Reviewer 1:

In the article, authors have described a technique to estimate the wind velocity from thermal images. The estimated wind velocity is comparable with the wind speed measurement qualitatively. But quantitatively, e.g. $R^2$ analysis, p-value) the agreement between them is poor. Also many details of the analysis are missing. Thus, I won't recommend the paper for publication in current form. Following are the itemized suggestions:

Answer:

Thank you for your detailed suggestions and comments. We have updated the manuscript according to your concerns.

Reviewer 1:

- How authors calibrate each pixel of the thermal image to physical distance, m? The pixel sizes vary based on the camera viewing angle. Pixels closer to the camera have smaller physical sizes compared to the further away. How authors have corrected this camera viewing distortion of the thermal image? What are the image sizes in physical distance?

Answer:

As stated in Line 321 referred to the revised track change version here and after, distortion showed the maximal deviation the calculated pixel size was 0.05 m. This means
that a pixel closer to the camera in the center of the image is maximally 0.05 m smaller in x- and y-direction than a pixel on the edge of the area of interest. The size of the hokey field is 91.4 x 55 m. The deviation was calculated using the number of pixels between the midfield line and the 22.85 m line and the latter one and the goal line for the x-direction. The same was calculated for the y-direction using the pixels located at the goal lines and at the center line. The pixels on the very edge of the field were excluded from the calculation as stated in line 319. We have further clarified this paragraph (line 314 - 324) and added the size of the areas of interest to the corresponding experiment overviews (Table 1).

Reviewer 1:

- How did authors perform the spatial average of the 15m x 15m patch? From figure 5, the velocities are estimated when temperature gradient is present. Does the patch size ensure no missing wind velocity estimate for the pixels considered? Why specific size of patch is chosen?

Answer:

The average was performed with a spatial mean function over the 15m x 15m patch. The A-TIV method ensures that less vacant grid cells are calculated compared to any TIV performed (see Figure 5b and improvement in vector spatial density relative to the missing values in figure 5a). The vacant grid cells of the TIVs resulting from higher perturbation filters are adapted to the predominant perturbation time scale by the weighted average resulting in the A-TIV output velocity grid. Therefore, the vacant grid cells of higher perturbation TIVs (30s and 20s) are not resulting in 0 or NA values for the spatial averaging. Small velocities (< 0.5 m/s) may not be displayed with the vectors in the figures. The patch size was chosen based on the available upwind area of the sonic anemometer within the field of view of the camera and the size and the calculated footprint (see answer to question 3 for the footprint calculation) To avoid corner effects i.e. from obstacles and to reflect the core of the calculated footprint, we decreased the size of the averaging area to 15 m x 15 m. We have added a clarification to the manuscript (line 338-342 and line 365-369) alongside with the new calculations of the flux footprint in response to reviewer question 3 below).

Reviewer 1:

- For comapraison with the sonic anemometer measurement, how the location of 15m x 15m patch is chosen? Does it lie inside the footprint of sonic anemometer? If not, why authors have not considered some flux footprint model, as descibed in Garai et al. 2013, Boundary Layer Meteorology?

Answer:

We have added new material (Appendix) for the estimation of the footprint of the sonic anemometer using the UMEP plugin for QGIS which allowed us to pick a suitable area for the averaging (Lindberg 2018). We have now prepared a more detailed view in the appendix. We have also referenced the Appendix in the text where needed.

Reviewer 1:

- What are the weights for averaging wind velocities from different perturbation? How the
weights are chosen?

Answer:
The TIVs within the A-TIV process resolve various perturbation scales (30s – 5s in the present manuscript). With decreasing perturbation time scale the amplitude of the perturbation is decreasing and the noise level of the camera is becoming more present in the perturbation signal. Therefore, we have decided that the resulting TIVs are weighted according to the perturbation scale its calculation is based on. In our case this is 30s – 20s – 10s and 5s in this manuscript meaning weights of sixfold, fourfold, twofold and onefold (based on perturbation filter divided by 5). We have rephrased the paragraph explaining this procedure in line 131 – 134.

Reviewer 1:
- Figure 7, how the difference is calculated? Not all the pixels result velocity estimation.

Answer:
The difference velocity fields were based on pixel to pixel subtraction. A-TIV creates an velocity estimation for each pixel. This may be that the velocity is 0 and hence no difference is estimated (white pixels). The purpose of this figure is to show the added information associated with the ATIV approach relative to the perturbation average window. Through the weighted averaging (see question 2), more pixels are resulting in a velocity estimation compared to any single TIV. Additionally, the differences between A-TIV and a single TIV result shows the need for the A-TIV.

Reviewer 1:
- Lower value of p-estimation means that the assumed null hypothesis does not hold. For present study, what is the null hypothesis, is it TIV corresponds to wind velocity? If so, then the reported small p-value means there is no correlation.

Answer:
Your statement is correct. This analysis used the null hypothesis (H0) that the means of the two distributions do not match. Hence the low p-value. We have adjusted H0 and report now a higher p-value as per common practice (line 263 - 265).

Reviewer 1:
- Figure 9a. Too many wiggly lines make it difficult to read. Authors should consider separate out the temperature and velocity curves in two separate figures.

Answer:
We have adjusted the figure and separated the lines into two plots.

Reviewer 1:
- Figure 9 shows that the TIV and sonic anemometer have some temporal lag. It also looks like the lag and wind direction are not constant for the time period considered.
How the authors account for these effects when calculating $R^2$, p-value and histograms for quantitative comparison? Also how these variable effects the comparison.

Answer:

Figure 9 may show a temporal lag due to multiple reasons:

- As mentioned in line 304 - 308 this lag may be caused by the resetting mechanism of the thermal camera.
- Essentially the sonic anemometer and the A-TIV do not necessarily reflect the same atmospheric turbulence. While the sonic anemometer mounted at 1.5 m height reflects a larger footprint the A-TIV only measures in-situ temperatures and spatial perturbation changes between images with the likelihood of measuring structures very close to the surface (within centimetres above the surface).

The p-value expresses only statistical (distribution) relationships between the measurements and not direct correlations. It is shown in the manuscript that the thermal imagery and A-TIV reflects the turbulent flow ~1.5 cm above ground whereas the sonic anemometer reflects a point measurement at 1.5 m above ground with a spatial footprint. Hence a direct correlation value such as the $R^2$ is lower because the A-TIV measurement cannot directly explain all atmospheric turbulence measurements of another sensor of a different type. We have added a dedicated paragraph in the discussion section explaining this (line 338-342).

Reviewer 1:

- Figure 10, What are the markers in the figure?

Answer:

The Figure caption states that the black arrows mark the "minimal resolvable wind speed of the lag-correlation method from the thermocouples"

Reviewer 1:

- Page 17, line 261: What do you mean by positive wind speed? Wind speed is always positive.

Answer:

We have rephrased the corresponding sentence (line 280).

Reviewer 1:

- For Turf-T2 why authors have not considered to have two thermocouple arrays on the two surfaces, instead of putting one thermocouple array in the mixed surface. As the surface properties are changing, a new boundary layer will start to develop. How authors account for that in the analysis.

Answer:
Due to limited amount of Thermocouples it was not possible to create two separate arrays. The Analysis shows that the lag correlations performed between temperature perturbations measured with “physical” thermocouples and with “virtual” thermocouples with the same surface cover result in very similar estimations. However, when the virtual array was moved in either surface cover the estimations no similarity was determined. Therefore, it was decided to separate the A-TIV calculations for both surfaces. An explanation has been added to the methods section (line 228-229).

Reviewer 1:

- Garai and Kleissl 2013, Journal of Turbulence, reported that the different temporal filtering result thermal structures corresponding different scale. How authors account for that when comparing averaged TIV from different temporal perturbation with sonic anemometer? The small flow structure giving the TIV 5s perturbation may not be registered at the sonic anemometer.

Answer:

The thermal interactions of structures with shorter time scales create a reduced amplitude in thermal perturbations measured by a sonic anemometer or measured by thermal imagery. Without a time-frequency decomposition on the signal, the sonic anemometer registers these as one time series integrating these scales and represented by a measured temperature or wind velocity. The TIV does not directly reflect this frequency composition and may reflect only certain frequencies due to the decomposition done in the calculation of the perturbation and the Hilbert-Huang Transform. Higher frequencies and sensor noise are neglected in this way. To mitigate this indirect focus and limitation to certain frequencies we introduced the A-TIV composition of multiple perturbation windows which then allows to reflect the compositions of multiple frequencies. In our case we decided to use 30, 20, 10 and 5s because our thermal sensors noise was only starting to interfere with the 5 s perturbation window. Therefore, we decided to use a weighted averaging system because it is expected that the noise level rises with the decrease of the temporal perturbation filter size. This is the case for any sensor used and is not limited to the thermal camera. This paragraph has also been added to the manuscript alongside with further comments on the running temporal filter size (line 343 – 353 and line 377 - 380)

Reviewer 1:

- English in the article is poor.

Answer:

Thank you for all your comments and suggestions. We have corrected multiple grammatical, wording and spelling mistakes.

References:
