Selecting and ranking interesting pairs of cyclic temporal granularities

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 (1) 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

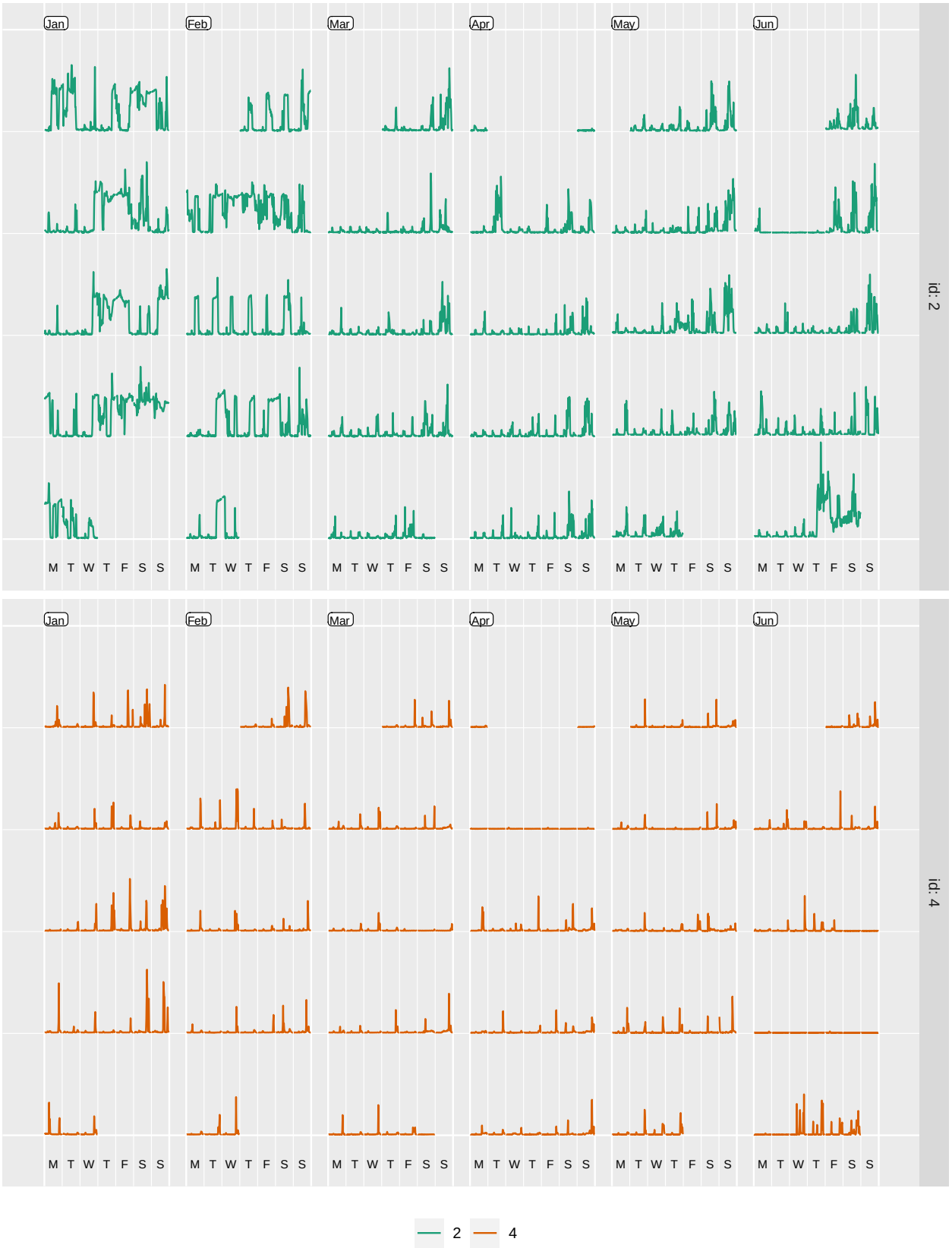


Figure 1: Calendar display.

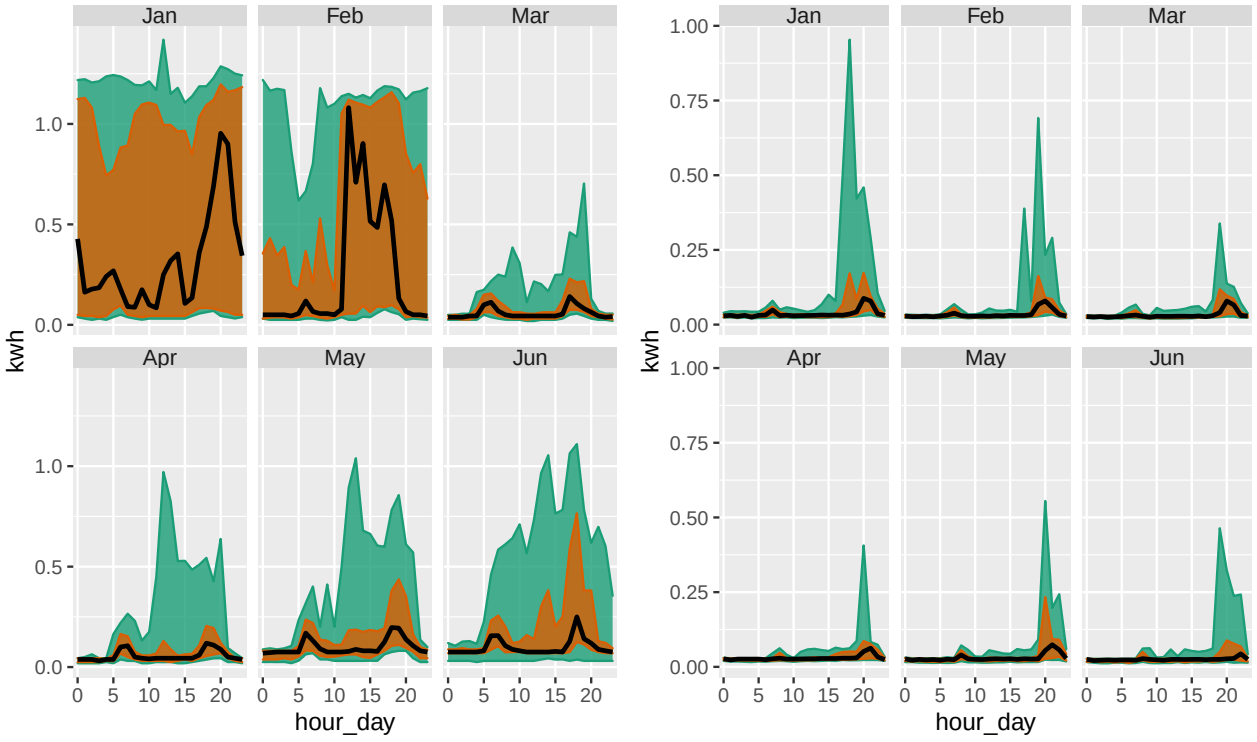


Figure 2: something

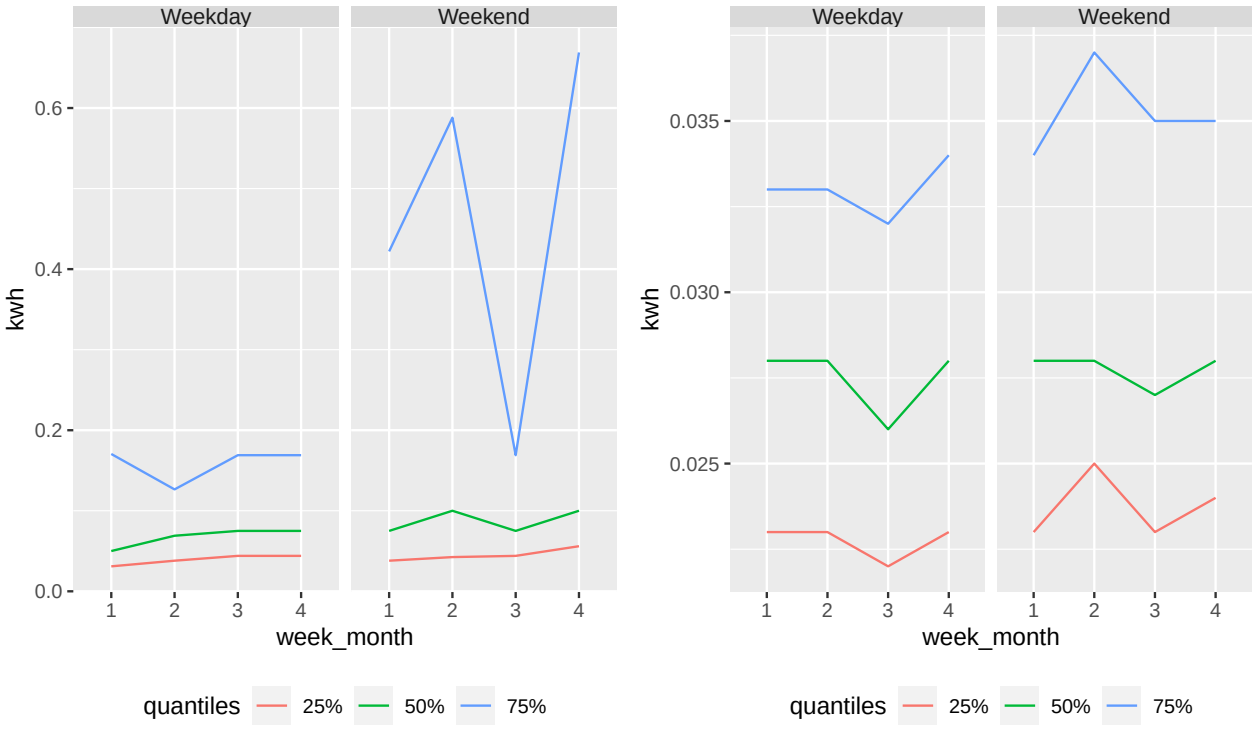


Figure 3: something2

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 Distance measure

We are interested in assessing structure in probability distributions of the measured variable across bivariate cyclic granularities. We propose a measure called Median Maximum Pairwise Distances (MMPD) 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 4. Each of these figures have the same panel design with 2 x-axis categories and 4 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 and different for different facet categories. Figure 4b exhibits an exact opposite situation where distribution between the x-axis categories within each facet is different but they are same across facets. Figure 4d 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 MMPD which stands for Median Maximum Pairwise Distances. In the next section we shall see how we go about computing this measure.

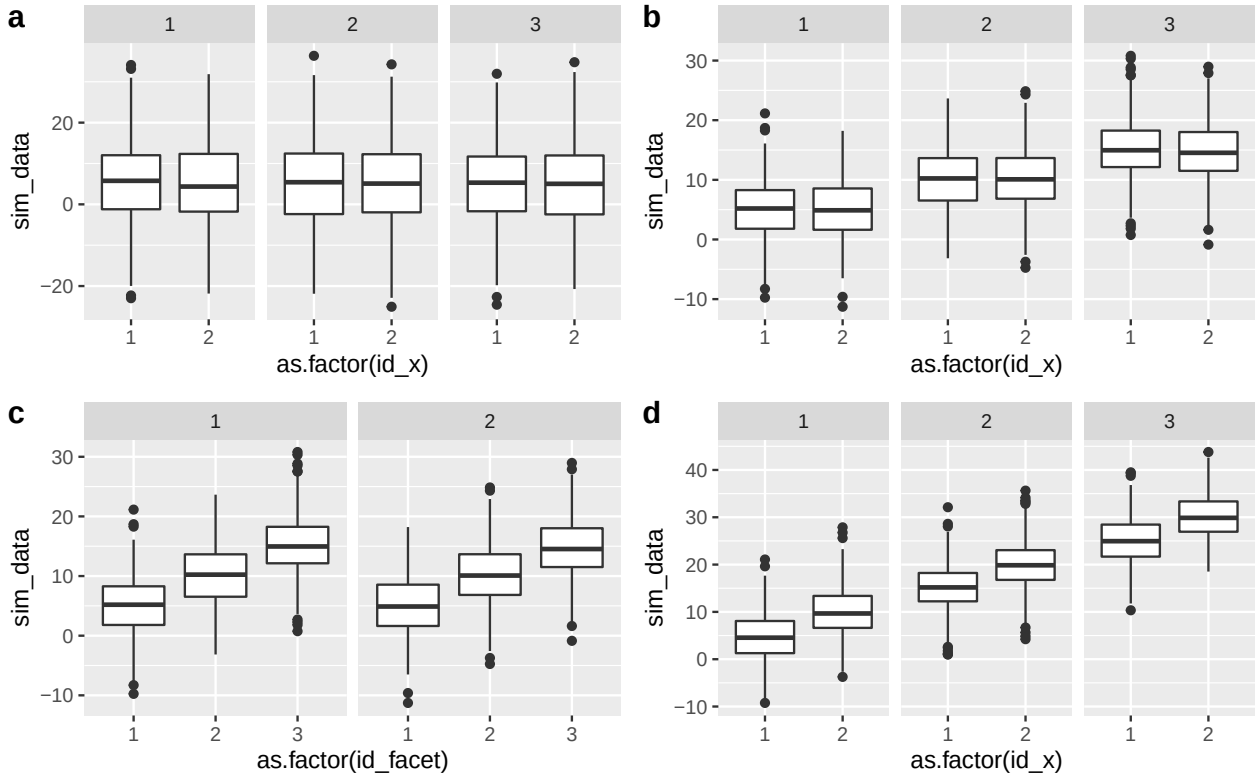


Figure 4: A graphical display with two categories mapped to x-axis and 4 categories mapped to facets with the distribution of a continuous random variable plotted on the y-axis. Display a is not interesting as the distribution of the continuous rv does not depend across x-axis or facet categories. Display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

2.2 Methodology

2.2.1 Characterising distribution

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.

2.2.2 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.

2.2.3 Notations Definitions

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 $(a_j b_k, a_{j'} b_{k'})$ b e 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 5 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 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 (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 $\lambda > 1$, the pairwise distances $d_{(jk,j'k')}$ are transformed based on the distance type as follows:

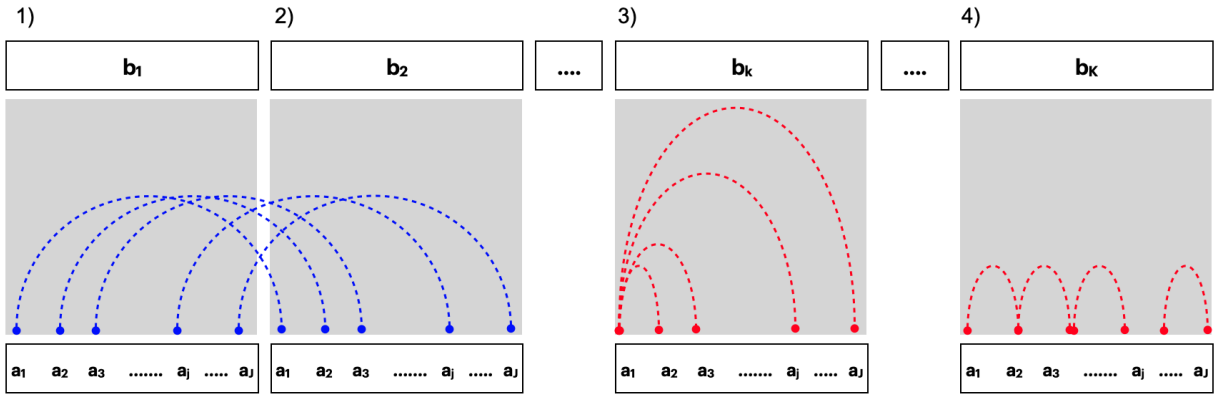


Figure 5: Within and between-facet distances shown for two cyclic granularities A and B, where A is mapped to x-axis and B is mapped to facets. The dotted lines represent the distances between different categories. Panel 1) and 2) show the between-facet distances. Panel 3) and 4) are used to illustrate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

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

The maximum weighted pairwise distances are defined as:

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

.

–>

3 Behavior of raw wpd (weighted pairwise distances)

Most of the behavior of the measure wpd was studied via simulation. The simulations explore how wpd performs under various designs and parameters and its limitations. To study the behavior of wpd, 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)
- 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)
- $nsim$ (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)

3.1 Null design

This section explores the behavior of wpd in designs where there is no difference in distribution between x and facet categories. We have considered different initial distributions to study the impact of initial distribution under the null setup. Since the measure wpd is essentially set up to detect differences in distributions irrespective of underlying distribution, it would be ideal if it has minimal dependency on the type, location and scale of the initial distribution. To that end, some pre-processing of the data is preferred to bring it to the same scale, location and type. The Normal Score Transform or NQT has been applied in various fields of geo-statistics in order to make most asymmetrical distributed real world measured variables more treatable and normal-like.

Following the work of Krzysztofowicz (1997) the empirical NQT involves the following steps: 1. Sort the sample of measured variable X from the smallest to the largest observation $x_{(1)}, \ldots, x_{(i)}, .., x_{(n)}$. 2. Estimate the cumulative probabilities $p_{(1)}, \ldots, p_{(i)}, .., p_{(n)}$. using a plotting position like $i/(n+1)$ such that $p_{(i)} = P(X \leq x_{(i)})$. 3. Transforming each observation $x_(i)$ of X into observation $y(i) = Q^{-1}(p(i))$ of the standard normal variate Y , with Q denoting the standard normal distribution and Q^{-1} its inverse, applying discrete mapping.

Further, we want to study the distribution of wpd for different nx and $nfacet$.

3.1.1 Simulation setup

Using two types of distributions, viz. normal and gamma (non-normal), we generated observations for each combination of nx and $nfacet$ from the following sets: $nx = \{2, 3, 5, 7, 14, 20, 31, 50\}$ and $nfacet = \{2, 3, 5, 7, 14, 20, 31, 50\}$ 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 λ is set to 0.67, since we want to up-weigh the within-facet distances and that of ω is set to 0, since there is no significant differences between distributions in the null case. Observations were generated for each type of distribution changing the shape and scale to study the effect of shape, scale and type of distribution on wpd. The set of distributions considered for this purpose is $N(0,1), N(5,1), N(0,5), \Gamma(0.5,1), \Gamma(2,1)$. Each of the scenario is run $nsim = 200$ times to see the distribution of wpd values for each scenario.

Figure 6 and 7 show distribution of wpd for different initial normal and gamma distribution to study the effect of changing location and scale on the distribution of wpd. It seems like the distribution changes across different facet and x levels but look the same for each panel, which implies wpd value is unaffected by the change in location and scale of the the normal distribution. Figure 8 shows how mean and sd of the distribution of wpd changes with the increasing x and facet levels. It seems like both mean and standard deviation are affected more by change in the x-axis levels than the facet levels. Figure 9 gives another way to look at the effect of changing facet and x levels on the distribution of wpd for an initial distribution $N(0,5)$. Clearly, the location of the distribution shifts to the right for increasing x levels and scale is different for low and high values of facet levels with a longer left tail for lower facet levels. We have considered alternate distribution, location and scale for this simulation setting, but do not present the results as they are behaviorally similar.

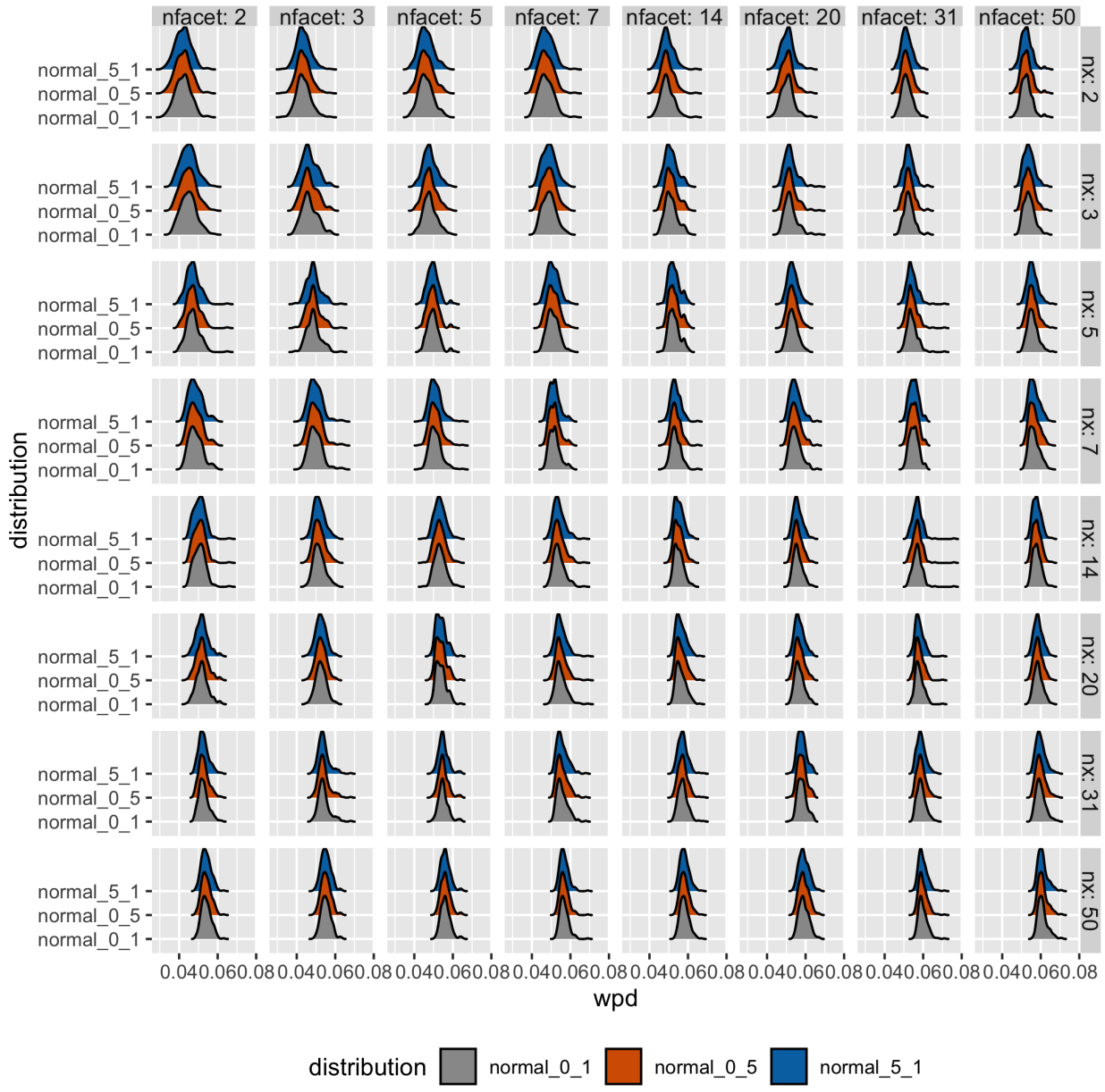


Figure 6: Ridge plots of raw wpd is shown for N(0,1), N(5,1) and N(0,5) distribution. The densities change across different facet and x levels but look same for each panel, which implies wpd value is unaffected by the change in mean and standard deviation of the normal distribution

Figure 7: Ridge plots of raw wpd is shown for Gamma(0.5,1), Gamma(2,1) distribution. The densities change across different facet and x levels but look same for the two distributions, which implies wpd value is unaffected by the change in the shape paramter of the gamma distribution

Table 1: Simulation setup for a panel with 3 facet levels and 2 x-axis levels for different designs starting from an initial distribution N(0, 1) for the combination (a_1, b_1).

x	facet	D_{var_f}	D_{var_x}	$D_{var_{all}}$
a_1	b_1	N(0, 1)	N(0, 1)	N(0, 1)
a_2	b_1	N(0, 1)	N(1, 1)	N(1, 1)
a_1	b_2	N(1, 1)	N(0, 1)	N(2, 1)
a_2	b_2	N(1, 1)	N(1, 1)	N(3, 1)
a_1	b_3	N(2, 1)	N(0, 1)	N(4, 1)
a_2	b_3	N(2, 1)	N(1, 1)	N(5, 1)

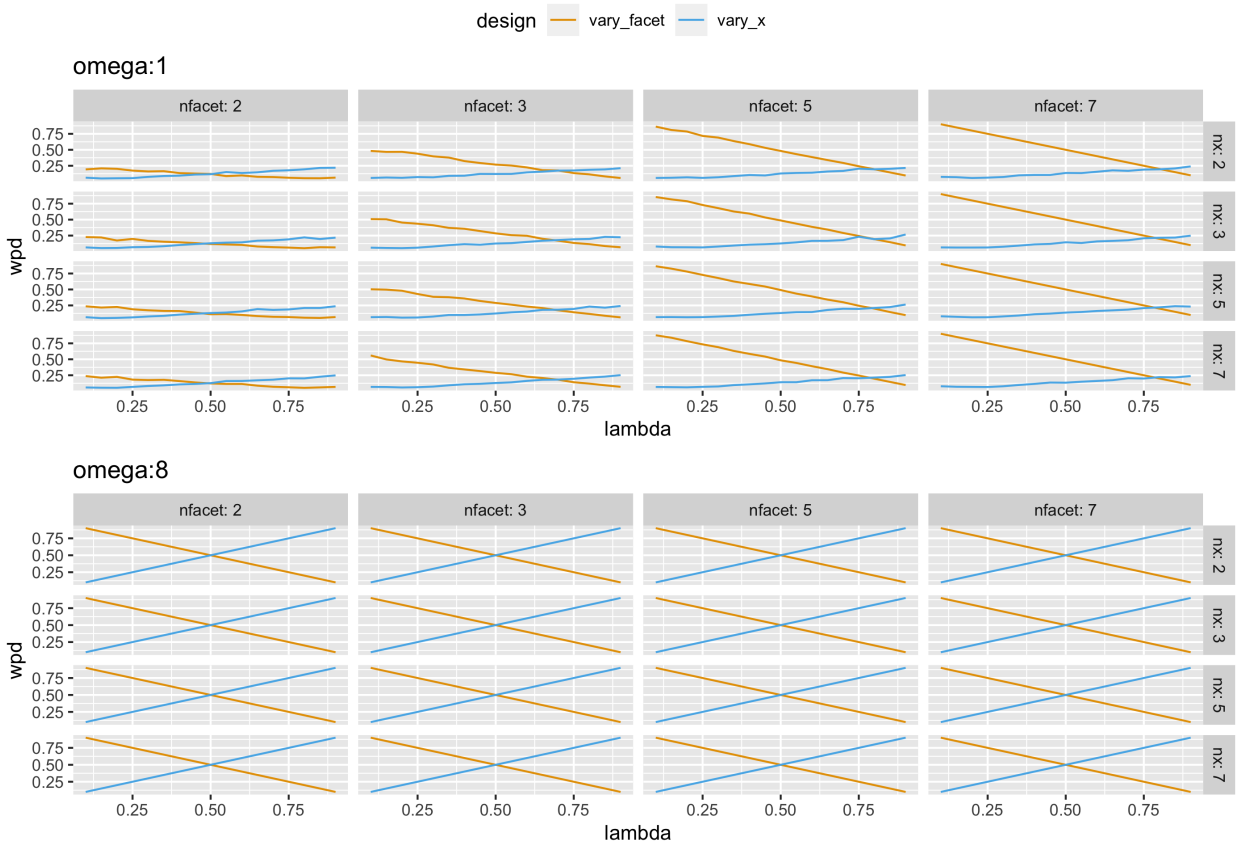
3.2 Alternate designs

This section explores the behavior of wpd in designs where there is infact difference in distribution between facet categories (D_{var_f}) or across x-categories (D_{var_x}) or both ($D_{var_{all}}$). Since it is established in the last section that after preprocessing the data through NQT, the initial distribution does not have a role to play in the values of wpd, we proceed with a N(0,1) distribution only. Supposably, the tuning parameter (λ) and increment parameter (ω) should impact the values of wpd.

3.2.1 Simulation setup

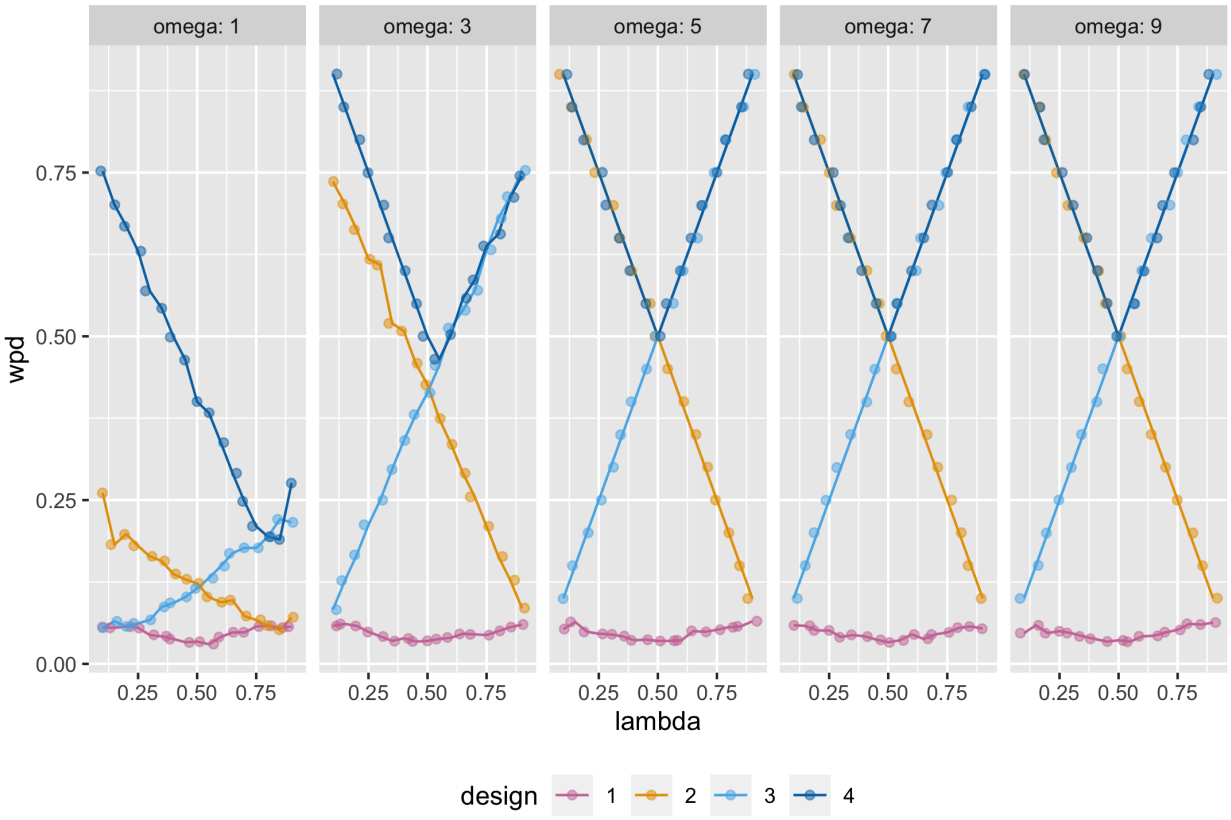
Using $\omega = \{1, 2, \ldots, 10\}$ and $\lambda = seq(from = 0.1, to = 0.9, by = 0.05)$, observations are drawn from a N(0,1) distribution for each combination of nx and $nfacet$ from the following sets: $nx = nfacet = \{2, 3, 5, 7, 14, 20, 31, 50\}$. $ntimes = 500$ is assumed for this setup as well. 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 μ_{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}). Table 1 shows an example of how initial distributions are assigned in a panel with $nfacet = 3$ and $nx = 2$ for different designs using $\omega = 1$.

We now discuss the effect of different tuning parameters λ and ω with the help of two alternate designs (D_{var_x} and D_{var_f}). Figure **??** displays different values of λ for a relatively small and higher ω for $nx = nfacet = \{2, 3, 5, 7, 9\}$. The λ for which the two designs intersect will then be chosen as the optimal λ that could then be weighed appropriately to up-weigh the within-facet distances and down-weigh the between-facet distances. The design D_{var_x} increases linearly with increasing λ, whereas D_{var_f} decreases linearly with increasing λ. This is expected as wpd has a linear relationship with λ by construction. The two designs mostly intersect at $\lambda = 0.5$ for a higher value of ω. Figure 11 shows how the value of λ changes with increasing ω for $\omega = \{1, 2, \ldots, 10\}$ and $\lambda = seq(from = 0.1, to = 0.9, by = 0.05)$.



How value of lambda changes with increasing omega and mean for fixed nx, nfacet and standard deviation?

How value of lambda changes with increasing omega and sd for fixed nx, nfacet and mean?



Distribution across facet and x categories

From Figure **??** and **??** shows the MMPD distribution is different for different levels of facets and x-axis levels. With increasing number of facets, the location of the distribution shift rightwards and with increasing x-axis levels, the scale of the distribution reduces.

Median and maximum distances are affected by the number of categories considered and hence the distance measure $MMPD_{raw}$, which is a combination of median and maximum would also be influenced by the number of levels. It would have higher values if C_i or C_j has higher levels. We would ideally want a higher value of the measure if there is significant difference between distributions across facet or x-axis categories, and not because the number of categories are higher. In Figures **??** and **??**, both the cyclic granularities C_i and C_j are considered such that their levels vary in the range $(2, 5, 7, 9, 14, 21, 30, 45)$. Each of these combinations is considered a panel for each of which $MMPD_{raw}$ has been constructed 100 times to compute the distribution.

Therefore, in order to compare $MMPD_{raw}$ 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.

3.3 Sample size

3.4 Number of permutations

4 Normalisation

In an attempt to make the densities **??** same across different levels, we want to start by making their scale and location same. In that regard, we shuffle the data multiple times and compute the sampling distribution

of $mmpd_{raw}$ for each combination of facet and x-axis levels.

The null hypothesis is that the two cyclic granularities do not differ on the outcome (i.e., that the outcome is observed independently of treatment assignment). When we permute the outcome values during the test, we therefore see all of the possible alternative treatment assignments we could have had and where the mean-difference in our observed data falls relative to all of the differences we could have seen if the outcome was independent of treatment assignment. While a permutation test requires that we see all possible permutations of the data (which can become quite large), we can easily conduct "approximate permutation tests" by simply conducting a vary large number of resamples. That process should, in expectation, approximate the permutation distribution.

Finally, we run some simulation experiments to see if normalisation works. If normalisation has worked, it should work for both x levle and facet-levels.

Thus these median maximum pairwise distances need to be normalized for different harmonies in a way that eliminates the effect of different levels, consequently enabling comparison across different harmonies. The Fisher–Tippett–Gnedenko theorem in the field of Extreme Value Theory states that the maximum of a sample of iid random variables after proper re-normalization can converge in distribution to only one of Weibull, Gumbel or Freschet distribution, independent of the underlying data or process.

4.1 Methodology

The mean and sd for each combination of facet and x levels could be computed by shuffling the data repeatedly and obtaining the distribution of $MMPD_{raw}$ for different combinations. If $MMPD_{raw}$ is then adjusted by the location and scale of this distribution, we could assume that the distribution of the resultant $MMPD_{norm}$ will have same distribution across different combinations of x and facet levels. All data is repeatedly shuffled $nperm$ times in random manner to obtain the measure $MMPD_{raw}$. Then the $MMPD_{norm}$ is obtained by scaling the observed value by the mean and sd of the distribution of $MMPD_{raw}$ obtained from the permuted data.

Step 7 in the Algorithm defined in Section **??** need to be revised in order to compute $MMPD_{norm}$ instead of $MMPD_{raw}$. Step 1-6 stays the same. The entire algorithm will have to repeated for all harmony pairs considered in the context.

A simulation study is conducted to see if this methodology is working.

4.2 Simulation study

The behavior of the measure is monitored and understood through simulation experiments. A single simulation consisting of computational operations on a panel generating MMPD the value of which represents to what extent a pair of cyclic granularities would be interesting when displayed in the design. In comparing different simulation runs, we distinguish between different distributions and simulation scenarios (also iterations may be).

4.2.1 Simulated designs

D1. Same distribution of the measured variable across all x-axis and facet categories D2. Different distribution across facet categories but same across x-axis categories D3. Different distribution across x-axis categories but same across facet categories D4. Different across both x-axis and facet categories

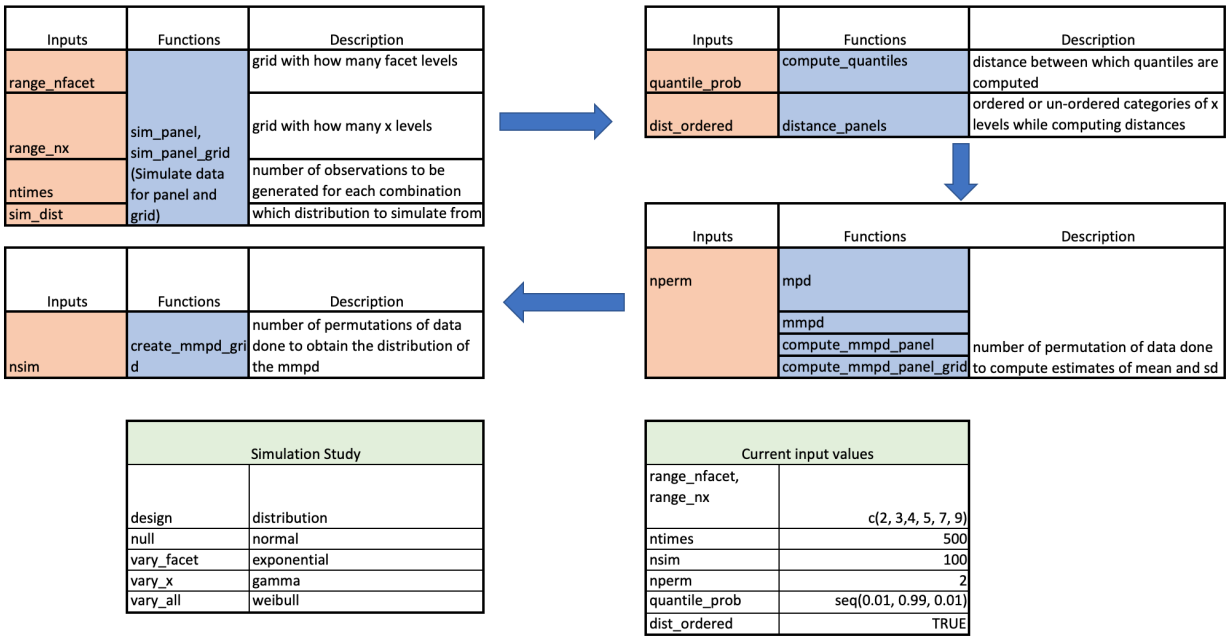
Each of these case scenarios tried against Normal and a non-normal (Gamma) distribution to check if underlying distribution has a role to play.

4.2.2 Environment

R version 4.0.1 (2020-06-06) is used with platform: x86_64-apple-darwin17.0 (64-bit) running under: macOS Mojave 10.14.6

4.2.3 Experimental set up

This section describes the design of simulation runs, in terms of the scenarios simulated, the number of permutations used to compute estimates and to display the distribution of MMPD.



Observations from Normal(5, 10) and Gamma(0.5, 0.2) are3 chosen in all the designs.

–> –>

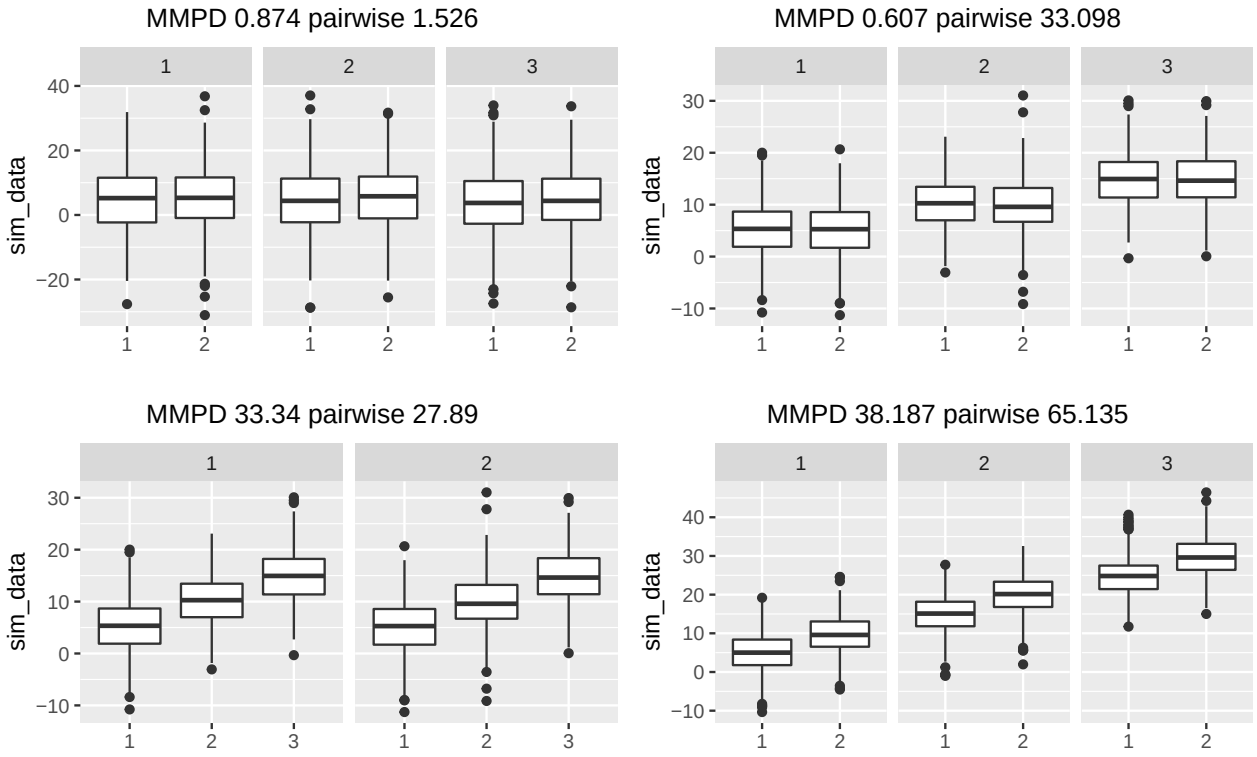
4.3 Results

4.3.1 Null distribution

Under the null hypothesis, all the combinations of x-axis and facets are obtained from the same distribution. We need to study the distribution of the mmpd values under the null hypothesis to see if comparison of their values across different x-axis and facet levels are at all possible. So, in turn it needs to be checked that if normalisation worked.

Currently, we see in Figure **??** that the normalisation works fine for each facet since the distributions for each column are equal irrespective of the distribution type. But it does not work along the x-axis.

4.3.2 Performance of normalised MMPD



5 The statistical test

5.1 Algorithm

for computation for all harmony pairs

Assumption: random permutation without considering ordering (global)

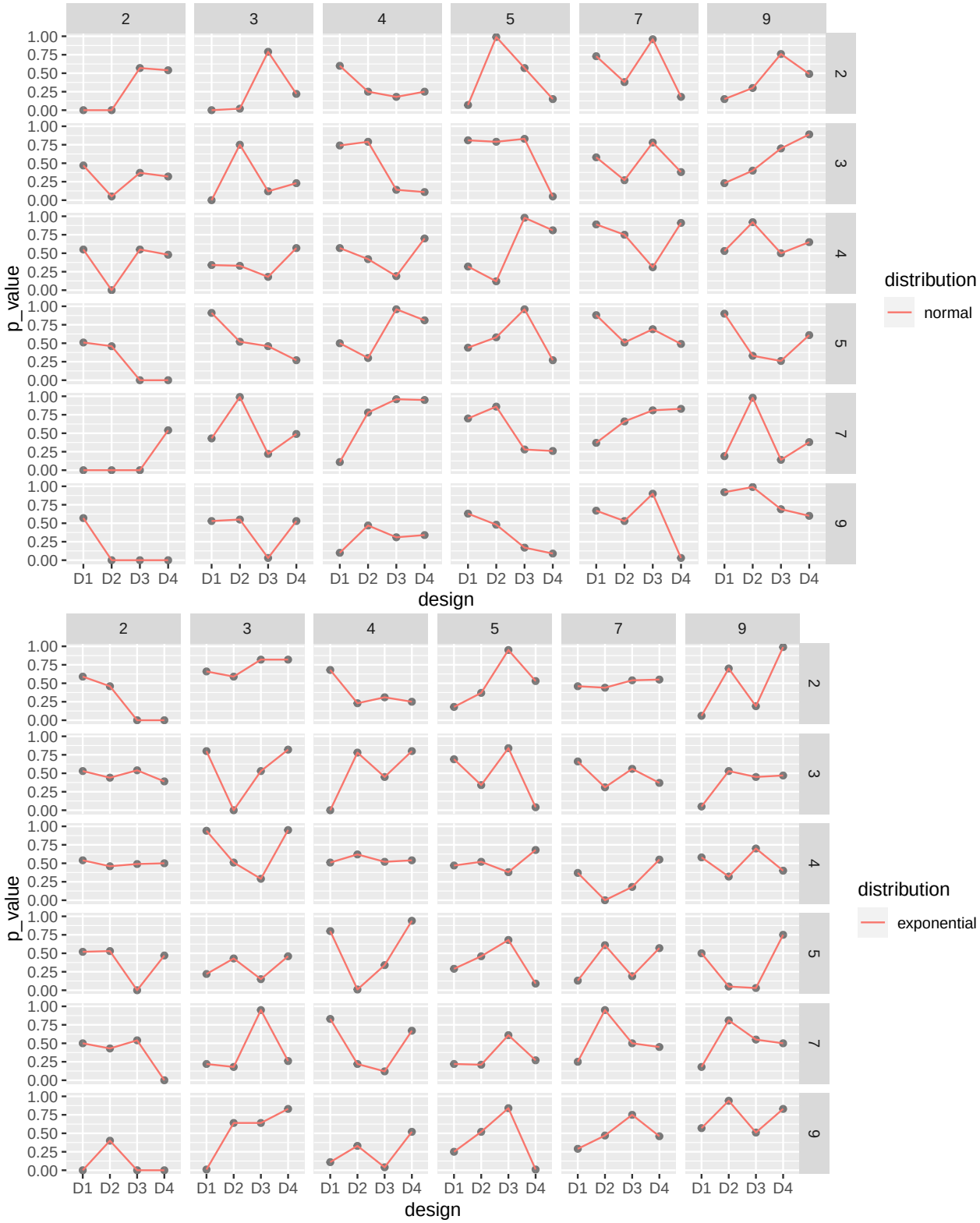
1. Given the data; $\{v_t : t = 0, 1, 2, \ldots, T-1\}$, the MMPD is computed and is represented by $MMPD_{obs}$.
2. From the original sequence a random permutation is obtained: $\{v_t^* : t = 0, 1, 2, \ldots, T-1\}$.
3. MMPD is computed for all random permutation of the data and is represented by $MMPD_{sample}$.
4. Steps (2) and (3) are repeated a large number of times M (e.g. 1000).
5. For each permutation, one $MMPD_{sample}$ value is obtained.
6. 95^{th} percentile of this $MMPD_{sample}$ distribution is computed and stored in $MMPD_{threshold}$.
7. If $MMPD_{obs} > MMPD_{threshold}$, harmony pairs are accepted. Only one threshold for all harmony pairs.

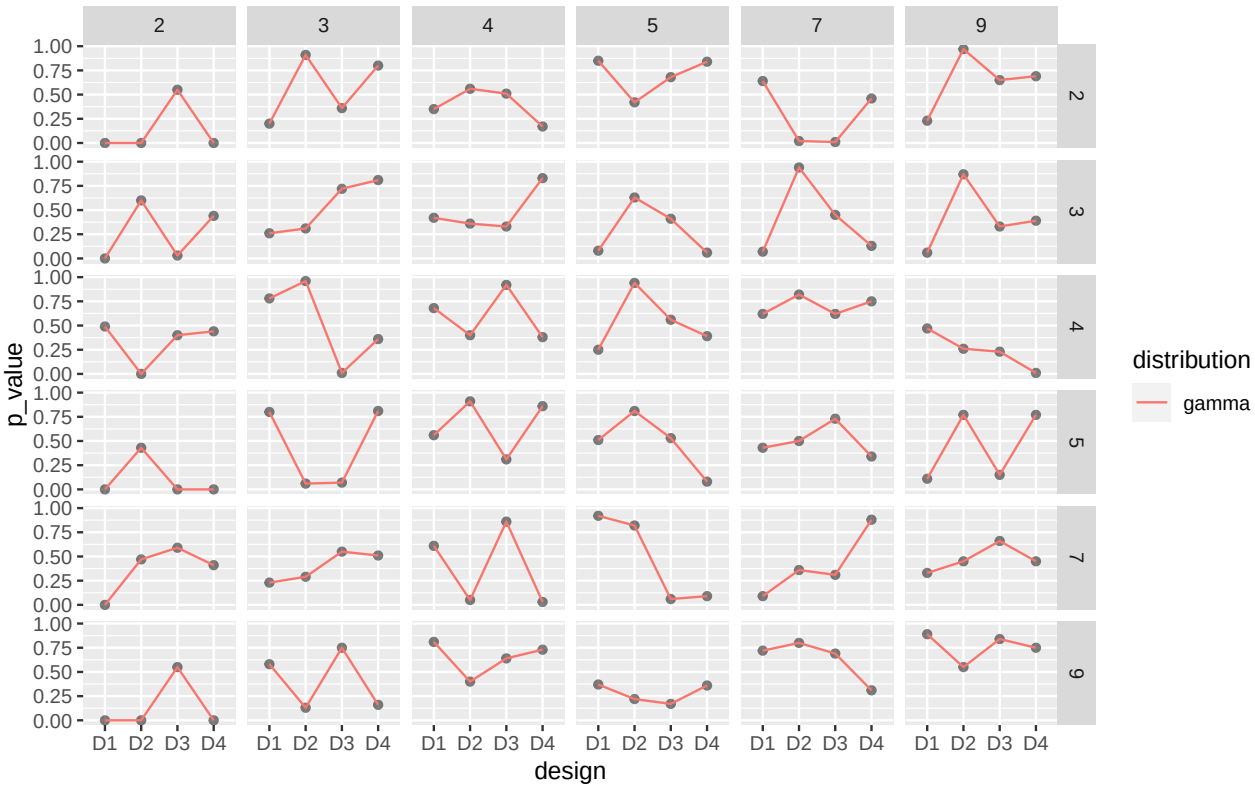
Pros: Considering thresholds global for all harmony pairs would imply less computation time.

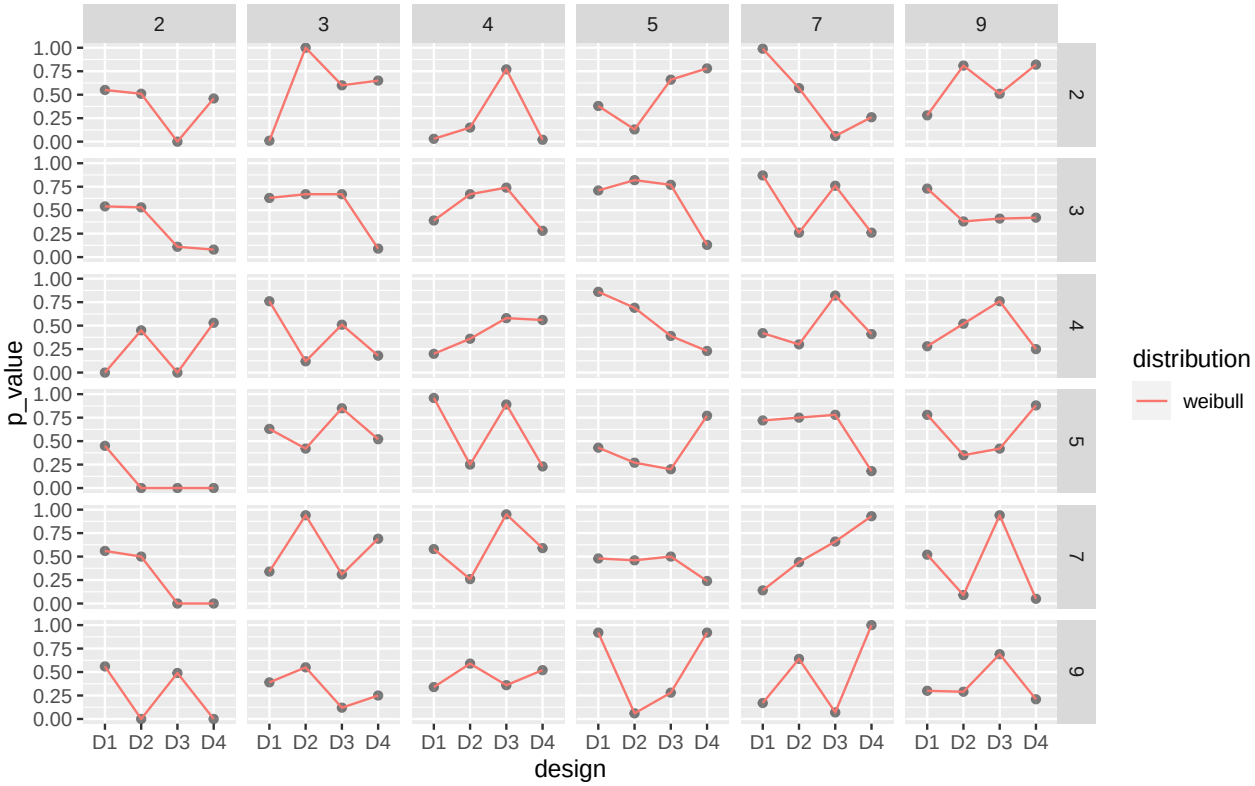
Cons: Only one threshold for all harmony pairs means we are assuming distribution of all harmonies pairs are similar, which might not be the case.But nevertheless, it is a good benchmark.

5.2 Size, power and confidence interval

5.2.0.1 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
p_value
1.00
0.75
0.50
0.25
0.00
D1
D2
D3
D4
design
distribution
gamma


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

6 Applications

6.1 Smart meter data of Australia

Smart meters provide large quantities of measurements on energy usage for households across Australia. One of the customer trials (Department of the Environment and Energy 2018) conducted as part of the Smart Grid Smart City project in Newcastle, New South Wales and some parts of Sydney provides customer wise data on energy consumption for every half hour from February 2012 to March 2014. The idea here is to show how to visualize the distribution of the energy consumption across different cyclic granularities in a systematic way to identify different behavioral patterns.

6.1.1 Cyclic granularities search and computation:

The tsibble object `smart_meter10` from R package `gravitas` (Gupta et al. 2019) consisting of `reading_datetime`, `customer_id` and `general_supply_kwh` denoting the index, key and measured variable of the tsibble is used to facilitate the systematic exploration. While trying to explore the energy behavior of these customers systematically across cyclic time granularities, the first thing to consider is which cyclic time granularities we can look at exhaustively. Let us consider conventional time deconstructions for a Gregorian calendar (second, minute, half-hour, hour, day, week, month, year). Since the interval of this tsibble is 30 minutes, the temporal granularities may range from half-hour to year. Considering 6 linear granularities half-hour, hour, day, week, month and year in the hierarchy table, $N_C = (6 * 5/2) = 15$. If N_C seem too large, the smallest and largest linear granularities could be considered to be removed from the hierarchy table. We remove half-year and year to have $N_C = (4 * 3/2) = 6$ and 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. Now that we have a list of cyclic granularities to look at, we should be able to compute the multiple-order-up granularities using Section **??**.

6.1.2 Screening and visualizing harmonies

From the search list, $N_C = 7$ cyclic granularities are chosen for which we would like to derive insights of energy behavior. Recalling the data structure $<C_i, C_j,$ `general_supply_kwh`$>$ for exploration $\forall i, j \in \{1, 2, \ldots, 7\}$, each of these 7 cyclic granularities can either be mapped to x-axis or to facet. Choosing 2 of the possible 7 granularities, which is equivalent to having ${}^7P_2 = 42$ candidates for visualization. Fortunately, harmonies can be identified among those 42 possibilities to narrow the search. **??** shows 16 harmony pairs after removing clashes and any cyclic granularities with levels more than 31, as effective exploration becomes difficult with many levels (Section **??**). The MMPD is also shown along with indicator (*) only when variation of measured variable across the harmony pair significant. Starting from 42 possible pairs of cyclic granularities to visualize, we are finally left with only 6, which is a very sizable number of displays for exploration.

Few harmony pairs are displayed in Figure 13 to illustrate the significance of MMPD, threshold and the impact of different distribution plots and reverse mapping. For each of Figure 13 (b) and (c), C_i is the circular granularity day-type (weekday/weekend) and C_j is hour of the day. The geometry used for displaying the distribution is chosen as area-quantiles and violins in Figure 13 (b and c respectively). Figure 13 (a) displays reverse mapping of C_i and C_j with C_i denoting hour of the day and C_j denoting day-type with distribution geometrically displayed as boxplots.

In Figure 13 (b), the black line is the median, whereas the purple band covers 25th to 75th percentile, the orange band covers 10th to 90th percentile and the green band covers 1st to 99th percentile. The first facet represents the weekday behavior while the second one displays the weekend behavior and energy consumption across each hours of the day is shown inside each facet. The energy consumption is extremely (positive- or right-) skewed with the 1st, 10th and 25th percentile lying relatively close whereas 75th, 90th and 99th lying further away from each other. This is common across both weekdays and weekends. For the first few

hours on weekdays, median energy consumption starts and continues to be higher for longer as compared to weekends.

Consider looking at violin plots instead of quantile plots to look at the same data in Figure 13(c). There is additional information that we can derive looking at the distribution. There is bimodality in the early hours of the day, implying both low and high energy consumption is probable in the early hours of the day both for weekdays and weekends. If we visualize the same data with reverse mapping of the cyclic granularities, then the natural tendency would be to compare weekend and weekday behavior within each hour and not across hours. For example in Figure 13(a), it can be seen that median energy consumption for the early morning hours is extremely high for weekdays compared to weekends. Also, outliers are more prominent in the latter part of the day. All of these indicate that looking at different distribution geometry or changing the mapping might shed lights on different aspect of the energy behavior for the same sample population.

If the data for all keys are visualized together, it might lead to Simpson's paradox, which occurs when one observation shows a particular behavior, but this behavior paradoxically becomes obscured by aggregation. For example in a particular neighborhood one household may have the least daily power consumption for a full week, yet still not be the household with the minimum weekly power consumption. This is an intuitive possibility, because heterogeneous `customer_id`'s with very different occupation or demographics will tend to have very different energy behavior and combining them together will somehow weaken any typical or extreme behavior. A strategy for analyzing multiple keys together could be to first group them basis time series or demographic features and then look at their energy behavior. This is beyond the scope of the current work.

This case study shows systematic exploration of energy behavior for a household to gain exhaustive insights on periodic behavior of the households.

6.2 T20 cricket data of Indian Premiere League

The method is not only restricted to temporal data, and can be generalized to many hierarchical granularities (with continuous and uni-directional nature). We illustrate this with an application to the sport cricket. Although there is no conventional time component in cricket, each ball can be thought to represent an ordering from past to future with the game progressing forward with each ball. In the Twenty20 format, an over will consist of 6 balls (with some exceptions), an inning is restricted to a maximum of 20 overs, a match will consist of 2 innings and a season consists of several matches. Thus, similar to time, there is a hierarchy where ball is nested within overs, overs nested within innings and innings within matches. The idea of cyclic granularities can be likewise mapped to this hierarchy. Example granularites then include ball of the over, over of the inning and ball of the inning. Although most of these cyclic granularities are circular in design of the hierarchy, in application of the rules some granularities are aperiodic. For example, in most cases an over will consist of 6 balls with some exceptions like wide balls or when an inning finishes before the over finishes. Thus, the cyclic granularity ball-of-over will be circular in most cases and aperiodic in others.

The Indian Premier League (IPL) is a professional Twenty20 cricket league in India contested by eight teams representing eight different cities in India. The ball by ball data for IPL season 2008 to 2016 is fetched from Kaggle. The `cricket` data set in the `gravitas` package summarizes the ball-by-ball data across overs and contains information for a sample of 214 matches spanning 9 seasons (2008 to 2016) such that each over has 6 balls, each inning has 20 overs and each match has 2 innings. This could be useful in a periodic world when we wish to compute any circular/quasi-circular granularity based on a hierarchy table which look like Table 2.

However, even if the situation is not periodic and a similar hierarchy can not be formed, it can be interesting to visualize the distribution of a measured variable across relevant cyclic granularities to shed light on the aperiodic behavior of a non-temporal data set similar to aperiodic events like formal meetings, workshops, conferences, school semesters in a temporal set up. There are many interesting questions that could possibly be answered with such a data set irrespective of the type of cyclic granularities.

First, it would be interesting to see if the distribution of total runs vary depending on if a team bats in the first or second innings. The Mumbai Indians (MI) and Chennai Super kings (CSK) appeared in final

Table 2: Hierarchy table for cricket where overs are nested within an inning, innings nested within a match and matches within a season.

linear (G)	single-order-up cyclic (C)	period length/conversion operator (K)
over	over-of-inning	20
inning	inning-of-match	2
match	match-of-season	k(match, season)
season	1	1

playoffs from 2010 to 2015. We take their example in order to dive deeper into this question. From Figure 14(a), it can be observed that for the team batting in the first inning there is an upward trend of runs per over, while there is no clear upward trend in median and quartile deviation of runs for the teams batting in the second inning. This seem to indicate that players feel mounting pressure to score more runs as they approach towards the end of the first inning. Whereas teams batting in the second inning have a set target in mind and are not subjected to such mounting pressure and may adopt a more conservative strategy, to score runs. Thus winning teams like CSK and MI seem to employ different inning strategies when it comes to their batting order.

Another interesting question could be: do runs per over decrease in the subsequent over if fielding (defending) was good in the previous over? For establishing the fielding quality, we apply an indicator function on dismissals (1 if there was at least one wicket in the previous over due to run out or catch, 0 otherwise). Runs in the current over is then the observation variable. Dismissals in the previous over can lead to a batsman adopting a more defensive play style. Figure 14(b) shows that no dismissals in the previous over leads to a higher median and quartile spread of runs per over as compared to the case when there has been at least one dismissal in the previous over.

Wickets per over are considered as an aperiodic cyclic granularity with wickets as an aperiodic linear granularity. These granularities do not appear in the hierarchy table since it is difficult to position them in a hierarchy. These are similar to holidays or special events in temporal data.

7 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.

8 Appendix

8.1 Null distribution

8.1.1 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.

8.1.2 Normalised maximum distances follow standard Gumbel distribution

8.1.3 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 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).

8.2 Power

8.3 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.

8.3.1 Varying distribution across facet

8.3.2 Varying distribution across x-axis

8.3.3 Varying distribution across both facets and x-axis

8.3.4 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.

Department of the Environment and Energy. 2018. *Smart-Grid Smart-City Customer Trial Data*. Australian Government, Department of the Environment; Energy: Department of the Environment; Energy, Australia. https://data.gov.au/dataset/4e21dea3-9b87-4610-94c7-15a8a77907ef.

Gupta, Sayani, Rob Hyndman, Di Cook, and Antony Unwin. 2019. *gravitas: Explore Probability Distributions for Bivariate Temporal Granularities*. https://CRAN.R-project.org/package=gravitas.

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.

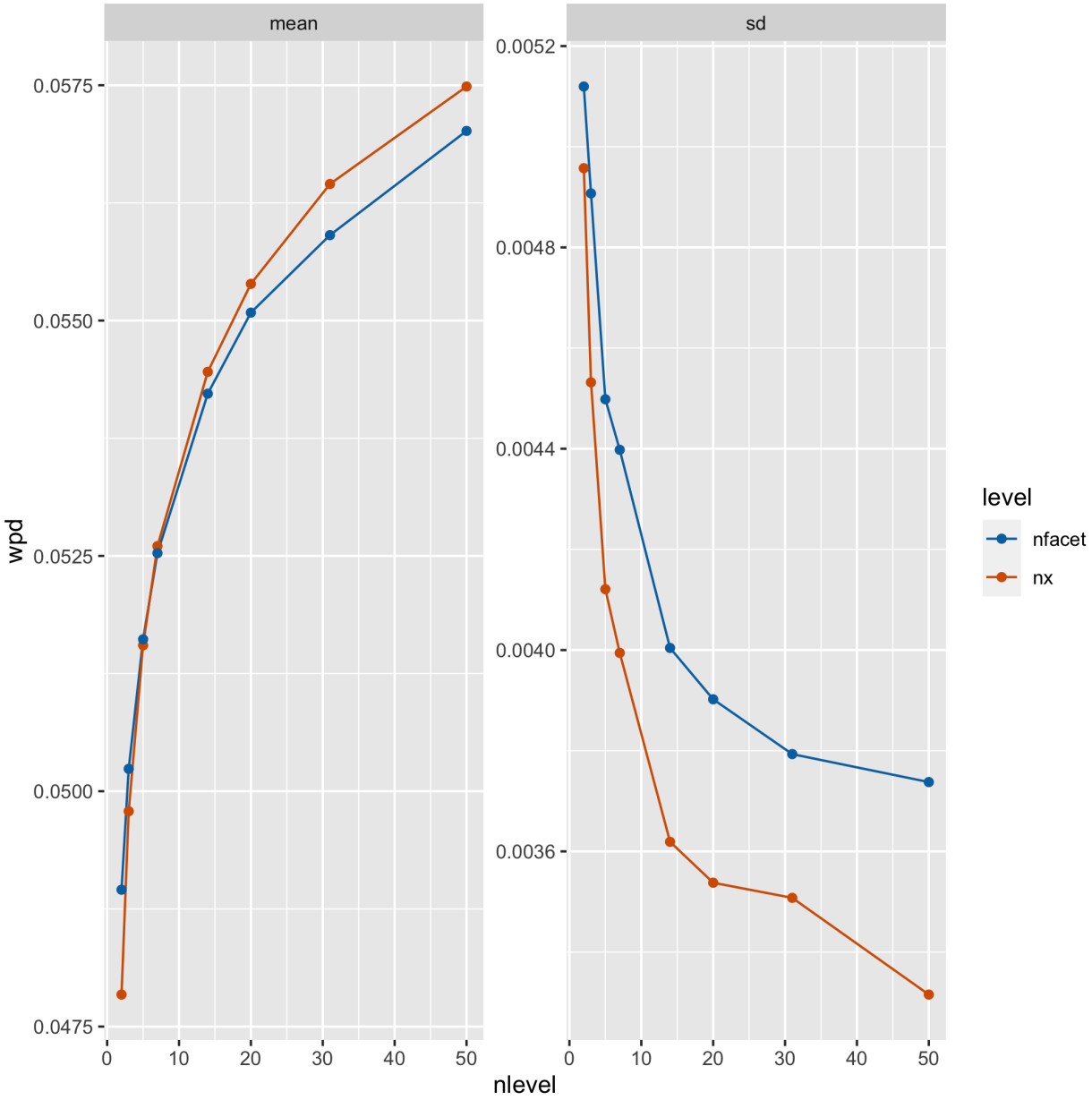


Figure 8: Movement of mean and sd for raw wpd is shown for different number of levels (nlevel) of x-axis and facets through line plots. Mean increases and sd decreases more sharply for increasing x-axis levels as compared to facet levels. It seems like both mean and standard deviation are affected more by change in the x-axis levels than the facet levels.

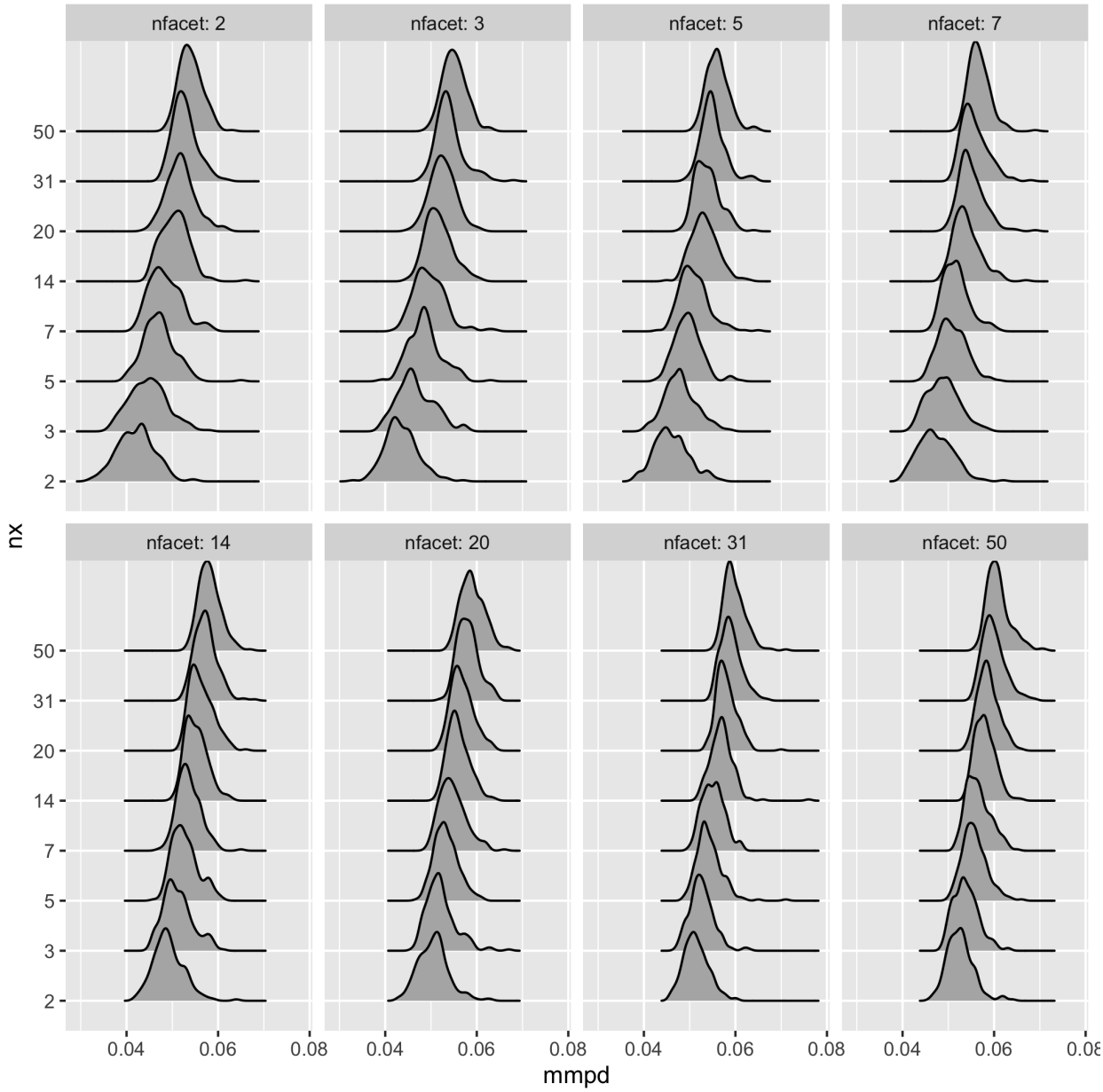


Figure 9: Ridge plots of raw wpd is shown for N(0,5) distribution. For each panel, it could be seen that the location shifts to the right for increasing x levels. Across each panel, the scale of the distribution seems to change for low/moderately lower values and higher values of nfacets and left tails are longer for lower facet levels.

Figure 10: Ridge plots of raw wpd is shown for G(0.5,1) and G(2,1) distribution after quantile transformation looks similar and hence is unaffected by change in location.

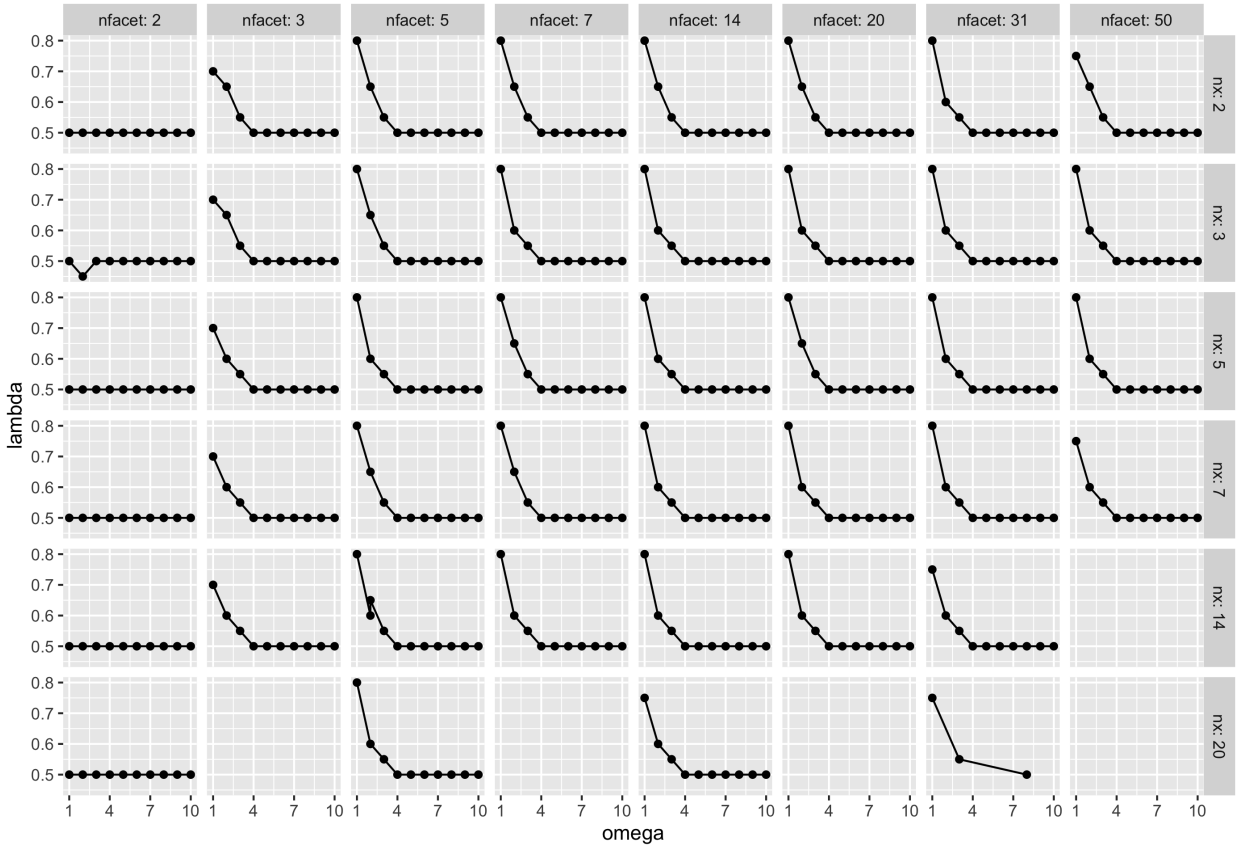


Figure 11: For most panels it is observed that the most common value of the tuning paramter for which the designs interact is 0.5, which implies any value greater than 0.5 could be chosen to up-weigh the within-facet distances and down-weigh the between-facet distances for most sitatuions.

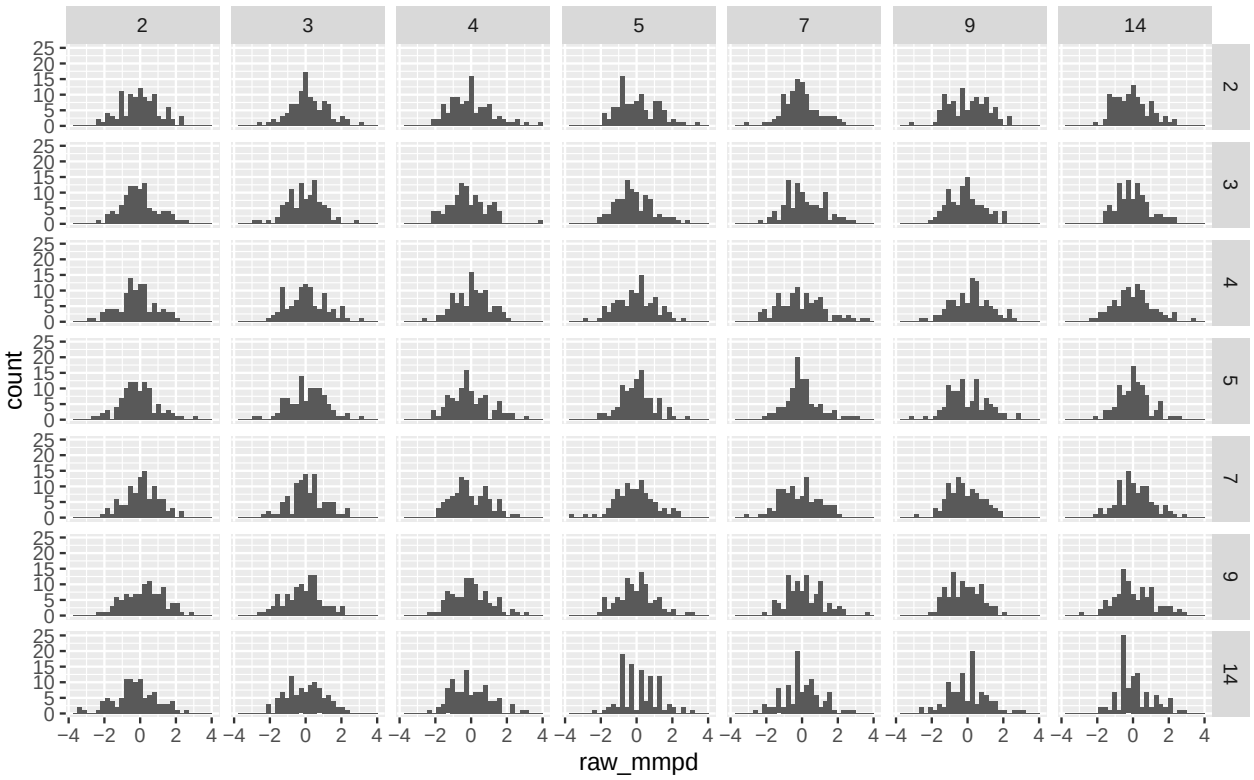


Figure 12: Distribution of $MMPD_{norm}$ across different levels of facet and x-axis. These distributions should look the same if normalisation worked.

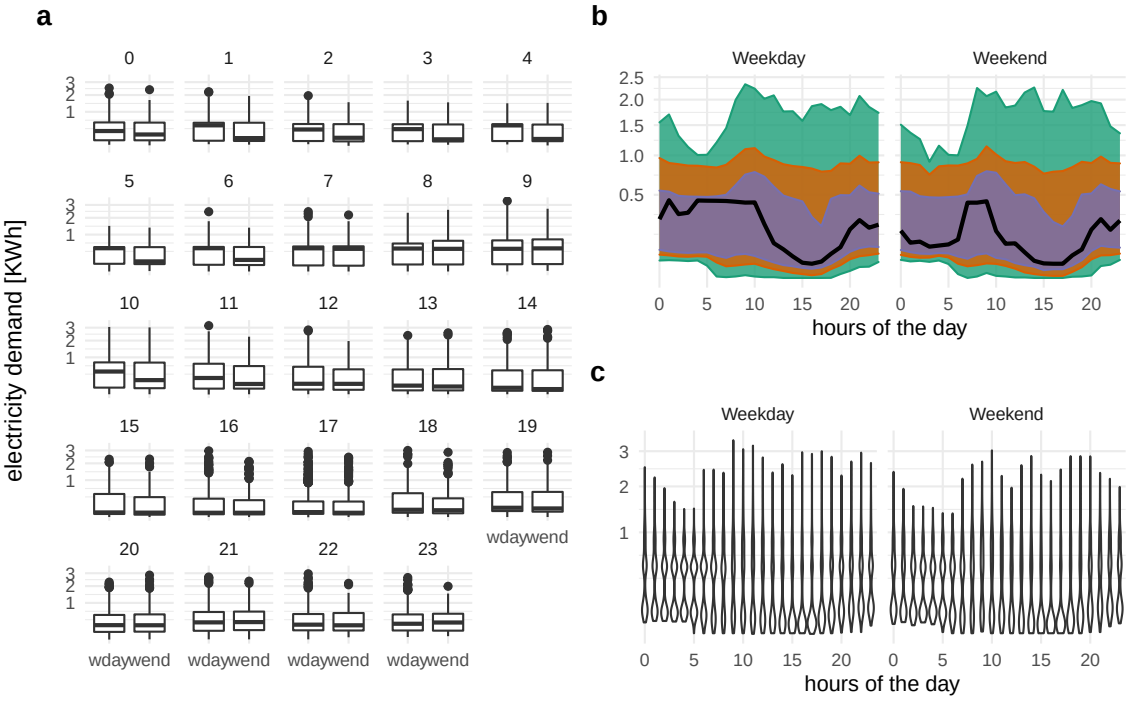


Figure 13: Energy consumption of a single customer shown with different distribution displays, and granularity arrangements. Two granularities are used: hour of the day (I) and weekday/weekend (II). Plot (a) shows granularity I facetted by granularity II, and plots (b), (c) shows the converse mapping. Plot (a) makes a comparison of usage by workday within each hour of the day using side-by-side boxplots. Generally, on a work day there is more consumption early in the day. Plots (b) and (c) examine the temporal trend of consumption over the course of a day, separately for the type of day. Plot (b) uses an area quantile to put the emphasis on the time series, for example, the median consumption over time shows prolonged usage in the morning on weekdays. Plot (c) uses a violin plot to place emphasis on distributional differences across hours. It can be seen that the morning use on weekdays is bimodal, some work days there is low usage, which might indicate the person is working from home and also having a late start.

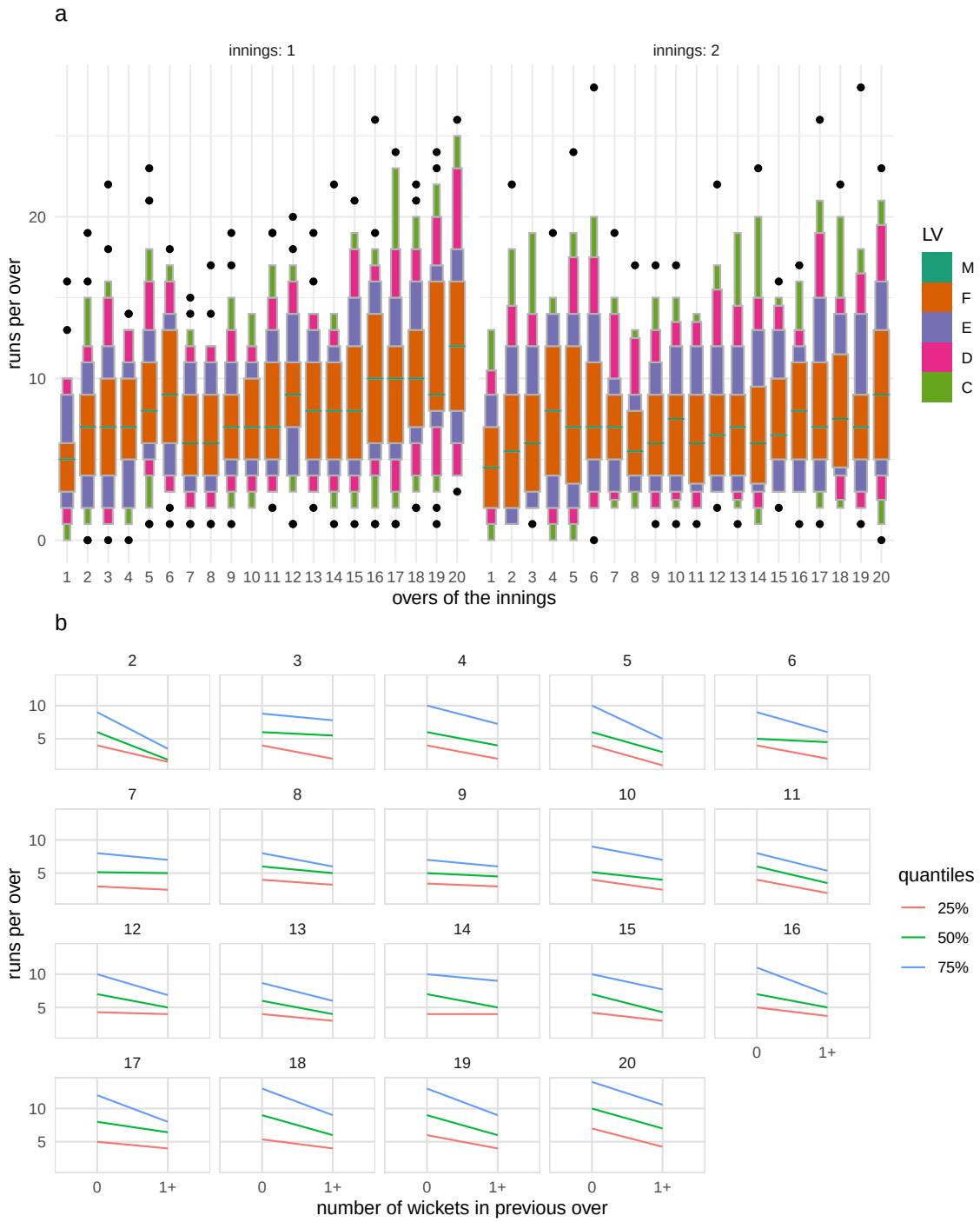


Figure 14: Runs per over shown with different distribution displays, and granularities. Plot (a) shows letter value plot across overs faceted by innings. For the team batting in the first innings there is an upward trend of runs per over, while there is no such pattern of runs for the teams batting in the second innings. Plot (b) shows quantile plot of runs per over across an indicator of wickets in previous over faceted by current over. This indicates that at least one wicket in the previous over leads to lower median run rate and quartile spread in the subsequent over.