Statin and Beta-blocking Agents in West Virginia Surface Waters

Joe Kingsbury

April 9, 2021

KEYWORDS beta-blockers, linear regression, pharmaceuticals, statins, water quality, waste water treatment plant.

Pharmaceutical pollution is a growing concern regarding the potential impacts it might have on aquatic communities. Thus, it is important to understand if there are ways to predict the environemental concentration of pharmaceutical residues and their active metabolites. With this in mind there are several questions we wish to try and answer regarding pharmaceuticals concentrations in the environment. Are traditional water quality variables (tempterature, flow, conductivity, etc.) strong predictors of the environmental concentratiosn of statins and beta-blockers? Are there significant differences between the various sampling positions at each site and are there differences in concentrations between cities?

STUDY AREA

Surface water grab samples were collected in 3 separate river systems, the Buckhannon, the West Fork, and the Tygart Valley River. All three river systems possess some kind of flow control mechanism and experience similar flow regimes. More specifically, the sampling locations were located in the West Virginia cities of Buckhannon, Weston, and Elkins (Figure 1). Each city has a primary wastewater treatment plant(WWTP) that handles wastewater from households, hospitals, and other various industrial sources.

METHODS

Experimental Design

Surface water grab samples and water quality readings were collected on a monthly basis from August-November 2020. Surface water samples were collected from the WWTP influent (raw sewage), the WWTP effluent (treated water), 150m upstream of the WWTP discharge pipe, and 150m downstream of the discharge pipe. Water quality measurements were taken at the upstream collection, downstream collection, and effluent discharge area (Figure 2). A total of 48 grab samples (N = 48) and a total of 36 water quality readings were taken (N = 36). Water quality was not recorded for the influent due to COVID-19 safety concerns.

Statisical Methods

In order to answer all of the previously mentioned research questions a series of linear models were used under the assumption that the environmental concentrations of all 4 pharmaceuticals are Guassian random variables:

$$y_{ator},\ y_{simv},\ y_{meto},\ y_{carv} \sim \mathsf{Gaussian}(\mu,\ \sigma^2)$$

The residuals for the pharmaceutical concentrations is also assumed to be a Gaussian distribution:

$$\epsilon_{pharm} \sim (0,\ \sigma^2)$$

Water Quality Estimators

We recorded several types of water quality measurements: flow(cfs), temperature(C°), conductivity (µS/cm), Total Dissolved Solids (TDS, mg/L), and pH. The month was also recorded each time samples were collected. The goal was to use see if there is a relationship between water quality parmaters and pharmaceutical concentrations. With this in mind the pharmaceutical concentrations were the response variable and the water quality measurements were the predictor variables with the assumption that the data was normally distributed and thus we used the following Gaussian linear model:

$$y_{pharm} = \beta_0 + \beta_{month} + \beta_{cfs} + \beta_{temperature} + \beta_{conductivity} + \beta_{TDS} + \beta_{pH} + \epsilon_{pharm}$$

This initial model was treated as a starting point to include all parameters but there was no guarantee it was the model of best fit and thus multiple models with fewer parameters than this one were tested. The β_{month} coefficient was treated as a categorical variable to include the months of August, September, October and November. The other vairalbes in the linear equation are all continuous variables. From here environmental variables were removed 1 at a time to create multiple models for Second-order Akaike Information Criteria (AICc) testing. Second order was chosen because the data set is small. Only models with an ($AIC\Delta \leq 2$) were further evaluated for best fit model. If multiple models met this threshold then they were further evualuated by their respective AICc weight and the model with the greatetest AICc weight was chosen. A model was selected for each of the 4 pharmaceuticals resulting in 4 separate linear models. Coefficients for all models were noted as signficiant if $\alpha \leq 0.05$ after performing a Student's t-test. All models were generated using base R and the linear model function.

Sites and Sampling Postion Differences

Grab samples were collected at 3 different wastewater treatement plants (WWTPs), also referenced as sites (s), and at each individual plant there were 4 different positions from which grab samples were taken. These 4 positions (p) are referenced as influent (raw sewage), effluent (treated water), upstream (of the discharge pipe), and downstream (downstream of the discharge pipe). Our response variable is the environmental concentration at all sites and positions and our predictor variables are the factors of site and position. In order to compare the differences in environmental concentrations at each site and at all 4 positions, contrasts between these two factors were calculated from the following linear model assuming Gaussian random variables:

$$y_{sp} = \beta_0 + \beta_{site} + \beta_{position} + \epsilon_{sp}$$

The residuals for our contrasts are also assumed to be Gaussian random variables:

$$\epsilon_{sp} \sim (0,\ \sigma^2)$$

For the contrasts we assume that the linear combinatin of multivariate randome variables are also Guassian:

$$XY_{sp} \sim \mathsf{Gaussian}(\mathsf{X}\mu, \mathsf{X}\Sigma\mathsf{X}')$$

X represents the contrast matrices, Y represents the slope coefficients from each model, and Σ represents the covariance matrices between the site and position parameters.

Testing for significance between these various factors we used a Wald's test setting the alpha-level at ($\alpha \leq 0.05$). The glht function was used in the package MULTCOMP in R to perform multiple contrast hypothesis tests. Significant variables should indicate that there are important differences between cites and/or sampling positions. Differences in the effluent and influent as well as differences between the upstream and downstream are key areas of interest.

Downstream Concentration Prediction Models

The final research question of interest is if there is a possibility of effectively predicting the environmental concentration of these pharmaceuticals downstream from the WWTPs. Using the downstream concentration of each pharmaceutical as the response variable and the upstream, influent, and effluent concentrations as

predictor variables the following Guassian linear model was examined as a potentially useful model:

$$y_{downstream} = \beta_{upstream} + \beta_{effluent} + \beta_{influent} + \beta_{effluent*influent} + \epsilon_{downstream}$$

An interaction term exists between influent and effluent since influent directly effects how much pharmaceutical waste is present in the treated effluent water. While it is not possible to accurately predict the rate of pharmaceutical waste removal the interaction is still significant and needs to be taken into account. Student's t-test was utilized to calculate both a test statistic and a p-value for each predictor variable as well s the overall p-value of the model ($\alpha \leq 0.05$). At the same time we are interested in seeing which predictor variables are significant in predicting changes in our downstream concentrations as well as seeing how much variance exists in our model by inspecting the confidence intervals. Confidence intervals were assessed at 95%.

RESULTS

For this study we collected a total of 36 water quality measurements and obtained a total of 48 1L grab samples across 3 wastewater treatment facilities at 4 different sampling positions per site. This project is intended to be a pilot study with these results showing potential relationships that justify further investigation. Significant results were obtained in all 3 analyses and they will be discussed in the order that they were initially presented for clarity.

Water Quality Models

Water quality variables displayed some predictability of environmental pharmaceutical concentrations. Using equation (1) in conjuction with a second-order AIC (AICc) the model of best fit was selected for each pharmaceutical under the condtion that ($\Delta\mathsf{AIC} \leq 2$). The results of the AICc (Table 1.) revealed that models containing flow (cfs), tempterature (C°), and conductivity (µS/cm) were the best fit. As seen in the model outputs not all predictor variables were significant even in the model of best fit (Table 2.). Carvedilol's best fitting model used solely the 'Month' variable, but it was ruled out because it was not likley to be a good predictor with only 4 months of data. 3 out of the 4 water quality predictors displayed an inverse relationship with the pharmaceutical concentrations with the exception of Metoprolol which displayed a positive relationship with conductivity (figure 3.). The significance of these findings is that flow and temperature appear to be key influencers on pharmaceutical concentration and by monitoring these

variables it might be possible to roughly estimate the current concentrations pharmaceutical concentrations in the environment without having to directly test the surface water. For further information on the model selection and the models chosen refer to tables 1 and 2.

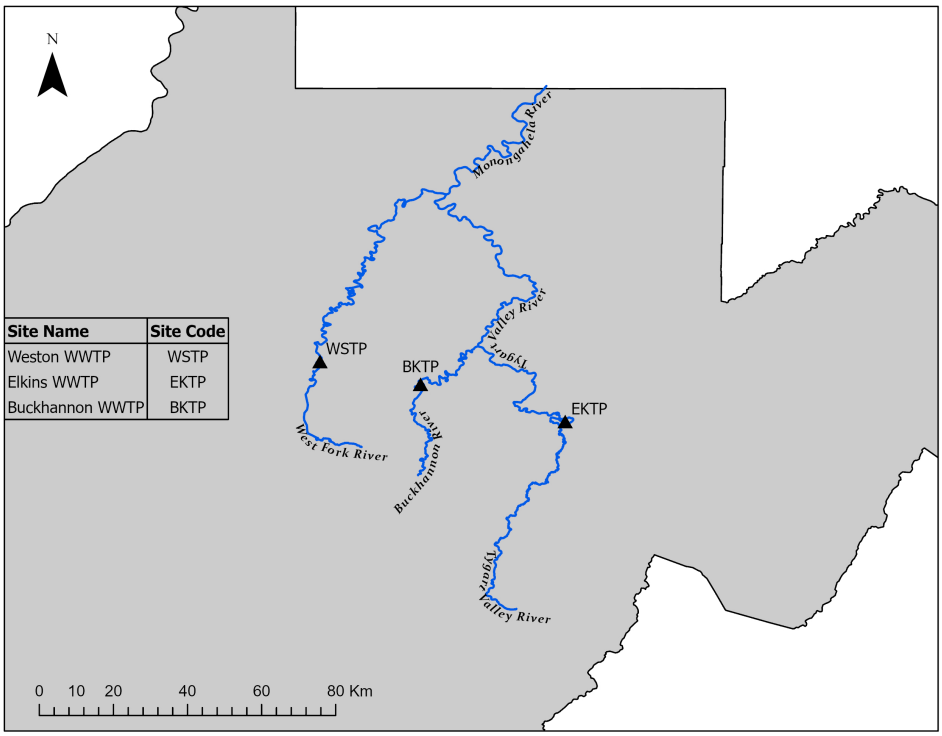
Site and Position Comparisons

Site and position contrasts revealed a plethora of interesting information as it relates to concentration differences between cites and between positions relative to the WWTP. All of the linear models used to calculate the contrasts displayed significant p-values ($\alpha \leq 0.05$)(Table 3.). At the same time influent concentrations for all 4 pharmaceuticals were significant indicating that the influent has much greater quanities of pharmaceuticals than the rest of the positions we measured. At the same time when we compared the influent to the effluent it often came back significant which indicates a large poportion of this waste is being treated by our WWTPs (Table 4.). Another key result was that there are significant differences between each WWTP in terms of pharmaceutical concentrations. This result most likely indicates that the input sources are varied between each plant or that one plant is more efficient at treating pharmaceutical waste than the others. For a full breakdown on all of these numerical results use tables 3 and 4.

Downstream Concentrations Models

Predicting downstream concentrations based on quantified inputs is an imporant part of determining pollution sources and areas of concern. Using the influent, effluent, and upstream concentrations to predict downstream concentrations provided a more accurate model than water quality parameters alone. While few individual predcitors were significant for any of the 4 models, the overall model for Atorvastatin and Metoprolol was significant (Table 5.). Visually comparing each model with their 95% confidence intervals makes it obvious which models do a better job accounting for the variance (Figure 4.). At the same time the average concentration value of Atorvastatin and Metoprolol was significantly higher than Simvastatin or Carvedilol. These results indicate that Metoprolol and Atorvastatin persist in higher concentrations in our surface waters. While these concentrations are in ng/L some aquatic organisms can still experience detrimental effects from even trace concentrations. As our populations increase it is reasonable to assume that the influent concentration will increase as more individuals excrete these pharmaceuticals leading to higher concentrations in our surface waters.

FIGURES AND TABLES



Site Name	Site Code
Weston WWTP	WSTP
Elkins WWTP	EKTP
Buckhannon WWTP	BKTP

Figure 1: Site map of study area.

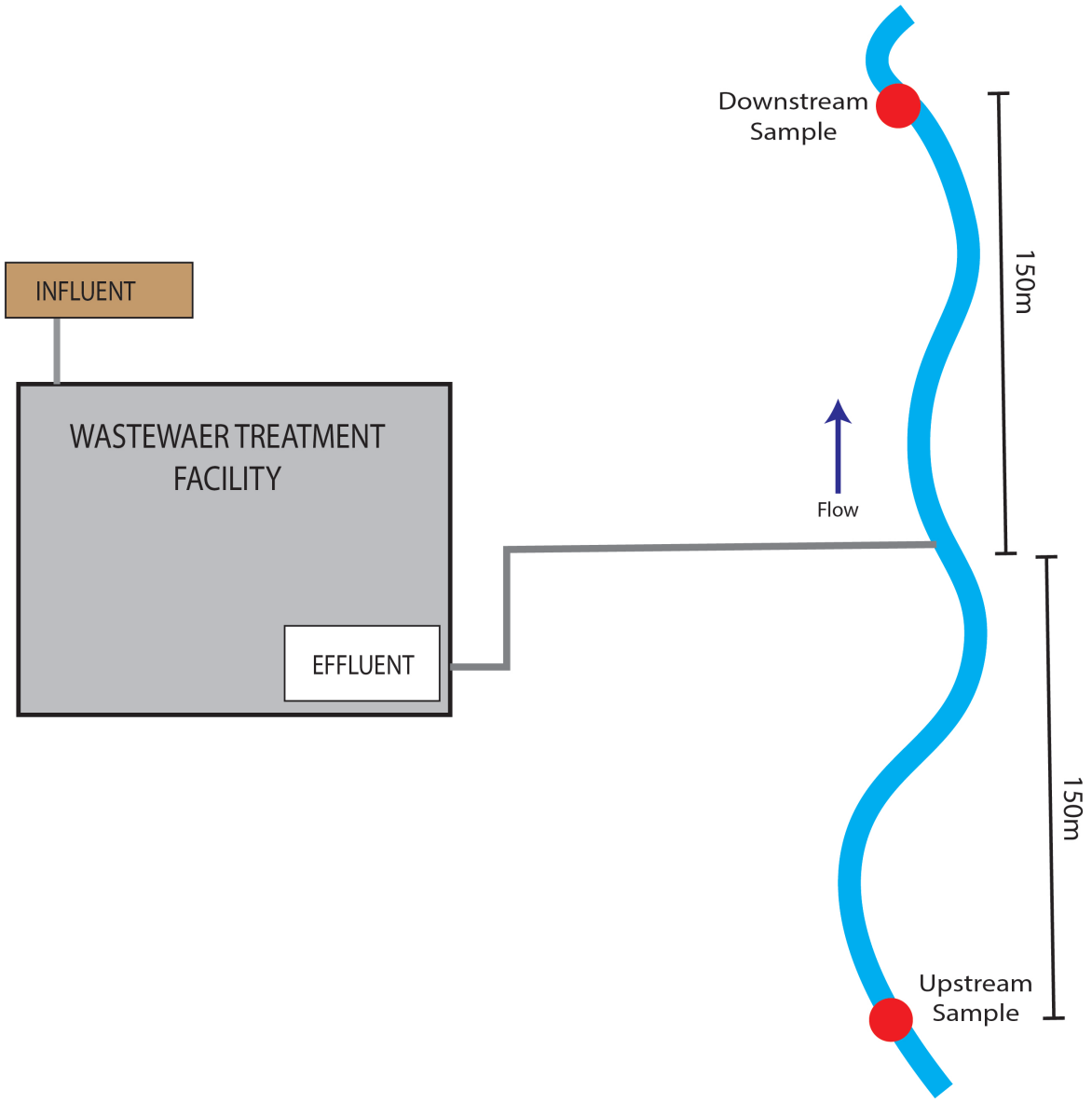


Figure 2: Diagram of collection sites at each waste water treatrement facility.

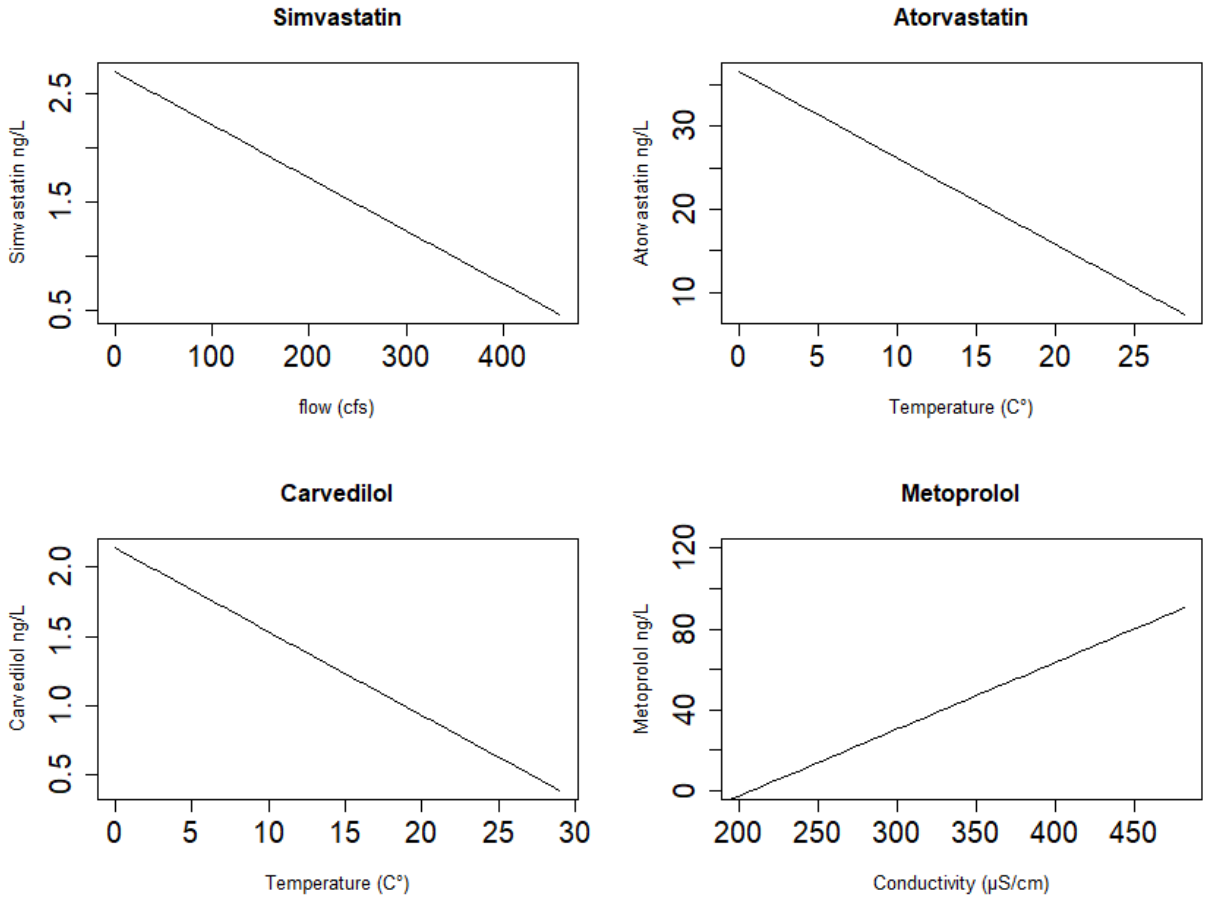
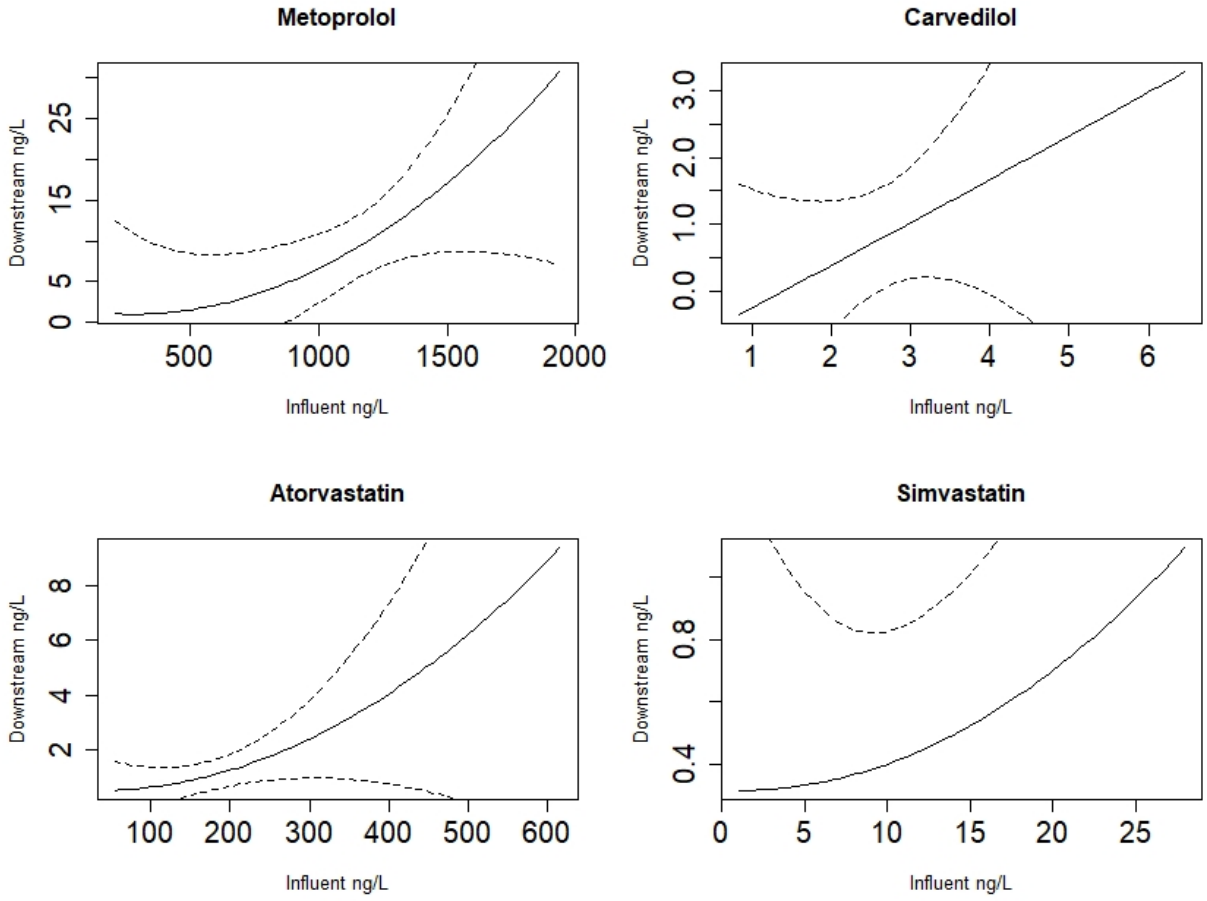


Figure 3: Environmental predictors relationship to each pharmaceutical based on model of best fit.



126

127 Figure 4. Predicted downstream concentrations of all 4 pharmaceuticals plotted against influent concentra-

128 tions. 95% confidence intervals are denoted by the dashed lines.

Metoprolol AICc Scores					Atorvastatin AICc Scores				
Model Names	K	AICc	ΔAICc	AICcWt	Model Names	K	AICc	ΔAICc	AICcWt
M10	3	364.60	0.00	0.84	M9	3	402.22	0.00	0.28
M11	5	367.95	3.35	0.16	M7	3	402.25	0.03	0.28
M3	8	376.03	11.43	0.00	M10	3	402.36	0.14	0.26
M6	8	377.56	12.97	0.00	M5	4	404.15	1.93	0.11
M2	9	379.04	14.44	0.00	M8	5	406.40	4.18	0.03
M1	10	382.32	17.73	0.00	M11	5	406.73	4.51	0.03
M9	3	389.17	24.57	0.00	M4	7	412.00	9.78	0.00
M7	3	389.74	25.14	0.00	M6	8	415.27	13.04	0.00
M4	7	391.55	26.95	0.00	M3	8	415.28	13.06	0.00
M5	4	391.67	27.07	0.00	M2	9	418.83	16.61	0.00
M8	5	395.02	30.42	0.00	M1	10	422.62	20.40	0.00
Carvedilol AICc Scores					**Simvastatin AICc Scores**				
Model Names	K	AICc	ΔAICc	AICcWt	Model Names	K	AICc	ΔAICc	AICcWt
M8	5	83.98	0.00	0.56	M7	3	216.29	0.00	0.37
M5	4	86.28	2.31	0.18	M10	3	217.18	0.89	0.24
M9	3	86.74	2.76	0.14	M9	3	217.22	0.93	0.23
M11	5	88.68	4.71	0.05	M5	4	218.79	2.50	0.11
M4	7	88.88	4.91	0.05	M11	5	221.50	5.21	0.03
M3	8	91.86	7.89	0.01	M8	5	221.80	5.51	0.02
M6	8	91.93	7.95	0.01	M4	7	227.12	10.83	0.00
M2	9	95.32	11.34	0.00	M6	8	230.40	14.11	0.00
M7	3	96.79	12.82	0.00	M3	8	230.43	14.14	0.00
M1	10	99.03	15.05	0.00	M2	9	233.73	17.45	0.00
M10	3	102.29	18.31	0.00	M1	10	237.59	21.30	0.00

Table 1. Simvastatin environmental model AICc scores and weights for each model.

Simvastatin ~ cfs				
Coefficients:	Estimate	Std. Error	t-value	p-value
(Intercept)	2.697	1.027	2.626	0.013
cfs	-0.005	0.005	-0.945	0.352
F-statistic:	0.892		p-value:	0.3515
Multiple R-squared:	0.026	Adjusted R-squared:		-0.003

Atorvastatin ~ cfs + Temperature				
Coefficients:	Estimate	Std. Error	t-value	p-value
(Intercept)	49.554	35.130	1.411	0.168
cfs	-0.054	0.071	-0.751	0.458
Temperature	-1.395	1.808	-0.772	0.446
F-statistic:	0.458		p-value:	0.636
Multiple R-squared:	0.027	Adjusted R-squared:		-0.032

Carvedilol ~ cfs + Temperature				
Coefficients:	Estimate	Std. Error	t-value	p-value
(Intercept)	2.125	0.425	5.000	0.000
cfs	0.001	0.001	1.693	0.100
Temperature	-0.083	0.022	-3.797	0.001
F-statistic:	11.12		p-value:	0.000
Multiple R-squared:	0.403	Adjusted R-squared:		0.366

Metoprolol ~ Conductivity				
Coefficients:	Estimate	Std. Error	t-value	p-value
(Intercept)	-37.434	13.165	-2.844	0.008
Conductivity	0.331	0.056	5.890	0.000
F-statistic:	34.69		p-value:	0.000
Multiple R-squared:	0.505	Adjusted R-squared:		0.490

Table 2. Linear model results for all 4 pharmaceuticals showing which evnironmental prediction models were the best fit and their associated predictors for each pharmaceutical.

Simvastatin ~ Site + Postion

Coefficients:	Estimate	Std. Error	t-value	p-value
(Intercept)	-0.120	1.771	-0.068	0.947
SiteElkins	-1.353	1.771	-0.764	0.449
SiteWeston	3.159	1.771	1.784	0.082
PositionEffluent	4.639	2.045	2.268	0.029
PositionInfluent	11.424	2.045	5.585	1.57E-06
PositionUpstream	0.053	2.045	0.026	0.980

F-statistic:	9.695	p-value:	3.29E-06
Multiple R-squared:	0.5358	Adjusted R-squared:	0.481

Atorvastatin ~ Site + Postion

Coefficients:	Estimate	Std. Error	t-value	p-value
(Intercept)	-5.98	40.462	-0.148	0.883
SiteElkins	-7.8462	40.462	-0.194	0.847
SiteWeston	29.3062	40.462	0.724	0.473
PositionEffluent	55.0592	46.721	1.178	0.245
PositionInfluent	261.7225	46.721	5.602	1.48E-06
PositionUpstream	-0.4092	46.721	-0.009	0.993

F-statistic:	8.71	p-value:	9.94E-06
Multiple R-squared:	0.509	Adjusted R-squared:	0.451

Carvedilol ~ Site + Postion

Coefficients:	Estimate	Std. Error	t-value	p-value
(Intercept)	1.345	0.419	3.208	0.003
SiteElkins	-0.790	0.419	-1.885	0.066
SiteWeston	-0.731	0.419	-1.745	0.088
PositionEffluent	0.422	0.484	0.871	0.389
PositionInfluent	1.974	0.484	4.079	1.97E-04
PositionUpstream	-0.089	0.484	-0.184	0.855

F-statistic:	5.584	p-value:	4.95E-04
Multiple R-squared:	0.399	Adjusted R-squared:	0.328

Metoprolol ~ Site + Position

Coefficients:	Estimate	Std. Error	t-value	p-value
(Intercept)	25.259	94.507	0.267	0.791
SiteElkins	2.496	94.507	0.026	0.979
SiteWeston	-60.177	94.507	-0.637	0.528
PositionEffluent	79.739	109.127	0.731	0.469
PositionInfluent	863.75	109.127	7.915	7.35E-10
PositionUpstream	-2.894	109.127	-0.027	0.979

F-statistic:	17.96	p-value:	1.71E-09
Multiple R-squared:	0.681	Adjusted R-squared:	0.643

Table 3. Linear model results for environmental concentrations of simvastatin, atorvastatin, carvedilol, and metoprolol where site and sampling position are the predictor variables.

Simvastatin Contrasts (1 == 0)				
Site	Estimate	Std. Error	t value	p-value
Elkins : Weston	4.513	1.771	2.547	0.015
Buckhannon : Weston	3.279	3.068	1.069	0.291
Buckhannon : Elkins	-1.234	3.068	-0.402	0.690
Postion				
Influent: Effluent	6.785	2.045	3.317	0.002
Upstream: Downstream	0.172	3.392	0.051	0.960

Atorvastatin Contrasts (1 == 0)				
Site	Estimate	Std. Error	t value	p-value
Elkins : Weston	37.150	40.460	0.918	0.364
Buckhannon : Weston	35.290	70.080	0.504	0.617
Buckhannon : Elkins	-1.866	70.082	-0.027	0.979
Postion				
Influent: Effluent	206.660	46.720	4.423	6.75E-05
Upstream: Downstream	5.571	77.478	0.072	0.943

Carvedilol (1 == 0)				
Site	Estimate	Std. Error	t value	p-value
Elkins : Weston	0.059	0.419	0.140	0.889
Buckhannon : Weston	-2.076	0.726	-2.860	0.007
Buckhannon : Elkins	-2.135	0.726	-2.941	0.005
Postion				
Influent: Effluent	1.553	0.484	3.208	0.003
Upstream: Downstream	-1.434	0.803	-1.787	0.081

Metoprolol Contrasts (1 == 0)				
Site	Estimate	Std. Error	t value	p-value
Elkins : Weston	-62.670	94.510	-0.663	0.511
Buckhannon : Weston	-85.440	163.690	-0.522	0.604
Buckhannon : Elkins	-22.760	163.690	-0.139	0.890
Postion				
Influent: Effluent	784.000	109.100	7.184	7.93E-09
Upstream: Downstream	-28.150	180.970	-0.156	0.877

Table 4. Contrast results using from linear models to determine differences between sites (WWTPs) and positions relative to the WWTP for all 4 pharmaceuticals in question. P-values are considered significant if ≤ 0.05.

Simvastatin

Coefficients:	Estimate	Std. Error	t-value	p-value
(Intercept)	0.294	0.539	0.545	0.603
upstream	-0.038	0.391	-0.096	0.926
effluent	-0.031	0.056	-0.544	0.603
influent	0.021	0.038	0.556	0.596
effluent: influent	0.002	0.003	0.468	0.654

F-statistic:	0.442	p-value:	0.775
Multiple R-squared:	0.202	Adjusted R-squared:	-0.254

Atorvastatin

Coefficients:	Estimate	Std. Error	t-value	p-value
(Intercept)	0.472	0.547	0.862	0.417
upstream	0.043	1.304	0.033	0.975
effluent	-0.002	0.004	-0.628	0.550
influent	0.001	0.001	0.634	0.546
effluent: influent	0.000	0.000	1.035	0.335

F-statistic:	6.284	p-value:	0.018
Multiple R-squared:	0.782	Adjusted R-squared:	0.658

Carvedilol

Coefficients:	Estimate	Std. Error	t-value	p-value
(Intercept)	-0.515	1.602	-0.321	0.758
upstream	1.019	0.430	2.372	0.050
effluent	0.371	1.054	0.352	0.735
influent	0.036	0.509	0.070	0.946
effluent: influent	0.008	0.457	0.017	0.987

F-statistic:	2.324	p-value:	0.1556
Multiple R-squared:	0.5705	Adjusted R-squared:	0.325

Metoprolol

Coefficients:	Estimate	Std. Error	t-value	p-value
(Intercept)	0.830	6.922	0.120	0.908
upstream	0.655	1.026	0.638	0.543
effluent	-0.089	0.080	-1.105	0.306
influent	0.001	0.010	0.139	0.893
effluent: influent	0.000	0.000	1.042	0.332

F-statistic:	4.188	p-value:	0.048
Multiple R-squared:	0.705	Adjusted R-squared:	0.537

Table 5. Linear model outputs for all 4 pharmaceuticals that treat the downstream concentration as the response variable and the other 3 positions at predictor variables. Interaction occurs between influent and enfluent concentrations since influent directly effects concentrations in treated waste water.