

AB plant and invert responses to HF

Cari D. Ficken

2019-04-01

The goal of this project is to compare and contrast the responses of plants and invertebrates to environmental gradients. It uses the following data:

1. **Plant occurrence data**
 - These are presence-absence data from ABMI monitoring sites sampled with the Terrestrial and Wetland protocols
 - Sites are restricted to those classified as wetlands (bogs, fens, marshes, SOWWs, wet meadows)
2. **Insect abundance data**
 - These are frequency data from ABMI monitoring sites sampled with the *Wetland protocol only*.
 - Insects were not identified at ABMI sites sampled with the Terrestrial protocol
3. **Climatic data**
 - These include standard measure of temperature and precip (e.g. evaporation, frost-free period, MAP, MAT...)
 - These are available from ABMI website
 - We currently only have data for one year (**which?**) and for sites sampled with the wetland protocol
4. **Human footprint data (HF)**
 - 250 m buffer around each site

1. Load vegetation and HF datasets

Keep only data from sites with data in both df's.

The vegetation file (**veg_pa**) includes vegetation occurrence (i.e. presence-absence) for all wetland ABMI sites in Alberta. The vegetation taxa are ID'd to species - we have excluded taxa ID'd to genus and we have also removed any finer-level classifications (e.g. subsp.).

The human footprint df (**hf**) provides area estimates (km²) of different specific human footprint types. Each type (**FEATURE_TY**) has been assigned a higher-order category (**HFCateogry**) based on ABMI's HF metadata. Sum the total disturbance of each HF Category at each site in **hf**.

Let's explore the distribution of sites. We have 1653 wetlands in total.

```
keepsites %>% distinct(Protocol,WetlandType,Site, Year, NRNAME) %>% group_by(Protocol) %>% tally()

## # A tibble: 2 x 2
##   Protocol      n
##   <fct>      <int>
## 1 Terrestrial  437
## 2 Wetland    1216

keepsites %>% distinct(Protocol,WetlandType,Site, Year, NRNAME) %>% group_by(WetlandType) %>% tally()

## # A tibble: 8 x 2
##   WetlandType      n
##   <fct>      <int>
## 1 Alkali Fen        1
## 2 Bog             321
## 3 Marsh            457
## 4 Poor Fen        235
```

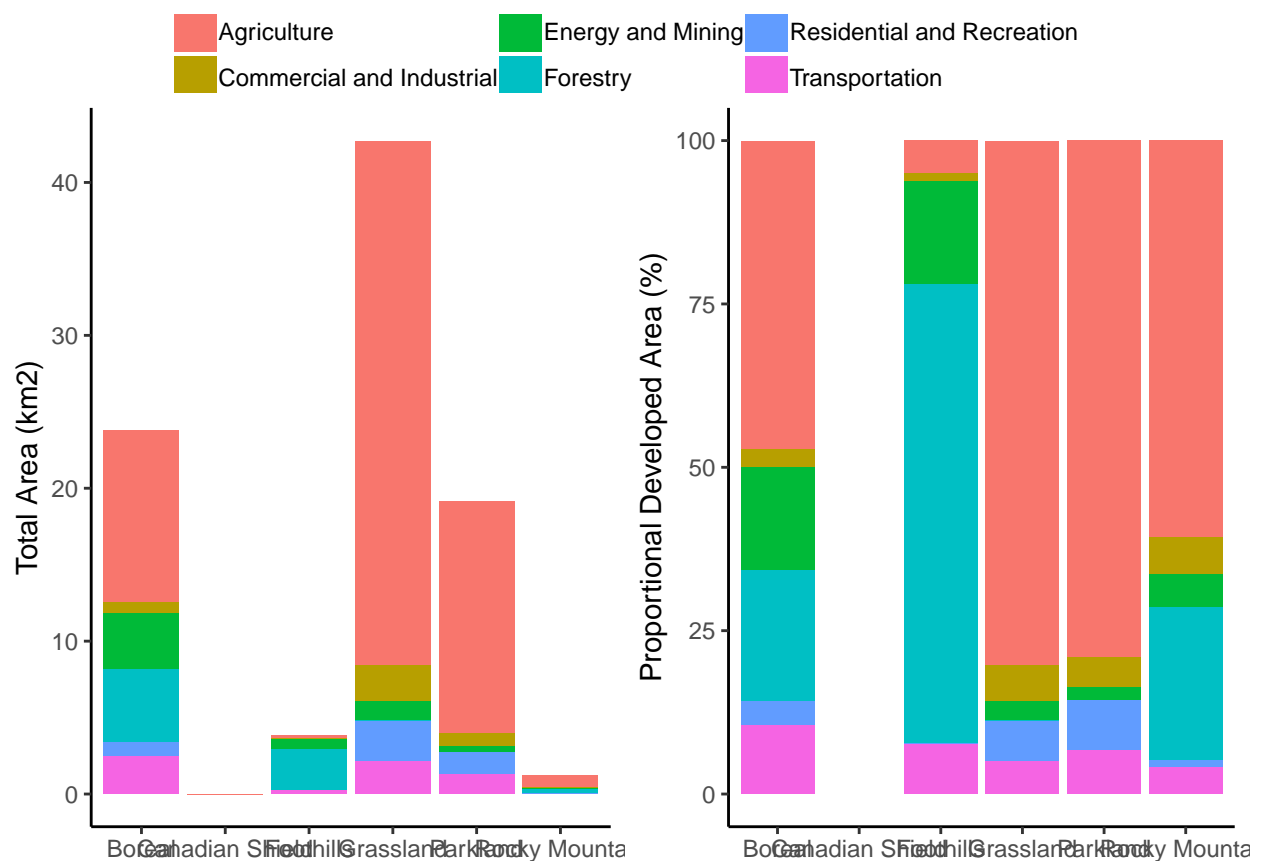
```
## 5 Rich Fen      85
## 6 Shallow Lake 181
## 7 Swamp        55
## 8 Wet Meadow   318
```

```
keepsites %>% distinct(Protocol,WetlandType,Site, Year, NRNAME) %>% group_by(NRNAME) %>% tally()
```

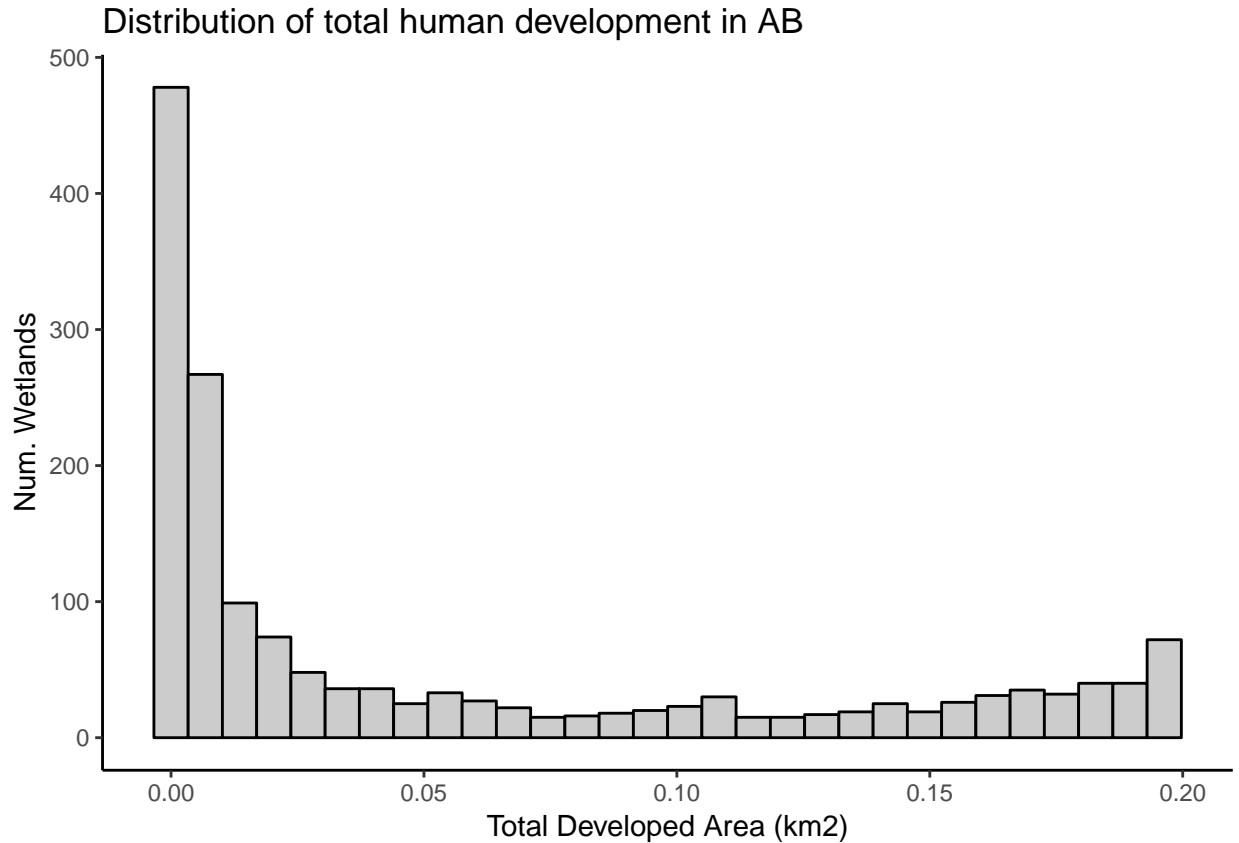
```
## # A tibble: 6 x 2
##   NRNAME      n
##   <fct>    <int>
## 1 Boreal    929
## 2 Canadian Shield 31
## 3 Foothills  97
## 4 Grassland 408
## 5 Parkland  160
## 6 Rocky Mountain 28
```

2. Examine the human footprint (HF) data

The plots below show the total land developed in each natural region, as divided by each HF Category. It also shows the proportion of total developed land in each NR that is attributed to each HF category. Note that wetlands in the Canadian Shield do not have any development in their immediate vicinity.



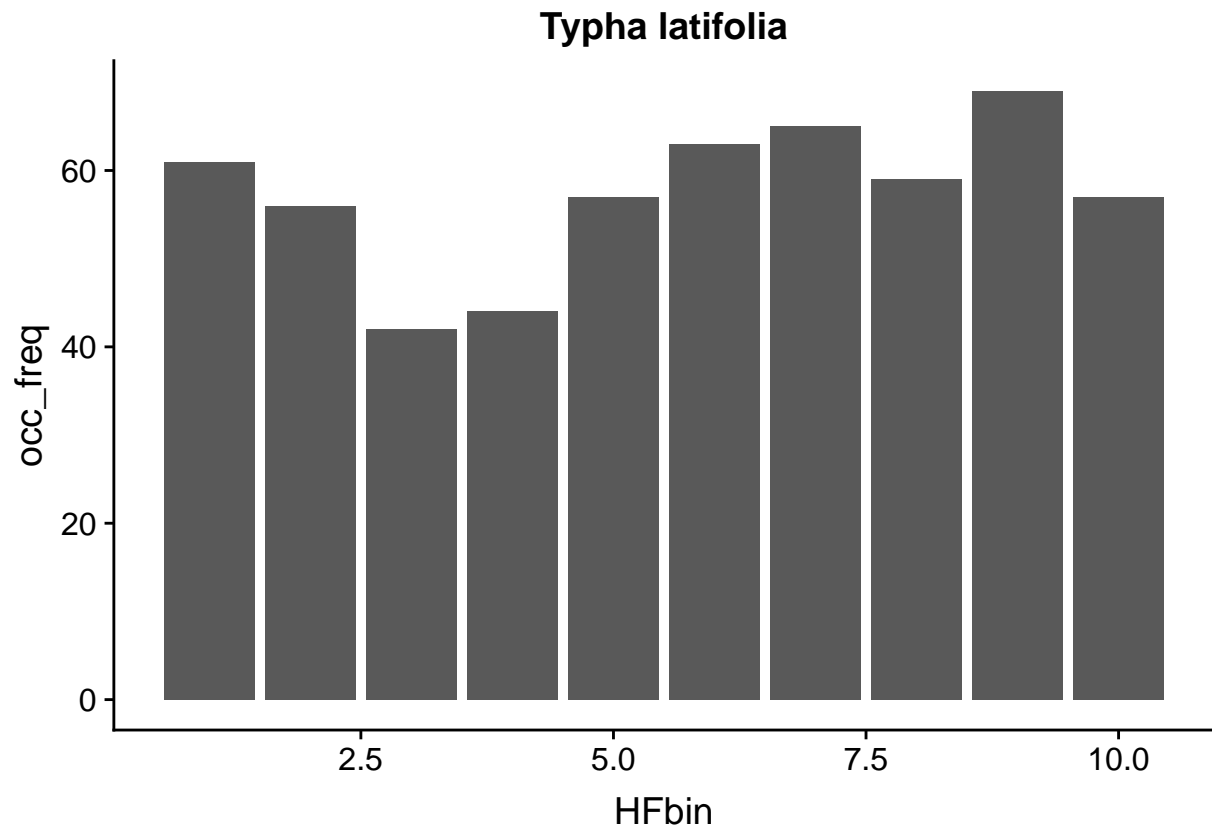
Create a new df with the total human development at each site (`hf_tot`). The histogram below shows the distribution of each HF Category across all focal sites.



3. Calculate species sensitivity index (SSI)

We can calculate the sensitivity of a species to HF by examining its total number of occurrences in each climate bin. To do so, we:

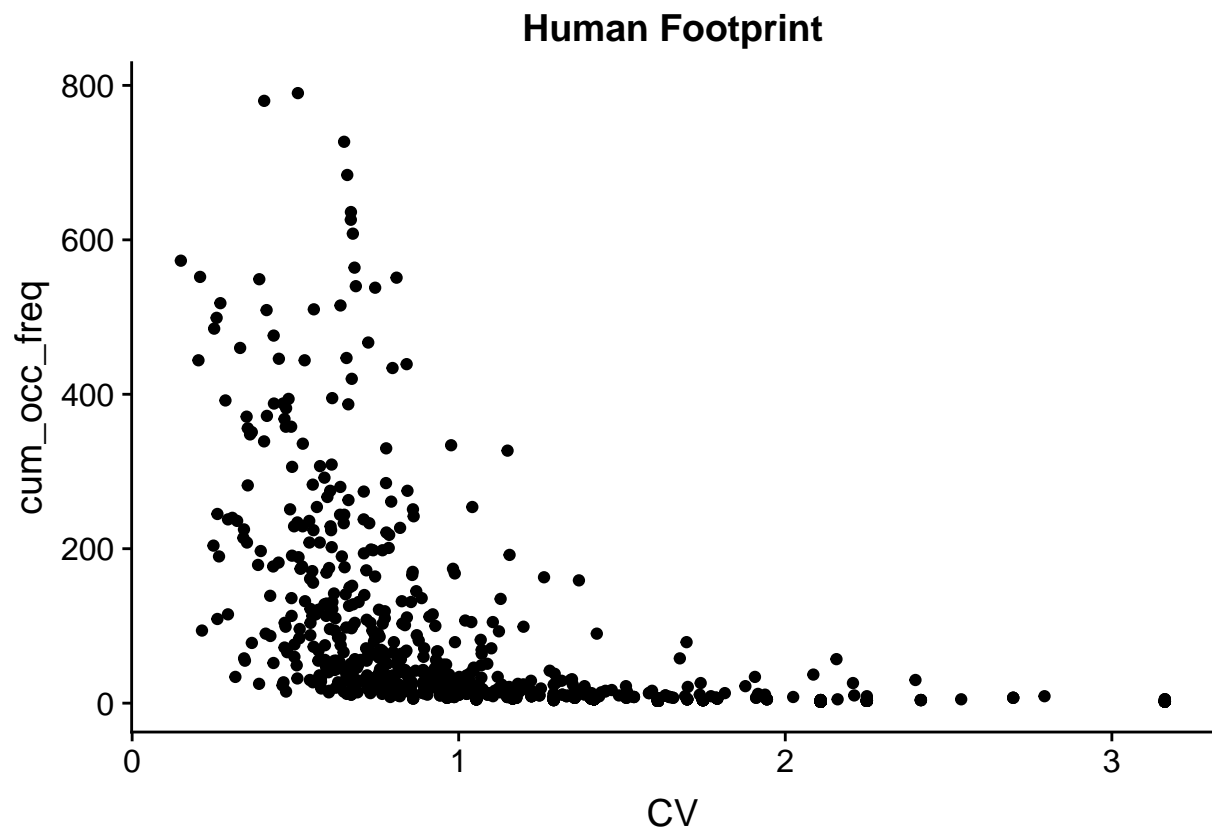
1. Bin the continuous HG gradient into 10 bins each with a similar number of sites (~140 sites) and assign each site to a bin
2. Join the vegetation df (`veg_pa`) to the HF df (`hf_tot`) to create `veg_hf`
3. Sum the number of occurrences of each species in each bin (`occ_freq`)
4. Calculate the coefficient of variation (CV) ($sd/mean$) for the `occ_freq` of each species across the full HF bin gradient. For example, below we can see the occurrence frequency distribution of *Typha latifolia* across the HF gradient.
 - Exclude species which occur only 1x in the full data set since these species will have very high sensitivity (N=34 sp which occur 1x)
 - Also be sure to expand the HF bin gradient for every species; that is we have to fill in `occ_freq` = 0 in some bins
4. The resulting df is called `sp_SSI`



3.1 Relationships between species occurrence frequency and CV:

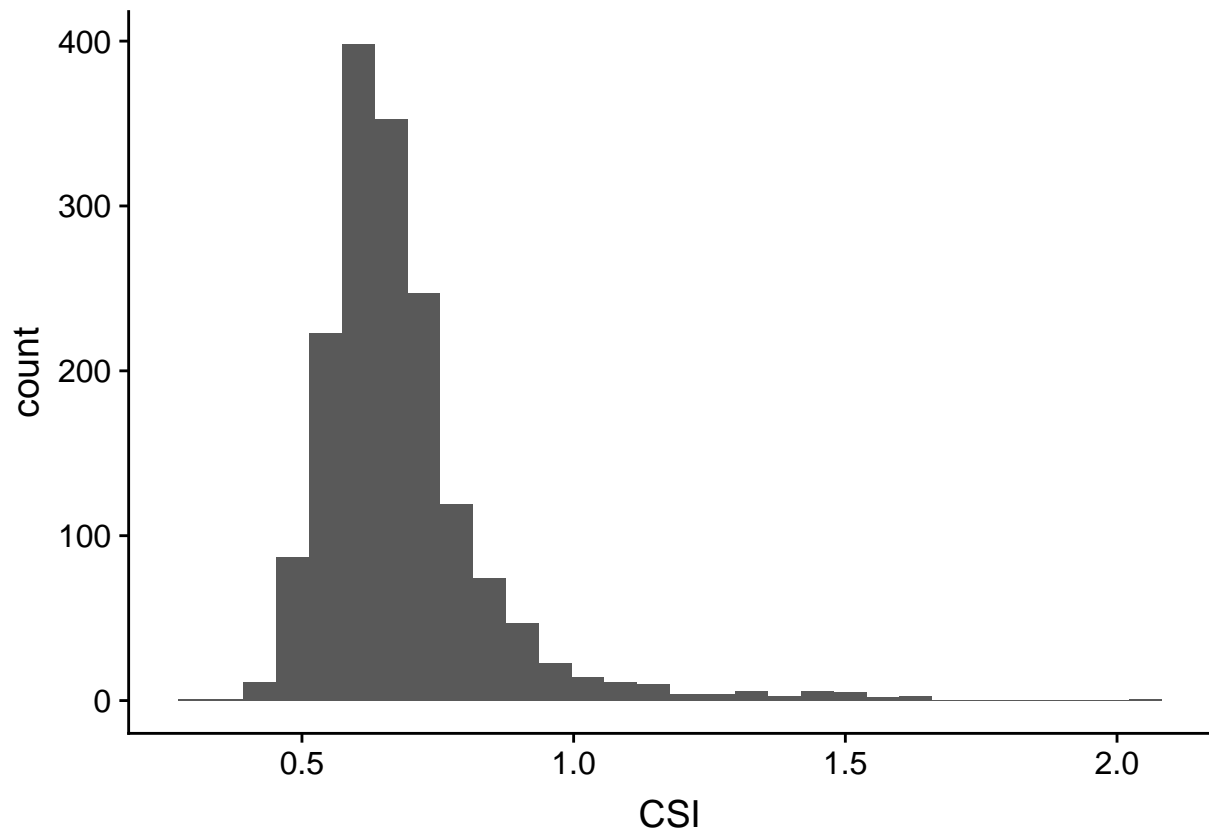
The resulting df is called `sp_SSI`. We can examine the relationship between SSI and occurrence frequency; we would expect common species to have low sensitivity to HF (low SSI).

```
## # A tibble: 6 x 2
##   Species      CV
##   <fct>      <dbl>
## 1 Abies balsamea 0.689
## 2 Abies bifolia 1.05
## 3 Acer negundo 1.79
## 4 Achillea alpina 0.756
## 5 Achillea millefolium 0.389
## 6 Acorus americanus 1.74
## Joining, by = "Species"
```



4. Calculate community sensitivity index (CSI)

Add SSI of each species (`sp_SSI`) to the vegetation df (`veg_pa`). Calculate the mean SSI at each site (`CSI`).



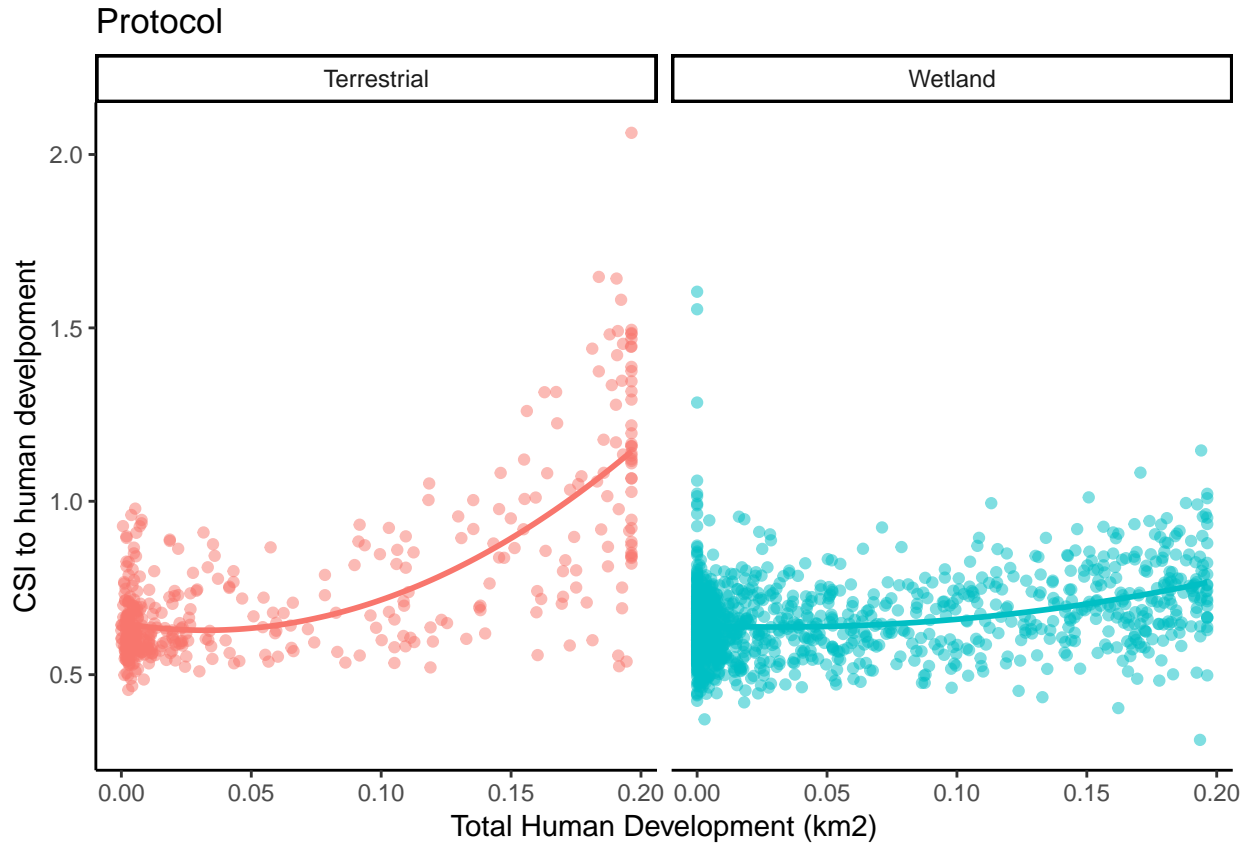
5. Examine patterns of CSI across different wetland/site classifications gradients

5.1 Comparisons between Terrestrial and Wetland Protocols

Compare the distribution of CSI values for sites sampled with wetland vs terrestrial ABMI protocols.

The Wetland protocol underestimates the CSI of communities at high development levels. Generally, though, communities show the highest sensitivity to development at really high development intensities. This may be because communities at high development levels are composed of ruderal species that are only found under these disturbed conditions; at low development intensities, perhaps communities are composed of more broadly distributed native species. That is, after ~0.05 km² of development, ruderals take over.

For sites sampled with the wetland protocol, there appears to be a more directional (positive) trend.

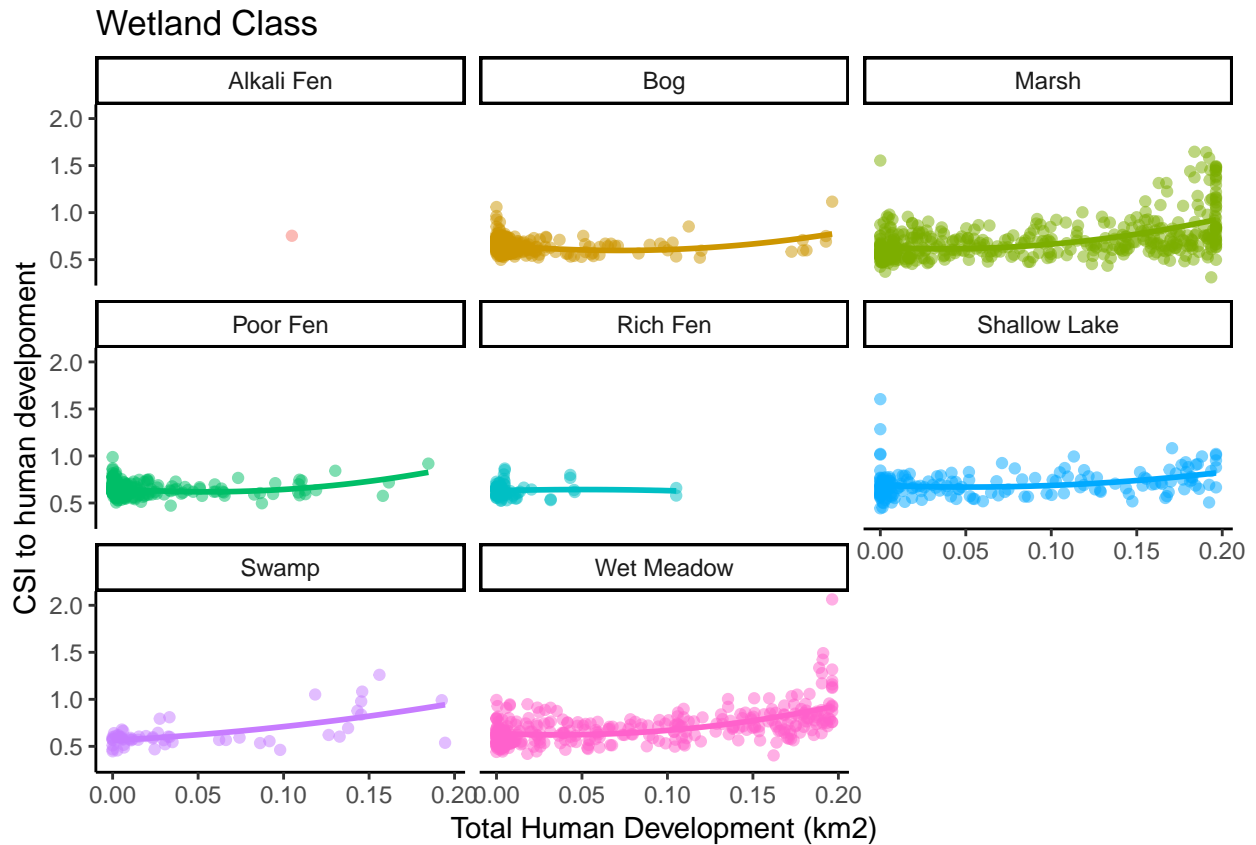


5.2 Comparisons among wetland classes

Patterns across wetland classes are difficult to disentangle due to vastly different sample sizes of each wetland class. Marshes, shallow lakes, and wet meadows have the most thorough representation across the full HF gradient. For these classes, there is slightly high sensitivity to development at both ends of the development spectrum. This largely parallels the pattern observed across the full dataset. For other wetland classes, sites are not distributed across the gradient sufficiently to really compare their sensitivity.

No single wetland class appears to have particularly high or low sensitivity overall.

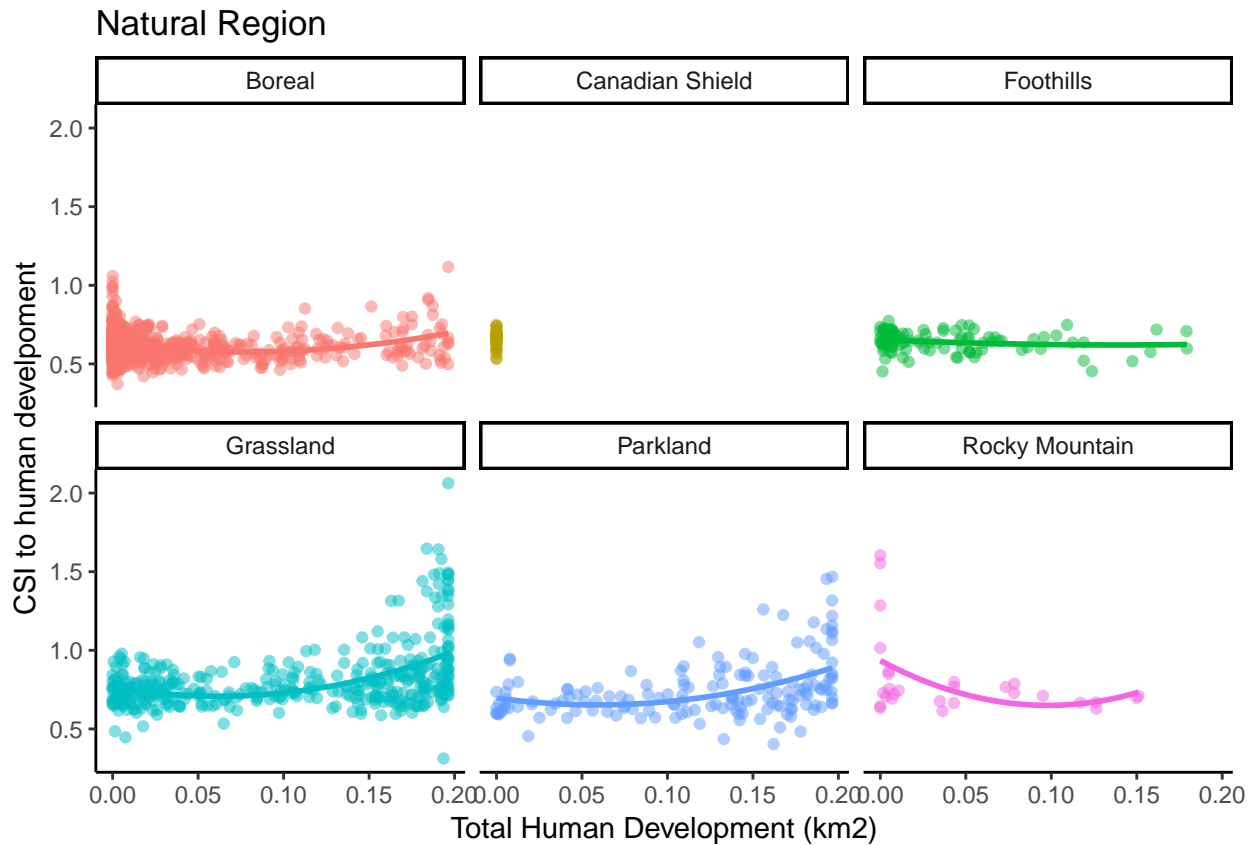
```
## # A tibble: 6 x 8
## # Groups:   Protocol, NRNAME, WetlandType, Site [5]
##   Protocol  NRNAME WetlandType Site  Year  CSI totdist_km2 HFbin
##   <fct>      <fct>  <fct>      <chr> <int> <dbl>      <dbl> <int>
## 1 Terrestrial Boreal Bog      1018  2011 0.853      0.112      8
## 2 Terrestrial Boreal Bog      1079  2013 0.597      0.00304    3
## 3 Terrestrial Boreal Bog      1178  2013 0.691      0.193     10
## 4 Terrestrial Boreal Bog      1178  2014 0.751      0.193     10
## 5 Terrestrial Boreal Bog       128  2014 0.634      0.0168    6
## 6 Terrestrial Boreal Bog       157  2014 0.523      0.0245    6
```



5.3 Comparisons among natural regions

Wetlands across the natural regions vary somewhat in their exposure to human development. For example, there is essentially no development in the Canadian shield and very little in the rocky mountain NR Sites in the boreal, foothills, grassland, and parkland NRs show a greater range in human development.

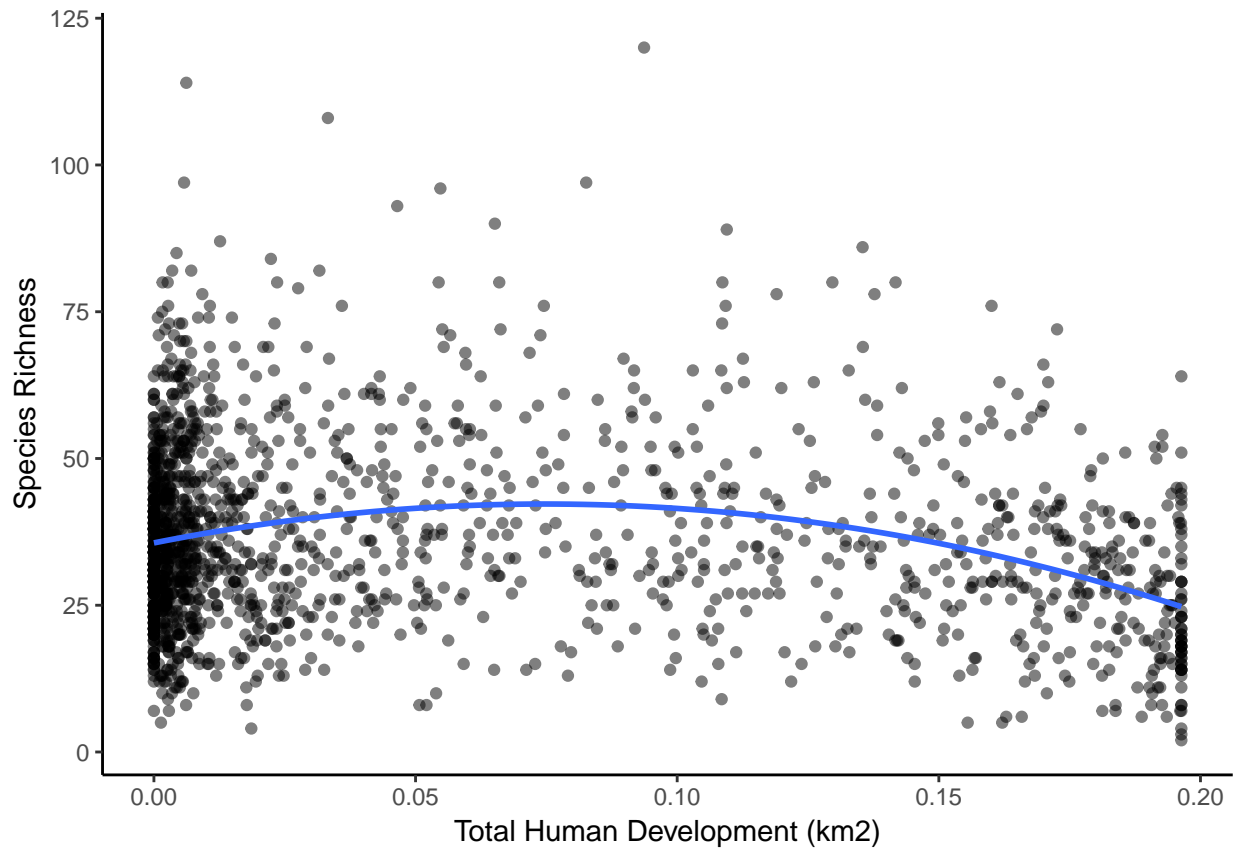
Rocky mountain wetlands have high sensitivity at low development intensities. Grassland wetlands maybe have higher overall sensitivity to human development than wetlands in the other NRs.



6. Patterns of species richness across HF gradient

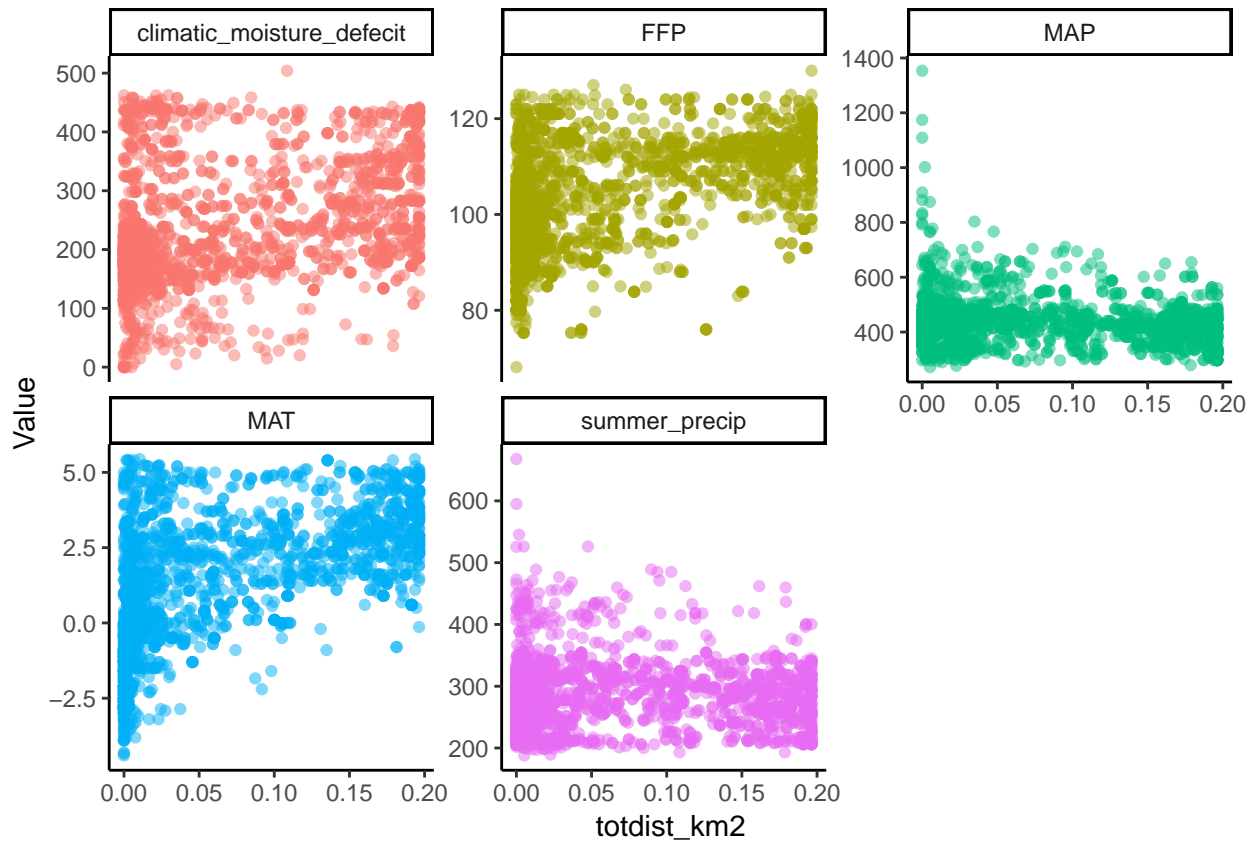
Channeling the Intermediate Disturbance Hypothesis here...

Species richness is somewhat elevated at intermediate human development intensities. This supports the findings of Mayor et al 2012 (*Regional boreal biodiversity peaks at intermediate human disturbance*) which uses ABMI data to examine this pattern only in the boreal.

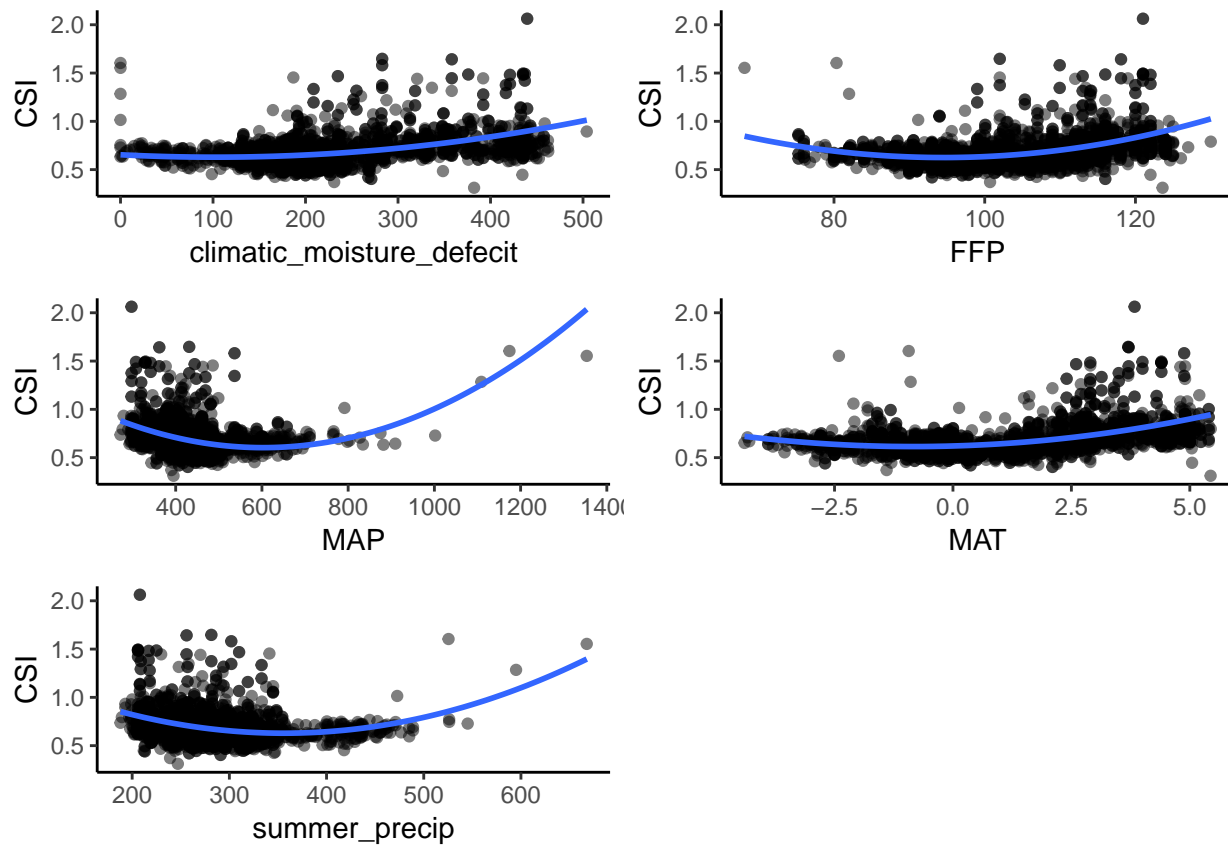


7. Examine patterns of CSI to HF across climatic gradients

Now we can examine whether CSI to disturbance is mediated by mean climatic conditions. First check to ensure that HF isn't correlated with climatic conditions: it definitely is not.



Next we can compare patterns of CSI and climatic conditoinis. There is an insufficient gradient in MAP or summer precip to really compare CSI across these variables. There are relationships between CSI and CMD (climatic moisture defecit), and with MAT: communities are relatively more sensitive to human development at either either ends of each climatic gradient. There is a general positive relationship between CSI and CMD, FFP, and MAT, indicating that communities are more sensitive to development in warmer, drier areas. This has important implications for climate change.

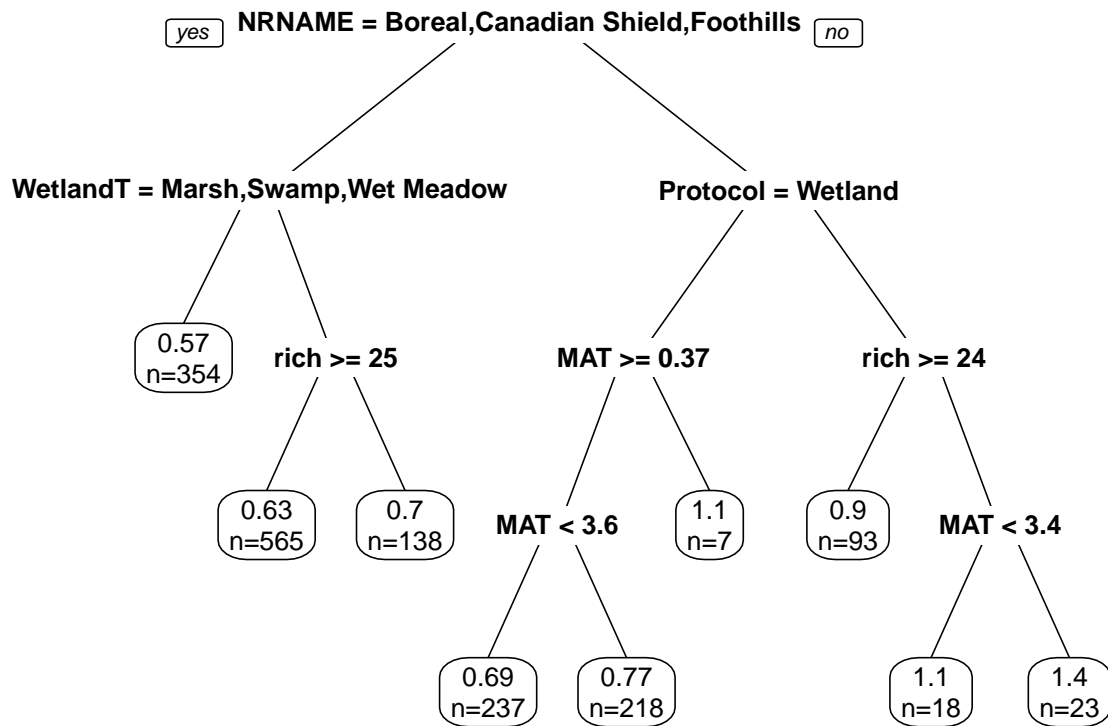


8. Choosing the most predictive variables

How can we choose the most appropriate variables to describe community sensitivity? Use regression trees and/or random forest. To predict the community sensitivity index (CSI), use the following predictors

- Protocol (terrestrial vs wetland)
- Natural region
- Wetland class
- Site-level species richness
- Climate variables

Regression tree predicting CSI - MAT



Now grow a random forest to compare the importance of each predictor.

```
##                               IncNodePurity
## Protocol                      7.860675
## NRNAME                       5.855853
## WetlandType                   1.774759
## rich                         6.512000
## climatic_moisture_defecit    4.308011
## MAT                          6.616138
## MAP                          3.100149
## FFP                          3.042378
## summer_precip                2.627107
## [1] 0.6962621
```

The final tree's pseudo- $R^2 = 0.7$.