* P and L denote Page and Line, respectively.

Reviewer #3

General comment:

why only 2 months of daily temperature data? why not precip? Please justify. Also, from tables 5 & 6 and from figures 6 & 7, the best method by far was the Quantile Search construct.

1- Regarding the justification, the data is justified in the Case study section as:
    - P11, L25-29: One of the objectives of this project is to produce daily air temperature map from point measurements to be used in crop growth simulations for assessing near-real time crop and irrigation water requirement. Considering the importance of June in the crop calendar of the study area which is the end of winter crops and beginning of summer crops especially maize, we applied the proposed methods to available dataset of this month.
    - P12, L10-13: To analyse the temporal variability of dependence structure, which is modelled by copula's parameter, the proposed bias correction methods are applied separately at each day in June 2014. Due to lack of availability of daily air temperature measurements in 2014 over the study area, copulas and marginal distributions are fitted to the eleven years series of the daily air temperature data.

2- Regarding the "best method", five methods are compared based on three criteria introduced in Section 2.4, P10, L13.

P2, L13:by whom?

- P2, L13-L14, this line is revised and references are inserted as: Currently, the Gamma and empirical distributions are specifically used for bias correction of precipitation data (Lafon et al. 2013; Kum et al. 2014) and the Gaussian distribution for bias correction of temperature data (Teutschbein and Seibert 2012).

P2, L19:ref?

- P2,L19, reference is inserted as: (Mao et al. 2015).

P2, L20 :where is 'here'? who is doing this?

- P2, L20, this sentence is revised as: In copula-based methods,…

P3, L5:add " ' ability"

- P3, L5, this is inserted as: evaluating the ability of these methods….

P3, L6: "ungauged" instead of "unvisited"?

- P2, L6-7: we defined "unvisited locations".
- In a weather station, rain gauge is used to measure rainfall and thermometer is used to measure temperature. In this paper, the variable of interest is air temperature.

P3, L7: add "the proposed methods with"

- P3, L7, this is inserted as: comparing the proposed methods with…

P3, L10: add "a"

- P2, L10,this is inserted as: …when a relatively low…

P3, L12: add "s"

- P3, L12, this is inserted as: …concept of copulas…

P3, L22: this, I do not follow, it will be one value the way it's written. rather 'two variables in space and time, such as u1(s,t) and u2(s,t) at common locations s and instants t'

- P3, L22-25, these lines are removed.
- P3, L22 and L26-27, the sentence is revised as: A bivariate copula describes the dependence structure between two variables at one point in time and space, where U_1 is treated as "true" variable and U_2 is biased variable.

P3, L28: Remove all "the" before copula names

- P3, L28, “the” are removed.

P4, L3: rather Z1(s,t) and Z2(s,t)?

- P4, L3, it is included as: the measurements from weather stations denoted by $z_1^{t,s}$, and weather forecasts denoted by $z_2^{t,s}$….

P4, L6:surely Cu2s,t(u1) including s as well?

- As the reviewer #1 suggested, the notations *s* and *t* are removed from copula formula.
- P4, L8-10 further explanations, however, are provided about s and t as: Throughout, the copulas vary time;. We assumed spatial stationarity during each moment in time to estimate $C_{u_2}(u_1)$.

P4, L6: s and t?

- P4, L6, this line is removed.
- P4, L2, further explanation is added as: The bias at a single moment t in time and location s in space is defined….

P4, L7:the dependence on u2 is lost in pu1 should be pu1/u2

- P4, L7, it is revised as: $p_{u_1|u_2}$

P4, L9 :Aha!! at last!! put this sentence up front as well

- P4, L9, this line is removed.
- P4, L2, further explanation is added as: The bias at a single moment t in time and location s in space is defined….

P4, L9: add “during each t”

- P4, L9, it is inserted into the text as: We assumed spatial stationarity during each moment in time to….

P4, L11: add “spatially”

- P4, L11, it is inserted into the text as: …unlikely to change spatially in a non-stationary way ….

P4, L14: add (i = 1, 2)

- P4, L14, it is inserted into the text.

P4, L15: where k = 4 or 5

- The dataset is changed. For eight weather stations, the daily minimum and maximum air temperatures are available for the periods 1-30 June 2004 to 2014. The quality of measurements and number of missing values, however, differ at each stations (Table 2). The proposed bias correction methods are used to predict bias-corrected values at each day in June 2014, separately. Then, in cross validation, marginal distributions and copulas are fitted to 77 pairs data when there are no any missing values at weather stations. Therefore, k=4 or 5 is not correct.

P4, L16: cannot be 'approximately' as ranks are increasing integers

- P4, L16, the reason for 'approximately' is that marginal quantiles obtained by rank-order-transformation in Eq (5) are not continuous.
- P4, L16, this line is revised as:… which are discrete and now approximately …

P4, L17: add “are prevented from occurring”

- P4, L16-17, it is inserted as: Extreme values that possibly exist in the observations, however, are prevented from occurring ….

P4, L18: “ungauged” instead of “unvisited”?

- P2, L6-7: we defined “unvisited locations”.
- In a weather station, rain gauge is used to measure rainfall and thermometer is used to measure temperature. In this paper, the variable of interest is air temperature.

P4, L19: you fit a 3rd degree polynomial which is notorious for 'flying away' at the ends - why not fit a spline?

- P4, L19, as the reviewer suggested, now spline is fitted to marginal quantiles and this line is revised as: …a spline is fitted to …

P4, L20:how many weather stations? Aha - I see 5 in Figure 1 - that's very low - especially to fit a copula day-by-day. do you pool the estimates to refine the estimated copulas? Please explain the steps more clearly

- P11, L25-29: One of the objectives of this project is to produce daily air temperature map from point measurements to be used in crop growth simulations for assessing near-real time crop and irrigation water requirement. Considering the importance of June in the crop calendar of the study area which is the end of winter crops and beginning of summer crops especially maize, we applied the proposed methods to available dataset of this month.
- P11, L29-31 and P12, L1-4: Eight weather stations (Table 2) were selected because they had a long range of air temperature measurements available and were well spread over the study area. Minimum and maximum distances between stations are 13 and 78 km, respectively (Figure 3). For all weather stations, the daily minimum and maximum air temperatures are available for the periods 1-30 June 2004 to 2014. The quality of measurements and number of missing values, however, differ at each stations (Table 2).
- P12, L10-13: To analyse the temporal variability of dependence structure, which is modelled by copula's parameter, the proposed bias correction methods are applied separately at each day in June 2014. Due to lack of availability of daily air temperature measurements in 2014 over the study area, copulas and marginal distributions are fitted to the eleven years series of the daily air temperature data.
- P4, L29-32 and P14, L18-19, goodness of fit test is done for copulas and p.values are listed in Table 4.
- P4, L13-32 and P13, L24-33 and P14, L1-19 , steps are explained.

P4, L24: add "for"

- P4, L24, it is done as: except for Student's

P4, L21: In Table 4 a different copula is fitted to each of 60 days of 5 observations, when there is relatively stationary behaviour of the observations in each month. Why not pool the monthly data and get a decent relationship? Why not go the whole hog and treat each of 12 months and get some macro- in stead of just disjointed micro-structure?

1- The dataset is changed. For eight weather stations, the daily minimum and maximum air temperatures are available for the periods 1-30 June 2004 to 2014. The quality of measurements and number of missing values, however, differ at each stations (Table 2). The proposed bias correction methods are used to predict bias-corrected values at each day in June 2014, separately. Then, in cross validation, marginal distributions and copulas are fitted to 77 pairs data when there are no any missing values at weather stations. Therefore, different copulas are fitted to more than 5 observations at each day.
2- Regarding the justification of dataset, the data is justified in the Case study section as:
   - P11, L25-29: One of the objectives of this project is to produce daily air temperature map from point measurements to be used in crop growth simulations for assessing near-real time crop and irrigation water requirement. Considering the importance of June in the crop calendar of the study area which is the end of winter crops and beginning of summer crops especially maize, we applied the proposed methods to available dataset of this month.
   - P11, L29-31 and P12, L1-4: Eight weather stations (Table 2) were selected because they had a long range of air temperature measurements available and were well spread over the study area. Minimum and maximum distances between stations are 13 and 78 km, respectively (Figure 3). For all weather stations, the daily minimum and maximum air temperatures are available for the periods 1-30 June 2004 to 2014. The quality of measurements and number of missing values, however, differ at each stations (Table 2).
   - P12, L10-13: To analyse the temporal variability of dependence structure, which is modelled by copula's parameter, the proposed bias correction methods are applied separately at each day in June 2014. Due to lack of availability of daily air temperature measurements in 2014 over the study area, copulas and marginal distributions are fitted to the eleven years series of the daily air temperature data.

P4: If you are doing cross-validation, 5 sites is very few

1- We want to emphasise on the problem definition as:

- P1, L23-24, the problem is defined as: Weather stations are often sparse and usually located at irregular positions. If their data are used for crop growth simulations, then their results at unvisited locations are likely to be uncertain.
- P2, L4-L8, further explanations are added as: Due to the coarse spatial resolution of weather forecasted gridded data, there is an apparent mismatch between measurements obtained from weather stations and weather forecast data. In this study, unvisited locations are grid points which do not contain an observation due to the relatively low number of weather stations in the study area. In order to obtain unbiased values, a bias correction method should be applied for these grid points before using the weather forecast data.

2- We were able increase the number of stations to eight and use the historical data at each day to deal with low number of observations as:
    - P11, L25-29: One of the objectives of this project is to produce daily air temperature map from point measurements to be used in crop growth simulations for assessing near-real time crop and irrigation water requirement. Considering the importance of June in the crop calendar of the study area which is the end of winter crops and beginning of summer crops especially maize, we applied the proposed methods to available dataset of this month.
    - P11, L29-31 and P12, L1-4: Eight weather stations (Table 2) were selected because they had a long range of air temperature measurements available and were well spread over the study area. Minimum and maximum distances between stations are 13 and 78 km, respectively (Figure 3). For all weather stations, the daily minimum and maximum air temperatures are available for the periods 1-30 June 2004 to 2014. The quality of measurements and number of missing values, however, differ at each stations (Table 2).
    - P12, L10-13: To analyse the temporal variability of dependence structure, which is modelled by copula's parameter, the proposed bias correction methods are applied separately at each day in June 2014. Due to lack of availability of daily air temperature measurements in 2014 over the study area, copulas and marginal distributions are fitted to the eleven years series of the daily air temperature data.

P5, L8: I'm sorry to say that I tried to trawl through Section 2.3 several times but gave up - it's too 'busy', complicated and not well explained.

- P5, L8-17, this section is completely revised.

P8, L2: Oh? how's that measured? x, y & e are spatial coordinates giving elevation above sea level at a point in plan. R increases at x and y increase, even over a flat surface where e is constant, and R is then constant along a circular path in plan???

- P8, L2-3, this sentence is removed.
- P7, L27-29, the variable R is replaced by the elevation.

P8, L17: turns into space divided by 5 Thiessen polygons

- P8, L17-19, it is explained that: To select the nearest neighbour to an unvisited location, the distance between two locations is calculated using three dimensional coordinates.
- P8, L15-16, further explanations are added as: In BCQM-II, one bivariate copula describes dependence structure between forecasted air temperature and its nearest neighbour, and another copula describes dependence structure between measurements and its nearest neighbour.

P9, L4: As I emphasize, because this QS is the best of all the methods, doing what you want to do, you should take the trouble to explain it more carefully, instead of using the many back references, which are annoying.

- P9, L4, this section is completely revised and we added Figure 2 for further illustration.

P10, L9: ??

- P10, L8-11, the paragraph is revised as: There are several methods that lead to the minimization of the fitness function (Burke and Kendall, 2014). In this study, we applied a genetic algorithm for doing the search. Details on this algorithm can be found in the literature (Sastry et al., 2013) and are beyond the scope of this paper. The sample code to implement in R, however, is given in the appendix 3.
- P18,L13 and P19, L33, the references are added.
- P21, L11-17, appendix 3 is added.

P12, L1: What is the logic here? you limit the observations to only 2 months!! you throw away the 'long range of air temperature measurements'!!
P12, L8: only 2 months?

- P11, L25-29: One of the objectives of this project is to produce daily air temperature map from point measurements to be used in crop growth simulations for assessing near-real time crop and irrigation water requirement. Considering the importance of June in the crop calendar of the study area which is the end of winter crops and beginning of summer crops especially maize, we applied the proposed methods to available dataset of this month.
- P11, L29-31, P12, L1-4: Eight weather stations (Table 2) were selected because they had a long range of air temperature measurements available and were well spread over the study area. Minimum and maximum distances between stations are 13 and 78 km, respectively (Figure 3). For all weather stations, the daily minimum and maximum air temperatures are available for the periods 1-30 June 2004 to 2014. The quality of measurements and number of missing values, however, differ at each stations (Table 2).
- P12, L10-13: To analyse the temporal variability of dependence structure, which is modelled by copula's parameter, the proposed bias correction methods are applied separately at each day in June 2014. Due to lack of availability of daily air temperature measurements in 2014 over the study area, copulas and marginal distributions are fitted to the eleven years series of the daily air temperature data.

P12, L13-14 and L18-20 : This is surely not a surprise, nor a disadvantage. you have two estimates of temperature: uniformly tiled ECMWF 'data' and point weather station data. If you believe the station values to be 'true' and useful for agriculture, then you transform the ECMWF values to station scale all over the tiles, which it seems you have not smoothed in Figs 6 & 7. Perhaps this paragraph should be relocated to the introduction?

- P12, L13-14 and L18-20, as suggested this paragraph is moved to the introduction, P2, L4-8.

P12, L27-30: how do you determine outliers? - just because they are low? - how about Station 1 June 20? - I've ringed it in the figure.

- Outlier were responsible for negative temperature in March. Based on our knowledge, negative temperature in March is a strange value in the study area. Aggarwal 2013 defined outlier as : "Outliers are also referred to as abnormalities, discordants, deviants, or anomalies in the data".
- P12, L27-30, this paragraph, however, is removed due to removing data in March. We now limit the implementation to data in June.

P13, L1-2: what do you mean? why the fuss?

- P13, L1-2, these lines are removed due to removing data in March.

P13, L7: why not use this information directly rather than model with copulas?

- P3, L1-11, the aim of this study is further explained as: In this study, we aim for:
  - estimating different conditional quantiles at different unvisited locations accounting for the temporal variability of the dependence structure.
  - evaluating the ability of these methods to predict the spatial variability of the bias-corrected daily air temperature at unvisited locations.
  - comparing the proposed methods with available bias correction methods, which are quantile mapping, expectation predictor and single quantile predictor.
  - providing a review and application of these methods for bias correction of the daily air temperature data when a relatively low number of observations are available.
- P13, L6-8, we revised it as: The average of bias for all stations and all days equals 4.5°C in June 2014. Since there is both spatial and temporal variability in the bias, we were not able to correct for bias at one day and an unvisited location using the average value of bias.

P13, L14: the treatment of the standard deviations is unfair - better to use Coefficient of variation?

- Standard deviation is the second moment of a distribution. The coefficient of variation is used by (Lafon et al. 2013) as one criteria in evaluating the performance of bias correction methods. As suggested, however, the coefficient of variation is considered in this study, too

- P13, L14-15, these lines are revised as: Figure 6 shows the mean, sample standard deviation, coefficient of variation, skewness and kurtosis for both observed and forecasted values at each day from all weather stations.
- P11, L5-6, these lines are revised as: …moments of the marginal distribution; mean, standard deviation (as well as coefficient of variation), skewness and kurtosis,…

P13, L27: Poly(3) bad choice - better to use splines which behave better at the ends. however see my note below Fig 5.

- P4, L19, as suggested by the reviewer, now spline is fitted to marginal quantiles and this line is revised as: …a spline is fitted to …
- P13, L27, it is revised as: the fitted spline are presented…

P13, L30: I cannot understand why you limit the analysis to these 2 months - what's the matter with the other 10? their behaviour might help the analysis ..

- P11, L25-29: One of the objectives of this project is to produce daily air temperature map from point measurements to be used in crop growth simulations for assessing near-real time crop and irrigation water requirement. Considering the importance of June in the crop calendar of the study area which is the end of winter crops and beginning of summer crops especially maize, we applied the proposed methods to available dataset of this month.
- P11, L29-31 and P12, L1-4: Eight weather stations (Table 2) were selected because they had a long range of air temperature measurements available and were well spread over the study area. Minimum and maximum distances between stations are 13 and 78 km, respectively (Figure 3). For all weather stations, the daily minimum and maximum air temperatures are available for the periods 1-30 June 2004 to 2014. The quality of measurements and number of missing values, however, differ at each stations (Table 2).
- P12, L10-13: To analyse the temporal variability of dependence structure, which is modelled by copula's parameter, the proposed bias correction methods are applied separately at each day in June 2014. Due to lack of availability of daily air temperature measurements in 2014 over the study area, copulas and marginal distributions are fitted to the eleven years series of the daily air temperature data.

P14, L28: why not say this up front in the introduction and abstract? it's far and away the best

- In Section 2.4, there are three criteria to evaluate five bias correction methods. The results are illustrated in Tables.
- P1, L16-21, the advantages of new methods are mentioned.

P14, L28: unfortunately QS is poorly explained in 2.3.6 with many back references

- The section 2.3.6 in P8-P10 is completely revised and we added Figure 2 for further illustration.

P15, L3:why not pool the daily values in each month?

- P11, L25-29: One of the objectives of this project is to produce daily air temperature map from point measurements to be used in crop growth simulations for assessing near-real time crop and irrigation water requirement. Considering the importance of June in the crop calendar of the study area which is the end of winter crops and beginning of summer crops especially maize, we applied the proposed methods to available dataset of this month.
- P11, L29-31 and P12, L1-4: Eight weather stations (Table 2) were selected because they had a long range of air temperature measurements available and were well spread over the study area. Minimum and maximum distances between stations are 13 and 78 km, respectively (Figure 3). For all weather stations, the daily minimum and maximum air temperatures are available for the periods 1-30 June 2004 to 2014. The quality of measurements and number of missing values, however, differ at each stations (Table 2).
- P12, L10-13: To analyse the temporal variability of dependence structure, which is modelled by copula's parameter, the proposed bias correction methods are applied separately at each day in June 2014. Due to lack of availability of daily air temperature measurements in 2014 over the study area, copulas and marginal distributions are fitted to the eleven years series of the daily air temperature data.

P15, L7:but these uncertainties are all pooled into SMAE, so why define R as in Eq (11)??

- P15, L7-8, this line is removed.

- P10, L21: The spatial mean absolute error (SMAE) is calculated at each weather station.
- P7, L29, the variable R in Eq (11) is removed.

P16, L23-25: poor choice - pool the days into months - you do not even look at serial ccs, which might have added value.

1- Regarding "poor choice", P4, L19, spline is used for fitting marginal distribution instead of a polynomial, as suggested.
2- Regarding "pool the days into months", P11-12 the justification of dataset is mentioned in the Case study section.
3- Regarding "serial ccs":
   - P2, L22-26, more explanations are added to the Introduction section as: A bias correction method proposed by Laux et al. (2011) employed bivariate conditional copulas to model dependence between the daily precipitation time series retrieved from a regional climate model and observations at three locations where data is available. In their method, however, a bivariate copula is fitted to daily time series at one location, ignoring the temporal variability of copula parameter as well as spatial dependency. In addition, the fitting is required to remove autocorrelation and heteroscedasticity which may exist in the time series (Laux et al. 2011).
   - P10, L29-31 and P11, L1-4, it is mentioned that: In addition, the correlation coefficient r between observed and bias-corrected values is calculated at each weather station as: $r^s = \frac{cov\{Z_1^s, \hat{Z}_1^s\}}{\sigma_{Z_1^s}\sigma_{\hat{Z}_1^s}}$; $Z_1^s = \{z_1^{s,1}, z_1^{s,2}, \dots, z_1^{s,T}\}$, $\hat{Z}_1^s = \{\hat{z}_1^{s,1}, \hat{z}_1^{s,2}, \dots, \hat{z}_1^{s,T}\}$, where Z_1^s is the measurement from weather stations , $\hat{Z}_1^s$ is the biased-corrected values obtained by cross-validation, and T is the number of time steps in time series. To compare the five bias correction methods based on the r, an correlation score (CS) is calculated at each weather station. The smallest CS indicates for the smallest r.

P17, L13: Add "One of"

- In Section 2.4, there are three criteria to evaluate five bias correction methods. The results are illustrated in Tables.

P24, L1, table title: ' ... at each day ...' I cannot understand why you do not pool the data for the months and analyse the other 10 months to get the full picture. I think the analysis is over-elaborate in some ways and woefully inadequate in others.

The justification of dataset is mentioned in the Case study section. Please see the answer for the similar comments, too.

P24, L2, table title: 16 out of 30 is only just 'most'! you should have given the modelling procedure something to chew on!

This table is completely revised.

P25, L2: define SAME

It is done.

P25: table title: please describe these in the heading

It is done.

P25, last row in table: these should be reported to 3 significant figures - 2 is too low

The new results will be reported in 3 digits.

P25, L7: change "better" to "best"

It is done.

P26, L5: change "better" to "best"

It is done.

P30, L3: correct km2

It is corrected.

P33, on the figure: outlier??why are these outliers? why not make all obs axes the same spread [14 - 33] to make visual comparison easier?

The figure in P33 is removed. We added new Figure in P34.

P35, figure: pooled moments from 5 sites? averaged over the 5 stations? you've destroyed the clear structure of Figure 3 by being over-complicated? Also, stdev is a poor comparison - rather report coefficient of variation - then the scales will be similar. By the way, the skewness of the data is close to its average of zero in March, but diverges in June. I don't think there are enough data to get a useful kurtosis, even though it averages around 2 ...

- The justification is mentioned in the Case study section. Please see the answer for the similar comments, too. Standard deviation is the second moment of a distribution. As suggested, however, the coefficient of variation is considered in this study, too. It is used by (Lafon et al. 2013) as one criteria in evaluating the performance of bias correction methods.
- The figure in P35 is removed. We added new Figure in P36.

P37, figure: why only 4 points from 5 gauges? in Fig 2 they are all there .... if these are quantiles on the day, why aren't they evenly spaced as i/(k+1), i = 1, .. ,k??because the relationships between Observed and Forecast are relatively stationary within the months, why not rank them monthly and fit a decent curve using least squares - I'm unhappy that you use polynomials - they go wild at the ends - rather use splines through the 150 (pooled from 30 days of 5 gauges') ranked points.

The figure in P37 is removed. We added new Figure in P38.
The justification is mentioned in the Case study section. Please see the answer for the similar comments, too.
Spline is used for fitting marginal distribution instead of a polynomial, as suggested.

P39, figure: for 3 selected days

The figure in P39 is removed.

P40, figure: for 3 selected days

The figure in P40 is removed. We added new Figure in P41. L3 is revised as suggested.