Working version of Random Forests For Paper

Jeff Hollister

Wednesday, April 23, 2014

This document represents the various modeling efforts we have tried for the HKM2014 paper. First step in this analyis is to set up the groups of variables. Four groups, plus the Trophic State response variable were initially indentified.

First, are the water quality variables from NLA.

```
##  [1] "WSA_ECO9"  "BASINAREA" "DEPTHMAX"  "ELEV_PT"   "DO2_2M"
##  [6] "PH_FIELD"  "COND"      "ANC"       "TURB"      "TOC"
## [11] "DOC"       "NH4"       "NO3_NO2"   "NTL"       "PTL"
## [16] "CL"        "NO3"       "SO4"       "CA"        "MG"
## [21] "Na"        "K"         "COLOR"     "SIO2"      "H"
## [26] "OH"        "NH4ION"    "CATSUM"    "ANSUM2"    "ANDEF2"
## [31] "SOBC"      "BALANCE2"  "ORGION"    "CONCAL2"   "CONDHO2"
## [36] "SECMEAN"   "TminW"     "TmaxW"     "TmeanW"
```

Second,are the GIS based variables

```
##  [1] "MaxDist"          "AlbersX"          "AlbersY"
##  [4] "LakeArea"         "LakePerim"        "ShoreDevel"
##  [7] "DDs40"            "DDs45"            "DDs50"
## [10] "DDs55"            "MaxLength"        "MaxWidth"
## [13] "MeanWidth"        "FetchN"           "FetchNE"
## [16] "FetchE"           "FetchSE"          "MaxDepthCorrect"
## [19] "VolumeCorrect"    "MeanDepthCorrect"
```

Third, are the landscape variables based on total area

```
##  [1] "ImpervAreaKm2_MaxDist" "WaterKm2_MaxDist"
##  [3] "IceSnowKm2_MaxDist"    "DevOpenKm2_MaxDist"
##  [5] "DevLowKm2_MaxDist"     "DevMedKm2_MaxDist"
##  [7] "DevHighKm2_MaxDist"    "BarrenKm2_MaxDist"
##  [9] "DeciduousKm2_MaxDist"  "EvergreenKm2_MaxDist"
## [11] "MixedForKm2_MaxDist"   "ShrubKm2_MaxDist"
## [13] "GrassKm2_MaxDist"      "PastureKm2_MaxDist"
## [15] "CropsKm2_MaxDist"      "WoodyWetKm2_MaxDist"
## [17] "HerbWetKm2_MaxDist"    "ImpervAreaKm2_300m"
## [19] "WaterKm2_300m"         "IceSnowKm2_300m"
## [21] "DevOpenKm2_300m"       "DevLowKm2_300m"
## [23] "DevMedKm2_300m"        "DevHighKm2_300m"
## [25] "BarrenKm2_300m"        "DeciduousKm2_300m"
## [27] "EvergreenKm2_300m"     "MixedForKm2_300m"
## [29] "ShrubKm2_300m"         "GrassKm2_300m"
## [31] "PastureKm2_300m"       "CropsKm2_300m"
## [33] "WoodyWetKm2_300m"      "HerbWetKm2_300m"
## [35] "ImpervAreaKm2_1500m"   "WaterKm2_1500m"
## [37] "IceSnowKm2_1500m"      "DevOpenKm2_1500m"
## [39] "DevLowKm2_1500m"       "DevMedKm2_1500m"
## [41] "DevHighKm2_1500m"      "BarrenKm2_1500m"
```

```
## [43] "DeciduousKm2_1500m"    "EvergreenKm2_1500m"
## [45] "MixedForKm2_1500m"     "ShrubKm2_1500m"
## [47] "GrassKm2_1500m"        "PastureKm2_1500m"
## [49] "CropsKm2_1500m"        "WoodyWetKm2_1500m"
## [51] "HerbWetKm2_1500m"      "ImpervAreaKm2_3000m"
## [53] "WaterKm2_3000m"        "IceSnowKm2_3000m"
## [55] "DevOpenKm2_3000m"      "DevLowKm2_3000m"
## [57] "DevMedKm2_3000m"       "DevHighKm2_3000m"
## [59] "BarrenKm2_3000m"       "DeciduousKm2_3000m"
## [61] "EvergreenKm2_3000m"    "MixedForKm2_3000m"
## [63] "ShrubKm2_3000m"        "GrassKm2_3000m"
## [65] "PastureKm2_3000m"      "CropsKm2_3000m"
## [67] "WoodyWetKm2_3000m"     "HerbWetKm2_3000m"
```

Fourth, the percent landscape variables (also inlcluding percent impervious)

```
##  [1] "PercentImperv_MaxDist" "PercentImperv_300m"
##  [3] "PercentImperv_1500m"   "PercentImperv_3000m"
##  [5] "WaterPer_MaxDist"      "IceSnowPer_MaxDist"
##  [7] "DevOpenPer_MaxDist"    "DevLowPer_MaxDist"
##  [9] "DevMedPer_MaxDist"     "DevHighPer_MaxDist"
## [11] "BarrenPer_MaxDist"     "DeciduousPer_MaxDist"
## [13] "EvergreenPer_MaxDist"  "MixedForPer_MaxDist"
## [15] "ShrubPer_MaxDist"      "GrassPer_MaxDist"
## [17] "PasturePer_MaxDist"    "CropsPer_MaxDist"
## [19] "WoodyWetPer_MaxDist"   "HerbWetPer_MaxDist"
## [21] "WaterPer_300m"         "IceSnowPer_300m"
## [23] "DevOpenPer_300m"       "DevLowPer_300m"
## [25] "DevMedPer_300m"        "DevHighPer_300m"
## [27] "BarrenPer_300m"        "DeciduousPer_300m"
## [29] "EvergreenPer_300m"     "MixedForPer_300m"
## [31] "ShrubPer_300m"         "GrassPer_300m"
## [33] "PasturePer_300m"       "CropsPer_300m"
## [35] "WoodyWetPer_300m"      "HerbWetPer_300m"
## [37] "WaterPer_1500m"        "IceSnowPer_1500m"
## [39] "DevOpenPer_1500m"      "DevLowPer_1500m"
## [41] "DevMedPer_1500m"       "DevHighPer_1500m"
## [43] "BarrenPer_1500m"       "DeciduousPer_1500m"
## [45] "EvergreenPer_1500m"    "MixedForPer_1500m"
## [47] "ShrubPer_1500m"        "GrassPer_1500m"
## [49] "PasturePer_1500m"      "CropsPer_1500m"
## [51] "WoodyWetPer_1500m"     "HerbWetPer_1500m"
## [53] "WaterPer_3000m"        "IceSnowPer_3000m"
## [55] "DevOpenPer_3000m"      "DevLowPer_3000m"
## [57] "DevMedPer_3000m"       "DevHighPer_3000m"
## [59] "BarrenPer_3000m"       "DeciduousPer_3000m"
## [61] "EvergreenPer_3000m"    "MixedForPer_3000m"
## [63] "ShrubPer_3000m"        "GrassPer_3000m"
## [65] "PasturePer_3000m"      "CropsPer_3000m"
## [67] "WoodyWetPer_3000m"     "HerbWetPer_3000m"
```

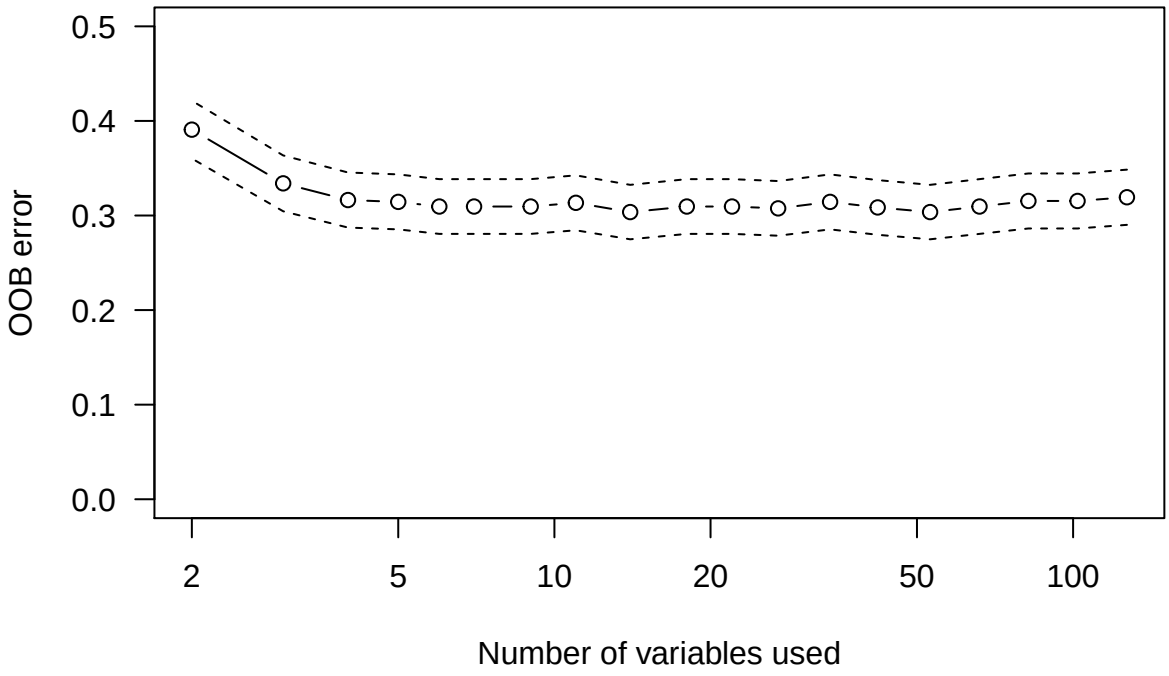
Lastly, the response variable

```
## [1] "TS_CHLA"
```

Chl *a* Trophic Status ~ All Variables + Landscape Total Area

This first random forest predicts Cholorophyl Trophic status using all NLA variables, all GIS variables except for the landscape variables summarized by total area.

```
##  Number.Variables
##  Min.   :  2.0
##  1st Qu.:  6.5
##  Median : 18.0
##  Mean   : 33.4
##  3rd Qu.: 47.5
##  Max.   :127.0
##
##                                                 Vars.in.Forest
##  SECMEAN + TURB                                               : 1
##  SECMEAN + TURB + PTL                                         : 1
##  SECMEAN + TURB + PTL + NTL                                   : 1
##  SECMEAN + TURB + PTL + NTL + TOC                             : 1
##  SECMEAN + TURB + PTL + NTL + TOC + WSA_ECO9                  : 1
##  SECMEAN + TURB + PTL + NTL + TOC + WSA_ECO9 + DEPTHMAX       : 1
##  (Other)                                                      :13
##       OOB             sd.OOB
##  Min.   :0.304   Min.   :0.0144
##  1st Qu.:0.309   1st Qu.:0.0145
##  Median :0.309   Median :0.0145
##  Mean   :0.316   Mean   :0.0145
##  3rd Qu.:0.315   3rd Qu.:0.0145
##  Max.   :0.391   Max.   :0.0153
##
```

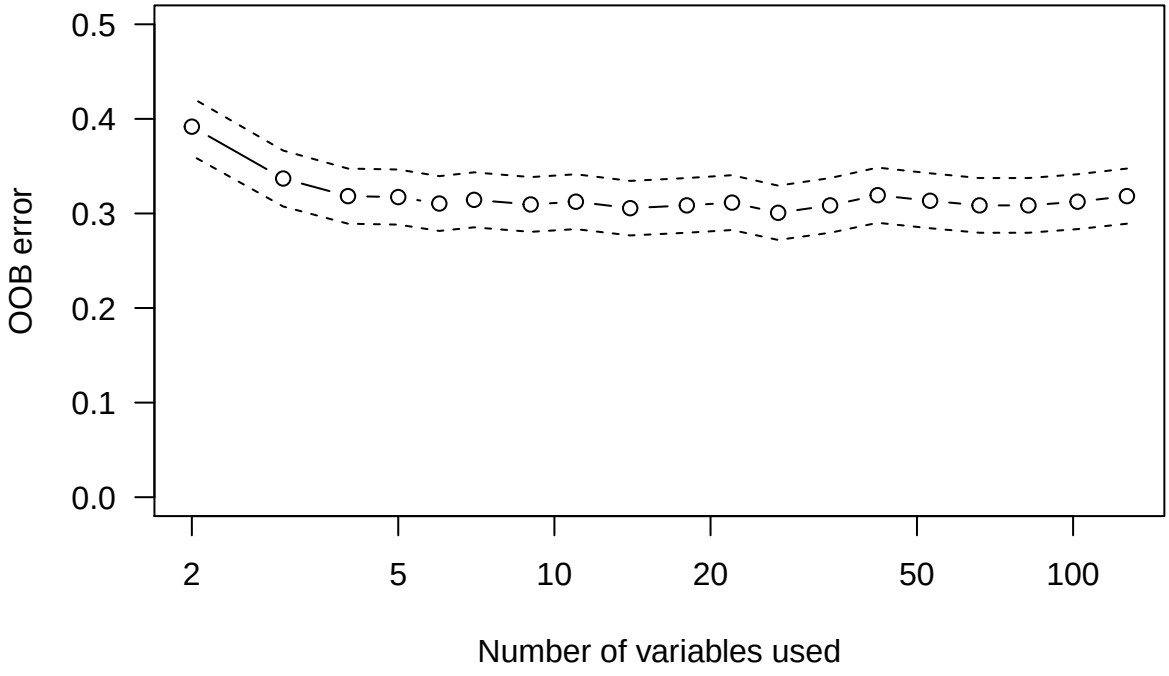


Chl *a* Trophic Status ~ All Variables + Landscape Percent

Our second model uses all NLA and GIS variables, and landscape var summarized by total percent.

```
##  Number.Variables
##  Min.   :  2.0
##  1st Qu.:  6.5
##  Median : 18.0
##  Mean   : 33.4
##  3rd Qu.: 47.5
##  Max.   :127.0
##
##                                                 Vars.in.Forest
##  SECMEAN + TURB                                        : 1
##  SECMEAN + TURB + PTL                                  : 1
##  SECMEAN + TURB + PTL + NTL                            : 1
##  SECMEAN + TURB + PTL + NTL + TOC                      : 1
##  SECMEAN + TURB + PTL + NTL + TOC + WSA_ECO9           : 1
##  SECMEAN + TURB + PTL + NTL + TOC + WSA_ECO9 + DEPTHMAX: 1
##  (Other)                                               :13
##       OOB             sd.OOB
##  Min.   :0.301   Min.   :0.0144
##  1st Qu.:0.308   1st Qu.:0.0145
##  Median :0.312   Median :0.0145
##  Mean   :0.317   Mean   :0.0146
```

```
##  3rd Qu.:0.318   3rd Qu.:0.0146
##  Max.   :0.392   Max.   :0.0153
##
```

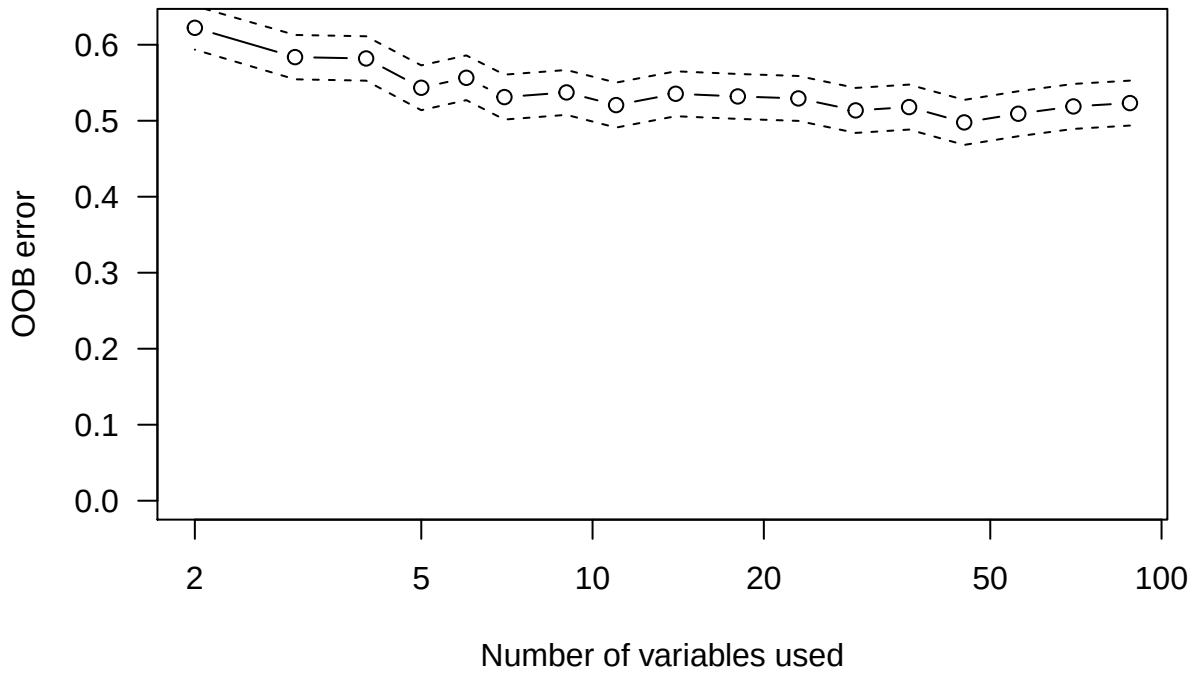


Chl *a* Trophic Status ~ GIS Only Variables + Landscape Total Area

This third model uses only GIS derived variables with landscape summarized by total area

```
##  Number.Variables
##  Min.   : 2.0
##  1st Qu.: 6.0
##  Median :14.0
##  Mean   :25.1
##  3rd Qu.:36.0
##  Max.   :88.0
##
##
##  EvergreenKm2_3000m + CropsKm2_3000m
##  EvergreenKm2_3000m + CropsKm2_3000m + AlbersX
##  EvergreenKm2_3000m + CropsKm2_3000m + AlbersX + CropsKm2_1500m
##  EvergreenKm2_3000m + CropsKm2_3000m + AlbersX + CropsKm2_1500m + MeanDepthCorrect
##  EvergreenKm2_3000m + CropsKm2_3000m + AlbersX + CropsKm2_1500m + MeanDepthCorrect + EvergreenKm2_150
##  EvergreenKm2_3000m + CropsKm2_3000m + AlbersX + CropsKm2_1500m + MeanDepthCorrect + EvergreenKm2_150
##  (Other)
##       OOB            sd.OOB
```

```
##  Min.   :0.498   Min.   :0.0144
##  1st Qu.:0.519   1st Qu.:0.0147
##  Median :0.531   Median :0.0148
##  Mean   :0.538   Mean   :0.0147
##  3rd Qu.:0.543   3rd Qu.:0.0148
##  Max.   :0.622   Max.   :0.0148
##
```

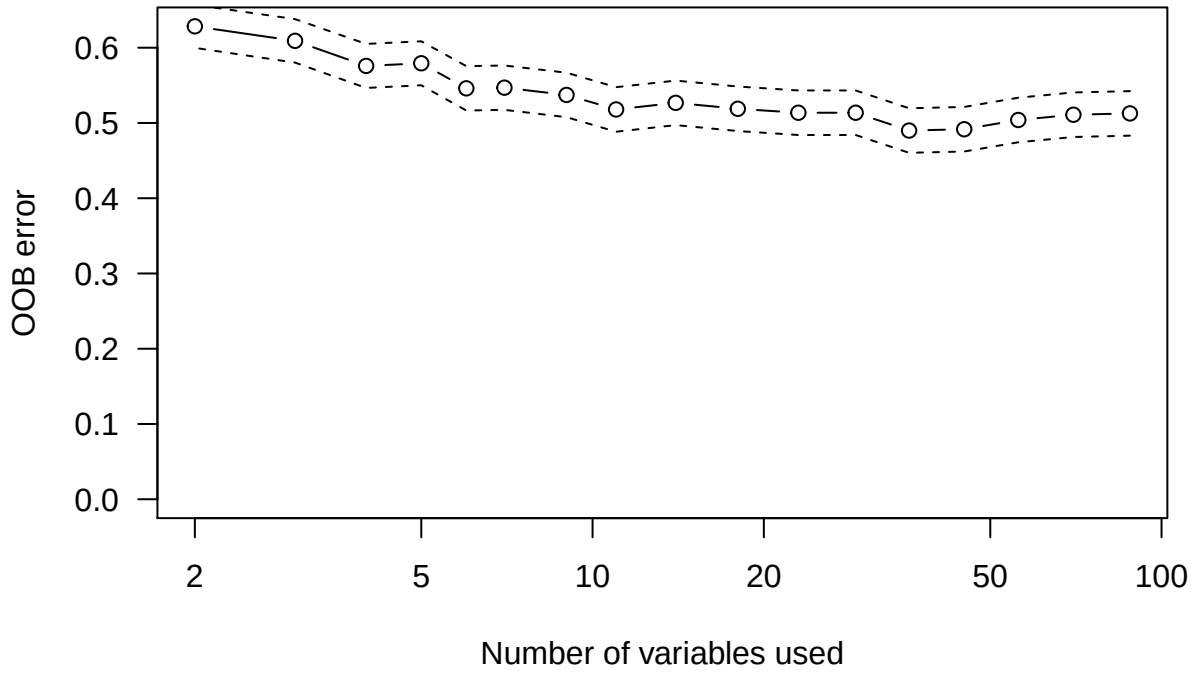


Chl *a* Trophic Status ~ GIS Only Variables +Landscape Percent

This fourth model uses only GIS derived variables with landscape summarized by total percent.

```
##  Number.Variables
##  Min.   : 2.0
##  1st Qu.: 6.0
##  Median :14.0
##  Mean   :25.1
##  3rd Qu.:36.0
##  Max.   :88.0
##
##
##  EvergreenPer_3000m + CropsPer_3000m
##  EvergreenPer_3000m + CropsPer_3000m + CropsPer_1500m
##  EvergreenPer_3000m + CropsPer_3000m + CropsPer_1500m + AlbersX
##  EvergreenPer_3000m + CropsPer_3000m + CropsPer_1500m + AlbersX + EvergreenPer_MaxDist
```

```
##  EvergreenPer_3000m + CropsPer_3000m + CropsPer_1500m + AlbersX + EvergreenPer_MaxDist + AlbersY
##  EvergreenPer_3000m + CropsPer_3000m + CropsPer_1500m + AlbersX + EvergreenPer_MaxDist + AlbersY + E
##  (Other)
##       OOB             sd.OOB
##  Min.   :0.490   Min.   :0.0143
##  1st Qu.:0.513   1st Qu.:0.0147
##  Median :0.519   Median :0.0148
##  Mean   :0.537   Mean   :0.0147
##  3rd Qu.:0.547   3rd Qu.:0.0148
##  Max.   :0.628   Max.   :0.0148
##
```

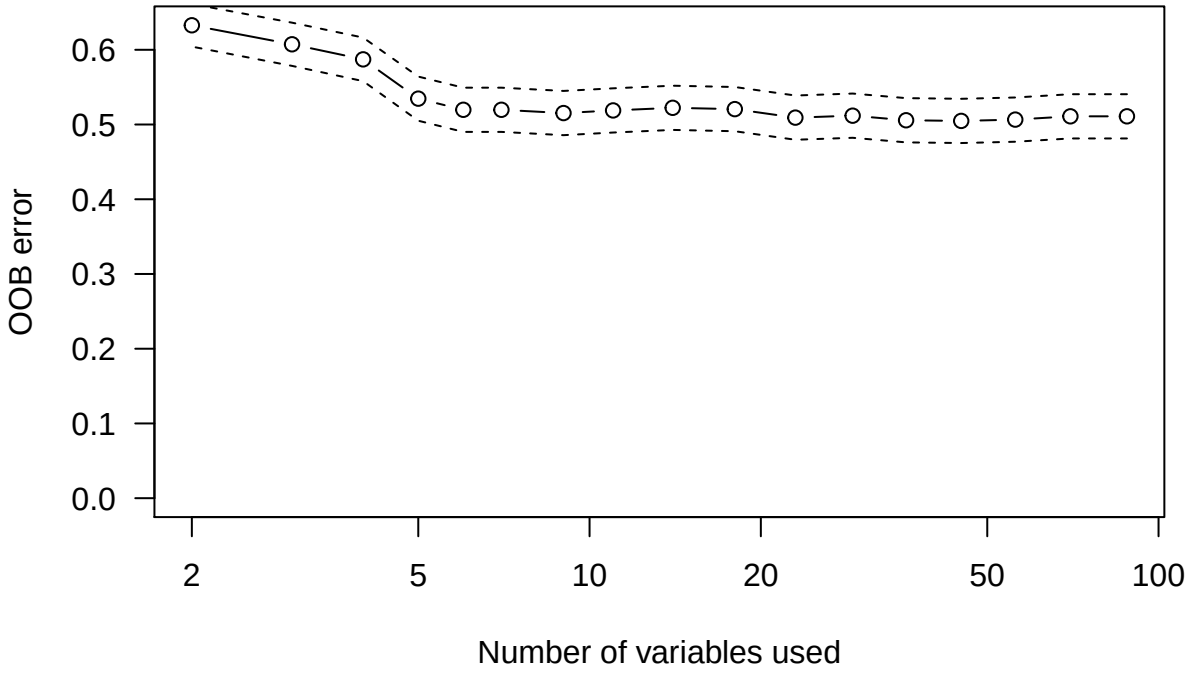


Chl *a* Trophic Status ~ GIS Only Variables + Landscape Total Area/Lake Area

This fith model uses only GIS derived variables with landscape summarized by total area normalized by area of the lake.

```
##  Number.Variables
##  Min.   : 2.0
##  1st Qu.: 6.0
##  Median :14.0
##  Mean   :25.1
##  3rd Qu.:36.0
##  Max.   :88.0
##
```

```
##
##  EvergreenKm2_3000m + AlbersX
##  EvergreenKm2_3000m + AlbersX + EvergreenKm2_MaxDist
##  EvergreenKm2_3000m + AlbersX + EvergreenKm2_MaxDist + CropsKm2_1500m
##  EvergreenKm2_3000m + AlbersX + EvergreenKm2_MaxDist + CropsKm2_1500m + MeanDepthCorrect
##  EvergreenKm2_3000m + AlbersX + EvergreenKm2_MaxDist + CropsKm2_1500m + MeanDepthCorrect + AlbersY
##  EvergreenKm2_3000m + AlbersX + EvergreenKm2_MaxDist + CropsKm2_1500m + MeanDepthCorrect + AlbersY +
##  (Other)
##       OOB             sd.OOB
##  Min.   :0.505   Min.   :0.0143
##  1st Qu.:0.511   1st Qu.:0.0148
##  Median :0.519   Median :0.0148
##  Mean   :0.532   Mean   :0.0147
##  3rd Qu.:0.522   3rd Qu.:0.0148
##  Max.   :0.633   Max.   :0.0148
##
```



Chl *a* Trophic Status ~ GIS Only Variables + Landscape Total Area/Lake Volume

This sixth model uses only GIS derived variables with landscape summarized by total area normalized by lake volume.

```
##  Number.Variables
```

```
##  Min.   : 2.0
##  1st Qu.: 6.0
##  Median :14.0
##  Mean   :25.1
##  3rd Qu.:36.0
##  Max.   :88.0
##
##
##  CropsKm2_3000m + CropsKm2_1500m
##  CropsKm2_3000m + CropsKm2_1500m + AlbersX
##  CropsKm2_3000m + CropsKm2_1500m + AlbersX + EvergreenKm2_3000m
##  CropsKm2_3000m + CropsKm2_1500m + AlbersX + EvergreenKm2_3000m + MeanDepthCorrect
##  CropsKm2_3000m + CropsKm2_1500m + AlbersX + EvergreenKm2_3000m + MeanDepthCorrect + CropsKm2_300m
##  CropsKm2_3000m + CropsKm2_1500m + AlbersX + EvergreenKm2_3000m + MeanDepthCorrect + CropsKm2_300m +
##  (Other)
##       OOB             sd.OOB
##  Min.   :0.505   Min.   :0.0134
##  1st Qu.:0.514   1st Qu.:0.0148
##  Median :0.519   Median :0.0148
##  Mean   :0.536   Mean   :0.0147
##  3rd Qu.:0.538   3rd Qu.:0.0148
##  Max.   :0.709   Max.   :0.0148
##
```