Robust $L_1$ orthogonal regression

J. Paul Brooks$^1$ and Edward L. Boone$^2$
Department of Statistical Sciences and Operations Research
Virginia Commonwealth University
1001 West Main Street
PO Box 843083
Richmond, VA, USA 23284-3083
Phone: +1(804)828-5853$^1$, +1(804)828-4637$^2$
Fax: +1(804)828-8785
Email: jpbrooks@vcu.edu$^1$, elboone@vcu.edu$^2$
$^1$Corresponding author

January 10, 2008
Abstract

Assessing the linear relationship between a set of continuous predictors and a continuous response is a well studied problem in statistics and is applied in many data mining situations. $L_2$ based methods such as ordinary least squares and principal components regression can be used to determine this relationship. However, both of these methods become impaired when multicollinearity is present. This problem becomes compounded when outliers confound standard multicollinearity diagnostics. This work proposes a $L_1$ orthogonal regression method ($L_1$OR) formulated as a nonconvex optimization problem. Solution strategies for finding globally optimal solutions are presented. A simulation study is conducted to determine the robustness of the method to outliers which shows that $L_1$OR is superior to ordinary least squares regression (OLS) and principal components regression (PCR) and is competitive with M-regression (M-R) in the presence of outliers. The new method outperforms OLS, PCR, and M-R on data from an environmental application.

Introduction and Background

Data miners are often posed with the problem of determining the relationship between several variables and a response variable. Standard techniques available are ordinary least squares regression (OLS), principle components regression (PCR), and non-parametric and semi-parametric regression. When outliers, or unusual observations, are present in data, these regression techniques become impaired. Methods such as M-Regression (M-R) use M estimates reduce the impact of outliers. These methods are not designed for developing errors-in-variables models in which both the predictors and the response have measurement error or are considered random components. An example of this would be measuring pH and Alkalinity in the field which usually have measurement error.

Orthogonal regression (OR) is used when uncertainty is known to be present in both independent and dependent variables. In contrast with OLS, where residuals are measured as the vertical distance of observations to the fitted surface, residuals are measured by the orthogonal distances to the fitted surface.
Previous Work on Robust Orthogonal Regression

The sensitivity of OR to outliers has been noted, and other investigators have worked to develop robust methods (Brown, 1982; Carroll and Gallo, 1982; Zamar, 1989). The work of Zamar (Zamar, 1989) includes the use of $S$ and $M$ estimates for OR.

Orthogonal regression can be formulated as equivalent to finding the last principal component, or the direction of minimum variation, in principal component analysis (PCA). Hence, any robust PCA method can be used for robust orthogonal regression. The two main approaches for robust PCA are (1) to find robust estimates of the covariance matrix (in traditional PCA, the principal components are eigenvectors of the covariance matrix) and (2) to use a robust measure of dispersion. Research in the former area includes (Campbell, 1980; Devlin et al, 1981; Galpin and Hawkins, 1987; Naga, 1990; Marden, 1999; Croux and Haesbroeck, 2000; Kamiya and Eguchi, 2001). The latter approach coincides with the work presented here. Robust estimates of dispersion in PCA have been investigated in (Li and Chen, 1985; Xie et al, 1993; Maronna, 2005).

Traditional Orthogonal Regression

Suppose we are given observations with continuous responses $(x_i, y_i) \in \mathbb{R}^d \times \mathbb{R}$, $i = 1, \ldots, n$. OR seeks to find an orthogonal projection of the data onto a hyperplane such that the orthogonal distances of the points $(x_i, y_i)$ to the hyperplane is minimized. We assume throughout this work that the means have been subtracted from samples so that the fitted hyperplane passes through the origin.

In OR, the sum of squared orthogonal distances of $(x_i, y_i)$ to the hyperplane defined by $b^T(x, y) = 0$ is minimized. Finding $b$ involves first optimizing the problem

$$\max_b \sum_{i=1}^n \| (x_i, y_i) - b^T(x_i, y_i)b \|_2^2,$$

subject to

$$b^Tb = 1.$$

The variables are in the vector $b \in \mathbb{R}^d$. The vector $b$ gives the direction of least variation in the data,
and so the best fitting hyperplane is orthogonal to \( b \) and is defined by the equation \( b^T (x_i, y_i) = 0 \).

In this paper, we present a new robust method for OR based on a robust distance measure, \( L_1 \) OR.

The direction of least variation in data is found by maximizing the \( L_1 \) distances of observations to their projection points in a linear subspace. The fitted hyperplane is orthogonal to the direction of least variation. The problem is formulated as a nonconvex optimization problem. We describe how globally optimal solutions can be derived based on a reformulation-linearization technique (RLT) developed by Sherali and Tuncbilek (Sherali and Tuncbilek, 1992). We present results of applying \( L_1 \) OR to simulated data that is contaminated with outliers and compare the results to standard methods for regression. The \( L_1 \) OR is applied to data collected for the evaluation of marine habitats, where uncertainty resides in both the dependent and independent variables.

**Minimizing \( L_1 \) residuals**

Suppose instead of maximizing the sum of the squared perpendicular distances, we maximize the sum of the \( L_1 \) distances. Note that the usual inner product that defines projections of points leads to the \( L_2 \) norm and that there is no inner product leading to the \( L_1 \) norm. We continue to use the usual inner product to define the projection of data onto the direction of least variation.

Maximizing the sum of the \( L_1 \) residuals is written as

\[
\max_b \sum_{i=1}^{n} \| (x_i, y_i) - b^T (x_i, y_i) b \|_1
\]

\[
= \max_{b} \sum_{i=1}^{n} \sum_{j=1}^{d-1} x_{ij} - b_j \left( \sum_{k=1}^{d-1} x_{ik} b_k + y_i b_d \right) + \sum_{i=1}^{n} \left| y_i - b_d \left( \sum_{k=1}^{d-1} x_{ik} b_k + y_i b_d \right) \right|.
\]

The objective function is nonlinear and nonconvex. As with \([L_2 OR]\), the optimal hyperplane is defined by \( b^T (x, y) = 0 \). Let \( r_i \) be the \( L_1 \) residual for observation \( i \). Also, let \( a = b + 1 \), where \( 1 \) is a vector of 1’s, so that all \( a_j \) variables are nonnegative. This substitution is necessary for our solution method which will be explained below. The math program can then be formulated as

\[
[L_1 OR] \max \sum_{i=1}^{n} \sum_{j=1}^{d+1} r_{ij},
\]
subject to

\[ r_i = (x_i, y_i) - (a - 1)^T (x_i, y_i)(a - 1) \quad \text{if } y_i = 0, \ i = 1, \ldots n, \]

\[ r_i = -(x_i, y_i) + (a - 1)^T (x_i, y_i)(a - 1) \quad \text{if } y_i = 1, \ i = 1, \ldots n, \]

\[ (a - 1)^T (a - 1) = 1, \]

\[ a \geq 0, \]

\[ a \leq 2, \]

\[ y_i \in \{0, 1\} \quad \text{if } i = 1, \ldots, n. \]

change \( y_i \) variables to \( z_{ij} \) The quantity 2 is a vector with each coordinate having value 2. The objective function is now linear, and each constraint is defined by a nonconvex function. The first two sets of constraints can be modeled by inequalities by introducing \( M \) constants.

To derive globally optimal solutions for \( [L_1 OR] \), we combine the use of branch and bound for integer programming with branch and bound for the reformulation-linearization technique (RLT) as described in (Sherali and Tuncbilek, 1992). Subproblem will refer to a linear mixed-integer program (MIP) that corresponds to a node in a branch and bound tree for the RLT.

The following is a summary of RLT applied to \( [L_1 OR] \):

- Select a subproblem to solve. Each subproblem is a linear MIP that relaxes the nonconvex constraints. If all subproblems have been solved, then the incumbent solution is optimal.

- If the solution satisfies the original nonconvex constraints, the current solution is feasible. Update the incumbent solution and objective value if appropriate.

- Fathom if (1) the solution satisfies the original constraints, (2) the subproblem is infeasible, or (3) the objective value for the subproblem is less than the incumbent objective value.

- Select a variable for branching, creating two subproblems.

We now describe the construction of the root subproblem for RLT. For each occurrence of \( a_j a_k \),
substitute a new variable $A_{jk}$ into the formulation. Also, add constraints of the form

\[
(2 - a_j)(2 - a_k) \geq 0 \quad j, k = 1, \ldots, d + 1; \ l = 1, \ldots, q, \\
(a_k - 0)(2 - a_j) \geq 0 \quad j, k = 1, \ldots, d + 1; \ l = 1, \ldots, q, \\
(a_k - 0)(a_j - 0) \geq 0 \quad j, k = 1, \ldots, d + 1; \ l = 1, \ldots, q,
\]

but again replace occurrences of $a_ja_k$ with $A_{jk}$. The presence of 0 in the constraints is to reflect the lower bounds on the $a_j$ variables; these lower bounds will be changed during the optimization algorithm as described below. The result is a linear MIP that is a relaxation of $[\mathcal{L}_1\text{OR}]$ (Sherali and Tuncbilek, 1992).

We now describe the branching procedure. The optimal solution to the relaxation is feasible for $[\mathcal{L}_1\text{OR}]$ if $A_{jk} = a_ja_k$ for all $j, k$. If this condition is not satisfied, then choose a variable $a_j$ with $A_{jk} \neq a_ja_k$ for some $k$ with current value $\bar{a}_j$ and create two new subproblems. One of the new subproblems will have constraints of the form

\[
(\bar{a}_j - a_j)(2 - a_k) \geq 0 \quad k = 1, \ldots, d + 1 \\
(a_k - 0)(\bar{a}_j - a_j) \geq 0 \quad k = 1, \ldots, d + 1 \\
(a_k - 0)(a_j - 0) \geq 0 \quad k = 1, \ldots, d + 1 \\
\quad a_j \leq \bar{a}_j
\]

Again, replace all occurrences of $a_ja_k$ with $A_{jk}$ to create linear constraints. The other new subproblem will have the linearized form of the constraints

\[
(2 - a_j)(2 - a_k) \geq 0 \quad k = 1, \ldots, d + 1 \\
(a_j - \bar{a}_j)(2 - a_k) \geq 0 \quad k = 1, \ldots, d + 1 \\
(a_k - 0)(a_j - \bar{a}_k) \geq 0 \quad k = 1, \ldots, d + 1 \\
\quad a_j \geq \bar{a}_j
\]

$a_k$ to $a_j$ in third set of constraints above As nodes in the branch and bound tree are traversed, the bounds for the $a_j$ variables are successively tightened. Sherali and Tuncbilek (Sherali and Tuncbilek, 1992) prove that the search for optimal solutions either terminates with a globally optimal solution in finite steps or else any accumulation point of solutions along an infinite branch
of the branch and bound tree is a globally optimal solution.

Simulation Study

A simulation study is conducted to assess the ability of $\mathcal{L}_1$ OR to detect patterns in data in the presence of outliers. A $3 \times 4$ full factorial simulation design is utilized by varying the number of contaminated observations (C) and contamination magnitude (m) for $\mathcal{L}_1$ OR, PCR, OLS and M-regression (M-R). Each method is run on 20 datasets with 30 observations under each treatment condition. For this study, C is varied in the following manner: low contamination, $C=1$, moderate contamination, $C=5$, and high contamination, $C=10$. The magnitude of contamination m is varied as: $m=1$, no contamination, $m=5$, low magnitude, $m=10$, moderate magnitude, $m=50$, large magnitude. In all cases the uncontaminated standard deviation is $\sigma = 0.1$.

$\mathcal{L}_1$ OR MIP subproblems are solved using CPLEX 10.1. CPLEX allows for specification of “indicator constraints” such as those in $\mathcal{L}_1$ OR through the function $CPXaddindconstr()$. If provable optimality is not achieved for MIP subproblems after 2 minutes, the best known integer feasible solution is used. We implemented our own branch and bound algorithm for applying RLT. Problems are solved on machines with $2 \times 2.6$GHz Opteron processors and 2GB RAM. OLS, PCR, and M-R models are derived using $lm()$, $pcr()$, and $rlm()$ functions, respectively, called in the R environment for statistical computing (http://www.r-project.org). The function $pcr()$ is contained in the $pls$ library, and $rlm()$ is in the $MASS$ library.

The data are sampled in the following manner:

- Generate uncontaminated data: $x_i \sim U[-1,1]$ and $y_i = x_i + \epsilon_i$ where $\epsilon_i \sim N(0,0.1)$, for $i = 1, \ldots, 30 - C$.
- Generate the contaminated data: $x_i \sim U[0.5,1]$ and $y_i \sim |N(0,m \times 0.1)|$ for $i = 31 - C, \ldots, 30$.

To evaluate the method’s ability to accurately fit the known underlying model the following distance or model discrepancy, $D$, is used:

$$D(\hat{f}, f) = \int_{-1}^{1} |\hat{f}(x) - f(x)|dx,$$
where $f$ is the known model and $\hat{f}$ is estimated model. Note that $D$ corresponds to the area between $f$ and $\hat{f}$. If the estimated model is close to the true model, then $D$ will be small. For each of the simulations $D$ is computed and recorded. Using these results the average model discrepancy, $\bar{D}$, and standard error is computed.

To analyze the simulation a three-way ANOVA with interactions is utilized with Tukey’s Honestly Significant Difference (HSD) multiple comparison procedure to determine where differences exist. The results of the ANOVA table are found in Table 1. The three-way interaction is significant, in that the p-value is less than 0.05. Since there is a three-way interaction all multiple comparisons must be completed in the three-way interaction space.

Tukey’s HSD multiple comparison procedure is performed on Method × m for each level of contamination. The results from this are found in Table 2. Any $\bar{D}$s that share a common subscript letter are not deemed significantly different at the 0.05 level. Table 2 shows that $L_1$ OR performs well when contamination is low $C=1$. It is not statistically different from any of the other methods. OLS and PCR perform poorly under low contamination when $m=50$. For moderate contamination, $C=5$, $L_1$ OR performs well except when $m=50$. In this case ($C=5,m=50$), M-R is statistically better. Under high contamination, $C=10$, $L_1$ OR performs the best in for each of the values of $m$. However, it is not statistically better than M-R in any case. From this simulation study we have evidence that $L_1$ OR is superior to OLS and PCR in the presence of outliers, and is competitive with M-R.

In addition to making statements about $L_1$ OR, Table 2 shows that poor performance of OLS and PCR in many cases. Hence, using a non-robust regression technique in the presence of outliers is not advised.

**An Environmental Example**

The pH and Alkalinity of the water in which the fish live is known to impact their overall health. Alkalinity is a measure of the ability of a solution to neutralize acids. Researchers expect that pH and Alkalinity to be highly correlated. However, the relationship of the two variables is difficult to estimate due to low variation in pH across streams and the presence of outliers. The dataset for
this example is a subset of values collected across the state of Ohio resulting in 45 observations. Various subsets of this dataset have been considered previously by Norton (1999), Lipkovich et al. (2002), Noble et al. (2004) and Boone et al. (2005) with varying degrees of success at estimating the relationship between Alkalinity and pH. For the purposes of this work both Alkalinity and pH have been normalized. Note that in this data both pH and Alkalinity have measurement error and hence an orthogonal regression method should be used.

Figure 1 shows the scatter plot of pH versus Alkalinity. There appears to be a linear relationship between Alkalinity and pH. Also notice the many outliers present in the data. The Pearson correlation coefficient for the relationship between pH and Alkalinity is $r = 0.5245$ which is biased down due to the outliers in the data. Furthermore, since the correlation is biased down extracting a pH-Alkalinity component and using that as a predictor would be less than optimal. Hence, a regression method is needed that is robust to outliers/influential points. The M-R as applied by Venables and Ripley (2002) in R would appear to be a good alternative estimation procedure for this problem. However, in this specific case M-R performs poorly as well.

Table 3 shows the summary of regression models from least OLS, PCR, M-R and $L_1$OR. Here the standard errors for PCR and $L_1$OR are standard bootstrap standard errors based on 1,000 bootstrap samples. Notice that the OLS and PCR coefficients are virtually identical and the M-R and $L_1$OR differ greatly from each other and OLS and PCR. Also notice that $L_1$OR produces a much steeper slope. Furthermore, the PCR regression would not be deemed statistically significant when a cut off of 0.02 is used. OLS, M-R and $L_1$OR regressions shows that the relationship between alkalinity and pH is statistically significant.

To further explore this phenomenon Figure 1 shows the plot of pH versus alkalinity with the fitted regression lines. As expected the OLS and PCR lines are identical. Furthermore, one can see the difference between the lines fitted by $L_1$OR and M-R. Notice that M-R seems to be greatly influenced by the two influential points at approximately (-6,0) and (-12,-0.9). In contrast, the $L_1$OR does not seem to be affected by these points. However, the $L_1$OR does seem to be mildly influenced by the points near (-9,-3.5).

Based on Figure 1 one could argue that the slope of the line through the data should have an even steeper slope. To achieve this one could simply omit the outliers and influential values. However, it
would be unreasonable to simply delete the outliers all together in that determining which values are outliers is subjective. Deleting outliers is not always straightforward, and detracts from the objectivity of an analysis.

Discussion

This work introduced a new $\mathcal{L}_1$ orthogonal regression technique that is designed to be robust to outliers. Via simulation, the method shows promise for being robust to outliers. Furthermore, the environmental example shows that the method produces results which are much more robust to outliers than the standard OLS and the orthogonal PCR methodology. Hence, this method gives data miners that deal with errors-in-variables data contaminated with outliers a robust alternative to PCR. Furthermore, this article method highlights the need to deal with outliers appropriately versus simply deleting them from consideration which is quite common when researchers “clean” their data.

Much work remains in proving $\mathcal{L}_1$OR computationally feasible. Future work includes developing solution techniques for the MIP subproblems and exploiting problem structure in the branch and bound procedure for the RLT technique. In the instances for this work, provable optimality for MIP subproblems is sometimes not achieved within 2 minutes; however, the best integer feasible solution at termination is usually optimal. In this study, we considered data sets with only one independent variable and one dependent variable. Future studies should investigate the efficacy of the method for data sets with more variables.

Traditional orthogonal regression ($\mathcal{L}_2$OR) can be formulated as a special case of PCA. The approach presented in this work for formulation and optimization can potentially be adapted to develop a robust method for PCA. A robust PCA algorithm would be useful for data miners that are frustrated with contaminated data and need methods to deal with it. Another possible extension is for a robust factor analysis procedure for analyzing categorical data.
References


<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>3</td>
<td>11.5840</td>
<td>3.8613</td>
<td>105.27</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>28.0840</td>
<td>14.0420</td>
<td>382.82</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>Method×C</td>
<td>6</td>
<td>3.3294</td>
<td>0.5549</td>
<td>15.13</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>m</td>
<td>3</td>
<td>103.6633</td>
<td>34.5544</td>
<td>942.04</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>Method×m</td>
<td>9</td>
<td>21.7423</td>
<td>2.4158</td>
<td>65.86</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>C×m</td>
<td>6</td>
<td>40.1508</td>
<td>6.6918</td>
<td>182.43</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>Method×C×m</td>
<td>18</td>
<td>6.1427</td>
<td>0.3412</td>
<td>9.30</td>
<td>&lt;.0001*</td>
</tr>
<tr>
<td>Error</td>
<td>912</td>
<td>33.4527</td>
<td>0.0366</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* indicates statistical significance at the 0.05 level.
Table 2: Tukey’s HSD multiple comparisons for by contamination level C for the simulation experiment. Within each contamination group, \( \bar{D} \) that share a common subscript are not significantly different at the 0.05 level.

<table>
<thead>
<tr>
<th>Method</th>
<th>m ( \bar{D} )</th>
<th>Method</th>
<th>m ( \bar{D} )</th>
<th>Method</th>
<th>m ( \bar{D} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>M-R</td>
<td>10  0.0206_a</td>
<td>( \mathcal{L}_1 ) OR</td>
<td>1     0.0254_a</td>
<td>( \mathcal{L}_1 ) OR</td>
<td>1     0.0249_a</td>
</tr>
<tr>
<td>( \mathcal{L}_1 ) OR</td>
<td>1     0.0223_a</td>
<td>OLS</td>
<td>1     0.0329_a</td>
<td>OLS</td>
<td>1     0.0360_a</td>
</tr>
<tr>
<td>OLS</td>
<td>1     0.0228_a</td>
<td>PCR</td>
<td>1     0.0329_a</td>
<td>PCR</td>
<td>1     0.0360_a</td>
</tr>
<tr>
<td>PCR</td>
<td>1     0.0228_a</td>
<td>M-R</td>
<td>1     0.0352_ab</td>
<td>M-R</td>
<td>1     0.0364_a</td>
</tr>
<tr>
<td>M-R</td>
<td>5     0.0254_a</td>
<td>( \mathcal{L}_1 ) OR</td>
<td>5     0.0682_ab</td>
<td>( \mathcal{L}_1 ) OR</td>
<td>5     0.1211_ab</td>
</tr>
<tr>
<td>( \mathcal{L}_1 ) OR</td>
<td>5     0.0295_a</td>
<td>M-R</td>
<td>5     0.0739_ab</td>
<td>M-R</td>
<td>5     0.1426_ab</td>
</tr>
<tr>
<td>M-R</td>
<td>50    0.0323_a</td>
<td>OLS</td>
<td>5     0.1131_ab</td>
<td>PCR</td>
<td>5     0.1852_abc</td>
</tr>
<tr>
<td>OLS</td>
<td>5     0.0336_a</td>
<td>PCR</td>
<td>5     0.1131_ab</td>
<td>( \mathcal{L}_1 ) OR</td>
<td>10    0.2214_abc</td>
</tr>
<tr>
<td>PCR</td>
<td>5     0.0336_a</td>
<td>M-R</td>
<td>50    0.1245_ab</td>
<td>M-R</td>
<td>10    0.2414_bc</td>
</tr>
<tr>
<td>( \mathcal{L}_1 ) OR</td>
<td>10    0.0385_a</td>
<td>( \mathcal{L}_1 ) OR</td>
<td>10    0.1331_ab</td>
<td>OLS</td>
<td>10    0.3885_c</td>
</tr>
<tr>
<td>OLS</td>
<td>10    0.0601_a</td>
<td>OLS</td>
<td>10    0.2381_b</td>
<td>PCR</td>
<td>10    0.3885_c</td>
</tr>
<tr>
<td>PCR</td>
<td>10    0.0601_a</td>
<td>PCR</td>
<td>10    0.2381_b</td>
<td>( \mathcal{L}_1 ) OR</td>
<td>50    0.9427_d</td>
</tr>
<tr>
<td>( \mathcal{L}_1 ) OR</td>
<td>50    0.1353_ab</td>
<td>( \mathcal{L}_1 ) OR</td>
<td>50    0.6474_c</td>
<td>M-R</td>
<td>50    1.0337_d</td>
</tr>
<tr>
<td>OLS</td>
<td>50    0.3175_b</td>
<td>OLS</td>
<td>50    1.3612_d</td>
<td>OLS</td>
<td>50    1.9565_e</td>
</tr>
<tr>
<td>PCR</td>
<td>50    0.3175_b</td>
<td>PCR</td>
<td>50    1.3612_d</td>
<td>PCR</td>
<td>50    1.9565_e</td>
</tr>
</tbody>
</table>
Table 3: Summary of Regression models for Alkalinity on pH using OLS, PCR, M-R and $L_1$OR. Note that the standard errors for PCR and $L_1$OR are bootstrap standard errors based on 1,000 bootstrap samples.

<table>
<thead>
<tr>
<th>Method</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>T-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>0.1877</td>
<td>0.0465</td>
<td>4.0366</td>
<td>0.0002</td>
</tr>
<tr>
<td>PCR</td>
<td>0.1877</td>
<td>0.0840</td>
<td>2.2345</td>
<td>0.0308</td>
</tr>
<tr>
<td>M-R</td>
<td>0.1376</td>
<td>0.0378</td>
<td>3.6402</td>
<td>0.0007</td>
</tr>
<tr>
<td>$L_1$OR</td>
<td>0.6017</td>
<td>0.0912</td>
<td>6.5976</td>
<td>0.0001</td>
</tr>
</tbody>
</table>
Figure 1: Scatter plot of IBI versus pH and IBI versus Alkalinity with OLS line (solid), PCR line (dotted), M-R line (longdash) and $L_1$ orthogonal regression line (dot-dash). Note: OLS and PCR lines overlay each other.