# An automatic approach to find all possible calendar effects (periodic patterns) for an univariate time series

## Contents

# 1 Introduction

Exploratory data analysis, as coined by John W. Tukey (Tukey 1965) involves many iterations of finding structures and patterns that allows the data to be informative. With temporal data available at finer scales, exploring periodicity and their relationships can become overwhelming with so many possible cyclic temporal granularities (Gupta et al. 2020) to explore.

Take the example of the calendar display of electricity smart meter data (**??**) used in Wang, Cook, and Hyndman (2020) for four households in Melbourne, Australia. The authors show how hour-of-the-day interact with weekday and weekends and then move on to use calendar display to show daily schedules. The calendar display has several components in it, which helps us look at energy consumption across hour-of-the-day, day-of-the-week, week-of-the-month, and month-of-the-year at once. Some interaction of these cyclic granularities could also be interpreted from this display. This is a great start to have an overview of the energy consumption. However, if one wants to understand the periodicities in energy behavior and how the periodicities interact in greater details, it is not easy to comprehend the interactions of some periodicities' from this display, due to the combination of linear and cyclic representation of time. For example, this display might not be the best to understand how hour-of-the-day varies and month-of-year varies across week-of-the-month. Further, it is not clear what all interactions of cyclic granularities should be read from this display as there could be many combinations that one can look at. Moreover, calendar effects are not restricted to conventional day-of-week or month-of-year deconstructions (Gupta et al. (2020)) and could include other cyclic granularities like hour-of-week or day-of-fortnight, which could potentially become useful depending on the context.

Moreover, there might be specific interactions that are interesting and others that are not and that too will vary with different households. For example, area distribution quantiles are plotted for household 2 and 4 in Figure 2a and b respectively. For the first household, the 75th and 90th percentile for Jan, Feb and July are very close, implying that energy usage for these months are generally on a much higher side due to the usage of air conditioners (in Jan and Feb) and heaters (in July). The energy consumption for household 2 is also higher relative to its own consumption for Jan, Feb and March but the 75th and 90th percentile are apart implying that contrary to the first household, the second household resorts to air conditioners and heaters much less regularly than the first one. Moreover, the 75th percentile distribution is not bimodal across hours of the day for the first household in those months, but the distribution looks similar for all months for the second household. Difference in the energy consumption seem to be varying both across month-of-year (facets) and hour-of-day (x-axis). And thus, both the cyclic granularities would deem important while studying the periodicities in the first household. However, it seems like energy consumption across hours of the day are not that different across different months for the second household. Differences seem to be more prominent across month-of-year (facets) than hour-of-day (x-axis). Again, look at **??** c and d, where energy consumption for these two households are plotted against (weekend/weekday, week-of-month). Here, for both households, the pattern of energy consumption vary across different weeks of the month irrespective of the fact it is a weekday or weekend. In that respect, the harmony pair (month-of-year, hour-of-day) seems to be more informative than (weekend/weekday, week-of-month) for the first household. It could be immensely useful to make the transition from all possible ways to only ways that could potentially be informative given a household.

The paper Gupta et al. (2020) describes how we can compute all possible combinations of cyclic time granularities. If we have $n$ periodic linear granularities in the hierarchy table, then $n(n-1)/2$ circular or quasi-circular cyclic granularities could be constructed. Let $N_C$ be the total number of contextual circular, quasi-circular and aperiodic cyclic granularities that can originate from the underlying periodic and aperiodic linear granularities. The mapping of the graphical elements chosen in the paper implies that, for a numeric response variable, the graphics display distributions across combinations of cyclic granularities, one placed at x-axis and the other on the facet. That essentially implies there are ${}^{N_C}P_2$ possible pairwise plots exhaustively, where each plot would display a pair of cyclic granularities. This is large and overwhelming for human consumption.

This is similar to Scagnostics (Scatterplot Diagnostics) by Tukey and Tukey (1988), which is used to discern meaningful patterns in large collections of scatterplots. Given a set of $v$ variables, there are $v(v-1)/2$ pairs

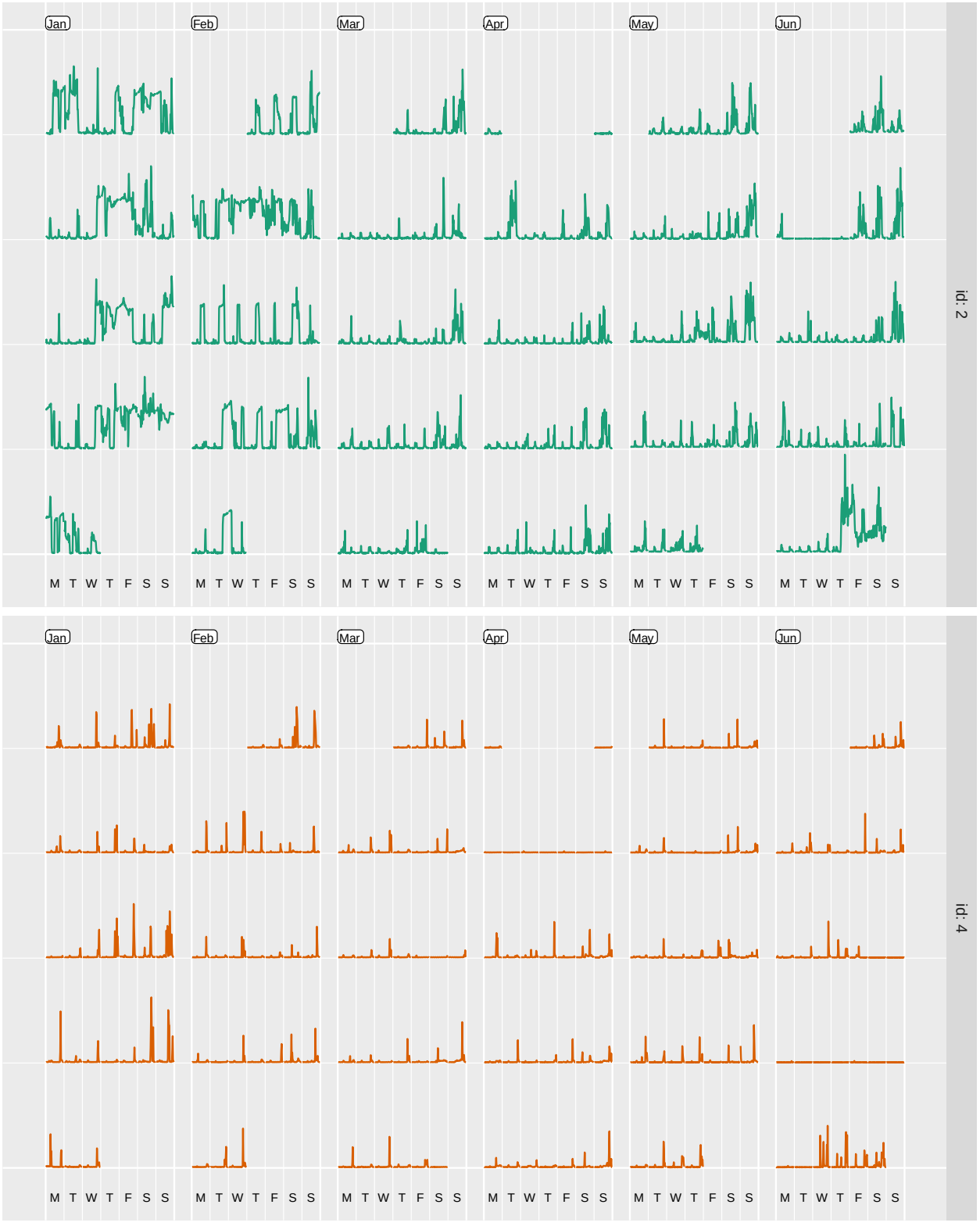



Figure 1: Calendar display.

(#fig:calendar- <- )

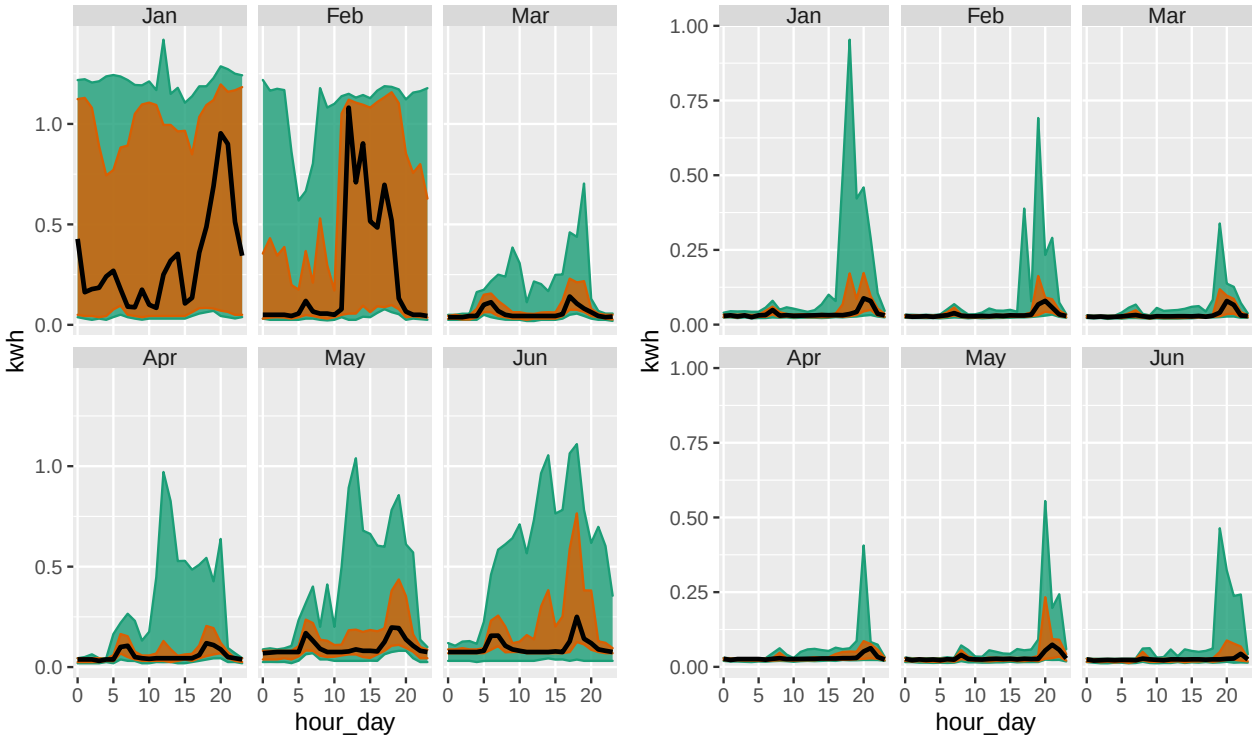



Figure 2: something

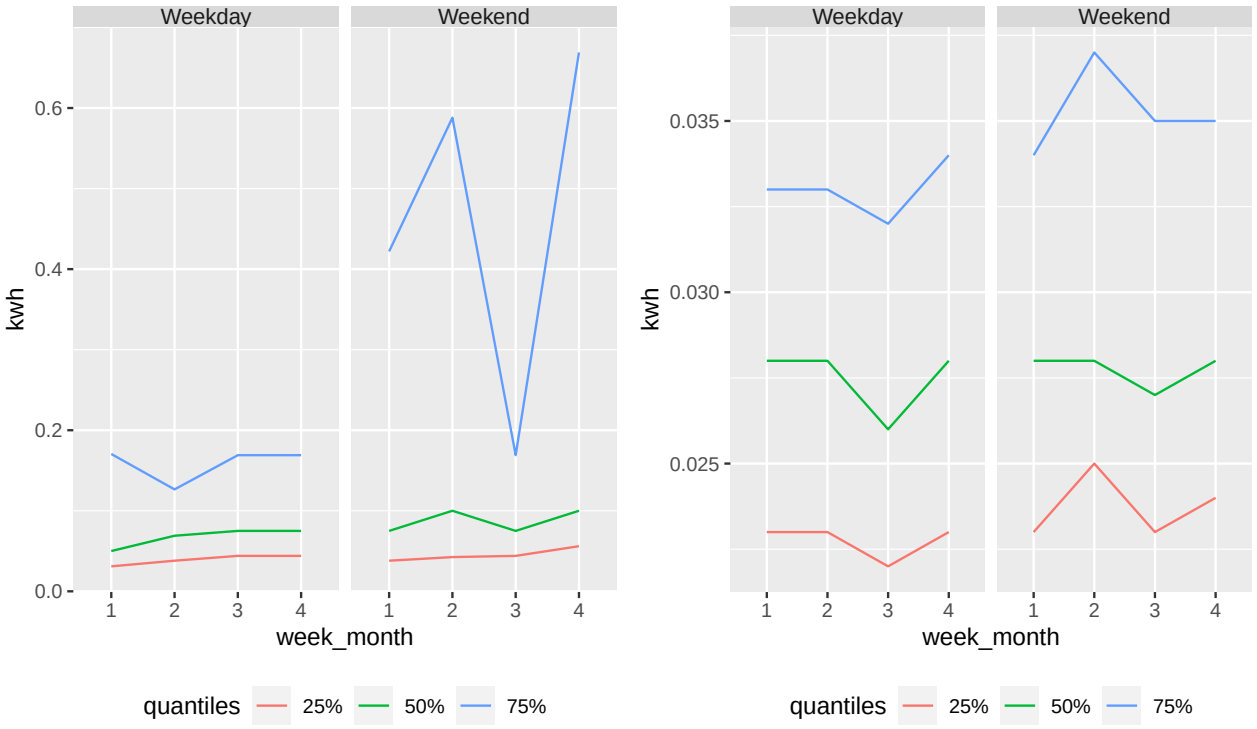



Figure 3: something2

of variables, and thus the same number of possible pairwise scatterplots. Therefore for even small $v$, the number of scatterplots can be large, and scatterplot matrices (SPLOMs) could easily run out of pixels when presenting high-dimensional data. Dang and Wilkinson (2014) and Wilkinson, Anand, and Grossman (2005) provides potential solutions to this, where few characterizations help us to locate anomalies for defining several measures aimed to detect anomalies in density, shape, trend, and other features in the 2D point scatters.

The paper (Gupta et al. (2020)) narrows down the search from ${}^{N_C}P_2$ plots by identifying pairs of granularities that can be meaningfully examined together (a "harmony"), or when they cannot (a "clash"). However, even after excluding clashes, the list of harmonies left could be enormous for exhaustive exploration. Hence, there is a need to reduce the search even further by including only those harmonies which are informative enough. Also, ranking the remaining harmony pairs based on how well they capture the variation in the measured variable could be potentially useful.

In this paper, we aim to build a new measure to follow through these two main objectives:

- To choose harmonies for which distributions of categories are significantly different
- To rank the selected harmonies from highest to lowest variation in the distribution of their categories.

# 2 The distance measure for quantifying patterns in harmonies

We are interested in assessing structure in probability distributions of the measured variable across bivariate cyclic granularities. We propose a measure called Weighted Maximum Pairwise Distances (wpd) to evaluate structure in such a design.

## 2.1 Idea

The principle employed for building a new metric is explained through a simple example explained in Figure **??**. Each of these figures have the same panel design with 2 x-axis categories and 4 facet levels. Figure **??**a has all x categories drawn from N(5, 10) distribution for each facet. It is not an interesting display particularly, as distributions do not vary across x-axis or facet categories. Figure **??**b has x categories drawn from the same distribution within a facet and different for different facet categories. Figure **??**b exhibits an exact opposite situation where distribution between the x-axis categories within each facet is different but they are same across facets. Figure **??**d takes a step further by varying the distribution across both facet and x-axis categories. If we are asked to rank the displays in order of importance from minimum to maximum, we might order it as a, b, c and then d. It might be argued that it is not clear if b should precede or succeed c. Gestalt theory suggests that when items are placed in close proximity, people assume that they are in the same group because they are close to one another and apart from other groups. Hence, displays that capture more variation within different categories in the same group would be important to bring out different patterns of the data. With this principle, display b could be considered less informative as compared to display c.

With reference to the graphical design in **??**, therefore the idea would be to rate a harmony pair higher if the variation between different levels of the x-axis variable is higher on an average across all levels of the facet variables. Thus the metric could be obtained by computing maximum pairwise distances between distributions of the continuous random variable across x-axis categories for all facets and then taking the median of those maximum pairwise distances across facets. This would help capture the average maximum difference in distribution of the measurement variable explained by the two cyclic granularities together. We call this metric wpd which stands for Median Maximum Pairwise Distances. In the next section we shall see how we go about computing this measure.

## 2.2 Characterising distributions

Each of the data subsets in the data structure have multiple observations and may vary widely across different subsets due to the structure of the calendar, missing observations or uneven locations of events in the time domain. The set of observations corresponding to each combination is assumed to be a sample from an unknown probability density function. While the whole population of observations has certain characteristics, we can typically never measure all of them. Often shape, central tendency, and variability are the common characteristics used to describe the distribution. Another way to describe the probability distribution is through quantiles. (Define quantiles here) Sample quantiles could be thought to estimate the population quantiles. But there are a large number of different definitions used for sample quantiles. The median-unbiased estimator is recommended (Rob's paper) because of its desirable properties of a quantile estimator and can be defined independently of the underlying distribution.

## 2.3 Distance between distributions

The most common divergence measure between distributions is the Kullback-Leibler (KL) divergence(Kullback and Leibler 1951) introduced by Solomon Kullback and Richard Leibler in 1951. The KL divergence, denoted $D(p(x), q(x))$ is a non-symmetric measure of the difference between two probability distributions $p(x)$ and $q(x)$ and is interpreted as the amount of information lost when $q(x)$ is used to approximate $p(x)$. Although the KL divergence measures the "distance" between two distributions, it is not a distance measure since it is not symmetric and does not satisfy the triangle inequality. The Jensen-Shannon divergence (Menéndez et al. 1997) based on the Kullback-Leibler divergence is symmetric and it always has a finite value. The square root of the Jensen-Shannon divergence is a metric, often referred to as Jensen-Shannon distance. Other common measures of distance are Hellinger distance, total variation distance and Fisher information metric.

In the context of this paper, the pairwise distances between the distributions of the measured variable are computed through Jensen-Shannon distance (JSD) which is based on Kullback-Leibler divergence and is defined by,

$$JSD(P||Q) = \frac{1}{2}D(P||M) + \frac{1}{2}D(Q||M)$$

where $M = \frac{P+Q}{2}$ and $D(P||Q) := \int_{-\infty}^{\infty} p(x) f(\frac{p(x)}{q(x)})$ is the KL divergence between distributions $p(x)$ and $q(x)$. Probability distributions are estimated through quantiles instead of kernel density so that there is minimal dependency on selecting kernel or bandwidth.

## 2.4 Definition

Consider two cyclic granularities $A$ and $B$, such that $A = \{a_j : j = 1, 2, \ldots, J\}$ and $B = \{b_k : k = 1, 2, \ldots, K\}$ with $A$ placed across x-axis and $B$ across facets. Let the pairwise distances between pairs $(a_j b_k, a_{j'} b_{k'})$ be denoted as $d_{(jk,j'k')} = JSD(a_j b_k, a_{j'} b_{k'})$. Pairwise distances could be within-facets or between-facets. Figure 4 illustrates how the within-facet or between-facet distances are defined. Pairwise distances are within-facets $(d_w)$ when $b_k = b_{k'}$, that is, between pairs of the form $(a_j b_k, a_{j'} b_k)$ as shown in panel (3) of Figure 4. If categories are ordered (like all temporal cyclic granularities), then only distances between pairs where $a_{j'} = (a_{j+1})$ are considered (panel (4)). Pairwise distances are between-facets $(d_b)$ when they are considered between pairs of the form $(a_j b_k, a_j b_{k'})$.

From Section 2.1, the idea is to put more weights on within-facet distances than between-facet distances. Hence, for a suitable tuning parameter $0 < \lambda < 1$, the pairwise distances $d_{(jk,j'k')}$ are transformed based on the distance type as follows:

$$d*_{(j,k),(j'k')} = \begin{cases} \lambda d_{(jk),(j'k')}, & \text{if } d = d_w \\ (1-\lambda) d_{(jk),(j'k')}, & \text{if } d = d_b \end{cases} \tag{1}$$

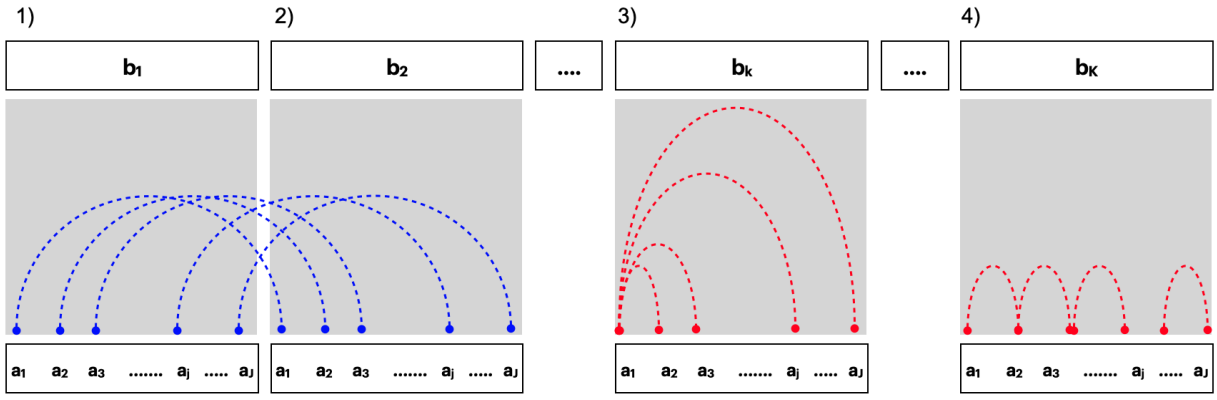



Figure 4: Within and between-facet distances shown for two cyclic granularities A and B, where A is mapped to x-axis and B is mapped to facets. The dotted lines represent the distances between different categories. Panel 1) and 2) show the between-facet distances. Panel 3) and 4) are used to illustrate within-facet distances when categories are un-ordered or ordered respectively. When categories are ordered, distances should only be considered for consecutive x-axis categories. Between-facet distances are distances between different facet levels for the same x-axis category, for example, distances between ($a_1$,$b_1$) and ($a_1$, $b_2$) or ($a_1$,$b_1$) and ($a_1$, $b_3$).

The maximum weighted pairwise distances are defined as:

$$wpd = max_{j,j',k,k'}(d*_{(jk),(j'k')}) \forall j, j' \in \{1, 2, \ldots, J\}, k, k' \in \{1, 2, \ldots, K\}$$

.

## 2.5 Properties

A simulation study is carried out to explore how *wpd* performs under various designs, its parameters and limitations. To this end, simulations were carried out for four different designs and the following factors that could potentially have an impact on the values of *wpd*:

- $nx$ (number of levels of x-axis)
- $nfacet$ (number of levels of facets)
- $\lambda$ (tuning parameter)
- $\omega$ (increment in each panel design)
- $dist$ (normal/non-normal distributions with different location and scale)
- $n$ (sample size for each combination of categories)
- $nsim$ (number of simulations)
- $nperm$ (number of permutations of data)

- *designs*
  $D_{null}$ (No difference in distribution)
  $D_{var_f}$ (Difference in distribution only across facets)
  $D_{var_x}$ (Difference in distribution only across x-axis)
  $D_{var_{all}}$ (Difference in distribution in both facets and x-axis)

Results are presented in two parts. The dependence of *wpd* on $nx$ and $nfacet$ under $D_{null}$ is presented here, which lays the foundation for the next section. The rest of the results that discusses the relationship of the *wpd* with other factors is presented in the Supplementary section of the paper.

### 2.5.1 Design of the simulation study

Observations are generated from a Gamma(2,1) distribution for each combination of $nx$ and $nfacet$ from the following sets: $nx = nfacet = \{2, 3, 5, 7, 14, 20, 31, 50\}$ to cover a wide range of levels from very low to moderately high. Each combination is being referred to as a *panel*. That is, data is being generated for each of the panels $\{nx = 2, nfacet = 2\}, \{nx = 2, nfacet = 3\}, \{nx = 2, nfacet = 5\}, \ldots, \{nx = 50, nfacet = 31\}, \{nx = 50, nfacet = 50\}$. For each of the 64 panels, $ntimes = 500$ observations are drawn for each combination of the categories. That is, if we consider the panel $\{nx = 2, nfacet = 2\}$, 500 observations are generated for each of the combination of categories from the panel, namely, $\{(1, 1), (1, 2), (2, 1), (2, 2)\}$. The values of $\lambda$ is set to 0.67 and values of raw wpd $wpd$ is obtained. This entire design applies to the null cases and hence there is no difference in distribution between any categories in the panel.

### 2.5.2 Results

Figure 5 shows the distribution of $wpd$ plotted across different nx and nfacet categories. Both shape and scale of the distributions change across panels. This is not desirable as it would mean we would not be able to compare $wpd$ across different $nx$ and $nfacet$ as each of them are drawn from distributions with different locations and scale. In Figure 6, we see how the median of $wpd_{raw}$ varies with the total number of distances $nx * nfacet$ for each panel. The median increases abruptly for lower values of $nx * nfacet$ and slowly for higher $nx * nfacet$.

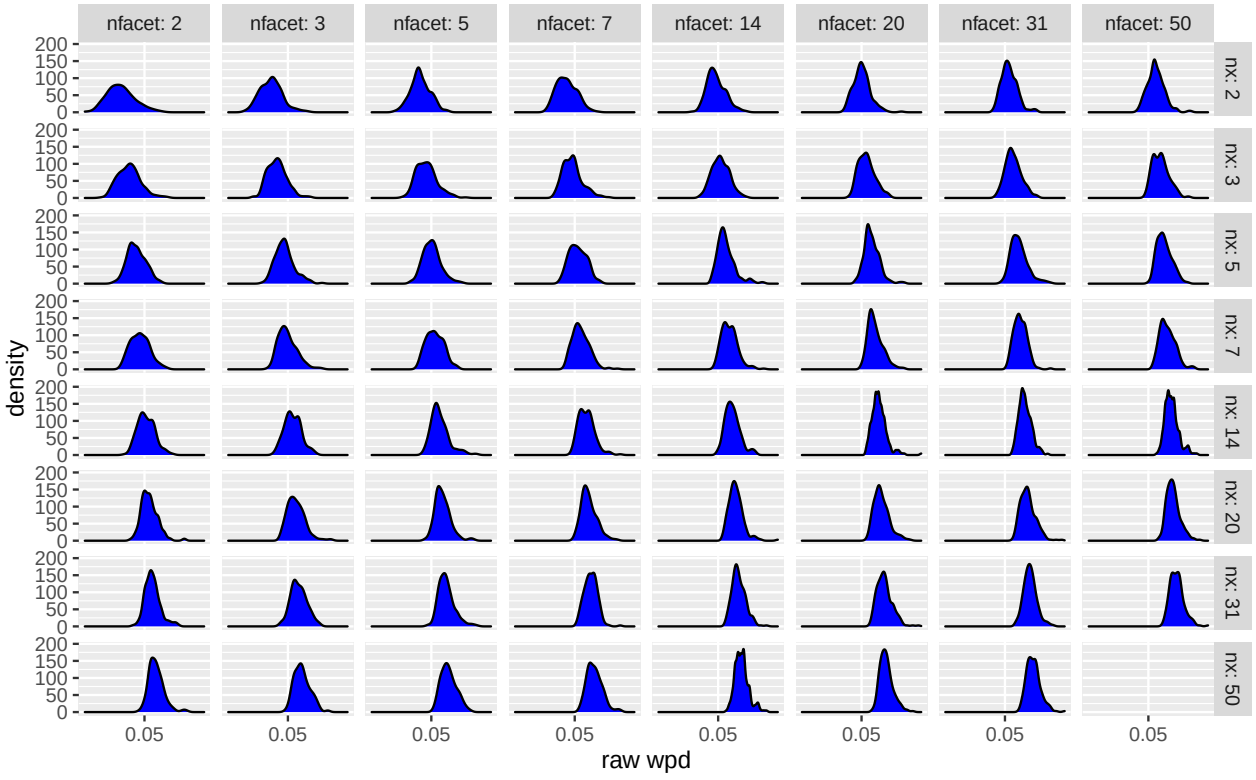



Figure 5: Distribution of raw wpd is plotted across different nx and nfacet categories. Both shape and scale of the distribution changes for different nx and nfacet categories.

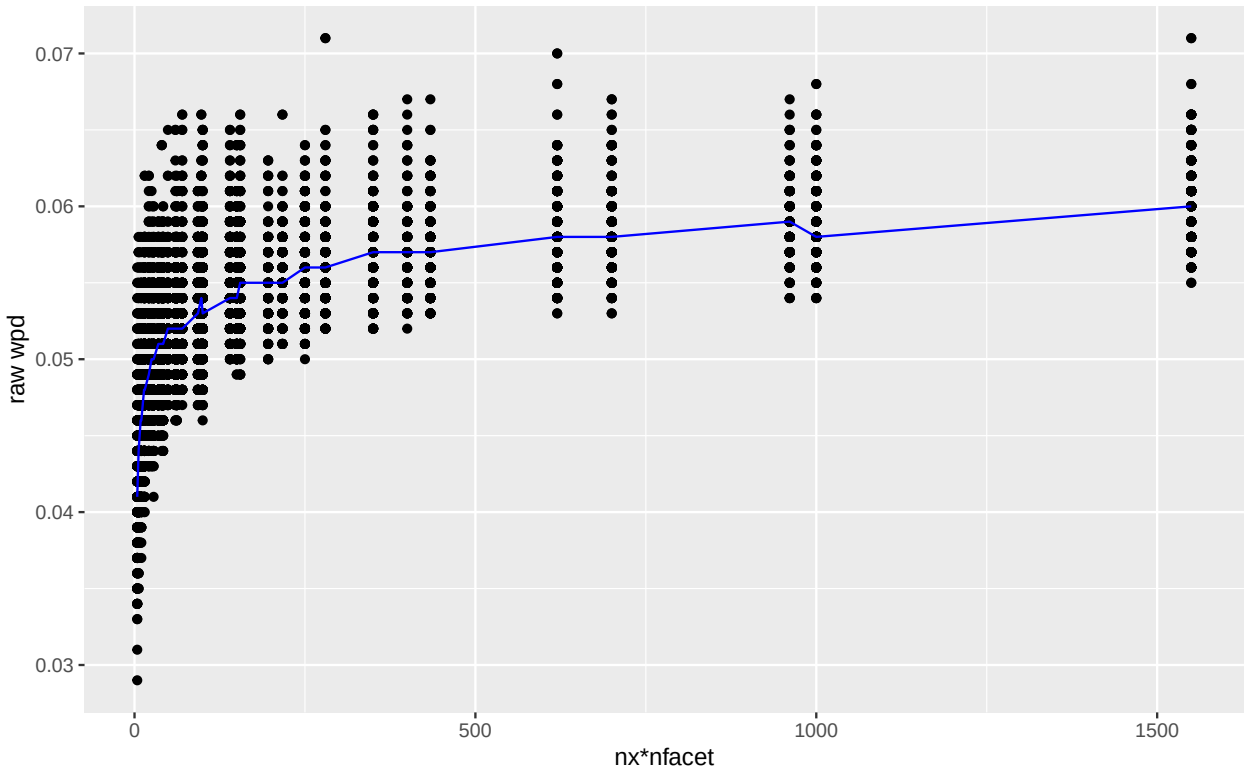



Figure 6: $wpd_{raw}$ is plotted against nx*nfacet and the blue line represents the median of the multiple values for each nx*nfacet levels.

# 3 Normalized distance measure

The distribution of wpd is different for different levels of facets and x-axis levels. This is because the statistic maximum which is used to define *wpd* is affected by the number of categories. The measure would have higher values if $A$ or $B$ has higher levels. However, we would ideally want a higher value of the measure only if there is significant difference between distributions across facet or x-axis categories, and not because the number of categories $J$ or $K$ is high. Therefore, in order to compare *wpd* across different combinations of facet and x-axis levels, we need to eliminate the impact of different levels of the facets and x-axis first and get a normalized measure. Henceforth we call the measure already discussed as $wpd_{raw}$ and the normalized measure as $wpd_{norm}$. The measure $wpd_{norm}$ could potentially lead to comparison of the measure across different panels and also identifying only the interesting panels from a data set. We discuss two approaches for normalization, both of which are based on the simulation results.

## 3.1 Simulation environment

Simulation studies were carried out to study the behavior of *wpd*, build the normalization method as well as compare and evaluate different normalization approaches. R version 4.0.1 (2020-06-06) is used with the platform: x86_64-apple-darwin17.0 (64-bit) running under: macOS Mojave 10.14.6 and MonaRCH, which is a next-generation HPC/HTC Cluster, designed from the ground up to address the computing needs of the Monash HPC community.

## 3.2 Methodology

We need a transformation on $wpd$ which will make it independent of the values of $nx * nfacet$. Two approaches have been employed for that purpose, the first one involves fitting a model and the latter involves a permutation method to make the distribution of the transformed $wpd$ similar across different $nx$ and $nfacet$.

### 3.2.1 Permutation approach

The permutation approach ensures that the distribution of the normalized distance measure has the same mean and standard deviation across all combinations of $nx_i$ and $nfacet_j$. The normalized distances through permutation is computed as follows: $x_k^{perm} = (x_k - mean_l(x_{k,l}))/sd_l(x_{k,l})$, $x_k^{perm}$ is the value of the $wpd_{perm}$ for the $k^{th}$ panel. Standardizing the variable $wpd_{perm}$ in this approach leads to $location = 0$ and $scale = 1$ for this variable.

While this works successfully to make the mean and standard deviation across different $nx$ and $nfacet$ (as seen in Figure **??**), it is computationally heavy and time consuming, and hence less user friendly when being actually used in practice. Hence, we propose another approach to normalization which is more approximate than exact but still has the same accuracy when compared to the permutation approach.

*incorporate from the document combining_normalisation_method.Rmd*

### 3.2.2 Modelling approach

*Linear model*

A log-linear model is fitted to see how the values of $wpd_{raw}$ changes with the values of $nx$ and $nfacet$. The model is of the form

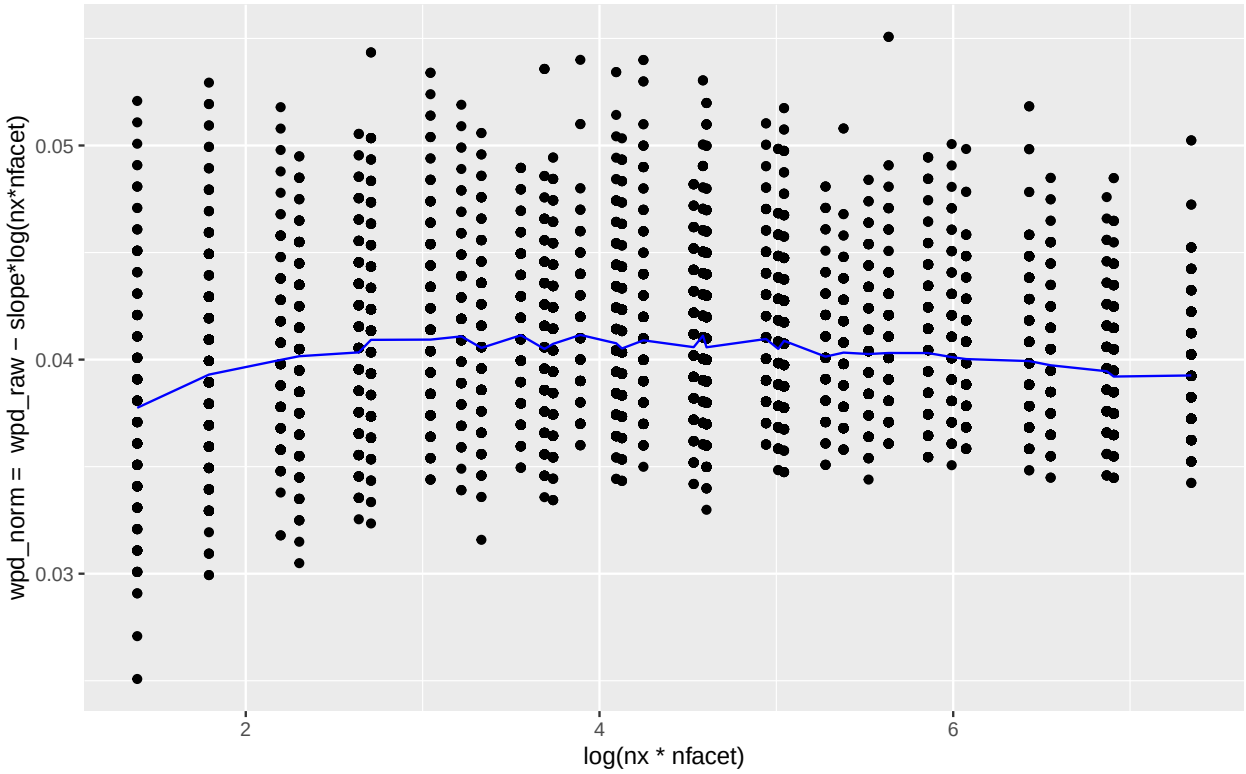
$$y_k = a + b * log(z_k) + e_k$$

, where, $y_k = median_l(x_{k,l})$ and $e_k$ are idiosyncratic errors. We have gone with the approach of fitting a linear regression model to estimate the parameters $a$ and $b$. The estimates and other model summary is given in **??**.

```
#>
#> Call:
#> lm(formula = actual ~ poly(log(`nx * nfacet`), 1, raw = TRUE),
#>     data = G21_median)
#>
#> Residuals:
#>        Min         1Q     Median         3Q        Max
#> -2.946e-03 -2.240e-04  5.135e-05  4.147e-04  1.014e-03
#>
#> Coefficients:
#>                                      Estimate Std. Error t value Pr(>|t|)
#> (Intercept)                         4.003e-02  4.171e-04   95.97   <2e-16
#> poly(log(`nx * nfacet`), 1, raw = TRUE) 2.826e-03  8.796e-05   32.13   <2e-16
#>
#> (Intercept)                         ***
#> poly(log(`nx * nfacet`), 1, raw = TRUE) ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Residual standard error: 0.0007881 on 32 degrees of freedom
#> Multiple R-squared:  0.9699, Adjusted R-squared:  0.969
#> F-statistic:  1032 on 1 and 32 DF,  p-value: < 2.2e-16
```

The final idea is to find a transformation on $wpd_{raw}$ which would remove the effect of $nx * nfacet$ on $wpd_{raw}$ and thus is defined as follows: $y^* = y - \hat{b} * log(z)$, where $y^*$ is the $median(wpd_{norm})$, $y$ is the $median(wpd_{raw})$, $\hat{b}$ is the estimated value of the parameter $b$, and $z = nx * nfacet$.

The above takes care of the mean and the heterogeneity of the median transformed measure. But, the original distribution will still have some dissimilarities in shape and location specially for small values of $nx$ and $nfacet$ as could be seen in **??**



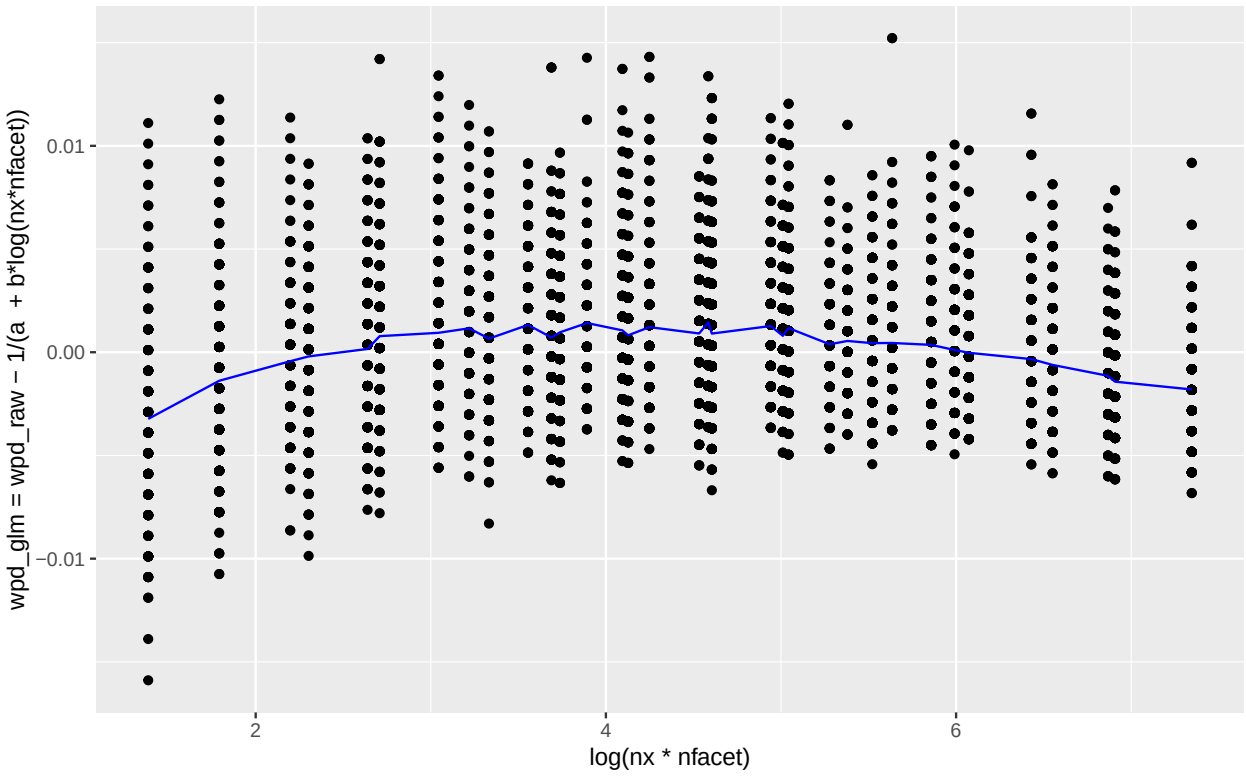

*Generalised linear model*

In the linear model approach, $wpd_{raw} \in R$ was assumed, whereas, $wpd_{raw}$, Jensen-Shannon Distance (JSD) lies between 0 and 1. Furthermore, JSD follows a Chi-square distribution, which is a special case of Gamma distribution and hence belongs to exponential family of distributions. Therefore, we can fit a generalized linear model instead of a linear model to allow for the response variable to follow a Gamma distribution. The inverse link is used when we know that the mean response is bounded, which is applicable in our case since $0 \leq wpd_{raw} \leq 1$.

We fit a Gamma generalized linear model with the inverse link which is of the form:

$$y = a + b * log(x) + e$$

, where $y = median(wpd_{raw})$, $x = nx * nfacet$. Let $E(y) = \mu$ and $a + b * log(x) = g(\mu)$ where $g$ is the link function. Then $g(\mu) = 1/mu$ and $\hat{\mu} = 1/(\hat{a} + \hat{b}log(x))$. The residuals from this model $(y - \hat{y}) = (y - 1/(\hat{a} + \hat{b}log(x)))$ would be expected to have no dependency on $x$. Thus, $wpd_{glm}$ is chosen as the residuals from this model and is defined as: $wpd_{glm} = wpd_{raw} - 1/(\hat{a} + \hat{b} * log(nx * nfacet))$.




```
#> [1] 1.003985
```

### 3.3 Combining normalizing approaches

We see that the transformation through the modeling approach leads to very similar distribution across high $nx$ and $nfacet$ (higher than 7) and not so much for lower $nx$ and $nfacet$. Hence, the computational load of permutation approach could be alleviated by using the modeling approach for the higher $nx$ and $nfacet$, however, it is important that we use the permutation approach for lower $nx$ and $nfacet$. However, it is difficult to compare the transformed $wpd$ from both of these approaches, since each of the variables is measured on a different scale (each of them have location 0). The transformed variables from the two approaches could be brought to the same scale so that for smaller categories, permutation approach is used and for larger categories, we can stick to modeling approach. These could be done through the following:

- Making the range of both the variables same by using min-max scaling method. In practice, however, we would only have one value of $wpd_{raw}$ which we need to transform using the modeling approach. Hence, min-max scaling approach could not be used here.
- Standardizing the variables and expressing scores at standard deviation units. Again in practice, however, we would only have one value of $wpd_{raw}$ which we need to transform using the modeling approach. Hence, standardizing scores could not be used here as we do not have the mean and standard deviation of a series while using transformation using modeling.
- Make the location and scale of both the approaches similar so that they could be compared. Please note that the range of values could be different in this case, however location and scale are brought to same levels.)

The measure $wpd_{glm}$ has location 0 and standard deviation ~ 0.003, whereas the measure $wpd_{permutation}$ which is a z-score, has a normal distribution with location 0 and standard deviation 1. To bring them

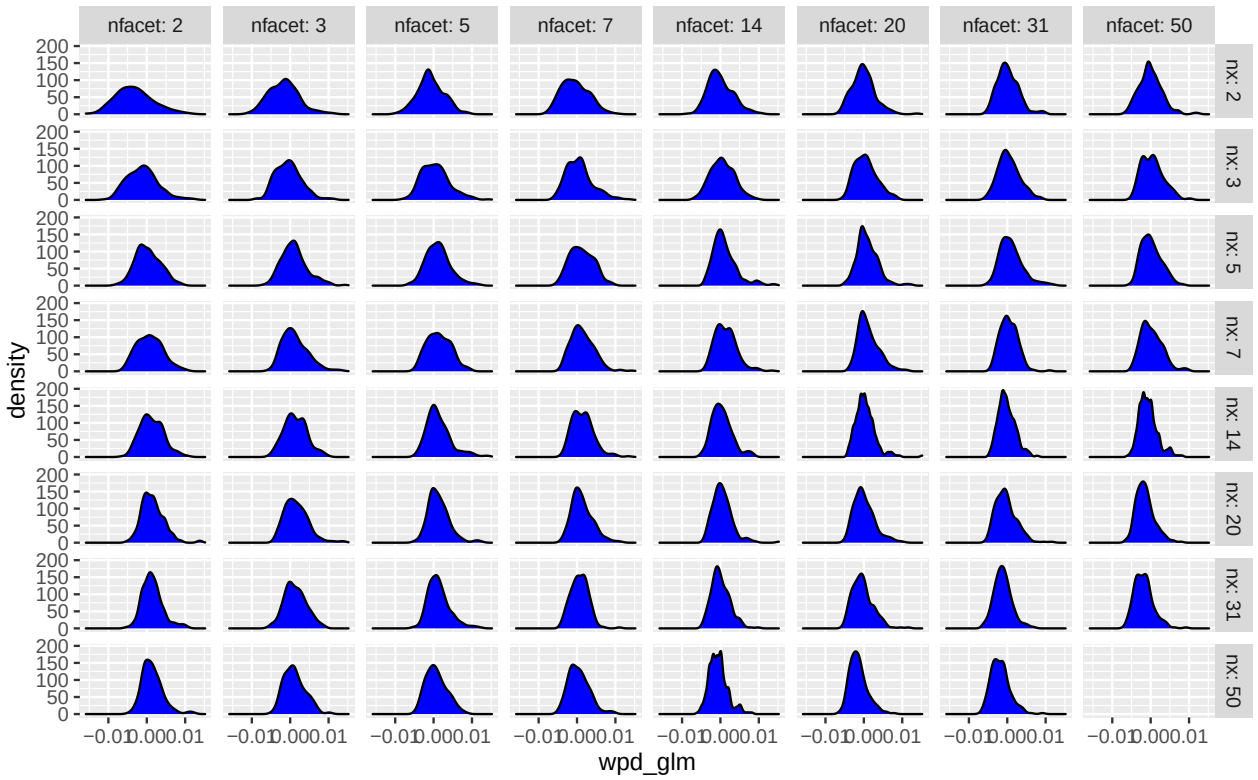



Figure 7: The distribution of $wpd_{glm}$ is plotted. The distributions are more similar across higher nx and nfacet and dissimilar for fewer nc and nfacets.

to the same scale, we have defined $wpd_{glm-scaled} = wpd_{glm} * 300$, which brings the standard deviation of $wpd_{glm-scaled}$ to almost 1, without changing the location.

The measure $wpd_{glm-scaled}$ seems to roughly follow a normal distribution except in the tails as could be seen in Figure 8 and the very method of permutation approach ensures that $wpd_{permutation}$ is also normally distributed. Further, they are brought to the similar scale and location and hence could be compared.

## 3.4 Properties

This section reports the results of a simulation study that was carried out to evaluate the behavior of $wpd_{norm}$. The behavior of $wpd_{norm}$ is explored in designs where there is in fact difference in distribution between facet categories ($D_{var_f}$) or across x-categories ($D_{var_x}$) or both ($D_{var_{all}}$). Using $\omega = \{1, 2, \ldots, 10\}$ and $\lambda = seq(from = 0.1, to = 0.9, by = 0.05)$, observations are drawn from a N(0,1) distribution for each combination of $nx$ and $nfacet$ from the following sets: $nx = nfacet = \{2, 3, 5, 7, 14, 20, 31, 50\}$. $ntimes = 500$ is assumed for this setup as well. Furthermore, to generate different distributions across different combination of facet and x levels, the following method is deployed - suppose the distribution of the combination of first levels of $x$ and $facet$ category is $N(\mu, \sigma)$ and $\mu_{jk}$ denotes the mean of the combination $(a_j b_k)$, then $\mu_{j.} = \mu + j\omega$ (for design $D_{var_x}$) and $\mu_{.k} = \mu + k\omega$ (for design $D_{var_f}$).

The tabulated values and graphical representations of the simulation results are provided in Appendix. The learning from the simulations are as follows: The values of $wpd_{norm}$ is least for $D_{null}$, followed by $D_{var_f}$, $D_{var_x}$) and $D_{var_{all}}$). This is a desirable result since the measure $wpd_{norm}$ was designed such that this relationship holds. Furthermore, the distribution of the measure $wpd_{norm}$ does not change for different facet and x categories. The distribution of $wpd_{norm}$ looks similar with at least the mean and standard of the distributions being uniform across panels. This means, $wpd_{norm}$ could be used to measure differences in distribution across panels. Also, note that since the data is processed using normal-quantile-transform, this

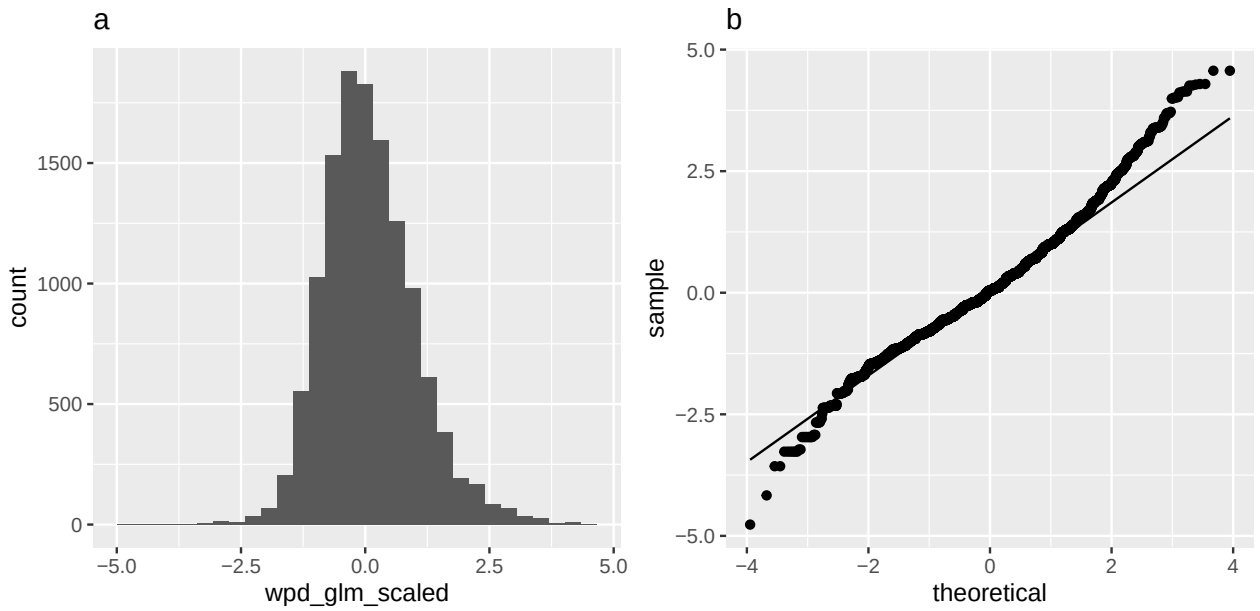



Figure 8: In panel a, the histogram of $wpd_{glm-scaled}$ is plotted. In parl b, the QQ plot is shown with the theoretical quantiles on the x-axis and $wpd_{glm-scaled}$ quantiles on the y-axis. The distribution looks symmetric and looks like normal except in the tails.

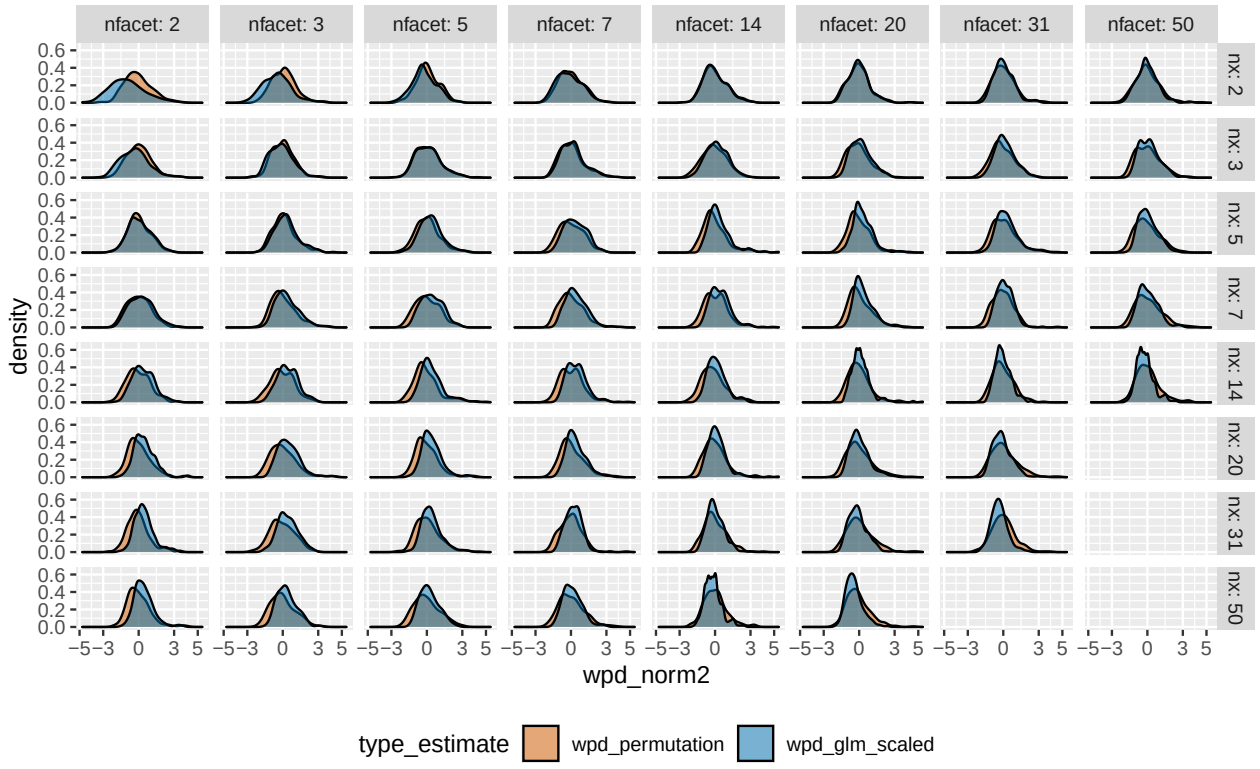


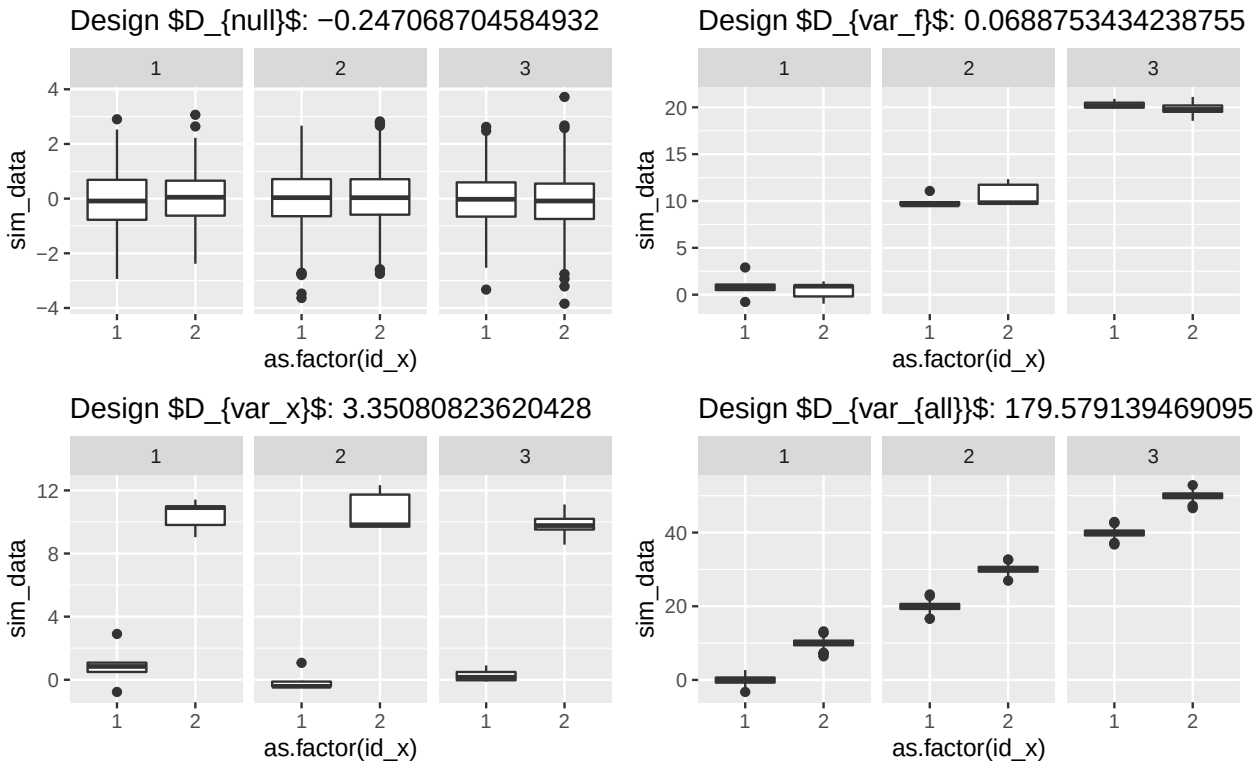

Figure 9: w$pd_permutation$ and $wpd_{glm-scaled}$ are plotted together on the same scale. They also have the same location and hence the values from these two approaches could be compared across panels. $wpd_{glm-scaled}$ would be used to normalise $wpd_{raw}$ for higher $nx$ and $nfacet$ and $wpd_{glm-scaled}$ would be used for smaller levels to alleviate the problem of computational time.

measure is independent of the initial distribution of the underlying data and hence is also comparable across different data sets. This is valid for the case when sample size *ntimes* for each combination of categories is at least 30 and *nperm* used for computing $wpd_{norm}$ is at least 100. More detailed results about the properties of $wpd_{raw}$ and $wpd_{norm}$ could be found in Appendix.

# 4 Ranking and selecting significant harmonies

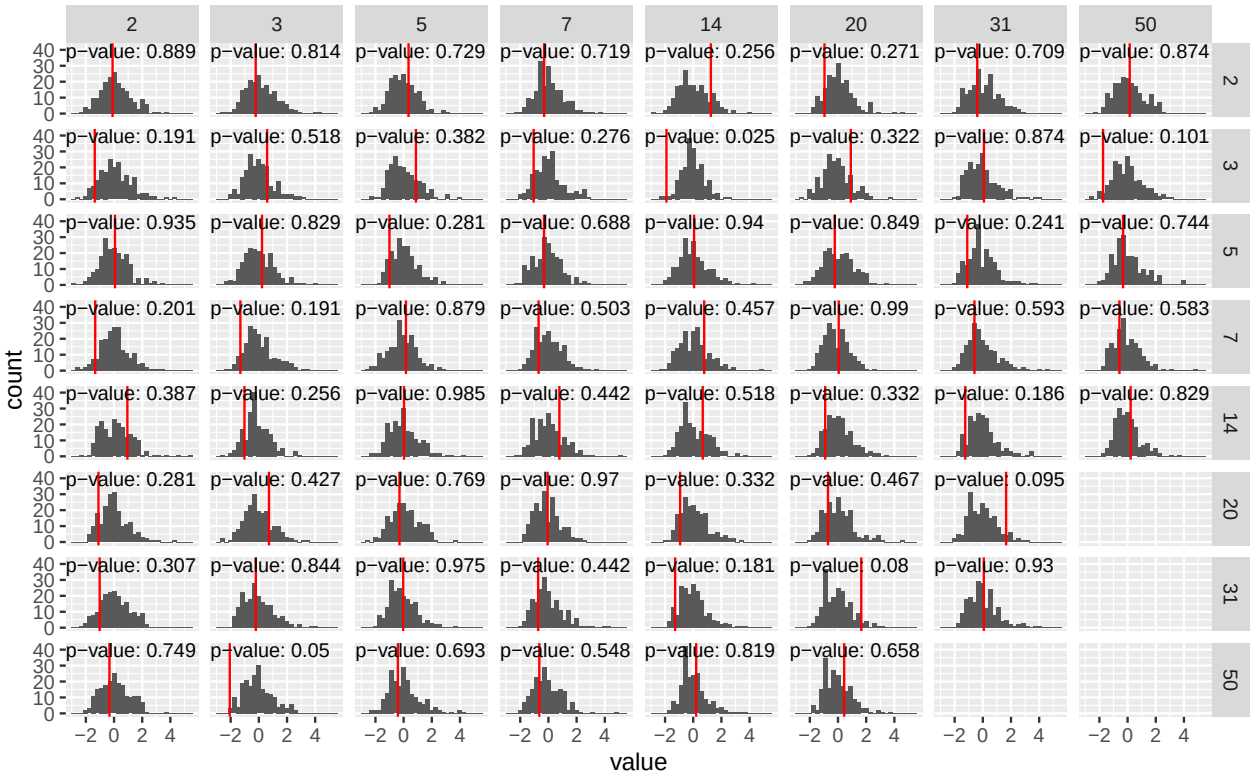
Complete randomness in the measured variable indicates that the process follows a homogeneous underlying distribution over the whole time series, which essentially implies there is no interesting distinction across any different categories of the cyclic granularities. We can remove the harmonies for which no interesting patterns are observed through a randomization permutation method. Essentially, the assumption is that under the null hypothesis, there is no difference in categories between the pair of cyclic granularities in the chosen harmony. This method is based on the generation of randomly chosen reassignments (permutations) of the data across different cyclic granularities and the computation of $wpd_{norm}$ for each of these reassignments. The percentages of times the theoretical distribution greater than or equal to the respective observed $wpd_{norm}$ values are calculated and are used to obtain the P value. The procedure for the permutation test is:

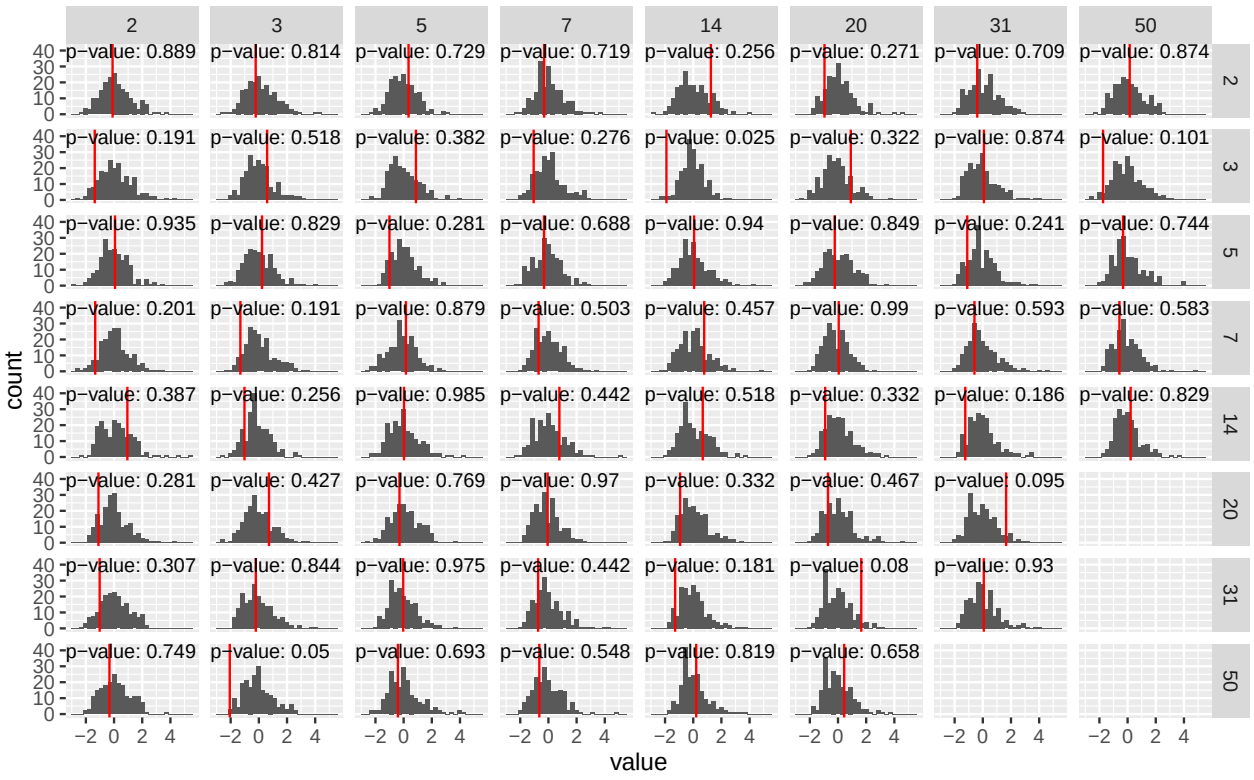
1. Given the data; $\{v_t : t = 0, 1, 2, \ldots, T-1\}$, the $wpd_{norm}$ is computed and is represented by $wpd_{obs}$.
2. From the original sequence a random permutation is obtained: $\{v_t^* : t = 0, 1, 2, \ldots, T-1\}$.
3. $wpd_{norm}$ is computed for the permuted sequence of the data and is represented by $wpd_{perm_1}$.
4. Steps (2) and (3) are repeated a large number of times M (M = 200).
5. For each permutation, one $wpd_{perm_i}$ is obtained. Define $wpd_{sample} = \{wpd_{perm_1}, wpd_{perm_2}, \ldots, wpd_{perm_M}\}$.
6. $95^{th}$ percentile of this $wpd_{sample}$ distribution is computed and stored in $wpd_{threshold}$.
7. If $wpd_{obs} > wpd_{threshold}$, harmony pairs are accepted. Only one threshold for all harmony pairs.



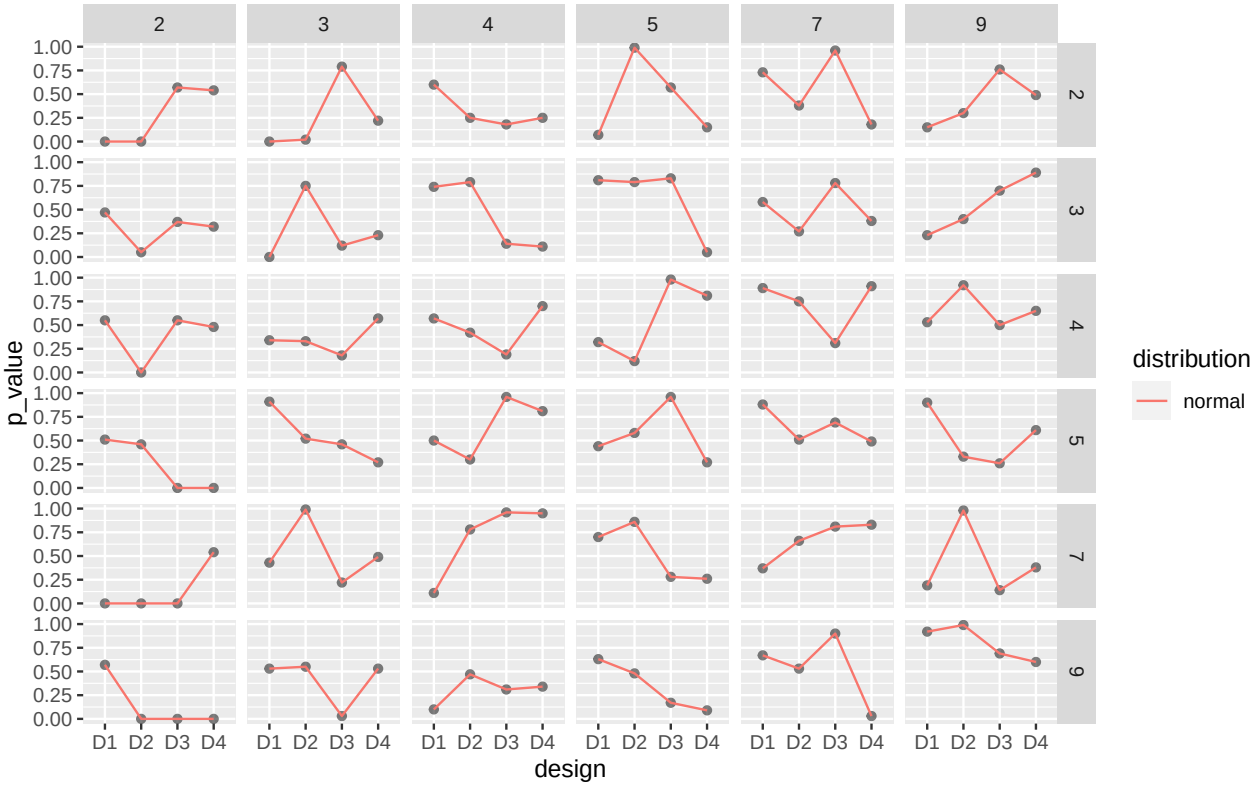
The p-value of the design $D_{null}$ is size. The p-value of other designs is power. Confidence interval of the

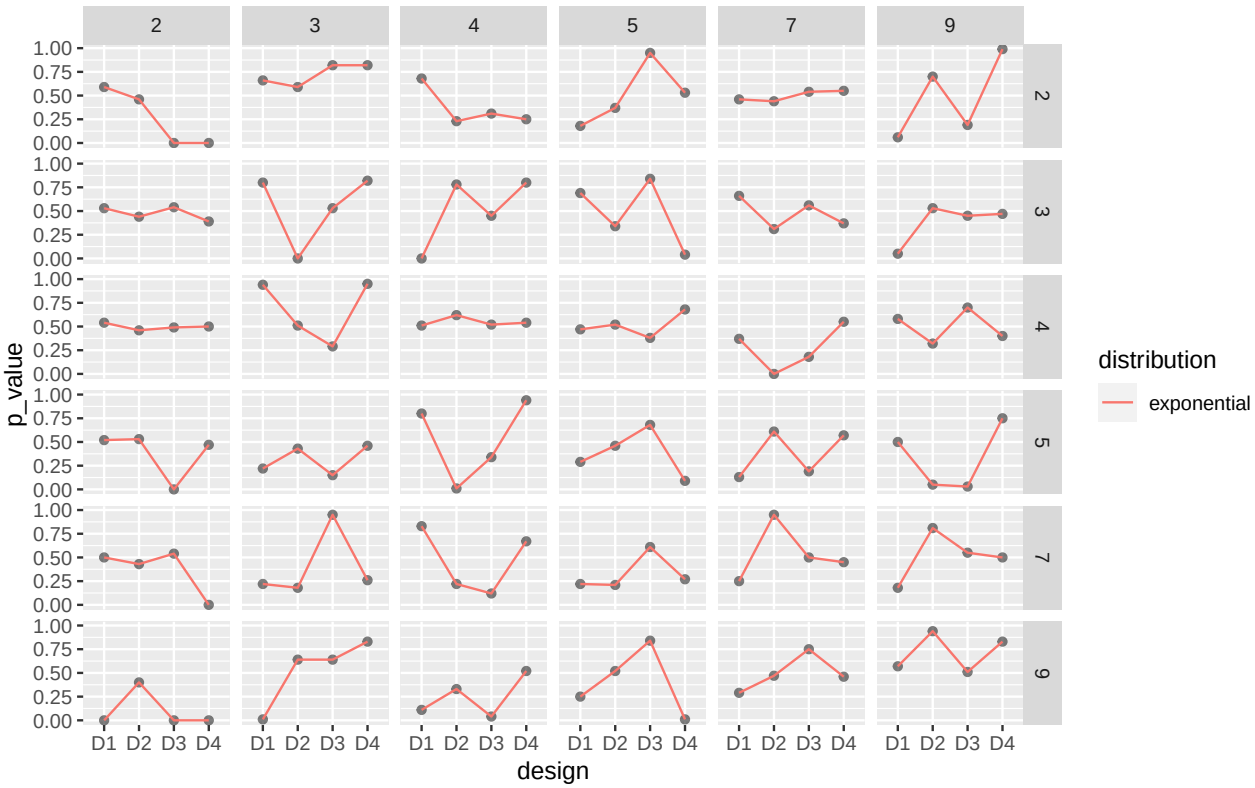
The p-value is almost always greater than 0.05, which means there is no significant differences between the different categories, which is true from the simulation design.

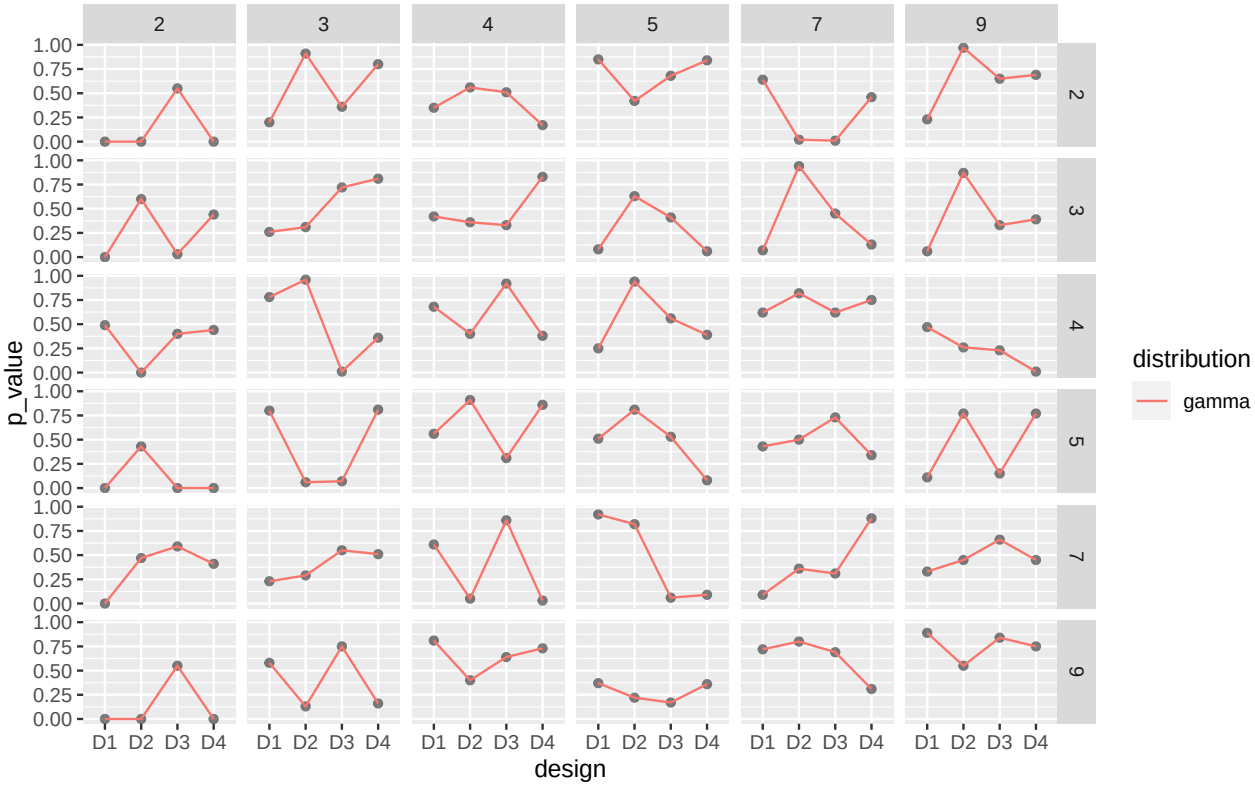


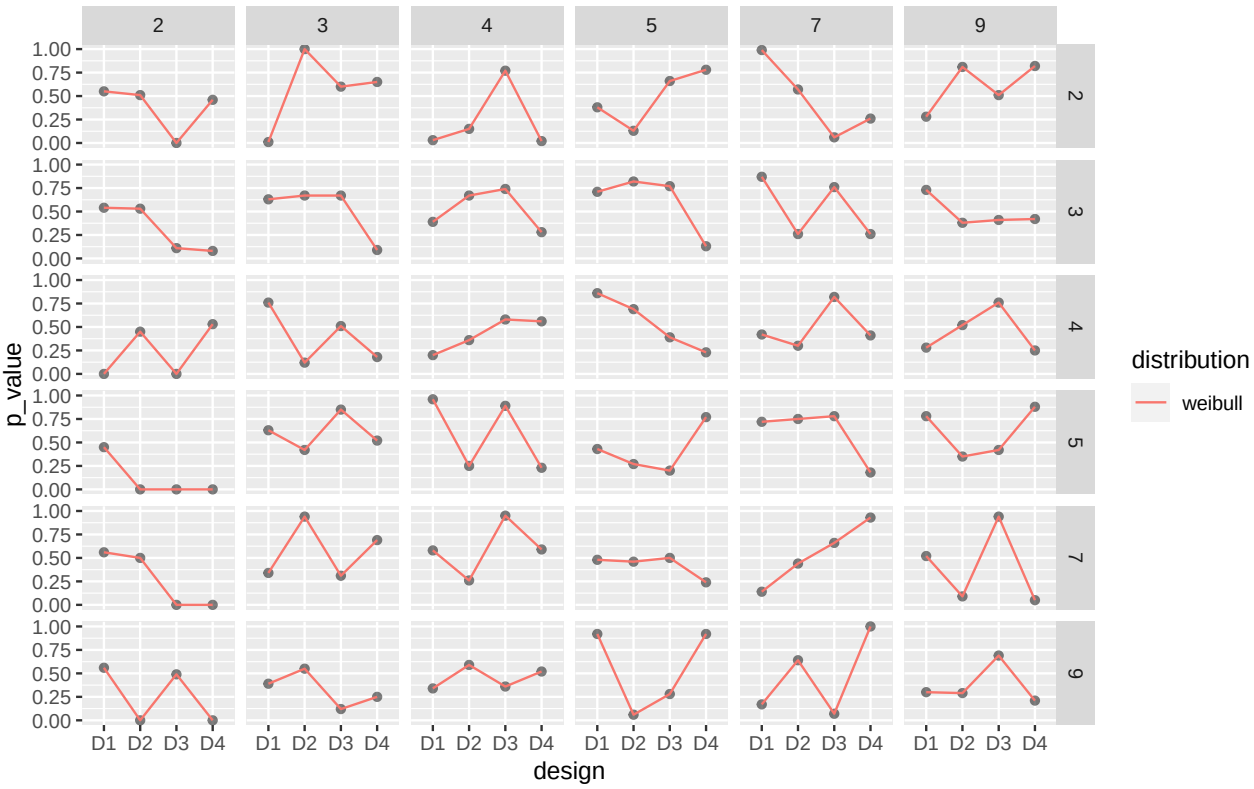



**4.0.0.1 Characteristics under different simulation designs** A set of simulation runs that are conducted and some outputs of which are reported.




2
3
4
5
7
9
1.00
0.75
0.50
0.25
0.00
p_value
D1 D2 D3 D4
design
distribution
exponential


2
3
4
5
7
9
1.00
0.75
0.50
0.25
0.00
p_value
D1 D2 D3 D4
design
distribution
gamma




# 5 Application to smart meter datasets

## 5.1 Residential smart meter dataset

The smart meter data set from four households in Melbourne referred in Wang, Cook, and Hyndman (2020) has been utilized to see the use of the distance measure proposed in the paper. The data contains half-hourly electricity consumption from Jan-2018 to Jun-2018 for each of the households, which is procured by them by downloading their data from the energy supplier/retailer. Demand data for two households (id 2 and 4) are shown in a linear time scale in 10. It is evident from the range of the demand data that these households vary in consumption levels as in their temporal patterns. In panel A (left), the linear representation of the entire time period is shown, whereas in panel B (right) a particular month is shown and furthermore a week has been highlighted to inspect if there is any daily or weekly periodic patterns in their behavior that is evident from the linear representation of the time series. We pick household id 2 and 4 for our analysis to see which periodic patterns are important in their behavior, if they are same or different and if it makes sense using our distance metric.

### 5.1.1 Data preprocessing

**Normal-quantile-transform**

Let $y_{i,t}$ denote the electricity demand for $i^{th}$ household for time period $t$. This is expected to have an assymmetrical distribution as could be seen in **??** and the Normal score transform has been applied to make it more treatable. Let $y*_{i,t}$ denote the normal-quantile transformed electricity demand for $i^{th}$ household for time period $t$.

**Characterizing empirical distributions through quantiles** ( write in the methodology maybe)

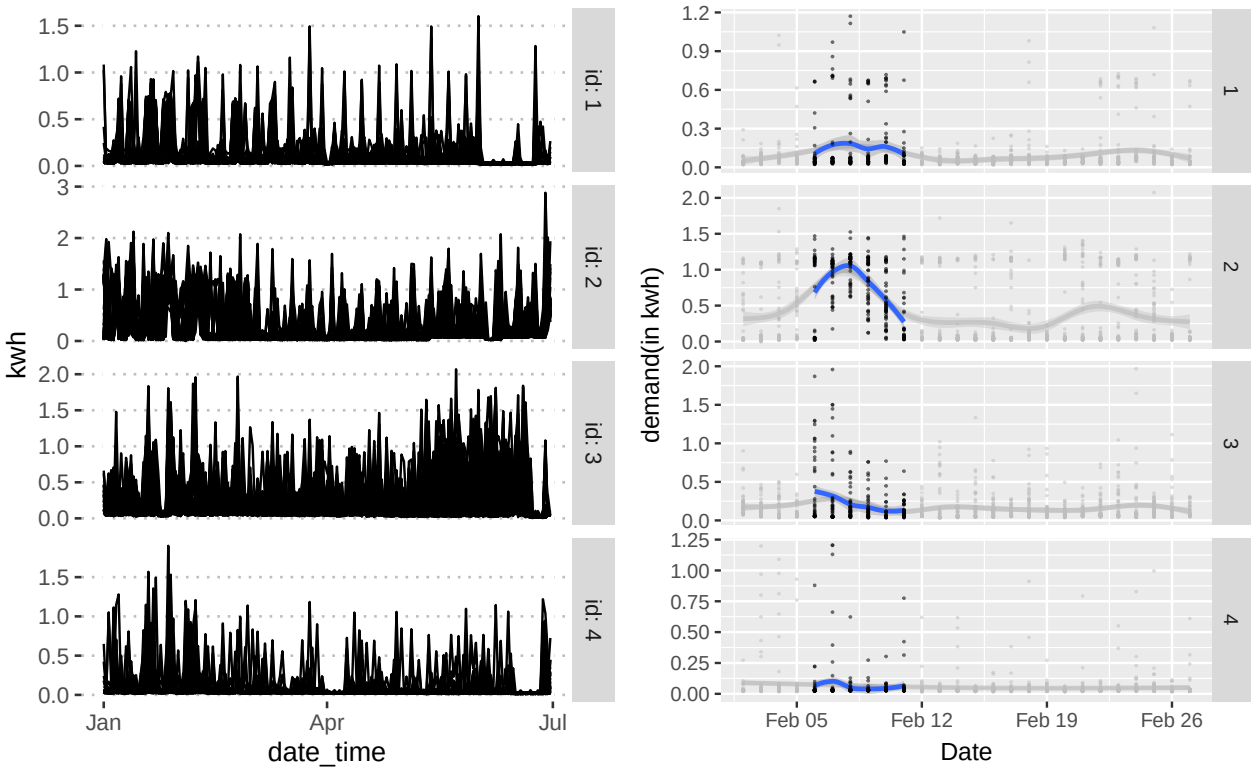
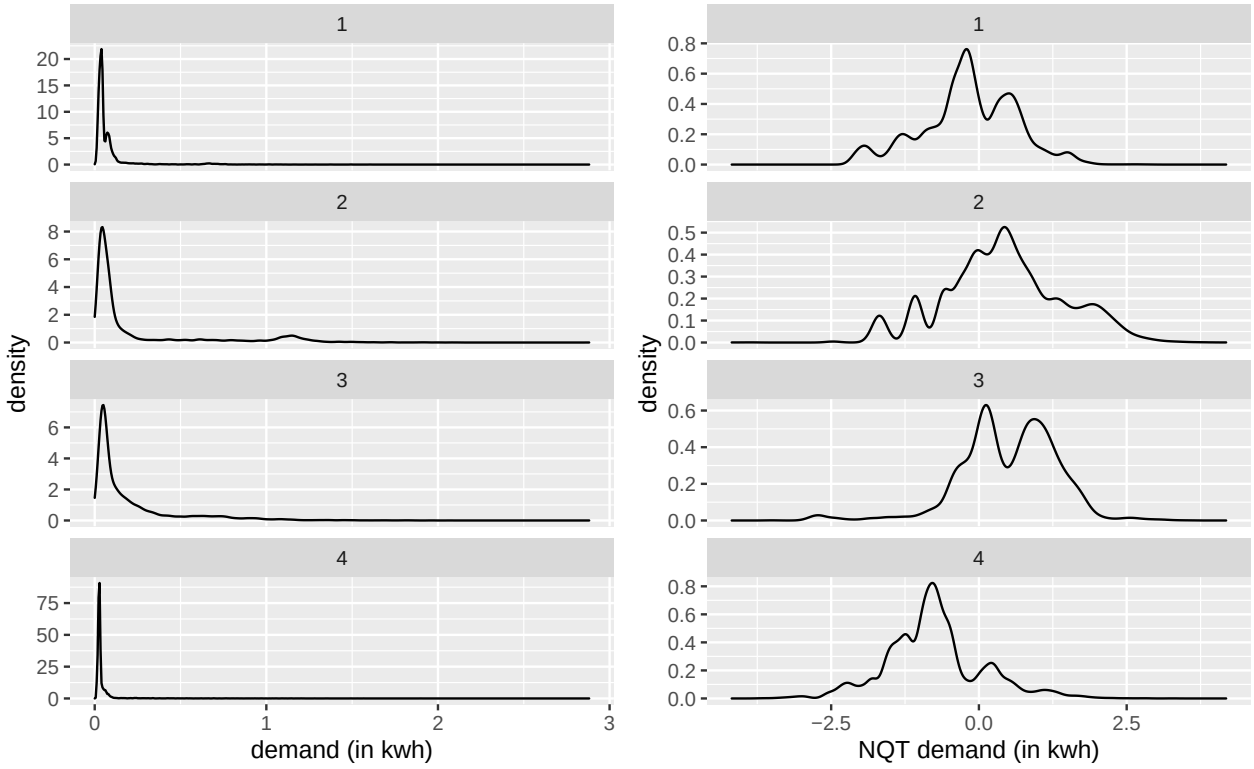



Figure 10: Electricity demand for four households are shown in different facets from Jan-18 to Jun-18 in Fig a and it has been zoomed in for Feb-18 in Fig b, where a week in Feb-18 has been highlighted. From the scales of Fig a, it is apparent that they have different level of consumption but all of them have some periodic behavior in terms of regular peaks and troughs. It is not clear which all periodic patterns exist. In fig b, periodic pattern is zoomed in for a month and we can see weekly patterns for the entire period and daily pattern for the highlighted week.

Let $A = \{a_j : j = 1, 2, \ldots, J\}$ be a cyclic granularity of interest. Let $q_{A,j}^{i,p}$ is the quantile with probability $p$ for the $j^{th}$ category of the cyclic granularity $A$ of the $i^{th}$ household. Sample quantiles were computed at $p = 0.01, 0.02, \ldots, 0.99$ for each household $i \in i = 1, 2, 3, 4$ for each category $j \in j = 1, 2, \ldots, J$ and all $A \in N_C$.

**Computing JS distances between harmonies for each categories of the cyclic granularity involved**




```
#> 22.168 sec elapsed
```

```
#> 17.695 sec elapsed
```

| facet_variable | x_variable | facet_levels | x_levels | wpd |
|---|---|---|---|---|
| day_month | hour_day | 31 | 24 | 146.239891 |
| hour_day | wknd_wday | 24 | 2 | 144.019534 |
| hour_day | day_week | 24 | 7 | 128.039580 |
| hour_day | week_month | 24 | 5 | 111.984797 |
| day_week | hour_day | 7 | 24 | 108.550915 |
| wknd_wday | hour_day | 2 | 24 | 82.654722 |
| hour_day | day_month | 24 | 31 | 64.307785 |
| day_week | week_month | 7 | 5 | 55.494311 |
| week_month | day_week | 5 | 7 | 53.863976 |
| week_month | hour_day | 5 | 24 | 49.106006 |
| day_month | wknd_wday | 31 | 2 | 30.372962 |
| wknd_wday | day_month | 2 | 31 | 27.357555 |
| wknd_wday | week_month | 2 | 5 | 7.121047 |
| week_month | wknd_wday | 5 | 2 | 6.493801 |

| facet_variable | x_variable | facet_levels | x_levels | wpd |
|---|---|---|---|---|
| day_week | hour_day | 7 | 24 | 147.453126 |
| hour_day | day_week | 24 | 7 | 141.823542 |
| hour_day | wknd_wday | 24 | 2 | 138.497045 |
| day_month | hour_day | 31 | 24 | 105.136735 |
| day_week | week_month | 7 | 5 | 91.216069 |
| week_month | day_week | 5 | 7 | 77.150700 |
| week_month | hour_day | 5 | 24 | 73.518788 |
| hour_day | day_month | 24 | 31 | 69.386583 |
| hour_day | week_month | 24 | 5 | 62.450988 |
| wknd_wday | day_month | 2 | 31 | 44.768545 |
| wknd_wday | hour_day | 2 | 24 | 43.919054 |
| day_month | wknd_wday | 31 | 2 | 34.989184 |
| week_month | wknd_wday | 5 | 2 | 6.641774 |
| wknd_wday | week_month | 2 | 5 | 3.971057 |

```
#> 2069.182 sec elapsed
```

```
#> 202.58 sec elapsed
```

| facet_variable | x_variable | facet_levels | x_levels | wpd |
|---|---|---|---|---|
| day_month | hour_day | 31 | 24 | 146.2399 |
| hour_day | wknd_wday | 24 | 2 | 144.0195 |
| hour_day | day_week | 24 | 7 | 128.0396 |

| facet_variable | x_variable | facet_levels | x_levels | wpd |
|---|---|---|---|---|
| day_week | hour_day | 7 | 24 | 147.4531 |
| hour_day | day_week | 24 | 7 | 141.8235 |
| hour_day | wknd_wday | 24 | 2 | 138.4970 |

### 5.2 Australian smart meter data set

# 6 Discussion points and future work

Exploratory data analysis involve many iterations of finding and summarizing patterns. With temporal data available at ever finer scales, exploring periodicity has become overwhelming with so many possible granularities to explore. This work refines the selection of appropriate pairs of granularities by identifying those for which the differences between the displayed distributions is greatest, and rating these selected harmony pairs in order of importance for exploration.

A future direction of work could be to look at more individuals/subjects and group them according to similar periodic behavior. Behaviors across different cyclic granularities would be different for different subjects and one way to find groups would be to actually locate clusters who have similar periodic behavior.

# 7 Appendix

## 7.1 Null distribution

#### 7.1.1 Size: Simulated same distribution for all combinations of categories for all harmony pairs.

Failure to reject the null hypothesis when there is in fact no significant effect.

#### 7.1.2 Normalised maximum distances follow standard Gumbel distribution

#### 7.1.3 Limiting distribution of median of normalised maximum distances is normal

Let a continuous population be given with cdf F(x) (cumulative distribution function) and median $\xi$ (assumed to exist uniquely). For a sample of size $2n+1$, let $\tilde{x}$ denote the sample median. The distribution of $\tilde{x}$,under certain conditions, to be asymptotically normal with mean $\xi$ and variance $\sigma_n^2 = \frac{1}{4}[f(\xi)]^2(2n+1)$, where $f(x) = F'(x)$ is the pdf (probability density function).

### 7.2 Power

### 7.3 Confidence interval

Failure to reject the null hypothesis when there is in fact a significant effect.

To estimate the sampling distribution of the test statistic we need many samples generated under the null hypothesis. If the null hypothesis is true, changing the exposure would have no effect on the outcome. By randomly shuffling the exposures we can make up as many data sets as we like. If the null hypothesis is true the shuffled data sets should look like the real data, otherwise they should look different from the real data. The ranking of the real test statistic among the shuffled test statistics gives a p-value.

#### 7.3.1 Varying distribution across facet

#### 7.3.2 Varying distribution across x-axis

#### 7.3.3 Varying distribution across both facets and x-axis

#### 7.3.4 Repeat all with varying facet and x-axis levels

*Conclusion*: The test should reject the null hypothesis if distributions are different.

Dang, T N, and L Wilkinson. 2014. "ScagExplorer: Exploring Scatterplots by Their Scagnostics." In *2014 IEEE Pacific Visualization Symposium*, 73–80.

Gupta, Sayani, Rob J Hyndman, Dianne Cook, and Antony Unwin. 2020. "Visualizing Probability Distributions Across Bivariate Cyclic Temporal Granularities," October. http://arxiv.org/abs/2010.00794.

Kullback, S, and R A Leibler. 1951. "On Information and Sufficiency." *Ann. Math. Stat.* 22 (1): 79–86.

Menéndez, M L, J A Pardo, L Pardo, and M C Pardo. 1997. "The Jensen-Shannon Divergence." *J. Franklin Inst.* 334 (2): 307–18.

Tukey, John W, and Paul A Tukey. 1988. "Computer Graphics and Exploratory Data Analysis: An Introduction." *The Collected Works of John W. Tukey: Graphics: 1965-1985* 5: 419.

Wang, Earo, Dianne Cook, and Rob J Hyndman. 2020. "Calendar-Based Graphics for Visualizing People's Daily Schedules." *Journal of Computational and Graphical Statistics*. https://doi.org/10.1080/10618600.2020.1715226.

Wilkinson, Leland, Anushka Anand, and Robert Grossman. 2005. "Graph-Theoretic Scagnostics." In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, 157–64. IEEE.