East San Joaquin Water Quality Coalition
1201 L Street Modesto, CA 95354
www.esjcoalition.org

MEMO

DATE: April 12, 2019

TO: Patrick Pulupa, Executive Officer

FROM: Parry Klassen, Executive Director East San Joaquin Water Quality Coalition

SUBJECT: Outlier Methodology

The East San Joaquin Water Quality Coalition (ESJWQC) is implementing its revised Waste Discharge Requirement General Order (ESJ Order), including working with its members to assist with implementing management practices that are protective of groundwater quality. Currently, the ESJWQC's focus is on educating growers about the need to implement practices that minimize the amount of excess nitrate that could potentially leach past the root zone into groundwater aquifers. Contamination of groundwater is problematic because groundwater serves as a source of drinking water for people supplied by public water systems and domestic wells in the Central Valley.

One of the issues addressed in the ESJ Order issued by the State Water Resources Control Board (See Order WQ 2018-0002) is the designation of "outliers." Outliers refer to member parcels or management units where the applied amount of nitrogen (A) as compared to the amount of nitrogen removed from harvested materials or sequestered in permanent tissue (R) differs substantially as compared to parcels or management units planted to the same crop in the coalition region. Alternatively, in the future, an outlier could refer to a member's parcels or management units where the A compared to R exceeds a target value identified by the ESJWQC and the Central Valley Water Board. The State Water Resources Control Board Order defines the outlier process as follows:

"Outliers will be identified by the Third Party annually based on the INMP Summary Report data submitted for that particular year. Eventually, it is our expectation that outliers will be determined with reference to the ranges for the multi-year A/R ratio and A-R difference target values developed by the Third Party and the Central Valley Water Board…… We will not specifically define the term in Appendix A, Modified Eastern San Joaquin Agricultural General WDRs, but will direct the Third Party to propose an approach and the Central Valley Water Board Executive Officer to approve that approach, after public notice and comment, that defines a set of Members with whom the Third Party will follow up. The Third Party may choose to apply that approach annually for a period of years to determine outliers, or the Third Party may propose and seek approval of a different approach each year."[1]

---

[1] State Water Resources Control Board Order WQ 2018-0002, page 52

East San Joaquin
WATER QUALITY COALITION
1201 L Street Modesto, CA 95354
www.esjcoalition.org

To determine if an irrigated parcel or management unit is an outlier, there is a need for certain crop-specific information: nitrogen applied and nitrogen removed. This information is reported in the Irrigation and Nitrogen Management Plan (INMP) Summary Report. Initially, members are required to submit an INMP Summary Report only for parcels in high vulnerability areas.

"The nitrogen management data collected by the third-party from individual Members will be aggregated by the township where the enrolled parcel is located and will not be associated with the Member or their enrolled parcel. For example, the third-party may have information submitted for 180 different parcels in a given township. At a minimum, the board would receive a statistical summary of those 180 data records describing the range, percentiles (10th, 25th, 50th, 75th, 90th), and any outliers for similar soil conditions and similar crops in that township. A box and whisker plot or equivalent tabular or graphical presentation of the data approved by the Executive Officer may be used."[2]

After reviewing several potential methods for identifying outliers (e.g., methods based on the variance of the data, Mahalanobis distance, Cooks D), the ESJWQC proposes that identification of outliers for the next several years be based on the Interquartile Range (IQR) method using the A/R ratio averaged over three years.[3],[4] The IQR is a distribution-free method that does not depend on an estimate of the mean or standard deviation of the data. It is not based on a statistical test and therefore can be performed across a wide range of sample sizes and data distributions. The method was developed by J. W. Tukey, a famous statistician, in 1977 (*Tukey, J. W. 1977. Exploratory Data Analysis, Addison-Wesley, Reading MA*). The IQR method has a long history of use in identifying outliers in a variety of different settings - including in the biomedical and environmental sciences fields. One of the advantages of the IQR method is that the data and the outliers can be easily visualized in a boxplot format. Although the method is relatively insensitive to skewness and extreme values, without modification a skewed data distribution can result in a boxplot and outliers that are an inaccurate representation of the data and outliers. For this reason, the ESJWQC is proposing to use a modification of the boxplot method developed in 2008 specifically to avoid the skewness issue.

The Interquartile Range of any data distribution is the difference between the 25th percentile value of the data and the 75th percentile value of the data. Using the standard approach, outliers are determined to be any data points outside of the $25^{th}/75^{th}$ percentile data values $\pm 1.5*IQR$. The IQR method provides both low and high outliers. The adjusted boxplot method (see attached publication) is based on the *medcouple* statistic (MC), a measure of the skewness of a univariate

[2] Attachment A to Order R5-2012-0016-R4 – Information Sheet page 24
[3] As indicated in the ESJ Order, the Coalition has the discretion to re-consider the outlier approach annually. At this time, we expect to use this approach for the next five (5) years. However, we reserve the right to propose a different approach prior to that time period if a different method is determined to be preferable.
[4] Hubert, M., and E. Vandervieren. 2008. An adjusted boxplot for skewed distributions. Computational Statistics and Data Analysis 52:5186-5201.

distribution. The MC is defined as the scaled median difference of the left and right halves of a distribution. If the distribution is skewed to the right, MC ≥ 0; if the distribution is skewed to the left, MC < 0. As developed in the Hubert and Vandervieren (2008) publication, outliers in a distribution skewed to the right (MC ≥ 0) are identified as any data points outside of the interval

$$[Q_1 - (1.5e^{-4MC}) * IQR; Q_3 + (1.5e^{3MN}) * IQR] \qquad \text{Eq. 1}$$

And for distributions skewed to the left (MC < 0), outliers are data points outside of the interval

$$[Q_1 - (1.5e^{-3MC}) * IQR; Q_3 + (1.5e^{4MN}) * IQR] \qquad \text{Eq. 2}$$

Where Q_1 and Q_3 are the 25th and 75th percentiles of the data respectively, and e is the natural exponential function. When the distribution of data is perfectly symmetrical, MC = 0 and the range becomes exactly the standard (non-modified) IQR outlier method. However, the A/R data for each crop are generally skewed to the right, therefore the data range specified in Eq. 1 is more likely to be the expression used to determine the number of outliers. The focus of the ESJWQC will be those outliers that are above the 75th percentile of the IQR.

The metric used to identify outliers will be the ratio of Nitrogen Applied (A) to Nitrogen Removed (R). A/R is a measure of agronomic efficiency. When A/R = 1.0, it indicates that the nitrogen applied to a parcel is the same as the amount of nitrogen removed from the parcel through harvested plant material and/or sequestered in permanent tissue (if applicable). The IQR will focus on parcels growing a specific crop across the coalition region. For example, for mature almonds (almonds > 4 years old) the IQR will be calculated using data from all mature almonds (almonds > 4 year) grown in the ESJWQC region. For annual crops, calculation of the IQR relies on the intended use of harvested material (when there is more than one potential end use). For example, parcels planted to corn for grain (N removed coefficient = 0.012) will have a different IQR than corn chopped for silage (N removed coefficient = 0.00378), as does wheat for grain (N removed coefficient = 0.02128) and wheat for silage (N removed coefficient = 0.00525).

Per the ESJ Order, members that self-certify will receive notification about their outlier parcels and will need to attend additional INMP self-certification training in person. The other option is having their next INMP certified by a specialist in irrigation and nitrogen management (See ESJ Order, p. 52.). They will also be the focus of additional ESJWQC education and outreach and be required to track current management practices and additional practices they intend to implement to protect groundwater quality.