* **Data**

48 observations (sites) in total

3 sites with ER\_sed >0

Only 45 sites were included in the analysis after removing missing values and positive ER\_sed values

* **Variables**

**target variable**:

ERsed\_gm2day

**regression variables**:

HOBO\_Temperature(water temperature),

log(TSS),velocity\_ms, log(slope) , discharge,

log(totdasqkm (total drainage area),

PctMxFst2019Ws (%forest) ,log(PctCrop2019Ws (%crop)),

log(D50\_m), aridity

**Notes**:

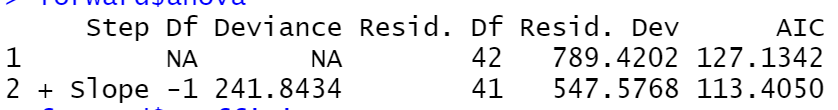
a) Log transfer for slope, D50\_m, totdasqkm and TSS(i.e., log(drained area), log(TSS))

b) remove positive ERsed values

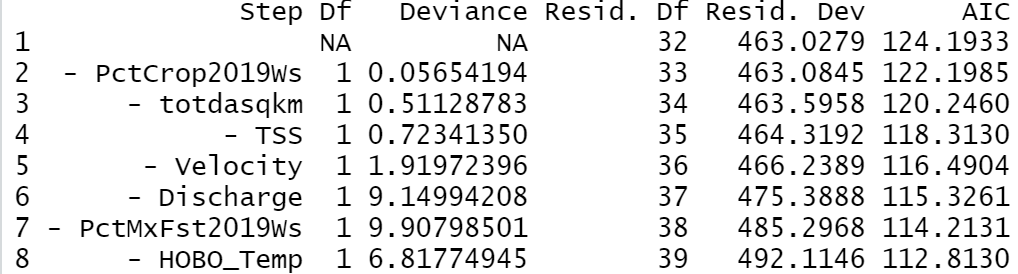
Scatter plots between ‘ERsed\_gm2day’ and the exploratory variable are saved in this folder [ERsed\_scatterplot](https://pnnl.sharepoint.com/:f:/r/teams/RC-2RiverCorridorSFA/Shared%20Documents/Manuscripts/2022_Spatial_Sed_Resp/MLR_results/ERsed/ERsed_scatterplot?csf=1&web=1&e=tfkUka)

* **Performs stepwise model selection by Akaike Information Criteria-AIC (both forward and backward selection)**

1. **Forward stepwise-****selected variables (variables to keep in model):**

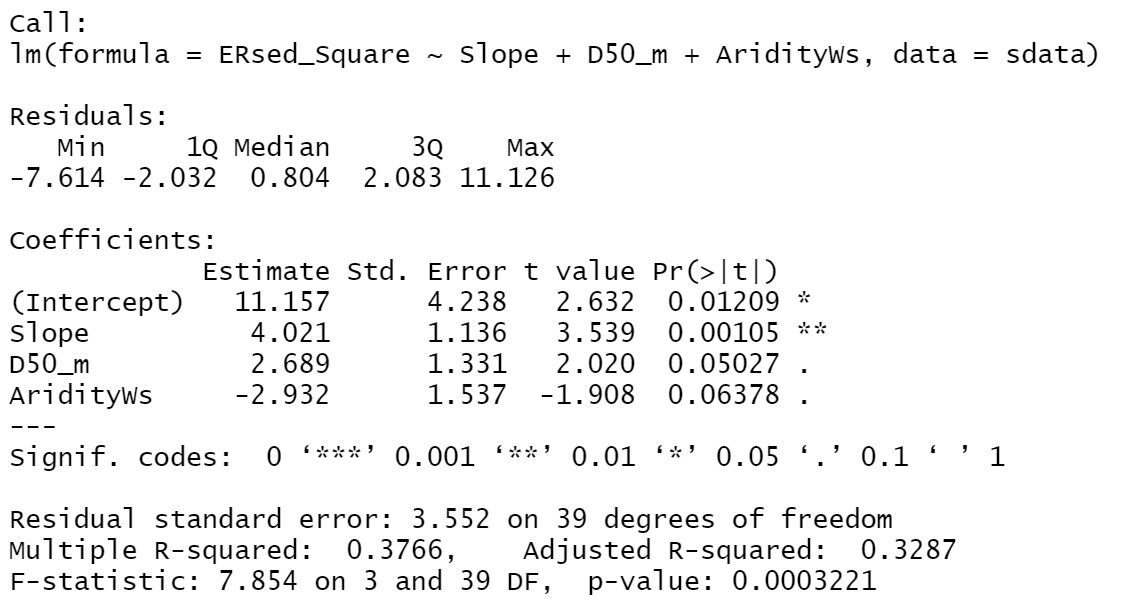


2) **backward stepwise-selected variables (variables to remove from model):**



Forward selection shows Slope is the only significant variable for ERsed, while backward selection shows Slope, AridityWs, and D50s are significant variable for ERsed

Summary statistics for the lm model fitted using the selected variables(i.e., Slope, AridityWs, and D50s)



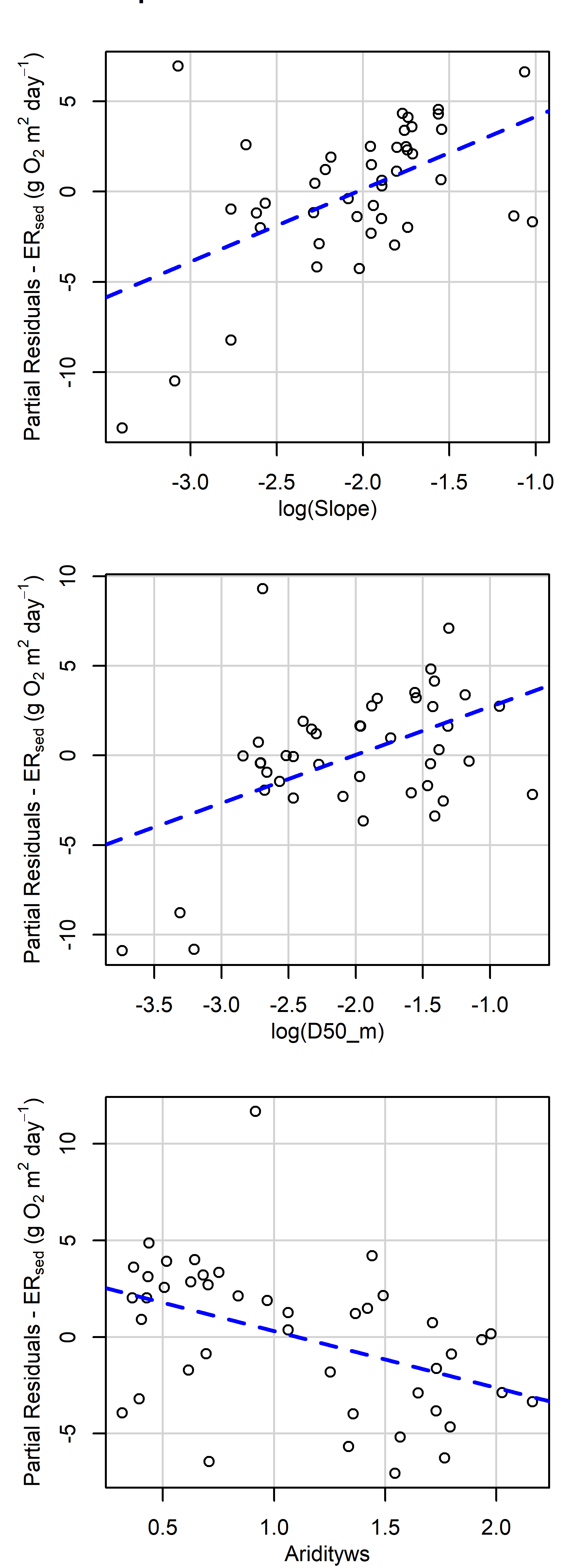
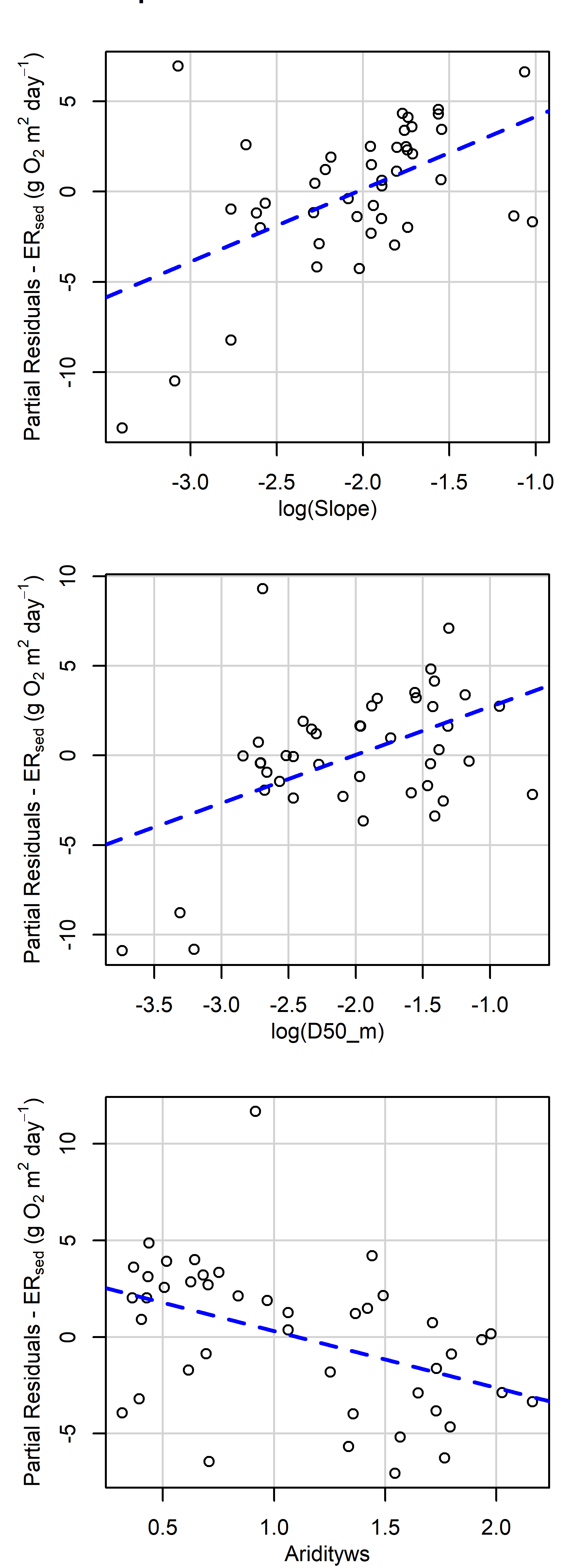
The R-squared value is 0.3766, meaning the fitted model can explain about 37.66% of the total variability. From the p values we can see that only the coefficients for ‘Slope’ is significant.

**Partial residual plots (generated using function ‘*crPlots()’* from *‘car’* package in R).**

Partial residual plots attempt to show the relationship between a given predictor , and the response variable (ERsed\_gm2day) given that other predictor variables are also in the model.

The steps to create partial residual plots:

1. Regress the response ‘ERsed\_gm2day’ on all predictors. Store the residuals of this model,
2. Calculate the partial residuals:
3. Plot  vs  (i.e., totdasqkm, Minidot\_Temperature, velocity\_ms, and AridityWs), adding a regression line.

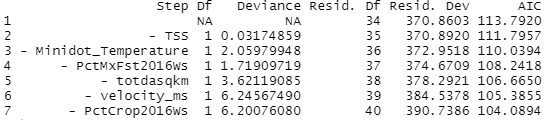
 

* **Log transform variables : "totdasqkm"( total drainage area), 'TSS', 'slope' , ‘D50\_m’ and PctCrop2016Ws (%crop)**

1. **Forward stepwise-selected variables (variables to keep in model):**

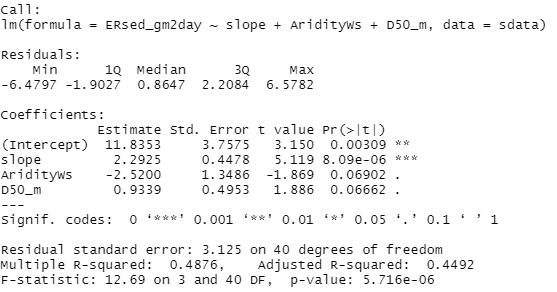


2) **backward stepwise-selected variables (variables to remove from model):**



Forward selection only returns one variable -slope, while backward selection returns 3 variables (i.e., slope, AridityWs, D50\_m)

Summary statistics for the lm model fitted using the selected variables(i.e., slope, AridityWs, D50\_m)



The R-squared value is 0.4876, meaning the fitted model can explain about 48.76% of the total variability. From the p values we can see that only the coefficient for slope is significant, the coefficient for D50\_m and Aridity is not significant .

**Partial residual plots (generated using function ‘*crPlots()’* from *‘car’* package in R).**

Partial residual plots attempt to show the relationship between a given predictor , and the response variable (ERsed\_gm2day) given that other predictor variables are also in the model.

The steps to create partial residual plots:

1. Regress the response ‘ERsed\_gm2day’ on all predictors. Store the residuals of this model,
2. Calculate the partial residuals:
3. Plot  vs  (i.e., totdasqkm, Minidot\_Temperature, velocity\_ms, and AridityWs), adding a regression line.

