Recommendations and Next Steps

There are several opportunities to refine and apply this new approach to rapid data collection and integrated analysis for various Reclamation and similar stakeholder needs. Future efforts should primarily include the following:

1. Further design improvements and rebuild of the non-magnetic cart for sake of being modular and more easily transported (e.g., shippable).
2. Explore the use of additional remote sensing data products, such as the following:
 - a. Embankment elevation profile data from LIDAR or photogrammetry to detect vulnerable or suspect narrow points and over-steepened embankment sections.
 - b. Day versus night time-lapse infrared imaging of embankments to identify wet areas.
3. Explore the potential for integration of Seepage Index values with more quantitative yet sparse piezometer array and heat-tracer seepage data along Truckee Canal in Nevada. This could be used to 1) help validate the approach of this study and resulting patterns in calculated Seepage Index values, and 2) see if modeled seepage losses can be better calibrated on a system-wide scale using Seepage Index values to help interpolate loss rates between monitoring stations.
4. Long-term time-lapse monitoring for FDEM and NDVI changes in order to prioritize repair efforts.
5. Test incorporation of additional geophysical data-types (e.g., time-domain EM)
6. Explore additional FDEM data metrics in order to attempt differentiation of shallow versus deep and focused versus distributed seepage patterns (based on divergent variations of different FDEM systems with different DOI, and perhaps looking at in-line gradients or “sharpness” of Seepage Index fluctuations).

In addition to the above recommendations, further testing will be required to evaluate and develop SVM and other classification algorithms as reliable tools for predicting seepage or other issues along canals using this study’s approach. This last point will require ongoing cooperative partnerships with various stakeholders, in order to provide verified seepage location and performance data for use as training sets. This information will also help to further evaluate the other data types and help to update and optimize analysis workflows.

References

- Geonics, 2018. Geonics Training: EM-61 EM-31 EM-34 EM-38. Available at: <http://geonicstraining.com/exploration-depth-geonics-equipment/> . Accessed September 23, 2018.
- Guyot, G., and FaGu, X., 1994. Effect of radiometric corrections on NDVI-determined from SPOT-HRV and Landsat-TM data, Remote Sensing of Environment, Vol 49, Issue 3. [https://doi.org/10.1016/0034-4257\(94\)90012-4](https://doi.org/10.1016/0034-4257(94)90012-4)
- Lancaster, D.M., 1952. MEASUREMENT OF SEEPAGE LOSSES FROM IRRIGATION CANALS, in Proceedings Summer ASCE Convention, Denver, Colorado. Available at https://www.usbr.gov/tsc/techreferences/hydraulics_lab/pubs/PAP/PAP-0015.pdf.
- NASA, 2000. Measuring Vegetation (NDVI & EVI), Web-based Arrticle. Available at: https://earthobservatory.nasa.gov/Features/MeasuringVegetation/measuring_vegetation_2.php. Accessed September 21, 2018.
- Wikipedia contributors, 2018a. Cluster analysis. Wikipedia, The Free Encyclopedia. Available at: https://en.wikipedia.org/w/index.php?title=Cluster_analysis&oldid=860247097 . Accessed September 21, 2018.
- Wikipedia contributors, 2018b. Machine learning. Wikipedia, The Free Encyclopedia. Available at: https://en.wikipedia.org/w/index.php?title=Machine_learning&oldid=860184347 . Accessed September 21, 2018.

Data Sets that Support the Final Report

Several electronic deliverables and data sets associated with this research are available:

- Share Drive folder name and path where data are stored:
<https://drive.google.com/drive/folders/1HCSA31Mvrkm1C7O6O0F5TuFuqZVL49H?usp=sharing>
- Point of Contact name, email, and phone: Justin B. Rittgers, 303-445-3010, jrittgers@usbr.gov
- Short description of the data:
 - Raw NDVI images used for this research
 - Extracted NDVI values along the canals studied
 - Matlab .M File scripts developed for performing calculations and supplementary figure generation
 - Output ASCII (.CSV) files containing:
 - Final FDEM and NDVI data attributes values interpolated along the master GPS track for each canal studied
 - Coordinates of each category of predicted seepage (e.g. “most likely,” “likely,” and “possible” seepage).
 - Supplementary data and results figures output from scripts
 - Google Earth™ .KMZ files that present overlays of the resulting Seepage Index values and corresponding predicted seepage locations
- Keywords: NDVI, Seepage.
- Approximate total size of all files: 493 Files, 46 Folders, 3.86 Gb

