

Development Change in Texas Eurycea Habitat

Annabelle White

Contents

Rationale and Research Questions	5
Dataset Information	6
Exploratory Analysis	8
Analysis	11
Question 1: Has development increased over time across Texas?	11
Question 2: Has development increased over time in Eurycea habitats?	11
Question 3: Has development increased over time in new species' habitats?	11
Question 4: Is development in Texas associated with base flow index?	13
Summary and Conclusions	14
References	15
Appendix	15

List of Tables

1	Table 1: StreamCat Base Flow Index	6
2	Table 2: Land Cover in Eurycea Catchments	6
3	Table 3: Development in Texas Catchments	7

List of Figures

1	Development in Eurycea Habitat over Time	8
2	Development in Undescribed Species Habitat over Time	8
3	Forest in Eurycea Habitat over Time	9
4	Forest in Undescribed Species Habitat over Time	9
5	Development in Texas Catchments over Time	11
6	Development in Eurycea Catchments over Time	12
7	Development in Undescribed Species Catchments over Time	12
8	Effect of Development on Base Flow Index in Texas Catchments	13

```
## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr  1.0.1
## v tibble  3.1.8      v dplyr  1.1.0
## v tidyr   1.3.0      v stringr 1.5.0
## v readr   2.1.3      v forcats 1.0.0
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
##
## Attaching package: 'lubridate'
##
##
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
##
##
## Attaching package: 'rlang'
##
##
## The following objects are masked from 'package:purrr':
##
##     %@%, flatten, flatten_chr, flatten_dbl, flatten_int, flatten_lgl,
##     flatten_raw, invoke, splice
```

Rationale and Research Questions

Genetic testing has delineated fourteen species of Eurycea cave salamanders in central Texas, three of which are undescribed (Devitt et al. 2019). These undescribed species are new to science and therefore enjoy no protections under the Endangered Species Act (ESA). The first of these species was the subject of a petition by Save our Springs Alliance (SOS) to the U.S. Fish and Wildlife Service (USFWS) for its protection (Datri 2021). This petition passed its initial review and USFWS is now considering the species for protection (FWS 2022). Anthropogenic development including land cover change and groundwater overdraft pose the largest threats to Eurycea species by destroying forests that protect their habitats and draining the groundwater in which they live (Datri 2021).

During the 2022-2023 school year, I worked with Rory Kuczek on a NSOE master's project for SOS to facilitate and strengthen future petitions by analyzing habitat threats to these undescribed species. Kuczek and I used MaxEnt and nonmetric multidimensional scaling (NMS) statistical methods to test the association of Eurycea species presences with 13 catchment-level environmental metrics from the StreamCat dataset. Both methods identified that undescribed Eurycea species 2 and 3 presences were most dependent on base flow index (BFI), a ratio of groundwater discharge to total surface flow.

Building on these results, I intend to test several research questions related to habitat degradation for Eurycea species 2 and 3:

1. Has development increased over time across Texas?
2. Has development increased over time in Eurycea habitats?
3. Has development increased over time in new species' habitats?
4. Is development in Texas associated with base flow index?

Answers to these questions should paint a clearer picture of development in Texas and in Eurycea habitat.

Dataset Information

The StreamCat dataset details over 600 natural and anthropogenic metrics for streams in the continental US at the catchment and watershed levels. The dataset includes robust web services such as a REST API and R package to facilitate data management. StreamCat is maintained by EPA as part of the National Aquatic Resource Surveys. Use of this dataset for analysis of Eurycea salamander species was recommended by John Fay.

The StreamCat package and API provide metrics listed by COMID at the catchment and watershed level, along with catchment and watershed area. The `sc_nlcd` function returns all NLCD metrics, in this format, for a given year and area of interest. Column names include the metrics' year of acquisition, which made it difficult to join the data into one dataframe and graph metrics over time.

To format the data as desired, I wrote a for loop which incrementally obtained NLCD from each year available, added a new "Year" column to record the year, altered the column names to remove the year, and then joined the resulting data to a growing dataframe of NLCD by year. The result is a large dataframe with all NLCD metrics for the specified area of interest (Texas) and all years available.

I subsequently wrangled this dataframe to remove duplicates and NA values, change the "Year" column into a date object, and select only the relevant columns: COMID, Species, Year, BFI (at the catchment level), and summed percents of developed and forested land cover.

Salamander presences were obtained from scientific literature (Devitt et al. 2019) and expert communication in government petitions (Datri 2021).

```
## [1] 8
```

Table 1: Table 1: StreamCat Base Flow Index

COMID	State	Watershed Area (km^2)	Catchment Area (km^2)	Catchment BFI (%)	Watershed BFI (%)
20007253	TX	160.9497	5.5665	11.0000	12.4137
20007337	TX	123.7698	9.0765	11.0000	11.0000
20006195	TX	31344.1605	0.1116	12.0000	37.5727
20005395	TX	6.3792	0.8622	10.0000	10.0000
20006847	TX	33860.1150	0.0504	11.0000	35.7625
20007553	TX	1073.1015	0.1701	12.0000	14.9827
20007325	TX	31682.8746	0.3753	12.0000	37.3049
20007071	TX	106.9029	4.7997	14.0000	11.7099
20005117	TX	6.1128	6.1128	11.0000	11.0000
20005979	TX	7.9641	1.3770	9.9725	8.9366

```
## 'summarise()' has grouped output by 'Year'. You can override using the
## '.groups' argument.
## 'summarise()' has grouped output by 'Year'. You can override using the
## '.groups' argument.
```

Table 2: Table 2: Land Cover in Eurycea Catchments

COMID	Species	Year	Base Flow Index (%)	Development (%)	Forest (%)
317741	sp3	2011-01-01	33.9521	0.22	5.42
317741	sp3	2019-01-01	33.9521	0.23	10.80

COMID	Species	Year	Base Flow Index (%)	Development (%)	Forest (%)
317741	sp3	2001-01-01	33.9521	0.21	5.39
317741	sp3	2013-01-01	33.9521	0.23	5.40
317741	sp3	2006-01-01	33.9521	0.22	5.40
317741	sp3	2008-01-01	33.9521	0.22	5.41
317741	sp3	2016-01-01	33.9521	0.23	5.30
317741	sp3	2004-01-01	33.9521	0.22	5.24
317745	sp3	2001-01-01	39.0042	28.29	0.96
317745	sp3	2011-01-01	39.0042	36.05	0.96

Table 3: Table 3: Development in Texas Catchments

Year	Mean Development (%)
2001-01-01	5.066942
2004-01-01	5.150763
2006-01-01	5.329008
2008-01-01	5.439416
2011-01-01	5.596450
2013-01-01	5.657407
2016-01-01	5.797787
2019-01-01	5.902000

Exploratory Analysis

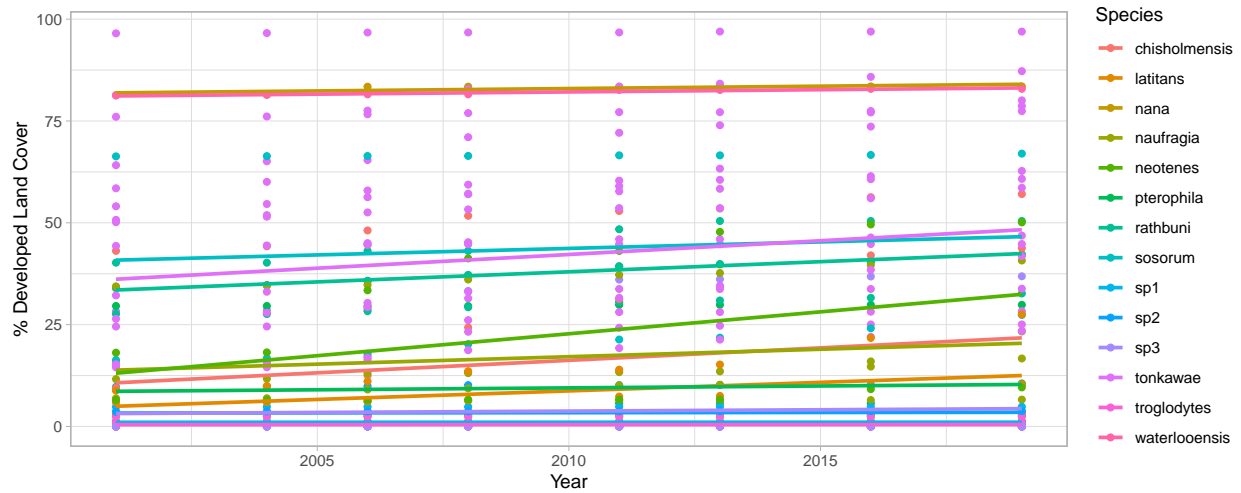


Figure 1: Development in Eurycea Habitat over Time

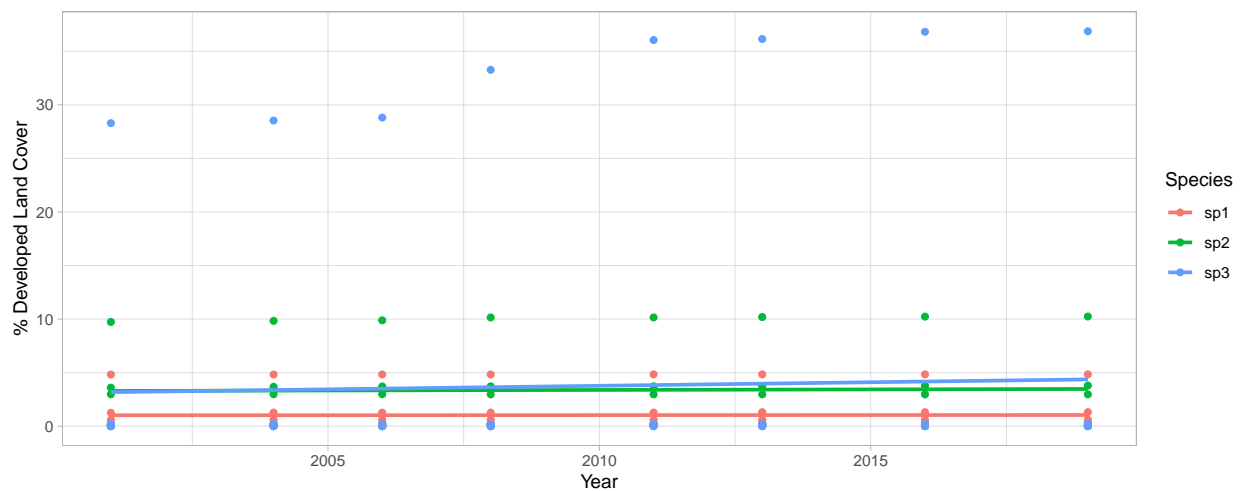


Figure 2: Development in Undescribed Species Habitat over Time

This first plot shows a clear upward trend in development within catchments from 2001 to 2019 for most Eurycea species. The slope of this trend is more extreme for some species, especially *E. neotenes*. (I attempted to configure a colorblind-friendly color palette, but symbolizing 14 colors is a tall task even for broader palettes.)

The second plot shows development over time for the three undescribed species. Initial takeaways are that sp. 3 has the highest percentage of development within its catchments from 2001 to 2019, followed by sp. 2 and sp. 1. It is also clear that known sites for these species have dramatic outliers, with one sp. 3 site routinely scoring at around 30-35% developed land cover, whereas most sites have 10% or less development.

Conversely, a plot of forested land cover over time shows a clear downward trend for most Eurycea species. This trend is most dramatic, again, for *E. neotenes*. Together, these plots suggest that most deforestation and development is occurring in *E. neotenes* catchments relative to other Eurycea species.

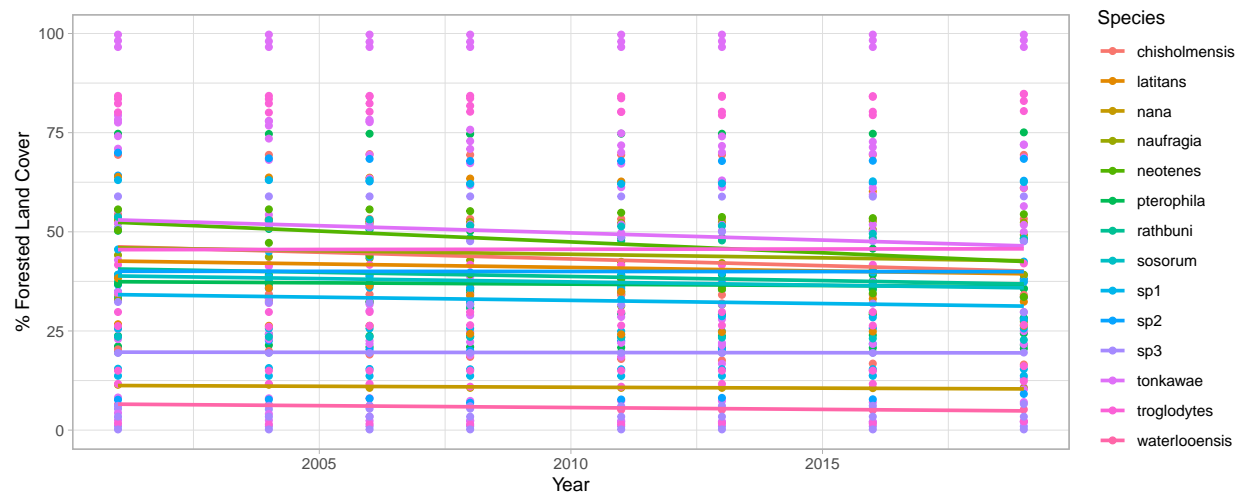


Figure 3: Forest in Eurycea Habitat over Time

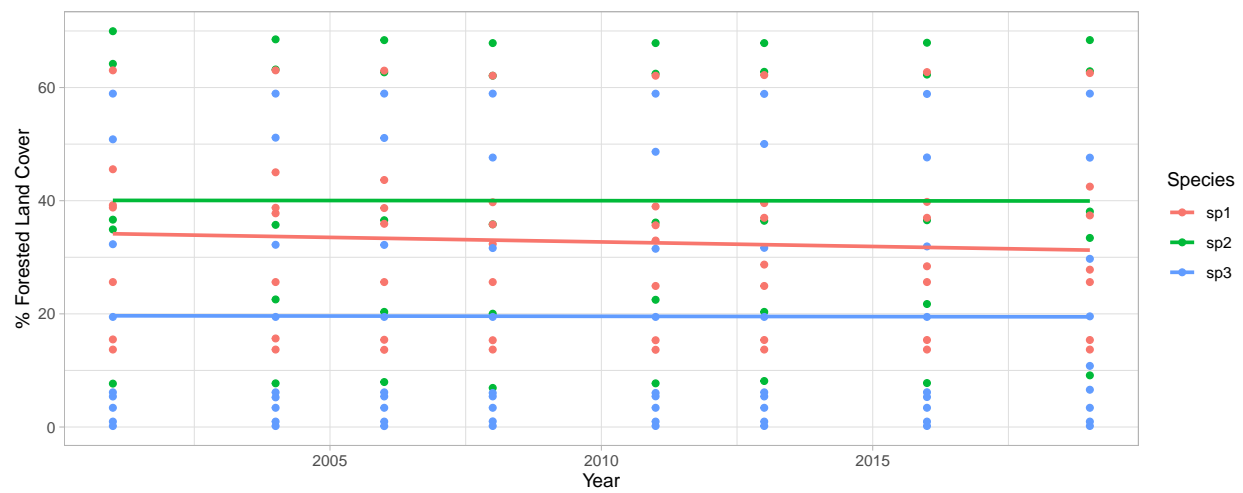


Figure 4: Forest in Undescribed Species Habitat over Time

The second plot again restricts this analysis to the undescribed species. Forest cover remains relatively constant for sp. 2 (20%) and sp. 3 (40%) over time, but shows a downward trend for sp. 1.

Analysis

I set out to test whether and how development had changed over time. I realized that time-series analysis requires data from continuous, regular time intervals. NLCD is only assessed every 2-3 years, which makes it poorly suited for this task. However, I soldiered on.

Question 1: Has development increased over time across Texas?

```
## Score = 28 , Var(Score) = 65.33334  
## denominator = 28  
## tau = 1, 2-sided pvalue =0.00083661
```

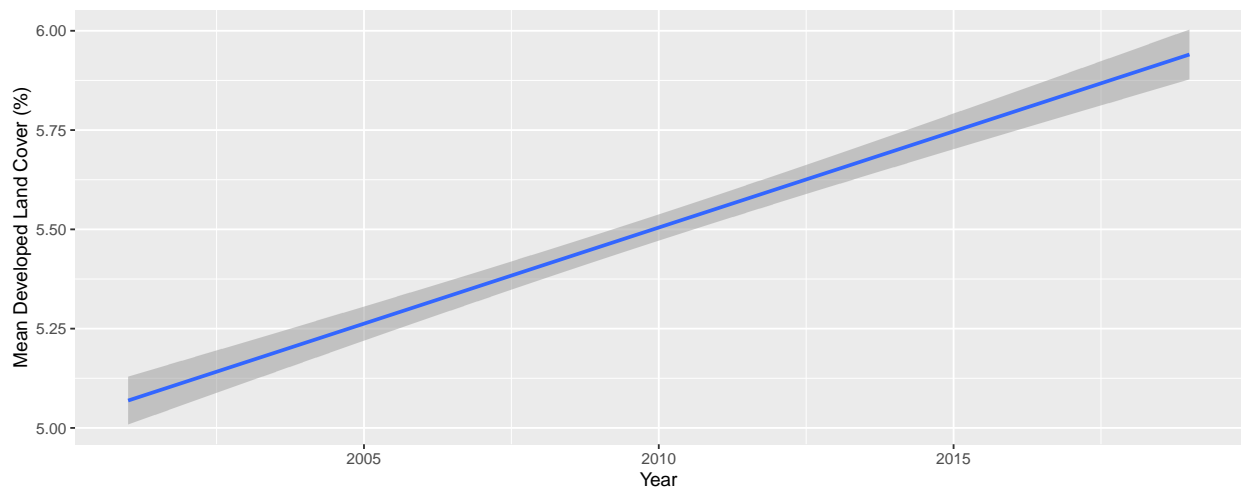


Figure 5: Development in Texas Catchments over Time

A simple Kendall time-series analysis of the data shows a clear upward trend in development throughout Texas stream catchments. The test achieves significance (two-sided p-value = 0.00083661).

Question 2: Has development increased over time in Eurycea habitats?

```
## Score = 643 , Var(Score) = 155478.1  
## denominator = 5833.617  
## tau = 0.11, 2-sided pvalue =0.10349
```

A simple Kendall time-series analysis of the data shows no trend in development throughout Eurycea stream catchments. The test fails to achieve significance (two-sided p-value = 0.10349).

Question 3: Has development increased over time in new species' habitats?

```
## Score = 105 , Var(Score) = 1590.771  
## denominator = 261.3274  
## tau = 0.402, 2-sided pvalue =0.0091197
```

A simple Kendall time-series analysis of the data shows a clear upward trend in development throughout unnamed species' stream catchments. The test achieves significance (two-sided p-value = 0.00091197).

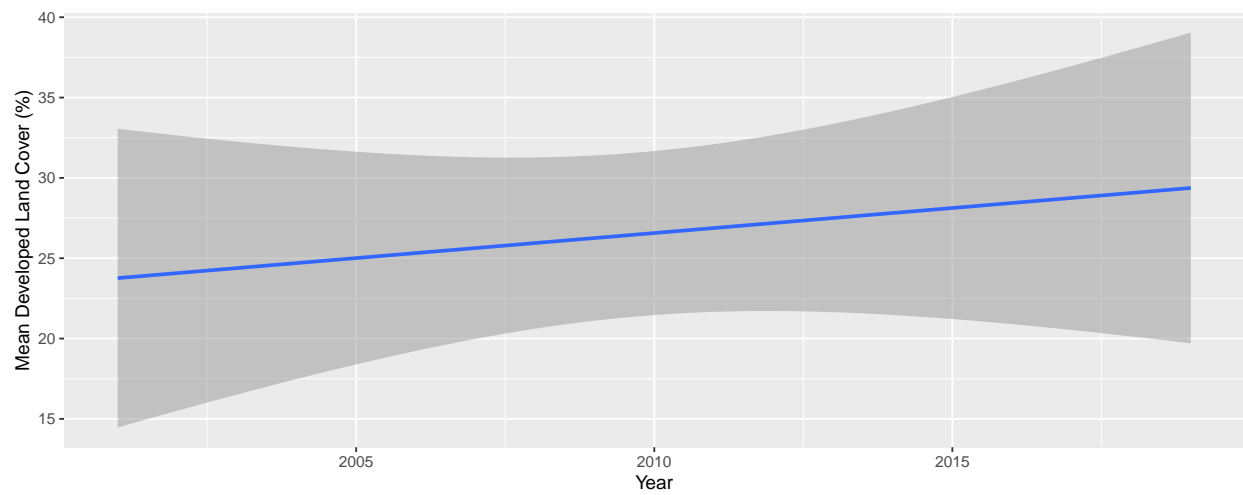


Figure 6: Development in Eurycea Catchments over Time

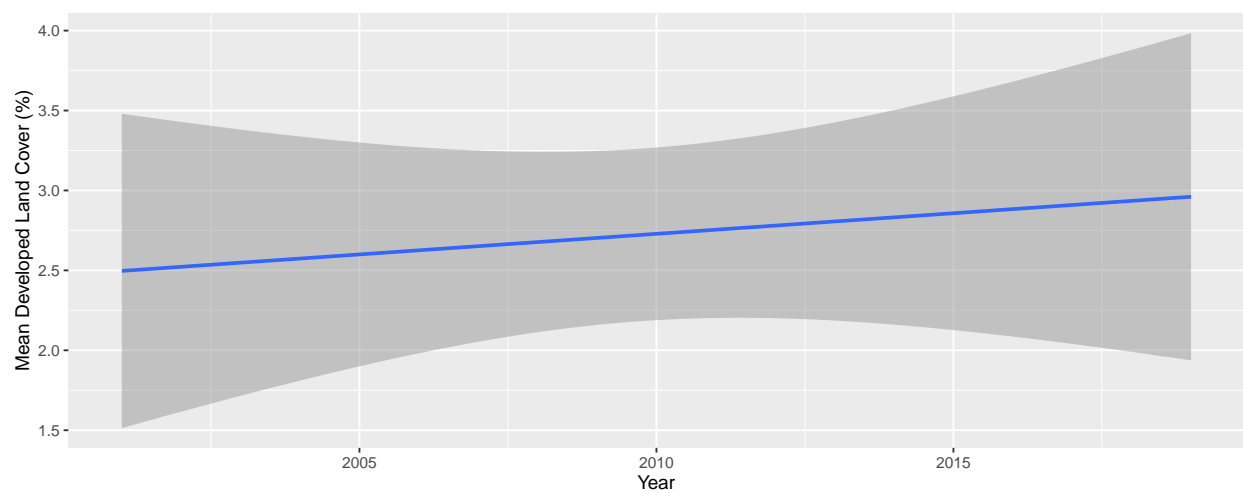


Figure 7: Development in Undescribed Species Catchments over Time

Question 4: Is development in Texas associated with base flow index?

```
##
## Call:
## lm(formula = BFI ~ PercentDevelop, data = Eurycea)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.310  -8.310  -2.235   6.078  42.463
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  21.3100159  0.0132758 1605.18  <2e-16 ***
## PercentDevelop -0.0301961  0.0009314  -32.42  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.75 on 769582 degrees of freedom
## Multiple R-squared:  0.001364, Adjusted R-squared:  0.001362
## F-statistic: 1051 on 1 and 769582 DF, p-value: < 2.2e-16
```

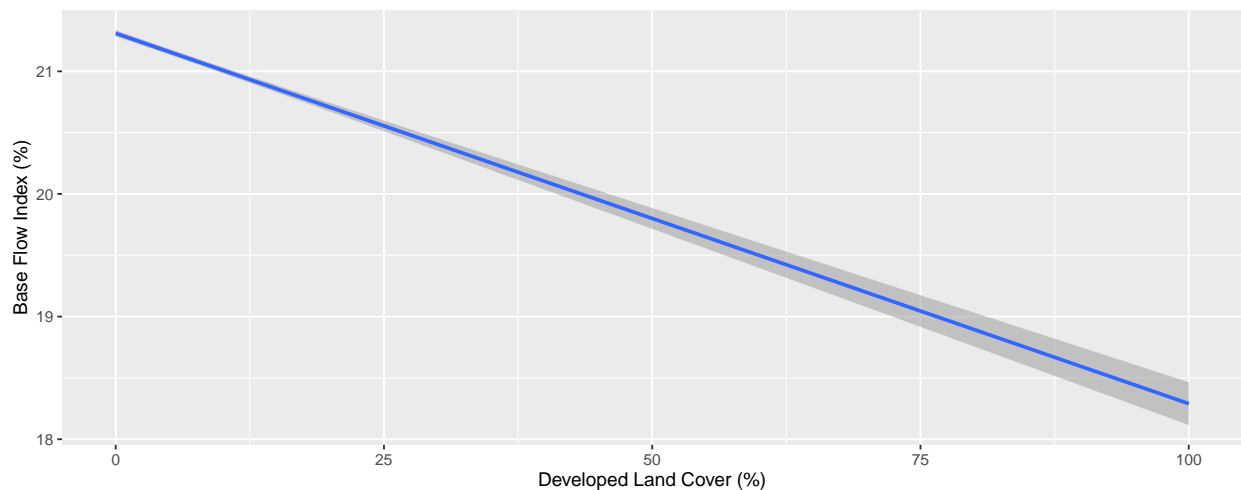


Figure 8: Effect of Development on Base Flow Index in Texas Catchments

This simple linear model suggests that base flow index and development cover are negatively correlated. For each 1% increase in development cover within a catchment, BFI decreases by 3%. The model achieves significance (p-value < 2.2e-16), but has an adjusted R-squared of less than 1%.

Summary and Conclusions

Despite the inconvenient format of the dataset, I was able to perform time-series analysis on the percentage of developed land cover within Texas stream catchments from 2001 to 2019. Almost all of my tests achieved significance. From these tests, I can conclude that:

(Question 1) the mean percent of developed land cover within Texas stream catchments has increased from 2001 to 2019;

(Question 2) the mean percent of developed land cover within stream catchments inhabited by all *Eurycea* species has not changed from 2001 to 2019;

(Question 3) the mean percent of developed land cover within stream catchments inhabited by undescribed *Eurycea* species has increased from 2001 to 2019;

and (Question 4) as developed land cover within a catchment increases, base flow index within the catchment decreases.

If true, these are sensible and useful results to see. It is broadly understood that Texas is increasingly developing its land as population centers like Austin and San Antonio expand. This development poses a threat to *Eurycea* habitat. Question 4's linear model evinces this threat: as development increases, base flow index decreases threefold. This is likely because developed areas build wells to extract groundwater for human consumption, reducing the amount of groundwater available to *Eurycea* in those catchments.

Across *Eurycea* as a genus, there is no clear trend in development change. This may be because these species occupy different areas within the Edwards-Trinity aquifer system. For example, *E. sosorum* and *E. waterlooensis* habitat is centered in Austin. These catchments may already be maximally developed and have little room to change. Other species, such as *E. troglodytes*, primarily occupy low-development catchments in western Hill Country. These catchments have not yet seen increased development, but may be targeted for development in future. Species like *E. neotenes* occupy suburban and ex-urban areas, which are at the forefront of development from urban sprawl. This could explain why *E. neotenes* shows the clearest upward trend in graphs of development change.

However, when subset to undescribed *Eurycea* species 1, 2, and 3, there is a clear upward trend in development change. This trend is primarily driven by outlier sites for species 3, which have increased from below 30% development to over 35% development. This suggests an increasing rate of development at at least one catchment inhabited by species 3, which may be cause for petitioning USFWS to protect this species.

There are several key weaknesses to this analysis. The adjusted R-squared of Question 4's linear model is minuscule. This suggests that while development may have a relation to base flow index, it is far from the only variable explaining rates of groundwater discharge. Additionally, the time-series data for questions 1-3 is choppy and irregular. This may limit the accuracy of the conclusions drawn.

Overall, this analysis paints a simple picture. Development is increasing in Texas, particularly in catchments occupied by undescribed *Eurycea* species. This development negatively impacts groundwater discharge, which is crucial for *Eurycea* habitat. It is clearly imperative to petition USFWS for the protection of these undescribed species.

GitHub repository: <https://github.com/AnnabelleWhite/ESM>

References

Datri, Crystal (2021). "Petition to the U.S. Fish and Wildlife Service to List the Pedernales River Springs Salamander as a Threatened or Endangered Species Under the Endangered Species Act."

Devitt, T. J., Wright, A. M., Cannatella, D. C., & Hillis, D. M. (2019). Species delimitation in endangered groundwater salamanders: Implications for aquifer management and biodiversity conservation. *Proceedings of the National Academy of Sciences*, 116(7), 2624-2633.

Service Completes Initial Reviews on Endangered Species Act Petitions for Four Species | U.S. Fish & Wildlife Service. 2022 Oct 18. FWSgov. [accessed 2023 May 1]. <https://fws.gov/press-release/2022-10/service-completes-initial-reviews-endangered-species-act-petitions-four>.

Weber M (2023). StreamCatTools: Streamcat Tools. R package version 0.1.1.9002, <https://usepa.github.io/StreamCatTools/>.

Appendix

```
# Set your working directory

# Load your packages
library(tidyverse);library(dataRetrieval);library(lubridate);library(remotes)
library(ggplot2);library(ggthemes)
library(rlang);library(Kendall);library(knitr)
#install_github("USEPA/StreamCatTools", build_vignettes=TRUE, force=TRUE)
library("StreamCatTools") # This imports EPA's tool for StreamCat data

# Pulling NLCD from StreamCat for every COMID in TX for every year available

years <- c(2001, 2004, 2006, 2008, 2011, 2013, 2016, 2019)
dates <- c('20010101', '20040101', '20060101', '20080101', '20110101',
           '20130101', '20160101', '20190101') # Best to have ymd format
length(years) # 8 years

nlcd <- data.frame() # Create empty dataframe to be filled

for (i in 1:8) { # For each year for which there are NLCD...
  nlcd_i <- sc_nlcd(years[i], state = "TX") # Create temporary df for this year
  nlcd_i$Year = dates[i] # Denote which year these data are from
  for (col in 1:ncol(nlcd_i)){
    colnames(nlcd_i)[col] <- sub(years[i], "", colnames(nlcd_i)[col])
  } # All the column names have the year in them. I want them removed
  nlcd <- rbind(nlcd, nlcd_i) # Adjoin to growing dataframe of all NLCD
  rm(nlcd_i) # Remove the temporary dataframe
}

#summary(nlcd) # Check the resulting dataframe
# All the fields starting with "PCT" should have a max of 100.00
# Year should run from min: 2001 to max: 2019
# I note there are a lot of NA's; we'll remove those later

# List of Eurycea species found in each COMID (from my MP)
species <- read.csv("species.csv", stringsAsFactors = TRUE)
```

```

#summary(species) # This gives a sense of which species have most presences

# Pull BFI for every COMID in TX (this is comparably easy)
bfi <- sc_get_data(metric = "BFI", state = "TX")
#summary(bfi) # Again we see a lot of NA's, to be removed later

# Assemble the final dataframe
nlcd_bfi <- merge(nlcd, bfi) # Intermediary df with the env. variables
env_vars <- merge(species, nlcd_bfi) # Adjoin to salamander presences

kable(head(bfi, n = 10),
       col.names = c("COMID", "State", "Watershed Area ($km^2$)",
                     "Catchment Area ($km^2$)", "Catchment BFI (%)",
                     "Watershed BFI (%)"),
       align = c('r', 'l'),
       caption = "Table 1: StreamCat Base Flow Index")

rm(nlcd_bfi);rm(bfi);rm(nlcd) # Remove extraneous datasets

Eurycea <-
  env_vars %>%
  distinct() %>%
  drop_na() %>%
  mutate(Year = ymd(Year)) %>%
  mutate(BFI = BFICAT) %>% # We're working at the catchment level only
  mutate(PercentDevelop = PCTURBLOCAT + PCTURBMDCAT + PCTURBHICAT + PCTURBOPCAT) %>%
  mutate(PercentForest = PCTCONIFCAT + PCTDECIDCAT + PCTMXFSTCAT) %>%
  select(COMID, Species, Year, BFI, PercentDevelop, PercentForest)

Eurycea.only <- # Dataframe of only salamander presences
  Eurycea %>%
  filter(Species != "background")

Eurycea.new <- # Dataframe of only undescribed species
  Eurycea %>%
  filter(Species %in% c("sp1", "sp2", "sp3"))

Dev.yearly <-
  Eurycea %>%
  group_by(Year) %>%
  summarise(Develop.Mean = mean(PercentDevelop))

Dev.yearly.Eurycea <- # Dataframe of only salamander presences
  Eurycea.only %>%
  group_by(Year, Species) %>%
  summarise(Develop.Mean = mean(PercentDevelop))

Dev.yearly.new <- # Dataframe of only undescribed species
  Eurycea.new %>%
  group_by(Year, Species) %>%
  summarise(Develop.Mean = mean(PercentDevelop))

```



```

kable(head(Eurycea.only, n = 10),
      col.names = c("COMID", "Species", "Year", "Base Flow Index (%)",
                    "Development (%)", "Forest (%)"),
      align = c('r', 'l'), # Align columns
      caption = "Table 2: Land Cover in Eurycea Catchments") # Title table

kable(Dev.yearly,
      col.names = c("Year", "Mean Development (%)"),
      align = c('r', 'l'),
      caption = "Table 3: Development in Texas Catchments")

ggplot(Eurycea.only, aes(x = Year,
                        y = PercentDevelop,
                        color = Species)) +
  geom_point() +
  # Add line of best fit for each species
  stat_smooth(method = "lm",
             formula = y ~ x,
             se = FALSE,
             alpha = 0.5) +
  xlab("Year") +
  ylab("% Developed Land Cover") +
  theme_light()

ggplot(Eurycea.new, aes(x = Year,
                      y = PercentDevelop,
                      color = Species)) +
  geom_point() +
  # Add line of best fit for each species
  stat_smooth(method = "lm",
             formula = y ~ x,
             se = FALSE,
             alpha = 0.5) +
  xlab("Year") +
  ylab("% Developed Land Cover") +
  theme_light()
# I could set ylim to remove outliers from this graph
# But I think those outliers are an important part of this story

ggplot(Eurycea.only, aes(x = Year,
                        y = PercentForest,
                        color = Species)) +
  geom_point() +
  # Add line of best fit for each species
  stat_smooth(method = "lm",
             formula = y ~ x,
             se = FALSE,
             alpha = 0.5) +
  xlab("Year") +
  ylab("% Forested Land Cover") +
  theme_light()
ggplot(Eurycea.new, aes(x = Year,

```

```

        y = PercentForest,
        color = Species)) +
    geom_point() +
    # Add line of best fit for each species
    stat_smooth(method = "lm",
                formula = y ~ x,
                se = FALSE,
                alpha = 0.5) +
    xlab("Year") +
    ylab("% Forested Land Cover") +
    theme_light()

Dev.yearly.ts <- ts(Dev.yearly$Develop.Mean,
                   start = c(2001,1,1), end = c(2019,1,1), frequency = 1)
Dev.yearly.trend <- Kendall(Dev.yearly$Year, Dev.yearly$Develop.Mean)
summary(Dev.yearly.trend) # p-value = 0.00083661
ggplot(Dev.yearly, aes(x = Year,
                      y = Develop.Mean)) +
    stat_smooth(method = "lm",
                formula = y ~ x,
                se = TRUE,
                alpha = 0.5) +
    xlab("Year") +
    ylab("Mean Developed Land Cover (%)")

Dev.yearly.Eurycea.ts <- ts(Dev.yearly.Eurycea$Develop.Mean,
                           start = c(2001,1,1), end = c(2019,1,1), frequency = 1)
Dev.yearly.Eurycea.trend <- Kendall(Dev.yearly.Eurycea$Year,
                                   Dev.yearly.Eurycea$Develop.Mean)
summary(Dev.yearly.Eurycea.trend) # p-value = 0.10349
ggplot(Dev.yearly.Eurycea, aes(x = Year,
                              y = Develop.Mean)) +
    stat_smooth(method = "lm",
                formula = y ~ x,
                se = TRUE,
                alpha = 0.5) +
    xlab("Year") +
    ylab("Mean Developed Land Cover (%)")

Dev.yearly.new.ts <- ts(Dev.yearly.new$Develop.Mean,
                       start = c(2001,1,1), end = c(2019,1,1), frequency = 1)
Dev.yearly.new.trend <- Kendall(Dev.yearly.new$Year,
                               Dev.yearly.new$Develop.Mean)
summary(Dev.yearly.new.trend) # p-value = 0.00091197
ggplot(Dev.yearly.new, aes(x = Year,
                          y = Develop.Mean)) +
    stat_smooth(method = "lm",
                formula = y ~ x,
                se = TRUE,
                alpha = 0.5) +

```

```

xlab("Year") +
ylab("Mean Developed Land Cover (%)")

bfidev.lm <- lm(data = Eurycea, BFI ~ PercentDevelop)
summary(bfidev.lm) # Shows significant negative effect
ggplot(Eurycea, aes(x = PercentDevelop,
                    y = BFI)) +
  stat_smooth(method = "lm",
              formula = y ~ x,
              se = TRUE,
              alpha = 0.5) +
  xlab("Developed Land Cover (%)") +
  ylab("Base Flow Index (%)")

```