

ROBUST CHAUVENET OUTLIER REJECTION

M. P. MAPLES, D. E. REICHART, N. C. KONZ, T. A. BERGER, A. S. TROTTER, J. R. MARTIN, D. A. DUTTON,
M. L. PAGGEN, R. E. JOYNER, C. P. SALEM

Department of Physics and Astronomy, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599

Accepted to the Astrophysical Journal Supplement Series

ABSTRACT

Sigma clipping is commonly used in astronomy for outlier rejection, but the number of standard deviations beyond which one should clip data from a sample ultimately depends on the size of the sample. Chauvenet rejection is one of the oldest, and simplest, ways to account for this, but, like sigma clipping, depends on the sample's mean and standard deviation, neither of which are robust quantities: Both are easily contaminated by the very outliers they are being used to reject. Many, more robust measures of central tendency, and of sample deviation, exist, but each has a tradeoff with precision. Here, we demonstrate that outlier rejection can be both very robust *and* very precise if decreasingly robust but increasingly precise techniques are applied in sequence. To this end, we present a variation on Chauvenet rejection that we call “robust” Chauvenet rejection (RCR), which uses three decreasingly robust/increasingly precise measures of central tendency, and four decreasingly robust/increasingly precise measures of sample deviation. We show this sequential approach to be very effective for a wide variety of contaminant types, even when a significant – even dominant – fraction of the sample is contaminated, and especially when the contaminants are strong. Furthermore, we have developed a bulk-rejection variant, to significantly decrease computing times, and RCR can be applied both to weighted data, and when fitting parameterized models to data. We present aperture photometry in a contaminated, crowded field as an example. RCR may be used by anyone at <https://skynet.unc.edu/rcr>, and source code is available there as well.

Subject headings: methods: statistical — methods: data analysis

1. INTRODUCTION

Consider a sample of outlying and non-outlying measurements, where the non-outlying measurements are drawn from a given statistical distribution, due to a given physical process, and the outlying measurements are sample contaminants, drawn from a different statistical distribution, due to a different, or additional, physical process, or due to non-statistical errors in measurement. Furthermore, the statistical distribution from which the outlying measurements are drawn is often unknown. Whether (1) combining this sample of measurements into a single value, or (2) fitting a parameterized model to these data, outliers can result in incorrect inferences.

There are a great many methods for identifying and either down-weighting (see §2.1) or outright rejecting outliers. The most ubiquitous, particularly in astronomy, is sigma clipping. Here, measurements are identified as outlying and rejected if they are more than a certain number of standard deviations from the mean, assuming that the sample is otherwise distributed normally. Sigma clipping, for example, is a staple of aperture photometry, where it is used to reject signal above the noise (e.g., other sources, Airy rings, diffraction spikes, cosmic rays, and hot pixels), as well as overly negative deviations (e.g., cold pixels), when measuring the background level in a surrounding annulus.

Sigma clipping, however, is crude in a number of ways, the first being where to set the rejection threshold. For example, if working with ≈ 100 data points, 2-sigma deviations from the mean are expected but 4-sigma devia-

tions are not, so one might choose to set the threshold between 2 and 4. However, if working with $\approx 10^4$ points, 3-sigma deviations are expected but 5-sigma deviations are not, in which case a greater threshold should be applied.

Chauvenet rejection is one of the oldest, and also the most straightforward, improvement to sigma clipping, in that it quantifies this rejection threshold, and does so very simply (Chauvenet 1863). Chauvenet's criterion for rejecting a measurement is:

$$NP(>|z|) < 0.5, \quad (1)$$

where N is the number of measurements in the sample, and $P(>|z|)$ is the cumulative probability of the measurement being more than z standard deviations from the mean, assuming a Gaussian distribution. We apply Chauvenet's criterion iteratively, rejecting only one measurement at a time for increased stability, but consider the case of (bulk) rejecting all measurements that meet Chauvenet's criterion each iteration in §5.¹ In either case, after each iteration (1) we lower N by the number of points that we rejected, and (2) we re-estimate the mean and standard deviation, which are used to compute each measurement's z value, from the remaining, non-rejected measurements.

However, both traditional Chauvenet rejection, as well as its more-general, less-defined version, sigma clipping, suffer from neither the mean nor the standard deviation

¹ Some care must be taken here: Since both the mean and the standard deviation change each iteration, measurements that were outlying can become not outlying (though the opposite is usually the case).

being “robust” quantities: Both are easily contaminated by the very outliers they are being used to reject. In §2, we consider increasingly robust (but decreasingly precise; see below) replacements for the mean and standard deviation, settling on three measures of central tendency (§2.1) and four measures of sample deviation (§2.2). We calibrate seven pairings of these, using uncontaminated data, in §2.3.

In §3, we evaluate these increasingly robust improvements to traditional Chauvenet rejection against different contaminant types: In §3.1, we consider the case of two-sided contaminants, meaning that outliers are as likely to be high as they are to be low; and in §3.2, we consider the (more challenging) case of one-sided contaminants, where all or almost all of the outliers are high (or low; we also consider in-between cases here). In §3.3, we consider the case of rejecting outliers from mildly non-Gaussian distributions.

In §3, we show that these increasingly robust improvements to traditional Chauvenet rejection do indeed result in increasingly accurate measurements, and they do so in the face of increasingly high contaminant fractions and contaminant strengths. But at the same time, these measurements are decreasingly precise. However, in §4, we show that one can make measurements that are both very accurate *and* very precise, by applying these techniques in sequence, with more-robust techniques applied before more-precise techniques.

In §5, we evaluate the effectiveness of bulk rejection, which can be significantly less demanding computationally. In §6, we consider the case of weighted data. In §7, we exercise both of these techniques with an astronomical example.

In §8, we show how RCR can be applied to model fitting, which first requires a generalization of this, traditionally, non-robust process. In §9, we compare RCR to Peirce rejection (Peirce 1852; Gould 1855), which is perhaps the next-most commonly used outlier-rejection technique (Ross 2003). Peirce rejection is a non-iterative alternative to traditional Chauvenet rejection, that can also be applied to model fitting, and that has a reputation of being superior to traditional Chauvenet rejection.

We summarize our findings in §10.

2. ROBUST TECHNIQUES

2.1. Measures of Central Tendency

There are a wide variety of, increasingly robust, ways to measure central tendency. For example, instead of the mean, one could use the WindsORIZED mean, in which the values in each tail of a distribution are replaced by the most extreme value remaining, before calculating the mean. Or, one could use the truncated mean, in which these values are instead simply discarded. In either case, such measures are a tradeoff, or a compromise, between robustness and precision, depending on what fraction of each side of the distribution is replaced or discarded: If 0% is replaced or discarded, these measures are just the mean, which is not robust, but is precise; in the limit that all but one value (or two, if there are an even number of values in the distribution) are replaced or discarded, these measures are equivalent to the median, which is more robust than the mean, but less precise.

In this paper, we are not trying to introduce a com-

promise between robustness and precision. Rather, we are attempting to have both by applying measures with differing properties in sequence. Consequently, we limit this investigation to the three most-common measures of central tendency, which already have a wide range of properties: the mean, the median, and the mode, which are increasingly robust, but decreasingly precise.

The mean and median are calculated in the usual ways: The mean is given by summing a data set’s values, and dividing by its number of values, N ; the median is given by instead sorting these values, and taking the middle value if N is odd, and the mean of the two middle values if N is even.

Given continuous data, the mode, however, can be defined in a variety of ways. We adopt an iterative half-sample approach (e.g., Bickel & Frühwirth 2005), and calculate the mode as follows. Sort the data, x_i , and for every index j in the first half of the data set, including the middle value if N is odd, let k be the largest integer such that:

$$k \leq j + 0.5N. \quad (2)$$

Of these (j, k) combinations, select the one for which $|x_k - x_j|$ is smallest. If multiple combinations meet this criterion, let j be the smallest of their j values and k be the largest of their k values. Restricting oneself to only the $k - j + 1$ values between and including j and k , repeat this procedure, iterating to completion. Take the median of the final $k - j + 1$ (typically two) values.

2.2. Measures of Sample Deviation

As with central tendency, there are a wide variety of measures of sample deviation. When we use the mean to measure central tendency, we use the standard deviation to measure sample deviation: Neither are robust, but both are precise.

However, when using more-robust measures of central tendency, like the median or the mode, we need to pair these with more-robust measures of sample deviation. For this, we use what we will call the 68.3-percentile deviation, which we define here in three increasingly robust ways.

The first way is to sort the absolute values of the individual deviations from the measure of central tendency (either the median or the mode), and then to simply take the 68.3-percentile value from this sorted distribution. This is analogous to the “median absolute deviation” measure of sample deviation, but with the 68.3-percentile value instead of the 50-percentile value (which we do to remain analogous to the standard deviation, in the limit of a Gaussian distribution).

This technique works well as long as less than 40% – 70% of the measurements are contaminated (see §3.1 and §3.2). However, sometimes a greater fraction of the sample may be contaminated. In this case, we model the 68.3-percentile deviation from the lower-deviation measurements.

Consider the case of N measurements, distributed normally and sorted by the absolute value of their deviations from μ (equal to either the median or the mode). If weighted uniformly (however, see §6), the percentile of the i th element is given by:

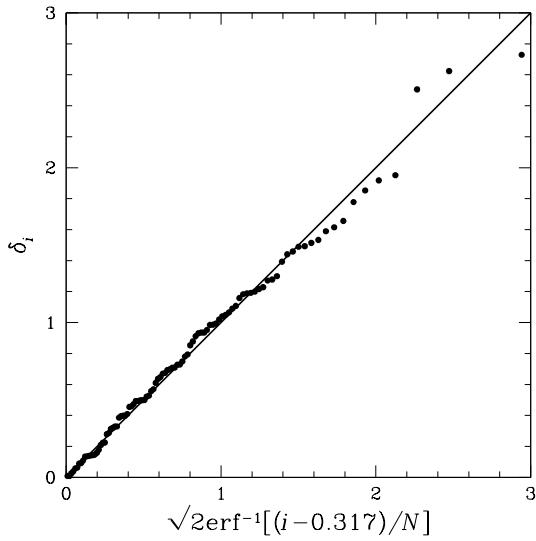


FIG. 1.— 100 sorted deviations from the median, all drawn from a Gaussian distribution of standard deviation $\sigma = 1$. The measured 68.3-percentile deviation is also ≈ 1 .

$$\frac{i - 1 + \Delta i}{N} = P\left(< \left|\frac{\delta_i}{\sigma}\right|\right), \quad (3)$$

where $P(<|\delta_i/\sigma|)$ is the cumulative probability of being within $|\delta_i/\sigma|$ standard deviations of the mean, δ_i is the i th sorted deviation, σ is the 68.3-percentile deviation, and $0 < \Delta i < 1$ is the bin center. We set $\Delta i = 0.683$ to yield intuitive results in the limit that $N \rightarrow 1$ and μ is known a priori (§6). Solving for δ_i yields:

$$\delta_i = \sigma \left[\sqrt{2} \operatorname{erf}^{-1} \left(\frac{i - 0.317}{N} \right) \right]. \quad (4)$$

Consequently, if plotted δ_i vs. $\sqrt{2} \operatorname{erf}^{-1}[(i - 0.317)/N]$, the distribution is linear, and the slope of this line yields σ (see Figure 1).

However, if a fraction of the sample is contaminated, the shape of the distribution changes: The slope steepens, and (1) if the value from which the deviations are measured (the median or the mode) still approximates that of the uncontaminated measurements, and (2) if the contaminants are drawn from a sufficiently broader distribution, the curve breaks upward (see Figure 2, upper left).² Consequently, we model the 68.3-percentile deviation of the uncontaminated measurements in three, increasingly accurate ways: (1) by simply using the 68.3-percentile value, as described above (e.g., Figure 2, upper right); (2) by fitting a zero-intercept line to the $\sqrt{2} \operatorname{erf}^{-1}[(i - 0.317)/N] < \sqrt{2} \operatorname{erf}^{-1}(0.683) = 1$ data, and using the fitted slope (e.g., Figure 2, lower left); and (3) by fitting a broken line of intercept zero (see Appendix A for fitting details) to the same data, and using the fitted slope of the first component (e.g., Figure 2, lower right).

² If the median or the mode no longer approximates that of the uncontaminated measurements, the curve can instead break downward, making the following three 68.3-percentile deviation measurement techniques decreasingly robust, instead of increasingly robust (see §3.2, Figure 17).

We then iteratively Chauvenet-reject the greatest outlier (§1), using either (1) the median or (2) the mode instead of the mean, and the 68.3-percentile deviation instead of the standard deviation.³ The effect of this on the data presented in Figure 2 can be seen in Figure 3, for each of our three, increasingly robust, 68.3-percentile deviation measurement techniques.

2.3. Calibration

Before further using these two more-robust measures of central tendency (§2.1) and three more-robust measures of sample deviation (§2.2) to Chauvenet-reject outliers, we calibrate these $2 \times 3 = 6$ more-robust techniques, using uncontaminated data. We also calibrate two less-robust, comparison techniques, using the mean and standard deviation (1) without and (2) with iterated Chauvenet rejection.

For each sample size $2 \leq N \leq 100$, as well as for $N = 1000$, we drew 100,000 samples from a Gaussian distribution of mean $\mu = 0$ and standard deviation $\sigma = 1$, and then recovered μ and σ using each technique. Averaged over the 100,000 samples, the recovered value of μ was always ≈ 0 , and the recovered value of σ was ≈ 1 in the limit of large N . However, all of the techniques, including the traditional, comparison techniques,⁴ underestimated σ in the limit of small N (see Figure 4).

In Figure 4, we plot correction factors by which measured standard and 68.3-percentile deviations need to be multiplied to yield the correct result, on average. We make use of these correction factors throughout this paper, to avoid overaggressive rejection (although this can still happen in sufficiently small samples; see §3.3.1).

3. ROBUST TECHNIQUES APPLIED TO CONTAMINATED DISTRIBUTIONS

We now evaluate the effectiveness (1) of the two traditional, less-robust techniques, and (2) of the $2 \times 3 = 6$ more-robust techniques, that we introduced in §2 at rejecting outliers from Gaussian (see §3.1 and §3.2) and mildly non-Gaussian (see §3.3) distributions. In §3.1, we consider the case of two-sided contaminants, meaning that outliers are as likely to be high as they are to be low. In §3.2, we consider the (more challenging) case of one-sided contaminants, where all or almost all of the outliers are high (or low); we also consider in-between cases here.

³ With the following exception: We never reject down to a sample of identical measurements. In the standard case of producing a single measurement from multiple, we always leave at least two distinct measurements. In the more general case of fitting a multiple-parameter model to multiple measurements (see §8), we always leave at least $M + 1$ distinct measurements, where M is the number of model parameters.

⁴ It is well known that although the variance can be computed without bias using Bessel's correction, the standard deviation cannot, and the correction depends on the shape of the distribution. For a normal distribution, without rejection of outliers, the correction is given by $\sqrt{\frac{N-1}{2}} \frac{\Gamma(\frac{N-1}{2})}{\Gamma(\frac{N}{2})}$, which matches what we determined empirically, and plot in the upper-left panel of Figure 4 (solid black curve).

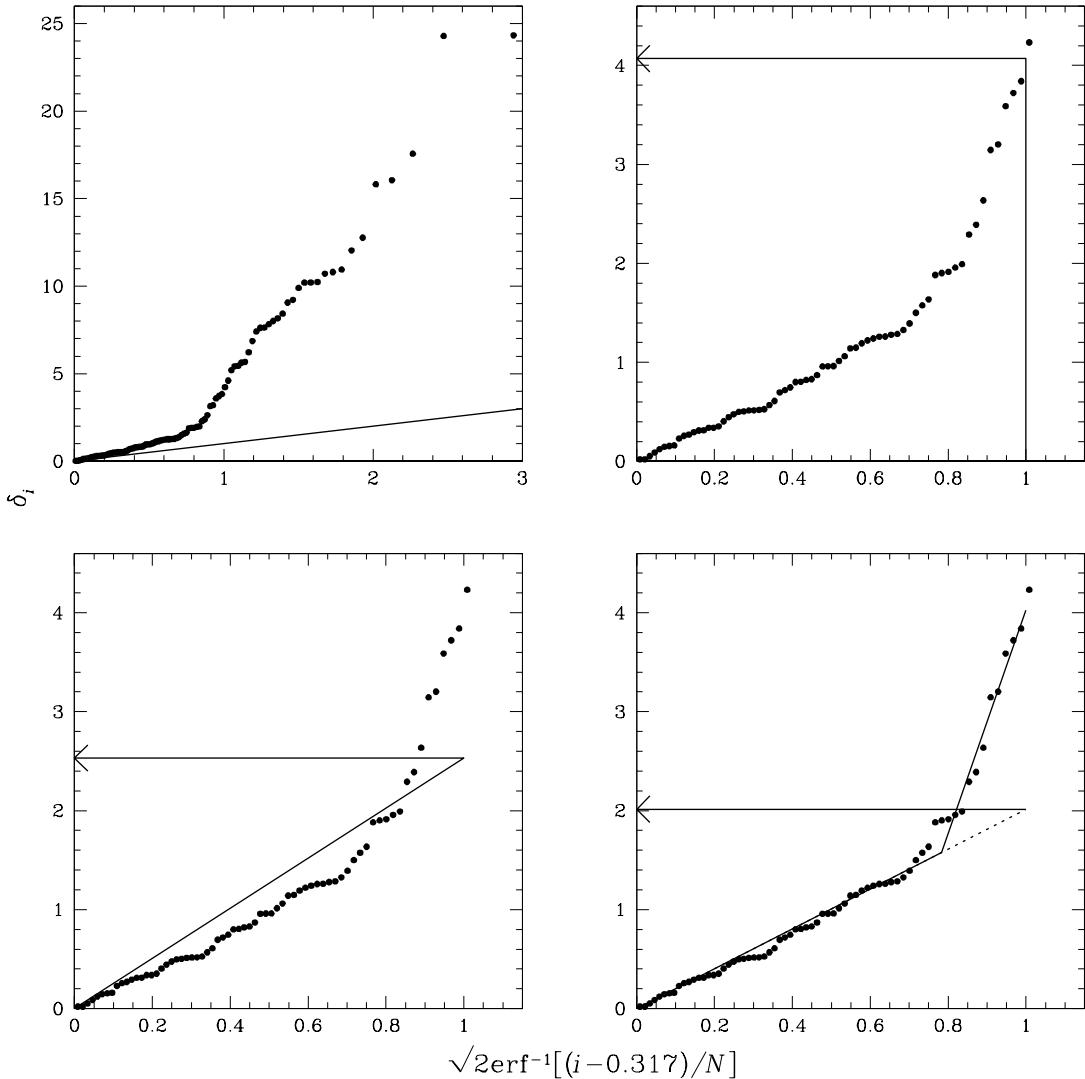


FIG. 2.— **Upper left:** 100 sorted deviations from the median, with fraction $f_1 = 0.5$ drawn from a Gaussian distribution of standard deviation $\sigma_1 = 1$, and fraction $f_2 = 0.5$, representing contaminated measurements, drawn from a Gaussian distribution of standard deviation $\sigma_2 = 10$. **Upper right:** Zoom-in of the upper-left panel, with the 68.3-percentile deviation measured using technique 1, yielding a pre-rejection value of $\sigma_1 = 4.07$. **Lower left:** Zoom-in of the upper-left panel, with the 68.3-percentile deviation measured using technique 2, yielding a pre-rejection value of $\sigma_1 = 2.53$. **Lower right:** Zoom-in of the upper-left panel, with the 68.3-percentile deviation measured using technique 3, yielding a pre-rejection value of $\sigma_1 = 2.01$. See Figure 3 for post-rejection versions and measured values.

3.1. Normally Distributed Uncontaminated Measurements with Two-Sided Contaminants

For sample sizes $N = 1000$, 100 , and 10 , we draw $f_1 N$ uncontaminated measurements from a Gaussian distribution of mean $\mu_1 = 0$ and standard deviation $\sigma_1 = 1$, and $f_2 N$ contaminated measurements, where $f_2 = 1 - f_1$. In this section, we model the contaminants as two-sided, meaning that outliers are as likely to be high as they are to be low. We draw contaminants from a Gaussian distribution of mean $\mu_2 = 0$ and standard deviation σ_2 , and add them to uncontaminated measurements, drawn as above.⁵ In the case of two-sided contaminants, the

⁵ In this paper, we draw our contaminants from Gaussian distributions, which ensures that we include some of the worst-case scenarios for rejecting outliers (from Gaussian uncontaminated-

mean, median, and mode are all three, on average, insensitive to outliers, even in the limit of a large fraction of the sample being contaminated ($f_2 \rightarrow 1$; see Figure 6).

measurement distributions) in our analyses. For example, with two-sided contaminants, the contaminated measurements are then also distributed as a Gaussian, of standard deviation $\sqrt{\sigma_1^2 + \sigma_2^2}$, which becomes increasingly difficult to distinguish from the uncontaminated-measurement distribution as $\sigma_2 \rightarrow \sigma_1$ (and of course as $\sigma_2 \rightarrow 0$). Furthermore, this contaminated-measurement distribution becomes increasingly difficult to distinguish from any, Gaussian uncontaminated-measurement distribution as $f_2 \rightarrow 1$. We have also experimented with non-Gaussian contaminant distributions, but always to similar, or greater, effect: While these outlier-rejection techniques do depend on the assumed shape of the uncontaminated-measurement distribution, they do not depend on the assumed shape of the contaminant distribution (other than strongly contaminated measurements are of course easier to identify than weakly contaminated measurements).

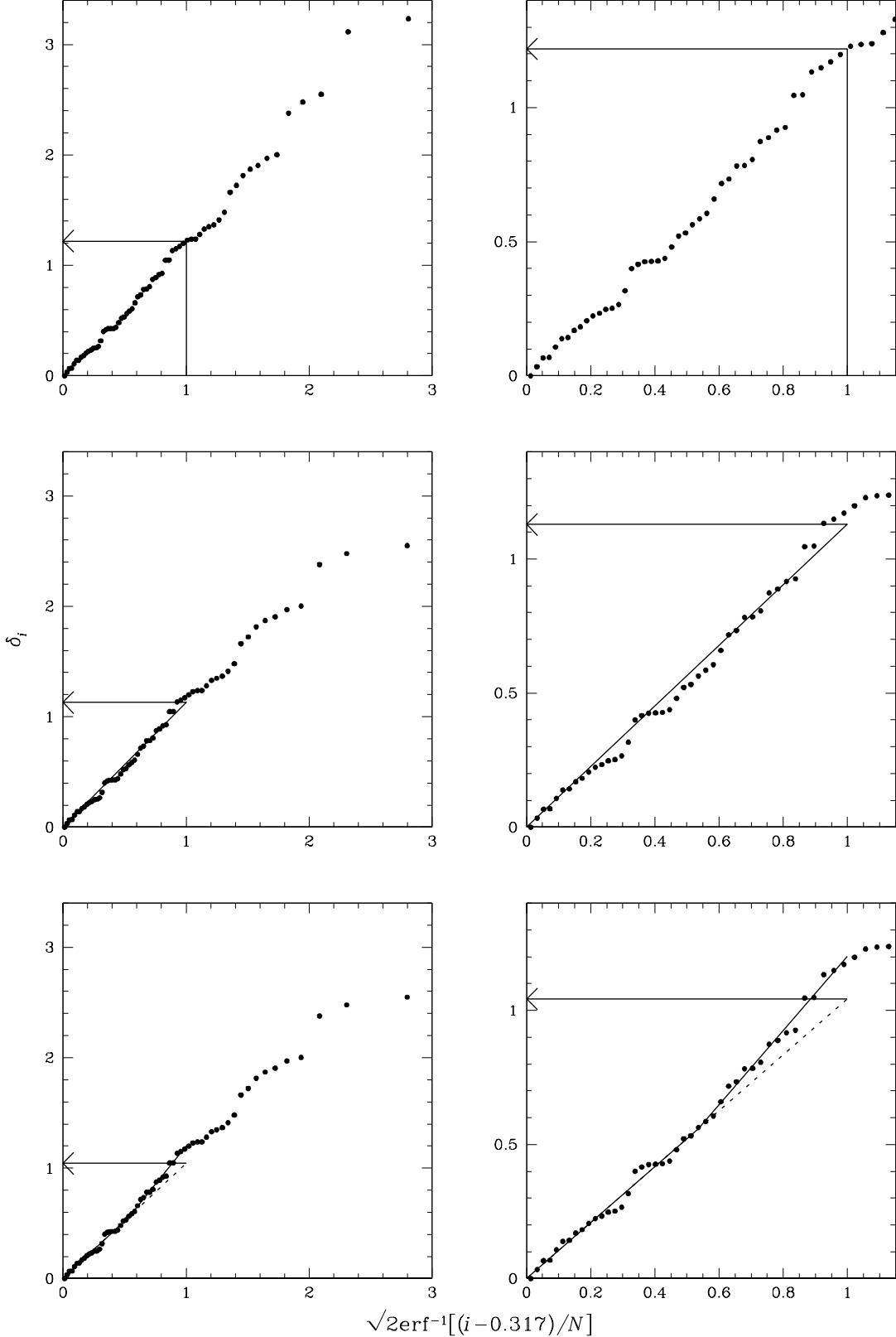


FIG. 3.— Figure 2, after iterated Chauvenet rejection. **Upper left:** Using the 68.3-percentile deviation from technique 1, yielding a final measured value of $\sigma_1 = 1.22$. **Upper right:** Zoom-in of the upper-left panel. **Middle left:** Using the 68.3-percentile deviation from technique 2, yielding a final measured value of $\sigma_1 = 1.13$. **Middle right:** Zoom-in of the middle-left panel. **Lower left:** Using the 68.3-percentile deviation from technique 3, yielding a final measured value of $\sigma_1 = 1.04$. **Lower right:** Zoom-in of the lower-left panel.

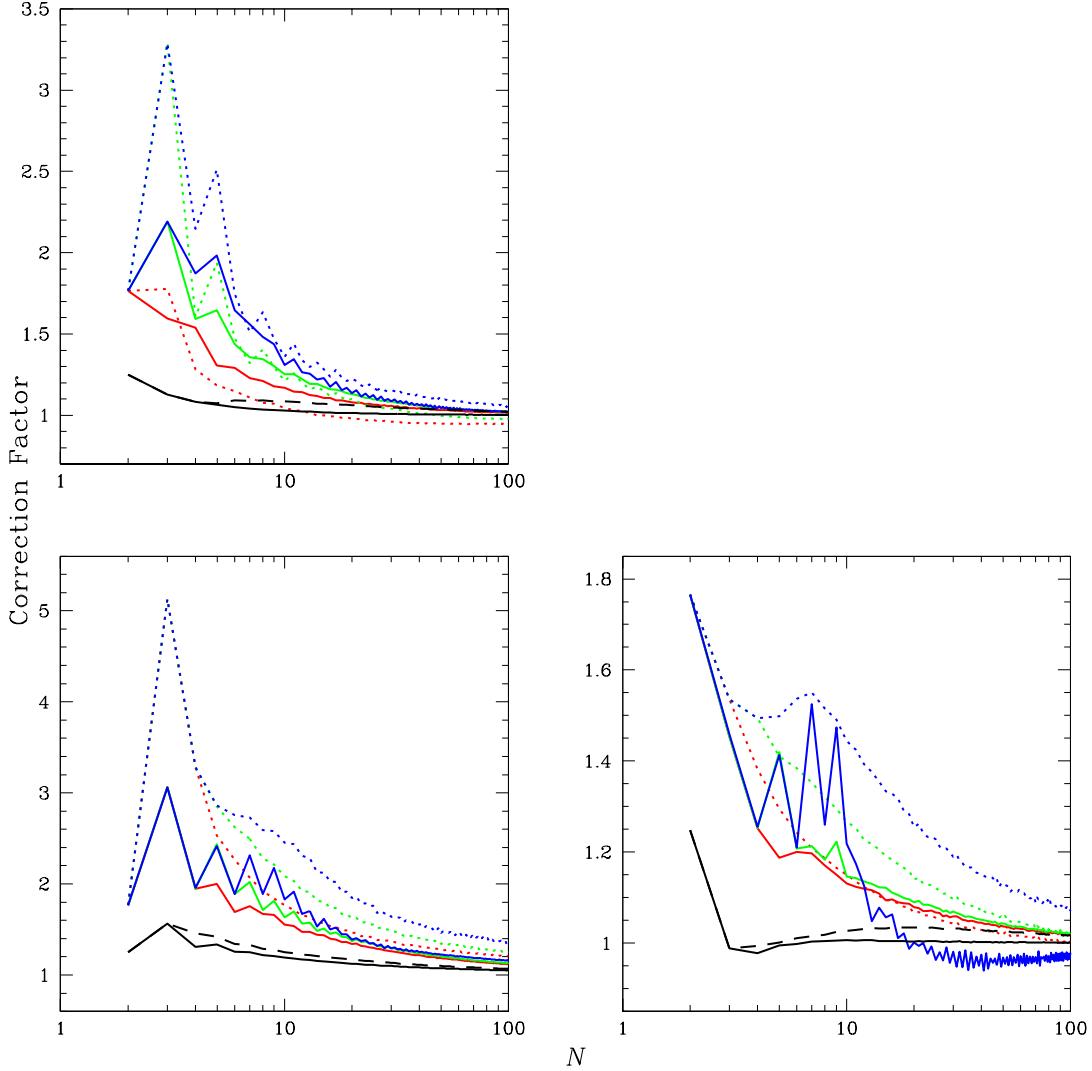


FIG. 4.— Correction factors by which standard and 68.3-percentile deviations, measured from uncontaminated data, need to be multiplied to yield the correct result, on average, and to avoid overaggressive rejection (although this can still happen in sufficiently small samples; see §3.3.1), (1) for the case of no rejection, using the mean and standard deviation (solid black curves; see Footnote 3); (2) for the case of Chauvenet rejection, using the mean and standard deviation (dashed black curves); (3) for the case of Chauvenet rejection, using the median and 68.3-percentile deviation as measured by technique 1 from §2.2 (solid red curves), as measured by technique 2 from §2.2 (solid green curves), and as measured by technique 3 from §2.2 (solid blue curves); and (4) for the case of Chauvenet rejection, using the mode and 68.3-percentile deviation as measured by technique 1 (dotted red curves), technique 2 (dotted green curves), and technique 3 (dotted blue curves). **Upper left:** For the simplest case of computing a single σ (standard or 68.3-percentile deviation), using the deviations both below and above μ (the mean, the median, or the mode; see §3.1). **Lower left:** For the case of computing separate σ below and above μ (σ_- and σ_+ , respectively) and using the smaller of the two when rejecting outliers (see §3.2). **Lower right:** For the same case, but using σ_- to reject outliers below μ and σ_+ to reject outliers above μ (see §3.1). Note that technique 3 defaults to technique 2 when the two are statistically equivalent (see Appendix A), or when fitting to fewer than three points (e.g., when $N < 4$ for the cases in the top row and when $N < 7$ for the median cases in the bottom row). Similarly, technique 2 defaults to technique 1 when fitting to fewer than two points (e.g., when $N < 3$ for the cases in the top row and when $N < 5$ for the median cases in the bottom row). Oscillations are not noise, but odd-even effects (e.g., with equally weighted data, when N is odd, use of the median always results in at least one zero deviation, requiring a larger correction factor). We use look-up tables for $N \leq 100$ and power-law approximations for $N > 100$ (see Appendix B).

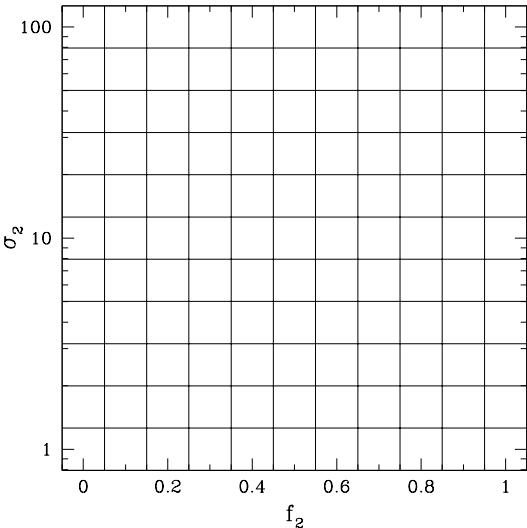


FIG. 5.— Blank contaminant strength (σ^2) vs. fraction of sample (f_2) figure. Each pixel corresponds to either a recovered quantity (μ_1 or σ_1) or the uncertainty in a recovered quantity ($\Delta\mu_1$ or $\Delta\sigma_1$), measured from 100 samples with contaminants modeled by f_2 and σ_2 . This figure is provided as reference, as axis information would be too small to be easily readable in upcoming figures.

Consequently, this is a good case to evaluate the effectiveness of our three, increasingly robust, 68.3-percentile deviation techniques. (We explore the more challenging case of one-sided contaminants in §3.2.)

For each technique and sample size, we draw 100 samples for each combination of $f_2 = 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$ and $\sigma_2 = 1, 1.6, 2.5, 4.0, 6.3, 10, 16, 25, 40, 63, 100$ (see Figure 5), and plot the average recovered μ_1 in Figure 6, the uncertainty in the recovered μ_1 in Figure 7, the average recovered σ_1 in Figure 8, and the uncertainty in the recovered σ_1 in Figure 9.

As expected with two-sided contaminants, the average recovered μ_1 is always ≈ 0 . However, the uncertainty in the recovered μ_1 , the average recovered σ_1 , and the uncertainty in the recovered σ_1 are all susceptible to contamination, especially when f_2 and σ_2 are large. However, our increasingly robust 68.3-percentile deviation measurement techniques are increasingly effective at rejecting outliers in large- f_2 samples, allowing σ_1 to be measured significantly more accurately, and both μ_1 and σ_1 to be measured significantly more precisely. Note that this is at a marginal cost: When applied to uncontaminated samples, our increasingly robust measurement techniques recover μ_1 and σ_1 with degrading precisions (Figures 7 and 9). This suggests that one can reach a point of diminishing returns; however, this is a drawback that we largely eliminate in §4. Given this, when Chauvenet-rejecting two-sided contaminants, we recommend using (1) the median (because it is just as accurate as the mode (in this case), more precise, and computationally faster) and (2) the 68.3-percentile deviation as measured by technique 3 from §2.2 (the broken-line fit). This technique is highlighted in Figures 6–9 with a bold outline.

3.2. Normally Distributed Uncontaminated Measurements with One-Sided Contaminants

We now repeat the analysis of §3.1, but for the more challenging case of one-sided contaminants, which we model by drawing values from only the positive side of a Gaussian distribution of mean $\mu_2 = 0$ and standard deviation σ_2 . This case is more challenging because even though the median is more robust than the mean, and the mode is more robust than the median, even the mode will be biased in the direction of the contaminants (see Figure 10), and increasingly so as the fraction of the sample that is contaminated increases (see Figures 11 and 12).

Furthermore, as μ (equal to the mean, the median, or the mode) becomes more biased in the direction of the one-sided contaminants, σ (equal to the standard deviation or the 68.3-percentile deviation, as measured by any of the techniques presented in §2.2) becomes more biased as well, (1) because of the contaminants, and (2) because it is measured from μ . However, σ can be measured with less bias, if measured using only the deviations from μ that are in the opposite direction as the contaminants (in this case, the deviations below μ ; Figure 11). Since the direction of the contaminants might not be known a priori, or since the contaminants might not be fully one-sided, instead being between the cases presented in §3.1 and §3.2, we measure σ both below and above μ ,⁶ and use the smaller of these two measurements when rejecting outliers (Figure 12). Note, using the smaller of these two measurements should only be done if the uncontaminated measurements are symmetrically distributed (see §3.3.1).

For the same techniques presented in §3.1, except now computing σ both below and above μ and adopting the smaller of the two, and for the same sample sizes presented in §3.1, we plot the average recovered μ_1 in Figure 13, the uncertainty in the recovered μ_1 in Figure 14, the average recovered σ_1 in Figure 15, and the uncertainty in the recovered σ_1 in Figure 16.

With one-sided contaminants, all four of these are susceptible to contamination, especially when f_2 and σ_2 are large. However, for a fixed σ -measurement technique, our increasingly robust μ -measurement techniques are increasingly effective at rejecting outliers in large- f_2 samples, allowing μ_1 and σ_1 to be measured both significantly more accurately and significantly more precisely. However, when μ_1 cannot be measured accurately, as is the case with the mean and the median when f_2 is large (Figures 10, 11, and 12), our (otherwise) increasingly robust σ -measurement techniques are *decreasingly* effective at rejecting outliers (see Figure 17). However, the mode can measure μ_1 significantly more accurately (Figures 10, 11, and 12), even when f_2 is large, though with decreasing effectiveness in the low- N limit. In any case, when μ_1 is measured accurately, all of these techniques are nearly equally effective, because σ_1 is measured on the nearly uncontaminated side of each sample's distribution. Given this, when Chauvenet-rejecting one-sided contaminants, we recommend using (1) the mode, and (2) the 68.3-percentile deviation as measured by technique 1 from §2.2 (the 68.3% value, because it is essen-

⁶ When computing σ below or above μ , if a measurement equals μ , we include it in both the below and above calculations, but with 50% weight for each (see §6).

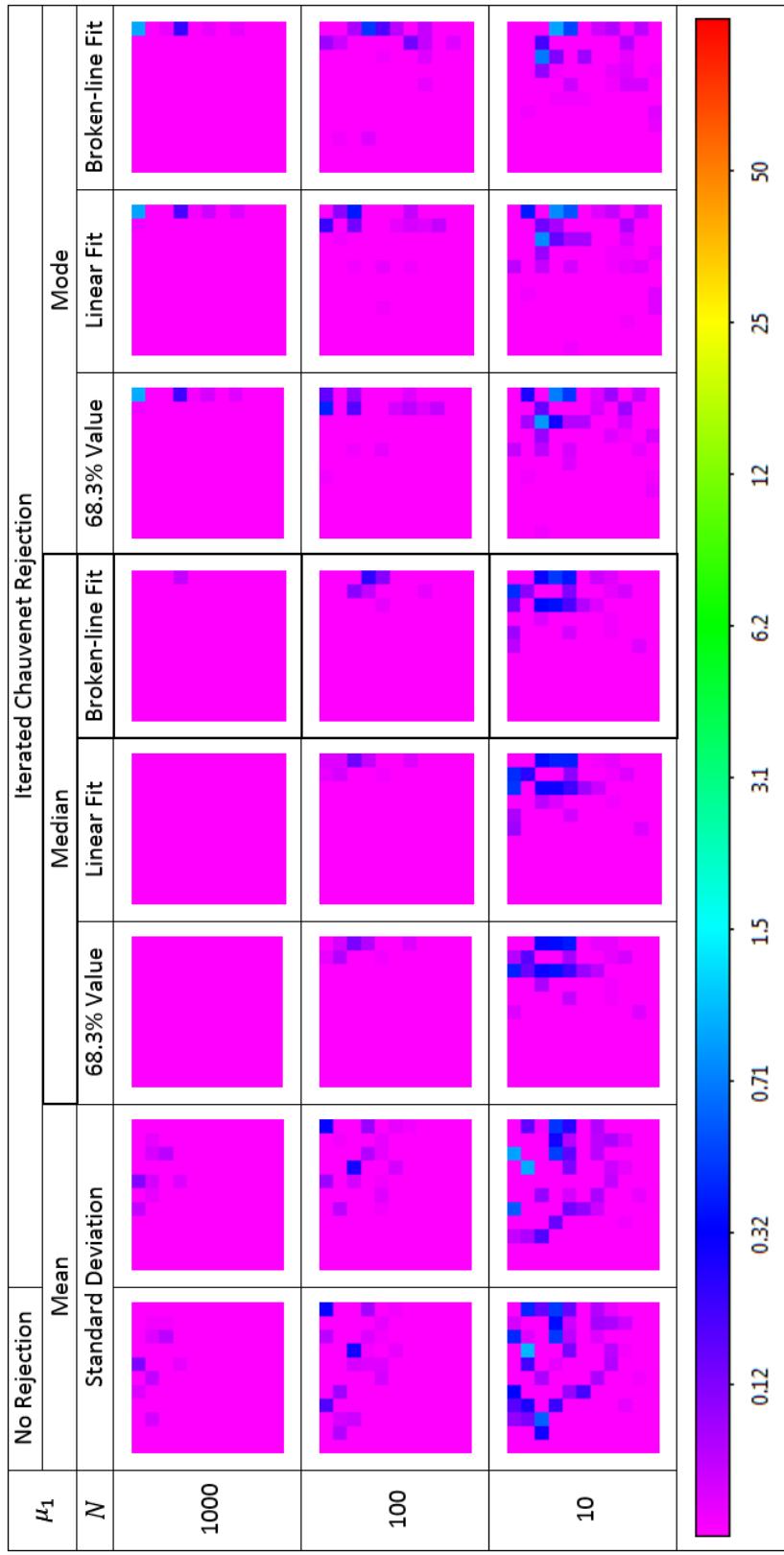


Fig. 6.— Average recovered μ_1 for increasingly robust measurement techniques and decreasing sample sizes (N), for two-sided contaminants. See Figure 5 for contaminant strength (σ_2) vs. fraction of sample (f_2) axis information. As expected with two-sided contaminants, the recovered values are ≈ 0 , independent of contaminant fraction or strength. Variation about zero is due to drawing only 100 samples, and is larger for larger values of f_2 and σ_2 , and for smaller values of N (see Figure 7). All things considered (Figures 6 – 9; §3.1), the best-performing technique for Chauvenet-rejecting two-sided contaminants is highlighted with a bold outline. The colors are scaled logarithmically, and cut off at 0.02, to match the upcoming figures, permitting direct comparison of colors between figures.

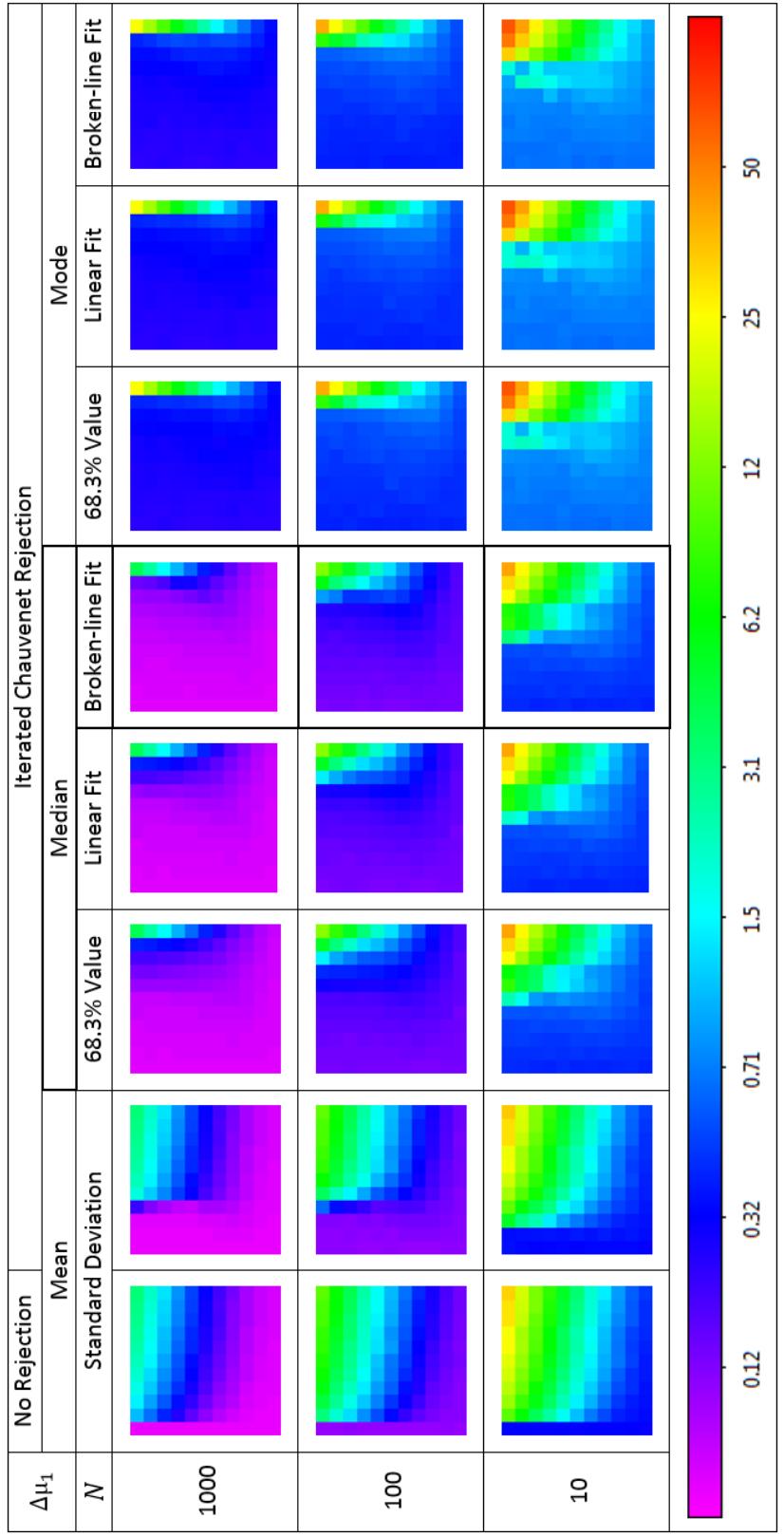


FIG. 7.— Uncertainty in the recovered μ_1 for increasingly robust measurement techniques and decreasing sample sizes (N), for two-sided contaminants. See Figure 5 for contaminant strength (σ_2) vs. fraction of sample (f_2) axis information. The effect of the contaminants, without rejection, can be seen in the first column: Larger contaminant fractions and strengths, as well as smaller sample sizes, result in less precise recovered values of μ_1 . However, increasingly robust measurement techniques are increasingly effective at rejecting outliers in large- f_2 samples, allowing μ_1 to be measured significantly more precisely. Note that this is at a marginal cost: When applied to uncontaminated samples ($f_2 = 0$), these techniques recover μ_1 with degrading precisions, of $\Delta\mu_1/\sigma_1 \approx 1.0N^{-1/2}$, $1.0N^{-1/2}$, $1.3N^{-1/2}$, $1.3N^{-1/2}$, $7.6N^{-1/2}$, $7.6N^{-1/2}$, and $7.6N^{-1/2}$, respectively. However, all things considered (Figures 6 – 9; §3.1), the best-performing technique for Chauvenet-rejecting two-sided contaminants is highlighted with a bold outline. The colors are scaled logarithmically, between 0.02 and 100.

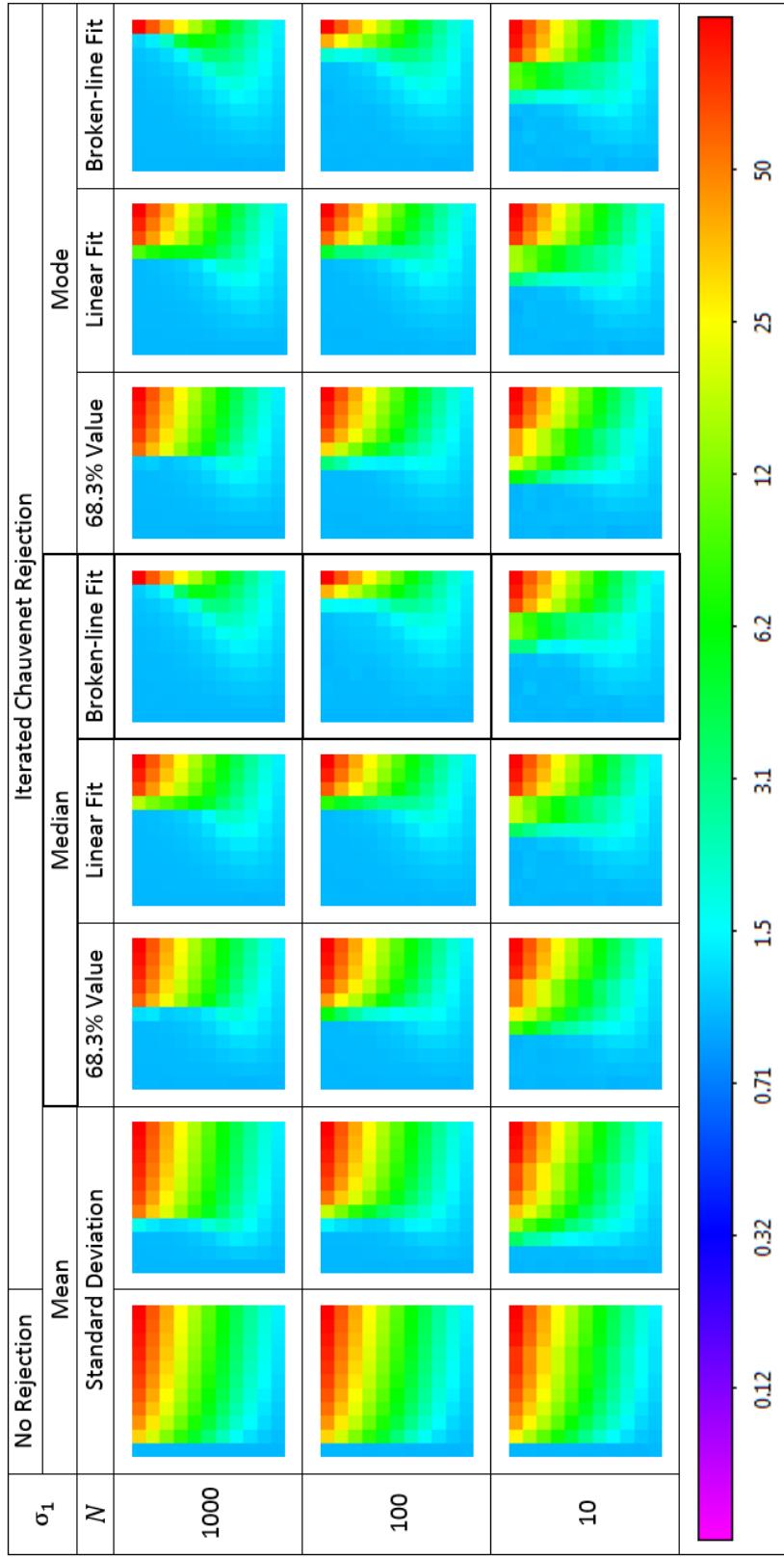


Fig. 8.— Average recovered σ_1 for increasingly robust measurement techniques and decreasing sample sizes (N), for two-sided contaminants. See Figure 5 for contaminant strength (σ_2) vs. fraction of sample (f_2) axis information. The effect of the contaminants, without rejection, can be seen in the first column: Larger contaminant fractions and strengths result in larger recovered values of σ_1 . However, increasingly robust measurement techniques are increasingly effective at rejecting outliers in large- f_2 samples, allowing σ_1 to be measured significantly more accurately. All things considered (Figures 6 – 9; §3.1), the best-performing technique for Chauvenet-rejecting two-sided contaminants is highlighted with a bold outline. The colors are scaled logarithmically, between 0.02 and 100.

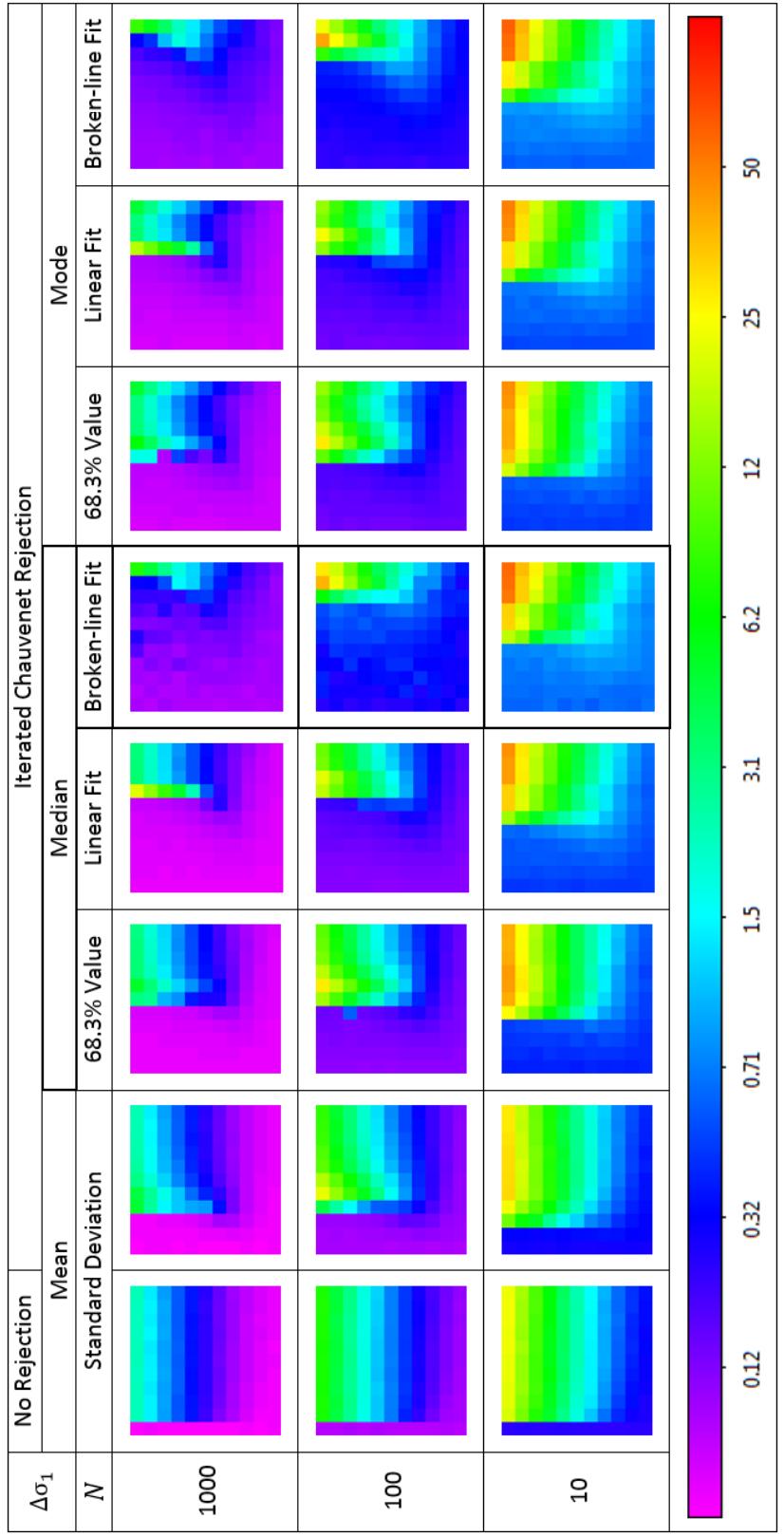


FIG. 9.— Uncertainty in the recovered σ_1 for increasingly robust measurement techniques and decreasing sample sizes (N), for two-sided contaminants. See Figure 5 for contaminant strength (σ_2) vs. fraction of sample (f_2) axis information. The effect of the contaminants, without rejection, can be seen in the first column: Larger contaminant fractions and strengths, as well as smaller sample sizes, result in less precise recovered values of σ_1 . However, increasingly robust measurement techniques are increasingly effective at rejecting outliers in large- f_2 samples, allowing σ_1 to be measured significantly more precisely. Note that this is at a marginal cost: When applied to uncontaminated samples ($f_2 = 0$), these techniques recover σ_1 with degrading precisions of $\Delta\sigma_1/\sigma_1 \approx 0.7N^{-1/2}$, $0.8N^{-1/2}$, $1.0N^{-1/2}$, $1.0N^{-1/2}$, $1.5N^{-1/2}$, $2.3N^{-1/2}$, $2.8N^{-1/2}$, respectively. However, all things considered (Figures 6 – 9; §3.1), the best-performing technique for Chauvenet-rejecting two-sided contaminants is highlighted with a bold outline. The colors are scaled logarithmically, between 0.02 and 100.

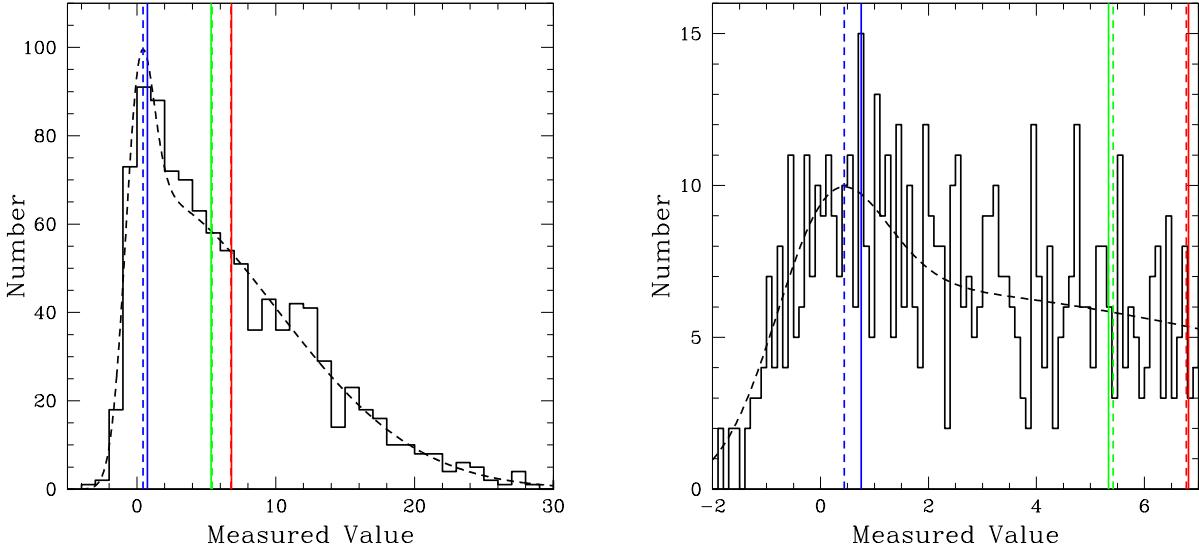


FIG. 10.— **Left:** 1000 measurements, with fraction $f_1 = 0.15$ drawn from a Gaussian distribution of mean $\mu_1 = 0$ and standard deviation $\sigma_1 = 1$, and fraction $f_2 = 0.85$, representing contaminated measurements, drawn from the positive side of a Gaussian distribution of mean $\mu_2 = 0$ and standard deviation $\sigma_2 = 10$, and added to uncontaminated measurements, drawn as above. The measurements have been binned, and the mean (solid red line), median (solid green line), and mode (solid blue line) have been marked. The dashed black curve marks the theoretical, or large-N, distribution, and for this the mean, median, and mode have also been marked, with dashed lines. **Right:** Zoom-in of the left panel, with smaller bins. A large f_2 was chosen to more clearly demonstrate that the mode is biased in the direction of the contaminants, albeit only marginally. Also, the sample mode differs from the theoretical mode more than the sample median and mean differ from the theoretical median and mean, due to “noise” peaks, caused by random sampling. This is typical, and why the mode, although significantly more accurate, is less precise.

tially as accurate as the other techniques (in this case), more precise,⁷ and computationally faster). This technique is highlighted in Figures 13 – 16 with a bold outline.

When Chauvenet-rejecting contaminants that are neither one-sided nor two-sided, but that are in-between these cases, with values that are both positive and negative, but not in equal proportion or strength, we recommend using the smaller of the below- and above-measured 68.3-percentile deviations, as in the one-sided case, but recommend using (1) the mode (which is just as effective as the median at eliminating two-sided contaminants (§3.1), but more effective at eliminating one-sided contaminants), and (2) the 68.3-percentile deviation as measured by technique 3 from §2.2 (the broken line fit, which is more effective than the other techniques at eliminating two-sided contaminants (§3.1) and essentially as effective at eliminating one-sided contaminants). This technique is highlighted in Figures 13 – 16 with a double outline.

3.3. Non-Normally Distributed Uncontaminated Measurements with Contaminants

In §3.1 and §3.2, we assumed that the uncontaminated measurements were drawn from a Gaussian distribution. Although this is often a reasonable assumption, sometimes one might need to admit the possibility of an asymmetric (see §3.3.1) or a peaked or flat-topped (see §3.3.2) distribution for the uncontaminated measurements.

⁷ As in the case of two-sided contaminants, when applied to uncontaminated samples, our increasingly robust measurement techniques recover μ_1 and σ_1 with degrading precisions (Figures 14 and 16), but again, this is a drawback that we largely eliminate in §4.

3.3.1. Asymmetric Uncontaminated Distributions

In this case, it is better to use the σ (equal to the standard deviation or the 68.3-percentile deviation, as measured by any of the techniques presented in §2.2) measured from the deviations below μ (equal to the mean, the median, or the mode) to reject outliers below μ , and the σ measured from the deviations above μ to reject outliers above μ , assuming that the distribution is only mildly non-normal, even if this means not always using the smaller of the two σ values, as can be done with normally distributed uncontaminated measurements (§3.2).

However, this weakens one’s ability to reject outliers, particularly when one-sided contaminants dominate the sample. Even if the uncontaminated measurements are not asymmetrically distributed, simply admitting the possibility can reduce one’s ability to remove contaminants, so this is a decision that should be made with care.

To demonstrate this, we repeated the analysis of §3.2, not changing the uncontaminated measurements, but changing the assumption that we made about their distribution, instead admitting the possibility of asymmetry. We then plotted the average recovered μ_1 , the uncertainty in the recovered μ_1 , the average recovered below-measured σ_{1-} , the uncertainty in the recovered σ_{1-} , the average recovered above-measured σ_{1+} , and the uncertainty in the recovered σ_{1+} , and compared these to those from §3.1 and §3.2.

As one might expect, (1) the plots for μ_1 , $\Delta\mu_1$, σ_{1-} , and $\Delta\sigma_{1-}$ resembled the one-sided contaminant results (Figures 13 – 16, respectively), and (2) the plots for σ_{1+} and $\Delta\sigma_{1+}$ resembled the two-sided contaminant results (Figures 15 and 16, respectively, but for about half as many measurements), where the latter can be less ef-

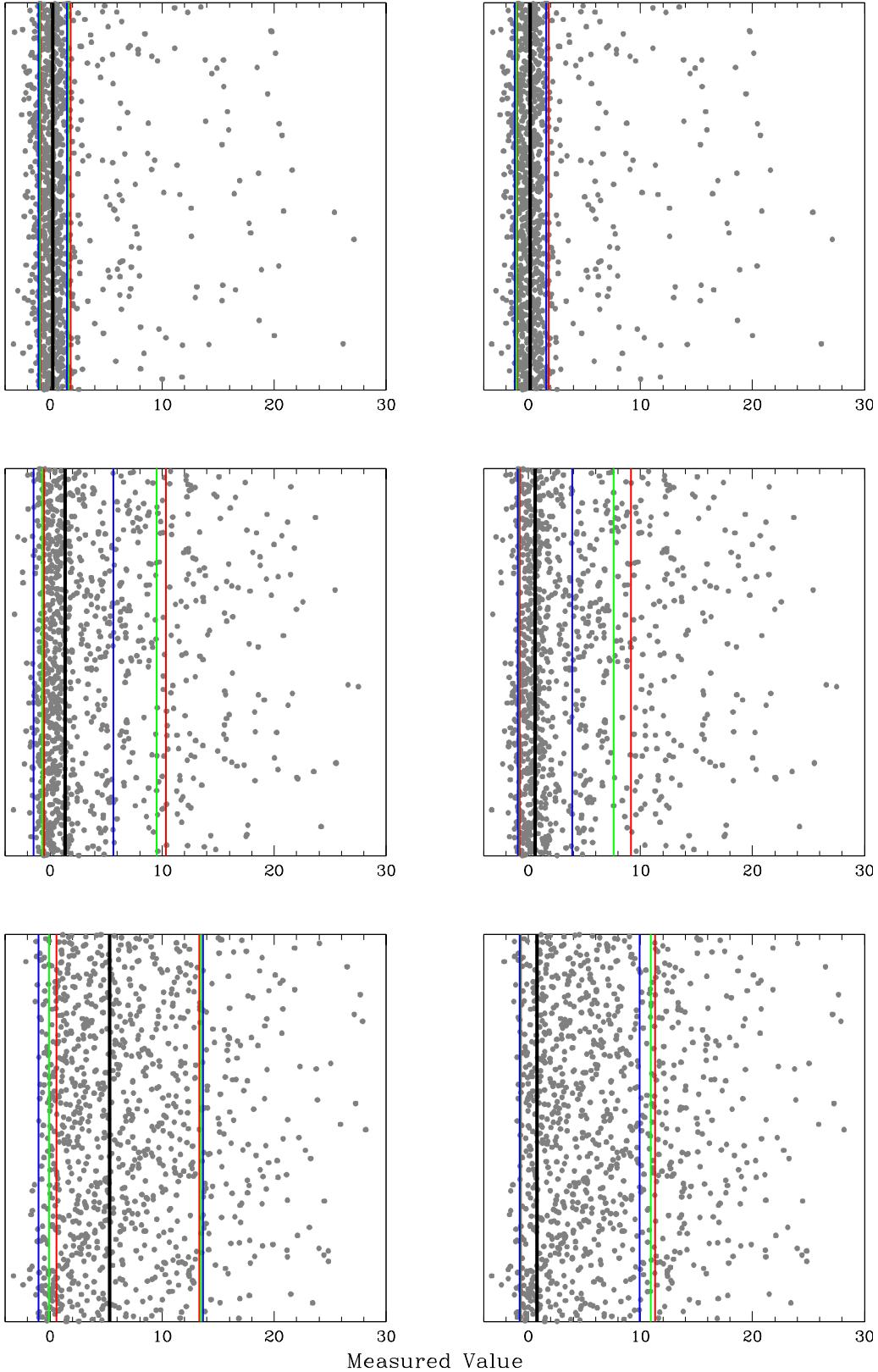


FIG. 11.— 1000 measurements, with fraction $f_1 = 1 - f_2$ drawn from a Gaussian distribution of mean $\mu_1 = 0$ and standard deviation $\sigma_1 = 1$, and fraction $f_2 = 0.15$ (top row), 0.5 (middle row), and 0.85 (bottom row), representing contaminated measurements, drawn from the positive side of a Gaussian distribution of mean $\mu_2 = 0$ and standard deviation $\sigma_2 = 10$, and added to uncontaminated measurements, drawn as above. **Left column:** Median (black line) and 68.3-percentile deviations, measured both below and above the median, using technique 1 from §2.2 (red lines), using technique 2 from §2.2 (green lines), and using technique 3 from §2.2 (blue lines). **Right column:** Same as the left column, except using the mode instead of the median. The mode performs better, especially in the limit of large f_2 . The 68.3-percentile deviation performs better when paired with the mode, and when measured in the opposite direction as the contaminants. See Figure 12 for post-rejection versions.

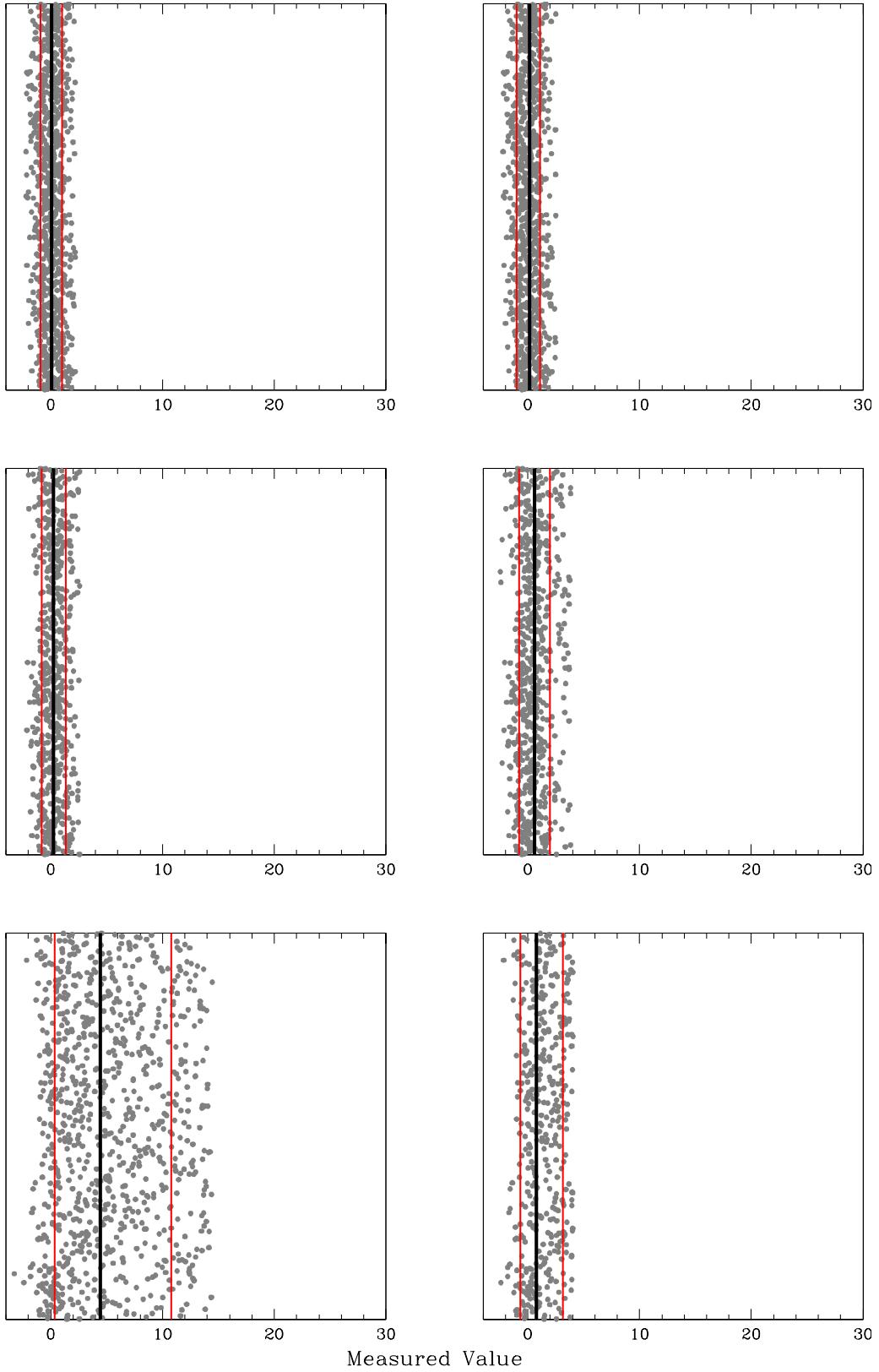


FIG. 12.— Figure 11, after iterated Chauvenet rejection, using the smaller of the below- and above-measured 68.3-percentile deviations, in this case as measured by technique 1 from §2.2. Techniques 2 and 3 from §2.2 yield similar post-rejection samples and μ and σ measurements. The mode continues to perform better in the limit of large f_2 .

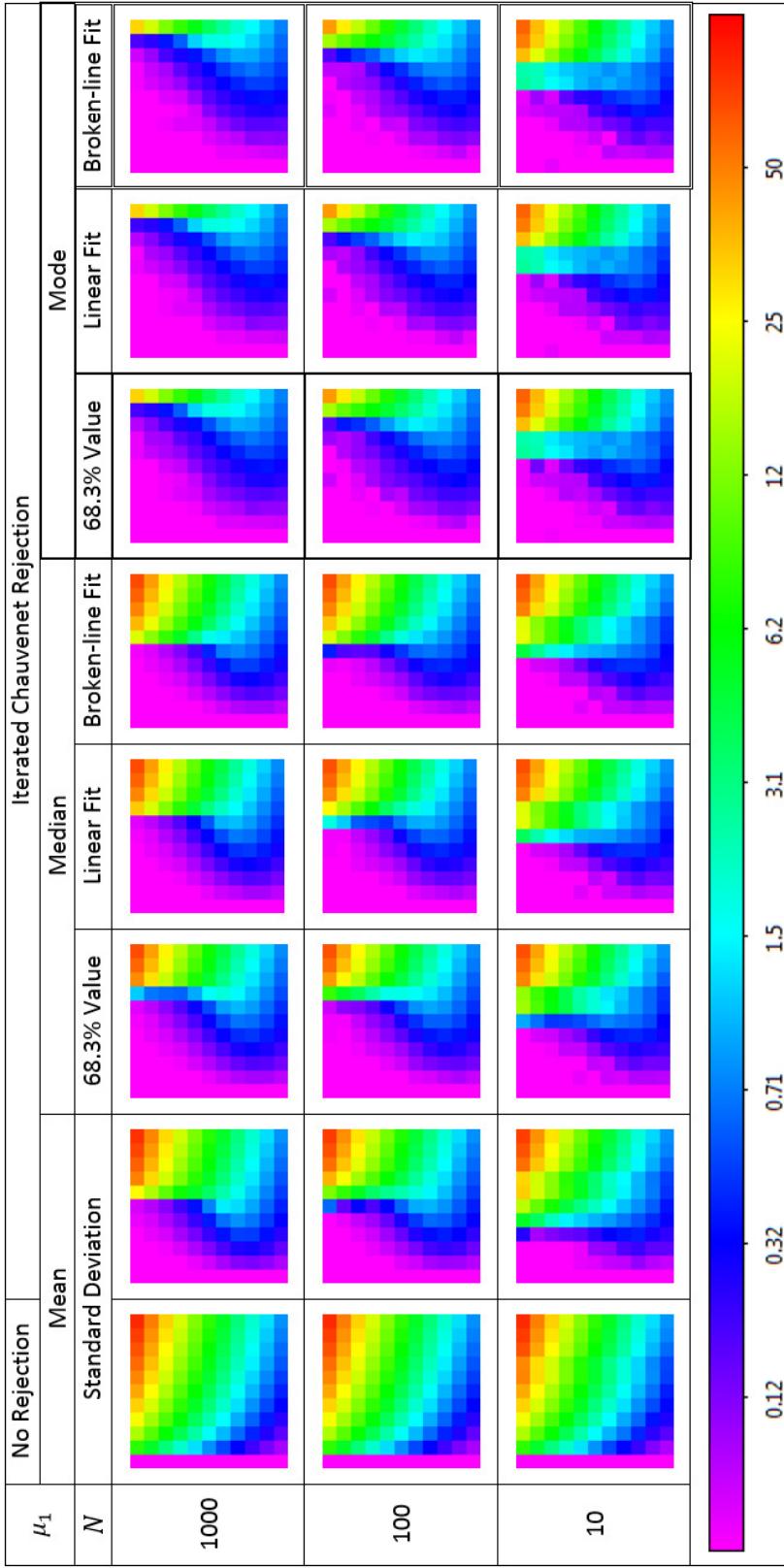


FIG. 13.— Average recovered μ_1 for increasingly robust measurement techniques and decreasing sample sizes (N), for one-sided contaminants. See Figure 5 for contaminant strength (σ_2) vs. fraction of sample (f_2) axis information. The effect of the contaminants, without rejection, can be seen in the first column. Larger contaminant fractions and strengths result in larger recovered values of μ_1 . However, for a fixed σ -measurement technique, our increasingly robust μ -measurement techniques are increasingly effective at rejecting outliers in large- f_2 samples, allowing μ_1 to be measured significantly more accurately. However, when μ_1 cannot be measured accurately, as is the case with the mean and the median when f_2 is large (Figures 10, 11, and 12), our (otherwise) increasingly robust σ -measurement techniques are *decreasingly* effective at rejecting outliers (see Figure 17). This can be seen in columns 3 – 5, which use the 68.3 percentile deviation as measured by technique 1 from §2.2, the 68.3 percentile deviation as measured by technique 2 from §2.2, and the 68.3 percentile deviation as measured by technique 3 from §2.2, respectively. However, the mode can measure μ_1 significantly more accurately (Figures 10, 11, and 12), even when f_2 is large, though with decreasing effectiveness in the low- N limit. In any case, when μ_1 is measured accurately, all of these techniques are nearly equally effective, because σ_1 is measured on the nearly uncontaminated side of each sample's distribution. All things considered (Figures 13 – 16; §3.2), the best-performing technique for Chauvenet-rejecting one-sided contaminants is highlighted with a bold outline, and the best-performing technique for Chauvenet-rejecting contaminants that are neither one-sided nor two-sided, but that are in-between these cases, is highlighted with a double outline. The colors are scaled logarithmically, between 0.02 and 100.

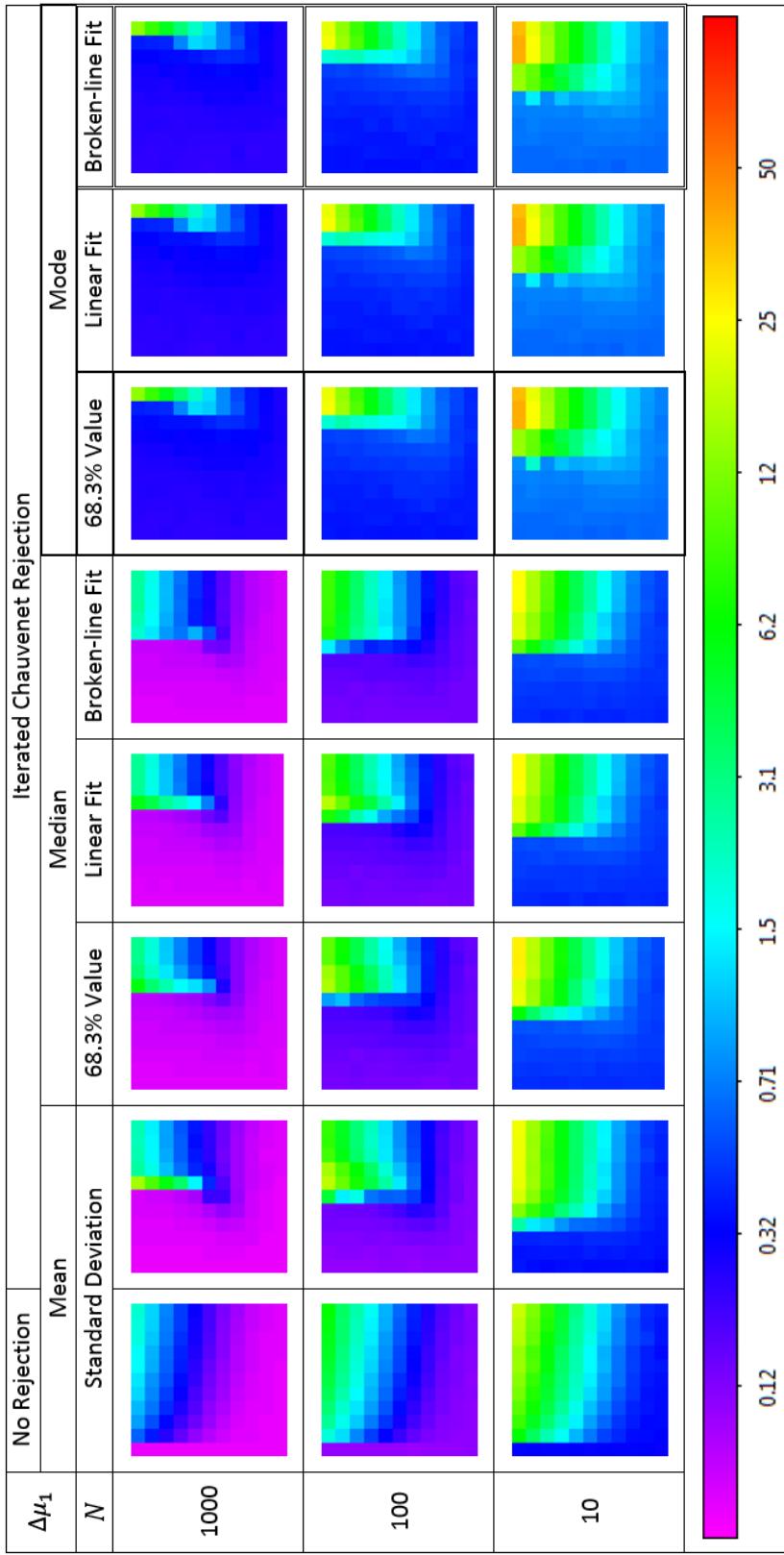


FIG. 14.— Uncertainty in the recovered μ_1 for increasingly robust measurement techniques and decreasing sample sizes (N), for one-sided contaminants. See Figure 5 for contaminant strength (σ_2) vs. fraction of sample (f_2) axis information. The effect of the contaminants, without rejection, can be seen in the first column: Larger contaminant fractions and strengths, as well as smaller sample sizes, result in less precise recovered values of μ_1 . However – to the degree that μ_1 can be measured accurately (Figure 13) – all of our Chauvenet rejection techniques are effective at removing outliers (and nearly equally so, since σ_1 is measured on the nearly uncontaminated side of each sample's distribution), allowing μ_1 to be measured significantly more precisely. Note that, as in the case of two-sided contaminants (Figure 7), when applied to uncontaminated samples ($f_2 = 0$), these techniques recover μ_1 with degrading precisions, of $\Delta\mu_1/\sigma_1 \approx 1.0N^{-1/2}$, $1.0N^{-1/2}$, $1.3N^{-1/2}$, $1.3N^{-1/2}$, $7.4N^{-1/2}$, and $7.4N^{-1/2}$, respectively. However, all things considered (Figures 13 – 16; §3.2), the best-performing technique for Chauvenet-rejecting one-sided contaminants is highlighted with a bold outline, and the best-performing technique for Chauvenet-rejecting contaminants that are neither one-sided nor two-sided, but that are in-between these cases, is highlighted with a double outline. The colors are scaled logarithmically, between 0.02 and 100.

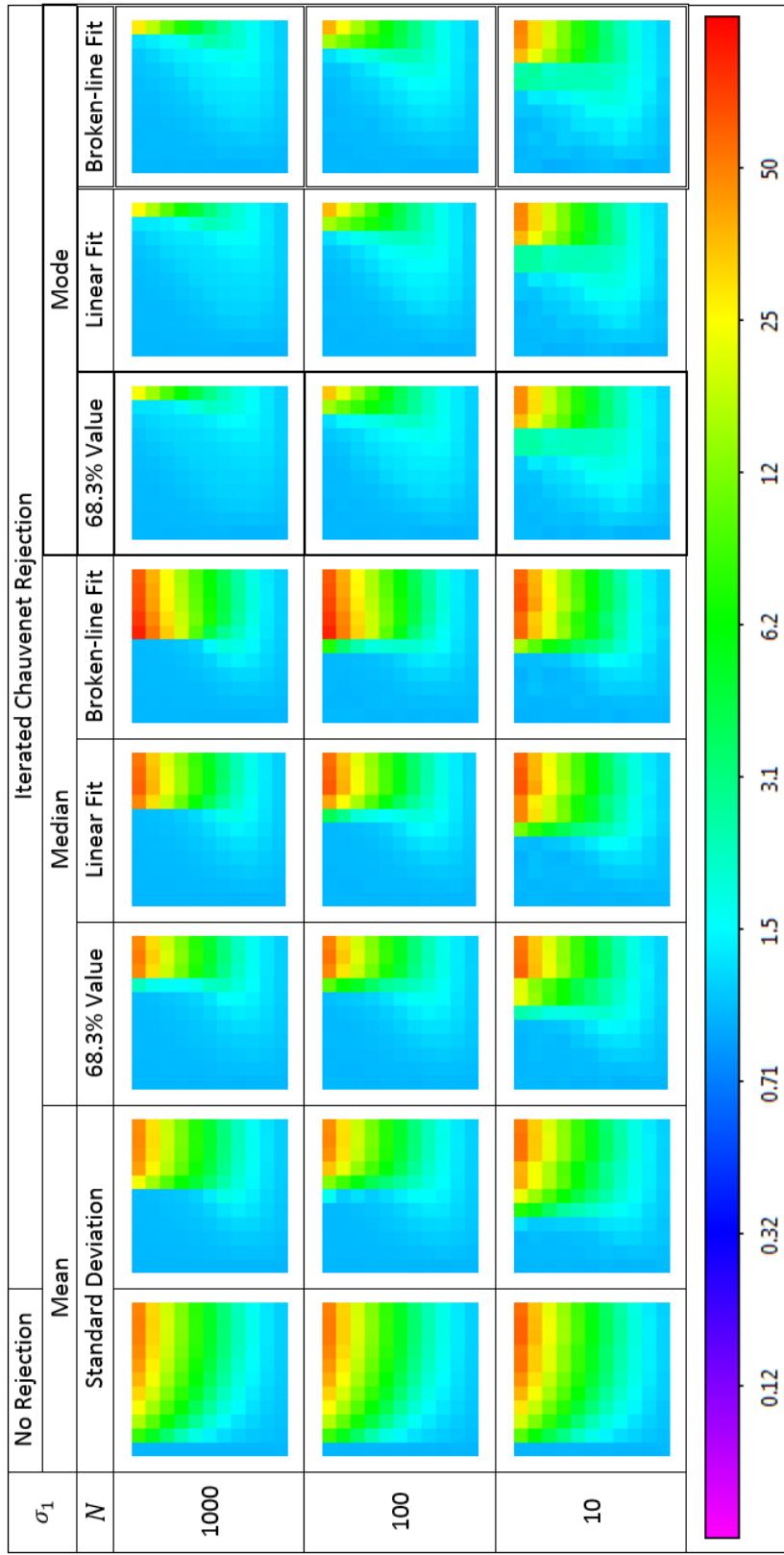


FIG. 15.— Average recovered σ_1 for increasingly robust measurement techniques and decreasing sample sizes (N), for one-sided contaminants. See Figure 5 for contaminant strength (σ_2) vs. fraction of sample (f_2) axis information. The effect of the contaminants, without rejection, can be seen in the first column: Larger contaminant fractions and strengths generally result in larger recovered values of σ_1 . However – to the degree that μ_1 can be measured accurately (Figure 13) – all of our Chauvenet rejection techniques are effective at removing outliers (and nearly equally so, since σ_1 is measured on the nearly uncontaminated side of each sample's distribution), allowing σ_1 to be measured significantly more accurately. All things considered (Figures 13 – 16; §3.2), the best-performing technique for Chauvenet-rejecting one-sided contaminants is highlighted with a bold outline, and the best-performing technique for Chauvenet-rejecting contaminants that are neither one-sided nor two-sided, but that are in-between these cases, is highlighted with a double outline. The colors are scaled logarithmically, between 0.02 and 100.

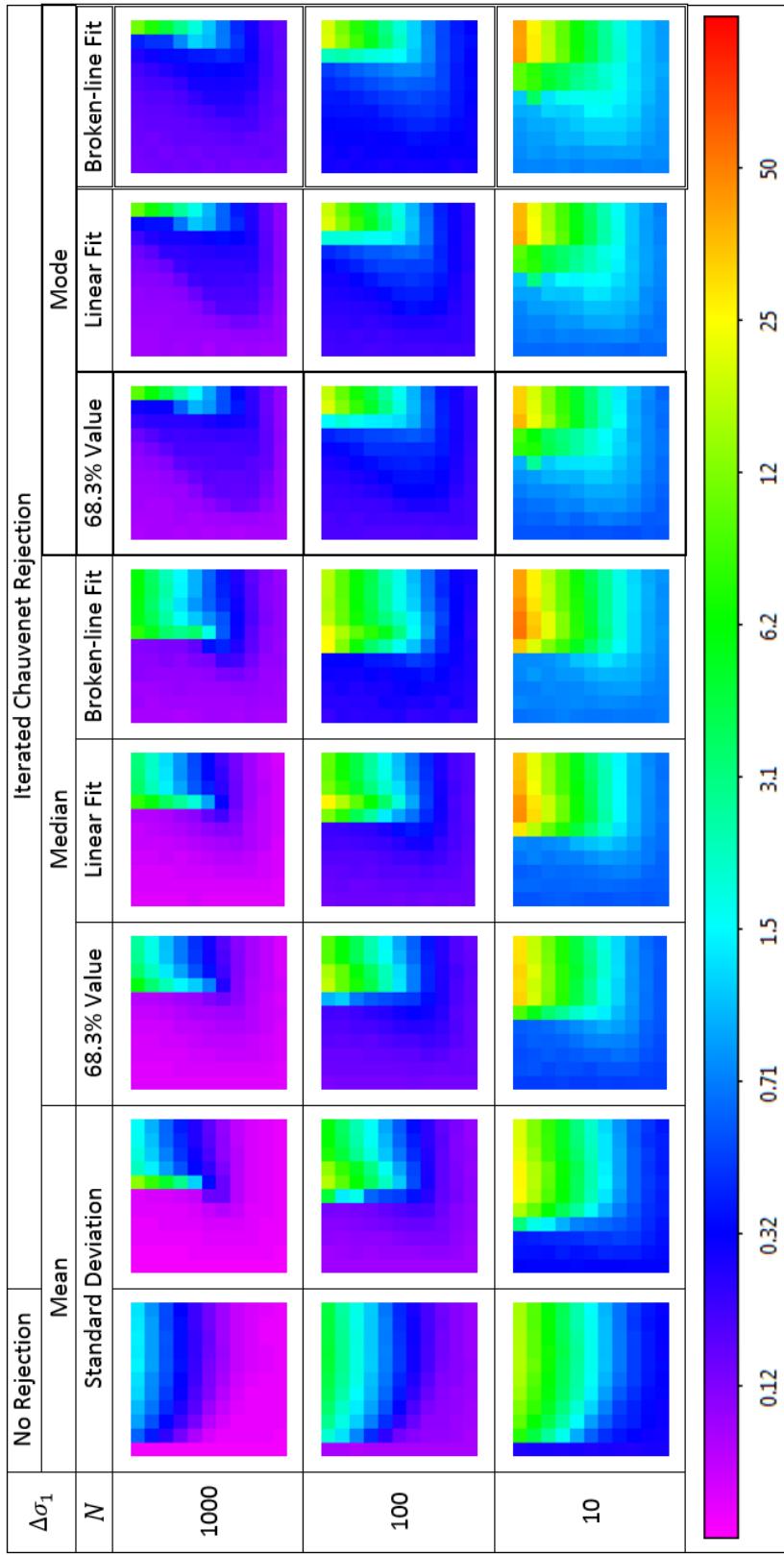


FIG. 16.— Uncertainty in the recovered σ_1 for increasingly robust measurement techniques and decreasing sample sizes (N), for one-sided contaminants. See Figure 5 for contaminant strength (σ_2) vs. fraction of sample (f_2) axis information. The effect of the contaminants, without rejection, can be seen in the first column: Larger contaminant fractions and strengths, as well as smaller sample sizes, generally result in less precise recovered values of σ_1 . However – to the degree that μ_1 can be measured accurately (Figure 13) – all of our Chauvenet rejection techniques are effective at removing outliers (and nearly equally so, since σ_1 is measured on the nearly uncontaminated side of each sample’s distribution), allowing σ_1 to be measured significantly more precisely. Note that, as in the case of two-sided contaminants (Figure 9), when applied to uncontaminated samples ($f_2 = 0$), these techniques recover σ_1 with degrading precisions of $\Delta\sigma_1/\sigma_1 \approx 0.8N^{-1/2}$, $1.3N^{-1/2}$, $2.7N^{-1/2}$, $2.5N^{-1/2}$, $2.8N^{-1/2}$, and $3.9N^{-1/2}$, respectively. However, all things considered (Figures 13 – 16; §3.2), the best-performing technique for Chauvenet-rejecting one-sided contaminants is highlighted with a bold outline, and the best-performing technique for Chauvenet-rejecting contaminants that are neither one-sided nor two-sided, but that are in-between these cases, is highlighted with a double outline. The colors are scaled logarithmically, between 0.02 and 100.

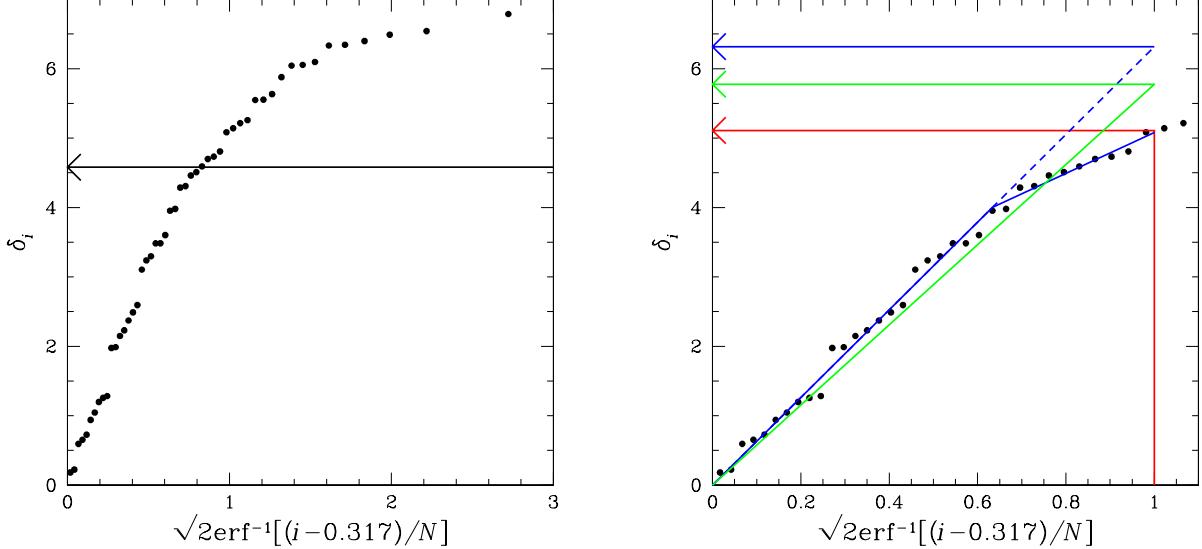


FIG. 17.— **Left:** Sorted deviations from below the median of 100 measurements. A fraction $f_1 = 0.15$ of these measurements are drawn from a Gaussian distribution of mean $\mu_1 = 0$ and standard deviation $\sigma_1 = 1$, and a fraction $f_2 = 0.85$, representing contaminated measurements, are drawn from the positive side of a Gaussian distribution of mean $\mu_2 = 0$ and standard deviation $\sigma_2 = 10$, and added to uncontaminated measurements, drawn as above. The standard deviation, measured below the median, is marked (black arrow). **Right:** Zoom-in of the left panel, but with the 68.3-percentile deviation, also measured below the median, using technique 1 from §2.2 (68.3% value, red), using technique 2 from §2.2 (linear fit, green), and using technique 3 from §2.2 (broken-line fit, blue), instead marked. In this case, the median significantly overestimates μ_1 , measuring 5.81 instead of 0, and consequently the curve breaks downward instead of upward. When this happens, our normally increasingly robust σ -measurement techniques are *decreasingly* accurate, measuring $\sigma_1 = 4.58$, 5.11, 5.78, and 6.32, respectively, instead of 1. In other words, these techniques are only increasingly robust if μ_1 is measured sufficiently accurately. This is the case with the mode, even when f_2 is large, but is not the case with the mean and the median when f_2 is large (Figures 10 and 11), even post-rejection (Figure 12).

fective in the limit of large f_2 and σ_2 (but still significantly more effective than traditional Chauvenet rejection). Since this case approximates both one-sided and two-sided results, when Chauvenet-rejecting contaminants, we recommend using (1) the mode and (2) the 68.3-percentile deviation as measured by technique 3 from §2.2 (the broken-line fit) for the same reasons that we recommend using this combination when rejecting in-between contaminants from normally distributed uncontaminated measurements (§3.2).

It should be noted that if we also change the uncontaminated measurements to be asymmetrically distributed, instead of merely admitting the possibility that they are asymmetrically distributed, the mean, median, and mode then mean different things, in the sense that they mark different parts of the distribution, even in the limit of large N and no contaminants. Furthermore, deviations, however measured, from each of these μ measurements likewise then mean different things. A deeper exploration of these differences, and of their effects on contaminant removal, is beyond the scope of this paper. However, as long as the asymmetry is mild, the effectiveness of this technique should not differ greatly from what has been presented here.⁸

It should also be noted that in the simpler case of two-sided contaminants, this technique differs very little from what has been presented in §3.1, except that σ_{1-} , $\Delta\sigma_{1-}$, σ_{1+} , and $\Delta\sigma_{1+}$ are each determined with about half as many measurements (the measurements on each quanti-

⁸ And what has been presented here is treating each side of the distribution as a pure Gaussian, but of different σ (which, technically, is a discontinuous approximation of the true distribution.)

tity’s side of μ_1).

Finally, it should be noted that this technique is less prone to runaway over-rejection than the techniques presented in §3.1 and §3.2. The calibration of these techniques that we introduced in §2.3 is intended to, and largely does, prevent this from happening, but it can still happen in the limit of very-low N , if two measurements happen to be unusually close together (in which case all other measurements are rejected). For uncontaminated, Gaussian-random data, and the techniques presented in §3.1 and §3.2, this happens $\approx 25\% - 33\%$, $\approx 3.6\% - 9.4\%$, and $\approx 0.30\% - 2.3\%$ of the time when $N = 5, 10$, and 20 , respectively. This is not surprising, given that very-low N distributions can be very non-Gaussian in appearance, in which case this, asymmetric technique may be more appropriate. In this case, again applied to uncontaminated, Gaussian-random data, runaway over-rejection happens only $\approx 0.014\%$ of the time when $N = 5$, and never when $N \geq 10$.

3.3.2. Peaked or Flat-Topped Uncontaminated Distributions

Consider the following generalization of the Gaussian (technically called an exponential power distribution):

$$p(\delta) = \frac{\sqrt{\kappa}}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}|\frac{\delta}{\sigma}|^{2\kappa}}, \quad (5)$$

which reduces to a Gaussian when $\kappa = 1$, but results in peaked (positive-kurtosis) distributions when $\kappa < 1$ and flat-topped (negative-kurtosis) distributions when $\kappa > 1$ (see Figure 18). The standard deviation of this distribution is $\sigma/\sqrt{\kappa}$.

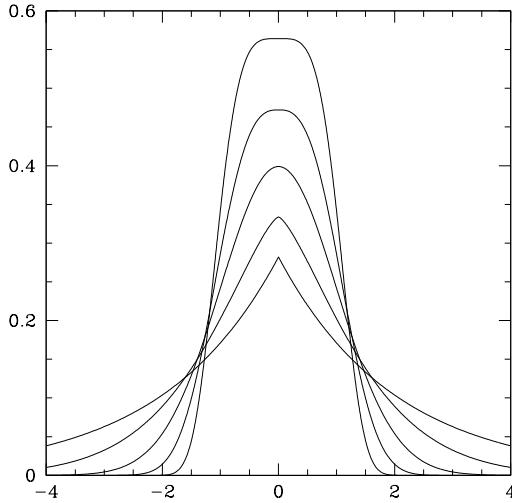


FIG. 18.— Exponential power distribution (Equation 5), for $\kappa = 0.5$ (peaked), 0.7, 1 (Gaussian), 1.4, and 2 (flat-topped).

For this distribution, Chauvenet’s criterion (Equation 1) implies that measurements are rejected if their deviations are greater than a certain number of $\sigma/\sqrt{\kappa}$ (standard deviations), instead of σ , as in the pure Gaussian case.

Furthermore, Equation 4 becomes:

$$\delta_i = \frac{\sigma}{\sqrt{\kappa}} \left[\sqrt{2} \operatorname{erf}^{-1} \left(\frac{i - 0.317}{N} \right) \right], \quad (6)$$

which is proportional to $\sigma/\sqrt{\kappa}$, instead of σ . Consequently, the techniques presented in this paper work identically if the uncontaminated measurements are distributed not normally but peaked or flat-topped – in this specific way.

Of course, not all peaked and flat-topped distributions are of this specific form. However, if only mildly peaked or flat-topped, this form is a good, first-order approximation, and consequently we conclude that the techniques presented in this paper are not overly sensitive to our assumption of Gaussianity, for the uncontaminated measurements.

We summarize all of the recommended, or best-option, robust techniques of §3 in Figure 19.

4. ROBUST CHAUVENET REJECTION: ACCURACY AND PRECISION

In general, we have found that the mode is just as accurate (in the case of two-sided contaminants) or more accurate (in the case of one-sided contaminants) than the median, yet the mode is up to ≈ 5.8 times less precise than the median, and up to ≈ 7.7 times less precise than the mean. We have also found that when μ (equal to the median or the mode) is measured accurately, our increasingly robust 68.3-percentile deviation measurement techniques are either equally accurate (in the case of one-sided contaminants) or increasingly accurate (in the case of two-sided contaminants), yet technique 3 (the broken-line fit) is up to ≈ 2.2 times less precise than technique 2 (the linear fit), up to ≈ 2.4 times less precise than technique 1 (the 68.3% value), and up to ≈ 3.6 times less

precise than the standard deviation.

Consequently, there appears to be a tradeoff between accuracy and precision. But can we have both? In this section, we demonstrate that we can, by applying (1) our robust improvements to traditional Chauvenet rejection (§3), and (2) traditional Chauvenet rejection (§1) in sequence. Traditional Chauvenet rejection uses the mean and the standard deviation, and is consequently the least robust of these techniques, but it is also the most precise, at least when not significantly contaminated by outliers. By applying our robust techniques first, we eliminate the outliers that most significantly affect traditional Chauvenet rejection, allowing us to then capitalize on its precision without its inaccuracy.

We demonstrate the success of this approach first using only our best-option robust techniques, for each of the following contaminant types:

- The median + technique 3 (the broken-line fit) is our best option *for two-sided contaminants*, which are contaminants that are both positive and negative, in equal proportion and strength (§3.1). We plot the average recovered μ_1 , the uncertainty in the recovered μ_1 , the average recovered σ_1 , and the uncertainty in the recovered σ_1 for this technique followed by traditional Chauvenet rejection in the third column of Figures 20 – 23, respectively.
- The mode + technique 1 (the 68.3% value) is our best option *for one-sided contaminants*, which are contaminants that are all positive (the case presented here) or all negative (§3.2). We plot the average recovered μ_1 , the uncertainty in the recovered μ_1 , the average recovered σ_1 , and the uncertainty in the recovered σ_1 for this technique followed by traditional Chauvenet rejection in the third column of Figures 24 – 27, respectively.
- The mode + technique 3 (the broken-line fit) is our best option (1) *for in-between cases*, in which contaminants are both positive and negative, but not in equal proportion or strength (§3.2), and/or (2) *if the uncontaminated distribution is taken to be asymmetric* (§3.3.1). The former case behaves very similarly to Figures 20 – 23 in the limit of two-sided contaminants, and very similarly to Figures 24 – 27 in the limit of (positive) one-sided contaminants. The latter case behaves very similarly to Figure 20 (μ_1), Figure 21 ($\Delta\mu_1$), Figure 22 (σ_{1-} and σ_{1+}), and Figure 23 ($\Delta\sigma_{1-}$ and $\Delta\sigma_{1+}$) in the limit of two-sided contaminants, and similarly to Figure 24 (μ_1), Figure 25 ($\Delta\mu_1$), Figure 26 (σ_{1-}), Figure 27 ($\Delta\sigma_{1-}$), Figure 22 (σ_{1+}), and Figure 23 ($\Delta\sigma_{1+}$), in the limit of (positive) one-sided contaminants. (Consequently, we will not plot these cases separately.)

In all cases, our best-option robust techniques followed by traditional Chauvenet rejection results in vastly improved precisions – comparable to those of traditional Chauvenet rejection when not significantly contaminated by outliers – with only small compromises in accuracy. The small compromises in accuracy, when they occur, are due to our best-option robust techniques not eliminating

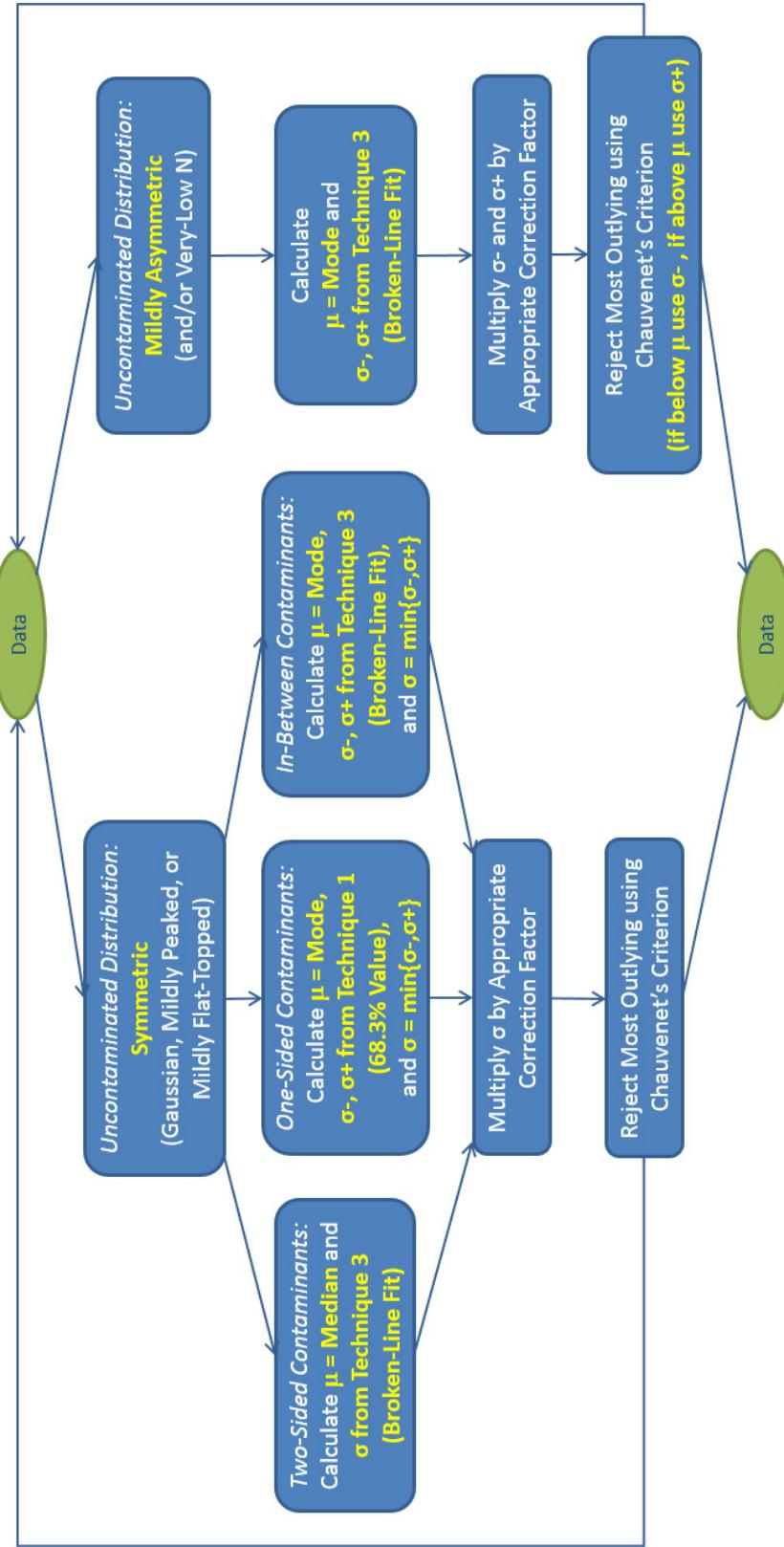


FIG. 19.— Best-option robust techniques for different uncontaminated distributions and contaminant types. The mildly asymmetric technique is also robust against runaway over-rejection in very small samples (§3.3.1). The most discrepant outlier is rejected each iteration, and one iterates until no outliers remain, μ and σ or σ_- and σ_+ , depending on whether the uncontaminated distribution is symmetric or asymmetric, and on the contaminant type) are recalculated after each iteration, and the latter is multiplied by the appropriate correction factor (see Figure 4) before being used to reject the next outlier.

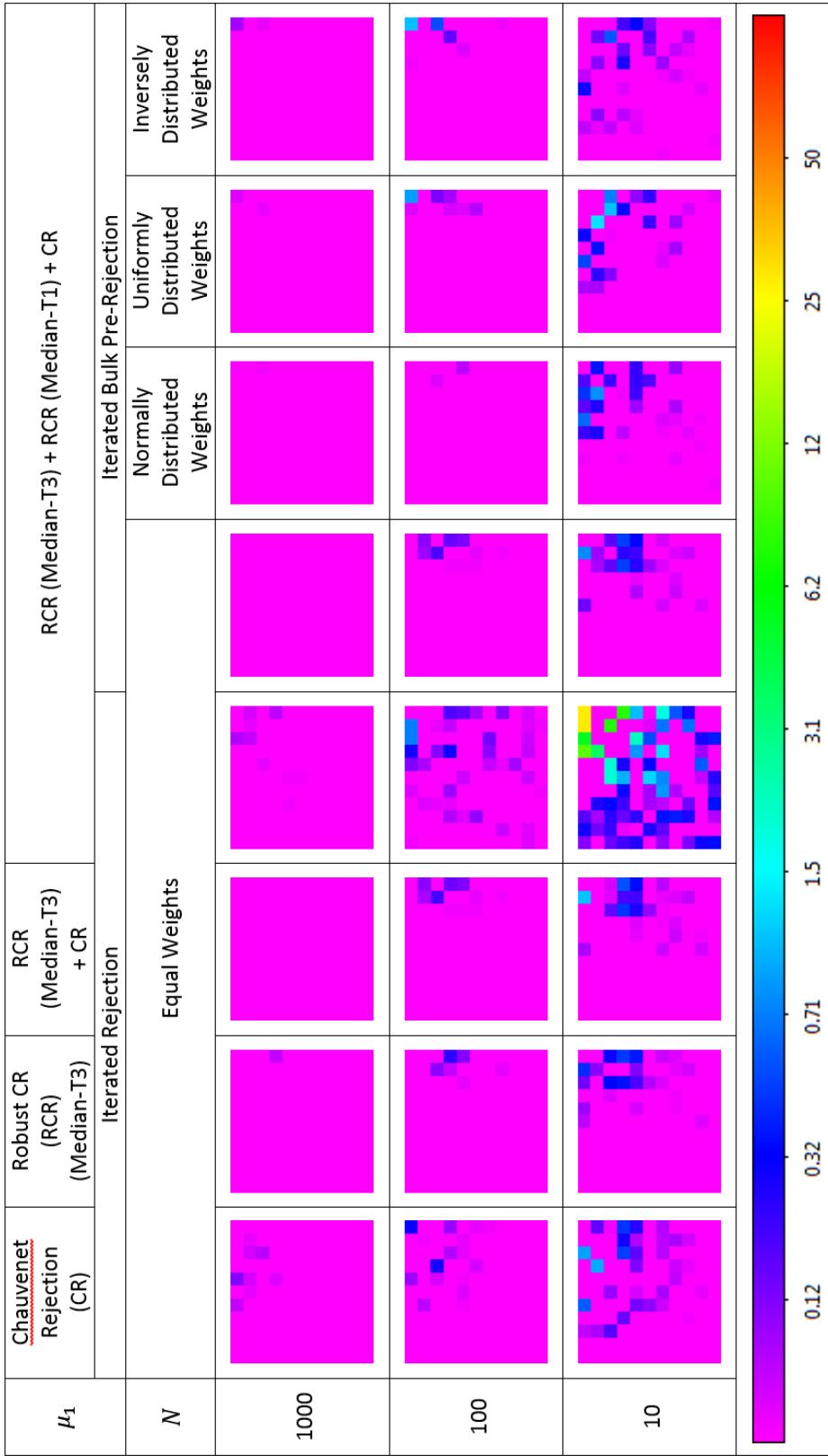


FIG. 20.— Average recovered μ_1 given two-sided contaminants, for, from left to right: (1) traditional Chauvenet rejection; (2) our best-option robust technique (§3.1, §4); (3) column 2 followed by column 1 (§4); (4) column 2 followed by our most-precise robust technique followed by column 3, multiplied by 100 (§4); (5) our best-option bulk pre-rejection technique followed by column 4 (§5); (6) same as column 5, except for weighted data, with weights distributed normally with standard deviation as a fraction of the mean $\sigma_w/\mu_w = 0.3$ (§6); (7) same as column 5, except for weights distributed uniformly from zero, corresponding to $\sigma_w/\mu_w \approx 0.58$ (§6); and (8) same as column 5, except for weights distributed inversely over one dex, corresponding to $\sigma_w/\mu_w \approx 0.73$ (§6). See Figure 5 for σ_2 vs. f_2 axis labels. The colors are scaled logarithmically, between 0.02 and 100.

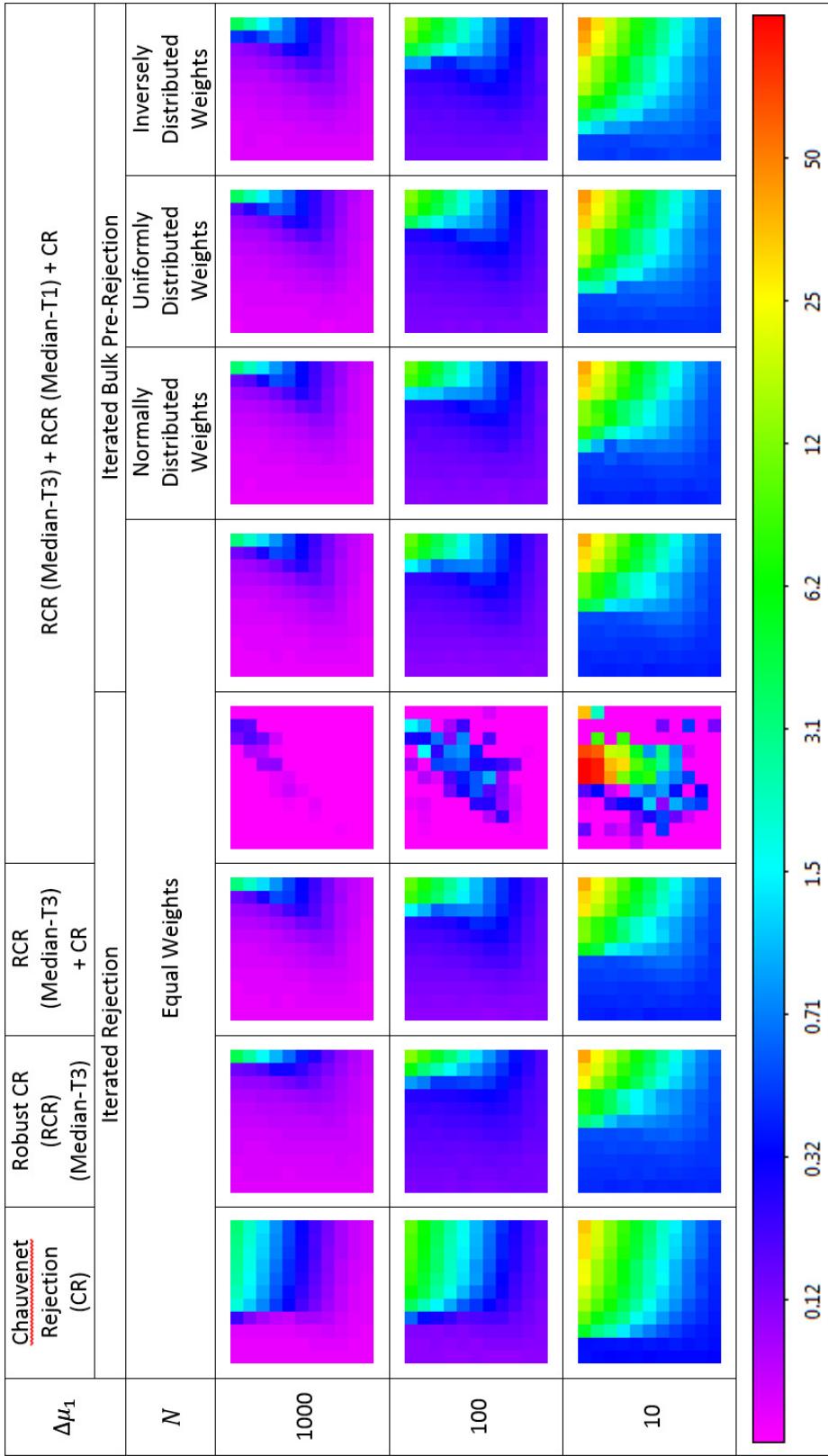


FIG. 21.— Uncertainty in recovered μ_1 given two-sided contaminants, for, from left to right: (1) traditional Chauvenet rejection; (2) our best-option robust technique (§3.1, §4); (3) column 2 followed by column 1 (§4); (4) column 2 followed by our most-precise robust technique followed by column 1, plotted as improvement over column 3, multiplied by 100 (§4); (5) our best-option bulk pre-rejection technique followed by column 4 (§5); (6) same as column 5, except for weighted data, with weights distributed normally with standard deviation as a fraction of the mean $\sigma_w/\mu_w = 0.3$ (§6); (7) same as column 5, except for weights distributed uniformly from zero, corresponding to $\sigma_w/\mu_w \approx 0.58$ (§6); and (8) same as column 5, except for weights distributed inversely over one dex, corresponding to $\sigma_w/\mu_w \approx 0.73$ (§6). See Figure 5 for σ_2 vs. f_2 axis labels. The colors are scaled logarithmically, between 0.02 and 100.

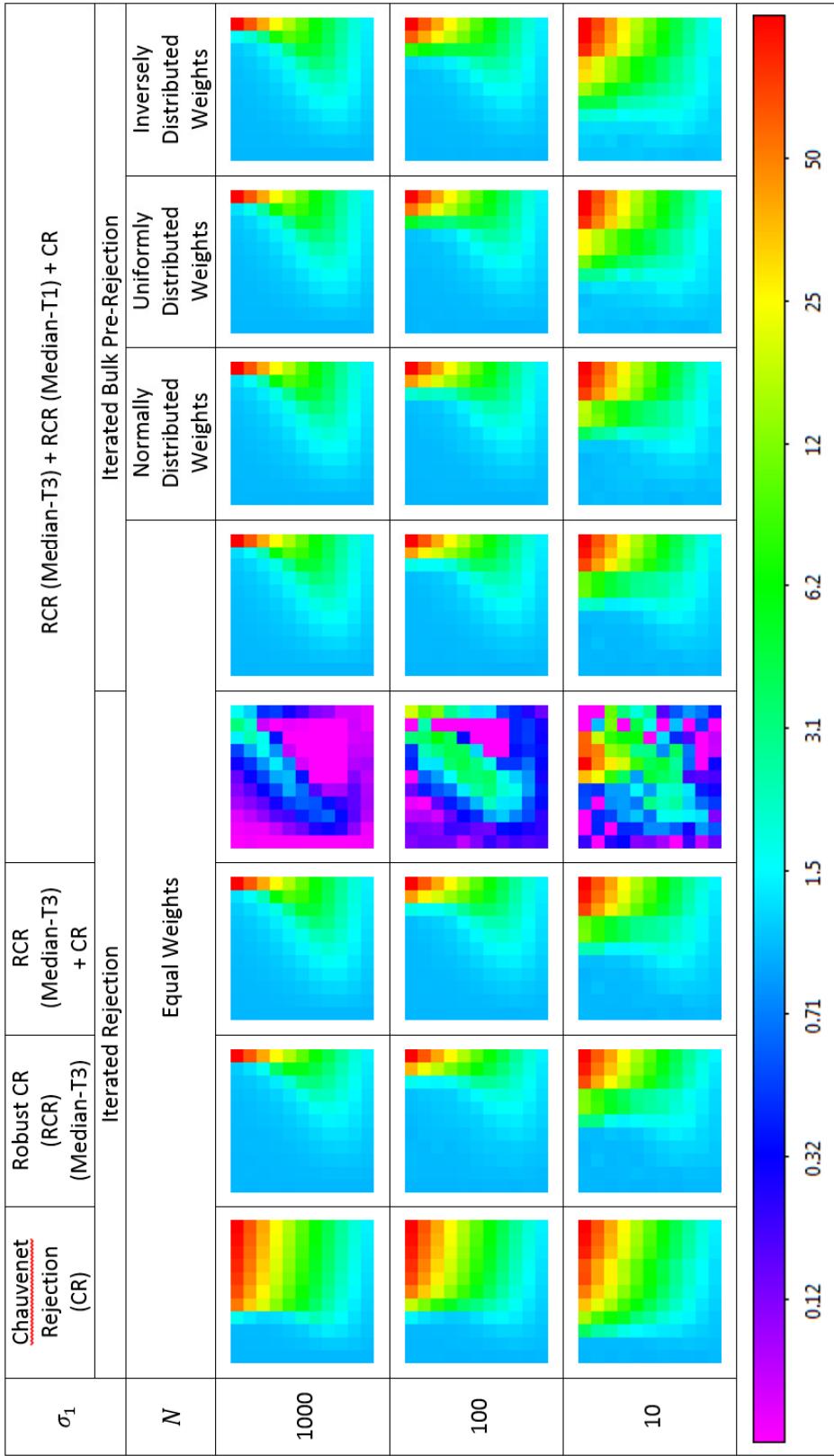


FIG. 22— Average recovered σ_1 given two-sided contaminants, for, from left to right: (1) traditional Chauvenet rejection; (2) our best-option robust technique (§3.1, §4); (3) column 2 followed by column 1 (§4); (4) column 2 followed by column 1, plotted by column 3, multiplied by 100 (§4); (5) our best-option bulk pre-rejection technique followed by column 4 (§5); (6) same as column 5, except for weighted data, with weights distributed normally with standard deviation as a fraction of the mean $\sigma_w/\mu_w = 0.3$ (§6); (7) same as column 5, except for weights distributed uniformly from zero, corresponding to $\sigma_w/\mu_w \approx 0.58$ (§6); and (8) same as column 5, except for weights distributed inversely over one dex, corresponding to $\sigma_w/\mu_w \approx 0.73$ (§6). See Figure 5 for σ_2 vs. f_2 axis labels. The colors are scaled logarithmically, between 0.02 and 100.

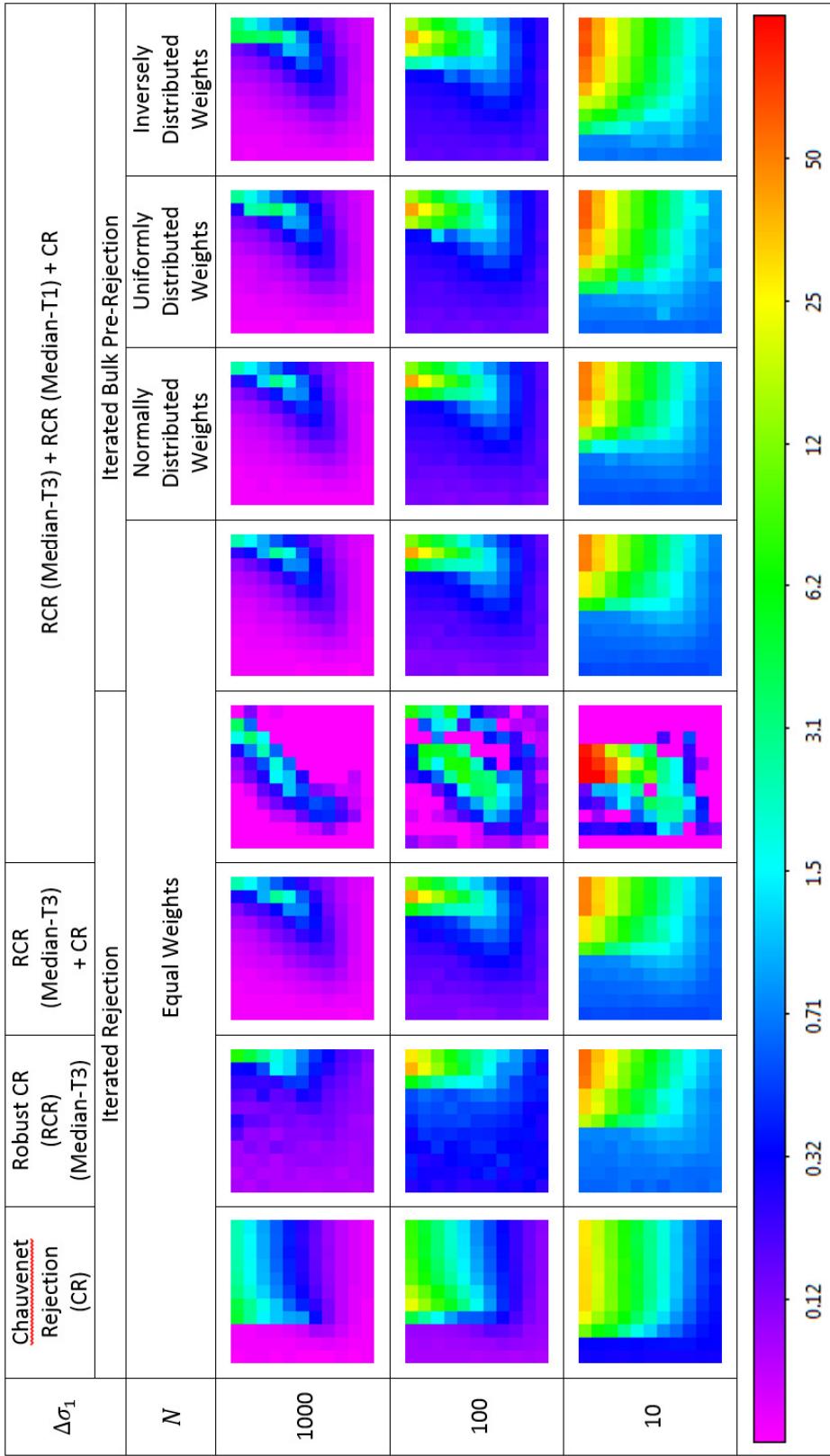


FIG. 23 — Uncertainty in recovered σ_1 given two-sided contaminants, for, from left to right: (1) traditional Chauvenet rejection; (2) our best-option robust technique (§3.1, §4); (3) column 2 followed by column 1 (§4); (4) column 2 followed by our most-precise robust technique followed by column 1, plotted by column 1, multiplied by 100 (§4); (5) our best-option bulk pre-rejection technique followed by column 4 (§5); (6) same as column 5, except for weighted data, with weights distributed normally with standard deviation as a fraction of the mean $\sigma_w/\mu_w = 0.3$ (§6); (7) same as column 5, except for weights distributed uniformly from zero, corresponding to $\sigma_w/\mu_w \approx 0.58$ (§6); and (8) same as column 5, except for weights distributed inversely over one dex, corresponding to $\sigma_w/\mu_w \approx 0.73$ (§6). See Figure 5 for σ_2 vs. f_2 axis labels. The colors are scaled logarithmically, between 0.02 and 100.

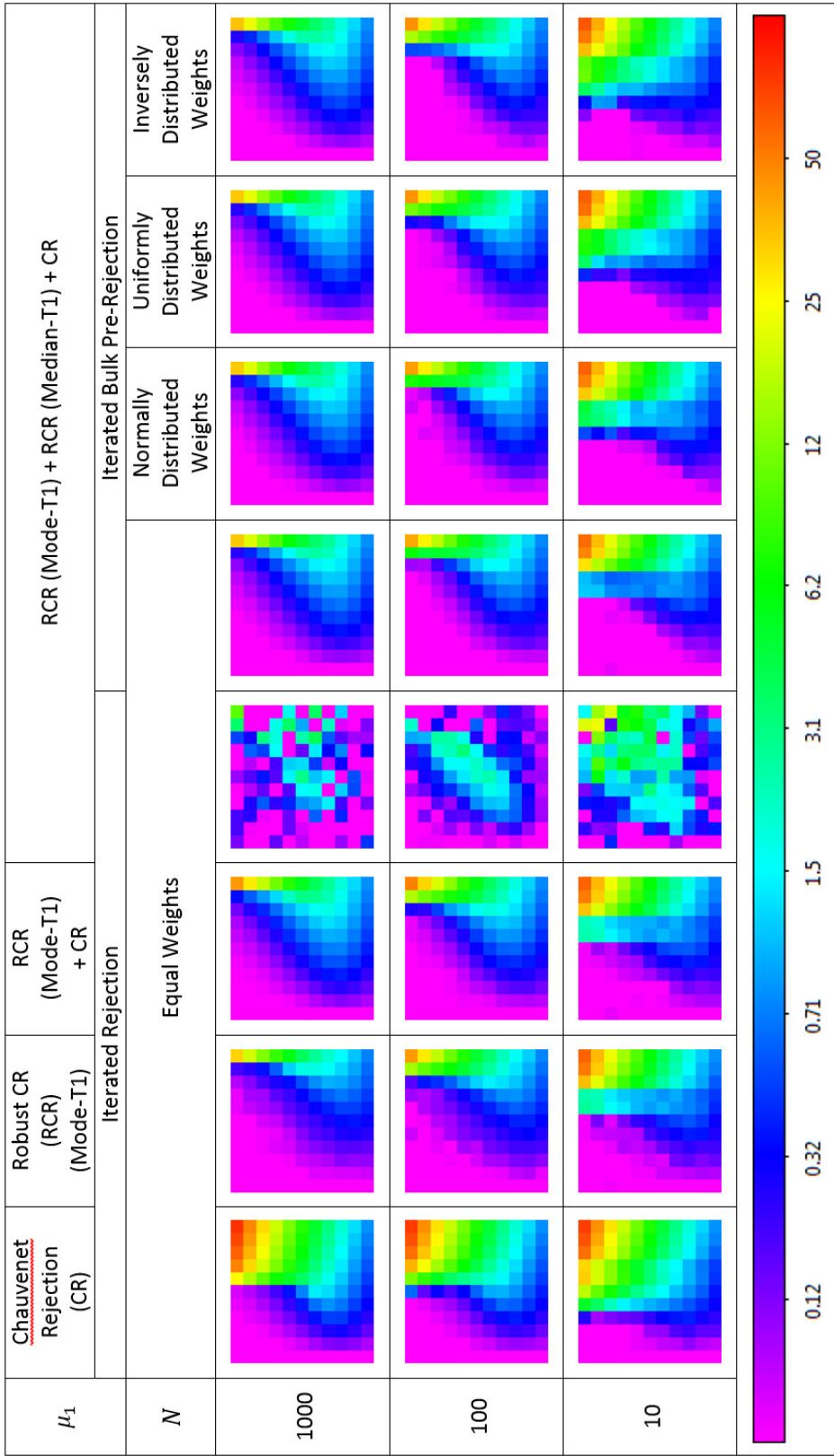


FIG. 24.— Average recovered μ_1 given one-sided contaminants, for, from left to right: (1) traditional Chauvenet rejection; (2) our best-option robust technique (§3.1, §4); (3) column 2 followed by column 1 (§4); (4) column 2 followed by column 1, plotted by column 3, multiplied by 100 (§4); (5) our best-option bulk pre-rejection technique followed by column 4 (§5); (6) same as column 5, except for weighted data, with weights distributed normally with standard deviation as a fraction of the mean $\sigma_w/\mu_w = 0.3$ (§6); (7) same as column 5, except for weights distributed uniformly from zero, corresponding to $\sigma_w/\mu_w \approx 0.58$ (§6); and (8) same as column 5, except for weights distributed inversely over one dex, corresponding to $\sigma_w/\mu_w \approx 0.73$ (§6). See Figure 5 for σ_2 vs. f_2 axis labels. The colors are scaled logarithmically, between 0.02 and 100.

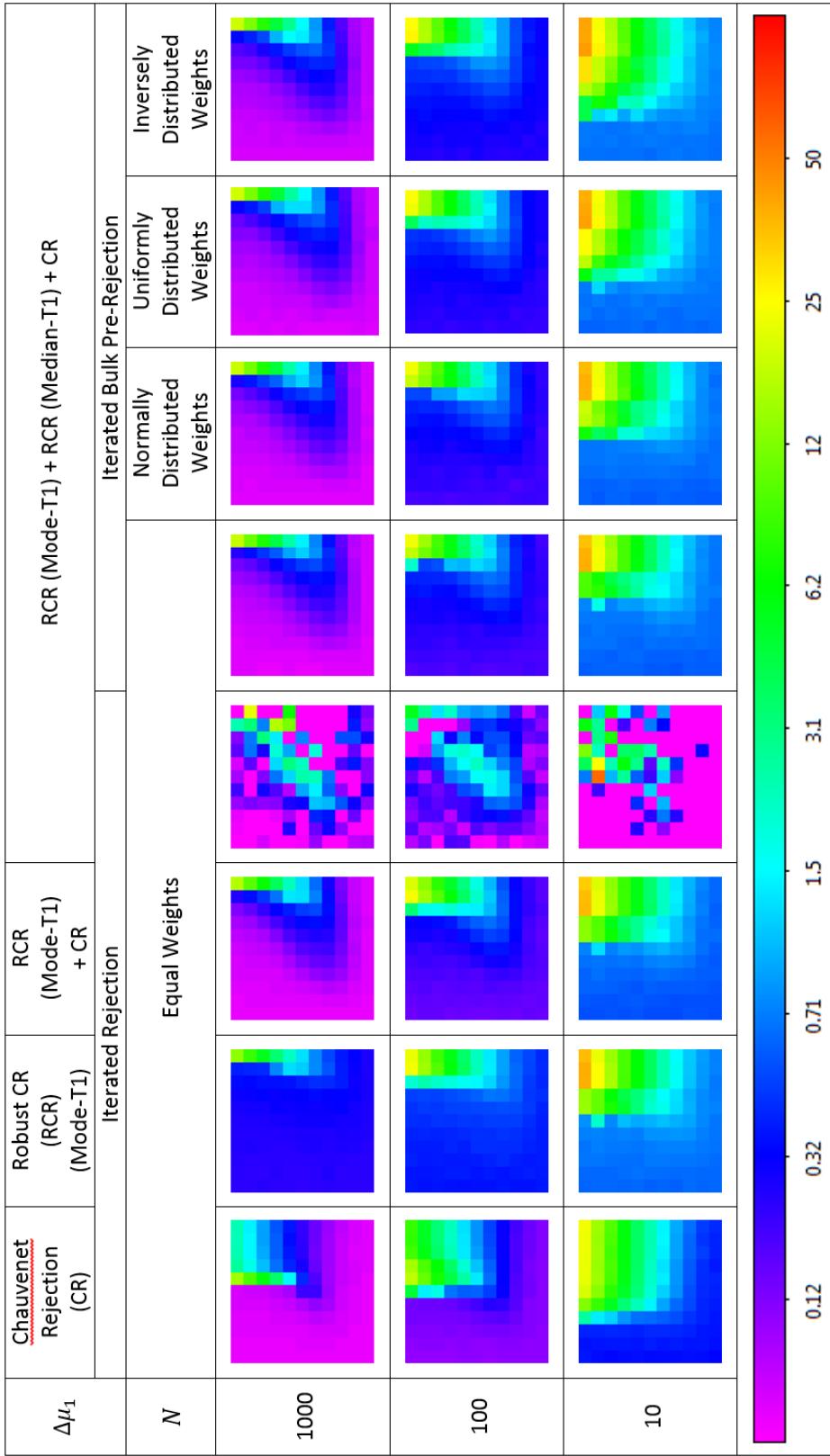


FIG. 25 — Uncertainty in recovered one-sided contaminants, for, from left to right: (1) traditional Chauvenet rejection; (2) our best-option robust technique (§3.1, §4); (3) column 2 followed by column 1 (§4); (4) column 2 followed by our most-precise robust technique followed over column 3, multiplied by 100 (§4); (5) our best-option bulk pre-rejection technique followed by column 4 (§5); (6) same as column 5, except for weighted data, with weights distributed normally with standard deviation as a fraction of the mean $\sigma_w/\mu_w = 0.3$ (§6); (7) same as column 5, except for weights distributed uniformly from zero, corresponding to $\sigma_w/\mu_w \approx 0.58$ (§6); and (8) same as column 5, except for weights distributed inversely over one dex, corresponding to $\sigma_w/\mu_w \approx 0.73$ (§6). See Figure 5 for σ_2 vs. f_2 axis labels. The colors are scaled logarithmically, between 0.02 and 100.

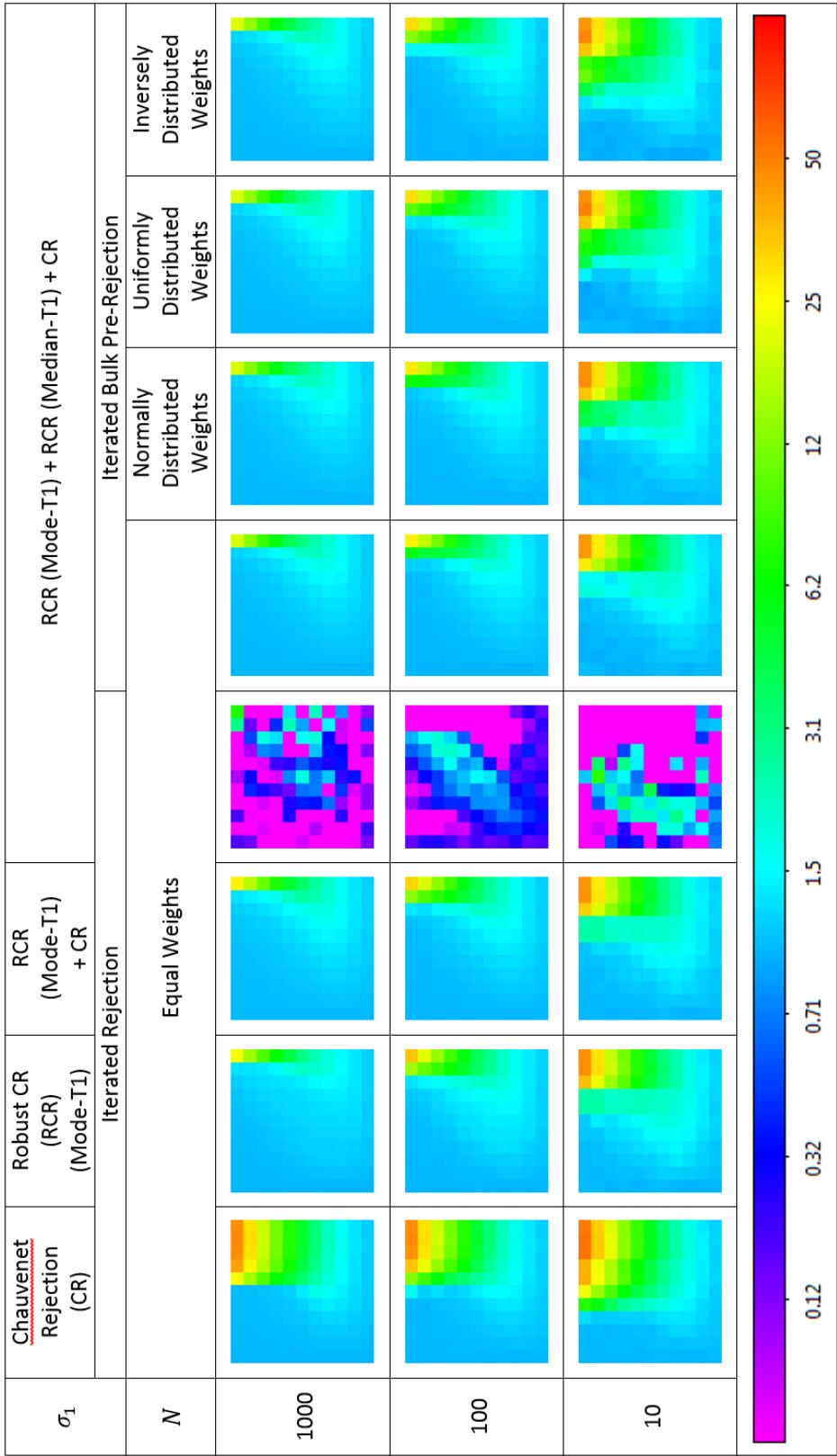


FIG. 26— Average recovered σ_1 given one-sided contaminants, for, from left to right: (1) traditional Chauvenet rejection; (2) our best-option robust technique (§3.1, §4); (3) column 2 followed by column 1 (§4); (4) column 2 followed by column 1, plotted as improvement over column 3, multiplied by 100 (§4); (5) our best-option bulk pre-rejection technique followed by column 4 (§5); (6) same as column 5, except for weights distributed normally with standard deviation as a fraction of the mean $\sigma_w/\mu_w = 0.3$ (§6); (7) same as column 5, except for weights distributed uniformly from zero, corresponding to $\sigma_w/\mu_w \approx 0.58$ (§6); and (8) same as column 5, except for weights distributed inversely over one dex, corresponding to $\sigma_w/\mu_w \approx 0.73$ (§6). See Figure 5 for σ_2 vs. f_2 axis labels. The colors are scaled logarithmically, between 0.02 and 100.

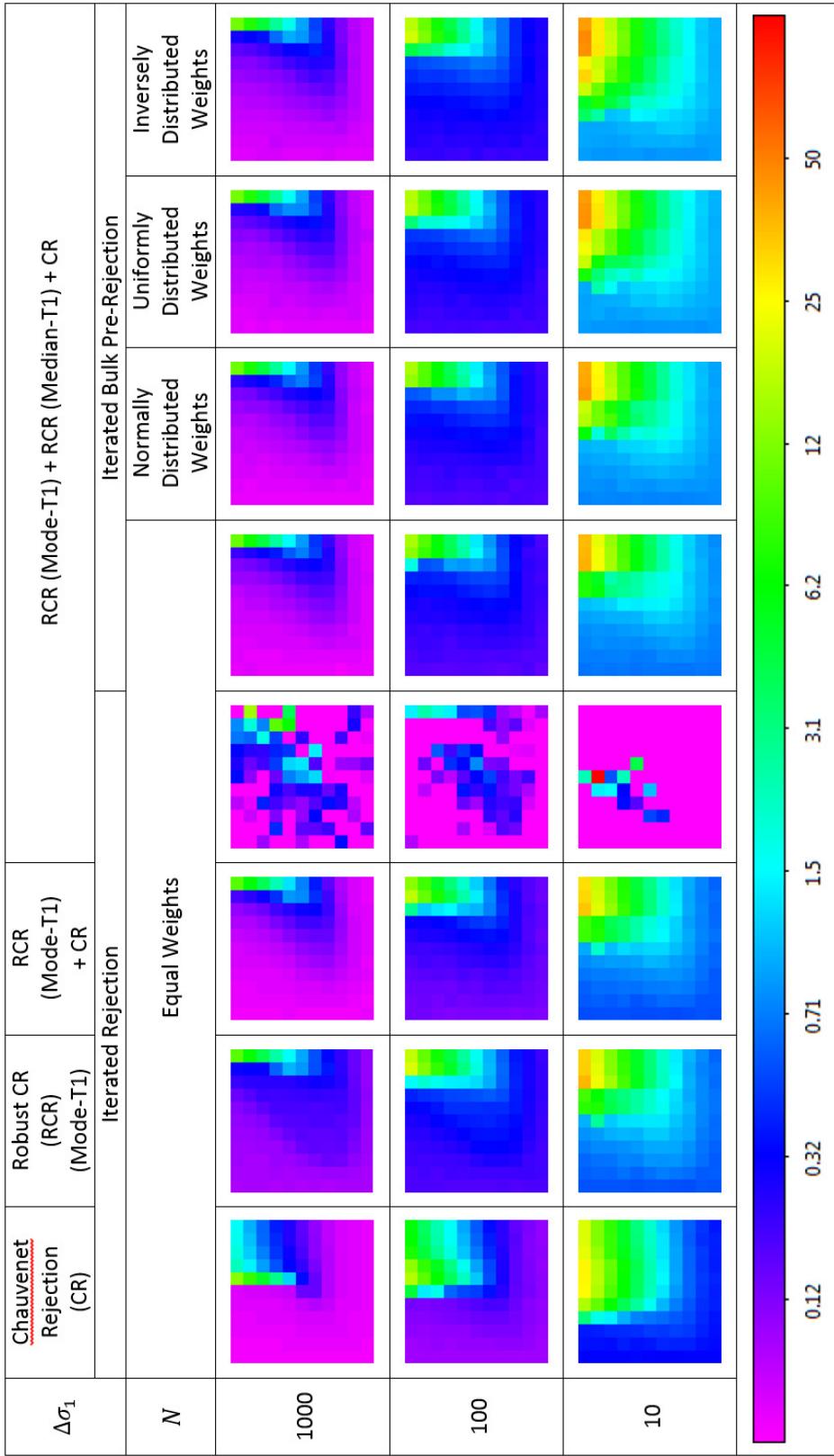


FIG. 27.— Uncertainty in recovered σ_1 given one-sided contaminants, for, from left to right: (1) traditional Chauvenet rejection; (2) our best-option robust technique (§3.1, §4); (3) column 2 followed by column 1 (§4); (4) column 2 followed by our most-precise robust technique followed by column 1, plotted by column 1, multiplied by 100 (§4); (5) our best-option bulk pre-rejection technique followed by column 4 (§5); (6) same as column 5, except for weighted data, with weights distributed normally with standard deviation as a fraction of the mean $\sigma_w/\mu_w = 0.3$ (§6); (7) same as column 5, except for weights distributed uniformly from zero, corresponding to $\sigma_w/\mu_w \approx 0.58$ (§6); and (8) same as column 5, except for weights distributed inversely over one dex, corresponding to $\sigma_w/\mu_w \approx 0.73$ (§6). See Figure 5 for σ_2 vs. f_2 axis labels. The colors are scaled logarithmically, between 0.02 and 100.

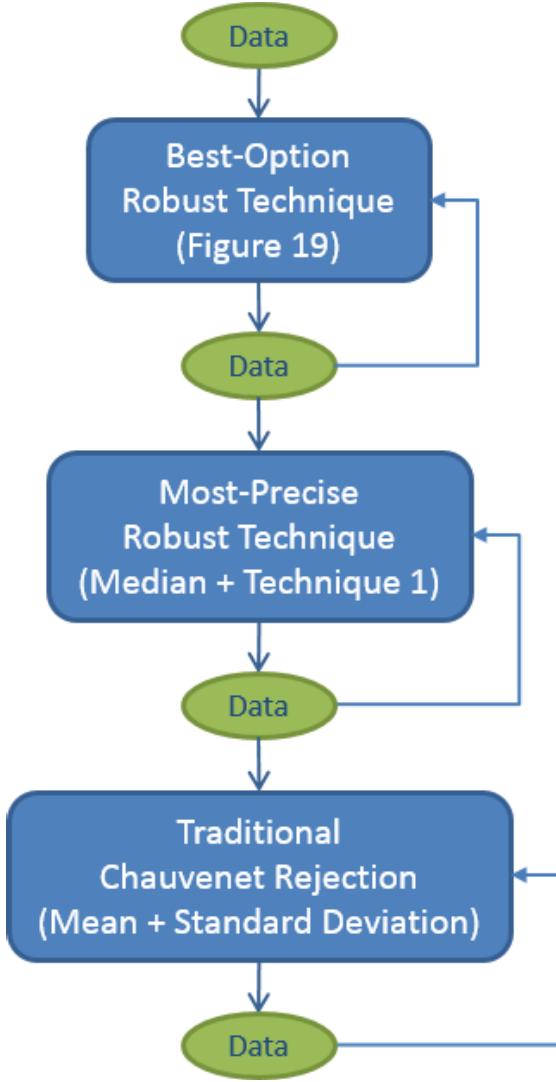


FIG. 28.— Flowchart of our algorithm, without bulk pre-rejection (see Figure 30). The most discrepant outlier is rejected each iteration, and one iterates until no outliers remain before moving on to the next step. μ and σ (or σ_- and σ_+ , depending on whether the uncontaminated distribution is symmetric or asymmetric, and on the contaminant type; Figure 19) are recalculated after each iteration, and the latter is multiplied by the appropriate correction factor (see Figure 29) before being used to reject the next outlier. μ and σ (or σ_- and σ_+) may be calculated in different ways in different steps, but how they are used to reject outliers depends on whether the uncontaminated distribution is symmetric or asymmetric, and on the contaminant type, and consequently does not change from step to step (Figure 19).

enough outliers before traditional Chauvenet rejection is applied.

We further improve this approach by sequencing (1) our best-option robust technique from above, (2) our most-precise robust technique – the median + technique 1 (the 68.3% value) – to eliminate more outliers before applying (3) traditional Chauvenet rejection (see Figure 28 for a flowchart). In nearly all cases, this either leaves the accuracies and the precisions the same, or improves them, by as much as $\approx 30\%$. These are worthwhile gains, particularly given the computational efficiency of the additional step, but they are also difficult to see given the logarithmic scaling that we use in Figures 20 – 27.

Consequently, we instead plot the improvement over column 3, multiplied by 100, in column 4.

Both of these sequencing techniques, as well as a bulk-rejection variant of the latter technique that we present in §5, require the calculation of new correction factors, which we do as in §2.3 and plot in Figure 29.

5. BULK REJECTION

So far, we have rejected only one outlier – the most discrepant outlier – at a time, recomputing μ and σ (or σ_- and σ_+ , depending on whether the uncontaminated distribution is symmetric or asymmetric, and on the contaminant type; Figure 19) after each rejection. This can be time-consuming, computationally, particularly with large samples, so now we evaluate the effectiveness of bulk rejection. In this case, we reject all measurements that meet Chauvenet’s criterion each iteration (however, see Footnote 3), recomputing μ and σ once per iteration instead of once per rejection.

However, bulk rejection works only if σ_1 is never significantly underestimated. If this happens, even if only for a single iteration, significant over-rejection can occur. Furthermore, each of the techniques that we have presented can fail in this way, under the right (or wrong) conditions:

- With one-sided contaminants, when μ_1 *cannot* be measured accurately (Figure 13), the standard deviation underestimates the 68.3-percentile deviation as measured by technique 1 (the 68.3% value), which underestimates the 68.3-percentile deviation as measured by technique 2 (the linear fit), which underestimates the 68.3-percentile deviation as measured by technique 3 (the broken-line fit; Figure 17). In this case, the latter technique overestimates σ_1 . However, the former three techniques can either overestimate σ_1 or underestimate it, sometimes significantly.
- With one-sided or two-sided contaminants, when μ_1 *can* be measured accurately, technique 3 (the broken-line fit) is as accurate (§3.2) or more accurate (§3.1) than the other techniques, but it is also the least precise (§4), meaning that it is as likely to underestimate σ_1 as overestimate it, and, again, sometimes significantly.

Note also that one can transition between these two cases: μ_1 often begins inaccurately measured but ends accurately measured, after iterations of rejections (Figures 11 and 12).

A solution that works in all cases is to measure σ_1 using both techniques 2 (the linear fit) and 3 (the broken-line fit), and adopt the larger of the two for bulk rejection. When μ_1 cannot be measured accurately, the deviation curve breaks downward, and the broken-line fit is the most conservative option (Figure 17). When μ_1 can be measured accurately, the deviation curve breaks upward, and the linear fit is a sufficiently conservative option (Figures 2 and 3). (Technique 1, the 68.3% value, is in this case a more conservative option, but can be overly conservative, bulk-rejecting too few points per iteration.)

We use the same μ -measurement technique as we use for individual rejection. Finally, once bulk rejection

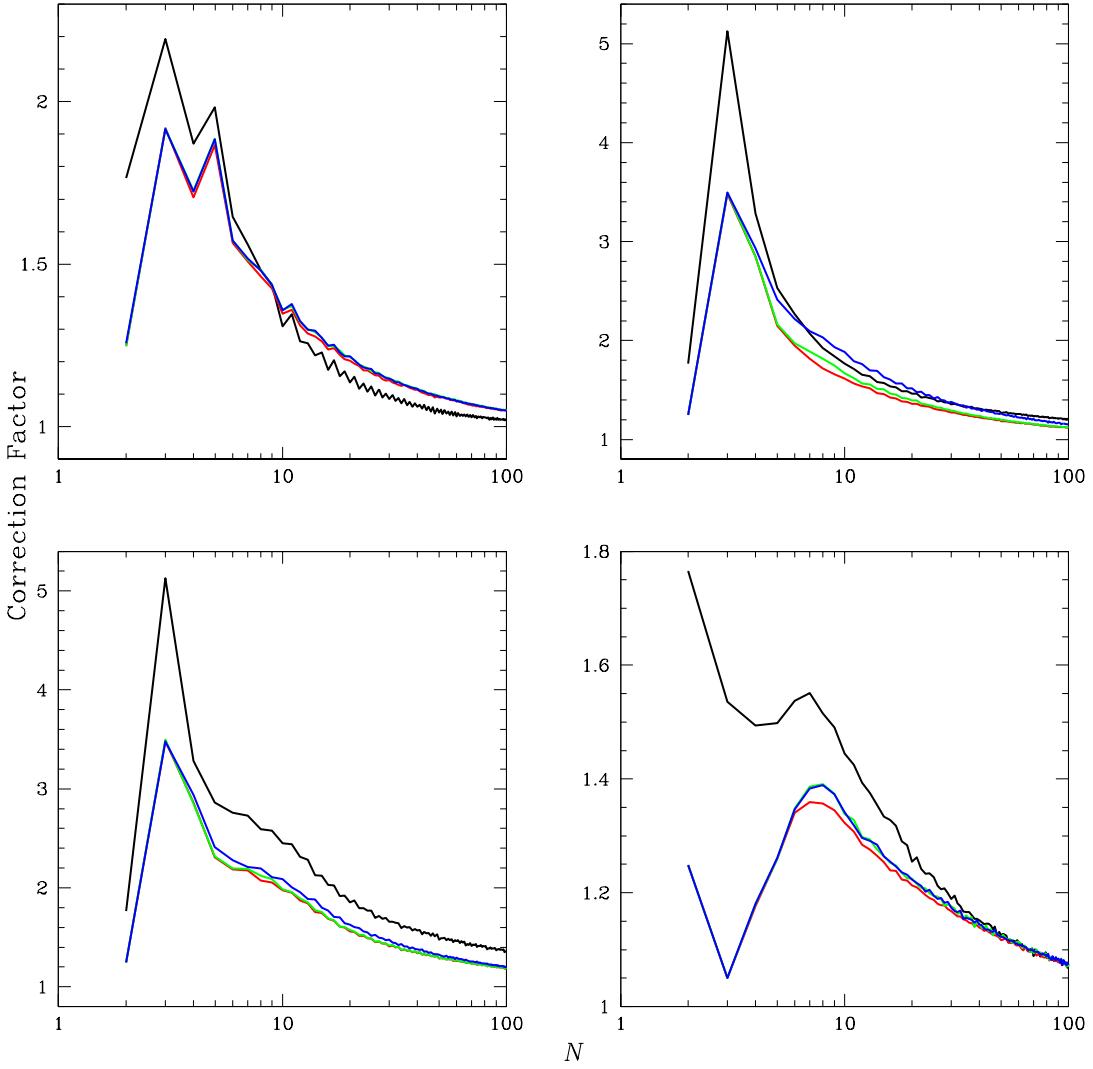


FIG. 29.— Correction factors by which standard and 68.3-percentile deviations, measured from uncontaminated data, need to be multiplied to yield the correct result, on average, and to avoid overaggressive rejection (although this can still happen in sufficiently small samples; see 3.3.1), (1) for the case of our best-option robust techniques (see below; black curves, from Figure 4); (2) for the case of (1) followed by traditional Chauvenet rejection (red curves); (3) for the case of (1) followed by our most-precise robust technique – the median + technique 1 (the 68.3% value) – followed by traditional Chauvenet rejection (green curves); and (4) for the case of bulk rejection (see §5) followed by (3) (blue curves). **Upper left:** For our best-option robust technique for two-sided contaminants – the median + technique 3 (the broken-line fit) – in which we compute a single σ using the deviations both below and above μ (§3.1). **Upper right:** For our best-option robust technique for one-sided contaminants – the mode + technique 1 (the 68.3% value) – in which we compute separate σ below and above μ (σ_- and σ_+ , respectively) and use the smaller of the two when rejecting outliers (§3.2). **Lower left:** For our best-option robust technique for in-between cases – the mode + technique 3 (the broken-line fit) – in which we also use the smaller of σ_- and σ_+ when rejecting outliers (§3.2). **Lower right:** For our best-option robust technique if the uncontaminated distribution is taken to be asymmetric – the mode + technique 3 (the broken-line fit) – in which we use σ_- to reject outliers below μ and σ_+ to reject outliers above μ (§3.3.1). We use look-up tables for $N \leq 100$ and power-law approximations for $N > 100$ (see Appendix B).

is done, we follow up with individual rejection, as described in the second to last paragraph of §4 (see Figure 30 for a flowchart). Individual rejection (1) is significantly faster now that most of the outliers have already been bulk pre-rejected, and (2) ensures accuracy with precision (§4). We plot the results in column 5 of Figures 20 – 27, and, desirably, they do not differ significantly from those of column 4. Speed-up times are presented in Table 1.

6. WEIGHTED DATA

We now consider the case of weighted data. In this case, the mean is given by:

$$\mu = \frac{\sum_{i=1}^N w_i x_i}{\sum_{i=1}^N w_i}, \quad (7)$$

where x_i are the data and w_i are the weights. When the

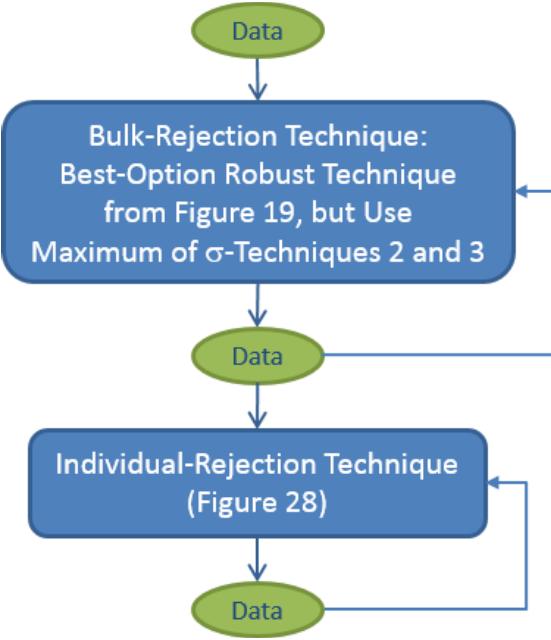


FIG. 30.— Flowchart of our algorithm, with bulk pre-rejection. The first step is bulk rejection, in which all outliers are rejected each iteration, and one iterates until no more outliers are identified. μ and σ (or σ_- and σ_+ , depending on whether the uncontaminated distribution is symmetric or asymmetric, and on the contaminant type; Figure 19) are recalculated after each iteration, and the latter is multiplied by the appropriate correction factor (Figure 29) before being used to reject more outliers. The second step is our individual-rejection algorithm (Figure 28), which ensures accuracy with precision (§4)

mean is measured from the sample, the standard deviation is given by:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N w_i (x_i - \mu)^2}{\sum_{i=1}^N w_i - \Delta \frac{\sum_{i=1}^N w_i^2}{\sum_{i=1}^N w_i}}}, \quad (8)$$

where $\Delta = 1$ when summing over data both below and above the mean, and we take $\Delta = 0.5$ when summing over data either only below or only above the mean.

To determine the weighted median, sort the data and the weights by x_i . First, consider the following, crude definition: Let j be the smallest integer such that:

$$\sum_{i=1}^j w_i \geq 0.5 \sum_{i=1}^N w_i. \quad (9)$$

The weighted median could then be given by $\mu = x_j$, but this definition would be very sensitive to edge effects. Instead, we define the weighted median as follows. Let:

$$s_j = \sum_{i=1}^j (0.5w_{i-1} + 0.5w_i), \quad (10)$$

where $w_0 = 0$, and let j be the smallest integer such that:

$$s_j \geq 0.5 \sum_{i=1}^N w_i. \quad (11)$$

The weighted median is then given by interpolation:

$$\mu = x_{j-1} + (x_j - x_{j-1}) \frac{0.5 \sum_{i=1}^N w_i - s_{j-1}}{s_j - s_{j-1}}, \quad (12)$$

where $s_0 = 0$.

To determine the weighted mode, we again follow an iterative half-sample approach (§2.1). For every j such that:

$$s_j \leq 0.5 \sum_{i=1}^N w_i, \quad (13)$$

let k be the largest integer such that:

$$s_k \leq s_j + 0.5 \sum_{i=1}^N w_i, \quad (14)$$

and for every k such that:

$$s_k \geq 0.5 \sum_{i=1}^N w_i, \quad (15)$$

let j be the smallest integer such that:

$$s_j \geq s_k - 0.5 \sum_{i=1}^N w_i. \quad (16)$$

Of these (j,k) combinations, select the one for which $|x_k - x_j|$ is smallest. If multiple combinations meet this criterion, let j be the smallest of their j values and k be the largest of their k values. Restricting oneself to only the $k - j + 1$ values between and including j and k , repeat this procedure, iterating to completion. Take the weighted median of the final $k - j + 1$ values.

To determine the weighted 68.3-percentile deviation, measured either from the weighted median or the weighted mode, sort the deviations $\delta_i = |x_i - \mu|$ and the weights by δ_i . Analogously to the weighted median above, first consider the following, crude definition: Let j be the smallest integer such that:

$$\sum_{i=1}^j w_i \geq 0.683 \sum_{i=1}^N w_i. \quad (17)$$

The weighted 68.3-percentile deviation could then be given by $\sigma = \delta_j$, but, again, this definition would be very sensitive to edge effects. Instead, we define the weighted 68.3-percentile deviation, for technique 1 (the 68.3% value), as follows. Let:⁹

$$s_j = \sum_{i=1}^j (0.317w_{i-1} + 0.683w_i), \quad (18)$$

where $w_0 = 0$, and let j be the smallest integer such that:

⁹ We center these not halfway through each bin, as we do for the weighted median and weighted mode, but 68.3% of the way through each bin. The need for this can be seen in the case of μ being known a priori, in the limit of one measurement having significantly more weight than the rest, or in the limit of $N \rightarrow 1$.

TABLE 1
TIME IN MILLISECONDS TO MEASURE μ_1 AND σ_1^a

Contaminant Type:	2-Sided		1-Sided		In-Between	
Post-Bulk Rejection Technique: ^b	RCR (Median-T3)	RCR (Mode-T1)	RCR (Mode-T1)	RCR (Mode-T3)	(2-Sided Limit)	(1-Sided Limit)
Corresponding Figures:	19 – 22		23 – 26		—	—
Bulk Pre-Rejection:	No	Yes	No	Yes	No	Yes
Corresponding Column:	4	5	4	5	—	—
$N = 1000$	73	29	160	5.6	160	59
$N = 100$	0.86	0.50	1.6	0.40	1.9	0.98
$N = 10$	0.027	0.030	0.042	0.037	0.047	0.042
					0.048	0.078

^a Averaged over the $11 \times 11 \times 100 = 121,000$ samples in each σ_2 vs. f_2 figure in columns 4 vs. 5 of Figures 20 – 23 (2-sided case), Figures 24 – 27 (1-sided case), and in corresponding (but similar looking, and hence unprinted; §4) figures for the in-between case, in both the 2-sided and 1-sided limits, using a single, AMD Opteron 6168 processor. Measuring the mode is $\approx 1.6N^{0.05}$ times slower than measuring the median, and technique 3 (the broken-line fit) is ≈ 1.2 times slower than technique 1 (the 68.3% value), but bulk pre-rejection is $\approx (N/7.8)^{0.21}$ (2-sided) to $\approx (N/12)^{0.73}$ (1-sided) times faster than no bulk pre-rejection, where N is the sample size. Time to completion is proportional to N^α , where $\alpha \approx 2$ (no bulk pre-rejection) or $1 < \alpha < 2$ (bulk pre-rejection), plus an overhead constant, which dominates when $N \lesssim 5 – 500$. In the case of weighted data (see §6), completion times are roughly $1 + 0.7N^{-0.4}$ times longer.

^b + RCR (Median-T1) + CR (§5)

$$s_j \geq 0.683 \sum_{i=1}^N w_i. \quad (19)$$

The weighted 68.3-percentile deviation, for technique 1, is then given by interpolation:

$$\sigma = \delta_{j-1} + (\delta_j - \delta_{j-1}) \frac{0.683 \sum_{i=1}^N w_i - s_{j-1}}{s_j - s_{j-1}}, \quad (20)$$

where $s_0 = 0$. For techniques 2 (the linear fit) and 3 (the broken-line fit), the 68.3-percentile deviation is given by plotting δ_i vs. $\sqrt{2}\text{erf}^{-1}(s_i / \sum_{i=1}^N w_i)$ and fitting as before (§2.2), except to weighted data (e.g., Appendix A).

Note that as defined here, all of these measurement techniques reduce to their unweighted counterparts (§2.1 and §2.2) when all of the weights, w_i , are equal.

Note also that the correction factors (§2.3) that one uses depend on the weights of the data. To this end, for each of the four scenarios that we consider in §4, corresponding to the four panels of Figure 29, we have computed correction factors for the case of bulk rejection (§5) followed by individual rejection as described in the second to last paragraph of §4, for five representative weight distributions: (1) all weights equal (see Figure 31, solid black curves – same as Figure 29, blue curves); (2) weights distributed normally with standard deviation as a fraction of the mean $\sigma_w/\mu_w = 0.1$ (Figure 31, solid red curves); (3) weights distributed normally with $\sigma_w/\mu_w = 0.3$ (Figure 31, solid green curves); (4) weights distributed uniformly from zero (i.e., low-weight points as common as high-weight points; Figure 31, solid blue curves), corresponding to $\sigma_w/\mu_w \approx 0.58$; and (5) weights distributed inversely over one dex (i.e., low-weight points more common than high-weight points, with the sum of the weights of the low-weight points as impactful as the sum of the weights of the high-weight points; Figure 31,

solid purple curves), corresponding to $\sigma_w/\mu_w \approx 0.73$.

The differences between these are small, but monotonically increasing with σ_w/μ_w , at each N . Furthermore, we have tried other-shaped weight distributions, but with similar σ_w/μ_w , to similar results: The small differences that we do see appear to be more about the effective width of these distributions – which can be easily measured from any sample of weighted measurements – than about the specific shape of these distributions.

Consequently, using these five representative weight distributions, we have produced empirical approximations, as functions of (1) N and (2) σ_w/μ_w of the $x_i = \sqrt{2}\text{erf}^{-1}(s_i / \sum_{i=1}^N w_i) < 1$ points, which can be used with any sample of similarly distributed weights (Figure 31, dashed curves; see Appendix B). We demonstrate these for the latter three weight distributions listed above in columns 6, 7, and 8, respectively, of Figures 20 – 27, and, desirably, they do not differ significantly from those of column 5, in which $\sigma_w/\mu_w = 0$, although there is some decrease in effectiveness in the low- N , high- σ_w/μ_w limit.

It is this combination of (1) sequencing robust improvements to traditional Chauvenet rejection with traditional Chauvenet rejection, to achieve both accuracy and precision (§4), (2) bulk pre-rejection, to significantly decrease computing times in large samples (§5), and (3) the ability to handle weighted data (§6) that we typically refer to as robust Chauvenet rejection (RCR).

7. EXAMPLE: APERTURE PHOTOMETRY

The Skynet Robotic Telescope Network is a global network of fully automated, or robotic, volunteer telescopes, scheduled through a common web interface.¹⁰ Currently, our optical telescopes range in size from 14 to 40 inches, and span four continents. Recently, we added Skynet’s first radio telescope, Green Bank Observatory’s 20-meter diameter dish, in West Virginia (Martin et al. 2018).

¹⁰ <https://skynet.unc.edu>

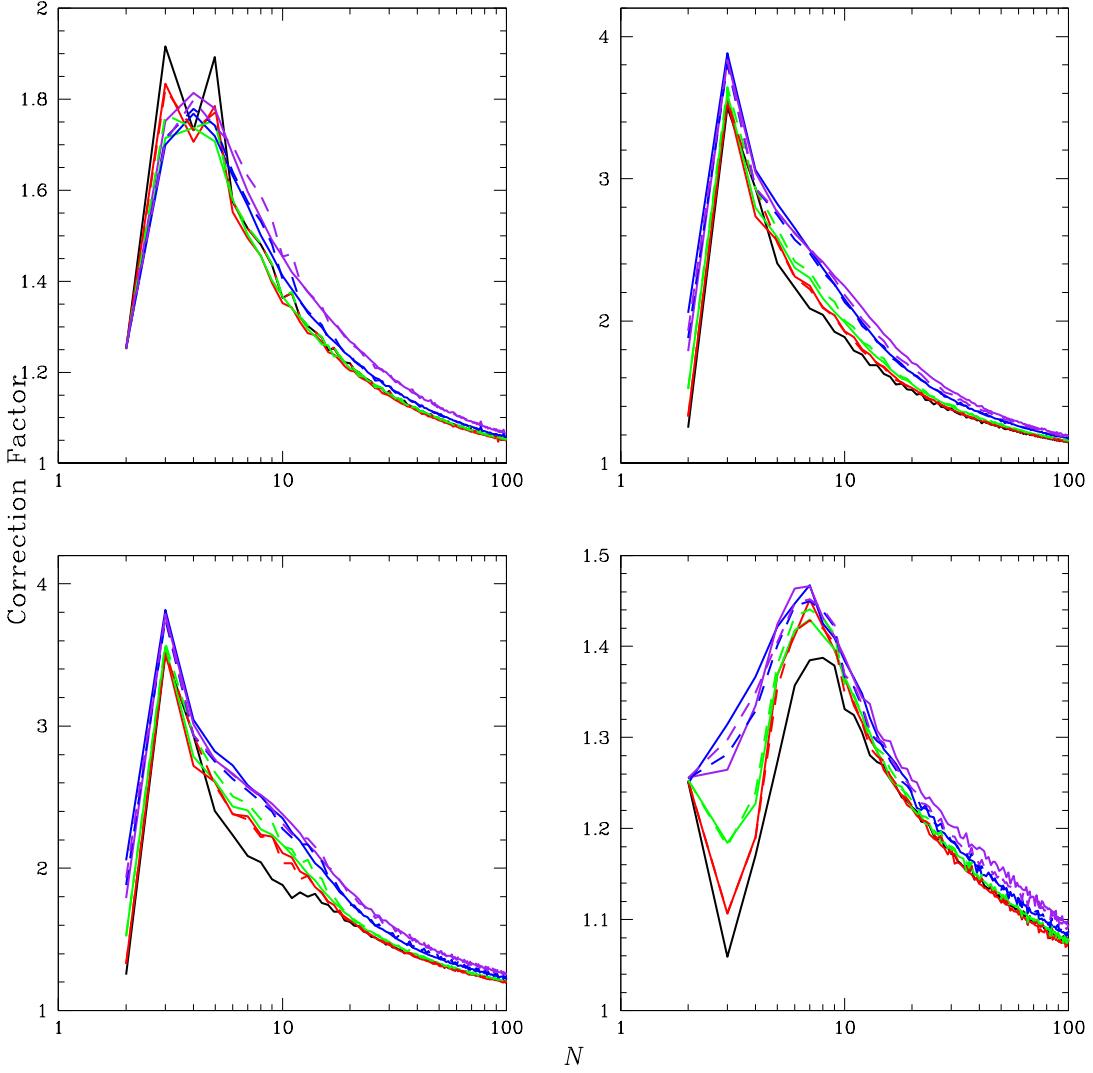


FIG. 31.— Same as the blue curves from Figure 29, but for five representative weight distributions: (1) all weights equal (solid black curves – same as the blue curves from Figure 29); (2) weights distributed normally with standard deviation as a fraction of the mean $\sigma_w/\mu_w = 0.1$ (solid red curves); (3) weights distributed normally with $\sigma_w/\mu_w = 0.3$ (solid green curves); (4) weights distributed uniformly from zero (i.e., low-weight points as common as high-weight points; solid blue curves), corresponding to $\sigma_w/\mu_w \approx 0.58$; and (5) weights distributed inversely over one dex (i.e., low-weight points more common than high-weight points, with the sum of the weights of the low-weight points as impactful as the sum of the weights of the high-weight points; solid purple curves), corresponding to $\sigma_w/\mu_w \approx 0.73$. From these, we have produced empirical approximations, as functions of (1) N and (2) σ_w/μ_w of the $x_i = \sqrt{2}\text{erf}^{-1}(s_i / \sum_{i=1}^N w_i) < 1$ points, which can be used with any sample of similarly distributed weights (dashed curves; see Appendix B).

We have been incorporating RCR into Skynet’s image-processing library, beginning with our single-dish mapping algorithm (Martin et al. 2018). Here, we use RCR extensively: (1) to eliminate contaminants during gain calibration; (2) to measure the noise level of the data along each scan, and as a function of time, to aid in background subtraction along the scans; (3) to combine locally fitted, background-level models into global models, for background subtraction along each scan; (4) to eliminate contaminants if signal and telescope-position clocks must be synchronized post facto from the background-subtracted data; (5) to measure the noise level of the background-subtracted data across each scan, and as a function of time, to aid in radio-frequency interfer-

ence (RFI) cleaning; and (6) to combine locally fitted models of the background-subtracted, RFI-cleaned signal into a global model, describing the entire observation. After this, we locally model and fit a “surface” to the background-subtracted, time-delay corrected, RFI-cleaned data, filling in the gaps between the signal measurements to produce the final image (e.g., see Figure 32). Furthermore, each pixel in the final image is weighted, equal to the proximity-weighted number of data points that contributed to its determination (e.g., Figure 32, lower right).

Here, we demonstrate another application of RCR: aperture photometry, in this case of the primary source, Cas A, in the lower-left panel of Figure 32. We have cen-

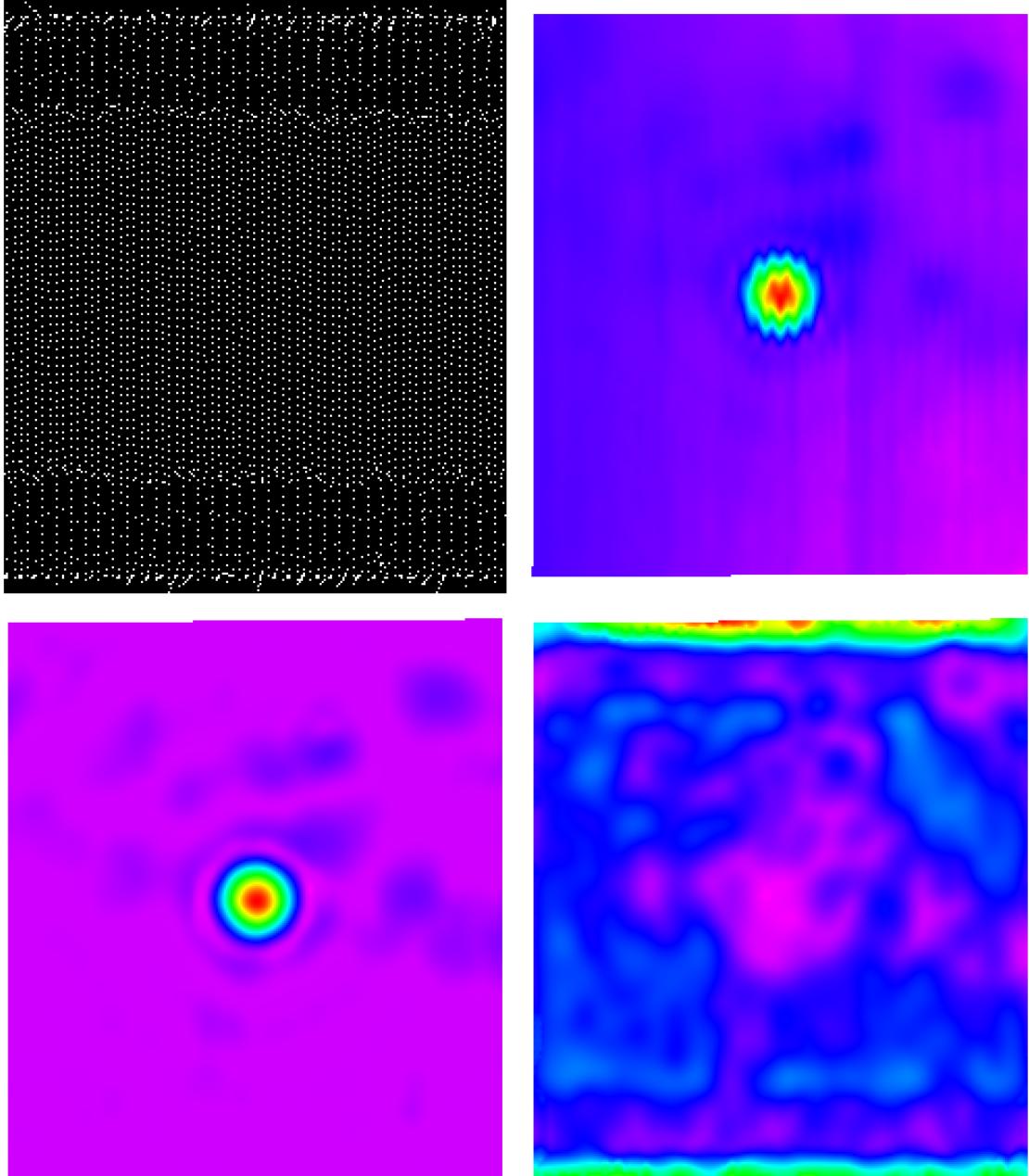


FIG. 32.— **Upper left:** Signal-measurement positions from an on-the-fly raster mapping of Cas A, made with Green Bank Observatory’s 20-meter diameter telescope, in L band (gaps at the top and bottom are due to the telescope jumping ahead to get back on schedule, after losing time reversing direction at the ends of scans). **Upper right:** Raw image, which has been surface modeled (to fill in the gaps between the signal measurements, without additionally blurring the image), but has not been background subtracted, time-delay corrected, or RFI cleaned. **Lower left:** Final image, which has been background subtracted, time-delay corrected, RFI cleaned, and then surface modeled. **Lower right:** Proximity-weighted number of data points that contributed to the surface model at each pixel. Weights are lower in the vicinity of signal, due to the RFI-cleaning algorithm (Martin et al. 2018). In the latter three panels, square-root scaling is used to enhance the visibility of fainter structures.

tered the aperture on the source, and have selected its radius to match that of the minimum between the source and its first Airy ring (see Figure 33). We sum all of the values in the aperture, but from each we must also subtract off the average background-level value, which we measure from the surrounding annulus.

The annulus we have selected to extend from the radius of the aperture to 10 beamwidths (Figure 33). However,

it is heavily contaminated, by the source’s Airy rings and diffraction spikes, and by other sources. This is a good case to demonstrate RCR, because (1) a large fraction, f_2 , of the pixels in the annulus are contaminated, and (2) they are strongly contaminated, σ_2 , compared to the background-noise level, σ_1 . It is also a good case to demonstrate bulk pre-rejection (§5), because there is a large number of pixels in the annulus, and to demon-

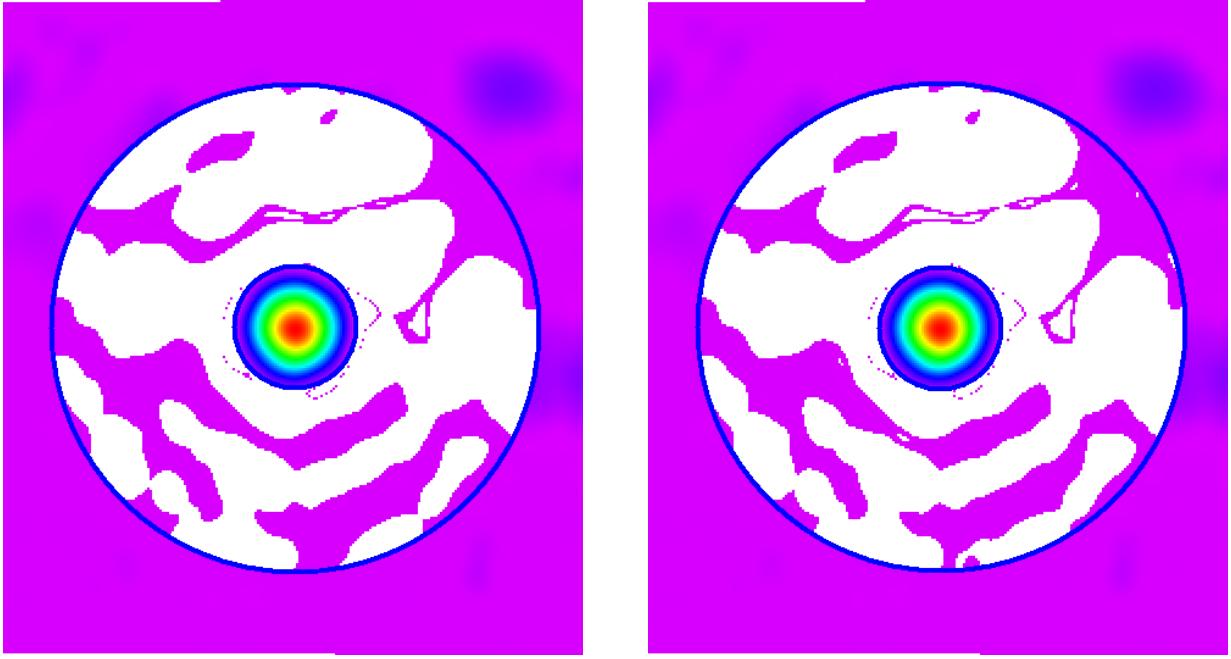


FIG. 33.— Same as the lower-left panel of Figure 32, except that contaminated pixels (contaminated by other sources, Airy rings, diffraction spikes, etc.) have been robust-Chauvenet rejected within an annulus in which we are measuring the background level, (1) assuming that the contaminants are one sided (left), and (2) assuming that the contaminants are an in-between case, with some negative contaminants as well (right).

strate RCR’s ability to handle weighted data (§6, Figure 32, lower right).

These are one-sided contaminants, so we follow bulk pre-rejection with “RCR (Mode – Technique 1) + RCR (Median – Technique 1) + CR” (§4, Figures 24 – 27). The rejected pixels have been excised from the left panel of Figure 33.

If one suspected an in-between case, with some negative contaminants as well, we would instead follow bulk pre-rejection with “RCR (Mode – Technique 3) + RCR (Median – Technique 1) + CR” (§4). The rejected pixels for this case have been excised from the right panel of Figure 33.

For these two cases, the post-rejection background level is measured to be -0.00002 ± 0.00047 and -0.00003 ± 0.00045 , respectively, which is a significant improvement over the pre-rejection value, 0.023 ± 0.040 (gain-calibration units).

It is also a significant improvement over what traditional Chauvenet rejection yields: 0.022 ± 0.036 , which is nearly identical to the pre-rejection value. I.e., traditional Chauvenet rejection fails to eliminate most of the outliers, resulting in biased, and additionally uncertain, photometry. In this case, traditional Chauvenet rejection is equivalent to sigma clipping with a 4.35σ threshold, given the number of pixels in the annulus (Equation 1). This demonstrates that something as fundamental to astronomy as aperture photometry can be improved upon, in the limit of contaminated, or crowded, fields.

Lastly, we point out that RCR has already been successfully employed by Trotter et al. (2017), who made many measurements of Cas A, and other bright radio sources, with Skynet’s 20-meter telescope, and calibrated these with measurements of Cyg A, observed as closely in time as possible, but not always on the same day.

RCR was used to reject measurements that were outlying, because of variations in the receiver’s gain between the primary and calibration observations. In some cases, in particular when the timescale between these observations was longer, up to 35% of these samples were contaminated, necessitating the use of RCR instead of traditional Chauvenet rejection/sigma clipping. (Trotter et al. additionally used RCR to eliminate occasional pointing errors when modeling systematic focus differences between these sources, from drift-scan data taken with a different, transit radio telescope.)

8. MODEL FITTING

So far, we have only considered cases where uncontaminated measurements are distributed, either normally (§3.1, §3.2) or non-normally (§3.3), about a single, parameterized value, y . In particular, we have introduced increasingly robust ways of measuring y , or to put it differently, of fitting y to measurements, namely: the mean, the median, and the mode (§2.1, §6). We have also introduced techniques: (1) to more robustly identify outlying deviations from y , for rejection (§2.2 – §3, §6); (2) to more precisely measure y , without sacrificing robustness (§4); and (3) to more rapidly measure y (§5).

In this section, we show that RCR can also be applied when measurements are distributed not about a single, parameterized value, but about a parameterized model, $y(\{x\} | \{\theta\})$, where $\{x\}$ are the model’s independent variables, and $\{\theta\}$ are the model’s parameters. But first, we must introduce new, increasingly robust ways of fitting $y(\{x\} | \{\theta\})$ to measurements, now given by $\left\{ \left\{ x_i^{+\sigma_{x+,i}} \right\}, y_i^{+\sigma_{y+,i}} \right\}$. Specifically, these will be generalizations of the mean, the median, and the mode, that reduce to these in the limit of a single-parameter fit, but that result in best-fit, or baseline, models from which

deviations can be calculated otherwise. Consequently, these will be able to replace the mean, the median, and the mode in the RCR algorithm, with no other modification to the algorithm being necessary.

8.1. Generalized Measures of Central Tendency

Usually, models are fitted to measurements by maximizing a likelihood function.¹¹ For example, if:

$$\sigma_{x-,i} \approx \sigma_{x+,i} \approx 0, \quad (21)$$

$$\sigma_{y-,i} \approx \sigma_{y+,i} \approx \sigma_{y,i}, \quad (22)$$

and

$$\chi^2 = \sum_i \left[\frac{y_i - y(\{x\} | \{\theta\})}{\sigma_{y,i}} \right]^2 \approx N - M, \quad (23)$$

where N is the number of independent measurements, and M is the number of non-degenerate model parameters, this function is simple: $\mathcal{L} \propto e^{-\chi^2/2}$, in which case maximizing \mathcal{L} is equivalent to minimizing χ^2 . If these conditions are not met, \mathcal{L} , and its maximization, can be significantly more involved (e.g., Reichart 2001; Trotter 2011). Regardless, such, maximum-likelihood, approaches are generalizations of the mean, and consequently are not robust.

To see this, again consider the simple case of the single-parameter model: $y(\{x\} | \{\theta\}) = y$. Minimizing Equation 23 with respect to y (i.e., solving $\partial\chi^2/\partial y = 0$ for y) yields a best-fit parameter value, and a best-fit model, of $y = (\sum_i y_i / \sigma_{y,i}^2) / (\sum_i 1 / \sigma_{y,i}^2) = (\sum_i w_i y_i) / (\sum_i w_i)$. This is just the weighted mean of the measurements (Equation 7), which is not robust.

One could imagine iterating between (1) maximizing \mathcal{L} to establish a best-fit model, and (2) applying robust outlier rejection to the deviations from this model, but given that (1) is not robust, this would be little better than iterating with traditional Chauvenet rejection, which relies on the weighted mean. Instead, we retain the RCR algorithm, but replace the weighted mean, the weighted median, and the weighted mode with generalized versions, maintaining the robustness, and precision, of each. We generalize the weighted mean as above, with maximum-likelihood model fitting. We generalize the weighted median and the weighted mode as follows.

First, consider the case of an M -parameter model where for any combination of M measurements, a unique set of parameter values, $\{\theta\}_j$, can be determined.¹² Furthermore, imagine doing this for all $1 \leq j \leq N!/[M!(N-M)!]$ combinations of M measurements,¹³

¹¹ Or, by maximizing the product of a likelihood function and a prior probability distribution, if the latter is available.

¹² In the event of redundant independent-variable information, fewer than M parameter values can be determined, and we address this case in §8.3.2. In the event of a periodic model, multiple M -parameter solutions can be determined (some equivalent to each other, some not), and we address this case in §8.3.3.

¹³ Or for as large of a randomly drawn (but without repetitions) subset of these as is computationally reasonable. We switch over to random draws, where each measurement is drawn in proportion to its weight, when $N!/[M!(N-M)!] > 20,000$. For $M = 2$, this corresponds to $N > 200$. For $M = 3$, this corresponds to $N > 50$.

and weighting each calculated parameter value by how accurately it could be determined (see §8.2). Our generalizations are then given by: (1) the weighted median of $\{\{\theta\}_j\}$, and (2) the weighted mode of $\{\{\theta\}_j\}$.

Although more sophisticated implementations can be imagined, here we define these quantities simply, and such that they reduce to the weighted median and the weighted mode, respectively, in the limit of the single-parameter model, just as the maximum-likelihood technique above reduces to the weighted mean in this limit:

- For the weighted median of $\{\{\theta\}_j\}$, we calculate the weighted median for each model parameter separately.
- For the weighted mode of $\{\{\theta\}_j\}$, we determine the half-sample for each model parameter separately, but then include only the intersection of these half-samples in the next iteration.¹⁴

We demonstrate these techniques, and the maximum-likelihood technique, for a simple, linear, but contaminated, model in Figure 34.

In Figure 35, we apply RCR as before (§4, §5), except that we no longer use the weighted mode, the weighted median, and the weighted mean to establish baseline *values* from which the deviations of the measurements can be determined. Rather, we use our generalizations of these, to establish baseline *functions* of $\{x\}$, of corresponding robustness and precision, from which these deviations can, as before, be determined.¹⁵ Even in the face of heavy contamination, this approach can be very effective at recovering the original, underlying correlation.

8.2. Implementation

In this section, we describe how M parameter values, $\{\theta\}_j$, can be calculated from M measurements, both in the simplest $M > 1$ case of a linear model (see §8.2.1), and in general (see §8.2.2).¹²

We also describe how uncertainties, and hence weights, can be calculated for each of these M parameter values.

¹⁴ In this case, iteration ends either: (1) as before, if the next intersection would be unchanged (§2.1, §6), or (2) if the next intersection would be null.

¹⁵ More model parameters means more degrees of freedom, and consequently artificially smaller deviations for the same number of measurements. To correct for this, we multiply our correction factors (Figures 29 and 31) by (from Equation 8):

$$\sqrt{\left(\sum_i w_i - \Delta \frac{\sum_i w_i^2}{\sum_i w_i} \right) / \left(\sum_i w_i - M \Delta \frac{\sum_i w_i^2}{\sum_i w_i} \right)} \quad (24)$$

where the sums are over the non-rejected measurements. (For N unweighted measurements, this corresponds to dividing by $\sqrt{N-M}$, instead of by $\sqrt{N-1}$, when calculating a (two-sided) standard deviation, but this correction applies to our 68.3-percentile deviation calculations as well.) This also prevents over-rejection rates (§3.3.1) from increasing with M .

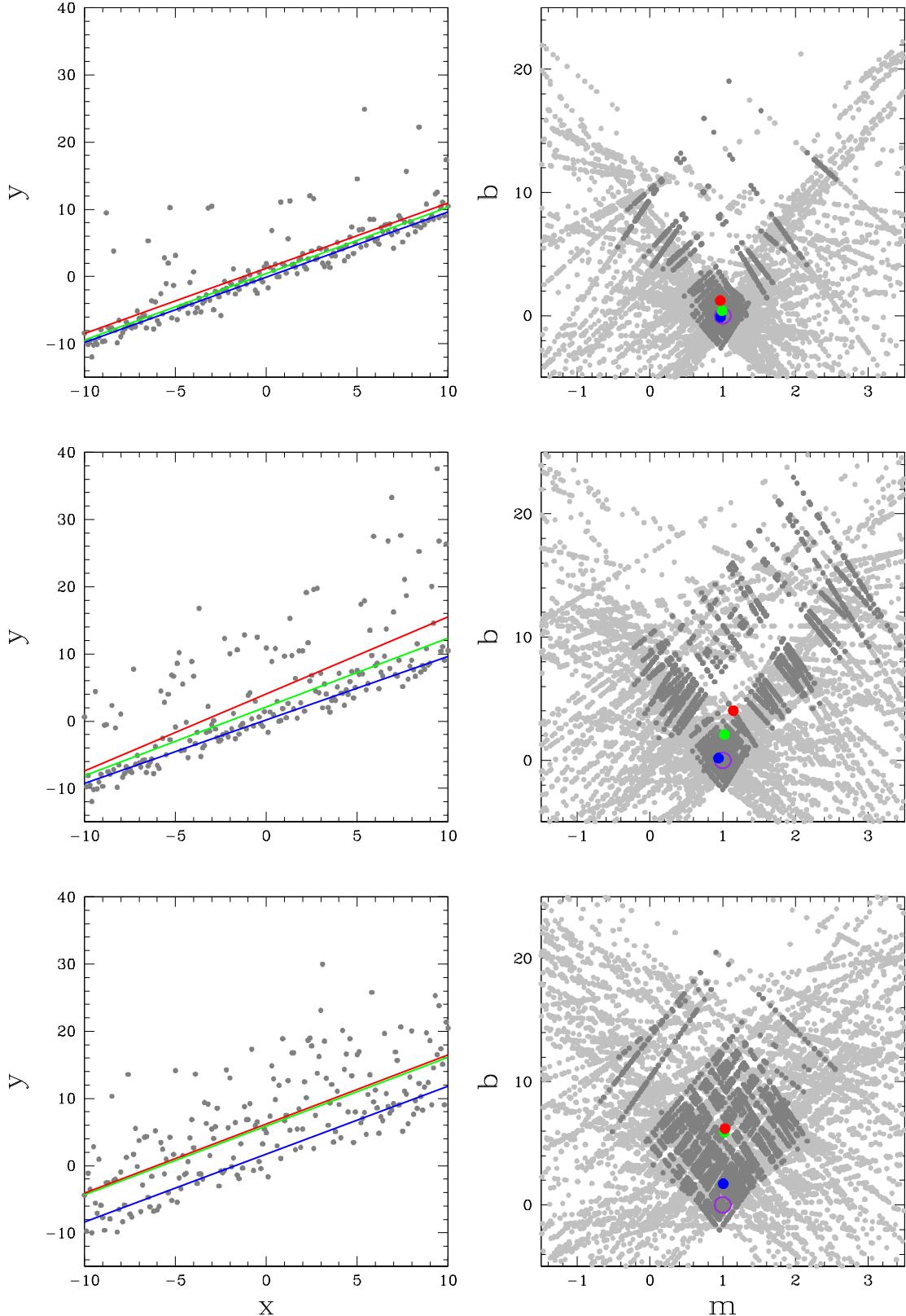


FIG. 34.— **Left column:** 201 measurements, with fraction $f_1 = 1 - f_2$ drawn from a Gaussian distribution of mean $y(x) = x$ and standard deviation 1, and fraction $f_2 = 0.15$ (top row), 0.5 (middle row), and 0.85 (bottom row), representing contaminated measurements, drawn from the positive side of a Gaussian distribution of mean zero and standard deviation 10, and added to uncontaminated measurements, drawn as above. **Right column:** Model solutions, $\{\theta\}_j$, calculated from each pair of measurements in the panel to the left, using $y(x) = b + m(x - \bar{x})$, with $\bar{x} = \sum_i w_i x_i / \sum_i w_i$ (see §8.3.5), and for model parameters b and m . Each calculated parameter value is weighted (see §8.2.1 or §8.2.2), and darker points correspond to models where the product of these weights is in the top 50%. The purple circle corresponds to the original, underlying model, and in both columns, blue corresponds to the weighted mode of $\{\theta\}_j$, green corresponds to the weighted median of $\{\theta\}_j$, and red corresponds to maximum-likelihood model fitting. The weighted mode of $\{\theta\}_j$ performs the best, especially in the limit of large f_2 . Maximum-likelihood model fitting performs the worst. See Figure 35 for post-rejection versions.

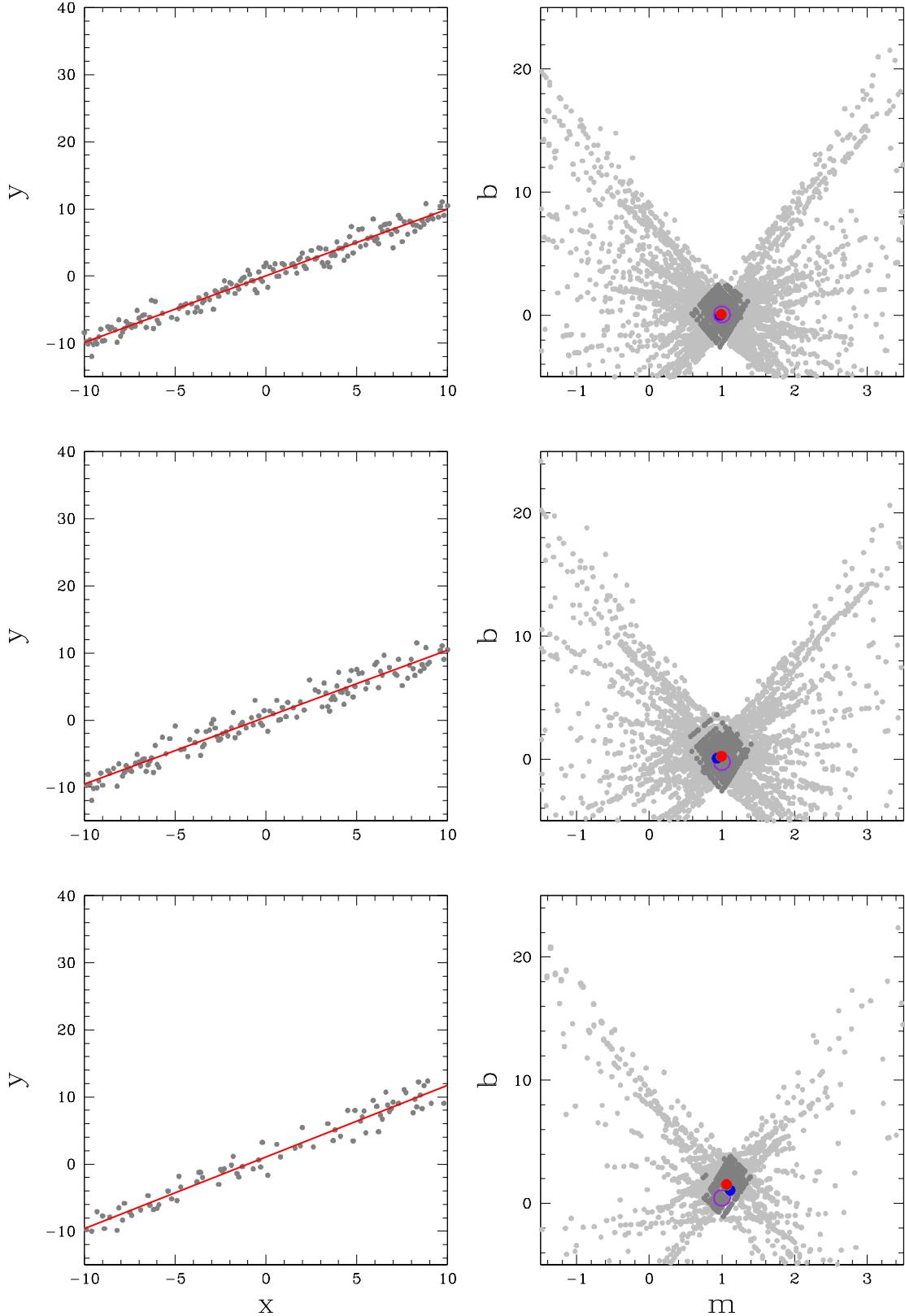


FIG. 35.—Figure 34, after RCR. Here, we have performed bulk rejection as in §5, but using our generalization of the mode instead of the mode, followed by individual rejection as in §4, using (1) our most-general robust technique for symmetrically distributed uncontaminated measurements – now consisting of our generalization of the mode + technique 3 (the broken-line fit) – followed by (2) our most-precise robust technique – now consisting of our generalization of the median + technique 1 (the 68.3% value) – followed by (3) traditional Chauvenet rejection, but using our generalization of the mean instead of the mean (e.g., Figures 30, 28, and 19). RCR proves effective, even in the face of heavy contamination.

This depends on the locations and weights of the M measurements, but it also depends on how one models their scatter, about the best-fit model to all of the non-rejected measurements. In §8.2.1 and §8.2.2, we present the simplest, and most common, model for this scatter, in which its RMS, at least for same-weight, uncontaminated measurements, is taken to be the same, or constant, at all locations, as it is in Figures 34 and 35. In §8.2.3, we consider non-constant RMS scatter, and present its most common case.

8.2.1. Simplest $M > 1$ Case: Linear Model with Constant RMS Scatter

Consider a linear model given by $y(x) = b + m(x - \bar{x})$. For any $M = 2$ of the N measurements, (x_1, y_1) and (x_2, y_2) , one can calculate $M = 2$ parameter values, given by:

$$m = \frac{y_2 - y_1}{x_2 - x_1} \quad (25)$$

and

$$b = \frac{x_2 - \bar{x}}{x_2 - x_1} y_1 - \frac{x_1 - \bar{x}}{x_2 - x_1} y_2. \quad (26)$$

The uncertainties in these values depend not only on the statistical uncertainties in y_1 and y_2 – which may or may not be known – but also on any systematic scatter in the measurements, at x_1 and x_2 .

Let $\sigma_y(x)$ be a to-be-specified model for the RMS scatter (statistical and/or systematic) of *average-weight*, uncontaminated measurements, about the best-fit model to all of the non-rejected measurements.

In the limit that $\sigma_y(x)$ is purely statistical, the RMS scatter of *any-weight*, uncontaminated measurements is then given by $(\bar{w}/w)^{1/2}\sigma_y(x)$, where w is measured weight, and \bar{w} is the average value of w for the uncontaminated measurements.

In the limit that $\sigma_y(x)$ is purely systematic, statistical error bars, and hence measured weights, do not matter, and consequently, an unweighted fit should be performed instead.¹⁶ Note, the same expression may be used for the RMS scatter, but in this case, all measured weights should be reset to a common value, such as $w = \bar{w} = 1$.¹⁷

Given this expression for the RMS scatter, Equations 25 and 26, and standard propagation of uncertainties, the uncertainties in m and b are then given by:

$$\sigma_m = \sqrt{\frac{\frac{\bar{w}\sigma_y^2(x_2)}{w_2} + \frac{\bar{w}\sigma_y^2(x_1)}{w_1}}{(x_2 - x_1)^2}} \quad (27)$$

¹⁶ Note, this is as much the case in §6 as it is here.

¹⁷ If in-between these two limiting cases, with statistical uncertainty greater than systematic scatter for some measurements, and less than it for the rest, one should also perform an unweighted fit. In this case, most measurements with statistical uncertainty \gg systematic scatter will be rejected as outlying (e.g., as outliers were rejected in Figure 35), but since these measurements are, by definition, of low measured weight, they were not going to significantly impact the fit anyway. However, if statistical uncertainties are known, one could then calculate new weights, given by $\left\{1 + [\sigma_i/\sigma_{sys}(x_i)]^2\right\}^{-1}$, where $\sigma_{sys}(x)$ is the RMS scatter of the non-rejected measurements about the unweighted fit, and then perform a weighted fit.

and

$$\sigma_b = \sqrt{\frac{\frac{\bar{w}\sigma_y^2(x_1)}{w_1}(x_2 - \bar{x})^2 + \frac{\bar{w}\sigma_y^2(x_2)}{w_2}(x_1 - \bar{x})^2}{(x_2 - x_1)^2}}. \quad (28)$$

Consequently, we weight m by $w_m \propto \sigma_m^{-2}$ and b by $w_b \propto \sigma_b^{-2}$. Since \bar{w} factors out, and is constant, it can be ignored.

In the simplest, and most common, case, $\sigma_y(x) = \sigma_y$ is also constant, as it is in Figures 34 and 35. In this case, it also factors out and can be ignored, yielding weights for m and b that depend only on the locations and weights of the $M = 2$ measurements from which they were calculated:¹⁸

$$w_m \propto \frac{(x_2 - x_1)^2}{w_1^{-1} + w_2^{-1}} \quad (29)$$

and

$$w_b \propto \frac{(x_2 - x_1)^2}{\frac{(x_1 - \bar{x})^2}{w_2} + \frac{(x_2 - \bar{x})^2}{w_1}}. \quad (30)$$

And again, (1) if all N of the measurements have the same weight, and/or (2) if σ_y is dominated by systematic scatter, these equations simplify even further, with $w_1 = w_2 = 1$.

Note, if a parameter’s value is known to be more or less probable a priori – i.e., if there is a prior probability distribution for that parameter – the $N!/[M!(N - M)!]$ weights that we calculate for that parameter (given by, e.g., Equation 29 or 30) should be multiplied by the prior probabilities of the $N!/[M!(N - M)!]$ values that we calculate for that parameter (given by, e.g., Equation 25 or 26), respectively, to up- or down-weight them accordingly, before calculating their generalized median or mode (§8.1).

8.2.2. General Case

Although many models can be solved for their M parameters analytically, given M measurements (e.g., as the linear model in §8.2.1 is solved for m and b , given two measurements), many models cannot be solved analytically. And even if a model can be solved analytically, this is not always easy to do, nor can all solvable models be anticipated in advance. Consequently, in general, we do this numerically, using the Gauss-Newton algorithm,¹⁹ which requires only that the user supply (1) the model, (2) its first partial derivative with respect to each model parameter, to construct its Jacobian, and (3) an initial guess, which usually has no bearing on the end result (however, see §8.3.3).

Furthermore, the uncertainty, σ_{θ_i} , in each calculated parameter value, θ_i , is straightforward to calculate, from

¹⁸ With non-linear models, \bar{w} and σ_y (if constant) also factor out and can be ignored. However, these weights, on the calculated parameter values, can also depend on the model parameters themselves (see §8.2.2).

¹⁹ With one modification: Each time an iteration results in a poorer fit, (1) we do not apply the increment vector, and (2) we shrink it by 50% in future iterations. This helps to ensure local convergence, in the case of periodic models (see §8.3.3).

the same matrix that lies at the heart of the Gauss-Newton algorithm, which, when $N = M$, is simply the inverse Jacobian, \mathcal{J}^{-1} .

Let $\vec{\sigma}_y = (\sigma_{y_1}, \dots, \sigma_{y_{N=M}})$ be an array of hypothetical errors in each of the M measurements, each drawn from a Gaussian of mean zero and standard deviation $(\bar{w}/w_i)^{1/2}\sigma_y(\{x\}_i)$ (§8.2.1). This corresponds to an array of errors in the calculated parameter values, given by $\mathcal{J}^{-1}\vec{\sigma}_y$. Next, imagine repeating these draws, and recalculating $\mathcal{J}^{-1}\vec{\sigma}_y$, an infinite number of times. Each σ_{θ_i} is then given by the RMS of these arrays' i th values. Mathematically, this is straightforward to calculate, and is equivalent to setting each $\sigma_{y_i} = (\bar{w}/w_i)^{1/2}\sigma_y(\{x\}_i)$ and calculating $(\sigma_{\theta_1}, \dots, \sigma_{\theta_M}) = \mathcal{J}^{-1}\vec{\sigma}_y$, except that terms in this matrix-vector multiplication are instead summed in quadrature (it is not difficult to show that in the case of the linear model of §8.2.1, this yields Equations 27 and 28.)

And as in §8.2.1, the weight of each calculated parameter value is then given by $w_{\theta_i} \propto \sigma_{\theta_i}^{-2}$.

And as in Equations 29 and 30, \bar{w} again factors out, and since constant, can be ignored. Likewise, if $\sigma_y(\{x\})$ can again be modeled as constant, it too factors out and can be ignored. (If $\sigma_y(\{x\})$ is not constant, it may be a function of $\{x\}$, as well as of the model parameters; we offer a common example in §8.2.3.)

However, unlike in Equations 29 and 30, and regardless of how $\sigma_y(\{x\})$ is modeled, each w_{θ_i} may now depend on the model parameters (through \mathcal{J}^{-1}). Note however, when calculating these weights, we do not use the calculated parameter values, $\{\theta\}_j$, from the corresponding M -measurement combination. Rather, we use those of the most recent baseline model, determined from taking the generalized mode, the generalized median, or the generalized mean of $\{\{\theta\}_j\}$ in the most recent iteration of the RCR algorithm (e.g., Figures 30 and 28). This should be a significantly more accurate representation of the underlying model than any individual $\{\theta\}_j$.

We also use these, significantly more accurate, parameter values as the starting point for the Gauss-Newton algorithm in the next iteration of the RCR algorithm. Only the starting point for the very first iteration need be supplied by the user.²⁰

8.2.3. Non-Constant RMS Scatter: Logarithmic Case

In general, $\sigma_y(\{x\})$ may not be constant, in which case a model must be provided for it by the user, just as a model must be provided for $y(\{x\})$ by the user. With no additional work, we can support models for $\sigma_y(\{x\})$ that are proportional to any function of (1) the independent

²⁰ As stated above, for most applications, the Gauss-Newton algorithm yields the same result, $\{\theta\}_j$, regardless of the initial guess. However, the generalized mode, median, or mean of $\{\{\theta\}_j\}$ also depends on each $\{\theta\}_j$'s corresponding weight, which does depend on the initial guess. Consequently, before beginning the RCR algorithm, and bulk rejecting outliers, we iteratively measure the generalized mode of $\{\{\theta\}_j\}$, without rejecting measurements, and with each iteration implying new weights for $\{\{\theta\}_j\}$, until we converge, from the user's initial guess, to a starting point for the RCR algorithm that is maximally consistent with the measurements.

variables, $\{x\}$, as well as (2) the model parameters. This is because we already support dependencies on both of these in the inverse Jacobian (§8.2.2). (As with σ_y in §8.2.1 and §8.2.2, the constant of proportionality factors out and can be ignored.)

How one models $\sigma_y(\{x\})$ depends on the problem at hand. As stated above, $\sigma_y(\{x\})$ can usually be modeled as constant and ignored. However, another common case arises when the user has a model $y(\{x\})$ that can be linearized. For example, exponential and power-law models can be linearized by taking a logarithm of both sides: e.g., $y(x) = be^{m(x-\bar{x})}$ becomes $\ln y(x) = \ln b + m(x - \bar{x})$, and $y(x) = b(x/e^{\ln x})^m$ becomes $\ln y(x) = \ln b + m(\ln x - \bar{\ln x})$.

This of course is fine, and even preferable, if the RMS scatter about $\ln y(\{x\})$ can be modeled as constant: i.e., if $\sigma_{\ln y}(\{x\}) = \sigma_{\ln y}$. However, often $\sigma_y(\{x\}) = \sigma_y$ is constant, in which case $\sigma_{\ln y}(\{x\})$ is then not constant, and consequently must be modeled.

In this case, $\sigma_{+\ln y}(\{x\}) \approx \ln[y(\{x\}) + \sigma_y] - \ln y(\{x\}) \rightarrow \sigma_y/y(\{x\})$ and $\sigma_{-\ln y}(\{x\}) \approx \ln y(\{x\}) - \ln[y(\{x\}) - \sigma_y] \rightarrow \sigma_y/y(\{x\})$ when $\sigma_y \ll y(\{x\})$, and $\sigma_{+\ln y} \rightarrow \ln[\sigma_y/y(\{x\})]$ and $\sigma_{-\ln y} \rightarrow \infty$ when $\sigma_y \gg y(\{x\})$. Since the $\sigma_y \ll y(\{x\})$ measurements are the most informative, one can approximate $\sigma_{\ln y} \approx \sigma_y/y(\{x\})$, which, conservatively, underestimates the weights for the less-informative, $\sigma_y \gtrsim y(\{x\})$ measurements.

In other words, logarithmic compression of constant RMS scatter results in smaller RMS scatter, and hence higher weights, for high- $\ln y(\{x\})$ measurements, and larger RMS scatter, and hence lower weights, for low- $\ln y(\{x\})$ measurements.

In the case of the linearized exponential model, Equations 29 and 30 then become:

$$w_m \propto \frac{(x_2 - x_1)^2}{w_1^{-1}y^{-2}(x_1) + w_2^{-1}y^{-2}(x_2)} \quad (31)$$

and

$$w_b \propto \frac{(x_2 - x_1)^2}{\frac{(x_1 - \bar{x})^2}{w_2 y^2(x_2)} + \frac{(x_2 - \bar{x})^2}{w_1 y^2(x_1)}}, \quad (32)$$

and in the case of the linearized power-law model, they instead become:

$$w_m \propto \frac{(\ln x_2 - \ln x_1)^2}{w_1^{-1}y^{-2}(x_1) + w_2^{-1}y^{-2}(x_2)} \quad (33)$$

and

$$w_b \propto \frac{(\ln x_2 - \ln x_1)^2}{\frac{(\ln x_1 - \bar{\ln x})^2}{w_2 y^2(x_2)} + \frac{(\ln x_2 - \bar{\ln x})^2}{w_1 y^2(x_1)}}. \quad (34)$$

Note, these equations depend not only on the independent variable, x , but now also on the model parameters, m and b , through $y(x)$, and consequently are evaluated as prescribed in the second-to-last paragraph of §8.2.2.

However, although the linearization of these, and other, models allows their parameters to be determined

analytically, as in Equations 25 and 26, instead of numerically as in §8.2.2, this really does not gain the user anything, given the speeds of modern computers. Instead, when possible, we recommend either leaving one's model in, or transforming one's model to, whatever form yields constant, or near-constant, RMS scatter about its best fit to the non-rejected measurements, and then simply applying the all-purpose (linear and non-linear) machinery of §8.2.2.²¹

8.3. Considerations, Limitations, and Examples

In this section, we present a few additional considerations and limitations, and examples. In §8.3.1 – §8.3.3, we consider special cases that while they do not change how we calculate M -parameter solutions, $\{\theta\}_j$, from M measurements (§8.2.2), they can affect how we calculate the generalized mode, and sometimes also the generalized median, from a full set of $N!/[M!(N-M)!]$ M -parameter solutions, $\{\{\theta\}_j\}$ (§8.1). In §8.3.4, we show that RCR becomes less robust as M increases, and this appears to be a fundamental limitation of our approach. And finally, in §8.3.5 and §8.3.6, we discuss the importance of good modeling practices, both in general, but also specifically to RCR.

8.3.1. Measurements that Cannot Be Described by the Model

Due to statistical and/or systematic scatter, some combinations of M measurements, even M uncontaminated measurements, might not map to any combination of values for a model's M parameters.

For example, consider an exponential model that asymptotes from positive values to zero as $x \rightarrow \infty$ (e.g., $y(x) = be^{m(x-\bar{x})}$, with $b > 0$ and $m < 0$), but with measurements, y_i , that are occasionally negative due to statistical and/or systematic scatter. Since all combinations of values for b and m yield only-positive or only-negative values for $y(x)$, if presented with an oppositely signed pair of measurements, our Gauss-Newton algorithm (§8.2.2) will instead run away to one of the following, limiting solutions, depending on the values of x_i , y_i , and \bar{x} : $m = -\infty$ or ∞ , and $b = -\infty$, 0 , ∞ , or the value of the positive measurement (if x_i happens to equal \bar{x}). Note, such solutions are easily flagged, since the fitted model does not (cannot) pass through all M of the measurements (e.g., resulting in a non-zero χ^2 value).

Although extreme, and not fully representative of the measurements that produced them, we do not exclude such solutions when calculating the weighted median of

²¹ That said, both approaches usually yield near identical results. Modeling exponential or power-law data with parameters b and m , instead of linearized data with $\ln b$ and m , yields (1) a different inverse Jacobian (§8.2.2), and (2) a different model for the RMS scatter, $\sigma_y(x)$ vs. $\sigma_{\ln y}(x)$. But together these yield the same expressions for $w_b = w_{\ln b}$ and w_m (up to factors of proportionality that do not matter). Consequently, the only difference is how concentrated the calculated parameter values, $\{\{\theta\}_j\}$, are, which does not affect the weighted median of $\{\{\theta\}_j\}$, but can affect the weighted mode of $\{\{\theta\}_j\}$: Using b instead of $\ln b$ favors lower values, but usually only marginally. This is known as choice of basis, which we return to §8.3.6.

$\{\{\theta\}_j\}$: To do so could bias the result (in this particular example, toward higher values of b and shallower values of m). At the same time, we *do* exclude such solutions when calculating the weighted mode of $\{\{\theta\}_j\}$, lest any of these parameter values be returned artificially (e.g., in this case, they could result in a meaningless, but statistically significant, overdensity of $b = 0$ values).

We demonstrate RCR applied to such an exponential model in Figures 36 and 37, and despite a fair number of $y_i < 0$ measurements at high- x values, it converges to an acceptable solution in all but the most contaminated case.

8.3.2. Combinations of M Measurements with Redundant Independent-Variable Information

Combinations of M measurements with redundant independent-variable information cannot be used to determine all M of a model's parameters. Furthermore, if, in this case, any of the model's parameters can be determined, they will be overdetermined.

For example, consider a planar model, constrained by three measurements. If these measurements happen to be co-linear, all three of the model's parameters cannot be determined. However, if this line happens to run parallel to one of the model's axes, at least one, and possibly two, of the model's parameters (i.e., the plane's slope along this axis, and the plane's normalization, if defined along this line) can be determined. But they will be overdetermined, given three measurements for only one or two parameters.

In the interest of simplicity, we discard these (usually rare) combinations completely, noting that uncontaminated measurements selected in this way are unlikely to be preferentially under- or over-estimates, and consequently their exclusion is unlikely to bias calculation of the weighed median of $\{\{\theta\}_j\}$, let alone of the weighted mode of $\{\{\theta\}_j\}$. However, more sophisticated implementations can also be imagined.

Note, such cases are also easily flagged, in that the Jacobian in §8.2.2 is not invertible (i.e., its determinant is zero).

8.3.3. Combinations of M Measurements that Can Be Described by Multiple Model Solutions

Periodic models require a bit more care, in that each combination of M measurements can be described by a countably infinite number of model solutions, including not only solutions that are equivalent to each other, but also shorter-period, overtone solutions that are not. Both can bias calculation of the weighted median of $\{\{\theta\}_j\}$, and of the weighted mode of $\{\{\theta\}_j\}$.

For example, consider the simple, periodic model $y(x) = b \sin m(x - x_0)$. The same measurements can result in model solutions that are equivalent to each other (1) by reflection about both the x and y axes, (2) by translation along the x axis, by multiples of $2\pi/m$, and/or (3) by translation along the x axis by odd multiples of π/m , in combination with a reflection about the x axis. Consequently, once the Gauss-Newton al-

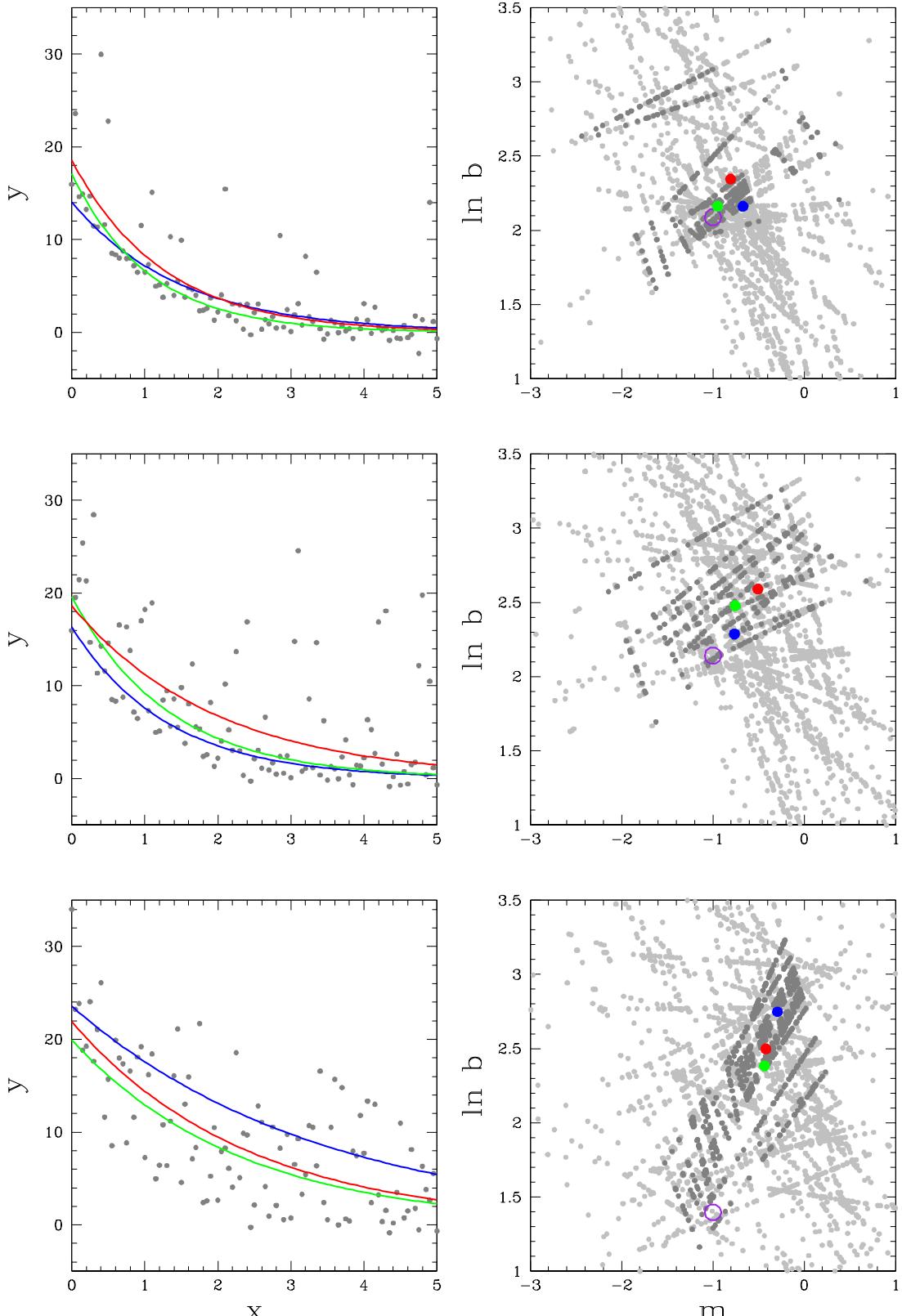


FIG. 36.— **Left column:** 101 measurements, with fraction $f_1 = 1 - f_2$ drawn from a Gaussian distribution of mean $y(x) = 10e^{-(x-0.5)}$ and standard deviation 1, and fraction $f_2 = 0.15$ (top row), 0.5 (middle row), and 0.85 (bottom row), representing contaminated measurements, drawn from the positive side of a Gaussian distribution of mean zero and standard deviation 10, and added to uncontaminated measurements, drawn as above. **Right column:** Model solutions, $\{\theta\}_j$, calculated from each pair of measurements in the panel to the left, using $y(x) = be^{m(x-\bar{x})}$, with $\bar{x} = \sum_i w_i x_i y^2(x_i) / \sum_i w_i y^2(x_i)$ (see §8.3.5), and for model parameters $\ln b$ and m (see §8.3.6). Each calculated parameter value is weighted (§8.2.2), and darker points correspond to models where the product of these weights is in the top 50%. The purple circle corresponds to the original, underlying model, and in both columns, blue corresponds to the weighted mode of $\{\theta\}_j$, green corresponds to the weighted median of $\{\theta\}_j$, and red corresponds to maximum-likelihood model fitting. The contaminants have a greater, relative, effect on the high- x /low- y measurements than on the low- x /high- y measurements, biasing the calculated models toward shallower slopes and higher normalizations (i.e., toward the upper right, in the panels on the right). The weighted mode of $\{\theta\}_j$ most successfully overcomes this bias as $f_2 \rightarrow 0.5$, but all three techniques fail as $f_2 \rightarrow 0.85$. See Figure 37 for post-rejection versions.

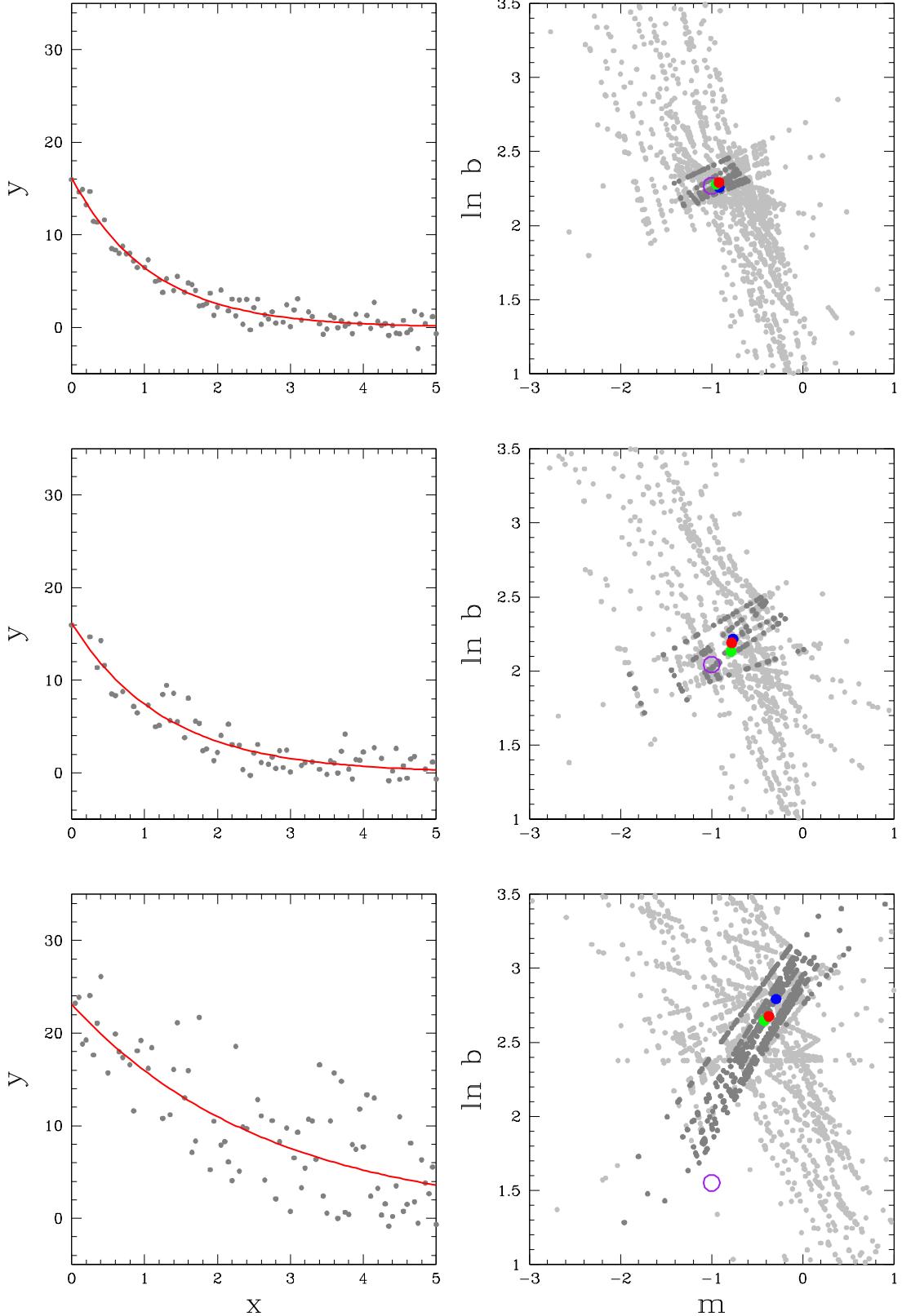


FIG. 37.— Figure 36, after RCR. Here, we have again performed bulk rejection as in §5, but using our generalization of the mode instead of the mode, followed by individual rejection as in §4, using (1) our most-general robust technique for symmetrically distributed uncontaminated measurements – now consisting of our generalization of the mode + technique 3 (the broken-line fit) – followed by (2) our most-precise robust technique – now consisting of our generalization of the median + technique 1 (the 68.3% value) – followed by (3) traditional Chauvenet rejection, but using our generalization of the mean instead of the mean (e.g., Figures 30, 28, and 19). RCR proves effective in the face of fairly heavy contamination, but is unable to overcome bias introduced by the contaminants (Figure 35) as $f_2 \rightarrow 0.85$.

gorithm (§8.2.2) finds one of these solutions, we give the user the option to map it to a designated simplest form. For example, with this model: (1) If $m < 0$, map $m \rightarrow -m$ and $b \rightarrow -b$; (2) then if $m|x_0| \geq 2\pi$, map $x_0 \rightarrow x_0 - \frac{2\pi x_0}{m|x_0|} \text{floor}\left(\frac{m|x_0|}{2\pi}\right)$; and (3) then if $m|x_0| \geq \pi$, map $x_0 \rightarrow x_0 - \frac{\pi x_0}{m|x_0|}$ and $b \rightarrow -b$.

In the case of shorter-period/higher- m , overtone solutions, which solution the Gauss-Newton algorithm finds depends on the initial guess that it is given. This is analogous to centroiding algorithms in astrometry. If a user clicks anywhere in a star's vicinity, such algorithms arrive at the same solution for the star's center. But if the user clicks too far away, another star's center will be found instead. We have modified the Gauss-Newton algorithm to help ensure local convergence (Footnote 19), but ultimately it is up to the user to make a reasonable (in this case, low- m) initial guess.

We demonstrate RCR applied to this model in Figures 38 and 39, using the same contamination fractions as in Figures 34 – 37. The combination of re-mapping equivalent solutions, and of making a reasonable initial guess, results in good outcomes through fairly high contamination fractions (however, see §8.3.4).

8.3.4. RCR Less Robust as M Increases

If a fraction, $1 - f$, of N measurements is uncontaminated, a smaller fraction, $(1 - f)^M$, of the corresponding $N! / [M!(N - M)!]$ model solutions, $\{\{\theta\}_j\}$, is uncontaminated. So, the higher the dimension of the model, and hence of the model's parameter space, the more difficult it becomes for our generalization of the mode, in particular, to latch on to a desirable solution. Or to put it another way, the higher M , the lower f beyond which RCR fails. This appears to be a fundamental limitation of our approach, and one that can be only partially mitigated by a (significantly) larger number of measurements.²²

This can be seen by the greater degree of scatter in the $M = 3$ parameter-space plots in Figure 38, compared to that of the $M = 2$ parameter-space plots in Figures 34 and 36, and by the fact that this greater degree of scatter could not be successfully resolved in the $f = 0.85$ row in Figure 39, despite the contaminants not biasing the calculated parameter values in a systematic direction, as they did in Figures 36 and 37. (See Figures 41 and 42 for another $M = 3$ example, with similar results.)

8.3.5. Avoid Introducing Unnecessary Correlations between Calculated Parameter Values through Good Model Design

Naturally, our generalization of the mode, in particular, is most effective if the uncontaminated subset of $\{\{\theta\}_j\}$ is maximally concentrated. However, this can depend on how wisely, or poorly, one constructs their model.

²² Other approaches can be envisioned, in which combinations of more than M measurements are used to calculate model solutions, with RCR employed at this stage as well, to reduce the fraction of these that are contaminated. However, this is beyond the scope of this paper.

For example, consider a linear model, given by $y(x) = b + m(x - \bar{x})$, with constant RMS scatter, $\sigma_y(x) = \sigma_y$. In this case, \bar{x} is usually given by $\bar{x} = \sum_i w_i x_i / \sum_i w_i$, which results in a largely uncorrelated, near-maximally concentrated distribution of, at least the highest-weight, b vs. m values (e.g., Figures 34 and 35).²³

However, a significantly different choice for \bar{x} would introduce a correlation between the calculated values of b and m , resulting in a dispersed, and hence not near-maximally concentrated, distribution (we demonstrate this for a different, but similar, case in the bottom row of Figure 40; see below). This can make our generalization of the mode, in particular, and hence RCR, less precise, and in this case, unnecessarily.

Note, this is not always the best expression for \bar{x} . For example, consider either an exponential model, given by $y(x) = be^{m(x-\bar{x})}$, or a power-law model, given by $y(x) = b\left(x/e^{\ln x}\right)^m$, with constant RMS scatter, $\sigma_y(x) = \sigma_y$. If $m < 0$, high- x measurements may be scatter-dominated and not contribute significantly to the fit, and consequently should not contribute significantly to \bar{x} (and vice versa if $m > 0$, with low- x measurements). However, if linearized (§8.2.3), resulting in $\ln y(x_i)$ vs. x_i data for the exponential model and $\ln y(x_i)$ vs. $\ln x_i$ data for the power-law model, all measurements would contribute to the fit, but with additional weights given by $\sigma_{\ln y}^{-2}(x_i) \propto y^2(x_i)$ (§8.2.3). Hence, we take $\bar{x} = \sum_i w_i x_i y^2(x_i) / \sum_i w_i y^2(x_i)$ for the exponential model, and $\bar{\ln x} = \sum_i w_i (\ln x_i) y^2(x_i) / \sum_i w_i y^2(x_i)$ for the power-law model (whether the model has been linearized or not).²⁴ We demonstrate the effectiveness of this in Figure 40.

Sometimes, however, these correlations cannot be avoided. For example, if presented with quadratic y_i vs. x_i data, one can design away correlations between two of the three pairings of the model's three parameters, but not between all three pairings simultaneously: If one models these data with $y(x) = b + m_1(x - \bar{x}) + m_2(x - \bar{x})^2$, with $\bar{x} = \sum_i w_i x_i / \sum_i w_i$, both (1) the highest-weight b vs. m_1 values and (2) the highest-weight m_1 vs. m_2 values will, for the most part, be uncorrelated, but the highest-weight b vs. m_2 values will be marginally (negatively) correlated (see Figure 41). Despite this, RCR is still effective through fairly high contamination fractions, which we demonstrate in Figure 42.

Of course, more sophisticated implementations can be imagined, in which one would not have to consider these correlations at all. For example, instead of determining (1) the weighted median of $\{\{\theta\}_j\}$ and especially (2) the weighted mode of $\{\{\theta\}_j\}$ along the given parameter-space coordinate system, as we do in this paper (§8.1), one could imagine doing this (or perhaps

²³ We calculate \bar{x} using only non-rejected measurements, and consequently, we update \bar{x} after each iteration of the RCR algorithm.

²⁴ Here, \bar{x} and $\bar{\ln x}$ additionally depend on model parameters, through $y(x)$. But as we do when calculating parameter weights in §8.2.3, we use the parameter values of the most recent baseline model, from the most recent iteration of the RCR algorithm. This should be significantly more accurate than using, say, individual y_i measurements for $y(x_i)$.

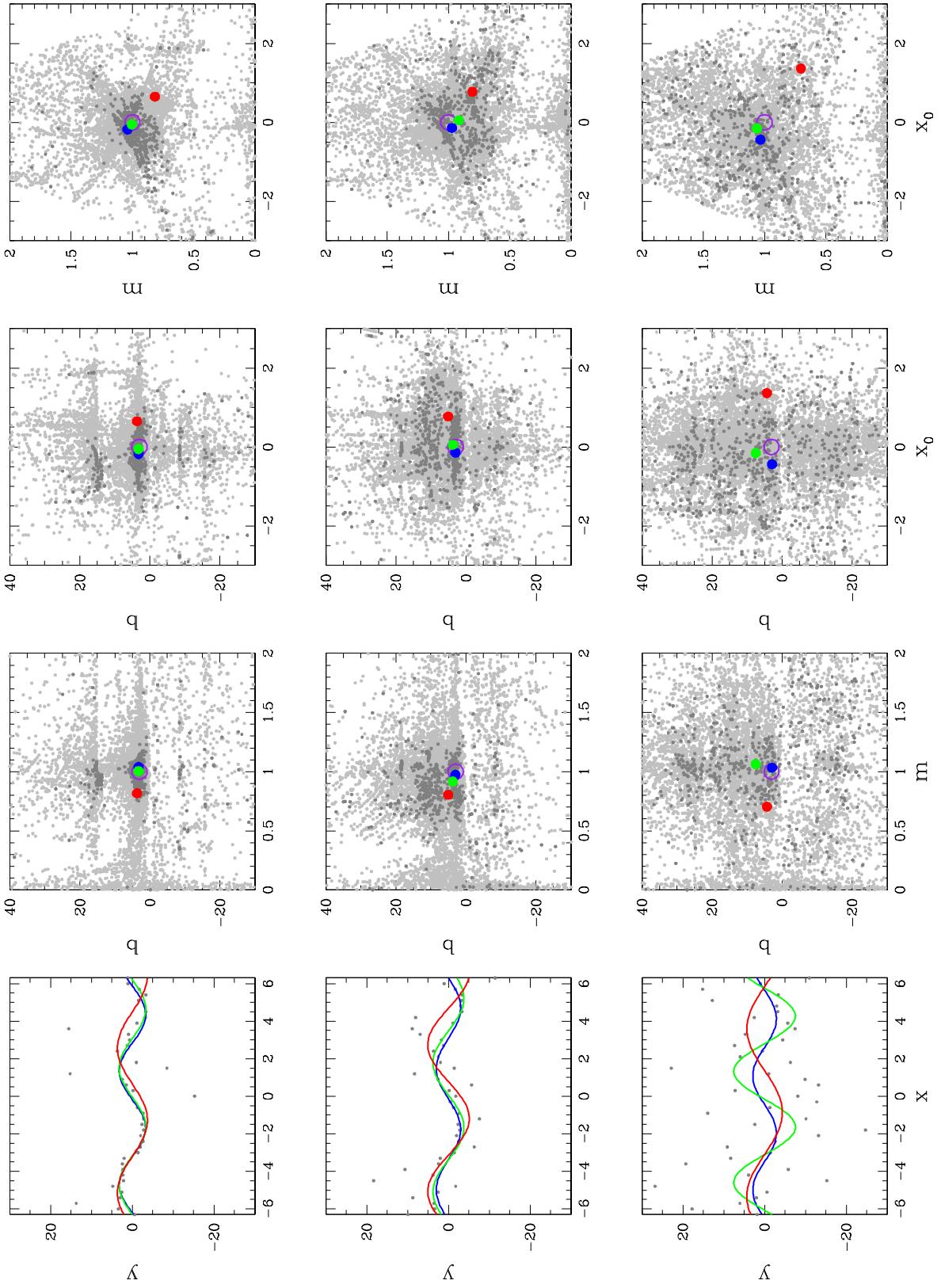


Fig. 38.— **Left column:** 43 measurements, with fraction $f_1 = 1 - f_2$ drawn from a Gaussian distribution of mean $y(x) = 3 \sin x$ and standard deviation $f_2 = 0.15$ (top row), 0.5 (middle row), and 0.85 (bottom row), representing contaminated measurements, drawn from a Gaussian distribution of mean zero and standard deviation 10 , and added to uncontaminated measurements, drawn as above. **Right columns:** Model solutions, $\{\theta_j\}$, calculated from each triplet of measurements in the panel to the left, using $y(x) = b \sin m(x - x_0)$, for model parameters b, m , and x_0 . Each calculated parameter value is weighted (§8.2.2), and darker points correspond to models where the product of these weights is in the top 50%. The purple circle corresponds to the original, underlying model, and in all columns, blue corresponds to the weighted mode of $\{\theta_j\}$; green corresponds to the weighted median of $\{\theta_j\}$, and red corresponds to maximum-likelihood model fitting. The weighted mode of $\{\theta_j\}$ performs the best, especially in the limit of large f_2 . Maximum-likelihood model fitting performs the worst. See Figure 39 for post-rejection versions.

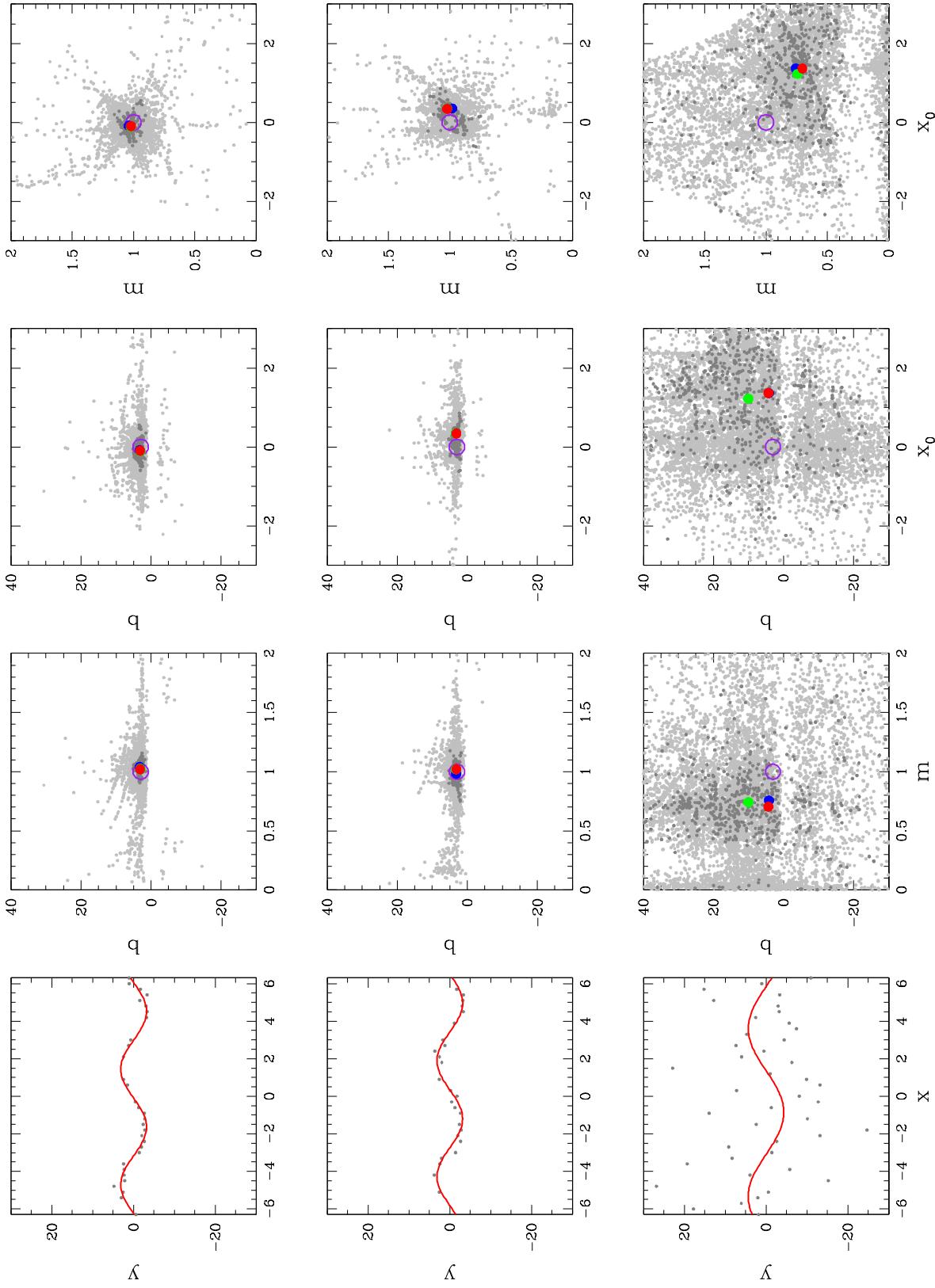


FIG. 39.— Figure 38, after RCR. Here, we have again performed bulk rejection as in §5, but using our generalization of the mode instead of the mean, followed by individual rejection as in §4, using (1) our most-general robust technique for symmetrically distributed uncontaminated measurements – now consisting of our generalization of the mode + technique 3 (the broken-line fit) – followed by (2) our most-precise robust technique – now consisting of our generalization of the mean + technique 1 (the 68.3% value) – followed by (3) traditional Chauvenet rejection, but using our generalization of the mean instead of the mean (e.g., Figures 30, 28, and 19). RCFR proves effective in the face of fairly heavy contamination, but is unable to overcome the greater fraction of contaminated model solutions (see §8, 3.4) as $f_2 \rightarrow 0.85$: $1 - (1 - 0.85)^2 = 0.996625$ for $M = 3$ v.s. $1 - (1 - 0.85)^2 = 0.9775$ for $M = 2$.

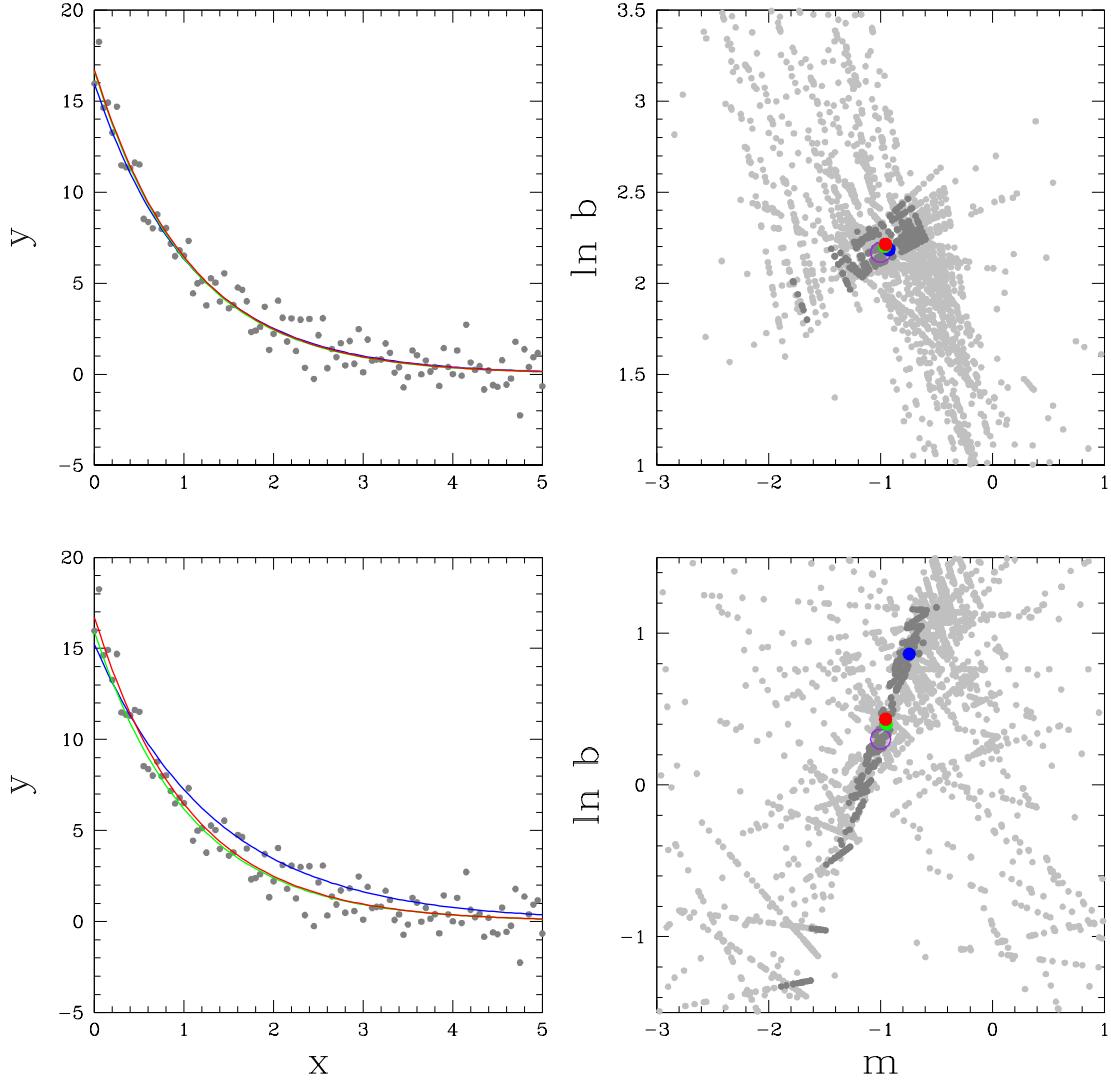


FIG. 40.— **Left column:** 101 uncontaminated measurements, drawn from a Gaussian distribution of mean $y(x) = 10e^{-(x-0.5)}$ and standard deviation 1. **Right column:** Model solutions, $\{\theta\}_j$, calculated from each pair of measurements in the panel to the left, using $y(x) = be^{m(x-\bar{x})}$, with \bar{x} calculated as described below, and for model parameters $\ln b$ and m (see §8.3.6). Each calculated parameter value is weighted (§8.2.2), and darker points correspond to models where the product of these weights is in the top 50%. The purple circle corresponds to the original, underlying model, and in both columns, blue corresponds to the weighted mode of $\{\theta\}_j$, green corresponds to the weighted median of $\{\theta\}_j$, and red corresponds to maximum-likelihood model fitting. **Top row:** Here, we additionally weight each x_i by $\sigma_{\ln y}^{-2}(x_i) \propto y^2(x_i)$ when calculating \bar{x} , which results in a fairly uncorrelated/fairly concentrated distribution for the highest-weight $\ln b$ vs. m values, and consequently, the weighted mode of $\{\theta\}_j$, in particular, is less susceptible to imprecision. **Bottom row:** Here, we do not additionally weight each x_i when calculating \bar{x} , which results in a strongly correlated/dispersed distribution for the highest-weight $\ln b$ vs. m values, and consequently, the weighted mode of $\{\theta\}_j$, in particular, is more susceptible to imprecision.

something else a bit more sophisticated) in a rotated, or even non-linearly transformed, coordinate system, with a principal axis determined (robustly) from the calculated parameter values and weights. This is beyond the scope of the current work, but would be a natural next investigation.

8.3.6. Make Good Basis Decisions

Another example of good model design is proper choice

of basis. For example, when fitting an exponential model, e.g., $y(x) = be^{m(x-\bar{x})}$, or a power-law model, e.g., $y(x) = b \left(x/e^{\ln x} \right)^m$, to measurements, one usually calculates $\ln b$ and m , instead of b and m (e.g., Figures 36, 37, and 40). This is called choice of basis. When performing maximum-likelihood model fitting, basis choices (like normalization choices; §8.3.5) do not affect the best fit, but good ones can yield more concentrated/symmetric

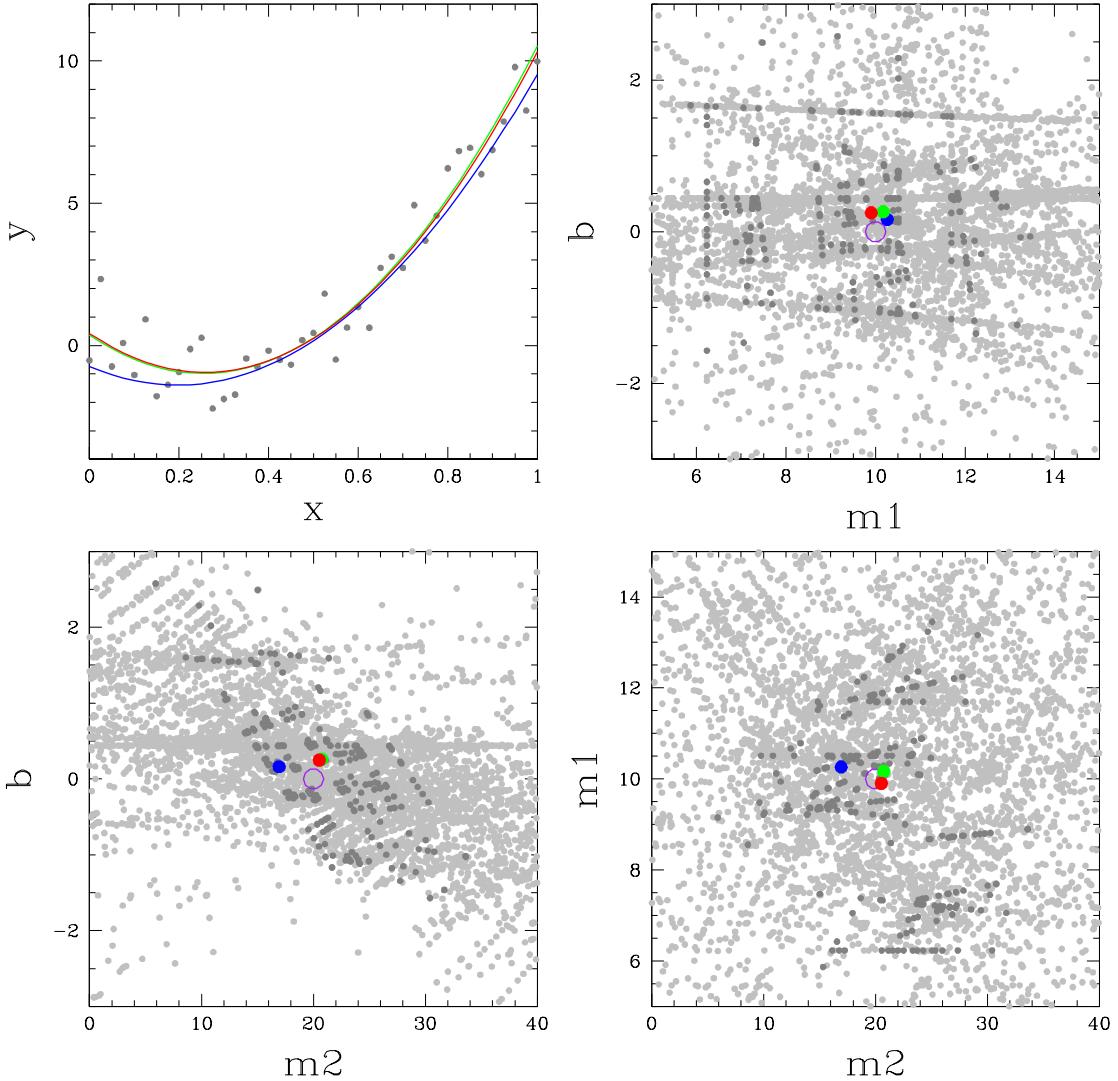


FIG. 41.— **Upper left:** 43 uncontaminated measurements, drawn from a Gaussian distribution of mean $y(x) = 10(x - 0.5) + 20(x - 0.5)^2$ and standard deviation 1. **Upper right and bottom row:** Model solutions, $\{\theta\}_j$, calculated from each triplet of measurements in the upper-left panel, using $y(x) = b + m_1(x - \bar{x}) + m_2(x - \bar{x})^2$, with $\bar{x} = \sum_i w_i x_i / \sum_i w_i$, and model parameters b , m_1 , and m_2 . Each calculated parameter value is weighted (§8.2.2), and darker points correspond to models where the product of these weights is in the top 50%. The purple circle corresponds to the original, underlying model, and in all panels, blue corresponds to the weighted mode of $\{\{\theta\}_j\}$, green corresponds to the weighted median of $\{\{\theta\}_j\}$, and red corresponds to maximum-likelihood model fitting. The highest-weight b vs. m_1 values, corresponding to linear $y_{eff}(x) \equiv y(x) - m_2(x - \bar{x})^2 = b + m_1(x - \bar{x})$, and the highest-weight m_1 vs. m_2 values, corresponding to linear $y_{eff}(x) \equiv [y(x) - b]/(x - \bar{x}) = m_1 + m_2(x - \bar{x})$, are largely uncorrelated, but the highest-weight b vs. m_2 values, corresponding to non-linear $y_{eff}(x) \equiv y(x) - m_1(x - \bar{x}) = b + m_2(x - \bar{x})^2$, are marginally, negatively correlated: Since $(x - \bar{x})^2$ is always positive, if m_2 is high, b tends to be low, to compensate.

probability distributions for the model's parameters, and consequently, more concentrated/symmetric error bars for these parameters.

Similarly, good basis choices can yield more concentrated/symmetric distributions of calculated parameter values, $\{\{\theta\}_j\}$. While this does not affect the weighed median of $\{\{\theta\}_j\}$, as in §8.3.5, it can affect the weighted mode of $\{\{\theta\}_j\}$.

That said, as long as one's uncontaminated measurements are not scatter-dominated, this is usually a very small effect, and multiple, equivalent parameterizations are perfectly acceptable.²⁵

Application of RCR to parameterized models is potentially a very broad topic, with applications spanning not only science, but all quantitative disciplines. Here, we have but scratched the surface with a few, simple

²⁵ Another example is using $\tan m$ instead of m for slopes when they are very large.

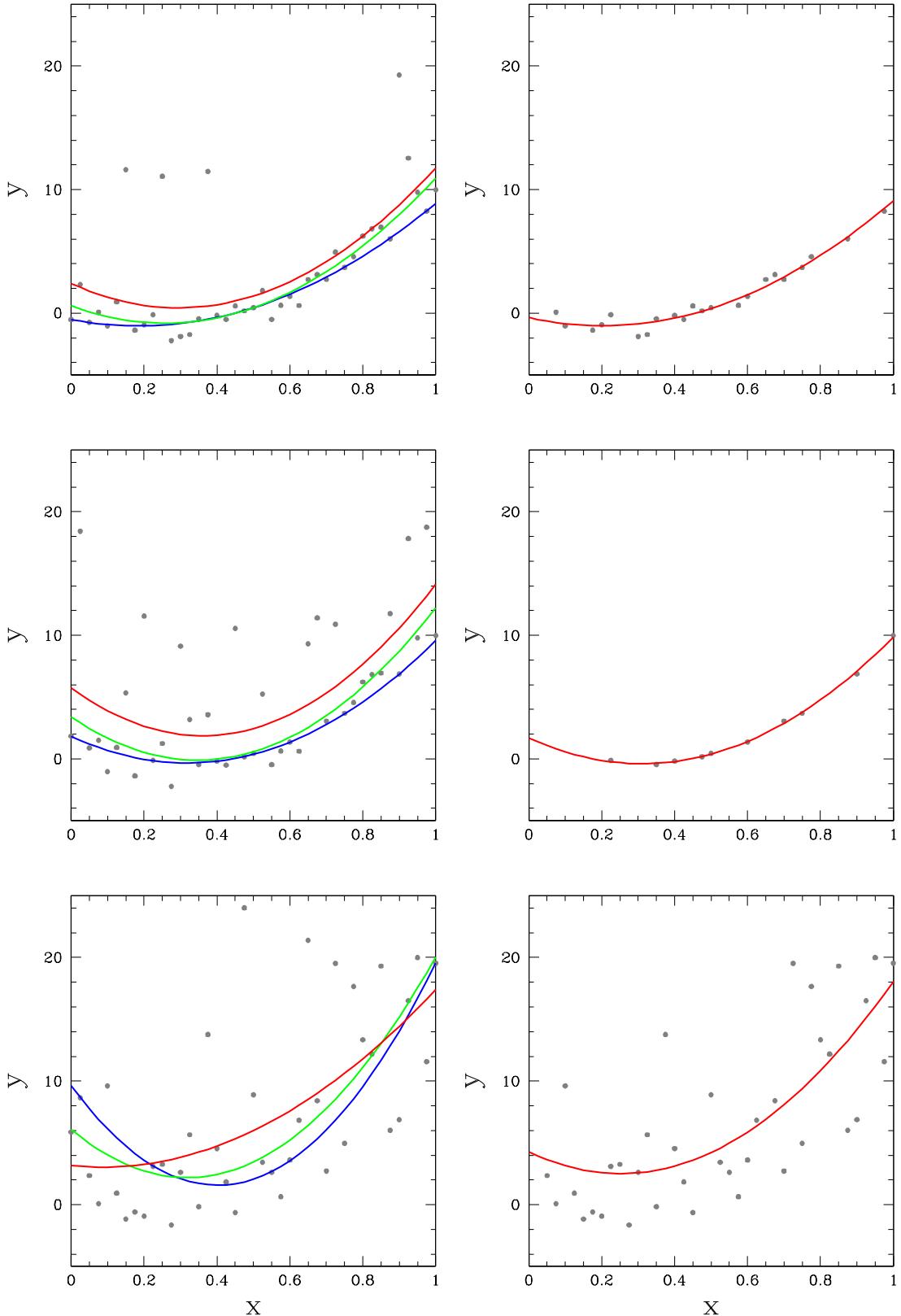


FIG. 42.— **Left column:** 43 measurements, with fraction $f_1 = 1 - f_2$ drawn from a Gaussian distribution of mean $y(x) = 10(x - 0.5) + 20(x - 0.5)^2$ and standard deviation 1, and fraction $f_2 = 0.15$ (top row), 0.5 (middle row), and 0.85 (bottom row), representing contaminated measurements, drawn from the positive side of a Gaussian distribution of mean zero and standard deviation 10, and added to uncontaminated measurements, drawn as above. Blue corresponds to the weighted mode of $\{\{\theta\}_j\}$, green corresponds to the weighted median of $\{\{\theta\}_j\}$, and red corresponds to maximum-likelihood model fitting. **Right column:** After RCR. Here, we have again performed bulk rejection as in §5, but using our generalization of the mode instead of the mode, followed by individual rejection as in §4, using (1) our most-general robust technique for symmetrically distributed uncontaminated measurements – now consisting of our generalization of the mode + technique 3 (the broken-line fit) – followed by (2) our most-precise robust technique – now consisting of our generalization of the median + technique 1 (the 68.3% value) – followed by (3) traditional Chauvenet rejection, but using our generalization of the mean instead of the mean (e.g., Figures 30, 28, and 19). RCR proves effective in the face of fairly heavy contamination, but is unable to overcome the greater fraction of contaminated models as $f_2 \rightarrow 0.85$ (§8.3.4).

examples.

9. PEIRCE REJECTION

Traditional Chauvenet rejection is sigma clipping plus a rule for selecting a reasonable number of sigma for the threshold, given N measurements (§1). It is straightforward to use, and as such has been adopted as standard by many government and industry laboratories, and is commonly taught at universities (Ross 2003). However, it is not the only approach one might take to reject outliers. For example, even Chauvenet (1863) deferred to Peirce's approach (1852; Gould 1855),²⁶ which has recently seen new life with its own implementation in the R programming language (Dardis 2012). Instead of assuming 0.5 in Equation 1, Peirce derives this value from probability theory, and finds (1) that it is weakly dependent on N , asymptoting to 0.5 as N increases, and (2) that it decreases with subsequent rejections.

However, unlike Peirce's approach, Chauvenet rejection is amenable to (1) N , (2) the mean, and (3) the standard deviation being updated after each rejection (§1). Peirce's approach requires all three of these quantities to remain fixed until all rejections have been completed, and as such Peirce rejection is less robust than Chauvenet rejection, at least when the latter is implemented iteratively, as we have done.

Furthermore, our correction factors (§2.3, §4 – §6) empirically account for the above, weak dependence on N , as well as for differences in implementation (e.g., our use of one-sided deviation measurements, our use of robust quantities, etc.)

In Figures 43 – 46, we compare Peirce rejection (1) to traditional Chauvenet rejection and (2) to RCR, for both two-sided and one-sided contaminants. Given our iterative implementation, and our correction factors, we find traditional Chauvenet rejection to be comparable to Peirce rejection when the contaminants are two-sided and N is low, and better than Peirce rejection otherwise. RCR is significantly better than both of these approaches.

10. SUMMARY

The most fundamental act in science is measurement. By combining multiple measurements, one can better

constrain a quantity's true value, and its uncertainty. However, measurements, and consequently samples of measurements, can be contaminated. Here, we have introduced, and thoroughly tested, an approach that, while not perfect, is very effective at identifying which measurements in a sample are contaminated, even if they constitute most of the sample, and especially if the contaminants are strong (making contaminated measurements easier to identify).

In particular, we have considered:

- Both symmetrically (§3.1) and asymmetrically (§3.2) distributed contamination of both symmetrically (§3.1, §3.2, §3.3.2) and asymmetrically (§3.3.1) distributed uncontaminated measurements, and have developed robust outlier rejection techniques for all combinations of these cases.
- The tradeoff between these techniques' accuracy and precision, and have found that by applying them in sequence, from more robust to more precise, both can be achieved (§4).
- The practical cases of bulk rejection (§5), weighted data (§6), and model fitting (§8), and have generalized the RCR algorithm accordingly.

Finally, we have developed a simple web interface so anyone can use the RCR algorithm.²⁷ Users may upload a data set, and select from the above scenarios. They are returned their data set with outliers flagged, and with μ_1 and σ_1 robustly measured. Source code is available here as well.

We gratefully acknowledge the support of the National Science Foundation, through the following programs and awards: ESP 0943305, MRI-R² 0959447, AAG 1009052, 1211782, and 1517030, ISE 1223235, hbcu-up 1238809, TUES 1245383, and STEM+C 1640131. We are also appreciative to have been supported by the Mt. Cuba Astronomical Foundation, the Robert Martin Ayers Sciences Fund, and the North Carolina Space Grant Consortium. We also thank the referee, and the editor, for comments that helped us to improve this paper considerably.

APPENDIX

A. BROKEN-LINE FIT THROUGH ORIGIN

Let $x_i = \sqrt{2}\text{erf}^{-1}[(i - 0.317)/N]$ and $y_i = \delta_i$. We model these data as a broken line that passes through the origin:

$$y = \begin{cases} \sigma_1 x, & \text{if } i \leq m \\ \sigma_1 x_m + \sigma_2(x - x_m), & \text{if } i \geq m \text{ and } x_i \leq 1 \end{cases}, \quad (\text{A1})$$

where σ_1 is the slope of the line for $i \leq m$, and our modeled 68.3-percentile deviation, and σ_2 is the slope of the line for $i \geq m$ and $x_i \leq 1$. We model the break to occur at x_m , instead of between points, for simplicity.

Let the fitness of this three-parameter model be measured by:

$$\chi_3^2 = \sum_{i=1}^{N'} [y(x_i | \sigma_1, \sigma_2, m) - y_i]^2, \quad (\text{A2})$$

²⁶ It is interesting to note that this topic, although statistical in nature, originates in the field of astronomy. Two of these publications are in the early volumes of the Astronomical Journal, and

the third, Chauvenet's "A Manual of Spherical and Practical Astronomy", was a standard reference for decades.

²⁷ <https://skynet.unc.edu/rcr>

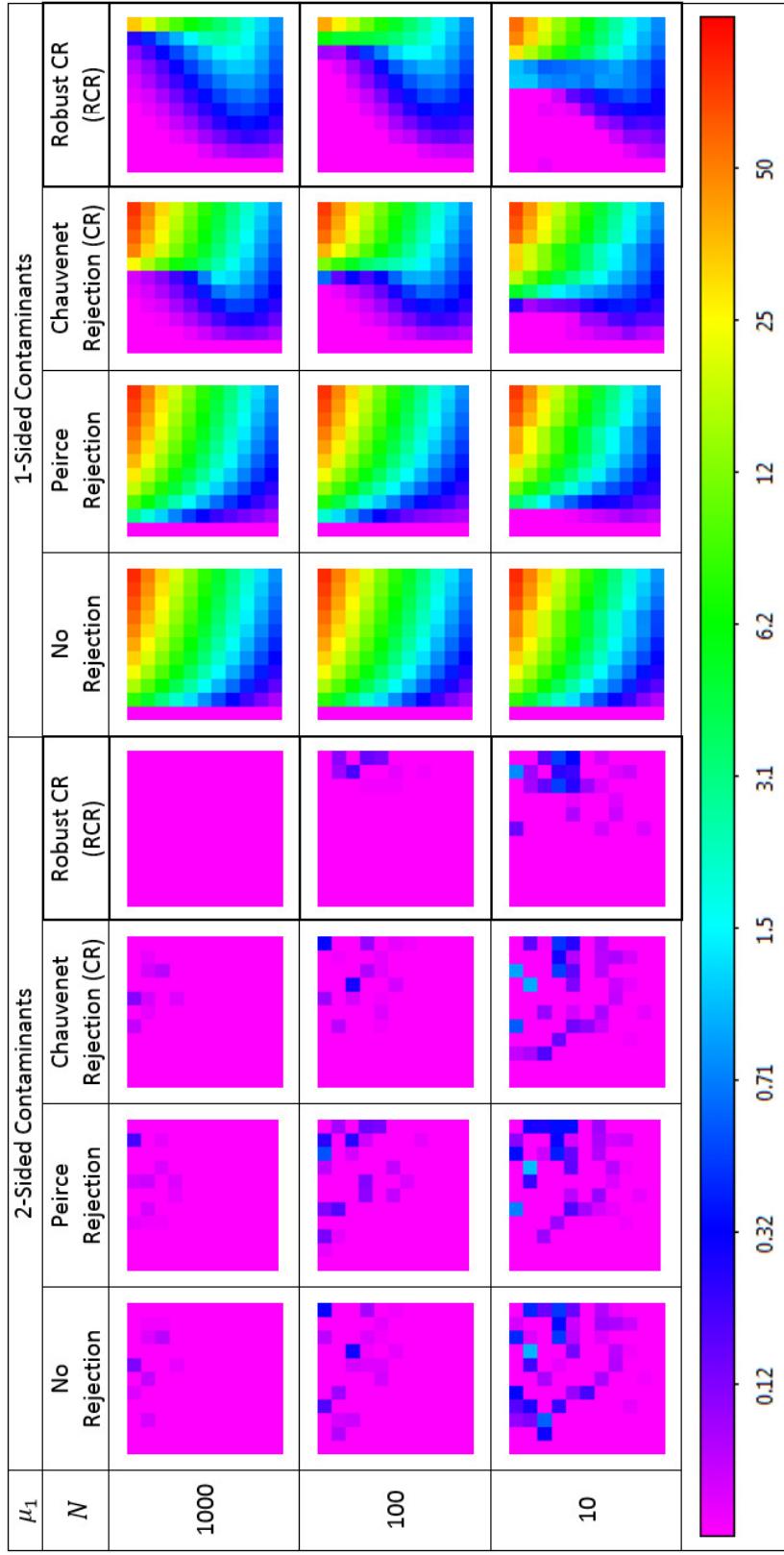


FIG. 43.— Average recovered μ_1 for (1) no rejection, (2) Peirce rejection, (3) traditional Chauvenet rejection, and (4) RCR, for two-sided contaminants (left) and one-sided contaminants (right). See Figure 5 for contaminant strength (σ_2) vs. fraction of sample (f_2) axis information. In the case of two-sided contaminants, all techniques recover $\mu_1 \approx 0$ (Figure 6). In the case of one-sided contaminants, traditional Chauvenet rejection, as implemented in this paper, is superior to Peirce rejection, and RCR (highlighted with a bold outline) is superior to traditional Chauvenet rejection. The colors are scaled logarithmically, between 0.02 and 100.

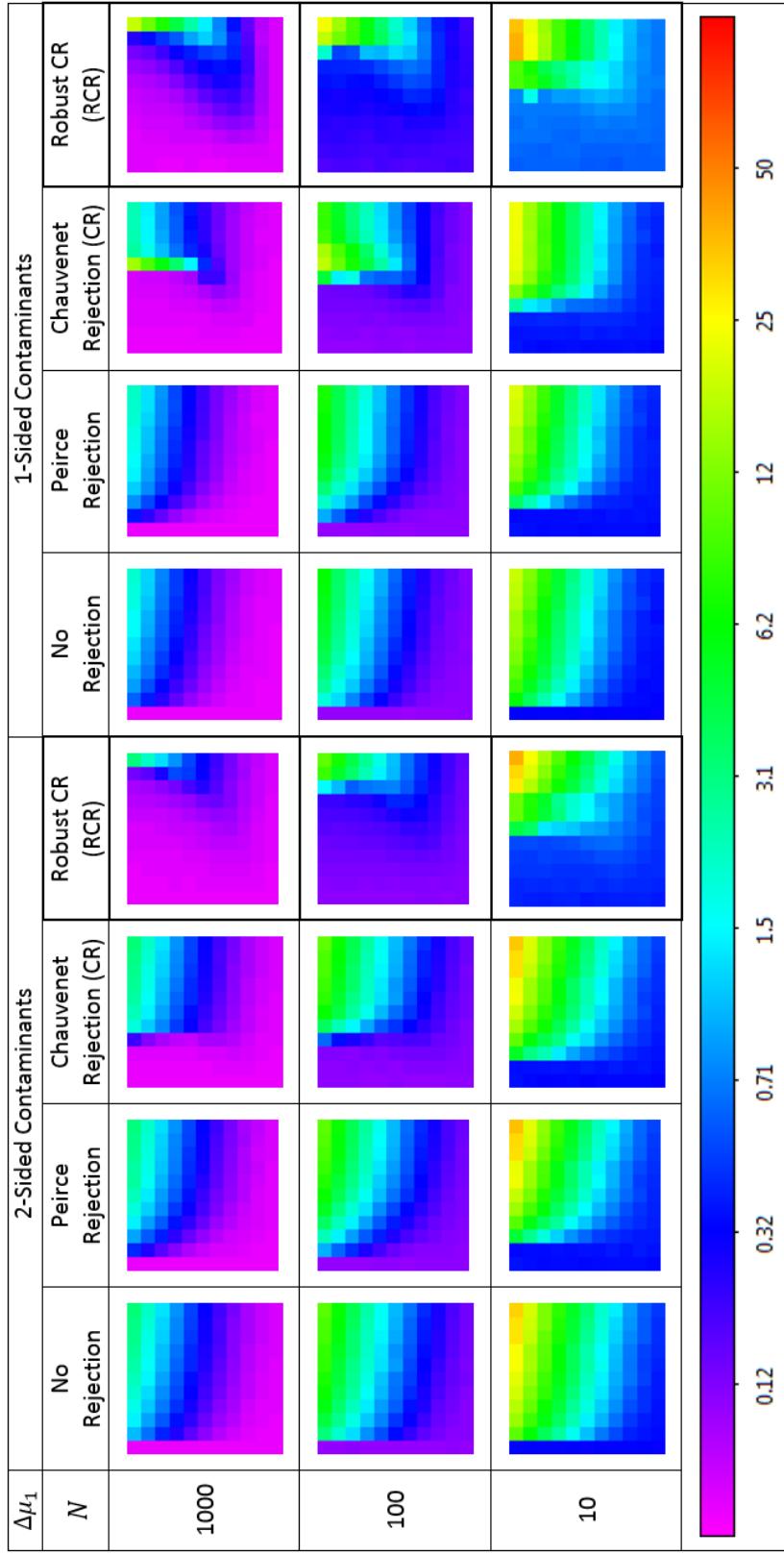


FIG. 44.— Average recovered $\Delta\mu_1$ for (1) no rejection, (2) Peirce rejection, (3) traditional Chauvenet rejection, and (4) RCR, for two-sided contaminants (left) and one-sided contaminants (right). See Figure 5 for contaminant strength (σ_2) vs. fraction of sample (f_2) axis information. In the case of two-sided contaminants, all techniques recover $\mu_1 \approx 0$ (Figure 6). In the case of one-sided contaminants, traditional Chauvenet rejection, as implemented in this paper, is superior to Peirce rejection, and RCR (highlighted with a bold outline) is superior to traditional Chauvenet rejection, albeit with marginally reduced precision at low N (§4). The colors are scaled logarithmically, between 0.02 and 100.

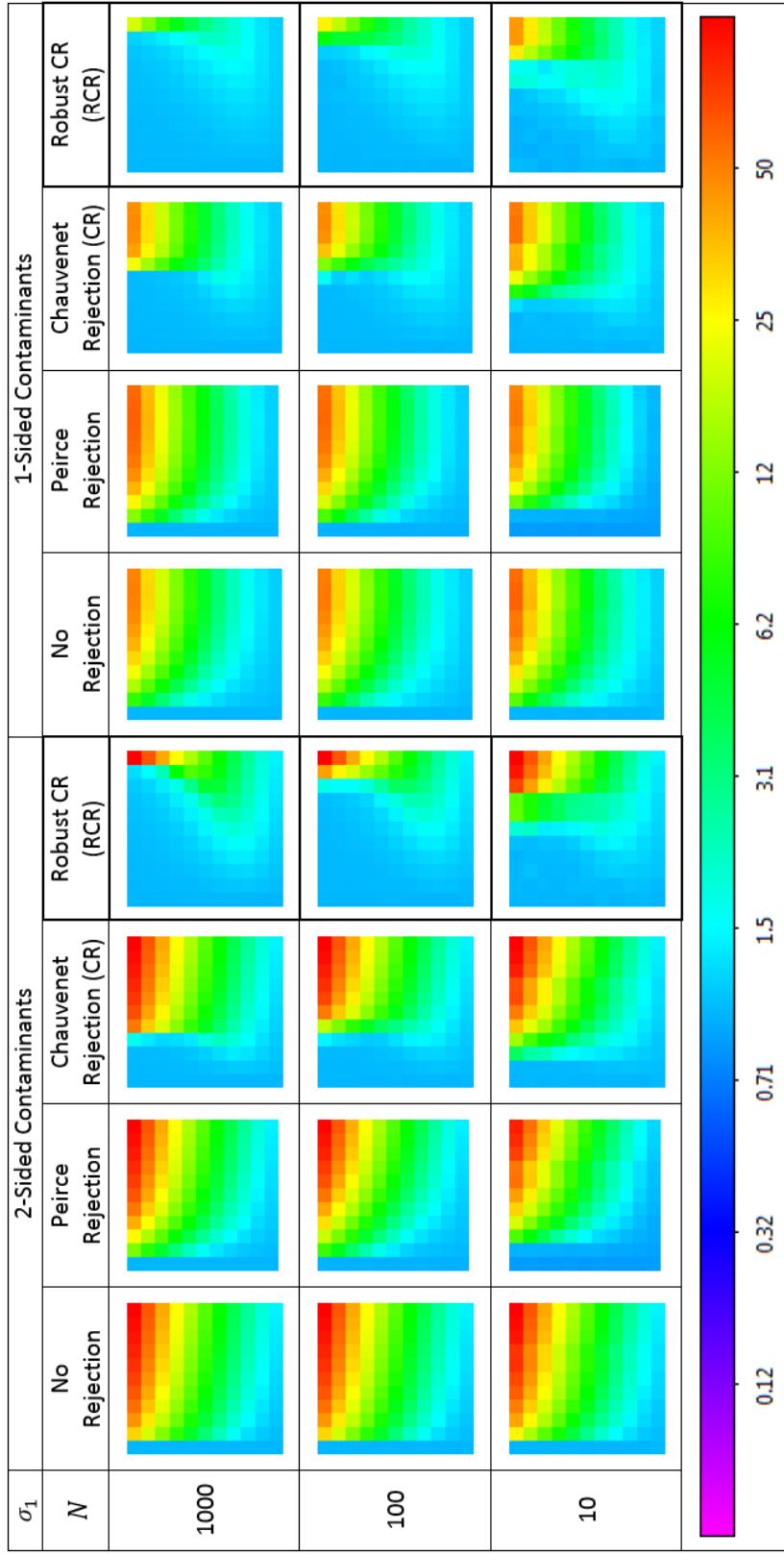


FIG. 45.— Average recovered σ_1 for (1) no rejection, (2) Peirce rejection, (3) traditional Chauvenet rejection, and (4) RCR, for two-sided contaminants (left) and one-sided contaminants (right). See Figure 5 for contaminant strength (σ_2) vs. fraction of sample (f_2) axis information. In the case of two-sided contaminants, all techniques recover $\mu_1 \approx 0$ (Figure 6). In the case of one-sided contaminants, traditional Chauvenet rejection, as implemented in this paper, is superior to Peirce rejection, and RCR (highlighted with a bold outline) is superior to traditional Chauvenet rejection. The colors are scaled logarithmically, between 0.02 and 100.

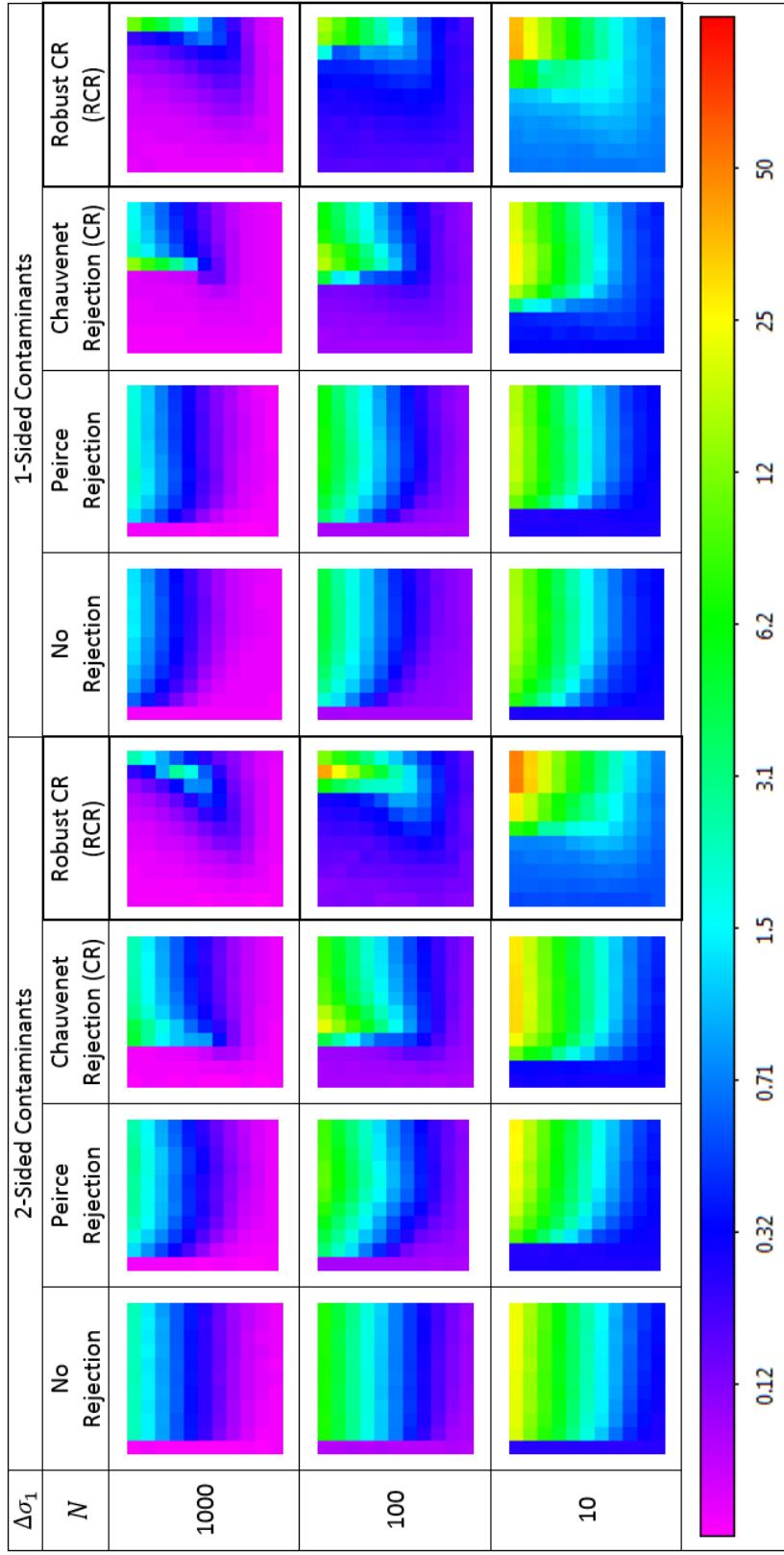


FIG. 46.— Average recovered $\Delta\sigma_1$ for (1) no rejection, (2) Peirce rejection, (3) traditional Chauvenet rejection, and (4) RCR, for two-sided contaminants (left) and one-sided contaminants (right). See Figure 5 for contaminant strength (σ_2) vs. fraction of sample (f_2) axis information. In the case of two-sided contaminants, all techniques recover $\mu_1 \approx 0$ (Figure 6). In the case of one-sided contaminants, traditional Chauvenet rejection, as implemented in this paper, is superior to Peirce rejection, and RCR (highlighted with a bold outline) is superior to traditional Chauvenet rejection, albeit with marginally reduced precision at low N (§4). The colors are scaled logarithmically, between 0.02 and 100.

where $N' = \text{floor}(0.683N + 0.317)$ is the number of points for which $x_i \leq 1$. Then for a given break point, m , the best fit is given by $d\chi_3^2/d\sigma_1 = d\chi_3^2/d\sigma_2 = 0$, yielding:

$$\begin{bmatrix} \sigma_1 \\ \sigma_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^m x_i^2 + x_m^2 \sum_{i=m+1}^{N'} 1 & x_m \sum_{i=m+1}^{N'} (x_i - x_m) \\ x_m \sum_{i=m+1}^{N'} (x_i - x_m) & \sum_{i=m+1}^{N'} (x_i - x_m)^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^m x_i y_i + x_m \sum_{i=m+1}^{N'} y_i \\ \sum_{i=m+1}^{N'} (x_i - x_m) y_i \end{bmatrix}. \quad (\text{A3})$$

We use a recursive partitioning algorithm to efficiently find the value of m for which χ_3^2 is minimized. We restrict $m > 1$ to avoid the following pathological case: If μ is measured by the median and N is odd, one of the measured values will always equal the median value, and consequently y_1 will always be zero; $m = 1$ would then imply $\sigma_1 = 0$, but without meaning.

Statistical equivalence to an unbroken-line fit through the origin (§2.2) can similarly result in spurious values of σ_1 : In this case, any fitted value of m is possible, and the lower it happens to be, the less well constrained σ_1 will be. Consequently, if statistically equivalent, we instead use the value of σ from the unbroken-line fit through the origin. We measure statistical equivalency by:

$$\frac{\chi_1^2 - \chi_3^2}{\chi_3^2} < f(N), \quad (\text{A4})$$

where χ_1^2 is the fitness of the one-parameter, unbroken-line fit, and $f(N)$ would be $\approx 2.3N'$ if the points to which we are fitting were statistically independent of each other, but because we are fitting to sorted data, this is not the case. Consequently, we determined $f(N)$ empirically: For each technique and sample size N , we drew 100,000 uncontaminated samples, measured $(\chi_1^2 - \chi_3^2)/\chi_3^2$ for each, sorted these values, and took the 68.3-percentile value (see Figure 47).

Otherwise, we only restrict σ_1 to be positive.

Finally, when fitting to weighted data (§6):

- x_i is instead given by $\sqrt{2}\text{erf}^{-1}(s_i / \sum_{i=1}^N w_i)$, where s_i is given by Equation 18, and w_i is the weight of the i th point;
- N' is instead given by the largest integer such that $s_{N'} \leq 0.683 \sum_{i=1}^N w_i$;
- Each term summed over i in χ_1^2 and χ_3^2 (Equation A2) and Equation A3 is multiplied by w_i ; and
- $f(N)$ instead depends on the weights of the data. To this end, for the three scenarios that we consider in §6 that make use of technique 3 (the broken-line fit), corresponding to all but the upper-right panel of Figure 31, we have computed $f(N)$ for the same five, representative weight distributions (see Figure 48, solid curves). From these, we have similarly produced empirical approximations, as functions of (1) N and (2) σ_w/μ_w of the $x_i = \sqrt{2}\text{erf}^{-1}(s_i / \sum_{i=1}^N w_i) < 1$ points, which can be used with any sample of similarly distributed weights (Figure 48, dashed curves; see Appendix B). We demonstrate these for the latter three weight distributions in columns 6 – 8, respectively, of Figures 20 – 27, and, desirably, they do not differ significantly from those of column 5, in which $\sigma_w/\mu_w = 0$.

B. EMPIRICAL APPROXIMATIONS FOR $F(N)$ AND CORRECTION FACTORS

For each scenario and technique presented in this paper, we calculated $f(N)$ (Appendix A) beyond $N = 1000$ and correction factors (Figures 4, 29, and 31) beyond $N = 100$, every 0.1 dex for an additional 1 – 2 dex, until it became computationally inefficient to continue.

For the cases involving data of equal weight, we fitted functions of N to these calculated values, yielding empirical approximations (Tables 2 – 5).

For the cases involving a distribution of weights, characterized by σ_w/μ_w of the $x_i = \sqrt{2}\text{erf}^{-1}(s_i / \sum_{i=1}^N w_i) < 1$ points (§6, Appendix A), we fitted functions of both (1) N and (2) σ_w/μ_w to the calculated values, yielding empirical approximations that can be used with any sample of similarly distributed weights.

For the “single σ ” scenario, appropriate for two-sided contaminants (§3.1), $f(N)$ is given by:

$$f(N) = f_1(N) 10^{10^{a_1(N)+b_1(N)\log_{10}(\frac{\sigma_w}{\mu_w})}}, \quad (\text{B1})$$

where $f_1(N)$ is the value of $f(N)$ for data of equal weight, corresponding to $\sigma_w/\mu_w = 0$, and $a_1(N \leq 7)$ and $b_1(N \leq 7)$ are listed in Table 6. For $7 < N \leq 1000$:

$$\begin{aligned} a_1(N) = & 0.2313(\log_{10} N)^6 - 3.02(\log_{10} N)^5 + 15.997(\log_{10} N)^4 - 43.713(\log_{10} N)^3 + 64.629(\log_{10} N)^2 \\ & - 49.976 \log_{10} N + 15.484 + 0.1513(-1)^N N^{-0.471}, \end{aligned} \quad (\text{B2})$$

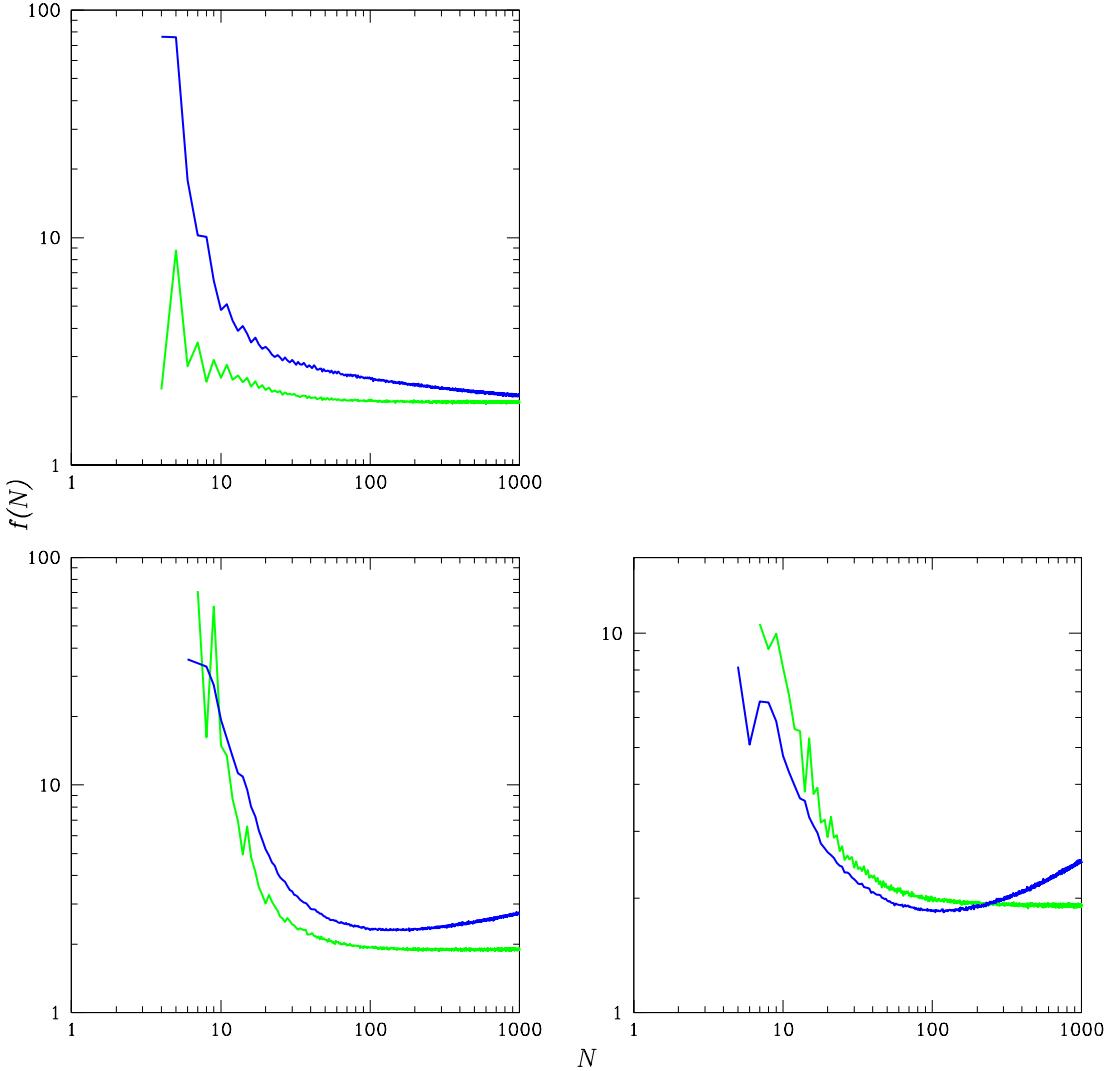


FIG. 47.— $f(N)$ vs. N , for sorted, but otherwise independent data, where deviations have been measured from (1) the median (green) and (2) the mode (blue). **Upper left:** For the simplest case of computing a single σ_1 , using the deviations both below and above μ (the median or the mode; §3.1). **Lower left:** For the case of computing separate σ_1 below and above μ and computing $(\chi_1^2 - \chi_3^2)/\chi_3^2$ for only the smaller of the two (§3.2). **Lower right:** For the same case, but computing $(\chi_1^2 - \chi_3^2)/\chi_3^2$ for either of the two, selected randomly (§3.3.1). Oscillations are not noise, but odd-even effects (e.g., with equally weighted data, when N is odd, use of the median always results in at least one zero deviation, resulting in a larger value of $(\chi_1^2 - \chi_3^2)/\chi_3^2$). We use look-up tables for $N \leq 1000$ and empirical approximations for $f(N > 1000)$ (see Appendix B).

and

$$b_1(N) = -0.3556(\log_{10} N)^6 + 3.7036(\log_{10} N)^5 - 14.932(\log_{10} N)^4 + 29.176(\log_{10} N)^3 - 28.81(\log_{10} N)^2 + 14.397 \log_{10} N - 2.64511. \quad (\text{B3})$$

For $N > 1000$, $f(N) = f_1(N)$.

For this scenario and the “bulk rejection (median) + RCR (median-T3) + RCR (median-T1) + CR” technique, the correction factor is given by:

$$CF(N) = \begin{cases} CF_1(N)10^{10^{a(N)+b(N)\log_{10}\left(\frac{\sigma_w}{\mu_w}\right)}} & N = 2, N = 4, N > 5 \\ CF_1(N)10^{-10^{a(N)+b(N)\log_{10}\left(\frac{\sigma_w}{\mu_w}\right)}} & N = 3, N = 5 \end{cases}, \quad (\text{B4})$$

where $CF_1(N)$ is the value of $CF(N)$ for data of equal weight, corresponding to $\sigma_w/\mu_w = 0$, and $a(N \leq 5)$ and

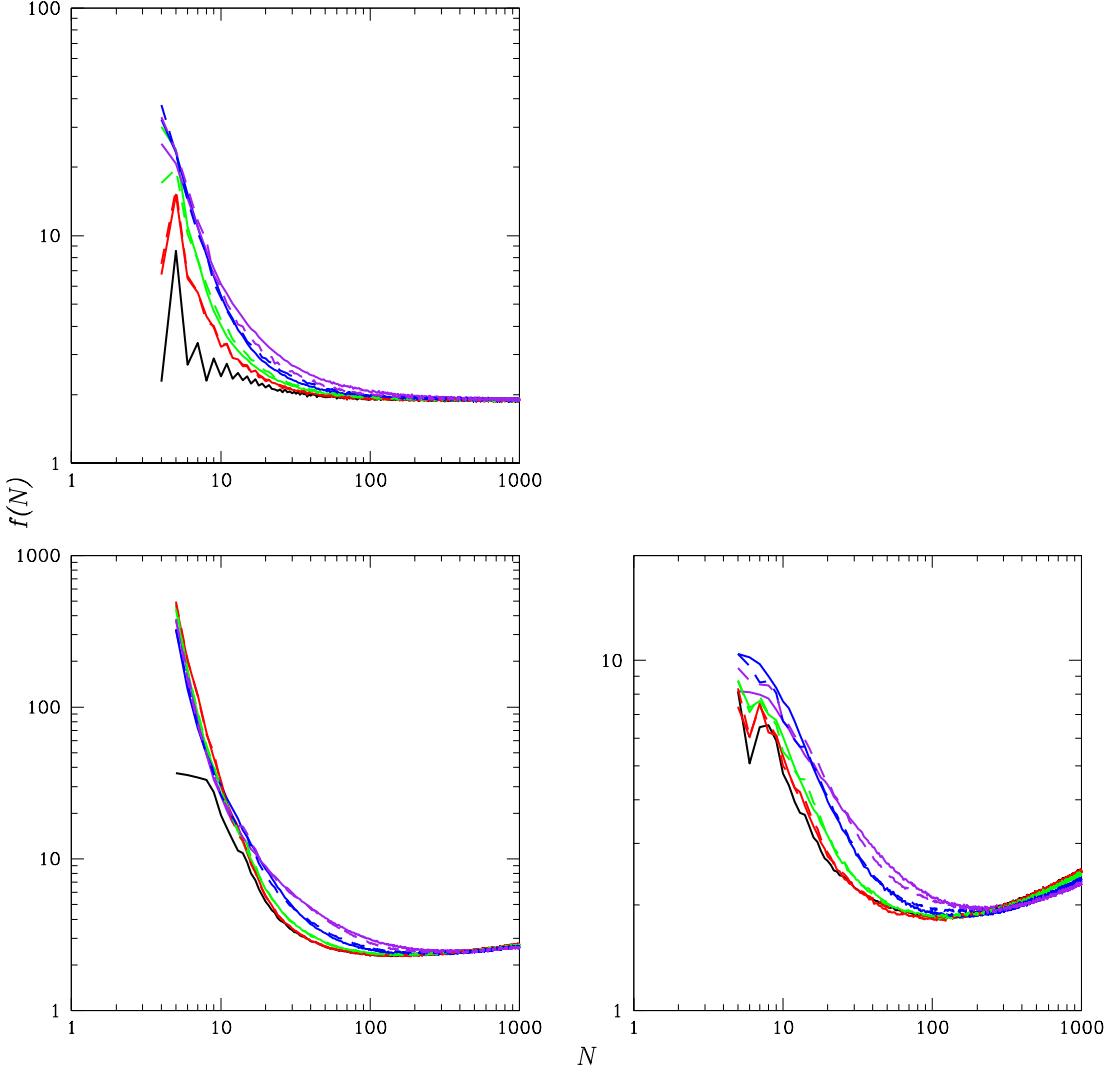


FIG. 48.— Same as the green, blue, and blue curves from the upper-left, lower-left, and lower-right panels of Figure 47, respectively, corresponding to what is needed for our best-option robust techniques (§4), but for five representative weight distributions: (1) all weights equal (solid black curves – same as the designated curves from Figure 47); (2) weights distributed normally with standard deviation as a fraction of the mean $\sigma_w/\mu_w = 0.1$ (solid red curves); (3) weights distributed normally with $\sigma_w/\mu_w = 0.3$ (solid green curves); (4) weights distributed uniformly from zero (i.e., low-weight points as common as high-weight points; solid blue curves), corresponding to $\sigma_w/\mu_w \approx 0.58$; and (5) weights distributed inversely over one dex (i.e., low-weight points more common than high-weight points, with the sum of the weights of the low-weight points as impactful as the sum of the weights of the high-weight points; solid purple curves), corresponding to $\sigma_w/\mu_w \approx 0.73$. From these, we have produced empirical approximations, as functions of (1) N and (2) σ_w/μ_w of the $x_i = \sqrt{2}\text{erf}^{-1}(s_i/\sum_{i=1}^N w_i) < 1$ points, which can be used with any sample of similarly distributed weights (dashed curves; see Appendix B).

$b(N \leq 5)$ are listed in Table 7. For $N > 5$:

$$a(N) = -0.7914 \log_{10} N + 0.0243 \quad (\text{B5})$$

and

$$b(N) = 0.1196 \log_{10} N + 4.5073. \quad (\text{B6})$$

For the “smaller of σ_- and σ_+ ” scenario, appropriate for one-sided contaminants and in-between cases (§3.2), $f(N \leq 264)$ is given by Equation B1, $f(264 < N \leq 567)$ is given by:

$$f(N) = \begin{cases} f_1(N) 10^{10^{a_1(N)+b_1(N)\log_{10}\left(\frac{\sigma_w}{\mu_w}\right)}} & a_1(N) + b_1(N) \log_{10}\left(\frac{\sigma_w}{\mu_w}\right) > a_2(N) + b_2(N) \log_{10}\left(\frac{\sigma_w}{\mu_w}\right) \\ f_1(N) & a_1(N) + b_1(N) \log_{10}\left(\frac{\sigma_w}{\mu_w}\right) = a_2(N) + b_2(N) \log_{10}\left(\frac{\sigma_w}{\mu_w}\right) \\ f_1(N) 10^{-10^{a_2(N)+b_2(N)\log_{10}\left(\frac{\sigma_w}{\mu_w}\right)}} & a_1(N) + b_1(N) \log_{10}\left(\frac{\sigma_w}{\mu_w}\right) < a_2(N) + b_2(N) \log_{10}\left(\frac{\sigma_w}{\mu_w}\right) \end{cases}, \quad (\text{B7})$$

TABLE 2
EMPIRICAL APPROXIMATIONS FOR $f(N > 1000)^a$

Scenario	Median	Mode
Single σ	1.90	$39.2519N^{-0.7969} + 1.8688$
Smaller of σ_- and σ_+	1.90	$1.3399N^{0.1765}$
Random of σ_- and σ_+	1.90	$1.2591N^{0.2052}$

^a For data of equal weight (Appendix A). The latter two approximations should be used with caution beyond $N \sim 10^6$.

TABLE 3
CORRECTION FACTORS FOR THE “SINGLE σ ” SCENARIO^a

Technique	Correction Factor
Mean-Standard Deviation (No Rejection)	$(1 - 0.2897N^{-1.033})^{-1}$
CR (Mean-Standard Deviation)	$(1 - 0.7240N^{-0.773})^{-1}$
RCR (Median-Technique 1)	$(1 - 1.7198N^{-1.022})^{-1}$
RCR (Median-Technique 2)	$(1 - 2.9442N^{-1.073})^{-1}$
RCR (Median-Technique 3)	$(1 - 4.2145N^{-1.153})^{-1}$
RCR (Mode-Technique 1)	$1 - \frac{0.1052}{(N/41.99)^{-0.5328} + (N/41.99)^{0.4130}}$
RCR (Mode-Technique 2)	$1 - \frac{0.05104}{(N/104.9)^{-3.2545} + (N/104.9)^{0.3444}}$
RCR (Mode-Technique 3)	$(1 - 2.1893N^{-0.803})^{-1}$
RCR (Median-T3) + CR	$(1 - 4.2134N^{-0.971})^{-1}$
RCR (Median-T3) + RCR (Median-T1) + CR	$(1 - 4.3185N^{-0.975})^{-1}$
Bulk Rejection (Median) + RCR (Median-T3) + RCR (Median-T1) + CR	$(1 - 3.5780N^{-0.942})^{-1}$

^a For data of equal weight (§2.3, §3.1, §4, §5). Appropriate for two-sided contaminants. Empirical approximations are for $N > 100$.

and $f(N > 567)$ is given by:

$$f(N) = f_1(N)10^{-10^{a_2(N)+b_2(N)\log_{10}\left(\frac{\sigma_w}{\mu_w}\right)}}, \quad (\text{B8})$$

where $a_1(N \leq 8)$ and $b_1(N \leq 8)$ are listed in Table 6.²⁸ For $8 < N \leq 1000$:

$$a_1(N) = -0.541(\log_{10} N)^5 + 4.6943(\log_{10} N)^4 - 15.407(\log_{10} N)^3 + 21.875(\log_{10} N)^2 - 11.211\log_{10} N - 0.3798, \quad (\text{B9})$$

$$b_1(N) = 0.1462(\log_{10} N)^3 - 4.2139(\log_{10} N)^2 + 14.366\log_{10} N - 10.658, \quad (\text{B10})$$

$$a_2(N) = 18.149(\log_{10} N)^3 - 149.27(\log_{10} N)^2 + 410.15\log_{10} N - 378.47, \quad (\text{B11})$$

and

$$b_2(N) = 26.945(\log_{10} N)^3 - 221.42(\log_{10} N)^2 + 606.91\log_{10} N - 553.89. \quad (\text{B12})$$

For $N > 1000$:

$$a_2(N) = 0.3861\log_{10} N - 2.5852 \quad (\text{B13})$$

and

$$b_2(N) = 0.0424\log_{10} N + 1.4479. \quad (\text{B14})$$

This approximation should be used with caution beyond $N \sim 10^6$.

For this scenario and the “bulk rejection (mode) + RCR (mode-T1) + RCR (median-T1) + CR” technique, the correction factor is given by:

$$CF(N) = \begin{cases} CF_1(N)10^{10^{a(N)+b(N)\log_{10}\left(\frac{\sigma_w}{\mu_w}\right)}} & N = 2, N = 3, N > 4, \\ CF_1(N) & N = 4 \end{cases} \quad (\text{B15})$$

where $a(N \leq 5)$ and $b(N \leq 5)$ are listed in Table 7. For $5 < N \leq 100$:

$$a(N) = -1.1937(\log_{10} N)^4 + 6.5268(\log_{10} N)^3 - 13.308(\log_{10} N)^2 + 11.432\log_{10} N - 4.4769 \quad (\text{B16})$$

and

$$b(N) = -1.4528(\log_{10} N)^3 + 5.3519(\log_{10} N)^2 - 5.33\log_{10} N + 2.2902 + 0.1879(-1)^N(\log_{10} N)^{0.9521}. \quad (\text{B17})$$

²⁸ When $N = 5$, technique 3 (the broken-line fit) always defaults to technique 2 (the linear fit) if the data are weighted equally (Figure 4), leaving $f_1(5)$ undefined. However, this is not always the

case if the data are not weighted equally. Consequently, we must define $f_1(5)$ before determining $a_1(5)$ and $b_1(5)$. To this end, we adopt $f_1(5) = 36.8534$, by extrapolation.

TABLE 4
CORRECTION FACTORS FOR THE “SMALLER OF σ_- AND σ_+ ” SCENARIO^a

Technique	Correction Factor
Mean-Standard Deviation (No Rejection)	$(1 - 0.5092N^{-0.514})^{-1}$
CR (Mean-Standard Deviation)	$(1 - 0.6939N^{-0.522})^{-1}$
RCR (Median-Technique 1)	$(1 - 1.3320N^{-0.549})^{-1}$
RCR (Median-Technique 2)	$(1 - 1.5058N^{-0.559})^{-1}$
RCR (Median-Technique 3)	$(1 - 1.0426N^{-0.443})^{-1}$
RCR (Mode-Technique 1)	$(1 - 0.5736N^{-0.265})^{-1}$
RCR (Mode-Technique 2)	$(1 - 0.7285N^{-0.279})^{-1}$
RCR (Mode-Technique 3)	$(1 - 0.8790N^{-0.264})^{-1}$
RCR (Mode-T1) + CR	$(1 - 1.7079N^{-0.602})^{-1}$
RCR (Mode-T3) + CR	$(1 - 2.8415N^{-0.630})^{-1}$
RCR (Mode-T1) + RCR (Median-T1) + CR	$(1 - 1.7453N^{-0.605})^{-1}$
RCR (Mode-T3) + RCR (Median-T1) + CR	$(1 - 2.9047N^{-0.633})^{-1}$
Bulk Rejection (Mode) + RCR (Mode-T1) + RCR (Median-T1) + CR	$(1 - 2.3525N^{-0.627})^{-1}$
Bulk Rejection (Mode) + RCR (Mode-T3) + RCR (Median-T1) + CR	$(1 - 3.3245N^{-0.650})^{-1}$

^a For data of equal weight (§2.3, §3.2, §4, §5). Appropriate for one-sided contaminants and in-between cases. Empirical approximations are for $N > 100$.

TABLE 5
CORRECTION FACTORS FOR THE “RANDOM OF σ_- AND σ_+ ” SCENARIO^a

Technique	Correction Factor
Mean-Standard Deviation (No Rejection)	$(1 - 0.4176N^{-1.293})^{-1}$
CR (Mean-Standard Deviation)	$(1 - 0.4482N^{-0.717})^{-1}$
RCR (Median-Technique 1)	$(1 - 2.0285N^{-1.021})^{-1}$
RCR (Median-Technique 2)	$(1 - 2.5569N^{-1.050})^{-1}$
RCR (Median-Technique 3)	$1 - \frac{0.06629}{(N/41.83)^{-2.8626} + (N/41.83)^{0.9580}}$
RCR (Mode-Technique 1)	$\begin{cases} 1.02187 - 0.00907 \log_{10} N & \text{if } 100 < N \leq 1000 \\ 1 - 0.03946N^{-0.2895} & \text{if } N > 1000 \end{cases}$
RCR (Mode-Technique 2)	$\begin{cases} 1.07422 - 0.02651 \log_{10} N & \text{if } 100 < N \leq 1000 \\ 1 - 0.01616N^{-0.2895} & \text{if } N > 1000 \end{cases}$
RCR (Mode-Technique 3)	$(1 - 3.4414N^{-0.849})^{-1}$
RCR (Mode-T3) + CR	$(1 - 3.2546N^{-0.840})^{-1}$
RCR (Mode-T1) + RCR (Median-T1) + CR	$(1 - 2.8989N^{-0.824})^{-1}$
Bulk Rejection (Mode) + RCR (Mode-T3) + RCR (Median-T1) + CR	$(1 - 3.1666N^{-0.833})^{-1}$

^a For data of equal weight (§2.3, §3.3.1, §4, §5). Appropriate for (mildly) asymmetric uncontaminated distributions. Empirical approximations are for $N > 100$.

For $N > 100$:

$$a(N) = -0.5408 \log_{10} N - 0.6482 \quad (\text{B18})$$

and

$$b(N) = 1.4154 + 0.3635(-1)^N. \quad (\text{B19})$$

For this scenario and the “bulk rejection (mode) + RCR (mode-T3) + RCR (median-T1) + CR” technique, the correction factor is given by Equation B15, where $a(N \leq 5)$ and $b(N \leq 5)$ are listed in Table 7. For $5 < N \leq 20$:

$$a(N) = -0.2683(\log_{10} N)^4 + 1.9174(\log_{10} N)^3 - 5.062(\log_{10} N)^2 + 5.452 \log_{10} N - 2.9999 \quad (\text{B20})$$

and

$$b(N) = 43.179(\log_{10} N)^6 - 331.85(\log_{10} N)^5 + 968.25(\log_{10} N)^4 - 1399.1(\log_{10} N)^3 + 1070.7(\log_{10} N)^2 - 415.81 \log_{10} N + 65.002 + 0.1365(-1)^N (\log_{10} N)^{2.4716}. \quad (\text{B21})$$

For $20 < N \leq 100$, $a(N)$ is given by Equation B20, and:

$$b(N) = 1.5144 \log_{10} N - 0.0448 + 0.1365(-1)^N (\log_{10} N)^{2.4716}. \quad (\text{B22})$$

For $N > 100$:

$$a(N) = -0.4282 \log_{10} N - 0.4412 \quad (\text{B23})$$

and

$$b(N) = 2.9881 + 0.7530(-1)^N. \quad (\text{B24})$$

TABLE 6
EMPIRICAL APPROXIMATION PARAMETER VALUES FOR $f(N \leq 8)$

Scenario: N	Single σ $a_1(N)$	$b_1(N)$	Smaller of σ_- and σ_+ $a_1(N)$	$b_1(N)$	Random of σ_- and σ_+ $a_1(N)$	$b_1(N)$
4	0.2024	0.4642
5	-0.2916	0.2603	-0.0828	-0.3003	-0.4664	2.1342
6	-0.0332	0.3638	-0.2675	-0.2443	-0.3124	1.0197
7	-0.1818	0.4547	-0.5588	-0.4097	-0.7502	0.5797
8	Equation B2	Equation B3	-0.8893	-0.4884	Equation B25	Equation B26

TABLE 7
EMPIRICAL APPROXIMATION PARAMETER VALUES FOR $CF(N \leq 5)$

Scenario: Technique:	Single σ Bulk Rejection (Median) + RCR (Median-T3) + RCR (Median-T1) + CR	Bulk Rejection (Mode) + RCR (Mode-T1) + RCR (Median-T1) + CR	Smaller of σ_- and σ_+ a(N) b(N)	Bulk Rejection (Mode) + RCR (Mode-T3) + RCR (Median-T1) + CR	Random of σ_- and σ_+ a(N) b(N)	
N	$a(N)$	$b(N)$	$a(N)$	$b(N)$	$a(N)$	$b(N)$
2	-3.1528	0.2739	-0.3984	1.0815	-0.3984	1.0815
3	-1.1913	0.4487	-1.1462	1.0699	-1.1094	1.5143
4	-1.0509	3.3825	-0.9543
5	-1.4145	0.1185	-1.1196	0.4597	-1.1446	0.3394

For the “random of σ_- and σ_+ ” scenario, appropriate for (mildly) asymmetric uncontaminated distributions (§3.3.1), $f(N \leq 190)$ is given by Equation B1, $f(190 < N \leq 305)$ is given by Equation B7, and $f(N > 305)$ is given by Equation B8, where $a_1(N \leq 7)$ and $b_1(N \leq 7)$ are listed in Table 6. For $7 < N \leq 1000$:

$$a_1(N) = 3.1767(\log_{10} N)^6 - 34.561(\log_{10} N)^5 + 152.16(\log_{10} N)^4 - 347.96(\log_{10} N)^3 + 435.59(\log_{10} N)^2 - 282.57 \log_{10} N + 73.696, \quad (B25)$$

$$b_1(N) = 5.8718(\log_{10} N)^4 - 47.049(\log_{10} N)^3 + 131.12(\log_{10} N)^2 - 150.24 \log_{10} N + 61.727, \quad (B26)$$

$$a_2(N) = -1.8953(\log_{10} N)^2 + 11.745 \log_{10} N - 19.36, \quad (B27)$$

and

$$b_2(N) = -2.7584(\log_{10} N)^2 + 17.078 \log_{10} N - 24.602. \quad (B28)$$

For $N > 1000$:

$$a_2(N) = -1.1827 \quad (B29)$$

and

$$b_2(N) = 1.8064. \quad (B30)$$

This approximation should also be used with caution beyond $N \sim 10^6$.

For this scenario and the “bulk rejection (mode) + RCR (mode-T3) + RCR (median-T1) + CR” technique, the correction factor is given by:

$$CF(N) = CF_1(N)10^{10^{a(N)+b(N)\log_{10}\left(\frac{\sigma_w}{\mu_w}\right)}}, \quad (B31)$$

where $a(N \leq 4)$ and $b(N \leq 4)$ are listed in Table 7. For $4 < N \leq 19$:

$$a(N) = -1.3993(\log_{10} N)^3 + 6.5746(\log_{10} N)^2 - 9.8844 \log_{10} N + 2.8572 \quad (B32)$$

and

$$b(N) = 4.0458(\log_{10} N)^2 - 6.4354 \log_{10} N + 2.7667. \quad (B33)$$

For $19 < N \leq 100$, $a(N)$ is given by Equation B32, and:

$$b(N) = 1.7394 \log_{10} N - 1.0435. \quad (B34)$$

For $N > 100$:

$$a(N) = -0.5989 \log_{10} N - 0.6097 \quad (B35)$$

and

$$b(N) = 1.4123 \log_{10} N - 0.3893. \quad (B36)$$

REFERENCES

- Bickel, D. R., and Frühwirth, R. 2006, Computational Statistics and Data Analysis, 50, 3500
- Chauvenet, W. 1863, A Manual of Spherical and Practical Astronomy, Vol. 2 (Philadelphia, PA: J. B. Lippincott Company)
- Dardis, C. 2012, Peirce, R-Forge (<https://r-forge.r-project.org/>)
- Gould, B. A. 1855, AJ, 83, 81
- Martin, J. R., Reichart, D. E., Dutton, D. A., Maples, M. P., Berger, T. A., Ghigo, F. D., Haislip, J. B., Trotter, A. S., Shaban, O. H., Barnes, L. M., Paggan, M. L., Gao, R. L., Salemi, C. P., Langston, G. I., Bussa, S., Duncan, J. A., White, S., Heatherly, S. A., Karlik, J. B., Johnson, E. M., Reichart, J. E., Foster, A. C., Kouprianov, V. V., Mazlin, S., and Harvey, J. 2018, ApJS, submitted
- Peirce, B. 1852, AJ, 45, 161
- Reichart, D. E. 2001, ApJ, 553, 235
- Ross, S. M. 2003, Journal of Engineering Technology, 2, 1
- Trotter, A. S. 2011, The Gamma-Ray Burst Afterglow Modeling Project: Foundational Statistics and Absorption & Extinction Models, PhD dissertation, University of North Carolina at Chapel Hill (Chapel Hill, NC: ProQuest/UMI)
- Trotter, A. S., Reichart, D. E., Egger, R. E., Styblova, J., Paggan, M. L., Martin, J. R., Dutton, D. A., Reichart, J. E., Kumar, N. D., Maples, M. P., Barlow, B. N., Berger, T. A., Foster, A. C., Frank N. R., Ghigo, F. D., Haislip, J. B., Heatherly, S. A., Kouprianov, V. V., LaCluyze, A. P., Moffett, D. A., Moore, J. P., Stanley, J. L., and White, S. 2017, MNRAS, 469, 1299