36-467 Midterm

Eu Jing Chua

10 October, 2018

Introduction

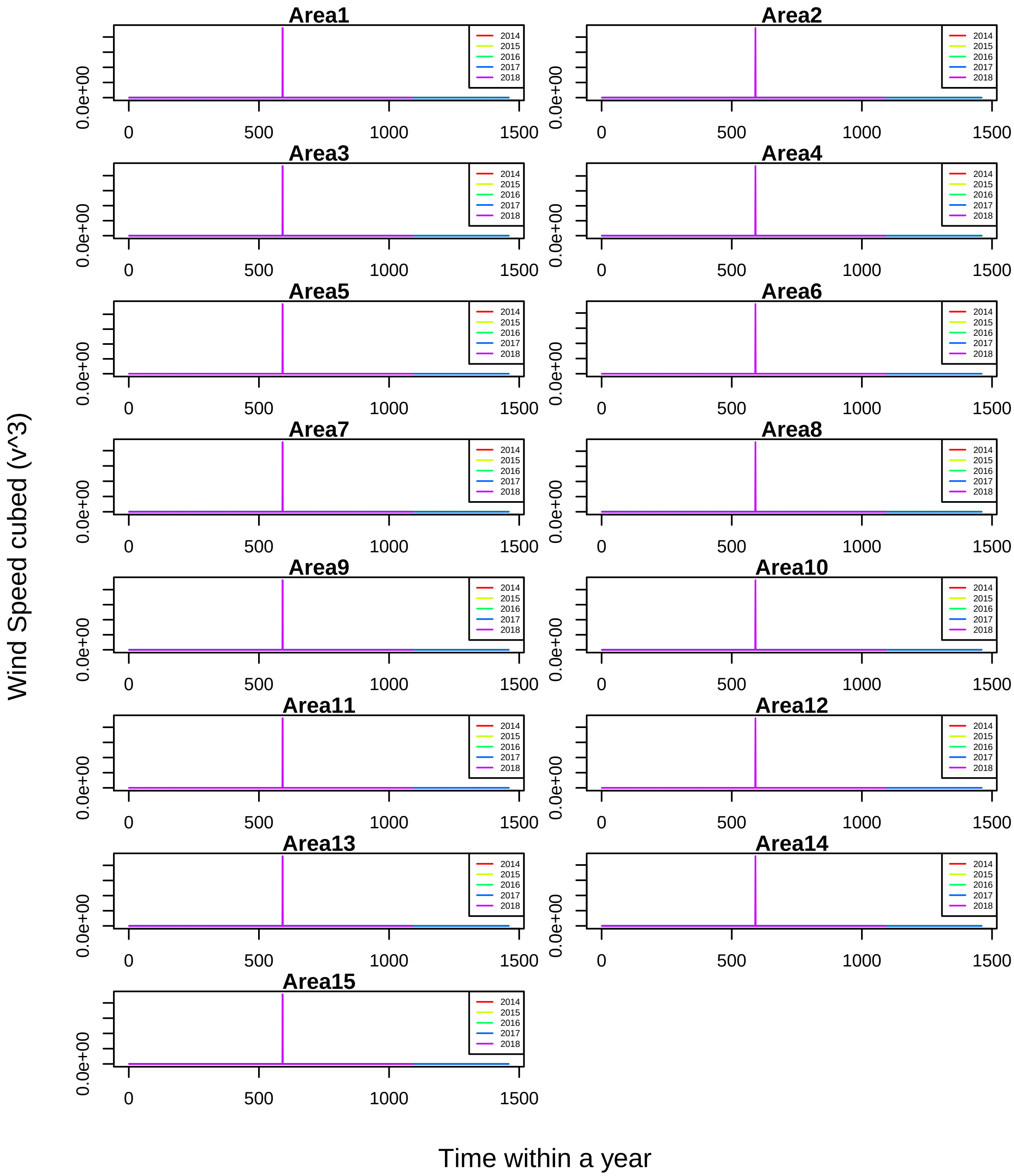
We seek to systematically find the optimal location for wind power. This location should have the highest windspeed v, as energy generated is proportional to v^3. We then also seek to find the estimated expected value of v^3 at this location, as well as the standard error for this estimate.

We also code the areas as follows:

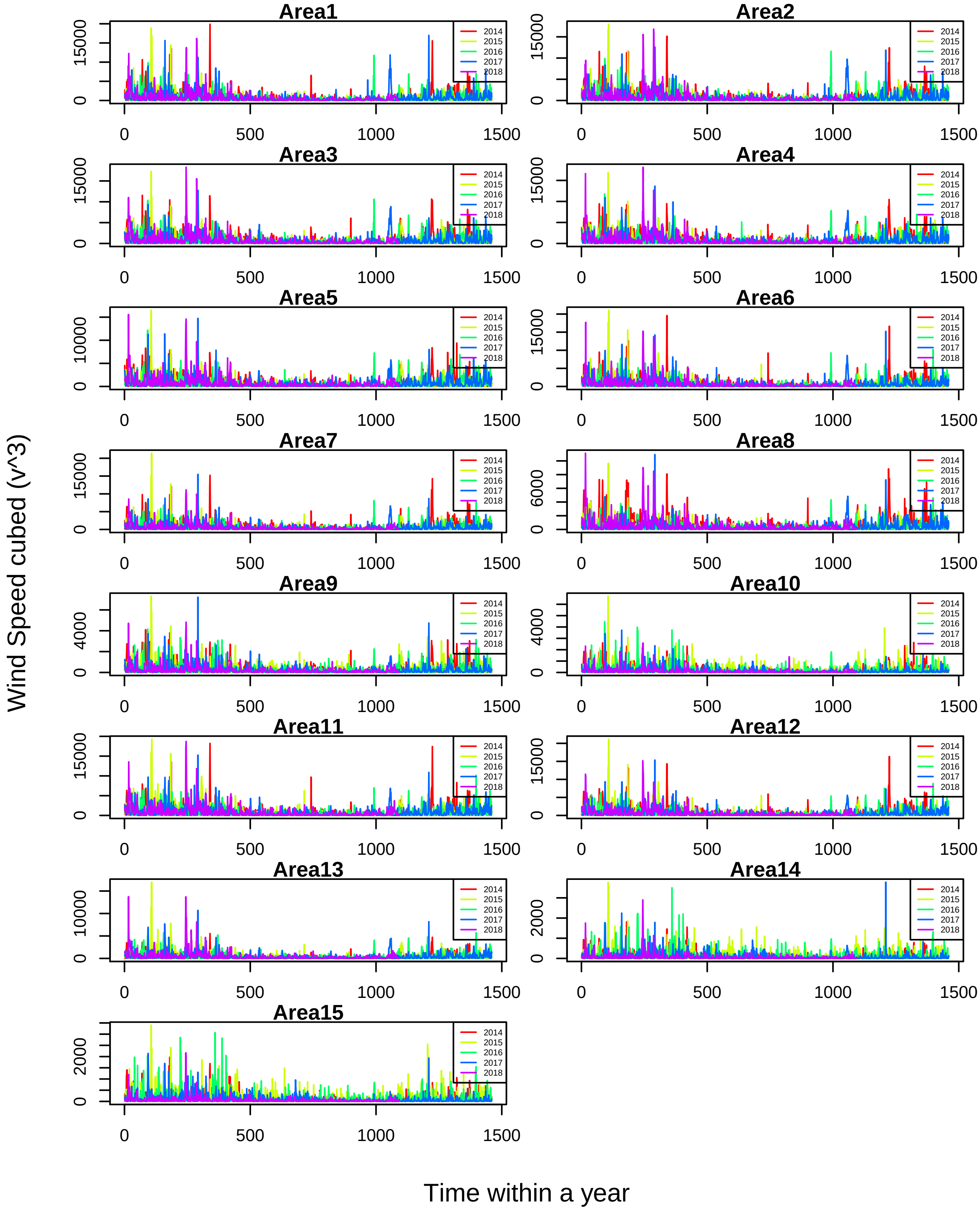
Label	Latitude	Longitude
Area1	41.0	-69.5
Area2	41.0	-70.0
Area3	41.0	-70.5
Area4	41.0	-71.0
Area5	41.0	-71.5
Area6	41.5	-69.5
Area7	41.5	-70.0
Area8	41.5	-70.5
Area9	41.5	-71.0
Area10	41.5	-71.5
Area11	42.0	-69.5
Area12	42.0	-70.0
Area13	42.0	-70.5
Area14	42.0	-71.0
Area15	42.0	-71.5

Exploratory Data Analysis

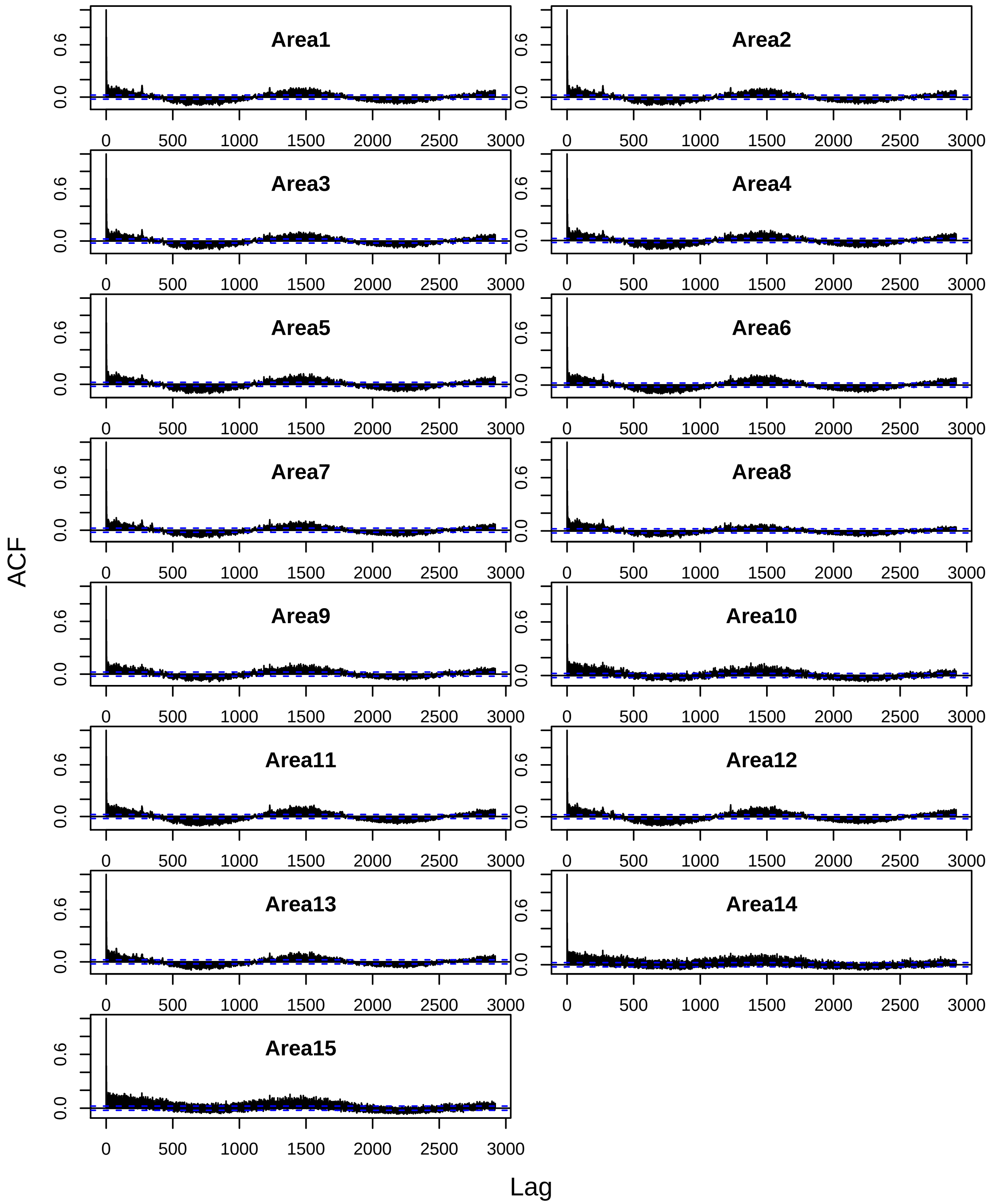
First, since we are interested in v^3, we apply a cubic transform to the windspeeds in the data set to make later analysis easier. Then we shall analyze if there are any annual patterns in the wind speeds in each area, across the years 2014 to 2018. We overlap the plots of cubed wind speeds for each year within a region to see if there seems to any seasonal trends:



However, there seems to be an obvious outlier in the data which effects got exaggerated by the cubic transform, so we remove it and re-plot:

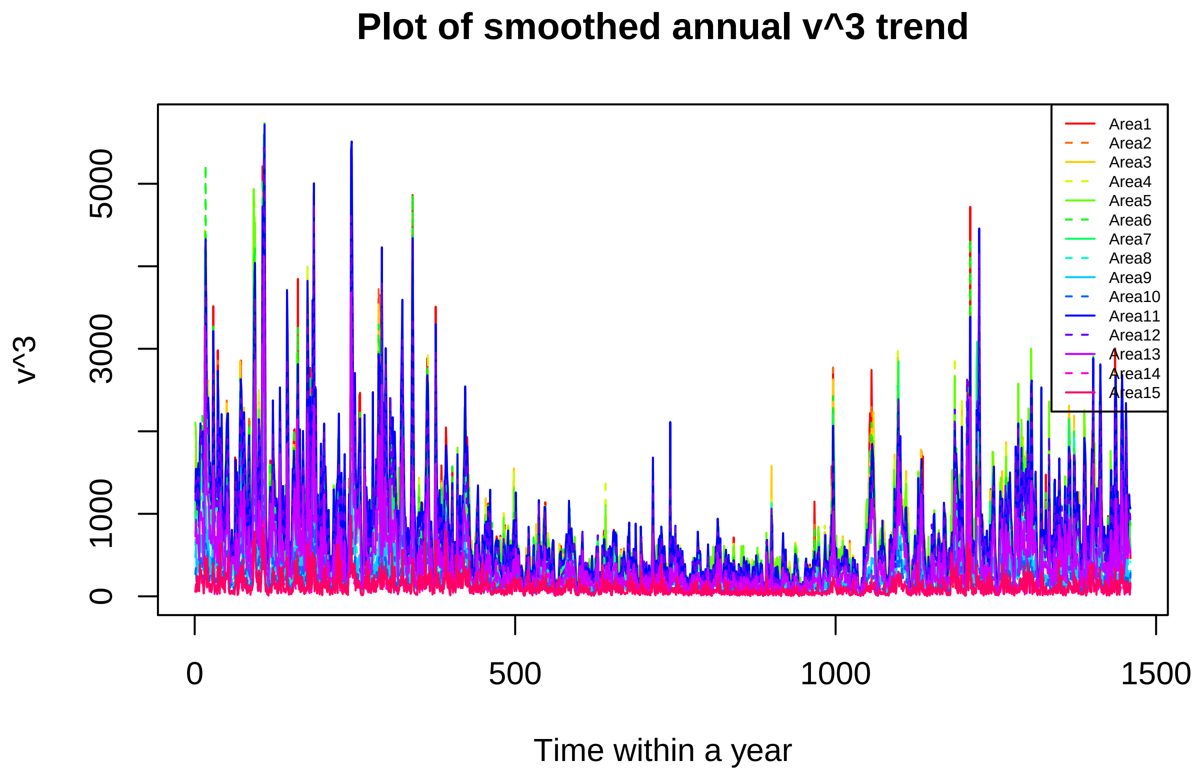


Now that we suspect that there might be an annual trend, we also plot the auto-correlation over time within each location to see if the correlation has a period nature, indicating a seasonal trend:



Detrending & Stationarity in Time

As seen from the auto-correlation, there is a seasonal trend for each location as the ACF is sinusoidal, with a period of roughly $365 * 4$ as there are four observations a day, hence stationarity over time cannot be assumed yet. However, by taking an average across the years for each location, we can smooth this seeming annual trend:



Now that we have determined an approximate trend for the windspeed cubed, v^3, across a year, we know that we cannot assume v^3 is stationary as the mean at any point in time in the year is seemingly non-constant. However, we can use the trend that we have just found to de-trend the data and analyze the noise around the trend, plotted as so:

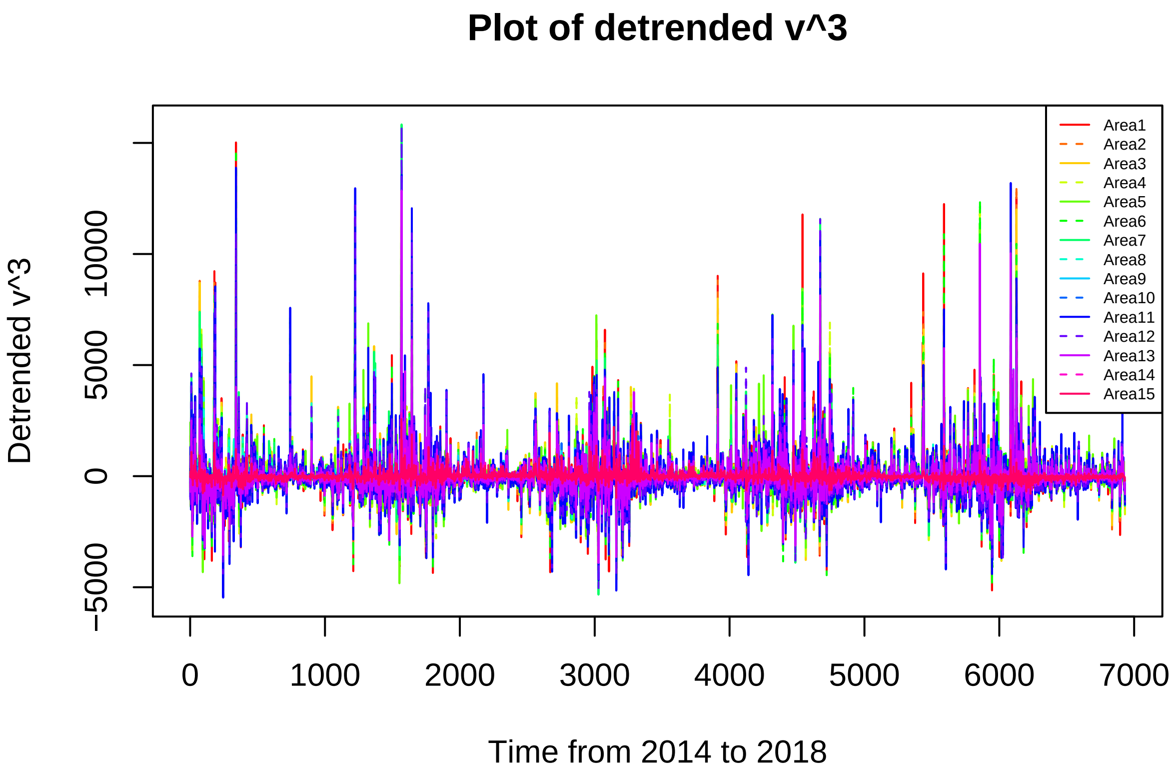
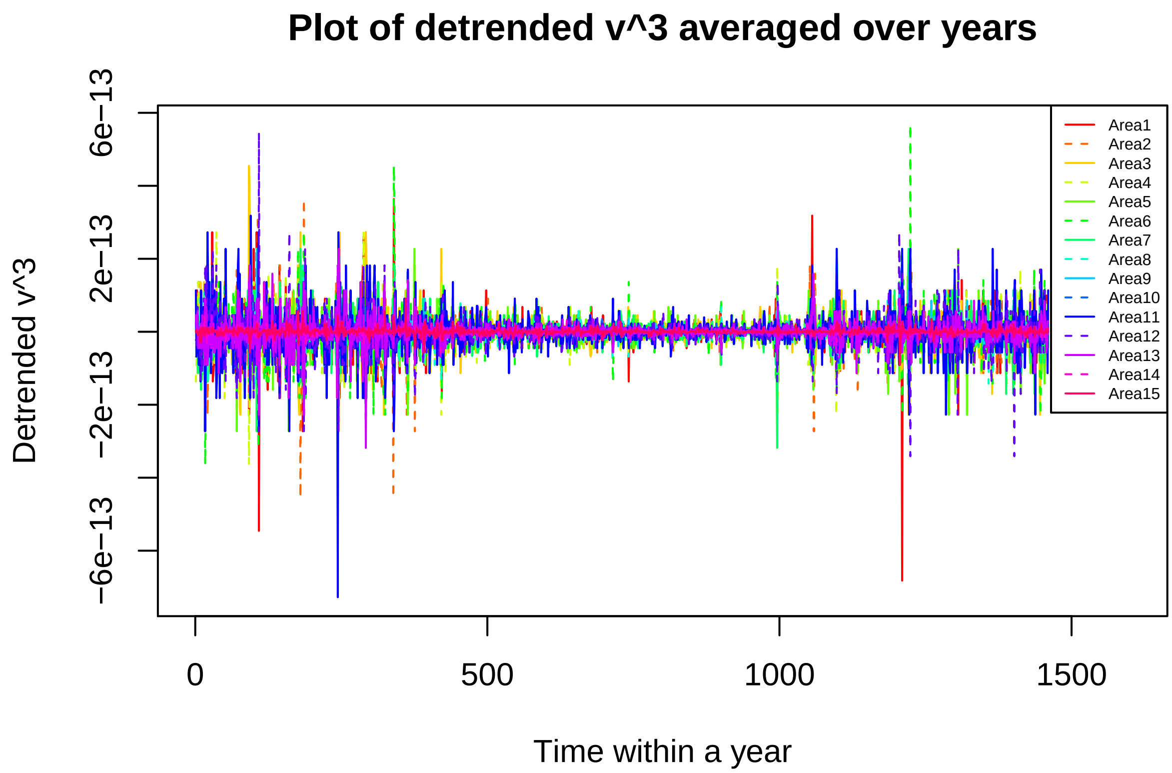


Table 2: Summary statistics of detrended wind speeds

	Area1	Area2	Area3	Area4	Area5	Area6	Area7	Area8
Min.	-5143.36	-4631.30	-5160.89	-4945.33	-4822.97	-4929.19	-5334.66	-2838.23
1st Qu.	-384.01	-375.51	-399.48	-395.49	-370.23	-389.12	-339.06	-245.31
Median	-88.87	-94.97	-102.04	-104.12	-93.09	-92.99	-89.91	-66.45
Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3rd Qu.	208.87	208.84	220.77	220.10	201.40	228.26	196.32	135.09
Max.	15021.85	13078.58	12951.85	12967.02	11678.51	15334.39	15831.72	8141.51
NA's	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

	Area9	Area10	Area11	Area12	Area13	Area14	Area15
Min.	-1949.10	-1601.20	-5468.40	-5063.28	-3907.56	-972.05	-977.35
1st Qu.	-152.98	-105.41	-416.96	-353.04	-220.03	-54.66	-58.51
Median	-40.39	-27.60	-107.51	-88.85	-58.42	-13.71	-14.70
Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3rd Qu.	81.82	52.68	243.36	203.57	115.06	23.50	26.68
Max.	5436.76	5056.90	13884.36	15721.05	12855.84	2895.03	2566.36
NA's	1.00	1.00	1.00	1.00	1.00	1.00	1.00

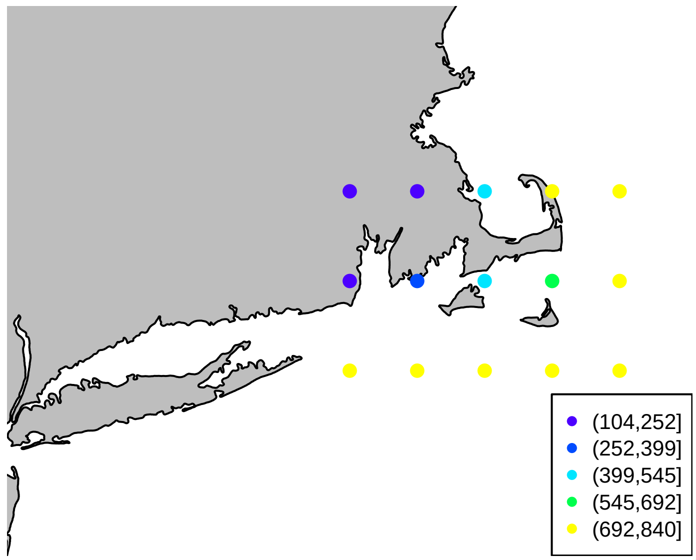


The plot of the averaged de-trended v^3 now appears to be more of a random scatter about 0, as also indicated by the summary statistics. We can now more safely make the assumption that this de-trended v^3 is stationary in time.

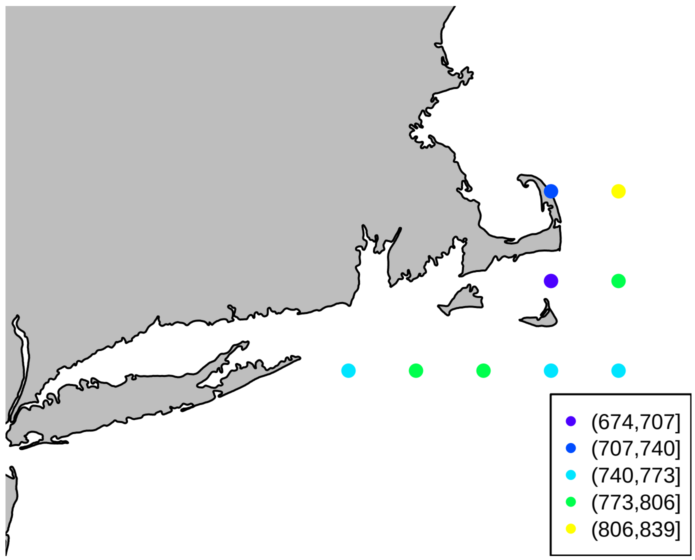
Spatial Kriging

It should be noted that the data was collected over a regular 3x5 grid, where the average annual wind speed at each point is as follows:

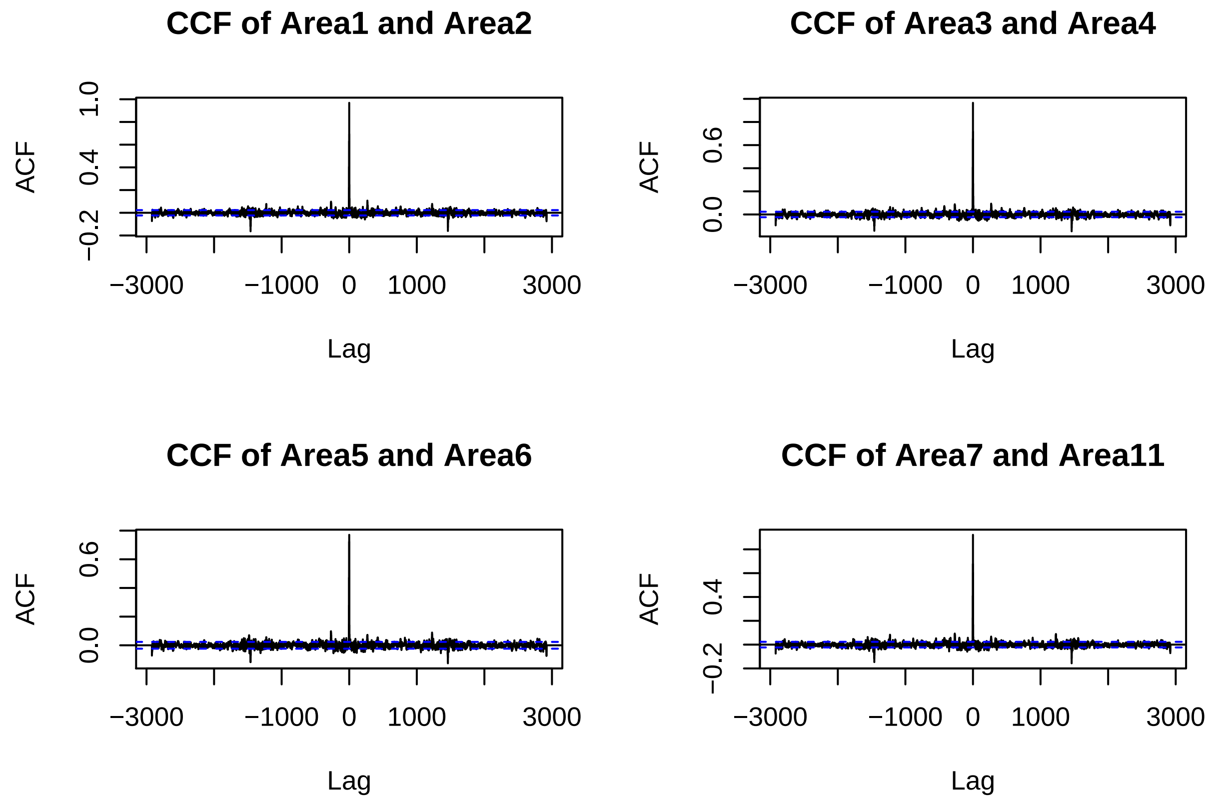
	-71.5	-71	-70.5	-70	-69.5
41	744.007346825	784.598137367	787.220436997	751.851018664	765.803428280
41.5	195.342699964	290.896228596	465.072866562	673.685141970	788.441082722
42	114.753577057	104.945648408	446.193745644	721.011561467	839.034215039



As we can visually observe, for the data that we have on the grid, locations at sea tend to have a higher average wind speed compared to locations on land. Although the data that we have is evenly spaced out on a grid, this stark difference in means between land windspeeds and sea windspeeds prevents us from assuming stationarity of the correlation function. In order to proceed, we can stratify the data into land and and sea windspeeds. Since we are interested in finding a location with higher windspeeds, we focus on the sea strata. In order to perform this stratification simply, we pick the locations that have higher than average (global) wind speeds.



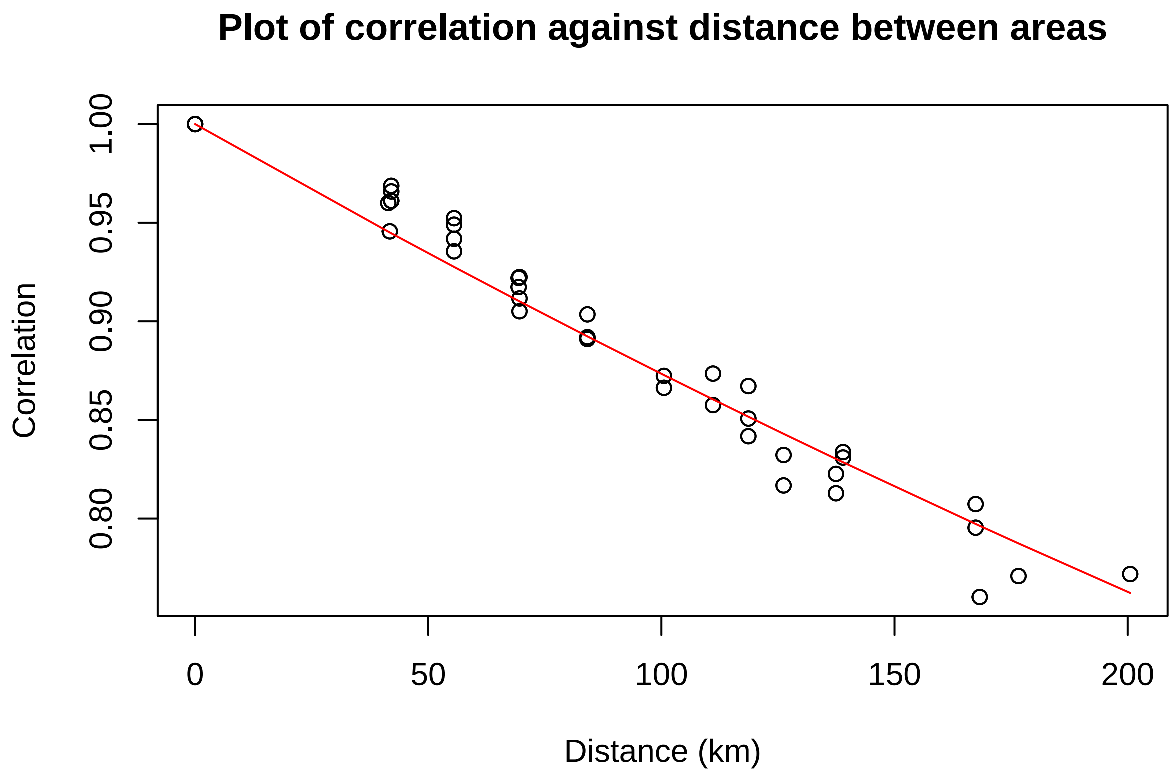
Among these locations, we assume that the correlation function has stationarity. Indeed, we can verify this by looking at the cross-correlation functions.



Although this does not represent all pairwise CCFs, we can see that there is no significant cross-correlation between these locations.

Now, we want to be able to estimate a correlation function so that we can make predictions with an optimal linear predictor on locations that we do not have data for. We do not want to restrict ourselves to just finding locations on this 3x15 grid and would like to see if there are other more optimal locations around the grid. Thus, we would like to assume, in addition, the correlation function is isotropic.

Assuming that the correlation function is stationary and isotropic among the sea locations, we can estimate the correlation with a parametric exponential decay.



By performing non-linear least squares, we get an estimated correlation distance length of 738.81km.

Now, we can use optimal linear prediction with our estimated correlation function to predict for new locations. Since we know for our coarse grid that the highest v^3 observed was at longitude -69.5 and latitude 42, we conduct a finer grid-search centered around this location in try find an optimal location while saving computation time.

	-69.75	-69.7	-69.65	-69.6	-69.55	-69.5	-69.45	-69.4	-69.35	-69.3	-69.25
41.75	750.22	760.33	770.04	779.00	786.91	793.53	798.73	802.53	805.01	806.34	806.72
41.8	755.68	765.95	775.83	784.94	792.89	799.39	804.31	807.66	809.61	810.38	810.22
41.85	760.79	771.54	781.93	791.51	799.78	806.29	810.80	813.44	814.54	814.47	813.56
41.9	765.23	776.65	787.86	798.36	807.43	814.16	818.04	819.48	819.30	818.13	816.32
41.95	768.60	780.63	792.69	804.49	815.27	822.95	825.45	824.76	822.91	820.55	817.91
42	770.53	782.64	794.96	807.43	820.00	832.63	829.70	826.76	823.79	820.78	817.71
42.05	770.84	782.27	793.72	804.91	815.11	822.25	824.27	823.18	820.97	818.33	815.46
42.1	769.68	779.88	789.82	799.08	806.98	812.63	815.56	816.19	815.33	813.59	811.35
42.15	767.39	776.22	784.67	792.35	798.83	803.71	806.80	808.22	808.32	807.45	805.91
42.2	764.30	771.89	779.07	785.55	791.07	795.39	798.42	800.19	800.85	800.62	799.69
42.25	760.68	767.22	773.37	778.93	783.71	787.58	790.46	792.35	793.34	793.53	793.06

With our optimal linear predictor applied to the original v^3 data, we find that the location with the highest wind speeds still corresponds to longitude -69.5 and latitude 42, the location from our original data with the highest wind speed.

Results

As we have just seen, using an optimal linear predictor on the sea locations we picked out, and using our parametric estimate of the correlation function with this model resulted in the longitude -69.5 and latitude 42 having the highest expected v^3 of 832.63 $(m/s)^3$, with a standard error of 47.027 $(m/s)^3$.