Evaluating operational AVHRR sea surface temperature data at the coastline using surfers

Robert J.W. Brewin[*,a,b], Lee de Mora[a], Oliver Billson[c], Thomas Jackson[a], Paul Russell[c], Thomas G. Brewin[d], Jamie D. Shutler[e], Peter I. Miller[a], Benjamin H. Taylor[a], Tim J. Smyth[a], James R. Fishwick[a]

[a]*Plymouth Marine Laboratory, Plymouth, Devon, United Kingdom*
[b]*National Centre for Earth Observation, Plymouth Marine Laboratory, Plymouth, Devon, United Kingdom*
[c]*School of Biological & Marine Sciences, University of Plymouth, Plymouth, Devon, United Kingdom*
[d]*Chatham & Clarendon Grammar School, Ramsgate, Kent, United Kingdom*
[e]*College of Life and Environmental Sciences, University of Exeter, Cornwall Campus, Penryn, Cornwall, United Kingdom*

Abstract

Sea surface temperature (SST) is an essential climate variable that can be measured routinely from Earth Observation (EO) with high temporal and spatial coverage. To evaluate its suitability for an application, it is critical to know the accuracy and precision (performance) of the EO SST data. This requires comparisons with co-located and concomitant *in situ* data. Owing to a relatively large network of *in situ* platforms there is a good understanding of the performance of EO SST data in the open ocean. However, at the coastline this performance is not well known, impeded by a lack of *in situ* data. Here, we used *in situ* SST measurements collected by a group of surfers over a three year period in the coastal waters of the UK and Ireland, to improve our understanding of the performance of EO SST data at the coastline. At two beaches near the city of Plymouth, UK, the *in situ* SST measurements collected by the surfers were compared with *in situ* SST collected from two autonomous buoys located $\sim$7 km and $\sim$33 km from the coastline, and showed good agreement, with discrepancies consistent with the spatial separation of the sites. The *in situ* SST measurements collected by the surfers around the coastline, and those collected offshore by the two autonomous buoys, were used to evaluate the performance of operational Advanced Very High Resolution Radiometer (AVHRR) EO SST data. Results indicate: (i) a significant reduction in the performance of AVHRR at retrieving SST at the coastline, with root mean square errors in the range of 1.0 to 2.0 °C depending on the temporal difference between match-ups, significantly higher than those at the two offshore stations (0.4 to 0.6 °C); (ii) a systematic negative bias in the AVHRR retrievals of approximately 1 °C at the coastline, not observed at the two offshore stations; and (iii) an increase in root mean square error at the coastline when the temporal difference between match-ups exceeded three hours. Harnessing new solutions to improve *in situ* sampling coverage at the coastline, such as tagging surfers with sensors, can improve our understanding of the performance of EO SST data in coastal regions, helping inform users interested in EO SST products for coastal applications. Yet, validating EO SST products using *in situ* SST data at the coastline is challenged by difficulties reconciling the two measurements, which are provided at different spatial scales in a dynamic and complex environment.

Key words: Sea surface temperature, Thermal Radiometry, Remote sensing, Validation, Coastline, Surfers

---

1. Introduction

Sea surface temperature (SST) is considered by the Global Climate Observing System as an essential climate variable (GCOS, 2011; Bojinski et al., 2014). It is a vital property of the aquatic system, controlling its physical (Moore et al., 1999; Nonaka and Xie, 2003), biological (Eppley, 1972; Pepin, 1991; Keller et al., 1999; Lazareth et al., 2003; Doney, 2006; Tittensor et al., 2010; Couce et al., 2012) and chemical (Lee et al., 2006; Kitidis et al., In press) environment. SST impacts the transfer of compounds between the ocean and atmosphere (Land et al., 2013; Takahashi et al., 2002), the distributions and foraging of many marine vertebrates (Frederiksen et al., 2007; Scales et al., 2014; Miller et al., 2015) and the regional and global climate (Sutton and Allen, 1997; Saji et al., 1999; Lea et al., 2000; Bader and Latif, 2003; Yu and Weller, 2007; Raitsos et al., 2011). It is also a variable that can be retrieved routinely, and operationally, with high spatial coverage and good temporal resolution using Earth Observation (EO), through measurements of radiation in the infrared (Llewellyn-Jones et al., 1984) and microwave (Wentz et al., 2000) portion of the electromagnetic spectrum from radiometers mounted on satellite platforms.

To evaluate the use of EO SST products for various operational applications, it is imperative to know the accuracy and precision of the data. This typically requires direct comparison of EO data with co-located and concomitant *in situ* data. In the open-ocean, our understanding of this accuracy and precision is generally high, due to a large network of *in situ* instruments on a variety of plat-

*Corresponding author. Plymouth Marine Laboratory, Plymouth, Devon, United Kingdom
Email address: robr@pml.ac.uk (Robert J.W. Brewin)

forms, resulting in a considerable number of co-incident *in situ* and EO SST measurements distributed over a wide geographical area (e.g. see Table 3 of Merchant et al., 2014). However, despite demonstrative evidence on the value of SST observations for monitoring of coastal seas (e.g. Goreau and Hayes, 1994; Mustard et al., 1999; Paerl and Huisman, 2008; Tang et al., 2003), the economic and ecological importance of coastal waters (Costanza et al., 1997, 2014; Tittensor et al., 2010) and their high sensitivity to human pressures and climate change (Jickells, 1998), the accuracy and precision of EO SST data at the coastline are not well known, impeded by a lack of *in situ* data resulting in few validation studies (Smit et al., 2013). The issue is complicated further by the increased complexities inherent in the retrieval of EO SST data at the coastline, for instance, from land contamination, from the complex coastal aerosol composition impacting the signal received by the satellite sensor (Thomas et al., 2002), from the heterogeneity of SST at the coastline in space and time, and from potential differences in the relationship between the skin temperature (the top 10-20 micrometre) measured by the satellite and the temperature at the depth typically measured *in situ* (hereafter we define SST as the temperature at 1 m depth (z), or SST(z) where z = 1 m, as defined by the Group for High Resolution Sea Surface Temperature, see GHRSST, 2017).

Acquiring *in situ* SST measurements in coastal regions, using conventional platforms such as research vessels, buoys and autonomous vehicles, is notoriously difficult and expensive, hampered by challenges such as: biofouling; vandalisation; wave damage; complex and shallow bathymetry; and strong tidal and coastal currents. This lack of *in situ* SST data at the coastline prohibits EO validation. New solutions are required to improve *in situ* sampling coverage of SST measurements at the coastline, and consequently our understanding of the accuracy and precision of EO SST products.

Building on the work of Brewin et al. (2015b), we present results from a three-year study in which a small group of recreational surfers, based primarily in the south west United Kingdom (UK), were tagged with temperature sensors that they used when surfing to measure SST *in situ* at the coastline. The SST data collected by the surfers, together with SST data collected from two oceanographic stations (L4 and E1, ~7 km and ~33 km from the coastline of Plymouth, UK, respectively) were compared with co-incident and co-located operational 1 km EO SST data from the Advanced Very High Resolution Radiometers (AVHRR), to improve our understanding of the accuracy and precision of EO SST products at the coastline and consequently their use for coastal applications.

2. Methods

2.1. Statistical tests

To compare the estimates of SST from two sources the following univariate statistical tests that are commonly used in comparisons between satellite and *in situ* data were used (e.g. Doney et al., 2009; Brewin et al., 2015c): the coefficient of determination (r^2); the absolute Root Mean Square Error (Ψ); absolute bias between the estimated and measured variable (δ); absolute centre-pattern (or unbiased) Root Mean Square Error (Δ); and the Slope (S) and Intercept (I) of a linear regression between the estimated and measured variables. The equations used to compute each statistic are provided in Appendix A.

2.2. Study Site: United Kingdom and Ireland

The chosen study sites were beaches around the coastline of the United Kingdom (UK) and Ireland (Fig. 1a). Like many coastal regions, the seas surrounding the UK and Ireland are sensitive to increasing human pressure and climate change (Nicholls et al., 2007; Wang et al., 2008), with implications for changes in marine biodiversity and productivity (Frost et al., 2016; Holt et al., 2016), and the monitoring of key environmental indicators such as SST (L'Hévéder et al., 2016). Whereas a few measurements were collected on the west coast of Ireland and south-east coast of the UK (Fig. 1a), the majority of SST data collected by the surfers were from the south-west coastline of the UK (Fig. 1a and b), in particular the coastline surrounding the city of Plymouth (Fig. 1c), which also hosts two oceanographic stations (Station L4 and E1) that form part of the Western Channel Observatory (http://www.westernchannelobservatory.org.uk/) run by Plymouth Marine Laboratory and the UK Marine Biological Association.

2.3. In situ datasets

2.3.1. SST collected by surfers at the coastline

Between the 5$^{\text{th}}$ January 2014 and the 8$^{\text{th}}$ Feburary 2017, five recreational surfers were equipped with a UTBI-001 Tidbit v2 Temperature Data Logger and a Garmin etrex 10 GPS, following methods described in Brewin et al. (2015b, see their Fig. 1). The Garmin GPS device was used to extract information on the location (latitude and longitude) of the surf session. It contains an EGNOS-enabled GPS receiver, has HotFix® satellite prediction and can track both GPS and GLONASS satellites simultaneously. The GPS device was stored in a water-resistant Aquapac inside a waist-bag worn by the surfer (typically under the wet-suit) and set to record GPS data at 1 Hz. The first and last five minutes of the GPS track were removed (approximately the time between switching on (off) the GPS and entering (exiting) the water), and the median latitude and longitude of the remaining data were extracted to derive information on the central location of data collection during the surfing session. In cases where the GPS device failed (e.g. battery depletion) or was not used, the central location (latitude and longitude) of the surf session was extracted immediately proceeding the surf session, using GIS software (https://itouchmap.com/latlong.html).

The Tidbit v2 temperature loggers were attached, using cable-ties, to the mid-point of each surfers leash (tether connecting the surfer to their surfboard) to ensure continuous contact with seawater when surfing, and measured temperature in the top metre of the water column (see Fig. 1 of Brewin et al., 2015b). Manufacturers state that the Tidbit v2 sensors have an accuracy of 0.2°C over a range of 0-50°C, a resolution of ~0.02°C at 25°C, a stability of ~0.1°C per year, a response time of 5 minutes in water, and a battery life of ~5 years at a >1 minute logging interval. To ensure good quality data collection, we monitored the performance of each sensor approximately every 6 months over the study period, by comparing the Tidbit v2 temperature loggers with a VWR1620-200 traceable digital thermometer (NIST/ISO calibrated, with an accuracy of 0.05°C at the range of 0 to 100°C and a resolution of 0.001°C) at 1°C intervals in the laboratory, from 6 to 25°C using a PolyScience temperature bath.

Figure 2a-d illustrates four laboratory comparisons between a Tidbit v2 sensor (10308732) and the VWR1620-200 traceable digital thermometer, and Fig. 2e-j show variations in statistical tests (Eq. A.1 to A.5) for each laboratory comparison, for the five Tidbit v2 sensors used in the study. Over the study period, all sensors performed within the manufacturers technical specifications, with high r^2, slopes (S) staying close to one, and intercepts close to zero for all laboratory comparisons (Fig. 2e, i and j). Root Mean Square Errors (Ψ) were <0.15°C for all sensors (Fig. 2f). When decomposing Ψ into its precision (Δ) and accuracy (δ) components, Ψ was dominated by a small systematic bias (δ) for all sensors (Fig. 2h). We used piecewise regression to model δ as a function of time (Fig. 2h) for each sensor, which was then used to correct any temperature data collected by each sensor. In cases where data were collected before the first laboratory comparison, or after the last, the correction (δ) was set at the closest laboratory comparison (rather than extrapolating the piecewise regression model outside of the time period it was developed for, see Fig. 2h). Having removed the systematic bias, the errors in each sensor were within the accuracy of VWR1620-200 traceable digital thermometer (<0.05°C see Fig. 2g). The piecewise regression model also improved the consistency between sensors, by correcting each sensor to the same common reference (see Appendix B and Fig. A1 for an example of deployment at the same location for two different sensors). Table 1 provides the number of times each sensor was used in a surfing session during the study period, and the duration of use for each sensor.

HOBOware software and HOBO USB Optic Base Station (BASE-U-4) were used by the surfer to launch the Tidbit v2 temperature logger prior to each session, and then to upload data post session. Temperature data were collected at 10 Hz during each surf. Temperature data were processed following a method building on that developed in Brewin et al. (2015a,b). Briefly, the assumption is made that the midpoint of the temperature data for each surf session occurred while the sensor was in the water. This assumption was checked manually for each surf session and found to hold when visually checked with available GPS data. The data were then divided into two equal halves around the mid-point. For the first half of the data, every data point was removed sequentially in time and the standard deviation was calculated incrementally, with the last data point representing the standard deviation of the midpoint (zero). For the second half of the data, this procedure was repeated but in reverse. The standard deviations for the two halves of the data were then recombined. The point at which the surfer began measuring SST (entered the water) was taken as the point when the standard deviation first fell below the bottom third percentile, and the point at which the surfer stopped measuring SST (exited the water) was taken as the last point of the session when the standard deviation was below the bottom third percentile. The bottom third percentile was chosen based on a visual comparison with the timing of the first and last waves caught by the surfer, as estimated from GPS data (see Brewin et al., 2015b). Appendix B illustrates an example of the processing method applied to a surf session at Tolcarne Beach in Newquay, UK (see Fig. A2).

The only difference with this method, to that described in Brewin et al. (2015b), is that a percentile was used rather than determining the start and end points according to when the standard deviation was less than 10 % of the largest standard deviation. We found that using a percentile was slightly more robust in cases where the temperature in the water was very stable, and the previous technique selected data before and after the surfer entered the water. All temperature measurements collected before and after the determined start and end points were excluded, and the median of the remaining data was considered as the SST for each session (see Appendix B, Fig. A2). Note that the median is resistant to outliers and thus fairly resilient to variations in the derived start and finish points. For example, the difference between the processing methods used here and that used by Brewin et al. (2015b) to determine SST was very small ($r^2 = 1.00$, $\Psi = 0.07$, $\Delta = 0.07$, $\delta = -0.02$, $S = 1.00$ and $I = -0.01$).

Appendix B, Fig. A3, shows a superposition of all temperature data acquired by the surfer during the study period, normalised such that the start and end of the surf is at the same point on the x-axis for each session. The plot demonstrates the temperature of the sensor in the sea is relatively stable compared with that before and after each surf. As discussed in Brewin et al. (2015b), the method assumes that the mid-point of the collected data occurred in the sea and that duration of data collection in the sea is longer than duration out of the water. We caution against the use of the method in cases where these assumptions are breached. The method is also designed specifically to determine the median SST of the session. The time of data collection (GMT) was taken as the mid-point (median) of all 10 Hz samples selected to compute SST.

In total, 297 surfing sessions took place during the study period, around the coastline of the United Kingdom (UK) and Ireland (Fig. 1a), most of which were in the south-west UK (Fig. 1b and c). The majority of surf sessions (233) took place at Wembury Beach (latitude = 50.316 °N, longitude = -4.085 °E) and Bovisand Beach (latitude = 50.332 °N, longitude = -4.122 °E) located close to each other and near to the city of Plymouth, UK. The majority of measurements were collected during conditions preferable for surfing. This typically involved breaking waves at the coastline in the range of 0.3-3.0 m, though some measurements were collected in calm sea during surfer paddle training. The SST data collected by the surfers are publicly available through the British Oceanographic Data Centre (Brewin et al., 2017).

2.3.2. SST from station L4 and E1

SST data were also acquired from two oceanographic stations in the Western Channel Observatory (WCO): station L4 (latitude = 50.250 °N, longitude = -4.217 °E) located ~7 km from the coastline and station E1 (latitude = 50.033 °N, longitude = -4.367 °E) located ~33 km from the coastline (Fig. 1c). At both stations an autonomous buoy is operated, equipped with a WET Labs Water Quality Monitor (WQM), which incorporates WET Labs' fluorometer-turbidity and Sea-Bird's CTD sensors, providing temperature, salinity, depth, dissolved oxygen, chlorophyll fluorescence, turbidity and backscattering data. The WQM are mounted on a marine-grade stainless steel cage and situated in a moon pool (an opening in the floatation) at a fixed depth of 1 m. The WQM records SST at hourly intervals, with an accuracy of 0.002°C at a range of -5 to 35 °C, and a resolution of 0.001°C. Further details on the operation of the autonomous buoy systems can be found in Smyth et al. (2010). Quality controlled datasets on SST were downloaded from the Western Channel Observatory website (http://www.westernchannelobservatory.org.uk/data/buoy/) between January 2014 and December 2016, with some gaps in the datasets from buoy maintenance and downtime.

2.4. AVHRR satellite observations

Operational AVHRR SST data were acquired through the UK Natural Environmental Research Council (NERC) Earth Observation Data Acquisition and Analysis Service (NEODAAS, http://www.neodaas.ac.uk/). This service is regularly used by the UK and European scientific communities, and has supported a wide variety of international research (see http://www.neodaas.ac.uk/publications.php). The AVHRR is a scanning sensor on-board the National Oceanic and Atmospheric Administration (NOAA) family of Polar Orbiting Environmental Satellites (POES). These platforms are sun synchronous, viewing the same location roughly twice a day (depending on latitude) due to a relatively wide swath (~2400 km). The AVHRR measures the radiance of the Earth at a suite of bands, including bands centred around 11 and 12 micrometers, measuring emitted thermal radiation. It is these bands that are principally used to derive SST.

The NEODAAS operational processing system is illustrated in Fig. 3. During the 15 minute period when each satellite is in range, a receiving station located in Dundee acquires High Resolution Picture Transmission (HRPT) passes over NW Europe and the Arctic, ~14 per day and ~4.6 of which cover the UK (see http://www.sat.dundee.ac.uk/coverage.html). The passes are immediately transmitted, via a fast internet link, from the receiving station to Plymouth Marine Laboratory for processing. The HRPT images are then processed to Level 3, which involves: georeferencing, using an orbital model together with ephemeris data from NOAA (Sandford and Stephenson, 1992) and an automated navigation adjustment that matches image features with a database of ground control points (Bordes et al., 1992); generation of a land mask using the University of Hawaii's Generic Mapping Tools (http://gmt.soest.hawaii.edu/) which is then overlaid on the georectified AVHRR image; application of a hybrid cloud mask, adapted from Saunders and Kriebel (1988), Thiermann and Ruprecht (1992), and Roozekrans and Prangsma (1988); application of a cloud proximity test to minimise cloud-edge effects and sub-pixel cloud contamination (Miller et al., 1997); implementation of the NEODAAS operational SST algorithm adapted from the standard NOAA method (Non-linear SST (NLSST) split-window equation using infrared channels 4 and 5, with modifications to correct for atmospheric water-vapour absorption; Miller et al., 1997); application of a quality control step by comparison with climatological weekly average Optimum Interpolation SST (OISST) provided by the US National Meteorological Centre (Reynolds and Smith, 1994; Reynolds et al., 2007), flagging any pixels that differ +2°C and −4°C from the climatology; and finally image transformation to Mercator projection (~ 1 km resolution), using the MODIS Swath-to-Grid Toolbox (MS2GT).

Additional details of the NEODAAS operational processing system can be found in Miller et al. (1997). SST images are available within 90 minutes of the start of acquisition.

NEODAAS provides data extractions for various regions. Here we used products provided between -15°E and 13°E and 47°N and 63°N, covering the study area (Fig. 3). Level 3 mapped scenes were acquired from NEODAAS between 5th January 2014 and the 8th February 2017, providing SST, latitude and longitude data for each pixel in the scene, and the time (GMT) of the overpass. In addition to using the individual satellite passes directly for comparison with *in situ* data, we also used daily mean composite products, produced using all the Level 3 passes available during a single day, for a given pixel.

2.5. Comparison of datasets

2.5.1. Comparison of in situ datasets

We first analysed differences in the *in situ* SST over the duration of the study period at three locations near the city of Plymouth in the UK; at Station E1; at Station L4; and at the coastline, using temperature measurements collected from two nearby beaches in Plymouth (Wembury Beach and Bovisand Beach). This was conducted qualitatively, by overlaying the SST time-series of the three datasets onto the same graph which was then inspected visually, and quantitatively, by matching (with a time difference of ≤1hr) co-incident SST measurements and through the application of statistical tests.

2.5.2. *Comparison of daily AVHRR products*

Next we compared daily AVHRR SST products, at the same three locations (Station E1, Station L4, and at the coastline (Wembury Beach and Bovisand Beach)), with the *in situ* data (daily median) over the duration of the study period. At L4 and E1 we extracted AVHRR SST data from a group of nine pixels centred on the location of the oceanographic buoys (see Fig. 6a) for each day in the time-series. At the coastline, we extracted data from six pixels that run along the coastline between the two beaches (see Fig. 6a) for each day in the time-series. For each group of pixels per day, we computed the median SST, the standard deviation and percentage of the group of pixels with SST data. To ensure reasonable homogeneity in the match-up site, required when comparing observations (*in situ* and satellite) representative of vastly different volumes of water, AVHRR data were discarded when the standard deviation of the group of pixels was greater than 1°C and where percentage of pixels with SST data was less than 50%.

As with the comparison of the three *in situ* datasets, we compared the daily AVHRR SST with the *in situ* data at each location qualitatively, by overlaying the satellite and *in situ* SST time-series at each location onto the same graph which was then inspected visually, and quantitatively, by comparing daily match-ups using statistical tests outlined in section 2.1.

2.5.3. *Validation of AVHRR satellite passes*

We matched all *in situ* data (at Station L4, Station E1 and SST measurements collected around the coastline of UK and Ireland by the surfers) to all available

Level 3 AVHRR SST satellite passes, within a time difference of ±12 h. As with the daily AVHHR data for E1 and L4, we extracted a group of nine pixels centred at each location. However, we only used the centre (closest) pixel in the comparison of satellite passes (rather than the median of the nine pixels), to ensure the closest spatial agreement between data. For the *in situ* data at the coastline (collected by the surfers), we used the closest pixel to the *in situ* measurement within a 1 km radius, to account for cases where the closest pixel was dominated principally by land (i.e. the *in situ* measurement was at the edge of a land pixel, see Fig. 4c for an example). As with the daily AVHRR data, the group of nine pixels were used to ensure reasonable homogeneity of the match-up region. Match-ups were discarded when the standard deviation of the group of pixels was greater than 1°C, and where percentage of the group of pixels with SST data was less than 33% (3 pixels needed to compute the standard deviation), which was lower than the daily AVHRR data (<50%), as typically, roughly half of the pixels were located on land when extracting the 9 pixels at the coastline (see Fig. 4c for an example). The absolute time difference (T) between the overpass of the satellite data and the *in situ* was recorded, to investigate the influence of T on statistical tests between datasets. Figure 4 illustrates an example of the match-up process for AVHRR satellite passes, for a relatively cloud free AVHRR SST image taken on the 20th April 2015 at 03:39 GMT (Fig. 4a), compared with SST data collected at Station E1 at 04:04 GMT (Fig. 4b) and by a surfer at Bovisand beach at 05:58 GMT (Fig. 4c).

3. Results

3.1. In situ comparison

Seasonal variations in the three *in situ* time-series are in good agreement visually (Fig. 5b, d and f). The warmest temperatures are observed during late summer and coolest in early March. Inter-annual differences are also generally consistent. For instance, an unusual decrease in SST in August 2014 was seen at both Station L4 and at the beaches, and sharp but brief increases in SST in June and July 2016 are consistent in all three datasets (Fig. 5). Although the L4 and E1 buoys collect data far more regularly (per hour) than the surfers, there are significant periods of time during the study period when one of the buoys were not operating, which was not the case for the surfer data.

Quantitative comparisons among the three time-series (with a time difference of $\leq$1hr) show that the data collected by the surfer explains $\geq$91 % of the variance in the Station L4 and E1 data, with a root mean square difference (Ψ) of 0.74 to 0.84°C (Fig. 5c and e). These statistical results are similar to those found when comparing the two oceanographic buoys (Fig. 5g). Yet, despite these similarities, there are systematic differences seen in the three datasets consistent with their spatial separation (Fig. 5a). Whereas the average bias (δ) between surfer and E1 data is quite low (-0.15°C, Fig. 5e), the autumn and early winter periods show systematically lower SST in the surfer data when compared with E1 (e.g. winter 2014/2015 and autumn 2016, see Fig. 5d). This is likely linked to the influence of the terrestrial environment on nearshore SST during this period. The land cools more rapidly in the autumn and early winter, owing to a lower heat capacity when compared with the ocean, potentially impacting nearshore SST. It may also be influenced by enhanced fresh water input during this period, and by the atmospheric cooling, with increased exchanges of heat between the atmosphere and ocean at the coastline caused by wave breaking. Furthermore, it is possible that enhanced vertical mixing at the coastline due to wave breaking could promote upwelling of colder water during autumn and winter storm conditions.

Both the surfer and the E1 SST data show systematically higher temperatures than that observed at L4 (with an average bias of between 0.33 and 0.40°C, Fig. 5c and g), particularly during the summer of 2015 (Fig. 5b and f). It is likely that Station L4 is less strongly stratified during the summer period when compared with E1, perhaps due to stronger tidal mixing (shallow bathyemetry) and estuarine outflow from Plymouth Sound. Higher SST in the summer of 2015 at the beaches, when compared with L4, may be related to more rapid warming of shallower water at the beaches during the day. Considering good agreement among the three SST datasets, with discrepancies generally consistent with expectations given their spatial separation and contrasting proximity to land, one can be confident using the surfer SST data for coastal applications.

3.2. AVHRR comparison of daily products

Figure 6 shows a comparison of the daily AVHRR SST data with the daily median *in situ* data at L4, E1 and the two beaches (Wembury and Bovisand). With the exception of a few outliers, likely caused from miss-classification of cloud-contaminated pixels (owing to a much lower SST characteristic of cloud-contamination), there is very good agreement between the AVHRR SST data and the *in situ* measurements at L4 and E1, with the satellite observations tracking tightly variations in the *in situ* data (Fig. 6d and f). At both L4 and E1, the AVHRR data explains 97 % of the variance in the *in situ* data, with a very low bias ($\delta = -0.04$°C), low errors (Ψ and Δ, ≤0.44°C), slopes (S) close to one and intercepts (I) close to zero (Fig. 6e and g).

At the coastline, however, the agreement between the AVHRR SST data and *in situ* data is not as good (Fig. 6b and c). The satellite observations do not track the *in situ* data as tightly over the course of the seasons (Fig. 6b) as they do at L4 and E1, and statistical tests between daily match-ups (Fig. 6c) are not so good when compared with the two offshore stations, with the AVHRR data explaining only 87 % of the variance in the *in situ* data, with a systematic negative bias ($\delta = -1.20$°C), lower precision ($\Delta = 1.08$°C), slopes less than one ($S = 0.89$) and an intercept (I) of 0.31. The results indicate a degradation in the performance of the AVHRR data at the coastline, when compared with Station L4 and E1.

3.3. AVHRR comparison of satellite passes

Scatter plots of AVHRR satellite passes and *in situ* SST data at Station L4, E1 and measurements collected around the coastline of UK and Ireland by the surfers, are shown in Fig. 7, for an absolute time difference (T) of <1 h, <3 h and <5 h. In general, the statistical performance of the AVHRR data at L4 (Fig. 7d, e, and f) and E1 (Fig. 7g, h, and i) are consistent with that in the comparison of daily AVHRR values (Fig. 6), with high coefficient of determination (>0.95), no biases ($\delta \sim 0$), slopes (S) close to one and intercepts (I) close to zero. The root mean square errors (Ψ), composed principally by the precision component (Δ) considering the biases were zero (Fig. 7), are slightly higher (Ψ = 0.52 to 0.54) than the daily AVHRR comparison at L4 (Ψ = 0.44, Fig. 6e), and higher at L4 (Ψ = 0.52 to 0.54) than at E1 (Ψ = 0.45 to 0.47).

Consistent with the daily AVHRR comparison, statistical tests of AVHRR and *in situ* data indicate a significantly better performance in AVHRR SST at the two offshore stations (L4 and E1) when compared with performance at the coastline (Fig. 7), with Ψ two to three times higher at the coastline than offshore (L4 and E1), a systematic negative bias in AVHRR at the coastline ($\delta = -0.39$ to -1.07°C), slopes less than one and generally high intercepts (Fig. 7a-c). At L4 and E1, there is an increase in Ψ from <1 h to <5 h. The same is shown at the coastline between <3 h and <5 h (Fig. 7b and c). Figure 8 shows Ψ plotted as a function of T at the coastline (beaches) and at L4 and E1. In all cases, there is a significant increase in Ψ with T. At E1 and L4, this increase is linear. At the beaches, there is a sharp increase after 3 hr, with Ψ significantly higher at 6 hr (confidence intervals do not overlap).

4. Discussion

The coastal zone is arguably one of the most precious marine environments on the planet, containing the highest level of marine biodiversity (Tittensor et al., 2010), a large proportion of the world's fish catch (Stewart et al., 2010), and supporting a wide range of human activities, from energy extraction (Gill, 2005) to waste disposal. It is also vulnerable to increasing human pressure and climate change (Jickells, 1998; Lotze et al., 2006; McGranahan et al., 2007). Adequate management of the coastal environment requires the monitoring of key environmental indicators like SST (Bojinski et al., 2014). Yet, the coastal environment is drastically under-sampled and the observational networks are not adequate to meet management needs. Due to the paucity of data in coastal systems, there is increasing reliance placed on using models. Yet, these models are often based on false assumptions and are usually not verified with field data (Livingston, 2014). New solutions are needed to increase the spatial and temporal sampling of *in situ* data in the coastal zone.

4.1. Monitoring SST at the coastline in situ using recreational citizens

Here, we utilised a small group of surfers who regularly immerse themselves in the coastal zone, to measure SST over a three year period. The SST collected by the surfers were found to be in good agreement with measurements collected at two nearby oceanographic stations giving confidence in the method (Fig. 5), with discrepancies consistent with the spatial separation of sampling locations. It has been estimated that in the region of 40 million measurements of SST per year could be acquired in the UK coastal zone by tagging surfers with temperature sensors (Brewin et al., 2015b). In the US there are an estimated ~3.3 million surfers who surf ~108 times per year (Thomas, 2012), suggesting a potential of an additional ~350 million measurements of SST per year in the US. Surfers often visit remote and uninhabited regions, countries with limited coastal monitoring infrastructure and capabilities, where few coastal observations have been collected, regions that are highly vulnerable to climate change (e.g. Latin America and the East Asia Pacific).

There are also many other recreational watersports beyond surfing, which involve direct interaction with the aquatic environment in regions that are difficult to measure using conventional platforms. It has been demonstrated that recreational divers (Boss and Zaneveld, 2003; Wright et al., 2016), kayakers (Bresnahan et al., 2016), stand-up paddle-boarders (Bresnahan et al., 2016) and recreational sailors (Lauro et al., 2014), could contribute significantly to data collection in the coastal zone. Considering many of these other recreational watersports occur in maritime conditions different to that of surfing (e.g. calm seas), integrating such observations with data from surfers could increase the range of environmental conditions sampled by citizens. With rapid improvements in technology, including: miniaturisation of sensors, wireless data transfer, cloud data storage and wireless communication, the feasibility of harnessing citizens for coastal monitoring is becoming a real option (Busch et al., 2016; Farnham et al., 2017). Integrating these observations with other developing *in situ* techniques, such as coastal gliders (Rudnick et al., 2004), autonomous beach buoy systems (Shively et al., 2016) and the tagging of marine vertebrates with sensors (Fedak, 2004), as well as traditional *in situ* methods from ships and buoys, would significantly enhance the spatial and temporal sampling of *in situ* data in the coastal zone.

4.2. Satellite remote sensing of SST

The combined spatial and temporal coverage of satellite remote sensing observations, and its synoptic capabilities, means it provides more observations of SST than any other technique over wide spatial scales, and has significantly impacted operational ocean forecasting (Donlon et al., 2007). Yet, satellite remote sensing of SST has certain limitations. Thermal radiation emitted from the ocean is impacted by clouds and is only representative of the first few millimeters (the skin) of the ocean, relying on algorithmic conversions and assumptions to derive SST (at 1 m depth in the ocean), which can then be compared with the *in situ* datasets collected at $\sim$ 1m depth. To maximise the use of satellite SST data, the accuracy and precision of the data must be determined, which requires direct comparison with co-located and concomitant *in situ* data. The lack of *in situ* SST observations at the coastline means to date, our knowledge of the accuracy and precision of satellite SST at the coastline is severely limited. In light of the next generation of satellite thermal sensors (e.g. ESA's Sentinel 3 programme with dual-view measurement capabilities and proposed high resolution thermal sensors) it is vital these *in situ* networks are improved, to maximise the use of satellite SST observations for long-term monitoring and operational coastal applications.

When compared with other AVHRR SST processing systems, the operational NEODAAS system works well in offshore waters (Station L4 and E1) with no systematic difference ($\delta \sim 0.0$, see Fig. 7). The centre-pattern root mean square error (Δ) in AVHRR data for E1 and L4 data varies between 0.45 and 0.51 °C respectively, within an hour absolute time difference (Fig. 7). When using the robust standard deviation between match-ups rather than Δ, calculated by scaling the median absolute deviation from the median (making it less sensitive to outliers), these values drop to 0.18 and 0.21 °C, which fall below the range (0.26 and

0.58 °C) presented in a global validation by Merchant et al. (2014, see their Table 3) for various AVHRR sensors, giving confidence in the operational AVHRR SST data provided by NEODAAS.

At the coastline we observe a significant degradation in the performance of AVHRR at retrieving SST (Figs. 6, 7 and 8), with significantly higher root mean square errors (Ψ) that at L4 and E1, in the range of 1.0 to 2.0°C (Fig. 8). This clearly limits the use of AVHRR SST data at the coastline for applications that require errors to be less than that in this range. This finding is consistent with that of Smit et al. (2013), who caution against the use of 4 km SST MODIS Terra and Pathfinder v5.2 products around the coastline of South Africa, and observed significant biases between the satellite and *in situ* datasets. Yet, for applications that don't require high accuracy and precision, AVHRR SST data at the coastline may still have some use. For instance, in August 2014 there was a significant reduction in SST in Plymouth coastal and offshore waters, of the order of 3 to 4 °C seen in the *in situ* and satellite observations (Figs. 6). The AVHRR SST data at the coastline captured this decrease (Fig. 6), which was larger than the errors reported in the validation.

Yet, for the majority of applications where error requirements in SST are lower than 1.0 °C, there needs to be a significant improvement in the satellite AVHRR SST processing systems at the coastline. Retrievals of SST at the coastline are inherently complex when compared with offshore waters, owing to factors such as land contamination (e.g. from tidal changes), land adjacency issues, complexities in atmospheric-correction (e.g. from coastal aerosols), potential changes in the conversions from skin temperature to SST (e.g. from more bubbles at the land-sea interface; Jessup et al., 1997; Eifler and Donlon, 2001), and errors in satellite georeferencing. With better coastal *in situ* networks, we can drastically increase the number of co-incident and concurrent satellite and *in situ* match-ups, which in addition to validation, may help improve algorithm development.

Even with more *in situ* data, validation of satellite retrievals of SST at the coastline are more challenging than in offshore waters. SST at the coastline can be notoriously heterogeneous, due to a variety of factors such as: freshwater runoff at the coastline (e.g. impact of land run-off as well as nearby rivers and estuaries); tidal stirring; exchanges of heat between the land and ocean; and wave breaking (Farmer and Gemmrich, 1996), resulting in gradients in SST within a 1 km pixel that may not be captured by the surfer. Figure 1d illustrates the coverage of a typical GPS track by a surfer within a mapped NEODAAS AVHRR SST pixel, highlighting large differences in the spatial sampling in SST by the surfer and by the satellite. In some cases, it may be that portion of the pixel the surfer is sampling (the shallow landward boundary) has a systematically different temperature than the average of the pixel. This difference could be higher (consistent with the negative bias we see in Fig. 6c and 7a-c) where the shallow landward boundary might heat up quicker than the average, or even lower, in cases where a colder landmass (or fresh water run-off) is significantly influencing the shallower landward boundary of the pixel (e.g. in Autumn). This spatial heterogeneity could be quantified by integrating high spatial resolution thermal observations (e.g. Landsat or from aircraft platforms) with the courser resolution AVHRR data, but would be limited by infrequent concurrent overpasses. This coastal heterogeneity also has a temporal component that is likely to be greater than in offshore waters. Figure 8 highlights a sharp jump in the root mean square error (Ψ) when increasing the absolute time difference (T) between the *in situ* and satellite data beyond three hours, emphasising a requirement to minimise T when validating SST retrievals at the coastline. This sharp increase may be related to the semi-diurnal tidal cycle in the region.

5. Conclusions

To evaluate the suitability of EO SST data for coastal applications, it is essential to know the accuracy and precision of the data. This involves matching co-located and concomitant *in situ* and EO SST data. Due to a limited number of *in situ* measurements, little is know about the accuracy and precision of the EO SST data at the coastline. Using *in situ* SST measurements collected by a group of surfers over a three year period in the coastal waters of the UK and Ireland, we evaluated the accuracy and precision of operational AVHRR SST data at the coastline. When compared with match-ups at two autonomous buoys ~7 km and ~33 km offshore, we observed a significant reduction in the performance of AVHRR at retrieving SST at the coastline. Root mean square errors at the coastline were in the range of 1.0 to 2.0 °C, depending on the temporal difference between match-ups, significantly higher than those at the two offshore stations (0.4 to 0.6 °C). For match-ups at the coastline we also observed a systematic negative bias in the AVHRR retrievals of roughly 1 °C, and an increase in root mean square error when the temporal difference between match-ups exceeded three hours.

Tagging recreational water-users, like surfers, with sensors has the potential to improve the spatial and temporal coverage of *in situ* measurements at the coastline. This can aid our understanding of the accuracy and precision of the EO data, improve algorithm development, and inform users interested in using EO SST products for coastal applications. However, when compared with offshore waters, comparing EO SST products with *in situ* SST at the coastline is challenging. The dynamic and inherently complex coastal environment is difficult to sample remotely and *in situ*, and it is more complicated to reconcile geophysical and spatial differences between the two types of SST observations. Yet, in the face of increasing human pressures and climate change, our coastal seas require careful monitoring. This can only be achieved through integrating observations from different sources, including new *in situ* sampling and EO.

A. Appendix A

To compare the estimates of SST from two sources the following univariate statistical tests were used.

A.1. Coefficient of determination (r^2)

The coefficient of determination (r^2) was taken to be the square of the Pearson correlation coefficient (or squared Pearson's product moment correlation) and was calculated according to

$$r^2 = \left\{ \frac{1}{N-1} \sum_{i=1}^{N} \left[\frac{X_i^M - \left(\frac{1}{N}\sum_{j=1}^{N} X_j^M\right)}{\{\frac{1}{N-1}\sum_{k=1}^{N}\left[X_k^M - \left(\frac{1}{N}\sum_{l=1}^{N} X_l^M\right)\right]^2\}^{1/2}} \right] \left[\frac{X_i^E - \left(\frac{1}{N}\sum_{m=1}^{N} X_m^E\right)}{\{\frac{1}{N-1}\sum_{n=1}^{N}\left[X_n^E - \left(\frac{1}{N}\sum_{o=1}^{N} X_o^E\right)\right]^2\}^{1/2}} \right] \right\}^2, \tag{A.1}$$

where, X is the variable (e.g. SST) and N is the number of samples. The superscript E denotes the estimated variable (e.g. from the satellite sensor) and the superscript M denotes the measured variable (e.g. measured *in situ*). Note that the Pearson correlation coefficient assumes a linear relationship between variables. The squared correlation coefficient may take any value between 0 and 1.0, with 1.0 indicating the estimated variable explains 100% of the variability in the measured variable.

A.2. Root Mean Square Error (Ψ)

The absolute Root Mean Square Error (Ψ) was calculated according to

$$\Psi = \left[\frac{1}{N} \sum_{i=1}^{N} \left(X_i^E - X_i^M \right)^2 \right]^{1/2}. \tag{A.2}$$

The Root Mean Square Error (Ψ) can be partitioned into the bias (δ), which represent the systematic difference between variables (accuracy), and the centre-pattern (or unbiased) Root Mean Square Error (Δ), which represents the random difference between two variables (precision), such that $\Psi = \sqrt{(\Delta^2 + \delta^2)}$. Compuation of δ and Δ are described next.

A.3. The bias (δ)

The absolute bias between the estimated and measured variable was expressed according to

$$\delta = \frac{1}{N}\sum_{i=1}^{N}\left(X_i^E - X_i^M\right). \qquad \text{(A.3)}$$

A.4. The centre-pattern Root Mean Square Error (Δ)

The absolute centre-pattern (or unbiased) Root Mean Square Error (Δ) was calculated according to

$$\Delta = \left(\frac{1}{N}\sum_{i=1}^{N}\{\left[X_i^E - \left(\frac{1}{N}\sum_{j=1}^{N}X_j^E\right)\right] - \left[X_i^M - \left(\frac{1}{N}\sum_{k=1}^{N}X_k^M\right)\right]\}^2\right)^{1/2}. \qquad \text{(A.4)}$$

It describes the error of the estimated values with respect to the measured ones, regardless of the average bias between the two distributions.

A.5. Slope (S) and Intercept (I) of a linear regression

The performance of a model with respect to *in situ* data can be tested using linear regression between the estimated variable (from the model) and the measured variable (*in situ* data), such that

$$X^E = X^M S + I. \qquad \text{(A.5)}$$

A slope (S) close to one and an intercept (I) close to zero is an indication that the model compares well with the *in situ* data.

B. Appendix B

In Appendix B we provide supporting information on the processing of the SST data collected by surfers in the study. We demonstrate an improved consistency between the Tidbit v2 sensors when correcting each sensor to the same common reference. Figure A1 shows data collection by two surfers at the same location using two different sensors at an overlapping time period in the water (purple shading). The systematic difference (δ) between sensor readings were reduced when correcting each sensor to the same common reference using the piecewise, bias-correction model (Fig. 2h).

We also provide supporting information illustrating the method used to process the data collected by surfers and derive SST (see Fig. A2). A superposition of all temperature data acquired by the surfer during the study period, normalised such that the start and end of the surf is at the same point on the x-axis for each session, is provided in Fig. A3. The plot highlights the stability of the temperature of the sensor in the sea compared with that before and after each surf.

Acknowledgements

We thank Gavin Tilstone and Vassilis Kitidis for help comparing the temperature sensor (Tidbit v2 temperature logger) used in the study with the VWR1620-200 traceable digital thermometer. We thank Giorgio Dall'Olmo and Dionysios Raitsos for encouragement, motivation and useful discussions. We thank Stefano Ciavatta for help with statistical interpretation. The work was initally presented at the 2016 European Space Agency Living Planet Symposium in Prague before being updated and written up. The work was supported by Plymouth Marine Laboratory and NERC's UK National Centre for Earth Observation. JS time was funded by the NERC project Candyfloss NE/K002058/1. We thank the NERC Earth Observation Data Acquisition and Analysis Service (NEODAAS) for the AVHRR imagery.

References

Bader, J., Latif, M., 2003. The impact of decadal-scale Indian Ocean sea surface temperature anomalies on Sahelian rainfall and the North Atlantic Oscillation. Geophysical Research Letters 30, 2169.

Bojinski, S., Verstraete, M., Peterson, T., Richter, C., Simmons, A., Zemp, M., 2014. The concept of essential climate variables in support of climate research, applications, and policy. Bulletin of the American Meteorological Society 95, 1431–1443.

Bordes, P., Brunel, P., Marsouin, A., 1992. Automatic adjustment of AVHRR navigation. Journal of Atmospheric and Oceanic Technology 9, 15–27.

Boss, E., Zaneveld, J. R. V., 2003. The effect of bottom substrate on inherent optical properties: Evidence of biogeochemical processes. Limnology and Oceanography 48, 346–354.

Bresnahan, P. J., Wirth, T., Martz, T. R., Andersson, A. J., Cyronak, T., D'Angelo, S., Pennise, J., Melville, W. K., Lenain, L., Statom, N., 2016. A sensor package for mapping pH and oxygen from mobile platforms. Methods in Oceanography 17, 1–13.

Brewin, R. J. W., de Mora, L., Jackson, T., Brewin, T. G., Shutler, J., 2015a. Annual time series of sea surface temperature (SST) measurements collected by a surfer at Wembury Beach, Plymouth, UK. Tech. rep., British Oceanographic Data Centre - Natural Environment Research Council, UK.

Brewin, R. J. W., de Mora, L., Jackson, T., Brewin, T. G., Shutler, J., 2015b. On the potential of surfers to monitor environmental indicators in the coastal zone. PLoS One 10 (7), e0127706.

Brewin, R. J. W., de Mora, L., Jackson, T., Brewin, T. G., Shutler, J., Billson, O., 2017. Sea surface temperature (SST) measurements collected by surfers around the southern UK and western Ireland coastline between 2014 and 2017. Tech. rep., British Oceanographic Data Centre - Natural Environment Research Council, UK.

Brewin, R. J. W., Sathyendranath, S., Müller, D., Brockmann, C., Deschamps, P.-Y., Devred, E., Doerffer, R., Fomferra, N., Franz, B. A., Grant, M., Groom, S., Horseman, A., Hu, C., Krasemann, H., Lee, Z.-P., Maritorena, S., Mélin, F., Peters, M., Platt, T., Regner, P., Smyth, T., Steinmetz, F., Swinton, J., Werdell, J., White III, G. N., 2015c. The Ocean Colour Climate Change Initiative: III. A round-robin comparison on in-water bio-optical algorithms. Remote Sensing Environment 162, 271–294.

Busch, J., Bardaji, R., Ceccaroni, L., Friedrichs, A., Piera, J., Simon, C., Thijsse, P., Wernand, M., van der Woerd, H. J., Zielinski, O., 2016. Citizen Bio-Optical Observations from Coast-and Ocean and Their Compatibility with Ocean Colour Satellite Measurements. Remote Sensing 8 (11), 1–19.

Costanza, R., d'Arge, R., de Groot, R., Farberk, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R. V., Paruelo, J., Raskin, R. G., 1997. The value of the world's ecosystem services and natural capital. Nature 387, 253–260.

Costanza, R., de Groot, R., Sutton, P., van der Ploeg, S., Anderson, S. J., Kubiszewski, I., Farber, S., Turner, R. K., 2014. Changes in the global value of ecosystem services. Global Environmental Change 26, 152–158.

Couce, E., Ridgwell, A., Hendy, E. J., 2012. Environmental controls on the global distribution of shallow-water coral reefs. Journal of Biogeography 39, 1508–1523.

Doney, S. C., 2006. Plankton in a warmer world. Nature 444, 695–696.

Doney, S. C., Lima, I. D., Moore, J. K., Lindsay, K., Behrenfeld, M. J., Westberry, T. K., Mahowald, N., M., G. D., Takahashi, T., 2009. Skill metrics for confronting global upper ocean ecosystem-biogeochemistry models against field and remote sensing data. Journal of Marine Systems 76, 95–112.

Donlon, C., Rayner, N., Robinson, I., Poulter, D. J. S., Casey, K. S., Vazquez-Cuervo, J., Armstrong, E., Bingham, A., Arino, O., Gentemann, C., May, D., 2007. The global ocean data assimilation experiment high-resolution sea surface temperature pilot project. Bulletin of the American Meteorological Society 88 (8), 1197–1213.

Eifler, W., Donlon, C. J., 2001. Modeling the thermal surface signature of breaking waves. Journal of Geophysical Research 106, 27163–27185.

Eppley, R. W., 1972. Temperature and phytoplankton growth in the sea. Fishery

Bulletin 70, 1063–1085.

Farmer, D. M., Gemmrich, J. R., 1996. Measurements of temperature fluctuations in breaking surface waves. Journal of Physical Oceanography 26, 816–825.

Farnham, D. J., Gibson, R. A., Hsueh, D. Y., McGillis, W. R., Culligan, P. J., Zain, N., Buchanan, R., 2017. Citizen science-based water quality monitoring: Constructing a large database to characterize the impacts of combined sewer overflow in New York City. Science of The Total Environment 580, 168–177.

Fedak, M., 2004. Marine animals as platforms for oceanographic sampling: a "winwin" situation for biology and operational oceanography. Memoirs of National Institute of Polar Research. Special issue 58, 133–147.

Frederiksen, M., Edwards, M., Mavor, R. A., Wanless, S., 2007. Regional and annual variation in black-legged kittiwake breeding productivity is related to sea surface temperature. Marine Ecology Progress Series 350, 137–143.

Frost, M. T., Bayliss-Brown, G., Buckley, P., Cox, M., Dye, S. R., Sanderson, W. G., Stoker, B., Withers Harvey, N., 2016. A review of climate change and the implementation of marine biodiversity legislation in the United Kingdom. Aquatic Conservation: Marine and Freshwater Ecosystems 26, 576–595.

GCOS, 2011. Systematic observation requirements from satellite-based data products for climate. Tech. rep., World Meteorological Organisation (WMO), 7 bis, avenue de la Paix, CH-1211 Geneva 2, Switzerland.

GHRSST, 2017. Group for High Resolution Sea Surface Temperature: Products. URL https://www.ghrsst.org/ghrsst-data-services/products/

Gill, A. B., 2005. Offshore renewable energy: ecological implications of generating electricity in the coastal zone. Journal of Applied Ecology 42, 605–615.

Goreau, T., Hayes, R. L., 1994. Coral bleaching and ocean "hot spots". Ambio 23, 176–180.

Holt, J., Schrum, C., Cannaby, H., Daewel, U., Allen, I., Artioli, Y., Bopp, L., Butenschon, M., Fach, B., Harle, J., Pushpadas, D., 2016. Potential impacts of climate change on the primary production of regional seas: a comparative analysis of five European seas. Progress in Oceanography 140, 91–115.

Jessup, A., Zappa, C., Loewen, M., Hesany, V., 1997. Infrared remote sensing of breaking waves. Nature 385, 52–55.

Jickells, T. D., 1998. Nutrient biogeochemistry of the coastal zone. Science 281, 217–222.

Keller, A. A., Oviatt, C. A., Walker, H. A., Hawk, J. D., 1999. Predicted impacts of elevated temperature on the magnitude of the winter-spring phytoplankton bloom in temperate coastal waters: A mesocosm study. Limnology and Oceanography 44, 344–356.

Kitidis, V., Brown, I., Hardman-Mountford, N., Lefévre, N., In press. Surface ocean carbon dioxide during the Atlantic Meridional Transect (1995-2013); evidence of ocean acidification. Progress in Oceanography.

Land, P. E., Shutler, J. D., Cowling, R. D., Woolf, D. K., Walker, P., Findlay, H. S., Upstill-Goddard, R. C., Donlon, C. J., 2013. Climate change impacts on sea-air fluxes of CO_2 in three Arctic seas: a sensitivity study using Earth observation. Biogeosciences 10, 8109–8128.

Lauro, F. M., Senstius, S. J., Cullen, J., Neches, R., Jensen, R. M., Brown, M. V., Darling, A. E., Givskov, M., McDougald, D., Hoeke, R., Ostrowski, M., Philip, G. K., Paulsen, I. T., Grzymski, J. J., 2014. The Common Oceanographer: Crowdsourcing the Collection of Oceanographic Data. PLoS Biology 12 (9), e1001947.

Lazareth, C. E., Putten, E. V., André, L., Dehairs, F., 2003. High-resolution trace element profiles in shells of the mangrove bivalve Isognomon ephippium: a record of environmental spatio-temporal variations? Estuarine, Coastal and Shelf Science 57, 1103–1114.

Lea, D. W., Pak, D. K., Spero, H. J., 2000. Climate Impact of Late Quaternary Equatorial Pacific Sea Surface Temperature Variations. Science 289, 1719–1724.

Lee, K., Tong, L., Millero, F. J., Sabine, C. L., Dickson, A. G., Goyet, C., Park, G. H., Wanninkhof, R., Feely, R. A., Key, R. M., 2006. Global relationships of total alkalinity with salinity and temperature in surface waters of the world's oceans. Geophysical Research Letters 33 (19).

L'Hévéder, B., Speich, S., Ragueneau, O., Gohin, F., Bryère, P., 2016. Observed and projected sea surface temperature seasonal changes in the Western English Channel from satellite data and CMIP5 multi-model ensemble. International Journal of Climatology 37, 2831–2849.

Livingston, R. J., 2014. Climate change and coastal ecosystems: long-term effects of climate and nutrient loading on trophic organization. CRC Press.

Llewellyn-Jones, D. T., Minnett, P. J., Saunders, R. W., Zavody, A. M., 1984.

Satellite multichannel infrared measurements of sea surface temperature of the NE atlantic ocean using AVHRR/2. Quarterly Journal of the Royal Meteorological Society 110 (465), 613–631.

Lotze, H. K., Lenihan, H. S., Bourque, B. J., Bradbury, R. H., Cooke, R. G., Kay, M. C., Kidwell, S. M., Kirby, M. X., Peterson, C. H., Jackson, J. B. C., 2006. Depletion, Degradation, and Recovery Potential of Estuaries and Coastal Seas. Science 312, 1806–1809.

McGranahan, G., Balk, D., Anderson, B., 2007. The rising tide: assessing the risks of climate change and human settlements in low elevation coastal zones. Environment and Urbanization 19, 17–37.

Merchant, C. J., Embury, O., Roberts-Jones, J., Fiedler, E., Bulgin, C. E., Corlett, G. K., Good, S., McLaren, A., Rayner, N., Morak-Bozzo, S., Donlon, C., 2014. Sea surface temperature datasets for climate applications from Phase 1 of the European Space Agency Climate Change initiative (SST CCI). Geoscience Data Journal 1, 179–191.

Miller, P., Groom, S., McManus, A., Selley, J., Mironnet, N., 1997. PANORMA: A semi-automated AVHRR and CZCS system for observation of coastal and ocean processes. In: Observations and Interactions, Proceesings of the Remote Sensing Society, Reading, September 1997.

Miller, P. I., Scales, K. L., Ingram, S. N., Southall, E. J., Sims, D. W., 2015. Basking sharks and oceanographic fronts: quantifying associations in the north-east Atlantic. Functional Ecology 29 (8), 1099–1109.

Moore, J., Abbott, M., Richman, J., 1999. Location and dynamics of the Antarctic Polar Front from satellite sea surface temperature data. Journal of Geophysical Research 104, 3059–3073.

Mustard, J. F., Carney, M. A., Sen, A., 1999. The Use of Satellite Data to Quantify Thermal Effluent Impacts. Estuarine, Coastal and Shelf Science 49, 509–524.

Nicholls, R. J., Wong, P. P., Burkett, V. R., Codignotto, J. O., Hay, J. E., McLean, R. F., Ragoonaden, S., Woodroffe, C. D., 2007. Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK,, Ch. Coastal systems and low-lying areas, pp. 315–356.

Nonaka, M., Xie, S., 2003. Covariations of sea surface temperature and wind over the Kuroshio and its extension: Evidence for ocean-to-atmosphere feedback. Journal of Climate 16 (9), 1404–1413.

Paerl, H. W., Huisman, J., 2008. Blooms like it hot. Science 320, 57–58.

Pepin, P., 1991. Effect of temperature and size on development, mortality, and survival rates of the pelagic early life history stages of marine fish. Canadian Journal of Fisheries and Aquatic Sciences 48 (3), 503–518.

Raitsos, D. E., Hoteit, I., Prihartato, P. K., Chronis, T., Triantafyllou, G., Y., A., 2011. Abrupt warming of the Red Sea. Geophysical Research Letters 38, L14601.

Reynolds, R. W., Smith, T. M., 1994. Improved global sea surface temperature analyses using optimum interpolation. Journal of Climate 7, 929–948.

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., Schlax, M. G., 2007. Daily High-Resolution-Blended Analysis for Sea Surface Temperature. Journal of Climate 20, 5473–5496.

Roozekrans, J. N., Prangsma, G. J., 1988. Processing and application of digital AVHRR imagery for land and sea surfaces. Final report of BCRS project no: TO-3.1 "digital thermal imagery of land and sea surfaces". Tech. rep., Royal Netherlands Meteorological Institute (KNMI).

Rudnick, D., Davis, R. E., Eriksen, C. C., Fratantoni, D. M., Perry, M. J., 2004. Underwater gliders for ocean research. Marine Technology Society Journal 38, 73–84.

Saji, N. H., Goswami, B. N., Vinayachandran, P. N., Yamagata, T., 1999. A dipole mode in the tropical Indian Ocean. Nature 401, 360–363.

Sandford, T. D. G., Stephenson, J., 1992. Orbital prediction for the NOAA satellite series. In: Remote sensing from Research to Operations: Proceedings of the Remote Sensing Society Annual Conference, University of Dundee, September 1992, 424-433.

Saunders, R. W., Kriebel, K. T., 1988. An improved method for detecting clear sky and cloudy radiances from AVHRR data. International Journal of Remote Sensing 9, 123–150.

Scales, K. L., Miller, P. I., Embling, C. B., Ingram, S., Pirotta, E., Votier, S. C., 2014. Mesoscale fronts as foraging habitats: composite front mapping reveals oceanographic drivers of habitat use for a pelagic seabird. Journal of the Royal Society Interface 11.

Shively, D. A., Nevers, M. B., Breitenbach, C., Phanikumar, M. S., Przybyla-Kelly, K., Spoljaric, A. M., Whitman, R. L., 2016. Prototypic automated continuous recreational water quality monitoring of nine Chicago beaches. Journal of Environmental Management 166, 285–293.

Smit, A. J., Roberts, M., Anderson, R. J., Dufois, F., Dudley, S. F., Bornman, T. G., Olbers, J., Bolton, J. J., 2013. A coastal seawater temperature dataset for biogeographical studies: large biases between in situ and remotely-sensed data sets around the coast of South Africa. PLoS One 8, e81944.

Smyth, T. J., Fishwick, J. R., Gallienne, C. P., Stephens, J. A., Bale, A. J., 2010. Technology, Design, and Operation of an Autonomous Buoy System in the Western English Channel. Journal of Atmospheric and Oceanic Technology 27, 2056–2064.

Stewart, K. R., Lewison, R. L., Dunn, D. C., Borkland, R. H., Kelez, S., Halpin, P. N., B., C. L., 2010. Characterizing Fishing Effort and Spatial Extent of Coastal Fisheries. PLoS ONE 5 (12), e14451.

Sutton, R. T., Allen, M. R., 1997. Decadal predictability of North Atlantic sea surface temperature and climate. Nature 388, 563–567.

Takahashi, T., Sutherland, S. C., Sweeney, C., Poisson, A., Metzl, N., Tilbrook, B., Bates, N., Wanninkhof, R., Feely, R. A., Sabine, C. Olafsson, J., Nojiri, Y., 2002. Global sea–air CO_2 flux based on climatological surface ocean pCO_2, and seasonal biological and temperature effects. Deep Sea Research Part II: Topical Studies in Oceanography 49 (9), 1601–1622.

Tang, D., Kester, D. R., Wang, Z., Lian, J., Kawamura, H., 2003. AVHRR satellite remote sensing and shipboard measurements of the thermal plume from the Daya Bay, nuclear power station, China. Remote Sensing of Environment 84 (4), 506–515.

Thiermann, V., Ruprecht, E., 1992. A method for detection of clouds using AVHRR infrared observations. International Journal of Remote Sensing 13, 1829–1841.

Thomas, A., Byrne, D., Weatherbee, R., 2002. Coastal sea surface temperature variability from Landsat infrared data. Remote Sensing of Environment 81, 262–272.

Thomas, G., August 2012. Surfonomics quantifies the worth of waves. The Washington Post 24, G1.

Tittensor, D. P., Mora, C., Jetz, W., Lotze, H. K., Ricard, D., Berghe, E. V., Worm, B., 2010. Global patterns and predictors of marine biodiversity across taxa. Nature 466, 1098–1101.

Wang, S., McGrath, R., Hanafin, J., Lynch, P., Semmler, T., Nolan, P., 2008. The impact of climate change on storm surges over Irish waters. Ocean Modelling 25 (1), 83–94.

Wentz, F. J., Gentemann, C., Smith, D., Chelton, D., 2000. Satellite measurements of sea surface temperature through clouds. Science 288, 847–850.

Wright, S., Hull, T., Sivyer, D. B., Pearce, D., Pinnegar, J. K., Sayer, M. D. J., Mogg, A. O. M., Azzopardi, E., Gontarek, S., Hyder, K., 2016. SCUBA divers as oceanographic samplers: The potential of dive computers to augment aquatic temperature monitoring. Scientific Reports 6, 1–8.

Yu, L., Weller, R. A., 2007. Objectively analyzed air-sea heat fluxes for the global ice-free oceans (1981-2005). Bulletin of the American Meteorological Society 88, 527–539.

Table 1: Details for each Tidbit v2 sensor of the number (N) of surfing sessions the sensor was used for during the study period and its duration of use.

Tidbit v2 sensor	N	Duration of use
10308732	141	5th Jan 2014 - 28th Nov 2015*
10551172	27	13th Sep 2014 - 6th Nov 2016#
10551173	35	12th Aug 2014 - 4th Jan 2017#
10551174	4	8th Jul 2015 - 7th Aug 2016#
10782552	90	28th Nov 2015 - 8th Feb 2017#

* Sensor ran out of battery after this date
# Sensor still operational at end of study

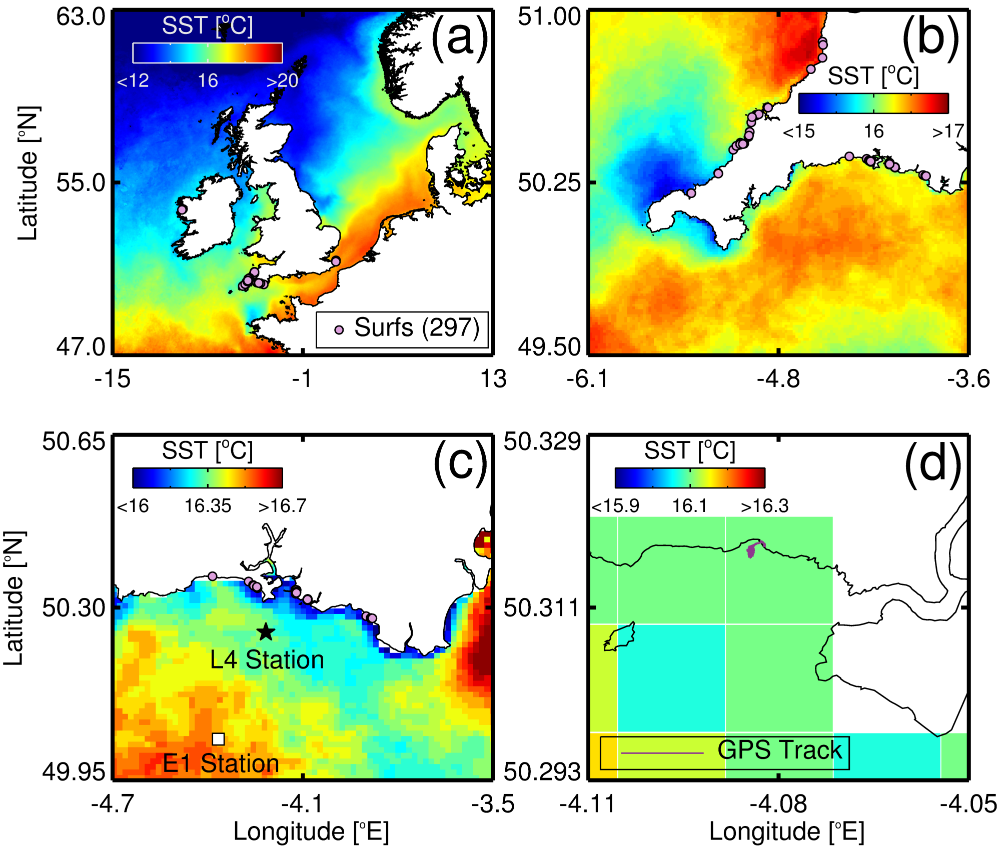

Figure 1: Study site and locations of sampling. (a) Shows the locations of the 297 surfing sessions where SST data were collected during the study in the UK and Ireland, overlain onto a NEODAAS AVHRR SST average composite image of September, averaged between the duration of the study (2014-2017). (b) Locations where the majority of samples were collected by the surfers around the south-west UK coastline, overlain onto the same September SST composite. (c) Sample locations near the city of Plymouth, UK, showing the position of two nearby oceanographic stations (Station L4 and E1) that form part of the Western Channel Observatory, all overlain onto the same September SST composite. (d) GPS track from a surf on the 20th September 2014, overlain onto the same September SST composite, to illustrate the coverage of a typical GPS track within a mapped NEODAAS AVHRR pixel.

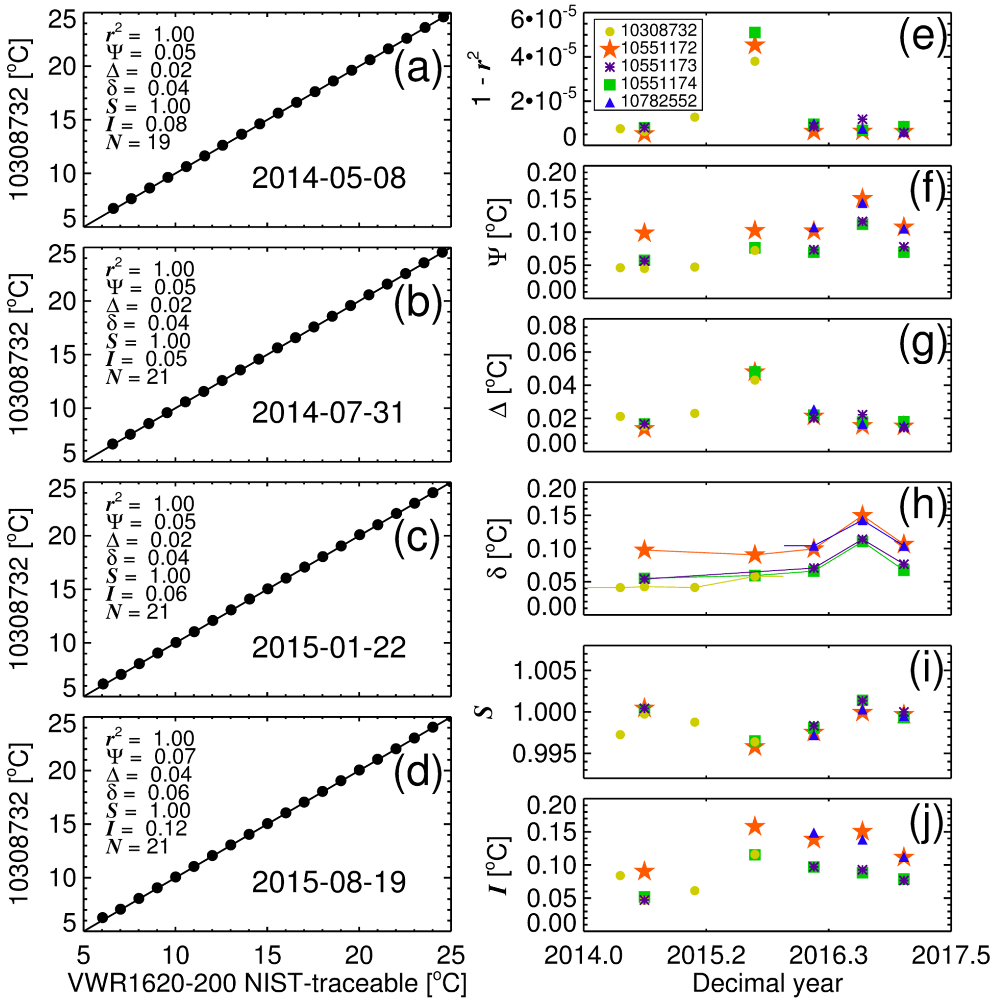

Figure 2: Laboratory comparisons between the Tidbit v2 sensors and a VWR1620-200 traceable digital thermometer, using a PolyScience temperature bath over the range from 6 to 25°C. (a-d) Illustrate four laboratory comparisons between Tidbit v2 sensor 10308732 and the VWR1620-200 traceable digital thermometer, and (e-j) show variations in statistical tests for each laboratory comparison, for the five Tidbit v2 sensors used in the study. Lines in (h) show the piecewise regression model used to correct the bias (δ) of each sensor over the time period of use. r^2 is the coefficient of determination, Ψ the root mean square error, δ the bias, Δ the centre-pattern (or unbiased) root mean square error, S the slope and I the intercept of a linear regression, and N the number of samples.

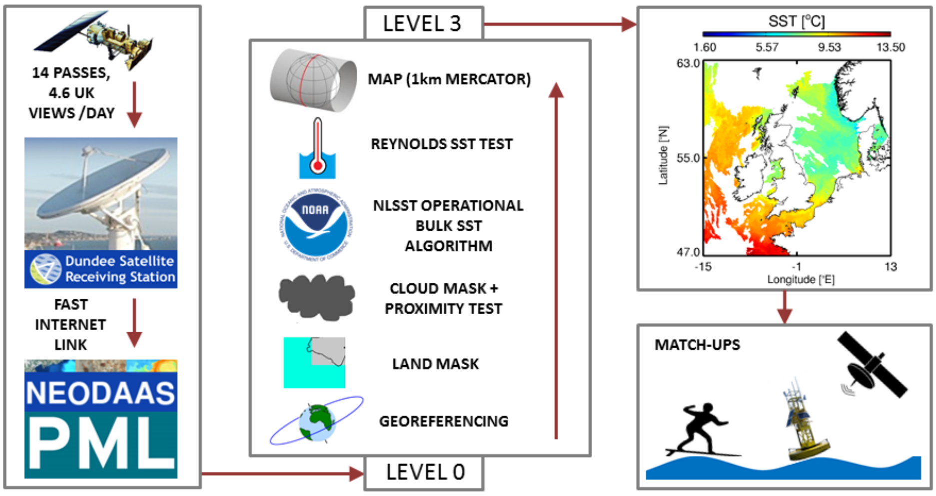

Figure 3: Schematic diagram of the NEODAAS system for producing the operational AVHRR SST products used in the study.

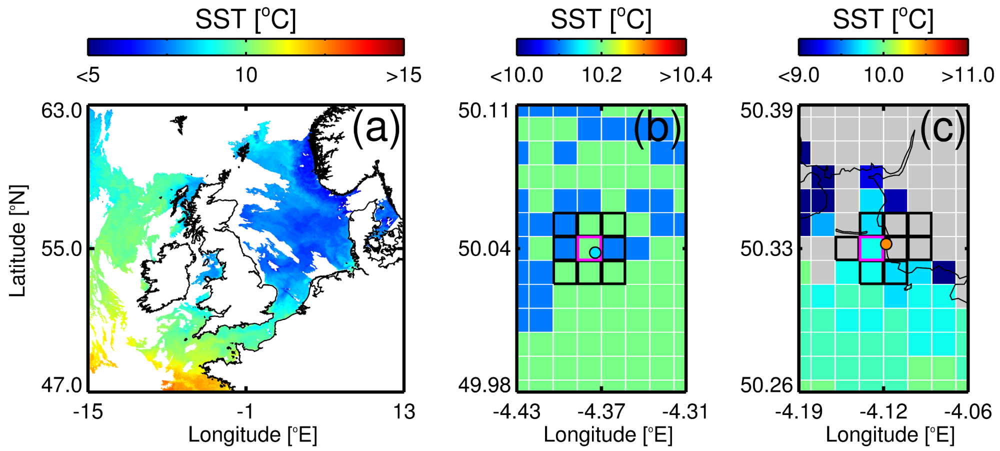

Figure 4: Example of the match-up process used in the study for Level 3 satellite passes. (a) Shows a relatively cloud free Level 3 AVHRR SST pass taken on the 20th April 2015 at 03:39 GMT, and processed by NEODAAS. (b) Shows the group of nine pixels in the AVHRR image centred on Station E1 (black and pink border) used to check homogeneity of the match-up region, with the centre pixel located closest to the E1 buoy (pink border) used for comparison with the E1 *in situ* data (circle and colour-coded to the same scale as the image) collected at 04:04 GMT on the 20th April 2015. (c) Shows the group of nine pixels (black and pink border) in the AVHRR image centred on Bovisand Beach, the location of a surfing session that took place on the 20th April 2015 at 05:58 GMT, that were used to check homogeneity of the match-up region, with the pixel with data located closest (<1 km) to the surf session (pink border) used for comparison with the *in situ* data (circle and colour-coded to the same scale as the image). Note that in this case, the closest pixel was actually dominated by land (i.e. the *in situ* measurement was at the edge of a land pixel) such that the next closest pixel with SST data within a 1 km radius was selected.

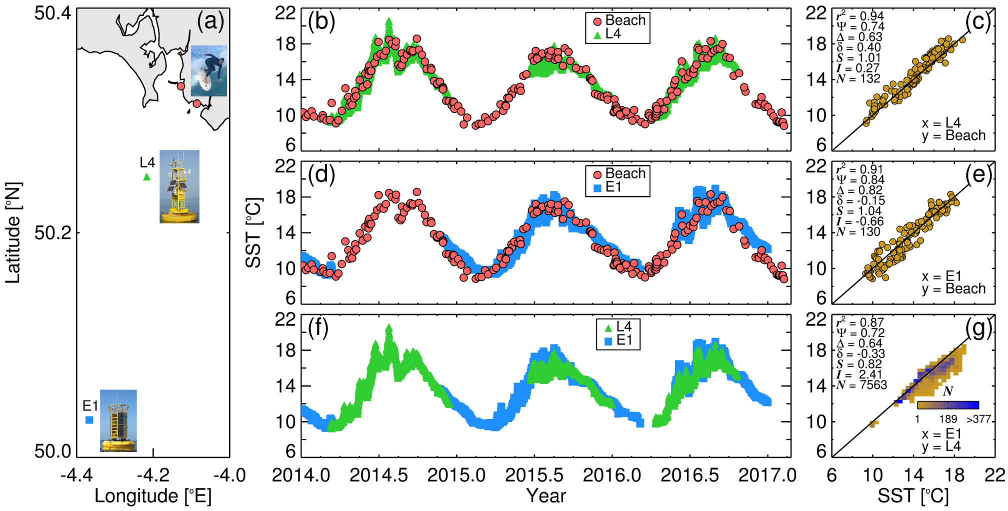

Figure 5: Comparison of *in situ* sea surface temperature (SST) datasets near Plymouth, UK. (a) Locations of SST data collected at the two beaches (Wembury and Bovisand), Station L4 and E1. (b) Time-series of SST acquired by the surfer at the two beaches overlain onto the SST data from Station L4. (c) Scatter plots of hourly match-ups between SST acquired by the surfer at the beaches and SST data from Station L4. (d) Time-series of SST acquired by the surfer at the beaches overlain onto the SST data from Station E1. (e) Scatter plots of hourly match-ups between SST acquired by the surfer at the beaches and SST data from Station E1. (f) Time-series of SST acquired at Station L4 overlain onto the SST data from Station E1. (g) Scatter plots of hourly match-ups between SST at L4 and E1. r^2 is the coefficient of determination, Ψ the root mean square error, δ the bias, Δ the centre-pattern (or unbiased) root mean square error, S the slope and I the intercept of a linear regression, and N the number of samples.

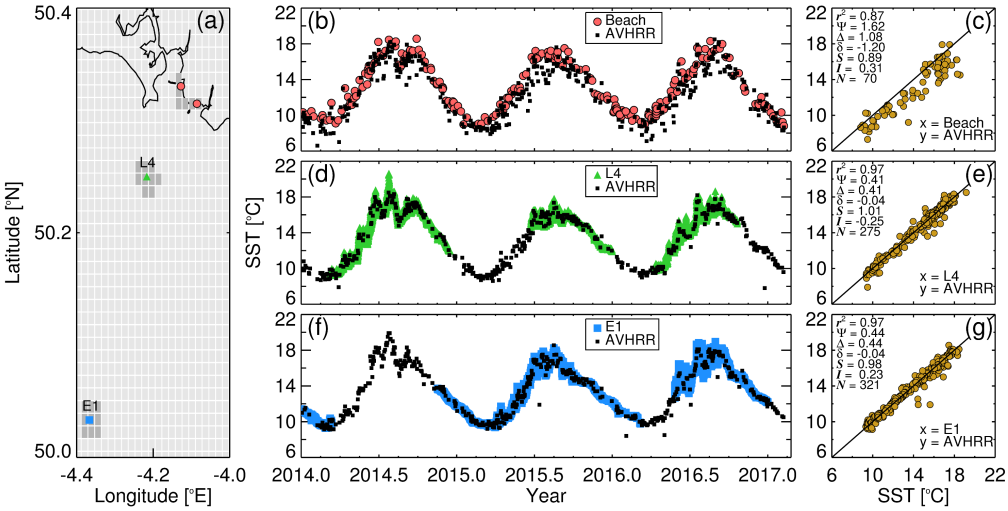

Figure 6: Comparison of daily Level 3 AVHRR and *in situ* sea surface temperature (SST) datasets near Plymouth, UK. (a) Locations of SST data collected at the two beaches (Wembury and Bovisand), at Station L4 and E1, and the group of pixels selected from the AVHRR data to be representative of the three locations (dark grey pixels). (b) Time-series of AVHRR Level 3 daily SST at the six pixels covering the two beaches overlain onto that acquired by the surfers *in situ* at the two beaches. (c) Scatter plots of daily match-ups between SST acquired *in situ* by the surfers and by AVHRR at the beaches. (d) Time-series of AVHRR SST overlain onto *in situ* SST at L4. (e) Scatter plots of daily match-ups between SST acquired *in situ* and by AVHRR at L4. (f) Time-series of AVHRR SST overlain onto *in situ* SST at E1. (g) Scatter plots of daily match-ups between SST acquired *in situ* and by AVHRR at E1. r^2 is the coefficient of determination, Ψ the root mean square error, δ the bias, Δ the centre-pattern (or unbiased) root mean square error, S the slope and I the intercept of a linear regression, and N the number of samples.

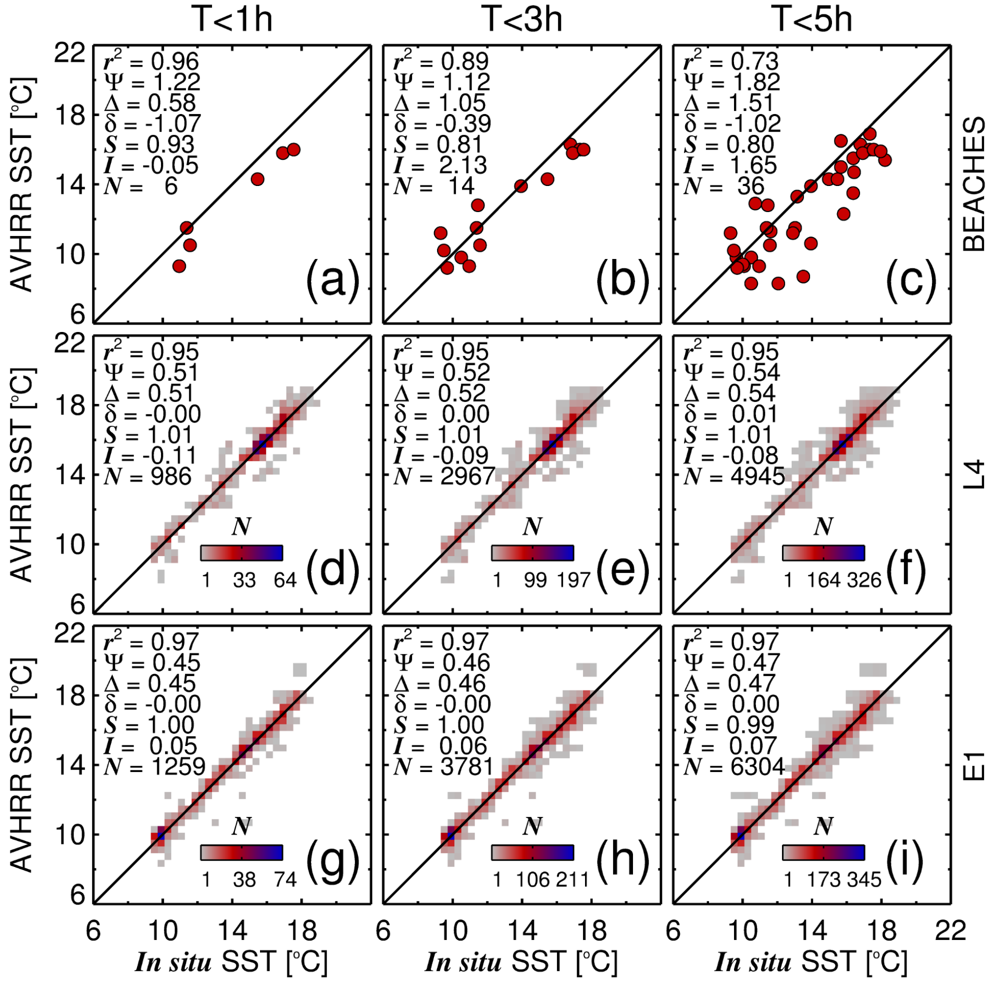

Figure 7: Scatter plots of Level 3 AVHRR satellite passes and *in situ* sea surface temperature (SST) data for an absolute time difference (T) of <1 h, <3 h and <5 h, at the coastline (a-c), at L4 (d-f) and at E1 (g-i). r^2 is the coefficient of determination, Ψ the root mean square error, δ the bias, Δ the centre-pattern (or unbiased) root mean square error, S the slope and I the intercept of a linear regression, and N the number of samples.

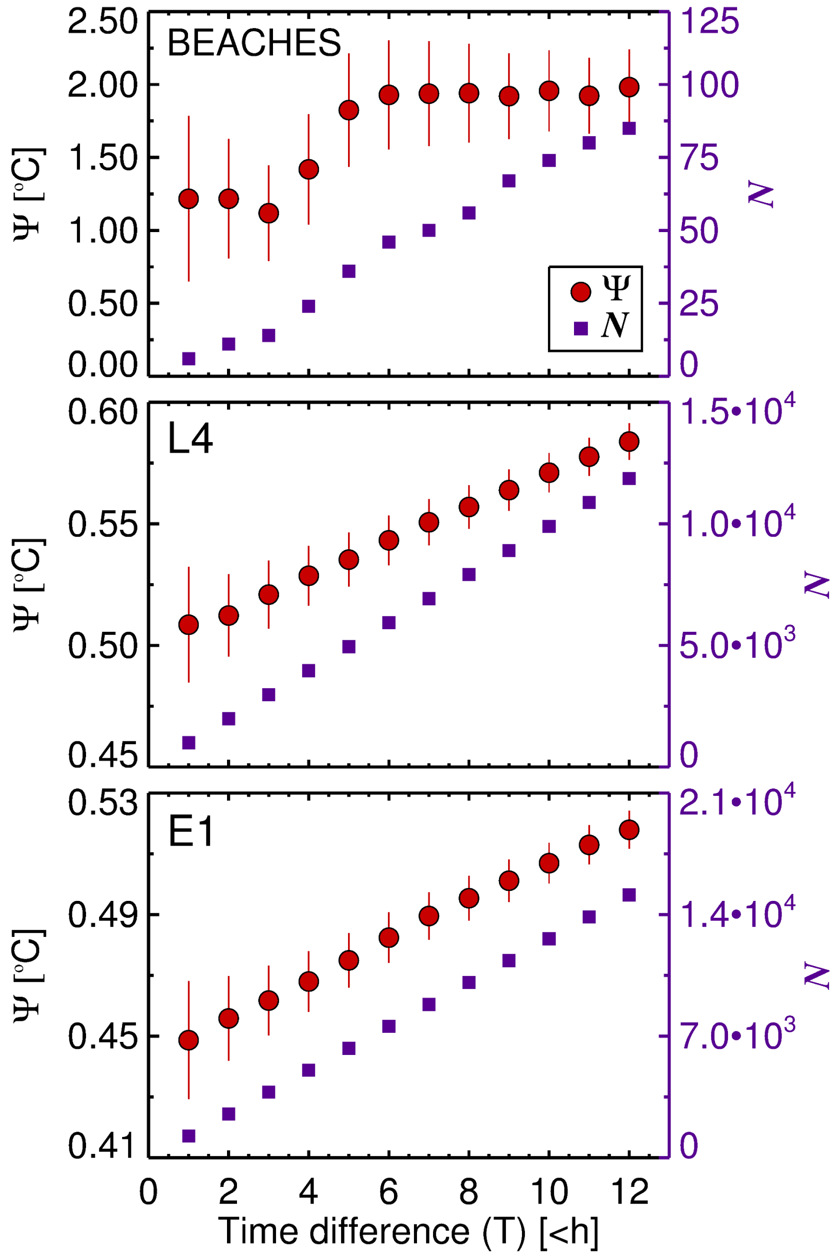

Figure 8: The root mean square error (Ψ) between Level 3 AVHRR satellite passes and *in situ* sea surface temperature (SST) data plotted as a function of the absolute time difference (T) at the coastline (beaches) and at L4 and E1. Confidence intervals (red lines) were computed based on the standard error of the mean and the *t*-distribution of the sample size.

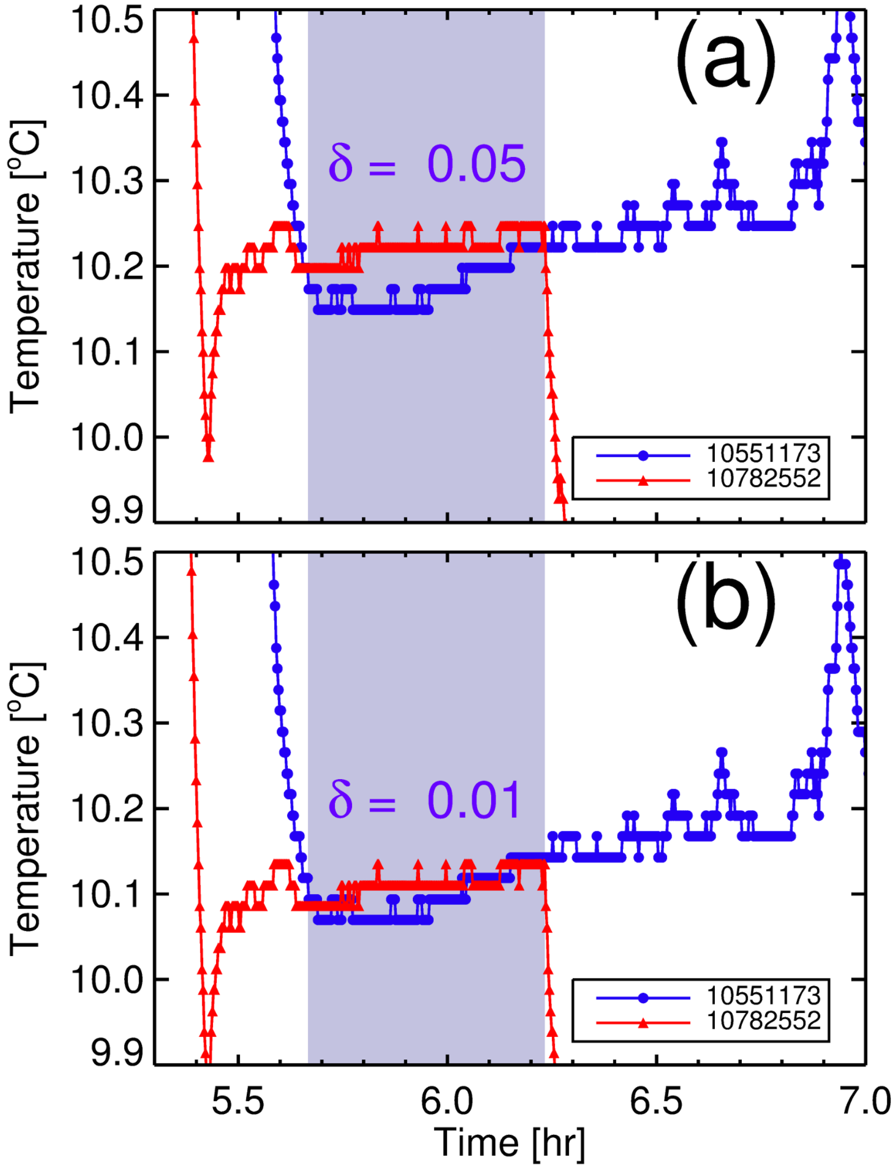

Figure A1: Comparison of temperature data collected by two surfers using two different Tidbit v2 sensors (10551173 and 10782552) at the same location (Bovisand Beach, Plymouth, UK) at an overlapping time period on the 14^{th} April 2016. (a) shows the raw comparison and (b) shows the comparison after application of the bias-correction model (piecewise regression model) such that each sensor was corrected to the same common reference. The systematic differences (δ) between the two sensors readings were reduced when correcting each sensor to the same common reference.

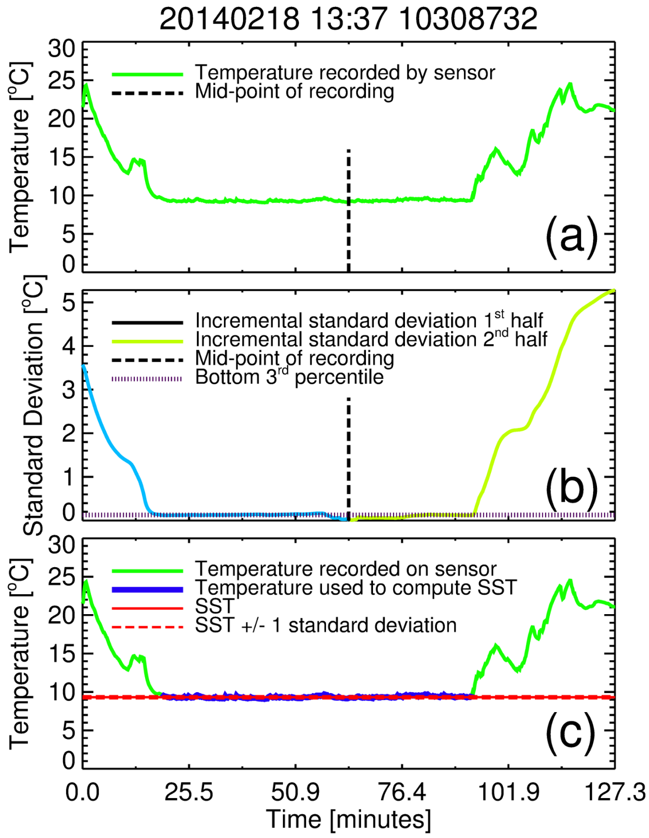

Figure A2: Illustration of the method used to process the data collected by a surfer and derive SST at Tolcarne Beach, Newquay, UK on the 18th February 2014. (a) Shows the raw temperature data collected by the surfer as a function of time, showing when the sensor was switched on (high temp), when the surfer was in the ocean (temperature stabalised around 9°C) and the rise in temperature as the surfer exited the water and uploaded the data. The midpoint of the surf is also shown. (b) Shows how the data were divided into two equal halves around the mid-point. For the first half of the data, every data point was removed sequentially in time and the standard deviation was calculated incrementally (light blue line), with the last data point representing the standard deviation of the midpoint (zero). For the second half of the data, this procedure was repeated but in reverse (light green line). The standard deviations for the two halves of the data were then recombined, and the bottom third percentile of the standard deviations were derived (purple dashed line). (c) The point at which the surfer began measuring SST (entered the water) was taken as the point when the standard deviation first fell below the bottom third percentile, and the point at which the surfer stopped measuring SST (exited the water) was taken as the last point of the session when the standard deviation was below the bottom third percentile. This data is shown in blue and is used to compute SST by taking the median of this data. Note that the median is resistant to outliers and thus fairly resilient to variations in the derived start and finish points.

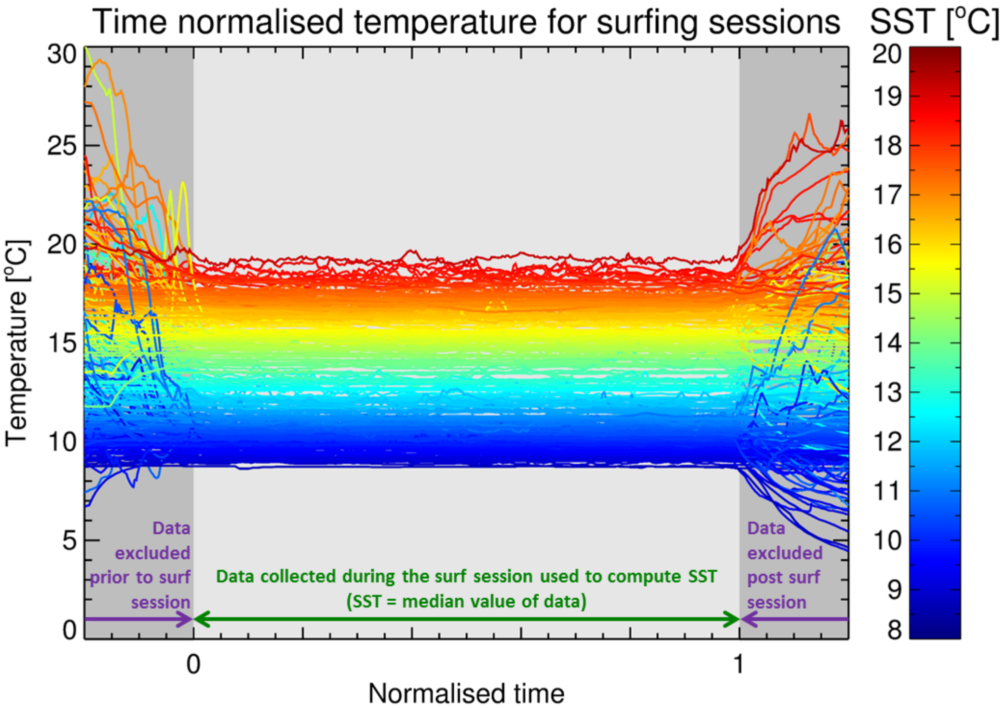

Figure A3: A superposition of all temperature data acquired by the surfer during the study period, normalised such that the start (0) and end (1) of the surfs are at the same point on the x-axis for each session. Data in dark grey were excluded and light grey included. The data in light grey were used to compute SST by taking the median of this data.