

# Stormwater Heatmap Technical Reference

Review Draft. Not for release.

Christian Nilsen, Geosyntec Consultants, Inc.

Emily Howe, The Nature Conservancy

Jamie Robertson, The Nature Conservancy

Last updated: 2020-05-22

# Contents

<b>Abstract</b>	<b>4</b>
<b>1 Introduction</b>	<b>5</b>
1.1 What is Stormwater? . . . . .	5
1.2 Urban stormwater runoff is a two-fold problem, impacting the quantity of water pulsing off the land, as well as the quality of that water. . . . .	6
1.3 Toxic Stormwater Impacts . . . . .	6
1.4 Moving forward- identifying where stormwater pollution is generated on the landscape .	7
1.5 Building an interactive tool to service stormwater manager needs: . . . . .	8
1.6 Open data helps us bound forward . . . . .	10
<b>2 High-resolution Land Cover</b>	<b>11</b>
2.1 Overview . . . . .	11
2.2 Land Cover Development . . . . .	12
2.3 Validation and Accuracy Assessment . . . . .	13
<b>3 Hydrology</b>	<b>16</b>
3.1 Overview . . . . .	16
3.2 Modeling approach . . . . .	16
3.3 Data Sources . . . . .	18
3.4 Verification of Results . . . . .	21
3.5 Spatially Aggregated Results . . . . .	29

<b>4 Water Quality Statistics</b>	<b>34</b>
4.1 Data Sources . . . . .	34
4.2 Spatial data . . . . .	35
4.3 Methods . . . . .	39
4.4 Results . . . . .	49
4.5 Predictions . . . . .	76
4.6 Discussion . . . . .	85
4.7 Generation of heatmap layers . . . . .	85
<b>5 References</b>	<b>87</b>
<b>A Regional HSPF Calibration Factors</b>	<b>90</b>
<b>B Instructions for Accessing Tabulated Hydrology Results via BigQuery</b>	<b>112</b>
Tabulated Results via BigQuery . . . . .	112

# List of Tables

2.2	Accuracy and kappa statistics for three derivatives of the seven-class land cover product.	15
4.2	Zinc - Water - Total . . . . .	41
4.3	Copper - Water - Total . . . . .	42
4.4	Nitrite-Nitrate - Water - Dissolved . . . . .	43
4.5	Lead - Water - Total . . . . .	44
4.6	Total Phosphorus - Water - Total . . . . .	45
4.7	Total Suspended Solids - Water - Total . . . . .	46
4.8	Total Phthalate - Water - Total . . . . .	46
4.9	Total PAH - Water - Total . . . . .	47
4.10	CPAH - Water - Total . . . . .	48
4.11	HPAH - Water - Total . . . . .	48

# List of Figures

3.1	HSPF PERLND Conceptual Model . . . . .	17
3.2	WRF Forecasting Grid . . . . .	19
3.3	Observed and simulated flow-duration curves for Madsen Creek, King County, WA . . .	28
3.4	Example flow duration curves of altered and forested land covers . . . . .	32
3.5	Summary of flow index values in study area . . . . .	33
4.1	All observations - outliers in place . . . . .	36
4.2	Quantile-quantile plots of COCs, log scale . . . . .	37
4.3	All observations - outliers removed . . . . .	38

# Abstract

The Stormwater Heatmap project springs forth from The Nature Conservancy's (TNC) Cities program, which strives to bring TNC's core mission – preserving and protecting the land and waters upon which all life depends – to urban areas. As the leading contributor of toxic pollutants to Puget Sound's streams, rivers, and marine waters, urban stormwater runoff is a key ecological problem generated by Washington's urban landscapes. Urban stormwater runoff has harmed virtually all urban and urbanizing streams and rivers, and delivered massive quantities of toxic contaminants to Puget Sound. As a result, the abundance and survival of aquatic species has declined. A recurrent question asked of stormwater managers, eco-toxicologists and ecologists alike, is how much stormwater intervention is needed, and where would you place it for efficient and effective treatment?

This “how much and where” question serves as the foundation of the stormwater heatmap project. The project quantitatively visualizes a pollution loading threat-map for the Puget Sound watershed. The stormwater heatmap uses land use, landcover, stormwater monitoring data, precipitation data, and hydrological modeling to predictively map stormwater pollution loading across the Puget Sound landscape, and quantifies the pollution load on a 1-m<sup>2</sup> spatial resolution. This scale allows managers and planners to aggregate data at multiple spatial scales, and allows us to “see” hotspots of pollution loading for a variety of monitored stormwater contaminants. With the pollution visualization layer in place, we can now begin to overlay social-ecological questions in order to develop a stormwater intervention action-map.

# **Chapter 1**

## **Introduction**

### **1.1 What is Stormwater?**

One of the primary terrestrial pressures on the Salish Sea estuarine and marine environment is urban stormwater runoff. When rainfall runs across hard, impervious surfaces, rather than soaking into the soil, it picks up and delivers toxic contaminants directly to nearby streams, rivers, and eventually the Salish Sea. In fact, for most toxic substances, surface runoff is the largest contributing source of loading to Puget Sound (Washington State Department of Ecology 2011).

Unfortunately, the Salish Sea's relationship with stormwater effluent is no outlier; stormwater is the fastest growing cause of surface water impairment in the United States as urbanization transitions forested and other natural landscapes to hard, impervious surfaces (United States Environmental Protection Agency 2019). Given that the Salish Sea is expected to house another 5 million people by 2040, stormwater interventions will be necessary in order to break the relationship between urbanization and stormwater-caused ecological degradation.

Fortunately, researchers have uncovered a variety of successful techniques to reduce stormwater impairment of surface and receiving waters, including street sweeping, pervious pavement, and green stormwater infrastructure wherein stormwater is filtered by soil and plant mixtures on its way between the streets and the sea. These interventions are costly (~\$65-132 billion is needed to restore Puget Sound to hydraulically function like a forest), but the costs of stormwater pollution are high as well: the sickening and deaths of Salish Sea organisms. Annual losses due to one contaminant (polycyclic aromatic hydrocarbon exposure) alone are estimated to be between \$4.4 to \$12.1 billion (King County 2014; Washington State Department of Ecology and Washington Department of Health 2012).

## **1.2 Urban stormwater runoff is a two-fold problem, impacting the quantity of water pulsing off the land, as well as the quality of that water.**

As a result of stormwater's twin problems, urban watersheds and marine receiving waters suffer from "urban syndrome"- a condition that results in low abundance and survival of sensitive aquatic and coastal species (Walsh et al. 2005). Virtually all urban streams and rivers in Puget sound have been harmed by stormwater pollution xxx (Booth et al. 2004).

### **1.2.1 Water Quantity**

Watersheds with as little as 5-10% impervious surface area- such as rooftops, roads, and paved parking areas- exhibit aquatic habitat degradation as a result of increased surface runoff (Walsh et al., 2005). This changes the timing, magnitude, and frequency of high flow events, making urban streams "flashier" than those with natural surrounding landcover conditions. These hydrological changes cause combined sewer overflow events, flooding, erosion, and scouring of stream and riverbeds. Flashy hydrology disrupts habitat structure and alters the ecology of freshwater ecosystems themselves, but also disrupts larger ecosystem processes in marine environments such as nutrient flux, organic matter processing, and ecosystem metabolism (Palmer and Ruhi 2019) While coastal food webs rely on rivers to deliver organisms, nutrients, and detritus from the land to the sea, these fluxes increasingly result in negative impacts, such as eutrophication, hypoxia, and harmful algal blooms.

### **1.2.2 Water Quality**

In addition to altering hydrological flow regimes in watersheds contributing to the Salish Sea, urban stormwater also delivers a suite of contaminants that severely impact the water quality of streams, rivers, estuaries, and the Salish Sea itself. Urban runoff contains complex and unpredictable mixtures of chemicals, including persistent organic pollutants (i.e. PCPs), heavy metals (i.e. copper, zinc), hydrocarbons (i.e. motor oil, tailpipe emissions, rubber tire particles), nutrients (nitrogen, phosphorous), pesticides, and pharmaceuticals . Toxic pollutants entering the Salish Sea may be metabolized in plant and animal tissues, bioaccumulated in tissues, incorporated into sediments, volatized, degraded, or conserved in marine waters.

## **1.3 Toxic Stormwater Impacts**

Researchers have documented toxic effects of stormwater exposure for a diverse range of aquatic and marine species, ranging from primary producers to high trophic-level predators. Some effects are sublethal, reducing species fitness and long-term survival. For example, heavy metal accumulation is common among marine macroalgae and eelgrass (*Zostera marina*), reducing photosynthetic function (Jarvis and Bielmyer-Fraser 2015; Lyngby and Brix 1984). Other sublethal impacts of stormwater on marine organisms include the reduction of byssus strength in marine mussels (Gaw, Thomas, and Hutchinson 2014), reduced olfactory function in juvenile salmonids (Baldwin, Tatara, and Scholz 2011),

reduced growth and lipid storage in juvenile Chinook (Meador, Sommers, and Ylitalo 2006), reduced pathogen resistance in juvenile salmon (Arkoosh et al. 2001), cardiotoxicity in juvenile fish (Incardona, Linbo, and Scholz 2011), decreased reproductive function and immune response in benthic fishes (Rice et al. 2000), seals (Anan et al. 2002), and Southern Resident Killer Whales (Kayhanian et al. 2012; Ross et al. 2000; WDFW 2011).

Some effects are acutely lethal, as is the case for adult coho salmon, where pre-spawn mortality rates in urban streams can be as high as 90% (Scholz et al. 2011). These fish end their years-long journey to the ocean and back with their bellies still full of unfertilized eggs, missing their single chance to spawn the next generation. For coho, it appears that pre-spawn mortality is linked to the transportation network, where contaminants, like tire wear leachates, are generated (Feist et al. 2017). Thus, development expansion and increasing use intensity of the built environment is significantly impacting the long-term viability of local Coho populations, with far-reaching ramifications for both freshwater and marine food webs alike. And while it is tempting to focus on lethal impacts to iconic species such as coho, road runoff is similarly lethal to lower trophic level organisms, such as mayfly larvae, sea urchins, and amphipods, which all play important roles in upholding marine, freshwater, and terrestrial food webs (Anderson et al. 2007; Kayhanian et al. 2012; McIntyre et al. 2015 ).

## 1.4 Moving forward- identifying where stormwater pollution is generated on the landscape

A much repeated phrase from stormwater managers is “how much and where” do we need to implement stormwater BMPs (Best Management Practices)? This is a difficult question to answer until we identify our ecological and social goals for stormwater management. The amount and spatial configuration of stormwater interception techniques will look very different depending on whether the goal is to meet permit regulations, recover coho salmon, or recover Southern Resident Killer Whales because biological organisms are susceptible to stormwater contaminants for different reasons, in different locations, at different scales, and at different points in time according to their life history traits (Levin, Howe, and Robertson, n.d.) Answering the “how much and where” question will therefore require integrating ecological data with stormwater monitoring and pollution loading data.

However, a promising starting place to answer the “how much and where” question is to build a predictive map quantifying levels of stormwater pollution loading across the landscape. This type of ‘threat’ heatmap can be coupled with ecological data to produce action maps for stormwater intervention.

To build the predictive stormwater pollution heatmap, we focused on three major steps:

**1. Landcover Refinement:** we generated a high resolution landcover dataset enabling landcover mapping at the 1-m<sup>2</sup> resolution. This is a critical level of resolution for urban runoff modeling because impervious surfaces so strongly drive hydrologic response, and therefore pollution loading. Thus, accurate mapping of impervious surfaces was needed to accurately calculate surface runoff.

**2. Hydrology:** we conducted continuous hydrology simulations for the 32 different hydrologic response units (HRU = combination of landcover, soils, & slope) found within the Puget Sound domain. Using regional precipitation datasets provided by the Climate Impacts Group, we modeled both current and future hydrology in order to assess how climate change will impact stormwater pollution loading across the landscape, generating more than 311 billion rows of data. This dataset alone provides

an efficient way to quickly model rainfall-runoff relationships across Puget Sound using the Western Washington Hydrology Model (WWHM).

**3. Pollution Statistics:** We used Bayesian statistical modeling to link stormwater monitoring data to land use and land cover characteristics to predict pollution concentrations across the landscape. These concentration predictions were then combined with hydrology output to generate the pollution load across the Puget Sound landscape at a 1-m<sup>2</sup> spatial resolution.

The resulting interactive tool enables users to visualize and aggregate stormwater pollution loading data at several spatial resolutions for local, watershed, and regional-scale planning. The project reveals that areas with high percent cover of impervious surfaces, such as hard cityscapes, as well as industrial and commercial zones, tend to produce higher pollutant loads than high-density residential, low-density residential, and rural areas. Transportation networks- roads and highways- generate very high levels of stormwater contaminants, especially those with higher traffic intensity. These high intensity roads can cut through lower-density areas, lighting up as pollution hotspots. Traffic behavior (i.e. congestion points) also plays a role, indicating that a combination of a static landscape structure and dynamic anthropogenic behavior layered atop that structure can combine to create stormwater pollution hotspots throughout the landscape.

Using the stormwater heatmap as a foundation, we can begin to integrate the ecological layers to understand exactly where on the landscape stormwater interventions will be most efficient and effective at breaking the link between urbanization and aquatic degradation. Examples of spatialized biotic response data generated by robust local monitoring programs include NOAA's MusselWatch, King County's Benthic-Index of Biotic Integrity (B-IBI), and NOAA's coho pre-spawn mortality monitoring. WDFW's Salmonscape data represents another source of ecological information, showing the timing and spatial habitat use of different species of salmonids. With respect to human well-being Front and Centered's Environmental Health Disparities Map offers data-driven insights on human health in the region.

## 1.5 Building an interactive tool to service stormwater manager needs:

The mapping tool is especially timely because the Washington Department of Ecology recently issued a new stormwater permit which increased the number of jurisdictions required to develop and implement stormwater management plans. Historically, just 4 stormwater permittees were required to submit detailed stormwater management plans and models. Now, 85 jurisdictions must develop stormwater management plans and be able to scientifically defend their prioritization and decision-making process.

In order to help move more jurisdictions towards this goal/requirement, The Nature Conservancy embarked on a Design Thinking project to better understand what tools those smaller Phase II cities and counties would need in order to meet permit requirements. The Design Thinking approach centers on a structured interview process that is human-centered. Through interviewing stormwater planners, engineers, and leaders throughout the Puget Sound region, this process identified “pinch” and “release” points that currently prevent or promote effective stormwater management. The Design Thinking project emphasized that a tool supporting stormwater management in the region requires the following elements:

- **Compelling Visuals:** tools should help stormwater managers tell a story to different audiences
- **Multiple Scales:** stormwater planning takes place at the parcel, neighborhood, watershed and regional scale. Data need to be flexibly aggregated at all scales.
- **Make it Mine-able:** serve as a data platform and resource for use with other tools. Land cover, soils, hydrology, and climate change impacts data would help meet multiple modeling needs.
- **Grounded in Science:** data and calculations should be apparent and meet current best practices

For many of the smaller jurisdictions, financial and personnel constraints are significant barriers to effective stormwater management and innovation. Most projects are opportunistic, and many stormwater management departments have only one employee.

This stormwater management tool is targeted for stormwater managers in order to 1) get the best available science and tools into the hands of decision-makers, 2) lower the costs for effective decision making and planning, and 3) improve Puget Sound water quality and recover ecological health.

Thus, in addition to the stormwater pollution loading data layer, the online mapping tool also includes data extraction capabilities and report modules that service requirements outlined by the Department of Ecology's stormwater permit. The modules can be flexibly applied at multiple scales, and include:

- Land Use/Land Cover
  - Land cover classification (% cover)
  - Land use
  - Imperviousness
- Hydrology
  - Hydrologic response units
  - Mean annual runoff
  - Flow-control metrics
  - Climate change impacts
- Pollutant loading
  - 25th, 50th, 75th concentration quantiles
  - 25th, 50th, 75th annual loading quantiles
- Other

- Age of development
- Estimated population

## 1.6 Open data helps us bound forward

Early adopters are currently testing the capabilities of the data, analyses, and approaches generated by the Stormwater heatmap project. King County is using the hydrology modeling output to prioritize and resize culverts for fish passage. Our Green/Duwamish is using the tool for stormwater management action planning, the City of Tacoma is using it for watershed prioritization, EPA's Office of Research and Development are integrating this work with the VELMA runoff model, and the City of Phoenix and Maricopa County are using it for a regional LID study.

The stormwater heatmap is an open-source tool, free for use by the public<sup>1</sup>. Code will be accessible on the stormwater heatmap Github repository.

The Stormwater heatmap can be used as a foundational layer to answer many social-ecological questions to benefit people and nature. It is built for you, with stormwater management in mind. Please use, modify, and distribute this work widely. See where you can take it!

---

<sup>1</sup>[Mozilla Public License 2.0] (<https://www.mozilla.org/en-US/MPL/2.0/>)

# Chapter 2

## High-resolution Land Cover

### 2.1 Overview

We produced a high spatial resolution 7-class land cover dataset of the Puget Sound trough below 1,500 meters elevation using a hybrid of automated classification and manual data overlays (Table 1). Here, the Puget Sound trough is defined as the areas of Washington state which drain into the Puget Sound or Straight of Juan de Fuca. This hybrid approach allowed us to produce a highly accurate land cover dataset appropriate for our stormwater modeling and using already available resources (data, software, computation, and personnel.) In Google Earth Engine (Gorelick et al., 2017), we created a six-class 1-meter land cover layer using the Naïve Bayes classifier (Google, 2019) and then used overlay decision rules to combine that layer with a 10-m land cover product from NOAA Coastal Change Analysis Program (C-CAP) (NOAA C-CAP, 2018a), water body and river polygons (Washington Department of Natural Resources, 2015), shoreline polygons (Esri, 2006), over-water structure polygons (Washington Department of Natural Resources, 2007), road polygons buffered from lines (U.S. Census Bureau, 2016), and building rooftops polygons (Microsoft, 2019). Lastly, we validated accuracy with an observed land cover point dataset from the Washington Department of Fish & Wildlife (Pierce, Jr., 2015).

Table 1. The number IDs and label names of the seven classes in the final land cover dataset.

<i>Class ID</i>	<i>Label Name</i>
1	Fine Vegetation
2	Medium Vegetation
3	Coarse Vegetation
4	Dirt/Barren
5	Water
6	Impervious Other
7	Impervious Roofs

## 2.2 Land Cover Development

Our first step was to create a 1-meter resolution land cover layer of the Puget Sound in Google Earth Engine using a Naïve Bayes classifier. We trained the classifier using a subset of a 1-meter resolution land cover dataset (NOAA C-CAP, 2018b) covering Snohomish County, Washington, produced by NOAA C-CAP as a proof-of-concept. Through trial-and-error, we decided upon and composited eleven bands from various sources into a single dataset to be trained. We created the 1-meter Red, Green, Blue and Near-infrared (NIR) reflectance bands by calculating the mean band values from NAIP aerial imagery (U.S. Department of Agriculture, 2019) collected over 2009 to 2019. The eleven composited bands included:

- Texture: calculated from NAIP NIR band as entropy, or how expected a pixel value is in the context of nearby pixels, in this case using a radius of 12 meters. For more information, see [https://developers.google.com/earth-engine/image\\_texture](https://developers.google.com/earth-engine/image_texture).
- Land use: polygons (Washington State Department of Ecology, 2010) scaled to 1-meter.
- Elevation: the USGS National Elevation Dataset ( arc-second) (U.S. Geological Survey, 2015).
- Hillshade: produced from the USGS National Elevation Dataset ( arc-second) (U.S. Geological Survey, 2015).
- GNDVI: Green Normalized Difference Vegetation Index calculated with NAIP imagery as:

$$\frac{(NIR - Green)}{(NIR + Green)}$$

- MSAVI: Modified Soil Adjusted Vegetation Index calculated with NAIP imagery as:

$$\frac{(2 \cdot NIR + 1) - \sqrt{(2 \cdot NIR + 1)^2 - 8(NIR - RED)}}{2}$$

- Binary MSAVI: a Boolean selection of MSAVI where:

$$Bin_{MSAVI} = \begin{cases} 1, & 0.5 \leq MSAVI < 0.9 \\ 0, & \text{otherwise} \end{cases}$$

- Green: Green NAIP band.
- Blue: Blue NAIP band.
- Red: Red NAIP band.
- NIR: NIR NAIP band.

The Naïve Bayes classification result was then smoothed with a dilation process (2-meter focal maximum analysis followed by a 2-m focal minimum analysis). The quality of this product varied among

the classes. Vegetated and impervious areas mapped well, but dirt/barren and water were inconsistent and often mistaken for each other or for impervious.

To improve the labels in those inconsistent locations, we used a decision rule and replacement approach using NOAA's 10-meter CCAP land cover dataset (NOAA C-CAP, 2018a), which covers western Washington including the Puget Sound. While this dataset tested high in accuracy already, 10-meters is not an adequate spatial resolution to inform many site-level analyses used in our subsequent stormwater modeling, particularly in highly urbanized areas where vegetation is present but dispersed. However, the CCAP product does well to capture land cover existing at large, continuous extents throughout the landscape, such as forests, vegetated fields, broad sandy beaches, waterbodies, and large parking lots. Recognizing this, we chose to substitute pixels of certain classes from the 1-m classification with the 10-m CCAP. We also used the CCAP water and dirt classes to help clean up shoreline areas using a 10-m focal maximum algorithm to grow the class landward. The decision and replacement approach is as follows for each 1-m pixel, where label ID numbers are listed in Table 1 above

#### Classification Pseudocode

```
IF Rooftop:  
    label = 7  
ELSE IF Dirt Road:  
    label = 4  
ELSE IF Road OR Over-Water Structure:  
    label = 6  
ELSE IF Inland waterbody or County shoreline:  
    label = 5  
ELSE IF 1-m Naïve Bayes label in (Fine vegetation, Medium vegetation, Coarse vegetation):  
    label = (1, 2, 3) respectively  
ELSE IF 1-m Naïve Bayes label = 10-m C-CAP label:  
    label = C-CAP label  
ELSE IF 10-m radius focal max of 10-m C-CAP in (4, 5):  
    label = (4,5) respectively  
ELSE IF 1-m Naïve Bayes = Dirt/Barren:  
    replace with 10-m C-CAP  
ELSE  
    10-m C-CAP
```

Our final step with the land cover creation was to limit the data to areas below 1,500-meters elevation to exclude permanent ice, e.g. glaciers on Mt. Rainier, which was not classified or necessary for our stormwater model.

## 2.3 Validation and Accuracy Assessment

We validated the predicted land cover dataset in Earth Engine by producing confusion matrices and testing accuracy and kappa statistics. Here, the accuracy test is simply calculated as the number of ground-truthed sample points which correctly match the predicted pixel label divided by the total number of sample points used. We also calculated a kappa statistic, which is generally thought to be

a more reliable test as it accounts for chance agreement of the observed and predicted labels. For both tests, zero (0) is lowest accuracy and 1 is highest.

We produced our ground-truth dataset of 34,699 points using a dataset of points and polygons provided by WDFW (Pierce, Jr., 2015). The original points data consisted of 44,882 randomly distributed locations across the Puget Sound which were assigned class names through visual inspection of 2013 NAIP imagery for the purpose of ground-truthing land cover data. The polygons data represent land cover in 2015 and were developed with a segmentation and decision tree approach. We prepared the validation points data with the following process:

#### Data preparation:

1. Subset the points by filtering out unlabeled points, points located outside our classification area (e.g. marine water), and points with labels that did not match labels of 2015 land cover polygons. This filtering retained 34,699 of the original 44,882 points.
2. Relabel names to six classes:
  1. FinalFineVegetation & FinalFineVegetation → Fine Vegetation
  2. FinalMedium → Medium Vegetation
  3. FinalCoarseVegetation → Coarse Vegetation
  4. FinalDirt → Barren/Dirt
  5. FinalWaterFinalWater → Water
  6. BuiltConflict & FinalBuilt & FinalBuiltBrownRed → Impervious Developed
3. In Earth Engine, remap the point labels to produce a second set of point features labeled Pervious and Impervious to also validate the dataset for its ability to map impervious surface.
4. We randomly selected 3,000 points to use from each of the six classes in the ground-truthed points data to reduce bias towards one or more classes.
5. At this point, the two datasets were ready for analysis. We refer to them respectively as: Ground-truthed Land Cover and Ground-truthed Impervious Surface.

Using those ground-truth points and our seven-class land cover product, we calculated accuracy and kappa statistics for three derivative products of the seven-class land cover dataset: Impervious, Six-class, and Four-class (Table 2).

- The **Impervious** derivative is a binary dataset where impervious includes rooftops, roads, and other features that force water to run off rather than percolate through the surface. These are labeled referred to as “Built” within the ground-truth points data. We remapped the “Impervious Roofs” and “Impervious Other” to the Impervious class, and all other classes were considered “Non-impervious.”
- The **Six-class** derivative includes all classes, but the two impervious classes are remapped to one single Impervious class. These classes are equivalent to the six classes labeled in the ground-truthed points data and the NOAA 10-m Puget Sound land cover dataset.
- The **Four-class** derivative includes only the Fine Vegetation, Coarse Vegetation, Water, and Impervious (Built) classes.

Table 2.2: Accuracy and kappa statistics for three derivatives of the seven-class land cover product.

County	Impervious		Six Classes		Four Classes	
	Accuracy	Kappa	Accuracy	Kappa	Accuracy	Kappa
Clallam	0.985	0.947	0.702	0.631	0.915	0.885
Island	0.987	0.942	0.537	0.447	0.912	0.877
Jefferson	0.977	0.931	0.698	0.629	0.911	0.880
King	0.971	0.898	0.635	0.567	0.929	0.904
Kitsap	0.976	0.923	0.725	0.655	0.946	0.926
Lewis	1.000	1.000	0.612	0.443	0.964	0.934
Mason	0.977	0.919	0.686	0.612	0.926	0.897
Pierce	0.973	0.912	0.626	0.548	0.888	0.848
San Juan	0.976	0.899	0.591	0.511	0.829	0.771
Skagit	0.983	0.926	0.655	0.585	0.911	0.880
Snohomish	0.958	0.860	0.646	0.581	0.926	0.901
Thurston	0.975	0.919	0.641	0.573	0.906	0.875
Whatcom	0.987	0.944	0.637	0.566	0.869	0.823

While the Impervious and Four-class derivatives test extremely well with both Accuracy and Kappa statistics, the Six-class derivative tests considerably lower. The Error (Confusion) matrices produced with the Six-class derivative dataset indicate that the Medium Vegetation and Dirt/Barren classes do not match well between our land cover product and the ground-truth points. Inspection of those matrices tells us that the Medium Vegetation ground-truth points were often labeled Fine or Coarse Vegetation in our predicted set. Similarly, the Dirt ground-truth points were often labeled Built or Fine Vegetation.

Some of the discrepancies likely occur because vegetation green-up (particularly in fine vegetation such as grass) can occur differently from year to year and cause mislabeling when comparing data from different periods. Where grass is green in one year's input imagery, it may appear as dirt or barren in another year's imagery if captured during a particularly dry period. These issues may result from error in the land cover labeling, ground-truth labeling, or actual changes in land cover that occurred between the input data dates (i.e. 2013 to 2017). Furthermore, dirt is often mistaken for impervious surfaces (and vice versa) because those surfaces are often built with dirt and stone, thus creating similar spectral reflectance.

Overall, we find the discrepancies to be low concern for the purpose of our stormwater modeling. The horizontal area of the Medium Vegetation and the Dirt/Barren classes in our land cover data is considerably smaller than the other classes, and their respective labeling discrepancies are to classes that are treated similarly to each of them respectively in the stormwater hydrology modeling (e.g. Medium Vegetation is subsequently remapped to Coarse Vegetation).

# **Chapter 3**

## **Hydrology**

### **3.1 Overview**

This document provides an overview of hydrology simulation methods and results for the Puget Sound Stormwater heatmap. Continuous hydrology simulation was performed using regional pre-calibrated parameters. Batched simulations were run for combinations of land cover, soils, and slopes across the Puget Sound domain. Results are stored in a cloud-based database. It is intended to be used in conjunction with data derived from the stormwaterheatmap or other geospatial data sources to quickly model rainfall-runoff relationships across Puget Sound.

### **3.2 Modeling approach**

The hydrologic modeling approach was developed to replicate as much as feasible, commonly applied continuous simulation hydrologic analysis for stormwater in Puget Sound. Ecology developed guidance for continuous simulation modeling as described in the Stormwater Manual for Western Washington (Department of Ecology 2014).

This guidance calls for the application of continuous simulation models based on the Hydrologic Simulation Program Fortran (HSPF). HSPF is a lumped-parameter rainfall-runoff model developed by the USGS and EPA. HSPF is generally used to perform analysis on hydrologic processes related to effects of land cover, interception, surface ponding and soil moisture retention. Although maintenance development of HSPF has not occurred since 1997, it is currently distributed by EPA under the Better Assessment Science Integrating Point and Non-point Sources (BASINS) analysis system. In Western Washington, application of HSPF to stormwater design is routinely performed through the Western Washington Hydrology Model (WWHM), a Windows-based graphical user interface program with built-in meteorologic data and modules specific to stormwater analysis.

HSPF contains a number of specialized modules that are not used by WWHM. These include modules related to snowmelt, sediment budgets, and specific water quality routines. The primary HSPF routines

used by WWHM are designated as IWATER (water budget for impervious land cover) and PWATER (water budget for pervious land cover). A graphical schematic of the PWATER routine is shown in Figure 3.1.

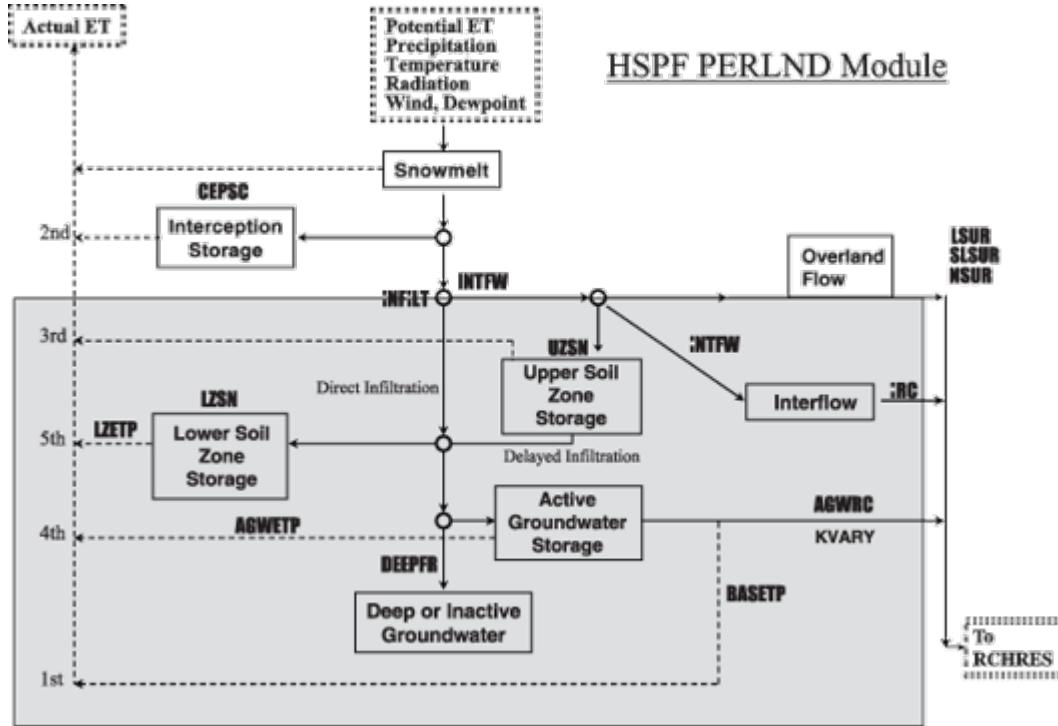


Figure 3.1: HSPF PERLND Conceptual Model

### 3.2.1 Hydrologic Response Units

Modeling was performed on discretized landscape units based on common soils, land cover, and slope characteristics known as hydrologic response units (HRUs). The HRU approach provides a computationally efficient method of pre-computing hydrologic response for later use. Results for a particular watershed can be calculated by summing or averaging the results for individual HRUs.

Each combination of parameters was modeled in separate batched simulations. HRUs were designated by a three-digit number according to the following convention:

- **First digit:** Hydrologic Soil Group Number ( $0 = A/B$ ,  $1 = C$ ,  $2 = \text{Saturated}$ )
- **Second digit:** Land cover ( $0 = \text{Forest}$ ,  $1 = \text{Pasture}$ ,  $2 = \text{Lawn}$ ,  $5 = \text{Impervious}$ ),
- **Third Digit:** Slope ( $0 = \text{Flat}$ ,  $1 = \text{Mod}$ ,  $2 = \text{Steep}$ )

For example, a site with Type C soils, with forested land cover, on a moderate slope would be represented by 101. This schema allowed for HRUs to be stored as an eight-bit unsigned integer on a Puget-Sound wide raster, minimizing storage size.

### **3.2.2 Regional Calibrated Parameters**

Regional calibration factors for the Puget Lowlands Ecoregion were developed by USGS in the 1990s (Dinicola 1990) and updated by Clear Creek Solutions for use within WWHM (Department of Ecology 2014). These parameters, referred to as the ‘default parameters’ by Ecology were used in this study and applied to individual HRUs. Parameters are provided in Appendix A.

### **3.2.3 Python Implementation**

To allow for parallel computations, we used a Python adaption of HSPF (PyHSPF<sup>1</sup>) developed by David Lambert with funding from the United States Department of Energy, Energy Efficiency & Renewable Energy, Bioenergy Technologies Office (Lampert 2019). PyHSPF is able to generate HSPF input files, run simulations, and provide HSPF compatible output. Similar to WWHM, we provided separate output files for three flow-paths: surface flow, interflow, and groundwater flow. In HSPF, these output classes are referred to as **SURO**, **INFW**, and **AGW0** respectively. We developed and ran individual PyHSPF models for each combination of HRU and Precipitation grid and generated output for each flow patch component. This resulted in 27,990 individual output files.

## **3.3 Data Sources**

### **3.3.1 Precipitation**

A region-wide, simulated precipitation dataset was provided by the University of Washington Climate Impacts Group. Methodology used to develop this dataset is documented in (Mauger et al. 2018). The dataset contains modeled hourly precipitation using the GFDL CM3 global climate model and the Representative Concentration Pathways (RCP) 8.5 scenario.

The GFDL model was chosen by CIG due to its ability to accurately model winter storm drivers, important for stormwater applications. Combined with the higher emissions scenario, this modeling scenario represents the upper end of expected future climate changes effects.

CIG downscaled GCM results using a statistical-dynamical approach to capture the anticipated changes in extreme events as well as the different drivers of rainfall that affect the Puget Sound Region. Regional simulations were performed using the Weather Research and Forecasting community mesoscale model. This resulted in hourly rainfall predictions at an approximately 12 km grid size across Puget Sound. Predictions were bias-corrected on a quantile-mapping basis (individual mean bias corrections for precipitation in each quantile range) using the historic (1970-2005) WRF data. The WRF Grid in our study area is shown in Figure 3.2.

### **3.3.2 Potential Evaporation**

Gridded potential evaporation estimates were acquired from the forcing data for the North American Land Data Assimilation System (NLDAS2) (NASA Goddard Earth Sciences Data and Information

---

<sup>1</sup><https://github.com/djlampert/PyHSPF>

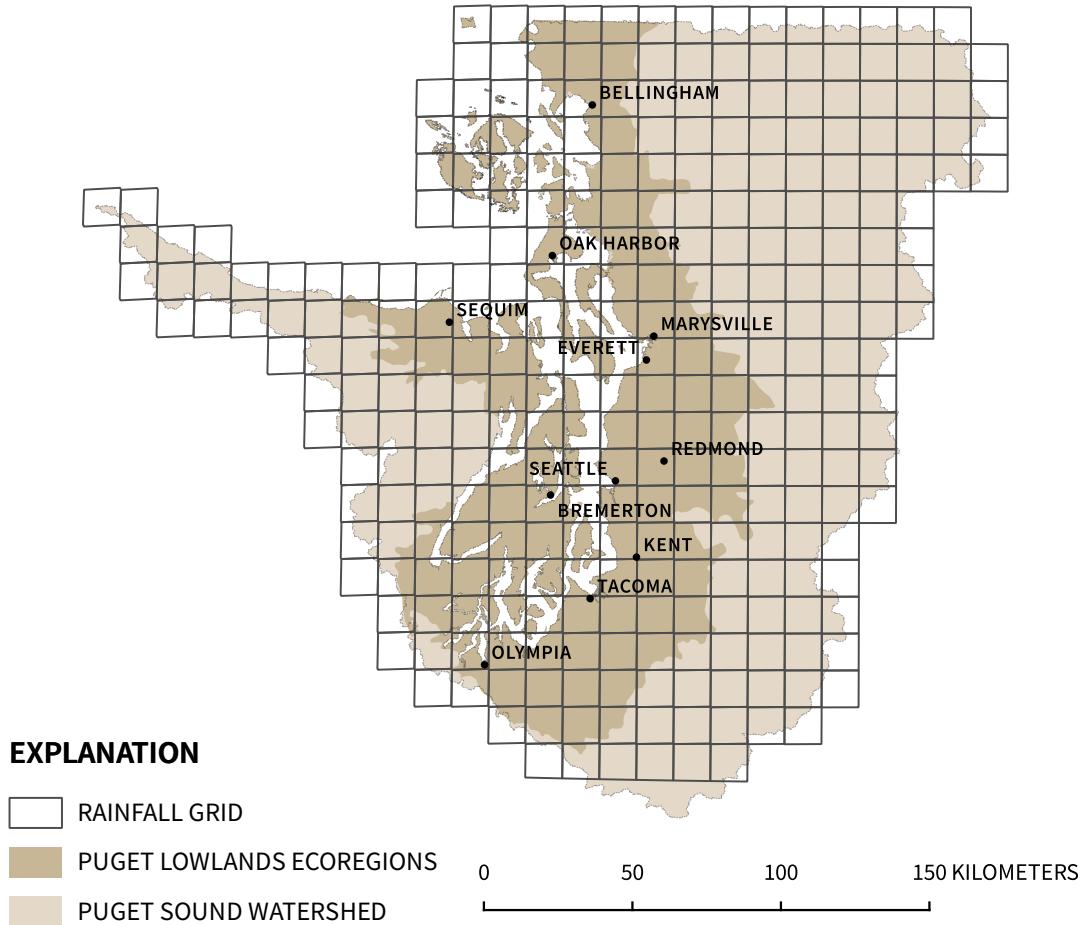


Figure 3.2: WRF Forecasting Grid

Services Center (GES DISC) 2019). This dataset combines multiple sources of observations to produce estimates of surface climate variables. Evaporation data was derived from the NCEP North American Regional Reanalysis, consisting of a retrospective dataset beginning January 1979 through December 2005. Data were acquired in degree grid spacing; at an hourly temporal resolution. Average monthly potential evaporation rates were calculated and resampled for each grid cell in the heatmap model domain.

### 3.3.2.1 Land Cover

Land cover was derived from the Nature Conservancy's high-resolution land cover data set. See Section 2 for details on land cover derivation.

Land cover values were remapped to equivalent HSPF land cover classes as shown below.

Derived Land Cover	HSPF Land cover class
Fine Vegetation	Grass
Medium Vegetation	Grass
Coarse Vegetation	Forest
Dirt/Barren	Grass
Water	Water
Impervious Other	Impervious
Impervious Roofs <sup>2</sup>	Impervious
NLCD Cropland	Pasture

Pasture landcover was derived from the US National Landcover Database (Yang et al. 2018) in areas outside of urban growth area boundaries.

### 3.3.2.2 Soils

**3.3.2.2.1 Gridded SSURGO Data** The primary source of soils data was the Gridded Soil Survey Geographic Database (gSSURGO), (Soil Survey Staff 2018). The gridded soils database contains 10-meter rasterized coverage of surface soils derived from National Cooperative Soil Survey (NCSS) maps. These maps are generally drawn at 1:24000 scale. NCSS designates soils by a “map-unit name,” which can be joined with other attribute data. Map units in the study area were joined with the soils component table, containing hydrologic-soil group designations. NCSS classifies hydrologic soil groups according to estimates of runoff potential. Soils are assigned to four groups (A, B, C, and D) and three dual classes (A/D, B/D, and C/D) as defined below:

- **Group A.** Soils having a high infiltration rate (low runoff potential) when thoroughly wet. These consist mainly of deep, well drained to excessively drained sands or gravelly sands. These soils have a high rate of water transmission.

---

<sup>2</sup>Roofs were designated impervious/flat slope

- **Group B.** Soils having a moderate infiltration rate when thoroughly wet. These consist chiefly of moderately deep or deep, moderately well drained or well drained soils that have moderately fine texture to moderately coarse texture. These soils have a moderate rate of water transmission.
- **Group C.** Soils having a slow infiltration rate when thoroughly wet. These consist chiefly of soils having a layer that impedes the downward movement of water or soils of moderately fine texture or fine texture. These soils have a slow rate of water transmission.
- **Group D.** Soils having a very slow infiltration rate (high runoff potential) when thoroughly wet. These consist chiefly of clays that have a high shrink-swell potential, soils that have a high water table, soils that have a claypan or clay layer at or near the surface, and soils that are shallow over nearly impervious material. These soils have a very slow rate of water transmission.

If a soil is assigned to a dual hydrologic group (A/D, B/D, or C/D), the first letter is for drained areas and the second is for undrained areas. Only the soils that in their natural condition are in group D are assigned to dual classes. In certain locations, data were augmented with the SSURGO Value added tables (Soil Survey Staff 2016) using the Potential wetland soil landscapes field.

**3.3.2.2.2 Oak Ridge National Laboratory HYSOGs250m** In areas where gSSURGO data were not available, we used the Global Hydrologic Soil Groups (HYSOGs250m) for Curve Number-Based Runoff Modeling developed by Oak Ridge National Laboratory (Ross, C.W., L. Prihodko, J.Y. Anchang, S.S. Kumar, W. Ji 2018). This dataset contains world-wide hydrologic soils groups derived at a 250 meter resolution from machine learning predictions. Hydrologic soil groups were given the same designation as the SSURGO data above.

**3.3.2.2.3 GAP/LANDFIRE DATA** To account for wetlands and saturated soils not included in the above datasets, we used the USGS GAP/LANDFIRE National Terrestrial Ecosystems data set, which includes nationwide vegetation and land cover data.

### 3.3.2.3 Slope

Slope values were calculated from the USGS National Elevation Dataset. Elevations were provided in 1/3 arc-second resolution (approximately 10-meters). Slope was calculated and classified into the following categories, consistent with Ecology guidance:

- \* Flat: < 5%
- \* Moderate: 5-15%
- \* Steep: > 15%

## 3.4 Verification of Results

Results were verified by comparing simulations to measured streamflow for a gaged watershed in King County. King County operates a stream gage on Madsen Creek, near Renton. The watershed above the gage site is approximately 2,000 acres, with about 25% imperviousness.

Daily streamflow data for the Madsen Creek watershed was provided by King County<sup>3</sup> for the period 1991-2010. We delineated the watershed above the gaging site using the USGS NHDPLus flow-conditioned raster (Moore et al. 2019). Using this watershed boundary, we extracted HRUs and associated areas from the stormwater heatmap HRU layer on Google Earth Engine. HRU results and areas are shown in Table 3.4.

Summary of HRUs and areas in Madsen Creek Watershed

hruName

sq.m

acres

soil

land cover

slope

hru000

2432168.00

601.001891

A/B

forest

flat

hru001

1228910.00

303.670347

A/B

forest

moderate

hru002

743614.00

183.751019

A/B

forest

steep

hru010

556435.10

---

<sup>3</sup>[https://green2.kingcounty.gov/hydrology/SummaryDataGraphs.aspx?G\\_ID=98](https://green2.kingcounty.gov/hydrology/SummaryDataGraphs.aspx?G_ID=98)

137.498118

A/B

pasture

flat

hru011

153862.00

38.020137

A/B

pasture

moderate

hru012

24682.16

6.099095

A/B

pasture

steep

hru020

909390.80

224.715351

A/B

lawn

flat

hru021

230017.00

56.838437

A/B

lawn

moderate

hru022

17757.89

4.388070

A/B  
lawn  
steep  
hru100  
93139.05  
23.015160  
C  
forest  
flat  
hru101  
9888.50  
2.443502  
C  
forest  
moderate  
hru102  
6064.56  
1.498585  
C  
forest  
steep  
hru110  
13569.95  
3.353207  
C  
pasture  
flat  
hru111  
2032.37  
0.502208  
C

pasture

moderate

hru112

230.02

0.056840

C

pasture

steep

hru120

118094.30

29.181745

C

lawn

flat

hru121

7885.95

1.948661

C

lawn

moderate

hru122

243.56

0.060185

C

lawn

steep

hru200

90542.94

22.373647

D

forest

flat  
hru201  
3724.20  
0.920269  
D  
forest  
moderate  
hru210  
22074.19  
5.454652  
D  
pasture  
flat  
hru211  
1212.49  
0.299612  
D  
pasture  
moderate  
hru220  
42232.11  
10.435781  
D  
lawn  
flat  
hru221  
2335.62  
0.577145  
D  
lawn  
moderate

hru250  
1477833.00  
365.180393  
D  
impervious  
flat  
hru251  
320216.80  
79.127303  
D  
impervious  
moderate  
hru252  
25874.91  
6.393828  
D  
impervious  
steep

Modeling results were then queried and aggregated from the BigQuery dataset as described in Appendix B. The same HRU values were also run in WWHM for comparison. Both the WWHM and BigQuery results were truncated to have the same period of record as the streamflow data. Only the surface runoff and interflow components were used in this analysis.

Figure 3.3 shows a comparison of the observed and simulation flow-durations for the Madsen Creek watershed.

Both the WWHM and PyHSPF results underpredict actual streamflow primarily because baseflow was not simulated. This is expected, since both models exclude groundwater contributions. However, the results show good agreement between both simulated datasets over the full duration of simulations. Note that the simulations use different precipitation (see Figure ?? ) datasets and are not expected to

### Flow Duration Curve– Madsen Creek, King County

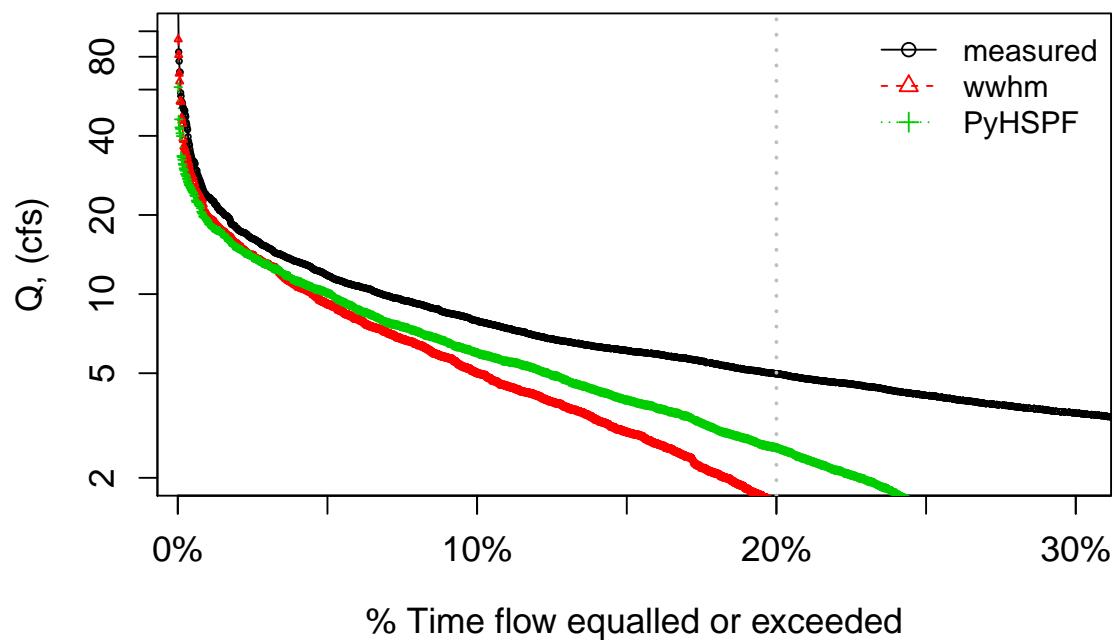
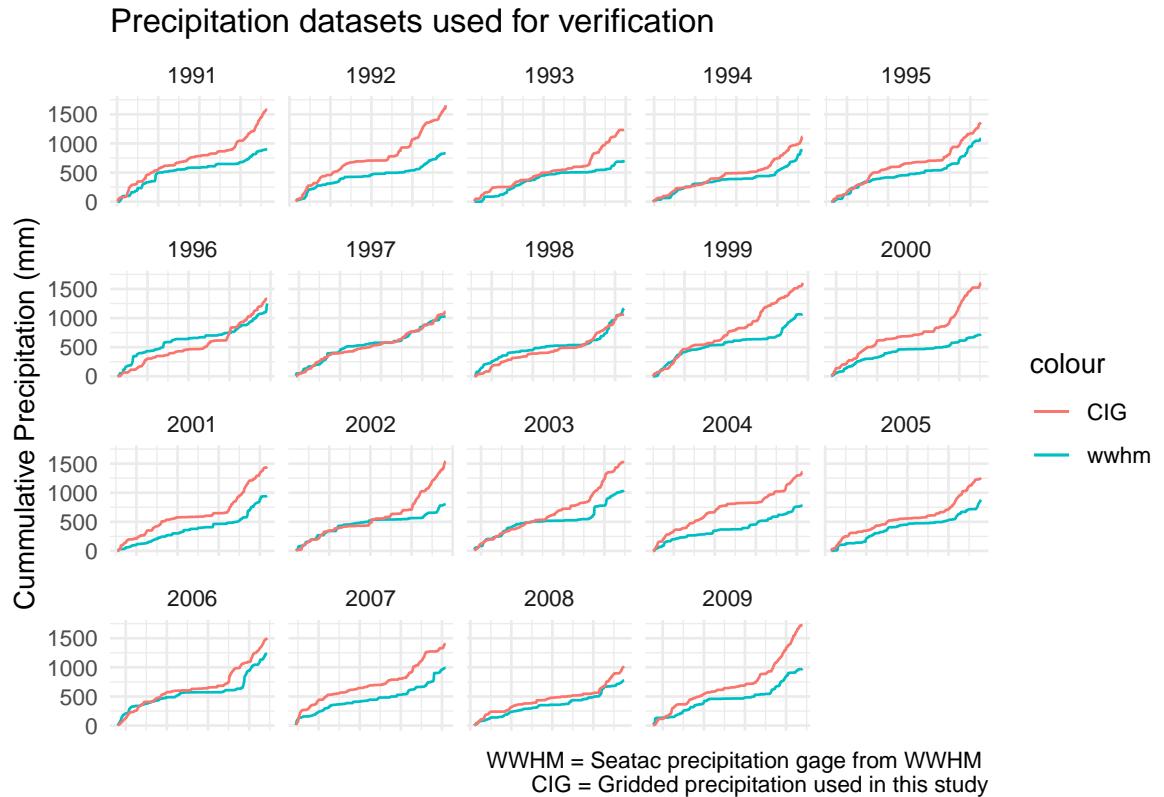


Figure 3.3: Observed and simulated flow-duration curves for Madsen Creek, King County, WA



match.

### 3.5 Spatially Aggregated Results

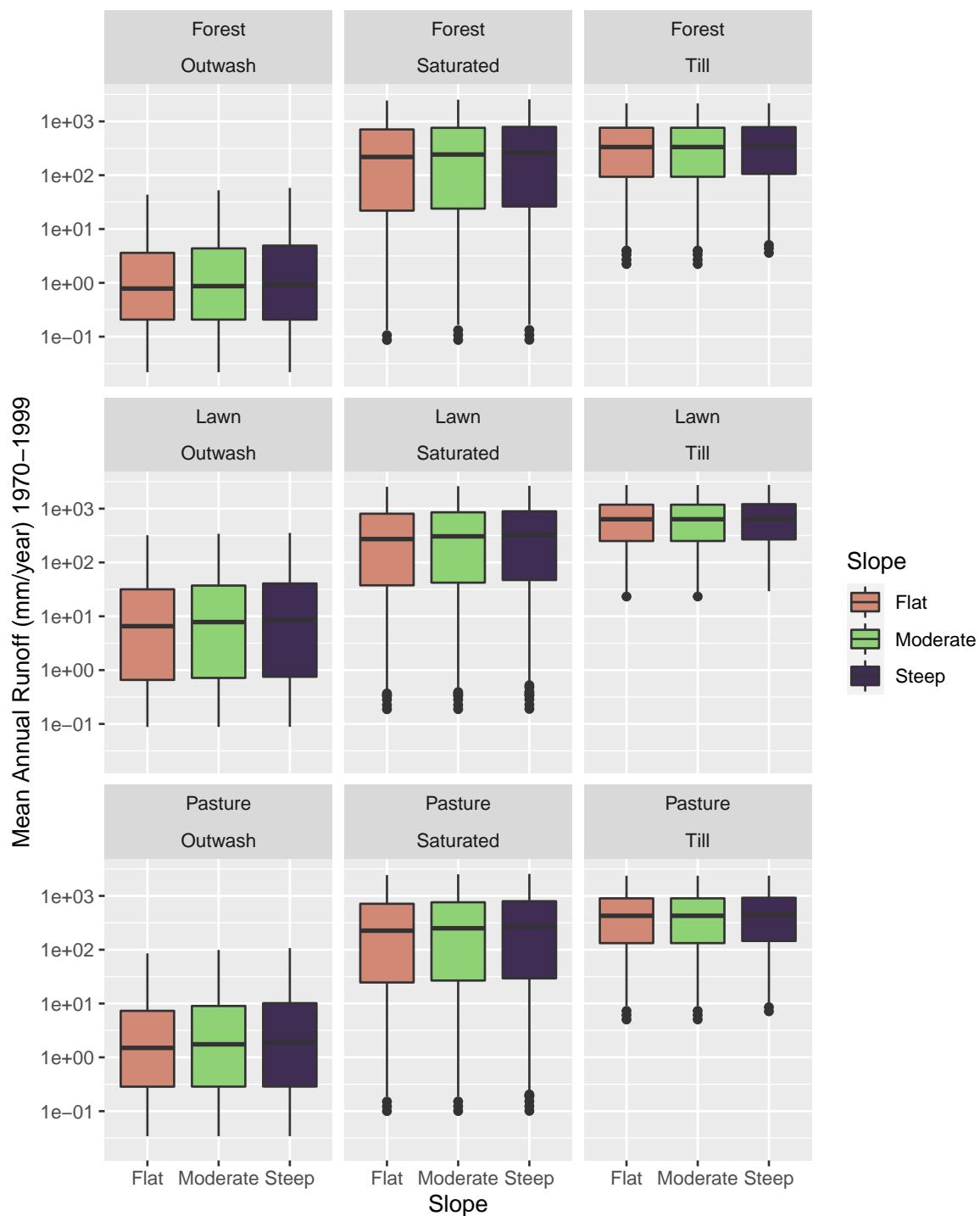
Since the PyHSPF model is a lumped parameter model, results can be calculated for HRU/precipitation grids individually and then aggregated after calculation.

The stormwater heatmap contains two spatial aggregates of hydrology results: Mean Annual Runoff for the historic period (1970-1999) and a new index, termed the Flow Duration Index.

#### 3.5.1 Mean Annual Runoff (1970-1999)

Mean annual runoff for each HRU/grid combination was aggregated from BigQuery for the historic period of record (1970-1999). Consistent with Ecology guidance for stormwater projects, only the surface flow components, **SUR0** and **IFWO** were used. **AGWO**, deep groundwater flow, was not included in this calculation.

Total runoff was calculated for each year/hru/grid combination in the period of record, then averaged by hru/grid combination.



### 3.5.2 Flow Duration Index

#### 3.5.2.1 Ecology Performance Standards

Ecology Stormwater Guidance includes flow-related performance standards to protect receiving waters from degradation caused by changes in the hydrologic regime due to development. These performance standards rely on flow-duration matching, whereby flow durations from developed land are required to match pre-developed flow-durations for a range of discharge values. The flow duration standard is intended to prevent flashy flows in receiving stream channels.

#### 3.5.2.2 Calculation of the Index

We developed an index representing the magnitude of change to the flow-duration curve between flow thresholds. Thresholds were chosen based on Ecology's LID and Flow Control Standards (Department of Ecology 2014), which require flow-duration matching over the range between 8 percent of the 2-year peak discharge (lower threshold of the LID standard) up to the 50-year peak discharge (upper threshold of the flow-control standard).

The flow discharge index is calculated by summing the discharge over the simulation period between a high-flow and low-flow threshold. Figure 3.4 illustrates the summation of flow-duration values used in calculating this index.

The flow duration index can be described by Equation (3.5.2.2).

$$\ln \left( \frac{\sum q_{current} \Delta t}{\sum q_{forest} \Delta t} + 1 \right) \text{ for: } \{ 0.06 \cdot Q_{2,forest} \leq q \leq Q_{50,forest} \}$$

Where  $q_{current}$  is the simulated discharge for current or altered conditions and  $q_{forest}$  is the predevelopment or forested conditions. One is added to this ratio and the logarithm is taken to produce an index that generally falls between 1 and 10. This index is then applied to hru/grid combinations in the stormwater heatmap to produce a spatially explicit mapping of flow alteration. Figure 3.5 shows a summary of flow index values used in the stormwater heatmap.

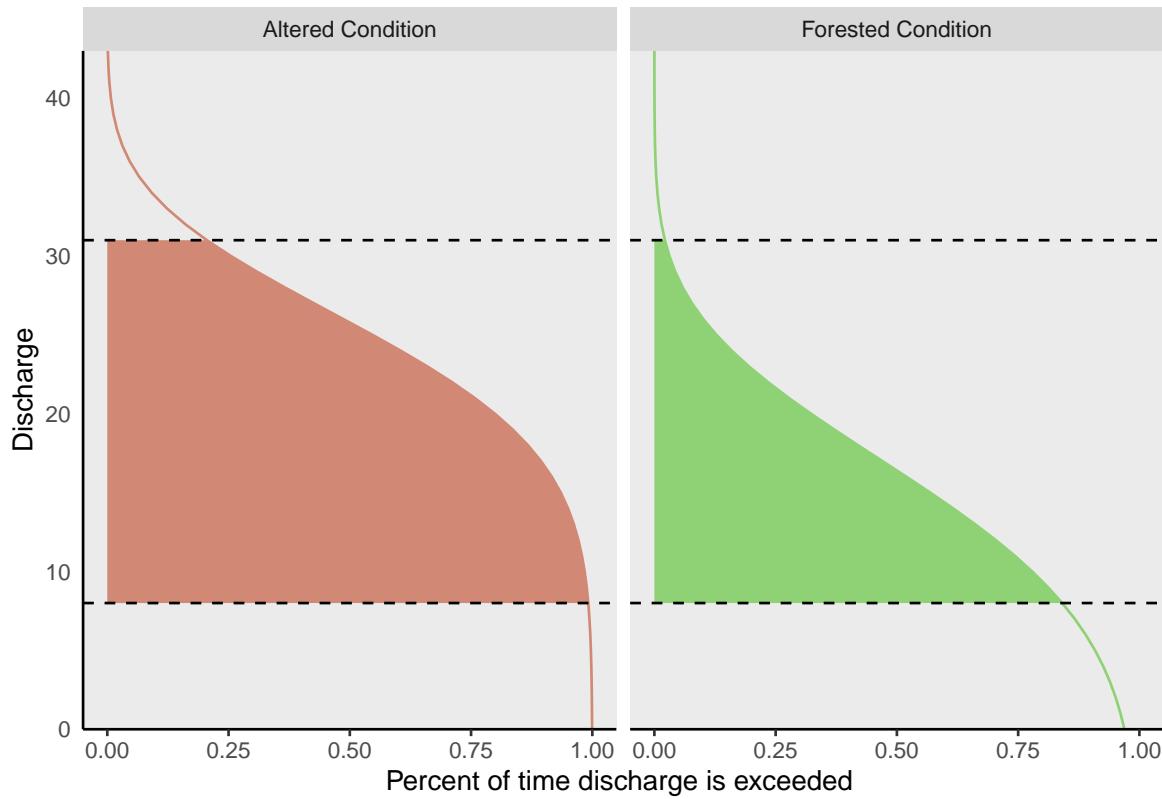


Figure 3.4: Example flow duration curves of altered and forested land covers

### Histogram of Flow Duration Index – All Puget Sound Locations 1970–1999 Modeling Results

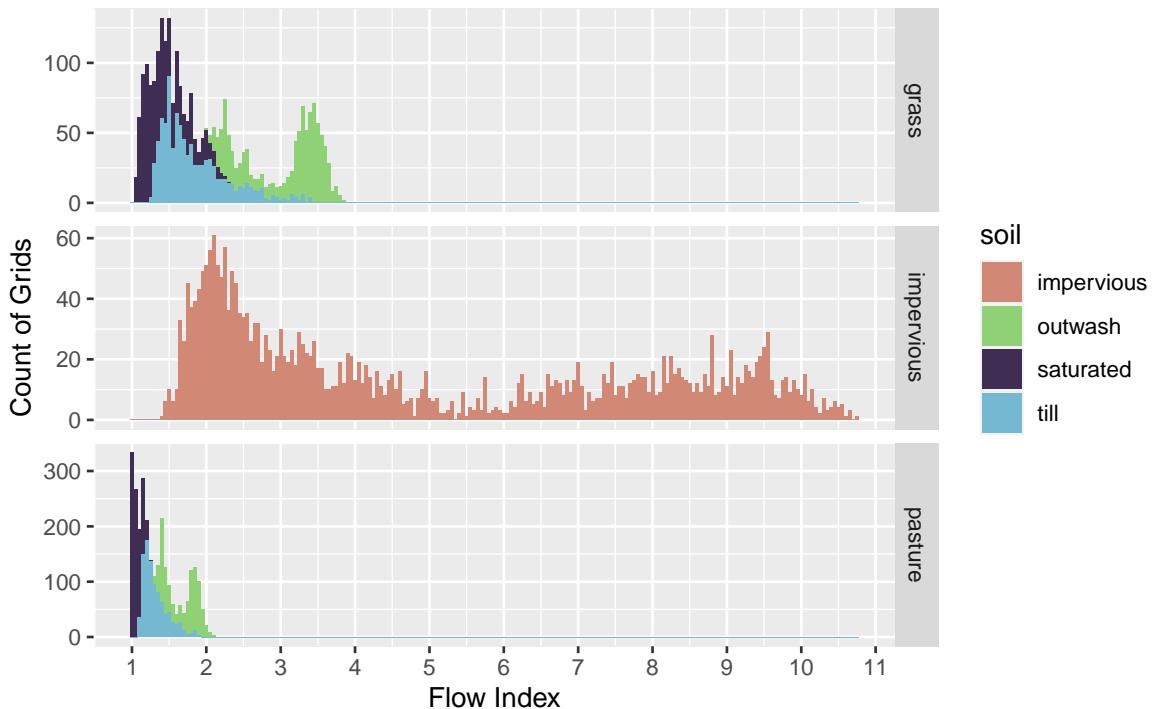


Figure 3.5: Summary of flow index values in study area

# Chapter 4

## Water Quality Statistics

We developed a spatial regression model to estimate concentrations for constituents of concern (COCs) in Puget Sound urban stormwater. We first used a linear mixed model to select spatial regression parameters. We then used a censored Markov chain Monte Carlo simulation to perform final regressions.

### 4.1 Data Sources

#### 4.1.1 Outfall Data

The primary source of measured stormwater data is the S8.D Municipal Stormwater Permit Outfall Data (referred to as the S8 Data in this document) provided by the Washington Department of Ecology (William Hobbs et al. 2015). Special Condition S8.D of the 2007-2012 Phase I Municipal Stormwater Permit required permittees to collect and analyze data to evaluate pollutant loadings of stormwater discharged from different land uses: high density (HD) residential, low density (LD) residential, commercial, and industrial. Phase I Permittees<sup>1</sup> collected water quality and flow data, sediment data, and toxicity information from stormwater discharges during storm events.

The stormwater outfall data is available from Ecology via an open-data api at: <https://data.wa.gov/Natural-Resources-Environment/Municipal-Stormwater-Permit-Outfall-Data/d958-q2ci>.

COCs analyzed in this study are:

- Zinc - Total
- Copper - Total
- Nitrite-Nitrate - Dissolved

---

<sup>1</sup>Cities of Tacoma and Seattle; King, Snohomish, Pierce and Clark counties; Ports of Tacoma and Seattle

- Lead - Total
- Total Phosphorus - Total
- Total Suspended Solids
- Total Phthalate
- Total Polycyclic Aromatic Hydrocarbons (PAH)
- Total Carcinogenic PAH (CPAH)
- Total High molecular weight PAH (HPAH)

We extracted data for these COCs, and performed minimal data cleaning. We filtered out rejected data (values with a R or REJ flag), removed replicates, and removed three data points that were obvious outliers. While our analysis is not overly sensitive to outliers, three parameters had reported data that were orders of magnitude higher than the rest of the data. One high-outlier value was removed for each of the following COCs: Total Suspended Solids, Nitrite-Nitrate, and Total Phosphorus.

Outliers were removed for Figure 4.1 shows data before outliers were removed. Figure 4.3 shows data after outliers were removed.

Quantile-quantile plots of COCs analyzed are shown in 4.2

## 4.2 Spatial data

For this study, we did not rely on the permittee's self-reported land use type to run regression models predicting pollution loading from land use. A visual scan of our land cover data layer versus self-reported land use types revealed little agreement among permittee definitions of the four land use types (high density residential, low density residential, commercial, industrial). Therefore, we compiled a suite of continuous and categorical landscape datasets from which to run prediction loading models. We divide these into land use and landscape data.

### 4.2.1 Land use

In order to employ a consistent analysis across different monitored watersheds we extracted land use data from Ecology's 2010 Statewide Land use data set<sup>2</sup>. Ecology generated the coverage from digital county tax parcel layers using Department of Revenue (DOR) two digit land use codes (see; WAC 458-53-030, Stratification of assessment rolls - real property).

### 4.2.2 Landscape data

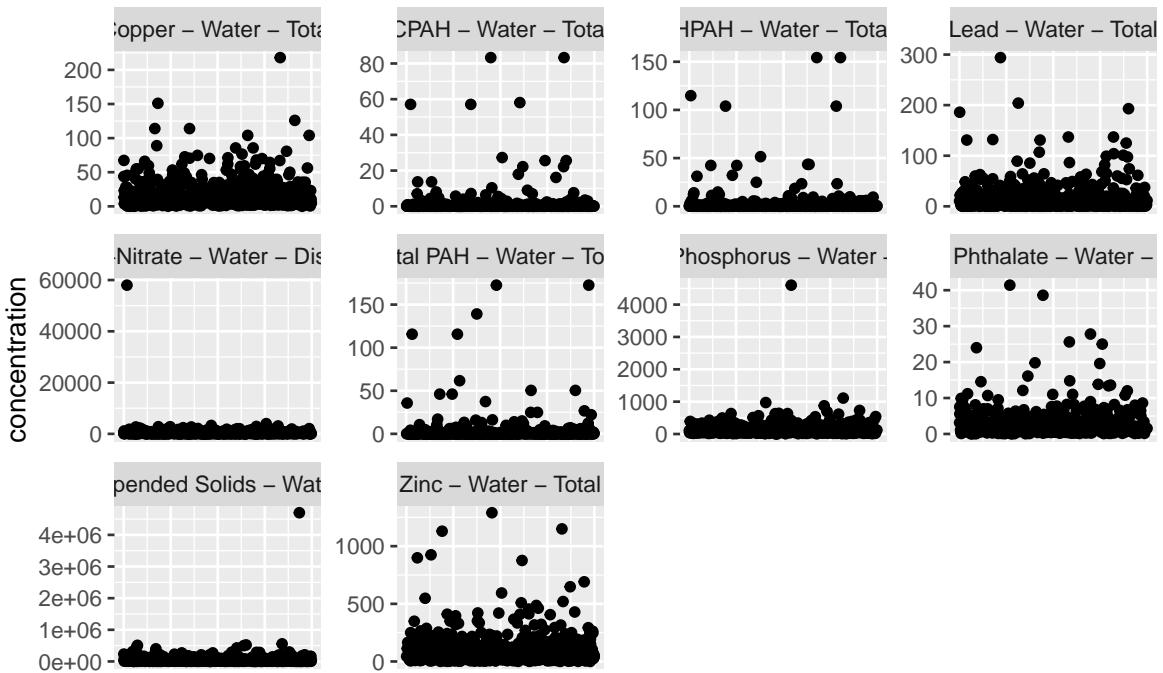
For each watershed contained in the S8 dataset, potentially relevant landscape data was extracted from the following sources below:

---

<sup>2</sup>See: [https://fortress.wa.gov/ecy/gispublic/DataDownload/ECY\\_CAD\\_Landuse2010.htm](https://fortress.wa.gov/ecy/gispublic/DataDownload/ECY_CAD_Landuse2010.htm) for more information

## All Observations

Data collected 2009–2013



Data source: Ecology, 2015

Figure 4.1: All observations - outliers in place

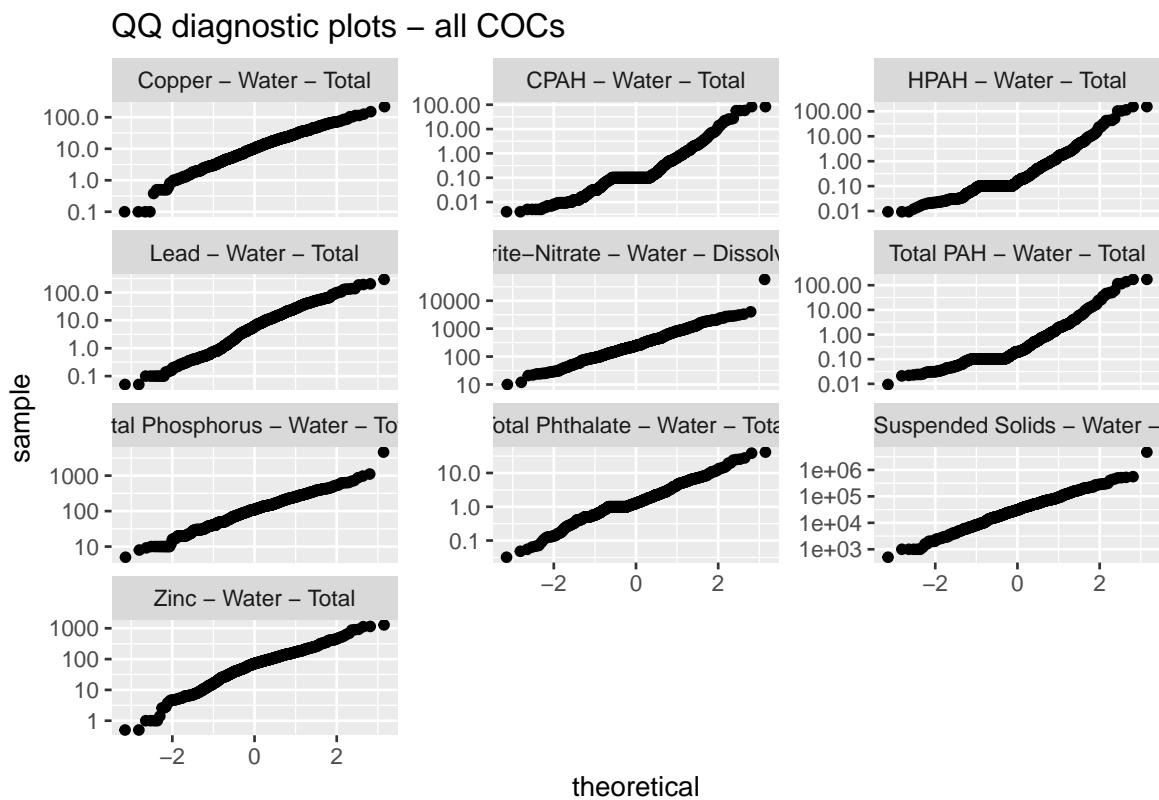
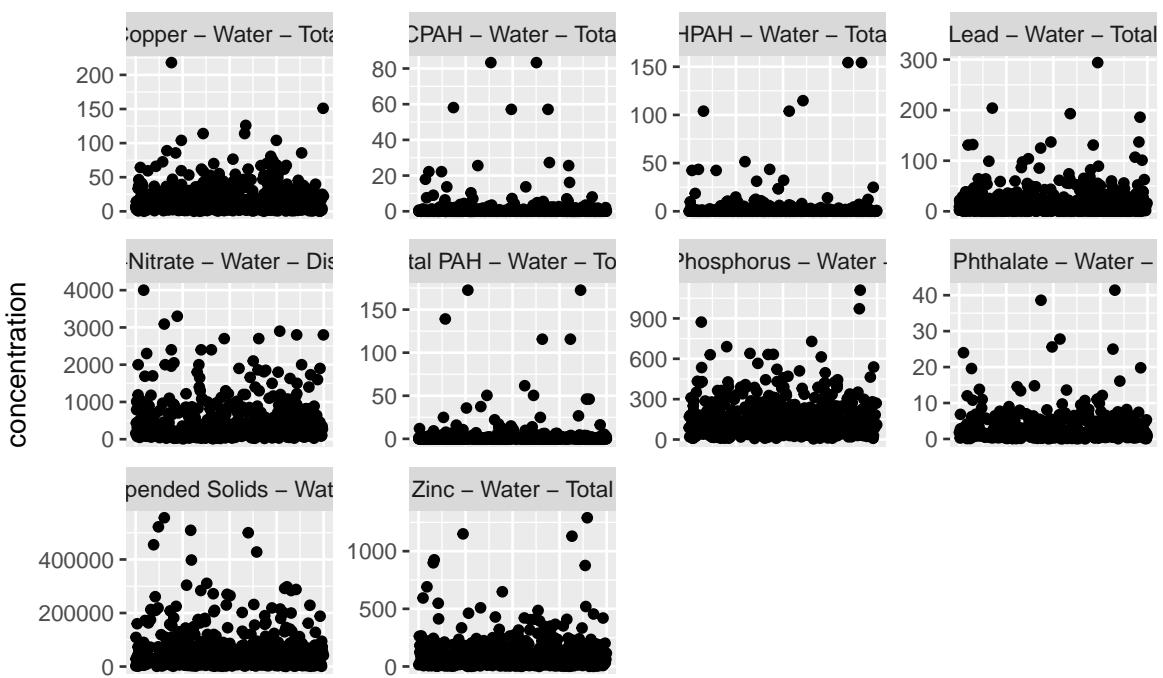


Figure 4.2: Quantile-quantile plots of COCs, log scale

## All Observations – Outliers Removed

Data collected 2009–2013



Data source: Ecology, 2015

Figure 4.3: All observations - outliers removed

Layer	ID	Source
Nighttime Lights	nighttime_lights	Global Radiance Calibrated Nighttime Lights van Donkelaar et al. 2018.
Particulate Matter 2.5 m	pm25	
Rooftop Density	roofs	Microsoft AI US Building Footprints
Imperviousness	impervious	TNC Puget Sound land cover
Age of Impervious Surface	change_year_index	Tsinghua Univer- sity FROM-GLC year of change to impervious surface
Logarithm of Population Density	logPopulation	CIESIN - Columbia University Gridded Population of the World, Version 4
Logarithm of Average Daily Traffic Volume	logTraffic	INRIX Traffic

## 4.3 Methods

### 4.3.1 Pre-processing of spatial data

In order to use the landscape data at an appropriate scale across the study area, spatial predictors were stacked and then convolved with a 100-meter gaussian kernel. This resulted in a “fuzzy” set of predictors that could apply across dataset boundaries. These values were then extracted for each monitored watershed. Values were scaled and centered for regression purposes.

### 4.3.2 Controlling for multicollinearity

To address multicollinearity, we calculate the variance inflation factor (VIF) and iteratively remove parameters with the highest VIF. We keep removing parameters one at a time until all VIF values are below 10.0. Table ?? shows final VIF factors for Zinc. Similar values were found for all other COCs and are not reported here.

Variance inflation factors Zinc - Water - Total

vif

change_year_index	
3.41	
nighttime_lights	
3.27	
logPopulation	
2.00	
logtraffic	
1.92	
LU_ratio	
1.11	

#### 4.3.3 Censored Data

All COCs had non-detect (left-censored) data present. Ecology flagged non-detect data and provided the reporting limit for each non-detect value. For purposes of model selection, non-detect values were substituted with the reporting limit. For regression, concentration was modeled as a two-parameter response variable according to the approaches detailed by Hadfield (2010) and Helsel (2012). We set the first parameter equal to the measured concentration and the second parameter equal to the reporting limit. We then specified a censored-gaussian prior distribution or censored-log-gaussian prior distribution.

#### 4.3.4 Model Selection

Four potential models were evaluated for each COC:

1. **linear** - Non-transformed concentrations with location random effects
2. **log-linear** - Log-transformed concentrations with location random effects
3. **linear seasonal** - Non-transformed concentrations with location random effects with the addition of a seasonal fixed-effect
4. **log-linear seasonal** - Log-transformed concentrations with location random effects with the addition of a seasonal fixed-effect

Seasonal factors were included in the model selection by designating an integer corresponding to the season in which data were collected: (*1 = Winter, 2 = Spring, 3 = Summer, 4 = Autumn*)

Model selection was performed through Backward elimination of random-effect terms followed by backward elimination of fixed-effect terms. Denominator degrees of freedom and F-statistics were calculated using Satterthwaite's method.

For each model we calculated the Akaike information criterion (AIC) estimator. The model with the lowest AIC was selected for regression.

Linear model results for each COC are shown in the following tables. In general, the model with the lowest AIC was selected to move on to the Bayesian analysis. Only predictors that are statistically

significant are shown. In some cases, no predictors were significant. In those cases, only the model intercept is reported.

#### 4.3.4.1 Zinc

**Model with lowest AIC:** log-linear  
**Predictors Selected:** change\_year\_index

Table 4.2: Zinc - Water - Total

	Dependent variable:			
	concentration	log(concentration)	concentration	log(concentration)
	(1)	(2)	(3)	(4)
change_year_index	59.200*** (16.100)	1.010*** (0.189)		1.000*** (0.187)
nighttime_lights			57.100*** (13.300)	
season2			4.190 (15.400)	-0.063 (0.099)
season3			115.000*** (18.200)	0.388*** (0.116)
season4			8.810 (12.600)	0.002 (0.080)
Constant	107.000*** (16.600)	4.040*** (0.198)	109.000*** (18.000)	4.010*** (0.202)
Observations	414	414	414	414
Log Likelihood	-2,546.000	-456.000	-2,513.000	-453.000
Akaike Inf. Crit.	5,100.000	920.000	5,041.000	921.000
Bayesian Inf. Crit.	5,116.000	936.000	5,069.000	949.000

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

#### 4.3.4.2 Copper

**Model with lowest AIC:** log-linear  
**Predictors Selected:** logtraffic,change\_year\_index

Table 4.3: Copper - Water - Total

	<i>Dependent variable:</i>			
	concentration	log(concentration)	concentration	log(concentration)
	(1)	(2)	(3)	(4)
change_year_index		0.476*** (0.164)		0.476*** (0.164)
logtraffic	34.365*** (5.987)	1.865*** (0.530)	34.584*** (5.983)	1.865*** (0.530)
season2			−0.715 (1.132)	
season3			6.383*** (1.334)	
season4			−0.680 (0.942)	
Constant	67.255*** (9.628)	5.065*** (0.819)	67.290*** (9.633)	5.065*** (0.819)
Observations	438	438	438	438
Log Likelihood	−1,569.401	−464.499	−1,551.354	−464.499
Akaike Inf. Crit.	3,146.801	938.998	3,116.708	938.998
Bayesian Inf. Crit.	3,163.130	959.409	3,145.284	959.409

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

#### 4.3.4.3 Nitrite-Nitrate

**Model with lowest AIC:** log-linear seasonal **Predictors Selected:** LU\_ratio

Table 4.4: Nitrite-Nitrate - Water - Dissolved

	Dependent variable:			
	concentration	log(concentration)	concentration	log(concentration)
	(1)	(2)	(3)	(4)
LU_ratio	79.943** (35.237)	0.159** (0.070)	80.833** (34.971)	0.160** (0.068)
season2			-109.123* (62.255)	-0.015 (0.103)
season3			16.237 (74.908)	0.483*** (0.124)
season4			-165.247*** (50.621)	-0.370*** (0.084)
Constant	496.135*** (90.949)	5.719*** (0.180)	575.779*** (94.915)	5.806*** (0.183)
Observations	406	406	406	406
Log Likelihood	-3,047.002	-480.894	-3,025.528	-459.825
Akaike Inf. Crit.	6,102.003	969.788	6,065.056	933.650
Bayesian Inf. Crit.	6,118.028	985.813	6,093.100	961.695

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

#### 4.3.4.4 Lead

**Model with lowest AIC:** log-linear (*seasonal effects not significant at model selection*)  
**Predictors Selected:** change\_year\_index,

#### 4.3.4.5 Total Phosphorus

**Model with lowest AIC:** log-linear seasonal  
**Predictors Selected:** logtraffic

Table 4.5: Lead - Water - Total

	<i>Dependent variable:</i>			
	concentration	log(concentration)	concentration	log(concentration)
	(1)	(2)	(3)	(4)
change_year_index	4.257*** (1.465)	1.015*** (0.269)	4.257*** (1.465)	1.015*** (0.269)
logPopulation			-3.009** (1.266)	
logtraffic	13.085** (5.394)		13.085** (5.394)	
Constant	30.478*** (8.177)	1.600*** (0.283)	30.478*** (8.177)	1.600*** (0.283)
Observations	417	417	417	417
Log Likelihood	-1,613.139	-539.748	-1,613.139	-539.748
Akaike Inf. Crit.	3,238.277	1,087.496	3,238.277	1,087.496
Bayesian Inf. Crit.	3,262.476	1,103.628	3,262.476	1,103.628

Note:

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Table 4.6: Total Phosphorus - Water - Total

	<i>Dependent variable:</i>			
	concentration	log(concentration)	concentration	log(concentration)
	(1)	(2)	(3)	(4)
logtraffic	214.009*** (51.608)	1.694*** (0.424)	214.280*** (52.601)	1.681*** (0.424)
season2			−12.043 (13.300)	−0.107 (0.088)
season3			95.375*** (15.928)	0.771*** (0.105)
season4			8.698 (11.117)	0.145** (0.073)
Constant	478.137*** (82.938)	7.207*** (0.682)	468.292*** (84.736)	7.082*** (0.683)
Observations	424	424	424	424
Log Likelihood	−2,557.061	−450.868	−2,525.908	−424.755
Akaike Inf. Crit.	5,122.122	909.735	5,065.815	863.509
Bayesian Inf. Crit.	5,138.321	925.934	5,094.163	891.857

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

#### 4.3.4.6 Total Suspended Solids

**Model with lowest AIC:** log-linear (*seasonal effects not significant at model selection*)

**Predictors Selected:** logtraffic

Table 4.7: Total Suspended Solids - Water - Total

	<i>Dependent variable:</i>			
	concentration	log(concentration)	concentration	log(concentration)
	(1)	(2)	(3)	(4)
logtraffic	63,941.480*** (24,613.700)	1.946*** (0.492)	63,941.480*** (24,613.700)	1.946*** (0.492)
Constant	152,052.500*** (39,520.290)	13.198*** (0.790)	152,052.500*** (39,520.290)	13.198*** (0.790)
Observations	415	415	415	415
Log Likelihood	−5,125.929	−616.491	−5,125.929	−616.491
Akaike Inf. Crit.	10,259.860	1,240.983	10,259.860	1,240.983
Bayesian Inf. Crit.	10,275.970	1,257.096	10,275.970	1,257.096

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

#### 4.3.4.7 Total Phthalate

**Model with lowest AIC:** log-linear (*seasonal effects not significant at model selection*)

**Predictors Selected:** none

Table 4.8: Total Phthalate - Water - Total

	<i>Dependent variable:</i>			
	concentration	log(concentration)	concentration	log(concentration)
	(1)	(2)	(3)	(4)
Constant	2.896*** (0.568)	0.418** (0.208)	2.896*** (0.568)	0.418** (0.208)
Observations	412	412	412	412
Log Likelihood	−1,171.609	−503.981	−1,171.609	−503.981
Akaike Inf. Crit.	2,349.218	1,013.961	2,349.218	1,013.961
Bayesian Inf. Crit.	2,361.281	1,026.024	2,361.281	1,026.024

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

#### 4.3.4.8 Total PAH

**Model with lowest AIC:** log-linear  
**Predictors Selected:** LU\_ratio

Table 4.9: Total PAH - Water - Total

	<i>Dependent variable:</i>			
	concentration	log(concentration)	concentration	log(concentration)
	(1)	(2)	(3)	(4)
LU_ratio		−0.161** (0.070)		−0.156** (0.068)
season2				−0.316** (0.141)
season3				−0.305* (0.166)
season4				0.036 (0.115)
Constant	0.592*** (0.151)	−1.444*** (0.180)	0.592*** (0.151)	−1.357*** (0.189)
Observations	412	412	412	412
Log Likelihood	−633.518	−595.652	−633.518	−594.163
Akaike Inf. Crit.	1,273.036	1,199.303	1,273.036	1,202.326
Bayesian Inf. Crit.	1,285.099	1,215.387	1,285.099	1,230.474

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

#### 4.3.4.9 Total CPAH (*seasonal effects not significant at model selection*)

**Model with lowest AIC:** linear **Predictors Selected:** logPopulation, logtraffic

#### 4.3.4.10 Total HPAH

**Model with lowest AIC:** linear (*seasonal effects not significant at model selection*)  
**Predictors Selected:** none

Table 4.10: CPAH - Water - Total

	<i>Dependent variable:</i>			
	concentration	log(concentration)	concentration	log(concentration)
	(1)	(2)	(3)	(4)
logPopulation	0.176*** (0.056)		0.176*** (0.056)	
logtraffic	-0.640*** (0.215)		-0.640*** (0.215)	
LU_ratio		-0.188*** (0.057)		-0.188*** (0.057)
Constant	-0.689** (0.332)	-2.173*** (0.147)	-0.689** (0.332)	-2.173*** (0.147)
Observations	412	412	412	412
Log Likelihood	-377.401	-578.552	-377.401	-578.552
Akaike Inf. Crit.	764.803	1,165.104	764.803	1,165.104
Bayesian Inf. Crit.	784.908	1,181.188	784.908	1,181.188

Note:

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Table 4.11: HPAH - Water - Total

	<i>Dependent variable:</i>			
	concentration	log(concentration)	concentration	log(concentration)
	(1)	(2)	(3)	(4)
LU_ratio		-0.174*** (0.065)		-0.174*** (0.065)
Constant	0.509*** (0.140)	-1.608*** (0.170)	0.509*** (0.140)	-1.608*** (0.170)
Observations	412	412	412	412
Log Likelihood	-597.161	-599.328	-597.161	-599.328
Akaike Inf. Crit.	1,200.321	1,206.656	1,200.321	1,206.656
Bayesian Inf. Crit.	1,212.384	1,222.740	1,212.384	1,222.740

Note:

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

### 4.3.5 Bayesian Regression

We performed censored bayesian regression using a Markov chain Monte Carlo approach. We use the MCMCglmm package in R (Helsel 2012) for analysis.

Models were run for 60,000 iterations with a burn-in interval of 10,000 and a thinning interval of 13. For each regression model, we specified either a censored Gaussian or a censored log-Gaussian prior distribution, depending on the model selection results. We specified a simple prior co-variance matrix where covariances between predictors was fixed at 1. We explored other prior variances were not sensitive to these changes.

## 4.4 Results

### 4.4.1 Zinc

#### 4.4.1.1 Regression Coefficients

Regression results for Total Zinc are summarized in Table 4.4.1.1.

Zinc - Water - Total - Bayseian Regression Results

Predictor

Posterior Mean

lower 95% CI

upper 95% CI

MCMC p value

effect

(Intercept)

4.039

3.596

4.48

0

fixed

change\_year\_index

1.011

0.613

1.44

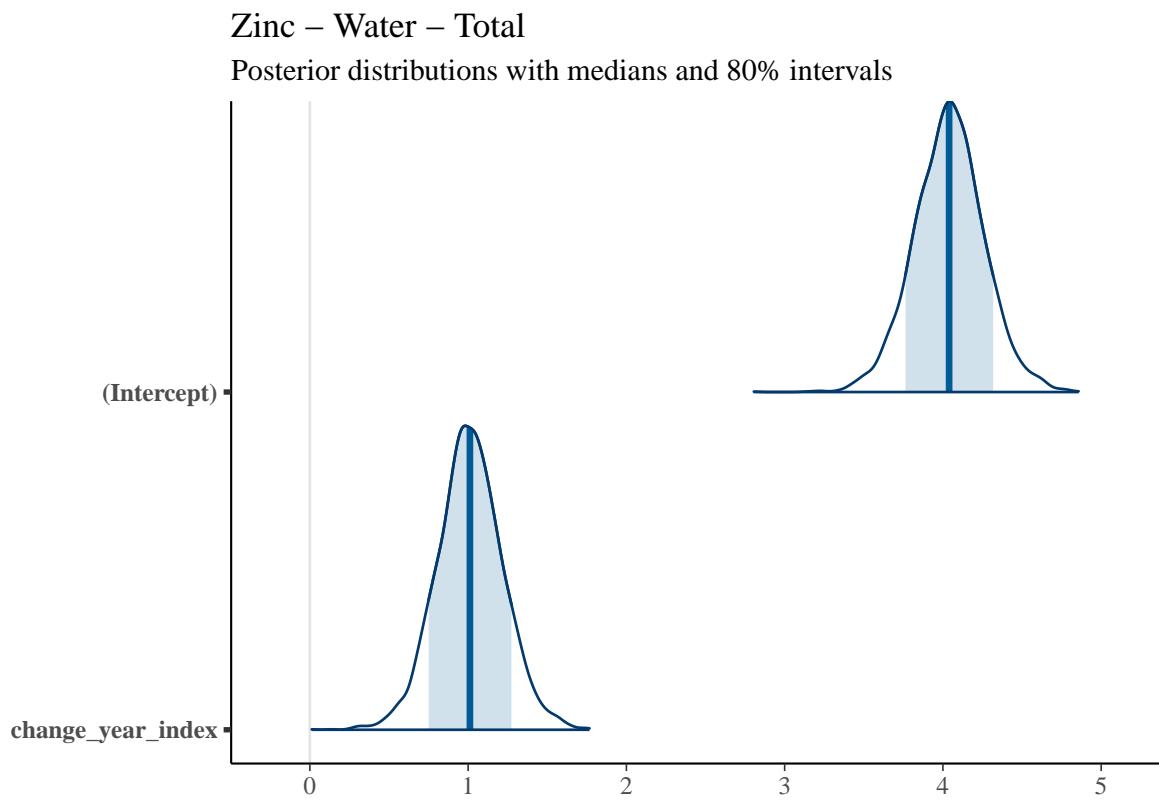
0

```
fixed  
Location  
0.551  
0.157  
1.10  
NA  
random  
units  
1.000  
1.000  
1.00  
NA  
residual
```

#### 4.4.1.2 Posterior Uncertainty Levels

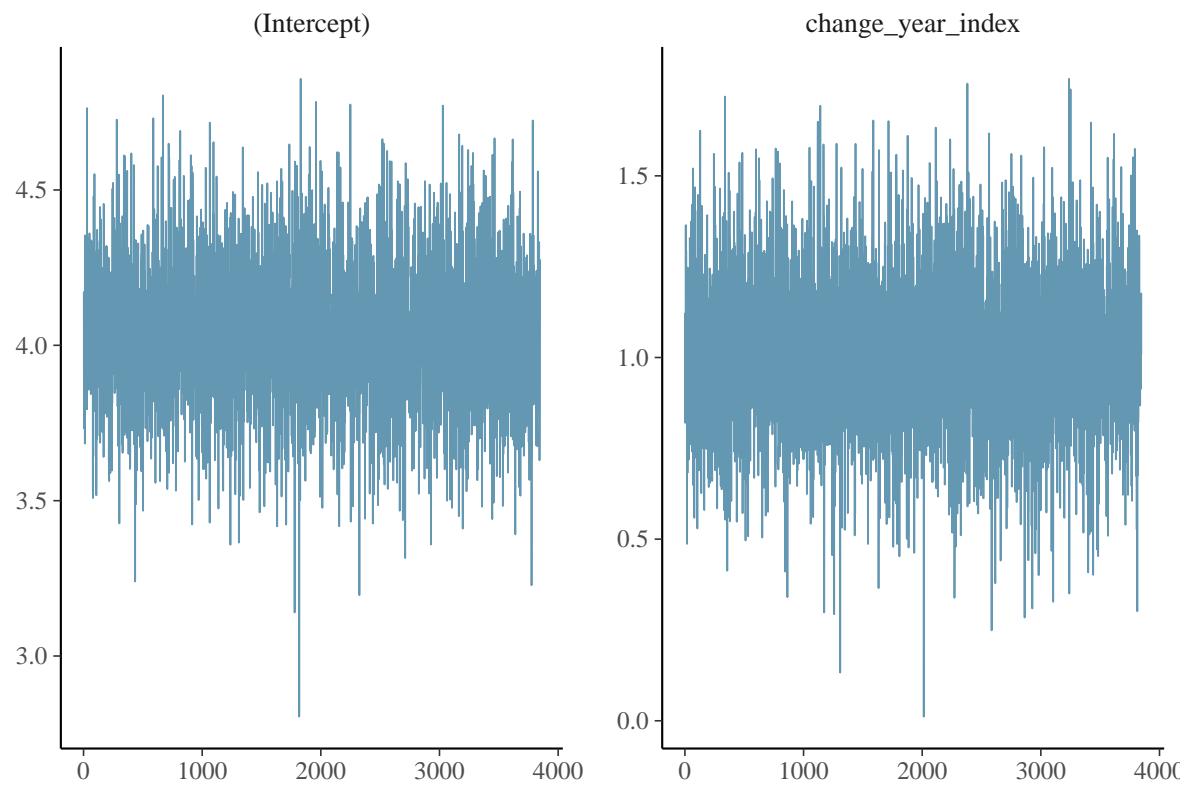
Estimated posterior density curves with 80% confidence intervals are shown in Figure ??.

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



#### 4.4.1.3 Model Diagnostics

Figure ?? shows the diagnostic trace plot of MCMC draws. This was used to verify prior distribution assumptions.



#### 4.4.2 Copper

##### 4.4.2.1 Regression Coefficients

Regression results for Total Zinc are summarized in Table 4.4.2.1.

Copper - Water - Total - Bayseian Regression Results

Predictor

Posterior Mean

lower 95% CI

upper 95% CI

MCMC p value

effect

(Intercept)

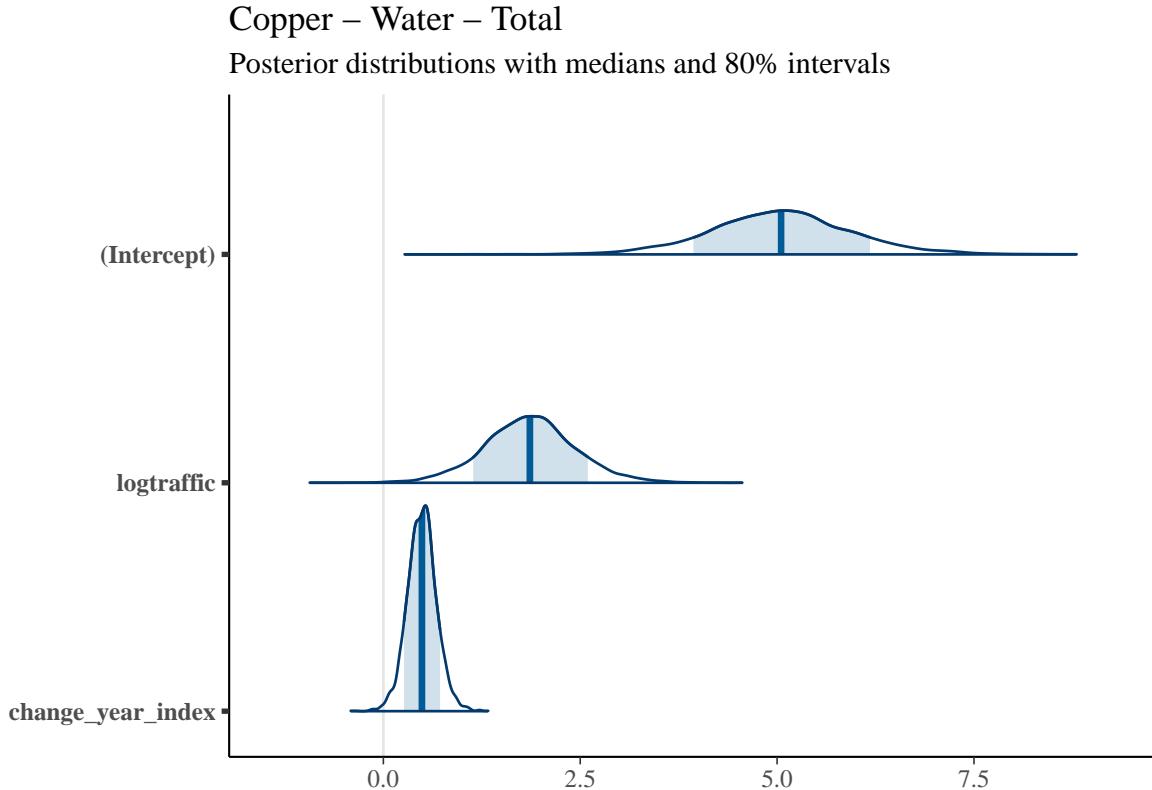
5.051

3.195  
6.805  
0.000  
fixed  
logtraffic  
1.860  
0.729  
3.082  
0.004  
fixed  
change\_year\_index  
0.490  
0.122  
0.864  
0.012  
fixed  
Location  
0.309  
0.082  
0.660  
NA  
random  
units  
1.000  
1.000  
1.000  
NA  
residual

#### 4.4.2.2 Posterior Uncertainty Levels

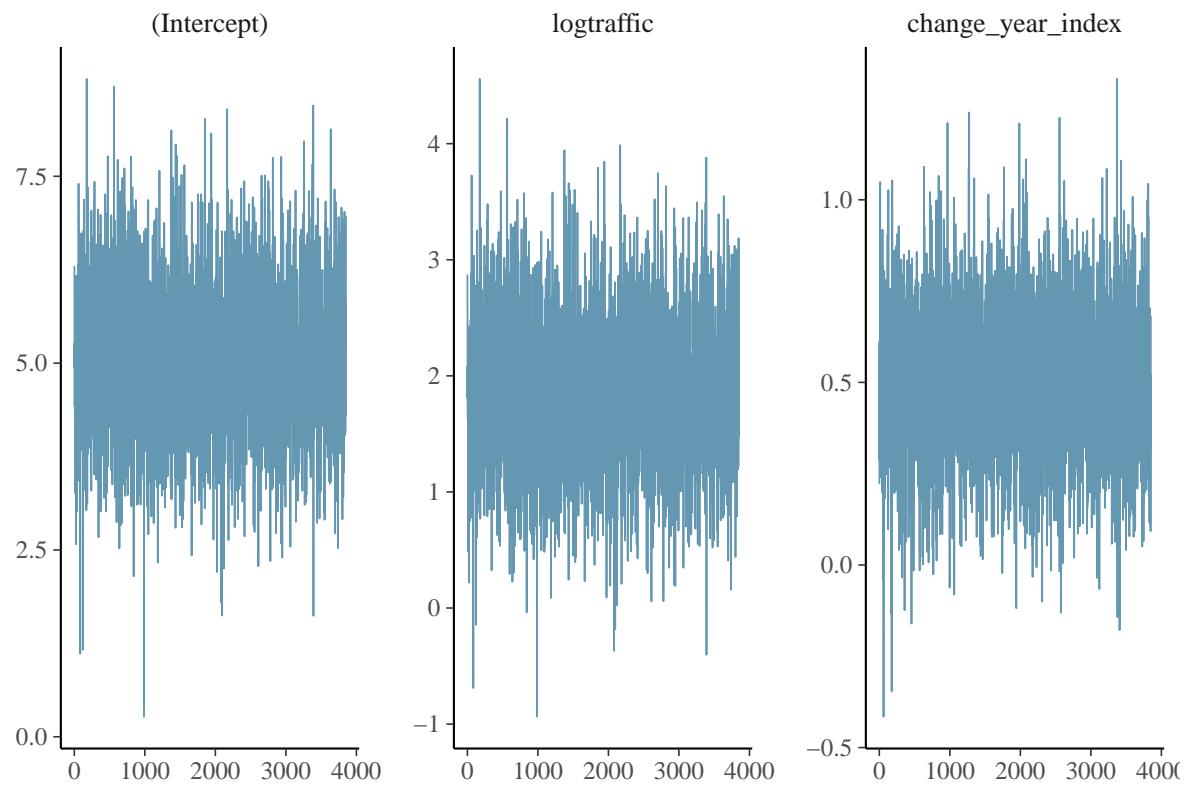
Estimated posterior density curves with 80% confidence intervals are shown in Figure ??.

```
## [1]
```



#### 4.4.2.3 Model Diagnostics

Figure ?? shows the diagnostic trace plot of MCMC draws. This was used to verify prior distribution assumptions.



#### 4.4.3 Nitrite-Nitrate

##### 4.4.3.1 Regression Coefficients

Regression results for Nitrite-Nitrate are summarized in Table 4.4.3.1.

Nitrite-Nitrate - Water - Dissolved - Bayseian Regression Results

Predictor

Posterior Mean

lower 95% CI

upper 95% CI

MCMC p value

effect

(Intercept)

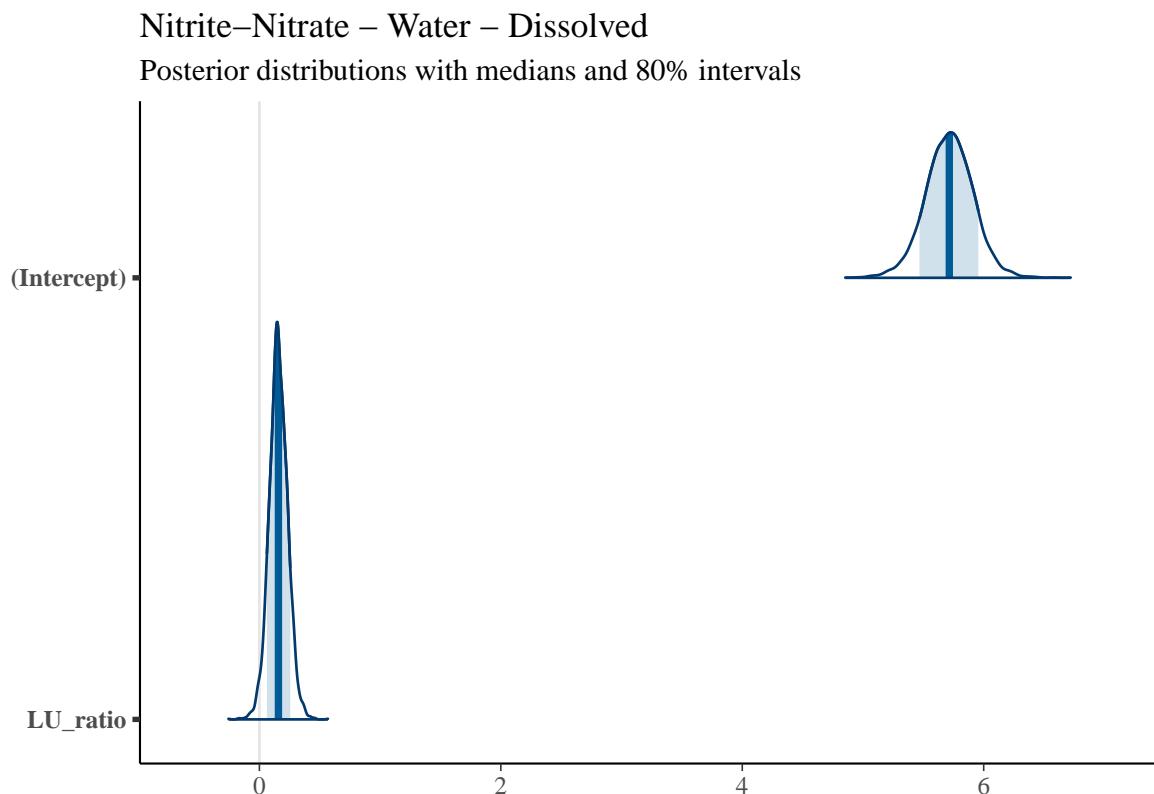
5.715

```
5.301  
6.087  
0.000  
fixed  
LU_ratio  
0.158  
-0.006  
0.302  
0.049  
fixed  
Location  
0.509  
0.157  
1.040  
NA  
random  
units  
1.000  
1.000  
1.000  
NA  
residual
```

#### 4.4.3.2 Posterior Uncertainty Levels

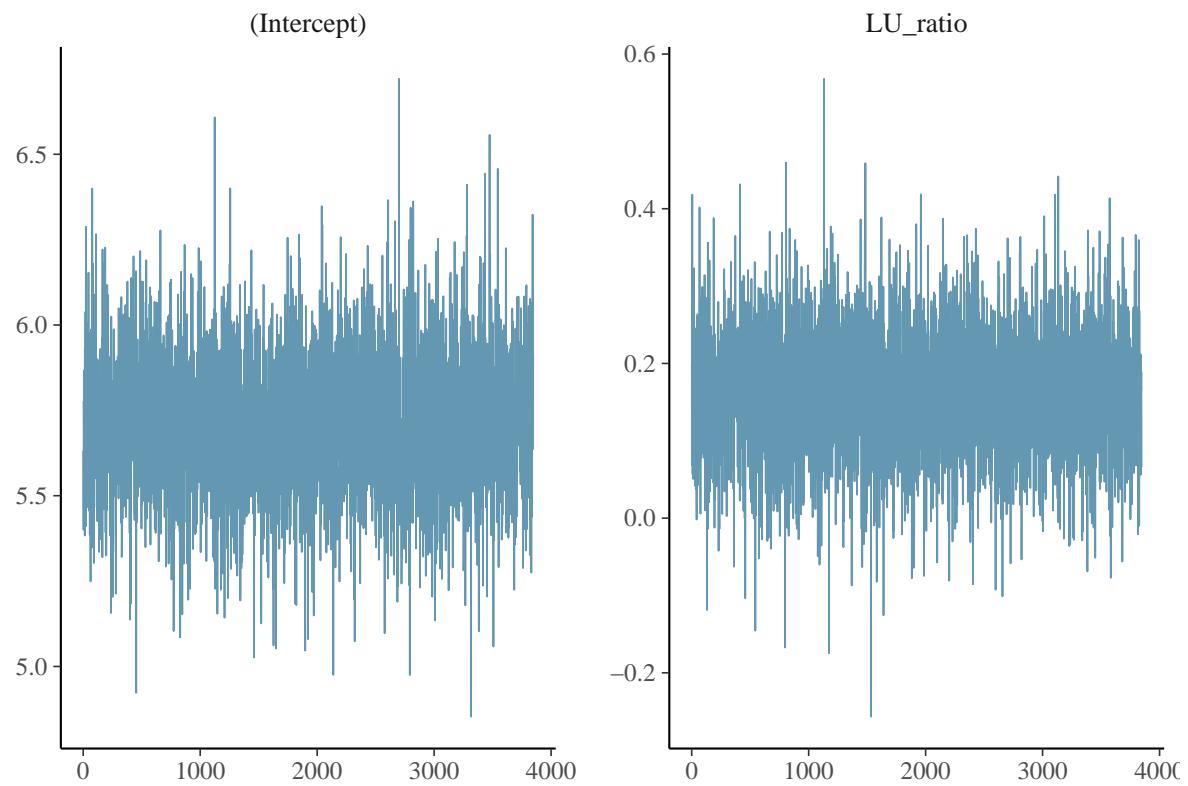
Estimated posterior density curves with 80% confidence intervals are shown in Figure ??.

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



#### 4.4.3.3 Model Diagnostics

Figure ?? shows the diagnostic trace plot of MCMC draws. This was used to verify prior distribution assumptions.



#### 4.4.4 Lead

##### 4.4.4.1 Regression Coefficients

Regression results for Total Lead are summarized in Table 4.4.4.1.

Lead - Water - Total - Bayseian Regression Results

Predictor

Posterior Mean

lower 95% CI

upper 95% CI

MCMC p value

effect

(Intercept)

1.314

```
0.718
1.98
0.002
fixed
impervious
0.873
0.257
1.50
0.009
fixed
Location
1.441
0.483
2.84
NA
random
units
1.000
1.000
1.00
NA
residual
```

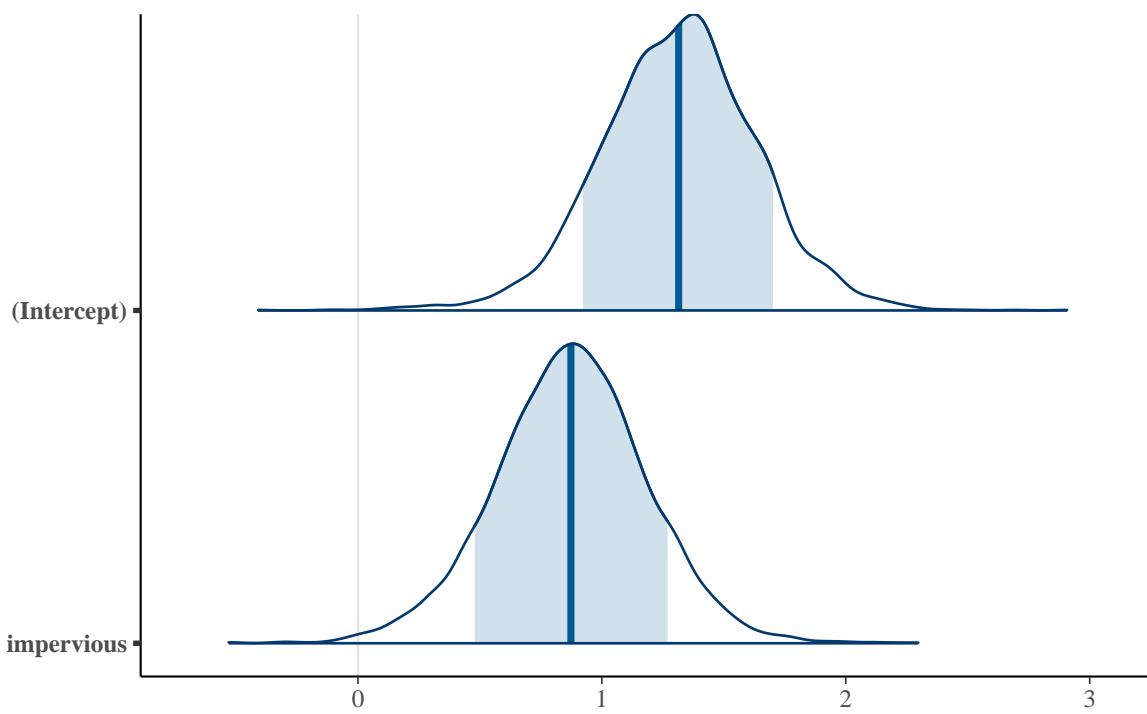
#### 4.4.4.2 Posterior Uncertainty Levels

Estimated posterior density curves with 80% confidence intervals are shown in Figure ??.

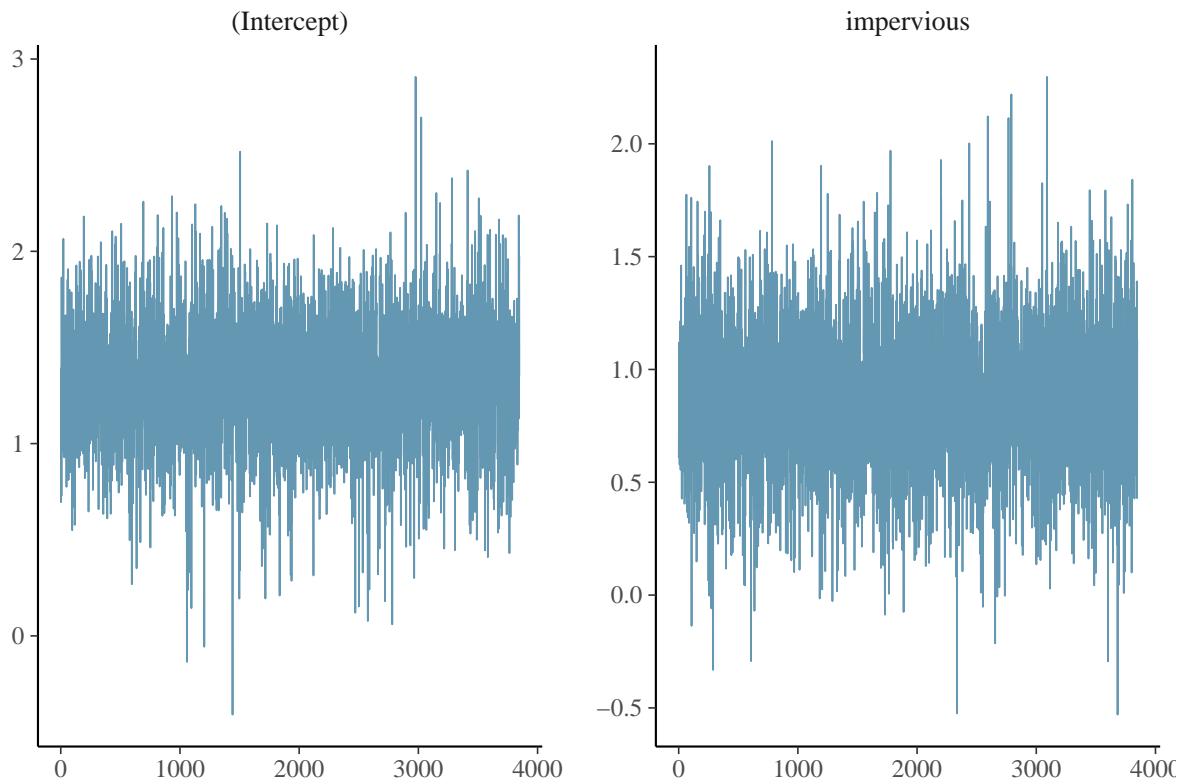
```
## [[1]]
```

## Lead – Water – Total

Posterior distributions with medians and 80% intervals



#### Model Diagnostics Figure ?? shows the diagnostic trace plot of MCMC draws. This was used to verify prior distribution assumptions.



#### 4.4.5 Total Phosphorus

##### 4.4.5.1 Regression Coefficients

Regression results for Total Phosphorus are summarized in Table 4.4.5.1.

Total Phosphorus - Water - Total - Bayseian Regression Results

Predictor

Posterior Mean

lower 95% CI

upper 95% CI

MCMC p value

effect

(Intercept)

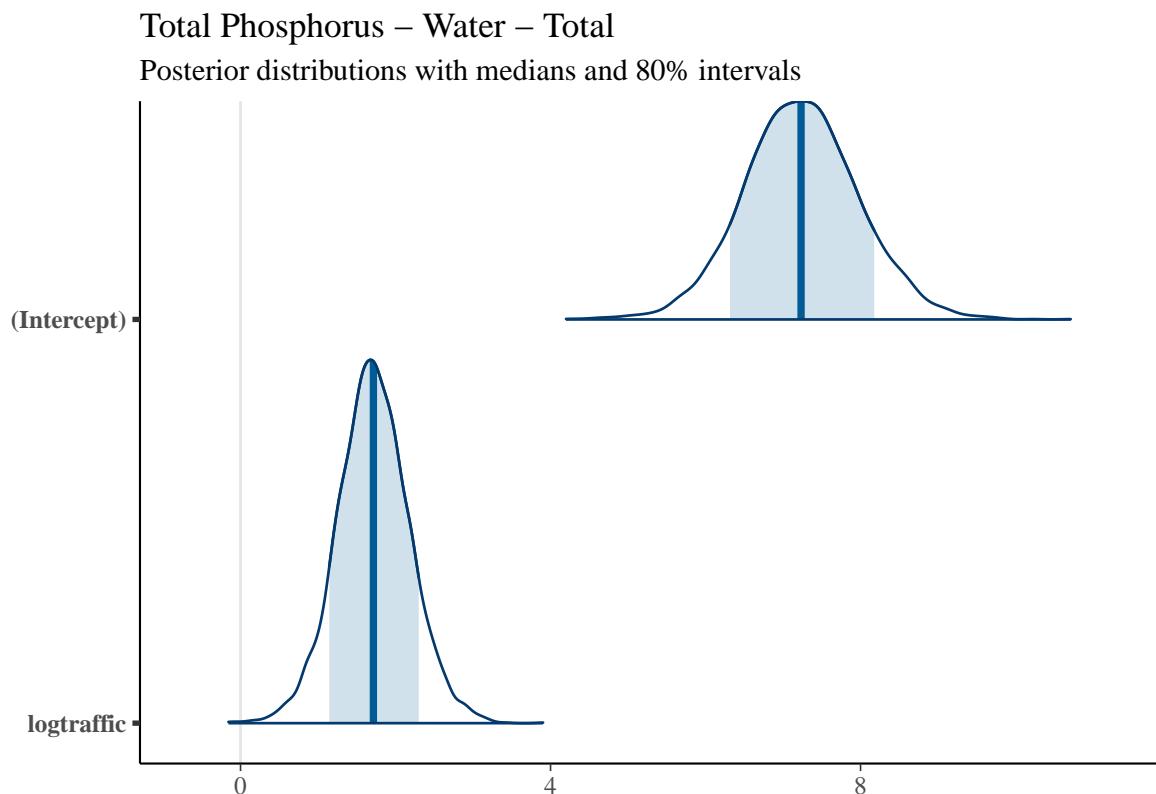
7.231

```
5.799  
8.734  
0.000  
fixed  
logtraffic  
1.714  
0.806  
2.649  
0.002  
fixed  
Location  
0.251  
0.065  
0.517  
NA  
random  
units  
1.000  
1.000  
1.000  
NA  
residual
```

#### 4.4.5.2 Posterior Uncertainty Levels

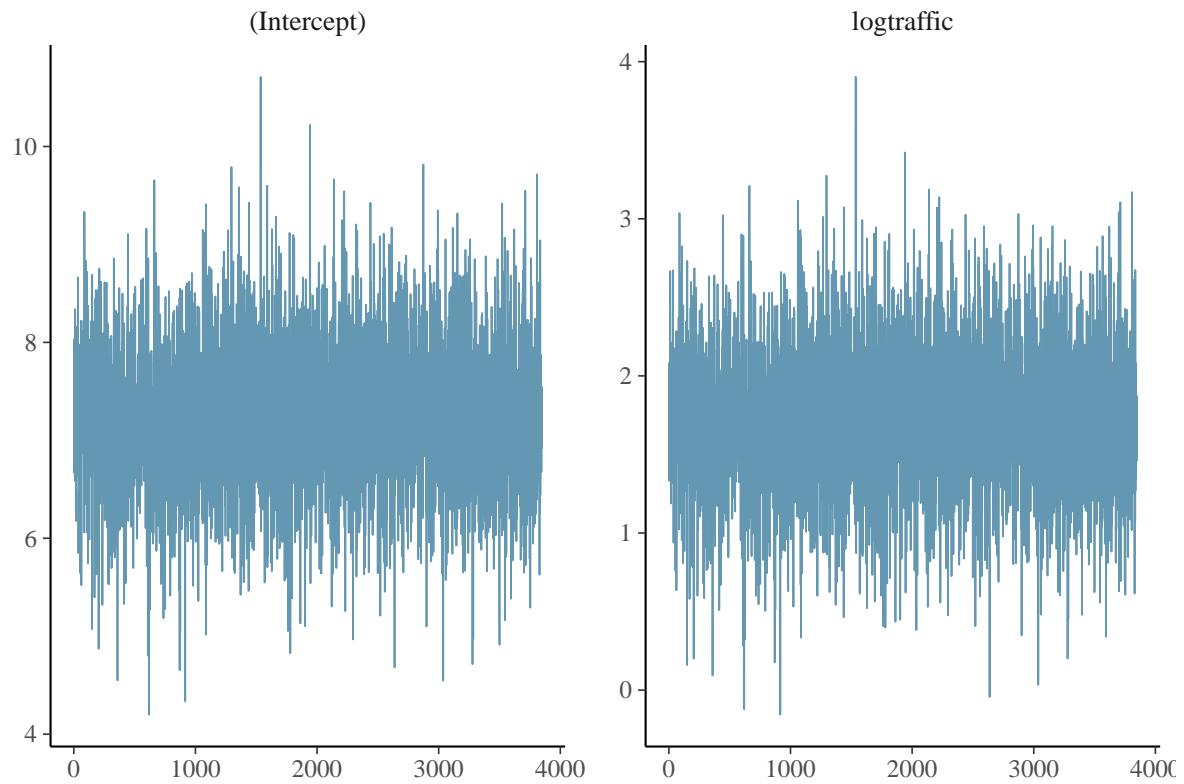
Estimated posterior density curves with 80% confidence intervals are shown in Figure ??.

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



#### 4.4.5.3 Model Diagnostics

Figure ?? shows the diagnostic trace plot of MCMC draws. This was used to verify prior distribution assumptions.



#### Total Suspended Solids #### Regression Coefficients  
 Regression results for Total Suspended Solids are summarized in Table 4.4.5.3.

#### Total Suspended Solids - Water - Total - Bayseian Regression Results

Predictor

Posterior Mean

lower 95% CI

upper 95% CI

MCMC p value

effect

(Intercept)

13.217

11.457

14.976

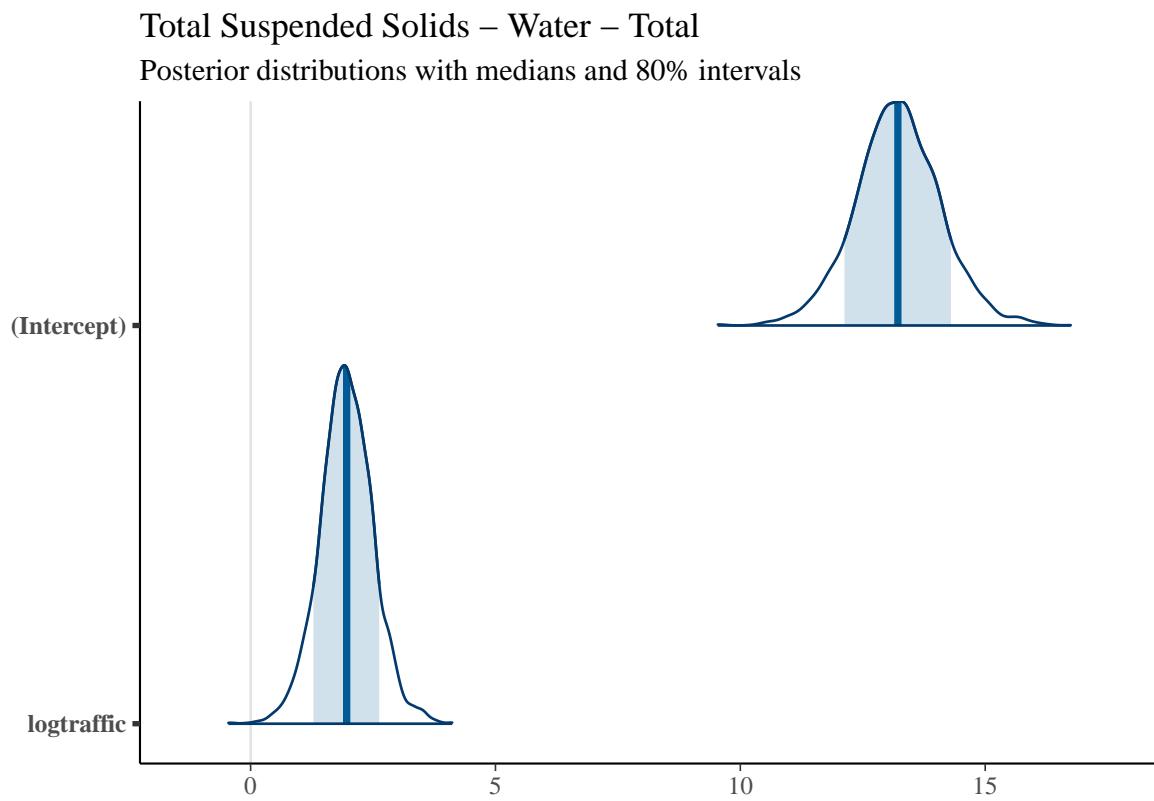
0.000

```
fixed  
logtraffic  
1.961  
0.885  
3.058  
0.002  
fixed  
Location  
0.349  
0.091  
0.707  
NA  
random  
units  
1.000  
1.000  
1.000  
NA  
residual
```

#### 4.4.5.4 Posterior Uncertainty Levels

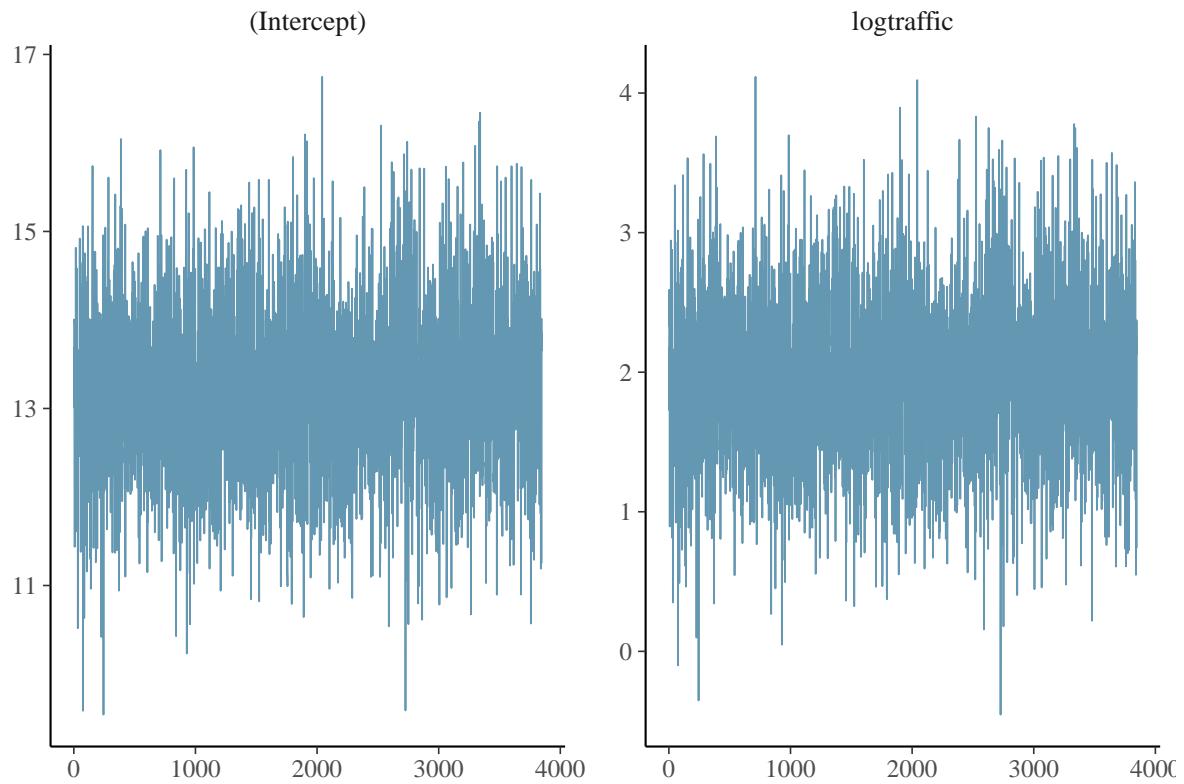
Estimated posterior density curves with 80% confidence intervals are shown in Figure ??.

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



#### 4.4.5.5 Model Diagnostics

Figure ?? shows the diagnostic trace plot of MCMC draws. This was used to verify prior distribution assumptions.



#### 4.4.6 Total PAH

##### 4.4.6.1 Regression Coefficients

Regression results for Total PAH are summarized in Table 4.4.6.1.

Total PAH - Water - Total - Bayseian Regression Results

Predictor

Posterior Mean

lower 95% CI

upper 95% CI

MCMC p value

effect

(Intercept)

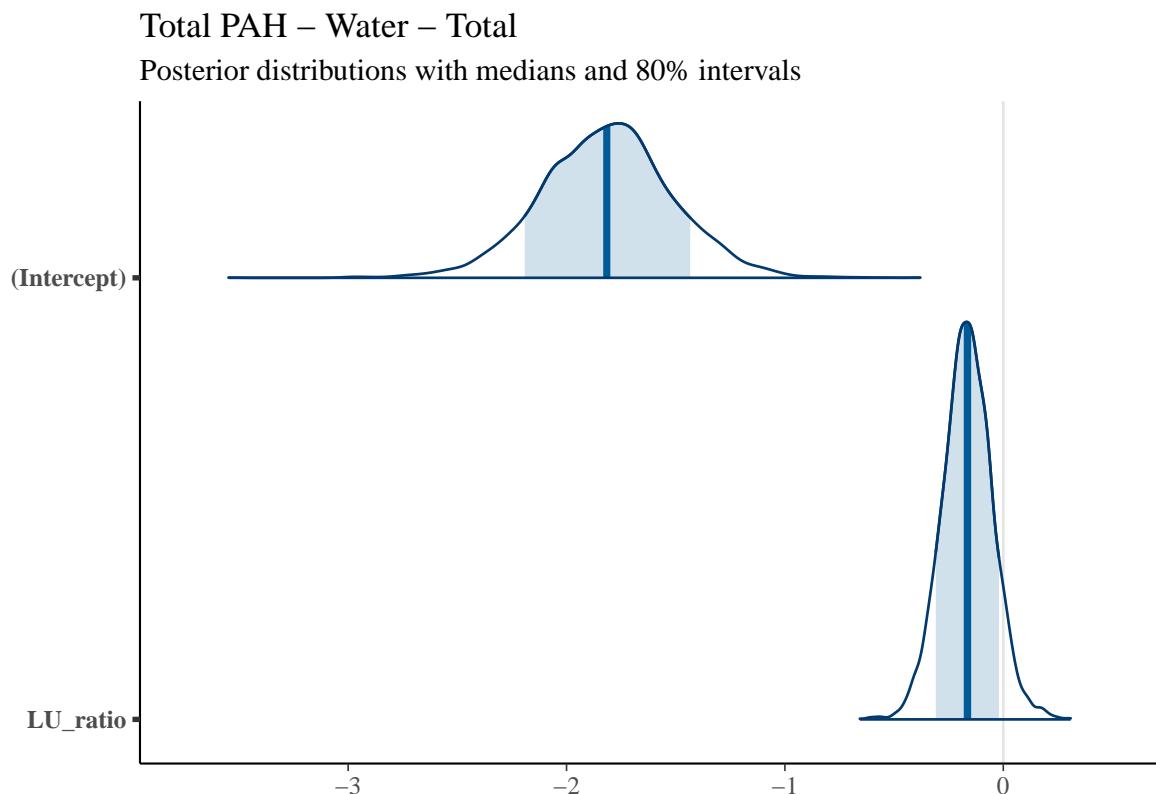
-1.815

```
-2.430
-1.243
0.00
fixed
LU_ratio
-0.164
-0.396
0.061
0.15
fixed
Location
1.226
0.400
2.480
NA
random
units
1.000
1.000
1.000
NA
residual
```

#### 4.4.6.2 Posterior Uncertainty Levels

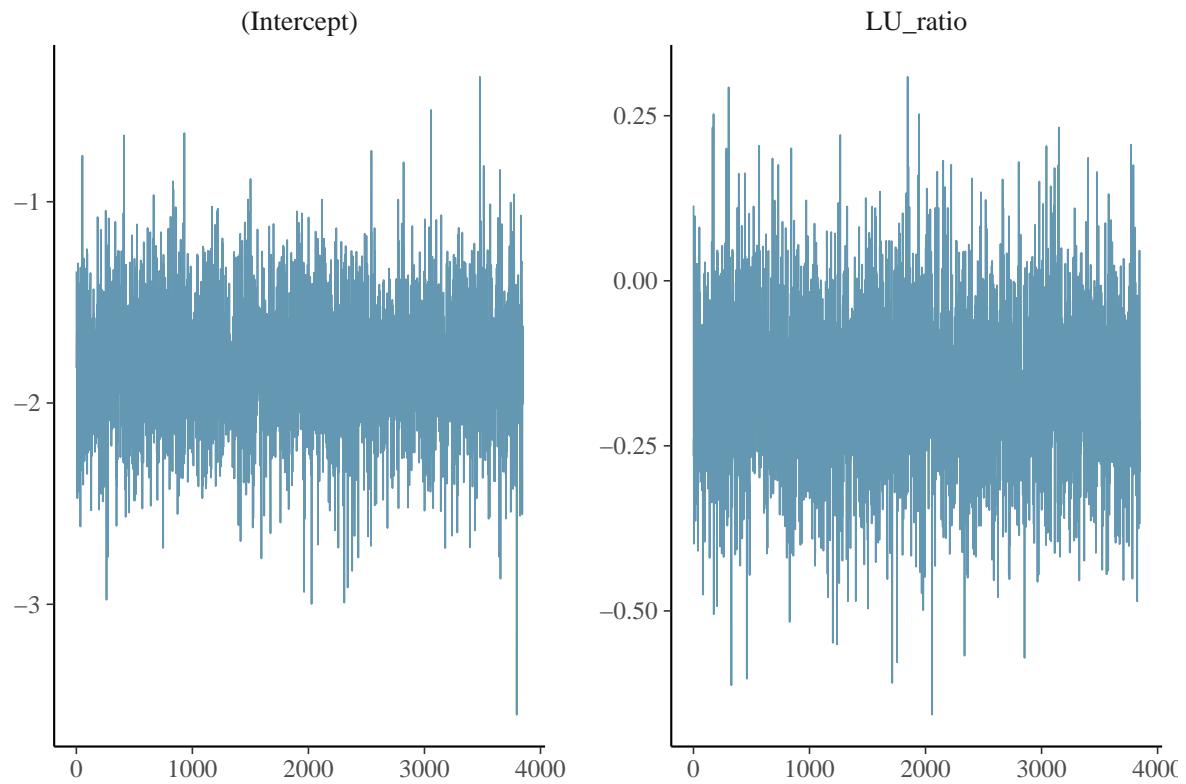
Estimated posterior density curves with 80% confidence intervals are shown in Figure ??.

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



#### 4.4.6.3 Model Diagnostics

Figure ?? shows the diagnostic trace plot of MCMC draws. This was used to verify prior distribution assumptions.



#### 4.4.7 Total CPAH

##### 4.4.7.1 Regression Coefficients

Regression results for Total CPAH are summarized in Table 4.4.7.1.

CPAH - Water - Total - Bayseian Regression Results

Predictor

Posterior Mean

lower 95% CI

upper 95% CI

MCMC p value

effect

(Intercept)

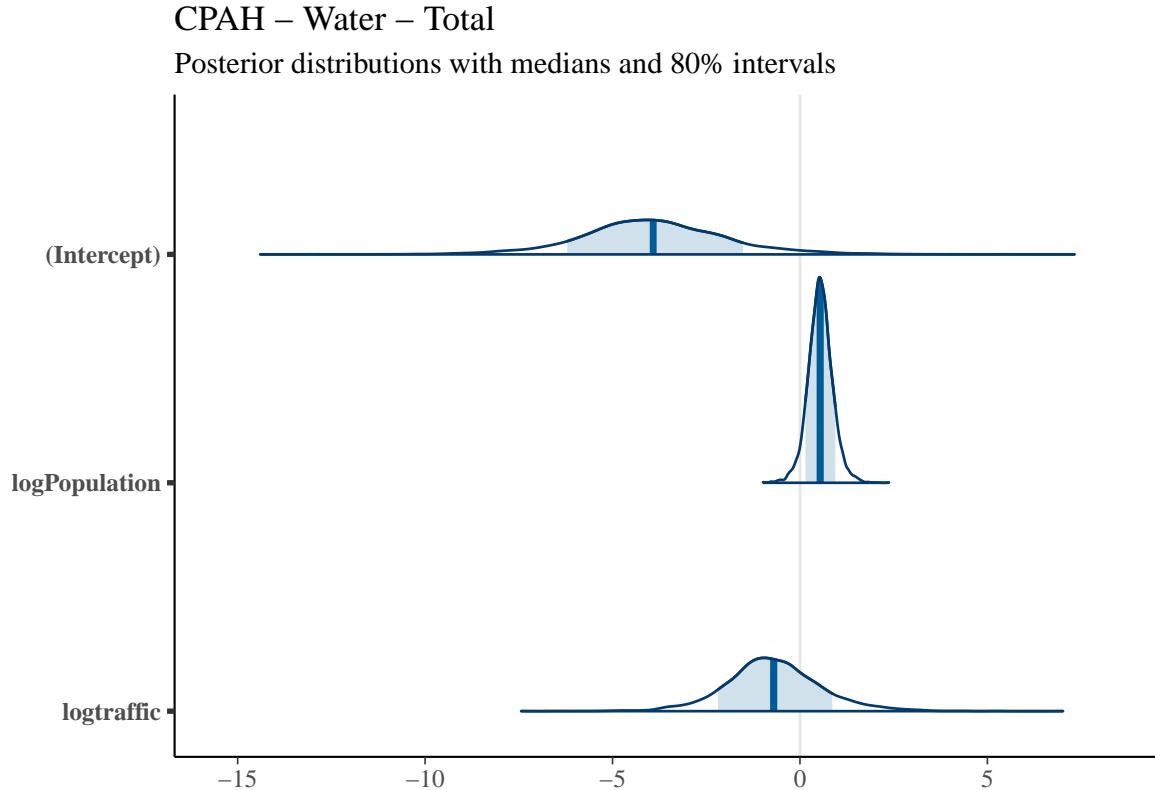
-3.916

-8.059  
-0.059  
0.060  
fixed  
logPopulation  
0.539  
-0.149  
1.178  
0.095  
fixed  
logtraffic  
-0.702  
-3.287  
1.957  
0.521  
fixed  
Location  
1.271  
0.298  
2.833  
NA  
random  
units  
1.000  
1.000  
1.000  
NA  
residual

#### 4.4.7.2 Posterior Uncertainty Levels

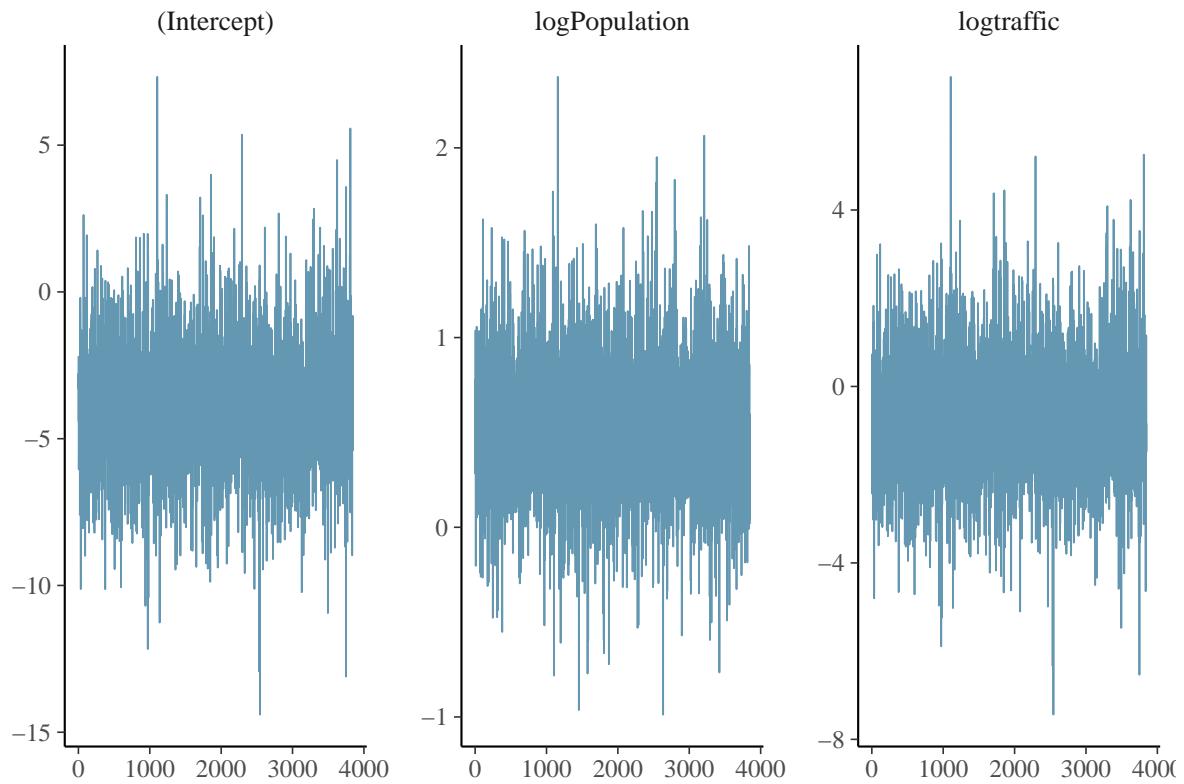
Estimated posterior density curves with 80% confidence intervals are shown in Figure ??.

```
## [1]
```



#### 4.4.7.3 Model Diagnostics

Figure ?? shows the diagnostic trace plot of MCMC draws. This was used to verify prior distribution assumptions.



#### 4.4.8 Total HPAH

##### 4.4.8.1 Regression Coefficients

HPAH - Water - Total - Bayseian Regression Results

Predictor

Posterior Mean

lower 95% CI

upper 95% CI

MCMC p value

effect

(Intercept)

-2.2172457

-2.9392675

```
-1.4083642
```

```
0.0002599
```

```
fixed
```

```
LU_ratio
```

```
-0.1939536
```

```
-0.4814508
```

```
0.0923844
```

```
0.1715623
```

```
fixed
```

```
Location
```

```
1.9186908
```

```
0.4360824
```

```
4.1605803
```

```
NA
```

```
random
```

```
units
```

```
1.0000000
```

```
1.0000000
```

```
1.0000000
```

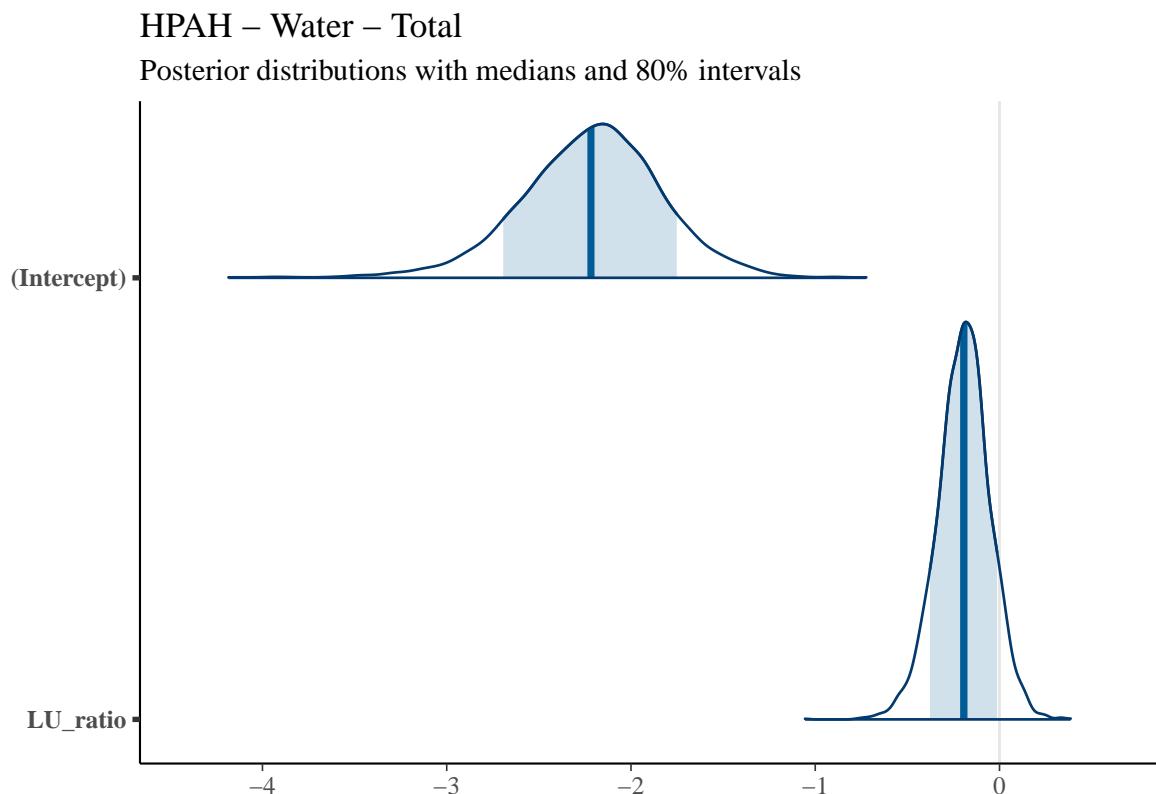
```
NA
```

```
residual
```

#### 4.4.8.2 Posterior Uncertainty Levels

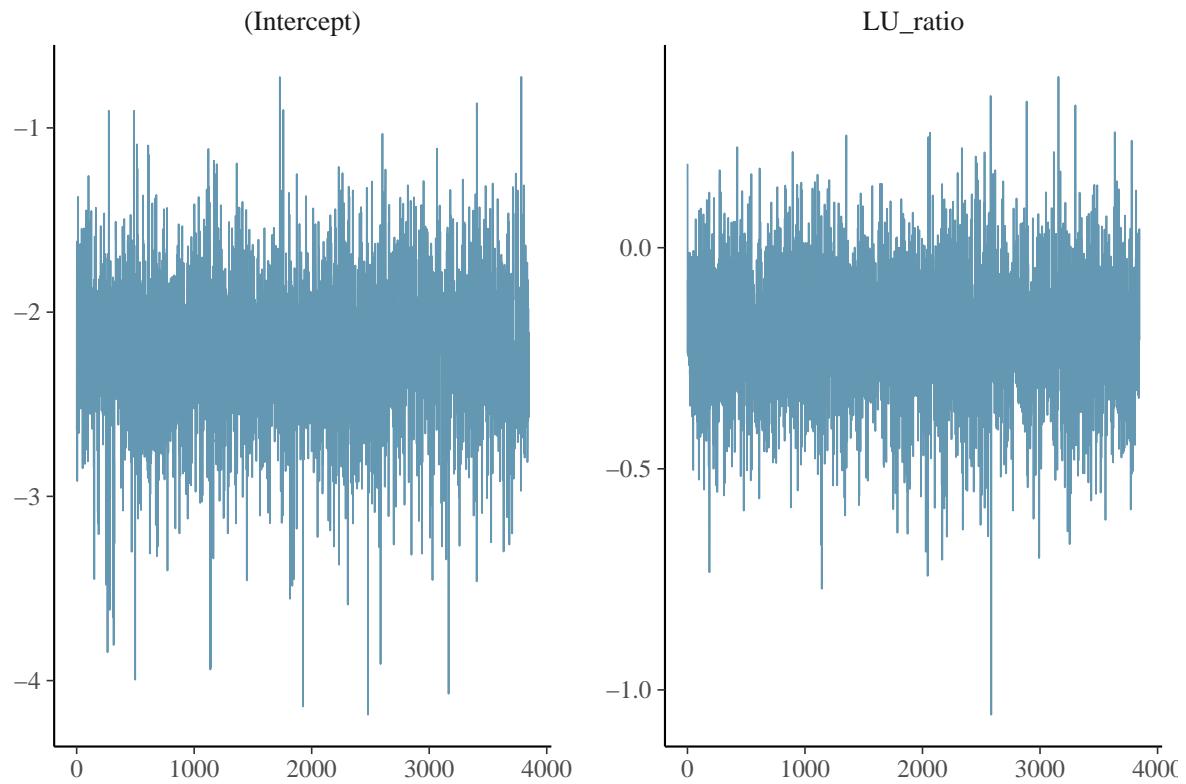
Estimated posterior density curves with 80% confidence intervals are shown in Figure ??.

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



#### 4.4.8.3 Model Diagnostics

Figure ?? shows the diagnostic trace plot of MCMC draws. This was used to verify prior distribution assumptions.



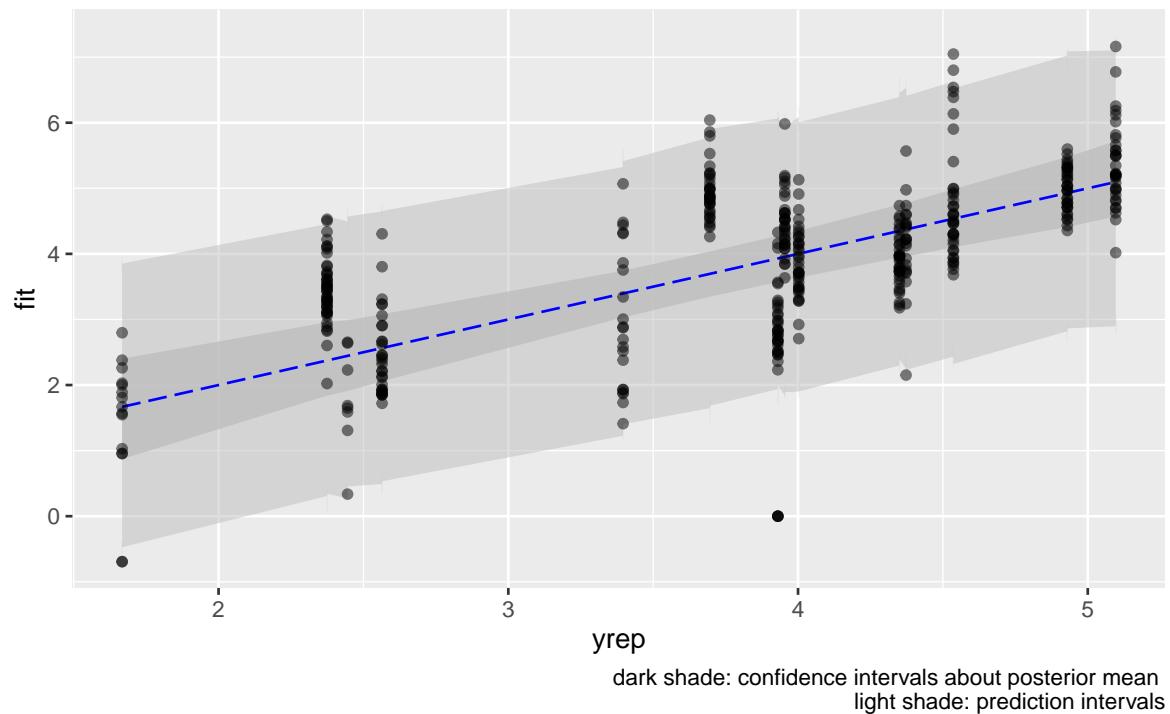
## 4.5 Predictions

We used the models developed above to generate predictions of pollutant concentrations to verify regression parameters. We first evaluated the posterior means of the predictors to evaluate data fit. We then simulated 100 draws from the full posterior distributions to evaluate the range of predictions. Results are show below.

```
## [1]
```

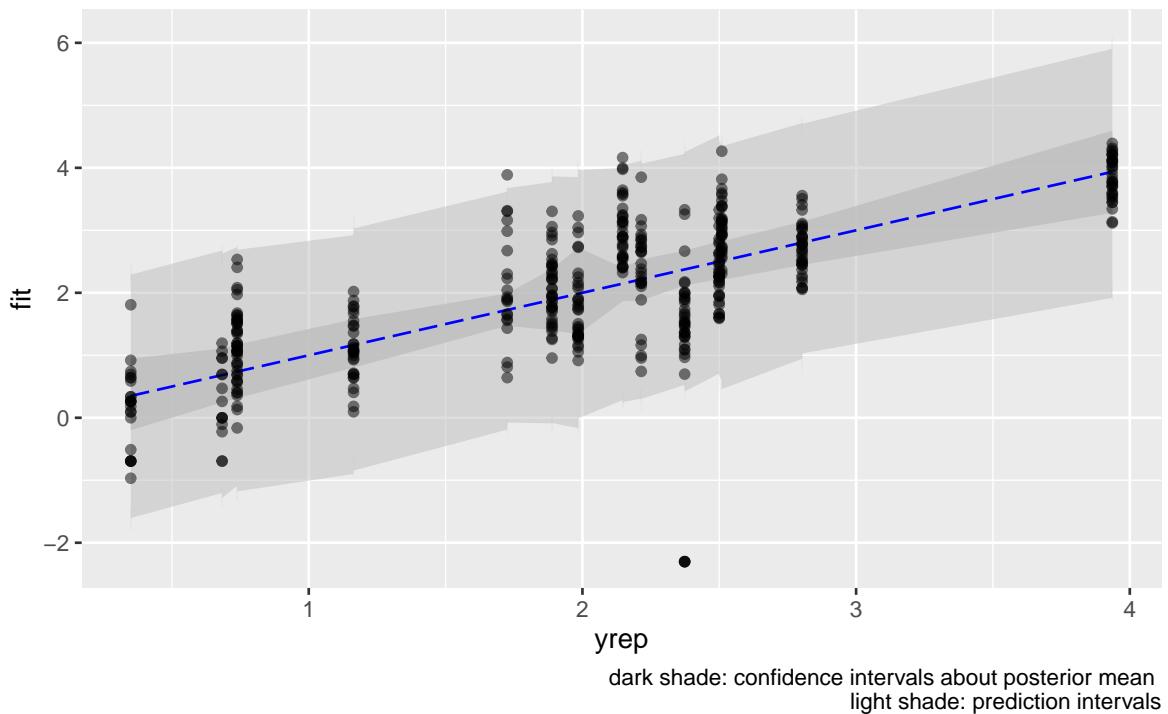
## Zinc – Water – Total

Simulated vs observed values.



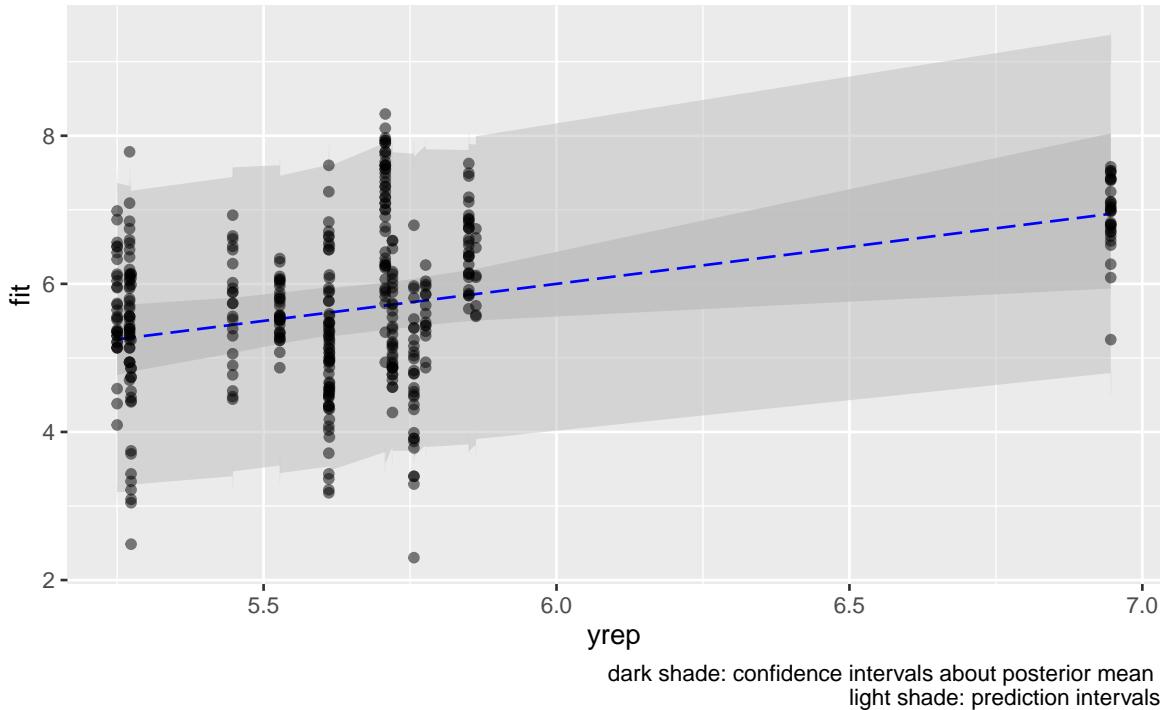
```
#> [[1]]
```

Copper – Water – Total  
Simulated vs observed values.



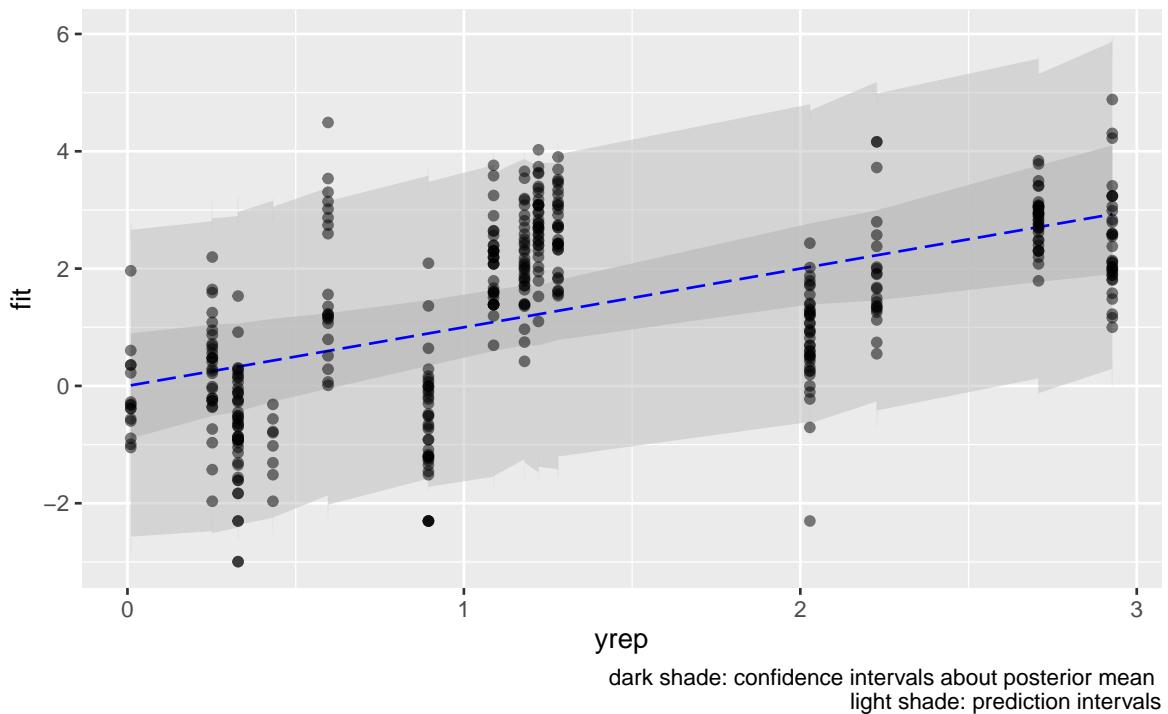
## [[1]]

Nitrite–Nitrate – Water – Dissolved  
Simulated vs observed values.



#  
## [[1]]

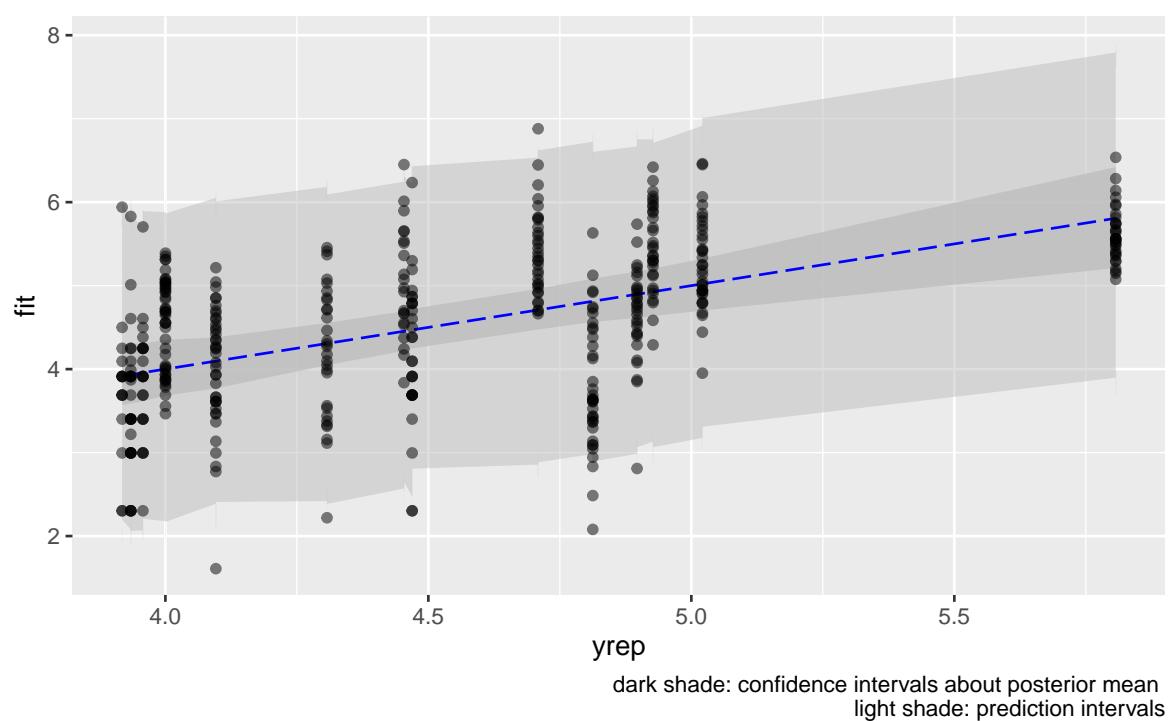
Lead – Water – Total  
Simulated vs observed values.



## [[1]]

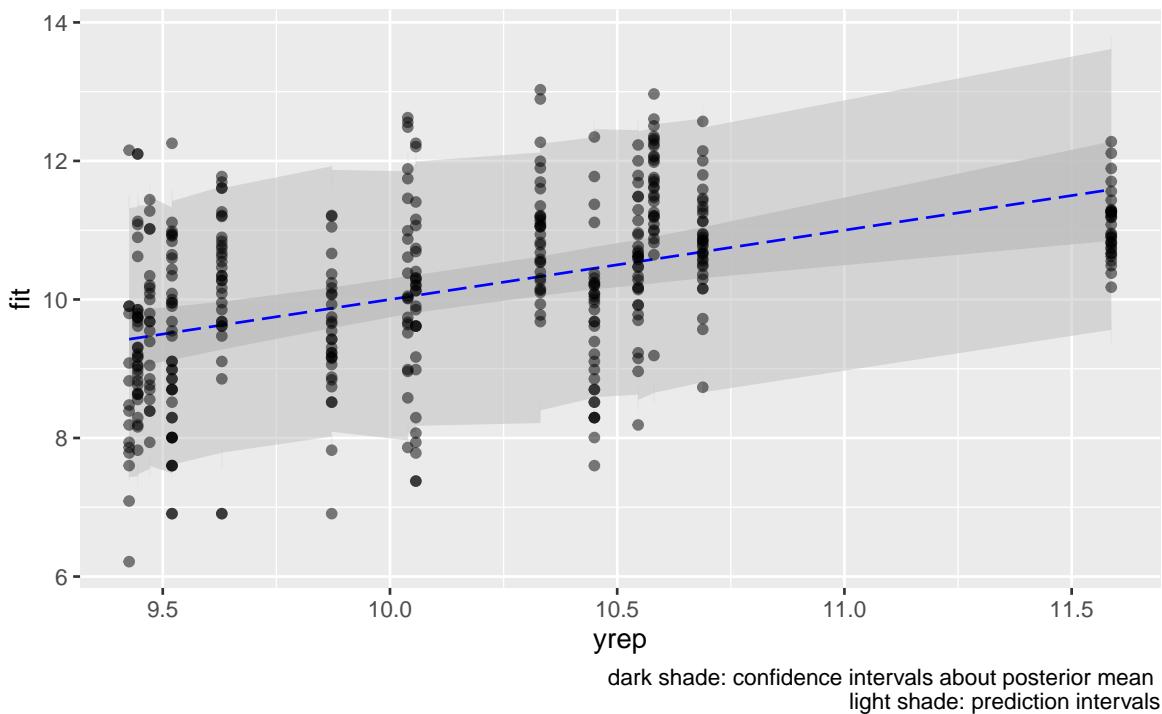
## Total Phosphorus – Water – Total

Simulated vs observed values.



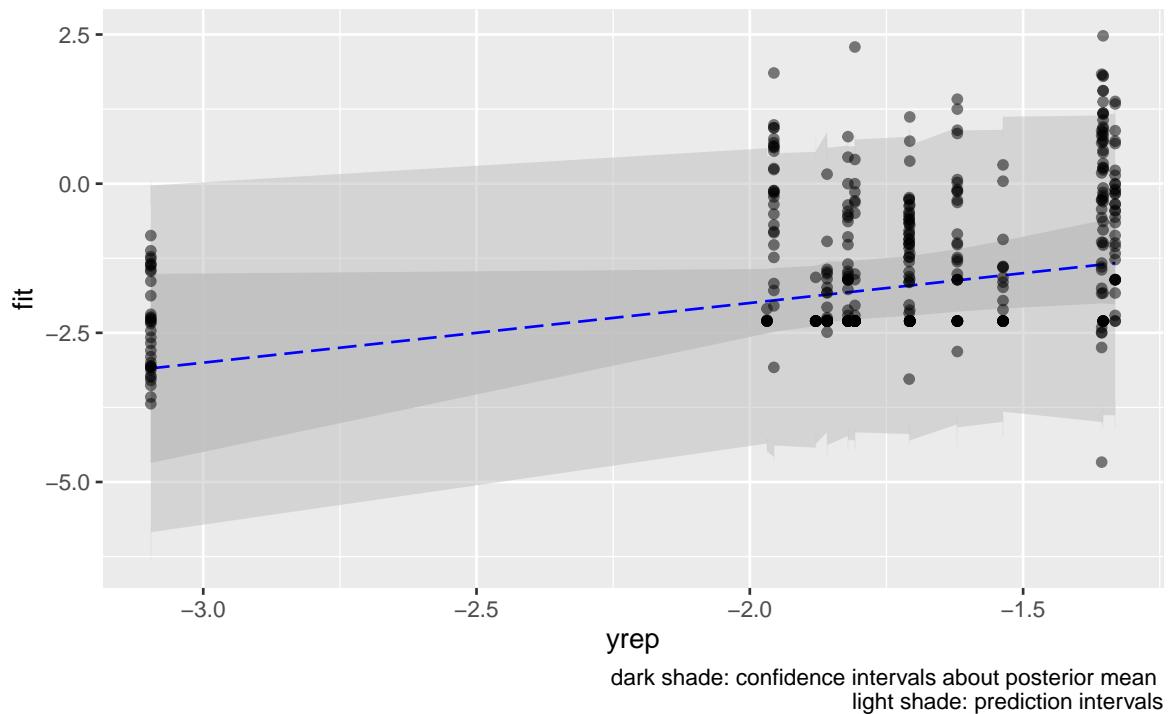
```
#> [[1]]
```

Total Suspended Solids – Water – Total  
Simulated vs observed values.



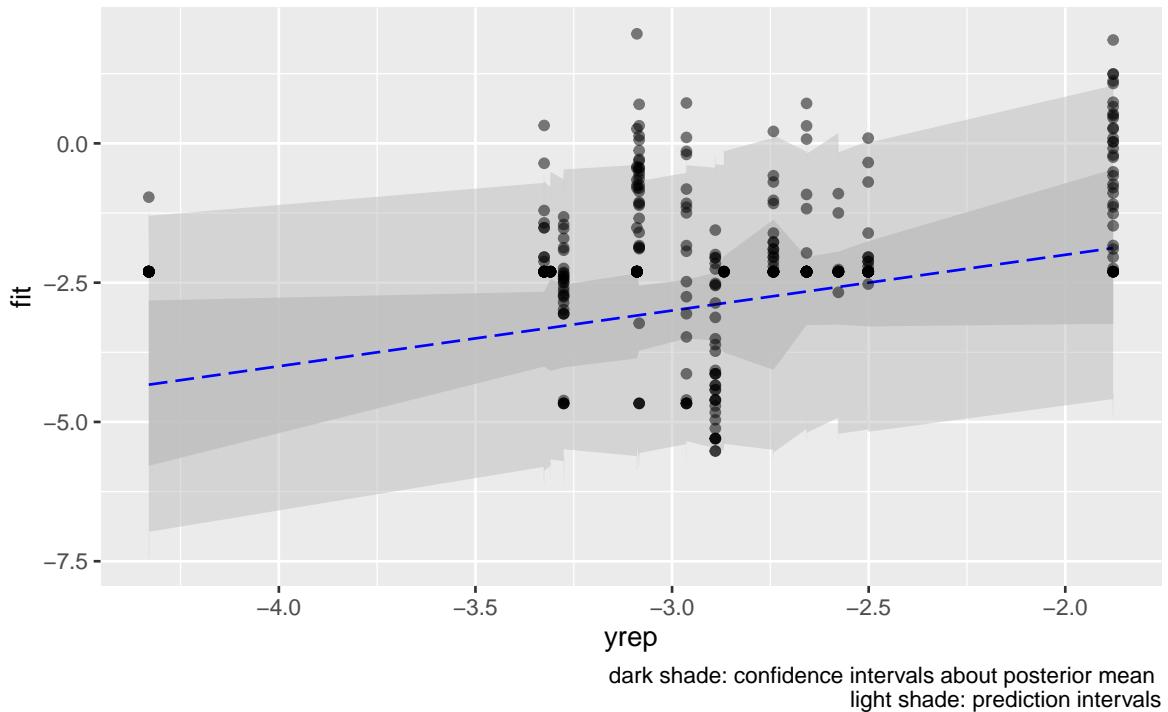
#  
## [[1]]

Total PAH – Water – Total  
Simulated vs observed values.



#  
## [[1]]

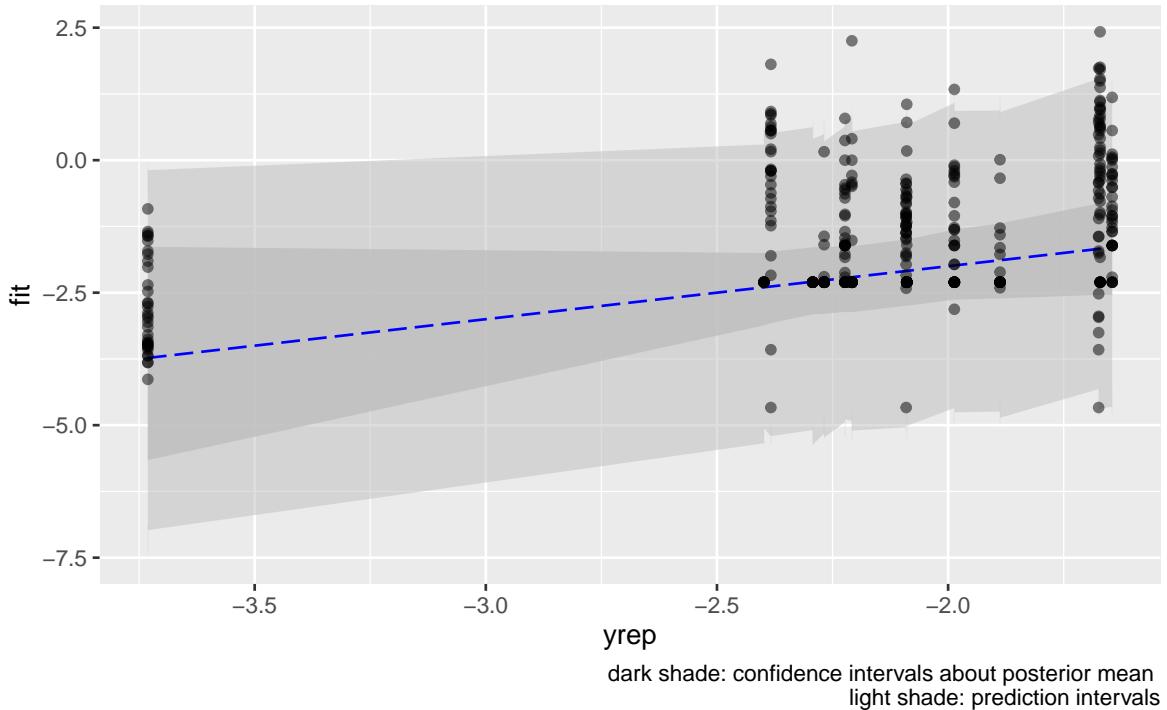
**CPAH – Water – Total**  
Simulated vs observed values.



## [[1]]

## HPAH – Water – Total

Simulated vs observed values.



## 4.6 Discussion

Our models appear to fit both the mean concentrations within each location and capture the full spread of the data within the bounds of the 95% confidence intervals. Model fit appears to be the strongest for metals (Zinc, Copper, Lead) and not as strong for organics (PAH, CPAH and HPAH). These may be improved through specifying a better prior distribution.

To our knowledge, this work represents the first effort to predict stormwater pollution loading from landscape characteristics, rather than from simple land use categories. In the stormwaterheatmap tool, we use these predictive models to spatially output predicted levels of commonly reported stormwater pollutants.

## 4.7 Generation of heatmap layers

To produce heatmap layers of pollutant concentration, we use the original spatial predictor layers along with a linear relationship between response and predictors. To avoid extrapolating beyond the available data, we first clamp the predictors to the high and low values observed with monitored

watersheds. We then reduce predictor layers into an array-valued image, where each pixel in the model domain contains an array of predictor values. We then multiply this layer by an array that contains the intercept of the regression relationship, along with coefficients.

For example for each COC, we have a linear regression in the form of:

$$y_{i,j} = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + \epsilon_{i,j}$$

Where  $\beta_0$  is the intercept and  $\beta_n$  are the regression coefficients. We express this as a 1-D array in the form of

$$\text{Coefficients} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_n \end{bmatrix} \cdot \text{predictors} \begin{bmatrix} 1 \\ x_1 \\ \dots \\ x_n \end{bmatrix}$$

The predictor layer is comprised of pixels with 1-D arrays of predictor values in the form of:

$$\begin{bmatrix} 1 \\ x_1 \\ \dots \\ x_n \end{bmatrix}$$

We then calculate the dot product of the two arrays and reduce the values to a single image. This is expressed in the same manner as the linear regression equation:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n$$

# Chapter 5

## References

- Anan, Y; Kunito T, T Ikemoto, R Kubota, I Watanabe, S Tanabe, N Miyazaki, and E Petrov. 2002. "Elevated concentrations of trace elements in Caspian seals (*Phoca caspia*) found stranded during the mass mortality event in 2000." *Environmental Contamination & Toxicology* 42: 354–62.
- Anderson, B, J Hunt, B Phillips, B Thompson, S Lowe, i K Tabersk, and R Carr. 2007. "Patterns and trends in sediment toxicity in the San Francisco Estuary." *Environmental Research* 105: 145–55.
- Arkoosh, M, E Clemons, P Huffman, A Kagley, E Casillas, N Adams, H Sanborn, T Collier, and J Stein. 2001. "Increased Susceptibility of Juvenile Chinook Salmon to Vibriosis after Exposure to Chlorinated and Aromatic Compounds Found in Contaminated Urban Estuaries." *Journal of Aquatic Animal Health* 13 (3): 257–68.
- Baldwin, D, C Tatara, and N Scholz. 2011. "Copper induced olfactory toxicity in salmon and steelhead: Extrapolation across species and rearing environments." *Aquatic Toxicology*. 101 (1): 295–97.
- Booth, Derek B, James R Karr, Sally Schauman, Christopher P Konrad, Sarah A Morley, Marit G Larson, and Stephen J Burges. 2004. "Reviving urban streams: Land use, hydrology, biology, and human behavior 1." *JAWRA Journal of the American Water Resources Association* 40 (5): 1351–64.
- Department of Ecology. 2014. "Stormwater Management Manual for Western Washington," 1192.
- Dinicola, Richard S. 1990. "Characterization and simulation of rainfall-runoff relations for headwater basins in western King and Snohomish Counties, Washington." *Water-Resources Investigations Report 89-4052*: 52 p. <https://pubs.er.usgs.gov/publication/wri894052>.
- Feist, Blake E., Eric R. Buhle, David H. Baldwin, Julann A. Spromberg, Steven E. Damm, Jay W. Davis, and Nathaniel L. Scholz. 2017. "Roads to ruin: Conservation threats to a sentinel species across an urban gradient." *Ecological Applications* 27 (8): 2382–96. <https://doi.org/10.1002/eap.1615>.
- Gaw, S, K Thomas, and T Hutchinson. 2014. "Sources, impacts and trends of pharmaceuticals in the marine and coastal environment." *Philosophical Transactions of the Royal Society B: Biological Sciences* 369: 20130572.
- Hadfield, Jarrod D, and Others. 2010. "MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package." *Journal of Statistical Software* 33 (2): 1–22.

- Helsel, Denise R. 2012. *Statistics for Censored Environmental Data*.
- Incardona, J, T Linbo, and N Scholz. 2011. "Cardiac toxicity of 5-ring polycyclic aromatic hydrocarbons is differentially dependent on the aryl hydrocarbon receptor 2 isoform during zebrafish development." *Toxicology and Applied Pharmacology* 257 (2): 242–49.
- Jarvis, T., and G. Bielmyer-Fraser. 2015. "Accumulation and effects of metal mixtures in two seaweed species." *Comparative Biochemistry and Physiology, Part C: Toxicology and Pharmacology* 171: 28–33.
- Kayhanian, M, B Fructman, J Gulliver, C Montanaro, E Ranieri, and S Wuertz. 2012. "Review of highway runoff characteristics: Comparative analysis and universal implications." *Water Research*. Vol. 46, pp." *Water Research*, 6609–24.
- King County. 2014. "Development of a Stormwater Retrofit Plan for Water Resources Inventory Area (WRIA) 9: Comprehensive Needs and Cost Assessment and Extrapolation to Puget Sound." Prepared by Simmonds J; Wright O, King County Department of Natural Resource; Parks, Water; Land Resources Division.
- Lampert, David. 2019. "PyHSPF." <https://github.com/djlampert/PyHSPF>.
- Levin, P, E Howe, and J Robertson. n.d. "Impacts of stormwater on coastal ecosystems: the need to match the scales of management objectives and solutions." *Philosophical Transactions of the Royal Society B: Biological Sciences*.
- Lyngby, J, and H Brix. 1984. "Seasonal and environmental variation in cadmium, copper, lead, and zinc concentrations in eelgrass (*Zostera marina L.*) in the Limfjord, Denmark." *Aquatic Botany* 54: 100–106.
- Mauger, G. S., J. S. Won, K. Hegewisch, C. Lynch, R. Lorente Plazas, Salathe, and E. P. Salathé Jr. 2018. "New Projections of Changing Heavy Precipitation in King County."
- McIntyre, J, J Davis, C Hinman, K Macneale, B Anulacion, N Scholz, and J Stark. 2015. "Soil bioretention protects juvenile salmon and their prey from the toxic impacts of urban stormwater runoff." *Chemosphere* 132: 213–19.
- Meador, J, F Sommers, and SC Ylitalo. 2006. "Altered growth and related physiological responses in juvenile Chinook salmon (*Oncorhynchus tshawytscha*) from dietary exposure to polycyclic aromatic hydrocarbons (PAHs)." *Canadian Journal of Fisheries and Aquatic Sciences* 63: 2364–76.
- Moore, Richard B, Lucinda D McKay, Alan H Rea, Timothy R Bondelid, Curtis V Price, Thomas G Dewald, and Craig M Johnston. 2019. "User's guide for the national hydrography dataset plus (NHDPlus) high resolution." US Geological Survey.
- NASA Goddard Earth Sciences Data and Information Services Center (GES DISC). 2019. "NLDAS-2: North American Land Data Assimilation System Forcing Fields."
- Palmer, Margaret, and Albert Ruhi. 2019. "Linkages between flow regime, biota, and ecosystem processes: Implications for river restoration." *Science* 365 (6459): eaaw2087.
- Rice, C, M Myers, M Willis, B French, and E Casillas. 2000. "From sediment bioassay to fish biomarker – connecting the dots using simple trophic relationships." *Marine Environmental Research* 50: 527–22.
- Ross, C.W., L. Prihodko, J.Y. Anchang, S.S. Kumar, W. Ji, N. P. Hanan. 2018. "Global Hydrologic Soil Groups (HYSGOs250m) for Curve Number-Based Runoff Modeling." <https://doi.org/10.3334/ORNLDAAC/1566>.

- Ross, P, Ellis G, M Ikonomou, L Barrett-Lennard, and R Addison. 2000. "High PCB concentrations in freeranging Pacific killer whales, *Orcinus orca*: Effects of age, sex and dietary preference." *Marine Pollution Bulletin* 40 (6): 504–15.
- Scholz, N, M Myers, S McCarthy, J Labenia, and J McIntyre. 2011. "Recurrent Die-Offs of Adult Coho Salmon Returning to Spawn in Puget Sound Lowland Urban Streams." *PLoS ONE* 6 (12): e28013.
- Soil Survey Staff. 2016. "National Value Added Look Up (valu) Table Database for the Gridded Soil Survey Geographic (gSSURGO) Database for the United States of America and the Territories, Commonwealths, and Island Nations served by the USDA-NRCS."
- . 2018. "Gridded Soil Survey Geographic (gSSURGO) Database for Washington." <https://gdg.sc.egov.usda.gov/>.
- United States Environmental Protection Agency. 2019. "National rivers and streams assessment 2008-2019: A collaborative survey." U.S. Environmental Protection Agency.
- Walsh, Christopher J., Allison H. Roy, Jack W. Feminella, Peter D. Cottingham, Peter M. Groffman, and Raymond P. Morgan. 2005. "The urban stream syndrome: Current knowledge and the search for a cure." *Journal of the North American Benthological Society* 24 (3): 706–23. <https://doi.org/10.1899/04-028.1>.
- Washington State Department of Ecology. 2011. "Toxics in Surface Runoff to Puget Sound: Phase 3 Data and Load Estimates." Prepared by Herrera Environmental Consultants, Inc.: Washington Department of Ecology.
- Washington State Department of Ecology and Washington Department of Health. 2012. "PAH Chemical Action Plan." Prepared Davies H, Stone A, Grice J, Patora K, Kadlex M, Delistraty D, Norton D,; White J.: Washington Department of Ecology; Washington Department of Health.
- WDFW. 2011. "Toxic Contaminants in Harbor Seal (*Phoca vitulina*) Pups from Puget Sound." Prepared by Noell M, Ross P, Jeffries S, Lance M: Washington Department of Fish; Wildlife.
- William Hobbs, Brandi Lubliner, Nataniel Kale, and Evan Newell. 2015. "Western Washington NPDES Phase I Stormwater Permit Final S8 . D Data Characterization." 15.
- Yang, Limin, Suming Jin, Patrick Danielson, Collin Homer, Leila Gass, Stacie M Bender, Adam Case, et al. 2018. "A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies." *ISPRS Journal of Photogrammetry and Remote Sensing* 146: 108–23.

## Appendix A

# Regional HSPF Calibration Factors

Soil  
Land Cover  
Slope  
LZSN  
INFILT  
LSUR  
SLSUR  
KVARY  
AGWRC  
INFEXP  
INFILD  
BASETP  
AGWETP  
CEPSC  
UZSN  
NSUR  
INTFW  
IRC  
LZETP  
IWAT

RETSC

A/B

Forest

Flat

5.0

2.00

400

0.050

0.3

0.996

2

2

0

0.00

0.20

0.50

0.35

0

0.7

0.70

NA

NA

A/B

Forest

Mod

5.0

2.00

400

0.100

0.3

0.996

2  
2  
0  
0.00  
0.20  
0.50  
0.35  
0  
0.7  
0.70  
NA  
NA  
A/B  
Forest  
Steep  
5.0  
2.00  
400  
0.150  
0.3  
0.996  
2  
2  
0  
0.00  
0.20  
0.50  
0.35  
0  
0.7  
0.70

NA  
NA  
A/B  
Pasture  
Flat  
5.0  
1.50  
400  
0.050  
0.3  
0.996  
2  
2  
0  
0.00  
0.15  
0.50  
0.30  
0  
0.7  
0.40  
NA  
NA  
A/B  
Pasture  
Mod  
5.0  
1.50  
400  
0.100  
0.3

0.996

2

2

0

0.00

0.15

0.50

0.30

0

0.7

0.40

NA

NA

A/B

Pasture

Steep

5.0

1.50

400

0.150

0.3

0.996

2

2

0

0.00

0.15

0.50

0.30

0

0.7

0.40

NA

NA

A/B

Lawn

Flat

5.0

0.80

400

0.050

0.3

0.996

2

2

0

0.00

0.10

0.50

0.25

0

0.7

0.25

NA

NA

A/B

Lawn

Mod

5.0

0.80

400

0.100

0.3  
0.996  
2  
2  
0  
0.00  
0.10  
0.50  
0.25  
0  
0.7  
0.25  
NA  
NA  
A/B  
Lawn  
Steep  
5.0  
0.80  
400  
0.150  
0.3  
0.996  
2  
2  
0  
0.00  
0.10  
0.50  
0.25  
0

0.7  
0.25  
NA  
NA  
C  
Forest  
Flat  
4.5  
0.08  
400  
0.050  
0.5  
0.996  
2  
2  
0  
0.00  
0.20  
0.50  
0.35  
6  
0.5  
0.70  
NA  
NA  
C  
Forest  
Mod  
4.5  
0.08  
400

0.100

0.5

0.996

2

2

0

0.00

0.20

0.50

0.35

6

0.5

0.70

NA

NA

C

Forest

Steep

4.5

0.08

400

0.150

0.5

0.996

2

2

0

0.00

0.20

0.30

0.35

6  
0.3  
0.70  
NA  
NA  
C  
Pasture  
Flat  
4.5  
0.06  
400  
0.050  
0.5  
0.996  
2  
2  
0  
0.00  
0.15  
0.40  
0.30  
6  
0.5  
0.40  
NA  
NA  
C  
Pasture  
Mod  
4.5  
0.06

400  
0.100  
0.5  
0.996  
2  
2  
0  
0.00  
0.15  
0.40  
0.30  
6  
0.5  
0.40  
NA  
NA  
C  
Pasture  
Steep  
4.5  
0.06  
400  
0.150  
0.5  
0.996  
2  
2  
0  
0.00  
0.15  
0.25

0.30

6

0.3

0.40

NA

NA

C

Lawn

Flat

4.5

0.03

400

0.050

0.5

0.996

2

2

0

0.00

0.10

0.25

0.25

6

0.5

0.25

NA

NA

C

Lawn

Mod

4.5

0.03

400

0.100

0.5

0.996

2

2

0

0.00

0.10

0.25

0.25

6

0.5

0.25

NA

NA

C

Lawn

Steep

4.5

0.03

400

0.150

0.5

0.996

2

2

0

0.00

0.10

0.15  
0.25  
6  
0.3  
0.25  
NA  
NA  
SAT  
Forest  
Flat  
4.0  
2.00  
100  
0.001  
0.5  
0.996  
10  
2  
0  
0.70  
0.20  
3.00  
0.50  
1  
0.7  
0.80  
NA  
NA  
SAT  
Forest  
Mod

4.0  
2.00  
100  
0.010  
0.5  
0.996  
10  
2  
0  
0.70  
0.20  
3.00  
0.50  
1  
0.7  
0.80  
NA  
NA  
SAT  
Forest  
Steep  
4.0  
2.00  
100  
0.100  
0.5  
0.996  
10  
2  
0  
0.70

0.20  
3.00  
0.50  
1  
0.7  
0.80  
NA  
NA  
SAT  
Pasture  
Flat  
4.0  
1.80  
100  
0.001  
0.5  
0.996  
10  
2  
0  
0.50  
0.15  
3.00  
0.50  
1  
0.7  
0.60  
NA  
NA  
SAT  
Pasture

Mod  
4.0  
1.80  
100  
0.010  
0.5  
0.996  
10  
2  
0  
0.50  
0.15  
3.00  
0.50  
1  
0.7  
0.60  
NA  
NA  
SAT  
Pasture  
Steep  
4.0  
1.80  
100  
0.100  
0.5  
0.996  
10  
2  
0

0.50

0.15

3.00

0.50

1

0.7

0.60

NA

NA

SAT

Lawn

Flat

4.0

1.00

100

0.001

0.5

0.996

10

2

0

0.35

0.10

3.00

0.50

1

0.7

0.40

NA

NA

SAT

Lawn

Mod

4.0

1.00

100

0.010

0.5

0.996

10

2

0

0.35

0.10

3.00

0.50

1

0.7

0.40

NA

NA

SAT

Lawn

Steep

4.0

1.00

100

0.100

0.5

0.996

10

2

0  
0.35  
0.10  
3.00  
0.50  
1  
0.7  
0.40  
NA  
NA  
Impervious  
Impervious  
Flat  
NA  
NA  
400  
0.010  
NA  
1  
0.10

Impervious

Impervious

Moderate

NA

NA

400

0.050

NA

NA

NA

NA

NA

NA

NA

NA

0.10

NA

NA

NA

1

0.08

Impervious

Impervious

Steep

NA

NA

400

0.100

NA

NA

NA

NA

NA

NA

NA

NA

0.10

NA

NA

NA

1

0.05

## Appendix B

# Instructions for Accessing Tabulated Hydrology Results via BigQuery

## Tabulated Results via BigQuery

Tabulated hydrology results are available via Google BigQuery, a cloud-based relational database that includes a distributed SQL engine. The data are located on the `tnc-data-v1` data bucket (sign-in required). The table is named `tnc-data-v1:hydrology.gfdl`. BigQuery supports several client libraries. See <https://cloud.google.com/bigquery/docs/reference/libraries> for a list of supported clients libraries.

Using R, the `tnc-data-v1` databucket can be accessed through a database connection using the DBI package:

### Schema

The table schema are shown in Table B.

BigQuery Table Schema for `tnc-data-v1:hydrology.gfdl`.

Fieldname

Type

Description

grid

STRING

WRF Grid ID Number

```

year
INTEGER
Year of Simulation

month
INTEGER
Month of Simulation

comp
STRING
HSPF Runoff component (AGWO, IFWO, SURO)

hru000 ... hru252
STRING
Runoff (mm) (one column for each HRU)

Datetime
TIMESTAMP
Simulation Hour (UTC)

simulation_day
INTEGER
Day of simulation (01-Jan-1970 = Day 1)

simulation_day
INTEGER
Day of simulation (01-Jan-1970 = Day 1)|
```

## Querying Tabulated Results

The data may be queried through Google Cloud Platform directly, or through a number of available software libraries. Queries are performed through standard SQL language. Some example queries are provided below.

Get all surface flow components from the SeaTac precipitation grid (ID16\_V7) for the years 1970-1999:

```

SELECT
  *
FROM
  `tnc-data-v1.hydrology.gfdl`
WHERE
  grid = "ID16_V7"
```

```

AND comp IN ('suro',
  'ifwo')
AND year BETWEEN 1970
AND 1999

```

Get the annual peak flow for surface flow components from the SeaTac precipitation grid (ID16\_V7) for the years 1970-1999:

```

SELECT
  max(hru250) as peakQ, Year,
FROM
  `tnc-data-v1.hydrology.gfdl`
WHERE
  grid = "ID16_V7"
  AND comp IN ('suro',
    'ifwo')
  AND year BETWEEN 1970
  AND 1999
Group by Year

```

## Querying Geometry

Google BigQuery supports PostGIS geometry functions (see [https://cloud.google.com/bigquery/docs/reference/standard-sql/geography\\_functions](https://cloud.google.com/bigquery/docs/reference/standard-sql/geography_functions) for instructions).

Grid geometries are available from the `tnc-data-v1.gfdl.geometry` table on Big Query. The table schema is as follows:

Fieldname	Type	Description
grid	STRING	WRF Grid ID Number
xy	GEOGRAPHY	Centroid of the grid (PostGIS point) )
geohash	String	PostGIS geohash string approximating grid boundary
geometry	STRING	Well known text format of the grid boundary

An example query to return the Grid ID covering the Seattle Center:

```

WITH
  SeattleCenter AS (
  SELECT
    ST_geogpoint(-122.35,
      47.62) AS location)
SELECT
  grid
FROM

```

```
`tnc-data-v1.gfdl.geometry`  
WHERE  
    ST_DWITHIN(ST_GEOGFROMTEXT(geometry),  
    (  
        SELECT  
            location  
        FROM  
            SeattleCenter),  
    0)
```

Returns the grid ID pertaining to this location:

Row	grid	1	ID16_V9
1			