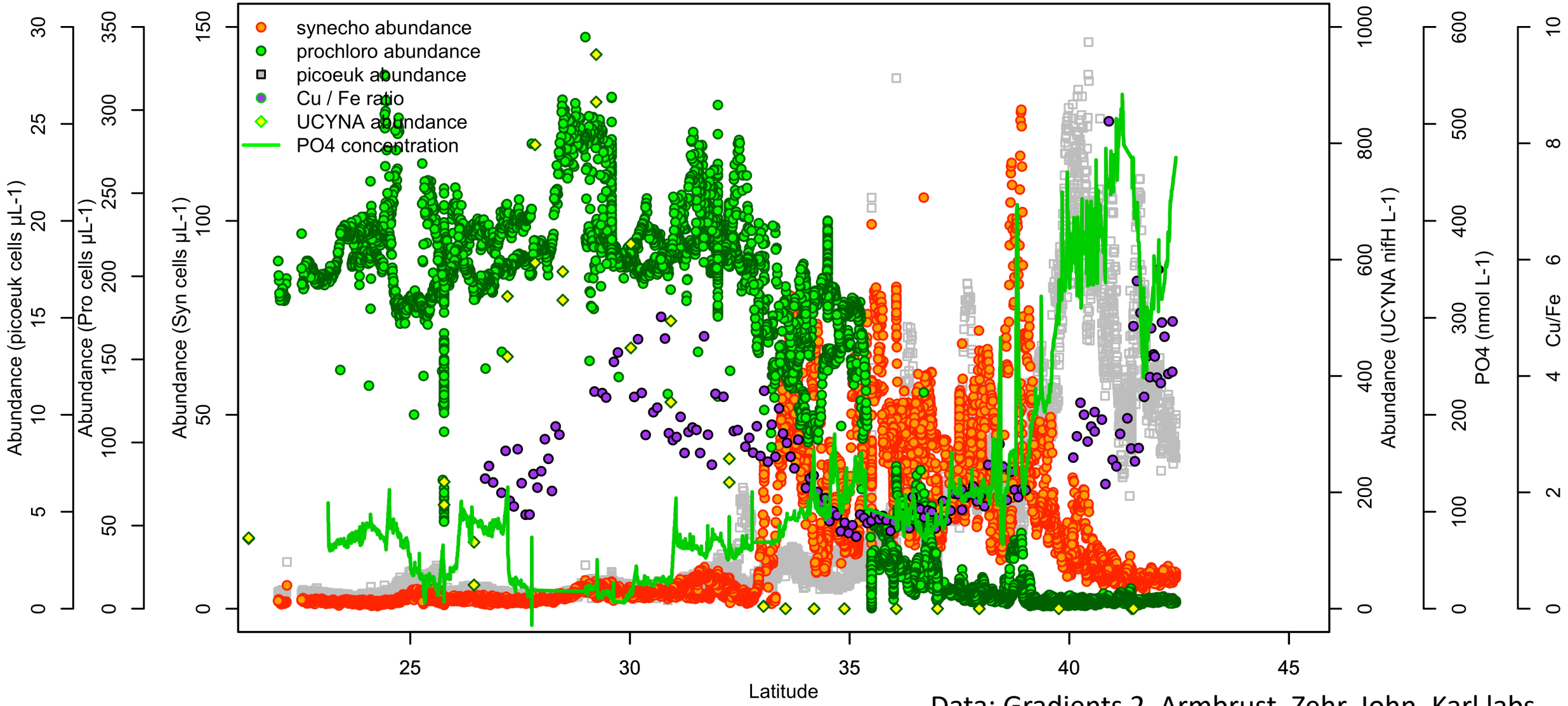
Explaining gradients in Gradients 2: statistical models of breakpoints

Jonathan Bradet-Legris
Andrew Irwin
Zoe Finkel
Dalhousie University
And the Gradients team!

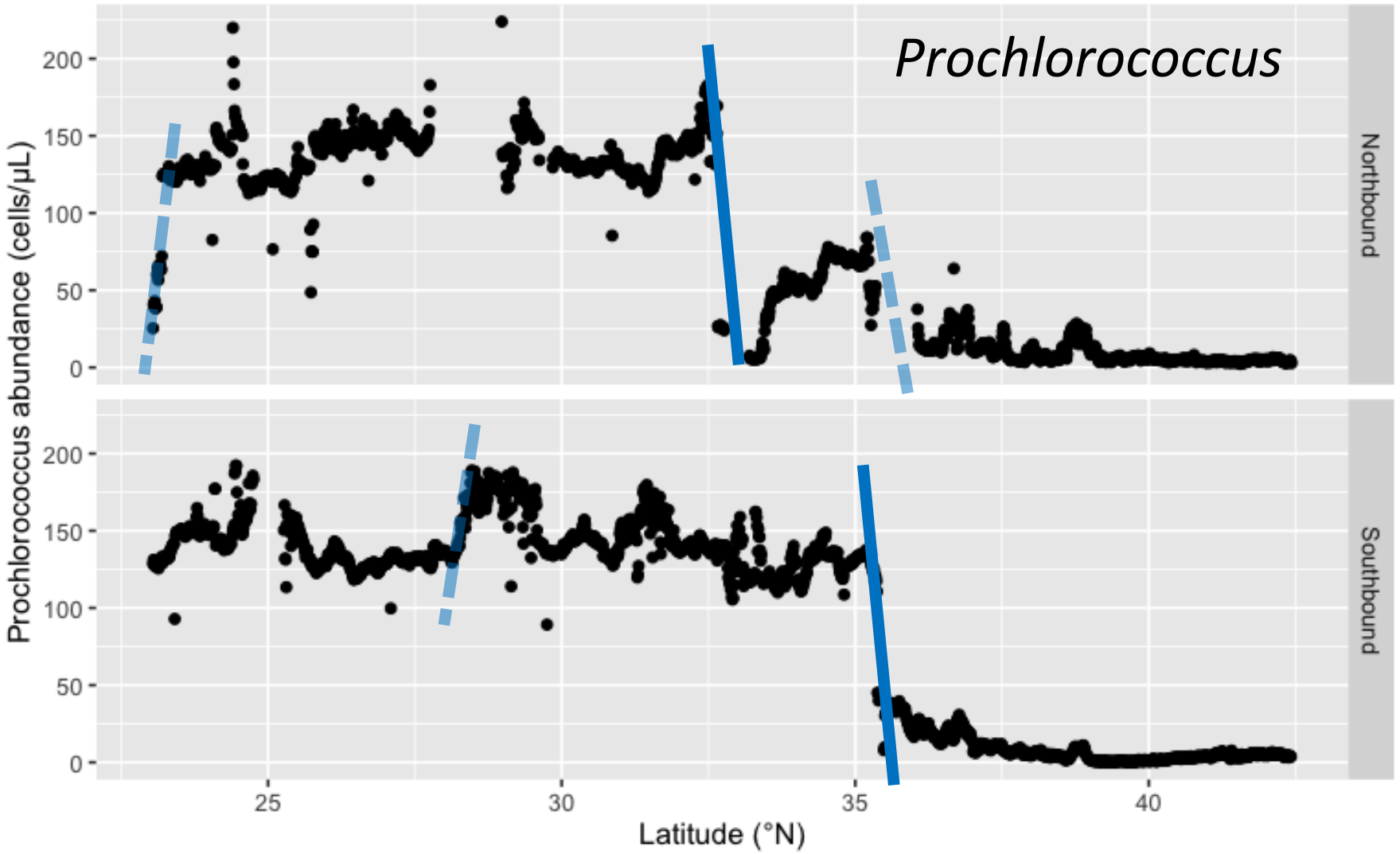
Gradients workshop, Seattle – October 2019

Detecting clines, breakpoints (4 ways)

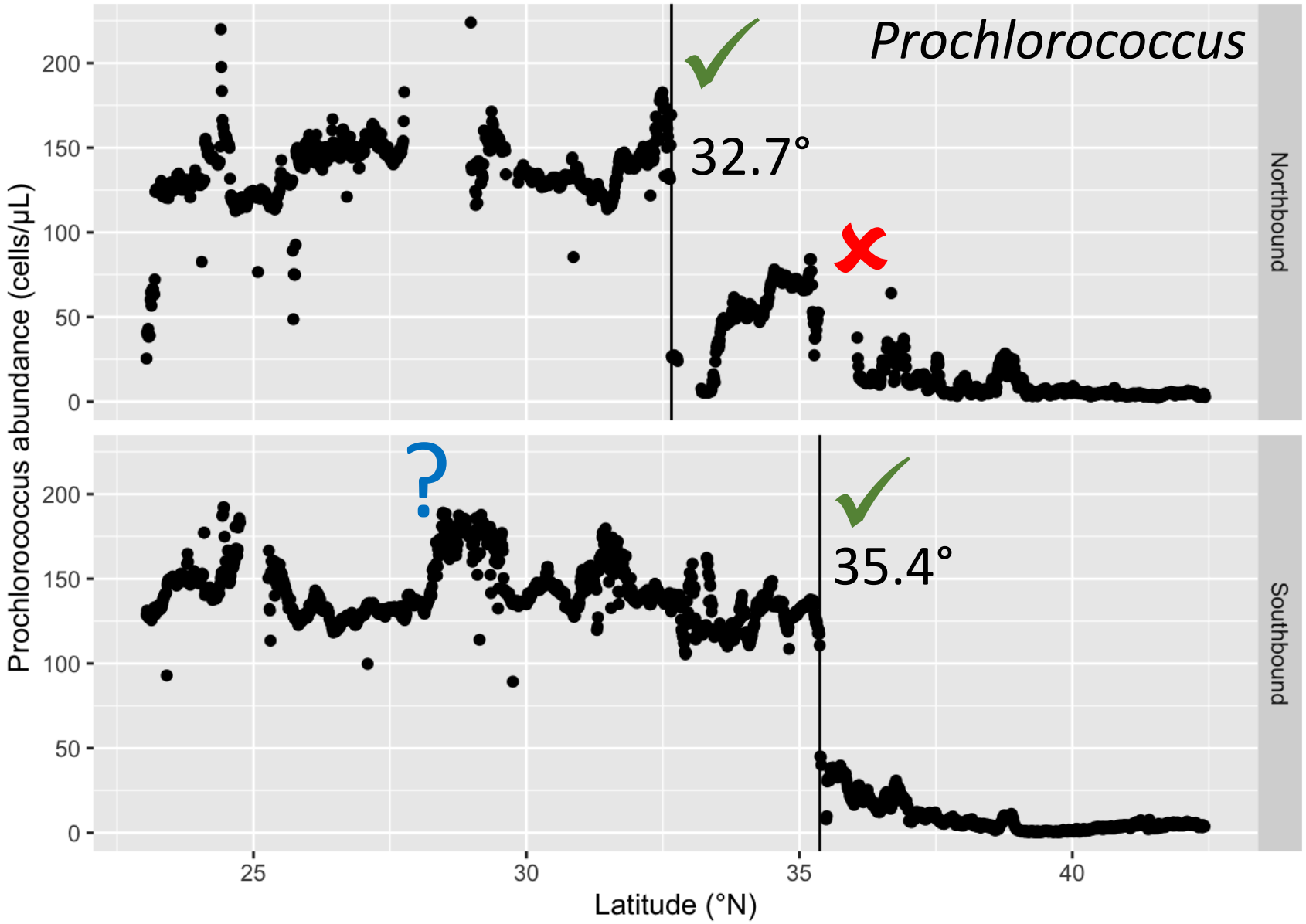


Data: Gradients 2. Armbrust, Zehr, John, Karl labs

Second derivative method: detects large local slopes

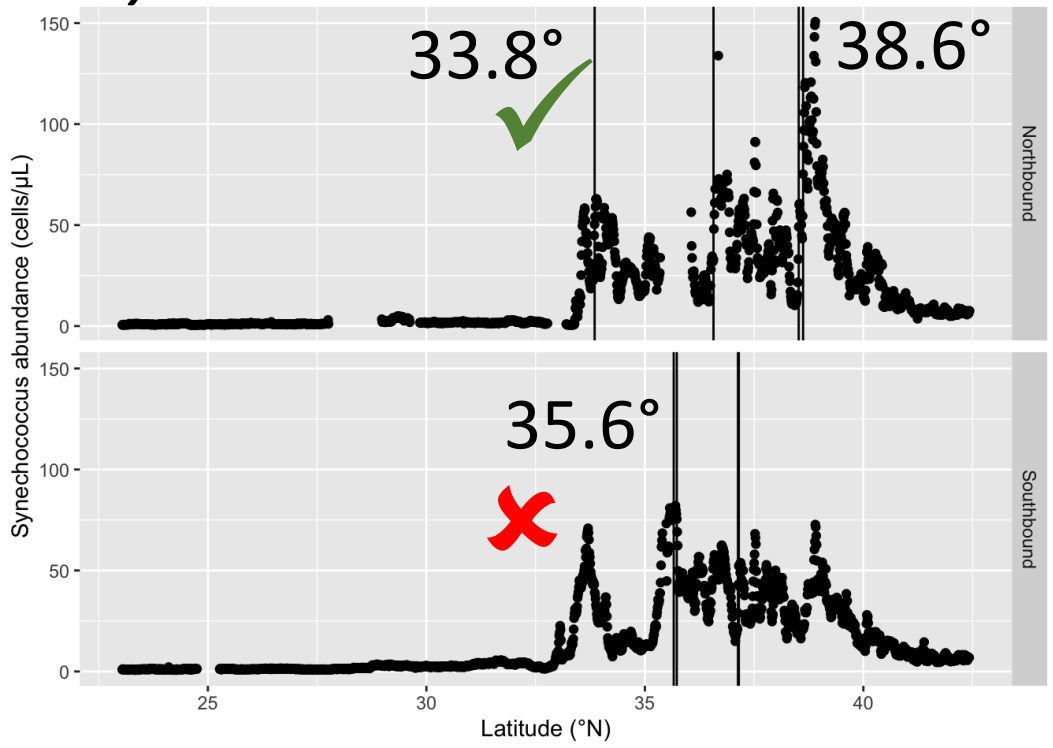


Second derivative method: results

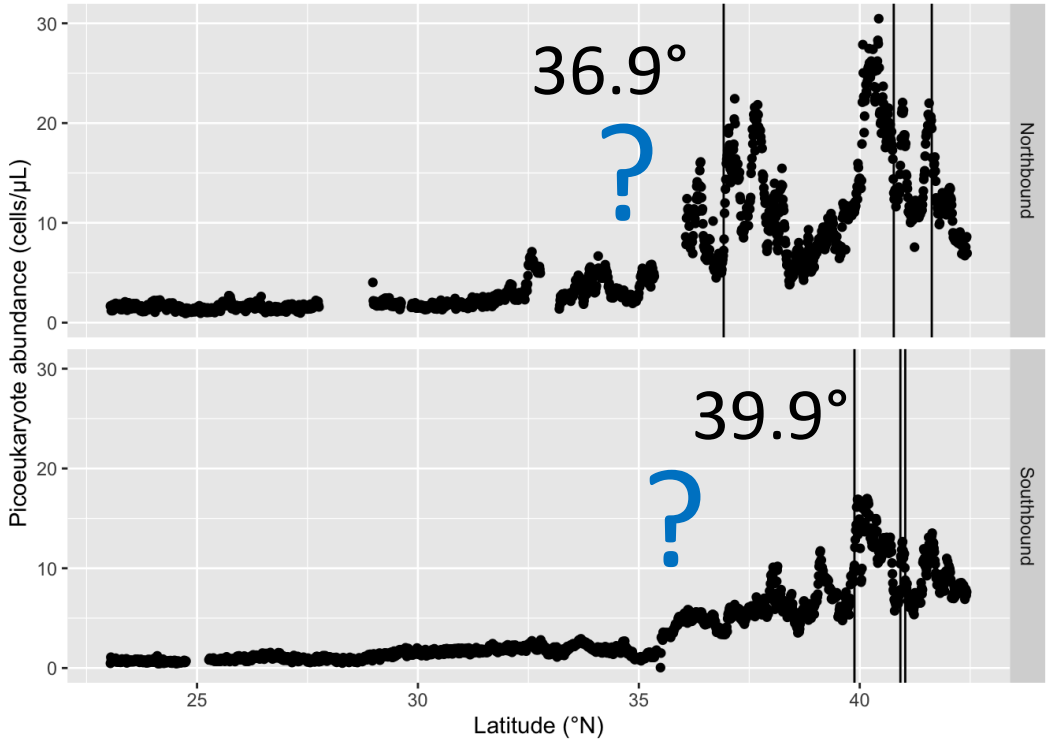


Second derivatives: results

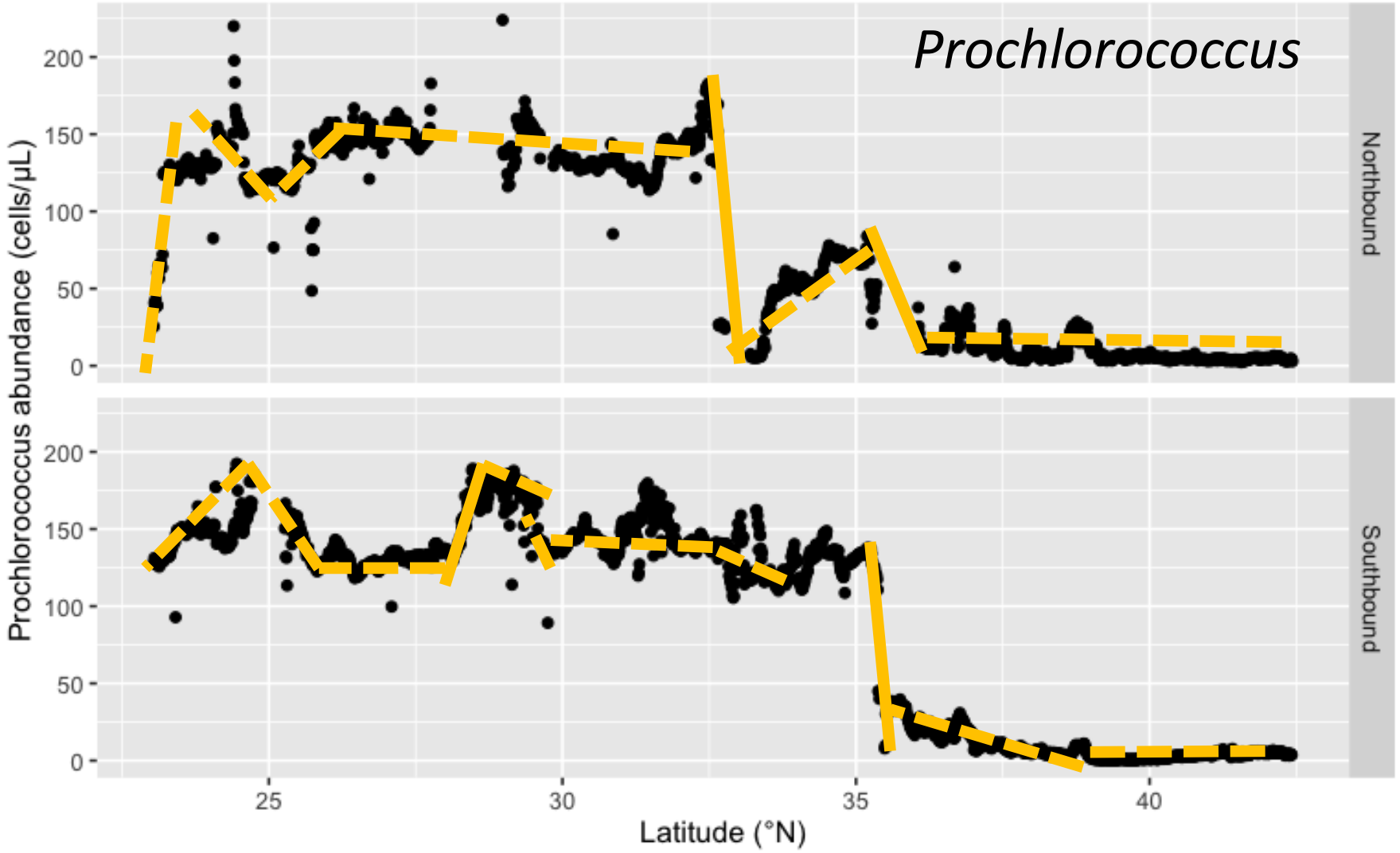
Synechococcus



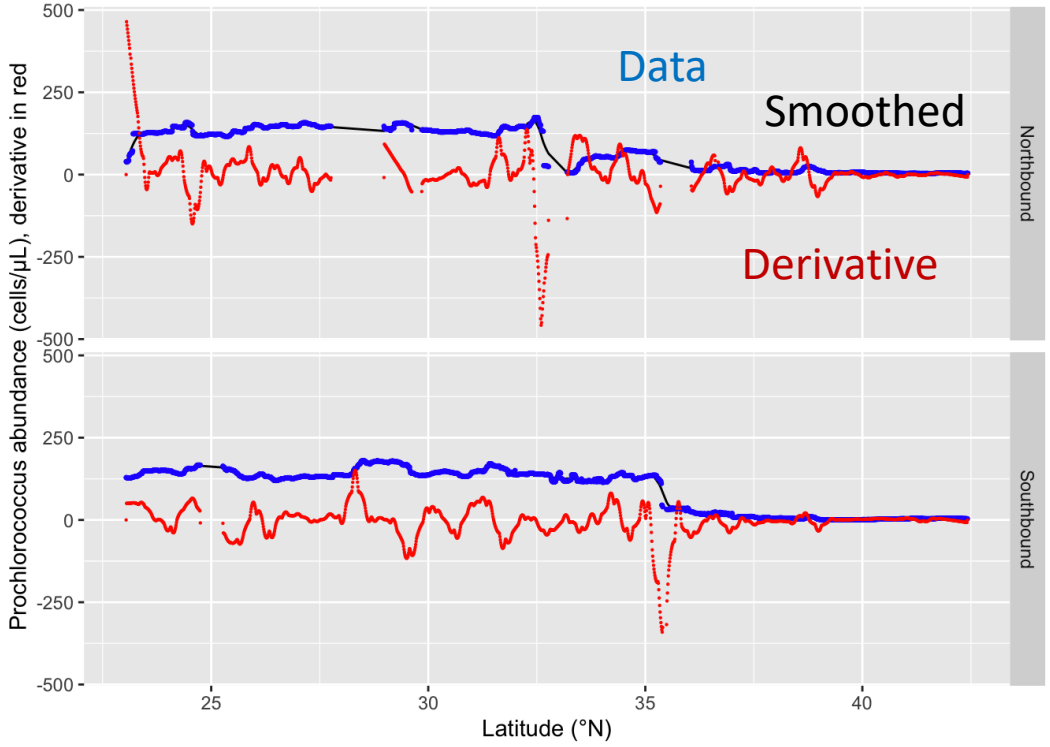
Picoeuks



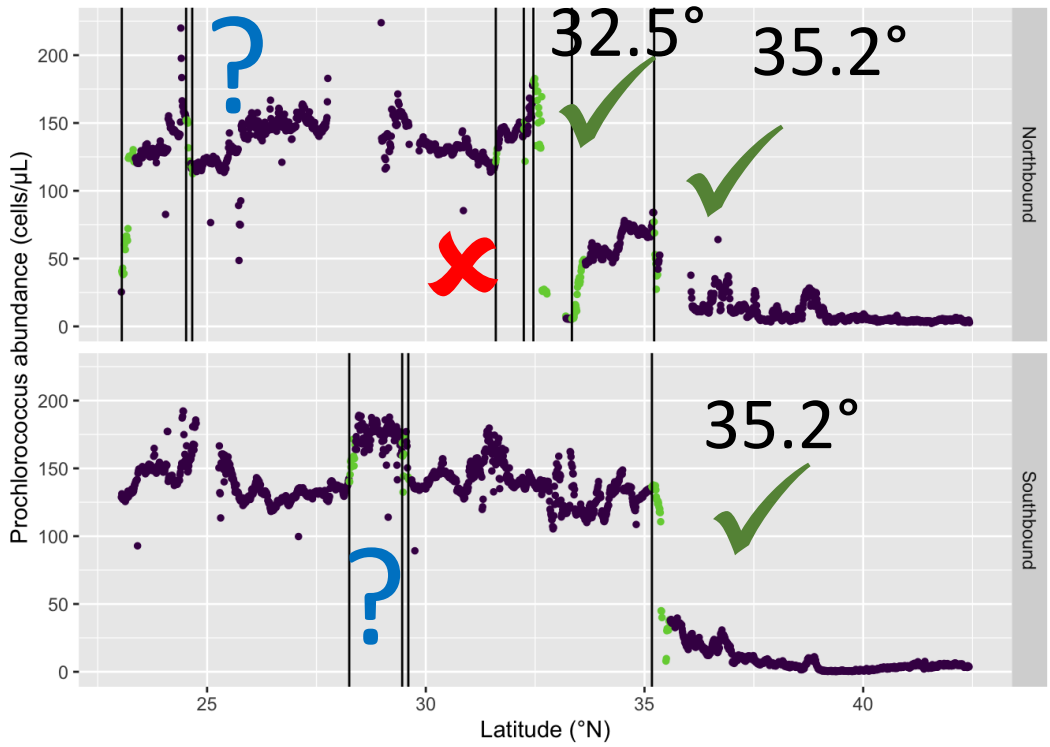
Loess – detects large local slopes



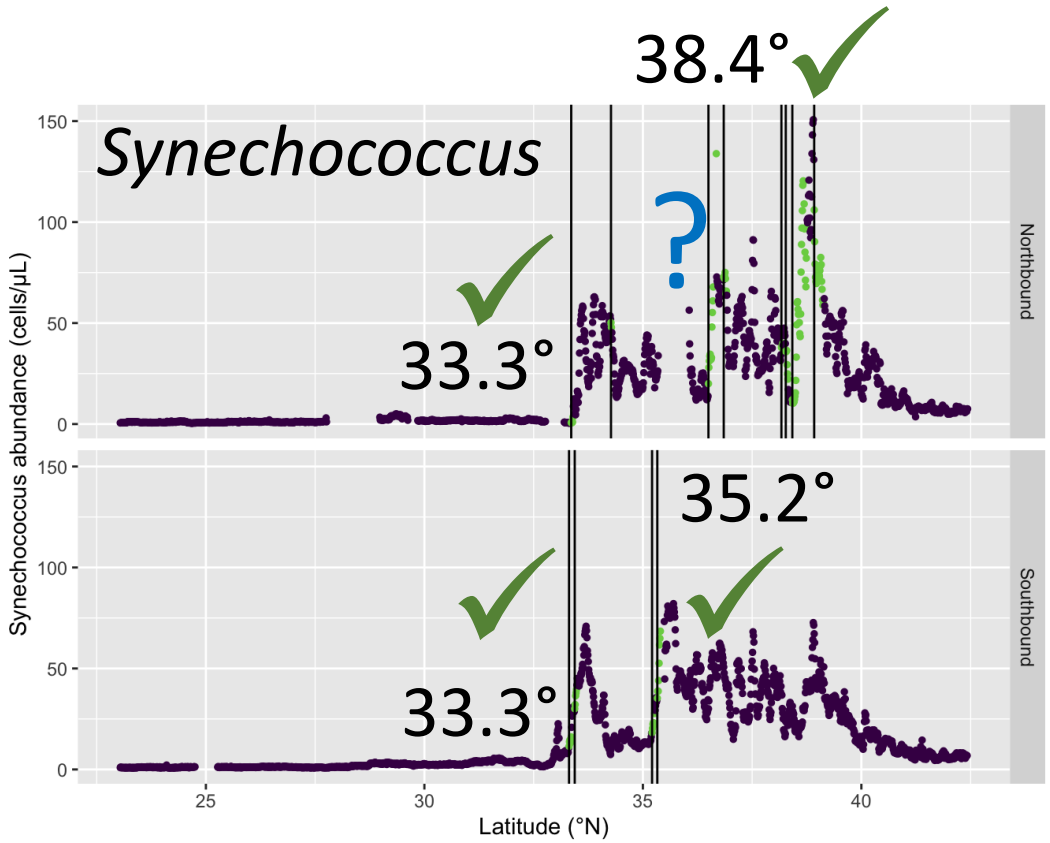
Loess – results

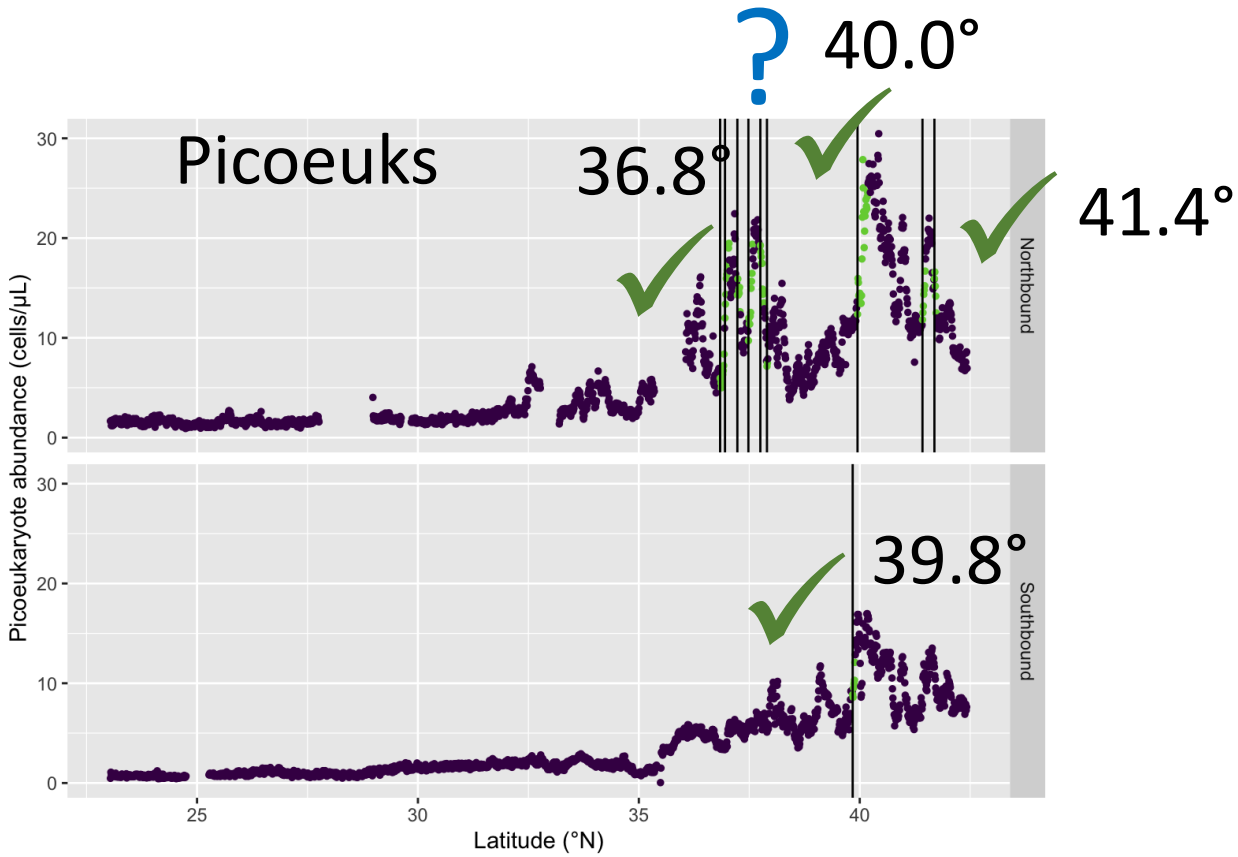


Prochlorococcus

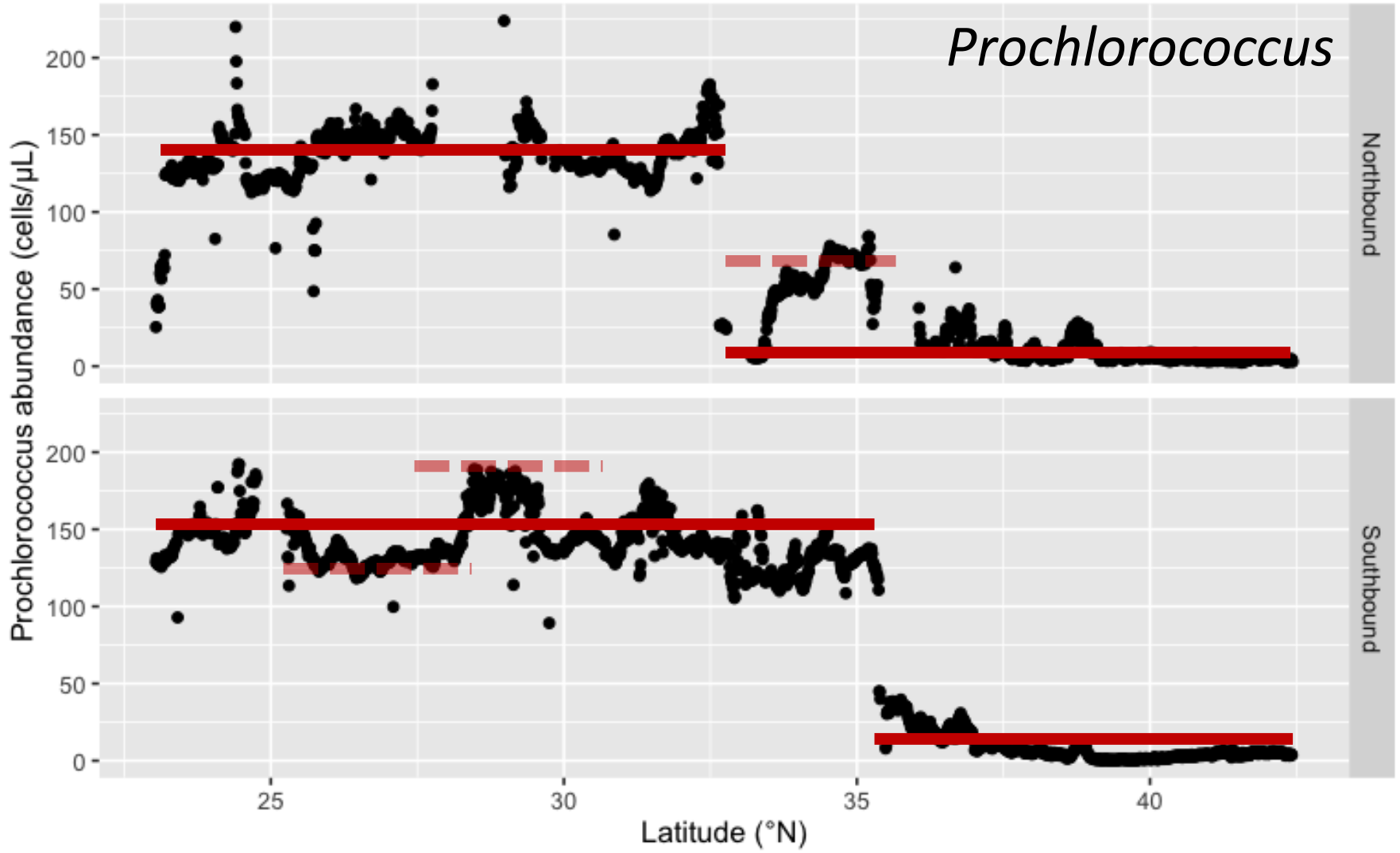


Loess – results

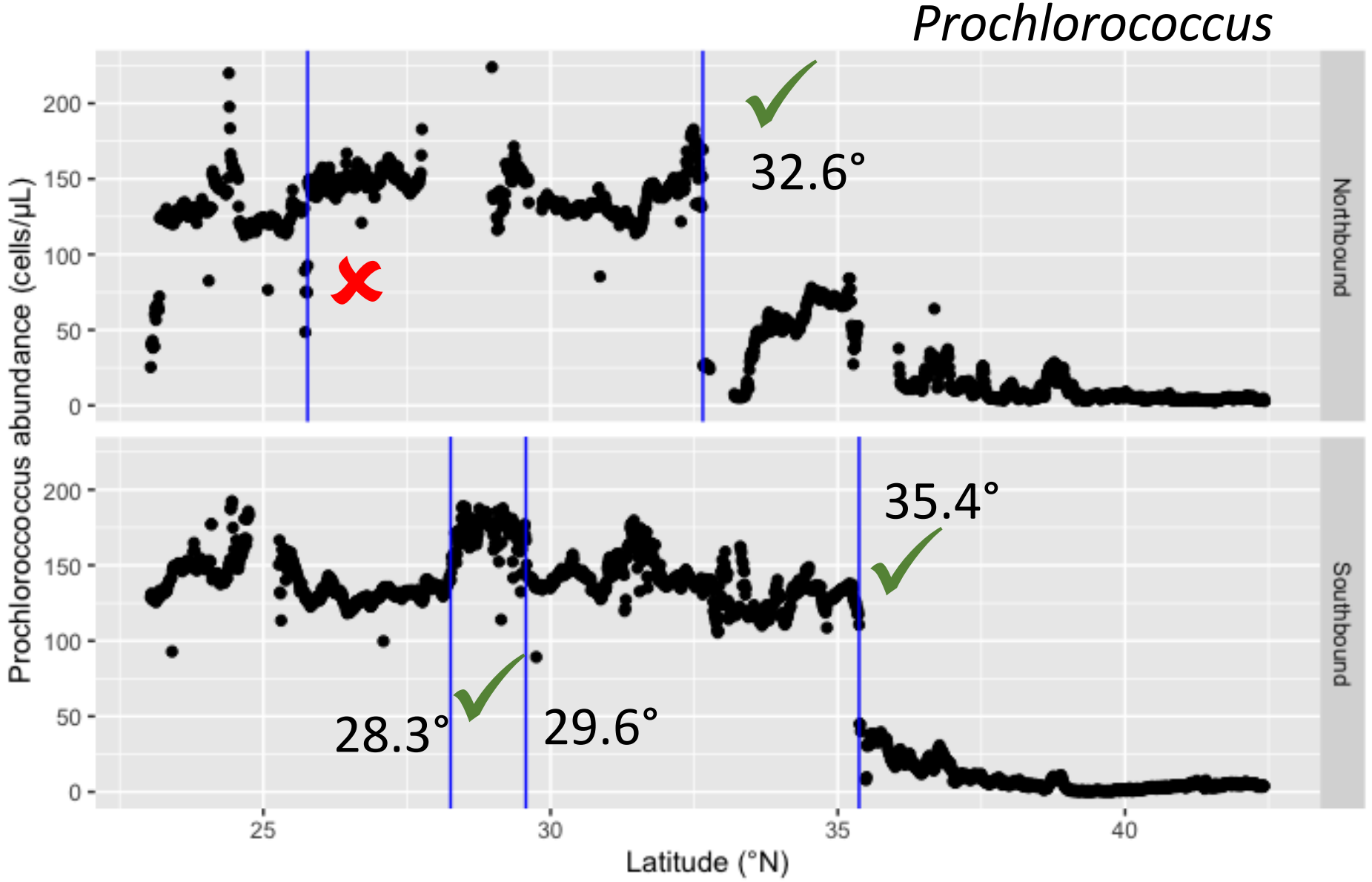




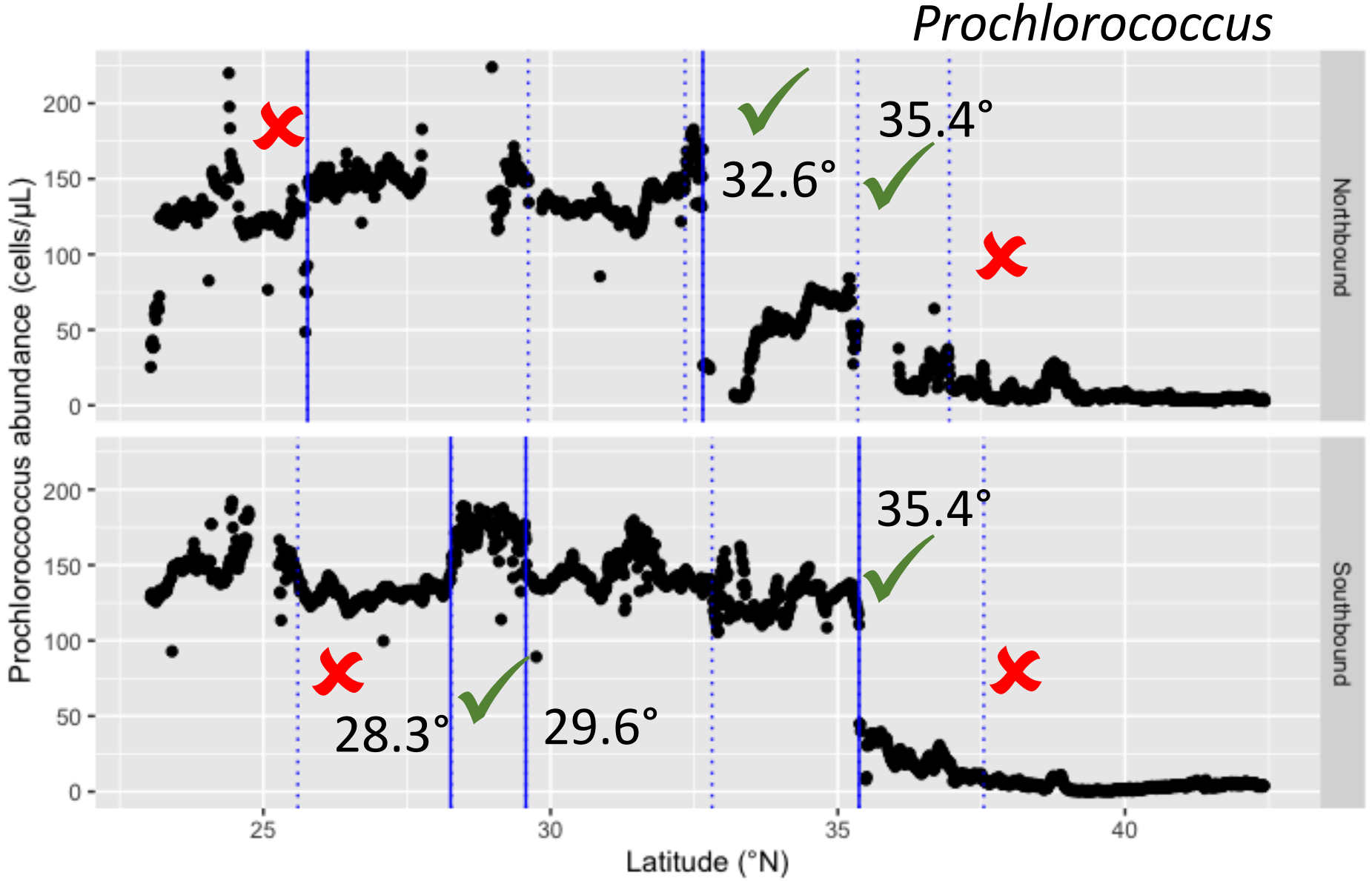
Variance inflation factors detects jumps in abundance



Variance inflation factors: results

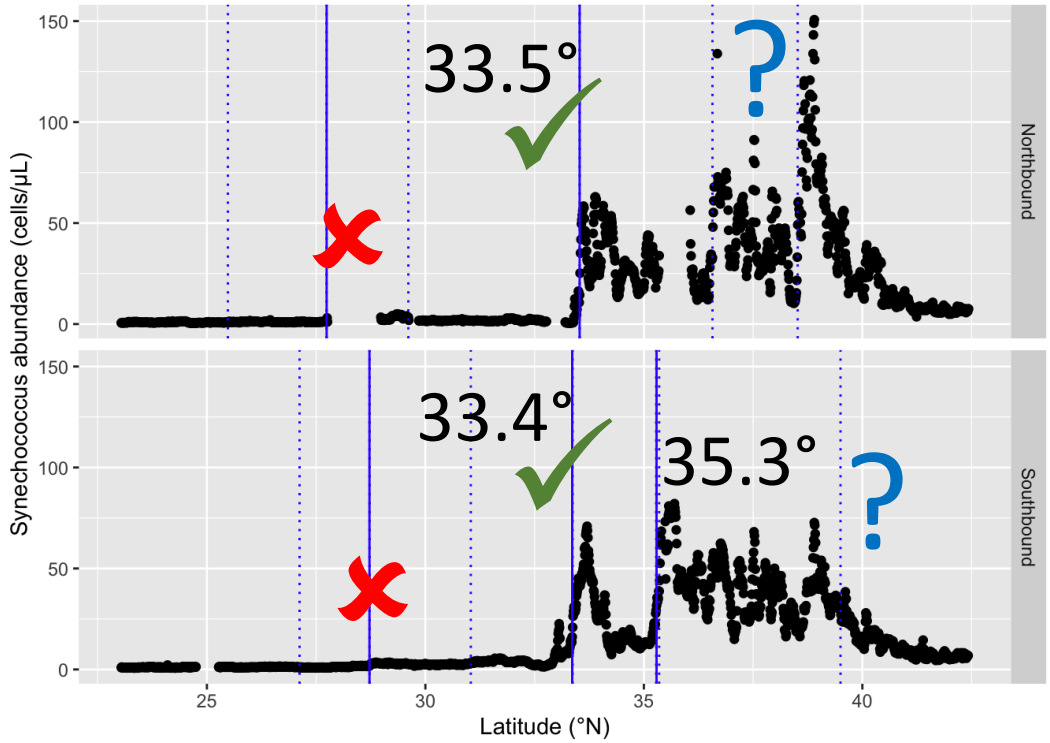


Variance inflation factors: results

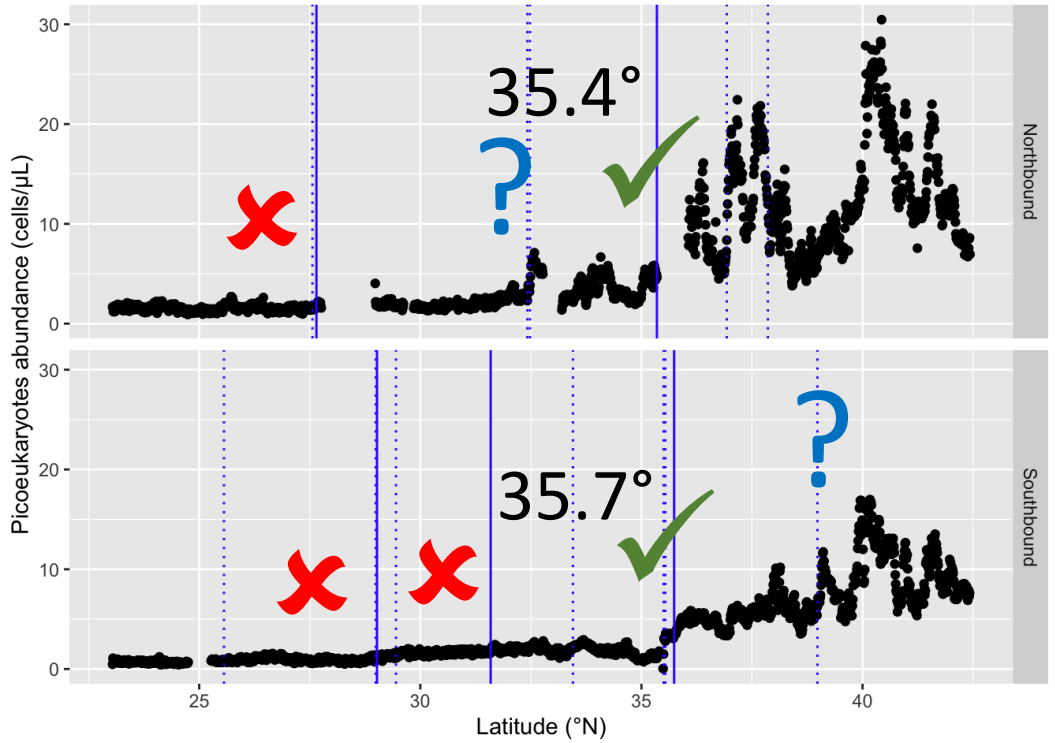


Variance inflation factors: results

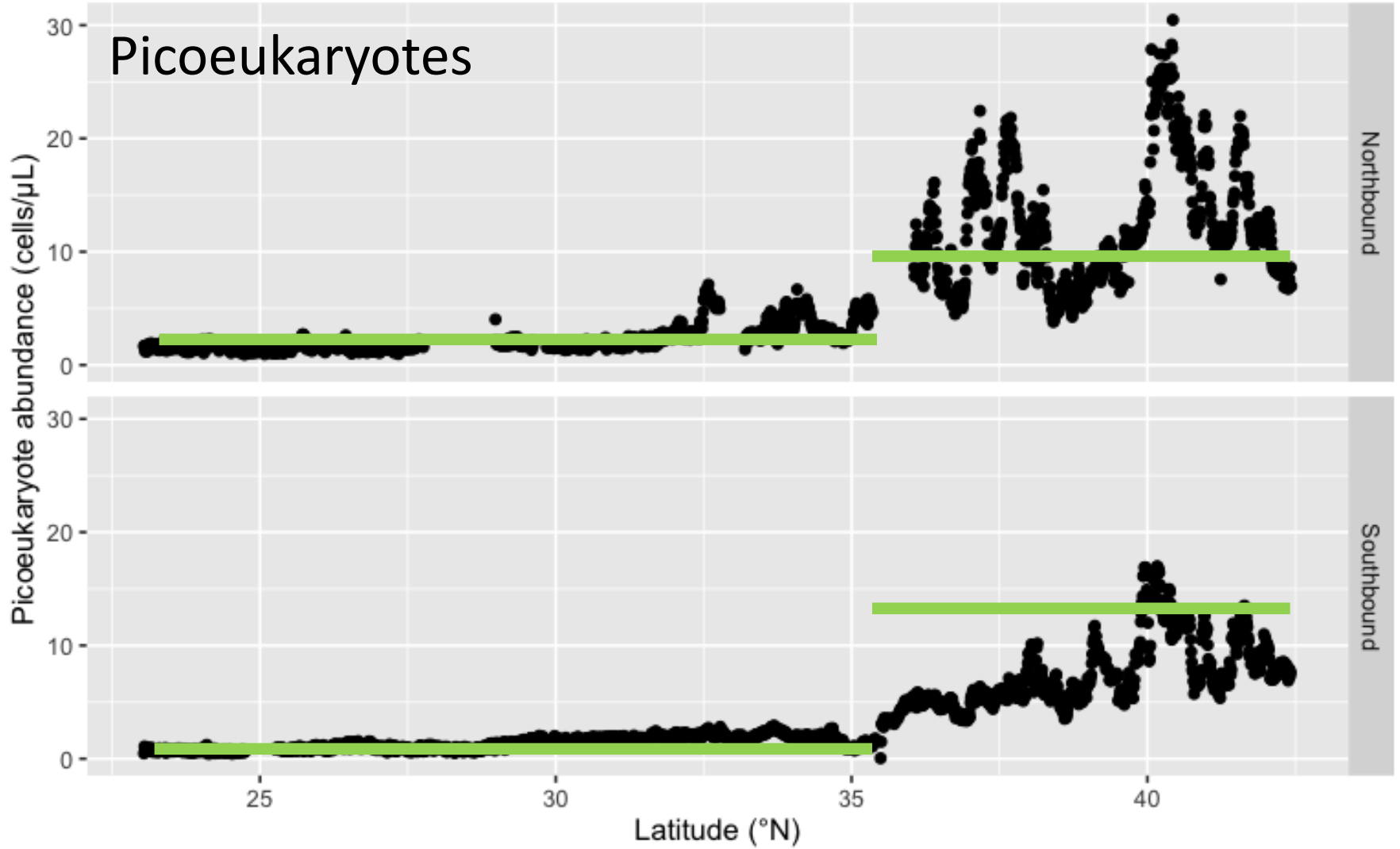
Synechococcus



Picoeuks

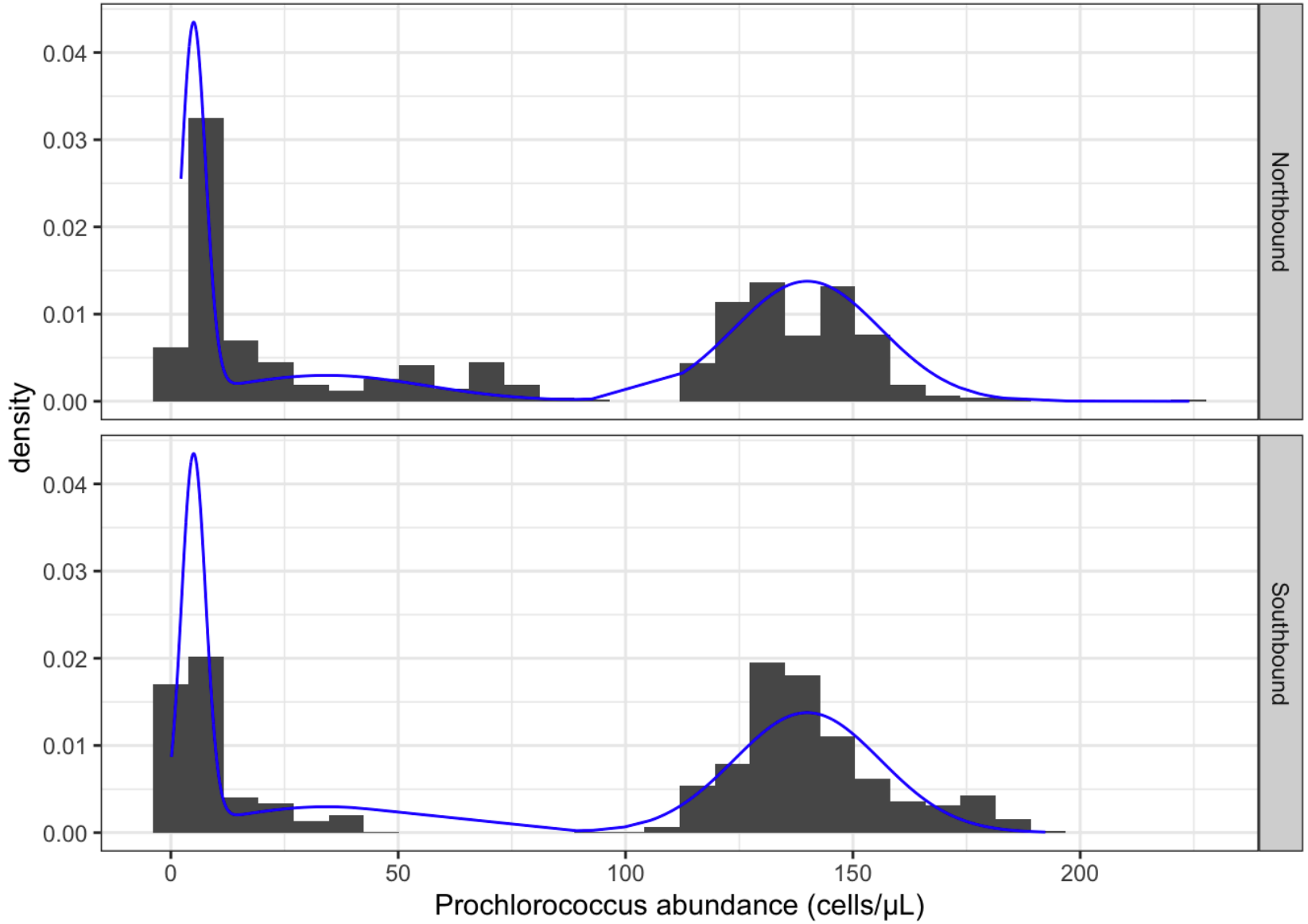


Gaussian mixtures – detects regime shifts



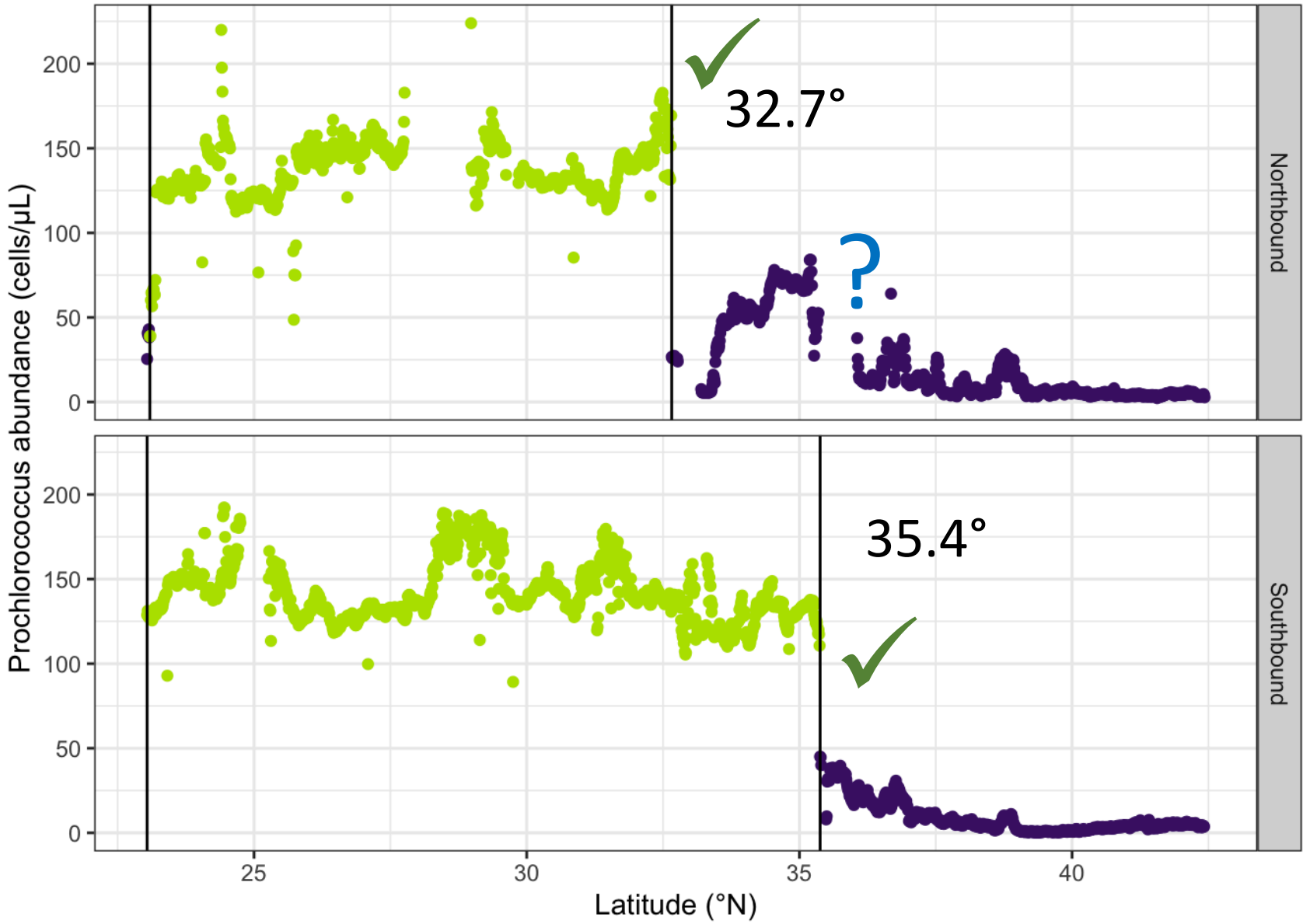
Gaussian mixtures: results

Prochlorococcus



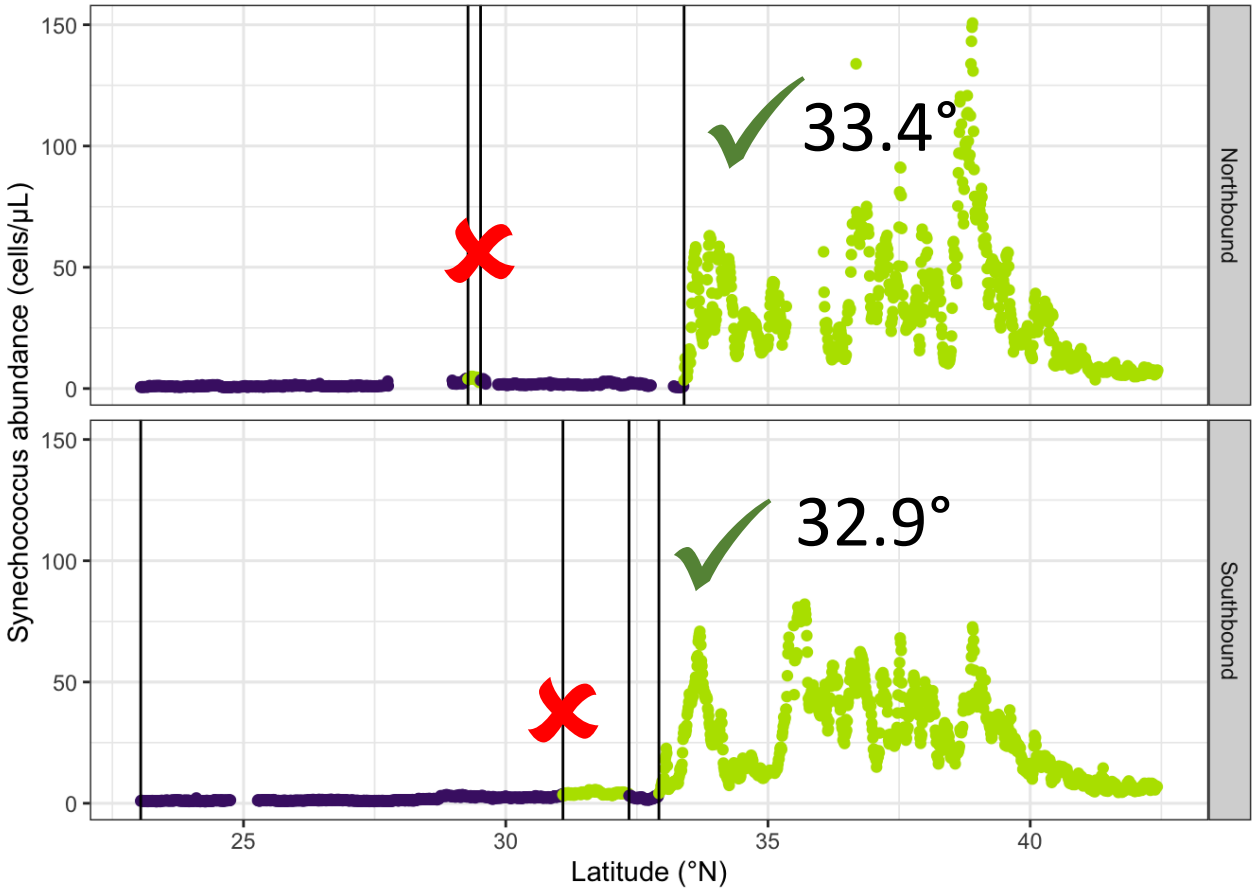
Gaussian mixtures: results

Prochlorococcus

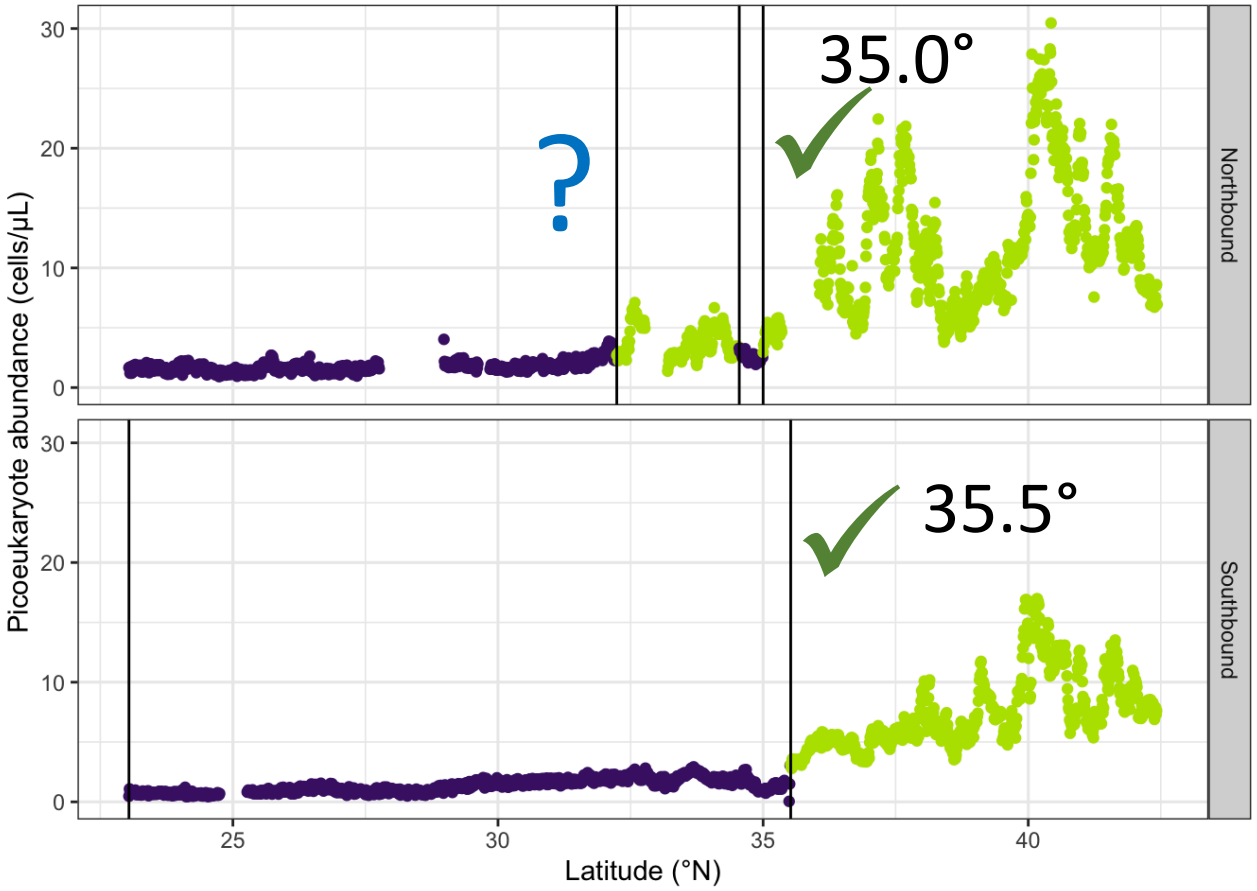


Gaussian mixture: results

Synechococcus



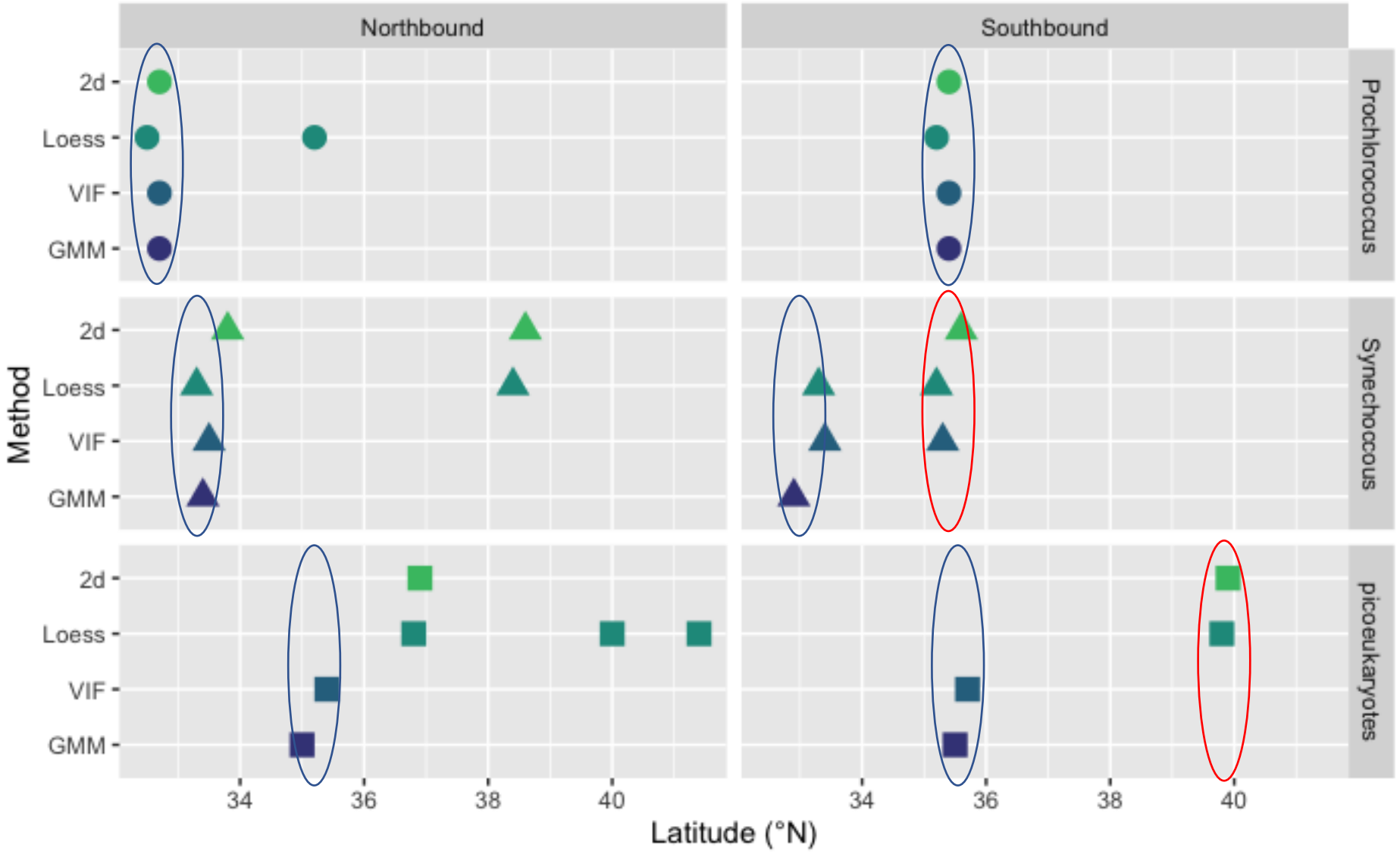
Picoeuks



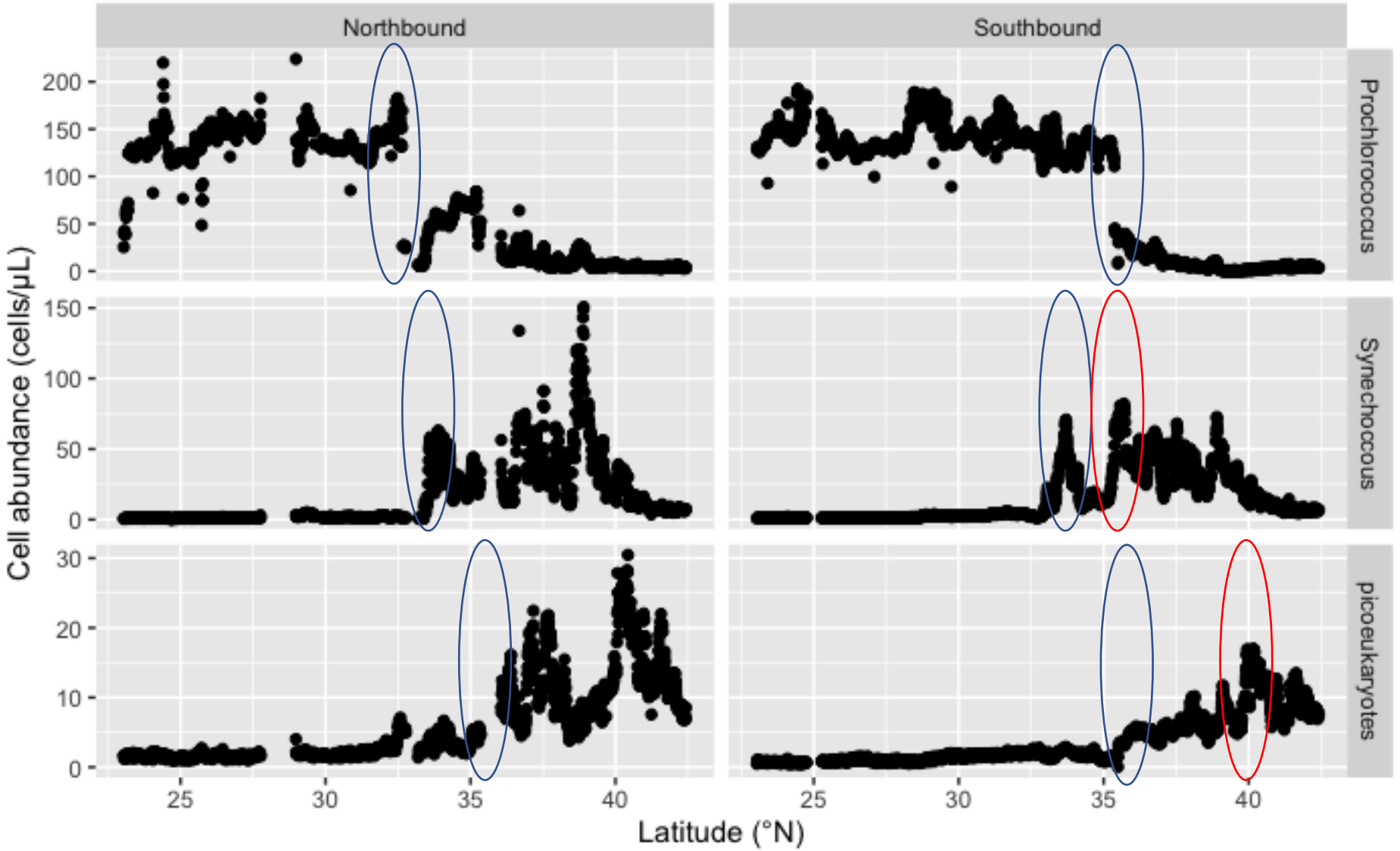
Breakpoint summary

	Second derivative		Loess		VIF		GMM	
	Northbound	Southbound	Northbound	Southbound	Northbound	Southbound	Northbound	Southbound
Prochloro-coccus	32.7	35.4	32.5 35.2	35.2	32.6 *35.4*	35.4	32.7	35.4
Synecho-coccus	33.8 38.6	35.6	33.3 38.4	33.3 35.2	33.5	33.4 35.3	33.4	32.9
Pico-eukaryotes	36.9	39.9	36.8 40.0 41.4	39.8	35.4	35.7	35.0	35.5

Breakpoint summary – highlighting species differences



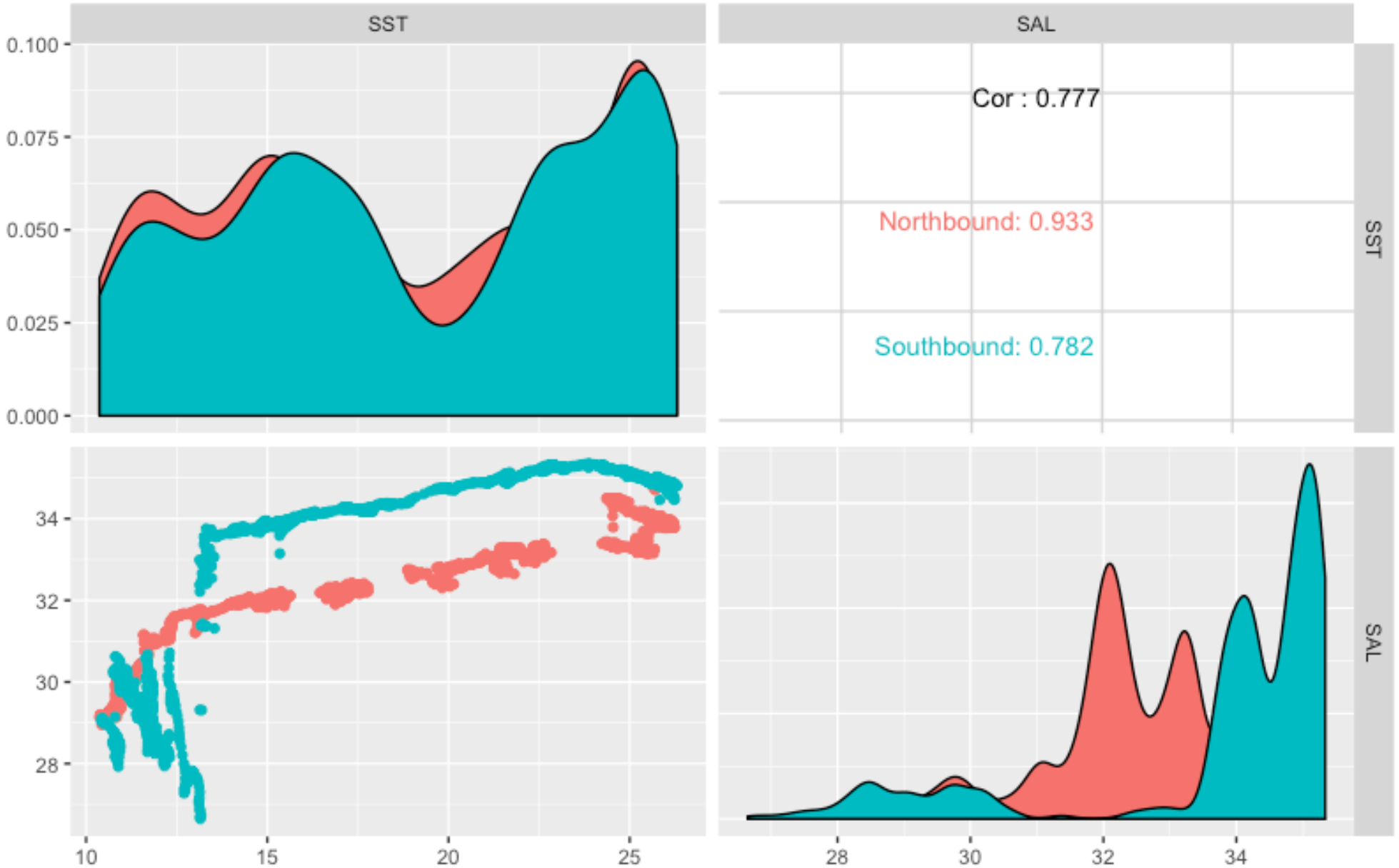
Breakpoint summary



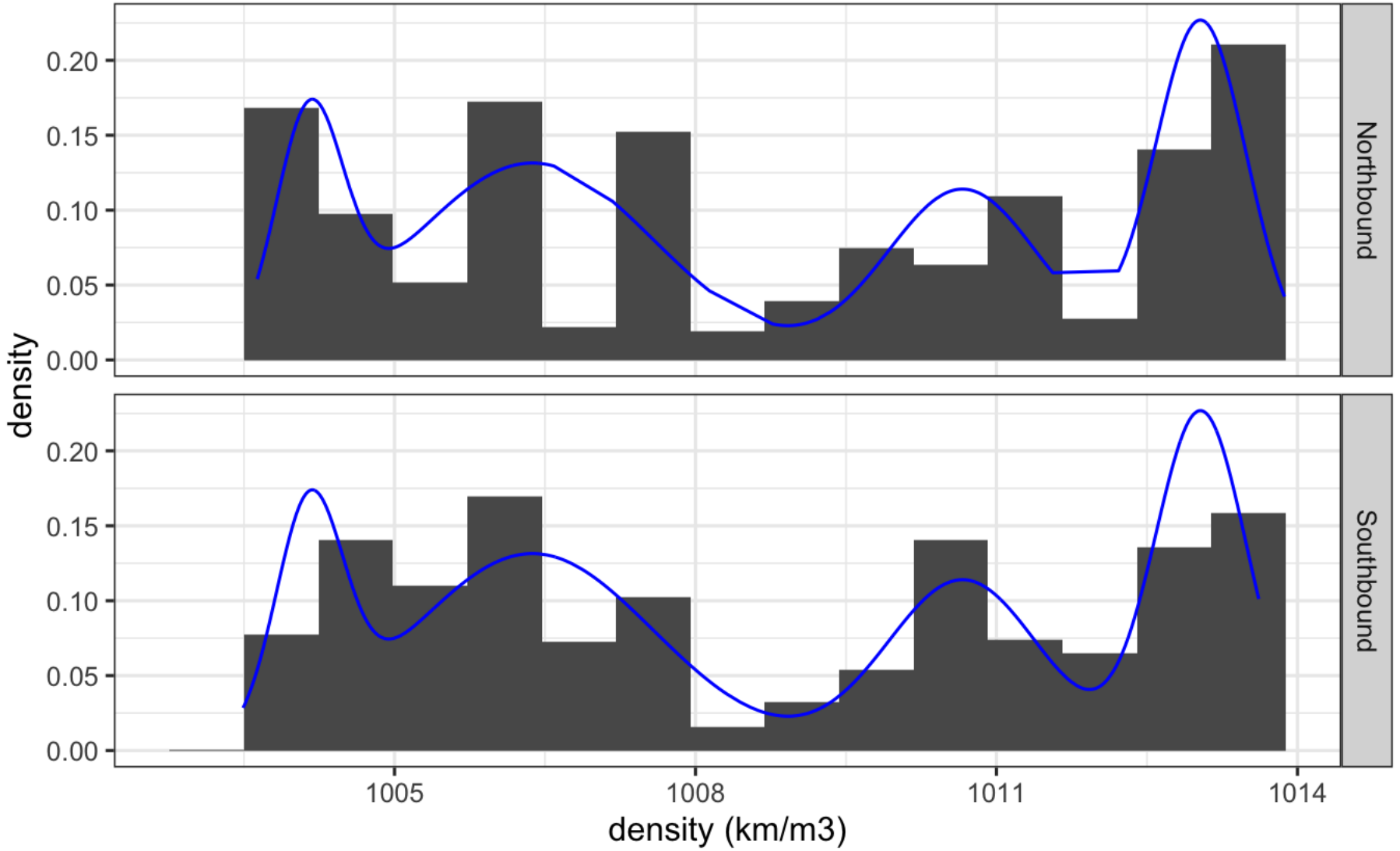
Recommendations & next steps…

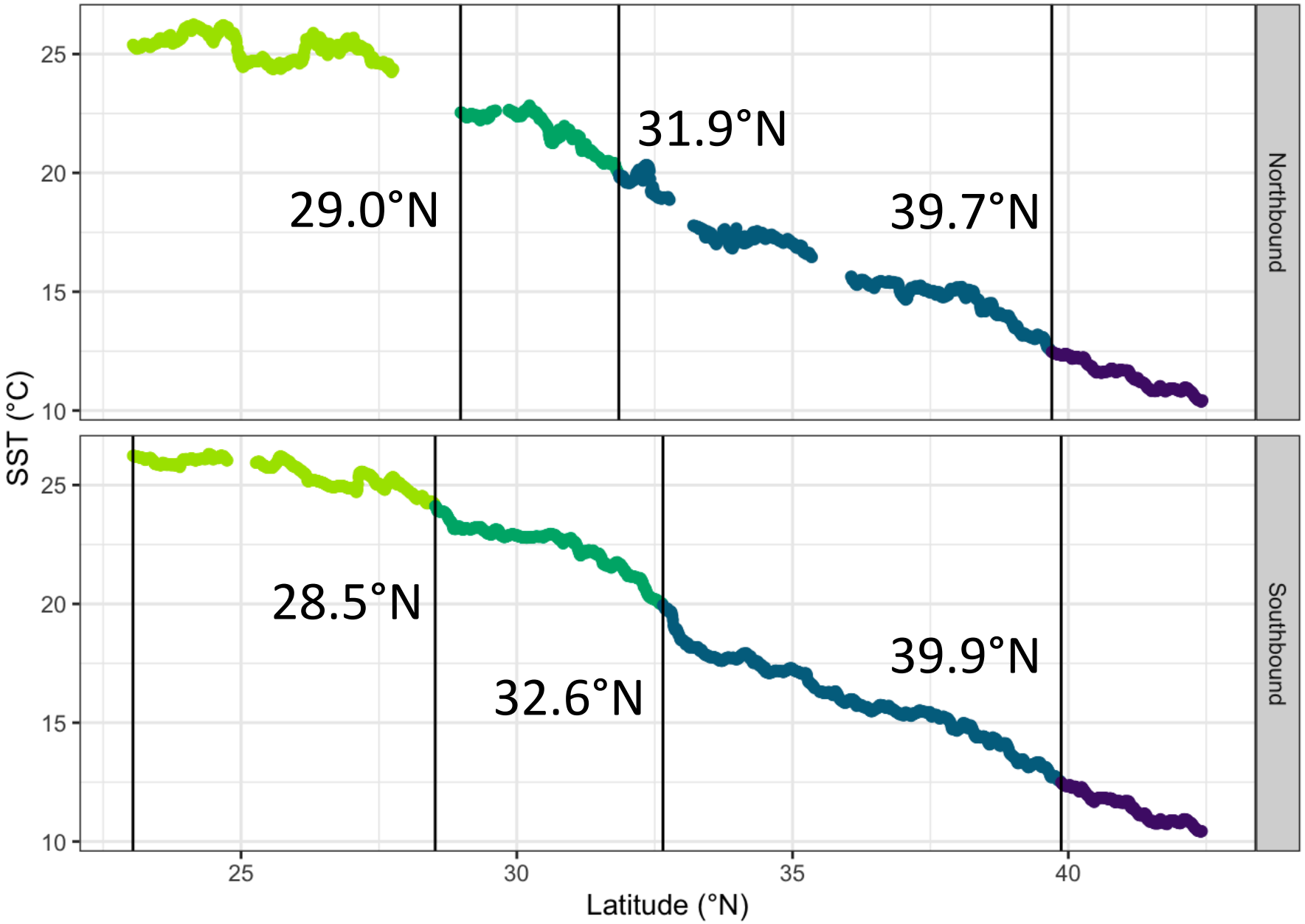
- Several methods capture transitions in populations in most cases
- Some manual intervention required
  - To set thresholds (on derivatives, number of breaks)
  - Removing duplicate / close transitions; transitions at boundaries
- Best models
  - Gaussian mixture model (least user influence)
  - Consensus of several models
- Gaussian mixture model captures the idea of different regimes with high or low cell abundance & variation around a mean
- **Agree on a common method and values to use across our projects?**

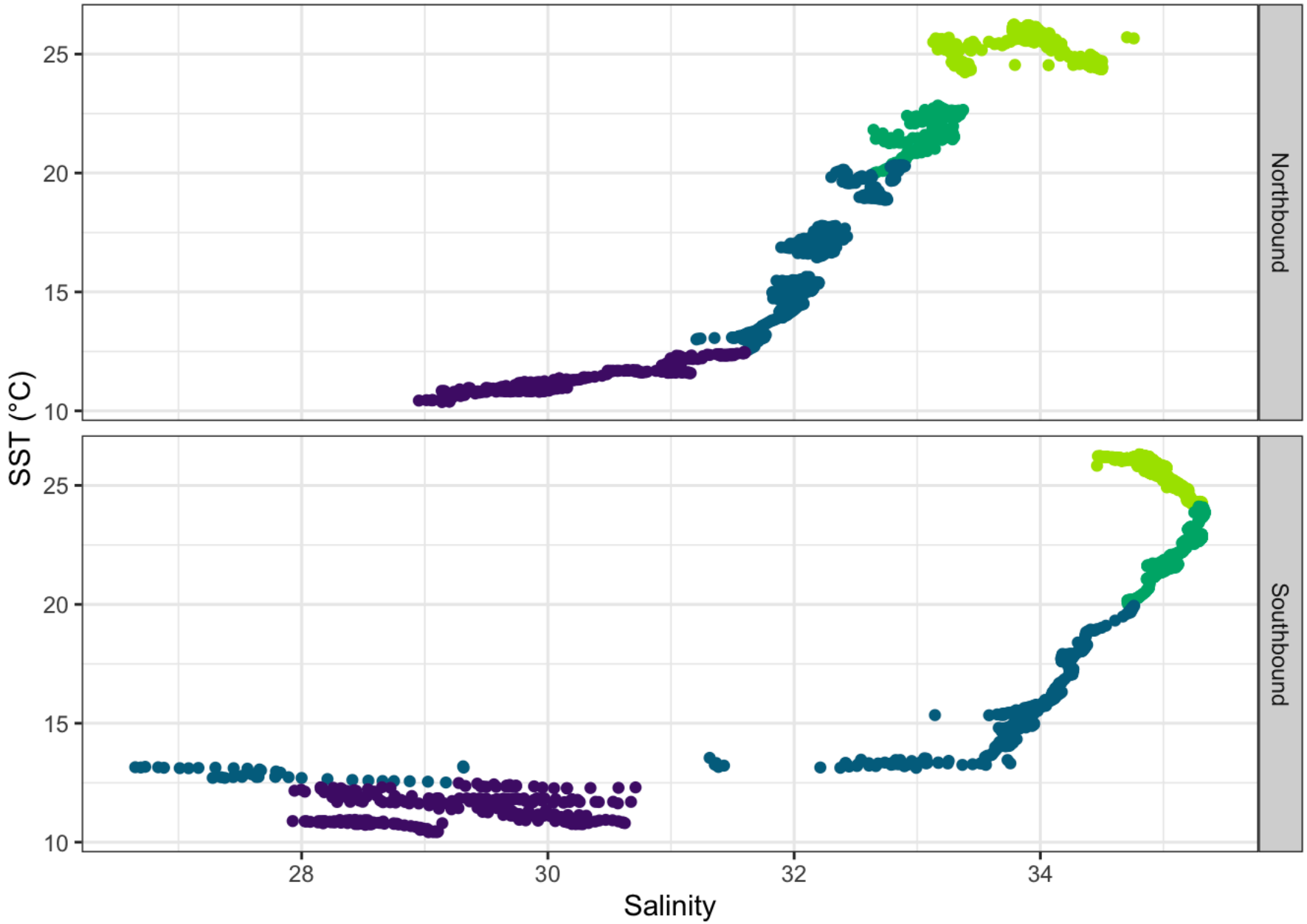
Finding temperature & salinity breakpoints



Finding temperature & salinity breakpoints




SST (°C)
25
20
15
10
29.0°N
31.9°N
39.7°N
Northbound
28.5°N
32.6°N
39.9°N
Southbound
25
30
35
40
Latitude (°N)


SST (°C)
25
20
15
10
Northbound
Southbound
28
30
32
34
Salinity

Breakpoint summary

	GMM	
	Northbound	Southbound
Prochloro-coccus	32.7	35.4
Synecho-coccus	33.4	32.9
Pico-eukaryotes	35.0	35.5
SST	29.0 31.9 39.7	28.5 32.6 39.9
Salinity	26.4 40.8	32.8 38.9 39.5