# Interactive comment on "Real time drought forecasting system for irrigation management" by A. Ceppi et al.

**Anonymous Referee #2**

Received and published: 6 February 2014

General comments
This manuscript proposes the feasibility of implementation of a real-time forecasting system to be used in irrigation management (PRE.G.I.). The forecast system relies on a meteorological (WRF) and a hydrological component (FEST-WB). The authors state that the models were calibrated and validated using in-situ measurement as latent heat flux and soil moisture. I found the paper potentially interesting but there is a need for improve.

First of all we thank the anonymous referee 2 for his appreciation of our research study.

One of my main concerns is how the model was validated.

When we speak about the validation of the model we refer to our hydrological model and not the WRF meteorological model. We carried out two years of calibration (2010 and 2011), comparing observed and simulated soil moisture data and one year validation in 2012. The WRF forecast output was supplied by Epson Meteo Centre and its features are described in Section 3.1.

In figure 7 the authors states that the rainfall forecasts shows better skills for more extreme precipitation thresholds (100 mm), however this is not completely true if not misleading.

The precipitation threshold of 20, 50 and 100 mm is referred to a cumulated period of 1-30 days not to one single day, hence a cumulated precipitation over an extended period cannot be considered as an extreme event.

As the Brier score is defined, the rarer an event it is easier to get a better BS without having any real improvement in the forecast skill.

This is true and, in fact, we highlight it in the text, when we mention: "*This reason can be explained with the occurred frequency of events which exceeds the 100 mm cumulated threshold in 1-30 days, rarely occurred during March-August 2012 and in general in the summer season in the Po Valley area, in comparison with the cumulated threshold (observed/forecasted) of 20 mm which is more frequently exceed*"; therefore a cumulated threshold of 20 mm in 1-30 days is more frequent and a worse BS is shown.
However, the aim of the Figure 7 is to show the reliability of the WRF meteorological model for a forecast horizon between 1-30 days during the growing season of 2012, where we found a high reliability (BS values lower than 0.15) within the first 10 days, even for a threshold of 20 mm cumulated in 10 days which is more frequent and therefore more difficult to predict.
We chose to show the performance over a cumulated variable period (1-30 days) after an investigation done with the landowner of Livraga field which is the real decision-maker: from his point of view he is more interested to know the reliability of a cumulated precipitation forecast over 7 days or 10 days and not if it is going to rain exactly on the 7th day or 10th day from the forecast initialization date.

In this respect I would recommend the authors to benchmark the model with different metrics that take into account a reference forecast as the climatology or the persistence. Just to name one, this is the case for the Brier Skill Score (BSS, see Mason 2004). In this way, some of the authors' statements need additional justification.

We cannot use the climatological data since weather data over our experimental test site (Livraga) were available only for 2 years (2010 and 2011), hence this period was too short to be assumed as reference. Further, statistical data coming from rain gauges located in the Po Valley are available only as monthly average and they are not so close to our experimental field.

However I think the manuscript can be improved with additional analysis of results to support that the proposed system will be useful for such a forecasting system. Therefore I think that the article is of relevance for readers of HESS, I do recommend some revisions –listed below- which could make the paper more complete and better understandable.

Particular comments

Page 15812, line 14: The reference of a forecast of up to 30 days as a long-range forecast is not totally accurate. Usually long range forecasts are associated a periods from 30 days to 2 years. As the authors are discussing about forecasts mainly between periods beyond 7 days and up to 30 days I would recommend to refer to that as Medium-range or Monthly forecasting.

The term medium-range forecast has been accepted.

Page 15815, lines 16-18: Here is not clear the source of the temperature and precipitation data. Are an output from the WRF or is observed data? Why at every 2 days? And not 1 or 10 days?

The meteorological forecast output (temperature and precipitation) are provided every 2 days by the Epson Meteo Centre: it is the time needed to run the combined GEPS-REPS system. However, the REPS-WRF used in this study, as mentioned in Section 3.1, has a forecast horizon of 30 days, the spatial resolution is 18 km while the temporal resolution is 12 hours. The meteorological model output becomes an input for the FEST-WB hydrological model to run soil moisture forecasts.

Then in the next sentence the authors state that the hydro model is initialized with observed data. Are referring to the same data from the previous statement? I do think that the entire paragraph need to be rephrased, please try to be more specific here.

To create initial conditions of soil moisture, as we mention in the text, the FEST-WB is initialized with observed data, provided by the ARPA Lombardy and the Meteonetwork-Epson Meteo Centre meteorological station network, of the previous day to the forecast run.

We accept your suggestion to rephrase better this paragraph.
However, to better understand this forecasting chain we report an example of each step of the operative chain:

- At 00:00 UTC on e.g. 20 June 2012 the WRF model is launched by Epson Meteo Centre
- At 12:00 UTC on 22 June 2012 the WRF model output are uploaded on our POLIMI server
- At 13:00 UTC on 22 June 2012: observed weather data of the previous day provided by the ARPA Lombardy and Meteonetwork-Epson Meteo Centre meteorological station network are available on the POLIMI server
- At 13:30 UTC on 22 June 2012 the FEST-WB model is launched with observed weather data of the previous day to produce initial conditions
- At 14:00 UTC on 22 June 2012, once initial conditions are obtained, the FEST-WB model is initialized with the WRF probabilistic forecasts to produce soil moistures forecasts

Page 15815, line 19: The acronym ARPA should be defined here instead of in page 15820.

This suggestion has been accepted.

Page 15815, lines 22-24: "In addition to observed and forecasted data, the knowledge of scheduled irrigation dates are fundamental to calculate the irrigation water input over the experimental field of Livraga." This is a general statement or the authors want to refer to the information used in the analysis? This sentence seems to be disconnected, please rephrase.

In the Muzza Consortium Basin where the Livraga maize filed is located, this is a mandatory information to produce real-time soil moisture forecasts. Since irrigation amounts are fixed and planned by the Muzza Consortium Basin (MBL), the landowner cannot irrigate his fields in different days than the ones scheduled by the MBL. Therefore, to calibrate and validate hydrological models, this information is necessary if we want to model the soil moisture considering planned irrigations too.

Page 15819 eq (2): For me it is not clear how eq(1) becomes eq(2) and how the stress threshold is defined. A clearer link between the two equations is necessary. Please explain in more detail the meaning of RAW and TAW and their link with the stress and water surplus threshold.
Page 15819 lines 12-13: the values of 0.23 and 0.33 are intended to be incorporated in eq(2)? I can't follow the construction of this thresholds.

According to FAO (Allen et al., 1998) the TAW is defined as: (field capacity – wilting point), while the RAW is defined as (field capacity – stress threshold).
From measurements done in-situ the soil texture of Livraga maize field is a silt loam, hence from the handbook of hydrology (Maidment, 1993), field capacity and wilting point are equal to 0.33 and 0.133 respectively.

Following the equation in Baroni et al. 2010, where RAW = p · TAW (eq.1), the equation 2 for a maize field, where p is assumed as 0.5, becomes:

Field capacity – stress threshold = p · (filed capacity – wilting point), hence the stress threshold we are looking for, is:

stress threshold = field capacity - p · (filed capacity – wilting point)

stress threshold = 0.33 – 0.5 · (0.33 - 0.133), then stress threshold is equal to 0.2315

Page 1589 line 20: "some observed data are missing: : :" Please be more specific.

As it is shown in Figures 2, 3, and 4 observed soil moisture and real evapotranspiration data plotted in red lines are missing. As mentioned in the text, storage battery problems during the three years project caused some lacking of data.

Page 15820 line 4: It's hard to see the contribution of the precipitation and irrigation separately. I suggest to use a stacked bar with two colors (one for each contribution) in the figures 2, 3 and 4 and enlarge the axis fonts -specially the horizontal axis- as it is difficult to read them in the printed version.

We agree with your suggestion and we will plot in different colors irrigation and precipitation amounts. Then, we are going to enlarge the axis fonts and, above all, we will use the same axis dimension for all the three figures as suggested in short comments by T. Caloiero.

Page 15820 paragraph between lines 9-13: Is this paragraph referring to Figure 3? If yes, I would recommend to swap this paragraph with the next in order to present the Figure first.

Yes, this paragraph refers to Figure 3, hence we accept to revise lines 9-17 and to present the Figure 3 first.

Section 4.2. I feel that this section could be reorganized and addressed in a better way. For instance, some results of the performance metrics are presented first than the metric is defined. This is the case of MAE and MRE that are already depicted in the previous section 4.1. Also MAE values are presented but this metric is not defined at all in the text. The Nash-Sutcliffe index is used (lines 16-20 page 15822) before the equation is defined (eq-4). Also the acronym related with this index should be homogenized (NS or ENS, or are different things?). I suggest to present first the performance metrics used. Sections 2 and 3 can be merged as a section named as data and methodology where all the metrics can be defined. Or preferably this metrics can be defined separately in an appendix.

As we answered to reviewer 1, we accept to extend the Section 3 and to rename it as "Models and methods", adding a paragraph for statistical indexes used in this study, where we first define the performance metric shown in the text.
We will also homogenize the acronym for Nash-Sutcliffe index as NS.

Page 15821 line 27: To be consistent with the reference list, Jolliffe, 2003 should be Jolliffe and Stephenson, 2003. Please check the citation or reference.

This was our mistake; the right citation is the one you propose.

Page 15824 line 2: It's not clear to me why the thresholds of 50 and 100 mm are equivalent? Please explain or rephrase.

To analyze the meteorological model performance we could have chosen different thresholds, but the reason to select these values of 50 and 100 mm is because they correspond to about a half and full irrigation amount over Livraga field, while the threshold value of 20 mm corresponds to common precipitation amounts in this geographical area.

Figure 7. In my opinion this is one of the weakest points of the paper. The authors states that the greater skill is observed for the forecasts of the extreme events. These results obtained in such a short period are only an artifact of the methodology used to assess the skill.
First of all, the three thresholds are considered not as daily precipitation values, but cumulated precipitation values in 1-30 days, hence 100 mm over a period of more days is not an extreme event.
Further, as we mentioned above, the aim of this Figure 7 is to value the reliability of the WRF forecast at the increasing of lead time during the analyzed growing season 2012 for different threshold of cumulated precipitation forecasts.

I would recommend the authors to assess the skill of these forecasts by using other metrics that take into account reference forecasts like the climatology as a benchmark.
As answered above, we cannot use the climatological data since weather data over our experimental test site (Livraga) were available only for 2 years (2010 and 2011), hence this period was too short to be assumed as reference forecast. Further, statistical data coming from rain gauges located in the surrounding area are available only as monthly average, and we cannot calculate a cumulated value over a variable time.

Page 15824 lines 19-21: This sentence is a little bit cryptic. Please consider rephrasing it.

The google map platform for the PREGI Project shows a colored dot over the Livraga test-bed. This dot can be colored in red and orange if the stress and surplus thresholds are respectively exceed, while if no of these two thresholds are exceed, no alert is forecasted and a green dot appears on the map.

Page 15824 lines 22-26. After reading this sentence, I'm not 100% sure if I understand how the probabilities were computed. The number of ensemble members exceeding the threshold is a daily value or it is accumulated over 30 days? Please explain.

The probability value (i.e. the number of ensembles out of 20) displayed in the colored dot shows the greater daily probability value over a period of 30 days. Red and orange values, representing stress and surplus thresholds, are shown only if 33% of probability (i.e. at least 7 ensembles out of 20) is detected, following the method already used in the MAP D-PHASE Project reported in Zappa et al., 2008.

Figure 8: What are the meaning of the yellow circle and the 60% value? Please add a clarification in the caption.
The Figure 8 is just an example of the web platform realized for the PREGI Project: a tool for the landowner of Livraga which was able to control in real time soil moisture forecasts over his maize field. As suggested by the reviewer 1, we are going to move it as supplementary data. However, the meaning of the Figure 8 is that there is a 60% of probability (i.e. 12 ensembles out of 20) to exceed the surplus threshold in next 30 days with the forecast simulation started on 31 August 2012.

Page 15825 lines 11-14: The authors state that "The two irrigations scheduled on 29 June and 14 July 2012 raised significantly soil moisture values in the following days above the water surplus threshold as shown in Fig. 10." However Figure 10 is showing the observed accumulated rainfall and forecasts. Are the authors referring here to Figure 9?
Yes, we refer to Figure 9. We are going to correct this statement.

Figure 9: Some of the text of this figure is in Italian and some in English. Please homogenize the language.
Ok, we are going to homogenize it in English.

Figure 10: As far I understood the extra irrigation water is not affecting the rainfall forecasts but is a deterministic value that is systematically added to both observed and forecasted information. I think that adding here the water added for irrigation is not necessary and can hide the real magnitude of the differences between the forecast and observations. Also it can be helpful to see in the plots the 25 and 75th percentile as in figure 9.
First, consider that an irrigation input is 108 mm, hence two irrigations are 216 mm. In Figure 10 cumulated amounts are shown and the median of the ensemble forecast is 256 mm while the observed value is 238 mm; hence subtracting 216 mm from total amounts, we found a difference of only 18 mm in 30 days between observed and forecasted cumulated vales. However, this plot represents only a quality estimation for the simulation initialized on 22 June 2012; as mentioned above, this section will be moved as supplementary material available during experimental test of the PREGI Project.

Page 15825 lines 15-18: The authors state that "The comparison between the REPS-WRF model forecast and the observed value at Livraga rain gauge (leaving out the two scheduled contributions coming from irrigation which are known a priori) shows a good agreement during the central phase of the maize growing season." How the authors drawn this conclusion? Is hard to see it from figure 10 as the magnitude of the irrigation is too high. Please consider to redraw Fig 10 with only the accumulated rainfall. Moreover, how the authors determine the good agreement? Is this measured somehow or is only a graphical estimation? Please provide further elements that sustain this conclusion, as this is one of the key questions.

The same as described above, we are going to move the last three figures as supplementary data.

Page 15826 lines 10-16: This paragraph is a quite general statement that is not supported in the paper. Moreover, I can't agree that the system presented in this paper "has a higher reliability in comparison with flood forecasting systems", at least I can't found any evidence of that in the paper. Please consider deleting or rephrasing this paragraph as in the present form is not completely accurate.

We think to review this paragraph supporting this statement in the text. In fact, from a meteorological point of view it is easier to predict the persistence of high pressure systems bringing drought episodes than severe perturbations producing floods. We show how in the growing season 2012 soil moisture forecasts are reliable till about 20 days, although a minor frequency of significant events has brought an improvement of the forecast performance score. However, since a flood prediction has a good reliability of 48-72 hours days as lead time (Thilen et.al., 2009), a drought forecasting system has very likely a higher reliability in comparison with the flood one.

Thielen, J., Bartholmes, J., Ramos, M.-H., and de Roo, A., The European Flood Alert System – Part 1: Concept and development, Hydrology and Earth System Sciences, 13, 125-140, 2009.

The reference to the Wilks, 2006 citation (Page 15821 line 26-27 and Page 15823 line15) is missing.

We add this citation in the reference list.

References: Mason, 2004: On using "climatology" as a reference strategy in the Brier and ranked probability skill scores. Mon. Wea. Rev. 132(7), 1891-1895.

As described above we cannot use the climatological value in this study.