INDEPENDENT VALIDATION OF NDFD-BASED SOLAR RADIATION FORECASTS

Richard Perez
ASRC, The University at Albany
251 Fuller Rd, Albany, NY
12203
Perez@asrc.cestm.albany.edu

Kathleen E. Moore
Integrated Environmental Data
255 Fuller Rd., Albany, NY
12203
moore@iedat.com

Paul W. Stackhouse Jr.
NASA Langley Research Center
21 Langley Blvd.
Hampton, VA 23681
Paul.w.stackhouse@nasa.gov

ABSTRACT

The objective of this study is to expand upon our initial evaluation of irradiance products derived from National Digital Forecast Data Base (NDFD, [1]) against ground truth data in Albany, NY, in two ways:

1. By analyzing multiple, climatically distinct, ground truth locations, and
2. By evaluating the performance of the forecasts in terms of extended area coverage.

1. METHODOLOGY

1.1 Recall on the solar radiation forecast model

The NDFD is assembled from the forecasts of local and regional National Weather Service (NWS) offices in collaboration with the National Center for Environmental Prediction (NCEP). Local forecasts are generated as a byproduct of national model outputs, mesoscale model runs and human input [2]. These local forecasts are then merged and assembled on a national grid. At present, forecast products include ambient temperature, dew point temperature, probability of precipitation, weather type, sky cover, wind speed and direction, significant wave height, quantitative precipitation forecasts, and snow amount. Most parameters are considered operational, but sky cover is still considered an experimental parameter by NWS.

Global irradiance, GHI, is modeled from the NDFD sky cover forecast, SK. This sky cover parameter is used as a modulator of clear sky global irradiance GHI_{clear} per equation (1) fitted to observations in Albany, NY [3]:

$$GHI = GHIclear \quad (1 - 0.87\ SK1.9) \qquad (1)$$

1.2 Individual ground-truth locations

Three stations from NOAA's SURFRAD [4] network were selected based upon their climatic/geographic distinction with Albany, NY. The stations include:

- Desert Rock, NV
- Goodwin Creek, MS
- Boulder, CO

The time period analyzed for the ground truth data and the time-coincident NDFD forecasts ranges from June 23, 2005 to September 29, 2005.

1.3 Extended area

A 3° longitude by 2° latitude area near Los Angeles, CA was selected based upon the existence of its large microclimatic variability – from the cloudy Pacific to the Mojave Desert (Figure 1).

Figure 1: Extended validation area

The objective of the extended area analysis is to ascertain whether forecasted cloud cover accounts for observed solar microclimate signatures. The "ground-truth" data in this case consist of satellite-derived irradiances for the same area.

The metrics used for extended area validation include: (1) RMSE and MBE at the center and at each corner of the selected area, and (2) the mean forecasted GHI map vs. the mean satellite-derived GHI map.

Forecast time frames: The forecasts considered for this analysis are generated at 13:00 GMT and span every 3 hour from 2 to 59 hours ahead, and every six hour from 65 to 155 hours ahead.

Hence, day time forecasts considered for the Pacific-time sites (extended area and Desert Rock) are at 10:00, 13:00 16:00 local time for the first 3 days and 10:00 and 16:00 local time for days 4 to 7. For Boulder, the times are respectively 11:00, 14:00 and 17:00 for the first three days and 11:00 and 17:00 beyond. For these locations the first considered forecast is 5 hours ahead.

For Goodwin Creek the considered local times are 9:00, 12:00 and 15:00 for days 1-3 and 12:00 and 18:00 for days 4-7. The first considered forecast is 2 hours ahead.

Forecasts are grouped in 5 categories for analysis purposes:

- Category 1: $\leq$ 5 hours ahead
- Category 2: > 5 hours ahead, same day
- Category 3: All hours second day
- Category 4: All hours, third day
- Category 5: All hours, 4th – 7th day.

2. RESULTS

2.1 Individual ground-truth locations

Observed trends: The relationship observed between NDFD-sky cover forecasts and measured GHI index (ratio of GHI to GHI_{clear}) is illustrated in Figure 2. The points

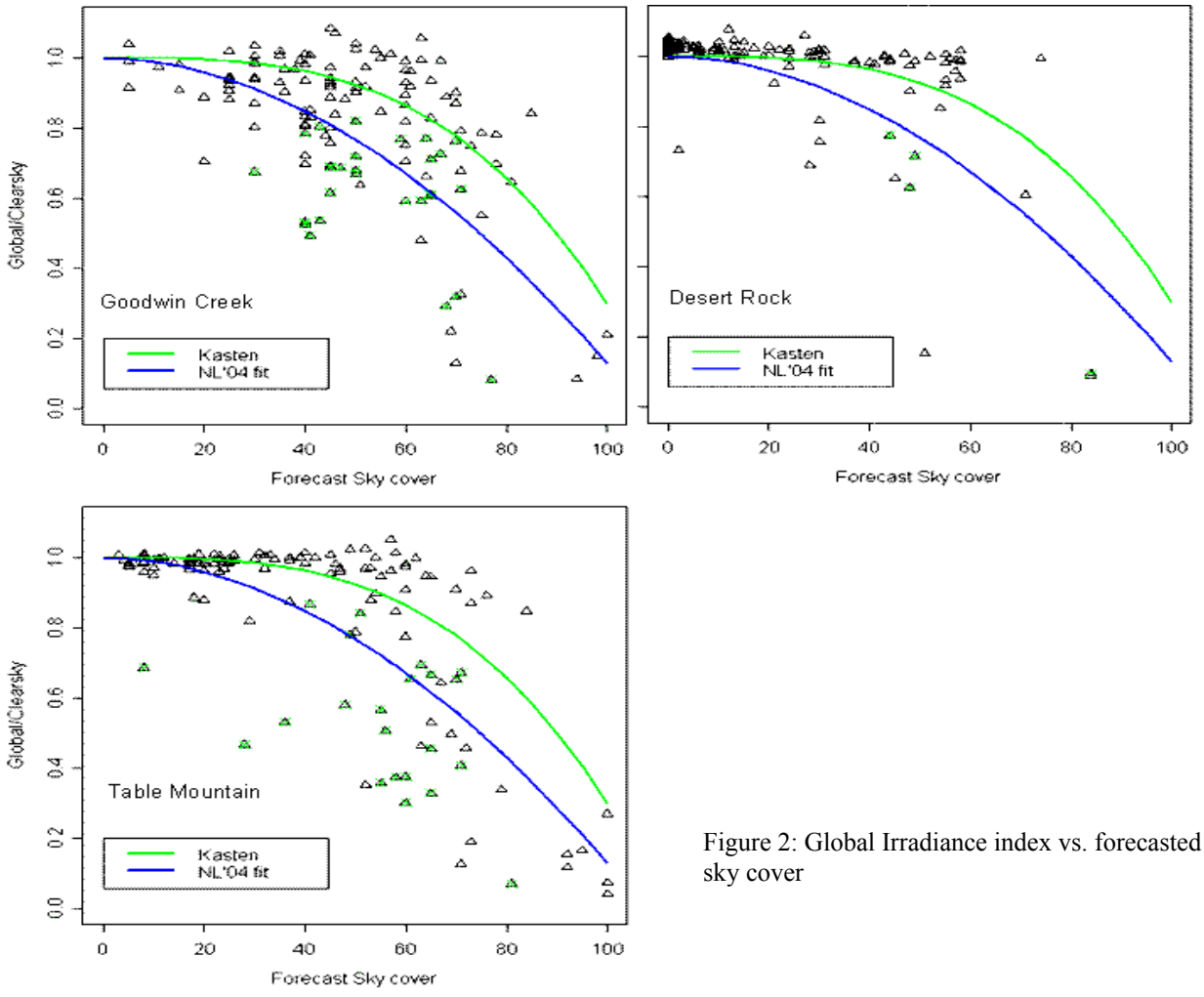

Figure 2: Global Irradiance index vs. forecasted sky cover

selected to illustrate this relationship include category 2 forecasts. The figure also includes plots for the original Kasten and Czeplak formula relating observed cloud cover and GHI index [5], and equation (1) derived from Albany data (labeled NL'04 fit).

The observed relationships are comparable for all sites and consistent with our initial analysis in Albany.

The points marked with crossed-triangles represent convective (partly cloudy) conditions as identified from the time series of ground measurements with hour-to-hour clearness index difference in excess of 0.3. Qualitatively, one observes a distinct relationship between steady and variable conditions. This observation could lead to a better parameterization of predicted sky cover if the stable-unstable distinction can be inferred from other NDFD forecasted parameters.

Table 1 summarizes the MBE and RMSE observed for each location and forecast category over the considered period. Three forecasts models are evaluated:

- A model based upon equation (1) previously fitted from Albany data [3]
- A model based upon The Kasten-Czeplak equation [4]
- A dependent model based on the best fit equation for each site -- equation (1) with fitted coefficients.

Model performance is benchmarked against a simple persistence model (constant clearness index).

In terms of bias, results confirm the tendency observed in Albany for the Kasten Czeplak formulation to overestimate predicted irradiance – this confirms the remark that forecasted cloud cover (modeled from the predicted status of the atmosphere) is different in nature from the cloud cover detected by human observers. The Albany-fitted function performs remarkably well against independent sites. Interestingly, persistence leads to a large positive bias at two of the sites for Category 2 forecasts (same day AM-PM forecasts); this may be an indication of typical diurnal patterns with afternoon cloud build-up at these locations during time period of this analysis.

In terms of dispersion (RMSE), all sky cover-to-irradiance forecast models perform similarly and do so noticeably better than the persistence benchmark.

Model performance may be qualitatively visualized in Figure 3, comparing next-day (category 3) forecasts to ground measurements. As above, note that the points marked with crossed triangles represent partly cloudy conditions as identified from the time series of ground measurements with a clearness index difference in excess of 0.3. It is apparent from the resulting pattern that forecasts could be improved if partly cloudy conditions could be identified by analysis of other forecasted parameters.

2.2 Extended area

Point-specific comparisons for each corner and the center of the extended area are summarized in Table 2. Comparing forecasts against satellite-derived "ground truth" data leads to results which are consistent with actual ground measurement comparisons. As would be expected, the magnitude of the RMSE increases with the amount of cloudiness at the considered location during the considered time period – e.g., contrast the very clear northwest point to the more cloudy Pacific Ocean southwest point. MBES for the Albany fit are in an acceptable range for all points, while the Kasten & Czeplak (K&C) relation overestimates as

TABLE 1

		Mean Bias Error					Root Mean Square Error				
		Cat. 1	Cat. 2	Cat. 3	Cat. 4	Cat. 5	Cat. 1	Cat. 2	Cat. 3	Cat. 4	Cat. 5
TABLE MOUNTAIN	Mean	412	713	408	525	530	412	713	408	525	530
	Albany-fit	20	-34	26	-3	-5	87	166	176	176	191
	Kasten-Czeplak	60	76	115	92	89	85	166	181	175	181
	Best-fit	9	8	-3	-10	-30	85	169	182	176	181
	Persistence	32	70	33	44	60	108	442	182	240	236
GOODWIN CREEK	Mean	528	683	489	596	747	528	683	489	596	747
	Albany-fit	-33	-54	16	-31	-36	122	156	174	164	210
	Kasten-Czeplak	56	70	115	78	98	115	149	173	155	203
	Best-fit	61	76	120	83	101	117	151	174	156	202
	Persistence	52	129	36	42	53	181	307	200	241	318
DESERT ROCK	Mean	305	816	556	606	599	305	816	556	606	599
	Albany-fit	30	-47	-8	-7	-1	45	110	124	116	118
	Kasten-Czeplak	43	2	29	28	32	50	105	118	113	112
	Best-fit	43	3	30	28	32	50	107	118	113	112
	Persistence	-11	5	-3	-3	-3	40	169	137	148	153

observed elsewhere.

Interestingly, the K&C relation leads to slightly better RMSE. The shape of the Albany fit is driven largely by scatter points below the K&C trend (e.g., see fig. 2) -- this minimizes the overall bias, but does little to improve overall RMSE. As mentioned above, this observation suggests that, if partly cloudy occurrences can be identified through ancillary forecasted parameters, a model using two trends, one closer to the K&C for homogeneous sky conditions, and one much lower for partly cloudy occurrences may prove effective.

The ability of forecasted cloud cover to account for regional microclimates may be qualitatively gauged in Fig. 4. The figure includes two maps assembled from the data points corresponding to next day forecasts (category 3). Region-wide features are similar, with a good capture of the Pacific coastal cloudiness. But the forecasts tend to gloss over micro-features and to privilege clear conditions near the coast for the considered time period.

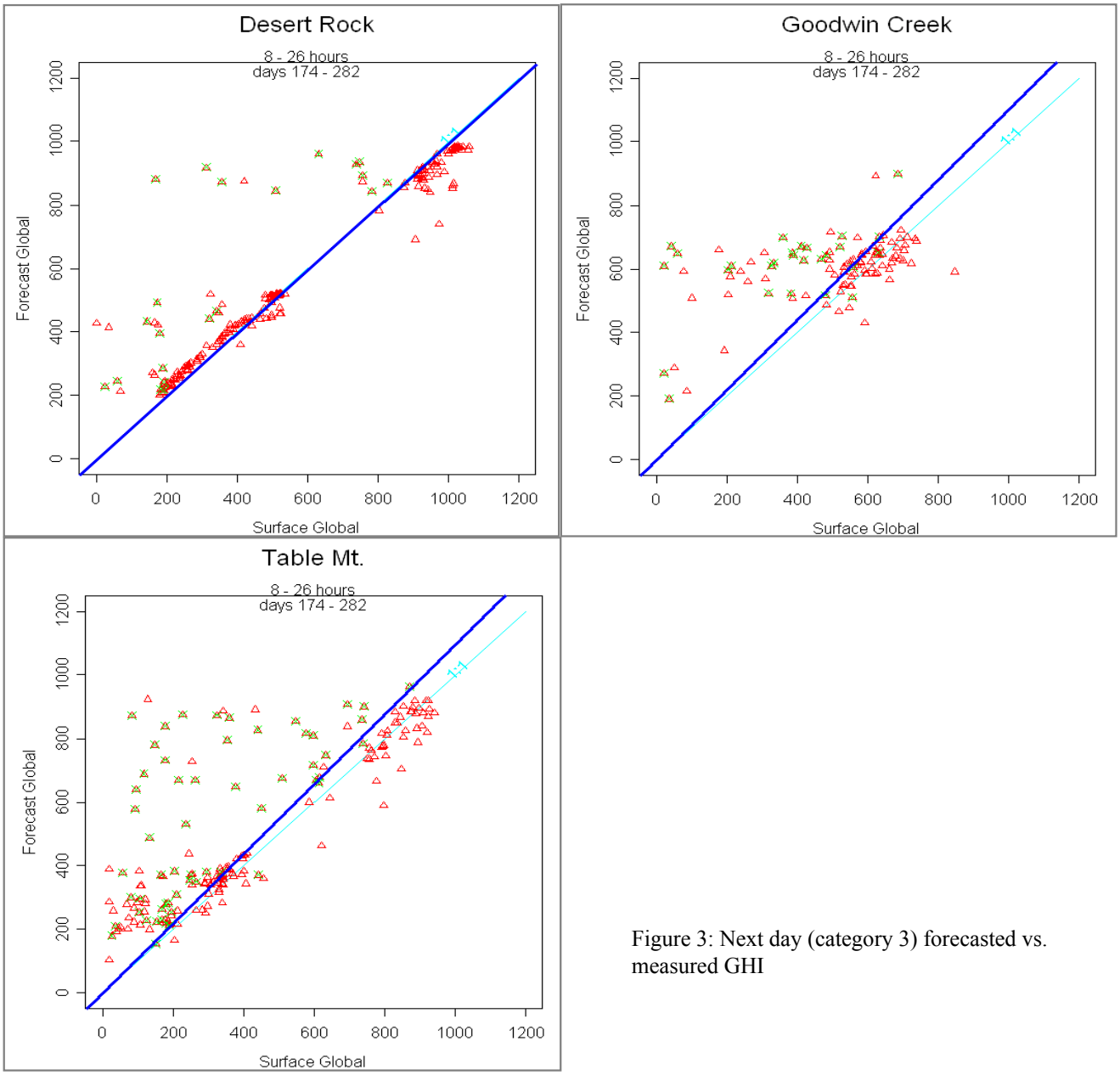

Figure 3: Next day (category 3) forecasted vs. measured GHI

TABLE 2

		Mean Bias Error					Root Mean Square Error				
		Cat. 1	Cat. 2	Cat. 3	Cat. 4	Cat. 5	Cat. 1	Cat. 2	Cat. 3	Cat. 4	Cat. 5
Southwest		647	383	488	487	417	647	383	488	487	417
	Albany-fit	42	14	12	12	36	225	151	201	211	214
	Kasten-Czeplak	135	79	98	98	107	213	145	189	197	200
Southeast		675	417	522	522	441	675	417	522	522	441
	Albany-fit	14	-22	-23	-25	10	200	133	176	189	196
	Kasten-Czeplak	107	44	64	63	83	189	117	162	172	180
Center		665	399	506	505	434	665	399	506	505	434
	Albany-fit	39	7	7	7	28	196	129	170	176	193
	Kasten-Czeplak	128	70	91	90	97	187	119	160	166	182
Northwest		746	472	584	582	505	746	472	584	582	505
	Albany-fit	-18	-47	-48	-47	-24	171	94	136	139	141
	Kasten-Czeplak	62	8	27	27	36	147	59	111	113	121
Northeast		689	434	540	537	457	689	434	540	537	457
	Albany-fit	43	-6	1	2	26	205	111	163	166	186
	Kasten-Czeplak	122	49	74	76	87	196	102	152	156	176

3. CONCLUDING REMARKS

The most important findings of this report are:

A model derived for Albany, NY works with a similar degree of accuracy for other climatically distinct locations, exhibiting little bias when compared to ground measurements and satellite estimates.

Forecasting errors quantified in terms of RMSE range from a little over 10% for category 1 forecasts to 25-40% for next day (category 3) forecasts, depending upon climate, with the best results observed for clear locations.

Cloud cover forecasts are found to account adequately for strong microclimatic features, with a possible tendency to overestimate sky cover in the clearest year-around locations, and underestimate it in locations which are clear only part of the year (e.g., contrast the Mojave Desert and the costal range in Fig. 4).

The shape of the sky cover-irradiance relationship is affected by the insolation conditions. Qualitatively, one observes a distinct relationship between steady and variable insolation conditions as defined by the hour-to-hour variation in ground-truth clearness index -- variable conditions are likely representative of convective, partly cloudy situations. Thus a better parameterization of predicted sky cover involving two trends – one for stable and one for convective conditions – could be envisaged if the stable-unstable distinction can be inferred from other forecasted parameters. This avenue will be explored in future work.

Finally, since NDFD's sky cover is still an experimental product, it is possible that its accuracy will be improved in the future, using, e.g., the feedback of ongoing efforts to improve cloud/irradiance forecasts, such as the International Energy Agency's Solar Heating and Cooling Programme's task 36.

4. ACKNOWLEDGEMENTS

This project was supported by NASA via SAIC contract # 4400113425

5. REFERENCE

1. NDFD, the National Forecast Database, (2005): National Weather Service, NOAA, Washington, DC. http://www.nws.noaa.gov/ndfd/

2. Snyder, W. R., (2005): Personal Communication. National Weather Service, Albany, NY

3. Perez, R. K. Moore, S. Wilcox, D. Renné and A. Zelenka (2005): Forecasting Solar Radiation: Preliminary Evaluation of an Approach Based upon the National Forecast Data Base, Proc. ISES World Congress, Orlando, FL

4. The SURFRAD Network: Monitoring Surface Radiation in the Continental United States, (2005): National Weather Service, NOAA, Washington, DC. http://www.srrb.noaa.gov/surfrad/

5. Kasten F. and G. Czeplak, (1979): Solar and Terrestrial Radiation dependent on the Amount and type of Cloud. Solar Energy 24, 177-189

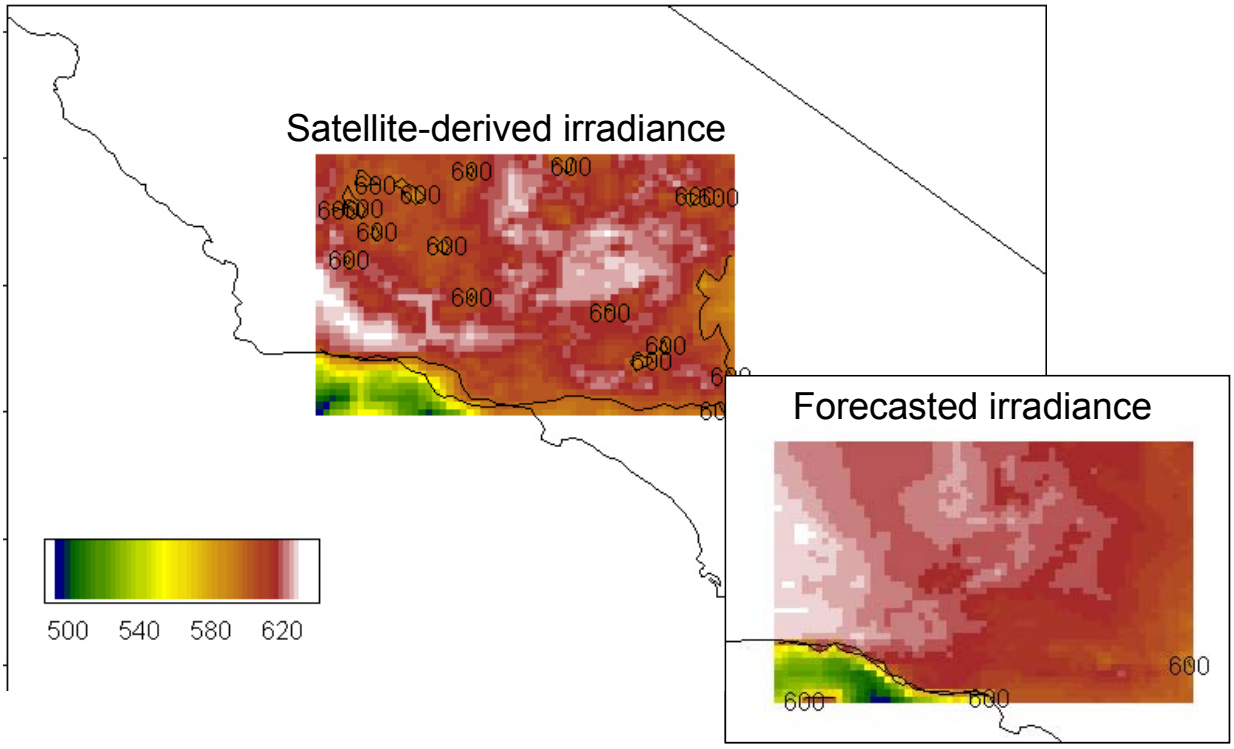

Figure 4: Average irradiances derived from satellite data and Category 3 (next-day) forecasts.