RC1: 'Comment on essd-2021-462', Anonymous Referee #1, 21 Feb 2022 reply

This study presents a high resolution (1/16-deg and 1-hr) DOISST using SEVIRI satellite as source and a model SST as first-guess, which will no doubt have many applications. Their analysis demonstrated that DOISST well represent the diurnal cycle with a low mean bias and RMSD of 0.4ºC. The seasonal features of SST diurnal cycle in MED area are described and compared with independent buoy SST and model simulation. My major concern is how the model SST at 1 m depth is used as the first-guess field. It is not clearly stated what depth the DOISST represent, and how the DOISST is validated with buoy SSTs at 0.2 m depth. The ideal case is that all three components are compared/generated at the same depth level. I recommend accept the manuscript after a major revision.

First, we thank the reviewer for the stimulating comments and questions. Overall, the reviewer helped us understand that several key aspects lacked clearness and detail and we do agree that they need to be introduced and described with more detail. This could partly be originated from having taken for granted some of these aspects, as e.g. the different types of SST definitions and some commonly adopted validation procedures. In any case, following the reviewer's comments, we worked to improve the level of detail.

We will start answering the more general comments/questions of the reviewer, then we will provide answers to the more specific comments.

"My major concern is how the model SST at 1 m depth is used as the first-guess field".

The first model layer is unfortunately centered at 1 m depth, so we used the SST at this level to produce an anomaly field that represents the difference between the observed hourly satellite and our first-guess. This anomaly is the variable that we interpolate over data voids using our Objective Interpolation (OI) scheme. This first-guess choice is a better alternative to the use of climatologies or previous analysis data, as operated by other schemes in producing daily SST L4 maps (see Marullo et al., 2014), since it gives the best estimate of hourly SSTs in the absence of any observation or in situ measurement. This choice simplifies the computation of the space-time covariance function that is used to weight the input observations within the OI algorithm. In fact, figure 3 of Marullo et al. (2014) shows the behavior of the correlation function versus time when either the hourly model or the SST daily climatology are subtracted. It is evident that, in the second case a strong daily component is present, while in the first case, the daily component is significantly reduced, allowing to state that the data are "nearly" free from the diurnal cycle. Anomalies observed at different times can thus be combined to better describe the diurnal warming patterns. Indeed, as we aim to retrieve in the most accurate way the spatial pattern of the surface warm anomalies (which evolves over time), it is desirable that observations that are closer in time are weighted more by the OI algorithm (through covariance) than distant ones.

Of course, we do agree that it would have been better to use a sub-skin model SST as first-guess but, at the present time, such a product is not available. 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.

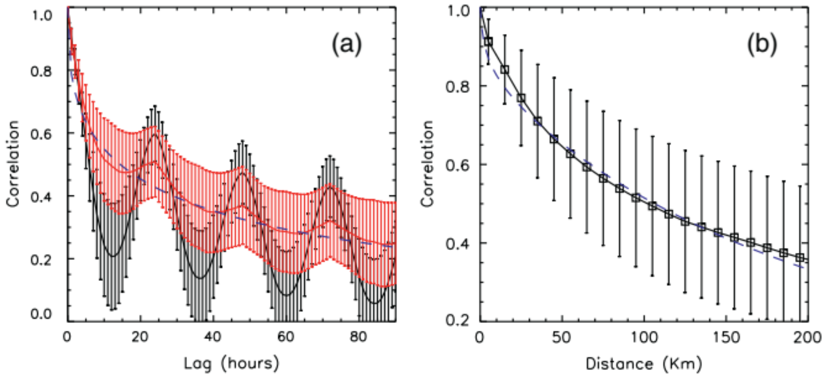

Fig. 3. Covariance structure function of the Mediterranean Sea estimated from summer 2011 SEVIRI data. (a) Temporal covariance: in red the covariance is computed using hourly SST anomaly field (hourly SEVIRI–hourly MFS); in black the covariance is obtained using hourly SST field after removal of mean daily SST (hourly SEVIRI–mean daily MFS). Vertical bars represent ± 1 standard deviation. (b) Spatial covariance function at $\Delta t = 0$ for the SST anomaly field. Dotted blue curves in (a) and (b) represent Eqs. (1) and (2) respectively.

These concepts have been included in the revised manuscript (see e.g. Abstract and Conclusions). In particular, we added Section 3.2, which better introduces the optimal interpolation method and the choice of a model output as 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.

"It is not clearly stated what depth the DOISST represent, and how the DOISST is validated with buoy SSTs at 0.2 m depth".

Actually, the depth information is intrinsically contained in the definition of the SST provided by DOISST, namely the sub-skin SST. "This product provides hourly mean maps (Level-4) of sub-skin SST" (as stated now at line 12), and "sub-skin SST is the temperature at the base of the conductive laminar sub-layer of the ocean surface, as defined by the Group of High Resolution SST (GHRSST) (line 91; see also figure 1, and Minnett et al., 2019). In practice, this is the temperature at ~1 mm depth" (line 93). Here, the lack of clarity could be due to the fact that we did not highlight that SEVIRI provides sub-skin SST (as also reported in the SEVIRI product user manual (PUM), see https://osi-saf.eumetsat.int/lml/doc/osisaf_cdop3_ss1_pum_msg_sst_data_record.pdf), and, being SEVIRI the predominant input into the optimal interpolation scheme, for consistency our DOISST provides sub-skin SST. In section 3.1, we added a paragraph that clarifies the depth represented by DOISST. We also added in Table 2 the depth information of the DOISST product. Finally, a clear reference to this depth was added in the Abstract too.

Of course, the differences in observations' representativeness make it difficult to compare in situ with satellite data, since there are no in situ instruments able to routinely measure skin/sub-skin sea surface temperatures. The commonly adopted validation procedure is in fact to use surface drifting buoys due to both their high accuracy and closeness to the sea surface, namely at ~20 cm. Of course, also these observations include a representativeness error when compared to sub-skin SST estimates. This concept has been introduced in the revised text (see lines 171-176).

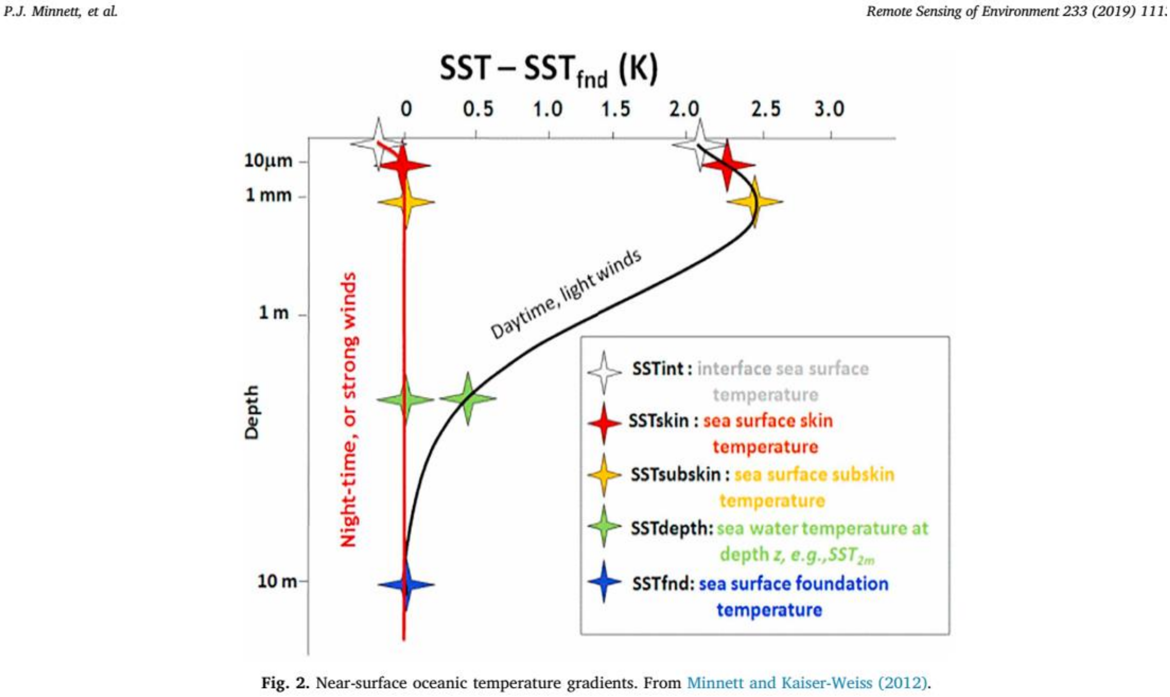

Figure 1. The different types of SST based on the GHRSST definitions (GHRSST - Products).

"The ideal case is that all three components are compared/generated at the same depth level."

We definitely agree with the referee; the ideal case would be if all data were generated and compared at the same depth. Unfortunately, SST data are instead acquired at different depths and with a variety of instruments of different efficiency and precision. As also suggested by the reviewer, 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. However, as already stated in a previous answer, any correction algorithm would introduce additional potential error sources rather than eliminate a (small) bias due to depth differences. That's why we preferred to leave the near surface temperature first-guess data uncorrected, and focus on optimising the corrections driven by available hourly satellite data.

This concept is introduced in the new Section 3.2 and Conclusions (lines 536-542) of the revised manuscript.

"L14, it is not clear why the model analysis is used as the first-guess".

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 thus allows to treat the hourly SEVIRI data as corrections to the hourly model data. These anomalies are generally small and mostly describe 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 much closer to original satellite data, avoiding the complications of setting up an additional preprocessing step just to improve the first-guess.

This concept is entirely reported in the new Section 3.2.

L15-16, it is not clear what is “any diurnal cycle”. “differences between satellite and model SST are free, or nearly free”. If this is the case, then why do we need DOISST analysis?

The anomaly we are looking at is computed as the difference between the satellite derived hourly SST field and the model hourly analysis (see Figure 2 from Marullo et al., 2014). This anomaly is the variable to be interpolated. If the model outputs were representing the same layer sensed by SEVIRI and model physics at the air-sea interface was accurately represented, this anomaly would be free from the diurnal cycle. See also previous answer to the question “*My major concern is how the model SST at 1 m depth is used as the first-guess field*”.

However, we do agree with the reviewer, that sentence is rather unclear. We re-written this sentence as follows (see lines 16-18): “The choice of using a model output as first-guess represents an innovative alternative to the commonly adopted climatologies or previous analyses, providing physically consistent estimates of hourly SSTs in the absence of any observation or in situ measurement”.

S. Marullo et al. / Remote Sensing of Environment 146 (2014) 11–23 15

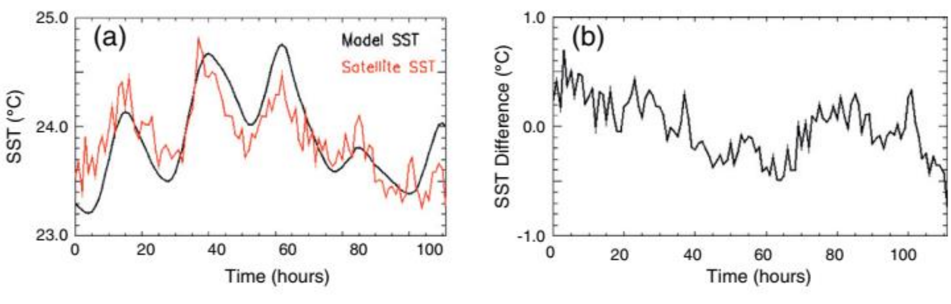

Fig. 2. Example of model first-guess and original SST SEVIRI data over a point in the Mediterranean Sea. (a) Model (black curve) and SEVIRI (red curve) SST time series. (b) Difference between the SEVIRI and model SST.

“L35, Does this mean, the SST analysis will be absent when it is rain?”

Satellite-based SST images are frequently, and usually, affected by several data voids since infrared and microwave sensors cannot “see” under cloudy and rainy conditions. Therefore, many applications require SST data to be processed up to what is generally called Level 4 (L4), namely gap-free fields. The DOISST is an example of such a product, obtained through statistical optimal interpolation technique.

"L71, "slightly less than that" => approximately"

Corrected.

"L155, SST at 1 m level. What depth does the satellite SST in section 2.2 represent? How is the model SST at deeper level used as a first-guess of the satellite SSTs near the skin level? "

The CMEMS Mediterranean daily product provides (nighttime) gap-free maps of foundation SST. By definition, foundation SST is the water temperature at a depth such that the daily variability induced by the solar irradiance is negligible. For this reason, a fixed depth can not be assigned a priori, as it changes continuously. A reference value is given at about 10 m (see GHRSST definitions) since, on average, the diurnal warming is not seen anymore at this depth.

The answer to the last question was given in the answer to the previous question *"The ideal case is that all three components are compared/generated at the same depth level"*.

"L158, how is the buoy SST at 20 cm level used to validate DOISST"

Here, the answer is the same as that given to the previous question: *"how the DOISST is validated with buoy SSTs at 0.2 m depth"*

"L166, delete an extra space".

Corrected.

L169, "between" => among?

Corrected

"Table 1, sub-skin SST, What is the level of sub-skin? please add a depth level".

Added in Table 2 of the revised manuscript.

"L188-190, the statement is not clear, and need to be clarify, particularly "allowing to interpolate SST anomalies using satellite data". Is the "anomaly" referenced to an hourly climatology, how is the climatology is defined?"

The new section 3.2 should hopefully clarify this point. However, as answered to the first question ("*My major concern is how the model SST at 1 m depth is used as the first-guess field*"), the choice of a model output as first-guess allows to use a covariance function which is practically monotonic decreasing, which, in turn, allows to interpolate SST anomalies at different times of the day (specifically, ± 24 hours around the interpolation time). This could not be achieved by using a climatology, since the covariance function would present local maxima and minima, oscillating during time. Then, no climatology is used, neither introduced in the text.

"L201-202, what is the difference between L3C SST and L3C sub-skin SST? How are the SSTs at different levels blended in DOISST"?

There is no difference. Indeed, L3C indicates the processing level (namely, single-sensor collated file) while sub-skin indicates the type of SST. In other words, OSI-SAF routinely processes SEVIRI measurements providing L3C maps of sub-skin SST, which are downloaded by our DOISST processing system and used to produce the DOISST product. This is explained in section 2.1. The blending of hourly L3C data is obtained through optimal interpolation, and detailed in section 3.2.

"L212, f(r,dt)=f(r)(dt) may not be appropriate. Delete "f(r)*(dt)="?"*

Deleted.

"L224, "no first-guess data are used", how is this possible as described in L227-234?"

We agree with the reviewer, this sentence is actually not clear. Overall, the first-guess is always subtracted to observations to create anomalies. However, as stated at line xx, after the interpolation, the first-guess is added back to the optimally interpolated anomalies to get the actual SST value. Then, if at least one observation is present within the spatial and temporal bounding box of the interpolation pixel, first-guess pixel value is corrected. We have rephrased the unclear sentence as follows: "This error ranges between 0-100%, meaning that the error is almost zero when an optimal number of observations is present within the space-time influential radius, while only first-guess data are used (i.e. no observations are found within the search radius) when the error is 100%." (see lines 312-314).

“L240, how is “co-located” defined, interpolated to the in situ location and time or rounded to a certain spatial and time resolution?”

As stated at line 337: “…the validation is carried out on hourly basis, building a matchup database by collecting the closest (in space) SST grid point to the in situ measurement within a symmetric temporal window of 30 minutes with respect to the beginning of each hour”. We substituted “validation” with “co-location” to make this sentence clearer.

“L272-273, How is the uncertainty of RMSD (±number) calculated?”

This was actually stated (now at lines at lines 348-349): “Validation statistics are quantified in terms of mean bias and Root-Mean-Square Difference (RMSD) from matchup temperature differences (namely, SST minus drifter). Each statistical parameter is associated with a 95% confidence interval computed through a bootstrap procedure (Efron 1994)”. For clearness, we added in each caption this reference.

“L295, is it possible the biases result from the first-guess of the model SST at different level?”

It is possible, but very marginally. Indeed, DOISST is 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 (lines 237-242).

“L337, define DWA earlier”

Thanks for notifying, this has been corrected.

“L347-347, will the underestimation of DWA in model affect the performance of DOISST since it is used as the first-guess?”

Overall, the DOISST improves the description of the diurnal cycle, including DWA estimates, with respect to a purely model-derived estimate. As shown in Section 4.2.2, while the model clearly underestimates diurnal amplitudes larger than 1 K, the DOISST is able to correctly reconstruct these amplitudes. This is likely due to two concurrent factors, the high accuracy of SEVIRI SST data and the fact that the Mediterranean area is particularly favorable in terms of clear sky conditions. Then, we could argue that the underestimation of modelled DWAs does not strongly impact the performance of DOISST. This concept was added in the revised manuscript (see line 574-577 in Conclusions).

“L385-386, it may be better to explain the reasons”.

We thank the Reviewer for this comment. Firstly, in our manuscript we documented that, in general, the model outputs tend to underestimate the SST diurnal warming (DW) with respect to the DOISST. Investigating the spatial variability of such underestimations, we expected the model to produce weaker diurnal cycles in the areas where this signal is known to be intense (see e.g. the southern Tyrrhenian and the east Mediterranean, north of Cyprus), in very good agreement with previous results described in Marullo et al. 2016. Roughly speaking, in open ocean contexts, the diurnal cycle is modulated by the relative role of insolation and wind-induced mixing. From figure 8, one can see that the DW tends to be larger in areas sheltered by the strongest wind systems or in correspondence of freshwater/lower salinity discharge areas (Zecchetto & De Biasio 2007, Minnet et al. 2019, Field 2007). In our study, this is visible in the southern Tyrrhenian, the east Mediterranean area north of Cyprus, in correspondence of the Po/Nile rivers deltas and also in the Alboran gyre (i.e. in correspondence of the low salinity Atlantic Water inflow). However, to document the mechanisms behind the DW is out of the scope of the present study, where we mostly present the DOISST production and quality assessment. We thus suggest the readers refer to previously published papers on this topic (duly documented in our manuscript) for further insights on the DW mechanisms (see lines 490-492 of the revised manuscript).

“L422, these depth information should be presented much earlier”

This is just a reminder. The definition of sub-skin and its depth is actually given much earlier, at line 90 and now in the Abstract too.

“L429, In MED area or over the global oceans?”

Those values have been found for validation in the global ocean. However, we also performed an intercomparison between OSTIA diurnal and DOISST in the Mediterranean Area (see revised paper, figures 6-7). Such analysis shows that DOISST yields a more accurate description of the diurnal variability than OSTIA in the Mediterranean Area.