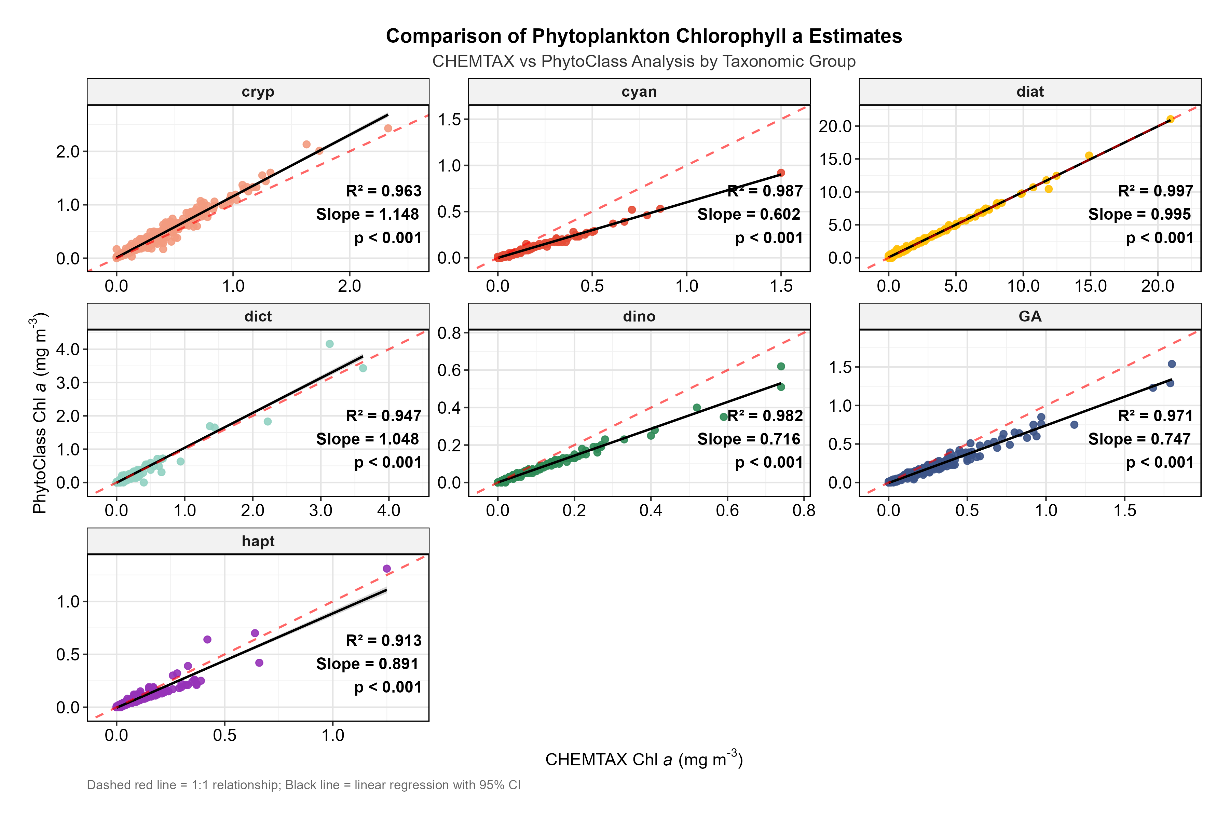
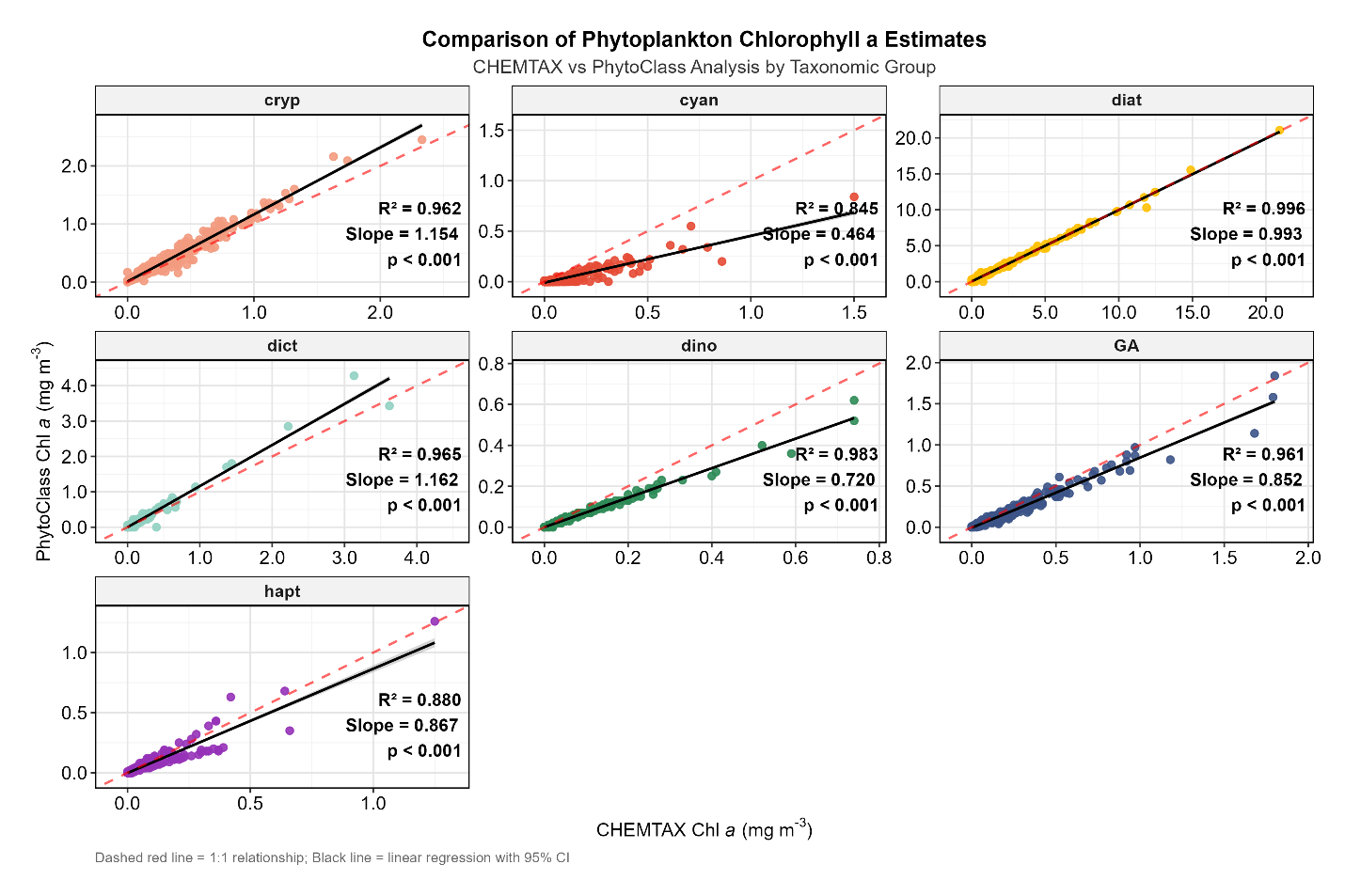
Phytoclass Trial log

* Min\_chem\_100\_noc1c2 (ran on July 22rd, 2025)
  + I meant to do 100% difference from my CHEMTAX manuscript values, but actually did CHEMTAX ratio/0.5 (min) and CR\*2 (max).
    - Think did to ensure min ratios did not = 0.
  + Deleted chlc1c2 as was causing confusion and large misclassifications.
  + Created my best comparison yet with high R2 for each group   
    (> 0.91) and slight phytoclass underestimates of:
    - Cyanobacteria
    - Dinoflagellates
    - Green Algae
  + Pro
    - Can nearly replicate CHEMTAX with easier method to run
  + Con
    - Defeats the purpose of Phytoclass - meant to not rely on specific input ratios with convergence on optimized ratios from wide min\_max range.
    - Optimized ratios meant to provide better results.
  + Trials
    - Iteratively introduce literature min and max for group to see how it changes.
  + Todo
    - Export converged ratios and analysis stats for comparison to CHEMTAX output ratios.
    - Try to set min\_max based on my field data rather than lit?
    - Should I try to get Phytoclass to exactly replicate CHEMTAX?
    - Run on clusters and see how changes.



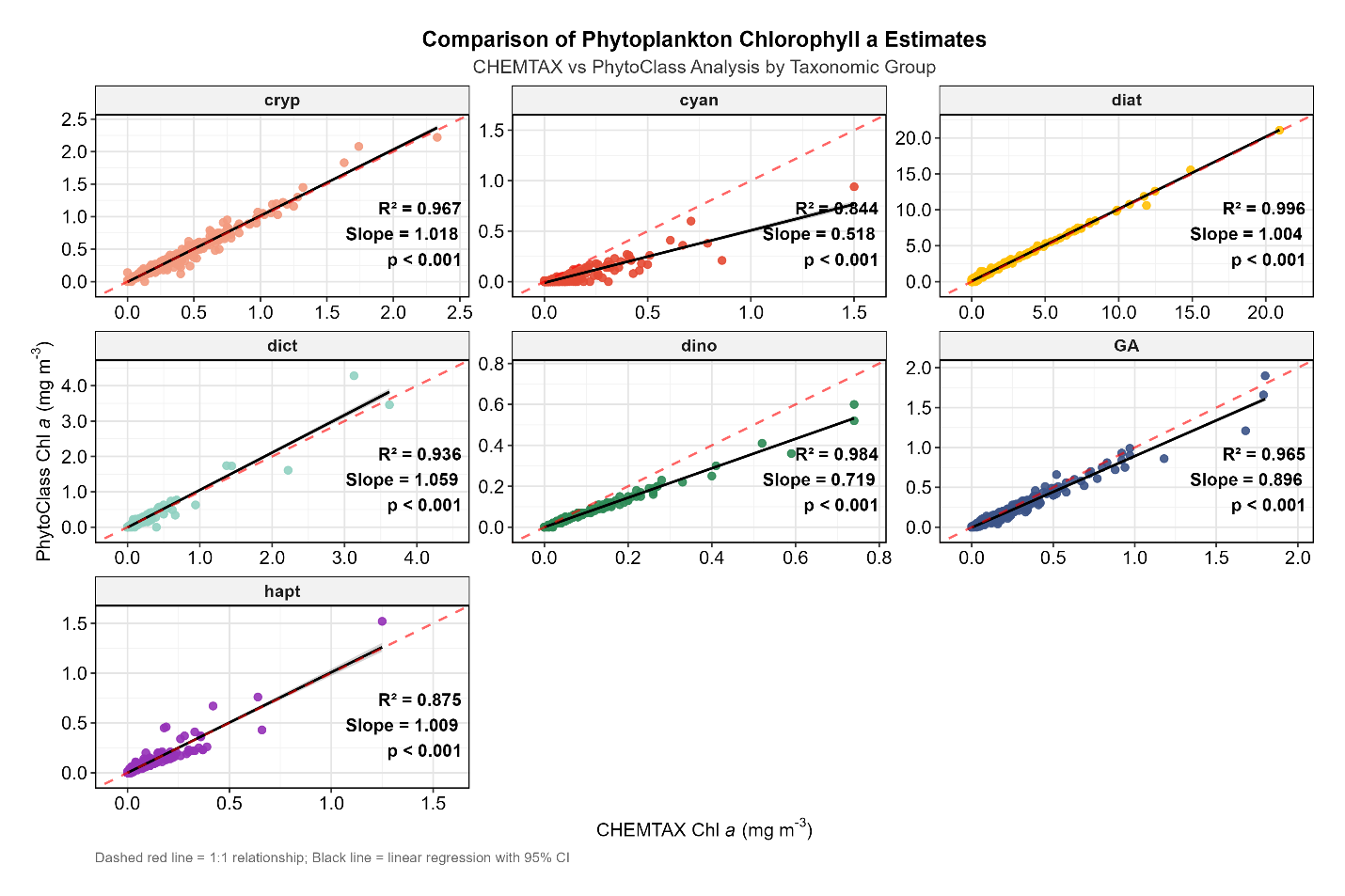
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| dataset | RMSE | condition\_number | type | sample\_size |
| autumn | 0.025509 | 1010.261 | Cluster | 102 |
| summer | 0.025884 | 666.2668 | Cluster | 120 |
| spring | 0.029665 | 727.2185 | Cluster | 113 |
| winter | 0.040668 | 1042.183 | Cluster | 76 |

* Min\_chem\_100\_noc1c2\_pras\_lit (ran on July 23rd, 2025)
  + Took the prasinophytes-3 literature min and max values.
  + Changed
    - Cyan – much poorer correlation and shallower slope
    - GA – similar correlation shallower slope
    - More divergence in haptos
      * Wonder what this is – two different types?
  + Are these differences just because CHEMTAX has wider ranges? How do I assess which is better?



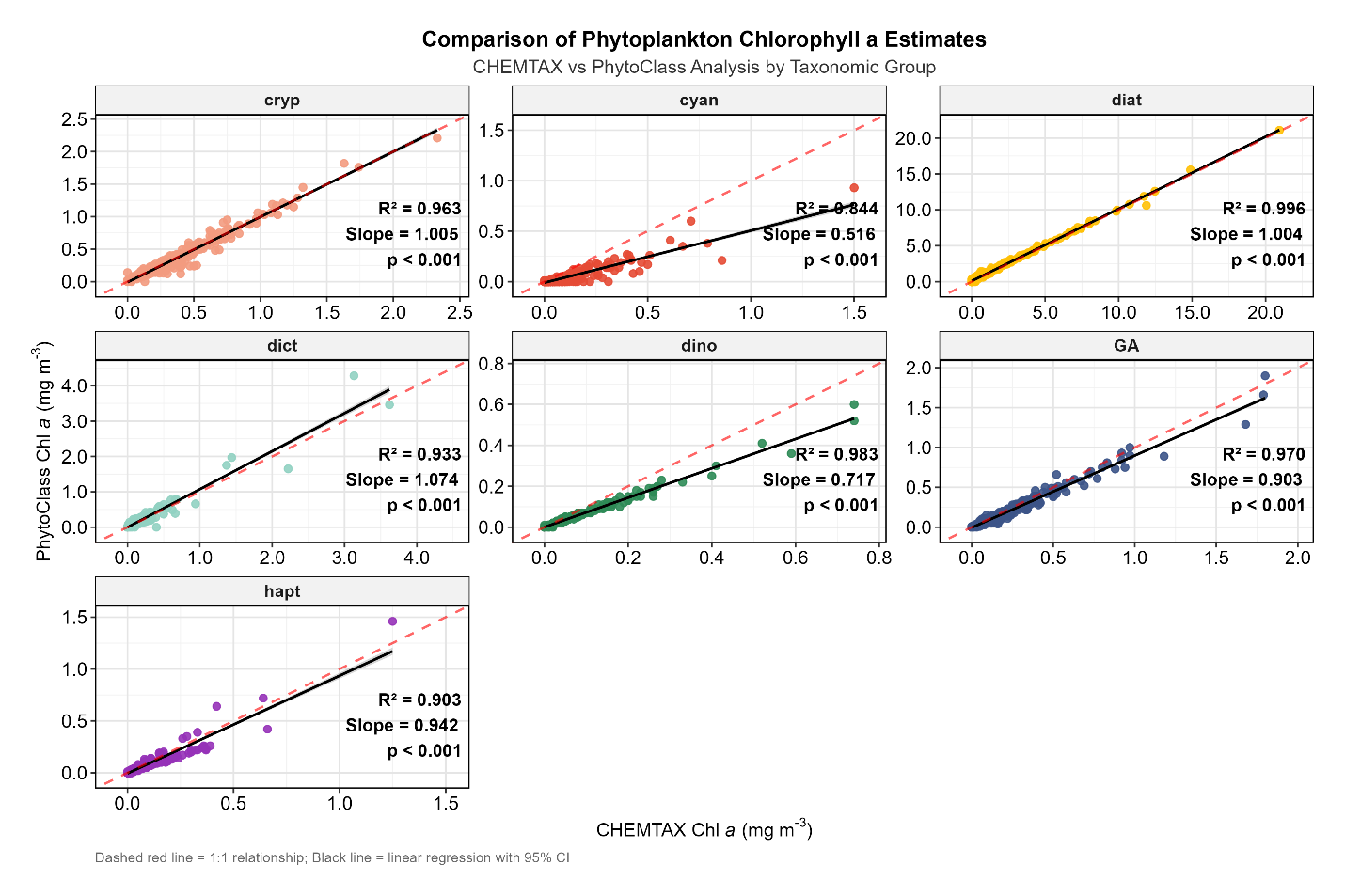
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| dataset | RMSE | condition\_number | type | sample\_size |
| autumn | 0.024956538 | 890.7726 | Cluster | 102 |
| summer | 0.025819509 | 700.1539 | Cluster | 120 |
| spring | 0.030029735 | 1773.824 | Cluster | 113 |
| winter | 0.040801521 | 1128.626 | Cluster | 76 |

* Min\_pras\_lit\_chem\_field (ran on July 24rd, 2025)
  + Using field data to constrain cryptophyte limits – leaving GA as lit ratios.
    - Our ratios for cryptophytes are considerably lower than those in the literature. For my CHEMTAX paper, I developed my own input ratios and have the ratio limits be 500%.
    - Here, I am going to try to take the min and max from the 10-year time-series.
      * Don’t like how I am using min\_max from entire time-series including periods when cryptophytes are not present – this could skew my min values low.
        + For Alloxanthin – filtered so that the concentration > 0 prior to calculating ratios. Modest change in min from 0.001 to 0.002.
    - Should also try using the literature values.
  + Field values created huge overestimations – nearly completely replaced diatoms!
    - Potentially a result of allowing such low ratios?
    - Allo converged at:
      * 0.002 for W,Sp,Su – confirms hypothesis
      * 0.16 for prior runs where mine 50% lower and max 100% higher than CHEMTAX (I think 0.12).
      * Try literature values – although lower range > our mean
* Min\_pras\_lit\_cryp\_lit (ran on July 24rd, 2025)
  + Literature min\_max for cryptophytes.
  + Do seem to work better!
  + Convergence at:
    - 0.17 (W)
    - 0.23 (Sp)
    - 0.18 (Su)
    - 0.19 (Au)
      * Some greater variability than the initial tests, which were static – this is good.
  + Outputs pretty comparable to CHEMTAX – best to use literature ratios.
  + Stats pretty comparable – slightly lower conditions numbers.



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| dataset | RMSE | condition\_number | type | sample\_size |
| autumn | 0.024448 | 881.507 | Cluster | 102 |
| summer | 0.02552 | 653.7504 | Cluster | 120 |
| spring | 0.028094 | 440.0372 | Cluster | 113 |
| winter | 0.040292 | 844.0409 | Cluster | 76 |

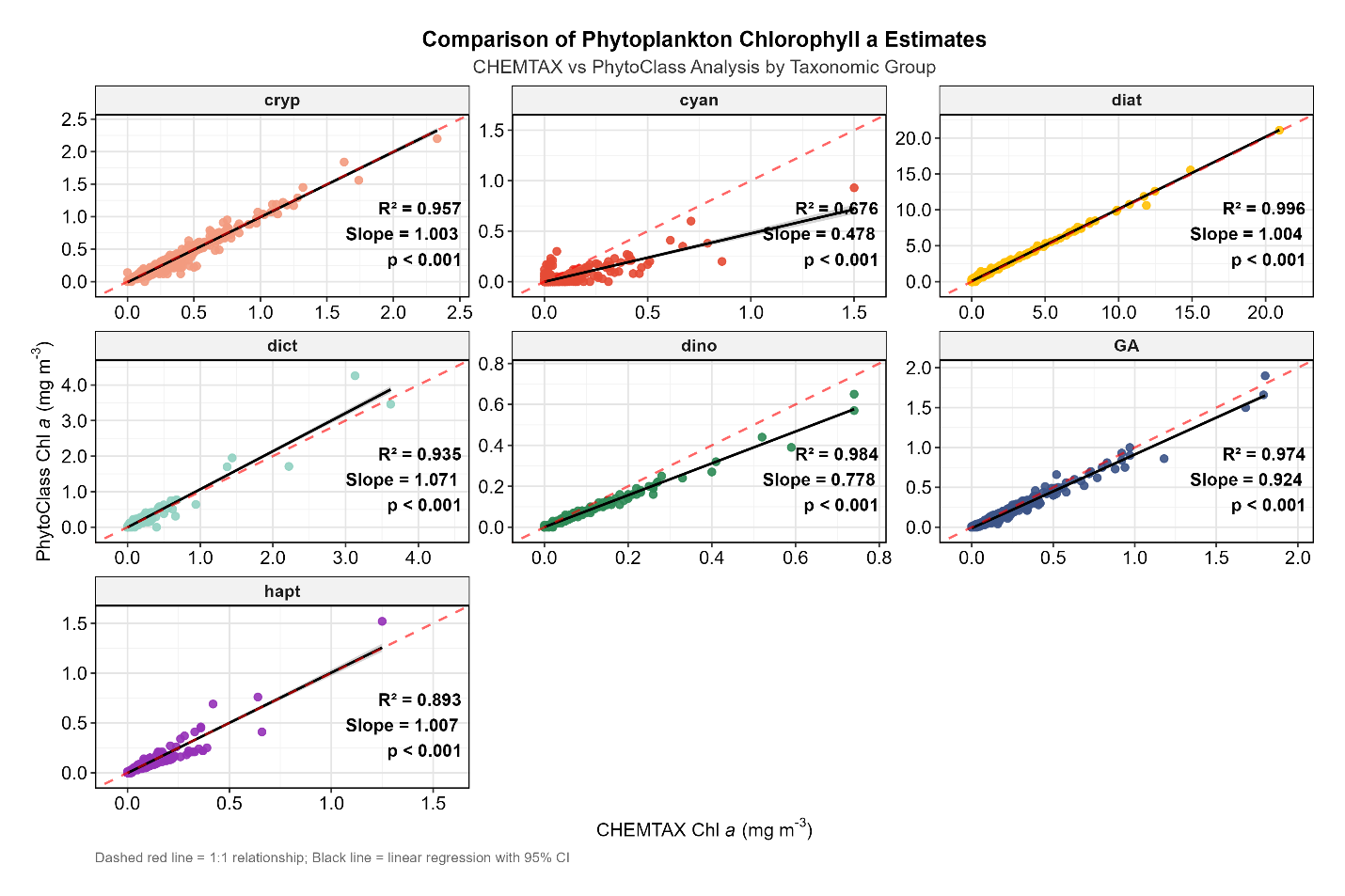
* Min\_pras\_lit\_cryp\_lit\_d\_lit (ran on July 24rd, 2025)
  + Trying to use diatom literature values
  + For now, should us the min and max for D1+D2 as both present.
  + Fuco convergence from 0.39-0.49 (low)
  + Not very different than prior runs, but should also compare to CHEMTAX.



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| dataset | RMSE | condition\_number | type | sample\_size |
| autumn | 0.024607 | 914.2549 | Cluster | 102 |
| summer | 0.025512 | 645.4107 | Cluster | 120 |
| spring | 0.028095 | 476.2789 | Cluster | 113 |
| winter | 0.039708 | 1012.665 | Cluster | 76 |

|  |  |  |
| --- | --- | --- |
| season | pigment | Zea |
| spring | Cyanobacteria | 1.279994 |
| summer | Cyanobacteria | 0.938914 |
| autumn | Cyanobacteria | 1.085118 |
| winter | Cyanobacteria | 0.32 |

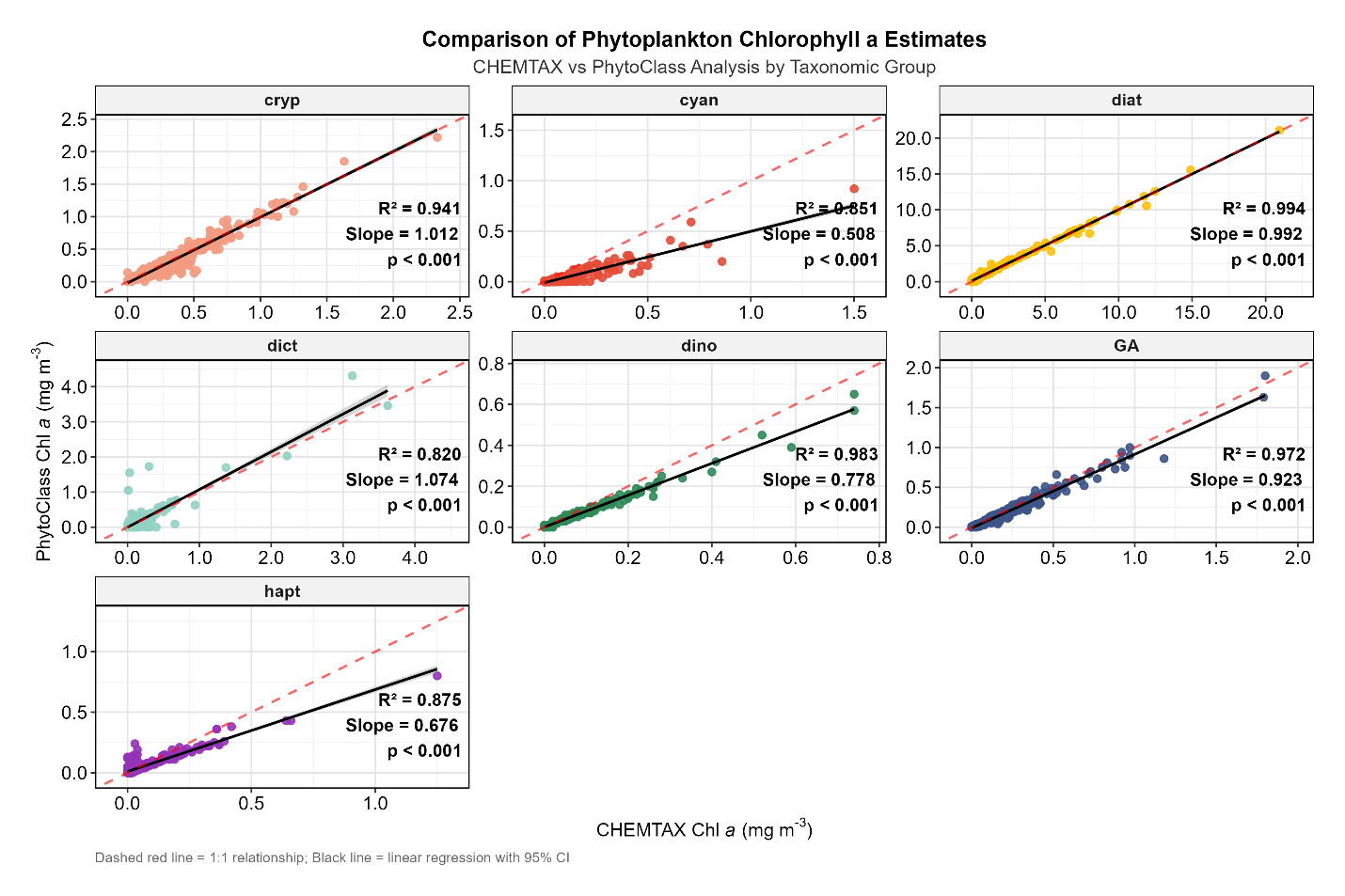
* Min\_lit\_to\_dino (ran on July 24rd, 2025)
* Subbing in literature peridinin
* Could be interesting as one of the groups showing biggest differences from chemtax
* Literature summary values are actually narrower than mine – suspect won’t make a difference
  + Could try cultured values that could be wider
* Difference with CHEMTAX could be that I am still using Chlc1c2, which would attribute more to Dinos during diatom bloom events – possibly.
* Output ratios the same across seasons (1.028) and very similar to prior runs (1.12)
  + Really just a minor group, so maybe why ratios so constrained?
* Try removal of c1c2 from CHEMTAX.
* Min\_lit\_to\_cyano
  + Skipping haptos and dictyos and moving to cyano to see if it changes the above underestimation.
  + Interesting to see how much it changed the cyano output -an even lower slope, more scatter and a separate low biomass grouping.
  + Everything else seems relatively comparable.
  + Slightly increased condition numbers.



|  |  |  |
| --- | --- | --- |
| season | pigment | Zea |
| spring | Cyanobacteria | 1.716736 |
| summer | Cyanobacteria | 0.952768 |
| autumn | Cyanobacteria | 1.460501 |
| winter | Cyanobacteria | 0.076 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| dataset | RMSE | condition\_number | type | sample\_size |
| autumn | 0.024709 | 885.5163 | Cluster | 102 |
| summer | 0.02564 | 656.658 | Cluster | 120 |
| spring | 0.028117 | 530.4062 | Cluster | 113 |
| winter | 0.038189 | 1300.181 | Cluster | 76 |

* Min\_lit\_to\_cyano\_hapt
* Accidentally overwrote previous outputs and statistics
  + Should fix so outputs = input min\_max names
* Using the literature hapto ratios changes dictyo and hapto quite a bit – as expected.
* Unexpected:
  + Loss of the tail in the hapto data.
  + More scatter in the cryp low magnitude data
* Also removed 2021-01-27 ad bad data point – either date got confused or not our data

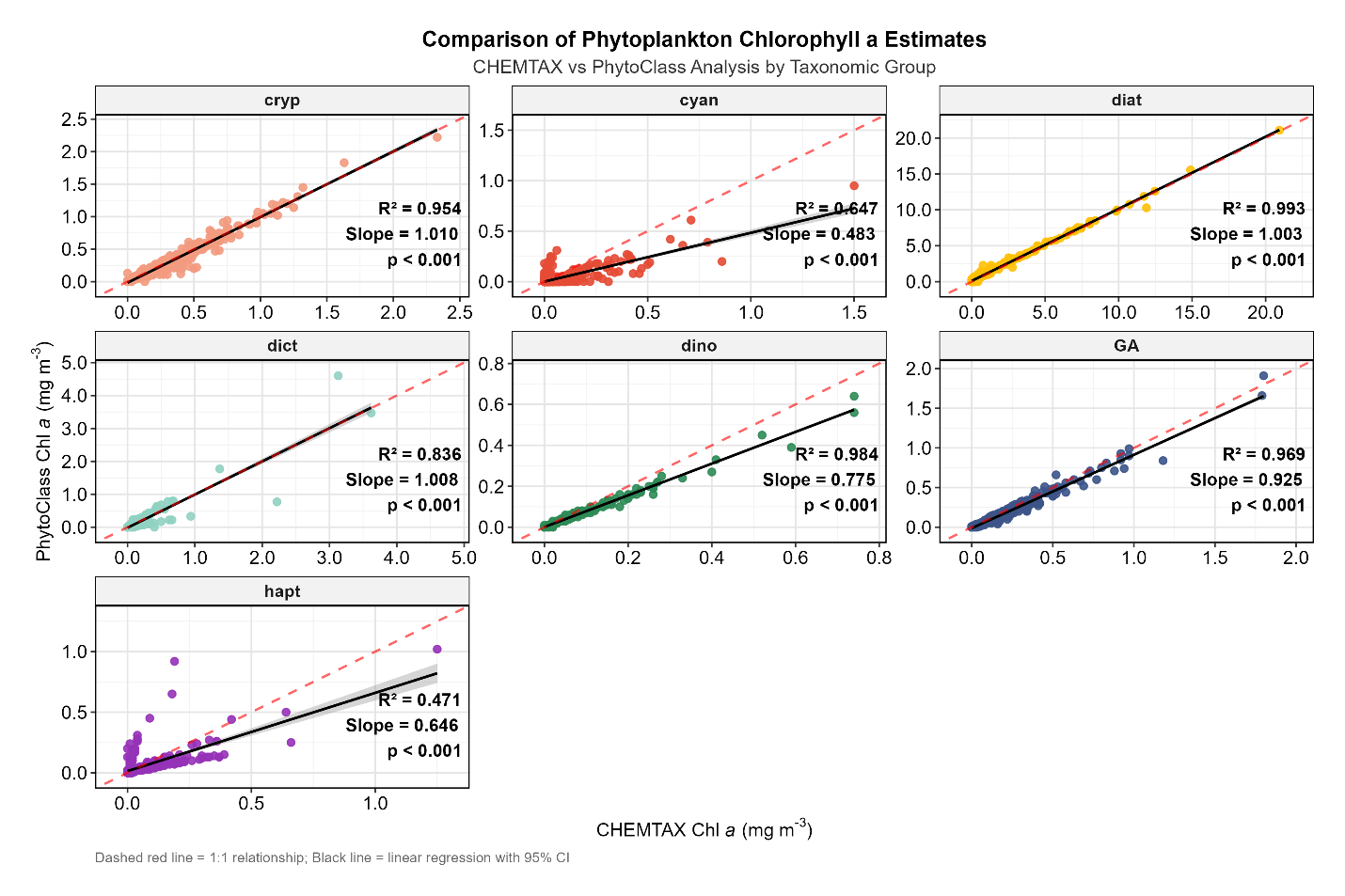


|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| dataset | RMSE | condition\_number | | | type | | sample\_size | |
| autumn | 0.024647 | 872.9896 | | | Cluster | | 102 | |
| summer | 0.025664 | 659.5772 | | | Cluster | | 120 | |
| spring | 0.027868 | 786.5599 | | | Cluster | | 119 | |
| winter | 0.034114 | 938.0621 | | | Cluster | | 81 | |
| season | pigment | | X19but | Fuco | | X19hex | | Tchla | |
| spring | Haptophytes | | 0.231663 | 0.697326 | | 0.901003 | | 1 | |
| summer | Haptophytes | | 0.063504 | 0.270851 | | 0.751452 | | 1 | |
| autumn | Haptophytes | | 0.059158 | 0.630137 | | 0.614562 | | 1 | |
| winter | Haptophytes | | 0.253623 | 0.736021 | | 0.09 | | 1 | |

|  |  |  |  |
| --- | --- | --- | --- |
| season | pigment | Zea | Tchla |
| spring | Cyanobacteria | 1.707293 | 1 |
| summer | Cyanobacteria | 0.96287 | 1 |
| autumn | Cyanobacteria | 1.158309 | 1 |
| winter | Cyanobacteria | 1.111131 | 1 |

1. **min\_lit\_to\_dino\_hapt\_dict\_trial\_higher\_min**

* Had to increase the dictyo min limit from 0.001 to 0.05 as they took everything.
* Other than that, everything are the summary min and max values.
* Seems like it sometimes does a better job of separating out Phaeocystis at times.
* Seems to be causing some confusion between dictyos and diatoms though – I think maybe the Fuco ratios are closer now?
  + Should play with this.
* Tail back on cyanos and generally downgrades concentrations
  + I think this tail represents spring bloom increases, which is NOT correct – what could this be?
* These will be very useful to have molecular data for.
* Make plot showing better correspondence with haptophytes from microscopy?



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| dataset | RMSE | condition\_number | type | sample\_size |
| autumn | 0.025518491 | 843.5694 | Cluster | 102 |
| summer | 0.025734565 | 709.8564 | Cluster | 120 |
| spring | 0.027753682 | 194.9607 | Cluster | 119 |
| winter | 0.037859504 | 2686.662 | Cluster | 81 |

OK, for now want to test introducing some new groups such as raphidophytes and D1 and D2.

Also, need to try with clustering rather than seasons to see what the outputs are.

**Min\_lit\_raph**

* Generally worked pretty well, but put relatively large concentrations during spring blooms (2021, 2018) – Why?
* In 2021 – very large bloom 2021-04-05 and viola was high (0.75), but ratio very low when considering all chla (21).
* Try upping viola ratio?
* Bumped up to 0.1 (for min). Up from 0.01.
  + Didn’t work
  + Bumping up again to 0.2
  + Didn’t change and all ratios are maxing out. Trying to see what happens if I double max ratio – changing min back to 0.1.
  + Increasing to max 0.6 maxed out again, but definitely decreases ratios. Trying to add Fuco
  + Seems to be a ratio maxing out problem… weird.

|  |  |  |  |
| --- | --- | --- | --- |
| season | pigment | Viol | Tchla |
| spring | Raphido | 0.319 | 1 |
| summer | Raphido | 0.319 | 1 |
| autumn | Raphido | 0.319 | 1 |
| winter | Raphido | 0.319 | 1 |

|  |  |  |  |
| --- | --- | --- | --- |
| season | pigment | Viol | Tchla |
| spring | Raphido | 0.6 | 1 |
| summer | Raphido | 0.6 | 1 |
| autumn | Raphido | 0.6 | 1 |
| winter | Raphido | 0.597922 | 1 |

**min\_lit\_raph\_fuco**

* Went back to literature ratios for raphidophytes and included fucoxanthin
* Did not seem to help at all – still showing presence during those blooms.
* I wonder if it is pulling from prasinophytes?
  + Doesn’t seem to be as there wasn’t any prasinoxanthin and chlb was low – although may have come from other greens?
  + I think maybe stealing from greens OR the diatoms in this year had viola…
  + When I look at the outputs, 2018 through 2021 just had higher raph and viola outputs. What was causing this? However, not an

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| season | pigment | Fuco | Viol | Tchla |
| spring | Raphido | 0.830305 | 0.319 | 1 |
| summer | Raphido | 0.667384 | 0.319 | 1 |
| autumn | Raphido | 0.727947 | 0.319 | 1 |
| winter | Raphido | 0.831 | 0.318181 | 1 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| season | pigment | Fuco | Viol | Tchla |
| spring | D1 | 0.430583 | 0 | 1 |
| spring | Raphido | 0.830305 | 0.319 | 1 |
| summer | D1 | 0.484415 | 0 | 1 |
| summer | Raphido | 0.667384 | 0.319 | 1 |
| autumn | D1 | 0.463224 | 0 | 1 |
| autumn | Raphido | 0.727947 | 0.319 | 1 |
| winter | D1 | 0.4126 | 0 | 1 |
| winter | Raphido | 0.831 | 0.318181 | 1 |

**min\_lit\_raph\_diat\_fuco\_viol**

* Going to try to give diatoms some viola.
* Causing huge raphido contributions – very odd.
* Maybe I just throttle the raphido ratios and make very limited range? Cut down the max by quite a bit?
* Going to just leave this for now.

Try best output compared to chemtax and run clustering approach.