



## Comparison of numerical weather prediction solar irradiance forecasts in the US, Canada and Europe

Richard Perez<sup>a,\*</sup>, Elke Lorenz<sup>b</sup>, Sophie Pelland<sup>c</sup>, Mark Beauharnois<sup>a</sup>, Glenn Van Knowe<sup>d</sup>,  
Karl Hemker Jr.<sup>a</sup>, Detlev Heinemann<sup>b</sup>, Jan Remund<sup>e</sup>, Stefan C. Müller<sup>e</sup>,  
Wolfgang Traunmüller<sup>f</sup>, Gerald Steinmayer<sup>g</sup>, David Pozo<sup>h</sup>, Jose A. Ruiz-Arias<sup>h</sup>,  
Vicente Lara-Fanego<sup>h</sup>, Lourdes Ramirez-Santigosa<sup>i</sup>, Martin Gaston-Romero<sup>j</sup>,  
Luis M. Pomares<sup>k</sup>

<sup>a</sup> Atmospheric Sciences Research Center, The University at Albany, 251 Fuller Rd., Albany, NY 12203, USA

<sup>b</sup> University of Oldenburg, Oldenburg, Germany

<sup>c</sup> Natural Resource Canada, Montreal, Canada

<sup>d</sup> MESO, Inc., Albany, New York, USA

<sup>e</sup> Meteotest, Zurich, Switzerland

<sup>f</sup> Bluesky Wetteranalysen, Atnang-Puchheim, Austria

<sup>g</sup> Austria Solar Innovation Center, Wels, Austria

<sup>h</sup> University of Jaen, Jaen, Spain

<sup>i</sup> CIEMAT, Madrid, Spain

<sup>j</sup> CENER, Sarriguren, Spain

<sup>k</sup> IrSOLaV, Madrid, Spain

Received 5 November 2012; received in revised form 8 May 2013; accepted 9 May 2013

Available online 17 June 2013

Communicated by: Associate Editor Frank Vignola

### Abstract

This article combines and discusses three independent validations of global horizontal irradiance (GHI) multi-day forecast models that were conducted in the US, Canada and Europe. All forecast models are based directly or indirectly on numerical weather prediction (NWP). Two models are common to the three validation efforts – the ECMWF global model and the GFS-driven WRF mesoscale model – and allow general observations: (1) the GFS-based WRF- model forecasts do not perform as well as global forecast-based approaches such as ECMWF and (2) the simple averaging of models' output tends to perform better than individual models.

© 2013 Elsevier Ltd. All rights reserved.

**Keywords:** Irradiance; Forecast; Validation; Solar resource; Numerical weather prediction

### 1. Introduction

Solar power generation is highly variable due its dependence on meteorological conditions. The integration of this fluctuating resource into the energy supply system

requires reliable forecasts of the expected power production as a basis for management and operation strategies. During the last years the contribution of solar power to the electricity supply has been increasing fast leading to a strong need for accurate solar power predictions (in Germany, for instance, the PV production already exceeds 40% of electrical demand on sunny summer days).

\* Corresponding author. Tel.: +1 5184378751.

E-mail address: [rperez@albany.edu](mailto:rperez@albany.edu) (R. Perez).

## Nomenclature

AEMet	Spanish Weather Service	KSI	Kolmogorov–Smirnov test integral
ARPS	Advanced Multiscale Regional Prediction System	MASS	Mesoscale Atmospheric Simulation System [model]
$CDF_{meas}$	cumulative measured frequency distribution	MAE	mean absolute error
$CDF_{pred}$	cumulative predicted frequency distribution	MBE	mean bias error
CENER	Centro Nacional de Energias Renovables (National Renewable Energy Center)	MOS	Model Output Statistics
CIEMAT	Centro de Investigaciones Energeticas, Medioambientales y Tecnologicas (Center for Research on Energy, Environment and Technology)	MSE	mean square error
DSWRF	downward shortwave radiation flux at the surface	NCEP	National Centers for Environmental Prediction
ECMWF	European Center for Medium Range Weather Forecasts [model]	NDFD	National Digital Forecast Database [model]
GEM	Global Environmental Multiscale [model]	NOAA	National Oceanic and Atmospheric Administration
GHI	global irradiance	$N$	number of evaluated prediction-measurement pairs
GFS	Global Forecast System [model]	NWP	numerical weather prediction
HIRLAM	High Resolution Limited Area Model	RMSE	root mean square error
$I_{mas}$	maximum possible irradiance value	SURFRAD	Surface Radiation Network NOAA
$I_{meas}$	measured irradiance	WRF	Weather Research and Forecasting [model]
$I_{pred}$	predicted irradiance	WRF-ASRC	WRF-Model from Atmospheric Sciences Research Center
IEA SHC	International Energy Agency Solar Heating and Cooling Programme	WRF-AWS	WRF Model from AWSTruepower
		WRF-Meteotest	WRF Model from Meteotest
		WRF-UJAEN	WRF Model from University of Jaen

Following this new and rapidly evolving situation on the energy market, substantial research effort is currently being spent on the development of irradiance and solar power prediction models, and several models have been proposed recently by research organizations as well as by private companies. Common operational approaches to short-term solar radiation forecasting include (1) numerical weather prediction (NWP) models that infer local cloud information – hence, indirectly, transmitted radiation – through the dynamic modeling of the atmosphere up to several days ahead (e.g., see Remund et al., 2008); (2) models using satellite remote sensing or ground based sky measurements to infer the motion of clouds and project their impact in the future. Earlier contributions by some of the authors have shown that satellite-derived cloud motion tends to outperform NWP models for forecast horizons up to 4–5 h ahead depending on location (e.g., Perez et al., 2010; Heinemann et al., 2006). Short-term forecasting using ground-based sky imagery with very high spatial and temporal resolution is suitable for intra-hour forecasting (Chow et al., 2011); (3) statistical time series models based on measured irradiance data are applied for very short term forecasting in the range of minutes to hours (e.g., see Pedro and Coimbra, 2012). In this paper we focus our attention on solar radiation forecasts based on NWP models which are most appropriate for day-ahead and multi-day forecast

horizons. Day-ahead predictions are of particular importance for application in the energy market, where day-ahead power trading plays a major role in many countries.

This article combines and discusses three independent validations of global horizontal irradiance (GHI) multi-day forecast models that were performed in the US, Canada and Europe in the framework of the IEA SHC Task 36 “Solar resource knowledge management” (<http://archive.iea-shc.org/task36/>). Comparing the performance of different models gives valuable information both to researchers, to rank their approaches and inform further model development, and to forecast users, to assist them in choosing between different forecasting products. It is important that a standardized methodology for evaluation is used for the comparison in order to achieve meaningful results when comparing different approaches. Therefore, a common benchmarking procedure has been set up in the framework of the IEA SHC Task 36. As a basis for the benchmarking we have prepared several ground measurement data sets covering different climatic regions and a common set of accuracy measures has been identified.

The paper first gives an overview of the different forecasting approaches. Then we present the ground measurement datasets used for the validation. Next, the concept of evaluation is introduced, and finally, we provide the results

of the forecast comparison along with a short discussion and conclusions.

## 2. Forecast models

The evaluation includes forecasts based on global, multiscale and mesoscale NWP models. Hourly site-specific forecasts are derived from direct NWP model output with different methods ranging from simple averaging and interpolation techniques to advanced statistical postprocessing tools and meteorologists' interpretation to combine the output of various NWP models. The models considered for this evaluation are listed below, along with the acronyms that will be used to present results:

1. The Global Environmental Multiscale (GEM) model from Environment Canada in its regional deterministic configuration (Mailhot et al., 2006).
2. An application of the European Centre for Medium-Range Weather Forecasts (ECMWF) model (Lorenz et al., 2009a,b).
3. Several versions of the Weather Research and Forecasting (WRF) model (Skamarock et al., 2005, 2008) initialized with Global Forecast System (GFS) forecasts (GFS, 2010) from the US National Oceanic and Atmospheric Administration's (NOAA) National Centers for Environmental Prediction (NCEP).
  - WRF-ASRC, a version used as part of an operational air quality forecasting program at the Atmospheric Sciences Research Center of the University of Albany (AQFMS, 2010).
  - WRF-AWS, a version of WRF operated at AWS Truepower in the US.
  - WRF-Meteotest, a version of WRF operated at Meteotest in Europe.
  - WRF-UJAEN, a version operated at the University of Jaén, Spain (Lara-Fanego et al., 2012).
4. The MASS model (Manobianco et al., 1996).
5. The Advanced Multiscale Regional Prediction System (ARPS) model (Xue et al., 2001).
6. The regional weather forecasting system Skiron (Kallos, 1997) operated and combined with statistical postprocessing based on learning machines at Spain's National Renewable Energy Center (CENER). (Skiron-CENER, Gastón et al., 2009).
7. The High Resolution Limited Area Model (HIRLAM, 2010) operational model from the Spanish weather service (AEMet) combined with a statistical postprocessing at CIEMAT (HIRLAM-Ciemat).
8. A model based on cloud cover predictions from the US National Digital Forecast Database, (NDFD) proposed by Perez et al. (2010).
9. BLUE FORECAST: statistical forecast tool of Bluesky based on the GFS predictions from NCEP.

10. Forecasts based on meteorologists' cloud cover forecasts by Bluesky (BLUESKY-meteorologists).

The first two models are directly based on global (planetary) NWP systems, respectively GEM, and ECMWF.

The native time step of the regional configuration of the GEM model and its ground resolution are 7.5 min and ~15 km, respectively. GEM forecasts of downward short-wave radiation flux at the surface (DSWRF) originating at 00:00Z and 12:00Z were de-archived by the Canadian Meteorological Centre at an hourly time step for this analysis. The de-archived forecasts cover North America and adjacent waters. As described by Pelland et al. (2011), the GEM solar forecasts were postprocessed by taking an average of the irradiance forecasts over a square region centered on the location of each site used in the validation. The size of this square grid was optimized for each station by selecting a size that minimized forecast root mean square error during a 1 year training period prior to the evaluation period used here.

ECMWF irradiance forecasts used here had a temporal resolution of 3 h and a spatial resolution of 25 km. The ECMWF site-specific, hourly data prepared for the present analysis according to Lorenz et al. (2009b) are obtained via time interpolation of the 3-hourly global clear sky indices. In addition, a bias correction that is dependent upon the cloud situation was performed for the European sites. This postprocessing was based on historic ground measured irradiance values for Germany and Switzerland, and on satellite derived irradiance values for Spain and Austria. For the US and Canadian sites no additional training to ground measured data was applied.

Models 3–7 are mesoscale models that use global weather models as an input to define regional boundary conditions, but add high resolution terrain and other features to produce higher resolution forecasts. In all cases analyzed here, the global weather model input is NOAA's GFS model. The GFS model dataset used for this project has a time resolution of 3 h and a nominal ground resolution of one by one degree (i.e., ~80 × 100 km in the considered latitude range). All the mesoscale models produce hourly output.

The WRF version of the model run by AWS Truepower as well as the MASS and ARPS models have a final ground resolution of 5 km. They are tested in two operational modes: with and without Model Output Statistics (MOS) postprocessing. The MOS process consists of integrating ongoing local irradiance measurements, when available, to correct localized errors from the numerical weather prediction process. This is a common operational forecasting practice: taking advantage of ongoing local surface and upper air measurements to deliver better forecasts.

The Advanced Research WRF model currently used in operational forecasting at the Atmospheric Sciences Research Center (WRF-ASRC) is a next-generation mesoscale numerical weather prediction system designed to serve both operational forecasting and atmospheric

research needs. It features multiple dynamical cores, a 3-dimensional variational (3DVAR) data assimilation system, and a software architecture allowing for computational parallelism and system extensibility. The operational version of this WRF model is version 3.2.1 and is run at a horizontal grid resolution of 12 km for domain encompassing the eastern section of the United States and Canada.

The two applications of WRF for Europe (Meteotest and U. Jaén) do not integrate postprocessing with ground measured values. WRF forecasts processed by the University of Jaén for a study region in Andalusia show a final spatial resolution of 3 km. The choice of the different parameterizations was based on a calibrating experiment for MM5, a former version of the WRF model, carried out for an optimum adjustment for the study region by Ruiz-Arias et al. (2008). WRF forecasts at Meteotest for Central Europe are processed with a grid size of 5 km × 5 km for the innermost domain. The forecasts are averaged using 10 × 10 model pixels around the point of interest corresponding to an area of 50 × 50 km.

Models 6 and 7 apply a postprocessing procedure to predictions of a mesoscale NWP model. CENER's solar global irradiance prediction scheme (model 6) is based on the regional weather forecasting system Skiron (Kallos, 1997), developed at the Hellenic National Meteorological Service, and operated with a final spatial resolution of 0.1° × 0.1°. The applied statistical postprocess is based on learning machines (Gastón et al., 2009). CIEMAT applies a bias correction to forecasts of the HIRLAM operational model of the Spanish weather service (AEMet) with a spatial resolution of 20 km × 20 km.

The statistical forecast tool BLUE FORECAST (model 9) is also based on the global GFS model. The original GFS forecasts with temporal resolutions of 3 and 6 h and spatial resolutions of 1° × 1° and 0.5° × 0.5° are integrated into a statistical postprocessing procedure using different methods of data mining such as ridge regression, automatic quadratic models or neural networks, based on meteorological inputs (see Natschläger et al., 2008).

The NDFD forecast does not provide irradiance per se, but cloud amount that extends up to 7 days ahead with a ground resolution of ~5 km and a time resolution of 3 h up to 3 days ahead and 6 h beyond that. The NDFD is also based on the GFS global model. GFS forecasts are individually processed by regional NOAA offices using mesoscale models and local observations and gridded nationally into the NDFD. The forecast cloud amounts are modeled into irradiance using an approach developed by Perez et al. (2010).

A similar approach is also operated by Bluesky for model 10. The meteorologists in the operational weather service use the results of several meteorological forecast models and combine these with their meteorological knowledge and forecasting experience. The result is cloud cover forecasts with hourly resolution in a first step. These are converted to solar irradiance forecasts using an

equation including the cloud cover coefficient and clear sky irradiances.

All forecasts are set to nominally originate at 00:00Z. In addition, some of the models are also tested with an origination time of 12:00Z. This 00:00Z common reference results in a slight performance handicap for the European validations compared to the North American validations; however as can be gauged from the results, e.g., by comparing the 00:00Z and 12:00Z performance, this is a very small effect.

### 3. Validation

The evaluation was performed for sites in the US, Canada and Europe covering different climatic conditions. These include Mediterranean climate in Southern Spain, humid continental climate in Canada, mostly continental climate in Central Europe and some high alpine stations in Switzerland, and finally arid, sub-tropical, semi-arid, and continental conditions in the US.

Because of operational contingencies not all the models could be validated at all the sites. Models 1–5 and 8 were validated in the US. Models 1, 2 and 3 (without MOS application) were validated against Canadian sites. Models 2, 3, 6, 7, 9 and 10 were validated in Europe. The common denominators to all the validations are (1) the ECMWF model and (2) the GFS-driven WRF model applied by various operators under slightly different conditions.

#### 3.1. Validation measurements

All benchmark measurement stations are part of networks operated by each considered country's weather services and include well maintained and quality controlled Class I instruments and data.

##### 3.1.1. United States

Validation measurements consist of hourly averaged global horizontal irradiance (GHI) recorded for a 1 year period (May 1st, 2009, through April 30th, 2010) at the seven stations of the SURFRAD network (SURFRAD, 2010). The stations are listed in Table 1.

Some of the models were only processed for a subset of the SURFRAD sites. The ARPS, MASS and WRF model processed by AWS Truepower could only be run at Desert Rock, Goodwin Creek and Penn State, while the WRF-ASRC model, run as part of the air quality forecast model, was only available for Goodwin Creek and Penn State.

All models were processed to deliver data up to 48 h ahead (next day and 2 day forecasts). The ECMWF forecasts were processed up to 3 days ahead, and the NDFD up to 7 days ahead.

##### 3.1.2. Canada

The three sites used for evaluating irradiance forecasts in Canada are listed in Table 2. The validation period runs from June 1, 2009 to May 31, 2010. The GEM, ECMWF

Table 1  
Location and climate type for the US sites.

Station	Latitude	Longitude	Elevation (m)	Climate
Goodwin Creek	34.25	89.87	98	Humid continental
Desert Rock	36.63	116.02	1107	Arid
Bondville	40.05	88.37	213	Continental
Boulder	40.13	105.24	1689	Semi-arid
Penn State	40.72	77.93	376	Humid continental
Sioux Falls	48.73	96.62	473	Continental
Fort Peck	48.31	105.1	643m	Continental

Table 2  
Location and climate type for the Canadian sites.

Station	Latitude (°)	Longitude (°)	Elevation (m)	Climate
Egbert	44.23	79.78	250	Humid continental
Bratt's Lake	50.20	104.71	580	Humid continental
Varenes	45.63	73.38	36	Humid continental

Table 3  
Location and climate type for the German sites.

Station	Latitude (°)	Longitude (°)	Elevation (m)	Climate
Fürstzell	48.55	−13.35	476	Continental
Stuttgart	48.83	−9.20	318	Continental
Würzburg	49.77	−9.97	275	Continental

Table 4  
Location and climate type for the Austrian sites.

Station	Latitude (°)	Longitude (°)	Elevation (m)	Climate
Linz	48.30	−14.30	266	Continental
Vienna	48.20	−16.37	171	Continental

and WRF-ASRC forecasts originating at 00:00Z were processed for forecast horizons of 0–48 h ahead, and compared to hourly average irradiance measured at the three

Table 5  
Location and climate type for the Swiss sites.

Station	Latitude (°)	Longitude (°)	Elevation (m)	Climate
Basel-Binningen	47.54	−7.58	316	Temperate Atlantic
Payerne	46.81	−6.94	490	Moderate maritime/continental
La Chaux-de-Fonds	47.09	−6.80	1018	Temperate Atlantic
Bern-Liebefeld	46.93	−7.42	565	Moderate maritime/continental
Buchs-Suhr	47.38	−8.08	387	Moderate maritime/continental
Napf	47.00	−7.94	1406	Moderate maritime/continental
Zürich SMA	47.38	−8.57	556	Moderate maritime/continental
Säntis	47.25	−9.34	2490	Alpine
St. Gallen	47.43	−9.40	779	Moderate maritime/continental
Genève-Cointrin	46.24	−6.12	420	Moderate maritime/continental
Sion	46.22	−7.34	482	Dry alpine
Montana	46.31	−7.49	1508	Alpine
Jungfrauoch	46.55	−7.99	3580	High alpine
Locarno-Magadino	46.16	−8.88	197	Warm temperate, humid
Weissfluhjoch	46.83	−9.81	2690	Alpine
Davos	46.81	−9.84	1590	Continental/alpine

ground stations. The mean of the individual forecast models was also evaluated against the ground station data to investigate whether this yields any benefits, as reported in the case of wind forecasts (Ernst et al., 2007).

In the case of WRF, forecasts were only available for two stations (Egbert and Varennes) for the last 2 months of the evaluation period (i.e. April 1, 2010 to May 31, 2010).

### 3.1.3. Europe

The selected data sets with hourly average values of measured irradiance for Europe cover four countries: Southern Germany, Switzerland including mountain stations, Austria, and Southern Spain. The period of evaluation for all sites and forecasting approaches is July 2007 to June 2008.

**3.1.3.1. German sites.** For the German sites (see Table 3) ground measurement data for three locations were provided by the German weather service (DWD). Forecast data of ECMWF, WRF-Meteotest, and BLUE FORECAST were considered for horizons up to 3 days ahead. Skiron-CENER forecasts were processed for 48 h.

**3.1.3.2. Austrian sites.** In addition to irradiance forecasts of ECMWF, WRF-Meteotest, Skiron-CENER and BLUE FORECAST, irradiance forecasts based on cloud cover

Table 6  
Location and climate type for the Spanish sites.

Station	Latitude (°)	Longitude (°)	Elevation (m)	Climate
Huelva	37.28	−6.91	19	Mediterranean
Córdoba	37.84	−4.85	91	Mediterranean
Granada	37.14	−3.63	687	Mediterranean

Table 7  
Overview of forecast validations.

	Forecast models – the number in () corresponds to the descriptive number in the text	Time horizon (days)
<i>Europe</i>		
Germany	ECMWF (2)	3
	WRF-Meteotest (3)	3
	SKIRON-CENER (6)	3
	BLUE FORECAST (9)	2
Switzerland	ECMWF (2)	3
	WRF-Meteotest (3)	3
	BLUE FORECAST (9)	3
Austria	ECMWF (2)	3
	WRF-Meteotest (3)	3
	CENER (6)	3
	BLUE FORECAST (9)	2
Spain	BLUESKY-Meteorologists (10)	2
	ECMWF (2)	3
	WRF-UJAEN (3)	3
	CENER (6)	3
	HIRLAM (7)	2
	BLUE FORECAST (9)	3
<i>USA</i>		
USA	GEM (1)	2
	ECMWF (2)	3
	WRF-ASRC(3)	2
	WRF-AWS <sup>a</sup> (3)	2
	MASS <sup>a</sup> (4)	2
	ARPS <sup>a</sup> (5)	2
	NDFD (8)	7
<i>Canada</i>		
Canada	GEM (1)	2
	ECMWF (2)	2
	WRF-ASRC (3)	2

<sup>a</sup> Models run both with and without MOS.

forecasts by the meteorologists' of Bluesky up to 48 h ahead were evaluated. The Austrian ground measurements (see Table 4) were recorded by BLUESKY in two locations.

**3.1.3.3. Swiss sites.** The models considered for the Swiss validation include ECMWF, WRF-Meteotest, and BLUE FORECAST. Ground measurements for sixteen sites are from the MeteoSwiss network. The sites considered for Switzerland cover a considerable variety in climatic conditions (see Table 5).

**3.1.3.4. Spanish sites.** Forecasts for Spain were processed based on the global ECMWF model and three different mesoscale models (WRF-Jaén, Skiron-CENER and

HIRLAM-CIEMAT). The three ground measurement stations (see Table 6) operated by the Spanish Weather Service AEMet are located in the South of Spain.

### 3.2. Overview of forecast model benchmarking tests

A summary of the models tested as part of this article is presented in Table 7. The ECMWF model and the GFS-driven WRF model are the only common denominators to all tests, noting that the WRF model was run by different entities in different countries, with slightly differing operational settings and was not available at some of the US and Canadian sites.

### 3.3. Concept of evaluation

To compare the performance of the different methods, hourly forecasts for the evaluation sites as provided by the different research groups and private companies were evaluated against hourly mean values of measured irradiance, regardless of the original spatial and temporal resolution of the underlying NWP models. The analysis presented focusses on the “end-use” accuracy of these site-specific, hourly irradiance predictions derived by the different forecast providers from gridded NWP data rather than on the evaluation of the direct NWP model output. To assess the performance of forecast algorithms, in general, a lot of different aspects have to be taken into account. In this paper, which aims at the inter-comparison of different models we focus on a few, basic measures of accuracy that are considered to be most relevant for the intended application of solar power prediction.

The validation metrics include the root mean square error, RMSE, to compare predicted irradiance  $I_{pred,i}$  to measured irradiance  $I_{meas,i}$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (I_{pred,i} - I_{meas,i})^2} \quad (1)$$

Here,  $N$  is the number of evaluated data pairs. The RMSE is often considered as the most important model validation metric in the context of renewable power forecasting. Because it is based on the square of differences between modeled and measured values, large forecast errors and outliers are weighted more strongly than small errors. It is suitable for applications where small errors are more tolerable and large forecast errors have a disproportionately high impact, which is the case for many aspects of grid management issues.

The MAE – the mean absolute error:

$$MAE = \frac{1}{N} \sum_{i=1}^N |I_{pred,i} - I_{meas,i}| \quad (2)$$

is a useful complement to the RMSE that is effective at quantifying the tightness of the measured-modeled scatter plot near the 1-to-1 line. In particular it is appropriate

Table 8  
Relative RMSE US.

% RMSE			Bondville	Boulder	Desert Rock	Fort Peck	Goodwin Creek	Penn State	Sioux Falls	Composite
Mean GHI ( $W m^{-2}$ )			335	374	466	326	363	298	328	356
Reference satellite model			21%	25%	15%	23%	20%	28%	22%	22%
Persistence	0:00Z	Day 1	59%	51%	29%	46%	51%	65%	51%	50%
GEM	0:00Z	Day 1	35%	38%	21%	30%	33%	38%	38%	33%
GEM	12:00Z	Day 1	33%	36%	20%	29%	33%	38%	36%	32%
ECMWF	0:00Z	Day 1	34%	38%	21%	32%	31%	39%	38%	33%
NDFD	0:00Z	Day 1	40%	44%	25%	38%	38%	48%	44%	40%
NDFD	12:00Z	Day 1	40%	43%	23%	37%	37%	45%	43%	38%
WRF-ASRC		Day 1	46%				43%	51%		44%
MASS	0:00Z	Day 1			31%		53%	67%		55%
MASS	12:00Z	Day 1			32%		55%	64%		54%
MAS-MOS	0:00Z	Day 1			24%		38%	44%		38%
MAS-MOS	12:00Z	Day 1			24%		38%	44%		38%
WRF-AWS	0:00Z	Day 1			25%		45%	54%		45%
WRF-AWS	12:00Z	Day 1			26%		47%	58%		47%
WRF-AWS-MOS	0:00Z	Day 1			23%		41%	47%		40%
WRF-AWS-MOS	12:00Z	Day 1			23%		41%	46%		40%
ARPS	0:00Z	Day 1			33%		54%	69%		56%
ARPS-MOS	0:00Z	Day 1			24%		43%	48%		42%
GEMS/ECMWF	0:00Z	Day 1	32%	37%	20%	30%	31%	37%	37%	32%
Persistence	0:00Z	Day 2	64%	57%	32%	49%	60%	72%	57%	56%
GEM	0:00Z	Day 2	37%	37%	21%	32%	35%	40%	37%	34%
GEM	12:00Z	Day 2	34%	36%	22%	30%	33%	39%	36%	33%
ECMWF	0:00Z	Day 2	38%	39%	22%	34%	34%	41%	39%	35%
NDFD	0:00Z	Day 2	43%	45%	27%	39%	40%	49%	45%	41%
NDFD	12:00Z	Day 2	42%	45%	25%	38%	39%	48%	45%	40%
WRF-ASRC		Day 2	50%				45%	55%		46%
MASS	0:00Z	Day 2			31%		57%	68%		56%
MASS	12:00Z	Day 2			31%		58%	66%		55%
MAS-MOS	0:00Z	Day 2			24%		40%	46%		39%
MAS-MOS	12:00Z	Day 2			24%		40%	46%		39%
WRF-AWS	0:00Z	Day 2			27%		47%	59%		47%
WRF-AWS	12:00Z	Day 2			26%		46%	58%		46%
WRF-AWS-MOS	0:00Z	Day 2			24%		42%	49%		41%
WRF-AWS-MOS	12:00Z	Day 2			23%		41%	48%		40%
ARPS	0:00Z	Day 2			33%		55%	70%		57%
ARPS-MOS	0:00Z	Day 2			25%		44%	50%		42%
GEMS/ECMWF	0:00Z	Day 2	35%	37%	20%	31%	33%	38%	37%	33%
Persistence	0:00Z	Day 3	67%	58%	32%	54%	63%	77%	58%	58%
ECMWF	0:00Z	Day 3	40%	41%	23%	35%	37%	45%	41%	37%
NDFD	0:00Z	Day 3	47%	46%	29%	39%	44%	54%	46%	44%
NDFD	12:00Z	Day 3	45%	46%	28%	38%	42%	51%	46%	42%
Persistence	0:00Z	Day 4	69%	59%	33%	54%	62%	79%	59%	59%
NDFD	0:00Z	Day 4	49%	46%	29%	39%	46%	55%	46%	44%
NDFD	12:00Z	Day 4	47%	46%	29%	38%	45%	55%	46%	44%
Persistence	0:00Z	Day 5	71%	59%	33%	52%	63%	78%	59%	59%
NDFD	0:00Z	Day 5	52%	47%	29%	41%	47%	58%	47%	46%
NDFD	12:00Z	Day 5	51%	47%	29%	40%	48%	58%	47%	45%
Persistence	0:00Z	Day 6	68%	59%	33%	54%	60%	78%	59%	59%
NDFD	0:00Z	Day 6	56%	49%	29%	43%	50%	61%	49%	48%
NDFD	12:00Z	Day 6	56%	50%	30%	42%	48%	59%	50%	48%
Persistence	0:00Z	Day 7	67%	60%	34%	54%	60%	75%	60%	59%
NDFD	0:00Z	Day 7	57%	51%	31%	45%	54%	61%	51%	50%
NDFD	12:00Z	Day 7	56%	51%	30%	44%	52%	59%	51%	49%

for applications with linear cost functions, that is, where the costs that are caused by a wrong forecast are proportional to the forecast error.

The MBE-mean bias error:

$$MBE = \frac{1}{N} \sum_{i=1}^N (I_{pred,i} - I_{meas,i}) \quad (3)$$

describes systematic deviations of a forecast. The agreement between the distribution functions of measured and predicted time series can be evaluated using the Kolmogorov–Smirnov test integral (KSI) (e.g., see Perez et al., 2010). We decided to use a robust interpretation of the KSI metric that simply describes the integrated absolute difference between the predicted and measured normalized

Table 9  
Relative RMSE Central Europe.

% RMSE			Fürstentzell	Stuttgart	Würzburg	Composite Germany	Linz	Wien	Composite Austria	Composite Switzerland
Mean GHI ( $\text{W m}^{-2}$ )			227	233	224	228	206	241	224	270
Reference satellite model										
Persistence	0:00Z	Day 1	66%	63%	61%	64%	71%	57%	64%	58%
ECMWF-OL	0:00Z	Day 1	40%	40%	42%	40%	50%	42%	46%	40%
BLUE FORECAST	0:00Z	Day 1	41%	42%	42%	42%	46%	43%	45%	41%
WRF-Meteotest	0:00Z	Day 1	48%	51%	57%	52%	64%	47%	55%	44%
CENER	0:00Z	Day 1	46%	51%	53%	50%	63%	53%	58%	
Meteorologists							55%	46%	50%	
Persistence	0:00Z	Day 2	74%	69%	68%	70%	78%	63%	70%	64%
ECMWF-OL	0:00Z	Day 2	41%	42%	42%	42%	52%	43%	47%	42%
BLUE FORECAST	0:00Z	Day 2	43%	45%	44%	44%	49%	41%	45%	42%
WRF-Meteotest	0:00Z	Day 2	51%	55%	59%	55%	64%	53%	59%	46%
CENER	0:00Z	Day 2	48%	54%	56%	53%	65%	54%	60%	
Meteorologists							55%	44%	49%	
Persistence	0:00Z	Day 3	75%	74%	71%	73%	78%	65%	72%	67%
ECMWF	0:00Z	Day 3	44%	46%	45%	45%	54%	47%	51%	43%
BLUE FORECAST	0:00Z	Day 3	45%	47%	45%	46%	51%	45%	48%	44%
WRF-Meteotest	0:00Z	Day 3	57%	62%	63%	61%	67%	58%	63%	51%

cumulative distributions  $CDF_{pred}$  and  $CDF_{meas}$  integrated over all irradiance levels  $I$  and normalized to 1,

$$KSI = \frac{1}{I_{\max}} \int_0^{I_{\max}} |CDF_{pred} - CDF_{meas}| dI \quad (4)$$

The evaluation of distribution functions is helpful e.g. for applications where decisions are related to threshold values. However, the KSI metric is less important for forecast evaluation than the other metrics introduced and is given here only for the Canadian and US sites, where a discretized version of Eq. (4) was used to evaluate the KSI metric.

The accuracy measures are calculated using only daytime hours ( $I > 0$ ) (i.e., night values with zero irradiance are excluded from the evaluation.) The evaluation results are grouped according to forecast days. For a model run at 00:00Z, the results for the first forecast day (intraday) integrate forecast horizons up to 24 h, the second forecast day (day-ahead) integrates forecast horizons from 25 to 48 h, and so on. The reason for grouping results according to forecast days rather than forecast hours is the strong dependency of forecast accuracy on the daytime caused by the daily course of irradiance.

Relative values of the error measures are obtained by normalization to the mean ground measured irradiance of the considered period.

As an additional quality check, forecasts often are compared to trivial reference models, which are the result of simple considerations and not of modeling efforts. It is worthwhile to implement and run a complex forecasting tool if it is able to clearly outperform trivial (i.e., self-evident) reference models. The most common such reference model for short term forecasts is persistence. Persistence consists of projecting currently and recently measured conditions into the future while accounting for solar geometry changes. Here, where we are inter-comparing NWP models

originating nominally at 00:00Z, designed for next and subsequent day forecasts, the benchmark persistence is obtained by determining the daily global clear sky index  $kt^*$  (ratio of measured irradiance to irradiance for clear sky conditions) from the last available day and projecting this index for all subsequent forecast days/hours.

Forecast skill can be gauged by comparing the forecast and reference (i.e., persistence) errors as follows:

$$MSE \text{ skill score} = (MSE_{ref} - MSE_{forecast}) / MSE_{ref} \quad (5)$$

where MSE is the mean square error (square of the RMSE as defined in Eq. (1)). A MSE skill score of one corresponds to a perfect model. Negative MSE skill scores indicate performance worse than persistence.

For the US sites, the satellite irradiance model developed by Perez et al. (2002) and used in the NSRDB (2005) and SolarAnywhere (2010) is used as a complementary reference to gauge the performance of the forecast models – note that this reference model is an “existing conditions” and not a forecast model.

Results of the forecast evaluation are provided at different levels of detail. Tables 8–21 give the different validation metrics for the single sites. (As an exception, for Switzerland with more than 15 stations and the same forecast models available for all stations, the average of the errors of the individual sites is given instead.) These detailed results allow for directly assessing and comparing the performance of different forecasting methods for a given location with its particular climatic conditions, which is of interest not only from the scientific point of view. Forecast users, e.g. a utility company or a plant operator, are also often interested in applying the forecasts and hence in the relevant information about forecast accuracy for a certain location or region.

In addition to the evaluation and model comparison for the single sites, all-site composite errors for the different

Table 10  
Relative RMSE SPAIN.

% RMSE			Cordoba	Granada	Huelva	COMPOSITE SPAIN
Mean GHI (W m <sup>-2</sup> )			443	409	407	420
Reference satellite model						
Persistence	0:00Z	Day 1	34%	36%	34%	35%
ECMWF-OL	0:00Z	Day 1	23%	23%	20%	22%
CENER	0:00Z	Day 1	26%	25%	26%	25%
WRF-UJAEN	0:00Z	Day 1	28%	27%	25%	26%
HIRLAM	0:00Z	Day 1	26%	32%	26%	29%
Persistence	0:00Z	Day 2	37%	39%	38%	38%
ECMWF-OL	0:00Z	Day 2	25%	22%	21%	23%
CENER	0:00Z	Day 2	30%	26%	27%	27%
WRF-UJAEN	0:00Z	Day 2	29%	29%	27%	28%
HIRLAM	0:00Z	Day 2	29%	36%	32%	33%
Persistence	0:00Z	Day 3	29%	41%	39%	40%
ECMWF	0:00Z	Day 3	29%	24%	22%	23%
WRF-UJAEN	0:00Z	Day 3	29%	30%	30%	30%
HIRLAM	0:00Z	Day 3	29%	39%	36%	35%

Table 11  
Relative RMSE Canada.

% RMSE			Egbert	Bratt's Lake	Varennes	Composite
Mean GHI (W m <sup>-2</sup> )			320	306	306	311
Reference satellite model						
Persistence	0:00Z	Day 1	52%	52%	58%	54%
GEM	0:00Z	Day 1	32%	31%	37%	33%
ECMWF	0:00Z	Day 1	32%	31%	35%	32%
WRF-ASRC <sup>a</sup>	0:00Z	Day 1	40%		44%	42%
GEM/ECMWF/WRF-ASRC <sup>a</sup>	0:00Z	Day 1	31%		33%	30%
GEM/ECMWF	0:00Z	Day 1	31%	29%	34%	31%
Persistence	0:00Z	Day 2	56%	57%	63%	59%
GEM	0:00Z	Day 2	33%	35%	38%	35%
ECMWF	0:00Z	Day 2	34%	35%	38%	36%
WRF-ASRC <sup>a</sup>	0:00Z	Day 2	43%		45%	44%
GEM/ECMWF/WRF <sup>a</sup>	0:00Z	Day 2	32%		36%	32%
GEM/ECMWF	0:00Z	Day 2	32%	33%	36%	34%

<sup>a</sup> The WRF model was only run on a 2 month data subset and results were prorated using the other models as a template.

evaluation regions (US, Canada, and Europe) are calculated by averaging the errors of the individual sites, in order to give a quick overview of model performances. For some of the models forecasts are available only for a subset of sites in a given region. For these models *i* an estimate of the all-site composite value, e.g. the  $RMSE_{all-site,i}^*$ , is prorated with the following equation:

$$RMSE_{all-sites,i}^* = RMSE_{subset,i} \frac{\sum_{j=1}^M RMSE_{all-sites,j}}{\sum_{j=1}^M RMSE_{subset,j}} \quad (6)$$

i.e. by multiplying the composite  $RMSE_{subset,i}$  for the subset of sites at which the forecast is available with the ratio of the average all-site composite RMSE to the average subset composite RMSE of all models *j* that are available for all sites. This estimate of the average performance is of course provided with some uncertainty. In particular, averaging over sites with different climatic conditions may result in biased overall estimates – note that this is also the reason why composite values for Northern and Southern

Europe are given separately. However, given the availability of the detailed site-specific results in Tables 8–21, we consider it to be a reasonable simplification.

#### 4. Results and discussion

An overview of the all-site composite relative RMSE values for the different study regions US, Canada, Central Europe and Spain is given in Figs. 1–4. Corresponding RMSE values for the single sites are given Tables 8–11 respectively.

Figs. 5–14 accordingly provide composite summaries for the MAE, MBE and KSI metrics, also completed by the detailed site specific results in Tables 12–15 for the MAE, 16–19 for the MBE and 20–21 for the KSI.

We first give a description and discussion of the US results, which include the largest number of different forecast models and also cover different climate zones. Next,

Table 12  
Relative MAE US.

% MAE			Bondville	Boulder	Desert Rock	Fort Peck	Goodwin Creek	Penn State	Sioux Falls	Composite
Mean GHI ( $W m^{-2}$ )			335	374	466	326	363	298	328	356
Reference satellite model			14%	17%	9%	16%	13%	18%	15%	14%
Persistence	0:00Z	Day 1	39%	34%	18%	29%	34%	44%	34%	33%
GEM	0:00Z	Day 1	24%	24%	11%	19%	21%	26%	24%	21%
GEM	12:00Z	Day 1	23%	23%	11%	18%	22%	26%	23%	21%
ECMWF	0:00Z	Day 1	21%	23%	11%	19%	20%	25%	23%	21%
NDFD	0:00Z	Day 1	26%	28%	14%	23%	23%	30%	28%	25%
NDFD	12:00Z	Day 1	26%	27%	14%	23%	23%	29%	27%	24%
WRF-ASRC		Day 1	30%				28%	34%		28%
MASS	0:00Z	Day 1			21%		39%	49%		40%
MASS	12:00Z	Day 1			22%		40%	47%		39%
MAS-MOS	0:00Z	Day 1			15%		27%	31%		27%
MAS-MOS	12:00Z	Day 1			15%		27%	32%		27%
WRF-AWS	0:00Z	Day 1			16%		29%	37%		29%
WRF-AWS	12:00Z	Day 1			16%		29%	39%		31%
WRF-AWS-MOS	0:00Z	Day 1			14%		28%	34%		28%
WRF-AWS-MOS	12:00Z	Day 1			14%		28%	33%		27%
ARPS	0:00Z	Day 1			23%		39%	49%		40%
ARPS-MOS	0:00Z	Day 1			15%		30%	34%		29%
GEMS/ECMWF	0:00Z	Day 1	21%	23%	11%	18%	19%	25%	23%	20%
Persistence	0:00Z	Day 2	44%	39%	19%	32%	41%	50%	39%	38%
GEM	0:00Z	Day 2	25%	24%	12%	20%	22%	27%	24%	22%
GEM	12:00Z	Day 2	23%	23%	12%	19%	21%	26%	23%	21%
ECMWF	0:00Z	Day 2	24%	24%	12%	21%	21%	27%	24%	22%
NDFD	0:00Z	Day 2	28%	29%	16%	24%	25%	32%	29%	26%
NDFD	12:00Z	Day 2	27%	29%	15%	24%	24%	30%	29%	25%
WRF-ASRC		Day 2	32%				29%	37%		30%
MASS	0:00Z	Day 2			21%		42%	49%		40%
MASS	12:00Z	Day 2			21%		43%	47%		40%
MAS-MOS	0:00Z	Day 2			15%		28%	33%		27%
MAS-MOS	12:00Z	Day 2			15%		28%	33%		27%
WRF-AWS	0:00Z	Day 2			17%		30%	39%		31%
WRF-AWS	12:00Z	Day 2			16%		29%	39%		30%
WRF-AWS-MOS	0:00Z	Day 2			15%		29%	34%		28%
WRF-AWS-MOS	12:00Z	Day 2			14%		29%	34%		28%
ARPS	0:00Z	Day 2			23%		40%	50%		41%
ARPS-MOS	0:00Z	Day 2			15%		31%	34%		29%
GEMS/ECMWF	0:00Z	Day 2	23%	23%	11%	19%	21%	26%	23%	21%
Persistence	0:00Z	Day 3	46%	40%	20%	36%	44%	54%	40%	40%
ECMWF	0:00Z	Day 3	25%	25%	12%	21%	23%	30%	25%	23%
NDFD	0:00Z	Day 3	31%	31%	17%	25%	28%	36%	31%	28%
NDFD	12:00Z	Day 3	30%	30%	16%	24%	26%	34%	30%	27%
Persistence	0:00Z	Day 4	49%	41%	20%	35%	43%	56%	41%	41%
NDFD	0:00Z	Day 4	34%	32%	17%	25%	30%	38%	32%	30%
NDFD	12:00Z	Day 4	32%	32%	17%	24%	29%	37%	32%	29%
Persistence	0:00Z	Day 5	50%	41%	20%	34%	44%	55%	41%	41%
NDFD	0:00Z	Day 5	36%	32%	17%	26%	31%	41%	32%	31%
NDFD	12:00Z	Day 5	35%	32%	17%	26%	31%	40%	32%	30%
Persistence	0:00Z	Day 6	48%	41%	20%	35%	42%	55%	41%	40%
NDFD	0:00Z	Day 6	39%	34%	17%	28%	34%	43%	34%	33%
NDFD	12:00Z	Day 6	39%	34%	18%	27%	32%	42%	34%	32%
Persistence	0:00Z	Day 7	47%	42%	21%	36%	41%	53%	42%	40%
NDFD	0:00Z	Day 7	40%	36%	18%	29%	37%	43%	36%	34%
NDFD	12:00Z	Day 7	40%	35%	18%	29%	35%	42%	35%	33%

the discussion is extended to the evaluation for Canada and Europe and some additional findings are highlighted.

RMSE all-site composite values for the US given in Fig. 1 show a considerable spread for the different models. They range between 32% and 47% for Day 1 forecasts and – showing only a slight increase – between 34% and 48% for Day 2 forecasts. The corresponding values of MAE

(Fig. 5) lie between 20% and 29% for Day 1 and between 22% and 31% for Day 2 forecasts.

Lowest MAE and RMSE values are found for the global model ECMWF and GEM irradiance forecasts. All considered mesoscale-model forecasts (WRF-AFS, WRF-ASRC, ARPS, MAS) as well as the NDFD based forecasts show larger forecast errors. This indicates some

Table 13  
Relative MAE Central Europe.

% MAE			Fürstzell	Stuttgart	Würzburg	Composite	Linz	Wien	Composite Austria	Composite Switzerland
Mean GHI ( $W m^{-2}$ )			227	233	224	228	206	241	224	270
Reference satellite model										
Persistence	0:00Z	Day 1	42%	40%	39%	41%	46%	36%	41%	39%
ECMWF-OL	0:00Z	Day 1	26%	26%	27%	26%	32%	26%	29%	26%
BLUE FORECAST	0:00Z	Day 1	26%	28%	28%	27%	28%	27%	28%	27%
WRF-Meteotest	0:00Z	Day 1	30%	32%	37%	33%	40%	29%	35%	28%
CENER	0:00Z	Day 1	29%	32%	33%	32%	43%	35%	39%	
Meteorologists										
Persistence	0:00Z	Day 2	48%	45%	44%	46%	50%	41%	46%	43%
ECMWF-OL	0:00Z	Day 2	27%	28%	28%	28%	34%	27%	31%	27%
BLUE FORECAST	0:00Z	Day 2	28%	30%	29%	29%	30%	27%	28%	28%
WRF-Meteotest	0:00Z	Day 2	32%	34%	38%	35%	40%	34%	37%	29%
CENER	0:00Z	Day 2	31%	34%	36%	34%	44%	35%	40%	
Meteorologists										
Persistence	0:00Z	Day 3	49%	48%	47%	48%	51%	42%	47%	45%
ECMWF	0:00Z	Day 3	29%	30%	29%	30%	35%	30%	32%	28%
BLUE FORECAST	0:00Z	Day 3	29%	31%	31%	30%	32%	30%	31%	30%
WRF-Meteotest	0:00Z	Day 3	36%	38%	40%	38%	42%	37%	40%	32%

Table 14  
Relative MAE Spain.

% MAE			Cordoba	Granada	Huelva	Composite Spain
Mean GHI ( $W m^{-2}$ )			443	409	407	420
Reference satellite model						
Persistence	0:00Z	Day 1	20%	19%	19%	19%
ECMWF-OL	0:00Z	Day 1	15%	13%	12%	13%
CENER	0:00Z	Day 1	16%	16%	17%	16%
WRF-UJAEN	0:00Z	Day 1	15%	14%	13%	14%
HIRLAM	0:00Z	Day 1	19%	25%	19%	21%
Day 2						
Persistence	0:00Z	Day 2	22%	21%	21%	22%
ECMWF-OL	0:00Z	Day 2	16%	13%	12%	14%
CENER	0:00Z	Day 2	18%	17%	17%	17%
WRF-UJAEN	0:00Z	Day 2	16%	15%	14%	15%
HIRLAM	0:00Z	Day 2	21%	27%	23%	24%
Day 3						
Persistence	0:00Z	Day 3	24%	23%	23%	23%
ECMWF	0:00Z	Day 3	16%	14%	13%	14%
WRF-UJAEN	0:00Z	Day 3	16%	15%	16%	16%
HIRLAM	0:00Z	Day 3	22%	30%	25%	24%

Table 15  
Relative MAE Canada.

% MAE			Egbert	Bratt's Lake	Varenes	Composite
Mean GHI ( $W m^{-2}$ )			320	306	306	311
Reference satellite model						
Persistence	0:00Z	Day 1	37%	37%	41%	38%
GEM	0:00Z	Day 1	23%	20%	25%	23%
ECMWF	0:00Z	Day 1	20%	19%	22%	21%
WRF-ASRC <sup>a</sup>	0:00Z	Day 1	27%		30%	28%
GEM/ECMWF/WRF-ASRC <sup>a</sup>	0:00Z	Day 1	21%		22%	20%
GEM/ECMWF	0:00Z	Day 1	21%	19%	23%	21%
Day 2						
Persistence	0:00Z	Day 2	41%	39%	46%	42%
GEM	0:00Z	Day 2	23%	22%	25%	23%
ECMWF	0:00Z	Day 2	22%	21%	25%	23%
WRF-ASRC <sup>a</sup>	0:00Z	Day 2	30%		32%	31%
GEM/ECMWF/WRF <sup>a</sup>	0:00Z	Day 2	23%		24%	22%
GEM/ECMWF	0:00Z	Day 2	22%	21%	24%	22%

<sup>a</sup> The WRF model was only run on a 2 month data subset and results were prorated using the other models as a template.

Table 16  
Relative MBE US.

% MBE			Bondville	Boulder	Desert Rock	Fort Peck	Goodwin Creek	Penn State	Sioux Falls	Composite
Mean GHI ( $W m^{-2}$ )			335	374	466	326	363	298	328	356
Reference satellite model			0%	0%	2%	0%	1%	2%	2%	1%
Persistence	0:00Z	Day 1	-2%	-1%	-2%	-2%	-1%	-1%	-1%	-1%
GEM	0:00Z	Day 1	8%	10%	2%	6%	8%	11%	10%	8%
GEM	12:00Z	Day 1	7%	7%	3%	5%	7%	12%	7%	7%
ECMWF	0:00Z	Day 1	6%	14%	5%	9%	6%	12%	14%	10%
NDFD	0:00Z	Day 1	-7%	-9%	-1%	4%	-6%	-8%	-9%	-5%
NDFD	12:00Z	Day 1	-9%	-10%	-1%	2%	-8%	-9%	-10%	-6%
WRF-ASRC		Day 1	9%				13%	13%		15%
MASS	0:00Z	Day 1			19%		34%	41%		37%
MASS	12:00Z	Day 1			18%		34%	40%		36%
MAS-MOS	0:00Z	Day 1			-1%		1%	0%		0%
MAS-MOS	12:00Z	Day 1			0%		-1%	0%		-1%
WRF-AWS	0:00Z	Day 1			1%		19%	23%		17%
WRF-AWS	12:00Z	Day 1			1%		18%	22%		16%
WRF-AWS-MOS	0:00Z	Day 1			1%		-1%	0%		0%
WRF-AWS-MOS	12:00Z	Day 1			0%		0%	0%		0%
ARPS	0:00Z	Day 1			20%		33%	41%		37%
ARPS-MOS	0:00Z	Day 1			0%		-2%	0%		-1%
GEMS/ECMWF	0:00Z	Day 1	7%	12%	4%	8%	7%	12%	12%	9%
Persistence	0:00Z	Day 2	-2%	-2%	-2%	-2%	-2%	-1%	-2%	-2%
GEM	0:00Z	Day 2	6%	7%	2%	5%	8%	11%	7%	7%
GEM	12:00Z	Day 2	6%	7%	3%	4%	6%	10%	7%	6%
ECMWF	0:00Z	Day 2	5%	14%	5%	9%	6%	10%	14%	9%
NDFD	0:00Z	Day 2	-8%	-9%	-3%	3%	-6%	-7%	-9%	-6%
NDFD	12:00Z	Day 2	-8%	-10%	-2%	2%	-8%	-9%	-10%	-6%
WRF-ASRC		Day 2	8%				14%	12%		14%
MASS	0:00Z	Day 2			18%		36%	40%		35%
MASS	12:00Z	Day 2			18%		37%	36%		34%
MAS-MOS	0:00Z	Day 2			-1%		-1%	-2%		-1%
MAS-MOS	12:00Z	Day 2			-1%		1%	-2%		-1%
WRF-AWS	0:00Z	Day 2			1%		18%	22%		15%
WRF-AWS	12:00Z	Day 2			1%		17%	21%		15%
WRF-AWS-MOS	0:00Z	Day 2			0%		-1%	-1%		-1%
WRF-AWS-MOS	12:00Z	Day 2			0%		-1%	-1%		-1%
ARPS	0:00Z	Day 2			20%		31%	41%		34%
ARPS-MOS	0:00Z	Day 2			0%		-3%	0%		-1%
GEMS/ECMWF	0:00Z	Day 2	6%	10%	4%	7%	7%	11%	10%	8%
Persistence	0:00Z	Day 3	-2%	-2%	-2%	-1%	-2%	-1%	-2%	-2%
ECMWF	0:00Z	Day 3	5%	13%	5%	9%	6%	10%	13%	9%
NDFD	0:00Z	Day 3	-6%	-9%	-4%	3%	-8%	-10%	-9%	-6%
NDFD	12:00Z	Day 3	-8%	-10%	-3%	1%	-8%	-10%	-10%	-7%
Persistence	0:00Z	Day 4	-2%	-2%	-2%	-1%	-3%	-2%	-2%	-2%
NDFD	0:00Z	Day 4	-5%	-7%	-3%	4%	-7%	-7%	-7%	-5%
NDFD	12:00Z	Day 4	-7%	-8%	-4%	4%	-9%	-9%	-8%	-6%
Persistence	0:00Z	Day 5	-2%	-2%	-2%	-1%	-2%	-2%	-2%	-2%
NDFD	0:00Z	Day 5	-4%	-5%	-3%	6%	-7%	-7%	-5%	-4%
NDFD	12:00Z	Day 5	-6%	-6%	-3%	4%	-8%	-7%	-6%	-5%
Persistence	0:00Z	Day 6	-2%	-2%	-2%	-1%	-2%	-2%	-2%	-2%
NDFD	0:00Z	Day 6	-3%	-4%	-1%	6%	-6%	-8%	-4%	-3%
NDFD	12:00Z	Day 6	-4%	-5%	-2%	6%	-8%	-6%	-5%	-4%
Persistence	0:00Z	Day 7	-2%	-2%	-2%	-1%	-2%	-2%	-2%	-2%
NDFD	0:00Z	Day 7	-2%	-5%	-1%	5%	-7%	-7%	-5%	-3%
NDFD	12:00Z	Day 7	-2%	-5%	-2%	5%	-8%	-6%	-5%	-3%

shortcomings in the selected mesoscale models' radiation and/or cloud schemes. Another reason might be the use of lateral boundary conditions from GFS, used to initialize all meso-scale models evaluated here. In recent work by Mathiesen and Kleissl (2011), the GFS model irradiance forecasts were found to have a similar performance to

those of the ECMWF model when applying a simple post-processing. This suggests that the performance difference noted here between the ECMWF and GEM model on the one hand and the different meso-scale models initialized with GFS on the other hand has more to do with the meso-scale models themselves than with the GFS boundary

Table 17  
Relative MBE Central Europe.

% MBE		Fürstzell	Stuttgart	Würzburg	COMPOSITE	Linz	Wien	Composite Austria	Composite Switzerland
Mean GHI ( $W m^{-2}$ )		227	233	224	228	206	241	224	270
Reference satellite model									
Persistence	0:00Z Day 1	−3%	−2%	−1%	−2%	−11%	−2%	−6%	−6%
ECMWF-OL	0:00Z Day 1	−1%	−4%	−4%	−3%	12%	2%	7%	0%
BLUE	0:00Z Day 1	0%	−4%	−1%	−1%	0%	1%	1%	−3%
FORECAST									
WRF-Meteotest	0:00Z Day 1	1%	0%	−1%	0%	28%	14%	21%	1%
CENER	0:00Z Day 1	7%	3%	8%	6%	21%	6%	14%	
Meteorologists	0:00Z Day 1					9%	−1%	0%	
Persistence	0:00Z Day 2	−3%	−3%	−2%	−3%	−11%	−2%	−7%	−7%
ECMWF-OL	0:00Z Day 2	−1%	−4%	−5%	−3%	12%	0%	6%	0%
BLUE	0:00Z Day 2	1%	−3%	−1%	−1%	1%	0%	1%	−1%
FORECAST									
WRF-Meteotest	0:00Z Day 2	4%	−1%	−11%	−5%	25%	2%	13%	1%
CENER	0:00Z Day 2	5%	0%	6%	4%	18%	1%	9%	
Meteorologists	0:00Z Day 2					8%	−2%	3%	
Persistence	0:00Z Day 3	−4%	−3%	−2%	−3%	−11%	−2%	−7%	−7%
ECMWF	0:00Z Day 3	−2%	−4%	−5%	−3%	13%	2%	8%	−1%
BLUE	0:00Z Day 3	2%	−3%	−1%	−1%	1%	3%	2%	−1%
FORECAST									
WRF-Meteotest	0:00Z Day 3	2%	−7%	−12%	−6%	24%	2%	14%	−1%

Table 18  
Relative MBE Spain.

% MBE		Cordoba	Granada	Huelva	Composite Spain
Mean GHI ( $W m^{-2}$ )		443	409	407	420
Reference satellite model					
Persistence	0:00Z Day 1	0%	1%	0%	0%
ECMWF-OL	0:00Z Day 1	−2%	2%	0%	0%
CENER	0:00Z Day 1	2%	2%	−4%	−1%
WRF-UJAEN	0:00Z Day 1	9%	7%	4%	6%
HIRLAM	0:00Z Day 1	−6%	−16%	−7%	−10%
Persistence	0:00Z Day 2	0%	1%	0%	0%
ECMWF-OL	0:00Z Day 2	−3%	2%	−1%	0%
CENER	0:00Z Day 2	−1%	1%	−3%	−1%
WRF-UJAEN	0:00Z Day 2	9%	6%	5%	7%
HIRLAM	0:00Z Day 2	−5%	−17%	−10%	−12%
Persistence	0:00Z Day 3	0%	1%	0%	0%
ECMWF	0:00Z Day 3	−2%	1%	0%	0%
WRF-UJAEN	0:00Z Day 3	9%	7%	5%	7%
HIRLAM	0:00Z Day 3	−7%	−18%	−9%	−9%

Table 19  
Relative MBE Canada.

% MBE		Egbert	Bratt's Lake	Varenes	Composite
Mean GHI ( $W m^{-2}$ )		320	306	306	311
Reference satellite model					
Persistence	0:00Z Day 1	−4%	−8%	−6%	−6%
GEM	0:00Z Day 1	2%	2%	−2%	1%
ECMWF	0:00Z Day 1	4%	4%	0%	3%
WRF-ASRC <sup>a</sup>	0:00Z Day 1	2%		−1%	0%
GEM/ECMWF/WRF-ASRC <sup>a</sup>	0:00Z Day 1	2%		0%	1%
GEM/ECMWF	0:00Z Day 1	3%	3%	−1%	2%
Persistence	0:00Z Day 2	−5%	−9%	−6%	−6%
GEM	0:00Z Day 2	1%	1%	−1%	1%
ECMWF	0:00Z Day 2	1%	5%	−1%	2%
WRF-ASRC <sup>a</sup>	0:00Z Day 2	0%		6%	2%
GEM/ECMWF/WRF <sup>a</sup>	0:00Z Day 2	−1%		3%	1%
GEM/ECMWF	0:00Z Day 2	1%	3%	−1%	1%

<sup>a</sup> The WRF model was only run on a 2 month data subset and results were prorated using the other models as a template.

Table 20  
KSI \* 100 US.

KSI * 100			Bondville	Boulder	Desert Rock	Fort Peck	Goodwin Creek	Penn State	Sioux Falls	Composite
Mean GHI ( $W m^{-2}$ )			335	374	466	326	363	298	328	356
Reference satellite model			1.2	0.5	1.0	0.8	0.9	1.2	0.5	0.9
Persistence	0:00Z	Day 1	1.7	3.1	2.0	2.0	1.6	2.0	3.1	2.2
GEM	0:00Z	Day 1	3.4	3.9	1.6	2.5	3.2	4.3	3.9	3.3
GEM	12:00Z	Day 1	3.4	3.2	1.6	2.1	3.7	4.0	3.2	3.0
ECMWF	0:00Z	Day 1	2.6	5.6	2.6	3.0	3.3	4.0	5.6	3.8
NDFD	0:00Z	Day 1	2.6	3.6	1.7	1.5	2.8	2.4	3.6	2.6
NDFD	12:00Z	Day 1	3.1	3.6	1.6	0.8	3.5	2.8	3.6	2.7
WRF-ASRC		Day 1	3.7				4.9	4.3		4.1
MASS	0:00Z	Day 1			7.9		11	12		11
MASS	12:00Z	Day 1			7.5		11	11		11
MAS-MOS	0:00Z	Day 1			0.6		3.1	2.2		2.1
MAS-MOS	12:00Z	Day 1			1.3		2.5	1.9		2.1
WRF-AWS	0:00Z	Day 1			0.9		7.3	7.6		5.6
WRF-AWS	12:00Z	Day 1			0.9		6.7	6.8		5.1
WRF-AWS-MOS	0:00Z	Day 1			1.4		3.1	2.9		2.6
WRF-AWS-MOS	12:00Z	Day 1			1.3		3.1	3.0		2.7
ARPS	0:00Z	Day 1			8.2		11	11		11
ARPS-MOS	0:00Z	Day 1			1.2		3.5	2.8		2.7
GEMS/ECMWF	0:00Z	Day 1	3.3	4.8	2.2	2.8	3.4	4.2	4.8	3.6
Persistence	0:00Z	Day 2	1.7	3.1	2.1	2.0	1.6	2.0	3.1	2.3
GEM	0:00Z	Day 2	3.4	3.6	1.6	2.5	3.3	4.0	3.6	3.1
GEM	12:00Z	Day 2	3.1	3.3	1.7	2.0	3.4	4.0	3.3	3.0
ECMWF	0:00Z	Day 2	2.3	5.4	2.5	3.0	3.2	3.7	5.4	3.6
NDFD	0:00Z	Day 2	3.1	3.8	2.0	1.2	3.0	2.3	3.8	2.7
NDFD	12:00Z	Day 2	3.1	3.9	1.8	0.9	3.4	2.7	3.9	2.8
WRF-ASRC		Day 2	3.0				5.0	3.9		4.0
MASS	0:00Z	Day 2			7.6		12	11		11
MASS	12:00Z	Day 2			7.4		12	10		11
MAS-MOS	0:00Z	Day 2			1.0		3.1	3.3		2.7
MAS-MOS	12:00Z	Day 2			1.1		2.5	2.9		2.3
WRF-AWS	0:00Z	Day 2			0.9		6.5	6.7		5.1
WRF-AWS	12:00Z	Day 2			1.0		6.1	6.6		4.9
WRF-AWS-MOS	0:00Z	Day 2			1.5		2.9	3.1		2.7
WRF-AWS-MOS	12:00Z	Day 2			1.3		2.9	3.2		2.6
ARPS	0:00Z	Day 2			8.1		10	11		10
ARPS-MOS	0:00Z	Day 2			1.1		3.3	2.9		2.6
GEMS/ECMWF	0:00Z	Day 2	3.1	4.4	2.2	2.7	3.5	4.0	4.4	3.5
Persistence	0:00Z	Day 3	1.8	3.1	2.1	2.0	1.6	2.1	3.1	2.2
ECMWF	0:00Z	Day 3	2.2	5.3	2.3	3.0	3.2	3.4	5.3	3.5
NDFD	0:00Z	Day 3	3.3	4.2	2.6	1.6	4.0	3.2	4.2	3.3
NDFD	12:00Z	Day 3	3.6	4.2	2.1	1.2	3.8	3.0	4.2	3.2
Persistence	0:00Z	Day 4	1.8	3.1	2.1	1.9	1.7	2.1	3.1	2.2
NDFD	0:00Z	Day 4	3.7	4.5	2.4	2.4	4.2	3.1	4.5	3.5
NDFD	12:00Z	Day 4	3.8	4.5	2.5	2.0	4.3	3.2	4.5	3.5
Persistence	0:00Z	Day 5	1.8	3.1	2.1	2.0	1.7	2.1	3.1	2.3
NDFD	0:00Z	Day 5	4.1	4.7	2.0	2.9	4.7	3.6	4.7	3.8
NDFD	12:00Z	Day 5	4.2	4.7	2.3	2.3	4.5	3.2	4.7	3.7
Persistence	0:00Z	Day 6	1.8	3.1	2.1	2.0	1.7	2.1	3.1	2.2
NDFD	0:00Z	Day 6	4.6	4.9	1.7	3.0	4.8	4.4	4.9	4.0
NDFD	12:00Z	Day 6	4.8	4.9	2.0	2.8	5.0	3.8	4.9	4.0
Persistence	0:00Z	Day 7	1.8	3.1	2.1	2.0	1.6	2.1	3.1	2.3
NDFD	0:00Z	Day 7	5.1	4.9	1.8	2.8	5.3	4.5	4.9	4.2
NDFD	12:00Z	Day 7	5.1	4.9	2.0	2.9	5.2	3.9	4.9	4.1

conditions. Additional detailed studies comparing, e.g., the performance of mesoscale models as a function of the boundary conditions from different global models, are required to confirm this assertion.

For most of the mesoscale models a version with and without MOS training was available. The MOS versions of the mesoscale models, of course, do very well in terms

of MBE (Fig. 9) since they are, in effect, calibrated in real time with ground measurements. But also with respect to RMSE and MAE a large improvement is found in comparison to the original forecasts. After the application of MOS the investigated mesoscale models showed a similar performance. However, when looking at the original mesoscale model forecasts without the statistical training, WRF

Table 21  
KS \* 100 CANADA.

KSI * 100			Egbert	Bratt's Lake	Varenes	Composite
Mean GHI ( $W m^{-2}$ )			320	306	306	311
Reference satellite model						
Persistence	0:00Z	Day 1	3.2	3.4	3.6	3.3
GEM	0:00Z	Day 1	2.6	1.7	3.1	2.4
ECMWF	0:00Z	Day 1	2.1	1.7	2.2	1.9
WRF-ASRC <sup>a</sup>	0:00Z	Day 1	1.5		0.7	1.0
GEM/ECMWF/WRF-ASRC <sup>a</sup>	0:00Z	Day 1	1.9		2.2	1.7
GEM/ECMWF	0:00Z	Day 1	2.4	1.8	2.8	2.2
Persistence	0:00Z	Day 2	3.2	3.4	3.5	3.3
GEM	0:00Z	Day 2	2.5	1.8	2.8	2.3
ECMWF	0:00Z	Day 2	1.6	2.1	2.1	1.7
WRF-ASRC <sup>a</sup>	0:00Z	Day 2	0.5		1.7	1.1
GEM/ECMWF/WRF <sup>a</sup>	0:00Z	Day 2	1.9		2.3	1.8
GEM/ECMWF	0:00Z	Day 2	2.3	2.0	2.7	2.2

<sup>a</sup> The WRF model was only run on a 2 month data subset and results were prorated using the other models as a template.

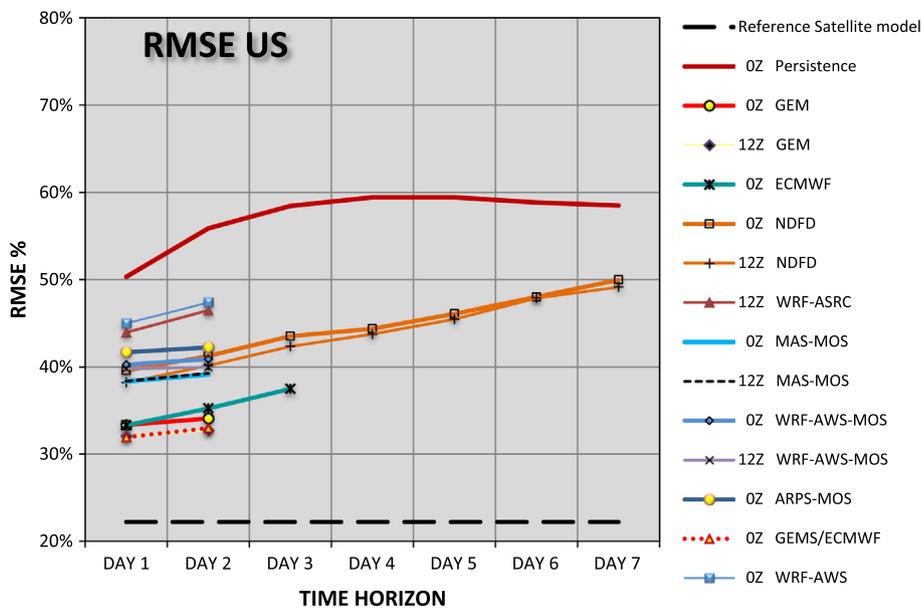


Fig. 1. Composite RMSE as a function of prediction time horizon – United States.

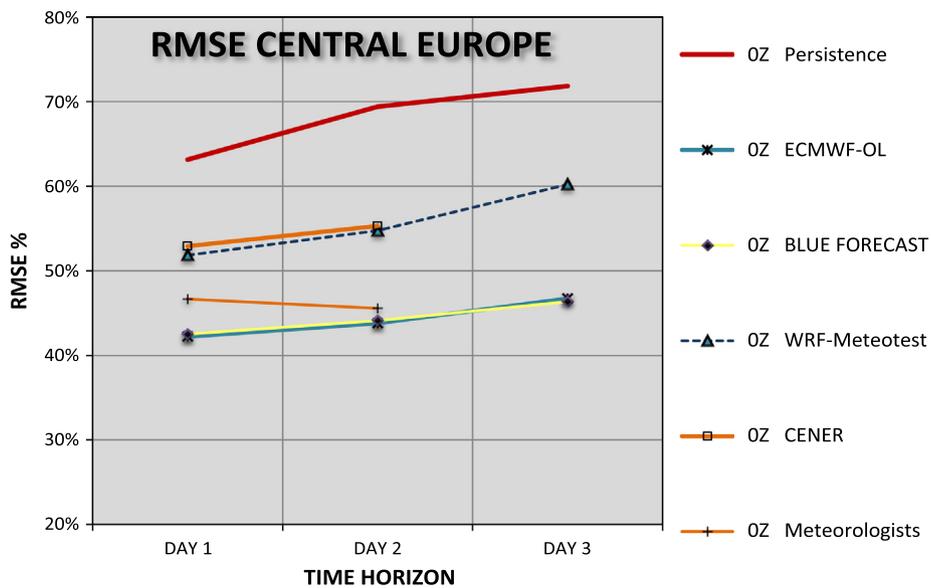


Fig. 2. Composite RMSE as a function of prediction time horizon, Central Europe.

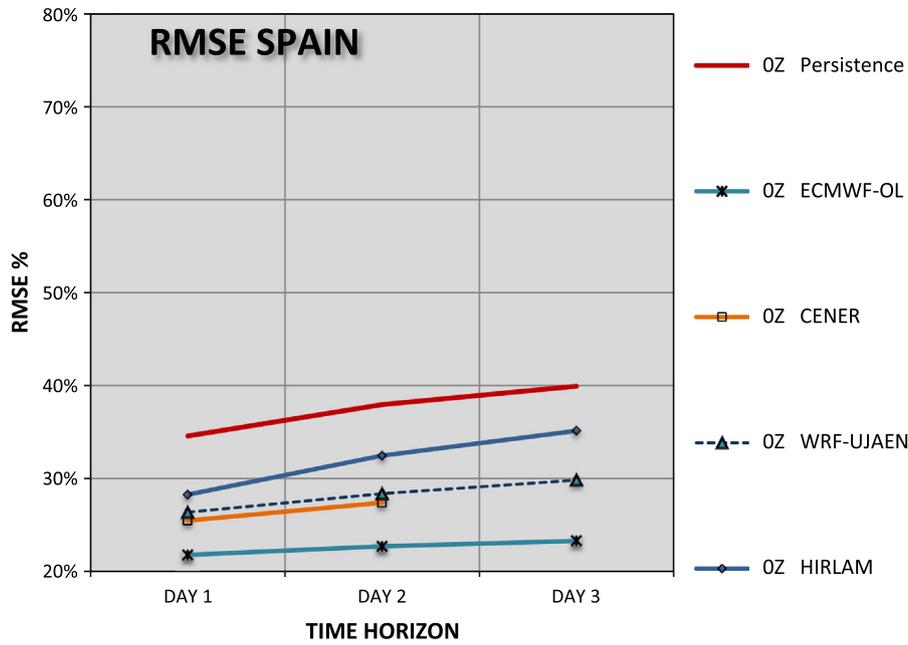


Fig. 3. Composite RMSE as a function of prediction time horizon, Spain.

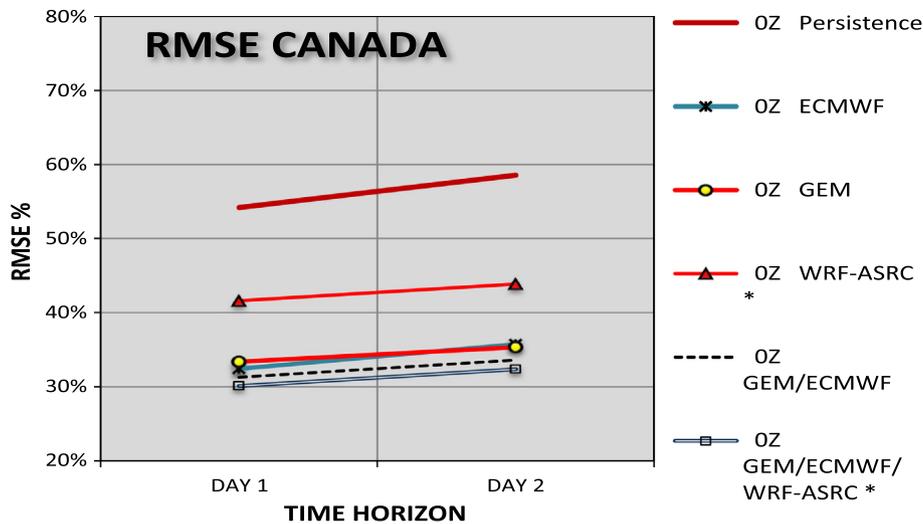


Fig. 4. Composite RMSE as a function of prediction time horizon, Canada.

clearly performs better than the other two models (MASS; ARPS) is found (see Tables 8, 12 and 16).

A comparison of the 00:00Z and 12:00Z runs (Figs. 1 and 5) shows slight advantages for the later 12:00Z runs for both RMSE and MAE (Figs. 1 and 5) as expected.

Almost all forecast models considered here outperform the persistence forecasts in terms of RMSE (Fig. 1) and MAE (Fig. 5), thus passing the basic test that confirms the skill of these forecasts with respect to trivial models. Exceptions are some pre-MOS models in the US evaluation (MASS and AEPS, see Tables 8, 12 and 16, not included in Figs. 1, 5 and 9). RMSEs and MAEs for persistence forecasts are significantly larger for Day 2 than for Day 1,

while for the other forecast models the increase in these error metrics is fairly modest.

There is a considerable variation of accuracy in terms of RMSE and MAE for the different sites and climates in the US (Tables 8 and 12), where in the following only models available for all sites are considered in the discussion. For an arid climate (Desert Rock, US) with many sunny days, relative RMSEs in the range of 20–25% for Day 1 forecasts, are considerably smaller than for the other sites for all investigated models, where the RSME values exceed 30%. Largest Day 1 RMSE values between 38% and 48% are found for Penn state with the lowest mean irradiance. Persistence shows a similar trend ranging from 29% for

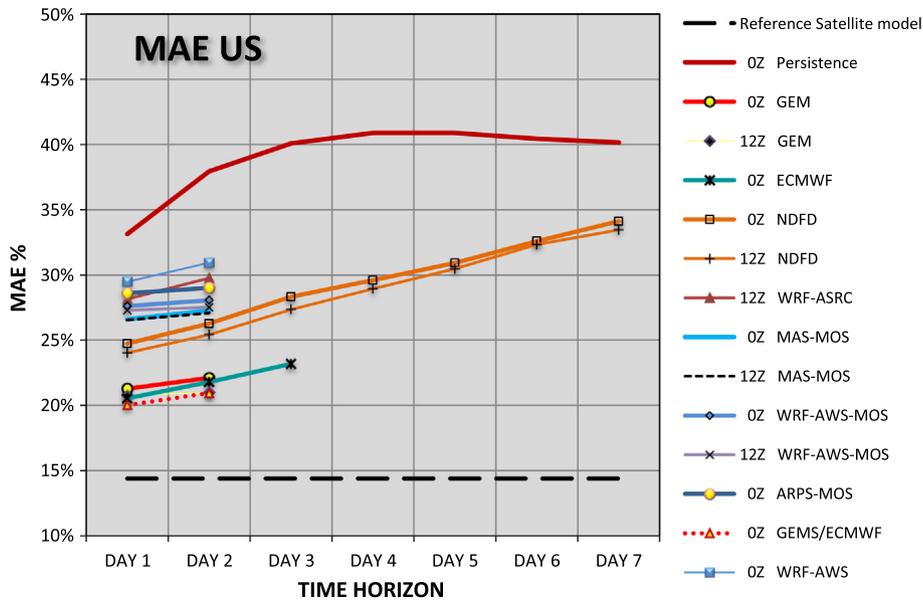


Fig. 5. Composite MAE as a function of prediction time horizon, USA.

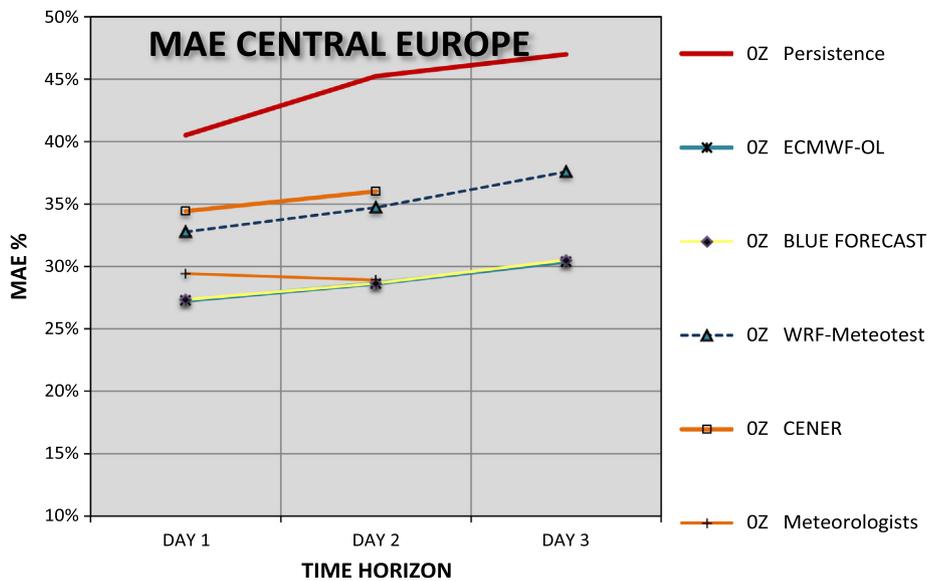


Fig. 6. Composite MAE as a function of prediction time horizon, Central Europe.

Desert Rock to 65% for Penn State. Forecasts' skill with respect persistence measured by the MSE skill score is lower for Desert Rock (e.g. for Day 1 forecasts: MSE skill score of 0.52 for GEM0Z and 0.37 for NDFD0Z) than for Penn State (for Day 1 forecasts: MSE skill score of 0.65 for GEM0Z and 0.52 for NDFD0Z).

Extending the model comparison from US to Canada (Figs. 4 and 8) and Europe (Figs. 2, 3, 6, and 7), the finding that ECMWF based irradiance forecasts show a higher accuracy than irradiance forecasts with WRF and the other investigated mesoscale models is confirmed. For Canada, like for the US, the performance of the Canadian GEM

model is similar to the performance of the ECMWF model. For the Central European evaluation (Figs. 2 and 6) the GFS-based statistical method BLUE FORECAST performs similarly to the ECWMF based forecasts. Good results were also achieved with a method using cloud cover forecasts by meteorologists, as shown in the evaluations for Austria (Tables 9 and 13). Especially for local weather phenomena, such as fog or orographic effects, this approach may be advantageous (see also Traummüller and Steinmüller, 2010). However, this method is restricted to areas well-known by the experts interpreting and combining different forecast models.

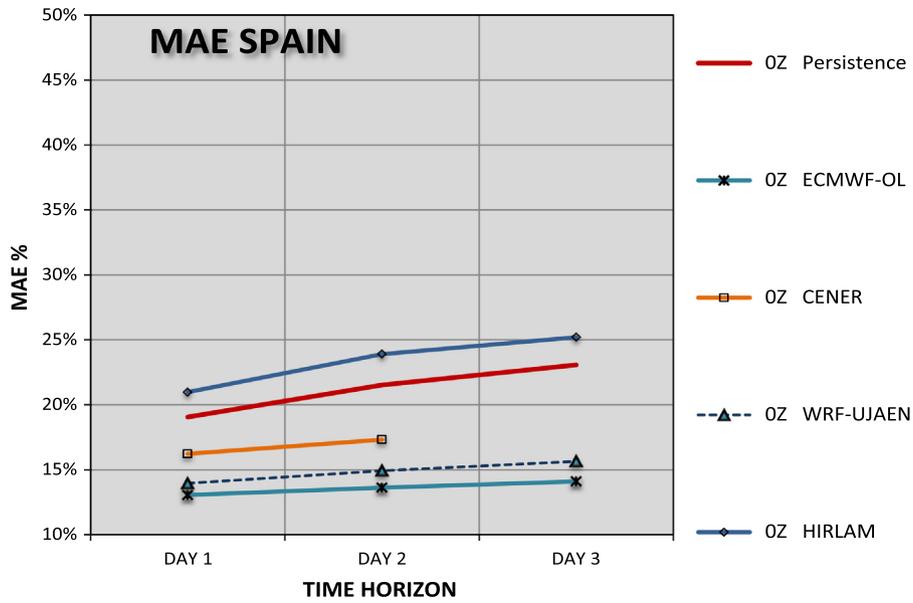


Fig. 7. Composite MAE as a function of prediction time horizon, Spain.

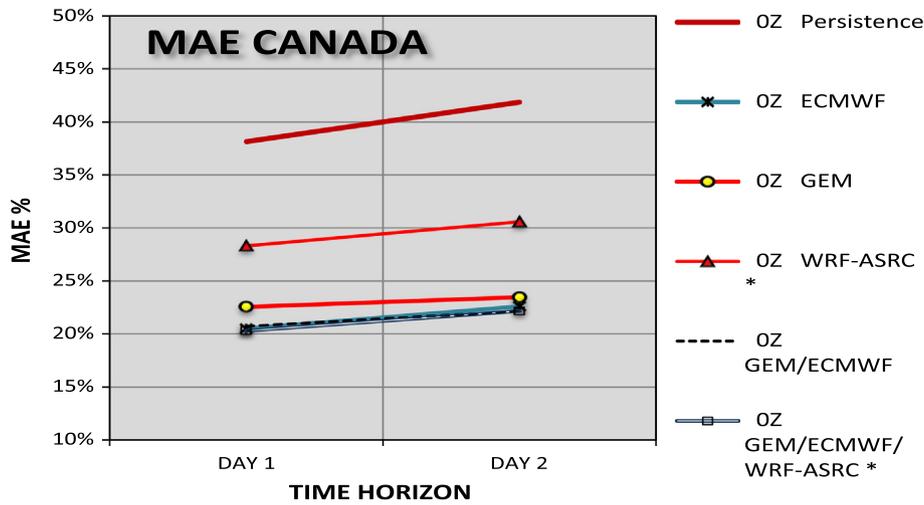


Fig. 8. Composite MAE as a function of prediction time horizon, Canada.

When looking at the inter-comparison between WRF and the other two mesoscale models in Europe (Figs. 2 and 3), it has to be considered that both WRF-meteotest and WRF – UJAEN did not include any adaptation to measured data, while the SKIRON based forecasts provided by CENER, showing a similar performance to WRF in terms of RMSE, and HIRLAM based forecasts included a statistical postprocessing. This suggests that without post-processing applied, forecasts with SKIRON and HIRLAM would show higher errors than the forecasts processed with WRF.

In addition to the evaluation of the single forecast models, a combination of some of the forecasts was investigated for the North American sites. The simple

averaging of the two best performing models – ECMWF and GEM – does slightly better than individual models in both the US and Canadian evaluations (Figs. 1 and 4). Furthermore, as shown in Fig. 4 and Table 11 for the Canadian sites, the average of the WRF, ECMWF and GEM models also outperforms the individual models in terms of RMSE and MAE, and outperforms the ECMWF/GEM combination even though the WRF model has higher RMSEs and MAEs than the other two models. Forecast errors of the different models are not fully correlated and partly compensate each other. These observations indicate that combining independently run forecast models is a worthwhile option for improving forecast performance.

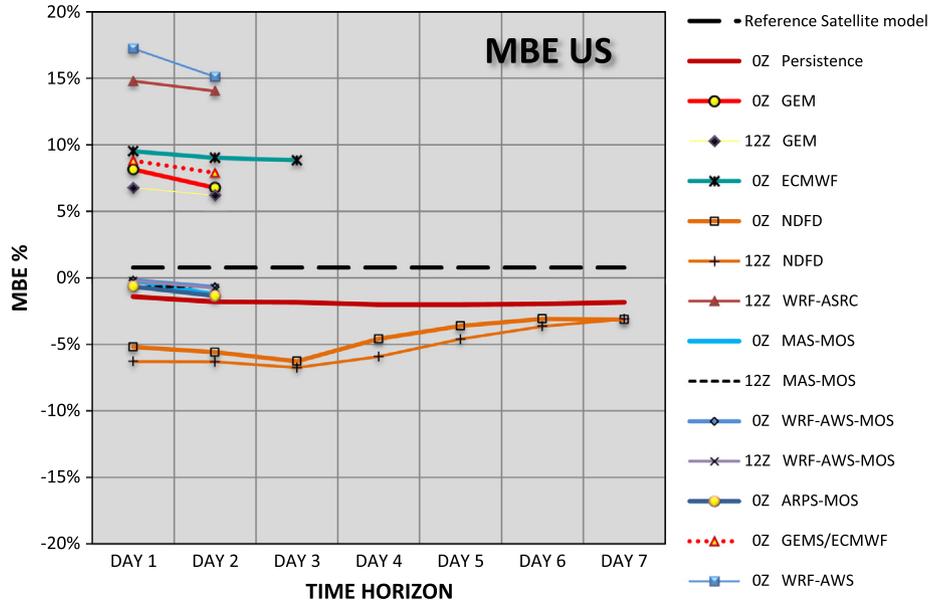


Fig. 9. Composite MBE as a function of prediction time horizon, USA.

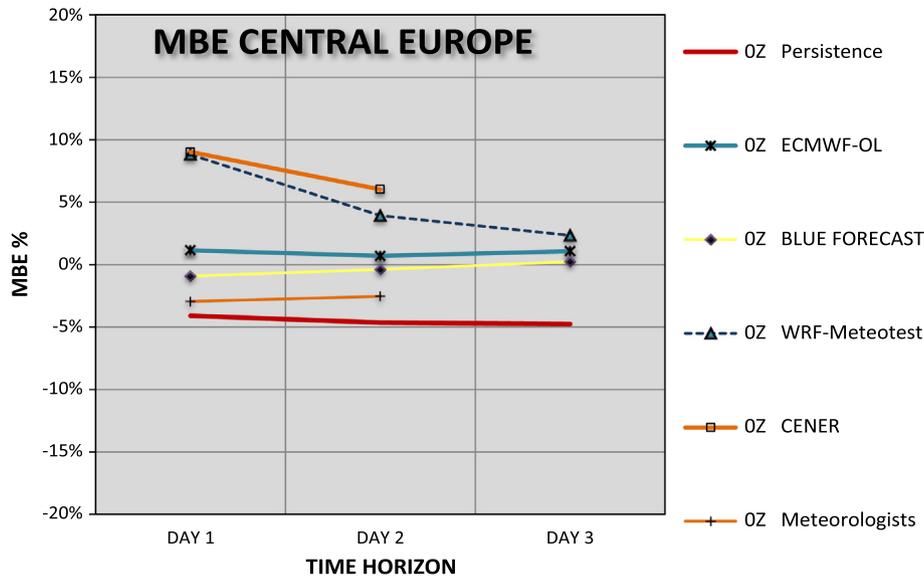


Fig. 10. Composite MBE as a function of prediction horizon, Central Europe.

With respect to the comparison of forecast performance for the different investigated regions we found lowest RMSE values in the range of 20% to 35% for the Mediterranean region Southern Spain (Fig. 3). For the Canadian stations with a humid continental climate, RMSE values between 30% and 45% are found (Fig. 4). For the US stations located in different climates (arid, sub-tropical, semi-arid, continental), RMSE values show a strong variation from station to station. All site-composite RMSE values for the US (Fig. 1) are similar to Canada. For the Central European stations with mostly continental climate and some alpine stations included average relative RMSE values range from 40% to 60% (Fig. 2).

### 5. Conclusions

We have presented three validation studies comparing NWP based irradiance multi-day forecast for the US, Canada and Europe. The focus of the comparison was on the end-use accuracy of the different models including global, multiscale and mesoscale NWP models as a basis and different postprocessing techniques to derive hourly site-specific forecasts ranging from very simple interpolation to advanced statistical postprocessing.

Two models are common to the three validation efforts – the ECMWF global model and the GFS-driven WRF mesoscale model that was run in different configurations

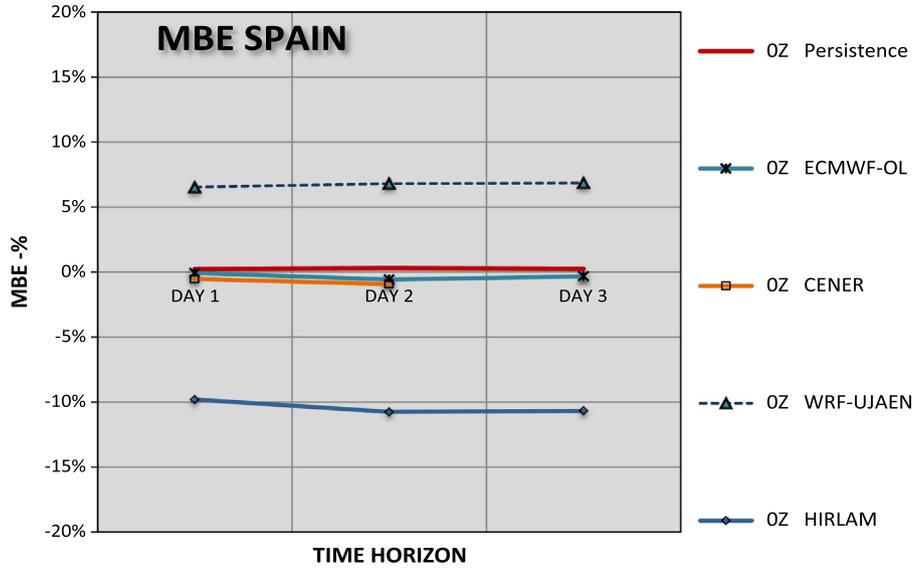


Fig. 11. Composite MBE as a function of prediction horizon, Spain.

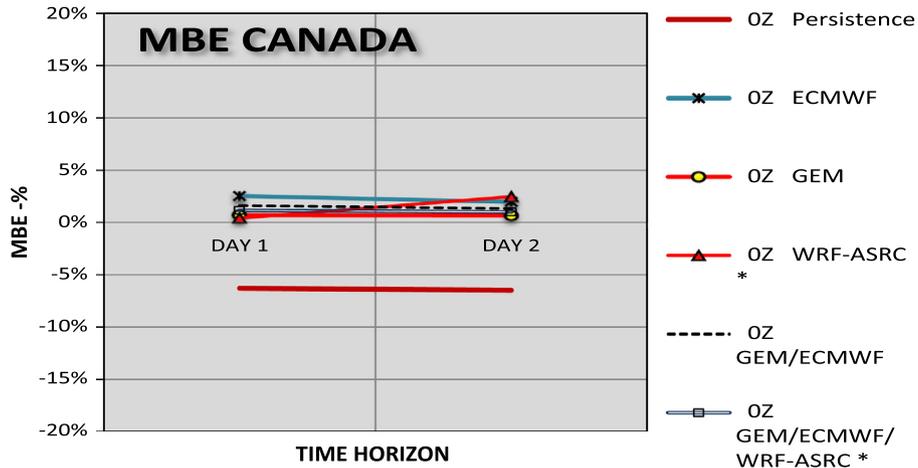


Fig. 12. Composite MBE as a function of prediction horizon, Canada.

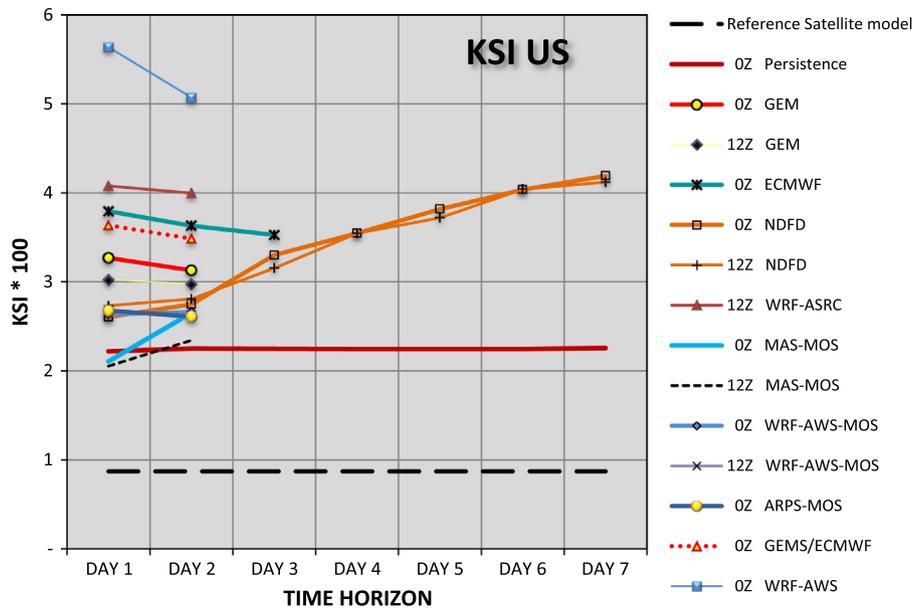


Fig. 13. Composite KSI \* 100 as a function of prediction time horizon, USA.

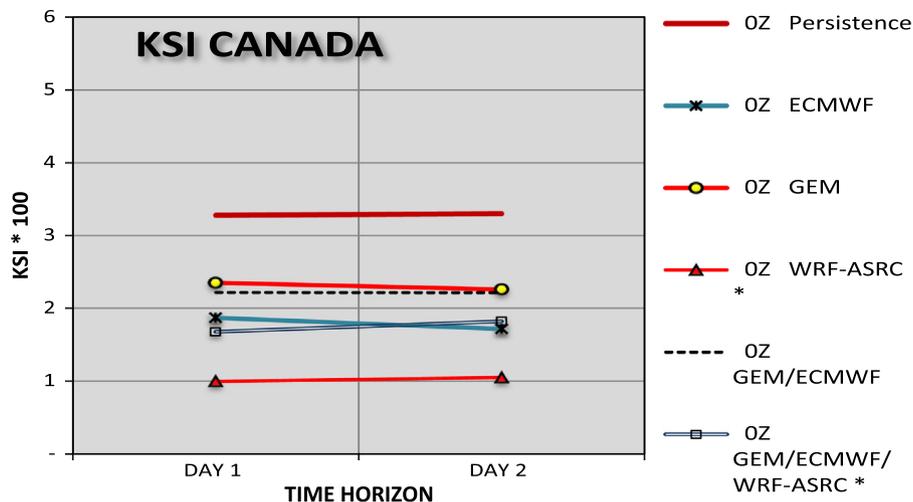


Fig. 14. Composite KSI \* 100 metric as a function of time horizon, Canada.

by various forecast providers – and allow the general observation that the global-model ECMWF forecasts perform significantly better than the GFS-based WRF-model forecasts. This trend is observed for all sites and different climatic conditions.

All other investigated meso-scale models available either for the US or for Europe showed even higher forecast errors than WRF. The potential of MOS to improve forecasts with large systematic deviations was shown for some of the mesoscale models in the US. A forecast performance similar to the ECMWF forecasts in North America was achieved with the Canadian GEM model and in Central Europe with a statistical tool based on GFS forecasts. Furthermore, it was found that simple averaging of models' output tends to perform better than individual models.

Currently, major research efforts are spent on irradiance forecasting, driven by the strong need for reliable solar power forecasts which is arising from the continuously increasing amount of solar power installed in many countries. Weather services and research groups are working on improving cloud parameterizations and radiation schemes in NWP models and investigating the use of ensemble prediction systems and rapid update models. Another focus of current research is the application of intelligent statistical methods like machine learning to improve or combine the output of NWP systems. Accordingly, evaluation and comparison of different approaches for irradiance forecasting will be continued and new comparison studies will reflect the new developments in this field.

#### Acknowledgements

This work was supported by NREL under Contract No. AGJ04029001 as a contribution to IEA SHC Task 36. Financial support for the comparisons at Canadian sites was provided by Natural Resources Canada through the ecoENERGY Technology Initiative, which is a component

of ecoACTION, the Canadian government's actions towards clean air and greenhouse gas emission reductions.

#### References

- AQFMS, 2010: Air Quality Forecast Modeling System. <<http://asrc.albany.edu/research/aqf/>>.
- Chow, C.W., Urquhart, B., Kleissl, J., Lave, M., Dominguez, A., Shields, J., Washom, B., 2011. Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. *Solar Energy* 85 (11), 2881–2893.
- Ernst, B., Oakleaf, B., Ahlstrom, M.L., Lange, M., Moehrlen, C., Lange, B., Focken, U., Rohrig, K., 2007. Predicting the wind. *Power and Energy Magazine*, IEEE 5 (6), 78–89. <http://dx.doi.org/10.1109/MPE.2007.906306>.
- Gastón, M., Lorenz, E., Lozano, S., Heinemann, D., Blanco, M., Ramírez, L., 2009. Comparison of global irradiance forecasting approaches. In: *SolarPACES Symposium*, September 15–18 in Berlin, Germany.
- GFS, 2010. Global Forecasting System. <<http://www.emc.ncep.noaa.gov/index.php?branch=GFS>>.
- Heinemann, D., Lorenz, E., Girodo, M., 2006. *Solar Irradiance Forecasting for the Management of Solar Energy Systems*, Solar 2006, Denver, CO, USA (07.07.06).
- HIRLAM, 2010: High Resolution Limited Area Model. <<http://hirlam.org/>>.
- IEA SHC Task 36, 2011. International Energy Agency, Solar Heating and Cooling Programme, Task 36 Solar Resource Knowledge Management, Subtask A, Model Benchmarking. <<http://www.iea-shc.org/task36/>>.
- Kallos, G., 1997. The regional weather forecasting system SKIRON. In: *Proceedings, Symposium on Regional Weather Prediction on Parallel Computer Environments*, 15–17 October 1997, Athens, Greece.
- Lara-Fanego, V., Ruiz-Arias, J.A., Pozo-Vázquez, D., Santos-Alamillos, F.J., Tovar-Pescador, J., 2012. Evaluation of the WRF model solar irradiance forecasts in Andalusia (southern Spain). *Solar Energy* 86 (8), 2200–2217. <http://dx.doi.org/10.1016/j.solener.2011.02.014>.
- Lorenz, E., Hurka, J., Heinemann, D., Beyer, H.G., 2009a. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. *IEEE Journal of Special Topics in Earth Observations and Remote Sensing* 2, 2–10.
- Lorenz, E., Remund, J., Müller, S.C., Traunmüller, W., Steinmaurer, G., Pozo, D., Ruiz-Arias, J.A., Lara Fanego, V., Ramirez, L., Romeo, M.G., Kurz, C., Pomares, L.M., Guerrero, C.G., 2009b. Benchmarking of different approaches to forecast solar irradiance. In: *Proceedings*

- of the 24th European Photovoltaic Solar Energy Conference, 21–25 September 2009, Hamburg.
- Mailhot, J., Bélair, S., Lefaivre, L., Bilodeau, B., Desgagné, M., Girard, C., Glazer, A., Leduc, A.M., Méthot, A., Patoine, A., Plante, A., Rahill, A., Robinson, T., Talbot, D., Tremblay, A., Vaillancourt, P.A., Zadra, A., 2006. The 15-km version of the Canadian regional forecast system. *Atmosphere – Ocean* 44, 133–149.
- Manobianco, J., Zack, J.W., Taylor, G.E., 1996. Workstation based real-time mesoscale modeling designed for weather support to operations at the Kennedy Space Center and Cape Canaveral Air Station. *Bulletin of the American Meteorological Society* 77, 653–672.
- Mathiesen, P., Kleissl, J., 2011. Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States. *Solar Energy* 85, 967–977.
- Natschläger, T., Traunmüller, W., Reingruber, K., Exner, H., 2008. Lokal optimierte Wetterprognosen zur Regelung stark umweltbeeinflussteter Systeme; SCCH, Blue Sky. In: Tagungsband Industrielles Symposium Mechatronik Automatisierung. Clusterland Oberösterreich GmbH/ Mechatronik-Cluster, pp. 281–284.
- NSRDB, 2005. <[http://rredc.nrel.gov/solar/old\\_data/nsrdb/1991-2005/](http://rredc.nrel.gov/solar/old_data/nsrdb/1991-2005/)>.
- Pedro, H., Coimbra, C., 2012. Assessment of forecasting techniques for solar power production with no exogenous inputs. *Solar Energy* 86 (2017), 2028.
- Pelland, S., Gallanis, G., Kallos, G., 2011. Solar and photovoltaic forecasting through postprocessing of the global environmental multiscale numerical weather prediction model. *Progress in Photovoltaics: Research and Applications* (November 22, 2011).
- Perez, R., Ineichen, P., Moore, K., Kmiecik, M., Chain, C., George, R., Vignola, F., 2002. A new operational satellite-to-irradiance model. *Solar Energy* 73 (5), 307–317.
- Perez, R., Kivalov, S., Schlemmer, J., Hemker Jr., K., Renné, D., Hoff, T., 2010. Validation of short and medium term operational solar radiation forecasts in the US. *Solar Energy* 84 (12), 2161–2172.
- Remund, J., Perez, R., Lorenz, E., 2008. Comparison of solar radiation forecasts for the USA. In: Proc. 23rd European Photovoltaic Solar Energy Conference, Valencia, Spain.
- Ruiz-Arias, J.A., Pozo-Vázquez, D., Sánchez-Sánchez, N., Montávez, J.P., Hayas-Barrú, A., Tovar-Pescador, J., 2008. Evaluation of two MM5-PBL parameterizations for solar radiation and temperature estimation in the South-Eastern area of the Iberian Peninsula. *Il Nuovo Cimento C* 31, 825–842.
- Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Barker, D.M., Wang, W., Powers, J.G., 2005. A Description of the Advanced Research WRF Version 2, NCAR TECHNICAL NOTE: NCAR/TN-468+STR. National Center for Atmospheric Research, Boulder, Colorado, USA.
- Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Barker, D.M., Duda, M.G., Huang, X.Y., Wang, W., Powers, J.G., 2008. A Description of the Advanced Research WRF Version 3. Mesoscale and Microscale Meteorology Division, Natl. Cent. Atmos. Res., NCAR/TN-475+STR, Boulder, Colorado.
- SolarAnywhere, 2010. <[www.solaranywhere.com](http://www.solaranywhere.com)>.
- SURFRAD, 2010. NOAA's Surface Radiation Network <<http://www.srrb.noaa.gov/surfrad/>>.
- Traunmüller, W., Steinmaurer, G., 2010. Solar Irradiance Forecasting, Benchmarking of Different Techniques and Applications of Energy Meteorology. Eurosun, Graz, Austria.
- Xue, M., Droegemeier, K.K., Wong, V., Shapiro, A., Brewster, K., Carr, F., Weber, D., Liu, Y., Wang, D.-H., 2001. The Advanced Regional Prediction System (ARPS) – a multiscale nonhydrostatic atmospheric simulation and prediction tool. Part II: model physics and applications. *Meteorology and Atmospheric Physics* 76, 134–165.