WORKING PAPERS

N° 1650

July 2025

"Testing mean densities with an application to climate change in Vietnam"

Camille Mondon, Thi-Huong Trinh, Josep Antoni Martín-Fernández and Christine Thomas-Agnan

Toulouse School of Economics

Testing mean densities with an application to climate change in Vietnam

July 1, 2025

Camille Mondon[1*], Huong Thi Trinh[2], Josep Antoni Martín-Fernández[3], Christine Thomas-Agnan[4]

[1*]Faculty of Mathematical Economics, Thuongmai University, Ho Tung Mau street, Hanoi, 10000, Vietnam.
[2]Toulouse School of Economics, University of Toulouse Capitole, Toulouse, 31000, France.
[3]Department IMAE, Universitat de Girona, Campus Montilivi. Edifici P4. Carrer Maria Aurèlia Capmany i Farnés, 61, Girona, E-17003, Spain.

*Corresponding author(s). E-mail(s): camille.mondon@tse-fr.eu;
Contributing authors: trinhthihuong@tmu.edu.vn; josepantoni.martin@udg.edu; christine.thomas@tse-fr.eu;

Abstract

Given samples of density functions on an interval (a, b) of $\mathbb{R}$, categorized according to a factor variable, we aim to test the equality of their mean functions both overall and across the groups defined by the factor. While the Functional Analysis of Variance (FANOVA) methodology is well-established for functional data, its adaptation to density functions (DANOVA) is necessary due to their inherent constraints of positivity and unit integral. To accommodate these constraints, we naturally use Bayes spaces methodology by mapping the densities using the centered log-ratio transformation into the $L_0^2(a, b)$ space where we can use FANOVA techniques. Many traditional contrasts in FANOVA rely on squared differences and can be reinterpreted as squared distances between Bayes perturbations within the densities space. We illustrate our methodology on a dataset comprising daily maximum temperatures across Vietnamese provinces between 1987 and 2016. Within the context of climate change, we first investigate the existence of a non-zero temporal trend of the densities of daily maximum temperature over Vietnam and then examine whether there is any regional effect on these trends. Finally, we explore odds ratio based interpretations allowing to describe the trends more locally.

Keywords: Analysis of variance, Density data, Functional data, Log ratio, Odds ratio, Bayes spaces

1 Introduction

Climate change remains a central focus of scientific inquiry, as its effects on weather patterns, ecosystems, and human livelihoods become increasingly pronounced (Hultgren et al., 2022). According to the World Meteorological Organization, the global mean temperature during the 2011–2020 period has been $1.10 \pm 0.12°$C higher than the average recorded between 1850 and 1900. Notably, weather-related events accounted for nearly 94% of all recorded disaster-induced displacements over the past decade (World Meteorological Organization, 2023). Among the various global climate indicators, near-surface temperature is particularly critical, as it directly influences human well-being and daily activities (Gubler et al., 2023; World Meteorological Organization, 2024; Schlenker and Roberts, 2009). Temperatures can be measured using various indicators, such as heat stress, growing degree days, killing degree days, temperature intervals, or the entire temperature distribution (Roberts et al., 2013; Vo et al., 2022; Espagne et al., 2019; Schlenker and Roberts, 2009; Hultgren et al., 2022). Indeed, considering the entire temperature distribution captures the full range of temperatures throughout each day and over multiple days, offering a more comprehensive and informative perspective (Hsiang et al., 2017; Trinh et al., 2023). Vietnam lies closer to the tropics than the equator and is significantly influenced by the East Sea, resulting in a predominantly tropical monsoon climate. The country's economy is heavily reliant on agriculture, supported by its fertile deltas, mountainous regions, and extensive coastline. Vietnam has been identified as one of the five countries most vulnerable to the impacts of climate change (World Bank Group and Asian Development Bank, 2021; Trinh et al., 2021). Since 1960, the country's mean annual temperature has risen by approximately 0.5°C to 0.7°C, with an estimated warming rate of 0.26°C per decade between 1971 and 2010 (World Bank Group and Asian Development Bank, 2021). Consequently, as in many other nations, analyzing climate change in Vietnam is of particular interest due to its significant implications for agriculture, especially rice cultivation (Tran and Nguyen, 2021; Trinh et al., 2021; Trinh, 2018)

With the increasing volume of recorded data, data science has seen the rise of new types of observations like functional data. Functional data analysis is currently a very active field of statistics, see Aneiros et al. (2022). Of particular interest for our application are density-valued data, which can be found in other areas of social sciences (for example age distributions, income distributions, expenditure distributions), and in other fields (for example particle size distributions). Density data require a specific treatment within the framework of functional data analysis, due to their inherent constraints, namely non-negativity and integration to one. It is typically assumed that each observation corresponds to a sampled continuous density function, belonging to an infinite-dimensional function space, as discussed in Petersen et al. (2022). A smoothing tool is necessary to fill the gap between the discrete data and the continuous temperature density objects and this procedure step is called preprocessing. Among the methods described in Petersen et al. (2022), we use the Bayes spaces approach first introduced in Egozcue et al. (2006), and later on developed in Van Den Boogaart et al. (2010, 2014).

In this paper, we focus on testing mean density functions, investigating their equality to a reference density as well as their equality across groups. Since density functions share similar constraints with compositional vectors, albeit in a continuous form, we build upon techniques from compositional data analysis (CoDA) but also from the functional analysis of variance (FANOVA) framework, introduced in Ramsay and Silverman (2005); Kokoszka and Reimherr (2017).

Before comparing group means, a first question can be to find whether the expected density is equal to a given reference, for example the uniform density. We adapt the one-sample test from the functional framework (see Zhang (2013)) to density functions in Section 4.1.

As outlined by Martín-Fernández et al. (2015), the analysis of grouped data typically begins with testing the equality of group means and a widely used approach for this purpose in CoDA is the multivariate analysis of variance (MANOVA) contrast, which needs to be adapted here to continuous objects. On the other hand, the FANOVA techniques must be adapted when applied to density functions because they reside in the constrained space $B^2(a,b)$. We present five test statistics in Section 4.2 for the problem of testing the equality of group means of density functions, which we call DANOVA for distributional analysis of variance.

When an ANOVA test leads to a rejection of the null hypothesis of equal group means, a related question of interest is the pairwise comparison of group means (see for example Martín-Fernández et al., 2015, who present additional techniques for interpreting the differences between the groups in CoDA). This question is treated in Section 4.3.

Finally, to go beyond the global comparisons of densities and do a more local analysis, we adapt a technique based on odds ratios from Maier et al. (2025) to infer the relative mass of the densities over specific intervals.

We apply the above tools to address different questions relative to climate change in Vietnam. We use the distributions of maximum temperatures in the provinces of Vietnam to compare the six administrative regions in terms of climate change. The dataset and its preprocessing are presented in Section 3. After constructing a trend slope density for each province summarizing its time evolution, we investigate maximum temperature distribution changes through these trend slope densities sample. A one-sample test comparing the mean slope density to the uniform distribution addresses the question of the existence of a climate change in the whole of Vietnam in Section 4.1. Then, an analysis of variance of the slope densities in Section 4.2 detects whether the climate change is the same across regions. Finally in Section 5, we compare the relative frequencies in different subintervals of the temperature distribution using infinitesimal odds ratios introduced in Maier et al. (2025).

2 Framework and reminders

2.1 Reminders on functional analysis of variance

In the classical framework of Functional Analysis of Variance, or simply FANOVA (as presented in Zhang, 2013, p. 144), we observe G independent functional samples, denoted by $(f_{g1}, \dots, f_{gn_g})$ for $1 \le g \le G$, from stochastic processes with values in $L^2(a,b)$, satisfying for $1 \le i \le n_g$ and $a \le x \le b$

$$f_{gi}(x) = f_g(x) + v_{gi}(x), \tag{1}$$

where $f_g(x) = \mathbb{E}(f_{gi})(x)$ is the unknown mean function in group g and the stochastic error process v_{gi} has mean 0 and common covariance operator. The total sample size is $n = \sum_{g=1}^{G} n_g$. The overall sample mean curve $\bar{f}_{\cdot\cdot}$ and the group sample mean curves $\bar{f}_{g\cdot}$ are respectively defined by

$$\bar{f}_{\cdot\cdot}(x) = \frac{1}{n}\sum_{g=1}^{G}\sum_{i=1}^{n_g} f_{gi}(x) \tag{2}$$

$$\bar{f}_{g\cdot}(x) = \frac{1}{n_g}\sum_{i=1}^{n_g} f_{gi}(x) \tag{3}$$

The pointwise between-group mean square error and the pointwise within-group mean square error at x are respectively defined by

$$\mathrm{SSB}(x) = \sum_{g=1}^{G} n_g \left(\bar{f}_{g\cdot}(x) - \bar{f}_{\cdot\cdot}(x)\right)^2 \text{ and} \tag{4}$$

$$\mathrm{SSW}(x) = \sum_{g=1}^{G}\sum_{i=1}^{n_g} \left(f_{gi}(x) - \bar{f}_{g\cdot}(x)\right)^2. \tag{5}$$

Ramsay and Silverman (2005) extend the classical F-test and propose the pointwise functional F-ratio to test the equality of the group mean curves at a given point x, using the local F-statistic

$$F(x) = \frac{\mathrm{SSB}(x)}{\mathrm{SSW}(x)}. \tag{6}$$

For a global assessment of the equality of the group mean curves on the whole interval of variation of their argument, Zhang and Chen (2007) and Zhang (2013) introduce L^2-norm based tests as well as F-type test statistics. Later on, Zhang and Liang (2014) propose the GPF test based on the integral of the pointwise F-ratio statistic and the $F_{\max}$ test based on the maximum of the pointwise F-test. It is then necessary to approximate the null distribution of these statistics. This can be achieved using a permutation based procedure as in Ramsay and Silverman (2005) or a bootstrap procedure as in Zhang et al. (2019).

2.2 Reminders on distributional data analysis and Bayes spaces

Egozcue et al. (2006) and Van Den Boogaart et al. (2010, 2014) define the Bayes spaces of probability density functions relative to a reference measure λ on an interval of $\mathbb{R}$, in a similar fashion as $L^2(\lambda)$ spaces. In the following, the measure λ will be Lebesgue measure on a finite interval $[a,b]$ and we will denote these spaces by $B^2(a,b)$. We consider the separable Hilbert space $L_0^2(a,b)$ of square-integrable functions with a zero integral on (a,b) equipped with the inner product $\langle f,g\rangle_{L^2} = \int_a^b fg\,\mathrm{d}\lambda$. For any measurable function $p : [a,b] \to \mathbb{R}$ that is positive almost everywhere and such that the function $\log(p)$ is integrable, we can define its centered log-ratio transform $\mathrm{clr}(p)$

$$x \in [a,b] \mapsto \log(p(x)) - \frac{1}{b-a}\int_a^b \log(p)(u)\mathrm{d}u. \tag{7}$$

Note that by construction $\mathrm{clr}(p) \in L_0^2(a,b)$. Conversely for each $f \in L_0^2(\lambda)$, the equivalence class $\mathrm{clr}^{-1}(f) = \{\alpha\exp(f), \alpha > 0\}$ contains positive functions that are equal almost everywhere, up to a multiplicative constant. Among them there is a unique probability density function p (thus satisfying $\int_a^b p(u)\,\mathrm{d}u = 1$) that we use to represent $\mathrm{clr}^{-1}(f)$.

Then the Bayes Hilbert space with Lebesgue reference measure on the interval $[a,b]$ is the set of probability density functions

$$B^2(a,b) = \left\{\mathrm{clr}^{-1}(f), f \in L_0^2(a,b)\right\} \tag{8}$$

equipped with the only separable Hilbert space structure $(\oplus, \odot, \langle\cdot,\cdot\rangle_{B^2})$ that makes the centered log-ratio transform an isometry between $(B^2(a,b), \oplus, \odot, \langle\cdot,\cdot\rangle_{B^2})$ and $(L_0^2(a,b), +, \cdot, \langle\cdot,\cdot\rangle_{L^2})$. The resulting addition $\oplus$ is called Aitchison perturbation ($\ominus$ denoting the negative perturbation) and the scalar multiplication $\odot$ is called Aitchison powering.

Following Van Den Boogaart et al. (2014), it is also possible to use the centered log-ratio transform in order to transport to $B^2(a,b)$ the Borel sets, as well as the expectation and covariance of $L_0^2(a,b)$-valued random variables. For a random density π in $B^2(a,b)$, one can define the expected value in the Bayes space:

$$\mathbb{E}^B(\pi) = \mathrm{clr}^{-1}(\mathbb{E}[\mathrm{clr}(\pi)]) \tag{9}$$

and the covariance operator, for $\phi \in B^2(a,b)$:

$$\begin{aligned} \mathrm{Cov}^B[\pi](\phi) &= \mathrm{clr}^{-1}(\mathrm{Cov}^B[\mathrm{clr}\,\pi](\mathrm{clr}\,\phi)) = \\ &\mathbb{E}^B\left[\langle \pi \ominus \mathbb{E}^B(\pi), \phi\rangle_{B^2} \odot \left(\pi \ominus \mathbb{E}^B(\pi)\right)\right]. \end{aligned} \tag{10}$$

Notations for distributional analysis of variance in Bayes spaces

In order to avoid confusion, we adopt a specific notation to distinguish ordinary (unconstrained) functions from densities in a Bayes space. We observe G independent density samples, denoted by $(\pi_{g1}, \dots, \pi_{gn_g})$ for $1 \le g \le G$, from stochastic processes with values in $B^2(a,b)$, satisfying for $1 \le i \le n_g$ and $a \le x \le b$

$$\pi_{gi}(x) = (\pi_g \oplus u_{gi})(x), \tag{11}$$

where $\pi_g(x) = \mathbb{E}^B(\pi_{gi})(x)$ is the unknown mean density in group g and the stochastic error process u_{gi} has mean equal to the uniform distribution on (a,b) and common covariance operator (defined by (10) in Bayes spaces). The total sample size is $n = \sum_{g=1}^G n_g$. As above the overall sample mean density is defined as

$$\bar{\pi}_{..}(x) = \frac{1}{n} \odot \bigoplus_{g=1}^{G} \bigoplus_{i=1}^{n_g} \pi_{gi}(x)$$

and the sample mean density in group g as

$$\bar{\pi}_{g.}(x) = \frac{1}{n_g} \odot \bigoplus_{i=1}^{n_g} \pi_{gi}(x).$$

Applying the FANOVA formulas to the clr-transformed densities, we adapt the FANOVA framework to densities and define the pointwise between-group mean square error and the pointwise within-group mean square error at x by

$$\mathrm{SSB}(x) = \sum_{g=1}^{G} n_g (\mathrm{clr}(\bar{\pi}_{g.})(x) - \mathrm{clr}(\bar{\pi}_{..})(x))^2 \tag{12}$$

$$\mathrm{SSW}(x) = \sum_{g=1}^{G} \sum_{i=1}^{n_g} (\mathrm{clr}(\pi_{gi})(x) - \mathrm{clr}(\bar{\pi}_{g.})(x))^2 \tag{13}$$

3 Data description and preprocessing

3.1 Description

The climate data used in this study comprises daily temperatures and precipitation, sourced from the Climate Prediction Center (CPC) database, developed by the National Oceanic and Atmospheric Administration (NOAA). This data provides historical records of the maximum and minimum temperatures for a 0.5-degree by 0.5-degree grid of latitude and longitude. For this study, we focus exclusively on maximum temperatures as they are a direct indicator of heat stress on both humans and the environment. The data spans the years 1987 to 2016. In order to address the discrepancy between the geographical coordinates and the municipalities of Vietnam (63 provinces), we calculate the average of all grid cells or the value of the four nearest localities to impute the missing data. This process results in 365 or 366 daily maximum temperature records for each province and year.

During the study period, maximum temperatures (hereafter denoted $T_{\max}$) range from -5°C to nearly 45°C. However, extreme cold and extreme heat are rare. For the purpose of our analysis, we focus on a temperature range of 12°C to 40°C. Temperatures below 12°C are adjusted upward to 12°C, while temperatures exceeding 40°C are capped at 40°C. As we can see on Figure 1, there are $n = 63$ provinces in Vietnam grouped into $G = 6$ socioeconomic regions. We use the following acronyms for the regions: NMM for Northern Midlands and Mountains region ($n_1 = 14$ provinces), RRD for Red River Delta region ($n_2 = 12$ provinces), NCC for North Central Coast region ($n_3 = 13$ provinces), CHR for Central Highlands region ($n_4 = 5$ provinces), SR for Southeast region ($n_5 = 6$ provinces) and MDR for Mekong Delta region ($n_6 = 13$ provinces). Trinh et al. (2021) highlight the fact that the regions have their own weather peculiarities. Some regions in the north have four seasons:

winter, spring, summer, and autumn, while some regions in the south experience two distinct seasons: the dry season (November to April) and the rainy season (May to October) (Trinh et al., 2021; World Bank Group and Asian Development Bank, 2021). For example, according to Trinh et al. (2021), during the period 1960-2000 winter temperatures rose faster than those of the summer, and temperatures in the northern zones rose faster than in the southern zones.

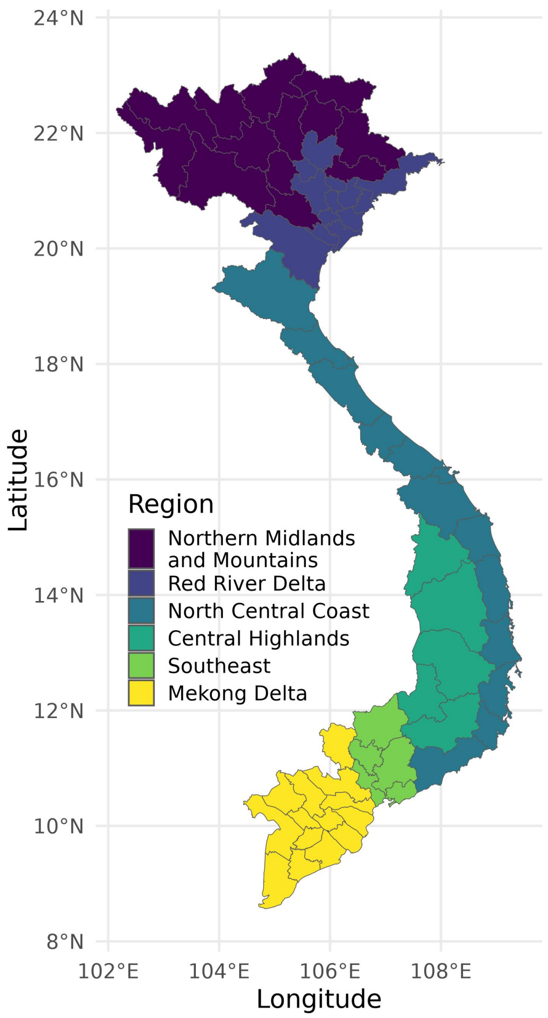

Fig. 1 Map of the provinces of Vietnam colored by region.

3.2 Smoothing

We will assume that the observed data arise from random densities, themselves sampled from an hypothesized and unknown distribution on a Bayes space. Since the elements of the Bayes spaces are density functions, whereas the observed data (usually samples of real-valued observations or histograms) is always discrete, a preprocessing step is necessary to transform the observed samples or histograms into a sample of probability density functions. We will follow the preprocessing procedure of Machalová et al. (2016, 2021) to transform our yearly samples of maximum temperatures at the province level into density functions that belong to a finite-dimensional subspace $\mathcal{C}_q^{\Delta}(a, b)$ of the Bayes Hilbert space $B^2(a, b)$, made of compositional splines. A compositional spline (CB-spline) is a probability density function whose logarithm is a polynomial spline.

The process starts by the choice of a basis of normalized B-spline functions for the space $\mathcal{S}_d^{\Delta}(a, b) \subset L^2(a, b)$, of polynomial splines or order d (degree less than or equal to $d-1$) and inside knots $\Delta = (\delta_1, \ldots, \delta_k)$, of dimension $d + k$ (see Schumaker, 1981, for a complete description). To accommodate the zero-integral constraint, Machalová et al. (2016) introduce the ZB-spline functions, denoted by $Z_\ell(x), 1 \leq \ell \leq d + k - 1$ for the subspace $\mathcal{Z}_d^{\Delta}(a, b) = \mathcal{S}_d^{\Delta}(a, b) \cap L_0^2(a, b)$ of dimension $d + k - 1$, the loss of one dimension being due to the zero-integral constraint. Finally the inverse clr of the ZB-spline basis functions, called the CB-spline basis functions, denoted by $C_\ell = \mathrm{clr}^{-1}(Z_\ell)$ generate the space $\mathcal{C}_d^{\Delta}(a, b)$, an Euclidean subspace of $B^2(a, b)$ of dimension $k + d - 1$, made of compositional splines on $[a, b]$ of order d with knots sequence Δ. The expansion (14) of a density π of the subspace $\mathcal{C}_d^{\Delta}(a, b)$ of $B^2(a, b)$ and the corresponding expansion (15) of its clr transform in the corresponding ZB-spline basis generating the space $\mathcal{Z}_d^{\Delta}(a, b)$ are then given by:

$$\pi(x) = \bigoplus_{\ell=1}^{d+k-1} [\pi]^{C_\ell} \odot C_\ell(x), \tag{14}$$

$$\mathrm{clr}(\pi)(x) = \sum_{\ell=1}^{d+k-1} [\pi]^{C_\ell} Z_\ell(x), \tag{15}$$

where $[\pi]^{C_\ell}$ is the ℓ-th coefficient of π in the CB-spline basis C. Note that the coefficients are the same in the two equations. The coefficients of the global sample mean and regional sample means are readily obtained by

$$\mathrm{clr}\, \bar{\pi}_{g.}(x) = \frac{1}{n_g} \sum_{i=1}^{n_g} \mathrm{clr}\, \pi_{gi}(x) \tag{16}$$

$$= \frac{1}{n_g} \sum_{i=1}^{n_g} \sum_{\ell=1}^{d+m-1} [\pi_{gi}]^{C_\ell} \cdot Z_\ell(x) \tag{17}$$

$$= \sum_{\ell=1}^{d+m-1} \left(\frac{1}{n_g} \sum_{i=1}^{n_g} [\pi_{gi}]^{C_\ell} \right) \cdot Z_\ell(x), \tag{18}$$

and similarly

$$\operatorname{clr} \bar{\pi}_{..}(x) = \sum_{\ell=1}^{d+m-1} \left(\frac{1}{G} \sum_{g=1}^{G} \frac{1}{n_g} \sum_{i=1}^{n_g} [\pi_{gi}]^{C_\ell} \right) \cdot Z_\ell(x). \tag{19}$$

As in Machalová et al. (2016), after summarizing the initial temperature samples using histograms, penalized least squares splines are applied to smooth the data, treating the histogram bin centers as the first coordinates and the corresponding relative frequencies as the second coordinates. Figure 2, which displays the province-level temperature densities grouped by region, highlights distinct regional patterns in the temperature distributions.

3.3 Temporal trend by province

To capture the temporal evolution of temperature density distributions across provinces within a given region (or within the whole Vietnam), we consider a simplified trend model in which each province's density evolves linearly over time in the Bayes space sense. More precisely, if π_{gi}^t now denotes the density in province i ($i = 1, \ldots, n$) at time t, the simple time evolution model (for each province) is the following density-on-scalar regression model:

$$\pi_{gi}^t(x) = [\alpha_{gi} \oplus (t \odot \beta_{gi}) \oplus \varepsilon_{git}](x), \tag{20}$$

where β_{gi} is the trend slope density (or simply slope density) for province i in region g. When the slope density of this model is the uniform distribution on (a, b), there is no observable trend in the evolution of temperature densities.

For estimating the parameters of (20), as in Talská et al. (2018), the clr transformation in applied to (20) and the resulting equation is that of a simple multivariate regression model. The fitted slopes seem to give a good approximation of the trend evolution since the coefficients of determination (defined in the context of density-on-scalar regression in Talská et al., 2018, p. 79) range from 63% to 97% (median of 94%) across provinces, with a better fit in the south. As often done in functional data, we will treat the resulting estimators of α_{gi} and β_{gi} as our sample density dataset (and therefore do not use the hat notation). In Section 5, we also use the trend density

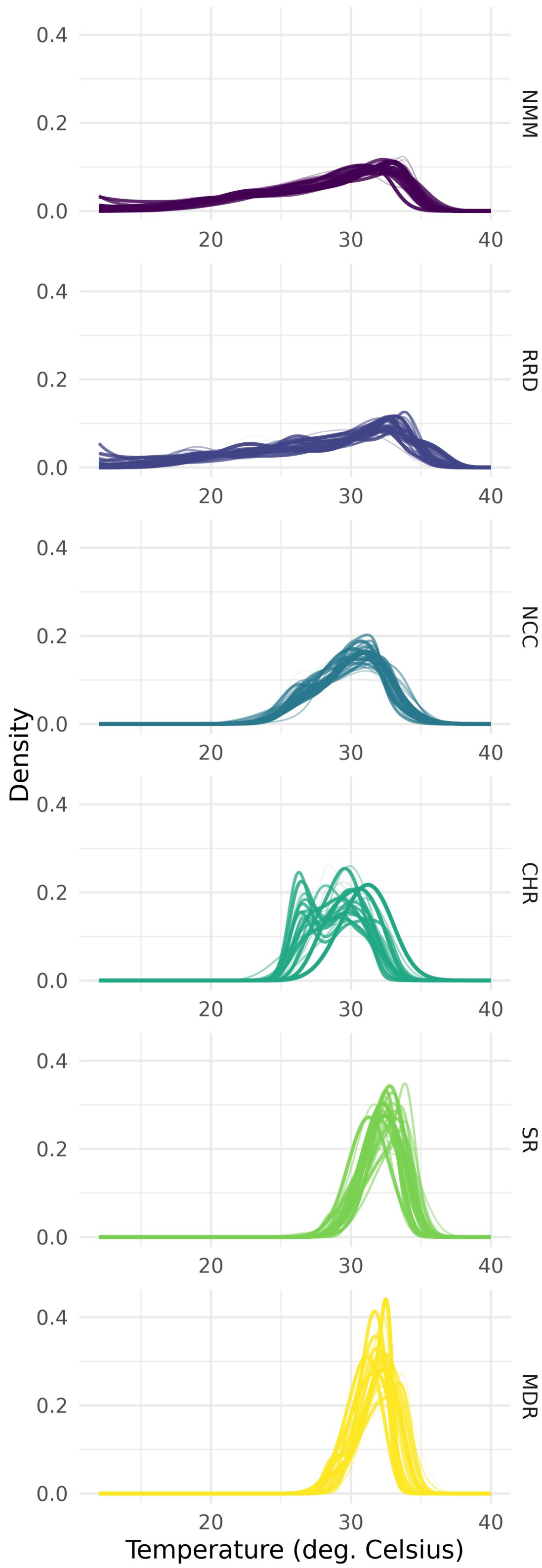

Fig. 2 Regional mean density of maximum temperature across the years 1987 to 2016.

functions $\tilde{\pi}_{gi}$ of this model defined by

$$\tilde{\pi}^t_{gi}(x) = [\alpha_{gi} \oplus (t \odot \beta_{gi})](x). \tag{21}$$

Figure 3 displays the province specific slope densities β_{gi} grouped by region.

4 Global tests

All testing procedures in this section involve hypotheses about the global behavior of mean densities in the whole interval $[a, b]$. In the application, the global tests of this section will be applied to the samples of trend slope densities in order to investigate the regional effect on climate change.

4.1 One-sample problem

In this section, we do not consider a group effect so that $G = 1$ and we temporarily drop the g index resulting in a sample $\pi_i, i = 1, \dots n$ distributed as π. We are given a reference density π_0 of particular interest and we wish to test the equality to π_0 of the mean density $\mathbb{E}^B(\pi)$:

$$H_0 : \mathbb{E}^B(\pi) = \pi_0. \tag{22}$$

Since

$$\begin{aligned} \mathbb{E}^B(\pi) = \pi_0 &\Leftrightarrow \operatorname{clr} \mathbb{E}^B(\pi) = \operatorname{clr} \pi_0 \\ &\Leftrightarrow \mathbb{E}(\operatorname{clr} \pi) = \operatorname{clr} \pi_0, \end{aligned} \tag{23}$$

the global one-sample test on mean densities is equivalent to a one sample test in the $L^2(a, b)$ space on the clr transformed densities $f_i = \operatorname{clr}(\pi_i)$ for which we may apply a classical one sample test for functional data as described in Chapter 5 of Zhang (2013). As in Zhang (2013), the functional sample $(f_1, \dots, f_n)$, is assumed to arise from model (1) (without the regional index g).

Two test statistics are available: the L^2-norm test statistic (Zhang (2013)) and the PC approach (Kokoszka and Reimherr (2017)). We now adapt them to our density framework.

L^2 approach. The L^2-norm statistic yields in the Bayes space

$$T_{\text{norm}} = n \| \operatorname{clr}(\bar{\pi}_{..}) - \operatorname{clr}(\pi_0) \|^2_{L^2_0(a,b)} \tag{24}$$

The asymptotic distribution of this statistic under the null is that of a linear combination of chi-squared statistics weighted by the eigenvalues of the covariance operator, see Kokoszka and Reimherr (2017). For the computation of the

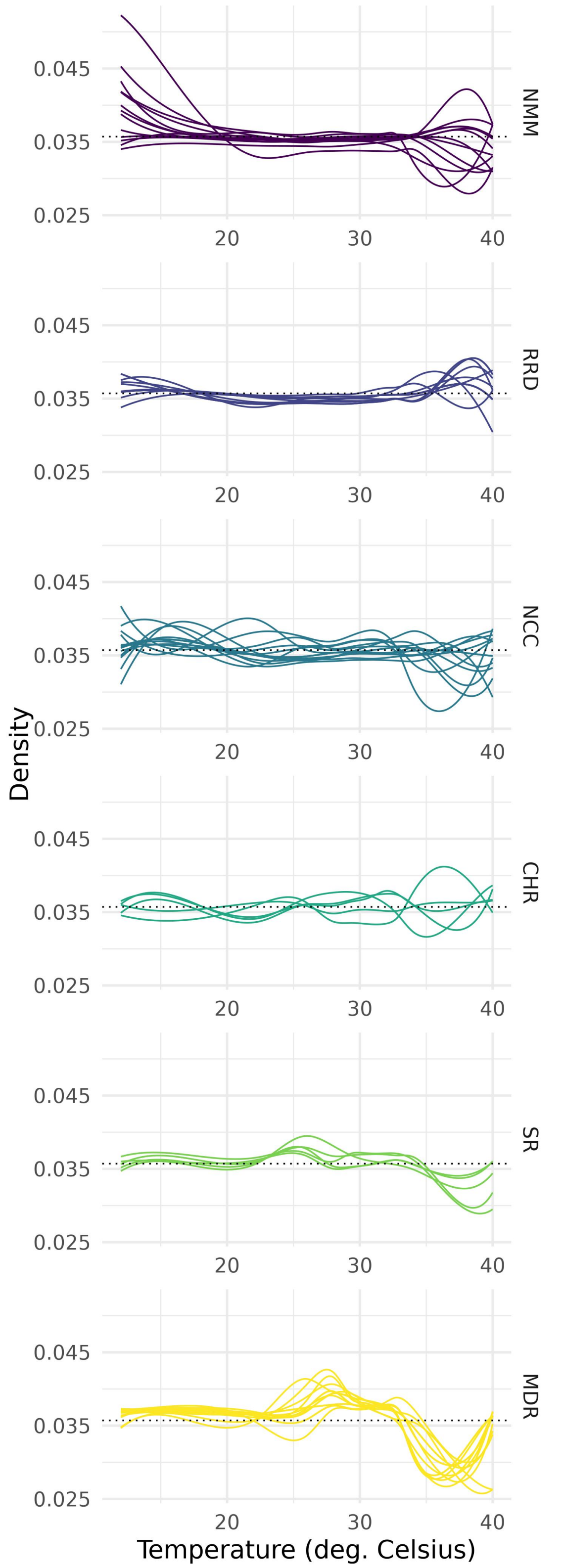

Fig. 3 Trends slopes of provinces grouped by region. The dotted line represents the uniform density on $[12°\mathrm{C}, 40°\mathrm{C}]$ which is the reference measure.

p-value of the test, the eigenvalues have to be estimated using FPCA and Kokoszka and Reimherr (2017) provide code for the computation of the L^2-norm.

PC approach. To eliminate the dependence of the limiting distribution upon the unknown eigenvalues, Kokoszka and Reimherr (2017) propose to truncate the Functional Principal Component Analysis (FPCA) and use a Hotelling-type test statistic. Keeping the first p terms, the statistic is given by

$$T_{\mathrm{PC}} = n\| \operatorname{clr}(\bar{\pi}_{..}) - \operatorname{clr}(\pi_0)\|^2_{S^{-1}} \tag{25}$$

$$= n\sum_{k=1}^{p} \frac{\langle \operatorname{clr}(\bar{\pi}_{..}) - \operatorname{clr}(\pi_0), v_k\rangle^2_{L^2_0(a,b)}}{\hat{\lambda}_k}, \tag{26}$$

where $\|.\|^2_{S^{-1}}$ is the squared norm weighted by the inverse of the covariance matrix of the clr coefficients, λ_k are the eigenvalues and v_k the eigenfunctions of the FPCA. Its limiting distribution under the null is then a simple chi-squared with p degrees of freedom. Kokoszka and Reimherr (2017) recommend using the smallest value of p for which the explained variance exceeds 85%, and to use the norm approach otherwise.

Application: evolution of temperature distributions over the period 1987-2016

Figure 4 displays the global mean trend which shows that, overall in Vietnam, there is a relative stability in the central range of temperatures [20°C, 32°C] while low temperatures (below 20°C) tend to become more frequent at the expense of very high ones above 32°C. The justification for this statement will be given in Section 5.

We perform the one-sample test for the trend slope sample of β_{gi}. To evaluate the existence of climate change, we choose π_0 to be the neutral element (uniform density on (a, b)), so that the null hypothesis represents the absence of climate change. Both tests in Table 4.1 confirm the existence of a non-uniform mean trend slope, indicative of global climate change. We use the following convention for p-values: *** indicate that the p-value is less than 1% (strong rejection), ** indicate that the p-value is less than 5% (medium rejection), * indicates that the p-value is less than 10% (weak rejection) and no star means that we cannot reject the null hypothesis based on the data.

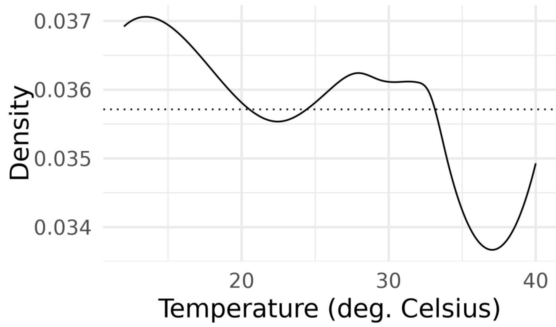

Fig. 4 Global mean trend. The dotted line represents the uniform density on $[12°\mathrm{C}, 40°\mathrm{C}]$, which is the reference measure.

Name	Statistic	p-value
PC	38	3.1e-08 ***
L^2	1.3	1.3e-06 ***

Table 1 Test statistics and p-values for the global one-sample problem.

Application: regional evolution of temperature distributions over the period 1987-2016

After testing the existence of global climate change, it may be interesting to perform a one-sample test in each region separately to see whether climate change does not affect some regions. The corresponding Šidák-adjusted p-values (see Abdi, 2010) are displayed in Table 2. This table strongly supports the existence of climate change in the Red River Delta (RRD) region and in the southern regions SR and MDR, the non-existence of climate change in the Central Highlands (CHR) region. The results are more contrasted in the Northern Midlands and Mountains (NMM) region and in the North Central Coast (NCC) region where the tests are either non-significant or very marginally significant. These results should be interpreted with caution due to the small sample size in each region.

Region	Sample size	PC	L^2
NMM	14	7.9e-02 *	0.12
RRD	12	1.6e-10 ***	1.5e-04 ***
NCC	13	8e-02 *	0.12
CHR	5	0.65	0.99
SR	6	1.8e-03 ***	4.2e-04 ***
MDR	13	0 ***	4.9e-08 ***

Table 2 Šidák-adjusted p-values for the regional one sample test.

4.2 Distributional ANOVA

Using the notations of Section 2, the objective of DANOVA is to test the null hypothesis H_0 that the G group mean densities $\mathbb{E}^B(\pi_g)$ are equal against the usual alternative that at least two means are different, where mean density here is understood in the Bayes space sense:

$$H_0 : \forall g, \quad \mathbb{E}^B(\pi_{gi}) = \mathbb{E}^B(\pi_{1i}). \tag{27}$$

Let us first note that for two densities π_1 and π_2 we have the following equivalence

$$\mathbb{E}^B(\pi_1) = \mathbb{E}^B(\pi_2) \Leftrightarrow \mathbb{E}(\mathrm{clr}\,\pi_1) = \mathbb{E}(\mathrm{clr}\,\pi_2), \tag{28}$$

since the clr of the expected density in the Bayes sense is equal to the expected value in the classical sense of the clr of the density in the corresponding L_0^2 space. Therefore, the assumption (27) is equivalent to

$$H_0 : \forall g, \quad \mathbb{E}(\mathrm{clr}\,\pi_{gi}) = \mathbb{E}(\mathrm{clr}\,\pi_{1i}). \tag{29}$$

In order to adapt FANOVA to DANOVA, we can thus simply apply the FANOVA techniques to the clr transformed densities in the functional space L_0^2. The R package `fdANOVA` (described in Górecki and Smaga, 2019) provides several FANOVA options and we focus on the following five test statistics. Their evaluation requires the computation of the following quantities: the pointwise variations $\mathrm{SSB}(x)$ and $\mathrm{SSW}(x)$ from (12) and (13), and pairwise negative perturbations between group means for the statistic proposed by Cuevas et al. (2004). The coefficients of the required clr transforms in a chosen ZB-basis are easily computed from the coefficients of the clr of $\pi_{gi}, \bar{\pi}_{g.}, \bar{\pi}_{..}$, the last two being obtained by (16) and (19). Let us briefly describe the five test statistics and try to provide when possible a Bayes space expression.

1. The L^2-norm based test statistic from Zhang and Chen (2007) is given in the framework of ANOVA by

$$\mathrm{L^2B} = \int_a^b \mathrm{SSB}(x)dx, \tag{30}$$

where $\mathrm{SSB}(x)$ is given by (4). Note that an alternative formula for L^2B when applied to clr of densities is directly given in the Bayes space by

$$\mathrm{L^2B} = \sum_{g=1}^{G} n_g \|\bar{\pi}_{g.} \ominus \bar{\pi}_{..}\|_{B^2}^2, \tag{31}$$

which shows that this statistic can be expressed as a norm and is therefore invariant under almost-everywhere equality. In the Górecki and Smaga (2019) R package, the `L2B` implementation uses an asymptotic distribution of the test statistic.

2. The F-type tests from Shen and Faraway (2004) and Zhang (2011) use both within and between variation:

$$\mathrm{FB} = \frac{\int_a^b \mathrm{SSB}(x)dx/(G-1)}{\int_a^b \mathrm{SSW}(x)dx/(n-G)} = \frac{\sum_{g=1}^{G} n_g \|\bar{\pi}_{g.} \ominus \bar{\pi}_{..}\|_{B^2}^2/(G-1)}{\sum_{g=1}^{G}\sum_{i=1}^{n_g} \|\pi_{gi} \ominus \bar{\pi}_{g.}\|_{B^2}^2/(n-G)}. \tag{32}$$

We choose the `FB` implementation (biased-reduced) for the computation of the p-value. Under the null hypothesis, this statistic is asymptotically equal to a linear combination of independent χ^2 for the numerator and the denominator, and can thus be approximated by a Fischer distribution (Zhang, 2011, Theorem 1 and equation (2.16)). The test statistic coincides with that of Van Den Boogaart et al. (2014) by the second equality in (32). However, Van Den Boogaart et al. (2014) use a bootstrap procedure rather than the above approximation.

3. The `CS` statistic from Cuevas et al. (2004) uses pairwise differences. When applied to clr of densities, due to the linearity of clr, we get

$$\mathrm{CS} = \sum_{g<g'} n_g \int_a^b \left(\mathrm{clr}\,\bar{\pi}_{g.}(x) - \mathrm{clr}\,\bar{\pi}_{g'.}(x)\right)^2 dx = \sum_{g<g'} n_g \|\bar{\pi}_{g.} \ominus \bar{\pi}_{g'.}\|_{B^2}^2 \tag{33}$$

The `CS` implementation in the `fdANOVA` package uses a heteroscedastic assumption and parametric bootstrap. Note that an alternative formula for CS is the sum of the Bayes norms of all pairwise differences between groups.

4. The GPF statistic from Zhang and Liang (2014) integrates the pointwise F-ratio instead of integrating separately the pointwise within and between variations:

$$\mathrm{GPF} = \int_a^b \frac{\mathrm{SSB}(x)/(G-1)}{\mathrm{SSW}(x)/(n-G)}dx. \tag{34}$$

Under the null hypothesis, this statistic is asymptotically equal to a linear combination of independent χ^2, which can be approximated by a χ^2-distribution (Zhang and Liang, 2014, Proposition 1). In the implementation of the `fdANOVA` package, `GPF` is divided by $b-a$, and the null distribution is modified accordingly. Górecki and Smaga (2018) (page 5) claim that the GPF test is more powerful than the F-type tests.

5. The `Fmax` statistic from Zhang et al. (2019) rather computes the supremum of the pointwise F-ratio instead of integrating it as in `GPF`:

$$F_{\max} = \sup_{x \in (a,b)} \frac{\mathrm{SSB}(x)/(G-1)}{\mathrm{SSW}(x)/(n-G)}. \tag{35}$$

We consider the implementation `Fmaxb` which bootstraps the distribution under the null hypothesis. The statistic $F_{\max}$ is the only one that is not invariant under almost-everywhere equality: changing one value of a density in the dataset might change the value of $F_{\max}$. Note also that the statistics GPF and $F_{\max}$ cannot be written straightforwardly in terms of norms in the Bayes space.

Application: regional evolution of temperature distribution over the period 1987-2016

Figure 5 displays the regional mean trend slope densities. For the central regions (CHR and NCC), this plot shows a relative stability of the maximum temperature distribution (the trend is not far from uniform). For the RRD region, the curve exhibits a noticeable bump above the uniform on the right tail. More details about the interpretation of these curves will be given in Section 5.

We now test whether these trend densities vary across regions. Table 3 summarizes the results of the global analysis of variance for which the above tests all conclude that there is a difference in the way the regional temperature densities evolve in time.

Name	Statistic	p-value
GPF	10	0 ***
Fmaxb	36	0 ***
CS	700	0 ***
L2B	130	0 ***
FB	9.9	0 ***

Table 3 Test statistics and p-values for DANOVA.

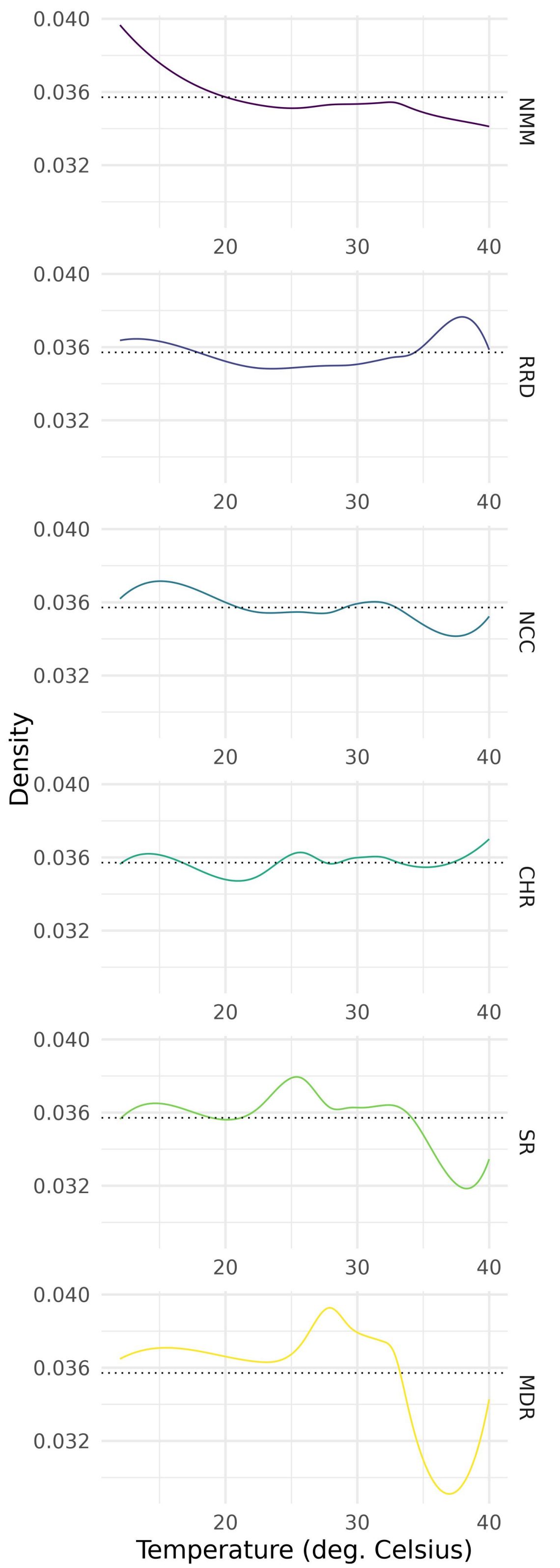

Fig. 5 Regional mean trend slopes. The dotted line represents the uniform density on $[12°\mathrm{C}, 40°\mathrm{C}]$ which is the reference measure.

4.3 Pairwise comparisons

When the test of equality of group mean densities is significant, conducting a post-hoc analysis of pairwise group mean comparisons can help identify which pairs differ. The null hypothesis for comparing groups g and g' is:

$$H_0: \quad \mathbb{E}^B(\pi_{gi}) = \mathbb{E}^B(\pi_{g'i}),$$

which is equivalent to

$$H_0: \quad \mathbb{E}(\operatorname{clr} \pi_{gi}) = \mathbb{E}(\operatorname{clr} \pi_{g'i}).$$

Testing whether the mean density in group g_1 is equal to the mean density in group g_2 can be viewed as a two-sample ANOVA test. A multiple testing correction is necessary when performing these tests for all pairs of groups.

Application: regional evolution of temperature distribution over the period 1987-2016

Table 4 displays the tests statistics with their Šidák-adjusted p-values for the five pairwise tests. We summarize and visualize these results on Figure 6 by drawing

1. a solid line between two regions when the equality of mean slope densities is rejected by all five tests
2. a dotted line between two regions when the equality of mean slope densities is rejected by some but not all of the five tests
3. no line between two regions when the equality of mean slope densities is rejected by all five tests.

The small number of dotted lines supports the fact that the five tests agree most of the time. The structure of Figure 6 is interestingly similar to the geographical structure of the regions (see the map on Figure 1) with solid lines between neighbouring regions. It shows that the Mekong Delta region differs statistically from most other regions in terms of climate change. Among the remaining regions, the Red River Delta region differs the most from the others in the sense that it is only connected by dotted lines.

5 Interval-wise interpretation

The global tests presented in the previous section, whether one-sample tests or analysis of variance, enable us to draw inferences about differences in the behavior of mean densities across the entire temperature range. Now, we aim to make more localized statements regarding regional differences in temporal evolution, particularly within specific temperature intervals.

Region 1	Region 2	L2B	FB	CS
NMM	RRD	1e-01 *	0.18	0 ***
NMM	NCC	1	1	1
NMM	CHR	0.97	0.99	0.96
NMM	SR	0.71	0.83	0.26
NMM	MDR	5.4e-09 ***	1.4e-06 ***	0 ***
RRD	NCC	2.5e-03 ***	8.6e-03 ***	0 ***
RRD	CHR	0.82	0.89	1
RRD	SR	8.7e-11 ***	3.9e-06 ***	0 ***
RRD	MDR	0 ***	0 ***	0 ***
NCC	CHR	0.99	1	0.99
NCC	SR	0.6	0.71	0.14
NCC	MDR	7.9e-13 ***	2.6e-09 ***	0 ***
CHR	SR	0.15	0.4	0.46
CHR	MDR	3.7e-11 ***	2.3e-07 ***	0 ***
SR	MDR	9e-04 ***	6.2e-03 ***	0 ***

Region 1	Region 2	GPF	Fmaxb
NMM	RRD	0.33	0.85
NMM	NCC	1	0.97
NMM	CHR	0.48	0.37
NMM	SR	3e-05 ***	0 ***
NMM	MDR	0 ***	0 ***
RRD	NCC	2.8e-02 **	0 ***
RRD	CHR	3.3e-02 **	0 ***
RRD	SR	0 ***	0 ***
RRD	MDR	0 ***	0 ***
NCC	CHR	0.98	0.96
NCC	SR	0.49	0.14
NCC	MDR	2e-13 ***	0 ***
CHR	SR	0.27	0.54
CHR	MDR	4.8e-11 ***	0 ***
SR	MDR	2.7e-06 ***	0.14

Table 4 Šidák-adjusted p-values for pairwise group mean comparison.

Unfortunately, the idea of adapting local tests from FANOVA does not work for two reasons. First of all, the meaning of a test of the equality of two mean densities evaluated at a given point x is unclear since densities are only defined almost everywhere, unless we impose continuity restrictions. The second reason is that the equivalence (28) between the equality of two mean densities and the equality of their clr transform is not valid anymore at the local level because the clr transform evaluated at x involves all values of the log density and not only its value at x.

For these reasons, we turn attention to some interpretation tools presented in Maier et al. (2024) based on odds ratios.

5.1 Infinitesimal odds ratios

We focus on subintervals $I \subset (a, b)$ and wish to assess the time evolution of the relative probability to be in one interval versus the other.

We can rely on (Maier et al., 2024, pp. 10 and 11, Proposition 3.1) for the definition of the infinitesimal odds ratios and their interpretation. The objective is to compare the value of a slope

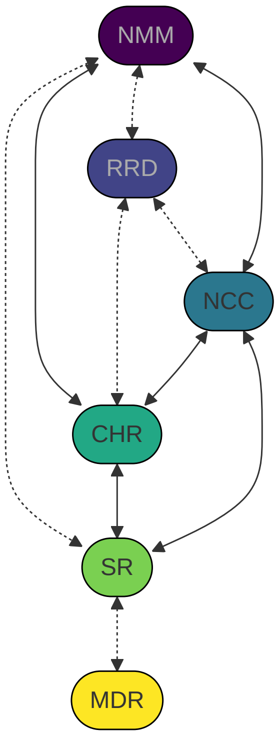

Fig. 6 Summary of pairwise tests between regions of Vietnam. Solid line: no test rejects the null hypothesis; dotted line: some tests reject the null hypothesis; absence of line: all tests reject the null hypothesis.

density β_{gi} and $\beta_{g.}$ (we call it β in this paragraph) at two different temperature values x and z in the temperature range. Let us define the relative frequency $\mathrm{OR}_{x|z}(\beta)$ of a density β at two points x and z by

$$\mathrm{OR}_{x|z}(\beta) = \frac{\beta(x)}{\beta(z)} \tag{36}$$

For $\beta = \beta_{gi}$ in model (20), let us show that this quantity is indeed an infinitesimal odds ratio relative to the trend densities $\tilde{\pi}_{it}(z)$. Indeed, by linearity of model (20), we have that for all t,

$$\frac{\beta_{gi}(x)}{\beta_{gi}(z)} = \frac{\tilde{\pi}^{t+1}_{gi}(x) \ominus \tilde{\pi}^{t}_{gi}(x)}{\tilde{\pi}^{t+1}_{gi}(z) \ominus \tilde{\pi}^{t}_{gi}(z)}$$

and therefore

$$\mathrm{OR}_{x|z}(\beta_{gi}) = \frac{\frac{\tilde{\pi}^{t+1}_{gi}(x)}{\tilde{\pi}^{t+1}_{gi}(z)}}{\frac{\tilde{\pi}^{t}_{gi}(x)}{\tilde{\pi}^{t}_{gi}(z)}} \tag{37}$$

is the ratio of the odds of x versus z at time $t+1$ by the odds of x versus z at time t. We thus see that the relative change of the odds of x versus z in the period $(t, t+1)$ is equal to $\frac{\beta(x)}{\beta(z)} - 1$. Note that this relative change formula is only valid within our simple linear trend model.

According to part (a) of Proposition 3.1 in Maier et al. (2024), if we observe that $\mathrm{OR}_{x|z}(\beta_{gi}) > 1$ for all x, z when x is in a given interval A and z in a given interval B for a given slope density β_{gi}, then we may conclude that conditional on the temperature being in A or B, the odds (according to the density $\tilde{\pi}^{t+1}_{gi}$) of being in A at time $t+1$ are larger than the odds (according to the trend density $\tilde{\pi}^{t}_{gi}$) of being in A at time t. Because the probability of an event is an increasing function of its odds, same is true for the corresponding probabilities so that there has been a mass transfer of the probability mass of the trend density from B to A between t and $t+1$.

Using the fact that $\mathrm{OR}_{x|z}(\beta_{gi}) > 1$ is equivalent to $\beta_{gi}(x) > \beta_{gi}(z)$, we are going to show that the curve of β_{gi} allows to draw conclusions about the temperature density changes as follows. For a given level $\tau = \beta_{gi}(x_0) > 0$, let the collection of intervals (or unions of intervals) $A_\tau(\beta_{gi})$ be defined by $A_\tau(\beta_{gi}) = \{x : \beta_{gi}(x) > \tau\}$. Let $A^c_\tau(\beta_{gi})$ be the complement of $A_\tau(\beta_{gi})$. Then we have for almost all $x \in A_\tau(\beta_{gi})$ and almost all $z \in A^c_\tau(\beta_{gi})$

$$\mathrm{OR}_{x|z}(\beta_{gi}) = \frac{\beta(x)}{\beta(x_0)} \frac{\beta(x_0)}{\beta(z)} > 1, \tag{38}$$

Therefore, we may say that the probability that it lies in $A_\tau(\beta_{gi})$ according to the trend density at time $t+1$ (i.e. under the distribution $\tilde{\pi}^{t+1}$) is higher than that according to the trend density at time t (i.e. under the distribution $\tilde{\pi}^{t}$). We will consider various values of the level τ increasing from the minimum to the maximum of the slope density and comment the corresponding interpretations.

5.2 Global change in temperature distribution over the period 1987-2016

Next we apply these interpretations to the global slope density $\beta_{\cdot\cdot}$. From (37) we can derive that

$$\mathrm{OR}_{x|z}(\bar{\beta}_{\cdot\cdot}) = \frac{\tilde{\pi}^{t+1}_{\cdot\cdot}(x) \ominus \tilde{\pi}^{t}_{\cdot\cdot}(x)}{\tilde{\pi}^{t+1}_{\cdot\cdot}(z) \ominus \tilde{\pi}^{t}_{\cdot\cdot}(z)} = \frac{\frac{\tilde{\pi}^{t+1}_{\cdot\cdot}(x)}{\tilde{\pi}^{t+1}_{\cdot\cdot}(z)}}{\frac{\tilde{\pi}^{t}_{\cdot\cdot}(x)}{\tilde{\pi}^{t}_{\cdot\cdot}(z)}} \tag{39}$$

On Figure 7, using a first level of 0.0362 represented by a horizontal line, we see that on average in Vietnam low temperatures below 18.2°C become relatively more frequent than temperatures above 18.2°C. Similarly, with a second level of 0.0355, we see that on average in Vietnam high

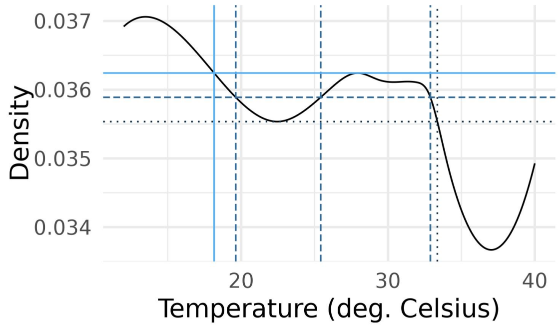

Fig. 7 Global mean slope density with three levels.

temperatures above 33.4°C become relatively more frequent than temperatures below 33.4°C. However these two statements need to be put into perspective due to the scarcity of observations at the extremes. More importantly, focusing now on the range $[19.6°\text{C}, 32.9°\text{C}]$ and the level 0.0359, we can say that the frequency of temperatures within the interval $[19.6°\text{C}, 25.4°\text{C}]$ has decreased over time relative to those in $[25.4°\text{C}, 32.9°\text{C}]$, supporting a global warming trend. Technically, this assertion corresponds to the interpretation of a conditional probability statement.

5.3 Regional changes in temperature distribution over the period 1987-2016

Similarly, when comparing the regional slope density $\beta_{g.}$ at two temperature points x and z, we use $\text{OR}_{x|z}(\beta_{g.})$. From (37) we can derive that

$$\text{OR}_{x|z}(\bar{\beta}_{g.}) = \frac{\tilde{\pi}^{t+1}_{g.}(x) \ominus \tilde{\pi}^{t}_{g.}(x)}{\tilde{\pi}^{t+1}_{g.}(z) \ominus \tilde{\pi}^{t}_{g.}(z)} = \frac{\frac{\tilde{\pi}^{t+1}_{g.}(x)}{\tilde{\pi}^{t+1}_{g.}(z)}}{\frac{\tilde{\pi}^{t}_{g.}(x)}{\tilde{\pi}^{t}_{g.}(z)}} \quad (40)$$

Figures 8, 9 and 10 display the slope densities of the six regions with some chosen values of level τ. We are able to group the regions in terms of the shape of their mean slope density.

Based on Figure 8, in mountainous regions, as for the global trend, the small and high values of τ demonstrate an increase of the frequency of low temperatures. Provided we focus on the medium range ($[22.6°\text{C}, 27.6°\text{C}]$ for NMM and $[21°\text{C}, 29.1°\text{C}]$ for NCC), the intermediate values of τ show a shift towards higher temperatures.

Based on Figure 9, the RRD and CHR regions display an increasing spread of their temperature distribution over time (small values of τ) and an increase of extremely high temperature

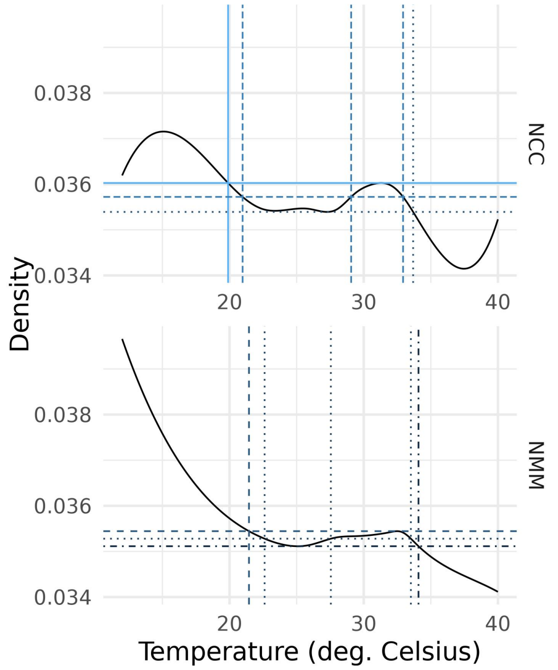

Fig. 8 Mean slope densities of North Central Coast (NCC) and Northern Midlands and Mountains (NMM) regions with levels.

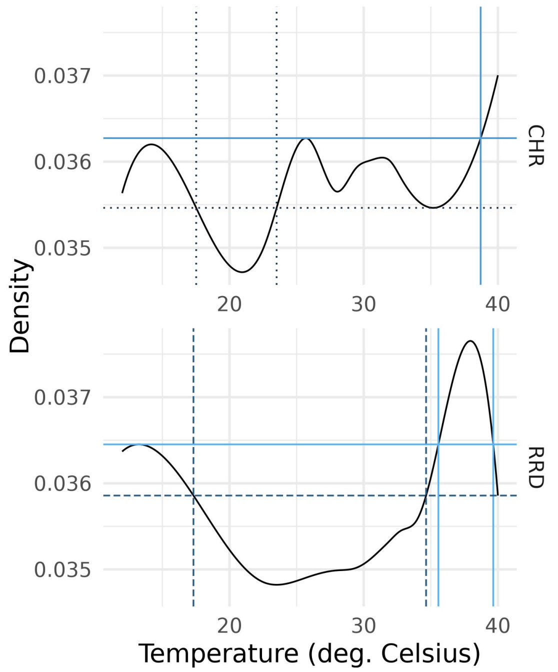

Fig. 9 Mean slope densities of Red River Delta (RRD) and Central Highlands (CHR) regions with levels.

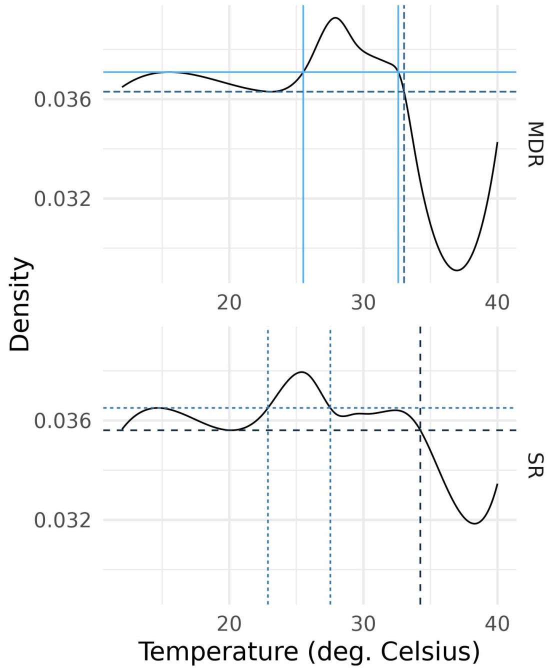

Fig. 10 Mean slope densities of Southeast (SR) and Mekong Delta (MDR) regions with levels.

events (high values of τ). Note that since temperature change is not detected in the CHR region (Table 3), its interpretation based on odds ratio is probably not reliable.

Based on Figure 10, the south in contrast shows a concentration in the medium range around 25°C for SR and 28°C for MDR at the expense of high temperatures.

While these groups reflect the geography (latitude and elevation) of Vietnam, they are also consistent with the groups suggested by the regional one-sample tests (Table 2) and the pairwise comparisons (Table 4).

6 Conclusion

We have adapted several functional data analysis tests to density functions in order to assess the equality of mean densities and to perform DANOVA tests in the framework of Bayes spaces. In our target application of temperature density evolution in Vietnam, the one-sample test allows to conclude that there is statistical evidence of a climate change in Vietnam in the sense that the trend slope density is not uniform. Furthermore, when considering the functions globally, most of the DANOVA statistics strongly reject the null hypothesis, that is, the equality of trend slope density across the Vietnamese regions. For more local interpretations, we rely on the infinitesimal odds ratio of Maier et al. (2025). These local investigations remain exploratory, and elevating them to formal tests presents an interesting avenue for future research. Needless to say that the proposed methodology can be applied in other application frameworks. Future research might try to reduce the uncertainty due to the low density of extreme temperatures, possibly weighting the domain of the Bayes space as in Talská et al. (2020). This methodology is promising for environmental studies, with a possible application for example to the relative concentrations of contaminants in grounds or rivers analyzed as density functions.

Acknowledgments

Part of this work was completed while the authors were visiting the Vietnam Institute for Advanced Study in Mathematics (VIASM) in Hanoi and the authors express their gratitude to VIASM. We also acknowledge funding from the French National Research Agency (ANR) under grant ANR-17-EURE-0010 (Investissements d'Avenir program), from the French National Association of Research and Technology (ANRT), grant CIFRE n°2020/0011, and from Spanish Ministerio de Ciencia e Innovación grant number PID2021-123833OB-I00.

Code & reproducibility

In order to implement DANOVA, we created the R package `ICSFun`, available at:

https://github.com/camillemndn/dda

This article and its figures are fully reproducible using the code available at:

https://github.com/camillemndn/danova

References

Abdi, H.: Holm's Sequential Bonferroni Procedure. In: Encyclopedia of Research Design, pp. 574–577. SAGE Publications, Inc., ??? (2010). https://doi.org/10.4135/9781412961288 . https://methods.sagepub.com/ency/edvol/encyc-of-research-design/chpt/holms-sequential-bonferroni-procedure Accessed 2025-06-24

Aneiros, G., Horová, I., Hušková, M., Vieu, P.: On functional data analysis and related topics. Journal of Multivariate Analysis **189**, 104861 (2022) https://doi.org/10.1016/j.jmva.2021.104861

Cuevas, A., Febrero, M., Fraiman, R.: An anova test for functional data. Computational Statistics & Data Analysis **47**(1), 111–122 (2004) https://doi.org/10.1016/j.csda.2003.10.021

Espagne, E., Diallo, Y., Marchand, S.: Impacts of Extreme Climate Events on Technical Efficiency in Vietnamese Agriculture. Number: c1221ee7-5311-4af0-b1b4-3972fc637a9f Publisher: Agence française de développement (2019). https://ideas.repec.org//p/avg/wpaper/en9476.html Accessed 2024-05-20

Egozcue, J.J., Díaz–Barrero, J.L., Pawlowsky–Glahn, V.: Hilbert Space of Probability Density Functions Based on Aitchison Geometry. Acta Mathematica Sinica, English Series **22**(4), 1175–1182 (2006) https://doi.org/10.1007/s10114-005-0678-2

Gubler, S., Fukutome, S., Scherrer, S.C.: On the statistical distribution of temperature and the classification of extreme events considering season and climate change—an application in Switzerland. Theoretical and Applied Climatology **153**(3-4), 1273–1291 (2023) https://doi.org/10.1007/s00704-023-04530-0

Górecki, T., Smaga, L.: fdANOVA: Analysis of Variance for Univariate and Multivariate Functional Data (2018). https://cran.r-project.org/web/packages/fdANOVA/index.html Accessed 2024-03-14

Górecki, T., Smaga, L.: fdANOVA: an R software package for analysis of variance for univariate and multivariate functional data. Computational Statistics **34**(2), 571–597 (2019) https://doi.org/10.1007/s00180-018-0842-7

Hultgren, A., Carleton, T., Delgado, M., Gergel, D.R., Greenstone, M., Houser, T., Hsiang, S., Jina, A., Kopp, R.E., Malevich, S.B., McCusker, K., Mayer, T., Nath, I., Rising, J., Rode, A., Yuan, J.: Estimating Global Impacts to Agriculture from Climate Change Accounting for Adaptation. SSRN Electronic Journal (2022) https://doi.org/10.2139/ssrn.4222020

Hsiang, S., Kopp, R., Jina, A., Rising, J., Delgado, M., Mohan, S., Rasmussen, D.J., Muir-Wood, R., Wilson, P., Oppenheimer, M., Larsen, K., Houser, T.: Estimating economic damage from climate change in the United States. Science **356**(6345), 1362–1369 (2017) https://doi.org/10.1126/science.aal4369

Kokoszka, P., Reimherr, M.: Introduction to Functional Data Analysis, 1st edn. Chapman and Hall/CRC, ??? (2017). https://doi.org/10.1201/9781315117416

Martín-Fernández, J.-A., Daunis-i-Estadella, J., Mateu-Figueras, G.: On the interpretation of differences between groups for compositional data. SORT-Statistics and Operations Research Transactions **39**(2), 231–252 (2015). Number: 2. Accessed 2024-03-11

Machalová, J., Hron, K., Monti, G.S.: Preprocessing of centred logratio transformed density functions using smoothing splines. Journal of Applied Statistics **43**(8), 1419–1435 (2016) https://doi.org/10.1080/02664763.2015.1103706

Maier, E.-M., Stöcker, A., Fitzenberger, B., Greven, S.: Additive Density-on-Scalar Regression in Bayes Hilbert Spaces with an Application to Gender Economics. arXiv. arXiv:2110.11771 [stat] (2024). https://doi.org/10.48550/arXiv.2110.11771 . http://arxiv.org/abs/2110.11771 Accessed 2025-02-25

Maier, E.-M., Stöcker, A., Fitzenberger, B., Greven, S.: Additive density-on-scalar regression in Bayes Hilbert spaces with an application to gender economics. The Annals of Applied Statistics **19**(1), 680–700 (2025) https://doi.org/10.1214/24-AOAS1979 . Publisher: Institute of Mathematical Statistics. Accessed 2025-03-18

Machalová, J., Talská, R., Hron, K., Gába, A.: Compositional splines for representation of density functions. Computational Statistics **36**(2), 1031–1064 (2021) https://doi.org/10.1007/s00180-020-01042-7

Petersen, A., Zhang, C., Kokoszka, P.: Modeling Probability Density Functions as Data Objects. Econometrics and Statistics **21**, 159–178 (2022) https://doi.org/10.1016/j.ecosta.2021.04.004

Ramsay, J.O., Silverman, B.W.: Functional Data Analysis. Springer Series in Statistics. Springer, New York, NY (2005). https://doi.org/10.1007/b98888

Roberts, M.J., Schlenker, W., Eyer, J.: Agronomic Weather Measures in Econometric Models of Crop Yield with Implications for Climate Change. American Journal of Agricultural Economics **95**(2), 236–243 (2013) https://doi.org/

10.1093/ajae/aas047

Schumaker, L.: Spline Functions: Basic Theory, 3rd edn. Cambridge University Press, ??? (1981). https://doi.org/10.1017/CBO9780511618994

Shen, Q., Faraway, J.: An F Test for Linear Models with Functional Responses. Statistica Sinica **14**(4), 1239–1257 (2004). Publisher: Institute of Statistical Science, Academia Sinica. Accessed 2025-02-11

Schlenker, W., Roberts, M.J.: Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. Proceedings of the National Academy of Sciences **106**(37), 15594–15598 (2009) https://doi.org/10.1073/pnas.0906865106

Trinh, T.-A., Feeny, S., Posso, A.: The Impact of Natural Disasters and Climate Change on Agriculture: Findings From Vietnam. In: Economic Effects of Natural Disasters, pp. 261–280. Elsevier, ??? (2021). https://doi.org/10.1016/B978-0-12-817465-4.00017-0

Talská, R., Menafoglio, A., Hron, K., Egozcue, J.J., Palarea-Albaladejo, J.: Weighting the domain of probability densities in functional data analysis. Stat **9**(1), 283 (2020) https://doi.org/10.1002/sta4.283 . Accessed 2024-01-06

Talská, R., Menafoglio, A., Machalová, J., Hron, K., Fišerová, E.: Compositional regression with functional response. Computational Statistics & Data Analysis **123**, 66–85 (2018) https://doi.org/10.1016/j.csda.2018.01.018 . Accessed 2025-03-05

Tran, T.D., Nguyen, T.V.: Climate Change Adaptation in Vietnam." In Climate Change Adaptation in Southeast Asia. Singapore: Springer Singapore, ??? (2021)

Trinh, T.A.: The Impact of Climate Change on Agriculture: Findings from Households in Vietnam. Environmental and Resource Economics **71**(4), 897–921 (2018) https://doi.org/10.1007/s10640-017-0189-5

Trinh, H.T., Thomas-Agnan, C., Simioni, M.: Discrete and Smooth Scalar-on-Density Compositional Regression for Assessing the Impact of Climate Change on Rice Yield in Vietnam. TSE Working Paper **23-1410** (2023)

Van Den Boogaart, K.G., Egozcue, J.J., Pawlowsky-Glahn, V.: Bayes linear spaces. SORT : statistics and operations research transactions **34**, 201–222 (2010)

Van Den Boogaart, K.G., Egozcue, J.J., Pawlowsky-Glahn, V.: Bayes Hilbert spaces. Australian & New Zealand Journal of Statistics **56**(2), 171–194 (2014) https://doi.org/10.1111/anzs.12074

Vo, T.H., Nguyen, V.T., Simioni, M.: Adaptation of Rice Yields to Heat Stress in Vietnam: New Insight from Heterogeneous Slopes Panel Data Modeling. (2022). https://hal.inrae.fr/hal-03715498 Accessed 2024-09-11

World Bank Group, Asian Development Bank: Climate Risk Country Profile: Vietnam. World Bank, ??? (2021). https://doi.org/10.1596/36367

World Meteorological Organization: The Global Climate 2011-2020: A decade of acceleration. Technical report, World Meteorological Organization (November 2023). https://wmo.int/publication-series/global-climate-2011-2020-decade-of-acceleration Accessed 2024-04-12

World Meteorological Organization: State of the Global Climate 2023. Technical Report WMO-No. 1347, World Meteorological Organization (March 2024). https://wmo.int/publication-series/state-of-global-climate-2023 Accessed 2024-04-12

Zhang, J.-T., Chen, J.: Statistical inferences for functional data. The Annals of Statistics **35**(3), 1052–1079 (2007) https://doi.org/10.1214/009053606000001505 . Publisher: Institute of Mathematical Statistics. Accessed 2025-02-11

Zhang, J.-T., Cheng, M.-Y., Wu, H.-T., Zhou, B.: A new test for functional one-way ANOVA with applications to ischemic heart screening. Computational Statistics & Data Analysis **132**, 3–17 (2019) https://doi.org/10.1016/j.csda.2018.05.004

Zhang, J.-T.: Statistical inferences for linear models with functional responses. Statistica Sinica **21**(3), 1431–1451 (2011) https://doi.org/10.5705/ss.2009.302 . Accessed 2024-11-19

Zhang, J.-T.: Analysis of Variance for Functional

Data, 0 edn. Chapman and Hall/CRC, ??? (2013). https://doi.org/10.1201/b15005

Zhang, J.-T., Liang, X.: One-Way ANOVA for Functional Data via Globalizing the Pointwise F-test. Scandinavian Journal of Statistics **41**(1), 51–71 (2014) https://doi.org/10.1111/sjos.12025