QUARTERLY JOURNAL OF THE ROYAL METEOROLOGICAL SOCIETY
RMetS Royal Meteorological Society

Measurements and models of the temperature change of water samples in Sea Surface Temperature buckets

Journal:	*QJRMS*
Manuscript ID	Draft
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Carella, Giulia; National Oceanography Centre, Marine Physics and Ocean Climate Morris, Andrew; National Oceanography Centre, Ocean Technology and Engineering Group Pascal, Robin; National Oceanography Centre, Ocean Technology and Engineering Group Yelland, Margaret; National Oceanography Centre, Marine Physics and Ocean Climate Berry, David; National Oceanography Centre, Marine Physics and Ocean Climate Morak-Bozzo, Simone; University of Reading, Dept. of Meteorology Merchant, Chris; University of Reading, Dept. of Meteorology; National Centre for Earth Observation Kent, Elizabeth; National Oceanography Centre,
Keywords:	sea surface temperature, climate change, observation bias, error model

Measurements and models of the temperature change of water samples in Sea Surface Temperature buckets

G. Carella * [1], A. K. R. Morris [1], R. W. Pascal [1], M. J. Yelland [1], D. I. Berry [1], S. Morak-Bozzo [2], C. J. Merchant [2,3] and E. C. Kent [1]

* Corresponding author

[1] National Oceanography Centre, Southampton UK.

[2] Department of Meteorology, University of Reading, Reading UK.

[3] National Centre for Earth Observation, Reading, UK.

Key words

sea surface temperature, climate change, observation bias, error model

Running Head

Temperature change of water samples in SST buckets.

Abstract

Uncertainty in the bias adjustments applied to historical sea surface temperature (SST) measurements made using buckets are thought to make the largest contribution to uncertainty in global surface temperature trends. Measurements of the change in temperature of water samples in wooden and canvas buckets used before World War 2 are compared with the predictions of models that have been used to estimate bias adjustments applied in widely-used gridded analyses of SST. The results show that the models are broadly able to predict the dependence of the temperature change of the water over time on the thermal forcing and the bucket characteristics: volume and geometry; structure and material. However, assumptions inherent in the derivation of the models are likely to affect their applicability. We observed that the water sample needed to be fairly vigorously stirred to agree with results from the model, which assumes well-mixed conditions. There were inconsistences between the model results and previous measurements made in a wind tunnel in 1951. The model assumes non-turbulent incident flow and consequently predicts an approximately square-root dependence on airflow speed. The wind tunnel measurements, taken over a wide range of airflows, showed a much stronger dependence. In the presence of turbulence the heat transfer will

increase with the turbulent intensity: for measurements made on ships the incident airflow is likely to be turbulent and the intensity of the turbulence is always unknown. Taken together these uncertainties are expected to be substantial and may represent the limiting factor for the direct application of these models to adjust historical SST observations. However, both the models and the observations indicate that the most important parameter driving temperature biases in historical bucket measurements is the difference between the water temperature and the wet-bulb temperature. Solar radiation is also important, but not examined in this paper.

1. Introduction

Global average surface temperature is the primary metric used to summarise the changing climate and underpins international policy to reduce carbon emissions (Rockström *et al.,* 2009; UNFCC, 2015). It is well-understood that to quantify, mitigate, and adapt to the many impacts of climate change, a range of measures of environmental change is needed (Briggs *et al.*, 2015). However, the long observational record of surface temperature remains an indispensible indicator of climate change, and a measure of direct relevance to societal interests via temperature impacts on health, food production and economies. Moreover, the ability of climate models to reproduce observed changes enables evaluation of climate model predictions: surface temperature, covering the past *ca.* 150 years, is the longest available observational record for such assessments (IPCC, 2013). Global Surface Temperature (GST) is usually constructed from near surface air temperature over land and sea surface temperature (SST) for the ocean (Kent *et al.*, 2016). Historical SST provides a lower boundary condition for reanalyses of past dynamics of the atmospheric circulation: centennial reanalyses such as the 20th Century Reanalysis (Compo *et al.*, 2011), and ERA-20C (Poli *et al.*, 2016), provide valuable resources for climate research and understanding the impacts of weather variability and climate change on the biosphere and human societies.

The greatest source of uncertainty in the long-term evolution in global average surface temperature arises from uncertainty in the bias adjustments applied to SST (Jones, 2016). Observations of SST show characteristic biases that depend on measurement method (Kent and Taylor, 2006; Kent and Kaplan, 2006; Kennedy *et al.*, 2011; Kent *et al.*, 2016). Changes in the observing system therefore lead to changing biases in SST regionally, and over time (Kennedy, 2014).

To construct accurate climate records of SST from observations in archives such as the International Comprehensive Ocean-Atmosphere Data Set (ICOADS, Freeman *et al.*, 2016), it is necessary to estimate these biases, make adjustments, and estimate the uncertainty in those adjustments (Kennedy *et al.*, 2011; Hirahara *et al.*, 2014; Huang *et al.*, 2015). The approaches taken vary, but there is agreement that the largest biases, and the largest uncertainties in the bias adjustments, are found in observations made from ships of the temperature of seawater samples taken with buckets (Kent *et al.*, 2016). The overall bias adjustment required in historic SST datasets therefore evolves as the proportion of observations made using buckets changes over time. Errors in both the bias adjustments and our knowledge of the mix of observations materially affect estimates of decadal scale variability through the historic record. The proportion of ships making bucket observations has decreased over time with the introduction of engine room intake and hull sensor measurements. The design, and therefore thermal properties, of the buckets used have also evolved. Broadly, the evolution over time of the type of buckets used to measure SST on ships was from wooden buckets (partly insulated), to canvas (uninsulated), and then to rubber or plastic buckets (typically well insulated) (Kent *et al.*, 2010).

The most-used historical SST gridded products make these adjustments for bucket bias in two different ways (Kent *et al.*, 2016). HadISST (Rayner *et al.*, 2003), HadSST3 (Kennedy *et al.*, 2011) and COBE-SST2 (Hirahara *et al.*, 2014) construct bias adjustments from weighted climatological monthly fields of estimates of bucket bias based on a physical model (Folland and Parker, 1995). ERSSTv4 (Huang *et al.*, 2015) makes bias adjustments to all ship observations based on night-time marine air temperature (NMAT) from the HadNMAT2 dataset (Kent *et al.*, 2013).

The factors affecting bucket measurements of SST are reasonably well-known (Kent *et al.*, 2016) and have been estimated using physical models developed by Folland and Parker (1995, hereafter FP95). The FP95 models, used in HadISST, HadSST3 and COBE-SST2, simulate the evaporative, direct, and radiative heat exchanges experienced by samples of water in buckets as a function of the bucket's structural and thermal characteristics (dimensions and material) as well as the airflow around the bucket. The contribution of each term in the model is expected to vary for different bucket types, and FP95 presents two different formulations designed to estimate heat exchange from wooden and canvas buckets. The FP95 models were coded in BASIC and have been converted to FORTRAN by Kent *et al.* (in prep).

There were few measurements available to FP95 to provide supporting validation for their models. Ashford (1948) compared temperature changes of water samples in 7 different types of bucket measured in a wind tunnel at a single wind speed. One of these buckets (the Met. Office Mark II) was a canvas bucket of the same type as that represented in FP95, the others were better-insulated buckets of various designs. FP95 concluded that their model could reproduce the temperature change of the Met Office Mark II canvas bucket with reasonable accuracy. However, in order to predict the measured temperature change, FP95 adjusted their canvas model, assuming free evaporation from the base and sides only. Moreover, Ashford (1948) only reported the rate of change of water temperature in the first minute, while it may have taken historical thermometers several minutes to equilibrate (FP95). Ashford (1948) did not make measurements with a wooden bucket. Roll (1951) made measurements in a wind tunnel of the characteristics of a single bucket type, the German scoop thermometer, at a wide range of wind speeds. FP95 did not develop versions of their model based on this type of bucket.

FP95 also described a comparison of their model output with the results of measurements made at sea of temperature change of water samples in canvas buckets, and again concluded that their model showed reasonable agreement.

The amount of data available to test the canvas FP95 bucket model was limited, and there were no measurements for temperature change for wooden buckets. In this paper we therefore compare measurements made in the laboratory of heat exchange from replicas of historical wooden and canvas buckets with the output of the FP95 model. The experimental setup and the implementation of the FP95 model are described in Section 2. The measurements are compared to the model predictions in Section 3 and, with insight from these comparisons, we review the wind tunnel results presented by Ashford (1948) and Roll (1951). Section 4 discusses the results and draws conclusions about the wider applicability of our measurements and the FP95 model.

2. Materials and methods

2.1 Description of the experimental setup

The buckets used in this study (Figure 1) are replicas of the Mk II Met Office canvas bucket and a 19th century wooden bucket similar to that modelled by FP95. Their structural characteristics are listed in Table I. The two buckets are of similar size (wood: 21.8 cm average inner diameter by 17.6 cm deep (up to a set water level), wider at top than at the bottom and a volumetric capacity ~ 6.6 l; canvas: 17.8 cm inner diameter by 19.4 cm deep (up to a set water level) and volumetric capacity ~ 4.8 l). The wooden bucket is made of oak 16 mm thick reinforced around the outside by two stainless steel bands. Only the sides of the canvas bucket are canvas: the base is wooden with a metal weight inside; the top is wooden with a metal spring-closing lid; the canvas is stitched and the top and base held in place with leather bands and metal pins. The masses of the wooden and canvas buckets when wet are ~ 3.3 kg and ~ 2.9 kg respectively.

Figure 2 illustrates the experimental setup. The experiments were performed in the National Oceanography Centre (Southampton, UK) Calibration Laboratory. This is kept at a roughly constant temperature of 20˚C, but the humidity is not controlled. A precision F250 thermometer was used to measure the water temperature (t) and a Vaisala probe was used to monitor the ambient air temperature (t_a) and the relative humidity (R). Data from the probes were logged every 2-3 s (alternate readings). The water temperature probe when not in use was left in a plastic container filled with water approximately in equilibrium with the ambient air temperature. A plastic bin was used to soak the buckets (the soaking time was about 4 min), which were then hung in front of a fan with three different speed settings (Table II). The centre of the fan was positioned about 0.5 m from the bucket.

The largest uncertainty in the ambient conditions comes from the airflow around the bucket. Because the bucket was fairly close to the fan relative to the bucket dimensions, the speed was not uniform around the bucket. The airflow was measured using a WindMaster ultrasonic anemometer (Gill Instruments Ltd.) for 30 s at each of six different positions: five positions in the vertical plane where the bucket would hang (centre of the bucket position and 0.5 m above, below, left and right) and at 0.35 m upwind from the centre of the bucket location. The airflow used in the implementation of the FP95 model was that measured where the centre of the bucket would be, with uncertainty derived from the standard deviation of measurements made in these surrounding locations.

Because the FP95 models assume the water sample is stirred, the water was mixed at all times using an automatic stirrer, connected to a power generator. The wooden bucket is open at the top (Figure 1). The top of the canvas bucket is a thick wooden disc with a hole for a metal lid, This lid was pushed inside the bucket by the plastic support of the stirrer during the measurements. The edge of the lid was in the water, but this is not expected to substantially affect the heat exchange as the metal lid was attached to the wooden top, limiting heat exchange by conduction. A ‘weak stirring’ regime, characterized by a mild but noticeable stirring, was created adopting an L-shaped metal piece as the stirrer; a ‘strong stirring’ regime was also implemented, where some tape was added to produce a sail-shaped stirrer. A hanging scale with precision of 0.01 kg was used to measure the mass of the filled bucket; the water level was also set and marked for each bucket and the bucket filled up to the level indicator. Finally, (clean) fresh water was used instead of salty water. The effect of salinity on latent heat of evaporation is well-known, and the vapour pressure over saline seawater is typically reduced by 2% compared to freshwater (Zeng *et al*., 1998).

Figure 3 shows thermal pictures of a replica Mk II Met Office canvas bucket (the type used by the UK Meteorological Office in the 1930s and 1940s (Ashford, 1948)), filled with water warmer than the ambient air temperature. The bucket is unstirred and the lid is shut. It is clear that the water in the bucket is cooling over time, with the cooling proceeding faster in the area facing the fan (located to the right of the bucket in these pictures). Initially the whole of the bucket is much warmer than the environment having been soaked in the warm water before exposure to the air. The structure of the bucket (rope handle, leather bands at top and bottom, stitched seam) can just be seen as cooler than the canvas body of the bucket containing the water. After 5 minutes the body of water can clearly be seen at higher temperature than the rest of the bucket, which is now colder than ambient temperature having cooled by evaporation. These images suggest that the non-canvas parts of the bucket are insulating and probably do not contribute strongly to the heat exchange which occurs almost exclusively through the canvas walls of the bucket.

2.2 Description of the model – experimental comparison procedure

In order to test FP95 heat exchange models we measured both the time and airflow dependence of the temperature of water in the buckets and compared the results with the model predictions. The FP95 model used is the laboratory version described by Kent *et al.* (in prep.). This laboratory version is similar to the full version used by FP95, although does not differentiate between the different ambient conditions expected during hauling and on-deck phases of measurement (for more information see Kent *et al.*, in prep.). Moreover, it sets the solar term to zero, as our measurements were taken indoors and away from windows. We have also excluded the salinity effect on the estimate of saturation vapour pressure, as we used fresh water. The bucket is modelled by FP95 as a cylinder in an incident airflow that is assumed to be non-turbulent. It is further assumed that the water in the bucket is well mixed (at temperature t [˚C]) and the bucket has been immersed in the sea for long enough to reach equilibrium. Inputs to the models are the airflow around the bucket for each fan speed, the ambient air temperature and humidity as measured, and the initial air-water temperature difference. For the canvas bucket, the rate of change of temperature as modelled by FP95 can be represented as

$$\frac{dt}{d\tau}=\frac{A}{c\mu}\left\{f_r h_r\left(t_a-t\right)+f_t h_t\left(t_a-t\right)+f_e h_e\left(e_a-e\right)\right\} \qquad (1)$$

where the ambient air temperature is t_a [˚C], the ambient vapour pressure is e_a and e is the saturation vapour pressure at t [both in hPa]. The transfer coefficients are h_r for longwave radiation, h_t for direct heat transfer and h_e for evaporative heat transfer [all in W m^{-2} K^{-1}]. h_e *is* $1.7h_r$. A [m^2] represents the total surface area of the bucket. The fraction of the surface area affected by longwave heat exchange, f_r, represents the sides and base. For the direct (f_t) and evaporative heat (f_e) exchange, the fraction is the same for both components, but is allowed to vary: the sides always contribute (up to the fill level) but the contribution of the top and base may be excluded, or included as required. Each of the transfer coefficients h_t and h_e depend on wind speed and the bucket geometry: slightly different values are used for the base and sides. FP95 explore different choices of f_t and f_e: heat exchange from the base, the top and the sides; heat exchange from the sides only; and heat exchange from the sides and the base (or the top), which is the final choice for FP95 as it gave the best agreement with Ashford (1948) results. c is the specific heat capacity [J kg^{-1} K^{-1}] and m is the effective mass [kg] of the bucket. In FP95 m is the combined mass of the bucket material and the water sample. The wooden bucket model is similar to that for the canvas bucket, but the thermal forcing experienced by the outside of the bucket walls acts to conduct heat through the wooden sides and base. The open top evaporates freely, but is assumed to experience a lower airflow as the water level is below the top of the bucket. More details can be found in FP95 and Kent *et al.* (in prep.).

Although FP95 have assumed little heat exchange from the top of the canvas bucket because of the lid, our experimental setup shows that for the water sample to be properly mixed, and for the measurement to be made, it requires the lid open (pushed down), permitting heat exchange from the upper water surface. On the other hand, the thermal images shown in Figure 3 suggest that most of the contribution to the overall heat loss is from the sides of the canvas bucket. The thick wooden base is not expected to make much contribution to the heat loss. However, if the top was open (it is not in these images) then exchange of heat from the open top is expected, although the airflow within the bucket would be rather small, limiting this effect. Therefore, when implementing the FP95 canvas bucket model in this study we have run the model assuming heat exchange from either the top and sides or from the sides only, with each included in the ensemble from which the model uncertainty range is calculated. In our implementation of FP95 we assume no contribution from the bucket

226 material to the effective mass and heat capacity of the canvas bucket. This seems justified by
227 Figure 3, as the images show that the temperature change largely affects only the water
228 sample: the non-canvas parts of the bucket quickly reach ambient temperature, suggesting that
229 the temperature change for the wooden and leather parts is superficial. The choice of the
230 effective mass, which sets the heat capacity, will scale the temperature change but will not
231 affect its functional dependence.

232 The models were initialized with measured ambient conditions (summarised in Table A1 in
233 the Appendix) and the appropriate bucket dimensions (Table I, other bucket properties are set
234 by the choice of the wooden or canvas model). The probe used to measure the water
235 temperature has a finite response time and typically took between 30 seconds and 1 minute to
236 reach equilibrium, less when the air and water temperatures were similar. Each experiment
237 was considered to start when the recorded water temperature reached a local maximum or
238 minimum (depending on whether the water was warmer or colder than the air). Uncertainty in
239 the equilibration temperature was estimated to be around 0.01˚C (much smaller than, for
240 example, the variation in the air temperature over each experiment) so the estimated
241 uncertainty is not sensitive to the value chosen. For each experiment, the uncertainties in the
242 model outcomes were expressed as an ensemble of 100 realisations. Each realisation was
243 randomly generated by forcing the model with samples of the measured ambient air
244 temperature, relative humidity, wind speed (mean and standard deviation as measured), of the
245 water temperature at time = 0 min (mean as measured, standard deviation of 0.01˚C) and of
246 the bucket diameter and water level. For the canvas bucket model the uncertainty in the
247 bucket geometry (mean as measured, standard deviation of 0.5 cm) was included to account
248 for the small variations in the initial mass of the water sample (for both buckets the standard
249 deviation over all the measurements of the mass of the water sample was about 0.05 kg). For

the canvas bucket, the uncertain contribution of evaporation from the bucket top is also included in the overall model uncertainty. For the wooden bucket model, the uncertainty in the geometry of the bucket mainly arises because of the uncertainty in the contribution of the bucket mass to the heat exchange of the water sample. In order to include also this component, the uncertainties in the bucket radius and water level were included in each realization (mean as measured at the half point of the bucket walls, standard deviation equal to the thickness of the bucket walls). For the wooden bucket, the uncertainty in the factor that accounts for the sheltering of water by the sides of the bucket from the effects of airflow (Kent *et al*., in prep.) is also included (to reproduce the range assumed by FP95 we assumed a mean of 0.875 and a standard deviation of 0.125 with upper limit of 1). The bucket was fairly full, the water level was about 2 cm below the top, so only a modest sheltering of the airflow would be expected. Finally, for the wooden bucket model, the uncertain thermal conductivity of wet oak is also included in the overall model uncertainty (mean of 0.3 W m^{-1} C^{-1} as assumed in FP95, standard deviation of 0.2 W m^{-1} C^{-1} with upper and lower limit defined by the thermal conductivity of dry oak (0.17 W m^{-1} C^{-1}) and water (0.6 W m^{-1} C^{-1}) respectively).

Leakage, determined by the change in mass, was largest for the canvas bucket, and decreased over time (0 - 3 minutes: ~ 0.05 kg min^{-1}, 4 - 20 minutes: ~ 0.04 kg min^{-1} and 20 minutes onwards ~ 0.03 kg min^{-1}). No significant leakage was measured for the wooden bucket. We included the changing mass in the canvas bucket model, but, as noted by Kent *et al.* (in prep.), the leakage makes very little difference as decreases in the surface area subject to heat exchange affect a decreasing volume of water, with little overall effect as long as the bucket remains fairly full.

Firstly, the evolution of the bucket temperature over time was measured for a set of experiments varying the temperature of the water in the plastic bin used for soaking the

274 bucket and from which the water sample is taken. The experiments were performed using the
275 two different stirring regimes ('strong' and 'weak') to test how different mixing conditions
276 may affect the heat exchange from the water sample (*dt*1 to *dt*3 in Table II). For each bucket,
277 the water temperature was measured for 15 min for three air-water temperature regimes and
278 each of these measurements was repeated three times. In the first set of experiments (*dt*1) the
279 initial water temperature (t_0) was warmer than the air temperature (t_a): t_0 - t_a ~ 5 ˚C. The
280 second set (*dt*2) has t_0 slightly colder than t_a: t_0 - t_a ~ -1 ˚C. In the third set (*dt*3) the water
281 temperature was colder again: t_0 - t_a ~ -5 ˚C. The fan was at its fastest setting, about 3.5 m s^{-1},
282 7 knots (u_3 see Table II), for all six experiments (three temperatures and two stirring
283 regimes).

284 Secondly, we measured the water temperature for 15 min for each of the four available
285 different airflows (*u*0 through to *u*4 in Table II) for an initial warm-water bucket temperature
286 difference of t_0 - t_a ~ 5˚C and under the strong stirring regime. Again, each set of
287 measurements was repeated three times.

288 **3. Results and discussion**

289 In this section we describe the results of the comparison of temperature change measured in
290 the laboratory and predicted by the models (3.1) for different degrees of mixing of the water
291 sample (3.1.1 and 3.1.2) and for different airflows (3.1.3). Also we present here the results of
292 the comparison with historical measurements in wind tunnels (3.2) made by Ashford (1948)
293 and by Roll (1951).

294 *3.1 Comparison of temperature change measured in the laboratory and predicted by the*
295 *models.*

296 3.1.1 Evolution of water temperature under strong stirring

Figure 4 shows the measured and modelled water temperature as a function of time for both the wooden and canvas buckets, for the range of three different initial water temperatures and also for the strong and weak stirring regimes (R Development Core Team, 2016). When the initial water temperature is warmer than ambient air temperature (set of experiments *dt*1) the water is cooled both directly and by evaporation. When the initial water temperature is slightly colder than the air temperature (set of experiments *dt*2) the water is warmed directly and cooled by evaporation. For these conditions the evaporation dominates and the water sample cools. When the water is significantly colder than ambient air (set of experiments *dt*3), the water is again being warmed directly and cooled by evaporation, this time with a net warming overall. As expected, the canvas bucket cools much more rapidly than the wooden bucket, despite their similar volumes. This feature is well reproduced in the model simulations. For both buckets the contribution of the uncertainty in the airflow (fan) speed explains a large portion of the overall model uncertainty: this is shown for each air-temperature regime by the error bars on the right of the plot, which represent the 95% confidence level uncertainty at time = 15 min computed from the ensemble generated accounting for the wind uncertainty only. For the canvas bucket, the remaining uncertainty is mostly due to the variations in the ambient relative humidity and air temperature; on the other hand, for the wooden bucket the biggest contribution to the remaining model uncertainty is represented by the uncertainty in the thermal conductivity of the bucket walls. The model estimates for the wooden bucket underestimate the observed temperature change for the strong stirring regime (Figure 4*a*), although the experimental results are close to the limits of the estimated model uncertainty. However, the rate of temperature change increases over the first few minutes of the 15-minute sampling period in both the measurements and the model (shown for the model in the inset in Figure 4*a*). A simple picture of temperature change would show a decreasing rate of temperature change over time as the water sample approaches equilibrium with its surroundings (as seen for the canvas bucket in Figure 4*b*). The model reproduces the measured behaviour well, and shows that the initial slow rate of temperature change is caused by the timescale for the conduction of heat through the walls of the wooden bucket. The water inside the bucket does not respond to the thermal forcing on the outside of the bucket until the temperature gradient within the bucket walls is established: once this occurs the temperature change of the water increases. In the 15-minute sampling period this effect dominates over the reduction in thermal forcing over time as the bucket sample reaches its equilibrium temperature.

In contrast the canvas bucket with strong stirring (Figure 4*b*) shows the expected decrease in the rate of temperature change over time, as already noted, and again the measurements and the model show the same general behaviour, with the modelled and measured temperature change agreeing at the 95% confidence level, although close to the limit of the estimated uncertainty in our experimental setup. As noted by FP95 and Farmer *et al.* (1989) the temperature in the canvas bucket will eventually asymptotically reach an “effective wet-bulb temperature” when the evaporative cooling is balanced by the warming from the atmosphere (Folland, 1991).

3.1.2 Evolution of water temperature under weak stirring

The effects of weaker stirring are explored in Figures 4*c* and 4*d.* If the water is not well-mixed the largest temperature changes will be expected near the water surface and the bucket walls. The temperature is measured in the centre of the bucket where a smaller temperature change would be expected, and this is what is observed. The observed temperature change under weak stirring is lower than under strong stirring and the measured temperature change for the wooden bucket remains in agreement with the model predictions under both low (set of experiments *dt*2 and *dt*3) and high thermal forcing (set of experiments *dt*1: warm initial water temperature), although the model assumes well-mixed conditions. The time evolution of the temperature change is unsteady compared with the better-mixed case (compare Figure 4*a* and 4*c*). For the canvas bucket, the difference due to reduced stirring is particularly noticeable for the high forcing case (*dt*1): here, an initial lower rate of temperature change is very obvious, similar to that observed for the wooden bucket and predicted by the wooden bucket model. This can again be explained by an initial setting up of temperature gradients in the water, in a similar way to the gradients established in the wooden bucket walls. As for the wooden bucket measurements, the weak stirring temperature change is unsteady.

3.1.3 Effect of airflow

The model heat exchange coefficients h_t and h_e depend approximately on the square root of the airflow, since the incident flow is assumed to be non-turbulent. Figure 5 shows the observed bucket temperature (grey dots) at time = 5 min for the various air (fan) speeds and the values predicted by the model (shading) for the wooden (Figure 5*a*) and the canvas (5*b*) bucket for water ~ 5 ˚C warmer than air temperature. When the fan was turned on (*u*1- *u*3), for each bucket, the observed dependence on airflow is similar to that assumed in the model, although for the wooden bucket the observed temperature change is either close to or, for some experiments, lies outside the limits of the estimated uncertainty range, as in Figure 4*a* for *dt*1. On the other hand, when the fan was turned off (*u*0), for both buckets the modelled and the observed temperature change do not agree within the range of the estimated uncertainty (Figure 5). FP95 models assume a Reynolds number always larger than one: this means that the situation when there is no airflow around the bucket is very uncertain but the temperature change will be small in these conditions. Finally, our experimental setup means that we cannot increase the speed of the airflow around the bucket beyond ~ 3.5 m s^{-1}, and the uncertainty in the speed is large.

3.2 Comparison with historical measurements in wind tunnels (Ashford 1948, Roll 1951).

3.2.1 Ashford (1948)

Measurements in a stronger airflow regime, about 9 m s^{-1}, were made by Ashford (1948, hereafter Ashford) for 7 different buckets. The results were presented as the rate of change of water temperature in the first minute plotted as a function of the water temperature *minus* wet-bulb temperature (Δt_{wb}). Plotted in this way buckets that evaporate strongly will show a curved relationship of temperature change with Δt_{wb} due to the Clausius-Claperyon relationship. When the water temperature is varied at the same ambient air temperature, as is the case for all the measurements we consider here, the wet-bulb temperature will be constant and e_a and t_a (Equation 1) are also constant. Δt_{wb} therefore varies linearly with variations in t, as does the direct heat exchange. However variations in e are non-linear and the relationship between temperature change and Δt_{wb} will be non-linear if the effects of evaporation are important. In contrast, buckets where the direct heat exchange dominates over evaporation, or under conditions where the air is close to saturation, will show a close to linear relationship when plotted in this way. Figure 6 shows measured values (from runs *dt*1 to *dt*3 in Table II) obtained with strong stirring as a function of Δt_{wb} (with wet-bulb temperature computed following the approach of Stull, 2011). The change of water temperature over the first minute exhibits different characteristic relationships with the Δt_{wb} according to the bucket thermal capacity. Both the wooden bucket (Figure 6*a*) and the uninsulated canvas bucket (Figure 6*b*) are characterized by a non-linear relationship in the model, because of evaporation (through the top for the wooden bucket and through the sides for the canvas bucket). In the first minute the measurements are nosier than our estimates of uncertainty, especially for the wooden bucket (Figure 6*a*). The measurements are fairly consistent with the model results for each bucket type but the non-linear relationship cannot be confirmed because of the noise. Also plotted in Figure 6*b* are the results from measurements with the same type of bucket by Ashford. The increased temperature change in the Ashford results is modest, despite the much greater airflow (~ 9 m s^{-1} *cf.* ~ 3.5 m s^{-1}), and the measurements agree well with the model. These results extend the range of airflows over which the canvas bucket model has been tested, and suggest that the wind speed dependence in these experiments is reasonably predicted by the model. We note that the Ashford measurements for the canvas bucket were used by FP95 as validation, but that here we have assumed a smaller heat capacity for the bucket (by excluding the contribution of the bucket itself, based on Figure 3 as discussed in Section 2.1). However, the modelled rate of temperature change at one minute as a function of Δt_{wb} for each of these different choices of the effective mass remains consistent with the Ashford measurements under either assumption.

The results presented by Ashford allow a comparison of the characteristics of a range of different bucket types and Figure 7 shows a selection of measurements reproduced from his Figure 2. Two types of bucket showed much greater temperature changes than the others: the canvas bucket as tested in the present study (Met Office Mk II) and the German scoop thermometer. A modern version of the German scoop is shown in Figure 1. The version tested by Roll (1951), and Ashford, is likely to be similar to this modern bucket. The capacity of the scoop is small (Table I) and it is mostly made of metal. A rubber buffer with an air cushion covers the sides. Older versions had a leather cover with felt filling, but we do not know which type was used by either Ashford or Roll (1951). The base is double-walled with cork insulation between. An integral thermometer, mechanically isolated to avoid breakage during use, means that the reading can be made immediately after hauling. Also plotted in Figure 7 are results from the new bucket design in versions with, and without, a lid. These new buckets were designed to minimise temperature change and show much lower rates of temperature change for a given water - wet-bulb temperature difference. Ashford describes the new bucket as canvas, but it has a copper vessel inside, which makes it partially insulated. The curvature of the lines becomes much less apparent for these buckets that show progressively smaller temperature change. This would be expected if the new designs were particularly effective at reducing heat loss by evaporation. Ashford reports that the temperature change was little affected if the outside of the bucket was wet or dry (note the Mk II canvas bucket cannot be kept dry). However, it may be that the curvature is simply not visible over the noise in the measurements for buckets with small rates of temperature change.

3.2.2 Roll (1951)

The German scoop thermometer was also studied in a wind tunnel at a range of wind speeds by Roll (1951, hereafter Roll). The measurements of temperature change after 1 minute ($\Delta t|_{1min}$) are presented in terms of a wind speed-dependent coefficient (β) and an equivalent air (θ_a) and water temperature (θ_b):

$$\Delta t\big|_{1\,\mathrm{min}} = \beta\left(\theta_a - \theta_b\right) \tag{2}$$

θ is defined following Rössler (1948):

$$\theta = t + \alpha\frac{e}{p} \tag{3}$$

to give:

$$\Delta t\big|_{1\,\mathrm{min}} = \beta\left[\left(t_a - t\right) + \frac{\alpha}{p}\left(e_a - e\right)\right] \tag{4}$$

where p is the atmospheric pressure [hPa] and α = 1560 [K]. Equation (4) is of similar form to Equation (1), if the small term for longwave radiation is neglected in the latter. We can then interpret the term β as a heat transfer coefficient. Figure 8*a* (measurements read from hand drawn Figure 2 in Roll) shows the wind speed dependence of β for the scoop, from 2 m s^{-1} to 19 m s^{-1}. Roll's results show a much stronger airflow dependence of β, (a power greater than 1), than that shown by the FP95 model (an approximate square-root dependence) which is tentatively confirmed for the canvas bucket by our measurements and those of Ashford (Figure 6*b*). Either an FP95-type model is not appropriate for the interpretation of Roll's measurements, or these measurements taken at higher wind speeds are indicating a stronger airflow dependence than the model, and also the canvas bucket measurements (both those of Ashford and our laboratory measurements).

The time evolution of the water temperature measured by Roll over the first 10 minutes is shown in Figure 8*b* for each of 8 different wind speeds and an air-water temperature difference of -10°C. The values plotted were read from Figure 1 in Roll: the original graph consists of hand drawn lines. Unfortunately Roll does not provide much information about the conditions under which the measurements were made. A small increase in the rate of temperature change over time is apparent at lower airflow speeds (2-8 m s^{-1}), as was seen with the wooden bucket, which might indicate that the behaviour of the scoop is comparable to the wooden bucket. There also seems to be some separation between the measurements taken at lower airflow speeds and those at higher speeds (10-19 m s^{-1}), which might indicate that conditions had changed over the course of the experiment.

One explanation for the stronger wind speed dependence might have been due to β having been estimated from measurements taken after one minute. At the start of exposure to the atmosphere, partly-insulated buckets take time to establish temperature gradients within the bucket walls (see Figure 4*a*) and if the timescale for this process depends on the airflow, which seems reasonable, then aliasing of this signal might cause an apparent increase in β with airflow. This was investigated (noting that the ambient environmental conditions are uncertain) but β as estimated from Figure 8*b* shows strong wind speed dependence throughout the first 10 minutes.

Despite the uncertainties around the Roll measurements it seems clear that the airflow dependence of temperature change measured in the wind tunnel is greater than that predicted by the FP95 model, which predicts an approximate square-root dependence (Kent *et al.,* in prep.). Whilst the dimensions, design and thermal properties of the scoop are rather different than those of the wooden bucket, all of these differences could be accounted for, and the wind speed functional dependence would remain similar.

FP95 assume that the incident flow is laminar. They note that turbulence in the incident flow would increase the heat transfer coefficient, and further note that turbulent incident flow was likely for measurements made on a ship. It is also likely for our measurements in the lab, and for the two sets of wind tunnel results, but the intensity of turbulence for each of these sets of measurements is unknown. At the higher wind speeds measured by Roll the incident flow would certainly have been turbulent and his stronger speed dependence could potentially be explained by an increasing intensity of turbulence with wind speed giving an increased heat transfer coefficient (Lowery and Vachon, 1975). This means that comparing measurements made in different wind tunnels, and even at different flow speeds within the same wind tunnel is difficult, and will reduce the confidence with which any derived heat exchange characteristics can be applied to measurements at sea where the intensity of the turbulence is always unknown.

4. Summary and Conclusions

Tests in the laboratory show that the FP95-type models used to estimate the biases in bucket-derived SST measurements work well, when conditions are similar to those assumed in the models. At the range of airflows tested (a maximum of ~ 3.5 m s^{-1}), the model for the canvas bucket predicted a temperature change within the estimated experimental uncertainty for a range of air-water temperature differences (Figure 4) and airflow speeds (Figure 5). For the wooden bucket, although close to the limit of the estimated uncertainty, the model slightly underestimates the observed temperature change. We conclude that the models are able to reasonably reproduce the temperature change measured for the two buckets. The model simulations helped us to understand an observed initial period of reduced temperature change for the wooden buckets (Figure 4*a*). This was caused by the time taken for heat to be conducted through the bucket walls, an effect included in the wooden bucket model. However, the assumptions made in the model derivation may in practice be rather limiting. Our measurements showed that if the sample is not vigorously stirred, then the temperature change will be much lower than when the water is well-mixed, particularly when the rate of temperature change is large. This was particularly obvious for the canvas bucket filled with water substantially warmer than the ambient air temperature (Figure 4*d*). However, the assumptions made in the model derivation may in practice be rather limiting. Our measurements showed that if the sample is not vigorously stirred, then the temperature change will be lower than for well-mixed conditions as assumed by the models, particularly when the rate of temperature change is large. This was particularly obvious for the canvas bucket filled with water substantially warmer than the ambient air temperature (Figure 4*d*). We reviewed the results of some previous measurements of temperature change for a range of different bucket types taken in wind tunnels (Ashford 1948; Roll 1951). Ashford made measurements using the same canvas bucket used in this study, but at a substantially higher airflow speed (~ 9 m s^{-1}). The temperature change for a given thermal forcing (defined as the water temperature *minus* wet-bulb temperature, Δt_{wb}) was only slightly larger than that measured in the lab at ~ 3.5 m s^{-1} (Figure 5*b*), suggesting that the approximate square-root dependence of the heat transfer on airflow speed used by FP95 was reasonable. This modest airflow dependence was however not supported by the results of Roll, who made measurements for a single bucket type (the German scoop, Figure 1) at a wide range of wind speeds. Roll's results showed a much larger increase in heat transfer with airflow, a dependence stronger than linear. A possible reason for such inconsistency in the airflow dependence of heat exchange was suggested by FP95: they note that any turbulence in the incident flow will act to increase their heat exchange coefficient. The strong increase in temperature change observed by Roll with increasing airflow could reasonably be explained by an increase in the turbulent intensity of the incident flow with airflow. This explanation however leads to the problematic conclusion that any estimates of heat transfer coefficients will be affected by the particular circumstances of the experimental, or shipboard, conditions.

Ashford took measurements of temperature change in a range of different bucket types. His results clearly showed a wide range of different heat exchange characteristics (Figure 7), as did our measurements for wooden and canvas buckets (Figure 6). The heat exchange characteristics are broadly predictable for each bucket type and depend on the geometry, size and degree of insulation.

The FP95 formulation is fairly straightforward to adapt for different bucket types. The cylindrical bucket geometry can be specified, as can the degree of insulation of the bucket walls. Modern buckets for which the outer surface would not remain wet could be modelled by setting $f_e < f_t$ in Equation 1. The heat transfer coefficients h_e and h_t are formulated based on a Nusselt number. There are empirical formulations for the Nusselt number that are likely to be applicable in a wider range of conditions (e.g. Churchill and Bernstein, 1977). However the problem of unknown intensity of turbulence in the incident flow, and how that turbulence might depend on local obstacles for any particular measurement, remains. Despite this, the models might be expected to be effective at estimating the relative rates of temperature change for different types of bucket.

We need to consider the impact of our conclusions on the FP95-derived bias adjustments used in HadISST, HadSST3 and COBE-SST2. FP95 were well aware of the difficulties associated with quantifying biases in historical SSTs and attempted to design their bias adjustment methodology to be robust to the uncertainties they identified. FP95 conclude that their bias adjustment fields are "fairly insensitive to uncertainties such as the size of the bucket or the details of its exposure on deck". This is because the parameters assumed to be characterized by the largest uncertainty in the model (i.e. the mix of bucket types and the assumed exposure time for uninsulated canvas buckets) are estimated such that the internal consistency of the observations is improved. The mix of bucket types (wooden or canvas) is calculated to improve the agreement between the adjusted SST and NMAT anomalies in the Tropics (FP95) and the exposure time for canvas buckets is adjusted to give more similar seasonal cycles before and after World War 2. The resulting adjustment fields are only weakly dependent on the highly uncertain airflow around the bucket, and show a much stronger dependence on the water temperature *minus* wet-bulb temperature (Kent *et al.,* in prep.). Constraining the uncertain parameters in FP95 models to improve the internal consistency of the data leads to reasonable large-scale estimates of the biases in historical SST bucket observations (Kent *et al*., 2016).

We conclude therefore that new measurements of temperature change of water samples in buckets made onboard ships at sea would be more valuable than additional measurements made, for example, in wind tunnels. However, it would be challenging to make enough measurements with different types of buckets, in different environmental conditions, and in differently-exposed locations on different types of ships to fully explore the dependencies. None of the measurements discussed in this paper consider the effects of solar radiation, but we note that the effect of solar radiation on bucket measurements made at sea is detectable and can be used to distinguish between observations made using buckets and those from other methods such as engine-room intakes (Carella *et al*., in prep).

A good approach to estimating bias adjustments for historical bucket measurements would be to directly estimate the adjustments from the observations themselves, guided by the dependencies shown by the physically-based models. From our results, and those of Ashford, Roll, and Kent *et al.* (in prep.) we conclude that the adjustments are likely to be strongly dependent on (Δt_{wb}), as are the FP95-derived fields used by HadISST, HadSST3 and COBE-SST2. The relationship between temperature change and Δt_{wb} will be scaled depending on bucket type and will vary with measurement protocols (relating to the way the measurement was made - including how quickly - and whether the bucket was sheltered from the sun or the wind and whether the sample was well-mixed). On a secondary level, the temperature change will also depend on ambient conditions not related to Δt_{wb} (including airflow speed, the intensity of turbulence in incident flow, and solar radiation). Such approaches have not been explored in the past but are now possible because of a much increased number of observations (Freeman *et al*., 2016), improved metadata (Carella *et al*., 2015) and increased computer capacity.

Acknowledgments

The canvas bucket was manufactured from the original Met Office engineering drawings by W. G. Lucas & Son Ltd. The authors wish to acknowledge use of the Ferret program for some graphics in this paper. Ferret is a product of NOAA's Pacific Marine Environmental Laboratory. (Information is available at http://ferret.pmel.noaa.gov/Ferret/)

Ashford OM. 1948. A new bucket for measurement of sea surface temperature. *Q. J. R. Meteorol. Soc*. 74: 99–104. DOI: 10.1002/qj.49707431916

Briggs S, Kennel CF, Victor DG. 2015. Planetary vital signs. *Nature Climate Change* 11: 969-70. DOI: 10.1038/nclimate2828

Carella G, Kent EC, Berry DI. 2015. A probabilistic approach to ship voyage reconstruction in ICOADS. *Int. J. Climatol*.. DOI:10.1002/joc.4492

Carella G, Kent EC, Berry DI, Merchant CJ, Morak-Bozzo S. Estimating Sea Surface Temperature measurement methods from characteristic differences in the diurnal cycle. *Geophys. Res. Lett*., in preparation

Churchill SW, Bernstein M. 1977. A Correlating Equation for Forced Convection From Gases and Liquids to a Circular Cylinder in Crossflow. *J. Heat Transfer, Trans. ASME* 99: 300 – 306. DOI: 10.1115/1.3450685

Compo GP, Whitaker JS, Sardeshmukh PD, Matsui N, Allan RJ, Yin X, Gleason BE, Vose RS, Rutledge G, Bessemoulin P, Brönnimann S, Brunet M, Crouthamel RI, Grant AN, Groisman PY, Jones PD, Kruk MC, Kruger AC, Marshall GJ, Maugeri M, Mok HY, Nordli Ø, Ross TF, Trigo RM, Wang XL, Woodruff SD, Worley SJ. 2011. The Twentieth Century Reanalysis Project. *Quart. J. Roy. Meteor. Soc.* 137: 1-28. DOI: 10.1002/qj.776

Folland CK. 1991. Sea temperature bucket models used to correct historical SST data in the Meteorological Office. Climate Research Technical Note No 14. Unpublished, available from: http://www.metoffice.gov.uk/hadobs/hadsst3/references/CRTN14_Folland1991.pdf

Folland CK, Parker DE. 1995. Correction of instrumental biases in historical sea-surface temperature data. *Q. J. R. Meteorol. Soc.* 121: 319–367. DOI: 10.1002/qj.49712152206

Freeman E, Woodruff SD, Worley SJ, Lubker, SJ, Kent EC, Angel WE, Berry DI, Brohan P, Eastman R, Gates L, Gloeden W, Ji Z, Lawrimore J, Rayner NA, Rosenhagen G, Smith SR. 2016. ICOADS Release 3.0: a major update to the historical marine climate record. *Int. J. Climatol*.. DOI: 10.1002/joc.4775

Hirahara S, Ishii M, Fukuda Y. 2014. Centennial-scale sea surface temperature analysis and its uncertainty. *J. Climate* 27: 57-75. DOI: 10.1175/JCLI-D-12-00837.1

Huang B, Banzon VF, Freeman E, Lawrimore J, Liu W, Peterson TC, Smith TM, Thorne PW, Woodruff SD, Zhang H-M. 2015. Extended Reconstructed Sea Surface Temperature Version 4 (ERSST.v4). Part I: Upgrades and Intercomparisons. *J. Climate* 28: 911–930. DOI: 10.1175/JCLI-D-14-00006.1

IPCC. 2013. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V and Midgley PM (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp. DOI: 10.1017/CBO9781107415324

Jones P. 2016. The Reliability of Global and Hemispheric Surface Temperature Records. *Adv. Atmos. Sci.* 33 (3): 269–282. DOI: 10.1007/s00376-015-5194-4

Kennedy JJ, Rayner NA, Smith RO, Parker DE, Saunby M. 2011. Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization, *J. Geophys. Res*. 116: D14104. DOI: 10.1029/2010JD015220.

Kennedy JJ. 2014. A review of uncertainty in in situ measurements and data sets of sea surface temperature. *Rev. Geophys.* 52:1–32. DOI: 10.1002/2013RG000434

Kent EC, Kaplan A. 2006. Toward Estimating Climatic Trends in SST, Part 3: Systematic Biases. *J. Atmos. Ocean. Tech*. 23 (3): 487-500. DOI: 10.1175/JTECH1845.1

Kent EC, Taylor PK. 2006. Toward Estimating Climatic Trends in SST, Part 1: Methods of Measurement. *J. Atmos. Ocean. Tech.* 23 (3): 464-475. DOI: 10.1175/JTECH1843.1

Kent EC, Kennedy JJ, Berry DI, Smith RO. 2010. Effects of instrumentation changes on ocean surface temperature measured in situ. *Wiley Interdisciplinary Reviews: Climate Change* 1 (5): 718-728. DOI: 10.1002/wcc.55

Kent EC, Rayner NA, Berry DI, Saunby M, Moat BI, Kennedy JJ, Parker DE. 2013. Global analysis of night marine air temperature and its uncertainty since 1880: the HadNMAT2 Dataset. *Journal of Geophysical Research* 118 (3): 1281-1298. DOI:10.1002/jgrd.50152

Kent EC, Kennedy JJ, Smith TM, Hirahara S, Huang B, Kaplan A, Parker DE, Atkinson CP, Berry DI, Carella G, Fukuda Y, Ishii M, Jones PD, Lindgren F, Merchant CJ, Morak-Bozzo S, Rayner NA, Venema V, Yasui S, Zhang H-M. 2016. A call for new approaches to quantifying biases in observations of sea-surface temperature, *Bull. Am. Meteorol. Soc.*. DOI: 10.1175/BAMS-D-15-00251.1

Kent EC, Berry DI, Carella G, Merchant CJ, Chiu JC. Models for the adjustment of observations of sea surface temperature made using buckets. *J.Tech.*, in preparation

Lowery GW, Vachon RI. 1975. The effect of turbulence on heat transfer from heated cylinders. *International Journal of Heat and Mass Transfer* 18 (11): 1229-1242. DOI: 10.1016/0017-9310(75)90231-8

Poli P, Hersbach H, Dee DP, Berrisford P, Simmons AJ, Vitart F, Laloyaux P, Tan DG, Peubey C, Thépaut JN, Trémolet Y. 2016. ERA-20C: An Atmospheric Reanalysis of the Twentieth Century. *Journal of Climate* 29 (11): 4083-4097. DOI: 10.1175/JCLI-D-15-0556.1

R Development Core Team. 2016. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*, Vienna, Austria.

Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A. 2003. Global Analyses of SST, Sea Ice and Night Marine Air Temperature Since the Late 19th Century. *Journal of Geophysical Research* 108 (D14): 4407. DOI: 10.1029/2002JD002670

Rockström J, Steffen W, Noone K, Persson Å, Chapin FS, Lambin EF, Lenton TM, Scheffer M, Folke C, Schellnhuber HJ, Nykvist B. de Wit CA, Hughes T, van der Leeuw S, Rodhe H, Sörlin S, Snyder PK, Costanza R, Svedin U, Falkenmark M, Karlberg L, Corell RW, Fabry VJ, Hansen J, Walker B, Liverman D, Richardson K, Crutzen P, Foley JA. 2009. A safe operating space for humanity. *Nature* 461 (7263): 472-5. DOI: 10.1038/461472a

Roll HU. 1951. Water temperature measurements on deck and in the engine room. *Ann. Meteor*. 4: 439–443

Rössler F. 1948. Wärmeübergan an nassen Oberflächen. Die Naturwissenschaften, 35 (7): 219-220

Stull R. 2011. Wet-bulb temperature from relative humidity and air temperature. *Journal of Applied Meteorology and Climatology* 50 (11): 2267–2269. DOI: 10.1175/JAMC-D-11-0143.1

UNFCCC. 2015. *Adoption of the Paris Agreement*. Report No. FCCC/CP/2015/L.9/Rev.1, http://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf

Zeng X, Zhao M, Dickinson R. 1998. Intercomparison of Bulk Aerodynamic Algorithms for the Computation of Sea Surface Fluxes Using TOGA COARE and TAO Data. *J. Climate* 11: 2628–2644. DOI: 10.1175/1520-0442(1998)011<2628:IOBAAF>2.0.CO;2

Tables

Bucket type	Bucket material	Water level [m]	Diameter (inner) [m]	Thickness [mm]	Bucket volume [l]	Bucket mass (wet & empty) [kg]
Wooden	Oak	0.176	0.218	16	6.6	3.30
Mk II canvas	Mixed	0.194	0.178	-	4.8	2.91
German scoop	Mixed	0.116	0.097	13	0.9	3.35

Table I. Structural characteristics of the buckets discussed in this study.

Experiment type	Bucket type	Stirring type	$t_0 - t_a$ [˚C]	Fan speed [m s^{-1}]
dt 1	Wooden & Canvas	Strong &Weak	~ 5	3.53 ± 0.49
dt 2	Wooden & Canvas	Strong & Weak	~ -1	3.53 ± 0.49
dt 3	Wooden & Canvas	Strong & Weak	~ -5	3.53 ± 0.49
u 0	Wooden & Canvas	Strong	~ 5	0.05 ± 0.02
u 1	Wooden & Canvas	Strong	~ 5	2.17 ± 0.37
u 2	Wooden & Canvas	Strong	~ 5	2.88 ± 0.44
u 3	Wooden & Canvas	Strong	~ 5	3.53 ± 0.49

Table II. Summary of each experiment. The table reports the bucket type, the stirring type, the approximate water temperature at time = 0 min *minus* the ambient air temperature $t_0 - t_a$ [˚C] and the fan speed [m s^{-1}] for each experiment. The uncertainty in the fan speed is reported at one standard deviation. For model simulations when the fan was turned off (*u*0) a small airflow speed (0.05 m s^{-1}) was assumed to be consistent with the measured fluctuations (standard deviation 0.02 m s^{-1}). For an extended summary see Table A1 in the Appendix.

Appendix

Experiment type	Bucket type	Stirring type	R [%]	t_a [˚C]	t_0 [˚C]	$t_0 - t_a$ [˚C]
dt 1	Wooden	Strong	60.1 ± 0.1	20.77 ± 0.07	25.09 ± 0.20	4.32 ± 0.21
dt 2	Wooden	Strong	71.9 ± 0.2	20.95 ± 0.07	19.38 ± 0.04	-1.57 ± 0.08
dt 3	Wooden	Strong	71.1 ± 2.1	21.01 ± 0.15	15.63 ± 0.13	-5.38 ± 0.20
dt 1	Wooden	Weak	59.0 ± 0.9	20.85 ± 0.13	25.27 ± 0.14	4.42 ± 0.20
dt 2	Wooden	Weak	72.0 ± 0.9	21.14 ± 0.25	19.39 ± 0.00	-1.75 ± 0.25
dt 3	Wooden	Weak	72.7 ± 0.3	21.04 ± 0.07	15.69 ± 0.15	-5.35 ± 0.17
u 0	Wooden	Strong	56.9 ± 0.4	22.15 ± 0.15	25.04 ± 0.09	2.89 ± 0.18
u 1	Wooden	Strong	61.7 ± 0.9	21.07 ± 0.18	25.04 ± 0.05	3.97 ± 0.19
u 2	Wooden	Strong	70.9 ± 0.3	20.65 ± 0.12	25.05 ± 0.07	4.40 ± 0.14
u 3	Wooden	Strong	60.1 ± 0.1	20.77 ± 0.07	25.09 ± 0.20	4.32 ± 0.21
dt 1	Canvas	Strong	69.5 ± 0.9	20.96 ± 0.17	25.00 ± 0.13	4.04 ± 0.22
dt 2	Canvas	Strong	64.3 ± 0.6	20.70 ± 0.08	19.27 ± 0.10	-1.43 ± 0.13
dt 3	Canvas	Strong	63.5 ± 0.6	20.93 ± 0.21	15.67 ± 0.07	-5.26 ± 0.22
dt 1	Canvas	Weak	60.2 ± 0.0	20.75 ± 0.07	25.01 ± 0.05	4.26 ± 0.09
dt 2	Canvas	Weak	64.9 ± 0.4	20.77 ± 0.22	19.34 ± 0.07	-1.43 ± 0.23
dt 3	Canvas	Weak	64.2 ± 0.6	20.84 ± 0.05	15.56 ± 0.14	-5.28 ± 0.15
u 0	Canvas	Strong	68.0 ± 1.3	22.10 ± 0.12	25.08 ± 0.11	2.98 ± 0.17
u 1	Canvas	Strong	70.6 ± 0.6	21.31 ± 0.21	24.83 ± 0.11	3.52 ± 0.24
u 2	Canvas	Strong	71.1 ± 0.3	20.93 ± 0.08	24.95 ± 0.09	4.02 ± 0.12
u 3	Canvas	Strong	69.5 ± 0.9	20.96 ± 0.17	25.00 ± 0.13	4.04 ± 0.22

Table A1. Extended summary of each experiment. R: relative humidity [%], t_a: ambient air temperature, t_0: water temperature at time = 0 min [˚C]. The experiment corresponding to each row in the table was repeated three times and was run for 15 min: the relative humidity and the ambient air temperature represents the mean over 15 min and all the repetitions; the water temperature at time = 0 represents the mean over all the repetitions. The uncertainty in each variable is reported at one standard deviation.

696 **Figures**

697

698 **Figure 1** (*a*) wooden bucket; (*b*) canvas bucket; (*c*) German scoop (modern version).

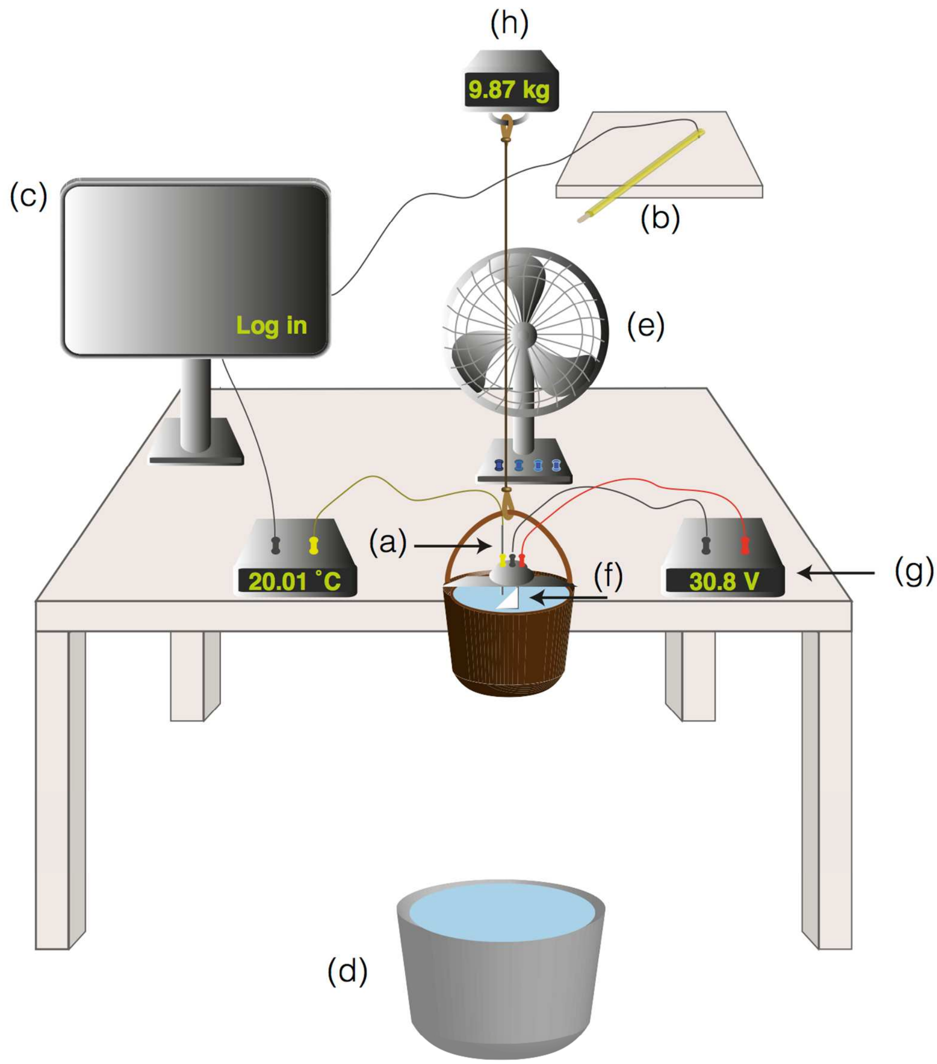

Figure 2. Illustration of the experimental setup: (*a*) precision thermometer (F250); (*b*) air temperature and relative humidity probe (Vaisala); (*c*) PC used for logging; (*d*) plastic bin (containing clean, freshwater) used to soak the buckets; (*e*) fan; (*f*) automatic stirrer; (*g*) power generator for the automatic stirrer; (*h*) hanging scale.

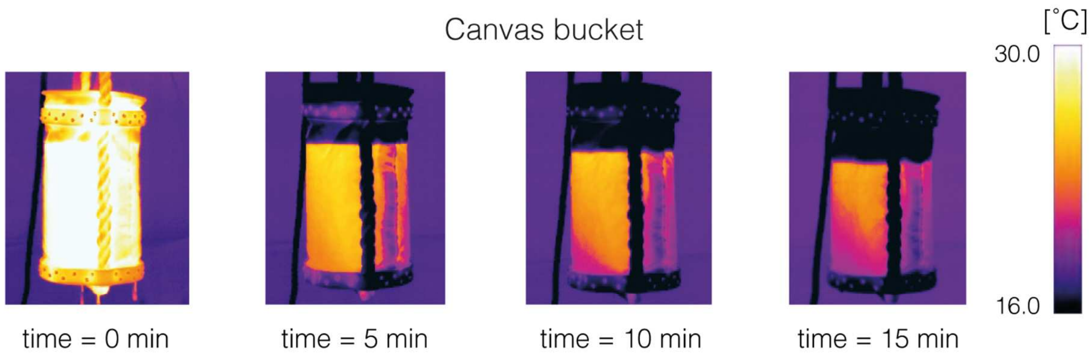

Figure 3. Thermal pictures taken at 5-minute intervals of the Met Office Mk II canvas bucket (Figure 1) filled with warm water hung in front of a fan positioned to the right in these images. The bucket is not stirred and the lid is shut.

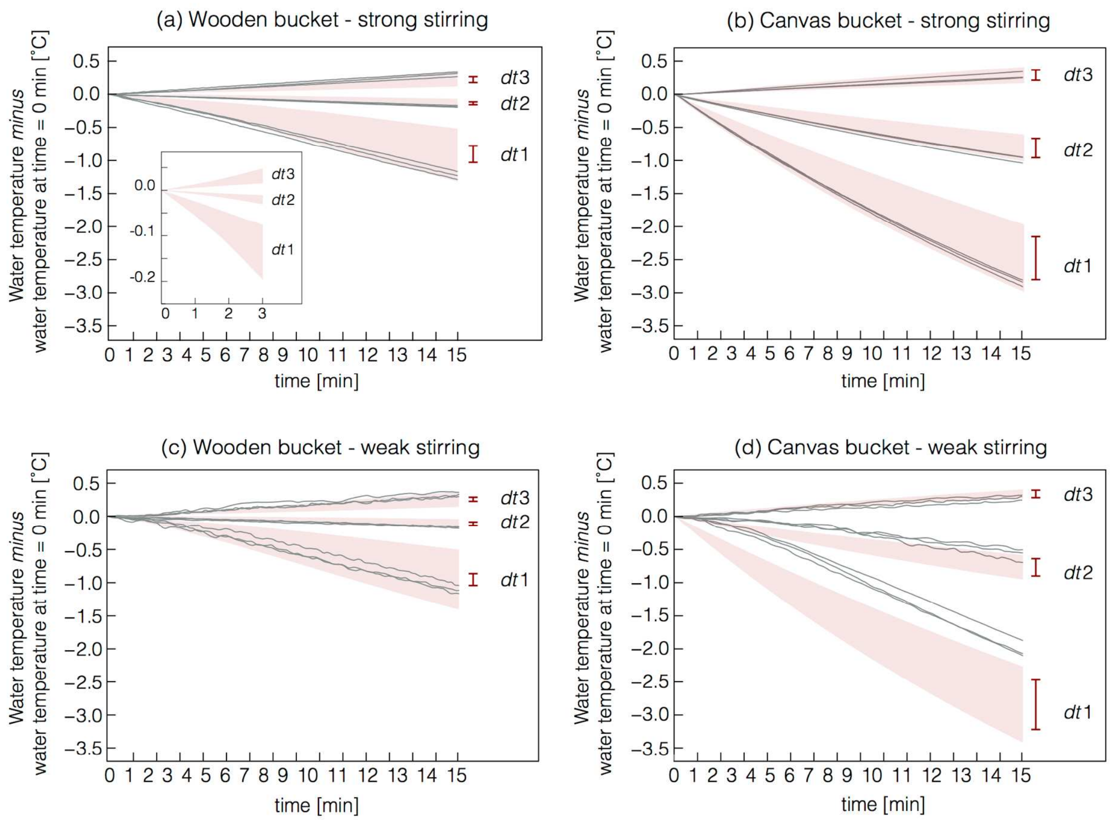

Figure 4. Measured (*black lines*) and modelled (*pink shading*) evolution of the water temperature over time. Shaded regions represent model uncertainty at 95% confidence level. Also shown is the wind speed only contribution to the model uncertainty at time = 15 min (red bars). Each panel shows three sets of measurements with different initial water temperatures. *dt*1: water ~ 5˚C warmer than ambient air temperature; *dt*2: water ~ 1˚C colder than ambient air temperature; *dt*3: water ~ 5˚C colder than ambient air temperature. (*a*) time evolution of water temperature – initial water temperature for wooden bucket, strong stirring. Inset shows expansion of first 3 minutes for *dt*1; (*b*): as (*a*) but for canvas bucket, strong stirring; (*c*) as (*a*) but for weak stirring; (*d*): as (*b*) but for weak stirring. Fan speed of ~ 3.5 m s^{-1} throughout.

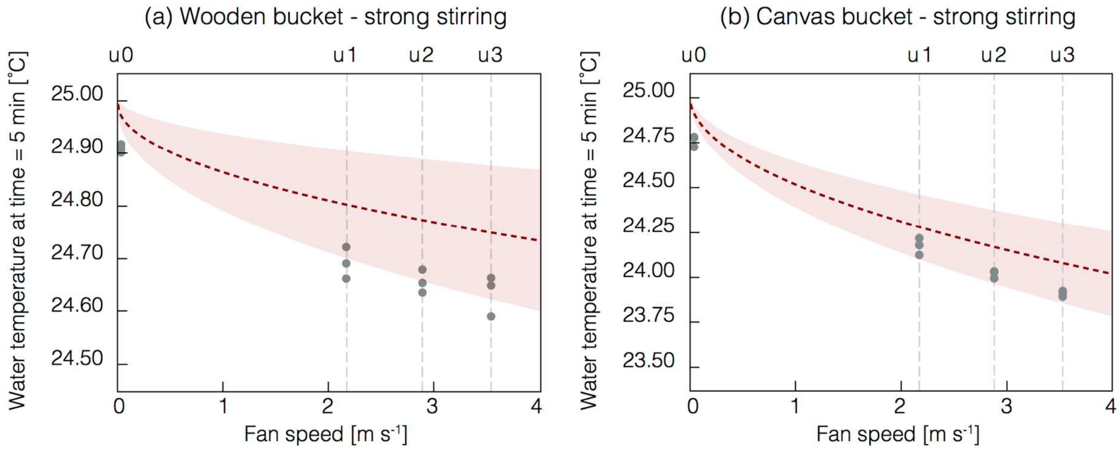

Figure 5. Measured (*dots*) and modelled (*dotted line and red shading*) water temperature at time = 5 min as a function of the air (fan) speed for t_0 - t_a ~ 5°C. Lines represent the median of the model output; shaded regions represent uncertainty at 95% confidence level in the model output. (*a*) wooden bucket, strong stirring; (*b*) canvas bucket, strong stirring. Note change to y-axis scales.

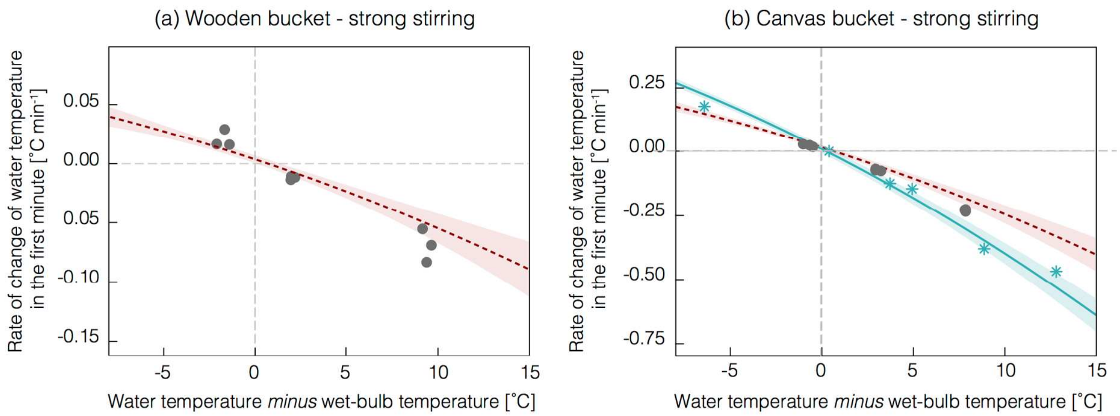

Figure 6. Measured (*dots*) and modelled (*dotted line and red shading*) rate of change of water temperature at time = 1 min as a function of the water temperature *minus* wet-bulb temperature difference. Also shown is the measured (*stars*) and modelled (*solid line and blue shading*) rate of change for the Ashford (1948) results with the Mk II Met Office bucket. Lines represent the median of the model output; shaded regions represent uncertainty at 95% confidence level in the model output. (*a*) wooden bucket, strong stirring; (*b*) canvas bucket, strong stirring. Fan speed of ~ 3.5 m s^{-1} throughout. Note change to y-axis scales.

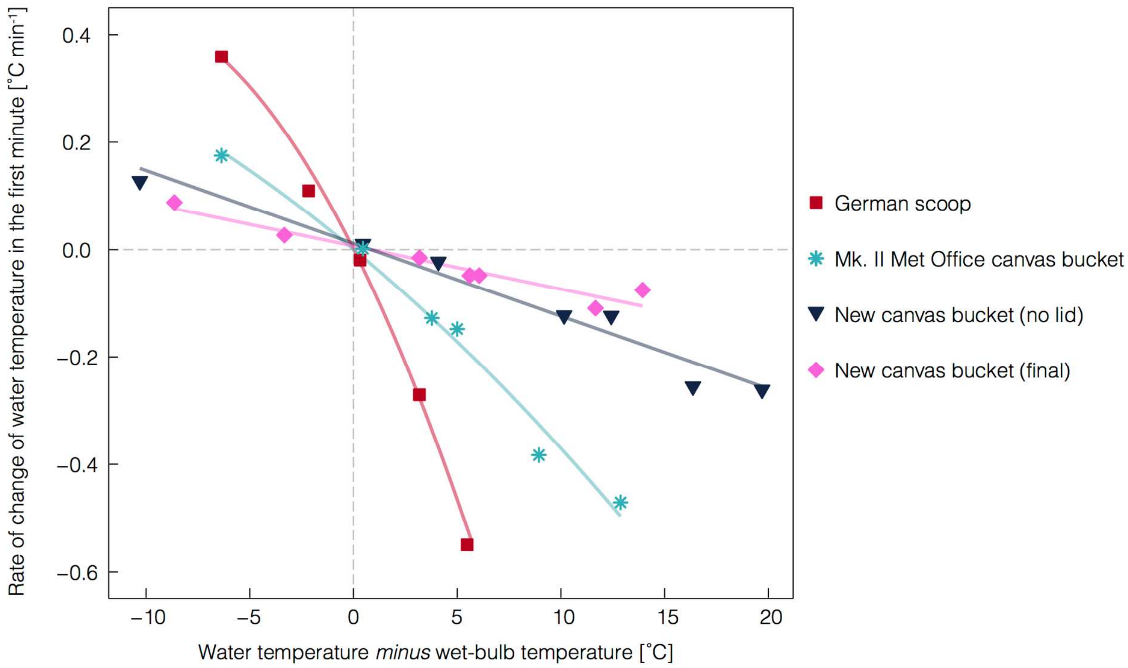

Figure 7. Reproduction of Ashford (1948) results, values have been read from Figure 2 in Ashford (1948). The plot shows the rate of change of water temperature at the first minute as a function of the initial water temperature *minus* the wet-bulb temperature for the German scoop (*red squares*), the Mk II Met Office canvas bucket (*blue stars*), the new canvas bucket (*pink diamonds*) and the new canvas bucket without the lid (*dark blue triangles*). Here the lines represent polynomial fit to the data, while in the original figure the lines were hand drawn.

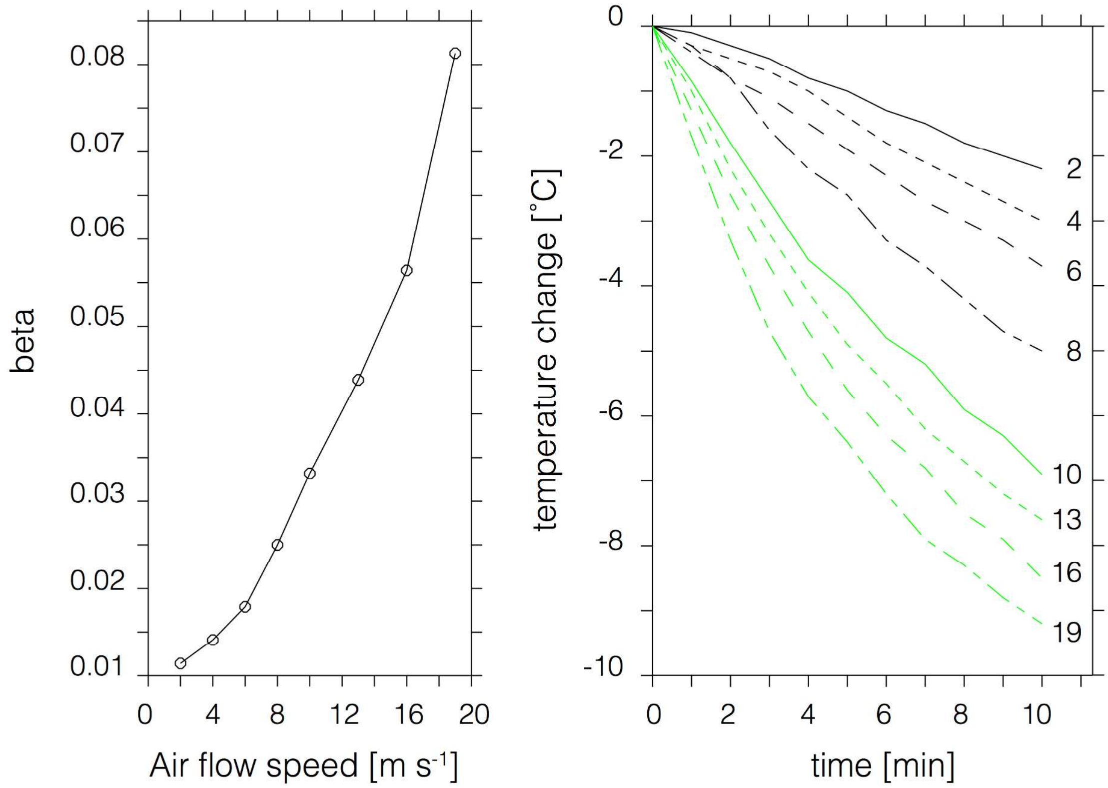

Figure 8. Reproduction of Roll (1951) results, values have been read from Figure 1 and 2 in Roll (1951). (*a*): β, Equation (2) as a function of airflow speed. (*b*): temperature change over 10 minutes at 8 different airflow speeds (as annotated for each line, m s^{-1}). Values were read from the original figures every minute.