# Appendix 1. Model Archival Summary for Sensor-Measured Nitrate plus Nitrite Concentration Along Indian Creek on July 23, 2014

This model archival summary summarizes the nitrate plus nitrite concentration, milligrams per liter as nitrogen ($NO_3$ + $NO_2$) model developed to correct 30-second sensor-measured $NO_3$ + $NO_2$ on July 23, 2014.

## Study Reach and Model Information

Study Reach: Indian Creek at Highway 69, Overland Park, Kansas (station number: 06893270; site identifier: Switzer) to Indian Creek near 106 Terrace, Overland Park, Kansas (station number: 385609094412600; site identifier: 24).
Location: Lat 38°55'14.07", long 94°42'20.02" to lat 38°56'9.21", long 94°41'26.29" referenced to North American Datum of 1983, Hydrologic Unit 10270104.

Equipment: A Hach Nitratax SC sensor was suspended from a kayak along the Indian Creek study reach on July 23, 2014. Readings from the sensor were recorded every 30 seconds.

Date model was created: June 1, 2016

Model calibration data period and application date: July 23, 2014.

## Model-Calibration Dataset

All data were collected using U.S. Geological Survey protocols (U.S. Geological Survey, variously dated; http://water.usgs.gov/owq/FieldManual/) and are stored in the National Water Information System database at *http://dx.doi.org/10.5066/F7P55KJN* and in King and others (2016). Ordinary least squares analysis was used to develop regression models between sensor- and laboratory-measured $NO_3$ + $NO_2$ concentrations for all sites along the reach using the open-source software package "R." Separate Nitratax sensors were used during the 2014 and 2015 spatial surveys. These sensors varied in the level of over- and/or under-estimation of $NO_3$ + $NO_2$. Therefore, all models developed to correct 30-second sensor-measured $NO_3$ + $NO_2$ are instrument specific.

The final selected regression model is based on 11 concurrent measurements of sensor-measured $NO_3$ + $NO_2$ and laboratory-measured $NO_3$ + $NO_2$ collected on July 23, 2014. Samples were collected during below-normal flow conditions. No samples were below laboratory detection limits. Summary statistics and the complete model-calibration dataset are provided below. Studentized residuals from the final model were inspected for values greater than 3 or less than negative 3. Values outside of that range were considered potential outliers and investigated. None of the $NO_3$ + $NO_2$ samples were deemed outliers.

## Nitrate plus Nitrite Sampling Details

All $NO_3$ + $NO_2$ samples for laboratory analysis were collected either using grab sample or vertically integrated at the centroid of flow every 0.5 km along the study reach. Samples were analyzed for $NO_3$ + $NO_2$ concentration at the Johnson County, Kansas Water Quality Laboratory.

## Model Development

Ordinary least squares regression analysis was done using R by examining laboratory-measured $NO_3$ + $NO_2$ concentrations as explanatory variables for correcting sensor-measured $NO_3$ + $NO_2$ concentrations. The distribution of residuals was examined for normality, and plots of residuals (the difference between the measured and computed values) as compared to computed $NO_3$ + $NO_2$ were examined for homoscedasticity (meaning that their departures from zero did not change substantially over the range of computed values).

Laboratory-measured $NO_3$ + $NO_2$ was selected to correct sensor-measured $NO_3$ + $NO_2$ based on residual plots, relatively high adjusted coefficient of determination (adjusted $R^2$), and relatively low model standard percentage error (*MSPE*). Values for all of the

aforementioned statistics and metrics were computed and are included below along with all relevant sample data and more in-depth statistical information.

## Model Summary

Summary of final regression analysis for corrected sensor-measured $NO_3 + NO_2$ concentration from Indian Creek at Highway 69, Overland Park, Kansas (station number: 06893270; site identifier: Switzer) to Indian Creek near 106 Terrace, Overland Park, Kansas (station number: 385609094412600; site identifier: 24) on July 23, 2014.

$NO_3 + NO_2$ correction-based model:

$$NO_x = 1.07 \times sensor - 0.302$$

where

$NO_x$ = corrected sensor-measured nitrate plus nitrite in milligrams per liter (mg/L); and,

*sensor* = uncorrected sensor-measured nitrate plus nitrite in milligrams per liter (mg/L).

R Output for the relation between sensor-measured and laboratory-measured nitrate plus nitrite concentrations at the Indian Creek study sites

## Model Statistics, Data, and Plots

Definitions for terms used in this output can be found at the end of this document.

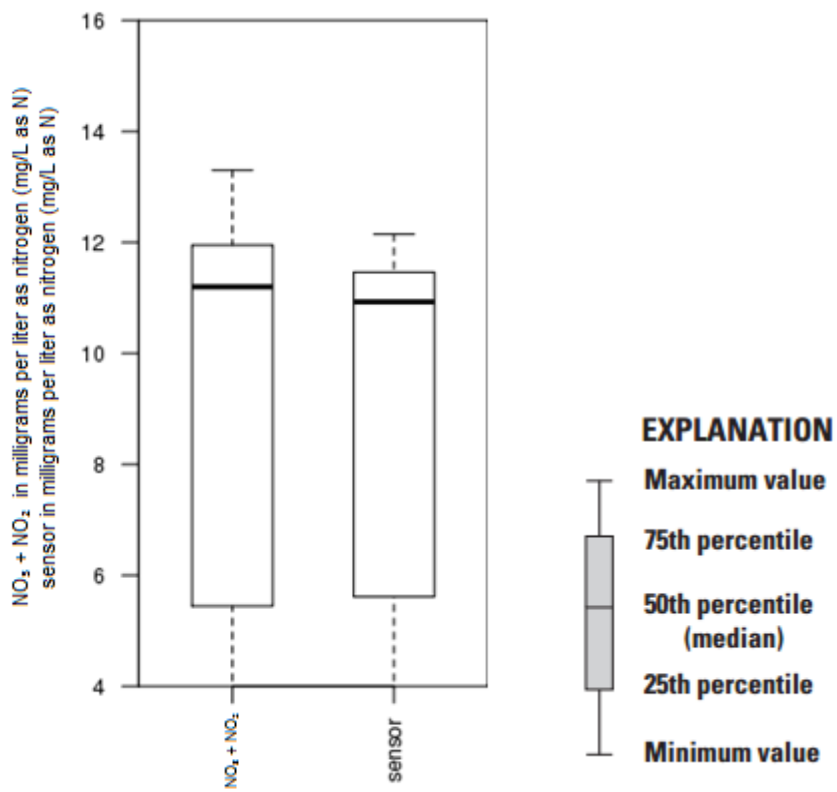
### Model

NOx = + 1.07 * sensor - 0.302

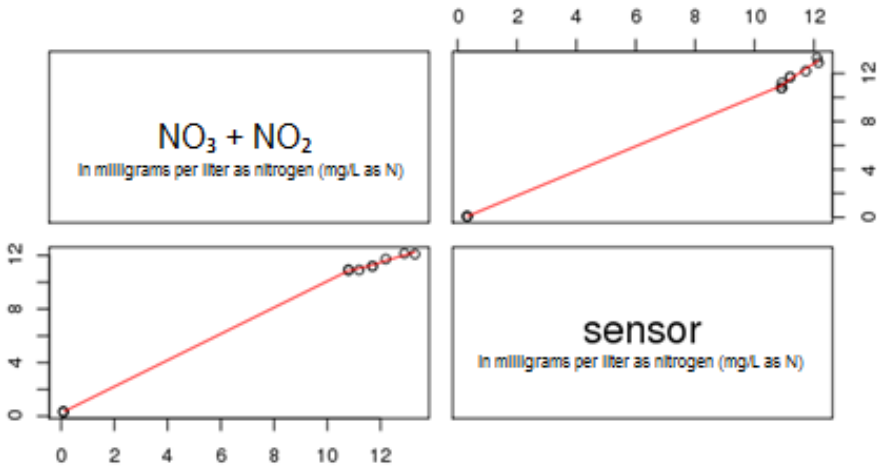
### Variable Summary Statistics

```
                NOx sensor
Minimum        0.08  0.313
1st Quartile   0.09  0.338
Median        11.20 10.900
Mean           8.62  8.370
3rd Quartile  12.20 11.700
Maximum       13.30 12.100
```

### Box Plots



## Exploratory Plots




Red line shows the locally weighted scatterplot smoothing (LOWESS).

## Basic Model Statistics

For a detailed explanation of the terms used below, refer to Helsel and Hirsch (2002).

```
Number of Observations                              11
Standard error (RMSE)                            0.356
Upper Model standard percentage error (MSPE)      4.13
Lower Model standard percentage error (MSPE)      4.13
Coefficient of determination (R²)                0.996
Adjusted Coefficient of Determination (Adj. R²) 0.996
```

## Explanatory Variables

```
            Coefficients Standard Error t value Probability(>|t|)
(Intercept)       -0.302         0.2110   -1.43          1.86e-01
sensor             1.070         0.0217   49.00          3.05e-12
```

## Correlation Matrix

```
          Intercept sensor
Intercept     1.000 -0.861
sensor       -0.861  1.000
```

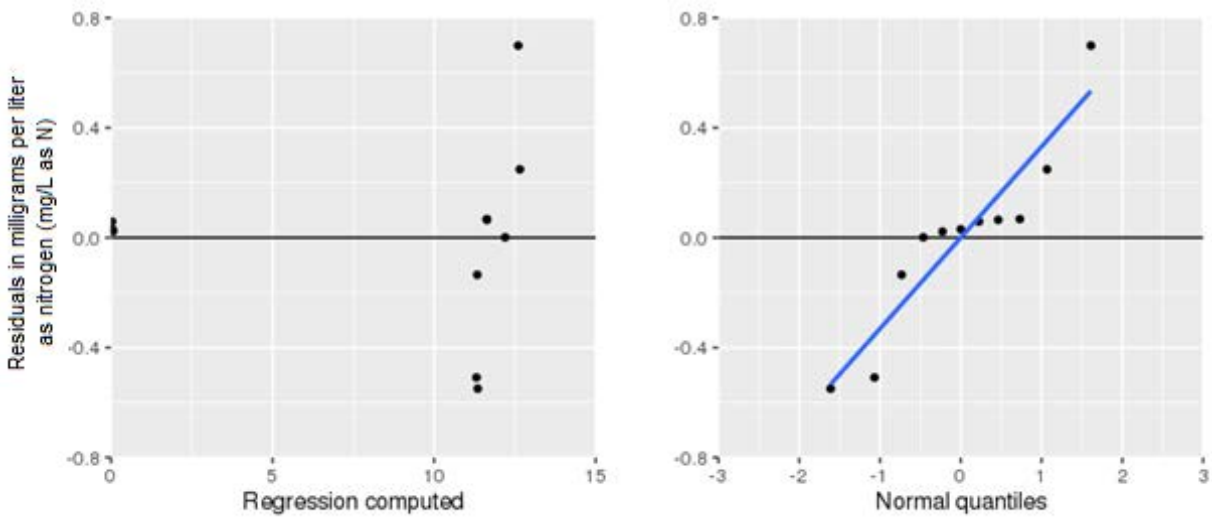
## Outlier Test Criteria

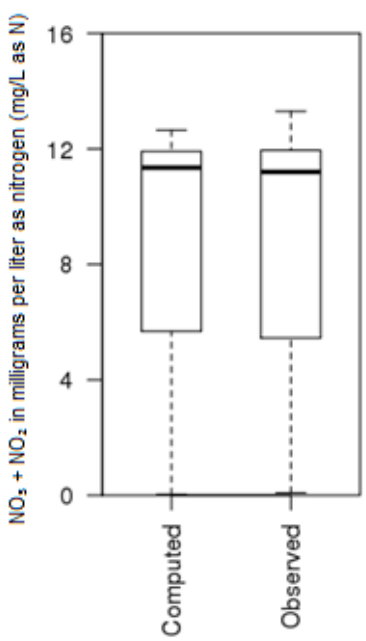
```
Leverage Cook's D   DFFITS
   0.273    0.106    0.603
```

## Flagged Observations

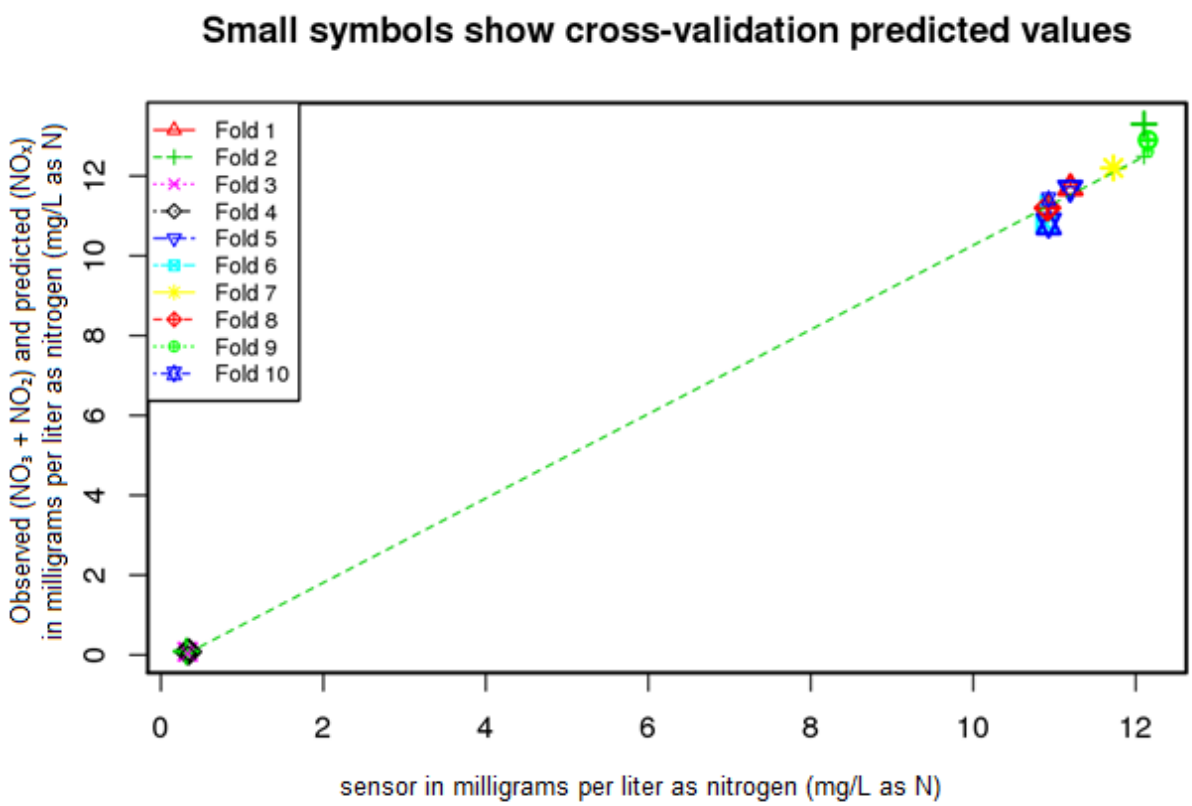
| | $NO_3+NO_2$ | Estimate | Residual | Standard Residual | Studentized Residual | Leverage | Cook's D | DFFITS |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.08 | 0.0494 | 0.0306 | 0.1050 | 0.0989 | 0.331 | 0.00273 | 0.0697 |
| 2 | 0.08 | 0.0578 | 0.0222 | 0.0762 | 0.0719 | 0.331 | 0.00144 | 0.0505 |
| 3 | 0.09 | 0.0314 | 0.0586 | 0.2010 | 0.1900 | 0.332 | 0.01010 | 0.1340 |
| 5 | 10.80 | 11.3000 | -0.5090 | -1.5200 | -1.6600 | 0.115 | 0.14900 | -0.5960 |
| 6 | 10.80 | 11.3000 | -0.5490 | -1.6400 | -1.8400 | 0.115 | 0.17500 | -0.6660 |
| 11 | 13.30 | 12.6000 | 0.6990 | 2.1200 | 2.8200 | 0.143 | 0.37300 | 1.1500 |

## Statistical Plots

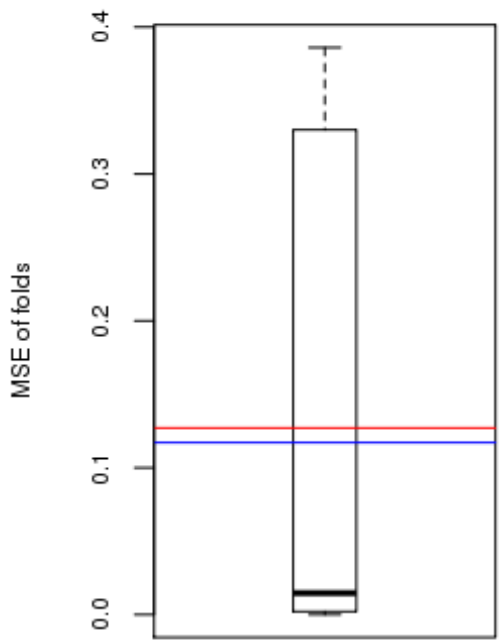


## Cross Validation




```
Minimum model standard error (MSE) of folds:  2.43e-06
                          Mean MSE of folds:  1.17e-01
                        Median MSE of folds:  1.46e-02
                       Maximum MSE of folds:  3.86e-01
(Mean MSE of folds) / (Model MSE):  9.22e-01
```




```
Red line - Model MSE

Blue line - Mean MSE of folds
```

## Model-Calibration Dataset

| | Time | $NO_3+NO_2$ | sensor | Computed NOx | Residual | Normal Quantiles | Censored Values |
|---|---|---|---|---|---|---|---|
| 0 | | | | | | | |
| 1 | 7/23/2014 6:50 | 0.08 | 0.33 | 0.0494 | 0.0306 | 0 | -- |
| 2 | 7/23/2014 6:50 | 0.08 | 0.338 | 0.0578 | 0.0222 | -0.226 | -- |
| 3 | 7/23/2014 7:25 | 0.09 | 0.313 | 0.0314 | 0.0586 | 0.226 | -- |
| 4 | 7/23/2014 7:30 | 11.7 | 11.2 | 11.6 | 0.065 | 0.464 | -- |
| 5 | 7/23/2014 8:15 | 10.8 | 10.9 | 11.3 | -0.509 | -1.07 | -- |
| 6 | 7/23/2014 8:15 | 10.8 | 10.9 | 11.3 | -0.549 | -1.61 | -- |
| 7 | 7/23/2014 8:55 | 11.2 | 10.9 | 11.3 | -0.135 | -0.732 | -- |
| 8 | 7/23/2014 9:37 | 11.7 | 11.2 | 11.6 | 0.0679 | 0.732 | -- |
| 9 | 7/23/2014 10:10 | 12.2 | 11.7 | 12.2 | 0.00135 | -0.464 | -- |
| 10 | 7/23/2014 10:53 | 12.9 | 12.1 | 12.7 | 0.249 | 1.07 | -- |
| 11 | 7/23/2014 11:16 | 13.3 | 12.1 | 12.6 | 0.699 | 1.61 | -- |

Definitions

**Cook's D:** Cook's distance (Helsel and Hirsch, 2002).
**DFFITS:** Difference in fits statistic (Helsel and Hirsch, 2002).
**Leverage:** An outlier's measure in the x direction (Helsel and Hirsch, 2002).
**LOWESS:** Locally weighted scatterplot smoothing (Cleveland, 1979; Helsel and Hirsch, 2002).
**MSE:** Model standard error (Helsel and Hirsch, 2002).
**$NO_3$+$NO_2$:** Nitrate plus nitrite, water, filtered, milligrams per liter as nitrogen (parameter code 00631).
**NOx:** Corrected sensor-measured nitrate plus nitrite, milligrams per liter as nitrogen (mg/L as N).
**MSPE:** Model standard percentage error (Helsel and Hirsch, 2002).
**Probability(>|t|):** The probability that the independent variable has no effect on the dependent variable (Helsel and Hirsch, 2002).
**RMSE:** Root mean square error (Helsel and Hirsch, 2002).
**Sensor:** Nitrate plus nitrite, water, in situ, milligrams per liter as nitrogen (parameter code 99133).
**t value:** Student's t value; the coefficient divided by its associated standard error (Helsel and Hirsch, 2002).

## References Cited


Cleveland, W.S., 1979, Robust Locally Weighted Regression and Smoothing Scatterplots: *J. Am. Stat. Assoc. 74*, 829-836.

Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources—Hydrologic analysis and interpretation: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. A3, 510 p.

King, L.R., Foster, G.M., and Graham, J.L., 2016, Spatial water-quality data for Indian Creek, Johnson County, Kansas, May 23, 2013, July 23, 2014, July 30, 2015 and August 26, 2015, *https://dx.doi.org/10.5066/F77W69CP*.

U.S. Geological Survey, variously dated, National field manual for the collection of water-quality data: U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chaps. A1–A9.

# Appendix 2. Model Archival Summary for Sensor-Measured Nitrate plus Nitrite Concentration Along Indian Creek on July 23, 2014

This model archival summary summarizes the nitrate plus nitrite concentration, milligrams per liter as nitrogen ($NO_3 + NO_2$) model developed to correct 30-second sensor-measured $NO_3 + NO_2$ on July 23, 2014.

## Study Reach and Model Information

Study Reach: Indian Creek near Metcalf Ave, Overland Park, Kansas (station number: 385632094395800; site identifier: 17) to Indian Creek at Sate Line Road, Leawood, Kansas [station number: 06893390; site identifier: State Line (1)].
Location: Lat 38°56'31.64", long 94°39'57.73" referenced to North American Datum of 1983, Hydrologic Unit 10270104.

Equipment: A Hach Nitratax SC sensor was suspended from a kayak along the Indian Creek study reach on July 23, 2014. Readings from the sensor were recorded every 30 seconds.

Date model was created: June 1, 2016

Model calibration data period and application date: July 23, 2014.

## Model-Calibration Dataset

All data were collected using U.S. Geological Survey protocols (U.S. Geological Survey, variously dated; http://water.usgs.gov/owq/FieldManual/) and are stored in the National Water Information System database at *http://dx.doi.org/10.5066/F7P55KJN* and in King and others (2016). Ordinary least squares analysis was used to develop regression models between sensor- and laboratory-measured $NO_3 + NO_2$ concentrations for all sites along the reach using the open-source software package "R." Separate Nitratax sensors were used during the 2014 and 2015 spatial surveys. These sensors varied in the level of over- and/or under-estimation of $NO_3 + NO_2$. Therefore, all models developed to correct 30-second sensor-measured $NO_3 + NO_2$ are instrument specific.

The final selected regression model is based on 18 concurrent measurements of sensor-measured $NO_3 + NO_2$ and laboratory-measured $NO_3 + NO_2$ collected on July 23, 2014. Samples were collected during below-normal flow conditions. No samples were below laboratory detection limits. Summary statistics and the complete model-calibration dataset are provided below. Studentized residuals from the initial model were inspected for values greater than 3 or less than negative 3. Values outside of that range were considered potential outliers and investigated. The laboratory-measured $NO_3 + NO_2$ concentration of 6.72 mg/L (corresponding sensor-measured $NO_3 + NO_2$ concentration of 11.04 mg/L) was deemed an outlier and was therefore removed from the final model. The three laboratory-measured $NO_3 + NO_2$samples immediately before this outlier ranged from 11.4 mg/L to 10.4 mg/L, similar to the corresponding sensor-measured $NO_3 + NO_2$ range of 12.2 mg/L to 11.6 mg/L.

## Nitrate plus Nitrite Sampling Details

All $NO_3 + NO_2$ samples for laboratory analysis were collected either using grab sample or vertically integrated at the centroid of flow every 0.5 km along the study reach. Samples were analyzed for $NO_3 + NO_2$ concentration at the Johnson County, Kansas Water Quality Laboratory.

## Model Development

Ordinary least squares regression analysis was done using R by examining laboratory-measured $NO_3 + NO_2$ concentrations as explanatory variables for correcting sensor-measured $NO_3 + NO_2$ concentrations. The distribution of residuals was examined for normality, and plots of residuals (the difference between the measured and computed values) as compared to computed $NO_3 + NO_2$ were examined for homoscedasticity (meaning that their departures from zero did not change substantially over the range of computed values).

Laboratory-measured $NO_3 + NO_2$ was selected to correct sensor-measured $NO_3 + NO_2$ based on residual plots, relatively high adjusted coefficient of determination (adjusted $R^2$), and relatively low model standard percentage error (*MSPE*). Values for all of the aforementioned statistics and metrics were computed and are included below along with all relevant sample data and more in-depth statistical information.

## Model Summary

Summary of final regression analysis for corrected sensor-measured $NO_3 + NO_2$ concentration Indian Creek near Metcalf Ave, Overland Park, Kansas (station number: 385632094395800; site identifier: 17) to Indian Creek at Sate Line Road, Leawood, Kansas [station number: 06893390; site identifier: State Line (1)].

$NO_3 + NO_2$ correction-based model:

$$NO_x = 0.939 \times sensor - 0.485$$

where

$NO_x$ = corrected sensor-measured nitrate plus nitrite in milligrams per liter (mg/L); and,

*sensor* = uncorrected sensor-measured nitrate plus nitrite in milligrams per liter (mg/L).

R Output for the relation between sensor-measured and laboratory-measured nitrate plus nitrite concentrations at the Indian Creek study sites

## Model Statistics, Data, and Plots

Definitions for terms used in this output can be found at the end of this document.

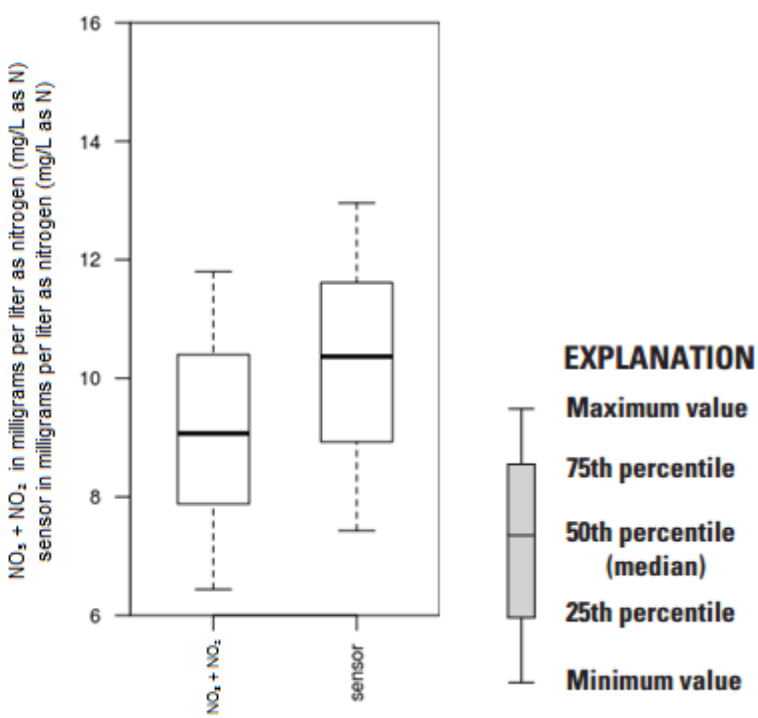
### Model

NOx = + 0.939 * sensor - 0.485

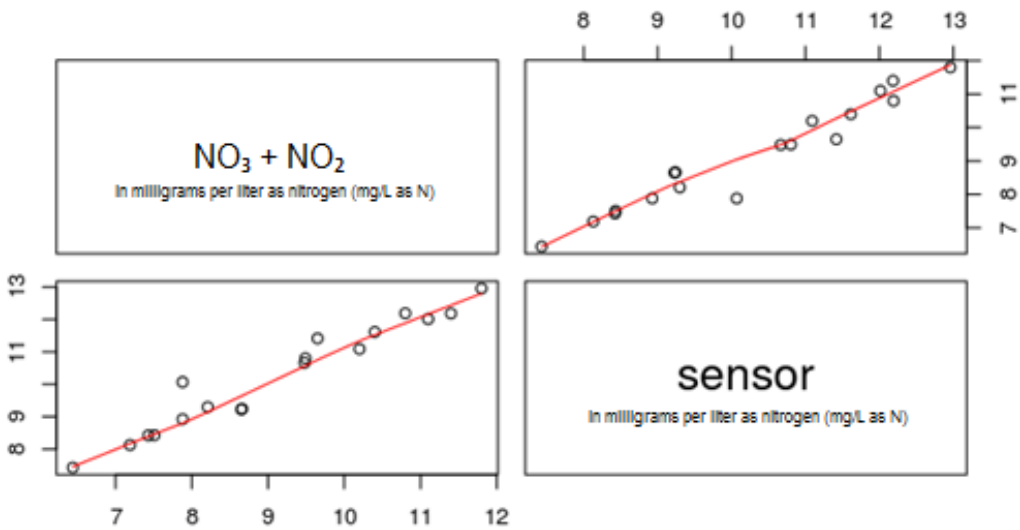
### Variable Summary Statistics

```
                NOx sensor
Minimum        6.44   7.43
1st Quartile   7.88   8.93
Median         9.07  10.40
Mean           9.12  10.20
3rd Quartile  10.40  11.60
Maximum       11.80  13.00
```

### Box Plots



## Exploratory Plots




Red line shows the locally weighted scatterplot smoothing (LOWESS).

## Basic Model Statistics

For a detailed explanation of the terms used below, refer to Helsel and Hirsch (2002).

| | |
|---|---|
| Number of Observations | 18 |
| Standard error (RMSE) | 0.389 |
| Upper Model standard percentage error (MSPE) | 4.26 |
| Lower Model standard percentage error (MSPE) | 4.26 |
| Coefficient of determination ($R^2$) | 0.944 |
| Adjusted Coefficient of Determination (Adj. $R^2$) | 0.94 |

## Explanatory Variables

| | Coefficients | Standard Error | t value | Probability(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | -0.485 | 0.5920 | -0.819 | 4.25e-01 |
| sensor | 0.939 | 0.0572 | 16.400 | 1.95e-11 |

## Correlation Matrix

| | Intercept | sensor |
|---|---|---|
| Intercept | 1.000 | -0.988 |
| sensor | -0.988 | 1.000 |

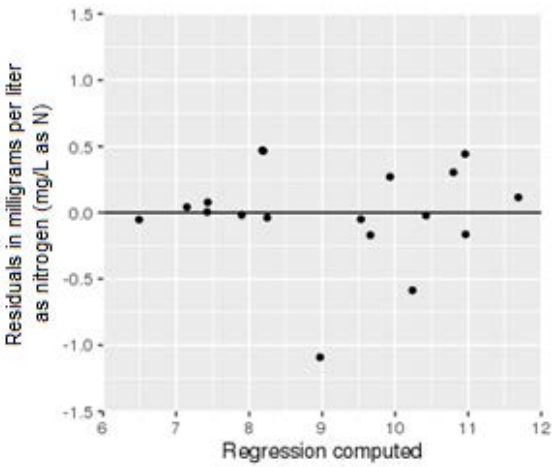
## Outlier Test Criteria

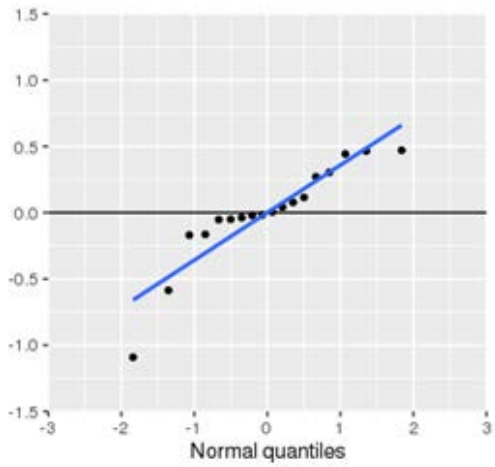
| Leverage | Cook's D | DFFITS |
|---|---|---|
| 0.167 | 0.106 | 0.471 |

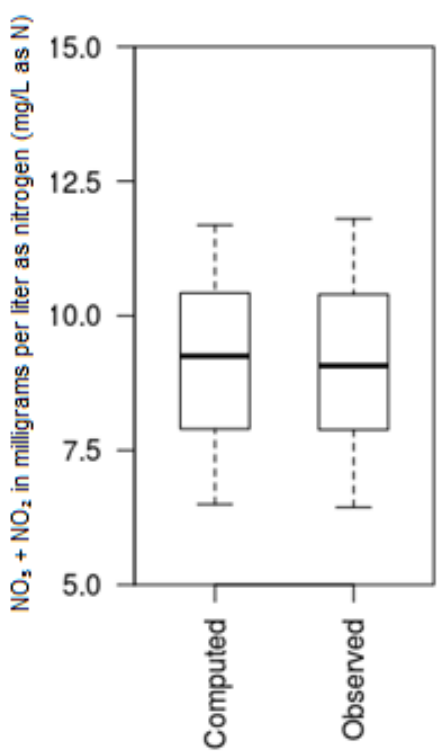
## Flagged Observations

| | $NO_3+NO_2$ | Estimate | Residual | Standard Residual | Studentized Residual | Leverage | Cook's D | DFFITS |
|---|---|---|---|---|---|---|---|---|
| 1 | 11.80 | 11.70 | 0.1150 | 0.334 | 0.324 | 0.2170 | 0.01540 | 0.1710 |
| 3 | 9.65 | 10.20 | -0.5870 | -1.580 | -1.660 | 0.0861 | 0.11700 | -0.5110 |
| 14 | 6.44 | 6.49 | -0.0526 | -0.154 | -0.149 | 0.2250 | 0.00344 | -0.0803 |
| 15 | 7.88 | 8.97 | -1.0900 | -2.890 | -4.050 | 0.0561 | 0.24800 | -0.9870 |
| 16 | 11.40 | 11.00 | 0.4430 | 1.230 | 1.250 | 0.1380 | 0.12100 | 0.5000 |

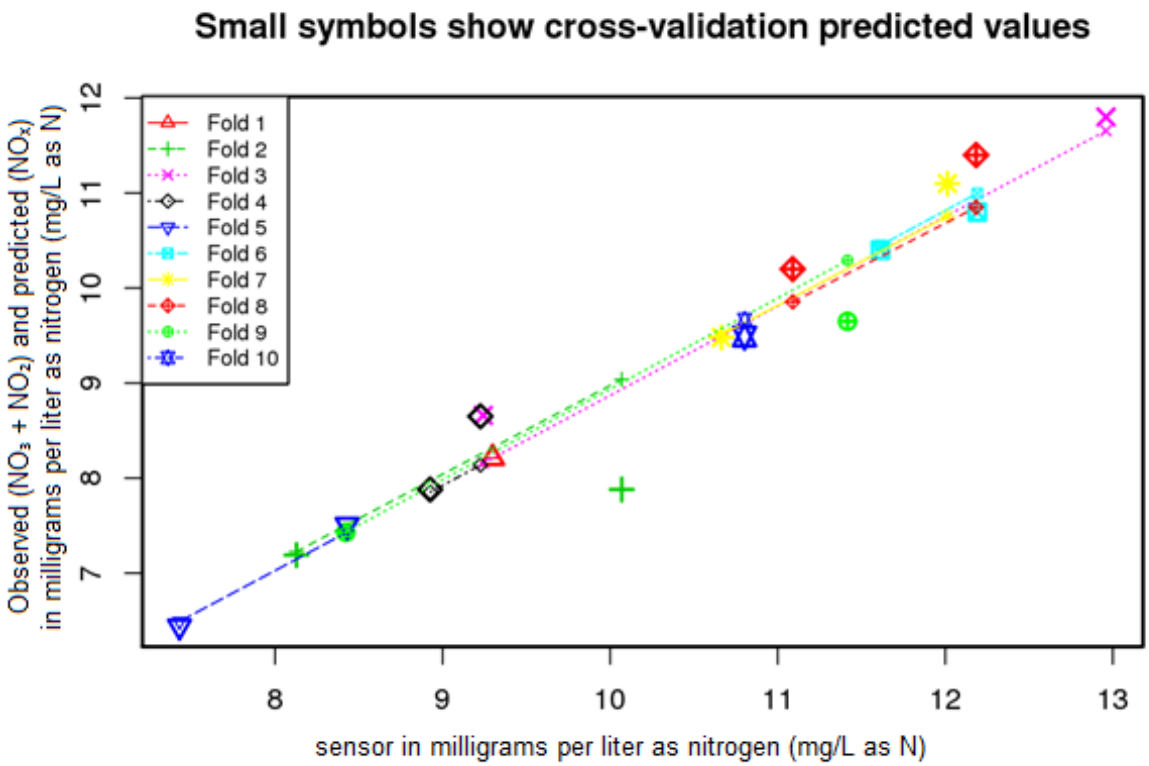
## Statistical Plots



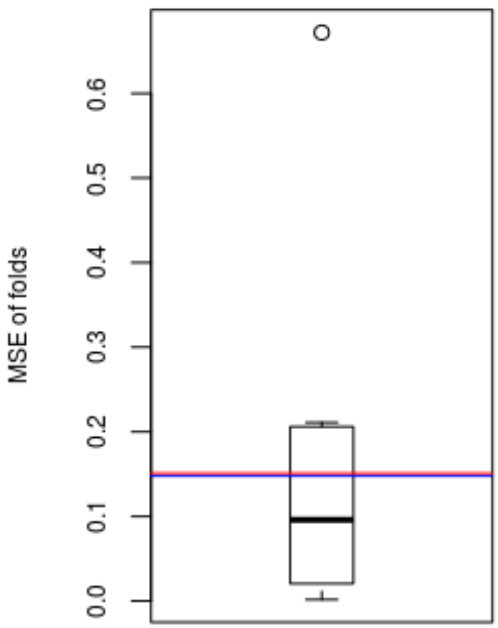



## Cross Validation




```
Minimum model standard error (MSE) of folds:  0.00166
                          Mean MSE of folds:  0.14800
                        Median MSE of folds:  0.09590
                       Maximum MSE of folds:  0.67200
(Mean MSE of folds) / (Model MSE):  0.97800
```




```
Red line - Model MSE

Blue line - Mean MSE of folds
```

## Model-Calibration Dataset

| | Time | $NO_3+NO_2$ sensor | Computed NOx | Residual | Normal Quantiles | Censored Values |
|---|---|---|---|---|---|---|
| 1 | 7/23/2014 15:06 | 11.8 | 13 | 11.7 | 0.115 | 0.502 | -- |
| 2 | 7/23/2014 15:40 | 10.8 | 12.2 | 11 | -0.164 | -0.849 | -- |
| 3 | 7/23/2014 16:10 | 9.65 | 11.4 | 10.2 | -0.587 | -1.35 | -- |
| 4 | 7/23/2014 16:37 | 10.2 | 11.1 | 9.93 | 0.27 | 0.666 | -- |
| 5 | 7/23/2014 17:00 | 9.49 | 10.8 | 9.66 | -0.17 | -1.07 | -- |
| 6 | 7/23/2014 17:30 | 9.48 | 10.7 | 9.53 | -0.0499 | -0.502 | -- |
| 7 | 7/23/2014 18:24 | 8.21 | 9.3 | 8.25 | -0.0377 | -0.351 | -- |
| 8 | 7/23/2014 18:40 | 8.66 | 9.24 | 8.2 | 0.465 | 1.35 | -- |
| 9 | 7/23/2014 18:40 | 8.65 | 9.23 | 8.18 | 0.47 | 1.84 | -- |
| 10 | 7/23/2014 19:02 | 7.88 | 8.93 | 7.9 | -0.0178 | -0.0689 | -- |
| 11 | 7/23/2014 19:23 | 7.43 | 8.42 | 7.42 | 0.0051 | 0.0689 | -- |
| 12 | 7/23/2014 19:23 | 7.51 | 8.43 | 7.43 | 0.0776 | 0.351 | -- |
| 13 | 7/23/2014 19:40 | 7.19 | 8.13 | 7.15 | 0.0417 | 0.208 | -- |
| 14 | 7/23/2014 20:05 | 6.44 | 7.43 | 6.49 | -0.0526 | -0.666 | -- |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 15 | 7/23/2014 20:10 | 7.88 | 10.1 | 8.97 | -1.09 | -1.84 | -- |
| 16 | 7/23/2014 20:25 | 11.4 | 12.2 | 11 | 0.443 | 1.07 | -- |
| 17 | 7/23/2014 20:40 | 11.1 | 12 | 10.8 | 0.304 | 0.849 | -- |
| 18 | 7/23/2014 20:53 | 10.4 | 11.6 | 10.4 | -0.0213 | -0.208 | -- |

Definitions

**Cook's D:** Cook's distance (Helsel and Hirsch, 2002).
**DFFITS:** Difference in fits statistic (Helsel and Hirsch, 2002).
**Leverage:** An outlier's measure in the x direction (Helsel and Hirsch, 2002).
**LOWESS:** Locally weighted scatterplot smoothing (Cleveland, 1979; Helsel and Hirsch, 2002).
**MSE:** Model standard error (Helsel and Hirsch, 2002).
**$NO_3$+$NO_2$:** Nitrate plus nitrite, water, filtered, milligrams per liter as nitrogen (parameter code 00631).
**NOx:** Corrected sensor-measured nitrate plus nitrite, milligrams per liter as nitrogen (mg/L as N).
**MSPE:** Model standard percentage error (Helsel and Hirsch, 2002).
**Probability(>|t|):** The probability that the independent variable has no effect on the dependent variable (Helsel and Hirsch, 2002).
**RMSE:** Root mean square error (Helsel and Hirsch, 2002).
**Sensor:** Nitrate plus nitrite, water, in situ, milligrams per liter as nitrogen (parameter code 99133).
**t value:** Student's t value; the coefficient divided by its associated standard error (Helsel and Hirsch, 2002).

## References Cited


Cleveland, W.S., 1979, Robust Locally Weighted Regression and Smoothing Scatterplots: *J. Am. Stat. Assoc. 74*, 829-836.

Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources—Hydrologic analysis and interpretation: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. A3, 510 p.

King, L.R., Foster, G.M., and Graham, J.L., 2016, Spatial water-quality data for Indian Creek, Johnson County, Kansas, May 23, 2013, July 23, 2014, July 30, 2015 and August 26, 2015, *https://dx.doi.org/10.5066/F77W69CP*.

U.S. Geological Survey, variously dated, National field manual for the collection of water-quality data: U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chaps. A1–A9.

# Appendix 3. Model Archival Summary for Sensor-Measured Nitrate plus Nitrite Concentration for Survey 1 Along Indian Creek on August 26, 2015

This model archival summary summarizes the nitrate plus nitrite concentration, milligrams per liter as nitrogen ($NO_3 + NO_2$) model developed to correct 30-second sensor-measured $NO_3 + NO_2$ for Survey 1 on August 26, 2015.

## Study Reach and Model Information

Study Reach: Indian Creek at Highway 69, Overland Park, Kansas (station number: 06893270; site identifier: Switzer) to Indian Creek at State Line Road, Leawood, Kansas [station number: 06893390; site identifier: State Line (1)]. Location: Lat 38°55'14.07", long 94°42'20.02" to lat 38°56'18.87", long 94°36'29.12" referenced to North American Datum of 1983, Hydrologic Unit 10270104.

Equipment: A Hach Nitratax SC sensor was suspended from a kayak along the Indian Creek study reach on August 26, 2015. Readings from the sensor were recorded every 30 seconds.

Date model was created: June 1, 2016

Model calibration data period and application date: August 26, 2015.

## Model-Calibration Dataset

All data were collected using U.S. Geological Survey protocols (U.S. Geological Survey, variously dated; http://water.usgs.gov/owq/FieldManual/) and are stored in the National Water Information System database at *http://dx.doi.org/10.5066/F7P55KJN* and in King and other (2016). Ordinary least squares analysis was used to develop regression models between sensor- and laboratory-measured $NO_3 + NO_2$ concentrations for all sites along the reach using the open-source software package “R.” Separate Nitratax sensors were used during the 2014 and 2015 spatial surveys. These sensors varied in the level of over- and/or under-estimation of $NO_3 + NO_2$. Therefore, all models developed to correct 30-second sensor-measured $NO_3 + NO_2$ are instrument specific.

The final selected regression model is based on 33 concurrent measurements of sensor-measured $NO_3 + NO_2$ and laboratory-measured $NO_3 + NO_2$ collected on August 26, 2015. Samples were collected during below-normal flow conditions. No samples were below laboratory detection limits. Summary statistics and the complete model-calibration dataset are provided below. Studentized residuals from the initial model were inspected for values greater than 3 or less than negative 3. Values outside of that range were considered potential outliers and investigated. The laboratory-measured $NO_3 + NO_2$ concentration of 17.5 mg/L (corresponding sensor-measured $NO_3 + NO_2$ concentration of 2.31 mg/L) was deemed an outlier and was therefore removed from the final model. This outlier was likely in result of the laboratory-measured $NO_3 + NO_2$ being sampled directly from the Tomahawk wastewater effluent pipe, whereas the sensor-measured $NO_3 + NO_2$ was recorded in the centroid of flow adjacent to, or just downstream of the effluent.

## Nitrate plus Nitrite Sampling Details

All $NO_3 + NO_2$ samples for laboratory analysis were collected using grab sample at the centroid of flow every 0.5 km along the study reach. Samples were analyzed for $NO_3 + NO_2$ concentration at the Johnson County, Kansas Water Quality Laboratory.

## Model Development

Ordinary least squares regression analysis was done using R by examining laboratory-measured $NO_3 + NO_2$ concentrations as explanatory variables for correcting sensor-measured $NO_3 + NO_2$ concentrations. The distribution of residuals was examined for normality, and plots of residuals (the difference between the measured and computed values) as compared to computed $NO_3 + NO_2$

were examined for homoscedasticity (meaning that their departures from zero did not change substantially over the range of computed values).

Laboratory-measured $NO_3 + NO_2$ was selected to correct sensor-measured $NO_3 + NO_2$ based on residual plots, relatively high adjusted coefficient of determination (adjusted $R^2$), and relatively low model standard percentage error (*MSPE*). Values for all of the aforementioned statistics and metrics were computed and are included below along with all relevant sample data and more in-depth statistical information.

## Model Summary

Summary of final regression analysis for corrected sensor-measured $NO_3 + NO_2$ concentration from Indian Creek at Highway 69, Overland Park, Kansas (station number: 06893270; site identifier: Switzer) to Indian Creek at State Line Road, Leawood, Kansas [station number: 06893390; site identifier: (1)].

$NO_3 + NO_2$ correction-based model:

$$NO_x = 0.908 \times sensor + 0.444$$

where

$NO_x$ = corrected sensor-measured nitrate plus nitrite in milligrams per liter (mg/L); and,

*sensor* = uncorrected sensor-measured nitrate plus nitrite in milligrams per liter (mg/L).

R Output for the relation between sensor-measured and laboratory-measured nitrate plus nitrite concentrations at the Indian Creek study sites

## Model Statistics, Data, and Plots

```
Definitions for terms used in this output can be found at the end of this
document.
```

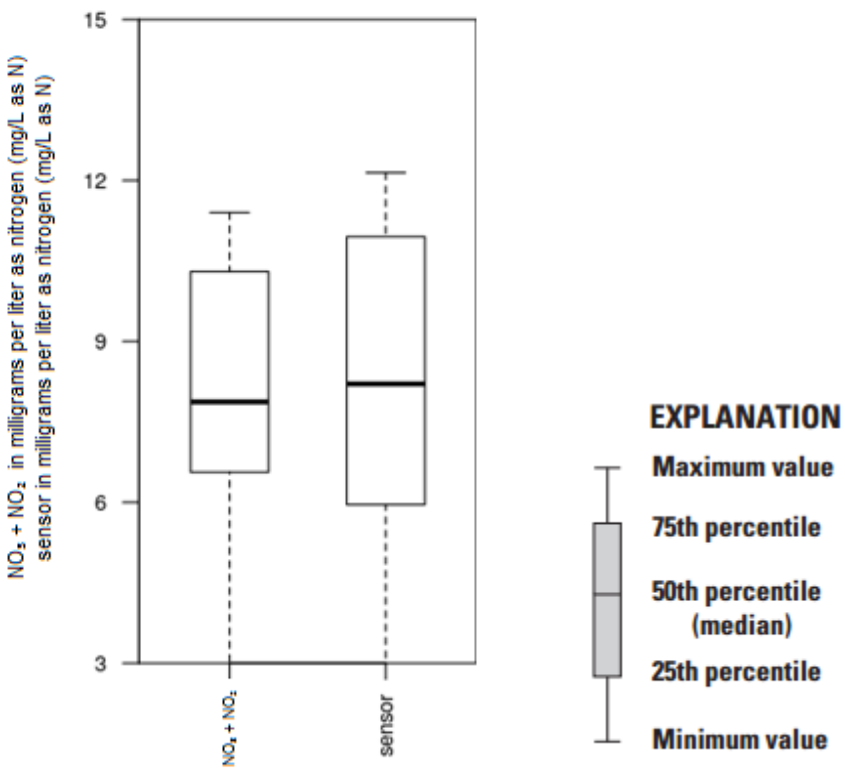
### Model

NOx = + 0.908 * sensor + 0.444

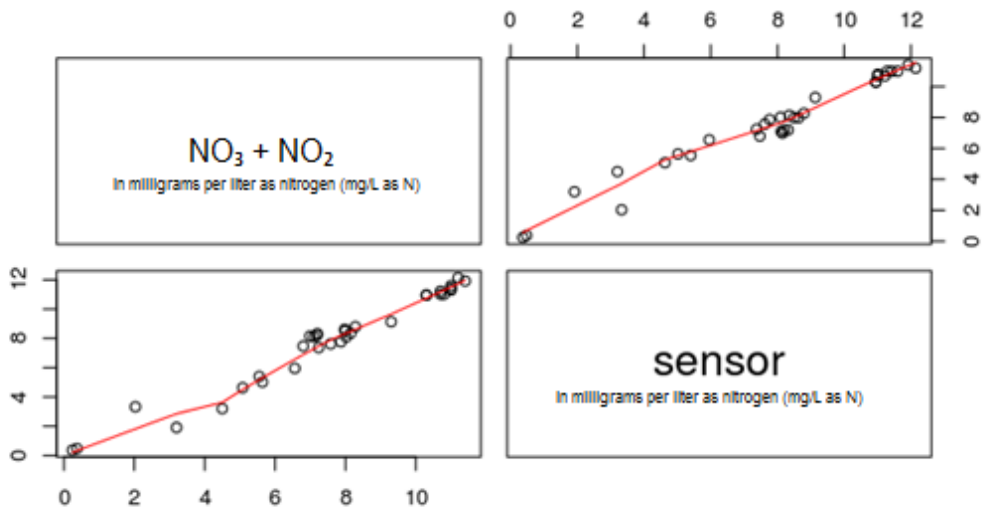
### Variable Summary Statistics

```
                NOx sensor
Minimum        0.24  0.371
1st Quartile   6.56  5.950
Median         7.87  8.200
Mean           7.55  7.830
3rd Quartile  10.30 10.900
Maximum       11.40 12.100
```

### Box Plots



## Exploratory Plots




Red line shows the locally weighted scatterplot smoothing (LOWESS).

## Basic Model Statistics

For a detailed explanation of the terms used below, refer to Helsel and Hirsch (2002).

```
Number of Observations                              33
Standard error (RMSE)                            0.563
Upper Model standard percentage error (MSPE)      7.45
Lower Model standard percentage error (MSPE)      7.45
Coefficient of determination (R²)                0.967
Adjusted Coefficient of Determination (Adj. R²) 0.966
```

## Explanatory Variables

| | Coefficients | Standard Error | t value | Probability(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.444 | 0.2570 | 1.73 | 9.40e-02 |
| sensor | 0.908 | 0.0303 | 29.90 | 1.91e-24 |

## Correlation Matrix

| | Intercept | sensor |
|---|---|---|
| Intercept | 1.000 | -0.924 |
| sensor | -0.924 | 1.000 |

## Outlier Test Criteria

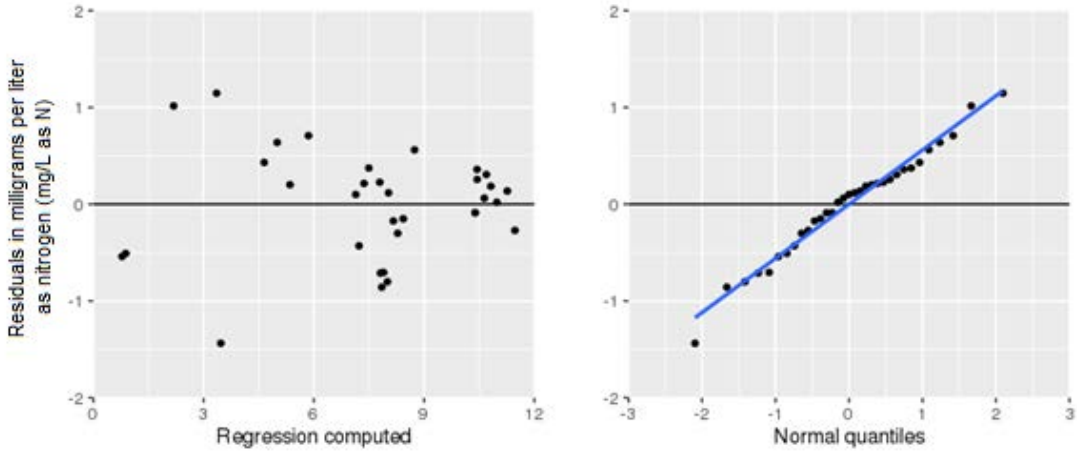
| Leverage | Cook's D | DFFITS |
|---|---|---|
| 0.0909 | 0.1057 | 0.3482 |

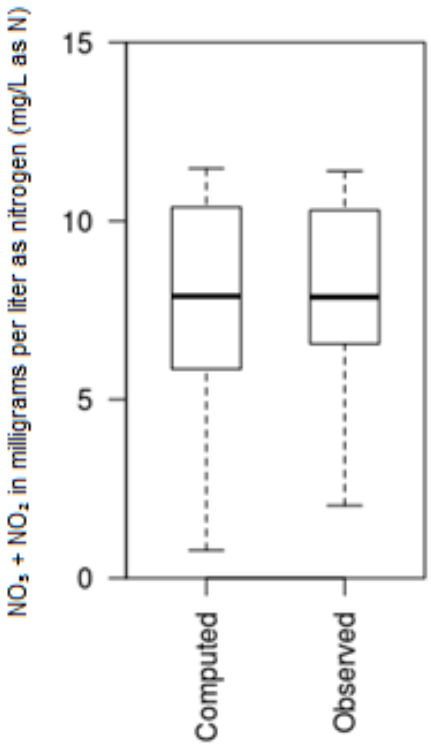
## Flagged Observations

| | $NO_3+NO_2$ | Estimate | Residual | Standard Residual | Studentized Residual | Leverage | Cook's D | DFFITS |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.37 | 0.879 | -0.509 | -1.00 | -1.00 | 0.1870 | 0.116 | -0.481 |

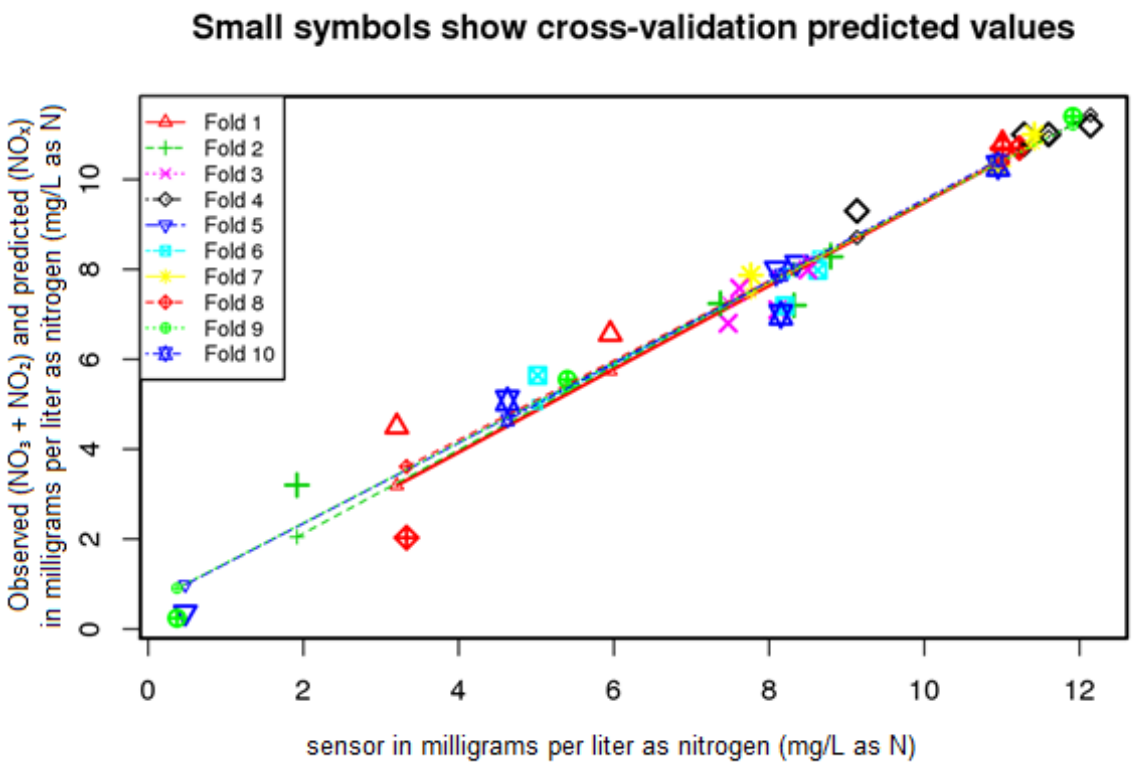
| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 2 | 0.24 | 0.781 | -0.541 | -1.07 | -1.07 | 0.1920 | 0.136 | -0.522 |
| 13 | 2.03 | 3.470 | -1.440 | -2.68 | -3.00 | 0.0891 | 0.350 | -0.939 |
| 28 | 4.50 | 3.350 | 1.150 | 2.14 | 2.28 | 0.0924 | 0.233 | 0.727 |
| 29 | 3.20 | 2.180 | 1.020 | 1.94 | 2.03 | 0.1320 | 0.285 | 0.791 |

## Statistical Plots

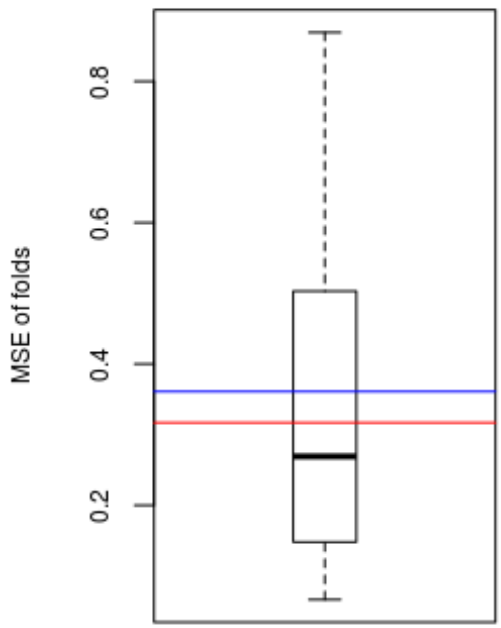


## Cross Validation




```
Minimum model standard error (MSE) of folds:  0.0665
                          Mean MSE of folds:  0.3610
                        Median MSE of folds:  0.2690
                       Maximum MSE of folds:  0.8690
(Mean MSE of folds) / (Model MSE):  1.1400
```




```
Red line - Model MSE

Blue line - Mean MSE of folds
```

## Model-Calibration Dataset

| | Time | $NO_3+NO_2$ sensor | Computed NOx | Residual | Normal Quantiles | Censored Values |
|---|---|---|---|---|---|---|
| 1 | 8/26/2015 6:25 | 0.37 | 0.479 | 0.879 | -0.509 | -0.846 | -- |
| 2 | 8/26/2015 6:55 | 0.24 | 0.371 | 0.781 | -0.541 | -0.959 | -- |
| 3 | 8/26/2015 7:00 | 11 | 11.3 | 10.7 | 0.307 | 0.646 | -- |
| 4 | 8/26/2015 7:40 | 10.8 | 11 | 10.4 | 0.36 | 0.742 | -- |
| 5 | 8/26/2015 7:40 | 10.7 | 11 | 10.4 | 0.256 | 0.556 | -- |
| 6 | 8/26/2015 8:00 | 11 | 11.6 | 11 | 0.02 | -0.152 | -- |
| 7 | 8/26/2015 8:20 | 11.2 | 12.1 | 11.5 | -0.271 | -0.556 | -- |
| 8 | 8/26/2015 8:40 | 10.7 | 11.2 | 10.6 | 0.0621 | -0.0756 | -- |
| 9 | 8/26/2015 9:20 | 11.4 | 11.9 | 11.3 | 0.136 | 0.152 | -- |
| 10 | 8/26/2015 9:29 | 11 | 11.4 | 10.8 | 0.186 | 0.228 | -- |
| 11 | 8/26/2015 9:48 | 10.3 | 10.9 | 10.4 | -0.0873 | -0.228 | -- |
| 12 | 8/26/2015 9:49 | 10.3 | 10.9 | 10.4 | -0.089 | -0.307 | -- |
| 13 | 8/26/2015 10:10 | 2.03 | 3.33 | 3.47 | -1.44 | -2.1 | -- |
| 14 | 8/26/2015 10:15 | 9.3 | 9.13 | 8.74 | 0.562 | 1.09 | -- |
| 15 | 8/26/2015 10:30 | 8.28 | 8.79 | 8.43 | -0.151 | -0.387 | -- |
| 16 | 8/26/2015 11:04 | 7.98 | 8.63 | 8.28 | -0.301 | -0.646 | -- |
| 17 | 8/26/2015 11:20 | 7.99 | 8.5 | 8.16 | -0.173 | -0.47 | -- |
| 18 | 8/26/2015 11:50 | 8.15 | 8.36 | 8.03 | 0.118 | 0.0756 | -- |
| 19 | 8/26/2015 12:16 | 8.02 | 8.09 | 7.79 | 0.227 | 0.47 | -- |
| 20 | 8/26/2015 12:40 | 7.87 | 7.76 | 7.5 | 0.374 | 0.846 | -- |
| 21 | 8/26/2015 13:00 | 7.58 | 7.62 | 7.37 | 0.215 | 0.387 | -- |
| 22 | 8/26/2015 13:20 | 6.8 | 7.47 | 7.23 | -0.43 | -0.742 | -- |
| 23 | 8/26/2015 13:40 | 7.24 | 7.37 | 7.14 | 0.101 | 0 | -- |
| 24 | 8/26/2015 13:50 | 6.56 | 5.95 | 5.85 | 0.708 | 1.42 | -- |
| 25 | 8/26/2015 14:11 | 5.55 | 5.4 | 5.35 | 0.202 | 0.307 | -- |
| 26 | 8/26/2015 14:25 | 5.64 | 5.02 | 5 | 0.638 | 1.23 | -- |
| 27 | 8/26/2015 14:30 | 5.08 | 4.63 | 4.65 | 0.432 | 0.959 | -- |
| 28 | 8/26/2015 15:02 | 4.5 | 3.2 | 3.35 | 1.15 | 2.1 | -- |
| 29 | 8/26/2015 15:30 | 3.2 | 1.92 | 2.18 | 1.02 | 1.66 | -- |
| 30 | 8/26/2015 15:55 | 7.2 | 8.32 | 8 | -0.801 | -1.42 | -- |
| 31 | 8/26/2015 16:10 | 7.1 | 8.11 | 7.81 | -0.712 | -1.23 | -- |
| 32 | 8/26/2015 16:40 | 7.19 | 8.2 | 7.89 | -0.705 | -1.09 | -- |
| 33 | 8/26/2015 16:50 | 6.99 | 8.15 | 7.85 | -0.858 | -1.66 | -- |

Definitions

**Cook's D:** Cook's distance (Helsel and Hirsch, 2002).
**DFFITS:** Difference in fits statistic (Helsel and Hirsch, 2002).
**Leverage:** An outlier's measure in the x direction (Helsel and Hirsch, 2002).
**LOWESS:** Locally weighted scatterplot smoothing (Cleveland, 1979; Helsel and Hirsch, 2002).
**MSE:** Model standard error (Helsel and Hirsch, 2002).
**$NO_3$+$NO_2$:** Nitrate plus nitrite, water, filtered, milligrams per liter as nitrogen (parameter code 00631).
**NOx:** Corrected sensor-measured nitrate plus nitrite, milligrams per liter as nitrogen (mg/L as N).
**MSPE:** Model standard percentage error (Helsel and Hirsch, 2002).
**Probability(>|t|):** The probability that the independent variable has no effect on the dependent variable (Helsel and Hirsch, 2002).
**RMSE:** Root mean square error (Helsel and Hirsch, 2002).
**Sensor:** Nitrate plus nitrite, water, in situ, milligrams per liter as nitrogen (parameter code 99133).
**t value:** Student's t value; the coefficient divided by its associated standard error (Helsel and Hirsch, 2002).

## References Cited


Cleveland, W.S., 1979, Robust Locally Weighted Regression and Smoothing Scatterplots: *J. Am. Stat. Assoc. 74*, 829-836.

Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources—Hydrologic analysis and interpretation: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. A3, 510 p.

King, L.R., Foster, G.M., and Graham, J.L., 2016, Spatial water-quality data for Indian Creek, Johnson County, Kansas, May 23, 2013, July 23, 2014, July 30, 2015 and August 26, 2015, *https://dx.doi.org/10.5066/F77W69CP*.

U.S. Geological Survey, variously dated, National field manual for the collection of water-quality data: U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chaps. A1–A9.

# Appendix 4. Model Archival Summary for Sensor-Measured Nitrate plus Nitrite Concentration for Survey 2 Along Indian Creek on August 26, 2015

This model archival summary summarizes the nitrate plus nitrite concentration, milligrams per liter as nitrogen ($NO_3 + NO_2$) model developed to correct 30-second sensor-measured $NO_3 + NO_2$ for Survey 2 on August 26, 2015.

## Study Reach and Model Information

Study Reach: Indian Creek at Highway 69, Overland Park, Kansas (station number: 06893270; site identifier: Switzer) to Indian Creek at State Line Road, Leawood, Kansas [station number: 06893390; site identifier: State Line (1)]. Location: Lat 38°55'14.07", long 94°42'20.02" to lat 38°56'18.87", long 94°36'29.12" referenced to North American Datum of 1983, Hydrologic Unit 10270104.

Equipment: A Hach Nitratax SC sensor was suspended from a kayak along the Indian Creek study reach on August 26, 2015. Readings from the sensor are recorded every 30 seconds.

Date model was created: June 1, 2016

Model calibration data period and application date: August 26, 2015.

## Model-Calibration Dataset

All data were collected using U.S. Geological Survey (U.S. Geological Survey, variously dated; http://water.usgs.gov/owq/FieldManual/) protocols and are stored in the National Water Information System database at *http://dx.doi.org/10.5066/F7P55KJN* and in King and others (2016). Ordinary least squares analysis was used to develop regression models between sensor- and laboratory-measured $NO_3 + NO_2$ concentrations for all sites along the reach using the open-source software package “R.” Separate Nitratax sensors were used during the 2014 and 2015 spatial surveys. These sensors varied in the level of over- and/or under-estimation of $NO_3 + NO_2$. Therefore, all models developed to correct 30-second sensor-measured $NO_3 + NO_2$ are instrument specific.

The final selected regression model is based on 18 concurrent measurements of sensor-measured $NO_3 + NO_2$ and laboratory-measured $NO_3 + NO_2$ collected on August 26, 2015. Samples were collected during below-normal flow conditions. No samples were below laboratory detection limits. Summary statistics and the complete model-calibration dataset are provided below. Studentized residuals from the initial model were inspected for values greater than 3 or less than negative 3. Values outside of that range were considered potential outliers and investigated. None of the $NO_3 + NO_2$ samples were deemed outliers.

## Nitrate plus Nitrite Sampling Details

All $NO_3 + NO_2$ samples for laboratory analysis were collected via grab sample at the centroid of flow every 0.5 km along the study reach. Samples were analyzed for $NO_3 + NO_2$ concentration at the Johnson County, Kansas Water Quality Laboratory.

## Model Development

Ordinary least squares regression analysis was done using R by examining laboratory-measured $NO_3 + NO_2$ concentrations as explanatory variables for correcting sensor-measured $NO_3 + NO_2$ concentrations. The distribution of residuals was examined for normality, and plots of residuals (the difference between the measured and computed values) as compared to computed $NO_3 + NO_2$ were examined for homoscedasticity (meaning that their departures from zero did not change substantially over the range of computed values).

Laboratory-measured $NO_3 + NO_2$ was selected to correct sensor-measured $NO_3 + NO_2$ based on residual plots, relatively high adjusted coefficient of determination (adjusted $R^2$), and relatively low model standard percentage error (*MSPE*). Values for all of the

aforementioned statistics and metrics were computed and are included below along with all relevant sample data and more in-depth statistical information.

## Model Summary

Summary of final regression analysis for corrected sensor-measured $NO_3 + NO_2$ concentration from Indian Creek at Highway 69, Overland Park, Kansas (station number: 06893270; site identifier: Switzer) to Indian Creek at State Line Road, Leawood, Kansas [station number: 06893390; site identifier: (1)].

$NO_3 + NO_2$ correction-based model:

$$NO_x = 0.968 \times sensor - 0.493$$

where

$NO_x$ = corrected sensor-measured nitrate plus nitrite in milligrams per liter (mg/L); and,

*sensor* = uncorrected sensor-measured nitrate plus nitrite in milligrams per liter (mg/L).

R Output for the relation between sensor-measured and laboratory-measured nitrate plus nitrite concentrations at the Indian Creek study sites

## Model Statistics, Data, and Plots

Definitions for terms used in this output can be found at the end of this document.

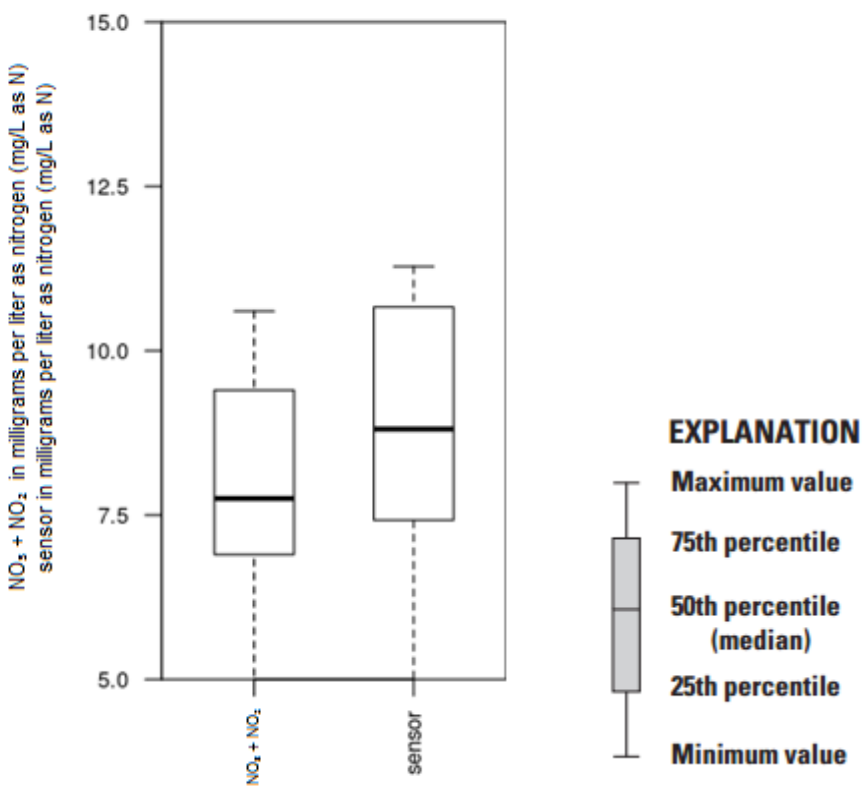
### Model

NOx = + 0.968 * sensor - 0.493

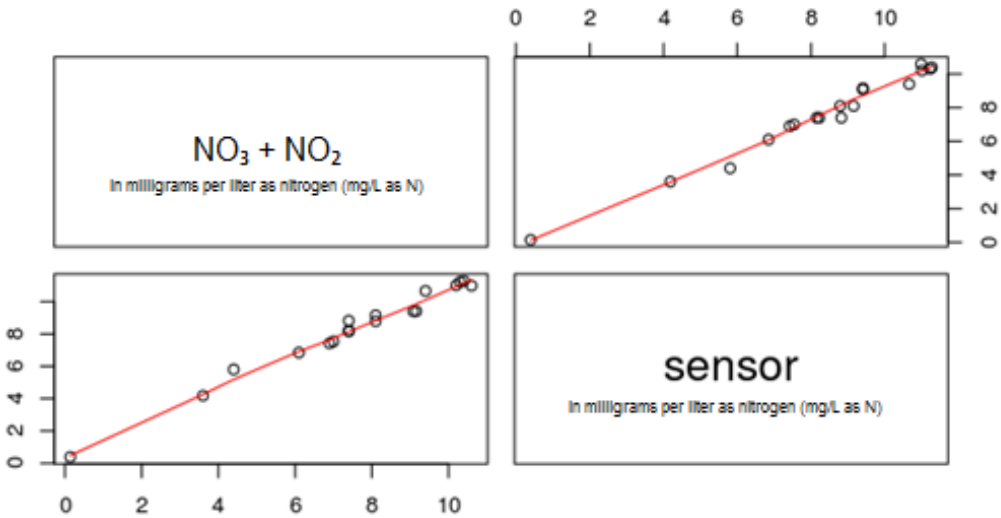
### Variable Summary Statistics

```
                NOx sensor
Minimum        0.14  0.395
1st Quartile   6.90  7.420
Median         7.75  8.810
Mean           7.54  8.300
3rd Quartile   9.40 10.700
Maximum       10.60 11.300
```

### Box Plots



## Exploratory Plots




Red line shows the locally weighted scatterplot smoothing (LOWESS).

## Basic Model Statistics

For a detailed explanation of the terms used below, refer to Helsel and Hirsch (2002).

| | |
|---|---|
| Number of Observations | 18 |
| Standard error (RMSE) | 0.362 |
| Upper Model standard percentage error (MSPE) | 4.81 |
| Lower Model standard percentage error (MSPE) | 4.81 |
| Coefficient of determination ($R^2$) | 0.983 |
| Adjusted Coefficient of Determination (Adj. $R^2$) | 0.982 |

## Explanatory Variables

| | Coefficients | Standard Error | t value | Probability(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | -0.493 | 0.2770 | -1.78 | 9.41e-02 |
| sensor | 0.968 | 0.0317 | 30.50 | 1.32e-15 |

## Correlation Matrix

| | Intercept | sensor |
|---|---|---|
| Intercept | 1.000 | -0.951 |
| sensor | -0.951 | 1.000 |

## Outlier Test Criteria

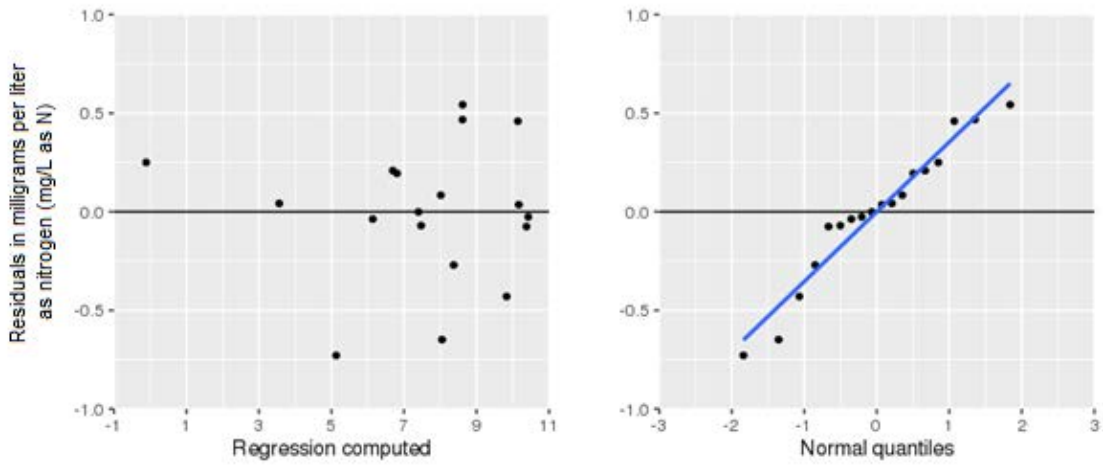
| Leverage | Cook's D | DFFITS |
|---|---|---|
| 0.167 | 0.106 | 0.471 |

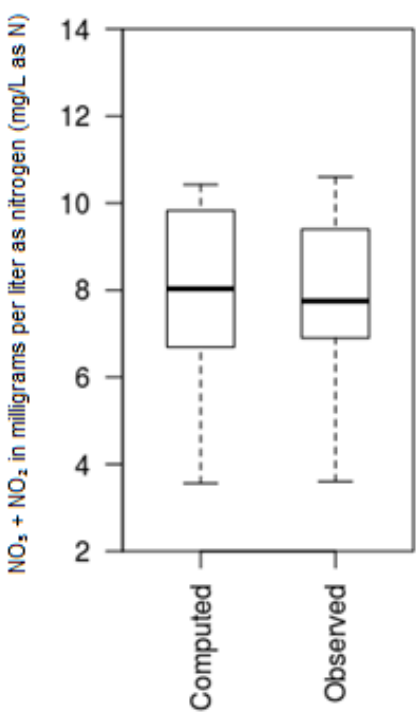
## Flagged Observations

| | $NO_3+NO_2$ | Estimate | Residual | Standard Residual | Studentized Residual | Leverage | Cook's D | DFFITS |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.14 | -0.11 | 0.2500 | 1.010 | 1.010 | 0.5340 | 0.5880 | 1.0800 |

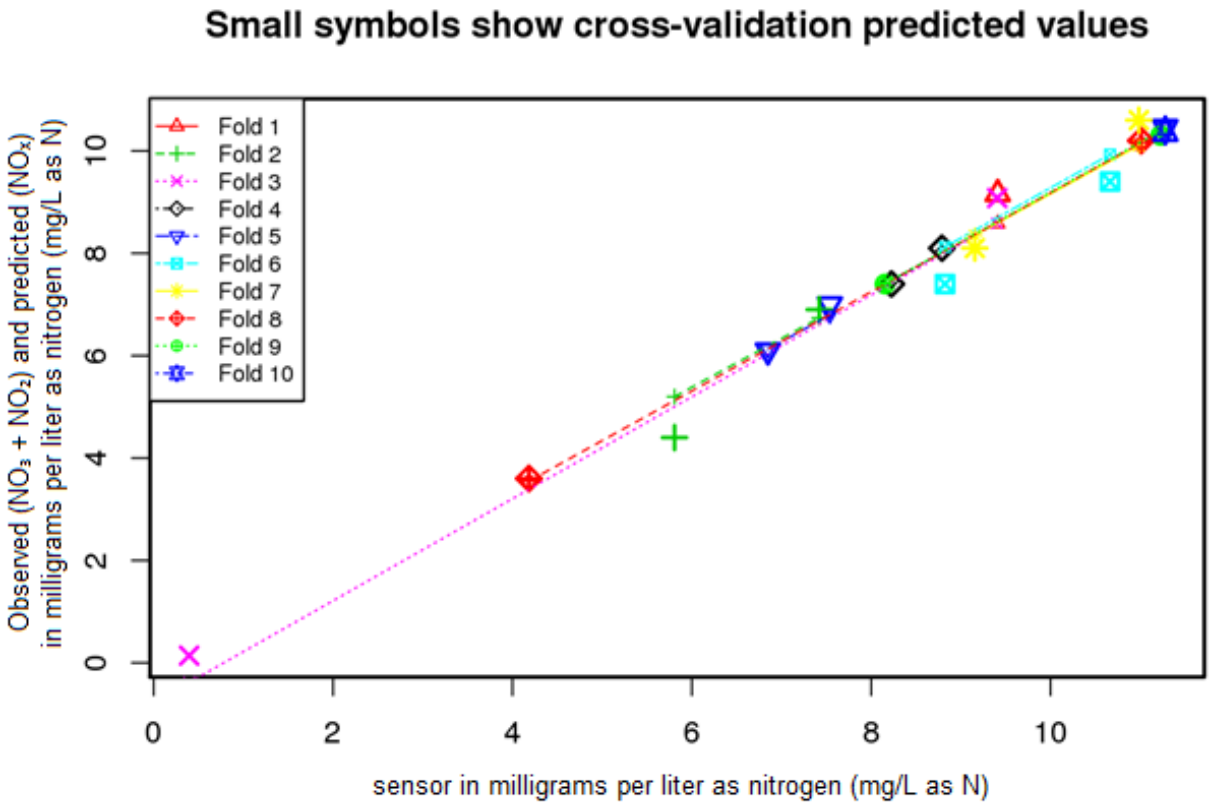
| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 6 | 10.60 | 10.10 | 0.4600 | 1.350 | 1.380 | 0.1110 | 0.1130 | 0.4890 |
| 15 | 4.40 | 5.13 | -0.7290 | -2.120 | -2.430 | 0.1030 | 0.2590 | -0.8220 |
| 16 | 3.60 | 3.56 | 0.0423 | 0.129 | 0.125 | 0.1850 | 0.0019 | 0.0597 |
| 18 | 7.40 | 8.05 | -0.6490 | -1.840 | -2.010 | 0.0577 | 0.1040 | -0.4980 |

## Statistical Plots

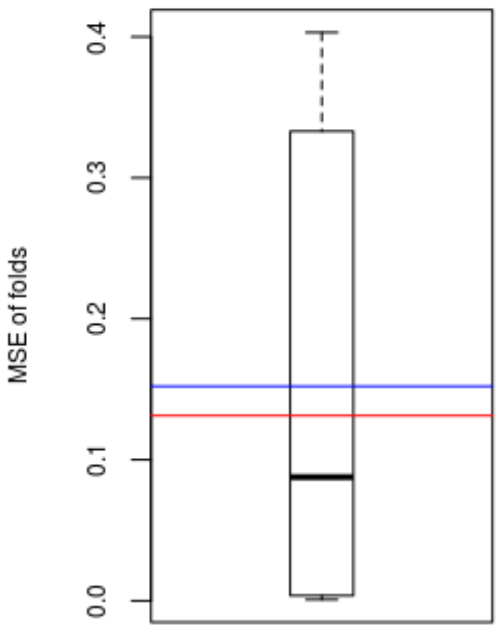


## Cross Validation




```
Minimum model standard error (MSE) of folds:  0.000888
                          Mean MSE of folds:  0.152000
                        Median MSE of folds:  0.087700
                       Maximum MSE of folds:  0.403000
(Mean MSE of folds) / (Model MSE):  1.160000
```




```
Red line - Model MSE

Blue line - Mean MSE of folds
```

## Model-Calibration Dataset

| | Time | $NO_3+NO_2$ | sensor | Computed NOx | Residual | Normal Quantiles | Censored Values |
|---|---|---|---|---|---|---|---|
| 1 | 8/26/2016 12:05 | 0.14 | 0.395 | -0.11 | 0.25 | 0.849 | -- |
| 2 | 8/26/2016 12:25 | 9.4 | 10.7 | 9.83 | -0.43 | -1.07 | -- |
| 3 | 8/26/2016 12:55 | 10.3 | 11.2 | 10.4 | -0.0751 | -0.666 | -- |
| 4 | 8/26/2016 13:25 | 10.2 | 11 | 10.2 | 0.0358 | 0.0689 | -- |
| 5 | 8/26/2016 13:49 | 10.4 | 11.3 | 10.4 | -0.0261 | -0.208 | -- |
| 6 | 8/26/2016 14:12 | 10.6 | 11 | 10.1 | 0.46 | 1.07 | -- |
| 7 | 8/26/2016 14:46 | 9.16 | 9.41 | 8.62 | 0.543 | 1.84 | -- |
| 8 | 8/26/2016 14:46 | 9.08 | 9.41 | 8.61 | 0.468 | 1.35 | -- |
| 9 | 8/26/2016 15:26 | 8.1 | 8.79 | 8.02 | 0.0842 | 0.351 | -- |
| 10 | 8/26/2016 15:54 | 7.4 | 8.23 | 7.47 | -0.0702 | -0.502 | -- |
| 11 | 8/26/2016 16:21 | 7.4 | 8.15 | 7.4 | -0.000733 | -0.0689 | -- |
| 12 | 8/26/2016 16:48 | 7 | 7.54 | 6.81 | 0.195 | 0.502 | -- |
| 13 | 8/26/2016 16:49 | 6.9 | 7.42 | 6.69 | 0.209 | 0.666 | -- |
| 14 | 8/26/2016 17:20 | 6.1 | 6.85 | 6.14 | -0.0369 | -0.351 | -- |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 15 | 8/26/2016 17:37 | 4.4 | 5.81 | 5.13 | -0.729 | -1.84 | -- |
| 16 | 8/26/2016 18:13 | 3.6 | 4.18 | 3.56 | 0.0423 | 0.208 | -- |
| 17 | 8/26/2016 18:37 | 8.1 | 9.16 | 8.37 | -0.27 | -0.849 | -- |
| 18 | 8/26/2016 18:57 | 7.4 | 8.82 | 8.05 | -0.649 | -1.35 | -- |

Definitions

**Cook's D:** Cook's distance (Helsel and Hirsch, 2002).
**DFFITS:** Difference in fits statistic (Helsel and Hirsch, 2002).
**Leverage:** An outlier's measure in the x direction (Helsel and Hirsch, 2002).
**LOWESS:** Locally weighted scatterplot smoothing (Cleveland, 1979; Helsel and Hirsch, 2002).
**MSE:** Model standard error (Helsel and Hirsch, 2002).
**$NO_3$+$NO_2$:** Nitrate plus nitrite, water, filtered, milligrams per liter as nitrogen (parameter code 00631).
**NOx:** Corrected sensor-measured nitrate plus nitrite, milligrams per liter as nitrogen (mg/L as N).
**MSPE:** Model standard percentage error (Helsel and Hirsch, 2002).
**Probability(>|t|):** The probability that the independent variable has no effect on the dependent variable (Helsel and Hirsch, 2002).
**RMSE:** Root mean square error (Helsel and Hirsch, 2002).
**Sensor:** Nitrate plus nitrite, water, in situ, milligrams per liter as nitrogen (parameter code 99133).
**t value:** Student's t value; the coefficient divided by its associated standard error (Helsel and Hirsch, 2002).

## References Cited


Cleveland, W.S., 1979, Robust Locally Weighted Regression and Smoothing Scatterplots: *J. Am. Stat. Assoc. 74*, 829-836.

Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources—Hydrologic analysis and interpretation: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. A3, 510 p.

King, L.R., Foster, G.M., and Graham, J.L., 2016a, Spatial water-quality data for Indian Creek, Johnson County, Kansas, May 23, 2013, July 23, 2014, July 30, 2015 and August 26, 2015, *https://dx.doi.org/10.5066/F77W69CP*.

U.S. Geological Survey, variously dated, National field manual for the collection of water-quality data: U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chaps. A1–A9.