Design and evaluation of calibrated and seamless ensemble weather forecasts for crop protection applications

Ivana Aleksovska, Laure Raynaud, Robert Faivre, François Brun, Marc Raynal

Design and Evaluation of Calibrated and Seamless Ensemble Weather Forecasts for Crop Protection Applications

IVANA ALEKSOVSKA,[a] LAURE RAYNAUD,[a] ROBERT FAIVRE,[b] FRANÇOIS BRUN,[c] AND MARC RAYNAL[d]

[a] *CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France*
[b] *INRAE, Castanet-Tolosan, France*
[c] *ACTA-French Technical Institute, Castanet-Tolosan, France*
[d] *Institut Français de la Vigne et du Vin (IFV), UMT SEVEN, Bordeaux, France*

(Manuscript received 24 July 2020, in final form 24 February 2021)

ABSTRACT: Agriculture is a highly weather-dependent activity, and climatic conditions impact both directly crop growth and indirectly diseases and pest developments causing yield losses. Weather forecasts are now a major component of various decision-support systems that assist farmers to optimize the positioning of crop protection treatments. However, properly accounting for weather uncertainty in these systems still remains a challenge. In this paper, three global and regional ensemble prediction systems (EPSs), covering different spatiotemporal scales, are coupled to a temperature-driven developmental model for grapevine moths in order to provide probabilistic forecasts of treatment dates. It is first shown that a parametric postprocessing of the EPSs significantly improves the prediction of treatment dates. Anticipating the need for phytosanitary treatments also requires seamless weather forecasts from the next hour to subseasonal time scales. An approach is presented to design seamless ensemble forecasts from the combination of the three EPSs used. The proposed method is able to leverage the increased performance of high-resolution EPS at short ranges, while ensuring a smooth transition toward larger-scale EPSs for longer ranges. The added value of this seamless integration on agronomic predictions is, however, difficult to assess with the current experimental setup. Additional simulations over a larger number of locations and years may be required.

KEYWORDS: Bias; Ensembles; Forecast verification/skill; Model output statistics; Agriculture; Decision support

1. Introduction

Numerical weather prediction (NWP) models are now widely used for operational forecasting over a wide range of spatiotemporal scales (Bauer et al. 2015). In addition, the inherent uncertainty of NWP forecasts has motivated the development of probabilistic forecasting based on ensemble prediction systems (EPSs) (Palmer 2019). Typically, a set of perturbed forecasts is designed in order to account for the different sources of error, regarding for instance initial conditions and the model formulation. Originally developed for global models (Molteni et al. 1996; Leutbecher and Palmer 2008), such systems have more recently become popular at the convective scale for short ranges (Bouttier and Raynaud 2018a; Hagelin et al. 2017). Several downstream applications also started to use these probabilistic forecasts (Fundel et al. 2019), including in particular agriculture (Christ et al. 2015), power grid management (Pinson et al. 2009), aircraft routing (Arbogast et al. 2015), hydrology (Bellier et al. 2017, 2018), and wildfires (Worsnop et al. 2019).

In agriculture, pests develop according to weather conditions (Moyer et al. 2016). Crop management operations such as applications of phytosanitary treatments are common practices to control diseases and pests, but they need to be properly positioned and applied according to some criteria that depend on the development of the pest population (Cooke et al. 2006). The precision in their positioning is even more important for example for biocontrol solutions, which use living organisms to prevent or reduce damages caused by pests, because of their lower levels of efficiency (Shipp and Clarke 1999). For that purpose, different decision support tools (DSTs) have been developed to predict the evolution of pest dynamics and support farmers in their decision-making (Pertot et al. 2017). They are complementary to field observations, and they provide an opportunity to anticipate the situation and to access variables that are difficult to observe. Most of these tools require skillful extended-range weather forecasts, which are most of the time given by deterministic models (Olatinwo et al. 2011, 2012). Probabilistic forecasts could provide valuable additional information to support decision-making, and hence contribute to a sustainable use of treatments through minimal and timely applications.

However, raw ensemble weather forecasts may have systematic errors regarding bias and dispersion, which can have a negative impact on the subsequent DSTs outputs. Several statistical postprocessing methods have been proposed to reduce these errors, that basically aim at estimating a statistical relationship between the observations and the predicted probability distributions. Once this relationship is established, it can be used to generate calibrated probabilistic forecasts from ensemble outputs. Postprocessing methods can be divided into two categories: parametric and nonparametric approaches. Parametric approaches are based on the assumption that the calibrated predictive distribution of a weather variable is given by a parametric density function, whose parameters

Corresponding author: Ivana Aleksovska, ivana.aleksovska@meteo.fr

DOI: 10.1175/WAF-D-20-0128.1

depend on ensemble members and are fitted over a training period (Gneiting 2014). Standard parametric approaches include nonhomogenous regression (NR)/ensemble model output statistics (EMOS) methods (Gneiting and Raftery 2005) and Bayesian model averaging (BMA) methods (Raftery et al. 2005). On the other hand, nonparametric methods remove any assumption on the variable distribution. One example is the quantile regression forest proposed by Taillardat et al. (2019).

Calibrated ensemble members can be generated from the postprocessed distributions, with a random or quantile sampling for instance. However, since the calibration is performed independently at each location and lead time, the spatiotemporal structure of ensemble members is lost. These dependencies, which are crucial to many users, can be recovered with copula methods (Schefzik 2011), that perform a reordering of the sampled values based on a given dependence structure. For instance, the ensemble copula coupling (ECC) approach (Schefzik et al. 2013) uses the rank structure of the raw ensemble members, while the Schaake shuffle approach (Clark et al. 2004) uses the rank structure of past observations.

As mentioned previously, DSTs rely on seamless weather trajectories, from the next hour to several days ahead, in order to make early decisions about crop protection (Calanca et al. 2011). Seamless members can be obtained from an extended-range EPS such as the ECMWF EPS (Pappenberger et al. 2013), but it may also be interesting to integrate higher-resolution EPSs available on shorter time scales for a better representation of small-scale phenomena. To the best of the authors' knowledge, seamless probabilistic multisource forecasting is still an open question. In this paper, the seamless combination of three EPSs with different spatiotemporal scales will be considered as a simple concatenation problem, which aims at matching short-range forecasts from a fine-resolution EPS to coarse-resolution forecasts from larger-scale EPSs. A similar approach was adopted by Wetterhall and Di Giuseppe (2018) in order to merge subseasonal to seasonal ECMWF ensemble forecasts for hydrological applications. Seamless integration methods should address two major challenges: ensuring smooth transitions (i.e., without temporal discontinuities) between the different EPSs considered, and providing enhanced performances with respect to the reference ECMWF EPS, at least at short ranges.

The goal of this paper is to examine the impacts of postprocessing and seamless EPS merging on both weather and agronomic outputs. Standard EMOS and ECC methods are used to generate calibrated distributions, while different strategies for designing seamless ensemble forecasts are proposed and evaluated in a univariate setting. The outline is as follows. Section 2 describes the NWP models and the agronomic use case, as well as the postprocessing and seamless integration methods. Section 3 presents the characteristics and performances of seamless calibrated ensemble forecasts on 2-m temperature, while the impact on agronomic outputs is discussed in section 4. Finally, section 5 provides conclusions and future works.

2. Material and methods

a. Ensemble weather forecasts

For the purpose of this work, three operational EPSs are considered: the convective-scale AROME-EPS developed at Météo-France, which covers a western Europe domain (Raynaud and Bouttier 2016), the global French ARPEGE-EPS (Descamps et al. 2015), and the global IFS-EPS of the European Centre for Medium-Range Weather Forecasts (ECMWF) (Palmer 2019). These models account for both initial and model uncertainties, and the AROME-EPS uses selected ARPEGE-EPS members as lateral boundary conditions (Bouttier and Raynaud 2018b). The main characteristics of these EPSs are given in Table 1.

b. Ensemble model output statistics

For postprocessing ensemble weather forecasts, the NR/EMOS univariate statistical method is used. Let us denote y the weather variable of interest and $x_1, \ldots, x_N$ the N forecasts corresponding to the N members of a given EPS. The NR/EMOS method assumes that the predictive distribution of y is a parametric density function p whose parameters depend on ensemble output statistics (Gneiting and Raftery 2005):

$$Y|x_1, \ldots, x_N \sim p(Y|x_1, \ldots, x_N). \tag{1}$$

The parameter of interest in this study is the 2-m temperature, which is commonly postprocessed using a Gaussian distribution (Gneiting et al. 2005). Its mean μ and variance σ^2 are assumed affine functions of the ensemble mean $\bar{x}$ and the ensemble variance s^2, respectively:

$$\mu = \alpha_0 + \alpha_1 \bar{x}, \quad \sigma^2 = \beta_0^2 + \beta_1^2 s^2,$$

where coefficients $\alpha_0, \alpha_1, \beta_0$, and β_1 are estimated from pairs of past raw forecasts-observations, by minimizing the mean continuous ranked probability score (CRPS) of the calibrated distribution over a training period, as proposed by Gneiting et al. (2005). In the remainder of the paper, the considered training period is a sliding window of 60 days prior to the EPS starting date (Hemri et al. 2014). The observations used are 2-m temperature measurements from the French real-time weather observation network.

A finite N-sample of calibrated members $\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_N$ is then obtained by randomly sampling the postprocessed distribution at each lead time. These members are finally reordered with the ECC matching procedure according to the raw ensemble template (Schefzik et al. 2013). This step ensures that the new trajectories present a realistic temporal consistency across lead times. Spatial and multivariate dependencies could also be addressed with the same approach; however, they are not considered in this study since the agronomic use case considered only requires univariate forecasts at independent locations.

c. Seamless ensemble forecasts

In this section, a simple design of seamless ensemble members, which performs a concatenation of the systems AROME-EPS, ARPEGE-EPS, and IFS-EPS over different spatiotemporal

TABLE 1. Characteristics of the three EPSs used (corresponding to their operational configurations over the period 1 Jun 2018–31 May 2019).

Characteristics model	AROME-EPS	ARPEGE-EPS	IFS-EPS
Size	12 perturbed members	34 perturbed members + 1 control forecast	50 perturbed members + 1 control forecast
Horizontal resolution	2.5 km	Stretched grid, 10 km over France	18 km
Vertical resolution (levels)	90	90	91
Initial time of forecast	2100 UTC	1800 UTC	0000 UTC
Lead times	51 h	102 h	360 h
Output frequency	1 h	3 h	3 h up to 144 h, 6 h beyond
Dynamical core	Nonhydrostatic	Hydrostatic	Hydrostatic
Archive grid resolution (regular lat/lon)	0.025° × 0.025°	0.1° × 0.1°	0.5° × 0.5°

scales, is described. Given the characteristics of these EPS, the seamless forecasts are started at 0000 UTC and constructed with the following pragmatic approach: the seamless ensemble is composed of the 12 AROME-EPS members only over the forecast period 0–48 h; for longer ranges each AROME-EPS member is matched to an ARPEGE-EPS and/or IFS-EPS member according to some assignment rules described below. This configuration is motivated by the higher performance of the AROME-EPS for its timeframe.

The simplest assignment is to randomly select, without repetition, members in the target sample, which contains members of ARPEGE-EPS and IFS-EPS, similarly to (Wetterhall and Di Giuseppe 2018). This method, hereafter denoted random neighbor (RN), will be used as a naive benchmark for the evaluation of more advanced strategies.

Another approach is to minimize the distance between matched forecasts in order to ensure a smooth transition between lead times. Given an appropriate distance measure d, d_{ij} denotes the distance between the ith member of sample 1 (with size N) and the jth member of sample 2 (the target sample, with size M). The nearest neighbor approach (hereafter denoted NN) assigns to each member of sample 1 the closest member of sample 2:

$$j_i^\star = \underset{j}{\operatorname{argmin}}\{d_{ij}\}, \quad i = 1, \ldots, N.$$

This method thus performs an individual optimization for each member. One drawback is that it allows for member replication, several members of sample 1 can be matched to the same member of sample 2, thus leading to an effective ensemble size smaller than N.

On the other hand, the Hungarian method (hereafter denoted HU) uses the Kuhn–Munkres algorithm (Kuhn 1955) to find the optimal bijective mapping between two samples, by minimizing the total distance of the assignment:

$$j^\star = \underset{j}{\operatorname{argmin}}\left\{\sum_{i=1}^{N} d_{ij}\right\}.$$

Contrary to the nearest neighbor approach, this method does not allow for member repetitions.

Two important aspects of this configuration may be mentioned. First, the size of the seamless ensemble is constrained by the AROME-EPS size (12 members), second the matching procedure between high-resolution and lower-resolution members is applied one or two times, depending on the selected members at 48 h. If an AROME-EPS member is connected at 48-h lead time to an IFS-EPS member, the matching procedure is only applied once for that member because IFS provides forecasts up to 15 days ahead. If it is connected to an ARPEGE-EPS member then the matching procedure will be repeated because ARPEGE provides forecast up to 96 h only. This second assignment will then perform the connection between this ARPEGE-EPS member and an IFS-EPS member.

Since the agronomic use case considers 2-m temperature forecasts at independent locations, the merging of EPS members applies to time series. The metric d chosen is then the dynamic time warping (DTW) distance, initially proposed by Berndt and Clifford (1994), since it is well suited for measuring the similarity between two time series. The DTW performs an optimal nonlinear alignment between the two sequences, which can provide a more intuitive similarity measure than the Euclidian distance since it allows for similar shapes to match even if they are out of phase. The type of the warping window, as well as its width, which controls the amount of temporal shift allowed, are tuning parameters. In the following, the Sakoe–Chiba window is used with a width of one hour (Sakoe et al. 1990). The DTW is implemented using the R algorithm proposed in (Giorgino et al. 2009) and applied over a period W prior to the merging time t_{merge}. This matching criterion is thus only based on the information prior to the connection time. Other strategies using information posterior to the connection time could also be considered but they are beyond the scope of this paper.

The three assignment procedures RN, NN, and HU are applied at $t_{\text{merge}} = 48$ h using calibrated AROME-EPS members as sample 1 and calibrated ARPEGE-EPS+IFS-EPS members as sample 2. If a second assignment is necessary it is performed at $t_{\text{merge}} = 96$ h, the ARPEGE-EPS members selected at 48 h are used as sample 1 and the remaining IFS-EPS members (i.e., those not chosen at 48 h) are used as sample 2. Note that all members, including the control members of both models, are considered for the matching.

d. *Probabilistic verification scores*

The verification of ensemble forecasts is performed with a comprehensive set of standard metrics, including bias and

root-mean-squared error (RMSE) of the ensemble mean and probabilistic scores presented below. Note that observation errors are not accounted for in the score computation.

1) Continuous ranked probability score

The CRPS is well suited to the verification of probabilistic forecasts of a continuous variable (Candille and Talagrand 2005; Candille et al. 2007). Let us denote F the cumulative distribution function (CDF) associated with an ensemble forecast and $\mathbf{1}_{\{y \le x\}}$ the CDF of a scalar observation y, the CRPS is defined as the quadratic difference between the forecast and observed CDFs:

$$\mathrm{CRPS}(F, y) = \int_{\mathbb{R}} [F(x) - \mathbf{1}_{\{y \le x\}}]^2 \, dx. \tag{2}$$

The CRPS measures the overall performance of an ensemble forecast, the smaller the score, the better the performance.

2) Receiver operating characteristics

The receiver operating characteristics (ROC) curve is used to determine how well a forecast discriminates between events and nonevents (Hanley and McNeil 1982). Given a binary event, the ROC curves plot the hit rate against the false alarm rate using increasing probability thresholds to make the yes/no decision. The area under the ROC curve (denoted hereafter AUC) summarizes the forecast skill. An area of 0.5 indicates no skill, the closer the AUC is to 1 the better the forecast skill.

3) Reliability diagram

For a given binary event, the reliability diagram plots the observed frequency of the event against the binned predicted probabilities. A forecast is reliable if the observed and predicted probabilities are consistent, in other words, if the reliability diagram is close to the diagonal.

4) Observation coverage

The observation coverage measures the percentage of observations that fall in the predicted range. The expected observation coverage percentage for a perfectly reliable ensemble of size N is calculated as $[(N - 1)/(N + 1)] \times 100\%$.

Statistical significance of scores differences between two forecasts is assessed with the Wilcoxon signed-rank test (Wilcoxon et al. 1970). Wilcoxon tests are performed using the significant level $\alpha = 0.05$.

e. *The agronomic use case*

In the French vineyard, two species of insects of the Lepidoptera's order, *Eupoecilia ambiguella* and *Lobesia botrana*, known as grapevine moth, are responsible for quantitative and qualitative damages (Thiéry et al. 2013). These insects evolve in four successive stages of development and may have several reproduction cycles, usually three and up to five per year according to climatic conditions and mainly temperatures. Second and third cycles usually are the most critical in terms of damages. To control grapevine moth, winegrowers have three types of treatments at their disposal: chemical products, biocontrol tools (trichograms and *Bacillus thuringiensis* toxins) (Thiéry et al. 2018) or sexual confusion. The correct positioning of the treatment depends on the mode of action of its active ingredient in response to the target (eggs, larvae, etc.). Since the number and precise timing of the development cycles are highly dependent on weather conditions and can be difficult to predict, mathematical models are often used in addition to field monitoring to describe their evolution and predict their attainment.

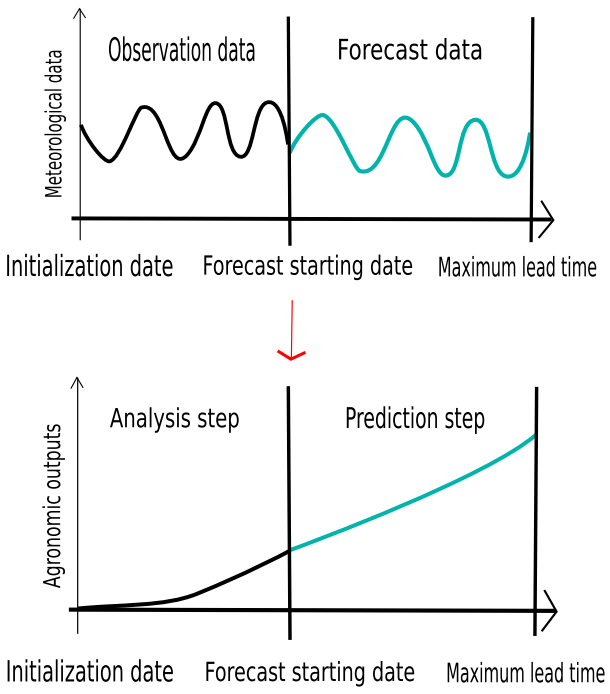

FIG. 1. Schematic representation of the analysis and prediction steps. During the analysis step the agronomic model is run with observation data from an initialization date to the forecast starting date. The model is then run using weather forecasts from the forecast starting date to the maximum lead time.

In this work we consider the EVA model described (in French) by Chavent (1983), and operationally used by the French Institute for Vine and Wine. EVA is a phenological model based on the thermal conditions of the grapevine moth development cycles. From the accumulation of the hourly temperature, the model calculates the proportion and accumulation of individuals at each stage over time. When it reaches a certain level, a fraction of the population moves on to the next stage. The timing of treatment is based on a developmental decision threshold. Based on discussions with experts, the two thresholds relevant to trigger the positioning of the treatment on eggs accumulation are commonly fixed at 2% for ovicide chemical treatments and 15% for biocontrol treatments (due to the lower persistence of the product).

EVA simulations consist of an analysis step and a prediction step. In the analysis step the model is run using time series of observed hourly temperatures from an initialization date to the forecast starting date. This procedure then provides the initial conditions for the prediction. In the prediction step the model is integrated using hourly temperature forecasts. A schematic representation for the analysis and the prediction steps is shown on Fig. 1. Ensemble EVA simulations are obtained by

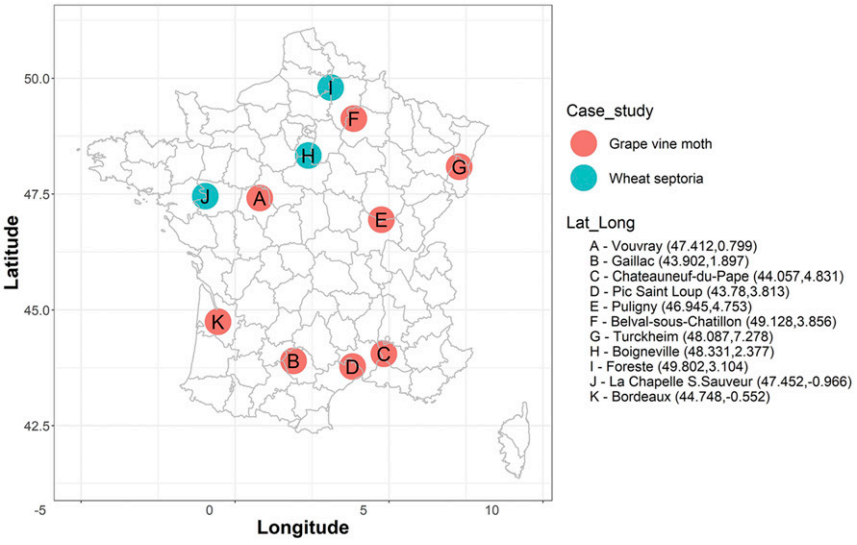

FIG. 2. Locations used for the weather and agronomic simulations. The red points correspond to vineyards, and the green points correspond to wheat fields. The latitude–longitude coordinates of each point are also provided.

running the prediction step with temperature forecasts from EPS members or from the seamless members, while initial conditions are the same for all EVA members. Note that hourly temperature is a direct output from AROME-EPS, while ARPEGE-EPS and IFS-EPS outputs are linearly interpolated to a 1-h frequency.

A probabilistic evaluation of the EVA ensemble forecasts is performed with the same measures used in NWP, applied to the predicted distribution of recommended treatment dates. It is important to note that an observation-based treatment date is calculated by running the EVA model with observed temperature over the entire season. This is the standard operational practice because field observations of the grapevine moth's stage series are often difficult to obtain. In the remainder of the paper results will be presented for the positioning of phytosanitary treatments on the second development cycle using the 15% threshold.

f. Experimental setup

Agronomic simulations and evaluations with the EVA model are performed for the years 2018 and 2019 at eight sites that correspond to famous French vineyards where the grapevine moth causes significant damages (Fig. 2). It is assumed that there is no spatial dependency between these locations. Initialization date for EVA simulation is set on 20 September of the previous year. From this initialization date to the starting date of the forecast, the model is integrated using temperature measurements from the French real-time meteorological observation network. EVA forecasts are run in the period ranging from 1 April to 19 September (for years 2018 and 2019), when the impact of the grape berry moth is the most significant.

Three additional sites corresponding to wheat fields that can be affected by tritiform septoria (STB), a major disease in Europe of winter wheat (Miedaner et al. 2013; Kollers et al. 2013; Suffert et al. 2011; Ghaffary 2011), have been considered for the purpose of meteorological evaluation. These 11 locations provide a wide variety of meteorological conditions in different parts in France. The verification of calibrated and seamless 2-m temperature ensemble forecasts is performed over a 1-yr period from 1 June 2018 to 31 May 2019 and uses observations from the French real-time meteorological network. Corresponding AROME-EPS, ARPEGE-EPS, and IFS-EPS forecasts are extracted from the operational archive at the nearest model grid points to the agronomic locations.

Seamless forecasts are designed using the DTW distance computed over the last $W = 7$ h before the merging time, following the performances obtained with different W values between 1 and 48 h.

3. NWP results

In this section we first present the impact of statistical postprocessing of temperature forecasts, and we then discuss the characteristics and probabilistic performances of the different seamless temperature forecasts.

a. Verification of calibrated forecasts

Figure 3 presents the bias and RMSE for 2-m temperature ensemble mean forecasts over a one year period, before and after postprocessing. As expected, the parametric calibration procedure allows for a noticeable improvement of the three EPS performances over the entire forecast range. These calibrated ensembles are used in the remainder of the paper to compute seamless forecasts.

b. Justification of seamless design

We recall that the proposed strategies for building seamless forecasts are based on two main choices: using the AROME-EPS only over the first 48 h and selecting subsequent ARPEGE-EPS

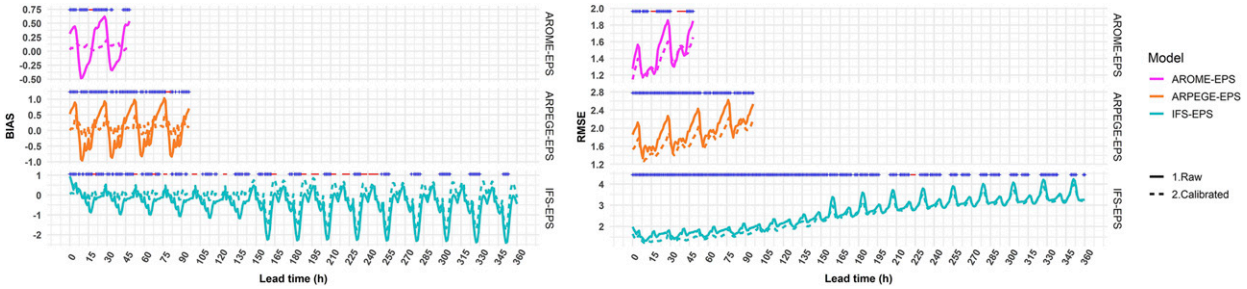

FIG. 3. (left) Bias and (right) RMSE of ensemble mean forecast for 2-m temperature, as a function of forecast lead time. Scores are computed for the 11 sites over the period 1 Jun 2018–31 May 2019 with raw (solid line) and calibrated (dashed line) ensembles. (top) AROME-EPS, (middle) ARPEGE-EPS, and (bottom) IFS-EPS. The plus (minus) signs at the top indicate that the performance of the calibrated ensemble is statistically better (worse) than the raw ensemble performance, according to the Wilcoxon test.

and IFS-EPS members based on their proximity to AROME-EPS members over the last 7 h (corresponding to forecast ranges 42–48 h). We provide here some objective results to justify this configuration.

Figure 4 presents the CRPS for the three calibrated EPSs over a 1-yr period. The AROME-EPS is shown to perform significantly better up to around 20-h lead time. From 24- to 48-h ranges, the AROME-EPS is better than ARPEGE-EPS and close to or slightly worse than the IFS-EPS. These results thus support the idea of opting for the AROME-EPS over its timeframe. Over the forecast window 48–96 h the ARPEGE-EPS performs worse than the IFS-EPS, despite its higher resolution. This thus justifies the choice of considering both ARPEGE-EPS and IFS-EPS members for the connection at 48 h, instead of ARPEGE-EPS members only.

c. *Characteristics of seamless forecasts*

1) ORIGIN OF SELECTED MEMBERS AT 48 H

As a first diagnostic, Table 2 presents the percentage of ARPEGE-EPS and IFS-EPS members selected for the connection at 48 h, for the three seamless designs and over different periods. With the random draw, 59.3% of the selected members come from the IFS-EPS, because of the larger size of this ensemble. On the other hand, the distance-based assignments provide more balanced selections, and some seasonal variations are observed. In spring and summer the IFS-EPS is more frequently chosen, while the ARPEGE-EPS is slightly preferred in winter. In autumn, both EPSs are almost equally likely. These variations reflect the mean distances between the AROME-EPS members and the ARPEGE-EPS and IFS-EPS members. It has been verified for instance that in winter the ARPEGE-EPS members are on average closer to the AROME-EPS members. Since the ARPEGE-EPS is used as lateral boundary conditions for the AROME-EPS, one could have expected a higher representation of ARPEGE-EPS members in this 48-h selection. However, it is likely that 2-m temperature mainly results from small-scale physical processes and only slightly depends on the large-scale coupling.

It can finally be noticed that the percentages obtained with the NN and HU methods are very similar. In addition, differences in the NN and HU selections, which are only due to the assignment rule (the distance criterion is the same), are rather

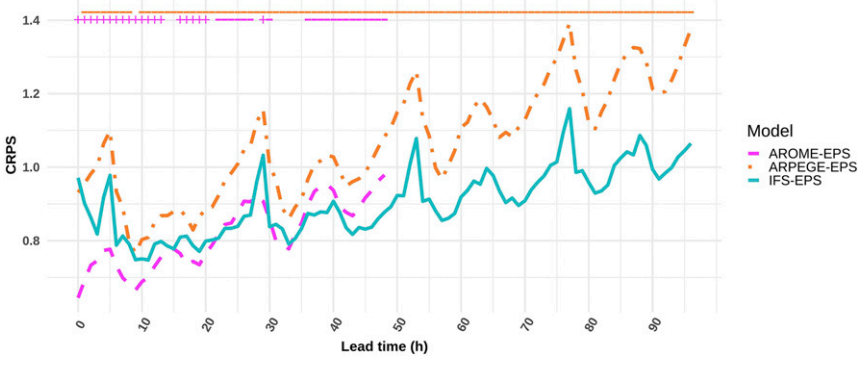

FIG. 4. CRPS of 2-m temperature forecasts as a function of forecast lead time. Scores are computed for the 11 sites over the period 1 Jun 2018–31 May 2019 for the three calibrated EPS. The plus (minus) signs at the top indicate that the performance of the AROME-EPS and the ARPEGE-EPS is statistically better (worse) than the IFS-EPS performance, according to the Wilcoxon test.

TABLE 2. Percentage of selected ARPEGE-EPS and IFS-EPS members with the different assignment strategies for the first matching step, over different time periods. Theoretical values for the random selection are 40.7% for ARPEGE-EPS and 59.3% for IFS-EPS, according to their respective ensemble size.

	ARPEGE-EPS			IFS-EPS		
Time period	RN (%)	NN (%)	HU (%)	RN (%)	NN (%)	HU (%)
1 Jun 2018–31 Aug 2018	40.7	45.65	45.41	59.3	54.35	54.59
1 Sep 2018–30 Nov 2018	40.7	49.60	49.10	59.3	50.40	50.90
1 Dec 2018–28 Feb 2019	40.7	53.00	52.50	59.3	47.00	47.50
1 Mar 2019–31 May 2019	40.7	46.27	45.87	59.3	53.73	54.13

small with an intersection between NN and HU selected members equal to 87% on average for this 48-h connection.

2) REPETITION OF MEMBERS FOR THE NN STRATEGY

As explained in section 2c, a major limitation of the nearest neighbor approach is the possible repetition of selected ARPEGE-EPS and IFS-EPS members, thus breaking the assumption of independent ensemble members. To examine this aspect, Fig. 5 presents the distribution of effective ensemble size (i.e., the number of independent ensemble members) after the first and second merging steps. It appears that most ensembles have between 9 and 12 independent members, while full 12-member ensembles after the first and second connections only represent 20% and 10% of the cases, respectively. Ensembles with less than 6 independent members are also observed in a few extreme cases.

d. Performance of seamless forecasts

The different seamless designs are evaluated using two criteria: the temporal discontinuities in temperature time series in the vicinity of the junction time, and the probabilistic performances up to 15 days ahead. Ideally, the best seamless design is the one that optimizes the two criteria, in other words, one that minimizes temporal discontinuities and maximizes the performance.

1) ANALYSIS OF AVERAGE DISCONTINUITIES

The ability of the seamless design strategies to generate smooth transitions between two EPS members is measured with the amplitude of temperature differences between lead times 48 and 49 h for the first assignment, and lead times 96 and 97 h for the second assignment. The distributions of these 1-h absolute temperature differences are shown in Fig. 6 for the three seamless designs. Differences obtained with the IFS-EPS members, which are seamless by construction, are taken as a reference. As expected, the largest discontinuities are obtained with the RN strategy, while the discontinuities obtained with the NN and HU assignments are much smaller and close to those observed in the seamless IFS-EPS forecasts. These results thus indicate that the mergings obtained with the NN and HU methods are able to provide realistic forecasts. Note also that smaller discontinuities are observed for the 96-h connection, especially for the RN strategy. This is because this second assignment is done only for the ARPEGE-EPS members, and the IFS-EPS members selected at 48 h contribute to reduce the average temperature differences.

2) CRPS

The probabilistic performances of the different seamless ensemble forecasts are assessed with respect to the IFS-EPS performance. Since the seamless ensembles have only 12 members, the IFS-EPS scores are computed using a random subset of 12 members. As expected from Fig. 4, the positive impact of seamless forecasts is the largest in the first 24 h, when forecasts from the AROME-EPS are used, with an average CRPS improvement of 9% over the first forecast day (Fig. 7). For longer lead times, the NN ensemble performs significantly worse because of its smaller effective ensemble size. The RN ensemble performs best for forecast days 2–4 because the number of IFS-EPS members, which perform better than ARPEGE-EPS members (Fig. 4), is larger. From forecast days 5 to 15, RN and HU ensembles have similar performances, with CRPS values not statistically different from the IFS-EPS CRPS. Examination of CRPS at individual locations (not

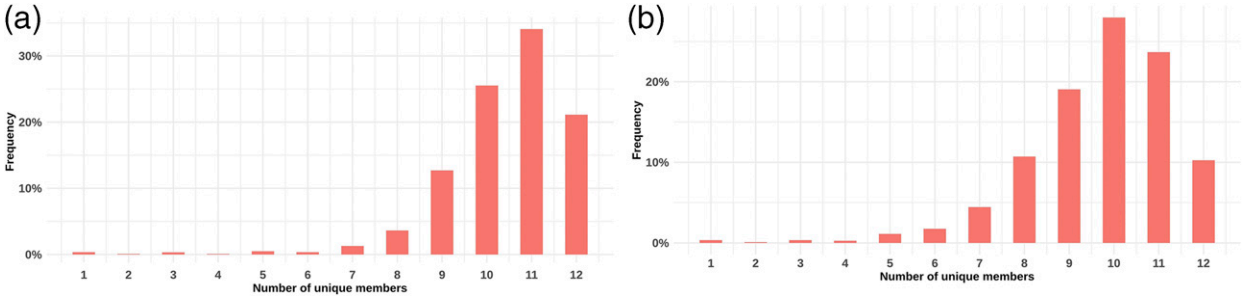

FIG. 5. Frequency histograms of the number of independent members in the NN ensembles, computed for the 11 sites over the period 1 Jun 2018–31 May 2019 (a) after the 48 h-merging step and (b) after the 96 h-merging step.

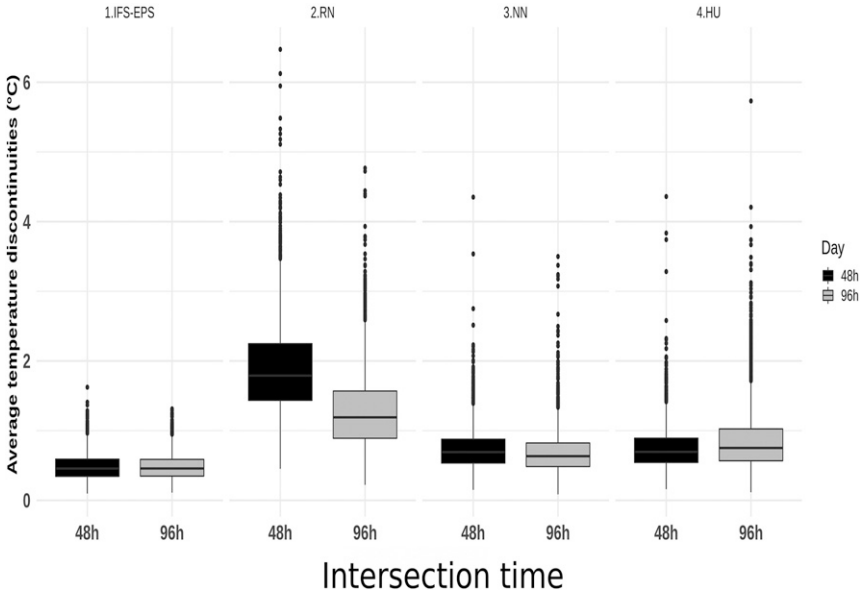

FIG. 6. Boxplots of absolute temperature forecast differences, averaged over ensemble members, between lead times 48 and 49 h (in black) and lead times 96 and 97 h (in gray), computed for the 11 sites over the period 1 Jun 2018–31 May 2019. These forecast differences are presented (from left to right) for IFS-EPS members and for the seamless RN, NN, and HU ensemble members.

shown) indicates that over the first 2 days using AROME-EPS leads to an improvement at 8 sites (over 11). At longer ranges (from days 7 to 15), the seamless HU and RN forecasts both improve on average the IFS-EPS performance at 7 sites.

3) AUC AND RELIABILITY

The short-range improvement of seamless forecasts is linked to both increased resolution (Fig. 8) and reliability (Fig. 9). Over the first 48 h, seamless forecasts improve the reliability of probabilities higher than 0.6 for the event 2-m temperature above 15°C. Over the whole 15 days forecast range, RN and HU ensembles are very close to the IFS-EPS, while the NN reliability is slightly worse, especially for small probabilities. Similar conclusions are obtained with thresholds 5° and 10°C.

Since the added value of using the AROME-EPS is mainly restricted to the first 24 h, an alternative seamless setup would be to perform the first merging at 24 h instead of 48 h. Figure 10 indicates that this earlier merging leads to noticeable improvements of the HU ensemble performance from day 2 to 4.

Overall, from an NWP point of view the HU seamless design is the only one to fulfill both smoothness and performance evaluation criteria. This seamless forecast is shown to efficiently

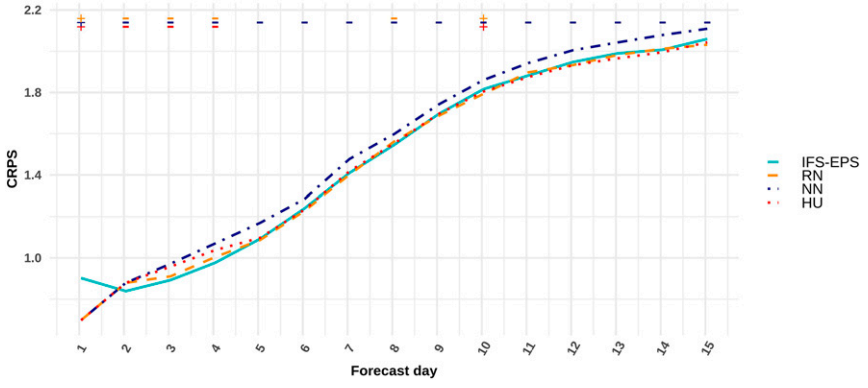

FIG. 7. CRPS of 2-m temperature forecasts computed for the 11 sites over the period 1 Jun 2018–31 May 2019, as a function of forecast day. Results for the random 12-member IFS-EPS and the three seamless ensembles are overlaid. The plus (minus) signs at the top indicate that the performance of a seamless ensemble is statistically better (worse) than the IFS-EPS performance, according to the Wilcoxon test.

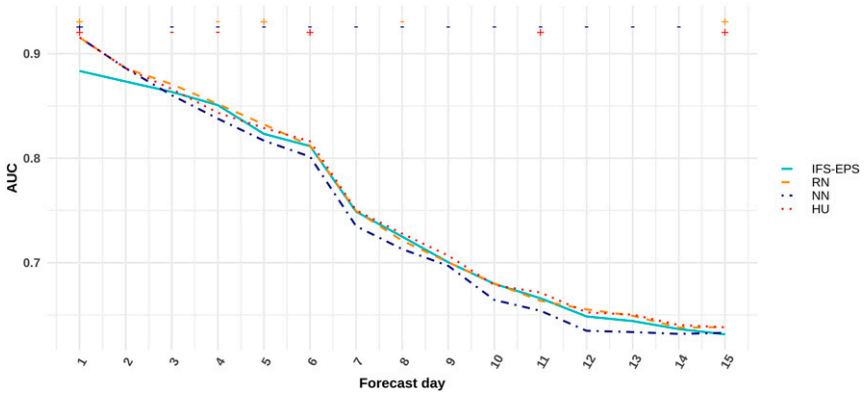

FIG. 8. AUC of 2-m temperature forecasts above 15°C, computed for the 11 sites over the period 1 Jun 2018–31 May 2019, as a function of forecast day. Results for the random 12-member IFS-EPS and the three seamless ensembles are overlaid. The plus (minus) signs at the top indicate that the performance of a seamless ensemble is statistically better (worse) than the IFS-EPS performance, according to the Wilcoxon test.

combine the higher performance of the convective-scale AROME-EPS at short ranges, while maintaining the performance of the IFS-EPS at longer ranges. In the next section, the impact of calibrated and seamless forecasts is examined on the performance of the agronomic simulations.

4. EVA results

In this section we focus on the development of the second generation of the grapevine moth. The treatment date is determined from the cumulative level of the egg stage using the threshold 15%, commonly used by experts for the application of biocontrol treatments. Five ensembles of 12 EVA simulations are compared, which are driven by 12 random temperature forecasts from the raw and calibrated IFS-EPS, and by the three 12-member seamless ensembles. To evaluate these predictions, scores commonly used for NWP EPS verification are applied to the distributions of treatment dates. These scores are presented as a function of the time interval between the observation-based treatment date and the forecast starting date. According to this approach and considering that the maximum lead time of probabilistic forecasts used for this work is 15 days, the results are provided for time intervals between 1 and 15 days. For long time intervals, the treatment threshold may not be reached by some members at the end of the forecast. For these members the model sets the treatment date to ''beyond.'' To perform the scores computation, these ''beyond'' dates are arbitrarily set up to 17. Finally, given the results of the meteorological evaluation, only seamless forecasts with a first merging at 24 h are shown in this section.

a. Bias

The bias of the ensemble mean treatment date is presented in Fig. 11. For short-term predictions corresponding to time intervals between 1 and 3 days, the bias is zero whatever the

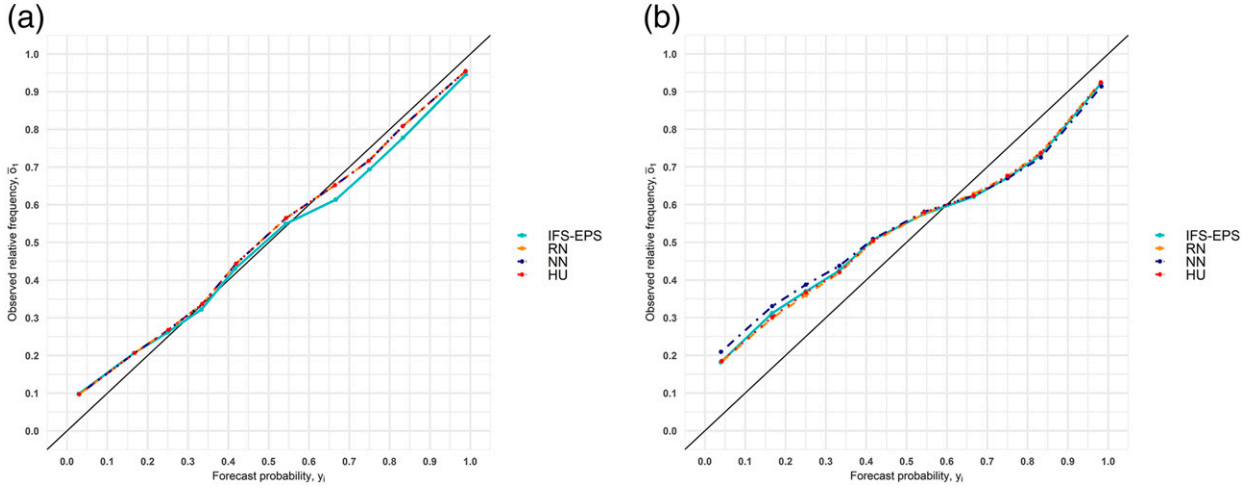

FIG. 9. Reliability diagram for 2-m temperature forecasts above 15°C, computed for the 11 sites over the period 1 Jun 2018–31 May 2019: lead times (a) 0–48 and (b) 0–360 h. Results for the random 12-member IFS-EPS and the three seamless ensembles are overlaid.

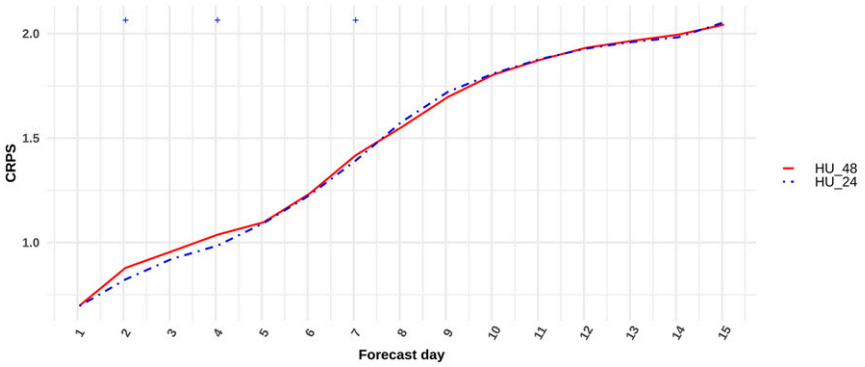

FIG. 10. CRPS of 2-m temperature forecasts computed for the 11 sites over the period 1 Jun 2018–31 May 2019, as a function of forecast day. Results are given for seamless HU ensembles with a first merging at 24 h (dashed–dotted line) and 48 h (solid line). The plus (minus) signs at the top indicate that the performance of the 24-h merging is statistically better (worse) than the performance of the 48-h merging, according to the Wilcoxon test.

temperature forecast used. For time intervals between 4 and 7 days, the bias slightly increases toward positive values (which means that the treatment date is slightly too late) and there is a small benefit from using calibrated temperature forecasts. This positive bias is in agreement with the negative temperature bias of raw IFS-EPS forecasts seen in Fig. 3. The advantage of calibration becomes larger for forecasts between one and two weeks before the treatment date, with bias values that remain close to zero.

b. CRPS

The impact of the calibration procedure is also clearly visible on the CRPS (Fig. 12), which is significantly improved for week-2 forecasts by almost 40%. This improvement is observed at all sites except Bordeaux. On the other hand, it is more difficult to conclude about the value and ranking of seamless forecasts, compared to calibrated IFS-EPS. Local CRPS at individual locations (not shown) indicate that the seamless HU week-2 forecasts lead to a slight improvement over the calibrated IFS-EPS at 6 sites (over 8). On the other hand, NN and RN showed an improvement over 5 sites.

c. Observation coverage

Finally, Fig. 13 shows the percentage of observation-based treatment dates that fall in the predicted range. Consistently with previous results, the prediction is almost perfect for short intervals and hardly sensitive to the meteorological forecast used up to a one-week interval. For longer time intervals, the observation coverage significantly decreases to less than 70%, which is lower than the theoretical value of 85% for a reliable 12-member ensemble, while calibrated and seamless forecasts help maintaining a high level of observation coverage, with values around 95% or more up to 11 days. The mean observation coverage over days 7–15 is slightly larger for the HU forecasts, with a value of 94.5% compared to values between 91% and 92% for the calibrated, RN and NN forecasts.

From an agronomic point of view, the impact of weather forecast improvements is rather limited up to one week before

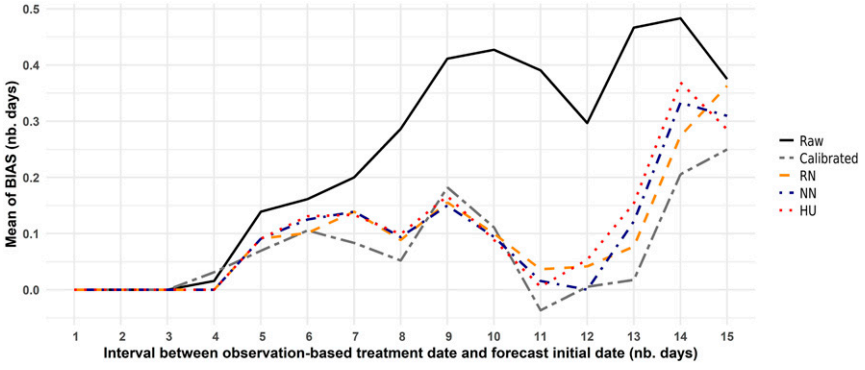

FIG. 11. Bias of the ensemble mean treatment date forecast, as a function of the time interval (in day) between the observation-based date and the forecast starting date. Scores are averaged over the eight vineyard sites and the period of interest 1 Apr–19 Sep for the years 2018 and 2019. Results from using raw IFS-EPS, calibrated IFS-EPS, and the three seamless ensemble forecasts are overlaid.

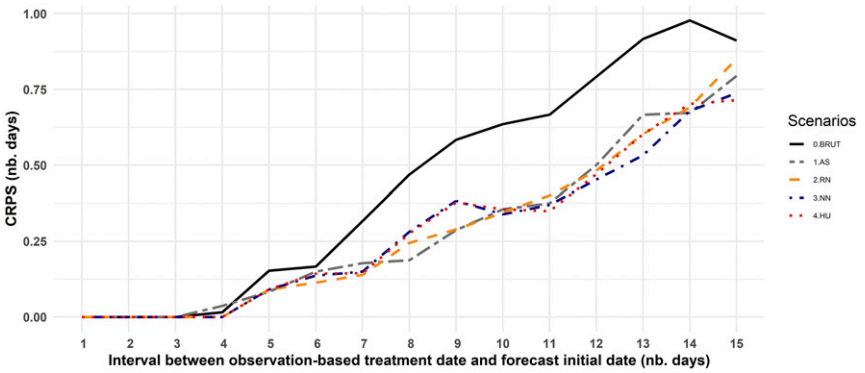

FIG. 12. CRPS for the forecast treatment date, as a function of the time interval (in day) between the observation-based date and the forecast starting date. Scores are averaged over the eight vineyard sites and the period of interest 1 Apr–19 Jun for the years 2018 and 2019. Results from using raw IFS-EPS, calibrated IFS-EPS, and the three seamless forecasts are overlaid.

the treatment date. This behavior may be specific to the EVA model. It is possible that the evolution rules based on the accumulation of daily temperatures and threshold effects contribute to reduce the impact of forecast accuracy and uncertainty at short ranges. On the other hand, the advantage of using calibrated forecasts is clearly visible between 7 and 15 days and this may help winegrowers to better anticipate the scheduling of treatments. It is, however, not possible to properly conclude on the added value of seamless forecasts with the current experimental setup. Additional simulations covering a larger number of locations and years may be required to detect a more robust signal.

5. Discussion and future works

The goal of this paper was to evaluate the potential of statistical postprocessing and seamless integration methods to generate skillful 2-m temperature ensemble forecasts for application to crop protection. Postprocessing is based on the well-known EMOS and ECC methods, while new strategies have been proposed to design seamless ensemble forecasts from the concatenation of several ensemble predictions with different spatiotemporal scales. The seamless methods aim at generating ensemble members from high-resolution AROME-EPS forecasts at short ranges, with a smooth transition toward larger-scale EPSs at longer ranges. The proposed approach performs minimum-cost associations between members from the different EPSs, according to a given distance and matching rules.

The evaluation is first performed from a meteorological point of view. It is shown that a naive random association of EPS members achieves satisfactory probabilistic performances, but with the inconvenience of generating large discontinuities in the vicinity of the matching time. On the other hand, the distance-based associations lead to smaller discontinuities and achieve good performances with the Hungarian matching algorithm.

Ensembles of agronomic forecasts are then issued from the calibrated and seamless weather forecasts. Statistical

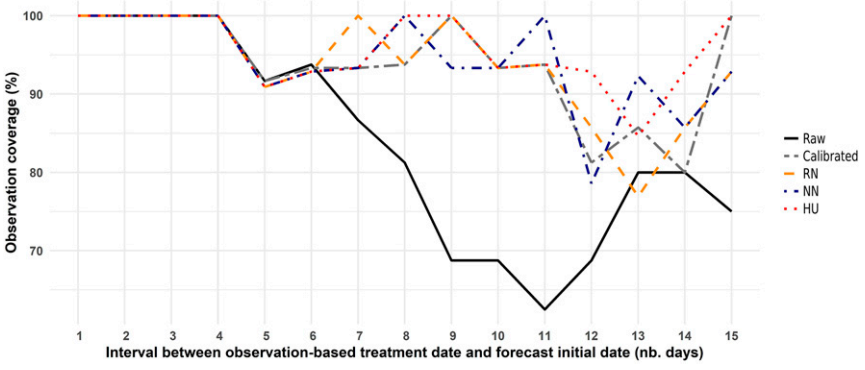

FIG. 13. Observation coverage, as a function of the time interval (in days) between the observation-based treatment date and the forecast starting date. The percentage values of observation coverage are averaged over the 8 vineyard sites and the period of interest 1 Apr–19 Sep for the years 2018 and 2019. Results from using raw IFS-EPS, calibrated IFS-EPS, and the three seamless forecasts are overlaid.

postprocessing strongly benefits to the quality of the treatment date prediction, especially between one and two weeks before the treatment date, which indeed allow a better anticipation of the spraying schedule for the winegrowers. However, the impact of seamless forecasts is more difficult to assess and may require a more thorough evaluation considering additional locations and years. Agronomic models that depend on less-predictable variables and that are more sensitive to short-range weather uncertainty will also be considered in future studies to better highlight the potential of the seamless ensembles.

The proposed method is generic and could be applied to other parameters. Since precipitation play a major role in the development of diseases and pests, calibration and seamless combination of precipitation forecasts is a natural step forward. The calibration of precipitation has already been considered in the work of Scheuerer (2014); Taillardat et al. (2019) for instance. The design of seamless predictions may require specific considerations especially regarding the choice of the distance metric to measure the similarity between precipitation forecasts which are subject to spatiotemporal intermittency.

Seamless forecasts are important to several applications, including in particular hydrology and energy production, and concern different time scales, e.g., from nowcasting to NWP or from NWP to seasonal forecasts. It would be straightforward to design seamless forecasts over wider time ranges with the integration of more EPSs and matching steps. For instance, additional connections to S2S forecasts (White et al. 2017) could be interesting for a longer-term management of agricultural practices.

The smoothness criterion is central in the proposed seamless design strategy, although it may not be crucial for this crop protection application that is mainly sensitive to the cumulative effect of some weather variables. However, it is relatively intuitive and of particular interest for other applications, such as the electricity production from renewable energies, which is very sensitive to rapid changes in the meteorological conditions.

Possible avenues for improving the seamless integration method include a relaxation of the ''AROME-EPS only'' assumption over the first 2 days with, for instance, an initialization of the seamless ensemble with a clustering of all available EPS members. It is also possible to develop a backward matching procedure, starting from all IFS-EPS members at long ranges with backward connections to shorter-range EPSs. This approach would preserve the full ensemble variability at long ranges (that may be the most relevant for crop protection). Another elegant way would be to build seamless blended ensemble forecasts from the aggregation of probabilistic forecasts and Schaake shuffle reconstruction, as proposed by Zamo (2016).

The extension of the seamless design to a multivariate context is an important future step since most applications require several meteorological input variables. There are at least two ways of doing this, either using the current univariate matching scheme based on the most important weather variable for the application or using a multivariate metric such as the Mahalanobis distance. This aspect will be the subject of future works.

Finally, an important question is whether it is more efficient to calibrate the input temperature forecasts or rather to directly calibrate the end-product. In the present case, the calibration of temperature is the preferred option because it first improves the construction of seamless forecasts by reducing the concatenation gaps. Second the calibration of treatment dates is, to the best of our knowledge, an unexplored area that may face several challenges. For instance, there is no guarantee that treatment dates follow a simple parametric distribution and the historical sample of observation-based treatment dates available for calibration is rather small because there are only three or four generations of grapevine moth per year.

Acknowledgments. This work is a part of a Ph.D. thesis that is supported by the French National Research Agency under the Investments for the Future Program, referred as ANR-16-CONV-0004, for DigitAg ''Digital Agriculture Convergence Lab'' (www.hdigitag.fr/en). It is also supported by the CASDAR-funded project METEOPREC ''the contribution of accurate forecasting to farmers'' (CASDAR are funded by the French Ministry of Agriculture) and the network Modeling and Data Analysis for Agriculture (www.modelia.org). This article has greatly benefited from the support of my colleague Michael Zamo for the EMOS post-processing. I also thank Frank Souverain, Sébastien Prat, and Pascal Raynaud for their help regarding the extraction of weather forecasts.

Data availability statement. The forecast data used for this work are taken from the Météo-France and ECMWF databases. The agronomic EVA model is the property of the French Institute for Vine and Wine.

REFERENCES

Arbogast, P., A. Hally, J. Cheung, J. Heijstek, A. Marsman, and J.-L. Brenguier, 2015: Recommendations on trajectory selection in flight planning based on weather uncertainty. *Proc. Fifth SESAR Innovation Days (SID2015)*, Bologna, Italy, Meteo-France, 17 pp., https://library.wmo.int/doc_num.php?explnum_id=4390.

Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525**, 47–55, https://doi.org/10.1038/nature14956.

Bellier, J., G. Bontron, and I. Zin, 2017: Using meteorological analogues for reordering postprocessed precipitation ensembles in hydrological forecasting. *Water Resour. Res.*, **53**, 10 085–10 107, https://doi.org/10.1002/2017WR021245.

——, I. Zin, and G. Bontron, 2018: Generating coherent ensemble forecasts after hydrological postprocessing: Adaptations of ECC-based methods. *Water Resour. Res.*, **54**, 5741–5762, https://doi.org/10.1029/2018WR022601.

Berndt, D. J., and J. Clifford, 1994: Using dynamic time warping to find patterns in time series. *KDD Workshop*, Seattle, WA, Vol. 10, 359–370.

Bouttier, F., and L. Raynaud, 2018a: Clustering and selection of boundary conditions for limited-area ensemble prediction. *Quart. J. Roy. Meteor. Soc.*, **144**, 2381–2391, https://doi.org/10.1002/qj.3304.

——, and ——, 2018b: Clustering and selection of boundary conditions for limited area ensemble prediction. *Quart. J. Roy. Meteor. Soc.*, **144**, 2381–2391, https://doi.org/10.1002/qj.3304.

Calanca, P., D. Bolius, A. Weigel, and M. Liniger, 2011: Application of long-range weather forecasts to agricultural decision problems in Europe. *J. Agric. Sci.*, **149**, 15–22, https://doi.org/10.1017/S0021859610000729.

Candille, G., and O. Talagrand, 2005: Evaluation of probabilistic prediction systems for a scalar variable. *Quart. J. Roy. Meteor. Soc.*, **131**, 2131–2150, https://doi.org/10.1256/qj.04.71.

——, C. Côté, P. Houtekamer, and G. Pellerin, 2007: Verification of an ensemble prediction system against observations. *Mon. Wea. Rev.*, **135**, 2688–2699, https://doi.org/10.1175/MWR3414.1.

Chavent, F., 1983: Etude de la dynamique des populations d'Eudémis de la vigne (Lobesia botrana Den. et Schiff.). Adaptation d'un modèle d'évolution aux conditions climatiques de Provence, mémoire de fin d'études.

Christ, E. H., P. J. Webster, G. D. Collins, V. E. Toma, and S. A. Byrd, 2015: Using precipitation forecasts to irrigate cotton. *J. Cotton Sci.*, **19**, 351–358, https://www.cotton.org/journal/2015-19/3/upload/JCS19-351.pdf.

Clark, M., S. Gangopadhyay, L. Hay, B. Rajagopalan, and R. Wilby, 2004: The Schaake shuffle: A method for reconstructing space–time variability in forecasted precipitation and temperature fields. *J. Hydrometeor.*, **5**, 243–262, https://doi.org/10.1175/1525-7541(2004)005<0243:TSSAMF>2.0.CO;2.

Cooke, B. M., and Coauthors, 2006: *The Epidemiology of Plant Diseases*. Vol. 2. Springer, 568 pp.

Descamps, L., C. Labadie, A. Joly, E. Bazile, P. Arbogast, and P. Cébron, 2015: PEARP, the Météo-France short-range ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **141**, 1671–1685, https://doi.org/10.1002/qj.2469.

Fundel, V. J., N. Fleischhut, S. M. Herzog, M. Göber, and R. Hagedorn, 2019: Promoting the use of probabilistic weather forecasts through a dialogue between scientists, developers and end-users. *Quart. J. Roy. Meteor. Soc.*, **145**, 210–231, https://doi.org/10.1002/qj.3482.

Ghaffary, M. T., 2011: Efficacy and mapping of resistance to Mycosphaerella graminicola in wheat. Ph.D. thesis, Wageningen University, https://library.wur.nl/WebQuery/wda/abstract/1964546.

Giorgino, T., and Coauthors, 2009: Computing and visualizing dynamic time warping alignments in R: The DTW package. *J. Stat. Software*, **31**, 1–24, https://doi.org/10.18637/jss.v031.i07.

Gneiting, T., 2014: Calibration of medium-range weather forecasts. ECMWF Tech. Memo. 719, ECMWF, 30 pp., https://doi.org/10.21957/8xna7glta.

——, and A. E. Raftery, 2005: Weather forecasting with ensemble methods. *Science*, **310**, 248–249, https://doi.org/10.1126/science.1115255.

——, ——, A. H. Westveld III, and T. Goldman, 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Wea. Rev.*, **133**, 1098–1118, https://doi.org/10.1175/MWR2904.1.

Hagelin, S., J. Son, R. Swinbank, A. McCabe, N. Roberts, and W. Tennant, 2017: The Met Office convective-scale ensemble MOGREPS-UK. *Quart. J. Roy. Meteor. Soc.*, **143**, 2846–2861, https://doi.org/10.1002/qj.3135.

Hanley, J. A., and B. J. McNeil, 1982: The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**, 29–36, https://doi.org/10.1148/radiology.143.1.7063747.

Hemri, S., M. Scheuerer, F. Pappenberger, K. Bogner, and T. Haiden, 2014: Trends in the predictive performance of raw ensemble weather forecasts. *Geophys. Res. Lett.*, **41**, 9197–9205, https://doi.org/10.1002/2014GL062472.

Kollers, S., and Coauthors, 2013: Genetic architecture of resistance to *Septoria tritici* blotch (*Mycosphaerella graminicola*) in European winter wheat. *Mol. Breed.*, **32**, 411–423, https://doi.org/10.1007/s11032-013-9880-6.

Kuhn, H. W., 1955: The Hungarian method for the assignment problem. *Nav. Res. Logist. Q.*, **2**, 83–97, https://doi.org/10.1002/nav.3800020109.

Leutbecher, M., and T. N. Palmer, 2008: Ensemble forecasting. *J. Comput. Phys.*, **227**, 3515–3539, https://doi.org/10.1016/j.jcp.2007.02.014.

Miedaner, T., Y. Zhao, M. Gowda, C. F. H. Longin, V. Korzun, E. Ebmeyer, E. Kazman, and J. C. Reif, 2013: Genetic architecture of resistance to *Septoria tritici* blotch in European wheat. *BMC Genomics*, **14**, 858, https://doi.org/10.1186/1471-2164-14-858.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119, https://doi.org/10.1002/qj.49712252905.

Moyer, M. M., D. M. Gadoury, W. F. Wilcox, and R. C. Seem, 2016: Weather during critical epidemiological periods and subsequent severity of powdery mildew on grape berries. *Plant Dis.*, **100**, 116–124, https://doi.org/10.1094/PDIS-12-14-1278-RE.

Olatinwo, R., T. Prabha, J. Paz, D. Riley, and G. Hoogenboom, 2011: The Weather Research and Forecasting (WRF) Model: Application in prediction of TSWV-vectors populations. *J. Appl. Entomol.*, **135**, 81–90, https://doi.org/10.1111/j.1439-0418.2010.01539.x.

——, ——, ——, and G. Hoogenboom, 2012: Predicting favorable conditions for early leaf spot of peanut using output from the Weather Research and Forecasting (WRF) Model. *Int. J. Biometeor.*, **56**, 259–268, https://doi.org/10.1007/s00484-011-0425-6.

Palmer, T., 2019: The ECMWF ensemble prediction system: Looking back (more than) 25 years and projecting forward 25 years. *Quart. J. Roy. Meteor. Soc.*, **145**, 12–24, https://doi.org/10.1002/qj.3383.

Pappenberger, F., F. Wetterhall, E. Dutra, F. Di Giuseppe, K. Bogner, L. Alfieri, and H. L. Cloke, 2013: Seamless forecasting of extreme events on a global scale. *Climate and Land Surface Changes in Hydrology*, E. Boegh et al., Eds., IAHS Publication, 3–10.

Pertot, I., and Coauthors, 2017: A critical review of plant protection tools for reducing pesticide use on grapevine and new perspectives for the implementation of IPM in viticulture. *Crop Prot.*, **97**, 70–84, https://doi.org/10.1016/j.cropro.2016.11.025.

Pinson, P., H. A. Nielsen, H. Madsen, and G. Kariniotakis, 2009: Skill forecasting from ensemble predictions of wind power. *Appl. Energy*, **86**, 1326–1334, https://doi.org/10.1016/j.apenergy.2008.10.009.

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, **133**, 1155–1174, https://doi.org/10.1175/MWR2906.1.

Raynaud, L., and F. Bouttier, 2016: Comparison of initial perturbation methods for ensemble prediction at convective scale. *Quart. J. Roy. Meteor. Soc.*, **142**, 854–866, https://doi.org/10.1002/qj.2686.

Sakoe, H., S. Chiba, A. Waibel, and K. Lee, 1990: Dynamic programming algorithm optimization for spoken word recognition. *Readings in Speech Recognition*, A. Waibel and K.-F. Lee, Eds., Elsevier Science, 159–165.

Schefzik, R., 2011: Ensemble copula coupling. M.S. thesis, Faculty of Mathematics and Informatics, University of Heidelberg, Germany.

——, and Coauthors, 2013: Uncertainty quantification in complex simulation models using ensemble copula coupling. *Stat. Sci.*, **28**, 616–640, https://doi.org/10.1214/13-STS443.

Scheuerer, M., 2014: Probabilistic quantitative precipitation forecasting using ensemble model output statistics. *Quart. J. Roy. Meteor. Soc.*, **140**, 1086–1096, https://doi.org/10.1002/qj.2183.

Shipp, J. L., and N. D. Clarke, 1999: Decision tools for integrated pest management. *Integrated Pest and Disease Management in Greenhouse Crops*, R. Albajes et al., Eds., Springer, 168–182.

Suffert, F., I. Sache, and C. Lannou, 2011: Early stages of *Septoria tritici* blotch epidemics of winter wheat: Build-up, overseasoning, and release of primary inoculum. *Plant Pathol.*, **60**, 166–177, https://doi.org/10.1111/j.1365-3059.2010.02369.x.

Taillardat, M., A.-L. Fougères, P. Naveau, and O. Mestre, 2019: Forest-based and semiparametric methods for the postprocessing of rainfall ensemble forecasting. *Wea. Forecasting*, **34**, 617–634, https://doi.org/10.1175/WAF-D-18-0149.1.

Thiéry, D., and Coauthors, 2013: Histoire de l'installation de quelques ravageurs. *Interactions Insectes-Plantes*, N. Sauvion et al., Eds., Quae, 623–662.

——, P. Louâpre, L. Muneret, A. Rusch, G. Sentenac, F. Vogelweith, C. Iltis, and J. Moreau, 2018: Biological protection against grape berry moths: A review. *Agron. Sustain. Dev.*, **38**, 15, https://doi.org/10.1007/s13593-018-0493-7.

Wetterhall, F., and F. Di Giuseppe, 2018: The benefit of seamless forecasts for hydrological predictions over Europe. *Hydrol. Earth Syst. Sci.*, **22**, 3409–3420, https://doi.org/10.5194/hess-22-3409-2018.

White, C. J., and Coauthors, 2017: Potential applications of Subseasonal-to-Seasonal (S2S) predictions. *Meteor. Appl.*, **24**, 315–325, https://doi.org/10.1002/met.1654.

Wilcoxon, F., S. Katti, and R. A. Wilcox, 1970: Critical values and probability levels for the Wilcoxon rank sum test and the Wilcoxon signed rank test. *Selected Tables in Mathematical Statistics*, Vol. 1, American Mathematical Society, 171–259.

Worsnop, R. P., M. Scheuerer, and T. M. Hamill, 2019: Extended-range probabilistic fire-weather forecasting based on ensemble model output statistics and ensemble copula coupling. *Mon. Wea. Rev.*, **148**, 499–521, https://doi.org/10.1175/MWR-D-19-0217.1.

Zamo, M., 2016: Statistical post-processing of deterministic and ensemble wind speed forecasts on a grid. Ph.D. thesis, University of Paris-Saclay, Gif-sur-Yvette, France, 166 pp.