TULIP Landmark Density (Average Distance between Landmarks) impact on Apollonius Error Distance

Problem Statement:

After optimizing the Alpha value for different regions, when tests were run for targets of various regions, it was observed that results for only North America and Europe were improved and for other regions, Error distance still mounted to thousand of Kilometers. One observation was that this is mainly due to instability of internet connectivity in these regions and better results in Europe and North America are due to better and stable connectivity in those areas. However, another thing observed during the analysis was that these Europe and N. America each also possess a large number of landmarks compared to other regions. So the impact of landmark's density can't be ignored. It was observed that most of the times poor results appear mainly because of the fact that no landmark is found in the close vicinity of the target and hence even the three final selected landmarks with minimum RTT may have RTT values above 300ms or so.

So to analyze this geographical distance impact on Error distance, I have conducted a region by region analysis in which I have done following main tasks:

- First, I found distance between different landmarks of a region (their inter-distance) and then calculated the average distance for each region. E.g. if there are 10 landmarks in a certain region, I calculated distance between a landmark and all other landmarks in that region and in this way inter-distance of all landmarks was calculated and then averaging out these results, I got one average distance value for each region.
- Then I took targets of each region, and run tests using the Apollonius geolocation technique for them (using average alpha value of that particular region). In this way I got error distance for each target and then we calculated average Error distance for each region.
- Now the analysis began. I wanted to investigate that what behavior will be there in error distance when the average geographical distance is higher and what it will be when average geographical distance is lower. Of course, an ideal situation was that when there is high geographical distance between landmarks, i.e. landmarks are far apart from each other then the error distance should be higher compared to the case when we have high number of landmarks in a particular region (i.e. we have low average geographical distance).
- Following table shows the average geographical distance and average error distance for each region.

Region	"Avg Distance of Landmarks"		"Avg Error Distance"
Europe	555.6404153		440.0434656
East Asia	841.0220395		2966.42115
South Asia	1006.405099		6379.043878
North America	1756.877965		804.486103
Latin America	1790.503606		5431.530291
Africa	3387.983944		4743.33271

For Europe and North America the distribution of landmarks within 500km of each other is seen below. the data comes from a spreadsheet . it is seen that n. America has more landmarks close together than Europe. This maybe since the N. American landmarks are clustered on the East Coast, West cost and around Chicago.

Histogram of number of landmarks within 500km of a landmark target
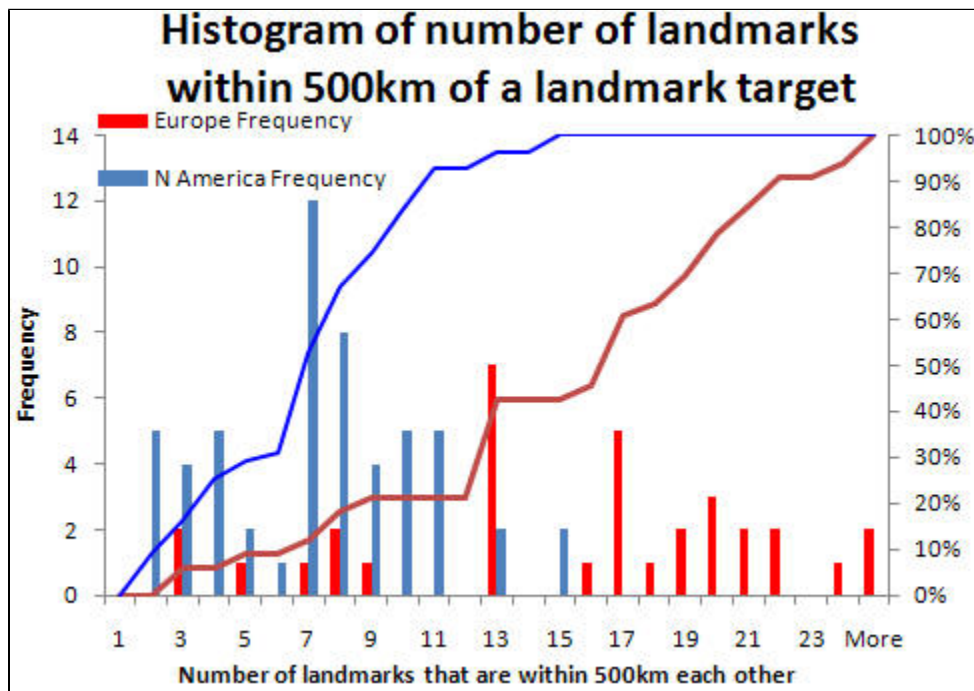
We also looked at the value of alpha to use in order to get the correct distance between the landmark and target target based on the measured minimum RTT. Histograms of these values is plotted below or N. America and Europe.
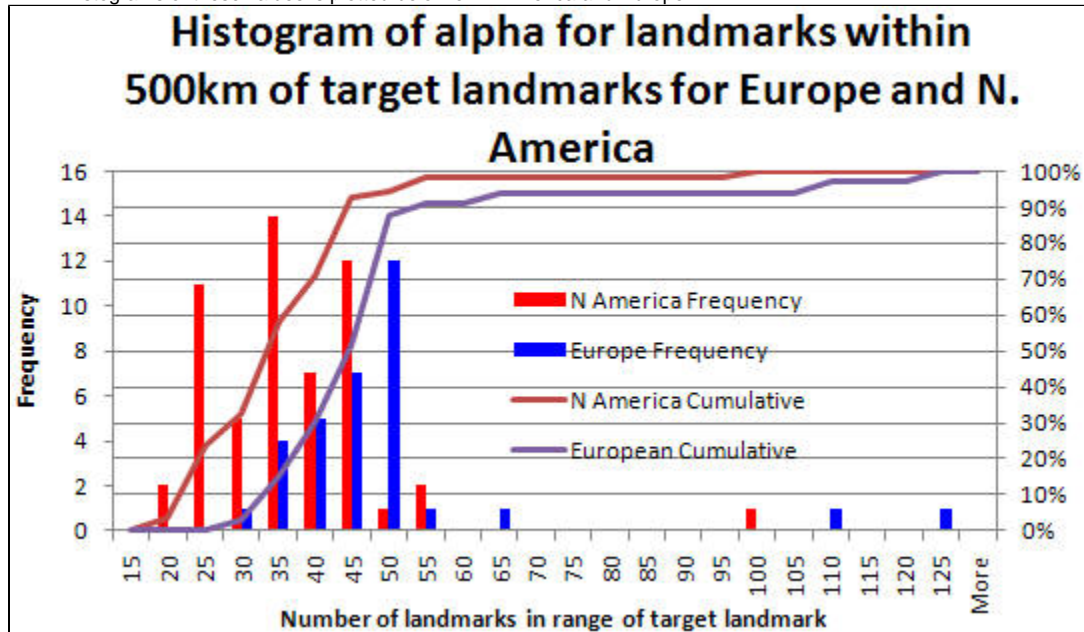

It is seen the the median alpha for N.america is ~ 0.35 while for Europe it is ~0.45.

Based on the above data, we plotted a graph to see the relationship between distance and error distance (as predicted by the Apollonius method) for each major region where we have a majority of the landmarks. The spreadsheet of the detailed results is available.
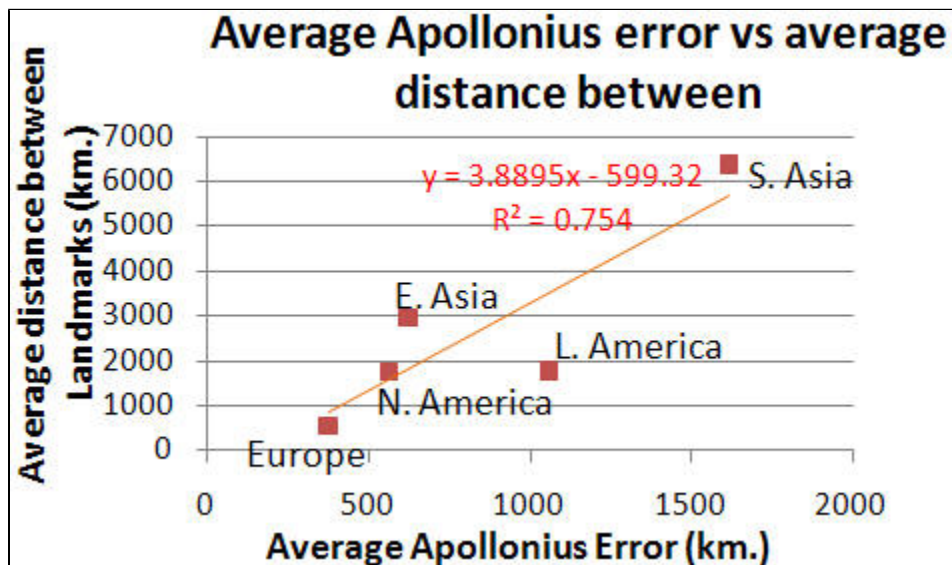
It is seen that as expected that the error is generally smaller if the landmarks are closer together. The same result shown in the form of bar graph below: is:

Distance Versus Error Distance
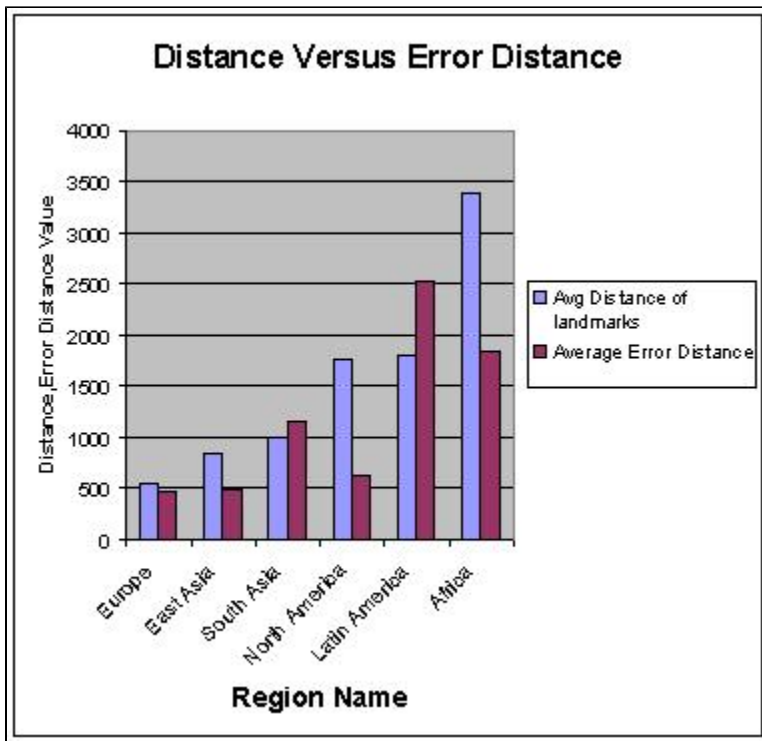
From the graph, we see that in Europe where the number of landmarks is high and there is less average distance between landmarks, then we get good results in terms of error distance. The trend looks fine when we move to next regions where average geographical distance is comparatively higher (error distance has also increased for these regions) but we see a different trend for North America where average geographical distance between landmarks is higher than other regions like East Asia and South Asia but yet, the Error distance is quite lower. In fact this result depicts that even though landmarks are geographically dispersed and not closely located in this region, the connectivity in this region is so better that we get quite reasonable results even then.

For all other regions, we see some impact of geographical distance and from this we can say that hopefully results for those regions can also improve if we could have more landmarks in those regions.

This hypothesis is further strengthened and proved if we see the results in the attached sheet Analysis Results. It's clear from the result that for all those targets for which Error Distance is less than 1000Km, the MinRTT of the first selected landmark is less than even 60ms. This means that if for any target we are able to find a nearest landmark (with MinRTT<60), then we will be getting good results and smaller this minRTT is better are the results. From the attached results, we can also see that for all those targets for which we got error distance above 2000Km, the MinRTT (column D) was above 100ms.

From these results, we can roughly conclude that if we mange to have our landmarks distributed in the world in such a way that we get low MinRTT for any target, the error distance can drastically reduce.