Using remote sensing and ancillary data to extend airborne electromagnetic resistivity surveys for regional permafrost interpretation.

Neal J. Pastick[1]*, Bruce K. Wylie[2], Burke J. Minsley[3], Torre Jorgenson[4], Lei Ji[5], Michelle A. Walvoord[3], Bruce Smith[3], Jared Abraham[3], and Joshua R. Rose[6]

[1]Stinger Ghaffarian Technologies, Inc., contractor to the U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center, Sioux Falls, SD 57198, USA. Work performed under USGS contract G10PC00044. [2]USGS EROS Center, Sioux Falls, SD 57198. [3]USGS, Denver, CO 80225. [4]Alaska Ecoscience, Fairbanks, AK 99709. [5]ASRC Research & Technology Solutions, contractor to the USGS EROS Center, Sioux Falls, SD 57198, USA. Work performed under USGS contract 08HQCN0007. [6]Yukon Flats National Wildlife Refuge, Fairbanks, AK 99701. *Corresponding author Tel: 605 594 6078; E-mail: njpastick@usgs.gov

C41B-0390

I. Introduction:

Permafrost has a significant impact on high latitude ecosystems and is spatially heterogeneous. However, only generalized maps of permafrost extent are available. Due to its impacts on carbon pools, subsurface hydrology, lake water levels, vegetation communities, and surface soil deformations, an understanding of spatial extents and depth of permafrost are critical for proper management and monitoring of these areas.

In this study, we propose a method for accurately extrapolating Airborne Electromagnetic Resistivity (AEM) for regional permafrost mapping in the Yukon Flats Ecoregion (YFE), Alaska, through the use of regression tree models. Electrical resistivity serves as the proxy for permafrost presence in the AEM extrapolation portion of this study, as electrical resistivity increases dramatically as soil freezes. This method uses resistivity values, and other relevant data to predict near surface (0-2.6m) electrical resistivity at a 30-m resolution within the YFE.

We also propose a piecewise regression model (Cubist) and a Presence/Absence active layer decision tree classification (See5) that use in-situ data and other relevant spatial data, to accurately estimate Active Layer Thickness (ALT) or thaw depth (0-122cm) at a 30-m resolution within the YFE.

II. Study Site:

The YFE is located approximately 100 miles north of Fairbanks, Alaska, and encompasses an area of 33,400 km². The topography of YFE consists mostly of a flat relatively low lying center with an anterior characterized by steep elevations (**Figure 1.1**). Throughout the YFE permafrost presence has been known to be widespread, but is discontinuous and of variable thickness (Williams, J.R. 1962).

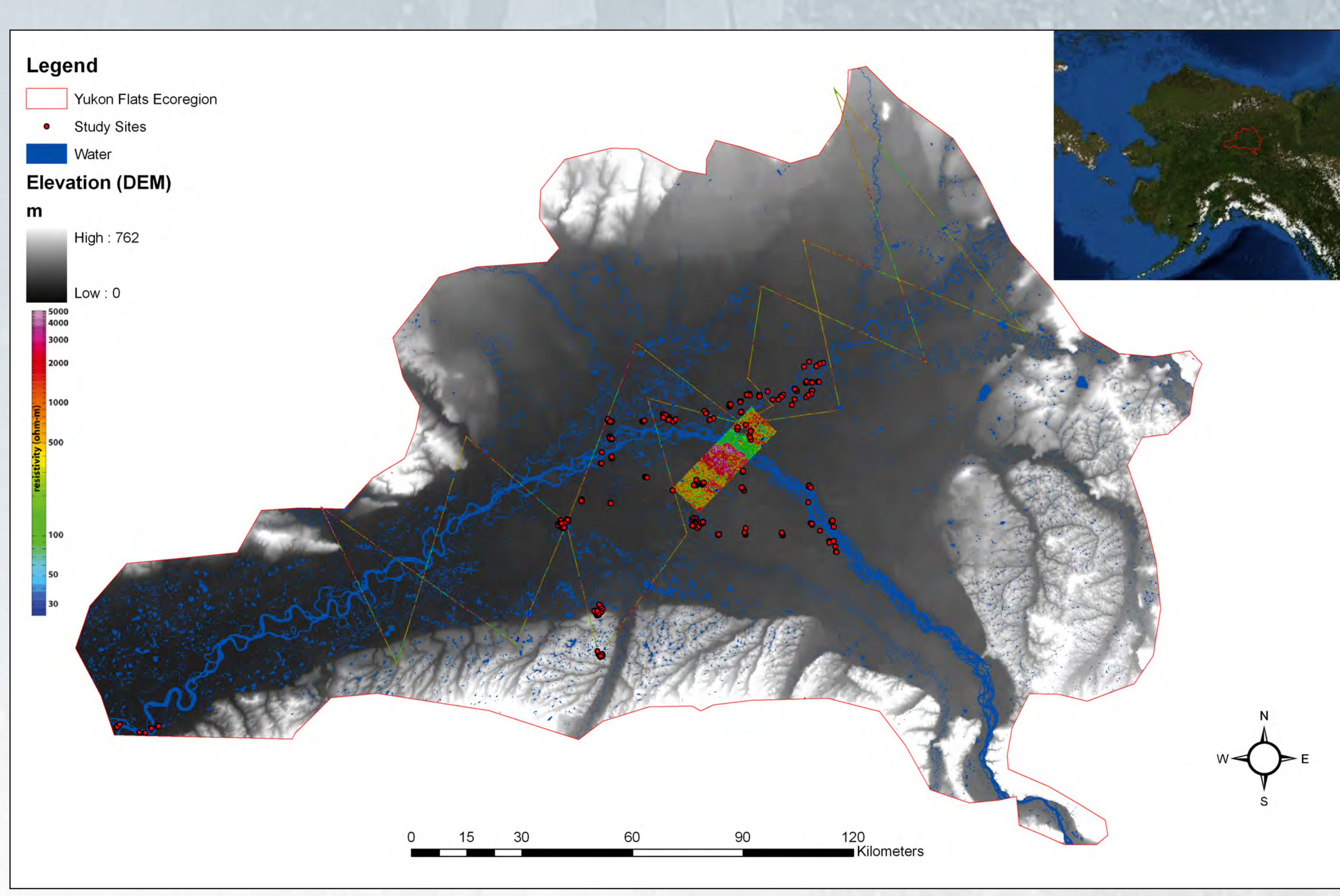


Figure 1.1 The Yukon Flats Ecoregion overlaid with NLCD 2001 (Water), AEM Flight Line Resistivity (0-2.6m), and a DEM (30m).

III. Data:

Airborne Electromagnetic Survey

The AEM survey was flown in June of 2010, over a portion of the YFE with measurements recorded in reconnaissance lines and a contiguous block area of coverage (**Figure 1.1**). Subsurface resistivity models were derived by inverting the AEM data (Abraham, Jared 2011). Inverted AEM survey values between 0 and 2.6 meters were selected as the dependent variable for various regression tree models that extrapolate resistivity, if they meet the criteria discussed below.

Landsat TM Mosaic

A Landsat image mosaic of the YFE was developed from 6 Landsat scenes taken from late August 2008 through early September 2008 (Ji et al. 2010).The averaged coefficient of variation values in a 3x3 window for Landsat Bands 3, 4, 5 were then calculated and served as a criterion for spatially homogeneous AEM values and in-situ data used for model development. Homogenous areas were chosen because study sites that contain homogeneity at a scale of several pixels of the selected sensor may be simply compared with satellite observations (Liang et al. 2002).

In-situ Data

Field observations used within this study were conducted during mid August to early September of the years 2009-2010. One set of observations involved using a 122cm probe to quantify ALT (n=5/site) within each sub-transect (30m). ALT measurements within each sub-transect were averaged to provide an estimate of overall site ALT.

The second set of observations used a 200cm probe and soil pits to quantify ALT (n=1/site). All field observations were combined in a spatial database and a value of 200cm was given to all active layer measurements (=0cm & >122cm) for model consistency. These field observations served as dependent variables within the regression tree and decision tree models that estimated ALT in the YFE.

Other Data/Model Input Variables (Table 1.1)

Table 1.1 Input variables, description, and % averaged usage within each model.

Input Variable	Description*		% Usage - AEM Extrapolation	Depth Extrapolation (<=122cm)	Presence/Absence of Active Layer Depth
NDII	Normalized difference infrared index (spectral indices)	$NDII = \frac{b4-b5}{b4+b5}$	0.739		0.22
Band 5	Landsat Band 5 (6-scence Landsat TM mosaic)		0.598		0.91
Soil Moisture	Estimated based on field measurements (22 samples), spectral and topographic variables using regression method		0.485		1
NDII7	Normalized difference infrared index (spectral indices)	$NDII7 = \frac{b4-b7}{b4+b7}$	0.47	0.58	1
NDWI	Normalized difference water index (spectral indices)	$NDWI = \frac{b2-b5}{b2+b5}$	0.44		0.59
Band 4	Landsat Band 4 (6-scence Landsat TM mosaic)		0.5		0.99
NDVI	Normalized difference vegetation index	$NDVI = \frac{b4-b3}{b4+b3}$	0.431		
NDWI7	Normalized difference water index (spectral indices)	$NDWI7 = \frac{b2-b7}{b2+b7}$	0.451	0.58	0.31
Band 6	Landsat Band 6 (6-scence Landsat TM mosaic)		0.474		1
GNDVI	Green normalize difference vegetation index (spectral indices)	$GNDVI = \frac{b4-b2}{b4+b2}$	0.438		0.65
Band 3	Landsat Band 3 (6-scence Landsat TM mosaic)		0.441		0.83
Biomass	30-m resolution aboveground biomass (AGB) dataset for the Yukon River Basin of Alaska and Canada using Landsat data and field observations acquired in recent years (Li *et al.* 2010)		0.423		0.9
SAVI	Soil adjusted vegetation (spectral indices). (L = 0.5 in this study)	$SAVI = \frac{(1+L)(b4-b3)}{b4+b3+L}$	0.431		0.79
EVI	Enhanced vegetation index (spectral indices) (G = 2.5, L= 1, C1 = 6, C2= 7.5)	$EVI = \frac{G(b4-b3)}{(b4+C_1b3-C_2b1)+L}$	0.415	0.58	0.46
DEM	30m digital elevation model		0.801	0.96	0.99
Band 2	Landsat Band 2 (6-scence Landsat TM mosaic)		0.412		1
Averaged Weighted Performance Anomalies	Reflects ecological changes that are caused by factors other than climate or site potential (Wylie *et al.* 2008)		0.747	0.34	0.8
CTI	Compound Topographic Index = f(slope & upstream contributing area)		0.253		0.39
Band 1	Landsat Band 1 (6-scence Landsat TM mosaic)		0.192		0.67
LST	Land surface temperature (Band 7 from 6-scene Landsat TM mosiac)		0.504	1	0.88
Yfslope	A long-baseline slope analysis (7.7 kilometers) which utilized topographic analysis tools to derive a flow-direction image from the underlying DEM. Slope was calculated between a given cell and the location 128 cells downstream. (Bliss, Norman. Personal Communication. 05 Nov 2011)		0.132	0.71	1
NLCD (2001) + Fire data (MTBS)	NLCD 2001 was used as a categorical variable within this study. Areas containing recent fires were removed (2008-2010) & Fires from 2000-2007 were give a unique Landcover value.		0.252		1
Geology Map	General Surface Geology Map (Williams, J.R., 1962)		0.356		0.73
In-Situ Data (Active Layer Depth Measurements)	See "*In-situ Data*" Section.			Served as dependent variable	Served as dependent variable
Airbone Electromagnetic Survey (LN(0-2.6m)*10)	See "*Airborne Electromagnetic Survey*" Section.		Served as dependent variable		
Final Resistivity (0-2.6m) Extrapolated Map	5 model averaged image (R^2 = 0.7132)				0.9

*In spectral indicies equations, b1, b2, etc. are the reflecatnces in Landsat TM bands 1, 2 etc.

IV.Methods/Modeling:

Resistivity-

- AEM values from homogenous areas were randomly split into a training set for model development and test data set.
- A total of three, 5-member committee regression tree models generated (CV <= 0.04), used a training data set (n = 8,848) and test data set (n = 988).
- A total of two, 5-member committee regression tree models generated (CV <= 0.08), used a training data set (n = 20,471) and test data set (n = 2,179).
- The output maps produced by the initial AEM extrapolation models were then compared to ensure accuracy, as discussed below in the results and validation section.

Active Layer Thickness & Presence/ Absence of Active Layer-

- ALT measurements (<=122cm) taken from homogenous areas (CV <= 0.10), served as the dependent variable within the ALT piecewise regression model. (n=99)
- A single, 3-rule regression tree model with use of instances was generated to estimate ALT measurements (0-122cm).
- ALT measurements (n=377) taken from homogenous and heterogeneous areas, served as the dependent variable within the Presence/Absence Active Layer decision tree model.
- A Presence/Absence Active Layer classification model was then generated and served to map ALT measurements >122cm.

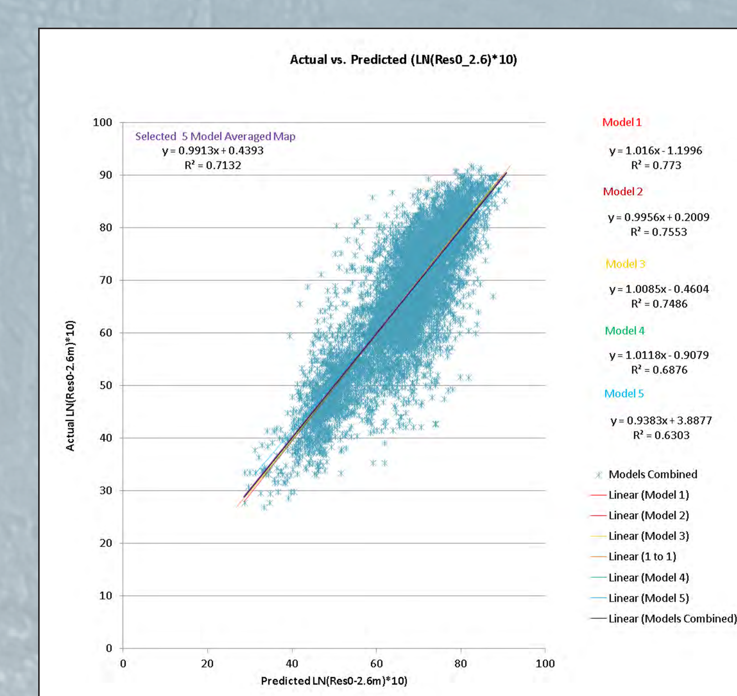


Figure 1.2 Actual vs. Predicted (LN(0-2.6m)*10) Resistivity Values for the 5 models generated. Natural logs of resistivity were used within the database, as relationships of resistivity are often distributed log-normally.

V. Results and Validation

Resistivity-

- In order to ensure model accuracies, all output images were compared to the original AEM flight lines.
  - The Mean Average Difference (MAD) values were calculated for all resistivity models/images.
  - The model/map with the lowest MAD (lowest difference with all AEM pixels) was chosen to serve as the final resistivity 0-2.6m extrapolation map.
- Once the final 5 model averaged image (MAD = 586 ohm-m) (**Figure 1.2 & Figure 1.3**) was selected, the standard deviation and mean of all images, at each pixel, were calculated so a coefficient of variation (uncertainty map) could be produced (**Insert in Figure 1.3**).
  - Important drivers for the model development included: DEM, Averaged Weighted Performance Anomalies, NDII, and others (**Table 1.1**).
  - Mean coefficient of variation values were also calculated by land cover (**Table 1.2**).

Active Layer Thickness & Presence/ Absence of Active Layer –

- A single 3-rule regression tree model (R^2 = .89) was generated for ALT extrapolation (<=122cm).
  - Important drivers for model development included: DEM, NDWI7, NDII7, and others (**Table 1.1**).
- In order to ensure accuracy and robustness of the model developed for ALT (<=122cm) a 10- fold cross validation was preformed due to the limited training dataset (n=99).
  - The estimated average results, on the hold-out cases of the 3-rule model, indicated an R^2 of 0.73.
- A Presence/Absence ALT (>122) decision tree classification (R^2 = 0.984) was then used to map the locations ALT values >122cm.
  - Important drivers for model development included: DEM, Soil Moisture, Band 6, Final Resistivity (0-2.6m) Extrapolated Map, and others (**Table 1.1**).
  - A 10-fold cross validation indicated an R^2 of 0.80 on hold-out cases.
- The final ALT (0-122cm) map produced from the combined 3-rule model and the presence/absence classification is shown in **Figure 1.4**.
- The map in **Figure 1.4** was then compared to field observations taken within the study area from the months of August and September (2010), to insure overall image accuracy (**Figure 1.5**).
  - Error metric calculations show that bias errors (MBE = -4.13 cm and rMBE = -3.79%) and absolute errors (MAE = 8.35 cm and rMAE = .02 %) are low.

Figure 1.3 The averaged 5 model/image resistivity (0-2.6m) extrapolation and uncertainty maps.

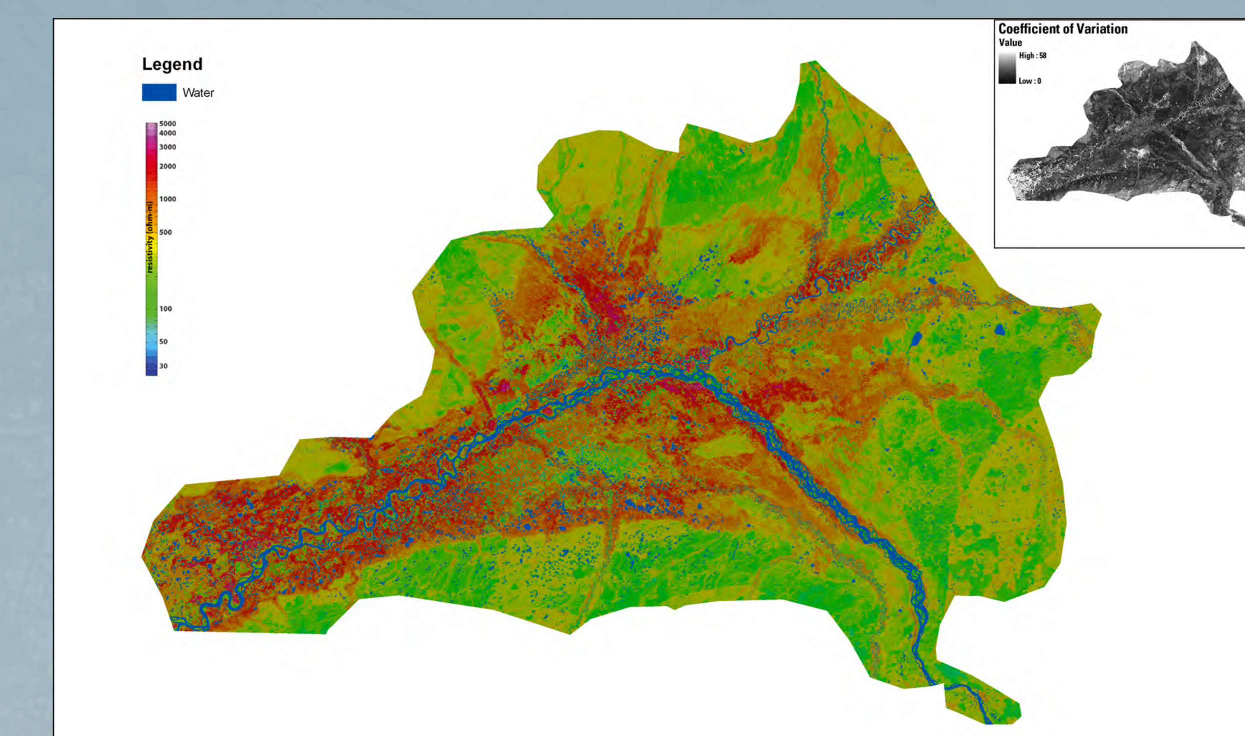


Figure 1.4 Estimated Active Layer Thickness, in the Yukon Flats. (Combined 3-Rule model with use of instances (<=122cm) & Presence/Absence Active Layer Decision Tree Model (>122cm)).

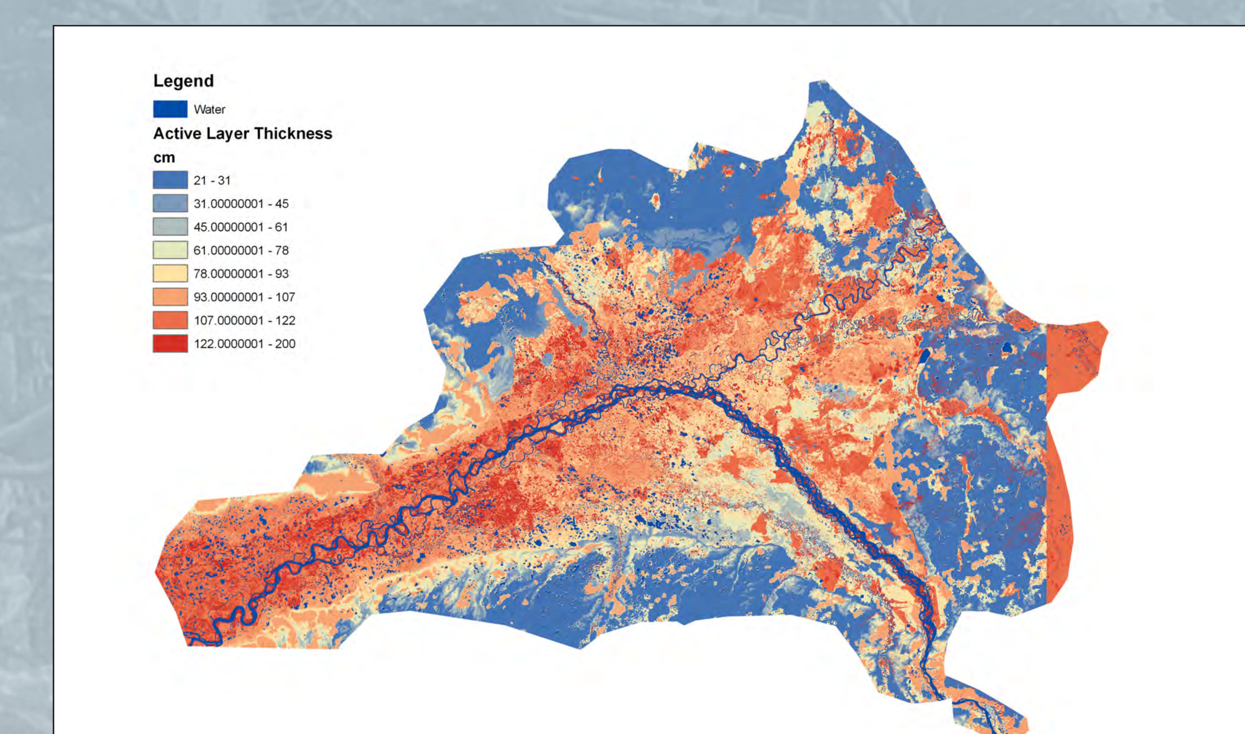
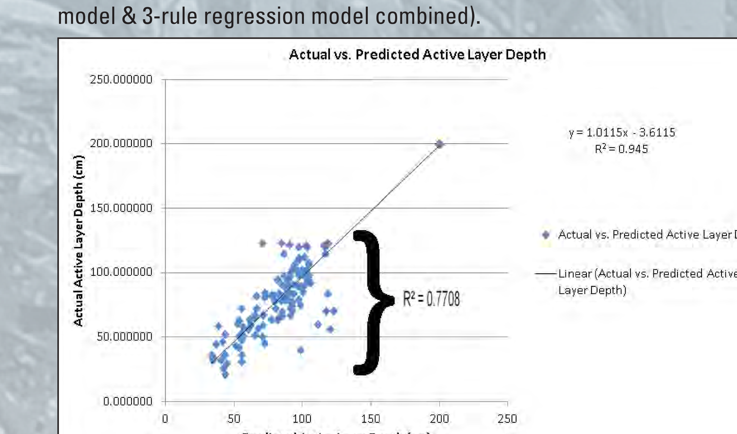


Table 1.2 Mean coefficient of variation and mean ALT by Land Cover.

VALUE	COUNT	Land_Cover	Mean_CV	Mean_ALT
2	8574205	2000-2007 Fires	5.73	60.99
11	2199587	Open Water	9.33	111.67
22	3706	Developed, Low Intensity	8.38	92.16
23	335	Developed, Medium Intensity	6.31	104.15
31	67161	Barren Land	8.84	112.20
41	6270890	Deciduous Forest	4.35	91.20
42	9243256	Evergreen Forest	4.62	76.18
43	2700619	Mixed Forest	4.81	76.05
51	40319	Dwarf Shrub	5.18	92.65
52	3299397	Shrub/Scrub	5.33	69.65
71	7580	Grassland/Herbaceous	3.95	46.62
72	19160	Sedge/Herbaceous	6.55	59.36
90	4047269	Woody Wetlands	5.34	87.95
95	640016	Emergent Herbaceous Wetlands	5.09	109.52

Figure 1.5 Actual Active Layer Thickness Measurements vs. Predicted Active Layer Thickness Measurements (Presence/Absence decision tree model & 3-rule regression model combined).



- The R^2 value is slightly inflated because ALT = 0 & >122cm were given the value of 200.
- Mean ALT values were then calculated by land cover. **(Table 1.2)**
  - Evergreen Forests (Black Spruce) are comprised of thinner active layers then Deciduous Forests.
  - Wetlands are comprised of thick active layers.

VI. Conclusion

We mapped near-surface (0-2.6m) resistivity and Active Layer Thickness (0-122cm) at a 30-m resolution for the Yukon Flats ecoregion using an Airborne Electromagnetic Survey, in-situ data, Landsat data, and other relevant spatial data. Accuracy assessment of the near-surface (0-2.6m) extrapolated resistivity map indicated a MAD of 586 ohm-m. Accuracy assessment of the ALT (0-122cm) map indicated that the rMAE and rMBE were 0.02% and -3.79%, respectively.

VII. Acknowledgements

This study is funded by the U.S. Geological Survey: Climate Land Use Research and Development Program. We thank all of those that have contributed to this project: Mark Waldrop (USGS), Kurtis Nelson (USGS), Jack McFarland (USGS), Birgit Peterson (ASRC), Dana Nossov (UAF), Mark Winterstein (UAF), and Jim Webster (Bush Pilot).

Reference:

Abraham, Jared, 2011, A promising tool for subsurface permafrost mapping—An application of airborne geophysics from the Yukon River Basin, Alaska: U. S. Geological Survey, Fact Sheet 2011-3133, 4 p.
Cubist and See5. *Rule Quest Research data mining tools* from http://www.rulequest.com on 20 May 2011.
Ji et al. 2010, Mapping aboveground biomass for interior Alaska using Landsat data and field measurements. http://lca.usgs.gov/lca/yukonflats_biomass/index.php on 01 Nov 2011.
Liang et al. 2002, Validating MODIS land surface products: methods and preliminary results. Remote Sensing of Environment 83:149-162.
Williams, J.R., 1962, Geologic reconnaissance of the Yukon Flats district, Alaska: U.S. Geological Survey Bulletin 1111-H, p. 289-331.
Wylie et al. 2008, Integrating modeling and remote sensing to identify ecosystem performance anomalies in the boreal forest, Yukon River Basin, Alaska. International Journal of Digital Earth, Vol. 1, No. 2, June 2008, 196-220.