Chapter 12
A Conformal Approach for Distribution-free Prediction of Functional Data

Matteo Fontana, Simone Vantini, Massimo Tavoni and Alexander Gammerman

Abstract Interval prediction has always been a complex problem to solve in the realm of Functional Data Analysis, and the solutions currently proposed to address this very important theoretical and applied issue are not satisfactory. In this contribution we propose a novel approach, based on a non-parametric forecasting approach coming from machine learning, called Conformal Prediction. In the scalar setting, the method is based on simple yet remarkable considerations about sample quantiles. After having stated in a formal way the issue of forecasting for functional data, we develop an algorithm that can be used to generate non-parametric prediction bands for a functional-on-scalar linear regression model. These forecasts are proven to be valid in a statistical sense (i.e., they guarantee a global coverage probability larger or equal to a given threshold) under a very minimal set of assumptions, and thus extremely useful in the statistical practice. The method is then tested on a real world application, namely ensemble emulations for climate economy models, very used in the climate change economics realm.

Matteo Fontana (✉)
MOX - Department of Mathematics, Politecnico di Milano, Italy,
e-mail: matteo.fontana@polimi.it

Simone Vantini
MOX - Department of Mathematics, Politecnico di Milano, Italy,
e-mail: simone.vantini@polimi.it

Massimo Tavoni
Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Italy
RFF-CMCC European Institute on Economics and the Environment (EIEE), Fondazione CMCC, Lecce, Italy, e-mail: massimo.tavoni@polimi.it

Alexander Gammerman
Computer Learning Research Center - Department of Computer Science, Royal Holloway, University of London, UK, e-mail: a.gammerman@rhul.ac.uk

12.1 Introduction

Functional Data Analysis (FDA) [16, 3] is a vibrant and dynamic field of statistics, that aims to develop theory and methods dealing with data points that come in the form of uni- or multivariate functions defined over a one or multi-dimensional domain. The typical mathematical setting of functional data consists in a random sample of independent real-valued univariate functions $Y_1(t), \ldots, Y_n(t)$ defined over the compact interval $T = [t_0, t_1] \subset \mathbb{R}$. The most common embedding of such functions is a Hilbert space such as L^2, but other embeddings are possible, such as Sobolev spaces ([18]) to deal with differential information or Bayes spaces for functional compositions [6].

Still open in FDA research is the the definition and computation of meaningful prediction bands. despite some attempts relying on a parametric setting have been performed [7, 2], it can be argued that the distributional assumptions at the core of these specific methods can't be tested in the statistical practice, which makes these forecasting methods not useful for many real world applications.

The main idea of the present work is to use a relatively new prediction framework, called Conformal Prediction (CP) [19], to develop a statistical method able to produce global prediction bands for functional data, while having guarantees on their statistical validity. Conformal Prediction was developed in the Machine Learning community as a method to define prediction intervals for Support Vector Machines [5]. It has been subsequently used by statisticians as a way to develop distribution-free prediction intervals for regression, both in the low [9] and high-dimensional [10] setting.

This contribution is structured as follows: in Section 12.2 we extend the CP framework to Functional Data Analysis, while in Section 12.3 we discuss the specific role of the Conformity Measure (CM) in predicting functional data in a conformal way, and how CMs define the shape of the prediction set calculated using CP. We then provide an application to a problem arising in climate change economics in Section 12.4.

12.2 Conformal Prediction for Functional Variables

Let us consider a regressive framework, in which we observe n data points

$$Z_1, \ldots, Z_n \sim P$$

where each Z_i, $i = 1, \ldots, n$ is IID. Each Z_i is a tuple $\{X_i, Y_i\}$ in $\mathbb{R}^p \times L^2$. Y_i is the response, a function embedded in $L^2[T]$ space, that is the space of L^2 functions defined over the interval $T = [t_0, t_1]$. X_i is instead a $p-$dimensional vector of covariates $X_i = \left[X_{i,1}, \ldots, X_{p,1}\right]$.

The objective is predicting with confidence a new (functional) value Y_{n+1} from a new vector of covariates X_{n+1} by using a regression operator of the form

$$\mu(x) = \mathbb{E}\left(Y | X = x\right), \; Y \in L^2[T], \; x \in \mathbb{R}^p. \tag{12.1}$$

In a more formal way, given a nominal type-1 error level $\alpha \in (0, 1)$, the idea is to identify sets C in $L^2[T]$ for which the following property holds:

$$\mathbb{P}\left(Y_{n+1} \in C(X_{n+1})\right) \geq 1 - \alpha \tag{12.2}$$

Please note that our objective is more ambitious than the one described in [8], where the authors aim to identify a prediction set not for Y_{n+1} itself, but for $\Pi\left(Y_{n+1}\right)$, its projection over a basis of $L^2[T]$.

The driving example in our case is Linear Function-on-Scalar regression , a specific case of a Functional Linear Model (FLM) [16] in which the covariates are scalars, the response is a function, and $\mu(x)$ is approximated using a t-dependent linear combination of the covariates x. Please note that in the specified setting we do not require μ to be the exact conditional mean: all the results presented still stand also in the case of poorly estimated or even mis-specified regression model.

A Conformal Prediction framework is quite apt to address the problem of identification of prediction sets as described by Equation 12.2. Moreover, the use of CP requires a very minimal set of assumptions (namely, the data tuples being IID).

The basic reasoning behind the developement of CP is due to an intuitive yet remarkable result about the probabilistic properties of sample quantiles. Let $U_1, \ldots, U_n$ be IID samples of a random scalar variable[1]. For a given level α and after having independentely sampled U_{n+1} from the same distribution, it can be noted that

$$\mathbb{P}\left(U_{n+1} \leq \hat{q}_{1-\alpha}\right) \geq 1 - \alpha \tag{12.3}$$

where $\hat{q}_{1-\alpha}$ is the sample quantile defined as

$$\hat{q}_{1-\alpha} = \begin{cases} U_{(\lceil (n+1)(1-\alpha) \rceil)}, & \text{if } \lceil (n+1)(1-\alpha) \rceil \leq n \\ +\infty & \text{otherwise.} \end{cases} \tag{12.4}$$

where $U_{(1)}, \ldots, U_{(n)}$ is a order statistic of $U_1, \ldots, U_n$. The previous result is valid not only in the univariate setting, where the order statistic is of straightforward definition, but also in the multivariate and functional case, after having defined a suitable notion of ordering. The statement in 12.3 is easy to verify: If $U_1, \ldots, U_n$ are exchangeable, the rank of U_{n+1} among the previous observations will be distributed as a discrete uniform over $\{1, \ldots, n, n+1\}$. As shown by [10], starting from this simple result, one can build with relative ease prediction sets for linear regression models, with the only assumption of the data being IID and that the regression operator being invariant to permutations of the training data.

We now provide an extension of the framework in [10] to the functional case. Let $y \in L^2$ be a new (trial) value. For each y we train the regression estimator $\hat{\mu}_y$ on the dataset $Z_1, \ldots, Z_n, (X_{n+1}, y)$. We then define the functional regression residuals

[1] Please note that the IID assumption is actually too strict, and the whole argument stands also under the weaker notion of exchangeability

$$\begin{aligned} E_{y,i} &= \left(Y_i - \hat{\mu}_y(X_i)\right), \ i = 1, \ldots, n \\ E_{y,n+1} &= \left(y - \hat{\mu}_y(X_{n+1})\right). \end{aligned} \tag{12.5}$$

Please note that, being Y_i and y objects in L^2, also the residuals are actually residual functions, not scalars. For this reason, the ranking step proposed at this point in [10] is now non-trivial, and its choice will have important consequences on the interpretability of the forecast bands provided. Let $\mathcal{R}\left(\left\{E_{y,1}, \ldots, E_{y,n}, E_{y,n+1}\right\}\right) : (L^2[T])^{n+1} \to \mathbb{R}^{n+1}$ be a functional able to provide a scoring criterion among $E_{y,1}, \ldots, E_{y,n}, E_{y,n+1}$, and $\rho(E_{y,i})$ the value of this score for the function $E_{y,i}$. Let us assume that the criterion behaves like a depth measure, assigning high values to points that are deep in the point cloud, and low values to shallow points. We can then build a statistic $\pi(y)$ by computing

$$\pi(y) = \frac{1}{n+1} \sum_{i=1}^{n+1} 1\left\{\rho(E_{y,i}) \leq \rho(E_{y,n+1})\right\} \tag{12.6}$$

where $1\{\cdot\}$ is the indicator function. Essentially, $\pi(y)$ represents the proportion of fitted residuals in the augmented sample that have a lower score than the last one $R_{y,n+1}$. By exchangeability of $Z_1, \ldots, Z_{n+1}$ and the permutational invariance of $\hat{\mu}$ we can see that $\pi(Y_{n+1}$ is distributed as a uniform defined over the set $\left\{\frac{1}{n+1}, \frac{2}{n+1}, \ldots, 1\right\}$. This means that

$$\mathbb{P}\left((n+1)\pi(Y_{n+1}) \leq \lceil (1-\alpha)(n+1) \rceil\right) \geq 1 - \alpha \tag{12.7}$$

which can be alternatively interpreted by saying that $1 - \pi(Y_{n+1})$ is a valid (conservative) $p-$value to test the null hypothesis $H_0 : Y_{n+1} = y$. By calculating such $p-$value over all possible values of $y \in L^2$ and then perform a thresholding to the desired level, we can define in a fairly straightforward way our conformal prediction interval for X_{n+1} in the following way.

$$C_{conformal}(X_{n+1}) = \left\{y \in L^2 : (n+1)\pi(Y_{n+1}) \leq \lceil (1-\alpha)(n+1) \rceil\right\} \tag{12.8}$$

The whole procedure is summarised in Algorithm 1 of [4]. [22] prove that conformal prediction sets defined as in 12.8 have valid finite-sample coverage: in fact, Vovk et al. proofs are valid for very general covariates and responses (i.e. elements of measurable sets), much more general than the relatively well-behaved L^2 functions.

While being theoretically valid, this version of Conformal Prediction is very hard and computationally intensive to be implemented: As noted by [22], the "search" step has to be performed on a regular grid on the set in which the $Y_i, i = 1, \ldots, n$ are embedded. Defining a regular grid in L^2 is a nontrivial task, and will reasonably yield a grid of points of such a size that will render any real-world implementation of the full version of CP non-feasible. For this reason, instead of the Full or Transductive Conformal Prediction framework, we will use the Split [8] or Inductive [15, 22] Conformal Prediction that is able to ease, by a smarter use of the training data, the computational burden associated to CP. The essential idea behind Split Conformal

Prediction is, as the name suggests, random splitting the training data in two samples, a proper training one and a calibration one. We present the Split Conformal prediction version of the method in Algorithm 2 of [4].

12.3 Functional Conformity Measures

To perform a Conformal Prediction task we need to choose a method to order residual curves. While for data lying in $\mathbb{R}$ this issue is straightforward, this is no longer the case even on the Euclidean plane $\mathbb{R}^2$. The most common idea to perform such ordering is the use of data depth, which allows also generalizations of univariate concepts such as medians and quantiles in the multivariate setting ([11] and references therein).

While many choices for a functional conformity or non-conformity measures can be performed, such as the use of small-ball probabilities ([3], specifically Chapter 4 and 13) used as a proxy for density in the L^2 case, we are focusing our attention to depth measures for functional data, and specifically the Band Depth ([12]).

As shown by Lopez-Pintado and Romo in [12] and by [20], the concept of depth in the functional setting can be effectively used to develop functional versions of univariate nonparametric techniques and visualization methods such as rank tests and boxplots. In fact, Band Depth is particularly apt to be used for identifying α-regions (in the sense of [11]) in the shape of bands. and thus very effective in reaching the goal of identifying prediction bands for functional data. Additional information about the choice of band depth can be found in [4]. The concept of band depth can be then effectively used as the ordering criterion in CP: we present the specific version of the algorithm in [4]

We have argued in favour of the validity of set predictions calculated through CP methods: it should also be noted that the bands identified using CP are valid also in a global sense, meaning that its coverage properties are valid for $\forall t \in T$. The proof, along some considerations about the pratical identification of the prediction set are presented in [4]

12.4 Application to IAM Ensemble Forecasting

As a test case for this newly developed method, we present an application to climate change economics, in which we create a statistically valid scenario and policy emulator via the use of functional on scalar linear models and the forecasting methods described in Section 12.3.

Climate Change is by far, the greatest policy challenge the humankind is facing: according to the last Intergovernmental Panel for Climate Change (IPCC) report [1] more decisive actions must be undertaken now, if we want to contain the average increase of the World surface temperature by 1.5°C, and avoid severe disruptions to the earth climate system. A fundamental tool to understand and explore the complex dynamics that regulates this phenomenon is the use of computer models, such as Integrated Assessment Models [14]. By integrating an economic and a climatic module, these models are able to simulate the profile of a variable of interest on a

given timescale (usually CO_2 over the next century). Predicting a quantity for such a long time scale is a notoriously hard task, with a great degree of uncertainty involved. Many efforts have been undertaken to model and control this uncertainty, such as the development of standardized scenarios of future developement, called Shared Socioeconomic Pathways (SSPs) [21, 17] or the use of model ensembles to tackle the issue of model uncertainty. All the details about the application can be found in [4], while an example of prediction bands generated using the described version of CP for functional data can be seen in Figure 12.1

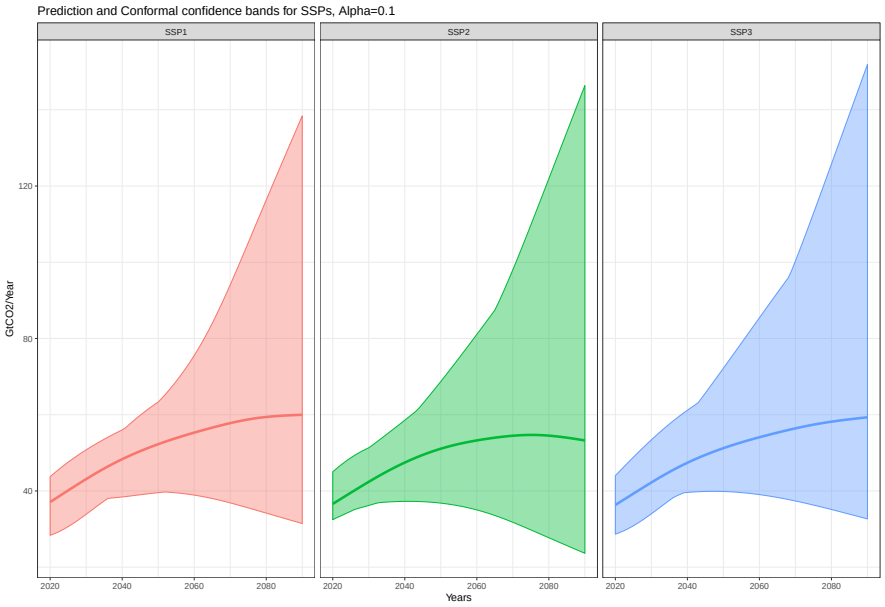

Fig. 12.1 Prediction Bands for SSP1, SSP2 and SSP3, $\alpha = 0.10$

References

[1] IPCC: Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. In Press (2018)

[2] Antoniadis, A., Brossat. X., Cugliari, J., Poggi, J.M.: A prediction interval for a function-valued forecast model: Application to load forecasting. International Journal of Forecasting **32**(3), 939–947 (2016)

[3] Ferraty, F., Vieu, P.: Nonparametric Functional Data Analysis: Theory and Practice. New York, Springer (2007)

[4] Fontana, M., Vantini, S., Tavoni, M., Gammerman, A., Vovk, V.: Distribution-Free Conformal Prediction of Functional Data, with an application to climate economy model interval forecasting. MOX Report, Politecnico di Milano (2019)
[5] Gammerman, A., Vovk, V., Vapnik, V.: Learning by Transduction. In: Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence. UAI'98, pp. 148–155. San Francisco, Morgan Kaufmann Publishers Inc. (1998)
[6] Hron, K., Menafoglio, A., Templ, M., Hrůzová, K., Filzmoser, P.: Simplicial principal component analysis for density functions in Bayes spaces. Computational Statistics & Data Analysis **94**, 330–350 (2016)
[7] Hyndman, R.J., Shahid Ullah, M.: Robust forecasting of mortality and fertility rates: A functional data approach. Computational Statistics & Data Analysis **51**(10), 4942–4956 (2007)
[8] Lei, J., Rinaldo, A., Wasserman, L.: A conformal prediction approach to explore functional data. Annals of Mathematics and Artificial Intelligence **74**(1), 29–43. (2015)
[9] Lei, J., Robins, J., Wasserman, L.: Distribution-Free Prediction Sets. Journal of the American Statistical Association **108**(501), 278–287 (2013)
[10] Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R.J., Wasserman, L.: Distribution-Free Predictive Inference for Regression. Journal of the American Statistical Association **113**(523), 1094–1111 (2018)
[11] Liu, R.Y., Parelius, J.M., Singh, K.: Multivariate analysis by data depth: descriptive statistics, graphics and inference. The Annals of Statistics **27**(3), 783–858 (1999)
[12] López-Pintado, S., Romo, J.: On the Concept of Depth for Functional Data. Journal of the American Statistical Association **104**(486), 718–734 (2009)
[13] Marangoni, G., Tavoni, M., Bosetti, V., Borgonovo, E., Capros, P., Fricko, O., et al.: Sensitivity of projected long-term CO2 emissions across the Shared Socioeconomic Pathways. Nature Climate Change **7**(2), 113–117 (2017)
[14] Nordhaus, W.D.: Rolling the 'DICE': an optimal transition path for controlling greenhouse gases, Resource and Energy Economics. **15**(1), 27–50 (1993)
[15] Papadopoulos, H., Proedrou, K., Vovk, V., Gammerman, A.: Inductive Confidence Machines for Regression. In: Machine Learning: ECML 2002. Lecture Notes in Computer Science, pp. 345–356. Springer, Berlin, Heidelberg (2002)
[16] Ramsay, J.O., Silverman, B.W.: Functional Data Analysis. 2nd ed. Springer series in statistics. New York (2005)
[17] Riahi, K. et al.: The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. Global Environmental Change **41**, 153–168 (2017)
[18] Sangalli, L.M., Secchi, P., Vantini, S., Veneziani, A.: A Case Study in Exploratory Functional Data Analysis: Geometrical Features of the Internal Carotid Artery. Journal of the American Statistical Association **104**(485), 37–48 (2009)

[19] Shafer, G., Vovk, V.: A Tutorial on Conformal Prediction. J Mach Learn Res **9**, 371–421 (2008)
[20] Sun, Y., Genton, M.G.: Functional Boxplots. Journal of Computational and Graphical Statistics **20**(2), 316–334 (2011)
[21] van Vuuren, D.P., Edmonds, J., Kainuma, M. et al.: The representative concentration pathways: an overview. Climatic Change **109**, 5 (2011)
[22] Vovk, V., Gammerman, A., Shafer, G.: Algorithmic learning in a random world. New York, Springer (2005)