



**MACQUARIE**  
University  
SYDNEY · AUSTRALIA

## Macquarie University PURE Research Management System

---

This is a post-peer-review, pre-copyedit version of an article published in  
Journal of Population Research, volume 38, issue 1, pages 1-24 (2021)

The final authenticated version is available online at:

<https://doi.org/10.1007/s12546-020-09250-0>

# Forecasting Australian subnational age-specific mortality rates

Han Lin Shang\*

Department of Actuarial Studies and Business Analytics  
Macquarie University

Yang Yang

Research School of Finance, Actuarial Studies and Statistics  
Australian National University

## Abstract

When modeling sub-national mortality rates, it is important to incorporate any possible correlation among sub-populations to improve forecast accuracy. Moreover, forecasts at the sub-national level should aggregate consistently across the forecasts at the national level. In this study, we apply a grouped multivariate functional time series to forecast Australian regional and remote age-specific mortality rates and reconcile forecasts in a group structure using various methods. Our proposed method compares favorably to a grouped univariate functional time series forecasting method by comparing one-step-ahead to five-step-ahead point forecast accuracy. Thus, we demonstrate that joint modeling of sub-populations with similar mortality patterns can improve point forecast accuracy.

*Keywords:* Multivariate functional principal component analysis; Hierarchical/grouped time series; Forecast reconciliation; Australian regional mortality rates

---

\*Postal address: Department of Actuarial Studies and Business Analytics, Level 7, 4 Eastern Rd, Macquarie University, Sydney, NSW 2109, Australia; Email: hanlin.shang@mq.edu.au; ORCID: <https://orcid.org/0000-0003-1769-6430>.

# 1 Introduction

Mortality is regarded as one of the most widely available measures of health conditions in a community. National and sub-national mortality patterns in terms of age, sex, and geographical distribution are of interest to epidemiologists, health care personnel, and those working in health and social policy, planning, and administration. Traditional studies of human mortality focus on population data at the national level. In recent years, academics and government stakeholders have been showing increasing interest in regional mortality improvements after realizing that sub-national forecasts of age-specific mortality rates are useful for informing social and economic policies within local regions. Thus, any improvement in the forecast accuracy of regional mortality rates would help determine the allocation of current and future resources at the national and sub-national levels. Many mortality modeling methods have been proposed since the publication of the Gompertz law in 1825 (see [Booth and Tickle, 2008](#), for a comprehensive literature review on mortality modeling and forecasting), but only a few approaches can forecast multiple regional mortality rates within a country simultaneously. Thus, proposing a multivariate functional time series method for producing accurate and coherent sub-national age-specific mortality forecasts is the primary motivation of our study.

Functional time series can arise by separating an almost continuous time record into consecutive intervals, such as the intraday trajectories of the S&P 500 index ([Shang, 2017](#)) and the monthly sea surface temperature in climatology ([Shang and Hyndman, 2011](#)). Alternatively, functional time series can also arise when the observations in a period can be considered together as finite realizations of an underlying continuous function, for example, age-specific mortality rates in demography (see [Hyndman and Ullah, 2007](#); [Chiou and Müller, 2009](#)). In either case, the functions obtained form a time series  $\{f_t, t \in \mathbb{Z}\}$ , where each  $f_t$  is a random function of a stochastic process and  $x \in \mathcal{I}$  represents a continuum bounded within a finite interval. In this study, we consider age-specific mortality rates observed annually as a functional time series whose continuum is the age of the population.

The increasing popularity of the functional time series has resulted in a rapid increase in the literature on functional time series modeling and forecasting. From a parametric perspective, [Bosq \(2000\)](#) and [Bosq and Blanke \(2007\)](#) proposed the functional autoregressive of order 1 (FAR(1)). They derived one-step-ahead forecasts that are based on a regularized form of the Yule-Walker equations. Later, FAR(1) was extended to FAR( $p$ ), under which the order  $p$  can be determined via [Kokoszka and Reimherr's \(2013\)](#) hypothesis testing procedure. To overcome the difficulty of infinite-dimensional parameter estimation, [Aue et al. \(2015\)](#) showed the asymptotic equivalence between a FAR and a vector autoregressive (VAR) model via a functional principal component analysis (FPCA) and proposed a forecasting method based on the VAR forecasts of principal component

scores. The approach of [Aue et al. \(2015\)](#) can also be viewed as an extension of [Hyndman and Shang \(2009\)](#). [Hyndman and Shang \(2009\)](#) forecasts principal component scores by a univariate time series forecasting method. Recently, [Klepsch and Klüppelberg \(2017\)](#) proposed the functional moving average (FMA) process and introduced an innovative algorithm to obtain the best linear predictor. [Klepsch et al. \(2017\)](#) extended the VAR model to the vector autoregressive moving average (VARMA) model. The VARMA model is a simpler estimation approach of the functional autoregressive moving average model. Recently, [Li et al. \(2020\)](#) considered long-range dependent curve time series and proposed a functional ARIMA model. From a nonparametric perspective, [Besse et al. \(2000\)](#) proposed the functional kernel regression method to model temporal dependence via a similarity measure characterized by semi-metric, bandwidth, and kernel functions. From a semi-parametric viewpoint, [Aneiros-Pérez and Vieu \(2008\)](#) put forward a semi-functional partial linear model that combines parametric and nonparametric models. The semi-functional partial linear model allows us to consider additive covariates and to use a continuous path in the past to predict a stochastic process's future values. Apart from the estimation of a conditional mean, [Hörmann et al. \(2013\)](#) consider a functional analog of the autoregressive conditional heteroskedasticity model for modeling conditional variance. In contrast, [Aue et al. \(2017\)](#) consider a functional analog of the generalized autoregressive conditional heteroskedasticity model. [Kokoszka et al. \(2017\)](#) considered a portmanteau test for testing autocorrelation under a functional generalized autoregressive conditional heteroskedasticity model.

In the functional time series literature, limited work has been conducted on jointly modeling and forecasting multiple sub-national mortality rates over some time. [Shang \(2016\)](#) attempted to simultaneously model and forecast state-specific mortality in Australia by determining a common trend across populations and sex-specific trends of each state. Another noticeable exception is the work by [Gao et al. \(2019\)](#) in which an additive factor model is fitted to functional principal component scores obtained in the initial dimension reduction of smoothed sub-national mortality data. However, both approaches mentioned above fail to explicitly consider significant correlations between sub-national populations in basis function decomposition and forecasting and thus lose valuable information that can be exploited to improve forecasting accuracy. Moreover, applying existing functional time series forecasting methods to sub-national mortality disaggregated by attributes such as sex, state, or ethnicity independently does not ensure coherence in the forecasts. That is, the forecasts of all subpopulations will not add up to the forecasts obtained by applying the technique to the national data. Hence, in practice, it is important to consider reconciliation approaches (see, e.g., [Stone et al., 1942](#); [Sefton and Weale, 2009](#)) in forecasting sub-national mortality rates.

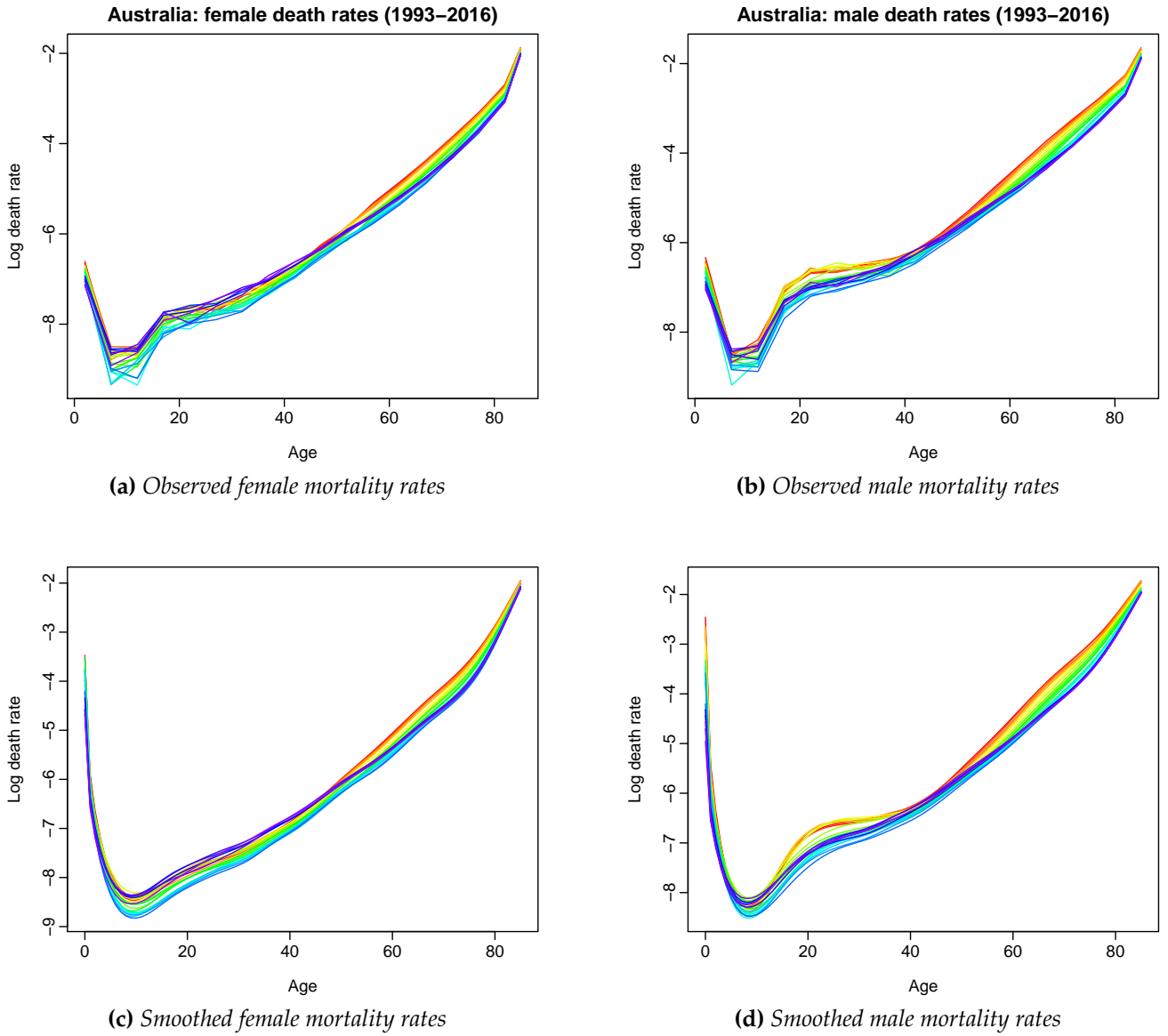
In this study, we apply the univariate and multivariate functional time series forecasting methods to the Australian national and sub-national age-specific mortality rates from 1993 to 2016 and compare their point forecast accuracies. To ensure forecasts of disaggregated series consistently aggregating to the national totals, we consider a group structure for mortality series. We apply the bottom-up, optimal combination, trace minimization reconciliation methods, and their forecast combination to reconcile forecasts. Evaluated by the mean absolute forecast error and the root mean square forecast error, we find that the trace minimization and forecast combination methods have similar superior point forecast accuracy. Our empirical study results prove that jointly modeling of populations with similar mortality patterns can improve point forecast accuracy.

The rest of this report is structured as follows. In Section 2, we describe the motivating dataset, which is the Australian national and sub-national age-specific mortality rates. In Section 3, we introduce the multivariate functional principal component regression for producing point forecasts. Forecast reconciliation methods for the grouped functional time series are discussed in Section 4. Reconciled point forecasts produced by univariate and multivariate forecasting methods are evaluated and compared in Section 5. Conclusions are presented in Section 6.

## 2 Australian age-specific mortality rates

### 2.1 Overview of data

Australia, as a developed country, faces increasing longevity risks and challenges to the sustainability of the pension, health, and aged care systems caused by an aging population(see, e.g., [Coulmas, 2007](#); [OECD, 2013](#)). To better understand the ongoing mortality improvements, we studied the Australian age-specific mortality rates in 1993-2016. We consider ages from 0 to 85 years in 5-year age groups and include all ages at and beyond 85 years in the last age group, thus considering 18 groups. Figure 1 presents rainbow plots of the age-specific mortality rates of females and males in the period studied. The “J” shaped death curves shown in the figure indicate typical mortality patterns that rapidly decrease in infant ages, and a steady increase in middle-to-old age groups. There is also a noticeable “accident hump” during the teenage years for the Australian male population. The rainbow plot is designed to display the distant past curves in red and the more recent curves in purple ([Hyndman and Shang, 2010](#)), indicating an overall reduction in the trend of total mortality rates for Australian people. However, death rates for males and females aged less than 50 have a sudden increase in 2012, and then gradually fall off with time, as shown in the bottom panels in Figure 1.



**Figure 1:** Functional time series graphical displays. The nonparametric smoothing method in (2.2) can be used to smooth age-specific mortality rates.

## 2.2 Nonparametric smoothing

The observed mortality rates contaminated by noise are shown in the top panel of Figure 1. To obtain smooth functions and deal with possibly missing values, we consider a penalized regression spline smoothing with monotonic constraint while preserving the shape of the log mortality curves (see also Hyndman and Ullah, 2007; Shang, 2016). For a group of populations, let  $y_t^j(x_i)$  be the log central mortality rates observed at the beginning of each year for year  $t = 1, 2, \dots, n$  at observed ages  $x_1, x_2, \dots, x_p$  where  $x$  denotes the age variable,  $p$  denotes the number of ages considered, and superscript  $j$  represents an individual population. The value of  $j$  depends on the number of populations considered. For example,  $j$  is binary when representing age-specific mortality rates of females and males for a particular area (at the national or sub-national level);  $j = 1, 2, \dots, 11$  if *Region* level (described in Section 2.3) total mortality rates are considered.

We assume that there is an underlying continuous and smooth function  $f_t^j(x)$  such that

$$y_t^j(x_i) = f_t^j(x_i) + \delta_t^j(x_i)\varepsilon_{t,i}^j,$$

where  $x_i$  represents the center of each age or age group,  $\{\varepsilon_{t,i}^j\}$  are independent and identically distributed random variables across  $t$  and  $i$  with a mean of zero and a unit variance, and  $\delta_t^j(x_i)$  measures the variability in mortality at each age in year  $t$  for the  $j^{\text{th}}$  population. Jointly,  $\delta_t^j(x_i)\varepsilon_{t,i}^j$  represents the smoothing error.

Let  $m_t^j(x_i) = \exp[y_t^j(x_i)]$  be the observed central mortality rates for age  $x_i$  in year  $t$ . Define  $N_t^j(x_i)$  to be the total  $j^{\text{th}}$  population of age  $x_i$  at 1<sup>st</sup> January of year  $t$ . The observed log mortality rate approximately follows a Poisson distribution with estimated variance

$$\left(\widehat{\delta}_t^j\right)^2(x_i) := \text{Var} \left\{ \ln[m_t^j(x_i)] \right\} \approx \frac{1}{m_t^j(x_i) \times N_t^j(x_i)}. \quad (1)$$

In line with an earlier study by [Hyndman and Ullah \(2007\)](#), we smooth log mortality rates using weighted penalized regression splines with a partial monotonic constraint for ages above 65 years, where the weights are equal to the inverse variance given in (1), that is,  $w_t^j(x_i) = 1/(\widehat{\delta}_t^j)^2(x_i)$ . The penalized regression spline can be written as

$$\widehat{f}_t^j(x_i) = \underset{\theta_t(x_i)}{\text{argmin}} \sum_{i=1}^M w_t^j(x_i) \left| y_t^j(x_i) - \theta_t(x_i) \right| + \tau \sum_{i=1}^{M-1} \left| \theta'_t(x_{i+1}) - \theta'_t(x_i) \right|,$$

where  $i$  represents different ages (grid points) in a total of  $M$  grid points,  $\theta'$  denotes the first derivative of smooth function  $\theta$ , and  $\tau$  is a smoothing parameter. Both  $\theta'$  and  $\theta$  can be approximated by a set of  $B$ -spline basis (see, e.g., [De Boor, 2001](#)). While the  $L_1$  loss function and the  $L_1$  roughness penalty are employed to obtain robust estimates, the monotonic increasing constraint helps to reduce the noise from the estimation of older ages (see also [He and Ng, 1999](#)).

### 2.3 Multivariate functional time series of mortality rates

The vast land area of Australia can be divided into six states: New South Wales (NSW), Victoria (VIC), Queensland (QLD), South Australia (SA), Western Australia (WA), and Tasmania (TAS), as well as two internal territories, namely, the Northern Territory (NA) and the Australian Capital Territory (NT) (ACT). Based on the remoteness, areas with similar characteristics of eight considered states and territories are often grouped into geographic classifications. We follow [Guan \(2018\)](#) and classify sub-national population statistics sourced from the Australian Bureau of Statistics (ABS) into 47

harmonized geographic areas based on the Statistical Division level of ABS. In this study, we adopt a structure similar to the PRMA classification (DPIE and DSHS, 1994). As shown in Table 1, we classify 47 areas into Capital City<sup>1</sup> and four broader regions, i.e., NSW Coast, Country Victoria, Regional Australia, and Remote Australia.

**Table 1:** *Remoteness classification in Australia.*

Capital City		Regional Australia		Remote Australia	
Area	Name	Area	Name	Area	Name
1	Sydney	16	North West NSW	35	West NSW
5	Melbourne	17	Central West NSW	36	South West Queensland
11	Brisbane	18	Murrumbidgee	37	Central West Queensland
12	Adelaide	19	Murray	38	Far North
13	Perth	20	Wimmera	39	North West
14	Greater Hobart	21	Mallee	40	Eyre
15	Canberra	22	Ovens-Murray	41	Northern SA
<b>NSW Coast</b>		23	Wide Bay-Burnett & Fitzroy	42	South Eastern WA
2	Hunter	24	Darling Downs	43	Central WA
3	Illawarra	25	Mackay & Northern	44	Pilbara
4	Mid-North Coast	26	Yorke & Lower North	45	Kimberley
<b>Country Victoria</b>		27	Murray Lands	46	Mersey-Lyell
6	Barwon	28	South East	47	Northern Territory
7	Western District	29	South West		
8	Central Highlands	30	Lower Southern WA		
9	Loddon & Goulbourn	31	Upper Southern WA		
10	Gippsland	32	Midlands		
		33	Northern Tasmania		
		34	Darwin		

To form an appropriate disaggregation structure for areas listed in Table 1, we collect seven capital cities and four broader regions into a group and name it by “*Region*”. Thus, we have a three-level hierarchy, from top to bottom, consisting of the country total (Australia), 11 *Regions*, and 47 *Areas*, respectively. Together with sex in each area, there are in total 177 series across three levels of disaggregation listed in Table 2.

States and territories of Australia diversify significantly in area size, geographical features, and population. Most of Australia’s population is concentrated in coastal (within 80 km from the coast wherever possible) regions, and the population within these regions is concentrated in capital cities in particular. In June 2010, around 14.3 million people, or approximately 64% of Australia’s population, lived in capital cities (including Darwin) (ABS, 2012). Of the remainder population, over 6 million people (about one-third of the population) live in what referred to as regional and

<sup>1</sup>The capital city of NT, Darwin, is grouped into Regional Australia.



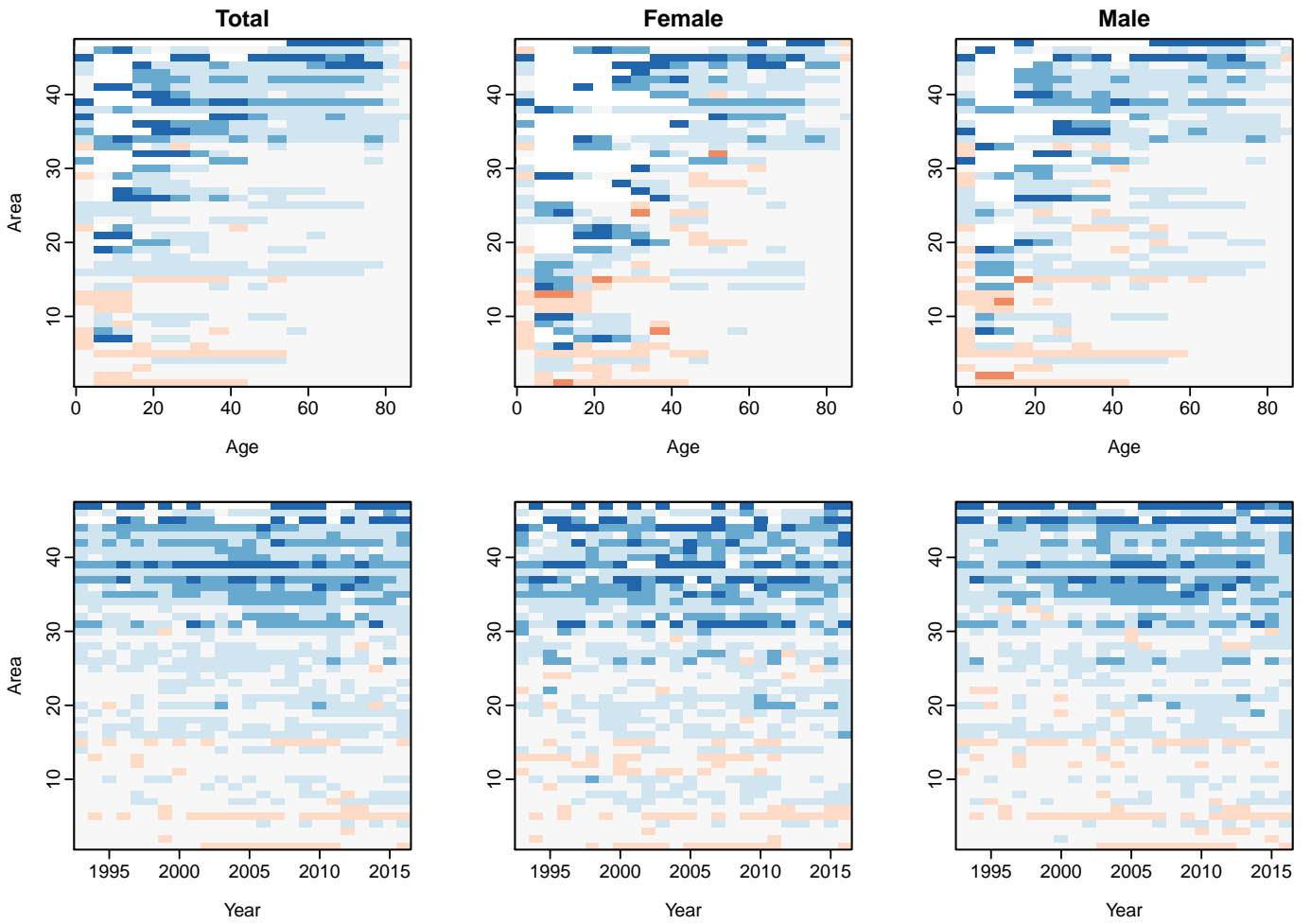
**Table 2:** *Hierarchy of Australian mortality rates.*

Level	Number of series
Australia	1
Sex	2
Region	11
Sex $\times$ Region	22
Area	47
Sex $\times$ Area	94
All	177

remote areas that are located with some distance away from major population centers and have extremely diverse geographical features (AIHW, 2003). Mortality rates for populations living in the metropolitan, regional, and remote areas are expected to be influenced by different geography, climate, and environmental conditions. For instance, AIHW (2007) reported mortality rates in remote areas that have very little accessibility of goods, services, and opportunities for social interaction went up to 1.76 times in capital cities. Another reason for inconsistent death rates among metropolitan and non-metropolitan populations is that many of the occupations in regional and remote areas involve higher levels of risk than other occupations (AIHW, 1998). For example, many regional and remote areas are primarily in the following high-risk industries: mining, transport, forestry, commercial fishing, and farming.

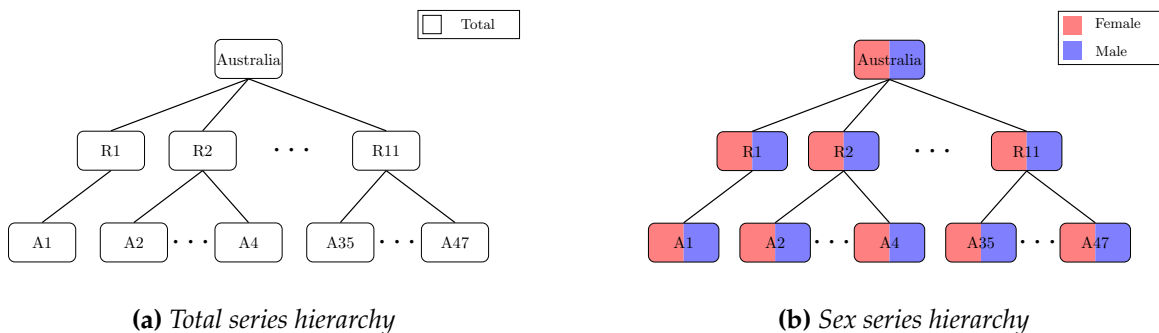
To visualize the differences between metropolitan and non-metropolitan mortality patterns, we divide the mortality rates of each area to death rates for Australia and then take the natural logarithm of the ratio. In Figure 2, positive log-ratios are represented in blue color, whereas negative values are represented in orange color, with the vertical axis indicating area numbers shown in Table 1. In the top row, the national mortality rates averaged over the period during 1993-2016 show that people living in capital cities have lower mortality rates than the country average, especially during the teenage years. In contrast, regional and remote areas, in general, have higher mortality rates for middle-to-old populations. The disadvantage of regional and remote areas regarding mortality is that people living in these areas have lower access to health services (AIHW, 1998). In the bottom row, the mortality rates for each area and year, averaged over all ages, highlight a consistent trend that metropolitan populations enjoy lower mortality rates than non-metropolitan populations from 1993 to 2016.

Figure 2 also shows that related populations tend to have similar mortality features. On the one hand, each of three categories, Capital City, Regional Australia, and Remote Australia, has specific mortality patterns explainable by living conditions of areas in that remoteness classification. On the other hand, females and males in each area across the country appear to show consistent mortality



**Figure 2:** Image plots showing log ratios of mortality rates. The top panel shows mortality rates averaged over the years, while the bottom panel shows mortality rates averaged over the ages.

movements during 1993-2016. Considering the distinct mortality patterns of metropolitan and non-metropolitan populations, we jointly model total death rates of areas within each of 11 *Regions* shown in Figure 3a, before collectively modeling total series for  $R_1, \dots, R_{11}$ . Due to the close correlation of female and male populations in each Region and Area, we jointly consider sex series (colored by red and blue) in every node of Figure 3b. Tree diagrams illustrating possible disaggregation of Australian national and sub-national mortality rates are shown in Figure 3. In Section 3, we will introduce the methodology of jointly modeling functional time series.



(a) Total series hierarchy

(b) Sex series hierarchy

**Figure 3:** Hierarchy tree diagrams for the Australian mortality rates.

### 3 Methodology

#### 3.1 Multivariate functional principal component analysis

The sub-national mortality rates of Australia can be organized into sets of multivariate functional time series based on the combinations of geographical disaggregation factors and sex. Following [Chiou et al. \(2014\)](#), we define  $\{f^{(l)}(x)\}_{l=1,\dots,\omega}$  as a set of smoothed sub-national mortality functions, with  $\omega$  representing the number of series under consideration. In our study,  $\omega \in \{1, 2, 11, 22, 47, 94\}$  as shown in [Table 2](#), and  $\omega = 1$  corresponds to the special case of univariate functional time series. For  $\omega \geq 2$ , define  $\mathbf{f}(x) = [f^{(1)}(x), \dots, f^{(\omega)}(x)]^\top$  as a vector in the Hilbert space. All elements of  $\mathbf{f}(x)$  in the study are square-integrable functions defined on the same interval of age  $\mathcal{I} = [0, 85]$ .

Let  $\mu^{(l)}(x) := E[f^{(l)}(x)]$  denote the mean function for the  $l^{\text{th}}$  subpopulation. For  $x, z \in \mathcal{I}$ , the covariance function of  $\mathbf{f}(x)$  has its  $(l, j)^{\text{th}}$  elements defined as follows, where  $1 \leq l, j \leq \omega$ :

$$C_{lj}(x, z) := E\{[f^{(l)}(x) - \mu^{(l)}(x)][f^{(j)}(z) - \mu^{(j)}(z)]\} = \text{Cov}[f^{(l)}(x), f^{(j)}(z)].$$

Let  $\mathbf{C}_l(x, z) = [C_{l1}(x, z), \dots, C_{l\omega}(x, z)]^\top$  and  $\mathbf{C}(x, z) = \{C_{lj}(x, z)\}$ . We define an integral operator  $\mathcal{A} : \mathbb{H} \rightarrow \mathbb{H}$  with the covariance kernel  $\mathbf{C}(x, z)$ , for any given  $\mathbf{f} \in \mathbb{H}$ , such that

$$(\mathcal{A}\mathbf{f})(x) = \int \mathbf{C}(x, z)\mathbf{f}(z)dz = \begin{pmatrix} \langle \mathbf{C}_1(x, z), \mathbf{f}(z) \rangle \\ \vdots \\ \langle \mathbf{C}_\omega(x, z), \mathbf{f}(z) \rangle \end{pmatrix},$$

where  $\langle \mathbf{C}_l(x, z), \mathbf{f}(z) \rangle = \sum_{j=1}^{\omega} \langle C_{lj}(x, z), f^{(j)}(z) \rangle$ . Since  $\mathcal{A}$  is linear, an eigenvalue  $\lambda$  and an eigenfunction  $\phi$  in  $\mathbb{H}$  satisfy  $(\mathcal{A}\phi)(x) = \lambda\phi(x)$ . Via Mercer's lemma, there exists orthonormal sequences  $\{\phi_k = [\phi_k^{(1)}, \dots, \phi_k^{(\omega)}]^\top\}_{k=1,2,\dots}$  of continuous functions in  $\mathbb{H}$ , and a non-increasing sequence  $\lambda_k$  of positive numbers, such that

$$\mathbf{C}(x, z) = \sum_{k=1}^{\infty} \lambda_k \phi_k(x) \phi_k(z),$$

with the  $(l, j)^{\text{th}}$  element  $C_{lj}(x, z)$  of  $\mathbf{C}(x, z)$

$$C_{lj}(x, z) = \sum_{k=1}^{\infty} \lambda_k \phi_k^{(l)}(x) \phi_k^{(j)}(z).$$

By the separability of Hilbert spaces, the Karhunen-Loève expansion of a stochastic process for the

$l^{\text{th}}$  subpopulation can be expressed as

$$f^{(l)}(x) = \mu^{(l)}(x) + \sum_{k=1}^{\infty} \beta_k^{(l)} \phi_k^{(l)}(x),$$

where  $\beta_k^{(l)}$  is the  $k^{\text{th}}$  principal component score given by the projection of  $f^{(l)}(x) - \mu^{(l)}(x)$  in the direction of eigenfunction  $\phi_k^{(l)}$ , that is,  $\beta_k^{(l)} = \langle f^{(l)}(x) - \mu^{(l)}(x), \phi_k^{(l)}(x) \rangle$ .

The proposed multivariate functional principal component analysis (MFPCA) as an extension of the widely used FPCA approach (for theoretical, methodological, and applied aspects of FPCA, see, e.g., [Hall, 2011](#); [Shang, 2014](#); [Wang et al., 2016](#)), explicitly incorporate cross-covariance between subpopulation curves. We use this dimension reduction technique to effectively summarize the main features of each infinite-dimensional curves considered by its finite key elements, and form a base of functional principal component regression.

### 3.2 Forecasting via functional principal component regression

A time series of smoothed functions corresponding to multiple subpopulations  $\{f_1(x), \dots, f_n(x)\}$  can be decomposed into orthogonal functional principal components and their associated scores as

$$\begin{aligned} f_t^{(l)}(x) &= \mu^{(l)}(x) + \sum_{k=1}^{\infty} \beta_{t,k}^{(l)} \phi_k^{(l)}(x) \\ &= \mu^{(l)}(x) + \sum_{k=1}^K \beta_{t,k}^{(l)} \phi_k^{(l)}(x) + e_t^{(l)}(x), \end{aligned} \quad (2)$$

where  $\mu^{(l)}(x)$  is the mean function for the  $l^{\text{th}}$  subpopulation;  $\{\phi_1^{(l)}(x), \dots, \phi_K^{(l)}(x)\}$  is a set of the first  $K$  functional principal components;  $\{\beta_1^{(l)}, \dots, \beta_K^{(l)}\}$ , with elements  $\beta_1^{(l)} = [\beta_{1,1}^{(l)}, \dots, \beta_{n,1}^{(l)}]^{\top}$ , denotes a set of principal component scores for the  $l^{\text{th}}$  subpopulation;  $e_t^{(l)}(x)$  denotes the model truncation error function with mean zero and finite variance for the  $l^{\text{th}}$  subpopulation; and  $K < n$  is the number of retained principal components.

Following [Shang and Hyndman \(2017\)](#), the number of retained components is determined as the minimum that reaches 95% of total variance explained by the leading components<sup>2</sup>, such that

$$K = \operatorname{argmin}_{K:K \geq 1} \left\{ \frac{\sum_{k=1}^K \hat{\lambda}_k}{\sum_{k=1}^{\infty} \hat{\lambda}_k \mathbb{1}_{\{\hat{\lambda}_k > 0\}}} \geq 0.95 \right\},$$

where  $\mathbb{1}\{\cdot\}$  represents the binary indicator function. Since all individual elements of  $C(x, z)$  share

<sup>2</sup>Several alternative methods are used for selecting the number of functional principal components  $K$ , such as those of [Chiou \(2012\)](#); [Yao et al. \(2005\)](#); [Rice and Silverman \(1991\)](#); [Hall and Vial \(2006\)](#).

the same set of eigenvalues, only one common  $K$  value is needed in dimension-reduction for all subpopulations.

Expansion (2) facilitates dimension reduction since the first  $K$  terms often provide a good approximation to the infinite sums, and thus the information contained in time series  $\{f_1^{(l)}(x), \dots, f_n^{(l)}(x)\}$  can be adequately summarized by the  $K$ -dimensional vector  $\{\beta_1^{(l)}, \dots, \beta_K^{(l)}\}$ . When collectively considering all  $\omega$  time series, truncating at the first  $K^{\text{th}}$  functional principal components yields the approximation in its matrix formulation as

$$\mathbf{f}_t(x) = \mathbf{\Phi}(x)\boldsymbol{\beta}_t^\top, \quad (3)$$

where  $\boldsymbol{\beta}_t = \{\beta_{t,1}^{(1)}, \dots, \beta_{t,K}^{(1)}, \beta_{t,1}^{(2)}, \dots, \beta_{t,K}^{(2)}, \dots, \beta_{t,1}^{(\omega)}, \dots, \beta_{t,K}^{(\omega)}\}$  is the vector of the basis expansion coefficients, and

$$\mathbf{\Phi}(x) = \begin{pmatrix} \phi_1^{(1)}(x) & \dots & \phi_K^{(1)}(x) & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & \phi_1^{(2)}(x) & \dots & \phi_K^{(2)}(x) & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 & \phi_1^{(\omega)}(x) & \dots & \phi_K^{(\omega)}(x) \end{pmatrix}_{\omega \times (K \times \omega)}.$$

For dense and regularly spaced functional time series, such as age-specific mortality rates, the mean function  $\hat{\boldsymbol{\mu}}(x) = \frac{1}{n} \sum_{t=1}^n \mathbf{f}_t(x)$  and the covariance function  $\hat{\mathbf{C}}(x, z)$  can be empirically estimated. Based on Eq. (3), we extract empirical basis functions from the empirical covariance function  $\hat{\mathbf{\Phi}}(x)$ , and then make  $h$ -step-ahead point forecasts as

$$\begin{aligned} \hat{\mathbf{f}}_{n+h|n} &= \mathbb{E}[\mathbf{f}_{n+h}(x) | \mathbf{f}_1(x), \dots, \mathbf{f}_n(x); \hat{\mathbf{\Phi}}(x)] \\ &= \hat{\boldsymbol{\mu}}(x) + \sum_{k=1}^K \hat{\boldsymbol{\beta}}_{n+h|n,k} \hat{\mathbf{\Phi}}(x), \end{aligned}$$

where the forecast principal component score  $\hat{\boldsymbol{\beta}}_{n+h|n,k}$  is obtained using the automated autoregressive integrated moving average (ARIMA) model-fitting algorithm of Hyndman and Khandakar (2008). More details about forecasting functional principal component scores can be found in Hyndman and Shang (2009).

We adopt a univariate time series forecasting method of Hyndman and Shang (2009) to obtain the forecast principal component score  $\hat{\boldsymbol{\beta}}_{n+h|n,k}$ . This univariate time series forecasting method can model a time series with a stochastic trend component. Since the yearly age-specific mortality rates

do not contain seasonality, the ARIMA has the general form of

$$(1 - \tau_1 B - \dots - \tau_p B^p)(1 - B)^d \beta_k = \varphi + (1 + \nu_1 B + \dots + \nu_q B^q) \mathbf{w}_k,$$

where  $\varphi$  represents the intercept,  $(\tau_1, \dots, \tau_p)$  denote the coefficients associated with the autoregressive component,  $\beta_k = \{\beta_{1,k}, \dots, \beta_{n,k}\}$  represents principal component scores,  $(\nu_1, \dots, \nu_q)$  denote the coefficients associated with the moving average component,  $B$  denotes the backshift operator,  $d$  denotes the differencing operator, and  $\mathbf{w}_k = \{w_{1,k}, \dots, w_{n,k}\}$  represents a white-noise error term. We use the automatic algorithm of [Hyndman and Khandakar \(2008\)](#) to choose the optimal autoregressive order  $p$ , moving average order  $q$ , and difference order  $d$ . The value of  $d$  is selected based on successive Kwiatkowski-Phillips-Schmidt-Shin (KPSS) unit root tests ([Kwiatkowski et al., 1992](#)). KPSS tests are used for testing the null hypothesis that an observable time series is stationary around a deterministic trend. We first test the original time series for a unit root; if the test result is significant, then we test the differenced time series for a unit root. The procedure continues until we obtain our first insignificant result. Once the order  $d$  is determined, the orders of  $p$  and  $q$  are selected based on the optimal AIC with a correction for a small finite sample size ([Akaike, 1974](#)). Having identified the optimal ARIMA model, the maximum likelihood method can be used to estimate the parameters.

## 4 Forecast reconciliation for grouped functional time series

### 4.1 Notation

For both hierarchies illustrated in [Figure 3](#), we denote a particular disaggregated series using the notation  $G * S$ , meaning the geographical area  $G$  and the sex  $S$ . For instance,  $R_1 * F$  denotes females in Region 1,  $A_1 * T$  denotes females and males in Area 1, and  $Australia * M$  denotes males in Australia. Let  $E_{G*S,t}(x)$  denote the exposure to risk for series  $G * S$  in year  $t$  and age  $x$ , and let  $D_{G*S,t}(x)$  be the number of deaths for series  $G * S$  in year  $t$  and age  $x$ . Then, age-specific mortality rate is given by  $R_{G*S,t}(x) = D_{G*S,t}(x) / E_{G*S,t}(x)$ . Dropping the age variable ( $x$ ) allows us to express the national and sub-national mortality series in a matrix multiplication as

$$\begin{array}{c}
R_{\text{Australia}*T,t} \\
R_{\text{Australia}*F,t} \\
R_{\text{Australia}*M,t} \\
R_{R1*T,t} \\
R_{R2*T,t} \\
\vdots \\
R_{R11*T,t} \\
R_{R1*F,t} \\
R_{R2*F,t} \\
\vdots \\
R_{R11*F,t} \\
R_{R1*M,t} \\
R_{R2*M,t} \\
\vdots \\
R_{R11*M,t} \\
R_{A1*T,t} \\
R_{A2*T,t} \\
\vdots \\
R_{A47*T,t} \\
R_{A1*F,t} \\
R_{A1*M,t} \\
R_{A2*F,t} \\
R_{A2*M,t} \\
\vdots \\
R_{A47*F,t} \\
R_{A47*M,t}
\end{array}
=
\begin{array}{c}
\frac{E_{A1*F,t}}{E_{\text{Australia}*T,t}} \quad \frac{E_{A1*M,t}}{E_{\text{Australia}*T,t}} \quad \frac{E_{A2*F,t}}{E_{\text{Australia}*T,t}} \quad \frac{E_{A2*M,t}}{E_{\text{Australia}*T,t}} \quad \frac{E_{A3*F,t}}{E_{\text{Australia}*T,t}} \quad \frac{E_{A3*M,t}}{E_{\text{Australia}*T,t}} \quad \dots \quad \frac{E_{A47*F,t}}{E_{\text{Australia}*T,t}} \quad \frac{E_{A47*M,t}}{E_{\text{Australia}*T,t}} \\
\frac{E_{A1*F,t}}{E_{\text{Australia}*F,t}} \quad 0 \quad \frac{E_{A2*F,t}}{E_{\text{Australia}*F,t}} \quad 0 \quad \frac{E_{A3*F,t}}{E_{\text{Australia}*F,t}} \quad 0 \quad \dots \quad \frac{E_{A47*F,t}}{E_{\text{Australia}*F,t}} \quad 0 \\
0 \quad \frac{E_{A1*M,t}}{E_{\text{Australia}*M,t}} \quad 0 \quad \frac{E_{A2*M,t}}{E_{\text{Australia}*M,t}} \quad 0 \quad \frac{E_{A3*M,t}}{E_{\text{Australia}*M,t}} \quad \dots \quad 0 \quad \frac{E_{A47*M,t}}{E_{\text{Australia}*M,t}} \\
\frac{E_{A1*F,t}}{E_{R1*T,t}} \quad \frac{E_{A1*M,t}}{E_{R1*T,t}} \quad 0 \quad 0 \quad 0 \quad 0 \quad \dots \quad 0 \quad 0 \\
0 \quad 0 \quad \frac{E_{A2*F,t}}{E_{R2*T,t}} \quad \frac{E_{A2*M,t}}{E_{R2*T,t}} \quad \frac{E_{A3*F,t}}{E_{R2*T,t}} \quad \frac{E_{A3*M,t}}{E_{R2*T,t}} \quad \dots \quad 0 \quad 0 \\
\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \dots \quad \vdots \quad \vdots \\
0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad \dots \quad \frac{E_{A47*F,t}}{E_{R8*T,t}} \quad \frac{E_{A47*M,t}}{E_{R8*T,t}} \\
\frac{E_{A1*F,t}}{E_{R1*F,t}} \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad \dots \quad 0 \quad 0 \\
0 \quad 0 \quad \frac{E_{A2*F,t}}{E_{R2*F,t}} \quad 0 \quad \frac{E_{A3*F,t}}{E_{R2*F,t}} \quad 0 \quad \dots \quad 0 \quad 0 \\
\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \dots \quad \vdots \quad \vdots \\
0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad \dots \quad \frac{E_{A47*F,t}}{E_{R8*F,t}} \quad 0 \\
0 \quad \frac{E_{A1*M,t}}{E_{R1*M,t}} \quad 0 \quad 0 \quad 0 \quad 0 \quad \dots \quad 0 \quad 0 \\
0 \quad 0 \quad 0 \quad \frac{E_{A2*M,t}}{E_{R2*M,t}} \quad 0 \quad \frac{E_{A3*M,t}}{E_{R2*M,t}} \quad \dots \quad 0 \quad 0 \\
\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \dots \quad \vdots \quad \vdots \\
0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad \dots \quad 0 \quad \frac{E_{A47*M,t}}{E_{R8*M,t}} \\
\frac{E_{A1*F,t}}{E_{A1*T,t}} \quad \frac{E_{A1*M,t}}{E_{A1*T,t}} \quad 0 \quad 0 \quad 0 \quad 0 \quad \dots \quad 0 \quad 0 \\
0 \quad 0 \quad \frac{E_{A2*F,t}}{E_{A2*T,t}} \quad \frac{E_{A2*M,t}}{E_{A2*T,t}} \quad 0 \quad 0 \quad \dots \quad 0 \quad 0 \\
\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \dots \quad \vdots \quad \vdots \\
0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad \dots \quad \frac{E_{A47*F,t}}{E_{A47*T,t}} \quad \frac{E_{A47*M,t}}{E_{A47*T,t}} \\
1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad \dots \quad 0 \quad 0 \\
0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad \dots \quad 0 \quad 0 \\
0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad \dots \quad 0 \quad 0 \\
0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0 \quad \dots \quad 0 \quad 0 \\
\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \dots \quad \vdots \quad \vdots \\
0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad \dots \quad 1 \quad 0 \\
0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad \dots \quad 0 \quad 1
\end{array}
\begin{array}{c}
R_{A1*F,t} \\
R_{A1*M,t} \\
R_{A2*F,t} \\
R_{A2*M,t} \\
\vdots \\
R_{A47*F,t} \\
R_{A47*M,t}
\end{array}
\begin{array}{c}
\mathbf{b}_t
\end{array}$$

$\underbrace{\hspace{15em}}_{\mathbf{R}_t} \quad \underbrace{\hspace{15em}}_{\mathbf{S}_t}$

or  $\mathbf{R}_t = \mathbf{S}_t \mathbf{b}_t$ , where  $\mathbf{R}_t$  is a vector containing all series at all levels of disaggregation,  $\mathbf{b}_t$  is a vector of the most disaggregated series, and  $\mathbf{S}_t$  shows how the two are related. We present a brief review of three methods for forecast reconciliation based on this equation in Sections 4.2 to 4.5.

## 4.2 The bottom-up (BU) method

A widely used approach to reconcile grouped functional time series is the bottom-up method (see, e.g., [Dangerfield and Morris, 1992](#); [Zellner and Tobias, 2000](#)). This method has the agreeable feature that it is intuitive and straightforward and always results in forecasts that satisfy the same group structure as the original data.

First, we adopt the multivariate functional mortality forecasting method to obtain  $h$ -step-ahead base forecasts for the most disaggregated series, denoting it by  $\hat{\mathbf{b}}_{n+h} = \left[ \hat{R}_{A_1 * F, n+h}, \hat{R}_{A_1 * M, n+h}, \hat{R}_{A_2 * F, n+h}, \hat{R}_{A_2 * M, n+h}, \dots, \hat{R}_{A_{47} * F, n+h}, \hat{R}_{A_{47} * M, n+h} \right]^\top$ . Then, we use empirical ratios to form the  $\mathbf{S}_t$  for all  $t \in [1993, 2016]$  (see also [Shang and Haberman, 2017](#)). Thus, we obtain reconciled forecasts for all series

$$\bar{\mathbf{R}}_{n+h} = \mathbf{S}_{n+h} \hat{\mathbf{b}}_{n+h},$$

where  $\bar{\mathbf{R}}_{n+h}$  denotes the reconciled forecasts.

## 4.3 The optimal combination (OP) method

Instead of considering only the bottom-level series, [Hyndman et al. \(2011\)](#) proposed a method in which base forecasts for all aggregated and disaggregated series are computed successively. Then the resulting forecasts are combined through linear regression. The reconciled forecasts are as close as possible to the base forecasts and aggregate consistently within the group.

The method is derived by writing the base forecasts as the response variable of the linear regression:

$$\hat{\mathbf{R}}_{n+h} = \mathbf{S}_{n+h} \boldsymbol{\beta}_{n+h} + \boldsymbol{\epsilon}_{n+h},$$

where  $\hat{\mathbf{R}}_{n+h}$  is a matrix of  $h$ -step-ahead base forecasts for all series, stacked in the same order as for the original data;  $\boldsymbol{\beta}_{n+h} = E[\mathbf{b}_{n+h} | \mathbf{R}_1, \dots, \mathbf{R}_n]$  is the unknown mean of the forecast distributions of the most disaggregated series; and  $\boldsymbol{\epsilon}_{n+h}$  represents the reconciliation error which is independent of past observations, with mean zero and variance-covariance matrix  $\boldsymbol{\Sigma}_h := \text{var}(\boldsymbol{\epsilon}_{n+h})$ .

We follow [Hyndman et al. \(2011\)](#) and [Hyndman et al. \(2016\)](#) and estimate the regression coefficient in a weighted least squares fashion:

$$\hat{\boldsymbol{\beta}}_{n+h} = \left( \mathbf{S}_{n+h}^\top \mathbf{W}^{-1} \mathbf{S}_{n+h} \right)^{-1} \mathbf{S}_{n+h}^\top \mathbf{W}^{-1} \hat{\mathbf{R}}_{n+h},$$

where  $\mathbf{W}$  is a weight matrix. To guarantee the invertibility of  $\mathbf{W}$ , we assume  $\mathbf{W} = c \times \mathbf{I}$ , where  $c$  is any constant and  $\mathbf{I}$  denotes the identity matrix. Thus, the reconciled forecasts can be obtained via the



ordinary least-squares method as

$$\begin{aligned}\bar{\mathbf{R}}_{n+h} &= \mathbf{S}_{n+h}\hat{\boldsymbol{\beta}}_{n+h} = \mathbf{S}_{n+h} \left[ \mathbf{S}_{n+h}^\top (\mathbf{c} \times \mathbf{I})^{-1} \mathbf{S}_{n+h} \right]^{-1} \mathbf{S}_{n+h}^\top (\mathbf{c} \times \mathbf{I})^{-1} \hat{\mathbf{R}}_{n+h} \\ &= \mathbf{S}_{n+h} \left( \mathbf{S}_{n+h}^\top \mathbf{S}_{n+h} \right)^{-1} \mathbf{S}_{n+h}^\top \hat{\mathbf{R}}_{n+h}.\end{aligned}$$

By construction, these are aggregate consistent and involve a combination of all the base forecasts. They are also unbiased since  $E[\bar{\mathbf{R}}_{n+h}] = \mathbf{S}_{n+h}\boldsymbol{\beta}_{n+h}$ .

#### 4.4 The trace minimization (MinT) method

The optimal combination method attempts only to minimize the variance of reconciliation errors in linear regression. Wickramasuriya et al. (2019) proposed an improved approach of finding coherent forecasts for the grouped functional time series. By minimizing the sum of variances of reconciliation errors, essentially minimizing trace of the variance matrix, the new approach produces coherent forecasts across the entire collection of time series. Since the variance-covariance matrix of out-of-sample reconciliation errors is often not known and not identifiable in practice, the MinT method attempts to approximate  $\boldsymbol{\Sigma}_h$  within-sample base forecast errors.

Define the  $h$ -step-ahead in-sample error of base forecasts as

$$\hat{\mathbf{e}}_{t+h} = \mathbf{R}_{t+h} - \hat{\mathbf{R}}_{t+h},$$

and the corresponding error of reconciled forecasts using information up to and including time  $t$  as

$$\tilde{\mathbf{e}}_{t+h} = \mathbf{R}_{t+h} - \bar{\mathbf{R}}_{t+h},$$

with  $t = 1, \dots, n - h$ . Wickramasuriya et al. (2019) proves that the variance-covariance matrix of  $\tilde{\mathbf{e}}_{t+h}$  can be written as

$$\text{var}[\mathbf{R}_{t+h} - \bar{\mathbf{R}}_{t+h} | \mathbf{R}_1, \dots, \mathbf{R}_t] = \mathbf{S}_{t+h} \mathbf{P} \mathbf{W}_h \mathbf{S}_{t+h}^\top \mathbf{P}^\top,$$

where  $\mathbf{P}$  is a projection matrix satisfying  $\mathbf{S}_{t+h} \mathbf{P} \mathbf{S}_{t+h} = \mathbf{S}_{t+h}$ , and  $\mathbf{W}_h = E[\hat{\mathbf{e}}_{t+h} \hat{\mathbf{e}}_{t+h}^\top | \mathbf{R}_1, \dots, \mathbf{R}_t]$  is the variance-covariance matrix of the  $h$ -step-ahead base forecast errors. By minimizing the trace of  $\mathbf{S}_{t+h} \mathbf{P} \mathbf{W}_h \mathbf{S}_{t+h}^\top \mathbf{P}^\top$ , we obtain

$$\mathbf{P} = \left( \mathbf{S}_{t+h}^\top \mathbf{W}_h^{-1} \mathbf{S}_{t+h} \right)^{-1} \mathbf{S}_{t+h}^\top \mathbf{W}_h^{-1},$$

leading to the reconciled forecasts from the MinT approach:

$$\bar{\mathbf{R}}_{n+h} = \mathbf{S}_{t+h} \left( \mathbf{S}_{t+h}^\top \mathbf{W}_h^{-1} \mathbf{S}_{t+h} \right)^{-1} \mathbf{S}_{t+h}^\top \mathbf{W}_h^{-1} \hat{\mathbf{R}}_{n+h}.$$

In this equation,  $\mathbf{W}_h$  is the covariance matrix of the base forecast errors. Our study is estimated by shirking off-diagonal entries of the unbiased sample covariance estimator of the in-sample 1-step-ahead base forecast errors towards targets on its diagonal; for more details together with alternative estimations of  $\mathbf{W}_h$ , consult Section 2.4 of [Wickramasuriya et al. \(2019\)](#).

#### 4.5 The forecast combination (Comb\_av) method

It is possible to combine the forecast mentioned above reconciliation methods to reduce bias, variance, and uncertainty of forecasts. [Shang and Haberman \(2020\)](#) recently considered a forecast combination method that computes the averaged forecast for horizon  $h$  as

$$\bar{\mathbf{R}}_{n+h}^{\text{comb}} = \sum_{g=1}^G w_g \bar{\mathbf{R}}_{n+h}^g,$$

where  $\bar{\mathbf{R}}_{n+h}^g$  denotes forecasts obtained via a reconciliation method, and  $\{w_1, \dots, w_g\}$  are weights summing to 1. Following [Shang and Haberman \(2020\)](#), we adopt a simple average combination method that applies equal weighting on the BU, OP and MinT reconciled forecasts as

$$\bar{\mathbf{R}}_{n+h}^{\text{Comb.av}} = \frac{1}{3} [\bar{\mathbf{R}}_{n+h}^{\text{BU}} + \bar{\mathbf{R}}_{n+h}^{\text{OP}} + \bar{\mathbf{R}}_{n+h}^{\text{MinT}}].$$

## 5 Empirical application results

The Multivariate functional time series forecasting method, and a univariate method of [Shang and Hyndman \(2017\)](#), are applied to Australian age-specific mortality rates to obtain base forecasts following the hierarchy structures in [Figure 3](#). We then conduct reconciliation via the BU, OP, MinT, and Comb\_av methods. To assess model and parameter stabilities over time, we consider an expanding window analysis of considered time series models (see [Zivot and Wang, 2006](#), Chapter 9 for details). Specifically, we initially use the observed mortality curves from 1993 to 2011 to produce one- to 5-step-ahead point forecasts. Through an expanding window approach, we re-estimate the parameters of considered models using the first 20 years of observations from 1993 to 2002 and make one- to 4-step-ahead forecasts using re-estimated models. The process is iterated with the sample size increased by one year until the end of the data period in 2016. This process produces

five one-step-ahead forecasts, four two-step-ahead forecasts, and so on, up to one five-step-ahead forecast. We evaluate the point forecast accuracy and report our study results.

## 5.1 Point forecast error measures

To evaluate the point forecast accuracy, we compute the mean absolute forecast error (MAFE), and the root mean squared forecast error (RMSFE). Because the raw Australian sub-national mortality rates cover 18 age groups and contain missing values for some areas, we define the MAFE for each series  $j$  as

$$\text{MAFE}_j(h) = \frac{1}{18 \times (6 - h)} \sum_{\zeta=h}^5 \sum_{i=1}^{18} \left| \left[ f_{n+\zeta}^{(j)}(x_i) - \widehat{f}_{n+\zeta}^{(j)}(x_i) \right] \mathbb{1}_{\{f_{n+\zeta}^{(j)}(x_i) > 0\}} \right|,$$

where  $f_{n+\zeta}^{(j)}(x_i)$  represents the actual holdout sample for the  $i^{\text{th}}$  age and  $\zeta^{\text{th}}$  curve of the forecasting period, and  $\widehat{f}_{n+\zeta}^{(j)}(x_i)$  is the corresponding point forecasts;  $\mathbb{1}_{\{f_{n+\zeta}^{(j)}(x_i) > 0\}}$  is to exclude missing death rate observations with  $\mathbb{1}_{\{\cdot\}}$  representing the binary indicator function. For each series  $j$ , the RMSFE is given by

$$\text{RMSFE}_j(h) = \sqrt{\frac{1}{18 \times (6 - h)} \sum_{\zeta=h}^5 \sum_{i=1}^{18} \left[ f_{n+\zeta}^{(j)}(x_i) - \widehat{f}_{n+\zeta}^{(j)}(x_i) \right]^2 \mathbb{1}_{\{f_{n+\zeta}^{(j)}(x_i) > 0\}}}.$$

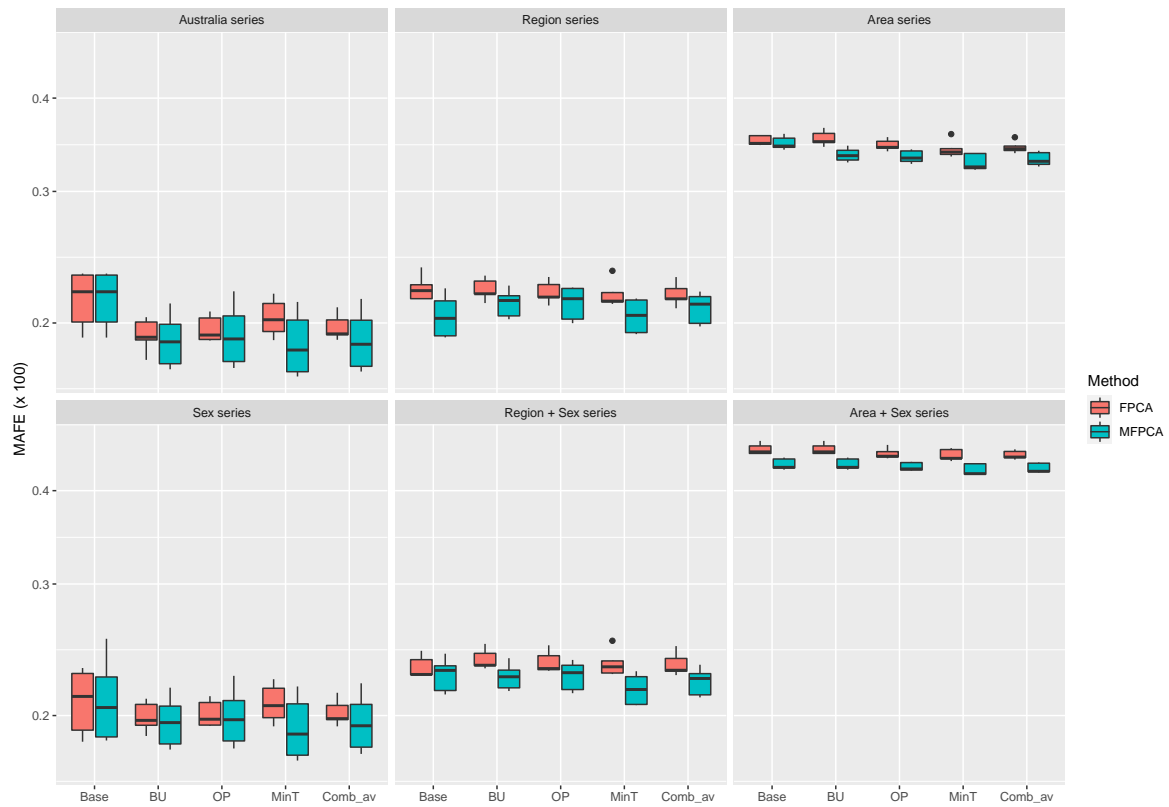
Averaging over five forecast horizons, we obtain measures of point forecast accuracies for all national and sub-national mortality series as

$$\text{Mean (MAFE)} = \frac{1}{5} \sum_{h=1}^5 \text{MAFE}(h), \quad \text{and} \quad \text{Mean (RMSFE)} = \frac{1}{5} \sum_{h=1}^5 \text{RMSFE}(h).$$

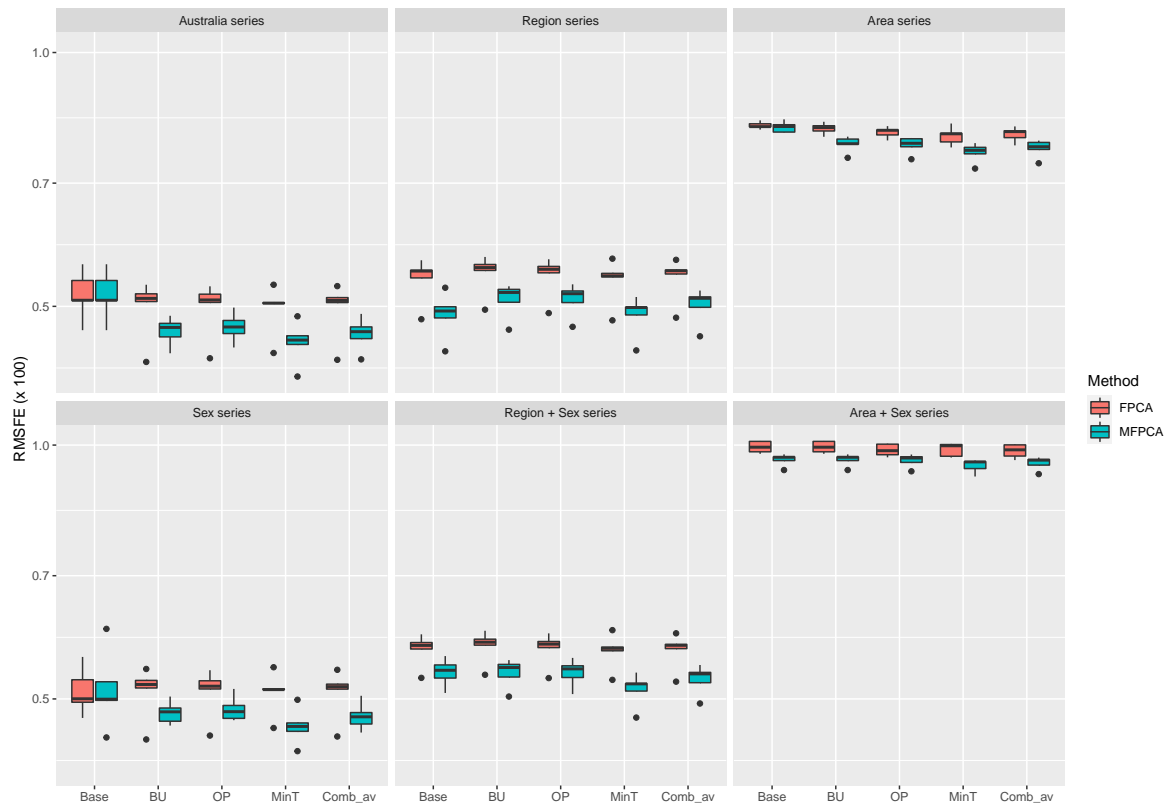
## 5.2 Point forecast results

We present the mean MAFEs ( $\times 100$ ) for the base and reconciled forecasts in Figure 4, and the mean RMSFE ( $\times 100$ ) in Figure 5, both by calculating the averages of all series at each disaggregation level of hierarchies considered. These forecasts are obtained by the proposed multivariate functional time series (MFPCA) method and its univariate (FPCA) counterpart. At all the sub-national levels, jointly modeling age-specific mortality rates produces more accurate point forecasts than considering each series individually using the FPCA method. The MFPCA method can extract common mortality features of strongly-related populations at a level of disaggregation. At the national level, where only one mortality time series exists, the FPCA model is equivalent to the MFPCA model with  $\omega = 1$ . (FPCA can be regarded as a special case of MFPCA when only a single series is considered.) Hence,

both methods yield the same forecasts for the Australia series.



**Figure 4:**  $MAFE (\times 100)$  in the holdout sample between the univariate and multivariate functional time series methods.

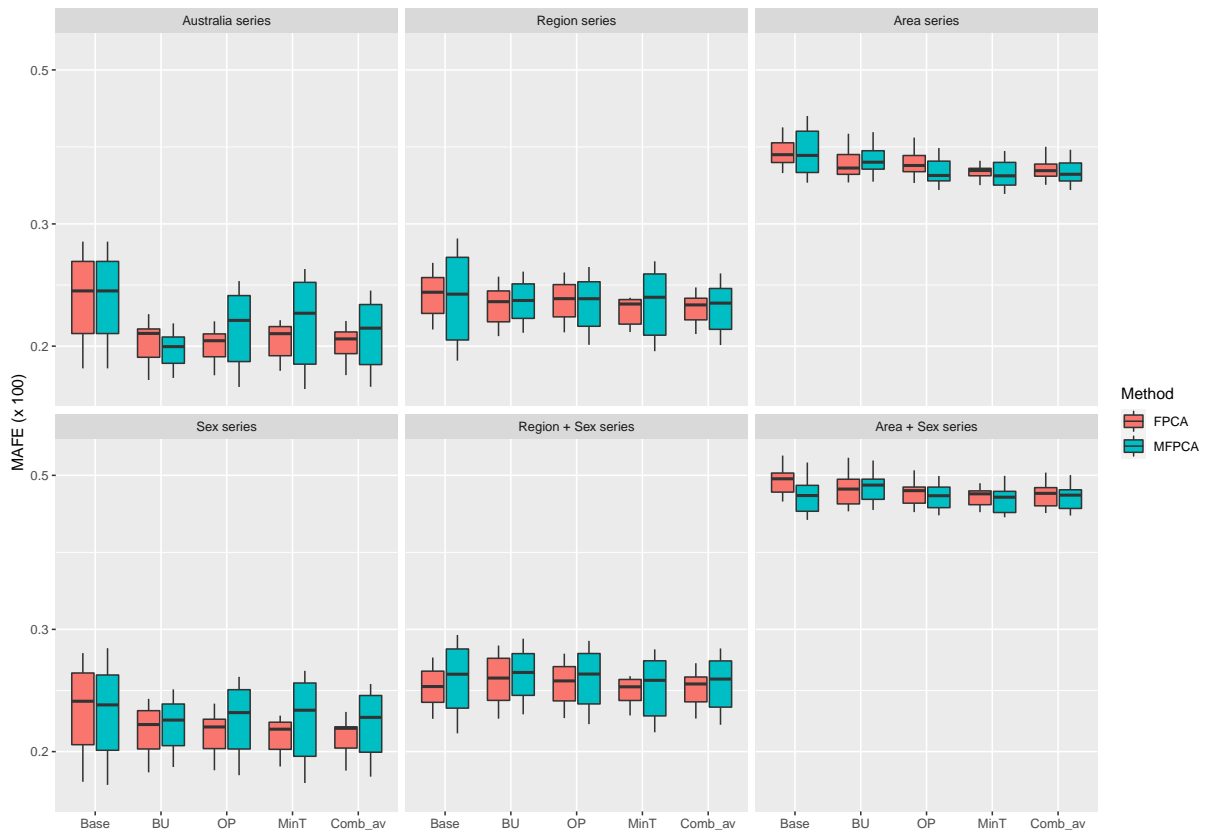


**Figure 5:**  $RMSFE (\times 100)$  in the holdout sample between the univariate and multivariate functional time series methods.

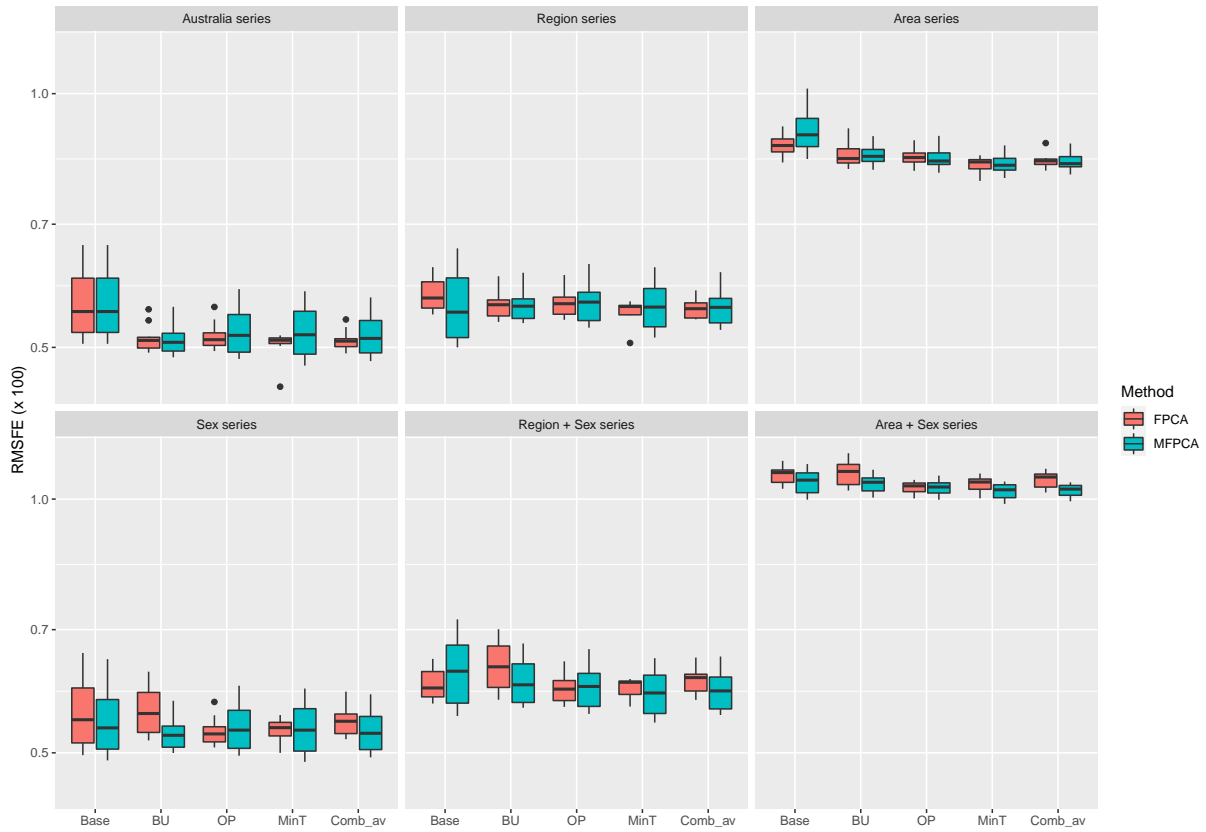
Figures 4 and 5 also show that pairing any of three reconciliation methods (BU, OP, and MinT), or using a mix of three (Comb\_av), with either MFPCA or FPCA improves the point forecast accuracy. Among the reconciliation methods considered, MinT yields the most accurate overall point forecasts when used with MFPCA. This result can be attributed to the MinT method incorporating all information from a full covariance matrix of forecast errors in obtaining a set of coherent forecasts. Both figures also indicate that forecast combination can help reducing point forecast errors, confirming findings of Shang and Hyndman (2017) and Shang and Haberman (2020).

### 5.3 Robustness check

We conduct an extended empirical application with one- to 10-step-ahead forecasts using the Australian age-specific mortality rates. We point out that the current dataset only covers death rates in regional and remote areas from 1993 to 2016. Moreover, most of the missing values in the dataset occur for the early years. Extending the forecasting horizon to  $h = 10$  produce less accurate forecasts than those in Section 5. We provide the mean MAFE and RMSFE results corresponding to the extended application in Figure 6 and Figure 7, respectively. It can be seen that the MFPCA method still outperforms the FPCA method in general.



**Figure 6:** MAFE ( $\times 100$ ) in the holdout sample between the univariate and multivariate functional time series methods;  $h = 1, \dots, 10$ .

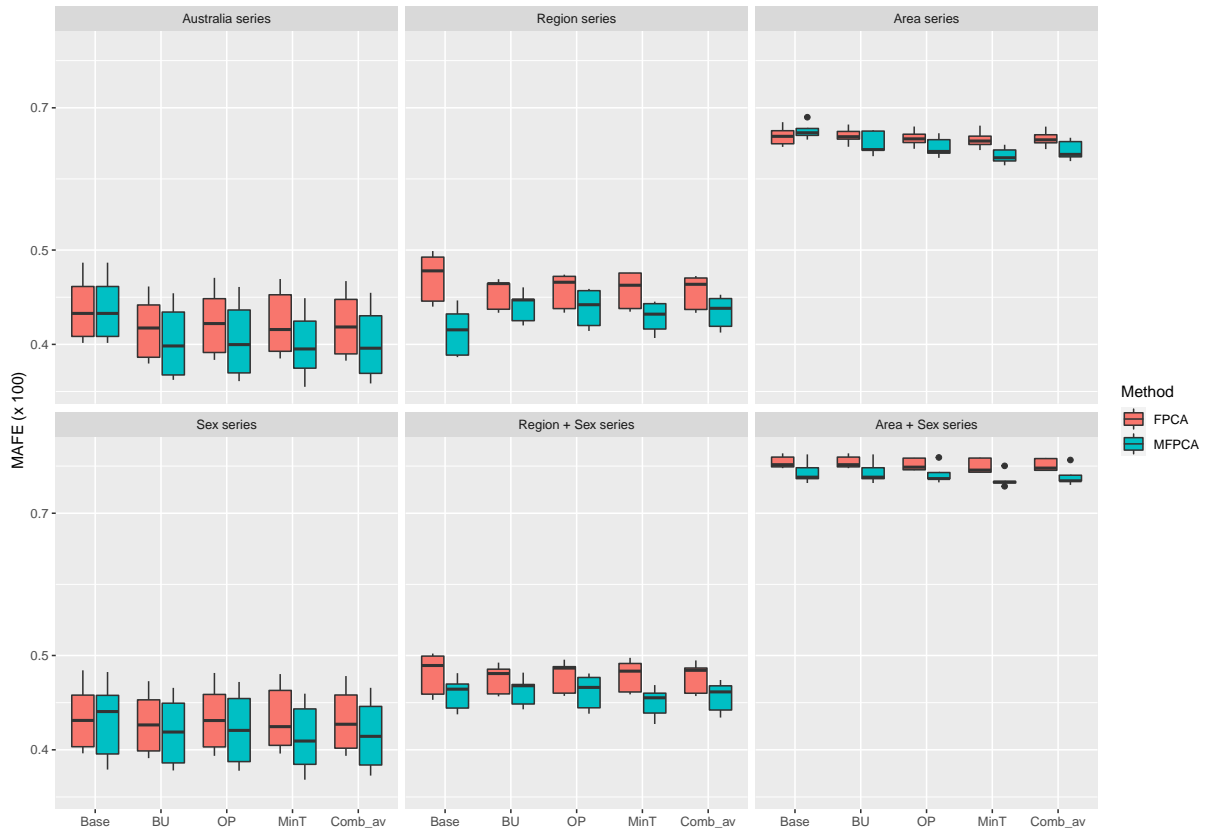


**Figure 7:** *RMSFE ( $\times 100$ ) in the holdout sample between the univariate and multivariate functional time series methods;  $h = 1, \dots, 10$ .*

Further, it is well known that mortality rates have different structures throughout ages. As shown in Figure 2, mortality rates for young people are generally more volatile than those for middle and old age. To check the robustness of our point forecasting results, we conduct an additional application using mortality rates for ages greater than 50. The mean MAFEs and RMSFEs of the robustness check are shown in Figure 8 and Figure 9, respectively. We can see that the MFPCA method consistently outperforms the conventional FPCA method in forecasting age-specific mortality rates for ages between 50-85. Similar results reported in Sections 5.2 and 5.3 indicate the usefulness of the MFPCA method in modeling and forecasting age-specific mortality rates in a hierarchical structure.

## 6 Conclusion

We extended the univariate functional time series forecasting method to a multivariate framework. The proposed multivariate functional time series method can be combined with reconciliation methods when applied to functional data formed by disaggregated series, such as sub-national age-specific mortality rates. The reconciliation methods, such as the bottom-up, optimal combination, trace minimization, and forecast combination methods, can improve point forecast accuracy. The complete approach described in this paper can be viewed as a grouped multivariate functional time



**Figure 8:**  $MAFE (\times 100)$  in the holdout sample between the univariate and multivariate functional time series methods, when mortality rates for ages between 50-95 are considered.



**Figure 9:**  $RMSFE (\times 100)$  in the holdout sample between the univariate and multivariate functional time series methods, when mortality rates for ages between 50-95 are considered.

series forecasting method.

The proposed method is applied to Australian age-specific mortality rates from 1993 to 2016. The out-of-sample forecasting results show the superior performance of our approach for point projections, compared with the conventional univariate functional forecasting method. Among the three forecast reconciliation methods, the trace minimization method shows the best performance in the study of Australian mortality data. A forecast combination method taking averages of the bottom-up, the optimal combination, and the trace minimization forecasts also has similar superior point forecast accuracy.

There are several ways in which the present paper can be further extended:

- 1) We considered two factors, namely sex, and geography, to disaggregate Australian sub-national mortality rates. It is possible to enrich the information contained in the group structure by differentiating these rates according to other factors, such as the cause of death ([Murray and Lopez, 1997](#); [Gaille and Sherris, 2015](#)), identification of Indigenous Australians (Aboriginal and Torres Strait Islander) ([Thomson, 1991](#); [Supramaniam et al., 2006](#)), and native-born or immigrant status ([Singh and Siahpush, 2001](#)). If appropriate data are available, we may attempt to extend the proposed grouped multivariate functional time series forecasting method to cause-specific mortality rates and Indigenous Australian mortality rates.
- 2) We have not considered the spatial dependence of Australian subnational populations. The spatial distance among populated areas as a covariate can be considered to incorporate the extent of correlation among populations, and thus build better disaggregation structures with homogeneous sub-groups. With an appropriate spatial distance-based clustering rule for sub-populations, we may develop more accurate mortality models and thus further improve the point forecast accuracy.
- 3) We only considered point forecasts in this study. It is also possible to compute interval forecasts to provide a precise estimate of the probability that the future realizations lie within a given range.
- 4) Optimal weights in forecast combination among the forecast reconciliation methods can also be a further pursuit to include more complicated algorithms such as cross-validation.



## References

- ABS (2012), 'Year Book Australia 2012', *Australian Bureau of Statistics* . URL: <https://www.abs.gov.au/AUSSTATS/abs@.nsf/allprimarymainfeatures/02B2D95C0F67E84CCA257A07001BFB46?opendocument>.
- AIHW (1998), *Health in rural and remote Australia*, Vol. AIHW Cat. No. PHE 6, Australian Institute of Health & Welfare, Canberra.
- AIHW (2003), *Rural, regional and remote health: A study on mortality*, Vol. AIHW cat. no. PHE 45, Australian Institute of Health & Welfare, Canberra.
- AIHW (2007), *Rural, regional and remote health: A study on mortality (2nd edition)*, Vol. Rural health series no. 8. Cat. no. PHE 95, Australian Institute of Health & Welfare, Canberra.
- Akaike, H. (1974), 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control* **19**(6), 716–723.
- Aneiros-Pérez, G. and Vieu, P. (2008), 'Nonparametric time series prediction: A semi-functional partial linear modeling', *Journal of Multivariate Analysis* **99**(5), 834–857.
- Aue, A., Horváth, L. and Pellatt, D. F. (2017), 'Functional generalized autoregressive conditional heteroskedasticity', *Journal of Time Series Analysis* **38**(1), 3–21.
- Aue, A., Norinho, D. D. and Hörmann, S. (2015), 'On the prediction of stationary functional time series', *Journal of the American Statistical Association: Theory and Methods* **110**(509), 378–392.
- Besse, P., Cardot, H. and Stephenson, D. (2000), 'Autoregressive forecasting of some functional climatic variations', *Scandinavian Journal of Statistics* **27**(4), 673–687.
- Booth, H. and Tickle, L. (2008), 'Mortality modelling and forecasting: A review of methods', *Annals of actuarial science* **3**(1-2), 3–43.
- Bosq, D. (2000), *Linear Processes in Function Spaces*, Lecture notes in Statistics, New York.
- Bosq, D. and Blanke, D. (2007), *Inference and Prediction in Large Dimensions*, John Wiley & Sons, West Sussex, England.
- Chiou, J.-M. (2012), 'Dynamical functional prediction and classification with application to traffic flow prediction', *The Annals of Applied Statistics* **6**(4), 1588–1614.
- Chiou, J.-M., Chen, Y.-T. and Yang, Y.-F. (2014), 'Multivariate functional principal component analysis: A normalization approach', *Statistica Sinica* **24**(4), 1571–1596.
- Chiou, J.-M. and Müller, H.-G. (2009), 'Modeling hazard rates as functional data for the analysis of cohort lifetables and mortality forecasting', *Journal of the American Statistical Association: Applications & Case Studies* **104**(486), 572–585.
- Coulmas, F. (2007), *Population Decline and Ageing in Japan – the Social Consequences*, Routledge, New

- York.
- Dangerfield, B. J. and Morris, J. S. (1992), 'Top-down or bottom-up: Aggregate versus disaggregate extrapolations', *International Journal of Forecasting* **8**(2), 233–241.
- De Boor, C. (2001), *A practical guide to splines*, Vol. 27 of *Applied Mathematical Sciences*, Springer, New York.
- DPIE and DSHS (1994), 'Rural, remote and metropolitan areas classification: 1991 census edition'.  
**URL:** <https://www.pc.gov.au/inquiries/completed/nursing-home-subsidies/submissions/subdr096/subdr096.pdf>
- Gaille, S. A. and Sherris, M. (2015), 'Causes-of-death mortality: What do we know on their dependence?', *North American Actuarial Journal* **19**(2), 116–128.
- Gao, Y., Shang, H. L. and Yang, Y. (2019), 'High-dimensional functional time series forecasting: An application to age-specific mortality rates', *Journal of Multivariate Analysis* **170**, 232–243.
- Guan, Q. (2018), Document on Creating Population-based Consistent Geography over 1981-2011, Technical report, Australian National University. Retrieved from [https://demography.cass.anu.edu.au/sites/default/files/docs/2018/11/Document\\_on\\_Creating\\_Population-based\\_Consistent\\_Geography\\_over\\_1981-2011.pdf](https://demography.cass.anu.edu.au/sites/default/files/docs/2018/11/Document_on_Creating_Population-based_Consistent_Geography_over_1981-2011.pdf) at February 10, 2020.
- Hall, P. (2011), Principal component analysis for functional data: Methodology, theory, and discussion, in F. Ferraty and Y. Romain, eds, 'The Oxford Handbook of Functional Data Analysis', Oxford University Press, New York, pp. 210–234.
- Hall, P. and Vial, C. (2006), 'Assessing the finite dimensionality of functional data', *Journal of the Royal Statistical Society (Series B)* **68**(4), 689–705.
- He, X. and Ng, P. (1999), 'COBS: Qualitatively constrained smoothing via linear programming', *Computational Statistics* **14**(3), 315–337.
- Hörmann, S., Horváth, L. and Reeder, R. (2013), 'A functional version of the ARCH model', *Econometric Theory* **29**(2), 267–288.
- Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G. and Shang, H. L. (2011), 'Optimal combination forecasts for hierarchical time series', *Computational Statistics and Data Analysis* **55**(9), 2579–2589.
- Hyndman, R. J. and Khandakar, Y. (2008), 'Automatic time series forecasting: the forecast package for R', *Journal of Statistical Software* **27**(3).
- Hyndman, R. J., Lee, A. and Wang, E. (2016), 'Fast computation of reconciled forecasts for hierarchical and grouped time series', *Computational Statistics and Data Analysis* **97**, 16–32.
- Hyndman, R. J. and Shang, H. L. (2009), 'Forecasting functional time series (with discussions)', *Journal of the Korean Statistical Society* **38**(3), 199–221.
- Hyndman, R. J. and Shang, H. L. (2010), 'Rainbow plots, bagplots, and boxplots for functional data',

- Journal of Computational and Graphical Statistics* **19**(1), 29–45.
- Hyndman, R. J. and Ullah, M. S. (2007), ‘Robust forecasting of mortality and fertility rates: A functional data approach’, *Computational Statistics & Data Analysis* **51**(10), 4942–4956.
- Klepsch, J. and Klüppelberg, C. (2017), ‘An innovations algorithm for the prediction of functional linear processes’, *Journal of Multivariate Analysis* **155**, 252–271.
- Klepsch, J., Klüppelberg, C. and Wei, T. (2017), ‘Prediction of functional ARMA processes with an application to traffic data’, *Econometrics and Statistics* **1**, 128–149.
- Kokoszka, P. and Reimherr, M. (2013), ‘Determining the order of the functional autoregressive model’, *Journal of Time Series Analysis* **34**(1), 116–129.
- Kokoszka, P., Rice, G. and Shang, H. L. (2017), ‘Inference for the autocovariance of a functional time series under conditional heteroscedasticity’, *Journal of Multivariate Analysis* **162**, 32–50.
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. and Shin, Y. (1992), ‘Testing the null hypothesis of stationarity against the alternative of a unit root’, *Journal of Econometrics* **54**(1-3), 159–178.
- Li, D., Robinson, P. M. and Shang, H. L. (2020), ‘Long-range dependent curve time series’, *Journal of the American Statistical Association: Theory and Methods* **in press**.
- Murray, C. J. L. and Lopez, A. D. (1997), ‘Alternative projections of mortality and disability by cause 1990-2020: Global burden of disease study’, *The Lancet* **349**(9064), 1498–1504.
- OECD (2013), *Pensions at a Glance 2013: OECD and G20 indicators*, Technical report, OECD Publishing. Retrieved from [http://dx.doi.org/10.1787/pension\\_glance-2013-en](http://dx.doi.org/10.1787/pension_glance-2013-en) at January 20, 2016.
- Rice, J. and Silverman, B. (1991), ‘Estimating the mean and covariance structure nonparametrically when the data are curves’, *Journal of the Royal Statistical Society (Series B)* **53**(1), 233–243.
- Sefton, J. and Weale, M. (2009), *Reconciliation of National Income and Expenditure: Balanced Estimates of National Income for the United Kingdom, 1920-1990*, Cambridge University Press, Cambridge.
- Shang, H. L. (2014), ‘A survey of functional principal component analysis’, *ASTA Advance in Statistical Analysis* **98**(2), 121–142.
- Shang, H. L. (2016), ‘Mortality and life expectancy forecasting for a group of populations in developed countries: A multilevel functional data method’, *The Annals of Applied Statistics* **10**(3), 1639–1672.
- Shang, H. L. (2017), ‘Forecasting intraday s&p 500 index returns: A functional time series approach’, *Journal of Forecasting* **36**(7), 741–755.
- Shang, H. L. and Haberman, S. (2017), ‘Grouped multivariate and functional time series forecasting: An application to annuity pricing’, *Insurance: Mathematics and Economics* **75**, 166–179.
- Shang, H. L. and Haberman, S. (2020), ‘Forecasting multiple functional time series in a group structure: an application to mortality’, *ASTIN Bulletin* **50**(2), 357–379.

- Shang, H. L. and Hyndman, R. J. (2011), 'Nonparametric time series forecasting with dynamic updating', *Mathematics and Computers in Simulation* **81**(7), 1310–1324.
- Shang, H. L. and Hyndman, R. J. (2017), 'Grouped functional time series forecasting: An application to age-specific mortality rates', *Journal of Computational and Graphical Statistics* **26**(2), 330–343.
- Singh, G. K. and Siahpush, M. (2001), 'All-cause and cause-specific mortality of immigrants and native born in the United States', *American Journal of Public Health* **91**(3), 392.
- Stone, R., Champernowne, D. G. and Meade, J. E. (1942), 'The precision of national income estimates', *The Review of Economic Studies* **9**(2), 111–125.
- Supramaniam, R., Grindley, H. and Pulver, L. J. (2006), 'Cancer mortality in aboriginal people in new south wales, australia, 1994–2002', *Australian and New Zealand Journal of Public Health* **30**(5), 453–456.
- Thomson, N. J. (1991), 'Recent trends in aboriginal mortality', *Medical journal of Australia* **154**(4), 235–239.
- Wang, J.-L., Chiou, J.-M. and Müller, H.-G. (2016), 'Functional data analysis', *Annual Review of Statistics and Its Application* **3**, 257–295.
- Wickramasuriya, S. L., Athanasopoulos, G. and Hyndman, R. J. (2019), 'Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization', *Journal of the American Statistical Association: Theory and Methods* **114**(526), 804–819.
- Yao, F., Müller, H.-G. and Wang, J.-L. (2005), 'Functional data analysis for sparse longitudinal data', *Journal of the American Statistical Association: Theory and Methods* **100**(470), 577–590.
- Zellner, A. and Tobias, J. (2000), 'A note on aggregation, disaggregation and forecasting performance', *Journal of Forecasting* **19**(5), 457–465.
- Zivot, E. and Wang, J. (2006), *Modeling Financial Time Series with S-PLUS*, Springer, New York.