**Question:** What predicts the occurrence of floating plant species?

**Response Variables** (n=3) (Each modelled separately):

* *Lemna minor* presence/absence
* *Spirodela polyrhiza* presence/absence
* *Wolffa* sp. presence/absence

**Predictor Variables** (n=13):

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Units** | **Code** | **Transformation** |
| Water body surface area | ha | size | log (x + 1) |
| Shoreline development index | - | shoreline | (xλ - 1) / λ |
| Total phosphorus, average | mg/L | totalP | (xλ - 1) / λ |
| pH, average | - | pH | (xλ - 1) / λ |
| Conductivity, average | µS/cm | cond | (xλ - 1) / λ |
| Secchi depth, average | m | secchi | (xλ - 1) / λ |
| Numb. water bodies within 1 km | - | lakes\_1km | log (x + 1) |
| Numb. water bodies within 10 km | - | lakes\_10km | (xλ - 1) / λ |
| Dist. to migratory water fowl habitat | km | waterfowl | log (x + 1) |
| Dist. to nearest lake w/ same species | km | dist\_occupied | (xλ - 1) / λ |
| Latitude | dec. degrees | latitude | (xλ - 1) / λ |
| Longitude | dec. degrees | longitude | (xλ - 1) / λ |
| Boat launch present? | - | boatlaunch | none |

**Excluded** (highly correlated)**:**

* Numb. water bodies within 5 km
* Dist. to nearest lake with any floating plant species
* Depth, maximum
* Alkalinity, average

**Model:**

* Logistic regression - generalized linear model with binomial error and logit link.

**Model formulations:**

* Full model (no interactions)
* Null model (intercept only)
* All combinations of 13 predictors (no interactions) [library MuMIn]
* Best model – model with lowest AICc

**Model diagnostics:**

* AICc
* Proportion deviance explained (Dev.null – Dev.residual) / Dev.null
* Likelihood-ratio based pseudo-R2 (compares fitted model to null model) [library MuMIn]
* 10-fold cross-validation prediction error [library boot]
* Partial residuals plots [library car]

**Model averaging:**

* All models with delta AICc < 2.0 [library MuMIn]
* Variable importance plots (Sum of wAICc over all models including the explanatory variable)

***Lemna minor***

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Formula** | **df** | **AICc** | **Proportion Deviance Explained** | **Pseudo-R2** | **CVerror** |
| Best (lowest AIC) | cond + waterfowl + latitude + dist\_occupied + secchi + totalP + lakes\_1km | 8 | 176.250 | 0.271 | 0.289 | 0.259 |
| Average (ΔAIC < 2) | cond + waterfowl + latitude + dist\_occupied + secchi + totalP + lakes\_1km + lakes\_10km + pH + longitude + shoreline | 12 | - | - | - | - |
| Full | All 13 predictors | 14 | 184.358 | 0.285 | 0.301 | 0.247 |
| Null | intercept only | 1 | 220.630 | 0 | 0 | - |

***Spirodela polyrhiza***

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Formula** | **df** | **AICc** | **Proportion Deviance Explained** | **Pseudo-R2** | **CVerror** |
| Best (lowest AIC) | cond | 2 | 158.884 | 0.051 | 0.046 | 0.178 |
| Average (ΔAIC < 2) | cond + secchi + size + lakes\_1km + waterfowl + pH + lakes\_10km + boatlaunch + longitude + shoreline | 12 | - | - | - | - |
| Full | All 13 predictors | 14 | 176.582 | 0.089 | 0.080 | 0.213 |
| Null | intercept only | 1 | 165.071 | 0 | 0 | - |

***Wolffia* sp.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Formula** | **df** | **AICc** | **Proportion Deviance Explained** | **Pseudo-R2** | **CVerror** |
| Best (lowest AIC) | boatlaunch + cond + secchi | 4 | 124.714 | 0.143 | 0.106 | 0.144 |
| Average (ΔAIC < 2) | boatlanch + cond + secchi + lakes\_10km + totalP + shoreline + lakes\_1km + waterfowl + latitude | 10 | - | - | - | - |
| Full | All 13 predictors | 14 | 137.212 | 0.196 | 0.142 | 0.184 |
| Null | intercept only | 1 | 137.900 | 0 | 0 | - |

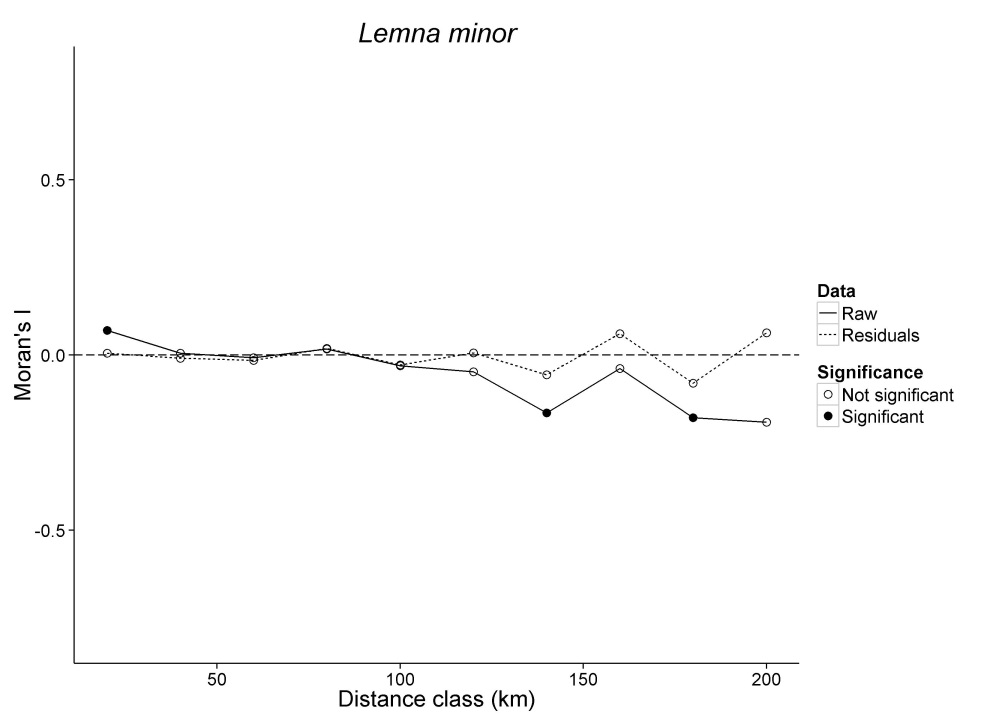
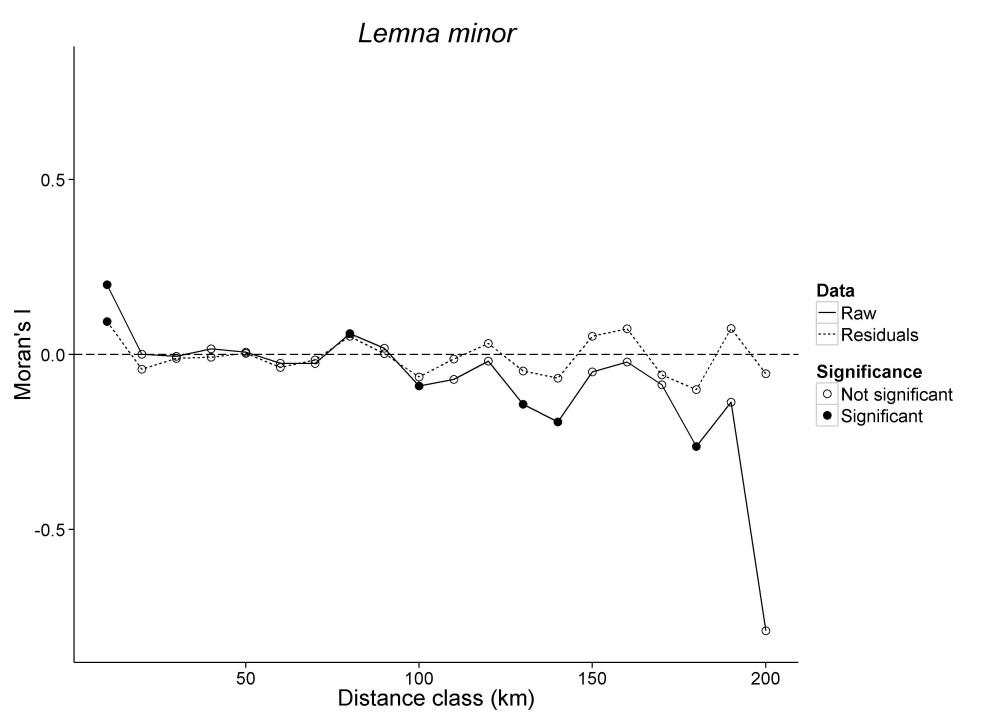
**Question:** Are model residuals spatially autocorrelated?

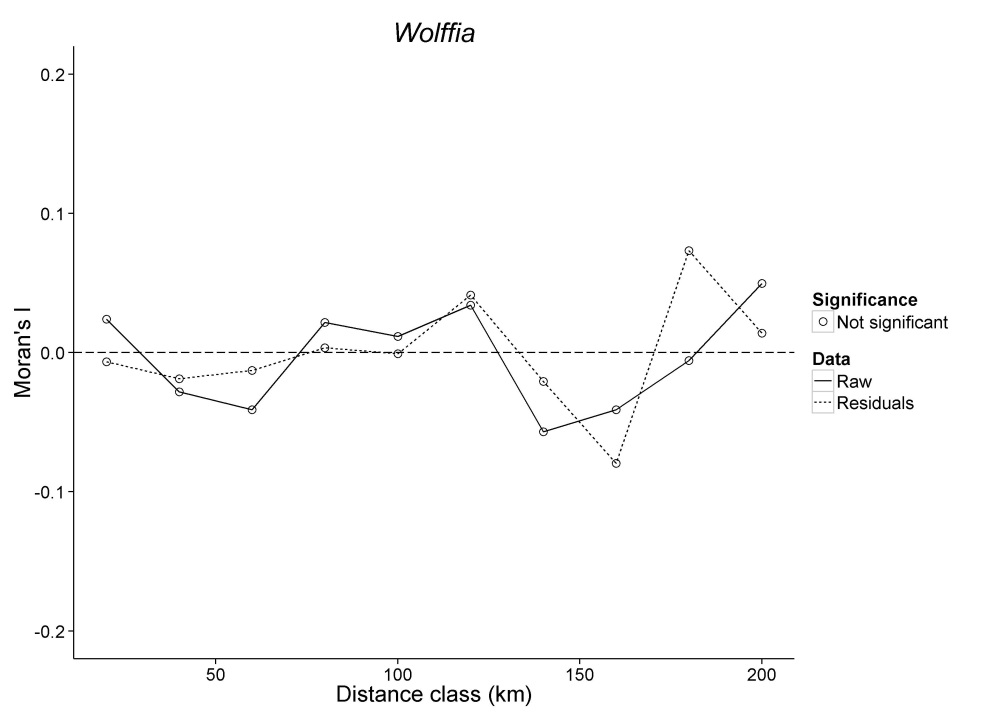
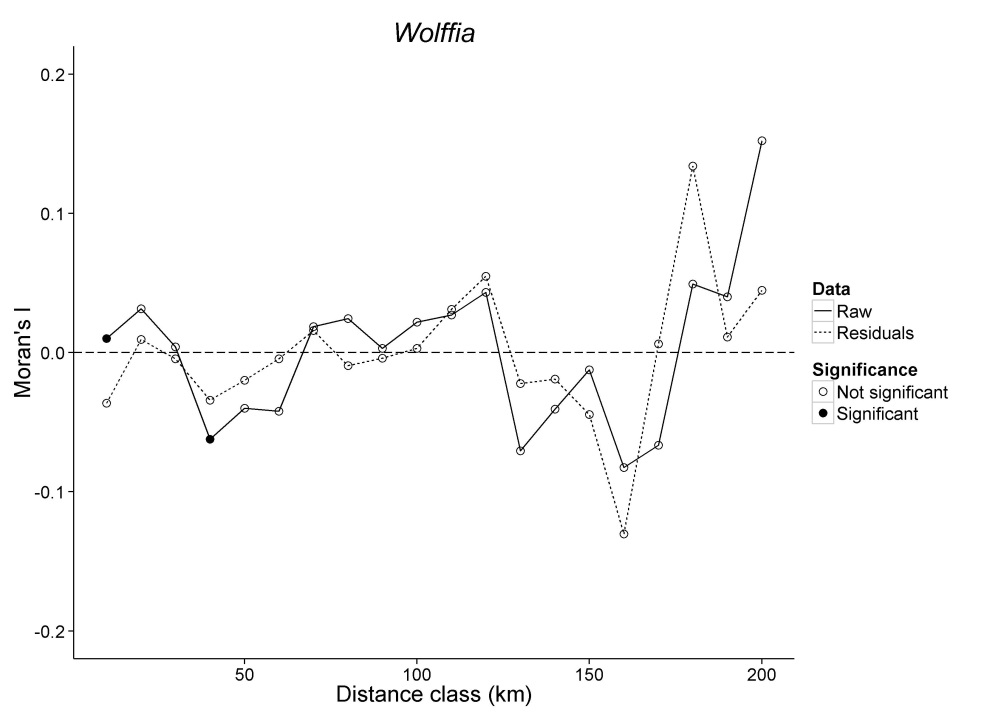
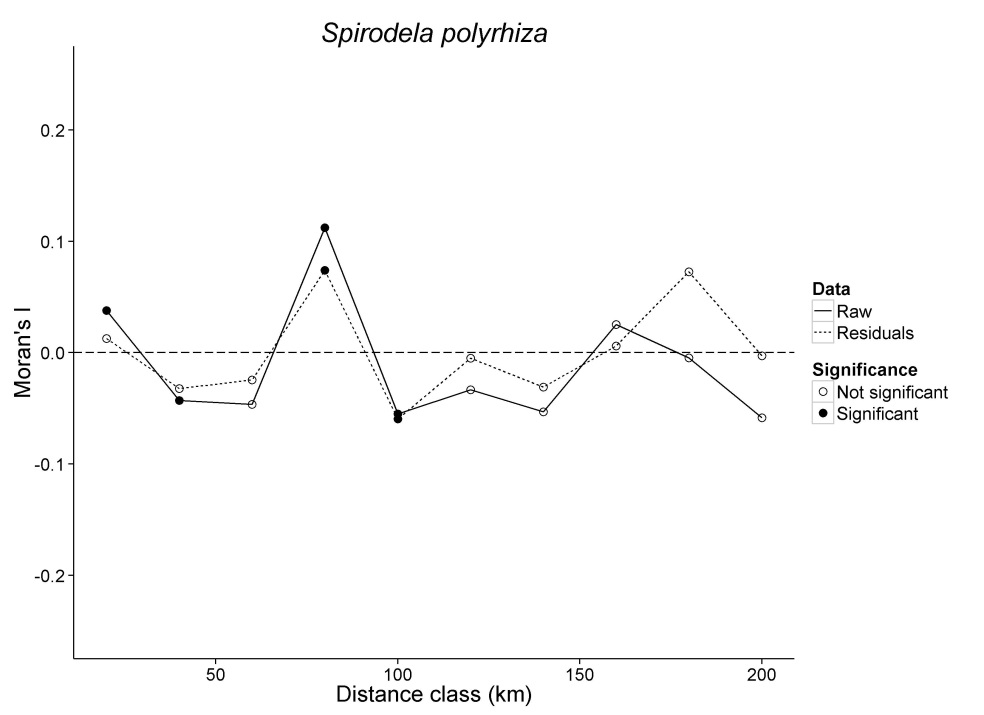
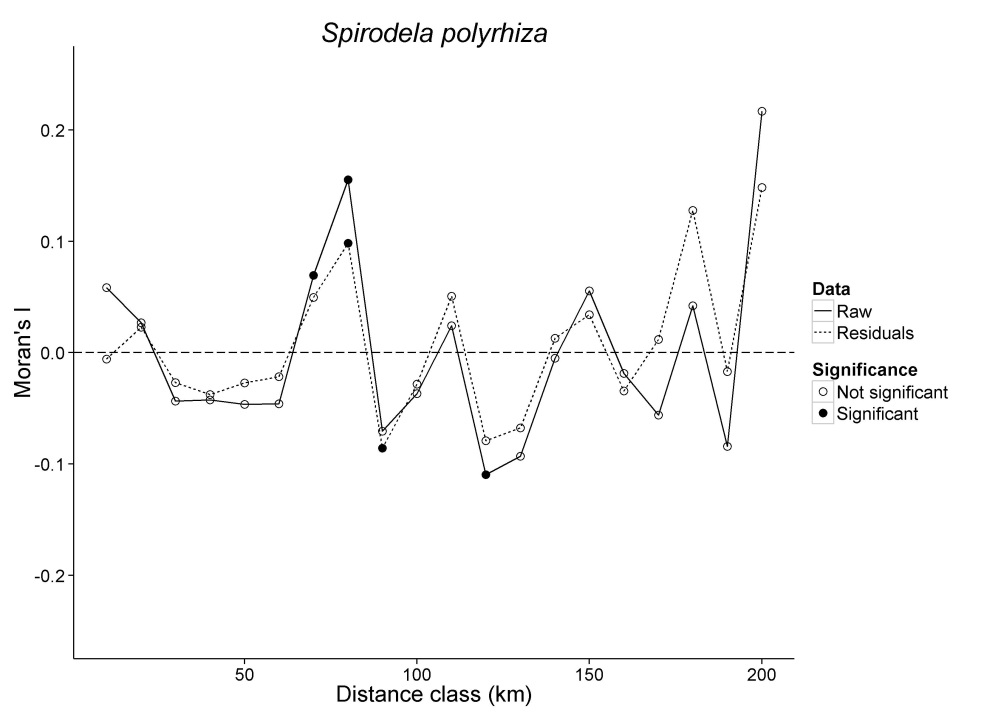
**Approach:** Spatial correlogram – plot Moran’s I for raw data and residuals at varying distance classes

**Justification:** If there is significant autocorrelation in raw data, then environmental factors may be spatially structured.

If the autocorrelation remains after fitting the model (i.e., for the residuals), then there may be an unmeasured factor that is spatially structured.

**Concern:** Increment used to specify distance classes *might* affect the results

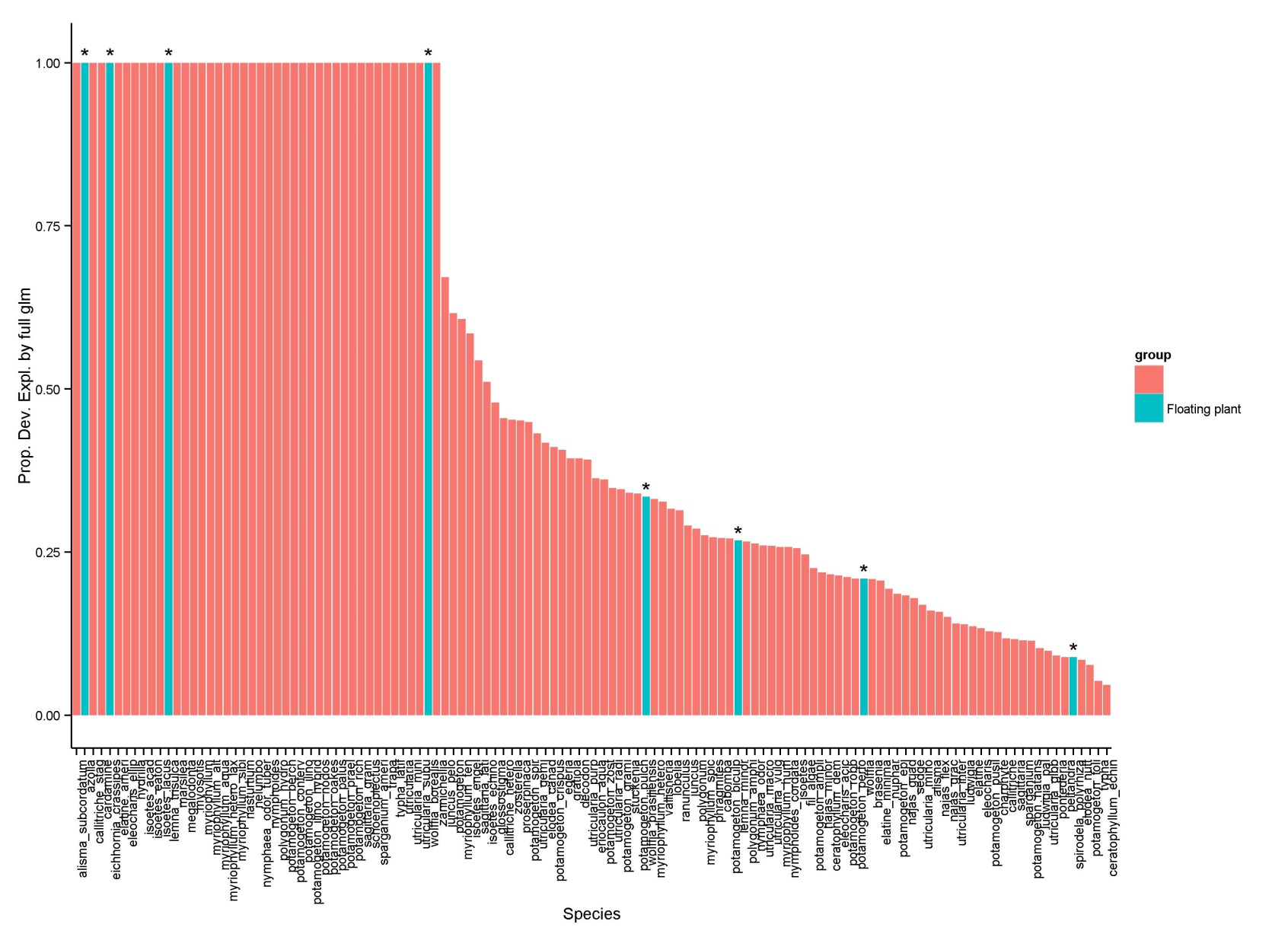






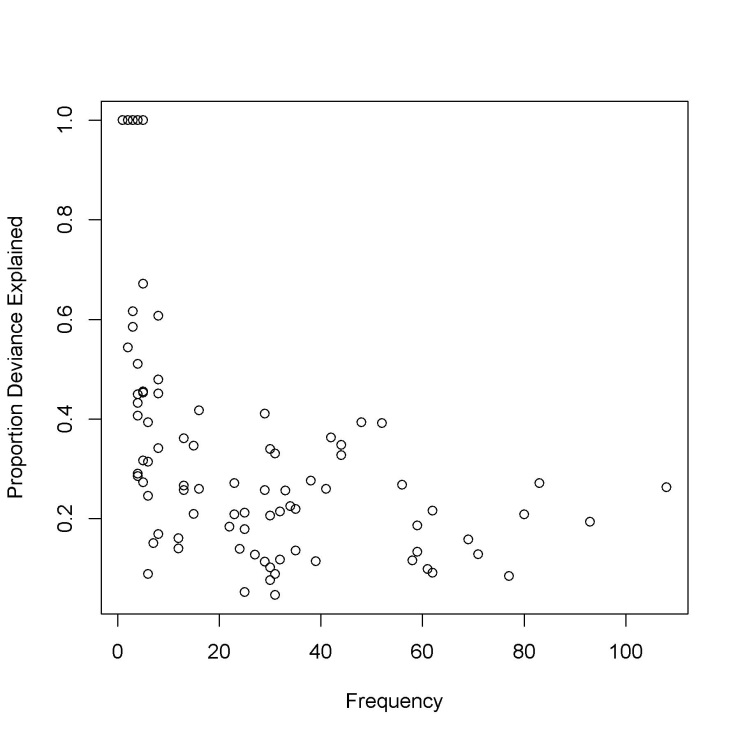
**Comparison to other species**

GLM (logistic regression) for all other species in CT – Full model (except *Dist. to nearest lake w/ same species*)

****

**Comparison to other species** (continued)

There appears be some dependence of proportion deviance explained on frequency of species occurrence (more common species are harder to predict?)



**Significant predictors** (As returned by summary(fitted.glm))

* 68 species without any significant predictors

|  |  |
| --- | --- |
| **Variable** | **# models w/ predictor as a signif. factor** |
| (Intercept) | 23 |
| longitude | 17 |
| COND\_avg | 16 |
| surfacearea\_ha | 11 |
| secchi\_avg | 10 |
| PH\_avg | 8 |
| boatlaunch | 7 |
| dist\_waterfowl | 6 |
| latitude | 5 |
| shoreline\_development | 5 |
| waterbodies\_1km | 5 |
| TOTP\_avg | 4 |
| waterbodies\_10km | 3 |
|  |  |
|  | **124 total models** |