Brian Hall and David Foster
8/2/2012

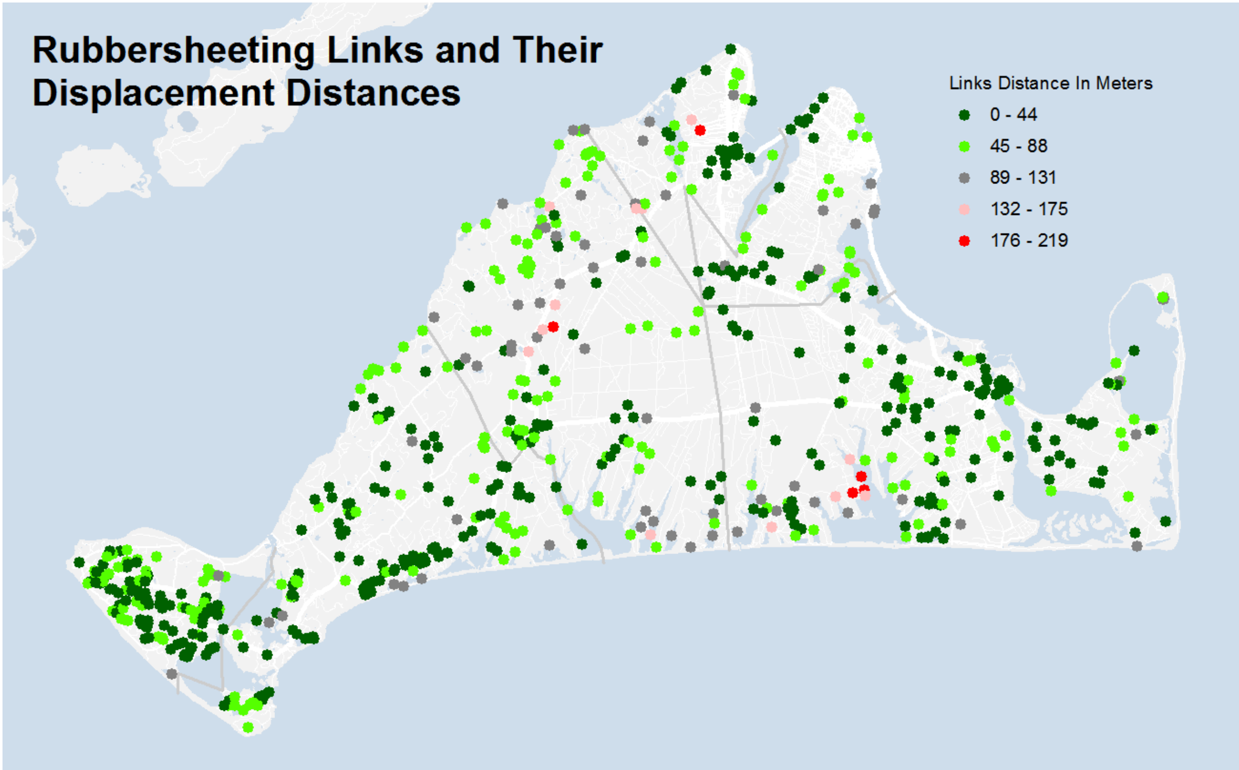
Methods used to digitize US Coastal Survey 1848 data at the Harvard Forest

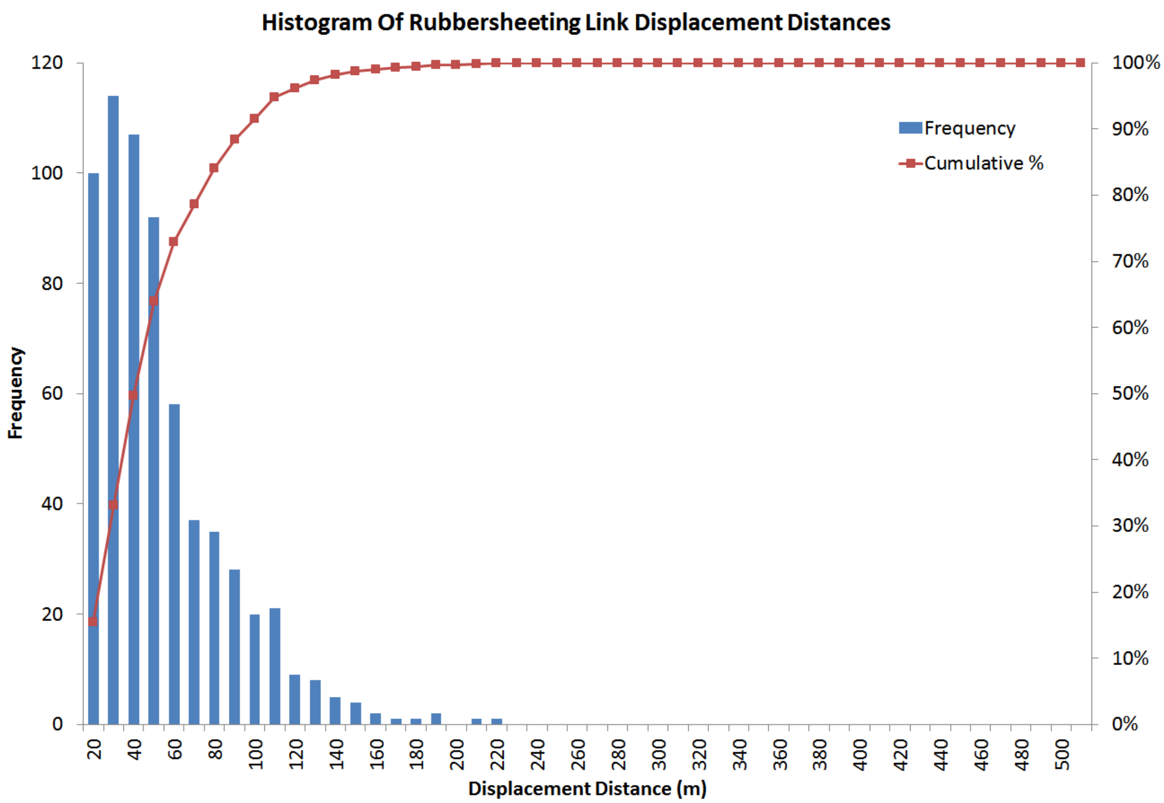
1) In 1998, Brian Hall (BH) and Ben Slater got photocopies of US Coastal Survey maps from copies housed in the National Archives in College Park, Maryland for the region between New York City to the north shore of Sandwich, MA.
2) BH colored in many of the features on the map photocopies to ease interpretation while digitizing and to leave a record of what features were digitized. These copies are housed in the Harvard Forest archives. The features were digitized with a digitizing table into State Plane, Massachusetts Mainland, NAD 83 projection using georeferencing "tics" from road intersections that are still existing and readily identifiable on a modern GIS road layer from MassGIS. There were approximately 6-10 tics per Coastal Survey map sheet; these tics and their real-world coordinates were marked on the copies archived at Harvard Forest. The digitized layers were then compared to the photocopies of original sheets to reduce errors of ommission.
3) Spatial inaccuracy can be introduced into a multi-step process such as this in a number of ways. Errors can be introduced during:
    a. the original field surveying,
    b. transcribing field notes to field maps,
    c. transcribing field maps to lithographic plates for original publication,
    d. photocopying which can distort images - BH did a test on letter-sized paper and found that lines can be skewed by as much as 0.125 inches which can translate to a very large distance on a scale map,
    e. selecting georeferencing tics (mismatching road junctions on the USCS and modern map, etc.)
    f. digitizing
    g. warping from the rubbersheeting process (can greatly improve areas but can also create more error in others)
    h. the selection of "true" locations on the modern GIS layers since the modern layers may not be accurate
4) In 2011, BH and David Foster began to use the data at a much more local scale than had been previously done. It was decided to improve the spatial accuracy of the USCS layers by rubbersheeting them to better fit the same feature as shown on modern GIS layers and georeferenced aerial photographs (from MassGIS, 2009). Improvements in the ESRI GIS software since the data was originally digitized in 1998 make rubbersheeting and error checking much faster, easier, and more accurate.
5) The data was rubbersheeted in ArcGIS 10.0 using the "Spatial Adjustment" toolbox with the rubbersheet method selected. BH created 646 rubbersheeting links across the island. Each link connects a still-existing feature on the USCS datalayers to its location on a modern GIS layer; we made the most likely assumption that the modern layer is more accurate. The links are spread more or less evenly across the island as readily identifiable features on both maps allowed (see map below). The Great Plains in the central part of Martha's Vineyard have very few links since there were few identifiable features in common between the two time periods; this is not a major problem because the area was

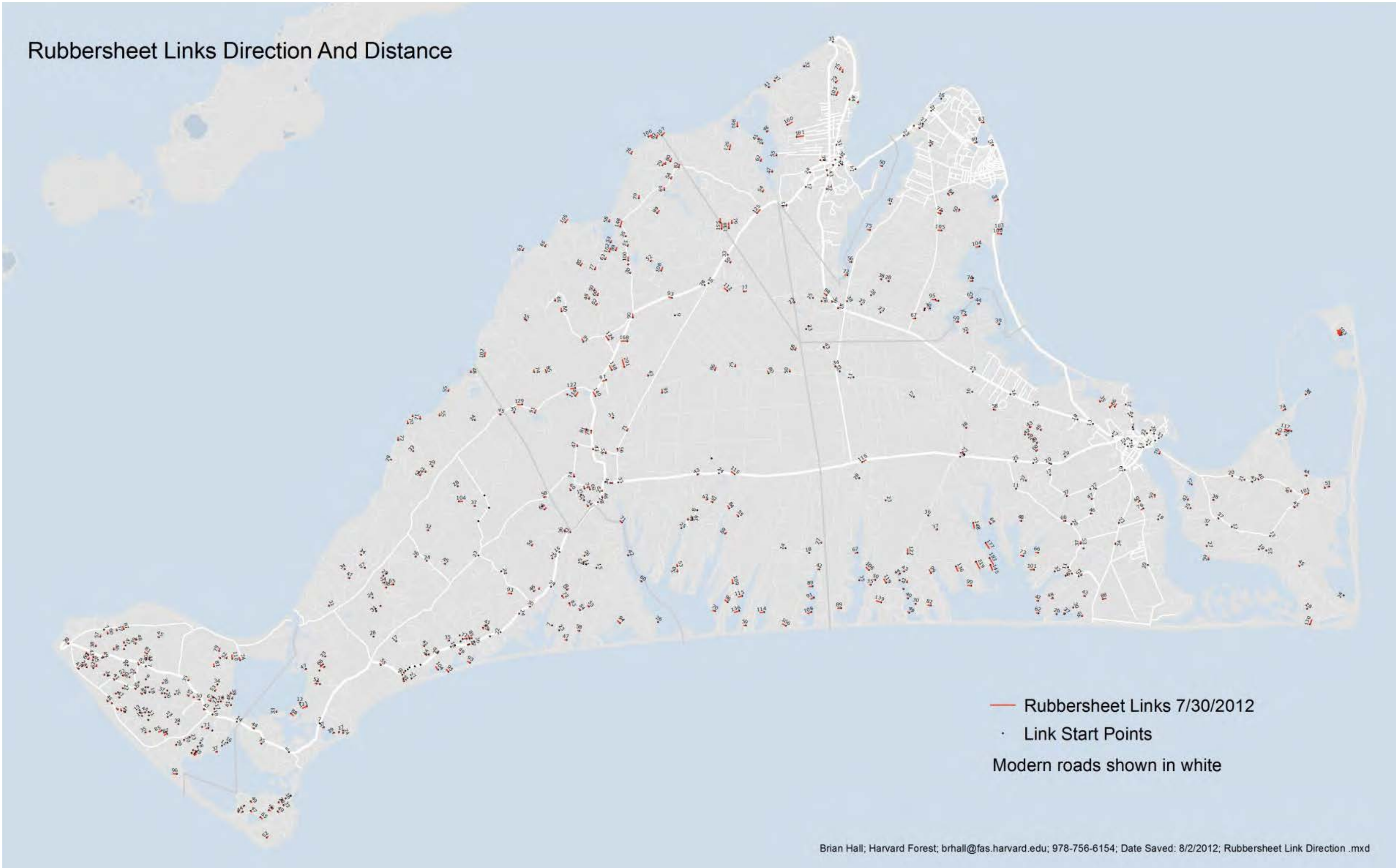
mostly wooded on the USCS maps so spatial inaccuracy associated with such a homogenous and dominant feature is not a major issue during analyses and data presentation. A variety of USCS features that are still visible on modern GIS layers were used, but preference was given to (shown in declining order):

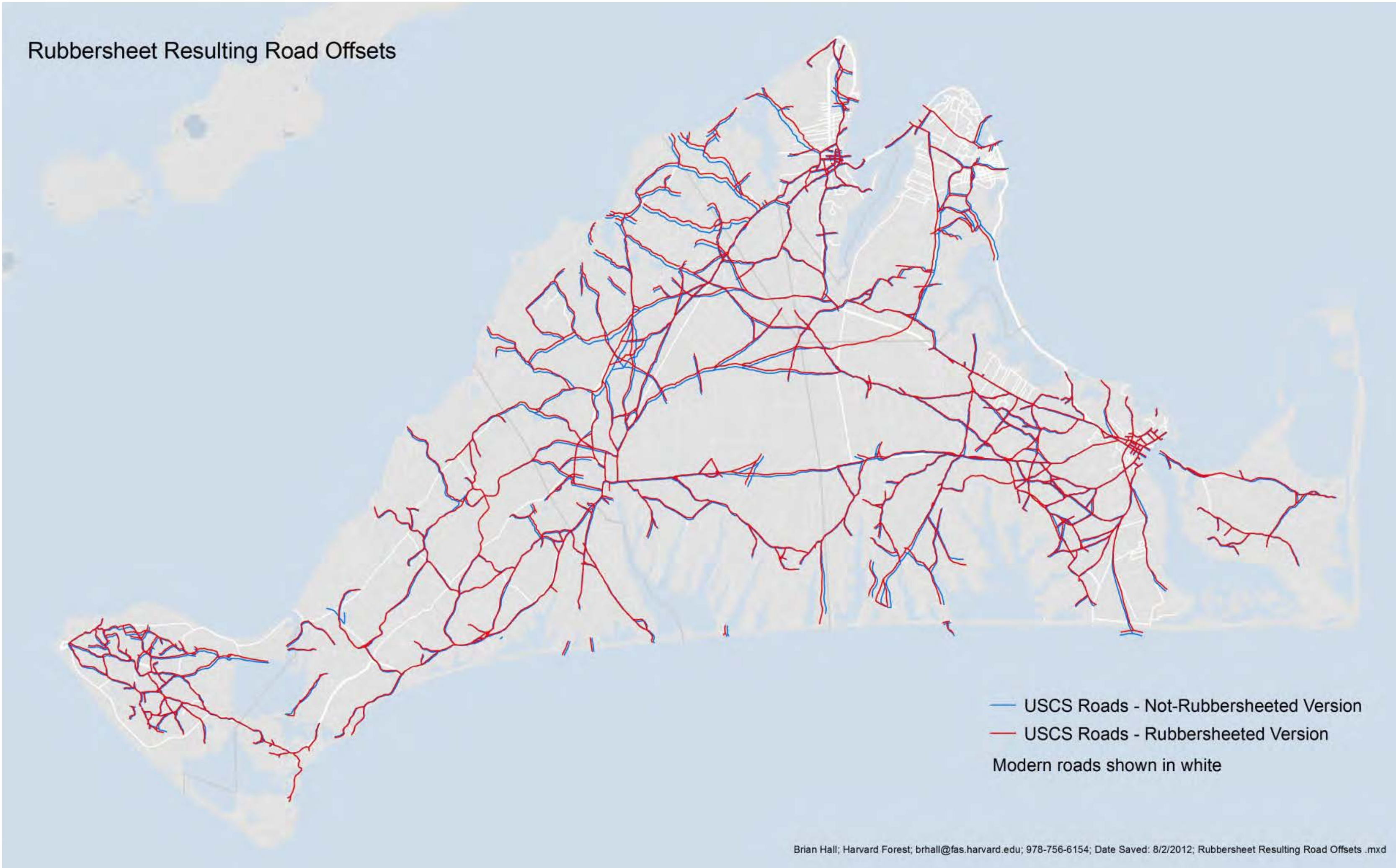
a. road intersections and stone walls that could be seen on aerial photos or the MassGIS road layer,
b. stonewalls that could be inferred from modern parcel boundary shapes,
c. ponds or wetlands with identifiable shoreline features (coves, points, etc),
d. approximate centers of ponds or wetlands
e. coastal pond shorelines or coastline features such as points or coves were used as a last resort in areas with no other comparable features. Although shorelines and coastlines certainly do change over time, it was felt that there would be more spatial error in the resulting dataset if we did not use these features for rubbersheeting.

6) The final rubbersheeting links had a mean distance of 48.6 m and were heavily skewed to the left (many more low values than higher values-see histogram), with half of the links less than 41 m. Important note: this does not mean that the digitized version of the USCS layers were, on average, 48.6 m off, what it means is that where BH thought the distance was too far off and needed to be corrected was on average 48.6 m. (David – this number is full of nuances and biases so it really is not useful; for example, I could have filled the map up with really short links or really long ones and made the average anything I wanted it to be…I think it is most honest, simple, and sufficient to say that the USCS maps were on average 8-15 meters off from modern GIS layers as determined by RMS error in the scans from NOAA.
7) In July, 2012, we obtained scans of the USCS map sheets from NOAA. Brian georeferenced these in ArcGIS 10.0 using the "Georeference" toolbar. Five to ten control points were used for each sheet. The sheets only had a RMS (Root Mean Square) error of 8-14 meters illustrating the high accuracy of the original USCS sheets and the greater accuracy possible with the interactive control point selection process available in more recent georeferencing software. If we were digitizing the maps for the first time today, we would use these georeferenced scans as our source instead of using the digitizing table as we did 14 years ago.
8) BH and David Foster agree that the rubbersheeting process greatly improved the agreement between the USCS datalayers and the locations of the same features as determined by modern GIS layers. Generally the rubbersheeted versions of the USCS data are more accurate than the georeferenced versions of the USCS scans that we received in July, 2012, so we decided to continue to use the rubbersheeted versions in future analyses; we will most likely use the scans for illustration purposes.


Rubbersheeting Links and Their Displacement Distances
Links Distance In Meters
0 - 44
45 - 88
89 - 131
132 - 175
176 - 219


Histogram Of Rubbersheeting Link Displacement Distances
Frequency
Cumulative %
Frequency
Displacement Distance (m)
0
20
40
60
80
100
120
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
20 40 60 80 100 120 140 160 180 200 220 240 260 280 300 320 340 360 380 400 420 440 460 480 500


Rubbersheet Links Direction And Distance
Rubbersheet Links 7/30/2012
Link Start Points
Modern roads shown in white
Brian Hall; Harvard Forest; brhall@fas.harvard.edu; 978-756-6154; Date Saved: 8/2/2012; Rubbersheet Link Direction .mxd


Rubbersheet Resulting Road Offsets
USCS Roads - Not-Rubbersheeted Version
USCS Roads - Rubbersheeted Version
Modern roads shown in white
Brian Hall; Harvard Forest; brhall@fas.harvard.edu; 978-756-6154; Date Saved: 8/2/2012; Rubbersheet Resulting Road Offsets .mxd