Tree Data Analysis

Tahmidul Islam

9/5/2020

Tree Data

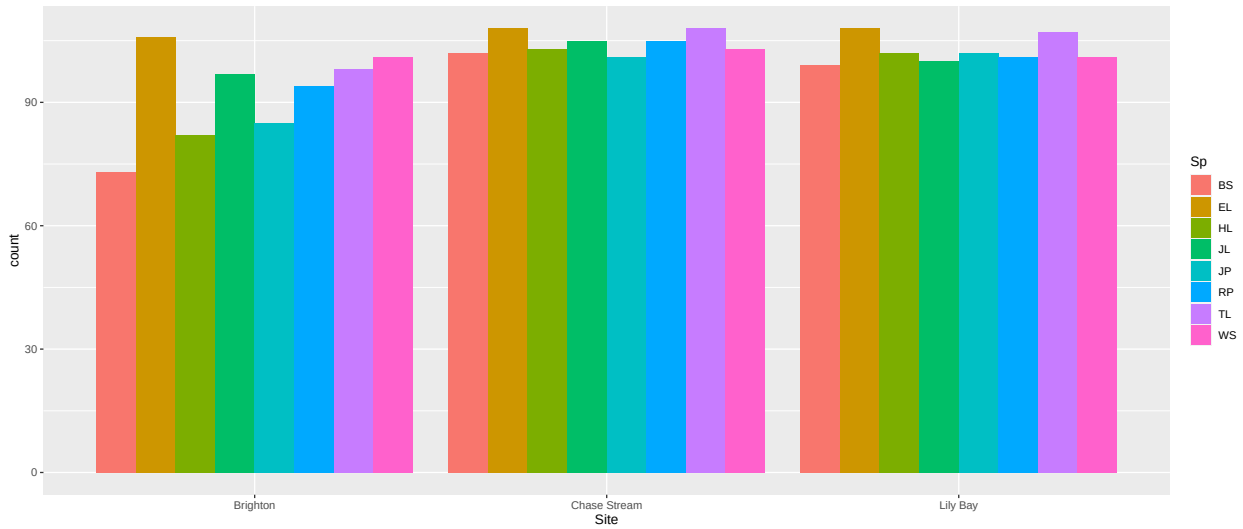
This dataset contains features of trees (height, width, etc.) from 3 sites: Brighton, Chase Stream and Lily Bay. There are 8 species in this sample.

Table: Data overview.

Site	Tree	Rep	Sp	Year	Height	id
Brighton	1	1	BS	5	4.5	11
Brighton	1	1	BS	10	10.4	11
Brighton	2	1	BS	5	6.8	21
Brighton	2	1	BS	10	13.1	21
Brighton	4	1	BS	5	4.4	41
Brighton	4	1	BS	10	10.3	41

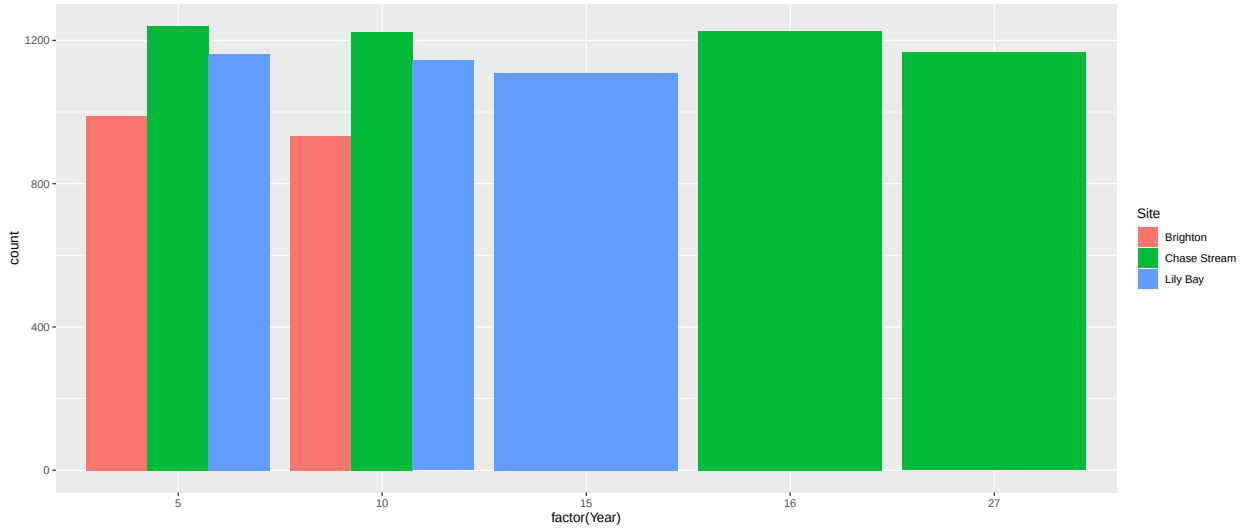
Table: Distribution of site and species.

Site	BS	EL	HL	JL	JP	RP	TL	WS
Brighton	73	106	82	97	85	94	98	101
Chase Stream	102	108	103	105	101	105	108	103
Lily Bay	99	108	102	100	102	101	107	101

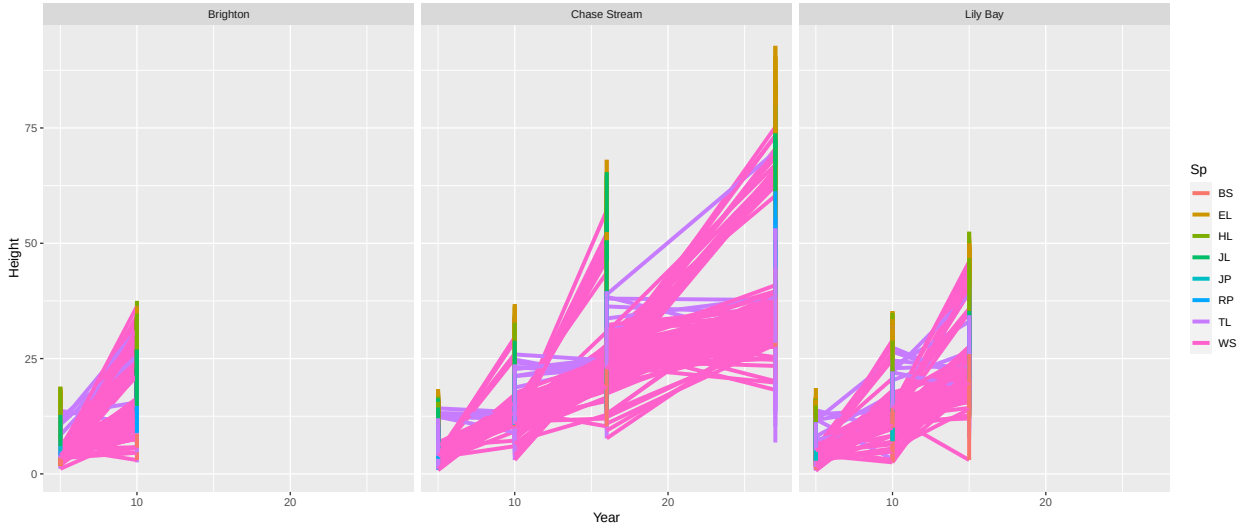


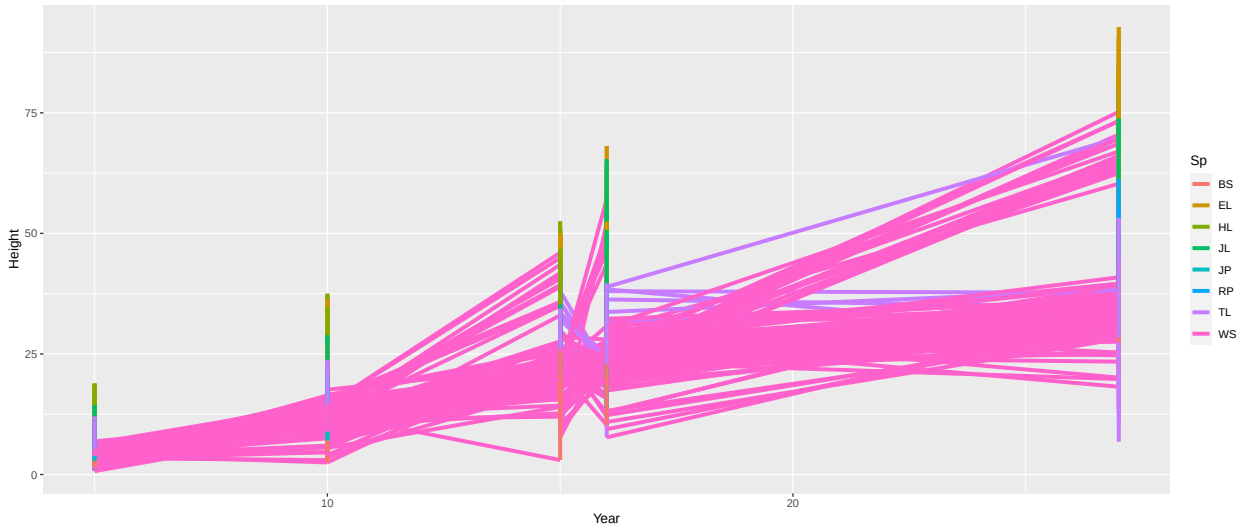
The measurements are collected few years apart: at 5, 10, 15, 16 and 27 years.

Site	Year	n
Brighton	5	989
Brighton	10	933
Chase Stream	5	1240
Chase Stream	10	1224
Chase Stream	16	1226
Chase Stream	27	1166
Lily Bay	5	1162
Lily Bay	10	1144
Lily Bay	15	1109



There is height measurement for years after 10 in Brighton and year after 15 in Lily Bay.





GP Model for Tree Height

Assuming the tree heights for each of the species as a function over time, we analyze the data in functional data analysis approach. For each species, we combine data from all three sites and model the mean functions of the tree heights. We have combined the variables `Tree` and `Rep` to create an `id` variable. Gaussian kernel has been for the estimation.

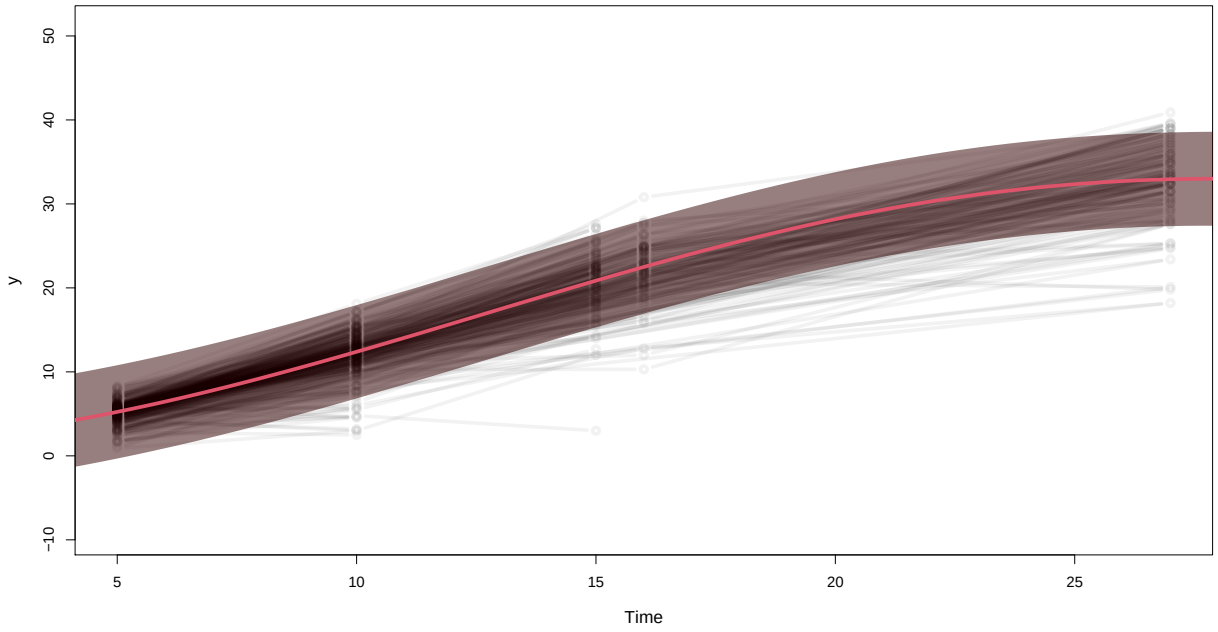


Figure 1: Species: BS

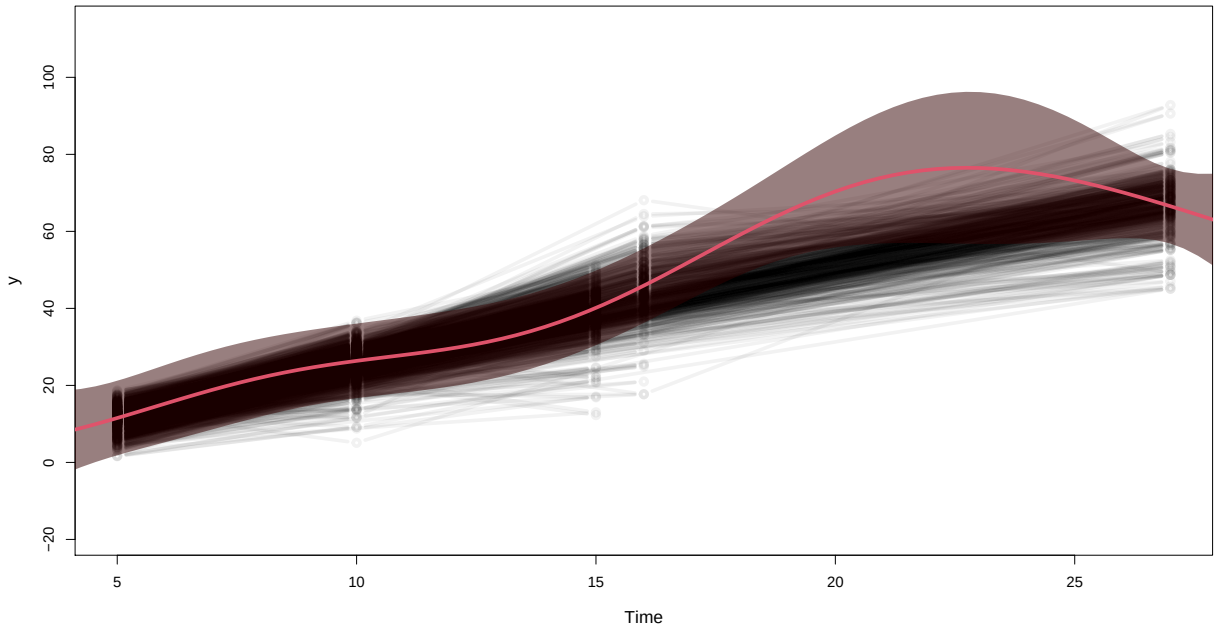


Figure 2: Species : EL

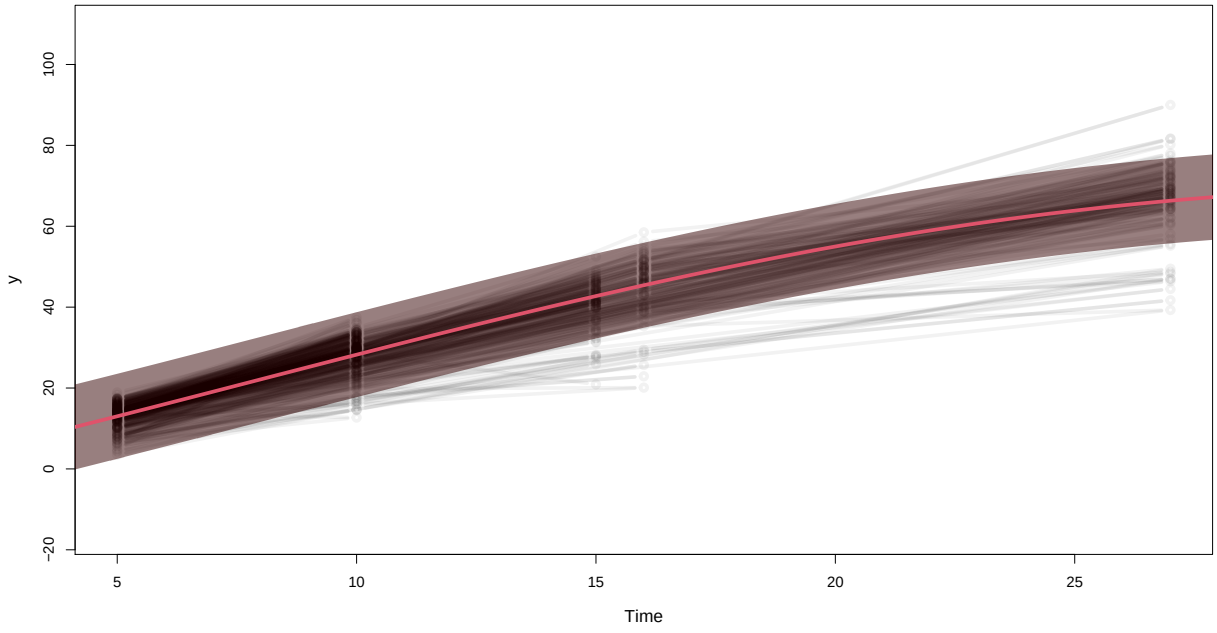


Figure 3: Species : HL

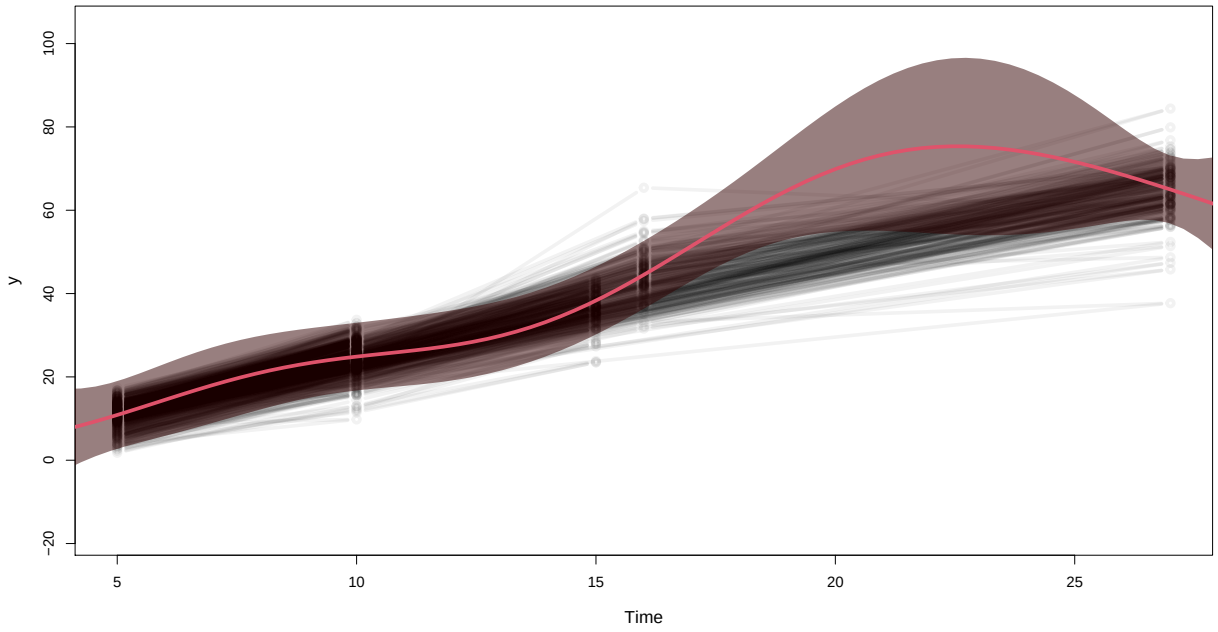


Figure 4: Species : JL

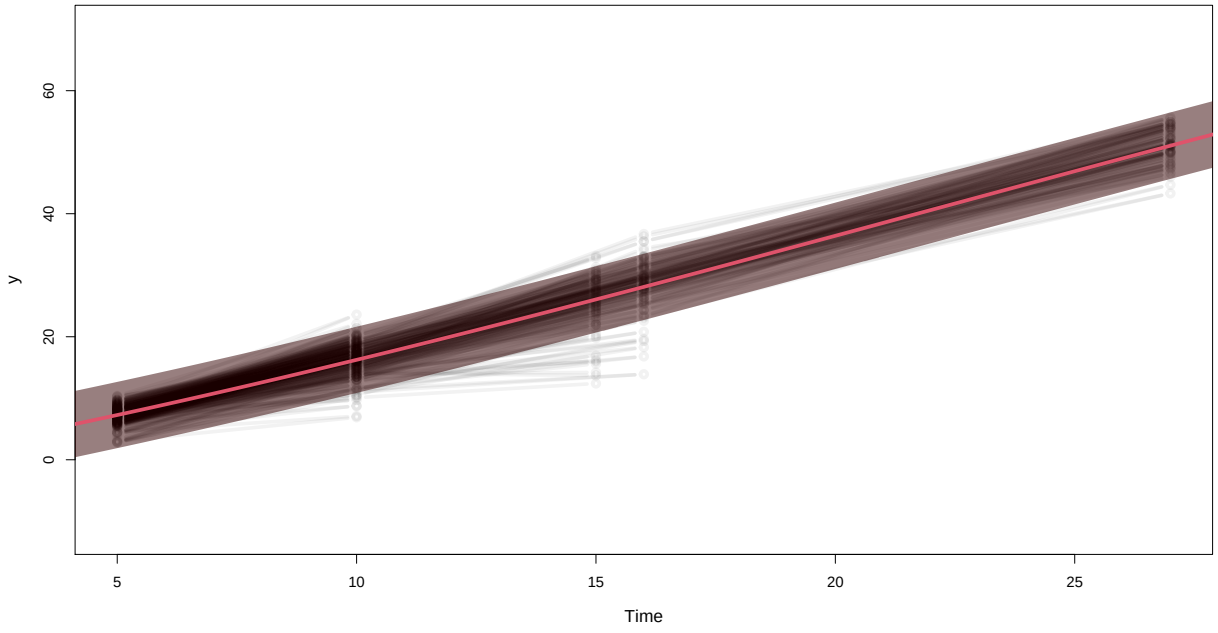


Figure 5: Species : JP

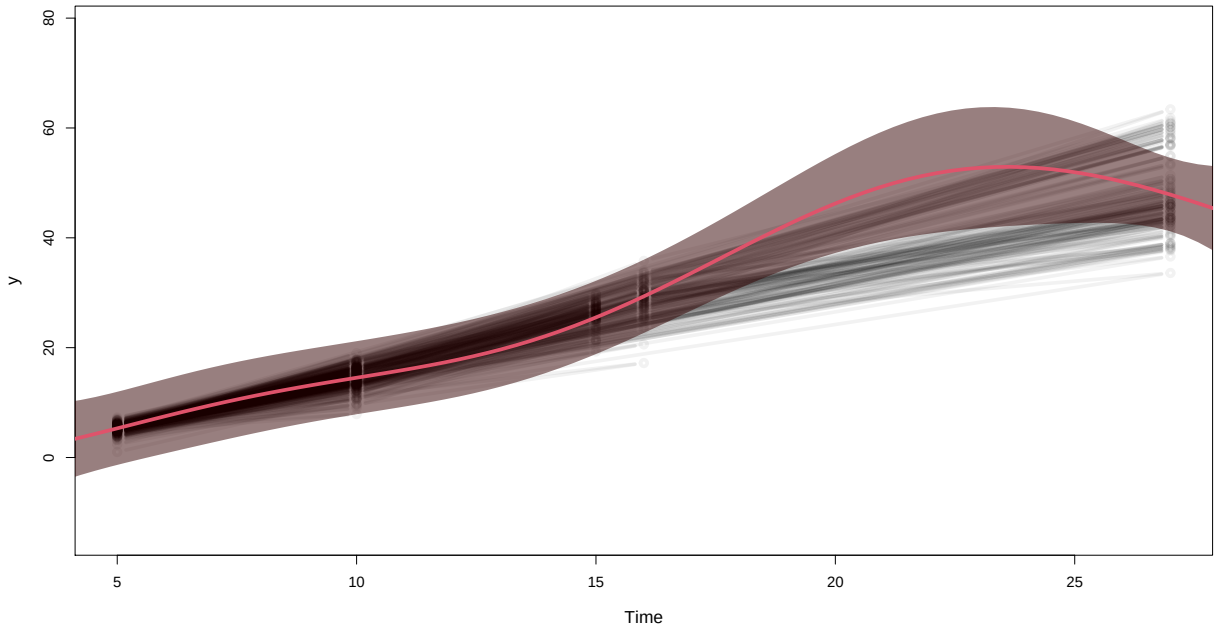


Figure 6: Species : RP

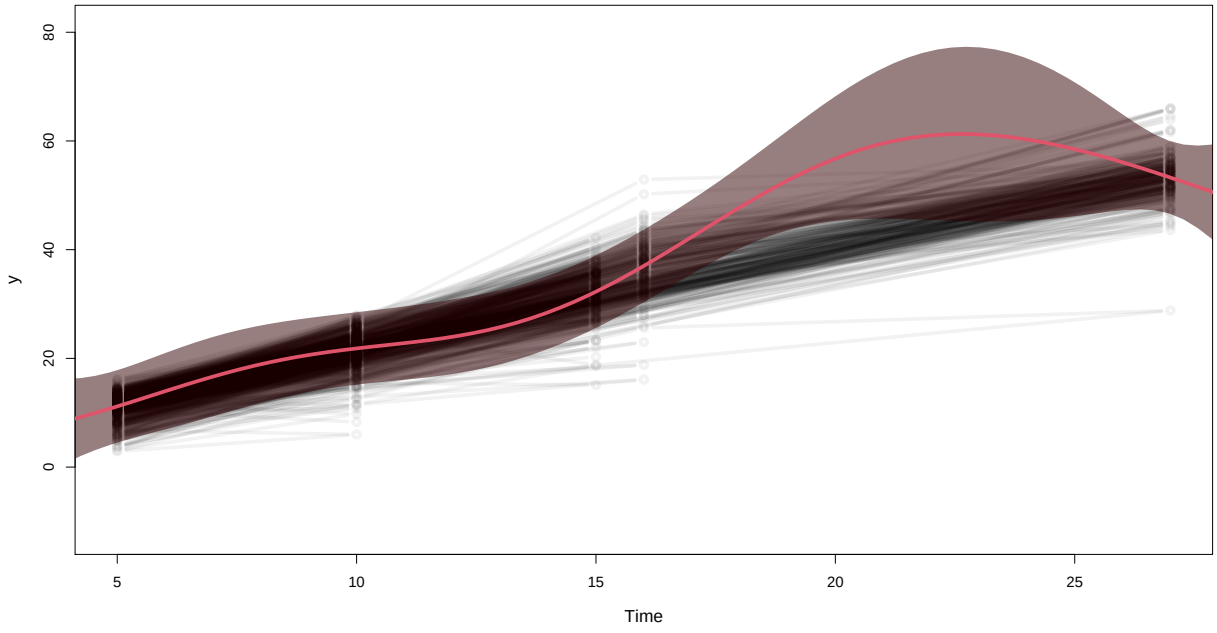


Figure 7: Species : TL

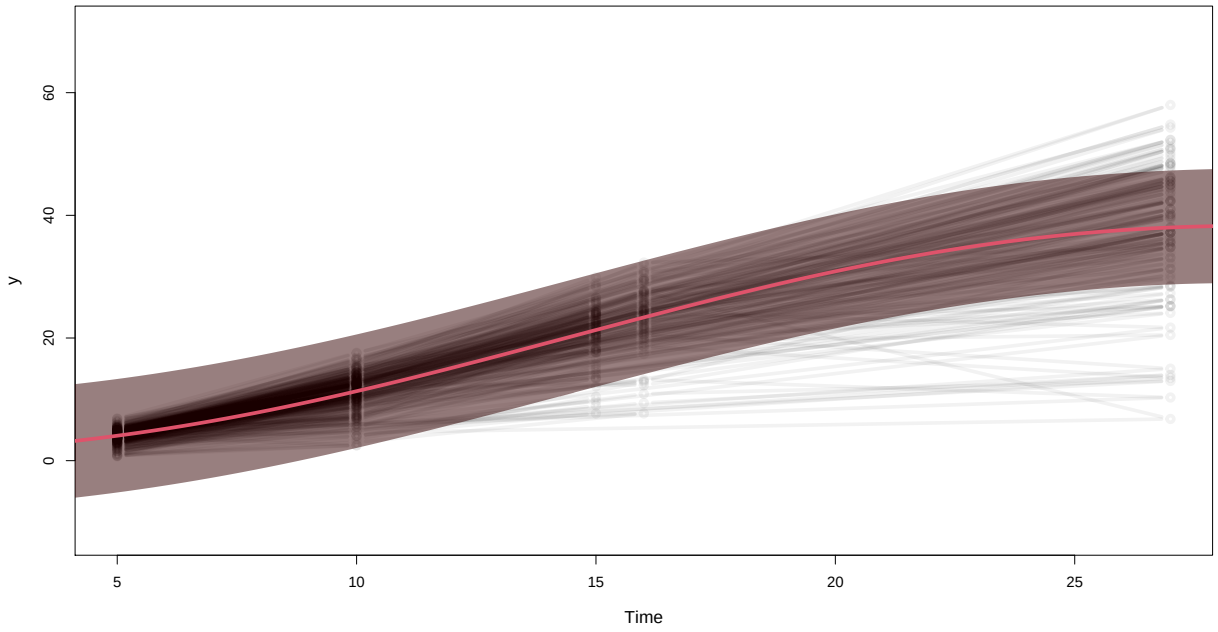


Figure 8: Species : WS

Site Specific Mean Height Estimation

We attempt to estimate the mean function for each of the species by the sites. The black curve denotes the overall mean function where the red, green and blue curves represent the site specific mean functions for the sites Brighton, Chase Stream and Lili Bay respectively. We observe some features in the tree data as well as some issues with the model fit. For example, we have found no site specific variation for the species BS, all the site specific curves coincide with the overall mean curve. It suggest there is no site specific effect for the species BS.

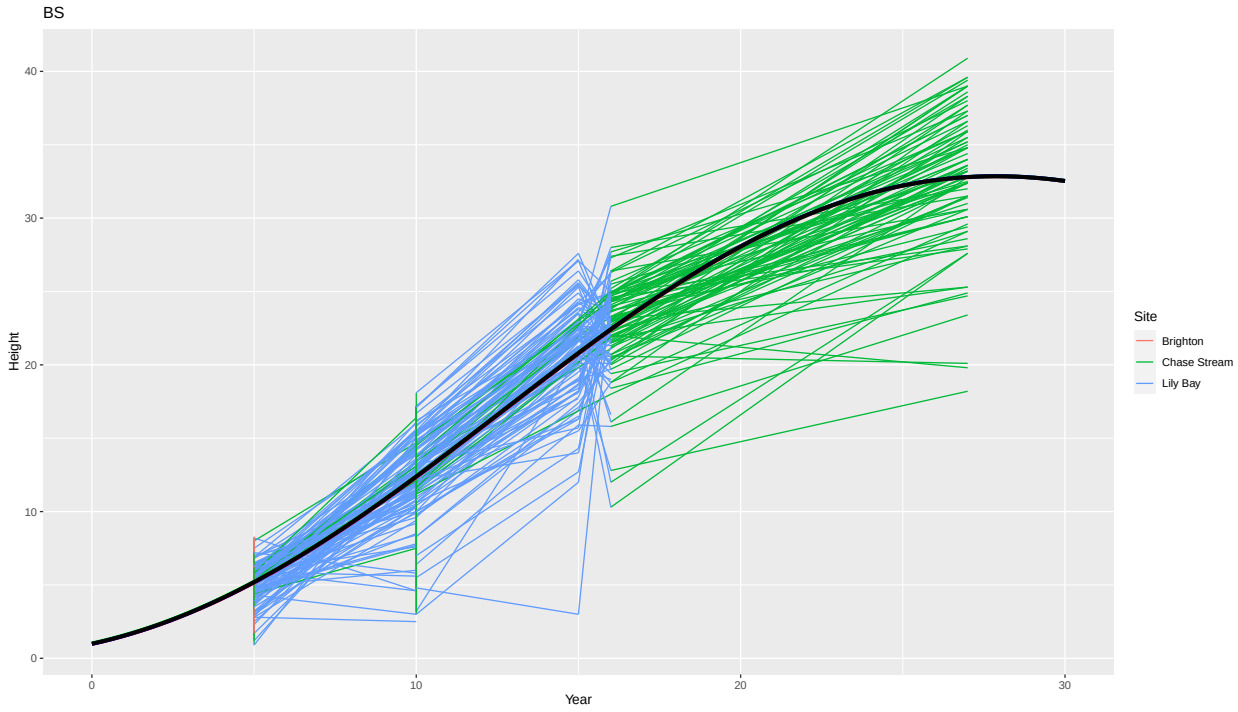


Figure 9: Species: BS

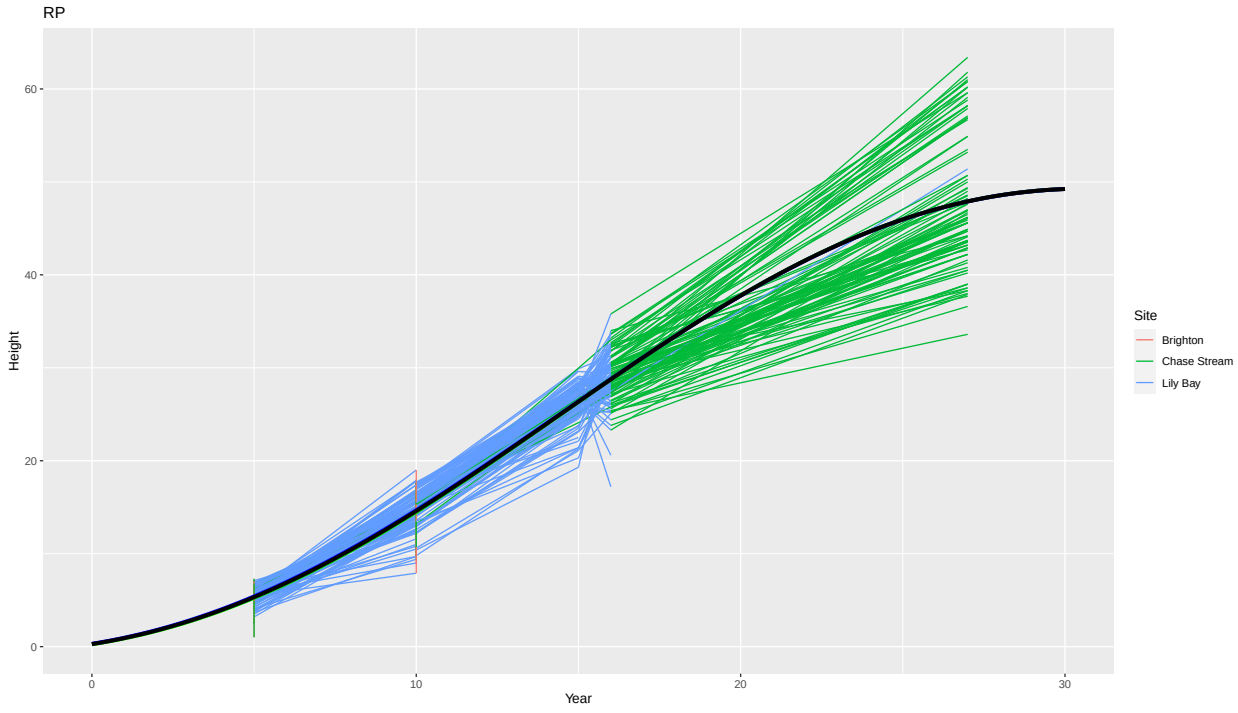


Figure 10: Species: BS

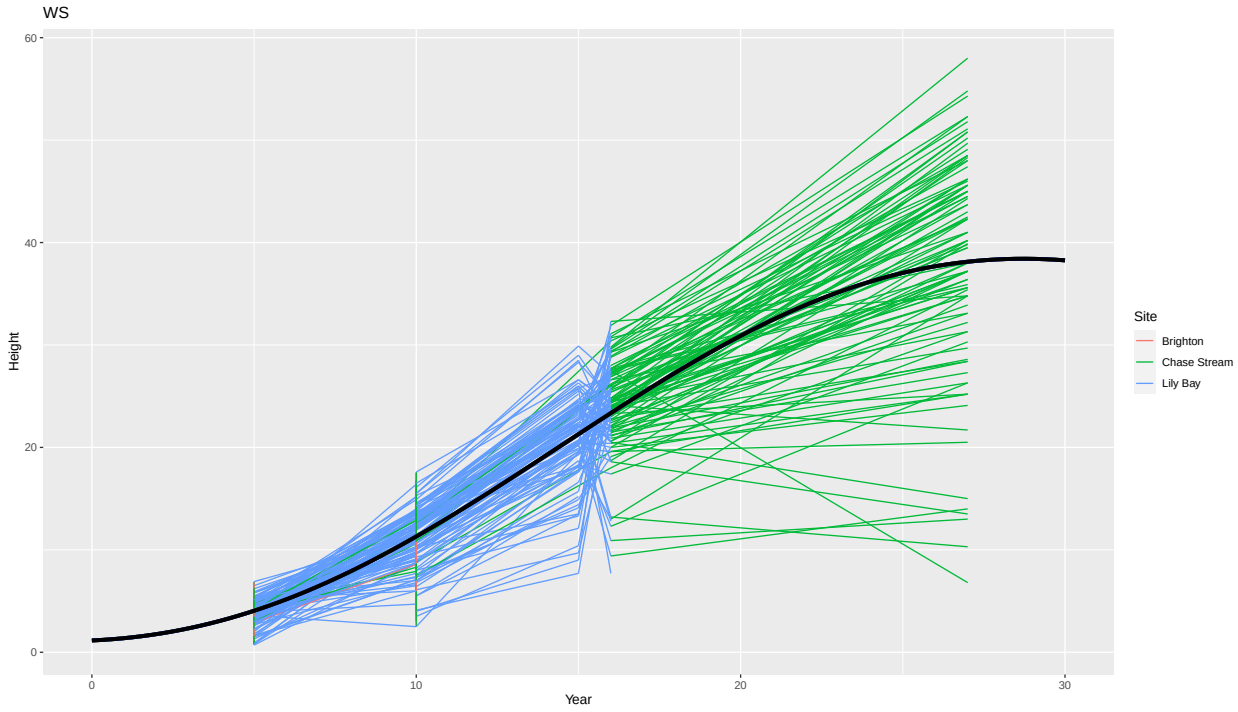


Figure 11: Species: BS

For EL and HL, we observev a nice separation of the site effects.

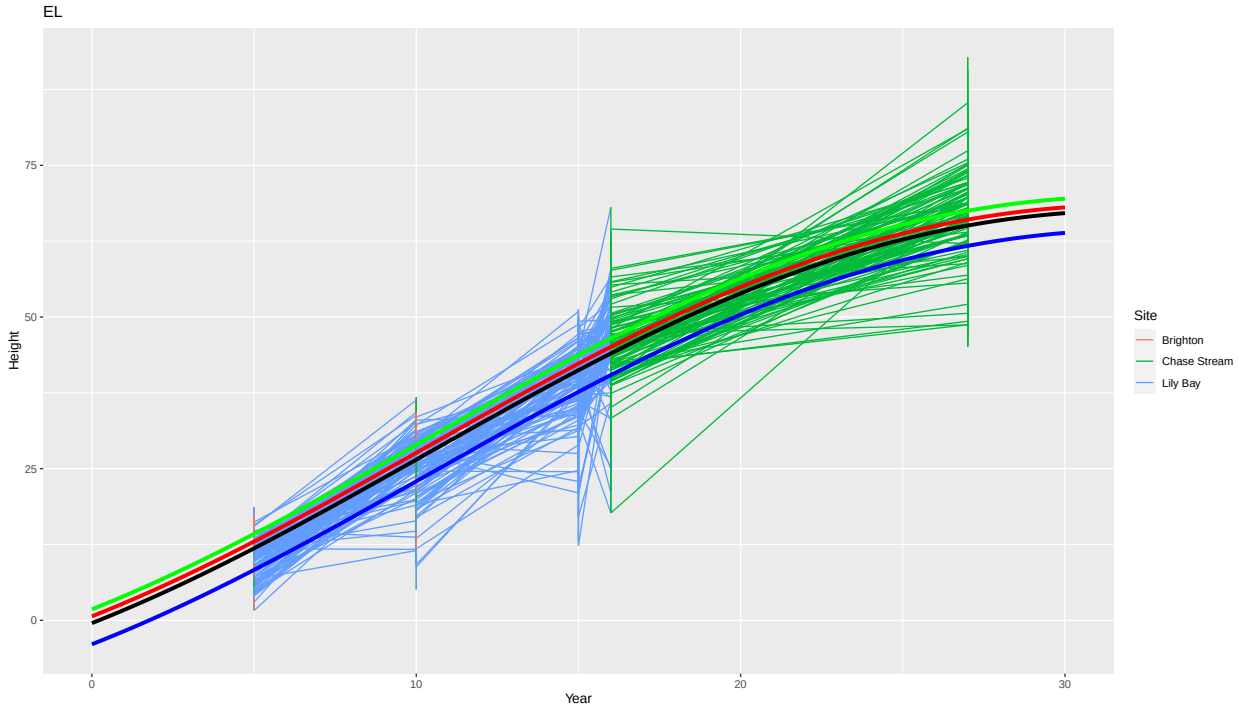


Figure 12: Species: EL

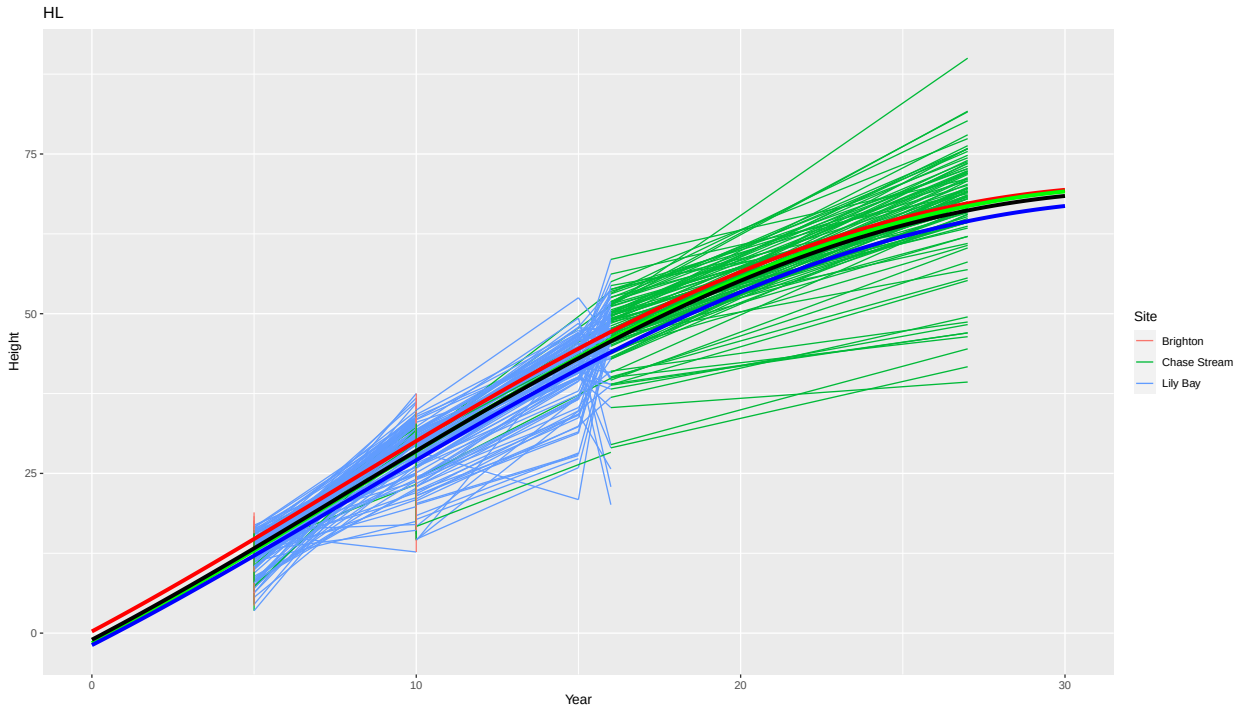


Figure 13: Species: EL

For species JL and TL, it appears that the overall model fit was not right in the first place. I am rerunning the optimization and model fit to check it.

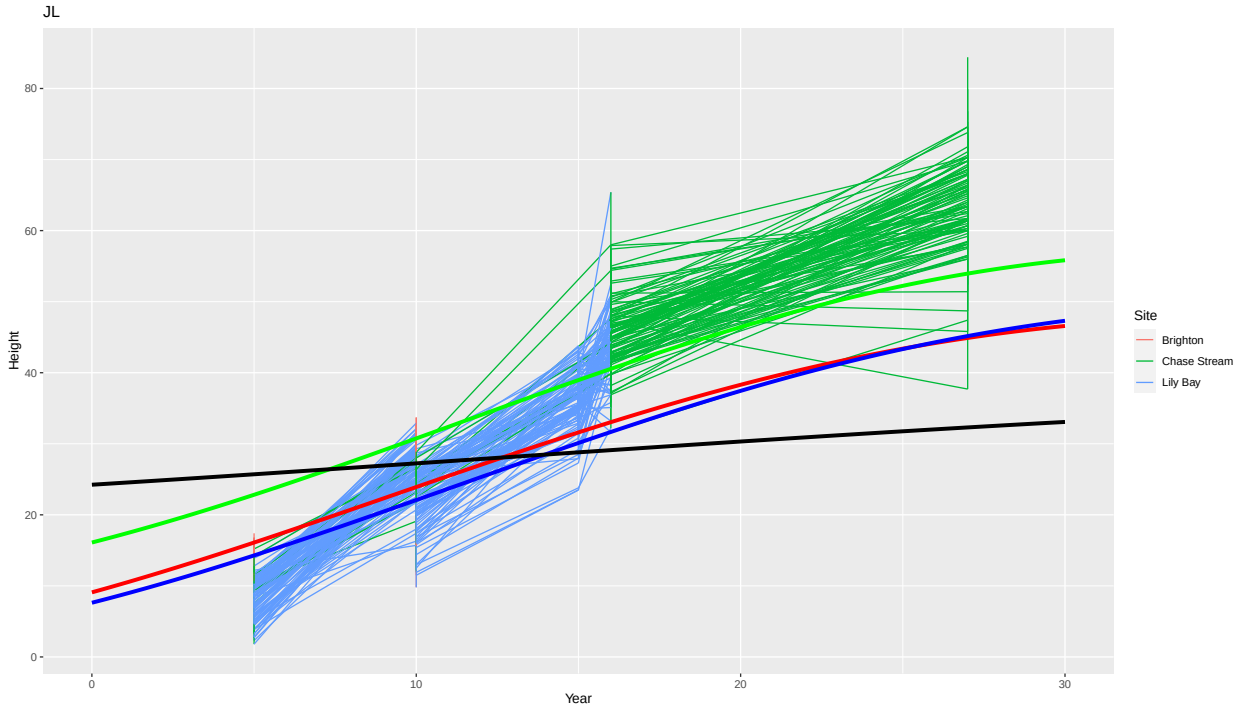


Figure 14: Species: JL

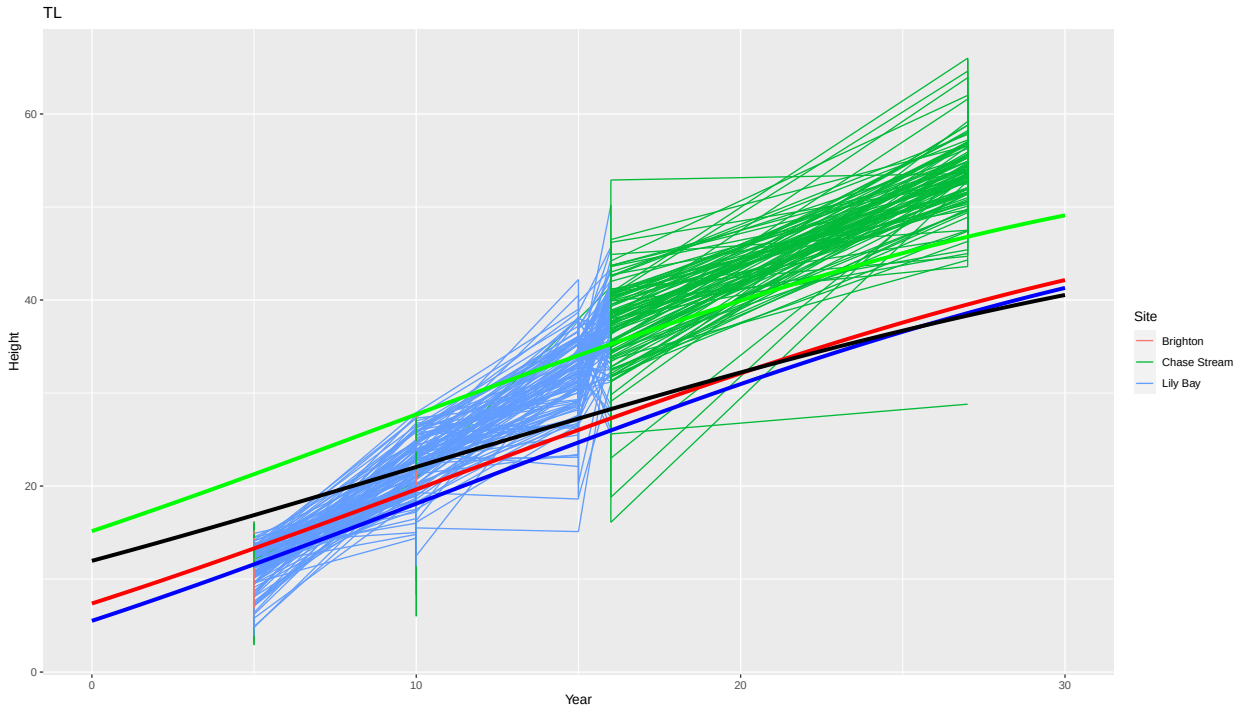


Figure 15: Species: JL

Also for JP, it appears that there are some issues with site specific fit for Brighton and Chase Stream.

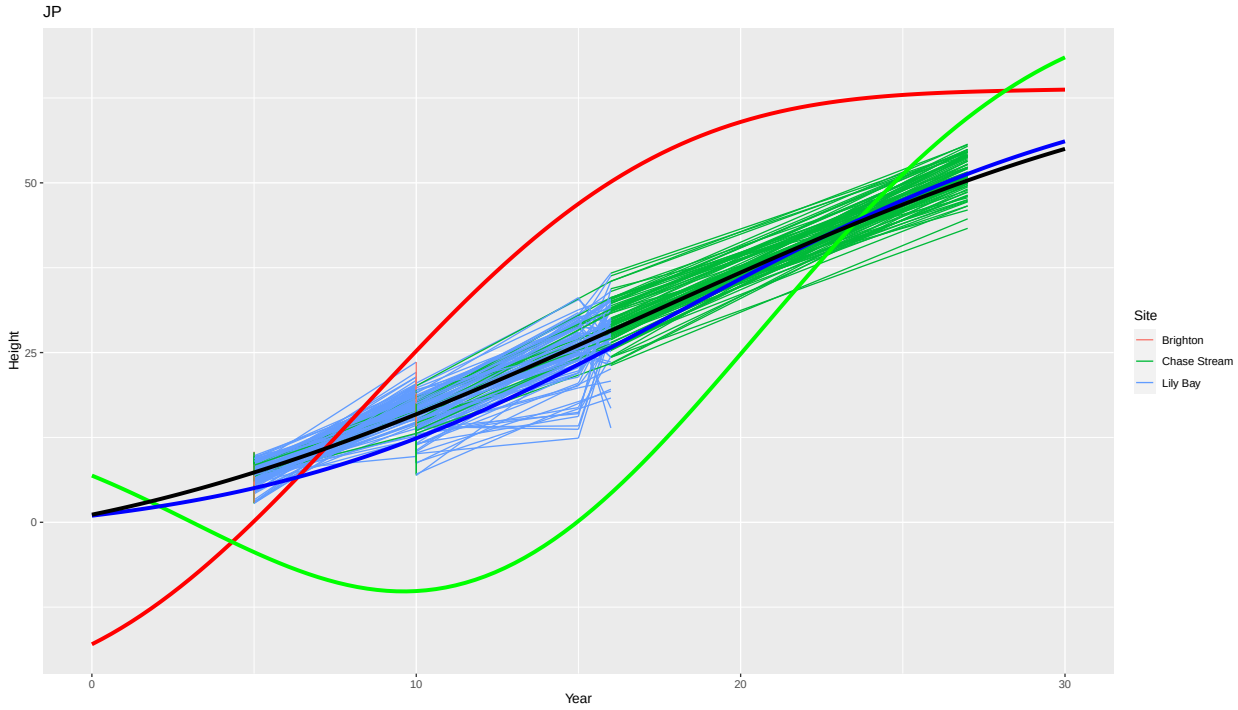


Figure 16: Species: JP