A new metric for automatic discovery of periodic patterns in 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

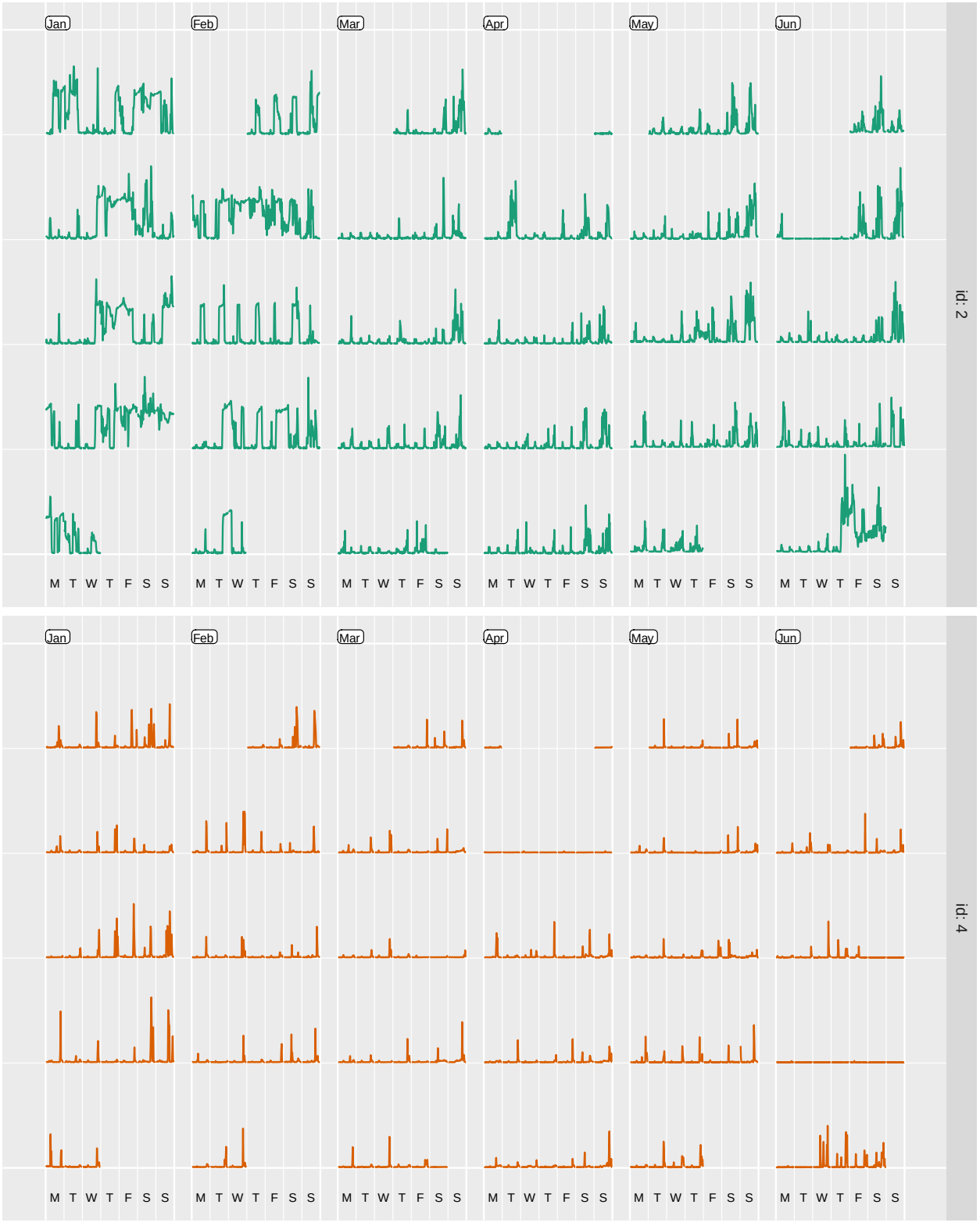


Figure 1: Calendar display.

(#fig:calendar- <-)

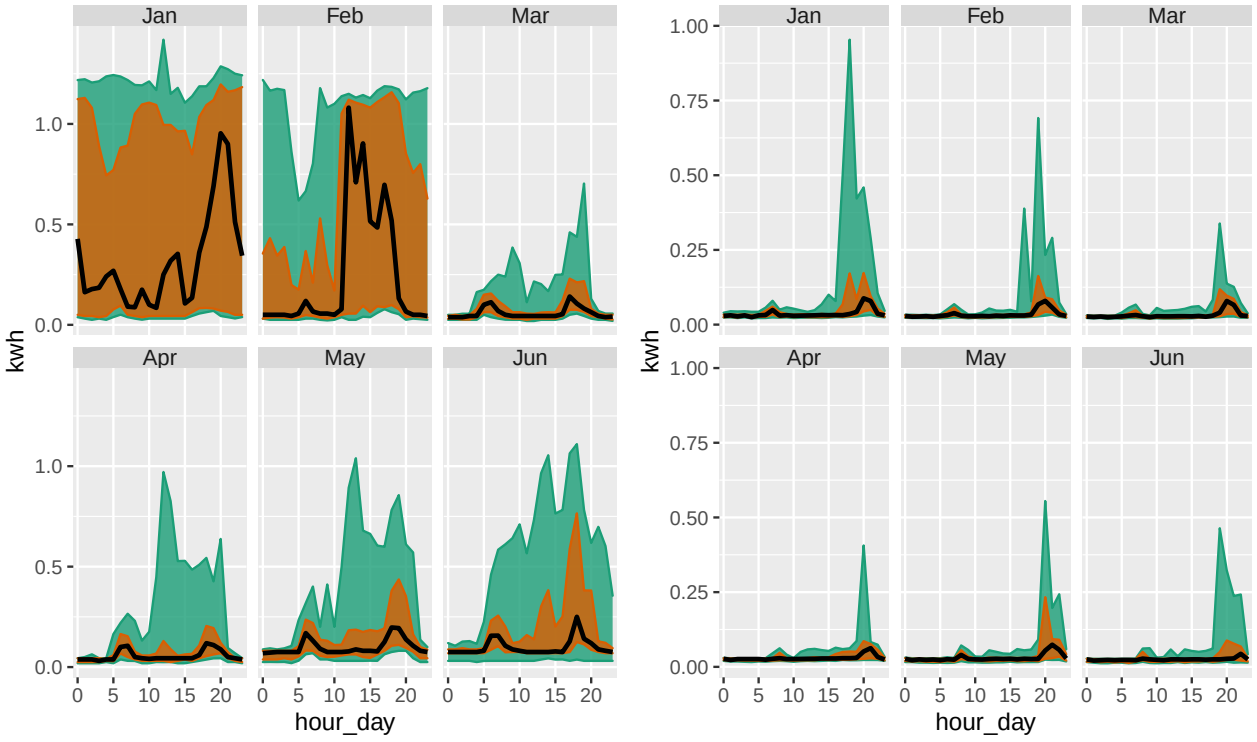


Figure 2: something

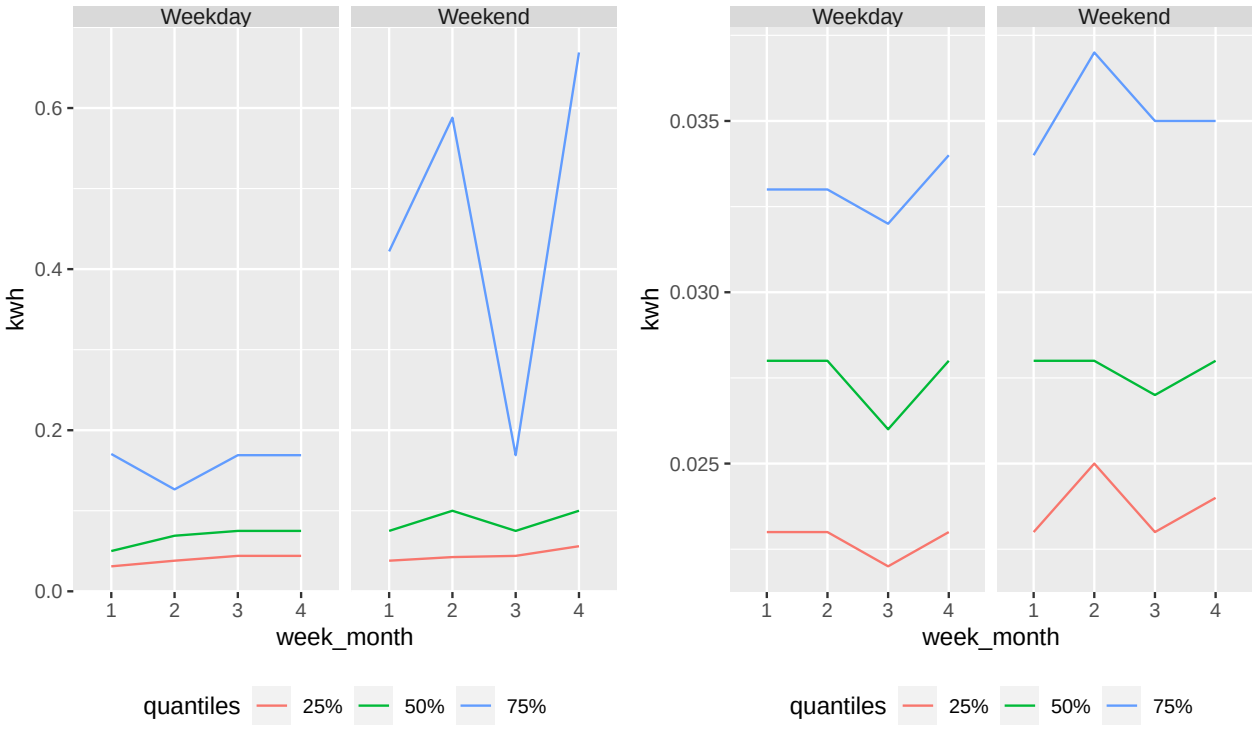


Figure 3: something2

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 A distance measure for quantifying patterns in harmonies

We are interested in assessing the structure in probability distributions of the measured variable across bivariate cyclic granularities. We propose a measure called Weighted Maximum Pairwise Distances (wpd) to quantify the structure in such a design. The principle employed towards this goal is explained through a simple example explained in Figure 4. Each of these figures have the same panel design with 2 x-axis categories and 3 facet levels. Figure 4a 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 4b has x categories drawn from the same distribution within a facet but the mean has been incremented by 5 units for every consecutive facets. Figure 4c exhibits an exact opposite situation where distribution between the x-axis categories within each facet is different but they are same across facets. For this situation, mean of only the x-axis categories are increased by 5 units for each consecutive category. Figure 4d takes a step further by varying the distribution across both facet and x-axis categories. If the displays are to be ranked in order of importance from minimum to maximum, then an obvious choice would be placing a followed by b, c and then d. It might be argued that it is not clear if b should precede or succeed c in the ranking. 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 in mind, display b is considered less informative as compared to display c. Hence, with reference to the graphical design in Gupta et al. (2020), 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.

Intuitively, while finding a structure or measuring the strength of patterns in Figure 4, it makes sense to look for within-group and between-group variation. Larger variation would imply stronger patterns, whereas small variation would imply the underlying structure is not changing within or between group. Thus, a distance measure aimed to capture this structure should ideally estimate these within-group and between-group variations. One of the potential ways to do this is to measure the distances between distributions of the continuous random variable measured within and between groups, weigh them basis if they are within or between groups and then take the maximum of those distances as an estimate of the strength of the structure. We call this metric *wpd* which stands for Weighted Pairwise Distances. This section starts with possible ways of characterizing distributions and computing distances between them and then describe in details how the measure *wpd* is defined. This is similar to Hyndman, Liu, and Pinson (2018) where the authors compute the Jensen Shannon distance between two density estimates by computing percentiles and stresses

the advantages to working with percentiles rather than the data directly in case of missing observations or even unsynchronized time series.

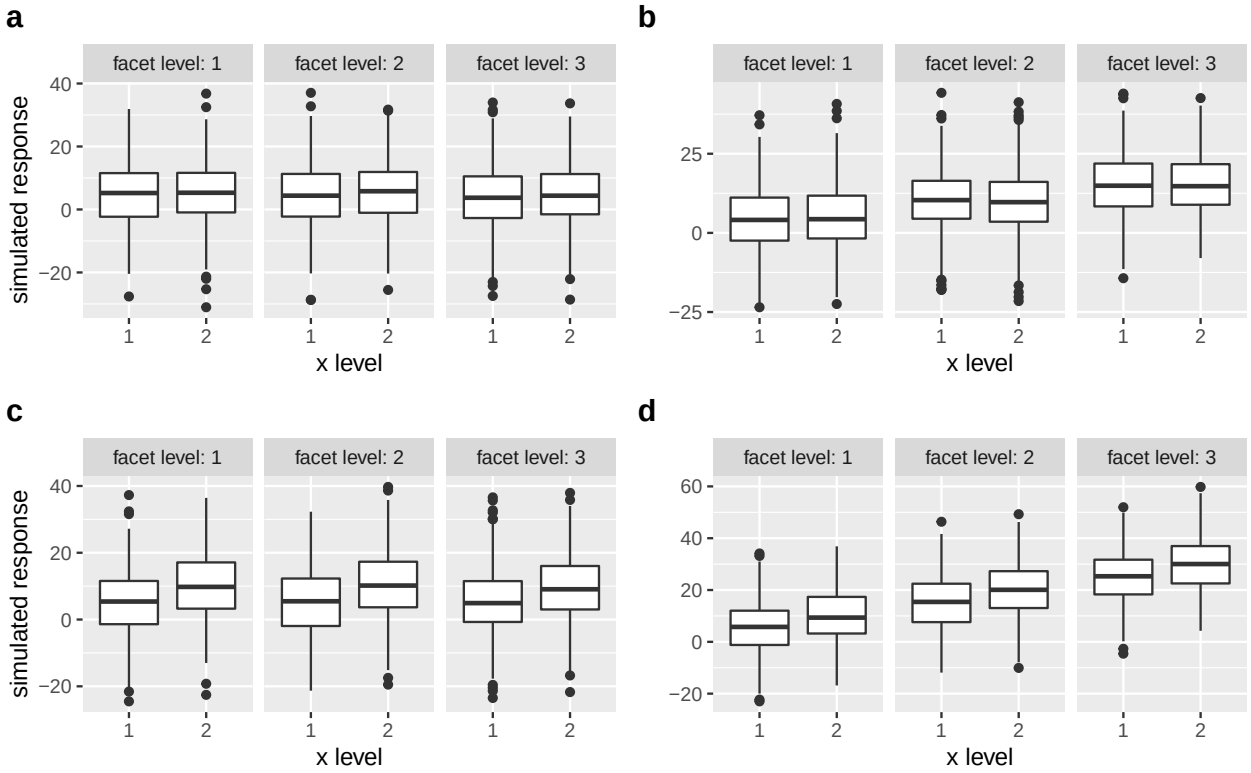


Figure 4: A graphical display with two categories mapped to x-axis and 3 categories mapped to facets with the distribution of a continuous random variable plotted on the y-axis. Display a is not interesting as the distribution of the variable does not depend on x or facet categories. Display b and c are more interesting than a since there is a change in distribution either across facets (b) or x-axis (c). Display d is most interesting in terms of displaying the strongest pattern as distribution of the variable changes across both facet and x-axis variable.

2.1 Notations

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 $v = \{v_t : t = 0, 1, 2, \ldots, T-1\}$ be a continuous variable observed across T time points. Let the four elementary designs as described in Figure 4 be D_{null} where there is no difference in distribution of v for A or B, D_{var_f} denotes the set of designs where there is difference in distribution of v for B and not for A. Similarly, D_{var_x} denotes the set of designs where difference is observed only across A. Finally, $D_{var_{all}}$ denotes those designs for which difference is observed across both A and B.

2.2 Characterising distributions

Multiple observations of v correspond to the subset $v_{jk} = \{s : A(s) = j, B(s) = k\}$. The number of observations and the structure might vary widely across subsets due to the structure of the calendar, missing observations or uneven locations of events in the time domain. Each $v_{jk} : j \in \{1, 2, \ldots, J\}, k \in \{1, 2, \ldots, K\}\}$ is assumed to be drawn from a continuous probability distribution and have certain characteristics. Often shape, central tendency, and variability are the common characteristics used to describe the distribution. Mean, median or mode are generally used to describe the center of the distribution, while range, standard

deviation, quantiles, standard errors and confidence intervals are often used to describe variability in the distribution. Quantiles are chosen as a way to characterize distributions in this paper.

The quantile of a distribution with probability p is defined as $Q(p) = F^{-1}(p) = inf\{x : F(x) > p\}$, $0 < p < 1$ where $F(x)$ is the distribution function. There are two broad approaches to quantile estimation, viz, parametric and non-parameteric. The benefit of using a non-parametric estimator is that there are less rigid assumptions made about the nature of the underlying distribution of the data. Sample quantiles could be used for estimation in a non-parametric setup. Hyndman and Fan (1996) describes the difficulty in defining sample quantiles and recommends the use of median-unbiased estimator because of "desirable properties of a quantile estimator and can be defined independently of the underlying distribution.". The `stats::quantile()` function in R Core Team (2019) could be used for practical implementation where type = 8 refers to the algorithm corresponding to the median-unbiased estimator. The default quantile chosen in this paper is percentiles computed for $p = 0.01, 0.02, \ldots, 0.99$, where for example, the 99^{th} percentile would be the value corresponding to $p = 0.99$ and hence 99% of the observations would lie below that.

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 by $D(q_1||q_2)$ is a non-symmetric measure of the difference between two probability distributions q_1 and q_2 and is interpreted as the amount of information lost when q_2 is used to approximate q_1. 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 between distributions are Hellinger distance, total variation distance and Fisher information metric.

In this paper, the pairwise distances between the distributions of the measured variable are computed through Jensen-Shannon distance (JSD), defined by,

$$JSD(q_1||q_2) = \frac{1}{2}D(q_1||M) + \frac{1}{2}D(q_2||M)$$

where $M = \frac{q_1+q_2}{2}$ and $D(q_1||q_2) := \int_{-\infty}^{\infty} q_1(x) f(\frac{q_1(x)}{q_2(x)})$ is the KL divergence between distributions q_1 and q_2.

Furthermore, these distances are distributed as chi-squared with m degrees of freedom (Menendez1997-in), where the continuous distribution is being discretized with m discrete values. Taking sample percentiles to approximate the integral would mean taking $m = 99$. As the degrees of freedom m get larger, the chi-square distribution approaches the normal distribution.

2.4 Computation

The distance measure *wpd* between two cyclic granularities A and B is aimed to capture the strength of the structure by estimating the maximum within-group and between-group variations. Furthermore, the intended aim of *wpd* is to capture differences in categories irrespective of the distribution from which the data is generated. Hence, as a pre-processing step,the raw data is normal quantile transformed so that the quantile of the transformed data follows a standard normal distribution. The steps employed for computing the distance measure is summarized as follows:

1. Perform NQT on the measured variable v_t to obtain v_t^*.
2. Fix harmony pair (A, B).

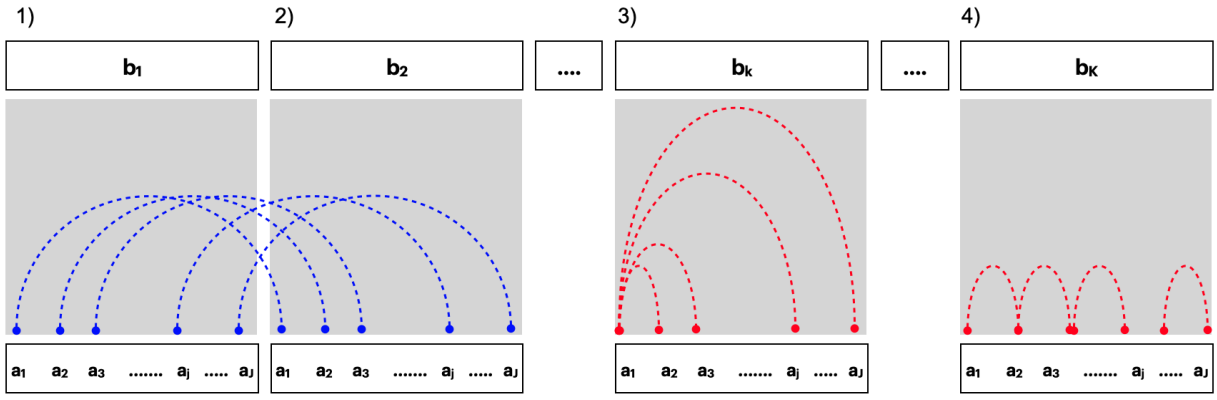


Figure 5: Within and between-facet distances shown for two cyclic granularities A and B, where A is mapped to x-axis and B is mapped to facets. The dotted lines represent the distances between different categories. Panel 1) and 2) show the between-facet distances. Panel 3) and 4) are used to illustrate 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).

3. Percentiles of v^*_{jk} are computed and stored in q_{jk}. Repeat for all pairs of categories of the form $(a_j b_k, a_{j'} b_{k'}) : j = 1, 2, \ldots, J\}, B = \{b_k : k = 1, 2, \ldots, K\}$.

4. The pairwise distances between pairs $(a_j b_k, a_{j'} b_{k'})$ denoted by $d_{(jk,j'k')} = JSD(q_{jk}, q_{j'k'})$ is computed.

5. The pairwise distances $d_{(jk,j'k')}$ is transformed using a suitable tuning parameter $(0 < \lambda < 1)$ depending on if they are within-facet(d_w) or between-facets(d_b) as follows:

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

5. The wpd is then computed as $wpd = max_{j,j',k,k'}(d*_{(jk),(j'k')}) \forall j, j' \in \{1, 2, \ldots, J\}, k, k' \in \{1, 2, \ldots, K\}$.

6. Repeat Steps 2-5 for all harmony pairs in N_C.

Pairwise distances could be within-facets or between-facets. Figure 5 illustrates how the within-facet or between-facet distances are defined. Pairwise distances are within-facets 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 5. 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 when they are considered between pairs of the form $(a_j b_k, a_j b_{k'})$. Number of between-facet distances would be ${}^K C_2 * J$ and number of within-facet distances are $K * (J-1)$ (ordered) and ${}^J C_2 * K$ (un-ordered). If the measure is intended to put more importance in pointing towards distributional differences between x categories, a $\lambda > 0.5$ should be chosen.

2.5 Properties

Simulations were carried out to explore the behavior of *wpd* under the following factors that could potentially impact the values of *wpd*:

- nx (number of levels of x-axis)
- $nfacet$ (number of levels of facets)
- λ (tuning parameter)
- ω (increment in each panel design)
- $dist$ (normal/non-normal distributions with different location and scale)
- n (sample size for each combination of categories)
- $designs$ (D_{null}, D_{var_f}, D_{var_x} and $D_{var_{all}}$)

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 details in the Supplementary section of the paper. They show that the designs D_{var_f} and D_{var_x} intersect at $\lambda = 0.5$ and hence for up-weighing designs of the form D_{var_x}, $\lambda = 0.67$ has been considered for computation of wpd in the rest of the paper.

2.5.1 Simulation design

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 wpd is obtained for each of the panels. This design corresponds to D_{null} as each combination of categories in a panel are drawn from the same distribution. Furthermore, the data is simulated for each of the panels $nsim = 200$ times, so that the distribution of wpd under D_{null} could be observed.

2.5.2 Results

Figure 6 shows the distribution of wpd plotted across different nx and $nfacet$ categories. Since under D_{null}, there is no difference in distributions across different categories, we expect the distance measure wpd to reflect that as well and have the same null distribution across categories. But Figure 6 shows that both the location and scale of the distributions change across panels. This is not desirable under D_{null} as it would mean comparisons of wpd values is not appropriate across different nx and $nfacet$. Figure 7 shows 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 $nx * nfacet$.

3 Normalization for the number of comparisons

The distribution of wpd is different for different levels of facets and x-axis levels. This is because the statistics maximum which is used to define wpd is affected by the number of comparisons (resulting pairwise distances). 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 a 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 number of comparisons and get a normalized measure. Henceforth, we call the normalized measure as wpd_{norm}. The measure wpd_{norm} could potentially lead to comparison of the measure across different panels and also help distinguishing the interesting panels from a data set. We discuss two approaches for normalization, both of which are based on the simulation results.

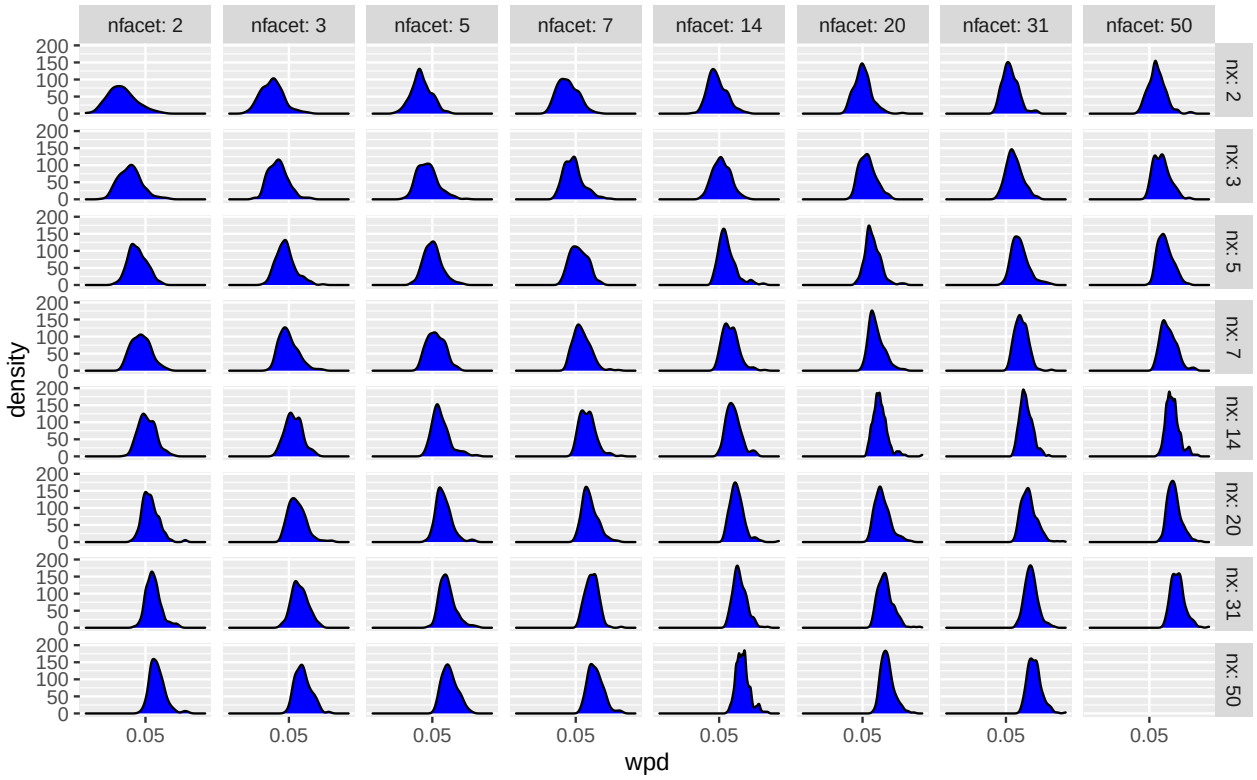


Figure 6: Distribution of *wpd* is plotted across different *nx* and $nfacet$ categories. Both shape and scale of the distribution changes for different nx and nfacet categories.

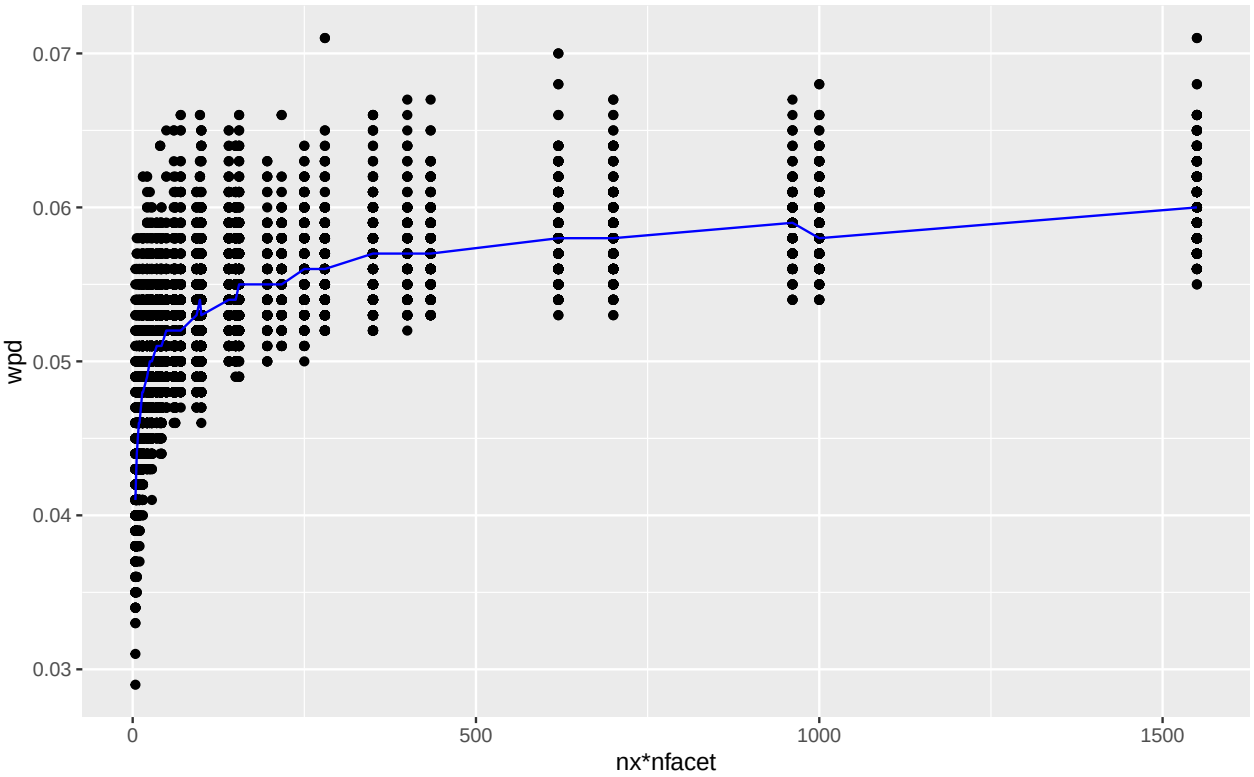


Figure 7: *wpd* is plotted against $nx * nfacet$ (the maximum number of pairwise comparisons) and the blue line represents the median of the multiple values for each $nx * nfacet$ levels. The median increases abruptly for lower values of $nx * nfacet$ and slowly for higher $nx * nfacet$.

3.1 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 for different comparisons.

3.1.1 Permutation approach

This method is somewhat similar in spirit to bootstrap or permutation tests, where the goal is to test the hypothesis that the groups under study have identical distributions. This method, in essence, accomplishes a different goal of making the location and scale of different groups (panels) same under D_{null}. The steps are as follows:

1. Compute wpd for a harmony pair (A, B) for the original measured variable v_t and store it in wpd^{orig}.
2. Consider a permutation of the original measured variable $v_t^{perm_1}$ and again compute wpd for the permuted data. Store it in wpd^{perm_1}.
3. Repeat Step 2 for a large number ($nperm = 200$) of random permutations of the data yielding $nperm$ values : $wpd^{perm_1}, wpd^{perm_2}, \ldots, wpd^{perm_{nperm}}$. Store the vector in wpd^{perm}.
4. Define $wpd_{norm}^{perm} = \frac{(wpd^{orig} - \bar{wpd^{perm}})}{sd(wpd^{perm})}$, where $\bar{wpd^{perm}}$ and $sd(wpd^{perm})$ are the mean and standard deviation of wpd^{perm} respectively.

Standardizing the variable wpd in the permutation approach ensures that the distribution of wpd_{norm}^{perm} has the same $mean = 0$ and $sd = 1$ across all comparisons under D_{null}. While this works successfully to make the location and scale similar across different nx and $nfacet$ (as seen in Figure 8), 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.

3.1.2 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|)
```

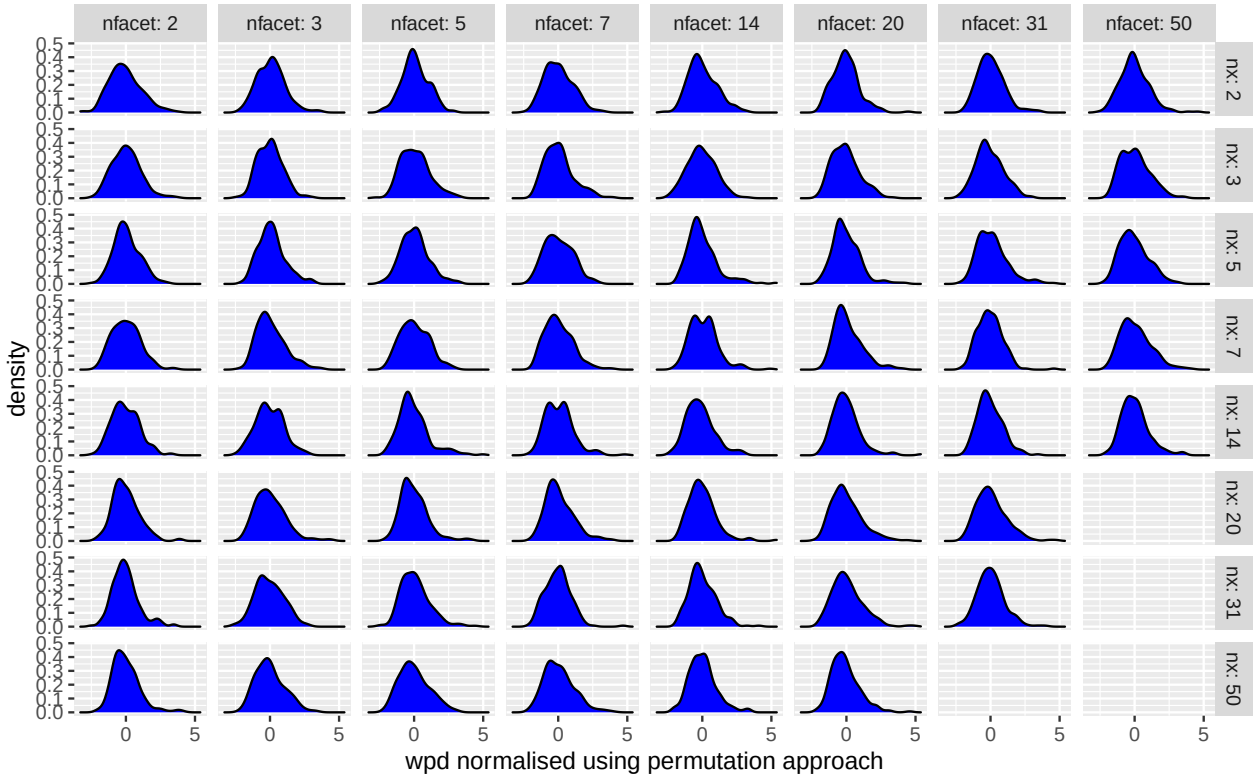
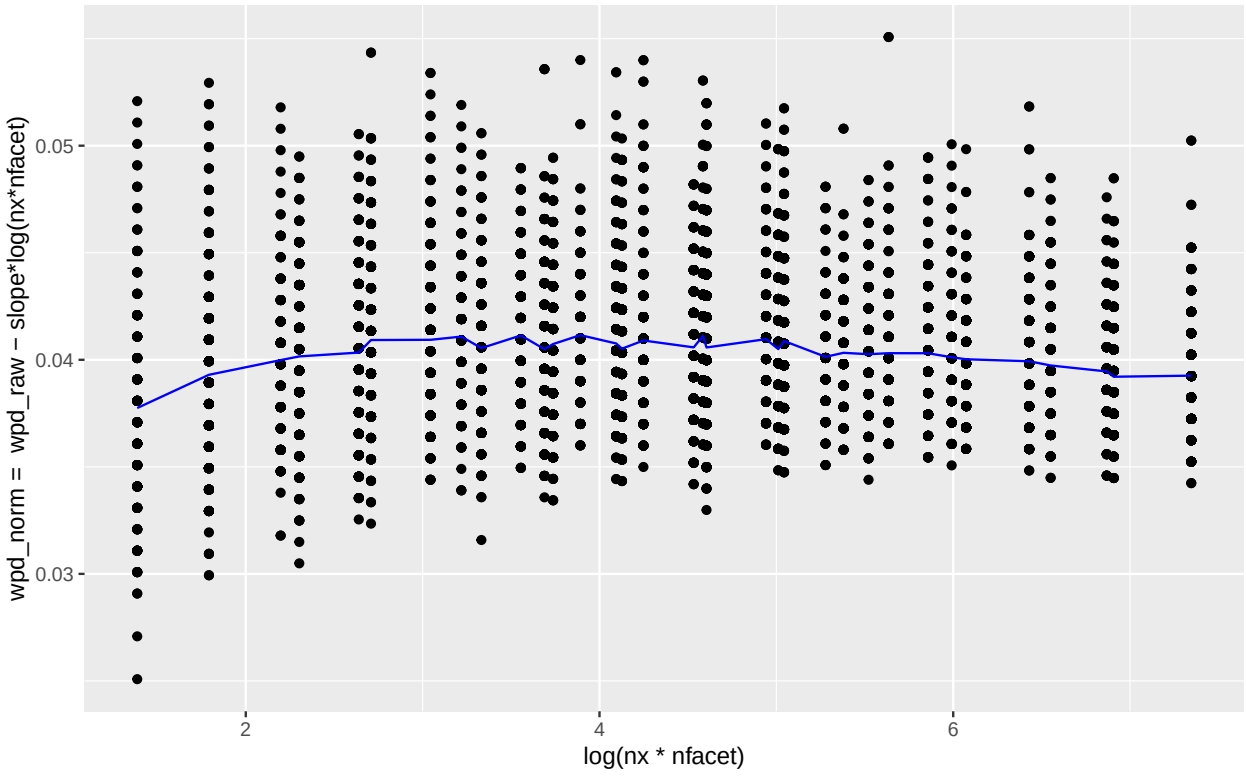


Figure 8: Distribution of wpd_{norm}^{perm} is plotted across different nx and $nfacet$ categories. Both shape and scale of the distributions are now similar for different panels under the null design.

```
#> (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 **??**



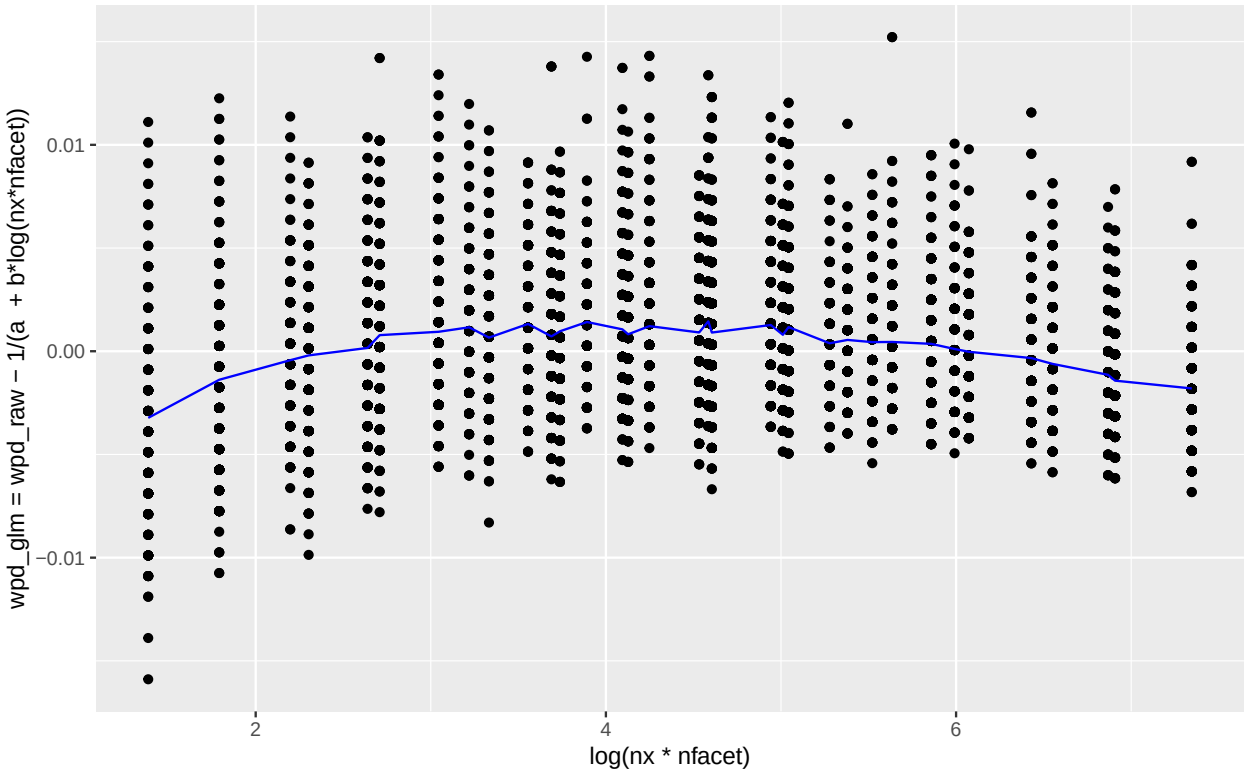
Generalised linear model

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
```

3.2 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.

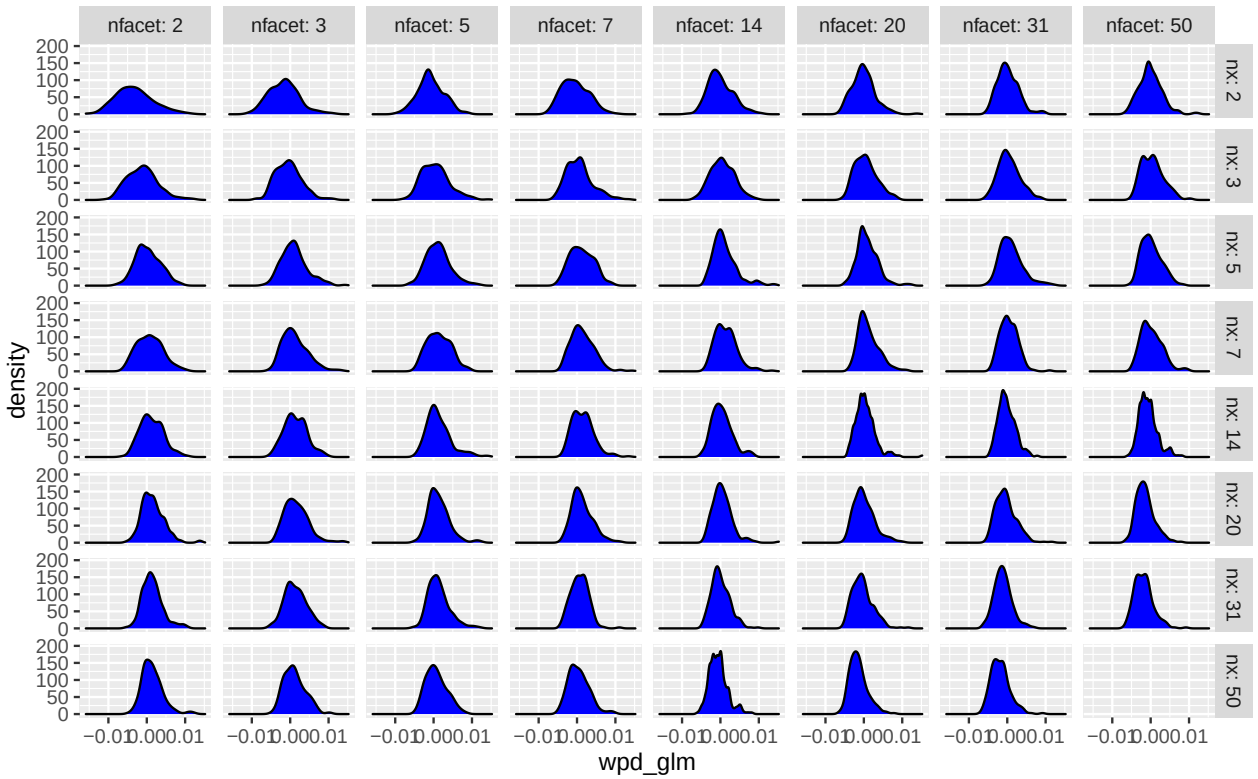


Figure 9: The distribution of wpd_{glm} is plotted. The distributions are more similar across higher nx and nfacet and dissimilar for fewer nc and nfacets.

- 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 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 10 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.

3.3 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: $nx = nfacet = \{2, 3, 5, 7, 14, 20, 31, 50\}$. $ntimes = 500$

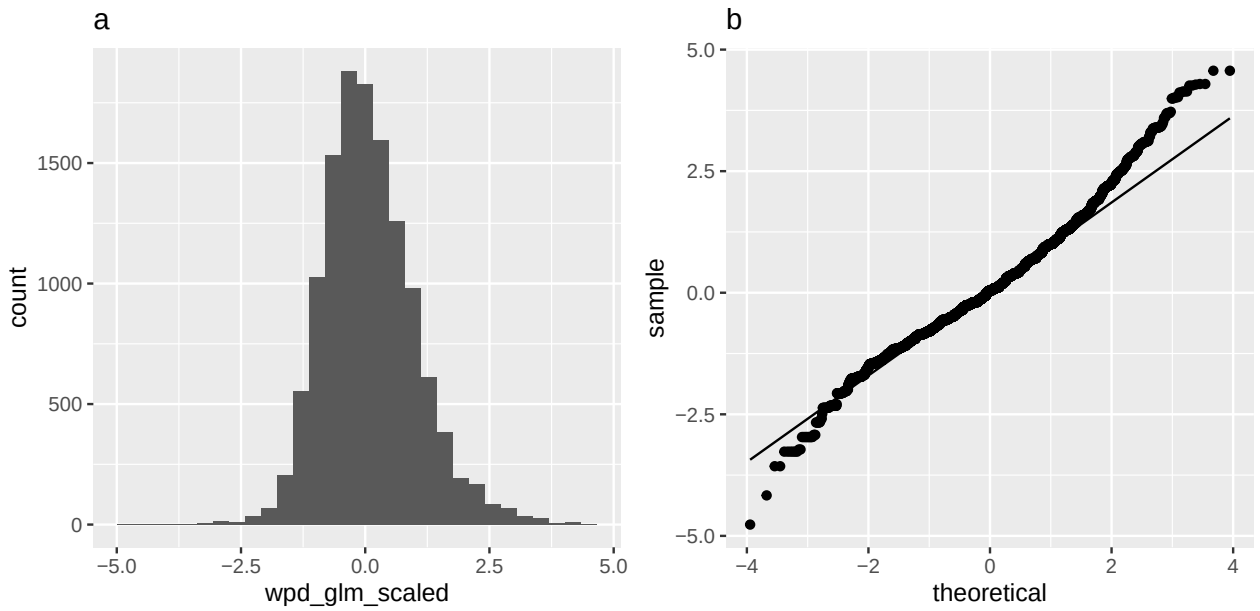


Figure 10: In panel a, the histogram of $wpd_{glm-scaled}$ is plotted. In parl b, the QQ plot is shown with the theoretical quantiles on the x-axis and $wpd_{glm-scaled}$ quantiles on the y-axis. The distribution looks symmetric and looks like normal except in the tails.

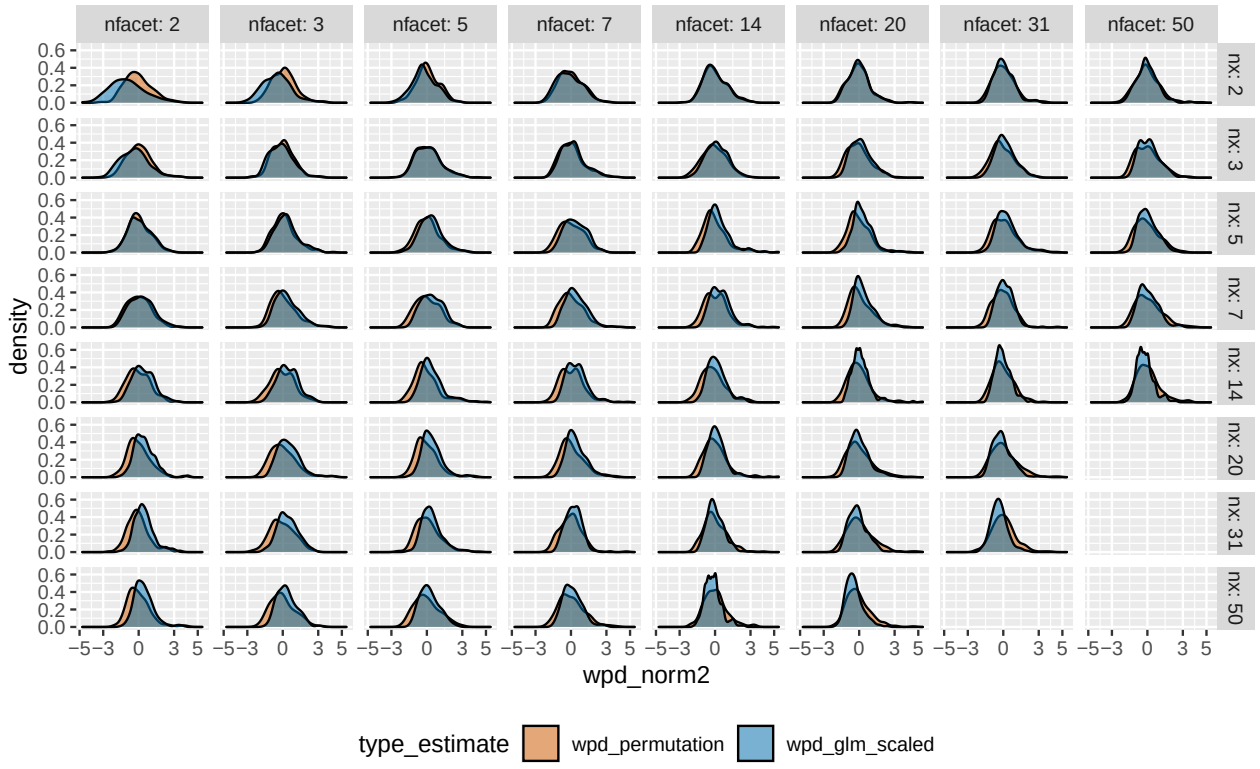


Figure 11: w$pd_permutation$ and $wpd_{glm-scaled}$ are plotted together on the same scale. They also have the same location and hence the values from these two approaches could be compared across panels. $wpd_{glm-scaled}$ would be used to normalise wpd_{raw} for higher nx and $nfacet$ and $wpd_{glm-scaled}$ would be used for smaller levels to alleviate the problem of computational time.

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 μ_{jk} denotes the mean of the combination $(a_j b_k)$, then $\mu_{j\cdot} = \mu + j\omega$ (for design D_{var_x}) and $\mu_{\cdot k} = \mu + k\omega$ (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 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

In this section, we provide a method to select important harmonies by eliminating all harmonies for which patterns are not significant across x or facet categories through randomization test. Randomization tests (permutation tests) generates a random distribution by re-ordering our observed data and allow to test if the observed data is significantly different from any random distribution. 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.

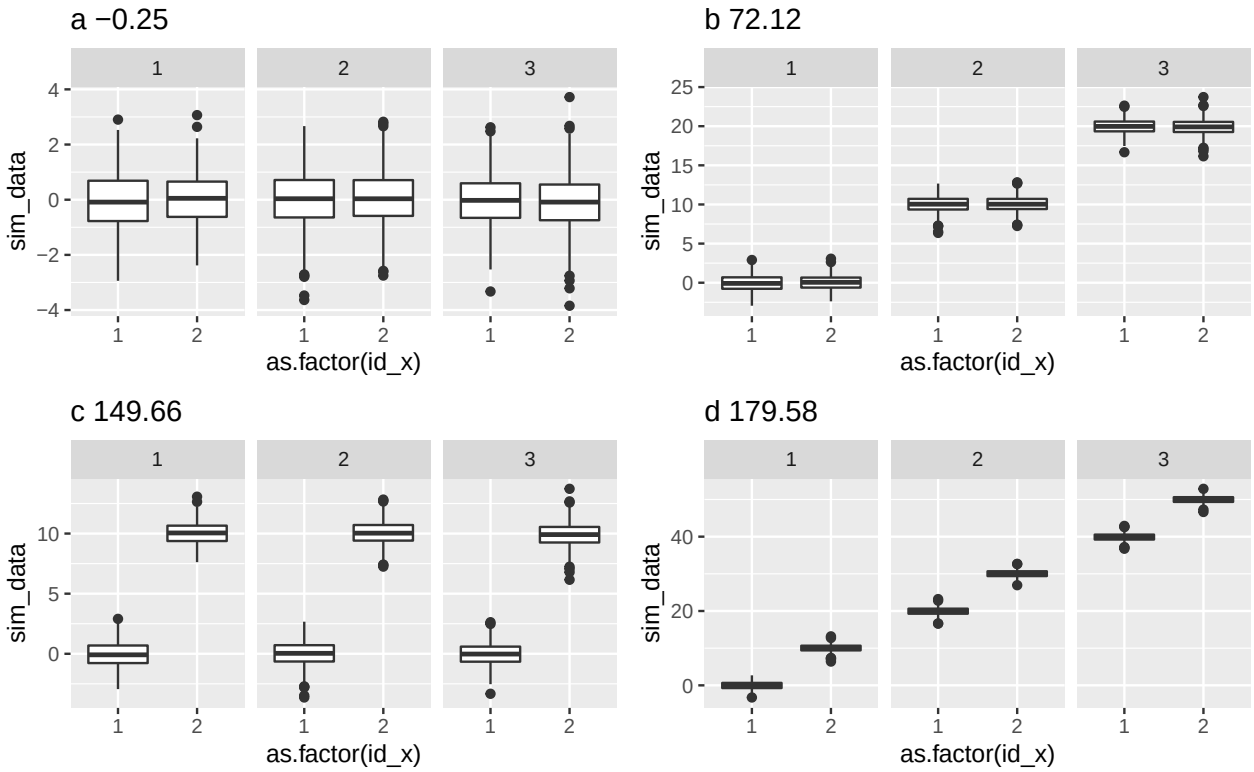
4.1 Choosing a threshold

Typically, a randomization test involves calculating a test statistic, randomly shuffling the data and calculating the test statistic several times to obtain a distribution of the test statistic. We will use this procedure to test if there is any interesting pattern captured by the harmonies, which essentially implies if wpd_{norm} is significantly different from zero. The percentages of times the wpd_{norm} obtained from the permuted data is greater than or equal to the observed wpd_{norm} is the p-value. The randomization test is described as follows:

- **Input:** All harmonies of the form $\{(A, B), A = \{a_j : j = 1, 2, \ldots, J\}, B = \{b_k : k = 1, 2, \ldots, K\}\}$ with A placed across x-axis and B across facets $\forall (A, B) \in N_C$.
- **Output:** Harmony pairs (A, B) for which wpd_{norm} is significant.

1. Fix harmony pair (A, B).
2. Given the data; $\{v_t : t = 0, 1, 2, \ldots, T-1\}$, the wpd_{norm} is computed and is represented by wpd_{obs}.
3. From the original sequence a random permutation is obtained: $\{v_t^* : t = 0, 1, 2, \ldots, T-1\}$.
4. wpd_{norm} is computed for the permuted sequence of the data and is represented by wpd_{perm_1}.
5. Steps (3) and (4) are repeated a large number of times M (M = 200).
6. For each permutation, one wpd_{perm_i} is obtained. Define $wpd_{sample} = \{wpd_{perm_1}, wpd_{perm_2}, \ldots, wpd_{perm_M}\}$.
7. Repeat Steps (1-6) for all harmony pairs.
8. 95^{th} percentile of wpd_{sample} obtained for all harmony pairs is computed and stored in $wpd_{threshold}$.
9. If $wpd_{obs_{A,B}} > wpd_{threshold}$, harmony pair (A, B) is selected, otherwise rejected.

4.2 Results



5 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.

6 Application to 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 is a tsibble containing 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 these households are shown in a linear time scale in Figure 12. It is evident from the range of the demand data that these households vary in consumption levels as well as in their temporal patterns. In the left panel of Figure 12 (a), 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 to select and rank the important harmonies.

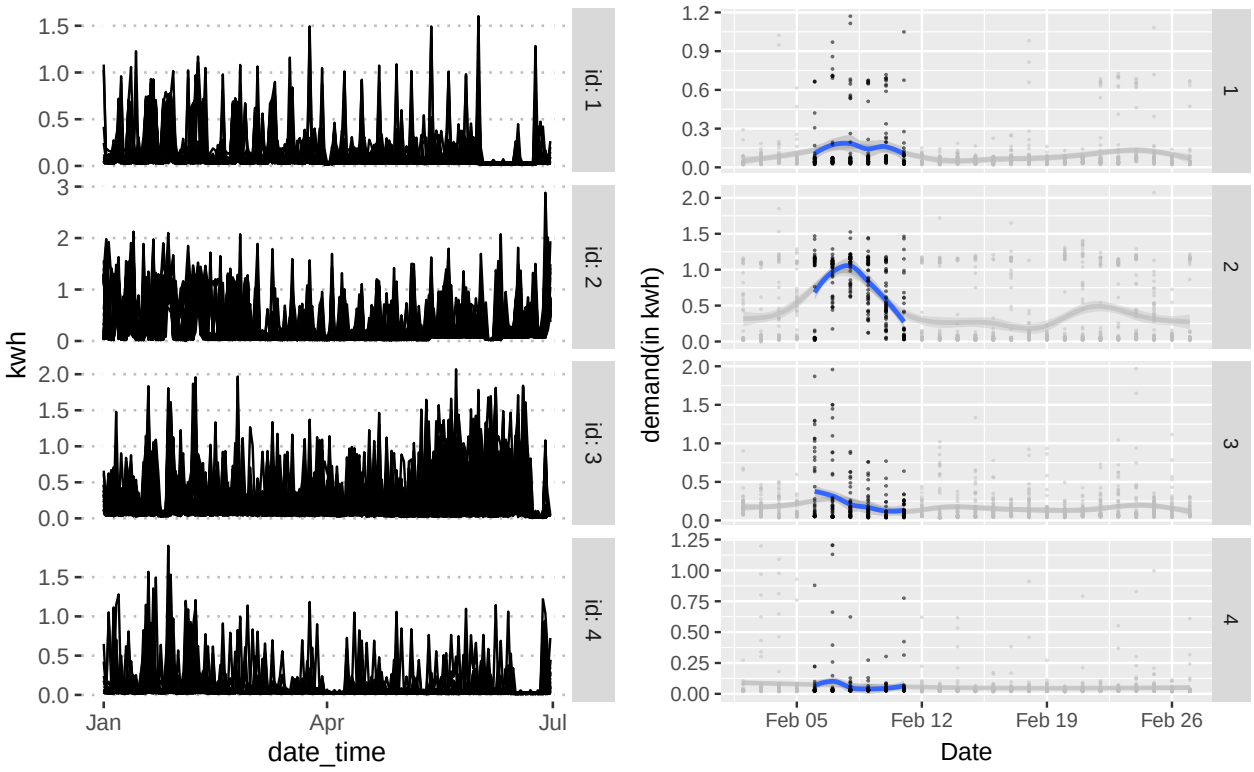
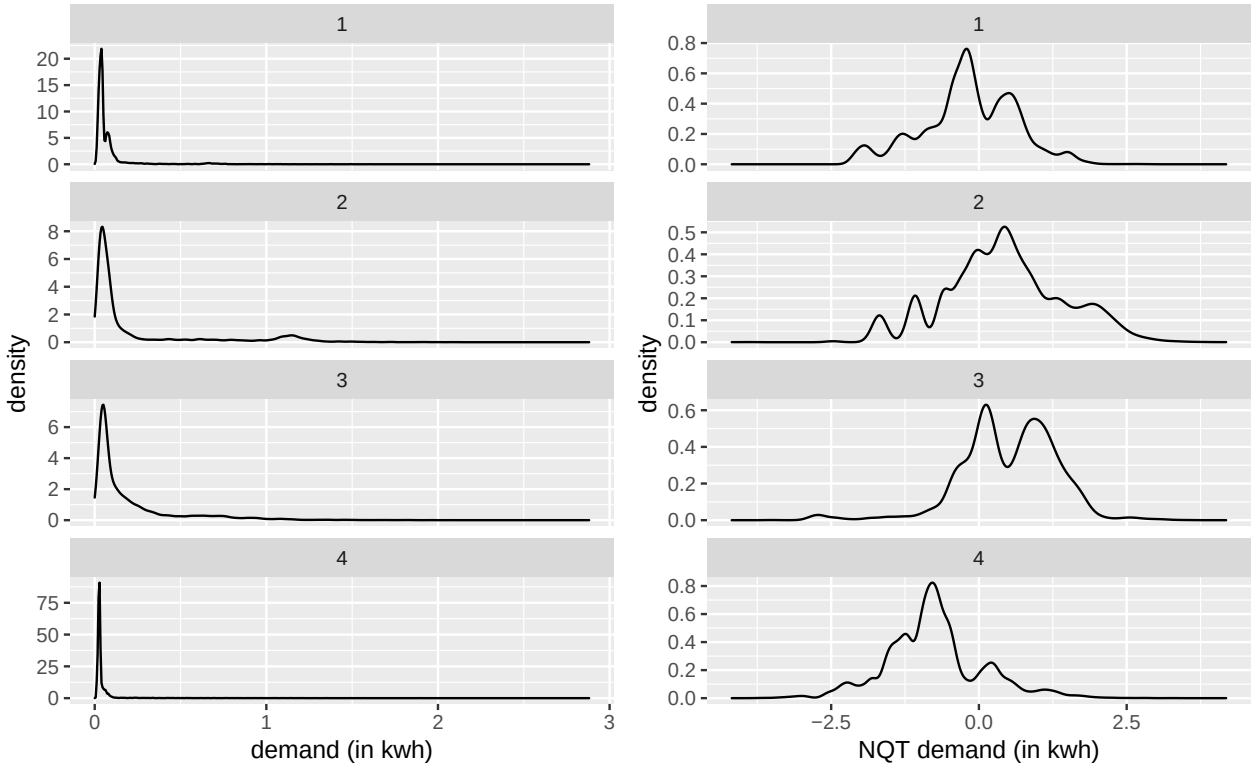


Figure 12: Electricity demand for four households are shown in different facets from Jan-18 to Jun-18 in Fig a and it has been zoomed in for Feb-18 in Fig b, where a week in Feb-18 has been highlighted. From the scales of Fig a, it is apparent that they have different level of consumption but all of 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 $y_{i,t}$ 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 symmetric. Let $y*_{i,t}$ denote the normal-quantile transformed electricity demand for i^{th} household for time period t. Technically, this $y*_{i,t}$ is the linear granularity corresponding to half-hour since the interval of this data is 30 minutes. We consider coarser linear granularities like hour, day, week and month from the commonly used Gregorian calendar. Considering 4 linear granularities hour, day, week, month in the hierarchy table, the number of cyclic granularities is $N_C = (4 * 3/2) = 6$. We obtain cyclic granularities namely "hour_day", "hour_week", "hour_month", "day_ week", "day_month" and "week_month", read as "hour of the day", etc. Further, we add cyclic granularity day-type("wknd wday") to capture weekend and weekday behavior. Thus, we consider 7 cyclic granularities in the set of cyclic granularities of interest defined by C_{N_C}. We then compute the set of harmonies H_{N_C} from C_{N_C} by choosing pairs of cyclic granularities that are not clashes using procedures described in (Gupta et al. 2020). The paper also shows how the number of harmonies could be identified among 42 pair of granularities by removing clashes. This section shows how we could refine the search further by only looking at significant harmonies.

Suppose $(A, B) \in H_{N_C}$ be a harmony pair where $A = \{a_j : j = 1, 2, \ldots, J\}$ and $B = \{b_k : k = 1, 2, \ldots, K\}$ with A placed across x-axis and B across facets. Suppose $q^{i,p}_{A,j}$ denote the quantiles with probability p for the of the i^{th} household for j^{th} category of the cyclic granularity A. Similarly, $q^{i,p}_{B,k}$ denotes the same for the k^{th} category of the cyclic granularity B. Sample quantiles were computed at $p = 0.01, 0.02, \ldots, 0.99$. Jensen-Shannon distances are computed between $q^{i,p}_{A,j}$ and $q^{i,p}_{B,k}$ for each $j \in J, k \in K$ to obtain within-facet and between-facet distances. A tuning parameter of $\lambda = 0.67$ has been considered to upweigh the within-facet distances and down-weigh the between facet distances and the maximum of them are obtained to compute wpd_{raw}. It is further normalized using the approach described in Section **??**. This entire process is repeated for all harmony pairs $(A, B) \in H_{N_C}$ and for each households $i \in i = \{1, 2, 3, 4\}$. The harmony pairs are then arranged in descending order and the important ones with significance level 1%, 5% and 10% are highlighted with ***, ** and * respectively.

Table **??** shows the rank of the harmonies for different households **??** shows the bubble chart for the four households with size as the value of wpd and color indicating how significant they are. This is a quick

summary of which harmonies are significant and also which cyclic granularities are important for the context.

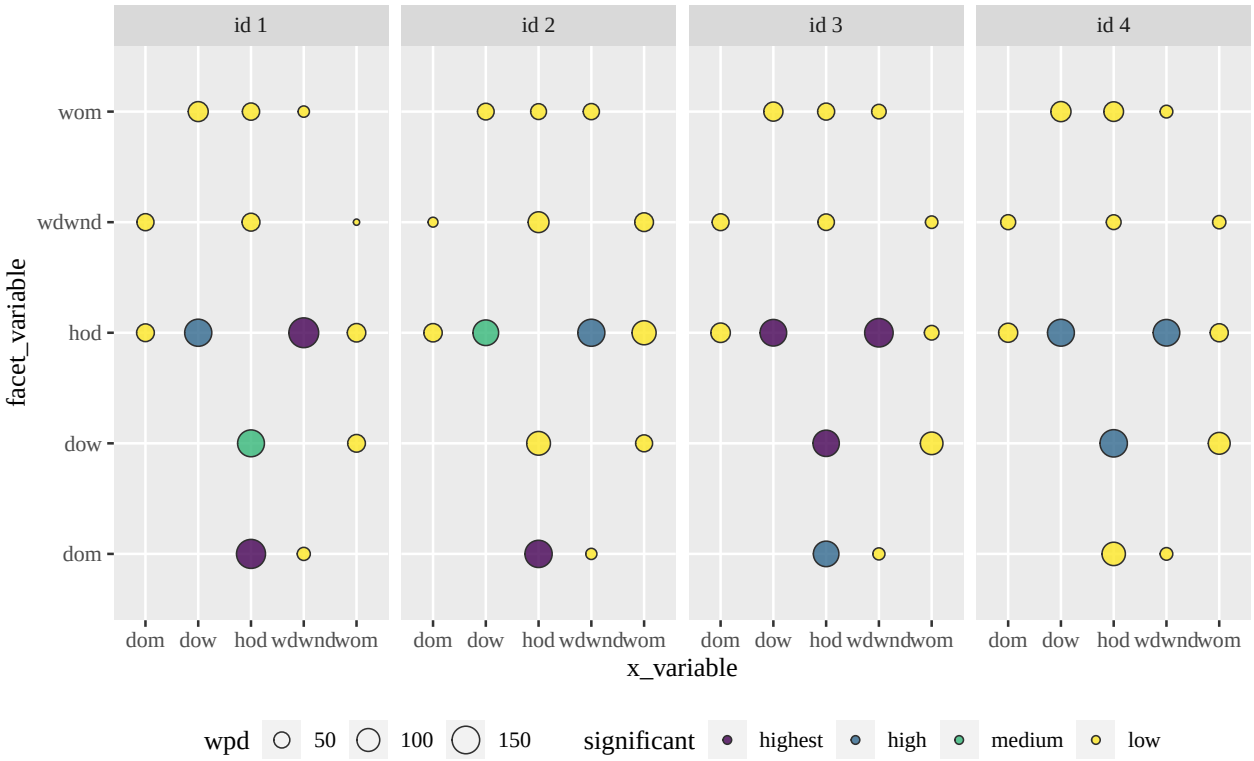


Figure 13: Harmony pairs are shown for all household ids. The size represent how important the pair is and the colour represents if the harmonies are at all important or not. Visualizing the pairs in this way helps us to see the important cyclic granularities along the x-axis and facet along with the information that which ones amongst them should be analyzed together.

Table 1: Ranking of the harmonies for the four households are shown. The first four harmonies are unanimously important for all households, after which there is differences in the rankings for most households.

facet variable	x variable	facet_levels	x levels	id 1	id 2	id 3	id 4
hod	wdwnd	24	2	1 ***	2 **	1 ***	3 **
dom	hod	31	24	2 ***	1 ***	4 **	4
hod	dow	24	7	3 **	3 *	2 ***	2 **
dow	hod	7	24	4 *	5	3 ***	1 **

7 Discussion

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.

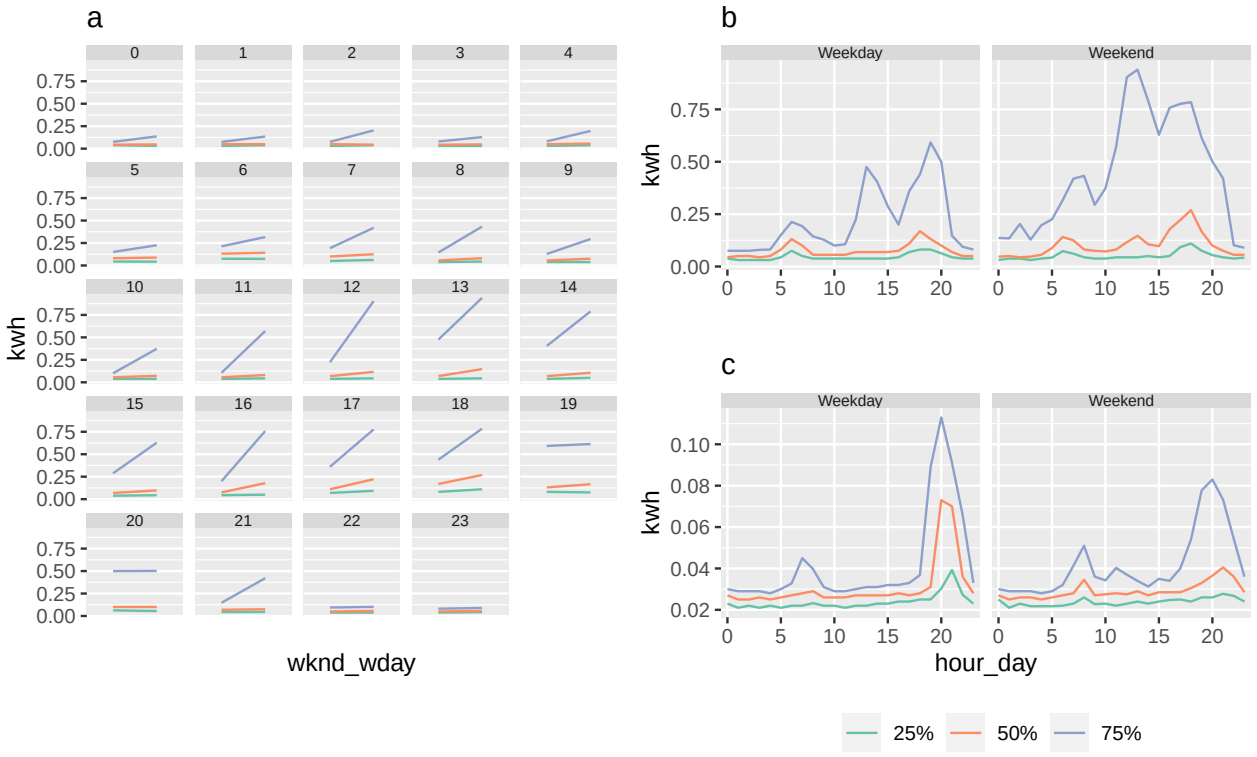


Figure 14: Distribution of energy demand shown for household id: 2 across hod in x-axis and wd-wnd in facets in a) and just the reverse in b). In c), distribution of energy demand for household id:4 shown across hod and wd-wnd. It can be seen that the differences in distributions are more apparent when viewed in a) as compared to b). It seems like there is more difference in the distributions of hod for b) compared to c). This also confers with the value of the normalised measure shown in Figure 11.

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.

8 Appendix

8.1 Smart meter application to 8 datasets from energy competition

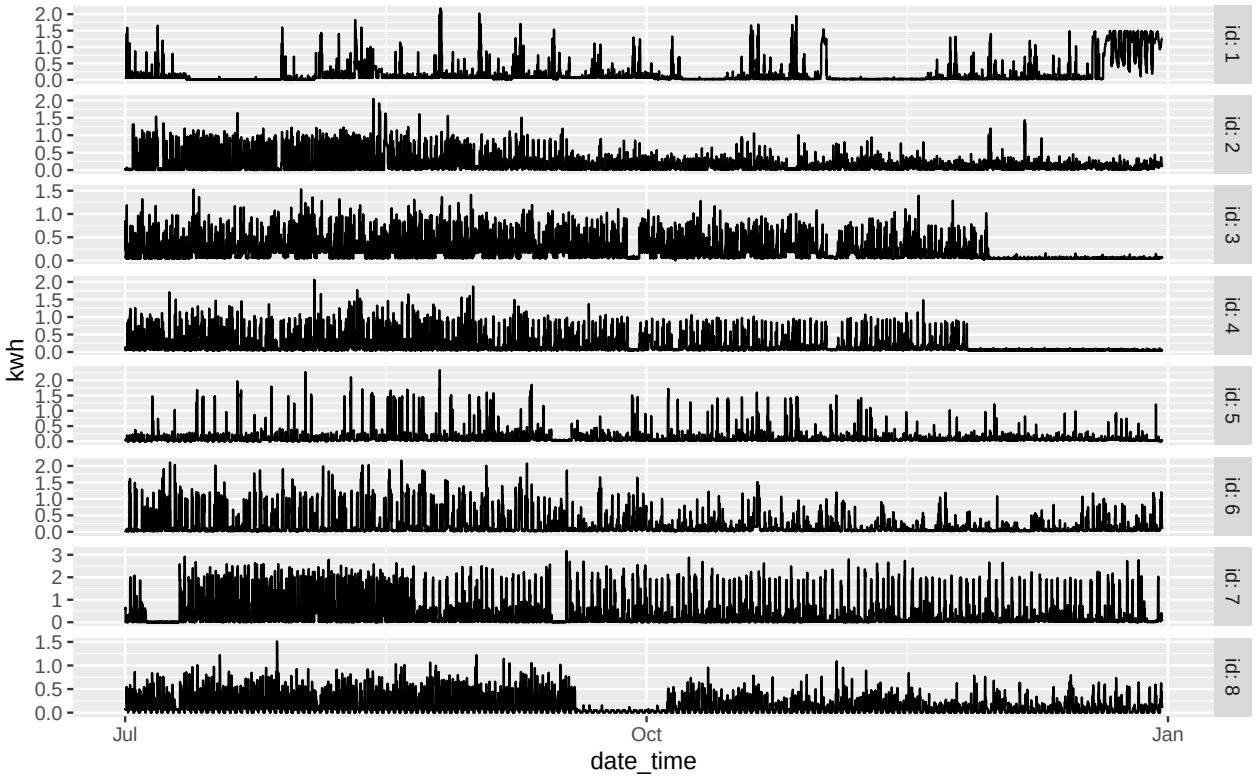


Figure 15: Electricity demand for four households are shown in different facets from Jan-18 to Jun-18 in Fig a and it has been zoomed in for Feb-18 in Fig b, where a week in Feb-18 has been highlighted. From the scales of Fig a, it is apparent that they have different level of consumption but all of 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.

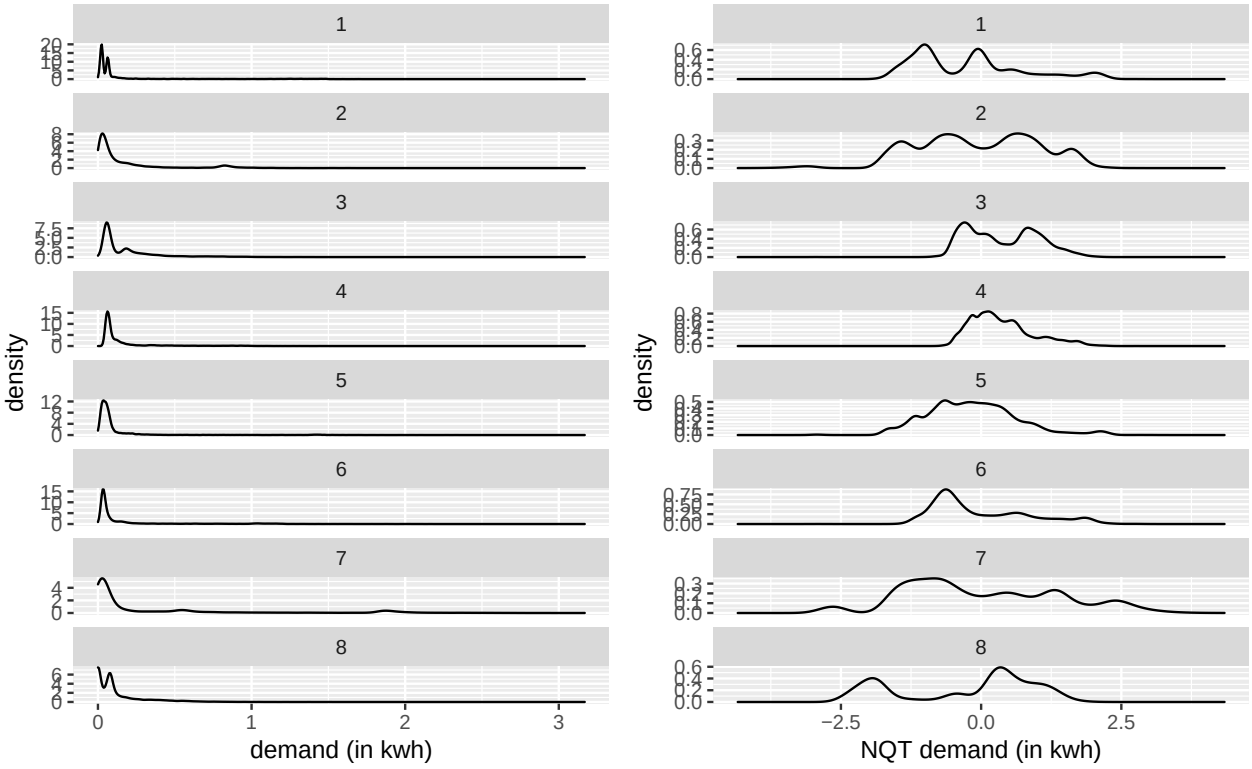


Table 2: Ranking of the harmonies for the four households are shown. The first four harmonies are unanimously important for all households, after which there is differences in the rankings for most households.

facet variable	x variable	facet_levels	x levels	id 1	id 2	id 3	id 4	id 5	id 6	id 7	id 8
hod	wdwnd	24	2	1 ***	2 *	1 **	2 **	3	1 **	3	3 *
dom	hod	31	24	2 ***	4	3 **	3 **	4	3 *	4	6
wdwnd	hod	2	24	3 **	10	7	7	6	8	8	10
hod	wom	24	5	4	9	6	5	5	5	5	5
wom	wdwnd	5	2	5	14	14	10	12	9	12	13
hod	dow	24	7	6	1 ***	2 **	1 ***	1 *	2 **	2 **	1 **
wdwnd	wom	2	5	7	12	13	8	7	7	10	12
dow	hod	7	24	8	3	4 **	4 **	2	4 *	1 ***	2 **
hod	dom	24	31	9	7	10	13	10	10	9	4
wom	dow	5	7	10	6	8	9	8	6	7	9
dow	wom	7	5	11	5	9	11	11	12	6	7
wom	hod	5	24	12	8	5	6	9	11	11	8
dom	wdwnd	31	2	13	13	11	12	14	14	14	14
wdwnd	dom	2	31	14	11	12	14	13	13	13	11

- check transformation and scalar approaches are same

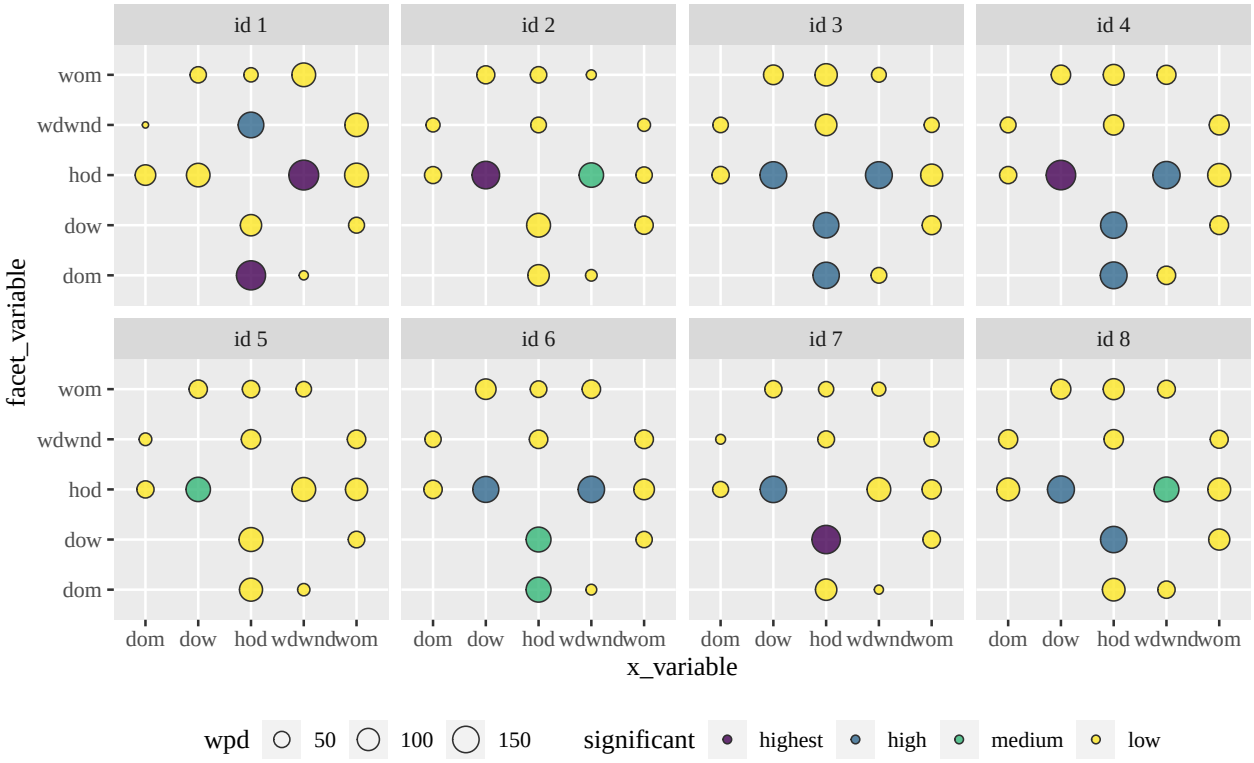


Figure 16: Harmony pairs are shown for all household ids. The size represent how important the pair is and the colour represents if the harmonies are at all important or not. Visualizing the pairs in this way helps us to see the important cyclic granularities along the x-axis and facet along with the information that which ones amongst them should be analyzed together.

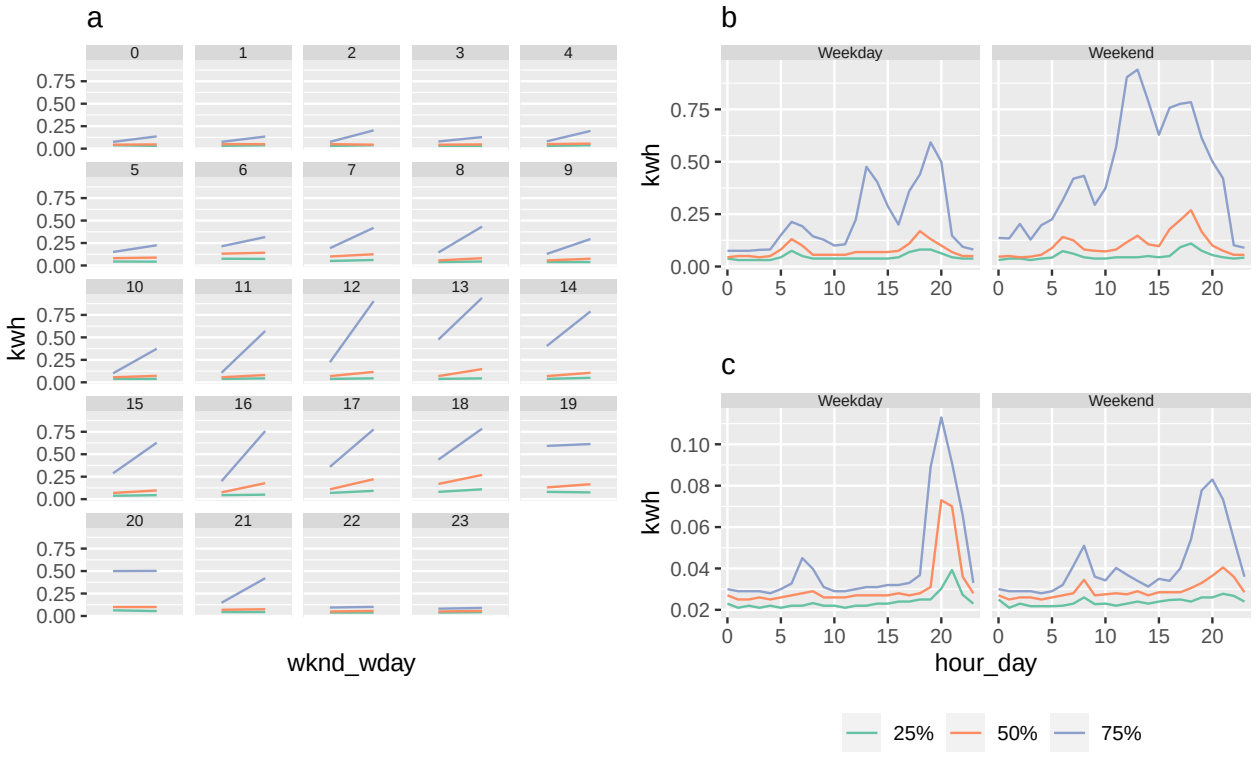
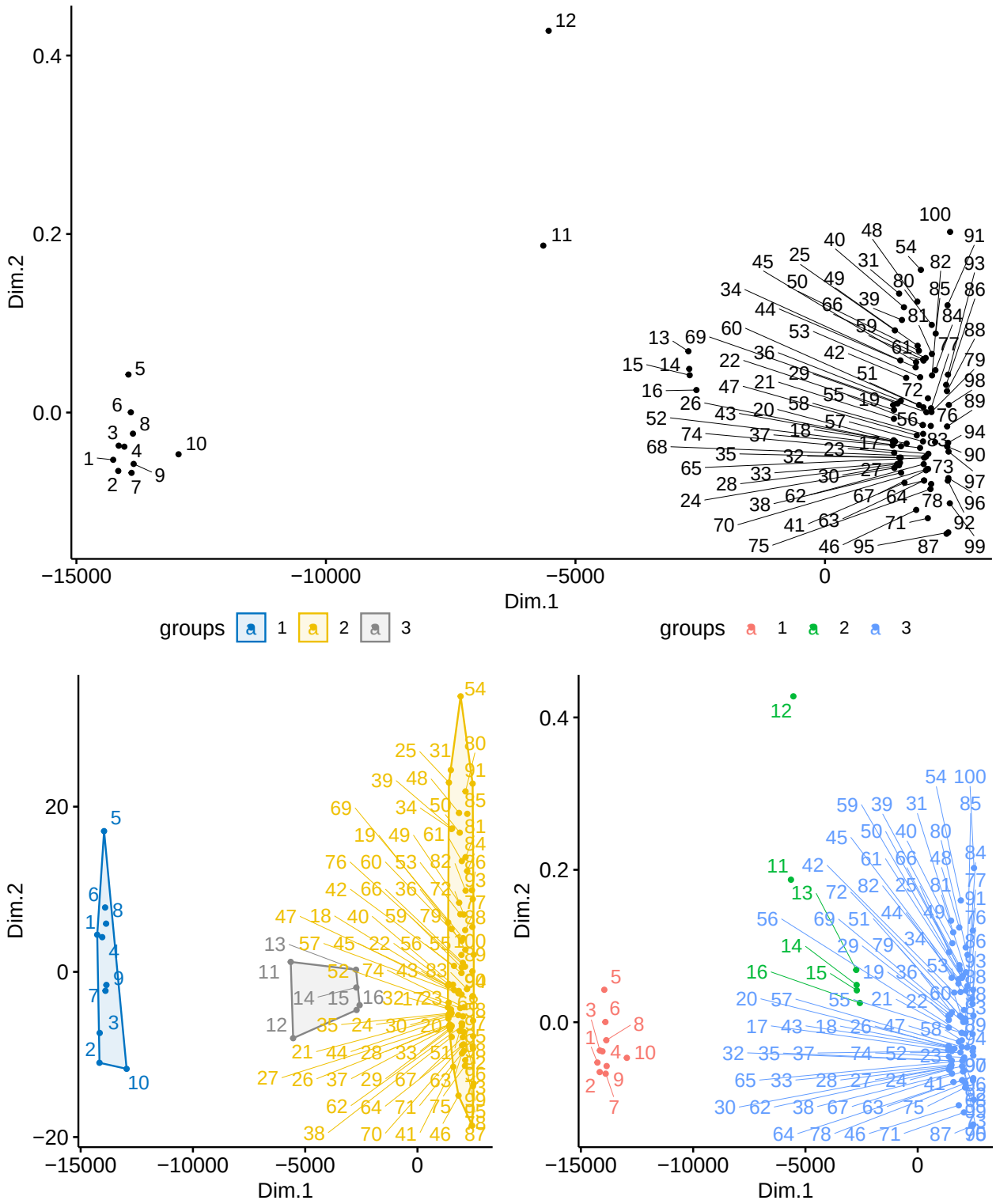


Figure 17: Distribution of energy demand shown for household id: 2 across hod in x-axis and wd-wnd in facets in a) and just the reverse in b). In c), distribution of energy demand for household id:4 shown across hod and wd-wnd. It can be seen that the differences in distributions are more apparent when viewed in a) as compared to b). It seems like there is more difference in the distributions of hod for b) compared to c). This also confers with the value of the normalised measure shown in Figure 11.



8.2 Null distribution

8.2.1 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.

8.2.2 Normalised maximum distances follow standard Gumbel distribution

8.2.3 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 ξ (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 ξ and variance $\sigma_n^2 = \frac{1}{4}[f(\xi)]^2(2n+1)$, where $f(x) = F'(x)$ is the pdf (probability density function).

8.3 Power

8.4 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.

8.4.1 Varying distribution across facet

8.4.2 Varying distribution across x-axis

8.4.3 Varying distribution across both facets and x-axis

8.4.4 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.

Hyndman, R. J., X. Liu, and P. Pinson. 2018. "Visualizing Big Energy Data: Solutions for This Crucial Component of Data Analysis." *IEEE Power and Energy Magazine* 16 (3): 18–25. https://doi.org/10.1109/MPE.2018.2801441.

Hyndman, Rob J, and Yanan Fan. 1996. "Sample Quantiles in Statistical Packages." *Am. Stat.* 50 (4): 361–65.

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.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

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.