

Supplemental Information for

**Addressing the Contribution of Indirect Potable Reuse to Inland Freshwater
Salinization**

Shantanu V. Bhide¹, Stanley B. Grant*^{1,2}, Emily A. Parker¹, Megan A. Rippy^{1,2}, Adil Godrej¹,
Sujay Kaushal³, Greg Prelewicz⁴, Niffy Saji⁴, Shannon Curtis⁵, Peter Vikesland⁶, Ayella Maile-
Moskowitz⁶, Mark Edwards⁶, Kathryn Lopez⁶, Thomas A. Birkland⁷, Todd Schenk^{2,8}

A Manuscript for Nature Sustainability

¹Occoquan Watershed Monitoring Laboratory, The Charles E. Via, Jr. Department of Civil and Environmental Engineering, Virginia Tech, 9408 Prince William Street, Manassas VA, 20110 USA.

²Center for Coastal Studies, Virginia Tech, 1068A Derring Hall (0420), Blacksburg, VA 24061 USA.

³Department of Geology & Earth System Science Interdisciplinary Center, University of Maryland, College Park MD, 20740 USA.

⁴Fairfax Water, 8560 Arlington Blvd., Fairfax VA 22031 USA.

⁵Fairfax County Stormwater Planning Division, Dept. Public Works and Environmental Services, 12000 Government Center Parkway, Suite 448., Fairfax VA 22035 USA.

⁶The Charles E. Via, Jr. Department of Civil and Environmental Engineering, Virginia Tech, 750 Drillfield Blacksburg, VA, 24061 USA.

⁷College of Humanities and Social Sciences, North Carolina State University, 2526 Hillsborough Street, Suite 102, Raleigh NC 27607

⁸Urban Affairs and Planning, School of Public and International Affairs, Virginia Tech, 212 Architecture Annex (0113), 140 Otey Street NW, Blacksburg VA 24061

*Corresponding author: stanleyg@vt.edu; ORCID ID: 0001-6221-7211

Section S1. MLR Model Validation and Stationarity Analysis

At stations ST10 and ST45 we validated the MLR model using the hold-out method. The entire dataset at these two stations (including timeseries of all relevant dependent and independent variables) were split into two groups: (a) calibration set (75%, used to train the MLR model) and (b) external validation (hold-out) set (25%, independent of the calibration set). The top-ranked MLR models (trained with the calibration set) were then used to generate instantaneous predictions (and associated errors) for sodium concentration and mass loads for the external validation period using USGS-Loadflex (Figures S2 and S3). The following statistical parameters were evaluated as part of the external validation at ST10 and ST45:

(a) root mean squared error of prediction (RMSEP):

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n_{EXT}} (y_{i,p} - y_i)^2}{n_{EXT}}} \quad (S1)$$

where n_{EXT} is the total number of data points in the external set, $y_{i,p}$ is the i^{th} predicted value and y_i is the corresponding observed value of log transformed sodium concentration, and

(b) prediction squared correlation coefficient (Q^2), which is a measure of the model's predictive power [74-76]:

$$Q^2 = 1 - \frac{\sum_{i=1}^{n_{EXT}} (y_{i,p} - y_i)^2 / n_{EXT}}{\sum_{i=1}^{n_C} (y_i - \bar{y}_C)^2 / n_C} \quad (S2)$$

where n_C is the total number of data points in the calibration set and \bar{y}_C is the mean of observed values in the calibration set.

Equation S2 satisfies the mathematical properties of ergodicity (observations in the external validation set are independent of each other) [74] and is corrected for the underlying assumption that the external dataset is not uniformly distributed over the entire range of the calibration dataset [74]. The metric used to gauge external validation is the difference between

hold-out R^2 and calculated Q^2 (equation S2). A large difference between the two values ($R^2 - Q^2 > 0.3$) indicates that the model might: (a) suffer from overfitting, (b) contain a large number of outliers or (c) include irrelevant covariates [75,123]. Both models at ST10 and ST45 performed satisfactorily ($R^2 - Q^2 < 0.3$). Metrics for external validation are reported in Table S1.

Table S1. External (hold-out) validation metrics

	ST10	ST45
Model performance metrics		
RMSEP	0.315	0.148
Hold-out R^2	0.15	0.81
Q^2	0.05	0.73
$R^2 - Q^2$	0.10	0.08

Because of the limited number of sodium measurements on UOSA's daily composite samples (N=68), to evaluate the MLR model for this dataset we used leave-one-out cross validation coefficient of determination (LOOCV- R^2), which was evaluated as follows: (Step 1) omit one data point and fit the MLR model to the remaining dataset and calculate model R^2 ; (Step 2) repeat Step 1 for all data points in the dataset; (Step 3) estimate LOOCV- R^2 by calculating the average of all R^2 values from Step 2. The LOOCV- R^2 for UOSA is 51.6%.

Stationarity. Both the MLR and copula analysis described in the main text assume that the time series are stationary [124]. While we know that sodium concentration in the reservoir is not stationary over the past 40 years (see rising trend in Figure 1b, main text), the relevant time series measurements may be stationary over the period encompassing the MLR and copula analyses (2010 to 2018). To test this idea, we conducted the following unit root tests [125]: (a) ADF test, (b) PP test, and (c) KPSS test. A time-series process, when mathematically expressed by a series of monomials each corresponding to a root, may contain a root equal to 1, called a unit root [126]. ADF and PP test the null hypothesis that a unit root is present in the time-series

sample (and hence the process is not stationary); the alternate hypothesis is that the time-series is stationary. A negative test statistic and a p-value < 0.05 indicates rejection of the null hypothesis and confirms the stationarity of the time-series. For the KPSS test, the null hypothesis is that the timeseries is stationary and the alternate hypothesis is that the timeseries contains a unit root. In this case, a p-value > 0.05 indicates lower evidence of rejecting the null hypothesis. The results presented in Table S2 confirm that the time-series data used for the copula analysis and MLR model generation is stationary. We did not perform these tests for UOSA's sodium concentration measurements as the sampling frequency was too sporadic and infrequent (see Figure S1).

Table S2. ADF, PP and KPSS test results.

Variable		ADF statistic	PP statistic	KPSS statistic
ln (Na)	ST10	-4.25**	-129.83**	0.33
	ST45	-5.73**	-157.83**	0.09
ln(Q_{Total})	ST10	-4.99**	-113.27**	0.21
	ST45	-5.43**	-167.81**	0.23
	UOSA	-8.53**	-632.71**	0.85
Specific Conductance ($\mu\text{S}/\text{cm}$)	ST45	-6.19**	-207.09**	0.20
	UOSA	-8.08**	-321.24**	0.12
Rainfall, maximum in the previous two weeks (inches)		-5.55**	-310.31**	0.25
Snow depth, maximum in the previous two weeks (inches)		-7.22**	-311.15**	0.07
No. of days below freezing in the previous two weeks		-7.10**	-82.37**	0.01

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

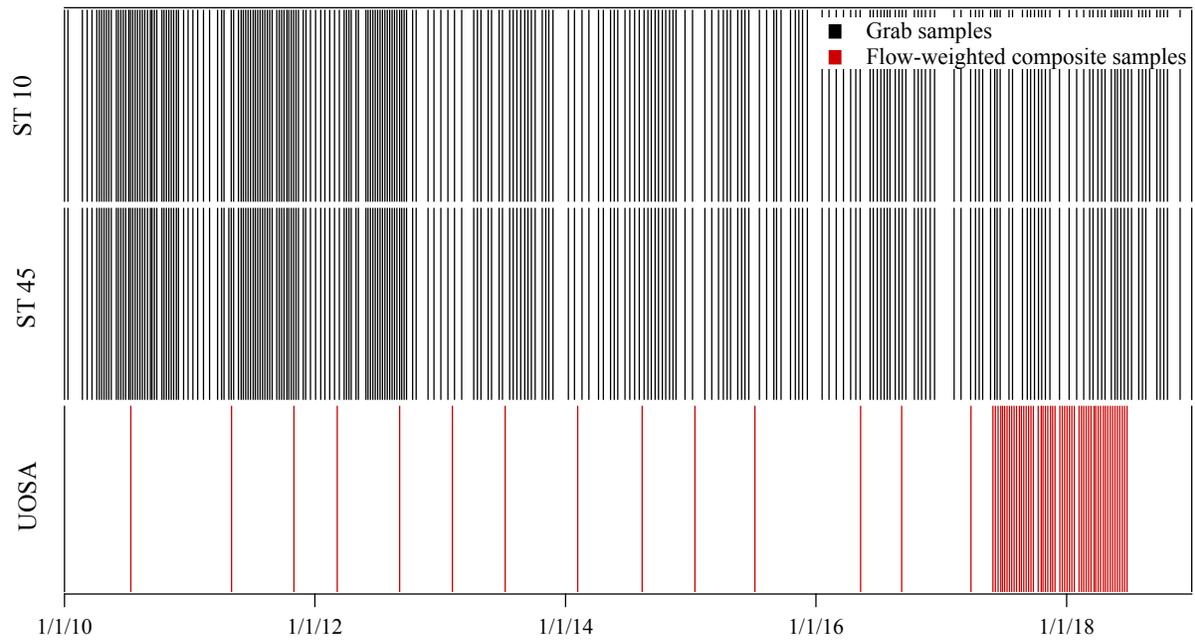


Figure S1. Schedule of sodium concentration measurements (vertical lines) available at ST10, ST45 and UOSA, respectively for the period 2010 – 2018. White spaces indicate days with no sodium concentration measurements. Black lines represent measurements collected on grab samples (ST10 and ST45) and red lines represent measurements on daily flow-weighted composite samples (UOSA).

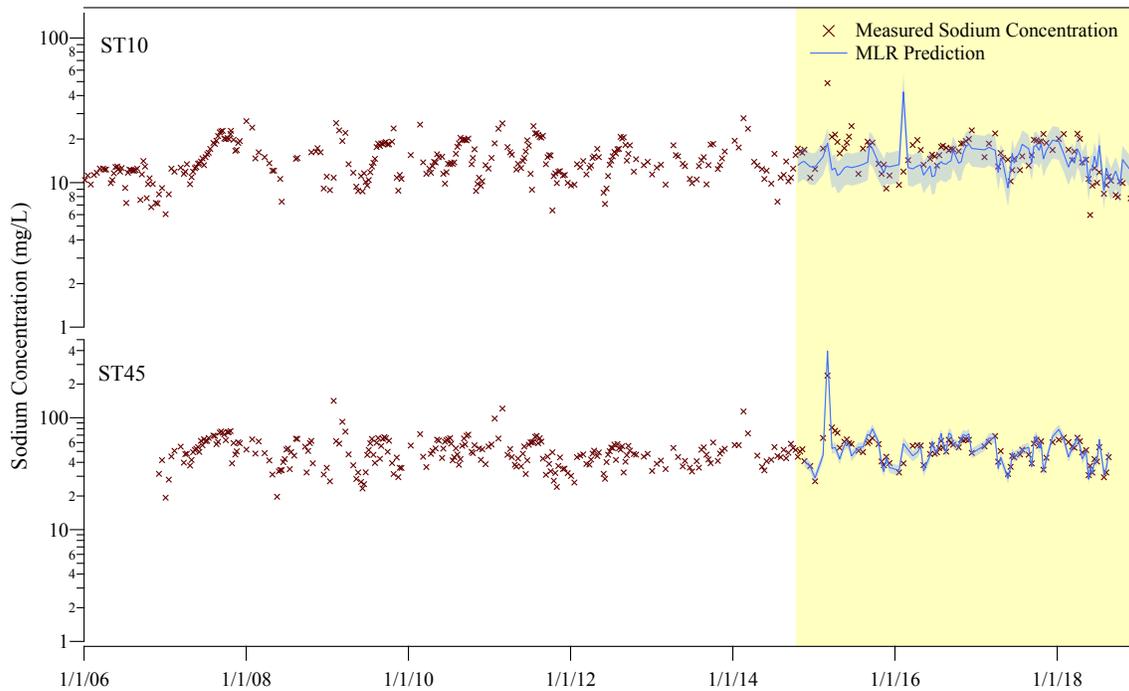


Figure S2. Observed values of sodium concentration (ST10 and ST45) and MLR model predictions from USGS-Loadflex for the validation period (tinted portion), along with standard error of prediction.

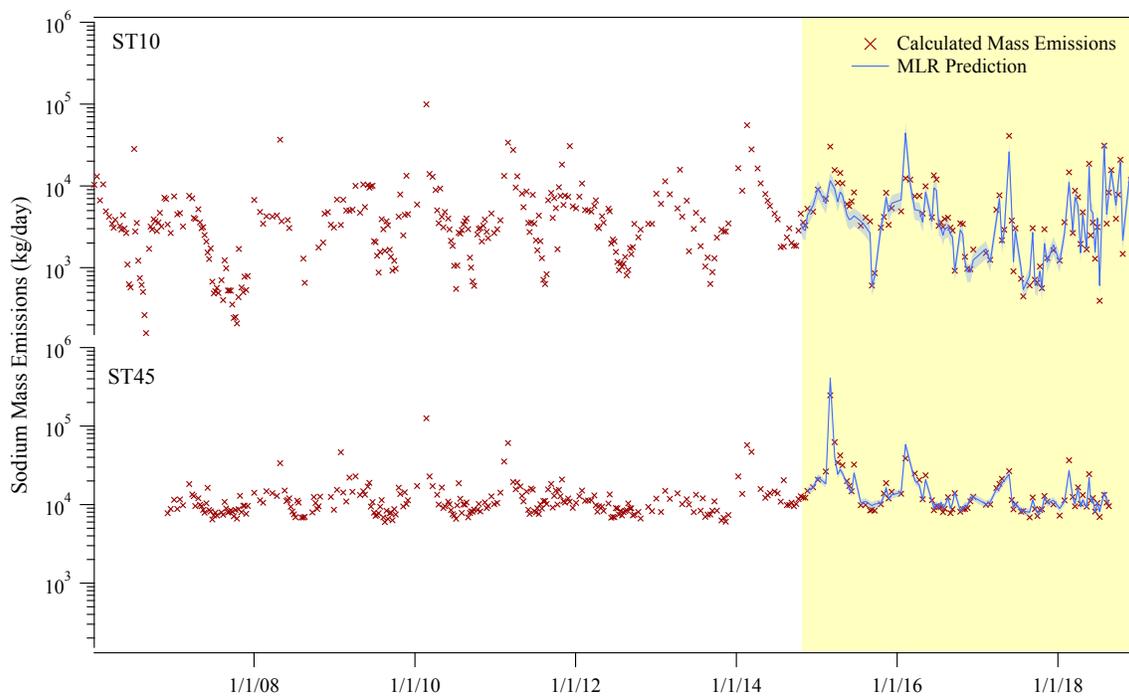


Figure S3. Calculated sodium mass emissions (from flow and sodium concentration measurements at ST10 and ST45) and MLR model predictions from USGS-Loadflex for the validation period (tinted portion), along with standard error of prediction.