Deepest Regression in Analytical Chemistry

P. J. Rousseeuw*,a, S. Van Aelstä†,a, B. Rambali†, and J. Smeyers-Verbekeb

aDepartment of Mathematics and Computer Science, U.I.A., Universiteitsplein 1, B-2610 Antwerp, Belgium
bPharmaceutical Institute, V.U.B., Laarbeeklaan 103, B-1090 Brussels, Belgium

Abstract

Recently the concept of regression depth has been introduced [1]. The deepest regression (DR) is a method for linear regression which is defined as the fit with the best depth relative to the data. In this paper we explain the properties of the DR and give some applications of deepest regression in analytical chemistry which involve regression through the origin, polynomial regression, the Michaelis-Menten model, and censored responses.

1 Introduction

In this paper we introduce the deepest regression estimator and apply it to some chemometrical data. We show that in some cases the deepest regression can be used to solve the problem of censured data. In Section 2 we motivate the notion of regression depth which forms the basis of the deepest regression estimator. The regression depth ranks all possible fits for a given data set. Fits with a higher regression depth fit the data better than do fits with a lower regression depth. Section 3 considers the deepest regression estimator. The deepest regression is the fit which has the best regression depth relative to the data. We give the formal definition of the deepest regression estimator and show that the deepest regression still performs well in the presence of outliers. We also discuss the computation

*Corresponding author. Tel: +32/3/820.2412, fax: +32/3/820.2421.
E-mail address: Peter.Rousseeuw@uia.ua.ac.be.
†Stefan Van Aelst is Research Assistant of the Fund for Scientific Research-Flanders (Belgium).
of the regression depth and the deepest regression. In Section 4 we give some applications of regression depth and deepest regression to chemometrical problems. In particular, we consider regression through the origin, generalized linear models, polynomial regression, the Michaelis-Menten model, and regression models with censored response.

2 Regression depth

Consider a data set $Z_n$ with $p - 1$ predictor variables $x = (x_{i1}, \ldots, x_{ip-1})$ and a response variable $y_i$ for $i = 1, \ldots, n$. In linear regression we want a fit of the form $y = \theta_1x_1 + \ldots + \theta_{p-1}x_{p-1} + \theta_p$ for some $\theta = (\theta_1, \ldots, \theta_p)^t \in IR^p$, where $\theta_1, \ldots, \theta_{p-1}$ are called the slope parameters and $\theta_p$ is the intercept. The residuals of the data points relative to the fit $\theta$ are denoted as $r_i = r_i(\theta) = y_i - \theta_1x_{i1} - \cdots - \theta_{p-1}x_{i,p-1} - \theta_p$. The regression depth of a candidate fit $\theta$ indicates how well $\theta$ fits the data. Any possible fit which can be tilted in some way until it becomes vertical (i.e. useless) without passing (or touching) any observations is called a nonfit. Nonfits are considered to be the worst possible fits to the data set. Note that a fit through at least one data point cannot be a nonfit since it always touches at least the point it goes through, when it is tilted.

Let us consider an example in simple regression (i.e. $p = 2$). Figure 1 shows a data set with 6 data points and three fits $\theta$, $\eta$ and $\xi$. Let us first look at the line $\eta$. If we consider the intersection between $\eta$ and the vertical line $x = v_\eta$ which is marked by a cross in Figure 1, then we see from the plot that $\eta$ can be tilted in the direction indicated by the arrow until it becomes the vertical line $x = v_\eta$ without passing/touching any observation. Therefore, $\eta$ is considered a nonfit for this data set. Note that all observations lie above $\eta$, hence $\eta$ is indeed a very bad fit to these data. In general, a line that lies completely above or below all observations is always a nonfit. Let us now look at the line $\xi$. If we tilt this line in the direction indicated by the arrow until it becomes the vertical line $x = v_\xi$, then again we do not pass/touch any observation, hence $\xi$ is also a nonfit. Note that $\xi$ is a nonfit although it does not lie completely above or below all data points, but $\xi$ is indeed a very bad fit to these data as can be seen from Figure 1.

The regression depth of a candidate fit $\theta$ is now defined as the smallest number of observations that has to be removed from the data in order to make $\theta$ a nonfit. Equivalently, the regression depth of a fit $\theta$ is the smallest number of data points that is passed (or touched) when tilting the fit $\theta$ in any way until it becomes vertical. The regression depth
Figure 1: Bivariate data set with two nonfits $\eta$ and $\xi$, and a fit $\theta$ with regression depth 2. Thus measures how well the data are balanced about the linear fit $\theta$. Note that the regression depth only depends on the $x$-values and the signs of the residuals, but not on the actual size of the residuals.

Let us consider the line $\theta$ in Figure 1. If we tilt $\theta$ in the direction indicated by the arrow until it becomes the vertical line $x = v_\xi$ then it passes observations 4 and 5. If we tilt $\theta$ in any other way until it becomes vertical, then we always pass at least two observations. Hence the regression depth of $\theta$ equals 2. The regression depth thus indicates that the line $\theta$ fits the data better than do the lines $\eta$ and $\xi$, as can be seen from the plot.

In the special case of $p = 1$ there are no $x$-values, and $Z_n = \{y_1, \ldots, y_n\}$ is a univariate dataset. In this case a nonfit is any fit which can go to infinity without passing any observation. Hence, the regression depth of a fit $\theta \in IR$ is the rank of $\theta$, when we rank from the outside inwards.

3 Deepest regression

3.1 Definition

As seen in the previous section, the regression depth measures the quality of any candidate fit. Fits with higher regression depth fit the data better than do fits with lower regression...
depth. Hence, the regression depth ranks all possible fits from worst (depth=0) to best (maximal depth). This immediately leads to the definition of the deepest regression (DR) estimator, which is the fit with maximal regression depth relative to the data set. If several fits have the same maximal regression depth, we will take their average.

The errors in the regression model are assumed to be independent, each having zero median. These are very weak conditions, e.g. the error distribution does not have to be symmetric, nor does it have to stay the same across different values of $x$. This model is very large and allows for skewed error distributions and heteroscedasticity. Note that other robust regression methods such as Least Median of Squares (LMS) and Least Trimmed Squares (LTS) (Rousseeuw 1984) or S-estimators (Rousseeuw and Leroy 1987) assume a much smaller regression model. In particular, these methods do not allow skewed error distributions or heteroscedasticity. Another difference is that the latter methods are “mode-seeking” which means that they search for a concentrated linear cloud that contains the majority of the data. On the other hand, the deepest regression is a “median-type” regression method that searches for the center of the linear cloud of (most of) the data which implies that the DR estimator moves more gradually and in a monotone way. For a univariate dataset $\{y_1, \ldots, y_n\}$ the deepest regression is the fit with maximal rank, hence it equals the sample median. The DR thus generalizes the univariate median to linear regression.

3.2 Robustness

To show the robustness of the DR, let us consider the breakdown value. The breakdown value (see e.g. [2]) of an estimator $T$ is the smallest fraction of points of the data set $Z_n$ that must be replaced by arbitrary values to make $T$ explode. The breakdown value of the deepest regression [1] is always at least $1/(p + 1)$. If the good data satisfy the conditions above, then the breakdown value converges (as $n$ goes to infinity) to $1/3$ in any dimension $p$ [1, 3]. This means that the deepest regression does not break down when at least 67% of the data come from the model given above, while the remaining data (i.e., up to 33% of the points) may be anything. A property related to the breakdown value is the exact fit property. For the deepest regression it holds that if at least 2/3 of the data points are exactly on a hyperplane, then the deepest regression fit equals this hyperplane whatever the other observations are. Other robustness properties (such as the influence function) are given in [3]. Asymptotics of the DR are given in [4, 5].

The robustness properties show that the DR is very different from least absolute devia-
Figure 2: Hertzsprung-Russell diagram of 47 stars in the direction of Cygnus with the deepest regression line DR and the least absolute deviations line \textit{LAD} which is attracted by the giant stars.

\textit{LAD} regression, which is another generalization of the univariate median to regression. The \textit{LAD} regression is defined as the fit that minimizes the sum of the absolute values of the residuals. However, \textit{LAD} regression has zero breakdown value due to its vulnerability to leverage points. (Leverage points are data points which have a large value of x). To illustrate this, we use a data set from astronomy, obtained from [2]. The data in Figure 2 form the Hertzsprung-Russell diagram of the star cluster CYG OB1, which contains 47 stars in the direction of Cygnus. In Figure 2 the logarithm of each star’s light intensity is plotted versus the logarithm of its surface temperature, where the log temperature is plotted from right to left. In astronomy a Hertzsprung-Russell diagram can e.g. be used to classify stars with regard to their age. In a Hertzsprung-Russell diagram there are two groups: the main sequence which follows a steep linear trend and the stars in the upper right corner which are red giants. In Figure 2 we see that the \textit{LAD} line of this data set is attracted by the four leverage points in the upper right corner (the giant stars), whereas the deepest regression line is robust and hence fits the majority (the main sequence stars).

Note that the least squares (LS) and \textit{LAD} estimators both have zero breakdown value,
which means that a single outlier in a data set can make the LS and LAD estimates completely useless. Outliers (even a single one) can attract the LS and LAD estimates such that their residuals become smaller than residuals of good data points. Therefore, using standardized LS or LAD residuals to detect outliers can cause good data points to be incorrectly identified as outliers while the true outliers remain hidden. For example, in Figure 2 we clearly see that the four giant stars in the upper right corner have a much smaller residuals corresponding to the LAD line than most of the stars in the main sequence. Hence, classical diagnostics based on the LAD estimates would not lead to the identification of the four giant stars as outliers. Therefore, it is recommended to use high breakdown robust estimators such as the DR estimator to estimate the linear trend of the good data points and use diagnostics based on the residuals corresponding to the robust estimates to identify outliers correctly. In two dimensional data sets, outliers can usually be detected by visual inspection of the data. However, automatic procedures might not allow visual inspection of the data. Examples throughout this paper show that consequences may be dramatic when the procedure is based on nonrobust methods such as LS or LAD. In higher dimensions \((p \geq 3)\) even visual inspection can be insufficient to detect outliers, hence the need for robust methods becomes even clearer in this case.

### 3.3 Computation

In \(p = 2\) dimensions the regression depth of a fit \(\theta = (\theta_1, \theta_2)^t\) can be computed in \(O(n \log n)\) time with the algorithm described in [1]. This algorithm first sorts the observations in \(O(n \log n)\) time such that \(x_1 \leq x_2 \leq \cdots \leq x_n\). The regression depth of \(\theta\) is then given by

\[
\min_{1 \leq i \leq n} \min \{L_i^+ + R_i^-, L_i^- + R_i^+\}
\]

where \(L_i^+ = \#\{j; x_j \leq x_i \text{ and } r_j \geq 0\}\) is the number of observations to the left of \((x_i, y_i)\) with a positive residual, \(R_i^- = \#\{j; x_j > x_i \text{ and } r_j \leq 0\}\) is the number of observations strictly to the right of \((x_i, y_i)\) with a negative residual, and \(L_i^-\) and \(R_i^+\) are defined accordingly. This minimum can be computed in \(O(n)\) time because it suffices to update \(L_i^+, L_i^-, R_i^-\), and \(R_i^+\) at each \(i = 1, \ldots, n\). Thus we obtain a total time complexity of \(O(n \log n)\). To compute the regression depth of a fit \(\theta\) in \(p = 3\) or \(p = 4\) dimensions, exact algorithms with time complexity \(O(n^{p-1} \log n)\) have been constructed [6]. These algorithms compute the smallest number of observations that has to be passed or touched when tilting \(\theta\) until it becomes a vertical hyperplane through \(p - 1\) data points. To overcome this high time complexity, the same paper also gives an approximate algorithm for datasets with large \(n\) and/or \(p\) that computes the regression depth of a fit in \(O(mp^3 + mpn + mn \log n)\) time,
which is quite feasible. This approximate algorithm randomly selects vertical hyperplanes, and for each of these vertical hyperplanes it computes the smallest number of observations that has to be passed or touched when tilting the fit $\theta$ until it becomes parallel to this vertical hyperplane. The overall minimum then approximates the regression depth of the fit $\theta$.

Note that the regression depth can only increase if we move the fit until it passes through an observation (while not changing the signs of the other residuals). Therefore, a naive exact algorithm for the deepest regression computes the regression depth of all $O(n^p)$ fits through $p$ observations and keeps the one(s) with maximal depth. This yields a total time complexity of $O(n^{2p-1} \log n)$ which is very slow for large $n$ and/or high $p$. Even if we use the approximate algorithm proposed in [6] to compute the depth of each fit, the time complexity remains very high. For simple regression, researchers in computational geometry have obtained exact algorithms of complexity $O(n \log^2 n)$ [7] and even $O(n \log n)$ [8], i.e. little more than linear time. To speed up the computation of the DR in more than two dimensions, the fast approximate algorithm MEDSweep has been constructed [9]. This is a sweeping algorithm (see [10] p. 246) based on deepest regression through the origin, which is explained in Section 4.1. The MEDSweep algorithm as well as the exact and approximate regression depth algorithms are available as stand-alone FORTRAN programs. The programs can be downloaded from our website http://win-www.uia.ac.be/u/statist/ where their use is explained.

4 Applications

4.1 Regression through the origin

The $DR$ can also be applied to regression through the origin, i.e. to the model $y = \theta_1 x_1 + \cdots + \theta_p x_p$. For regression through the origin the regression depth of a fit $\theta = (\theta_1, \ldots, \theta_p)$ through the origin is defined as the smallest number of observations that needs to be passed when tilting $\theta$ in any direction until it becomes a vertical fit through the origin. Using this definition, the $DR$ through the origin is defined as before. In the case of regression through the origin we can transform the data and the model to a regression with intercept, for which the previous properties continue to hold (see [1]). When $p = 1$ the model becomes $y = \theta_1 x$, i.e. a line through the origin. The $DR$ fit then reduces to the simple formula $\hat{\theta}_1 = \text{median}(y_i/x_i)$ where observations with $x_i = 0$ are not counted.
To illustrate the deepest regression in the case of regression through the origin we consider the calibration data set shown in Figure 3. The data represent a calibration line measured in peak area (AS) for cadmium from graphite furnace atomic absorption spectrometry (GFAAS). The concentrations are in ng/ml. From Figure 3 we clearly see that the least squares (LS) line through the origin is attracted by the outlier at 24 ng/ml while the deepest regression line through the origin is robust and fits the good data points.

Figure 3: Calibration data set with an outlier at 24 ng/ml and the corresponding least squares line (LS) and deepest regression line (DR) through the origin.

4.2 Regression depth of a general function

As seen in Section 2, the regression depth of a fit $\theta$ relative to a given dataset only depends on the $x$-values and the signs of the residuals. Therefore this definition can also be applied to more general models. For example, suppose we have a regression fit of the form

$$y = f(x_1, \ldots, x_{p-1})$$

for some real function $f$. Denote the residuals as $r_i(f) = y_i - f(x_{i1}, \ldots, x_{i,p-1})$ for $i = 1, \ldots, n$. Now any vertical hyperplane can be denoted as $x_1u_1 + \cdots + x_{p-1}u_{p-1} = v$ where
\( u = (u_1, \ldots, u_{p-1}) \) is a vector with \( \|u\| = 1 \) and \( v \) is any real number. For any vertical hyperplane consider the total number of observations with \( x_{i1}u_1 + \cdots + x_{i,p-1}u_{p-1} < v \) and \( r_i(f) \geq 0 \) and with \( x_{i1}u_1 + \cdots + x_{i,p-1}u_{p-1} > v \) and \( r_i(f) \leq 0 \). Then the regression depth of the fit \( f \) is the minimum of this number over all possible vertical hyperplanes (i.e. over all possible \( u \) and \( v \)) which do not contain any observation. Note that for a linear fit \( f(x_1, \ldots, x_{p-1}) = \theta_1x_1 + \cdots + \theta_{p-1}x_{p-1} + \theta_p \) this gives the number of observations passed when tilting the fit \( \theta \) until it becomes vertical, \( x_{i1}u_1 + \cdots + x_{i,p-1}u_{p-1} = v \), hence in the case of linear regression this definition of regression depth coincides with the definition of Section 2.

The regression depth has the following monotone invariance property.

**Proposition 1** Suppose we have a data set \( Z_n = \{(x_{i1}, \ldots, x_{i,p-1}, y_i); i = 1, \ldots, n\} \) and a strictly monotone real function \( g \). Denote \( y'_i = g(y_i) \) and \( Z'_n = \{(x_{i1}, \ldots, x_{i,p-1}, y'_i); i = 1, \ldots, n\} \). Then it holds for any function \( f \) that the regression depth of \( f \) relative to the data set \( Z_n \) equals the regression depth of the fit \( g(f) \) relative to the data set \( Z'_n \).

This property of regression depth allows us to deal with several interesting models, as shown in the following examples.

### 4.3 Generalized linear models

Suppose we want a regression fit of the form

\[
y = g(\theta_1x_1 + \cdots + \theta_{p-1}x_{p-1} + \theta_p)
\]

with \( g \) a link function. Denote \( r_i(g\theta) = y_i - g(\theta_1x_{i1} + \cdots + \theta_{p-1}x_{i,p-1} + \theta_p) \), then the regression depth of the (nonlinear) fit \( g\theta \) is given in Section 4.2. Using this definition of depth, we can now compute the deepest generalized linear regression defined as in Section 2 and denote it by \( DR_g(Z_n) \).

From the monotone invariance of the regression depth (Proposition 1) it follows that the deepest regression is equivariant for monotone transformations of the response \( y_i \). This monotone equivariance does not hold for LAD or other estimators such as least squares and least trimmed squares [11].

**Proposition 2** Consider \( Z_n = \{(x_{i1}, \ldots, x_{i,p-1}, y_i); i = 1, \ldots, n\} \) and a strictly monotone link function \( g \). Put \( \tilde{y}_i = g^{-1}(y_i) \) and denote the deepest linear regression of the transformed data \((x'_{i1}, \tilde{y}_i)\) as \( \hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_p)' \). Then the deepest generalized linear regression to the original
Typical examples of $g$ include the logarithmic, the exponential, the square root, the square and the reciprocal transformation. Proposition 2 shows that the deepest regression fit $DR_g(Z_n)$ can be obtained by computing the deepest linear regression (e.g. using the MEDSWEEP algorithm) of the transformed data.

4.4 Polynomial regression

Consider a dataset $Z_n = \{(x_i, y_i); i = 1, \ldots, n\} \subset IR^2$. Polynomial regression wants to fit the data by $y = \theta_1 x + \theta_2 x^2 + \cdots + \theta_k x^k + \theta_{k+1}$ where $k$ is called the degree of the polynomial.

The residuals of $Z_n$ relative to the fit $\theta = (\theta_1, \ldots, \theta_{k+1})^t$ are denoted as $r_i = r_i(\theta) = y_i - \theta_1 x - \cdots - \theta_k x^k - \theta_{k+1}$.

We could consider this to be a multiple linear regression problem with regressors $x = (x, x^2, \ldots, x^k)^t$ and determine the depth of a fit $\theta$ as in Section 2. But we know that the joint distribution of $(x^t, y)$ is degenerate (i.e. it does not have a density), so many properties would not hold in this case.

A better way to define the depth of a polynomial fit $f(x) = \theta_1 x + \theta_2 x^2 + \cdots + \theta_k x^k + \theta_{k+1}$ is to use the definition of the regression depth of general functions given in Section 4.2. Note that in this case $x$ is univariate. We denote the corresponding deepest polynomial of degree $k$ by $DR_k(Z_n)$. It can be shown that with this definition of depth the deepest polynomial regression has a positive breakdown value of approximately 1/3, so it is robust to vertical outliers as well as to leverage points.

Figure 4 shows a calibration data set. The data represent a calibration line measured in peak height (A) for aluminum from graphite furnace atomic absorption spectrometry (GFAAS). The concentrations are in ng/ml. We computed the least squares and deepest quadratic through the origin. From Figure 4 we clearly see that the least squares fit LS is influenced by the inaccurate measurement at 120 ng/ml. The resulting least squares fit is convex while the majority of the data points suggest a concave parabola. On the other hand, the deepest regression parabola DR is robust and yields a concave parabola corresponding to the majority of the data points.
Figure 4: Calibration data set with an inaccurate observation at 120 ng/ml and the corresponding least squares quadratic fit LS and deepest quadratic fit DR.

4.5 Michaelis-Menten model

In the field of enzyme kinetics, the steady-state kinetics of the great majority of the enzyme-catalyzed reactions that have been studied are adequately described by a hyperbolic relationship between the concentration $s$ of a substrate and the steady-state velocity $v$. This relationship is expressed by the Michaelis-Menten model

$$v = \frac{v_{\text{max}}s}{K_m + s}$$

(3)

where the constant $v_{\text{max}}$ is the maximum velocity and $K_m$ is the Michaelis constant. The Michaelis-Menten model is nonlinear, and has been linearized by rewriting it in the following three ways:

$$\frac{v}{s} = \frac{v_{\text{max}}}{K_m} - \frac{1}{K_m}v$$

(4)

$$\frac{1}{v} = \frac{1}{v_{\text{max}}} + \frac{K_m}{v_{\text{max}}} \frac{1}{s}$$

(5)

$$\frac{s}{v} = \frac{K_m}{v_{\text{max}}} + \frac{1}{v_{\text{max}}}s$$

(6)

which are known as the Scatchard equation [12], the double reciprocal equation [13] and the Woolf equation [14]. Each of the three relations (4), (5), (6) can be used to estimate
the constants \( v_{\text{max}} \) and \( K_m \). In general the three relations yield different estimates for the constants \( v_{\text{max}} \) and \( K_m \), because the error terms are also transformed in a nonlinear way. Cressie and Keightley [15] compared these three linearizations of the Michaelis-Menten relation in the context of hormone-receptor assays, and concluded that both the double reciprocal equation (5) and the Woolf equation (6) work fairly well. This leaves us with the choice between these two linearizations, and the original nonlinear form (3).

Theorem 1 shows that applying the deepest regression to the Woolf equation (6) yields the same estimates for \( v_{\text{max}} \) and \( K_m \) as the deepest regression applied to the double reciprocal equation (5). This resolves the ambiguity. Moreover, if we transform this solution to the original \((s,v)\)-space, then we obtain the deepest regression fit to the original nonlinear Michaelis-Menten model given by (3).

**Theorem 1** Let \( Z_n = \{(s_i, v_i); i = 1, \ldots, n\} \subset \mathbb{R}^2 \) with \( s_i > 0 \) for all \( i = 1, \ldots, n \), and denote the DR fit of the double reciprocal equation as \( \text{DR}(\{(\frac{1}{s_i}, \frac{1}{v_i}); i = 1, \ldots, n\}) = (\hat{\theta}_1, \hat{\theta}_2)^t \). Then the DR fit of the Woolf equation satisfies

\[
\text{DR}(\{(s_i, \frac{s_i}{v_i}); i = 1, \ldots, n\}) = (\hat{\theta}_2, \hat{\theta}_1)^t.
\]

Moreover, the deepest regression fit for the original Michaelis-Menten equation (3) is exactly given by

\[
v = \frac{s}{\hat{\theta}_2 s + \hat{\theta}_1}.
\]

In all three cases we obtain the same \( \hat{v}_{\text{max}} = 1/\hat{\theta}_2 \) and \( \hat{K}_m = \hat{\theta}_1/\hat{\theta}_2 \).

**Example:** In assays for hormone receptors the Michaelis-Menten equation describes the relationship between the amount \( B \) of hormone bound to receptor and the amount \( F \) of hormone not bound to receptor. These assays are used e.g. to determine the cancer treatment method (see [15]). Equation (3) now becomes

\[
B = \frac{B_{\text{max}} F}{K_D + F}.
\]

The parameters of interest are the concentration \( B_{\text{max}} \) of binding sites and the dissociation constant \( K_D \) for the hormone-receptor interaction. Figure 5a shows the Woolf plot of data from an estrogen receptor assay obtained by [16]. Note that this dataset clearly contains one outlier. In the plot we indicated the deepest regression line \( \frac{F}{B} = 0.567 + 0.0215 F \). In Figure 5b we show the double reciprocal plot with its deepest regression line \( \frac{1}{B} = 0.0215 + 0.567 \frac{1}{F} \). Thus in both cases we obtain the same estimated values \( \hat{B}_{\text{max}} = 46.556 \) and \( \hat{K}_D = 26.398 \),
Figure 5: (a) Woolf plot of the Cressie and Keightley data with the deepest regression line $\frac{F}{B} = 0.567 + 0.0215 F$ and the least squares fit $\frac{F}{B} = 0.728 + 0.0191 F$; (b) double reciprocal plot of the data with the deepest regression line $\frac{1}{B} = 0.0215 + 0.567 \frac{1}{F}$ and the least squares fit $\frac{1}{B} = 0.0203 + 0.596 \frac{1}{F}$. 
Figure 6: Plot of the Cressie and Kightley data. Superimposed are the deepest regression (DR), the transformed LS fit according to the Woolf equation (LSw) and the transformed LS fit according to the reciprocal equation (LSr).

which are comparable to the least squares estimates $\hat{B}_{\text{max}} = 45.610$ and $\hat{K}_D = 24.079$ obtained from the Woolf equation based on all data except the outlier. On the other hand, least squares applied to the full data gives $\hat{B}_{\text{max}} = 52.290$ and $\hat{K}_D = 38.048$ based on the Woolf equation, and $\hat{B}_{\text{max}} = 49.354$ and $\hat{K}_D = 29.436$ based on the double reciprocal equation. These estimates are quite different. This can also be seen in Figure 6, which shows the data points and the estimated curves in the original (F,B) space. Here, LSw resp. LSr stand for the LS fit according to the Woolf equation and the reciprocal equation. With least squares, we see that in both cases the estimates $\hat{B}_{\text{max}}$ and $\hat{K}_D$ are highly influenced by the outlying observation (both $\hat{B}_{\text{max}}$ and $\hat{K}_D$ come out too high) which may lead to wrong conclusions, e.g. when determining a cancer treatment method. Note that the outlier lies outside the plot region in the direction of the arrow.
4.6 Censored data

Regression depth only depends on the $x$-values and the signs of the residuals, which allows us to apply DR to data with censored responses. If we do not know the actual response of an observation but we do know the sign of its residual, then we can still compute the deepest regression. Hence, we do not have to delete the observation from the data set. We simply set the missing responses equal to an arbitrary large value $c$ (relative to the measured responses) multiplied by the sign of the residual corresponding to the observation. If we now compute the deepest regression, it follows that the residuals corresponding to observations with missing response have the desired sign.

When we also need to estimate the scale of the residuals, we use a robust scale estimator such that the scale estimate is not affected by the large residuals. The $Q_n$ scale estimator [17] has a 50% breakdown value and a high statistical efficiency, which makes it an appropriate choice. After computing the $Q_n$ scale $\hat{\sigma}$, we replace all positive residuals larger than $2\hat{\sigma}$ by $2\hat{\sigma}$, and all negative residuals below $-2\hat{\sigma}$ by $-2\hat{\sigma}$. The corresponding response now becomes the sum of the fitted value and this ‘truncated residual’. Finally, we compute the least squares fit to this new data set. (This construction corresponds to computing a so-called one-step Huber-type M-estimator starting from a robust initial estimate, see e.g. [18, 19].) Since the least squares fit to the final data yields confidence intervals and corresponding $p$-values, we can perform variable selection to obtain a significant regression model.

We will illustrate this method on data from an experimental design to study the granulation process in a fluidized powder bed [20]. In particular, the effect of spray rate, inlet air flow rate, inlet air temperature and inlet air humidity was investigated. The response variable was the geometric mean of the granule size. After a first analysis, the explanatory variables were reduced to powder bed moisture content ($x_1$) and spray rate ($x_2$). However, due to overwetting of the powder bed 8 responses could not be measured, but their corresponding residual to the fit would be positive. We have set their responses equal to 1 000 000 and computed the deepest regression fit where we included linear terms, quadratic terms and interaction in the model. Variable selection yielded a significant regression model with linear terms, the interaction term and a quadratic term for spray rate. Figure 7 shows the truncated DR residuals of the observations. Data points with missing response are plotted as diamonds. The solid horizontal lines are at $2\hat{\sigma}$ and $-2\hat{\sigma}$. We see that the residuals of the points with missing response were set equal to $2\hat{\sigma}$. Also some negative residuals were truncated by this procedure, which reduces the effect of possible outliers on the regression
Figure 7: Adjusted DR residuals of the observations for a model with linear terms, the interaction term and a quadratic term for spray rate. Data points with missing response are plotted as diamonds.

model. The coefficients of the final fit are shown in Table 1. We see that all parameters are highly significant.

5 Conclusions

In this paper we have seen that regression depth and deepest regression can be useful to solve several problems in analytical chemistry. The deepest regression method is widely applicable and makes very weak assumptions, which makes it an interesting approach for problems where conventional methods are not suitable or applicable. The regression depth is based on the $x$-values and the signs of the residuals. Therefore, the $DR$ is robust to outliers and can be applied when some actual responses are missing but the sign of their residuals is known. The $DR$ was also applied to the Michaelis-Menten model of enzyme kinetics, where it resolves a long-standing ambiguity.
Table 1: Estimated coefficients of the final analysis of the granulation process. The response variable is the geometric mean of the granule size. The independent variables are the bed moisture content ($x_1$) and spray rate ($x_2$).

### Appendix

**Proof of Proposition 1.** Denote $r_i(f, (x_i, y_i)) = y_i - f(x_i)$ and $r_i(g(f), (x_i, y'_i)) = y'_i - g(f(x_i))$. For a monotone increasing function $g$ it follows that

$$\text{sign}(r_i(f, (x_i, y_i))) = \text{sign}(y_i - f(x_i))$$

$$= \text{sign}(g(y_i) - g(f(x_i)))$$

$$= \text{sign}(r_i(g(f), (x_i, y'_i))).$$

Since only the responses are transformed, it follows from the general definition of regression depth in Section 4.2 that

$$r_{\text{depth}}(f, Z_n) = r_{\text{depth}}(g(f), Z'_n).$$

For monotone decreasing functions the proof is similar.

**Proof of Proposition 2.** This follows immediately from Proposition 1 which implies that $r_{\text{depth}}(\theta, \tilde{Z}_n) = r_{\text{depth}}(g_\theta, Z_n)$ for all possible fits $\theta$. □

**Proof of Theorem 1.** We first prove that the Woolf equation and the double reciprocal equation yield the same estimates for $v_{\text{max}}$ and $K_m$ when using the $DR$. We will show that when $s_i > 0$ for all $i = 1, \ldots, n$ it holds for every $\theta = (\theta_1, \theta_2)'$ that $r_{\text{depth}}((\theta_1, \theta_2)', \{(\frac{1}{s_i}, \frac{1}{v_i}); i = 1, \ldots, n\}) = r_{\text{depth}}((\theta_2, \theta_1)', \{(s_i, \frac{s_i}{v_i}); i = 1, \ldots, n\})$. This follows from

$$r_i((\theta_1, \theta_2)', \{(\frac{1}{s_i}, \frac{1}{v_i}); i = 1, \ldots, n\}) = \frac{1}{v_i} - \theta_1 \frac{1}{s_i} - \theta_2$$

$$= \frac{1}{s_i}(\frac{s_i}{v_i} - \theta_1 - \theta_2 s_i)$$

$$= \frac{1}{s_i}r_i((\theta_2, \theta_1)', \{(s_i, \frac{s_i}{v_i}); i = 1, \ldots, n\}).$$
Since \( s_i > 0 \) for all \( i = 1, \ldots, n \) we have

\[
\text{sign}(r_j((\theta_1, \theta_2)^t, \{(\frac{1}{s_i} \frac{1}{v_i}; i = 1, \ldots, n)})) = \text{sign}(r_j((\theta_2, \theta_1)^t, \{(s_i, \frac{s_i}{v_i}); i = 1, \ldots, n)}))
\]

for all \( j = 1, \ldots, n \), and switching the \( x \)-values from \( \frac{1}{s_i} \) to \( s_i \) reverses their order. Therefore, according to the definition of regression depth in Section 2 both depths are the same.

We now show that we obtain the deepest regression solution to the original Michaelis-Menten equation. For this we will show that it holds for every \( \theta = (\theta_1, \theta_2)^t \) that

\[
r\text{depth}(((\theta_1, \theta_2)^t, \{(\frac{1}{s_i} \frac{1}{v_i}; i = 1, \ldots, n)})) = r\text{depth}(g(\frac{1}{\theta_2}, \frac{1}{\theta_1}), \{(s_i, v_i); i = 1, \ldots, n)}))
\]

where

\[
g(\eta_1, \eta_2)(\{(s_i, v_i); i = 1, \ldots, n)} = \frac{m \eta_1}{\eta_2 + \eta_1}. \quad \text{We have that} \quad \frac{1}{v_i} - (\theta_1 \frac{1}{s_i} + \theta_2) < 0 \implies v_i - \frac{s}{\theta_2 + \theta_1} = v_i - \frac{s/\theta_2}{s + \theta_1/\theta_2} > 0 \quad \text{and vice versa. This implies that} \quad r_i(((\theta_1, \theta_2)^t, \{(\frac{1}{s_i} \frac{1}{v_i}; i = 1, \ldots, n)})) \quad \text{and} \quad r_i(g(\frac{1}{\theta_2}, \frac{1}{\theta_1}), \{(s_i, v_i); i = 1, \ldots, n)})) \quad \text{have opposite sign. Moreover,} \quad \frac{1}{s_i} \quad \text{is a monotone (decreasing) transformation of} \quad s_i. \quad \text{Hence, according to the general definition of regression depth both depths are the same.} \]

\[
\]

**References**


