

¹ **Robust Bivariate Error Detection in Skewed Data with**
² **Application to Historical Radiosonde Winds**

³ Ying Sun,¹ Amanda S. Hering,² and Joshua M. Browning²

January 6, 2016

⁴ **Abstract**

⁵ Quality control methods for multivariate observations are generally based on using robust
⁶ estimates of parameters for a particular distribution, and that particular distribution is usually
⁷ the multivariate normal (MVN). However, many multivariate data generating processes do not
⁸ produce elliptical contours, and in such cases, error detection using the MVN distribution can lead
⁹ to legitimate observations being erroneously flagged. In this work, we propose non-parametric
¹⁰ and parametric methods for identifying errors in skewed bivariate data. In the first method, we
¹¹ remove potential outliers by assigning each bivariate observation a depth score and remove those
¹² observations that fall beyond a given threshold. In the second method, we first develop robust
¹³ estimators for the parameters in a bivariate skew- t (BST) distribution, and these parameters
¹⁴ are used in either distance-based or contour-based approaches to flag observations as potential
¹⁵ outliers. We test the performance of these methods in simulation against a more common MVN
¹⁶ outlier detection method. Finally, we show how our methods can be used in practice with
¹⁷ radiosonde launches of horizontal and vertical wind components measured at 8 vertical pressure
¹⁸ levels in which we demonstrate differences in the number of flagged outliers across pressure levels
¹⁹ and through time.

²⁰ **Some keywords:** Outliers, Radiosonde Winds, Skewed Distributions

²¹ **Short title:** Robust Bivariate Error Detection in Skewed Data

¹Division of Computer, Electrical and Mathematical Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia,
E-mail: ying.sun@kaust.edu.sa

²Applied Mathematics and Statistics, Colorado School of Mines, Golden, CO 80401, USA.
E-mail: {ahering, jbrownin}@mines.edu

22 **1 Introduction**

23 Detecting multivariate outliers is an inherently difficult problem since an observation may not
24 be considered an outlier in any one given dimension, but it could be unusual when considered
25 jointly across all of its dimensions. One common approach to detecting multivariate outliers when
26 the number of variables is small is to use robust estimation of the parameters of a multivariate
27 normal (MVN) distribution, such as Rousseeuw and Van Driessen (1999) or Peña and Prieto
28 (2001), since the outliers themselves can influence the parameter estimates. Then, a MVN
29 based algorithm to detect outliers is applied using the robust estimates of location and scale in
30 Mahalanobis distance (Filzmoser et al. 2005; Filzmoser et al. 2008). Good overviews are given
31 in Rousseeuw and Leroy (2003) and Maronna et al. (2006). However, many processes do not
32 fit the MVN distribution profile and may have heavy tailed and/or skewed distributions. An
33 alternative multivariate distribution to the MVN is the multivariate skew- t (MST) distribution.
34 It is flexible enough to fit variations in the third and fourth moments of a distribution and has
35 the MVN distribution as a special, central case (Azzalini and Capitanio 2014).

36 Another approach to outlier detection in multivariate data is to project the observations into
37 a lower-dimensional subspace with a method such as principal component analysis and then
38 summarize the observations in terms of a single monitoring statistic like Hotelling's T^2 or the
39 squared prediction error (SPE) (e.g., Wold et al. 1987; Schölkopf et al. 1998). The distributions
40 of the monitoring statistics depend on the assumption that the data-generating process is either
41 multivariate normal or locally linear, but these assumptions do not always hold, making it difficult
42 to determine the threshold at which to declare that an observation is an outlier (see Kazor et
43 al. 2016). The dimension reduction approach is particularly popular in process monitoring where
44 the number of variables being monitored is large. However, when a process of interest has only
45 a few variables of interest, flagging outliers in the original space of observations is preferred
46 since this allows researchers to quickly identify possible causes of the error. For this reason,
47 transforming the observations with a multivariate copula is also not ideal (Joe 2015).

48 There is an important difference between extremes of a distribution and outliers, as described
49 by Reimann et al. (2005) for univariate data. Extremes of a distribution are simply observations

50 that are consistent with the distribution but will appear if a large enough sample is taken.
51 Outliers, or errors, are observations that are inconsistent with the distribution and were generated
52 from possibly multiple different processes. Our goal is not to remove extremes but to identify
53 likely errors, specifically with application to winds that are observed in two-dimensional space
54 with the horizontal u -component and vertical v -component. Indeed, as Gupta et al. (2014) point
55 out in their overview of outlier detection methods in temporal data, outlier detection methods
56 for different data types are not trivial to generalize, and while the methods described herein
57 can be applied quite generally to any set of bivariate observations, we evaluate and assess the
58 methods with particular interest in applying them to radiosonde launches.

59 Radiosondes are instruments that are attached to weather balloons and are released twice
60 daily at over 700 stations around the globe. They are the only direct measurements of the upper
61 atmosphere and include pressure, temperature, dewpoint, and winds as the balloon rises through
62 the atmosphere at a standard set of pressure levels. An additional 1,300 historical stations
63 exist that are no longer in use, and these records date back to the 1920's and contain many
64 millions of observations. As with any such large and complex archive, many errors are present.
65 Both systematic errors, such as differences in units of measure, changing instrumentation, and
66 discrepancies in station locations, and random errors, such as data entry and transmission errors,
67 can occur. In particular for wind, random errors can occur due to imprecise tracking of the
68 ascending balloon or by the motion of the balloon relative to the atmosphere (WMO 2008). We
69 focus this work on identifying random errors in the historical archive (i.e., data collected prior to
70 1980) since older records cannot be externally verified with numerical weather prediction models.

71 Most quality assurance systems for finding random errors in the historical radiosonde archive
72 have focused on temperature (Durre et al. 2008; Anderson et al. 2016; and Browning and Her-
73 ing 2016) as temperature is an important variable in studying climate change. However, the
74 radiosonde winds are also important as they are used in data assimilation products, such as Na-
75 tional Centers for Environmental Prediction (NCEP) reanalyses (Kalnay et al. 1996; Kanamitsu
76 et al. 2002), and these are used as boundary conditions in global and regional climate models.
77 The radiosonde winds can also be used in studies of wind climate (Jury and Pathack 1991; Frank

78 and Landberg 1997; Brönnimann and Luterbacher 2004); events such as severe windstorms and
79 the Dust Bowl (Klimowksi et al. 2003; Brönnimann et al. 2009); and low-level jets (Walters
80 and Winkler 2001). Quality control of winds in current launches are typically basic and include
81 checking that observations are within reasonable bounds; removing observations if only one of
82 the pair is recorded; and checking for long strings of repeated values (Durre et al. 2004). One
83 exception is Wartenburger et al. (2013) who use neighboring stations to identify random errors in
84 temperature, wind speed, wind direction, and geopotential height in records from 1923 to 1966;
85 however, each variable is assessed individually, not jointly. Ultimately, they note that their wind
86 errors should be interpreted cautiously, and they still inspect the wind observations visually.

87 In addition to the bivariate nature of wind, each launch of a radiosonde produces a function
88 of wind over pressure level. Figure 1 shows the u and v components of wind plotted as a function
89 of pressure for yearly averaged launches over 50 years from 1962 to 2011 (left) and over 200
90 individual launches in 1962 (right) at the Denver, Colorado station. In the yearly averages, there
91 is clearly a shift in observed winds, which may indicate a change in instrumentation or slight
92 shift in station location, representing a systematic change. In the individual launches, there is a
93 substantial amount of variability with an increase in spread in the mid-pressure levels, which is
94 consistent with the height of the jet stream. The nonparametric bivariate wind densities at each
95 pressure level for over 35,000 launches are shown in Figure 2. The longer tails on the right-hand
96 side of the distributions are especially evident at the middle pressure levels, which is consistent
97 with Figure 1 (right). Our goal here is to focus on identifying random errors in the record, not
98 systematic shifts (see wind homogenization work by Gruber and Haimberger (2008) and Ramella
99 Pralungo and Haimberger (2014)).

100 In this paper, we compare two established methods and propose two new methods for handling
101 both the bivariate and skewed nature of this data. The aforementioned parametric method based
102 on the BVN is applied (Filzmoser et al. 2005); a nonparametric method based on the bagplot
103 in Rousseeuw et al. (1999) that preserves asymmetries in the bivariate density is tested; and
104 finally, two methods based on the parametric form of the BST distribution are developed. The
105 bagplot is constructed by computing Tukey's half space depth for each observation (Tukey 1975),

and errors are flagged and removed by identifying those observations that are greater than the distance from the point with the greatest depth to the edge of the region containing the largest 50% of depth values multiplied by an appropriate factor. Although reviewed as a method for outlier identification in Hubert et al. (2015), to our knowledge, this is the first time this method has been applied to wind data. The new parametric approaches we develop involve fitting the BST distribution to each pressure level, and observations are flagged using either their squared Mahalanobis distances or the $100(1 - \alpha)\%$ contour of smallest geometrical size. The squared Mahalanobis distances from a BST follow a scaled F -distribution, and their contours are elliptical but not skewed. Thus, the advantage of the contour approach is that the contours produced are skewed and elliptical, and we name this method the skew elliptical contour (SEC). The scaled F -distribution and the SEC of the BST both depend on estimating the parameters in the BST. The tail-heaviness parameter in the BST is very difficult to estimate in the presence of outliers, so we develop a robust M -estimator for the parameters of this distribution.

To compare these methods, we design a simulation study in which we develop a model for simulating winds as a function of pressure level that allows us to simulate realistic wind profiles of radiosonde launches. These launches are then contaminated with different types of likely errors, and thus we are able to know whether each observation is an error or not. We evaluate the effect of changing skewness in simulated data on the ability of the methods to identify outliers. We record both the number of correctly classified errors (true positives) and incorrectly identified true observations (false negatives), with the goal of having high values of the former and low values of the latter.

This paper is organized as follows: in Section 2, we explain the Filzmoser et al. (2005) BVN and Rousseeuw et al. (1999) bagplot method, and we introduce our new methods based on the BST. In Section 3, we describe the robust estimation of the BST parameters and present the results of a simulation study in which we investigate the choice of a tuning parameter. In Section 4, we describe and report the results of the outlier simulation study, and Section 5 displays the results of applying the methods to the radiosonde winds at the Denver launch station. Section 6 concludes, describes future work, and outlines additional considerations to be taken into account

134 when applying such random error detection methods to the entire radiosonde archive.

135 2 Methods

136 In this section, we discuss different outlier detection methods for bivariate data. In Section 2.1, we
137 review the method proposed by Filzmoser et al. (2005) for data following a multivariate normal
138 distribution, where the squared Mahalanobis distance with robust estimates is used as the test
139 statistic to detect multivariate outliers. In Section 2.2, we review the bagplot by Rousseeuw et
140 al. (1999) based on Tukey's bivariate data depth. It is rank-based and does not need distributional
141 assumptions. Motivated by the skewed and heavy-tailed distribution of wind, we develop two
142 approaches to detect outliers with the bivariate skew- t (BST) distribution in Section 2.3, using the
143 Mahalanobis distance whose distribution is determined by the tail-heaviness parameter, and the
144 corresponding bivariate skewed elliptical contours (SEC) which takes the skewness into account.
145 Since both approaches involve fitting the BST distribution, robust estimation is necessary in the
146 presence of outliers and is presented in Section 3. Throughout this section, we assume that G is
147 an absolutely continuous probability distribution in \mathbb{R}^2 , and $\{\mathbf{Y}_1, \dots, \mathbf{Y}_n\}$ is a random sample
148 from G , where $\mathbf{Y}_i = (U, V)_i^T$, $i = 1, \dots, n$.

149 2.1 Bivariate Normal

Filzmoser et al. (2005) assume that the underlying data-generating process is a multivariate normal distribution and use Mahalanobis distance to measure the distance of observations from the center of the distribution, as follows:

$$MD_i = ((\mathbf{Y}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}))^{1/2},$$

and it is well-known that $MD_i^2 \sim \chi_p^2$, where p is the dimension of the observations, and in our case $p = 2$. However, in place of the classical estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, they use the Minimum Covariance Determinant (MCD) approach proposed by Rousseeuw (1985). For this method, the location and covariance matrix are estimated using only the observations of size h that minimizes

the determinant of the sample covariance. They use $h \approx 0.75n$, where n is the sample size. This approach is computationally fast (Rousseeuw and Van Driessen 1999), and robust distances are then formed as follows:

$$RD_i = ((\mathbf{Y}_i - \boldsymbol{\mu}_r)' \boldsymbol{\Sigma}_r^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_r))^{1/2},$$

150 where $\boldsymbol{\mu}_r$ and $\boldsymbol{\Sigma}_r$ are the robust estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

Then, a threshold with which to classify the observations as outliers is required. As the sample size increases, it is expected to observe more extreme values from the distribution, which are desirable to retain, so this threshold is adjusted for various sample sizes and data dimension (n and p). They let $F_n(u)$ represent the empirical distribution function of the squared robust distances, and $F(u)$ is the distribution function of χ_p^2 . As n gets large, $F_n(u) \rightarrow F(u)$, so the tails can be compared to detect outliers. To measure deviations in the tails of these distributions, first define $\delta = \chi_{p;1-\alpha}^2$ for a small α . We use $\alpha = 0.025$. Then, deviation in the upper tail of the distribution is measured with

$$p_n(\delta) = \sup_{u \geq \delta} (F(u) - F_n(u))^+,$$

where $+$ indicates the positive differences. A critical value to distinguish between outliers and extremes is defined as

$$\alpha_n(\delta) = \begin{cases} 0, & p_n(\delta) \leq p_{crit}(\delta, n, p), \\ p_n(\delta), & p_n(\delta) > p_{crit}(\delta, n, p), \end{cases}$$

and the threshold value is defined to be $c_n(\delta) = F_n^{-1}(1 - \alpha_n(\delta))$. Filzmoser et al. (2005) find $p_{crit}(\delta, n, p)$ through simulation of 1,000 observations for specific combinations of n and p . In experimenting with $p \leq 10$, they find that the empirical relationship

$$p_{crit}(\delta, n, p) = \frac{0.24 - 0.003p}{\sqrt{n}}$$

151 can be used to quickly choose p_{crit} . This method is implemented and available in the `mvoutlier`
152 package in R.

153 [Ying—we need to talk about the Benjamini-Hochberg adjustment that we use

154 instead. Should we do it here? We could use the following text:] In skewed distributions,
155 this method flags a substantial number of observations in the heavy tail as outliers, making this
156 threshold much too low. Thus, we compute the following probability: $P(\chi_p^2 > RD_i^2)$, that a
157 χ^2 random variable with p degrees of freedom exceeds the squared robust distance, associating
158 each observation with a “ p -value.” To control the expected false discovery rate, we apply the
159 Benjamini-Hochberg adjustment for dependent hypothesis tests by ordering all of the p -values
160 from smallest to largest as $(p_{(1)}, p_{(2)}, \dots, p_{(m)})$. Then, we find the largest j such that $p_{(j)} < \frac{j\alpha}{m \cdot C_m}$,
161 where $C_m = \sum_{i=1}^m 1/i$, and flag the observations corresponding to $p_{(1)}, p_{(2)}, \dots, p_{(j)}$ as outliers
162 (Benjamini and Hochberg 1995; Benjamini and Yekutieli 2001). This approach improves the
163 performance of the BVN method in both the outlier simulation and the case study.

164 2.2 Bivariate Depth

165 The bivariate normal method detects outliers that deviate from the BVN distribution. When the
166 underlying distribution is unknown, a nonparametric rank-based method is appropriate. How-
167 ever, we cannot simply extend univariate order statistics to the multivariate setting because
168 of the absence of a natural ordering for multi-dimensional Euclidean space. Data depth is an
169 important concept for multivariate data ordering and for understanding the close relationships
170 among ranking quantiles, outlyingness, and robustness. There are many available notions of data
171 depth, for example, the Mahalanobis depth (Mahalanobis 1936), the Tukey halfspace location
172 depth (Tukey 1975), the Oja depth (Oja 1983), the simplicial depth (Liu 1990), the majority
173 depth (Singh 1991), and the likelihood depth (Fraiman and Meloche 1999), many of which were
174 reviewed by Liu (1999). The general idea is that one can compute the depth values of all the
175 observations for a given depth notion, and order them according to the decreasing depth values.
176 Then, the first order statistic, $\mathbf{Y}_{[1]}$, associated with the largest depth value is defined as the
177 median, which is the most central or the deepest observation. The order statistics, $\mathbf{Y}_{[1]}, \dots, \mathbf{Y}_{[n]}$,
178 induced by data depth start from the most central data point and move outwards in all direc-
179 tions. The implication is that a larger rank is associated with a more outlying position, or less
180 representative observation with respect to the data cloud.

Here we present the Tukey half-space depth (Tukey 1975) in \mathbb{R}^2 . The half-space depth of \mathbf{y} w.r.t. G is defined to be

$$D(G; \mathbf{y}) = \inf_H \{P(H) : H \text{ is a closed half-space in } \mathbb{R}^2 \text{ and } \mathbf{y} \in H\},$$

and its sample version is denoted by $D(G_n; \mathbf{y})$, where G_n is the empirical distribution function of the sample $\{\mathbf{Y}_1, \dots, \mathbf{Y}_n\}$ and is given by the smallest number of data points contained in a closed half-plane of which the boundary line passes through \mathbf{y} . [Should we define $P(H)$ in equation for $D(G; \mathbf{y})$?] Then, the sample $100\alpha\%$ ($0 < \alpha < 1$) central region is naturally defined as the convex hull containing the most central α proportion of the deepest sample points. In particular, the sample 50% central region is

$$C_{0.5} = \text{convex hull}\{\mathbf{Y}_{[1]}, \dots, \mathbf{Y}_{[\lceil n/2 \rceil]}\},$$

where $\lceil n/2 \rceil$ is the smallest integer not less than $n/2$. Rousseeuw et al. (1999) proposed the bagplot to inflate the boundary of the 50% central region by a factor of 3 as the fence to detect outliers. Figure 3 shows a set of bivariate observations with the bivariate bagplot overlaid, where the central blue dot is the median which has the highest Tukey's depth value; the convex hull in red contains the 50% of observations with the highest Tukey's depth values; and the distance from the median to each vertex on the red convex hull is inflated by 3 to get the outer blue convex hull. Note that the bagplot naturally preserves asymmetry in the bivariate distribution.

2.3 Bivariate Skew-T

Following similar notation to Azzalini and Capitanio (2003), the MST distribution is as follows:

$$f_{\mathbf{Y}}(\mathbf{y}) = 2 t_p(\mathbf{y}; \nu) T_1 \left\{ \boldsymbol{\alpha}' \boldsymbol{\omega}^{-1} (\mathbf{y} - \boldsymbol{\xi}) \left(\frac{\nu + p}{Q_{\mathbf{y}} + \nu} \right)^{1/2}; \nu + p \right\},$$

where $\boldsymbol{\omega} = \text{diag}(\omega_{11}, \dots, \omega_{pp})^{1/2}$ for ω_{ij} the i th, j th entry of $\boldsymbol{\Omega}$, a $p \times p$ scale matrix; $\boldsymbol{\xi}$ is the location vector;

$$Q_{\mathbf{y}} = (\mathbf{y} - \boldsymbol{\xi})' \boldsymbol{\Omega}^{-1} (\mathbf{y} - \boldsymbol{\xi});$$

189 $t_p(\cdot; \nu)$ is the density of a p -dimensional t -distribution with ν degrees of freedom; and
190 $T_1(\cdot; \nu + p)$ is the distribution of the univariate t -distribution with $\nu + p$ degrees of freedom. A
191 p -dimensional random vector following this distribution is denoted $\mathbf{Y} \sim St_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, \nu)$. When
192 $\boldsymbol{\alpha} = (0, \dots, 0)_p^T$ and $\nu = \infty$, then the multivariate normal distribution with mean $\boldsymbol{\xi}$ and variance-
193 covariance $\boldsymbol{\Omega}$ is recovered. For the bivariate case with $p = 2$, there are 8 unknown parameters to
194 estimate.

195 **2.3.1 F-Distance**

196 Our first method is based on the quadratic form, $Q = (\mathbf{Y} - \boldsymbol{\xi})' \boldsymbol{\Omega}^{-1} (\mathbf{Y} - \boldsymbol{\xi})$. In this case,
197 $Q \sim p \cdot F(p, \nu)$, and this property has been used to construct Healy plots (Healy 1968) for
198 assessing the fit of data to the MST by comparing this quadratic form to quantiles from the
199 associated F -distribution. The distribution of Q depends on the degrees of freedom parameter,
200 and this parameter's estimation is extremely sensitive to the presence of outliers. In Section 3,
201 we describe how to robustly estimate this parameter. However, given robust estimates of the
202 parameters, denoted $\boldsymbol{\xi}_r$, $\boldsymbol{\Omega}_r$, $\boldsymbol{\alpha}_r$, and ν_r , we construct $Q_r = (\mathbf{Y} - \boldsymbol{\xi}_r)' \boldsymbol{\Omega}_r^{-1} (\mathbf{Y} - \boldsymbol{\xi}_r)$ and flag
203 **the observation as an outlier if this value exceeds $F_{1-\alpha}(p, \nu_r)/p$, which is the scaled**
204 **quantile of an $F(p, \nu_r)$ distribution with α lying under the curve and to the right.**
205 Note that the univariate α is the Type I error rate that we set to 0.025, and $\boldsymbol{\alpha}$ is the $p \times 1$
206 vector in the MST that controls the direction and strength of skewness in each dimension of the
207 distribution.

208 [Ying—actually, here is how we flag outliers, just verify and then we can replace
209 the bold above with this instead:] and for each observation we compute $P(F(p, \nu_r)/p > Q_r)$,
210 the probability that a scaled F -distribution with p and ν_r degrees of freedom exceeds Q_r . To flag
211 outliers, we follow the same Benjamini-Hochberg approach as is applied for the BVN method,
212 setting the Type I error rate to $\alpha = 0.025$.

213 **2.3.2 Skew Elliptical Contours**

214 The disadvantage of the approach based on the quadratic form described above is that its distri-
215 bution is invariant to the skewness of the multivariate distribution, represented by $\boldsymbol{\alpha}$. We want
216 to flag outliers that are outside the main cloud of points differently depending on the direction
217 that they lay with respect to the center. Those that are in the “long skewed tail” of the distri-
218 bution should be less likely to be flagged as outliers than those that may be relatively close to
219 the center but in the “short skewed tail,” as illustrated in Figure 4. The right-hand plot shows
220 the contours produced based on the F -distribution, and even though the red point is outside the
221 area of highest density, it would not be flagged as unusual. However, in the left-hand plot, if the
222 contours are skewed to match the density of the observations, then the red observation would be
223 flagged.

Thus, we need to find the region $R_{BST} \subset \mathbb{R}^2$ of smallest geometrical size such that $P(\mathbf{Y} \in R_{BST}) = (1 - \alpha)$. This is equivalent to stating that the solution must be of the type

$$R_{BST} = \{\mathbf{y} : f_{BST}(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, \nu) \geq f_0\},$$

where f_{BST} is the pdf of the BST, and f_0 is a value ensuring that $P(\mathbf{Y} \in R_{BST}) = (1 - \alpha)$. R_{BST} is a convex set since the BST density is concave, and the goal is to find a suitable f_0 . Azzalini and Capitanio (2014) state that an exact solution for f_0 does not seem feasible in the multivariate skew-normal case, and they construct an approximation based on the χ_p^2 solution for the MVN. Soriano (2007) extends this approximation for the MST case, and these implementations can be found in the `sn` package in R. Thus, we flag observations as outliers when they are outside of the approximate region of smallest geometrical size for $\alpha = 0.025$ based on robust estimates of the BST parameters, defined as

$$\tilde{R}_{BST} = \{\mathbf{y} : f_{BST}(\mathbf{y}; \boldsymbol{\xi}_r, \boldsymbol{\Omega}_r, \boldsymbol{\alpha}_r, \nu_r) \geq f_0\}.$$

224 [Ying—we need to update the text used to describe how observations are flagged.]

225 3 Robust Estimation in Bivariate Skew-T

226 The MST density allows for the modeling of skewed and heavy-tailed distributions, and estimates
 227 of the parameters are typically computed via maximum likelihood. While maximum likelihood
 228 estimators have nice properties, they are not guaranteed to be robust to outliers, and the “tail-
 229 heaviness” parameter, ν , is very sensitive to the presence of outliers. In particular, the estimate
 230 of ν is negatively biased when outliers are present. While the univariate skew- t density has been
 231 used as the error term in regression models to estimate covariate coefficients robustly (Azzalini
 232 and Genton 2008; Azzalini and Capitanio 2014), here, our goal is to construct robust estimates
 233 of the parameters in the BST directly, which has not been done for the ST distribution.

To understand the robustness of a particular estimator, it is common to look at the influence functions (IF) for each of the parameters (Huber and Ronchetti 2009). Letting $\boldsymbol{\theta} \in \mathbb{R}^d$ represent the vector of d parameters, the influence function for the estimator of the j th element of $\boldsymbol{\theta}$ is defined as

$$IF_j(\mathbf{y}) = \lim_{t \rightarrow 0^+} \frac{\hat{\theta}_j(t\Delta_{\mathbf{y}} + (1-t)G) - \hat{\theta}_j(G)}{t}, \quad j = 1, \dots, d,$$

234 where $\hat{\theta}_j$ is the estimator of interest, G is the true underlying distribution, and $\Delta_{\mathbf{y}}$ is the distri-
 235 bution of a point mass at \mathbf{y} , where \mathbf{y} is a $p \times 1$ vector and would include the case where \mathbf{y} is
 236 univariate. Thus, the influence function $IF_j(\mathbf{y})$ provides a measure of how an estimator changes
 237 due to contamination at the point \mathbf{y} . An estimator is generally considered robust if its influence
 238 function is bounded.

239 Huber and Ronchetti (2009) discuss the general class of M -estimators; these are estimators
 240 computed by minimizing some function $\rho(\mathbf{y}; \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. Note that maximum likelihood
 241 estimators (MLEs) belong to this broad class; this can be seen by choosing $\rho(\mathbf{y}; \boldsymbol{\theta}) = -\ell(\boldsymbol{\theta}; \mathbf{y})$,
 242 where $\ell(\cdot)$ is the joint log-likelihood function of the data. Assuming that the density, $g(\mathbf{y})$, exists,
 243 the influence function for such estimators can be computed as

$$IF_j(\mathbf{y}) = -\frac{\psi_j(\mathbf{y}; \hat{\boldsymbol{\theta}})}{\int \frac{\partial \psi_j(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_j} g(\mathbf{x}) d\mathbf{x}},$$

244 where

$$\psi_j(\mathbf{y}; \boldsymbol{\theta}) = \frac{\partial \rho(\mathbf{y}; \boldsymbol{\theta})}{\partial \theta_j}. \quad (1)$$

245 For the MST distribution, the derivatives of the negative log-likelihood for all parameters have
246 closed forms except for ν , which can be computed numerically (Azzalini and Capitanio 2003).
247 The solid line in Figure 6 shows the influence functions for the MLEs of all four parameters in
248 a univariate skew- t . Both $\hat{\omega}_{MLE}$ and $\hat{\nu}_{MLE}$ appear to suffer from strongly unbounded influence
249 functions. Lucas (1997) also shows that the influence functions for the scale and degrees of
250 freedom in a univariate student- t distribution are unbounded when they are estimated from the
251 data.

252 To derive a robust M -estimator for the parameters of the skew- t distribution, a natu-
253 ral approach is to consider a redescending M -estimator, where the derivative of ρ satisfies
254 $\lim_{z \rightarrow \pm\infty} \rho'(z) = 0$. We propose the following redescending M -estimator for the skew- t dis-
255 tribution by redefining $\rho(\cdot)$ as a function of $z = -\ell(\boldsymbol{\theta}; \mathbf{y})$,

$$\rho_r(z) = \begin{cases} z, & z \leq k, \\ 2k - k \cdot \exp(-z/k + 1), & z > k, \end{cases} \quad (2)$$

256 where $k > 0$ is a constant controlling the threshold at which observations are considered ex-
257 treme. Figure 5 shows the trajectory of the function $\rho_r(z)$ for $k = 2$. This particular choice
258 of $\rho_r(\cdot)$ in Equation (2) has advantages. First of all, it makes no adjustment to observations
259 with small (and hence reasonable) negative log-likelihoods, and it is continuously differentiable.
260 Furthermore, when the negative log-likelihood goes to infinity, $\lim_{z \rightarrow \infty} \rho_r(z) = 2k$ as shown in
261 Figure 5 where the limiting value is 4, and $\lim_{z \rightarrow \infty} \rho'_r(z) = 0$ indicating that the M -estimator is
262 indeed redescending. The corresponding IFs of all the estimators are strongly redescending to 0
263 as shown in dashed lines in Figure 6. The function $\rho_r(\cdot)$ could be one of many, although we do
264 not explore other alternatives here. The optimality of redescending M -estimators for the skew- t
265 distribution requires future research, especially for ω and ν . Similar discussions for the location
266 parameter can be found in Shevlyakov et al. (2008).

267 An alternative approach to robust estimation with redescending M -estimators is trimmed like-

268 lihood, as described in Hadi and Luceño (1997). Rather than bounding the negative log-likelihood
 269 directly, a fixed percent of observations with the largest individual negative log-likelihood values,
 270 $-\ell(\boldsymbol{\theta}; \mathbf{y}_i)$, are removed. We considered this method as well; but in a BST distribution, some very
 271 unusual values still occur in the heavy tail of the distribution (see left panel of Figure 7), and
 272 removing such values completely results in an overestimation of ν .

As a trade-off, we must determine an appropriate choice of k in Equation (2), and we want to choose a value that will work equally well in a variety of circumstances. We proposed and tested multiple approaches to choosing k dynamically and present two of these here. The left panel of Figure 7 shows a sample simulated dataset that has been contaminated with outliers, which are marked in red, in the heavy tail. First, we find the k associated with the minimum of the density of the observations' likelihood values, denoted k_{min} . To do this, the non-robust maximum likelihood estimates are found, and the likelihood of each observation based on these estimates is computed to form the following set: $\{\ell(\hat{\boldsymbol{\theta}}; \mathbf{y}_1), \ell(\hat{\boldsymbol{\theta}}; \mathbf{y}_2), \dots, \ell(\hat{\boldsymbol{\theta}}; \mathbf{y}_n)\}$. Then, the nonparametric density of this set of likelihood values is constructed, as shown in the center panel of Figure 7 with the dashed line showing the minimum value of the density. In this case, the k chosen would be $k_{min} = 13.3$. A second approach to selecting k is to construct the empirical cumulative density function (ECDF) of the set of non-robust individual likelihood values, denoted $\hat{\mathcal{L}}(\hat{\boldsymbol{\theta}}; \mathbf{y})$. Then, k_{deriv} , is chosen such that the following criteria hold,

$$\hat{\mathcal{L}}'(\hat{\boldsymbol{\theta}}; \mathbf{y}) < \epsilon_1 \text{ and } |\hat{\mathcal{L}}''(\hat{\boldsymbol{\theta}}; \mathbf{y})| < \epsilon_2.$$

273 where we set $\epsilon_1 = 0.01$ and $\epsilon_2 = 0.001$. The right panel in Figure 7 illustrates this choice of k
 274 for this dataset, which is 12.7. We perform a simulation study to evaluate the choice of k on the
 275 mean square error (MSE) of the robust parameter estimates and present the study design and
 276 results in the next two subsections.

²⁷⁷ **3.1 Robust BST Simulation Design**

²⁷⁸ In this section, we perform a simulation study and vary features that we expect to have
²⁷⁹ an impact on the robust estimation of the BST parameters. The amount of variability in the
²⁸⁰ distribution; the shape and behavior of the tails; the direction of outlier contamination; and
²⁸¹ the percent of outlier contamination are all features that can influence the parameter estimates.
²⁸² For example, based on pilot studies, we know that the estimation of ν is quite sensitive to
²⁸³ the presence of outliers, and those distributions that exhibit more variability have even more
²⁸⁴ difficulty in estimating ν . Below are the steps we take to simulate data:

- ²⁸⁵ 1. Simulate a dataset of size $n \in \{100, 250, 500, 750, 1000\}$ from a BST with $\xi = (16.51, -1.22)^T$
²⁸⁶ and one of the following 5 levels of variability: $\Omega_1 = I_2$.

$$\Omega_2 = \begin{pmatrix} 22.4 & -3.9 \\ -3.9 & 25.9 \end{pmatrix}, \quad \Omega_3 = \begin{pmatrix} 85.6 & 10.5 \\ 10.5 & 63.8 \end{pmatrix},$$

$$\Omega_4 = \begin{pmatrix} 147.9 & 38.9 \\ 38.9 & 210.0 \end{pmatrix}, \quad \text{or} \quad \Omega_5 = \begin{pmatrix} 171.8 & 44.6 \\ 44.6 & 242.7 \end{pmatrix}$$

²⁸⁷ and one of the following sets of remaining parameter values:

- ²⁸⁸ (a) Symmetric (MVN): $\alpha = (0, 0)^T$, $\nu = \infty$
²⁸⁹ (b) Observed (OBS): $\alpha = (1, -1)^T$, $\nu = 10$
²⁹⁰ (c) Extreme (EX): $\alpha = (6, 0)^T$, $\nu = 5$

²⁹¹ The values of the parameters chosen for the simulation are based on the MLEs of the BST
²⁹² for the u and v components at the Denver launch station. The Ω 's are chosen based on
²⁹³ particular pressure levels, with Ω_2 being the least variable level and Ω_5 being the most
²⁹⁴ variable level. The others are chosen to fill in within this range. Set (a) above corresponds
²⁹⁵ to multivariate normal observations; set (b) values are similar to what was in the observed
²⁹⁶ data; and set (c) is more skewness and tail-heaviness than observed.

297 2. Contaminate either 0%, 5%, or 10% of the observations with one of two types of outliers.

298 The first type introduces errors by shifting the simulated value a distance that is chosen
299 uniformly at random between 12 and 16 units anywhere around the main cloud of points.

300 The second method differs in that it samples a direction such that the outliers occur in
301 the heavy tail of the distribution. Outliers of this type may occur if there is a common
302 prevailing wind direction. The left panel of Figure 7 depicts an example wherein 10% of the
303 observations are contaminated with outliers in the heavy tail of an EX skewed distribution.

304 For each of 250 datasets simulated under a given setting, as described above, the following steps

305 are performed:

1. The simulated and contaminated data is denoted $\mathbf{X} = (\mathbf{u}, \mathbf{v})' \in \mathbb{R}^{n \times 2}$, which follows the BST, $\mathbf{X} \sim ST_2(\boldsymbol{\xi}, \boldsymbol{\Omega}_i, \boldsymbol{\alpha}, \nu)$. The variability of the simulated data impacts the estimation, so we first center and scale the data. Let $\mathbf{a} = (M_u, M_v)'$ be the vector of medians of \mathbf{u} and \mathbf{v} . The respective medians are first subtracted from the data, and then the data is rescaled by multiplying it by \mathbf{A} where \mathbf{A}^{-1} is

$$\mathbf{A}^{-1} = \begin{pmatrix} MAD(\mathbf{u}) & 0 \\ 0 & MAD(\mathbf{v}) \end{pmatrix},$$

306 and the $MAD(\cdot)$ is the median absolute deviation, a robust estimate of scale. Thus, the
307 centered and scaled data is $\mathbf{Y} = \mathbf{A}(\mathbf{x} - \mathbf{a}) = \mathbf{Ax} - \mathbf{Aa}$. Note, that the true parameter
308 values change when the affine transformation is applied. Thus, the distribution of \mathbf{Y} is
309 $\mathbf{Y} \sim ST_2(\mathbf{A}(\boldsymbol{\xi} - \mathbf{a}), \mathbf{A}\boldsymbol{\Omega}_i\mathbf{A}, \boldsymbol{\alpha}, \nu)$. Generally speaking, the value of $\boldsymbol{\alpha}$ would change as
310 well for such a transformation, but when \mathbf{A} is diagonal, $\boldsymbol{\alpha}$ is unchanged, and the proof
311 is given in the Appendix. Here, we will denote these transformed parameters as follows:
312 $\mathbf{Y} \sim ST_2(\boldsymbol{\xi}', \boldsymbol{\Omega}', \boldsymbol{\alpha}, \nu)$.

- 313 2. The parameters are estimated robustly with either k_{min} , k_{deriv} , or $k = 10$. In a preliminary
314 simulation, we estimated the parameters robustly for every k in the set $\{2, 4, \dots, 26\}$. We
315 found that $k = 10$ or $k = 12$ was often chosen as the value of k that minimized the MSE of

316 the parameters, but in practice, we recommend the dynamic choice of k since the simulated
317 data that we generate here cannot cover all possible parameter settings. In particular, we
318 noticed that larger values of k were chosen as the sample size increased.

319 3. For each simulated dataset, the true parameter values and the estimates are saved. The
320 MSE (variance of the estimator plus the bias squared) across all 250 simulations is computed
321 and then averaged across all of the parameters except the degrees of freedom. It does not
322 make sense to compute the bias of the degrees of freedom parameter, $\hat{\nu} - \nu$, under the
323 symmetric simulation setting (a) above since $\nu = \infty$. Thus, the median of the estimated
324 degrees of freedom parameter is reported instead.

325 3.2 Robust BST Simulation Results

326 Complete simulation results for $n = 500$ are given in Table 1, and findings for the other sample
327 sizes are similar. Across all of the simulation settings, the amount of variability present in
328 the data does not appear to influence results as long as the robust standardization described
329 above is performed. With no outliers, all three choices of k perform similarly across the first
330 seven parameters, but using k_{min} yields the best estimate of ν , regardless of the shape of the
331 distribution. Using k_{deriv} results in an overestimation of ν by about 3 units for the OBS data
332 and by 1 unit in the EXT case, and $k = 10$ performs worse than that. The difference is due
333 to the fact that, on average, the value of k chosen with k_{min} is much larger, as it should be,
334 and it increases substantially as the strength of the skewness in the underlying simulated data
335 increases.

336 Once the simulated data is contaminated with outliers, the placement of the outliers, the
337 number of outliers, and the strength of the skewness of the distribution all play a role in how
338 well each choice of k performs in the parameter estimation. The MSE of the first seven parameters
339 is approximately the same for all three choices of k for 5% All, 5% Angle, and 10% All in the SYM
340 and OBS datasets. However, the EXT skew data has better MSE for the parameters with k_{deriv}
341 and $k = 10$, and the OBS skewness with 10% Angle contamination shows some preference for
342 these two choices as well. In estimating ν in the presence of contamination, we see the following:

- 343 • Higher outlier contamination makes ν more difficult to estimate.
- 344 • Outliers in the tail of the distribution make ν more difficult to estimate than when they
345 are placed all around the main cloud of points.
- 346 • All three choices of k result in a median estimate of ν that is smaller than the true value with
347 some exceptions when the distribution is symmetric and contamination does not exceed
348 5%.
- 349 • A choice of k_{min} results in the worst estimation of ν when outliers are present. Both k_{deriv}
350 and $k = 10$ produce similar estimates of ν , with $k = 10$ performing better in some instances.

351 Based on these results, we proceed to use k_{deriv} to choose k dynamically in the estimation,
352 which gives more flexibility over using a static value such as $k = 10$ but loses very little in terms
353 of overall accuracy of the parameter estimation. In particular, we see that the value of k_{deriv} that
354 is chosen increases with sample size and the strength of the skewness of the data and decreases
355 by a small amount as the outlier contamination increases, all of which are desirable features.
356 Additionally, the parameter estimates are not strongly influenced by changing k by one or two
357 units.

358 4 Simulation Studies

359 **[Ying's section to modify. Note that Figure 8 will need to be updated to reflect
360 changes in the outlier models.]** In this section, we simulate radiosonde launches and con-
361 taminate them with realistic types of errors. Then, we apply all of the methods presented herein
362 to determine which one has the highest true negative and lowest false positive rates.

363 4.1 Outlier Simulation

364 To compare the methods introduced in Section 2, we generate 500 launches at 8 pressure levels
365 from 16-dimensional skew- t distributions with one of 3 types of skewness and degrees of freedom.
366 We use MVN to denote a multivariate normal distribution with no skewness and infinite degrees

367 of freedom, OBS to denote a multivariate skew-*t* (MST) distribution with similar skewness to
 368 what is observed in the Denver radiosonde data and degrees of freedom 10, and EX to denote a
 369 MST with more extreme skewness than observed in the data and degrees of freedom 5. For each
 370 distribution type, we consider one basic model (model 1) without contamination and 3 outlier
 371 models (models 2-4) with 5% and 10% contamination. The top panel of Figure 8 shows one
 372 realization from model 1 for each of the 3 distribution types. We contaminate an entire launch
 373 in model 2, random higher levels of a launch in model 3 and random half levels in model 4 as
 374 shown in the bottom panel of Figure 8 for the OBS distribution. Model details are described as
 375 follows:

- 376 1. Model 1 is a basic model without outliers: $\mathbf{X}_i = (\mathbf{Y}_1^T, \dots, \mathbf{Y}_8^T)_i^T$, $i = 1, \dots, 500$, generated
 377 from one of the 3 distribution types, where $\mathbf{Y}_j = (U, V)_j^T$ is a bivariate vector at level j ,
 378 $j = 1, \dots, 8$. In models 2–4 below, the u and v components are always randomly selected
 379 to be contaminated. We only use \mathbf{X}_i for simplicity.
- 380 2. Model 2 includes contamination of an entire launch: $\mathbf{Z}_i = \mathbf{X}_i + s_i d_i \sigma_j K$, where s_i is 1
 381 with probability τ (100 $\tau\%$ contamination), and 0 otherwise, d_i is a sequence of random
 382 variables independent of s_i taking values 1 and -1 with probability $1/2$, $K = 6$ is a constant
 383 indicating the contamination size, and σ_j is the standard deviation at level j .
- 384 3. Model 3 is only contaminated at random higher levels of a launch : $\mathbf{Z}_i(j) = \mathbf{X}_i(j) + s_i d_i \sigma_j K$,
 385 only if $j \geq \ell_i$, where ℓ_i is a random number uniformly generated from $\{2, \dots, 7\}$.
- 386 4. Model 4 is similar to model 3 but contaminated by randomly selecting 4 out of 8 levels.

387 Although data are generated as entire launches, we apply outlier detection methods level by
 388 level in order to identify bivariate outliers rather than the entire launch. For each model and
 389 distribution combination, we consider all the methods introduced in Section 2: the bivariate
 390 normal (BVN) with Type I error $\alpha = 0.025$, the bivariate data depth (BDP) with the inflating
 391 factor of 3, the bivariate skew-*t* with parameter estimated by trimmed data (BSTtrim), where
 392 97.5% of sample points that have the largest depth values on each level are kept, and the bivari-
 393 ate skew-*t* with robust estimators (BSTrobust). For BSTtrim and BSTrobust, two approaches

394 detailed in Section 3 are applied, the one based on the scaled F distribution of the Mahalanobis
395 distance (F), and the one using skewed elliptical contours (SEC). To show the necessity of the
396 robust estimation, we also include the results from the bivariate skew- t with direct parameter
397 estimation (BST), and with the true parameters (BSTtrue).

398 The performance of the outlier detection is summarized by true positives (TP) and false
399 positives (FP). In this simulation, TP is the percentage of the time that one method correctly
400 flags outliers, and FP is the percentage of the time that one method incorrectly flags non-
401 outliers. For example, when data is not contaminated with any outliers as shown in Table 2, TP
402 is computed as the percentage of simulations that each method detects no outliers; when data
403 is contaminated with outliers, TP is defined as the averaged percentage of correctly detected
404 outliers out of the total number of outliers over 1000 simulations. Therefore, a high TP with a
405 low FP is an indicator of a better method. Values in Table 3 are percentages of the averaged TP
406 and FP with 10% and 5% contamination, and all the standard errors (not shown) are fairly small.
407 It shows that all the methods perform better when the contamination proportion is smaller.

408 Overall, the BVN method flags too few outliers with low TP and FP, possibly due to the
409 adjustment where the multivariate normal distribution is assumed. The nonparametric BDP
410 approach works well and consistently across different outlier models, but does not do as well as a
411 parametric approach, when the parametric model fits. One may consider to adjust the inflating
412 factor in a bagplot according to the underlying distribution. For example, in the univariate case,
413 Hubert and Vandervieren (2006) adjusted the boxplot using a robust measure of skewness, and
414 for functional data, Sun and Genton (2012) proposed a simulaiton-based method to adjust the
415 factor in a functional boxplot (Sun and Genton 2011). However, in this simulation study, we
416 keep this factor fixed to show the effect of the skewness and heavy-tailness.

417 The results for the BST method with true parameters show the best performance of the
418 F -distance and the SEC approaches, where the TP is close to 100% for outlier models and
419 FP retains the nominal level of 2.5%. In contrast, the BST method with direct MLEs does very
420 poorly, indicating that it is crucial to estimate the parameters of the bivariate skew- t distribution
421 in a robust fashion in order to effectively detect outliers. Between the two robust methods we

422 have considered, the BSTrobust method performs better with higher TP but flags too many true
423 values, whereas the BSTtrim method only shows some robustness in parameter estimation due to
424 the misspecified trimmed proportion. Using the same parameter estimates in a bivariate skew- t
425 distribution, the approach based on SECs is generally better than the one based on a scaled
426 F distribution, especially when more skewness is present as illustrated in Section 2.3. For the
427 BSTrobust method, $k = 10$ is used as demonstrated in Section ?? that the performance of the
428 outlier detection is not sensitive to the choice of k .

429 5 Case Study

430 Several databases housing radiosonde launch measurements are maintained, such as the Compre-
431 hensive Historical Upper-Air Network (CHUAN) (Stickler et al. 2010) and the Integrated Global
432 Radiosonde Archive (IGRA) (Durre et al. 2004). The National Center for Atmospheric Research
433 (NCAR) maintains the Upper Air Database (UADB) archive of radiosonde data that contains
434 more launch stations and reaches further back in time than the IGRA (DSS/CISL/NCAR 2014).
435 Each of these records contain many millions of observations and could use the procedures de-
436 veloped herein to either remove or down weight suspected erroneous observations. To illustrate
437 our methods, we analyze a set of 35,816 launches from the UADB archive at the Denver, CO
438 station beginning on March 3, 1962 and ending on October 31, 2011, spanning a fifty year period.
439 Figure 2 shows the nonparametric density of all 35,816 launches at each of 8 standard pressure
440 levels.

441 As the radiosonde rises through the atmosphere, it becomes more likely for an error to occur,
442 and errors are also more likely at higher wind speeds. In addition, the number of outliers may
443 vary over time. Random outliers are more likely to occur in later years as Ramella Pralungo and
444 Haimberger (2014) note that early observations are subject to “fair weather sampling bias,” since
445 before radar tracking began in the 1960’s, theodolites were used, which only worked when there
446 was good visibility, and the winds were not too strong. Ramella Pralungo and Haimberger (2014)
447 also perform their homogenization analysis independently for each pressure level since larger
448 biases tend to occur for high wind speeds, and pressure levels are characterized by distinctly

449 different mean wind speeds. Thus, for the Denver launch station, we subset the observations
450 by pressure level for each five year time period and for a given season, yielding 320 subsets of
451 observations. For sixteen of these subsets, less than 100 observations are recorded, and these are
452 excluded from our analysis. The remaining 304 subsets have sample sizes ranging from 121 to
453 971 with a median size of 731, and only four subsets have less than 400 observations.

454 For each of the 304 subsets of observations, we apply all four methods of outlier detection,
455 namely (i) BVN; (ii) Depth; (iii) BST F; and (iv) BST SEC. We use $\alpha = 0.025$ with the
456 Benjamini-Hochberg adjustment for multiple dependent testing for the BVN and both BST
457 approaches. The Depth method flags the most outliers, 0.21% of the observations with the BST
458 SEC next with 0.15%, then BVN with 0.11%, and BST F having the fewest flagged with 0.01%
459 percent. Figure 9 illustrates the proportion of observations that are flagged as outliers by each
460 method across time and for each pressure level. From this, we see that the BVN and Depth
461 methods flag a greater proportion of outliers between the years of 1971 and 1981, particularly
462 for the higher pressure levels (i.e., closer to the surface). An additional increase in flags occurs
463 between 2001 and 2006 at the 100 Mb pressure level across all three methods, but again BVN and
464 Depth flag more observations in this time period at additional pressure levels. In many instances,
465 all of the methods are flagging the same obviously erroneous observations, but Figure 10 shows
466 two illustrative subsets wherein there are differences. In both cases, the BVN method is flagging
467 more observations in the heavier right-hand tail than is likely to be necessary. At the same time,
468 observations in the shorter left-hand tail are flagged by the SEC method but are missed by the
469 F-based distance approach. We should note that we had not modified the threshold for the BVN
470 approach that is suggested by Filzmoser et al. (2005), the BVN method would have flagged a
471 total of 1.32% of the observations as outliers instead of 0.11%.

472 Finally, we fit a logistic regression model for the proportion of flagged observations in each
473 of the 304 subsets of observations using the robust estimates of α_1 , α_2 , and ν as predictors.
474 This model was fit for each of the BVN, BST F, and BST SEC methods, and the coefficients
475 along with their associated p -values are given in Table 4. Here, we see that as skewness of the
476 density increases (either $\hat{\alpha}_1$, $\hat{\alpha}_2$, or both), all three methods have significantly higher odds of

477 flagging an observation as an outlier; however, the BST-SEC method appears to be the least
478 affected by skewness as only one of the skewness parameters is significant, and the strength of
479 the significance is the lowest among the three methods. Interestingly, the estimated degrees of
480 freedom is only significant for the BVN method, and the effect is negative. In other words, as
481 the estimated degrees of freedom increases, (meaning that the kurtosis decreases), the odds of
482 flagging an outlier decreases as well. Both the BST-F and BST-SEC methods are not significantly
483 influenced by the kurtosis of the distribution.

484 6 Conclusions

485 We have applied our methods to one launch station in Denver, and the performance of the SEC-F
486 method produces the best results, but the ultimate goal is to extend and apply these methods with
487 the entire radiosonde archive. Future work will require us to find the best subsets of observations
488 to which to apply the method of choice so that the effect of a potentially changing climate will not
489 be a factor. In addition, the optimization required to find the robust BST parameters needs to
490 be improved to make it computationally feasible to apply to the entire archive. While the depth
491 method works reasonably well in simulation, it does not perform as well in the case study, but it
492 does have the advantage of being fast and simple to execute. On the other hand, the physically
493 interpretable robust BST parameters can be used to study differences in wind climatology over
494 space and time.

495 Additional future work will be to explore using the entire vertical column to check for errors.
496 If we use a multivariate skew- t distribution as a parametric model, then there would be 169
497 parameters to estimate, most of which come from requiring a 16×16 Ω matrix for the bivariate
498 observations collected for each of the 8 vertical pressure levels. In general, if p is the number
499 of variables, and q is the number of pressure levels, then there will be $0.5(pq)^2 + 2.5(pq) + 1$
500 parameters to estimate in the MST. The reliability of the robust parameter estimates needs to
501 be tested for a random variables of such higher dimensions. However, there is structure in the
502 variability as it increases to the mid-pressure levels and then decreases, as shown in Figure 1,
503 that can potentially be described with a smooth function, allowing us to parameterize Ω and

504 reduce its dimension. Another nonparametric approach to explore is to extend the bivariate
 505 depth with a functional depth across pressure levels, weighted by the number of observations at
 506 each level, as proposed in Hubert et al. (2015).

507 Acknowledgements

508 The authors would like to thank Adelchi Azzalini, Douglas Nychka, Steve Worley, and Joey
 509 Comeaux for discussions and ideas that helped to direct and focus this work.

510 Appendix

511 *Proposal:* If $\mathbf{Y} \sim ST_2(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, \nu)$, $\mathbf{X} = \mathbf{AY} + \mathbf{a}$, where \mathbf{A} and \mathbf{a} are known constants, and \mathbf{A} is
 512 diagonal, then $\mathbf{X} \sim ST_2(\mathbf{A}\boldsymbol{\xi} + \mathbf{a}, \mathbf{A}\boldsymbol{\Omega}\mathbf{A}, \boldsymbol{\alpha}', \nu)$.

513 *Proof:* By definition, for any \mathbf{A} and \mathbf{a} , $\mathbf{X} \sim ST_2(\mathbf{A}\boldsymbol{\xi} + \mathbf{a}, \mathbf{A}\boldsymbol{\Omega}\mathbf{A}, \boldsymbol{\alpha}', \nu)$, where

$$\boldsymbol{\alpha}' = \frac{1}{(1 + \boldsymbol{\alpha}^T(\bar{\boldsymbol{\Omega}} - \mathbf{B}\boldsymbol{\Omega}_x^{-1}\mathbf{B}^T)\boldsymbol{\alpha})^{1/2}} \cdot \omega_x \boldsymbol{\Omega}_x^{-1} \mathbf{B}^T \boldsymbol{\alpha}, \quad (3)$$

514 where $\boldsymbol{\Omega}_x = \mathbf{A}\boldsymbol{\Omega}\mathbf{A}$; $\mathbf{B} = \omega^{-1}\boldsymbol{\Omega}\mathbf{A}^T$; $\omega = [\text{diag}(\boldsymbol{\Omega}_{11}, \boldsymbol{\Omega}_{22})]^{1/2}$; $\omega_x = [\text{diag}(\boldsymbol{\Omega}_{x,11}, \boldsymbol{\Omega}_{x,22})]^{1/2}$; and
 515 $\bar{\boldsymbol{\Omega}} = \omega^{-1}\boldsymbol{\Omega}\omega^{-1}$.

516 First, the second term in the denominator of the expression in Equation (3) is equal to zero.

$$\begin{aligned} \boldsymbol{\alpha}^T(\bar{\boldsymbol{\Omega}} - \mathbf{B}\boldsymbol{\Omega}_x^{-1}\mathbf{B}^T)\boldsymbol{\alpha} &= \boldsymbol{\alpha}^T(\omega^{-1}\boldsymbol{\Omega}\omega^{-1} - \omega^{-1}\boldsymbol{\Omega}\mathbf{A}[\mathbf{A}^{-1}\boldsymbol{\Omega}^{-1}\mathbf{A}^{-1}]\mathbf{A}\boldsymbol{\Omega}\omega^{-1})\boldsymbol{\alpha} \\ &= \boldsymbol{\alpha}^T(\omega^{-1}\boldsymbol{\Omega}\omega^{-1} - \omega^{-1}\boldsymbol{\Omega}\omega^{-1})\boldsymbol{\alpha} \\ &= 0. \end{aligned}$$

517 So, $\frac{1}{(1 + \boldsymbol{\alpha}^T(\bar{\boldsymbol{\Omega}} - \mathbf{B}\boldsymbol{\Omega}_x^{-1}\mathbf{B}^T)\boldsymbol{\alpha})^{1/2}} = 1$. Now, if the second part of the expression in Equation 3 is equal
 518 to $\boldsymbol{\alpha}$, then we are done. First, note that $\mathbf{B}^T = \mathbf{A}\boldsymbol{\Omega}\omega^{-1}$, and

$$\begin{aligned} \omega_x &= \begin{bmatrix} a_{11}^2 \boldsymbol{\Omega}_{11} & 0 \\ 0 & a_{22}^2 \boldsymbol{\Omega}_{22} \end{bmatrix}^{1/2} \\ &= \begin{bmatrix} a_{11} \boldsymbol{\Omega}_{11}^{1/2} & 0 \\ 0 & a_{22} \boldsymbol{\Omega}_{22}^{1/2} \end{bmatrix}. \end{aligned}$$

519 Then,

$$\begin{aligned}\omega_x \mathbf{\Omega}_x^{-1} \mathbf{B}^T \boldsymbol{\alpha} &= \omega_x [\mathbf{A} \mathbf{\Omega} \mathbf{A}^T]^{-1} \mathbf{A} \mathbf{\Omega} \omega^{-1} \boldsymbol{\alpha} \\&= \omega_x \mathbf{A}^{-1} \mathbf{\Omega}^{-1} \mathbf{A}^{-1} \mathbf{A} \mathbf{\Omega} \omega^{-1} \boldsymbol{\alpha} \\&= \omega_x \mathbf{A}^{-1} \omega^{-1} \boldsymbol{\alpha} \\&= \begin{bmatrix} a_{11} \mathbf{\Omega}_{11}^{1/2} & 0 \\ 0 & a_{22} \mathbf{\Omega}_{22}^{1/2} \end{bmatrix} \begin{bmatrix} \frac{1}{a_{11}} & 0 \\ 0 & \frac{1}{a_{22}} \end{bmatrix} \begin{bmatrix} \mathbf{\Omega}_{11}^{-1/2} & 0 \\ 0 & \mathbf{\Omega}_{22}^{-1/2} \end{bmatrix} \boldsymbol{\alpha} \\&= \begin{bmatrix} \mathbf{\Omega}_{11}^{1/2} & 0 \\ 0 & \mathbf{\Omega}_{22}^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{\Omega}_{11}^{-1/2} & 0 \\ 0 & \mathbf{\Omega}_{22}^{-1/2} \end{bmatrix} \boldsymbol{\alpha} \\&= \mathbf{I} \boldsymbol{\alpha} = \boldsymbol{\alpha}.\end{aligned}$$

520 References

- 521 Anderson, A. N., Browning, J. M., Comeaux, J., Hering, A. S., and Nychka, D. (2016), “A
522 simulation study to compare statistical quality control methods for error detection in his-
523 torical radiosonde temperatures,” *International Journal of Climatology*, Now Online: DOI
524 10.1002/joc.4327.
- 525 Azzalini, A. and Capitanio, A. (2003), “Distributions generated by perturbation of symmetry
526 with emphasis on a multivariate skew-*t* distribution,” *Journal of the Royal Statistical Society,
527 Series B*, 65: 367–389.
- 528 Azzalini, A. and Capitanio A. (2014), *The Skew-Normal and Related Families*, Cambridge Uni-
529 versity Press: New York, NY.
- 530 Azzalini, A. and Genton, M. G. (2008), “Robust likelihood methods based on the skew-*t* and
531 related distributions,” *International Statistical Review*, 76: 106–129.
- 532 Benjamini, Y. and Hochberg, Y. (1995) “Controlling the false discovery rate: A practical and
533 powerful approach to multiple testing,” *Journal of the Royal Statistical Society, Series B*, 57:
534 289–300.
- 535 Benjamini, Y. and Yekutieli, D. (2001) “The control of the false discovery rate in multiple testing
536 under dependency,” *The Annals of Statistics*, 29:1165–1188.
- 537 Brönnimann, S. and Luterbacher, J. (2004) “Reconstructing northern hemisphere upper-level
538 fields during World War II,” *Climate Dynamics*, 22: 499–510.
- 539 Brönnimann, S., Stickler, A., Griesser, T., Ewen, T., Grant, A. N., Fischer, A. M., Schraner, M.,
540 Peter, T., Rozanov, E., and Ross, T. (2009) “Exceptional atmospheric circulation during the
541 ‘Dust Bowl’,” *Geophysical Research Letters*, 36, L08802.
- 542 Browning, J. M. and Hering, A. S. (2016), “Simultaneous treatment of random and systematic
543 errors in the historical radiosonde temperature archive,” *In Preparation*.
- 544 Data Support Section/Computational and Information Systems Laboratory/National Center for
545 Atmospheric Research/University Corporation for Atmospheric Research (2014) NCAR Up-
546 per Air Database, 1920-ongoing. Research Data Archive at the National Center for Atmo-
547 spheric Research, Computational and Information Systems Laboratory.
548 <http://rda.ucar.edu/datasets/ds370.1/>. Accessed 06 01 2016.

- 549 Durre, I., Vose, R. S., and Wuertz, D. B. (2004), "Overview of the integrated global radiosonde
 550 archive," *Journal of Climate*, 19: 53–68.
- 551 Durre, I., Vose, R. S., and Wuertz, D. B. (2008), "Robust automated quality assurance of
 552 radiosonde temperatures," *Journal of Applied Meteorology and Climatology*, 47: 2081–2095.
- 553 Filzmoser, P., Reimann C., and Garrett R. G. (2005), "Multivariate outlier detection in explo-
 554 ration geochemistry," *Computers and Geosciences*, 31: 579–587.
- 555 Filzmoser, P., Maronna, R., and Werner, M. (2008), "Outlier identification in high dimensions,"
 556 *Computational Statistics and Data Analysis*, 52: 1694–1711.
- 557 Frank, H. P. and Landberg, L. (1997), "Modelling the wind climate of Ireland," *Boundary Layer
 558 Meteorology*, 85: 359–377.
- 559 Fraiman, R. and Meloche, J. (1999), "Multivariate L -estimation," *Test*, 8, 255–317.
- 560 Gruber, C. and Haimberger, L. (2008) "On the homogeneity of radiosonde wind time series,"
 561 *Meteorologische Zeitschrift*, 17: 631–643.
- 562 Gupta, M., Gao, J., Aggarwal, C. C., and Han, J. (2014), "Outlier detection for temporal data:
 563 A survey," *IEEE Transactions on Knowledge and Data Engineering*, 25: 1–20.
- 564 Hadi, A. S. and Luceño, A. (1997), "Maximum trimmed likelihood estimators: A unified ap-
 565 proach, examples, and algorithms," *Computational Statistics and Data Analysis*, 25: 251–
 566 272.
- 567 Healy, M. J. R. (1968), "Multivariate normal plotting," *Applied Statistics*, 17: 157–161.
- 568 Huber, P. J. and Ronchetti, E. M. (2009) *Robust Statistics*. Wiley and Sons: Hoboken, NJ.
- 569 Hubert, M., Rousseeuw, P. and Segaert, P. (2015), "Multivariate functional outlier detection,"
 570 *Statistical Methods and Applications*, Now Online: DOI 10.1007/s10260-015-0297-8.
- 571 Hubert, M. and Vandervieren E. (2006), "An adjusted boxplot for skewed distributions," *Tech-
 572 nical report*, TR-06-11, KU Leuven.
- 573 Joe, H. (2015) *Dependence Modeling with Copulas*. CRC Press: Boca Raton, FL.
- 574 Jury, M. R. and Pathack, B. (1991), "A study of climate and weather variability over the Tropical
 575 Southwest Indian Ocean," *Meteorology and Atmospheric Physics*, 47: 37–48.
- 576 Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha,
 577 S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J.,
 578 Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.
 579 (1996), "The NCEP/NCAR 40-year reanalysis project," *Bulletin of the American Meteorolo-
 580 gical Society*, 437–471.
- 581 Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S., Hnilo, J. J., Fiorino, M., and Potter, G. L.
 582 (2002), "NCEP-DOE AMIP-II Reanalysis (R-2)," *Bulletin of the American Meteorological
 583 Society*, 1631–1643.
- 584 Kazor, K., Holloway, R., Cath, T., and Hering, A. S. (2016) "Comparison of linear and non-
 585 linear dimension reduction techniques for automated process monitoring of a decentralized
 586 wastewater treatment facility," *In Revision at Stochastic Environmental Research and Risk
 587 Assessment*.
- 588 Klimowksi, B. A., Bunkers, M. J., Hjelmfelt, M. R., and Covert, J. N. (2003), "Severe convective
 589 windstorms over the Northern High Plains of the United States," *Weather and Forecasting*,
 590 18: 502–519.

- 591 Liu, R. (1990), "On a notion of data depth based on random simplices," *The Annals of Statistics*,
 592 18, 405-414.
- 593 Liu, R. Y., Parelius, J. M. and Singh, K. (1999), "Multivariate analysis by data depth: descriptive
 594 statistics, graphics and inference," *The Annals of Statistics*, 27, 783-858.
- 595 Lucas, A. (1997), "Robustness of the Student-*t* based *M*-estimator," *Communications in
 596 Statistics—Theory and Methods*, 26.5: 1165–1182.
- 597 Mahalanobis, P. C. (1936), "On the generalized distance in statistics," *Proceedings of National
 598 Academy of Science of India*, 12, 49-55.
- 599 Maronna, R., Martin, R. D., and Yohai, V. J. (2006), *Robust Statistics: Theory and Methods*,
 600 Wiley: New York, NY.
- 601 Oja, H. (1983), "Descriptive statistics for multivariate distributions," *Statistics & Probability
 602 Letters*, 1, 327-332.
- 603 Peña, D. and Prieto, F. J. (2001), "Multivariate outlier detection and robust covariance matrix
 604 estimation," *Technometrics*, 43: 286–300.
- 605 Ramella Pralungo, L. and Haimberger, L. (2014) "A Global Radiosonde and tracked-balloon
 606 archive on sixteen pressure levels (GRASP) going back to 1905 – Part 2: Homogeneity
 607 adjustments for pilot balloon and radiosonde wind data," *Earth System Science Data*, 6:
 608 297–316.
- 609 Reimann, C., Filzmoser, P., and Garrett, R. G. (2005), "Background and threshold: Critical
 610 comparison of methods of determination," *Science of the Total Environment*, 346: 1–16.
- 611 Rousseeuw, P. J. (1985), "Multivariate estimation with high breakdown point," in Grossmann,
 612 W., Pflug, G., Vincze, I. Wertz, W. (Eds.), *Mathematical Statistics and Applications*, Vol.
 613 B, Akadémiai Kiadó: Budapest, Hungary, pp. 283–297.
- 614 Rousseeuw, P. J. and Leroy, A. M. (2003), *Robust Regression and Outlier Detection*, Wiley: New
 615 York, NY.
- 616 Rousseeuw, P.J. and Ruts, I. (1996), "Bivariate location depth," *Applied Statistics—Journal of the
 617 Royal Statistical Society, Series C*, 45: 516–526.
- 618 Rousseeuw, P. J., Ruts, I., and Tukey, J. W. (1999), "The bagplot: A bivariate boxplot," *The
 619 American Statistician*, 53: 382–387.
- 620 Rousseeuw, P. J. and Van Driessen, K. (1999), "A fast algorithm for the minimum covariance
 621 determinant estimator," *Technometrics*, 41: 212–223.
- 622 Schölkopf, B., Smola, A., and Müller, K. B. (1998) "Nonlinear component analysis as a kernel
 623 eigenvalue problem," *Neural Computation*, 10: 1299–1319.
- 624 Shevlyakov, G., Morgenthaler, S., and Shurygin, A. M. (2008), "Redescending M-estimators,"
 625 *Journal of Statistical Planning and Inference*, 138: 2906–2917.
- 626 Singh, K. (1991), "A notion of majority depth," *unpublished manuscript*, Rutgers University.
- 627 Soriani, N. (2007), "La distribuzione t asimmetrica: Analisi discriminante e regioni di toller-
 628 anza," *Tesi di laurea, Facoltá di Scienze Statistiche, Universitá di Padova*, Available online
 629 at <http://tesi.cab.unipd.it/7115/>.
- 630 Stickler, A., Grant, A. N., Ewen, T., Ross, T. F., Vose, R. S., Comeaux, J., Bessemoulin, P.,
 631 Jylhä, K., Adam, W. K., Jeannet, P., Nagurny, A., Sterin, A. M., Allan, R., Compo, G. P.,
 632 Griesser, T., and Brönnimann, S. (2010) "The comprehensive historical upper air network

- 633 (CHUAN)," *Bulletin of the American Meteorological Society*, 91: 741–751.
- 634 Sun, Y. and Genton, M. G. (2011), "Functional boxplots," *Journal of Computational and Graph-*
635 *ical Statistics*, 20, 316-334.
- 636 Sun, Y. and Genton, M. G. (2012), "Adjusted functional boxplots for spatio-temporal data
637 visualization and outlier detection," *Environmetrics*, 23, 5464.
- 638 Tukey, J. W. (1975), "Mathematics and the picturing of data," *Proceedings of the International*
639 *Congress of Mathematics*, Vancouver, Canada, 2: 523–531.
- 640 Walters, C. K. and Winkler, J. A. (2001), "Airflow configurations of warm season southerly
641 low-level wind maxima in the Great Plains. Part I: Spatial and temporal characteristics and
642 relationship to convection," *Weather and Forecasting*, 16: 513–530.
- 643 Wartenburger, R., Brönnimann, S., and Stickler, A. (2013) "Observation errors in early historical
644 upper-air observations," *Journal of Geophysical Research*, 118: 12012–12028.
- 645 Wold, S., Esbensen, K., and Geladi, P. (1987) "Principal component analysis," *Chemometrics*
646 *and Intelligent Laboratory Systems*, 2: 37–52.
- 647 World Meteorological Organization (WMO) (2008) *Guide to Meteorological Instruments and*
648 *Methods of Observation*, 8, 7 ed., World Meteorological Organization.

Table 1: Results of simulation study to evaluate the impact of the choice of k on the robust parameter estimation for $n = 500$. When the median estimated ν exceeds 1,000, we note this with $> 1K$.

Average MSE of Parameters Excluding ν										
Ω_i	No Outliers		5% All		5% Angle		10% All		10% Angle	
	k_{min}	k_{deriv}	$k = 10$	k_{min}	k_{deriv}	$k = 10$	k_{min}	k_{deriv}	$k = 10$	k_{min}
SYM	1	0.27	0.27	0.17	0.22	0.19	0.19	0.06	0.08	0.05
	2	0.26	0.26	0.16	0.23	0.22	0.20	0.05	0.07	0.05
	3	0.27	0.27	0.19	0.22	0.21	0.20	0.07	0.10	0.06
	4	0.28	0.28	0.13	0.20	0.25	0.21	0.05	0.07	0.05
	5	0.28	0.28	0.15	0.22	0.18	0.25	0.21	0.05	0.05
OBS	1	0.06	0.06	0.06	0.07	0.06	0.08	0.07	0.06	0.06
	2	0.06	0.06	0.06	0.09	0.07	0.11	0.08	0.13	0.13
	3	0.06	0.07	0.06	0.07	0.07	0.06	0.05	0.10	0.10
	4	0.06	0.07	0.08	0.08	0.07	0.06	0.07	0.10	0.12
	5	0.05	0.06	0.06	0.07	0.06	0.06	0.09	0.11	0.12
EXT	1	0.24	0.28	0.33	0.41	0.36	0.41	0.33	0.92	0.66
	2	0.29	0.32	0.36	0.44	0.30	0.28	0.38	0.90	0.61
	3	0.28	0.32	0.36	0.33	0.25	0.23	0.43	0.42	0.44
	4	0.28	0.33	0.39	0.49	0.32	0.30	0.48	0.41	0.46
	5	0.30	0.33	0.39	0.51	0.33	0.30	0.41	0.36	0.35
Median of Estimated ν										
SYM	1	No Outliers	$k = 10$	k_{min}	k_{deriv}	$k = 10$	k_{min}	k_{deriv}	$k = 10$	k_{min}
	2	$> 1K$	$> 1K$	$> 1K$	$> 1K$	$> 1K$	4.84	742.47	$> 1K$	3.77
	3	$> 1K$	$> 1K$	$> 1K$	$> 1K$	$> 1K$	4.75	6.67	$> 1K$	3.72
	4	$> 1K$	$> 1K$	$> 1K$	$> 1K$	$> 1K$	5.46	$> 1K$	$> 1K$	3.69
	5	$> 1K$	$> 1K$	$> 1K$	$> 1K$	$> 1K$	10.15	4.40	6.30	3.66
OBS	1	11.47	13.97	17.14	4.45	5.42	5.56	3.43	4.24	4.97
	2	10.94	13.83	16.96	4.21	5.43	5.57	3.30	4.12	5.04
	3	11.16	13.50	16.50	4.53	5.57	5.75	3.55	4.42	5.10
	4	11.26	14.24	17.47	4.07	5.21	5.26	3.26	4.91	4.91
	5	10.73	13.62	15.58	4.16	5.29	5.29	3.21	4.06	4.99
EXT	1	5.10	6.04	7.04	3.03	3.67	4.20	2.50	3.02	3.75
	2	5.13	6.05	7.18	3.04	3.76	4.31	2.49	2.94	3.76
	3	5.13	6.05	7.28	3.25	3.76	4.29	2.53	3.04	3.74
	4	5.11	5.99	7.16	2.89	3.58	4.19	2.46	2.93	3.71
	5	5.12	6.09	7.17	2.85	3.53	4.13	2.47	2.94	3.75
Average k Chosen										
SYM	1	No Outliers	$k = 10$	k_{min}	k_{deriv}	$k = 10$	k_{min}	k_{deriv}	$k = 10$	k_{min}
	2	13.68	11.22	10	10.04	9.47	10	11.64	10.37	9.78
	3	13.74	11.10	10	10.13	9.46	10	11.84	10.53	9.81
	4	13.59	11.16	10	10.06	9.51	10	11.19	10.60	9.92
	5	13.67	11.16	10	10.30	9.55	10	12.22	10.60	9.79
OBS	1	16.19	11.95	10	10.99	10.07	10	12.62	10.88	10.47
	2	16.36	11.92	10	11.12	10.04	10	13.01	11.04	10.49
	3	16.04	11.85	10	10.94	10.08	10	12.25	10.66	10.53
	4	15.92	11.66	10	11.25	10.06	10	13.29	11.28	10.48
	5	15.97	11.73	10	11.27	10.04	10	13.46	11.14	10.46
SYM	1	19.60	12.38	10	12.33	10.84	10	14.64	11.85	10.59
	2	19.64	12.29	10	12.49	10.83	10	14.78	12.02	11.43
	3	19.73	12.37	10	11.84	10.76	10	14.09	11.71	10.60
	4	19.89	12.51	10	12.60	10.88	10	14.96	12.05	11.62
	5	19.47	12.32	10	12.67	10.96	10	15.09	12.00	11.56

Table 2: Simulation results for Model 1, which contains no outliers.

Model	Distribution	BVN	BDP	BST (F)				BST (SEC)			
				mle	true	trim	robust	mle	true	trim	robust
1	MVN	TP	67.4	73.0	0	0	0	0	0	0	0
		FP	0.08	0.01	2.33	2.52	3.27	2.33	2.33	2.50	3.34
	OBS	TP	0	0.1	0	0	0	0	0	0	0
		FP	1.78	0.37	2.52	2.52	4.29	5.23	2.52	2.50	4.63
	EX	TP	0	0	0	0	0	0	0	0	0
		FP	2.70	1.32	2.51	2.49	4.36	6.24	2.47	2.45	4.63

Table 3: Simulation results for $\tau = 5\%$ and $\tau = 10\%$.

τ	Model	Distribution	BVN	BDP	BST (F)				BST (SEC)				
					mle	true	trim	robust	mle	true	trim	robust	
5%	2	MVN	TP	92.8	93.7	92.9	100	98.8	99.8	93.8	100	99.2	99.9
			FP	0.01	0.01	0	2.53	0.27	1.64	0.01	2.51	0.28	1.64
		OBS	TP	93.1	88.3	96.0	99.3	97.3	99.5	89.0	99.8	98.1	99.8
			FP	0.22	0.18	0.30	2.50	0.72	2.74	0.14	2.48	0.64	2.65
	3	EX	TP	89.0	62.0	93.3	96.2	87.8	98.0	69.5	98.7	91.9	99.2
			FP	0.93	0.56	1.13	2.51	1.26	3.75	0.45	2.47	1.13	3.61
		MVN	TP	91.8	96.8	93.7	100	99.3	99.9	97.0	100	99.6	100
	4	OBS	TP	95.9	93.6	96.7	99.4	98.4	99.7	94.5	99.8	99.1	99.8
			FP	0.71	0.91	0.33	2.48	2.33	4.09	0.84	2.46	1.37	4.14
		EX	TP	93.1	75.3	94.5	95.9	92.2	98.2	82.2	98.9	96.2	99.4
10%	2	MVN	TP	97.3	98.7	93.8	100	99.6	99.9	99.0	100	99.1	99.9
			FP	0.12	0.15	0.01	2.49	1.49	3.08	0.15	2.50	0.27	1.62
		OBS	TP	97.7	97.0	96.4	99.3	99.1	99.7	97.9	99.8	98.0	99.8
			FP	0.55	0.63	0.34	2.53	2.18	4.04	0.53	2.50	0.66	2.68
	3	EX	TP	95.0	85.9	94.2	96.1	95.7	98.5	90.5	98.7	91.6	99.1
			FP	1.43	1.14	1.23	2.50	2.58	4.90	0.99	2.47	1.14	3.61
		MVN	TP	79.2	91.1	48.4	100	80.4	99.4	47.9	100	80.2	99.6
	4	OBS	TP	79.8	94.6	31.5	99.4	67.5	98.7	34.2	99.8	68.8	99.1
			FP	0.09	0.23	0.02	2.51	0.07	1.14	0.01	2.48	0.06	1.01
		EX	TP	75.1	91.3	11.8	96.2	38.0	95.3	20.0	98.7	48.7	97.3
		MVN	TP	86.0	92.4	74.5	100	91.6	99.7	74.0	99.5	91.8	99.8
		OBS	TP	87.5	96.0	64.1	99.3	85.8	99.2	65.8	99.8	86.7	99.5
			FP	0.50	0.30	0.61	2.51	1.46	3.08	0.57	2.49	1.50	3.05
		EX	TP	83.7	93.3	36.4	96.3	64.8	97.0	47.9	98.9	73.1	98.7
	4	MVN	TP	92.7	93.1	93.5	100	98.7	99.8	93.4	100	99.1	99.9
			FP	0.22	0.30	0.18	2.50	0.26	1.63	0.01	2.50	0.27	1.62
		OBS	TP	92.9	95.9	87.9	99.3	97.2	99.5	88.5	99.8	99.1	99.8
			FP	0.22	0.30	0.18	2.50	0.72	2.76	0.15	2.49	0.27	2.68
		EX	TP	88.5	93.2	60.9	96.0	87.2	98.0	68.6	98.6	91.6	99.1
		MVN	TP	0.91	1.11	0.55	2.49	1.27	3.74	0.46	2.45	1.14	3.61

Table 4: Logistic regression coefficients and significance for each method. Any p -values that are significant at the $\alpha = 0.025$ levels are italicized.

Predictor	BVN		BST-F		BST-SEC	
	Coefficient	<i>p</i> -value	Coefficient	<i>p</i> -value	Coefficient	<i>p</i> -value
$\hat{\alpha}_1$	0.65	<i>0.0000</i>	0.68	<i>0.0043</i>	0.17	<i>0.0199</i>
$\hat{\alpha}_2$	-0.12	0.4002	1.13	<i>0.0041</i>	0.01	0.9659
$\hat{\nu}$	-2.0×10^{-4}	<i>0.0015</i>	-1.4×10^{-5}	0.7998	1.5×10^{-5}	0.2216

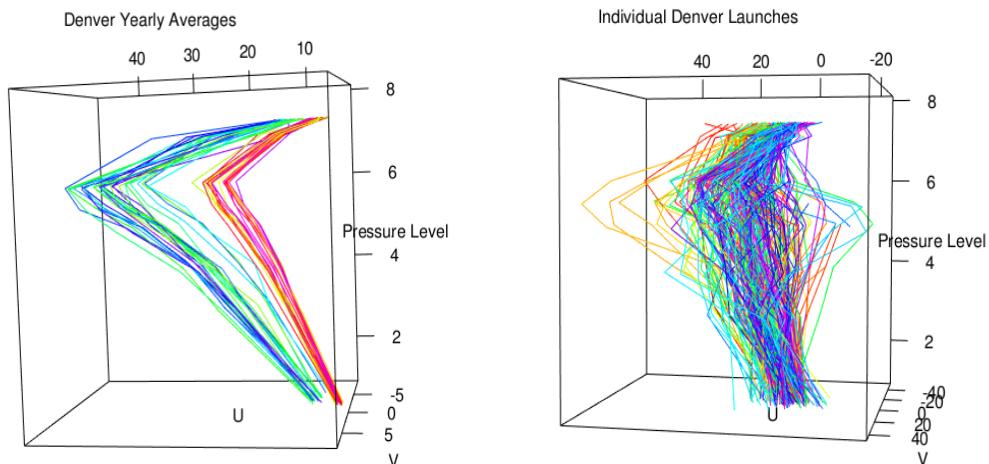


Figure 1: On the left are yearly averages from 1962 to 2011 of the wind profiles at Denver with the colors of the profiles moving from red (1962) to purple (2011) through the rainbow. On the right are 230 launches in 1962 at the Denver site.

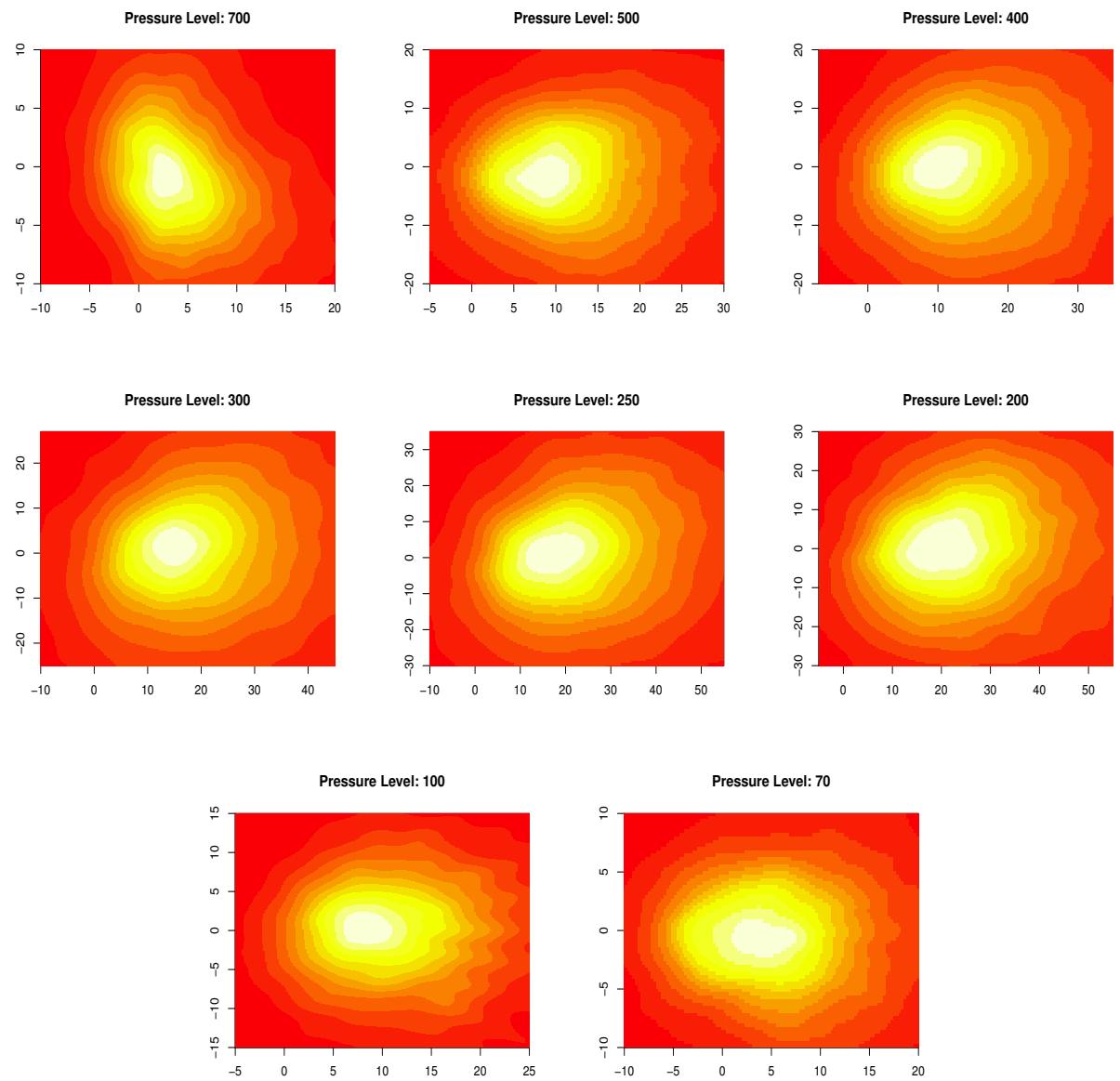


Figure 2: Bivariate nonparametric density estimates of the wind distribution by pressure level.

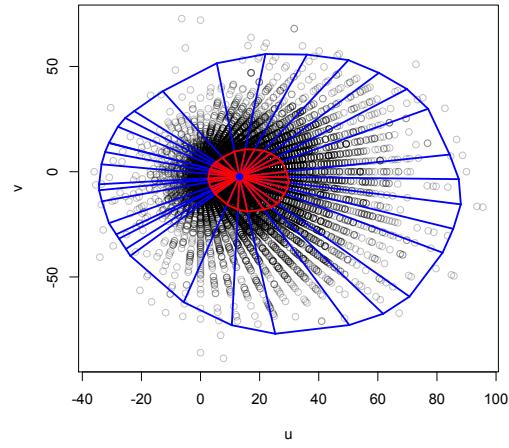


Figure 3: A set of bivariate observations with the bivariate bagplot overlaid. The central blue dot has the highest Tukey's depth and is the median. The convex hull that captures the 50% of observations with the highest Tukey's depths is in red, and the distance from the median to each vertex on the red convex hull is inflated by 3 to get the outer blue convex hull.

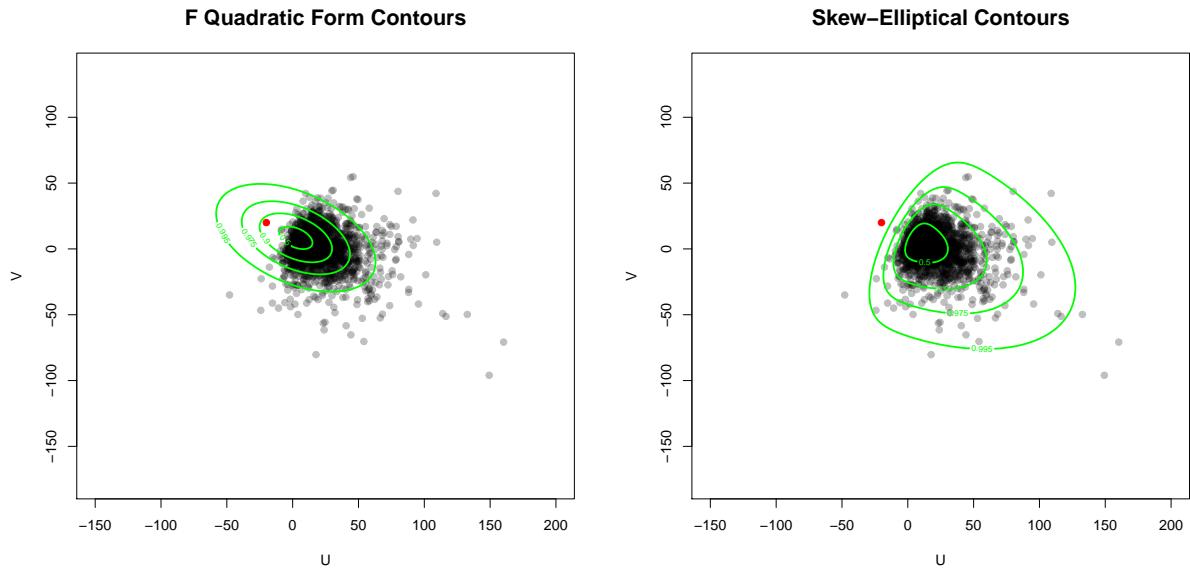


Figure 4: Differences in regions wherein outliers would be flagged with traditional distance (left) or skew-elliptical contours (right).

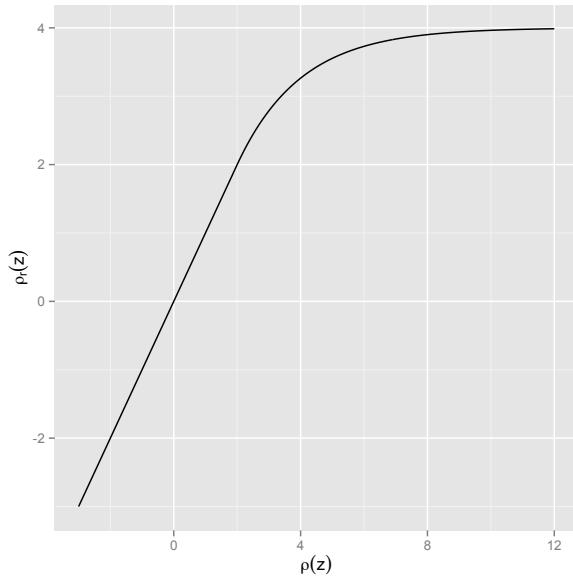


Figure 5: Trajectory of $\rho_r(z)$ in Equation (2) with $k = 2$.

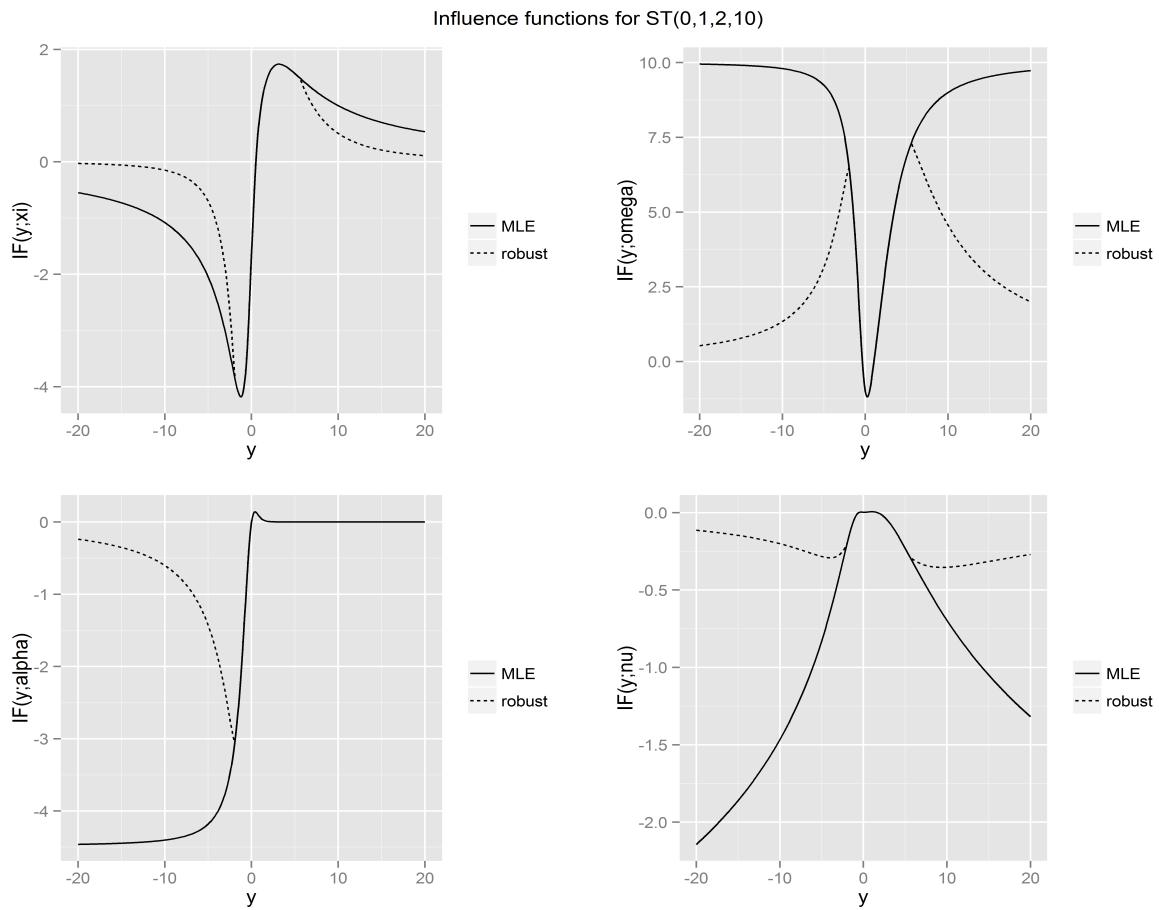


Figure 6: Influence functions for the MLEs of the parameters of the univariate skew- t distribution (solid lines) and the robust MLEs (dashed lines) with $\xi = 0$, $\omega = 1$, $\alpha = 2$, and $\nu = 10$.

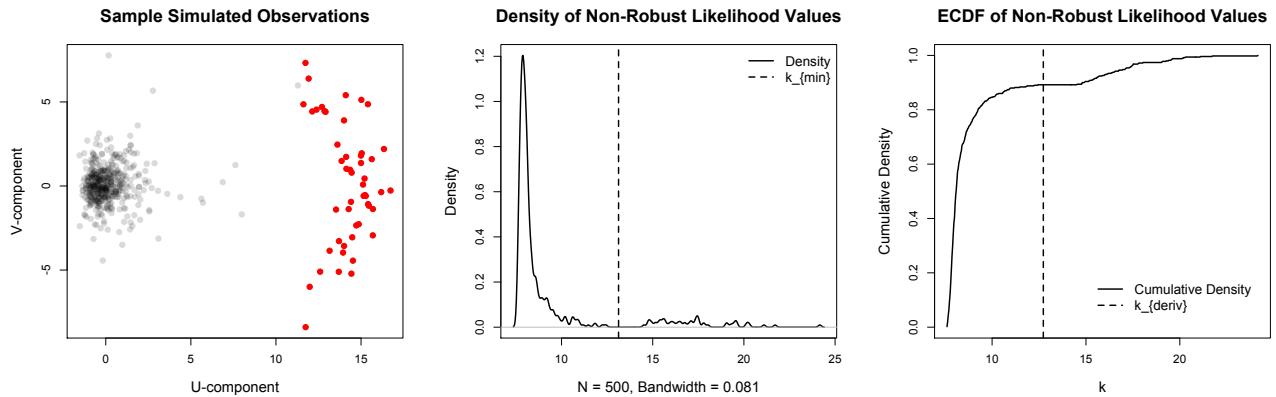


Figure 7: Illustrations of automated methods to choose k dynamically. Left: simulated dataset that has been contaminated with outliers, centered, and standardized. Center: Density of the individual observations' likelihood values. Right: Empirical cumulative density of the individual observations' likelihood values with choice of k based on the first and second derivatives of the ECDF falling within pre-specified tolerances.

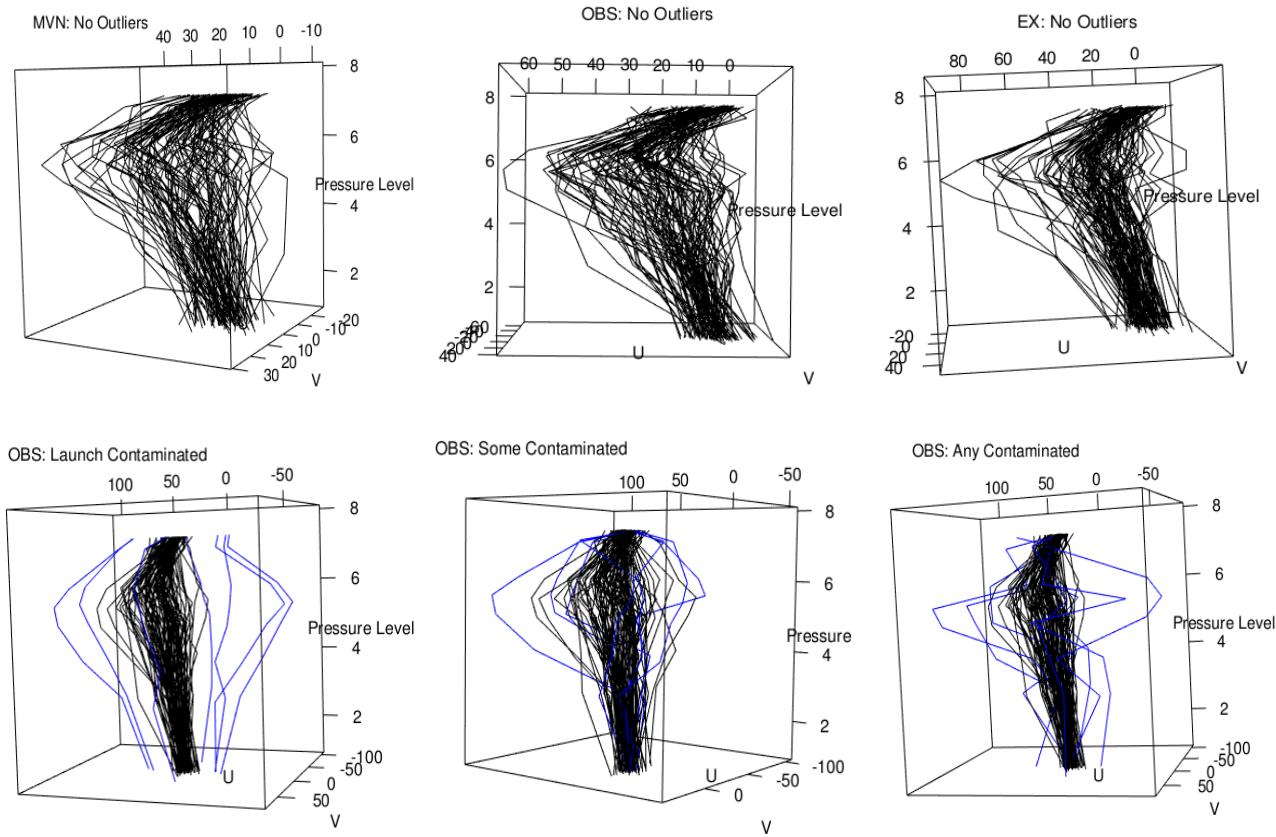


Figure 8: Top panel: one realization from model 1 for each of the 3 distribution types. Bottom panel: one realization from models 2-4 for the OBS distribution.

Figure 9: Proportion of observations flagged by pressure level and time with BVN (top); depth (second); BST F (third); and BST SEC (bottom).

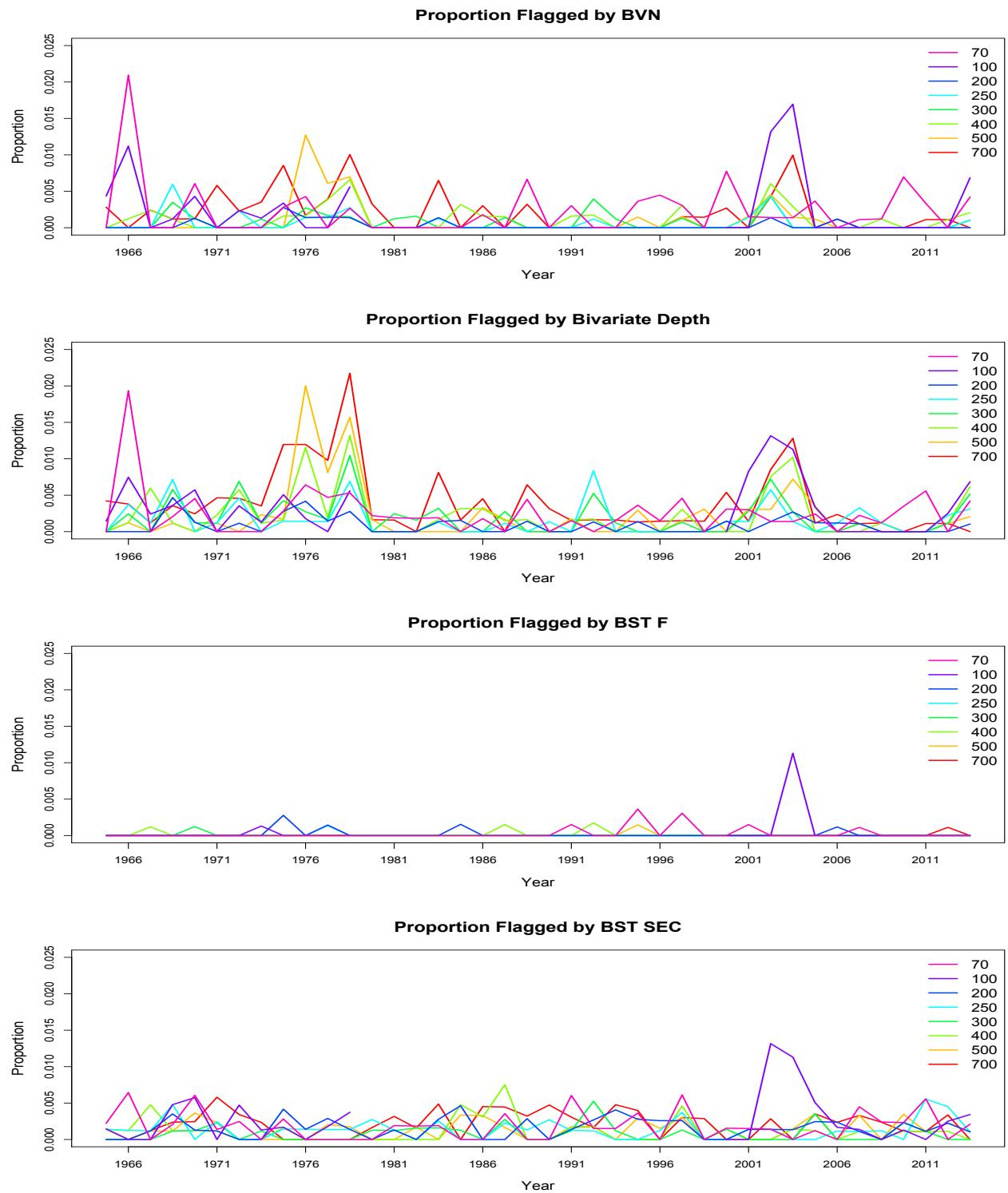


Figure 10: Select subsets of u and v observations with outliers flagged by each method shown in red.

