Dear Reviewer,

It was likewise a pleasure for the authors of this work to receive Your welcoming reviews and encouraging comments. We wish that the positive but careful approach exercised by You were used more often in the community!

My responses are given in blue.

Line 32: "sizes" to "size". - done

Line 33: "durations" to "duration". - done

Line: 34 remove "now". - done

Line 37: remove "already" - done

Line 46: "depend" to "depends". - done

Line 100: "of" to "is a" - the DJI camera XT2 is du al-sensor, i.e. it combines both a thermal and an RGB sensor in one body. That is why I wrote "thermal sensor of XT2 is FLIR Tau 2".

Line 123: What do you mean by "decimated"? Did you cherry pick every 30th image, or did you use a window mean on every second of data? I'm working on a dataset right now and these differences are showing up in the final results. - The original 30 Hz thermal sequences were downsampled to 1 Hz, i.e. every 30th image was kept. Averaging over the 30 images recorded each second is unnecessary in my opinion, as there is no noise that needs elimination on that level - and even harmful, as small-scale structures would become lost. I hope to upgrade to the 10 Hz data analysis in a future study, but this will require a reduction of errors emerging in image co-registration. The upper limit on meaningful frequencies is, of course, imposed also by the random thermal noise of the camera, and may be lower than 10 Hz.

Line 162: "An own"? Perhaps replace with "We developed a method,.." or something of the sort. - done

Line 174-175: Incomplete sentence "...the perturbed and template images is achieved." - corrected; the sentence reads now "The process is continued until convergence between the perturbed and template images is achieved";

Figure 3: Does the colorbar in (d) go with (c)? Or are these colors just representing shape? It seems like the structure annotated in (d) has a mean negative temperature, but based in the colorbar in (c) this structure would have a mean positive temperature. Maybe just make a note in the caption that says (c) only represents the shape of the structure. – added to the figure 3 caption: The colors in (c) are only to tell the structures apart and do not correspond to the color bars in the other panels. Line 244: "1-m" - done

Line 278: Comma "..., as well as" - done

Line 292: put the equation in paratheses $(=\langle w'T' \rangle)$ - done

Line 316: Comma "complex," - done

Line 341: "The information provided in Fig. 5 is visualized spatially in Fig. 6." Isn't figure 5 the spatial standard deviation of T' written in line 329, and figure 6 is the temporal standard deviation of each pixel? –Actually, this is the same quantity, T', presented in Fig. 5 as a statistical distribution, and in Fig. 6 spatially by plotting the maps of its per-pixel standard deviation. Section 3.4: This is my favorite section. Good work! - Thanks a lot for the good words! Line 430: Just make sure you're consistent with putting your variables in italics - done Line 472: "14-m" Use a dash when using numbers as an adjective. - done

Line 523: Recording was at 30 Hz? Or are you saying the inertial subrange is captured with 10 Hz measurements? I would say that the inertial subrange in the atmospheric boundary layer starts at frequencies much smaller than 1 Hz (actually at around 0.01 Hz), 10 Hz is already well into the dissipation subrange. The camera does record at 30 Hz, but this is certainly an excessive rate for the atmospheric applications.

Kind regards,

Pavel Alekseychik