R Notebook

# Spearman Correlations

### Watershed Characteristics

Soil Erodibility Factor and Mean Soil Erodibility were significantly correlated with the other watershed characteristics (r > 0.5), and therefore, were removed from the analysis.

Watershed Chracteristics

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Area | Karst | Precip | Soil Erodibility Factor | Mean Soil Erodibility | Physiographic Province | Average Watershed Slope | Soil Group | Ecoregion | Ecoregion4 | Ecoregion3 |
| Area | 1.00 |  |  |  |  |  |  |  |  |  |  |
| Karst | 0.21 | 1 |  |  |  |  |  |  |  |  |  |
| Precip | -0.11 | -0.17 | 1 |  |  |  |  |  |  |  |  |
| Soil Erodibility Factor | -0.14 | 0.22 | 0.39 | 1 |  |  |  |  |  |  |  |
| Mean Soil Erodibility | -0.03 | 0.11 | 0.34 | 0.64 | 1 |  |  |  |  |  |  |
| Physiographic Province | 0.14 | 0.42 | -0.71 | -0.2 | -0.3 | 1 |  |  |  |  |  |
| Average Watershed Slope | 0.13 | -0.19 | -0.42 | -0.76 | -0.73 | 0.27 | 1 |  |  |  |  |
| Soil Group | -0.07 | -0.12 | 0.04 | 0.06 | 0.04 | -0.05 | -0.13 | 1 |  |  |  |
| Ecoregion | -0.06 | 0.26 | 0.17 | 0.58 | 0.47 | -0.02 | -0.62 | 0.1 | 1 |  |  |
| Ecoregion4 | -0.08 | -0.26 | -0.37 | -0.56 | -0.56 | 0.28 | 0.64 | 0.01 | -0.67 | 1 |  |
| Ecoregion3 | -0.06 | 0.16 | -0.35 | -0.3 | -0.43 | 0.42 | 0.44 | -0.02 | -0.29 | 0.78 | 1 |

## Alteration x Alteration

In both scenarios 3-Day Minimum (cfs/sqmi), Extreme Low Flow Frequency, and Extreme Low Flow Duration were ignored because they were not present in the Phase 1 analysis or the Middle Potomac study. In subsequent analyses, these three alteration metrics were removed.

In the Baseline/Current scenario Flashiness was significantly correlated (r > 0.50) with MH21, DH17, High Pulse Count, and 3-Day Maximum (cfs/sqmi). However, in the Baseline/Impervious scenario Flashiness was only significantly correlated with High Pulse Count and 3-Day Maximum (cfs/sqmi). DH17 and MH21 were significantly correlated (r = 0.86). MH21 was excluded becuase it was slightly more correlated with Flashiness (r = -0.40). Flashiness and DH17 were retained for further analysis.

In both scenarios, Low Pulse Duration was not significantly correlated with any of the other alteration metrics. Therefore, Low Pulse Duration was retained for further analyses.

### Baseline/Current Scenario

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | High flow index, MH21 (days) | High flow duration, DH17 (days) | High pulse count (#) | Flashiness (ratio) | Low pulse duration (days) | 3-day maximum (cfs/sqmi) | 3-day minimum (cfs/sqmi) | Extreme low flow frequency | Extreme low flow duration |
| High flow index, MH21 (days) | 1.00 |  |  |  |  |  |  |  |  |
| High flow duration, DH17 (days) | 0.93 | 1 |  |  |  |  |  |  |  |
| High pulse count (#) | -0.81 | -0.8 | 1 |  |  |  |  |  |  |
| Flashiness (ratio) | -0.88 | -0.89 | 0.91 | 1 |  |  |  |  |  |
| Low pulse duration (days) | 0.47 | 0.46 | -0.41 | -0.43 | 1 |  |  |  |  |
| 3-day maximum (cfs/sqmi) | -0.82 | -0.83 | 0.91 | 0.95 | -0.41 | 1 |  |  |  |
| 3-day minimum (cfs/sqmi) | -0.08 | 0 | 0.03 | 0.03 | -0.04 | 0.12 | 1 |  |  |
| Extreme low flow frequency | -0.14 | -0.21 | 0.25 | 0.25 | -0.05 | 0.2 | -0.72 | 1 |  |
| Extreme low flow duration | 0.45 | 0.41 | -0.48 | -0.48 | 0.31 | -0.51 | -0.1 | -0.11 | 1 |

### Baseline/Impervious Scenario

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | High flow index, MH21 (days) | High flow duration, DH17 (days) | High pulse count (#) | Flashiness (ratio) | Low pulse duration (days) | 3-day maximum (cfs/sqmi) | 3-day minimum (cfs/sqmi) | Extreme low flow frequency | Extreme low flow duration |
| High flow index, MH21 (days) | 1.00 |  |  |  |  |  |  |  |  |
| High flow duration, DH17 (days) | 0.86 | 1 |  |  |  |  |  |  |  |
| High pulse count (#) | -0.60 | -0.53 | 1 |  |  |  |  |  |  |
| Flashiness (ratio) | -0.40 | -0.39 | 0.84 | 1 |  |  |  |  |  |
| Low pulse duration (days) | 0.07 | -0.01 | -0.1 | -0.16 | 1 |  |  |  |  |
| 3-day maximum (cfs/sqmi) | -0.21 | -0.22 | 0.7 | 0.73 | -0.09 | 1 |  |  |  |
| 3-day minimum (cfs/sqmi) | 0.05 | 0.05 | -0.18 | -0.33 | 0.36 | -0.26 | 1 |  |  |
| Extreme low flow frequency | 0.06 | 0.07 | 0.28 | 0.3 | -0.15 | 0.35 | -0.31 | 1 |  |
| Extreme low flow duration | 0.00 | -0.03 | -0.27 | -0.31 | 0.37 | -0.22 | 0.48 | -0.38 | 1 |

## Alteration x Watershed Characteristic

In both scenarios Low Pulse Duration was not significantly correlated (-0.5 < r < 0.5) with any watershed characteristics.

In the Baseline/Current scenario, DH17 was significantly correlated with Soil Erodibility Factor, Mean Soil Erodbility. However, DH17 was not signficantly correlated with any watershed characteristics in the Baseline/Impervious scenario.

In the Baseline/Current scenario, Flashiness was significantly correlated with FCODEgraphic Province and Average Watershed Slope. However, Flashiness was only significanly correlated with Average watershed Slop in the Baseline/Impervious scenario. In both scenarios Flashiness was negatively correlated with Average Watershed Slope but the relationship was not as strong in the Baseline/Impervious scenario.

### Baseline/Current Scenario

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Area | Karst | Precip | Soil Erodibility Factor | Mean Soil Erodibility | Physiographic Province | Average Watershed Slope | Soil Group |
| High flow index, MH21 (days) | 0.07 | -0.30 | -0.16 | -0.63 | -0.49 | -0.01 | 0.66 | -0.07 |
| High flow duration, DH17 (days) | 0.12 | -0.22 | -0.24 | -0.67 | -0.50 | 0.07 | 0.68 | -0.10 |
| High pulse count (#) | -0.16 | 0.28 | 0.37 | 0.76 | 0.59 | -0.14 | -0.74 | 0.06 |
| Flashiness (ratio) | -0.11 | 0.24 | 0.44 | 0.79 | 0.64 | -0.21 | -0.81 | 0.06 |
| Low pulse duration (days) | -0.06 | -0.17 | -0.12 | -0.34 | -0.23 | 0.02 | 0.27 | 0.00 |
| 3-day maximum (cfs/sqmi) | -0.11 | 0.30 | 0.37 | 0.80 | 0.65 | -0.14 | -0.81 | 0.00 |

### Baseline/Impervious Scenario

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Area | Karst | Precip | Soil Erodibility Factor | Mean Soil Erodibility | Physiographic Province | Average Watershed Slope | Soil Group |
| High flow index, MH21 (days) | -0.05 | -0.49 | 0.43 | 0.04 | 0.12 | -0.49 | -0.01 | 0.07 |
| High flow duration, DH17 (days) | -0.07 | -0.47 | 0.33 | -0.03 | 0.06 | -0.39 | 0.05 | -0.02 |
| High pulse count (#) | -0.12 | 0.41 | -0.07 | 0.50 | 0.39 | 0.23 | -0.42 | -0.09 |
| Flashiness (ratio) | -0.14 | 0.32 | 0.28 | 0.65 | 0.45 | -0.03 | -0.56 | -0.04 |
| Low pulse duration (days) | 0.01 | 0.01 | -0.14 | 0.00 | 0.00 | 0.11 | -0.05 | 0.06 |
| 3-day maximum (cfs/sqmi) | -0.15 | 0.22 | 0.16 | 0.53 | 0.47 | 0.02 | -0.49 | 0.07 |
| ## Alteration x %Impervious Cover |  |  |  |  |  |  |  |  |
| The percent change in Impervious C | over was | held cons | tant in th | e |  |  |  |  |
| Baseline/Impervious scenario, and | therefore | , correla | tions will | not |  |  |  |  |
| exist between the Alteration metri | cs and th | e percent | change in |  |  |  |  |  |
| Impervious cover. Low Pulse Durati | on was no | t signfic | antly corr | elated with |  |  |  |  |
| the Percentage of Impervious Cover | Change i | n the Bas | eline/Curr | ent Scenario. |  |  |  |  |
| Percent Impervious Cover Change wa | s negativ | ely corre | lated with | DH17 and |  |  |  |  |
| positively correlated with Flashin | ess in th | e Baselin | e/Current | scenario. |  |  |  |  |

### Baseline/Current Scenario

|  |  |
| --- | --- |
|  | Impervious Cover (%) |
| High flow index, MH21 (days) | -0.88 |
| High flow duration, DH17 (days) | -0.89 |
| High pulse count (#) | 0.84 |
| Flashiness (ratio) | 0.91 |
| Low pulse duration (days) | -0.46 |
| 3-day maximum (cfs/sqmi) | 0.87 |

# Loess Curves

The color gradient in all loess plots belwo is based on the percentage of impervious cover in the Current scenario. The percentage of impervious cover in each watershed from the Baseline scenario is approximately zero. To create the Impervious Cover scenario, the percentage of impervious cover in each watershed was increased by 10%.

Categorical variables, like FCODE (FCODEgraphic Province) and Soil, are not appropriate for Loess curves. However, I left these plots in because they have the potental to inform patterns in the other plots.

In general, the Baseline/Impervious Loess Curves show similar but weaker trends to the Baseline/Current.

## Flashiness

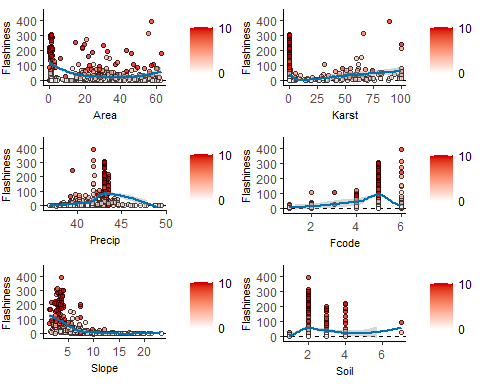
Flashiness in the Baseline/Current scenario had a slight concave unimodal distribution when plotted against Area. This pattern did not exist for the Baseline/Impervious scenario. Most likely the pattern observed in the Baseline/Current had to do with a development bias. The percent change in impervious cover can change more rapidly in small areas. For example, a equally sized parking lots could be added to a small watershed and a large watershed. Relative to the the large watershed, the proportion of area disturbed by the parking lot is greater in the smaller watershed. Therfore, minor land cover changes that occur in small Baseline/Current watersheds may have a large influence on Flashiness values. When land cover changes are held constant in the Baseline/Current scenario, the concave unimodal distribution disappears.

The relationship beteen Karst and Flashiness appears to be slightly stronger in the Baseline/Impervious scenario. As the percentage of Karst in a watershed increases, Flashiness may also increase.

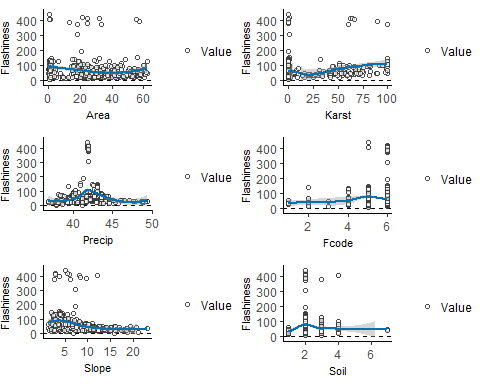
In both scenarios, Percipitation had an interesting convex unimodal distribution when plotted against Flashiness. Although Precipitation was not sigificantly correlated with FCODEgraphic Province (r = -0.45), the most obvious explanation for this unimodal distiribution would be that certain areas receive more rain than others. One way to look into this further would be to dividethe Ridge and Valley FCODEgrapic Province (FCODE = 62) into two seperate provinces. My current assumption is that Ridges receive more rain than the other provinces but this is being masked by grouping Ridges with Valleys.

Slope was negatively correlated with Flashiness in both scenarios (See correlation section). Development is more prominent in flat areas (low slope), therefore, in the Baseline/Current scenario the negative correlation was assumed to be an artifact development bias. Although the Baseline/Imperious scenario correlation was weaker, there does appear to be a negative correlation between Slope and Flashiness, even if the percent change in impervious cover is a constant.

### Baseline/Current Scenario



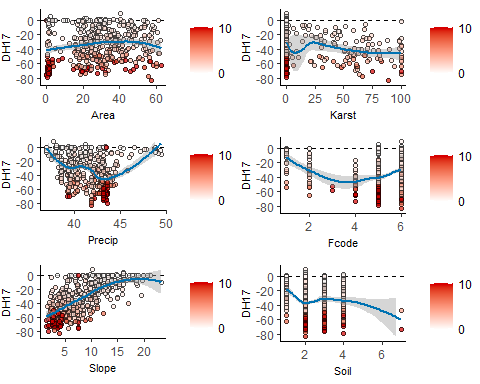
### Baseline/Impervious Scenario

 ## DH17 Area and Karst had similar distirbutions in both scenarios but the values were generally lower and variability was generally smaller in the Baseline/Impervious scenario.

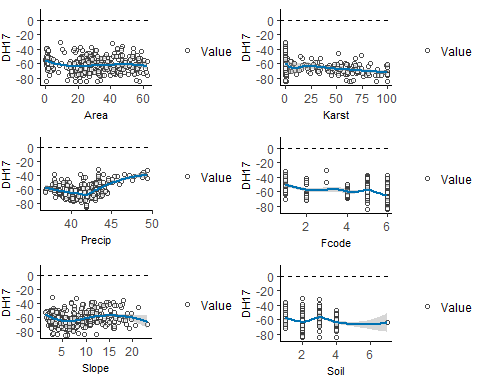
The Baseline/Current scenario had a concave, almost, biomodal distribution between Precipitation and DH17. This pattern was dramatically reduced in the Baseline/Impervious scenario. There did appear to be a positive relationship between Perceiptation and DH17 for percipitation values greater than 42.

The was a strong positive correlation between Slope and DH17 in the Baseline/Current scenario. However, this relationship did not exist in the Baseline/Impervious scenario. As was stated in the Flashiness section, the relationship observed in the Baseline/Current scenario may be due to a development bias. Development occurs more frequently in flat areas.

### Baseline/Current Scenario



### Baseline/Impervious Scenario

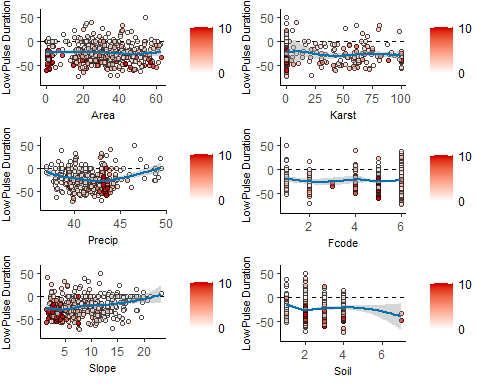
 ## Low Pulse Duration

Area and Karst had similar distirbutions in both scenarios but the values were generally lower and variability was generally smaller in the Baseline/Impervious scenario.

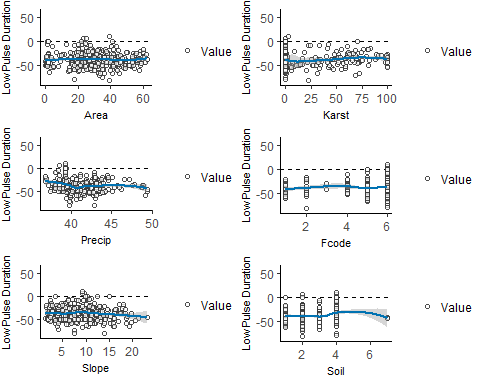
In the Baseline/Current scenario, the relationship between Precipiation and Low Pulse Duration was a slightly concave unimodal distribution. There was no relationship in the Basline/Impervious scenario and there was generally less variance.

There was a positive relationship between Slope and Low Pulse Duration in the Baseline/Current scenario. There was no relationship observed between Slope and Low Pulse Duration in the Basline/Impervious scenario.

### Baseline/Current Scenario



### Baseline/Impervious Scenario

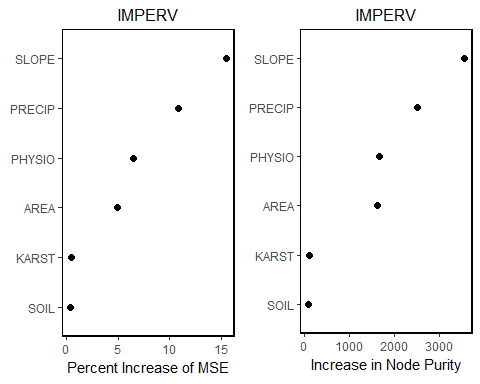
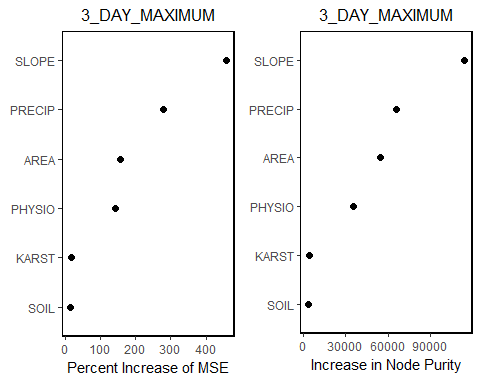
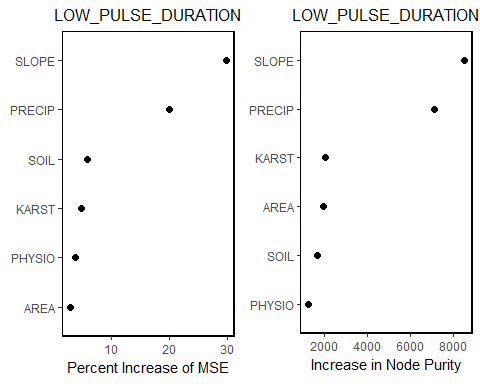
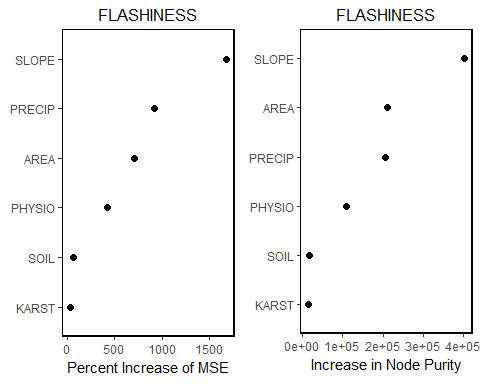
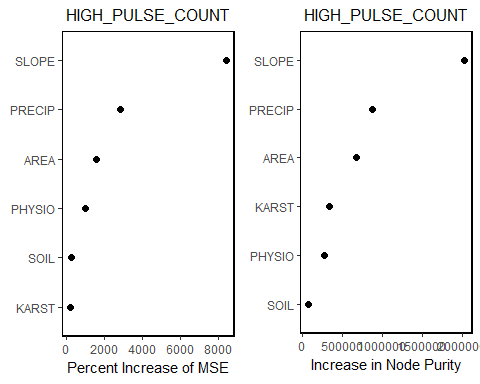
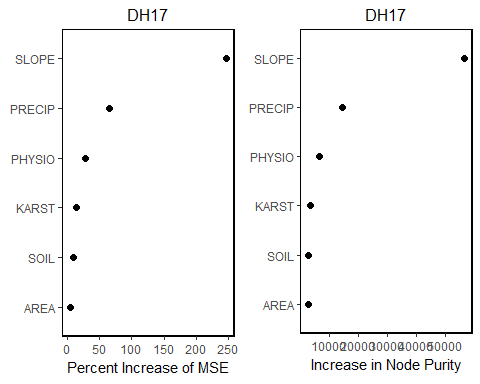
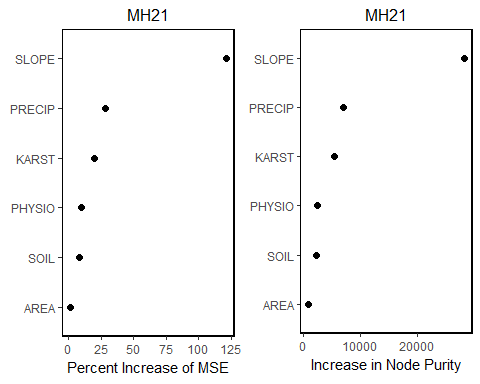


# Random Forest

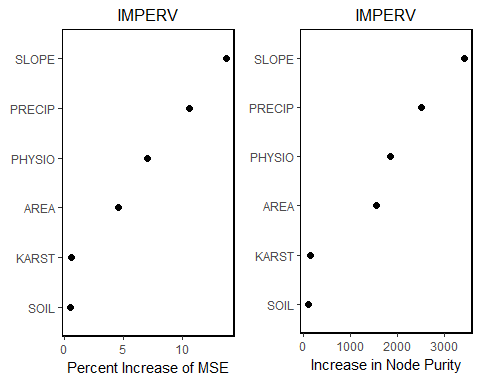
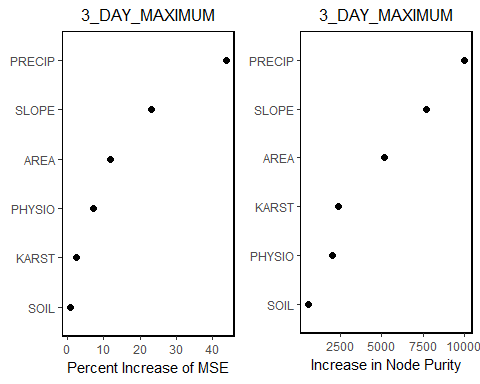
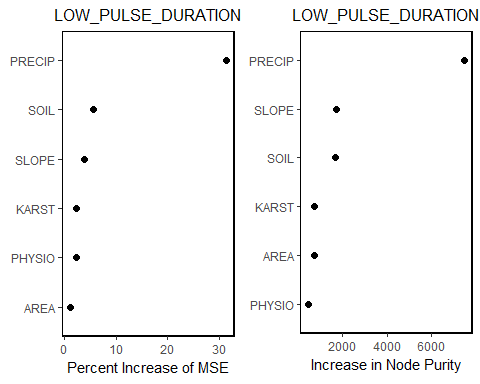
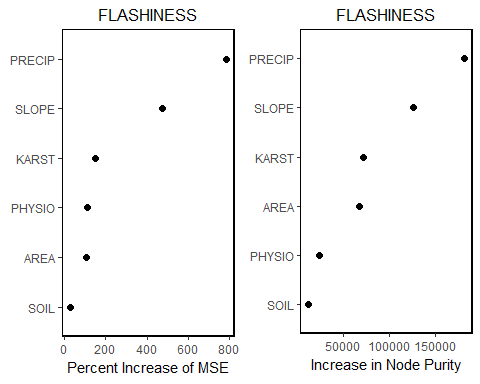
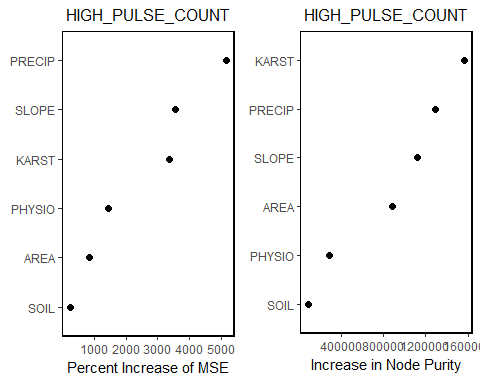
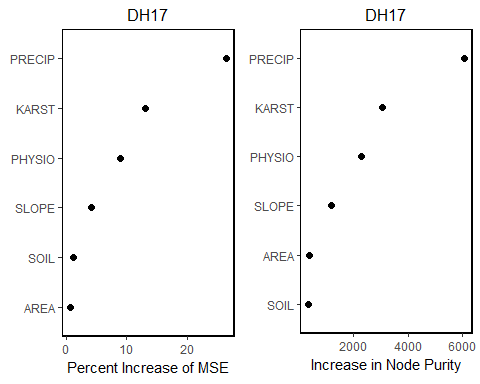
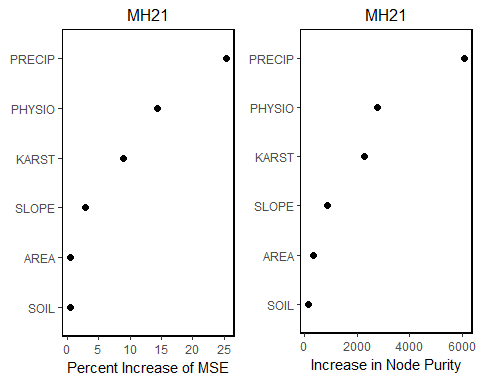
In the Baseline/Current scenario, Slope was consistently the most important watershed characteristic for predicting Alteration Metric values. Precipiation and Area were also important in the Flashiness Random Forest model. Similaraly, Precipiation was the second most import preditive variable in the Low Pulse Duration Random Forest model.

In the Baseline/Impervious scenario, Slope dropped out as the most important predictive variable and was replaced with Precipiation. Karst was the second most important predictive variable for the DH17 Random Forest model. For the Flashiness Random Forest model, Slope and Karst were also important predictive variables. Slope was the second most important predictive variable for the Low Pulse Duration Random Forest model but appeared to be relatively weak.

### Baseline/Current Scenario



### Baseline/Impervious Scenario



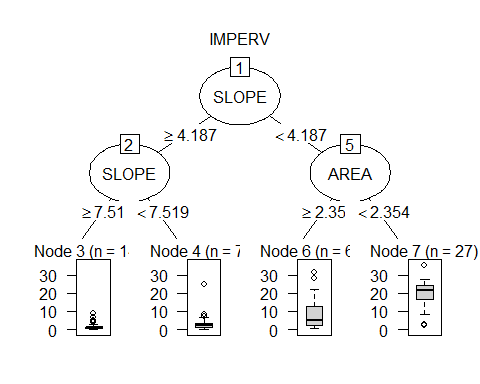
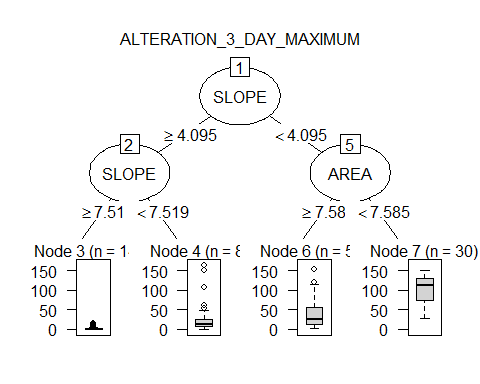
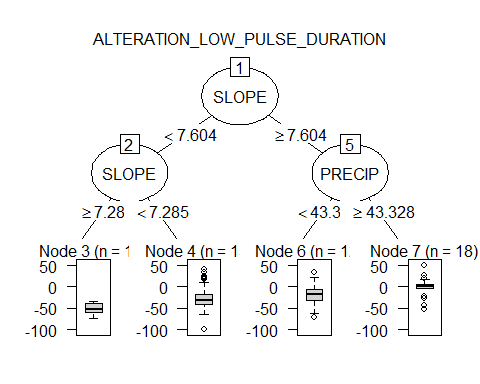
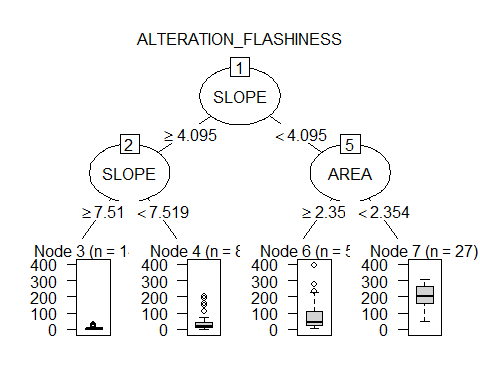
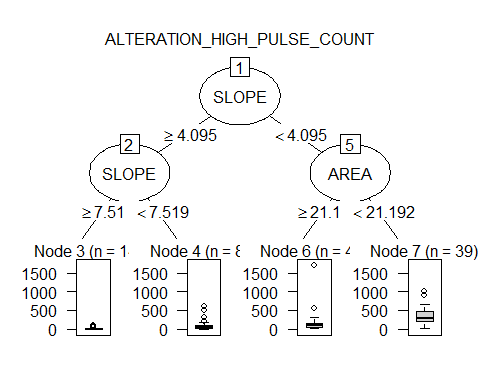
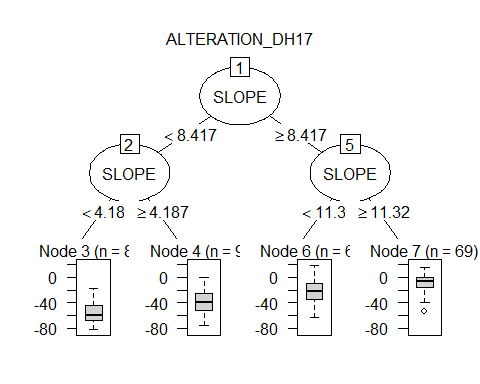
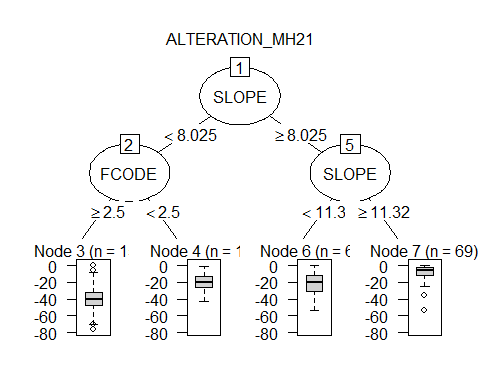
# Rpart

The *caret* package function train was used to create the rpart trees. A maxdepth of 4 was used. Addtionally, trainControl was set to method = "repeatedcv", number = 10, repeats = 100; in other words a 10 fold cross validation was performed 100 times.

### Baseline/Current Scenario

## Loading required package: grid

## Loading required package: lattice



### Baseline/Impervious Scenario

