Revisions / Rebuttals to Ms. SE-D-19-03074

Updated: Monday 2/3/2020

Contents

General Feedback

We have provided revisions and rebuttals to reviewer feedback to manuscript no. SE-D-19-03074. We sincerely thank the reviewers for their insights, suggestions, and taking the time to help us improve our work. The comments were extremely insightful and allowed us to strengthen our draft manuscript. All of the feedback was valuable.

Responses to reviewer concerns are provided below in the order they appear, and corresponding inline edits/references have been highlighted in a revised manuscript (submitted as a supplementary document), where feedback for Reviewer #1 is highlighted in green, Reviewer #2 in magenta, Reviewer #3 in cyan, and overlapping feedback for multiple reviewers in gray. The highlighted supplementary manuscript document will help the reviewers navigate the submission to easily find our edits in response to their feedback. Both variations of the manuscript are provided.

We would like to provide the following general clarifications before addressing each specific concern:

- Most scientific contributions steadily complement, confirm, or refute theories and approaches.
- We attempt to explain upfront that each categorical approach to the problem of estimating sky energy has its own merits and demerits (and after reviewer comments, we now better explain advantages and disadvantages of each approach). The work we have accomplished is scientifically sound, valuable to the solar energy community (especially given the overlaps), and useful for downstream building performance simulations, illumination studies, and rendering applications today. We have confirmed and complemented an initial data-driven approach by prior authors, expanded findings over a wider, more useful spectrum of energy, predicted energy for entire hemispherical skies, and provided a refined, efficient pipeline for both researchers and engineers to employ on clear skies specifically. We also used a much more comprehensive data set [*], which spanned an entire year, to help accommodate for seasonal turbidity variation.
- Our proposed methods are intended for real-time solutions needed yesterday [†] to help combat climate change in the (built-environment) building performance sector. Our methods confirm relationships between look of sky and underlying energy, and more importantly, can be used to reconstruct useful estimates based on low resolution monitoring with ubiquitous, cheap technology. We purposely omit atmospheric measurements for this purpose. In future studies, AOD and other measurements can be included as training/input features to increase accuracy, however the strength of our approach is demonstrating what can be accomplished without this data.

Signed by all authors as follows:

Joseph Del Rocco [‡], Paul D. Bourke, Charles B. Patterson, Joseph T. Kider Jr.

Joseph Del Rocco
University of Central Florida

[*] Kider Jr, J. T., Knowlton, D., Newlin, J., Li, Y. K., & Greenberg, D. P. (2014). ACM (TOG), 33(6), 1-12.
[†] Butler, D. (2008). Architects of a low-energy future... Nature, 452(7187), 520-524.
[‡] Corresponding author

Reviewer #1

Concern: clear sky only

The biggest issue of this work is that the method is only for clear sky condition, which significantly limits its application values. Please also assess the other conditions or clearly indicate its application in other conditions.

Response

This is a very reasonable, expected concern of this work.

As mentioned, this work is a continuation of work we published last year with SPIE (Del Rocco, 2018). We initially tested all of our data (in a less refined pipeline) and found that while 1 of the 4 regression models did moderately well on cloudy skies, most of them did not. Yet all 4 final regression models did very well on clear skies. So we made the decision to focus on clear skies with regression models for this work, perform more experiments, validate our work with a radiative transfer package, compute radiance for every single pixel of a sky image (sradmaps), utilize our HDR data, and clean our data further (we lowered error a few percent by accounting for lens warp, fixed a few bugs in data processing and visualizations of data (as seen in the whole sky error maps), etc.). In general, we refined and expanded our work. We choose to initially focus on clear-sky data experiments. We have included some mixed and cloudy-sky results that demonstrate the challenges and future research opportunities.

In our Conclusion section, we now mention initial results we are having with deep learning neural networks and our "mixed" (unified) dataset of all data. Our ongoing work involves investigating patterns in unified or scattered sky data with neural networks. We will not include that work in this paper, as it will take the rest of this year to complete and requires more experimentation.

We now discuss this point upfront in the new Section 3.1.

The title of our paper was deliberately chosen to emphasize this point, but we are willing to modify it if another title better describes our work.

Concern: innovations of this work

The innovation of the paper is not clear to the reviewer. As the authors indicated, some work, such as Tohsing et al. 2014, has proposed similar work, applying similar machine learning models to a wider spectrum has limited contributions to the existing body of research.

Response

Text has been edited and expanded in the Abstract and on lines 39, 48, 53, 56. The "bullet list" of contributions we provided at the end of Section 2.3 starting at 147 has also been edited. Among those listed, how about the prediction of high resolution (non-visible) radiometric signals from low resolution (visible) photometric measurements? This is not obvious and somewhat unexpected. Demonstrating the work holds for non-visible spectra is relevant to many applications and non-trivial since many absorption bands occur only in the IR.

Concern: correspondence of radiance distribution to sky image

While the input features are well described, how is the radiance distribution measured and corresponded to each point on a sky image is vague. Please better describe it.

Response

Sky capture details of the validated radiance measurements and photos of the sky are described in (Kider et al., 2014), cited multiple times. We can provide even more detail from this work on the capture process if the reviewers request additional information. We choose only to focus on our new contributions in this manuscript.

We now better explain the correlation between specific locations of a sky capture photo and the radiance measurements in sections 3, 3.2, and 3.3.

Concern: figure clarifications

Some explanations and descriptions are needed for figures, such as Figs. 1, 2, 3, and 4, to help readers better understand the problem formulation and sky image setups. For example, what do those points in the left figure of Fig. 2 mean? What's the relationship between those points and lines in the right plot of Fig 2? Also, axis title should be indicated in some figures, such as the right plot of Figure 3.

Response

Figure 1 slightly modified and description expanded. Figure 2 has now been re-plotted and its description rewritten. Figure 4(b) plot now has axes, and description rewritten. Figure 5 expanded to better explain sampling methods. Other figure descriptions throughout the paper were expanded as well.

Concern: feature importance analysis

The feature importance analysis should be better described. And more advanced feature selection method should be used to analyze and select features in this regression problem.

Response

This is now better described in Section 4. The figure with features individually listed has also been updated to reflect one less feature to tighten our explanation.

Concern: RGB values from multiple exposures as feature input

Why didn't the authors input RGB values from images of different exposures to the machine learning models, which might improve the accuracy?

Response

As mentioned in 4.1, this was precisely our thought, that HDR insights would improve the learning. Capturing HDR data involved extra work, was not trivial, and dramatically increased our data storage and offline processing. In reality, as the results of this experiment are described on line 426, multiple exposures didn't help at all... at least for <u>clear skies</u>. We believe this is because clear skies are too uniform in color gradient, regardless of exposure. We have a suspicion that it will help for clouds however, since the cloud colors scale differently between exposures. We our using multiple exposures in our ongoing cloudy sky work.

Concern: address specific bad predictions

(P12, L284): Please explain reasons for the bad predictions.

Response

There is no way to know for sure, but in similar work by other authors, the worst predictions (highest deviations) often occur within or near the sun's corona, where radiance values are traditionally higher and more erratic compared to the rest of the sky. This is common. In fact many authors prior to Remi Chauvin's team, simply excluded a 20% circumsolar region to account for errors that often occur there due to problems like lens flare, etc. All of our whole sky error plots show this pattern. We added clarifications on line 269 and 419.

Concern: address noise in distribution between 1 and 5 nm resolution

(P12, L303): Why is the noise in spectral radiance distribution reduced when decreasing the spectral resolution from 1 to 5 nm?

Response

The word "noise" was a poor choice and has different mental models in different disciplines. We meant frequent oscillations of the radiance curve. We have corrected this explanation on 440 and Figure 16 description.

Concern: how many models and computation time for sradmaps

(P10, L259): How many models are needed to generate a sradmap? Please also report the computational time.

Response

Only 1 model is needed to produce an sradmap of a sky. We stated this on line 452. We now clarify this on line 395 and Figure 7 as well.

We state computation time of sradmaps on line 454. We now elaborate on optimizations that can be taken to improve speed further on the next line. We now reiterate the computation time on line 495.

Concern: dataset URL

Please give a link to this publicly available dataset.

Response

Available from our project website (on Github): https://github.com/spectralskylight

The underlying data will also be hosted at UCF's Showcase of Text, Archives, Research and Scholarship (STARS) system repository (http://stars.library.ucf.edu) for long term storage and access by researchers. STARS is committed to providing access to individuals with disabilities (http://stars.library.ucf.edu/accessibility.html). Administered by the UCF Libraries, STARS is available to host and promote research, creative activity, and institutional outputs to ensure persistent access to research work. Data will be stored on STARS for at least 10 years as an additional mechanism to keep the data open and public.

Concern: libRadtran accuracy

libRadtran is used to validate the proposed method, therefore, its accuracy should be reported first.

Response

More explanation and citations to peer-reviewed publications have been provided on lines 56, 103, and 469.

Reviewer #2

Concern: names before acronyms

Please make sure that you define names before acronyms even it is defined in the abstract. For example line 33: HDR, and line 37: RGB.

Response

Defined acronyms on lines 27, 31, 36, 63, Figure 1, 84, 86, 87, 98, 132, 163, 287, 385.

Concern: figure 2 axis

y-axis is missing for the distributions plot in Figure-2.

Response

Provided now.

Concern: better compare approaches to this problem

The authors criticize analytic-based methods, physics-based methods, and some other data-based models in many places in the article, but they do not present any real comparison with these techniques in the body of the paper (score values, application on the same data, ...). How the authors assume the superiority of their methods over the others?

Response

Thank you for pointing out this tone in our paper. It was not our intention. We intended to describe related work and various approaches. This is a very normal categorization of approaches and we were just outlining advantages and disadvantages for each.

Nowhere do we assume superiority of our method over other scientists. Rather, we explain in Section 2 that each categorical approach to the problem of estimating sky energy has its own merits and demerits. We have certainly improved upon the initial approach by Tohsing and other authors of data-driven approaches to this problem. We captured and considered much more data spanning an entire year, confirmed and replicated findings over a broader spectrum of energy (useful for many fields/problems, not just atmospheric science), predicted energy for entire hemispherical skies, and provided a refined, efficient pipeline for researchers and engineers. However, as mentioned and pointed out by Reviewer 3, data-driven approaches are not perfect solutions either.

We now better explain this throughout Section 2.

Concern: over-fitting and cross-validation

The main drawback of supervised methods (methods used in this work) is the over-fitting effect when the models memorize the training data rather than discovering the actual pattern (fictitious optimism), especially with many inputs variables like the case of this work; 13 entries. However, the authors use 10-fold cross-validation as a preventative measure against over-fitting which, in my opinion, does not work well usually, especially with the constraint of limiting the study in periods of clear skies. How it is the distribution of the clear sky atmospheric radiation distributions (if for example considered a vector variable)? low variability data has a negative effect on the 10-fold cross-validation. I strongly recommend that authors discuss and clarify this point.

Response

Overfitting is a problem for all machine learning problems; it is not unique to supervised methods or our method. We have taken a couple of measures to test and prevent overfitting to the best of our ability described in detail below. Since this was a concern - we validate our predictions with an external 3rd party library, libRadtran, a validated, well-cited radiative transfer software package that computes atmospheric data.

In general, if a model is underfit you can add complexity by adding features, and if overfit you can decrease complexity by removing features. If this concern is directed at linear regression specifically, we point out that only one of our models is a linear regressor (LNR). Even so, LNR was the only model to use polynomial expansion, which greatly "artificially" expands the input feature set to add complexity, and yet was still the worst performing and clearly not overfit.

If a model learns training data too well, it may perform poorly when tested on unseen data. To combat overfitting, we performed the two "gold standards" of machine learning projects, which was to (1) split all data into a training set and a holdout set (the holdout was never used during the tuning of any models; it also consisted of entire skies worth of samples, to prevent training on parts of a particular sky and predicting on other parts of that same sky; this was never done), and (2) 10-fold cross-validation (CV) to help the models better understand the dataset without showing them holdout data. We run cross-validation to demonstrate accuracy is not biased to any training set and has similar performance with our test set. Without cross-validation, models may be skewed or overfit to particular subsets of the data. K-fold cross validation is a standard technique to detect overfitting. However, there is no guarantee that k-fold cross-validation removes overfitting; we acknowledge this point. Additionally, we shuffle the order of the data and ensure no duplicates in folds. CV allows one to partition the data once, train/test and holdout, as opposed to separate train, test, and holdout partitions.

We felt the supervised approach was inherent to our data. We have photos of skies and 81 radiance measurements (curves/distributions) per sky. The measurements are already ground truths in themselves. We are trying to learn the pattern (if it exists) between what the sky looks like (and capture time) that produces those measurements. Perhaps this can be configured as an unsupervised problem, but that didn't seem obvious to us.

Our final feature set is actually 12 (sample altitude was found to be completely unimportant and thus dropped; we corrected the typo and feature figure in the manuscript). We could also drop sample azimuth as the EDA showed it was the next least important feature, however a test with and without it revealed a 1-2% RMSD difference, so we kept it. We now mention that this feature could also be dropped. Before we discovered solar-point-angle (SPA) from (Chauvin et al., 2015), sample azimuth and altitude were more important. Clearly the position of a sample in the sky isn't as important, once SPA is included. We improve the generalizability of our approach by removing irrelevant input features and testing the impact of each to produce the simplest possible model.

We know that tree (ensemble) estimators are more prone to overfitting than other regressors, so to further address overfitting, we used a Random Forest Regressor (RFR) specifically, which harnesses randomness to decrease variance in lieu of some bias. With this in mind, we then discovered Extra Trees Regressor which introduces even more randomness for an even bigger tradeoff. The very purpose of both of these regressors is to combat overfitting. Interestingly, ETR performed slightly better than RFR, even with more randomness. You can read more about them here:

- Kocev, D., Vens, C., Struyf, J., Džeroski, S., 2013. Tree ensembles for predicting structured outputs. Pattern Recognition 46, 817–833.

- Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Machine learning 63, 3–42.

Is it still possible that our models are slightly overfit in one way or another? Of course. But all machine learning projects are susceptible to this. The only way to know for sure is to continue to test the models on even more unseen data over time. In our case, we have already captured more data (Kider, 2014) than any other author of a data-driven approach to this problem prior to us, and our dataset spans an entire year to help with seasonal variation. We welcome a validation study by future authors that tests our pipeline on more data. Prior authors in this domain have validated their work on a much smaller subset of data or part of the sky, such as just the zenith coordinate, or some arbitrary coordinate. We validated our results against holdout ground truth measurements as well as libRadtran.

Concern: outputs as categorical variables (?)

I do not see how the outputs "atmospheric radiation distributions" are converted into categorical variables to deal with the suggested methods.

Response

We're not entirely sure we understand this concern, so please feel free to clarify further if the following response does not satisfy.

This work involves regression analysis for which the final output is a vector of data representing a curve (throughout the paper we usually call it a distribution), as opposed to the outputs of classifiers, which produce a single output or one-hot vector that needs to be decoded to represent a category. Our regressors output a vector of data representing decimal radiance values at each specific wavelength of interest of the electromagnetic spectrum (350-1780 nm) (spectral radiance). There are no categories and no values which need to categorized. There is only a curve, which is a vector of decimal values.

Our method is by no means perfect (and we measure and report the error), but we did show that this it is possible to predict these curves for the unseen clear skies of our dataset. We compared our predicted curves on that data with actual measurements of 81 points per sky that was heldout, to compute our RMSD, R2, and ratio errors. We also compared our predicted curves to those generated by a popular, validated radiative transfer package, libRadtran.

Many machine learning algorithms have a larger set of inputs that are whittled down to a smaller set of outputs, but we had the exact opposite problem, a handful of input features (that we engineered to 12) to predict a curve/distribution (our vector output) of 1430 (350-1780nm) output values. This is different from the well-known machine learning problem of taking many feature input extracted from an image database and outputting a single (or one-hot vector) classification.

Some of these points were explained more in our previous publication, and because we are continuing that work, we didn't feel the need to repeat those explanations. In fact, it could be seen as unethical for us to repeat those explanations here.

However, given your concern, we have refined various lines of text in Section 4 to better explain this.

Concern: preliminary results

lines 202: "Preliminary results encouraged us to focus on the following regression models ..." what are these preliminary results? I think this is a key detail for clarity.

Response

We have made edits to Section 4 to address this concern.

That line simply means we initially considered as many regression models as we could find at first available in Weka toolkit (briefly), and then scikit-learn library (ultimately), quickly tested the ones we could without refining anything, and pursued the ones with promise. Many of the models forced a single decimal output value (not a vector), which didn't fit well with our problem formulation. So we focused on the vector output regression models. We initially tried 10 of them without any tuning, and obviously some of them performed better than others from the start, we chose 2 of them at first as the ones to focus on (KNN (proximity based) and a tree based one), and we included a standard linear regressor (LNR) as a baseline that we assumed would not perform well. In Section 4 we also already say, "*10 separate regression models were trained and tested, including: linear, Ridge (Hoerl and Kennard, 1970), Lasso (Tibshirani, 1996), ElasticNet (Zou and Hastie, 2005), Lars, KNN, RandomForest (Kocev et al., 2013), ExtraTrees (Geurts et al., 2006), etc.*"

Machine learning is a huge toolkit of algorithms/heuristics/statistics/models. One has to throw the data at a wide variety of techniques before finding something interesting. It's possible to tune and tweak hyper-parameters of every single algorithm you come across, and perhaps improve performance, but that can take a long time, and the final results may not improve much or be over overfit. We took an Occam's razor approach - try many models up front and see which ones have promise, then tune them by engineering new inputs, tweaking hyper-parameters, double-checking and cleaning data for any anomalies, and basically investigate why error is occurring to see if we can address it without exposing the holdout data. That is what we did with the 4 final models we chose, ETR, RFR, KNR, LNR.

Preliminary results of cloudy sky data in particular is now discussed in the new Section 3.1.

Concern: address weaknesses of methods

A paragraph where the authors discuss their perceived weaknesses in their methods and results will be great.

Response

Weaknesses of each categorical approach are provided in Section 2. Weakness of our pixel sampling technique is mentioned in Section 3.3, line 236. Improvements that could complement this work are now better spelled out in the Introduction and Conclusion. Future and ongoing work is also discussed in the Conclusion to account for work not provided in this paper.

The most obvious weakness of our dataset is that it was captured at a single location on Earth. But that is by no means unique to us. Most authors are only using data from a single site. It would have been nice to train on sky images and corresponding radiance measurements from multiple measuring stations from multiple sites. This can still be done, but not in a timely manner. It requires serious coordination, cooperation, and/or considerable normalization of data for it to work in the same pipeline. We are currently working with other researchers to expand this dataset to 3 additional locations.

Concern: open-source and dataset

I see that the authors are following the good practice of making the codes and data publicly available. I strongly recommend authors to put their data and code on a GitHub repository, in this way the visibility of the research will be increased and other researchers may pull requests and contribute to the development and impact of the models.

Response

Available from our project website (on Github): https://github.com/spectralskylight

The underlying data will also be hosted at UCF's Showcase of Text, Archives, Research and Scholarship (STARS) system repository (http://stars.library.ucf.edu) for long term storage and access by researchers. STARS is committed to providing access to individuals with disabilities (http://stars.library.ucf.edu/accessibility.html). Administered by the UCF Libraries, STARS is available to host and promote research, creative activity, and institutional outputs to ensure persistent access to research work. Data will be stored on STARS for at least 10 years as an additional mechanism to keep the data open and public.

Reviewer #3

Concern: effect of AOD and particle changes

the modeled radiance will systematically deviate from that measured when AOD would increase and/or the nature of aerosol particles would change

The spectral distortion effects become more apparent under elevated aerosol content because aerosols of different sizes, shapes, composition scatter the light in a different manner. Large particles cause the scattering function becomes more forward-lobed, while small particles cause the importance of side scatter increases. The ripple structure of scattering function is significantly suppressed when aerosol population shows wide range of particle sizes, but some specific signatures of scattering function are amplified if particle size distribution is quite narrow. These signatures smoothly disappear when transitioning from non-absorbing to absorbing materials. The above effects could be marginal if AOD values are low enough compared to those of ROD, but some effects may become decisive when AOD exceeds that of ROD. It seems the authors did not made the experiments under elevated AODs (e.g. for AOD¿0.4 at nominal wavelength of 500 nm). The author should validate their method for high values of AOD.

Response

We are aware of how particles scatter energy from a microphysics perspective. The question is, how much does turbidity affect the spectral radiance distributions across clear skies specifically, without affecting the appearance of the skies? (Hosek and Wilkie, 2012) show renderings of clear skies with varying turbidity factors (from T2-8), which align with reference photographs, and higher T-factor skies are visibly different. We are all aware of what an 8 T-factor sky looks like (not that clear). (Willers and Viljoen, 2016) in their study of the effect of the atmosphere on the color coordinates of sunlit surfaces, state that aerosol attenuation in the atmosphere has a relatively weak spectral variation compared to molecular absorption. The older (Steven and Unsworth, 1977) observations found departures due to variation in atmospheric turbidity to be small.

The (Eltbaakh et al., 2012) overview of attenuation by aerosols found inconsistencies in various turbidity metrics, and stated that the metrics are hard to distinguish from water vapor, are weather and climate dependant, seasonal, diurnal, geographically affected, and in general, not independent of optical air mass. We captured measurements over a longer span of time than many authors. As mentioned, *"453 total sky captures were taken over 16 days between 2012-2013, covering all four seasons, dawn to dusk, and various sky covers, for a total of over 36000 individual spectral radiance measurements."* And our capture location, Ithaca, New York, United States, is more scattered and overcast than clear on average. This suggests that at least some of our measurements do contain seasonal, diurnal clear sky turbidity factors that are affecting the look of the clear skies.

Our model predicts high resolution radiometric distributions from low resolution photometric measurements, which ultimately reflect a combination of atmospheric intricacies. Our model has learned trends that are relatively linear in their coefficients, regardless of season, time of day, exposure, color model and image compression, as shown in our experiments. All of this suggests that highly turbid clear skies won't dramatically affect spectral variation, so long as some are included in training data.

As replied to Reviewer 2 above, the most obvious weakness of our dataset is that it was captured at a single location on Earth. We are currently working with other researchers to expand this dataset to 3 additional locations. But in general, we believe that scans of the sky from a weathering station over the course of one year can yield a good enough model for an entire region.

As stated, AOD, and any other atmospheric measurement desired, can be included as a training/input feature in our pipeline without hassle. If anything, including such a measurement should slightly improve predictions and make them more robust if used in other locations and climates, as an additional weight is tied to the expected radiance. Though, it may make no difference at all under most clear sky conditions. We mentioned this, but now further clarify this in our Introduction. We didn't capture AOD. This does not discount our progress for clear skies. We provide an additional stepping-stone to future researchers, and our work demonstrates exact opportunities for future work.

Concern: not applicable to cloudy skies

I don't believe the method developed by the authors is applicable to inhomogeneous cloud arrays with isolated clouds randomly distributed over the sky (as author indicated in Conclusions.

clouds scatter in geometric optics regime because water droplets are very large compared to the visible wavelengths. This has a large unknown impact on colors, because the optical signal detected is a superposition of color-neutral signal from clouds and spectrally-dependent light scattering in a cloud-free atmospheric windows (or in an undercloud atmosphere). The resulting radiance is then difficult to predict under elevated turbidity conditions.

Response

Again, we agree, for the regression models and input features presented in this work.

Our regression model approach is currently not unified. Separating clear, scattered, and overcast skies has been discussed in many papers (clear-sky index, fractional cloud cover, R/B ratio, colormetric and spectral combined metric, etc.). There are two valid procedural approaches, either categorize the entire sky into buckets of clear, scattered, overcast, few, broken, etc. and pass to an appropriate model, or separate clear from cloudy bits of the sky and pass samples/colors to separate models. If our methods help for estimations on clear skies only, that is progress. For the applications we mentioned, it is completely viable (and at this point necessary) for multiple models to be used to process all skies.

Note that we previously showed that our simple RFR regression model made some progress on scattered skies. We now refer and include some of this work to better highlight this point.

It is also possible of course that a data-driven approach could find a unified solution. If you imagine a training data set from many points on Earth, and including atmospheric measurements, it is very possible that a more complicated (deep learning in nature) data-driven approach could learn these non-linear relationships. There are so many measuring stations and satellites these days, that a concerted collaboration to collect and normalize such data for a comprehensive dataset could lead to this solution. We mention this in the Conclusion.

The training and prediction can also be divided easily per spectra, perhaps between visible and non-visible, or some other partitioning. Toshing et al. used a model per wavelength. Other (fewer) partitions are possible.

This does not discount our progress for clear skies. We provide an additional stepping-stone to future researchers, and our work demonstrates exact opportunities for future work.

Concern: method description is too qualitative/objective

The method is described mostly qualitatively, but more objective details are missing. The authors should be more specific in describing the methodology. The graphics of something like the flow diagram would be welcome.

Response

A more detailed explanation of machine learning methods and feature selection is now provided in Section 4. A graphical flow diagram was provided in our SPIE paper. A simpler flow diagram is now provided in Section 4.

Concern: UniSky Simulator

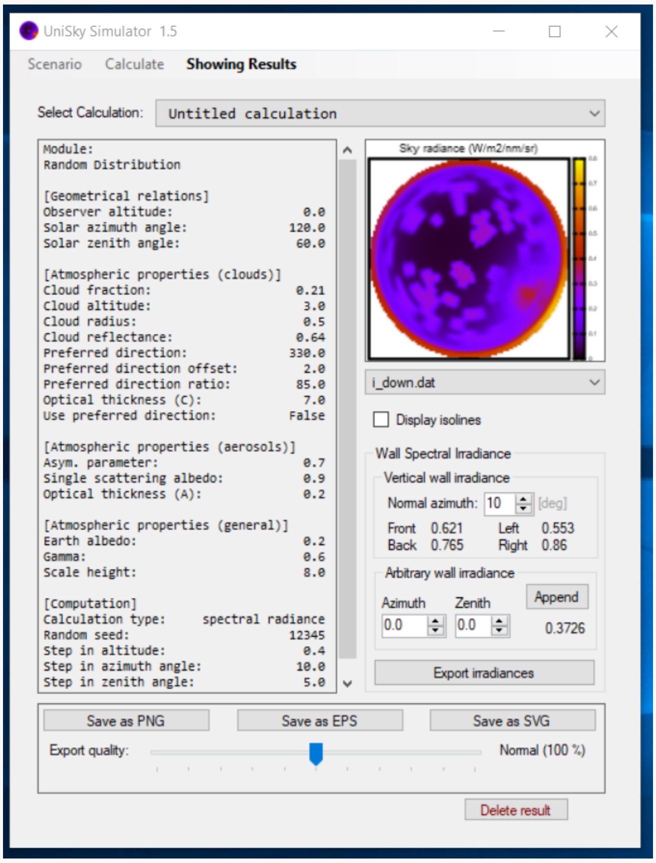
line 91: Did the authors made any computations using the UniSky Simulator? The theory is complex, but the tool is extremely fast because of parallelization implemented.

Response

UniSky Simulator (and Kocifaj's work in general) looks very interesting and detailed. We cited several papers from the author regarding physically-based solutions. As mentioned, given consistent, timely, accurate atmospheric measurements, an application can use them to simulate a sky for use in downstream calculations (building control system decisions, renders, etc.). Is such a pipeline of using highly accurate atmospheric measurements to estimate spectral radiance (not irradiance) distributions being used by any building control systems today, in order to make daylighting and HVAC decisions?

We installed UniSky Simulator and tested it. When we run the default scenarios provided, it runs very fast and the results look interesting. We are not atmospheric experts, but we assume these generated skies have been validated against real skies along the same atmospheric measurements.

However, we don't understand how to use this tool with our dataset and our current data-driven solution. We purposely did not measure the inputs required, such as AOD, cloud fraction, reflectance and altitude, ground albedo, scale height, gamma, etc. We don't believe UniSky Simulator is relevant to our current proposed solution. We are willing to work with Kocifaj et al. to understand and parse our dataset to extract metrics they are interested in perhaps for validation purposes. We developed a viewer and exporter for the data, so specific subsets can easily be collated and exported.



A run of UniSky Simulator.

Concern: conform to SI units

line 128: the common unit is W/m2/nm/sr. I am aware the authors indicated the field of view of spectro-radiometer is 1 deg, but the authors should use of basic (SI) units to avoid misunderstanding.

Response

Updated on line 159, nomenclature table, and Figure 2 description.

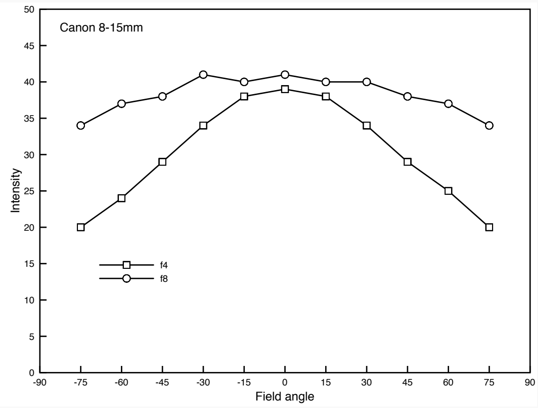
Concern: sensitivity of lens as function of zenith angle

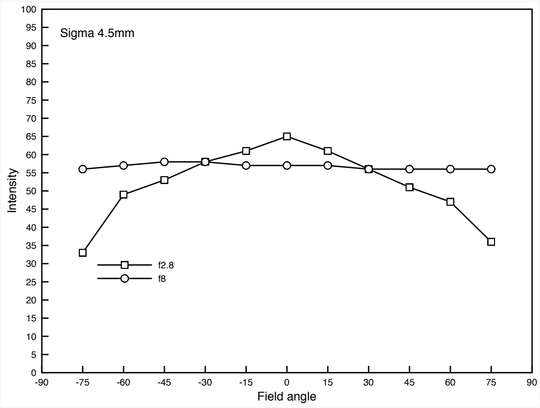
Fig 3b: what's about the relative sensitivity as a function of zenith angle? A fisheye lens system causes that the intensity of a reference point-source of light measured by a photosensitive chip may change with the incidence angle (see e.g. Fig. 3 and Fig. 5 in Monthly Notices of the Royal Astronomical Society 453, 819-827, 2015). Is this effect incorporated to the data processing?

Response

This is true. This of course doesn't affect our results, as all captures (split for training and validation) were taken by a single lens. However, as likely implied, the sensitivity per angle ratio could be included as a training and prediction feature, which might make the models more robust when used on sky photos captured between different lenses and apertures. It might also contribute nothing. It is a question worth investigating in the future.

This sensitivity is actually f-stop (aperture) dependant. Here are plots from an experiment just now to measure the sensitivity of two lenses, a Sigma 4.5mm and Canon 8-15mm (as specified, we used a Sigma 8mm in 2014). As mentioned in Section 4.1 and Figure 8 description, we used the wider f/4 aperture photos while working with the entire sky (versus the sun), which comes from experiment results by (Stumpfel et al, 2004).





Although this experiment is limited in scope and results can't be compared directly, it may indicate that smaller apertures have less variation in sensitivity. However, several of our experiments show (color model of pixels, compression vs near raw, multiple exposure) that our results are not much affected by capture technique. We state this several times in Section 4 and once in the Conclusion. We believe this is due to the seemingly "uniform" color gradient of clear skies, irrespective of capture. Our high-dynamic-range (HDR) experiment in particular showed that using sampled sky colors from 4 separate exposures, which have a much greater affect on sampled color than sensitivity per angle (literally much darker or lighter than any sensitivity variation), had very little effect on our results. As mentioned, we were expecting more of an impact from such data, but using a single exposure seems to produce the same results... on clear skies. We mention that we expect multiple exposures to provide more insight on the relationship between scattered and overcast skies and their underlying energy, but this is currently unknown.