

A feature-based procedure for detecting technical outliers in water-quality data from in situ sensors

Priyanga Dilini Talagala^{1,2}, Rob J. Hyndman^{1,2}, Catherine Leigh^{1,3,4}, Kerrie Mengersen^{1,4}, Kate Smith-Miles^{1,5}

¹ARC Centre of Excellence for Mathematics and Statistical Frontiers (ACEMS), Australia

²Department of Econometrics and Business Statistics, Monash University, Australia

³Institute for Future Environments, Science and Engineering Faculty, Queensland University of Technology, Australia

⁴School of Mathematical Sciences, Science and Engineering Faculty, Queensland University of Technology, Australia

⁵School of Mathematics and Statistics, University of Melbourne, Australia

Key Points:

- Our feature-based procedure starts by applying different statistical transformations to water-quality data to highlight outliers in high dimensional space
- Density and distance-based unsupervised outlier scoring techniques were applied to detect outliers due to technical issues with the sensors
- An approach based on extreme value theory was then used to calculate outlier thresholds

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as doi: [10.1029/2019WR024906](https://doi.org/10.1029/2019WR024906)

Corresponding author: Priyanga Dilini Talagala, dilini.talagala@monash.edu

Abstract

Outliers due to technical errors in water quality data from *in situ* sensors can reduce data quality and have a direct impact on inference drawn from subsequent data analysis. However, outlier detection through manual monitoring is infeasible given the volume and velocity of data the sensors produce. Here we introduce an automated procedure, named oddwater, that provides early detection of outliers in water-quality data from *in situ* sensors caused by technical issues. Our oddwater procedure is used to first identify the data features that differentiate outlying instances from typical behaviours. Then, statistical transformations are applied to make the outlying instances stand out in a transformed data space. Unsupervised outlier scoring techniques are applied to the transformed data space and an approach based on extreme value theory is used to calculate a threshold for each potential outlier. Using two datasets obtained from *in situ* sensors in rivers flowing into the Great Barrier Reef lagoon, Australia, we show that oddwater successfully identifies outliers involving abrupt changes in turbidity, conductivity and river level, including sudden spikes, sudden isolated drops and level shifts, while maintaining very low false detection rates. We have implemented this oddwater procedure in the open source R package `oddwater`.

1 Introduction

Water-quality monitoring traditionally relies on water samples collected manually. The samples are then analyzed within laboratories to determine the water-quality variables of interest. This type of rigorous laboratory analysis of field-collected samples is crucial in making natural resources management decisions that affect human welfare and environmental conditions. However, with the rapid advances in hardware technology, the use of *in situ* water-quality sensors positioned at different geographic sites is becoming an increasingly common practice used to acquire real-time measurements of environmental and water-quality variables. Though only a subset of the required water-quality variables can be measured by these sensors, they have several advantages. Their ability to collect large quantities of data and to archive historic records allows for deeper analysis of water-quality variables to improve understanding about field conditions and water-quality processes (Glasgow, Burkholder, Reed, Lewitus, & Kleinman, 2004). Near-real-time monitoring also allows operators to identify and respond to potential issues quickly and thus manage the operations efficiently. Further, the use of *in situ* sensors can greatly reduce the labor involved in field sampling and laboratory analysis.

Water-quality sensors are exposed to changing environments and extreme weather conditions, and thus are prone to errors, including failure. Automated detection of outliers in water-quality data from *in situ* sensors has therefore captured the attention of many researchers both in the ecology and data science communities (Archer, Baptista, & Leen, 2003; Hill, Minsker, & Amir, 2009; Koch & McKenna, 2010; McKenna, Hart, Klise, Cruz, & Wilson, 2007; Raciti, Cucurull, & Nadjm-Tehrani, 2012). This problem of outlier detection in water-quality data from *in situ* sensors can be divided into two sub-topics according to their focus: (1) identifying errors in the data due to issues unrelated to water events per se, such as technical aberrations, that make the data unreliable and untrustworthy; and (2) identifying real events (e.g. rare but sudden spikes in turbidity associated with rare but sudden high-flow events). Both problems are equally important when making natural resource management decisions that affect human welfare and environmental conditions. Problem 1 can also be considered as a data preprocessing phase before addressing Problem 2.

In this work we focus on Problem 1, i.e. detecting unusual measurements caused by technical errors that make data unreliable and untrustworthy, and affect performance of any subsequent data analysis under Problem 2. According to Yu (2012), the degree of confidence in the sensor data is one of the main requirements for a properly defined

70 environmental analysis procedure. For instance, researchers and policy makers are un-
71 able to use water-quality data containing technical outliers with confidence for decision
72 making and reporting purposes because erroneous conclusions regarding the quality of
73 the water being monitored could ensue, leading, for example, to inappropriate or unnec-
74 essary water treatment, land management or warning alerts to the public (Kotamäki et
75 al., 2009; Rangeti, Dzwaïro, Barratt, & Otieno, 2015). Missing values and corrupted data
76 can also have an adverse impact on water-quality model building and calibration pro-
77 cesses (Archer et al., 2003). Early detection of these technical outliers will limit the use
78 of corrupted data for subsequent analysis. For instance, it will limit the use of corrupted
79 data in real-time forecasting and online applications such as on-line drinking water-quality
80 monitoring and early warning systems (Storey, Van der Gaag, & Burns, 2011), predict-
81 ing algal bloom outbreaks leading to fish kill events and potential human health impacts,
82 forecasting water level and currents, etc. (Archer et al., 2003; Glasgow et al., 2004; Hill
83 & Minsker, 2006). However, because data arrive near continuously at high speed in large
84 quantities, manual monitoring is highly unlikely to be able to capture all the errors. These
85 issues have therefore increased the importance of developing automated methods for early
86 detection of outliers in water-quality data from *in situ* sensors (Hill et al., 2009).

87 Different statistical approaches are available to detect outliers in water-quality data
88 from *in situ* sensors. For example, Hill and Minsker (2006) addressed the problem of out-
89 lier detection in environmental sensors using regression-based time series models. In this
90 work they addressed the scenario as a univariate problem. Their prediction models are
91 based on four data-driven methods: naive, clustering, perceptron, and Artificial Neural
92 Networks (ANN). Measurements that fell outside the bounds of an established predic-
93 tion interval were declared as outliers. They also considered two strategies: anomaly de-
94 tection (AD) and anomaly detection and mitigation (ADAM) for the detection process.
95 ADAM replaces detected outliers with the predicted value prior to the next predictions
96 whereas AD simply uses the previous measurements without making any alteration to
97 the detected outliers. These types of data-driven methods develop models using sets of
98 training examples containing a feature set and a target output. Later, Hill et al. (2009)
99 addressed the problem by developing three automated anomaly detection methods using
100 dynamic Bayesian networks (DBN) and showed that DBN-based detectors, using ei-
101 ther robust Kalman filtering or Rao-Blackwellized particle filtering, outperformed that
102 of Kalman filtering.

103 Another common approach for detecting outliers in environmental sensor data is
104 based on residuals (the differences between predicted and actual values). Due to the abil-
105 ity of ANNs to model a wide range of complex non-linear phenomena, Moatar, Fessant,
106 and Poirel (1999) used ANN techniques to detect anomalies such as abnormal values,
107 discontinuities and drifts in pH readings. After developing the pH model, the Student
108 t-test and the cumulative Page–Hinkley test were applied to detect changes in the mean
109 of the residuals to detect measurement error occurring over short periods of time. The
110 work was later expanded to a multivariate scenario with some additional water-quality
111 variables including dissolved oxygen, electrical conductivity, pH and temperature (Moatar,
112 Miquel, & Poirel, 2001). Their proposed algorithm used both deterministic and stochas-
113 tic approaches for the model building process. Observed data were then compared with
114 the model forecasts using a set of classical statistical tests to detect outliers, demonstrat-
115 ing the effectiveness and advantages of the multimodel approach. Later, Archer et al.
116 (2003) proposed a method to detect failures in the water-quality sensors due to biofoul-
117 ing based on a sequential likelihood ratio test. Their method also had the ability to pro-
118 vide estimates of biofouling onset time, which was useful for the subsequent step of out-
119 lier correction.

120 A common feature of all of the above methods is that they are usually employed
121 in a supervised or semi-supervised context and thus require training data pre-labeled with
122 known outliers or data that are free from the anomalous features of interest. In many

123 cases, however, not all the possible outliers are known in advance and can arise spon-
124 taneously as new outlying behaviors during the test phase. In such situations, supervised
125 methods may fail to detect those outliers. Semi-supervised methods are also unsuitable
126 for certain applications due to the unavailability of training data containing only typ-
127 ical instances that are free from outliers (Goldstein & Uchida, 2016). The datasets that
128 we consider in this paper suffer from both of these limitations highlighting the need for
129 a more general approach.

130 This paper develops a method for detecting technical outliers in water-quality data
131 derived from *in situ* sensors. Prior work by Leigh et al. (2019) emphasises the impor-
132 tance of different anomaly types and end-user needs and provides the starting point for
133 constructing a framework for automated anomaly detection in high frequency water-quality
134 data from *in situ* sensors. Their work briefly introduced unsupervised feature based meth-
135 ods for detecting technical-outliers in such data. The present paper differs substantially
136 from Leigh et al. (2019) as (1) the unsupervised feature based procedure we present for
137 detecting technical-outliers in high frequency water-quality data measured by *in situ* sen-
138 sors is its sole focus (2) the unsupervised feature based procedure is fully elaborated in
139 both details and depth and (3) the experimental results are enhanced through empha-
140 sis on the multivariate capabilities of the unsupervised feature based procedure. Further-
141 more, we focus on outliers involving abrupt changes in value, including sudden spikes,
142 sudden isolated drops and level shifts (high priority outliers as described in Leigh et al.
143 (2019)) rather than the broader suite considered by Leigh et al. (2019).

144 First, we present in detail our unsupervised feature based procedure that provides
145 early detection of technical outliers in water-quality data from *in situ* sensors. Rule-based
146 methods are also incorporated into the procedure to flag occurrences of impossible, out-
147 of-range, and missing values. Second, we provide a comparative analysis of the efficacy
148 and reliability of both density- and nearest neighbor distance-based outlier scoring tech-
149 niques. Third, we introduce an R (R Core Team, 2018) package, `oddwater` (Talagala &
150 Hyndman, 2019b) that implements the feature-based procedure and related functions.
151 Further, to facilitate reproducibility and reuse of the results presented in this paper, we
152 have made all of the code and associated datasets available on zenodo (Talagala & Hyn-
153 dman, 2019a).

154 Our feature-based procedure has many advantages: (1) it can take the correlation
155 structure of the water-quality variables into account when detecting outliers; (2) it can
156 be applied to both univariate and multivariate problems; (3) the outlier scoring techniques
157 that we consider are unsupervised, data-driven approaches and therefore do not require
158 training datasets for the model building process and can be extended easily to other time
159 series from other sites; (4) the outlier thresholds have a probabilistic interpretation as
160 they are based on extreme value theory; (5) the approach has the ability to deal with
161 irregular (unevenly spaced) time series; and (6) it can easily be extended to streaming
162 data. In contrast to a batch scenario, which assumes that the entire dataset is available
163 prior to the analysis with the focus on detecting complete events, the streaming data sce-
164 nario gives many additional challenges due to high velocity, unbounded, nonstationary
165 data with incomplete events (Hill et al., 2009; Talagala, Hyndman, Smith-Miles, Kan-
166 danaarachchi, & Muñoz, 2019). In this paper, although our `oddwater` procedure is in-
167 troduced as a batch method, it can easily be extended to streaming data such that it can
168 provide near-real-time support using a sliding window technique.

169 2 Materials and Methods

170 Our unsupervised feature-based procedure for detecting outliers in water-quality
171 data from *in situ* sensors has six main steps (Figure 1), and the structure of this section
172 is organised accordingly. For easy reference, we named our unsupervised feature-based

173 procedure as oddwater procedure, which stands for **Outlier Detection in Data from WA-**
174 **TER**-quality sensors.

175 **Figure 1.** Unsupervised feature-based procedure, named oddwater procedure for outlier de-
176 tection in water quality data from *in situ* sensors. Squares represents the main steps involved.
177 Circles correspond to input and output.

178 2.1 Study region and data

179 To evaluate the effectiveness of our oddwater procedure we considered a challeng-
180 ing real-world problem of monitoring water-quality using *in situ* sensors in a natural river
181 system. This is challenging because the system is susceptible to a wide range of envi-
182 ronmental, biological and human impacts that can lead to variation in water-quality and
183 affect the technological performance of the sensors. For comparison, we evaluated two
184 study sites, Sandy Creek and Pioneer River (PR), both in the Mackay-Whitsunday re-
185 gion of northeastern Australia (Mitchell, Brodie, & White, 2005). These two rivers flow
186 into the Great Barrier Reef lagoon and have catchment areas of 1466 km² and 326 km²,
187 respectively. In this region, the wet season typically occurs from December to April and
188 is dominated by higher rainfall and air temperatures, whereas the dry season typically
189 occurs from May to November with lower rainfall and air temperatures (McInnes et al.,
190 2015). The sensors at these two sites are housed within monitoring stations on the river
191 banks. Water is pumped from the rivers to the stations approximately every 60 or 90
192 minutes to take measurements of various water-quality variables that are logged by the
193 sensors. Here we focused on three water-quality variables: turbidity(NTU), conductiv-
194 ity (strictly, specific conductance at 25⁰C; μ S/cm) and river level (m).

195 The water-quality data obtained from *in situ* sensors located at Sandy Creek were
196 available from 12 March 2017 to 12 March 2018. The data set included 5402 recorded
197 points. These time series were irregular (i.e. the frequency of observations was not con-
198 stant) with a minimum time gap of 10 minutes and a maximum time gap of around 4
199 hours. The data obtained from Pioneer River were available from 12 March 2017 to 12
200 March 2018, and included 6303 recorded points. Many missing values were observed dur-
201 ing the initial part of all three series, i.e. turbidity, conductivity and river level, at Pi-
202 oneer River. With the help of a group of water-quality experts who were familiar with
203 the study region and with over 40 years of combined knowledge of river water quality,
204 observations were labeled as outliers or not, with the aim of evaluating the performance
205 of the procedure. Our Shiny web application available through the *oddwater* R package
206 was used during the labeling process to pinpoint observations and provide greater visual
207 insight into the data. Using this interactive visualization tool and expert knowledge, the
208 ground-truth labels were decided by consensus vote.

209 2.2 Apply rule-based approaches

210 Following Thottan and Ji (2003), we incorporated simple rules into our oddwater
211 procedure to detect outliers such as out-of-range values, impossible values (e.g. negative
212 values) and missing values, and labeled them prior to applying the statistical transfor-
213 mations introduced in Section 2.4.

214 If a sensor reading was outside the corresponding sensor detection range, it was marked
215 as an outlier. Negative readings are also inaccurate and impossible for river turbidity,
216 conductivity and level. We therefore imposed a simple constraint on the algorithm to
217 filter these values and mark them as outliers. Missing values are also frequently encoun-

218 tered in water-quality sensor data (Rangeti et al., 2015). We detected missing values by
 219 calculating the time gaps between readings. If a gap exceeded the maximum allowable
 220 time difference between any two consecutive readings, the corresponding time stamp was
 221 then marked as an outlier due to missingness. Here the maximum allowable time differ-
 222 ence was set at 180 minutes, given that the water-quality measurements were set to be
 223 taken at most every 90 minutes (measurements were often taken at higher frequencies
 224 during high-flow events, e.g. every 10-15 minutes, and occasionally as one-off measure-
 225 ments at times of interest to water managers).

226 2.3 Identify data features

227 After labeling out-of-range, impossible and missing values as outliers, further in-
 228 vestigation was done with the remaining observations. We initiated this investigation by
 229 identifying common characteristics or patterns of the possible types of outliers in water-
 230 quality data that would differentiate them from typical instances or events. For turbid-
 231 ity, for example, “extreme” deviations upward are more likely than deviations downward
 232 (Panguluri, Meiners, Hall, & Szabo, 2009). The opposite is true for conductivity (Tut-
 233 mez, Hatipoglu, & Kaymak, 2006). Further, in a turbidity time series, a sudden isolated
 234 upward shift (spike) is a point outlier (a single observation that is surprisingly large, in-
 235 dependent of the neighboring observations (Goldstein & Uchida, 2016)), but if the sud-
 236 den upward shift is followed by a gradually decaying tail then it becomes part of the typ-
 237 ical behavior. For river level, rates of rise are often fast compared with fall rates. In gen-
 238 eral, isolated data points that are outside the general trend are outliers. Further, nat-
 239 ural water processes under typical conditions generally tend to be comparatively slow;
 240 sudden changes therefore mostly correspond to outlying behaviors. Hereafter, these char-
 241 acteristics will be referred to as ‘data features’.

242 2.4 Apply statistical transformations

243 After identifying the data features, different statistical transformations were ap-
 244 plied to the time series to highlight different types of outliers focusing on sudden isolated
 245 spikes, sudden isolated drops, sudden shifts, and clusters of spikes (Table 1) that devi-
 246 ate from the typical characteristics of each variable (Leigh et al., 2019).

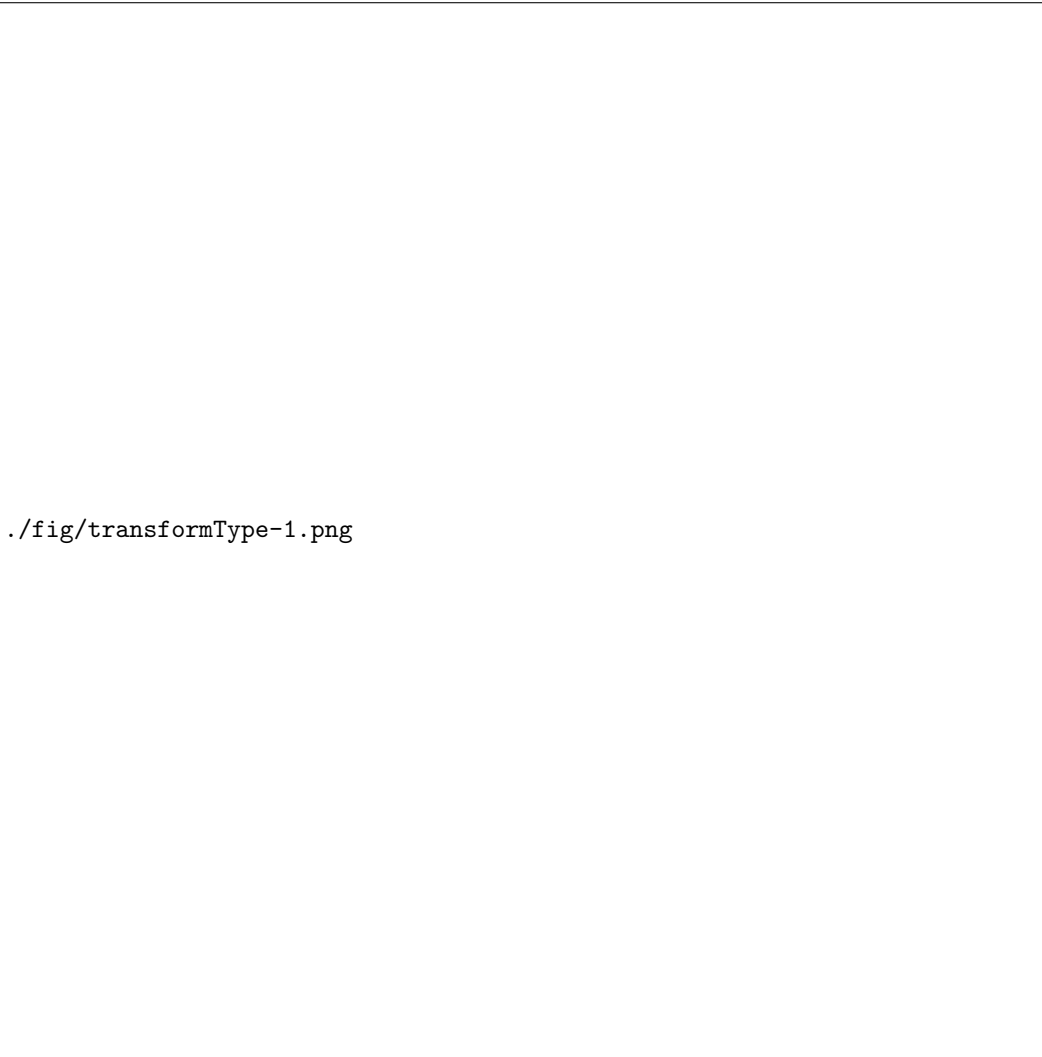
250 In this work, we considered the outlier detection problem in a multivariate setting.
 251 By applying different transformations on water-quality variables, we converted our orig-
 252 inal problem of outlier detection in the temporal context to a non-temporal context through
 253 a high dimensional data space with three dimensions defined by the three variables: tur-
 254 bidity, conductivity and river level. Different transformations were applied on different
 255 axes of the three dimensional data space resulting in different data patterns. We eval-
 256 uated the performance of the transformations (Dang & Wilkinson, 2014) using the max-
 257 imum separability of the two classes: outliers and typical points in the three dimensional
 258 data space. To provide a better visual illustration, in Figure 2 we present only the two
 259 dimensional data space defined by turbidity and conductivity; however, our actual data
 260 space is three dimensional. In this work our focus was to evaluate whether each point
 261 in time is an outlier or not such that an alarm could be triggered in the presence of an
 262 outlier. However, it was not our interest to investigate which variable(s) is (are) respon-
 263 sible for the outlier in time. Therefore, in Figure 2, a point is marked as an outlier in
 264 the two dimensional space if at least one variable corresponding to that point was labelled
 265 as an outlier by the water-quality experts.

266 When the transformation involves both the current value, Y_t , and the lagged value,
 267 Y_{t-1} , (as in the first difference and first derivative) both the outlier and immediate neigh-
 268 bour are highlighted in the transformed space. For example, if an outlier occurs at time
 269 point t , then the two values derived from the first derivative transformation ($(y_t - y_{t-1})$
 270 and $(y_{t+1} - y_t)$) are highlighted as outlying values, because they both involve y_t . There-

247 **Table 1.** Transformation methods used to highlight different types of outliers in water-quality
 248 sensor data. Let Y_t represent an original series from one of the three variables: turbidity, conduc-
 249 tivity and level at time t .

Data Feature	Requirement	Possible Transformation	Formula
High variability of the data.	Stabilize the variance across time series and make the patterns more visible (e.g. level shifts)	Log transformation	$\log(y_t)$
Isolated spikes (in both positive and negative directions) that are outside the general trend are considered as outliers. Under typical behavior, sudden upward (downward) shifts are possible for turbidity (conductivity), but their rate of fall (rise) is generally slower than the rate of rise (fall).	Separate isolated spikes from the general upward/downward trend patterns.	First difference	$\log(y_t/y_{t-1})$
Missing values in the data. the maximum allowable time difference between observations is 180 minutes.	Identify missing values.	Time gap	Δt
Data are unevenly spaced time series.	Handle irregular time series.	First derivative (Data points with large gaps will get small value. Large gaps indicate the lack of information to make a claim regarding the points.)	$x_t = \log(y_t/y_{t-1})/\Delta t$
One sided derivative Extreme upward trend in turbidity and level under typical behavior.	Separate spikes from typical upward trends.	<i>Turbidity or level</i>	$\min\{x_t, 0\}$
Extreme downward trend in conductivity under typical behavior.	Separate isolated drops from typical downward trends.	<i>Conductivity</i>	$\max\{x_t, 0\}$
High or low variability in the data.	Detect change points in variance.	Rate of change	$(y_t - y_{t-1})/y_t$
Natural processes are comparatively slow. Sudden changes (upward or downward movements) typically correspond to outlying instances.	Detect sudden changes (both upward and downward movements)	Relative difference	$y_t - (1/2)(y_{t-1} + y_{t+1})$

271 fore, each outlying instance is now represented by two consecutive values under the first
272 derivative or first difference transformation. As a result, one outlying instance is now rep-
273 resented by two points in the transformed data space (Figure 2(c, d)). The goal of the
274 one sided derivative transformation is to select only one high value as a representative
275 point for each outlying instance. However, the high values obtained could correspond
276 to either the actual outlying time point or the neighboring time point, because each trans-
277 formed value is derived from two consecutive observations. For example, in the data ob-
278 tained from Sandy Creek, the one sided derivative transformation (Figure 2(e)) clearly
279 separates all of the target outlying instances from the typical points using only one point
280 for each outlying instance, shown as either red triangles (corresponding to outliers) or
281 green squares (corresponding to the immediate neighbours of outliers). The second rep-
282 resentative member of each outlying instance mingles with the typical points, allowing
283 only one point to stand out on behalf of the corresponding outlying instance. If the pri-
284 mary focus of detecting technical outliers is to alert managers of sensor failures, then it
285 will be inconsequential if the alarm is triggered either at the actual time point correspond-
286 ing to the outlier or at the next immediate time point. However, if the purpose is dif-
287 ferent, such as producing a trustworthy dataset by labeling or correcting detected out-
288 liers, then additional conditions should be imposed to ensure that the time points de-
289 clared as outliers correspond to the actual outlying points and not to their immediate
290 neighboring points.



291 **Figure 2.** Bivariate relationships between transformed series of turbidity and conductivity
 292 measured by *in situ* sensors at Sandy Creek. In each scatter plot, outliers determined by water-
 293 quality experts are shown in red, while typical points are shown in black. Neighboring points are
 294 marked in green. (a) Original series, (b) Log transformation, (c) First difference, (d) First deriva-
 295 tive, (e) One sided derivative, and (f) Rate of change, (g) Relative difference (for original series),
 296 (h) Relative difference (for log transformed series). In each scatter plot, data are normalised such
 297 that they are bounded by the unit hypercube.

298 2.5 Calculate outlier scores

299 We considered eight commonly used, unsupervised outlier scoring techniques for
 300 high dimensional data involving nearest neighbor distances or densities of the observa-
 301 tions and applied them to the three dimensional data space defined by the three vari-
 302 ables: turbidity, conductivity and river level. Methods based on k -nearest neighbor dis-
 303 tances (where $k \in \mathbb{Z}^+$) were the NN-HD algorithm (details of this algorithm, which was
 304 inspired by HDoutliers algorithm (Wilkinson, 2018) are provided in Supporting Infor-
 305 mation), KNN-AGG and KNN-SUM algorithms (Angiulli & Pizzuti, 2002; Madsen, 2018)
 306 and Local Distance-based Outlier Factor (LDOF) algorithm (Zhang, Hutter, & Jin, 2009),
 307 which calculate the outlier score under the assumption that any outlying point (or out-

308 lying clusters of points) in the data space is (are) isolated; therefore the outliers are those
309 points having the largest k -nearest neighbor distances. In contrast, the density based Lo-
310 cal Outlier Factor (LOF) (Breunig, Kriegel, Ng, & Sander, 2000), Connectivity-based
311 Outlier Factor (COF) (Tang, Chen, Fu, & Cheung, 2002), Influenced Outlierness (IN-
312 FLO) (Jin, Tung, Han, & Wang, 2006) and Robust Kernel-based Outlier Factor (RKOF)
313 (Gao, Hu, Zhang, Zhang, & Wu, 2011) algorithms calculate an outlier score based on how
314 isolated a point is with respect to its surrounding neighbors, and therefore, the outliers
315 are those points having the lowest densities (see Supporting Information for detail). Each
316 algorithm assigns outlier scores for all of the data points in the high dimensional space
317 that describe the degree of outlierness of the individual data points such that outliers
318 are those points having the largest scores (Kriegel, Kröger, & Zimek, 2010; Shahid, Naqvi,
319 & Qaisar, 2015). This step allowed us to set a data driven threshold (Section 2.6) for
320 the outlier scores to select the most relevant outliers (Chandola, Banerjee, & Kumar, 2009).

321 2.6 Calculate outlier threshold

322 Following Schwarz (2008), Burrige and Taylor (2006) and Wilkinson (2018), we
323 used extreme value theory (EVT) to calculate a separate outlier threshold for each set
324 of outlier scores calculated using a given unsupervised outlier scoring technique (intro-
325 duced in Section 2.5) and assign a bivariate label for each point either as an outlier or
326 typical point. Thus, 8 outlier scoring techniques resulted 8 different thresholds for a given
327 dataset. The threshold calculation process started from a subset of data containing 50%
328 of observations with the smallest outlier scores, under the assumption that this subset
329 contained the outlier scores corresponding to typical data points and the remaining sub-
330 set contained the scores corresponding to the possible candidates for outliers. Follow-
331 ing Weissman’s spacing theorem (Weissman, 1978), the algorithm then fit an exponen-
332 tial distribution to the upper tail of the outlier scores of the first subset, and computed
333 the upper $1-\alpha$ (in this work α was set to 0.05) points of the fitted cumulative distri-
334 bution function, thereby defining an outlying threshold for the next outlier score. From
335 the remaining subset, the algorithm then selected the point with the smallest outlier score.
336 If this outlier score exceeded the cutoff point, all the points in the remaining subset were
337 flagged as outliers and searching for outliers ceased. Otherwise, the point was declared
338 as a non-outlier and was added to the subset of the typical points. The threshold was
339 then updated by including the latest addition. The searching algorithm continued un-
340 til an outlier score was found that exceeded the latest threshold (Schwarz, 2008). We per-
341 formed this threshold calculation under the assumption that the distribution of outlier
342 scores produced by each of the eight unsupervised outlier scoring techniques for high di-
343 mensional data was in the maximum domain of attraction of the Gumbel distribution,
344 which consists of distribution functions with exponentially decaying tails including the
345 exponential, gamma, normal and log-normal (Embrechts, Klüppelberg, & Mikosch, 2013).

346 2.7 Performance evaluation

347 In this paper, we focused on high priority outliers as described in Leigh et al. (2019)
348 in which importance ranking of different outlier types was done by taking into account
349 the end-user goals and the potential impact of outliers going undetected. However, it is
350 beyond the scope of this paper to discuss in detail the different types of outliers and their
351 importance ranking. For more detail, we refer the reader to Leigh et al. (2019). We per-
352 formed an experimental evaluation on the accuracy and computational efficiency of our
353 oddwater procedure with respect to the eight outlier scoring techniques using the dif-
354 ferent transformations (Table 1) and different combinations of variables (turbidity, con-
355 ductivity and river level). These experimental combinations were evaluated with respect
356 to common measures for binary classification based on the values of the confusion ma-
357 trix, which summarizes the false positives (FP; i.e. when a typical observation is mis-
358 classified as an outlier), false negatives (FN; i.e. when an actual outlier is misclassified

359 as a typical observation), true positives (TP; i.e. when an actual outlier is correctly clas-
 360 sified), and true negatives (TN; i.e. when an observation is correctly classified as a typ-
 361 ical point). In this work, false positives and false negatives are equally undesirable as false
 362 positives may demand unnecessary and/or expensive actions for corrections and refine-
 363 ment, and false negatives greatly reduce confidence in the data and results derived from
 364 them. The measures we considered include accuracy

$$accuracy = (TP + TN)/(TP + FP + FN + TN), \quad (1)$$

365 which explains the overall effectiveness of a classifier; and geometric-mean

$$GM = \sqrt{TP * TN}, \quad (2)$$

366 which explains the relative balance of TP and TN of the classifier (Sokolova & Lapalme,
 367 2009). According to Hossin and Sulaiman (2015), these measures are not enough to cap-
 368 ture the poor performance of the classifiers in the presence of imbalanced datasets where
 369 the size of the typical class (positive class) is much larger than the outlying class (neg-
 370 ative class). The datasets obtained from *in situ* sensors were highly imbalanced and neg-
 371 atively dependent (i.e. containing many more typical observations than outliers). There-
 372 fore, we used three additional measures that are recommended for imbalanced problems
 373 with only two classes (i.e. typical and outlying) by Ranawana and Palade (2006): the
 374 negative predictive value

$$NPV = TN/(FN + TN), \quad (3)$$

375 which measures the probability of a negatively predicted pattern actually being nega-
 376 tive; positive predictive value

$$PPV = TP/(TP + FP), \quad (4)$$

377 which measures the probability of a positively predicted pattern actually being positive;
 378 and optimized precision, which is a combination of accuracy, sensitivity and specificity
 379 metrics (Ranawana & Palade, 2006). The optimized precision is calculated as

$$OP = P - RI, \quad (5)$$

380 where

$$P = S_p N_n + S_n N_p \quad (6)$$

$$RI = |S_p - S_n|/(S_p + S_n) \quad (7)$$

$$S_p = TN/(TN + FP) \quad (8)$$

$$S_n = TP/(TP + FN) \quad (9)$$

384 and N_p and N_n represent the proportion of positives (outliers) and negatives (typical)
 385 within the entire dataset).

386 To evaluate the performance of our oddwater procedure, we incorporated additional
 387 steps after detecting the outlying time points using the outlying threshold based on EVT.
 388 This was done because the time points declared as outliers by the outlying threshold could
 389 correspond to either the actual outlying points or to their neighbors. Once the time points
 390 were declared as outliers, the corresponding points in the three dimensional space were
 391 further investigated by comparing their positions with respect to the median of the typ-
 392 ical points declared by the oddwater procedure. This step allowed us to find the most
 393 influential variable for each outlying point. For example, in Figure 2(e) the isolated point
 394 in the first quadrant is an outlier in the two dimensional space due to the outlying be-
 395 havior of the conductivity measurement. This allowed us because the deviation of this
 396 point from the median of the typical points (around (0, 0)) happens primarily along the
 397 conductivity axis. In contrast, the four isolated points in the third quadrant are outliers
 398 due to the outlying behavior of the turbidity measurement because the deviations of the

399 four points from the median of the typical points (around $(0, 0)$) happen primarily along
400 the turbidity axis. After detecting the most influential variable for each outlying instance
401 in the three dimensional space, further investigations were carried out separately for each
402 individual outlying instance with respect to the most influential variable detected. This
403 allowed us to see whether the outlying instance was due to a sudden spike or a sudden
404 drop by comparing the direction of the detected points with respect to the mean of its
405 two immediate surrounding neighbors and itself. These additional steps in the oddwater
406 procedure allowed us to trigger an alarm at the actual outlying point in time if the
407 neighboring points were declared as outliers instead of the actual outliers. However, we
408 acknowledge that these additional steps select only the most influential variable, not all
409 of the influential variables in the presence of more than one influential variable. The ad-
410 ditional steps were incorporated solely to measure the performance of the oddwater pro-
411 cedure. In practice, because the goal is to trigger an alarm in an occurrence of a tech-
412 nical outlier, it is inconsequential if the alarm is triggered either at the actual time point
413 or at the immediate neighbouring time points corresponding to the actual outlier. As
414 such, users of the oddwater procedure can ignore these additional steps.

415 Using the outlier threshold, our oddwater procedure assigns a bivariate label (ei-
416 ther as outlier or typical point) to each observed time point and thereby creates a vec-
417 tor of predicted class labels. That is, if a time point is declared as an outlier by oddwa-
418 ter procedure, then that could be due to at least one variable in the dataset. We also
419 declared each time point as an outlier or not based on the labels assigned by the water-
420 quality experts. At a given time point, if at least one variable was labeled as an outlier
421 by the water-quality experts then the corresponding time point was marked as an out-
422 lier, thereby creating a vector of ground-truth labels. Then, the performance measures
423 were calculated based on these two vectors of ground-truth labels and predicted class
424 labels. Thus, this performance evaluation was done with respect to the algorithm's abil-
425 ity to label a point in time as an outlier or not (i.e. a point in time is an outlier if the
426 observed value for any one or more of the three variables measured at that point in time
427 are outliers).

428 2.8 Software implementation

429 The oddwater procedure was implemented in the open source R package `oddwater`
430 (Talagala & Hyndman, 2019b), which provides a growing list of transformation and out-
431 lier scoring methods for high dimensional data together with visualization and perfor-
432 mance evaluation techniques. In addition to the implementations available through `oddwater`
433 package, `DDoutlier` package (Madsen, 2018) was also used for outlier score calculations.
434 We measured the computation time (mean execution time) using the `microbenchmark`
435 package (Mersmann, 2018) for different combinations of algorithms, transformations and
436 variable combinations on 28 core Xeon-E5-2680-v4 @ 2.40GHz servers. We also devel-
437 oped an R Shiny web application (available via `oddwater` R package) to provide inter-
438 active visual analytic tools to gain greater insight into the data and perform preliminary
439 investigations of the relationships between water-quality variables at different sites. To
440 facilitate reproducibility of the results presented herein, we have archived a snapshot of
441 version 0.7.0 of the R package on zenodo (Talagala & Hyndman, 2019a) along with the
442 code and datasets used. The latest version and on-going development of the `oddwater`
443 R package are available from Github (<https://github.com/pridiltal/oddwater>).

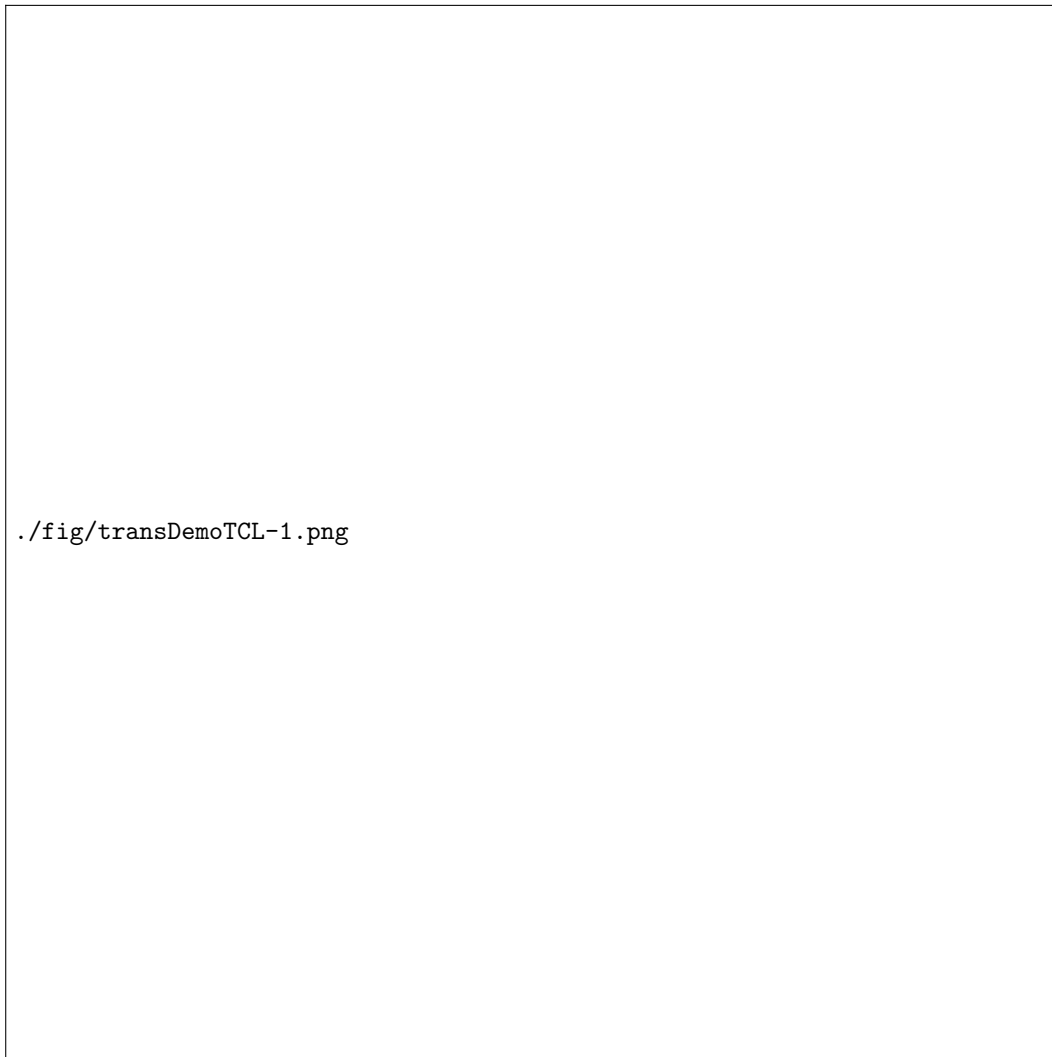
444 3 Results

445 3.1 Analysis of water-quality data from *in situ* sensors at Sandy Creek

446 A negative relationship was clearly visible between the water-quality variables tur-
447 bidity and conductivity and also between conductivity and river level measured by *in situ*
448 sensors at Sandy Creek (Figures 3 (a-i, b-i, c-i) and 4(a,c)). Further, no clear separa-

449 tion was observed between the target outliers and the typical points in the original data
 450 space (Figure 4(a-c)). However, a clear separation was apparent between the two sets
 451 of points once the one sided derivative transformation (an appropriate transformation
 452 for unevenly spaced data) was applied to the original series (Figures 4(d-f) and 3 (a-
 453 ii, b-ii, c-ii)).

454 KNN-AGG and KNN-SUM algorithms performed on all three water-quality vari-
 455 ables together using the one sided derivative transformation gave the highest OP (0.83)
 456 and NPV values(0.9996), which are the most recommended measurements for negatively
 457 dependent data where the focus is more on sensitivity (the proportion of positive pat-
 458 terns being correctly recognized as being positive) than specificity (Ranawana & Palade,
 459 2006).



460 **Figure 3.** Time series for turbidity (NTU) (a-i), conductivity ($\mu\text{S}/\text{cm}$) (b-i) and river level
 461 (m) (c-i) measured by *in situ* sensors at Sandy Creek. Transformed series (one sided derivatives)
 462 of turbidity (NTU) (a-ii), conductivity ($\mu\text{S}/\text{cm}$) (b-ii) and river level (m) (c-ii) measured by *in*
 463 *situ* sensors at Sandy Creek. In each plot outliers determined by water-quality experts are shown
 464 in red, while typical points are shown in black. Neighboring points are marked in green.



465 **Figure 4.** Top panel (a–c): Bi-variate relationships between original water-quality variables
 466 (turbidity (NTU), conductivity ($\mu\text{S}/\text{cm}$) and river level (m)) measured by *in situ* sensors at
 467 Sandy Creek. Bottom panel (d–f): Bi-variate relationships between transformed series (one sided
 468 derivative) of turbidity (NTU), conductivity ($\mu\text{S}/\text{cm}$) and river level (m) measured by *in situ*
 469 sensors at Sandy Creek. In each scatter plot, outliers determined by water-quality experts are
 470 shown in red, while typical points are shown in black. Neighboring points are marked in green.

475 Based on OP values, the one sided derivative transformation outperformed the first
 476 derivative transformation (Table 2, rows 1–2 compared to rows 3–4). Further, the distance-
 477 based outlier detection algorithms NN-HD, KNN-AGG and KNN-SUM outperformed
 478 all others (Table 2, rows 1–10 compared to rows 11–48). Among the three methods, the
 479 performance of k -nearest neighbor distance-based algorithms were only slightly higher
 480 (OP = 0.83) than the NN-HD algorithm (OP = 0.80), which is based only on the near-
 481 est neighbor distance. The algorithm combinations with the two highest OP values also
 482 had highest NPV (0.9996) and PPV (approximately 0.83). Furthermore, considering river
 483 level for the detection of outliers in the water-quality sensors slightly improved the per-
 484 formance (OP = 0.83). Among the analysis with transformed series, LOF with the first
 485 derivative transformation performed the least well (OP = 0.25). For most of the outlier
 486 detection algorithms (KNN-SUM, KNN-AGG, NN-HD, COF, LOF and INFLO) the poor-

471 **Table 2.** Performance metrics of outlier detection algorithms performed on multivariate water-
 472 quality time series data (T, turbidity; C, conductivity; L, river level) from *in situ* sensors at
 473 Sandy Creek, arranged in descending order of OP values. See Sections 2.7-8 for performance
 474 metric codes and details.

i	Variables	Transformation	Method	Accuracy	GM	OP	PPV	NPV	Time (mean)
1	T-C-L	One sided Derivative	KNN-AGG	0.9994	164.23	0.83	0.83	0.9996	404.0
2	T-C-L	One sided Derivative	KNN-SUM	0.9994	164.23	0.83	0.83	0.9996	186.8
3	T-C	First Derivative	NN-HD	0.9991	146.87	0.80	0.57	0.9996	45.0
4	T-C	First Derivative	KNN-AGG	0.9989	146.86	0.80	0.50	0.9996	415.8
5	T-C	One sided Derivative	NN-HD	0.9996	146.91	0.80	1.00	0.9996	112.9
6	T-C	One sided Derivative	KNN-AGG	0.9994	146.90	0.80	0.80	0.9996	411.7
7	T-C	One sided Derivative	KNN-SUM	0.9994	146.90	0.80	0.80	0.9996	190.4
8	T-C-L	First Derivative	KNN-AGG	0.9993	127.22	0.60	1.00	0.9993	404.4
9	T-C-L	First Derivative	KNN-SUM	0.9993	127.22	0.60	1.00	0.9993	188.9
10	T-C	First Derivative	KNN-SUM	0.9993	103.88	0.50	1.00	0.9993	189.5
11	T-C	First Derivative	LDOF	0.9991	103.87	0.50	0.67	0.9993	17444.7
12	T-C	One sided Derivative	LDOF	0.9991	103.87	0.50	0.67	0.9993	17253.8
13	T-C-L	First Derivative	NN-HD	0.9991	103.87	0.44	1.00	0.9991	52.5
14	T-C-L	First Derivative	INFLO	0.9965	103.74	0.44	0.12	0.9991	1107.9
15	T-C-L	First Derivative	COF	0.9987	103.86	0.44	0.50	0.9991	5939.8
16	T-C-L	First Derivative	RKOF	0.9963	103.73	0.44	0.12	0.9991	369.7
17	T-C-L	One sided Derivative	NN-HD	0.9991	103.87	0.44	1.00	0.9991	118.2
18	T-C-L	One sided Derivative	INFLO	0.9985	103.85	0.44	0.40	0.9991	1113.6
19	T-C-L	One sided Derivative	COF	0.9987	103.86	0.44	0.50	0.9991	5787.4
20	T-C-L	One sided Derivative	LDOF	0.9985	103.85	0.44	0.40	0.9991	17261.9
21	T-C-L	One sided Derivative	LOF	0.9985	103.85	0.44	0.40	0.9991	516.9
22	T-C-L	One sided Derivative	RKOF	0.9976	103.80	0.44	0.20	0.9991	370.5
23	T-C-L	Original series	KNN-AGG	0.9989	103.87	0.44	0.67	0.9991	391.6
24	T-C-L	Original series	INFLO	0.9974	103.79	0.44	0.18	0.9991	1070.7
25	T-C-L	Original series	LDOF	0.9987	103.86	0.44	0.50	0.9991	17156.9
26	T-C-L	Original series	RKOF	0.9985	103.85	0.44	0.40	0.9991	354.0
27	T-C	First Derivative	INFLO	0.9983	73.43	0.28	0.20	0.9991	1194.9
28	T-C	First Derivative	COF	0.9991	73.46	0.28	1.00	0.9991	5991.8
29	T-C	First Derivative	LOF	0.9987	73.44	0.28	0.33	0.9991	512.3
30	T-C	First Derivative	RKOF	0.9983	73.43	0.28	0.20	0.9991	363.2
31	T-C	One sided Derivative	INFLO	0.9987	73.44	0.28	0.33	0.9991	1207.0
32	T-C	One sided Derivative	COF	0.9987	73.44	0.28	0.33	0.9991	5880.8
33	T-C	One sided Derivative	LOF	0.9969	73.38	0.28	0.08	0.9991	511.3
34	T-C	One sided Derivative	RKOF	0.9961	73.35	0.28	0.06	0.9991	368.3
35	T-C	Original series	KNN-AGG	0.9989	73.45	0.28	0.50	0.9991	405.1
36	T-C	Original series	INFLO	0.9974	73.40	0.28	0.10	0.9991	1143.6
37	T-C	Original series	LDOF	0.9987	73.44	0.28	0.33	0.9991	17022.9
38	T-C	Original series	RKOF	0.9985	73.44	0.28	0.25	0.9991	351.8
39	T-C-L	First Derivative	LDOF	0.9989	73.45	0.25	1.00	0.9989	17323.2
40	T-C-L	First Derivative	LOF	0.9989	73.45	0.25	1.00	0.9989	517.1
41	T-C-L	Original series	NN-HD	0.9987	73.44	0.25	0.50	0.9989	48.6
42	T-C-L	Original series	KNN-SUM	0.9989	73.45	0.25	1.00	0.9989	177.3
43	T-C-L	Original series	COF	0.9989	73.45	0.25	1.00	0.9989	5931.7
44	T-C-L	Original series	LOF	0.9989	73.45	0.25	1.00	0.9989	505.0
45	T-C	Original series	NN-HD	0.9987	0.00	0.00	0.00	0.9989	41.7
46	T-C	Original series	KNN-SUM	0.9989	0.00	0.00	NaN	0.9989	184.6
47	T-C	Original series	COF	0.9989	0.00	0.00	NaN	0.9989	5896.4
48	T-C	Original series	LOF	0.9989	0.00	0.00	NaN	0.9989	502.7

487 est performances were associated with the untransformed original series, having the low-
 488 est OP and NPV values, highlighting how data transformation can improve the ability
 489 of outlier detection algorithms while maintaining low false detection rates.



490 **Figure 5.** Classification of outlier scores produced from different algorithms as true negatives
 491 (TN), true positives (TP), false negatives (FN), false positives (FP). The top three panels (i,
 492 ii, iii) correspond to the original series (turbidity, conductivity and river level) measured by *in*
 493 *situ* sensors at Sandy Creek. The target outliers (detected by water-quality experts) are shown
 494 in red, while typical points are shown in black. The remaining panels (a-h) give outlier scores
 495 produced by different outlier detection algorithms for high dimensional data when applied to the
 496 transformed series (one sided derivative) of the three variables: turbidity, conductivity and level.
 497 Through different outlier scoring algorithms (Panel a - h), we are evaluating whether each point
 498 in time is an outlier or not. Therefore, from Panel a-h, if the outlier scoring algorithm is effective,
 499 then there should be either TP or TN at each point in time when either a red triangle is plotted
 500 in at least one of the three panels (i- iii), or black dots are plotted in all of the top three panels
 501 (i - iii), respectively. Because outlier scores are non negative and are mostly clustered near zero,
 502 with some occasional high values, a square root transformation was applied to reduce skewness of
 503 the data in Panel (a) to (h).

504 The three outlier detection algorithms that demonstrated the highest level of ac-
 505 curacy (NN-HD, KNN-AGG and KNN-SUM) also outperformed the others with respect
 506 to computational time. NN-HD algorithm required the least computational time. Among
 507 the remaining two, the mean computational time of KNN-AGG (≈ 400 milliseconds) was
 508 twice that of KNN-SUM's (< 200 milliseconds). LOF and its extensions (INFLO, COF
 509 and LDOF) demonstrated the poorest performance with respect computational time ($>$
 510 500 milliseconds on average).

511 Only KNN-SUM and KNN-AGG assigned high scores to most of the targeted out-
 512 liers in turbidity, conductivity and level data transformed using the one-sided derivative
 513 (Figure 5(a,b)). For each outlying instance, however, the next immediate neighboring
 514 point was assigned the high outlier score instead of the true outlying point. After de-
 515 termining the most influential variable using the additional steps of the algorithm (Sec-
 516 tion 2.7), adjustments were made to correct this to the actual outlier. Because of this
 517 correction, the first orange triangle for the True Positive in Figure 5(a – h), for instance,
 518 is always plotted next to the high outlier score (corresponding to the neighboring point),
 519 pointing to the actual outlier instead of the neighbouring point. The outlier scores pro-
 520 duced by LOF and COF (Figure 5(d,e)) were unable to capture the outlying behaviors
 521 correctly and demonstrated high scattering. In comparison to other outlier scoring al-
 522 gorithms, KNN-SUM algorithm displayed a good compromise between accuracy and com-
 523 putational efficiency (Table 2).

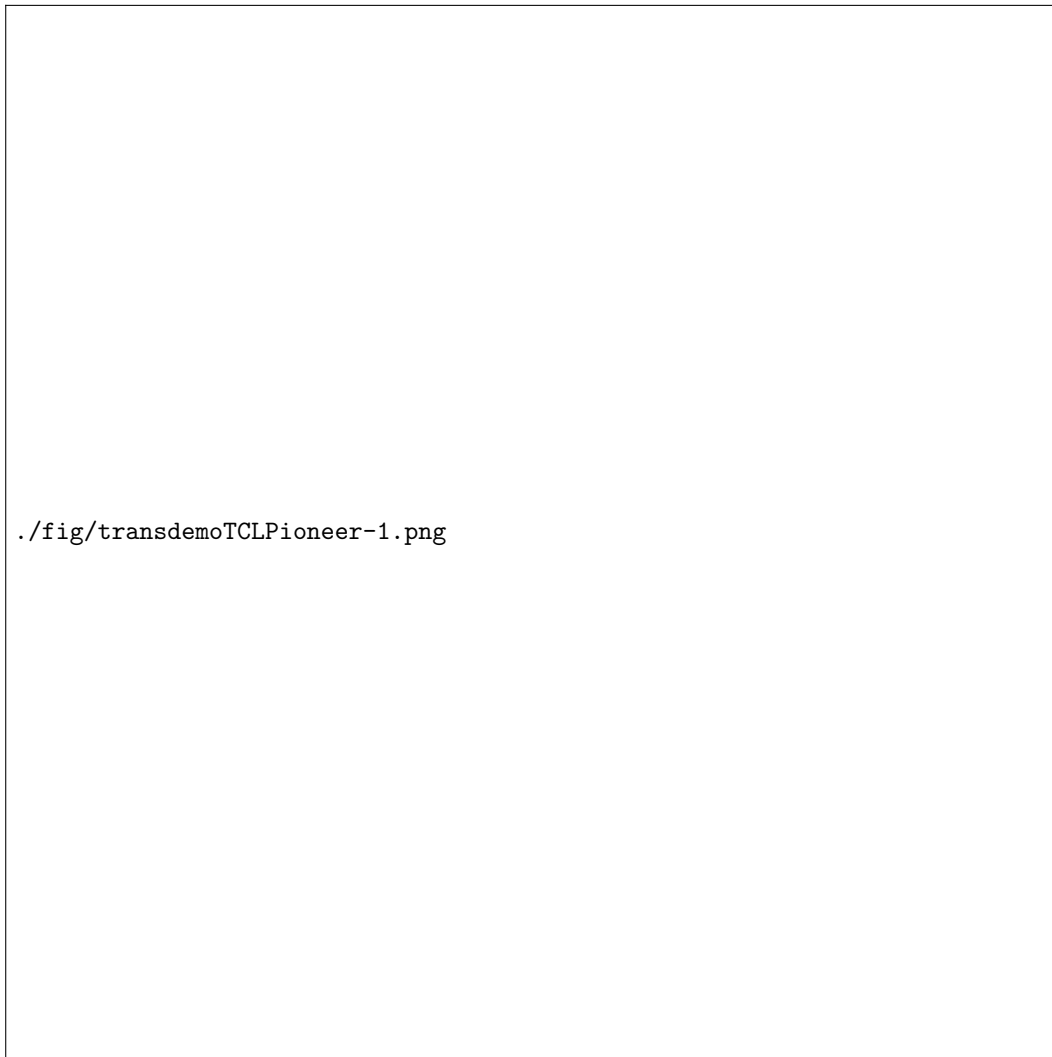
524 3.2 Analysis of water-quality data from *in situ* sensors at Pioneer River

525 Compared to Sandy Creek where the river level is mostly less than 1 meter with
 526 occasional bursts of atypical spikes and flow events resulting in levels up to 14.8 meters
 527 (Figure 3 (c-i)), Pioneer River is much deeper with the river level ranging between 13.9
 528 and 16.5 metres during the period of study (Figure 6 (c-i)). Two small dense clusters
 529 of points gathered around zero were observed for all three variables from late March to
 530 mid April in 2017 (Figure 6). These co-occurrences of values around zero are atypical
 531 behaviour and may have been due to technical issues with the sensor equipment. These
 532 type of anomalies can be easily detected by incorporating rule based methods.

533 Some of the target outliers in the data obtained from the *in situ* sensors at Pio-
 534 nner River only deviated slightly from the general trend (Figure 6 (a-i)), making out-
 535 lier detection challenging. A negative relationship was clearly visible between turbidity
 536 and conductivity (Figure 7(a)), however, the relationship between level and conductiv-
 537 ity was complex (Figure 7(c)). Most of the target outliers were masked by the typical
 538 points in the original space (Figure 7(a-c)). Similar to Sandy Creek, data obtained from
 539 the sensors at Pioneer River showed good separation between outliers and typical points
 540 under the one sided derivative transformation (Figures 7(d-f) and 6 (a-ii, b-ii, c-ii)).
 541 However, the sudden spikes in turbidity labeled as outliers by water-quality experts could
 542 not be separated from the majority by a large distance and were only visible as a small
 543 group (micro cluster (Goldstein & Uchida, 2016)) in the boundary defined by the typ-
 544 ical points (Figure 7(d, e)).

545 From the performance analysis, it was observed that turbidity and conductivity to-
 546 gether produced better results (Table 3, rows 1–8) than when combined with river level,
 547 which tended to reduce the performance (i.e. generating lower OP and NPV values) while
 548 increasing the false negative rate (Table 3, rows 9–13). KNN-AGG and KNN-SUM (Ta-
 549 ble 3, rows 2–3) had the highest accuracy (0.9978), highest geometric means (492.8012),
 550 highest OP (0.88) and highest NPV (0.9984). Despite the challenge given by the small
 551 spikes which could not be clearly separated from the typical points, KNN-AGG, KNN-
 552 SUM and NN-HD with one sided derivatives of turbidity and conductivity still detected
 553 some of those points as outliers while maintaining low false negative and false positive
 554 rates. Similar to Sandy Creek, NN-HD (< 200 milliseconds on average) and KNN-SUM

555 (< 230 milliseconds on average) demonstrated the highest computational efficiency for
556 the data obtained from Pioneer River.



557 **Figure 6.** Time series for turbidity (NTU) (a-i), conductivity ($\mu\text{S}/\text{cm}$) (b-i) and river level
558 (m) (c-i) measured by *in situ* sensors at Pioneer River. Transformed series (one sided deriva-
559 tives) of turbidity (NTU) (a-ii), conductivity ($\mu\text{S}/\text{cm}$) (b-ii) and river level (m) (c-ii) measured
560 by *in situ* sensors at Pioneer River. In each plot, outliers determined by water-quality experts are
561 shown in red, while typical points are shown in black. Neighboring points are marked in green.



562 **Figure 7.** Top panel (a–c): Bi-variate relationships between original water-quality variables
563 (turbidity (NTU), conductivity ($\mu\text{S}/\text{cm}$) and river level (m)) measured by *in situ* sensors at Pi-
564 oneer River. Bottom panel (d–f): Bi-variate relationships between transformed series (one sided
565 derivative) of turbidity (NTU), conductivity ($\mu\text{S}/\text{cm}$) and river level (m) measured by *in situ*
566 sensors at Pioneer River. In each scatter plot, outliers determined by water-quality experts are
567 shown in red, while typical points are shown in black. Neighboring points are marked in green.



568 **Figure 8.** Classification of outlier scores produced from different algorithms as true negatives
 569 (TN), true positives (TP), false negatives (FN), false positives (FP). The top two panels (i and
 570 ii) correspond to the original series (turbidity and conductivity) measured by *in situ* sensors at
 571 Pioneer River. The target outliers (detected by water-quality experts) are shown in red, while
 572 typical points are shown in black. The remaining panels (a–h) give outlier scores produced by
 573 different outlier detection algorithms for high dimensional data when applied to the transformed
 574 series (one sided derivative) of the two variables: turbidity and conductivity. Through differ-
 575 ent outlier scoring algorithms (Panel a - h), we are evaluating whether each point in time is an
 576 outlier or not. Therefore, from Panel a-h, if the outlier scoring algorithm is effective, then there
 577 should be either TP or TN at each point in time when either a red triangle is plotted in at least
 578 one of the two panels (i - ii), or black dots are plotted in both of the top two panels (i - ii), re-
 579 spectively. Because outlier scores are non negative and are mostly clustered near zero, with some
 580 occasional high values, a square root transformation was applied to reduce skewness of the data
 581 in Panel (a) to (h).

582 **Table 3.** Performance metrics of outlier detection algorithms performed on multivariate water-
 583 quality time series data (T, turbidity; C, conductivity; L, river level) from *in situ* sensors at
 584 Pioneer River, arranged in descending order of OP values. See Sections 2.7-8 for performance
 585 metric codes and details.

i	Variables	Transformation	Method	Accuracy	GM	OP	PPV	NPV	Time (mean)
1	T-C	One sided Derivative	NN-HD	0.9976	492.76	0.88	0.89	0.9984	136.5
2	T-C	One sided Derivative	KNN-AGG	0.9978	492.80	0.88	0.91	0.9984	478.8
3	T-C	One sided Derivative	KNN-SUM	0.9978	492.80	0.88	0.91	0.9984	222.2
4	T-C	First Derivative	NN-HD	0.9978	480.08	0.86	0.95	0.9981	182.0
5	T-C	First Derivative	KNN-AGG	0.9978	480.08	0.86	0.95	0.9981	488.5
6	T-C	First Derivative	KNN-SUM	0.9978	480.08	0.86	0.95	0.9981	225.3
7	T-C	First Derivative	INFLO	0.9971	479.92	0.86	0.86	0.9981	1525.0
8	T-C	First Derivative	RKOF	0.9970	479.88	0.86	0.84	0.9981	430.4
9	T-C-L	One sided Derivative	KNN-AGG	0.9975	492.72	0.86	0.91	0.9981	465.2
10	T-C-L	One sided Derivative	KNN-SUM	0.9975	492.72	0.86	0.91	0.9981	214.5
11	T-C-L	First Derivative	RKOF	0.9951	485.82	0.85	0.68	0.9979	425.9
12	T-C-L	First Derivative	KNN-AGG	0.9975	480.00	0.84	0.95	0.9978	478.0
13	T-C-L	First Derivative	KNN-SUM	0.9975	480.00	0.84	0.95	0.9978	220.0
14	T-C	First Derivative	COF	0.9978	473.58	0.84	0.97	0.9979	7908.2
15	T-C	First Derivative	LDOF	0.9978	473.58	0.84	0.97	0.9979	23435.7
16	T-C	First Derivative	LOF	0.9975	473.51	0.84	0.92	0.9979	594.4
17	T-C	One sided Derivative	INFLO	0.9973	473.47	0.84	0.90	0.9979	1559.9
18	T-C	One sided Derivative	COF	0.9976	473.54	0.84	0.95	0.9979	7505.5
19	T-C	One sided Derivative	LDOF	0.9975	473.51	0.84	0.92	0.9979	22986.0
20	T-C	One sided Derivative	LOF	0.9975	473.51	0.84	0.92	0.9979	596.9
21	T-C	One sided Derivative	RKOF	0.9960	473.16	0.84	0.75	0.9979	419.7
22	T-C	Original Series	INFLO	0.9973	473.47	0.84	0.90	0.9979	1498.5
23	T-C-L	First Derivative	COF	0.9975	473.51	0.83	0.97	0.9976	7910.7
24	T-C-L	First Derivative	LDOF	0.9975	473.51	0.83	0.97	0.9976	23357.7
25	T-C-L	One sided Derivative	NN-HD	0.9975	473.51	0.83	0.97	0.9976	131.9
26	T-C	Original Series	NN-HD	0.9976	466.96	0.83	0.97	0.9978	171.0
27	T-C	Original Series	KNN-AGG	0.9970	466.81	0.83	0.88	0.9978	468.7
28	T-C	Original Series	KNN-SUM	0.9970	466.81	0.83	0.88	0.9978	211.6
29	T-C	Original Series	COF	0.9978	467.00	0.83	1.00	0.9978	7617.6
30	T-C	Original Series	LDOF	0.9978	467.00	0.83	1.00	0.9978	22910.4
31	T-C	Original Series	LOF	0.9978	467.00	0.83	1.00	0.9978	579.1
32	T-C	Original Series	RKOF	0.9963	466.66	0.83	0.80	0.9978	401.9
33	T-C-L	First Derivative	NN-HD	0.9973	473.47	0.82	0.95	0.9976	167.1
34	T-C-L	One sided Derivative	INFLO	0.9971	473.43	0.82	0.92	0.9976	1418.8
35	T-C-L	One sided Derivative	COF	0.9973	473.47	0.82	0.95	0.9976	7497.9
36	T-C-L	One sided Derivative	LDOF	0.9973	473.47	0.82	0.95	0.9976	23090.7
37	T-C-L	One sided Derivative	RKOF	0.9952	472.97	0.82	0.71	0.9976	422.1
38	T-C-L	First Derivative	INFLO	0.9975	466.92	0.81	1.00	0.9974	1398.3
39	T-C-L	First Derivative	LOF	0.9975	466.92	0.81	1.00	0.9974	600.7
40	T-C-L	One sided Derivative	LOF	0.9965	466.70	0.81	0.85	0.9974	596.1
41	T-C-L	Original Series	NN-HD	0.9973	466.88	0.81	0.97	0.9974	163.0
42	T-C-L	Original Series	KNN-AGG	0.9967	466.73	0.81	0.88	0.9974	456.3
43	T-C-L	Original Series	KNN-SUM	0.9967	466.73	0.81	0.88	0.9974	201.4
44	T-C-L	Original Series	INFLO	0.9975	466.92	0.81	1.00	0.9974	1372.8
45	T-C-L	Original Series	COF	0.9975	466.92	0.81	1.00	0.9974	7707.2
46	T-C-L	Original Series	LDOF	0.9975	466.92	0.81	1.00	0.9974	127337.1
47	T-C-L	Original Series	LOF	0.9975	466.92	0.81	1.00	0.9974	580.9
48	T-C-L	Original Series	RKOF	0.9955	466.47	0.81	0.74	0.9974	406.8

4 Discussion

We introduced a new procedure, named oddwater procedure for the detection of outliers in water-quality data from *in situ* sensors, where outliers were specifically defined as due to technical errors that make the data unreliable and untrustworthy. We showed that our oddwater procedure, with carefully selected data transformation methods derived from data features, can greatly assist in increasing the performance of a range of existing outlier detection algorithms. Our oddwater procedure and analysis using data obtained from *in situ* sensors positioned at two study sites, Sandy Creek and Pioneer River, performed well with outlier types such as sudden isolated spikes, sudden isolated drops and level shifts, while maintaining low false detection rates. As an unsupervised procedure, our approach can be easily extended to other water-quality variables, other sites and also to other outlier detection tasks in other application domains. The only requirement is to select suitable transformation methods according to the data features that differentiate the outlying instances from the typical behaviors of a given system.

Studies have shown that transforming variables affects densities, relative distances and orientation of points within the data space and therefore can improve the ability to perceive patterns in the data which are not clearly visible in the original data space (Dang & Wilkinson, 2014). This was the case in our study where no clear separation was visible between outliers and typical data points in the original data space but a clear separation was obtained between the two sets of points once the one-sided derivative transformation was applied to the original series. Having this type of a separation between outliers and typical points is important before applying unsupervised outlier detection algorithms for high dimensional data because the methods are usually based on the definition of outliers in terms of distance or density (Talagala, Hyndman, Smith-Miles, Kandanarachchi, & Muñoz, 2019). Most of the outlier detection algorithms (KNN-SUM, KNN-AGG, NN-HD, COF, LOF and INFLO) performed least well with the untransformed original series, demonstrating how data transformation methods can assist in improving the ability of outlier detection algorithms while maintaining low false detection rates.

In our modified algorithm, the NN-HD algorithm, we did not incorporate the clustering step of the HDoutliers algorithm because the data obtained from the two study sites are free from micro clusters (Talagala, Hyndman, & Smith-Miles, 2019) and therefore free from the masking problem. Because the datasets have only local and global outliers, incorporating a clustering step that forms small clusters using a small ball with a fixed radius (the Leader Algorithm in Wilkinson (2018)) does not significantly change the structure of the data points in the high dimensional data space. Furthermore, because NN-HD has the additional requirement of isolation in addition to clear separation between outlying points and typical points, it performed poorly in comparison to the two KNN distance-based algorithms (KNN-AGG and KNN-SUM) which are not restricted to the single most nearest neighbor (Talagala, Hyndman, & Smith-Miles, 2019). For the current work, k was set to 10, the maximum default value of k in Madsen (2018), because too large a value of k could skew the focus towards global outliers (points that deviates significantly from the rest of the dataset) alone (Zhang et al., 2009) and make the algorithms computationally inefficient. On the other hand, too small a value of k could incorporate an additional assumption of isolation into the algorithm, as in the NN-HD algorithm where $k = 1$. Among the analyses using transformed series, LOF with the first derivative transformation performed the least well, which could also be due to its additional assumption of isolation (Tang et al., 2002). However, using the same k across all algorithms may bias direct comparison because the performance of the algorithms can depend on the value of k and algorithms can reach their peak performance for different choices of k (Campos et al., 2016). Therefore, performing an optimisation to select the best k is non trivial and we leave it for future work.

We took the correlation structure between the variables into account when detecting outliers given some were apparent only in the high dimensional space but not when

639 each variable was considered independently (Ben-Gal, 2005). A negative relationship was
640 observed between conductivity and turbidity and also between conductivity and level
641 for the Sandy Creek data. However, for Pioneer River, no clear relationship was observed
642 between level and the remaining two variables, turbidity and conductivity. This could
643 be one reason why the variable combination with river level gave poor results for the Pi-
644 oneer River dataset, while results for other combinations were similar to those of Sandy
645 Creek. The one-sided derivative transformation outperformed the derivative transfor-
646 mation. This was expected, because in an occurrence of a sudden spike or isolated drop,
647 the first derivative assigns high values to two consecutive points, the actual outlying point
648 as well as the neighboring point, and therefore increases the false positive rate (because
649 the neighboring points that are declared to be outliers actually correspond to typical points
650 in the original data space). Therefore, to detect technical outliers in water-quality data
651 from Sandy Creek and Pioneer River, the one sided derivative transformation is recom-
652 mended because it outperformed the other transformations during the comparative anal-
653 ysis. For Sandy Creek, all three water-quality variables together with the one-sided deriva-
654 tive transformation is recommended. However, for Pioneer River, the use of river level
655 is not advisable due its complex relationships with the other variables and its tempo-
656 ral variability. For both rivers, the use of KNN-SUM algorithm is recommended because
657 it provides a good compromise between accuracy and computational efficiency.

658 In this study, our goal was to detect suitable transformations, combinations of vari-
659 ables, and the algorithms for outlier score calculation for the data from two study sites.
660 Results may depend on the characteristics of the time series (site and time dependent
661 for example), and what is best for one site may not be the best for another site. There-
662 fore, care should be taken to select transformations most suitable for the problem at hand.
663 According to Dang and Wilkinson (2014), any transformation used on a dataset must
664 be evaluated in terms of a figure of merit (i.e. a numerical quantity used to character-
665 ize the performance of a method, relative to its alternatives). For our work on detect-
666 ing outliers, the figure of merit was the maximum separability of the two classes gener-
667 ated by outliers and typical points. However, we acknowledge that the set of transfor-
668 mations that we used for this work was relatively limited and influenced by the data ob-
669 tained from the two study sites. Therefore, the set of transformations we considered (Ta-
670 ble 1) should be viewed only as an illustration of our oddwater procedure for detecting
671 outliers. We expect that the set of transformations will expand over time as the oddwa-
672 ter procedure is used for other data from other study sites and for applications to other
673 fields.

674 For the current work, we selected transformation methods that could highlight abrupt
675 changes in the water-quality data. We hope to expand the ability of oddwater procedure
676 so that it can detect other outlier types not previously targeted but commonly observed
677 in water-quality data (e.g. low/high variability, drift etc. as per Leigh et al. (2019)). One
678 possibility is to consider the residuals at each point, defined as the difference between
679 the actual values and the fitted values (similar to Schwarz (2008)) or the difference be-
680 tween the actual values and the predicted values (similar to Hill and Minsker (2006)),
681 as a transformation and apply outlier detection algorithms to the high dimensional space
682 defined by those residuals. Here the challenge will be to identify the appropriate curve
683 fitting and prediction models to generate the residual series. In this way, continuous sub-
684 sequences of high values could correspond to other kinds of technical outliers such as high
685 variability or drift. However, the range of applications and the space of the transforma-
686 tions are extremely diverse, which makes it challenging to provide a structured formal
687 vision that covers all of the possible transformations that could be considered. The trans-
688 formations we present in this paper were mainly chosen as appropriate to the data col-
689 lected from Sandy Creek and Pioneer River. We observed that different transformations
690 can lead to entirely different data structures and that the selection of suitable transfor-
691 mations is directed by the data features and typical patterns imposed by a given appli-
692 cation. Domain specific knowledge plays a vital role when selecting suitable transforma-

693 tions and, as such, defining structured guidelines for the selection of suitable transfor-
694 mations remains problematic.

695 Not surprisingly, NN-HD algorithm required the least computational time given
696 the outlying score calculation only involves searching for the single most nearest neigh-
697 bors of each test point (Wilkinson, 2018). The mean computational time of KNN-AGG
698 was twice as high as that of KNN-SUM because the KNN-AGG algorithm has the addi-
699 tional requirement of calculating weights that assign nearest neighbors higher weight
700 relative to the neighbors farther apart (Angiulli & Pizzuti, 2002). LOF and its exten-
701 sions (INFLO, COF and LDOF) required the most computational time; all four algo-
702 rithms involve a two step searching mechanism at each test point when calculating the
703 corresponding outlying score. This means that at each test point each algorithm searches
704 its k nearest neighbors as well those of the detected nearest neighbors for the outlier score
705 calculation (Breunig et al., 2000; Jin et al., 2006; Tang et al., 2002; Zhang et al., 2009).

706 Assessing performance of the detection methods based on the classification criteria,
707 while traditional, has limitations. During performance evaluation, we observed that
708 some outliers were detected by all the approaches, some were detected as outliers only
709 by certain methods and some were identified by no method. Therefore, incorporating
710 ensemble methods as proposed in Unwin (2019) would assist in selecting the best per-
711 forming approaches for a particular outlier type and enable further insight into the re-
712 sults obtained from the oddwater procedure.

713 We hope to extend our multivariate outlier detection framework into space and time
714 so that it can deal with the spatio-temporal correlation structure along branching river
715 networks. Further, in the current paper, we have introduced our oddwater procedure as
716 a batch method. However, due to the unsupervised nature of our oddwater procedure
717 it can be easily extended to a streaming data scenario with the help of a sliding window
718 of fixed length. A streaming data scenario always demands a near-real-time support. There-
719 fore, one significant challenge is to find efficient methods that allow us to update out-
720 lier scores taking account of the newest observations and removing the oldest observa-
721 tions introduced by overlapping sliding windows, rather than recalculating scores cor-
722 responding to observations which are not affected by either new arrivals or the oldest
723 observations (that are no longer covered by the latest window). Further work will be needed
724 to investigate the efficient computation of regenerating nearest neighbours in a data stream-
725 ing context.

726 Notation

727 ***FP*** False Positives (i.e. when a typical observation is misclassified as an outlier)

728 ***FN*** False Negatives (i.e. when an actual outlier is misclassified as a typical observa-
729 tion)

730 ***TP*** True Positives (i.e. when an actual outlier is correctly classified)

731 ***TN*** True Negatives (i.e. when an observation is correctly classified as a typical point)

732 Acknowledgments

733 Funding for this project was provided by the Queensland Department of Environ-
734 ment and Science (DES) and the ARC Centre of Excellence for Mathematical and Sta-
735 tistical Frontiers (ACEMS). The authors would like to acknowledge the Queensland De-
736 partment of Environment and Science; in particular, the Great Barrier Reef Catchment
737 Loads Monitoring Program for the data, and the staff from Water Quality and Inves-
738 tigation for their input. We thank Ryan S. Turner and Erin E. Peterson for several valu-
739 able discussions regarding project requirements and water quality characteristics. Fur-

740 ther, this research was supported in part by the Monash eResearch Centre and eSolutions-
 741 Research Support Services through the use of the MonARCH (Monash Advanced Re-
 742 search Computing Hybrid) HPC Cluster. We would also like to thank David Hill and
 743 other anonymous reviewers for their valuable comments and suggestions. The datasets
 744 used for this article are available in the open source R package *oddwater* (Talagala &
 745 Hyndman, 2019b).

746 References

- 747 Angiulli, F., & Pizzuti, C. (2002). Fast outlier detection in high dimensional spaces.
 748 In *European conference on principles of data mining and knowledge discovery*
 749 (pp. 15–27).
- 750 Archer, C., Baptista, A., & Leen, T. K. (2003). Fault detection for salinity sensors
 751 in the columbia estuary. *Water Resources Research*, *39*(3).
- 752 Ben-Gal, I. (2005). Outlier detection. In *Data mining and knowledge discovery hand-*
 753 *book* (pp. 131–146). Springer.
- 754 Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). Lof: identifying
 755 density-based local outliers. In *Acm sigmod record* (Vol. 29, pp. 93–104).
- 756 Burrige, P., & Taylor, A. M. R. (2006). Additive outlier detection via extreme-
 757 value theory. *Journal of Time Series Analysis*, *27*(5), 685–701.
- 758 Campos, G. O., Zimek, A., Sander, J., Campello, R. J., Micenková, B., Schubert,
 759 E., ... Houle, M. E. (2016). On the evaluation of unsupervised outlier detec-
 760 tion: measures, datasets, and an empirical study. *Data Mining and Knowledge*
 761 *Discovery*, *30*(4), 891–927.
- 762 Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey.
 763 *ACM computing surveys (CSUR)*, *41*(3), 15.
- 764 Dang, T. N., & Wilkinson, L. (2014). Transforming scagnostics to reveal hidden
 765 features. *IEEE transactions on visualization and computer graphics*, *20*(12),
 766 1624–1632.
- 767 Embrechts, P., Klüppelberg, C., & Mikosch, T. (2013). *Modelling extremal events:*
 768 *for insurance and finance*. Springer Berlin Heidelberg. Retrieved from
 769 <https://books.google.com.au/books?id=BX0I2pICfJUC>
- 770 Gao, J., Hu, W., Zhang, Z. M., Zhang, X., & Wu, O. (2011). Rkof: robust kernel-
 771 based local outlier detection. In *Pacific-asia conference on knowledge discovery*
 772 *and data mining* (pp. 270–283).
- 773 Glasgow, H. B., Burkholder, J. M., Reed, R. E., Lewitus, A. J., & Kleinman, J. E.
 774 (2004). Real-time remote monitoring of water quality: a review of current ap-
 775 plications, and advancements in sensor, telemetry, and computing technologies.
 776 *Journal of Experimental Marine Biology and Ecology*, *300*(1-2), 409–448.
- 777 Goldstein, M., & Uchida, S. (2016). A comparative evaluation of unsuper-
 778 vised anomaly detection algorithms for multivariate data. *PloS one*, *11*(4),
 779 e0152173.
- 780 Hill, D. J., & Minsker, B. S. (2006). Automated fault detection for in-situ environ-
 781 mental sensors. In *Proceedings of the 7th international conference on hydroin-*
 782 *formatics*.
- 783 Hill, D. J., Minsker, B. S., & Amir, E. (2009). Real-time bayesian anomaly detection
 784 in streaming environmental data. *Water Resources Research*, *45*(4).
- 785 Hossin, M., & Sulaiman, M. (2015). A review on evaluation metrics for data classi-
 786 fication evaluations. *International Journal of Data Mining & Knowledge Man-*
 787 *agement Process*, *5*(2), 1.
- 788 Jin, W., Tung, A. K., Han, J., & Wang, W. (2006). Ranking outliers using symmet-
 789 ric neighborhood relationship. In *Pacific-asia conference on knowledge discov-*
 790 *ery and data mining* (pp. 577–593).
- 791 Koch, M. W., & McKenna, S. A. (2010). Distributed sensor fusion in water qual-
 792 ity event detection. *Journal of Water Resources Planning and Management*,

- 137(1), 10–19.
- 793 Kotamäki, N., Thessler, S., Koskiahho, J., Hannukkala, A. O., Huitu, H., Huttula, T.,
794 ... Järvenpää, M. (2009). Wireless in-situ sensor network for agriculture and
795 water monitoring on a river basin scale in southern finland: Evaluation from a
796 data users perspective. *Sensors*, 9(4), 2862–2883.
- 797 Kriegel, H.-P., Kröger, P., & Zimek, A. (2010). Outlier detection techniques. *Tuto-*
798 *rial at KDD*, 10.
- 799 Leigh, C., Alsibai, O., Hyndman, R. J., Kandanaarachchi, S., King, O. C., McGree,
800 J. M., ... others (2019). A framework for automated anomaly detection in
801 high frequency water-quality data from in situ sensors. *Science of The Total*
802 *Environment*, 664, 885–898.
- 803 Madsen, J. H. (2018). Ddoutlier: Distance and density-based outlier detection
804 [Computer software manual]. Retrieved from [https://CRAN.R-project.org/](https://CRAN.R-project.org/package=DDoutlier)
805 [package=DDoutlier](https://CRAN.R-project.org/package=DDoutlier) (R package version 0.1.0)
- 806 McInnes, K., Abbs, D., Bhend, J., Chiew, F., Church, J., Ekstrm, M., ... Whetton,
807 P. (2015). *Wet tropics cluster report: Climate change in australia projections*
808 *for australia's nrm regions*. CSIRO.
- 809 McKenna, S. A., Hart, D., Klise, K., Cruz, V., & Wilson, M. (2007). Event de-
810 tection from water quality time series. In *World environmental and water re-*
811 *sources congress 2007: Restoring our natural habitat* (pp. 1–12).
- 812 Mersmann, O. (2018). microbenchmark: Accurate timing functions [Com-
813 puter software manual]. Retrieved from [https://CRAN.R-project.org/](https://CRAN.R-project.org/package=microbenchmark)
814 [package=microbenchmark](https://CRAN.R-project.org/package=microbenchmark) (R package version 1.4-4)
- 815 Mitchell, C., Brodie, J., & White, I. (2005). Sediments, nutrients and pesticide
816 residues in event flow conditions in streams of the mackay whitsunday region,
817 australia. *Marine Pollution Bulletin*, 51(1-4), 23–36.
- 818 Moatar, F., Fessant, F., & Poirel, A. (1999). ph modelling by neural networks. appli-
819 cation of control and validation data series in the middle loire river. *Ecological*
820 *Modelling*, 120(2-3), 141–156.
- 821 Moatar, F., Miquel, J., & Poirel, A. (2001). A quality-control method for physical
822 and chemical monitoring data. application to dissolved oxygen levels in the
823 river loire (france). *Journal of Hydrology*, 252(1-4), 25–36.
- 824 Panguluri, S., Meiners, G., Hall, J., & Szabo, J. (2009). Distribution system water
825 quality monitoring: Sensor technology evaluation methodology and results. *US*
826 *Environ. Protection Agency, Washington, DC, USA, Tech. Rep. EPA/600/R-*
827 *09/076, 2772*.
- 828 R Core Team. (2018). R: A language and environment for statistical computing
829 [Computer software manual]. Vienna, Austria. Retrieved from [https://www.R-](https://www.R-project.org/)
830 [project.org/](https://www.R-project.org/)
- 831 Raciti, M., Cucurull, J., & Nadjm-Tehrani, S. (2012). Anomaly detection in wa-
832 ter management systems. In *Critical infrastructure protection* (pp. 98–119).
833 Springer.
- 834 Ranawana, R., & Palade, V. (2006). Optimized precision-a new measure for classi-
835 fier performance evaluation. In *Evolutionary computation, 2006. cec 2006. iee*
836 *congress on* (pp. 2254–2261).
- 837 Rangeti, I., Dzwaïro, B., Barratt, G. J., & Otieno, F. A. (2015). Validity and errors
838 in water quality dataa review. In *Research and practices in water quality*. In-
839 Tech.
- 840 Schwarz, K. T. (2008). *Wind dispersion of carbon dioxide leaking from underground*
841 *sequestration, and outlier detection in eddy covariance data using extreme*
842 *value theory*. University of California, Berkeley.
- 843 Shahid, N., Naqvi, I. H., & Qaisar, S. B. (2015). Characteristics and classification of
844 outlier detection techniques for wireless sensor networks in harsh environments:
845 a survey. *Artificial Intelligence Review*, 43(2), 193–228.
- 846 Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance mea-
847

- 848 sures for classification tasks. *Information Processing & Management*, 45(4),
849 427–437.
- 850 Storey, M. V., Van der Gaag, B., & Burns, B. P. (2011). Advances in on-line drink-
851 ing water quality monitoring and early warning systems. *Water research*,
852 45(2), 741–747.
- 853 Talagala, P. D., & Hyndman, R. J. (2019a). A feature-based procedure for detecting
854 technical outliers in water-quality data: R package oddwater v.0.7.0 [Computer
855 software manual]. Zenodo. doi: 10.5281/zenodo.3378211
- 856 Talagala, P. D., & Hyndman, R. J. (2019b). oddwater: Outlier detection in data
857 from water-quality sensors [Computer software manual]. Retrieved from
858 <https://github.com/pridiltal/oddwater> (R package)
- 859 Talagala, P. D., Hyndman, R. J., & Smith-Miles, K. (2019). Anomaly detection in
860 high dimensional data. *arXiv preprint arXiv:1908.04000*.
- 861 Talagala, P. D., Hyndman, R. J., Smith-Miles, K., Kandanaarachchi, S., & Muñoz,
862 M. A. (2019). Anomaly detection in streaming nonstationary temporal data.
863 *Journal of Computational and Graphical Statistics* (Accepted), 1–28.
- 864 Tang, J., Chen, Z., Fu, A. W.-C., & Cheung, D. W. (2002). Enhancing effective-
865 ness of outlier detections for low density patterns. In *Pacific-asia conference on*
866 *knowledge discovery and data mining* (pp. 535–548).
- 867 Thottan, M., & Ji, C. (2003). Anomaly detection in ip networks. *IEEE Transactions*
868 *on signal processing*, 51(8), 2191–2204.
- 869 Tutmez, B., Hatipoglu, Z., & Kaymak, U. (2006). Modelling electrical conductivity
870 of groundwater using an adaptive neuro-fuzzy inference system. *Computers &*
871 *geosciences*, 32(4), 421–433.
- 872 Unwin, A. (2019). Multivariate outliers and the o3 plot. *Journal of Computational*
873 *and Graphical Statistics*, 1–11.
- 874 Weissman, I. (1978). Estimation of parameters and large quantiles based on the k
875 largest observations. *Journal of the American Statistical Association*, 73(364),
876 812–815.
- 877 Wilkinson, L. (2018). Visualizing big data outliers through distributed aggregation.
878 *IEEE transactions on visualization and computer graphics*, 24(1), 256–266.
- 879 Yu, J. (2012). A bayesian inference based two-stage support vector regression frame-
880 work for soft sensor development in batch bioprocesses. *Computers & Chemical*
881 *Engineering*, 41, 134–144.
- 882 Zhang, K., Hutter, M., & Jin, H. (2009). A new local distance-based outlier de-
883 tection approach for scattered real-world data. In *Pacific-asia conference on*
884 *knowledge discovery and data mining* (pp. 813–822).

Figure 1.

Author Manuscript

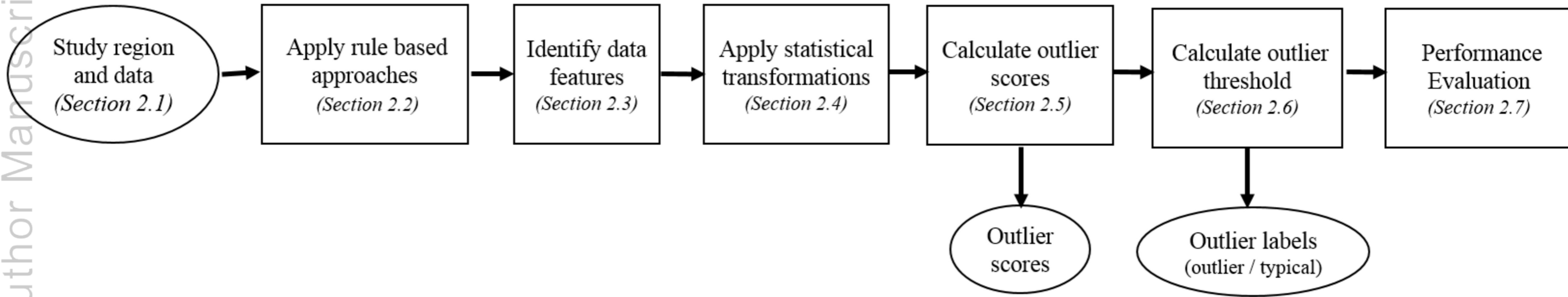


Figure 2.

Author Manuscript

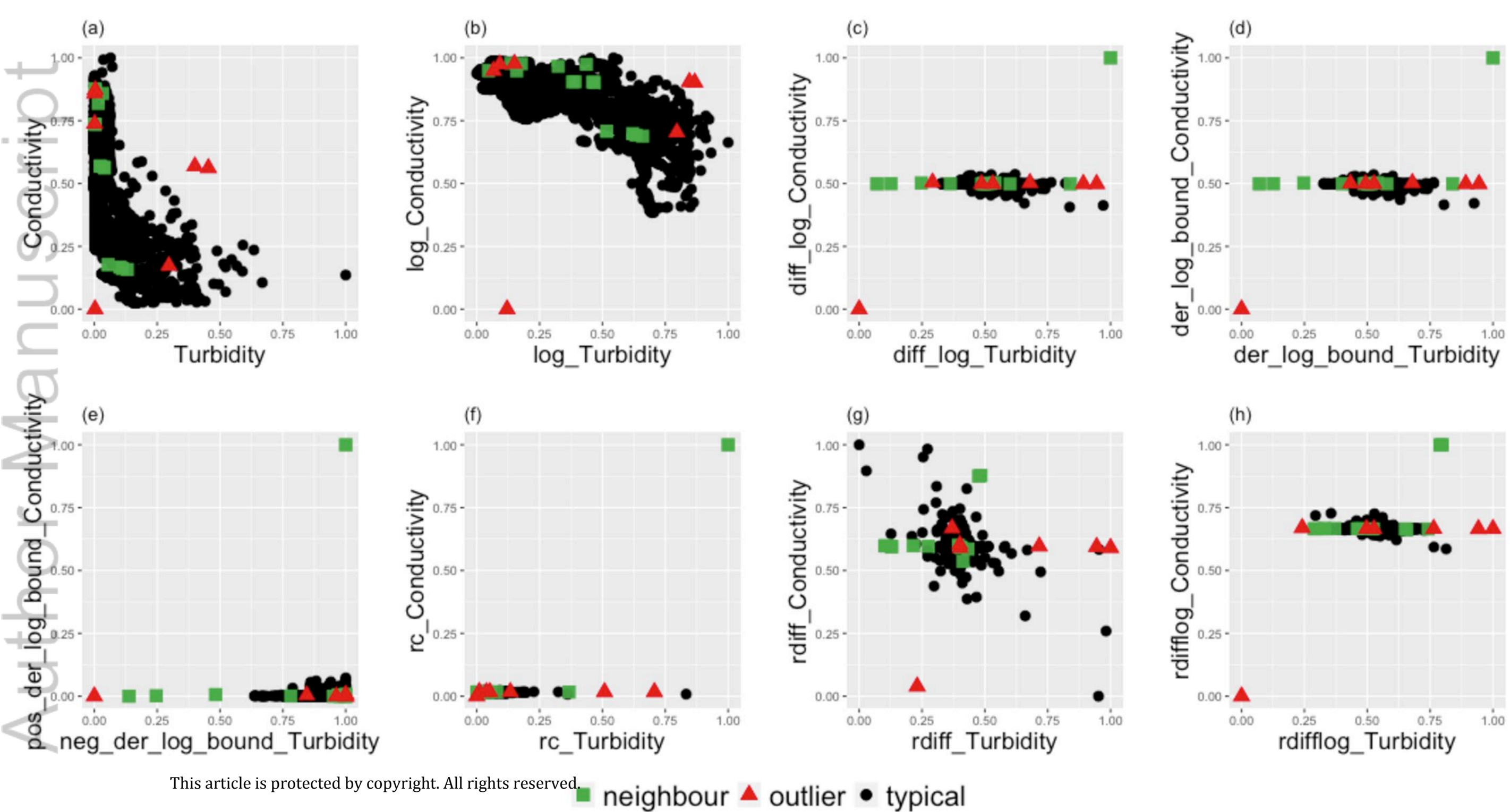


Figure 3.

Author Manuscript

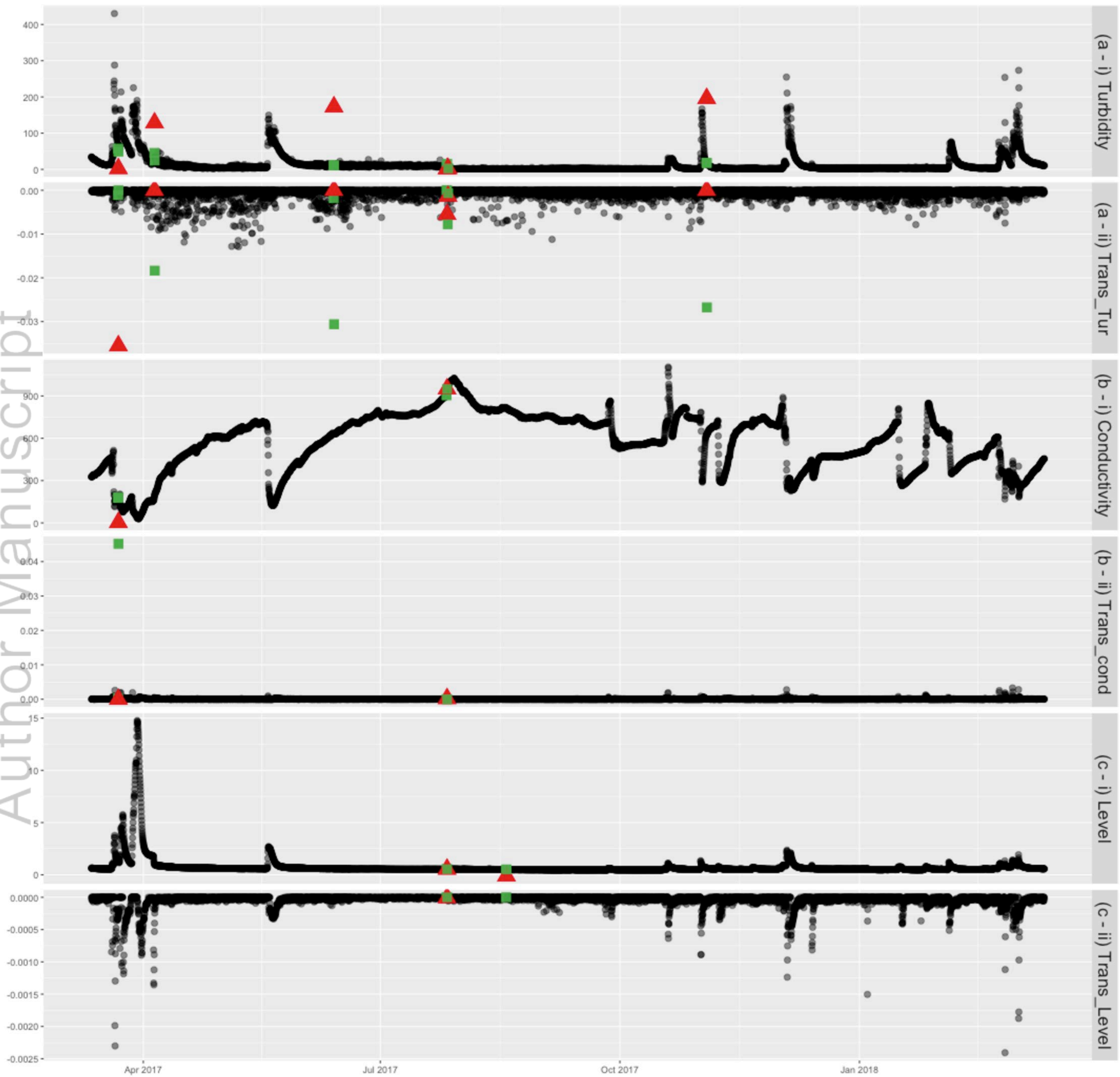


Figure 4.

Author Manuscript

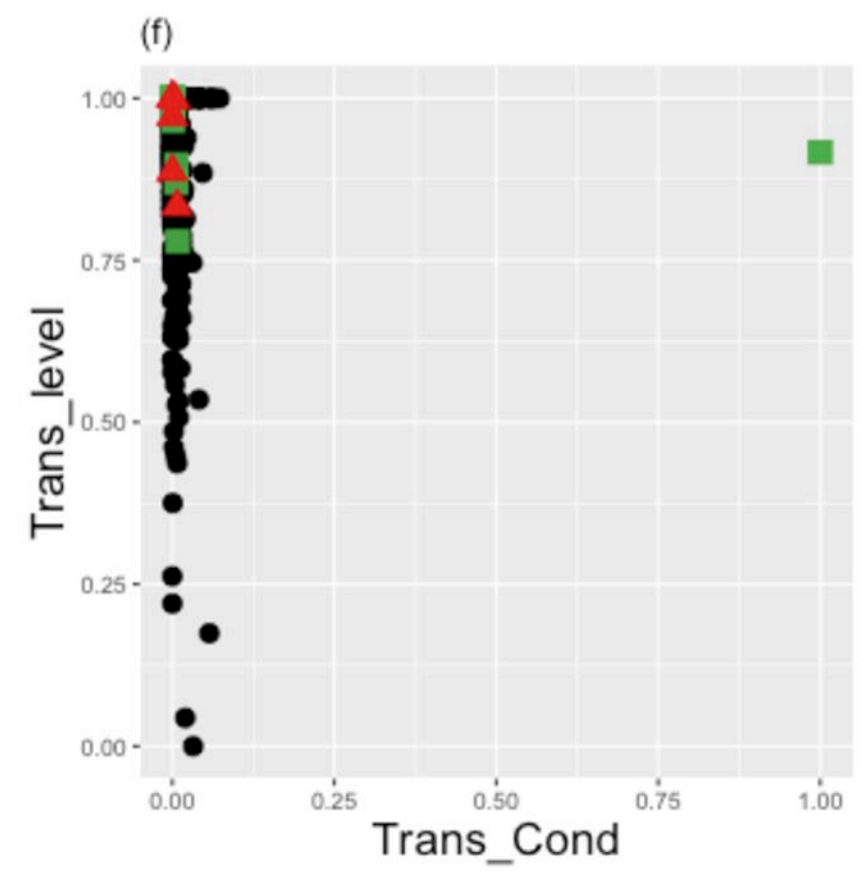
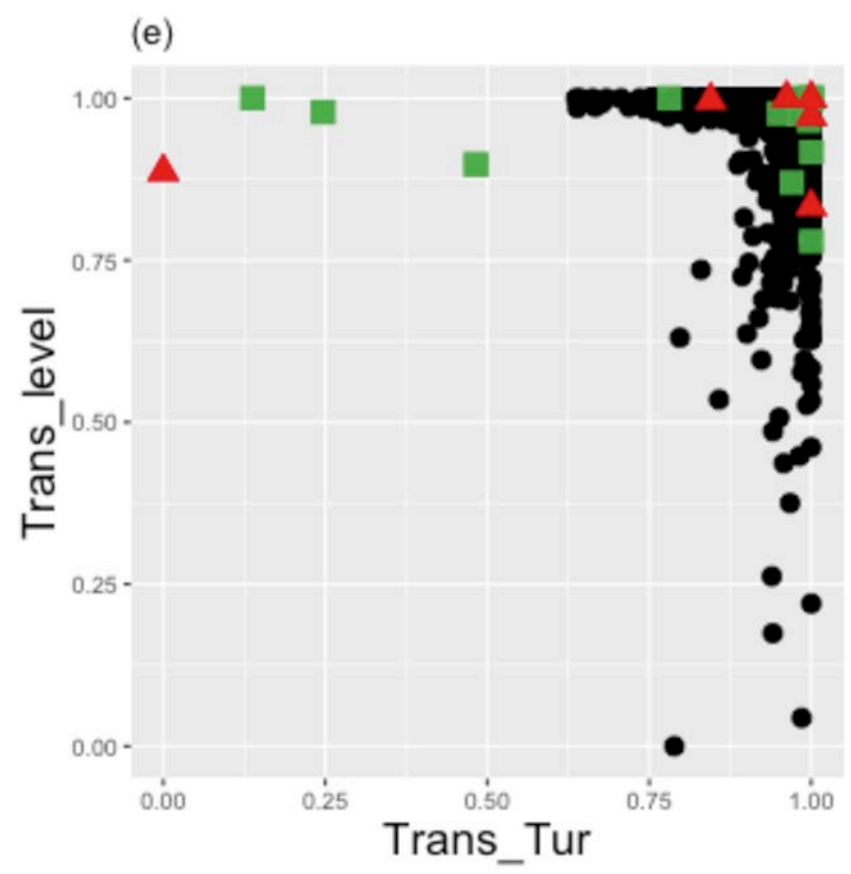
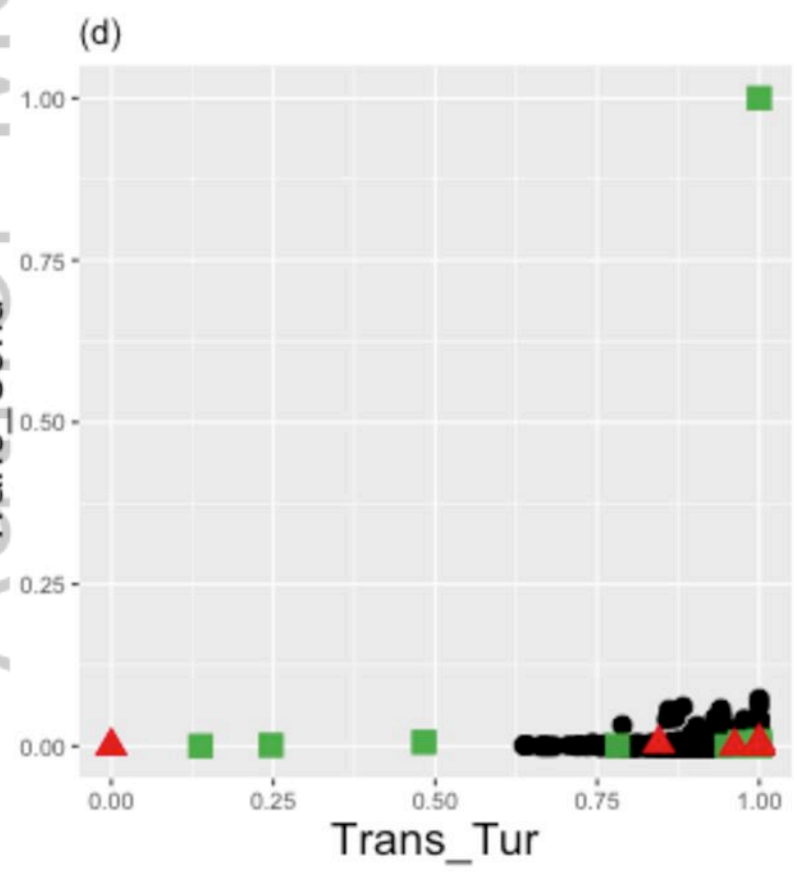
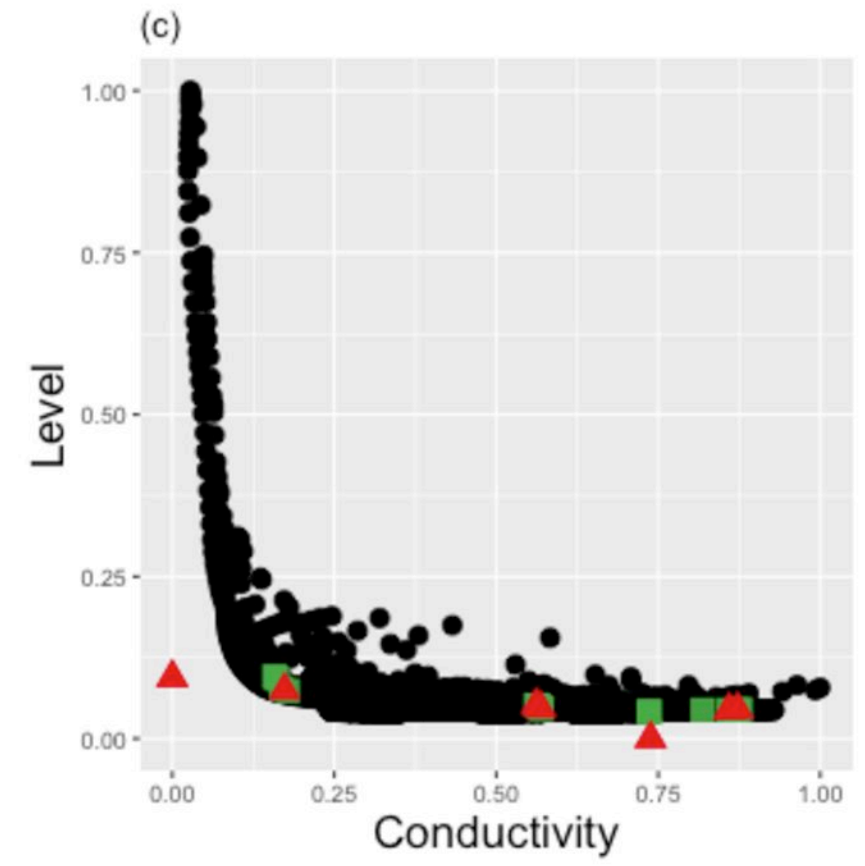
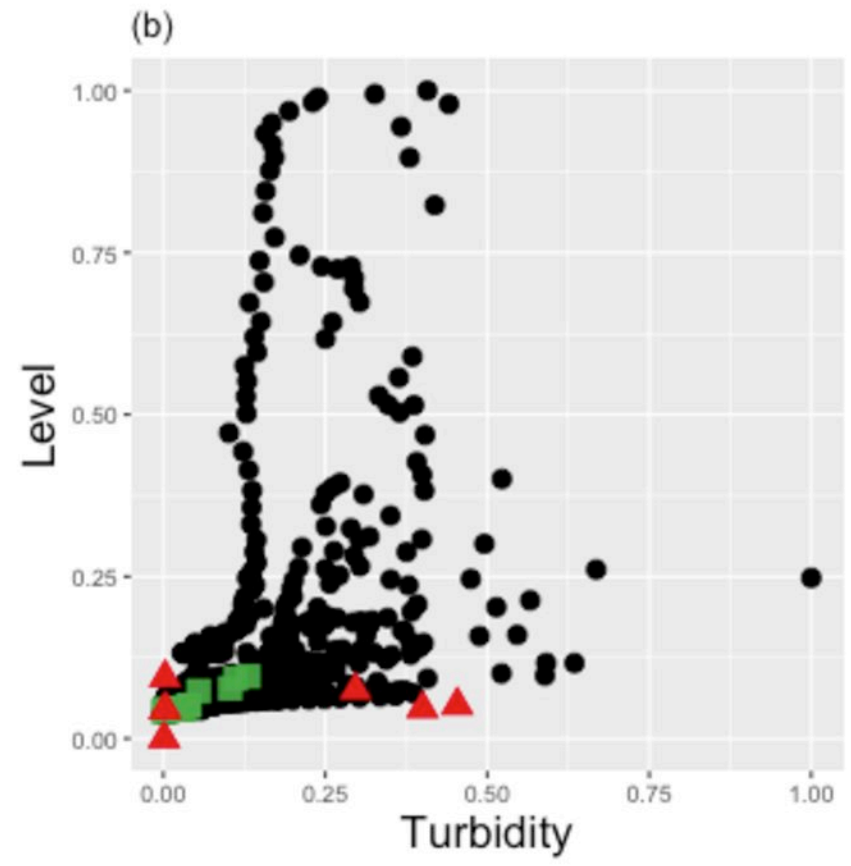
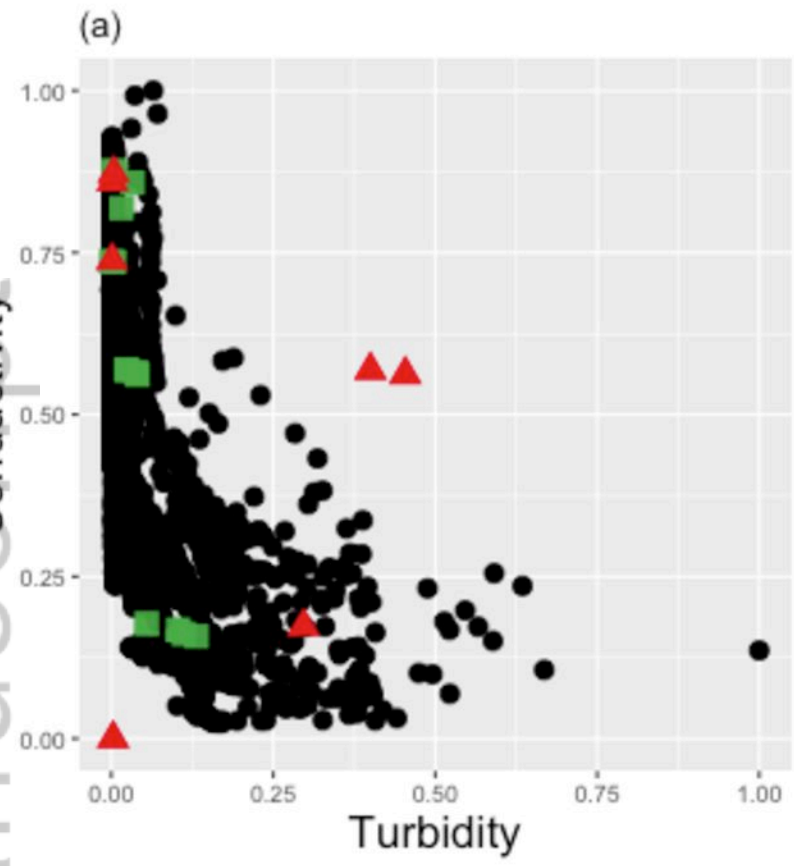


Figure 5.

Author Manuscript

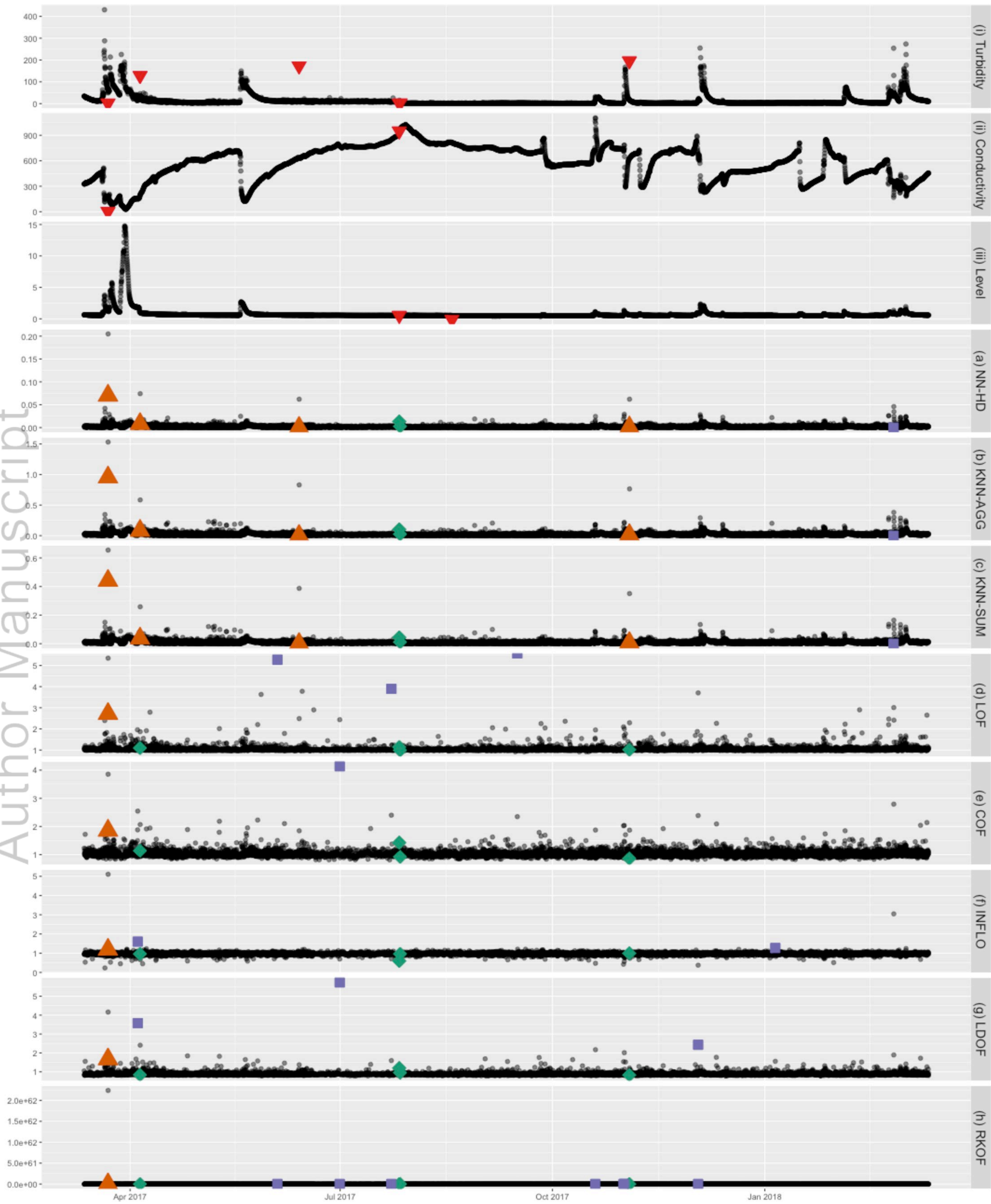


Figure 6.

Author Manuscript

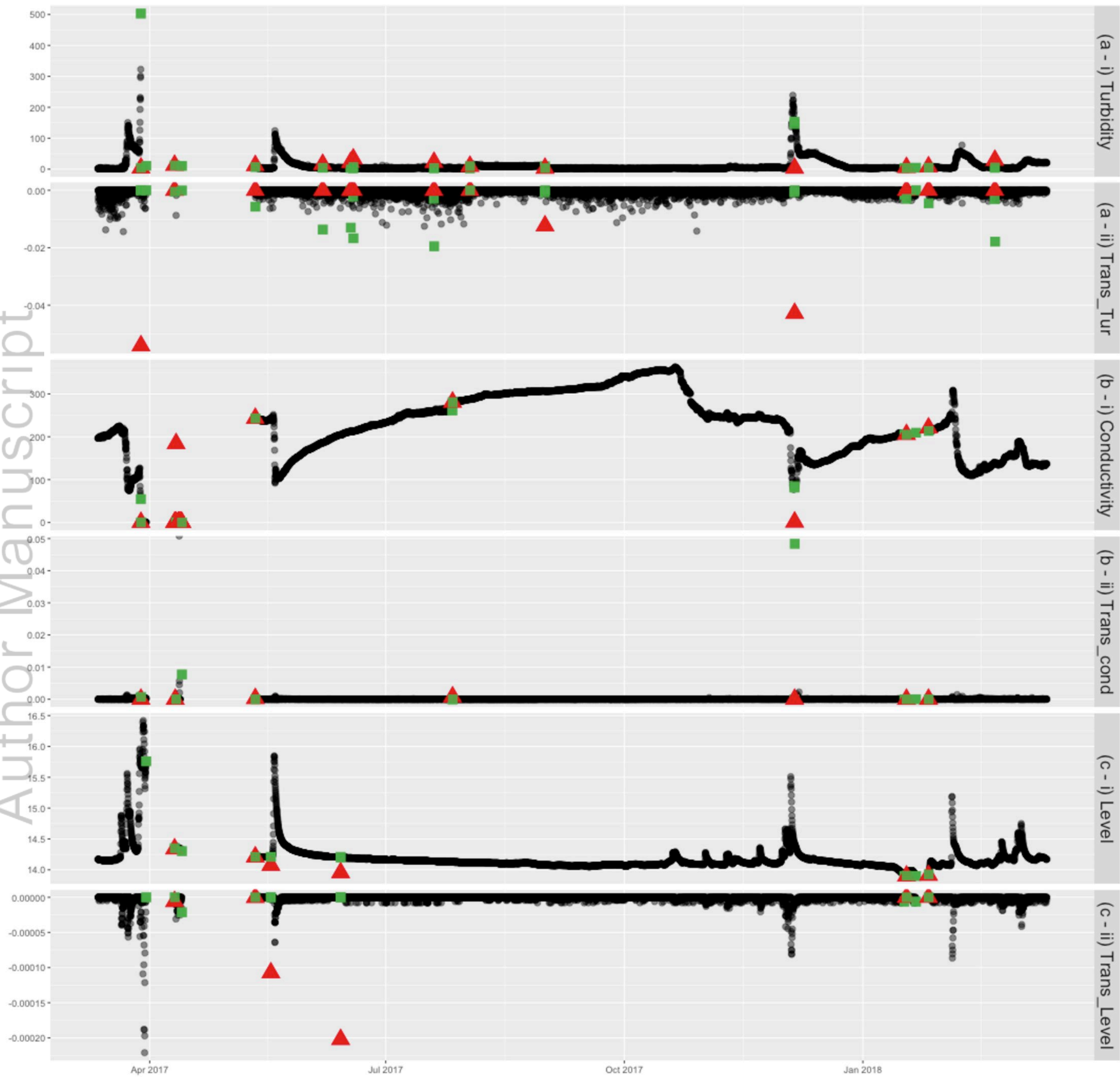


Figure 7.

Author Manuscript

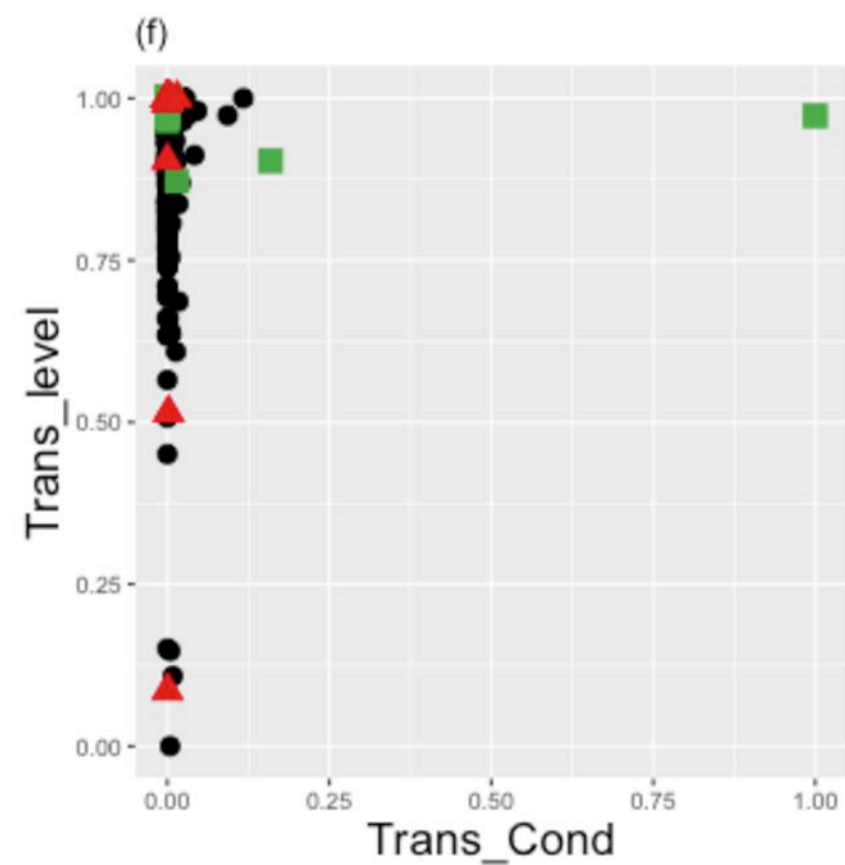
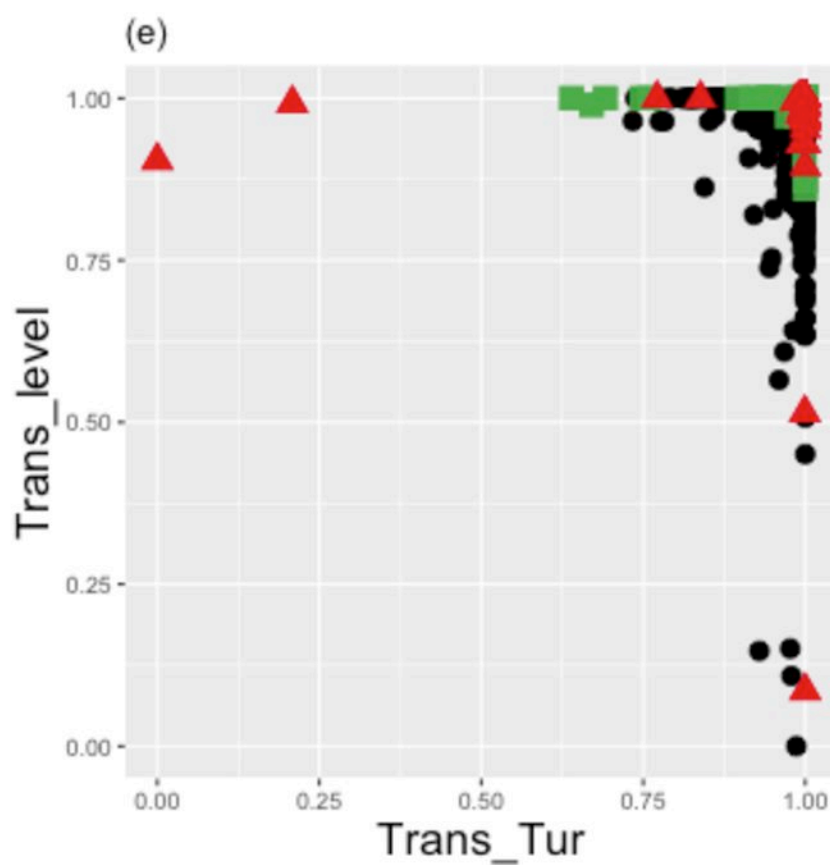
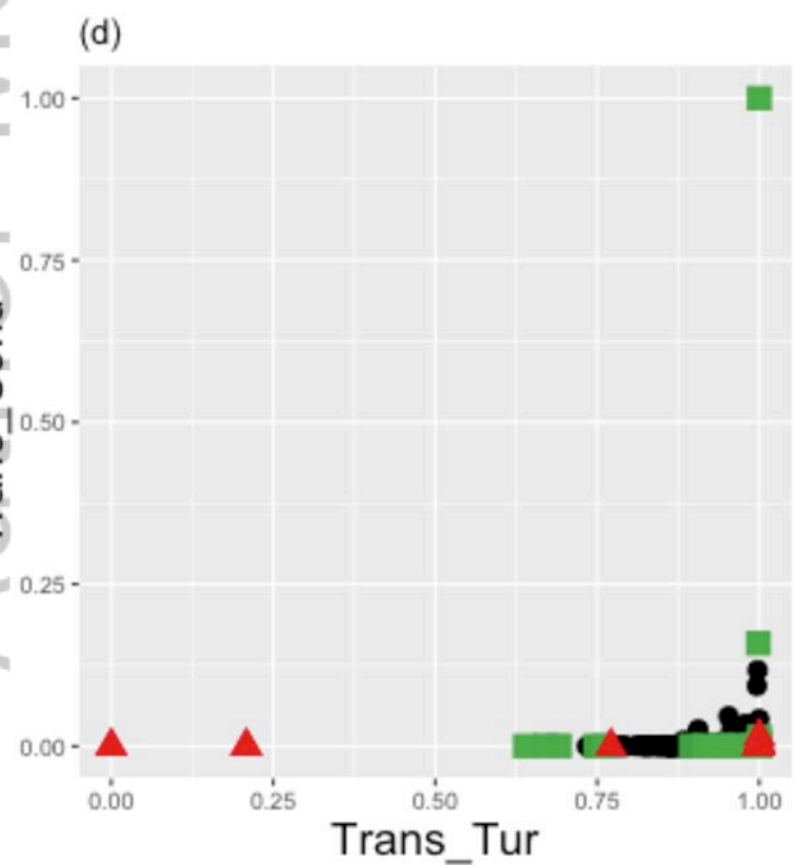
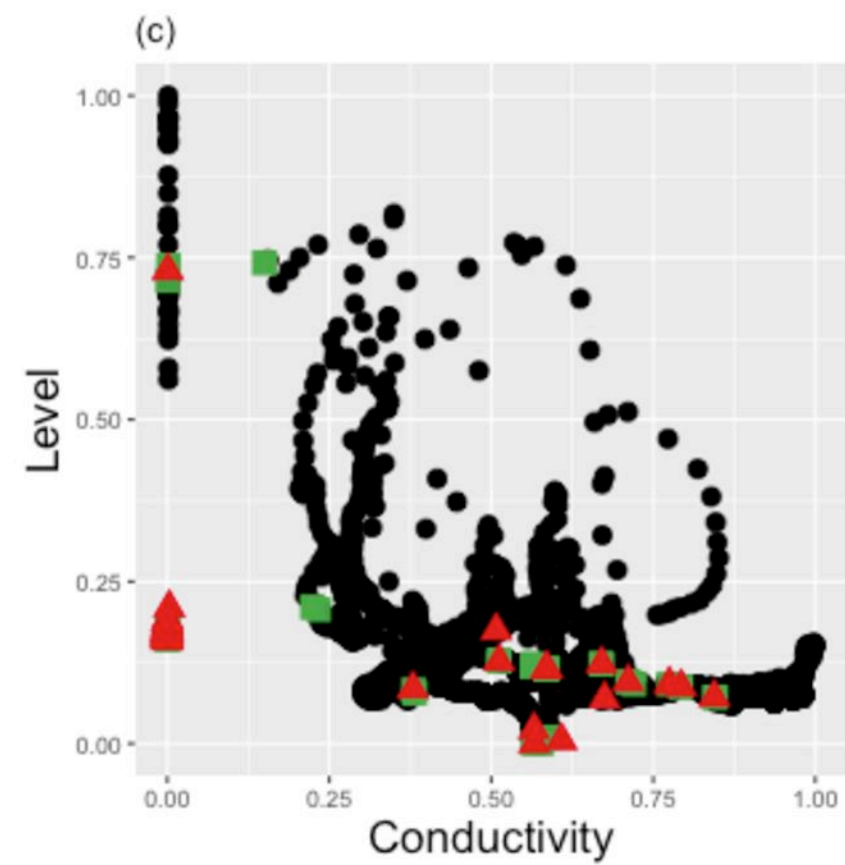
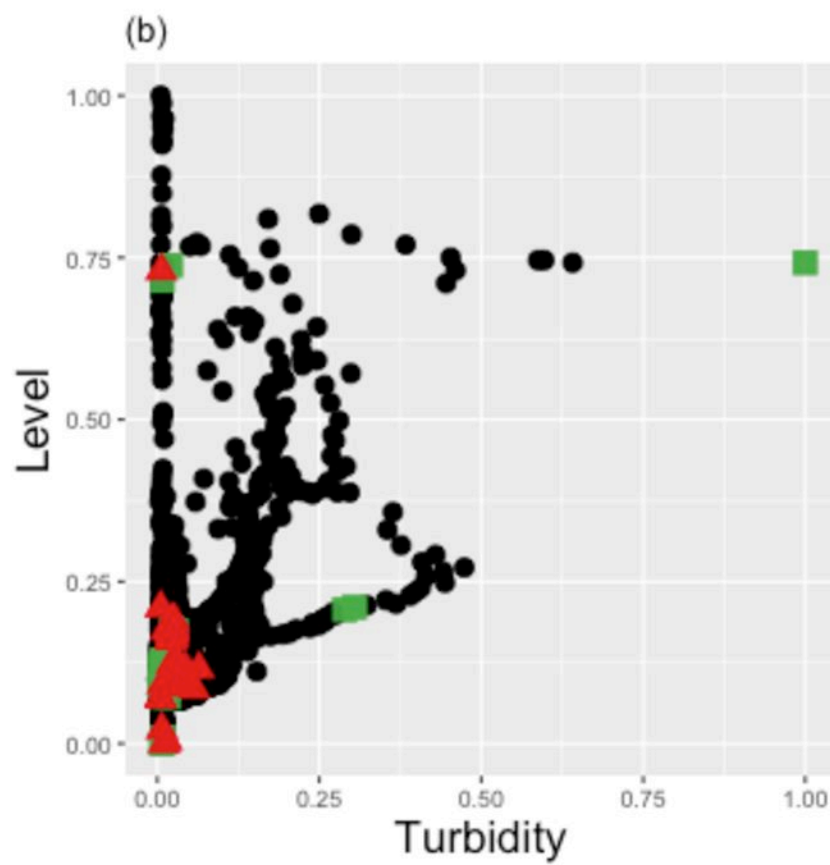
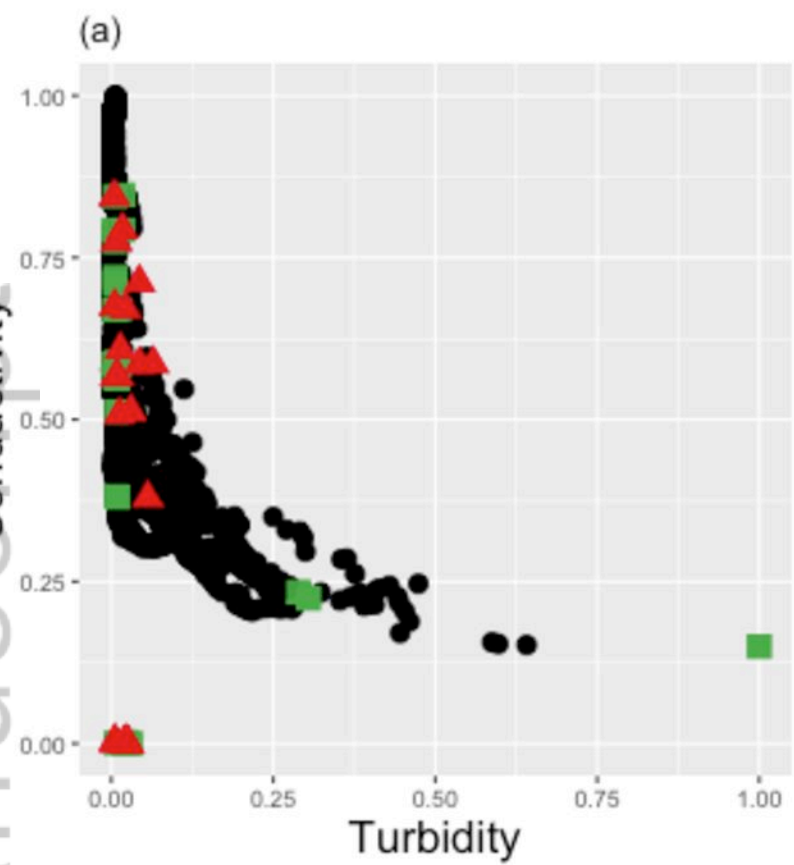
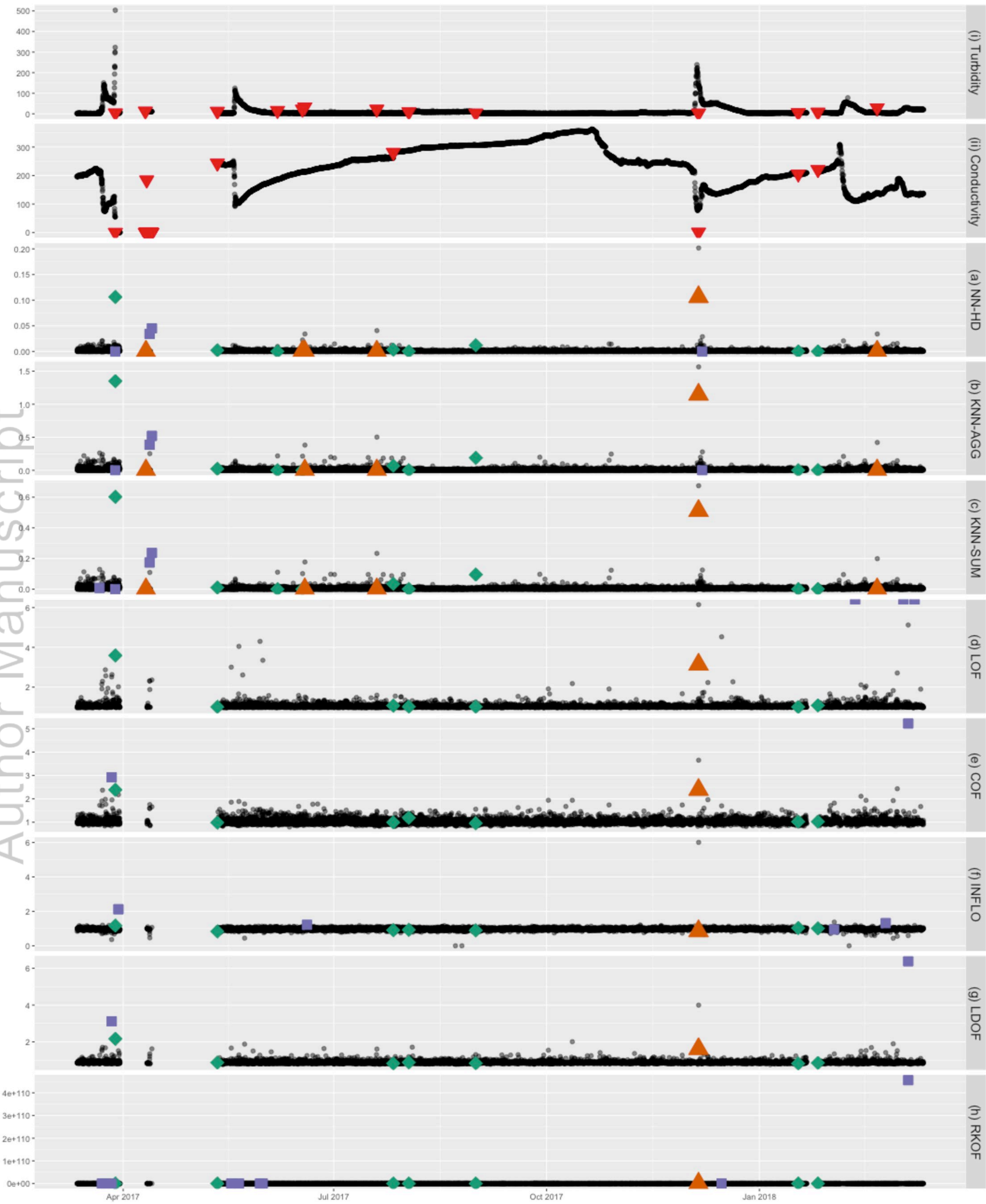
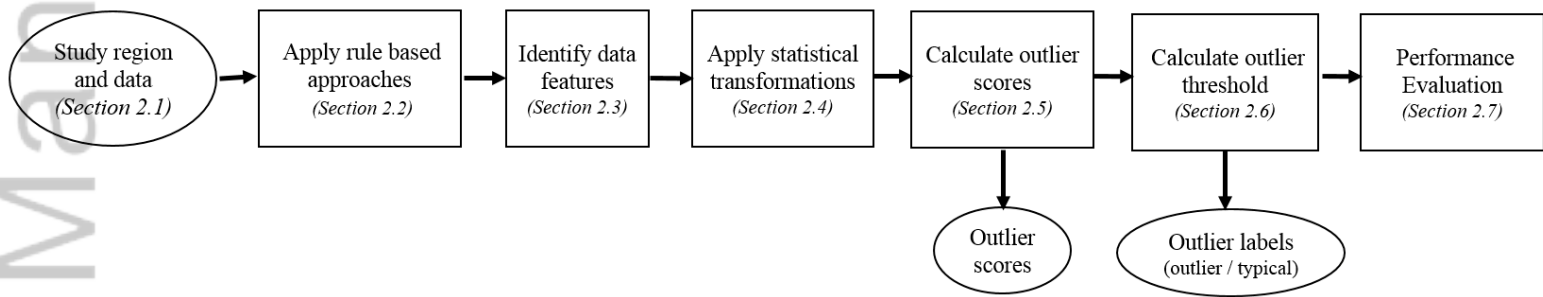


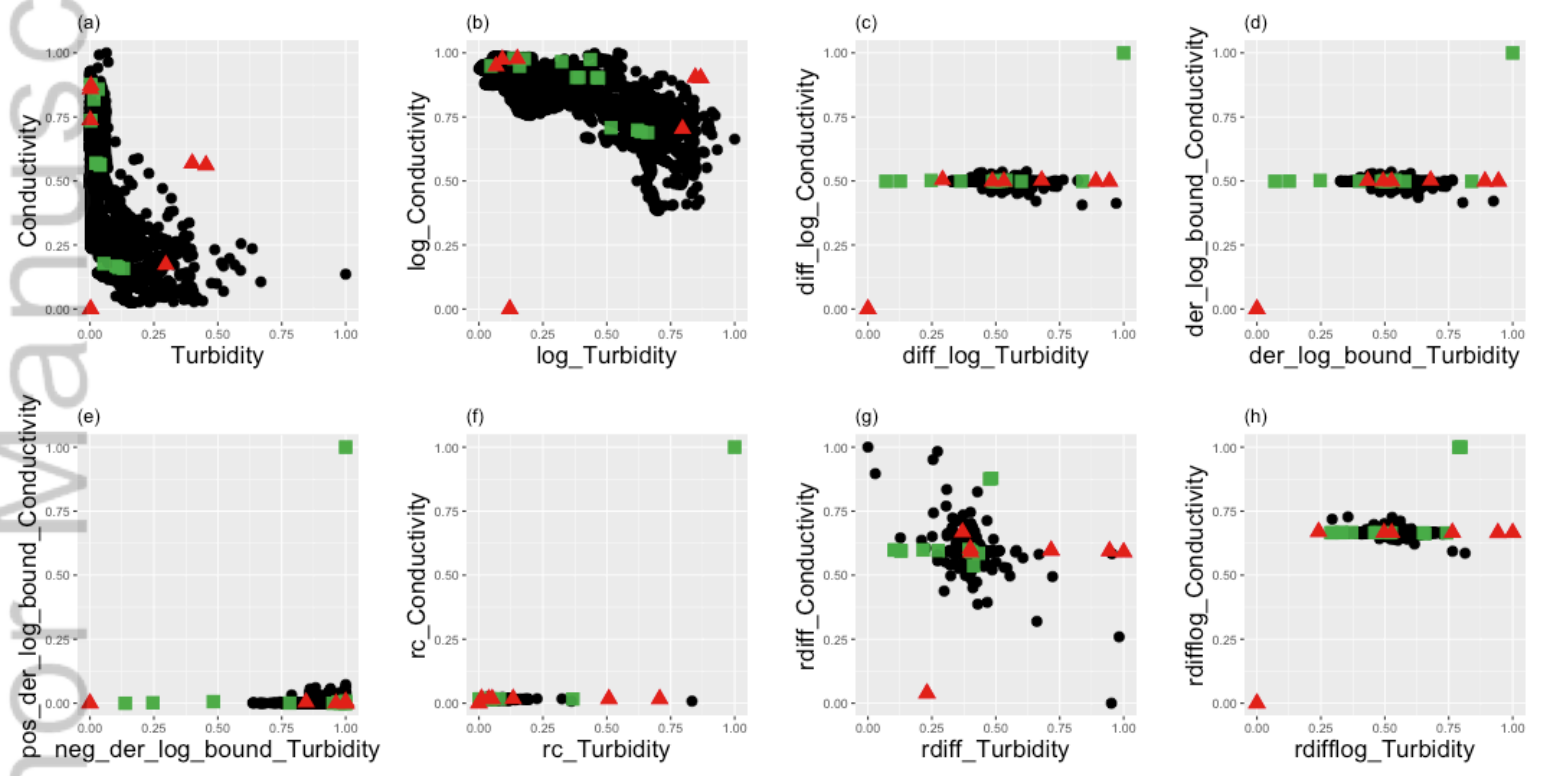
Figure 8.

Author Manuscript

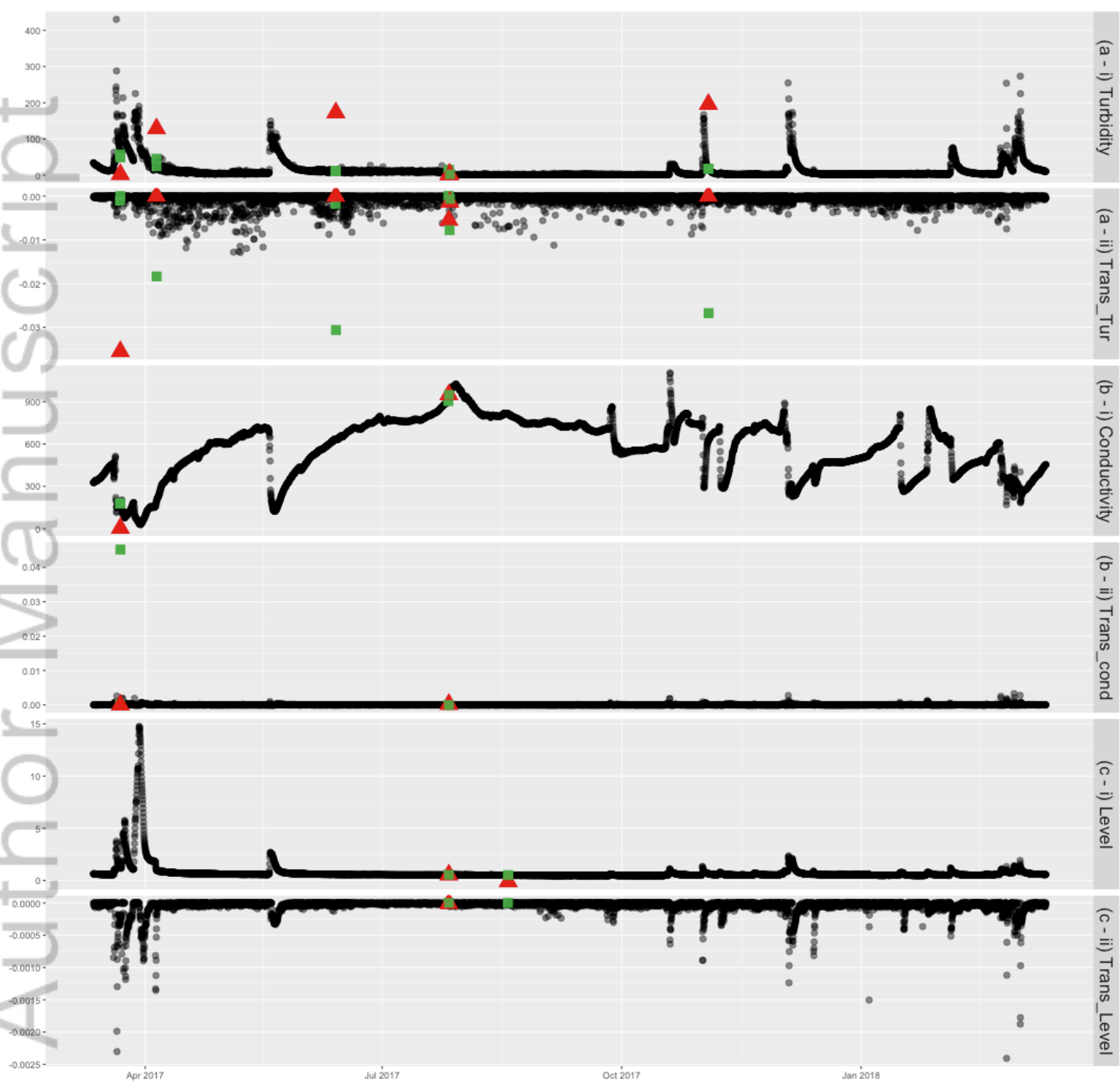




2019WR024906-f01-z-.png

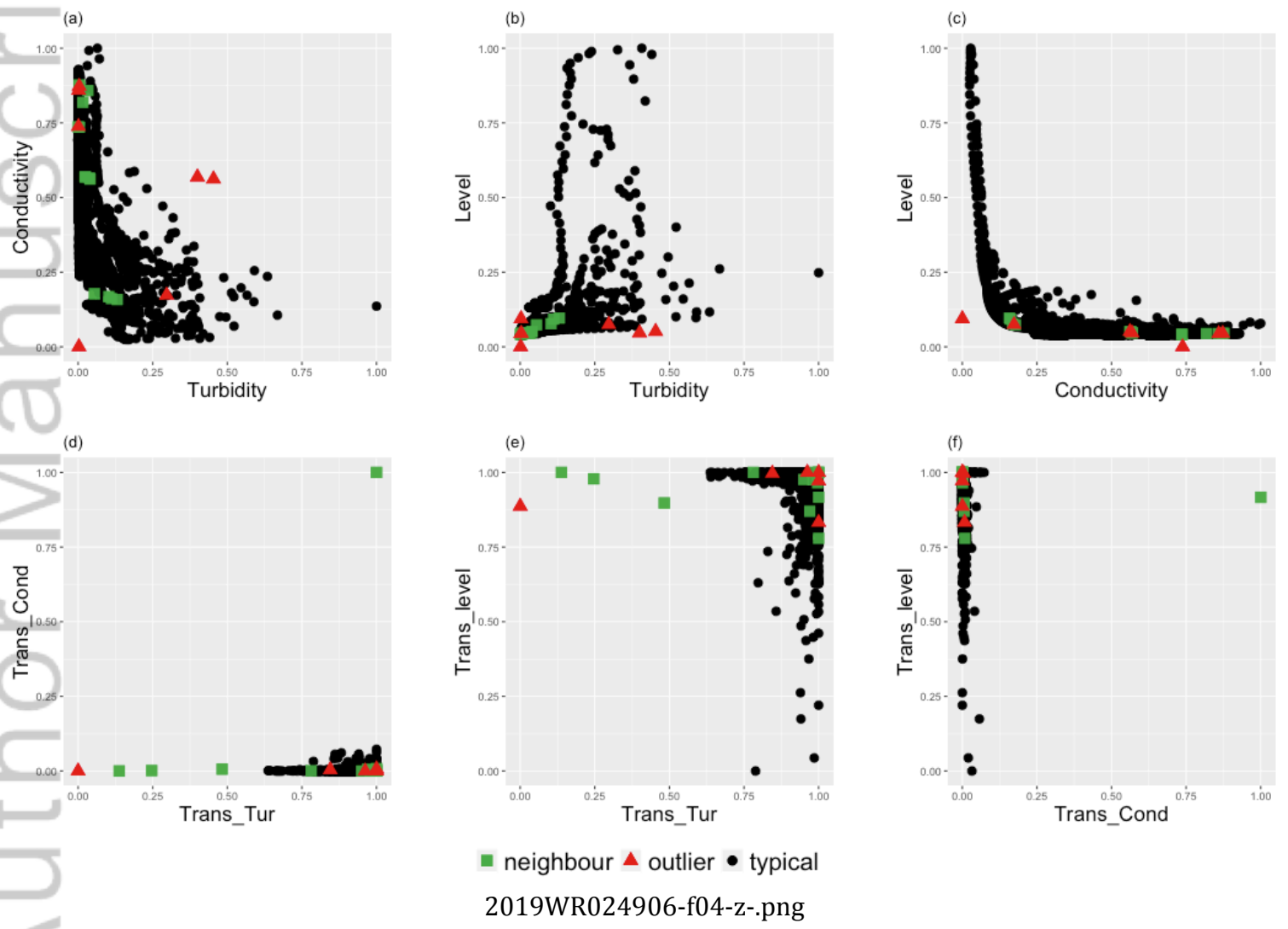


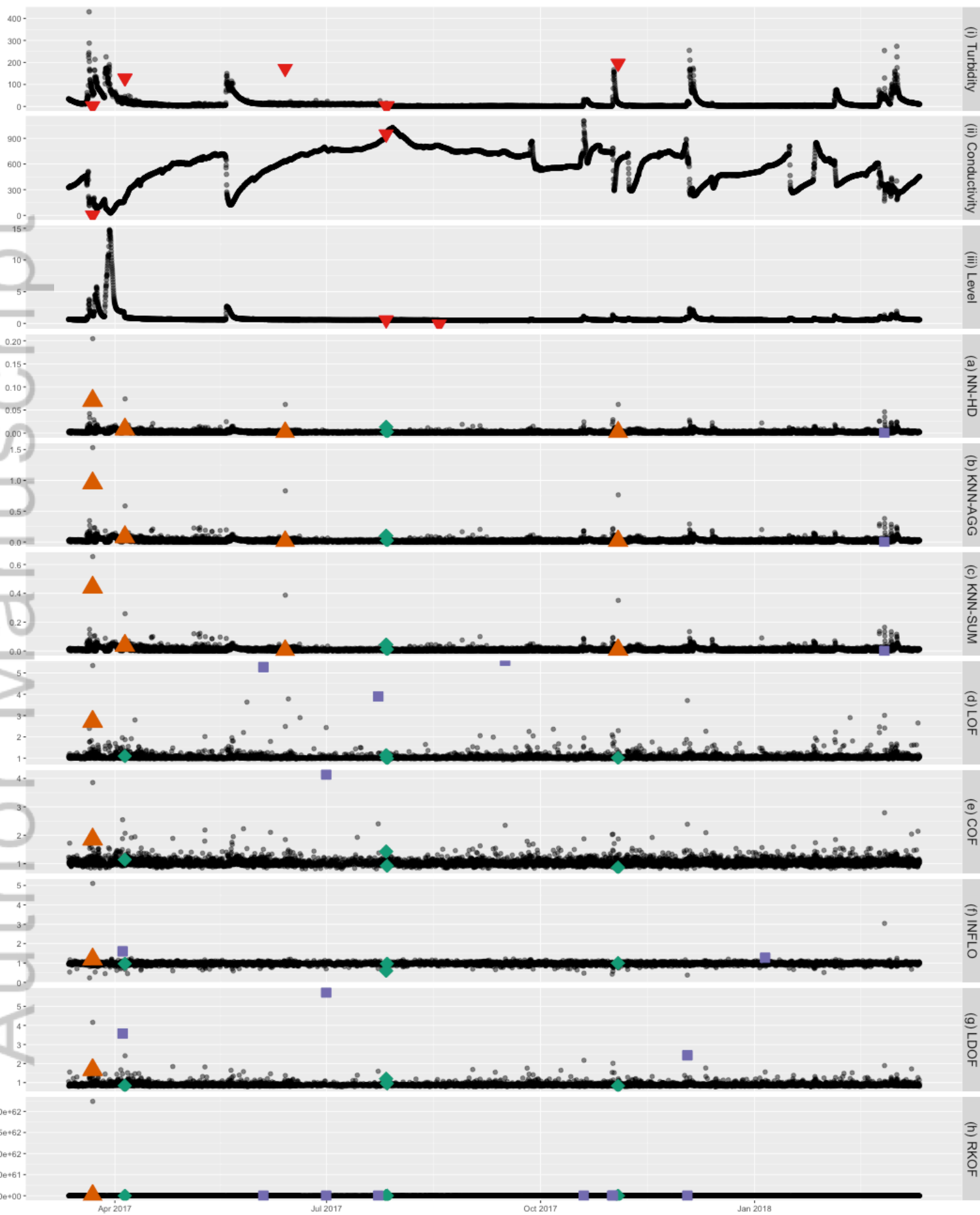
■ neighbour ▲ outlier ● typical
2019WR024906-f02-z-.png



■ neighbour ▲ outlier ● typical

2019WR024906-f03-z-.png

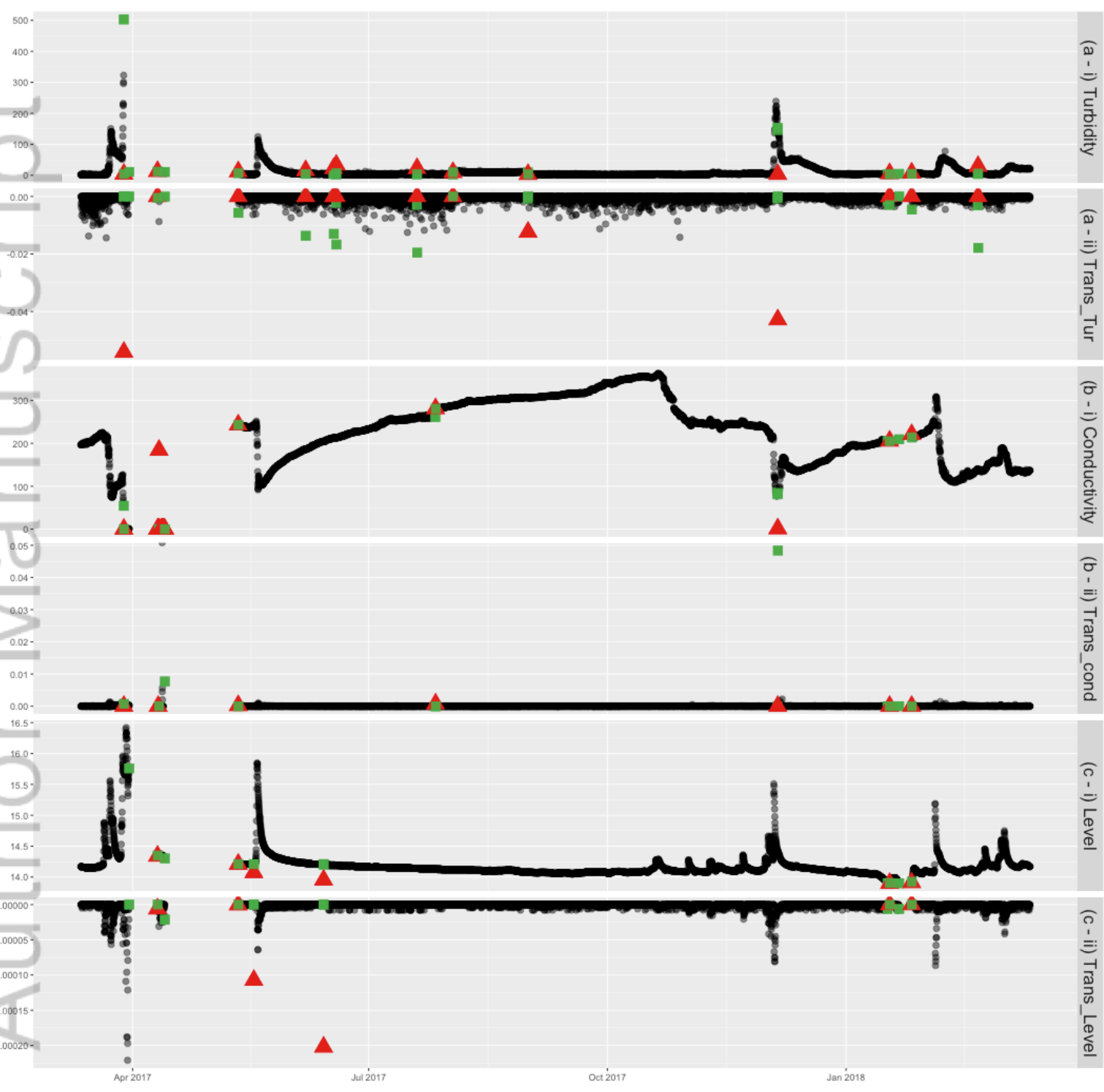




• typical • TN ◆ FN
▼ outlier ▲ TP ■ FP

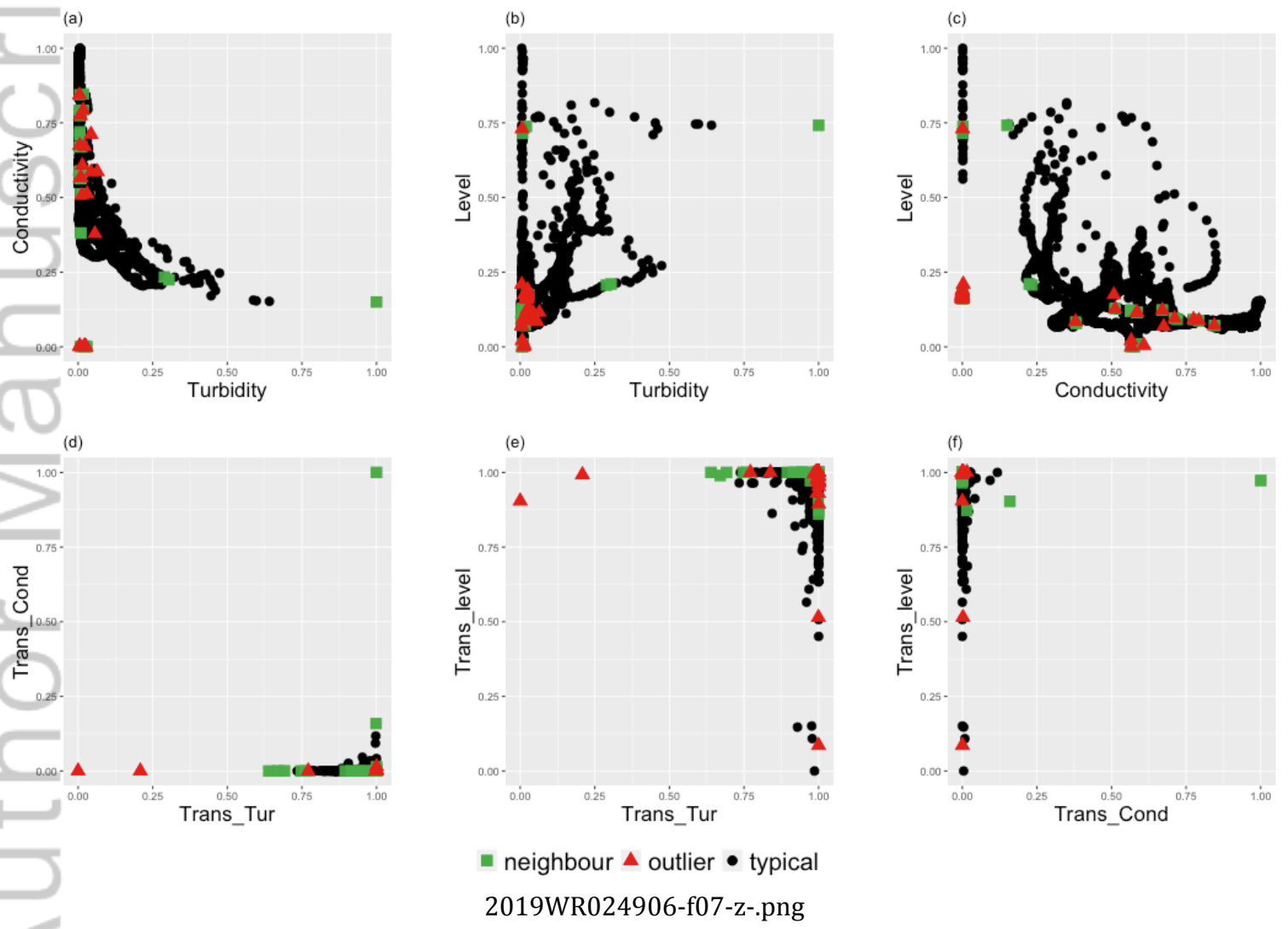
2019WR024906-f05-z-.png

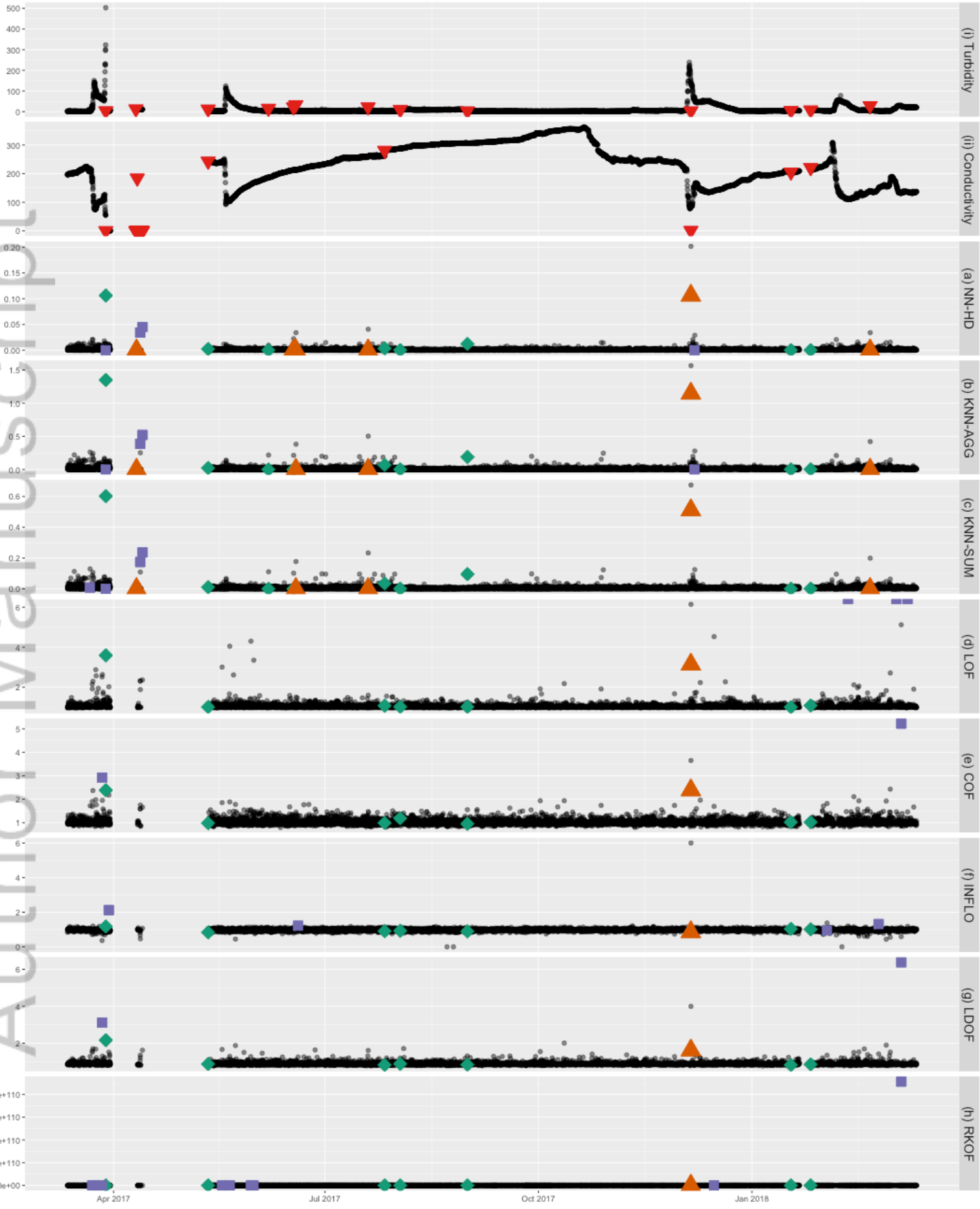
This article is protected by copyright. All rights reserved.



■ neighbour ▲ outlier ● typical

2019WR024906-f06-z-.png





• typical • TN ◆ FN
▼ outlier ▲ TP ■ FP



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Talagala, PD; Hyndman, RJ; Leigh, C; Mengersen, K; Smith-Miles, K

Title:

A Feature-Based Procedure for Detecting Technical Outliers in Water-Quality Data From In Situ Sensors

Date:

2019-11-06

Citation:

Talagala, P. D., Hyndman, R. J., Leigh, C., Mengersen, K. & Smith-Miles, K. (2019). A Feature-Based Procedure for Detecting Technical Outliers in Water-Quality Data From In Situ Sensors. WATER RESOURCES RESEARCH, 55 (11), pp.8547-8568.

<https://doi.org/10.1029/2019WR024906>.

Persistent Link:

<http://hdl.handle.net/11343/286581>

File Description:

Accepted version