Global ocean synoptic thermocline gradient, isothermal-layer depth, and other upper ocean parameters

Peter C. Chu & Chenwu Fan

Different from the existing global ocean climatological datasets of isothermal layer (ITL) depth \((h)\), a global ocean synoptic dataset of \(h\) along with other parameters has been established from temperature profiles of the National Centers for Environmental Information (NCEI) world ocean database 1961–2017. The exponential leap-forward gradient method was used to identify \(h\) and thermocline gradient \((G)\) from each temperature profile measured by expendable bathythermograph (XBT) and conductivity, temperature, depth (CTD) instruments. Due to quality and vertical resolution of the profiling data, the numbers of \((G, h)\) pairs are 446,811 out of 964,942 CTD profiles and 755,086 out of 2,303,433 XBT profiles. With the given \(h\), other parameters such as ITL heat content \((H_{ITL})\), sea surface temperature \((SST)\), temperature below ITL \((T_m)\), and quality measures \((Q, I)\) indices are provided. Altogether, the dataset contains 1,201,897 temporally and horizontally varying sets of \((G, h, H_{ITL}, SST, T_m, Q\text{-index}, I\text{-index})\). Note that we added 200° to the longitude. This synoptic dataset is located on the NCEI website for public use.

Background & Summary

Existing global datasets of isothermal layer (ITL) depth \((h)\) are all climatologies on regular grids with 1° horizontal resolution and standard World Ocean Atlas (WOA) vertical depths, either computed from 3D gridded temperature climatology\(^1\), \(^2\) or identified from observational temperature profiles and then averaged in 1° or 2° cells\(^3\), \(^4\). Such datasets provide climatological monthly mean gridded values. However, many processes occurring in the ITL cause \(h\) to change on diurnal, seasonal, to interannual time scales\(^5\), \(^6\), \(^7\), which cannot be represented by the existing climatological monthly mean datasets.

Besides, underneath the ITL, there exists a thermocline with a strong gradient \((G)\). Both \(h\) and \(G\) are important parameters in studying climate variability, interaction between atmosphere and oceans, and ocean prediction\(^8\). For example, the thermocline affects the heat exchange between the ITL and a deeper layer, and changes the ITL heat content \((H_{ITL})\), influences the evolution of the sea surface temperature \((SST)\), and in turn the heat/moisture fluxes and upward long wave radiation at the surface\(^9\). Thus, establishment of a synoptic dataset of \((h, G, H_{ITL}, SST)\) from temperature profiles becomes urgent.

The traditional approach to determine \(h\) from temperature profiles is because of a single near-zero gradient in the ITL using difference and gradient methods. The difference method requires the deviation of temperature from its value at a reference depth \((z_{ref})\) to be smaller than a certain fixed value, which varies among 1.0 °C\(^1\), \(^10\), 0.8 °C\(^1\), \(^11\), \(^12\), 0.2 °C\(^1\), \(^13\), \(^14\), to 0.1 °C\(^1\), \(^15\). The reference level changes from near the surface\(^11\) to 10 m depth\(^4\), \(^15\). The gradient method requires the vertical temperature gradient \(\partial T/\partial z\) to be smaller than a certain fixed value, which varies from 0.015 °C/m\(^1\), \(^16\) to 0.025 °C/m\(^1\), \(^17\), \(^18\). A recent approach is to use the transition from the near-zero gradient in the ITL to the non-zero gradient in the thermocline to determine \(h\). This leads to a maximum curvature method\(^19\), \(^20\), \(^21\). Large errors may occur in noisy profile data since the curvature involves the calculation of second derivative versus depth\(^20\), \(^22\). To improve the curvature method, the optimal linear fitting\(^23\) and maximum angle\(^24\) methods were developed with capability of handling noisy profile data. However, these two methods are iterative and not as straightforward as earlier methods such as the difference, gradient, and maximum curvature methods. Furthermore, none of these methods determines \(G\).

Naval Ocean Analysis and Prediction Laboratory, Department of Oceanography, Naval Postgraduate School, Monterey, CA, 93943, USA. Correspondence and requests for materials should be addressed to P.C.C. (email: pchu@nps.edu)
Recently, the exponential leap-forward gradient (ELG)\textsuperscript{25} was developed to archive \((G, h, \text{SST})\) and temperature below the ITL \((T_m)\) simultaneously from a temperature profile, and has been verified as optimal among all the existing methods with the highest skill score using the \(Q\)-index\textsuperscript{22} and the lowest Shannon information entropy (representing the least uncertainty)\textsuperscript{25}. Vertical integration of temperature profile from the surface \((z = 0)\) down to the base of the ITL \((z = -h)\) leads to the ITL heat content \((H_{\text{ITL}})\). The benefit of using \(H_{\text{ITL}}\) rather than heat content within fixed depths such as \(H_{700}\) for the upper 700 m is that \(H_{\text{ITL}}\) is fully determined by the upper ocean mixed layer dynamics and directly interacts with the atmosphere but \(H_{700}\) is not. Dynamics is more complicated for \(H_{700}\) than for \(H_{\text{ITL}}\).

We analysed the global ocean CTD and XBT temperature profiles (1961–2017), downloaded from the NCEI website https://www.nodc.noaa.gov/OC5/WOD/pr_wod.html in May 2018, with the ELG method. After that, the \((G, h)\) data pairs were generated by the ELG method from 446,811 out of 964,942 CTD profiles and 755,086 out of 2,303,433 XBT profiles. The lower number of \((G, h)\) data pairs compared to the observational temperature profiles is due to their quality and vertical resolution. Nine criteria are used for the data filtering such as too few data points between 10 m and 40 m, too few total observational points, maximum depth less than 20 m and so on. Please see MATLAB function (getgradient.m). The NCEI/WOD observational XBT profiles (no correction) were used because the attached algorithm is for analysing observational temperature profiles with real vertical data distribution (not on the standard depths). The quality of the \((G, h)\) data were identified by both \(Q\)-index\textsuperscript{22} and identification index (see the Technical Verification Section). Altogether, the dataset contains 1,201,897 sets of \((G, h, H_{\text{ITL}}, \text{SST}, T_m, Q\text{-index}, I\text{-index})\), and is located at the NCEI website for public use.

**Methods**

**ELG Method.** The ELG method\textsuperscript{25} was used to process observational temperature profile data to obtain \((G, h)\). It contains four steps: (1) estimating the ITL gradient (near-zero), (2) identifying the thermocline gradient \(G\), (3) to computing the vertical gradient at each depth (non-dimensionalized by \(G\)), and (4) to determining \(h\) with a given threshold (or user input) to separate the near-zero gradient layer (i.e., the ITL) and the non-zero gradient layer (i.e., the thermocline). Figure 1 illustrates the procedures of this method.

---

**Fig. 1** Characteristics of the isothermal layer and thermocline. (a) Temperature profile with the illustrated \(z_{(0.1)}\) and \(z_{(0.7)}\). (b) Vertical gradient and the defined \(z_1\). (c) Gradient ratio.
Step 1. Temperature profile data are often noisy, sometimes with unrealistically high vertical gradients near the surface. The reference levels \((z_{\text{ref}})\) of \((0\ m, -3\ m, -10\ m)\) are used for the difference method criteria in determination of \(h\) to reduce such noise. Thus, a layer with depth deeper than \(z_{\text{ref}}\), say 20 m, represents the vertical temperature variation of the upper layer. A depth \((z_1)\) with a minimum gradient

\[
\frac{\Delta T}{\Delta z} = Z_{\text{min}} \geq -20\ m
\]

(1)
is identified within this layer (i.e., \(0 \geq z \geq -20 \text{ m}\)). This minimum gradient at \(z_1\) is the best representation of the ITL gradient (Fig. 1a).

**Step-2.** Let an observational temperature profile starting from \(z_1\) to any depth \(z_k\) be represented by \([T(z_k), k = 1, 2, \ldots, b]\) with \(z_b\) the bottom of the profile. The vertical temperature difference from \(z_1\) to \(z_b\), \(T_d = T(z_1) - T(z_b)\), is the variation of temperature across the ITL, thermocline, and deep layer. Since the vertical gradient is strongest in the thermocline and weakest in the ITL (Fig. 1b,c), it is reasonable to assume that the vertical temperature difference inside the ITL is within 10% of \(T_d\). Since the deep layer is usually not vertically well mixed, the temperature variation is more in the deep layer than in the ITL. Thus, the main part of the thermocline can be roughly identified between the upper depth \([z(0.1)]\) with 10% temperature difference to \(z_1\) comparing to \(T_d\)

\[
T(z_1) - T(z(0.1)) = 0.1T_d, \tag{2a}
\]

and the lower depth \([z(0.7)]\) with 70% temperature difference to \(z_1\) comparing to \(T_d\) (Fig. 1a).

\[
T(z_1) - T(z(0.7)) = 0.7T_d. \tag{2b}
\]
Let \((N_{th} + 1)\) be the number of vertical data points for the closed interval \([z(0.1), z(0.7)]\) with \(T_{0} = T(z(0.1)), T_{N_{th}} = T(z(0.7))\), and \((T_i, i = 0, 1, 2, \ldots, N_{th} - 1)\) in between. The ordinary vertical gradients are calculated from the depth \(z(0.1)\) to each depth through \(z(0.7)\),

\[
G_i = \left| \frac{T(z(0.1)) - T(z_i)}{z_i - z(0.1)} \right|, \quad i = 1, 2, \ldots, N_{th}.
\]

(3)

Here, the gradient data \([G_1, G_2, \ldots, G_{N_{th}}]\) are not normally distributed. Thus, the overall feature [i.e., thermocline gradient \((G)\)] is represented by the median,

\[
G = \text{Median}[G_1, G_2, \ldots, G_{N_{th}}].
\]

(4)
If the computed $G$ is extremely small,

$$G \leq 0.001 \, ^\circ\text{C m}^{-1}$$

the thermocline vanishes. Since the temperature difference between $z_{0.7}$ and $z_b$ is only 30% of $T_d$, and the vertical distance of neighbouring data points increases with depth below $z(0.7)$, the gradient below $z(0.7)$ is also extremely small. The ITL extends to the bottom of the profile $z_b$.

**Step 3.** Let $N_p$ be the number of the vertical data points between $z_i$ and $z_{0.7}$ (inclusive), and let $N = \lfloor \log_{10}(N_p) \rfloor$ with the bracket indicating the integer part of the real number inside. $N$ is much smaller than $N_p$. Starting from $z_i$, the $(N + 1)$ exponential leap-forward gradients (ELGs) are calculated at each depth $z_k$ [between $z_i$ and $z_{0.7}$].

---

**Fig. 6** Histograms showing decadal variation for the thermocline gradient ($G$) in comparison to the climatology. (a) 1961–1970. (b) 1971–1980. (c) 1981–1990. (d) 1991–2000. (e) 2001–2010. (f) 2011–2017. (g) WOA annual 1° × 1°. (h) WOA annual 0.25° × 0.25°.
Where the computation stops if $z_k + Z_T$ is equal to or deeper than $z_b$. It is also noted that $D_n$ is an identifier and not a numerical value. The averaged value among $(N + 1)$ gradients $[D_1 T(z_k), D_1 T(z_k), \ldots, D_N T(z_k)]$ is computed by

$$G_z(z_k) = \frac{\sum_{n=1}^{N} D_n T(z_k)}{N},$$

which represents the gradient effectively at the depth $z_k$ with capability to filter out noises in the gradient calculation.
Step 4. Since \( G(z_k) \approx 0 \) if \( z_k \) in the ITL, \( G(z_k) = G \) if \( z_k \) in the thermocline,

\[
R(z_k) \equiv G(z_k)/G = \begin{cases} 1, & z_k \text{ in the ITL} \\ -1, & z_k \text{ in the thermocline} \end{cases} \quad (8)
\]

A threshold of the gradient ratio, \( R(z_k) \), is needed to determine if \( z_k \) is in the ITL or the thermocline. We make it a user input parameter. Since there is no layer between the ITL and thermocline, and gradient in the thermocline is near 1, this threshold should be reasonably near 1 to be included in the thermocline. Here, 0.8 is suggested, but readers can change it in their practice.

**Data filtering.** The ELG method requires a certain quality of the profile data. A temperature profile is flagged if it has the following features: (1) number of data points \( \leq 2 \) between 10 and 40 m depths, (2) total number of data points \( \leq 5 \), (3) maximum depth \( < 20 \) m (not reaching thermocline), (4) upmost data point deeper than 50 m (no way to identify the ITL), (5) variance above 20 m \( > \) variance below 20 m, (6) range \( (T_{max} - T_{min}) < 1.0 \) °C, (7) too small thermocline gradient \( (<0.001 \text{ °C/m}) \) (no thermocline), (8) big change \( (>5\) °C) between two neighboring data points. Finally, 446,811 out of 964,942 CTD profiles and 755,086 out of 2,303,433 XBT profiles are not flagged (i.e., normal). Thus, 1,201,897 sets of \((G, h, H_{ITL}, SST, T_{wo}, Q\text{-index}, I\text{-index})\) have been established. Note that SST here is the temperature at the ITL base \((z = -h)\). The bulk SST is easily obtained from \(H_{ITL}\) and \(h\) or from the first depth level of the profile.

**Data Traced Back to the WOD Profiles.** The dataset contains 1,201,897 sets of \((G, h, H_{ITL}, SST, T_{wo}, Q\text{-index}, I\text{-index})\). Each set has exactly the same metadata as the original temperature profile in the WOD such as time, location (longitude, latitude), country code, WOD cruise identifier, and WOD unique cast. Note that we added 200° to the WOD latitudes. With these parameters, each set of \((G, h, H_{ITL}, SST, T_{wo}, Q\text{-index}, I\text{-index})\) is easy to trace back to the original WOD temperature profile.

**Use of WOD CTD (2013-14) for Illustration.** The WOD dataset (1961–2017) contains 964,942 CTD and 2,303,433 XBT profiles with one file for one profile. We uploaded the WOD CTD temperature data (2013–2014) with 55,554 profiles in the folder named ‘wod20132014CTD’ and the derived dataset of \((G, h, H_{ITL}, SST, T_{wo}, Q\text{-index}, I\text{-index})\) named ‘WODCTD1314ELGtemp.nc’ to the Naval Postgraduate School website http://faculty.nps.edu/pcchu/ocean_data.htm (Item 12) for readers to practise. The following MATLAB codes are ready to use for this sub-dataset. It is easy to change dataset name to process other profile data.

**Data Records**

This global dataset for ocean synoptic thermocline gradient, isothermal-layer depth, and other upper ocean parameters26 is publicly available at the NOAA/NCEI data repository as a NetCDF file, which includes data citation, dataset identifiers, metadata, and ordering instructions. The dataset is located at the NCEI website (https://doi.org/10.25921/dgak-7a43) for public use.

**Technical Validation**

The key parameter of this dataset is \( h \). The error at the data point is defined by the difference between the fitted and observed temperatures. The whole temperature profile data fitted to a single linear function (thick lines in Fig. 2) represents the maximum error since it disregards the existence of ITL and thermocline. Such a maximum error is represented by the total sum of the square error \((\text{SSE}_T)\). After the ITL depth is identified, two lines are used to fit the observational data to get the fitted profile with the first one (near zero gradient in the ITL) from the top to the ITL base (circles in Fig. 2) and the second one (non-zero gradient in the thermocline) from the ITL base to the bottom of the profile. Such fitting has the errors in the ITL represented by the sum of the square error in ITL \((\text{SSE}_{ITL})\) and in the thermocline represented by the sum of the square error in thermocline \((\text{SSE}_{TH})\). The identification index (called \( I\text{-index} \)) is defined by

\[
I_{ITL} = \sqrt{1 - (\text{SSE}_{ITL} + \text{SSE}_{TH})/\text{SSE}_T}. \quad (9)
\]

If an ITL exists, but is not identified \((h = 0)\) (Fig. 2a), \(\text{SSE}_{ITL} = 0, \text{SSE}_{TH} = \text{SSE}_T\) which gives \( I_{ITL} = 0 \). If the identified \(h\) is shorter than the real one (Fig. 2b), \(\text{SSE}_{ITL} > 0, \text{SSE}_{TH} < \text{SSE}_T\) which leads to \( 0 < I_{ITL} < 1 \). If the

| Year       | Mean (°C/m) | Standard Deviation (°C/m) | Skewness | Kurtosis |
|------------|-------------|---------------------------|----------|----------|
| 1961–1970  | 0.244       | 0.178                     | 1.23     | 4.40     |
| 1971–1980  | 0.199       | 0.162                     | 1.39     | 5.15     |
| 1981–1990  | 0.162       | 0.150                     | 1.77     | 6.61     |
| 1991–2000  | 0.130       | 0.133                     | 2.29     | 9.47     |
| 2001–2010  | 0.124       | 0.132                     | 2.35     | 9.69     |
| 2011–2017  | 0.121       | 0.134                     | 2.55     | 10.90    |
| WOA (1°)   | 0.034       | 0.022                     | 0.89     | 3.90     |
| WOA (0.25°)| 0.034       | 0.023                     | 1.04     | 11.10    |

Table 2. Decadal variation of statistical characteristics of the global thermocline gradient (\(G\)) in comparison to climatology.
identified \( h \) is the same as the real one (Fig. 2c), \( \text{SSE}_\text{ITL} = 0 \), \( \text{SSE}_\text{ITL} = 0 \), \( \text{SSE}_T > 0 \); which makes \( \text{ITL} = 1 \). If the identified \( h \) is longer than the real one (Fig. 2d), \( \text{SSE}_\text{ITL} > 0 \), \( \text{SSE}_\text{ITL} = 0 \), \( \text{SSE}_T < \text{SSE}_T \); which leads to \( 0 < \text{ITL} < 1 \). If the identified \( h \) reaches the bottom of the thermocline [i.e., at \( z_{0.7} \)] (Fig. 2e), \( \text{SSE}_\text{ITL} = 0 \), \( \text{SSE}_\text{ITL} = \text{SSE}_T \); which gives \( \text{ITL} = 0 \). The histograms of \( \text{ITL} \) show high quality of the ELG method for identifying ITL depth using the WOD/XBT data with the mean \( \text{ITL} = 0.884 \) (Fig. 3a), the WOD/CTD data with the mean \( \text{ITL} = 0.843 \) (Fig. 3b). However, the quality is lower with the climatology dataset such as World Ocean Atlas (WOA) 2013, downloaded from the website: https://www.nodc.noaa.gov/OC5/woa13/woa13data.html. The mean \( \text{ITL} = 0.788 \) using the WOA-13 on 1° resolution (WOA13-1°) (Fig. 3c) and 0.755 using the WOA-13 on 0.25° resolution (WOA13-0.25°) (Fig. 3d). The low scores may be caused by the coarser vertical resolution in WOA dataset than in the WOD/CTD and WOD/XBT. Besides, the ELG method has the highest score of the commonly used Q-index.22,23 Quite a few zero values for \( \text{ITL} \) in Fig. 3 indicate that no ITL or thermocline can be identified from those profiles.

Figure 4 shows the spatial and temporal distributions of the derived 446,811 sets (from CTD) and 755,086 sets (from XBT) of \((G, h, \text{ITL}, \text{SST}, T_m, \text{Q-index}, \text{I-index})\). The longitude range of this dataset has the same range as the WOD from \(-180^\circ\) to \(180^\circ\). The values of the horizontal axis in Fig. 4a,b had 200° added for the sake of plotting. The dataset covers the global oceans pretty well except the Southern Ocean, where there are data-poor regions. The WOD XBT (Fig. 4c) and CTD (Fig. 4d) temperature profile data increase drastically after 1970.

Differently from the existing climatological datasets of \( h \), this dataset provides variabilities on various time scales. Here, we take decadal variability as an example for illustration. To do so, we divided the time from 1961 to 2017 into six periods: 1961–1970, 1971–1980, 1981–1990, 1991–2000, 2001–2010, and 2011–2017. There is no clear pattern for skewness and kurtosis for \( \text{ITL} \) during 1961–1970 to 2011–2017. All the statistical parameters are comparable between the synoptic and climatological datasets except the mean and standard deviation of \( \text{ITL} \), which were much lower in the climatology datasets such as \(0.034°C/m, 0.022°C/m\) in the WOA13-1°, and \(0.034°C/m, 0.023°C/m\) in the WOA13-0.25°.

### Code Availability

We present custom codes with the MATLAB to determine \((G, h, \text{ITL})\) from an individual temperature profile (see ‘Supplementary Material’). The main program is ‘ThermolineMLD.m’, which contains four Matlab functions: ‘validata.m’ to filter out bad profiles, ‘getgradient.m’ to calculate the vertical gradient between \( z(0.1) \) and \( z(0.7) \), ‘ELGCore.m’ to calculate the ELGs from \( z \) to \( z_{0.7} \), and ‘Iindex.m’ to calculate the Identification index \((I_{\text{ITL}})\) for the technical validation (see the Technical Validation Section). The code ‘getgradient.m’ can use temperature or potential density profiles to obtain the ITL and thermocline gradient or mixed layer and pycnocline gradient, but only the work with temperature profiles will be explained in this data descriptor. Interested readers may use our MATLAB codes to analyse the WOD 2013–14 CTD temperature profiles and to get the derived dataset \((G, h, \text{SST}, T_m, \text{ITL}, \text{Q-Index}, \text{I-index})\).

### References

1. Obata, A., Ishizaka, J. & Endoh, M. Global verification of critical depth theory for phytoplankton bloom with climatological in situ temperature and satellite ocean color data. J. Geophys. Res. 101, 20657–20667 (1996).
2. Monterey, G. & Levitus, S. Seasonal Variability of Mixed Layer Depth for the World Ocean. Report No. NOAA Atlas NESDIS-14, 1–102 (NOAA, 1997).

| Year Range | Mean (10^9 J/m²) | Standard Deviation (10^9 J/m²) | Skewness | Kurtosis |
|------------|------------------|---------------------------------|----------|----------|
| 1961–1970  | 27.9             | 25.8                            | 2.29     | 9.48     |
| 1971–1980  | 28.9             | 31.7                            | 2.66     | 13.90    |
| 1981–1990  | 44.3             | 43.0                            | 1.49     | 5.95     |
| 1991–2000  | 49.5             | 44.1                            | 1.82     | 13.10    |
| 2001–2010  | 52.0             | 47.5                            | 1.95     | 15.30    |
| 2011–2017  | 54.3             | 46.3                            | 1.38     | 7.28     |
| WOA (1°)   | 30.6             | 30.3                            | 1.41     | 4.43     |
| WOA (0.25°)| 31.1             | 31.2                            | 1.36     | 4.27     |

Table 3. Decadal variation of statistical characteristics of the global isothermal layer heat content \((\text{HITL})\) in comparison to climatology.
Acknowledgements
Outstanding efforts of Dr. Alexandra Grodsky to publish the dataset at the NOAA/NCEI are highly appreciated.

Author Contributions
P.C.C. developed the method, designed the project, conducted the data quality control, and prepared the Data Descriptor. C.W.F. developed the code, and helped prepare the Data Descriptor.

Additional Information
Supplementary information is available for this paper at https://doi.org/10.1038/s41597-019-0125-3.

Competing Interests: The authors declare no competing interests.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

The Creative Commons Public Domain Dedication waiver http://creativecommons.org/publicdomain/zero/1.0/ applies to the metadata files associated with this article.

This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2019