Assessing uncertainty for decision-making in climate adaptation and risk mitigation

Paolo Reggiani1 | Ezio Todini2 | Oleksiy Boyko1 | Roberto Buizza3

1Department of Civil Engineering, University of Siegen, Siegen, Germany
2Hon. President of the Italian Hydrological Society, Bologna, Italy
3Scuola Superiore Sant’Anna (Pisa) and Centre for Climate Change Studies and Sustainable Actions (3CSA, Pisa), Pisa, Italy

Correspondence
Paolo Reggiani, Department of Civil Engineering, University of Siegen, Paul Bonatz Strasse 9-11, Siegen, 57068, Germany.
Email: paolo.reggiani@uni-siegen.de

Present address
Paolo Reggiani, Department of Civil Engineering, University of Siegen, Paul Bonatz Strasse 9-11, Siegen, 57068, Germany

Abstract
Future water availability or crop yield studies, tied to statistics of river flow, precipitation, temperature or evaporation over medium to long-term horizons, are becoming frequent in climate impact and risk analysis. During the last two decades, access to multi-system integration of climate models has given rise to the concept of using model ensembles to issue probabilistic climatological projections. These probabilistic projections have not yet been exploited to the full extent in decision support, and are still used to mainly quantify uncertainty bands only for selected climate variables and indicators. One of the reasons of this limited use is the fact that the multi-system ensemble dispersion is sub-optimal and does not provide an accurate and reliable representation of the predictive probability density, which is essential for rational decision support under uncertain conditions. The aims of this paper are twofold. First, it seeks to highlight the potential benefits of using climate projections in conjunction with Bayesian paradigms towards educated decision-making. Second, it discusses how to appropriately formulate probabilistic forecasts by coherently integrating information contained in climate projection ensembles with observations to improve the estimation of the probability density function of future climate states. The results show that the proposed Bayesian approach yields unbiased and sharper predictive distributions for temperature with respect to using the unprocessed ensemble distribution. It also yields improved predictive densities with respect to the Reliability Ensemble Averaging (REA) method.

KEYWORDS
Bayesian forecasting, climate projections, decision-making, model ensembles, risk assessment, uncertainty

1 | INTRODUCTION

Estimating the extent of global warming due to mounting atmospheric greenhouse gas concentrations, and its impact on bio-physical systems is a global challenge. Because of the non-stationarity induced by climatic change (CC), extrapolation into the future is hardly possible using classical statistic techniques that rely on the assumption of stationarity. The assessment of climate change-induced risk (Aven and Renn, 2015) and the planning of mitigating action (Strzepek et al., 2011) relies on multiple numerical integration of Earth-system models,
which simulate as realistically as possible the interaction of atmosphere, land, ocean and sea-ice processes.

In these models, the physical processes are described by a system of highly non-linear governing equations, simulated on a three-dimensional grid with a finite spatial resolution that ranges between 10 and 50 km (IPCC, 2013). Due to the chaotic nature of the Earth-system (Lorenz, 1963), the model integrations are sensitive to the initial and boundary conditions of the slow-evolving components (e.g., the ocean state or sea-ice, which have a long memory and, if initialized wrongly, could have an impact also on the multi-annual integrations performed in climate studies) used in the simulation, as well as specific pre-defined carbon emission scenarios (Palmer, 2000). They are also sensitive to model approximations, for example, the fact that some key components of the Earth-system such as the sea-ice, vegetation, or the closure of the fresh-water cycle, are still simulated in a rather crude way.

The relative impact of the different sources of uncertainty depends on the climate projection length, with model-induced ones becoming dominant as time progresses (Hawkins and Sutton, 2011). More specifically, the sources of uncertainties can be separated into three main categories: internal variability, model uncertainty and scenario uncertainty, while initial condition uncertainties linked to the slow-evolving components are important only for the first few years of integration. Considering these three dominant sources of uncertainty in climate projections, Hawkins and Sutton (2011) concluded that internal variability is dominant in the first decade, while afterwards model uncertainty dominates. Scenario uncertainties play a relevant role in the long range, and can become dominant for certain regions and projections that go beyond few decades. Scenario uncertainties can be considered as part of the so-called boundary uncertainties.

Several multi-model ensemble simulations of future climate have been performed as part of the Climate Model Inter-comparison Project (CMIP), which to date includes multiple executive phases (Meehl et al., 2000, 2007; Eyring et al., 2012; Taylor et al., 2012). Results of these CMIP projects contribute to the scientific basis for the various IPCC reports. The time integration of Earth-system models driven by different RCPs leads to an ensemble of projections that can be used to estimate the response of the Earth climate to their forcing (IPCC, 2013). Note that the term ‘projection’ is used instead of ‘prediction’ because future integrations that extend beyond few years are mainly driven by the particular adopted radiative forcing scenarios, rather than by the initial conditions, as it is the case, for example, with weather predictions. In this work, we view projections as pseudo-measurements of future unknown quantities that are intrinsically uncertain.

Common attention is directed towards the use of such projections to extract policy-relevant indicators about expected changes as guidance for appropriate mitigation and adaptation measures, and related investments. Such measures should be based on a decision-making process that takes the projections’ uncertainty into account. For instance, the investments costs required to mitigate crop yield losses or increase the protection level of river or sea dikes to counteract climate change impacts can be estimated more accurately when using reliable and sharp forecast distributions (Gneiting and Raftery, 2004) of future cost/loss estimates. The latter can be estimated from the climate variables projected by the Earth-system ensembles, and do by no means constitute deterministic quantities.

To complicate matters, validating the reliability of climate projections retrospectively with a temporal range longer than, say, 20 years, can only be done selectively. The impact of progressive model improvements on the accuracy of climate projections (Tebaldi and Knutti, 2007; Knutti and Sedláček, 2012) is difficult to gauge compared to weather forecasting, where the prognostic skill is retrospectively estimated over many independent cases covering different seasons, in terms of verification metrics (Matheson and Winkler, 1976; Wilks, 1995; Bröcker and Smith, 2007) and supported by the shorter (from days to months ahead, compared to years and decades for climate) prediction horizon and abundant observations (Buizza and Leutbecher, 2015; Buizza, 2019).

Assessing the uncertainty of climate projections on the other hand has been steadily evolving across successive CMIP phases. Three development stages can be recognized so far.

1) Räisänen (1997) was among the first to advocate the need for a quantitative methodology to objectively inter-compare climate model output, while only few years later a first workable proposal for probabilistic elaboration of ensemble climate model output was presented by Räisänen and Palmer (2001) using the 17-member climate projection ensemble of CMIP2. In this approach the probability of exceedance of a given threshold temperature and precipitation changes between the present and a future time window were simply calculated as the relative frequency of models within the ensemble that predicted such change. The probabilities of exceedance were calculated for individual model cells or over the entire simulation domain. A verification of the assessment was performed employing cross-validation with a perfect model study. The Brier score, a widely diffused metric (Wilks, 1995) used to assess the accuracy of probabilistic forecasts of discrete events, was used for skill assessment. The same approach was later extended to the
analysis and quantification of climatic extreme events (Palmer and Räisänen, 2002).

II) The second development stage in assessing the uncertainty of climate projections starts with the introduction of the Reliability Ensemble Average (REA) concept by Giorgi and Mearns (2002). The REA consists of estimating the weighted average for a prognosticated climatic variable, whereby the performance of the model over a historic period serves as guidance for weight assignment. A model is rewarded, if bias and model mean are close to the ensemble mean. The larger the bias and the further predictions of a model depart from the mean, the more it is penalized. In contrast to the model performance assessment of type I), the REA invokes model performance over past observations. The REA is applied by Giorgi and Mearns (2003) in the same fashion as first advocated by Räisänen and Palmer (2001) and assigns exceedance probabilities of future climate threshold events, temperature in particular.

III) The third stage, leading to contemporary climate post-processing methods, is the introduction of Bayesian inference on the probability density function of the statistical distribution parameters used in the climate ensemble, as proposed by Tebaldi et al. (2004, 2005). As for the other two methods, the Bayesian formulation assumes statistical independence of model scenarios, while acknowledging that such dependencies implicitly exist because the Earth-system models use similar data for their initialization and the definition of the boundary conditions (Knutti et al., 2010; Deser et al., 2012). The authors apply Bayes’ theorem to a prior distribution of parameters and the joint series of past observations and simulated historical and future model output. Said data are assumed to follow Gaussian distributions, while the respective distribution parameters, bias, mean and inverse variance, are considered to be Gamma or Gauss distributed. The assumption of statistical independence leads to the formulation of a joint posterior probability distribution, from which the parameters are drawn with the help of a multi-system ensemble sampler. The inferred mean and variance for individual models serve the specification of ensemble-mean and variance for projected climatic variables by suitably weighting the future data as indicated by the REA approach.

A Bayesian method, as discussed under III), is also used in the present work. It is worth reminding that Smith et al. (2009) have further extended this kind of Bayesian approach from univariate to multivariate by involving neighbouring cells in a regionally spatialized setting. Furrer et al. (2007) applied it to analyze the spatial variability of climate change signal within a Bayesian inference framework by developing a multivariate hierarchical Bayes structure. Tebaldi and Sansò (2009) employed a multivariate Bayesian formulation to the analysis of decadal averages of temperature and precipitation. Buser et al. (2009) adopted Bayesian parameter inference for regional climate projection analysis, in which the non-stationarity of observed and future climate are modelled as a linear trend. The scaling parameters and bias of the linear model are estimated by sampling from the posterior distribution on the basis of specific priors. Rougier (2007); Rougier and Sexton (2007) presented a Bayesian analysis framework to evaluate the occurrence probability of future climate model ensemble projections. Similar to what we propose here, model (or climate simulator) projections are envisaged in their analysis as error-affected pseudo-measurements of future climate.

Approaches I–III provide relatively sophisticated measures of the multi-model ensemble spread, which is assumed to represent the uncertainty of future climate (Déqué et al., 2007).

As we already mentioned earlier on climate projection length, the spread of a multi-system ensemble is a mixture of uncertainties linked to ill known initial and boundary conditions, including the ones linked to RCP choice and model uncertainties. The chaotic evolution of the Earth-system leads to their flow-dependent amplification. In summary, uncertainties are partly attributable to knowledge gaps, model representation deficiencies and random effects like erratic emission pattern, practical limits to model resolution and inaccurate initial or boundary conditions (Von Storch and Zwiers, 1999; Abramowitz et al., 2019). The time evolution of this uncertainty into the future is visualized in Figure 1. The uncertainty is quantified in terms of the ensemble statistical distributions, with mean and variance over a past or future temporal window. Note that our notion of evaluating average climate requires estimating a statistic that is a fairly accurate approximation of the true climatic mean relative to the “present climate” baseline period. This concept still acknowledges that climate may well change over geologic timescales and thus the long-term (geologic) mean differs from the one of “present climate”. The assumption of ergodicity and stationarity of the climate process over selected temporal windows moreover ensures that sufficiently frequent observations converge toward the true climatic mean.

Given all above, a naive use of the ensemble spread of climate model projections as indicator of compound climate uncertainty could be misleading when prognosticating the statistics of present or future climate and is thus conducive to sub-optimal decision-making (Varis et al., 2004; Fowler et al., 2007; Clark et al., 2016). In analogy to weather forecasting, also climate projections aim to estimate the probability distribution of future, not yet observed climate variables, conditional on multi-model ensemble (MME) projections, rather than relying
only on the ensemble spread as a measure of uncertainty (Palmer and Räisänen, 2002). Such conditional distribution is defined as *predictive density*.

The *predictive density* is a widely-used concept in the weather forecasting (Whitaker and Hamill, 2006) and the hydrological (Krzysztofowicz, 1999) community, but has not been adopted in climate science, albeit the concept has been introduced in this field by Rougier (2007) and Stephenson *et al.* (2012). While in practice there is no need for an MME projection to assess the predictive density, which is obtainable also for a single ensemble projection, the availability of multiple projections is advantageous (Hagedorn *et al.*, 2005; Weigel *et al.*, 2010), because, as in Bayesian Model Averaging (BMA) (Raftery and Madigan, 1997; Bhat *et al.*, 2011), it allows marginalizing uncertainty effects due to individual models by estimating the probability density conditional on the ensemble of all models. In the marginalization, the weighting factors multiplying the predictive densities for each model as the likelihood of a model being correct, either reflects our prior knowledge (for instance total ignorance corresponds to uniform weights) or can be derived *a posteriori* through Bayesian inference.

Another topic pivotal to assessing the uncertainty of ensemble projections is non-stationarity, manifested by transitory statistical moments. This precludes the use of stationarity and ergodicity assumptions, which are essential to estimate the moments. One way of overcoming this problem is to assume weak stationarity and ergodicity over a time window sufficiently short to guarantee that the CC effects would only marginally affect moments and probability distribution parameters, but sufficiently long to allow for a reliable estimation. For instance, a time window of 20–30 years with a central reference year, could be effectively used to this end. Once the weighted stochastic dependency between individual models of ensemble prediction and reality for a baseline period has been confirmed retrospectively, it is assumed to hold, while the chaotic climate system is propagated ahead by the physical model equations. This assumption of temporal invariance of the observation-prediction dependency across windows is commonly accepted (Piani *et al.*, 2005; Collins *et al.*, 2006; Buser *et al.*, 2009; Sansom *et al.*, 2016). An alternative handling of non-stationarity is given by Buser *et al.* (2009), who extrapolates linear trends found over the control period into the future through Bayesian inference. Also in his case invariance of the linear relationship between control period and the future must be assumed.

In summary, the central objective of this work is to estimate the predictive density concerning climate projections, while demonstrating its benefits for economic value estimation in forthcoming decision-making for practitioners. The manuscript is structured into nine sections. In Section 2, we discuss predictive uncertainty, Section 3 is devoted to Bayesian decision-making, Section 4 addresses non-stationarity, Sections 5 and 6 introduce uni- and multivariate conditioning, Section 7 data and methods, Section 8 presents the results and Section 9 concludes. Supplementary information is provided in Appendices.
2 | UNCERTAINTY SPECIFICATION

The predictive density is defined as the probability distribution of a quantity, conditional on all presently available information (Hamill and Whitaker, 2006):

\[ f(y|\bar{y}) \]  

where \( y \) is the vector of climatic observations (past or future) and \( \bar{y} \) is a vector of predictions of \( y \). These become the known quantities, which enhance our knowledge by conditioning the probability distribution of the observations \( y \). Future, not yet observed occurrences, are indicated with \( y^* \), and their respective predictions with \( \bar{y}^* \).

Potential future material losses due to adverse climate will not be affected by model predictions \( \bar{y}^* \), but caused by actual future occurrences \( y^* \). For instance, flood damages will not necessarily be caused because a climate simulator predicts water overtopping the levees, but only if this is going to occur in reality. The goal is to increase our knowledge on \( y^* \) by estimating its probability distribution conditioned on error-affected information \( \bar{y}^* \) provided by the Earth-system ensembles. Projections \( \bar{y}^* \) must therefore be considered as proxies of reality, effectively error-affected pseudo-measurements of \( y^* \) (Rougier and Goldstein, 2014). These proxies are not real, but are quantities that emulate reality and have some informative content to be extracted by conveniently conditioning the density (1). Climate projections are therefore error-affected estimates of the future Earth-system states serving as decisional knowledge support in reducing uncertainty through bias-removal and sharpening (Gneiting et al., 2007) of the probability density of \( y^* \) conditional on \( \bar{y}^* \). This conditional density is essential for decision-making in any impact analysis. An example of the conditional density for a bi-variate model-observation process is shown in Figure 2. One could argue that if a perfect model existed, issuing probabilistic forecasts could be avoided. The only perfect model is reality, implying that one would need to build a 1:1 scale model, which is obviously infeasible. When extrapolating into the future, initial and boundary conditions as well as model performances may vary, affecting the predictions. Therefore, all climate projections represent reality imperfectly and must be supplemented with the predictive density (1).

3 | DECISION SUPPORT AND RISK

The main objective of predictions/projections is to allow for correct and robust estimation of the expected benefits or losses descending from decisions, usually taken on the basis of loss minimization. Losses are described by non-linear and/or discontinuous utility functions \( U(a^*, y) \) of the decision \( a^* \) and of one or multiple triggering variables, for example, temperature, cumulative precipitation or a river stage. When the utility functions are non-linear and continuous, and decisions are taken under conditions of uncertainty, the expected value of the utility function does not coincide with the utility function computed in the predicted expected value of the triggering function, that is, \( E[U(a^*, y)] \neq U(a^*, E[y]) \). For this reason it is essential that the expected value of the utility is estimated correctly by weighting each possible realization of the trigger by its specific predictive probability of occurrence and integrating it over the field of existence. Therefore the expectation of the utility can only be correctly estimated given the probability distribution of \( y \).

The effect can be demonstrated on hand of an example concerning the assessment of crop losses (Tebaldi and Lobell, 2008) under CC, with crop production affected by temperature-enhanced evaporation around 2050. Effects of precipitation are not taken into account. The productivity utility function is described by the red line in Figure 3. Losses are equal to loss of productivity. Assuming a climate projection with an annual mean temperature \( T = 13 \, ^{\circ}C \), crop mortality increases, as a first species of crops cannot bear the resulting water-stress and, as a result, productivity collapses. Additional temperature increase will also affect other types of crops and losses grow until the mean temperature \( T = 15 \, ^{\circ}C \) is reached, above which additional types of crops become
affected. On reaching \( T > 17 \) °C crop growth falters entirely.

Our assumed projection for the year 2050 is either a deterministic prediction of the expected temperature (in this case \( \mu_T = 15.8 \) °C) or a probabilistic one, expressed in terms of the probability distribution, which has been derived using several prognostic models, and represented here as a Gaussian distribution \( N(\mu_T = 15.8 \) °C, \( \sigma_T^2 = 1.0 \)°C²) (Figure 3). If the deterministic prediction is used to estimate the expected economical losses, a mean \( \mu_T = 15.8 \) °C temperature yields a \( 44.4 \) M$ productivity loss, which severely underestimates actual expected losses because uncertainty remains unaccounted for. In other words, presence of uncertainty calls for caution, because actual occurrences may differ substantially from the predicted mean loss.

To obtain a correct estimate of the expected losses, the integral of the cost function and the predictive density (1) is needed. One evaluates all possible losses, weighting each by its probability of occurrence through the predictive density, and marginalizing the uncertainty through expectation as in (2).

\[
E[U] = 55.8 \text{M} \quad (4)
\]

The resulting expected losses (\( 55.8 \) M$) estimated under conditions of uncertain knowledge on future temperature are substantially larger than those obtained on a purely deterministic basis (\( 44.4 \) M$).

Summarizing, model forecasts must be perceived as tools aimed at increasing decision makers’ knowledge on future events and supporting decisions. They must not be interpreted as real future outcomes to be compared to real quantities such as thresholds, but instead assumed as error-affected measures of future outcomes, which allow improving the prediction over the quantity of interest via conditioning of the predictive density. The density (1) is then used to assess the risk forthcoming from the decision. One evaluates all possible losses, weighting each by its probability of occurrence through the predictive density, and marginalizing the uncertainty through expectation as in (2).

### 4 Non-stationarity of Climatic Processes

Climate change analysis requires to characterize the statistical properties of an average year, while the driving processes are non-stationary. Although common, but rarely acknowledged (Von Storch and Zwiers, 1999), the statistical characterization of precipitation, temperature, evapotranspiration and discharges is always estimated over time windows of set length, for which one must postulate weak stationarity and ergodicity. For a system to be weakly stationary, it is necessary to assume that the long-term non-stationary effects do not substantially alter the statistical properties of the variables of interest over the analysis window, therefore allowing to take the window-average statistics as representative. Ergodicity applies to a dynamical system whose trajectories cover all points in space so that the mean over the state space can be substituted by the temporal mean over a single trajectory (Stephenson et al., 2012). Figure 4 shows an example of a non-stationary long-term stochastic process with two selected windows, wherein the process is assumed weakly stationary and ergodic.

For a baseline period of length \( n_b \) years and a prognostic period of length \( n_p \), with average statistical properties of the central year that are representative for the whole window. By assuming weak stationarity and ergodicity, we imply that the first two or three moments estimated in time will be representative of the statistical properties of a generic year within the window. When dealing with variables on a seasonal or annual basis,
we may postulate that the underlying probability distribution is Gaussian with time-independent first-order moment $\mu$ and second-order moment $\sigma^2$, which depend only on the temporal difference $\Delta t$ between two analysis points. Alternatively, when the variables of interest are sampled at shorter time intervals than a year, say season, months or days, Normality of the distributions is not granted. The same applies to certain variables other than temperature, such as precipitation, that are generally non-Gaussian. For example, projections of precipitation are known to exhibit stronger autocorrelation than observed rainfall (Ines and Hansen, 2006). Regression approaches like Ensemble Model Output Statistics (EMOS; Gneiting et al., 2005; Kharin et al., 2012), which rely on multiple ensemble projections vs. classical short and medium-range forecasts is not advised for use because of the need for stochastic modelling of the model-observations dependency. An example of such dependency is shown in Figure 2.

5 | CONDITIONING OF VARIATES

5.1 | Stochastic properties of climate projections

The principal difference between handling climate projections vs. classical short and medium-range forecasts is that presently available projections may be capable of preserving the probability distributions of the simulated Earth-system variables, but not their observed auto- and cross-correlation structure (Giorgi and Francisco, 2000; Palmer, 2000; Vidale et al., 2003; Déqué et al., 2007), a characteristic named asynchronicity (Stoner et al., 2013). For example, projections of precipitation are known to exhibit stronger autocorrelation than observed rainfall (Ines and Hansen, 2006). Regression approaches like Ensemble Model Output Statistics (EMOS; Gneiting et al., 2005; Kharin et al., 2012), which rely on multiple regression between ensemble projections and observations, are not advised for use because of the need for stochastic modelling of the model-observations dependency. An example of such dependency is shown in Figure 2.

5.2 | Standard univariate conditioning

Standard conditioning in a multi-variate process provides the distribution of a random variable, given that the remaining variables are held constant. The Normal variate $\eta = [\eta_1, ..., \eta_n]$, conditional on the $j$-th ensemble member output $\hat{\eta}_j = [\hat{\eta}_{ij}, ..., \hat{\eta}_{nj}]$ for every time step $i$ of the $n$-dimensional array of temporal steps, has density: 

$$f(\eta) \sim N(\mu_\eta, \sigma_\eta^2)$$

$$\mu_\eta = E(\eta) = \frac{1}{n_b} \sum_{i=1}^{n_b} \eta_i; \quad \sigma_\eta^2 = \frac{1}{n_b-1} \sum_{i=1}^{n_b} (\eta_i - \mu_\eta)^2$$ (5a)

$$f(\hat{\eta}_j) \sim N(\mu_{\hat{\eta}_j}, \sigma_{\hat{\eta}_j}^2)$$

$$\mu_{\hat{\eta}_j} = E(\hat{\eta}_j) = \frac{1}{n_b} \sum_{i=1}^{n_b} \hat{\eta}_{ij}; \quad \sigma_{\hat{\eta}_j}^2 = \frac{1}{n_b-1} \sum_{i=1}^{n_b} (\hat{\eta}_{ij} - \mu_{\hat{\eta}_j})^2$$ (5b)

$E(\cdot)$ being the expectation over $i = 1, ..., n$ time steps. Similarly, the projected series $\hat{\eta}^*$ and $\hat{\eta}^*$ map into Gaussian variates $\eta^*$ and $\eta^*$ with stationary parameters $\mu_\eta^*, \mu_{\hat{\eta}_j}^*, \sigma_\eta^2$, and $\sigma_{\hat{\eta}_j}^2$. The Greek symbols notation is kept in place from here onwards to recall that those variates are truly Gaussian.
\begin{equation}
    f_j(\eta_i | \hat{\eta}_{ij}) \sim N\left(\mu_{\eta_i | \hat{\eta}_{ij}}, \sigma^2_{\eta_i | \hat{\eta}_{ij}}\right)
\end{equation}

\forall i \in 1, \ldots, n \text{ and } \forall j \in 1, \ldots, m, \text{ with parameters (Mardia et al., 1979)}:

\begin{equation}
    \begin{align*}
        &\mu_{\eta_i | \hat{\eta}_{ij}} = E\left(\eta_i | \hat{\eta}_{ij}\right) = \mu_\eta + \rho_{\eta \hat{\eta}_j} \frac{\sigma_{\eta}}{\sigma_{\hat{\eta}_j}} (\hat{\eta}_{ij} - \mu_\eta),
        \\
        &\sigma^2_{\eta_i | \hat{\eta}_{ij}} = Var\left(\eta_i | \hat{\eta}_{ij}\right) = \sigma^2_\eta \left(1 - \rho^2_{\eta \hat{\eta}_j}\right)
    \end{align*}
\end{equation}

where $\rho_{\eta \hat{\eta}_j}$ is the Pearson correlation. This standard formulation, employed by several authors (Hirsch, 1982; Wood and Schaake, 2008) to process streamflow forecasts, seems inappropriate for climate projections as explained previously. Similarly, standard correlation should be used even less as a quantity to be preserved when extrapolating the model-observation dependency into the distant future.

### 5.3 Weak univariate conditioning

Instead, the preference is for distribution matching (DM) (Wood et al., 2002; Drusch et al., 2005; Boë et al., 2007; Themeßl et al., 2011), regularly used for bias correcting or downsampling Earth-system model output or remotely sensed data against observations. DM (here denoted with $\hat{}$) is a form of conditioning, which does not depend on correlation, and hence, we call weak conditioning. One can weakly condition two arbitrarily distributed variables by applying probability matching empirically or by fitting appropriate parametric distributions $F$ and $G$ and then inverting one of the two (Hirsch, 1982; Hashino et al., 2007; Wood and Schaake, 2008). This approach calibrates one distribution against the other. By setting the cdf of the observations equal to that of a predictor:

\begin{equation}
    G(x) = F(\hat{x})
\end{equation}

and inverting $G$:

\begin{equation}
    \hat{x} = G^{-1}(F(\hat{x})) = T(\hat{x})
\end{equation}

we obtain a transformation $T$ and hence a new variable with the same moments of the variate it is calibrated against. If the variates are Gaussian, $T$ becomes:

\begin{equation}
    \hat{x} = T(\hat{x}) = \mu_x + \sigma_x \frac{\hat{x} - \mu_x}{\sigma_x}
\end{equation}

with distribution:

\begin{equation}
    f(\hat{x}) \sim N(\mu_x, \sigma^2_x)
\end{equation}

Subsequently, the predictive density of the conditioned variable for model $j$ at time $i$ is Gaussian:

\begin{equation}
    f_i,j(\eta_i | \hat{\eta}_{ij}) \sim N\left(\mu_{\eta_i | \hat{\eta}_{ij}}, \sigma^2_{\eta_i | \hat{\eta}_{ij}}\right)
\end{equation}

with parameters:

\begin{equation}
    \begin{align*}
        &\mu_{\eta_i | \hat{\eta}_{ij}} = E\left(\eta_i | \hat{\eta}_{ij}\right) = \mu_\eta + \rho_{\eta \hat{\eta}_j} \frac{\sigma_{\eta}}{\sigma_{\hat{\eta}_j}} \left(\hat{\eta}_{ij} - \mu_\eta\right),
        \\
        &\sigma^2_{\eta_i | \hat{\eta}_{ij}} = Var\left(\eta_i | \hat{\eta}_{ij}\right) = \sigma^2_\eta \left(1 - \rho^2_{\eta \hat{\eta}_j}\right)
    \end{align*}
\end{equation}

where $E(\cdot : \cdot)$ and $Var(\cdot : \cdot)$ are taken over all possible realizations of $\eta_i$ at time $i$. We note that in (13) the mean is independent of the correlation, while the variance appears linearly dependent on $\rho_{\eta \hat{\eta}_j}$ and becomes larger for $\rho_{\eta \hat{\eta}_j} < 1$ than the variance for standard conditioning in (7), where the correlation appears squared. The statistics of the prediction residuals for DM are estimated by evaluating the moments of the difference between observations and forecasts:

\begin{equation}
    \begin{align*}
        &E_{(i)} \left[\eta_i - E\left(\eta_i | \hat{\eta}_{ij}\right)\right] = \mu_\eta - \mu_{\eta \hat{\eta}_j} = 0, \\
        &Var_{(i)} \left[\eta_i - E\left(\eta_i | \hat{\eta}_{ij}\right)\right] = \sigma^2_\eta + \sigma^2_{\eta \hat{\eta}_j} - 2\rho_{\eta \hat{\eta}_j} \sigma_{\eta} \sigma_{\hat{\eta}_j} = 2\sigma^2_\eta \left(1 - \rho_{\eta \hat{\eta}_j}\right)
    \end{align*}
\end{equation}

noting that $Var_{(i)} \left[\left(\eta_i - \hat{\eta}_{ij}\right)\right] = \sigma^2_{\eta \hat{\eta}_j} = \sigma^2_\eta$. The conditional distributions on future temperature given model projections, is obtained in total analogy and yields a series of probability distributions of the observations $\eta_i | \hat{\eta}_{ij}$, that are weakly conditional ($\cdot$) on each projection $\hat{\eta}_{ij}$. For this the mean $\mu_\eta$ and variance $\sigma^2_\eta$ derived for the BW in Equation (13) are used, as the bias with respect to observations and the variance of observations are assumed to hold into the future (Collins et al., 2006). This hypothesis could be relaxed by imposing for example a linear trend on the mean and variance, as proposed by Buser et al. (2009), if observations suggest this type of climate evolution. Given the lack of evidence for such a clear trend in our study case, we have refrained from doing so.

Weak conditioning by matching Gaussian distributions has also the advantage that one retains a parametric expression of type (12) over the entire distribution, fully including...
the tails. This is not the case for quantile matching (Maraun, 2013; Stoner et al., 2013), where one relies on piecewise interpolation among probability quantiles of models and historical observations. Such interpolation may turn out to be inadequate for post-processing of climate projections, especially for the distribution tails, where few support points are available. This is relevant when extending the relationship, obtained over the observed period, to the unobserved warmer future.

6 | RETROSPECTIVE PREDICTION AND PROJECTIONS

As explained in Section 5.3, we use weak conditioning for retrospective predictions and projections. A prediction/projection is performed by estimating the density \( f(\eta_i) \) at time \( i \). To obtain the predictive density for multiple predictors, one can adopt the concept of mixture distribution, a weighted mean of univariate conditional distributions \( f(\eta_i : \eta_j) \forall j \in 1,...,m \):

\[
    f(\eta_i : \eta_j) \sim E(j) \left[ f_j(\eta_i : \eta_{i,j}) \right] = \sum_{j=1}^{m} w_j f_j(\eta_i : \eta_{i,j}) \quad (15)
\]

The similarity sign indicates that the predictive distribution is approximated by the mixture, while the bold symbol \( \eta_j \) indicates the \( m \)-dimensional vector \( [\hat{\eta}_{i,1}, ..., \hat{\eta}_{i,m}] \) of predictions \( j \) at time \( i \). Expressions for mean and variance of the mixture distribution are given in Supplement S2. The weights \( w_j \) are the probability of the \( j \)-th ensemble member being a valid prediction and can be estimated either by Bayesian inference, as in eq. (S1.5),

| Model          | Source Organization                                      |
|---------------|---------------------------------------------------------|
| ACCESS1-0     | Bureau of Meteorology, CSIRO Australia                 |
| BCC-CSM1-1    | standard resolution model, Beijing Climate Center, China |
| BCC-CSM1-1m   | moderate resolution model, Beijing Climate Center, China |
| BNU-ESM       | Beijing National University, China                      |
| CanESM2       | Canadian Centre for Climate Modelling and Analysis, Canada |
| CCSM4         | Community Climate System Model V4, NCAR-UCAR, USA      |
| CESM1-CAM5    | Community Earth System Model V1, Community Atmosphere Model V5, USA |
| CNRM-CM5      | Centre National de Recherches Météorologiques, France   |
| CSIRO-Mk3-6-0 | CSIRO Australia                                        |
| GFDL-ESM2M    | Geophysical Fluid Dynamics Laboratory, USA              |
| GFDL-ESM2G    | Russian Academy of Sciences, Russia                     |
| INMCM4        | Institut Pierre Simon Laplace, France                   |
| IPSL-CM5A-MR  | Institut Pierre Simon Laplace, France                   |
| MIROC5        | Model for Interdisciplinary Research on Climate, Japan   |
| MIROC-ESM     | Russian Academy of Sciences, Russia                     |
| MIROC-ESM-CHEM| Russian Academy of Sciences, Russia                     |
| MPI-ESM-MR    | Max Planck Institute, Germany                           |
| MRI-CGCM3     | Meteorological Research Institute, Japan                |
| NorESM1-M     | Norwegian Climate Center, Norway                        |
which provides a posterior distribution or by point estimation, such as maximum likelihood estimation (MLE), for a unique set of weights. The weights must be non-negative and add up to unity, \( w_j \geq 0 \) and \( \sum_j w_j = 1 \). Here we use two point estimation approaches to assess the weights: (a) Bayesian Model Averaging (BMA) and (b) Uniform Weighting (UW).

### 6.1 Bayesian model averaging

First, one can apply the classical BMA approach (Raftery et al., 2005), which assumes non-uniform weights \( w_j \) in (15) and maximizes the log-likelihood of observations:

\[
\text{max}_{w_j} \log \mathcal{L}(w_1, \ldots, w_m) = \max_{w_j} \sum_{i=1}^{n} \log \sum_{j=1}^{m} w_j f_j(\eta_i; \hat{\eta}_j)
\]

subject to \( \sum_{j=1}^{m} w_j = 1; w_j \geq 0 \)  \( \text{(16)} \)

From the perspective of Bayesian inference, MLE is a special case of maximum a posteriori (MAP) estimation that assumes a uniform prior distribution of weights. We also note that the BMA formulation assumes weighting of distributions of independent variates. As mentioned earlier, in multi-model ensemble projections like CMIP5, dependency is implicitly present (Knutti et al., 2017) due to similar setups used for all models of the inter-comparison experiment. In principle this would require relaxing the independence assumption to avoid underestimation of the compound predictive variance. Nevertheless, we continue to assume independence, as in classical BMA. Therefore the predictive mean and variance of the mixture distribution for the case of independent variates \( \mu_{\eta_i; \hat{\eta}_j} \) are (see S2):

\[
\begin{align*}
\mu_{\eta_i; \hat{\eta}_j} &= \sum_{j=1}^{m} w_j \mu_{\eta_i; \hat{\eta}_j} = \mathbf{w}^T \mu_{\eta_i; \hat{\eta}_j} \\
\sigma_{\eta_i; \hat{\eta}_j}^2 &= \sum_{j=1}^{m} w_j \sigma_{\eta_i; \hat{\eta}_j}^2 + \sum_{j=1}^{m} w_j \left( \mu_{\eta_i; \hat{\eta}_j} - \mu_{\eta_i; \hat{\eta}_j} \right)^2 \\
\end{align*}
\]  \( \text{(17)} \)

The BMA mixture weights are estimated using Expectation Maximization (EM) (Dempster et al., 1977) or the Newton–Rhapson method to maximize the log-likelihood.

### 6.2 Uniform weighting (UW)

Alternatively to BMA, one can apply UW when mixing the \( m \) univariate predictive densities \( f_j(\eta_i; \hat{\eta}_j) \). UW is equivalent to no prior preference on models and is a
special case of (15). By assigning an uniform prior to all model predictive densities, one attributes the same likelihood to all components of the mix. The expressions for the mixture can be derived from (15) by setting \( w_j = 1/m \), while the parameters are as in (17) with \( w_j = 1/m \). It is self-explanatory that UW makes MLE redundant, as weights are predetermined.

7 | DATA AND METHODS

The main steps of the implementation of the proposed processor is data preprocessing and calibration of the uncertainty processor over a 27-year historical reference period, performance of necessary verifications through a retrospective analysis and application of the processor for predictions in two 21-year prognostic windows in the future. First, we start explaining the data elaboration for a selected study region.

7.1 | Observed temperatures

Our selected study region covers a rectangular area including the entire Po river basin, northern Italy, as in Figure 5.

The area extent is from (46.75°N, 6.5°E) on the N-W corner to (44.0°N, 12.5°E) in the S-E. Instead of ground observations we use gridded temperatures at 0.25° × 0.25° (~31 × 31 km) spatial resolution of the ECMWF ERA5 reanalysis (Hersbach et al., 2020) as observation proxies, which we aggregate by spatial averaging over the entire River Po basin shape indicated in Figure 5. Spatial averaging weights the temperature of the cells on the basin shape edge by the area portion that lies within the basin boundaries, while temperatures are altitude-corrected from the ERA5 grid cell elevation to a common reference height using the dry adiabatic lapse rate. ERA5 output starts 1979 and is updated continuously. The product provides analyzed fields of global atmospheric state variables.

7.2 | Climate model projections

Next we process the output of the CMIP5 climate model ensemble. CMIP5 control simulations start on January 1, 1901 and end on December 31, 2005. The prognostic period continues from January 1, 2006 through December 31, 2100. We select a \( n_b = 27 \) years baseline window (BW) 1979–2005, two \( n_p = 21 \) years prognostic windows (PW), 2040–2060

**FIGURE 6** BW, summer (JJA) mean seasonal observed temperature \( \eta_i \), unprocessed ensemble output \( \bar{\eta}_i \), ensemble mean \( E_i(\bar{\eta}_i) \), predictive mean \( \mu_i(\bar{\eta}_i) \), BMA (top) and UW (bottom). The shadowed areas indicate the 50–95% credible intervals
(PW1) and 2080–2100 (PW2), and two representative carbon pathway scenarios, RCP4.5 and RCP8.5, four prognostic cases in total. Daily temperature output series from the 19 models listed in Table 1 are cropped from the global CMIP5 temperature output (at their native spatial resolution, which differs between models) for the River Po basin study area (see Figure 5). This set of models was chosen because it constitutes the most complete set of daily climate data for the historical and the two RCP scenarios we could access, and for which the orography data was available. Similarly to observations, we average the simulated temperature data spatially over the basin by scaling the temperature with the adiabatic lapse rate from the native model grid cell topographic height to a common reference elevation and then back to the ground, obtaining basin-average series of daily data. These can be aggregate in time, to get monthly, seasonal or yearly series for the historical and the four prognostic windows.

7.3 | Calibration

We apply the processor by first conditioning observations on the simulated daily climate data and obtaining $m = 19$ conditional univariate distributions for the baseline window (BW). For temperature the probability distributions are close to Gaussian. For this reason, no further intermediate transformation to Gaussian distribution is needed. Then we mix the univariate conditional distributions for two cases: (a) non-uniform BMA weights and (b) uniform weighting. The phase of conditioning the distributions and estimating the weights is called the calibration of the processor. Related details and results are given below.

7.3.1 | Two-step procedure

First, the distributions of the daily climate model ensemble output are matched against those of observations. Thence the series of 27 years of daily data are first grouped into months. As a result we obtain series of $n_m \times 27$ daily values for each month of the BW with $n_m$ the number of days of the month. For each of the $j$ models we obtain an uni-variate weakly conditional predictive distribution (12) of daily data, which is quasi-Gaussian and parameterized for each month through

| Definition | Description | BW | PW1 | PW2 | PW1 | PW2 |
|------------|-------------|----|-----|-----|-----|-----|
| Bayesian model averaging | $\mu_{\eta_i} = E_i(\mu_{\eta_i})$ | DJF (mean of means) | 0.24 | 1.34 | 1.93 | 1.70 | 3.54 |
| | | MAM | 8.79 | 10.37 | 11.0 | 10.94 | 13.16 |
| | | JJA | 18.73 | 21.23 | 21.8 | 21.79 | 25.12 |
| | | SON | 10.01 | 11.68 | 12.44 | 12.27 | 14.64 |
| | $\sigma^2_{\eta_i} = \text{Var}_i(\mu_{\eta_i})$ | DJF (variance of means) | 2.16 | 2.28 | 2.29 | 3.08 | 3.24 |
| | | MAM | 1.69 | 1.97 | 2.0 | 2.94 | 3.45 |
| | | JJA | 1.36 | 1.61 | 1.76 | 4.12 | 4.57 |
| | | SON | 1.26 | 1.53 | 1.6 | 2.65 | 2.88 |
| Uniform weighting | $\mu_{\eta_i} = E_i(\mu_{\eta_i})$ | DJF (mean of means) | 0.24 | 1.51 | 2.03 | 1.97 | 3.65 |
| | | MAM | 8.79 | 10.31 | 10.72 | 10.7 | 12.73 |
| | | JJA | 18.73 | 20.81 | 21.47 | 21.53 | 24.67 |
| | | SON | 10.01 | 11.61 | 12.22 | 12.08 | 14.4 |
| | $\sigma^2_{\eta_i} = \text{Var}_i(\mu_{\eta_i})$ | DJF (variance of means) | 2.41 | 2.6 | 2.65 | 3.21 | 3.44 |
| | | MAM | 1.95 | 2.28 | 2.44 | 3.22 | 3.98 |
| | | JJA | 1.71 | 2.24 | 2.43 | 4.37 | 5.26 |
| | | SON | 1.52 | 1.9 | 1.97 | 2.93 | 3.49 |
conditional mean and variance as in (13). The stochastic relationship between model and observations depends on the mean and variance of observations and the Pearson correlation among distribution-matched data.

Second, the univariate conditional distributions of daily data are weighted linearly as in Equation (15) for each of the 12 months of the year. The non-uniform BMA weights \( w_j \) are found by MLE (Equation (16)). Table 2 summarizes the BMA weights estimated for the BW and each month of the year. We note that models are assigned changing non-zero weights across the year. The predictive mean \( \mu_{\eta_i} \) and variance \( \sigma_{\eta_i}^2 \) for the four seasons (DJF, MAM, JJA, SON) of the 27 years of the BW are estimated via Equation (17) with non-uniform and uniform weights. The series of predictive means and variances, which determine the credible interval, are shown in the two panes of Figure 6 for summer (JJA). The top pane conjointly shows observed seasonal means, the unprocessed ensemble output, the predictive means and the 50–95% credible intervals for BMA, the lower pane for UW. In both panes the raw ensemble mean is plotted for comparison. One notices that the predictive mean has been bias-corrected by BMA and UW. For UW, the temperature behaves similarly to the ensemble mean, while BMA assigns different weights to different models, and therefore the mean is different from the ensemble mean. Also, the predictive distribution for UW exhibits a slightly larger variance than BMA, as indicated by values in Table 3. The results for all seasons are provided as supplementary material sections S4–S10.

8 | RESULTS

8.1 | Verification

We apply the calibrated processor to predict the observations retrospectively over the historic BW. Figure 7 concisely explains the added value of the Bayesian processor against simply using the spread of an unprocessed ensemble projection as measure of climatic uncertainty. The image shows the probability density functions (pdf) of observed temperatures (blue) and the ensemble predictions of 19 models (light grey). The dark grey curve is the distribution of the ensemble average prior to conditioning. Calibration matches model output with observations through DM (see Equation (13)) and mixes the conditioned distributions by linear weighting. The posterior

![Figure 7](https://example.com/figure7.png)
mixed distribution is shown in red, and represents the predictive density of mean temperature for the BW, BMA weighting (solid), respectively UW (dashed). One notices a slightly more dispersed UW with respect to BMA, ought to the fact that UW considers all model projections to be equally likely, while BMA attributes non-zero likelihood to a limited number of ensemble members. The variance of the predictive density is estimated in accordance with the law of total variance (Mahler and Dean, 2001; Weiss, 2005):

\[
\text{Var}_i(\eta) = E_i \left[ \sigma^2_{\eta_i, \eta} \right] + \text{Var}_i(\mu_{\eta_i, \eta})
\]  

(18)

and equals the mean of the predictive variances for individual years of the BW plus the dispersion of the means. The corresponding values are summarized in Table 3.

Figure 8 demonstrates the effect of conditioning the densities in Figure 7 by visualizing the distributions of the means. The conditioned pdfs of the mean temperature over the window are shuffled together and have all the same mean of observations for each season (respectively, 0.24, 8.79, 18.73, 10.01 °C) because the processor is an unbiased estimator. The mixed distributions of the BMA and UW means are slightly more dispersed than the distributions of the mean for individual models, because the composition leads to a total variance of the mean according to (18) that is higher than the variance of the mean for individual models.

8.2 Performance testing

If the predictive mean is an adequate predictor of observations, the probabilities of observations, conditional on predictions, should be distributed like the Gaussian observations. Gaussianity can be verified graphically by plotting the predictive conditional probabilities of observations,

\[
P(\eta_i : \mu_{\eta_i, \eta}) = \int_{-\infty}^{\eta_i} f(\xi ; \mu_{\eta_i, \eta}) d\xi
\]

(19)

arranged in ascending order versus the Weibull plotting position \(i/(n + 1)\). Figure 9 visualizes the probability plot with dots closely around the bisection line. The Kolmogorov–Smirnov two-sample test at the 95% significance level, which is indicated as the range within the two limiting dashed lines, is successfully passed. The raw

FIGURE 8  BW, four seasons, conditional pdfs of mean temperature given individual models (light grey), pdfs of mean of predictive means over the window, BMA (solid red) and UW (dashed red), pdf of observed mean temperatures (blue). The green pdfs are the temperature means distributions obtained for the same data applying REA (continuous; Giorgi and Mearns, 2002) and REA-T (dashed; Tebaldi et al., 2005)
ensemble mean, used as a predictor, shows a clear departure from the Normal distribution, mainly attributable to the biased unprocessed model predictions, and fails the test.

8.3 Prediction/projection

For prediction/projection we apply Equation (15) in the two prognostic windows PW1 and PW2 by adopting the same weights that were derived for the BW. The weighted predictive mean $\mu_{\eta_i}^{\ast}$ and variance $\sigma_{\eta_i}^{\ast}$ (note the asterisk indicating the future period) for the two posterior distributions is given by (17). The results are visualized in Figure 10 for non-uniform weighting of the RCP4.5 scenario. Figure 11 shows an analogous projection using UW for processing the more pessimistic RCP8.5 scenario. In both figures the 21-year predictive windows PW1 and PW2 are indicated by two vertical lines, while the 27-year BW is shown on the left hand side. The dashed black horizontal lines indicate the mean of means $E_{i} \left[ \mu_{\eta_i} ; \eta_i \right]$. The values of the mean of means and variance of means for the two prognostic

FIGURE 9 BW, four seasons, probability plot testing the normality of residuals for BMA (blue), UW (red) and the raw ensemble (green). Data pairs from BMA and UW, except the unprocessed ensemble, are comfortably within the 95% confidence interval testing positive at the $\alpha = .05$ significance level, represented by the range limited by the two dashed lines.
windows and the selected scenarios are summarized in Table 3 and indicate a clearly increasing trend in terms of mean and variance towards the future. Overall, as one would expect, the UW predictive mean is much more adherent to the raw ensemble mean than BMA in PW1, and tends to converge to the ensemble mean in PW2.

Figure 12 relates to Figures 7 and 8 and shows the distributions of temperature means for individual unconditional models and the pdfs of the predictive mean of means calculated over PW1, PW2 and RCP4.5 respectively RCP8.5. The figure also visualizes the distribution of temperature means for the benchmarking methods REA and REA-T, which we discuss below. Also here the bias-correction and sharpening of the distribution with respect to using the ensemble average (dark grey line) is evident for all PWs.

8.4 | Benchmarking

The present approach is benchmarked against uncertainty assessment methods widely acknowledged in the climatological literature (Giorgi and Mearns, 2002; Tebaldi et al., 2005; Tebaldi and Lobell, 2008; Knutti et al., 2010). We first compute the distributions of temperature differences $\Delta T$ between BW and PW1 or PW2 for RCP4.5, respectively RCP8.5. Given that our distributions are Gaussian, the distribution of $\Delta T$ is also Gaussian:

$$f(\Delta T) = N\left(\mu_{\Delta T}, \sigma_{\Delta T}^2\right)$$

with mean of means and variances of means for the windows BW, PW1 and PW2 given in Table 3. The
covariance is generally considered zero due to uncorrelated retrospective and future model output.

The density $f(\Delta T)$ is compared against those obtained with two published approaches: the deterministic Reliability Ensemble Average (REA; Giorgi and Mearns, 2002) and its probabilistic extension (Tebaldi et al., 2005), which we denote as REA-T. The REA evaluates the weighted average for temperature, by accepting model performance over a historic period as guidance for model weighting. A model is rewarded, if bias and model mean are close to the ensemble mean. The REA-T (Tebaldi et al., 2005) starts out from the original REA and assumes a linear correlation between retrospective and future predictions. For both cases we adopt the probabilistic parameter estimation (Tebaldi et al., 2005) based on Bayesian inference. The parameters are drawn from the posterior distribution with the aid of a Gibbs sampler.

In Figure 13, REA and REA-T results for the indicator distributions are plotted against those obtained by Bayesian distribution mixing and the distribution obtained from the unprocessed ensemble. We notice a considerably smaller variance, that is, higher confidence, on the temperature indicator distributions for REA-T with respect to BMA and UW, while the results of REA are very similar to the much more dispersed pdf obtained from the raw ensemble. The smaller variance of REA-T with respect to all other approaches are due to the underlying assumption of correlation between retrospective and future model predictions, which is known to artificially reduce the variance of the difference distribution (Benjamin and Cornell, 1970).
**FIGURE 12**  Pdfs of temperature means for individual models (light grey) and the ensemble average (dark grey), pdfs of the mean of predictive means for BMA (solid red) and UW (dashed red), Winter (DJF). The unbiased estimation of the mean and sharpening of the density due to conditioning and mixing versus the ensemble average is visible. UW weighting shows a marginally higher dispersion than BMA. The green pdfs are the distributions of means obtained using REA (continuous) and REA-T (dashed) and are close to the pdf of the ensemble average.

**FIGURE 13**  Distributions of the climate change indicator $\Delta T$ for PW1 and PW2, RCP4.5 and RCP8.5, Winter (DJF). The densities refer to the Bayesian processor, BMA (red solid) and UW (red dashed), the REA (green solid) and REA-T (green dashed) and the unprocessed ensemble (blue dotted).
Now, we revert to Figure 8, which displays the comparison with the pdfs of temperature means evaluated using REA and REA-T for the BW. One notices that both density curves (solid and dashed green) are left-biased (MAM, JJA, SON) and right biased for Winter (DJF), as they heavily lean on the (biased) ensemble mean due to the way the likelihood function is formulated. The predictive density obtained with the present approach is effectively bias-corrected thanks to prior conditioning on observations. A sharpening of the pdf for BMA/UW is recognizable for JJA and SON, whereas for the remaining seasons sharpness is comparable among the two approaches. Figure 12 shows the pdfs of temperature means for the two prognostic windows and both RCPs. The corresponding distributions (green solid and dashed lines) strongly resemble the distribution of the model means (dark grey line), which leads us to conclude that the REA or REA-T predictive skill is comparable to simply using the ensemble spread as indicator of uncertainty. Finally, we note that the REA approach is sensible to the prior parameter $\lambda_0$, which we have set equal to the inverse variance of observations (Giorgi and Mearns, 2002). By choosing a larger $\lambda_0$ value one could get a less biased posterior distribution, but we did not see any physical basis for choosing other than the variance of observations.

9 | DISCUSSION AND CONCLUSIONS

The primary use of the proposed Bayesian processor is quantifying climate-related risks and supporting decision-makers in climate impact mitigation. As shown in Section 3, a properly calibrated predictive conditional density describing future uncertainty can lead to considerable cost savings when planning mitigating actions. In climate-related applications, standard conditioning should not be applied, because Earth-system models are not conceived to preserve the temporal correlation but rather to project only the statistical properties of the observed climatic variables into the future. This is possible by assuming the climatic process as weakly stationary over selected time windows that are sufficiently long to allow the estimation of the statistical parameters and at the same time sufficiently short as to allow negligence of trends effects.

Climate projections based on multi-model ensembles are affected by large prognostic uncertainty. This uncertainty encumbers the decision process needed to identify most effective mitigating actions. In the past two decades, different approaches have been developed to improve the uncertainty estimation and to reduce it. These approaches are based on formulations with following characteristics: The spread of the model ensemble is taken as a measure of future climate uncertainty without considering prior conditioning on observations (Räisänen and Palmer, 2001). Later on Bayesian parameter inference on the stochastic dependency between models and observation is used (Giorgi and Mearns, 2003; Tebaldi et al., 2005; Buser et al., 2009). The posterior parameters are obtained by direct sampling from the posterior distribution. Commonly, the likelihood of the Bayesian formulation accounts for the bias between model projections and observations.

No inference-based and implemented Bayesian approach formulates climate uncertainty in terms of a predictive probability distribution, which estimates the uncertainty of a retrospective or future reality given climatic projections. Very few approaches acknowledge and address the issue of non-stationarity of the climatic process. As alternative for the estimation of the prognostic uncertainty we propose a Bayesian distribution mixing method, which ties multiple model outputs to observations over a historic period, while assuming preservation of stochastic model-observation dependencies into the future. Because of the non-stationarity of the Earth climate, assumptions of weak stationarity of the process within time windows of appropriate length need to be made that allow assessing the average probability distribution over the window. The predictive mean and variance are obtained by combining the conditional distributions for all models through a Bayesian weighted average of distributions, either based on uniform weighting or on weights estimated by maximizing the probability of observations over the BW.

We also recall that here we aggregate ensemble members and estimate the predictive density at each time step first and time-average over the windows afterwards, while the reported inference-based approaches, among which REA and REA-T, do first the temporal averaging over the window and then estimate a single set of distribution parameters for the combined means. Since for any data we can always apply a transformation that makes the variable statistics Gaussian, the outcome can be expressed in terms of the mean and the variance of the predictive density, the probability distribution of observations given multi-model ensemble output. The predictive mean is an unbiased estimator of past observations and future, yet unobserved reality, whereas the predictive variance has been considerably reduced with respect to the commonly adopted simple multi-model treatment, thus demonstrating that the Bayesian mixture distribution approach has an improved information content.

The predictive variance reduction leads to a more precise estimation of losses due to climate-induced impacts. As a consequence, it can support effective investment decisions and probabilistic planning on possible mitigating actions and climate adaptation. A comparison against the
performance of commonly-adopted uncertainty assessment approaches shows a net benefit of the proposed Bayesian mixture of multi-model distributions, with clear gains in terms of a sharper predictive distribution. The enhanced prognostic capability can lead to more reliable impact analysis with accuracy gains that lie in the order of multiple tenths percent, as we have shown in the example discussed in Section 3. For governments or large lending institutions involved in financing climate change mitigation and adaptation actions, the adoption of more accurate and reliable uncertainty assessment tools means long-term cost savings. This is especially true for very large adaptation projects like coastal or riverine protection that can reach up to multiple times national GDP.

ACKNOWLEDGEMENTS
We would like to thank Claudia Tebaldi for sharing the R-code of her work (Tebaldi et al., 2005). We also acknowledge ECMWF for making ERA5 data available and the CMIP5 community for providing access to the climate output ensembles. Open Access funding enabled and organized by Projekt DEAL.

ORCID
Paolo Reggiani https://orcid.org/0000-0001-9989-9885

REFERENCES
Abramowitz, G., Herger, N., Gutmann, E., Hammerling, D., Knutti, R., Leduc, M., Lorenz, R., Pincus, R. and Schmidt, G.A. (2019) ESD reviews: model dependence in multi-model climate ensembles: weighting, sub-selection and out-of-sample testing. Earth System Dynamics, 10(1), 91–105.
Aven, T. and Renn, O. (2015) An evaluation of the treatment of risk and uncertainties in the IPCC reports on climate change. Risk Analysis, 35, 701–712.
Benjamin, J.R. and Cornell, C.A. (1970) Probability, Statistics and Decision for Civil Engineers. In: Academic Press. New York: McGraw Hill.
Berger, J. (1985) Statistical decision theory and Bayesian analysis. New York: Springer-Verlag.
Bernardo, J.M. and Smith, A.F.M. (1994) Bayesian Theory. Chichester, England: John Wiley and Sons.
Bhat, K.S., Haran, M., Terando, A. and Keller, K. (2011) Climate projections using Bayesian model averaging and space–time dependence. Journal of Agricultural, Biological, and Environmental Statistics, 16(4), 606–628.
Boë, J., Terray, L., Habets, F. and Martin, E. (2007) Statistical and dynamical downscaling of the seine basin climate for hydro-meteorological studies. International Journal of Climatology, 27(12), 1643–1655.
Bröcker, J. and Smith, L.A. (2007) Increasing the reliability of reliability diagrams. Monthly Weather Review, 22, 651–662.
Buizza, R. (2019) Introduction to the special issue on “25 years of ensemble forecasting”. Quarterly Journal of the Royal Meteorological Society, 145(S1), 1–11.
Buizza, R. and Leutbecher, M. (2015) The forecast skill horizon. Quarterly Journal of the Royal Meteorological Society, 141(693), 3366–3382.
Buser, C.M., Künsch, H.R., Lüthi, D., Wild, M. and Schär, C. (2009) Bayesian multi-model projection of climate: bias assumptions and interannual variability. Climate Dynamics, 33, 849–868.
Clark, M.P., Wilby, R.L., Gutmann, E.D., Vano, J.A., Gangopadhyay, S., Wood, A.W., Crop, H.J., Prudhomme, C., Arnold, J.R. and Brekke, L.D. (2016) Characterizing uncertainty of the hydrologic impacts of climate change. Current Climate Change Reports, 2(2), 55–64.
Collins, M., Booth, B.B.B., Harris, G.R., Murphy, J.M., Sexton, D.M. H. and Webb, M.J. (2006) Towards quantifying uncertainty in transient climate change. Climate Dynamics, 27, 127–147.
De Groot, M. (1970) Optimal Statistical Decisions. New York: John Wiley Sons.
Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society Series B, 39(1), 1–38.
Déqué, M., Rowell, D.P., Lüthi, D., Giorgi, F., Christensen, J.H., Rockel, B., Jacob, D., Kjellström, E., de Castro, M. and van den Hurk, B. (2007) An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections. Climatic Change, 8(1), 53–70.
Deser, C., Phillips, A., Bourdette, V. and Teng, H. (2012) Uncertainty in climate change projections: the role of internal variability. Climate Dynamics, 38, 527–546.
Draper, D. and Krnjajić, M. (2010), Calibration results for Bayesian model specification, Technical report, Dept. of Applied Mathematics and Statistics, University of California, Santa Cruz.
Drusch, M., Wood, E.F. and Gao, H. (2005) Observation operators for the direct assimilation of TRMM microwave imager retrieved soil moisture. Geophysical Research Letters, 32 (L15403). https://agupubs.onlinelibrary.wiley.com/doi/epdf/10.1029/2005GL023623.
Economou, T., Stephenson, D.B., Rougier, J.C., Neal, R.A. and Mylne, K.R. (2016) On the use of bayesian decision theory for issuing natural hazard warnings. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 472 (2194), 20160295.
Eyring, V., Bony, S., Meehl, G.A., Senior, C.A., Stevens, B., Stouffer, R. J. and Taylor, K.E. (2012) Overview of the coupled model intercomparision project phase 6 (CMIP6) experimental design and organization. Geoscientific Model Development, 9, 1937–2016.
Fowler, H.J., Blenkinsop, S. and Tebaldi, C. (2007) Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. International Journal of Climatology, 27(12), 1547–1578.
Furrer, R., Knutti, R., Sain, S., Nychka, D. and Meehl, G.A. (2007) Spatial patterns of probabilistic temperature change projections from a multivariate Bayesian analysis of climate model output. Geophysical Research Letters, 34(L06711). https://agupubs.onlinelibrary.wiley.com/doi/epdf/10.1029/2006GL027754.
Giorgi, F. and Francisco, R. (2000) Uncertainties in regional climate change prediction: a regional analysis of ensemble simulations with the HADCM2 coupled AOGCM. Climate Dynamics, 16(2), 169–182.
Giorgi, F. and Mearns, L.O. (2002) Calculation of average, uncertainty range and reliability of regional climate changes from AOGCM simulations via the ‘reliability ensemble averaging’ (REA) method. Journal of Climate, 15, 1141–1158.

Giorgi, F. and Mearns, L.O. (2003) Probability of regional climate change calculated using the reliability ensemble average (REA) method. Geophysical Research Letters, 30, 1629–1632.

Gneiting, T., Balabdaoui, F. and 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. and Raftery, A. E. (2004), Tech. rep. 463: Strictly proper scoring rules, prediction, and estimation, Technical report, Dept. of Statistics, University of Washington, Seattle, WA. Available at: www.stat.washington.edu/tech.reports

Gneiting, T., Raftery, A.E., Westveld, A.H. and Goldman, T. (2005) Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. Monthly Weather Review, 133(5), 1098–1118.

Hagedorn, R., Doblas-Reyes, F.J. and Palmer, T.N. (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting—i. basic concept. Tellus A, 57(3), 219–233.

Hamill, T.M. and Whitaker, J.S. (2006) Probabilistic quantitative precipitation forecasts based on reforecast analogs: theory and application. Monthly Weather Review, 134, 3209–3229.

Hashino, T., Bradley, A.A. and Schwartz, S.S. (2007) Evaluation of bias-correction methods for ensemble streamflow volume forecasts. Hydrology and Earth System Sciences, 11(2), 939–950.

Hawkins, E. and Sutton, R. (2011) The potential to narrow uncertainty in projections of regional precipitation change. Climate Dynamics, 37, 407–418.

Hersbach, H., Bell, B. and Berrisford, P.e.a. (2020) The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049.

Hirsch, R.M. (1982) A comparison of four streamflow record extension techniques. Water Resources Research, 18(4), 1081–1088.

Ines, A.V.M. and Hansen, J.W. (2006) Bias correction of daily GCM rainfall for crop simulation studies. Agricultural and Forest Meteorology, 138, 44–53.

IPCC (2013), Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. Available at: www.climatechange2013.org

Kharin, V.V., Boer, G.J., Merryfield, W.J., Scinocca, J.F. and Lee, W.-S. (2012) Statistical adjustment of decadal predictions in a changing climate. Geophysical Research Letters, 39, L23825. https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2012GL052647.

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J. and Meehl, G.A. (2010) Challenges in combining projections from multiple climate models. Journal of Climate, 23(10), 2739–2758.

Knutti, R. and Sedláček, J. (2012) Robustness and uncertainties in the new CMIP5 climate model projections. Nature Climate Change, 3, 369–373.

Knutti, R., Sedláček, J., Sanderson, B.M., Lorenz, R., Fischer, E.M. and Eyring, V. (2017) A climate model projection weighting scheme accounting for performance and interdependence. Geophysical Research Letters, 44, 1909–1918.

Krzyżtofowicz, R. (1999) Bayesian theory of probabilistic forecasting via deterministic hydrologic model. Water Resources Research, 35(9), 2739–2750.

Lorenz, E.N. (1963) Deterministic non-periodic flow. Journal of the Atmospheric Sciences, 20, 130–141.

Mahler, H.C. and Dean, C.G. (2001) Chapter 8: Credibility. In: Foundations of Casualty Actuarial Science, 4th edition. N. Fairfax Drive, Suite 250, Arlington, VA: Casualty Actuarial Society, pp. 525–526.

Maraun, D. (2013) Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. Journal of Climate, 26, 2137–2143.

Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979) Multivariate Analysis. London: Probability and Mathematical Statistics, Academic Press.

Matheson, J.E. and Winkler, R.L. (1976) Scoring rules for continuous probability distributions. Man of Science, 22, 1087–1096.

Meehl, G.A., Boer, G.J., Covey, C., Latif, M. and Stouffer, R.J. (2000) The coupled model intercomparison project (CMIP). Bulletin of the American Meteorological Society, 81, 313–318.

Meehl, G.A., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J.F.B., Stouffer, R.J. and Taylor, K.E. (2007) The WCRP CMIP3 multimodel dataset: a new era in climate change research. Bulletin of the American Meteorological Society, 88, 1383–1394.

Palmer, T.N. (2000), Predicting uncertainty in forecasts of weather and climate, Technical Report Rep. Prog. Phys. 63, IOP Science. Available at: https://iopscience.iop.org/article/10.1088/0034-4885/63/2/201

Palmer, T.N. and Räisänen, J. (2002) Quantifying the risk of extreme seasonal precipitation events in a changing climate. Nature, 415, 512–514.

Piani, C., Frame, D.J., Stainforth, D.A. and Allen, M.R. (2005) Constraints on climate change from a multi-thousand member ensemble of simulations. Geophysical Research Letters, 32 (L23825). https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2005GL024452.

Raftery, A.E., Gneiting, T., Balabdaoui, F. and Polakowski, M. (2005) Using Bayesian model averaging to calibrate forecast ensembles. Monthly Weather Review, 133(5), 1155–1174.

Raftery, A. E., Madigan, D., A., J. and Hoeting (1997), ‘Bayesian model averaging for regression models’, Journal of the American Statistical Association, 92, 179–191.

Räisänen, J. (1997) Objective comparison of patterns of CO2 induced climate change in coupled GCM experiments. Climate Dynamics, 13, 197–211.

Räisänen, J. and Palmer, T.N. (2001) A probability and decision-model analysis of a multimodel ensemble of climate change simulations. Journal of Climate, 14, 3212–3226.

Rougier, J. (2007) Probabilistic inference for future climate using an ensemble of climate model evaluations. Climatic Change, 81 (3–4), 247–264.

Rougier, J. and Goldstein, M. (2014) Climate simulators and climate projections. Annual Review of Statistics and Its Application, 1 (1), 103–123.

Rougier, J. and Sexton, D.M.H. (2007) Inference in ensemble experiments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 365(1857), 2133–2143.
Sansom, P.G., Ferro, C.A.T., Stephenson, D.B., Goddard, L. and Mason, S.J. (2016) Best practices for postprocessing ensemble climate forecasts. Part I: selecting appropriate recalibration methods. *Journal of Climate*, 29(20), 7247–7264.

Smith, R.L., Tebaldi, C., Nychka, D.W. and Mearns, L.O. (2009) Bayesian modeling of uncertainty in ensembles of climate models. *Journal of the American Statistical Association*, 104, 97–116.

Stephenson, D.B., Collins, M., Rougier, J.C. and Chandler, R.E. (2012) Statistical problems in the probabilistic prediction of climate change. *Environmetrics*, 23(5), 364–372.

Stoner, A.M.K., Hayhoe, K., Yang, X. and Wuebbles, D.J. (2013) An asynchronous regional regression model for statistical downscaling of daily climate variables. *International Journal of Climatology*, 33(11), 2473–2494.

Strzpekek, K., McCluskey, A., Boehlert, B., Jacobsen, M. and IV, C. F. (2011), *Climate variability and change: a basin scale indicator approach to understanding the risk to water resources development and management*, Technical report, The World Bank, 1818 H Street, NW Washington, DC 20433 USA.

Taylor, K.E., Stouffer, R.J. and Meehl, G.A. (2012) Summary of the CMIP5 experiment design. *Bulletin of the American Meteorological Society*, 93, 485–498.

Tebaldi, C. and Knutti, R. (2007) The use of the multi-model ensemble in probabilistic climate projections. *Journal of the American Statistical Association*, 365A, 2053–2075.

Tebaldi, C. and Lobell, D.B. (2008) Towards probabilistic projections of climate change impacts on global crop yields. *Geophysical Research Letters*, 35(L08705). https://agupubs.onlinelibrary.wiley.com/doi/epdf/10.1029/2008GL033423.

Tebaldi, C., Mearns, L.O., Nychka, D. and Smith, R.L. (2004) Regional probabilities of precipitation change: a Bayesian analysis of multimodel simulations. *Geophysical Research Letters*, 31(L24213). https://agupubs.onlinelibrary.wiley.com/doi/epdf/10.1029/2004GL021276.

Tebaldi, C. and Sansò, B. (2009) Joint projections of temperature and precipitation change from multiple climate models: a hierarchical Bayesian approach. *Journal. Royal Statistical Society*, 172A, 83–106.

Tebaldi, C., Smith, R.W., Nychka, D. and Mearns, L. (2005) Quantifying uncertainty in projections of regional climate change: a Bayesian approach to the analysis of multimodel ensembles. *Journal of Climate*, 18, 1524–1540.

Themessl, M.J., Gobiet, A. and Leuprecht, A. (2011) Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. *International Journal of Climatology*, 31(10), 1530–1544.

Varis, O., Kajander, T. and Lemmeli, R. (2004) Climate and water: from climate models to water resources management and vice versa. *Climatic Change*, 66(3), 321–344.

Vidale, P. L., Luthi, D., Frei, C., Seneviratne, S. I. and Schar, C. (2003) Predictability and uncertainty in a regional climate model. *Journal of Geophysical Research: Atmospheres*, 108(D18). https://agupubs.onlinelibrary.wiley.com/doi/epdf/10.1029/2002JD002810.

Von Storch, H. and Zwiers, F.W. (1999) *Statistical Analysis in Climate Research*. Cambridge, UK: Cambridge University Press.

Weigel, A.P., Knutti, R., Liniger, M.A. and Appenzeller, C. (2010) Risks of model weighting in multimodel climate projections. *Journal of Climate*, 23(15), 4175–4191.

Weiss, N.A. (2005) *A Course in Probability*. Boston: Addison-Wesley.

Whitaker, S. and Hamill, T.M. (2006) Probabilistic quantitative precipitation forecasts based on reforecast analogs: theory and application. *Monthly Weather Review*, 134(11), 3209–3229.

Wilks, D.S. (1995) *Statistical methods in the atmospheric sciences*, 2nd edition. Academic Press.

Wood, A.W., Maurer, E.P., Kumar, A. and Lettenmaier, D.P. (2002) Long-range experimental hydrologic forecasting for the eastern United States. *Journal of Geophysical Research*, 107(D20), 1–13.

Wood, A.W. and Schaake, J.C. (2008) Correcting errors in streamflow forecast ensemble mean and spread. *Journal of Hydrometeorology*, 9(1), 132–148.

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Reggiani P, Todini E, Boyko O, Buizza R. Assessing uncertainty for decision-making in climate adaptation and risk mitigation. *Int J Climatol.*, 2021;41:2891–2912. https://doi.org/10.1002/joc.6996