

Real-time spectral radiance estimation of hemispherical clear skies with machine learned regression models

Joseph Del Rocco^{a,*}, Paul D. Bourke^b, Charles B. Patterson^c and Joseph T. Kider^a

^a*IST, School of Modeling, Simulation, & Training, University of Central Florida, Orlando, FL, USA*

^b*University of Western Australia, Crawley WA, Australia*

^c*Full Sail University, Winter Park, FL, USA*

ARTICLE INFO

Keywords:

sky radiance
spectral radiance
all sky
machine learning
building performance
HDR

ABSTRACT

Whole sky spectral radiance distribution measurements are difficult and expensive to obtain, yet important for real-time applications of radiative transfer, building performance, physically based rendering, and photovoltaic panel alignment. This work presents a validated machine learning approach to predicting spectral radiance distributions (350-1780 nm) across the entire hemispherical sky, using regression models trained on high dynamic range (HDR) imagery and spectroradiometer measurements. First, we present and evaluate various measured and engineered machine learning features. Next, we perform experiments comparing regular and HDR imagery, sky sample color models, and spectral resolution. Finally, we present a tool to generate spectral radiance maps for use in building performance and spectral rendering pipelines. The spectral radiance of 81 sample points across the sky are estimated to within 7.5% RMSD at 1 nm resolution. Results are validated against libRadtran and measured spectroradiometer data.

1. Introduction

Multispectral atmospheric radiance distributions (UV, IR and visible spectra) for the entire sky are often simplified into a single downwelling irradiance measurement, mainly because whole sky spectral radiance is difficult and expensive to measure in real-time, and complicated to model. Yet they are still very much needed for accurate calculations in real-time applications of building performance (Hensen and Lamberts, 2012; Chandrasekhar, 2013; Jakica, 2017), environmental science (López-Álvarez et al., 2008), PV alignment (Smith et al., 2016), and physically based rendering (Jakob, 2010; Hosek and Wilkie, 2012; Satyilm̄s et al., 2016). Unlike irradiance, spectral radiance is directional and should be available for any point in the sky, as simulations are effected by the angle of incidence of sky energy and collision surface.

We present a data-driven, machine learning approach to estimate spectral radiance for any point in a clear sky to within acceptable tolerances for real-time applications. We use HDR photographs of the sky and validated spectral radiance measurements captured throughout an entire year by a custom sky scanning framework (Kider et al., 2014), to train models that learn a relationship between capture time, look of sky, and its underlying energy (350-1780 nm). The primary contribution of our research is the reconstruction of high-dimensional atmospheric spectral radiance for the entire sky, including non-visible spectra, UV and near IR, given only a low-dimensional RGB photograph of the sky and its capture time.

Notable previous data-driven approaches to model sky-light include Tohsing et al. (2014), Saito et al. (2016), and

López-Álvarez et al. (2008); Cazorla et al. (2008a,b). Tohsing et al leveraged ground-based sky radiance photographs and a non-linear regression model per wavelength to reconstruct the visible spectrum. Saito et al. used total ozone column readings, camera color matching functions, and a linear algebra approach. Cazorla et al. used neural networks, genetic algorithms, and regression models. Our proposed approach considers and estimates a much wider spectral range, and does so for the entire sky. In contrast to more traditional atmospheric science models, we purposely omit aerosol particulate and trace gas measurements to test viability of our method for real-time applications (e.g. commodity building monitoring systems, residential solar installations, rendering pipelines, etc.), which may or may not have access to such data to compute a physically-based approach. This work focuses on clear skies specifically; previous work (Del Rocco et al., 2018) included preliminary analysis of cloudy skies.

In this work, four separate regression models are developed with a machine learning approach by feeding a combination of measured and engineered features from correlated sky imagery and spectral radiance. A series of experiments are performed to test model effectiveness and efficiency with regards to changes in exposure, sky sample color model, and spectrum resolution. A tool is developed using a model to predict spectral radiance distributions for the entire sky, at 1 nm resolution, to within 7.5% RMSD overall. Results are validated against libRadtran (Emde et al., 2016).

The remainder of this paper is organized as follows. First, related work is presented in Section 2. Our measurements and engineered data is detailed in Section 3. We explain our methods and experiments in Section 4, results in Section 5, and validations in Section 6. Finally, conclusions and future work are presented in Section 7.

*Corresponding author.

 jdelrocco@ist.ucf.edu (J. Del Rocco)

ORCID(s):

Nomenclature

$L_{e\Omega\lambda}$	spectral radiance distribution ($W/m^2/sr/nm$)
($P\theta, P\phi$)	sky point of interest (azimuth, altitude) ($^\circ$)
($S\theta, S\phi$)	sun location (azimuth, altitude) ($^\circ$)
(x, y)	sky image pixel coordinate
σ	standard deviation
SPA	sun point angle ($^\circ$)
ETR	extra trees regression model
RFR	random forest regression model
KNR	k-nearest-neighbor regression model
LNR	linear regression model
R^2	coefficient of determination score [-1, 1]
RMSD	root mean squared deviation (%)

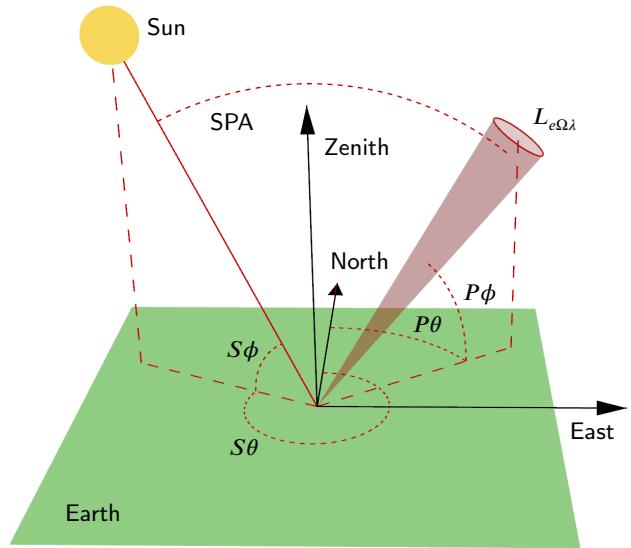


Figure 1: Atmospheric spectral radiance ($L_{e\Omega\lambda}$) is measured from the ground by a custom sky scanning system at sky coordinates ($P\theta, P\phi$) (azimuth, altitude). The sun is located at coordinates ($S\theta, S\phi$). The central angle between sun and sky point of interest is denoted as SPA.

2. Related work

Skylight itself has been studied for well over one hundred years (Strutt, 1871; Mie, 1908). Skylight simulation models typically fall into one of three categories. Early work often simplified solar and sky models by simulating luminance distributions and salient color characteristics with simple analytical equations. Later, the atmospheric science and computer graphics communities proposed brute-force physically-based simulations of light transport in the atmosphere. More recently, researchers have attempted to model skylight with data-driven approaches, which often measure, process, and quantify large sets of data and search for correlations. Modern atmospheric measuring systems installed at labs around the world are powerful and accurate, but often expensive and slow, and thus commodity sky scanning systems are more feasible for modern building performance solutions.

2.1. Analytical methods

Analytical skylight models fit parametric functions to observations of the sky (Pokrowski, 1929; Kittler, 1994). Such models were standardized by the CIE to calculate the spatial distribution of skylight, and are based on measurements of luminance, indirect sky irradiance, and direct solar radiance. Early analytical approaches include the Intermediate Sky by Nakamura et al. (1985) and the BRE average sky by Littlefair (1981). Lee Jr (2008) studied overcast skies to find meridional consistencies. Cordero et al. (2013) studied albedo effect on radiance distributions (both upwelling and downwelling). One of the most popular analytical models is the all-weather model by Perez et al. (1993), which formulated a mathematical equation with five coefficients to model sky luminance. This model was extended by Preetham et al. (1999) to calculate sky color values by fitting equations to a brute-force physically-based simulation. Hosek and Wilkie (2012) made several improvements including ground albedo, more realistic turbidity, and the handling of spectral components independently. Igawa and Nakamura (2001) and Yao et al. (2015) also improved the Perez all-sky model. All of

these models produce realistic looking results, but often suffer from inaccuracies (Zotti et al., 2007; Kider et al., 2014; Bruneton, 2016).

2.2. Physically-based methods

Physically-based skylight methods produce the highest quality results of simulating skylight. They directly calculate the transfer of solar radiation in the atmosphere through the radiative transfer equation. They also directly calculate the composition of the atmosphere through Rayleigh and Mie scattering, and polarization. The atmospheric research community developed programs such as 6SV (Vermote et al., 2006), SMARTS2 (Gueymard et al., 1995), MODTRAN (Berk et al., 2014), and SBDART (Ricchiazzi et al., 1998), which produce accurate results, but often at high computational cost unsuitable for real-time applications. They also tend to focus on luminance and irradiance. libRadtran (Mayer and Kylling, 2005; Emde et al., 2016) is an accurate program used in the atmospheric science community that includes many radiative transfer solvers to estimate atmospheric spectral radiance (among other solar and sky properties) across the UV, visible, and IR spectrum, and is highly configurable. We use it to validate our approach. libRadtran often requires using aerosol parameters and distributions (Hess et al., 1998; Holben et al., 1998) to produce the most accurate simulations. An alternative approach includes an even more complicated multi-scattering computation (Kocifaj, 2009, 2012, 2015). The computer graphics community also has developed numerous Monte Carlo based approaches (Nishita et al., 1993, 1996; Haber et al., 2005; Jarosz, 2008) to solve radiative transfer with the rendering equation (Kajiya, 1986). These methods produce pleasing visual results and often approximate the complicated scattering calculations with phase

substitutions by Henyey and Greenstein (1941) or Cornette and Shanks (1992).

2.3. Data-driven methods

Data-driven approaches systematically gather measurements and search for a relationship to model and simulate skylight. This includes capturing high-dynamic range (HDR) imagery (Stumpfel et al., 2004) and image-based lighting to produce luminance values for the sky directly from captured photographs.

The most relevant work to our own comes from Tohsing et al. (2014), who used 1143 separate machine learned regression models (one per color component (RGB) per wavelength of the visible spectrum (380-760 nm)) to estimate whole sky radiance. The authors used 113 samples from one clear sky day (all skies captured within a 12 day period). Whole sky scans took 12 minutes to complete, and thus a synthetic image was used for color sampling. Our approach captures and predicts a wider spectrum (350-1780 nm), including some UV and IR, and uses data over an entire year to account for seasonal variation.

Saito et al. (2016) improved upon the work of Sigernes et al. (2008) to estimate sky radiance, specifically “*without any training sets*,” by using an equation of total ozone column and raw sky image RGB counts. They focused on the zenith of the sky (single point) and estimated spectral radiance for a subset of visible wavelengths (430-680 nm). A notable contribution is the color matching functions, which took into account camera lens wavelength dependence, vignetting, and CMOS noise, and were used for cloud detection in Saito and Iwabuchi (2016).

Artificial neural networks (ANN), genetic algorithms, and pseudoinverse linear regression models were used in various projects by López-Álvarez et al. (2008); Cazorla et al. (2008a,b). They also used a custom sky scanner. Their models focused on visible spectra with a final dataset of 40 samples. More recently, Satylmýs et al. (2016) used an ANN to model certain properties of skylight.

Chauvin et al. (2015) used a custom sky imaging framework for irradiance and cloud detection for the purposes of concentrating solar plant technology. A noted contribution was their observation of the importance of the circumsolar region, in opposition of many sky models, and the central angle between sun position and sky point of interest (SPA). Their research was used for intrahour forecasting to improve solar resource acquisition (Nou et al., 2018).

Our research: (1) reconstructs the spectral radiance of the sky utilizing high resolution imagery, (2) accounts for seasonal and datetime variation with captures throughout the year, (3) adjusts for lens warp, (4) predicts a wide, useful spectrum of energy (350-1780 nm) at 1 nm resolution, (5) predicts non-visible spectrum energy with indirect visible data, (6) does so for the entire hemispherical sky, (7) considers multiple exposure imagery, color model, and spectral resolution, (8) considers real-time constrained applications, (9) compares multiple regression models, and (10) validates against a modern atmospheric radiative transfer package.

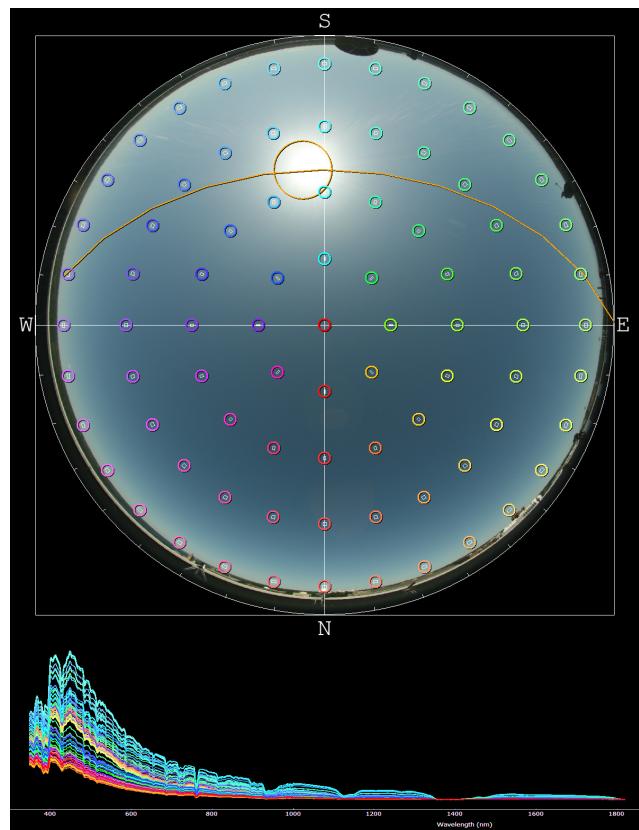


Figure 2: A single sky capture consisted of high-resolution HDR imagery (8 exposures) and 81 spectral radiance measurements between 350-2500 nm (350-1780 nm used). Colors merely correlate radiance distribution with sky location; radius of colored circles is not to scale with sampled pixel area.

3. Measurements and data

Measurements in this work come from the sky scanner discussed in detail by Kider et al. (2014). This framework captured high-resolution HDR imagery of the sky (8 exposures), along with atmospheric spectral radiance distributions (350-2500 nm) from 81 sample points in concentric circle patterns across the sky. Measurements were taken from the ground. The spectral radiance distributions were measured in $\text{W} / \text{m}^2 / \text{nm}$ with an ASD FieldSpec Pro spectroradiometer through a 1° solid angle fore-optic (Malthus and MacLellan, 2010), and were validated against NASA data sets (Kider et al., 2014). The multiple exposure photographs of the sky were captured in both CR2 (raw) and JPG formats consecutively at 4368 x 2912 pixels with a commodity Canon 5D digital single-lens reflex (DSLR) full-frame camera, with Sigma 8 mm f/3.5 EX DG circular fisheye lens, and Kodak Wratten neutral density filter. JPG quality level was set to 100. We automated the process with libgphoto2, which took approximately 40 s to capture all exposures and formats of photographs of the sky. Irradiance was also measured, but ignored for the purposes of this work.

All measurements were taken at a single site location, (42.44344, -76.48163) decimal degrees, rooftop of Frank Rhodes Hall, Cornell University, Ithaca, New York, USA.

453 total captures were taken over 16 days between 2012–2013, covering all four seasons, dawn to dusk, and various sky covers. About 25% of those captures consisted of full clear skies (0 octas), from which 6006 individual clear sky samples were used for this work. Scattered and overcast skies were ignored for this work. A complete table listing of all usable captures can be found in Del Rocco et al. (2018). This data is available to the public through our lab website.

Hemispherical sky coordinates in this work are specified in (azimuth, altitude) coordinates, where azimuth is an angle Eastward from North, and altitude is (90° – zenith). Sky imagery is vertically flipped due to capture orientation.

3.1. Lens linearity

Because our work involved mapping hemispherical sky coordinates to 2D pixel coordinates, and vice versa, it was important to accurately model the behavior of the fisheye lens employed. In a perfect circular fisheye lens, often called a "tru-theta" lens, equal increments in radius on the fisheye image correspond to equal angle increments of the respective field rays. Actual fisheye lenses typically exhibit some form of non-linearity, even those lenses designed to be linear (Bourke, 2016). Although more important with variegated skies (scattered, overcast, etc.), a measurement difference of even a single degree can result in sampling pixels in or out of the sun's corona. The standard ideal lens equation for mapping hemispherical sky coordinates to 2D center offset coordinates can be written as:

$$(x, y) = \frac{2 \cdot \text{zenith}}{\text{fisheyefov}} \cdot (\cos(\text{azimuth}), \sin(\text{azimuth})). \quad (1)$$

The following procedure was employed to measure the relationship between field angle and position on the fisheye image.

1. A close and distant vertical feature in the fisheye image was chosen. The zero parallax position of the lens is the position along the lens axis where those features stay aligned despite rotations perpendicular to the lens axis.
2. A clear narrow object in the image was chosen as a reference point and aligned with the center of fisheye image.
3. The lens is rotated in 5° steps from 0 to 90° , and a photograph taken.
4. For each photograph, the distance of the reference point from the center was measured.

For our Sigma 8 mm f/3.5 fisheye lens, this resulted in the following non-linear curve (plotted in Fig. 3), which was then used to alter zenith of sky coordinates ($r = \text{zenith}$):

$$r' = 0.7230r + 0.0252r^2 - 0.0499r^3 - 0.0004325r^4. \quad (2)$$

3.2. Sky color

Color at a particular location in the sky is a fairly subjective measure. What our eyes detect, what instruments measure, and how that data is processed, differs dramatically.

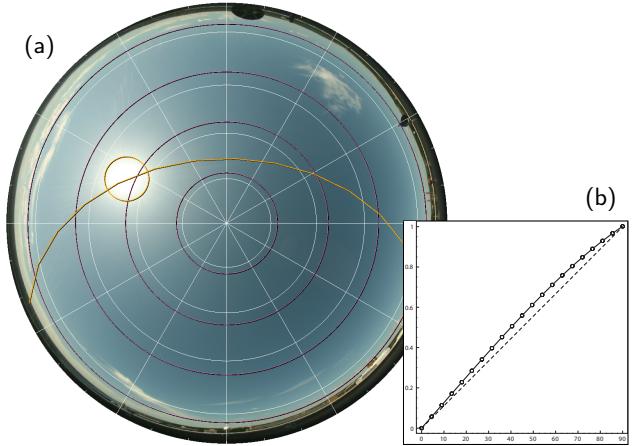


Figure 3: (a) shows the visual difference in linearity of actual lens (pink) vs ideal lens (white). Sun path and position after lens correction is overlaid (orange). (b) shows the actual lens sample points (solid) vs ideal (dashed), used to fit Eq. 2.

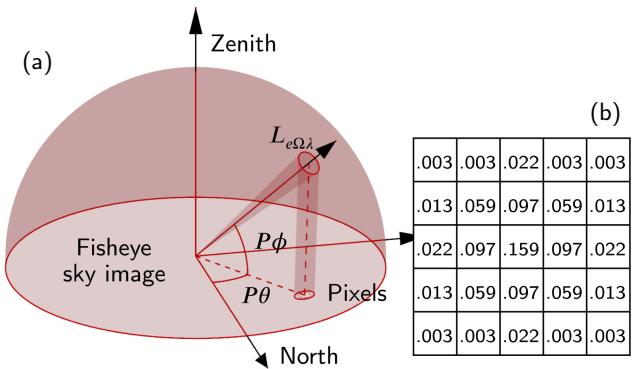


Figure 4: (a) illustrates a steridian capture area projected onto a sky image, the bounds of which contain the sky color pixels of interest. (b) shows the weights of a 5x5 Gaussian convolution matrix applied to the pixels in those bounds.

Nevertheless, our research investigates the relationship between sky color and energy distribution, and thus a metric must be used.

To obtain sky color at specific points in the sky, we projected the bounds of a 1° solid angle (same as fore-optic we used when measuring radiance) onto the 2D sky images captured with our digital camera (multiple images for the HDR experiment), and then sampled the pixels with a square convolution of similar width to the radius (Fig. 4). In image processing, a convolution is an algorithm that involves sliding a matrix of weights or values (the kernel) over a set of pixels, and performing some calculation in order to produce a new set of pixels (Parker, 2010). Such convolutions are used to implement a wide variety of image filters like blurring, edge highlighting, etc. We used a Gaussian convolution, in particular, to blend the pixel colors together, weighting pixels closer to the center higher than pixels near the edges of the projected bounds. We are aware that a square convolution does not account for all pixels in a projected circular area

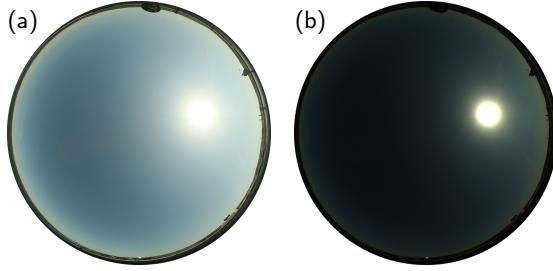


Figure 5: 05/27/2013 09:00 1s exposure of sky as a more traditional, camera processed, compressed JPG (a), and as a minimally processed, uncompressed TIFF (b). (a) is closer to what humans see when looking at the sky; (b) is a more accurate representation of what the sensor detects.

exactly (in fact the projected circle becomes an increasingly oblong ellipse as altitude decreases). A rectangular convolution kernel would likely provide better coverage of the pixels in the projected bounds. Our kernel was chosen for real-time efficiency and overlap with existing image processing techniques and libraries, most of which use a square kernel. The weights of our Gaussian kernels were generated with the following equation (Fisher et al., 1996):

$$\text{kernel}(x, y) = \frac{1}{2\pi\sigma^2} \cdot e^{-\frac{x^2+y^2}{2\sigma^2}}, \quad (3)$$

with kernel dimensions relative to the bounds of the convolution, and a standard deviation (σ) of half the radius.

3.3. Raw vs. digital positive

As mentioned, we captured photographs in a Canon CR2 (raw) format and a more traditional, camera processed, compressed JPG file format. Raw images contain much more capture information in a pre-interpolated format, before debayering, noise filtering, color space conversions, gamma correction, etc. In our previous work, we worked with the compressed JPG captures, which were smaller and faster to process (Del Rocco et al., 2018). For this work, we strove for accuracy of recorded color values and interpolated the raw photographs as uncompressed TIFFs, using camera white balance, but no other post-processing options that digital cameras use to produce images closer to what humans see (e.g. gamma correction, additive brightness, exposure shift, etc.). We used rawpy to read and process the raw images (Riechert, 2018; LibRaw). Fig. 5 shows the difference. Our previous work already showed that it is possible to infer a relationship between look of sky and spectral radiance using compressed imagery. The consistency of raw photograph interpolation may be more crucial than the specific parameters used.

4. Methods and experiments

The research question for this work asks whether it is possible (or not) to estimate the atmospheric radiance distribution of a clear sky given only a picture of the sky and its capture timestamp. In other words, is there a relationship between the time of day, what a commodity camera

sees in the sky, and its underlying spectral energy, despite the fact that we know solar radiation is absorbed and scattered by atmospheric particles at certain wavelengths? We used machine learning to help us search for such a relationship, but given the sheer magnitude of machine learning approaches (statistical models, artificial neural networks, support vector machines, etc.), we limited the scope of this research to regression models. Predicting a curve is more of a regression rather than a classification or clustering problem. More than 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. WEKA toolkit (Hall et al., 2009) was used to discover possible candidate models, but ultimately all machine learning models were configured and processed with scikit-learn in Python (Pedregosa et al., 2011).

Imperative to all machine learning algorithms is the quality and organization of training and testing data. Our data was distilled down to 6006 sky samples, each representing a single point in a clear sky along with corresponding spectral radiance measurement and capture timestamp. Our collection of sky samples was randomized and divided into an 80:20 train:test ratio. 10-fold cross-validation was used during training to dampen the effects of outliers (Picard and Cook, 1984; Kohavi et al., 1995). The samples from four arbitrary skies, selected at random, served as absolute hold-outs (Table 1); i.e., no samples from these skies were ever used for training or preliminary testing. These techniques are often employed to help minimize overfitting through data leakage.

Table 1
Four holdout test skies selected at random.

Date	Time	Part of Day	Season	Sky Cover
05/26/2013	15:15	Afternoon	Spring	CLR
05/27/2013	10:15	Morning	Spring	CLR
07/26/2013	13:15	Midday	Summer	CLR
09/24/2013	15:39	Afternoon	Fall	CLR

Each sky sample was engineered into an array of input and output features. Assuming a relationship between time of day, sky color, and radiance distribution, one of our goals was to discover the minimum set of features needed to show that relationship. From the raw measurements of capture timestamp, sample azimuth and altitude, sky color, and corresponding spectral radiance measurement, we engineered the additional features: sun azimuth and altitude, sun-point-angle (SPA), quarter, month, week, day and hour. The capture timestamp was “binned” into these discrete groupings to help the models internally correlate data throughout the year, as opposed to treating the timestamp as a single number or string (Macskassy and Hirsh, 2003). Sun position was computed with the solar position algorithm provided by NREL (Reda and Andreas, 2004). SPA comes from the insights of Chauvin et al. (2015), in contrast to our initial approach of culling all samples within a 20° circumsolar region

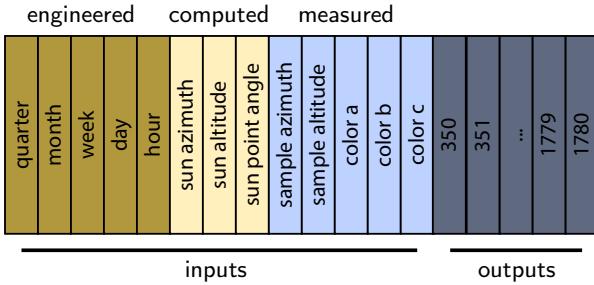


Figure 6: A single sky sample consists of 13 input features and 1430 output features (the spectral radiance curve between 350-1780 nm). Capture timestamp was binned into separate features; sun azimuth and altitude were computed via NREL sun position algorithm; sample azimuth and altitude were inherent to sky scanning logic; and a three component color was extracted from sky image relative to color model.

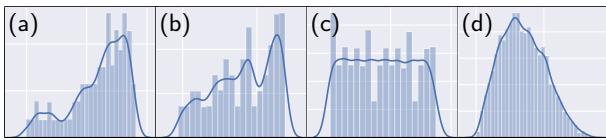


Figure 7: Histograms of input features: sun azimuth (a), sun altitude (b), sample azimuth (c), and SPA (d). (d) was much more significant than (c), likely because it combined the positions of both sun and sample points into a single feature.

(Del Rocco et al., 2018), like authors Saito et al. (2016) and Tohsing et al. (2014). The final input and output features of each sky sample are shown in Fig. 6.

Various exploratory data analysis (EDA) techniques were employed to gauge the significance of each possible input feature, including: histograms, correlation matrix, feature importance, etc. (Yu, 1977). While sky colors were found to be the most significant features, sample azimuth was found to be the least. As Fig. 7 shows, 81 samples per capture evenly distributed across the sky resulted in a nearly flat distribution of sample azimuth values. The SPA feature was a combination of both sun and sample locations in a single feature, and was much more significant.

Preliminary results encouraged us to focus on the following regression models: linear regression (LNR) (provided only as a baseline), k-nearest-neighbors (KNR), random forest (RFR), and extra-trees (ETR). Although the performance of RFR and ETR are often comparable, in prior experiments we found that RFR performed significantly better on scattered cloudy skies (Del Rocco et al., 2018). RFR and ETR are ensemble (decision tree) regression models, which implement a set of “if-then-else” rules internally for both training and prediction. This is why the model file sizes are so large. For all four of our models, tuning was done automatically with scikit-learn’s GridSearch algorithm, though some hyperparameters were tuned manually, including the number of trees and maximum tree depth of ensemble models.

Four separate error metrics were used to evaluate performance, including: coefficient of determination score (R^2),

mean bias deviation (MBD), root mean squared deviation (RMSD), and ratio of measured and predicted radiance curves. MBD and RMSD come from Iqbal (2012). Prior authors used MBD for single wavelength results (Tohsing et al., 2014; Cazorla et al., 2008a), but we found RMSD to be more appropriate for a spectrum of wavelengths. The R^2 metric from scikit-learn is calculated as follows:

$$R^2(t, p) = 1 - \frac{\sum_{i=1}^N (t_i - p_i)^2}{\sum_{i=1}^N (t_i - \bar{t}_i)^2}, \quad (4)$$

where (t, p) is a (truth, prediction) pair, N is the number of radiance distributions, and $\bar{t} = \frac{1}{N} \sum_{i=1}^N t_i$. Note that this metric can be negative, despite the name R^2 .

Along with a useful tool for viewing and exporting correlated sky data with many options, we developed a framework of Python scripts to send datasets through our machine learning pipeline of training, testing and plotting. The main script takes parameters such as: model type, dataset of sky samples, pseudo-random number seed, number of cpu cores to use, cross-validation amount, and model specific hyperparameters like polynomial expansion, maximum tree depth, etc. All of our code is 100% cross-platform, open-source and freely available to the public through our lab website.

4.1. High-dynamic range imagery

Simultaneously capturing the sun and sky with photography is difficult due to the range of illumination and intensity of the sun vs. sky, as well as the temporal changes that occur. Our dataset of sky imagery followed the capture approach proposed by Stumpfel et al. (2004). We took eight to nine photographs (depending on the time of day) to capture ~ 17 stops of dynamic range. Fig. 8 shows the difference in exposures captured. The top row (f/16 f-stop) accounts for the solar region and intensity of the sun. The bottom row (f/4 f-stop) captures the indirect skylight.

This experiment was designed to test the differences and effectiveness of using HDR imagery vs. a single exposure. For each sky sample’s color feature, we used the RGB color values from exposures 5-8 (f/4 f-stop) as input features for model training and prediction. Exposures 1-4 were ignored for this experiment. Although there are algorithms to merge multiple exposures into a single image for sampling, we simply sampled each exposure separately and used each sampled color as a separate input feature.

4.2. Color model

Colors are qualia for combinations of electromagnetic energy within the range of wavelengths visible to humans (the visible spectrum). The human eye detects energy with the use of retinal rods and cones and the brain merges the results into what we call a color (Kinney, 1958). Modeling the values of these colors is a field of research in and of itself (Koenderink, 2010). And yet, we are attempting to estimate spectral radiance using color values as a primary feature. This begs the research question: which color model best represents the underlying energy? Digital all-sky cameras typi-

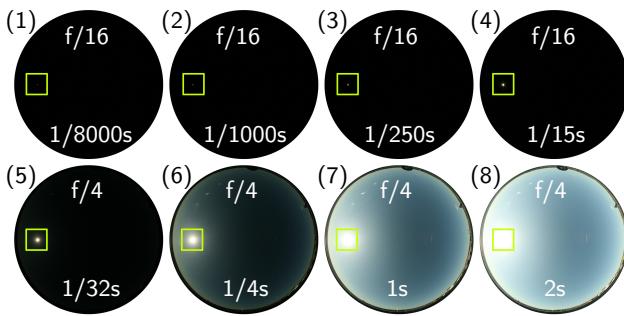


Figure 8: 8 exposures were taken to account for high dynamic range of sun + sky photography. F-stop f/4 captures (5-8) were used for this work. 1 s exposure (7) was used for non-HDR experiments. Yellow squares highlight sun location.

cally store measurements with trichromatic RGB color models (e.g. sRGB, Adobe RGB, ProPhotoRGB, etc.), but do so mostly for historical reasons relating to technology. There are a variety of other tristimulus color models that attempt to capture more of the color space detectable by the average human (Poynton, 1995; Stone, 2015), many of which derive from the CIE 1931 RGB and XYZ color space definitions (Wright, 1929). However, it is unclear which model is most beneficial for machine learning algorithms processing sky images.

For this experiment, we compared the overall training and predictive effectiveness of our models while only changing the color model used for each sky sample's color feature. Four separate color models were tested: sRGB (Stokes et al., 1996) (the default), HSV (Smith, 1978), HSL (Joblove and Greenberg, 1978), and LAB (Robertson et al., 1977). All other features were fixed. Because our commercial digital camera captured skies in an sRGB format, we then converted to the other color models using algorithms provided by the Python colormath module. The resulting datasets were fed through our machine learning pipeline separately.

4.3. Spectral resolution

This work will be used in a real-time setting, both simulated and cyber-physical, therefore model size and processing speed is important. For applications that predict a general quantity of energy in certain parts of the spectrum, it may be reasonable to limit the resolution of spectral data used during model training and prediction. Certainly the visual difference and area under the curve (amount of energy) between a 1 nm and 10 nm resolution curve is not significant. A spectral resolution experiment was designed to find the smallest model and dataset that still predicted with acceptable accuracy, by training and testing models using spectral resolutions of 1, 5, 10, 15 and 20 nm. Note that some pure spectral colors exist entirely within a 15 nm range, and therefore resolution should not be diminished too much if color information is important. Fig. 9 shows the visual difference of the five resolutions for a single measured radiance distribution. Depending on the downstream application, there is still plenty of useful information at lower resolutions.

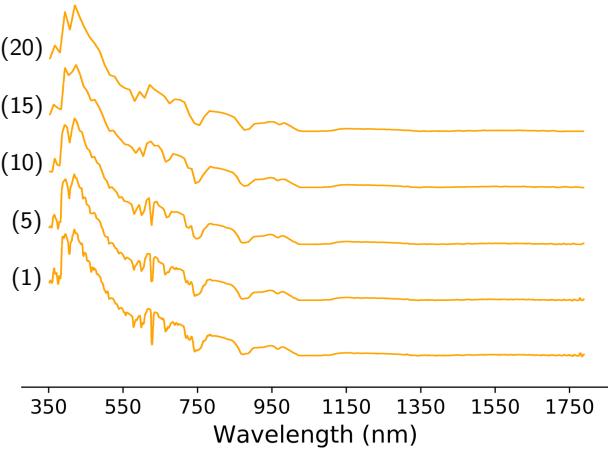


Figure 9: 05/26/2013 15:15 sample 24 (90° azimuth, 12.12° altitude) plotted at 5 different resolutions, 1, 5, 10, 15 and 20 nm, labeled accordingly.

This experiment was run on a Dell XPS 8920 PC with Intel 4 Core i7-7700K 4.20 GHz CPU and 16 GB of RAM. The operating system was x64-bit Microsoft Windows 10 Enterprise. All manually executable applications (i.e. ignoring OS services) were closed at the time of the experiment. Five runs were executed per resolution size and the timings averaged.

4.4. sradmap

Ideally our models should be able to generalize across the space between the sky samples used for machine learning. The same input features shown in Fig. 6 can be collected for any pixel of a sky image, and then fed through our models to produce a lookup file (map) with radiance predictions per pixel. We call this resulting file a spectral radiance map (sradmap). Although the primary purpose of these files is to provide a map between pixel location and spectral radiance prediction, each prediction can be summed, normalized, and plotted against a false-color map to help visualize the topology of the data.

The name sradmap is an homage to radmap by Anselmo and Lauritano (2003), a supplementary tool for the daylight simulator RADIANCE (Ward, 1994). In the building performance space, our sradmap generator can be integrated into daylight simulators, energy modelers, and parametric design tools like RADIANCE, EnergyPlus (Crawley et al., 2001), SUSTAIN (Greenberg et al., 2013), and Ladybug Tools (Roudsari et al., 2013). In the rendering space, sradmaps can be sampled from renderers like Mitsuba (Jakob, 2010) and Hy- perion (Burley et al., 2018), for use in scenes with natural daylighting.

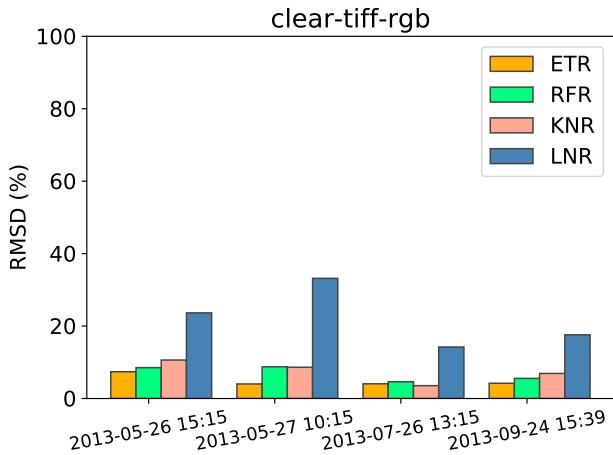


Figure 10: Model results on each of the four holdout test skies listed in Table 1. ETR performed the best, with an error of 4-7.5% RMSD. LNR was by far the worst performing.

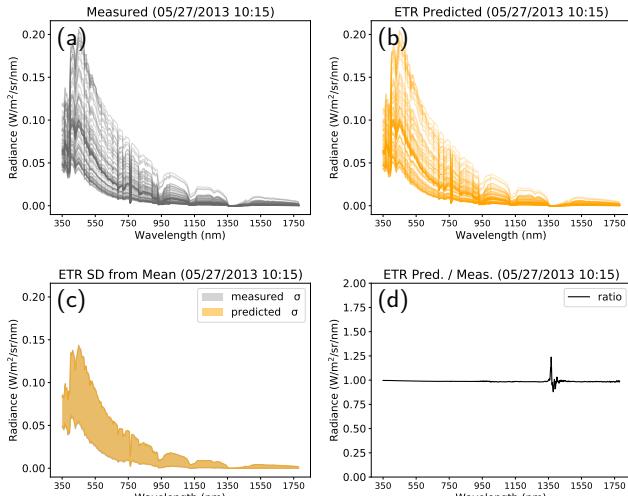


Figure 11: Whole sky results for holdout sky 05/27/2013 10:15 with ETR model; no ground truth sky samples from this capture were used for training. (a) and (b) show the measured and predicted spectral radiance distributions; (c) shows the standard deviation from mean for both measured and predicted distributions; and (d) is the overall ratio between the two. Note the error in the ratio is within the absorption band near 1350 nm, where radiance is extremely small.

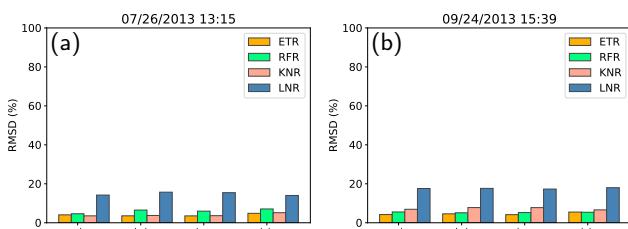


Figure 12: Sky color model made little to no difference in training and prediction results. (a) and (b) show RMSD results on 07/26/2013 13:15 and 09/24/2013 13:15 respectively.

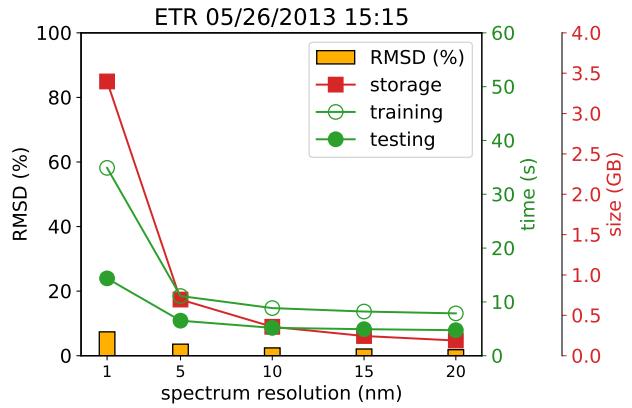


Figure 13: Limiting resolution to 5 nm drastically decreases model size, improves computation speed, and even increases prediction success, likely because the radiance curve is less noisy. Further reductions yield diminishing returns.

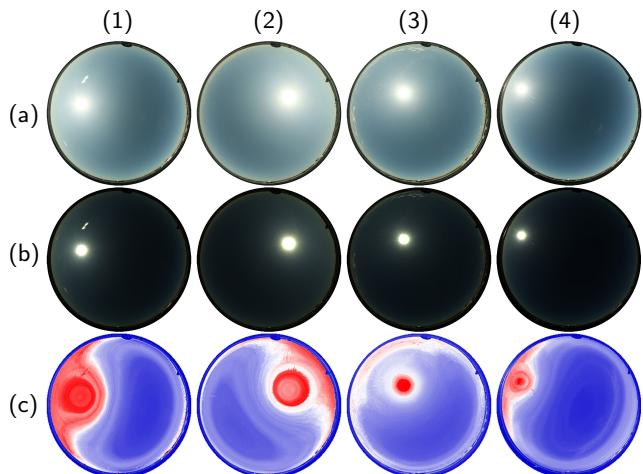


Figure 14: Columns (1-4) are the holdout test skies in Table 1, in respective order. Rows (a) and (b) show traditional, camera processed JPG and minimally processed TIFF captures, respectively. Row (c) shows the sradmaps generated for skies in row (b); we predict spectral radiance (350-1780 nm) for each pixel of each image, sum the distribution, and visualize with false-color map.

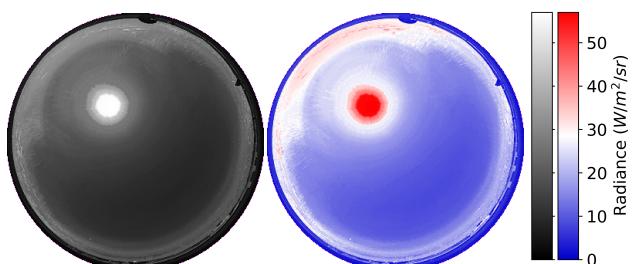


Figure 15: False-colored sradmap visualizations for holdout test sky 07/26/2013 13:15. Each pixel plotted is a summation of an entire spectral radiance distribution (350-1780 nm).

Real-time spectral radiance estimation of clear skies

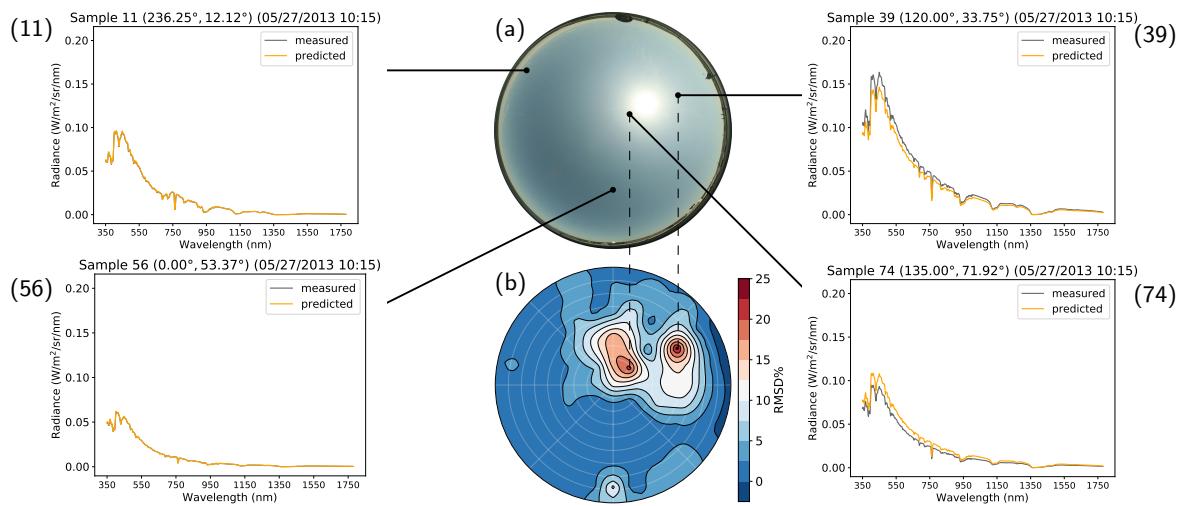


Figure 16: ETR results of four radiance predictions on holdout test sky 05/27/2013 10:15. (a) shows the camera processed JPG sky capture for convenience (the model was trained on TIFF data). (b) shows RMSD error across the entire sky. Radiance for samples (11), (56), (39) and (74) are pinpointed at their location in the sky. Samples (39) and (74) were the two worst predictions, with RMSD errors of 23.63% and 21% respectively.

5. Results

Three of the four final regression models (ETR, RFR, KNR) resulted in very high R^2 scores and acceptably low RMSD error on all holdout test skies listed in Table 1. As expected, the baseline LNR model resulted in relatively poor predictions across all test skies, with an overall error of 14–24% RMSD. By contrast, ETR resulted in 4–7.5% RMSD. For test sky 07/26/2013 13:15, three of the four models predicted within 4% RMSD. In general, the tree-based models (ETR and RFR) perform better than the nearest-neighbor model (KNN). RMSD results for all models on each test sky are shown in Fig. 10.

Fig. 11 shows a comparison of all 81 measured and ETR predicted radiance distributions, their standard deviations, and overall averaged ratio between measured and predicted on test sky 05/27/2013 10:15. The standard deviations of measured and predicted are nearly identical, and the averaged ratio is nearly 1.0. Note the noisy error in the ratio graph resides within an H_2O and CO_2 absorption band, where atmospheric radiance is extremely small (Lacis and Hansen, 1974), and measurements are susceptible to noise.

For the same holdout test sky (05/27/2013 10:15), Fig. 16 shows ETR prediction error across the entire hemispherical sky, and highlights the two worst spectral radiance predictions (23.63% and 21% RMSD). Two other predictions selected at random are shown for comparison. A vast majority of the 81 samples are predicted to within 1% RMSD. Note that even with “high” error, predicted curves align with ground truth measurements in terms of shape. The models therefore have learned the wavelength relative intensities of the sky in accordance with capture time, sun location, etc. This is consistent with nearly all predicted results; while the magnitudes per wavelength sometimes deviate, the general shapes each predicted curve is accurate.

Although we were expecting some insight from providing multiple exposures of sky images, results seem to indicate that HDR data, at least for clear skies, does not improve model prediction. All HDR runs resulted in similar error to non-HDR runs. This may be because clear sky color changes are so “uniform” throughout the day, that multiple exposures lack significance. In other words, all provided exposures may have had the same color change trends. We suspect that HDR data will be more significant in predicting scattered and overcast skies, as the color variations of clouds are less uniform across exposures.

Results of our color experiment (Fig. 12) seem to indicate that color model is irrelevant to our method. This implies that our method can be used with any representation of color, as the trends in color across the sky are similar regardless of format. It is unclear if using color data initially captured in an sRGB format somehow restricted the range of the other color models after conversion. In other words, would initially capturing the sky in a color model that maps to a larger color space be better?

The results of the spectral resolution experiment (Fig. 13) show the benefits of decreasing spectral resolution from 1 to 5 nm. Not only does prediction accuracy improve, but model sizes (particularly the large ensemble models), as well as model training and prediction times, decrease significantly. The improvements in prediction accuracy are likely due to several factors: less noise in the spectral radiance distribution, and an overall simpler prediction problem (fewer outputs to predict). Dataset size also decreases with reduced resolution, but is eclipsed by the largest model sizes. Beyond 5 nm resolution, further reductions result in diminishing returns. This is an important find for real-time applications, which may operate on limited embedded hardware.

We note here that results between the minimally pro-

Real-time spectral radiance estimation of clear skies

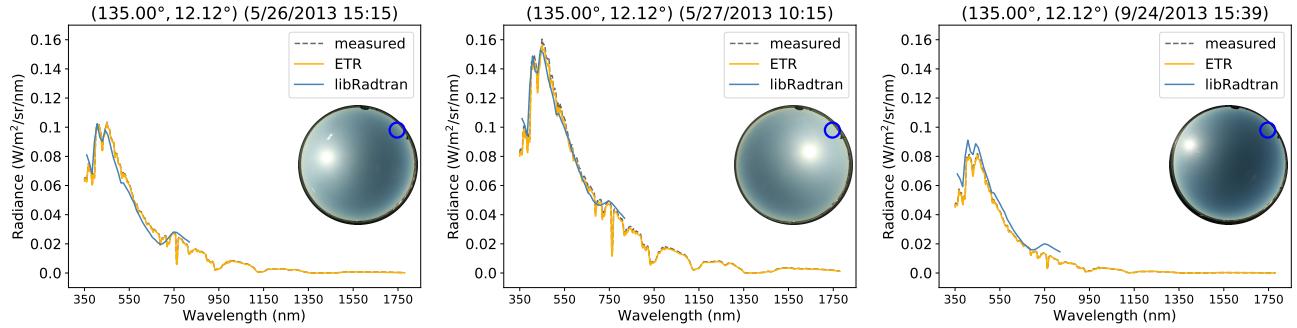


Figure 17: Spectral radiance at (33.75° azimuth, 12.12° altitude), circled, for three of the four holdout test skies. Spectroradiometer measured, ETR model predicted, and libRadtran radiance distributions are plotted. ETR model was trained on TIFF data (camera processed JPGs shown for convenience).

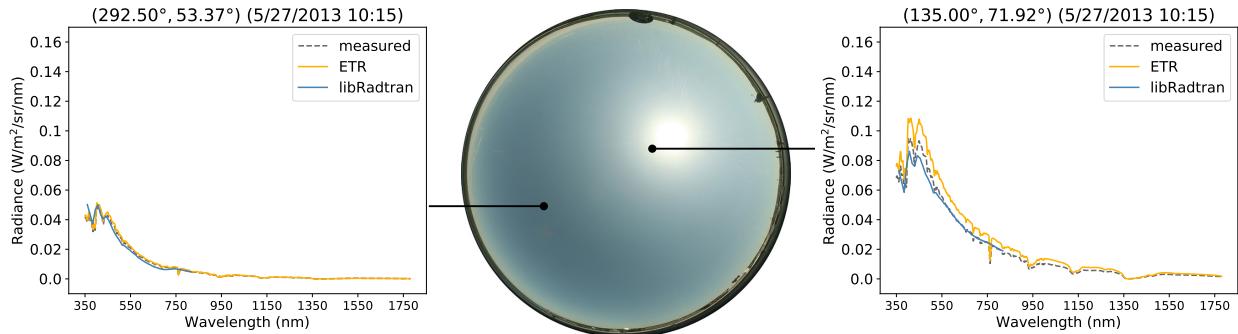


Figure 18: Spectral radiance for two sky samples of holdout test sky 05/27/2013 10:15. Spectroradiometer measured, ETR model predicted, and libRadtran radiance distributions are plotted.

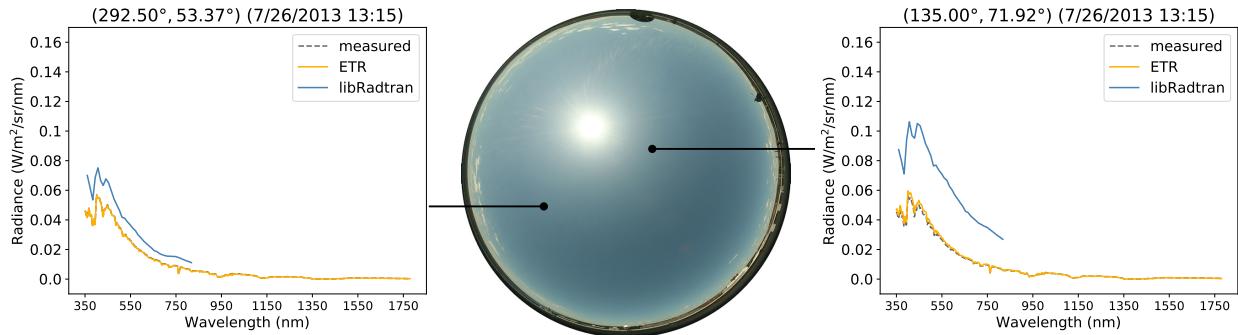


Figure 19: Spectral radiance for two sky samples of holdout test sky 07/26/2013 13:15. Spectroradiometer measured, ETR model predicted, and libRadtran radiance distributions are plotted. libRadtran calculated radiance deviates from both ETR predictions and measured ground truth data, likely because of the lack of needed atmospheric configuration data.

cessed, uncompressed TIFF sky images and traditional, camera processed, compressed JPG sky images, were roughly the same. TIFF color data resulted in slightly better results ($\sim 1\%$) on some skies, though this may be within the standard deviation of prediction error and machine learning random fluctuation. Since the TIFF images (~ 35 MB) are at least 1000% larger than the JPG images (~ 2.5 MB) compressed with quality level 100, and the results are similar, we recommend the use of JPG images in real-time applications of our method.

Spectral radiance files (sradmmaps) are the culminating

output of our method. They are generated by extracting and engineering features per pixel of test skies (Table 1) and feeding them through one of our models. Linear scale false-color visualizations of ETR model predicted sradmmaps are shown in Fig. 14 and Fig. 15. Test sky images were first scaled down to a resolution of 333x333 pixels, to anticipate real-time processing speeds. sradmap generation, visualization, and logged output took ~ 20 s to complete on the same machine specified in Subsection 4.3; embedded hardware would likely take longer, though various optimizations can be made to the process.

6. Validation

First, no samples from our holdout test skies (Table 1), chosen at random, were used during training or preliminary testing of any model. Machine learning projects often use this method to validate a model's ability to generalize over unforeseen data. The results presented in Fig. 10, Fig. 11, and Fig. 16 show that our models have this ability. The results of our additional experiments show that our method is robust against implementation details such as image compression, exposure, and color model.

Next, the sradmaps presented in Fig. 14 and Fig. 15 are the result of using every pixel per test sky. These maps demonstrate that our models have the ability to generalize across the entire hemisphere (i.e. predict spectral radiance for every point in the sky) even when trained on a mere skeleton of samples (81 concentric 1° steradians). Note that most of the sky is unaccounted for by the skeleton, including points beyond the variance of sun and sky coordinates. sradmaps contain predictions for the entire sky.

Finally, we compare ETR model predictions with ground truth measurements and the atmospheric spectral radiance distributions computed by libRadtran (Emde et al., 2016). libRadtran was configured the same for all four holdout test skies. In other words, no sky-specific data (atmospheric measurements, parameters, or ranges) were specified per test sky (Hess et al., 1998; Holben et al., 1998). Fig. 17 and Fig. 18 show that libRadtran spectral radiance for three of our four holdout test skies were in alignment with both ETR model predictions and ground truth measurements. However, for test sky 07/26/2013 13:15, libRadtran deviates from both ETR predictions and ground truth measurements (Fig. 19). All tested samples for this sky show similar deviations in magnitude, but not curve shape. As mentioned, libRadtran requires accurate atmospheric data for its calculations. Because such data was not configured, and because our predictions are closer to ground truth measurements, it is possible that our ETR model learned the sky specific atmospheric conditions libRadtran needed in order to compute accurately. In particular, we note the high-level cirrus clouds along the horizon, which might indicate ice crystals in the atmosphere, and account for deviations between data-driven predictions and physically-based model calculations.

7. Conclusions

Whole sky spectral radiance distributions are needed for accurate computations in a variety of applications, and yet they are often oversimplified. Real-time capable models are needed to estimate them to within acceptable tolerances. We presented a solution that: (1) took photographs of the entire hemispherical sky and measured the incoming radiance at various points, (2) used those measurements and modern machine learning methods to train regression models, and (3) used those models to predict atmospheric spectral radiance (350–1780 nm) at 1 nm resolution for the entire sky, given a photo of a clear sky and its capture timestamp. The whole sky RMSD error of our predicted results for all four

holdout test skies, none of which were used for training, was below 7.5%, and most of the predicted spectral radiance distributions were in line with libRadtran's.

Our results show that image compression, color model, and exposure of clear sky imagery have little to no effect on our method. This implies that our solution is robust and less likely to be affected by implementation details. We also showed that our models have the ability to generalize across the hemispherical space between measured sky samples, allowing for atmospheric spectral radiance predictions for every point in a sky image.

Our trained models can be used as-is, with similarly exposed and oriented sky photos. And our methods can be reproduced to train models using new datasets. Various sky scanning systems exist which can be employed to provide more training data. Existing correlated sky imagery and spectral radiance datasets from around the world can (and should) be used with our method. Once normalized, such comprehensive datasets could lead to even more robust models (e.g. more variations of sky turbidity), especially if coupled with additional input features from GOES satellite measurements and triangulated atmospheric measuring station data. Site location coordinates and/or elevation could also be investigated as input features when using multi-site data, as there may be properties specific to regions on Earth.

Although many downstream applications of our research are possible, one immediately viable option is a building monitoring system equipped with all-sky camera that adjusts smart glazing and kinetic facades in response to spectral radiance across the entire non-occluded sky. Such a system would automatically harness (or attenuate) light and heat with more fine-grain control and accuracy than one that operates on a single downwelling measurement, and would be much more affordable and efficient than a live sky scanning system. We hope our research motivates the building performance community to further refine such a system.

Additional work should include a more focused effort on predicting spectral radiance from scattered and overcast skies using similar methods. Clear skies are only part of the required solution. Scattered skies account for the bulk of our training data, publicly available, and in general account for the more interesting energy scattering scenarios. Newer machine learning approaches (e.g. deep learning) should be used on our dataset specifically. Color sampling could be improved with rectangular convolution kernels, which capture the projected solid angle area (ellipse) precisely. More spectral radiance measurements within the circumsolar region would likely improve accuracy, as the bulk of the energy of clear skies is accounted for within that area.

Portions of this work were presented at SPIE Optics and Photonics for Information Processing XII (Del Rocco et al., 2018).

References

- Anselmo, F., Lauritano, A., 2003. Evaluation of the solar energy potential in urban settings by irradiation map production, in: 2nd International

- Radiance Workshop, Radiance Community, Berkeley, California. pp. 1–1. <https://www.radiance-online.org/community/workshops/>.
- Berk, A., Conforti, P., Kennett, R., Perkins, T., Hawes, F., Van Den Bosch, J., 2014. MODTRAN 6: A major upgrade of the MODTRAN radiative transfer code, in: 2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), IEEE. pp. 1–4. <https://doi.org/10.1109/WHISPERS.2014.8077573>.
- Bourke, P., 2016. Lens correction and distortion. Professional web site: <http://paulbourke.net/dome/fisheyecorrect>.
- Bruneton, E., 2016. A qualitative and quantitative evaluation of 8 clear sky models. *IEEE transactions on visualization and computer graphics* 23, 2641–2655. <https://doi.org/10.1109/TVCG.2016.2622272>.
- Burley, B., Adler, D., Chiang, M.J.Y., Driskill, H., Habel, R., Kelly, P., Kutz, P., Li, Y.K., Teece, D., 2018. The design and evolution of Disney's Hyperion renderer. *ACM Transactions on Graphics (TOG)* 37, 33. <https://doi.org/10.1145/3182159>.
- Cazorla, A., Olmo, F., Alados-Arboledas, L., 2008a. Using a sky imager for aerosol characterization. *Atmospheric Environment* 42, 2739–2745.
- Cazorla, A., Olmo, F.J., Alados-Arboledas, L., 2008b. Development of a sky imager for cloud cover assessment. *JOSA A* 25, 29–39.
- Chandrasekhar, S., 2013. Radiative transfer. Courier Corporation.
- Chauvin, R., Nou, J., Thil, S., Grieu, S., 2015. Modelling the clear-sky intensity distribution using a sky imager. *Solar Energy* 119, 1–17. <https://doi.org/10.1016/j.solener.2015.06.026>.
- Cordero, R.R., Damiani, A., Ferrer, J., Rayas, J., Jorquera, J., Tobar, M., Labbe, F., Laroze, D., 2013. Downwelling and upwelling radiance distributions sampled under cloudless conditions in Antarctica. *Applied Optics* 52, 6287. URL: <https://www.osapublishing.org/abstract.cfm?URI=ao-52-25-6287>, doi:10.1364/AO.52.006287.
- Cornette, W.M., Shanks, J.G., 1992. Physically reasonable analytic expression for the single-scattering phase function. *Applied optics* 31, 3152–3160. <https://doi.org/10.1364/AO.31.003152>.
- Crawley, D.B., Lawrie, L.K., Winkelmann, F.C., Buhl, W.F., Huang, Y.J., Pedersen, C.O., Strand, R.K., Liesen, R.J., Fisher, D.E., Witte, M.J., et al., 2001. Energyplus: creating a new-generation building energy simulation program. *Energy and buildings* 33, 319–331. [https://doi.org/10.1016/S0378-7788\(00\)00114-6](https://doi.org/10.1016/S0378-7788(00)00114-6).
- Del Rocco, J., Patterson, C.B., Dhrif, H., Kider, J.T., 2018. Learning and estimating whole sky visible, vnir, swir radiance distributions from a commercial camera, in: Optics and Photonics for Information Processing XII, SPIE. p. 107510F. <https://doi.org/10.1117/12.2321295>.
- Emde, C., Buras-Schnell, R., Kylling, A., Mayer, B., Gasteiger, J., Hamann, U., Kylling, J., Richter, B., Pause, C., Dowling, T., et al., 2016. The libRadtran software package for radiative transfer calculations (version 2.0.1). *Geoscientific Model Development*, 1647–1672.
- Fisher, R., Perkins, S., Walker, A., Wolfart, E., 1996. Hypermedia image processing reference. John Wiley & Sons Ltd. <https://www.worldcat.org/title/hypermedia-image-processing-reference/oclc/43115444>.
- Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. *Machine learning* 63, 3–42.
- Greenberg, D., Pratt, K., Hencey, B., Jones, N., Schumann, L., Dobbs, J., Dong, Z., Bosworth, D., Walter, B., 2013. Sustain: An experimental test bed for building energy simulation. *Energy and Buildings* 58, 44–57. <https://doi.org/10.1016/j.enbuild.2012.11.026>.
- Gueymard, C., et al., 1995. SMARTS2 A: simple model of the atmospheric radiative transfer of sunshine: algorithms and performance assessment. Florida Solar Energy Center (FSEC) Cocoa, FL. <https://www.nrel.gov/grid/solar-resource/smarts.html>.
- Haber, J., Magnor, M., Seidel, H.P., 2005. Physically-based simulation of twilight phenomena. *ACM Transactions on Graphics (TOG)* 24, 1353–1373. <https://doi.org/10.1145/1095878.1095884>.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The weka data mining software: an update. *ACM SIGKDD explorations newsletter* 11, 10–18. <https://doi.org/10.1145/1656274.1656278>.
- Hensen, J.L., Lamberts, R., 2012. Building performance simulation for design and operation. Routledge.
- Heney, L.G., Greenstein, J.L., 1941. Diffuse radiation in the galaxy. *The Astrophysical Journal* 93, 70–83.
- Hess, M., Koepke, P., Schult, I., 1998. Optical properties of aerosols and clouds: The software package opac. *Bulletin of the American meteorological society* 79, 831–844.
- Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12, 55–67.
- Holben, B.N., Eck, T.F., Slutsker, I., Tanre, D., Buis, J., Setzer, A., Vermote, E., Reagan, J.A., Kaufman, Y., Nakajima, T., et al., 1998. Aeronet—A federated instrument network and data archive for aerosol characterization. *Remote sensing of environment* 66, 1–16.
- Hosek, L., Wilkie, A., 2012. An analytic model for full spectral sky-dome radiance. *ACM Transactions on Graphics (TOG)* 31, 95. <https://doi.org/10.1145/2185520.2185591>.
- Igawa, N., Nakamura, H., 2001. All sky model as a standard sky for the simulation of daylit environment. *Building and Environment* 36, 763–770. [https://doi.org/10.1016/S0360-1323\(00\)00062-7](https://doi.org/10.1016/S0360-1323(00)00062-7).
- Iqbal, M., 2012. An introduction to solar radiation. Elsevier.
- Jakica, N., 2017. State-of-the-art review of solar design tools and methods for assessing daylighting and solar potential for building-integrated photovoltaics. *Renewable and Sustainable Energy Reviews*.
- Jakob, W., 2010. Mitsuba renderer. [Http://www.mitsuba-renderer.org](http://www.mitsuba-renderer.org).
- Jarosz, W., 2008. Efficient Monte Carlo methods for light transport in scattering media. Citeseer.
- Joblove, G.H., Greenberg, D., 1978. Color spaces for computer graphics, in: ACM SIGGRAPH Computer Graphics, ACM. pp. 20–25. <https://doi.org/10.1145/965139.807362>.
- Kajiya, J.T., 1986. The rendering equation, in: ACM SIGGRAPH computer graphics, ACM. pp. 143–150. <https://doi.org/10.1145/15886.15902>.
- Kider, J.T., Knowlton, D., Newlin, J., Li, Y.K., Greenberg, D.P., 2014. A framework for the experimental comparison of solar and skydome illumination. *ACM Transactions on Graphics (TOG)* 33, 180. <https://doi.org/10.1145/2661229.2661259>.
- Kinney, J.A.S., 1958. Comparison of scotopic, mesopic, and photopic spectral sensitivity curves. *JOSA* 48, 185–190. <https://doi.org/10.1364/JOSA.48.000185>.
- Kittler, R., 1994. Some qualities of scattering functions defining sky radiance distributions. *Solar Energy* 53, 511–516. [https://doi.org/10.1016/0038-092X\(94\)90131-K](https://doi.org/10.1016/0038-092X(94)90131-K).
- Kocev, D., Vens, C., Struyf, J., Džeroski, S., 2013. Tree ensembles for predicting structured outputs. *Pattern Recognition* 46, 817–833.
- Kocifaj, M., 2009. Sky luminance/radiance model with multiple scattering effect. *Solar Energy* 83, 1914–1922. <https://doi.org/10.1016/j.solener.2009.07.004>.
- Kocifaj, M., 2012. Angular distribution of scattered radiation under broken cloud arrays: an approximation of successive orders of scattering. *Solar Energy* 86, 3575–3586. <https://doi.org/10.1016/j.solener.2012.06.022>.
- Kocifaj, M., 2015. Unified model of radiance patterns under arbitrary sky conditions. *Solar Energy* 115, 40–51. <http://linkinghub.elsevier.com/retrieve/pii/S0038092X15000894>.
- Koenderink, J.J., 2010. Color for the Sciences. The MIT Press.
- Kohavi, R., et al., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Ijcai, Montreal, Canada. pp. 1137–1145.
- Lacis, A.A., Hansen, J., 1974. A parameterization for the absorption of solar radiation in the earth's atmosphere. *Journal of the atmospheric sciences* 31, 118–133. [https://doi.org/10.1175/1520-0469\(1974\)031%3C0118:APFTA0%3E2.0.CO;2](https://doi.org/10.1175/1520-0469(1974)031%3C0118:APFTA0%3E2.0.CO;2).
- Lee Jr, R.L., 2008. Measuring overcast colors with all-sky imaging. *Applied optics* 47, H106–H115.
- LibRaw, 2018. libraw, raw image decoder. <https://www.libraw.org>.
- Littlefair, P.J., 1981. The luminance distribution of an average sky. *Lighting Research & Technology* 13, 192–198. <https://doi.org/10.1177/096032718101300402>.
- López-Álvarez, M.A., Hernández-Andrés, J., Romero, J., Olmo, F.J., Cazorla, A., Alados-Arboledas, L., 2008. Using a trichromatic CCD camera for spectral skylight estimation. *Applied optics* 47, H31–H38. URL: <https://www.osapublishing.org/abstract.cfm?uri=ao-47-34-h31>.

- Macskassy, S.A., Hirsh, H., 2003. Adding numbers to text classification, in: Proceedings of the twelfth international conference on Information and knowledge management, ACM. pp. 240–246. <https://doi.org/10.1145/956863.956910>.
- Malthus, T., MacLellan, C., 2010. High performance fore optic accessories and tools for reflectance and radiometric measurements with the ASD FR3 spectroradiometer, in: ESA Hyperspectral Workshop 2010 “From Chris/Proba to PRISMA and EnMAP and beyond”, European Space Agency. p. EP101432. <http://hdl.handle.net/102.100.100/108223>.
- Mayer, B., Kylling, A., 2005. The libRadtran software package for radiative transfer calculations. *Atmospheric Chemistry and Physics* 5, 1855–1877.
- Mie, G., 1908. Beiträge zur optik trüber medien, speziell kolloidaler mettallösungen. *Annalen der physik* 330, 377–445.
- Nakamura, H., Oki, M., Hayashi, Y., 1985. A study on the estimation of the relative frequency of occurrences of the clear sky, the intermediate sky and the overcast sky in japan. *Journal of Light & Visual Environment* 9, 2_22–2_31. https://doi.org/10.2150/jlve.9.2_22.
- Nishita, T., Dobashi, Y., Nakamae, E., 1996. Display of clouds taking into account multiple anisotropic scattering and sky light, in: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, ACM. pp. 379–386. <https://doi.org/10.1145/237170.237277>.
- Nishita, T., Sirai, T., Nakamae, E., Tadamura, K., 1993. Display of the earth taking into account atmospheric scattering, in: SIGGRAPH, p. 175. <https://doi.org/10.1145/166117.166140>.
- Nou, J., Chauvin, R., Eynard, J., Thil, S., Grieu, S., 2018. Towards the intrahour forecasting of direct normal irradiance using sky-imaging data. *Heliyon* 4, e00598. <https://doi.org/10.1016/j.heliyon.2018.e00598>.
- Parker, J.R., 2010. Algorithms for image processing and computer vision. John Wiley & Sons. <https://www.worldcat.org/title/algorithms-for-image-processing-and-computer-vision/oclc/813666888>.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., 2011. Scikit-learn: Machine learning in python. *Journal of machine learning research* 12, 2825–2830.
- Perez, R., Seals, R., Michalsky, J., 1993. All-weather model for sky luminance distribution—preliminary configuration and validation. *Solar energy* 50, 235–245. [https://doi.org/10.1016/0038-092X\(93\)90017-I](https://doi.org/10.1016/0038-092X(93)90017-I).
- Picard, R.R., Cook, R.D., 1984. Cross-validation of regression models. *Journal of the American Statistical Association* 79, 575–583. <https://doi.org/10.1080/01621459.1984.10478083>.
- Pokrowski, G., 1929. Über einen scheinbaren mie-effekt und seine mögliche rolle in der atmosphärenoptik. *Zeitschrift für Physik* 53, 67–71.
- Poynton, C.A., 1995. A guided tour of colour space, in: New Foundation for Video Technology: The SMPTE Advanced Television and Electronic Imaging Conference, SMPTE. pp. 167–180. <https://doi.org/10.5594/M00840>.
- Preetham, A., Shirley, P., Smits, B., 1999. A practical analytic model for daylight, in: Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, ACM. ACM Press/Addison-Wesley Publishing Co.. pp. 91–100. <https://doi.org/10.1145/311535.311545>.
- Reda, I., Andreas, A., 2004. Solar position algorithm for solar radiation applications. *Solar energy* 76, 577–589. <https://doi.org/10.1016/j.solener.2003.12.003>.
- Ricchiazzi, P., Yang, S., Gautier, C., Sowle, D., 1998. Sbdart: A research and teaching software tool for plane-parallel radiative transfer in the earth's atmosphere. *Bulletin of the American Meteorological Society* 79, 2101–2114. [https://doi.org/10.1175/1520-0477\(1998\)079<2101:SARATS>2.0.CO;2](https://doi.org/10.1175/1520-0477(1998)079<2101:SARATS>2.0.CO;2).
- Riechert, M., 2018. rawpy (0.13.1), RAW image processing for Python. <https://pypi.org/project/rawpy>.
- Robertson, A.R., et al., 1977. The CIE 1976 color-difference formulae. *Color Research & Application* 2, 7–11. <https://doi.org/10.1002/j.1520-6378.1977.tb00104.x>.
- Roudsari, M.S., Pak, M., Smith, A., et al., 2013. Ladybug: A parametric environmental plugin for Grasshopper to help designers create an environmentally-conscious design, in: Proceedings of the 13th international IBPSA conference held in Lyon, France Aug, pp. 3128–3135. https://www.ibpsa.org/proceedings/bs2013/p_2499.pdf.
- Saito, M., Iwabuchi, H., 2016. Cloud Discrimination from Sky Images Using a Clear-Sky Index. *Journal of Atmospheric and Oceanic Technology* 33, 1583–1595. URL: <http://journals.ametsoc.org/doi/10.1175/JTECH-D-15-0204.1>.
- Saito, M., Iwabuchi, H., Murata, I., 2016. Estimation of spectral distribution of sky radiance using a commercial digital camera. *Applied Optics* 55, 415. <https://doi.org/10.1364/AO.55.000415>.
- Satylmýs, P., Bashford-Rogers, T., Chalmers, A., Debattista, K., 2016. A machine-learning-driven sky model. *IEEE computer graphics and applications* 37, 80–91. <https://doi.org/10.1109/MCG.2016.67>.
- Sigernes, F., Holmes, J.M., Dyrland, M., Lorentzen, D.A., Svenøe, T., Heia, K., Aso, T., Chernouss, S., Deehr, C.S., 2008. Sensitivity calibration of digital colour cameras for auroral imaging. *Optics express* 16, 15623–15632.
- Smith, A.R., 1978. Color gamut transform pairs. *ACM SIGGRAPH Computer Graphics* 12, 12–19. <https://doi.org/10.1145/965139.807361>.
- Smith, C.J., Forster, P.M., Crook, R., 2016. An all-sky radiative transfer method to predict optimal tilt and azimuth angle of a solar collector. *Solar Energy* 123, 88–101.
- Stokes, M., Anderson, M., Chandrasekar, S., Motta, R., 1996. A standard default color space for the internet-sRGB. Microsoft and Hewlett-Packard Joint Report <https://www.w3.org/Graphics/Color/sRGB>.
- Stone, E., 2015. Completely painless programmer's guide to XYZ, RGB, ICC, xyY, and TRCs. Personal website <https://ninedegreesbelow.com/photography/xyz-rgb.html>.
- Strutt, J.W., 1871. On the light from the sky, its polarization and colour. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 41, 107–120.
- Stumpfel, J., Tchou, C., Jones, A., Hawkins, T., Wenger, A., Debevec, P., 2004. Direct hdr capture of the sun and sky, in: Proceedings of the 3rd International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa, ACM. pp. 145–149. <https://doi.acm.org/10.1145/1029949.1029977>.
- Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.
- Tohsing, K., Schrempf, M., Riechelmann, S., Seckmeyer, G., 2014. Validation of spectral sky radiance derived from all-sky camera images—a case study. *Atmospheric Measurement Techniques* 7 (2014), Nr. 7. <https://doi.org/10.15488/5>.
- Vermote, E., Tanré, D., Deuzé, J., Herman, M., Morcrette, J., Kotchenova, S., 2006. Second simulation of a satellite signal in the solar spectrum-vector (6SV). 6S User Guide Version 3, 1–55.
- Ward, G.J., 1994. The RADIANCE lighting simulation and rendering system, in: Proceedings of the 21st annual conference on Computer graphics and interactive techniques, ACM. pp. 459–472. <https://doi.org/10.1145/192161.192286>.
- Wright, W.D., 1929. A re-determination of the trichromatic coefficients of the spectral colours. *Transactions of the Optical Society* 30, 141. <https://doi.org/10.1088/1475-4878/30/4/301>.
- Yao, W., Li, Z., Zhao, Q., Lu, Y., Lu, R., 2015. A new anisotropic diffuse radiation model. *Energy Conversion and Management* 95, 304–313. <https://doi.org/10.1016/j.enconman.2015.01.016>.
- Yu, C.H., 1977. Exploratory data analysis. *Methods* 2, 131–160. <https://dx.doi.org/10.1093/obo/9780199828340-0200>.
- Zotti, G., Wilkie, A., Purgathofer, W., 2007. A critical review of the preetham skylight model, in: WSCG '07: Short Communications Proceedings, Václav Skala-UNION Agency. pp. 23–30. <http://hdl.handle.net/11025/11160>.
- Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67, 301–320.

Appendix A - 05/26/13 15:15¹

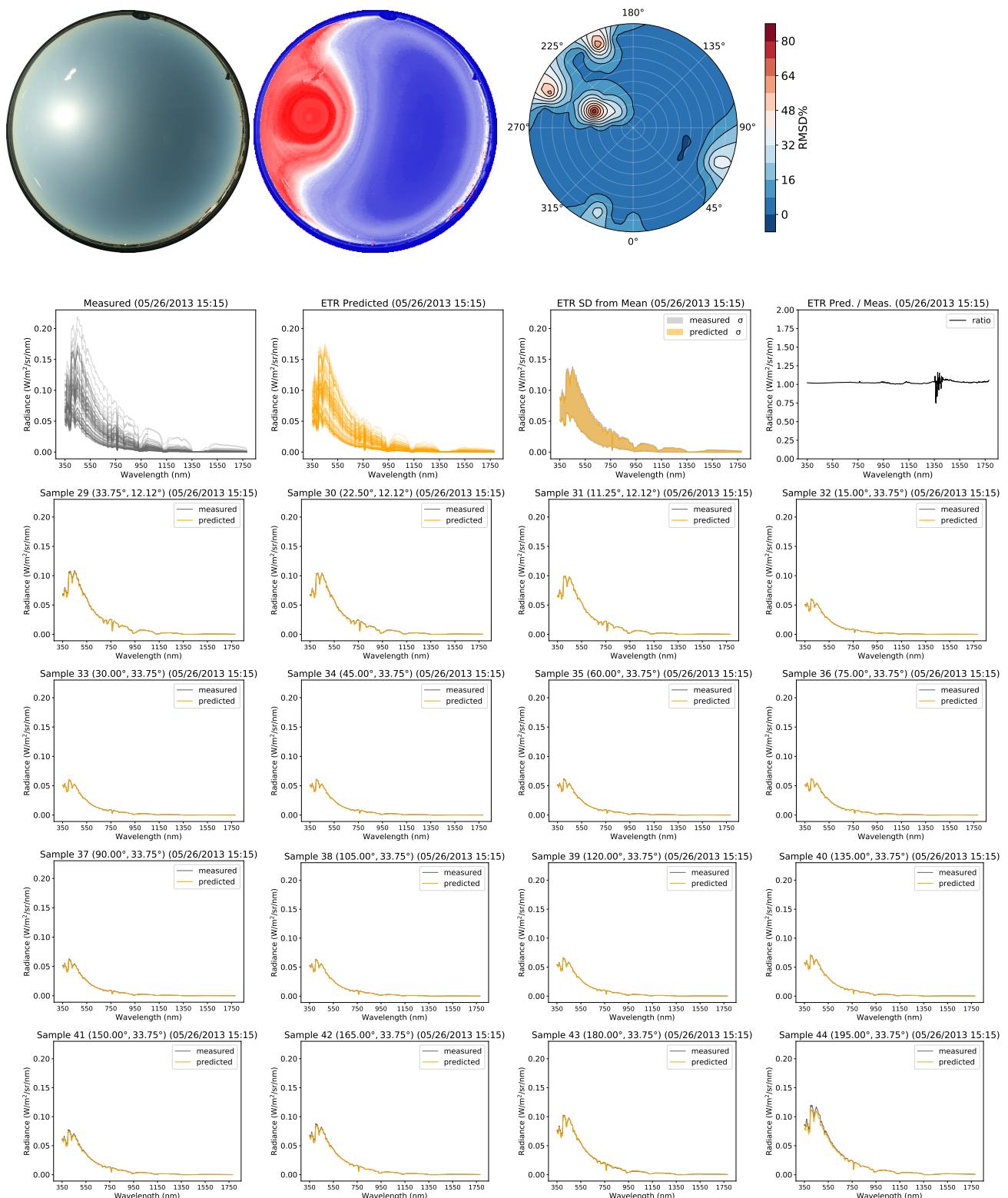
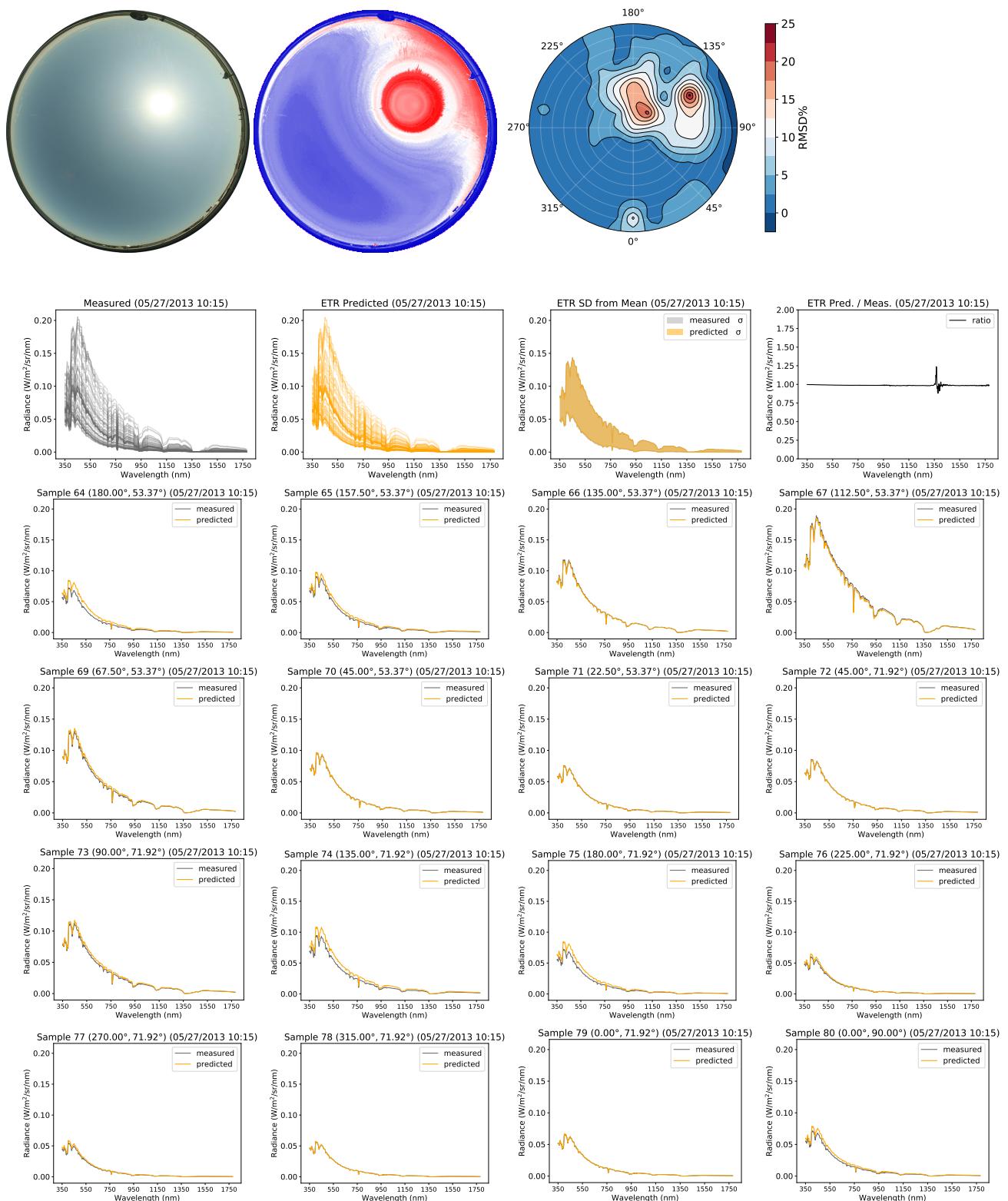


Figure: Whole sky results and 16 of 81 spectral radiance predictions for holdout sky 05/26/2013 15:15 with ETR model.

¹Provided for reviewing purposes only.

Appendix B - 05/27/13 10:15²**Figure:** Whole sky results and 16 of 81 spectral radiance predictions for holdout sky 05/27/2013 10:15 with ETR model.²Provided for reviewing purposes only.

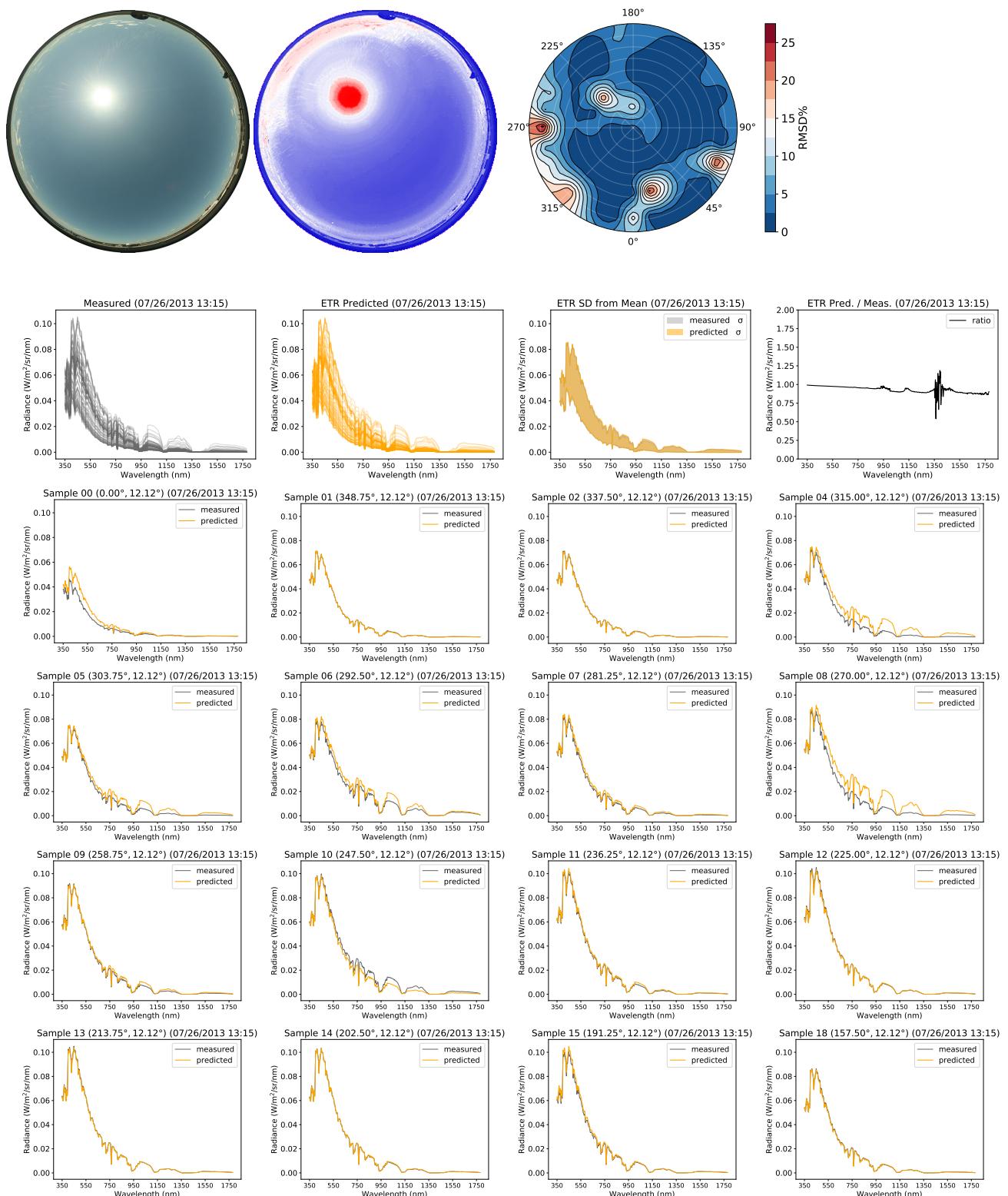
Appendix C - 07/26/13 13:15³


Figure: Whole sky results and 16 of 81 spectral radiance predictions for holdout sky 07/26/2013 13:15 with ETR model.

³Provided for reviewing purposes only.

Appendix D - 09/24/13 15:39 ⁴

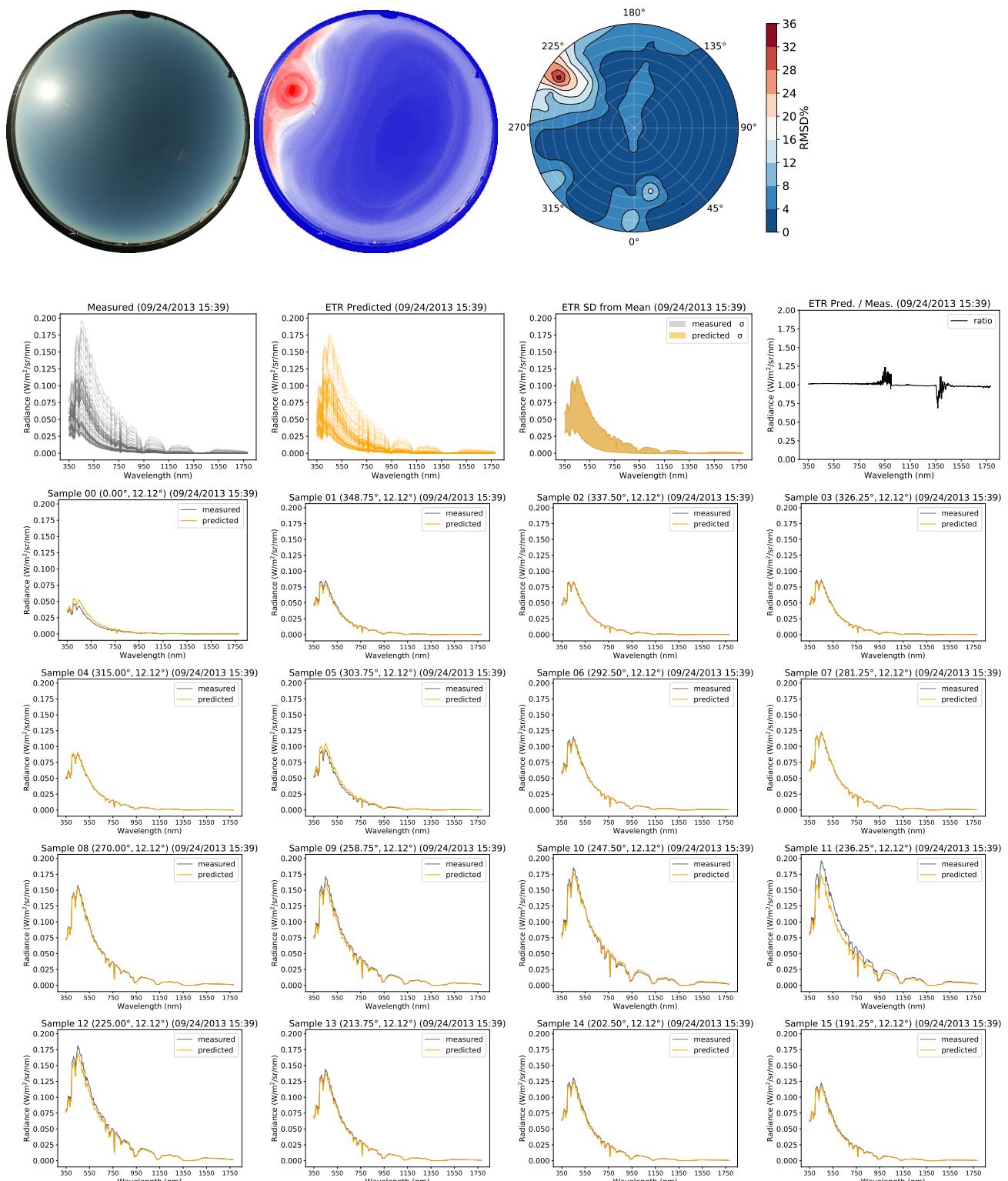


Figure: Whole sky results and 16 of 81 spectral radiance predictions for holdout sky 09/24/2013 15:39 with ETR model.

⁴Provided for reviewing purposes only.