

I respectfully disagree with most points but ESSD, as a journal, does not promote scientific contention or discussion. It publishes “original research data and data collections” for two purposes: a) make datasets a reliable, fully citable resource upon which other research can build; and b) reward data providers by establishing their priority and recognition. This review slights the data; I find nothing about data availability, organization, etc. Not obvious to this author that reviewer understands ESSD requirements? Not possible nor useful to make changes to manuscript text based on comments from this reviewer.

Review expresses appreciation of inexpensive environmentally-friendly sensors and deployments easily applicable in many regions; good. Review also seems to appreciate suggestions for scientific relevance; also good.

Most comments easily dispensed. Reviewer needs only to read carefully or watch example video. Wheel size, rotation, color etc. had no influence. Look at video. UV used to check visible data; read text following line 280. Supposed near-surface temperature profiles assume a base-point or starting point; this manuscript demonstrates that ‘base-points’ vary as a function of surface type, surface temp, gravel, vegetation, season, shading, etc; check example video. One of these sensors (TMP 117) often serves as primary reference; does one need to validate primary sensor against itself? Swiss and US sensors perfectly validated (see Figs 5, 6). Information about sensor manufacture and performance, rarely available for weather service sensors, included on freely-available data sheets. Sensor outputs compared directly to NOAA Climate Reference station data; only 3 exist in Montana. I use closest (25 km) station (Figure 7). Reviewer needs to assess Figure 11; first (only?) season-long inter-comparison between surface data and CRN data. If reviewer knows better free open-access source this author would like to see it. (Please note: ESSD does not allow CC-BY-NC contributions.)

Of course data only cover specific routes to and from a small city: one person on a bike. I make that point frequently and repeatedly throughout manuscript. I include vegetation and development (e.g. Demuzere 2022) categories, to facilitate comparisons with other studies. Multiple bikes carrying like sensors? Therein lies great promise. None of those cyclists will purchase standard Campbell \$15k sensor packages (plus data access fees?) nor try to mount commercial sensors on a bicycle (several failed examples exist). But they might follow guidance provided here to purchase and mount \$250 sensor packages of remarkable capability. As I contend in manuscript: bicycles should represent key component of future observing systems

Urban regions undoubtedly contain “roads/paved surfaces, but also include trees, grassland, buildings, and possibly rivers and lakes”. This dataset includes all of the above, geolocated at very high resolution, measured repeatedly over a full year’s worth of insolation, shading, snow cover, etc. So far, no other observations have covered more than a few warm days. What happens in deep shade, in daily- or seasonally-varying shade, when shade coincides with gravel, over grassland or on grassland under shade, or on paved surfaces under snow or ice? One can address these questions, accurately, with these sensors and this bicycle. Do we need more measurements in more regions under more conditions? Yes, as called for explicitly here! Those regions, if judiciously-selected, should allow multiple full-season measurements over pavement, gravel, shade, across all times of day. Until multiple cyclists accept the challenge, this represents a unique open-access single-cyclist survey.

Not sufficient to build “quantitative relationships with satellite remote sensing data”? This manuscript spends several paragraphs on plausible discrepancies between periodic (once or twice per month, only useful under cloud-free conditions) remote sensing data at (rarely) 10 m resolution versus these data at 4 m resolution under all weather, sky and environmental conditions. Review apparently missed those discussions? This author clearly asks a reverse

question: how will remote sensors validate their products against high-res ground truth?
Challenges ahead! Based on open-access data sources! What about slight GPS mis-locations?

As clearly mentioned, cooler surface temperatures of gravel surfaces long recognized. But, based on what data? Includes shade? Snow? Real world data, or generated artificially in a laboratory? Year-long or only on warmest days? Reviewer knows free open-access source for temperatures of pervious versus impervious surfaces under all conditions? If researchers want to explore UHI factors using an atlas of building- and vegetation-influenced surface temperatures, this data might make a worthy contribution? Reviewer entirely silent on this aspect. No comments on GPS registration issues, on daily patterns if shading, on snow?

I thank this reviewer for time and effort.

This manuscript presents review challenges. Sensors different to 'normal', deployed on a bicycle, potentially relevant across (at minimum) engineering (e.g. pavement durability), climate (causes and prevalence of urban heat islands/archipelagoes), hydrology (altered run-off), ecology (vegetation, migration corridors), etc. Reviewer wants consistency in uncertainty terms? Within topic (e.g. atmospheric temperature) I cite clear sensor uncertainties, list explicit correlation coefficients, include data sheets that allow readers make their own checks, and - for field checks - report as many intercomparisons as possible. If reviewer knows or prefers a different approach, one that covers temperature, lux, %, etc., I would gladly hear it.

Reviewer's comment about distance to standard weather station data: "In general, if a station is more than 10 miles or 100 feet different in elevation from the target location it is not considered a comparable record" provides neither assistance nor guidance. Can reviewer provide valid calibrated available data to back up this 'rule of thumb'? Data from NOAA's Climate Reference Network (check description and map), as used here, adhere to quality rather than artificial distance-based rules.. Weather services deserve enormous credit for maintaining reliable data sources amidst constant land use change (multilane highways, neighboring hotels, train stations, vast parking lots, runways, terminal buildings, constant traffic, nearby water, etc. If we adopted reviewer's 'rule-of'-thumb' qualification, we would relegate most regions, particularly mountain regions, of our world as data voids? To specify distance boundaries without citation or validation of source seems not helpful.

Reviewer apparently prefers less detail, discounts local sources, but wants more general discussion around uses of weather station data? Confusing. In any case: not appropriate topics for ESSD publication? I challenge reviewer to generate a second source that matches CRN data as well as match displayed in Fig 10 (Figure references refer to revised manuscript).

I likewise question reviewer's assertion that other better methods exist, even if one focuses solely on air temperatures. Drones, at least those that I know, generally require modification (at cost of additional weight) to measure air or pavement temperatures and operate within tight battery (15-20 minute) and weight limits; they seem unlikely to work in places or on patterns that bicycles can achieve. If reviewer knows alternate technology by which to observe and report underlying temperatures at better than 5 m

resolution, by which to take valid calibrated measurements across various surfaces, under vegetation canopies, along streams or across grasslands, at many times of day and in many seasons, free of flow distortion, I welcome suggestions! Manuscript as submitted addresses current satellite resolution (less than presented here by at least factors of 2) but, intentionally, did not call out very poor satellite overpass (repeat) patterns which themselves limit local application (Section 4). One hopes that bicycle sensor wavelengths and measurement patterns eventually match satellite verification needs. Very local GPS discrepancies (mis-registrations?), as exposed here, might offer a mutual challenge? Referee offers no discussion of GPS issues.

Reviewer suggests specially-constructed sites in “laboratory setting that provides an environment that could be controlled for wind, lighting, humidity, and other ambient atmospheric conditions,”. . . Pomeranz et al. (2000, cited, working from US EPA and DoE funding), looked at pavement albedo at local sites around LBNL. They say upfront that albedo played almost no role in observed daily (two data points, both around 50C) surface measurements. They found no reliable relationship between albedo and pavement age. They converged, using observations (collected entirely on single sunny afternoon) plus models, on maximum albedo-induced surface heating of 0.6C over realistic range of pavement albedos. Too small to have serious impact on surface temperatures as reported here; they have never, to my knowledge, published data nor follow-up. A group at Arizona State University built an enviable fixed site array consisting of multiple surface plots covered by decent instrumentation but, again, published no data. Will data offered in this manuscript stand up to scrutiny by those or other investigators? Reviewer mentions apparent mismatch recorded by these data and typical UHI (sun-warmed pavement heats overlying atmosphere) assumptions but assigns difference to wind or relative humidities. I include RH values in every file; with no apparent relation to any other variable. Users can check this assumption (e.g. Figure 15) or explore alternate hypotheses. Like most bicyclists I ‘feel’ wind (ahead, from either side, or behind), mostly its extreme variability. After several hundred rides at all times of day and in all seasons, exposed and oriented in all directions, I sincerely doubt any influence of local winds; users can - with my cautions - make their own explorations (as described in legend for Fig 2). Because disconnect between pavement heating (often 25C, occasionally 30C, warmer than air in summer, colder than adjacent atmosphere by ~14C in winter) and overlying air temperature represents a provocative feature of these data, lack of attention by this reviewer proves disappointing. What do we miss? Wider deeper roads? Sub-surface infrastructure? Buildings, parking lots,

vehicles? Larger expanses or abundance of any of these? Might these data prove valid and useful in raising such questions?

Many people would absolutely carry measurement systems on bikes; they invariably express curiosity and enthusiasm about my system. Bicyclists on campuses or in cities could, by modest efforts, achieve substantial improvements in understanding local surfaces, shade, daily heating, etc. Think of multiple bikes operating in multiple seasons in urban settings, sharing data. My bicycle-friendly proclivities contradict reviewer's pessimistic expectations. Cities devote substantial funds to bicycle infrastructure; each bike trip represents a success. How might bikes, e-bikes or E-scooters impact pavements, traffic and measurements? What if E-scooters automatically reported pavement and local air temperatures at completion of each rental? Review provides minimal assistance. I prefer to focus on accolades: extensive, extremely detailed, easily reproduced, well documented. Exactly what we hoped for ESSD.

I de-emphasized particular bicycle; moved description to Appendix. Emphasized temperatures of surface pavements and relation to UHI effects. Added slightly to UHI and bicycle literature. Please remember that under my explicit filters (free open access, measure both air and surface temperatures, cover range of seasons and road/path conditions, validate to best of ability) very little prior work qualifies.

Thanks for perspective and good efforts.