Quantile-adaptive probabilistic forecast combining

Ke Yu[a,*], James W. Taylor[a], Xiaochun Meng[b]

[a] *Saïd Business School, University of Oxford, Park End Street, Oxford, OX1 1HP, UK*
[b] *School of Management, University of Bath, Claverton Down, Bath, BA2 7AY, UK*

Abstract

Combining forecasts of cumulative probability distributions (CDFs) allows aggregation of the available information to improve accuracy. The linear opinion pool is commonly used, but it can yield overdispersed distributional forecasts. An alternative, leading to lower dispersion, is to average the quantiles of the CDF, which can be viewed as horizontal CDF averaging, with the averaging of probabilities in the linear opinion pool viewed as vertical averaging. Empirical results show that horizontal and vertical averaging can each be preferable for different parts of the CDF. For example, one method might be better for tail quantiles, while the other is better for central quantiles. To address this, we develop a method that transitions between vertical and horizontal averaging across the CDF. It relates to angular averaging, which is a recent proposal that performs aggregation along lines at an angle. Our new method averages along lines with slopes that smoothly transition across the CDF. The method is quantile-adaptive in the sense that the slopes of the lines vary across the quantiles, or equivalently, across the probabilities. We set the lines to emanate from a small number of fixed points, which are the parameters of the method. Viewing the lines as rays, we term the method *radial averaging*. Our theoretical results show that the method has the versatility to generate CDF forecasts that are sharper than horizontal averaging, and less sharp than vertical averaging. Our empirical results provide support for the new approach.

Keywords: decision analysis, combining forecasts, probability distributions, quantiles

1. Introduction

The combination of individual forecasts has played a crucial role in the development of forecasting practice since the seminal work of Reid (1968) and Bates and Granger (1969), who demonstrated that combining forecasts often produces more accurate predictions than individual forecasts alone. Timmermann (2006) and Lichtendahl and Winkler (2020) study theoretically the factors that determine the advantages of point forecast combinations, such as the correlation between forecast errors and the relative size of the forecast error variances. Empirical evidence from diverse fields, such as economics and finance (Stock & Watson, 1999, 2004; Safari & Davallou, 2018; Ziel & Weron, 2018), meteorology (Liu et al., 2020), and epidemiology (Nikolopoulos et al., 2021), has illustrated in a compelling way the efficacy of forecast combination. McAndrew et al. (2021) and X. Wang et al. (2023) provide recent reviews of the extensive literature on forecast combinations.

While early research focused primarily on combining point forecasts, increasing recognition of the importance of quantifying uncertainty to support decision-making has led to interest in probabilistic forecasting. Probabilistic forecasts, which provide full predictive distributions rather than single point forecasts, offer richer information about future uncertainties. Our interest in this paper is in aggregating forecasts of probability distributions. The linear opinion

*Corresponding author
Email addresses: ke.yusdu@hotmail.com (Ke Yu), james.taylor@sbs.ox.ac.uk (James W. Taylor), xiaochun.meng@bath.edu (Xiaochun Meng)

pool (Stone, 1961) is the most common approach to aggregating the probabilistic judgements of multiple experts, but while the linear opinion pool offers simplicity and ease of implementation, it faces limitations. One significant issue is that dispersion tends to increase under linear pooling. Other forms of combinations include nonlinear pools (Gneiting & Ranjan, 2013; Bassetti et al., 2018), which transform individual forecasts before combining, and trimmed opinion pools (Jose et al., 2014; Grushka-Cockayne et al., 2017), which provide a flexible way to combine forecasts in the sense that the dispersion can be controlled by trimming forecasts.

As an alternative to the averaging of probabilities in the linear opinion pool, Lichtendahl et al. (2013) study the averaging of quantiles, which leads to a sharper, and potentially better calibrated, probabilistic forecast. This can be viewed as horizontal averaging of the forecasts of the cumulative distribution functions (CDF), with the linear opinion pool viewed as vertical averaging.

Our empirical results show that horizontal and vertical averaging can each be preferable for different parts of the CDF. For example, horizontal averaging might be better for tail quantiles, with vertical averaging better for central quantiles. A further empirical finding is that the variation in the relative performance of the two approaches is continuous across the CDF. To address these issues, we develop a method that relates to angular averaging (Taylor & Meng, 2025), which is a recent proposal that performs aggregation along lines at an angle. Our new method averages along lines with slopes that smoothly transition across the CDF. An appealing property of our new proposal is that it has the versatility to yield CDF forecasts that are sharper or less sharp than those produced by vertical and horizontal averaging.

Implementation of the new approach involves averaging the pool of CDF forecasts along lines with different slopes that are determined by a finite number of points. We refer to these as *focal points*, and with the lines radiating as *rays* from the focal points, we call our new method *radial averaging*. With different positioning and numbers of the focal points, radial averaging encompasses many forms. We demonstrate several of these forms, and show the appealing properties of them in certain scenarios. Vertical, horizontal and angular averaging are all special cases of radial averaging, and thus radial averaging offers a high degree of flexibility for different data. Among our theoretical results, we show that the output of the new method is indeed a CDF. In our empirical study with Covid mortality CDF forecasts, we present a computationally efficient way to implement radial averaging. A method for choosing the focal points is proposed. We remark that the focus in this paper is on equally weighted averaging, as this allows us to compare more clearly vertical, horizontal, angular and radial methods. However, it is straightforward to extend our proposed radial averaging to the performance-based weighted averaging (see, for instance, X. Wang et al., 2023), inventory-cost-based weighted averaging (see, for instance, S. Wang et al., 2024) or decision-cost-based weighted averaging (see, for instance, Stratigakos et al., 2024). Furthermore, our proposed method can be integrated with other

existing methods such as the trimming-based method of Jose et al. (2014).

The rest of the paper is organized as follows. Section 2 reviews vertical, horizontal and angular averaging. In Section 3, we present our radial averaging method. Section 4 presents theoretical properties of the method. In Section 5, we refine angular combining by improving its implementation and parameter optimization, and describe the implementation and parameter optimization for radial averaging. An empirical study of forecasting Covid mortality is presented in Section 6.

2. A brief introduction to vertical, horizontal and angular averaging

Throughout this paper, we consider n forecasters, each providing a continuous CDF F_i, $i \in \{1, \ldots, n\}$, representing their belief about the probability distribution of a continuous random variable X. The goal is to average the CDFs F_i into a single aggregated CDF.

Taking the simple arithmetic mean of the CDFs, i.e., $\hat{F}_v(x) = (F_1(x) + \cdots + F_n(x))/n$, is a straightforward and widely used approach for CDF forecast combining. The probabilities are averaged across forecasters for a given outcome x, and thus the aggregation is performed in a vertical sense. Hence, this approach can be termed *vertical averaging*. Alternatively, one can consider averaging the quantiles, i.e., $\hat{F}_h^{-1}(\tau) = (F_1^{-1}(\tau) + \cdots + F_n^{-1}(\tau))/n$, for a given probability $\tau \in [0, 1]$. This can be viewed as *horizontal averaging*. Among the theoretical results presented by Lichtendahl et al. (2013), they show that horizontal averaging produces a sharper CDF than vertical averaging. Recently, Taylor and Meng (2025) propose a new approach, termed *angular averaging*, which generalizes vertical and horizontal averaging by allowing the averaging to be performed at an angle lying between the horizontal and vertical case. A more detailed review of angular averaging is provided in Section S3 of the Supplementary Material.

In Figure 1, we illustrate these three forms of CDF forecast averaging for $n = 2$ individual CDFs that are truncated normal distributions with means $\mu_1 = 3$ and $\mu_2 = 3.5$, and standard deviations $\sigma_1 = \sigma_2 = 0.1$, and are constrained to the interval $[2.5, 4]$. For angular combining, the averaging is performed along straight lines at an angle θ to the horizontal, as shown in the right panel of Figure 1. The angle is restricted to lie between 0 and 90°, implying that the angled straight lines are vertical or have a non-positive slope, which ensures that there is only one point of intersection of each of these lines with the graph of each CDF F_i.

3. Radial averaging

The relative performance of horizontal and vertical averaging has been found to vary across the quantiles of the CDF, as shown by Taylor and Meng (2025), and as we find in our empirical study of Section 6. Furthermore, the relative performance of these two averaging approaches varies gradually across the quantiles. This motivates us to propose a method that allows the averaging to be performed at an angle that evolves smoothly across the quantiles. This amounts

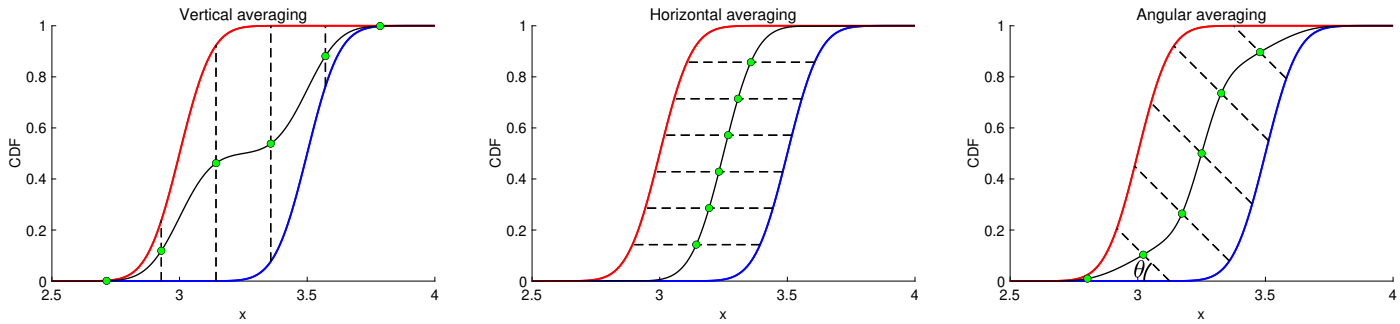


Figure 1: An illustration of vertical, horizontal and angular averaging. F_1 and F_2 are shown as the red and blue curves. F_1 and F_2 are averaged along the dashed line segments, and the green points are the midpoints of the line segments. The resulting CDF forecasts are shown as the black curves.

to smooth transitioning between horizontal and vertical averaging.

In Section 3.1, we consider the case of just one transition. This can either be a transition from horizontal in the left tail to vertical in the right tail, or from vertical in the left tail to horizontal in the right. We use the notation H-V and V-H to indicate these two averaging patterns, and use similar abbreviated notations for other averaging patterns. Section 3.2 considers the case of two transitions, with the two possibilities being either H-V-H or V-H-V. More than two transitions and a general definition of radial averaging are provided in Sections 3.3 and 3.4, respectively. Throughout this section, we assume that the support of every random variable whose CDF is F_i is a compact interval, and illustrate our ideas using the same two truncated normal distributions that we considered in Figure 1. This assumption is consistent with practical applications, because many real-world variables are naturally bounded above and below. We will also discuss the case of an unbounded support in Section 3.3.

3.1. One transition between vertical and horizontal averaging

In Figure 2, we present the two forms of radial averaging that have one transition. The left panel of the figure shows H-V radial averaging, whereby averaging is performed along a ray that is horizontal for the extreme left tail (i.e., the lowest probability levels), and transitions smoothly to become vertical for the extreme right tail (i.e., the highest probability levels). The right panel of Figure 2 illustrates V-H radial averaging, which involves a smooth transition from vertical averaging for the left tail to horizontal averaging for the right tail.

The two panels each show averaging performed along dashed rays emanating from a single point. As we have indicated in Figure 2, we refer to this as a *focal point*. We use the term *rays* for the dashed rays radiating from the focal point, which has led us to call the method *radial averaging*. In each panel, the midpoints of the black dashed line segments are indicated by green points, producing the radial average CDF, denoted by $\hat{F}_{r,o}$, where the subscript r stands for radial averaging, and o represents the focal point. More specifically, with rays radiating from the focal point o, at varying angle θ to the x-axis, the radial average forecast is given by $\hat{F}_{r,o}\left(\left(x_1(\theta)+x_2(\theta)\right)/2\right) = \left(F_1(x_1(\theta))+F_2(x_2(\theta))\right)/2$, where $x_i(\theta)$ is the x-coordinate of the

intersection point of a ray with the graph of the i-th individual CDF.

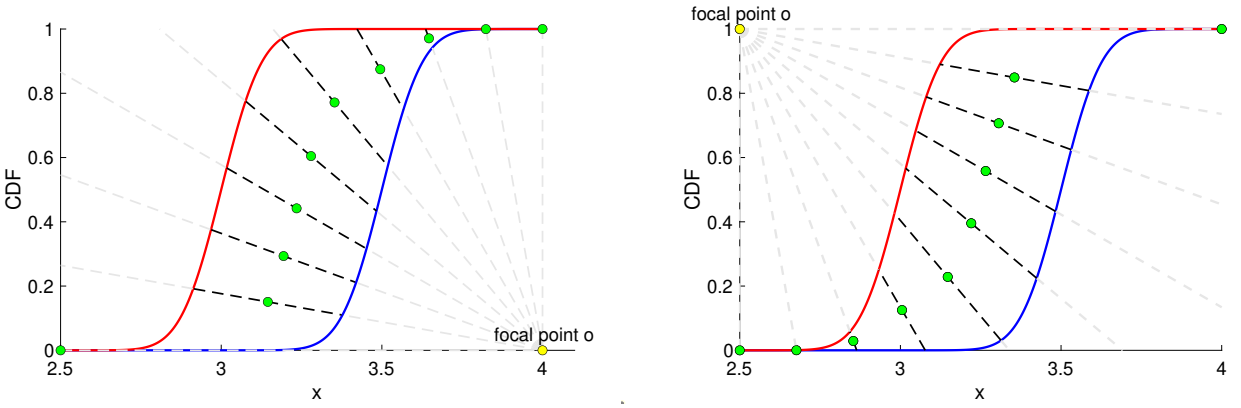


Figure 2: In the left panel, illustration of H-V radial averaging with the focal point $o = (4, 0)$, and in the right panel, V-H radial averaging with the focal point $o = (2.5, 1)$. The CDFs are averaged along the dashed rays. The green points are the midpoints of all black dashed line segments, and they are on the graph $\{(x, \hat{F}_{r,o}(x)) \mid x \in \mathbb{R}\}$ of the radial average forecast.

The essential feature of our radial averaging is that it transitions between vertical and horizontal averaging. The number of transitions is controlled by focal points, that is, the number of focal points equates to the number of transitions. When we use one focal point, we impose one transition, as we have shown in Figure 2. In Sections 3.2 and 3.3, we show how two and three focal points enable two and three transitions, respectively.

We remark that this type of averaging with one focal point is useful when the individual CDFs have an overly long left tail and an overly short right tail, or when they have an overly long right tail and an overly short left tail. Moreover, when the individual densities are symmetric about the mean, the density associated with the vertical, horizontal or angular average is also symmetric. On the contrary, the density associated with the radial average presented in Figure 2 is asymmetric, which demonstrates that our radial averaging can be advantageous in predicting random variables with asymmetric features (see, for instance, Hua & Zhang, 2008).

3.2. Two transitions between vertical and horizontal averaging

To enable two transitions, we use two focal points for radial averaging. Consider, for instance, a pair of focal points $o := (o_x, o_y)$ and $o' := (o'_x, o'_y)$ given by $o_x = o'_x = 3.25$, and $o_y = 0$, $o'_y = 1$, illustrated in the left panel of Figure 3. In this case, radial averaging smoothly moves from horizontal averaging in the left tail to vertical averaging around the center, and then to horizontal averaging in the right tail. We refer to this as H-V-H radial averaging. The radial average forecast is given by

$$\hat{F}_{r,o,o'}(x) := \begin{cases} \hat{F}_{r,o}(x), & x \le o_x, \\ \hat{F}_{r,o'}(x), & x > o'_x, \end{cases} \tag{1}$$

where $\hat{F}_{r,o}$ is the radial average CDF based on focal point o, as discussed in Section 3.1.

As a natural alternative, radial averaging can also be performed with the averaging moving

from vertical averaging in the left tail to horizontal averaging around the central quantiles, and then to vertical averaging in the right tail. We refer to this case as V-H-V radial averaging. The right panel of Figure 3 displays an example of such radial averaging with two focal points $o = (2.5, 0.5)$ and $o' = (4, 0.5)$. The formula of the resulting CDF $\hat{F}_{r,o,o'}$ is similar to that of the previous case.

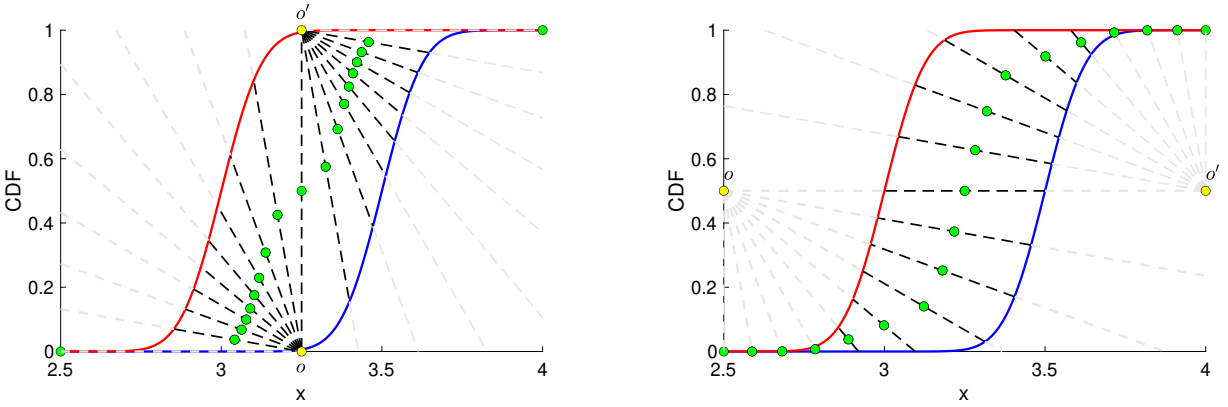


Figure 3: Two examples of radial averaging with two focal points. The left panel shows an example of H-V-H radial averaging, and the right panel shows an example of V-H-V radial averaging.

Radial averaging with two focal points can reduce to radial averaging with one focal point. Denote $a_i := \max\{x \mid F_i(x) = 0\}$, $b_i := \min\{x \mid F_i(x) = 1\}$. In this example, $a_1 = a_2 = 2.5$, $b_1 = b_2 = 4$. For the left panel of Figure 3, when $o_x = o'_x = \min_i a_i$, we obtain the one-transition case V-H shown by the right panel of Figure 2, and when $o_x = o'_x = \max_i b_i$, we obtain the one-transition case H-V shown by the left panel of Figure 2. Similarly, moving the focal points on the right panel of Figure 3 to the lines $y = 0$ or $y = 1$ can also lead to radial averaging with one focal point. (By convention, we say that the vertical axis is the y-axis.)

A potential advantage of horizontal averaging over vertical averaging is that the variance of the random variable obtained from horizontal averaging is lower, implying that when individual CDFs are overdispersed, horizontal averaging can improve the calibration. Similarly, vertical averaging has the potential advantage over horizontal averaging that it delivers a less sharp CDF, which is useful when aggregating underdispersed CDF forecasts. One may guess that the sharpness of radial averaging is between vertical and horizontal averaging, because radial averaging involves both of them. Interestingly, with suitable positioning of the focal points, H-V-H radial averaging can produce a sharper CDF than horizontal averaging, while V-H-V radial averaging can produce a CDF that is not as sharp as vertical averaging. We present and prove this more formally in Section 4.3.

3.3. More transitions between vertical and horizontal averaging

Having established radial averaging with one and two transitions, we can naturally extend the framework to involve more transitions. The construction is very similar to that in Sections 3.1 and 3.2, and thus in this subsection we only briefly illustrate a case with three transitions, with

a focus on the demonstration of how it encompasses the cases presented in the previous two subsections.

Figure 4(a) illustrates an example of radial averaging with three focal points o, o' and o''. In this example, as in the previous two subsections, we let one focal point o be on the $x = \min_i a_i$ line and another focal point o'' on the $y = 1$ line. For the remaining focal point o', we set its x-coordinate to be the same as that of o'', and y-coordinate to be the same as that of o. As a consequence, once the focal point o' is determined, the other two focal points are determined.

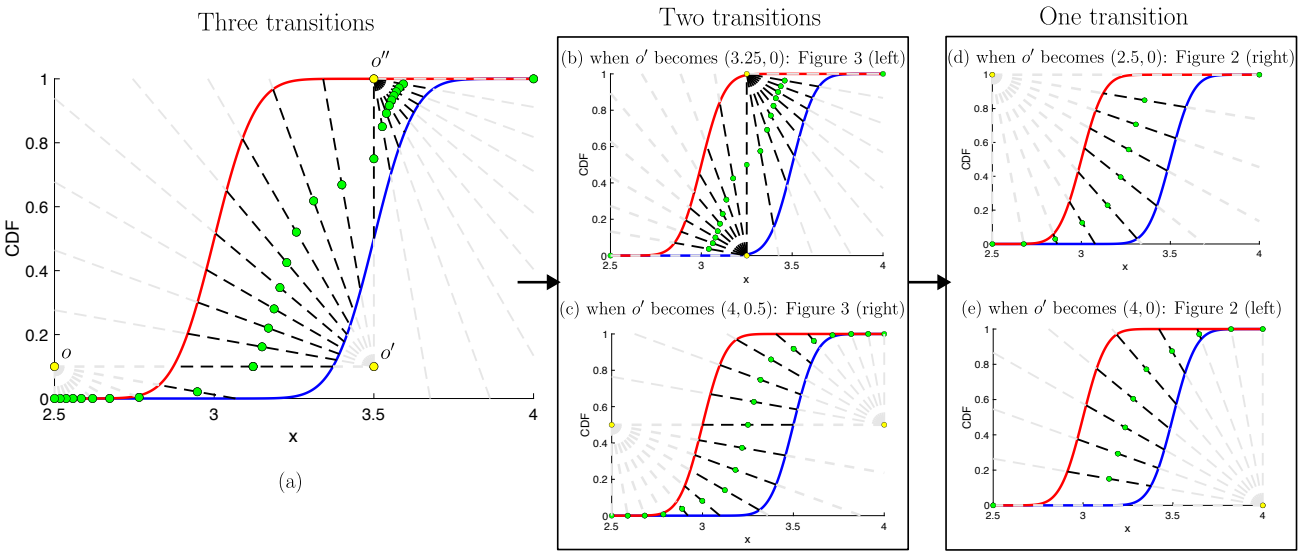


Figure 4: Illustration of radial averaging with three focal points (panel (a)) and the cases where it becomes radial averaging with two focal points (panels (b) and (c)) and one focal point (panels (d) and (e)) described in the previous two subsections.

The radial averaging with three focal points effectively encompasses the cases of one and two focal points in Sections 3.1 and 3.2, respectively. This can be seen by noting that if the focal point o' is moved from $(3.5, 0.1)$ in Figure 4(a) to $(3.25, 0)$, we obtain Figure 4(b), which is the two-focal-point case shown in the left panel of Figure 3. Also, if instead the focal point o' is moved from $(3.5, 0.1)$ in Figure 4(a) to $(4, 0.5)$, we get Figure 4(c), which is the same case of two focal points in the right panel of Figure 3. In addition, if o' is moved from $(3.5, 0.1)$ in Figure 4(a) to $(2.5, 0)$ or $(4, 0)$, we get the two cases of one focal point in Figures 4(d) and 4(e), respectively, which are the same as the cases we considered in Figure 2.

3.4. Other possible forms and definition of radial averaging

Having presented the illustrative examples in Sections 3.1, 3.2 and 3.3, we shall end Section 3 with a more general and comprehensive definition of radial averaging, showing the diverse possibilities that the method allows.

In all the examples we have shown so far, radial averaging performs horizontal or vertical averaging for the extreme left tail and extreme right tail. Radial averaging also allows one to consider other possibilities. For simplicity, we start the illustration of these possibilities with

only one focal point. The first possibility is that the extreme left (and/or right) tail of the CDFs may be averaged along a ray with a negative slope that is not horizontal or vertical, as shown in the left panel of Figure 5. This happens if the focal point lies outside the rectangle formed by the lines $y = 0$, $y = 1$, $x = \min_i a_i$ and $x = \max_i b_i$. We will hereafter refer to it as *the rectangle*. When the focal point in the left panel is moved toward the bottom-right or top-left corner until moving to a position that is infinitely far away from the graphs of the CDFs, the slopes of all rays converge to be identical. This case of radial averaging becomes angular averaging. Similarly, when the focal point converges to $(+\infty, 0)$ or $(4, -\infty)$, radial averaging converges to horizontal or vertical averaging, respectively. A formal statement of this is provided in Proposition 1 in Section 4.2.

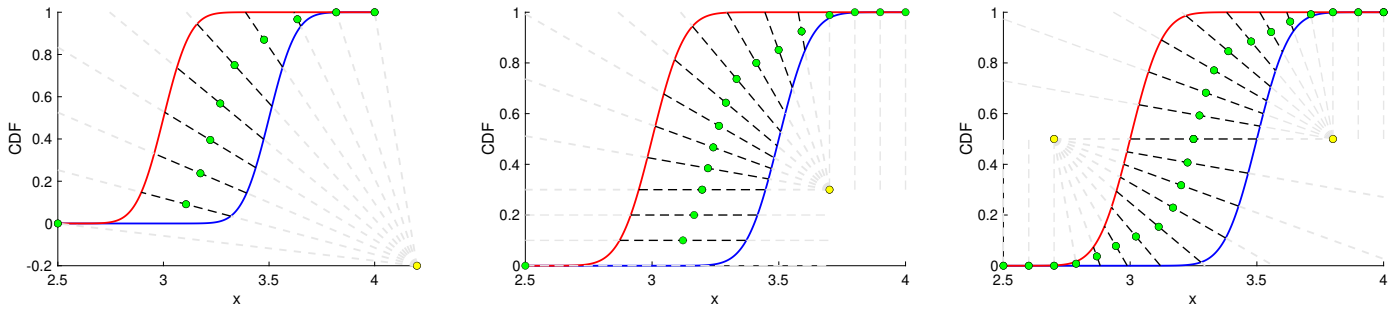


Figure 5: Other possible examples of radial averaging. Left: radial averaging that does not include horizontal and vertical averaging. Middle: radial averaging with vertical and horizontal averaging being performed for an interval of x and an interval of probabilities. Right: V-H-V for random variables with unbounded supports.

For the one focal point case, another possibility is that for the left and right tails, vertical averaging is performed for an interval of x, and horizontal averaging is performed for an interval of probabilities, as illustrated in the middle panel of Figure 5. This happens if the focal point lies inside the rectangle. Furthermore, we emphasize that the importance of this possibility is also that it allows the application of radial averaging to random variables with unbounded supports. The right panel of Figure 5 illustrates how V-H-V in the right panel of Figure 3 can be achieved when the random variable has unbounded support, which is considered in Theorem 3 in Section 4.3.

We end this section by providing a general definition of radial averaging that unifies all the possibilities. Radial averaging, in essence, is a method of CDF aggregation where the CDFs are averaged along rays that are uniquely determined by a set of focal points. Connecting these focal points forms a point, a line segment or a staircase (e.g., ⌟⌜) joining a few vertical and horizontal line segments. Every line segment intersects the graphs of all CDFs. The rays emanate from the focal points and are vertical or non-positively sloped. These ensure that radial averaging produces a CDF; details for this are provided in Section 4.1. Slightly more extra care is needed, as in Figure 5, when handling the tails if the focal point at the bottom-left or (and) the focal point at the top-right is (are) inside or outside the rectangle.

4. Theoretical results

In this section, we provide theoretical results to support our proposed new approach. Section 4.1 shows that radial averaging produces a non-decreasing, right-continuous function, the value of which approaches 0 as $x \to -\infty$ and approaches 1 as $x \to \infty$, i.e., a CDF. The convergence of the PDF and CDF produced by radial averaging to those of other methods is the focus of Section 4.2. Section 4.3 gives insight into the sharpness of the CDF resulting from radial averaging. Proofs of the main theoretical results are presented in Appendix A, while the proof of Proposition 1 is deferred to the Supplementary Material.

4.1. Ensuring the average is a CDF

The validity of all probabilistic forecast averaging methods discussed in this paper, i.e., vertical, horizontal, angular and radial averaging, is guaranteed by the following fundamental theorem, which provides a sufficient condition under which the average forecast is a CDF.

Theorem 1. *Let L be a set of straight lines which are vertical or have non-positive slopes, and the individual CDFs are averaged along them. Denote by $\mathcal{C}$ the set of all points of intersection of pairs of lines in L. If for all $(x, y) \in \mathcal{C}$, we have that $y \geq \max_i F_i(x)$, or $y \leq \min_i F_i(x)$, and every point on the graph of F_i is on a line in L, then the averaging yields a CDF.*

To better illustrate the theorem, we call the interior of the region enclosed by all CDFs, i.e., $\{(x, y) \in \mathbb{R}^2 \,|\, \min_i F_i(x) < y < \max_i F_i(x)\}$, the *inadmissible region*, and refer to its complement as the *admissible region*, as illustrated in Figure 6. The theorem indicates that the lines along which the CDFs are averaged not intersecting in the inadmissible region is a sufficient condition for producing a CDF.

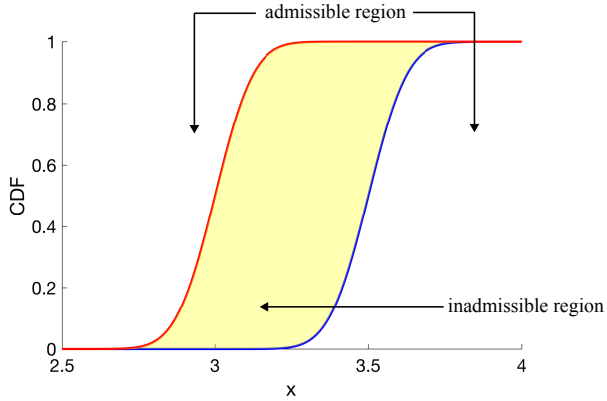


Figure 6: Illustration of the admissible and inadmissible regions for the points of intersection of the straight lines.

There are various possibilities for such lines. It is easy to see that those used by vertical, horizontal and angular averaging satisfy the requirement, because they do not even have any points of intersection. By the definition of radial averaging provided in Section 3.4, the rays also do not intersect in the inadmissible region, since the focal points do not lie in the inadmissible

region.

Furthermore, the key idea of angular averaging and radial averaging is to express all the straight lines or rays with one or a few parameters. For angular averaging, the parameter is the angle; for radial averaging, the parameters are the focal points. Once we know the parameter(s), all lines or rays are uniquely determined. In turn, the average forecast is uniquely determined. This parametrization enables theoretical analysis and practical implementation. The following result follows immediately.

Corollary 1. *Given a set of focal points, radial averaging produces a unique CDF.*

4.2. Convergence results of the PDF and CDF

Although vertical, horizontal, and angular averaging are intuitively limiting cases of radial averaging, in this section, we formally prove that both the CDF $\hat{F}_{r,o}$ and PDF $\hat{f}_{r,o}$ converge to these cases when the focal point is moved infinitely far away in specific directions. To obtain the results, consider the rays radiating from the focal point $o = (o_x, o_y)$ with the angle θ, as shown in Figure 7, ranging from 0 to $\pi/2$. The rays and F_i lead to a function x_i^o of $\theta \in [0, \pi/2]$, as shown in the figure. In the figure and in the rest of the section, the superscript o is omitted without ambiguity. Then, we have that $\hat{F}_{r,o}\left(\left(x_1(\theta) + x_2(\theta)\right)/2\right) = \left(F_1(x_1(\theta)) + F_2(x_2(\theta))\right)/2$.

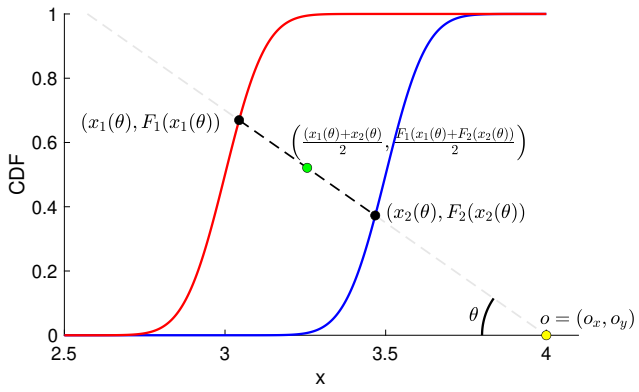


Figure 7: The expression of points of intersection and midpoints using θ.

For $\theta \in (0, \pi/2]$, $x_i(\theta)$ is the x-coordinate of the point of intersection of the ray with angle θ and the graph of F_i. Recall that F_i's are assumed to be continuous. Therefore, each ray with a negative slope and the graph of each F_i intersect at a unique point. This means that the functions x_i's are single-valued on $(0, \pi/2]$. However, a horizontal ray may have no point of intersection or infinitely many points of intersection with the graph of F_i. If F_i is strictly increasing on $\mathbb{R}$, the horizontal ray on the x-axis has no point of intersection with the graph of F_i. If F_i is not strictly increasing, i.e., given the focal point $o = (o_x, o_y)$, with $o_y \in [0, 1]$, if $F_i(x) = o_y$ for all $x \in [z_1, z_2]$, where $z_1 < z_2$, then the horizontal ray emanating from o intersects the graph of F_i at infinitely many points. We define $x_i(0)$ by $x_i(0) := \lim_{\theta \to 0} x_i(\theta)$. Therefore, the function $x_i : [0, \pi/2] \to \mathbb{R} \cup \{\pm\infty\}$ is single-valued.

Proposition 1. *Suppose every CDF admits a continuous probability density function f_i. Then* $\hat{f}_{r,o}\left(\frac{1}{n}\sum_{i=1}^n x_i(\theta)\right) = \left(\sum_{i=1}^n \frac{f_i(x_i(\theta))(F_i(x_i(\theta))-o_y)}{\tan\theta+f_i(x_i(\theta))}\right)\Big/\left(\sum_{i=1}^n \frac{F_i(x_i(\theta))-o_y}{\tan\theta+f_i(x_i(\theta))}\right)$. *When $o_x \to \infty, o_y \to -\infty$ such that $o_y/o_x \to c_o < 0$, or $o_x \to -\infty, o_y \to \infty$ such that $o_y/o_x \to c_o < 0$, we have that $\hat{f}_{r,o}$ converges pointwise to the density function of the angular average with angle* $\arctan(-c_o)$. *When $o_x \to \pm\infty$ such that $o_y/o_x \to 0$, $\hat{f}_{r,o}$ converges pointwise to the density function of the horizontal average. When $o_y \to \pm\infty$ such that $o_x/o_y \to 0$, $\hat{f}_{r,o}$ converges pointwise to the density function of the vertical average. Moreover, the convergence also holds for the CDFs.*

We display in Figure S1 of the Supplementary Material several examples of the resulting CDF and PDF corresponding to different focal points.

4.3. Sharpness

Probabilistic forecasts seek to approximate the (unknown) distribution of the random variable of interest. This distribution reflects the underlying uncertainty inherent in the data generating process. Its sharpness may vary: it will be sharp when future outcomes are relatively predictable, or dispersed when uncertainty is high. Therefore, it is useful for a combining method to have the flexibility to generate combined forecasts with different levels of dispersion. As established in the following two theorems, radial averaging offers this flexibility in sharpness: radial average forecasts can be sharper than the horizontal average (Theorem 2), and less sharp than the vertical average (Theorem 3).

Theorem 2. *Let $F_1, \ldots, F_n$ be the CDFs of symmetric and log-concave distributions from the same location-scale family, and have the same mean μ and different scale parameters. Then (i) $\hat{F}_h^{-1}(p) < \hat{F}^{-1}_{r,(\mu,0),(\mu,1)}(p)$, for $0 < p < 1/2$; $\hat{F}_h^{-1}(p) > \hat{F}^{-1}_{r,(\mu,0),(\mu,1)}(p)$, for $1/2 < p < 1$; and $\hat{F}_h^{-1}(p) = \hat{F}^{-1}_{r,(\mu,0),(\mu,1)}(p)$, for $p = 1/2$. (ii) The radial average with focal points $(\mu, 0)$ and $(\mu, 1)$ are sharper than the horizontal average.*

Well-known examples of log-concave distributions include uniform, normal, logistic and Laplace distributions. The following result does not require that the distributions are log-concave or from the same location-scale family, but explicitly assumes the specific ordering of the densities around the mean. Any set of individual CDFs that satisfies the assumptions in Theorem 2 also satisfies the milder assumptions of Theorem 3.

Theorem 3. *Let $F_1, \ldots, F_n$ be CDFs with the same mean μ, finite variances, and symmetric density functions $f_1, \ldots, f_n$. Assume that every density function is non-decreasing on $(-\infty, \mu)$. Assume that there exists $D > 0$ such that for all $d \in (0, D)$, $f_1(x) > \cdots > f_n(x)$ for all $x \in [-d+\mu, d+\mu]$. Then for any $d \in (0, D)$, letting the focal points o and o' be $(-d+\mu, 1/2)$ and $(\mu+d, 1/2)$, respectively, we have (i) $\hat{F}_v(x) < \hat{F}_{r,o,o'}(x)$ for $x \in (-d+\mu, \mu)$; $\hat{F}_v(x) > \hat{F}_{r,o,o'}(x)$ for $x \in (\mu, \mu+d)$; and $\hat{F}_v(x) = \hat{F}_{r,o,o'}(x)$ for $x \in \{\mu\} \cup (-\infty, -d+\mu] \cup [d+\mu, +\infty)$. (ii) The radial average with focal points o and o' is less sharp than the vertical average.*

In this section, we use the variance as a measure of sharpness. This choice follows the settings of Lichtendahl et al. (2013) and Gneiting et al. (2007), where predictive distributions are assumed to be continuous with finite second moments and variance is used to summarize dispersion. For heavy-tailed or strongly asymmetric distributions where variance may not exist or may not reflect level concentration, alternative notions of sharpness may be considered, such as the width of central prediction intervals.

To conclude this section, we provide an alternative perspective on the proposed method. Although radial averaging is defined by averaging along rays with varying slopes, the radial average forecast can be obtained by only averaging vertically, or only averaging horizontally, through the use of a transformation. This is closely related in spirit to generalized linear combining, which was introduced by Dawid et al. (1995) and developed by Ranjan and Gneiting (2010) and Gneiting and Ranjan (2013). In a generalized linear pool, each CDF forecast is first transformed by a monotone link function applied to its CDF values, and then the transformed values are averaged across forecasts. Radial averaging can also be viewed as a linear pooling scheme carried out after a common transformation of the individual CDFs: each CDF is transformed by an operator, the transformed functions are combined by vertical or horizontal averaging, and finally the inverse operator is applied to obtain the CDF forecast. We provide details of this in Section S1.3 of the Supplementary Material.

5. Implementation and parameter optimization for angular and radial averaging

In this section, we describe improvements to the approach proposed by Taylor and Meng (2025) for implementing angular averaging, and optimizing its parameter. We then use this as a basis for describing the implementation and parameter optimization for radial averaging.

5.1. Improved implementation of angular averaging

Angular averaging cannot be implemented using closed-form expressions. We follow Taylor and Meng (2025) by using a pragmatic numerical approach that involves computing the points of intersection of the CDFs with a finite number of angled lines, followed by interpolation of these points. Although our implementation approach for angular averaging is similar to the method of Taylor and Meng (2025), our approach is computationally more efficient. In this section, we first describe their implementation approach, and then present our approach.

Denote by a the minimum of the lower bounds of the distributional forecasts produced by individual forecasting teams, and by b the maximum of the upper bounds of the distributional forecasts. Taylor and Meng (2025) implement angular averaging by using the 1001 equally-spaced points on the line segment joining the two points $(a, 0)$ and $(b, 1)$. For illustration, several such points are shown as the red stars in the left panel of Figure 8, which shows a set of 26 CDF forecasts for mortality from our empirical analysis of Section 6. Taylor and Meng

(2025) then consider 1001 angled lines, each of which passes through one of the points, with the given slope, and average the pool of CDF forecasts along each of 1001 angled lines. We refer to this approach as *Implementation Method 1*.

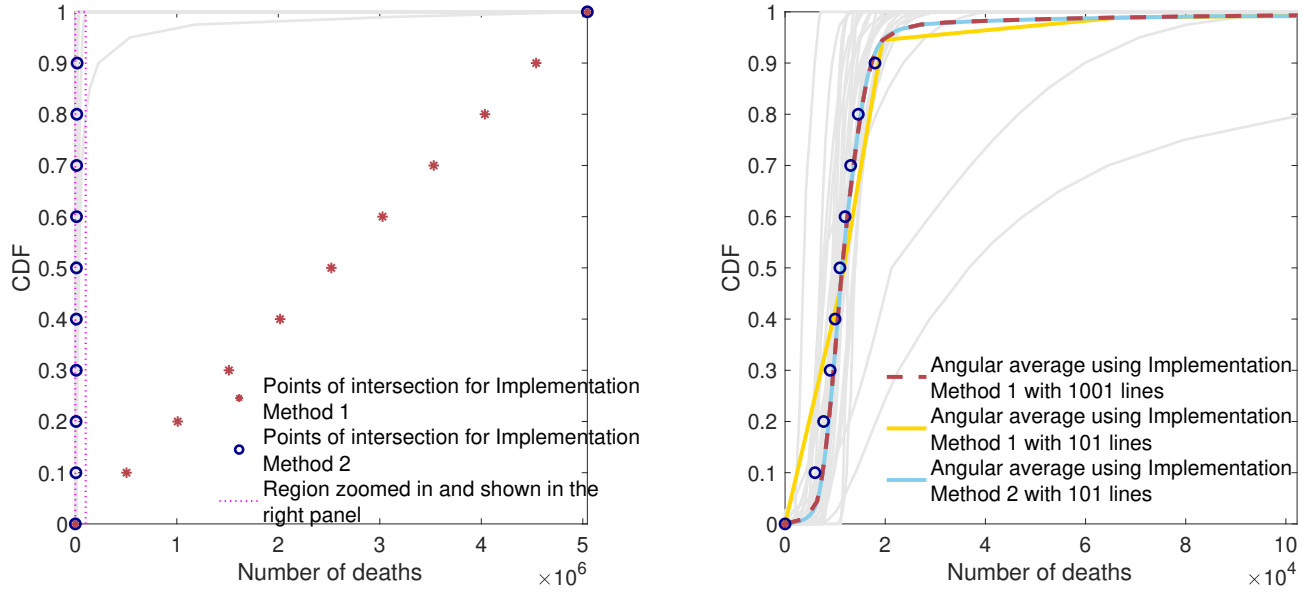


Figure 8: Illustration of two methods to implement radial averaging using CDF forecasts from our empirical study of Section 6. The left panel presents the 26 4-week ahead individual CDF forecasts for the U.S. national series in week 69. Implementation Method 1 uses the red starred points to compute the angular average, while Implementation Method 2 uses the dark blue points. The right panel is a zoomed-in view of the narrow area marked by the dotted rectangle in the left panel. The difference between the two methods is illustrated in the right panel, which shows the angular average CDFs with $\theta = 89°$.

Observe that, in our empirical study, forecasting teams occasionally submitted outlying CDF forecasts possibly due to software errors, record errors or incorrect model assumptions. For example, consider the 26 CDF forecasts in the left panel of Figure 8. All the CDFs, except one, lie within the dotted rectangle, whereas one outlying CDF stretches much further to the right of the plot. The implication of this is that, when implementing angular averaging with a large angle, the vast majority of the 1001 angled lines will intersect 25 of the 26 CDFs at points on the $y = 1$ line. This means that nearly all the computations relate only to the extreme upper quantiles.

We propose an alternative approach, referred to as *Implementation Method 2*, which provides implementation of angular averaging with substantially less computational time. In this method, we replace the 1001 points in Implementation Method 1 by the following 101 points: $(q^v_{j/100}, j/100)$, for $j = 0, \ldots, 100$, where q^v_τ, for $\tau \in (0, 1)$, denotes the quantile of the vertical average at probability level τ, and q^v_0 and q^v_1 denote the lower and upper bounds of the vertical average. We remark that $q^v_0 = a$, and $q^v_1 = b$ by the definition of vertical averaging. Note also that the quantiles of the vertical average span the entire interval $[a, b]$, and this is why we use

them to construct the family of angled lines.

The dark blue points in the left panel of Figure 8 correspond to 11 of the 101 points in Implementation Method 2. In comparison with the red starred points, the dark blue points lie closer to the graphs of the great majority of the CDFs in the dotted rectangular region. In the right panel of Figure 8, we have zoomed in on the left extreme of the left panel (i.e., the dotted rectangular region; note the change of scale on the x-axes of the left and right panels). This panel presents the two different implementation methods for angular averaging with $\theta = 89°$. The first issue to note in this panel is that reducing the number of points in Implementation Method 1 is unwise, as it leads to a notably different CDF. The second issue to note is that, by using Implementation Method 2 with 101 points, we have been able to achieve almost the same CDF as using Implementation Method 1 with 1001 points.

As the computational efficiency highly depends on the number of angled lines considered, our method achieves accurate estimation but only requires about 10% of the computational time required by Implementation Method 1, which was used by Taylor and Meng (2025).

5.2. Implementation of radial averaging

Like angular averaging, our radial averaging approach cannot be implemented using closed-form expressions. For radial averaging, we use the same 101 points from our Implementation Method 2 for angular averaging, and average along vertical or non-positively-sloped rays that pass through each of the 101 points and the focal point(s), the selection of which will be discussed in Section 5.4. The radial average CDF is produced by first computing the points of intersection between the individual CDFs and the rays; then, for a given ray, the points of intersection are averaged; and finally, we linearly interpolate between the points of intersection.

5.3. Improved parameter optimization for angular averaging

For angular averaging, the parameter is the angle θ, which can take any real value in $[0, 90°]$. The angle can be optimized by evaluating a finite set of candidate angles and selecting the one that minimizes the loss over the training period. Taylor and Meng (2025) rescale the interval $[a, b]$ of each set of CDFs to have unit length and consider all integer degrees between 0° and 90°. However, these angles are very susceptible to the scaling and outlying CDFs. Indeed, the angle θ that is applied to one set of CDFs may not be comparable to the same angle θ that is applied to another set of CDFs when the supports of them have different scales. To address this, we propose to consider angular averaging at multiple different candidate *relative angles*, which are not affected by the scale of the supports of CDFs and outlying forecasts, and are comparable across different sets of CDFs. We achieve this using the quantiles of the vertical average. Specifically, we draw 101 lines joining $(q^v_0, 1)$ and $(q^v_{1-j/100}, 1 - j/100)$, for $j = 0, \ldots, 100$. The corresponding 101 angles are $0, \ldots, 100$ relative degrees. Three examples of relative angles are illustrated in Figure 9, which shows a set of 32 CDF forecasts for mortality

from our empirical analysis of Section 6. A horizontal line corresponds to 0 relative degrees, and a vertical line corresponds to 100 relative degrees. The relative angles are not affected by rescaling of the x-axis, because when the CDFs are rescaled horizontally by a factor, the quantiles of the vertical average that are used to define the relative angles will also change by the same factor.

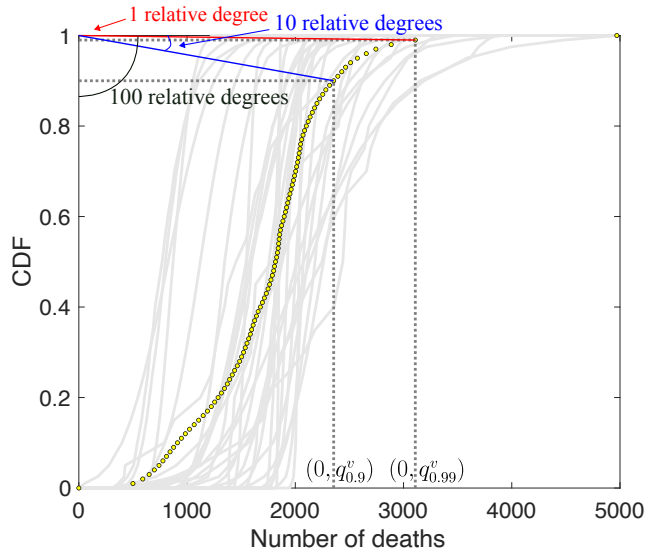


Figure 9: Illustration of the relative angles for angular averaging using CDF forecasts from our empirical study of Section 6. The plot shows 32 1-week ahead CDF forecasts for California in week 40. The 101 yellow points have x-coordinates that are the lower and upper bounds and quantiles of the vertical average at probability levels $0.01 \ldots, 0.99$ and y-coordinates that are the corresponding probabilities. Each point corresponds to a relative angle.

We remark that the choice of 101 candidate relative angles already provides sufficient resolution, since the resulting averaged CDF varies smoothly as the angle changes. We provide an example illustrating this in Figure S3 of the Supplementary Material.

5.4. Parameter optimization for radial averaging

For radial averaging, the parameters are the focal points. We optimize the focal points by defining a grid containing a collection of candidate points, and choosing the best focal points based on their performance for the training set. We employ this method for parameter optimization because, as we have mentioned in Section 5.2, the radial average CDF cannot be expressed as an explicit function of the parameters. Consequently, there is no closed-form expression for the loss defined by the scoring rule as a function of the parameters. This precludes the use of gradient-based optimization techniques, which require derivatives of the objective function with respect to the parameters.

Similar to our definition of relative angles, the grid is constructed based on the quantiles of the vertical average. Specifically, the grid for each focal point is defined to contain a finite set of points which have x-coordinates $q^v_{j/m}$, for $m = 10, j = 0, \ldots, 10$, y-coordinates $0, 0.1, \ldots, 1$ under the constraint that the points are in the admissible region. This relative location allows

for meaningful correspondence across different weeks, since the mortalities in different weeks from our empirical analysis of Section 6 have different scales.

5.5. Additional details for the practical use of radial averaging

Building on the parameter optimization method described in the previous subsection, we now discuss further details and several practical considerations for determining both the number and location of the focal points.

In terms of the number of focal points, we suggest using up to three focal points for the following reasons. First, a large number of parameters typically increases the risk of overfitting, especially when the historical data are limited, as is the case with Covid data. Second, recall that radial averaging is defined as a method that transitions between vertical and horizontal averaging, and the number of focal points decides the number of transitions. Our empirical results in Section 6.2 show that three focal points would be sufficient to produce all desirable patterns of transitions. Third, using up to three focal points essentially requires optimizing only one of them, and the remaining focal points are determined automatically. As illustrated in Section 3.3 and Figure 4, the radial average forecast is fully determined by o', since the other focal points share the same x-coordinate or the same y-coordinate with o'. Finally, the location of o' inherently indicates the number of focal points (one, two or three) employed, which was also illustrated in Figure 4. This is in analogy to angular averaging, where 0 degrees automatically implies that horizontal averaging is performed. As such, the optimization can be simplified without the need to optimize the number of focal points separately. In conclusion, all we need to do is just to optimize o'.

To determine the optimal o', we evaluate the in-sample loss over a finite grid of candidate locations for o', and the optimal one is selected as the one that minimizes the loss. We provide a visual illustration in Figure 10 for the grid we construct. For comparison, the grid in two weeks from our empirical analysis of Section 6 are displayed. As described in Section 5.4, the grid is constructed using the quantiles of the vertical average forecast, i.e., the x-coordinates of the black points are in $\{q_0^v, q_{0.1}^v, \ldots, q_1^v\}$ and the y-coordinates are in $\{0, 0.1, \ldots, 1\}$, and the points should not fall into the yellow inadmissible region to ensure that averaging yields a CDF, as discussed in Section 4.1.

A practical consideration that arises in optimizing over the grid is that some candidate locations may be admissible only in certain weeks, implying that their corresponding losses are not available for all weeks in the training set. This makes it difficult to evaluate all candidates on a common basis and to identify the optimal focal point. To provide a concrete example, comparing the two panels of Figure 10, note that points on the boundaries of the rectangle remain admissible throughout, while the interior points may be admissible only in certain weeks. The boundary points can therefore be used as reference points for imputing the missing losses:

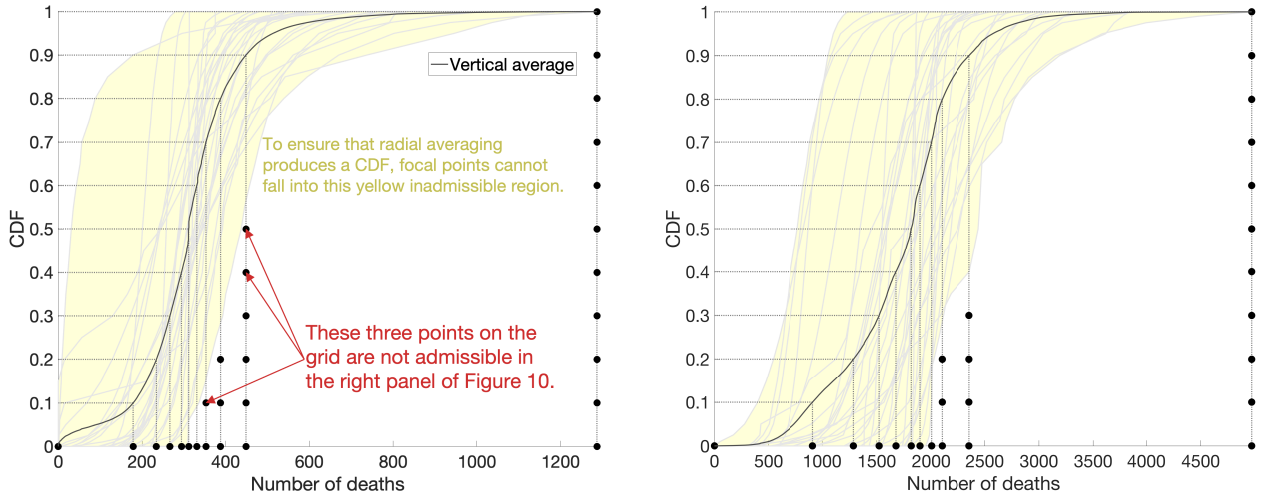


Figure 10: Illustration of the grid for o' for California in weeks 50 (left) and 40 (right). Dotted lines indicate quantiles of the vertical average, which are used to construct grid points. Yellow regions represent the inadmissible region, as in Figure 6, where focal points cannot be located.

we simply fill the missing loss with the average of the losses corresponding to the projections of each inadmissible point onto the bottom and right boundaries, so that the loss for every candidate can be averaged across the entire training period.

As a final remark, we would like to mention that the computational cost can be reduced by tailoring the grid to the specific characteristics of the data. The distribution of Covid mortality forecasts is typically right-skewed with a long right tail. This means that the region above the CDFs (that is, the intersection of the *epigraphs*) within the rectangle is negligibly small in many weeks. As such, in our empirical study, we only consider focal points below the CDFs, as shown in Figure 10. Section S1.4 of the Supplementary Material provides a visualization for the radial average forecast generated by each grid point.

6. Empirical analysis: Covid mortality forecasting

6.1. Data description

In April 2020, the U.S. COVID-19 Forecast Hub (Cramer et al., 2022) was launched to address the urgent need for accurate and timely forecasts during the Covid pandemic. Governments, public health agencies and healthcare systems required reliable predictions to guide policy and planning. However, the diversity of modeling methods and the lack of a unified framework for integrating forecasts posed challenges for decision-makers. The hub was designed to fill this gap by creating a centralized platform for collecting, evaluating and aggregating predictions from different academia, industry and government-affiliated teams around the world. The types of forecasting approaches varied widely, and included data-driven models, mechanistic models and hybrid models. This diversity of modeling techniques motivates the use of aggregation methods to form a synthesis of the information in the different forecasts. In this empirical study, we

compare vertical, horizontal, angular and radial averaging.

In this paper, we consider forecasts of weekly incident Covid mortality over one- to four-week prediction horizons, produced from forecast origins for the 84-week period between June 6, 2020 and January 8, 2022 (referred to as weeks 1 to 84). To evaluate the forecasts, we use actual mortality recorded on January 15, 2022, for the 84-week period between June 13, 2020 and January 15, 2022 (referred to as weeks 2 to 85). We consider data for the following 52 locations: each of the 50 states, Washington D.C., and the national level (i.e., total for the U.S.). The mortality forecasts submitted by the teams are in the form of quantiles at the following 23 probability levels: $0.01, 0.025, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80,$ $0.85, 0.90, 0.95, 0.975$ and 0.99. Because in this paper we are interested in distributional forecasts, we construct continuous, piecewise-linear CDFs from the quantiles via linear interpolation. Details for this are provided in Section S2.1 of the Supplementary Material.

A key achievement of the Hub has been creating ensemble forecasts by combining forecasts submitted by eligible contributing teams. The number of teams that submitted forecasts of incident deaths, and were screened for inclusion in the ensemble model, increased from an average across the 52 locations of around five in week 1 to around 35 in most weeks, and decreased to around 10 in week 84. We include the same teams included in the ensembles produced by the Hub.

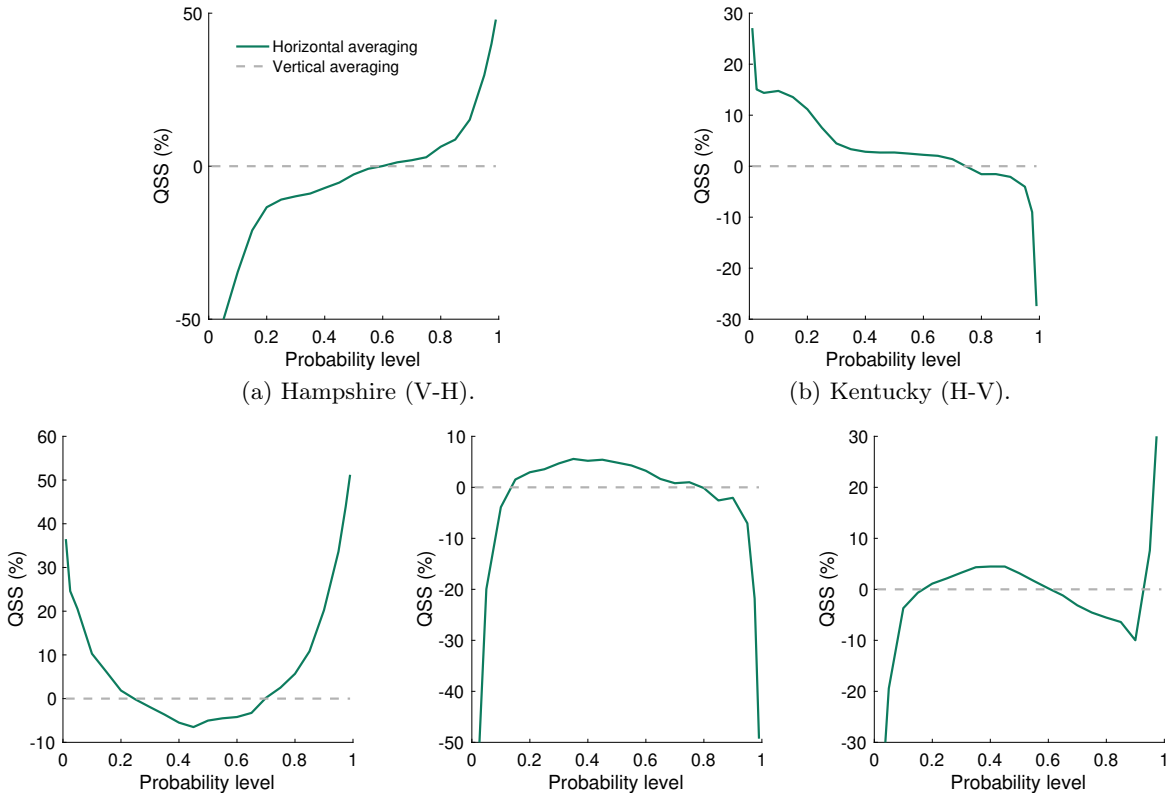
Since the forecasts submitted to the Hub are expressed as collections of predicted quantiles q_τ with probability level τ, we evaluate forecasts of the CDFs using the following multiple quantile score (MQS), which has been used in studies of the data from the COVID-19 Forecast Hub (see, for example, Ray et al., 2023): $\text{MQS}(\{q_\tau\}_{\tau\in\mathcal{T}}, y) \coloneqq \frac{1}{|\mathcal{T}|}\sum_{\tau\in\mathcal{T}} 2\text{QS}\left(\{q_\tau\}_{\tau\in\mathcal{T}}, y\right)$, where the quantile score (QS) is given by $\text{QS}\left(\{q_\tau\}_{\tau\in\mathcal{T}}, y\right) \coloneqq (\tau - \mathbb{1}(y \leq q_\tau))(y - q_\tau)$, q_τ denotes the quantile at probability level τ, y is the observation, and $\mathcal{T}$ is the set of the 23 probability levels considered by the Hub. The MQS is a proper scoring rule (Jose & Winkler, 2009).

We employ an expanding window approach to split the data into training and test sets. Initially, we use the first half of the dataset for training. For each subsequent forecast origin, the training set expands to include all data up to the forecast origin. This method ensures that the parameters are estimated on progressively more data without using future observations for training. For each location $s \in \{1, \ldots, 52\}$, each lead time $h \in \{1, \ldots, 4\}$ and each forecast origin in $w \in \{43, \ldots, 84\}$, we estimate averaging method parameters by minimizing the MQS averaged over the in-sample periods.

6.2. In-sample analysis

As mentioned in Section 5.3, the first half of the dataset serves as the initial training set. In this subsection, we present key observations obtained from it, and select the appropriate forms of radial averaging to implement for this dataset based on these in-sample findings.

For horizontal and vertical averaging, for each of the 23 probability levels, we compute the QS averaged across the initial training set and the four lead times for each location. The quantile skill score (QSS) is the percentage by which a method has a lower QS than a reference method. Throughout this empirical study, we select the vertical average as the reference. Higher values of the QSS indicate greater forecast accuracy. Figure 11 displays plots of the QSS for horizontal averaging for five out of the 52 locations. Each of the five panels shows one of the following patterns: V-H, H-V, H-V-H, V-H-V, V-H-V-H. Recall that in Section 3, we used the notation V-H to represent radial averaging that performs vertical averaging for the left tail and horizontal averaging for the right tail. Here, with a slight abuse of notation, we use V-H to indicate that vertical averaging performs better for the left tail, while horizontal averaging performs better for the right tail, and similar abbreviated notations are used for other types of relative performance. If the relative performance pattern V-H is observed, it suggests that the V-H averaging method is desirable. We inspected the QSS for all 52 locations, and found that almost all exhibited one of the five patterns presented in Figure 11.



(a) Hampshire (V-H). (b) Kentucky (H-V).

(c) Utah (H-V-H). (d) Arkansas (V-H-V). (e) Kansas (V-H-V-H).

Figure 11: QSS for horizontal averaging for five locations. QSS uses vertical averaging as benchmark, and is averaged over the initial training sample and the four lead times for five locations.

Based on this QSS result observed from the initial training sample, we select the forms of radial averaging that will be used throughout the out-of-sample period. It would be desirable if there is an averaging method that has the flexibility to produce all these five patterns of averaging, so that we can conveniently adopt it for all locations. Indeed, recall that in the

five panels of Figure 4, we illustrated that the radial averaging with the three focal points has the flexibility to produce all these five patterns, with suitable positioning of the focal points. This indicates that the version of radial averaging with three focal points shown in Figure 4 is suitable for this dataset, and thus we implement it for our empirical study. As we pointed out in Section 5.5, the problem of optimizing the locations of the focal points reduces to optimizing the location of one focal point. We employ the grid search approach as described in Sections 5.4 and 5.5 to optimize this focal point.

6.3. Out-of-sample results

Table 1 presents the MQS results. To summarize the results for the 52 series, and in view of the MQS being larger for higher mortality series, in Table 1, we report the MQS results for the following groupings of the series: the single series of the U.S. at the national level; the 17 locations with the highest cumulative Covid mortality in the final week of our dataset; the 17 locations with the next highest cumulative deaths in the final week; and the 17 locations with the lowest cumulative deaths in the final week. The MQS presented in the table is the MQS averaged across all forecast origins in the out-of-sample period, all lead times and all locations in each grouping. We also compute the multiple quantile skill score (MQSS) which is defined as the percentage by which the MQS is lower than that of the vertical average. To average over the MQSS, we compute the geometric mean of the ratios of the score for each method to the score for vertical averaging, and then subtract this from 1. In each column, bold highlights the best performing method. Horizontal averaging is the most accurate method for the U.S. series. For the other four groupings of series, angular averaging is the second most accurate method, while radial averaging produce the best results.

Table 1: For the four averaging methods, MQS and MQSS averaged across the four lead times and out-of-sample periods for the different groupings of the 52 locations.

	MQS					MQSS (%)				
	U.S.	High	Medium	Low	All	U.S.	High	Medium	Low	All
Vertical	779.50	57.77	23.79	7.57	44.13	0.00	0.00	0.00	0.00	0.00
Horizontal	**682.35**	60.93	24.65	7.86	43.67	**12.46**	-7.19	-1.78	-5.72	-4.51
Angular	707.03	57.31	23.43	7.40	42.41	9.30	1.60	2.36	2.11	2.17
Radial	693.32	**56.63**	**23.36**	**7.26**	**41.86**	11.06	**2.28**	**4.08**	**5.09**	**3.97**

Table 2 presents the MQS and MQSS of each method for each lead time. Radial averaging performs the best for averaging 1- and 2-week ahead CDF forecasts. When the lead time is 3 or 4 weeks, radial averaging has the highest MQSS for all groupings of the series, except the U.S. national level series, for which horizontal averaging is the most accurate.

To visualize the comparative accuracy of the averaging methods for each state, in Figure 12, we present the choropleth maps for the MQSS of horizontal, angular and radial averaging

Table 2: For the four averaging methods and each lead time, MQS and MQSS averaged across the out-of-sample periods for the different groupings of the 52 locations.

Lead time		MQS					MQSS (%)				
(week)		U.S.	High	Medium	Low	All	U.S.	High	Medium	Low	All
1	Vertical	594.45	47.76	21.09	6.86	36.18	0.00	0.00	0.00	0.00	0.00
	Horizontal	554.50	51.36	22.46	7.05	37.10	6.72	-10.60	-3.70	-3.72	-5.70
	Angular	556.78	46.62	**21.04**	6.79	35.05	6.34	3.14	1.67	0.85	1.98
	Radial	**553.10**	**46.00**	21.13	**6.76**	**34.79**	**6.96**	**5.31**	**2.36**	**1.50**	**3.15**
2	Vertical	693.12	53.85	23.67	7.42	41.10	0.00	0.00	0.00	0.00	0.00
	Horizontal	622.74	57.83	25.21	7.58	41.60	10.15	-10.00	-4.71	-4.30	-5.96
	Angular	625.88	52.64	23.53	7.29	39.32	9.70	3.08	1.75	1.76	2.35
	Radial	**618.71**	**50.67**	**23.34**	**7.19**	**38.44**	**10.74**	**7.82**	**3.98**	**4.32**	**5.49**
3	Vertical	825.60	**60.08**	23.50	7.68	45.71	0.00	0.00	0.00	0.00	0.00
	Horizontal	**711.97**	63.13	23.80	7.79	44.66	**13.76**	-6.61	0.12	-4.65	-3.31
	Angular	738.89	60.17	**22.98**	7.51	43.85	10.50	**1.24**	3.01	2.20	2.32
	Radial	722.55	60.34	23.03	**7.33**	**43.55**	12.48	0.42	**3.73**	**6.05**	**3.61**
4	Vertical	1004.81	**69.39**	26.89	8.33	53.52	0.00	0.00	0.00	0.00	0.00
	Horizontal	**840.21**	71.40	27.15	9.03	51.33	**16.38**	-3.25	0.47	-8.41	-3.24
	Angular	906.56	69.80	26.18	8.02	51.44	9.78	**1.70**	3.09	3.52	2.92
	Radial	878.90	69.51	**25.94**	**7.76**	**50.64**	12.53	0.86	**6.13**	**8.03**	**5.20**

for each of the 50 states and D.C. Red-shaped regions indicate states where the particular averaging method is less accurate than vertical averaging, whereas blue-shaded regions indicate better performance than vertical averaging. The results show that angular averaging and radial averaging yield more accurate forecasts for most locations than vertical averaging, and radial averaging has the best performance among the four methods.

From Figure 12, it seems that radial averaging underperforms in a few states. The seemingly poor performance is due to a few highly outlying forecasts submitted by a single forecaster for these states on rare occasions, rather than any systematic pattern in the data. This forecaster produced such rare, extreme forecasts in two separate weeks, with the first occurrence markedly inflating the average MQS of our method. However, when similar outlying CDF forecasts reappeared, radial averaging could have adjusted its parameter and performed well. We provide a detailed investigation of these outlying forecasts and the average forecasts in their presence in Section S2.2 of the Supplementary Material. To complement the average MQS comparisons, which can be affected by such occasional extremes, Table S1 in the same section reports the percentage of out-of-sample weeks in which our method outperforms the benchmark.

As we mentioned in Section 1, radial averaging can be performed after outlier-handling mechanisms, such as trimming (Jose et al., 2014), specifically mean-based exterior trimming. A proportion of CDF forecasts with the smallest and largest predictive means are excluded before radial averaging of the remaining CDF forecasts. This will make radial averaging more robust

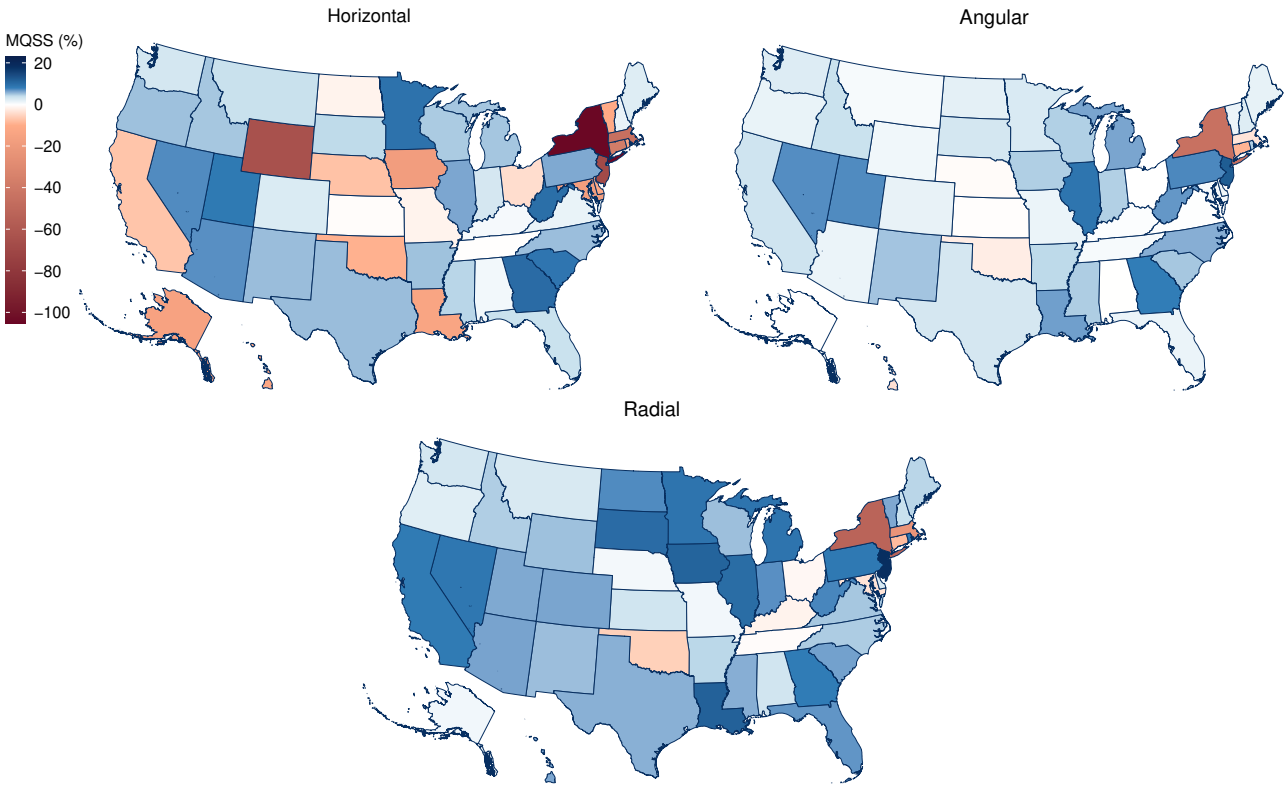


Figure 12: Choropleth maps showing the MQSS of each of the 50 states and D.C. for horizontal, angular and radial averaging. The colors indicate the predictive accuracy compared with the vertical averaging, with red indicating worse performance and blue indicating better performance.

to unpredictable idiosyncratic errors while retaining its strengths on regular data.

Figure 13 confirms that radial averaging does indeed draw on the advantages of the different averaging methods. The figure displays the average QSS across all 52 locations, all lead times and all weeks in the out-of-sample period at the 23 individual probability levels. For low probability levels (below 0.4), radial and angular averaging achieve the most accurate forecasts in terms of QSS. For intermediate quantiles at probability levels 0.4-0.8, the QSS of radial averaging is close to that of vertical averaging, which is the best performing method in this range. For the right tail, radial averaging has similar performance to horizontal averaging, and largely outperforms the other two.

Table 3 compares the dispersion of the four averaging methods based on the interquartile range (IQR) and the width of the 95% interval averaged across the four lead times, the forecast origins in the out-of-sample period, and locations in each grouping. The radial average forecast exhibits the narrowest 95% intervals for the U.S. national series, medium mortality locations, low mortality locations and the grouping consisting of all locations. The IQRs for horizontal, angular and radial averaging are similar. In contrast, vertical averaging yields the largest IQRs for almost all groupings and considerably wider 95% intervals for all groupings.

The calibration of probabilistic forecasts can be assessed using a reliability diagram (Murphy & Winkler, 1977), showing how well predicted probabilities align with observed frequencies.

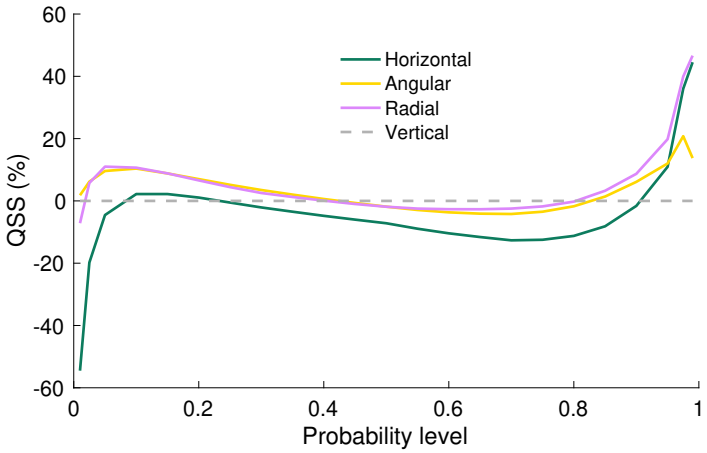


Figure 13: The QSS averaged across all 52 locations and four lead times for the 23 probability levels for the out-of-sample period for the four averaging methods.

Table 3: The IQR and the width of 95% intervals averaged across the four lead times, the forecast origins in the out-of-sample period, and locations in each group for the four methods.

	IQR					Width of 95% interval				
	U.S.	High	Medium	Low	All	U.S.	High	Medium	Low	All
Vertical	3021.4	175.8	66.6	22.1	144.6	12298.4	983.2	349.0	104.5	706.2
Horizontal	2392.4	136.2	58.2	22.4	116.9	7717.6	423.7	186.8	88.2	376.8
Angular	2356.4	139.3	56.3	20.1	115.8	8665.4	630.9	273.9	90.6	492.1
Radial	2457.0	137.8	55.5	19.5	116.8	7029.6	472.8	181.5	61.8	369.3

Figure 14 presents the reliability diagram based on quantile forecasts at the 23 probability levels for the out-of-sample period. The diagonal line represents perfectly calibrated forecasts. Vertical averaging is shown to demonstrate worse calibration than the other methods for most probability levels. The results for radial and angular averaging are similar, and they perform the best for higher probabilities. For lower probabilities, horizontal averaging has the best calibration. However, horizontal averaging provides an overly short left tail, as shown by the green curve falling above the diagonal line for the left tail.

In addition to the quantitative comparison presented above, it is also informative to visualize the CDF forecasts provided by different combining methods directly. An overall view of the average CDF forecasts across all weeks in the out-of-sample period for a state is presented in the Supplementary Material (Figure S7).

6.4. Computational time

We now report the computational time for implementation, parameter optimization and out-of-sample evaluation of our proposed radial averaging and the improved angular averaging. All experiments were implemented in Matlab on a machine equipped with a 3.40 GHz Intel Core i7 CPU and 32 GB of RAM.

The primary computational task is generating CDF forecasts associated with all candidate

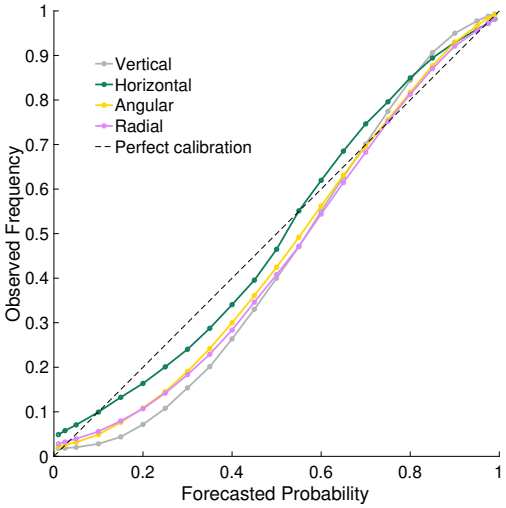


Figure 14: Reliability diagram for the four averaging methods, showing calibration averaged across the out-of-sample period and four lead times for all 52 locations.

parameter values. For all approximately 17,500 sets of CDFs in our study (52 locations, 4 lead times, 84 weeks), the total time to generate all angular average forecasts was 37.6 minutes, whereas our proposed radial averaging required only 8.7 minutes. The computational complexity of both methods grows linearly with the number of forecasters, because the methods are implemented by linearly interpolating individual CDFs. In our empirical study, the number of forecasters is typically around 35. This linear scaling suggests that the increase in runtime would be manageable, making the approaches suitable for large-scale forecast combination problems.

Once the average CDF forecasts are obtained, the remaining computations are highly efficient. They comprise in-sample parameter optimization by minimizing the MQS and evaluating the MQS associated with the optimized parameter over the out-of-sample period. Overall, this procedure required only about 7.5 seconds for angular averaging, and nearly identical time for radial averaging.

7. Conclusion

Our work in this paper was motivated by the relative performance of horizontal, vertical and angular averaging varying across the quantiles. We developed radial averaging as a method that addresses this by adapting angular averaging to allow the angle to vary across the quantiles. The method draws on the advantages of vertical, horizontal and angular averaging. We prove theoretically that our method is both proper and flexible. It is proper in the sense that it produces a CDF forecast. It is flexible in the sense that it can generate CDF forecasts that are sharper or less sharp than all of those generated by the other three methods in certain scenarios. Furthermore, it encompasses the other three averaging methods. Although our method allows different regions of the CDF forecasts to be averaged in different ways, notably, it does not require cumbersome optimization for every region of the CDF locally. The only parameters of

our method are the focal points. Empirical comparison with these methods provides support for our new form of CDF forecast aggregation.

Appendix A. Proofs of theoretical results

Proof of Theorem 1. Denote by $\hat{F}_r : \mathbb{R} \to [0,1]$ the averaged function. To prove that $\hat{F}_r$ is non-decreasing, we show that if $x_2 > x_1$, then $\hat{F}_r(x_2) \geq \hat{F}_r(x_1)$. Suppose the straight line (referred to as ℓ_1) where the point $(x_1, \hat{F}_r(x_1))$ lies is $y = a_1x + b_1$ or $x = c_1$, and the straight line (referred to as ℓ_2) where the point $(x_2, \hat{F}_r(x_2))$ lies is $y = a_2x + b_2$ or $x = c_2$. In the following, we only consider $y = a_1x + b_1$ and $y = a_2x + b_2$ with $a_1, a_2 < 0$, and the case when one or both of the lines are vertical or horizontal can be proved by similar arguments.

Denote by $p_{ij} = (x_{ij}^p, y_{ij}^p)$ the point of intersection between the graph of F_i and ℓ_j, for $i = 1, \ldots, n, j = 1, 2$. Then $(x_j, \hat{F}_r(x_j)) = \left(1/n \sum_{i=1}^n x_{ij}^p, 1/n \sum_{i=1}^n y_{ij}^p\right)$. Denote by $\bar{\ell}_2$ the line segment on ℓ_2 constrained by the end points $(\min_i\{x_{i2}^p\}, \max_i\{y_{i2}^p\})$ and $(\max_i\{x_{i2}^p\}, \min_i\{y_{i2}^p\})$.

We claim that if $x_2 > x_1$, then the line segment $\bar{\ell}_2$ lies to the right of and above ℓ_1:

$$\left\{(x,y) \in \mathbb{R}^2 \mid y = a_2x + b_2, x \in [\min_i\{x_{i2}^p\}, \max_i\{x_{i2}^p\}]\right\} \subset \{(x,y) \in \mathbb{R}^2 \mid y > a_1x + b_1\}.$$

We prove by contradiction. Assume that $\bar{\ell}_2$ lies on, or to the left of and below ℓ_1, i.e.,

$$\left\{(x,y) \in \mathbb{R}^2 \mid y = a_2x + b_2, x \in [\min_i\{x_{i2}^p\}, \max_i\{x_{i2}^p\}]\right\} \subset \{(x,y) \in \mathbb{R}^2 \mid y \leq a_1x + b_1\}.$$

Then since F_i's are non-decreasing, it follows that $x_{i2}^p \leq x_{i1}^p$ for all $i \in \{1, \ldots, n\}$, and thus $x_2 = 1/n \sum_{i=1}^n x_{i2}^p \leq 1/n \sum_{i=1}^n x_{i1}^p = x_1$, which is a contradiction.

Therefore, the facts that the line segment $\bar{\ell}_2$ lies to the right of and above ℓ_1 and that F_i's are non-decreasing yield that $x_{i2}^p \geq x_{i1}^p$ and $y_{i2}^p \geq y_{i1}^p$ for all $i \in \{1, \ldots, n\}$. This proves that $\hat{F}_r(x_2) = 1/n \sum_{i=1}^n y_{i2}^p \geq 1/n \sum_{i=1}^n y_{i1}^p = \hat{F}_r(x_1)$.

Now we show that $\hat{F}_r$ is right-continuous. $\hat{F}_r$ is right-continuous at c if $\lim_{x \to c^+} \hat{F}_r(x) = \hat{F}_r(c)$. Let $\ell_k \in L$ be the line that passes through $(c, \hat{F}_r(c))$. Denote by x_{ik}^p the x-coordinate of the point of intersection between F_i and ℓ_k, for $i = 1, \ldots, n$. It follows that $c = \frac{1}{n} \sum_{i=1}^n x_{ik}^p$. For every $i \in \{1, \ldots, n\}$, since F_i is right-continuous, for every $\epsilon/n > 0$, there exists $\delta_i > 0$ such that for all x_i satisfying

$$x_{ik}^p < x_i < x_{ik}^p + \delta_i, \tag{A.1}$$

we have $\left|\frac{1}{n}F_i(x_i) - \frac{1}{n}F_i(x_{ik}^p)\right| < \frac{\epsilon}{n}$. Then $\left|\frac{1}{n}\sum_{i=1}^n F_i(x_i) - \frac{1}{n}\sum_{i=1}^n F_i(x_{ik}^p)\right| \leq \sum_{i=1}^n \left|\frac{1}{n}F_i(x_i) - \frac{1}{n}F_i(x_{ik}^p)\right|$ $< \epsilon$.

Let $L' \subset L$ be the family of straight lines such that for any $\ell' \in L'$, the x-coordinate of the point of intersection of ℓ' and the graph of F_i is in $(x_{ik}^p, x_{ik}^p + \delta_i)$. Let $\bar{\ell} \in L'$ be the line which intersects the graph of F_i at points to the right of and above the points of intersection of other lines in L' and the graph of F_i, and denote by $x_{ik}^p + \delta_i^*$ the x-coodinate of point of

intersection between $\bar{\ell}$ and the graph of F_i, where $\delta_i^* \le \delta_i$. Let $\delta := \frac{1}{n}\sum_{i=1}^n \delta_i^*$. Then $c+\delta$ is the average of all $x_{ik}^p + \delta_i^*$. It follows that all x satisfying $c < x < c+\delta$ must be the average of some x_i's satisfying (A.1). So we can conclude that, for all $\epsilon > 0$, there exists δ, such that for all x satisfying $c < x < c+\delta$, we have that $\left|\hat{F}_r(x) - \hat{F}_r(c)\right| < \epsilon$.

Finally, we prove that $\lim_{x\to-\infty}\hat{F}_r(x) = 0$ and $\lim_{x\to+\infty}\hat{F}_r(x) = 1$. Suppose $q = (x, y), q_1 = (x_1, y_1), \ldots, q_n = (x_n, y_n)$ are points on the graphs of $\hat{F}_r, F_1, \ldots, F_n$, respectively, and they are on the same straight line. Then we have that $x = (x_1 + \cdots + x_n)/n$. As $x \to +\infty$, at least one of the $x_i \to +\infty$. Suppose w.l.o.g. that $x_a \to +\infty$. Then $y_a \to 1$, and it follows that $y_i \to 1$ for all i since all lines have non-positive slope or are vertical. The limit of $\hat{F}_r(x)$ as $x \to -\infty$ can be proved similarly. □

Proof of Theorem 2. (i) We first prove the case where two CDFs are averaged, and then generalize to $n > 2$ CDFs. Assume w.l.o.g. that the CDFs F_1 and F_2 have mean $\mu = 0$ and scale parameters b_1, b_2 respectively such that $0 < b_1 < b_2$. Denote by f_1 and f_2 the corresponding density functions. Denote by Φ and ϕ the CDF and PDF, respectively, of the random variable in the same location-scale family with mean μ and scale parameter 1. Consider an arbitrary ray passing through $(\mu, 0)$ with slope in $(-\infty, 0)$, and its intersections with the graph of F_1 and F_2 are $\left(x_1, \Phi\left(\frac{x_1}{b_1}\right)\right)$ and $\left(x_2, \Phi\left(\frac{x_2}{b_2}\right)\right)$, where $x_2 < x_1 < 0$. The midpoint of them are $\left(\frac{x_1+x_2}{2}, \frac{\Phi\left(\frac{x_1}{b_1}\right)+\Phi\left(\frac{x_2}{b_2}\right)}{2}\right) := (x_0, y_0)$, which is on the radial average.

We need to show that

$$\frac{b_1\Phi^{-1}(y_0) + b_2\Phi^{-1}(y_0)}{2} < \frac{x_1+x_2}{2}, \tag{A.2}$$

which is equivalent to $b_1\Phi^{-1}(y_0) - x_1 < x_2 - b_2\Phi^{-1}(y_0)$. We illustrate our idea for the proof in Figure A.15. The inequality we need to prove is that the orange line segment on the right is shorter than the orange line segment on the left.

If it holds that

$$f_2(z_2) < f_1(z_1) \tag{A.3}$$

for any $z_2 \in [b_2\Phi^{-1}(y_0), x_2]$ and any $z_1 \in [x_1, b_1\Phi^{-1}(y_0)]$, then by the Mean Value Theorem, and the fact that $\Phi\left(\frac{x_2}{b_2}\right) - y_0 = y_0 - \Phi\left(\frac{x_1}{b_1}\right)$, the desired result (A.2) follows. Then, recall that all log-concave densities f are strongly unimodal (Ibragimov, 1956). Since we assume that the densities are symmetric about the mean, both f_1 and f_2 are nondecreasing on $(-\infty, \mu]$. Therefore, to prove (A.3), it suffices to show that

$$f_2(x_2) < f_1(x_1). \tag{A.4}$$

To prove (A.4), we define $m(x) := \frac{\phi(x)}{\Phi(x)}$. We claim that m is non-increasing. Indeed, m is

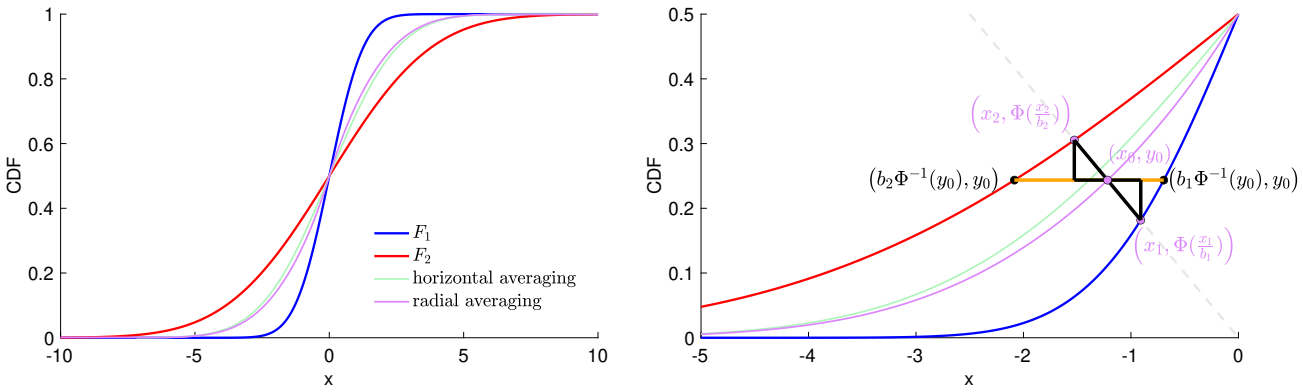


Figure A.15: An illustration of the idea for the proof of (i). The left panel shows that the quantiles of radial averaging are always smaller in absolute value than those of horizontal averaging at all probability levels. The right panel displays a ray and its points of intersections with the graphs of F_1 and F_2, and the midpoint (x_0, y_0) of the two points of intersection. The two black points have the same y-coordinate as the midpoint. Since the two black triangles are congruent, what we need to prove is that the orange line segment on the left is longer than the orange line segment on the right.

non-increasing if and only if for any $x_1 < x_2$,

$$\phi(x_2)\Phi(x_1) - \phi(x_1)\Phi(x_2) = \int_0^\infty (\phi(x_2)\phi(x_1 - \alpha) - \phi(x_1)\phi(x_2 - \alpha))\, \mathrm{d}\alpha \leq 0,$$

which holds since the integrand is non-positive by properties of log-concave function; see, for instance, Lemma 1(iv) of An (1997).

It follows from $\Phi\left(\frac{x_2}{b_2}\right) \geq \Phi\left(\frac{x_1}{b_1}\right)$ that $\frac{x_2}{b_2} \geq \frac{x_1}{b_1}$. Then $\frac{\phi\left(\frac{x_2}{b_2}\right)}{\Phi\left(\frac{x_2}{b_2}\right)} \leq \frac{\phi\left(\frac{x_1}{b_1}\right)}{\Phi\left(\frac{x_1}{b_1}\right)}$, and thus $\frac{\phi\left(\frac{x_2}{b_2}\right)}{\phi\left(\frac{x_1}{b_1}\right)} \leq \frac{\Phi\left(\frac{x_2}{b_2}\right)}{\Phi\left(\frac{x_1}{b_1}\right)} = \frac{x_2}{x_1} < \frac{b_2}{b_1}$. Note that $\frac{x_2}{x_1} \to \frac{b_2}{b_1}$ as $x_1, x_2 \to -\infty$, i.e., as the slope of the ray goes to 0. Therefore, we have shown that $\frac{\phi\left(\frac{x_2}{b_2}\right)}{\phi\left(\frac{x_1}{b_1}\right)} < \frac{b_2}{b_1}$, which is (A.4).

This concludes our proof for the statement that $\hat{F}_h^{-1}(p) < \hat{F}^{-1}_{r,(\mu,0),(\mu,1)}(p)$, for all $0 < p < 1/2$. Since both individual distributions are assumed to be symmetric about μ, and thus both horizontal and radial averages are also symmetric about μ. It follows immediately that $\hat{F}_h^{-1}(p) > \hat{F}^{-1}_{r,(\mu,0),(\mu,1)}(p)$, for all $1/2 < p < 1$, and $\hat{F}_h^{-1}(p) = \hat{F}^{-1}_{r,(\mu,0),(\mu,1)}(p)$, for $p = 1/2$.

Now we extend to the case where $F_1, \ldots, F_n$ (with scale parameters $b_1 < \cdots < b_n$) are averaged. Consider again an arbitrary negatively-sloped ray passing through $(\mu, 0)$ that intersect the individual CDFs at points $(x_1, y_1), \ldots, (x_n, y_n)$. Suppose the midpoint of these points of intersection is (x_0, y_0). Then we need to prove that

$$\frac{b_1\Phi^{-1}(y_0) + \cdots + b_n\Phi^{-1}(y_0)}{n} < \frac{x_1 + \cdots + x_n}{n}. \tag{A.5}$$

Suppose w.l.o.g. that $x_0 \leq x_j < \cdots < x_1$, and $x_n < \cdots < x_{j+1} < x_0$, for some $1 < j < n$. It follows that $y_0 \geq y_j > \cdots > y_1$, and $y_n > \cdots > y_{j+1} > y_0$. Since y_0 is the mean of $y_1, \ldots, y_n$, we have $(y_0 - y_1) + \cdots + (y_0 - y_j) = (y_{j+1} - y_0) + \cdots + (y_n - y_0)$. Then, by the same

arguments as above for the case when $n = 2$, the density functions satisfy $f_\xi(z_\xi) < f_\zeta(z_\zeta)$ for any $z_\xi \in \left[b_\xi \Phi^{-1}(y_0), x_\xi\right]$, any $z_\zeta \in \left[x_\zeta, b_\zeta \Phi^{-1}(y_0)\right]$, any $\xi \in \{j+1, \ldots, n\}$ and any $\zeta \in \{1, \ldots, j\}$. By the Mean Value Theorem,

$$\left(b_1 \Phi^{-1}(y_0) - x_1\right) + \cdots + \left(b_j \Phi^{-1}(y_0) - x_j\right) < \left(x_{j+1} - b_{j+1} \Phi^{-1}(y_0)\right) + \cdots + \left(x_n - b_n \Phi^{-1}(y_0)\right). \tag{A.6}$$

Rearranging (A.6) yields (A.5). Statement (i) follows immediately by applying the symmetry argument.

(ii) Let X have CDF F and finite second moment. The variance of X is given by

$$\begin{aligned} \mathrm{Var}(X) &= \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \int_{-\infty}^{\infty} \left(x^2 - 2x\mathbb{E}(X) + (\mathbb{E}(X))^2\right) f(x)\mathrm{d}x \\ &= \int_0^1 F^{-1}(p)^2 \mathrm{d}p - 2\mathbb{E}(X) \int_0^1 F^{-1}(p)\mathrm{d}p + (\mathbb{E}(X))^2 = \int_0^1 \left(F^{-1}(p) - \mathbb{E}(X)\right)^2 \mathrm{d}p, \end{aligned} \tag{A.7}$$

where the third equality follows since $p = F(x)$ and $\mathrm{d}p = f(x)\mathrm{d}x$. Since $F_1, \ldots, F_n$ are assumed to be the CDFs of log-concave distributions, which have finite moments of any order, the distributions given by vertical, horizontal and radial averaging all have finite variances. Then by (i) and (A.7), the variance of the radial average forecast is smaller than that of the horizontal average forecast. □

We remark that m we defined in the proof is closely related to the well-studied Mills ratio.The Mills ratio is defined by $R(x) \coloneqq \frac{S(x)}{\psi(x)}$, where $S(x) \coloneqq 1 - \Psi(x)$ is the survival function, and Ψ and ψ denote the CDF and PDF, respectively. The ratio of the density function to the CDF of log-concave distributions is studied by Bagnoli and Bergstrom (2005).

Proof of Theorem 3. Assume w.l.o.g. that $\mu = 0$. (i) Consider an arbitrary ray originating from point $(-d, 1/2)$ and intersecting the graphs of $F_1, \ldots, F_n$ at $(x_1, y_1), \ldots, (x_n, y_n)$, respectively. Denote the midpoint of the n points of intersection by (x_0, y_0). To prove (i), it suffices to prove the statement $\hat{F}_v(x) < \hat{F}_{r,(-d,1/2),(d,1/2)}(x)$ for all $x \in (-d, 0)$, which is equivalent to $F_1(x_0) + \cdots + F_n(x_0) < y_1 + \cdots + y_n$, that is,

$$(y_n - F_n(x_0)) + \cdots + (y_{j+1} - F_{j+1}(x_0)) < (F_1(x_0) - y_1) + \cdots + (F_j(x_0) - y_j). \tag{A.8}$$

The other statement of (i) follows by symmetry.

In the following, we suppose w.l.o.g. that $x_0 \leq x_j < \cdots < x_1$, and $x_n < \cdots < x_{j+1} < x_0$, for some $1 < j < n$. Under assumptions on the density functions, $f_\xi(z_\xi) < f_\zeta(z_\zeta)$ for any $z_\xi \in [x_\xi, x_0]$, any $z_\zeta \in [x_0, x_\zeta]$, any $\xi \in \{j+1, \ldots, n\}$ and any $\zeta \in \{1, \ldots, j\}$.

For any $\xi \in \{j+1, \ldots, n\}$, by the Mean Value Theorem, there exists some $w_\xi \in [x_\xi, x_0]$ such that $f_\xi(w_\xi)(x_0 - x_\xi) = F_\xi(x_0) - y_\xi$. Similarly, for any $\zeta \in \{1, \ldots, j\}$, there exists some $w_\zeta \in$

$[x_0, x_\zeta]$ such that $f_\zeta(w_\zeta)(x_\zeta - x_0) = y_\zeta - F_\zeta(x_0)$. Multiplying the left- and right-hand sides of $(x_1-x_0)+\cdots+(x_j-x_0) = (x_0-x_{j+1})+\cdots+(x_0-x_n)$ by $f_j(x_0)$ and $f_{j+1}(x_0)$, respectively, yields that $f_j(x_0)\{(x_1 - x_0) + \cdots + (x_j - x_0)\} > f_{j+1}(x_0)\{(x_0 - x_{j+1}) + \cdots + (x_0 - x_n)\}$. Therefore,

$$\begin{aligned} &f_1(w_1)(x_1 - x_0) + \cdots + f_j(w_j)(x_j - x_0) > f_j(x_0)\{(x_1 - x_0) + \cdots + (x_j - x_0)\} \\ >&f_{j+1}(x_0)\{(x_0 - x_{j+1}) + \cdots + (x_0 - x_n)\} > f_{j+1}(w_{j+1})(x_0 - x_{j+1}) + \cdots + f_n(w_n)(x_0 - x_n), \end{aligned}$$

which proves (A.8).

(ii) When the second moment of random variable X exists, then $\frac{1}{2}\mathbb{E}(X^2) = -\int_{-\infty}^{0} xF(x)\mathrm{d}x + \int_0^{\infty} x(1-F(x))\mathrm{d}x$. Then it follows from (i) that, when the focal points are $(-d, 1/2)$ and $(d, 1/2)$, the variance of the radial average forecast is larger than that of the vertical average forecast. □

References

An, M. Y. (1997). Log-concave probability distributions: Theory and statistical testing. *Duke University Dept of Economics Working Paper*(95-03).

Bagnoli, M., & Bergstrom, T. (2005). Log-concave probability and its applications. *Economic Theory*, *26*(2), 445–469.

Bassetti, F., Casarin, R., & Ravazzolo, F. (2018). Bayesian nonparametric calibration and combination of predictive distributions. *Journal of the American Statistical Association*, *113*(522), 675–685.

Bates, J. M., & Granger, C. W. (1969). The combination of forecasts. *Journal of the Operational Research Society*, *20*(4), 451–468.

Cramer, E. Y., Huang, Y., Wang, Y., Ray, E. L., Cornell, M., Bracher, J., Brennen, A., Castro Rivadeneira, A. J., Gerding, A., House, K., Jayawardena, D., Kanji, A. H., Khandelwal, A., Le, K., Niemi, J., Stark, A., Shah, A., Wattanachit, N., Zorn, M. W., ... Consortium, U. C.-. F. H. (2022). The United States COVID-19 forecast hub dataset. *Scientific Data*. Retrieved from `https://doi.org/10.1038/s41597-022-01517-w` doi: 10.1101/2021.11.04.21265886

Dawid, A. P., DeGroot, M. H., Mortera, J., Cooke, R., French, S., Genest, C., Schervish, M., Lindley, D. V., McConway, K., & Winkler, R. L. (1995). Coherent combination of experts' opinions. *Test*, *4*, 263–313.

Gneiting, T., Balabdaoui, F., & Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *69*(2), 243–268.

Gneiting, T., & Ranjan, R. (2013). Combining predictive distributions. *Electronic Journal of Statistics*, *7*, 1747–1782.

Grushka-Cockayne, Y., Jose, V. R. R., & Lichtendahl Jr, K. C. (2017). Ensembles of overfit and overconfident forecasts. *Management Science*, *63*(4), 1110–1130.

Hua, Z., & Zhang, B. (2008). Improving density forecast by modeling asymmetric features: An application to S&P500 returns. *European Journal of Operational Research*, *185*(2), 716–725.

Ibragimov, I. A. (1956). On the composition of unimodal distributions. *Theory of Probability & Its Applications*, *1*(2), 255–260.

Jose, V. R. R., Grushka-Cockayne, Y., & Lichtendahl, K. C. (2014). Trimmed opinion pools and the crowd's calibration problem. *Management Science*, *60*(2), 463–475.

Jose, V. R. R., & Winkler, R. L. (2009). Evaluating quantile assessments. *Operations Research*, *57*(5), 1287–1297.

Lichtendahl, K. C., Grushka-Cockayne, Y., & Winkler, R. L. (2013). Is it better to average probabilities or quantiles? *Management Science*, *59*(7), 1594–1611.

Lichtendahl, K. C., & Winkler, R. L. (2020). Why do some combinations perform better than others? *International Journal of Forecasting*, *36*(1), 142-149.

Liu, Z., Jiang, P., Zhang, L., & Niu, X. (2020). A combined forecasting model for time series: Application to short-term wind speed forecasting. *Applied Energy*, *259*, 114137.

McAndrew, T., Wattanachit, N., Gibson, G. C., & Reich, N. G. (2021). Aggregating predictions from experts: A review of statistical methods, experiments, and applications. *Wiley Interdisciplinary Reviews: Computational Statistics*, *13*(2), e1514.

Murphy, A. H., & Winkler, R. L. (1977). Reliability of subjective probability forecasts of precipitation and temperature. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *26*(1), 41–47.

Nikolopoulos, K., Punia, S., Schäfers, A., Tsinopoulos, C., & Vasilakis, C. (2021). Forecasting and planning during a pandemic: Covid-19 growth rates, supply chain disruptions, and governmental decisions. *European Journal of Operational Research*, *290*(1), 99–115.

Ranjan, R., & Gneiting, T. (2010). Combining probability forecasts. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *72*, 71–91.

Ray, E. L., Brooks, L. C., Bien, J., Biggerstaff, M., Bosse, N. I., Bracher, J., Cramer, E. Y., Funk, S., Gerding, A., Johansson, M. A., et al. (2023). Comparing trained and untrained probabilistic ensemble forecasts of COVID-19 cases and deaths in the United States. *International Journal of Forecasting*, *39*(3), 1366–1383.

Reid, D. J. (1968). Combining three estimates of gross domestic product. *Economica*, *35*(140), 431–444.

Safari, A., & Davallou, M. (2018). Oil price forecasting using a hybrid model. *Energy*, *148*, 49–58.

Stock, J. H., & Watson, M. W. (1999). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In R. F. Engle & H. White (Eds.), *Cointegration, causality and forecasting: A festschrift in honour of clive granger* (pp. 1–44). Oxford: Oxford University Press.

Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, *23*(6), 405–430.

Stone, M. (1961). The opinion pool. *The Annals of Mathematical Statistics*, 1339–1342.

Stratigakos, A., Pineda, S., & Morales, J. M. (2024). Decision-focused linear pooling for probabilistic forecast combination. *International Journal of Forecasting*.

Taylor, J. W., & Meng, X. (2025). Angular combining of forecasts of probability distributions. *Management Science*, Advance online publication. https://doi.org/10.1287/mnsc.2024.05558.

Timmermann, A. (2006). Forecast combinations. *Handbook of Economic Forecasting*, *1*, 135–196.

Wang, S., Kang, Y., & Petropoulos, F. (2024). Combining probabilistic forecasts of intermittent demand. *European Journal of Operational Research*, *315*(3), 1038–1048.

Wang, X., Hyndman, R. J., Li, F., & Kang, Y. (2023). Forecast combinations: An over 50-year review. *International Journal of Forecasting*, *39*(4), 1518–1547.

Ziel, F., & Weron, R. (2018). Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. *Energy Economics*, *70*, 396-420.