Escherichia coli and Dissolved Oxygen Trends in the Upper Llano River Watershed, Texas *

Michael Schramm *Texas A&M University AgriLife Research and Extension Service, Texas Water Resources Institute*
Tyson Broad *Texas Tech University, Llano River Field Station*
Tom Arsuffi *Texas Tech University, Llano River Field Station*

This document provides an introduction to R Markdown, argues for its benefits, and presents a sample manuscript template intended for an academic audience. I include basic syntax to R Markdown and a minimal working example of how the analysis itself can be conducted within R with the `knitr` package.

Keywords: Llano River, E. coli, Dissolved Oxygen, Water Quality

Contents

Introduction

The Upper Llano River Watershed includes the North and South Llano rivers and is located in the central Texas hill country (Figure 1). Both rivers have high water quality are supportive of diverse aquatic communities and sustain locally important recreational opportunities. Although both rivers meet existing water quality standards, local stakeholders voiced concern about changing land uses, proliferation of invasive species, and pressures on water supply as local populations grow. These concerns lead to the formation of local stakeholder organizations and the development of the Upper Llano River Watershed Protection Plan (ULRWPP). The URLWPP, accepted by the Environmental Protection Agency in Summer 2016, identified and recorded those key concerns raised by local stakeholders. Amongst the surface water quality issues identified in the plan were concerns about low dissolved oxygen (DO) and the elevated indicator bacteria, *Escherichia coli* (*E. coli*). Since 2016, implementation efforts in the watershed to reduce *E. coli* bacteria loads and potential contributors to depressed DO have expanded. Amongst these efforts are enrolling agricultural producers into Natural Resource Conservation Service (NRCS) conservation plans or prescribed grazing plans intended to reduce runoff, protect riparian areas, increase soil health, and reduce livestock time spent in streams. Although limited water quality sampling has occurred, this report intends to provide an understanding of bacteria and dissolved oxygen trends through 2016. The purpose of this report is to evaluate changes in water quality concentrations over time using simple statistical regression methods and to evaluate potential correlations with dissolved oxygen concentrations.

*A technical report prepared in cooperation with the Texas State Soil and Water Conservation Board under funding from a US EPA 319(h) nonpoint source protection grant

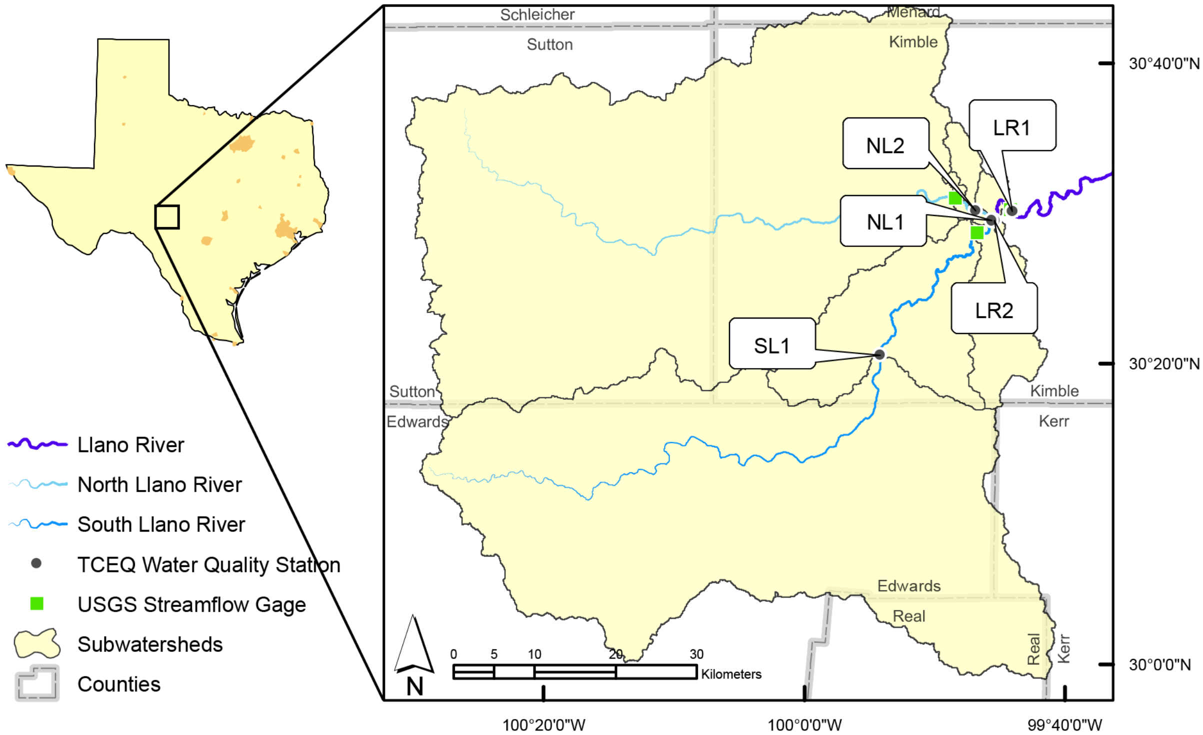

Figure 1: Study area

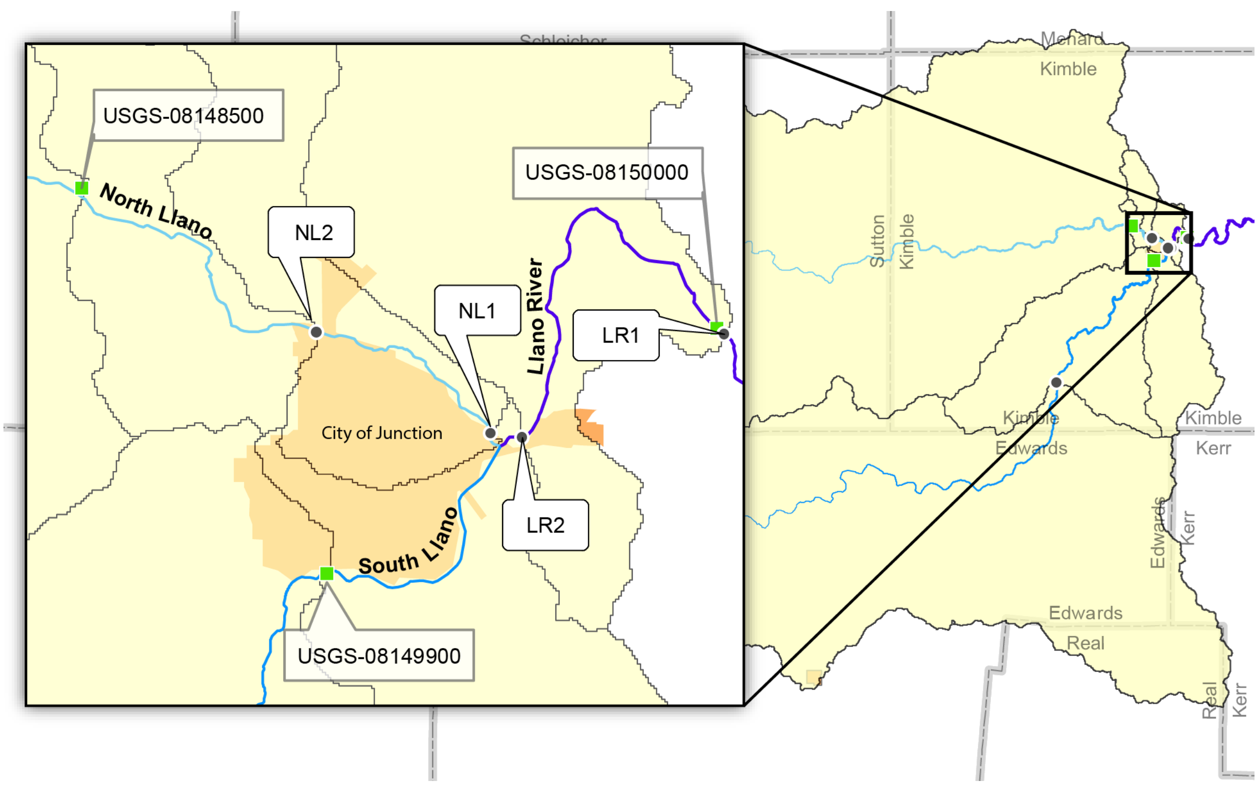

Figure 2: Detailed map of stream gage and monitoring station locations near Junction, Tx

Table 1: USGS gage summary information

USGS Site No.	Q_{min}	Q_{max}	$\bar{Q}$	$\tilde{Q}$	Drainage Area (km^2)	Name	Period of Record
08148500	0.0	13000	37.39	16.0	2332	North Llano River near Junction, TX	2001-2016
08149900	23.4	4160	70.88	49.1	2277	South Llano River at Flat Rock Ln at Junction, TX	2012-2016
08150000	31.8	45900	146.11	95.6	4809	Llano River Near Junction TX	2001-2016

Methods

Study Location and Data Sources

The Upper Llano Watershed includes the catchments of both the North Llano and South Llano River (Figure 1). For purposes of this report, the watershed was extended downstream to capture water quality trends of the combined flows from both watershed. The project watershed was 4,809 km^2 (1,188,330 acres). We obtained mean daily streamflow from three gages in the study area from the USGS National Water Information System (NWIS) Figure (1, Figure 2, and Table 1). The Texas Commission on Environmental Quality (TCEQ) water quality data was obtained through the Clean Rivers Program data tool. Although a number of water quality monitoring stations occur within the watershed, stations with less than one year of data were omitted because they could not provide meaningful insight on temporal concentration trends. Records from five water quality stations were included in the analysis (Figure 1, Figure 2, and Table 2). Although NL2 includes a limited number of available sample events, it was included because sampling at the long-term station NL1 was recently discontinued due to site access issues. Water quality monitoring data used in this report were collected by TCEQ, Lower Colorado River Authority, Texas State University, and other Clean Rivers Program Partners. The collection of these data are conducted under TCEQ and EPA approved Quality Assurance Project Plans.

Streamflow Estimation

Mean daily streamflows for ungaged water quality monitoring sites (LR2, NL1, NL2, SL1) were developed using the Drainage-Area Ratio (DAR) method. DAR provides a simple method of estimating streamflows at ungaged sites, in particular for nested subswatersheds (Hirsch (1979), Ries and Friez (2000), and Emerson, Vecchia, and Dahi (2005). DAR calculates streamflow discharge as:

$$Y_i = X_i(\frac{A_y}{A_x})^{\phi}$$

where Y_i is the calculated discharge at the ungaged site at time i, X_i is discharge measured at the gaged site at time i, while $(\frac{A_y}{A_x})$ is the ratio of ungaged and gaged catchment areas. Values used for $(\frac{A_y}{A_x})$ are indicated in Table 2.

In most applications, ϕ is assumed to equal one (Asquith, Roussel, and Vrabel 2006). However, when applied to Texas streams, ϕ = 1 results in substantial over- or under-estimation at the largest magnitude streamflows (Asquith, Roussel, and Vrabel 2006). Therefore, we applied values for ϕ based on streamflow percentile values as indicated by Asquith, Roussel, and Vrabel (2006) (Table 3). To facilitate interpretation of hydrological conditions as each station, flow duration curves were plotted and converted to load duration curves overlaid with water quality measurements (Morrison and Bonta 2008). The load duration curves were developed using Texas's primary contact recreation criterion for E. coli bacteria in surface freshwater bodies of 126 cfu/100 mL. The duration curves provide a visual analysis of flow conditions at stations, as well as an understanding of what flow conditions water quality concentrations and loads exceed established water quality standards.

Correlation Analysis

Monotonic relationships between dissolved oxygen concentrations, log transformed streamflow, and stream temperature were evaluated graphically and with Kendall's Tau (Helsel and Hirsch 2002). Censored values and outliers were not filtered due to the resistance of rank-methods such as Kendall's Tau to these values. Correlations were considered significant at the $p < 0.05$ level. All statistical analysis and plots were completed in R, versions 3.4.0 (R Core Team 2016).

Concentration and Load Trends

We evaluated trends in bacteria concentrations collected at each station with simple linear regressions of log transformed concentrations against date (Helsel and Hirsch 2002). Dissolved oxygen concentrations were evaluated using untransformed values regressed against date. The null hypothesis of zero slope over time was rejected at $p < 0.05$. Scatter plots of concentrations and dates are provided to aid result interpretation. We verified linear regression assumptions of normality with qqplots and by plotting model residuals against dates. LR2 and NL1 were the only stations with enough data to estimate concentrations and loads using regression based techniques. We estimated instream mean daily bacteria loads using the Weighted Regressions on Time, Discharge and Season (WRTDS) approach with the EGRET package in R (R.M. Hirsch, Moyer, and Archfield 2010; R.M. Hirsch and De Cicco 2014). We used WRTDS outputs to describe trends in concentration and loads over time. WRTDS is a regression-based approach that models log transformed water quality constituents concentrations as a function of time, flow, and season. The functional model is:

$$ln(E.coli) = \beta_0 + \beta_1 T + \beta_2 ln(Q) + \beta_3 cos(2\pi t)$$

Where T is decimal time and included in the model as an annual term (β_1) or seasonal term (β_3, β_4), and Q is flow. WRTDS differs from other regression based estimators, such as LOADEST, by using a weighted regression to estimate coefficients for each combination of Q and t. The coefficeints are used to estimate daily concentrations, daily loads, and flow-normalized loads and concentrations. Flow-normalized values are derived from removal of variation in concentration based on random variations in discharge. Further details of WRTDS are outlined in (R.M. Hirsch, Moyer, and Archfield 2010). These flow normalized values are useful to understand how changes in land practices over time affect instream constituent concentrations and loads. In particular flow normalized values are less influenced by increases in loads solely attributed to increases in stormflow runoff. Flow-normalized values provide a useful information that are not captured by regression on concentration values alone, which can be influenced by changes in precipitation and runoff and might not reflect progress in land management activities. We estimated instream dissolved oxygen concentrations at LR and NL1 using generalized additive mixed models (GAMMs) (Wood 2008). GAMMs are regression models that use non-linear smooth functions on covariates. Dissolved oxygen concentrations were not expected to be a function of flow, but a function of stream temperature, season, long-term trend, and instream metabolic processes. Therefore, we utilized GAMMs to describe dissolved oxygen concentration because it allows the inclusion of specified terms and random effects. The generalized GAMM model used to fit dissolved oxygen concentrations is:

$$DO = \beta_0 + \beta_1 ln(Q) + \beta_2 Year + \beta_3 Month + \beta_4 Temperature$$

Where Q is flow, $Year$ is year term, $Month$ is the seasonal month term, and $Temperature$ is the water temperature term. Appropriateness of GAMM model fits were evaluated by inspecting model residuals following procedures described by Zuur et al (2009).