A Semi-Parametric Method for Robust Multivariate Error Detection in Skewed Functional Data with Application to Historical Radiosonde Winds

Ying Sun,[1] Amanda S. Hering,[2] and Douglas Nychka,[3]

May 21, 2014

Abstract

Quality control methods for multivariate data are generally based on using robust estimates of parameters for a particular distribution, and that particular distribution is usually the multivariate normal (MVN). However, many multivariate data generating processes do not produce elliptical contours, and in such cases, error detection using the MVN distribution would lead to many legitimate observations being erroneously flagged. In this work, we develop a semi-parametric method for identifying errors in skewed multivariate data that also has a functional component. In the first step, we remove potential outliers by assigning each multivariate function a depth score and remove those observations that fall beyond a given threshold. The remaining observations are used to estimate the parameters in a multivariate skew-t (MST) distribution, and this estimated distribution is used in assigning all observations a probability of having been generated from this MST. We test the performance of this two-step approach in simulation against a more common MVN method that we adapt for functional data. Finally, we show how our method can be used in practice with radiosonde launches of horizontal and vertical wind components measured at 8 vertical pressure levels.

[1]Department of Statistics, Ohio State University, Columbus, OH 43210, USA. 614.292.2866, E-mail: sunwards@stat.osu.edu

[2]Applied Mathematics and Statistics, Colorado School of Mines, Golden, CO 80401, USA. 303.273.3880, E-mail: ahering@mines.edu

[3]Computational Information Systems Laboratory, National Center for Atmospheric Research, Boulder, CO 80305, USA. 303.497.1000. E-mail: nychka@ucar.edu

1 Introduction

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

For example, the wind vector with horizontal, u, and vertical, v, components does not typically follow a multivariate normal distribution. A set of simulated observations based on the MST fit to a real dataset is given in Figure 1. Classically estimated MVN contours are overlaid on the left, and robustly estimated MVN contours are in the center plot (Rousseeuw 1985; Rousseeuw and Van Driessen 1999). The Filzmoser et al. (2005) method based on robust Mahalanobis distances flags all of the red dots in the center plot as outliers when they are not. The contours based on MST maximum likelihood estimation are given in the right-hand plot, and only one potential outlier is flagged with our initial non-parametric sweep of the observations.

We should point out the difference between extremes of a distribution and outliers, as described by Reimann et al. (2005) for univariate data. Extremes of a distribution are simply observations that are consistent with the distribution but will appear if a large enough sample is taken. Outliers, or errors, are observations that are inconsistent with the distribution and were generated from possibly multiple different processes. Our goal is not to remove extremes but to identify likely errors. In particular, we are motivated by the problem of error detection in winds recorded in historical radiosonde launches in which extreme values are still of interest, but the errors are prevalent throughout the record.

Radiosondes are instruments that are attached to weather balloons and are released twice daily at stations around the globe. Measurements of pressure, temperature, dewpoint, and winds are recorded as the balloon rises through the atmosphere at a standard set of pressure levels. Many more historical stations exist than are currently in use today, and these records can date back to the 1920's. Both systematic errors, such as differences in units of measure, changing instrumentation, and discrepancies in station locations, and random errors, such as data entry and transmission errors, can occur. We focus this work on identifying random errors in the historical archive (i.e., data collected prior to 1980) as statistical methods are less applicable for dealing with systematic errors (Parker and Cox 2007), and older records cannot be externally verified with numerical weather prediction models. Most quality assurance systems for finding random errors have focused on temperature (Durre et al. 2008; Hering et al. 2014) as temperature is an important variable in studying climate change. However, the radiosonde winds are also important as they are used in data assimilation products, such as NCEP reanalyses (Kalnay, et al. 1996; Kanamitsu et al. 2002), and these are used as boundary conditions in global and regional climate models. The radiosonde winds can also be used in studies of wind climate (Jury and Pathack 1991; Frank and Landberg 1997); severe windstorms (Klimowksi et al. 2003); and low-level jets (Walters and Winkler 2001).

The National Climatic Data Center (NCDC) houses the Integrated Global Radiosonde Archive (IGRA) (Durre et al. 2004) and performs some basic quality checks for wind such as checking that observations are within reasonable bounds; removing observations if only one of the pair is recorded; and checking for long strings of repeated values. The National Center for Atmospheric Research (NCAR) maintains another archive of radiosonde data that contains more launch stations and reaches further back in time. Both records contain many, many millions of observations and could use the procedures developed herein to either remove or down weight suspected erroneous observations.

In addition to the bivariate nature of wind, each launch of a radiosonde produces a function of wind over pressure level. Figure 2 shows the u and v components of wind plotted as a function of pressure for yearly averaged launches over 50 years from 1962 to 2011 (left) and over 200 individual launches in 1962 (right) at the Denver station. In the yearly averages, there is clearly a shift in observed winds, which may indicate a change in instrumentation or slight shift in station location, representing a systematic change. In the individual launches, there is a substantial amount of variability with an increase in spread in the mid-pressure levels. Figure 3 plots the wind densities observed at each pressure level, and the longer tails on the right-hand side of the distribution are especially evident at the middle pressure levels, which is consistent with Figure 2 (right).

In this paper, we design a method for handling both the multivariate, skewed, and functional aspects of this data. First, each launch is assigned a weighted depth score (Rousseuw and Ruts 1996), and a median functional curve along with the 50% central region are defined. Then, initial errors are flagged and removed by identifying those observations that are greater than the distance from the median to the 50% central region multiplied by the appropriate factor. After this initial sweep, the MST distribution is fit to each pressure level, and all observations are assigned a *p*-value based on the value of its fitted bivariate cumulative density function. In simulations, the effect of skewness on the method is tested as well as the effect of different assignments of depth to each observation. We compare our approach to methods based on the MVN distribution, such as Filzmoser et al. (2005), and 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.

We'll see about this next part. In this paper, we also develop a model for simulating winds as a function of pressure level that can be used to simulate the wind of artificial radiosonde launches. These launches are then contaminated with errors so that each observation is either a known error or not. We also theoretically evaluate the effect of changing skewness in the MST on the ability of both it and the MVN approaches to identify outliers. This effect is verified in simulation.

This paper is organized as follows: in Section 2, we briefly explain the Filzmoser et al. (2005) method and describe how we extend it to handle functional data. We also introduce our new method. In Section 3, we describe and report the results of our simulation study, and Section 4 displays the results of applying both methods to the radiosonde winds at one launch station. Section 5 concludes and outlines additional considerations to be taken into account when applying this approach to the entire archive.

2 Simulation

2.1 Multiplicative Factor for Depth Method

For this simulation, we wanted to determine what factor to multiply the central 50% convex curve by in the nonparametric weighted functional depth approach such that no more than 0.3% of true observations are flagged. To do this, three types of functional data are simulated. All three have the same center and spread, but the skewness and degrees of freedom used in to simulate the functional launches from the multivariate skew-t distribution differ, as follows:

(1) MVN: The skewness is set to zero for all 8 pressure levels, and the degrees of freedom is set to infinity.

(2) OBS: The skewness for each pressure level is set to what is estimated for the Denver station launches, and the degrees of freedom is set to 10.

$$\alpha_D = (2.16, 1.44, 1.35, 0.95, 0.60, 0.31, 2.62, 3.02, -0.03, -0.81, -0.82, -0.89, -0.82, -0.75, -0.31, 0.39)$$

(3) EX: The skewness for each pressure level is chosen to be more skewed than the Denver station launches, and the degrees of freedom is set to 5.

$$\alpha_{ex} = (3, 4, 5, 6, 7, 8, 9, 10, 3, 4, 5, 6, 7, 8, 9, 10)$$

One thousand datasets of each of the three types above is simulated. Then, the weighted functional depth is computed for each set of simulated launches. The factor by which to multiply the 50% central convex hull is chosen such that no more than 0.3% of true observations are flagged as outliers, and this factor is stored across all simulated datasets. This process is repeated for the following number of launches: $(500, 1000, 1500, 2000, 2500, 3000)$.

Figure 4 shows the results of this simulation study with boxplots of the chosen factor displayed for each number of launches and data type. Overlaid on each boxplot with a red diamond is the value of the factor averaged across the 1,000 simulated datasets. In Figure 5, an exponential decay function is overlaid such that for each type of data and any number of launches, the appropriate factor can be selected. Letting F be the multiplier and n be the number of launches, the general form of the equation that was fit is

$$F = \min(\bar{\mathbf{F}}) + (\max(\bar{\mathbf{F}}) - \min(\bar{\mathbf{F}})) \cdot \exp\{\beta(n - 500)\},$$

where $\bar{\mathbf{F}}$ is the vector of mean factors for each number of launches. The fitted functions are as follows:

(1) MVN: $F = 1 + 0.472445 \cdot \exp\{-0.0037578(n - 500)\}$

(2) OBS: $F = 1.471 + 0.560031 \cdot \exp\{-0.0018624(n - 500)\}$

(3) EX: $F = 2.0525 + 0.939452 \cdot \exp\{-0.0011780(n - 500)\}$

To conclude, if we desire to choose one factor that will be applied universally, I would choose 3 if a small number of launches is used and 2 if a larger number of launches is taken, based on the right-hand panel of Figure 4 so that we can account for any potential skewness in the underlying distribution. However, in practice, we could allow this multiplier to be chosen adaptively for any number of launches and estimated distribution parameters.

2.2 Adjusting Thresholds

In this subsection, I will describe the simulation study that was performed to find the adjusted thresholds by which to classify a squared distance as either an “outlier” or a “true observation.”

1. Simulate a 16-dimensional MVST launch based on the parameters that we use in the “big simulation.”

(a) Estimate the MVST parameters in the simulated launch for a given pressure level, i.e. in 2 dimensions.**

**Note, in practice, a sweep of the observations with the depth method should be done first to remove unusual observations.

(b) Given these estimated parameters, we compute the distances:

$D_i = ((\mathbf{x}_i - \hat{\xi}_i)'\hat{\Omega}^{-1}(\mathbf{x}_i - \hat{\xi}_i))^{1/2}$. These squared distances follow a scaled F distribution, as follows: $D_i^2 \sim 2 \cdot F(2, \hat{\nu})$.

(c) Get original threshold—based on scaled F distribution, and $\alpha = 0.02$.

This is $\delta = 2 \cdot F_{1-\alpha}(2, \hat{\nu})$.

(d) Get adjusted threshold—based on simulated MVST in 2-D.

i. Compute maximum positive deviation between $G(u)$ and $G_n(u)$ in the upper tail, where $G_n(u)$ is the empirical cdf of the squared distances, and $G(u)$ is the cdf of the scaled F distribution, as follows:

$p_n(\delta) = \sup_{u \geq \delta}(G(u) - G_n(u))^+$.

ii. Simulate 1,000 MVST realizations of 500 2-D observations based on estimated MVST parameters in (a).

A. Find the empirical cdf of the squared distances in each simulated dataset, $G_i(u)$.

B. Save the maximum positive deviation between $G_i(u)$ and $G(u)$ in the upper tail (larger than δ).

C. Save the 95% quantile of these maximum positive deviations, denoted $p_{crit}(\delta, n, p, \nu)$.

iii. Then, α is adjusted as follows:

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

iv. The threshold value is then $c_n(\delta) = G_n^{-1}(1 - \alpha_n(\delta))$.

(e) Repeat for each pressure level.

2. Repeat step 1 1,000 times. For each instance, save the 1,000 original and adjusted thresholds at each pressure level.

3. Take average of thresholds across all 1,000 16-D simulated datasets.

4. Repeat steps 1 through 3 for each type of simulated dataset (MVN, OBS, and EX).

3 Skew-T Marginals

Let Y be a multivariate d-dimensional skew-t random variable, so $Y \sim St_d(\xi, \Omega, \alpha, \nu)$. Then, for a single component, Y_r, for $r \in \{1, \ldots, d\}$, we have that

$$Y_r \sim St(\xi_r, \omega_{rr}, \alpha'_r, \nu),$$

where α'_r is given by the following:

$$\alpha'_r = \frac{\alpha_r + \bar{\Omega}_{rr}^{-1}\bar{\Omega}_{rs}\alpha_s}{(1 + \alpha_s^T \bar{\Omega}_{ss\cdot r}\alpha_s)^{1/2}}.$$

This assumes that Y and its parameters can be partitioned as follows:

$$Y = \begin{pmatrix} Y_r \\ Y_s \end{pmatrix}, \quad \xi = \begin{pmatrix} \xi_r \\ \xi_s \end{pmatrix}, \quad \alpha = \begin{pmatrix} \alpha_r \\ \alpha_s \end{pmatrix}, \quad \text{and}$$

$$\Omega = \begin{pmatrix} \Omega_{rr} & \Omega_{rs} \\ \Omega_{sr} & \Omega_{ss} \end{pmatrix},$$

and for $\omega = \text{diag}(\Omega_{11}, \Omega_{22}, \ldots, \Omega_{kk})^{1/2}$ and $\bar{\Omega} = \omega^{-1}\Omega\omega^{-1}$,

$$\bar{\Omega}_{rr}^{-1} = (\bar{\Omega}_{rr})^{-1}$$

and

$$\bar{\Omega}_{ss\cdot r} = \bar{\Omega}_{ss} - \bar{\Omega}_{sr}\bar{\Omega}_{rr}^{-1}\bar{\Omega}_{rs}.$$

In particular, if we have a bivariate skew-t, then Y_r is 1×1 and Y_s is also 1×1. I haven't found it anywhere specific, but it looks like this formula could also apply to finding the bivariate marginal distribution from a 16-dimensional skew-t as well.

4 Methods

A list of methods and short summaries of each one:

- Filzmoser et al. (2005): Good paper. Makes a clear distinction between identifying extreme values of a particular distribution and outliers that are generated from a secondary process. Assumes the primary underlying process is a multivariate normal distribution, and uses Mahalanobis distance to measure the distance of observations from the center of the distribution, as follows:

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

  but in place of the classical estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, uses 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).

  The observations are assumed to be multivariate normal since then the squared Mahalanobis distances follow a chi-square distribution. Thus, the deviation between the chi-square distribution and the MD_i values in the upper tail of the distribution is used in computing the threshold for outliers, but first, they choose a critical value based on distinguishing between outliers and extremes, as derived through simulation under an assumption of no outliers. They choose critical values based on various sample sizes and data dimension (n and p), so the critical value can be adapted to the problem at hand. This is important since the amount of scatter inherent in a multivariate sample of size 100×10 is much greater than that observed in a sample of size $10,000 \times 2$.

  The authors go on to present a visualization technique for the identified outliers. Not only do they use different plotting characters for observations within different quantiles of the multivariate distribution, but they also color the plotting characters based on the kind of outlier, i.e., red for unusually large values and blue for unusually small values. They plot the observations spatially (since they have spatial data), and they are able to identify spatial patterns in the outliers.

  The authors have an `R` package called `mvoutlier` that will perform the methods that they propose. The data that they use for illustration is included in the package.

  Points of interest:

  - Generally speaking, wind data will not be multivariate normal. Can we instead use a multivariate skew-t distribution to perform a similar routine? If so, the Mahalanobis distances will no longer have a chi-square distribution. So, we will need to know what distribution to compare the Mahalanobis distances to.

    - As an exploratory step, we can plot the identified outliers as a time series plot to see if outliers tend to cluster together.

- Peña and Prieto (2001)
- Rousseeuw and Zomeren (1990)

5 Simulation Study

6 Illustration

7 Conclusions

We have applied our method to one launch station in Denver, but this will be extended and tested with the entire archive. This work is meant to be the testing ground for this approach, and future work will handle how the method should be applied over the course of very long records in which the effect of changing climate may become important.

Notes:

Meeting Notes with Doug 9/5/13

Doug suggests that we make two problems out of this

1. First, develop the method in a test/application sense. Make sure it makes sense to Joey and that we haven't overlooked anything.
2. Apply the method to a big part of the archive. This will be a big data problem, and we may run into some unexpected challenges.

He thinks that the first we should target initially to *JASA-Case Studies* and then the second to an atmospheric journal or something like *Environmentrics*.

For literature, Doug thinks we should try to see what NCDC does. There may be a set of literature out there on how they QC the other radiosonde variables. If we can find that, then we have a documented QC procedure that we can compare with. I haven't found it yet, but I haven't looked specifically into the NCDC documentation.

I probed Doug to get his thoughts on methods we should explore. He said that he really doesn't know, but he threw out the following:

- Do a "buddy check." He says to exploit the structure of the data. Check nearby stations (which I'm not sure will be helpful since wind is so variable, and the stations don't tend to be very close), or use a functional data approach using pressure levels.
- Can we say anything analytically, in other words can we demonstrate properties of our QC method without having to do massive simulations? Can we write how things will vary in terms of a formula?

- Avoid finding a particular threshold for flagging outliers, but instead give each observation a probability of being an outlier.
- If we are going to fit a skewed distribution, use multiple stations, determine if there is a common distribution among them, adjusting for latitude and longitude. Think of strategies for combining information from across stations to fit (for example), a multivariate skew-t distribution.

Meeting Notes with Doug 3/21/14

Another method that we could use would be to do a nonparametric bivariate density estimate. Then, we can estimate the density on a grid of points. Starting with the maximum of the density, we can begin to make horizontal slices through the density and figure out how many points are still captured, thereby making 50% median contours down to 1% contours. One advantage to this over the nonparametric functional approach is that we could capture contours that are not convex.

Questions/Issues that Doug asked:

1. Does the multivariate skew-t distribution have a closed form for its contours? Even if not, we could also report the contour that a particular observation lies on instead of computing the area under the curve of the bivariate cdf and getting a p-value as Ying has suggested in her document.
2. He thinks that we should try to do something analytical. For example, try to calculate the mismatch between the number of outliers that the MVN identifies versus the skew-t. Perhaps it is a simple function of the skewness parameter. We could do this under the assumption that the parameters are known, then do a Monte Carlo experiment to see how this performs when we estimate the parameters. Can we then develop some kind of rule of thumb? For example, if skewness is beyond 3 in absolute value, you would expect to mis-identify 15% of outliers if the MVN distribution is used.
3. Doug is a big proponent of resampling launches to use in the simulation study. He thinks that if we try to build a model from which to simulate launches, that the question may become ... if the simulation model is so good, then why didn't we just use the simulation model in our outlier detection scheme? In addition, he thinks it will be hard to model the bivariate relationships in u and v vertically. He also thinks that our first approach should be to look at each of the pressure levels separately and save the vertical component for later. He doesn't know how, for example, one might use a set of basis functions in 2 dimensions to model the (u, v) in the vertical.

One of my thoughts is the following: I looked through the NCDC website to see how they do their initial quality control for their radiosonde archive. They barely mention wind. When, they do, only 3 things are mentioned:

1. Ranges on the variables are checked: wind speed must be in $[0, 150]$ m/s, and wind direction must be in the interval $[0°, 360°]$.
2. If one of wind speed or direction is missing, then both are thrown out.
3. If there is a long string of repeated values in either speed or direction, then those values were also removed.

I still need to look at some of the papers that they reference on their website to make sure that I'm not missing anything with respect to wind, but it appears that they have focused most of their quality control efforts on pressure, geopotential height, and temperature. I propose instead to show an example of clearly skewed data and show how our method identifies outliers as opposed to the multivariate normal based method in Filmoser et al. (2005). Their method is available in an `R` package called `mvoutlier`, so it should be pretty easy to implement for comparison.

References

Azzalini, A. and Capitanio A. (2014) *The Skew-Normal and Related Families*, Cambridge University Press: New York, NY.

Durre, I., Vose, R. S., and Wuertz, D. B. (2004) "Overview of the integrated global radiosonde archive," *Journal of Climate*, 19: 53–68.

Durre, I., Vose, R. S., and Wuertz, D. B. (2008) "Robust automated quality assurance of radiosonde temperatures," *Journal of Applied Meteorology and Climatology*, 47: 2081–2095.

Filzmoser, P., Reimann C., and Garrett R. G. (2005) "Multivariate outlier detection in exploration geochemistry," *Computers and Geosciences*, 31: 579–587.

Filzmoser, P., Maronna, R., and Werner, M. (2008) "Outlier identification in high dimensions," *Computational Statistics and Data Analysis*, 52: 1694–1711.

Frank, H. P. and Landberg, L. (1997) "Modelling the wind climate of Ireland," *Boundary Layer Meteorology*, 85: 359–377.

Hering, A. S., Bell, A. N., Comeaux, J., and Nychka, D. (2014) "A simulation study to compare statistical quality control methods for error detection in historical radiosonde temperature," *In Progress.*

Jury, M. R. and Pathack, B. (1991) "A study of climate and weather variability over the Tropical Southwest Indian Ocean," *Meteorology and Atmospheric Physics*, 47: 37–48.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D. (1996) "The NCEP/NCAR 40-year reanalysis project," *Bulletin of the American Meteorological Society*, 437–471.

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S., Hnilo, J. J., Fiorino, M., and Potter, G. L. (2002) "NCEP-DOE AMIP-II Reanalysis (R-2)," *Bulletin of the American Meteorological Society*, 1631–1643.

Klimowksi, B. A., Bunkers, M. J., Hjelmfelt, M. R., and Covert, J. N. (2003) "Severe convective windstorms over the Northern High Plains of the United States," *Weather and Forecasting*, 18: 502–519.

Maronna, R., Martin, R. D., and Yohai, V. J. (2006) *Robust Statistics: Theory and Methods*, Wiley: New York, NY.

Parker, D. E. and Cox, D. I. (2007) "Towards a consistent global climatological rawinsonde data-base," *International Journal of Climatology*, 15: 473–496.

Peña, D. and Prieto, F. J. (2001) "Multivariate outlier detection and robust covariance matrix estimation," *Technometrics*, 43: 286–300.

Reimann, C., Filzmoser, P., and Garrett, R. G. (2005) "Background and threshold: Critical comparison of methods of determination," *Science of the Total Environment*, 346: 1–16.

Rousseeuw, P. J. (1985) "Multivariate estimation with high breakdown point," in Grossmann, W., Pflug, G., Vincze, I. Wertz, W. (Eds.), *Mathematical Statistics and Applications, Vol. B*, Akadémiai Kiadó: Budapest, Hungary, pp. 283–297.

Rousseeuw, P. J. and Leroy, A. M. (2003) *Robust Regression and Outlier Detection*, Wiley: New York, NY.

Rousseuw, P.J. and Ruts, I. (1996) “Bivariate location depth,” *Applied Statistics–Journal of the Royal Statistical Society, Series C*, 45: 516–526.

Rousseeuw, P. J. and Van Driessen, K. (1999) “A fast algorithm for the minimum covariance determinant estimator,” *Technometrics*, 41: 212–223.

Walters, C. K. and Winkler, J. A. (2001) “Airflow configurations of warm season southerly low-level wind maxima in the Great Plains. Part I: Spatial and temporal characteristics and relationship to convection,” *Weather and Forecasting*, 16: 513–530.

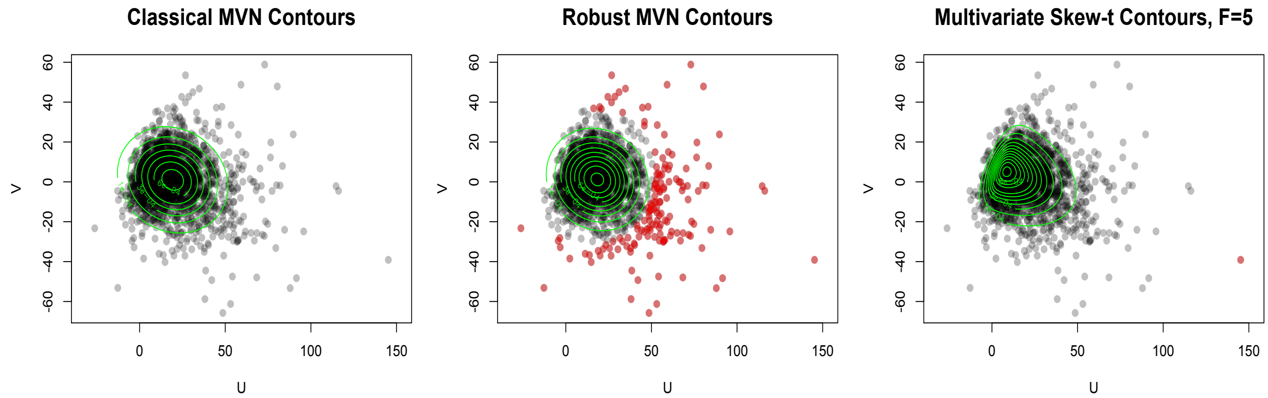

Figure 1: Scatterplots of u and v components simulated from a MST distribution with classic MVN (left), robust MVN (center), and MST (right) contours overlaid. For the robust MVN, the outliers that are flagged by the Filzmoser et al. (2005) method are in red. For the MST, the nonparametric multiplier is 5.

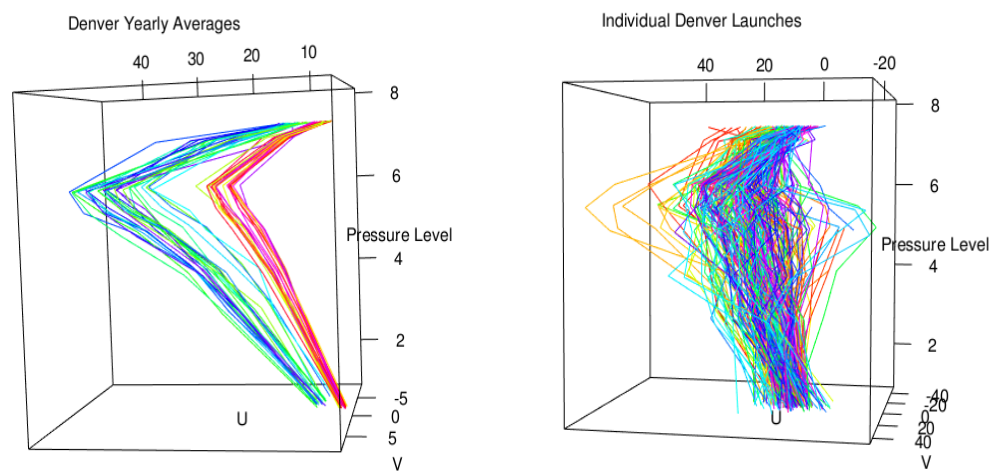

Figure 2: On the left are yearly averages from 1962 to 2011 of the wind profiles at Denver. On the right are 230 launches in 1962 at the Denver site.

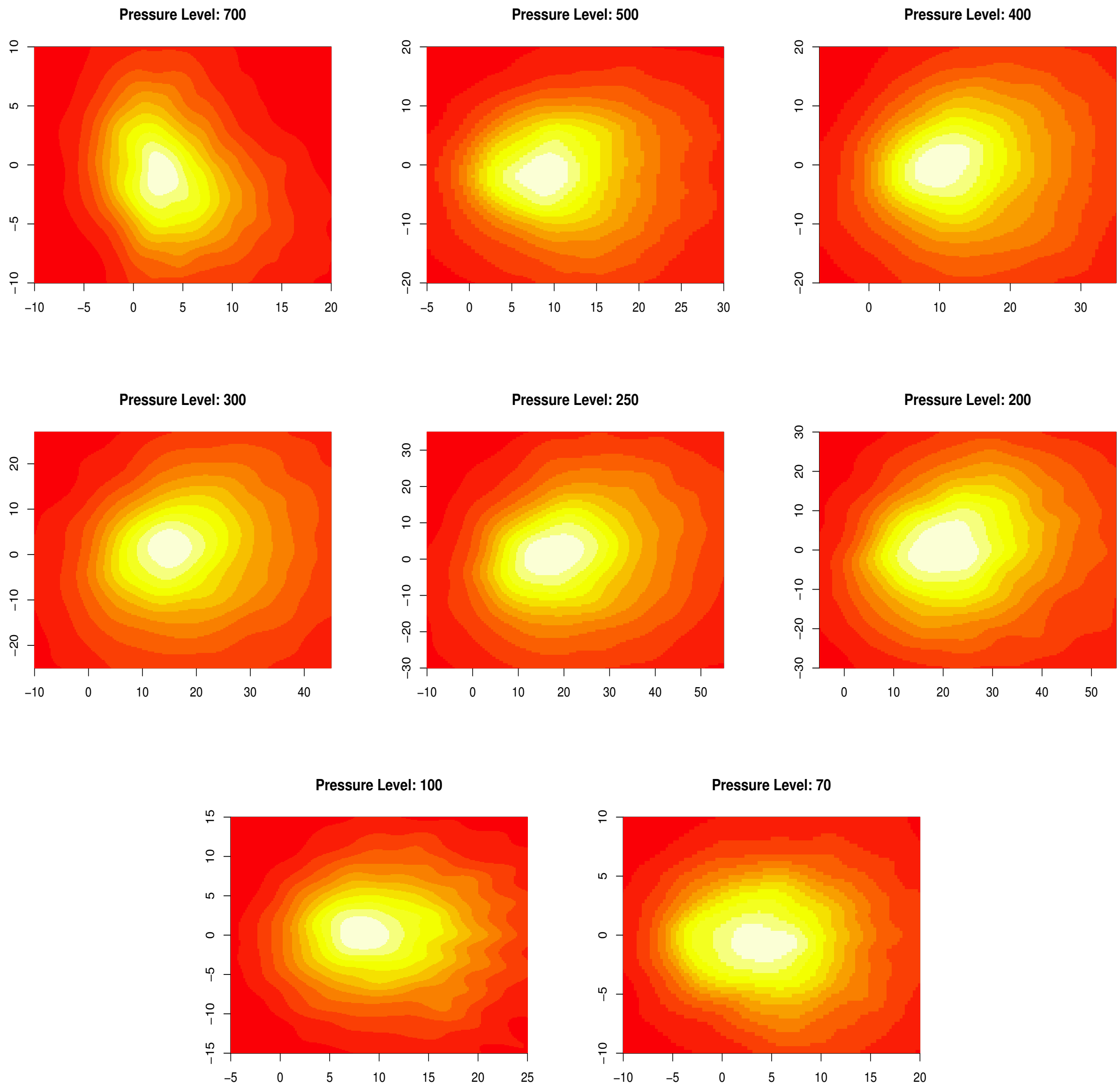

Figure 3: Bivariate density estimates of the wind distribution by pressure level.

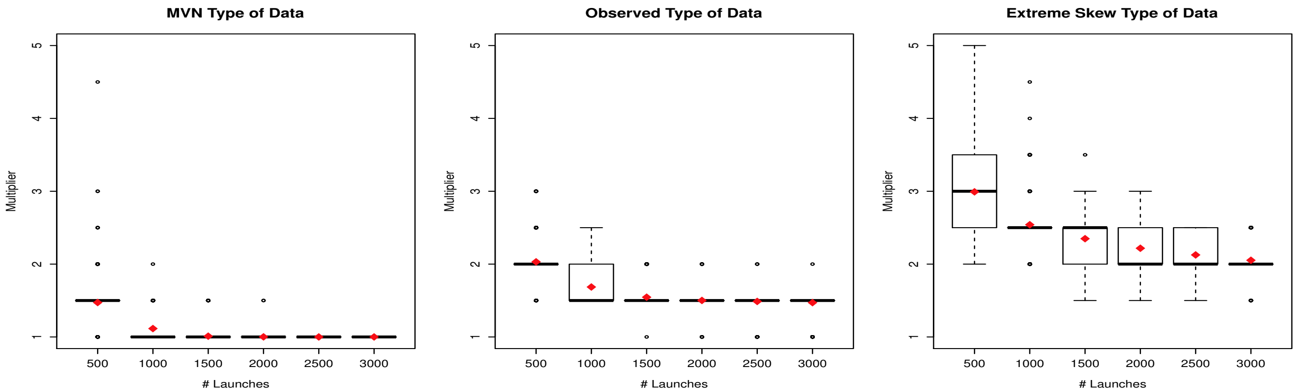

Figure 4: Boxplots of the multiplier selected for each sample size across multivariate normal data (left), data similar to what is observed among the Denver launches (center), and data whose skewness is more extreme than the Denver data (right). Overlaid on each boxplot in red is the average multiplier for each sample size.

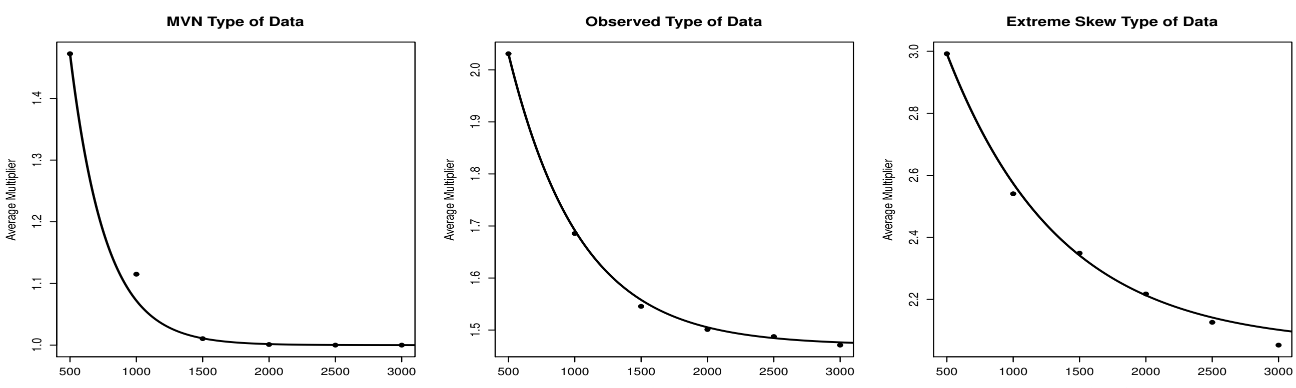

Figure 5: Fitted exponential decay functions to the factor multiplier as a function of the number of launches for each type of simulated data.