The paper describes a new sea-surface temperature product merging SEVIRI data and results from a data-assimilative model. The important novelty is that this product resolves the diurnal cycle and provides full fields without gaps (level 4). The authors also include a quite detailed comparison with in situ observations. While this paper describes the results for the years 2019 and 2020, this data set is continuously updated and the results for the year 2021 are also available.

The main questions that I asked myself while reading the manuscript are:

A. As there are different depths for the different types of SST (skin-temperature, bulk temperature, foundation temperature) which depth level is the DOISST targeting by this product? I understood that the model and SEVIRI data have different reference depths. Should there not be first a conversion/adjustment, so that the temperature is comparable? Maybe interpreting some of the conclusions within this context would be useful.

B. Comparison: I would have expected a comparison to show that DOISST is better (compared to in situ observations) than other observational products. However, the author compared the new product to a model solution. Are there other data L4 products available (resolving the diurnal cycle) based on SST data from geostationary satellites? In any case, the authors also compare the accuracy of their product (relative to drifters) to the accuracy of the SEVIRI data (at exactly the same location) which already shows some quite favorable results.

I recommend publications after minor revisions.
Minor comments:

1. line 106: assessment of the MED DOISST product covers two complete years (2019-2020)

Please clarify earlier in the manuscript the time coverage of the data product and the time coverage of the assessment.

2. degree K (line 19, abstract) or degree C (line 39, introduction). Can you please use the same units?

3. an overview table with all products would be useful, including resolution (time and space) and coverage (time and space) and reference depth (e.g. skin, subskin, foundation temperature,...), even if the study uses a subset of the input data set. This table could also include the new dataset.

4. typesetting of the equation should be improved and follow the style of other Copernicus papers.

5. page 9: "All these parameters have been deduced from a statistical analysis of the satellite SST data"

Please give more information about how you choose the particular parameters (a, c, d, decorrelation spatial length R, decorrelation time length T). In particular, what objective criterion was used to decide that these parameters are appropriate?

6. page 10, line 250: "At each step of decreasing n, data that falls out of the interval I = [mean(delta) - n sigma, mean(delta) + n sigma] are flagged. The process starts for n=10 and stops at n=3." If the data is outside of the interval for n=3, why would one also check for n=10? But I guess that delta (the difference, and the mean and standard deviation) also depends on n by selecting a different subset for different n. I think that this should be clarified in the proposal.

7. line 295: "The two diurnal cycles are practically coincident between 17:00 and 06:00, while they are biased by ~0.1 K between sunrise and 16:00, coherently with the DOISST bias oscillation (Fig. 3). This bias could be related to skin SST getting warmer faster than 20 cm temperature"
I suggest you replace "20 cm temperature" by "temperature at 20 cm depth". I am not sure if "coincident" is the right word. What about saying that the bias is close to zero (DOISST and drifter temperature) as you do not show the diurnal cycles of DOISST and drifter temperature individually.