Geophysical Research Letters

RESEARCH LETTER

10.1002/2016GL071334

Key Point:

- The paper presents a workflow allowing fast mapping of permafrost in a way which is both consistent with AEM-data and knowledge of an expert

Correspondence to:

M. L. Gulbrandsen,
mats.lundh@nbi.ku.dk

Citation:

Gulbrandsen, M. L., B. J. Minsley, L. B. Ball, and T. M. Hansen (2016), Semiautomatic mapping of permafrost in the Yukon Flats, Alaska, *Geophys. Res. Lett.*, 43, 12,131–12,137, doi:10.1002/2016GL071334.

Received 4 OCT 2016

Accepted 14 NOV 2016

Accepted article online 18 NOV 2016

Published online 9 DEC 2016

Semiautomatic mapping of permafrost in the Yukon Flats, Alaska

Mats Lundh Gulbrandsen¹ , Burke J. Minsley² , Lyndsay B. Ball² , and Thomas Mejer Hansen¹
¹Solid Earth Physics, Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark, ²Crustal Geophysics and Geochemistry Science Center, U.S. Geological Survey, Denver, Colorado, USA

Abstract Thawing of permafrost due to global warming can have major impacts on hydrogeological processes, climate feedback, arctic ecology, and local environments. To understand these effects and processes, it is crucial to know the distribution of permafrost. In this study we exploit the fact that airborne electromagnetic (AEM) data are sensitive to the distribution of permafrost and demonstrate how the distribution of permafrost in the Yukon Flats, Alaska, is mapped in an efficient (semiautomatic) way, using a combination of supervised and unsupervised (machine) learning algorithms, i.e., Smart Interpretation and K-means clustering. Clustering is used to sort unfrozen and frozen regions, and Smart Interpretation is used to predict the depth of permafrost based on expert interpretations. This workflow allows, for the first time, a quantitative and objective approach to efficiently map permafrost based on large amounts of AEM data.

1. Introduction

Permafrost is a state where the temperature in a subsurface layer stays below the freezing point throughout the year for at least two consecutive years and is hence mostly common in the arctic regions. As we experience higher surface temperatures due to changing climate, an increasing number of research projects address the issue of how warming will affect permafrost and the potential consequences of continued permafrost thaw. *Froese et al.* [2008] argue that permafrost can be quite stable and resilient to warmer climate under certain environmental conditions. However, *Vaughan et al.* [2013] state that a consequence of increasing surface temperatures related to global warming will be thawing of permafrost and increased formation of taliks, i.e., unfrozen zones in regions of permafrost. These changes in permafrost can have impact on several different factors such as infrastructure foundations, hydrogeological processes, climate change feedback, ecology, and the environment. Hydrogeological processes can be expected to change through enhanced surface water and groundwater interaction through taliks, and the contribution of groundwater to streamflow can be altered [Walvoord and Striegl, 2007; Bense et al., 2009]. According to studies by *Knorr et al.* [2005], *Schuur et al.* [2009], *Koven et al.* [2011], and *Schuur et al.* [2015], the large amounts of carbon stored in permafrost soils will have a significant impact on climate, acting as a positive feedback when emitted to the atmosphere due to thawing of permafrost. Thawing of permafrost can also affect the areal extent of wetlands [Avis et al., 2011] and fire frequency and intensity [Mack et al., 2011], having substantial ecological influence [O'Donnell et al., 2011]. It is therefore important to be able to map the distribution of permafrost, both laterally and with depth, over large areas.

In a study by *Minsley et al.* [2012] it is shown that remotely sensed airborne electromagnetic (AEM) data are sensitive to the subsurface distribution of permafrost. The process of manually interpreting the distribution of permafrost from AEM data is, however, very time-consuming and subjective. In addition, it is impossible to include all available data in the mapping process, especially considering the increasingly large AEM data sets being collected. In this study we demonstrate a workflow based on a combination of supervised and unsupervised learning. It consists of a combination of a few manual “expert” picks of the base of frozen soil and two machine-learning techniques: Smart Interpretation and K-means clustering. This workflow allows for a much faster mapping of the distribution of permafrost in a quantitative and objective way that includes all available data in the process and at the same time makes sure that the final map is both consistent with the data and the expert knowledge. The workflow is demonstrated on a 300 km² block of AEM data within the Yukon Flats, Alaska (see Figure 1), acquired by using the Fugro RESOLVE system along closely spaced flight lines. The Yukon Flats is sited in an area of interior Alaska near the boundary of continuous permafrost to the north and more discontinuous permafrost to the south, and is hence an important location for examining the dynamics of permafrost [Minsley et al., 2012].

Figure 1. The Yukon Flats study area, displayed together with the surface (0–1.2 m) resistivity values from the inversion of the obtained AEM data. Line **A** represents the AEM flight line 10020, and line **B** represents the combined flight lines 10142 (southwest of Fort Yukon) and 10140 (northeast of Fort Yukon). Coordinates are UTM Zone 6N.

2. Methods

The workflow is built up of two parts. First, the location frozen soil is found by using unsupervised K-means clustering [see, e.g., Bishop, 2006; Everitt et al., 2001]. Then, the thickness of the permafrost is estimated by using supervised regression, also known as Smart Interpretation (SI) [Gulbrandsen et al., 2015].

2.1. Study Area

The surface geology of the study area within the Yukon Flats (Figure 2) consists of Quaternary alluvial fan, terrace, and floodplain deposits of the Yukon River that overlay extensive Pliocene-Pleistocene lacustrine silts and clays [Williams, 1962; Clark et al., 2009]. The alluvial sediments are primarily coarse-grained gravels that extend to a depth of approximately 30–40 m. This geological setting is expected to exhibit a significant contrast in electrical resistivity, with gravels and other coarse-grained materials being more resistive than silts and clays [Hoekstra et al., 1975]. Nearby ground-based electrical resistivity measurements [Froese et al., 2005] support this model and suggest that large-scale AEM surveys can be useful in delineating these features over a large area.

2.2. AEM Survey

AEM data were acquired in the Yukon Flats area of interior Alaska in 2010 [Ball et al., 2011; Minsley et al., 2012] by using the Fugro RESOLVE frequency domain system. The RESOLVE system comprises six coil pairs that operate between approximately 0.4 and 140 kHz, and the instrument is towed with nominal 30 m ground clearance beneath a helicopter. Data are output approximately every 3 m along the flight lines, with nearly 500,000 data points for the entire 1800 line-km survey. Here we focus on a subset of about half of the total data sets that were collected along closely spaced flight lines that cover an area of approximately 300 km² (Figure 1). At every data location, the AEM data were inverted to recover a one-dimensional model of electrical resistivity in the subsurface by using the EM1DFM inversion algorithm [Farquharson et al., 2003]. The total depth of investigation, up to 130 m in some areas, was evaluated by using the method of Oldenburg and Li [1999]. These inverted resistivity models provide the main input to the semiautomated interpretation steps described below.

2.3. K-Means Clustering

K-means clustering is an unsupervised machine learning technique used to group a data set into a predefined number of clusters, K . In this study the clustering algorithm is used to sort the zones of unfrozen sediments from the regions with permafrost; therefore, the predefined number of clusters is $K = 2$. The data set in this case constitutes 243555 1D resistivity models derived from the inverted AEM data [Ball et al., 2011; Minsley et al., 2012], which are represented by a “point” in a high-dimensional model space. The algorithm’s job is to sort these points into frozen and unfrozen classes. The iterative procedure starts out by choosing two random samples from the high-dimensional input model space to represent the centers in each of the clusters. All the other points are then assigned to the cluster they are closest to (the Euclidean distance is used). Now the mean value of each cluster is computed and will represent the new cluster-centers. Again, all points are reassigned a cluster based on the distance measure and the new centers. This procedure continues as an iterative process until the clusters converge. Mathematically, the K-means clustering algorithm seeks to minimize the objective function:

Figure 2. Map of AEM flight lines displayed with surface geology [Williams, 1962].

$$J = \sum_{n=1}^N \sum_{k=1}^K r_{nk} \|x_n - \mu_k\|^2$$

where N is the number of points, K is the number of clusters, r_{nk} is an indicator variable being either 1 or 0, indicating whether point n , located at position x_n , belongs to cluster k or not, respectively. Parameter μ_k represents the cluster center of cluster k . The steps of the iterative process explained above are steps in minimizing the objective function, and the process converges when a minimum of the function is reached.

2.4. Smart Interpretation

Smart Interpretation is based on a linear regression technique and is a method developed for geological layer modeling [Gulbrandsen *et al.*, 2015]. Here we make use of this method extended from linear to second-order polynomial regression. The method is based on two steps. In the first step an expert interprets a physical interface such as the boundary between two lithological units, or the base of permafrost, by “picking” this surface on a cross section of inverted AEM data. While the expert is doing this manual picking the algorithm tries to learn the relation between the manually interpreted points and multiple available attributes. Specifically, the relationship is defined by the coefficients of a polynomial that are determined through linear regression, and attributes include quantifiable information such as resistivity values, elevation, geographic coordinates, and terrain. After a few picks, when the algorithm has learned the interpretation-attribute relation, the second part of the algorithm is to predict what the expert would interpret throughout the whole survey, which is simply done by using the relation learned in the first step. In this study the SI is used to interpret the permafrost thickness (by picking the base of frozen silt) of the block within Yukon Flats, close to Fort Yukon, Alaska. The attributes used in this study are the resistivity values, the terrain, and an extracted attribute representing the depth where the resistivity values cross 100 Ω m.

3. Results and Comparison

In Figure 3, the result from using the clustering procedure on one cross section is shown. The black dashed line in Figure 3a shows the distribution of permafrost according to the manual interpretation by Minsley *et al.*

Figure 3. (a) The resistivity profile B from Figure 1 (taken from Minsley *et al.* [2012]). The black and red dashed lines represent the base of permafrost and the gravel-silt contact, respectively, as interpreted by Minsley *et al.* [2012]. (b) The same profile as Figure 3a where all resistivity layers below depth of investigation are removed. (c) The K-means clustering predicted taliks. (d) The K-means clustering predicted permafrost. Coordinates are UTM Zone 6N.

[2012]. Where this line vanishes, the soil is assumed to be fully unfrozen throughout the entire vertical extent. Figures 3c and 3d show the models that represent taliks and permafrost, respectively, as predicted by the K-means clustering algorithm. It is seen that this automated approach is consistent with the manual interpretation. From the figure it is clear that the subsurface below both Twelvemile Lake and the Yukon River is interpreted to be unfrozen. This is consistent with the concept that the subsurface does not freeze beneath large lakes and rivers that are too deep to fully freeze during winter. Figure 4 shows a map of the spatial distribution of permafrost zones resulting from the K-means clustering, in the whole study area displayed together with all the lakes in the area. The figure shows that the results produced by the clustering algorithm are consistent with the distribution of lakes, which is a strong indication that the clustering algorithm performs satisfactorily. Note that the figure only displays the outlines of lakes, and not the Yukon River, its side channel, or the Porcupine River in the northwest part of the block. It should also be noted that there are no data of lakes from the northeast part of the survey.

There are several different unsupervised clustering techniques that probably also would be able to achieve the results achieved by the K-means clustering algorithm in this study. No other methods have been tested, since the K-means clustering method provides the desired results and is easy to implement. If, however, the geology of the study area was more complex (e.g., different frozen/unfrozen resistivity thresholds throughout the study area) more sophisticated clustering methods might be required.

Next, the taliks are sorted from the permafrost regions, and the result of the automated base of permafrost interpretation on cross section B from Figure 1 is shown on Figure 5b. The black dots (seven in total) represent the manual interpretation points that constitute the first step of the SI method. Based on the learned relation between these points and the attributes, the SI method predicts the red line to represent the base of the permafrost, which is very consistent with the manual interpretation shown in Figure 5a. The main benefit of using SI is that it allows a quantitative approach to interpretation that can take all data into account, while

Figure 4. Distribution of permafrost in the study area, as resulting from the K-means clustering algorithm together with the distribution of lakes in the area (white polygons) [Rover *et al.*, 2012]. The colors represent the surface resistivity values of the regions assumed to have permafrost, and the black color is assigned zones of unfrozen soil. Coordinates are UTM Zone 6N.

being trained by an expert. In practice, this allows much faster, and less qualitative, interpretation than the manual approach. In Figure 5b, the models indicating taliks are removed, and it is seen that these have the same location as where the black line in Figure 5a disappears, i.e., where the soil is interpreted to be completely unfrozen. In the middle of the cross section (about 574 km) Figure 5a, has a question mark, indicating the uncertainty in the manual interpretation as to whether this area is frozen or not. Even though the automated approach indicates that this region most likely is unfrozen, this uncertainty is also expressed by leaving a few models as frozen. Figures 5c and 5d show the interpreted base of permafrost on the AEM flight line highlighted with letter **A** in Figure 1 (from Minsley *et al.* [2012] and the SI method, respectively). The prediction in Figure 5d is made by using the interpretation-attribute relation learned on the cross section in Figure 5b.

By comparing the interpretations of the base of permafrost on the two cross sections displayed in Figure 5, it can be concluded that the semiautomatic approach provides objective results that are consistent with the manual interpretation. The SI method can be applied to all of the flight lines in the survey by using the interpretation-attribute relation learned from the seven interpretation points shown in Figure 5b. This allows for a prediction of the depth to the base of permafrost to be estimated across an entire area (see Figure 6). From the figure, it is clear that the permafrost is thickest in the southwest part of the block (100–110 m thick), thins towards northeast and the Yukon River, and thickens again on the other side of the river. This behavior is consistent with the results from the study by Minsley *et al.* [2012].

The detailed map of permafrost extent and thickness (Figure 6), which to our knowledge is the first of its kind, is an important data product that may be of use for numerous other studies. This result could for instance be used to compute the storage, and hence potential release, of CO₂ in the area, and in that way play a role in state of the art climate research. Results could also be interpreted together with maps of historical fire events and other remote sensing or geophysical data sets to better understand the impact of fire on permafrost distributions [Nossov *et al.*, 2013; Minsley *et al.*, 2016]. Knowledge of permafrost thickness and extent is also valuable for hydrogeological modeling [Walvoord *et al.*, 2012] and investigations of the role of subsurface flow on observed changes in lake area extent [Jepsen *et al.*, 2013; Chen *et al.*, 2014].

The workflow presented here not only gives the opportunity to achieve a reliable map of permafrost in a much faster way than previously available, but also opens up for the possibility of more frequent mapping of permafrost using AEM data collected over the same area. This would give rise to a completely new insight in the actual *change* in the distribution of permafrost, and not only a snapshot of the present state, potentially contributing to a much better understanding of the dynamics of permafrost, and the effects of global warming.

Figure 5. (a) The base of permafrost (black dashed line) as interpreted by Minsley *et al.* [2012]. (b) The interpretation points (black points) and the resulting predicted base of permafrost (the red line). The cross section represents the AEM flight line 10142 and highlighted with letter B (southwest of Yukon Flats) in Figure 1. (c) The manually interpreted base of permafrost along flight line 10020 (highlighted with letter A in Figure 1) from Minsley *et al.* [2012]. (d) The same cross section as Figure 5c with the predicted base of permafrost using SI. Coordinates are UTM Zone 6N.

Figure 6. The final map of permafrost distribution and configuration, where the colors represent the thickness of the permafrost. Coordinates are UTM Zone 6N.

4. Conclusions

In this study we demonstrate a workflow for fast and accurate mapping of the distribution of permafrost, in an area of discontinuous permafrost that is rich with taliks, based on an AEM survey. The workflow is based on a combination of manual picking of the base of permafrost and two machine-learning techniques: K-means clustering and Smart Interpretation. This combination ensures that all data are included in the mapping procedure and that the resulting map is both consistent with the geophysical data and the interpretations made by the expert. Although the manual interpretations require expert input, the overall result provides consistency between the final map, the data, and the expert, and the semiautomatic nature of this process, allowing fast results, suggests that this workflow provides a quantitative, fast, and useful approach to mapping of permafrost.

Acknowledgments

This study is partly financed by the Danish High Technology Foundation (journal number: 113-2013-1) and partly by the University of Copenhagen, and as far as we understand there is no potential conflict of interest. The AEM data used in this study can be accessed at <http://pubs.usgs.gov/of/2011/1304/> [Ball et al., 2011].

References

- Avis, C. A., A. J. Weaver, and K. J. Meissner (2011), Reduction in areal extent of high-latitude wetlands in response to permafrost thaw, *Nat. Geosci.*, 4(7), 444–448, doi:10.1038/ngeo1160.
- Ball, L. B., Smith, B. D., Minsley, B. J., Abraham, J. D., Voss, C. I., Astley, B. N., Deszcz-Pan, M., and Cannia, J. C., 2011, Airborne electromagnetic and magnetic geophysical survey data of the Yukon Flats and Fort Wainwright areas, central Alaska, June 2010: U.S. Geol. Surv. Open-File Rep. 2011–1304, 21 pp.
- Bense, V. F., G. Ferguson, and H. Kooi (2009), Evolution of shallow groundwater flow systems in areas of degrading permafrost, *Geophys. Res. Lett.*, 36, L22401, doi:10.1029/2009GL039225.
- Bishop, C. M. (2006), *Pattern Recognition and Machine Learning*, Springer, New York.
- Chen, M., J. C. Rowland, C. J. Wilson, G. L. Altmann, and S. P. Brumby (2014), Temporal and spatial pattern of thermokarst lake area changes at Yukon Flats, Alaska, *Hydrol. Process.*, 28(3), 837–852, doi:10.1002/hyp.9642.
- Clark, A., C. E. Barker, and E. P. Weeks (2009), Drilling and testing the DOI-04-1A coalbed methane well, Fort Yukon, Alaska, U.S. Geol. Surv. Open File Rep., 2009–1064, 69 pp.
- Everitt, B. S., S. Landau, and M. Leese (2001), *Cluster Analysis*, Arnold, London.
- Farquharson, C. G., D. W. Oldenburg, and P. S. Routh (2003), Simultaneous 1D inversion of loop-loop electromagnetic data for magnetic susceptibility and electrical conductivity, *Geophysics*, 68(6), 1857–1869, doi:10.1190/1.1635038.
- Froese, D. G., D. G. Smith, and D. T. Clement (2005), Characterizing large river history with shallow geophysics: Middle Yukon River, Yukon Territory and Alaska, *Geomorphology*, 67(3–4), 391–406, doi:10.1016/j.geomorph.2004.11.011.
- Froese, D. G., J. A. Westgate, A. V. Reyes, R. J. Enkin, and S. J. Preece (2008), Ancient permafrost and a future, warmer Arctic, *Science*, 321(5896), 1648, doi:10.1126/science.1157525.
- Gulbrandsen, M. L., T. Bach, K. S. Cordua, and T. M. Hansen (2015), *Localized Smart Interpretation—A Data Driven Semi-automatic Geological Modeling Method*, ASEG-PESA, Perth, Australia.
- Hoekstra, P., P. V. Sellmann, and A. Delaney (1975), Ground and airborne resistivity surveys of permafrost near Fairbanks, Alaska, *Geophysics*, 40(4), 641–656, doi:10.1190/1.1440555.
- Jepsen, S. M., C. I. Voss, M. A. Walvoord, B. J. Minsley, and J. Rover (2013), Linkages between lake shrinkage/expansion and sublacustrine permafrost distribution determined from remote sensing of interior Alaska, USA, *Geophys. Res. Lett.*, 40, 882–887, doi:10.1002/grl.50187.
- Knorr, W., I. C. Prentice, J. I. House, and E. A. Holland (2005), Long-term sensitivity of soil carbon turnover to warming, *Nature*, 433, 298–301, doi:10.1038/nature03226.
- Koven, C. D., B. Ringeval, P. Friedlingstein, P. Ciais, P. Cadule, D. Khvorostyanov, G. Krinner, and C. Tarnocai (2011), Permafrost carbon-climate feedbacks accelerate global warming, *Proc. Natl. Acad. Sci. U.S.A.*, 108(36), 14,769–14,774, doi:10.1073/pnas.1103910108.
- Mack, M. C., M. S. Bret-Harte, T. N. Hollingsworth, R. R. Jandt, E. A. G. Schuur, G. R. Shaver, and D. L. Verbyla (2011), Carbon loss from an unprecedented Arctic tundra wildfire, *Nature*, 475(7357), 489–492, doi:10.1038/nature10283.
- Minsley, B. J., et al. (2012), Airborne electromagnetic imaging of discontinuous permafrost, *Geophys Res Lett*, 39, L02503, doi:10.1029/2011GL050079.
- Minsley, B. J., N. J. Pastick, B. K. Wylie, D. R. N. Brown, and M. Andy Kass (2016), Evidence for nonuniform permafrost degradation after fire in boreal landscapes, *J. Geophys. Res. Earth Surf.*, 121, 320–335, doi:10.1002/2015JF003781.
- Nossov, D. R., M. T. Jorgenson, K. Kielland, and M. Z. Kanevskiy (2013), Edaphic and microclimatic controls over permafrost response to fire in interior Alaska, *Environ. Res. Lett.*, 8(3), 035013, doi:10.1088/1748-9326/8/3/035013.
- O'Donnell, J. A., J. W. Harden, A. D. McGuire, M. Z. Kanevskiy, M. T. Jorgenson, and X. Xu (2011), The effect of fire and permafrost interactions on soil carbon accumulation in an upland black spruce ecosystem of interior Alaska: Implications for post-thaw carbon loss, *Global Change Biol.*, 17, 1461–1474, doi:10.1111/j.1365-2486.2010.02358.x.
- Oldenburg, D. W., and Y. Li (1999), Estimating depth of investigation in dc resistivity and IP surveys, *Geophysics*, 64(2), 403–416, doi:10.1190/1.1444545.
- Rover, J., L. Ji, B. K. Wylie, and L. L. Tieszen (2012), Establishing water body areal extent trends in interior Alaska from multi-temporal Landsat data, *Remote Sens. Lett.*, 3(7), 595–604, doi:10.1080/01431161.2011.643507.
- Schuur, E. A. G., J. G. Vogel, K. G. Crummer, H. Lee, J. O. Sickman, and T. E. Osterkamp (2009), The effect of permafrost thaw on old carbon release and net carbon exchange from tundra, *Nature*, 459(7246), 556–559, doi:10.1038/nature08031.
- Schuur, E. A. G., et al. (2015), Climate change and the permafrost carbon feedback, *Nature*, 520(7546), 171–179, doi:10.1038/nature14338.
- Vaughan, D. G., et al. (2013), Observations: Cryosphere, in *Climate Change 2013: The Physical Science Basis, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T. F. Stocker et al., pp. 317–382, Cambridge Univ. Press, Cambridge, U. K., and New York.
- Walvoord, M. A., and R. G. Striegl (2007), Increased groundwater to stream discharge from permafrost thawing in the Yukon River basin: Potential impacts on lateral export of carbon and nitrogen, *Geophys. Res. Lett.*, 34, L12402, doi:10.1029/2007GL030216.
- Walvoord, M. A., C. I. Voss, and T. P. Wellman (2012), Influence of permafrost distribution on groundwater flow in the context of climate-driven permafrost thaw: Example from Yukon Flats Basin, Alaska, United States, *Water Resour. Res.*, 48, W07524, doi:10.1029/2011WR011595.
- Williams, J. R. (1962), Geologic reconnaissance of the Yukon Flats District Alaska, U.S. Geol. Surv. Bull., 1111-H, pp. 289–331.