Reconceptualizing psychopathology as extreme deviations from a normative reference model

Charlotte Fraza[1,2], Mariam Zabihi[1,2,5,] Christian F. Beckmann[1,2,4], Andre F. Marquand[1,2,3]
[1] Donders Institute for Brain, Cognition, and Behavior, Radboud University, Nijmegen, the Netherlands
[2] Department of Cognitive Neuroscience, Radboud University Medical Center, Nijmegen, the Netherlands
[3] Department of Neuroimaging, Centre for Neuroimaging Sciences, Institute of Psychiatry, King's College London, London, UK
[4] Centre for Functional MRI of the Brain (FMRIB), Nuffield Department of Clinical Neurosciences, Wellcome Centre for Integrative Neuroimaging, University of Oxford, Oxford, UK
[5]MRC Unit for Lifelong Health & Ageing, University College London (UCL), London, United Kingdom

Corresponding Author - Charlotte Fraza, Charlotte.fraza@donders.ru.nl

Abstract
When considering psychiatric disorders as deviations from a 'norm', the tails of the distribution are considerably more informative for understanding risk for, stratifying, and predicting such disorders. Yet, most statistical methods developed for psychiatry with neuroimaging, focus on modeling the bulk of the distribution and fail to adequately capture extreme values occurring in the tails. To address this, we propose a new framework that combines normative models with multivariate extreme value statistics to accurately model extreme deviations of a reference cohort for individual participants. Similar to the employment of normative growth charts in pediatric medicine to track a child's weight in relation to their age, where an extremely high or low value can be indicative of ill health, normative models can be used with neuroimaging measurements to quantify individual neurophenotypic deviations from a reference cohort. In this article, we begin with a non-technical introduction to the fundamental principles of extreme value statistics, to accurately map the tails of the normative distribution for biological markers, including multivariate tail dependence across markers. Afterward, we give a demonstration of our method to the UK Biobank dataset and demonstrate how extreme values can be used to estimate risk and detect atypicality accurately. This new framework provides a valuable tool for accurately modeling extreme deviations in neurobiological data, which could provide us with more accurate and effective diagnostic tools for neurological and psychiatric disorders.

Introduction
In many domains, understanding the extreme values of a given variable is of great importance since they have the most significant impact on outcomes of interest (1–3). In psychiatry, we are often similarly interested in the extreme values of a distribution. There are several reasons for this. First, many neurological and psychiatric disorders are characterized by patterns of deviations in specific regions or circuits of the brain (4–6). These deviations may be difficult to detect using traditional statistical methods that focus on the mean or median values of the data. However, they may be more readily apparent in the extreme values of the data, which reflect significant deviations from the norm. For

example, we know that atypical brain development underlies many brain disorders (7,8). If we understand early deviations in brain development better, we could provide risk factors for developing several mental disorders. For example, early extreme deviations in frontal cortex development from the norm could be a risk factor for attention deficit disorder (9). Or, an extreme sudden decrease of dopamine cells in the substantia nigra could be an early biomarker for Parkinson's disease (7). Small variations in brain development are generally not cause for concern, as they are within the normal range of individual differences. However, extreme deviations from normal brain development in multiple brain regions can be indicative of underlying neurological or psychiatric conditions that can affect a person's ability to function in their daily life. By accurately mapping these extreme deviations, researchers can gain a better understanding of the brain's structure and function and identify potential risk factors for psychiatric disorders such as depression, schizophrenia, and ADHD. Thus, we can appreciate why mapping these extreme deviations becomes so important. Second, many neurological and psychiatric disorders are complex and heterogeneous, with multiple subtypes and comorbidities (10,11). By focusing on the tails of the distribution, researchers can identify distinct subtypes of the disorder that are characterized by unique multivariate patterns of extreme deviations in brain activity or structure. Finally, the tails of the distribution might also provide particularly useful information for gauging prognosis and treatment response. For instance, people with more extreme values in particular brain regions may respond better to certain types of treatment or may have worse outcomes overall.

Reliable estimation of the tails of a distribution poses a challenge, owing to several factors. Firstly, data in the tails of a distribution are, by definition, sparse, which makes it difficult to acquire a sufficient number of observations to draw meaningful conclusions (12). Secondly, classical approaches in statistics are designed to focus on the central part of the distribution rather than the tails. This makes it difficult to capture the subtle nuances of the tail behavior, leading to imprecise estimation of tail probabilities (13). Inaccurate estimation of tail probabilities can have substantial implications, particularly in fields such as neuroimaging and psychiatry, where underestimating or overestimating the likelihood of such events can lead to errors in risk assessment and treatment decisions. Accurate estimation of tail probabilities is, therefore, critical for obtaining reliable estimates of risk and optimal clinical decision-making.

In this paper, we propose a novel framework for understanding and exploring patterns in extreme brain deviations for the individual, based on extreme value statistics and normative modeling. Normative modeling provides statistical estimates of the distribution of a given brain measure across the population and can be used to estimate individual risk or deviation scores (8,14). Combining the strength of normative modeling with extreme value statistics will allow us to accurately estimate extreme deviation risk scores for the individual. The combination of these methods enables us to establish brain-behavior mappings that are more sensitive to the multivariate structure underlying the extreme deviations from an expected pattern (e.g., normative brain development). We predict that the accurate understanding of small probabilities of extreme events is crucial for understanding several types of behavior and psychiatric disorders. The main goal of the paper is a proof of principle, to demonstrate that multivariate extreme value statistics is feasible for neuroimaging data. Our main contributions are: (i) Developing the first framework for multivariate extreme value statistics for neuroimaging data. (ii) Applying and demonstrating it on a large imaging dataset - UK biobank. (iii) Establishing clear underlying

tail pairwise dependence between several brain phenotypes. (iv) Summarizing the tail pairwise dependence using principal components based on the extreme values, henceforth called extreme principal components. (v) Relating these extreme principal components to several behavioral phenotypes. First, we give a didactic introduction to the theory underlying normative modelling and extreme value statistics, to contextualize our approach. Then, we develop an exploratory statistical framework of multivariate extremes for individual participants. Finally, we validate the approach using the UK biobank.

1. Theoretical background

1.1 Normative models

Normative modeling is a method created for understanding and modeling the heterogeneity in the population. Similar to growth charts, which are used to model the distribution of height or weight of children according to their age and sex, normative models can be used to model neuroimaging data according to a set of clinical or demographic covariates (e.g. age, sex, and site) (14,15). The output gives us the mean and centiles of variation of the data as a function of the covariates. Using normative models, we shift away from the case-control setting and can make inferences of deviation from a reference cohort for each individual. Normative models have been used extensively to create large reference cohorts, establish individual deviation scores from these reference cohorts, and link the deviation scores to diverse psychiatric problems (4–6,16,17). We want to use the deviation scores from various structural and functional brain measures for 'correlation to' and 'prediction of' the psychiatric behavioral patterns. However, it is unclear what a good approach to relate the numerous deviation scores in individual neurobiological measures to psychopathology at the subject level is, especially if deviations are observed in multiple correlated variables.

1.2 Conceptualizing multivariate deviations using a watershed model

Most models of mental disorders conceptualize them as the emerging end product of aggregated causes, from disruptions in genetic pathways, and brain networks, to the environment (18). For example, multiple neural system dysfunctions have been recorded that can lead to disruptions in attention, memory, and emotion regulation (19). Small individual disturbances to certain nodes in the brain network can cause combined effects that surpass a liability threshold for a specific or multitude of mental disorders. A complex psychiatric state such as schizophrenia, major depression, and anxiety disorder is then thought to be caused by the combined heterogeneous deviations from typical network functioning or by one local disruption in the network. To illustrate this point, we give an analogy inspired by hydrology, where extreme value statistics are commonly used, see Figure 1. In order to understand the effects of a flooding event, predicting the frequency and intensity of extreme values at different nodes on the river becomes much more important than predicting the change in the mean, because the combined extreme events are what cause the most damage (20). For example, when we consider the flooding of a river basin or watershed, it becomes crucial to map the dependency structure of the tails at several gauging stations. The extreme data from the gauging stations are correlated because they are part of a connected river network. We illustrate this in Figure 1, where we give a visual representation of the different types of extreme data that can exist at individual gauging stations that together cause a flood. It could be that one river in the network is having local water elevation, due to, for example, snow melting (Figure 1. A). This local water elevation then causes flooding downstream at the river basin. Or it could be that the

water elevation is raised at multiple locations due to a more generalized effect, e.g. a storm across the land (Figure 1. B and C). When we try to understand the flooding of the river basin it is important that we model the extremal dependence structure between the different gauging stations and water flow correctly, as it is not only the local extreme events but also a co-occurrence of multivariate extremes that determines how severe the flooding will be (21,22).

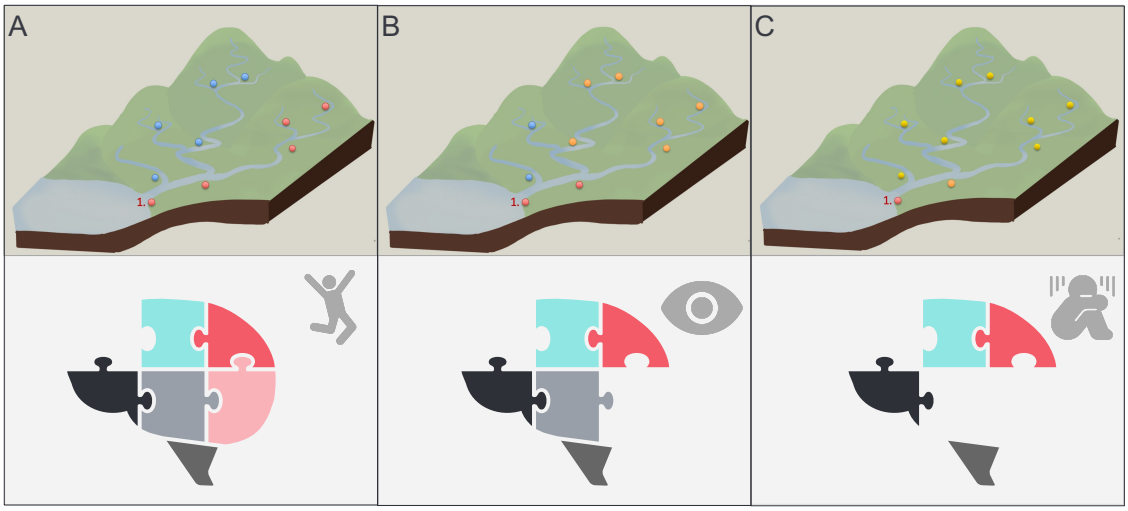


Figure 1 - Visual example of different types of extreme river elevations that can cause local flooding at location 1. In panel A local elevation along one water bank gives rise to the flooding of the river basin at location 1. In panels B and C, it is the multivariate extreme elevation of the rivers in combination that gives rise to the flooding at location 1. Likewise, extremes in behavior can be understood as being the end product of several large deviations in certain brain regions or networks.

Following Cannon's watershed model (18), we can understand the connection with psychopathology. Pathophysiological brain states can manifest either due to local brain disruptions in certain brain regions or nodes in the network or via a more widespread disruption in connected brain systems. The widespread disruption can cause psychopathology by a 'downstream' aggregation of different effects. Thus, there are many different ways that small causal effects (the tributaries) can be aggregated to yield the full-blown expression of a disorder (the river delta). The neural correlates of psychopathology can be understood as manifesting at the extreme end of continuous axes of variation across the population. Furthermore, these extreme deviations may be spread across multiple brain systems and may not necessarily overlap across individuals. Additionally, these deviations may not be reflected in only one data modality.

1.3 Introduction to Extreme Value Statistics

We will rely on extreme value statistics to accurately model probabilities of the tails of the distribution (23). We start with exploring a univariate extreme value model fitting a generalized Pareto distribution with the 'peaks over threshold' method (see next paragraph). In general, the univariate peaks over threshold method serves as a building block for the multivariate extreme value statistics method. There are two main motivations for adding the univariate method. First, depending on the question of the user the univariate approach could be sufficient. For example, if one suspects that the extremes in *one* brain region are most predictive of behavior, such as the amygdala in generalized anxiety disorder (24). Second, threshold selection for multivariate extreme value statistics is done by looking at individual model fit in univariate space, which will be further discussed in the paper.

For the multivariate extreme values, this work employs a framework that is specifically created for exploring extreme tail dependence when the dimensions become large, such as when we look at multiple brain regions or voxels simultaneously. We employ the tail pairwise dependence matrix (TPDM) (25) to understand the dependence structure in the extremes, which is similar to correlation or covariance principal components analysis (PCA), but specified for the tails. This is motivated by the fact that, whilst covariance PCA estimates variation about the mean, it is not well-suited to modeling multivariate tail dependency for the extremes, because dependence in the bulk of a multivariate distribution may not be equivalent to dependence in the tails (25). The TPDM is similar to a covariance matrix, but instead of describing the dependence structure around the mean of the distribution, the TPDM captures the dependence structure in the joint tails. Afterwards, the TPDM can be used with an eigendecomposition which gives us an ordered basis of eigenvectors based on the extremes, and values that capture, in order, the most important vectors for explaining the extreme dependence structure (25,26). We call the method 'TPDM PCA' or 'extreme PCA' and the associated principal components (PCs) 'TPDM PCs' or 'extreme PCs'. The mathematical derivation of the TPDM is described in detail in Supplement 1. Similar to standard PCs the TPDM PCs can be interpreted by looking at the individual contributions of each variable to the TPDM PCs, for example, certain regions of interest (ROIs) in neuroimaging data. This gives an indication of which variables are most important in understanding the extreme deviations.

2. Materials and methods

2.1 Multivariate extreme normative modelling

In Figure 2, we present the full overview of all the steps in the multivariate extreme value framework. We will briefly explain the framework here. For all the details, refer to the method sections in Supplement 1. First, we estimated a normative model for all the IDPs using age, sex, and site as our covariates. We applied a warped Bayesian Linear Regression (BLR) model, which helped us to predict the neuroimaging data from the covariates and provided us with predictive confidence intervals for each prediction. Afterward, we obtained the deviation or z-scores, which we combined into one large matrix for all the IDPs. We transformed the marginals, which will be explained in more detail below. Then, we thresholded the data at the 95-quantile in polar coordinates. Afterwards, the TPDM was calculated, based on the extreme quantiles of the data. On this matrix, we performed an eigenvalue decomposition, which gave us an eigenbasis that explains most of the extremal structure in the first few eigenvectors. Finally, we correlated the first few extreme PCs back to behavior, to establish a brain-behavior relationship.

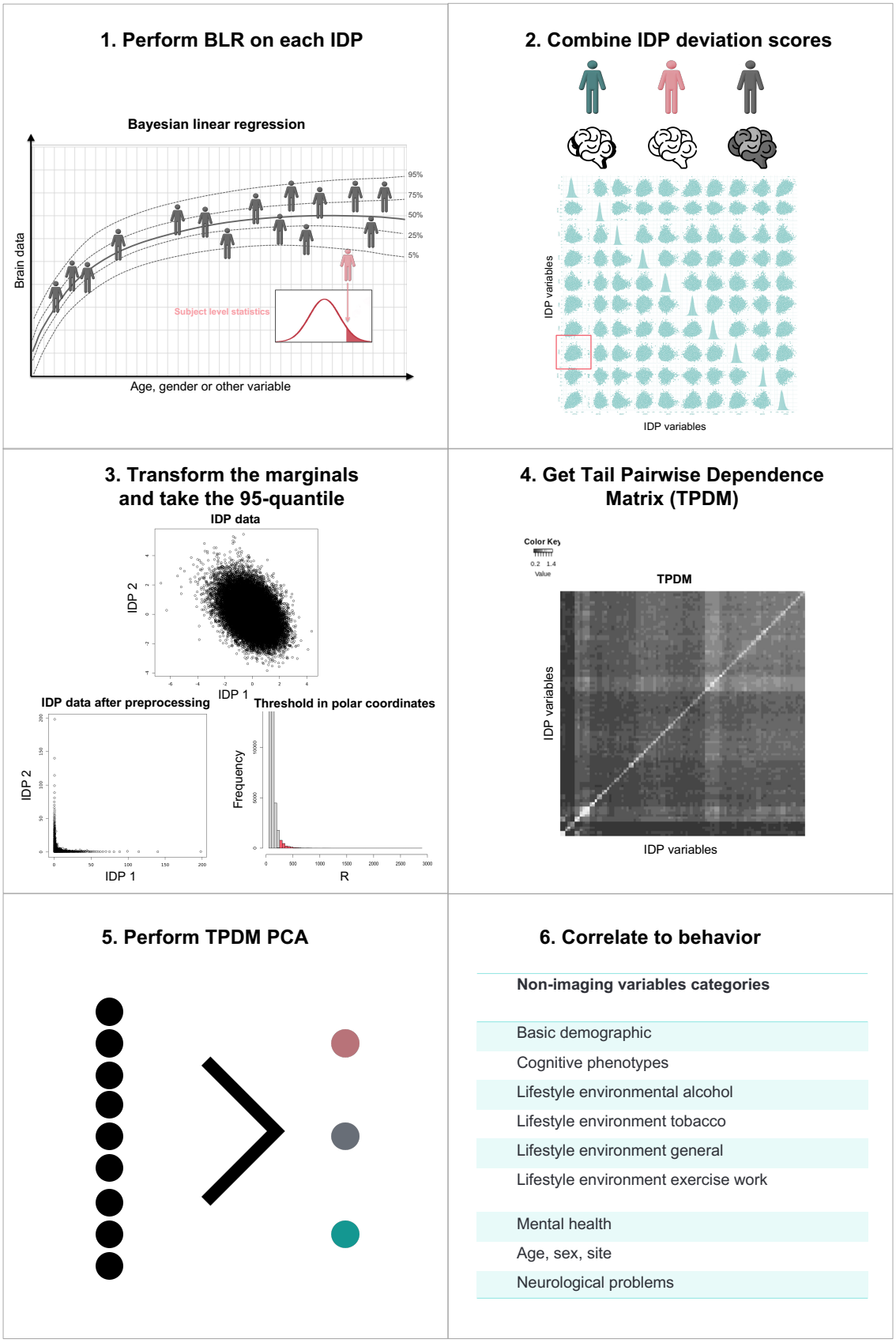


Figure 2 - Overview of the proposed extreme normative modeling approach showing all the steps in the pipeline. 1. Estimate a normative model with likelihood-warped Bayesian Linear regression for each (neuro) image-derived phenotype (IDP). This allows us to create deviation or z-scores from the mean trend for each participant. 2. Combine the z-scores of all the participants and each IDP into one matrix. 3. Transform the marginals of the data such that the marginal distribution is varying with tail index = 2 and threshold the data in polar coordinates at the 95-quantile. *4. Calculate the tail pairwise*

dependence matrix (TPDM) based on the outer 95-quantile of the z-scores. This matrix captures the extreme structure in the outer tails. 5. Perform an eigenvalue decomposition on the TPDM, creating an independent basis of which the first few eigenvectors contain most of the information of the extremal structure. 6. Correlate the first few extreme principal components to behavior to indicate how multivariate extreme values influence behavior. See the text for full details.

2.2 Extreme value theory

If we consider the entire distribution of deviation scores, we can approximate it as being composed of two components: the bulk of the distribution and the tails. Usually, we can fit a parametric model to the bulk of the distribution, such as a Gaussian distribution. However, if we are interested in the far upper and lower tails or the extreme quantiles of a distribution, we have to apply extreme value statistics, which allows us to model the tail probabilities accurately (23,27). The extreme events in the tails have a small probability to occur but usually exercise a large influence on relevant outcomes. However, the small occurrence or sparseness of measurements in the tails makes it difficult to model their probabilities accurately.

In the univariate case, two methods can be employed, i.e. the 'block maxima' approach or the 'exceedances' (a.k.a. 'peaks over threshold') approach, see Figure 3 A. and B. The block maxima approach involves dividing the data into blocks of a certain size and then finding the maximum value in each block. These maximum values are then used to create a statistical model that estimates the likelihood of extreme events occurring. The block maxima approach, although easy to employ, is not very efficient, because it ignores large parts of the data that could be relevant (for example, the second and third-highest values in each block which are potentially also informative about extreme behavior). Furthermore, this approach is mainly specified for data that have distinct time or spatial blocks with multiple measurements. Therefore, we employ the peaks over threshold approach. Instead of looking at the maximum value in each block, this method focuses on the values that exceed a certain threshold. These "peak" or extreme values are then used to create a statistical model that estimates the likelihood of extreme events occurring. The probability distribution of this method converges to a generalized Pareto distribution (GPD), which is defined by the shape ξ and scale $\tilde{\sigma}$ parameter. The mathematical details of both approaches are fully discussed in Supplement 1.

With univariate extreme value statistics, we can model the tails at one location or one IDP. However, in many circumstances, we are interested in multiple variables simultaneously, which calls for a multivariate approach. For example, when there is a strong tail dependence between certain locations in the brain or between different imaging phenotypes derived from the data. This can be done using multivariate extremes, where our objective is to characterize and understand the dependence structure in the far upper joint tail, see Figure 3 C. and D. In the bivariate case, we can fit a bivariate extreme value distribution, which tries to estimate the probability of the variable being large whilst the other is also large, see Figure 3 C. We again have to decide on an appropriate threshold similar to the univariate case (23). In this paper, we want to look at multivariate extremes beyond two dimensions. There are several methods to model the multivariate joint tails, for a full overview see (22). In our case, the chosen multivariate extreme value method is based on PCA (25), called the tail pairwise dependence matrix (TPDM), see Figure 3 D. We can also perform an eigendecomposition on the TPDM, giving us a lower-dimensional linear subspace that accounts for most of the extremal dependence. For full mathematical details for deriving the TPDM see Supplement 1.

To further understand and visualize the top extreme PCs we can perform an eigendecomposition (25). This eigendecomposition can be applied to the TPDM yielding an orthonormal basis $\widehat{U}\widehat{D}\widehat{U}^T = \hat{\Sigma}_Z$, with D the diagonal matrix of eigenvalues $(\lambda_1, \dots, \lambda_l)$ and $U = (u_1, \dots, u_l)$ the matrix of eigenvectors. We select the top eigenvectors by ordering the eigenvalues that explain the largest scale, which is similar to ordering standard eigenvalues which explain the most variance for covariance or correlation PCA. Finally, we correlate the extreme PCs with behavior to investigate how the extreme PCs map back to the behavioral phenotype.

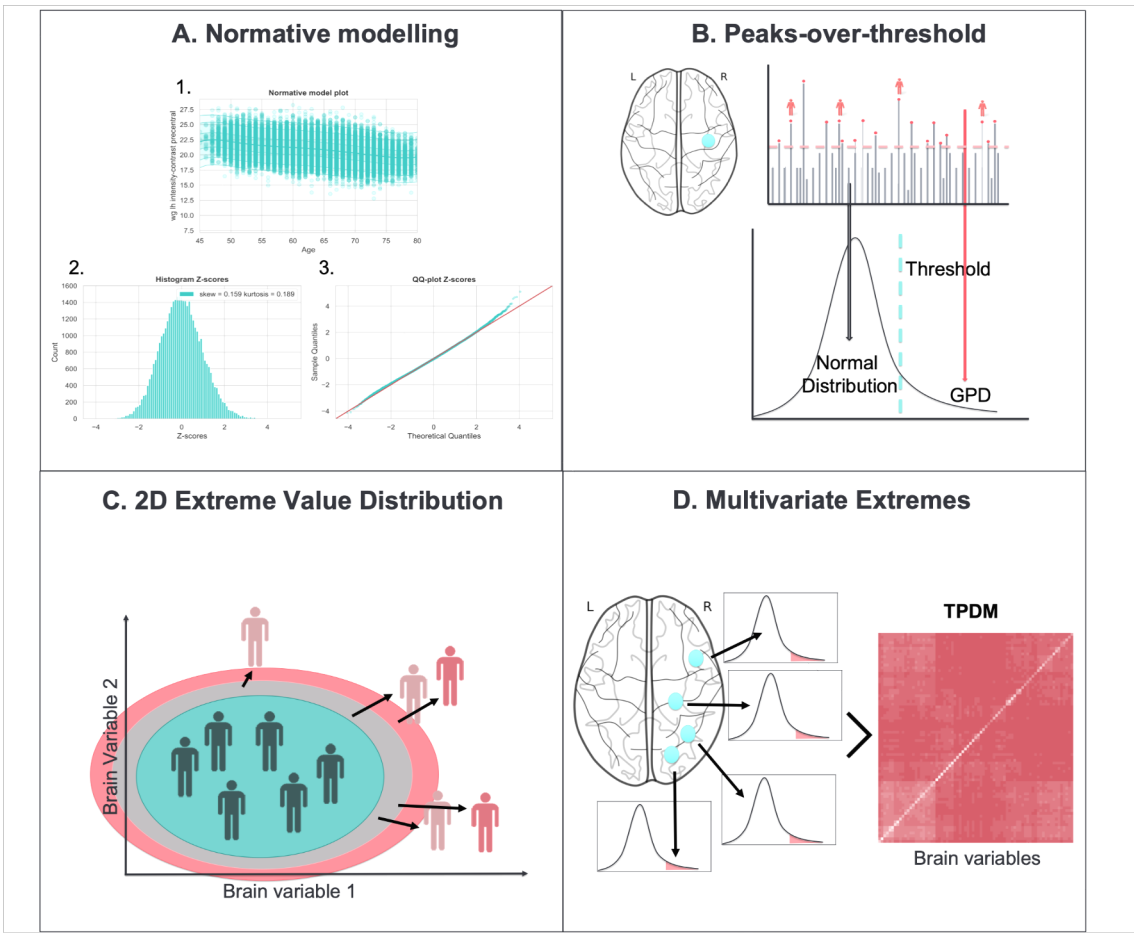


Figure 3 – Showing the different types of extreme value statistics that can be used with normative modeling. A. Normative model created for one IDP (27012) modeled as a function of age using a Bayesian Linear Regression (BLR) model. 1. Shows the normative model fit with the centiles of variation. 2. Shows the histogram of the deviation or z-scores. 3. Shows the quantile-quantile (QQ) plot of the z-scores. B. Extreme value statistics using the peaks over threshold method on the z-scores. In general, the bulk of the distribution is modelled by a normal distribution and the tails converge to a Generalized Pareto Distribution (GDP). C. The extreme value bivariate distribution for two dimensions. D. Modelling multivariate extremes for more than two variables. Here, we summarize the tail dependence with a so-called tail pairwise dependence matrix (TPDM), similar to a PCA for the bulk of the distribution.

2.3 Data Sample and Method Application

The data used in the paper came from the UK Biobank imaging data set (28,29). We used the image-derived phenotypes (IDPs), which are pre-processed imaging data, generated by the image-processing pipeline developed and run on behalf of the UK Biobank (30). These are summary measures of several imaging modalities, including structural, functional, and diffusion modalities, making them ideally suited for evaluating new methods

and theories. For the full details on the pre-processing steps and the design of the study, the following document can be read 'UK Biobank Brain Imaging Documentation' (31). We used data from 43,893 participants between 40 and 80 years old, with around 53% females. For the exact distribution of the age range and the number of participants currently in our dataset, see Supplement 1. For this study, we pre-processed the IDPs using FUNPACK (32). We used 2006 IDPs in total, based on several neuroimaging measures, including structural, functional, and diffusion MRI. We split the available IDP data into a training set and a test set with a 10-90 split, giving us 4405 participants for training and 39488 participants for testing. We want a larger amount of testing subjects to implement the new multivariate extreme value framework. For the UK Biobank, the dataset is very large, thus the number of training subjects is still acceptable; however, with a smaller dataset, we would recommend a 50-50 split. We fitted a normative BLR model with a Sinarcsinh likelihood warping and Powell's conjugate direction method. All the modeling parameters and decisions were extensively tested and validated previously (6,33). Briefly, for the covariates, we used age, sex and fixed effects for the sites to predict the IDP values. The inclusion of these covariates ensures that the interpretation of the final z-scores is largely independent of these factors. We created a B-spline basis transformation for the age covariate with 5-knots to incorporate possible non-linearity. We evaluated the performance of the BLR model and the likelihood warping by looking at the explained variance, MSLL, kurtosis, and skewness, see Supplement 1. If fitted properly the skewness and kurtosis should be around zero. Afterward, we collected the deviation or *z*-scores for each participant and all the IDPs into one *z*-scores matrix.

For the behavioral analysis, the UK Biobank dataset provided an extensive number of behavioral, clinical, lifestyle, and cognitive variables (28). These non-imaging derived phenotypes (nIDPs) were categorized into seven different categories: cognitive phenotypes (199 variables), lifestyle environmental alcohol (9 variables), lifestyle environment tobacco (4 variables), lifestyle environment general (44 variables), lifestyle environment exercise work (37 variables), mental health (29 variables), neurodegenerative disorders (6 variables).

3. Results

3.1 Univariate extreme value theory results

In this section, we will demonstrate the peak-over-threshold method for one IDP and show how an optimal threshold can be decided and the tails accurately modeled. In Figure 4, we show the different possible thresholds for one IDP, white matter hyperintensity (WMH). To decide between the different thresholds, we used graphical tools such as the mean residual life plot, and shape and scale plots, which are explained and visualized in Supplement 1.

We set the threshold at 0.95 quantile, as the graphical tools showed this to be low enough to give us an adequate amount of data and high enough to estimate the GPD parameters accurately. Once the threshold is decided, we can estimate the parameters using a maximum likelihood optimization. We optimized the profile log-likelihood to decide the scale and shape parameters. The more peaked the log-likelihood is the higher our confidence in our estimates. In Supplement 1, we show several model validation plots, such as the QQ-plot and return level plot, yielding straight plots. The straight line indicates that the fitted function is appropriate for the data (34). The scale and shape parameter estimates are: threshold: 1.9613, exceedances: 397, scale estimate 0.425 ± 0.03 (SE), shape estimate -0.065 ± 0.05 (SE).

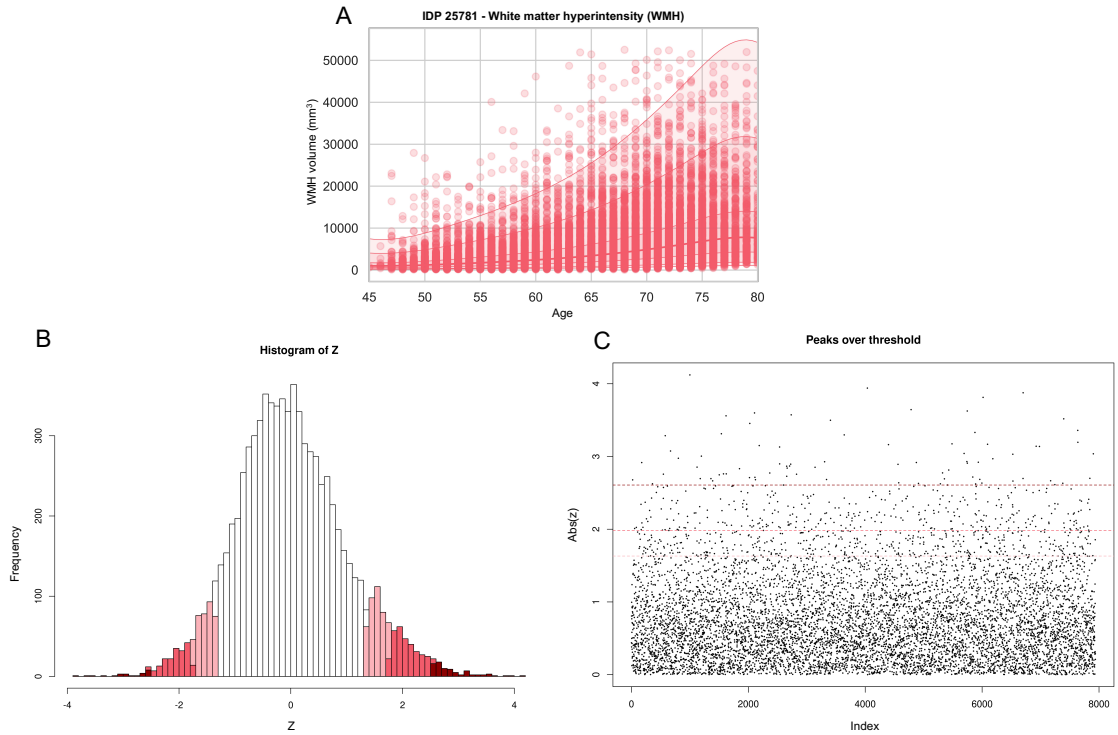


Figure 4 - Showing the univariate peak over threshold method for one IDP. A. Warped BLR normative model result for the white matter hyperintensity IDP, demonstrating the mean and centiles of variation. B. Histogram of the z-scores showing different thresholds that could be employed in the peaks over threshold method. C. Scatterplot for the peaks over threshold method demonstrating three possible thresholds of the 90th, 95th and 99th percentiles. We see that the sparsity of the points increases with higher threshold values, increasing the possible bias of the model fit.

3.2 Multivariate extreme value theory results

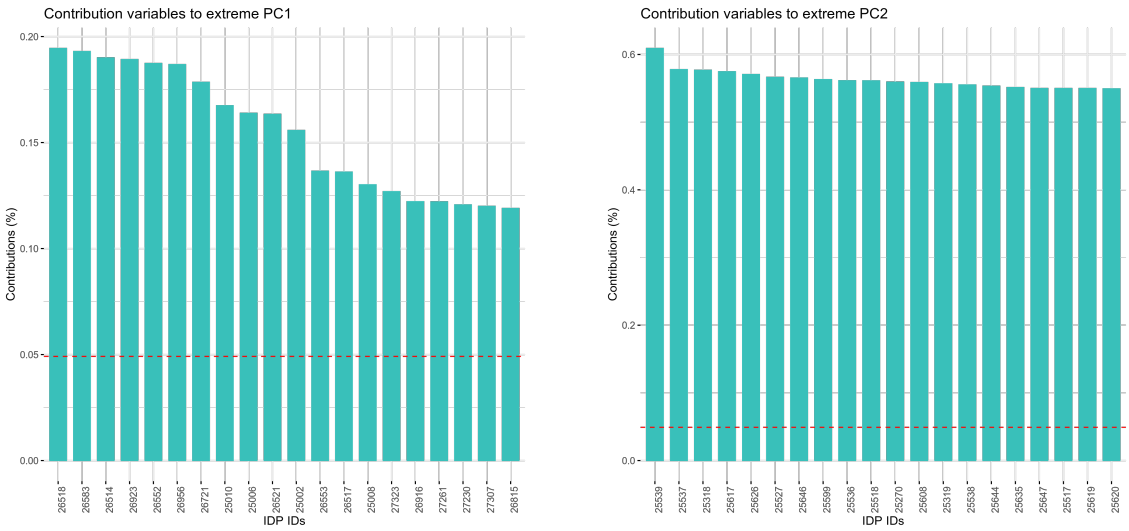
3.2.1 Thresholding Procedures

We estimated the multivariate tail dependence in the deviations or *z*-scores. We modeled the TPDM and applied an eigendecomposition to try to understand the extremal dependence. For the details of the used transformation and threshold choice, we refer the reader to Supplement 1. Briefly, to apply the method proposed by (25), we transformed the marginal distribution for each IDP, such that it fits the framework that requires z-scores to be regularly varying. To achieve this, we transformed the marginal distribution using a rank transform. Then, we applied the nonparametric marginal transformation proposed by (25) $G(\boldsymbol{z}) = \sqrt{-log\hat{F}(\boldsymbol{z})}$, with $\hat{F}$ the original marginal distribution of the z-scores. Afterward, to be able to threshold in multiple dimensions, we transformed the variables into polar coordinates $(r, \boldsymbol{\omega})$ and selected the threshold, r_0, as the 95th quantile, giving $r_0 = 247$. This threshold was selected after looking at several univariate extreme value diagnostic plots. From the $n_{exc} = 1960$ values exceeding this threshold the TPDM is calculated. Further details surrounding the mathematical formulation of these steps are provided in Supplement 1.

3.2.1 Interpretation of Extreme Principal Components

We performed a standard eigendecomposition of the TPDM or $\hat{\Sigma}_{\boldsymbol{z}}$ to obtain the eigenvectors (i.e. extreme PCs). Two of the goals of standard PCA are to extract the most important information and to analyze the structure of this information in terms of the original variables (35). Similar to looking at the variance explained in covariance-PCA, we

can look at the amount of scale that the first number of extreme eigenvalues explains. We plotted the scree plot of the scale (see Supplement 1), which shows that the first extreme PC explains 17.68% of the scale and the second extreme PC explains 3.72% of the scale. We selected the first two extreme PCs, as our informative basis. Afterwards, we try to understand this basis in terms of its original variables by estimating the individual contributions. We define the loadings for each component l: $f_l = u_l * \sqrt{\lambda_l}$, with f the loading, u the eigenvector and λ the eigenvalue. The contribution of each IDP d to the component l is then calculated by: $ctr_{d,l} = \frac{f_{d,l}^2}{\lambda_l}$, with λ_l the eigenvalue of the l-th component (35). The higher the contribution value the more that particular IDP contributes to the extreme component. In Figure 5, we plotted the highest contributions of the IDPs to the first two extreme PCs. A line is added to denote the average contribution $(1/l)$. The first extreme PC is dominated by regional and tissue volume IDPs and cortical area IDPs and the second extreme PC is dominated by white matter tract diffusivity IDPs, showing that in the extreme tails these variables vary the most together and explain the largest amount of scale for all the IDP data.



Row.names	IDP.short.name	Category.name	contrib.1.top.
26518	aseg_global_volume_TotalGray	regional and tissue volume	0.19
26583	aseg_rh_volume_Cortex	regional and tissue volume	0.19
26514	aseg_global_volume_BrainSeg	regional and tissue volume	0.19
26923	aparc-pial_lh_area_TotalSurface	cortical area	0.19
26552	aseg_lh_volume_Cortex	regional and tissue volume	0.19
26956	aparc-pial_rh_area_TotalSurface	cortical area	0.19
26721	aparc-Desikan_lh_area_TotalSurface	cortical area	0.18
25010	IDP_T1_SIENAX_brain-unnormalised_volume	regional and tissue volume	0.17
25006	IDP_T1_SIENAX_grey_unnormalised_volume	regional and tissue volume	0.16
26521	aseg_global_volume_EstimatedTotalIntraCranial	regional and tissue volume	0.16
25002	IDP_T1_SIENAX_peripheral_grey_unnormalised_volume	regional and tissue volume	0.16
26553	aseg_lh_volume_CerebralWhiteMatter	regional and tissue volume	0.14
26517	aseg_global_volume_SubCortGray	regional and tissue volume	0.14
25008	IDP_T1_SIENAX_white_unnormalised_volume	regional and tissue volume	0.13
27323	aparc-DKTatlas_rh_volume_superiorfrontal	regional and tissue volume	0.13
26916	aparc-Desikan_rh_volume_superiorfrontal	regional and tissue volume	0.12
27261	aparc-DKTatlas_rh_area_superiorfrontal	cortical area	0.12
27230	aparc-DKTatlas_lh_volume_superiorfrontal	regional and tissue volume	0.12
27307	aparc-DKTatlas_rh_volume_lateralorbitofrontal	regional and tissue volume	0.12
26815	aparc-Desikan_lh_volume_superiorfrontal	regional and tissue volume	0.12

Row.names	IDP.short.name	Category.name	contrib.1.top.
25539	IDP_dMRI_ProbtrackX_MD_str_r	WM tract diffusivity	0.61
25537	IDP_dMRI_ProbtrackX_MD_slf_r	WM tract diffusivity	0.58
25318	IDP_dMRI_TBSS_L3_Anterior_corona_radiata_R	WM tract diffusivity	0.58
25617	IDP_dMRI_ProbtrackX_L2_slf_l	WM tract diffusivity	0.57
25626	IDP_dMRI_ProbtrackX_L3_atr_r	WM tract diffusivity	0.57
25527	IDP_dMRI_ProbtrackX_MD_ifo_l	WM tract diffusivity	0.57
25646	IDP_dMRI_ProbtrackX_L3_str_l	WM tract diffusivity	0.57
25599	IDP_dMRI_ProbtrackX_L2_atr_r	WM tract diffusivity	0.56
25536	IDP_dMRI_ProbtrackX_MD_slf_l	WM tract diffusivity	0.56
25518	IDP_dMRI_ProbtrackX_MD_atr_r	WM tract diffusivity	0.56
25270	IDP_dMRI_TBSS_L2_Anterior_corona_radiata_R	WM tract diffusivity	0.56
25608	IDP_dMRI_ProbtrackX_L2_ifo_l	WM tract diffusivity	0.56
25319	IDP_dMRI_TBSS_L3_Anterior_corona_radiata_L	WM tract diffusivity	0.56
25538	IDP_dMRI_ProbtrackX_MD_str_l	WM tract diffusivity	0.56
25644	IDP_dMRI_ProbtrackX_L3_slf_l	WM tract diffusivity	0.55
25635	IDP_dMRI_ProbtrackX_L3_ifo_l	WM tract diffusivity	0.55
25647	IDP_dMRI_ProbtrackX_L3_str_r	WM tract diffusivity	0.55
25517	IDP_dMRI_ProbtrackX_MD_atr_l	WM tract diffusivity	0.55
25619	IDP_dMRI_ProbtrackX_L2_str_l	WM tract diffusivity	0.55
25620	IDP_dMRI_ProbtrackX_L2_str_r	WM tract diffusivity	0.55

Figure 5 - List of the top 20 contributions of the different IDPs to the first and second extreme PCs. We can see that the first extreme PC is largely dominated by IDPs relating to whole brain tissue volume and area. The second extreme PC is largely dominated by IDPs relating to white matter tract diffusivity.

3.2.2 Extreme PCs correlation with behavior

We wanted to correlate the extreme PCs to behavioral variables to ensure that the model can be used for behavioral predictions in clinical samples. In Figure 6, we present the

Manhattan plots of the p-values for the univariate Spearman correlation between the nIDPs and the first two extreme PCs. In general, we detected 43 significant Bonferroni corrected p-values for the first PC and 130 for the second. We can see that the extreme deviation scores link strongly to behavior.

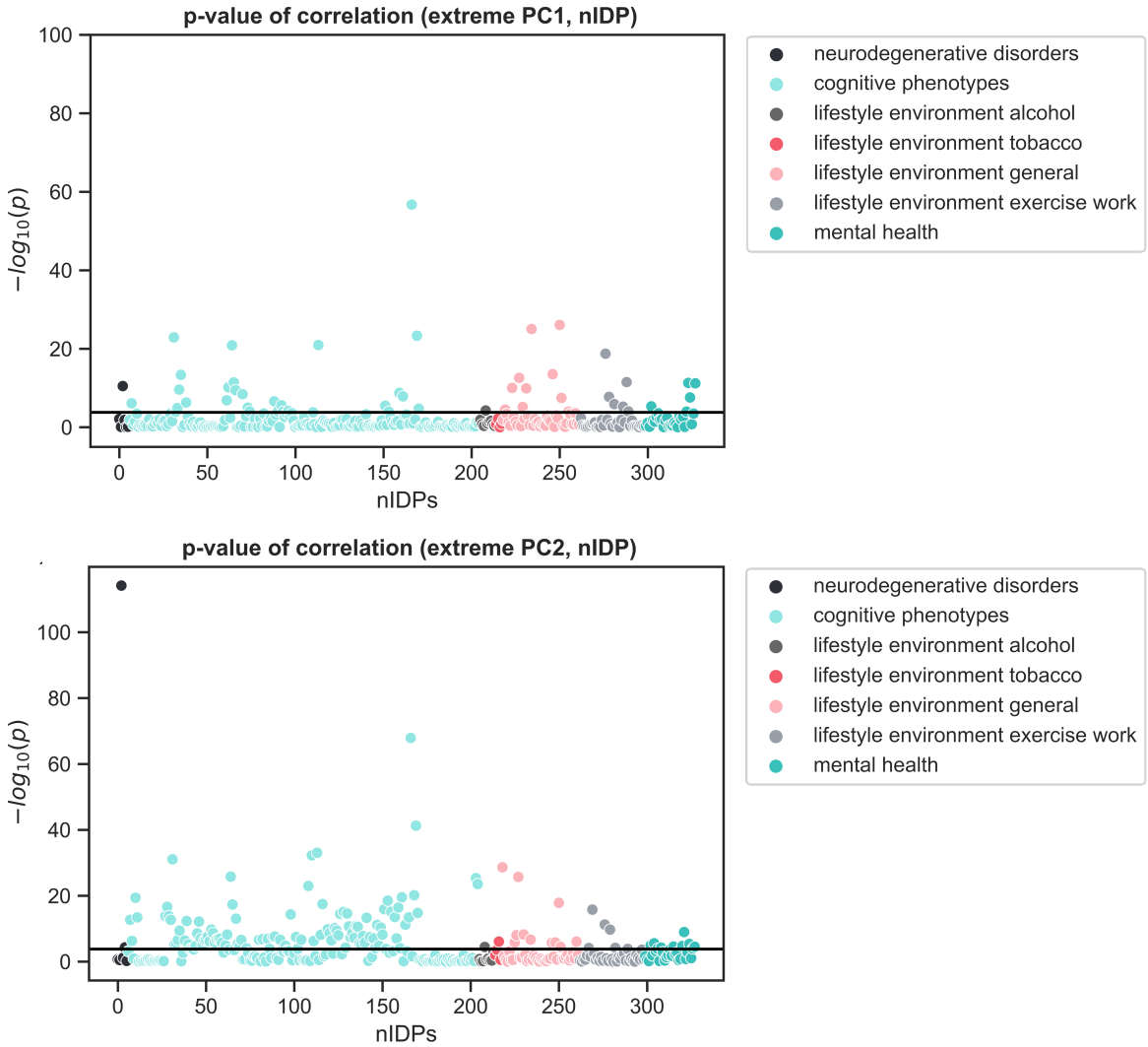


Figure 6 - Showing two Manhattan plots of the log p-values for the Spearman correlation between the nIDPs and the first two extreme PCs. The black line demonstrates the Bonferroni-corrected p-value threshold. Thus, values passing this line indicate a significant PC-nIDP correlation. We can see that mostly cognitive measures and general lifestyle nIDPs correlate with the first two extreme PCs.

Discussion

In this paper, we present a novel framework to characterize multivariate extreme tail behavior of brain imaging data, which is predicated on re-conceptualizing psychopathology as deviations from a normative pattern of functioning, where such deviations may be spread across multiple measures and have a complex correlation structure. We propose to account for the dependency structure that can be found in the tails through an eigendecomposition that can provide a good basis to further understand the tail dependence. The approach is used to identify participants deviating from the normative model in multivariate space.

In general, normative models are used to capture and understand the heterogeneity in the population by modeling individual deviations from a reference cohort. In this study, we extended the normative method and presented a new approach for exploring extreme tail behavior by summarizing it into a matrix of pairwise extremal dependence. We showed that: (i) the extremal dependence structure in the tail can be captured by a tail pairwise dependence matrix (TPDM); (ii) that the scale of the TPDM can be summarized by a low number of extreme principal components; (iii) and (iv) we observed a strong correlation between behavioral phenotypes and the first two extreme principal components. Furthermore, we demonstrated how to implement the multivariate extreme value analysis on a large dataset from the UK Biobank, using image-derived phenotypes (IDPs). The multivariate extreme method presented in this paper could offer an alternative to methods mostly focusing on mean deviation scores.

The use of normative modeling creates a suitable structure to examine disorders in individual participants without dividing the group into cases vs. controls. Rather than categorizing participants into "healthy" or "unhealthy," this approach allows for the understanding of disease as an extreme form of normal behavior or as an unusual deviation from typical functioning, which may be unique to the individual. Additionally, normative modeling allows for inferences to be drawn, regarding how behavioral variables and diagnostic labels, correspond to biological factors. Mental disorders are highly heterogeneous because they can take many different forms and have a range of symptoms, severity levels, and durations. The specific type of disorder, the person's age, sex, ethnic heritage, life experiences, genetics, and external conditions are some of the variables that can contribute to this heterogeneity (15). People who are depressed, for instance, may exhibit symptoms like sadness, diminished interest in activities, exhaustion, and adjustments to appetite and sleep patterns. However, from person to person, there can be significant variations in the degree of these symptoms as well as the precise combination of symptoms (36). Likewise, those who suffer from anxiety disorders may exhibit symptoms like excessive worry, agitation, and panic attacks, but the severity and frequency of these symptoms can vary greatly based on the person and the type of anxiety disorder they are suffering from (37). Therefore, it is important to map this heterogeneity accurately and have methods that are sensitive to these individual differences.

In our study, we have demonstrated that mental disorders can be understood as the extremum of the normal multivariate range. To this end, we propose a novel approach for understanding the underlying dependence structure between different neuroimaging modalities. Previous research has attempted to measure tail dependence between variables, as seen in (17). Our results show that the tail probability dependence matrix (TPDM) method is an effective tool for this purpose in neuroimaging data. The TPDM has been proven to be similar to a covariance matrix (18), a well-established tool in neuroimaging research for understanding the relationships between different brain areas and modalities. However, the covariance matrix only captures variance around the mean, which does not fully correspond to tail dependence. As a result, the TPDM is an alternative choice, specifically designed to capture extremal dependence in neuroimaging data. Additionally, the TPDM can be subjected to eigendecomposition, similar to covariance matrices, allowing us to identify the most important variables in quantifying the dependence structure. Overall, our proposed approach offers a powerful tool for uncovering new insights into the underlying relationships between different neuroimaging modalities and for discovering new biological markers of disease.

For extreme value statistics, one major consideration is that users have to carefully consider the questions they want the model to answer. The extreme value statistics method assumes apriori that there is more information of interest in the tails than in the bulk of the distribution. If users have a question pertaining to the general population and 'healthy' (or 'average') functioning of the brain, an extreme value approach would probably not be the most optimal. For these questions, classical approaches that focus on central tendency could be more appropriate. Therefore, we recommend users carefully assess the nature of their question and asses if it is focused on the mean or the tails of the distribution. The extreme value method is designed to give a more sensitive analysis of the multivariate tails of the distribution and is especially beneficial in applications where accurate quantification of risk is essential. Lastly, if one wants to make conclusions for the entire population one could design appropriate mixture models combining a general fit for the bulk of the distribution and an extreme value fit to the tails. In this case, giving adequate weight to the extremes and the bulk of the distribution is important.

In conclusion, we presented a comprehensive framework for understanding and visualizing the dependency structure in the tails of different neuroimaging variable distributions. The method presented here is general and can be applied to several types of neuroimaging data. In the future, we want to implement the extreme multivariate framework for understanding several psychiatric disorders. In general, one needs to train a reference normative model only once and then this can be used as the baseline to further understand smaller clinical datasets. Therefore, we expect that this method would be an excellent choice in a clinical setting, which often has smaller samples and where extreme brain deviations and their underlying dependence could have a significant effect on extreme behavior taking us further in the quest for individualized and precision medicine.

Acknowledgments This research was supported by grants from the European Research Council (ERC, grant "MENTALPRECISION" 10100118) and the Dutch Organisation for Scientific Research (VIDI grant 016.156.415). This research has been conducted using the UK Biobank resource under application number 23668.

Conflict of interest The authors declare that they have no conflict of interest.

Supplement 1

Sample

For the exact distribution of the age range and the number of participants currently in our dataset, see Figure 7.

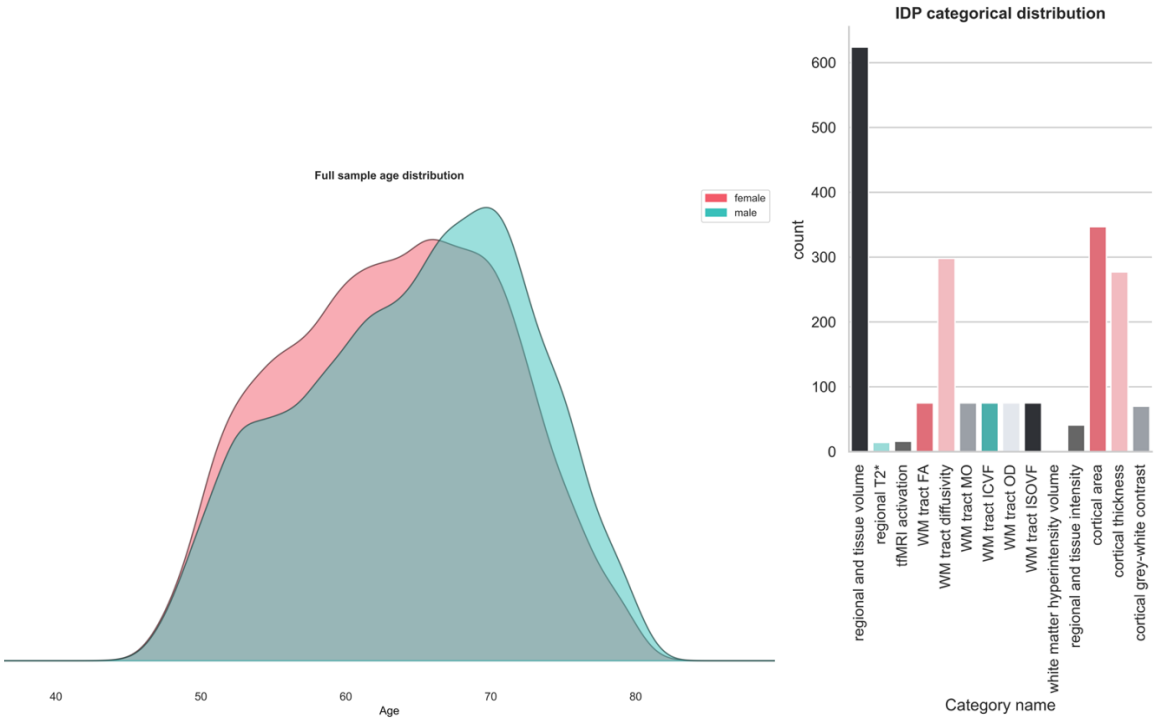


Figure 7 – On the right - showing the age and sex distribution of the IDPs used in this study from the UK biobank. On the left - showing the distribution of the different IDP categories present in the UK Biobank dataset

Normative model formulation

We estimated all the normative models using Python 3.8.3 and the PCNtoolkit version 0.20. We performed a normative model with the Bayesian linear regression model (BLR) and likelihood warping. For a full explanation and details of the mathematical framework behind this model see (38).

Here we will give a short overview of the mathematics behind the method. We define $\boldsymbol{y} = (y_{nd}) \in \mathbb{R}^{N\times D}$ with y_{nd} the d-th IDP of the n-th subject. The covariates are collected into one matrix $\boldsymbol{x} = (x_{nm}) \in \mathbb{R}^{N\times M}$, where x_{nm} is the m-th covariate of the n-th subject:

$$\boldsymbol{y} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1d} \\ y_{21} & y_{22} & \cdots & y_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nd} \end{bmatrix} \text{ and } \boldsymbol{x} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

We used the covariates age, sex, and site. Here the sites are denoted by $s \in \{1, \dots, S\}$ with all subjects having the same sites. For each IDP we fitted a separate model. To keep the notation concise, we will concentrate on one specific IDP labelled d and drop the subscript. Thus, for every IDP we denote $\boldsymbol{y} = (y_1, \dots, y_N)^T$ and take the set of independent variables $\boldsymbol{x}_n = (x_{n1}, \dots, x_{nM})^T$. For every subject, we specified the model as follows:

$$\varphi(y_n) = \boldsymbol{w}^T \phi(\boldsymbol{x}_n) + \epsilon_s \tag{1}$$

Where, $\boldsymbol{w}^T$ is the estimated vector of weights and $\phi(\boldsymbol{x})$ is a basis expansion of the covariate vector $\boldsymbol{x_n}$. In our case, a cubic B-spline basis expansion with 5 evenly spaced knots was

chosen. Empirically, this was enough to capture the curvature in space caused by the age covariate. $\epsilon = \mathcal{N}(0, \beta^{-1})$ is a Gaussian noise distribution with mean zero and noise precision term β (the reciprocal of the variance). $\varphi(y_n)$ is a likelihood warping function used to accommodate non-Gaussianity of the residuals of the data in the original space. For the likelihood warping a Sinarcsinh function was employed:

$$\varphi(y_n, \boldsymbol{\gamma})_{Sinarcsinh} = \sinh\left(c * \operatorname{arcsinh}(y_n) - d\right) \tag{2}$$

With $\boldsymbol{\gamma} = (c, d)$ the identified parameters for the warping function. This method has been shown to be able to model Gaussian as well as non-Gaussian distributions (Fraza et al 2021). If non-Gaussianity is present in the data, there are other techniques besides likelihood warping that one can consider. A pre-transformation of the dependent variable, like a Box-Cox or log-transform, is one illustration. Finding the right transformation beforehand that is optimal across various datasets, however, can be quite difficult. The warped BLR model thus has the additional benefit of eliminating the extra step of selecting an appropriate transformation for each dataset by incorporating this in the model through likelihood warping (38,39). We captured the site variation using a fixed-effects model, according to (40). We performed the optimization using Powell's conjugate direction method by minimizing the negative log-likelihood.
Afterward, the *z*-scores were calculated in the warped space for each subject, *n*, and IDP, *d*, in the test set as:

$$z_{nd} = \frac{y_{nd} - \hat{y}_{nd}}{\sqrt{\sigma_d^2 + (\sigma_*^2)_d}} \tag{3}$$

Where, y_{nd} is the true response, $\hat{y}_{nd}$ is the predicted mean, σ_d^2 is the estimated noise variance (reflecting uncertainty in the data), and $(\sigma_*^2)_d$ is the variance attributed to modeling uncertainty, for the full derivations see (14,38). We evaluated the model fits according to several model criteria: explained variance (R^2), mean squared log-loss (MSLL), and skew and kurtosis. Together these criteria allowed us to assess the central tendency, performance of the warping function, as well as overall model fit. Afterward, we used the *z*-scores to estimate the extreme value distributions.

Extreme value statistics

Most of the univariate and multivariate extreme value statistics were performed in R version 4.1.0. There are several dedicated packages for extreme value theory created in R, see https://cran.r-project.org/web/views/ExtremeValue.html. We defined z_{nd} as our actual measured values or deviation scores for each participant and IDPs. These values are realizations of the random variables $Z_1, \ldots, Z_{nd}$. We use uppercase letters for random quantities and lowercase letters for actual measured values to differentiate between the two. Although Z_{nd} is a random quantity, meaning it could take on a range of different values until measured, certain values are more probable than others. Z_{nd} is assumed to have a probability distribution that assigns probabilities to the various possible values or ranges of values that it could take.
First, we can estimate the probability function of the tails using the block maxima approach. We consider a set of independent variables Z_n for each variable d with a common distribution function F. We want to focus on the statistical behavior of the maxima for each variable:

$$\mathrm{M}_n = \max\left(Z_1, \dots, Z_n\right) \tag{4}$$

Usually, for the block-maxima approach, Z_n is measured with some kind of logical time frame, such as a time-block in fMRI, one full scanning session for MRI, or another repeated time measure in hours, months, and or years, see Figure 8. Alternatively, one can also define a block in the spatial domain such as a region of interest (ROI) in the brain (or in our case one IDP). In this case, M_n represents the maxima for each of those spatial blocks or timeframes. We want to know the probability distribution function F, which can be done by estimating it from observed data. It has been proven that the probability distribution of maxima M_n converges to a generalized extreme value distribution, when rescaled with parameters b_n and a_n (41,42):

$$\mathrm{P}\left\{\left(\frac{M_n - b_n}{a_n}\right) \leq z\right\} \rightarrow G(z), \quad n \rightarrow \infty \tag{5}$$

For constants $a_n > 0,\ b_n \epsilon R$ and z the upper end of the tails.

$$G(z) = \exp\left\{-\left[1 + \xi\frac{(z-\mu)}{\sigma}\right]^{-\frac{1}{\xi}}\right\} \tag{6}$$

The probability distribution $G(z)$ is defined by the shape ξ, location μ and scale σ parameters. The extreme value distributions can be further separated into Gumbel, Fréchet, and Weibull families. For convenience, we listed the full theorems used in this paper underlying the extreme value theory from (Coles, S., Bawa, J., Trenner, L., & Dorazio, P., 2001) below. What is notable about this result is that the only possible limit for the distributions of M_n, regardless of the population distribution F, is the generalized extreme value distribution. Therefore, the theorem serves as an extreme value version of the central limit theorem (23).

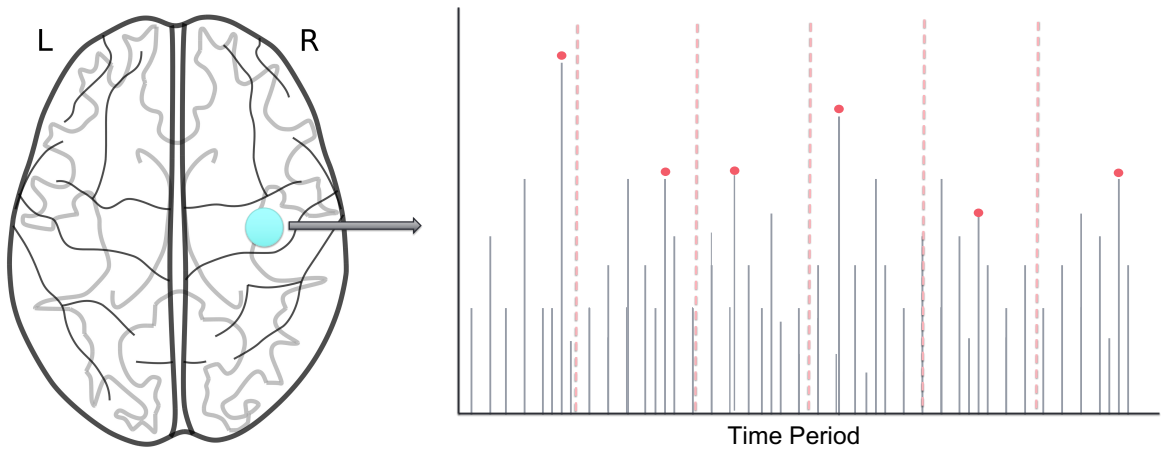


Figure 8 – Extreme value statistics using the block maxima approach. In general, if we use the block maxima approach to model the tails of a distribution. The model converges to a generalized extreme value (GEV) distribution.

The Generalized Extreme Value Distribution

Theorem 1: If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that:

$$\mathrm{P}\left\{\left(\frac{M_n - b_n}{a_n}\right) \leq z\right\} \to G(z), \quad n \to \infty \tag{7}$$

Where G is a non-degenerate distribution fucntion, then G belongs to one of the following families:

$$\mathrm{I: G(z)} = \exp\left\{-\exp\left[-\left(\frac{z-b}{a}\right)\right]\right\}, -\infty < z < \infty; \tag{8}$$

$$\mathrm{II: G(z)} = \begin{cases} 0, & z \leq b; \\ \exp\left\{-\left(\frac{z-b}{a}\right)^{-\alpha}\right\}, & z > b; \end{cases} \tag{9}$$

$$\mathrm{III: G(z)} = \begin{cases} \exp\left\{-\left[-\left(\frac{z-b}{a}\right)^{\alpha}\right]\right\}, & z < b; \\ 1, & z \geq b; \end{cases} \tag{10}$$

for parameters $a > 0$, b and, in the case of families II and III, $a > 0$. The extreme value distributions consist of three categories, which are commonly referred to as the Gumbel, Fréchet, and Weibull families, known as types I, II, and III, respectively. Each family is characterized by a location parameter (b) and a scale parameter (a). The Frechet and Weibull families also have a shape parameter (α). The Gumbel, Fréchet, and Weibull families can be merged into a unified family of models called the generalized extreme value (GEV) family of distributions.

Theorem 1.1: If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that:

$$\mathrm{P}\left\{\frac{M_n - b_n}{a_n} \leq z\right\} \to G(z), \quad n \to \infty \tag{11}$$

For a non-degenerate distribution function G, then G is a member of the GEV family:

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \tag{12}$$

Defined on $\{z: 1 + \frac{\xi(z-\mu)}{\sigma} > 0\}$, where $-\infty < \mu < \infty,\ \sigma > 0$ and $-\infty < \xi < \infty$.

An alternative method that takes multiple extreme values into account is the peaks over threshold approach:

$$M_n = kth\ largest\ of\ (Z_1, \dots, Z_n) \tag{13}$$

With k an appropriately large threshold. The probability distribution of this method converges to a generalized Pareto distribution:

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}} \tag{14}$$

which is defined by the shape ξ and scale $\tilde{\sigma} = \sigma + \xi(k - \mu)$ parameters. For more details on the underlying theory see the supplement or (23). In this method a threshold must be

specified, that is not too stringent as then there will be too few values to model the tails of the distribution accurately, but also not too low, as then the shape and scale parameters will be incorrect. One way to approach this problem is by using graphical tools with several thresholds and deciding on the best threshold based on their outcomes.

The Generalized Pareto Distribution

Theorem 2: Let $Z_1, Z_2, \dots, Z_n$ be a sequence of independent random variables with common distribution function F, and:

$$M_n = \max(Z_1, \dots, Z_n) \tag{15}$$

suppose that F satisfies Theorem 1.1, so that for large n,

$$\mathrm{P}\{M_n \leq z\} \approx G(z) \tag{16}$$

Where

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \tag{17}$$

For some $\mu, \sigma > 0$ and ξ. Then, for a large enough threshold k, the distribution function of $(Z - k)$, conditional on $Z > k$, is approximately:

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}} \tag{18}$$

Defined on $\{y: y > 0 \; and \; \left(1 + \frac{\xi y}{\tilde{\sigma}}\right) > 0\}$, where

$$\tilde{\sigma} = \sigma + \xi(k - \mu)$$

This family of distributions is called the generalized Pareto family.

Framework for the tail pairwise dependence structure

We will describe the framework used in this paper for understanding the tail pairwise dependence structure. The method is based on the framework of regular variation. A random vector is multivariate regularly varying if its joint tail decays like a power function (34). We assume $\boldsymbol{Z}$ is a regularly varying random vector. In general, for multivariate regular variation it is convenient to consider the variables in polar coordinates:

$$\boldsymbol{Z} \to \left(\|\boldsymbol{Z}\|, \frac{\boldsymbol{Z}}{\|\boldsymbol{Z}\|}\right) = (r, \boldsymbol{\omega}) \tag{19}$$

With $\boldsymbol{\omega} \in \mathbb{R}^N$ a location on the unit sphere $\mathbb{S}$ and r is the length of the vector (34). Then for a given A, a set of large values away from the origin, we can calculate the probability of $\boldsymbol{Z}$ as:

$$P(\boldsymbol{Z} \in A) \approx \int_{(r,\boldsymbol{\omega}) \in A} \alpha r^{-\alpha-1} dr dH(\boldsymbol{\omega}) \tag{20}$$

With $\approx$ signifying "approximately proportional to", $\alpha > 0$, r the magnitude or radial component of the location, $\boldsymbol{\omega}$ the location on the unit sphere $\mathbb{S}$, and H the angular measure on the unit sphere $\mathbb{S}$, see Figure 9. In the integrant, r follows a power-law distribution according to α the reciprocal of the shape parameter of the GPD (23), (26).

Tail pairwise dependence matrix TPDM

The angular measure H in lower dimensions can be modeled parametrically. However, as the dimension size increases there is not sufficient information in the extremes to capture

H. Instead of fully modelling the dependence structure contained in H an approximation can be made with the tail pairwise dependency matrix or TPDM. If we assume $\boldsymbol{Z}$ to be a regularly varying random vector with a common tail index α and angular measure H. Then each element in the TPDM $\Sigma_{\boldsymbol{z}}$ is specified according to:

$$\sigma_{i,j} = \int_{\mathbb{S}} \omega_i \omega_j dH(\boldsymbol{\omega}), \quad i,j = 1, \dots, d \tag{21}$$

The TPDM has similar interpretation properties as a standard covariance matrix only specified for the extremes (25). Thus, if two variables are independent in the outer tails $\sigma_{i,j} = 0$ and the diagonal element $\sigma_{i,i} = b^2$, with b the scale. The TPDM is positive definite and symmetric and therefore the eigenvalues are positive and the eigenvalues real (26).

Estimation of the TPDM

We estimated the dependence in the tails via the TPDM for $\boldsymbol{z} = (z_{nd}) \in \mathbb{R}^{N \times D}$. For this framework, the variables should be heavy-tailed with a common tail index $\alpha = 2$. The framework for regular variation outlined earlier necessitates that all variables exhibit heavy-tailed behavior, with a shared tail index denoted by α. To ensure this we first transformed the marginals. Transforming the marginals is a common procedure in extreme value statistics (34). We first used a rank order transformation to obtain $\hat{F}$, the estimated marginal cumulative distribution function (CDF) from the data. The rank order transformation smoothens effects of non-normality, outliers, unequal variances etc, but other transformations can be considered, see (34). Then we transformed our $\hat{F}$ with $\mathrm{G}(\boldsymbol{z}) = \sqrt{-log\hat{F}(\boldsymbol{z})}$ the CDF of a Fréchet random variable (25). Afterward, a polar coordinate transformation is applied and a threshold of the 95th quantile was chosen for the kth largest radius, see Figure 9. Finally, we estimated the TPDM according to:

$$\hat{\sigma}_{ij} = 2n_{exc}^{-1} \sum_{n=1}^{N} \omega_{ni}\omega_{nj} \mathbb{1}\,(r_n > r_0) \tag{22}$$

Where r_0 the threshold for the radial component and n_{exc} the number of observations that pass the threshold.

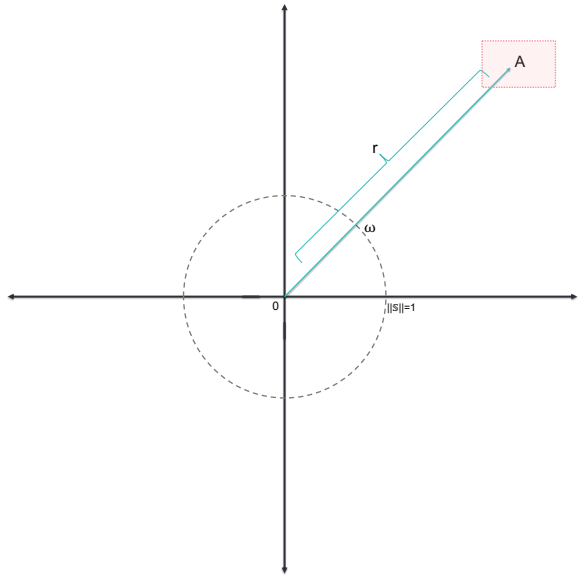


Figure 9 - Illustration of the applied polar transformation of the data. the value indicated by the radial component r becomes independent from the angular component ω. Adapted from (26).

Bayesian Linear Regression model fit

We evaluated the performance of the BLR model and the likelihood warping by looking at the explained variance, MSLL, kurtosis, and skewness, see Figure 10.

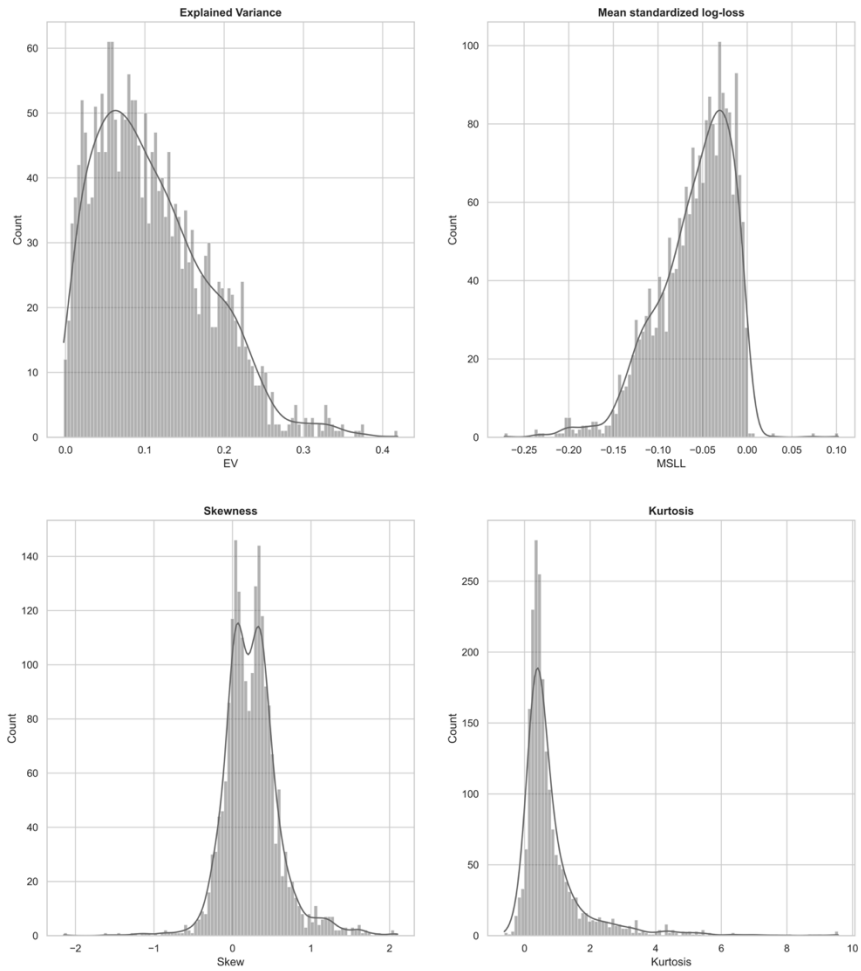


Figure 10 - Showing the performance measures for the different IDPs using a warped BLR model. In general, for an optimal model fit the skew and kurtosis would both be distributed around zero.

Accurate threshold selection

One graphical tool that can be employed is the mean residual life plot (MRL-plot) (43), which involves plotting the value of a threshold, k, against the 'mean excess', i.e. the exceedances at k, minus k. In Figure 11, we present the MRL plot for the IDP WMH. To help interpret this plot, it is assumed that the threshold should be placed where the plot starts diverging from the linear trend. In this case, we can observe that the plot is relatively linear until the 0.95 mark.

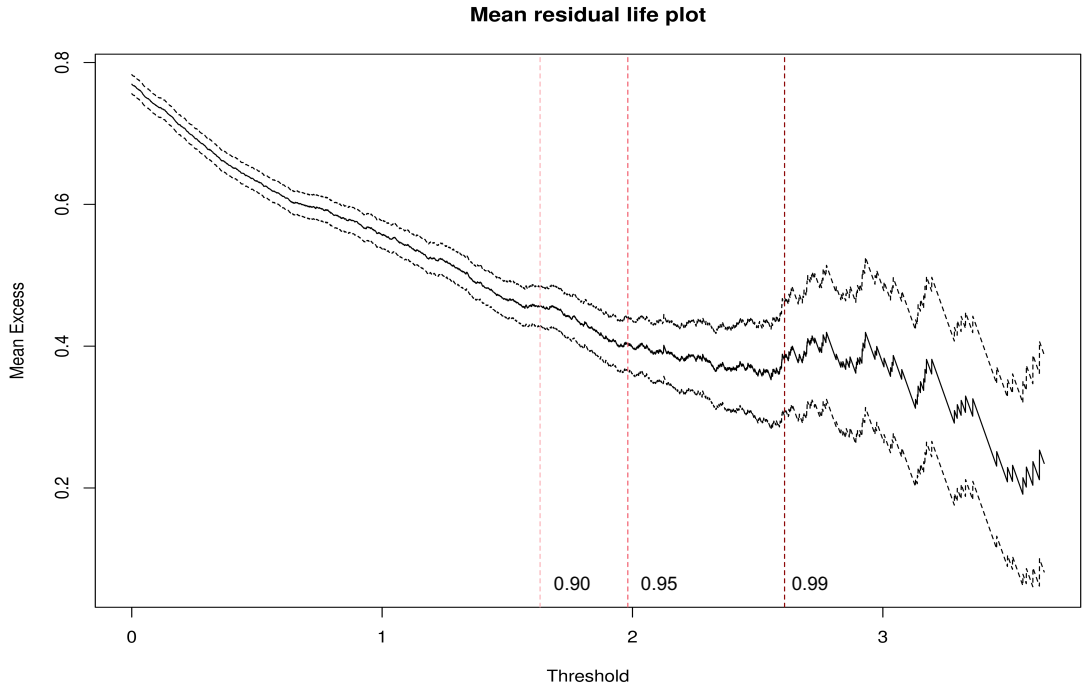


Figure 11 – Showing the mean residual life plot of the WMH IDP data. We can observe that the plot is relatively linear up to the 0.95 threshold. Indicating that this could be a good threshold.

Another graphical tool that can be employed is the parameter or threshold stability plot (44), which involves plotting model parameters as a function of the threshold. We look at different estimates of the parameters for the GDP and decide on an optimal threshold based on the lowest value where the plots are approximately stable. In Figure 12, we show the parameter threshold stability plot of the modified scale and shape parameters. The dots demonstrate the scale and shape parameters for the GPD at different thresholds. The lines show the 95% confidence interval. We can see that the threshold should be around the 0.9 to 0.95 mark.

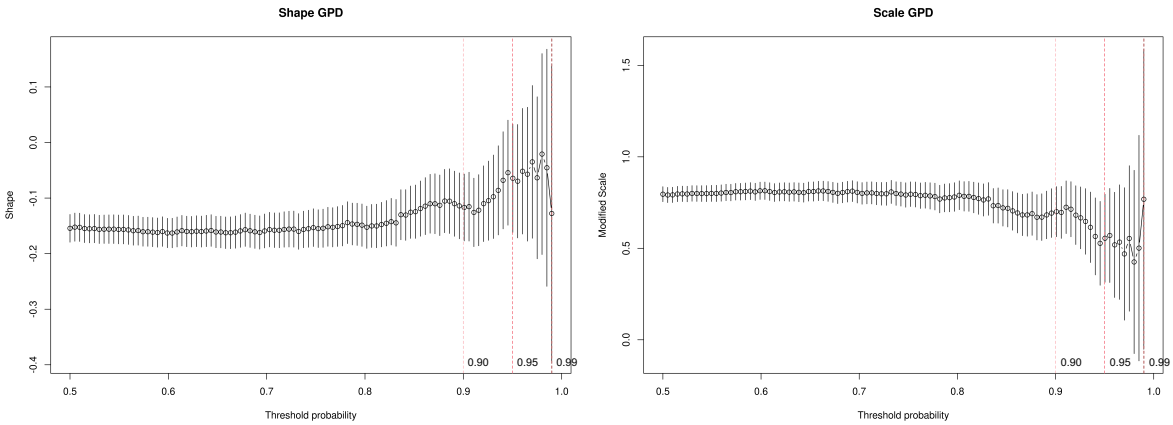


Figure 12 - Showing the parameter threshold stability plot of the modified scale and shape parameters. The dots demonstrate the scale and shape parameters for the generalized Pareto distribution (GPD) at different thresholds. The lines show the 95% confidence interval. An optimal threshold is chosen at the point where the parameters are both relatively stable in this case around the 95th percentile.

Choosing the optimal threshold for the TPDM estimation includes all the same difficulties with selecting a threshold as for univariate extreme value statistics. Commonly, it is selected with a visual inspection of all the graphs presented in the univariate result section. However, when the dimensions become large this becomes infeasible. Therefore, as suggested by (26) we selected a couple of IDPs and manually went through a few thresholds, see supplement Figure 13, Figure 14 and Figure 15. We decided the 0.95 quantile was a good common threshold above which the examined parameters seemed relatively constant.

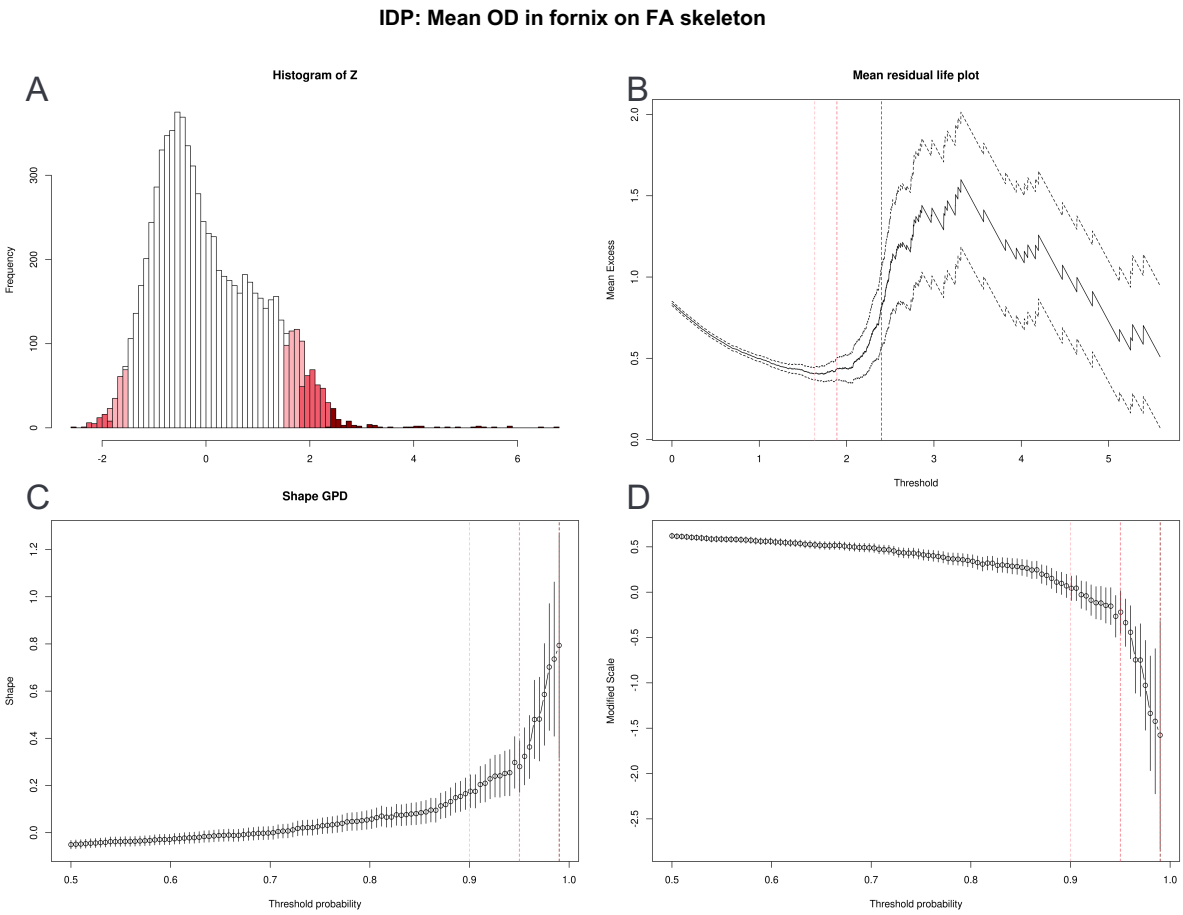


Figure 13 - Displaying various threshold selection results for IDP 25397 - mean OD in fornix on FA skeleton. A. illustrates the histogram of the z-scores, demonstrating different thresholds that could be employed in the peaks over threshold method. B. displays the mean residual life plot, showing a relatively linear plot up to the 0.95 threshold, indicating that this could be a good threshold. C. and D. show the parameter threshold stability plot of the modified scale and shape parameters. The dots demonstrate the scale and shape parameters for the Generalized Pareto distribution (GPD) at different thresholds. The lines show the 95% confidence interval. An optimal threshold is chosen at the point where both parameters are relatively stable, in this case, around the 95th percentile.

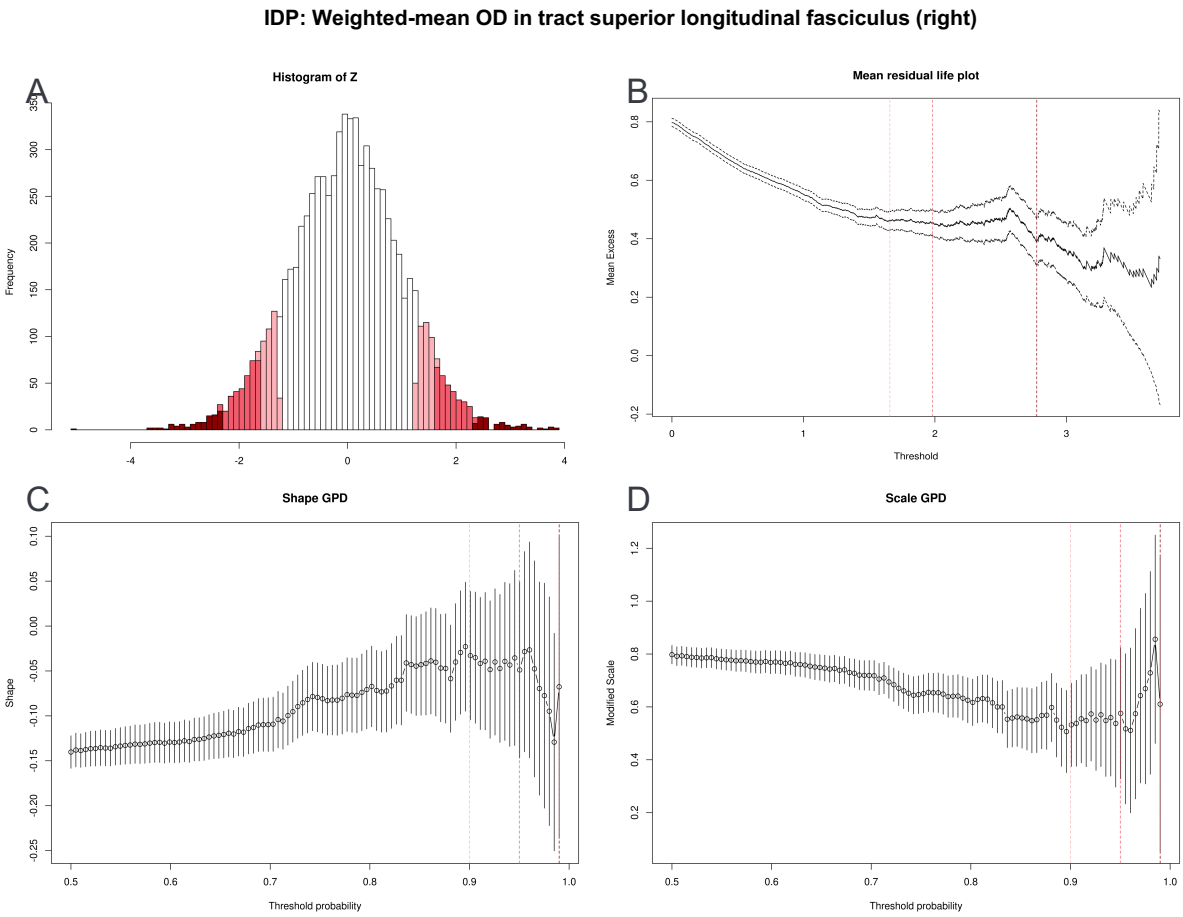


Figure 14 - Displaying various threshold selection results for IDP 25699 - weighted-mean OD in tract superior longitudinal fasciculus (right). A. illustrates the histogram of the z-scores, demonstrating different thresholds that could be employed in the peaks over threshold method. B. displays the mean residual life plot, showing a relatively linear plot up to the 0.95 threshold, indicating that this could be a good threshold. C. and D. show the parameter threshold stability plot of the modified scale and shape parameters. The dots demonstrate the scale and shape parameters for the Generalized Pareto distribution (GPD) at different thresholds. The lines show the 95% confidence interval. An optimal threshold is chosen at the point where both parameters are relatively stable, in this case, around the 95th percentile.

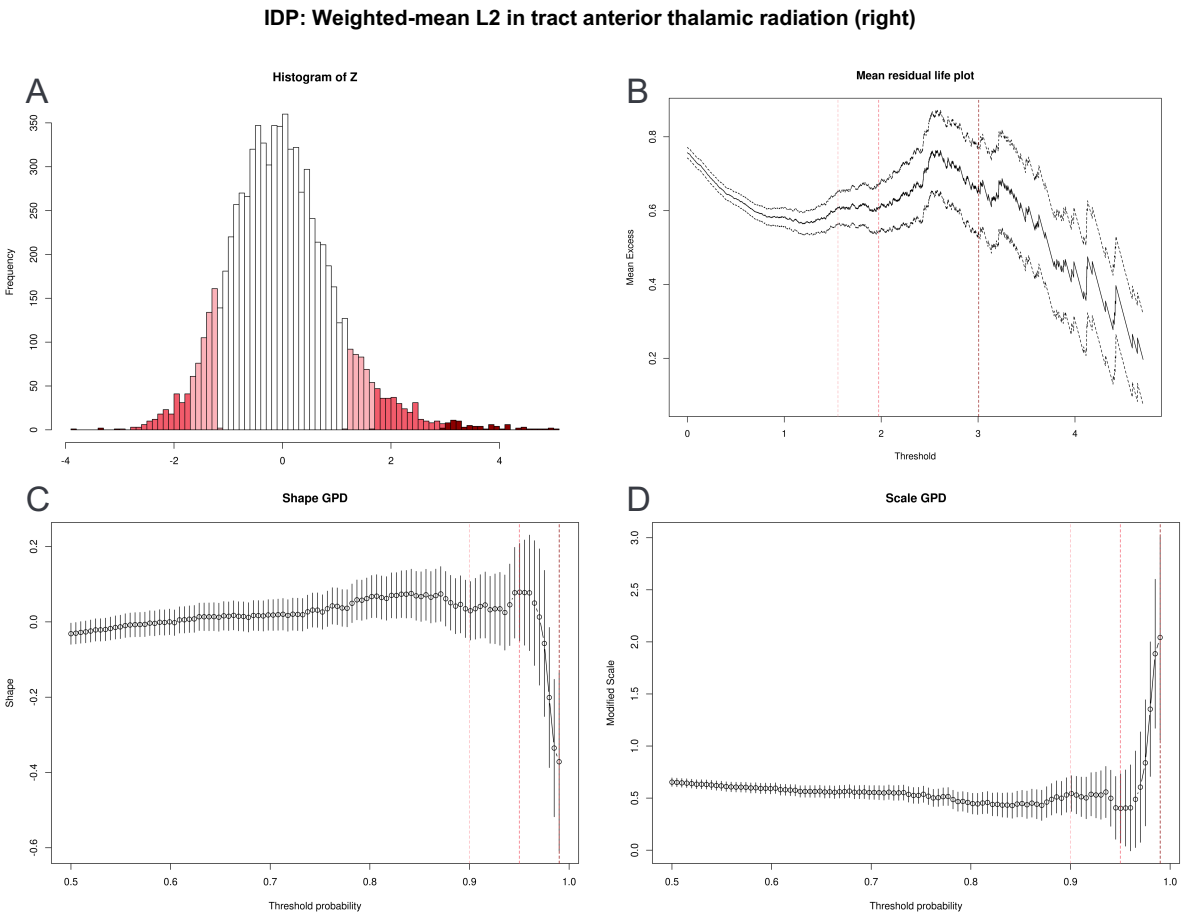


Figure 15 - Displaying various threshold selection results for IDP 25599 - Weighted-mean L2 in tract anterior thalamic radiation (right). A. illustrates the histogram of the z-scores, demonstrating different thresholds that could be employed in the peaks over threshold method. B. displays the mean residual life plot, showing a relatively linear plot up to the 0.95 threshold, indicating that this could be a good threshold. C. and D. show the parameter threshold stability plot of the modified scale and shape parameters. The dots demonstrate the scale and shape parameters for the Generalized Pareto distribution (GPD) at different thresholds. The lines show the 95% confidence interval. An optimal threshold is chosen at the point where both parameters are relatively stable, in this case, around the 95th percentile.

Extreme value theory results

In Figure 16, we show the data from two IDPs in the original space and after the marginal transformations have been applied. Furthermore, we show the histogram of r exceeding the 0.95 quantile threshold. It can be noted that the variables fully lie on their respective axis after the marginal transformation.

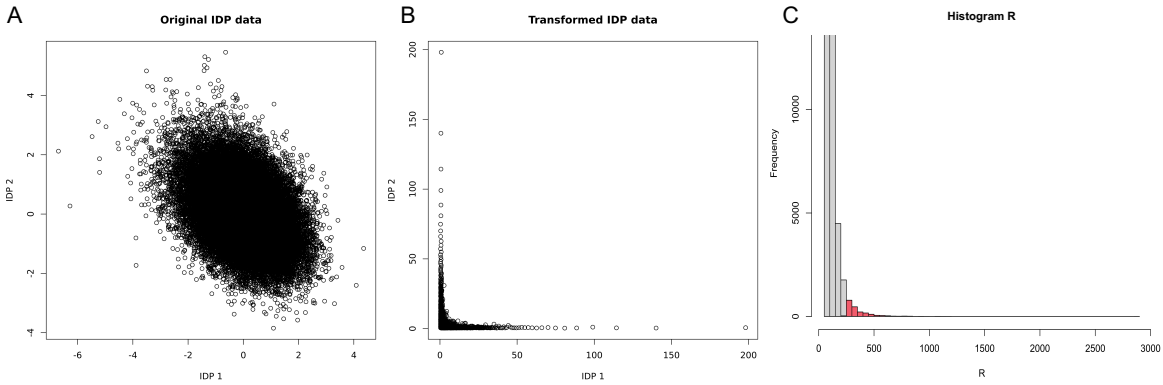


Figure 16 - Illustration of the applied marginal transformation to the IDP data. A. Scatterplot of two IDPs in their original dimensions. B. Scatterplot of the IDP data after transforming the marginal, such that the marginal is regularly varying with α=2. C. Histogram showing the r polar coordinate and the 0.95 largest components.

The $n_{exc} = 1960$ data is then used to estimate the TPDM $\widehat{\Sigma}_{\mathbf{z}}$. In, Figure 17, we give an example of the found TPDM separated per IDP modality.

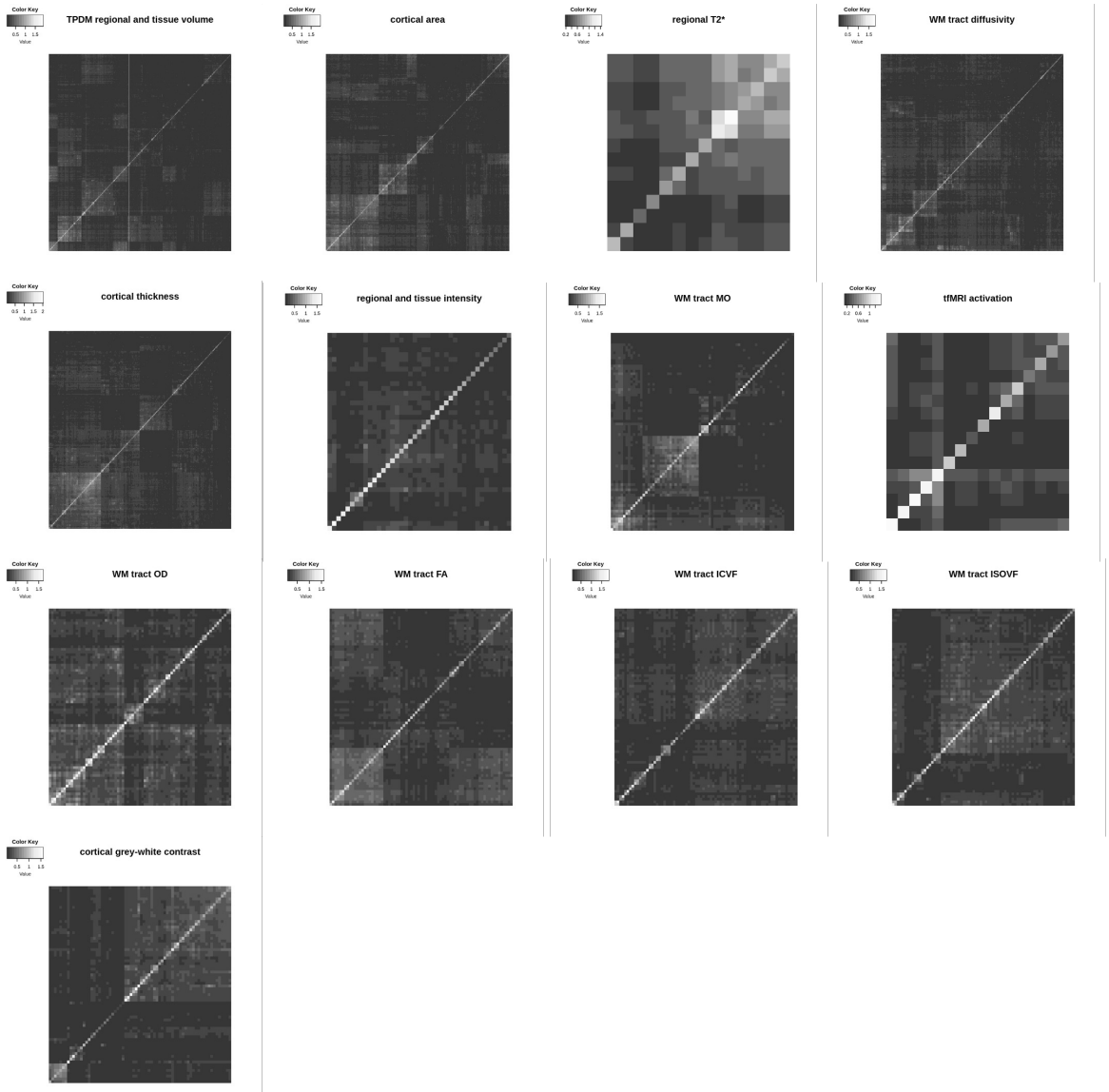
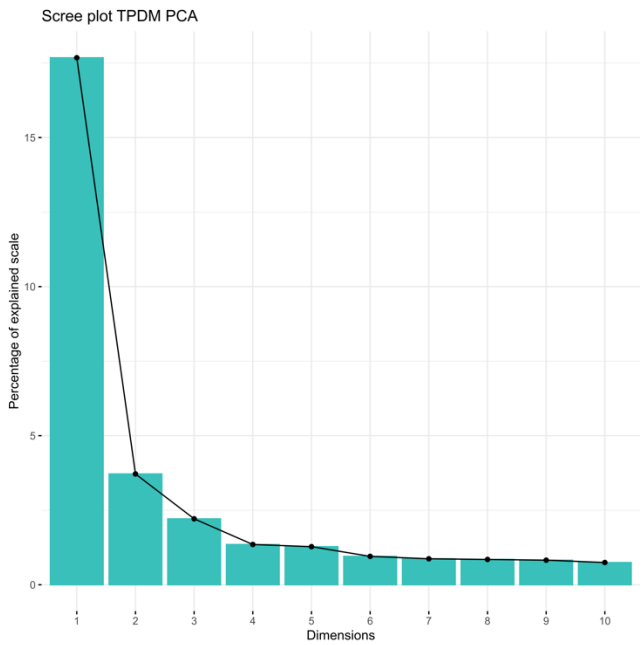


Figure 17 - TPDMs separated per IDP modality.

Interpretation of extreme principal components

We performed a standard eigendecomposition of $\hat{\Sigma}_{\boldsymbol{z}}$ or the TPDM to obtain the eigenvectors (i.e. extreme PCs). Similar to covariance-PCA, we can look at the amount of scale that the first number of eigenvalues explain. In Figure 18, we plotted the scree plot of the scale, which shows that most of the scale is explained by the first two eigenvectors. In Figure 19, we show the correlation circle for the first two extreme PCs.



	eigenvalue	scale percent	cumulative scale percent
Dim.1	359.73	17.68	17.68
Dim.2	75.63	3.72	21.39
Dim.3	45.02	2.21	23.61
Dim.4	27.48	1.35	24.96
Dim.5	26.01	1.28	26.23
Dim.6	19.38	0.95	27.19
Dim.7	17.73	0.87	28.06
Dim.8	17.27	0.85	28.91
Dim.9	16.79	0.83	29.73
Dim.10	15.19	0.75	30.48
Dim.11	13.03	0.64	31.12
Dim.12	12.54	0.62	31.74
Dim.13	11.42	0.56	32.3
Dim.14	10.64	0.52	32.82
Dim.15	10.22	0.5	33.32
Dim.16	9.85	0.48	33.81
Dim.17	9.47	0.47	34.27
Dim.18	9.22	0.45	34.72
Dim.19	9.11	0.45	35.17
Dim.20	8.95	0.44	35.61

Figure 18 - Showing the scree plot and table of scale values of the eigenvalue decomposition for the TPDM matrix. We can see that a large percentage of the scale is explained by the first extreme component.

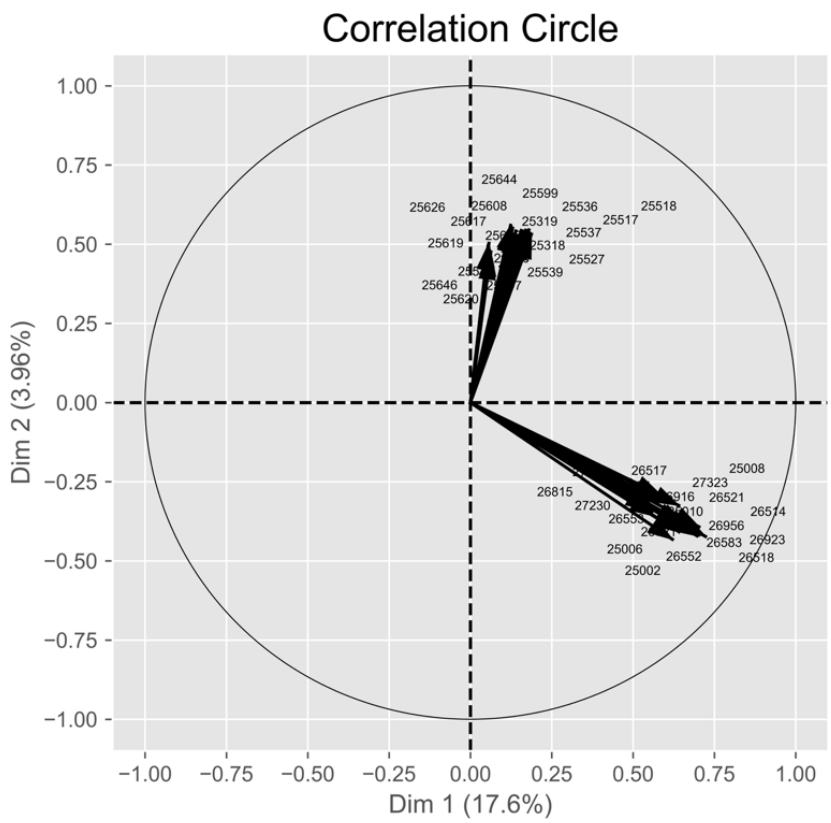


Figure 19 – Showing the correlation circle with the loadings of the top 40 contributing variables on the extreme principal components 1 and 2.

Removed IDPs

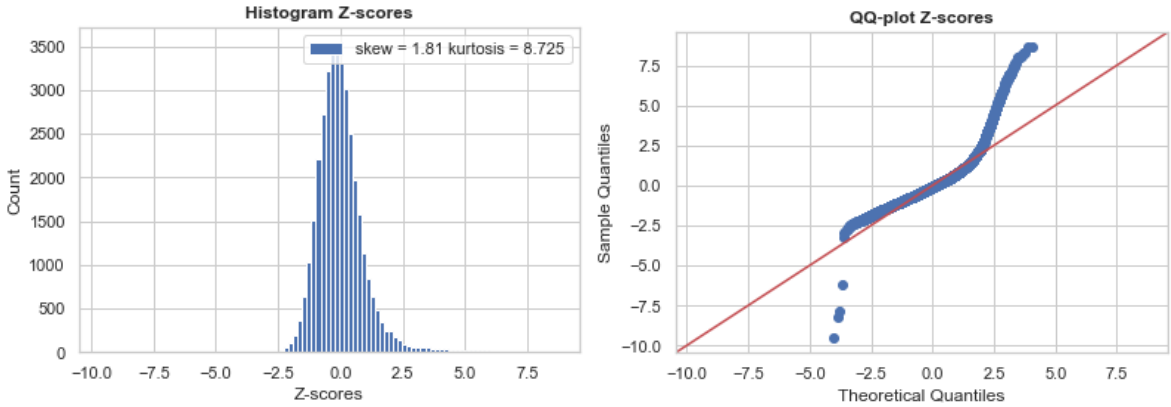


Figure 20 - Showing an example of a badly fitted IDP with a high kurtosis. In the analysis, we have removed these IDPs from further processing.

References

1. Zhao X, Scarrott C, Oxley L, Reale M. Extreme value modelling for forecasting market crisis impacts. Applied Financial Economics. 2010 Jan 1;20:63–72.

2. Poon SH, Rockinger M, Tawn J. Extreme Value Dependence in Financial Markets: Diagnostics, Models, and Financial Implications. The Review of Financial Studies. 2004;17(2):581–610.

3. Marsh TJ, Dale M. The UK Floods of 2000–2001: A Hydrometeorological Appraisal. Water and Environment Journal. 2002;16(3):180–8.

4. Zabihi M, Oldehinkel M, Wolfers T, Frouin V, Goyard D, Loth E, et al. Dissecting the Heterogeneous Cortical Anatomy of Autism Spectrum Disorder Using Normative Models. Biol Psychiatry Cogn Neurosci Neuroimaging. 2019 Jun;4(6):567–78.

5. Wolfers T, Doan NT, Kaufmann T, Alnæs D, Moberget T, Agartz I, et al. Mapping the Heterogeneous Phenotype of Schizophrenia and Bipolar Disorder Using Normative Models. JAMA Psychiatry. 2018 Nov 1;75(11):1146–55.

6. Rutherford S, Fraza C, Dinga R, Kia SM, Wolfers T, Zabihi M, et al. Charting brain growth and aging at high spatial precision. Baker CI, Taschler B, Esteban O, Constable T, editors. eLife. 2022 Feb 1;11:e72904.

7. Insel TR. Mental disorders in childhood: shifting the focus from behavioral symptoms to neurodevelopmental trajectories. JAMA. 2014 May 7;311(17):1727–8.

8. Marquand AF, Kia SM, Zabihi M, Wolfers T, Buitelaar JK, Beckmann CF. Conceptualizing mental disorders as deviations from normative functioning. Mol Psychiatry. 2019;24(10):1415–24.

9. Shaw P, Eckstrand K, Sharp W, Blumenthal J, Lerch JP, Greenstein D, et al. Attention-deficit/hyperactivity disorder is characterized by a delay in cortical maturation. Proc Natl Acad Sci U S A. 2007 Dec 4;104(49):19649–54.

10. Allsopp K, Read J, Corcoran R, Kinderman P. Heterogeneity in psychiatric diagnostic classification. Psychiatry Res. 2019 Sep;279:15–22.

11. Feczko E, Fair DA. Methods and Challenges for Assessing Heterogeneity. Biological Psychiatry. 2020 Jul 1;88(1):9–17.

12. Goix N, Sabourin A, Clémençon S. Sparse representation of multivariate extremes with applications to anomaly detection. Journal of Multivariate Analysis. 2017 Sep 1;161:12–31.

13. Davison AC, Huser R. Statistics of Extremes [Internet]. Rochester, NY; 2015 [cited 2023 Mar 11]. Available from: https://papers.ssrn.com/abstract=2598025

14. Marquand AF, Rezek I, Buitelaar J, Beckmann CF. Understanding Heterogeneity in Clinical Cohorts Using Normative Models: Beyond Case-Control Studies. Biol Psychiatry. 2016 Oct 1;80(7):552–61.

15. Marquand AF, Wolfers T, Mennes M, Buitelaar J, Beckmann CF. Beyond Lumping and Splitting: A Review of Computational Approaches for Stratifying Psychiatric Disorders. Biol Psychiatry Cogn Neurosci Neuroimaging. 2016 Sep;1(5):433–47.

16. Wolfers T, Beckmann CF, Hoogman M, Buitelaar JK, Franke B, Marquand AF. Individual differences v. the average patient: mapping the heterogeneity in ADHD using normative models. Psychol Med. 2020 Jan;50(2):314–23.

17. Lv J, Di Biase M, Cash RFH, Cocchi L, Cropley VL, Klauser P, et al. Individual deviations from normative models of brain structure in a large cross-sectional schizophrenia cohort. Mol Psychiatry. 2021 Jul;26(7):3512–23.

18. Cannon TD, Keller MC. Endophenotypes in the genetic analyses of mental disorders. Annu Rev Clin Psychol. 2006;2:267–90.

19. Cannon TD, Gasperoni TL, van Erp TG, Rosso IM. Quantitative neural indicators of liability to schizophrenia: implications for molecular genetic studies. Am J Med Genet. 2001 Jan 8;105(1):16–9.

20. Asadi P, Davison AC, Engelke S. Extremes on river networks. Ann Appl Stat [Internet]. 2015 Dec 1 [cited 2022 Feb 16];9(4). Available from: http://arxiv.org/abs/1501.02663

21. Chiapino M, Sabourin A. Feature Clustering for Extreme Events Analysis, with Application to Extreme Stream-Flow Data. In: Appice A, Ceci M, Loglisci C, Masciari E, Raś ZW, editors. New Frontiers in Mining Complex Patterns. Cham: Springer International Publishing; 2017. p. 132–47. (Lecture Notes in Computer Science).

22. Engelke S, Ivanovs J. Sparse Structures for Multivariate Extremes. arXiv:200412182 [math, stat] [Internet]. 2020 May 8 [cited 2022 Feb 15]; Available from: http://arxiv.org/abs/2004.12182

23. Coles, S., Bawa, J., Trenner, L., & Dorazio, P. An Introduction to Statistical Modeling of Extreme Values [Internet]. Londen: Springer; 2001 [cited 2022 May 20]. Available from: http://link.springer.com/book/10.1007/978-1-4471-3675-0

24. Martin EI, Ressler KJ, Binder E, Nemeroff CB. The Neurobiology of Anxiety Disorders: Brain Imaging, Genetics, and Psychoneuroendocrinology. Psychiatr Clin North Am. 2009 Sep;32(3):549–75.

25. Cooley D, Thibaud E. Decompositions of Dependence for High-Dimensional Extremes. arXiv:161207190 [stat] [Internet]. 2018 Apr 25 [cited 2022 Mar 1]; Available from: http://arxiv.org/abs/1612.07190

26. Jiang Y, Cooley D, Wehner MF. Principal Component Analysis for Extremes and Application to U.S. Precipitation. Journal of Climate. 2020 Aug 1;33(15):6441–51.

27. De Haan, L., Ferreira, A., & Ferreira, A. Extreme value theory: an introduction. New York: Springer.; 2006.

28. Miller KL, Alfaro-Almagro F, Bangerter NK, Thomas DL, Yacoub E, Xu J, et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. Nature Neuroscience. 2016;19(11):1523–36.

29. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLOS Medicine. 2015 Mar 31;12(3):e1001779.

30. Alfaro-Almagro F, Jenkinson M, Bangerter NK, Andersson JLR, Griffanti L, Douaud G, et al. Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. NeuroImage. 2018 Feb 1;166:400–24.

31. Smith S, Fidel Alfro A, Miller K. UK Biobank Brain Imaging Documentation. [Internet]. 2020. Available from: biobank.ctsu.ox.ac.uk/crystal/docs/brain_mri.pdf

32. McCarthy P. Funpack. 2020.

33. Fraza CJ, Dinga R, Beckmann CF, Marquand AF. Warped Bayesian linear regression for normative modelling of big data. NeuroImage. 2021 Dec 15;245:118715.

34. Resnick, S. I. Heavy-tail phenomena: probabilistic and statistical modeling. Springer Science & Business Media; 2007.

35. Abdi H, Williams LJ. Principal component analysis: Principal component analysis. WIREs Comp Stat. 2010 Jul;2(4):433–59.

36. Buch AM, Liston C. Dissecting diagnostic heterogeneity in depression by integrating neuroimaging and genetics. Neuropsychopharmacology. 2021 Jan;46(1):156–75.

37. Nandi A, Beard JR, Galea S. Epidemiologic heterogeneity of common mood and anxiety disorders over the lifecourse in the general population: a systematic review. BMC Psychiatry. 2009 Jun 1;9(1):31.

38. Fraza CJ, Dinga R, Beckmann CF, Marquand AF. Warped Bayesian linear regression for normative modelling of big data. Neuroimage. 2021 Dec 15;245:118715.

39. Dinga R, Fraza CJ, Bayer JMM, Kia SM, Beckmann CF, Marquand AF. Normative modeling of neuroimaging data using generalized additive models of location scale and shape [Internet]. bioRxiv; 2021 [cited 2022 May 20]. p. 2021.06.14.448106. Available from: https://www.biorxiv.org/content/10.1101/2021.06.14.448106v1

40. Kia SM, Huijsdens H, Rutherford S, Dinga R, Wolfers T, Mennes M, et al. Federated Multi-Site Normative Modeling using Hierarchical Bayesian Regression. bioRxiv. 2021 May 30;2021.05.28.446120.

41. Fisher RA, Tippett LHC. Limiting forms of the frequency distribution of the largest or smallest member of a sample. Mathematical Proceedings of the Cambridge Philosophical Society. 1928 Apr;24(2):180–90.

42. Gnedenko B. Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire. Annals of Mathematics. 1943;44(3):423–53.

43. Davison AC, Smith RL. Models for Exceedances over High Thresholds. Journal of the Royal Statistical Society Series B (Methodological). 1990;52(3):393–442.

44. Scarrott C, MacDonald A. A review of extreme value threshold estimation and uncertainty quantification. Revstat Statistical Journal. 2012 Mar 1;10:33–60.