Reviewer 2

GENERAL COMMENTS:

The manuscript presents a reprocessed version of previously published XBT data. The previous version(s) did not have the best available calibrations and/or quality control applied, and this new version apparently does. As such, I find the effort worthwhile - it is always good to have a version of a dataset that can be considered "final".

The manuscript could be made stronger by:

1. Adding an uncertainty estimate against an independent data source (Argo?) that validates the new data to be in better agreement with such reference than the previous version;
2. Adding a use case that shows what can be done with the new version that could not already be done with the old one(s). E.g. can we detect temperature trends now with better confidence (or after fewer years) than before?

These two items should make the points that yes, the new data is better, and yes, it was actually worth the effort. The present manuscript describes the methods in sufficient detail, but does not make these points.

The clarity of the manuscript could be improved by a copy/line editor authorized to make more than just minor language editing. Can the journal provide such services, for a fee if need be?

There are substantial problems with the dataset and the metadata that comes with it. None of these problems are unusual or difficult to correct, but they do need correction. For a manuscript that lays claim to high-quality metadata, the present state of the underlying dataset is not acceptable. Comments below list my findings in detail.

A: We thank the reviewer for the comments and suggestions, especially on the dataset which allowed us to substantially improve it. We just started our data publishing service with ERDDAP and we are still learning. Please see below our detailed manuscript and dataset review reaction, and please also have a look at our answers to reviewer #1. A new REP dataset version is provided at https://doi.org/10.13127/rep_xbt_1999_2019.2

We would like to precise that our objective was to release for the first time the complete dataset with comprehensive documentation of the new processing procedure. The REP dataset provides for the first time the raw profiles with calibration correction and the full metadata information (i.e. probe type, ship speed, launch height). A new automatic Quality Control and a new interpolation procedure have also been applied.

The XBT dataset available from SeaDataNet infrastructure consists of only interpolated profiles without calibration correction applied (Line 22) that have been quality controlled following (Line 26) Manzella et al. (2003, 2007). We performed a REP-SDN comparison to prove to the users how a different data processing (calibration, QC, interpolation) might affect the final interpolated profiles.

The lack of metadata information about the XBT probe type characterizes the main marine data infrastructures since in the past these metadata were not considered crucial for data re-use and integration with other data types. Cowley et al. (2021) report that only 50% of the World Ocean Database contains XBT probe type and manufacturer information, owing to the application of intelligent metadata algorithms to recreate them. The need for such information to reduce the uncertainty in the computation of the Ocean Heat Content indicator has also been widely reported in literature and it has been one of the motivations for this data review. We will revise the Med OHC estimation once the data description paper has been finalized. In fact, the present data description paper is already very long and rich in details that we would like to consolidate before any further data analysis.

We decided to add the uncertainty specification based on the nominal instrument accuracy provided by the manufacturer, in agreement with Atkinson et al. (2014) and Cowley et al. (2021). The depth and temperature uncertainties are equal for all REP XBT profiles being gathered with probe types produced by Sippican, so we inserted them in the file global attributes (please check it here http://oceano.bo.ingv.it/erddap/info/REP_XBT_1999_2019_v2_metadata/index.html) to not make the dataset heavier.

attribute	NC_GLOBAL	depth_uncertainity	String	depth<=230m: 4.6m;depth>230m: 2% (Table 2 from Cowley R et al., 2021 https://doi.org/10.3389/fmars.2021.689695)
attribute	NC_GLOBAL	TEMPET01_uncertainty	String	XBT = 0.10 deg C; XCTD = 0.02 deg C (Table 2 from Cowley R et al., 2021 https://doi.org/10.3389/fmars.2021.689695)

Reseghetti et al. (2018) and Bordone et al. (2020) performed XBT-CTD and XBT-Argo intercomparisons. The XBT profiles used are included in the REP dataset but they passed through a different QC and interpolation procedure that could slightly modify the results. We consider their results on XBT uncertainties estimation valid and we plan to update them soon but without expecting substantial changes. These values are consistent with the uncertainty values that we specified in the new manuscript version (Cowley et al., 2021; Tables 1 and 2).

We checked with the editorial office and every paper, once accepted, receives English language copy editing. However, we tried to improve the paper readability also thanks to the reviewers' suggestions.

SPECIFIC COMMENTS:

Manuscript:

Ll. 25-34: I recommend taking the URLs out of the text and putting them in footnotes. The one that breaks on the end-of-line cannot be used "as is", but requires hand-editing - that too should be corrected.

A: We realize that the pdf conversion of the manuscript did not preserve the correct URL links. We will take care of this with the editorial office during the next reviewing steps. We will also check your suggestion of inserting URLs as footnotes with the editorial office.

I had to read these sentences twice to understand which dataset was the original, and which was the new one that was being described in this article, and why there were three links instead of two. I recommend clarifying by e.g.:

> assigning names "ORIGINAL" and "REPROCESSED" to these, and using these names throughout the manuscript
> removing one of the two links to the reprocessed data (keep the doi one)

A: We apologize if the abstract is not straightforward in describing the two dataset versions:

- *the original dataset (raw data with calibration correction) has never been published before;*
- *the SeaDataNet dataset version (SDN) does not include raw data but only post-processed interpolated ones according to Manzella et al. (2003, 2007). It does not have calibration information and complete metadata description, i.e. no probe type, fall rate coefficients, manufacturer, ship speed, launch height);*
- *We decided to remove the SDN URL from the abstract, as suggested by the reviewer, and leave it in section 5.1;*
- *the reprocessed (REP) dataset has been prepared starting from the raw data and all available information in the operational log sheets.*

We prefer to keep the names REP and SDN datasets and improve their explanation in the abstract.

We removed the third link at our ERDDAP webpage since it is confusing.

L. 30: Bias and RMS difference against what - between the old and the new versions? Is there any evidence that the new dataset is better than the old one, i.e. that bias and RMS against the truth is now smaller?

A: We computed the bias and RMSD between the SDN and the REP versions with the objective not to prove that one version is better than the other but that they are different due to the new Quality Control procedure, the calibration correction and the new interpolation applied. The new interpolation technique (Barker and McDougall, 2020) has been selected because it recreates values closer to the true measured ones rather than the other two methods considered (Section 4.4). The final result is that the REP profiles are different from the SDN ones, especially in the surface layer from June to November when the thermocline settles. We believe this is crucial information to the data users.

If the reviewer intends as the truth, the nearest Argo or CTD profiles, we did not provide this comparison here. This analysis will be included in a next paper. The aim of the present data description paper is to publish for the first time the original dataset with full metadata description, which allows the users to utilize the XBT profiles for their applications and also to test alternative Quality Control procedures. We also provide a new calibrated/ QCed and interpolated data version with complete documentation of each processing step (each QC test applied to each measurement corresponds to an exit value and all test results are then mapped to a quality flag).

Reseghetti et al. (2018) e Bordone et al. (2020), as data providers, used some of the original data that are also contained in the REP dataset, for their comparison with CTDs and Argo profiles, but only here these profiles are shared openly with full metadata information, allowing the actual transparency and replicability of their work. We consider their data intercomparison still valid even if the XBT data have been processed differently. We will update next their results using our REP data version without expecting substantial changes but having the awareness that our QC and interpolation procedure are completely documented.

L. 5: Define acronym ENEA.

A: Done

L. 28: Define acronym SDN.

A: Done

L. 85: Define acronym DAQ.

A: Done

L. 198: Define acronym ZAMAK.

A: Done

L. 207: Define acronym CSIRO properly.

A: Done

L. 211: Define acronym CNR-ISMAR

A: Done

There are inconsistencies between the numbers of profiles reported in table 1 versus what is in the actual dataset. Please correct or clarify:

- Summing up the second-to-last column of table 1, I expect 3917 profiles in the SeaDataNet repository.
- Clicking on the link in the abstract led me to a data download that ultimately gave me 3662 individual files. Is there a reason these numbers do not match?
- Summing up the last column of table 1, I expect 3757 profiles. The downloaded REP dataset seems to contain 3754.

A: We apologize for the confusion with the numbers that we clarify hereafter:

- *Table 1 does not refer to SDN but to the REP dataset*
- *if you go to the SeaDataNet portal, using the saved query at the URL https://cdi.seadatanet.org/search/welcome.php?query=1866&query_code=%7B4E510DE6-CB22-47D5-B221-7275100CAB7F%7D, you get 3661 profiles due to the bounding box selection which cannot filter out precisely the REP dataset*

displayed in Figure 1. Moreover, not all the profiles in the REP dataset have been disseminated through SeaDataNet. We specify in Section 5.1 that the REP vs SDN comparison has been performed on the 3104 matching profiles.

- *We eliminated the second-to-last column of table 1 since it is not used in the manuscript and it is confusing.*
- *The numbers in the last column of table 1 have been modified. We found a bug on cruise_id so three profiles coming from the same cruise were skipped. This issue has been resolved now. New numbers include some profiles that have been re-introduced in the dataset following the reviewer's #1 suggestion.*

Ll. 146-150: Edit this sentence/paragraph for better English language:

A: The text has been modified as suggested by the reviewer and new details have been inserted to answer some of the questions posed earlier:

"Bordone et al. (2020) compared XBT profiles from SOOP activities in the Ligurian and Tyrrhenian Sea with quasi contemporaneous (± 1 day) and co-located (distance smaller than 12 km) Argo profiles. The XBT profiles used by Bordone et al. (2020) are included in the REP dataset but they went through a different QC and interpolation procedure that could slightly modify their results. In the 0-100 m layer, the mean T difference was 0.24 °C (the median 0.09 °C) and the Standard Deviation (SD) was 0.67 °C. Below 100 m depth, the XBT measurements were on average 0.05 °C warmer than the corresponding Argo values (mean and median were almost coincident) and the SD was 0.10°C. This last SD value agrees with the manufacturer specification and the T uncertainty value reported by Cowley et al. (2021), which has been assigned to the REP data. The values estimated by Bordone et al. (2020) for the surface and sub-surface layer (depth < 100 m) are instead affected by both the XBT (4.6 m) and Argo (2.4 dbar) depth uncertainty estimation, meaning that a small variation in depth could correspond to a large variation in temperature especially when the seasonal thermocline develops, so that the comparison with Argo values would not be significant. The specified uncertainties are independent of the systematic error or bias affecting the XBT temperature and depth measurements, that have been corrected in the REP dataset applying the Cheng et al. (2014) correction scheme."

Ll. 140-145: These quoted uncertainties are consistent with each other. State so.

A: We thank the reviewer for this suggestion, the text has been modified accordingly.

Ll. 146-150: Are these uncertainty estimates for XBT data using the dataset presented here, or are these different data? Has such a comparison been made for the data presented here?

A: *The text has been modified and new details have been inserted to answer this question. Please consider also our previous answers.*

L. 311: What does the word "imported" mean? Imported where?
L. 312: What does the word "collection" mean? If there is a specific meaning that only ODV users can understand, please explain.

A: We thank the reviewer for pointing this out, both words "imported" and "collection" are specific terms of ODV functionalities. We deleted the phrase in the revised manuscript to avoid confusion.

Dataset:

The dataset comprises ~3800 ocean temperature profiles (i.e. observations of temperature as a function of depth). This is not a particularly large dataset, and it is therefore reasonable that a user would like to download everything at once. From a quick back-of-the-envelope calculation, I assume that the size of the entire dataset (incl. metadata) should be a few hundred megabytes. However, when I tried to download the entire dataset with the settings below, the server failed with either error message 500 or 502. I assume it ran out of memory when I requested:

> http://oceano.bo.ingv.it/erddap/tabledap/REP_XBT_1999_2019.html
> requesting every variable
> requesting full time period (1999-2019)
> requesting either file type .ncCF or .ncCFMA

A: We thank the reviewer for this important comment. We increased the RAM of the dedicated Virtual Machine and improved its set up.

I then downloaded subsets (final ~6 months) of data, and these files do not make prudent use of memory (100-300 MB for 34 profiles). In particular, data types were unnecessarily large (e.g. floating-point variables when smaller integers would suffice), and there were many cases where information was redundant (e.g. ship names repeated for every data point, rather than once per profile). I recommend making changes that will reduce the total file size to less than ~500 MB, such that a user can "get everything at once". I recommend the following changes to save space (but please use your own good judgment - not all of this might work as intended):

Convert to 8-bit integers (instead of 16): DEPTH_FLAGS_QC, POSITION_SEADATANET_QC, TEMPET01_FLAGS_QC, TIME_SEADATANET_QC. You only ever have values 0-9, why make space for 32000?

A: We made the changes suggested by the reviewer.

Convert to 32-bit floats (instead of 64): depth, DEPTH_INT, TEMPET01, TEMPET01_INT. Single-precision is sufficient for temperature (~millionth of a degree) and depth (~0.1 mm).

A: Done

Convert to 16-bit integer (and apply corrections below) to TEMPET01_TEST_QC

A: Done

The following variables should have the same dimensionality as latitude and longitude (i.e. one per profile; should not be repeated for every data point):

POSITION_SEADATANET_QC, SDN_BOT_DEPTH, SDN_CRUISE, SDN_EDMO_CODE, TIME_SEADATANET_QC, cruise_id, institution,

institution_edmo_code, pi_name, platform_code, platform_name, platform_type, source, wmo_platform_code, url_metadata

A:Done

Consider eliminating the following: DEPTH_INT_SEADATANET_QC and TEMPET01_INT_SEADATANET_QC (the ...INT variable doesn't really need a QC flag, assuming that only "good" input data were used for the interpolation) DEPTH_TEST_QC (you already have DEPTH_FLAGS_QC, one is enough)

A: A fundamental requirement for interoperability is the understanding of data quality on a point by point basis. This is achieved by tagging every measurement with a single-byte encoded label (flag) incorporated as CF ancillary variables. These are linked to the geophysical variable through the "ancillary_variables" attribute in the parent variable set to the name of the ancillary variable.

In TEMPET01_INT_SEADATANET_QC we used a flag for "interpolated_value" to highlight when the profile has been "reconstructed" in correspondence of layers larger than 3m with BAD or PROBABLY BAD (not used in the interpolation) data in the calibrated profile.

DEPTH_TEST_QC includes all the QC test exit values, while in DEPTH_FLAGS_QC the test exit values have been mapped to quality flags. This is done analogously for TEMP01_TEST_QC and TEMP01_FLAGS_QC.

area should be a single global attribute, not a 17*Nobs array (!!!)

A: Done

The profile_id variable should be replaced with a 16-bit integer in ragged array representation (instead of 136 bits of redundant text)

A: Done

The naming of the variables in the dataset can be improved: some names are capitalized, others are not. Can you make all the same, or have a logic which ones are capitalized (e.g. the lat/lon/time coordinates plus depth and temperature)?

A: ERDDAP manages the spatial and temporal features of each dataset in such a way that they have specific names and units. This makes it easier to identify datasets with relevant data, to request spatial and temporal subsets, to make images with maps or time-series, and to save data in geo-referenced file types (e.g., .esriAscii and .kml). In tabledap, a depth variable (if present) always has the name "depth" and the units "m" below sea level. Locations below sea level have positive depth values.

TEMPET01 seems to be the primary scientific variable, but its name is not human-readable. Can this be changed to "TEMPERATURE"?

A: We understand the reviewer's concern but the variable name comes from the adoption of P01 SDN vocabulary, we cannot modify it. Please have a look at https://vocab.seadatanet.org/v_bodc_vocab_v2/browse.asp?order=conceptid&formname=search&screen=0&lib=p01&v0_0=TEMPET01&v1_0=conceptid%2Cpreflabel%2Caltlabel%2Cdefinition%2Cmodified&v2_0=0&v0_4=&v1_4=modified&v2_4=9&v0_5=&v1_5=modified&v2_5=10&x=0&y=0&v1_6=&v2_6=&v1_7=&v2_7=

Some variables seem to copy input data from SeaDataNet. If they are just duplicates (I have not checked if they are), is it really necessary to include these here? If we want to include them, can they at least be named consistently (at present, some start with "SDN_..." while others have "...SEADATANET..." somewhere in the middle)?

A: None data from SeaDataNet are copied in the REP dataset. We adopted SeaDataNet standards and vocabularies that suggest these variables' names.

There is some poor wording in the metadata of the primary temperature data, which ought to be improved. This is about the use of the words, "raw" and "calibrated". In my understanding, TEMPET01 is calibrated data at the original resolution in space/time, and TEMPET01_INT is data with the same calibration but interpolated onto a consistent depth grid. There are no "raw" data in these files. How about:

TEMPET01: long_name = 'Calibrated seawater temperature at original vertical resolution'

TEMPET01_INT: long_name =' Calibrated seawater temperature interpolated on standard depth levels'

The 'comment' attributes under CALIB and TEMPET01 should reflect this wording (i.e. get rid of "raw"). Also, the equations shown in the 'comment' attributes should use the actual variable names.

A: The metadata have been corrected as suggested by the reviewer, please have a look here http://oceano.bo.ingv.it/erddap/info/REP_XBT_1999_2019_v2_metadata/index.html

In the CF conventions, the attribute "standard_name" is the preferred mechanism by which a user (human or computer) finds out which physical quantity is in a variable. Therefore, I recommend that all variables for which such name definitions exist, should use one. In the present version, this is not Done consistently. In particular:

- All temperature variables should have a standard_name attribute set to, "sea_water_temperature"
- The variables DEPTH_FLAGS_QC, DEPTH_INT_SEADATANET_QC, DEPTH_TEST_QC have a wrong standard_name. Should be corrected as per next item.
- All QC variables should have one of two options for standard name: either simply, "quality_flag" or the standard_name of the corresponding data variable, followed by " status_flag", as in, "sea_water_temperature status_flag" or, "depth status_flag"

A: The standard_name is not mandatory in SeaDataNet guidelines but we added it to each QC variable as suggested by the reviewer.

I was expecting the depth and DEPTH_INT variables to have different lengths (and likewise for the matching temperature data and QC flags). You could save some disk space by not zero-padding the (shorter) interpolated ones.

A: The use of an equal TST size for all variables that require it depends on an ERDDAP limitation. ERDDAP reads the various dimensions and associates all the variables that have

only either INSTANCE or TST_D or MAXZ. This explains why DEPTH_INT and "depth" have the same lengths.

TECHNICAL CORRECTIONS:

Manuscript:

L. 35: Switch order of "Interoperable", "Accessible"

A: Done

L. 151: This is not "hard to describe". Bringas and Goni did it, didn't they? Were their findings used, e.g. by correcting the fall rate equation to account for the initial velocity estimated from drop height? If so, where is this documented in the metadata? Or were their reported depth errors used in some sort of error estimate?

A: Bringas and Goni (2015, BG15) carried out an accurate laboratory study in cylindrical tanks filled with water to describe the XBT motion after its impact with the sea surface as a function of the launch height. The BG15 algorithm corrects the depth value calculated by standard FRE when the launch height is different from about 3.0 - 4.0 m, but it is only an approximation of what happens in operational conditions that does not take into account the turbulence produced by the ship's wake, a perturbation that depends on the ship speed, its draft and the distance from the side of the hull. We found in the literature an unpublished communication from Gilson, Roemmich and Johnson (2008) that illustrates what happens when XBTs are launched from ships moving at different speeds, but it does not include any specific description about their behavior in the surface layer.

In our opinion, BG15 is a proper correction when the ship speed is close to zero, which is not the case for the majority of profiles in the REP dataset (Figure R1), so we preferred not to apply it to the REP dataset depth values and further investigate this issue in our next studies.

Figure R1 shows the distribution of launch heights per probe type in the REP dataset, indicating that about 70% of the drops were from platforms at ~ 10-11 m height.

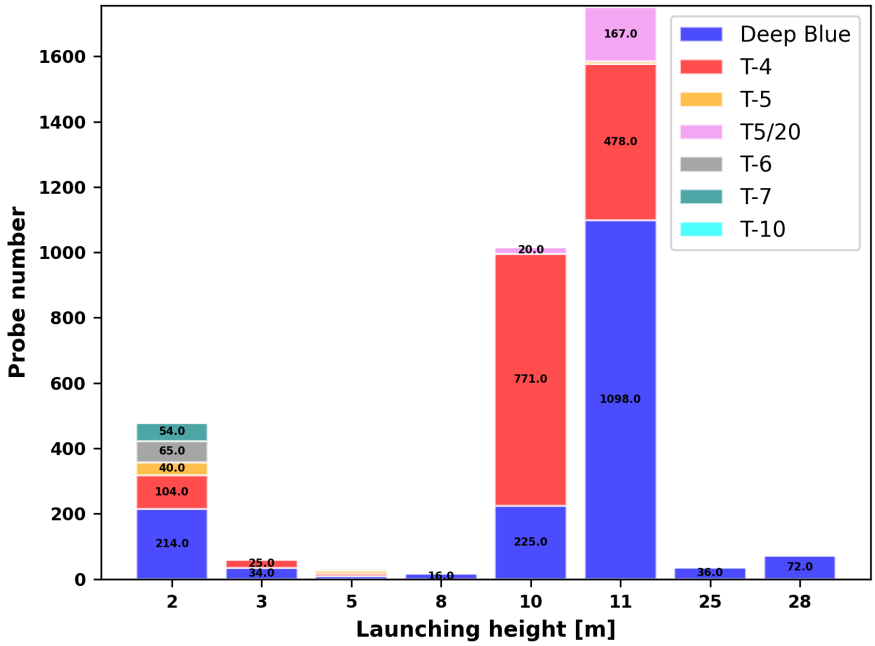

Figure R1 - Distribution of XBT launch heights above the sea level per probe type in the REP dataset.

L. 788: Change "point" to "profile"

A: Done

Dataset:

Something is wrong with the TEMPET01_TEST_QC variable. I assume it should encode, bitwise, the various QC tests. Assuming that there are <=16 tests, the variable type should then be a 16-bit integer (not a 64-bit float). The meaning of each bit should be explained in an attribute "flag_masks", not "flag_meanings", see http://cfconventions.org/cf-conventions/v1.6.0/cf-conventions.html#flags (section 3.5) for the difference between the two, and the values need to be re-computed (maybe I misunderstood, but the present values make no sense to me). In addition, it is unclear how these correspond to the values listed in table 2 of the manuscript, and in my data version, the content of flag_masks and flag_meanings have different lengths (15 vs. 13 entries).

A: The variable TEMPET01_TEST_QC has been written as a 16-bit integer. Adopting the SDN convention, the QC variable has the mandatory attributes:

- *flag_values: a list of all the flag values used in the encoding scheme*
- *flag_meanings: a list of the meanings associated with the codes in flag_values as space-delimited strings with internal spaces replaced by underscores.*

Their different lengths depend on the fact that in the netcdf file the exit values 571, 572, 581, 582 are mentioned explicitly, while in Table 2 appear as 57# and 58#. We specified "(# = 1 or 2)" in Table 2.

The present files have the 'cf_role' attribute assigned to the time variable. Since you actually have a variable "profile_id", I think this variable should have the cf_role attribute instead of time, or am I missing the logic here?

A: The reviewer's suggestion is right, the new dataset has the cf_role assigned to profile_id.

There are metadata entries that are presently global attributes, but they should be variables (or attributes) specific to each profile. These are factually incorrect at present: bathymetric_information, IMO_number, last_good_depth_according_to_operator, last_latitude_observation, last_longitude_observation, launching_height, max_acquisition_depth, max_recorded_depth, probe_manufacturer, probe_serial_number, recorder_types, ship_speed, fall_rate_equation_Coeff_1, fall_rate_equation_Coeff_2

I am unsure if this also applies to the following global attributes (please check): ices_platform_code, id, source_platform_category_code, sourceUrl, wmo_inst_type

A: An url_metadata is associated to the entire dataset: each profile is identified through the corresponding profile_id and cruise_id variable. We inserted in Appendix A an example on how to retrieve these info.

I recommend changing the dataset title (on the website and inside the files) exactly as follows (remove "of", correct "Tyrrhenian", and capitalize "Seas"): Reprocessed XBT dataset in the Ligurian and Tyrrhenian Seas (1999-2019)

A: Done

In the global attribute 'summary', correct spelling of "Expendible" to "Expendable"

A: Done

Reg. the fall rate coefficients:

They are presently given in global attributes, as if one set of coefficients applied to all probes. These change between probes; they have to be profile-specific.
The units are spelled wrong in the global attributes:
There needs to be a space between "m" and "s" (else, it is milliseconds)
The exponents behind "s" need to be negative

A: Done

The "coordinates" attributes are used incorrectly: DEPTH_INT is a coordinate and should not have such an attribute (I think, but correct me if I am wrong)

A: We agree with the reviewer, only geophysical variables must have “coordinates" attribute as mandatory.

TEMPET01_INT (and the other _INT variables except DEPTH_INT) should not list "depth" in the coordinates attribute, but rather "DEPTH_INT". This is important, because it defines the vertical position of the data - the way it is right now, you are telling the user that TEMPET01_INT data are coming from the wrong depth!

A: We agree with the reviewer and we applied this suggestion.