

Sensitivity Analysis of Bias in Satellite Sea Surface Temperature Measurements

Mike Eichhorn* Yuri A.W. Shardt* Joseph Gradone**
Ben Allsup**

* *Department of Automation Engineering, Technical University of Ilmenau, Helmholtzplatz 5, 98693 Ilmenau, Germany (e-mail: (mike.eichhorn, yuri.shardt)@tu-ilmenau.de).*

** *Teledyne Webb Research, 49 Edgerton Drive, North Falmouth, MA 02556, USA (www.webbresearch.com).*

Abstract: The satellite sea surface temperature (SST) measurement is based on the detection of ocean radiation using microwave or infrared wavelengths within the electromagnetic spectrum. The radiance of individual wavelengths can be converted into brightness temperatures for using in SST determination. The calibration and validation of the determined SST data require reference measurements from in-situ observations. These in-situ observations are from various platforms such as ships, drifters, floats and mooring buoys and require a high measurement accuracy. This paper presents an investigation about the possibility of using a glider as in-situ platform. A glider is a type of autonomous underwater vehicle (AUV) which can log oceanographic data over a period of up to one year by following predetermined routes. In contrast to buoys, a glider allows a targeted investigation of regional anomalies in SST circulations. To assess the quality of SST observations from a glider, logged data from a glider mission in the Atlantic Ocean from 2018 to 2019 and corresponding satellite SST data were used. The influence of variables (e.g. measurement depth, latitude, view zenith angle, local solar time) of the bias between satellite and glider SST data was investigated using sensitivity analysis. A new and efficient distribution-based method for global sensitivity analyzes, called PAWN, was used successfully. Interested readers will find information about its operation principle and the usage for passive observations where only “given-data” are available.

Keywords: Sensitivity analysis, Satellite applications, Bias analysis, Underwater glider, Sea surface temperature, PAWN method

1. INTRODUCTION

The sea surface temperature (SST) is an important variable in oceanography, meteorology and climatology since they determine both the thermal circulation in the oceans and heat exchange with the Earth’s atmosphere. SST is also an essential indicator for hurricane forecasting. The first detection of SST occurred from ships using a bucket and a thermometer in the late eighteenth century. For nearly four decades, geostationary meteorological satellites have been used for an exhaustive detection of the SST in the oceans.

The investigated SST data in this paper came from the geostationary weather satellite GOES-16. This satellite uses a multi-channel imaging radiometer called Advanced Baseline Imager (ABI). The ABI scans the Earth’s surface with 16 spectral bands, ranging from visible to infrared (Schmit et al. (2018)). The thermal infrared SST measurements have a good resolution and accuracy. On the other hand, the infrared band is sensitive to the presence of clouds and other atmospheric effects. This requires in-situ observations to calibrate the SST data and to fill the gaps in the SST maps in the presence of clouds. An underwater glider could be a possible platform to provide these reference data. In this paper, logged data from a Teledyne

Webb Research Slocum glider were used. Since a glider cannot measure the temperature exactly on the water surface and at any time (only during surfacings) it is necessary to analyze the influence of the glider measurement on the SST bias (defined as difference between satellite and glider SST measurements). Thus, the objective of this paper is to present a sensitivity analysis to determine the influence of chosen variables on the SST bias.

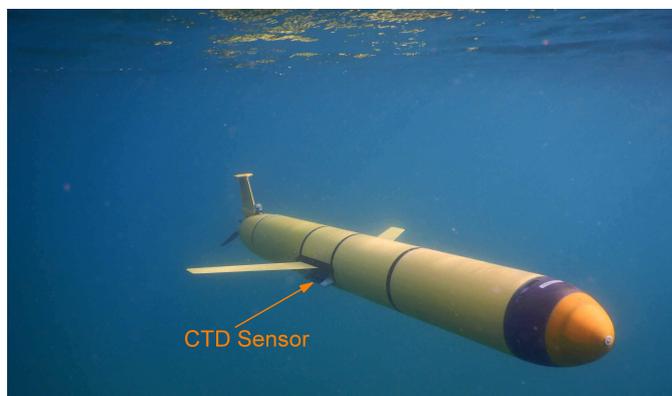


Fig. 1. Underwater picture of the Slocum glider Silbo18 (Teledyne Marine (2019))

2. BACKGROUND

2.1 Slocum glider

The Slocum glider, built by Teledyne Webb Research (Teledyne Marine (2020a)), is an Autonomous Underwater Vehicle (AUV) designed as sensor platform to operate up to 1000 m depth for monitoring oceans. Slocum gliders have proven their efficiency and robustness in the collection of oceanographic data in the last two decades (Schofield et al. (2007); Glenn et al. (2011); Schofield et al. (2018); Merckelbach et al. (2019)). The general specifications of a Slocum glider are shown in Table 1.

Table 1. General Specifications of Slocum glider (Teledyne Marine (2020b))

Parameter	Value
Power	Alkaline/Rechargeable/Lithium
Range	350-1200 km/700-3000 km/3000-13000 km
Endurance	15-50 days/1-4 months/4-18 months
Depth Range	(4 to 150 m) or (40 to 1000 m)
Navigation	GPS, Altimeter, Dead Reckoning
Communication	RF Modem, Iridium (RUDICS), ARGOS
Horizontal Speed	Buoyancy Engine: 0.35 Cruise, up to 0.5 m/s Thruster: up to 1.0 m/s
Mass	55 - 70 kg
Dimensions	Length: 1.5 m; Hull Diameter: 0.22 m

The glider which was used for the analysis in this paper is equipped with a conductivity-temperature-depth (CTD) sensor from Sea-Bird Scientific (Sea-Bird Scientific (2020)) called Slocum Glider CTD. The Slocum Glider Payload CTD is a low-power profiling instrument with high accuracy designed for integration into Slocum gliders. The sensor is fitted under the wing on the starboard-side of the glider (see Fig. 1). The CTD samples continuously every 2 s (0.5 Hz) the temperature with an accuracy of ± 0.002 °C in calibration range of 1-32 °C. During the surface periods the glider can use the on-board GPS for referencing the logged sensor data to its position. The reference system (map projection) for GPS is WGS84 which corresponds to the map projection used in GOES-16.

2.2 Data sources

The satellite SST data came from the satellite GOES-16 operated by NASA and the National Oceanic and Atmospheric Administration (NOAA). The required data for the region of interest with latitude of 14.5°N-28.25°N and longitude of 65.25°W-14.75°W from May 2018 to July 2019 were downloaded from NASA Physical Oceanography Distributed Active Archive Center (PO.DAAC) (NOAA Center For Satellite Applications And Research (STAR) (2019)) from the THREDDS data server as netCDF-4 files. In addition to the SST data, the quality level was also downloaded. The sum of all satellite data is 53.2 GB. The spatial resolution of the datasets is 0.02° (latitude) \times 0.02° (longitude) (equivalent to 2 km \times 2 km at nadir) with a temporal resolution of 1 hour which results in a $2,526 \times 688 \times 10,965$ matrix. The datasets have the level L3C. This means that the data have the processing Level 3 (L3) and includes collated data from a single sensor (C) mapped on a defined space-time grid. This dataset is a gridded version of a L2P dataset. The preprocessed (P) L2P dataset is derived from the brightness temperatures of the ABI infrared channels using the NOAA Advanced Clear-Sky Processor for Ocean (ACSPO) (Petrenko et al. (2010)) and a non-linear SST regression algorithm (Petrenko et al. (2014)). L2 or L3 datasets do not use additional in-situ observations in interpolation algorithms to fill gaps where no satellite measurements are available. This is carried out for a L4 dataset to generate gap free maps. Detailed information about the individual processing levels can be found in (GHRSSST Science Team (2012)). A SST field from the used data with gaps where clouds or atmospheric effects impeded an observation is shown in Fig. 2.

In-situ temperature data from the Silbo18-542 mission was used to evaluate satellite SST data. This mission is part of the Challenger Glider Mission (Dobson et al. (2013); RUCOOL (2020)) and was executed from the Canary Islands to the Virgin Islands from 2018-05-25 to 2019-07-16. This was the longest successfully completed mission in which a Slocum glider covered a distance of 6,359 km during 417 days at sea. Fig. 2 shows Silbo's track from east to west in the southern part of the North Atlantic Gyre.

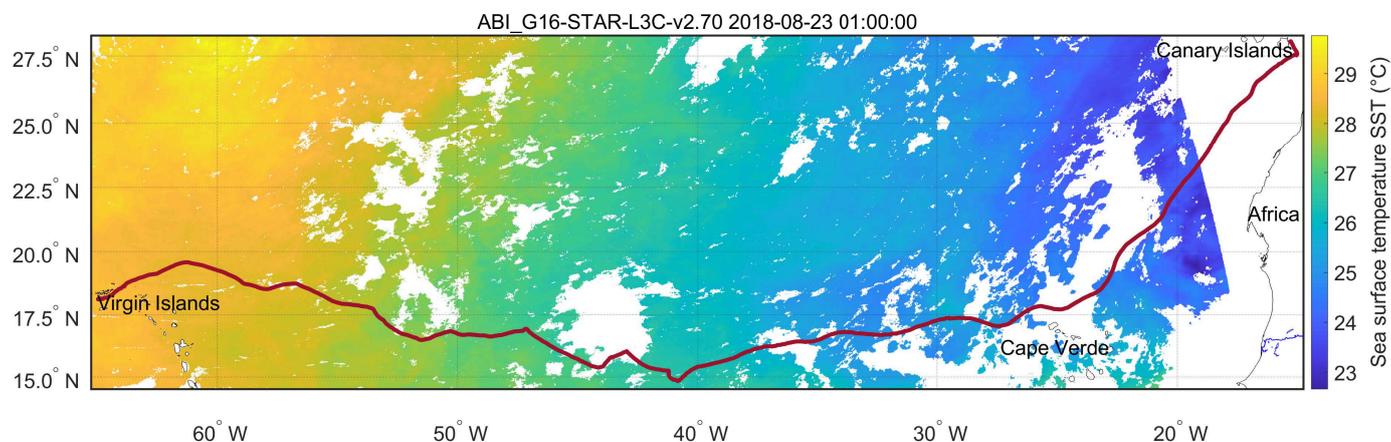


Fig. 2. Silbo's track (bordeaux red line) from the Canary Islands to the Virgin Islands from 2018-05-25 to 2019-07-16 and a L3-SST field from the 2018-08-23 at 01:00:00 UTC. This SST field is from GHRSSST NOAA/STAR GOES-16 ABI (NOAA Center For Satellite Applications And Research (STAR) (2019)).

During the 14 month mission, the glider completed roughly 1,000 segments (dives followed by a surfacing) and logged 6,545,864 CTD measurements. For most of the mission the glider was programmed to fly two yos, starting from the surface and diving to 980 m, climbing to 25 m, diving again to 980 m and returning to the surface. During the surface period of normally 20 min the glider sends the logged data via satellite to the Dockserver, the glider communication center, and receives new instructions for command and control for the next dive period. A comparison of the glider data with the satellite data was only possible after June 2018 where the glider crossed the coverage area of GOES-16 near the Tropic of Cancer (see Fig. 2 for details).

2.3 Data preparation

Only satellite data with the quality level 5 were used for the analysis. This is the “best” quality level and corresponds to “clear-sky” pixels. A space-time interpolation has been carried out to determine satellite SST data at the same position and time where the glider was surfacing. To avoid a distortion of the interpolated satellite data on the one hand and to extract as many samples as possible on the other hand, a combination of a nearest neighbor and a linear interpolation were used. The linear interpolation is used when a sample lies between two valid pixels. Otherwise, nearest neighbor is used. This combination allows the extraction of 20% more samples than by using only linear interpolation without significant other statistical parameters of the bias (mean, standard deviation, median, robust standard deviation, skewness, kurtosis). These statistical parameters are used in NOAA SST Quality Monitor (SQUAM) (Dash et al. (2010); SQUAM (2020)) to monitor the quality of L2-L4 SST products (see also Fig. 7). The resulting error in nearest neighbor interpolation as a result of the time (1 hour) and spatial (2 km × 2 km) resolution is acceptable in the operation area of the North Atlantic Ocean. This can be seen in the curve progression of the measurement temperature of the glider in the upper plot of Fig. 3. Apart from two phenomena (peaks in curve) on the 14th September 2018 (16.75°N 31.6-31.9°W) and on the 23th April 2019 (18°N 55°W) the temperature changes only slightly between neighboring surface periods. For information; the first phenomenon is the result of Hurricane Helene which crossed the track of the glider two days before (Cangialosi, John P. (2018)).

A glider SST value was determined at every surface period using the mean of five temperature samples \overline{SST} where the associated measurement depth was the lowest and less than one meter (see (8)). The mean values for depth, position and time of the five samples were used for analysis and interpolation. The upper plot of Fig. 3 shows the extracted satellite and glider SST data as well as the corresponding measurement depths of the glider. The resulting bias defined as:

$$Bias = SST_{GOES-16} - SST_{Glider} \quad (1)$$

is shown in the bottom plot of Fig. 3. This plot also includes the curves of the mean biases using samples from entire day, daytime and nighttime.

2.4 Data evaluation

In Fig. 3, the evaluation of the bias trend during the mission shows a high negative bias in the first and the end

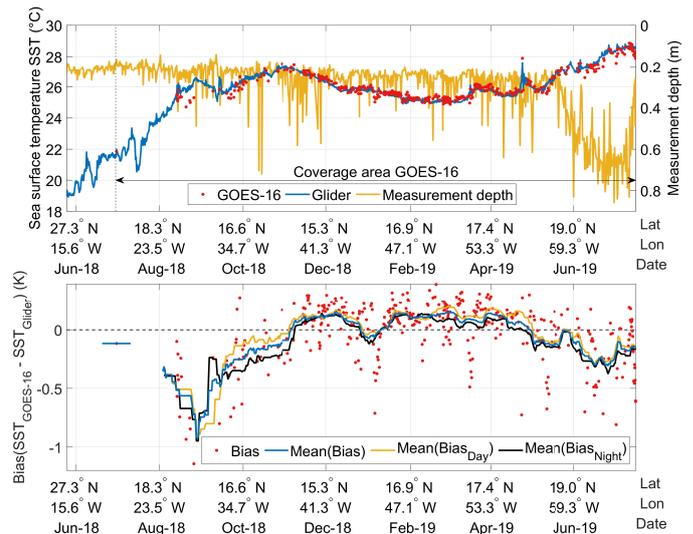


Fig. 3. *Upper plot:* Sea surface temperature (SST) from the satellite GOES-16 and the glider Silbo18 and the corresponding measurement depth for the total mission from May 2018 - July 2019. *Bottom plot:* Bias between the satellite and the glider SST and the curves of the mean biases using samples from entire day, daytime and nighttime determined with a moving mean with a sliding window of 21 days.

part of the mission. This correlates with the summertime period from June till September. The mean bias of the daytime curve lie over and of the nighttime curve under the mean bias curve. This is the result of (i) the diurnal warming effect where solar radiation warms the surface layer of the ocean during the daytime and (ii) the cool skin effect where a heat flux exists from the ocean to the atmosphere throughout the day (see Casey (2002); Zhang et al. (2019)). Another interesting fact is a possible correlation between the measurement depth of the glider and the bias at the end of the mission. This would mean that a glider cannot always accurately measure the SST. In the next sections, it will be analyzed if this is the case. Moreover, additional causes (latitude of the samples, view zenith angle of the satellite sensor) for the resulting bias between the satellite and the glider SST, which are used in SST regression algorithms (Petrenko et al. (2014)) and in the NOAA SST Quality Monitor (SQUAM) (Dash et al. (2010); SQUAM (2020)) will be analyzed. The following variables x_i which are based on the visual inspection of Fig. 3 (Month, LST, Depth) and a literature study about main causes of SST bias (LAT, VZA) are included in the analysis:

- *Latitude* of the samples (LAT)
- *View zenith angle* of the satellite sensor (VZA)
- *Measurement depth* of the glider (Depth)
- *Month* of the sample dates (Month)
- *Local solar time* of the UTC sample time (LST)

The high accuracy of the temperature sensor and its fitting position on the glider (see Section 2.1) excluded a temperature measurement error as a possible cause of SST bias. Appendix A includes the scatter plots, curves of the moving mean and standard deviation and histograms of the used variables.

3. METHODS

This section presents methods to determine how variations in the output of a system can be assigned to the variations of its inputs. This technique is called sensitivity analysis (SA) and is an important tool for development, improvement, validation and evaluation of complex mathematical models in Earth's and environmental science, engineering and economics. In this paper the SA was used to determine and to rank the influencing variables of the SST bias behavior. The presented methods are only a fraction of possible methods. For a detailed overview about SA methods, the reader is referred to (Iooss and Lemaitre (2015); Iooss and Saltelli (2016); Pianosi et al. (2016)).

3.1 Visual inspection

The visual inspection of graphs is the easiest way to extract information of the input-output relation of a system. In case of multiple inputs with high sensitivities to the output, such an approach cannot figure out the individual input-output relations. There exist multiple methods for displaying the data (Shardt (2015)). The art is to find a compromise between a maximal information gathering by an easy data interpretation, and avoiding information overload.

The pairwise comparison of time series was carried out in Section 2.4 (Month - Bias, Measurement depth - Bias) to find possible relations between variables and bias. The samples of output against samples of an examined input can be plotted in a scatter plot to show their relationship. This is shown in Appendix A for the included variables and the bias (middle part) as well as for the satellite and the glider SST samples with colored markers to indicate the corresponding variable value (left part). The calculation of a trend line for the samples in a scatter plot using regression or moving average can help to show a pattern or trend more clearly in case of a high data fluctuation (right part).

3.2 Correlation and regression methods

A possible sensitivity measure is *Pearson's correlation coefficient* r (Trauth (2015)). This coefficient is a measure of the linear correlation between a variable x_i and the output y and is defined as:

$$r_{x_i y} = \frac{\sum_{j=1}^n (x_{ij} - \bar{x}_i)(y_j - \bar{y})}{\sqrt{\sum_{j=1}^n (x_{ij} - \bar{x}_i)^2} \sqrt{\sum_{j=1}^n (y_j - \bar{y})^2}} \quad (2)$$

$$= \frac{\sum_{j=1}^n (x_{ij} - \bar{x}_i)(y_j - \bar{y})}{(N-1)s_{x_i}s_y}$$

where N is the sample size, x_{ij} and y_j are the individual data samples, \bar{x}_i and \bar{y} are the sample means and s_{x_i} and s_y are the standard deviations. The value of $r_{x_i y}$ lies between -1 and 1. Its magnitude corresponds to the sensitivity $S_i = |r_{x_i y}|$. This correlation coefficient can also be determined using a linear regression model $y = a_i + b_i x_i$. The standardized regression coefficient β_i of the linear coefficient b_i is defined as:

$$\beta_i = b_i \frac{s_{x_i}}{s_y} \quad (3)$$

and is exactly equal to *Pearson's correlation coefficient* $r_{x_i y}$. This coefficient is applicable as sensitivity measure for a linear input-output relationship, but not for nonlinear or seasonal relationships.

3.3 PAWN method

For a better understanding of the influence of the variables x_i , especially the measurement depth of the glider, on the bias, a distribution based method for global sensitivity analysis (GSA) is carried out. The method used is based on cumulative distribution functions (CDFs) and is described in detail in Pianosi and Wagener (2015). The key idea is to analyze the differences between the conditional and unconditional CDFs of the output y using the Kolmogorov-Smirnov (KS) statistic:

$$KS(x_i) = \max_y |F_y(y) - F_{y|x_i}(y|x_i)| \quad (4)$$

as a measure of their distances. This distance correlates to the sensitivity of the input x_i to the output y . The unconditional CDF $F_y(y)$ is the result when all inputs vary simultaneously in defined ranges, whereas the conditional CDFs $F_{y|x_i}$ is the result when all inputs vary except x_i which is fixed at a defined value, called conditioning point. The KS statistic provides a curve of sensitivity over all defined conditioning points $\bar{x}_i^{(1)}, \bar{x}_i^{(2)}, \dots, \bar{x}_i^{(n)}$ of x_i . A new density-based sensitivity index, called PAWN was presented in Pianosi and Wagener (2015) where a defined statistic *stat* such as the *maximum*, *mean* or *median* extracts a single value S_i from the KS curve for every input x_i :

$$\hat{S}_i = \text{stat}_{k=1, \dots, n} \hat{KS}(\bar{x}_i^{(k)}) \quad \text{where} \quad (5)$$

$$\hat{KS}(\bar{x}_i^{(k)}) = \max_y |\hat{F}_y(y) - \hat{F}_{y|x_i}(y|x_i = \bar{x}_i^{(k)})|$$

This value varies between 0 and 1 and is applicable as sensitivity measure. A low value of S_i implies a smaller influence of x_i on y . The original method in Pianosi and Wagener (2015) requires a tailored sampling strategy where three parameters (number of unconditional output samples N_u , number of conditional output samples N_c , number of conditioning points n) have to be specified. This requirement limits their application on *active experiments* and excludes applications where existing datasets are available in cases of previous studies (Pianosi and Wagener (2018)) or *passive observations*. This condition exists in this paper, where only logged glider data are available.

An alternative approximation procedure of the PAWN method which can handle generic datasets is presented in (Pianosi and Wagener (2018); SAFE Toolbox (2019)) and is used in this paper. In this procedure each input x_i is split in n equally spaced intervals \mathcal{I}_k between its upper and lower bounds. The samples in these intervals are the conditional samples. The PAWN sensitivity index is so defined as:

$$\hat{S}_i = \text{stat}_{k=1, \dots, n} \hat{KS}(\mathcal{I}_k) \quad \text{where} \quad (6)$$

$$\hat{KS}(\mathcal{I}_k) = \max_y |\hat{F}_y(y) - \hat{F}_{y|x_i}(y|x_i \in \mathcal{I}_k)|$$

The number of the conditional output samples N_c corresponds with the number of samples in the individual

conditioning intervals for each input x_i . In case of a uniformly distributed dataset, this value is approximately equal to N/n for all inputs. The mean value of the number of samples in all intervals for all inputs was used in the MATLAB code in SAFE Toolbox (2019) to set a value for N_u . To identify non-influential parameters in a GSA a *dummy parameter* was used in Pianosi and Wagener (2018). A dummy parameter has no influence on the output y and should have a sensitivity index of zero. The dummy parameter can be detected using two resamples of the empirical unconditional CDF $\hat{F}_y(y)$:

$$\hat{S}_{dummy} = \hat{KS}(x_{dummy}) = \max_y \left| \hat{F}_y^1(y) - \hat{F}_y^2(y) \right| \quad (7)$$

Detailed information about the dummy parameter is given in Pianosi and Wagener (2018). It is important to evaluate the robustness of the determined sensitivity indices and the dummy parameter in case of non-normal distributed datasets or small sample sizes. Bootstrapping is a usual method to solve this task. Thereby the PAWN sensitivity indices in (6) were determined N_b times using different bootstrap resamples. This results in a distribution of N_b PAWN indices $\hat{S}_i^1 \dots \hat{S}_i^{N_b}$ for each variable x_i . The mean and its confidence intervals for each distribution were finally determined and characterize the sensitivity indices.

4. RESULTS

A time period from August 2018 till July 2019 was chosen for the sensitivity analysis. During this period, the glider surfaced 764 times whereby the satellite had 401 possible observations. The Sensitivity Analysis for Everybody (SAFE) Toolbox for MATLAB (Pianosi et al. (2015)) was used to determine the PAWN indices for three datasets. The determined sensitivity indices for the variables calculated with *Pearson's correlation coefficient* (PCC) and the PAWN method are shown in Table 2. Fig. 4 shows the PAWN indices for the three datasets.

Table 2. Ranking of the variables derived from PCC and the PAWN indices for the datasets. The gray level in the fields corresponds with the ranking order.

Variable	N = 401		N = 1 991		N = 21 604	
	PCC	PAWN	PCC	PAWN	PCC	PAWN
VZA	-0.117	0.414	-0.126	0.353	-0.245	0.398
Month	-0.156	0.329	-0.157	0.297	-0.152	0.275
LAT	-0.307	0.236	-0.304	0.202	-0.255	0.175
LST	0.060	0.186	0.066	0.142	0.037	0.135
Depth	-0.184	0.387	-0.174	0.259	-0.059	0.069

The first dataset includes $N = 401$ samples using the mean value of five temperature samples \overline{SST} for the i -th surface period where the associated measurement depth was the lowest and less than one meter:

$$\begin{aligned} [Z^{all}, I^{all}] &= \{z_i < 1.0 \text{ m}\} \quad SST^{all} = \vartheta_i(I^{all}) \\ [Z^{sort}, I^{sort}] &= \text{sort}(Z^{all}) \\ SST^5 &= \left\{ SST^{all} \left(I_{(1)}^{sort} \right), \dots, SST^{all} \left(I_{(5)}^{sort} \right) \right\} \quad (8) \end{aligned}$$

$$\overline{SST} = \frac{1}{|SST^5|} \sum_{j=1}^{|SST^5|} SST_j^5$$

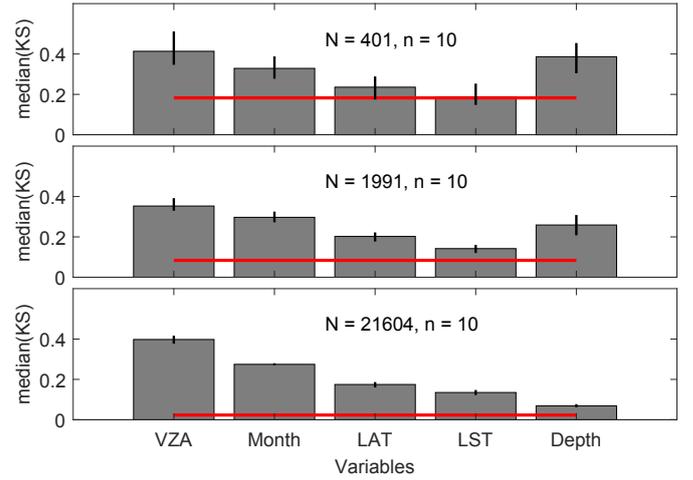


Fig. 4. PAWN sensitivity indices from the three datasets. The black vertical lines show the 95% confidence intervals. The red line shows the KS of the dummy variable. [Experimental setup: stat = median; number of bootstrap resamples $N_b = 50$].

The second dataset includes the five individual samples for each surface period which results in a dataset with $N = 1,991$ samples (SST^5 in (8)). The sensitivity index of the *Depth* shows for the first two datasets a high influence on the bias which confirms the results in Section 2.4. Since the confidence interval (CI) of the *Depth* is overlapped with the CI's of *LAT*, *VZA* and *Month* it is not possible to make a clear ranking. The partial correlation coefficients of these variables and the *Depth* lie also between 0.32 - 0.38 and are statistically significant. This could mean that the high sensitivity index of the *Depth* results not from itself but rather from its correlations. Since the glider logged temperature data from the surface to 980 m it was possible to include additional data at the surface and up to one meter depth in the third dataset with $N = 21,604$ samples (SST^{all} in (8)). The SA for this dataset shows the smallest influence of the *Depth* on the bias. *Pearson's correlation coefficient* can not verify this ranking in detail. The reason is the nonlinear and seasonal (*Month* and *LST*) behavior of the variables (see right part in Appendix A). Fig. 5 shows the result of the PAWN method for the third dataset in detail. It shows clearly that the less influential variables on the bias are *LST* and *Depth* for which the glider is responsible. The absorption of the radiance by the atmospheric water vapor is responsible for the high sensitivities for *VZA*, *Month* and *LAT*. Water vapor content depends on the latitude and the air temperature. The sum of water vapor correlates with the line of sight length between satellite and ocean surface.

Since the number of the samples in the individual conditioning interval and thus the number of conditional output samples N_c for the individual variables can vary widely (e.g. for *Depth* first interval (0.139-0.225 m): $N_c = 402$, second interval (0.225-0.310 m): $N_c = 6,076$, see also histograms in Appendix A), an additional test using the third dataset with $N_u = N_c = 100, 200$ and 400 was carried out. The results in Fig. 6 show the same ranking as the third dataset with the non-uniform distributed samples.

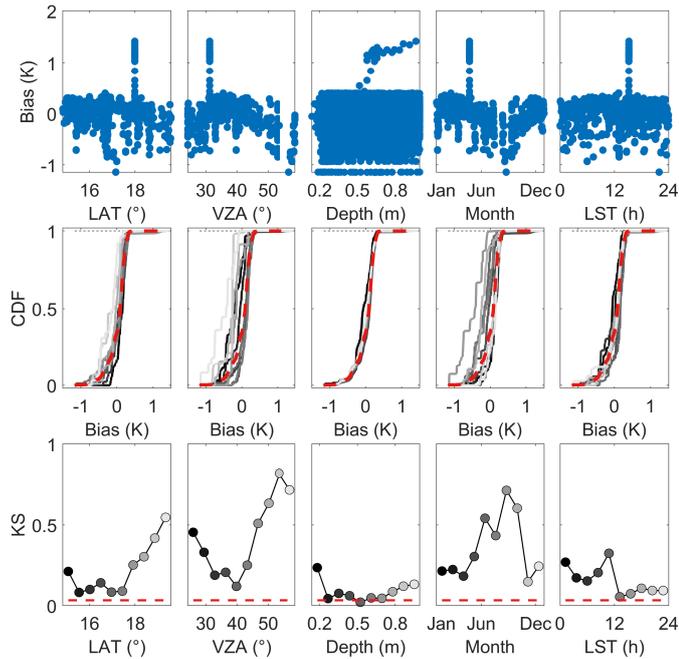


Fig. 5. *Top panels:* Scatter plots of the bias from the third dataset (SST^{all} in (8)) against the corresponding variable. *Middle panels:* Cumulative distribution functions (CDFs) of the bias. The red dashed line is the empirical unconditional distribution function $\hat{F}_{Bias}(\cdot)$ of the bias and the gray lines are the empirical conditional distribution functions $\hat{F}_{Bias|x_i}(\cdot)$. *Bottom panels:* Kolmogorov-Smirnov statistic $\hat{K}S(\mathcal{I}_k)$ at different conditioning intervals \mathcal{I}_k of x_i . The red dashed line shows the KS of the dummy variable.

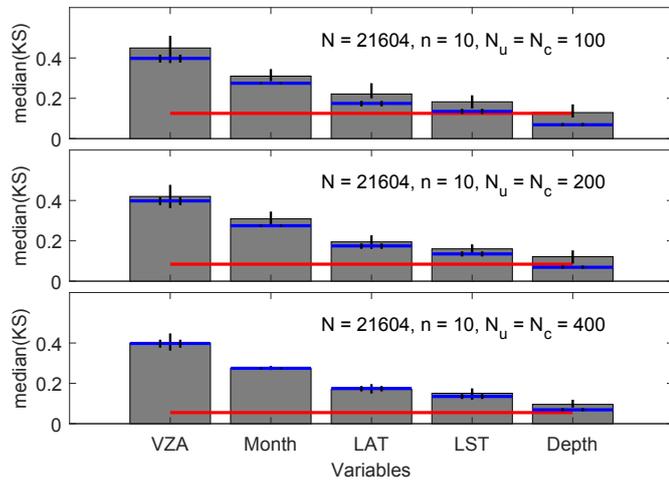


Fig. 6. PAWN sensitivity indices from the third dataset (SST^{all} in (8)) using subsamples of the individual conditioning intervals with a defined sample size. The black vertical lines show the 95% confidence intervals. The red line shows the KS of the dummy variable. The blue lines and the two black vertical lines correspond to the sensitivity indices using all available samples of the individual conditioning intervals (bottom plot in Fig. 4) [Experimental setup: stat = median; number of bootstrap resamples $N_b = 50$; $N_u = N_c = 100, 200$ and 400].

This shows the robustness of the PAWN method when using a non-uniform dataset.

Finally, a global histogram based on the NOAA SST Quality Monitor (SQUAM) (Dash et al. (2010); SQUAM (2020)) was created (see Fig. 7) where all data, daytime and nighttime, of the first dataset are included. The quality of the statistical parameters are compatible with used in-situ platforms (drifter + tropical moorings) in SQUAM.

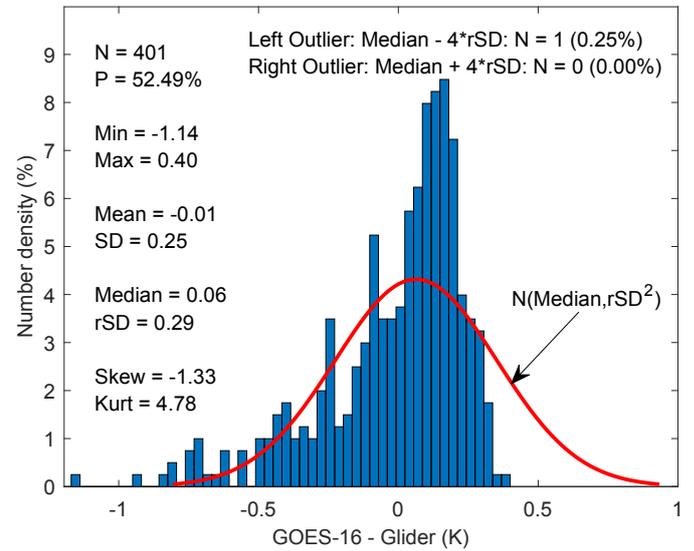


Fig. 7. Global histogram of Bias between the satellite GOES16 and the glider Silbo18.

5. CONCLUSION

This paper analyzed the influence of chosen variables of the bias between satellite and in-situ glider SST data. In practice used basic methods and a simple and efficient method called PAWN were used to solve the required task. The analyzed logged temperature data from a Slocum glider at the surface up to one meter depth shows that the measurement depth does not have a significant influence on the bias, which turns the glider into an additional in-situ platform for calibration and validation of satellite SSTs. It can be targeted to regions to analyze anomalies and phenomena as well as at the margin of the satellite coverage area.

REFERENCES

- Cangialosi, John P. (2018). Hurricane Helene. URL https://www.nhc.noaa.gov/data/tcr/AL082018_Helene.pdf.
- Casey, K.S. (2002). Daytime vs nighttime AVHRR sea surface temperature data: A report regarding Wellington et al. (2001). *Bulletin of Marine Science*, 70(1), 169–175. URL <https://www.ingentaconnect.com/content/umrsmas/bullmar/2002/00000070/00000001/art00013>.
- Dash, P., Ignatov, A., Kihai, Y., and Sapper, J. (2010). The sst quality monitor (squam). *Journal of Atmospheric and Oceanic Technology*, 27(11), 1899–1917. doi: 10.1175/2010JTECH0756.1.

- Dobson, C., Mart, J., Strandkov, N., Kohut, J., Schofield, O., Glenn, S., Jones, C., Webb, D., Barrera, C., and Ramos, A. (2013). The challenger glider mission: A global ocean predictive skill experiment. In *2013 OCEANS - San Diego*, 1–8. doi:10.23919/OCEANS.2013.6741236.
- GHRSSST Science Team (2012). The Recommended GHRSSST Data Specification (GDS) 2.0, document revision 5. URL <https://www.ghrsst.org/wp-content/uploads/2016/10/GDS20r5.pdf>.
- Glenn, S., Schofield, O., Kohut, J., McDonnell, J., Ludescher, R., Seidel, D., Aragon, D., Haskins, T., Handel, E., Haldeman, C., Heifetz, I., Kerfoot, J., Lemus, E., Lichtenwalner, S., Ojanen, L., Roarty, H., Carvalho, F., Lopez, A., Martin, A., Jones, C., Webb, D., Miller, J., Lewis, M., McLean, S., Martins, A., Barrera, C., Ramos, A., and Fanjul, E. (2011). The trans-atlantic slocum glider expeditions: A catalyst for undergraduate participation in ocean science and technology. *Marine Technology Society Journal*, 45(1), 52–67. doi:10.4031/mts.45.1.12. URL <https://doi.org/10.4031/mts.45.1.12>.
- Iooss, B. and Lemaitre, P. (2015). A review on global sensitivity analysis methods. In G. Dellino and C. Meloni (eds.), *Uncertainty Management in Simulation-Optimization of Complex Systems: Algorithms and Applications*, 101–122. Springer US, Boston, MA. doi:10.1007/978-1-4899-7547-8_5.
- Iooss, B. and Saltelli, A. (2016). Introduction to sensitivity analysis. In R. Ghanem, D. Higdon, and H. Owhadi (eds.), *Handbook of Uncertainty Quantification*, 1–20. Springer International Publishing, Cham. doi:10.1007/978-3-319-11259-6_31-1.
- Merckelbach, L., Berger, A., Krahnemann, G., Dengler, M., and Carpenter, J.R. (2019). A Dynamic Flight Model for Slocum Gliders and Implications for Turbulence Microstructure Measurements. *Journal of Atmospheric and Oceanic Technology*, 36(2), 281–296. doi:10.1175/JTECH-D-18-0168.1.
- NOAA Center For Satellite Applications And Research (STAR) (2019). GHRSSST NOAA/STAR GOES-16 ABI L3C Western Atlantic SSTskin v2.70 dataset in GDS2. doi:10.5067/GHG16-3UO27.
- Petrenko, B., Ignatov, A., Kihai, Y., and Heidinger, A. (2010). Clear-Sky Mask for the Advanced Clear-Sky Processor for Oceans. *Journal of Atmospheric and Oceanic Technology*, 27(10), 1609–1623. doi:10.1175/2010JTECHA1413.1.
- Petrenko, B., Ignatov, A., Kihai, Y., Stroup, J., and Dash, P. (2014). Evaluation and selection of SST regression algorithms for JPSS VIIRS. *Journal of Geophysical Research: Atmospheres*, 119(8), 4580–4599. doi:10.1002/2013JD020637.
- Pianosi, F., Beven, K., Freer, J., Hall, J.W., Rougier, J., Stephenson, D.B., and Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, 79, 214 – 232. doi:https://doi.org/10.1016/j.envsoft.2016.02.008.
- Pianosi, F., Sarrazin, F., and Wagener, T. (2015). A matlab toolbox for global sensitivity analysis. *Environmental Modelling & Software*, 70, 80 – 85. doi:10.1016/j.envsoft.2015.04.009.
- Pianosi, F. and Wagener, T. (2015). A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. *Environmental Modelling & Software*, 67, 1 – 11. doi:10.1016/j.envsoft.2015.01.004.
- Pianosi, F. and Wagener, T. (2018). Distribution-based sensitivity analysis for a generic input-output sample. *Environmental Modelling & Software*, 108, 197 – 207. doi:https://doi.org/10.1016/j.envsoft.2018.07.019.
- RUCOOL (2020). The challenger glider mission. URL <https://challenger.marine.rutgers.edu/>.
- SAFE Toolbox (2019). PAWN Method - The MATLAB code to implement the new strategy. URL https://www.safetoolbox.info/wp-content/uploads/2018/08/Pianosi_2018_matlab_code.zip.
- Schmit, T.J., Lindstrom, S.S., Gerth, J.J., and Gunshor, M.M. (2018). Applications of the 16 spectral bands on the Advanced Baseline Imager (ABI). *Journal of Operational Meteorology*, 06(04), 33–46. doi:10.15191/nwajom.2018.0604.
- Schofield, O., Aragon, D., Jones, C., Kohut, J., Roarty, H., Saba, G., Yi, X., and Glenn, S. (2018). Maturing glider technology providing a modular platform capable of mapping ecosystems in the ocean. In E. Delory and J. Pearlman (eds.), *Challenges and Innovations in Ocean In Situ Sensors - Measuring Inner Ocean Processes and Health in the Digital Age*, 173–181. Elsevier. URL <http://www.sciencedirect.com/science/article/pii/B9780128098868120010>.
- Schofield, O., Kohut, J., Aragon, D., Creed, L., Graver, J., Haldeman, C., Kerfoot, J., Roarty, H., Jones, C., Webb, D., and Glenn, S. (2007). Slocum gliders: Robust and ready. *Journal of Field Robotics*, 24(6), 473–485. doi:10.1002/rob.20200.
- Sea-Bird Scientific (2020). Slocum Glider CTD. URL <https://www.seabird.com/auv-rov-sensors/slocum-glider-ctd/family?productCategoryId=54654921153>.
- Shardt, Y.A.W.. (2015). *Statistics for Chemical and Process Engineers : A Modern Approach*. Springer, Berlin, Heidelberg. doi:10.1007/978-3-319-21509-9.
- SQUAM (2020). SST Quality Monitor 2.1. URL <https://www.star.nesdis.noaa.gov/sod/sst/squam/index.php>.
- Teledyne Marine (2019). Slocum glider Silbo18 near Saint Thomas, U.S. Virgin Islands on 2019-07-16.
- Teledyne Marine (2020a). Autonomous Underwater Gliders. URL <http://www.teledynemarine.com/autonomous-underwater-gliders/>.
- Teledyne Marine (2020b). Slocum G3 Glider Brochure. URL http://www.teledynemarine.com/Lists/Downloads/Teledyne%20Webb%20Research%20Brochure%20G3%202017_pages.pdf.
- Trauth, M. (2015). *MATLAB® Recipes for Earth Sciences*. Springer, Berlin, Heidelberg.
- Zhang, H., Beggs, H., Ignatov, A., and Babanin, A.V. (2019). Nighttime Cool Skin Effect Observed from Infrared SST Autonomous Radiometer (ISAR) and Depth Temperatures. *Journal of Atmospheric and Oceanic Technology*. doi:10.1175/JTECH-D-19-0161.1.

Appendix A. INCLUDED VARIABLES IN SENSITIVITY ANALYSIS

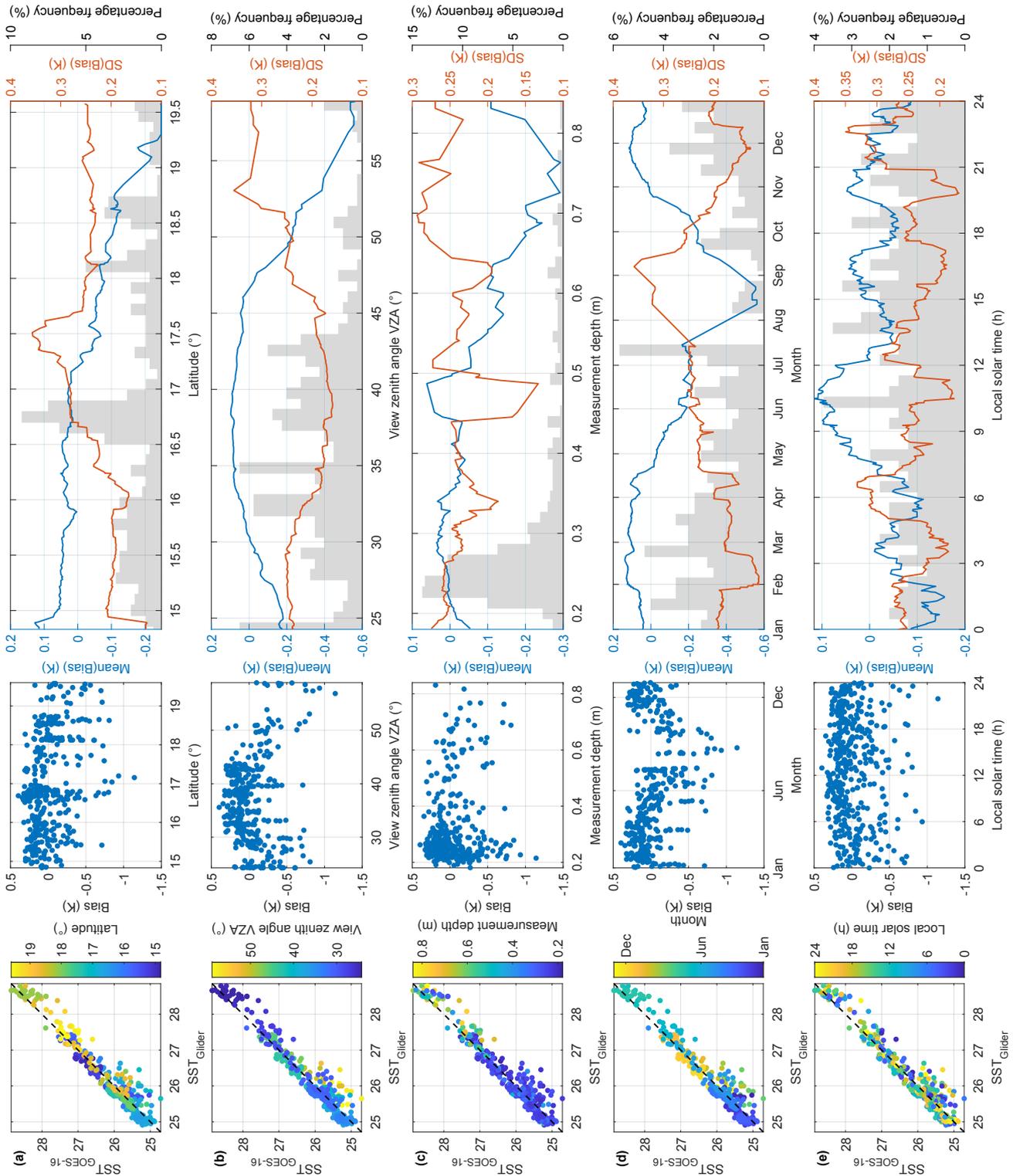


Fig. A.1. Scatter plots, curves and histograms of the used variables (a) Latitude, (b) View zenith angle, (c) Measurement depth, (d) Month and (e) Local solar time from the first dataset (\overline{SST} in (8)). *Left plot:* Scatter plot of the satellite SST samples against the glider SST samples with markers colored proportionally to the corresponding variable. The black dashed line is the 1:1 line. *Middle plot:* Scatter plot of the bias against the corresponding variable. *Right plot:* Moving mean (blue line) and standard deviation (red line) curves by sliding a window with the moving value as the ordinate and the corresponding variable in chronological order as abscissa and a percentage frequency histogram (gray bars) of the samples for the corresponding variable. The total area under the histogram is 100 % which corresponds to the sum of the individual bar frequencies.