What can be proven by functional test statistics: The R package GET

based on joint work with Mari Myllymäki and many others

Tomáš Mrkvička

Faculty of Economics
University of South Bohemia

Examples and basics

Examples of functional test statistic

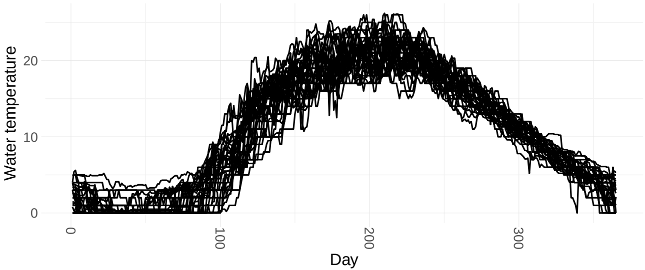


FIGURE – Annual water temperature curves sampled at the water level of Rimov reservoir in Czech Republic every day from 1979 to 2014.

The functional test statistic is the regression coefficient β computed at every day of the year. $Temperature(Day, Year) \sim \beta_0(Day) + \beta_1(Day) * Year$
This is function-on-scalar regression, but we put no assumption on the model, regarding distribution and homogeneity of the distribution. Here winter days are distributed differently than summer days.
Question : In which days of a year do we observe warming ?

Examples of functional test statistic

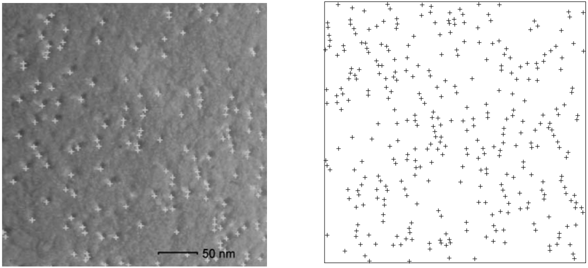


FIGURE – Electron micrographs of intramembraneous particles.

The functional test statistic is the transformation of Ripley's K function - The expected number of points up to a distance r divided by intensity of points.
Question : Which particle distances deviates from complete spatial randomness ?

Examples of functional test statistic

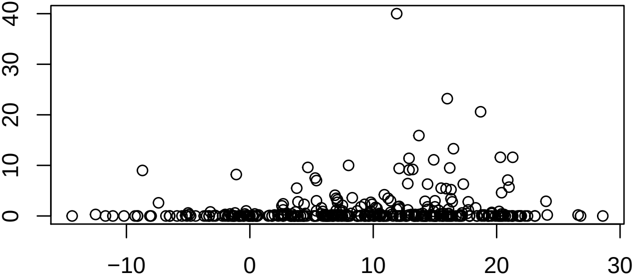


FIGURE – The temperature and precipitation of the first day in every month during 1989 - 2018, measured in Prague.

The functional test statistic is a 2D test statistic informing about correlation between two variables in the form of 2D quantile-quantile plot.
Question : Which combination of quantiles deviates from independence ?

Monte Carlo test / Permutation test

• We observe test statistic $T_0(r)$.
• Assume we can generate replicates $T_1(r), \ldots, T_s(r)$ of the test statistic under the null hypothesis. Either by reproducing data (Monte Carlo) or permuting data.
• Sort the function from most extreme to least extreme using a measure M.
$T_i(r) \prec T_j(r)$, i.e. T_i is more extreme than T_j, iff $M_i < M_j$.
• Global p-value is $p = \sum_{i=0}^{s} \mathbf{1}(M_i \leq M_0) \big/ (s+1)$.

Global envelope test / central region

• If we are able to order the functions $T_0(r), \ldots, T_s(r)$, we are able to produce the central region.

1. Denote $m_{(\alpha)} \in \mathbb{R}$ the largest of the M_i such that the number of $i : M_i < m_{(\alpha)}$ is less or equal αs ;
2. Denote $[\mathbf{T}_{\mathrm{low}}^{(\alpha)}(r) = \min\{T_i(r) : M_i \geq m_{(\alpha)}\}, \mathbf{T}_{\mathrm{upp}}^{(\alpha)}(r) = \max\{T_i(r) : M_i \geq m_{(\alpha)}\}]$ the $100(1-\alpha)\%$ central region.

• We are interested ONLY in ordering satisfying the intrinsic graphical interpretation (IGI).

• The 95% central region can be used to produce the global envelope test constructed from $T_0(r)$ test statistic of the data and $T_1(r), \ldots, T_s(r)$ resamples under null hypothesis.

Intrinsic graphical interpretation (IGI)

Intrinsic graphical interpretation with respect to the ordering $\prec$ induced by measure M satisfies :

1. [$T_i(r) < \mathbf{T}^{(\alpha)}_{\mathrm{low}}(r)$ or $T_i(r) > \mathbf{T}^{(\alpha)}_{\mathrm{upp}}(r)$ for some r iff $M_i < m_{(\alpha)}$] for every $i = 1, \ldots, s$;
2. [$\mathbf{T}^{(\alpha)}_{\mathrm{low}}(r) \leq T_i \leq \mathbf{T}^{(\alpha)}_{\mathrm{upp}}(r)$ for all r iff $M_i \geq m_{(\alpha)}$] for every $i = 1, \ldots, s$.

Global envelope test for temperature data

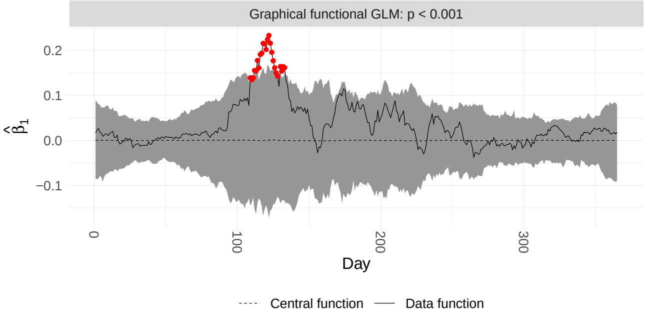


FIGURE – The output of the global ERL envelope test ($p < 0.001$) for testing the effect of the year on the temperatures. The grey area represents the 95% global ERL envelope and the red dots show the days where the data function exceeds the envelope.

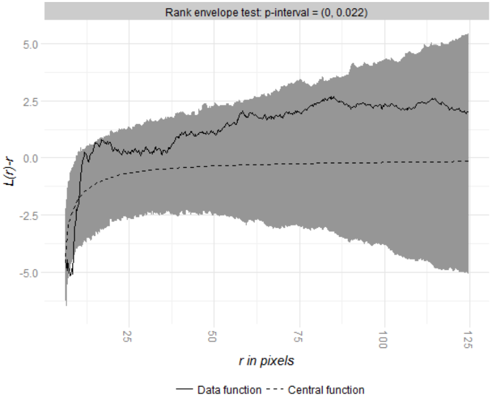
Global envelope test

- Implementation in R package GET. (Myllymäki M., Mrkvička T. : GET : Global envelopes in R. arXiv:1911.06583 [stat.ME])
- Provides p-value.
- Provides envelopes which specifies r distances leading to the rejection.
- Helps to detect reasons why the data contradict the null hypothesis.
- Does not suffer from the uneven variance and asymmetry of $T(r)$.
- It is a completely nonparametric procedure - It gives exactly same weight to every r.
- It does not rely on assumptions about distribution of the test statistic, homogeneity of the test statistic across the image, so it can be used for any test statistic.

Another perspective

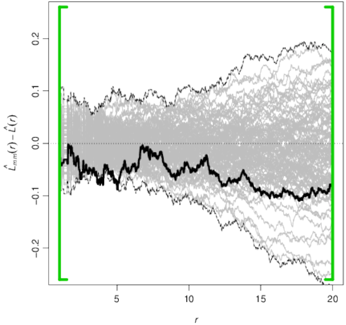
- The functions $T_i(r)$ have to be discretized, then $T_i(r)$ is a multivariate vector of dimension d.
- Then we have d point-wise tests and
- Global envelope test is multiple test comparison procedure with family wise error rate (FWER).
- The global envelope test uses the whole correlation structure between the test, since whole functions are resampled. Thus it is very powerful multiple testing procedure.
- It can be treated as the p-min multiple testing procedure, with resolved ties problem of p-mins. The solution is based on Extreme rank length measure (ERL) or Continuous rank measure (Cont) or Area rank measure (Area).
- In FDA the F-max test is also often used $F^{\max} = \max_{r \in I} T(r)$.
- The statistic $T(r)$ should be pivotal statistic, i.e. the distribution of the statistic must not change for different r values, in order to have same contribution of every r in F-max.
- But even basic statistics like t or F are not pivotal, but only second order pivotal statistic.

Global envelope test for goodness-of-fit of intramembraneous particles



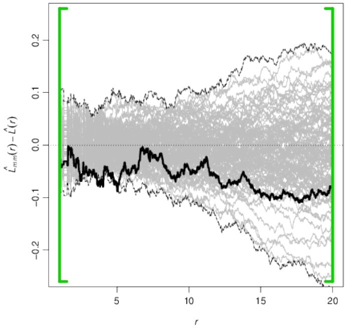
- The null model is hard-core point pattern model
- It indicates the deviation for short distances and suggests that a soft core model should be used.

Multiple testing problem of the conventional envelope test



- Ripley (1977)
  - introduced the envelope test
  - mentioned that the frequency of committing the type I error in the envelope test may be higher than for a single distance test
- Loosmore & Ford (2006)
  - demonstrated by an example that the type I error probability may be unreasonably high ≈ 0.74 !
- L & F (2006) recommended not to use the envelope test.

Multiple testing problem of the conventional envelope test



$s = 99$, type I error appr. ≈ 0.40

- Ripley (1977)
  - introduced the envelope test
  - mentioned that the frequency of committing the type I error in the envelope test may be higher than for a single distance test
- Loosmore & Ford (2006)
  - demonstrated by an example that the type I error probability may be unreasonably high ≈ 0.74 !
- L & F (2006) recommended not to use the envelope test.

Global envelope test - Testing independence

The global envelope test shows that days with high precipitation and small temperature appear less often than under independence (Blue area) and that days with high precipitation and average temperature appear more often than under independence (Red area).

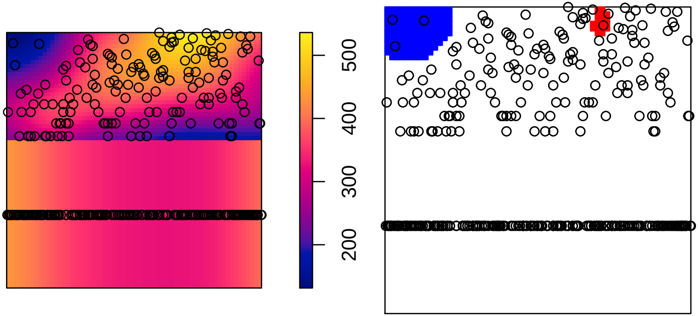


FIGURE – 2D global envelope test for the independence of temperature and precipitation. The test statistic is intensity of 2D qq-plot.

GET :package

GET : Global envelopes in R. : Myllymäki M., Mrkvička T. can be downloaded from CRAN (https://cran.r-project.org/package=GET).
Content :

1. General GET - graphical multiple testing with control of FWER
2. Functional GLM / Functional ANOVA (which groups differ)
3. Goodness-of-fit testing in spatial statistics
4. Functional box plot and central region
5. Functional clustering
6. Local independence test
7. Testing difference between 2 or more CDF
8. Local spatial correlation test
9. Still to be developed :)

Combining

Combining of different sources

Various perspectives can be addressed simultaneously,

1. different functional test statistic can be combined
2. different magnitudes and different shapes of the functions
   - raw functions - magnitude
   - shifting each curve to have a zero mean and normalizing the centered curves by their L_2 norms - shape
3. different components of multivariate functional data together with their covariance

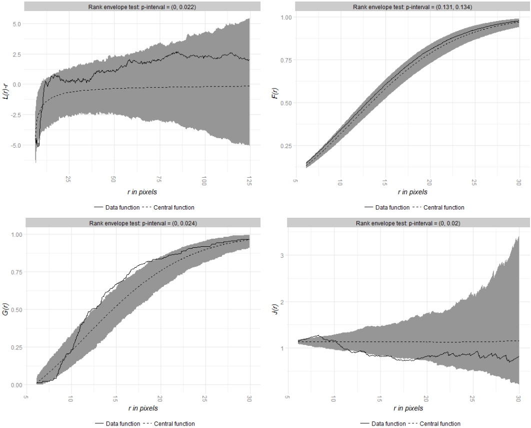
Thus assume we have k functional test statistic $T^1, \ldots, T^k$.
The central region is then computed for long vector

$$\mathbf{T}_i = (T_{i1}^1, \ldots, T_{im}^1, \ldots, T_{i1}^k, \ldots, T_{im}^k), \quad i = 0, \ldots, s.$$

The desired perspectives of ordering are considered and treated equally in such a combined ordering due to the ranking nature of proposed ordering.

Testing with several summary functions

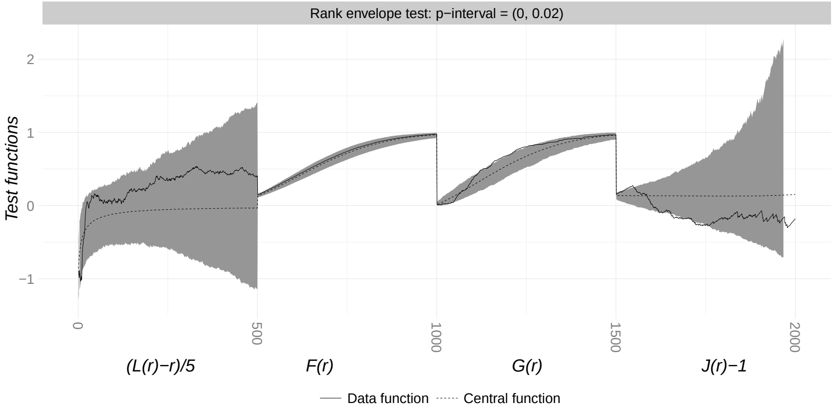
Schladitz et al. 2003 and many others recommend to explore also other summary functions to reveal also other features of the point pattern



Rank envelope tests of the intramembranous particles for testing a hard-core null model with 2499 simulations

Again the problem of multiple testing

Solution is combined rank envelope test with several test functions.



Combined rank envelope tests of the intramembranous particles for testing a hard-core null model with 9999 simulations.

Joint functional clustering for population growth curves

• The size of population growth and shape of the population growth is of interest when clustering the countries.
• To perform clustering, one needs to define the dissimilarity matrix.

• Assume that we observe functions $T_i(r), i = 1, \dots, s$ in fixed set of points $r_1, \dots, r_d$.
• Construct the new set of functional differences $D_f = \{T_i(r) - T_i'(r), i, i' = 1, \dots, s\}$.
• Apply a chosen IGI ordering to D_f.
• Finally, the elements of the dissimilarity matrix $d_{ii'} = 1 - M_{ii'}$, where $M_{ii'}$ is measure, which induce the chosen ordering, of $T_i(r) - T_i'(r)$.

Joint functional clustering for population growth curves

Our proposed procedure then consists from the following steps :

1. Choose the appropriate data sources (e.g. raw data, derivative and second derivative).
2. Choose the functional ordering which allows for intrinsic graphical interpretation and which gives the same weight to every chosen source (e.g. the studentised maximum ordering, the area rank ordering).
3. Compute the dissimilarity matrix from the set of differences.
4. Apply partitioning around medoids.
5. Plot the resulted clusters together with their central region with intrinsic graphical interpretation.

Joint functional clustering for population growth curves

- Raw curves of population growth

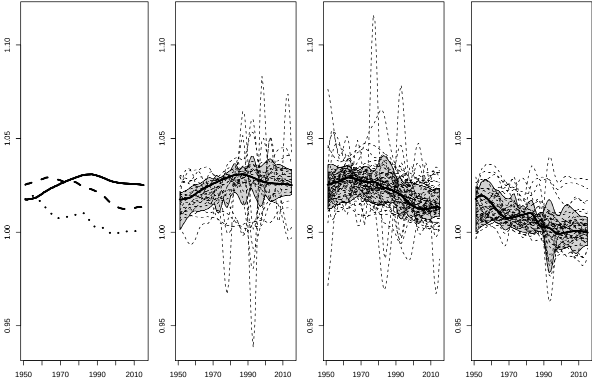


FIGURE – The functional clustering with the 'Area' measure. Light grey area shows 50% central region of each cluster. The left panel and solid curve in 3 right panels shows cluster medoids.

Joint functional clustering for population growth curves

- Sequentially transformed population growth (centered and scaled functions which reveal the shape)

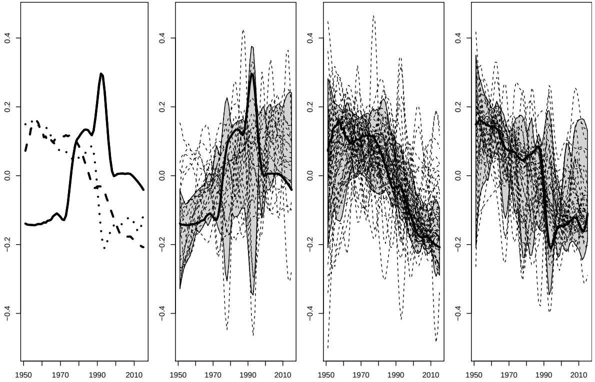


FIGURE – The functional clustering with the 'Area' measure. Light grey area shows 50% central region of each cluster. The left panel and solid curve in 3 right panels shows cluster medoids.

Composite case

Composite hypothesis !

• Often we need to estimate some parameters of the null model, then the null hypothesis is composite and all the tests become usually conservative.
• Solution is to use Two-stage Monte Carlo test (Baddeley et al. 2017)
• We have the data X_1 and estimate of parameters θ_1.
• We simulate $X_2, \dots, X_{s+1}$ and estimate parameters $\theta_2, \dots, \theta_{s+1}$.
• Perform Monte Carlo test for X_1 performing new simulations from θ_1 and obtain p_1.
• Perform s Monte Carlo tests, where X_i is compared with new set of simulations from θ_i and obtain p_i.
• Compare p_1 with $p_2, \dots, p_{s+1}$ and obtain adjusted p.

Adjusted global envelope

How to obtain an adjusted global envelope ?

- By taking 5% quantile from raw $p_0, \ldots, p_{s+1}$ we obtain an adjusted α.
- Then by drawing envelope with adjusted α gives adjusted global envelope.
- The adjustment procedure is rather time consuming. s^2 simulations are required.

Combining several adjusted directional quantile envelope tests

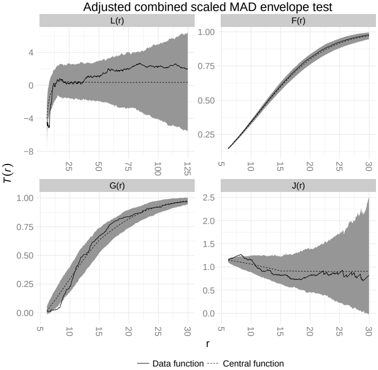


FIGURE – The adjusted combined directional quantile MAD envelope test for testing the Strauss hard core process using $L(r) - r$, $F(r)$, $G(r)$ and $J(r)$ functions. The grey area shows the adjusted combined 95% directional quantile MAD envelope based on $s = 499$ simulations.

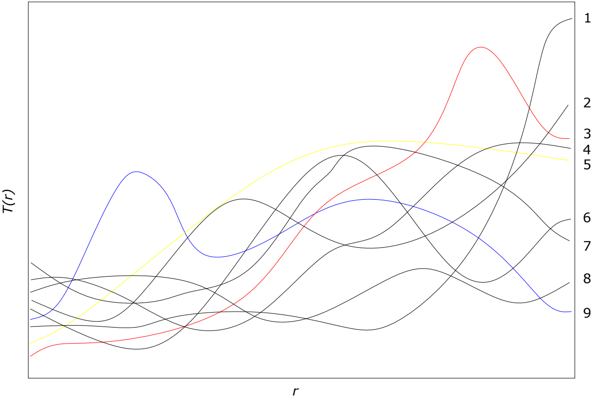
IGI ordering

GET :package - IGI ordering

1. Extreme rank ordering
2. Extreme rank length ordering
3. Continuous rank ordering
4. Area rank ordering
5. Studentized ordering
6. Directional quantile ordering

Choice of the type of global envelopes

- Large number of functions/simulations, $s \to$ rank, erl, cont, area lead to the same output, any choice OK
- with a low number of functions, the measure playes a role



Most extreme one : Rank : 1, 3, 4, 5, 7 and 9 ; ERL 5 ; Cont 9 ; Area 9 ; Unscaled 1

Extreme rank ordering - IGI

Let $r_{1j}, r_{2j}, \dots, r_{sj}$ be the raw ranks of $T_{0j}, T_{2j}, \dots, T_{sj}$, such that the smallest T_{ij} has rank 1. In the case of ties, the raw ranks are averaged. The two-sided pointwise ranks are then calculated as $R_{ij} = \min(r_{ij}, s + 1 - r_{ij})$.
The extreme rank R_i of the vector $\mathbf{T}_i$ is defined as the minimum of its pointwise ranks, namely

$$R_i = \min_{k=1,\dots,m} R_{ik}, \tag{1}$$

- Problem - ties
- It is equivalent to the p-min approach in multiple testing literature.

Extreme rank length ordering - IGI - Solving ties

Consider now the vectors of pointwise ordered ranks $\mathbf{R}_i = (R_{i[1]}, R_{i[2]}, \dots, R_{i[m]})$, where $\{R_{i[1]}, \dots, R_{i[m]}\} = \{R_{i1}, \dots, R_{im}\}$ and $R_{i[k]} \leq R_{i[k']}$ whenever $k \leq k'$. The extreme rank length measure of the vectors $\mathbf{R}_i$ is equal to

$$E_i = \frac{1}{s+1} \sum_{i'=0}^{s} \mathbf{1}(\mathbf{R}_{i'} \prec \mathbf{R}_i) \tag{2}$$

where

$$\mathbf{R}_i' \prec \mathbf{R}_i \quad \iff \quad \exists\, n \leq m : R_{i'[k]} = R_{i[k]} \ \forall\ k < n,\ R_{i'[n]} < R_{i[n]}.$$

Continuous rank ordering - IGI - Solving ties

The continuous rank measure is

$$C_i = \min_{j=1,\dots m} c_{ij} / \lceil s'/2 \rceil,$$

where c_{ij} are the pointwise continuous ranks defined as the continuous position of i-th function between one above and one bellow. If it is the most extreme rank it is defined as exponential distance to the second most extreme rank.

Area rank ordering - IGI - Solving ties

The area rank measure combines both approaches

$$A_i = \frac{1}{\lceil s'/2 \rceil m} \sum_j \min(R_i, c_{ij}),$$

where
$R_i = \min_j \{R_{ij}\}$ and R_{ij} are two-sided pointwise ranks defined above.

Sources of information for Rank, ERL, Continuous rank and Area rank

measure	extreme rank	Integration of extremes	Value of extremes
Rank (p-min)	Yes	No	No
ERL	Yes	Yes	No
Cont rank	Yes	No	Yes
Area rank	Yes	Yes	Yes
F - max	No	No	Yes

TABLE – Sources of information.

Problems with loosing power

	Ties	Inhomogeneity and non-normality of $\epsilon(r)$	Inhomogeneity of correlation structure of $\epsilon(r)$
F-max	No	Yes	No
p-min	Yes	No	No
ERL	No	No	Yes
Cont	No	Yes	No
Area	No	Partial	Partial

TABLE – Problems of different measures.

Studentised ordering - IGI - approximating limiting case of extreme rank ordering

The global studentized envelope approximates the distribution of $T_{\cdot k}, k = 1, \ldots, m$, by the expectation $\overline{T_{\cdot k}}$ and the standard deviation $\mathrm{sd}(T_{\cdot k})$. The studentized measure is

$$S_i = \max_k \left| \frac{T_{ik} - \overline{T_{\cdot k}}}{\mathrm{sd}(T_{\cdot k})} \right|. \tag{3}$$

The $100(1-\alpha)\%$ global studentized envelope induced by S_i is defined by

$$\mathbf{T}^{(l)}_{\mathrm{low}\,k} = \overline{T_{\cdot k}} - S_{(\alpha)}\mathrm{sd}(T_{\cdot k}) \quad \text{and} \quad \mathbf{T}^{(l)}_{\mathrm{upp}\,k} = \overline{T_{\cdot k}} + S_{(\alpha)}\mathrm{sd}(T_{\cdot k}) \quad \text{for } k = 1, \ldots, m, \tag{4}$$

where $S_{(\alpha)}$ is taken according to the point 1. of IGI.

Directional quantile ordering - IGI - approximating limiting case of extreme rank ordering

The global directional quantile envelope uses the expectation $\overline{T_{\cdot k}}$, $\beta\%$ upper $\widetilde{T}_{\cdot k}$ and lower $\underset{\sim}{T}_{\cdot k}$ quantiles to approximate the distributions. The directional quantile measure is

$$D_i = \max_k \left(\mathbf{1}(T_{ik} \geq \overline{T_{\cdot k}}) \frac{T_{ik} - \overline{T_{\cdot k}}}{|\widetilde{T}_{\cdot k} - \overline{T_{\cdot k}}|} + \mathbf{1}(T_{ik} < \overline{T_{\cdot k}}) \frac{\overline{T_{\cdot k}} - T_{ik}}{|\underset{\sim}{T}_{\cdot k} - \overline{T_{\cdot k}}|} \right), \qquad (5)$$

The $100(1-\alpha)\%$ global directional quantile envelope induced by D_i is defined by

$$\mathbf{T}^{(l)}_{\mathrm{low}\, k} = \overline{T_{\cdot k}} - D_{(\alpha)} |\underset{\sim}{T}_{\cdot k} - \overline{T_{\cdot k}}| \quad \text{and} \quad \mathbf{T}^{(l)}_{\mathrm{upp}\, k} = \overline{T_{\cdot k}} + D_{(\alpha)} |\widetilde{T}_{\cdot k} - \overline{T_{\cdot k}}| \quad \text{for } k = 1, \ldots, m, \qquad (6)$$

where $D_{(\alpha)}$ is taken according to the point 1. of IGI.

False discovery rate envelope

FWER / FDR - multiple testing view

- We test m hypotheses. (The functions are discretized in m points.)

TABLE – Possible outcomes from m hypothesis tests.

	Accept	Reject	Total
True null hypothesis	U	V	m_0
False null hypothesis	T	S	m_1
Total	W	ρ	m

- FWER is defined to be $\mathbb{P}(V \geq 1)$.
- FDR is defined to be $\mathbb{E}(Q)$, where $Q = \frac{V}{\rho}$ whenever $\rho > 0$, and $Q = 0$ if $\rho = 0$.
- FWER provides more strict control than FDR. Only in the case of all hypothesis being true null (i.e. $m_0 = m$), the FDR control implies FWER control.

- In order to find all hypothesis which should be rejected the FDR control is more practical, therefore we are working now on FDR envelopes.
- Checking the global hypothesis, FWER is used.

FDR envelope

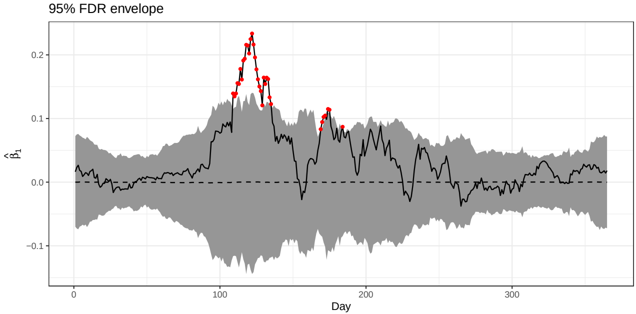


FIGURE – FDR envelope test for testing influence of the year on the Rimov reservoir temperature.

More days in a year are identified to be significant.

Continuous min-max envelope

Assume now, that we have s resamples of the complete null hypothesis, i.e. $m_0 = m$, which gives us s test vectors $\mathbf{T}_i = (T_{i1}, \ldots, T_{im}), i = 1, \ldots, s$. Let us define the lower and upper bounds of the envelope $\mathcal{E}_\gamma$ for the two-sided alternative hypotheses. For γ being an integer, $\gamma = 1, 2, 3, \ldots, \lfloor s/2 \rfloor$, the bounds are defined by

$$\mathcal{E}^{\text{low}}_{\gamma,k} = \min{}^{\gamma}\{T_{ik} : i = 1, \ldots, s\}, \quad \mathcal{E}^{\text{upp}}_{\gamma,k} = \max{}^{\gamma}\{T_{ik} : i = 1, \ldots, s\}$$

for $k = 1, \ldots, m$, where $\min^\gamma$ and $\max^\gamma$ denote the γ-th smallest and largest values, respectively. For a non-integer $\gamma > 1$, we define the bounds of $\mathcal{E}_\gamma$ as the average of neighboring integer γs :

$$\mathcal{E}^{\text{low}}_{\gamma,k} = \mathcal{E}^{\text{low}}_{[\gamma],k} + (\gamma - [\gamma]) \cdot (\mathcal{E}^{\text{low}}_{[\gamma]+1,k} - \mathcal{E}^{\text{low}}_{[\gamma],k}), \quad \mathcal{E}^{\text{upp}}_{\gamma,k} = \mathcal{E}^{\text{upp}}_{[\gamma],k} - (\gamma - [\gamma]) \cdot (\mathcal{E}^{\text{upp}}_{[\gamma],k} - \mathcal{E}^{\text{upp}}_{[\gamma]+1,k}).$$

For $0 < \gamma < 1$, we define the envelope $\mathcal{E}_\gamma$ as an extension of $\mathcal{E}_1$:

$$\mathcal{E}^{\text{low}}_{\gamma,k} = \mathcal{E}^{\text{low}}_{1,k} + \log\gamma \cdot (\mathcal{E}^{\text{upp}}_{1,k} - \mathcal{E}^{\text{low}}_{1,k}), \quad \mathcal{E}^{\text{upp}}_{\gamma,k} = \mathcal{E}^{\text{upp}}_{1,k} - \log\gamma \cdot (\mathcal{E}^{\text{upp}}_{1,k} - \mathcal{E}^{\text{low}}_{1,k}).$$

FDR envelope

For a fixed $\Gamma = \mathcal{E}_{\gamma}^{C}$, we can observe only $\rho(\Gamma)$, the number of positives, together with the corresponding numbers of acceptances, $W(\Gamma)$. The number of false positives $V(\Gamma)$ is not observable, but it can be estimated according to Storey 2002 as

$$\mathbb{E}\left[\frac{V(\Gamma)}{\rho(\Gamma)}\right] \approx \frac{\mathbb{E}V(\Gamma)}{\max(\rho(\Gamma),1)} = \pi_0 \frac{\mathbb{E}\rho^0(\Gamma)}{\max(\rho(\Gamma),1)}, \tag{7}$$

where $\rho^0(\Gamma)$ is the number of positives under complete null hypothesis. Unfortunately the Storey's approach has serious problems for highly correlated test statistics with estimation of π_0.

Benjamini Hochberg Adaptive two stage procedure (ATS)

1. Use the linear step-up procedure at level $\alpha^* = \alpha/(1+\alpha)$. Let r_1 be the number of rejected hypotheses. If $r_1 = 0$ do not reject any hypothesis and stop ; if $r_1 = m$ reject all m hypotheses and stop ; otherwise continue.
2. Let $\hat{\pi}_0 = (m - r_1)/m$.
3. Use the linear step-up procedure with $\alpha' = \alpha/\hat{\pi}_0$.

The linear step-up procedure uses m p-values $\{p_1, \ldots, p_m\}$ correspondent to the m hypotheses and the ordered p-values $p_{[1]} \leq p_{[2]} \leq \cdots \leq p_{[m]}$. It rejects k hypotheses corresponding to the first k smallest p-values, where $k = \max\{i : p_{[i]} \leq i\alpha/m\}$. If such k does not exists, it rejects no hypothesis.

In the first step of the algorithm, the numbers of rejections are estimated assuming that $\pi_0 = 1$. In the second step, an estimate for π_0 is found, and the third step finds the rejections utilizing the estimate of π_0.

This algorithm is not possible to be used for envelope construction, it only estimates the hypotheses which should be rejected.

Adaptive two-stage envelope (ATSE)

1. Find the largest rejection region Γ_{γ^*} for which
$$\frac{\mathbb{E}\rho^0(\Gamma_{\gamma^*})}{\max(\rho(\Gamma_{\gamma^*}),1)} \leq \alpha/(1+\alpha).$$
Let r_1 be the number of rejected hypotheses for Γ_{γ^*}. If $r_1 = 0$, do not reject any hypothesis, take $\Gamma_\gamma = \Gamma_{\gamma^*}$ and stop ; if $r_1 = m$ reject all m hypotheses, take $\Gamma_\gamma = \Gamma_{\gamma^*}$ and stop ; otherwise continue.
2. Let $\hat{\pi}_0 = (m - r_1)/m$.
3. Find the largest rejection region Γ_γ for which
$$\frac{\mathbb{E}\rho^0(\Gamma_\gamma)}{\max(\rho(\Gamma_\gamma),1)} \leq \alpha/\hat{\pi}_0.$$

As the estimate of $\mathbb{E}\rho^0(\Gamma_\gamma)$ in the steps 1. and 3. of the algorithm we use
$$\mathbb{E}\rho^0(\Gamma_\gamma) \cong m\tilde{\gamma}, \tag{8}$$
$\tilde{\gamma}$ is the pointwise p-value correspondent to Γ_γ. That is, $\tilde{\gamma} = 2\gamma/s$ in the two-sided case. Note that there is equality in Equation (8) for an integer γ.
We did not prove the conservativeness of ATSE algorithm.

Iterative adaptive two-stage envelope (IATSE)

1. Find the largest rejection region Γ_{γ^*} for which
$$\frac{\mathbb{E}\rho^0(\Gamma_{\gamma^*})}{\max(\rho(\Gamma_{\gamma^*}),1)} \leq \alpha.$$
Let r be the number of rejected hypotheses for Γ_{γ^*}. If $r=0$, do not reject any hypothesis, take $\Gamma_\gamma = \Gamma_{\gamma^*}$ and stop ; otherwise continue.
2. Let $\tilde{\gamma}^*$ be the pointwise p-value corresponding to the Γ_{γ^*} and
$$\hat{\pi}_0 = \min\left(1, \frac{(m-r)/m}{1-\tilde{\gamma}^*}\right). \tag{9}$$
3. Find the largest rejection region Γ_γ for which
$$\frac{\mathbb{E}\rho^0(\Gamma_{\gamma})}{\max(\rho(\Gamma_{\gamma}),1)} \leq \alpha/\hat{\pi}_0.$$

Note here that the ATSE method controls the estimate of π_0 by assuming a greater significance level in the first step, in the same way as the ATS algorithm. On the other hand, the IATSE method adds the expected number of false null hypotheses to the estimate of π_0.

Conservativeness

Theorem

The estimate $\hat{\pi}_0$ of the IATSE algorithm is a conservative estimate of π_0.

Functional GLM

Functional GLM

- What if we have nuisance factors ? The data cannot be considered exchangeable even under the null hypothesis.
- Consider the functional GLM (Winkler et al., 2014)

$$\mathbf{Y}(r) = \mathbf{X}(r)\beta(r) + \mathbf{Z}(r)\gamma(r) + \epsilon(r),$$

  where the argument $r \in R$ determines the argument of the vector, or voxel of an image.
- No exact permutation schemes, but many approximative ones : Freedman-Lane (1983) was deduced to be one of the best by Winkler et al. (2014) [1].
- Again F-statistic is often used. If there is inhomogeneity and non-normality in the error term, the quantiles of the F-statistic vary and the p-min / global envelope tests can have higher power than the F-max test [2].

---

1. Winkler, Ridgway, Webster, Smith, Nichols (2014) Permutation inference for the general linear model. Neuroimage 92, 381–397.
2. Mrkvička, T., Myllymäki, M., Kuronen, M. and Narisetty, N. N. (2021) New methods for multiple testing in permutation inference for the general linear model. Statistics in Medicine 41(2), 276–297. doi :10.1002/sim.9236

Global envelopes for functional ANOVA & GLM

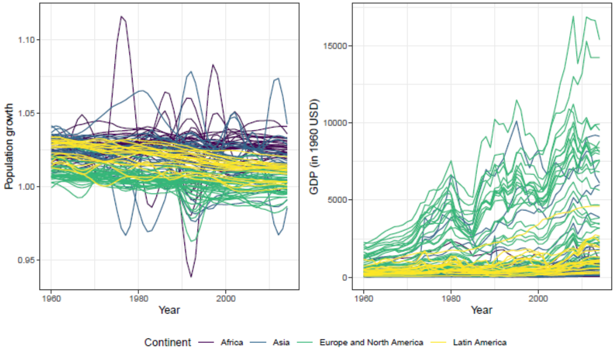
- One is not restricted to the use of F-statistic as the test statistic !
- A possibility is to use the $\hat{\beta}(r)$-coefficients [3]
- We have proposed two different versions for a discrete factor (function `graph.flm` of GET) :
  - if any of the group deviates from the overall functional mean, which groups deviate from the mean and where the groups deviate from the mean
  - a test statistic in the sense of post-hoc comparison in the ANOVA model : This test determines, at the global significance level, which groups differ from each other and where they differ
- These test statistics are similar to those proposed for us first for the case of the one-way functional ANOVA [4]

---

3. Mrkvička, Roskovec & Rost (2021) A nonparametric graphical tests of significance in functional GLM. Methodology and Computing in Applied Probability 23, 593-612. DOI: 10.1007/s11009-019-09756-y

4. Mrkvička, T., Myllymäki, M., Jílek, M. and Hahn, U. (2020) A one-way ANOVA test for functional data with graphical interpretation. Kybernetika 56 (3), 432-458. DOI : 10.14736/kyb-2020-3-0432

Population growth example & FDR envelopes



Population growth and GDP (in 1960 USD) curves for 1960–2014 for 112 countries from the four continents. (Source : World bank)

Population growth example & FDR envelopes

- We explore the functional general linear model (GLM) for the population growth across years 1960-2014 with categorical covariate (continent), continuous covariate (GDP), and interactions (GDP with respect to the continent).
- For this purpose, we applied three tests, and on each of them, we applied the proposed FDR control (IATSE) to obtain all years which are significant for the studied covariate.

Categorical factor

First, to test the effect of the continent, we assume the main effects model

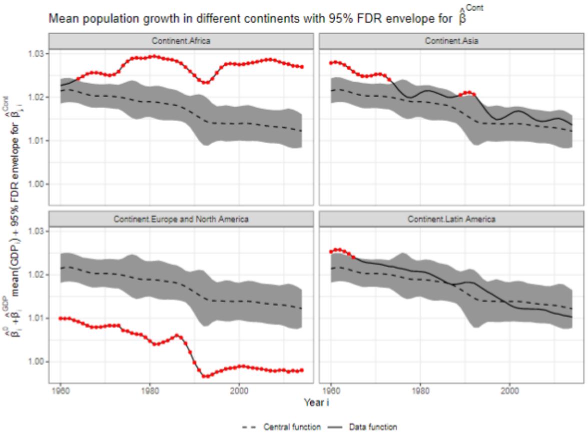
$$\text{population growth} \sim \text{Continent} + \text{GDP},$$

and we used the test statistic

$$\boldsymbol{\beta}^{\text{Cont}} = (\underbrace{\beta^{\text{Cont}}_{1,1960}, \dots, \beta^{\text{Cont}}_{1,2014}}_{\text{1: Africa}}, \underbrace{\beta^{\text{Cont}}_{2,1960}, \dots, \beta^{\text{Cont}}_{2,2014}}_{\text{2: Asia}}, \underbrace{\beta^{\text{Cont}}_{3,1960}, \dots, \beta^{\text{Cont}}_{3,2014}}_{\text{3: Europe and North America}}, \underbrace{\beta^{\text{Cont}}_{4,1960}, \dots, \beta^{\text{Cont}}_{4,2014}}_{\text{4: Latin America}}),$$

where $\beta^{\text{Cont}}_{1,i}, \dots \beta^{\text{Cont}}_{4,i}$ are the parameters of the univariate GLM model for year i, computed with condition $\sum_{j=1}^{4} \beta^{\text{Cont}}_{j,i} = 0$.

Categorical factor

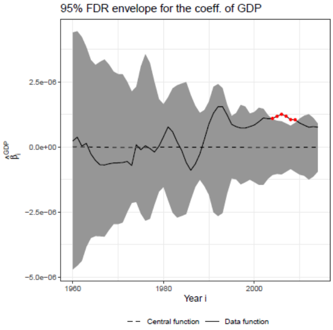


The advantage of this test statistic with respect to the F statistic is that we directly test which continents differ from the mean. (F-statistic based tests reject for all years. ATS+β-statistic does not tell about directions.)

Continuous factor

Second, to test the effect of the GDP, we assumed the same main effects model as above and we used simply the test statistic

$$\boldsymbol{\beta}^{\text{GDP}} = (\beta_{1960}^{\text{GDP}}, \ldots, \beta_{2014}^{\text{GDP}}). \tag{10}$$



Note : Rather decreasing variability of the test statistics, which makes no problem to the procedure itself, but it is worth noting for the interpretation.

Interactions

Third, to test the effect of the interactions, we assume the factorial model

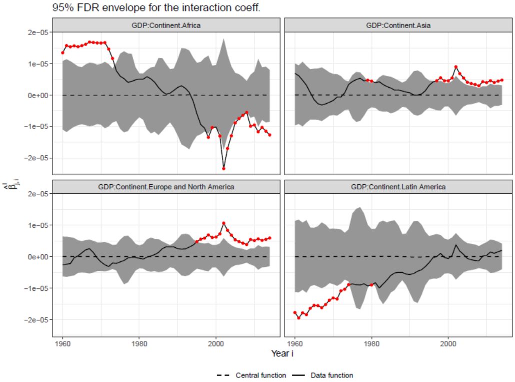
$$\text{population growth} \sim \text{Continent} * \text{GDP},$$

and we used the test statistic

$$\boldsymbol{\beta}^I = (\underbrace{\beta^I_{1,1960}, \dots, \beta^I_{1,2014}}_{\text{1: Africa}}, \underbrace{\beta^I_{2,1960}, \dots, \beta^I_{2,2014}}_{\text{2: Asia}}, \underbrace{\beta^I_{3,1960}, \dots, \beta^I_{3,2014}}_{\text{3: Europe and North America}}, \underbrace{\beta^I_{4,1960}, \dots, \beta^I_{4,2014}}_{\text{4: Latin America}}), \tag{11}$$

where $\beta^I_{1,i}, \dots \beta^I_{4,i}$ are interaction parameters of the univariate GLM model for year i, computed with condition $\sum_{j=1}^4 \beta^I_{j,i} = 0$.

Interactions



- The population growth of countries in Africa is significantly positively influenced by GDP up to 1974, whereas it is significantly less influenced than the whole world from 1998 onwards.
- The population growth of Asian countries is more positively influenced by GDP than the whole world from 1996 onwards. The same holds for Europe and North America.
- On the other hand, the population growth in Latin American countries is negatively influenced by GDP up to 1974.

References & Thank you for your attantion

- Dai W., Mrkvička T., Sun Y., Genton M. C. (2020). *Functional Outlier Detection and Taxonomy by Sequential Transformations*, Computational statistics and data analysis , 149, 16960.
- Dai W., Athanasiadis S., Mrkvička T. (2021). *A new functional clustering method with joined dissimilarity sources and graphical interpretation*, in Computational statistics : Intex open.
- Dvořák J., Mrkvička T. (2021). *Graphical tests of independence for general distributions*, Computational statistics
- Mrkvička T., Myllymäki M., Hahn U. (2017). *Multiple Monte Carlo Testing with Applications in Spatial Point Processes*, Statistics and computing 27/5, 1239–1255.
- Mrkvička T., Roskovec T., Rost M. (2019). *A nonparametric graphical tests of significance in functional GLM*, Methodology and computing in applied probability.
- Mrkvička T., Myllymäki M., Jílek M., Hahn U. (2020). *A one-way ANOVA test for functional data with graphical interpretation*, Kybernetika.
- Mrkvička T., Myllymäki M., Kuronen M., Narisetty N. (2021) : New methods for multiple testing in permutation inference for the general linear model, https ://arxiv.org/abs/1906.09004.
- Myllymäki, M., Mrkvicka, T., Seijo, H., Grabarnik, P. and Hahn, U. (2017). *Global envelope tests for spatial processes.* JRSSB.