RC2: 'Comment on essd-2021-462', Anonymous Referee #2, 22 Feb 2022 reply

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.

DOISST is indeed the result of a blending of SEVIRI sub-skin SSTs, representative of a depth of 1 mm, and modeled SSTs at 1 m. Then, the DOISST effective depth does, in principle, vary between 1 mm up to 1 m, depending on how many satellite observations enter the interpolation. As diurnal warming is significantly reduced under cloudy conditions, however, the difference between the SST at 1 m and the sub-skin SST will be much smaller when SEVIRI observations are not present. For this reason, we can define the DOISST product as representative of sub-skin values. We added this concept in Section 3.1 of the revised manuscript (lines 237-242). We also added in Table 2 the depth level of the DOISST product.

Concerning the second question, we definitely agree with the referee. The ideal case would be if all data were generated and compared at the same depth. Unfortunately, the first model layer is centered at 1 m depth, while sub-skin SST is, by definition, representative of a depth of 1 mm. In principle, it could be possible to correct all the data, bringing them all to the same depth before any comparison or merging, by applying some model (see e.g. Zeng et al., 1999). However, any correction algorithm would have added potential uncontrolled error sources (e.g., related to ancillary data and/or to model assumptions) and implied significant additional operational efforts. For these reasons, rather than trying to correct the first-guess bias, we preferred to leave it uncorrected, and focus on optimising the corrections driven by available hourly satellite data. This concept has been added in Summary and Conclusions (lines 536-542).

Additionally, a new section (3.2 in the revised manuscript) should clarify this point. As written from line 258: “The choice of using a model output as first-guess represents the best alternative to the use of climatologies or previous analyses, as usually done by other schemes to produce daily SST L4 maps, since the model provides physically consistent estimates of hourly SSTs in the absence of any observation or in situ measurement (Marullo et al., 2014). In fact, the model takes into account the effect of air-sea interactions by imposing external forcings that drive momentum and heat exchanges at the upper boundary. As such, it is able to reproduce at least part of the diurnal warming effects, that are driven by the forcing diagnosed from atmospheric model analyses. Using the model output as a first-guess means we are treating the hourly satellite data as corrections to the hourly model data. These anomalies are generally small and mostly drive corrections to the spatial patterns, while displaying a reduced diurnal cycle. Anomaly data from different times of the day can thus be more “safely” used to build

the interpolated field at each reference time (with different weights). Unfortunately, the first model layer is at 1 m depth, which means that it will generally underestimate the diurnal cycle anyway. While 1D models could in principle be used to better reproduce sub-skin SST from model data, the approach presented here is focusing on providing estimates that are as close as possible to the original satellite data, avoiding the complications of setting up an additional preprocessing step just to improve the first-guess."

*Zeng, X., Zhao, M., Dickinson, R. E., & He, Y. (1999). A multi-year hourly sea surface skin temperature dataset derived from the TOGA TAO bulk temperature and wind speed over the tropical Pacific. Journal of Geophysical Research, 104, 1525–1536.

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.

We do agree with the reviewer and thus added the global operational diurnal L4 SST OSTIA product in our intercomparison exercise. In particular, as OSTIA ingests the in situ data we used as reference for the validation (which would not be independent for OSTIA), we included OSTIA diurnal in section 4.2.2, which is dedicated to the reconstruction of diurnal warming amplitudes (DWAs) from different sources (DOISST, Model, SEVIRI and In situ data). The first part of the validation (section 4.2.1) is indeed mainly thought to assess the accuracy of the DOISST product against an independent in situ data source, and the inclusion of modelled SST data is thought to evaluate the DOISST performance with respect to the model, which is used as first-guess.

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.

Clarified (see lines 114-117).

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

Corrected.

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.

An overview Table has been added (i.e., Table 1 of the revised manuscript).

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

This has been corrected.

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?

This has been clarified (see lines 302-303): "All these parameters have been derived in Marullo et al. (2014), deduced from a nonlinear least square fit between the estimated temporal and spatial correlations."

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.

This sentence is actually unclear. It has been re-written. See lines 339-346 of the revised manuscript.

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".

Replaced.

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.

Figure 4 shows the mean diurnal SST cycle as reconstructed by DOISST, model and drifters, while the bias is shown in Figure 3. However, ae have substituted coincident with unbiased.