Is this normal? A new projection pursuit index to assess a sample against a multivariate null distribution

Annalisa Calvi[1], Ursula Laa[2], Dianne Cook[3]

ARTICLE HISTORY
Compiled February 1, 2025

[1] School of Mathematics, Monash University,
[2] Institute of Statistics, University of Natural Resources and Life Sciences, Vienna,
[3] Department of Econometrics and Business Statistics, Monash University,

ABSTRACT
Many data problems contain some reference or normal conditions, upon which to compare newly collected data. This scenario occurs in data collected as part of clinical trials to detect adverse events, or for measuring climate change against historical norms. The data is typically multivariate, and often the normal ranges are specified by a multivariate normal distribution. The work presented in this paper develops methods to compare the new sample against the reference distribution with high-dimensional visualisation. It uses a projection pursuit guided tour to produce a sequence of low-dimensional projections steered towards those where the new sample is most different from the reference. A new projection pursuit index is defined for this purpose. The tour visualisation also includes drawing of the projected ellipse, which is computed analytically, corresponding to the reference distribution. The methods are implemented in the R package, `tourr`.

KEYWORDS
exploratory data analysis; statistical graphics; dimension reduction; high-dimensional visualisation; climate change; clinical trials

1. Introduction

Linear projections are useful in many aspects of statistical analysis of multivariate data, and especially useful for visualising the data. A linear projection provides a dimension reduction while maintaining interpretability. For example, a biplot (Gabriel 1971; Gower and Hand 1996) shows the structure creating the maximum variance in the data, and

CONTACT: Annalisa Calvi. Email: annalisa.calvi@monash.edu. Ursula Laa. Email: ursula.laa@boku.ac.at. Dianne Cook. Email: dicook@monash.edu.

also visualizing the projection matrix to learn which variables contribute to it. We might find clusters of outliers that were hiding in high dimensions.

More generally, projection pursuit (Friedman and Tukey 1974; Jones and Sibson 1987; Huber 1985) defines a quantitative criterion for the interestingness of a projection (a projection pursuit index), and searches the space of possible projections for the most interesting one to display. We can also define sequences of interpolated linear projections to better understand a multivariate distribution. Animating a randomly selected interpolated sequence of linear projections is called a grand tour (Asimov 1985; Buja and Asimov 1986; Buja et al. 2005; Cook et al. 2006; Lee et al. 2022). The combination of these two approaches would then use a projection pursuit index to select interesting projections, but display them via an interpolated path to provide context. This is called a guided tour (Cook et al. 1995).

The question is whether we can use these techniques to assess new data samples in the context of an established normal, such as a specific multivariate normal distribution. In physics, the normal distribution may describe experimental results, or a global fit for a selected model, and we might want to compare to sets of other models. In medical applications, the normal distribution might summarize historic data of a healthy population and we compare it to samples from new patients. In outlier detection we might use robust measures to define the normal distribution and look for anomalies.

This paper describes a new projection pursuit index to find projections where a new sample is most distant from the existing normal distribution, that is integrated with the guided tour to visualise the ways that the data departs from the reference distribution. This work also expands on current outlier detection methods such as Kandanaarachchi and Hyndman (2021) and Loperfido (2018) by providing integrated multivariate visual-

isation to examine the outlyingness in different projections. It also provides the ability to examine potential outliers interactively - when they are outlying in different ways it can interfere with detection numerically, but visually one might group different types of outlyingness, and operate the outlier detection on subsets.

The paper is organised as follows. Section 2 provides more context for the methods and visualisation. Section 3 provides the details of the new index, and rendering projections of the reference ellipse. Section 4 describes the implementation. Two examples, one medical and one on climate extremes, illustrate usage in Section 5.

2. Background

Following the ideas expressed in Cook and Swayne (2007), to compare a new sample with an existing norm, like a multivariate normal distribution, in higher than two dimensions, you would use two samples of points. The norm is represented by points on the surface of a p-dimensional ellipsoid, with size corresponding to a chosen level curve of the density function of the reference distribution. This can be generated using these steps:

1. Simulate a sample of observations ($\boldsymbol{x}$, which are p-D vectors) from $N_p(\boldsymbol{\mu}, \Sigma)$.
2. Transform each observation to have unit distance from the mean, $\frac{\boldsymbol{x}^\top}{||\boldsymbol{x}^\top||}$. The sample is now uniform on the surface of the hypersphere.
3. Transform the shape from a sphere to an ellipsoid using the specific variance-covariance.
4. Shift the resulting sample to center it on the mean vector, and resize to correspond to the desired density level. This could also be treated like a confidence ellipse, where observations outside are more than, say, three standard deviations from

the sample mean.

New observations can be visually compared with this ellipsoid by plotting them together using a tour, making sure that any transformations used on the reference distribution are also applied during pre-processing. Figure 1 illustrates this process for 2-D.

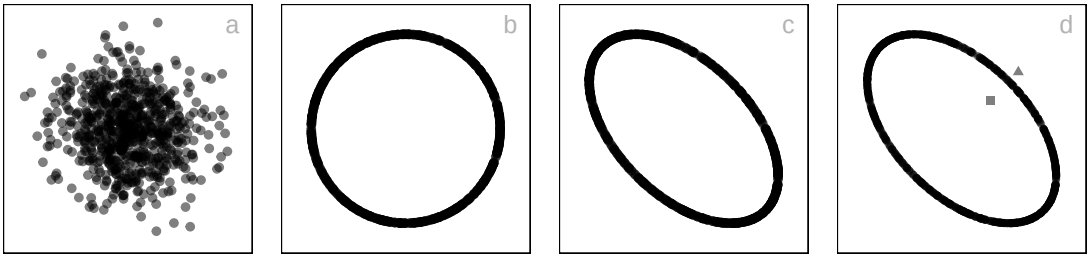
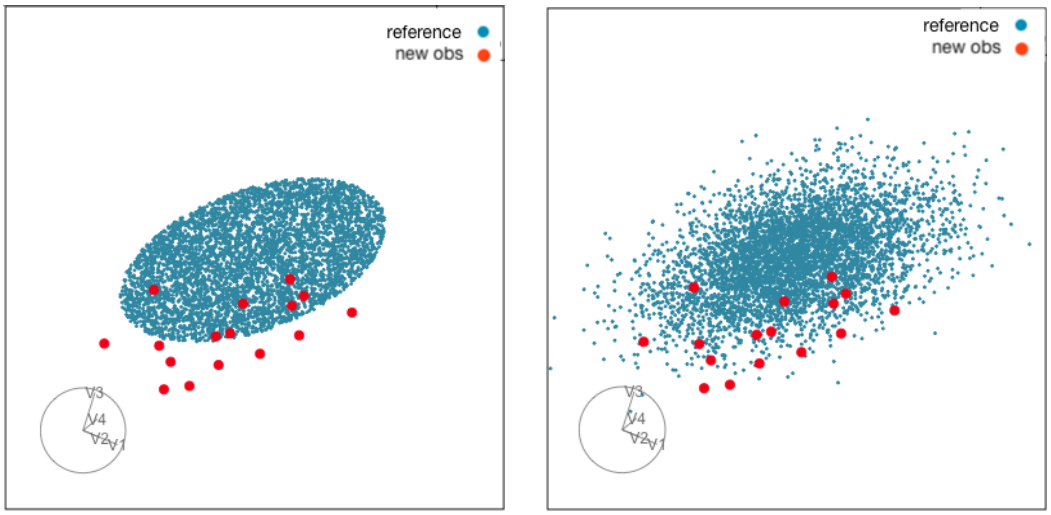


Figure 1. Simulating a uniform sample on a sphere involves sampling from a multivariate normal (a) and transforming each observation to have length equal to 1 (b). A 95% confidence ellipsoid is generated by transforming the sphere relative to a specified variance-covariance matrix, and sizing using a critical value (c), and new observations (triangle, square) can be visually assessed to be inside or outside by plotting with the ellipsoid (d).



(a) Reference ellipse (b) Reference sample

Figure 2. Illustration of existing procedures for a 4-D problem: compare new sample with (a) points on the surface of a p-D ellipse representing the reference distribution, (b) a sample of observations simulated from the reference distribution. A single 2-D projection from a tour. (The circle and projected axes represent projection coefficients, and animated gifs of these tours are available in the supplementary material.) Both approaches are useful, but need the refinements described in this paper.

This approach for comparing new observations with a reference distribution can be

applied to any p-D problem. Figure 2 illustrates the procedures for a 4-D problem. The new sample of observations is compared against the reference distribution in two ways (a) against a 4-D ellipse corresponding to a level curve of density function of the reference distribution and (b) against a sample simulated from the reference distribution. Here we show a single 2-D projection from a tour of the full p-D. The circle and line segments indicate the coefficients of the 4×2 basis used to project the data in these plots. The difference between the new sample of points (dark/red) from the sample is in the bottom left to top right direction, corresponding mostly to the V3 axis direction. Roughly orthogonal to this is a combination of V1 and V2, so the difference between new and reference as indicated by this projection is due to V4 contrasted with V1 and V2. The ellipse makes it clearer see but the sample can be useful, too, particularly if the reference distribution is not a multivariate normal distribution. (The supplementary material includes animated gifs showing these data in a tour, allowing one to see many different projections of the 4-D points.)

While this is a flexible and useful approach, there are two problems: (1) the ragged edges of the points marking the projected ellipse make it difficult to compare the new sample precisely against acceptable ranges, and (2) there is no projection pursuit index that can guide the tour towards the directions (projections) where the samples are most different from the reference.

What would be better, and what is developed in this paper, is to analytically define the confidence ellipsoid and display the projected ellipsoid as a geometric shape rather than a sample of points. Further, a new projection pursuit index that is based on flagging observations that are outside, and would steer the tour to projections that reveal the extent of the difference. These new procedures are described in the next section.

3. Anomaly index

3.1. *Projecting an ellipsoid*

Let $\boldsymbol{x}$ be a p-D vector. A p-D ellipsoid corresponding to a given variance-covariance (Σ) and mean vector ($\boldsymbol{\mu}$) is described by the equation

$$(\boldsymbol{x} - \boldsymbol{\mu})\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})^T = c^2 \tag{1}$$

where c is a constant that depends on a specific confidence level.

Let P be a $(p \times 2)$ matrix, whose columns form an orthonormal basis for the 2-D projection space. Let $\boldsymbol{y}$ be a 2-D vector, and let $\boldsymbol{\mu}_p := \boldsymbol{\mu}P$ denote the projected mean.

Theorem. *The projection of the p-D ellipsoid decribed in Equation 1 onto the 2-D projection space described by P has the equation*

$$(\boldsymbol{y} - \boldsymbol{\mu}_p)(P^T\Sigma P)^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_p)^T = c^2. \tag{2}$$

Proof. The projection of a hyperellipsoid onto 2-D space is an ellipse. The boundary of this 2-D ellipse is the orthogonal projection of select points on the boundary of the hyperellipsoid. For $\boldsymbol{x}$ on the hyperellipsoid, its orthogonal projection $\boldsymbol{x}P$ is on the boundary of the projected ellipse if and only if the tangent plane to the hyperellipsoid at $\boldsymbol{x}$ is perpendicular to the projection plane described by P. As the hyperellipsoid is a level curve, described by Equation (1), this means the gradient of the left hand side of Equation (1) at $\boldsymbol{x}$ is parallel to the projection plane. Note that any vector parallel to

the projection plane can be expressed as $2\boldsymbol{s}P^T$ for some $\boldsymbol{s} \in \mathbb{R}^2$.

Therefore, the boundary of the ellipse consists of the points $\boldsymbol{x}P$ for which $\boldsymbol{x}$ satisfies

$$2\boldsymbol{s}P^T = \nabla \left[(\boldsymbol{x} - \boldsymbol{\mu})\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})^T \right] = 2(\boldsymbol{x} - \boldsymbol{\mu})\Sigma^{-1} \tag{3}$$

for some $\boldsymbol{s} \in \mathbb{R}^2$. Rearranging this equation gives $\boldsymbol{x} - \boldsymbol{\mu} = \boldsymbol{s}P^T\Sigma$. Making this substitution in Equation (1) yields

$$c^2 = (\boldsymbol{x} - \boldsymbol{\mu})\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})^T = \boldsymbol{s}P^T\Sigma P\boldsymbol{s}^T \tag{4}$$

We denote points on the boundary of the projected ellipse by $\boldsymbol{y}$, so that $\boldsymbol{y} = \boldsymbol{x}P$ for some $\boldsymbol{x}$ satifying Equation (3). Then, as $\boldsymbol{\mu}_p = \boldsymbol{\mu}P$ by definition, $\boldsymbol{y} - \boldsymbol{\mu}_p = (\boldsymbol{x} - \boldsymbol{\mu})P = \boldsymbol{s}P^T\Sigma P$, again substituting $\boldsymbol{x} - \boldsymbol{\mu} = \boldsymbol{s}P^T\Sigma$. We then substitute $(\boldsymbol{y} - \boldsymbol{\mu}_p)(P^T\Sigma P)^{-1}$ for $\boldsymbol{s}$ in Equation (4). From this we can compute the analog equation for the projection as

$$(\boldsymbol{y} - \boldsymbol{\mu}_p)(P^T\Sigma P)^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_p)^T = c^2 \tag{5}$$

as claimed. □

This means the matrix $(P^T\Sigma P)^{-1}$ defines the ellipse in the 2-D projection. In general c could be any constant, but typically we would select it as a quantile of the χ^2

distribution, so that the size of the ellipse corresponds to a selected probability.

3.2. *Index specification*

Mahalanobis distance (Mahalanobis 1936) measures the distance from the center $\boldsymbol{\mu}$ in units of standard deviations, and is computed using the variance-covariance matrix Σ.

$$d_M(\boldsymbol{x}) = \sqrt{(\boldsymbol{x}-\boldsymbol{\mu})\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})^T} \tag{6}$$

To define a measure of an interesting projection we propose to use the average Mahalanobis distance in the projection, for a subset of points, W. The set of points could be chosen in different ways, but the default is those that are outside the specified ellipsoid in p-D, $W = \{\boldsymbol{x} : (\boldsymbol{x}-\boldsymbol{\mu})\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})^T > c^2\}$. Alternatives could be to select a set of observations with the largest Mahalanobis distance, manually select observations or possibly a group of points identified via clustering of the extremes.

We define our new index as

$$\sum_{\boldsymbol{w}\in W} (\boldsymbol{w}-\boldsymbol{\mu})P(P^T\Sigma P)^{-1}P^T(\boldsymbol{w}-\boldsymbol{\mu})^T. \tag{7}$$

3.3. *Additional considerations*

If the observations in W are primarily departing from the normal range in the same direction, the index will be expected to perform well in finding this average direction.

However, if the observations have very different departures from the norm, it may be useful to break them into groups, and separately optimize on these subsets. One could consider clustering these observations using angular distance to find groups of observations that have similar directions of departure. This is illustrated in the second example, see Section Section 5.2.

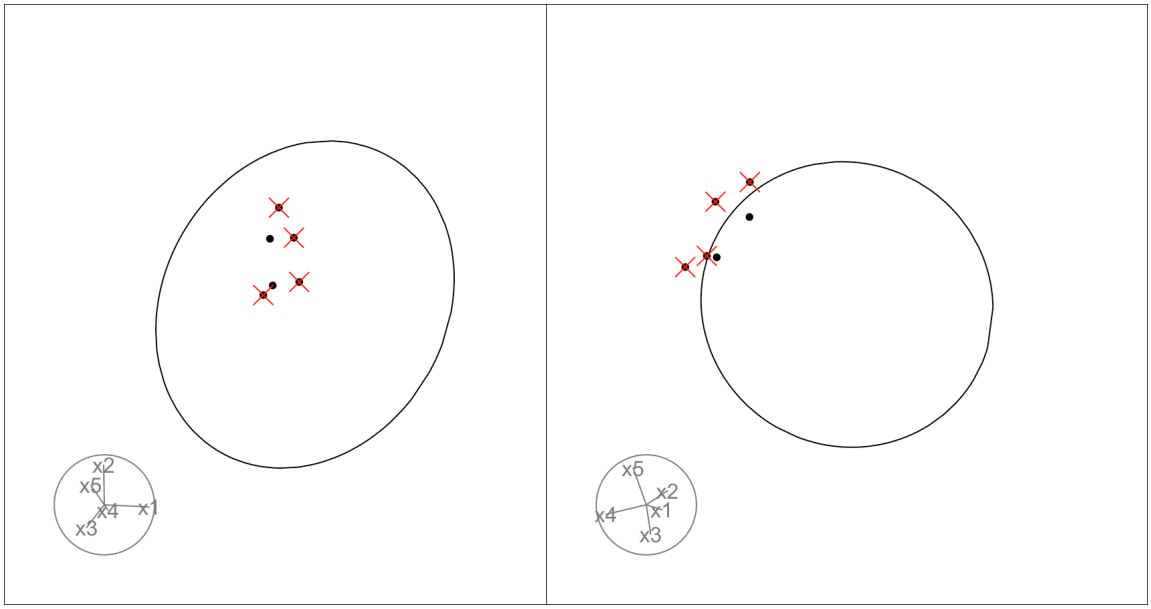
4. Implementation

This is implemented in the `tourr` (Wickham et al. 2011, 2024) package, where the projected ellipsoid can be drawn for each projection. The guided tour will take arguments specifying the data, and the null variance-covariance matrix, as shown in the example code.

```
library(tourr)
animate_xy(samp, guided_anomaly_tour(anomaly_index(),
  ellipse=vc_null), ellipse=vc_null,
  axes = "bottomleft", half_range=5, center=FALSE)
```

A random projection as well as the final view after optimization are shown in Figure Figure 3. The optimal projection (b) can be used to understand what combination of variables is driving the difference from the reference distribution.

To optimize the index any optimization algorithm implemented in the `tourr` package can be used. The focus here is on the final projection identified, and direct interpretation of index values is not considered. It may however be interesting to compare different optimization strategies and compare the index values along the optimization paths, as available in Zhang et al. (2021).



(a) Random projection (b) Optimal projection

Figure 3. Two projections of simulated example data corresponding to the sample code: (a) random projection where sample is inside the 2-D ellipse, (b) optimal projection from index, showing most of the sample outside. A red cross indicates that the point is outside the p-D ellipse. The optimal projection uses mostly variables x_4, x_5, which is expected because these are the two directions where the sample most differs from the norm.

5. Examples

5.1. *Health: liver function tests*

This example is motivated by a problem posed during consulting with a pharmaceutical company. The data shown here is simulated, for privacy purposes, but reflects the patterns seen in the clinical data. Liver function tests commonly provide measurements on albumin, protein, bilirubin, gamma-glutamyl- transferase (GGT), aspartate aminotransferase (AST), alkaline phosphatase (ALP) and alanine aminotransferase (ALT). There are normal ranges on these measurements reported by Lala, Zubair, and Minter (2023) and listed in Table 1.

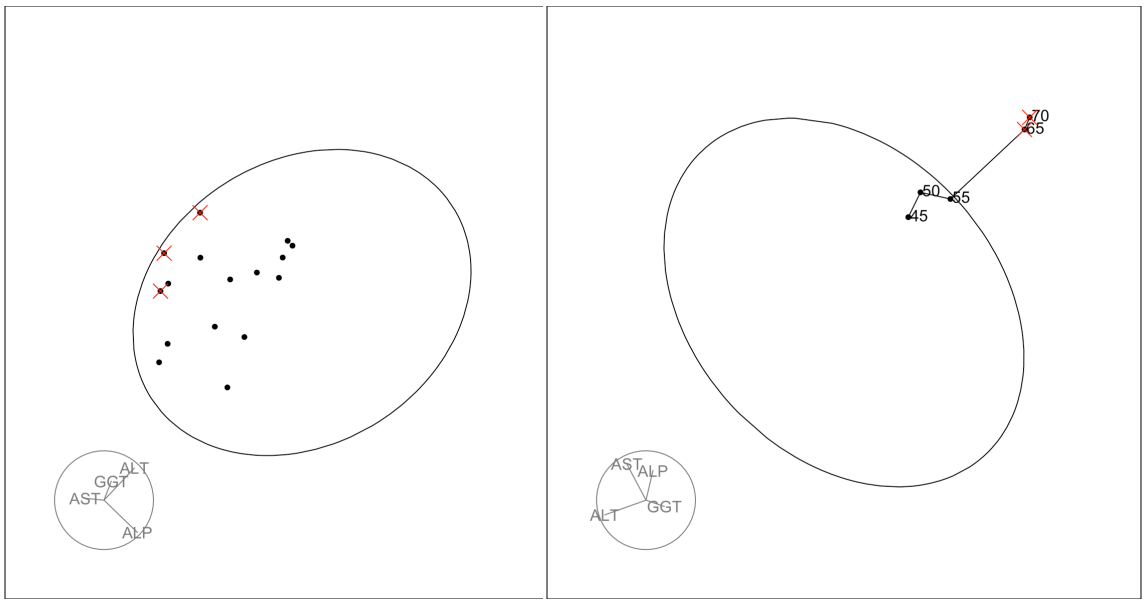
Table 1. Normal ranges provided for liver function tests.

	albumin (g/L)	protein (g/L)	bilirubin (µmol/L)	GGT (IU/L)	AST (IU/L)	ALP (IU/L)	ALT (IU/L)
min	35	60	0	2	5	30	5
max	50	80	20	44	30	120	40

These measurements are also likely correlated, based on guidance like the *ratio of AST to ALT of 2:1 indicates possible alcohol abuse*. Although this was provided by the pharmaceutical company, correlation between these measurements for normal patients is not readily available. When a correlation matrix for normal patients is provided this would allow construction of the null ellipse upon which to examine new samples.

Figure 4 illustrates two examples. The first is similar to the consulting project. A sample of liver test scores for new patients was provided in order to examine their values relative to the normal range. Here only four tests are used, GGT, AST, ALP, ALT. Plot (a) shows a projection of this sample relative to the normal range. Three of the patients are

outside the confidence ellipse but all of the patients are located away from the mean. What has been typical in the past is to compute normal values based on tests of healthy young males. The samples provided for the project were all recorded on women. We see that this sample has slightly lower ALT and ALP, which is consistent with what is reported in Lala, Zubair, and Minter (2023).



(a) Sample of female patients (b) Longitudinal profile of an aging patient

Figure 4. Two projections of simulated example data corresponding to common liver tests: (a) sample of female patients, (b) longitudinal test results for a single patient. Red cross indicates observation is outside the 4-D confidence ellipse. The female patients tend to have lower ALP and ALT than the normal range. As the patient ages, the level of ALP increases and ALT decreases.

The second example shows a longitudinal record of a single patient, measured at ages 45, 50, 55, 65 and 70. The lines connect the records in time order. The projection corresponds to the maximum from a projection-pursuit guided tour using the anomaly index:

$$P = \begin{bmatrix} 0.371 & -0.128 \\ -0.388 & 0.732 \\ 0.140 & 0.599 \\ -0.832 & -0.297 \end{bmatrix}$$

From the axes representation in Figure 4 (b) of this projection, we can see that the direction that profile extends is primarily contrasting ALT (fourth row) and ALP (third row). This is consistent with aging, where ALP increases and ALT decreases.

5.2. *Robust statistics: weather extremes*

This example is motivated by the weather data example from Mayrhofer and Filzmoser (2023). We illustrate using the anomaly index to compare potential outliers with a reference normal distribution that is derived using robust methods on the original data. The data contains 16 numerical measurements that are averages across the three summer months June, July and August, reported for 68 years (1955 - 2022) (see Mayrhofer and Filzmoser (2023) for more details).

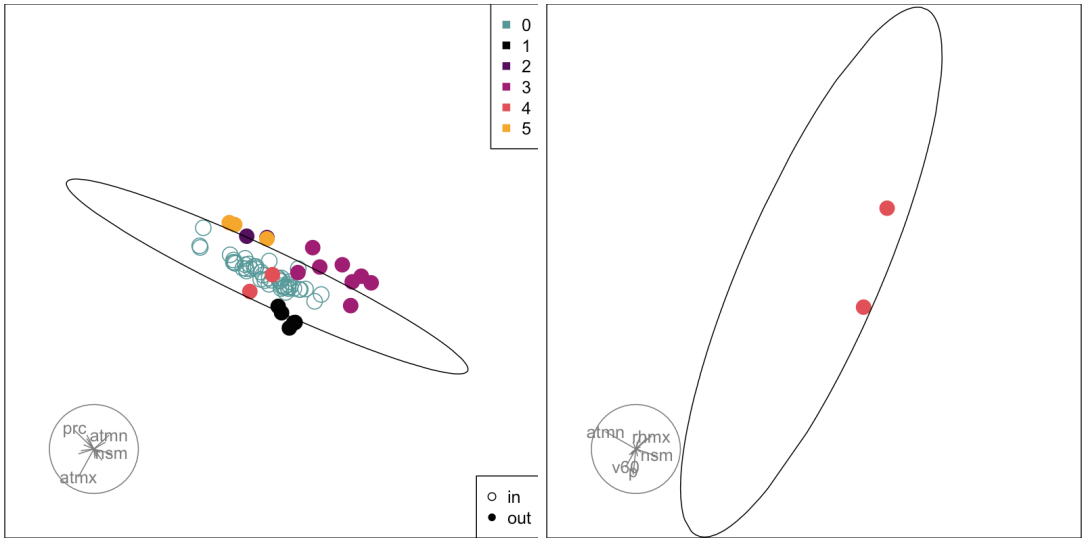
To estimate the underlying normal distribution the data is first centered and scaled using the median and the median absolute deviation (MAD), before applying the minimum covariance determinant (MCD) estimator (Rousseeuw 1985) using the implementation in Maechler et al. (2024). The MCD estimates for the mean vector and the variance-covariance matrix are then used to define the reference normal distribution.

Here we will consider points to be outlying if they are more than 5σ away from that mean value. With $p = 16$ this corresponds to a Mahalanobis distance of 60 or larger.

This will identify 20 out of the 68 observations as outlying. Since these are outlying in different combinations of variables, as found in Mayrhofer and Filzmoser (2023), we will further separate the outlying points into clusters based on similarity in direction. The similarity is computed by first normalizing observations to have length 1 and then apply k-means clustering with Euclidean distance. We use the Dunn index (Halkidi, Batistakis, and Vazirgiannis 2002), computed using Hennig (2024), to select the preferred number of clusters, $k = 5$.

The new anomaly index is first applied to the full dataset, such that the final projection will be found by averaging distance of points in many directions, see Figure 5 (a). It provides a global picture showing where the outlying points differ from the normal distribution. We can see that the contrast between maximum (atmx) and minimum (atmn) temperature is primarily used in the direction where points are most outlying. In the orthogonal direction the variables precipitation (prc) and number of days with maximum temperature above 25°C (nsm) can be used to separate the clusters 2, 3 and 5. With this solution the two points in cluster 4 cannot be resolved, they fall inside the ellipse. To better understand what is different for those two years we separately run the anomaly index for this subset, see Figure 5 (b). From the result we see that the maximum temperature is not relevant for this cluster, but that they have low values for the minimum temperature. We also see that there is still some averaging, and the preferred solution is not placing either point on the outside of the projected ellipse.

To compare our results to what is obtained by outlier detection methods, we reproduce one of the results from Mayrhofer and Filzmoser (2023) in Figure 6. On top we can see the cluster assignments for all points that we have identified as outliers using the p-dimensional Mahalanobis distance.



(a) All outlying points clustered (b) Focus on cluster 4

Figure 5. Examining weather anomalies using the anomaly index guided tour to assess what combinations of variables are creating the extremes. Anomalies were grouped based on similarity in direction, and colour indicates cluster membership. In (a) the anomaly index was computed on all outlying points, and most clusters (except for cluster 4) are outlying points are somewhat orthogonal to the normal ellipse. This extremeness is primarily due to temperature min (`atmn`) and max (`atmx`) variables as read from the axes. In (b) only the two points in cluster 4 are used for computing the index. These two observations are primarily extreme in the minimum temperature (`atmn`) variable, and have lower than normal minima.

For example we see that cluster 4 contains the years 1962 and 1996, and for these two years the result from Mayrhofer and Filzmoser (2023) also indicates unusually low values for the minimum teperature (atmn), while there is disagreement between outlyingness in other variables. Note that compared to the years 1965 and 1966 they have unusually low minimum temperature (atmn) while having average values for the number of days with maximum temperature above 25°C (nsm), and this contrast is also captured in the projection.

When looking at the overall picture (Figure 5 (a)), we can see that the linear projection is showing the difference between minimum (atmn) and maximum temperature (atmx), which is where clusters are outlying in different directions. The direction of cluster 1 is where the average maximum temperature (atmx) is unusually higher compared to the minimum temperature (atmn), while clusters 3 corresponds to years where the minimum temperature is unusually close to the maximum temperature. This combined information cannot be resolved with the cell detection algorithm, but we see that for example the most recent years (cluster 3) were tagged for having unusually high values for both the minimum (atmn) and the maximum temperature (atmx).

6. Conclusion

This paper has provided a new projection pursuit index for comparing a sample of data against a multivariate normal reference distribution. It has also provided an analytical result for drawing 2-D projections of the p-D ellipse corrsponding to the normal reference distribution. These combined provide new ways to visualise multivariate data, for this particular scenario.

The work is related to outlier detection methods (see, Mayrhofer and Filzmoser 2023, for

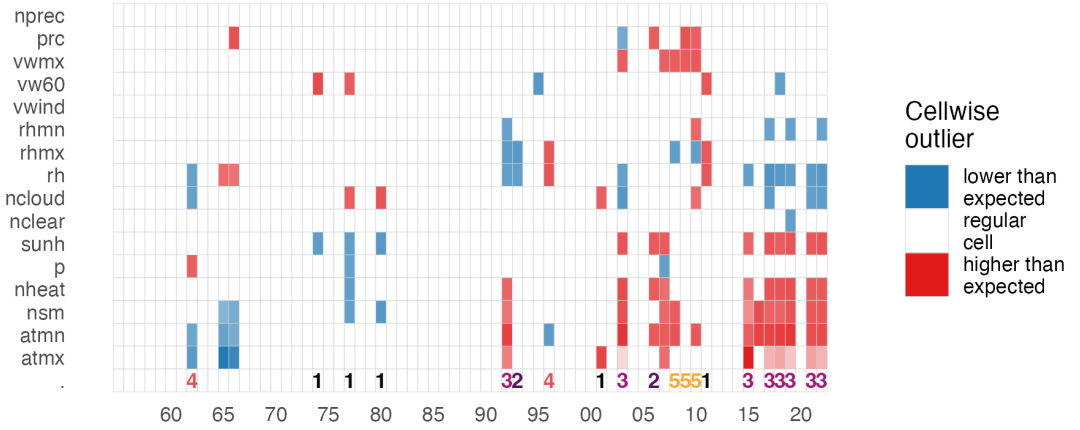


Figure 6. Reproduction of the results from Mayrhofer and Filzmoser (2023), showing which cells have been identified as outlying by their Shapley Cell Detector algorithm (Mayrhofer and Filzmoser 2023). Red cells indicate values higher than expected, blue cells values lower than expected. For years that have been identified as outliers in our approach we show the cluster assignment in the last row.

example), particularly those that use robust statistics, as illustrated in the weather example. Methods discussed in Rousseeuw, Raymaekers, and Hubert (2018) and Donoho and Gasko (1992) describe detecting outlyingness using projections of the data. When these projections are collected and used as a large set, they define a high-dimensional confidence region of an indeterminate shape. There is no analytical definition. However, we could explore the data relative to the high-dimensional shapes using the same approach described in Section 2. You would generate points on the surface of the irregular shape, overlay the data on this to visualise with a tour.

The methods on directional outlyingness or Stahel-Donoho outlyingness, described above, lend themselves to new potential 1-D projection pursuit indexes. These would integrate nicely into a 1-D projection pursuit guided tour to interactively visualise the potential outliers relative to the rest of the sample.

Another potential direction of this work is with model-based clustering using Gaussian

mixtures (Fraley and Raftery 2002), where ellipses form the basis of the model. Currently, the `mclust` software (Scrucca et al. 2023) has the capacity to show 2-D ellipses for two variables, or axis parallel 2-D ellipses for p-D. The equations developed here for drawing the 2-D projection of a p-D ellipse would provide more versatile visualisation for examining the model-based clustering fits.

Acknowledgements

The work of first author Calvi was supported by the ResearchFirst undergraduate research program at Monash University. German Valencia advised on the project, in relation to physics applications.

Supplementary Materials

The full paper, code and data to reproduce the work, are publicly available at https://github.com/uschiLaa/anomaly_ppi.

References

Asimov, Daniel. 1985. "The Grand Tour: A Tool for Viewing Multidimensional Data." *SIAM Journal of Scientific and Statistical Computing* 6 (1): 128–43.

Buja, Andreas, and Daniel Asimov. 1986. "Grand Tour Methods: An Outline." *Computing Science and Statistics* 17: 63–67.

Buja, Andreas, Dianne Cook, Daniel Asimov, and Catherine Hurley. 2005. "Computational Methods for High-Dimensional Rotations in Data Visualization." In *Handbook of Statistics: Data Mining and Visualization*, edited by C. R. Rao, E. J. Wegman, and J. L. Solka, 391–414. Amsterdam, The Netherlands: Elsevier/North-Holland.

Cook, Dianne, Andreas Buja, Javier Cabrera, and Catherine Hurley. 1995. "Grand Tour and Projection Pursuit." *Journal of Computational and Graphical Statistics* 4 (3): 155–72. https://doi.org/10.1080/10618600.1995.10474674.

Cook, Dianne, Eun-Kyung Lee, Andreas Buja, and Hadley Wickham. 2006. "Grand Tours, Projection Pursuit Guided Tours and Manual Controls." In *Handbook of Data Visualization*, edited by C.-H. Chen, W. Härdle, and A. Unwin. Berlin: Springer.

Cook, Dianne, and Deborah F. Swayne. 2007. *Interactive and Dynamic Graphics for Data Analysis: With R and GGobi*. Use r! New York: Springer-Verlag. https://doi.org/10.1007/978-0-387-71762-3.

Donoho, David L., and Miriam Gasko. 1992. "Breakdown Properties of Location Estimates Based on Halfspace Depth and Projected Outlyingness." *The Annals of Statistics* 20 (4): 1803–27. https://doi.org/10.1214/aos/1176348890.

Fraley, Chris, and Adrian Raftery. 2002. "Model-Based Clustering, Discriminant Analysis, Density Estimation." *Journal of the American Statistical Association* 97: 611–31.

Friedman, Jerome H., and John W. Tukey. 1974. "A Projection Pursuit Algorithm for Exploratory Data Analysis." *IEEE Transactions on Computing C* 23: 881–89.

Gabriel, K. Ruben. 1971. "The Biplot Graphical Display of Matrices with Applications to Principal Component Analysis." *Biometrika* 58: 453–67.

Gower, John C., and David J. Hand. 1996. *Biplots*. London: Chapman; Hall.

Halkidi, Maria, Yannis Batistakis, and Michalis Vazirgiannis. 2002. "Clustering Validity Checking Methods: Part II." *ACM SIGMOD Record* 31 (September). https://doi.org/10.1145/601858.601862.

Hennig, Christian. 2024. *Fpc: Flexible Procedures for Clustering*. https://CRAN.R-project.org/package=fpc.

Huber, Peter J. 1985. "Projection Pursuit (with Discussion)." *Annals of Statistics* 13:

435–525.

Jones, M. C., and Robin Sibson. 1987. "What Is Projection Pursuit? (With Discussion)." *Journal of the Royal Statistical Society, Series A* 150: 1–36.

Kandanaarachchi, Sevvandi, and Rob J. Hyndman. 2021. "Dimension Reduction for Outlier Detection Using DOBIN." *Journal of Computational and Graphical Statistics* 30 (1): 204–19. https://doi.org/https://doi.org/10.1080/10618600.2020.1807353.

Lala, Vasimahmed, Muhammad Zubair, and David A. Minter. 2023. "Liver Function Tests." https://www.ncbi.nlm.nih.gov/books/NBK482489/.

Lee, Stuart, Dianne Cook, Natalia da Silva, Ursula Laa, Nicholas Spyrison, Earo Wang, and H. Sherry Zhang. 2022. "The State-of-the-Art on Tours for Dynamic Visualization of High-Dimensional Data." *WIREs Computational Statistics* 14 (4): e1573. https://doi.org/10.1002/wics.1573.

Loperfido, Nicola. 2018. "Skewness-Based Projection Pursuit: A Computational Approach." *Computational Statistics & Data Analysis* 120: 42–57. https://doi.org/https://doi.org/10.1016/j.csda.2017.11.001.

Maechler, Martin, Peter J. Rousseeuw, Christophe Croux, Valentin Todorov, Andreas Ruckstuhl, Matias Salibian-Barrera, Tobias Verbeke, Manuel Koller, Eduardo L. T. Conceicao, and Maria Anna di Palma. 2024. *Robustbase: Basic Robust Statistics*. http://robustbase.r-forge.r-project.org/.

Mahalanobis, Prasanta Chandra. 1936. "On the Generalised Distance in Statistics." *Sankhya A* 80 (Suppl 1): 1–7. https://doi.org/10.1007/s13171-019-00164-5.

Mayrhofer, Marcus, and Peter Filzmoser. 2023. "Multivariate Outlier Explanations Using Shapley Values and Mahalanobis Distances." *Econometrics and Statistics*. https://doi.org/10.1016/j.ecosta.2023.04.003.

Rousseeuw, Peter J. 1985. "Multivariate Estimation with High Breakdown Point." *Math-*

ematical Statistics and Applications 8 (283-297): 37.

Rousseeuw, Peter J., Jakob Raymaekers, and Mia Hubert. 2018. "A Measure of Directional Outlyingness with Applications to Image Data and Video." *Journal of Computational and Graphical Statistics* 27 (2): 345–59. https://doi.org/10.1080/10618600.2017.1366912.

Scrucca, Luca, Chris Fraley, T. Brendan Murphy, and Adrian E. Raftery. 2023. *Model-Based Clustering, Classification, and Density Estimation Using mclust in R*. Chapman; Hall/CRC. https://doi.org/10.1201/9781003277965.

Wickham, Hadley, Dianne Cook, Heike Hofmann, and Andreas Buja. 2011. "Tourr: An R Package for Exploring Multivariate Data with Projections." *Journal of Statistical Software* 40 (2). https://doi.org/10.18637/jss.v040.i02.

Wickham, Hadley, Dianne Cook, Nicholas Spyrison, Ursula Laa, H. Sherry Zhang, and Stuart Lee. 2024. *Tour Methods for Multivariate Data Visualisation*. https://CRAN.R-project.org/package=tourr.

Zhang, H. Sherry, Dianne Cook, Ursula Laa, Nicolas Langrené, and Patricia Menéndez. 2021. "Visual Diagnostics for Constrained Optimisation with Application to Guided Tours." *The R Journal* 13 (2): 624–41. https://doi.org/10.32614/RJ-2021-105.