Estuarine Assimilation: Regression Analyses

Sean Kinard

2024-06-11

Contents

1 Report Summary

1.0.1 Background:

The research project aims to quantify the contributions of estuarine-derived nutrients to coastal streams across a natural precipitation gradient in South-Central Texas. Estuarine assimilation in both inconspicuous migrants and freshwater taxa is investigated, exploring their ecological dynamics in relation to climate, geography, and anthropogenic factors. Through multiple regression analysis, the aim is to elucidate the intricate relationships between environmental variables such as annual rainfall, estuary distance, and euryhaline biomass, shedding light on the mechanisms driving estuarine assimilation within freshwater ecosystems.

1.0.2 Approach

These analyses begin with an overview of linear relationships between environmental predictors as well as their individual effects on estuarine assimilation (EA) at different scales of comparison within site fish and invertebrate communities. Influential species with high EA values are then identified using sorting within tables, followed by an exploration of the option of correcting for non-proportional stable isotope sampling by simulating communities using quantitative biomass survey data.

In the second half of this report, multiple regression modeling is employed to examine estuarine assimilation (EA) within freshwater taxa across diverse habitats. Variables such as annual rainfall, estuary distance, and the proportion of euryhaline biomass within the community are utilized as predictors. The predictive models are evaluated using Root Mean Square Error (RMSE), and comparisons are made between observed data and model predictions. Additionally, visualizations are created to present site community EA plotted against annual rainfall, facilitating an assessment of the effectiveness of different modeling approaches. Through these methods, the aim is to uncover the intricate dynamics of EA and its relationship with environmental variables and habitat connectivity.

1.0.3 Key Findings:

1.0.3.1 Environmental Influences There is a consistent decline in EA with increasing annual rainfall and estuary distance, suggesting that regions with higher rainfall and greater distance from estuaries tend to have lower EA levels among freshwater species. This is likely due to dilution effects from rainfall runoff and limited accessibility of estuarine resources in distant areas.

1.0.3.2 Role of Euryhaline Biomass A positive correlation is observed between euryhaline biomass and EA, indicating that higher abundance of euryhaline taxa may enhance estuarine resource utilization by freshwater species. Further investigation is needed to understand the underlying mechanisms driving this relationship.

1.0.3.3 Impact of Habitat Connectivity Sites closer to estuaries exhibit higher EA levels, highlighting the importance of maintaining or enhancing connectivity between freshwater and estuarine habitats to support ecosystem functioning and biodiversity preservation.

1.0.3.4 Implications for Dam Effects Dams located near estuaries in Semi-Arid climates are likely to cause significant disruptions to estuarine connectivity, leading to pronounced declines in EA among freshwater taxa. Integrating these findings into conservation and management strategies can inform decisions aimed at preserving and restoring estuarine connectivity in coastal streams.

1.0.4 Relevance:

These findings and predictive models offer valuable insights into the dynamics of estuarine assimilation (EA) within freshwater taxa, with implications for ecosystem management and conservation. By identifying

key environmental factors influencing EA levels, such as annual rainfall, estuary distance, and euryhaline biomass, these models provide a basis for proactive conservation strategies. They can inform decisions regarding habitat restoration and management, helping to preserve estuarine connectivity and promote ecosystem resilience in the face of anthropogenic disturbances. Additionally, integrating EA predictions into management frameworks can facilitate more sustainable resource use and biodiversity conservation in coastal regions. Overall, this research contributes to a better understanding of ecosystem dynamics and has practical applications for ecological management worldwide.

2 Estuarine Assimilation Versus Site Characteristics (Annual Rainfall, Bay Distance, and Elevation)

2.1 Setup

This code chunk contains a set of functions designed to analyze the relationship between Estuarine Assimilation (EA) and site variables (annual rainfall, distance to bay, and elevation) across various taxonomic groups. The functions facilitate data preparation, linear regression analysis, and visualization of the EA versus Rainfall relationship. Specifically, the functions compute linear regression statistics, generate base plots, and perform analysis for each taxonomic group of interest. Overall, this code chunk streamlines the process of exploring and understanding the impact of rainfall and geographic features on estuarine assimilation within different ecological communities.

```r
#------------------------------------------------------------------------------
# Setup: EA Versus Rainfall (Within Taxonomic Groups)
#------------------------------------------------------------------------------
# table function
table_lm_stats <- function(my_data, predictor, response) {
  d_temp <- my_data
  colnames(d_temp) <- str_replace_all(colnames(d_temp), predictor, 'predictor')
  colnames(d_temp) <- str_replace_all(colnames(d_temp), response, 'response')

  lm(formula = response ~ predictor,
     data = d_temp) %>%
    summary()  %>%
    broom::tidy() %>%
    filter(term == 'predictor') %>%
    mutate(predictor = str_replace_all(term, 'predictor', predictor),
           response = response) %>%
    select(-term) %>%
    mutate(
      signif = case_when(
        p.value >= 0.01 & p.value < 0.05 ~ "*",
        p.value < 0.01 ~ "**",
        T ~ ""  )) %>%
    select(contains('response'), contains('predictor'), everything())
}

# base plot: EA versus Rainfall
plot_ea_lm_explore <- function(xdata, xpredictor, xresponse) {
  p_temp <- xdata
  colnames(p_temp) <- str_replace_all(colnames(p_temp), xpredictor, 'PRE')
  colnames(p_temp) <- str_replace_all(colnames(p_temp), xresponse, 'RES')

  p_temp %>%
    ggplot(aes(x=PRE, y=RES)) +
    facet_wrap(~XX) +
    stat_poly_eq(label.x=.5, label.y=.95, formula=y~x,
                 color='black', use_label(c("adj.R2","p")), size=4) +
    geom_point(size=3, color='blue', fill='skyblue', shape=21, alpha=.5) +
    geom_point(size=3, color='blue', fill=NA, shape=21) +
    labs(x=str_to_title(xpredictor),
         y=str_to_title(xpredictor)) +
    theme_bw(base_size=12) +
```

```r
    geom_smooth(data = . %>% filter(p.value < 0.1 & p.value >= 0.05),
                method = "lm", se = FALSE,
                color = "blue", lwd=.5, lty=2)  +
    geom_smooth(data = . %>% filter(p.value < 0.05),
                method = "lm", se = FALSE,
                color = "blue", lwd=.5, lty=1)
}

# Generate table and plot for x_group of comparison
ea_lm_explore <-function(x_data, x_group, n_sites=1) {
  temp_data <- x_data %>% add_rain() %>% add_sitevars()
  colnames(temp_data) <- str_replace_all(colnames(temp_data), x_group, 'XX')

  # list widespread groups
  widespread <- temp_data %>%
    group_by(XX) %>%
    summarize(n_samples = length(EA_XX_mu)) %>%
    filter(n_samples>n_sites) %>%
    pull(XX)

  # prepare predictor response cross
  my_predictor <- c('annualrain','baydist_km','elev_site_m')
  my_response <- c('EA_XX_mu')
  lm_cross <- crossing(my_predictor, my_response)

  # empty regression table
  t_unfit <- temp_data %>%
    filter(XX %in% widespread) %>%
    group_by(XX) %>%
    nest() %>%
    mutate(lm_cross = list(lm_cross)) %>%
    unnest(lm_cross)

  # Table: Regression Statistics
  t_fit <- t_unfit %>%
    mutate(lm = pmap(list(data, my_predictor, my_response), table_lm_stats)) %>%
    unnest(lm) %>%
    ungroup() %>%
    select(-contains('my'))

  colnames(t_fit) <- str_replace_all(colnames(t_fit), x_group, 'XX')

  # visualize
  p_EA_lm_explore <- t_fit %>%
    unnest(data) %>%
    group_by(response, predictor) %>%
    nest() %>%
    mutate(p_ea_lm = pmap(list(data, predictor, response), plot_ea_lm_explore))

  colnames(t_fit) <- str_replace_all(colnames(t_fit), 'XX', x_group)

  output <- list(figure = p_EA_lm_explore,
                 table = t_fit)
```

```r
    return(output)
}

format_table <- function(x) {
  x %>%
      select(-c(data, response)) %>%
  ungroup() %>%
  gt(groupname_col = 'predictor') %>%
  fmt_number(columns = where(is.numeric), decimals = 3)
}
```

2.2 Site Characteristic Regression Pairs

2.2.1 Setup

```r
my_vars <- c('annualrain', 'baydist_km', 'elev_site_m')

d_site_vars <- tibble(site_code = my_sites) %>%
  add_rain() %>%
  add_sitevars() %>%
  filter(site_type != 'Estuary') %>%
  select(any_of(my_vars))

library(GGally)
p_site_vars <- d_site_vars %>%
  rename('Annual Rain (cm)' = annualrain,
         'Bay Distance (km)' = baydist_km,
         'Elevation (m)' = elev_site_m) %>%
  ggpairs() +
  theme_bw(base_size=12)

t_site_vars <- crossing(y = my_vars, x = my_vars) %>%
  filter(y!=x) %>%
  mutate(data = list(d_site_vars)) %>%
  mutate(lm = pmap(list(data, x, y), table_lm_stats)) %>%
  unnest(lm) %>%
  select(-c(x, y, data))
```

2.2.2 Figure: Site Characteristic Regression Pairs

The lines in the plot represent fitted regression lines illustrating the relationship between pairs of variables. Correlation coefficients (Corr:) are displayed to denote the strength and direction of the linear relationship between variables, with '*' symbols indicating significance levels.
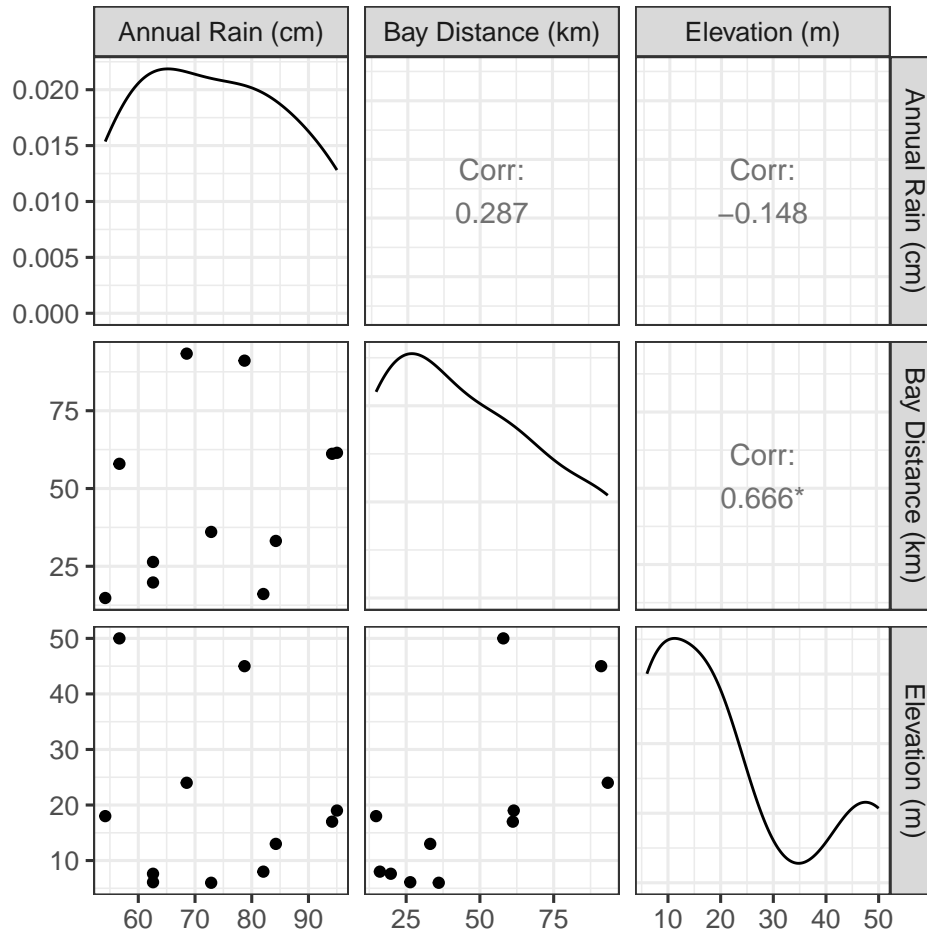
2.2.3 Table: Site Characteristic Regression Pairs

Linear regression statistics for pairings between site characteristic variables (rainfall, distance, elevation).

response	predictor	estimate	std.error	statistic	p.value	signif
annualrain	baydist_km	0.143	0.160	0.897	0.393	
annualrain	elev_site_m	−0.140	0.310	−0.450	0.663	
baydist_km	annualrain	0.573	0.638	0.897	0.393	
baydist_km	elev_site_m	1.252	0.468	2.679	0.025	*
elev_site_m	annualrain	−0.158	0.350	−0.450	0.663	
elev_site_m	baydist_km	0.354	0.132	2.679	0.025	*

2.2.4 Conclusions: Site Characteristics Regression Pairs

In the pairs plots of environmental predictors, no significant relationship is observed between rainfall and other site characteristics, namely distance-to-bay and elevation. As anticipated, a positive correlation is evident between elevation and distance-to-bay.

8

2.3 Widespread Taxonomic Order

2.3.1 Setup

```
#-------------------------------------------------------------------------------
# EA Versus Site Characteristics: Within Widespread Orders
#-------------------------------------------------------------------------------
widespread_orders <- ea_lm_explore(x_data = iso_order,
                                   x_group= 'order',
                                   n_sites = 8)
```

2.3.2 Figure: EA Versus Rainfall Within Widespread Taxonomic Orders

Linear regression of average estuarine assimilation (EA) within widespread taxonomic orders versus annual rainfall. EA is estimated for each order using Bayesian mixing models with d13C and d34S stable isotope data in pre-requisite analyses.

```
p_EA_v_rain_order <- widespread_orders$figure %>% filter(predictor == 'annualrain')
p_EA_v_rain_order$p_ea_lm[[1]] +
  labs(y = 'EA (%)',
       x = 'Annual Rain (cm)')
```
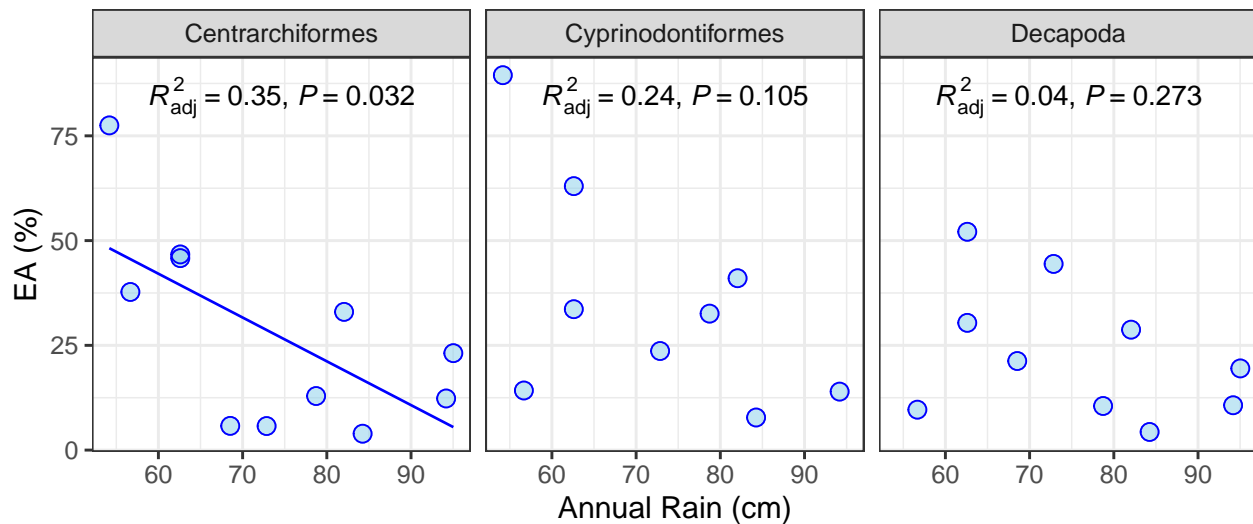

2.3.2.1 Interpretation: EA Versus Rainfall Within Widespread Taxonomic Orders

The analysis indicates that Centrarchiformes, such as sunfish and bass, experience reduced estuarine assimilation (EA) with increasing annual rainfall, suggesting that higher rainfall negatively impacts their assimilation of estuarine nutrients or resources. This relationship points to potential environmental or ecological factors influenced by rainfall that uniquely affect Centrarchiformes.

Conversely, for other widespread taxonomic orders like Cyprinodontiformes (including killifish, livebearers, pupfish, and topminnows) and Decapoda (including shrimp, crayfish, and crabs), there is no statistically significant linear relationship with annual rainfall. This implies that variations in annual rainfall do not systematically affect the EA of these taxa, and their estuarine assimilation is likely influenced by other factors rather than rainfall.

2.3.3 Figure: EA Versus Distance-To-Estuary Within Widespread Taxonomic Orders

Linear regression of average estuarine assimilation (EA) within widespread taxonomic orders versus the distance from the site to its downstream estuary. EA is estimated for each order using Bayesian mixing models with d13C and d34S stable isotope data in pre-requisite analyses.

```
p_EA_v_bdis_order <- widespread_orders$figure %>% filter(predictor == 'baydist_km')
p_EA_v_bdis_order$p_ea_lm[[1]] +
  labs(y = 'EA (%)',
       x = 'Distance To Estuary (km)')
```
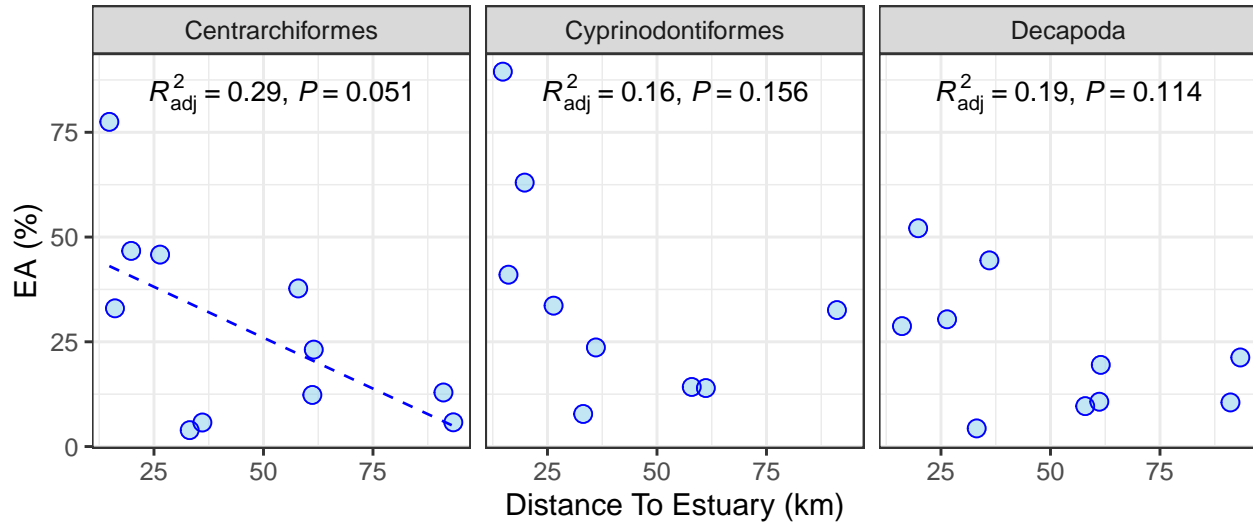

2.3.3.1 Interpretation: EA Versus Distance-To-Estuary Within Widespread Taxonomic Orders
The analysis reveals that Centrarchiformes show a negative correlation between estuarine assimilation (EA) and proximity to the bay, similar to the pattern observed with annual rainfall, although this relationship is marginally non-significant (p = 0.051). This suggests that as the distance to the bay increases, EA for Centrarchiformes tends to decrease, mirroring the effect of decreasing annual rainfall on EA.

In contrast, other common taxonomic orders such as Cyprinodontiformes (including killifish, livebearers, pupfish, and topminnows) and Decapoda (including shrimp, crayfish, and crabs) do not show any statistically significant linear relationships with annual rainfall. This indicates that, for these taxa, annual rainfall does not have a consistent or predictable impact on their EA, and their assimilation patterns are not influenced by variations in rainfall.

2.3.4 Figure: EA Versus Elevation Within Widespread Taxonomic Orders

Linear regression of average estuarine assimilation (EA) within widespread taxonomic orders versus site elevation above sea level. EA is estimated for each order using Bayesian mixing models with d13C and d34S stable isotope data in pre-requisite analyses.

```r
p_EA_v_elev_order <- widespread_orders$figure %>% filter(predictor == 'elev_site_m')
p_EA_v_elev_order$p_ea_lm[[1]] +
  labs(y = 'EA (%)',
       x = 'Elevation (m)')
```
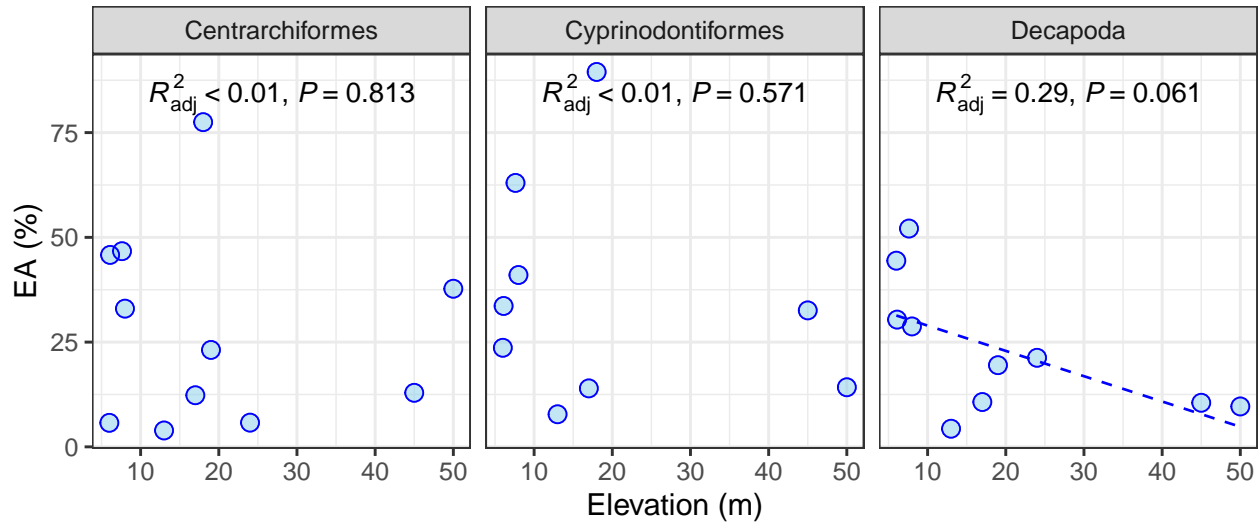

2.3.4.1 Interpretation: EA Versus Elevation Within Widespread Taxonomic Orders Decapoda exhibit a tendency where higher elevations correspond to lower levels of estuarine assimilation (EA). This suggests that as elevation increases, there is a decrease in the amount of estuarine assimilation among Decapoda species. On the other hand, both Centrarchiformes and Cyprinidontiformes demonstrate random patterns in their relationship between EA and elevation. This indicates that for these taxa, there is no discernible trend or consistent association between EA and elevation across different elevation levels.

2.3.5 Table: Regression Statistics for EA Vs. Site Characteristics (Within Widespread Orders)

linear regression statistics for estuarine assimilation versus annual rainfall (cm/yr), estuary distance (km), or Elevation (m) within widespread taxonomic orders.

```
widespread_orders$table %>% format_table()
```

order	estimate	std.error	statistic	p.value	signif
annualrain					
Centrarchiformes	−1.045	0.412	−2.539	0.032	*
Decapoda	−0.455	0.386	−1.177	0.273	
Cyprinodontiformes	−1.099	0.590	−1.862	0.105	
baydist_km					
Centrarchiformes	−0.486	0.216	−2.250	0.051	
Decapoda	−0.301	0.170	−1.774	0.114	
Cyprinodontiformes	−0.529	0.333	−1.588	0.156	
elev_site_m					
Centrarchiformes	−0.123	0.506	−0.244	0.813	
Decapoda	−0.604	0.278	−2.174	0.061	
Cyprinodontiformes	−0.343	0.577	−0.594	0.571	

2.3.6 Conclusions: EA vs Site Characteristics within Taxonomic Orders

The analysis reveals distinct patterns of estuarine assimilation (EA) across various site characteristics within different taxonomic orders. Centrarchiformes (sunfish and bass) demonstrate a negative relationship between EA and both annual rainfall and proximity to the bay, suggesting that these environmental factors negatively impact their nutrient assimilation. Conversely, Cyprinodontiformes (killifish, livebearers, pupfish, and topminnows) and Decapoda (shrimp, crayfish, and crabs) do not exhibit significant linear relationships between EA and either rainfall or distance to the bay, implying that their EA is influenced by other factors. Additionally, Decapoda show a negative correlation between EA and elevation, indicating decreased EA at higher elevations, while Centrarchiformes and Cyprinodontiformes display no consistent patterns with elevation, further highlighting the varied ecological responses among these taxonomic groups.
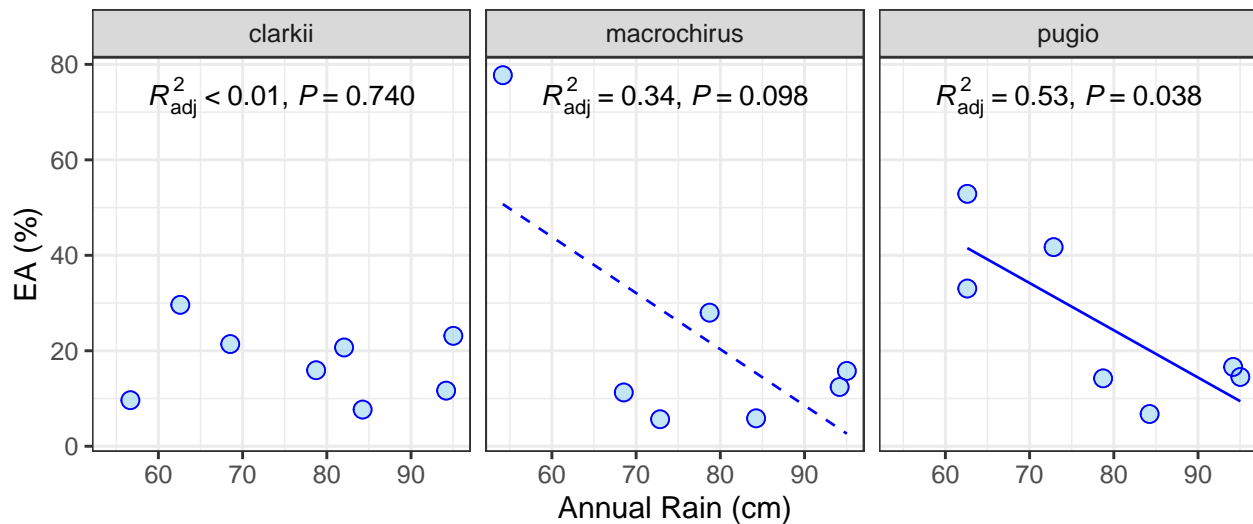
2.4 Widespread Taxonomic Species

2.4.1 Setup

```
#-------------------------------------------------------------------------------
# EA Versus Site Characteristics: Within Widespread species
#-------------------------------------------------------------------------------
widespread_species <- ea_lm_explore(x_data = iso_species,
                                    x_group= 'species',
                                    n_sites = 6)
```

2.4.2 Figure: EA Versus Rainfall Within Widespread Taxonomic Species

Linear regression of average estuarine assimilation (EA) within widespread taxonomic species versus annual rainfall. EA is estimated for each species using Bayesian mixing models with d13C and d34S stable isotope data in pre-requisite analyses.

```
p_EA_v_rain_species <- widespread_species$figure %>% filter(predictor == 'annualrain')
p_EA_v_rain_species$p_ea_lm[[1]] +
  labs(y = 'EA (%)',
       x = 'Annual Rain (cm)')
```


2.4.2.1 Interpretation: EA Versus Rainfall Within Widespread Taxonomic Species   Examining estuarine assimilation (EA) versus annual rainfall within widespread taxonomic species (species collected from at least 6 out of 9 streams), which include *Procambarus clarkii*, *Lepomis macrochirus*, and *Palaemonetes pugio*, reveals that only *Palaemonetes* exhibits a statistically significant linear relationship with annual rainfall. Specifically, *Palaemonetes* EA is negatively related to annual rainfall, suggesting that higher rainfall reduces their estuarine assimilation. It appears that missing data for *Lepomis macrochirus* may be affecting its statistical significance; filling in these gaps could potentially enhance the observed relationship, aligning with the negative EA-rainfall trend noted in its taxonomic order (Centrarchiformes). This indicates that accurate and complete data are crucial for understanding the environmental factors influencing EA in these species.

2.4.3 Figure: EA Versus Distance-To-Estuary Within Widespread Taxonomic Species

Linear regression of average estuarine assimilation (EA) within widespread taxonomic species versus the distance from the site to its downstream estuary. EA is estimated for each species using Bayesian mixing models with d13C and d34S stable isotope data in pre-requisite analyses.

```
p_EA_v_bdis_species <- widespread_species$figure %>% filter(predictor == 'baydist_km')
p_EA_v_bdis_species$p_ea_lm[[1]] +
  labs(y = 'EA (%)',
       x = 'Distance To Estuary (km)')
```
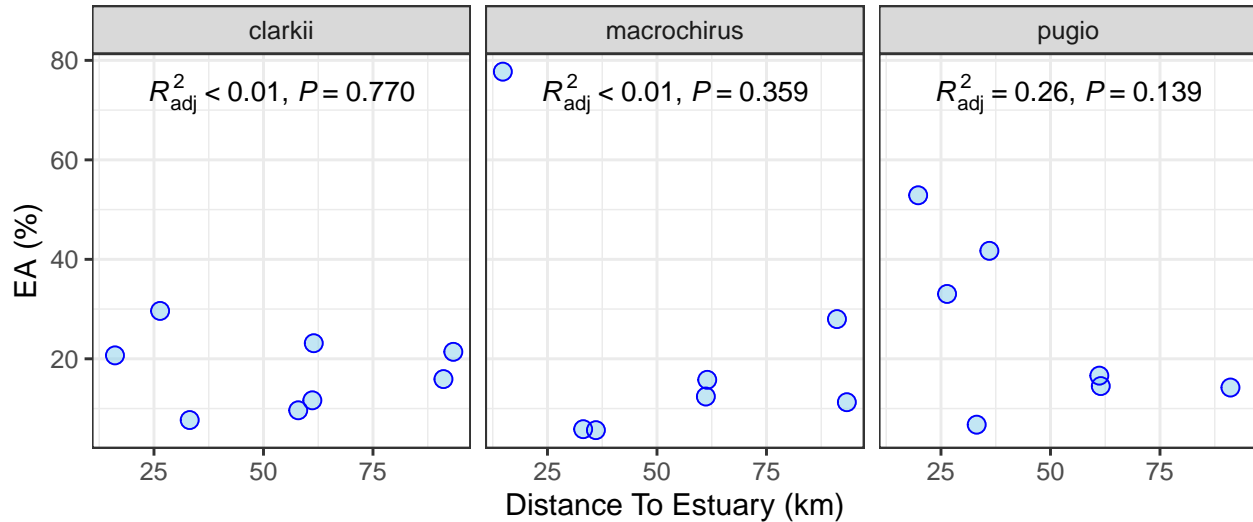

2.4.3.1 Interpretation: EA Versus Distance-To-Estuary Within Widespread Taxonomic Species No
statistically significant relationships were detected within widespread species between estuarine assimilation and distance to estuary.

2.4.4 Figure: EA Versus Elevation Within Widespread Taxonomic Species

Linear regression of average estuarine assimilation (EA) within widespread taxonomic species versus site elevation above sea level. EA is estimated for each species using Bayesian mixing models with d13C and d34S stable isotope data in pre-requisite analyses.

```
p_EA_v_elev_species <- widespread_species$figure %>% filter(predictor == 'elev_site_m')
p_EA_v_elev_species$p_ea_lm[[1]] +
  labs(y = 'EA (%)',
       x = 'Elevation (m)')
```
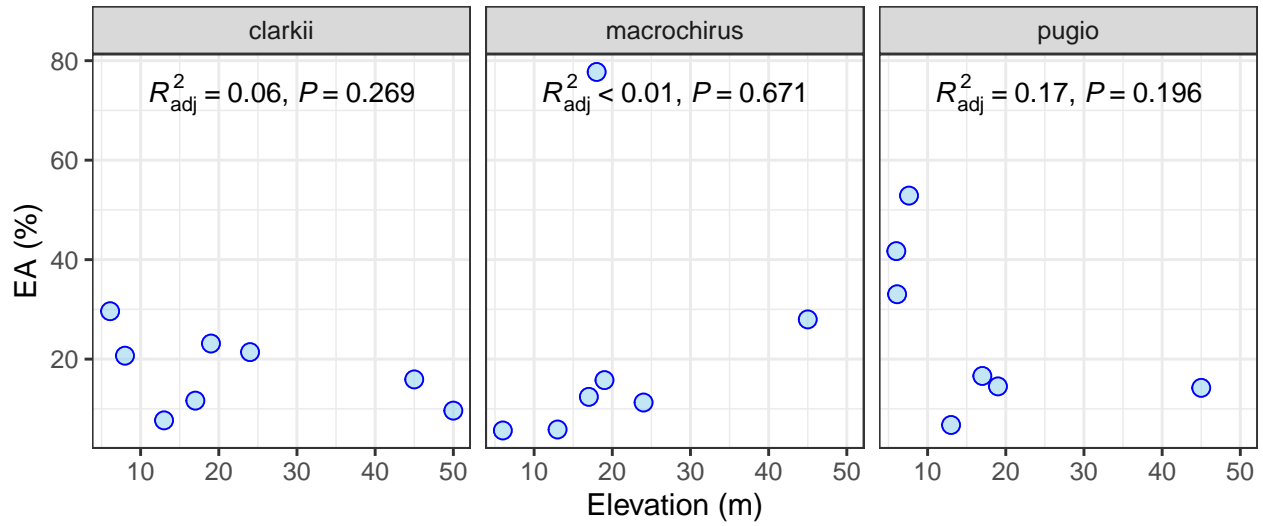

2.4.4.1 Interpretation: EA Versus Elevation Within Widespread Taxonomic Species
No statistically significant relationships were detected within widespread species between estuarine assimilation and elevation.

2.4.5 Table: Regression Statistics for EA Vs. Site Characteristics (Within Widespread Species)

linear regression statistics for estuarine assimilation versus annual rainfall (cm/yr), estuary distance (km), or Elevation (m) within widespread taxonomic species.

```
widespread_species$table %>% format_table()
```

species	estimate	std.error	statistic	p.value	signif
annualrain					
clarkii	−0.075	0.217	−0.347	0.740	
macrochirus	−1.177	0.581	−2.028	0.098	
pugio	−0.990	0.353	−2.801	0.038	*
baydist_km					
clarkii	−0.033	0.107	−0.306	0.770	
macrochirus	−0.354	0.350	−1.010	0.359	
pugio	−0.418	0.237	−1.760	0.139	
elev_site_m					
clarkii	−0.205	0.169	−1.217	0.269	
macrochirus	0.413	0.915	0.451	0.671	
pugio	−0.691	0.463	−1.492	0.196	

2.4.6 Conclusions: EA vs Site Characteristics within Taxonomic Species

Examining estuarine assimilation (EA) versus annual rainfall within widespread species (*Procambarus clarkii*, *Lepomis macrochirus*, and *Palaemonetes pugio*) reveals that only *Palaemonetes* shows a statistically significant negative relationship, suggesting higher rainfall reduces their EA. Missing data for *Lepomis macrochirus* might affect its statistical significance; filling these gaps could enhance the observed relationship, aligning with the negative EA-rainfall trend in Centrarchiformes. Accurate data are crucial for understanding EA in these species. No significant relationships were found between EA and distance to the estuary or elevation within these widespread species, indicating that these factors do not systematically influence their estuarine assimilation. This suggests that other environmental or biological factors may be more critical in determining EA for these taxa.
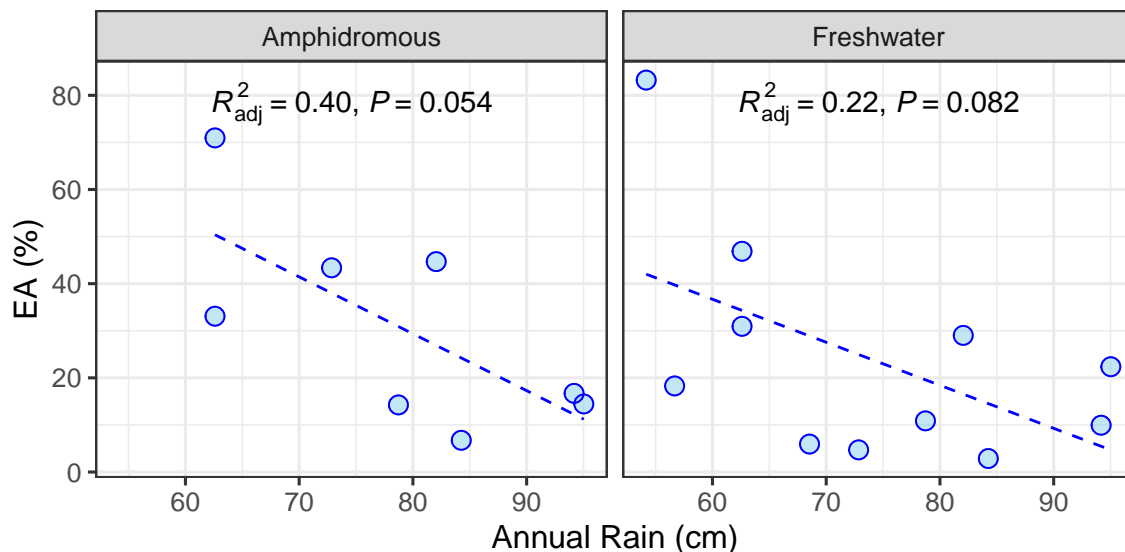
2.5 Transient Type

2.5.1 Setup

```
#-------------------------------------------------------------------------------
# EA Versus Site Characteristics: Within Widespread transient
#-------------------------------------------------------------------------------
widespread_transient <- ea_lm_explore(x_data = iso_transient,
                                      x_group= 'transient',
                                      n_sites = 5)
```

2.5.2 Figure: EA Versus Rainfall Within Widespread Transient Types

Linear regression of average estuarine assimilation (EA) within Transient Type versus annual rainfall. EA is estimated for each transient using Bayesian mixing models with d13C and d34S stable isotope data in pre-requisite analyses.

```
p_EA_v_rain_transient <- widespread_transient$figure %>% filter(predictor == 'annualrain')
p_EA_v_rain_transient$p_ea_lm[[1]] +
  labs(y = 'EA (%)',
       x = 'Annual Rain (cm)')
```


2.5.2.1 Interpretation: EA Versus Rainfall Within Widespread Transient Types
Among the transient types, isotope samples from freshwater and amphidromous species were widely distributed across the region: freshwater (n=9) and amphidromous (n=8), along with euryhaline (n=4) and catadromous (n=2). Linear regression analysis, with a p-value threshold of 0.1, demonstrates a significant negative correlation between estuarine assimilation (EA) and annual rainfall for both amphidromous and freshwater species. This indicates that higher rainfall levels are linked to reduced EA in these types, emphasizing the impact of precipitation patterns on estuarine dynamics. These findings also imply potential ecological consequences for these species in response to alterations in rainfall regimes.

2.5.3 Figure: EA Versus Distance-To-Estuary Within Widespread Transient Types

Linear regression of average estuarine assimilation (EA) within Transient Type versus the distance from the site to its downstream estuary. EA is estimated for each transient using Bayesian mixing models with d13C and d34S stable isotope data in pre-requisite analyses.

```
p_EA_v_bdis_transient <- widespread_transient$figure %>% filter(predictor == 'baydist_km')
p_EA_v_bdis_transient$p_ea_lm[[1]] +
  labs(y = 'EA (%)',
       x = 'Distance To Estuary (km)')
```
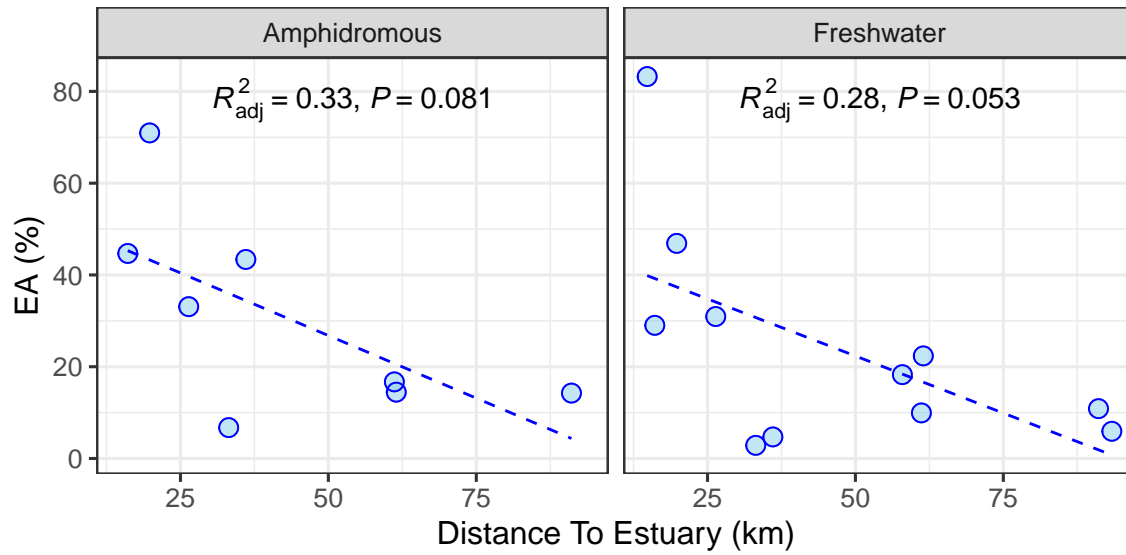

2.5.3.1 Interpretation: EA Versus Distance-To-Estuary Within Widespread Transient Types

Patterns of estuarine assimilation (EA) in relation to the distance to the nearest estuary mirror those observed with annual rainfall. Linear regression analysis, with a p-value threshold of 0.1, reveals that both amphidromous and freshwater groups display a significant negative correlation between EA and the distance to the nearest estuary. This suggests that as the distance to the estuary increases, EA tends to decrease for these species, potentially indicating a reduced availability of estuarine resources or increased energetic costs associated with accessing estuarine habitats. Understanding these relationships is crucial for assessing the ecological dynamics of these species in coastal environments.

2.5.4 Figure: EA Versus Elevation Within Transient Type

Linear regression of average estuarine assimilation (EA) within Transient Type versus site elevation above sea level. EA is estimated for each transient using Bayesian mixing models with d13C and d34S stable isotope data in pre-requisite analyses.

```
p_EA_v_elev_transient <- widespread_transient$figure %>% filter(predictor == 'elev_site_m')
p_EA_v_elev_transient$p_ea_lm[[1]] +
  labs(y = 'EA (%)',
       x = 'Elevation (m)')
```
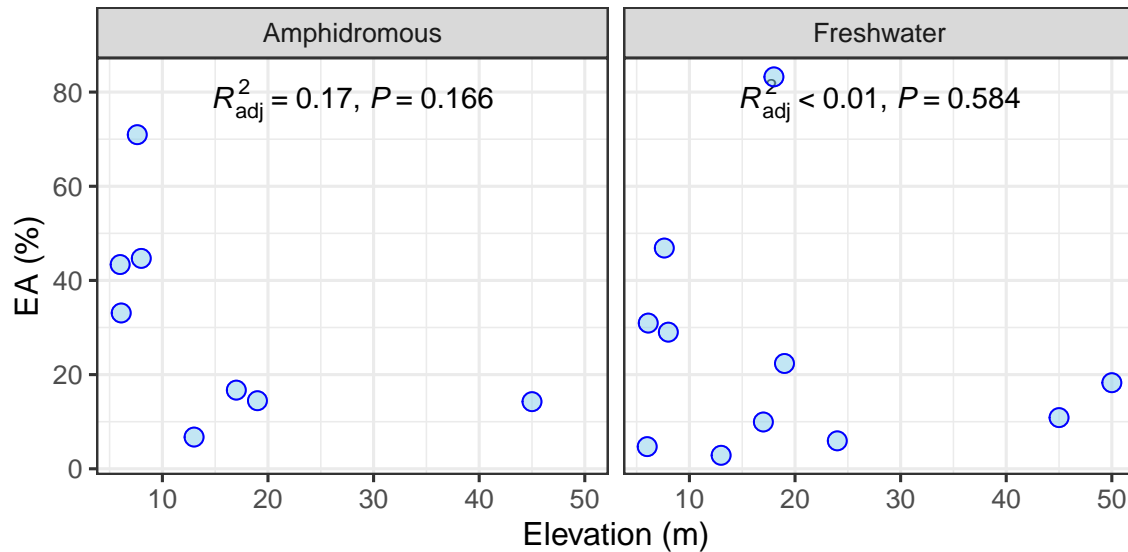

2.5.4.1 Interpretation: EA Versus Elevation Within Widespread Transient Types There are no statistically significant linear relationships between EA and elevation within freshwater or amphidromous transient groups.

2.5.5 Table: Regression Statistics for EA Vs. Site Characteristics (Within Widespread Transient Types)

linear regression statistics for estuarine assimilation versus annual rainfall (cm/yr), estuary distance (km), or Elevation (m) within widespread transient types.

```
widespread_transient$table %>% format_table()
```

transient	estimate	std.error	statistic	p.value	signif
annualrain					
Freshwater	-0.914	0.466	-1.959	0.082	
Amphidromous	-1.208	0.505	-2.393	0.054	
baydist_km					
Freshwater	-0.498	0.224	-2.221	0.053	
Amphidromous	-0.546	0.260	-2.097	0.081	
elev_site_m					
Freshwater	-0.292	0.515	-0.567	0.584	
Amphidromous	-0.900	0.571	-1.575	0.166	

2.5.6 Conclusions: EA vs Site Characteristics within Taxonomic transient

Our findings reveal significant negative correlations between EA and annual rainfall for both amphidromous and freshwater species, indicating that higher rainfall levels are associated with reduced EA. The negative correlation between EA and distance to the nearest estuary suggests that accessibility to estuarine habitats influences EA, with greater distances potentially leading to decreased EA due to reduced resource availability or increased energetic costs. However, the absence of statistically significant linear relationships between EA and elevation within freshwater or amphidromous transient groups suggests that elevation may not significantly influence EA in these species. Overall, understanding these complex relationships is essential for informing conservation and management strategies aimed at protecting estuarine ecosystems and the species that rely on them.

2.6 Section Summary: EA vs Site Predictors (Within Order, Species, Transient Groups)

The examination of estuarine assimilation (EA) across various taxonomic orders unveils nuanced responses shaped by site-specific characteristics. Within Centrarchiformes, a negative correlation emerges between EA and both annual rainfall and proximity to the bay, indicating adverse impacts on nutrient assimilation likely due to heightened runoff and reduced salinity near the bay mouth. Conversely, Cyprinodontiformes and Decapoda do not manifest significant linear relationships with rainfall or bay proximity, suggesting multifaceted influences on their EA, possibly related to species-specific habitat preferences or physiological adaptations.

Furthermore, Decapoda demonstrate a notable negative correlation between EA and elevation, implying diminished EA at higher elevations, potentially linked to reduced access to nutrient-rich estuarine waters. In contrast, Centrarchiformes and Cyprinodontiformes exhibit less clear trends with elevation, reflecting the complexity of ecological responses within these taxa.

When exploring EA versus annual rainfall within widespread species like *Procambarus clarkii* (red swamp crayfish), *Lepomis macrochirus* (bluegill sunfish), and *Palaemonetes pugio* (daggerblade grass shrimp), only *P. pugio* exhibits a statistically significant negative relationship with rainfall, suggesting a dampening effect on EA likely attributable to increased freshwater input. The lack of significant relationships between EA and distance to the estuary or elevation underscores the intricate interplay of multiple environmental factors shaping EA dynamics in these species.

In transient taxonomic groups, such as amphidromous and freshwater species, significant negative correlations between EA and annual rainfall highlight the vulnerability of these species to changes in precipitation patterns. The observed negative correlation between EA and distance to the nearest estuary underscores the importance of habitat accessibility in influencing EA dynamics, with increased distances potentially imposing energetic costs or limiting access to crucial resources. However, the absence of statistically significant linear relationships between EA and elevation within these transient groups suggests a more nuanced relationship with terrain features, warranting further investigation.

3 High EA / Influential EA species

3.1 Setup

The goal of this section is to present and interpret mean estuarine assimilation (EA) estimates for various species across different stream sites, highlighting their reliance on estuarine habitats. By arranging species in descending order of EA values and organizing site columns by annual precipitation, the analysis aims to elucidate patterns of habitat utilization and ecological adaptation among euryhaline, amphidromous, and freshwater species. This approach combines quantitative data with ecological interpretation to explore the crucial role of estuarine ecosystems in supporting diverse taxa in arid coastal streams.

```r
t_species <- iso_species %>%
  filter(site_code %in% my_streams) %>%
  mutate(genus_species = paste(genus, species, sep=' ')) %>%
  group_by(genus_species) %>%
  mutate(species_EA = mean(EA_species_mu)) %>%
  ungroup() %>%
  fix_site_order() %>%
  arrange(desc(species_EA)) %>%
  select(site_code, genus_species, transient_type, EA_species_mu) %>%
  pivot_wider(names_from = site_code, values_from = EA_species_mu) %>%
  rename(Species = genus_species, Transient = transient_type)
```

3.2 Table species estuarine assimilation (EA) mean estimates

This table presents the mean estuarine assimilation estimates for species at each stream site. Species are arranged in descending order of EA. Columns display 2-letter site abbreviations, organized from left to right in ascending order of annual precipitation.

```
t_species %>%
  gt() %>%
  fmt_number(decimals=1) %>%
  fmt_missing(columns = everything(), missing_text = "")
```

Species	Transient	TR	SF	PL	PD	MR	GC	WM	AR	EM
Cyprinodon variegatus	Euryhaline	86.1								
Fundulus grandis	Euryhaline	82.2								
Oreochromis aureus	Euryhaline	67.4								
Poecilia latipinna	Euryhaline	83.6	10.8	41.3						
Macrobrachium ohione	Amphidromous			44.7						
Trinectes maculatus	Amphidromous			35.1						
Fundulus notatus	Freshwater				32.7					
Gambusia affinis	Freshwater	80.6	21.1			23.8	7.8	13.8		
Herichthys cyanoguttatus	Freshwater	74.0	9.1		16.8	20.4			6.1	
Lepomis cyanellus	Freshwater		37.7		20.9			11.1	16.2	30.0
Lepomis macrochirus	Freshwater	77.7			28.0	5.7	5.8	12.4	11.3	15.8
Palaemonetes pugio	Amphidromous				14.2	41.7	6.8	16.6		14.5
Lepomis megalotis	Freshwater			33.0			8.1	18.8	2.9	27.6
Ictalurus punctatus	Freshwater				18.9				15.7	
Lepomis humilis	Freshwater						12.5	21.9		
Noturus gyrinus	Freshwater						8.6			24.5
Lepomis gulosus	Freshwater						6.7	21.7		20.8
Procambarus clarkii	Freshwater		9.6	20.7	15.9		7.7	11.7	21.4	23.1
Etheostoma gracile	Freshwater							12.9		
Lepomis auritus	Freshwater				14.2				6.9	

3.3 Conclusions: Species EA

Euryhaline species such as Cyprinodon variegatus, Fundulus grandis, and Poecilia latipinna are well adapted to arid environments, migrating between freshwater and estuarine systems. Stable isotope data show they assimilate most of their carbon and sulfur from estuarine habitats, despite being caught in freshwater. With EA values exceeding 80%, these species rely heavily on estuaries, demonstrating resilience to flash floods and salt-water intrusion during droughts. This highlights the critical ecological role of transitional zones in supporting euryhaline taxa in arid coastal streams.

Amphidromous species are absent in the most arid sites but appear in transitional and humid sites, indicating a minimum rainfall threshold for their presence. These species, such as Macrobrachium ohione, Trinectes maculatus, and Palaemonetes pugio, have lower EA values than euryhaline species, yet their EA values are higher in more arid regions compared to humid ones. This suggests that while amphidromous taxa require a certain level of annual rainfall, their dependency on estuaries is greater in arid environments.

Furthermore, the inclusion of freshwater species with high EA values, such as Oreochromis aureus (68%) and Fundulus notatus (33%), highlights the dynamic nature of estuarine ecosystems as transitional habitats supporting a mix of freshwater and marine taxa. Despite their primary association with freshwater environments, these species demonstrate a substantial utilization of estuarine habitats, indicating the ecological connectivity and importance of estuaries as ecotones supporting diverse assemblages of species.

Lastly, the variability in EA values among certain species, like gambusia affinis (8-80%), Herichthys cyanoguttatus (6-78%), Lepomis cyanellus (11-38%), Lepomis macrochirus (6-77%), and Lepomis megalotis (3-33%), suggests a range of ecological strategies and preferences within estuarine environments. This variability may reflect species-specific adaptations to local environmental conditions, resource availability, or niche partitioning dynamics, further emphasizing the complexity of estuarine ecosystems and the diverse interactions among species within them.

4 Biomass-Adjusted Community Estuarine Assimilation

4.1 Setup

In this section, I attempt to correct biases caused by uneven sampling during stable isotope collections, which do not accurately reflect the relative abundances of species in a community. Initially, I calculate the biomass percentage of fish and invertebrates for each monitoring survey event. Then, I combine this data with estimates of estuarine assimilation (EA). However, some species lack EA values due to insufficient sampling. To address this, I impute EA values (mean and standard deviation) from the species' transient type. Next, I simulate communities by drawing values from a normal distribution using EA mean and standard deviation. The number of samples drawn equals the integer of relative biomass multiplied by 200. So for each event, there are 200 draws, with draws per species determined by their biomass relative to the total biomass of fish and invertebrates. Finally, I use linear regression to analyze the relationship between EA within transient groups and annual rainfall, similar to the analysis using unaltered data.

```r
#-------------------------------------------------------------------------------
# Setup
#-------------------------------------------------------------------------------
dc <- read_csv(here('03_public', 'output', 'biomass_add_EA.csv'))

# TR missing EA data for invertebrates, calculate site_EA_mu from fish for imputation
dc <- dc %>%
  group_by(site_code) %>%
  mutate(EA_site_mu = mean(EA_guild_mu, na.rm=T),
         EA_site_sd = mean(EA_guild_sd, na.rm=T)) %>%
  group_by(site_code) %>% # regional average sd for imputing missing values (n=1)
  mutate(site_sd = mean(EA_guild_sd, na.rm=T)) %>%
  ungroup()

# prepare for simulations
d_empty <- dc %>%
  filter(collection_period %in% c('2019-Q4', '2020-Q1')) %>%
  mutate(mu = ifelse(is.na(EA_species_mu), EA_transient_mu, EA_species_mu),
         sd = ifelse(is.na(EA_species_sd), EA_transient_mu, EA_species_sd)) %>%
  select(-starts_with('EA')) %>%
  rename(EA_mu = mu, EA_sd = sd) %>%
  select(site_code, collection_period, lowest_taxon, EA_mu, EA_sd, biomass_mu,
         biomass_percent) %>%
  group_by(site_code, collection_period, lowest_taxon) %>%
  nest()

#-------------------------------------------------------------------------------
# Simulate Communities
#-------------------------------------------------------------------------------
sim_ea <- function(xdata) {
  # number of pulls = the integer of relative biomass (as proportion) multiplied by 200.
  # In other words n=grams of lowest taxon per 200 grams of biomass for that sample event.
  # In this way there are 200 total draws for each event,
  # and the draws per species are based on their biomass relative to the biomass
  # of fish and invertebrates for that event.
  r_pulls <- xdata$biomass_percent*200%>%ceiling()
  # take random samples from a normal distribution
  # mean and sd are taken from simmr aa estimates for species > guild
  EA_sim <- rnorm(n = r_pulls, mean = xdata$EA_mu, sd = xdata$EA_sd)
  output <- tibble(EA = EA_sim)
```

```r
    return(output) }

# # Commented to save time rendering report:
# d_fill <- d_empty %>%
#   mutate(sim_tib = map(data, sim_ea)) %>%
#   unnest(sim_tib) %>%
#   unnest(data)

d_fill <- read_csv(here::here('03_public', 'output',
                              '15_simulated_biomass_weighted_communities_with_EA.csv'))

# Add taxonomic and transient categories
d_fill <- left_join(
  d_fill,
  dc %>%
    select(lowest_taxon, order, family, genus, species, guild,
           transient, is_diadromous) %>%
    unique())

#-------------------------------------------------------------------------------
# Calculate Biomass-Adjusted Community Mean EA
#-------------------------------------------------------------------------------
d_site_fresh_EA_mu <- d_fill %>%
  filter(transient_type == 'Freshwater') %>%
  group_by(site_code) %>%
  summarize(site_fresh_EA_mu = mean(EA, na.rm=T),
            site_fresh_EA_sd = sd(EA, na.rm=T))

d_transient_EA_mu <- d_fill %>%
  group_by(site_code, transient_type) %>%
  summarize(transient_EA_mu = mean(EA, na.rm=T),
            transient_EA_sd = sd(EA, na.rm=T))

p_simulated <- d_site_fresh_EA_mu %>%
  add_rain() %>%
  add_sitevars() %>%
  ggplot(aes(x=annualrain, y=site_fresh_EA_mu, color = is_dam)) +
  stat_poly_eq(label.x=.5, label.y=.95, formula=y~x,
               color='black', use_label(c("adj.R2","p")), size=4) +
  geom_point(size=3, color='blue', fill='skyblue', shape=21, alpha=.5) +
  geom_point(size=3, color='blue', fill=NA, shape=21) +
  theme_bw(base_size=12) +
  # geom_smooth(method = "lm", se = FALSE,
  #             color = "blue", lwd=.5, lty=2) +
  labs(x = 'Annual Rain (cm)', y = 'Group Mean EA (%)')
```

4.2   Figure: Resident EA Versus Rainfall

Linear regression of average estuarine assimilation (EA) within widespread taxonomic species versus annual rainfall. EA is estimated for each order using bayesian mixing models with d13C and d34S stable isotope data in pre-requisite analyses.

`p_simulated`
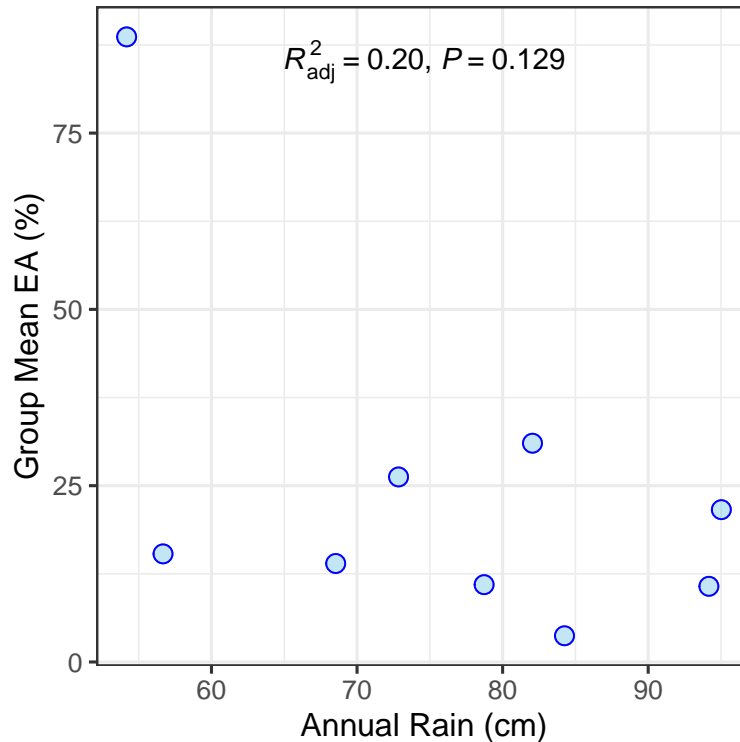
4.3   Conclusions: Biomass-Adjusted Community EA Estimates

Relationships between EA and annual rainfall are similar to those observed with the unaltered stable isotope data, though statistical significance slightly diminishes when using simulated biomass-weighted communities. This discrepancy may result from inadequate stable isotope sampling and the loss of nuanced EA profiles during imputation. For instance, some freshwater taxa show extraordinary EA compared to others within the same event. While imputation provides a useful approximation, it may not fully capture individual species' variability. Enhancing sample size and diversity, and refining imputation methods, could improve the accuracy of these analyses and clarify the relationships between EA and environmental variables like annual rainfall.

5 Multiple Regression: EA Versus Site Traits

5.1 Setup

In this section, I use multiple regression to develop a model that predicts the average estuarine assimilation (EA) within freshwater taxa at a site. The model considers site predictors such as annual rainfall, estuary distance, and elevation, as well as the proportion of euryhaline biomass within communities, expressed as a percentage. The goal is to identify the best predictors, understand their relationships, and evaluate the effects of dams, particularly the Calallen Dam. This dam is likely to block EA transfer upstream due to its proximity to the nearest estuary, low elevation, and semi-arid climate.

To prepare the data, I first filtered it to include only transient types and calculated mean EA values. I then summarized and averaged biomass percentages for euryhaline species across different collection periods. Next, I merged the EA and biomass data with site variables, including rainfall and elevation, and imputed missing biomass estimates using linear models based on abundance data.

In building the model, I developed a comprehensive model that incorporated annual rainfall, estuary distance, elevation, and euryhaline biomass as predictors. I used stepwise selection via the stepAIC function to pinpoint the most significant predictors. Multiple linear models were generated, and their performance was evaluated using metrics such as R-squared, adjusted R-squared, p-values, AIC, and other statistics.

For the analysis, I evaluated and ranked the models based on AIC to determine the best-fitting models. Additionally, I visualized the relationships between the predictors and EA through scatter plots and regression lines, enhancing the understanding of these relationships. Lastly, I examined the possibility of an interaction between rainfall and estuary distance using visualization.

```r
# ------------------------------------------------------------------------------
# Multiple regression
# ------------------------------------------------------------------------------


# ------------------------------------------------------------------------------
# Data prep
# ------------------------------------------------------------------------------
mr_EA <- iso %>% filter(dataset == 'transient_type') %>%
  rename(transient_type = m_group, EA_transient_mu = mean, EA_transient_sd=sd,
         site_code = site) %>%
  rename(transient = transient_type) %>%
  filter(transient %in% c('Freshwater')) %>%
  select(site_code, transient, EA_transient_mu, `2.5%`, `97.5%`) %>%
  pivot_wider(names_from = transient, values_from = EA_transient_mu, names_prefix = 'EA_')

mr_biomass <- dc %>%
  filter(collection_period %in% c('2020-Q1', '2019-Q4')) %>%
  group_by(site_code, transient, collection_period) %>%
  summarize(total_biomass_percent = sum(biomass_percent)) %>%
  group_by(site_code, transient) %>%
  summarize(mean_total_biomass_percent = mean(total_biomass_percent, na.rm=T)) %>%
  pivot_wider(values_from = mean_total_biomass_percent, names_from = transient,
              names_prefix = 'b_') %>%
  mutate(b_Euryhaline = b_Amphidromous + b_Catadromous + b_Euryhaline)

d_mr <- left_join(mr_EA, mr_biomass) %>%
  add_rain() %>%
  add_sitevars()

# Impute missing population estimate data at Calallen Dam
```

```r
iso_raw <- read_csv(here('03_public', 'output', 'isotope_CNS_2020_01_clean.csv'))

# predict transient biomass using transient proportion based on abundance data at streams
d_bio_pred_prep <- iso_raw %>%
  filter(site_code %in% c(my_streams, 'UN', 'LN')) %>%
  filter(guild %in% c('Fish', 'Invertebrate')) %>%
  group_by(site_code, transient_type) %>%
  summarize(n_samples = length(carbon)) %>%
  pivot_wider(names_from = transient_type, values_from = n_samples) %>%
  group_by(site_code, Freshwater) %>%
  summarize(transient_sum = sum(Amphidromous, Catadromous, Euryhaline, na.rm=T)) %>%
  ungroup() %>%
  mutate(abun_tp = transient_sum / (transient_sum + Freshwater)) %>%
  left_join(d_mr %>% select(site_code, b_Euryhaline))

# model
m_bio_pred_prep <- lm(b_Euryhaline ~ abun_tp, data = d_bio_pred_prep)
t_bio_pred_prep <- summary(m_bio_pred_prep) %>% tidy()

fill_biomass <- d_bio_pred_prep %>%
  add_predictions(model=m_bio_pred_prep, var='predicted') %>%
  mutate(b_Euryhaline = ifelse(is.na(b_Euryhaline), predicted, b_Euryhaline)) %>%
  add_rain() %>%
  add_sitevars()

# fill missing b_euryhaline at calallen dam
d_mr <- d_mr %>%
  select(-b_Euryhaline) %>%
  left_join(fill_biomass %>%
  select(site_code, b_Euryhaline))

# ------------------------------------------------------------------------------
# Model building and stepwise selection
# ------------------------------------------------------------------------------
my_full_formula <- formula(
  EA_Freshwater~ annualrain + elev_site_m + baydist_km + b_Euryhaline)
full_astep <- lm(my_full_formula, data =  d_mr)
astep <- stepAIC(full_astep, direction='both')

# ------------------------------------------------------------------------------
# Define and summarize multiple linear models
# ------------------------------------------------------------------------------
d_lm <- tibble(
  x_string = c(
    "EA_Freshwater~ annualrain",
    "EA_Freshwater~ b_Euryhaline",
    "EA_Freshwater~ baydist_km",
    "EA_Freshwater~ elev_site_m",
    "EA_Freshwater~ annualrain + baydist_km",
    "EA_Freshwater~ annualrain + elev_site_m",
    "EA_Freshwater~ annualrain + b_Euryhaline",
    "EA_Freshwater~ elev_site_m + b_Euryhaline",
    "EA_Freshwater~ baydist_km + b_Euryhaline",
```

```r
    "EA_Freshwater~ baydist_km + elev_site_m",
    "EA_Freshwater~ annualrain + elev_site_m + baydist_km"))

extract_lm_smry <- function(a_string) {
  a_string %>%
  formula() %>%
  lm(data=d_mr) %>%
  glance() %>%
  select(r.squared, adj.r.squared, p.value, AIC, nobs) }

table_lm_basdist_AIC <- d_lm %>%
  mutate(lm_smry = map(x_string, extract_lm_smry)) %>%
  unnest(lm_smry) %>%
  arrange(AIC)

# ---------------------------------------------------------------------
# Generate best fit models and summarize
# ---------------------------------------------------------------------
make_lm_stats_AIC <- function(xdata) {
  lm(formula(x_string), data = xdata) %>%
    tidy() }

best_formulas <- table_lm_basdist_AIC %>%
  slice(1:5) %>%
  select(x_string, adj.r.squared, p.value, AIC) %>%
  nest_by(formula = x_string) %>%
  mutate(lm_stats = list(lm(as.formula(formula), data = d_mr) %>% tidy())) %>%
  ungroup() %>%
  unnest(data) %>%
  rename(Eq.adj.r2=adj.r.squared, Eq.p=p.value) %>%
  unnest(lm_stats) %>%
  select(-std.error, -statistic,-x_string) %>%
  filter(term %in% c('annualrain', 'baydist_km', 'elev_site_m',
                     'b_Euryhaline')) %>%
  mutate(term = str_replace_all(term, 'annualrain', 'Rain'),
         term = str_replace_all(term, 'baydist_km', 'Distance'),
         term = str_replace_all(term, 'elev_site_m', 'Elevation'),
         term = str_replace_all(term, 'b_Euryhaline', 'Euryhaline'),
         formula = str_replace_all(formula, 'annualrain', 'Rain'),
         formula = str_replace_all(formula, 'baydist_km', 'Distance'),
         formula = str_replace_all(formula, 'elev_site_m', 'Elevation'),
         formula = str_replace_all(formula, 'b_Euryhaline', 'Euryhaline'),
         formula = str_replace_all(formula, 'EA_Freshwater~ ', 'EA ~ ')) %>%
  pivot_wider(names_from=term, values_from = estimate) %>%
  rename(coef.p=p.value) %>%
  select(-coef.p, everything(), coef.p) %>%
  arrange(AIC)

rank1 <- table_lm_basdist_AIC %>% slice(1) %>% pull(x_string) %>% formula()
best_fit <- lm(rank1, data = d_mr)
table_best_fit_equation <- best_fit %>% tidy()

rain_eury_model <- lm(
```

```r
  formula('EA_Freshwater ~ annualrain + b_Euryhaline'),
  data = d_mr)

env_fit_model <- lm(
  formula('EA_Freshwater~ annualrain + baydist_km'),
  data = d_mr)


# --------------------------------------------------------------------------
# Predictor response cross preparation
# --------------------------------------------------------------------------
x_predictor <- c('annualrain','baydist_km','elev_site_m', 'b_Euryhaline')
y_response <- c('EA_Freshwater')
z_cross <- crossing(x_predictor, y_response)

t_best_fit <- z_cross %>%
  mutate(data = list(d_mr)) %>%
  group_by(x_predictor, y_response) %>%
  mutate(lm = pmap(list(data, x_predictor, y_response), table_lm_stats)) %>%
  unnest(lm) %>%
  ungroup() %>%
  select(-contains('my'))


# --------------------------------------------------------------------------
# Visualization
# --------------------------------------------------------------------------
p_EA_lm_explore <- t_best_fit %>%
  unnest(data) %>%
  group_by(response, predictor) %>%
  nest() %>%
  mutate(p_ea_lm = pmap(list(data, predictor, response), plot_ea_lm_explore))

plot_best_predictors <- d_mr %>%
  select(annualrain, baydist_km, elev_site_m, b_Euryhaline, EA_Freshwater) %>%
  pivot_longer(cols=c(annualrain, baydist_km, elev_site_m, b_Euryhaline),
               names_to = 'predictor', values_to = 'value') %>%
  left_join(t_best_fit) %>%
  mutate(predictor = case_when(
     predictor == 'annualrain' ~ 'Annual Rainfall (cm)',
     predictor == 'baydist_km' ~ 'Estuary Distance (km)',
     predictor == 'elev_site_m' ~ 'Elevation (m)',
     T ~ 'Transient Biomass (%)')) %>%
  ggplot(aes(x=value, y = EA_Freshwater)) +
  facet_wrap(~predictor, scales = 'free_x') +
  stat_poly_eq(label.x=.5, label.y=.95, formula=y~x,
               color='black', use_label(c("adj.R2","p")), size=4) +
  geom_point(size=3, color='blue', fill='skyblue', shape=21, alpha=.5) +
  geom_point(size=3, color='blue', fill=NA, shape=21) +
  theme_bw(base_size=12) +
  geom_smooth(data = . %>% filter(p.value < 0.1 & p.value >= 0.05),
              method = "lm", se = FALSE,
              color = "blue", lwd=.5, lty=2)  +
  geom_smooth(data = . %>% filter(p.value < 0.05),
              method = "lm", se = FALSE,
```

```r
                  color = "blue", lwd=.5, lty=1) +
  labs(x= element_blank(), y = 'Freshwater EA (%)')

plot_best_fit_interaction <- d_mr %>%
  create_site_group() %>%
  ggplot(aes(x=baydist_km, y = EA_Freshwater,
             fill=site_group, color = site_group, shape = site_group)) +
  geom_point(alpha = .5, shape = 21, size = 3, show.legend = T) +
  geom_point(color='black', fill=NA, size = 3, show.legend = T) +
  geom_smooth(method = 'lm',
              aes(lty=site_group), se=F, lwd=.8, lty=2, show.legend=F) +
  geom_text_repel(aes(label=site_code), color='black', fill=NA) +
  scale_x_log10() +
  scale_shape_manual('', values = c(21,22,23)) +
  scale_color_manual('', values=c('tan', 'green3', 'blue'))+
  scale_linetype_manual('', values = 1:3) +
  scale_fill_manual('', values=c('tan', 'green3', 'blue')) +
  labs(y= "EA in Freshwater Taxa (%)",
       x= "Estuary Distance (km)") +
  theme_bw(base_size=12)
```

5.2 Table: Freshwater EA Prediction Model Rankings

Multiple regression equations and coefficient statistics for the top five AIC-ranked models predicting average percentage estuarine assimilation (EA) within freshwater taxa at each site (n=11). Equation statistics include the adjusted R^2 (Eq.adj.r²), *p*-value (Eq.p), and AIC score. Coefficient estimates and their associated *p*-values (coef.p) are also provided.

```
best_formulas %>% gt() %>% fmt_number(columns = where(is.numeric), decimals = 3) %>%
  fmt_missing(missing_text = '')
```

formula	Eq.adj.r2	Eq.p	AIC	Rain	Euryhaline	Distance	Elevation	coef.p
EA ~ Rain + Distance	0.385	0.058	100.076	−0.685				0.152
EA ~ Rain + Distance	0.385	0.058	100.076			−0.399		0.102
EA ~ Rain + Euryhaline	0.332	0.082	100.996	−0.740				0.136
EA ~ Rain + Euryhaline	0.332	0.082	100.996		0.624			0.153
EA ~ Distance	0.282	0.053	101.077			−0.498		0.053
EA ~ Distance + Elevation	0.292	0.103	101.628			−0.709		0.045
EA ~ Distance + Elevation	0.292	0.103	101.628				0.595	0.320
EA ~ Rain + Elevation + Distance	0.315	0.141	101.807	−0.583				0.299
EA ~ Rain + Elevation + Distance	0.315	0.141	101.807				0.261	0.690
EA ~ Rain + Elevation + Distance	0.315	0.141	101.807			−0.507		0.185

5.3 Table: Best Freshwater EA Prediction Model

Multiple regression statistics for the best-fit (lowest AIC ranked) model, including parameter coefficient estimates with their standard error, F-statistic, and *p*-value.

```
table_best_fit_equation %>%
  filter(! str_detect(term, 'Intercept')) %>%
  gt() %>% fmt_number(columns = where(is.numeric), decimals = 3)
```

term	estimate	std.error	statistic	p.value
annualrain	−0.685	0.432	−1.584	0.152
baydist_km	−0.399	0.216	−1.846	0.102

5.3.1 Interpretation: Freshwater EA prediction models

Multiple regression analysis identified significant predictors of average estuarine assimilation (EA) within freshwater taxa across 11 study sites.

The top-performing model, with an adjusted R-squared value of 0.385, included annual rainfall and estuary distance as predictors. For each centimeter increase in annual rainfall, EA decreased by 0.685%, and for each kilometer increase in estuary distance, EA decreased by 0.39% (p = 0.058).

The second-best model, with an adjusted R-squared value of 0.332, highlighted annual rainfall and euryhaline biomass percentage as predictors. It revealed a 0.74% decrease in EA per centimeter increase in rainfall and a 0.624% increase in EA per percentage increase in euryhaline biomass (p = 0.082).

Estuary distance emerged as the sole predictor in the third-ranked model, with an adjusted R-squared value of 0.292. It showed a significant negative effect on EA, with a decrease of 0.709% per kilometer increase (p = 0.053).

These results underscore the significance of environmental factors, particularly annual rainfall and estuary distance, in shaping EA within freshwater taxa across diverse habitats. Further investigation into the influence of euryhaline biomass percentage is warranted, despite its marginal significance in the regression models.

5.4 Figure: Freshwater EA predictors

Univariate regression analysis depicting the relationship between mean estuarine assimilation (EA) within freshwater taxa at each site and site predictors, along with the euryhaline percentage of total community biomass. The top displays the adjusted R-squared and p-value for the linear regression model. Dotted lines represent equations with p-values below 0.1, while solid lines represent equations with p-values below 0.05.

`plot_best_predictors`
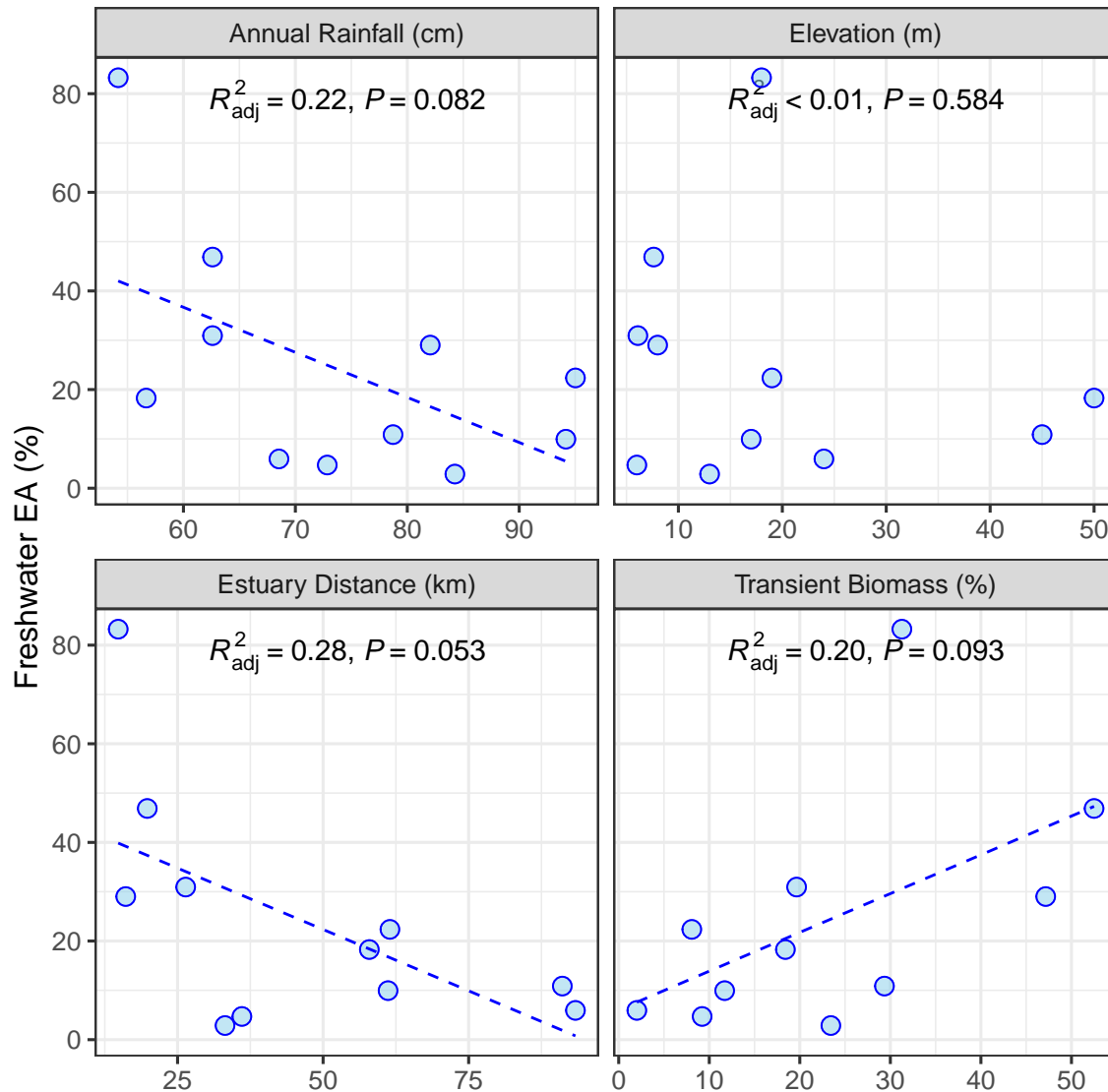
5.4.1 Interpretation: Freswhater EA predictors

As annual rainfall and estuary distance increase, we observe a consistent decline in mean estuarine assimilation (EA) within freshwater taxa. This trend suggests that areas with higher rainfall and farther proximity to estuaries experience reduced EA levels among freshwater species. Consequently, sites characterized by low estuary distance and low rainfall exhibit the highest EA within the study region, likely due to increased accessibility to estuarine resources and lower dilution effects from rainfall runoff.

Furthermore, the positive correlation between transient biomass and EA indicates that euryhaline species serve as indicators of estuarine exploitation by freshwater stream taxa. The presence of higher transient biomass signifies greater utilization of estuarine resources by freshwater species. However, the exact

mechanism driving this relationship remains ambiguous. It could be speculated that freshwater species indirectly exploit estuarine-derived nutrients through predation on euryhaline species or directly consume estuarine resources.

In summary, the highest levels of EA within the study region are found in sites characterized by low estuary distance, low annual rainfall, and a high proportion of euryhaline taxa.

5.5 Figure: Freshwater EA prediction model interaction

Regression analysis depicting the relationship between estuarine assimilation (EA) in freshwater taxa and estuary distance, grouped by precipitation category. This segmentation into three categories of rainfall allows for the observation of variations in the slope of the relationship between annual rainfall and estuary distance.

```
plot_best_fit_interaction
```
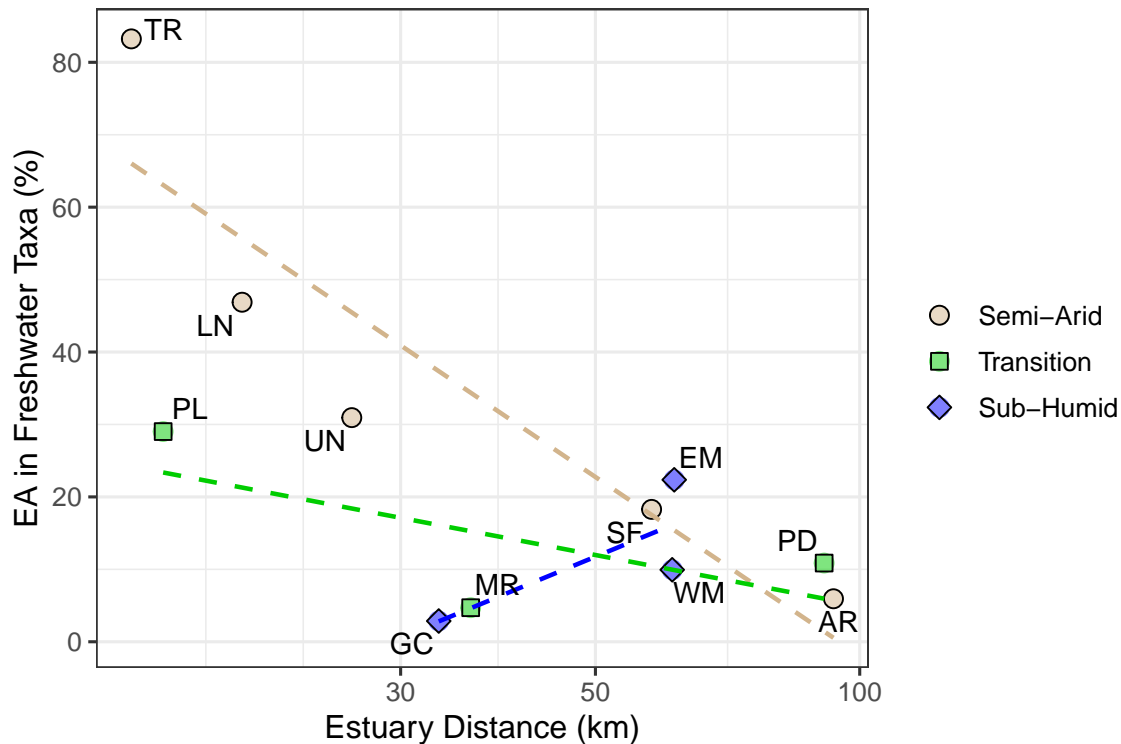

5.5.1 Interpretation: Freshwater EA predictor interaction

Semi-Arid sites exhibited the most pronounced and negative slope, indicating a substantial decline in estuarine assimilation (EA) with increasing distance from estuaries. This suggests that freshwater taxa in Semi-Arid regions are highly reliant on estuarine habitats, with their EA decreasing sharply as they move away from these environments. Transition sites, on the other hand, displayed a more moderate slope, indicating a less dramatic decline in EA with increasing estuary distance. This implies that freshwater taxa in Transition regions may have a somewhat lower dependence on estuarine habitats compared to Semi-Arid areas, yet still rely to some extent on these environments for certain ecological functions.

However, it's important to note that these results are based on very few data points, and none were statistically significant. These interpretations are purely visual and should be interpreted with caution.

5.6   Conclusions: Predicting EA Using Multiple Site Predictors

The multiple regression analysis illuminates the complex dynamics of estuarine assimilation (EA) within freshwater taxa across diverse habitats, revealing a nuanced interplay between environmental variables and EA levels. Our findings indicate a consistent decline in EA with increasing annual rainfall and estuary distance, suggesting that regions with higher rainfall and greater distance from estuaries tend to harbor lower EA levels among freshwater species. This trend may be attributed to dilution effects from rainfall runoff and limited accessibility of estuarine resources in distant areas.

Furthermore, our analysis identifies a positive correlation between euryhaline biomass and EA, suggesting that higher abundance of euryhaline taxa may enhance estuarine resource utilization by freshwater species. However, further investigation is needed to understand the underlying mechanisms driving this relationship.

Additionally, the negative impact of estuary distance on EA underscores the crucial role of habitat connectivity in facilitating estuarine assimilation by freshwater taxa. Sites closer to estuaries exhibit higher EA levels, highlighting the importance of maintaining or enhancing connectivity between freshwater and estuarine habitats to support ecosystem functioning and biodiversity preservation.

In conclusion, our regression models provide valuable tools for assessing the potential impacts of dams on estuarine connectivity within coastal streams. Specifically, they enable prediction of how dams may alter natural estuarine connectivity, with implications for freshwater taxa across habitats. Our findings suggest that dams located near estuaries in Semi-Arid climates are likely to cause the most significant disruptions to estuarine connectivity. These regions, where freshwater taxa heavily rely on estuarine habitats, may experience pronounced declines in EA due to dam construction. Moving forward, integrating our models into conservation and management strategies can inform decisions aimed at preserving and restoring estuarine connectivity. Prioritizing efforts to mitigate dam impacts on EA levels can enhance the resilience of freshwater ecosystems in the face of anthropogenic disturbances.

6 Dam Effects

6.1 Setup

The multiple regression analyses have yielded predictive models for estimating estuarine assimilation in coastal streams, utilizing variables such as annual rainfall, estuary distance, and the proportion of euryhaline biomass within the community. In this study, we assess two selected predictive models (Rain + Distance and Rain + Euryhaline) in relation to observed data and their implications for dam effects.

To evaluate the performance of these models, we compare their predictions to observations using Root Mean Square Error (RMSE). Additionally, we visualize site community EA plotted against Annual Rainfall, presenting both observations and predictions. This comparison allows us to assess the effectiveness of a purely environmental model versus one incorporating euryhaline biomass as an indicator of estuarine exploitation within a stream community.

```r
iso2 <- read_csv(here('03_public', 'output', 'CS_mix_out_region.csv'))

table_UN_prediction <- iso2 %>%
  rename(site_code = m_group) %>%
  filter(dataset == 'site') %>%
  rename(EA_Freshwater = mean, EA_transient_sd=sd) %>%
  select(site_code, EA_Freshwater, `2.5%`, `97.5%`) %>%
  left_join(
    fill_biomass %>%
      select(site_code, elev_site_m, annualrain, baydist_km, b_Euryhaline)) %>%
  add_predictions(model=rain_eury_model, var='Pre_Bio') %>%
  add_predictions(model=env_fit_model, var = 'Pre_Env') %>%
  mutate(Pre_Bio = ifelse(Pre_Bio < 0, 0, Pre_Bio),
         Pre_Env = ifelse(Pre_Env < 0, 0, Pre_Env)) %>%
  rename(Observed=EA_Freshwater) %>%
  add_is_dam()

# Table
t_prediction <- table_UN_prediction %>%
  select(site_code, is_dam, annualrain, baydist_km, b_Euryhaline,
         Pre_Bio, Pre_Env, Observed, `2.5%`, `97.5%`) %>%
  arrange(is_dam, annualrain) %>%
  rename(Site = site_code, Dam = is_dam, Rainfall = annualrain,
         Distance = baydist_km, Euryhaline = b_Euryhaline,
         'Rain+Eury' = Pre_Bio, 'Rain+Dist' = Pre_Env)

# Visualize
d_prep_cal <- table_UN_prediction %>%
  pivot_longer(cols= c(Pre_Bio, Pre_Env),
               names_to = 'type',
               values_to = 'x_val') %>%
  select(site_code, baydist_km, annualrain, type, x_val, b_Euryhaline,
         Observed, `2.5%`, `97.5%`) %>%
  add_is_dam() %>%
  mutate(type = ifelse(type == 'Pre_Bio', 'Rain+Euryhaline',
                       'Rain+Distance'))

t_RMSE <- d_prep_cal %>%
  mutate(diff2 = (x_val-Observed)^2) %>%
  group_by(type, is_dam) %>%
```

```r
  summarize(RMSE = sqrt(mean(diff2))) %>%
  mutate(annualrain = 90, x_val = 75,
    my_label = paste('RMSE', round(RMSE,0), sep = ' = '))

p_dam <- d_prep_cal  %>%
  ggplot(aes(x=annualrain)) +
  facet_grid(type~is_dam) +
  geom_text(aes(y = x_val, label=my_label),
            data = t_RMSE) +
  geom_smooth(aes(y=x_val), se=F, method='loess',
              span=1.2, lty=2, lwd=.4, color = 'red') +
  geom_errorbar(aes(ymin=`2.5%`, ymax=`97.5%`), width = 0.2) +
  geom_point(aes(y=x_val), shape = 21, size=3,
             fill = 'pink', color = 'red') +
  geom_label(aes(y=Observed, label=site_code, fill=b_Euryhaline),
             size=4,
             show.legend = T, color = 'blue') +
  scale_color_manual(values = c('blue', 'red')) +
  scale_fill_viridis_c('Euryhaline\n(%)',
                       direction=-1, begin = .6, end = 1) +
  scale_shape_manual(values=c(21, 22)) +
  labs(x= 'Annual Rainfall (cm)' , y='Mean Community EA (%)') +
  theme_bw(base_size=15) +
  theme(legend.position = 'top') +
  theme(legend.background = element_rect(colour = 'white', fill = 'white',
                                         linetype='solid'))
```

6.2 Figure: Dam Effects Vs Predicted EA

Comparison of predicted (red) and observed (yellow-blue) mean community estuarine assimilation (EA) against annual rainfall, stratified by dam-presence and model type. Observation points are colored from yellow to blue, indicating increasing euryhaline biomass percentage within a site. Predictions are generated using multiple regression models (either Rain + Distance or Rain + Euryhaline). Root Mean Square Error (RMSE) for each panel is shown in the upper right corner.

`p_dam`
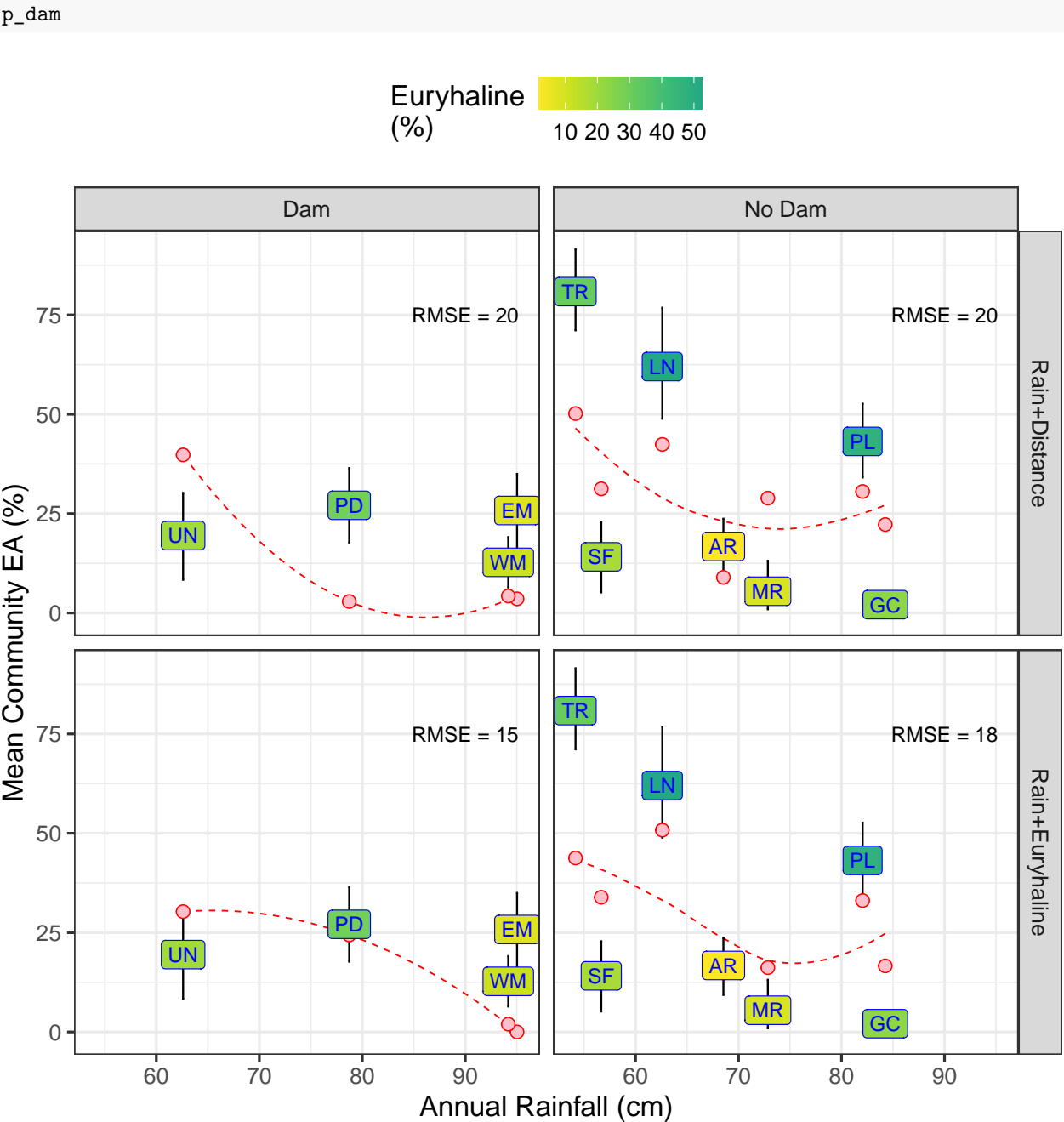
6.3   Table Dam Effects Vs Predicted EA

Comparison table detailing sites grouped by dam presence, environmental predictors (annual rainfall in cm, estuary distance in km, euryhaline biomass percentage), predictions from the top two AIC-ranked multiple regression models, observed mean EA for each site, and corresponding 95% confidence intervals.

```
t_prediction %>% gt(groupname_col = 'Dam') %>%
  fmt_number(decimals = 2)
```

Site	Rainfall	Distance	Euryhaline	Rain+Eury	Rain+Dist	Observed	2.5%	97.5%
Dam								
UN	62.60	26.39	19.65	30.29	39.78	19.48	8.35	30.23
PD	78.73	91.09	29.35	24.41	2.89	27.09	17.70	36.47
WM	94.17	61.16	11.70	1.98	4.27	12.74	6.41	19.10
EM	95.02	61.48	8.07	0.00	3.55	25.79	16.20	34.98
No Dam								
TR	54.17	14.81	31.27	43.78	50.18	80.87	71.12	91.53
SF	56.67	57.94	18.41	33.90	31.24	14.13	5.17	22.83
LN	62.60	19.79	52.50	50.79	42.42	62.01	48.89	76.84
AR	68.53	93.34	2.00	14.89	8.97	16.74	9.31	23.72
MR	72.85	36.05	9.21	16.19	28.90	5.47	0.93	13.13
PL	82.05	16.09	47.16	33.07	30.57	43.17	34.08	52.68
GC	84.25	33.15	23.43	16.63	22.25	2.09	0.46	5.00

6.4   Conclusions: Dam Effects

Multiple regression models provided valuable insights into the complex dynamics of estuarine assimilation (EA) within freshwater taxa across diverse habitats. The second-ranked model, as indicated by the Akaike Information Criterion (AIC), demonstrated superior predictive performance with a lower Root Mean Square Error (RMSE) (Rain+Euryhaline, RMSE = 18) compared to the second-ranked model (Rain+Distance, RMSE=20). This outcome underscores the significance of integrating euryhaline biomass percentage as an indicator of estuarine assimilation within the community, enhancing the overall accuracy of the predictive model.

Upon closer examination of the Calallen dam sites (UN and LN), intriguing patterns emerged. Both models consistently exhibited a trend of overpredicting observed EA above the dam (UN) while underpredicting EA below the dam (LN). This observation aligns with the expected consequence of dam presence at Calallen, which is anticipated to disrupt estuarine assimilation to the site above the dam (UN), resulting in an observed EA substantially lower than expected. Our models effectively captured this anticipated consequence, reaffirming the natural patterns in EA within this study region.